Check for updates

JOURNAL OF QUATERNARY SCIENCE (2026) 1–12 ISSN 0267-8179. DOI: 10.1002/jqs.70091

JQS Journal of Quaternary Science QRA Quaternary Research Association

Quantifying user variability in the traditional X-ray ice-rafted-debris counting technique

NEIL MCDONALD,[1]* TOM BRADWELL,[1] S. LOUISE CALLARD[2] and JAIME L. TONEY[3]
[1]Biological and Environmental Sciences, University of Stirling, Stirling, UK
[2]School of Geography, Politics and Sociology, Newcastle University, Newcastle upon Tyne, UK
[3]School of Geographical & Earth Sciences, University of Glasgow, Glasgow, UK

Received 27 November 2025; Revised 5 June 2026; Accepted 9 June 2026

ABSTRACT: The flux of icebergs is most commonly constrained in the palaeo-record by manually counting gravel clasts in sediment core X-ray images—a trusted method in Quaternary science and palaeoceanography for nearly 40 years. In this study, we explore the reproducibility (user consistency and between-user variability) of the traditional clast-counting technique, something that has rarely been quantified. A participatory study was undertaken whereby 14 individuals counted gravel clasts (>2 mm) on X-radiographs in cm intervals of two different glaciomarine sediment cores: (i) a matrix-supported subaqueous diamicton facies and (ii) a clast-poor to clast-moderate laminated facies. Pairwise comparisons of all 14 users' counts using Spearman's rho indicate statistically significant relationships in all but one case ($p < 0.05$). A strong correlation (rho = 0.895; $p < 0.001$) was also identified when comparing individual users' average clast counts on both images, indicating high internal counting consistency by users but significant differences in terms of overall counts. Substantial variation was also observed between different users' counts on the same image. In the diamicton facies, the standard deviation of all users' counts per cm was 1.91 (range: 0–4.3), while in the laminated facies, the standard deviation per cm was 1.28 (range: 0–3.67). Additionally, several users recorded a lower average clast count in the clast-rich diamicton facies and a higher count in the clast-poor laminated facies relative to other participants. These results may have significant implications when it comes to interpreting past iceberg flux and ice-marginal dynamics. As a way forward, we present a new rigid protocol to improve transparency, consistency and reproducibility for studies implementing this technique. Our findings have implications regarding how manual clast counts should be interpreted and implemented but also strengthen the case for automated clast-counting algorithms to reduce user bias and improve the consistency of IRD detection using X-radiographic images. © 2026 John Wiley & Sons, Ltd.

KEYWORDS: glaciomarine; iceberg calving; ice-rafted debris; sedimentology; user variability; X-radiography

Introduction

Ice-rafted debris (IRD) refers to sediment transported by icebergs and deposited in lakes or oceans. IRD comprises any sediment that reaches the glacier terminus and becomes incorporated into calving icebergs; thus, IRD can encompass the entire spectrum of grain sizes (Andrews, 2000). However, it is typically characterised by sand or gravel-sized sediment (Ruddiman, 1977; Grobe, 1987; Scourse et al., 2009; Bartels et al., 2017; Jennings et al., 2022; Ahn et al., 2024).

A range of different approaches can be used to identify IRD, including sieving, laser particle-size analysis, image greyscale analysis and clast counting or indexing using X-ray images (manual or automated) (Connolly and Ewing, 1965; Grobe, 1987; Hansen et al., 2015; Bartels et al., 2017; Reilly et al., 2019; McDonald et al., 2022; Jasper et al., 2024). Each of the different sampling strategies can bias IRD outputs in different ways, with some methods overestimating and others underestimating IRD in certain circumstances (McKay et al., 2022). This study investigates the variability and reproducibility of one of the most common IRD quantification techniques—clast counting on X-radiographs (XR) (Grobe, 1987). This method involves counting clasts on XR images (>2 mm) within horizontal cm-intervals. An increase in the relative abundance of gravel clasts is taken as a representative increase in ice-rafting, while a decrease indicates reduced ice-rafting (Grobe, 1987; Fig. 1).

Variations of the original method described by Grobe (1987) are occasionally implemented; for example, the grain-size threshold can be reduced to 1 mm for determining IRD, allowing increased detection (Grobe et al., 1990; Yamamoto et al., 2008; Weber et al., 2014; Hillenbrand et al., 2021), and the horizontal sampling interval can be increased to 2 cm or 5 cm (Callard et al., 2018; Jennings et al., 2022; Pesjak et al., 2023; Girard et al., 2024). Others go further, distinguishing gravel-size classifications, such as Smith et al. (2021), who quantify gravels in bins of 2–8 mm and 8 mm or greater. However, the traditional approach involves counting clasts >2 mm on XR images in 1 cm intervals (e.g. Grobe, 1987; Graham et al., 2017).

Supervised and automated clast counting/indexing routines using XR or X-ray computed tomography (CT) are becoming increasingly popular (e.g., Bartels et al., 2017, 2018; Reilly et al., 2019; Cederstrøm et al., 2021; Jasper et al., 2024). It is assumed that these techniques will reduce, user bias in clast counting. However, these routines require human training and/or validation, which is subject to error and bias, potentially affecting the accuracy of clast counts. Additionally, manual clast counting (Grobe, 1987) is still commonly

*Correspondence: N. McDonald, as above.
Email: neil.mcdonald@stir.ac.uk

© 2026 The Author(s). *Journal of Quaternary Science* Published by John Wiley & Sons Ltd.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

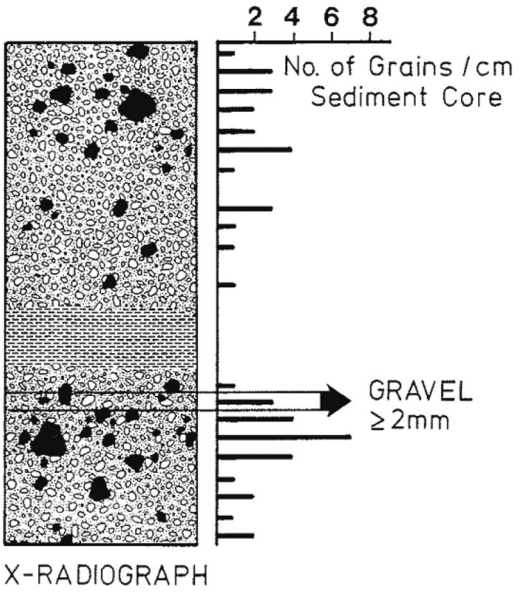


Figure 1. Illustration of how to perform clast counting manually from X-radiographs. Reproduced from Grobe (1987).

implemented in IRD studies (Graham et al., 2017; Jennings et al., 2022; Girard et al., 2024; López-Quirós et al., 2024). Therefore, it is important to understand the factors that may result in variations in clast counts between users and determine the variability and robustness of manual clast counting. In this study, we investigate the variability or uncertainty in the traditional approach and, in doing so, quantify observer bias in the manual X-ray IRD counting (Grobe, 1987) method.

Why might variability and inconsistency occur?

The XR clast-counting (>2 mm) method has been consistently implemented for nearly 40 years for the identification of IRD as part of studies reconstructing climate and oceanic circulation changes, ice-sheet growth and decay, and floating-ice margin (sea ice or ice shelf) presence and removal (e.g., Rack et al., 1996; Dowdeswell et al., 2000; Knies et al., 2001; Andrews et al., 2016; Weber et al., 2014; Graham et al., 2017; Jennings et al., 2022; López-Quirós et al., 2024; Callard et al., 2025). However, when a user implements clast counting, they are faced with many decisions that might influence the results, for example:

1. Is this object a clast? Could it be a shell fragment, an imaging artefact or an amalgamation of finer grains?
2. Should clasts be counted in all intervals or only once if they overlap the sampling grid and appear in more than a single interval?
3. Is this object a single clast or two (or more) overlapping clasts? (e.g., Hogan et al., 2016; McKay et al., 2022).
4. Is it most appropriate to measure the longest or shortest axis of each grain?
5. Is it necessary to physically measure every clast >2 mm? Or is size estimation from a scale bar or grid appropriate?

It is conceivable that these decisions might affect the variability and consistency of the technique, and experience level alone may not be enough to overcome them. None of these factors are specifically addressed in the literature and, at only 3 pages long, the original paper does not provide details. The only questions that can perhaps be answered from Grobe (1987) is that it appears that gravels were only counted in the first centimetre slice in which they appear (or the slice in which they mostly appear). This is deduced directly from a figure in the article (Fig. 1), but is not stated in the text. However, the figure is quite schematic and leaves some uncertainty in the counting process. Without strictly addressing this and other concerns outlined above, it is reasonable to assume that there is potentially a lack of consistency among researchers who use this technique.

The possibility of user bias in IRD clast-counting studies has only recently attracted attention (McKay et al., 2022; Jasper et al., 2024). McKay et al. (2022) found very strong agreement between two users with Pearson's correlation coefficients of 0.961 and 0.927 (when comparing clast counts in 2-cm bins). They suggest some factors for discrepancies between users such as overlapping grains, low-radiodensity clasts overlying a matrix comprising intense burrowing or lithological complexity. They also call for greater standardisation of criteria between working groups for determining IRD.

Jasper et al. (2024) used two different teams of researchers to produce two individual data sets on the same XR images (grain size >1 mm). These outputs were generally well matched with some exceptions, with a Pearson's correlation coefficient of 0.388, $p < 0.05$, when comparing clast counts in 1-cm bins (increasing to 0.555 when compared in 5-cm bins, likely due to differences in the cm-interval in which clasts were counted). However, in some horizons, discrepancies between the two teams' clast counts differed by as much as 16 clasts (Jasper et al., 2024). This indicates only moderate agreement, with considerable variation between users. We aim to provide greater methodological standardisation by investigating whether a larger participatory study could provide insight into why these variations occur, for example, whether sediment type or user experience affects clast counts.

