The CRONUS-Earth Project: A Synthesis

Fred M. Phillips^{a,*}, David C. Argento^b, Greg Balco^c, Marc W. Caffee^d, John Clem^e, Tibor J. Dunai^f, Robert Finkel^g, Brent Goehring^h, John C. Gosseⁱ, Adam M. Hudson^j, A.J. Timothy Jull^{j,k}, Meredith A. Kelly^l, Mark Kurz^m, Devendra Lal^{†,n}, Nathaniel Lifton^o, Shasta M. Marrero^{a,p}, Kunihiro Nishiizumi^q, Robert C. Reedy^r, Joerg Schaefer^s, John O.H. Stone^b, Terry Swanson^b, and Marek G. Zreda^t

^aEarth & Environmental Science Department, New Mexico Tech, Socorro NM 87801 USA

^bEarth & Space Sciences Department, University of Washington, Seattle, WA 98195 USA

^cBerkeley Geochronology Center, 2455 Ridge Road, Berkeley, CA 94709, USA

^dDepartment of Physics, Purdue University, West Lafayette, IN 47907-1396, USA

^eBartol Research Institute, Department of Physics and Astronomy, University of Delaware, Newark 19716

^fDepartment of Geosciences, University Cologne, 50939 Cologne, Germany

^gLawrence Livermore National Laboratory, Center for Accelerator Mass Spectrometry, Livermore, CA 94550 USA

^hDepartment of Earth and Environmental Sciences, Tulane University, New Orleans, LA 70118, USA

ⁱDepartment of Earth Sciences, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada

^jDepartment of Geosciences, University of Arizona, Tucson, AZ 85721, USA

^kNSF Arizona AMS Laboratory, University of Arizona, Tucson, AZ 85721, USA

^lDepartment of Earth Sciences, Dartmouth College, Hanover NH 03755 USA

^mWoods Hole Oceanographic Institute, Department of Marine Chemistry & Geochemistry, Woods Hole, MA 02543 USA

ⁿScripps Institution of Oceanography, Geological Research Division, La Jolla CA 92093, USA

^oDepartment of Earth, Atmospheric, and Planetary Sciences and Department of Physics and Astronomy, Purdue University, West Lafayette, IN 47907, USA

^pSchool of Geosciences, University of Edinburgh, Geography Building, Edinburgh, EH8 9XP, UK

^qSpace Sciences Laboratory, University of California, Berkeley, CA 94720-7450, USA

^rPlanetary Science Institute, 1700 E Fort Lowell, Suite 106, Tucson, AZ 85719-2395, USA

^sGeochemistry, Lamont-Doherty Earth Observatory (LDEO), Palisades, NY 10964, USA

^tDepartment of Hydrology and Water Resources, University of Arizona, Tucson, AZ 85721, USA

[†]Deceased

*Corresponding author: ^aEarth & Environmental Science Department, New Mexico Tech, 801 Leroy Place, Socorro NM 87801 USA. *E-mail address:* phillips@nmt.edu

Abstract

Geological surface-exposure dating using cosmogenic-nuclide accumulation became a practical geochronological endeavor in 1986, when the utility of ¹⁰Be, ²⁶Al, ³⁶Cl, and ³He were all demonstrated. In response to the lack of a common basis for quantifying analytical consistency and calibrating cosmogenic-nuclide production, the CRONUS-Earth Project in the U.S. was started in 2005, along with a European partner project, CRONUS-EU. The goal of the CRONUS-Earth Project was to improve the accuracy and precision of terrestrial cosmogenic nuclide dating in general, focusing especially on nuclide production rates and their variation with altitude, latitude, and time, and to attempt to move from empirically based methods to ones with a stronger basis in physics. The CRONUS-Earth Project conducted extensive intercomparisons of reference materials to attempt to quantify analytical reproducibility at the community level. We found that stated analytical uncertainties nearly always underestimate the actual degree of variability, as quantified by the over-all coefficient of variation of the intercalibration data. The average amount by which the actual coefficient of variation exceeded the

analytical uncertainty was a factor of two (100%), but ranged from 15% to 300% depending on the nuclide and material. Coefficients of variation ranged from 3-4% for ^{10}Be to 6-8% for ^{36}Cl , ^{14}C , and ^{21}Ne , to 5-11% for ^{26}Al . Both interlaboratory bias and within-laboratory excess spread of the data played a role in increasing variability above the stated analytical uncertainties. The physical basis for cosmogenic nuclide production was investigated through numerical modeling and the measurement of energy-dependent neutron cross sections for nuclide interactions. We formulated new, physically based, scaling models, denoted LSD and LSDn, by generalizing global numerical simulations of cosmic-ray processes. The CRONUS-Earth Project identified new geological calibration sites, including one at low latitude and high elevation (Huanacané, Peru), and replicated nuclide measurement at numerous laboratories. At many sites multiple nuclides were measured, providing much more confidence in the equivalence of surface-exposure ages calculated from differing nuclides. The data were interpreted using an original cosmogenic-nuclide calculator, CRONUSCalc, that incorporates the new physically based scaling. The new data and model produced significantly better fits than previous efforts, but do not fully resolve apparent spatial variations in production rates. The CRONUS-Earth and CRONUS-EU Projects have provided a firm foundation for assessing the strengths and weaknesses of cosmogenic-nuclide analytical methods, adjusted the AMS standards for ^{10}Be and consequently revised the half-life, and have provided improved calibration data sets and interpretative tools.

Keywords

Terrestrial cosmogenic nuclides; Beryllium-10; Aluminum-26; Chlorine-36; Carbon-14; Helium-3; production rate; interlaboratory comparison; scaling model

1. Introduction

On 7 April 1912, Victor Franz Hess, an Austrian physicist, ascended to an altitude of about 5000 m in a hydrogen-filled balloon. He carried with him electroscopes that demonstrated that the ‘background’ radiation increased markedly with height in the atmosphere (Carlson, 2012). For this investigation he would receive the Nobel Prize in Physics in 1936. The mysterious radiation he observed would soon be termed ‘cosmic radiation’ (to distinguish it from terrestrial sources of radiation) and for the next 40 years high-altitude balloons would provide physicists with their principal means of performing experiments in high-energy particle physics. In 1946, Willard Libby demonstrated that cosmic-ray neutrons produced abundant ^{14}C by reaction with atmospheric nitrogen (Libby, 1946) and that the ^{14}C could be used for geophysical dating (Arnold and Libby, 1949). This success led to the realization that rare nuclides produced by interaction with cosmic radiation might be used for the investigation of earth-surface processes. In 1955 Raymond Davis and Oliver Schaeffer attempted to measure, with mixed results, ^{36}Cl measured by means of decay counting to determine the surface-exposure age of an outcrop in Colorado, USA (Davis and Schaeffer, 1955). In the late 1970’s, accelerator mass spectrometry (AMS) and other accelerators were used to measure cosmogenic (cosmic-ray-produced) nuclides (Raisbeck et al., 1978a; Raisbeck et al., 1978b; Muller et al., 1978; Elmore et al., 1979; Raisbeck et al., 1983). However, analytical limitations

prevented routine application of terrestrial cosmogenic nuclides (TCN) until 1986, when the practicality of ^{36}Cl , ^{10}Be , ^{26}Al , and ^3He were all demonstrated within one year (Craig and Poreda, 1986; Kurz, 1986; Nishiizumi et al., 1986; Phillips et al., 1986). This enhanced capability was due to improvements in accelerator mass spectrometry (AMS) and noble-gas mass spectrometry.

Victor Hess' findings in 1912 prefigured both the potential and the challenges of terrestrial cosmogenic nuclide geochronology. By showing that the earth was blanketed by high-energy radiation arising from outer space, his research led to the realization that this radiation flux could provide a chronometer for earth-surface events. But his discovery that the radiation increased markedly with elevation demonstrated that some framework for comparing observations from different points on the Earth's surface would be required. The development of this conceptual framework remains a central challenge for the terrestrial cosmogenic nuclide (TCN) community, and was a primary CRONUS goal.

In 1991 Devendra Lal synthesized a range of atmospheric data related to cosmic-ray fluxes, which he originally measured and compiled for his Ph.D. dissertation (Lal, 1958), into polynomial equations that could be used to scale cosmogenic nuclide production anywhere on Earth to a common reference of high latitude and sea level. This scaling was a foundation of TCN research for the next decade. However, in the early 2000's the TCN community began to assess the extensive neutron-monitor data that had been collected by cosmic-ray physicists since Lal's dissertation work. This effort produced scaling results that differed quite significantly from those of Lal (1991) (Desilets and Zreda, 2001; Dunai, 2000; Lifton et al., 2005).

The need to assess numerous scaling formulations led to a realization that neither the methods nor the calibration data then available to the TCN community were up to the task. This realization motivated the formation of the European "CRONUS-EU Project", funded by the European Union (Stuart and Dunai, 2009) and the North American "CRONUS-Earth Project", funded by the U.S. National Science Foundation (described in this paper). "CRONUS" stands for "Cosmic-Ray-produced NUclide Systematics", and is also the name of the Greek god of time.

The overall goal of the CRONUS-Earth Project was to improve the accuracy and precision of cosmogenic dating, focusing especially on nuclide production rates and their variation with altitude, latitude, and time. The overall initiative comprised the following focus areas:

- i) analytical and sample-processing intercomparison
- ii) physics and measurement of cosmic-ray fluxes
- iii) artificial target experiments for testing scaling of cosmic-ray fluxes
- iv) improvement of nuclear-reaction cross sections applicable to TCN studies
- v) geological calibration of TCN production rates
- vi) development of a comprehensive numerical model to optimize production-parameter values and interpret data.

These focus areas are described in the following sections. While the motivation for these collective developments was to improve TCN research, the particle physics and scaling results have significant implications for the application of cosmogenic nuclides in extraterrestrial and atmospheric experiments.

2. Findings

2.1 Analytical and Sample-Preparation Intercomparison

Testing of global production-rate scaling formulations or assertions of precision of ages must be founded on quantitative knowledge of the true precision and accuracy of the nuclide analyses (and also other model inputs) on which the evaluations are based. Counting and other analytical uncertainties are a major source of variability(Elmore et al., 1984; Palonen and Tikkanen, 2007),but other factors external to these procedures could contribute to both analytical bias and the dispersion of measurements. Among these factors are uncertainties in sample weight, spike (carrier) weight, spike concentration, or spike isotope ratio. Differing procedures for mineral separation, leaching of contaminants, and ancillary chemical analyses can all contribute to interlaboratory variation. Processing blanks are also a critical requirement for analytical accuracy and quality control, but blank processing procedures that are incommensurate with sample procedures, or incorrect procedures for blank subtraction, can contribute to analytical bias. Some sources of variability may also be present in the AMS or the MS laboratory, particularly temporal drift of the instrumentation, if not adequately monitored, and improperly characterized AMS standards.

