Evaluating student involvement in sea-level data rescue

Patrick McLoughlin[1], Glenn Nolan[2], Kieran Hickey[3] and Gerard D. McCarthy[1]

[1] *Irish Climate Analysis and Research Units (ICARUS), Department of Geography, Maynooth University, Maynooth, Ireland*
[2] *Marine Institute, Rinville, Ireland*
[3] *University College Cork, Cork, Ireland*

Introduction

As the climate crisis continues to escalate, expanding our understanding of past climate and extremes has become more critical than ever. Historical sea-level data hold significant value, providing essential insights into coastal processes such as flooding, erosion and sea-level rise – factors that directly influence climate change assessments and adaptation efforts (Talke and Jay, 2013; Pugh *et al.*, 2021). The preservation and digitisation of historical climate records, often referred to as climate data rescue, play a crucial role in filling gaps in our knowledge. This challenge affects many parts of the world, such as Ireland, Asia, South America and Africa, where long-term tidal records remain scarce or incomplete. Despite their importance, many sea-level records remain undigitised, limiting accessibility. In particular, the digitisation of historical tidal records is a complex process that requires specialised skills and precise attention to detail (Talke and Jay, 2017; Latapy *et al.*, 2023).

In this context, climate data rescue – preserving and digitising historical records – emerges as a critical strategy to broaden access to data and inform both scientific research and policy decisions (Allan *et al.*, 2016; World Meteorological Organization, 2016; Mateus *et al.*, 2021). It presents a unique opportunity to engage students and the wider public in meaningful data management and climate science activities (Williams *et al.*, 2022). Student involvement in data rescue projects not only helps preserve valuable data but also fosters a deeper appreciation for climate science and data management, leading to successful outcomes (Phillips *et al.*, 2018; Ryan *et al.*, 2018). For example, a similar initiative at the National University of Ireland, Galway, engaged 357 participants in digitising over 775 000 historical Irish daily air temperature records. A post-project survey found that 87% valued the data, 88% had a positive experience and 42% showed interest in climate-related careers after participating, highlighting the initiative's impact (Mateus *et al.*, 2021). However, unlike that project, digitising tidal records was more complex due to their graphical nature and therefore required careful planning.

Tidal records are often represented graphically on charts called 'marigrams'. These charts display the rise and fall of sea levels or tides, over a specific period (e.g. Figure 1). They often show multiple tidal traces on a single sheet, with each trace representing sea-level fluctuations over a specific period, typically a day. While other types of historical data, such as written records in ledgers, can be transcribed manually with relative ease, tidal records require advanced techniques due to the precise timing needed for measurements and the challenge of resolving inconsistencies in the data (Latapy *et al.*, 2023).

The study involves geography undergraduate students from the third-year 2024–2025 cohort at Maynooth University, tasked with digitising constructed training datasets and more complex real historical tidal data. This assignment was part of Maynooth University's GY369 Oceanography (Phase 1) and GY310B Geography Research Workshops (Phase 2) modules. The tidal records used for student training include both training datasets (generated from modern tide gauge data) and historical datasets, selected to provide practical experience in digitising and analysing different types of tidal data.

- Kilrush Harbour: Modern Marine Institute tidal data, formatted to resemble historical marigrams, used to train students (Phase 1).
- Dún Laoghaire Harbour (1925): The training tidal data were created using real historical tidal data (McLoughlin *et al.*, 2024a, 2024b) (Phase 2A – Training). Students digitised both the training and historical tidal data (Phase 2B), which were derived from the same dataset.

The datasets are designed to both train students and assess their accuracy in digitising tidal records. This study primarily focuses on evaluating digitisation skills while developing a robust error-checking benchmark to ensure data quality. It also identifies common digitisation challenges, offers insights into students' understanding of tidal data and helps prepare them for future contributions to climate and weather research through improved training in oceanography and climate science. By engaging students in digitising both historical and modern tidal data, this study not only supports climate data preservation but also advances systematic quality control, enhances citizen

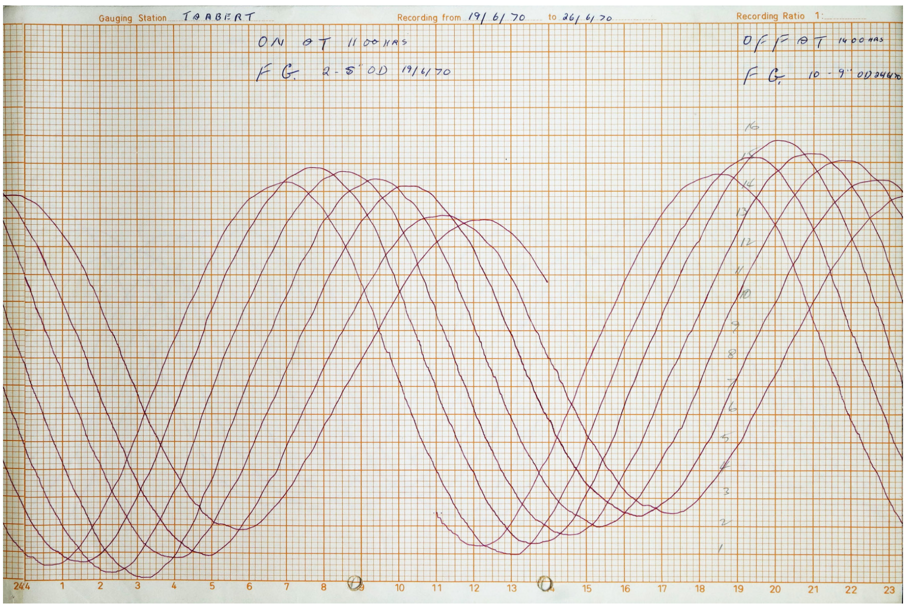

Figure 1. Example of a marigram (visual trace of tidal levels), showing approximately 1 week of daily tidal traces from 19 to 26 June 1970, in Tarbert, County Kerry, Ireland.

RMetS Royal Meteorological Society

science education and lays the groundwork for potential future machine learning applications. Building on Edwards' (2010) concept of the 'vast machine' – the global infrastructure behind climate knowledge – this project trains students to digitise and standardise historical sea-level data, preparing them to participate in global climate data systems. Accurate digitisation of tidal records is essential for calculating realistic weather-driven tidal residuals – such as storm surges and coastal flooding – thereby advancing our understanding of weather impacts on coastal environments.

Motivation for assignment setting

This study builds on prior student-led tidal data digitisation efforts in climate and oceanography courses, which revealed key challenges in maintaining data quality and consistency.

