Schmidt Hammer exposure dating (SHED)

DOI:
10.1016/j.quageo.2017.12.003

Document Version
Accepted author manuscript

Schmidt Hammer exposure dating (SHED): Calibration procedures, new exposure age data and an online calculator

Tomkins, M.D., Huck, J.J., Dortch, J.D., Hughes, P.D., Kirkbride, M.P., and Barr, I.D.

Abstract

Recent research has established Schmidt Hammer exposure dating (SHED) as an effective method for dating glacial landforms in the UK. This paper presents new data and discussion to clarify and to evaluate calibration procedures. These make a distinction between Schmidt Hammer drift following use (*instrument calibration*), and variation between both individual Schmidt Hammers and between user strategies when utilising age-calibration curves (*age calibration*). We show that while test anvil methods are useful for verifying that Schmidt Hammers maintain their standard R-values, they are inappropriate for instrument calibration except for the hardest natural rock surfaces (R-values: ≥ 70). A range of surfaces were tested using 3 N-Type Schmidt Hammers, which showed that existing anvil calibration procedures led to consistent overestimation of R-values by up to 17.9%. In contrast, new calibration procedures, which are based on the use of a calibration point which lies within the range of R-values measured in the field [Dortch *et al.* 2016, Quat. Geochron., 35, 67-68], limit variance to maximum of 4.4% for surfaces typically tested by Quaternary researchers (R-values: 25 - 60). Moreover, these new calibration procedures are more appropriate for age calibration as they incorporate operator variance through choice of sampling location. New calibration procedures are used to compile an updated age-calibration curve based upon 54 granite surfaces (R^2 = 0.94, p < 0.01) from across Scotland, NW England and Ireland. The inclusion of a further 29 terrestrial cosmogenic nuclide (TCN) exposure ages extends the calibration period to 0.8 – 23.8 ka, covering the entire post-Last Glacial Maximum (LGM) history of the British-Irish Ice Sheet. To facilitate comparison between studies, an online calculator is made available at http://shed.earth for Schmidt Hammer instrument and age calibration and SHED exposure age calculation. The SHED-Earth calculator provides a rapid and accessible means of exposure age calculation to encourage wider and more consistent application of SHED throughout the British Isles.

Introduction

In a recent study by Tomkins *et al.* (2016), a statistically significant relationship (R^2 = 0.81, p < 0.01) was observed between the terrestrial cosmogenic nuclide (TCN) exposure ages and Schmidt Hammer rebound values (R-values) of 25 granitic surfaces from Scotland and NW England (Phillips *et al.*, 2008; Small *et al.*, 2012; Wilson *et al.*, 2013; Kirkbride *et al.*, 2014). These data indicate that granite can weather linearly over significant spatial scales for regions of similar climate (Tomkins *et al.*, 2016). The associated calibration curve was applied to undated glacial erratics (n = 31) on Shap Fell, NW England and generated a deglacial age of 16.5 ± 0.5 ka, a result which corroborates with existing methods (Wilson *et al.*, 2013) and is of comparable accuracy and precision to proximal TCN exposure ages. Using the University of Manchester calibration boulder (Dortch *et al.*, 2016), a recent study in the Mourne Mountains, Northern Ireland (Barr *et al.*, 2017), applied the Tomkins *et al.* (2016) SHED calibration curve to undated granite surfaces and generated a deglacial chronology that was consistent with existing interpretations of post-Last Glacial Maximum (LGM) glaciation in that region (Wilson, 2004; McCabe *et al.*, 2007; McCabe and Williams, 2012). As there were previously

no published numerical ages to constrain the chronology of glaciation (Barr *et al*., 2017), these new data provide an important geochronological control on Lateglacial and Younger Dryas ice dynamics in the Mournes. While SHED age estimates were generally younger (more recent) than established chronologies, perhaps reflecting climatic or lithological variation, the SHED approach was able to differentiate clearly between different phases of glaciation and is considered a viable method for constraining the extent of this region's glaciers during the Younger Dryas (Barr *et al*., 2017).

However, Winkler and Matthews (2016) contend that the real potential of SHED has been undermined by inappropriate calibration procedures utilised in Tomkins *et al*., (2016) and presented in Dortch *et al*., (2016). Here, we present new data and discussion to evaluate this issue in a robust and quantitative way. To assess the effectiveness of these calibration procedures and by association, the suitability of our regional calibration curve (Tomkins *et al*., 2016), we have compiled an updated calibration curve with new exposure age data from the Holocene (Kirkbride *et al*., 2014), Younger Dryas (Small and Fabel, 2016) and Lateglacial Interstadial (Everest and Kubik, 2006; Finlayson *et al*., 2014). In addition, we include early post-LGM (18 – 24 ka) exposure ages from Wexford (Ballantyne and Stone, 2015) and Donegal, Ireland (Ballantyne *et al*., 2007; Clark *et al*., 2009). These new data (n = 29) extend the calibration period to 0.8 – 23.8 ka, covering the entire post-Last Glacial Maximum (LGM) history of the British-Irish Ice Sheet (Clark *et al*., 2012).

While this regional calibration curve is unlikely to be globally applicable, as long-term weathering rates exhibit systematic variability between diverse climatic regimes (Riebe *et al*., 2004; von Blanckenburg *et al*., 2015), our methods can be used to develop similarly robust SHED calibration curves in other well-dated regions. This study aims to enable this by:

1. Testing and clarifying Schmidt Hammer calibration procedures. We make a distinction between *instrument calibration* i.e. correcting for Schmidt Hammer drift following use, and *age calibration* i.e. correcting different Schmidt Hammers and user strategies to a verifiable standard prior to the utilisation of our regional calibration curve.
2. Presenting new evidence to update and reinforce our regional SHED calibration curve.
3. Providing an online calculator for Schmidt Hammer instrument and age calibration and SHED exposure age calculation at http://shed.earth to encourage wider and more consistent application of SHED throughout the British Isles.

Instrument calibration

In their comment on Dortch *et al*. (2016), Winkler and Matthews (2016) criticise our calibration procedures as being unnecessary, impractical and less accurate than existing methods (Proceq, 2004; Aydin, 2009). In addition, the authors argue that our term 'standardised R-values' creates confusion as it does not differentiate clearly between instrument calibration and converting R-values into age information when developing a SHED-calibration curve.

Existing ISRM calibration procedures recommend the use of the test anvil to account for R-value drift following intensive use (Aydin, 2009). The test anvil should yield R-values in the range of 81 ± 2 for N-type Schmidt Hammers which have specified impact energies of 2.207 Nm (Proceq, 2004). Schmidt Hammers should be calibrated before and after use to generate a correction factor (CF) which should be applied to all readings as follows:

$$\mathrm{CF} = \frac{\text{Specified standard value of the anvil}}{\text{Average of ten readings on the anvil}}$$

Test anvils are constructed with vertically guided impact points made of steel as hard as that of the plunger tip (Aydin, 2009) and thus amplify variation between pre- and post-use calibration values and variation between different Schmidt Hammers (McCarroll, 1987). An implicit assumption in this calibration procedure is that the difference (%) between the specified anvil standard and the average of 10 readings is consistent throughout the operational range of the Schmidt Hammer (Compressive strength 10 - 70 N/mm²; Proceq, 2004). However, very few natural rock surfaces generate R-values at the upper end of this operational range (c.f. Table 1 in Goudie, 2006; 17/110 entries record R-values ≥ 60, 1/110 entries record R-values ≥ 70). As such, it is clearly necessary to evaluate the effectiveness of the standard calibration procedures (Proceq, 2004; Aydin, 2009) on surfaces of varying hardness to establish their validity (Winkler and Matthews, 2016).