The original method was exceptionally quick when first introduced in the 1980s, being substantially faster and less labour-intensive than sieving (Grobe, 1987). However, with the introduction of new, automated or supervised routines that allow for clast counting/clast indexing from X-ray images (Bartels et al., 2017, 2018; Reilly et al., 2019; Cederstrøm et al., 2021; Jasper et al., 2024) or grey value standard deviation (McDonald et al., 2022), this may no longer be the case. Automated routines are, in some cases, near instantaneous or in other cases significantly faster than traditional on-screen clast counting. Therefore, we also aim to quantify the average time required to complete manual clast counting on sediment core XR images in this study.

The implications of variability in X-radiograph clast counting

The XR clast-counting technique is typically used to infer relative past iceberg flux—increases and decreases in past iceberg activity (e.g., Grobe, 1987; Rack et al., 1996; Grobe and Mackensen, 1992; Weber et al., 2014; Graham et al., 2017). However, it is also commonly used to determine other glaciomarine conditions, such as grounding-line proximity and ice-marginal dynamics (e.g., Hogan et al., 2016; Sheldon et al., 2016; Smith et al., 2017, 2021; Jennings et al., 2022).

The proximity to marine ice-sheet margins can sometimes be reconstructed from the relative abundance of gravel clasts (or IRD) in X-radiographs (e.g., Callard et al., 2018; Smith et al., 2019; Syring et al., 2020; Bradwell et al., 2021). Clast-rich, massive to stratified sandy units may indicate intense iceberg rafting or rainout in the proximal zone, while laminated to massive muddy units that range from clast-poor to clast-rich reflect varying intensities of iceberg rafting in more distal settings, relative to background sedimentation (e.g., meltwater plumes, turbidites and bottom currents—Hogan et al., 2016; O'Regan et al., 2021; Smith and Hogan, 2025;

© 2026 The Author(s). *Journal of Quaternary Science* Published by John Wiley & Sons Ltd.

10991417, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jqs.70091 by Enlighten Team - University Of Glasgow , Wiley Online Library on [06/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Ó Cofaigh et al., 2025). However, other proxies such as sedimentary structures, foraminifera, shear strengths and stratigraphic associations are also critical when inferring ice-marginal proximity (Anderson, 1999; Smith et al., 2019). Without such context, distinguishing between intense iceberg rafting in a more distal setting and ice-proximal deposition can be difficult. Nevertheless, user bias in XR clast counting remains an important source of potential inconsistency when interpreting glacier proximity in the marine realm.

Quantifying the abundance of gravel clasts on X-ray images is becoming an increasingly common technique for constraining the presence, or absence, of past ice shelves (Reilly et al., 2019; O'Regan et al., 2021; Smith et al., 2021; Jennings et al., 2022; Clark et al., 2024; Callard et al., 2025). Ice shelves typically filter gravel clasts, depositing IRD—in this case ice-shelf rafted debris (Powell 1984)—close to the grounding line, resulting in a marked absence of gravel (or very rare) clasts in more distal sub-ice-shelf sediments (Domack and Harris, 1998; Smith et al., 2017, 2019, 2021; Reilly et al., 2019; Jennings et al., 2022).

Similarly, shorefast–sea-ice margins can block IRD delivery resulting in units that are gravel-clast-free (or very rare in IRD) (Dowdeswell et al., 2000). Seasonal sea-ice signals can also be differentiated by the rhythmic nature of gravel-sized IRD using XR clast counting—often with mud–diamicton alternating sequences (Dowdeswell et al., 2000). Conversely, the removal or absence of perennial floating-ice margins and the opening or presence of iceberg calving bays may be differentiated based on more persistent and abundant IRD signals (Hogan et al., 2016; Smith et al., 2019; Reilly et al., 2019).

Significant variability between users undertaking XR clast counting may lead to over- or underestimations of iceberg flux and misinterpretations of palaeo-environmental conditions. This matters because the tracking of past marine-terminating glacier and ice-stream deglaciation is essential for projecting the long-term behaviour of contemporary marine-terminating glaciers and ice streams. It is also critical for predicting how glaciers and ice streams will respond to the loss of their floating-ice margins. If users are inconsistent in their quantification of IRD, misinterpretations of sediments are possible and thus reconstructions of past ice-stream behaviour could be skewed, leading to erroneous inferences regarding the nature of sedimentation and deglaciation more broadly.

This study aims to improve our understanding of how user variation (including experience level) can impact the reliability of the technique. The results will allow clearer insight into how best to interpret and implement XR clast counting to most reliably quantify IRD (using the Grobe (1987) method). This is additionally important, given that consistent manual clast counting is imperative for the calibration of automated routines. Specifically, in this paper we:

1. Robustly quantify, for the first time, the degree of user variability in clast counting (as an IRD proxy) in different sediment facies in a participatory study of 14 individuals.
2. Examine whether user-experience levels affect the consistency of clast counting on X-radiographs using 3 different groups (experienced, moderate experience and no experience).
3. Determine the time taken to undertake clast counting on two sediment core images from different glaciomarine facies.
4. Provide a more rigorous protocol to improve the robustness of clast counting as an IRD proxy.

Methods

Clast-counting participatory study

A comparative study of multiple different researchers undertaking clast counting on two different XR images (Fig. 2), 93 cm and 100 cm in length, was undertaken. The piston cores selected in the comparative study are from glaciomarine sediments deposited in the central Minch, NW Scotland, during the last glacial cycle (Fig. 2). Both sediment facies are very poor in marine carbonate (and/or organic material); this was key for this study, as the inclusion of shells can make the clast-counting process significantly more challenging. The two cores differ considerably in clast abundance—one is a clast-rich but matrix-supported subaqueous diamicton (JC123-032PC-section-8); the other is a clast-poor to clast-moderate strongly laminated 'ice-proximal' glaciomarine facies (JC123-031PC-section-7). Both facies were originally described by Bradwell et al. (2021).

Selecting these two different sediment types should help to understand the effect of different sediment facies on clast counting (i.e., laminated vs. massive sediment structure, with different clast abundances). XR-positive images were used for the clast-counting analysis and prepared following the methodology outlined in McDonald et al. (2022) to enhance image clarity and contrast. Each count was conducted independently and there was no communication between users during the study.

Participatory analysis

For the participatory study, a horizontal 1 cm grid was added to the images to divide the core up into slices, allowing a like-for-like comparison between all participants (Fig. 2). Individuals were asked to complete a small questionnaire to understand their educational background and previous experience level. 15 participants were approached with varying

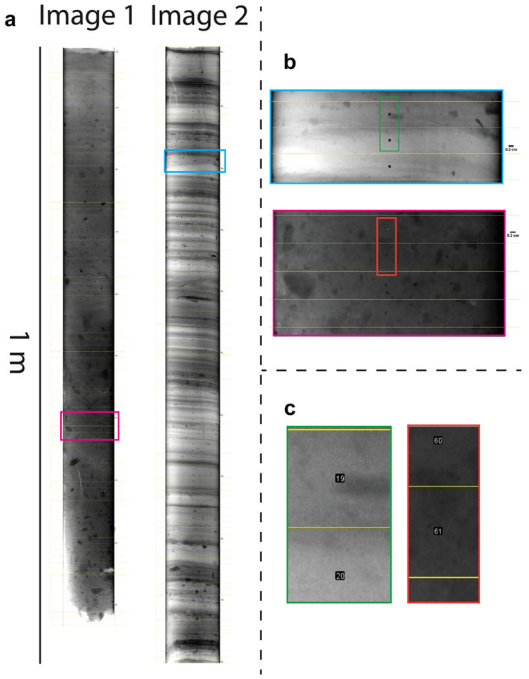


Figure 2. Sediment facies and core images provided to participants in this study. Panel (a) shows the two images in full with the sedimentary facies. Panel (b) scale bar visible on images; the scale bar appeared in the 10 cm interval on the right-hand side of core images (0.2 cm). Panel (c) pre-labelled depth intervals to ensure consistency between different users. Inset images are colour-coordinated. [Color figure can be viewed at wileyonlinelibrary.com]

10991417, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jqs.70091 by Enlighten Team - University Of Glasgow , Wiley Online Library on [06/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

degrees of expertise and experience levels. All 15 were at a minimum enroled on a PhD programme. Unfortunately, one person's data could not be used, as they misread the instructions and provided a count of all large, discernible sand grains on the XR images (<2 mm) as well as gravel clasts (>2 mm). The remaining 14 participants had varying expertise and experience levels. Finally, a validation count was undertaken using a rigid protocol, outlined in this study. This count was undertaken blind (by the lead author) without reference to the results of the other participants. The validation count and the mean-of-the-mean values for all participants' counts per cm on each image provide the best indication of the 'true' count; these were used to assess overall accuracy.

For analysis purposes, the participants were grouped into categories based on expertise/previous experience. These groups are labelled 'experienced' (participants 1–5), 'moderate experience' (participants 6–10) and 'no experience' (participants 11–14). Participants were provided with a short background to the study, with instructions on how to perform the task and a link to the Grobe (1987) paper (Supplementary files—Appendix A).