For instance, in 2023, postgraduate students in the Climate Change MSc programme worked with tidal records from Tivoli, Cork. While some students produced accurate results, others struggled with scaling and interpreting the data, leading to inconsistencies. Similarly, an undergraduate project involving 54 students digitising tidal records from Dublin Port found that only 22 of the resulting datasets met high-quality standards. A major issue in both cases was the absence of a reliable method for verifying digitised data. Attempts to benchmark the Dublin Port data against predicted tidal patterns – based on mathematical tidal harmonics – proved inadequate for accurate validation.

To address these challenges, we developed a two-phase approach that integrates real-time quality control and cross-referencing with verified tidal records. In Phase 1, students digitised training images based on modern tidal data from Kilrush (2019–2021). Phase 2 extended this approach by generating training images from historical tidal records of Dún Laoghaire Harbour, previously digitised in another project (McLoughlin *et al*., 2024a), and by digitising the original historical images. Students worked with both training datasets – derived from historical records and the original archival data – enabling hands-on practice, cross-validation and improved accuracy (McLoughlin *et al*., 2024b). This method, described in the next section, establishes a replicable model that includes an assessment of quality for future digitisation projects.

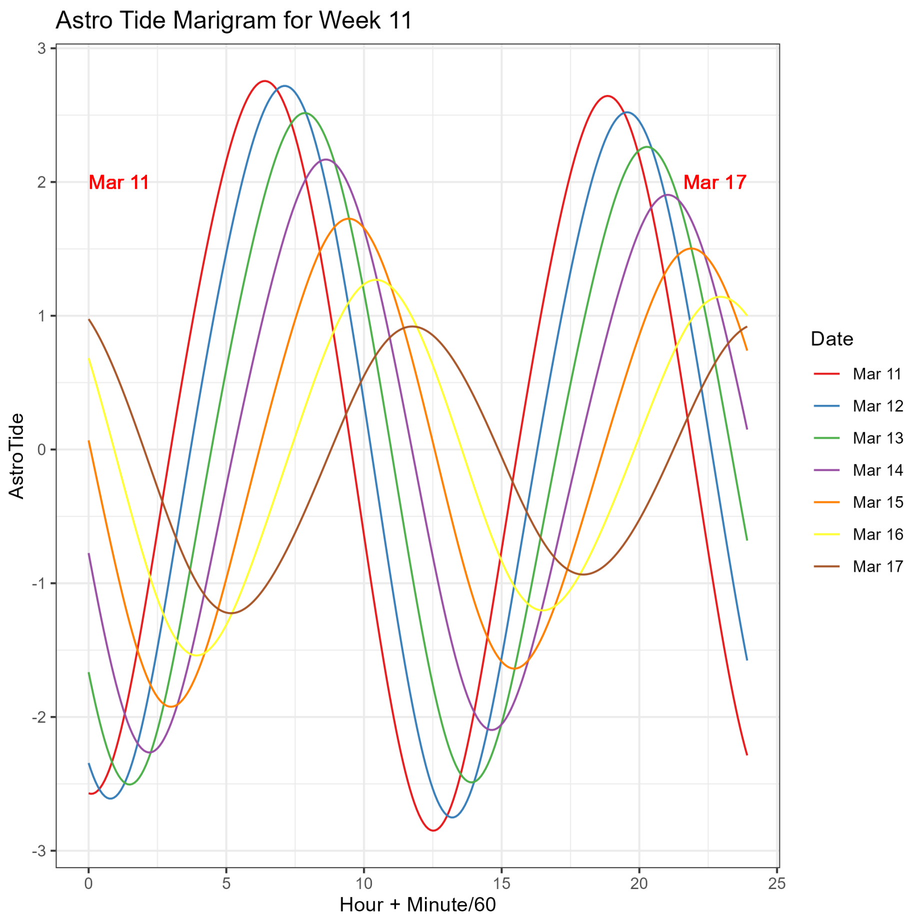

Figure 2. Tidal chart (marigram) generated for Kilrush Harbour, showing 1 week of tidal data from 11 to 17 March 2020. The X-axis represents hours, and the Y-axis represents metres relative to Ordnance Datum Malin (ODM).

Application of assignment in the classroom

Working with a modern training dataset: Kilrush Harbour (Phase 1)

In the first phase, students worked with tidal data from Kilrush Harbour (2019–2021). These modern data were downloaded from the Irish Marine Institute's ERDDAP server and replotted to resemble a historical marigram with seven daily traces on a single image (Figure 2). This setup allowed for immediate comparisons of student-digitised data with known tide levels. Students extracted high and low tidal points from these marigrams. Each student was provided with 4 weeks' worth of data and tasked with identifying the highest and lowest tidal points using WebPlotDigitizer – software that allows users to extract data points from images or plots (Rohatgi, 2024).

Before starting, students attended a workshop on key tidal concepts. These included tidal range (the height difference between high and low tides) and tidal interval (the time between consecutive tides). They also learned about tidal reference levels, including mean tide level (MTL), which is the average height of the tide over a specific period, as well as mean high water (MHW) and mean low water (MLW), which represent the average heights of the highest and lowest tides, respectively. As a supplementary exercise, students calculated these values from their digitised data; however, no results from this task are reported, as the main project focused solely on digitisation.

Having extracted the tidal data using the WebPlotDigitizer software (Rohatgi, 2024), students entered it into Microsoft Excel, where they recorded the tidal heights and corresponding times for further analysis. Excel was used to calculate key measurements, such as tidal range and intervals, using a pre-prepared template. Excel templates are commonly employed in the digitisation of historical data (Inayatillah *et al*., 2022). Students also calculated the mean tide level (MTL) and average high and low water levels.