To test this, we sampled a range of surfaces using three N-type Schmidt Hammers of varying age and usage (New Proceq: NPC, New NovaTest: NT, Old Proceq: OPC), all with specified impact energies of 2.207 Nm. 30 R-values were generated for each surface by the same operator and followed the procedures outlined by Viles *et al.* (2011). The sampling strategy was consistent for each test, while each surface was sampled on the same day to minimise the effect of variability in rock (Sumner and Nel, 2002) or surface moisture content (Viles *et al.*, 2011). Carborundum pre-treatment was performed for each surface to minimise potential errors resulting from variable surface roughness. Sample information and results are presented in Table 1.

The mean test anvil R-values varied significantly (NPC = 84 ± 0, NT = 79.3 ± 0.7, OPC = 67.5 ± 2.6), reflecting the varying usage of each Schmidt Hammer. However, these data demonstrate that the difference (%) between the Schmidt Hammers as recorded using the test anvil (NT = 5.6%, OPC = 19.6%) is not maintained throughout the operational range of the Schmidt Hammer (Fig. 1). Instead, the difference (%) between each tool decreases significantly as the surface R-value decreases. Of particular note is the consistency of mean R-values for the NPC and NT for the 3 weakest surfaces (R-values ≤ 48). The variance between the NPC and NT for these surfaces is not statistically significant as determined by Students T-tests (Table 1). The calibration procedures of Aydin (2009) were applied to the NT and OPC, with correction factors of 1.059 and 1.244 respectively, and compared to the values generated by the NPC (Table 2). To provide a baseline of measurement, we explicitly assume that the values generated by the NPC are a 'true' measure of the rock surface R-value. However, we acknowledge that a 'true' surface R-value is indeterminable (McCarroll, 1987) as this would require validation through repeated Schmidt Hammer testing which in turn could be affected by R-value drift following use, operator variance or even variance between new Schmidt Hammers (McCarroll, 1987). Ideally, we would have tested each Schmidt Hammer on surfaces with specified R-values but with the exception of the test anvil, these are not available. However, the NPC generated consistent values on the test anvil (84 ± 0) and we use these data as the basis for our calculations. Existing recalibration procedures result in consistent R-value overestimation in the typical 25 - 60 R-value operational range (Goudie, 2006), equivalent to 3.4 - 7.2% for the NT (Fig. 2B) and 9.2 – 17.9% for the OPC (Fig. 2A).

Next, we evaluate the calibration procedures of Dortch *et al.* (2016) using a calibration surface within the range of the sample data. Schmidt Hammers were recalibrated using the University of Manchester calibration boulder (Doddington Sandstone boulder; 1.8 m x 0.7 m x 0.7 m (L x W x H); NPC R-value 47.4 ± 1.9; NT R-value 47.6 ± 2.1, OPC R-value 43.5 ± 1.8) giving CFs of 0.994 and

1.167 for the NT and OPC respectively (Dortch *et al*., 2016). Recalibrated R-values correspond more closely to baseline NPC values, differing by a maximum of 2.9% for the NT (Fig. 2B) and 4.4% for OPC (Fig. 2A) within the typical 25 - 60 R-value operational range (Table 2). It must be noted that the data diverge towards the upper limit of the tools' operational range, as this calibration underestimates R-values for the two hardest surfaces by 3.4 - 6.1% for the NT and 6.6 - 12.4% for the OPC. However, from a combined total of 295 samples reported by Goudie (2006), Tomkins *et al*. (2016), and Barr *et al*. (2017), 277 or 93.9% have mean R-values within the range 25 - 60. For these surfaces, the variation between the NPC and the NT-OPC is reduced to a maximum of just 4.4% as compared to 17.9% for the ISRM method.

These data demonstrate that R-value data should be recalibrated using a calibration point which is within the range of sample data (Dortch *et al*., 2016). A boulder of sufficient size (Sumner and Nel, 2002; Demirdag *et al*., 2009), that is free of surface discontinuities (Williams and Robinson, 1983) and lichen (Matthews and Owen, 2008) and is easily accessible would be ideal. The University of Manchester calibration boulder as used in Tomkins *et al*. (2016) and presented in Dortch *et al*. (2016) is ideally suited for instrument calibration, as it is within the range of surfaces typically tested by Quaternary researchers (R-value: 48.08 ± 0.82). For clarity, we are not advocating that researchers must use the University of Manchester calibration boulder for *instrument calibration*, although we do encourage users to perform *age calibration* using this surface in order to test or utilise the regional calibration curve (*see below*). However, we are advocating that researchers use a *comparable* surface to perform instrument calibration i.e. one that returns R-values within the range of field data, is free of surface discontinuities and is easily accessible. Moreover, researchers should follow our sampling methodology and perform carborundum pre-treatment to ensure a smooth, debris-free surface. Users should record 30 R-values perpendicular to the tested surface to reduce the risk of frictional sliding of the plunger tip (Viles et al., 2011), with single impacts separated by at least a plunger width (Aydin, 2009). As the SH is sensitive to rock (Sumner and Nel, 2002) and surface moisture content (Viles et al., 2011), we recommend sampling in dry conditions. For very hard rock surfaces (R-values: ≥ 70), the test anvil method may be effective as variation between Schmidt Hammers as recorded on the anvil is probably representative of variability on sampled rock surfaces (Table 2). However, while we do not dispute the value of the test anvil in verifying that Schmidt Hammers maintain their standard R-values and for prompting cleaning or repair (Winkler and Matthews, 2016), it is clear that instrument calibration using the test anvil will significantly overestimate R-values for the vast majority of rock surfaces tested by Quaternary researchers (R-values: ≤ 60).

Age calibration

In their comment on Dortch *et al*., (2016), Winkler and Matthews (2016) are correct in their assertion that "Standardisation of R-values is irrelevant for the construction of these calibration curves because their accuracy relies on the quality of those specific control points and a consistent sampling design throughout data collection." However, unlike previous studies, which generate localised age calibration curves (i.e. single valley), the goal of the SHED project was to encourage researchers to test and use our calibration curve on undated landforms at the regional scale and compare results with independent dating methods (e.g. TCN, ^{14}C, OSL) to evaluate its effectiveness as a geochronological tool. Thus, "age calibration" relates to the *utilisation*, not the *construction*, of calibration curves. As a result, the standardisation of different Schmidt Hammers (McCarroll, 1987)

and different user strategies (Viles *et al*., 2011) to a verifiable standard is necessary to limit potential errors in SH exposure age estimates.

Age calibration could be undertaken using the test anvil but this is rejected for two reasons. Firstly, as with instrument calibration, the difference between Schmidt Hammers as recorded using the test anvil is unlikely to be replicated on surfaces typically tested by Quaternary researchers and would likely result in significant R-value overestimation. Secondly, the test anvil procedures of McCarroll (1987; 1994) and Aydin (2009), and calibration guidelines from Proceq (2004), consider instrument error but do not account for the uncertainty generated by user variance. Viles *et al*., (2011) state that "operator variance … may also be an issue even for well-established techniques such as the Schmidt Hammer". Variance between users, due to choice of sampling location or operating procedure, cannot be discounted as a source of error in R-value data and subsequent SH exposure age estimates. As a result, the use of a natural rock surface (University of Manchester calibration boulder) is more appropriate for age calibration as it simulates field sampling conditions and permits R-value variation due to (1) choice of sampling location and (2) sampling strategy. While "micro-scale inhomogeneity of the sandstone" will result in larger uncertainties than the test anvil methods (Winkler and Matthews, 2016), we consider this marginal increase in uncertainty to be insufficient to offset the considerable advantages of (1) incorporating operator variance and (2) enabling recalibration without significant overestimation.