To establish a basis for evaluating interlaboratory consistency, the CRONUS-Earth and CRONUS-EU Projects produced intercomparison materials (ICM) and analyzed them in multiple laboratories. These are referred to as “intercomparison materials” rather than “standards” because they are not derived from sources of rigorously known isotope ratio or concentration, and thus their concentrations can only be established by consensus. The ICM’s, their sources, and their concentrations are listed in Table 1. The numbers of laboratories and replicates varied. For some nuclides and materials the numbers were considered sufficient to adequately characterize the nuclide concentration. These are reported as weighted means and standard errors of the mean. Others had smaller numbers of replicates and the averages were considered not to be definitive. For these, the unweighted mean and standard deviation are quoted (see Table 1).

Well-characterized materials		Weighted Mean and Standard Error of Mean Concentration (atoms/g)	Citation
Nuclide	Material		
¹⁰ Be	CRONUS-A	$(3.303 \pm 0.008) \times 10^7$	1
¹⁰ Be	CRONUS-N	$(2.075 \pm 0.012) \times 10^5$	1
³ He	CRONUS-P	$(5.02 \pm 0.06) \times 10^9$	2
²¹ Ne	CREU-1	$(3.48 \pm 0.10) \times 10^8$	3
Incompletely characterized materials		Unweighted Mean and Standard Deviation of Concentration (atoms/g)	Citation
Nuclide	Material		
²⁶ Al	CRONUS-A	$(1.43 \pm 0.07) \times 10^8$	1
²⁶ Al	CRONUS-N	$(1.05 \pm 0.11) \times 10^6$	1
¹⁴ C	CRONUS-A	$(6.93 \pm 0.44) \times 10^5$	1

1. Jull et al. (2015)

2. Blard et al. (2015)

3. Vermeesch et al. (2015)

Table 1. Consensus values for the nuclide concentrations in the CRONUS intercomparison materials. These are the values that the authors of these studies suggest for use as intercomparison materials.

2.1.1 ¹⁰Be Intercomparison Materials

The CRONUS-Earth Project manufactured two intercomparison materials for ¹⁰Be (half-life 1.36 ± 0.07 Ma) analysis, known as the ‘A’ and ‘N’ materials. Full results have been published by Jull et al. (2015), from which this account is summarized. The A material is composed of purified quartz and derives from sandstone with long exposure time from moderate elevation in Antarctica and has a relatively high concentration of ¹⁰Be and other nuclides. The N material, also quartz, is derived from beach sand from northern Australia and has a lower ¹⁰Be concentration that is typical of many field TCN samples. Analyses of the A material by 13 laboratories yielded a consensus concentration of $(3.303 \pm 0.008) \times 10^7$ atoms ¹⁰Be g⁻¹. The N material was analyzed by 12 laboratories and yielded a weighted mean concentration of $(2.075 \pm 0.012) \times 10^5$ atoms ¹⁰Be g⁻¹ (Jull et al., 2015).

In addition to the ICM’s described above, field samples collected by the CRONUS-Earth Project were also analyzed by multiple laboratories and can be used as ad-hoc intercomparison materials. These field samples were rigorously homogenized and split and sent to the participating laboratories. The two sites for which samples were analyzed by three or more laboratories were Promontory Point (Lifton et al., 2015), with six samples analyzed by four laboratories, and the Huanacané II moraines at the Quelccaya Ice Cap in southern Peru (Phillips et al., 2015c), with nine samples analyzed by three laboratories (see below for site descriptions). If potential variability due to sample-to-sample factors (e.g., variable snow cover or erosion) are kept in mind, these field samples can also be used for intercomparisons. Results are summarized in Table 2.

2.1.2 ^{26}Al Intercomparison Materials

The same A and N quartz separates described above are also the ^{26}Al ICM's (half-life 705 ± 17 ka). Seven laboratories measured the A material, with the resulting mean concentration $(1.43 \pm 0.07) \times 10^8$ atoms $^{26}\text{Al} \text{ g}^{-1}$ (uncertainty is standard deviation). Five laboratories measured the N material, yielding $(1.05 \pm 0.11) \times 10^6$ atoms $^{26}\text{Al} \text{ g}^{-1}$. In addition, five laboratories measured six samples from the Promontory Point field site (Table 2). Only one laboratory measured ^{26}Al at the Huancané II moraines, thus it cannot be used for interlaboratory comparison.

2.1.3 ^{36}Cl Intercomparison Materials

Two ^{36}Cl ICM's have been processed by Tibor Dunai at the University of Cologne (^{36}Cl half-life 301 ± 2 ka). These are relatively new and have not been formally published, and sufficient analytical results have not yet been compiled for statistical analysis. One ICM is named "CoCal-N" and consists of calcite. The calcite was obtained from a lag deposit adjacent to a calcite vein in the Namib Desert, Namibia. The material was sampled as large (5-15 cm) calcite crystals. Interiors of crystals were fully transparent (Iceland-spar quality) and inclusion-free (inclusion-bearing material was discarded). Weathering surfaces on crystals were removed by nitric acid leaching. The cleaned material was crushed and sieved and split using a rotation splitter. Material should be used 'as is'; no further cleaning or leaching is required. The other ICM is named "CoFsp-N" and consists of feldspar, also obtained from the Namib Desert. The feldspar came from a single large feldspar crystal from a pegmatite outcrop. The material was washed, and visually inspected. Any material having visible traces of other minerals (quartz, mica, tourmaline) was discarded. Material was crushed, sieved and sequentially leached (Kohl and Nishiizumi, 1992). Leached material was split using a rotation splitter. This material should be used as is, requiring no further cleaning or leaching. A splitter should be used to obtain further aliquots, since the albite and potassium feldspar components of the perthitic texture etched at different rates and the fine fraction may differ in composition from the coarser material.

Since sufficient ^{36}Cl data are not available for the intended intercomparison materials, CRONUS field data were used for this purpose. Seven basalt samples from the Tabernacle Hill basalt flow (Lifton et al., 2015) were subjected to whole-rock ^{36}Cl analysis by three laboratories. The samples varied from outcrop to outcrop in both Cl and Ca content and therefore cannot be directly compared in terms of ^{36}Cl concentration. For this analysis, therefore, the measured concentrations were normalized by the calculated ^{36}Cl production rate based on the measured chemical composition of each sample (Table 2). See Appendix A.5 for details.

2.1.4 ^{14}C Intercomparison Material

The CRONUS-Earth "A" intercomparison material, described above in section 2.1.1, was also used for intercomparison of ^{14}C analyses (half-life 5730 ± 30 a). Aliquots

of CRONUS A were measured by five laboratories, yielding a mean concentration of $(6.93 \pm 0.44) \times 10^5$ atoms ^{14}C (g quartz) $^{-1}$ (Jull et al., 2015).

2.1.4 ^3He Intercomparison Material

The CRONUS-Earth Project prepared an intercomparison material consisting of pyroxene grains for cosmogenic ^3He analyses (^3He is a stable nuclide). It is designated “CRONUS-P”. The raw material was isolated from a dolerite boulder collected from about 2500 m elevation in the Dry Valleys of Antarctica (77.9° S) (Blard et al., 2015; Schaefer et al., 2015). The material was crushed, leached, subject to heavy liquid isolation of pyroxene, and further purified on a Frantz magnetic separator (Schaefer et al., 2015). Aliquots of CRONUS-P were analyzed by six laboratories, yielding a mean concentration of $(5.02 \pm 0.06) \times 10^9$ atoms ^3He g $^{-1}$ (Blard et al., 2015). Helium-3 produced by low-energy neutron absorption by Li is negligible in comparison to the cosmogenic ^3He .

2.1.5 ^{21}Ne Intercomparison Material

The CRONUS-EU Project has provided an intercomparison material for cosmogenic ^{21}Ne analyses (^{21}Ne is a stable nuclide). It is named “CREU-1” and was prepared from quartz eroded from a vein onto the land surface in the Atacama Desert of Chile (930 m elevation and ~19° S) (Vermeesch et al., 2015). It was prepared by crushing, sieving, leaching, and homogenizing. Aliquots were analyzed by five laboratories, yielding a mean concentration of $(3.48 \pm 0.10) \times 10^8$ atoms ^{21}Ne g $^{-1}$.

2.2 Results and Discussion of Intercomparisons

One major objective of the intercomparison study was to provide well-characterized intercomparison materials that could be used for ongoing interlaboratory standardization efforts. To accomplish this, the results from all of the laboratories participating were evaluated and averaged. The results for the various intercomparison materials are given in Table 1. These values should be used as the basis for evaluation of future isotopic analyses of the ICM.

In addition to the consensus values for the nuclide concentrations of the materials, in this paper we perform a more comprehensive evaluation of analytical reproducibility and sources of uncertainty based on both the CRONUS intercomparison materials and field samples that were extensively replicated. The results of this evaluation are summarized in Table 2. The table lists for each intercomparison study the average analytical uncertainty as a percentage of the mean reported analytical value and the coefficient of variation of the reported data. The “Mean and One Standard Deviation” values in this table are the simple arithmetic mean and the one-sigma standard deviation calculated directly from the analytical values, with each analysis weighted equally. These values may sometimes differ from those given in the source publications because of variations in statistical procedure, such as use of error-weighted means or means based on laboratory averages rather than individual analyses. One-sigma statistics are consistently used throughout this paper.

Table 2 also lists the difference between the coefficient of variation and the average analytical uncertainty. If the reported analytical uncertainties are encompassing all of the actual sources of variation in the data sets, the coefficient of variation should approximately equal the average analytical uncertainty. This is clearly not the case. For all of the intercomparisons, the coefficient of variation exceeds the average analytical uncertainty. On average, coefficients of variation are approximately twice the average analytical uncertainty for most materials, ranging from 15% larger for ^{10}Be and ^{26}Al measurements on the CRONUS ‘N’ material to more than 4 times larger for ^{14}C on the CRONUS ‘A’ and ^{21}Ne on the CREU-1 materials. On average, the coefficients of variation are about twice as large as the reported analytical uncertainties.

This excess, most notable for ^{14}C and ^{21}Ne , is an indication that in most cases the reported analytical uncertainties are underestimating the actual variability of the results from the measured materials, geologic parameters (e.g., ages, erosion rates) determined using the concentration data likewise may be systematically underestimating the uncertainty. Provided that the uncertainties are appropriately quantified,

Table 2. Summary statistics on CRONUS intercomparison materials (ICM) and on field sites with multiple laboratories preparing samples. All deviations and related statistics reported as one sigma.