Statistical analyses

Each user's counts on the two images were initially explored using the Shapiro–Wilk test. This test compares the observed data for each participant to a univariate normal distribution. As count data were almost unanimously found to be not normally distributed (only 2 of 28 tests returned acceptable normal distributions at $p = 0.05$; Fig. 3), the results of each participant were compared using a scatter matrix and statistically analysed using the non-parametric Spearman's rank correlation (rho) test. The following questions were investigated:

1. Whether the clast-counting method is consistently replicated (i.e., highly correlated) by different users, when counting clasts in cm intervals on the same images.
2. Whether the clast-counting method is consistently replicated (i.e., highly correlated) by the same user on different images. That is to say, whether individual users produce internally consistent results when quantifying gravel abundances on two different images.

Scatter matrices were produced to compare all counts on each image between different users (91 individual comparisons for each image) using GGally in R (Schloerke et al., 2025) to provide a visualisation of the strength of relationships and a statistical test of any relationship found using Spearman's rank correlation. The standard deviation was also selected to assess the variability of all 14 users counts over every cm interval on both images. This serves to determine the spread of users results around the mean, allowing insight into the variability of the technique.

Results

When analysing the results of 14 participants undertaking XR clast counting on two different glaciomarine facies (after Grobe, 1987), statistically significant correlations are identified between all users almost unanimously (Figs. 4 and 5). When comparing the counts of all 14 users with each other, making a total of 91 pairwise comparisons, Image 1 (the diamicton facies) returned Spearman's rho values from 0.190 to 0.721, a very wide range. However, only one comparison returned a p-value that was not statistically significant (at $\alpha = 0.05$).

A similar trend was observed for Image 2 (the laminated facies), where all 91 comparisons returned statistically

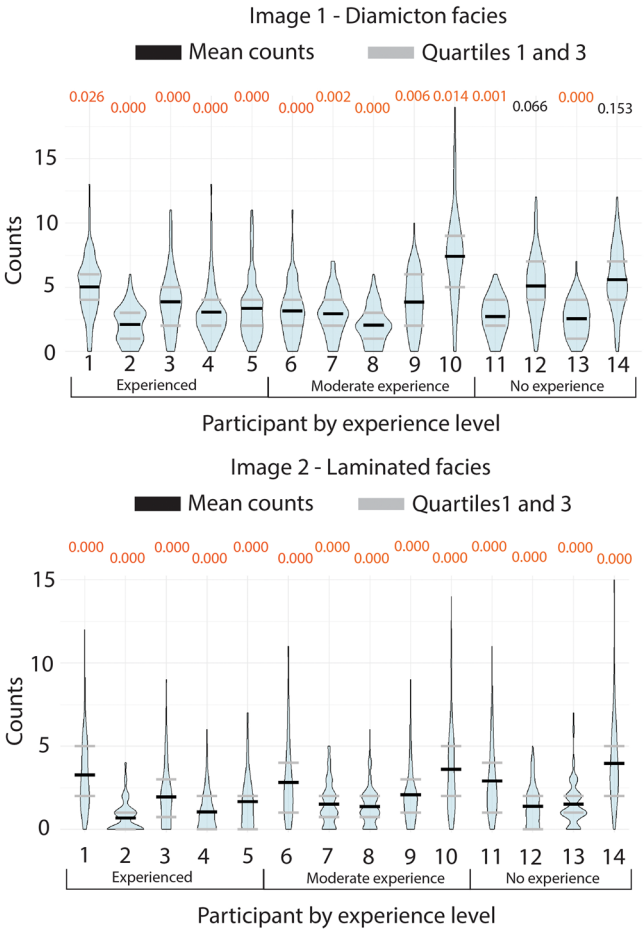


Figure 3. Participant count distributions illustrated by frequency. Wider violin sections indicate more frequent user counts. For example, a wider violin at 5 (*Y*-axis = counts) for Participant 1 than at 10 indicates that this user counted 5 clasts more frequently than 10 clasts. Mean numbers of counts per image (or stratigraphic facies) are indicated by dark bars, and quartiles 1 and 3 are indicated by light bars. Shapiro–Wilk tests of normality p-values are indicated by orange- and black-coloured values above each violin. Orange-coloured values indicate that the data are not normally distributed; black values indicate normally distributed data. [Color figure can be viewed at wileyonlinelibrary.com]

significant p-values ($\alpha < 0.05$), with Spearman's rho values ranging widely from 0.252 to 0.786, indicating relatively strong phasing relationships overall between different users. Interestingly, the weakest correlation, with a Spearman's rho of 0.252, comes from the two most experienced users (Participants 1 and 2), while the users with the lowest-level experience maintain strong correlations with most participants in this study.

When analysing the statistical variability between counts, the results of different users are stark (Figs. 6 and 7). Greater absolute count variability is observed in the diamicton facies (Image 1), with an average standard deviation of 1.91 per cm (standard deviation range = 0–4.3) (Fig. 7). On the other hand, the laminated facies (Image 2) returned an average standard deviation of 1.28 per cm (standard deviation range = 0–3.7) (Fig. 7). In general, the greater the clast abundance, the greater the absolute variability (higher standard deviation). Due to the generally strong correlations, but high variability in clast counting between different users, it can be determined that all participants' clast counts were relatively consistent with regard to phasing (broadly increasing and decreasing simultaneously) but not necessarily consistent in terms of absolute count (magnitude).

© 2026 The Author(s). *Journal of Quaternary Science* Published by John Wiley & Sons Ltd.

10991417, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jqs.70091 by Enlighten Team - University Of Glasgow , Wiley Online Library on [06/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

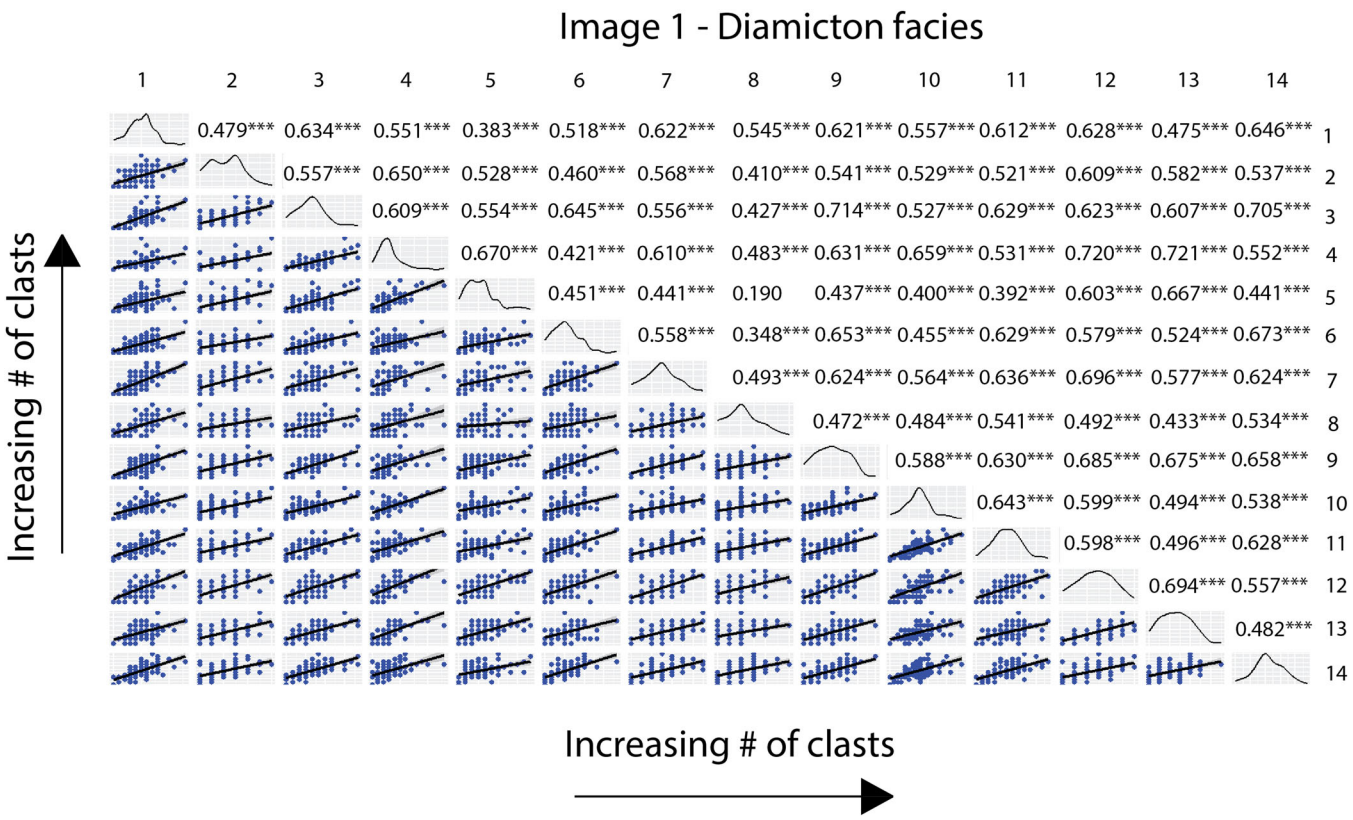


Figure 4. Cross-comparison (scatterplot matrix) of different users in the diamicton facies (Image 1). Values relate to Spearman's rho correlation coefficient, *p-value < 0.05; **p-value < 0.01; and ***p-value < 0.001. Participants are numbered 1–14 based on experience level (from high to low). [Color figure can be viewed at wileyonlinelibrary.com]

