Edinburgh

Student Journal of Science

Volume 1
Issue 2

ISSN 3049-7930

October 2024

"Any sufficiently advanced technology is indistinguishable from magic"

Arthur C. Clark

Edinburgh Student Journal of Science
Volume 1
Issue 2
October 2024
ISSN 3049-7930

Published by the University of Edinburgh
Edited and Produced by Jack L Smith

journals.ed.ac.uk/esjs

Edinburgh Student Journal of Science

Volume 1, Issue 2, October 2024

Board of Editors

Editorial Team for this Issue

Cover Illustration

The illustration on the journal cover of each issue is created based on one or more articles from within the issue which the editorial team found to be particularly well-written and engaging.

This issue features a redesigned cover to reflect the journal's improved operating structure, and to create a clean and consistent design for future issues, consolidating the journal's intended plans and impact. The image on this cover was created with the help of Microsoft Designer, however, in the future we hope to feature artwork from students; if you are interested in designing a future cover please email us!

Submitting to the Journal

We welcome submissions from later-year undergraduate, Master's, and recently graduated students within the College of Science and Engineering at the University of Edinburgh. Submissions should be no longer than 1500 words and should summarise work from academic/independent research projects, internships, or summer projects. Full details about submissions criteria and guidelines can be found on our website.

The Edinburgh Student Journal of Science is a student-run peer-reviewed journal for students studying science at the University of Edinburgh. It is an opportunity for students to publish formal summaries of their academic or independent research in a professional-style journal and get experience with the publishing and peer-review processes.

The journal is issued every four months to coincide with the end of both academic semesters for academic projects, and the end of the summer break for summer projects or internships. We are continuously open for submissions, and a typical timeline of the submission process can be found online.

The next issue of the Edinburgh Student Journal of Science is expected to be published in February and for more information, templates, and guidance on submitting an article, please visit our website: journals.ed.ac.uk/esjs.

Contents

Physics and Astronomy

Biological Sciences

Physics and Astronomy

Investigating the Inharmonicity of Piano Strings

Eleanor Roy*[1]

[1] School of Physics and Astronomy, University of Edinburgh

Open Access

Received
11 Sep 2024

Revised
29 Sep 2024

Accepted
03 Oct 2024

Published
24 Oct 2024

Abstract

In classical physics, ideal string vibrations are modelled using the harmonic series, yet real mode frequencies deviate from the integer set of harmonics due to string stiffness, known as inharmonicity. This work investigates this phenomenon in five piano strings on a Yamaha GB1 Grand Piano. Using the audio software *Audacity* for spectral analysis, the inharmonicity coefficient is determined experimentally and theoretically based on string properties. These values are comparable with previous research, offering insights into tuning techniques to minimise dissonant inharmonicity. Possible innovations from inharmonicity research are explored, with suggestions such as the temperature-dependent self-tuning systems for acoustic pianos.

DOI: 10.2218/esjs.9815 **ISSN** 3049-7930

Introduction

Rationale

Piano strings are tuned to account for inharmonicity, a deviation from harmonic frequencies caused by string stiffness (Heetveld *et al.* 1984). Understanding these patterns informs tuning strategies to improve sound quality. Advancing inharmonicity research could further modernise acoustic pianos, revolutionising music and science.

Mode Frequencies & Inharmonicity

The fundamental frequency is the natural frequency at which an object vibrates when struck or plucked, producing the lowest and most dominant pitch in the sound it generates. It is the lowest frequency of a vibrating object, with harmonics being exact integer multiples of this frequency, forming the harmonic series. Piano strings, however, exhibit resonant frequencies that deviate from these harmonics, known as in-harmonic partials (Nave 2020). This deviation occurs because the theory of harmonic frequencies assumes ideal strings with no stiffness, which is never the case in the real world. The degree of this deviation is called inharmonicity (Cohen 1984).

Inharmonicity, resulting from string stiffness, is crucial in tuning. For example, the sixteenth harmonic, equivalent to four octaves above another note, is approximately a semitone higher than an ideal string's harmonic due to the progressive sharpening of mode frequencies. This sharpening means that the note has a higher frequency than the original. The effect is influenced by the string's diameter-to-length ratio

*Student Author

(Shankland *et al.* 1939; Berg 2020). A higher ratio leads to greater sharpening because increased stiffness from a thicker string causes the vibrational frequencies to deviate further from their ideal harmonic values.

Controversially, a small amount of inharmonicity may be desirable for the piano's distinctive sound, as it adds warmth and richness to the tone, contributing to the instrument's characteristic timbre. However, excessive inharmonicity can degrade sound quality by making the pitch unclear and less harmonious (Campbell *et al.* 1994).

Theory

Harvey Fletcher's research on piano string inharmonicity has been fundamental in acoustics (Fletcher 1964; Pierce 1983). Fletcher's equations remain highly influential and widely applied in the field today.

Fletcher derived Equation 1 for the frequency, f_n, of the n^{th} mode of a string by considering the equation of motion of a circular string fixed at either end, accounting for the tension and elastic stiffness causing a restoring force, and the principles of energy conservation in the bending and stretching of the string:

$$f_n = nF(1+Bn^2)^{1/2} \qquad \text{for} \quad n = 1, 2, 3, ... \tag{1}$$

where F and B are two constants that can be obtained from an accurate measurement of the frequencies of any two modes. In particular, B is the inharmonicity coefficient with units of $\mathrm{m}^{-2}\mathrm{s}^{2}$.

In this paper, Equation 1 will be used to calculate the experimental inharmonicity coefficient. Although specific literature values are unavailable, the theoretical and experimental data from Fletcher (1964) on an upright Hamilton piano will provide a critical reference for comparison.

To calculate the theoretical inharmonicity coefficient, B, we use the equations from Fletcher (1964) for solid-steel and copper-wound steel strings, considering relevant physical properties. The analysis includes three copper-wound steel strings (named C1–C3) and two solid-steel strings (C4 & C5) which are discussed in more detail later. For steel strings, the formula is:

$$B = 3.95 \times 10^{10}\left(\frac{d^2}{l^4 {f_0}^2}\right) \tag{2}$$

where d is the diameter, l is the length, and f_0 represents fundamental frequency for the first harmonic. For copper-wound steel strings, a similar approach applies. Assuming only the steel core contributes to the inharmonicity, the formula becomes:

$$B = 4.6 \times 10^{10}\left(\frac{d^4}{D^2 l^4 {f_0}^2}\right) \tag{3}$$

with d the inner steel diameter, and D the outer diameter including the copper winding.

In the real world, f_0 and F can be approximated as nearly equal at the fundamental mode because the effects of inharmonicity are minimal at the lowest frequency — the impact of string stiffness is relatively small, meaning the actual frequency, F, closely matches the theoretical harmonic frequency, f_0. C1–C5 have relatively low fundamental frequencies (inharmonicity is most pronounced at higher modes), meaning the difference between f_0 and F is negligible. These calculations estimate inharmonicity for both string types, enabling comparison with experimental values.

Methods

On a piano, C1 to C5 correspond to C notes across different octaves: C1 is the lowest, in the deep bass range; C2 and C3 are higher in the bass; C4 is Middle C, the central reference; and C5 is one octave above Middle C in the mid-treble range. C1–C3 are copper-wound strings with an inner diameter, d, outer diameter, D, and length, l, while C4 & C5 are steel-only strings, having only a diameter, d, and length, l.

After a preliminary analysis of strings C1–C8, the first five strings (C1–C5) were selected for their clearly identifiable modes, with frequency peaks most easily visualised using *Audacity* software for accurate frequency determination.

The inharmonicity coefficient, B, was determined using two complementary methods: experimental and theoretical. The experimental method involved measuring the frequencies, f_n, across multiple modes and calculating B via Equation 1. This was done by plucking the strings, recording frequencies with *Audacity*, and performing Fourier analysis to decompose the recorded sound signal into its constituent frequencies, allowing for the identification of fundamental and overtone frequencies. Based on these data points, the inharmonicity coefficient is calculated for both solid-steel and copper-wound strings.

In the theoretical approach, physical properties (diameter, d, length, l, and outer diameter, D, for copper-wound strings C1–C3) were measured, and Equations 2 & 3 used to calculate B. However, even in this method, the fundamental frequency, f_0, must be experimentally obtained. Thus, the theoretical approach, while relying on physical measurements, still depends on f_0 for the calculation.

Equation 1, derived by Fletcher, relates the frequencies of different vibration modes to the string's physical properties to determine B. Measurements of string length, diameter, and frequency were used, and linear regression was applied to a transformed version of this equation to calculate B.

Results & Discussion

As seen in Figure 1, the experimental values align closely with the theoretical values for most strings, exhibiting minimal deviation. For string C5, the discrepancy is slightly higher, likely due to measurement uncertainties, factors affecting higher frequencies, and increased inharmonicity from greater stiffness and tension in shorter, higher-pitched strings. Additionally, Equation 2 shows that the terms are to the power of 4, meaning any small measurement discrepancies in length or frequency can result in significant errors in the calculation.

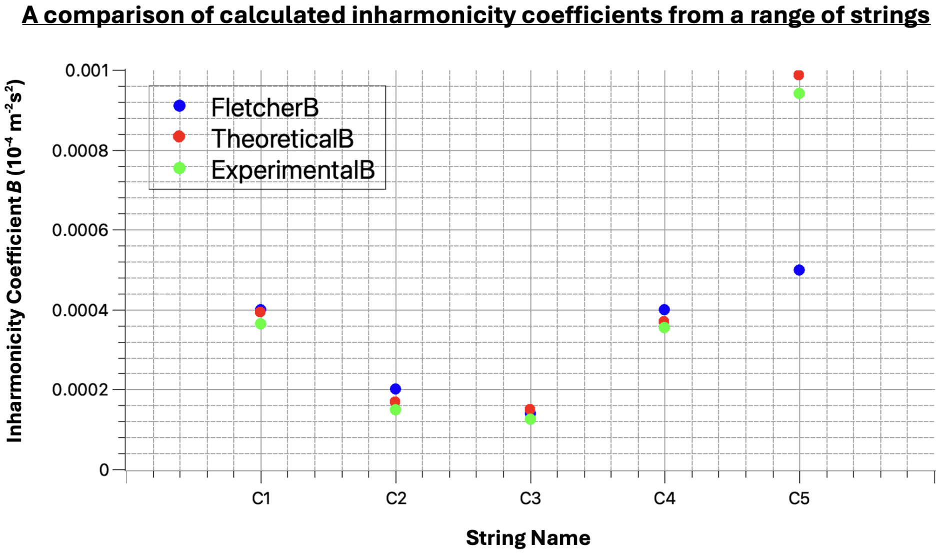


Figure 1: Graph showing experimental and theoretical inharmonicity coefficients as well as those taken from Fletcher (1964). *NB:* error bars are too small to be seen in the image.

These findings indicate a trend of increasing inharmonicity for steel-only strings (C4 & C5) as string numbers rise, attributed to progressive sharpening of partials with increasing mode numbers. Conversely, inharmonicity decreases for copper-wound strings (C1–C3), due to the decreasing thickness of the copper winding, forming the characteristic shape shown in Figure 1.

The graph provides a comparison with Fletcher's values, serving as a key benchmark. Despite deriving from a different piano over 60 years ago, the experimental data aligns well with these results. Variations are expected due to differences in piano construction, string materials, and lengths. Nonetheless, the trend of higher inharmonicity in shorter and thicker strings remains consistent. As string number increases, string length decreases, leading to greater stiffness in higher-numbered strings, while the lower, thicker strings (due to the copper-winding) exhibit increased inharmonicity due to their larger diameter.

The deviation observed in Fletcher's predicted value for C5, which is lower than this result, suggests differences in the energy distribution along the string during vibration, especially at higher frequencies. As strings vibrate, energy distributes across multiple modes, but in shorter, higher-pitched strings like C5, increased stiffness and tension confine energy to the string's boundaries. This confinement leads to more pronounced inharmonicity, as energy transfer becomes less efficient, resulting in a frequency shift. Additionally, variations in string construction, such as winding density, material properties, or tension adjustments, can alter vibrational responses, affecting inharmonicity measurements. Despite these differences, the overall consistency with Fletcher's findings reinforces the reliability of this study, even when accounting for variations in instruments and experimental conditions.

Conclusions

This work investigated the inharmonicity coefficient of five piano strings using both theoretical and experimental methods. The results showed consistency with each other and aligned with Fletcher's findings. A key observation was that higher strings exhibit increased inharmonicity due to the progressive sharpening of partials with increasing mode number, influenced by the string's diameter-to-length ratio (Shankland *et al.* 1939), which rises as string length decreases.

Historically, inharmonicity was overcome by tuning octaves by ear. Today, advancements in technology and a deeper understanding of inharmonicity have led to electronic tuning devices, paving the way for self-tuning acoustic pianos. With further development, temperature-dependent self-tuning systems could be globally commercialised, enabling modern mechanisms to correct harmonic deviations and establishing self-tuning acoustic pianos as instruments of the future.

Acknowledgements

I sincerely thank Murray Campbell for his exceptional supervision and support throughout this project, especially during the COVID-19 lockdown, which required remote work. I am grateful for his guidance and for enabling me to conduct the experiment on my own piano while providing invaluable mentorship.

References

Berg, R. E. 'Sound' *Accessed 4th Feb 2021* (2020)
Campbell, M. and Greated, C. 'The Musician's Guide to Acoustics' (Oxford University Press; 1994)
Cohen, E. A. 'Some Effects of Inharmonic Partials on Interval Perception' Music Perception **1** 3 (1984)
Fletcher, H. 'Normal Vibration Frequencies of a Stiff Piano String' The Journal of the Acoustical Society of America **36** 1 (1964)
Heetveld, V. and Rasch, R. A. 'String Inharmonicity and Piano Tuning' The Journal of the Acoustical Society of America **76** S1 (1984)
Nave, R. 'Fundamental and Harmonics' *Accessed 2nd Feb 2021* (2020)
Pierce, J. R. 'The Science of Musical Sound' (Scientific American Library; 1983)
Shankland, R. S. and Coltman, J. W. 'The Departure of the Overtones of a Vibrating Wire From a True Harmonic Series' The Journal of the Acoustical Society of America **10** 3 (1939)

Probing the Disk Kinematics of M31 with DESI

Struan Stevenson*[1]

[1] School of Physics and Astronomy, University of Edinburgh

Open Access

Received
12 Sep 2024

Revised
30 Sep 2024

Accepted
02 Oct 2024

Published
24 Oct 2024

Abstract

M31's stellar rotation curve and disk line-of-sight dispersion is presented using kinematic data from a novel DESI M31 survey in combination with similar datasets. After removing the foreground, the disk region is isolated and the velocity field deprojected. A flat ring model is fit to the velocity field of M31 out to 30 kpc using a maximum-likelihood routine. The stellar rotation curve is found to flatten to $\sim$ 220 km/s, and a disk line-of-sight velocity dispersion of $\sim$ 60 km/s is derived. The results extend M31's rotation curve to 30 kpc and support the picture in which M31 has thicker disk than the Milky Way.