Nuclide	Material or Locality	Type	Number Labs	Number Analyses	Average Analytical Uncertainty	Average Analytical Uncert (% mean)	Data Mean and One St Dev	Data Coeff Variation (%)	Coeff Variation Minus Average Analytical Uncertainty (%)	Overdispersion (%)	Notes
^{10}Be	CRONUS-A	ICM	13	29	$5.68 \times 10^5 \text{ at/g}$	1.66	$(3.42 \pm 0.102) \times 10^7 \text{ at/g}$	3.00	1.34	1.9	1
^{10}Be	CRONUS-N	ICM	12	23	$7.60 \times 10^5 \text{ at/g}$	3.50	$(2.17 \pm 0.087) \times 10^5 \text{ at/g}$	4.00	0.50	3.3	2
^{10}Be	Huancane II	Field	3	25	$9.67 \times 10^5 \text{ at/g}$	1.89	$(5.11 \pm 0.155) \times 10^5 \text{ at/g}$	3.02	1.13	2.2	3
^{10}Be	Promontory Pt	Field	4	39	$5.36 \times 10^5 \text{ at/g}$	2.08	$(2.57 \pm 0.102) \times 10^5 \text{ at/g}$	3.97	1.89	3.0	4
^{26}Al	CRONUS-A	ICM	7	13	$4.23 \times 10^5 \text{ at/g}$	2.98	$(1.42 \pm 0.075) \times 10^5 \text{ at/g}$	5.28	2.30	3.6	2
^{26}Al	CRONUS-N	ICM	5	10	$9.19 \times 10^5 \text{ at/g}$	8.64	$(1.06 \pm 0.108) \times 10^5 \text{ at/g}$	10.1	1.46	0.0	2
^{26}Al	Promontory Pt	Field	4	25	$8.64 \times 10^5 \text{ at/g}$	5.04	$(1.71 \pm 0.864) \times 10^5 \text{ at/g}$	10.7	5.66	10	4
^{36}Cl	Tabernacle Hill	Field	3	24	1020 yr	5.11	$(2.00 \pm 0.16) \times 10^4 \text{ yr}$	8.03	2.92	6.2	5
^{14}C	CRONUS-A	ICM	5	23	$1.25 \times 10^4 \text{ at/g}$	1.82	$(6.89 \pm 0.527) \times 10^5 \text{ at/g}$	7.65	5.83	7.1	2
^{14}C	Promontory Pt	Field	2	19	$1.56 \times 10^4 \text{ at/g}$	4.44	$(3.51 \pm 0.224) \times 10^5 \text{ at/g}$	6.39	1.95	3.9	6
^3He	CRONUS-P	ICM	6	62	$1.27 \times 10^5 \text{ at/g}$	2.52	$(5.02 \pm 0.192) \times 10^5 \text{ at/g}$	3.82	1.30	3.1	7
^{21}Ne	CREU-1	ICM	5	50	$7.78 \times 10^5 \text{ at/g}$	2.26	$(3.44 \pm 0.244) \times 10^5 \text{ at/g}$	7.08	4.82	7.1	8

1. Data from Juli et al. (2015), sample 8a omitted

2. Data from Juli et al. (2015)

3. Data from Phillips et al. (2015), normalized for elevation differences and topographic shielding, samples 15 and 16 LDEO omitted

4. Data from Lifton et al. (2015), normalized for topographic shielding

5. Data from Marrero et al. (2015), normalized for compositional heterogeneity, elevation differences, and topographic shielding

6. Data from Lifton et al. (2015), normalized for topographic shielding

7. Data from Blard et al. (2015)

8. Data from Vermeesch et al. (2015), GFZ b3 (K088) omitted

the uncertainty-weighted mean concentration provides the best estimate of the actual concentration (Bevington and Robinson, 1992). However, if the actual sources of uncertainty are not adequately quantified, then the weighted mean may no longer provide an accurate estimate.

In order to more fully quantify the unaccounted-for sources of variability we employ a statistical model that includes both the stated analytical uncertainty and an additional unspecified source of variability (Galbraith and Laslett, 1993; Galbraith et al., 1999). This methodology simultaneously solves for both the “central value” of the data set and the “overdispersion” of the data set (i.e., the component of the spread that cannot be attributed to the analytical uncertainty alone) and should provide a more robust estimate of the actual composition than the weighted mean if the data set is characterized as overdispersed (Vermeesch, 2009). We have not reported the central values because in

all cases they differed insignificantly from the mean values, indicating that the amount of excess spread of the data was not so large that it invalidated the standard statistical model. The magnitude of the calculated overdispersion ranges from zero percent (for ^{26}Al measurements on the CRONUS-N material) to 10% (again, for ^{26}Al , this time from the Promontory Point field data set). This comparison highlights the role of the assigned analytical uncertainty in the magnitude of the overdispersion. The average analytical uncertainty reported for the CRONUS-N data, 8.64%, is similar to the coefficient of variation of the analyses, 10.1%. The coefficient of variation of the Promontory Point data set is quite similar, 10.7%, but the average reported analytical uncertainty was only 5.04%. The Promontory Point data therefore appear overdispersed whereas the CRONUS-N data do not.

The data in Table 2 contain both good and bad news for practitioners of cosmogenic-nuclide dating. First and foremost, the dispersion is not large enough to invalidate the standard statistical model, so the mean is statistically significant. The next good news lies in the comparison of ICM and field data. For cases where the ICM and the field samples contain approximately the same nuclide concentrations, the coefficients of variation for the two data sets are always very similar (e.g., the coefficient of variation for ^{10}Be for the CRONUS-N material was 4.00%, while that for the similar-concentration Promontory field data was 3.97%).

The sources of variability in analytical results from field studies can be conceptualized as arising from three principal sources: (1) counting statistics of the sample material, blanks, and standards, as well as other random sources of uncertainty including scales, pipettes, etc., (2) systematic error due to inaccuracies in subsidiary measurements (e.g., inaccurate scales or an inaccurate isotope ratio in a spike) or differences in sample preparation procedure that result in differing concentrations of the nuclide in the final material, or inadequacies in the standardization of the mass spectrometer, and (3) sample-to-sample differences resulting from either variations in environmental conditions (e.g., snow or sand cover) or chemical/mineralogical heterogeneities that affect the analyses. The results from the field-versus-laboratory comparison demonstrate that the third source of variability is small compared to the other two, at least for these particular field sites. They also demonstrate that the CRONUS calibration sites do yield analytical results that are as consistent as can be reasonably achieved, given the state of the art.

The bad news is that typical reported analytical uncertainties significantly underestimate the actual magnitude of analytical variability. In many cases the magnitudes of the reported analytical uncertainty and the overdispersion are approximately equal, indicating that the actual uncertainties are underestimated by a factor of approximately 40%. For a few (e.g., ^{26}Al at Promontory Point, ^{14}C in the CRONUS-A material, and ^{21}Ne in CREU-1) the overdispersion is significantly larger than the average reported uncertainties, and thus dominates the total uncertainty.

A goal of the CRONUS-Earth Project was to reduce age uncertainties to below five percent (i.e., using empirically based estimation of analytical reproducibility rather than often-underestimated laboratory-reported uncertainties). For this to be achieved, basic sample reproducibility must be well below five percent, assuming that scaling and other sample variables introduce some amount of variability. In general, the quoted analytical uncertainties are within this bound (^{26}Al is the main exception, with ^{36}Cl and

¹⁴C somewhat higher than is desirable). As evidenced by the coefficients of variation of the various laboratory and field data sets, this requirement is only partially been achieved. The coefficient of variation for ¹⁰Be, for example, is consistently below four percent, as is that for the ³He laboratory ICM. The coefficients of variation of the other nuclides range between 6 and 10%. Appendix A provides a detailed statistical analysis of all of the available analytical intercomparison data sets from both laboratory and field studies, with the goal of identifying the sources of variability (Jull et al., 2015; Lifton et al., 2015; Phillips et al., 2015c; Blard et al., 2015; Schaefer et al., 2015; Vermeesch et al., 2015).

2.2.1 Evaluation of intercomparison results

We will conduct this evaluation in the context of a goal of cosmogenic age accuracy and precision less than 5%, and thus of analytical concentrations smaller than that value. The principal tools for evaluation is the coefficient of variation of the various data sets listed in Table 2 and the results of the detailed statistical evaluation in Appendix A. We will review the nuclides in order of increasing coefficient of variation (CoV).

In the following discussion we refer to within-laboratory variability to reflect reproducibility among aliquots of the same sample measured by a given laboratory, and inter-laboratory reproducibility to reflect the differences in the nuclide concentrations of a given ICM among the different participating laboratories. For noble gases where samples were processed and measured at the same location, the source of the interlaboratory variability can be linked directly to the individual labs. However, for the radionuclides, where target preparation is often done at labs unrelated to the AMS facility, the inter-laboratory and within-laboratory variations result from a combination of sources at the target preparation laboratory and the AMS laboratory.

The smallest coefficients of variation are calculated for ¹⁰Be. These values are 3.0% for the high-level CRONUS-A ICM and 4.0% for the more typical level CRONUS-N material, close to the 5% goal. However, the accomplishment is more impressive when a smaller sub-group of laboratories is considered. If only the self-selected group of five laboratories that analyzed three or more replicates is evaluated (see Appendix A), the CoV for the CRONUS-A material goes down to 1.5%. This accomplishment is quite significant, indicating that experienced laboratories can achieve analytical reproducibility that is well below the minimal 5% level and that is consistent with the reported analytical uncertainties. Similarly, for the groups of four and three laboratories, respectively, that analyzed the Promontory Point and Huancané field sample sets, the CoV's were 3.8 and 3.0%. There was a small-to-moderate amount of overdispersion in both the overall laboratory and field data sets that apparently mostly resulted mostly from underestimation of random error, but also to a lesser extent was attributable to interlaboratory bias. Although the goal of a calibration data set root-mean-squared error of less than 5% was not met (see below), the analytical results are generally encouraging.

Helium-3 is apparently in the same general range, with a CoV of 3.82% for the CRONUS-P material. However, this is a notably 'hot' ICM, and whether similar reproducibility can be obtained for ICM's closer to typical sample concentrations remains to be demonstrated. The data are significantly overdispersed, with the main cause apparently interlaboratory bias rather than internal scatter.

Following this is a group of nuclides whose reproducibility falls in the range of 6 to 8%, including ^{36}Cl , ^{14}C , and ^{21}Ne . The statistics for two of these, ^{36}Cl and ^{14}C , were notably from field datasets with typical concentrations for field samples. The ^{36}Cl data were moderately overdispersed, with the origin fairly clearly being interlaboratory bias (Appendix A). However, the moderate CoV of the ^{36}Cl ages overlies a larger scatter of the Cl and ^{36}Cl concentrations on which the normalized concentrations are based (Lifton et al., 2015). This larger scatter collapses into smaller normalized dispersion because the Cl concentration, ^{36}Cl concentration, and ^{36}Cl production rate are all strongly correlated, and since the production rate increases as the Cl and ^{36}Cl concentrations increase, the spread of the resultant ages is damped. This analytical phenomenon is clearly contributing to the interlaboratory bias and needs further investigation. The variation also clearly relates to a relatively high level of uncertainty in predicting ^{36}Cl production due to low-energy neutron absorption by ^{35}Cl (see Appendix B.2 and Marrero et al., 2015c).