It is vitally important for the development of SHED, and the trust of the geomorphological community, that unrealistically precise estimates are avoided in the literature. However, while errors in instrument and age calibration could influence SH exposure age estimates, it is clear that the largest uncertainties in SHED are a consequence of limited control points for age-calibration curve construction (Winkler, 2009; Matthews and Winkler, 2011). This is exacerbated by geological uncertainty associated with TCN exposure ages (Heyman *et al*., 2011) which may adversely affect calibration curves based on sparse and isolated control points (Tomkins *et al*., 2016). As a result, local 'R-value to age' calibration curves with limited age control points are unlikely to be applicable on a wider regional scale. Even robust age calibration curves (R^2 = 0.94, p = < 0.01), based upon significant exposure age data sets (n = 54), must avoid unrealistically precise estimates of surface exposure age if the Quaternary community at large is to take up the Schmidt Hammer and integrate it with radiometric dating methods. To that end, the instrument and age calibration procedures outlined here are suitable for the geomorphological community as they work effectively on surfaces typically tested by Quaternary researchers (R-values: 25 - 60) and minimise potential errors introduced by variation between Schmidt Hammers and between user strategies.

As a result, we provide the following recommendations:

1. *Instrument calibration* - Users can account for R-value drift following use using a suitable surface before and after sampling following the methods presented in Dortch *et al.* (2016).

2. *Age calibration* – Users can account for variation between Schmidt Hammers and between user strategies by calibrating their Schmidt Hammer using the University of Manchester calibration boulder to standardise R-values to our regional calibration curve.

3. *Exposure-age calculation* – Users can input instrument and age calibration values and raw R-value data into the SHED-Earth online calculator (http://shed.earth) to generate SH

exposure ages and 1σ uncertainties based on the updated calibration curve presented in this study.

4. *Developing an independent regional age calibration curve* – To generate a new exposure age to R-value calibration curve in a similar well-dated region, first select a suitable surface for age calibration. This surface should be tested before all field-testing to minimise errors due to variation between Schmidt Hammers and between user strategies. The location of this surface should be published to encourage wider use (Dortch *et al.*, 2016). Next, select a suitable surface for instrument calibration. This surface should be used before and after all field-testing to account for R-value drift. Finally, proceed to develop a calibration dataset for your region.

Updating the UK SHED calibration curve

These instrument and age calibration procedures are used to compile an updated age-calibration curve based on upon 54 granite surfaces (R^2 = 0.94, p = < 0.01) from across Scotland, NW England and Ireland. This calibration curve comprises a further 29 TCN exposure age control points (Figs. 3) and includes new Holocene and Younger Dryas exposure ages from moraine crests in Coire an Lochain, Cairngorms (Fig. 4A; n = 5, 0.8 – 5.5 ka; Kirkbride *et al.*, 2014) and on Rannoch Moor, Scottish Highlands (Fig. 4B; n = 5, 11.2 – 12.7 ka; Small and Fabel, 2016), in addition to Lateglacial exposure ages from Glen Einich, Cairngorms (Fig. 4C; n = 6, 14.3 – 16.8 ka; Everest and Kubik, 2006) and Glen Iorsa, Arran (n = 2, 15.8 – 16.7 ka; Finlayson *et al.*, 2014). Dated surfaces from Glen Einich were not included in the previous calibration curve (Tomkins *et al.*, 2016) due to their coarse-grained surface texture and poor internal 'exposure age to R-value' consistency (R^2 = 0.19, p = 0.39). However, on further analysis, we note that four of the six exposure ages were within 1σ uncertainty of the original calibration regression. In addition, their lack of internal consistency is probably best accounted for by TCN exposure age uncertainty (± 1.1 - 1.6 ka). For transparency, all of these data are included. Finally, we include early post-LGM exposure ages from Blackstairs Mountain, Wexford (n = 2, 23.3 – 23.4 ka; Ballantyne and Stone, 2015) and Bloody Foreland, Donegal (Fig. 4D; n = 8, 18.2 – 23.8 ka; Ballantyne *et al.*, 2007; Clark *et al.*, 2009). These data fit the trend established at early (> 20 ka) post-LGM sites from Buchan (Phillips *et al.*, 2008) and demonstrate the wide applicability of this calibration curve throughout the British Isles. As a result, the comparatively 'young' SHED exposure age estimates from the Mourne Mountains (Barr *et al.*, 2017) appear unlikely to reflect climatic variation.

All calibration curve exposure ages were calculated using the online calculators formerly known as the CRONUS-Earth online calculators (http://hess.ess.washington.edu/math/, Wrapper script 2.3, Main calculator 2.1, constants 2.3, muons 1.1; Balco *et al.*, 2008). Exposure ages are based on the time-dependent Lm scaling (Lal, 1991; Stone, 2000) and assuming 0 mm ka^{-1} erosion. While there are no reliable estimates of surface erosion rates for rock surfaces in the British Isles (Ballantyne, 2010), erosion rates for most crystalline glaciated rock surfaces are usually low (0.1 – 0.3 mm ka^{-1}; André, 2002). As such, assuming 0 mm ka^{-1} of surface erosion is the most suitable approach as rates of surfaces lowering are likely negligible (André, 2002) and should not be estimated without supplementary data. Exposure ages are based on the Loch Lomond production rate (LLPR; Fabel *et al.*, 2012) of 4.02 ± 0.18 atoms g^{-1} a^{-1}. The LLPR is based on ^{10}Be concentrations from erratic boulders on the terminal moraine of the Younger Dryas Loch Lomond glacier advance (Fabel *et al.*, 2012), the timing of which is independently constrained by ^{14}C ages (MacLeod *et al.*, 2011). The LLPR

is the default production rate for the SHED-Earth online calculator although we also include the option to calculate ages based on the Glen Roy production rate (GRPR; Small and Fabel, 2015) and the primary calibration dataset of Borchers et al., (2016) for comparison. The full exposure age calibration dataset is available in Appendix 1.

In total, this updated calibration curve (Fig. 5) is based on 54 dated surfaces from across Scotland, NW England and Ireland, and definitively demonstrates a clear correlation between exposure ages and recalibrated R-values (R^2 = 0.94, p = < 0.01). For individual SH exposure age estimates that fall within the operational range of our calibration curve (0.8 – 23.8 ka), this technique generates typical errors of ~1.4 ka, reflecting the uncertainty introduced by (1) recorded scatter in SH R-values and (2) intrinsic uncertainty associated with calibration curve TCN ages. However, in aggregate, internally-consistent SH exposure age datasets (e.g. n = 30) can be of comparable precision to TCN ages (Tomkins et al., 2016), as counting statistics can be used to consolidate probability when numerous ages are obtained. The addition of new exposure ages to the calibration curve has changed the slope of the calibration curve regression (Tomkins *et al*., 2016: y = -0.4881x + 34.834, updated curve: y = -0.5678x + 37.692). To evaluate the significance of this change, we recalibrated the Shap Fell data presented in Tomkins *et al*., (2016). Using the original calibration regression, this data generated a mean exposure age of 16.5 ± 0.5 ka and provided a limiting age for the south-westerly retreat of ice towards the mountains of the Lake District (Wilson, 2016). Using the updated calibration regression, the arithmetic mean and mean absolute deviation of this dataset is 16.36 ± 0.60 ka (n = 31). This estimate is consistent with the youngest LLPR TCN exposure age from Shap Fell of 16.42 ± 0.98 ka (Wilson *et al*., 2013). As such, the application of the updated calibration curve to these data has no impact on the conclusions of Tomkins *et al*., (2016).