Digitising tidal data from marigrams: Phase 2A – training and Phase 2B – historical data digitising for Dún Laoghaire Harbour

In the second phase, students from the GY310B Research Workshops module participated in training activities using tidal charts from Dún Laoghaire Harbour for the year 1925. Some students had previously completed the GY369 Oceanography module (Phase 1), which provided foun-

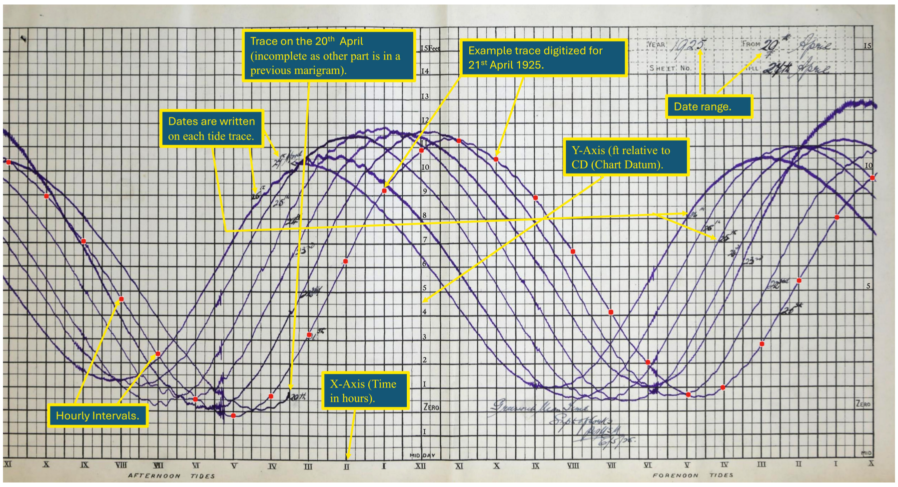

Figure 3. Example of a tidal chart (marigram) from Dún Laoghaire, Dublin, Ireland, for 20–27 April 1925, showing hourly intervals digitised on a trace (April 21). In this image, the X-axis is oriented from right to left, which differs from the usual left-to-right format seen in other charts. This reversed orientation is important to note, as it may require students to adjust their data extraction process, especially when aligning the chart with other data sources that follow the conventional left-to-right format. The X-axis represents hours, and the Y-axis represents feet relative to the Chart Datum (McLoughlin et al.*, 2024a, 2024b). The date is written in pen on each trace.*

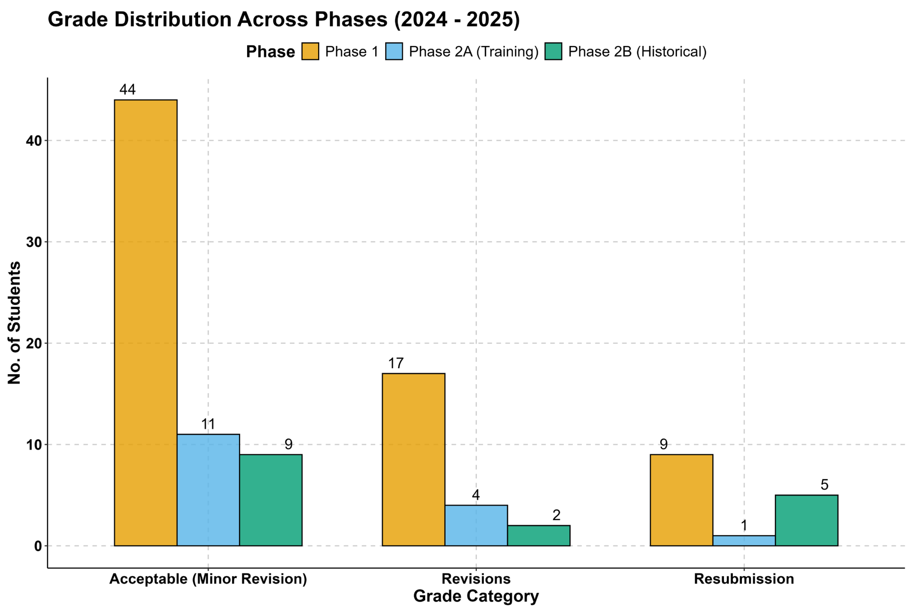

Figure 4. Results from Kilrush (Phase 1) and Dún Laoghaire Harbour (Phase 2A and 2B). The 'Acceptable (Minor Revisions)' category includes data that is usable with only minor issues, such as missing points or slight misidentifications. 'Revisions' refers to data with more noticeable errors, such as misplaced tidal points or hourly intervals. 'Resubmission' denotes data that is largely inaccurate, including significant errors or cases where no submission was made.

dational knowledge in tidal data analysis. Specifically, 10 of the 16 students in Phase 2 had undertaken Phase 1. Each student digitised two tide charts (marigrams): one from a training dataset, which was based on predicted tide levels derived from the historical Dún Laoghaire dataset (1925) (Phase 2A – Training), and another from actual historical records (McLoughlin *et al.*, 2024b) (Phase 2B – Historical Data Digitising). The training dataset (Phase 2A) used model-predicted tide levels for learning purposes, while the historical records (Phase 2B) consisted of real archival data.

Students also used WebPlotDigitizer (Rohatgi, 2024) for extracting data from both the training and historical images to record tidal levels at hourly intervals. This differed from the first phase, where students only recorded the highest and lowest tide levels each day instead of capturing data at hourly intervals. No tidal reference calculations were undertaken. In this phase, students digitised each trace at hourly intervals (Figure 3). Many historical images show bends and distortions (see McLoughlin *et al.*, 2024a) caused by damage to the original paper charts and by the way they were scanned or photographed. These distortions include tilting from angled cameras, warping from uneven paper and even slight stretching or bending from camera lenses. While these issues could affect the precision of measurements, they did not matter for our purposes, as the main goal was to teach digitisation skills.

Two students independently digitised the same tidal charts in Phase 1 only; this was not repeated in Phase 2. While this replication offered some insight into digitisation consistency, any discrepancies were systematically addressed through automated error correction using specialised R code, minimising the impact of individual variation.

Resources and students supports

To facilitate learning, a variety of resources were provided based on previous cohort experiences, emphasising their role in supporting student success in citizen science initiatives (Harlin *et al.*, 2018; O'Donnell, 2023). Essential resources included tutorial videos, detailed data extraction manuals, colour-coded marigram images and immediate tutor support for queries. Additionally, a questionnaire was developed in Phase 1 to assess educational impact and gather feedback on student learning experiences. Materials, including tutorial videos, were made available in advance, allowing for immediate feedback and troubleshooting as tasks were introduced (McLoughlin *et al.*, 2025).

Corrections and quality control measures

To ensure the accuracy and reliability of the digitised tidal data, a series of quality control measures was implemented by the lead author. Students were not responsible for quality checking or validating their data; their role was limited to manual digitisation.

- Kilrush data review: Tidal data digitised and submitted by students were compared with predicted tide levels using R software. Discrepancies between expected and recorded values were automatically flagged, visually inspected using R plots, logged and corrected by the author.
- Dún Laoghaire data review: Similarly, digitised data were checked by the author against predicted tide levels and historical records (McLoughlin *et al.*, 2024b). Discrepancies were flagged, visualised and logged for identification and correction.