The ^{14}C data on CRONUS-A are strongly overdispersed relative to quoted analytical uncertainties. The culprit appears to mainly be underestimation of random within-laboratory sources of variability, but interlaboratory bias apparently is also making a subsidiary contribution (Appendix A). The ^{21}Ne data, on CRONUS-P, shows a similar large overdispersion relative to the quoted analytical uncertainties. Here the cause is reversed from ^{14}C : interlaboratory bias appears to predominate, with a smaller contribution from underestimation of within-laboratory scatter. For all three of these nuclides (^{36}Cl , ^{14}C , and ^{21}Ne), significant improvement will have to be made to reach the 5% age-uncertainty goal, requiring that the analytical spread be cut roughly in half. This would appear to be achievable, although not without intensive effort.

The nuclide with the lowest level of reproducibility is ^{26}Al , with CoV's for the 'hot' CRONUS-A material (concentration of 1.43×10^8 atoms $^{26}\text{Al} \text{ g}^{-1}$ is higher than the great majority of natural samples) of 5.3% and for the CRONUS-N and Promontory datasets, having more typical sample concentrations, between 10 and 11%. The data (with the exception of CRONUS-N, which had more realistic average uncertainty reported) are strongly overdispersed. This overdispersion apparently arises in approximately co-equal amounts from random within-laboratory and systematic between-laboratory sources (Appendix A). This lack of reproducibility is particularly unfortunate because ^{26}Al is commonly used in conjunction with ^{10}Be in burial dating. For this purpose, the ratio of the concentrations of the two nuclides is important, and poor effective precision in ^{26}Al will negate the potentially high precision of the corresponding ^{10}Be . Aluminum-26 analysis will clearly require intensive effort if it is to approach the 5% goal. Considering that the ^{10}Be chemistry in the same samples achieved better reproducibility, a primary source of overdispersion (both via within-laboratory and between laboratory sources) may arise from the AMS measurement of $^{26}\text{Al}/^{27}\text{Al}$. The use of gas-filled magnets offers hope of significant improvements to the results from AMS measurement of ^{26}Al (Fifield et al., 2007). Measures to improve analytical reproducibility are discussed in Phillips et al. (2015a).

2.3 Scaling of the Contemporary Cosmic Hadron Flux

Spatial scaling of cosmogenic-nuclide production has principally been based on measurements of the contemporary cosmic-ray flux; the data sources vary between proposed scaling models. The original model of Lal (1991) was largely based on photographic records of spallation events (“stars”) in stacks of photoemulsions, but also included a range of other data that have never been fully documented. Later models reformulated the Lal model with fairly minor corrections or additions (Balco et al., 2008; Stone, 2000). Another family of models, however, used data from the extensive neutron-monitor network maintained by cosmic-ray physicists (Desilets and Zreda, 2003; Desilets et al., 2006; Dunai, 2001; Lifton et al., 2005). These models produced scaling of production with elevation in regions of high magnetic strength (i.e., equatorial) that differ significantly from the Lal-based models. Recently, as part of the CRONUS-Earth Project, Lifton et al. (2014) have introduced yet another approach to scaling, based on numerical simulations of cosmic-ray modulation by the geomagnetic field and the atmosphere that are calibrated using a relatively limited number of atmospheric differential flux measurements for neutrons and protons. Finally, Argento et al. (2015a, 2015b) have also used physics-based modeling to investigate scaling phenomena.

The new scaling models developed by Lifton et al. (2014) represent a major advance in terrestrial cosmogenic-nuclide applications. Previous scaling models have been largely empirical, based on the supposition that cosmogenic-nuclide production rates for all nuclides are proportional to some form of monitored cosmic-ray counts. The models of Lifton et al. (2014) are based on analytical approximations fit to nuclear-physics simulations of the cosmic-ray cascade by Sato and Niita (2006) and Sato et al. (2008), using the PHITS model. The analytical approximations agree well with the atmospheric and land-surface measurements of the neutron-energy spectrum by Goldhagen et al. (2004), Kowatari et al. (2005), and Nakamura et al. (2005). By incorporating measured or estimated energy-dependent neutron cross sections (see Reedy, 2013), the “nuclide-dependent” model (LSDn) can scale production on the basis of individual nuclides rather than treating all nuclides equivalently. This new approach carries the promise that, as understanding of the underlying nuclear physics improves, so too will the accuracy of the scaling.

This range of scaling models, yielding in many cases distinctly different spatial distributions of cosmogenic nuclide production and modulation by fluctuations in the geomagnetic field, has highlighted the need for performing direct measurements of contemporary cosmogenic nuclide production that can be related to contemporary measurements of the cosmic-ray flux. There is also a clear need for better understanding of the relation between the responses of cosmic-particle detectors and cosmogenic-nuclide production. These needs stimulated a series of experiments by CRONUS-Earth investigators. One was to investigate the relation between cosmic-ray neutron fluxes and neutron-monitor responses. The second was to emplace a series of targets for cosmogenic nuclide production on an elevation transect in Hawaii, an area of strong geomagnetic modulation. The nuclides measured were selected to be sensitive to production by the high-energy portion of the neutron and proton flux (^{10}Be produced from oxygen in water), to a wide range of energies (^3He production from oxygen in water), and to low-energy neutrons (^{32}P production from sulfur). Owing to the long-term nature of these investigations, results are still pending.

The implications of the neutron-monitor-response investigations have been summarized by Lifton et al. (2014). Briefly, neutron monitors produce a larger internal neutron cascade in response to an interaction with high-energy neutrons relative to low-energy neutrons, an effect known as multiplicity. The hardness of the neutron spectrum (the proportion of high-energy neutrons to low-energy ones) increases as geomagnetic rigidity increases toward the equator. This results in neutron monitors registering a more rapid increase of counts with increasing altitude than the overall flux. Lifton et al. (2014) conclude that this causes neutron-monitor-based scaling models to overestimate production rates at high altitude and low latitude.

2.4 Neutron-Beam Experiments to Determine Nuclear Reaction Cross-Sections

The cosmic-ray hadrons (mostly neutrons) that pass through Earth's surface have energies ranging from GeV per nucleon to thermal (Clem et al., 2004; Gordon et al., 2004; Heidebreder et al., 1971). The cross sections for production of cosmogenic nuclides of interest for geochronology vary as a function of the particle energy. If modeling of nuclide production is to be based on physics, as opposed to the traditional purely empirical approach (e.g., Lal, 1991), accurate knowledge of the energy-dependence of the reaction cross sections is a necessity.

In most cases of interest the cross sections for proton-induced reactions are well known (Nishiizumi et al., 2009b), but data on those for neutron-induced reactions are sparse, largely because of the difficulty of producing energetic (especially monoenergetic) neutron beams in the laboratory. However, the hadronic cosmic-particle flux at ground level is strongly dominated by secondary neutrons, so the lack of data on neutron-induced reactions is a serious limitation.

The CRONUS-Earth Project therefore sponsored measurements of reaction cross sections in laboratory neutron fluxes of varying energies. The results to date have been reported by Nishiizumi et al. (2009); Reedy (2013), and Caffee et al. (2013). These were done in three experimental settings: (1) quasi-monoenergetic neutrons generated using the iThemba LABS cyclotron in Somerset West, South Africa, (2) neutrons having a 'white' energy spectrum generated by the LANSCE (Los Alamos Neutron Science Center) facility at Los Alamos National Laboratory in New Mexico, USA, and (3) monoenergetic (~100 to 390 MeV) neutrons at the Research Center for Nuclear Physics (RCNP), Osaka University, Japan. Targets of SiO₂ and other compounds and elements important for terrestrial cosmogenic-nuclide production were irradiated, dissolved and purified, and nuclide concentrations measured by AMS.

To date, results have been reported for the production of ¹⁰Be and ²⁶Al by reactions on O and Si (Caffee et al., 2013). The cross section for the O(n,x)¹⁰Be neutron reaction below ~200 MeV was 3 to 5 times higher than the corresponding proton reaction, whereas the neutron reaction for Si(n,x)²⁶Al gave cross sections very similar to the proton one. The ²⁶Al/¹⁰Be ratio from the 'white' neutron experiments (i.e., a broad neutron energy spectrum) was close to four, compared to the observed terrestrial ratio of about seven, demonstrating the energy-dependence of the nuclide production. Target dissolutions and analyses are continuing and should ultimately yield additional data on ¹⁰Be and ²⁶Al cross sections, as well as those for ³⁶Cl and ⁵³Mn.

2.5 Physics-Based Modeling of Cosmogenic Nuclide Production

One of the original goals of the CRONUS-Earth Project was to develop a stronger physical basis for the computation of cosmogenic ages. This was partly accomplished by means of the LSD and LSDn scaling methods (Lifton et al., 2014) that employ analytical fits to the outputs of numerical simulations of the cosmic-ray cascade. The analytical fits greatly reduce the computational burden, but they also place significant limitation on the range of problems that can be addressed. The Project therefore also supported independent numerical modeling experiments designed to address specific issues pertinent to cosmogenic nuclide dating.

These results are given by Argento et al. (2015a, 2015b). They employed the galactic cosmic-ray energy spectra of Usoskin and Kovaltsov (2006) at the top of the atmosphere, as modified by the geomagnetic cutoff calculations of O'Brien (2008). The modified spectra were used as the source radiation for simulations in MCNPX, a general purpose, relativistic, Monte Carlo-style radiation-transport code. This approach allowed the flexibility to address a range of specific issues.

One issue investigated was the effect of modeled angular-dependence of the cosmic-ray flux (as opposed to the commonly assumed $\cos^{2.3}\theta$ dependence) on topographic-shielding corrections. They found that the near-horizontal cosmic-ray flux was greater than that calculated by the standard formulation. For steep shielding angles ($>45^\circ$), this increased the magnitude of the topographic shielding by up to 25%. At lesser, more commonly encountered, shielding angles, the differences are still significant, and may contribute to scatter in the calibration data and other data set. They present trigonometric fits for the angular dependence as a function of atmospheric depth and cutoff rigidity.

They were able to simulate the shift in the cosmic-ray energy spectrum with depth in the atmosphere and used this as the basis for calculating nuclide-dependent production rates with depth (Argento et al., 2015a). At high-mountain elevations the production-rate ratios can differ from sea-level ones by as much as 15%. Their results are similar to the generalized ones of Lifton et al. (2014), but not identical. Lifton et al. (2014) have quantified the differences in the models. Understanding the causes of the differences and how well they correspond to empirically measured values should be a priority for future research.

Argento et al. (2015a) also examined the depth-dependence of cosmogenic nuclide production below the land surface. Their simulations show that the production attenuation length is both depth dependent and nuclide dependent. Nuclide production ratios will thus vary with depth. The variation (relative to surface production) at 200 g cm^{-2} depth can be as large as 20% for ^{36}Cl produced from K, but for other nuclides is 6% or less. They also found that for ^{36}Cl from K the profile of production with depth showed a flattening just below the surface, as predicted by Masarik and Reedy (1995), but other nuclides did not. The difference between the ^{36}Cl -from-K reaction and those of other nuclides is the low energy threshold (~ 1 MeV) for this reaction. Given the complexities of the nuclear cascade in this depth region, the use of simple exponential models for the depth-dependence of production should be reexamined.