SHED online calculator

A key objective of the SHED project is to make our calibration curve accessible to Quaternary researchers and thus enable wider and more consistent application of the technique to undated landscapes (c.f. Barr *et al*., 2017). To that end, we present a tool for Schmidt Hammer instrument and age calibration and SHED exposure age calculation (available at http://shed.earth). SHED-Earth performs the following functions:

1. *Instrument calibration* – Users can input raw R-value data in chronological order (related to the time of sampling, not the SHED chronology) and the R-values of their instrument calibration surface before and after sampling. R-values will be corrected assuming linear R-value drift (Tomkins *et al*., 2016). This procedure is most effective when periods between calibration tests are short (McCarroll, 1987). While we encourage users to record 30 R-values per surface to ensure statistically significant results, the tool will also operate on variable sample sizes (c.f. Table 2 in Niedzielski *et al*., 2009).

2. *Age calibration* – Users can input the mean value recorded for the University of Manchester calibration boulder and the tool will correct each R-value using a correction factor (%). Users who have not completed age calibration using the University of Manchester calibration boulder should use the default value (R-value: 48.08 ± 0.82). This is the mean R-value generated by the Proceq N-type Schmidt Hammer used to generate the original calibration curve (Tomkins *et al*., 2016). As such, no correction for variation between different Schmidt Hammers or between user strategies will be made. Although variance between Schmidt

Hammers is usually small for surfaces with R-values of ≤ 60, and should be minimal if Schmidt Hammers are calibrated on a regular basis, variance can exceed ~10% for older Schmidt Hammers (Table 2; OPC = 5.2 – 12.3%) and should be accounted for.

3. *Exposure-age calculation* – Recalibrated mean R-values and the mean absolute deviation for each sample are calculated and are used to generate Schmidt Hammer exposure ages and 1σ uncertainties for each sampled surface using the updated granite calibration curve presented in this paper (Fig. 5).

User inputs include sample IDs and locations (latitude/longitude), which are stored in a database for monitoring of site usage. User data (R-values and SH exposure ages) are not recorded. The analysis codes are compiled in Python and are available for users to access. With the exception of age calibration using the University of Manchester calibration boulder, users can sample deposits, perform instrument calibration and generate exposure ages and uncertainties independently. SHED-Earth further streamlines this dating technique by providing a rapid and accessible means of exposure age calculation. It is anticipated that as new regional calibration curves are generated in similar well-dated regions, they will be made available on SHED-Earth. Finally, for researchers developing their own regional TCN to R-value calibration curves, we are happy to host your data and make your curve an available option for other users.

Conclusions

Despite extensive research over the last ~40 years (c.f. Table 1; Tomkins *et al*., 2016), previous studies have failed to inspire the Quaternary community at large to take up the Schmidt Hammer and integrate these data with newer dating methods. The recommendations in Dortch *et al*., (2016) were intended to encourage Quaternary researchers to test and utilise our calibration curve and generate their own curves in suitably dated regions. Winkler and Matthews (2016) criticize Tomkins *et al*., (2016) and Dortch *et al*., (2016) for instrument and age calibration but fail to address the significantly larger uncertainties associated with age-calibration curve construction. Our calibration procedures produce realistic uncertainties, incorporate operator variance and are more effective on surfaces typically tested by Quaternary researchers, making them more appropriate than previous calibration procedures which are not specifically designed for the Quaternary community or for SHED. While we acknowledge that robust Schmidt Hammer calibration procedures are necessary to generate reliable data (McCarroll, 1987; McCarroll, 1994), it is evident that of greater concern in the application of SHED is the use of isolated age control points. The largest uncertainties are a consequence of limited data points for age-calibration curve construction, which is exacerbated by geological uncertainty associated with TCN exposure ages (Heyman *et al*., 2011). To accommodate this uncertainty, calibration curves should be based on statistically large datasets to minimise individual exposure age uncertainty (c.f. Tomkins *et al*., 2016). Our methods take a conservative view, incorporate larger uncertainties and still produced a robust age calibration curve for granite surfaces in the UK.

We hope that clear instrument and age calibration procedures, new exposure age data (n = 29) and the availability of an online-calculator which streamlines calibration and SHED exposure age calculation (http://shed.earth), will provide further encouragement for Quaternary researchers. The calibration dataset is now substantial (n = 54) and is applicable over the timeframe of 0.8 – 23.8 ka.

While we acknowledge that further work is necessary to apply the technique more widely to undated landforms (e.g. Barr *et al*., 2017), we believe that the current calibration is fully usable and encourage researchers to test and utilize it.

References

André, M. F. (2002). Rates of postglacial rock weathering on glacially scoured outcrops (Abisko–Riksgränsen area, 68°N). *Geografiska Annaler: Series A, Physical Geography*, *84*(3-4), 139-150. https://doi.org/10.1111/j.0435-3676.2002.00168.x

Aydin, A. (2009). The ISRM Suggested Methods for Rock Characterization, Testing and Monitoring: 2007-2014. *Int. J. Rock Mech. Min. Sci.*, *46*, 627–634. https://doi.org/10.1007/978-3-319-07713-0

Ballantyne, C. K., McCarroll, D., & Stone, J. O. (2007). The Donegal ice dome, northwest Ireland: dimensions and chronology. *Journal of Quaternary Science*, *22*, 773–783. https://doi.org/10.1002/jqs

Ballantyne, C. K., & Stone, J. O. (2015). Trimlines, blockfields and the vertical extent of the last ice sheet in southern Ireland. *Boreas*, *44*, 277–287. https://doi.org/10.1111/bor.12109

Barr, I. D., Roberson, S., Flood, R., & Dortch, J. (2017). Younger Dryas glaciers and climate in the Mourne Mountains, Northern Ireland. *Journal of Quaternary Science*, *32*(1), 104–115. https://doi.org/10.1002/jqs.2927

Borchers, B., Marrero, S., Balco, G., Caffee, M., Goehring, B., Lifton, N., Nishiizumi, K., Phillips, F., Schaefer, J., & Stone, J. (2016). Geological calibration of spallation production rates in the CRONUS-Earth project. *Quaternary Geochronology*, 31, 188-198. https://doi.org/10.1016/j.quageo.2015.01.009

Clark, J., Cabe, A. M. M. C., Schnabel, C., Clark, P. U., Freeman, S., Maden, C., & Xu, S. (2009). 10Be chronology of the last deglaciation of County Donegal, northwestern Ireland. *Boreas*, *38*, 111–118. https://doi.org/10.1111/j.1502-3885.2008.00040.x

Clark, J., McCabe, A. M., Bowen, D. Q., & Clark, P. U. (2012). Response of the Irish Ice Sheet to abrupt climate change during the last deglaciation. *Quaternary Science Reviews*, *35*, 100–115. https://doi.org/10.1016/j.quascirev.2012.01.001

Demirdag, S., Yavuz, H., & Altindag, R. (2009). The effect of sample size on Schmidt rebound hardness value of rocks. *International Journal of Rock Mechanics and Mining Sciences*, *46*(4), 725–730. https://doi.org/10.1016/j.ijrmms.2008.09.004

Dortch, J. M., Hughes, P. D., & Tomkins, M. D. (2016). Schmidt hammer exposure dating (SHED): Calibration boulder of Tomkins *et al*. (2016). *Quaternary Geochronology*, *35*, 67–68. https://doi.org/10.1016/j.quageo.2016.06.001

Everest, J., & Kubik, P. (2006). The deglaciation of eastern Scotland: Cosmogenic ^{10}Be evidence for a Lateglacial stillstand. *Journal of Quaternary Science*, *21*(1), 95–104. https://doi.org/10.1002/jqs.961