Comprehensive instructions for this error-checking process can be found in the relevant links provided in the Data availability statement, including code for generating tidal images (training data), as described by McLoughlin (2025).

Findings

Student success and challenges in tidal data digitisation

A total of 70 participants engaged in Phase 1 of the module, which focused on digitising tidal data from the Kilrush Training dataset. In Phase 2, 16 participants worked with tidal data from Dún Laoghaire Harbour, divided into two sub-phases: Phase 2A, which involved training data, and Phase 2B, which used real historical tidal images. These findings provide valuable insights into students' abilities to digitise historical tidal data.

As shown in Figure 4, 44 out of 70 students (approximately 63%) in Phase 1, 11 out of 16 students (approximately 69%) in Phase 2A and 9 out of 16 students (approximately 56%) in Phase 2B produced work that was deemed 'Acceptable (Minor Revisions)'. Each student's work was evaluated against a verified reference dataset, with deviations under 5cm considered 'minor' errors for correctly identified tidal highs and lows in Phase 1. These outputs were mostly accurate, with deviations for high or low tide points typically under 1–2cm. For hourly interval digitisation of historical charts, overall accuracy generally remained within 3cm, accounting for image distortion and trace clarity. Minor issues that required attention included simple adjustments such as correcting naming conventions, fixing minor date errors, or – in Phase 1 – correcting the identification of a few tidal points. Similarly, in Phases 2A and 2B, minor issues included misidentifying or inaccurately digitising small sections of a trace. However, the overall work was of high quality and suitable for use.

In addition, 17 students in Phase 1, 4 in Phase 2A and 2 in Phase 2B produced data that required more substantial revisions (categorised as 'Revisions'). Despite these issues, their outputs were still considered suitable for use once corrections were applied, demonstrating a solid understanding of the digitisation process. Common issues in Phase 1 included missing a small number of tidal points (typically no more than 3–4), formatting issues in Excel and minor inaccuracies in point placement. In Phases 2A and 2B, additional challenges included misplacement of hourly intervals in the data and formatting errors, such as incorrect date entries in Excel. In Phase 2B, where students digitised historical images, the primary issue was the misidentification of overlapping traces. Some students correctly digitised certain traces but mixed up parts of others where overlaps occurred, which was expected due to the increased complexity of the Dún Laoghaire images.

A notable challenge was the percentage of students submitting work that required resubmission due to significant errors or lack of submission (categorised as 'Resubmission'). In Phase 1, 13% of students (9 out of 70) submitted work deemed unsuitable and categorised as 'Resubmission'. This proportion decreased to 6% (1 out of 16) in Phase 2A but increased to 31% (5 out of 16) in Phase 2B. Due to the small sample sizes, these findings should be interpreted with caution, and no formal statistical tests were conducted. Of the 16 students in Phase 2, 10 had previously completed Phase 1. These students achieved a slightly higher average total score (13.28 out of 15, or 88.5%) compared to those without prior experience (12.5 out of 15, or 83.3%) – a difference of 0.78 points, which corresponds to 5.2% of the total possible score – based only on those who submitted all required work. This suggests that while prior experience provided a slight advantage, students without it still performed well, indicating that the training was effective overall. This increase in resubmissions for Phase 2B was expected due to the more complex nature of the Dún Laoghaire images, which featured overlapping traces and coarser grid intervals. The primary reasons for resubmission across all phases included calibration errors (incorrectly scaling the X and Y axes of the image), consistent misidentification or misdating of traces and Excel formatting issues such as incorrect date and datetime entries.

The analysis revealed common student errors across all phases, including Excel formatting issues, digitisation mistakes, calibration errors and challenges with overlapping traces in historical data. In Phase 1, some students uniquely misidentified tidal highs or lows, especially on days with only three

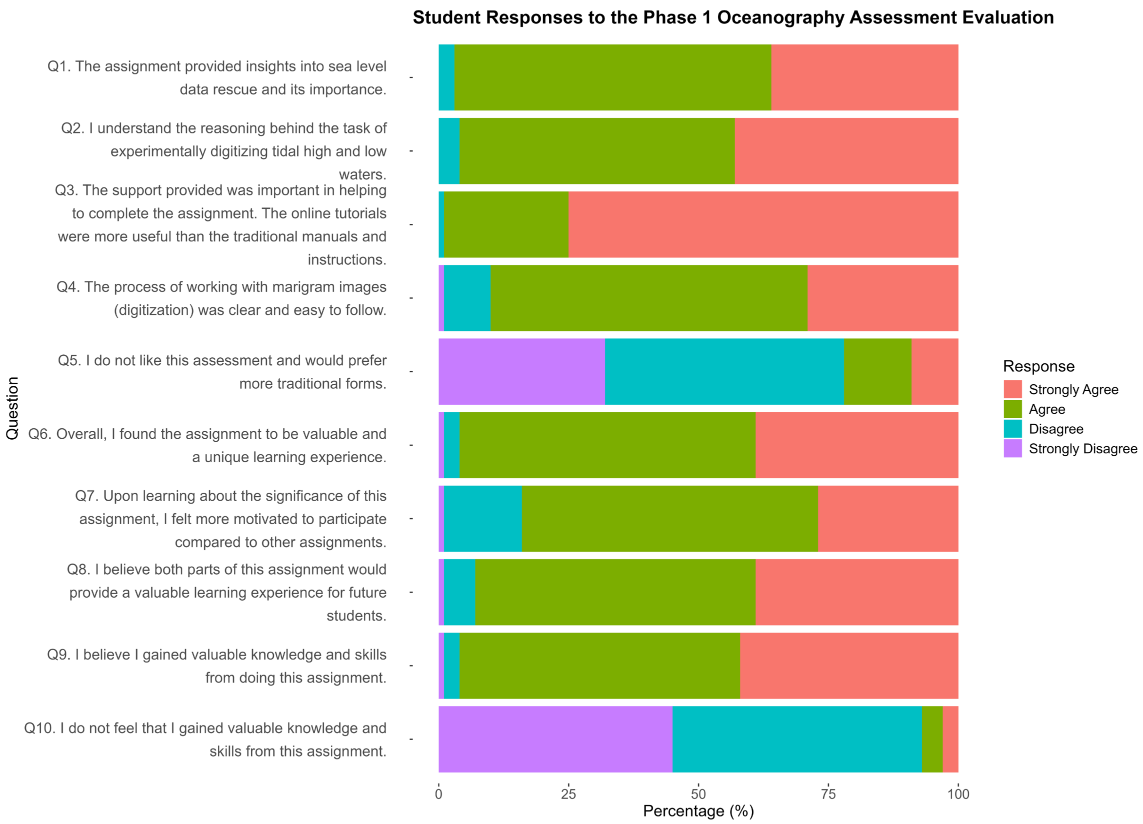

Figure 5. Questionnaire results (Phase 1) on student perceptions of the oceanography assignment's teaching and learning process.

tidal points, where the next point extended into the following day. These results highlight the need for robust error-checking methods. Comparing student work with verified datasets helps pinpoint specific weaknesses and guide improvements in digitisation. Furthermore, most students who performed well in Phase 2A also did well in Phase 2B, indicating the training dataset effectively prepared them for the complexities of real historical data.