DOI: 10.2218/esjs.9976 **ISSN** 3049-7930

Introduction

The internal motions of stars within galaxies are challenging to measure, but offer significant insight into a galaxy's evolution, matter distribution, and immigration history. The typical structures which form in a spiral galaxy include a bulge, disk, and stellar halo. However, external influences, such as immigrating satellites, can leave a unique distribution of stellar substructure in the form of streams and shells. Mapping out the overall velocity field of a galaxy provides a powerful tool for investigating not only its physical properties but also its accretion history. Due to its proximity, our nearest galactic neighbour M31 provides a detailed view of resolved galaxy dynamics in-action. In contrast to the Milky Way, M31 can be captured from an outsider perspective, allowing its disk to be fully observed while unobscured by a central bulge or dust, and hence more intuitively analysed. Surveys such as the Pan-Andromeda Archaeological Survey (McConnachie *et al.* 2009) have revealed a complex map of substructure within the halo of M31, revealing the pathways of chaotic mergers (McConnachie *et al.* 2018).

The rotation curve of M31 has been the subject of extensive investigation (Rubin *et al.* 1970; Roberts *et al.* 1975; Carignan *et al.* 2006), often probed via HI, Hα, and C0 emission lines, and continues to be a subject of interest (Chemin *et al.* 2009; Sofue 2015). Robust tilted ring models have been developed (Begeman 1989; Corbelli *et al.* 2010) which translate line-of-sight velocities into kinematic properties of the host galaxy, and continue to be advanced (Di Teodoro *et al.* 2015). One such kinematic property, the rotation curve, can be used to determine the mass profile of M31, including a constraint on the dark matter distribution (Zhang *et al.* 2024).

Shortly after the debut of the Dark Energy Spectroscopic Instrument (DESI), a powerful wide-field spectroscopic surveyor, Dey *et al.* (2023) completed a short survey campaign on the region around M31 and found evidence for a wealth of substructure in the halo of M31. Using the kinematics of substructures such as the Great Stellar Stream (a prominent stellar stream feature in the halo of M31), the immigration history and enclosed mass of M31 was inferred. However, there remains limited investigation into the kinematics of the disk within this DESI dataset despite thousands of stars captured within this region. Thus, the primary aim of this paper is to explore the DESI dataset further to develop a better understanding of M31's disk kinematics.

In this report, DESI data of M31 obtained by Dey *et al.* (2023) is decontaminated of Milky Way foreground stars and merged with kinematic datasets of planetary nebulae (Merrett *et al.* 2006) and globular clusters (Caldwell *et al.* 2016). The region around the disk is isolated and the velocity field deprojected. A flat ring model is fit, using a Gaussian mixture likelihood distribution in a maximum-likelihood routine, returning the stellar rotation curve and line-of-sight velocity dispersion profile.

*Student Author

Data & Analysis

Data

This paper uses data published in Dey *et al.* (2023), consisting of a short survey of M31 performed by DESI. Although short, this survey captured 11,554 astronomical targets in the region of M31, including 10,414 stars belonging to either M31 or the Milky Way foreground, representing one of the most comprehensive kinematic investigations of M31 to date. While Dey *et al.* (2023) probed the kinematics of the stellar halo, the kinematics of M31's disk remained unexamined. Members of the dataset belonging to M31 were isolated under the criteria: RVS_WARN = 0 and $\log(g) < 3.8$. The former criterion flags any targets which were poorly fitted in the RVS pipeline used to assign stellar parameters to DESI spectra (Koposov *et al.* 2011; Koposov 2019), and the latter removes most of the Milky Way foreground. Milky Way foreground stars consist of disk dwarfs, which have a strong surface gravity, $\log(g)$, relative to the mostly red giant M31 targets. The exact boundary was inferred via the ratio of two fitted Gaussians on a histogram over the stellar surface gravity.

The DESI focal plane spans a diameter of 3.2° (Cooper *et al.* 2023), which is able to capture M31 in its entirety with only a few exposures. This focal plane is evenly populated with 5000 spectral fibres, which naturally leads to an underdensity in the inner disk region of M31 due to its small area on the sky. The region was therefore strengthened by the addition of two similar but smaller datasets: one of planetary nebulae surveyed by Merrett *et al.* (2006) and another of globular clusters surveyed by Caldwell *et al.* (2016). To analyse galactic kinematics in the coordinate system of M31, the dataset was tangentially projected assuming M31 is centred on (RA, DEC) = $10^\circ.6847$, $41^\circ.26875$ (Dey *et al.* 2023). The line-of-sight velocity field was then made relative to the centre of M31 assuming it moves with a recessional velocity of -300 km/s (Dey *et al.* 2023), where all velocities were first converted to the Galactic Standard of Rest (GSR) frame.

Flat Ring Model

The main method of this paper consists of a 'flat ring model', which assumes the galactic disk is perfectly flat and investigates the kinematics of concentric rings at increasing radii. The velocity field was first deprojected assuming a constant tilt of 77° from the face-on position (Dey *et al.* 2023). After specifying the number of rings, an algorithm determines the position and width of each ring such that they are equally populated by stars. For each ring, a sinusoid was fit to the line-of-sight velocities over azimuthal angle in a maximum-likelihood routine, optimised using the Nelder-Mead method. To account for the presence of substructure contaminants, which do not follow the sinusoidal profile of the disk, an outlier likelihood distribution was chosen (Hogg *et al.* 2010). The use of a bimodal Gaussian mixture to represent a likelihood distribution containing both a sinusoid and a background of outlier points allowed substructure to be accounted for when fitting to the disk (at the cost of additional free parameters). In addition to optimising sinusoidal and outlier parameters, the Gaussian width of the sinusoidal distribution was allowed to vary and optimised, representing the line-of-sight velocity dispersion in each ring.

Results and Discussion

The results of the flat ring model are shown in Figure 1 for a set of 25 equally populated rings up to 30 kpc from the centre of M31. The top panel of Figure 1 shows the stellar rotation curve which adopts the expected profile, flattening to $\sim$ 220 km/s, consistent with prior studies (Rubin *et al.* 1970; Sofue 2015; Zhang *et al.* 2024). Also plotted is both a HI rotation curve (dashed) and a potential model computed by the GALA software package (solid). The separation between the stellar and gaseous/model curves can be attributed to the effects of asymmetric drift, a naturally occurring phenomenon arising from collisionless nature of stars (Binney *et al.* 2008). Further investigation would include a calculation of the asymmetric drift and derivation of M31's mass distribution.

The central panel of Figure 1 shows the line-of-sight velocity dispersion in the disk over radii. The initial spike at low radii can be attributed to the bulge which is not rotationally supported. Towards larger radii the dispersion decreases as the disk population becomes dominant, but plateaus to $\sim$ 60 km/s. These results contribute further evidence to the picture in which M31 has a thicker disk than the Milky Way, the thin disk of which has a dispersion of $\sim$ 20 km/s (Vieira *et al.* 2022). If we adopt the theory that immigration events are a main contributor in thickening a galactic disk (Abadi *et al.* 2003), this

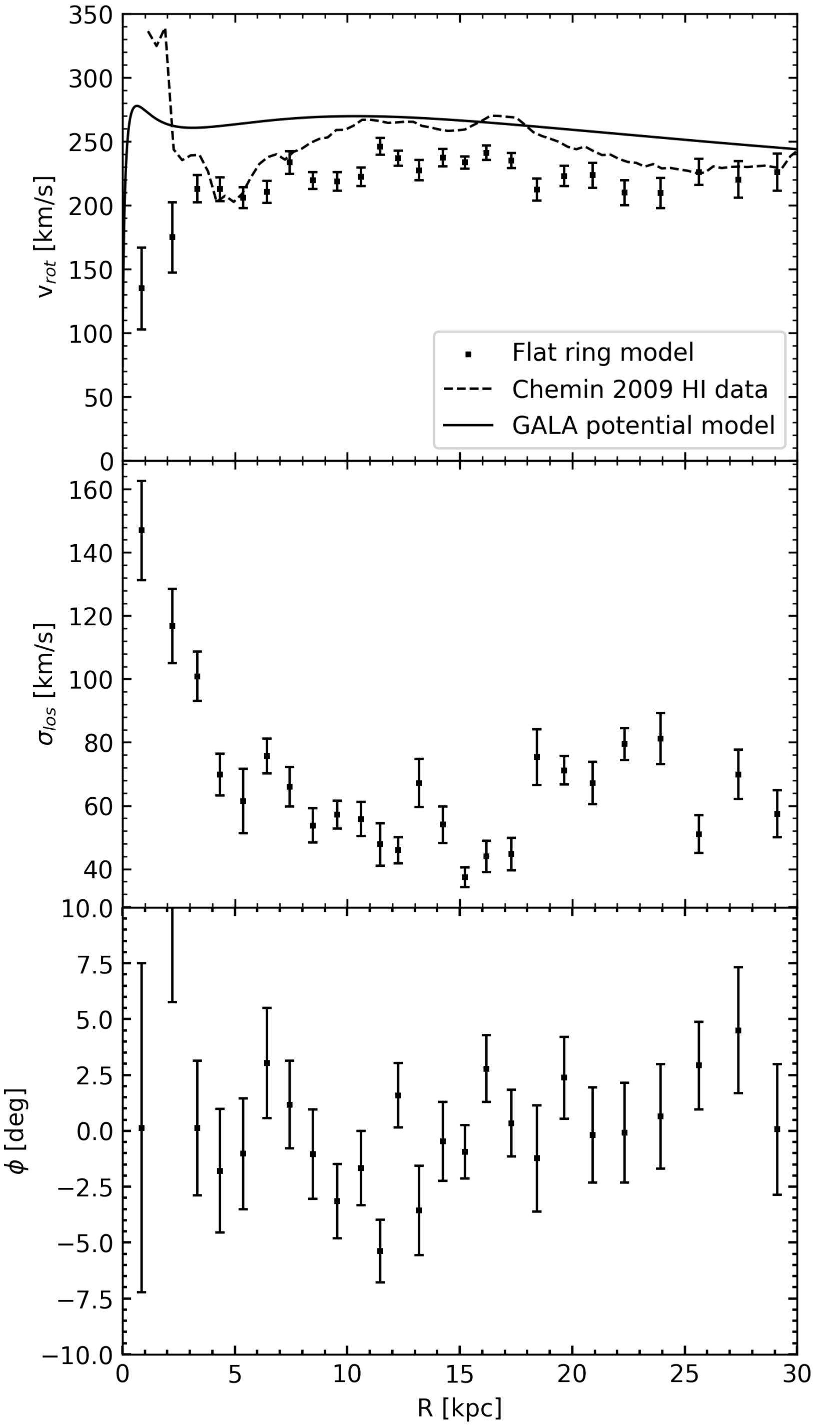


Figure 1: Results of a flat ring model applied to DESI, planetary nebulae and globular cluster kinematic datasets of M31, for 25 rings up to 30 kpc from the galactic centre. **Top:** The stellar rotation curve of this work (squares) compared with an HI rotation curve (dashed) obtained by Chemin *et al.* (2009) and a potential model rotation curve (solid) computed by GALA (Price-Whelan 2017). **Middle:** The line-of-sight velocity dispersion profile. **Bottom:** The azimuthal offset of each ring's sinusoidal line-of-sight velocity profile relative to the semi-major axis of M31 projected on the sky.

could suggest M31 has undergone a more violent accretion history than the Milky Way.

The lower panel of Figure 1 shows the azimuthal offset of the sinusoidal line-of-sight velocity profile in each ring, relative to the semi-major axis of M31 projected on the sky. For a perfectly flat disk this value should always be zero, as the line-of-sight velocity profile for a flat rotating disk is at its maximum/minimum along the semi-major axis, which alludes to the fact that the disk is warped in its inner (R $<$ 5 kpc) and outer (R $>$ 20 kpc) region. In further research, a tilted-ring model which varies ring inclination and position angle is recommended to return more robust ring parameters.

Conclusions

In this work a flat ring model was fit to the disk of M31 using a new survey by DESI in conjunction with previous surveys. The combined line-of-sight velocity field was deprojected and the disk of M31 split into 25 equally populated concentric rings. Using a Gaussian mixture likelihood distribution, a sinusoid was fit to disk star velocities over azimuthal angle, determining the rotational velocity, deprojected line-of-sight velocity dispersion and azimuthal offset of each ring. The stellar rotation curve of M31 was extended to 30 kpc and found to flatten at $\sim$ 220 km/s, agreeing with prior studies. The line-of-sight dispersion profile of the disk was derived and found to be $\sim$ 60 km/s, supporting the picture in which M31's disk is thicker than the Milky Way's and the possible explanation that this is due to a more violent accretion history. Further investigation would include a calculation of the asymmetric drift and subsequent derivation of the mass distribution, as well as a tilted ring model to map out any warps in the disk of M31 and provide a more robust rotation curve.

Acknowledgements

This work was undertaken as part of a Master's project for the author's MPhys (Hons) Astrophysics degree. The author would like to thank Dr. Sergey Koposov for his support and advice while supervising the project.

References

Abadi, M. G. *et al.* 'Simulations of Galaxy Formation in a ΛCold Dark Matter Universe. I. Dynamical and Photometric Properties of a Simulated Disk Galaxy' The Astrophysical Journal **591** 2 (2003)

Begeman, K. G. 'HI Rotation Curves of Spiral Galaxies. I. NGC 3198.' Astronomy & Astrophysics **223** (1989)

Binney, J. and Tremaine, S. 'Galactic Dynamics: Second Edition' (Princeton University Press; 2008)

Caldwell, N. and Romanowsky, A. J. 'Star Clusters in M31. VII. Global Kinematics and Metallicity Subpopulations of the Globular Clusters' The Astrophysical Journal **824** 1 (2016)

Carignan, C. *et al.* 'The Extended H I Rotation Curve and Mass Distribution of M31' The Astrophysical Journal **641** 2 (2006)

Chemin, L. *et al.* 'H I Kinematics and Dynamics of Messier 31' The Astrophysical Journal **705** 2 (2009)

Cooper, A. P. *et al.* 'Overview of the DESI Milky Way Survey' The Astrophysical Journal **947** 1 (2023)

Corbelli, E. *et al.* 'A Wide-Field H I Mosaic of Messier 31. II. The Disk Warp, Rotation, and the Dark Matter Halo' Astronomy & Astrophysics **511** (2010)

Dey, A. *et al.* 'DESI Observations of the Andromeda Galaxy: Revealing the Immigration History of Our Nearest Neighbor' The Astrophysical Journal **944** 1 (2023)

Di Teodoro, E. M. and Fraternali, F. '^{3D}BAROLO: A New 3D Algorithm to Derive Rotation Curves of Galaxies' Monthly Notices of the Royal Astronomical Society **451** 3 (2015)

Hogg, D. W. *et al.* 'Data Analysis Recipes: Fitting a Model to Data' arXiv e-prints (2010)

Koposov, S. E. 'RVSpecFit: Radial Velocity and Stellar Atmospheric Parameter Fitting' Astrophysics Source Code Library (2019)

Koposov, S. E. *et al.* 'Accurate Stellar Kinematics at Faint Magnitudes: Application to the Boötes I Dwarf Spheroidal Galaxy' The Astrophysical Journal **736** 2 (2011)

McConnachie, A. W. *et al.* 'The Remnants of Galaxy Formation From a Panoramic Survey of the Region Around M31' Nature **461** 7260 (2009)

McConnachie, A. W. *et al.* 'The Large-Scale Structure of the Halo of the Andromeda Galaxy II. Hierarchical Structure in the Pan-Andromeda Archaeological Survey' The Astrophysical Journal **868** 1 (2018)

Merrett, H. R. *et al.* 'A Deep Kinematic Survey of Planetary Nebulae in the Andromeda Galaxy Using the Planetary Nebula Spectrograph' Monthly Notices of the Royal Astronomical Society **369** 1 (2006)

Price-Whelan, A. M. 'Gala: A Python Package for Galactic Dynamics' The Journal of Open Source Software **2** 18 (2017)

Roberts, M. S. and Whitehurst, R. N. 'The Rotation Curve and Geometry of M31 at Large Galactocentric Distances.' The Astrophysical Journal **201** (1975)

Rubin, V. C. and Ford W. Kent, J. 'Rotation of the Andromeda Nebula From a Spectroscopic Survey of Emission Regions' The Astrophysical Journal **159** (1970)

Sofue, Y. 'Dark Halos of M 31 and the Milky Way' Publications of the Astronomical Society of Japan **67** 4 (2015)

Vieira, K. *et al.* 'Milky Way Thin and Thick Disk Kinematics With Gaia EDR3 and RAVE DR5' The Astrophysical Journal **932** 1 (2022)

Zhang, X. *et al.* 'The Rotation Curve and Mass Distribution of M31' Monthly Notices of the Royal Astronomical Society **528** 2 (2024)

A Topological Investigation of the Cosmic Web Formation

Eva Maria Staikou*[1], Marcos Pellejero†[2]

[1] University of Glasgow
[2] School of Physics and Astronomy, University of Edinburgh

Open Access

Received
17 Sep 2024

Revised
05 Oct 2024

Accepted
10 Oct 2024

Published
24 Oct 2024

Abstract

N-body simulations of the evolution of the large-scale structure of the universe (the Cosmic Web) were run while altering attractive gravity laws and initial conditions, in order to infer which properties of the universe are revealed by its current large-scale topology. The Cosmic Web was found to develop regardless of the gravitational alterations made. Significantly altering the initial conditions from Gaussian distributions was found to eliminate the Cosmic Web. Thus, we determine that small Gaussian-like perturbations are required in otherwise uniform initial conditions in the early universe, under an attractive gravitational force, for the Cosmic Web to develop.