2.6 Cosmogenic-Nuclide Dating Calculators

One of the goals of the CRONUS-Earth Project was to provide the cosmogenic nuclide community with easy-to-use computational tools for interpreting cosmogenic data that would be accessible on line. We dealt with this in a two-step fashion. The first was the implementation of the calculator described by Balco et al. (2008). This provided a relatively complete treatment of processes affecting ^{10}Be and ^{26}Al , but was limited to these nuclides. The second step was a more comprehensive on-line calculator that would treat all five of the commonly employed nuclides in a consistent fashion. It is named ‘CRONUScalc’ and is described by Marrero et al. (2015a).

The CRONUScalc program directly incorporates much of format and function of the $^{26}\text{Al}/^{10}\text{Be}$ code by Balco et al. (2008), but extends the program to include multi-nuclide capability (^{10}Be , ^{26}Al , ^{36}Cl , ^3He , ^{14}C , with ^{21}Ne in testing), full propagation of uncertainties, new scaling methods, and the ability to work with single samples at depth. The new CRONUScalc online interface (<http://web1.ittc.ku.edu:8888/1.0/>) performs scaling, shielding, and surface exposure age calculations. The code behind the online interface also includes functions to perform production/accumulation calculations, calibrations, and surface and depth-profile exposure ages.

The new CRONUScalc code and accompanying calculator were designed to incorporate all the major contemporary advances in the understanding of cosmogenic nuclide production and to treat all the commonly used nuclides in a completely consistent fashion. Perhaps the most important addition is the new scaling framework by Lifton et al. (2014), which allows for a physics-based calculation of the incoming cosmic-ray flux as well as nuclide-dependent scaling frameworks that incorporate the individual nuclide reaction cross sections. Other key features of CRONUScalc include differences in numerical integration methods, muon reparameterization, default reference production rates derived from the new CRONUS-Earth calibration data set, and more comprehensive uncertainty calculations. The model for depth-dependence of production rates used in exposure-age calculations eliminates some approximations and simplifications used in the 2008 online calculator, so exposure-age calculations can be applied to subsurface samples. CRONUScalc now uses an updated geomagnetic history as well as the new CRONUS-Earth production rates (Borchers et al., 2015; Marrero et al., 2015c; this paper). Balco et al. (2008) implemented a muon production model by Heisinger et al. (2002 a, b) that calculates the production from muon reactions using physical parameters determined from laboratory irradiation experiments. The parameters in the Heisinger muon model have been shown to overpredict the production (Braucher et al., 2003; Braucher et al., 2011; Kim and Englert, 2004; Borchers et al., 2015), so CRONUScalc uses the Heisinger model with parameters calibrated from an original CRONUS-Earth deep profile (Borchers et al., 2015; this paper, Figure 1) for ^{10}Be and ^{26}Al and previously published profiles for ^{36}Cl -K (Evans, 2001) and ^{36}Cl -Ca (Stone et al., 1998). CRONUScalc allows the user to incorporate more complete uncertainties for samples by allowing uncertainties on every input parameter and reporting internal and external uncertainties separately.

The online interface for the code provides access to the surface exposure and depth-profile dating capabilities of CRONUScalc. The spreadsheet-style input format and downloadable results make it easy to run samples without knowledge of coding. The new interface also allows users to enter an email address so results can be sent after the

647 samples have been completed without needing to leave a browser open during the
648 calculations. To maintain confidentiality, the server computer memory is deleted daily
649 and the only information that is saved (type of sample, abbreviated location) is used to
650 collect anonymous usage statistics on the number and type of samples.

651 The source code is available under the GNU General Public License agreement
652 and can be downloaded (<https://bitbucket.org/cronusearth/cronus-calc>) and modified by
653 advanced users to deal with specific atypical scenarios or to improve sampling or
654 chemistry approaches by predicting outcomes. The code was originally written in
655 MATLAB, but can be run in the open-source program Octave.

656 2.7 Geological Calibration of Cosmogenic Production-Rate Models

659 Empirical calibration of production-rate models using data from suitable
660 geological localities is the centerpiece for useful application of cosmogenic nuclides to
661 problems in the earth sciences. The CRONUS participants developed a dual
662 categorization of sites. “Primary calibration sites” were limited to a small number of
663 localities that received very high ratings in all, or nearly all, of the categories listed
664 above. This process was conducted by consensus of the CRONUS-Earth working group
665 and concluded in 2010. The sites accepted as primary consisted of four localities that
666 were sampled as part of the CRONUS-Earth Project and an additional four previously
667 sampled localities from earlier studies. Only these sites were used for the actual
668 production-rate calibration. Other high-quality sites were subsequently published by
669 other investigators but unfortunately could not be included in this calibration process.
670 The other category was “secondary sites”. These were sites that were of very good
671 quality, but were not deemed as secure as the primary sites because of uncertainties in
672 several of the selection criteria. These consisted of 11 localities. Three were sampled as
673 part of the CRONUS-Earth Project and the remainder was garnered from previous
674 studies. These were used to assess the uncertainties associated with the ages calculated
675 from the production-rate/scaling-model pairs calibrated using the primary data set. The
676 primary and secondary localities were selected by consensus of the entire CRONUS-
677 Earth membership by being voted on at the annual meetings of the project.

678 2.7.1 Primary Calibration Sites

680 2.7.1.1 Lake Bonneville Sites

683 Two localities associated with the abandoned shorelines of ancient Lake
684 Bonneville were sampled as part of the CRONUS-Earth Project and were included in the
685 primary calibration data set. These have been described by Lifton et al. (2015). Lake
686 Bonneville was a ‘pluvial’ lake in central Utah, USA, that repeatedly expanded during
687 intervals of glacial maxima during the Pleistocene. It reached an all-time high elevation
688 during Marine Isotope Stage 2. At this time, wave action cut prominent benches (the
689 ‘Bonneville shoreline’) into resistant metasedimentary rocks along the northern shore of
690 the lake. A large number of radiocarbon measurements, corroborated by U/Th ages,
691 indicate that the Bonneville shoreline was abandoned 18.36 ± 0.3 ka before 2010 CE (this
692 notation reflects 60 years added to the calibrated radiocarbon age referenced to 1950,

generated using INTCAL09 (Reimer et al., 2009)). The radiocarbon ages were on a variety of materials, including charcoal, wood, gastropod shells, shoreline tufa, and sub-lacustrine tufa deposited in caves (Lifton et al., 2015). At a locality named Promontory Point ('PPT') CRONUS-Earth members sampled quartzite outcrops from a shoreline platform >100 m wide. The sampled surfaces still retained wave polish. Four laboratories measured both ^{10}Be and ^{26}Al on these samples and two laboratories measured ^{14}C (Lifton et al., 2015). Promontory Point is therefore the most thoroughly replicated site. It appears to be very strong in all respects.

A second site associated with Lake Bonneville was also sampled by CRONUS-Earth. After Lake Bonneville withdrew from the Bonneville shoreline due to catastrophic erosion of the sill, it re-stabilized at the lower 'Provo shoreline'. After stabilization, a basaltic volcano erupted beneath the lake water and formed a sub-aerial lava flow now named Tabernacle Hill ('TAB'). The Provo shoreline, dated at the site by radiocarbon on tufa, is incised into the flow. Closely limiting radiocarbon ages have constrained the date of the eruption to 18.26 ± 0.3 ka before 2010 CE (Lifton et al., 2015). Very well-preserved primary flow-top morphology indicates minimal surface erosion at the site sampled. Samples from this site were analyzed for whole-rock ^{36}Cl (3 laboratories), ^{36}Cl in Ca-feldspar separates (1 laboratory), and ^3He in olivine separates (1 laboratory).

2.7.1.2 Huancané II Moraines, Quelccaya Ice Cap, Southern Peru

The Huancané moraines are on the west side of the Quelccaya Ice Cap in the southeastern Peruvian Andes (Mercer and Palacios, 1977). A detailed description of the geological setting and of CRONUS-related sampling at this site can be found in Phillips et al. (2015c) and Kelly et al. (2015). The moraines are located at $\sim 13.9^\circ \text{S}$ and 4850 m elevation and thus provide much-needed data on TCN production at low latitude and high elevation. The Huancané II moraines record a major expansion of the Quelccaya ice cap during the Younger Dryas stade (Kelly et al., 2012). The site is particularly suitable for calibration because the glaciers advanced into valley floors occupied by extensive wetlands that produced peat deposits. Peat was found both underlying and overlying the moraines in two valleys. Numerous radiocarbon ages on this peat produced concordant ages, limiting the moraine deposition to 12.3 ± 0.1 ka before 2010 CE.

The moraines are littered with large boulders of Miocene rhyolite. From the perspective of cosmogenic-nuclide calibration, the only significant issue with these boulders is surface erosion. Observation indicated significant surface erosion by granular disintegration, which was estimated to average 7 cm for the boulders sampled (Kelly et al., 2015; Phillips et al., 2015c); resulting in a $\sim 4\%$ reduction in nuclide concentrations. Samples from 10 boulders were analyzed for ^{10}Be , ^{26}Al , and ^{36}Cl . Beryllium-10 was analyzed by three laboratories and the other two nuclides by only one.

2.7.1.3 Isle of Skye and Western Highlands of Scotland

By the Bølling-Allerød interstade, the full-glacial ice sheet covering northern Scotland had shrank to virtually nothing (Gheorghiu et al., 2012). During the following Younger Dryas stade, cirque glaciers readvanced and reoccupied corries in the Western Highlands and Isles, known in Scotland as the Loch Lomond readvance. Boulders

deposited as the Loch Lomond glaciers rapidly retreated at the end of the Younger Dryas are chronologically well constrained and suitable for TCN production-rate calibration. Details of sites sampled and results can be found in Marrero et al. (2015d). The Loch Lomond readvance has been dated by a combination of direct radiocarbon measurements on peat in cores that either rest on Loch Lomond till or contain distal Loch Lomond outwash, and by means of tephrocorrelation with Icelandic tephra found in these cores, especially the Vedde ash, which is found in both the local cores and in Greenland ice cores. These enable the readvance to be correlated with the NGRIP ice core, constraining the deposition of the boulders sampled to 11.7 ± 0.3 ka before 2010 CE. Recent radiocarbon measurements by Bromley et al. (2014) have supported an earlier age for deglaciation; these are discussed in Marrero et al. (2015d).

The latitude of the sites ranged from 57.18° to 57.24° N and the elevation from 146 to 702 m. A variety of lithologies were sampled, including granite, Torridon sandstone, gabbro, and gneiss. Boulders were sampled at three sites on the Isle of Skye and analyzed for the following nuclides: Coire Fearchair (7^{10}Be , 2^{26}Al , 4^{36}Cl), Kyleakin Pass (2^{10}Be), Coire a Ghrunnda (6^{36}Cl). On the adjacent mainland of Scotland three additional sites were sampled: Corrie nan arr (4^{10}Be , 4^{26}Al , 3^{14}C), and Maol Chean-Dearg, (6^{10}Be , 6^{26}Al , 7^{14}C). The Scotland sites were particularly valuable because they allowed sampling for multiple nuclides and because they present a control on production at nearly the same age as the Huancané site, but at low elevation and high latitude.