Student perceptions: learning, engagement and preferences (Phase 1)

Regarding the perceptions of the oceanography assignment, 67 students completed the Phase 1 questionnaire, with the majority finding the task valuable for their learning. Specifically, 39% of students strongly agreed that the assignment was valuable (Question 6), and 36% strongly agreed that it provided insights into sea-level data rescue (Figure 5), with 61% agreeing. Additionally, 96% agreed or strongly agreed (Question 9) that the task helped them develop valuable skills related to the experimental digitisation of tidal data, which is essential for understanding oceanographic data. Furthermore, approximately 90% agreed or strongly agreed that the process of working with marigram images (digitisation) was clear and easy to follow, although 10% disagreed or strongly disagreed with its clarity (Question 4), highlighting areas for improvement in the task design. Figure 5 summarises the Phase 1 questionnaire responses.

Among all respondents, 84% agreed or strongly agreed that the assignment increased their motivation to participate compared to other assignments, suggesting that the unique nature of this task was a key factor in student engagement. Additionally, 46% of students disagreed, and 31% strongly disagreed that they preferred traditional learning methods over the current non-traditional assignment format (Question 5).

Moreover, 93% of students agreed or strongly agreed that this assignment would provide a valuable learning experience for future students, indicating sustained interest in data rescue activities. Thirty-nine percent strongly agreed that the assignment would provide a unique learning experience for future students, with 54% agreeing that it would offer a valuable learning experience (Question 8), showing strong appreciation for the novelty of the task. Of the 67 students, 99% strongly agreed or agreed that the video tutorials were more effective than traditional manuals (Question 3).

These findings (Table 1) suggest that, while some students may prefer traditional learning methods, the assignment was perceived by the majority as a valuable learning experience that increased engagement and highlighted the importance of tidal and climate data rescue. While these results reflect self-reported perceptions gathered through the questionnaire and not directly measured learning outcomes, they nonetheless offer useful insights into student engagement and perceived skill development.

Table 1

Summary of some key student perceptions of the oceanography assignment's teaching and learning process based on the questionnaire.

Contributor group	*Perceptions*
All respondents ($N=67$)	• Positive learning experience: 39% ($n=26$) strongly agreed that the assignment was valuable (Question 6) • Insightful assignment: 36% ($n=24$) strongly agreed, 61% ($n=41$) agreed, 3% ($n=2$) disagreed that the assignment provided insights into sea-level data rescue and its importance (Question 1) • Clear process: 28% ($n=19$) strongly agreed, 61% ($n=41$) agreed, 9% ($n=6$) disagreed, 1% ($n=1$) strongly disagreed that the process of digitising tidal waters was clear and easy to follow (Question 4) • Support and motivation: 75% ($n=50$) strongly agreed that the support provided was useful, and tutorials were more helpful than traditional manuals (Question 3) • Preference for traditional learning: 9% ($n=6$) strongly agreed, 13% ($n=9$) agreed, 46% ($n=31$) disagreed and 31% ($n=21$) strongly disagreed with the statement, suggesting a preference for the current non-traditional learning format over traditional methods (Question 5)
Strongly agree respondents (varied n)	• Useful learning: 42% ($n=28$) strongly agreed that the assignment helped them gain knowledge (Question 9) • Valuable future experience: 39% ($n=26$) strongly agreed that the assignment would provide a valuable learning experience for future students (Question 8)
Agree respondents (varied n)	• Clear process and instructions: 61% ($n=41$) agreed that the digitisation task was clear (Question 4) • Engagement with content: 57% ($n=38$) agreed that the assignment motivated participation (Question 7)
Disagree respondents (varied n)	• Lack of clarity: 9% ($n=6$) disagreed that the digitisation task was clear (Question 4) • Preference for traditional learning: 46% ($n=31$) disagreed with the statement that they prefer more traditional methods, suggesting they favoured the current non-traditional assignment format (Question 5)
Strongly disagree respondents (varied n)	• Preference for traditional learning: 31% ($n=21$) strongly disagreed with preferring traditional methods over the non-traditional format (Question 5) • Limited learning value: 45% ($n=30$) strongly disagreed that they did not gain valuable knowledge from the assignment (Question 10)

For the full visual distribution of response categories (Strongly Agree, Agree, Disagree, Strongly Disagree), see Figure 5.

Discussion, recommendations and conclusion

The results from the previous section highlight both the potential of students and the challenges in digitising tidal data, which affect broader data rescue efforts. While many students successfully digitised tidal data in all phases, several encountered challenges, including issues with software calibration, Excel formatting and mixing up overlapping traces in Phase 2B. In Phase 1, which focused on extracting daily high and low tidal points, some students incorrectly identified these points, especially on days with three tidal points instead of four. These challenges compromised data accuracy and quality, emphasising the need for targeted interventions. This study presents preliminary digitisation error estimates as a first step towards a detailed, station-specific error budget for historical tidal data. We compared student digitisation results from Kilrush and Dún Laoghaire Harbour training datasets with predicted tides and validated historic Dún Laoghaire records against a trusted reference (McLoughlin *et al.*, 2024b). A 1cm target accuracy was set for synthetic training data. Residuals from one student's Kilrush digitisation (Phase 1), shown in Figure 6, ranged from −0.83cm to +0.47cm, with a standard deviation of ~0.33cm. Absolute errors – mostly classified as 'Acceptable (Minor Revision)' (<5cm) as shown in Figure 4 – provide a quantitative framework for accuracy, reflecting the performance of students whose errors typically remained within ±1cm overall in the training datasets.

Preliminary comparisons showed that student digitisations of synthetic data had errors under 1cm, while historical tide record digitisations – validated against available archival station records – had errors between 1 and 3cm due to image quality. A 1cm target was set for training data, with up to 3cm considered acceptable for historical images. However, 1cm remains the ideal accuracy for historical sea-level data, which require distortion corrections (McLoughlin *et al.*, 2024a). This study addresses digitisation errors only, excluding datums or tide gauge calibration uncertainties.