DOI: 10.2218/esjs.9988 **ISSN** 3049-7930

Introduction

The large-scale structure of the universe, known as the Cosmic Web (CW) (Bond *et al.* 1996), is characterised by: knots, collapsed regions that host galaxies and halos; sheets and filaments, regions collapsed in one or two dimensions respectively that connect the knots; and voids, under-dense regions in between (Martizzi *et al.* 2019).

Modelling matter as collisionless fluid described by continuous phase space density $f(\mathbf{x}, \mathbf{v})$, a function of position, $\mathbf{x}$, and velocity, $\mathbf{v}$, the evolution of the CW can be described by the Vlasov-Poisson system of equations (see e.g., Rein 2007):

$$\frac{\partial f}{\partial t} + \nabla_{\mathbf{x}} f \cdot \frac{\mathbf{v}}{a^2} - \nabla_{\mathbf{v}} f \cdot \frac{\nabla_{\mathbf{x}}\phi}{a} = 0 \tag{1}$$

$$\rho(\mathbf{x}, t) = m_{\mathbf{x}} \int f(\mathbf{x}, \mathbf{v}, t) \mathrm{d}^3 v \tag{2}$$

$$\nabla^2_{\mathbf{x}}\phi = -4\pi G \rho_0 \delta \tag{3}$$

where a is the cosmic scale factor, ϕ is the gravitational potential, ρ is the local density, ρ_0 is the average density, and δ is the fractional difference between ρ and ρ_0. Equation 3 is known as the Poisson equation (Poisson 1827). This system of equations cannot be analytically solved, thus simulations are needed to solve the equations numerically.

This work uses N-body simulations to study the evolution of the CW while varying the laws of gravity and initial universe conditions, in order to infer which properties of the universe are revealed by the current large-scale topology of the universe. This is done by trying to 'break' the CW by making two classes of modifications to the simulations: modifying the Poisson equation, and altering the initial conditions of the universe. 2D simulations were used since they require significantly less computing power than 3D simulations, as to maximise the amount of alterations tested.

The code for the 2D N-body simulations used was based on code written by Stücker (2019). The initial form of the simulation started with a uniform grid of particles, which were then given small perturbations to their positions as initial conditions, randomly selected from a Gaussian distribution. The Vlasov-Poisson system of equations was then solved numerically, along with equations 4 and 5 which

*Student Author
†Corresponding academic contact: mpelleje@roe.ac.uk

describe the evolution of the particle position and momentum vectors, X^i and P^i, for a particle i in the weak-field non-relativistic limit (Angulo *et al.* 2022), to calculate the evolution of the potential, ϕ, with fractional density changes, δ. Thus, the evolution of the position of each particle was determined based on the evolution of ϕ:

$$\frac{\mathrm{d}X^i}{\mathrm{d}t} = \frac{P^i}{m} = \frac{P_i}{ma^2} \tag{4}$$

$$\frac{\mathrm{d}P_i}{\mathrm{d}t} = -m\frac{\partial \phi}{\partial X^i} \tag{5}$$

Poisson Equation Modifications

The first method used to attempt to prevent the CW from forming in the simulations was to make modifications to the Poisson equation. Alterations made to this equation changed the strength of gravity in different ways, while keeping the force attractive. The two main modifications to the Poisson equation that were tested are called the Different Derivative Model (DDM) and the Different Response Model (DRM).

The DDM tested changing the strength of gravity by changing the order of the derivative in the Poisson equation, thus varying the value of n in:

$$\nabla^n_{\mathbf{x}}\phi = -4\pi G\rho_0\delta \cdot R^{2-n} \tag{6}$$

where R is a constant used to ensure the units are the same on both sides of the equation. Increasing n results in an accelerated and stronger gravitational response to density changes, while decreasing n results in a decelerated and weaker gravitational response.

The DRM instead changes the strength of gravity by changing the dependence of the field on δ in the Poisson Equation, thus varying the value of n in:

$$\nabla^2_{\mathbf{x}}\phi = -4\pi G\rho_0\delta^n \tag{7}$$

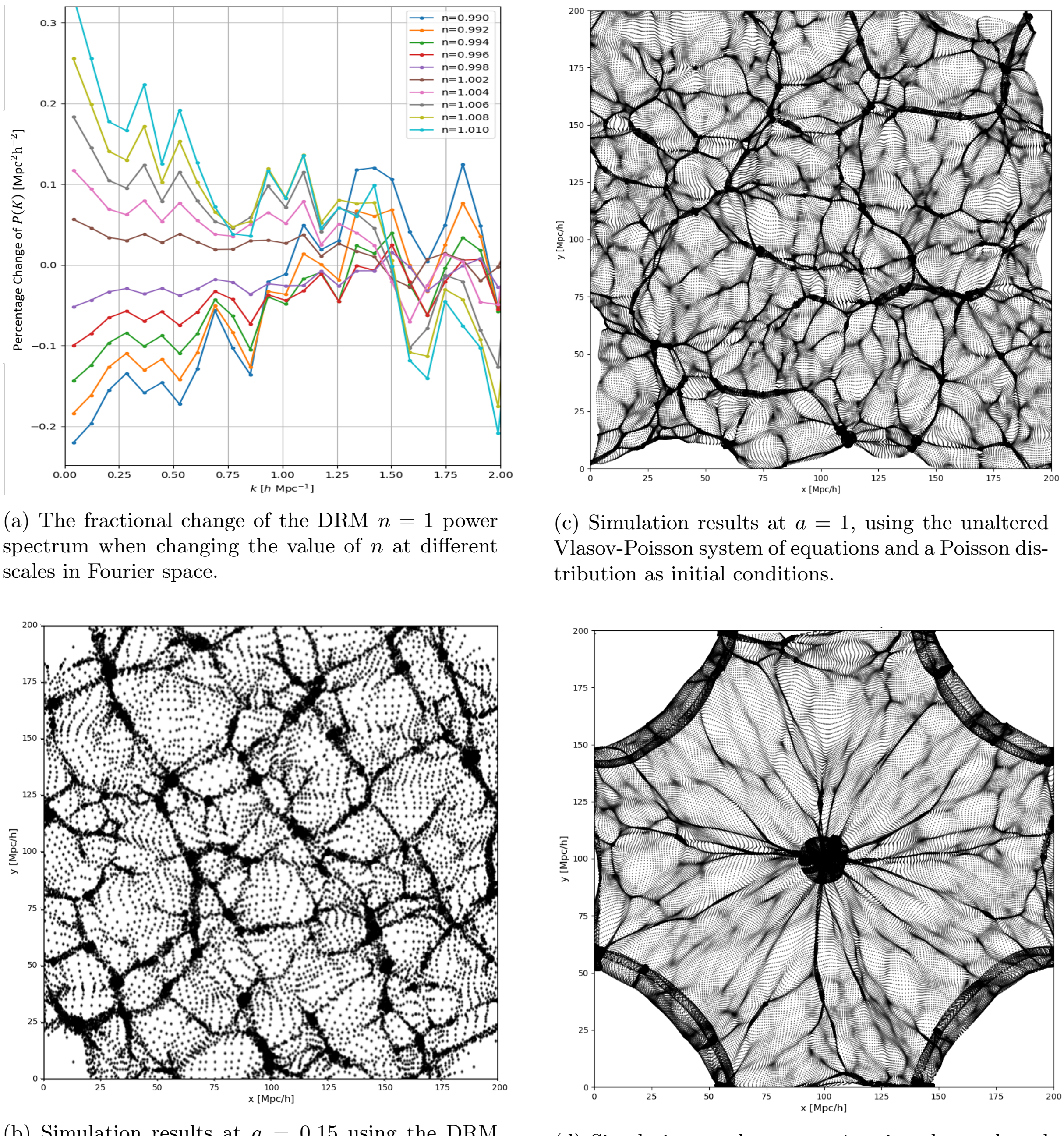
where again, larger values of n result in stronger gravitational responses to density changes.

Running simulations using both models for various values of n from $a = 0.02$ to $a = 1$ and beyond, a clear pattern emerges. Starting from the Gaussian initial conditions, a structure with the same topology as the CW would develop, which would eventually be destroyed as the knots would merge into larger collapsed regions which would start absorbing the filaments. The main difference between these simulations was the time at which the CW developed, and when it was subsequently destroyed. While the CW in our initial simulation is fully formed at $a = 1$, increasing n caused it to develop earlier, and decreasing n caused it to develop at higher values of a. For example, using $n = 1.25$ in a DRM simulation caused the CW to have fully developed by $a = 0.15$, which can be seen in Figure 1b. Based on these results, we can conclude that changing the strength of gravity cannot prevent the CW from eventually forming. This also shows that the topology of the CW alone cannot tell us which of these models represents our universe, and thus other cosmological observations are needed to put constrains on which models are possible, such as the age of the oldest stars in the universe.

The differences between these models were quantified through the calculation and comparison of their power spectra and their probability distribution function statistics. The comparison of the power spectra for different values of n to the original value of n in both models showed a different absolute fractional change between positive and negative changes to n of the same magnitude. An example of this for the DRM can be seen in Figure 1a.

Initial Conditions Modifications

The second method used to attempt to prevent the CW from forming in the simulations was to change the initial conditions provided to the original simulation, which solves the unaltered Vlasov-Poisson system of equations. Instead of randomly selecting small perturbations from a Gaussian distribution as initial conditions for each particle, different distributions were used to select the initial perturbations of the particles.



(a) The fractional change of the DRM $n = 1$ power spectrum when changing the value of n at different scales in Fourier space.

(c) Simulation results at $a = 1$, using the unaltered Vlasov-Poisson system of equations and a Poisson distribution as initial conditions.

(b) Simulation results at $a = 0.15$ using the DRM with $n = 1.25$, using a Gaussian distribution as initial conditions.

(d) Simulation results at $a = 1$, using the unaltered Vlasov-Poisson system of equations and a Binomial distribution ($n = 10$, $p = 0.9$) as initial conditions.

Figure 1: Multiple plots showing simulation results and statistics for three of the alterations made to the simulation.

Using a Poisson distribution and a Random/Uniform distribution to select particle initial conditions had the same result: a structure with the same topology as the CW having developed at $a = 1$, which only had small density differences with Gaussian distribution initial condition results. Using a Binomial distribution with a high p value to select initial conditions, which has a very small overlap with a Gaussian distribution, resulted in the CW not forming in the simulation at any value of a. Two examples of the simulation results using different initial condition distributions can be seen in Figure 1c and 1d.

The last initial condition modification used in order to take a further step away from the Gaussian distribution was using images to define the initial conditions. This was done by firstly converting the images to gray-scale, which assigns each pixel a number describing where it is in the black-white spectrum. Each particle was then assigned the pixel corresponding to its location index, and the magnitude of its perturbation was the gray-scale number of the pixel. This resulted in the grid of particles at $a = 0.02$ looking similar to a less resolved or pixelated version of the image. Using this method of defining particle initial conditions also prevented the CW from developing at any point in the simulation.

These results leads us to the conclusion that small perturbations, following a distribution similar to a Gaussian, to otherwise uniform initial conditions are required for the CW to form. Thus, the current topology of the CW reveals information about the initial conditions of the universe, billions of years ago.

Conclusions

From the altered simulations run in this project, we can conclude that the strength of gravity does not affect the topology of the Cosmic Web. This large-scale structure develops regardless of the type or strength of gravitational alterations made. The formation of the Cosmic Web can be prevented by altering the initial conditions given to the simulation to ones significantly different from a Gaussian distribution, such as a Binomial distribution with a high p value. In conclusion, small Gaussian-like perturbations to otherwise uniform initial conditions of particles (under the influence of an attractive gravitational force) are required in order to generate the Cosmic Web. This tells us that the current topology of the Cosmic Web carries information about the early universe's conditions.

Acknowledgements

I would like to thank my supervisor, Marcos Pellejero, for all his help in completing this project. I would also like to acknowledge the support received from The Royal Astronomical Society for this project, in the form of the RAS Undergraduate Research Bursary.

References

Angulo, R. E. and Hahn, O. 'Large-Scale Dark Matter Simulations' Living Reviews in Computational Astrophysics **8** 1 (2022)

Bond, J. R. *et al.* 'How Filaments of Galaxies Are Woven Into the Cosmic Web' Nature **380** 6575 (1996)

Martizzi, D. *et al.* 'Baryons in the Cosmic Web of IllustrisTNG - I: Gas in Knots, Filaments, Sheets, and Voids' Monthly Notices of the Royal Astronomical Society **486** 3 (2019)

Poisson, S. 'Mémoire Sur La Théorie Du Magnétisme En Mouvement' (L'Académie royale des sciences de l'Institut de France 1827)

Rein, G. 'Collisionless Kinetic Equations From Astrophysics - the Vlasov-Poisson System' in Handbook of Differential Equations: Evolutionary Equations (2007)

Stücker, J. O. 'The Complexity of the Dark Matter Sheet' (2019)

Investigating Enhanced Electrical Conductivity in Ice VII

Elvita Meskauskaite*[1], Miriam Peña Alvarez†[1], Israel Osmond[1]

[1] School of Physics and Astronomy, University of Edinburgh

Open Access

Received
19 Sep 2024

Revised
29 Sep 2024

Accepted
08 Oct 2024

Published
24 Oct 2024

Abstract

At ambient conditions, the electrical conductivity of H_2O is around 0.01 S/m, whereas in the deep interiors of Neptune and Uranus ($\approx$ 5000 K and $\approx$ 300 GPa), conductivity increases by three orders of magnitude. So far, few experimental studies have considered the conductivity properties of ice at high-P-T conditions. In this work, we evaluate how the conductivity of H_2O changes upon compression at 300 K, and show that at 10 GPa, ice VII exhibits a maximum in conductivity of $1.43 \pm 0.47 \times 10^{-5}$ S/m. We propose that time is an important variable in impedance spectroscopy measurements during the pressure-induced transitions: fluid → phase VI → phase VII, between 0.5 and 4 GPa.