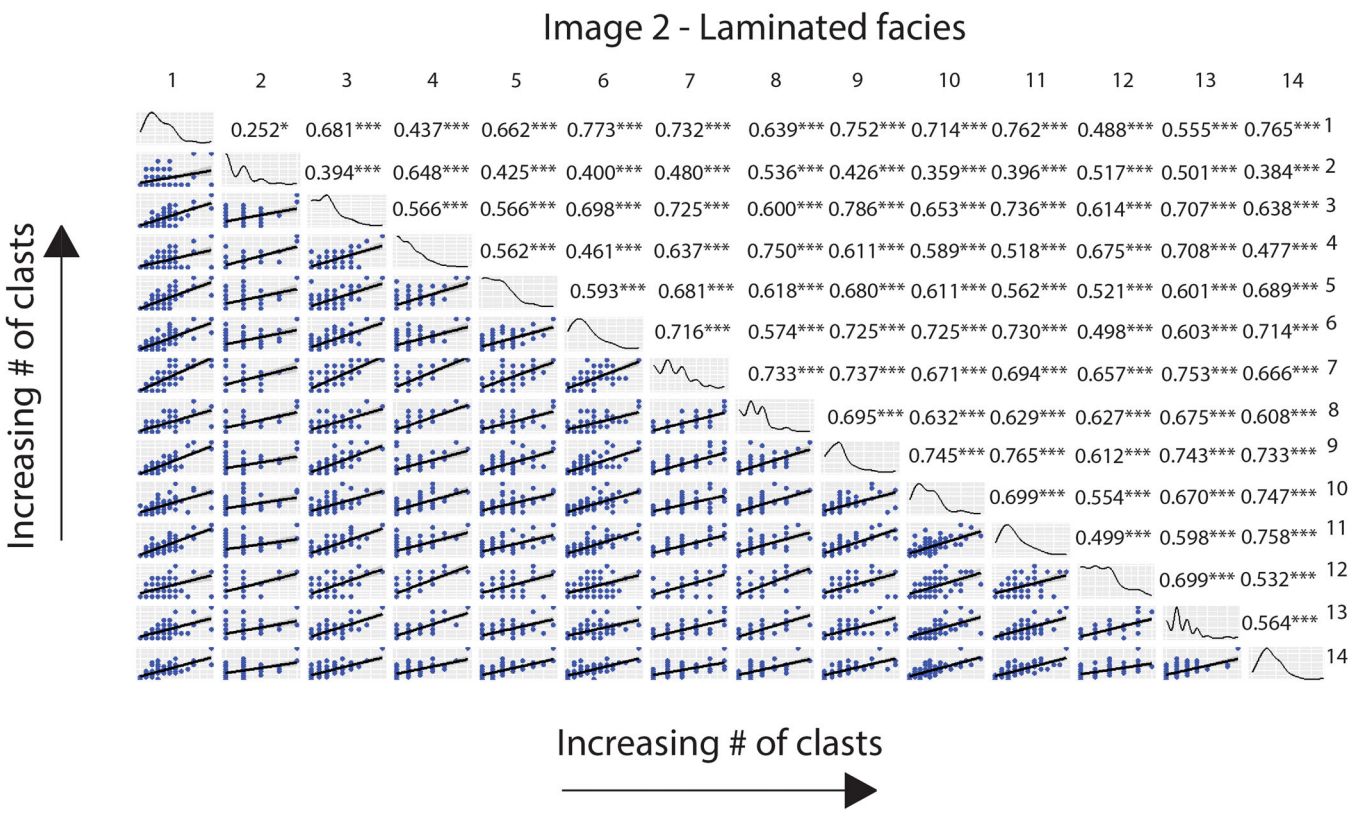


Figure 5. Cross-comparison (scatterplot matrix) of different users in the laminated facies (Image 2). Values relate to Spearman's rho correlation coefficient; *p-value < 0.05; **p-value < 0.01; and ***p-value < 0.001. Participants are numbered 1–14 based on experience level (from high to low). [Color figure can be viewed at wileyonlinelibrary.com]

© 2026 The Author(s). *Journal of Quaternary Science* Published by John Wiley & Sons Ltd.

10991417, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jqs.70091 by Enlighten Team - University Of Glasgow , Wiley Online Library on [06/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

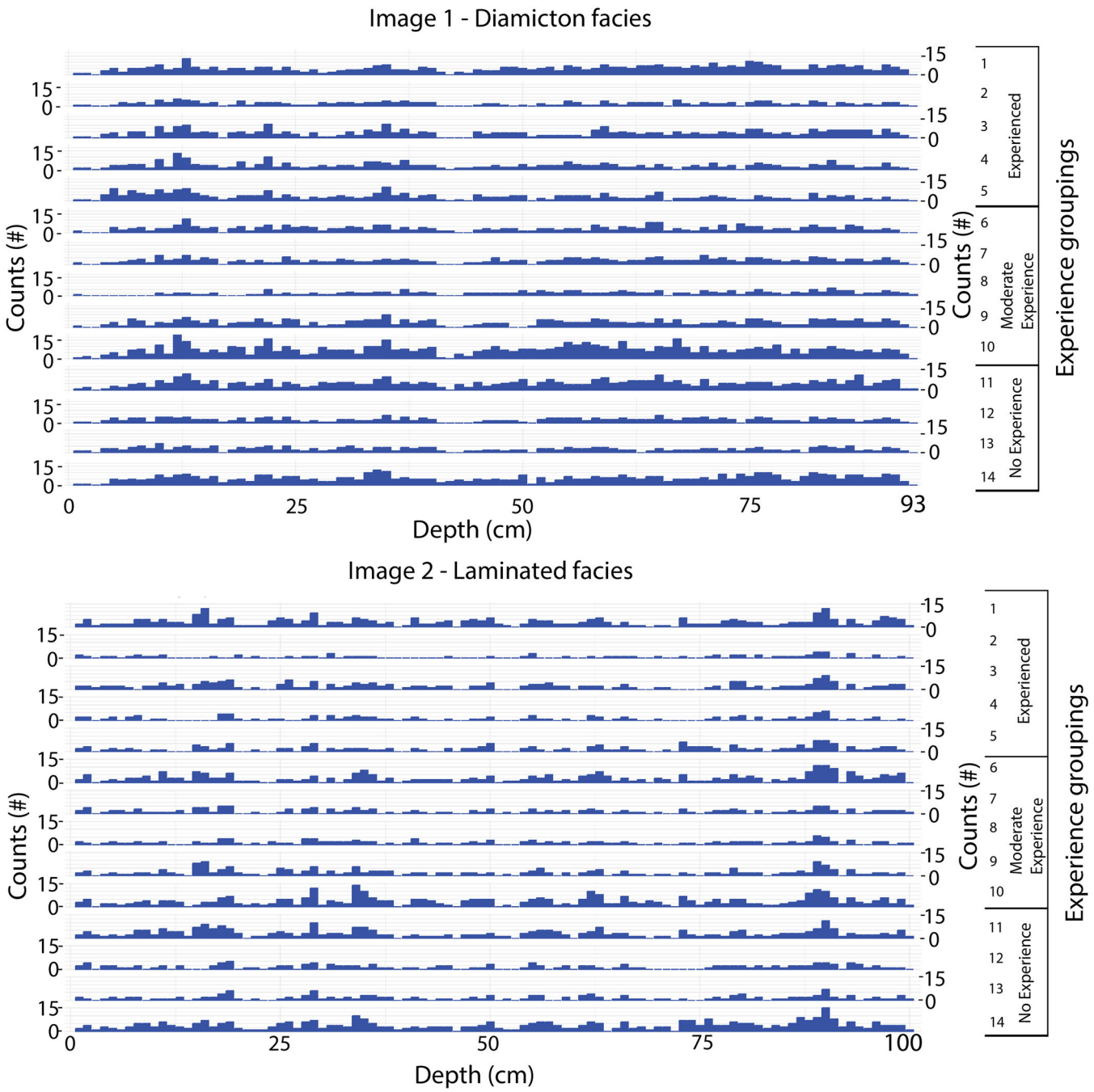


Figure 6. Participant counts by core depth. The *Y*-axis is ordered by level of experience. Note the consistent *Y*-axis scale between users for each image. [Color figure can be viewed at wileyonlinelibrary.com]

The major discrepancy between user results stems from some individuals being more conservative counters than others (Figs. 3, 6 and 8). Thus, when comparing results from two different facies, the average counts from different users can vary significantly. The mean clast counts per cm slice equalled 3.76 in the diamicton (Image 1) and 2.13 in the laminated facies (Image 2). However, there is significant overlap when comparing the counts of different users, with some users counting fewer clasts in the diamicton facies and more clasts in the laminated facies than others (Figs. 6 and 8). For example, participants 1, 6, 10, 11 and 14 counted more clasts per cm slice in the laminated facies than participants 2, 8, 12 and 13 counted in the diamicton facies per cm slice (Fig. 8).

A reasonable degree of variability was observed with the time taken to complete the task. For Image 1 (diamicton facies), the mean time to undertake counting on the entire image was 22 min and 8 s over 93 cm intervals. Image 2 (laminated facies) had a lower mean completion time of 16 min and 31 s over a longer core section (100 cm). This provides the first recorded insight into the time taken to perform on-screen manual clast counting (Fig. 9).