Previous studies reinforce these findings and underscore the importance of precise digitisation. Wöppelmann *et al.* (2008) showed long-term stability at the Brest tide gauge, highlighting the need to minimise digitisation errors when physical benchmarks are reliable. Talke *et al.* (2020) reported early tidal records with uncertainties of several centimetres, often due to outdated systems or degraded reference points. Even with modern tools, uncertainty estimation remains essential for quality control. McInnes *et al.* (2024) showed that common student digitisation errors can be detected and corrected by comparing results with predicted tides.

Future work should create detailed error budgets for each station and time period by combining student digitisation results with historical data uncertainties. A dynamic error model – tracking how uncertainty varies over time or by source – can inform re-digitisation priorities and data confidence. These insights are crucial for Bayesian analyses, where the quality of the data influences the assumptions built into the models. The error budget presented here lays the groundwork for this broader effort.

Training individuals in the accurate digitisation of historical tidal records is crucial for reconstructing long-term sea-level trends that directly inform climate change assessments, weather variability studies and coastal adaptation planning. These high-quality datasets enhance understanding of climate-driven sea-level variability and extreme weather phenomena, such as storm surges, while complementing other environmental datasets – including fluvial records – thereby supporting integrated modelling and improved prediction of diverse climate and weather hazards.

Based on our findings, with 63% success in Phase 1, 69% in Phase 2A and 56% in Phase 2B producing acceptable work (with minor revisions), it is clear that most students grasped the task through the workshops and online tutorials. The findings indicate that prior training provided a modest performance boost, but students without previous experience still performed nearly as well, demonstrating the overall effectiveness of the training programme.

Digitising tidal data (marigrams) is more complex than digitising other types of data, such as tabulated records, due to the intricate nature of tidal curves and overlapping traces. This greater complexity highlights the technical challenges of working with graphical data, a contrast to previous studies focused on transcription-based data (Ryan *et al.*, 2018; Mateus *et al.*, 2020; Mateus *et al.*, 2021). This type of data may require more pretraining and error-checking to achieve success rates comparable to more established data rescue initiatives (Hawkins *et al.*, 2019).

To address these challenges, educational institutions should provide structured digitisation training for complex datasets. Strengthening collaborations, fostering interdisciplinary learning (Aryee *et al.*, 2024) and securing funding will enhance quality and reach. Automated error-checking and tutorial videos would further improve accuracy and efficiency. Effective training and standardised protocols are essential to maintain data integrity (van der Velde *et al.*, 2017).

We have established a model for digitisation, where the student or citizen scientist first digitises training tidal data images, followed by real historical images from Dún Laoghaire. If successful, the user progresses to digitising real data; if not, they revert to the training data (Figure 7). These data are systematically benchmarked against known answers, ensuring reliable quality checks.

This methodology applies to three key areas: (1) digitisation projects that offer resources and methods to train students, citizen scientists, or professionals; (2) educator-led training exercises that provide hands-on experience with digitisation techniques; and (3) the structured preparation of datasets that may support future automation efforts. While not implemented in this study, future work could explore machine learning techniques – particularly image recognition – as a tool to assist with marigram digitisation. Such tools could complement manual workflows by increasing efficiency and consistency. However, the complexity and variability of historical marigrams currently pose challenges for automation, and further research is needed to assess feasibility. Due to current complexities, machine learning development is beyond the scope of this study but remains a promising avenue for future research. Student feedback from our questionnaire for Phase 1 indicates that the learning resources were valuable, helping them develop critical skills and insights. However, to ensure digitisation projects produce meaningful climate datasets, more intensive training and resources are needed. Training focused on digitising daily high and low tidal points (astronomical tides) and hourly intervals. Students were not asked to analyse non-tidal residuals like storm surge, which cause real deviations from predicted tides, and so were not guided to distinguish these from digitisation errors. Future materials should address this to better

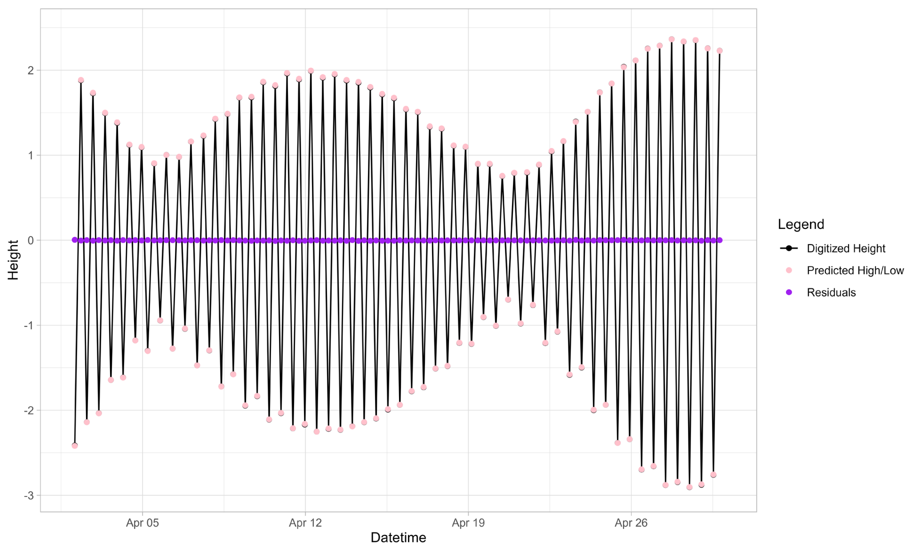

Figure 6. Comparison between digitised high and low tide points from a student's digitisation (black dots, almost fully hidden beneath pink dots due to excellent accuracy) and predicted tide levels (pink dots) at Kilrush Harbour. Purple dots represent the residuals – the quantitative differences between predicted and digitised values – offering a clear visualisation of the error distribution. This residual analysis supports preliminary estimates of digitisation error and serves as an initial step towards developing a station-specific error budget.