DOI: 10.2218/esjs.9994 **ISSN** 3049-7930

Introduction

Voyager 2 revealed that the magnetic fields of Uranus and Neptune are non-dipolar and non-axisymmetric, a phenomenon still poorly understood today (Stone *et al.* 1989). It is hypothesised that the interiors of these planets, composed of CH_4, NH_3, and H_2O, host H_2O in a superionic phase under extreme conditions ($\approx$ 5000 K and $\approx$ 300 GPa) (Hubbard 1997; Teanby *et al.* 2020). In this phase, hydrogen ions diffuse rapidly through a *bcc* (body-centred cubic) lattice, resulting in electrical conductivities exceeding 10 S/m (Cavazzoni *et al.* 1999). Shock compression experiments suggest that superionic ice possesses sufficient conductivity to generate the magnetic fields observed by Voyager 2 (David *et al.* 1959). Convective motions driven by internal heat may induce electric currents, contributing to the planetary dynamo. Understanding this dynamo process requires further study of the electrical conductivity of ices (Helled *et al.* 2020).

Sugimura *et al.* (2012) were the first to experimentally study superionic ice using a diamond anvil cell (DAC) at $T = 739$ K, $P = 56$ GPa, and $T = 749$ K, $P = 62$ GPa, reporting a conductivity of 10 S/m. Other methods, such as shock compression (Millot *et al.* 2018) and X-ray diffraction (Prakapenka *et al.* 2021) have also explored superionic ice, however, due to the experimental challenges at extreme pressure-temperature (P-T) conditions, its properties remain underexplored. Lin *et al.* (2005) investigated the melting behaviour of H_2O under high-P-T, identifying a triple point between fluid H_2O, ice VII, and superionic ice using Raman spectroscopy and X-ray diffraction. This work focuses on developing a technique to examine the electrical properties of H_2O as it compresses to ice VII at room temperature.

Upon compression at 300 K, a series of phase transitions in H_2O has been confirmed: fluid H_2O transitions to ice VI at 1 GPa, and further compression to 2.2 GPa forms ice VII (Noguchi *et al.* 2016). Hydrogen-bond symmetrisation leads to a transition from ice VII to a disordered phase (VII') at 40 GPa, with phase X forming at 66 GPa (Wolanin *et al.* 1997), as observed by X-ray diffraction, neutron studies, and density-functional calculations (Sugimura *et al.* 2008; Komatsu *et al.* 2024).

Phase VI has a tetragonal structure with disordered hydrogen atoms, allowing proton movement (Kamb 1965). Ice VII has a *bcc* structure with a denser hydrogen-bonded network (Kamb *et al.* 1964). Proton transport in the *bcc* oxygen lattice of ice is responsible for its electrical conductivity (Decroly *et al.* 1957; Jaccard 1959) and it involves transitions along hydrogen bonds, creating ionic defects (H_3O^+ and OH^-)

*Student Author

†Corresponding academic contact: Miriam.Pena.Alvarez@ed.ac.uk

and Bjerrum defects (L or D) from bond to bond displacements (Jaccard 1964). Theoretical studies suggest ice VII's conductivity maximum is due to the transition from rotational to ionic defect carriers (Iitaka 2013).

The maximum conductivity of ice at 10 GPa and 300 K reported by Okada *et al.* (2014) has not been reproduced, however, between 20 and 40 GPa at 300 K, conductivity is found to be inconsistent with the work of Sugimura *et al.* (2012), which suggests that electrode geometry or impedance spectroscopy methods may affect the obtained conductivity. Using the methods of Okada *et al.* (2014), we investigate conductivity behaviour at pressure-induced transitions: fluid → phase VI → phase VII, and their time dependence. Comparison with literature values allows us to identify the limiting factors in these methods. Additionally, we present electrical conductivity measurements of H_2O under pressures from 0.4 to 15 GPa at 300 K, using a diamond anvil cell (DAC) along with Raman and impedance spectroscopy. A conductivity maximum for ice VII is observed at $P_c = 10$ GPa and 300 K.

Methods

The DAC was prepared with a 200 μm diamond culet. A rhenium (Re) gasket was pre-indented and laser-drilled to create a sample chamber for Milli-Q water, insulated with Al_2O_3 and low-viscosity epoxy for electrical measurements. Tungsten (W) and gold (Au) electrodes were sputtered using the Korvus HEX magnetron sputtering system, with thicknesses of 1600 Å and 800 Å, respectively.

Ahmed *et al.* (2010) showed that the electrode's active area affects AC impedance in the low-frequency region. Initially, no peak in conductivity was observed with smaller active-area electrodes, as shown in the left panel of Figure 1(a). Redesigning the electrodes to increase proton transport pathways within ice grain boundaries, as shown in the right panel, led to improved results. Electrical contact between the electrodes and the cell was avoided. Impedance measurements were taken using a Hioki "LCR Meter 3536" with a frequency range of 4-8 MHz at 1 V for each pressure step.

Pressure monitoring was done via Raman spectroscopy using a 514.35 nm Argon laser and a PyLoN:100 CCD camera from Princeton Instruments. A 300 lines/mm grating was used to measure the OH stretching mode, and for pressures above 5 GPa, the Raman shift of the stressed diamonds' high-frequency edge was measured using an 1800 lines/mm grating (Hanfland *et al.* 1985).

Results & Discussion

Figures 1(b) and (c) show the pressure dependence of the OH stretching mode during compression. Phase transitions were identified based on (Hsieh *et al.* 2015), using the dominance and Raman shift of symmetric and asymmetric OH stretching modes. As fluid H_2O transitions to phase VI, the asymmetric OH stretching mode is less prominent, allowing the Raman shift of the in-phase symmetric stretch mode to be used for monitoring pressure beyond this transition.

Figure 1(d) shows a Nyquist plot obtained during a 10-minute compression run with a defined semicircle arc and a Warburg impedance tail. The arc diameter is found to increase with pressure, meaning that the bulk resistance increases. It is unexpected that above 2.3 GPa, one still identifies a Warburg tail, as the disappearance of the tail would suggest that the sample transitions from a more diffusive phase to a less diffusive phase, i.e., fluid H_2O → phase VI, VII. We infer that the observation of the Warburg tail is a result of the sample not achieving a steady state. This means that the impedance measurement does not reflect the equilibrium properties of the sample and is affected by factors such as an ongoing phase transition, rearrangement of the hydrogen bond network, or non-linear responses arising from the polarisation of the sample close to electrodes. However, during a decompression run, a sufficient amount of time was given for the sample to equilibrate, hence the disappearance of the tail past 1 GPa, which is in agreement with the fluid H_2O → phase IV transition. Figure 1(e) shows impedance spectra for a decompression run where the bulk resistance arc is less defined due to a greater effect of the electrical double layer (EDL) capacitance.

In phases VI and VII, complex impedance spectra were limited in the low-frequency region due to the EDL capacitance (Ariyoshi *et al.* 2022) as shown in Figure 1(f). Hence, an equivalence circuit fitting technique was used. This method relies on selecting circuit components to form a circuit which reflects the electrical properties of the tested system (Wang *et al.* 2021). The equivalence circuit with the resistor

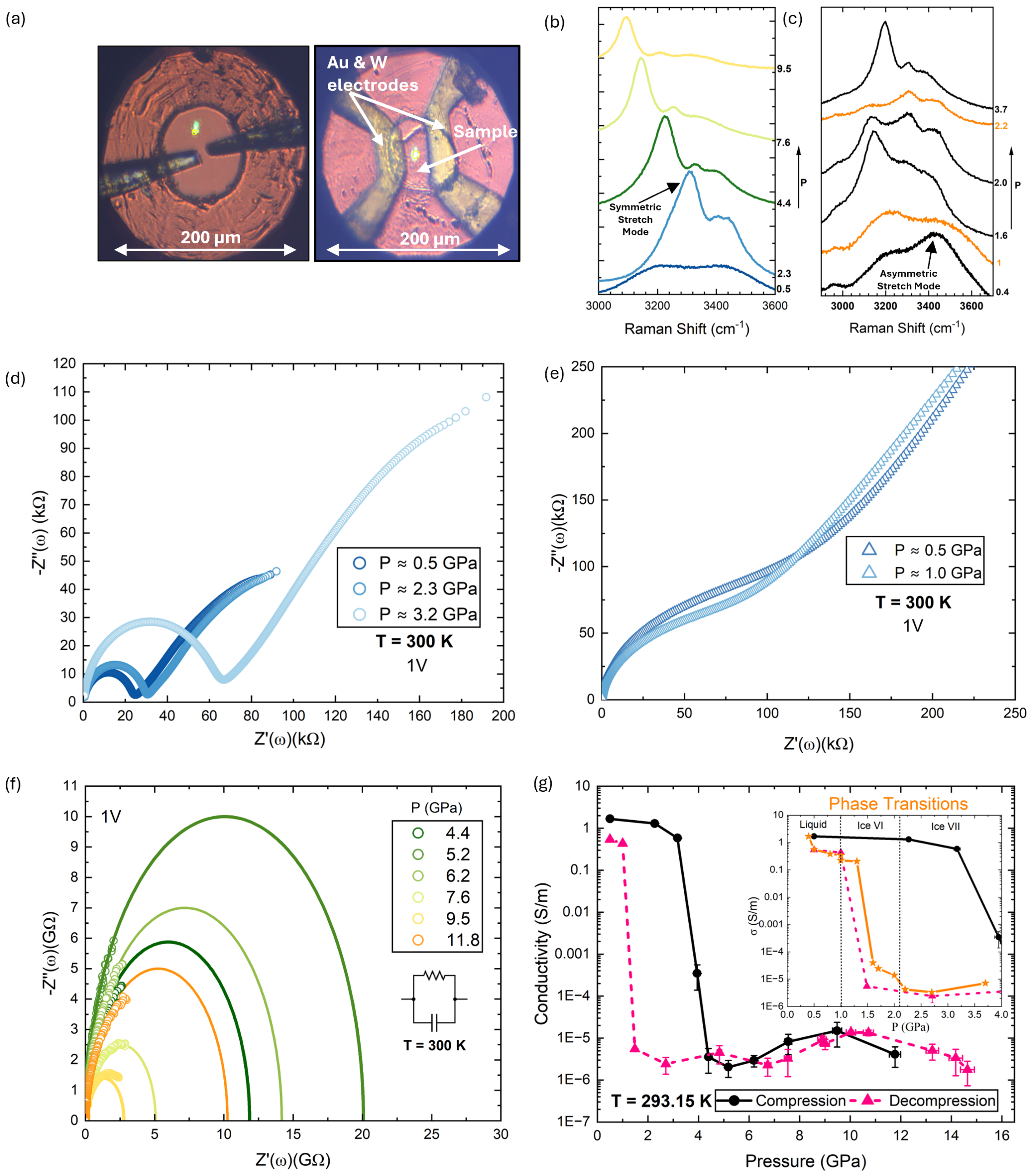


Figure 1: **(a)** Microscopic images of the sample and electrodes in the DAC. Left: Electrode geometry which did not produce a maximum in conductivity. Right: Improved electrode geometry with a greater active area. The sample chamber shows a phase transition from fluid H_2O → ice VI. **(b)** The OH stretching modes are colour-coded to match the complex impedance spectra shown in (d) and (f) for a compression run. **(c)** Raman spectra in the region of OH stretching bands as a function of pressure during the "Phase Transitions" run shown in the inset of (g). **(d)** Nyquist plot of complex impedance spectra for 0.5 GPa $< P <$ 3.2 GPa, which covers the transitions: fluid H_2O → ice VI → ice VII. **(e)** Nyquist plot of complex impedance spectra for a decompression run. **(f)** An equivalence circuit fitting of complex impedance spectra for pressure range, 4.4 GPa $< P <$ 11.8 GPa. The equivalence circuit was chosen to be a resistor in parallel to a capacitor. Resistance values for each pressure step obtained with these fits were used to calculate conductivity values plotted in the following panel. **(g)** Conductivity is calculated from the resistance and dimensions measured for each run and each pressure for both compression and decompression cycles. Inset: Conductivity values in orange represent a repeated compression run, for obtaining more data points covering the phase transition region.

(R) and capacitor (C) in parallel was chosen; see inset of Figure 1(f). The impedance equation of the equivalence circuit used for the fitting is:

$$Z = \frac{R}{1 + (i\omega)^p RC} \tag{1}$$

where the fitting parameter, $p = 1$, represents an ideal capacitor behaviour and is chosen for simplifying the data analysis. Angular frequency, ω, is expressed as $\omega = 2\pi f$, where f is frequency.

The conductivity of ice, σ, was determined by equation $\sigma = d/(R \times l \times t)$, where d is the distance between electrodes, l is the diameter of the sample chamber, and t is the *in situ* thickness of the sample. R is the bulk sample resistance which is determined from the diameter of the arc taken from the fitting shown in Figure 1(f).

In Figure 1(g), we have presented the calculated conductivity for all of our pressure points as a function of pressure. A maximum in conductivity of $1.43 \pm 0.47 \times 10^{-5}$ S/m is observed during compression and decompression runs at $P_c \approx 10$ GPa and 300 K. Okada *et al.* (2014) obtained a maximum in conductivity between 10^{-5} and 10^{-6} S/m at 10 GPa, which is in agreement with our result. Between 2 and 4 GPa, in which the transition VI $\rightarrow$ VII is present, they obtained $10^{-7} < \sigma < 10^{-5}$ S/m. However, in the same pressure region, as shown in the inset of Figure 1(g), we obtained higher conductivity by two orders of magnitude. Okada *et al.* (2014) did not investigate fluid $H_2O \rightarrow$ phase VI transition, therefore, our results cannot be compared. However, in this transition region below 1 GPa, fluid H_2O exhibits conductivity of 0.01–1 S/m, which agrees with literature values of Jones (2002). Positive conductivity dependence of pressure observed by Zheng *et al.* (1997) is inconsistent with our result; on the contrary, the conductivity was found to decrease.

Sugimura *et al.* (2012) obtained $10^{-5} < \sigma < 10^{-4}$ S/m from 20 to 40 GPa at 300 K, and $\sigma \approx 10^{-5}$ S/m at 40 GPa, whereas Okada *et al.* (2014) obtained $10^{-6} < \sigma < 10^{-5}$, and $10^{-7} < \sigma < 10^{-6}$, respectively. These results differ by around two orders of magnitude which is significant when interpreted in the context of superionic conduction.

Given our discrepancies at the phase transitions, and those between previous works by Sugimura *et al.* (2012) and Okada *et al.* (2014), we propose that the inconsistency in the results is due to the electrode geometry and the sample not reaching a steady state. Our preliminary results using electrode configuration shown in the left panel of Figure 1(a) would cause a drift in conductivity and omission of the maximum of conductivity upon compression. However, once the configuration was changed as shown in the right panel of Figure 1(a) the maximum was achieved, indicating that electrode geometry has a significant effect on impedance spectroscopy technique. Secondly, the aforementioned steady-state must also be achieved in order to obtain reliable impedance measurements. However, the effect on conductivity on a single ice phase structure is dependent on pressure. Thus, if a mixture of phases is present, small inevitable pressure gradients may fluctuate the overall conductivity of the sample as well.