2.7.1.4 Macaulay valley, South Island, New Zealand

The Macaulay valley is on the west side of the South Island of New Zealand. The Macaulay River drains into Lake Tekapo. About two-thirds of the distance to the headwaters is an extensive, relatively planar boulder field resulting from a large debris flow that came to rest on top of an organic-rich horizon containing well-preserved wood from shrubs. Ten radiocarbon measurements on this wood have provided an age of 9.69 ± 0.05 ka before 2010 CE on the depositional event. Seven large greywacke boulders were sampled for calibration of ^{10}Be and ^{14}C production rates, as detailed by Putnam et al. (2010). These boulders were at 43.57° S and about 1025 m elevation. The ^{14}C production rate at the same site has been independently calibrated by Schimmelpfennig et al. (2012).

2.7.1.5 Additional calibration sites

The localities described above that were sampled as part of the CRONUS-Earth Project did not provide adequate calibration material for all of the nuclides. In particular, material for calibration of ^3He and ^{14}C was very limited. Results from non-CRONUS, previously published, studies were therefore “imported” to fill this need.

Helium-3 Sites. A reassessment of ^3He calibration sites was published by Goehring et al. (2010). Sites were selected from this compilation by moving less desirable sites (i.e., those that suffered from problems with possible erosion or poor independent ages) into the secondary data set. Specifically, sites from Iceland by Licciardi et al. (2006), from Argentina by Ackert et al. (2003), and from Oregon and

Idaho in the western United States from Cerling and Craig (1994) were added for ^3He production calibration.

Carbon-14 Sites. In addition to the ^{14}C analyses performed on the samples from the CRONUS-Earth calibration sites described above, a set of analyses performed by Nathaniel Lifton (Purdue University) during and prior to the CRONUS-Earth Project on low-erosion samples that had been exposed for long enough to reach saturation were included in the ^{14}C calibration data set. These have been summarized in Borchers et al. (2015) and will be described in more detail in a subsequent publication. They were obtained from elevation transects in the western United States (White Mountains and Death Valley, both in California, USA), and the Andes in northern Chile.

Muogenic-production calibration. In addition to the major spallation reactions (and in the case of ^{36}Cl , low-energy neutron-absorption production reactions as well), many nuclides are also produced by direct reactions with high and low energy secondary cosmic-ray muons and by reactions with other particles released by muon reactions (Charalambus, 1971; Stone et al., 1998; Heisinger et al., 2002a; Heisinger et al., 2002b; Braucher et al., 2011). To calibrate the muogenic production of ^{10}Be and ^{26}Al , these nuclides were measured in samples from a 27-m deep vertical core drilled in sandstone at Beacon Heights, in the McMurdo Dry Valleys, Antarctica, drilled and processed by John O. Stone, University of Washington. The site was chosen for the low surface erosion rate ($< 10 \text{ cm/Myr}$), which results in high nuclide concentrations and makes the calibration insensitive to details of the erosion history. Three laboratories, the University of Washington (Seattle, USA), Centre Européen de Recherche et d'Enseignement de Géosciences de l'Environnement (CEREGE; Aix-en-Provence, France), and Dalhousie University (Halifax, Canada) measured samples from the core. The results of these measurements are shown in Figure 1. The data from the University of Washington were used as primary data and the less comprehensive (in terms of number of samples) data from CEREGE and Dalhousie were used for the secondary data set. The data are available in the data supplement to Borchers et al. (2015).

Chlorine-36 is also produced in significant amounts by muon reactions with calcium and potassium (Stone et al., 1998). To calibrate the muogenic ^{36}Cl production parameters, a deep core in Antarctica in dolerite was attempted. However, technical problems with the drilling operations prevented completion of this drill hole and so pre-CRONUS depth profiles from Australia in limestone (Stone et al., 1998) and granite (Evans, 2001) were used for this purpose.

Sites published subsequent to the CRONUS-Earth Project. Sites to be included in the primary and secondary CRONUS data sets went through a final selection process by vote at a face-to-face meeting open to all Project members in the summer of 2010. To avoid possible ad-hoc incorporation of additional data sets by a minority of the Project participants, and because the processes of formatting the large amount of raw data for

Figure 1. Concentrations of ^{10}Be (top) and ^{26}Al (bottom) in the Beacon Heights sandstone core, Antarctica, as a function of shielding depth. Data are from three laboratories, as described in the text. Red, black, and blue lines in the left-hand panels represent fits to, respectively, spallation, slow-muon, and fast-muon production reactions. Figure was constructed by John O. Stone, University of Washington, USA.

the nuclide samples, quality control, and inverse modeling for calibration were time consuming, the selection process was closed at that point. Since the summer of 2010, several high-quality calibration studies have been published (Amidon, 2011; Blard et al., 2013a; Blard et al., 2013b; Fenton et al., 2011; Fenton et al., 2013; Foeken et al., 2012; Schimmelpfennig et al., 2011a; Schimmelpfennig et al., 2011b), but could not be incorporated for these reasons. The procedural decision was made before these data were available and does not constitute any reflection on their utility.

2.7.2 Secondary Data Sites

As a general principle of scientific investigation, it is desirable to be able to calibrate a model using one set of data and then to test the model using an independent data set; the secondary data sets filled this testing role. These generally consisted of previously published studies that were assessed by the members of the CRONUS-Earth Project to contain reliable data, but often were not quite as secure (particularly with regard to independent chronological constraints) as the primary calibration data. Other

factors that were considered were the extent of complications due to snow cover, vegetation, landscape erosion, and so on. In addition to previously published sites, four sites were sampled as part of the CRONUS-Earth Project and used as secondary data. These were the Littleton-Bethlehem moraine in New England, USA, glacial erratic boulders on islands in Puget Sound, a Recess Peak-age moraine in the Sierra Nevada, and landslide deposits from Norway. These will be described first, followed by secondary data set grouped by nuclide. Additional information on the samples used is given by Borchers et al. (2015) and full descriptions of the sites are provided in the original sources cited below.

2.7.2.1 Littleton-Bethlehem moraine complex, New England, USA

The Littleton-Bethlehem moraine complex was deposited during a pause or small readvance of the Laurentide Ice Sheet during its retreat across New England after the Last Glacial Maximum. During retreat, a series of proglacial lakes were formed that were subsequently partially filled with varved sediment. The counting of these varves provided one of the earliest quantitative indications of the Quaternary time scale in North America (Antevs, 1922). More recently, the varve chronology has been supplemented with numerous radiocarbon dates (Ridge et al., 2012). This combination provides a very secure framework within which to place the deposition of particular features related to glacial retreat. Balco et al. (2009) have assigned an age of 13.84 ± 0.25 cal ka to the samples from the Littleton-Bethlehem moraine and described the sites and samples in detail.

The moraine complex was sampled along the Connecticut River Valley between 43.0° and 43.5° N and elevations between 130 and 300 m. The boulders are gneiss and granite. Quartz separates were analyzed by two laboratories for ^{10}Be and ^{26}Al and potassium-feldspar separates for ^{36}Cl by one laboratory. The region is humid and heavily forested, so the data were placed in the secondary, rather than primary, data set due to boulder-surface coverage by moss and soil and shielding from cosmic radiation by forest biomass.

2.7.2.2 Puget Sound, Northwestern USA

Puget Sound was exposed at the end of the last glaciation by retreat of the Cordilleran Ice Sheet. The islands in the Sound were initially below sea level but were rapidly raised by isostatic rebound. Radiocarbon dating of both marine and terrestrial materials on either side of the transition closely constrains the initial age of exposure to 15.0 ± 0.5 cal ka (Marrero et al., 2015b; Swanson and Caffee, 2001). Chlorine-36 measurements on erratics of a variety of lithologies were used to estimate production rates by Swanson and Caffee (2001). However, the production rates thus obtained were markedly larger than those for other studies. To better understand the origins of this discrepancy, CRONUS-Earth members revisited the area and resampled the same, or similar, outcrops. Twelve samples from four localities were analyzed for ^{36}Cl by two laboratories. Additionally, three samples from one locality were analyzed for ^{10}Be in order to check for any nuclide-specific anomalies. The sampling has been described and the results are in publication by Marrero et al. (2015b).

2.7.2.3 Baboon Lakes moraine, Sierra Nevada, California, USA

As described in Phillips et al. (2015b), numerous cosmogenic studies have previously been conducted in the Sierra Nevada of California. Several of these have been included in the secondary data set. Only three new samples were collected as part of the CRONUS-Earth Project. These were from a Recess Peak-age moraine at the Baboon Lakes, in the Bishop Creek drainage in the Eastern Sierra. The glacial advance at this site was dated by radiocarbon using macrofossils from a lacustrine core within the moraine limits (Clark and Gillespie, 1997). Three granodiorite boulders were sampled and analyzed for ^{36}Cl in several mineral fractions and for ^{10}Be (Phillips et al., 2015b). These were all used as part of the various secondary data sets, except for the ^{36}Cl measurement in the biotite fraction, which was used to help constrain the rate of ^{36}Cl production from low-energy neutron absorption by Cl (Marrero et al., 2015c).

2.7.2.4 ^{10}Be secondary data set

The ^{10}Be secondary data set contains 14 samples from the Littleton-Bethlehem moraine complex as well as the three samples each from the Puget Sound and Baboon Lakes sites that have been described above. In addition, it contains 10 samples from a full-glacial moraine in the New Zealand Alps (Putnam et al., 2010) and 10 samples from retreat of the Tioga 4 (latest glacial maximum) glacier in the Sierra Nevada, California (Nishiizumi et al., 1989; Phillips et al., 2015b). Eight samples from a Younger Dryas moraine in western Norway and 10 from a mid-Holocene landslide in the same area, sampled as part of the CRONUS-Earth Project, are included (Goehring et al., 2012a, b). Finally, there are two sets of sets of samples from the Andes, eight from the Puerto Bandera moraines in Argentina (Kaplan et al., 2011) and seven from the Breque moraines (Quebrada Rurec) in Peru (Farber et al., 2005). The entire data set contains 70 samples from 10 sites.

A dataset from Titcomb Lakes in the Wind River Range, Wyoming, USA, from Gosse et al. (1995), was originally included in the secondary data. However, subsequent efforts to determine the method of standardization used for the AMS analyses of ^{10}Be have not yet yielded conclusive results. Provided the standardization can be established, this data set should be included in subsequent calibration efforts.

2.7.2.5 ^{26}Al secondary data set

The ^{26}Al secondary data set is the same as the ^{10}Be secondary one, except that it does not contain samples from the Breque moraines in Peru, the New Zealand moraine, the Norway landslide, and the Puerto Bandera moraines in Argentina. This data set is fairly small, with 30 samples from four sites.