Finlayson, A., Fabel, D., Bradwell, T., & Sugden, D. (2014). Growth and decay of a marine terminating sector of the last British–Irish Ice Sheet: a geomorphological reconstruction. *Quaternary Science Reviews*, *83*, 28-45. https://doi.org/10.1016/j.quascirev.2013.10.009

Golledge, N. ., Fabel, D., Everest, J. D., Freeman, S., & Binnie, S. (2007). First cosmogenic ^{10}Be age constraint on the timing of Younger Dryas glaciation and ice cap thickness, western Scottish Highlands. *Journal of Quaternary Science*, *22*(8), 785–791. https://doi.org/10.1002/jqs

Goudie, A. S. (2006). Progress in Physical Geography The Schmidt Hammer in geomorphological research. *Progress in Physical Geography*, *30*, 703–718. https://doi.org/10.1177/0309133306071954

Gunnell, Y., Jarman, D., Braucher, R., Calvet, M., Delmas, M., Leanni, L., Bourlès, D., Aumaître, G., & Keddaouche, K. (2013). The granite tors of Dartmoor, Southwest England: Rapid and recent emergence revealed by Late Pleistocene cosmogenic apparent exposure ages. *Quaternary Science Reviews*, *61*, 62–76. https://doi.org/10.1016/j.quascirev.2012.11.005

Heyman, J., Stroeven, A. P., Harbor, J. M., & Caffee, M. W. (2011). Too young or too old: Evaluating cosmogenic exposure dating based on an analysis of compiled boulder exposure ages. *Earth and Planetary Science Letters*, *302*(1–2), 71–80. https://doi.org/10.1016/j.epsl.2010.11.040

Kirkbride, M., Everest, J., Benn, D., Gheorghiu, D., & Dawson, A. (2014). Late-Holocene and Younger Dryas glaciers in the northern Cairngorm Mountains, Scotland. *The Holocene*, *24*(2), 141–148. https://doi.org/10.1177/0959683613516171

Lal, D. (1991). Cosmic ray labeling of erosion surfaces: in situ nuclide production rates and erosion models. Earth and Planetary Science Letters, 104(2-4), 424-439. https://doi.org/10.1016/0012-821X(91)90220-C

Matthews, J. A., & Owen, G. (2008). Endolithic lichens, rapid biological weathering and Schmidt Hammer R-values on recently exposed rock surfaces: Storbreen glacier foreland, Jotunheimen, Norway. *Geografiska Annaler, Series A: Physical Geography*, *90*(4), 287–297. https://doi.org/10.1111/j.1468-0459.2008.00346.x

Matthews, J. A., & Winkler, S. (2011). Schmidt-hammer exposure-age dating (SHD): Application to early Holocene moraines and a reappraisal of the reliability of terrestrial cosmogenic-nuclide dating (TCND) at Austanbotnbreen, Jotunheimen, Norway. *Boreas*, *40*(2), 256–270. https://doi.org/10.1111/j.1502-3885.2010.00178.x

McCabe, A. M., Clark, P. U., & Clark, J. (2007). Radiocarbon constraints on the history of the western Irish ice sheet prior to the Last Glacial Maximum. *Geology*, *35*, 147–150. https://doi.org/10.1130/G23167A.1

Mccabe, A. M., & Williams, G. D. (2012). Timing of the East Antrim Coastal Readvance: phase relationships between lowland Irish and upland Scottish ice sheets during the Last Glacial Termination. *Quaternary Science Reviews*, *58*, 18–29. https://doi.org/10.1016/j.quascirev.2012.10.012

McCarroll, D. (1987). The Schmidt hammer in geomorphology: five sources of instrument error. *Br. Geomorphol. Res. Group Tech. Bull.*, *36*, 16–27.

McCarroll, D. (1994). The Schmidt Hammer as a measure of degree of rock surface weathering and terrain age. In C. Beck (Ed.), *Dating in Exposed and Surface Contexts* (pp. 29–45). Albuquerque: University of New Mexico Press.

Niedzielski, T., Migon, P., & Placek, A. (2009). A minimum sample size required from Schmidt hammer measurements. *Earth Surface Processes and Landforms*, *34*, 1713–1725. https://doi.org/10.1002/esp

Phillips, W. M., Hall, A. M., Ballantyne, C. K., Binnie, S., Kubik, P. W., & Freeman, S. (2008). Extent of the last ice sheet in northern Scotland tested with cosmogenic ^{10}Be exposure ages. *Journal of Quaternary Science*, *23*(2),101-107. https://doi.org/ 10.1002/jqs.1161

Proceq. (2004). *Operating Instructions Betonprüfhammer N/NR- L/LR*. Schwerzenbach.

Riebe, C. S., Kirchner, J. W., & Finkel, R. C. (2004). Erosional and climatic effects on long-term chemical weathering rates in granitic landscapes spanning diverse climate regimes. *Earth and Planetary Science Letters*, *224*(3), 547-562. https://doi.org/10.1016/j.epsl.2004.05.019

Small, D., Fabel, D. (2015). A Lateglacial ^{10}Be production rate from glacial lake shorelines in Scotland. *Journal of Quaternary Science*, 30, 509–513. https://doi.org/10.1002/jqs.2804

Small, D., & Fabel, D. (2016). Was Scotland deglaciated during the Younger Dryas? *Quaternary Science Reviews*, *145*, 259–263. https://doi.org/10.1016/j.quascirev.2016.05.031

Small, D., Rinterknecht, V., Austin, W., Fabel, D., Miguens-Rodriguez, M., & Xu, S. (2012). In situ cosmogenic exposure ages from the Isle of Skye, northwest Scotland: Implications for the timing of deglaciation and readvance from 15 to 11ka. *Journal of Quaternary Science*, *27*(2), 150–158. https://doi.org/10.1002/jqs.1522

Stone, J. O. (2000). Air pressure and cosmogenic isotope production. Journal of Geophysical Research: Solid Earth, 105(B10), 23753-23759. https://doi.org/10.1029/2000JB900181

Sumner, P., & Nel, W. (2002). The effect of rock moisture on Schmidt hammer rebound: Tests on rock samples from Marion Island and South Africa. *Earth Surface Processes and Landforms*, *27*(10), 1137–1142. https://doi.org/10.1002/esp.402

Tomkins, M. D., Dortch, J. M., & Hughes, P. D. (2016). Schmidt Hammer exposure dating (SHED): Establishment and implications for the retreat of the last British Ice Sheet. *Quaternary Geochronology*, *33*, 46–60. https://doi.org/10.1016/j.quageo.2016.02.002

Viles, H., Goudie, A., Grab, S., & Lalley, J. (2011). The use of the Schmidt Hammer and Equotip for rock hardness assessment in geomorphology and heritage science: A comparative analysis. *Earth Surface Processes and Landforms*, *36*(3), 320–333. https://doi.org/10.1002/esp.2040

Von Blanckenburg, F., Bouchez, J., Ibarra, D. E., & Maher, K. (2015). Stable runoff and weathering fluxes into the oceans over Quaternary climate cycles. *Nature Geoscience*, *8*(7), 538. https://doi.org/10.1038/ngeo2452

Williams, R. B. G., & Robinson, D. A. (1983). The effect of surface texture on the determination of the surface hardness of rock using the schmidt hammer. *Earth Surface Processes and Landforms*, *8*(3), 289–292. https://doi.org/10.1002/esp.3290080311

Wilson, K. R. (2004). The last glaciation in the Western Mourne Mountains, Northern Ireland. *Scottish Geographical Journal*, *120*, 199–210. https://doi.org/10.1080/00369220418737203

Wilson, P., Lord, T., & Rodés, Á. (2013). Deglaciation of the eastern Cumbria glaciokarst, northwest England, as determined by cosmogenic nuclide (^{10}Be) surface exposure dating, and the pattern and significance of subsequent environmental changes. *Cave and Karst Science*, *40*(1), 22–27.