Conclusion

We have experimentally found a maximum in electrical conductivity of ice VII at 10 GPa and 300 K, for both compression and decompression runs, using a DAC, through Raman and Impedance spectroscopy techniques. We suggest that electrode geometry and the effect of time-dependent phase transitions on impedance spectroscopy are responsible for the inconsistencies in obtained conductivity in the phase transition dominant pressure region and literature results.

Acknowledgements

I would like to thank Dr Miriam Peña Alvarez and Dr Israel Osmond for their help and support throughout the project, as well as the journal editors and reviewers for their feedback. This project was made possible and funded by the Carnegie Scholarship and the UKRI Future Leaders Fellowship No. Mrc-Mr/T043733/1.

References

Ahmed, R. and Reifsnider, K. 'Study of Influence of Electrode Geometry on Impedance Spectroscopy' in 2010 8th International Conference on Fuel Cell Science, Engineering and Technology (The American Society of Mechanical Engineers; 2010)

Ariyoshi, K. *et al.* 'Electrochemical Impedance Spectroscopy Part 1: Fundamentals' Electrochemistry **90** 10 (2022)

Cavazzoni, C *et al.* 'Superionic and Metallic States of Water and Ammonia at Giant Planet Conditions' Science **283** 5398 (1999)

David, H. G. and Hamann, S. D. 'The Chemical Effects of Pressure. Part 5. The Electrical Conductivity of Water at High Shock Pressures' Transactions of the Faraday Society **55** (1959)

Decroly, J. *et al.* 'Caractére de la Conductivité électrique de la Glace' Helvetica Physica Acta **30** (1957)

Hanfland, M and Syassen, K 'A Raman Study of Diamond Anvils Under Stress' Journal of Applied Physics **57** 8 (1985)

Helled, R. *et al.* 'Uranus and Neptune: Origin, Evolution and Internal Structure' Space Science Reviews **216** 3 (2020)

Hsieh, W. and Chien, Y. 'High Pressure Raman Spectroscopy of H2O-CH3OH Mixtures' Scientific Reports **5** 1 (2015)

Hubbard, W. 'Neptune's Deep Chemistry' Science **275** 5304 (1997)

Iitaka, T 'Simulating Proton Dynamics in High-Pressure Ice' The Review of High Pressure Science and Technology **23** 2 (2013)

Jaccard, C. 'Étude Théorique Et Expérimentale Des Propriétés électriques de la Glace' (1959)

Jaccard, C. 'Thermodynamics of Irreversible Processes Applied to Ice' Physik der kondensierten Materie **3** 2 (1964)

Jones, R. 'Measurements of the Electrical Conductivity of Water' IEE Proceedings-Science, Measurement and Technology **149** 6 (2002)

Kamb, B. 'Structure of Ice VI' Science **150** 3693 (1965)

Kamb, B. and Davis, B. L. 'Ice VII, the Densest Form of Ice' Proceedings of the National Academy of Sciences **52** 6 (1964)

Komatsu, K. *et al.* 'Hydrogen Bond Symmetrisation in D2O Ice Observed by Neutron Diffraction' Nature Communications **15** 1 (2024)

Lin, J.-F. *et al.* 'Melting Behavior of H2O at High Pressures and Temperatures' Geophysics Research Letters **32** 11 (2005)

Millot, M. *et al.* 'Experimental Evidence for Superionic Water Ice Using Shock Compression' Nature Physics **14** 3 (2018)

Noguchi, N. and Okuchi, T. 'Self-Diffusion of Protons in H2O Ice VII at High Pressures: Anomaly Around 10 GPA' The Journal of Chemical Physics **144** 23 (2016)

Okada, T. *et al.* 'Electrical Conductivity of Ice VII' Scientific Reports **4** 1 (2014)

Prakapenka, V. B. *et al.* 'Structure and Properties of Two Superionic Ice Phases' Nature Physics **17** 11 (2021)

Stone, E. C. and Miner, E. D. 'The Voyager 2 Encounter With the Neptunian System' Science **246** 4936 (1989)

Sugimura, E. *et al.* 'Compression of H 2 O Ice to 126 GPA and Implications for Hydrogen-Bond Symmetrization: Synchrotron X-ray Diffraction Measurements and Density-Functional Calculations' Physical Review B **77** 21 (2008)

Sugimura, E. *et al.* 'Experimental Evidence of Superionic Conduction in H2O Ice' The Journal of Chemical Physics **137** 19 (2012)

Teanby, N. A. *et al.* 'Neptune and Uranus: Ice or Rock Giants?' Philosophical Transactions of the Royal Society A **378** 2187 (2020)

Wang, S. *et al.* 'Electrochemical Impedance Spectroscopy' Nature Reviews Methods Primers **1** 1 (2021)

Wolanin, E. *et al.* 'Equation of State of Ice VII Up to 106 GPA' Physical Review B **56** 10 (1997)

Zheng, H. *et al.* 'The Electrical Conductivity of H2O at 0.21-4.18 GPA and 20-350 C' Chinese Science Bulletin **42** 12 (1997)

A JWST Pure-Parallel Search for the First Galaxies With PANORAMIC

Hanna Golawska*[1], Derek McLeod†[1]

[1] School of Physics and Astronomy, University of Edinburgh

Open Access

Received
20 Sep 2024

Revised
15 Oct 2024

Accepted
16 Oct 2024

Published
24 Oct 2024

Abstract

The James Webb Space Telescope offers unparalleled capabilities for observing the most distant objects in the universe, allowing us to study the first galaxies to form. In this work, the Lyman-break technique and spectral energy distribution fitting were used to analyse the JWST PANORAMIC survey and identify 26 robust galaxy candidates at redshifts greater than 9.5, including three around redshift 14.5. These candidates will make excellent targets for follow-up spectroscopy. The derived UV luminosity function at $z = 14.5$ is consistent with literature determinations at similar redshifts and implies a modest evolution in the number density of galaxies from $z \sim 14$ to $z \sim 11$. Future research based on this work could determine when the first galaxies formed, confirming or challenging current theories of cosmic evolution.

DOI: 10.2218/esjs.9993 **ISSN** 3049-7930

Introduction

As light travels at a finite speed, looking further out means looking back in time. The light from the earliest-formed galaxies travels billions of light years to reach us, and in the process gets redshifted due to the expansion of the universe. Only the largest, most powerful, and technically advanced space telescope can detect them: the James Webb Space Telescope (JWST). Operating for just over two years, JWST has already revolutionised the field of astronomy as its deep infrared vision allows us to observe galaxies at extreme redshifts, such as JADES-GS-z14-0, the earliest and most distant spectroscopically confirmed galaxy known so far. At a redshift of $z = 14.32$, it was formed 290 million years after the Big Bang, likely making it one of the very first galaxies to form (Carniani *et al.* 2024). Studying this and other similar galaxies gives invaluable insight into the formation and evolution of galaxies and the whole universe.

While the method involved in this study (photometry) requires analysing only a small part of the galaxy's spectrum of light, spectroscopy demands the time-consuming process of taking its explicit spectrum and comparing emission lines of various elements to laboratory experiments. Photometry is widely used to quickly select galaxy candidates for a spectroscopic confirmation of their redshifts. The large amounts of data released by the JWST until now led to discoveries of many high-redshift ($z > 9.5$) galaxy candidates (e.g., Bouwens *et al.* 2023; Harikane *et al.* 2023a) and spectroscopic confirmation of some of them (Harikane *et al.* 2023b; Castellano *et al.* 2024). Nevertheless, so far, there are only around 30 spectroscopically confirmed galaxies at redshifts greater than 9.5 in total (Harikane *et al.* 2024), making the predictions about the early universe highly uncertain. More data is necessary to better understand the initial conditions of galaxy evolution and place constraints on the period in the history of the universe when they began forming. Therefore, the main motivation for this project is expanding the sample of high-redshift ($z > 9.5$) galaxies for robust statistical analysis. Such analysis often involves measuring the UV luminosity function, which quantifies the abundance of galaxies as a function of the magnitude in the UV part of the spectrum (M_{UV}). Luminosity-weighted integral of the UV luminosity function gives the UV luminosity density, which can be converted to the star formation rate density (star formation rate per unit comoving volume). This can be later compared to various models of galaxy evolution to investigate whether our understanding of how galaxies evolved in the early universe is backed up by observations. The time frame of this study only allowed us to calculate the UV luminosity function at a redshift of $z = 14.5$, leaving the remainder of the analysis for future work.

*Student Author
†Corresponding academic contact: derek.mcleod@ed.ac.uk

One of the most effective and common methods to select high-redshift galaxies is the Lyman-break technique (e.g., McLeod *et al.* 2023). It relies on the fact that radiation at wavelengths shorter than the Lyman limit at 1216 Å is almost completely absorbed by the intergalactic medium (Dunlop 2012). Due to cosmological redshift, this limit is moved to longer wavelengths for galaxies far away from us. By using images taken with two different filters, one above the limit and one below, a Lyman-break galaxy can be discovered since it will only be detected in one of the images. This specific feature allows us to quickly select interesting candidates from large galaxy samples for more time-consuming spectroscopic confirmation.

To further reduce the sample size and only include the most likely candidates, spectral energy distribution (SED) fitting can be implemented. The SED is a plot of how the energy emitted by an object varies as a function of wavelength. The SED of a galaxy depends on various physical properties, such as redshift, age, stellar mass and dust extinction (Conroy 2013). SED fitting comes down to comparing the observed SEDs of individual galaxies to thousands of simulated templates based on various models and selecting the solution that provides the best match to observations. The most robust candidates will have a strong match with one high-redshift solution and a weak match with all low-redshift solutions (Pacifici *et al.* 2023).

The PANORAMIC survey (Williams *et al.* 2024) used in this analysis consists of 40 uncorrelated NIRCam (Near Infrared Camera) pointings, each of them utilising the F115W, F150W, F200W, F277W filters and at least two of the following: F356W, F410M and F444W. Each filter is centred at a different wavelength, for example, F115W is centred around 1.15 micrometres. Their transmission curves and other relevant information can be found at JWST (2017). Each pointing covers an area of about 10 square arcminutes, giving a total area of 0.1 square degree, making this study one of the largest JWST-based searches for high-redshift galaxies to date.

Methods

Firstly, the positions of light sources in all images of the survey were determined and their aperture photometry (flux measurement) performed with SourceExtractor (Bertin *et al.* 1996) in a 0.3" diameter aperture. SourceExtractor, run in dual mode, allows for the detection of objects in one image and their flux measurements in another. The F200W filter images were used for object detection as the spectrum of a high-redshift galaxy is expected to peak around the wavelength covered by this filter. The constructed multiwavelength photometry catalogues were cut to include only those objects that were bright in the F200W filter (with a signal-to-noise ratio, SNR, greater than 5) but were not visible in F115W (SNR $\leq$ 2). This is because for a Lyman-break galaxy at $z = 9.5$, the limit is expected to be observed at 1.28 micrometres, and the F115W filter lies just below this wavelength. An additional detection in any filter at wavelengths longer than the detection filter was introduced as a further criterion to rule out interlopers (objects that are definitively not high-redshift galaxies), such as snowballs (data artefacts caused by large cosmic ray impacts), which would typically appear in only one filter (Regan 2024).

The resulting catalogues were then input into LePHARE (Arnouts *et al.* 2011), a piece of software used for SED fitting. First, it produced spectral libraries with millions of simulated SEDs of galaxies with varying redshift, age, dust extinction etc., and later compared these to every object in each catalogue, looking for a best-fitting and second-best fitting solution. Further cuts were applied to the output catalogues to only leave galaxies that had a good-fitting high-redshift ($z > 9.5$) primary solution and a much less probable secondary solution. Besides the redshift, LePHARE also calculated the magnitude in the UV part of the spectrum which was then used to plot the UV luminosity function that was calculated following the method in Donnan *et al.* (2023).

Finally, the remaining galaxies were visually inspected to remove any that were positioned too close to the edge of an image to give reliable SNR calculations, were found to be artefacts, or did not appear to be a galaxy but rather a part of a foreground object.

Results and Discussion

Aperture photometry, SED fitting, and visual inspection led to the identification of 26 robust galaxy candidates at redshifts greater than $z = 9.5$, including three around redshift 14.5 (meaning the universe was at $\sim 2\%$ of its current age when they emitted the light we observe now):

1. PAN+53.17489-27.94433 at $z = 14.7$ — a very bright and compact object that, while the high-redshift solution is heavily favoured over any low-redshift solution, may be a dwarf star interloper. Future plans are to verify whether it is a star by fitting the photometry with stellar SED templates.

2. PAN+334.25047+0.37907 at $z = 14.5$ — even brighter than the previous candidate, meaning that it can be confirmed with single object spectroscopy as it requires as little as 4-5 hours of observation time. It is extended so a star interloper can be ruled out. After consulting the Mikulski Archive for Space Telescopes*, it has been found to lie in close proximity to the SSA22 protocluster at $z = 3.09$. While a $z = 3$ solution is heavily disfavoured by the SED fit, we cannot definitively rule out the possibility it is a faint member of this protocluster without spectroscopy. Nonetheless, such discovery would still be interesting, as a similar case occurred with the CEERS-93316 galaxy, initially predicted to be at a redshift of $z = 16.4$ (Donnan *et al.* 2023), but later revealed via spectroscopy to have a redshift of $z = 4.9$ (Arrabal Haro *et al.* 2023).

3. PAN+26.09351+17.25871 at $z = 13.4$ — a rather faint object that, fortunately, lies in an image where two more candidates were detected, allowing them to be treated with multi-object spectroscopy.

The brightest of these galaxies will be included in JWST Cycle 4 proposal (observations between July 2025 and June 2026) to confirm its redshift spectroscopically with NIRSpec (Near-Infrared Spectrograph). If confirmed, it might become one of the most distant objects ever known as the current record for the most faraway spectroscopically confirmed galaxy stands at a redshift of $z = 14.32$.

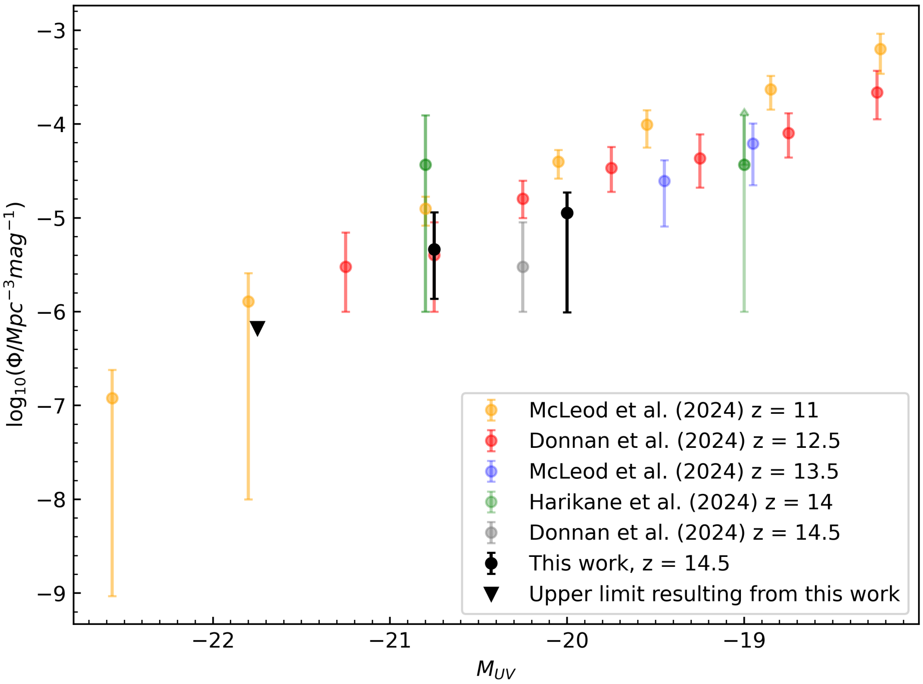


Figure 1: The evolving UV luminosity function at $z \geq 11$ including the three highest-redshift galaxy candidates found in this study binned according to UV magnitude (black points). The downward arrow indicates an upper limit — as no $z \sim 14.5$ galaxies were found with magnitudes smaller than -21.75, this signifies that their number density is smaller than 1 galaxy per the comoving volume encompassed by the whole survey. These results are consistent with previous literature at similar redshifts (McLeod *et al.* 2023; Donnan *et al.* 2024; Harikane *et al.* 2024).