Discussion

Reproducibility of X-ray clast counting

Overall, given the variation between respondents, we find that clast counting is a semi-quantitative discriminator of IRD using XR images, and is consistent with regard to identifying overall trends. In studies with very small sample sizes (n = 2), McKay et al. (2022) and Jasper et al. (2024) found differing agreement between users. McKay et al. (2022) recorded very high statistical agreement, whereas Jasper et al. (2024) recorded lower correlations, similar to

10991417, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jqs.70091 by Enlighten Team - University Of Glasgow , Wiley Online Library on [06/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

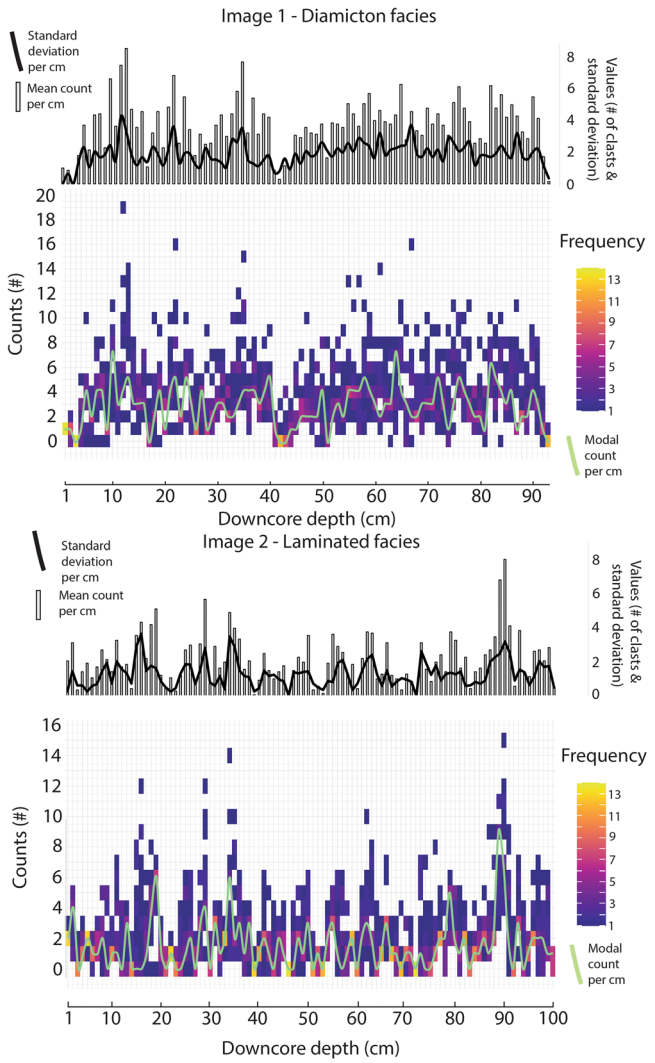


Figure 7. User variability of clast counting on both X-radiographs. The black line indicates the standard deviation of all participants' counts per cm; bars indicate the mean of all users' counts for each cm slice on Image 1 (top) and Image 2 (bottom). Heatmaps indicate the variability between users, with warmer colours indicating a greater number of users. The green line indicates modal counts per cm interval. [Color figure can be viewed at wileyonlinelibrary.com]

those in this study. It should be noted that both McKay et al. and Jasper et al. report Pearson's correlation coefficients, whereas this study uses Spearman's rho to explore correlations. Parametric and non-parametric correlation coefficients are not directly equivalent (Chen and Popovich, 2002). In addition to this, the stronger correlations ($r > 0.9$) identified by McKay et al. (2022) may have resulted from both users using 1-cm thick sediment core slabs, clear communication between users, similar training histories or similar levels of experience, or the fact that, in some marine sediment cores (e.g., more distal glaciomarine facies), clast counting is not as challenging as other sediment types (e.g., more proximal glaciomarine facies).

The counts in McKay et al. (2022) were undertaken on two cores that appear to be *potentially* more variable in IRD abundance than the cores selected in this study, with some horizons devoid of gravel clasts and others very clast rich (particularly in core LC42). In the images used in this study, IRD is consistently present throughout the sediment facies, albeit more variable in core 31PC. It is clear from this study that user counts are more consistent in zones with less IRD and less consistent in units with more IRD (Fig. 7). This may have perhaps played a small role in the higher correlations seen in McKay et al. (2022).

Jasper et al. (2024) found similar levels of agreement to the correlations recorded in this study, albeit using a clast size threshold of 1 mm to determine clasts and two teams working in parallel to produce the data sets. Their cores appear to have had a fairly consistent spread of IRD and non-IRD features, such as pyrite-infilled burrows, which *may* have led to lower correlation coefficients when binned in 1-cm intervals ($r = 0.4$), more comparable to the typical correlation coefficients returned in this study (rho = 0.4–0.7).

Does expertise increase the accuracy of X-radiographic clast counting?

It appears that expertise bears little impact on the accuracy and consistency of clast counting. This is evident in both the distributions and correlations, and also in the variability between average counts per image between users of varying experience (Figs. 3–7). This is a key finding of this research. However, if the variability in clast counting is not a function of expertise, it must be related to other factors.

Figure 10 provides an indication of consistency between different users; specifically, the plot shows the relationship between different participants' mean counts in both facies (Spearman's rho = 0.895; $p < 0.001$). We find that users closest to the trendline (3, 5, 7, 8, 9, 10, 11, 12 and 13) are more consistent in their method and therefore more consistent in their counts between facies. In addition to this, Participants 3 and 9 returned counts very close to the validation count and the overall mean value for both images, indicating the highest accuracy counts in the study (Fig. 10). However, this also suggests that expertise does not have a significant impact on the consistency (and potentially) and the accuracy of the clast-counting method using two sedimentologically different glaciomarine cores. We suggest, therefore, that ensuring a strict methodical approach is more important than experience.

Given this between-user consistency, it is clear that participants can be generally classed as 'high' or 'low' counters in this study. For example, participants 1, 10, 11 and 14 can be classed as 'high counters' and participants 2, 4, 7, 8, 12 and 13 can be classed as 'low counters' (Fig. 10). There is consistency between both images in this regard, with a strong, statistically significant correlation. This shows that the Grobe method is reliable when identifying IRD trends between all users, but much less reliable at quantifying the magnitude of IRD.

This study also indicates that user experience does not impact the applicability of the method. When reporting clast counts on the laminated facies, the two most experienced users had the lowest correlation with each other (Spearman's rho = 0.252). This was an unexpected result, but probably stems from each user's differing interpretations of what they consider as clasts on XR images. Participant 2 appears to have almost exclusively registered clasts that appear very dark, and therefore, have a higher radiodensity, relative to the matrix sediment on XR (Fig. S1). This may be because their expertise in other settings led them to interpret low-radiodensity (less dark on XR) clasts as non-IRD.

Alternatively, other factors such as operating with lower screen brightness and differences in eyesight between different users may lead to 'high' versus 'low' counting. Participant 1 appears to have been more generous with the 2 mm threshold, counting clasts that are close to and possibly below the 2 mm threshold. Participant 2 appears to have been more conservative by only counting clasts clearly >2 mm. On the other hand, the least-experienced user (Participant 14) had a high degree of correlation with all participants in the laminated facies, including the most-experienced users. This suggests that

© 2026 The Author(s). *Journal of Quaternary Science* Published by John Wiley & Sons Ltd.

10991417, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jqs.70091 by Enlighten Team - University Of Glasgow , Wiley Online Library on [06/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

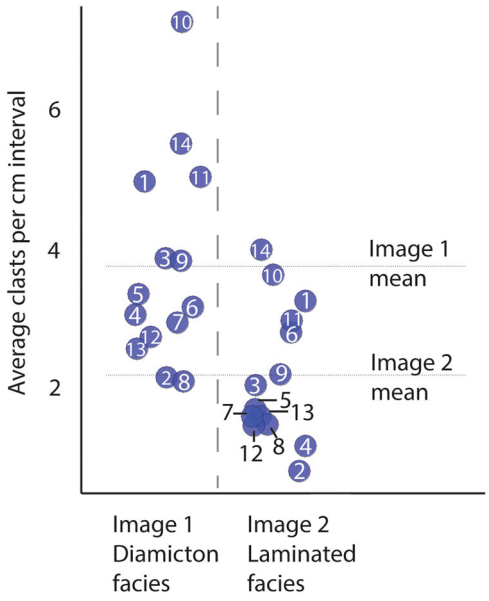


Figure 8. Average clasts counted per cm interval for each user. Numbers refer to the experience ranking of different users. Mean clast counts per cm are indicated by dotted lines (values: Image 1 mean clast per cm = 3.76; Image 2 mean clast per cm = 2.13). Diamicton facies and laminated facies have statistically significantly different clast counts (Mann–Whitney U value = 34, $p < 0.05$). [Color figure can be viewed at wileyonlinelibrary.com]

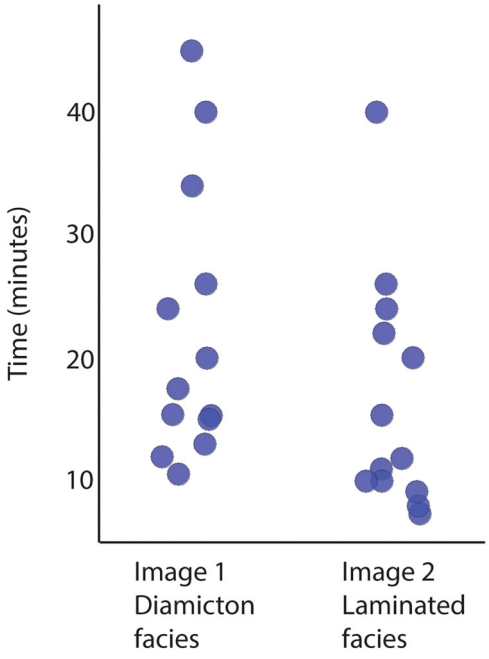


Figure 9. Time taken to perform a task on Image 1 and Image 2, respectively. Scatter points represent users. [Color figure can be viewed at wileyonlinelibrary.com]

regardless of expertise, with only introductory training, any early-career analytical scientist can undertake XR clast counting and generate consistent results.