Winkler, S. (2009). First attempt to combine terrestrial cosmogenic nuclide (^{10}Be) and Schmidt hammer relative-age dating: Strauchon Glacier, Southern Alps, New Zealand. *Central European Journal of Geosciences*, *1*(3), 274–290. https://doi.org/10.2478/v10085-009-0026-3

Winkler, S., & Matthews, J. A. (2014). Comparison of electronic and mechanical Schmidt hammers in the context of exposure-age dating: Are Q- and R-values interconvertible? *Earth Surface Processes and Landforms*, *39*(8), 1128–1136. https://doi.org/10.1002/esp.3584

Winkler, S., & Matthews, J. A. (2016). Inappropriate instrument calibration for Schmidt-hammer exposure-age dating (SHD) – A comment on Dortch *et al*., Quaternary Geochronology 35 (2016), 67–68. *Quaternary Geochronology*, *36*, 102–103. https://doi.org/10.1016/j.quageo.2016.08.009

Acknowledgements

We would like to thank C. Ballantyne, J. Clark, J. Everest and D. Small for providing sample photographs which were essential when conducting fieldwork. D. Fabel also kindly provided unpublished calibration data for the LLPR which permitted exposure age recalibration. We would also like to thank R. Pope, T. Tonkin and D. Tomkins for fieldwork assistance and J. Nudds for advice on suitable surfaces to sample. This study was partially funded by the University of Manchester SEED Fieldwork Support Fund. Hughes and Dortch would like to thank the University of Manchester Research Stimulation Fund.

Figure Captions

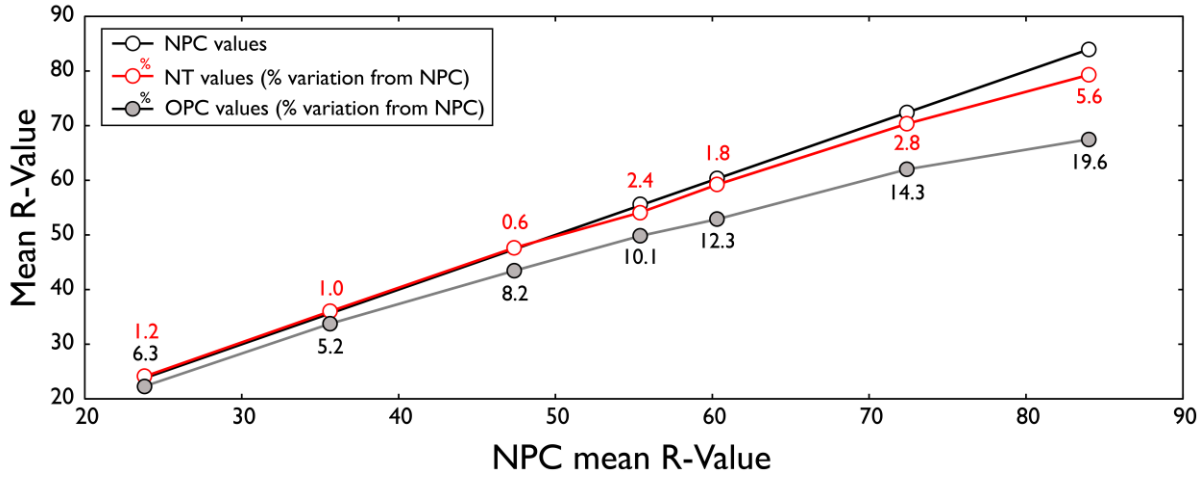

Fig. 1. Uncalibrated mean R-values for tested surfaces using the NPC, OPC and NT Schmidt Hammers.

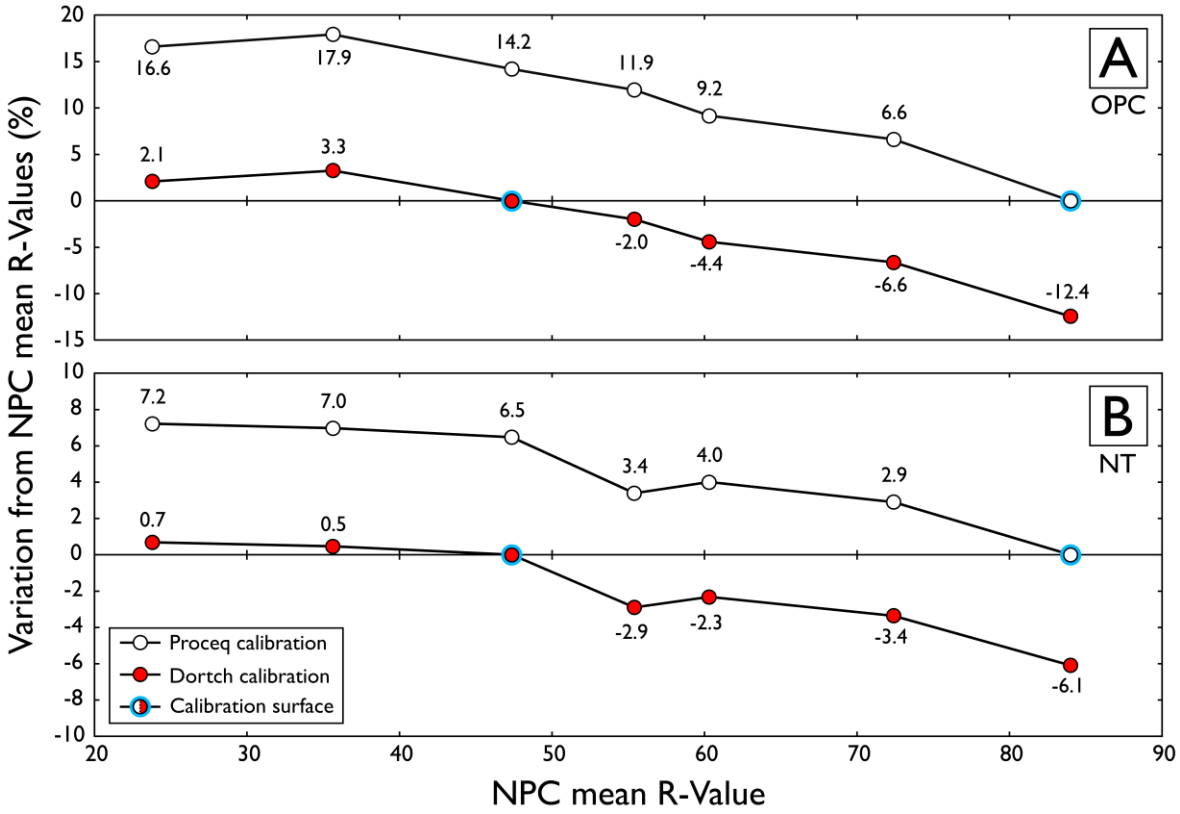

Fig. 2. Recalibrated mean R-values for tested surfaces using Proceq (2004) and Dortch et al. (2016) calibration procedures for the (A) OPC and (B) NT Schmidt Hammers.