The UV luminosity function (the number density of galaxies per unit comoving volume per unit magnitude, Φ, as a function of the magnitude in the UV part of the spectrum, M_{UV}) including these three galaxies is presented in Figure 1. The results are in agreement with literature at similar redshifts and

*Can be found at https://archive.stsci.edu suggest a mild evolution in the UV luminosity function from $z = 14$ to $z = 11$. Data points corresponding to different redshifts appear to lie near a single line, indicating that the luminosity function remains relatively constant over time.

Secondary solutions produced by LePHARE typically corresponded to low-redshift galaxies with very high reddening due to dust, or extreme emission lines. While many interlopers were removed in the SED fitting process, the only way to definitively confirm our candidates is through spectroscopy. Other possible interlopers are dwarf stars, whose spectral features may imitate a Lyman break. Setting LePHARE to compare the observed photometry to stellar libraries could eliminate some of these, but this approach will require newer, updated stellar libraries with sufficient wavelength coverage to match the JWST NIRCam observations, which extend to ~ 5 micrometres. This work is planned to be done in the future.

Conclusions

In summary, photometric methods such as Lyman-break selection and SED fitting were implemented to search an area of roughly 0.1 square degree for high-redshift galaxy candidates, identifying 26, including three at $z = 14.5$. The spectrum of the most robust and bright candidate will be taken in the future to confirm its redshift. The calculated UV luminosity function at $z = 14.5$ aligns with previous studies and exhibits a moderate evolution between $z = 14.5$ and $z = 11$. This function can later be used to calculate the UV luminosity density and star formation rate density, helping us to test and refine galaxy evolution models and gain a better understanding of the cosmic history.

Acknowledgements

I would like to thank my project supervisor, Dr. Derek McLeod, for his invaluable insight, guidance and support. I am grateful to Dan Magee for creating the JWST NIRCam data reduction software PENCIL used to produce the PANORAMIC data (Magee in prep.) and to Dr. Ryan Begley for sharing his Visual Inspection Tool[†] that notably facilitated the last phase of my analysis. This research project was funded by the Carnegie Trust for the Universities of Scotland.

References

Arnouts, S. and Ilbert, O. 'LePHARE: Photometric Analysis for Redshift Estimate' Astrophysics Source Code Library (2011)

Arrabal Haro, P. *et al.* 'Confirmation and Refutation of Very Luminous Galaxies in the Early Universe' Nature **622** 7984 (2023)

Bertin, E. and Arnouts, S. 'SExtractor: Software for Source Extraction' Astronomy and Astrophysics Supplement Series **117** 2 (1996)

Bouwens, R. *et al.* 'UV Luminosity Density Results at $z > 8$ From the First JWST/NIRCam Fields: Limitations of Early Data Sets and the Need for Spectroscopy' Monthly Notices of the Royal Astronomical Society **523** 1 (2023)

Carniani, S. *et al.* 'Spectroscopic Confirmation of Two Luminous Galaxies at a Redshift of 14' Nature **633** (2024)

Castellano, M. *et al.* 'JWST NIRSpec Spectroscopy of the Remarkable Bright Galaxy GHZ2/GLASS-z12 at Redshift 12.34' The Astrophysical Journal **972** 2 (2024)

Conroy, C. 'Modeling the Panchromatic Spectral Energy Distributions of Galaxies' Annual Review of Astronomy and Astrophysics **51** 1 (2013)

Donnan, C. T. *et al.* 'The Evolution of the Galaxy UV Luminosity Function at Redshifts $z \sim 8-15$ From Deep JWST and Ground-Based Near-Infrared Imaging' Monthly Notices of the Royal Astronomical Society **518** 4 (2023)

Donnan, C. T. *et al.* 'JWST PRIMER: A New Multi-field Determination of the Evolving Galaxy UV Luminosity Function at Redshifts $z \simeq 9 - 15$' Monthly Notices of the Royal Astronomical Society **533** 3 (2024)

Dunlop, J. S. 'Observing the First Galaxies' in The First Galaxies: Theoretical Predictions and Observational Clues (Springer Berlin Heidelberg; 2012)

[†] Can be found at `https://github.com/ryanbegley96/visual_inspection_tool`

Harikane, Y. *et al.* ‘A Comprehensive Study of Galaxies at $z \sim 9 - 16$ Found in the Early JWST Data: Ultraviolet Luminosity Functions and Cosmic Star Formation History at the Pre-reionization Epoch’ The Astrophysical Journal Supplement Series **265** 1 (2023)

Harikane, Y. *et al.* ‘Pure Spectroscopic Constraints on UV Luminosity Functions and Cosmic Star Formation History From 25 Galaxies at $z_{\text{spec}} = 8.61 - 13.20$ Confirmed With JWST/NIRSpec’ The Astrophysical Journal **960** 1 (2023)

Harikane, Y. *et al.* ‘JWST, ALMA, and Keck Spectroscopic Constraints on the UV Luminosity Functions at $z \sim 7 - 14$: Clumpiness and Compactness of the Brightest Galaxies in the Early Universe’ arXiv e-prints (2024)

JWST ‘NIRCam Filters - JWST User Documentation’ *Accessed 19th Sep 2024* (2017)

McLeod, D. J. *et al.* ‘The Galaxy UV Luminosity Function at $z \sim 11$ From a Suite of Public JWST ERS, ERO, and Cycle-1 Programs’ Monthly Notices of the Royal Astronomical Society **527** 3 (2023)

Pacifici, C. *et al.* ‘The Art of Measuring Physical Parameters in Galaxies: A Critical Assessment of Spectral Energy Distribution Fitting Techniques’ The Astrophysical Journal **944** 2 (2023)

Regan, M. ‘Detection and Flagging of Showers and Snowballs in JWST’ *JWST Technical Report; Accessed 17 Sep 2024* (2024)

Williams, C. C. *et al.* ‘The PANORAMIC Survey: Pure Parallel Wide Area Legacy Imaging With JWST/NIRCam’ arXiv e-prints (2024)

Leveraging Machine Learning for Photometric Redshift Estimation of JWST Galaxies

Julie Kalná*[1], Ryan Begley†[1], Callum Donnan[1]

[1] School of Physics and Astronomy, University of Edinburgh

Open Access

Received
20 Sep 2024

Revised
16 Oct 2024

Accepted
21 Oct 2024

Published
24 Oct 2024

Abstract

With the launch of JWST, the volume and complexity of astronomical data are increasing, a trend that will continue with future instruments such as SKA and Euclid. It is inevitable that data-driven methods will become more prominent alongside model-driven analysis. This research utilises machine learning, specifically a kernelised local linear regression model, in photometric redshift predictions of galaxies observed with JWST. With a spectroscopic dataset of 4605 galaxies, we trained the model and achieved a deviation of $\sigma_{\mathrm{dz}} = 0.018$ and a catastrophic outlier rate of $f_{\mathrm{outlier}} = 3.5\%$. These results demonstrate high accuracy and computational efficiency, highlighting the potential of machine learning for astronomical data analysis, in particular for large-scale surveys.

DOI: 10.2218/esjs.9996 **ISSN** 3049-7930

Introduction

Astronomical redshift measures the shift in light wavelengths due to the expansion of the universe, providing key information about galaxy distances and recession velocities, as well as insight into the evolution and large-scale structure of the universe. There are two primary approaches taken for measuring redshift: spectroscopic redshift (z_{spec}) and photometric redshift (z_{phot}). Spectroscopic redshifts are obtained by taking the explicit spectrum of light from an object and comparing prominent emission lines in the spectrum to known rest-frame wavelengths. These measurements are very precise with small uncertainties (e.g., $\delta z \lessapprox 0.01$); however, they can be time-consuming and resource-intensive, especially with large-scale surveys. On the other hand, photometric techniques simply require measurements of an object's flux taken through multiple broadband filters. Some common methods used for photometric redshift estimations are spectral energy distribution (SED) fitting (Brammer *et al.* 2008) and the Lyman-break technique (Dunlop 2012). These techniques enable efficient measurements on large sets of data, often used to compile samples of galaxies for spectroscopic follow-up observations.

The James Webb Space Telescope (JWST) has provided data that looks further back in time than previously possible, allowing some of the fundamental questions about the universe to be addressed. By operating primarily in the infrared, objects from the early universe whose light has been reddened due to cosmological expansion are now detectable. JWST also offers enhanced resolution and sensitivity, enabling deeper observations of fainter, distant objects (McElwain *et al.* 2023; Wang 2024). JWST is providing vast sets of high-quality data ready to be analysed, for which machine learning can offer insights into complicated patterns that may be hard to identify otherwise (Baron 2019).

This research focuses on the implementation of a supervised machine learning model to allow scalability of z_{phot} measurements for galaxies observed by JWST. The goal is to develop a data-driven approach that is efficient and accurate in handling large datasets as well as free from model-dependent biases (Hainline *et al.* 2024). This research demonstrates the potential for enhancing and accelerating data analysis methods during the big data era of astronomy (Zhang *et al.* 2015).

In this work, we first provide a brief background on machine learning and its applications in redshift astronomy. We then describe the observational data and the machine learning model used in this study

*Student Author
†Corresponding academic contact: rbeg@roe.ac.uk

before presenting and discussing the results.

Machine Learning Background

The central idea in supervised machine learning involves a mapping function, f, that relates feature vectors, X, to assigned labels, Y:

$$X \xrightarrow{f} Y \tag{1}$$

Because f is unknown and often not rudimentary, we must derive an estimated function, f_{est}, such that $f_{\text{est}} \sim f$. This is done by finding a function, f_{data}, for a training dataset $(X_{\text{data}}, Y_{\text{data}})$ where Y_{data} has known values. By then assuming that $f_{\text{est}} \sim f_{\text{data}}$, any data point that has a feature vector $\mathbf{x} \in X$ can now be assigned a label $\mathbf{y} \in Y$.

The estimated function depends on parameters, α, that define the model. To find optimal parameters for α, an objective function $J(\alpha)$ is defined which quantifies how well the model's predicted labels match the true labels and is typically composed of two parts:

$$J(\alpha) = L(\alpha, X_{\text{data}}, Y_{\text{data}}) + \lambda R(\alpha) \tag{2}$$

The loss term $L(\alpha, X_{\text{data}}, Y_{\text{data}})$ measures the discrepancy between the predicted labels and the true labels. The regularisation term $R(\alpha)$ penalises model complexity to prevent overfitting.

The hyperparameter λ balances the trade-off between the terms. By minimising $J(\alpha)$ with respect to α, we aim to find the parameters that result in the smallest error on the training data while maintaining the model's ability to generalise to new, unseen data. The model performance is assessed on a testing data set which is withheld during training.

Machine learning approaches aimed at inferring redshift estimates from photometry have previously been implemented to assist in obtaining z_{phot} of galaxies in other large astronomical surveys such as the Sloan Digital Sky Survey (SDSS) (Beck *et al.* 2016). Various approaches have been utilised, including artificial neural networks (Reis *et al.* 2012; Brescia *et al.* 2014) and random forests (Carliles *et al.* 2010). These models are trained on large spectroscopic datasets to predict the redshift of objects based solely on their photometric properties. A local linear regression model was used for low redshifts in the Pan-STARRS1 survey and achieved a standard deviation of $\sigma_{\text{dz}} = 0.0299$ and an outlier rate of $f_{\text{outlier}} = 4.30\%$ (Tarrío *et al.* 2020; a definition of these metrics is provided in the 'Results' section).

However, linear regression can be limiting when trying to fit data that may not necessarily be linearly separable. This work therefore extends the approach taken in Tarrío *et al.* (2020) so that it can be applied to higher redshifts relevant to galaxies detected by JWST by adapting the local linear regression and implementing a kernel function. At higher redshifts, galaxies exhibit more complex spectral energy distributions due to factors such as evolution in galaxy properties and the effects of cosmic expansion (Dunlop 2012). This leads to non-linear relationships between photometric colours and redshift, necessitating more sophisticated models to capture these complexities. Kernel functions are used to help capture non-linear relationships in data by mapping the data to a higher dimensional feature space where non-linear data may become linearly separable (e.g., Hofmann *et al.* 2008).

Method

In this study, a kernelised local linear regression model was used. In the context of redshift, a local linear regression model identifies galaxies close to each other in the colour space and fits the model to the nearest neighbours. The kernel function was implemented to help with capturing more complex, non-linear relationships between the JWST galaxies.

The galaxy photometry used to train the model originates from two surveys; PRIMER (e.g., Dunlop *et al.* 2021) and JADES (e.g., Eisenstein *et al.* 2023; Rieke *et al.* 2023a). From these surveys we use catalogues spanning three fields; PRIMER/UDS, PRIMER/COSMOS and JADES/GOODS-S (the reader is referred to Begley *et al.* (2024) for a detailed outline of these catalogues). The input data, S_n, consisted of magnitudes taken from 8 JWST NIRCam filters (F090W, F115W, F150W, F200W, F277W, F356W, F410M, F444W; Rieke *et al.* 2023b) and 3 Hubble Space Telescope (HST) HST/ACS WFC filters (F435W, F606W, F814W; Clampin *et al.* 2000).

For all consecutive filters, a colour was calculated by taking the difference of magnitudes between the adjacent filters, defining a 10-dimensional feature space consisting of feature vectors $\mathbf{x}^{(t)} \in \mathbb{R}^{10}$ for the model to operate in. The labels are defined by the z_{spec} of each galaxy, $\mathbf{y}^{(t)} \in \mathbb{R}$. Thus, the input data for the model is $S_n = \{(\mathbf{x}^{(t)}, \mathbf{y}^{(t)}) \mid t = 1, \ldots, n\}$, where $n = 4605$ galaxies. The full set of n galaxies is split by a 4:1 ratio into training and testing sets, ensuring stratified splitting such that the training and testing data groups are representative of the overall data.