2.7.2.6 ^{36}Cl secondary data set

The ^{36}Cl secondary data set contains a wide range of samples. They include numerous samples from the Sierra Nevada, whose setting has been re-evaluated by Phillips (2015) and Phillips et al. (2015b). These include the new data from the Baboon Lakes moraine, described above, potassium-feldspar and quartz separates from late-glacial moraine boulders by Evans et al. (1997), and whole-rock samples from late-glacial moraines (Phillips et al., 2009). Potassium-feldspar separates from the Littleton-Bethlehem moraine complex are included (Marrero et al., 2015c), as are whole-rock data from the Puget Sound study described above (Marrero et al., 2015b). Quartz separates from Younger Dryas age moraines in Scotland by Evans (2001) are part of this data set. Finally, the original calibration data set of Phillips et al. (2001), which included both lava-flow and moraine samples from a wide range of locations, was incorporated. At 85 samples from 20 sites, ^{36}Cl has the largest secondary data set of any nuclide. The selection of the ^{36}Cl secondary data set is further discussed in Appendix B.

2.7.2.7 ^3He secondary data set

The ^3He secondary data were taken from the compilation in Goehring et al. (2010). The rocks sampled were all lava flows. A substantial number are from sites in the western United States: California (Cerling and Craig, 1994), Idaho (Poreda and Cerling, 1992), Oregon (Cerling and Craig, 1994; Licciardi et al., 1999), and Hawaii (Blard et al., 2006). Additional sites are located in the Canary Islands (Dunai and Wijbrans, 2000) and Sicily (Blard et al., 2006). Altogether there are 40 samples from 13 localities.

2.7.2.8 ^{14}C secondary data set

Analyses of ^{14}C are as yet very limited in comparison to the other nuclides. All suitable ^{14}C data were used in the primary calibration data set; there is no secondary ^{14}C data set.

2.7.3 Calibration methodology

A detailed description of the calibration methodology is contained in Borchers et al. (2015). Briefly, the nuclide production rates were calibrated by calculating nuclide concentrations for all samples in the primary data set for that nuclide, using the independently determined exposure age for the sample and the appropriate scaling relations, implemented by CRONUScalc (Marrero et al., 2015a). The independent age was treated as uncertain within the bounds of the assigned age uncertainty, along with the nuclide concentration. The measured nuclide concentration for each sample was then subtracted from the calculated concentration. The best-fit solution was identified by orthogonal-distance regression (ODR) in which both the concentration misfit and the age misfit were minimized. The problem was formulated as a nonlinear least-squares minimization and numerically solved using the Levenberg-Marquardt method (Aster et al., 2012). It is important to note that because the ODR approach allows both the calculated sample concentration and the age to vary within the stated statistical bounds,

the calculated concentrations are not required to be unbiased (i.e., positive and negative deviations from the measured value do not have to sum to zero). Furthermore, sites with very small age bounds may strongly affect the distribution of calculated ages, since sites with wider age bounds may be forced into large age deviations in order to compensate for the relatively fixed ages of the small-bound sites.

The muogenic production parameters were calibrated by means of an iterative approach in which the calibration of the spallation production rate was alternated with calibration of the muogenic production parameters by fitting to the Beacon Heights core until the two converged. For chlorine, an additional iterative step to calibrate the low-energy neutron pathway was also included, in addition to the two muogenic production calibration steps (Marrero et al., 2015c).

Although the mathematical procedure minimizes the misfit of the predicted to the measured concentrations, residuals remain. If the nuclide-production systematics embodied in the model are adequate, and the data put into the model (e.g., sample elevations, topographic shielding data, geomagnetic history) are accurate, the predicted and measured concentrations should agree within statistically determined bounds. Two statistics were used to assess the goodness of fit. The first was the reduced-chi-squared, essentially a summation of the squared error divided by the sample variance, all divided by the degrees of freedom for the particular minimization. In general, if the spread of the predictions (relative to the measurements) is consistent with the uncertainties assigned to the parameters, the reduced chi squared should be close to unity. The p-value associated with this statistic is the probability of obtaining a misfit as large as the calculated one if the model, the parameter values in the model, and the uncertainties assigned to the parameters are all correct. p-values larger than 5% are generally considered to indicate that the model provides statistically valid predictions (Aster et al., 2012).

In general, the single most important parameter uncertainty is the nuclide-concentration analytical uncertainty. As discussed at length above, the various CRONUS analytical intercomparison efforts generally showed that the statistics of aggregates of nuclide-concentration data from numerous laboratories were ‘overdispersed’ (i.e., the spread of the aggregate data was larger than would be predicted based on the reported analytical standard deviations). The calibration effort (Borchers et al., 2015) took advantage of the CRONUS intercalibration initiatives to assign more realistic analytical uncertainties to the measurements that went into the calibration, based on the results of the intercomparisons. For each sample the calibration assigned the larger of the following two uncertainty estimates: (1) the laboratory-reported sample analytical uncertainty, or (2) the experimental coefficient of variation obtained from all of the samples analyzed for the appropriate laboratory intercomparison study (Section 2.1). The nuclide-specific details will be discussed below.

Provided that a statistically justified model (in the sense of combined model physics, parameters, and parameter uncertainties) is obtained, the uncertainty of the calibrated model parameters can then be calculated from the goodness of fit of the model to the data (Aster et al., 2012). However, the outcome of the goodness-of-fit testing was that few nuclide calibrations yielded p-values greater than 5%, and the statistical validity of most of those that did was compromised by inadequate numbers of calibration sites (Borchers et al., 2015). The implication is that the specification of production-rate uncertainties based on the results of the calibration of the primary data sets is not

statistically justifiable. Production-parameter uncertainties are a necessity for the application of terrestrial cosmogenic nuclides to geochronology. This problem was addressed by performing an empirical estimation of production-parameter uncertainties based on data sets external to the calibration (i.e., the secondary data sets).

With an adequate number of independent data, the predictive capability of the nuclide-production models can be assessed without reference to the primary calibration data sets. This was accomplished using the secondary data sets. Unlike the primary data set used for the calibrations, the secondary data were evaluated in terms of calculated ages rather than concentrations, largely because we wished to evaluate uncertainties in the calculated ages. First, the calculated ages were normalized by dividing by the independently established ages of the samples. Samples with calculated ages in perfect agreement with the independent age would thus have a normalized value of unity; numerals to the right of the decimal point represent the fractional error. The root mean squared error (RMSE) was then calculated for the entire secondary data set (i.e., based on the difference between unity and the normalized sample age). Next, the ages were recomputed with increasing values of the uncertainty attributed to the production-rate parameter (the value for the production rate itself was left unchanged). The increased production-rate uncertainty propagated into the calculated age as increased age uncertainty. The production-rate uncertainty was increased until the mean of the normalized computed age uncertainties equaled the RMSE of the ages in data set. Given an adequate number of sites and samples in the secondary data sets, the use of the production-parameter uncertainty calculated in this way should provide conservative and realistic “external” uncertainties for computed ages.

We feel that the production-parameter uncertainties we have calculated are likely to be conservative, meaning that they are more likely to overestimate the actual uncertainty than to underestimate it, for two reasons: (1) Most new data were incorporated into the primary calibration data sets, leaving the secondary data sets largely populated with older studies from the literature. Given improvements in analytical methodology with time, current studies are likely to give more accurate and reproducible results than the older studies used to estimate the uncertainty. (2) Only studies with the highest degree of certainty in the independent age and geological history were included in the primary calibration data sets. Studies that the CRONUS-Earth participants thought might have independent age uncertainties in excess of the stated limits, or questions regarding issues such as sample coverage or erosion, were placed in the secondary data set. Furthermore, target preparation and MS methods have continued to improve, so it is possible that earlier results may have uncharacterized systematic errors. Any errors in the exposure history of samples in the secondary data set resulting from these factors would then increase the calculated production-parameter uncertainty.

In the following section we will discuss in detail the calibration results and uncertainties for ^{10}Be , inasmuch as it had the best global coverage and the most accurate and precise analytical data, as well as being the most widely used. We will attempt to identify sources of variability or error, in the calibration of this nuclide. Similar, though less detailed, discussions for ^{26}Al , ^{36}Cl , ^3He , and ^{14}C are presented in Appendix B. These nuclides will be discussed below more succinctly. Readers can refer to Appendix B and calibration papers for those nuclides for further discussion.

2.7.4 Calibration Results

2.7.4.1 ^{10}Be calibration results

As discussed in section 2.6.1, the ^{10}Be calibration data set consisted of samples from four localities. The primary data set contained 107 samples distributed among the four sites. Complete data are available in spreadsheet form in the appendix of Borchers et al. (2015). The analytical uncertainty used in the calibration was in general derived from the interlaboratory comparison study of Jull et al. (2015) rather than the laboratory-reported uncertainty. The analytical uncertainty was interpolated, based on the measured ^{10}Be concentration, between 3.6% at 2.1×10^5 atoms g^{-1} and 2.3% at 347×10^5 atoms g^{-1} , unless the reported uncertainty was greater, in which case it was used. The average assigned uncertainty was 4.2%. The results are shown graphically in Figure 2 and a statistical summary of the results is given in Table 3 of this paper. To be consistent with Borchers et al. (2015), the results are presented as normalized ^{10}Be concentrations (measured divided by calculated concentrations).