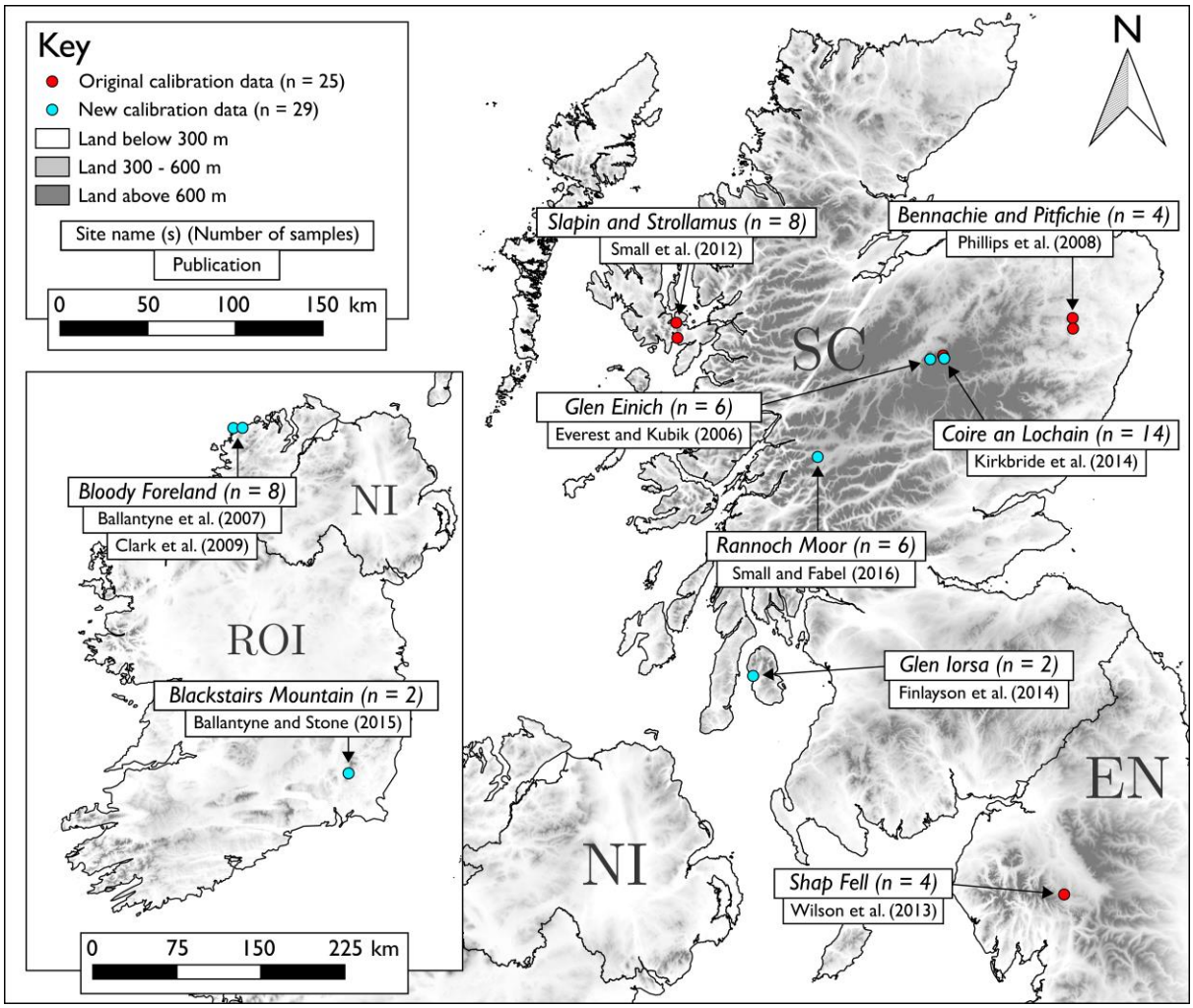

Fig. 3. Sample map showing the location of original (n = 25) and new calibration surfaces (n = 29) from across Scotland and NW England, including new sampled surfaces from Coire an Lochain (Kirkbride *et al.*, 2014), Rannoch Moor (Small and Fabel, 2016), Glen Einich (Everest and Kubik, 2006) and Glen Iorsa (Finlayson *et al.*, 2014). New early (> 20 ka) post-LGM samples from Ireland on Blackstairs Mountain, Wexford (Ballantyne and Stone, 2015) and at Bloody Foreland, Donegal (Ballantyne *et al.*, 2007; Clark *et al.*, 2009) are shown inset.

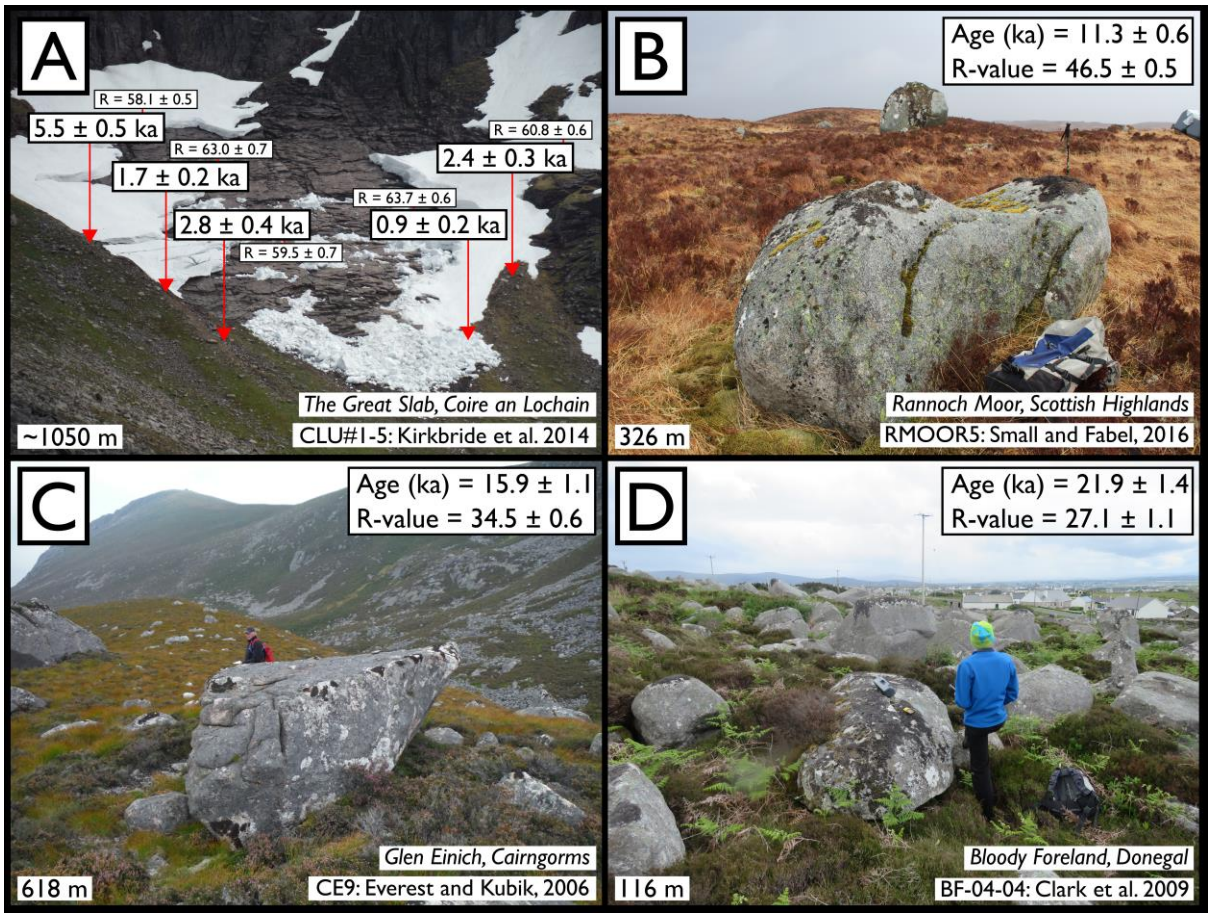

Fig. 4. Sample photos for (A) Holocene, (B) Younger Dryas, (C) Lateglacial Interstadial and (D) early post-LGM samples, displaying LLPR exposure ages (Fabel *et al.*, 2012), calibrated mean R-values, and sample elevations. The spread of exposure ages in Coire an Lochain (A) likely reflects the variable exposure of cliff surfaces to cosmogenic radiation (Kirkbride *et al.*, 2014).