This study was undertaken on two images that had low biogenic marine carbonate material content, that is, proximal glaciomarine sediment with significant fine muddy silts due to meltwater delivery (Bradwell et al., 2019, 2021). In more shell-rich sediments, it is possible that user expertise may play an important role in providing more accurate results. However, given the discrepancies between experienced users seen here, this remains uncertain.

Implications of user variability in the XR clast-counting technique

The high degree of variability between users highlights the potential for misinterpretation of certain sediments. Although the facies examined in this study differ significantly when considering average clast counts per cm across all users (Mann–Whitney $U = 34$, $p < 0.05$), there remains substantial overlap between individual users' counts (Fig. 8). For instance, when analysing users' counts from Image 2, 'high counters' often produced average clast counts per cm closer to the mean value of Image 1 (3.76), while 'low counters' on Image 1 reported values closer to—or even below—the overall mean for Image 2 (2.13). As a result, interpretations of iceberg flux may be subject to under- or overestimation, and this may affect palaeoenvironmental interpretations depending on user bias.

Moreover, when analysing the results of Image 2, the two most experienced users—Participant 1 and Participant 2—recorded mean clast counts of 3.28 and 0.69 clasts per cm, respectively. Although there is a weak statistically significant correlation between these users (Spearman's rho = 0.252, $p < 0.05$), the deviation in clast counts could potentially have important implications for palaeo-environmental interpretation. For example, very clast-poor laminated facies with intervals of IRD absence (as recorded by Participant 2) could perhaps be interpreted as evidence of ice-shelf cover (e.g., Smith et al., 2021; Jennings et al., 2022). In contrast, a more clast-rich laminated sediment unit (with high and persistent gravel clast abundances of 3.28 clasts per cm as recorded by Participant 1) would typically suggest the absence of a perennial floating-ice margin (Ó Cofaigh and Dowdeswell, 2001; Hogan et al., 2016; Smith et al., 2019; Jennings et al., 2022). The repetitive increases and decreases (or rhythmicity) of their gravel counts could also suggest former seasonal sea-ice cover (Dowdeswell et al., 2000; Fig. S1). Such discrepancies highlight how user-to-user variability could generate competing interpretations of the same deposits, especially when evaluating the potential presence of floating-ice margins. Significant differences in user counts—both across the two images and between the most experienced users on the laminated facies (Image 2)—clearly illustrate how differences in clast counts could translate into conflicting reconstructions of ice-shelf or sea-ice cover, calving-bay conditions and ice-margin proximity (Fig. S1).

A protocol to improve consistency between users when undertaking the Grobe (1987) IRD method

Numerous respondents reported differing methods for gravel classification and regarding which slice (or cm interval) clasts were counted in. Several users anecdotally testified to undertaking subtly different approaches. For example, some users indicated that for clasts that overlap different cm slices, they:

1. Counted all clasts in each sample slice (whole or partial).
2. Counted clasts in the interval in which they mostly appeared—also implemented by participant groups in Jasper et al. (2024).
3. Counted clasts only within the first slice in which they appeared.

This highlights the need for a more systematic approach when implementing clast counting.

Most respondents considered the axis of interest for clasts to be the a-axis (longest visible axis) (as performed by McKay et al., 2022), although, to be considered a gravel clast (>2 mm), the b-axis should technically be considered (the intermediate axis), as the b-axis would be the critical

© 2026 The Author(s). *Journal of Quaternary Science* Published by John Wiley & Sons Ltd.

10991417, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jqs.70091 by Enlighten Team - University Of Glasgow , Wiley Online Library on [06/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

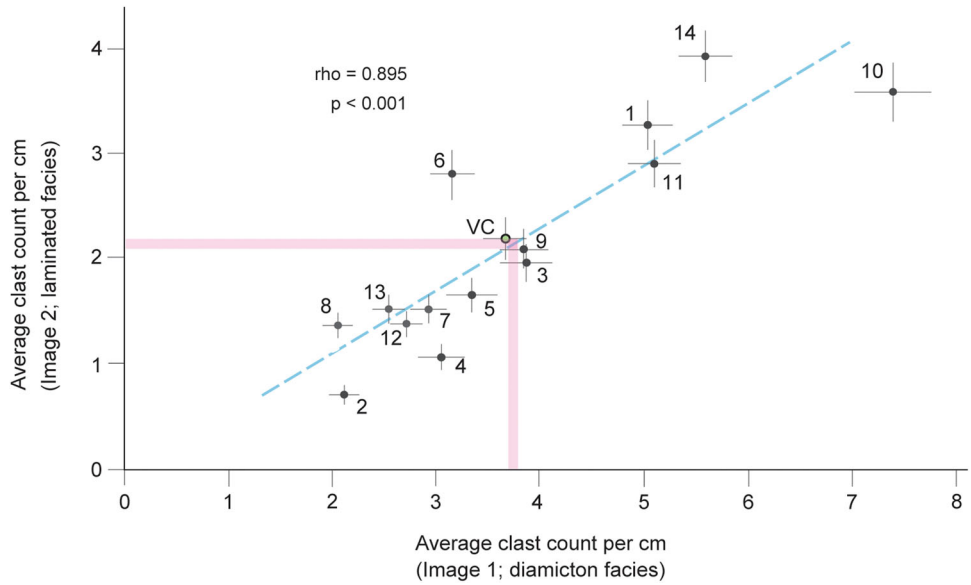


Figure 10. Comparison of average clast counts for each user on both images. This figure highlights that most users replicate their counting method consistently in different facies, but some users count more conservatively than others. The pink shaded area indicates the mean counts of 14 participants. Error bars show one standard error for each participant's mean count for each image. VC = validation count undertaken adhering to the newly devised protocol. Participants 3 and 9 are closest to the overall mean of all users' counts and the validation count, suggesting higher accuracy for these users. [Color figure can be viewed at wileyonlinelibrary.com]

dimension for a gravel to be classified via sieving. However, XR images are only 2-dimensional and do not allow investigation of the entire clast. As glacially derived clasts can often be bullet-shaped and elongate, it seems reasonable to consider the longest axis as the critical dimension. Most respondents also reported estimating clast size with the naked eye, while others reported measuring the clasts using a reference filter or measuring with a ruler. These factors are considered as the main reasons for the high variability (standard deviation) between clast counts.

There are also fundamentally challenging factors such as the fact that clasts may become wholly or partially 'hidden' behind others on XR images (Hogan et al., 2016; McKay et al., 2022); this is also more likely when sediment is rich in IRD. This may have played a role in the higher standard deviation in the diamicton facies where there are more clasts and subsequently more chance of clasts 'overlapping' (Figs. 2 and 7). However, the standard deviation becomes inherently greater when more clasts are present, as user biases due to different operating decisions become amplified. That is to say, even when clasts are sparsely distributed but in high abundances (e.g., some horizons of the laminated facies), there is a greater spread around the mean, as any factors that result in underestimations/overestimations of clast abundance become enhanced. This is exemplified in Fig. S1, where different decisions between participants 1 and 2 result in substantially different IRD signals for the same core section (especially at 15–20 cm, 40–45 cm and 95–100 cm depth).

A few clear suggestions are provided below, by way of a protocol, which, if routinely implemented, may help to improve the consistency of the Grobe (1987) method.

1. Crucially, only one observer should count the gravel clasts of cores in any given study. Also following the guidelines presented by McKay et al. (2022), these counts should be validated by another user. This should avoid large discrepancies, even between experienced users, such as those seen between high and low counters.
2. Screen brightness should be maximised to facilitate maximum visibility of all clasts, whilst remaining comfortable.
3. Gravel classification should be determined based on the length of clasts (the longest visible axis).
4. Clasts should be counted in the sampling interval in which they appear most predominantly. Clasts that evenly overlap sampling intervals should only be counted in the first sampling interval in which they appear. Failing to adhere to this rule may result in overcounting and may also result in the reported number of clasts being biased by the grid location.
5. All clasts visibly close to the 2 mm threshold should be measured using a measuring tool. We draw the analogy that when conducting clast-fabric analysis, clast size would not simply be 'eyeballed' but rather measured. Discretion should be used, as there is no need to measure clasts that are clearly longer than 2 mm.

Additionally, to avoid 'high' versus 'low' counting errors, horizons containing numerous ambiguous objects (potential clasts) should be checked against a small physical sediment subsample. However, we appreciate that this destructive approach may not be feasible for all studies, and thus, this step is deliberately not included in the above protocol. Furthermore, combining clast counting (>2 mm) with different IRD proxy records (e.g., sieving, LPSA or automated approaches) can enhance and improve IRD interpretations.

Final considerations

It is well accepted that the core diameter can impact gravel counting as an IRD proxy (McKay et al., 2022). The results from this study add further uncertainties, given that some users consistently count higher or lower than each other. Our findings further add to the existing observation that clast counting on XR images cannot be used to quantify IRD as a proportion of weight or volume such as can be achieved using sieving or X-ray CT (Cederstrøm et al., 2021). However, despite the variability, manual clast counts serve as a consistent *proxy* (or semi-quantitative measure) for clast percentage abundances, and therefore, can be used to infer

© 2026 The Author(s). *Journal of Quaternary Science* Published by John Wiley & Sons Ltd.

10991417, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jqs.70091 by Enlighten Team - University Of Glasgow , Wiley Online Library on [06/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

relative increases or decreases in iceberg abundance, iceberg delivery and/or ice-marginal dynamics.

By adhering to the above protocol, better consistency will be achieved between users. We also call for more transparency when undertaking this particular method, whereby the specific rules implemented should be clearly stated when using the Grobe method.