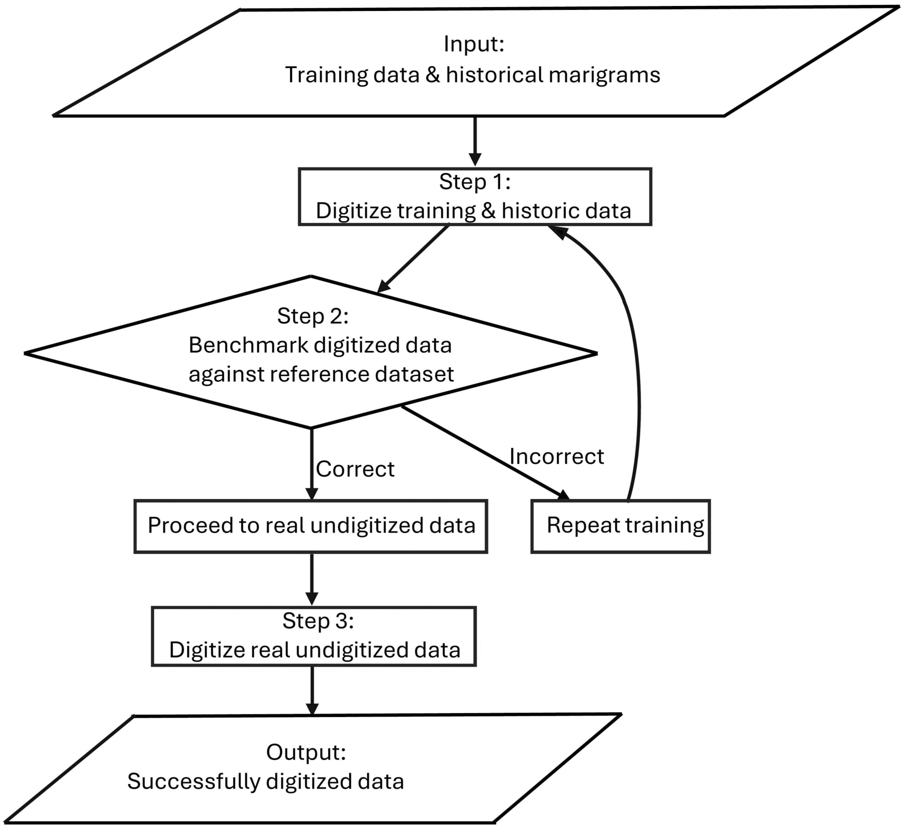

Figure 7. A workflow chart illustrating the digitisation process, including training, benchmarking against reference datasets and progressing to real data digitisation based on accuracy.

interpret and link such deviations to weather events. Embedding citizen science frameworks and partnering with professionals or retirees can enhance engagement and support research on storm surges, sea-level rise and climate change. While 84% of students felt more motivated, 93% expressed interest in this assignment as a model for future work, suggesting sustained engagement with data rescue activities. This study did not include formal pre- and post-assessments. Future research could incorporate such methods to more rigorously evaluate student learning outcomes – particularly in tidal processes, data interpretation and the climate relevance of historical datasets.

A case study from the Williamstown tide gauge in Port Phillip Bay, Melbourne, illustrates the value of comprehensive training (McInnes *et al*., 2024). This study extended historical sea-level records by nearly 100 years, yielding valuable insights for sea-level trend analysis and climate adaptation planning. While this success highlights the potential of student-led digitisation projects, our findings suggest that achieving comparable outcomes would require more extensive training.

To ensure that digitisation projects involving complex tidal data contribute meaningfully to climate research, it is essential to adopt a comprehensive training approach. Systematic error-checking is also crucial to maintain high data quality. We have developed a training method applicable to students, citizen scientists and educators, which may also support future machine learning model development – a step currently limited by the complex visual nature of marigrams. By refining this approach and fostering collaboration among participants, future digitisation efforts can preserve high-quality historical tidal data and deepen our understanding of sea-level trends and climate impacts. By improving digitisation accuracy through targeted training, this approach strengthens the foundation for producing reliable climate and weather data from historical archives, crucial for climate resilience research.

Acknowledgements

We extend our sincere thanks to the students of the GY369 Oceanography module (2024 and 2025 cohorts) and the GY310B Geography Research Workshops module for their contributions, which informed and supported the development of this article. The authors appreciate the insightful suggestions provided by the Editor and the reviewers.

Author contributions

Patrick McLoughlin: Conceptualization; writing – original draft; software; writing – review and editing. **Glenn Nolan:** Writing – review and editing. **Kieran Hickey:** Writing – review and editing. **Gerard D. McCarthy:** Writing – review and editing; conceptualization.

Funding information

PM is funded (Grant-Aid Agreement No. CS/21/006) with the support of the Marine Institute and funded under the Marine Research Programme by the Irish Government. GM is supported by the A4 Project (Aigéin, Aeráid, agus athrú Atlantaigh), with the support of the Marine Institute, under the Marine Research Programme, funded by the Irish Government grant no.: (PBA/CC/18/01) and iCRAG, with the support of Science Foundation Ireland (grant no. 13/RC/2092_P2). The open access publication fee was covered by Maynooth University via an institutional agreement with Wiley.

Conflict of interest statement

The authors declare they have no conflict of interest in relation to this work undertaken.

Data availability statement

The video tutorials designed to support students in completing the assignment are available on Zenodo (https://zenodo.org/records/14905518). The code for generating the training data (Kilrush and Dún Laoghaire) and the error-checking scripts for the Dún Laoghaire (including the historical reference dataset) and Kilrush datasets are available on GitHub (https://github.com/PatrickJMcLoughlin/Evaluating-Student-Involvement-in-Sea-Level-Data-Rescue-/tree/main).

References

Allan R, **Endfield G**, **Damodaran V** ***et al***. 2016. Toward integrated historical climate research: the example of atmospheric circulation reconstructions over the Earth. *WIREs Clim. Change* **7**(2): 164–174. https://doi.org/10.1002/wcc.379 (accessed 11 November 2024).

Aryee JNA, **Mensah C**, **Ayitah DE** ***et al***. 2024. Empowering the next-generation climate scientists: insights from the 2023 Forecasters' WEB exchange program/hackathon at KNUST, Ghana. *Weather*. **80**(3): 79–85. https://doi.org/10.1002/wea.7633

Edwards PN. 2010. A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming. MIT Press: Cambridge, MA.

Harlin J, **Kloetzer L**, **Patton D** ***et al***. 2018. Turning students into citizen scientists, in Citizen Science: Innovation in Open Science, Society and Policy. Hecker S, Haklay M, Bowser A *et al*. (eds). UCL Press: London, pp 410–428. https://www.jstor.org/stable/j.ctv550cf2.35 (Accessed 10 October 2024).