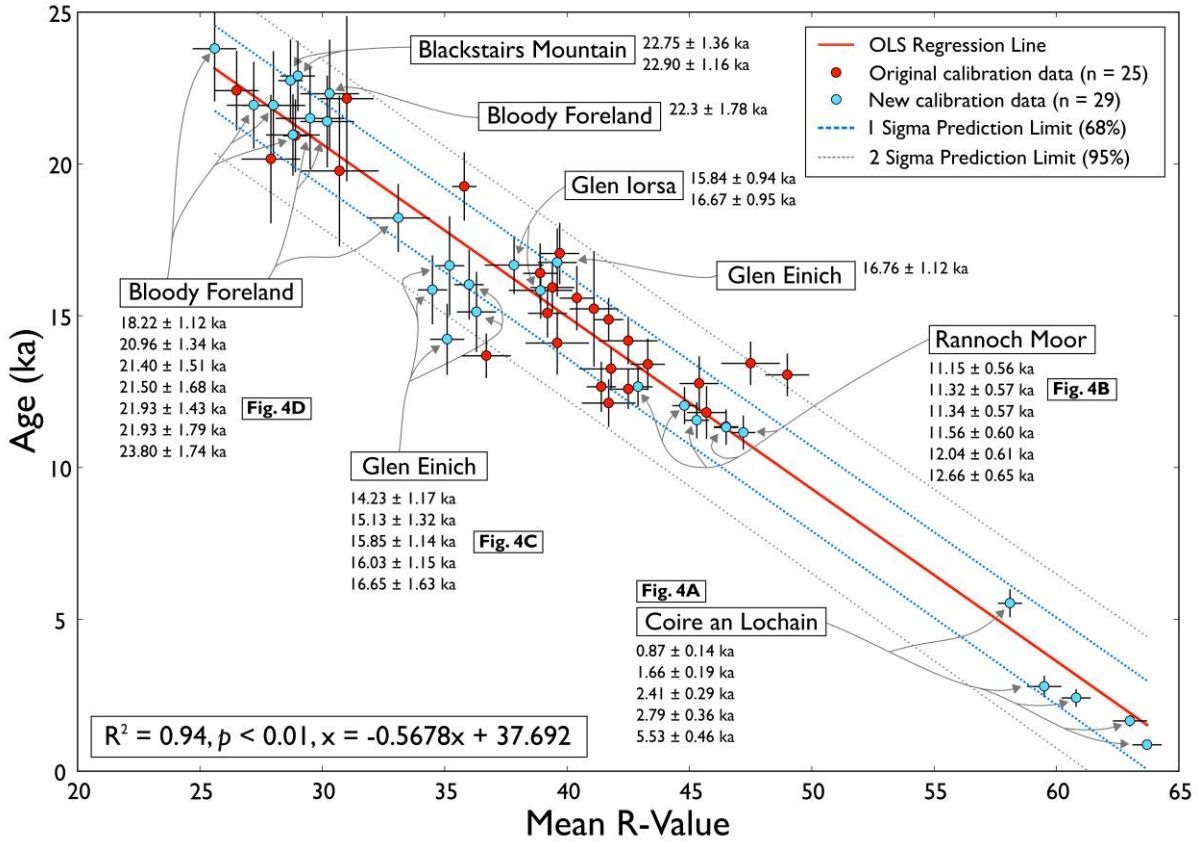

Fig. 5. Updated regional calibration curve for the British Isles (n = 54), displaying the least squares regression line (red), 1σ (blue) and 2σ (grey) prediction limits, and sample exposure ages for each new calibration site.

Table Captions

Table 1. Information on tested surfaces, raw R-value data for the NPC, NT and OPC Schmidt Hammers and T-test results.

				Mean R Values [b]						Difference from NPC (%)		T-test Results [c]		T-test Interpretations [d]	
Tested Surface	Location [a]	Latitude (°)	Longitude (°)	NPC	±	NT	±	OPC	±	NT	OPC	NT	OPC	NT	OPC
Test anvil	Arthur Lewis Building (interior)	-	-	84.0	0.0	79.3	0.7	67.5	2.6	5.6	19.6	< 0.01	< 0.01	H_1	H_1
Borrowdale Volcanic Group erratic	Old Quadrangle	53.465740	-2.234289	72.4	2.4	70.4	2.7	62.0	1.5	2.8	14.3	0.02	< 0.01	H_1	H_1
Polished Sandstone boulder	Bridgeford Street (rock garden)	53.466687	-2.235000	60.3	1.5	59.2	2.7	52.9	1.7	1.8	12.3	0.14	< 0.01	H_0	H_1
Marble pillar	Arthur Lewis Building (exterior)	53.466589	-2.235202	55.4	1.3	54.1	1.2	49.8	0.7	2.3	10.0	< 0.01	< 0.01	H_1	H_1
Doddington Sandstone boulder	Bridgeford Street (rock garden)	53.466639	-2.234881	47.4	1.9	47.6	2.1	43.5	1.8	-0.6	8.2	0.67	< 0.01	H_0	H_1
Concrete block	Bridgeford Street	53.466516	-2.234713	35.6	2.5	36.0	1.8	33.8	2.5	-1.0	5.2	0.61	0.02	H_0	H_1
Breezeblock	Arthur Lewis Building (interior)	-	-	23.8	1.4	24.1	1.7	22.3	1.2	-1.3	6.3	0.59	< 0.01	H_0	H_1

[a] University of Manchester properties, [b] Uncertainty estimates (±) are the mean absolute deviation, [c] *p* -values of two sample Student t-tests assuming unequal variance, [d] H_1 = the difference between the two population means is statistically significant, H_0 = the difference between the two population means is not statistically significant

Table 2. Calibration results using the Proceq (2004) and Dortch et al. (2016) calibration procedures for the OPC and NT Schmidt Hammers and comparison with baseline NPC R-values.

Proceq calibration [a]		Dortch calibration [a]		Proceq variance (%) [f]		Dortch variance (%) [f]		
NT [b]	OPC [c]	NT [d]	OPC [e]	NT	OPC	NT	OPC	Preferred calibration method [g]
84.0	84.0	78.9	73.6	0.0	0.0	-6.1	-12.4	*Proceq*
74.5	77.2	70.0	67.6	2.9	6.6	-3.4	-6.6	*Proceq*
62.7	65.8	58.9	57.6	4.0	9.2	-2.3	-4.4	Dortch
57.3	62.0	53.8	54.3	3.4	11.9	-2.9	-2.0	Dortch
50.4	54.1	47.4	47.4	6.5	14.2	0.0	0.0	Dortch
38.1	42.0	35.8	36.8	7.0	17.9	0.5	3.3	Dortch
25.5	27.8	24.0	24.3	7.2	16.6	0.7	2.1	Dortch

[a] Using correction factors of [b]1.05882, [c]1.24444, [d]0.994, [e]1.08972, [f] with respect to mean NPC R-values, [g] based on minimising the % variation between recalibrated R-values and the NPC