In addition, more clarity must be provided in the literature as to what constitutes a clast in a sediment core XR and what does not constitute a clast, for example, as observed in the discrepancies between 'high' and 'low' counters. This will be more common in sediments with many ambiguous features with a radiodensity higher than that of the surrounding sediment matrix, increasing the propensity for inconsistencies. This emphasises the need for greater clarification and potentially standardisation of how different objects appear on sediment core X-ray images.

Finally, the results of this study strengthen the case for the use of automated strategies to determine IRD and to pursue further research into how these different techniques may be improved in different glaciomarine settings. Many of these automated methods (e.g., Reilly et al., 2019; McDonald et al., 2022) can be significantly quicker to implement. In the meantime, however, the framework provided here will go some way towards minimising the inconsistencies in the manual clast-counting technique, which will in turn help to improve training for machine-learning models based on manually counted data sets (e.g., Jasper et al., 2024).

Conclusions

In a participatory study of 14 individuals, the manual clast-counting IRD method (Grobe, 1987) was found to be variable among different users, but an internally consistent technique for quantifying gravel-sized IRD in two different sediment facies. Statistically significant relationships were observed in all but one case when undertaking pairwise comparisons of all 14 users' counts using Spearman's rho ($p < 0.05$). Although the phasing (or frequency of clasts per cm) was strongly correlated between nearly all users, there was significant variability between users' total counts. This internal consistency (rho = 0.895), but between-user variability indicates that some users are systematically more conservative counters than others for various reasons.

The factors that result in this user variation include different approaches when clasts overlap on XR images (Hogan et al., 2016; McKay et al., 2022); clasts appear in more than one sampling interval (1 cm slice); when determining between non-clastic high-radiodensity objects and low-radiodensity clasts; different approaches to determining clasts greater than the 2 mm threshold ('eyeballing' vs. measuring) are used; and when potential differences in users' screen brightness and eyesight are present. Expert does not play a significant role, with statistically significant count correlations observed between both experienced and inexperienced participants in this study. These findings highlight that user-dependent inconsistencies in XR clast counting introduce significant uncertainties into reconstructions of past ice-stream dynamics, particularly surrounding interpretations of former ice-shelf, sea-ice and calving-bay settings. A new protocol is presented for researchers to follow in order to increase count consistency.

Overall, this study suggests that XR clast counting, when applied manually, should only be considered a semi-quantitative measure of IRD abundance. Innovations in automated approaches to quantify IRD, particularly Artificial Intelligence, provide exciting new approaches that may act to mitigate the observer biases discovered in this study.

Acknowledgements. Collection of the marine sediment cores in this study was funded by the UK Natural Environment Research Council consortium grant BRITICE-CHRONO NE/J009768/1. Marine sediment core X-radiographs were captured at BOSCORF (part of the UK National Oceanography Centre) and GEOTEK Ltd (Daventry, UK), and we thank all participants for their input. We thank the BGS Marine Operations team and the NOC/NMFSS coring team for their help during JC123. We also thank the master and crew of the RRS *James Cook* (JC123). We would also like to thank all participants who took part in this study for their time and effort. This work was funded by the Natural Environment Research Council as part of an IAPETUS2 Doctoral Training Partnership (N. McDonald; NERC DTP NE/S007431/1).

Conflicts of Interest—The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Supporting information

Additional supporting information can be found in the online version of this article.

McDonald_et_al_JQS_supplementary_files.

References

Ahn, Y., Joe, Y.J., Jang, K., Kim, J.H., Son, Y.J., Forwick, M. et al. (2024) Post-glaciation depositional changes in Wijdefjorden, northern Svalbard, using grain-size end-member modelling. *Marine Geology*, 472, 107306. Available from: https://doi.org/10.1016/j.margeo.2024.107306

Anderson, J.B. (1999) *Antarctic Marine Geology*. Cambridge: Cambridge University Press. Available from: https://doi.org/10.1017/CBO9780511759376

Andrews, J. (2000) Icebergs and iceberg rafted detritus (IRD) in the North Atlantic: facts and assumptions. *Oceanography*, 13(3), 100–108. Available from: https://doi.org/10.5670/oceanog.2000.19

Andrews, J.T., Stein, R., Moros, M. & Perner, K. (2016) Late Quaternary changes in sediment composition on the NE Greenland margin (~73° N) with a focus on the fjords and shelf. *Boreas*, 45, 381–397. Available from: https://doi.org/10.1111/bor.12169

Bartels, M., Titschack, J., Fahl, K., Stein, R. & Hebbeln, D. (2018) Wahlenbergfjord, eastern Svalbard: a glacier-surrounded fjord reflecting regional hydrographic variability during the Holocene? *Boreas*, 47, 1003–1021. Available from: https://doi.org/10.1111/bor.12325

Bartels, M., Titschack, J., Fahl, K., Stein, R., Seidenkrantz, M.-S., Hillaire-Marcel, C. et al. (2017) Atlantic Water advection vs. glacier dynamics in northern Spitsbergen since early deglaciation. *Climate of the Past*, 13, 1717–1749. Available from: https://doi.org/10.5194/cp-13-1717-2017

Bradwell, T., Fabel, D., Clark, C.D., Chiverrell, R.C., Small, D., Smedley, R.K. et al. (2021) Pattern, style and timing of British–Irish Ice Sheet advance and retreat over the last 45 000 years: evidence from NW Scotland and the adjacent continental shelf. *Journal of Quaternary Science*, 36(5), 871–933. Available from: https://doi.org/10.1002/jqs.3296

Bradwell, T., Small, D., Fabel, D., Smedley, R.K., Clark, C.D. & Saher, M.H. et al. (2019) Ice-stream demise dynamically conditioned by trough shape and bed strength. *Science Advances*, 5, eaau1380. Available from: https://doi.org/10.1126/sciadv.aau1380

Callard, S.L., Ó Cofaigh, C., Benetti, S., Chiverrell, R.C., Van Landeghem, K.J.J., Saher, M.H. et al. (2018) Extent and retreat history of the Barra Fan Ice Stream offshore western Scotland and

© 2026 The Author(s). *Journal of Quaternary Science* Published by John Wiley & Sons Ltd.

10991417, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jqs.70091 by Enlighten Team - University Of Glasgow , Wiley Online Library on [06/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

northern Ireland during the last glaciation. *Quaternary Science Reviews*, 201, 280–302. Available from: https://doi.org/10.1016/j.quascirev.2018.10.002

Callard, S.L., Ó Cofaigh, C., Lloyd, J.M., A. Smith, J., Gebhardt, A., Kanzow, T. et al. (2025) Ocean driven retreat of the Northeast Greenland Ice Stream following the Last Glacial Maximum. *Nature Communications*, 16(1), 10961. Available from: https://doi.org/10.1038/s41467-025-66671-2

Cederstrøm, J.M., Van der Bilt, W.G., Støren, E.W. & Rutledal, S. (2021) Semi-automatic ice-rafted debris quantification with computed tomography. *Paleoceanography and Paleoclimatology*, 36(9), e2021PA004293. Available from: https://doi.org/10.1029/2021PA004293

Chen, P.Y. & Popovich, P.M. (2002) *Correlation: parametric and nonparametric measures*, Series on Quantitative Applications in Social Sciences 07-139. Thousand Oaks, California: Sage University Papers, 95 pp. Available from: https://doi.org/10.4135/9781412983808.n1

Clark, R.W., Wellner, J.S., Hillenbrand, C.-D., Totten, R.L., Smith, J.A. & Miller, L.E. et al. (2024) Synchronous retreat of Thwaites and Pine Island glaciers in response to external forcings in the presatellite era. *Proceedings of the National Academy of Sciences*, 121(11), e2211711120. Available from: https://doi.org/10.1073/pnas.2211711120

Conolly, J.R. & Ewing, M. (1965) Ice-rafted detritus as a climatic indicator in Antarctic deep-sea cores. *Science*, 150, 1822–1824. Available from: https://doi.org/10.1126/science.150.3705.1822

Domack, E.W. & Harris, P.T. (1998) A new depositional model for ice shelves, based upon sediment cores from the Ross Sea and the MaC. Robertson shelf, Antarctica. *Annals of Glaciology*, 27, 281–284. Available from: https://doi.org/10.3189/1998AoG27-1-281-284

Dowdeswell, J.A., Whittington, H., Jennings, R.J., Andrews, Mackensen & Marienfeld (2000) An origin for laminated glacimarine sediments through sea-ice build-up and suppressed iceberg rafting. *Sedimentology*, 47, 557–576. Available from: https://doi.org/10.1046/j.1365-3091.2000.00306.x

Girard, J., Reilly, B.T., St-Onge, G., Lagroix, F., Montero-Serrano, J.C. & Stoner, J.S. et al. (2024) Paleomagnetic secular variations in North Greenland around 81° N over the last 6,000 years. *Geochemistry, Geophysics, Geosystems*, 25, e2024GC011620. Available from: https://doi.org/10.1029/2024GC011620

Graham, A.G.C., Kuhn, G., Meisel, O., Hillenbrand, C.D., Hodgson, D.A., Ehrmann, W. et al. (2017) Major advance of South Georgia glaciers during the Antarctic Cold Reversal following extensive sub-Antarctic glaciation. *Nature Communications*, 8, 14798. Available from: https://doi.org/10.1038/ncomms14798

Grobe, H. (1987) A simple method for the determination of ice-rafted debris in sediment cores. *Polarforschung*, 57(3), 123–126. Available from: https://doi.org/10.0013/epic.11658.d001

Grobe, H. & Mackensen, A. (1992) Late Quaternary climatic cycles as recorded in sediments from the Antarctic continental margin. *Antarctic Research Series*, 56, 349–376. Available from: https://doi.org/10.1029/AR056p0349