The model operates in a local space by identifying k-nearest neighbours of a data point by taking Euclidean distances in the normalised 10-dimensional colour space. This newly defined decision boundary is used for fitting the model at that data point. The kernel mapping function used for the model was a Gaussian radial basis function (RBF), which is standard in kernelised regression due to its ability to handle non-linear relationships:

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2) \tag{3}$$

where x and x' are feature vectors and γ is a hyperparameter that defines the width of the RBF. This implicitly computes an inner product in an infinite-dimensional feature space without explicitly computing the coordinates. Now calculating the kernel weights, α:

$$\alpha = (K + \lambda I)^{-1} \mathbf{y} \tag{4}$$

where λ is the regularisation hyperparameter introduced earlier, I is the identity matrix, and y is the vector of labels allows for the prediction. It follows that the predicted z_{phot} is:

$$z_{\text{phot}} = \sum_{i=1}^{k} \alpha_i K(\mathbf{x}, \mathbf{x}_i) \tag{5}$$

where the hyperparameter k is the number of nearest neighbours. The model refines the set of nearest neighbours by excluding those with large residuals based on prediction errors, $\delta_{z_{\text{phot}}}$:

$$\delta_{z_{\text{phot}}} = \sqrt{\frac{\sum_{j=1}^{k} \left(z_{\text{spec}}^{(j)} - (\alpha K)^{(j)} \right)^2}{k}} \tag{6}$$

During this refinement neighbors with residuals greater than $3 \times \delta_{z_{\text{phot}}}$ are excluded. The computation for α is then repeated with the remaining $l < k$ galaxies. This ensures that outliers or poorly matching data points do not negatively affect the prediction.

Hyperparameters were tuned through Bayesian optimisation (e.g., Snoek *et al.* 2012; Shahriari 2016) with a five-fold cross-validation strategy (Stone 1974) to ensure the model generalises well to unseen data. Bayesian optimisation utilises an acquisition function to determine the next point within the hyperparameter space to search by balancing exploration and exploitation. This reduces the number of evaluations required for optimisation and hence improves the model's performance without excessive computational cost. In this model, there are in total three hyperparameters that were tuned: λ, γ, and k.

The model was optimised on a random subset of training data and then run ten times with constant optimised hyperparameters using different random seeds to test for robustness. Each run included 921 testing galaxies, resulting from the 4:1 split of the total 4605 galaxies.

Results

The model's performance was evaluated using three key metrics: mean squared error (MSE), standard deviation of redshift deviations, σ_{dz}, and the fraction of catastrophic outliers, f_{outlier}. The standard deviation σ_{dz} was calculated based on the normalised median absolute deviation (NMAD) as $\sigma_{\text{dz}} =$

$1.4826 \times \sigma_{\text{MAD}}(dz)$, where $dz = (z_{\text{phot}} - z_{\text{spec}})/(1 + z_{\text{spec}})$. The factor 1.4826 scales the median absolute deviation (MAD) to be align with the standard deviation under the assumption of a normal distribution. The fraction of catastrophic outliers was measured as the fraction of galaxies with $|dz| > 0.15$, as typically used (Hildebrandt *et al.* 2010; Jones *et al.* 2020).

On the test set, the average σ_{dz} over the ten trials reached 0.018, with small deviations. The catastrophic outlier rate reached $f_{\text{outlier}} = 3.5\%$.

In Figure 1, we present the results of our model. The top plot shows the normalised redshift deviation dz as a function of the galaxy's z_{spec}. Most points cluster around $dz = 0$, indicating good agreement between the predicted z_{phot} and the true z_{spec}. The bottom plot displays the predicted z_{phot} versus the z_{spec}, showing that the predictions closely follow the ideal $z_{\text{phot}} = z_{\text{spec}}$ line.

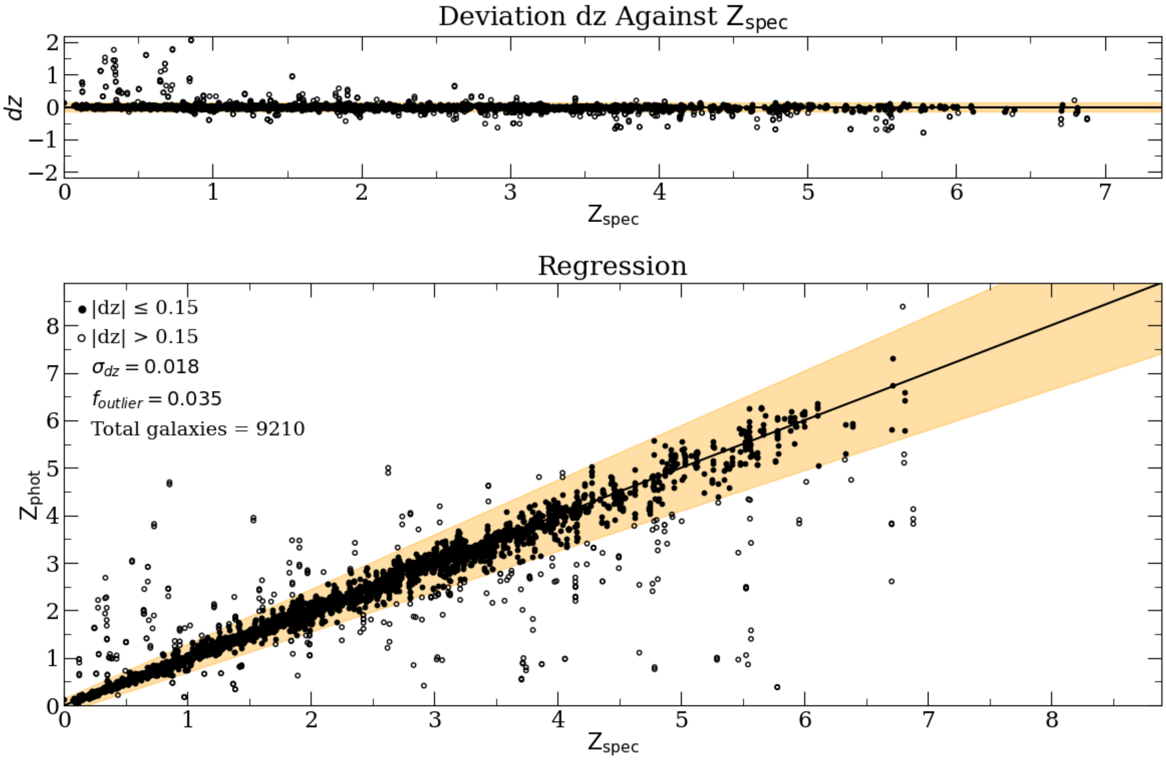


Figure 1: **Top**: The deviation of $dz = (z_{\text{phot}} - z_{\text{spec}})/(1 + z_{\text{spec}})$ as a function of the galaxy z_{spec}. **Bottom**: The regression plot of z_{phot} versus z_{spec}. Both plots contain data from the 10 runs with varying random seeds, each run containing a total of 921 testing galaxies. The solid circles mark galaxies within the defined threshold of $|dz| < 0.15$, while the hollow circles mark galaxies outside this range. The threshold for the outliers is also shaded in orange on both plots.

These results are comparable to individual SED runs from existing methods of code for obtaining z_{phot} and competitive with the latest model-based z_{phot} measurements (Wang *et al.* 2023; Hainline *et al.* 2024).

Discussion

The low value of $\sigma_{\text{dz}} = 0.018$ shows that the model's predictions are tightly clustered around the actual values, and the overall prediction error is low. Further evaluation of the model's performance was carried out by examining the probability density function (PDF) of the normalised redshift error $dz = (z_{\text{phot}} - z_{\text{spec}})/(\delta_{z_{\text{phot}}})$. Ideally, the model's errors would be normally distributed, closely following a standard Gaussian curve, indicating that the model's prediction errors are well-behaved and consistent with statistical expectations. There was a slight bias in the scaled data with the probability trailing off towards negative values along with a few extreme outliers which were sigma clipped. This suggests that the estimated errors $\delta_{z_{\text{phot}}}$ for z_{phot} might need to be reevaluated on whether they give a fair assessment of the estimation accuracy, especially if the PDF were not to improve for a larger data set.

An immediate improvement in the robustness and achieved accuracies of the model could be gained from a larger spectroscopic training sample. Such improvements are imminent with ongoing spectroscopic observations from JWST. This would increase the training size which, in turn, should continue to improve the evaluation metrics as suggested by the learning curve of the model. In the meantime, there are various other steps to be taken, for example evolving the model to utilise subsampling, allowing the model to train and predict on galaxies with missing features, similar to the approach taken in Tarrío *et al.* (2020). This would increase the training data size and result in a more versatile model with fewer restraints placed on the data.

This project offers a baseline for utilising machine learning in handling large-scale datasets efficiently which will be vital for future JWST surveys such as CosmosWeb as well as for next generation large-scale surveys like Euclid or LSST.

Conclusion

This paper demonstrated the potential of supervised machine learning in predicting z_{phot} with both accuracy and computational efficiency. Using a kernelised local linear regression model we achieve a standard deviation, $\sigma_{\mathrm{dz}} = 0.018$, and catastrophic outlier rate, $f_{\mathrm{outlier}} = 3.5\%$, on a relatively sparse spectroscopic data set of 4605 galaxies. The input vectors for the model were constructed in a 10-dimensional colour space using flux data from JWST and HST surveys. Hyperparameters were tuned using Bayesian optimisation and a 5-fold cross validation strategy was implemented to ensure generalisability of the model. An RBF kernel was used to allow for linear regression of non-linearly separable data. The paper also explored potential further modifications of the method which may improve its capabilities such as refining the model to handle missing data. It should be expected for the model to only improve from here as more data becomes available. This research demonstrates the potential of further applications with similar machine learning models in large astronomical surveys, not just limited to JWST, but possibly also for various upcoming large-scale surveys.

Acknowledgments

The author wishes to thank and acknowledge the support of Dr. Ryan Begley and Callum Donnan who supervised the project. The project was funded by the Carnegie Trust for the Universities of Scotland.

Software Availability

The code written for this project is available at:
`https://github.com/Juli-Kalna/JWST-ML-Galaxy-Photo-Z`

References

Baron, D. ‘Machine Learning in Astronomy: A Practical Overview’ arXiv e-prints (2019)

Beck, R. *et al.* ‘Photometric Redshifts for the SDSS Data Release 12’ Monthly Notices of the Royal Astronomical Society **460** 2 (2016)

Begley, R. *et al.* ‘The Evolution of [OIII]+Hβ Equivalent Width From z$\simeq$ 3 − 8: Implications for the Production and Escape of Ionizing Photons During Reionization’ arXiv e-prints (2024)

Brammer, G. B. *et al.* ‘EAZY: A Fast, Public Photometric Redshift Code’ The Astrophysical Journal **686** 2 (2008)

Brescia, M. *et al.* ‘A Catalogue of Photometric Redshifts for the SDSS-DR9 Galaxies’ Astronomy & Astrophysics **568** (2014)

Carliles, S. *et al.* ‘Random Forests for Photometric Redshifts’ The Astrophysical Journal **712** 1 (2010)

Clampin, M. *et al.* ‘The Advanced Camera for Surveys’ in UV, Optical, and IR Space Telescopes and Instruments (International Society for Optics and Photonics; 2000)

Dunlop, J. S. *et al.* ‘PRIMER: Public Release IMaging for Extragalactic Research’ *JWST Proposal Cycle 1* (2021)

Dunlop, J. S. ‘Observing the First Galaxies’ in The First Galaxies: Theoretical Predictions and Observational Clues (Springer Berlin Heidelberg; 2012)

Eisenstein, D. J. *et al.* 'Overview of the JWST Advanced Deep Extragalactic Survey (JADES)' arXiv e-prints (2023)

Hainline, K. N. *et al.* 'The Cosmos in Its Infancy: JADES Galaxy Candidates at Z > 8 in GOODS-S and GOODS-N' The Astrophysical Journal **964** 1 (2024)

Hildebrandt, H. *et al.* 'PHAT: PHoto-Z Accuracy Testing' Astronomy & Astrophysics **523** (2010)

Hofmann, T. *et al.* 'Kernel Methods in Machine Learning' The Annals of Statistics **36** 3 (2008)

Jones, E. and Singal, J. 'Tests of Catastrophic Outlier Prediction in Empirical Photometric Redshift Estimation With Redshift Probability Distributions' Publications of the Astronomical Society of the Pacific **132** 1008 (2020)

McElwain, M. W. *et al.* 'The James Webb Space Telescope Mission: Optical Telescope Element Design, Development, and Performance' Publications of the Astronomical Society of the Pacific **135** 1047 (2023)

Reis, R. R. R. *et al.* 'The Sloan Digital Sky Survey Co-add: A Galaxy Photometric Redshift Catalog' The Astrophysical Journal **747** 1 (2012)

Rieke, M. J. *et al.* 'JADES Initial Data Release for the Hubble Ultra Deep Field: Revealing the Faint Infrared Sky With Deep JWST NIRCam Imaging' arXiv e-prints (2023)

Rieke, M. J. *et al.* 'Performance of NIRCam on JWST in Flight' Publications of the Astronomical Society of the Pacific **135** 1044 (2023)

Shahriari, B. 'Practical Bayesian Optimization With Application to Tuning Machine Learning Algorithms' (2016)

Snoek, J. *et al.* 'Practical Bayesian Optimization of Machine Learning Algorithms' arXiv e-prints (2012)

Stone, M. 'Cross-Validatory Choice and Assessment of Statistical Predictions' Journal of the Royal Statistical Society: Series B (Methodological) **36** 2 (1974)

Tarrío, P. and Zarattini, S. 'Photometric Redshifts for the Pan-STARRS1 Survey' Astronomy & Astrophysics **642** (2020)

Wang, B. *et al.* 'UNCOVER: Illuminating the Early Universe–JWST/NIRSpec Confirmation of Z > 12 Galaxies' The Astrophysical Journal Letters **957** 2 (2023)

Wang, J. 'Eye Beyond the Sky' (Springer Nature Singapore; 2024)

Zhang, Y. and Zhao, Y. 'Astronomy in the Big Data Era' Data Science Journal **14** (2015)

Biological Sciences

Exploration of the Computational Predicted Internal Tagging Preferred Protein Properties

Ruoyu Chen*[1], Lukas Gerasimavicius[2]

[1] School of Biological Sciences, University of Edinburgh
[2] Institute of Genetics and Cancer, University of Edinburgh

Open Access

Received
23 Sep 2024

Revised
03 Oct 2024

Accepted
15 Oct 2024

Published
24 Oct 2024

Abstract

Protein tags are commonly used in many biological experiments, and adding a tag to an intolerant sequence position can significantly damage the functions of a protein and affect the outcome of an experiment. Recently, an in-house computational prediction method, TagScore, was developed which uses sequence homology to identify non-conservative regions of proteins permissive of tagging. The properties of the proteins that are predicted by TagScore to prefer internal tagging were explored using gene enrichment strategies. Proteins that prefer internal tagging were found to be related to GTPase-associated proteins and sequence features with disordered and polar regions, as predicted by TagScore.

DOI: 10.2218/esjs.9999 **ISSN** 3049-7930

Introduction

Gene fusion techniques are commonly used in tag insertion experiments, such as the use of green fluorescent protein tags for cell localisation, polyhistidine-tags for protein purification, and FLAG peptide tags (a synthetic polypeptide tag consisting of eight amino acids) for protein detection and purification, which all involve the construction of fusion proteins. In the majority of scientific literature concerning protein tags, these tags are placed at the N- or C-terminus of the protein of interest (POI). Some of the reasons may be conventional, for instance, if a given tag was already previously used at a certain terminus, it is likely that new proteins will be tagged at the same site with the same tag, following established practices from the past. Other reasons are theoretical, for example, that the end of the protein rarely includes active sites (Osuna 2021).