Hawkins E, **Burt S**, **Brohan P** ***et al***. 2019. Hourly weather observations from the Scottish Highlands (1883–1904) rescued by volunteer citizen scientists. *Geosci. Data J.* **6**(2): 160–173. https://doi.org/10.1002/gdj3.79

Inayatillah A, **Haigh ID**, **Brand JH** ***et al***. 2022. Digitising historical sea level records in the Thames Estuary, UK. *Sci. Data* **9**:

167. https://doi.org/10.1038/s41597-022-01223-7

Latapy A, **Ferret Y**, **Testut L** ***et al***. 2023. Data rescue process in the context of sea level reconstructions: an overview of the methodology, lessons learned, up-to-date best practices and recommendations. *Geosci. Data J.* **10**: 396–425. https://doi.org/10.1002/gdj3.179 (Accessed 11 November 2024).

Mateus C, **Potito A**, **Curley M**. 2020. Reconstruction of a long-term historical daily maximum and minimum air temperature network dataset for Ireland (1831-1968). *Geosci. Data J.* **7**(2): 102–115. https://doi.org/10.1002/gdj3.92 (Accessed 11 November 2024).

Mateus C, **Potito A**, **Curley M**. 2021. Engaging secondary school students in climate data rescue through service-learning partnerships. *Weather* **76**(4): 113–118. https://doi.org/10.1002/wea.3841

McInnes KL, **O'Grady JG**, **Hague BS** ***et al***. 2024. Digitizing the Williamstown, Australia tide-gauge record back to 1872: insights into changing extremes. *J. Geophys. Res. Oceans* **129**(8): e2024JC020908. https://doi.org/10.1029/2024JC020908 (Accessed 11 November 2024).

McLoughlin PJ. 2025. Evaluating Student Involvement in Sea Level Data Rescue. (Computer Software) GitHub. (online) https://github.com/PatrickJMcLoughlin/Evaluating-Student-Involvement-in-Sea-Level-Data-Rescue-/tree/main (Accessed 8 March 2025).

McLoughlin PJ, **McCarthy GD**, **Nolan G** ***et al***. 2025. Tutorial (Video) Resources to Support Student Training in Tidal Data Digitization. Zenodo: Geneva, Switzerland. Version v1. https://zenodo.org/records/14905518 (Accessed 10 March 2025).

McLoughlin PJ, **McCarthy GD**, **Nolan G** ***et al***. 2024a. The accurate digitization of historical sea level records. *Geosci. Data J.* **11**: 790–805. https://doi.org/10.1002/gdj3.256 (Accessed 11 November 2024).

McLoughlin P, **McCarthy GD**, **Nolan G** ***et al***. 2024b. The Accurately Digitized Dún Laoghaire Harbour Marigram (Sea Level) Records (1925–1931) [Dataset]. PANGAEA, Data Publisher for Earth & Environmental Science, University of Bremen: Bremen, Germany. https://doi.org/10.1594/PANGAEA.967078.

O'Donnell C. 2023. Exploring the experiences of K-12 science teachers using citizen science to engage student learners: an interpretive phenomenological analysis. Doctoral dissertation. Northeastern University, College of Professional Studies: Boston, MA. https://repository.library.northeastern.edu/files/neu:4f190w852/fulltext.pdf (Accessed 10 October 2024).

Phillips C, **Walsh D**, **O'Regan K** ***et al***. 2018. Assessing citizen science participation skill for altruism or university course credit: a case study analysis using cyclone center. *Citizen Sci. Theory Pract.* **3**(1): 6, 1–13. https://doi.org/10.5334/cstp.111 (Accessed 11 October 2024)

Pugh DT, **Bridge E**, **Edwards R** ***et al***. 2021. Mean sea level and tidal change in Ireland since 1842: a case study of Cork. *Ocean Sci.* **17**(6): 1623–1637. https://os.copernicus.org/articles/17/1623/2021/os-17-1623-2021.pdf (Accessed 5 October 2024)

Rohatgi, A. 2024. *WebPlotDigitizer* (Version 5.2). https://automeris.io/ (Accessed 18 October 2024).

Ryan C, **Duffy C**, **Broderick C** ***et al***. 2018. Integrating data rescue into the classroom. *Bull. Am. Meteorol. Soc.* **99**: 1757–1764. https://doi.org/10.1175/BAMS-D-17-0147.1 (Accessed 18 October 2024)

Talke SA, **Jay DA**. 2013. Nineteenth century North American and pacific tidal data: lost or just forgotten? *J. Coast. Res.* **29**(6a): 118–127. https://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?params=/context/cengin_fac/article/1078/&path_info=jay_nineteenth_century_north_american_and_pacific.pdf (Accessed 15 November 2024).

Talke SA, **Jay DA**. 2017. Archival Water-Level Measurements: Recovering Historical Data to Help Design for the Future. Civil and Environmental Engineering Faculty Publications and Presentations. 412. PDXScholar. Portland State University Library: Portland, OR. http://archives.pdx.edu/ds/psu/21294 (Accessed 5 November 2024).

Talke SA, **Mahedy A**, **Jay DA** ***et al***. 2020. Sea level, tidal and river flow trends in the Lower Columbia River Estuary, 1853–present. *J. Geophys. Res. Oceans* **125**(3): e2019JC015656. https://doi.org/10.1029/2019JC015656

van der Velde T, **Milton DA**, **Lawson TJ** ***et al***. 2017. Comparison of marine debris data collected by researchers and citizen scientists: Is citizen science data worth the effort? *Biol. Conserv.* **208**: 127–138. https://doi.org/10.1016/j.biocon.2016.05.025

Williams J, **Matthews A**, **Bradshaw E**. 2022. Using citizen science to digitise 3 million hand-written tide-gauge data entries. *EGU General Assembly 2022*, Vienna, Austria, 23–27 May 2022, EGU22-740, https://meetingorganizer.copernicus.org/EGU22/EGU22-740.html (Accessed 13 June 2025).

Wöppelmann G, **Pouvreau N**, **Coulomb A** ***et al***. 2008. Tide gauge datum continuity at Brest since 1711: France's longest sea-level record. *Geophys. Res. Lett.* **35**(22): L22605. https://doi.org/10.1029/2008GL035783 (Accessed 13 June 2025).

World Meteorological Organization. 2016. Climate Services for Supporting Climate Change Adaptation: Supplement to the Technical Guidelines for the National Adaptation Plan Process. WMO-No. 1170. World Meteorological Organization: Geneva, Switzerland.

Correspondence to: Patrick Mcloughlin
patrick.mcloughlin.2014@mumail.ie

doi: 10.1002/wea.7757