Grobe, H., Fütterer, D. & Spieß, V. (1990) Oligocene to Quaternary sedimentation processes on the Antarctic continental margin, ODP Leg 113, Site 693. *Proceedings ODP Scientific Results*, 113, 693. Available from: https://doi.org/10.2973/odp.proc.sr.113.193.1990

Hansen, M.A., Passchier, S., Khim, B.K., Song, B. & Williams, T. (2015) Threshold behavior of a marine-based sector of the East Antarctic Ice Sheet in response to early Pliocene ocean warming. *Paleoceanography*, 30(6), 789–801. Available from: https://doi.org/10.1002/2014PA002704

Hillenbrand, C.-D., Crowhurst, S.J., Williams, M., Hodell, D.A., McCave, I.N., Ehrmann, W. et al. (2021) New insights from multi-proxy data from the West Antarctic continental rise. *Quaternary Science Reviews*, 257, 106842. Available from: https://doi.org/10.1016/j.quascirev.2021.106842

Hogan, K.A., Ó Cofaigh, C., Jennings, A.E., Dowdeswell, J.A. & Hiemstra, J.F. (2016) Deglaciation of a major palaeo-ice stream in Disko Trough, West Greenland. *Quaternary Science Reviews*, 147, 5–26. Available from: https://doi.org/10.1016/j.quascirev.2016.01.018

Jasper, C.E., Dyer, B., Reilly, B.T., Williams, T., Hemming, S. & Raymo, M.E. (2024) A 3.3-million-year record of Antarctic iceberg rafted debris and ice sheet evolution quantified by machine learning. *Paleoceanography and Paleoclimatology*, 39, e2024PA004897. Available from: https://doi.org/10.1029/2024PA004897

Jennings, A., Reilly, B., Andrews, J., Hogan, K., Walczak, M., Jakobsson, M. et al. (2022) Modern and early Holocene ice shelf sediment facies from Petermann Fjord and northern Nares Strait. *Quaternary Science Reviews*, 283, 107460. Available from: https://doi.org/10.1016/j.quascirev.2022.107460.

Knies, J., Kleiber, H.-P., Matthiessen, J., Müller, C. & Nowaczyk, N. (2001) Marine ice-rafted debris records constrain maximum extent of Saalian and Weichselian ice-sheets. *Global and Planetary Change*, 31, 45–64. Available from: https://doi.org/10.1016/S0921-8181(01)00112-6

López-Quirós, A., Junna, T., Davies, J., Andresen, K.J., Nielsen, T., Haghipour, N. et al. (2024) Retreat patterns and dynamics of the former Norske Trough ice stream (NE Greenland): an integrated geomorphological and sedimentological approach. *Quaternary Science Reviews*, 325. 108477, Available from: https://doi.org/10.1016/j.quascirev.2023.108477

McDonald, N., Bradwell, T., Callard, S.L., Toney, J.L., Shreeve, B. & Shreeve, J. (2022) Automated characterisation of glaciomarine sediments using X-ray computed laminography. *Quaternary Science Advances*, 5, 100046. Available from: https://doi.org/10.1016/j.qsa.2021.100046

McKay, R., Albot, O., Dunbar, G.B., Lee, J.I., Lee, M.K., Yoo, K.C. et al. (2022) A comparison of methods for identifying and quantifying ice-rafted debris on the Antarctic margin. *Paleoceanography and Paleoclimatology*, 37, e2021PA004404. Available from: https://doi.org/10.1029/2021PA004404

Ó Cofaigh, C. & Dowdeswell, D. A. (2001) Laminated sediments in glacimarine environments: diagnostic criteria for their interpretation. *Quaternary Science Reviews*, 20(13), 1411–1436. Available from: https://doi.org/10.1016/S0277-3791(00)00177-3

Ó Cofaigh, C., Lloyd, J.M., Callard, S.L., Gebhardt, C., Streuff, K.T., Dorschel, B. et al. (2025) Shelf-edge glaciation offshore of northeast Greenland during the last glacial maximum and timing of initial ice-sheet retreat. *Quaternary Science Reviews*, 359, 109326. Available from: https://doi.org/10.1016/j.quascirev.2025.109326

O'Regan, M., Cronin, T.M., Reilly, B., Alstrup, A.K.O., Gemery, L., Golub, A. et al. (2021) The Holocene dynamics of Ryder Glacier and ice tongue in north Greenland. *The Cryosphere*, 15, 4073–4097. Available from: https://doi.org/10.5194/tc-15-4073-2021

Pesjak, L., McMinn, A., Chase, Z. & Bostock, H. (2023) Sea ice and productivity changes over the last glacial cycle in the Adélie Land region. *Climate of the Past*, 19, 419–437. Available from: https://doi.org/10.5194/cp-19-419-2023

Powell, R.D. (1984) Glacimarine processes and inductive lithofacies modelling. *Marine Geology*, 57, 1–52. Available from: https://doi.org/10.1016/0025-3227(84)90194-4

Rack, F.R., Finndin, R. & Moran, K. (1996) Analysis and interpretation of X-ray images of sediment cores from Hole 910D, Yermak Plateau: preliminary results. In: Thiede, J., Myhre, A.M., Firth, J.V., Johnson, G.L. & Ruddiman, W.F. eds *Proceedings of the Ocean Drilling Program, Scientific Results*, 151, pp. 377–388. Available from: https://doi.org/10.2973/odp.proc.sr.151.129.1996

Reilly, B.T., Stoner, J.S., Mix, A.C., Walczak, M.H., Jennings, A., Jakobsson, M. et al. (2019) Holocene break-up and reestablishment of the Petermann Ice Tongue. *Quaternary Science Reviews*, 218, 322–342. Available from: https://doi.org/10.1016/j.quascirev.2019.06.023

Ruddiman, W.F. (1977) Late Quaternary deposition of ice-rafted sand in the subpolar North. *Atlantic (lat 40° to 65° N). GSA Bulletin*, 88(12), 1813–1827. Available from: https://doi.org/10.1130/0016-7606(1977)88<1813:LQS>2.0.CO;2

Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E. et al. (2025) GGally: Extension to 'ggplot2'. R package version 2.4.0. Available at: https://ggobi.github.io/ggally/

Scourse, J.D., Haapaniemi, A.I., Colmenero-Hidalgo, E., Peck, V.L., Hall, I.R., Austin, W.E.N. et al. (2009) Growth, dynamics and deglaciation of the last British–Irish ice sheet: the deep-sea ice-rafted detritus record. *Quaternary Science Reviews*, 28(27–28),

© 2026 The Author(s). *Journal of Quaternary Science* Published by John Wiley & Sons Ltd.

10991417, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jqs.70091 by Enlighten Team - University Of Glasgow , Wiley Online Library on [06/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3066–3084. Available from: https://doi.org/10.1016/j.quascirev.2009.08.009

Sheldon, C., Jennings, A., Andrews, J.T., Ó Cofaigh, C., Hogan, K., Dowdeswell, J.A. et al. (2016) Ice stream retreat following the LGM and onset of the west Greenland current in Uummannaq Trough, west Greenland. *Quaternary Science Reviews*, 147, 27–46. Available from: https://doi.org/10.1016/j.quascirev.2016.01.019

Smith, J.A. & Hogan, K.A. (2025) Glaciomarine sediments and processes. In: Elias, S. (Ed.) *Encyclopedia of Quaternary Science*, 3rd edition. Amsterdam: Elsevier, pp. 380–397. Available from: https://doi.org/10.1016/B978-0-323-99931-1.00225-7

Smith, J.A., Andersen, T.J., Shortt, M., Gaffney, A.M., Truffer, M., Stanton, T.P. et al. (2017) Sub-ice-shelf sediments record history of twentieth-century retreat of Pine Island Glacier. *Nature*, 541, 77–80. Available from: https://doi.org/10.1038/nature20136

Smith, J.A., Graham, A.G.C., Post, A.L., Hillenbrand, C.D., Bart, P.J. & Powell, R.D. (2019) The marine geological imprint of Antarctic ice shelves. *Nature Communications*, 10, 5635. Available from: https://doi.org/10.1038/s41467-019-13496-5

Smith, J.A., Hillenbrand, C.D., Subt, C., Rosenheim, B.E., Frederichs, T., Ehrmann, W. et al. (2021) History of the Larsen C Ice Shelf reconstructed from sub-ice-shelf and offshore sediments. *Geology*, 49, 978–982. Available from: https://doi.org/10.1130/G48503.1

Syring, N., Lloyd, J.M., Stein, R., Fahl, K., Roberts, D.H., Callard, L. et al. (2020) Holocene interactions between glacier retreat, sea ice formation, and Atlantic water advection at the inner Northeast Greenland continental shelf. *Paleoceanography and Paleoclimatology*, 35, e2020PA004019, Available from: https://doi.org/10.1029/2020PA004019

Weber, M.E., Clark, P.U., Kuhn, G., Timmermann, A., Sprenk, D., Gladstone, R. et al. (2014) Millennial-scale variability in Antarctic ice-sheet discharge during the last deglaciation. *Nature*, 510, 134–138. Available from: https://doi.org/10.1038/nature13397

Yamamoto, M., Okino, T., Sugisaki, S. & Sakamoto, T. (2008) Late Pleistocene changes in terrestrial biomarkers in sediments from the central Arctic Ocean. *Organic Geochemistry*, 39(6), 754–763. Available from: https://doi.org/10.1016/j.orggeochem.2008.04.009

© 2026 The Author(s). *Journal of Quaternary Science* Published by John Wiley & Sons Ltd.

10991417, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jqs.70091 by Enlighten Team - University Of Glasgow , Wiley Online Library on [06/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License