Most protein labels in contemporary research are added haphazardly, and have a few drawbacks because tagging a protein will make it different from its native form as a fusion protein (Yofe *et al.* 2016; Weill *et al.* 2019) or unable to perform their proper biological functions (Yofe *et al.* 2016; Ki *et al.* 2020). In protein engineering, internal tagging is often necessary, especially when the N- and C-termini of the protein cannot tolerate a tag (van Zwam *et al.* 2024). In addition, internal labelling is critical for several other important reasons when: the termini of the protein are buried or are functionally associated (Zordan *et al.* 2015); there is a need for multiple tagging (Dhar *et al.* 2020); internal tagging is resistant to

*Student Author

proteolytic degradation for some proteins (Bäckström *et al.* 1994); the peptide needs to be structurally and functionally stabilised (Barthelmes *et al.* 2011); and when some functions of specific peptide are considered (Park *et al.* 2014).

Our as yet unpublished computational method, which we refer to as the TagScore method, was developed for predicting the best place to put a tag within a protein sequence using evolutionary information, which predicts the tagging sites both for human proteins and mouse proteins. This method is based on a simple principle of searching for regions of non-evolutionary conservatism by multiple sequence alignment (MSA) that may represent adaptive changes in a particular species or in a particular environment. Unlike conserved regions, which often contain critical active or binding sites, the non-conserved regions are more tolerant to sequence modifications, such as the insertion of protein tags, due to their lower functional importance.

Linkers of a protein are supposed to be a proper place to tolerate tag sequence insertions, which often connect two adjacent functional domains in a protein (George *et al.* 2002). These regions, because of their flexibility, can tolerate or adapt to the insertions of different gene sequences without disrupting the basic function of the protein, sometimes even acquiring new functions such as enhanced stability (Coyote-Maestas *et al.* 2020; Ford *et al.* 2020; Zane *et al.* 2023). For example, Lanthanide-binding tags (LBTs) could be incorporated into three different loops into the interleukin-1β (IL01β) protein without any impact of the overall fold of the protein or binding affinity of the LBTs tag (Barthelmes *et al.* 2011). In the case of Ras GTPase activating protein p120-RasGAP, the catalytic domain in its C-terminus promotes guanosine triphosphate (GPT) hydrolysis and the SH2, SH3, PH and CalB/C2 domains in the N-terminus allow functions such as cell migration and proliferation (Pamonsinlapatham *et al.* 2009).

Our TagScore method, using evolutionary conservation from MSA to infer protein positions tolerant of insertions, was run for every human and mouse protein, generating a dataset of 19,708 unique human protein-coding genes with computational predictions for tag tolerability. Based on the observed alignments and insertions in homologous proteins, each protein residue position was annotated with a TagScore that is scaled in the range 0-1 and represents the probability of tolerating an insertion at the given position. Additional features, such as relative solvent accessibility (RSA) and the AlphaFold modelling quality metric predicted local distance difference test score (pLDDT) were also annotated for comparison. The score was utilised both at the per-residue level and at gene-level. At the gene level, the per-residue TagScores were used to compare tag tolerability across three distinct protein sequence locations — at the N-terminus, the C-terminus, or internally. For each location class, a residue position with the highest TagScore was chosen to represent the gene tag tolerance at that location and, for each protein, the location class (N-terminus, C-terminus or internal) with the highest TagScore was chosen to annotate the protein as being the most tolerant of insertions at that location.

In this study, to gain a deeper understanding of the potential application of TagScore method in specific biological contexts, gene enrichment analyses were performed to explore whether proteins with high tag tolerance are clustered in certain specific biological processes or pathways. This analysis helped reveal which biologically functional proteins might be more suitable for tag insertion, thus providing more precise biological information for protein engineering.

Method

Based upon whether tags are predicted to be the most favourable according to the highest TagScore across the three classes: N-score, C-score, and internal score, the proteins were classified into three groups.

To clarify the biological function and signalling pathways associated with internal tagging preferred tagging genes, gene ontology (GO), Reactome pathway (Gillespie *et al.* 2022) and sequence feature enrichment analysis were conducted by the online resource DAVID v2024q2, accessed on 5th September 2024 (Sherman *et al.* 2022). GO provides comprehensive and computable knowledge concerning gene functions and products (Aleksander *et al.* 2023), which includes three functional categories: biological process (BP), cellular components (CC) and molecular function (MF).

Input data consisted of a list of gene identifiers of genes of the proteins that prefer internal tagging in 'UNIPROT_ACCESSION' format which was the accepted standard of DAVID. The 'Gene_Ontology' functional annotation tool within DAVID was selected and three main GO categories ('GOTERM_BP_DIRECT', 'GOTERM_CC_DIRECT', 'GOTERM_MF_DIRECT') were selected to retrieve comprehensive functional

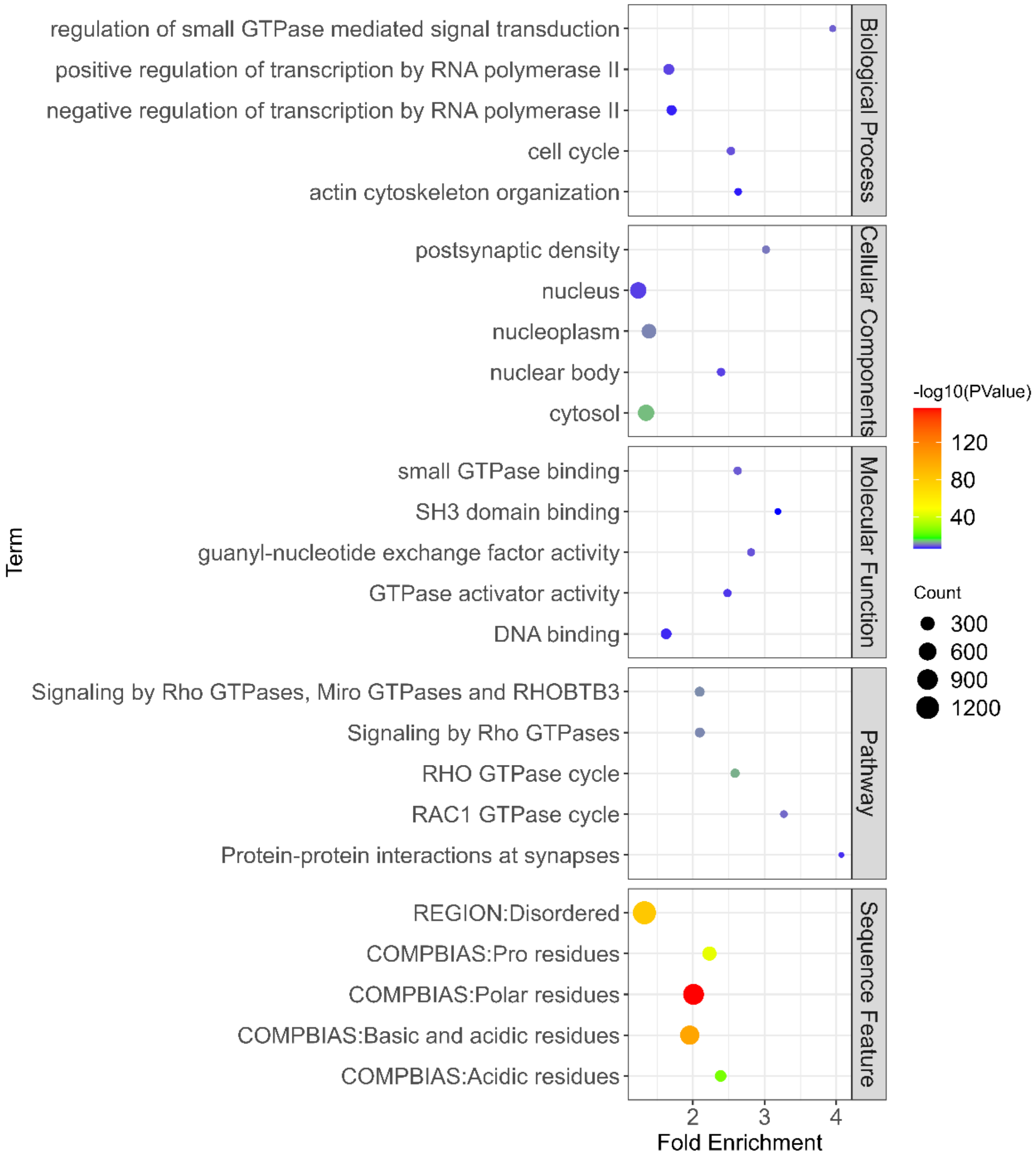


Figure 1: Bubble plot of the enrichment analysis of internal tagging preferred genes. The x- and y-axes indicate the different categories and fold enrichment, respectively. The size of the bubbles represents the number of genes enriched and the significance is shown coloured by $-\log_{10}(P)$, with red indicating the increase in significance.

data for the input genes. For pathway analysis, 'Pathways' options were selected, and from the list of available databases, 'REACTOME_PATHWAY' was chosen to explore the specific signalling pathways related to our input gene set. 'UP_SEQ_FEATURE' was selected for sequence feature enrichment analysis. All the analyses used the default parameters. The results were generated through the 'functional annotation clustering' option. A p-value of $P < 0.05$ was considered statistically significant and terms were selected using an Benjamini-Hochberg false discovery rate threshold of 0.05. Enrichment plots were generated according to fold enrichment, count read, and $-\log_{10}(P)$ through 'ggplot2' 3.5.1. All the code and enrichment results in DAVID online resource are provided at the end of this paper.

Results and Discussion

Biological features of proteins that tend to tolerate tags at different sites have not been previously investigated in detail. To verify if there are any common features of the internal tagging group, genes were used to accomplish enrichment analysis separately using the DAVID online data analysis tool and these results are visualised in Figure 1.

The results show that the top biological processes attributed to the genes of the proteins that prefer internal tagging included 'regulation of small GTPase mediated signal transduction', 'positive *and* negative regulation of transcription from RNA polymerase II promoter', 'cell cycle' and 'actin cytoskeleton organisation'. Their cellular components were mostly attributed to the terms indicating localisation in the nucleus and cytoplasm, and the molecular function terms were largely attributed to the 'small GTPase binding' and 'GTPase activator activity'. The internal group was suggested to be associated with GTPase-related signalling pathways as well as biological processes.

From the REACTOME pathway analysis for the internal group, 30 pathways were significantly enriched ($P < 0.01$, FDR < 0.05). The top 5 pathways were primarily associated with the GTPase related cycle, including 'Signalling by Rho GTPase, Miro GTPases and RHOBTB3', 'Signalling by Rho GTPases', and 'RHO GTPase cycle'. Sequence feature analysis was significantly enriched in disordered and polar residue terms for the internal tagging group, which are usually exposed to the surface of the protein, generating more possibility of tag tolerance compared to buried positions. Overall, gene enrichment analysis showed that genes which prefer internal tagging are more likely to be related to GTPase associated proteins.

There are several possible properties of the proteins in the internal tagging group that make them much better targets for internal tagging, such as the longer protein length on average and a higher disordered sequence content, as predicted by pLDDT. When both protein termini are involved in important molecular functions, it may be a good approach to attempt and insert tags internally. GTPase-associated proteins should be one of the more numerous groups that arise as multi-domain proteins.

Code Availability

The code written for this project is available at:
`https://github.com/B238522-2023/property_analysis.git`.

Acknowledgements

I am deeply grateful to Prof. Joseph A Marsh for his invaluable guidance and profound impact on this project. I also extend my sincere thanks to Dr. Lukas Gerasimavicius for his generous support in data analysis and project research. Additionally, I appreciate the encouragement and support from my family and friends.

References

Aleksander, S. A. *et al.* 'The Gene Ontology Knowledgebase in 2023' Genetics **224** 1 (2023)

Bäckström, M *et al.* 'Insertion of a HIV-1-Neutralizing Epitope in a Surface-Exposed Internal Region of the Cholera Toxin B-subunit' Gene **149** 2 (1994)

Barthelmes, K. *et al.* 'Engineering Encodable Lanthanide-Binding Tags Into Loop Regions of Proteins' Journal of the American Chemical Society **133** 4 (2011)

Coyote-Maestas, W. *et al.* 'Targeted Insertional Mutagenesis Libraries for Deep Domain Insertion Profiling' Nucleic Acids Research **48** 2 (2020)

Dhar, P. *et al.* 'Genetically Engineered Protein Based Nacre-Like Nanocomposites With Superior Mechanical and Electrochemical Performance' Journal of Materials Chemistry A **8** 2 (2020)

Ford, R. C. *et al.* 'Linker Domains: Why ABC Transporters 'Live in Fragments No Longer'' Trends in Biochemical Sciences **45** 2 (2020)

George, R. A. and Heringa, J. 'An Analysis of Protein Domain Linkers: Their Classification and Role in Protein Folding' Protein Engineering, Design and Selection **15** 11 (2002)

Gillespie, M. *et al.* 'The Reactome Pathway Knowledgebase 2022' Nucleic Acids Research **50** D1 (2022)

Ki, M.-R. and Pack, S. P. 'Fusion Tags to Enhance Heterologous Protein Expression' Applied Microbiology and Biotechnology **104** 6 (2020)

Osuna, S. 'The Challenge of Predicting Distal Active Site Mutations in Computational Enzyme Design' WIREs Computational Molecular Science **11** 3 (2021)

Pamonsinlapatham, P. *et al.* 'p120-Ras GTPase Activating Protein (RasGAP): A Multi-interacting Protein in Downstream Signaling' Biochimie **91** 3 (2009)

Park, A. *et al.* 'CRISPR/Cas9 Allows Efficient and Complete Knock-In of a Destabilization Domain-Tagged Essential Protein in a Human Cell Line, Allowing Rapid Knockdown of Protein Function' PLoS ONE **9** 4 (2014)

Sherman, B. T. *et al.* 'DAVID: A Web Server for Functional Enrichment Analysis and Functional Annotation of Gene Lists (2021 Update)' Nucleic Acids Research **50** W1 (2022)

Weill, U. *et al.* 'Assessment of GFP Tag Position on Protein Localization and Growth Fitness in Yeast' Journal of Molecular Biology **431** 3 (2019)

Yofe, I. *et al.* 'One Library to Make Them All: Streamlining the Creation of Yeast Libraries via a SWAp-Tag Strategy' Nature Methods **13** 4 (2016)

Zane, L. *et al.* 'Peptide Linker Increased the Stability of Pneumococcal Fusion Protein Vaccine Candidate' Frontiers in Bioengineering and Biotechnology **11** (2023)

Zordan, R. E. *et al.* 'Avoiding the Ends: Internal Epitope Tagging of Proteins Using Transposon Tn7' Genetics **200** 1 (2015)

Van Zwam, M. C. *et al.* 'IntAct: A Nondisruptive Internal Tagging Strategy to Study the Organization and Function of Actin Isoforms' PLOS Biology **22** 3 (2024)

Journal Acknowledgements

My thanks to all those who have helped with the production of the second issue of this journal, and those who have aided its improvement since the initial issue earlier this year, especially members of the editorial board.

I would also like to again acknowledge the support of Edinburgh Diamond and their team in enabling the publication of the journal, and for all their help and support behind the scenes. Finally, thank you to the PhD students who serve as reviewers for the journal, even though their expertise was not required for this particular issue.

Jack L Smith, Editor-in-Chief