The secondary data set for ^{10}Be is described in section 2.6.2 and consists of 10 globally distributed sites with a total of 89 samples. Complete data are available in spreadsheet form in the appendix of Borchers et al. (2015). The results are shown graphically in Figure 3 and a statistical summary of the results is given in Table 3. The secondary data set results are presented in terms of normalized ages (calculated ^{10}Be age divided by independent age), but the relative sense of displacement is the same (i.e., for a

Scaling Method	Average Analytic Unc	Lifton-Sato Nucl-Dep (LSDn or Sa)	Lifton-Sato Flux-Based (LSD or Sf)	Lal-Stone (St)	Lal 1991 Nish 1989 (Lm)	Lifton (Li)	Dunai (Du)	Desilets (De)
Shorthand Notation								
Production Rate (atoms $\text{g}^{-1} \text{yr}^{-1}$)		3.92	4.09	4.01	4.00	4.06	3.70	3.69
Production Rate Uncertainty		± 0.31	± 0.35	± 0.33	± 0.32	± 0.57	± 0.58	± 0.58
Reduced Chi-Squared		1.64	1.54	1.53	1.49	3.69	4.40	4.59
p-value		5.1×10^{-5}	4.1×10^{-4}	4.6×10^{-5}	1.0×10^{-3}	1.0×10^{-14}	1.0×10^{-14}	1.0×10^{-14}
Root Mean Squared Error (by sample; %)		7.9	7.0	7.1	6.9	15.8	18.0	18.2
Average Absolute Global Error (%)		6.5	5.4	5.7	5.5	14.7	17.2	17.2
Average Global Norm Conc (by sample)		1.043	1.035	1.038	1.038	1.090	1.105	1.106
Average Norm Concentration (by site)		1.025	1.017	1.024	1.022	1.073	1.087	1.090
St Dev Global Norm Concentration	0.042	0.067	0.061	0.060	0.057	0.130	0.146	0.148
Averaged Normalized Concentration								
NZ Average		0.977	0.964	0.980	0.971	1.057	1.066	1.082
PPT Average		1.090	1.086	1.063	1.056	1.168	1.184	1.192
SKY Average		1.066	1.031	1.072	1.074	1.180	1.221	1.215
HKY Average		1.062	1.028	1.071	1.073	1.176	1.217	1.212
SCOT Average		1.074	1.056	1.063	1.060	1.167	1.192	1.195
Peru Average		0.972	0.991	0.980	0.989	<i>0.892</i>	<i>0.880</i>	<i>0.877</i>
Standard Deviations of Normalized Conc								
NZ St Dev	0.037	0.015	0.014	0.015	0.014	0.016	0.016	0.016
PPT St Dev	0.039	0.042	0.041	0.040	0.040	0.044	0.045	0.045
SKY St Dev	0.075	0.053	0.052	0.053	0.053	0.059	0.061	0.061
HKY St Dev	0.041	0.042	0.038	0.046	0.046	0.046	0.046	0.046
SCOT St Dev	0.041	0.030	0.029	0.031	0.031	0.034	0.035	0.035
Peru St Dev	0.033	0.049	0.050	0.049	0.050	0.045	0.044	0.044
Mean	0.044	0.038	0.037	0.039	0.039	0.041	0.041	0.041

Table 3. Primary ^{10}Be data set: production rates, uncertainties, and statistical evaluation. The SLHL production rates, reduced chi-squared, and p-values are from Borchers et al. (2015), other values from this paper. Values are in terms of normalized ^{10}Be concentration and are unitless. Site-average normalized values more than 7% above the average concentration are bolded and ones more than 7% below the concentration are italicized. References for the scaling methods employed are in the text.

site at which the production rate is underestimated, both the normalized concentration and the normalized age will plot above the line representing unity) and the magnitude of the displacement is similar. The references for the scaling methods are as follows: LSD and LSDn, Lifton et al. (2014); St, Lal (1991) with atmospheric depth from Stone (2000); Lm, Lal (1991) with geomagnetic scaling based on Nishiizumi et al. (1989); Li, Lifton et al. (2005); Du, Dunai (2001); and De, Desilets et al. (2006). See Table 4 and associated text in Balco et al. (2008) for additional detail on the differences among these models.

Figure 4 shows the distribution of the primary and secondary ^{10}Be sample sites with respect to latitude and altitude. The range of likely latitudes and altitudes is covered well, but sites are lacking at high latitude combined with high altitude and at low latitude combined with low altitude. These should constitute global regions to target in future calibration studies. Figures 5 and 6 illustrate the dependence of the normalized ^{10}Be age on latitude, elevation, and independent age.

Evaluation of the primary calibration data set: Table 3 contains two notable results. The first is that the neutron-monitor-based scaling methods (Li, Du, De) yield worse fits to the data than do the alternative methods. Whereas the reduced chi-squared values for the neutron-monitor-based methods average about 4, those for the other methods are about 1.5. The p-values are correspondingly much lower for the neutron-monitor-based methods. The RMSE's of the neutron-monitor methods are in the range of 16 to 18%, whereas those of the other methods are about 7%. Notably, the average normalized concentrations (i.e., bias) of the neutron-monitor-based methods are about 10% high, while the other methods are generally only 2 to 4% high.

Figure 2(b) is informative as to the cause of the poor fits. It shows that the normalized concentrations, using Du scaling, for the low-to-moderate elevation sites (those on the left-hand side of the diagram) are uniformly excessive, but those for Huancané, the highest elevation site, are uniformly low. A similar pattern is also observed for Du scaling in Figure 6(b), which combines primary and secondary data. Sites above about 3,500 m tend to plot low and those below that elevation plot high. Similar patterns are observed when the results from the other neutron-monitor-based methods are plotted. This pattern is an outcome of an excessive range of scaling for these methods. The scaling methods produce a difference in scaling between

Figure 2. Normalized ^{10}Be concentrations (measured concentration divided by calculated concentration) for the primary calibration data set. Red lines separate sample sites; dotted black lines individual samples at each site. (a) Comparison of the classic Lal-Stone method (St) with the Lifton-Sato nuclide-dependent scaling (LSDn or Sa). (b) Comparison of the Lal-Stone method (St) with the neutron-monitor based Dunai scaling (Du). Note that the uncertainties associated with the calculated concentrations are not fully propagated uncertainties, which were not computed, but are simply proportional to the measured concentration uncertainties.

1151

1152
1153

1154
1155
1156
1157
1158
1159

Figure 3. Normalized ^{10}Be ages (^{10}Be age divided by independent age) for the secondary data set. Blue lines separate sample sites. (a) Comparison of the classic Lal-Stone method (St) with the Lifton-Sato nuclide-dependent scaling (LSDn or Sa). (b) Comparison of the Lal-Stone method (St) with the neutron-monitor-based Dunai scaling (Du). Note that the uncertainties associated with the calculated concentrations are not

1160 fully propagated uncertainties, which were not computed, but are simply proportional to
 1161 the measured concentration uncertainties.
 1162

1163
 1164 Figure 4. Latitude and altitude distribution of the ^{10}Be primary and secondary data sets.
 1165

	Average Analytic Unc	Lifton-Sato Nucl-Dep	Lifton-Sato Flux-Based	Lal-Stone	Lal 1991 Nish 1989	Lifton	Dunai	Desilets
Production Rate (atoms g ⁻¹ y ⁻¹)		3.92	4.09	4.01	4.00	4.06	3.70	3.69
Root Mean Squared Error (by sample; %)		11.1	10.7	11.0	10.5	16.7	18.7	18.7
Root Mean Squared Error (by site; %)		8.0	7.4	8.0	7.6	13.7	15.9	15.7
Average Absolute Global Error (%)		7.9	7.4	7.7	7.4	14.1	16.1	16.0
Average Absolute Error by Site (%)		8.0	7.5	8.1	7.8	15.1	17.7	17.3
Average Global Norm Age (by sample)		1.072	1.054	1.060	1.057	1.138	1.159	1.158
Average Global Norm Age (by site)		1.062	1.045	1.054	1.052	1.124	1.146	1.143
St Dev Global Norm Age (by sample)	0.043	0.085	0.091	0.092	0.087	0.092	0.095	0.097
<i>Averaged Normalized Ages</i>								
New Zealand		1.010	0.993	0.994	0.983	1.087	1.099	<i>1.109</i>
Sierra Nevada		<i>1.118</i>	<i>1.130</i>	<i>1.128</i>	<i>1.116</i>	<i>1.152</i>	<i>1.168</i>	<i>1.177</i>
New England		<i>1.116</i>	1.089	1.048	1.046	<i>1.199</i>	<i>1.220</i>	<i>1.215</i>
Norway moraine		1.046	1.005	1.083	1.085	<i>1.172</i>	<i>1.201</i>	<i>1.204</i>
Norway landslide		1.055	1.014	1.087	1.089	<i>1.177</i>	<i>1.210</i>	<i>1.215</i>
Baboon Lakes		0.972	0.985	0.966	0.969	0.984	1.010	1.005
Puget Lowland		1.067	1.033	1.030	1.035	<i>1.149</i>	<i>1.183</i>	<i>1.168</i>
Puerto Bandera		1.025	0.994	0.988	0.998	<i>1.108</i>	<i>1.142</i>	<i>1.133</i>
Breque		<i>1.148</i>	<i>1.165</i>	<i>1.165</i>	<i>1.147</i>	1.088	1.085	1.062
<i>Standard Deviations</i>								
New Zealand	0.029	0.012	0.012	0.012	0.011	0.012	0.013	0.013
Sierra Nevada	0.158	<i>0.104</i>	<i>0.102</i>	<i>0.105</i>	<i>0.097</i>	<i>0.119</i>	<i>0.116</i>	<i>0.120</i>
New England	0.023	0.065	0.063	0.063	0.061	0.070	0.071	0.071
Norway moraine	0.023	0.046	0.044	0.046	0.046	0.051	0.052	0.053
Norway landslide	0.026	<i>0.096</i>	<i>0.093</i>	<i>0.102</i>	<i>0.103</i>	<i>0.110</i>	<i>0.113</i>	<i>0.114</i>
Baboon Lakes	0.013	0.047	0.048	0.047	0.047	0.048	0.047	0.049
Puget Lowland	0.023	0.073	0.071	0.071	0.070	0.077	0.079	0.079
Puerto Bandera	0.022	0.029	0.028	0.028	0.028	0.031	0.032	0.032
Breque	0.039	0.075	0.076	0.083	0.074	0.072	0.075	0.072
Mean	0.040	0.061	0.060	0.062	0.060	0.065	0.067	0.067

Table 4. Statistical summary for the secondary ¹⁰Be data set. Values are in terms of normalized age and are unitless. In the top box the smallest values in each column are bolded. For the site-average normalized ages, values greater than 1.10 are italicized and ones less than 0.90 are bolded. Site standard deviations significantly larger than the site mean are italicized and those less are bolded.

sea level and 5,000 m that is too large to be consistent with the data from these elevations in the primary data set and the optimization program “splits the difference”, yielding overscaled production rates (and hence low calculated concentrations) at high elevation and underscaled ones at low. This finding is consistent with that of Balco et al. (2008). Lifton et al. (2014) have attributed this shortcoming of the neutron-monitor-based methods to inadequate compensation for the effects of ‘neutron multiplicity’ (the tendency to generate multiple counts for a single neutron impact event), which increases with increasing elevation. These findings strongly argue that the four alternative scaling methods are preferred to the neutron-monitor-based ones. The discussion below therefore will be confined to the four non-neutron-monitor-based methods (i.e., LSDn, LSD, St, and Lm).

The second notable result from Table 3 is that none of the scaling methods produced a statistically adequate fit of the calculated concentrations to the data (i.e., reduced chi-squared values close to unity and p-values greater than 0.05). The closest was the Lal (1991) scaling, with reduced chi-squared of 1.49 and p-value of 1×10^{-3} . This result is certainly not unique: no previous calibration efforts have yielded statistically consistent results either (e.g., Balco et al., 2008). However, it should be noted that the degree of misfit is not large. For example, uniformly increasing the assigned analytical uncertainties by 25% (i.e., increasing the average uncertainty from 4.2% to 5.2%) would bring the reduced chi-squared down to unity for the four non-

neutron-monitor-based scaling methods. The differences between the goodness-of-fit parameters for these

Figure 5. Normalized ^{10}Be age, for the Lifton-Sato nuclide-dependent scaling (LSDn or Sa), for primary and secondary data sets as a function of (a) latitude, (b) elevation, and (c) independent age. Primary samples are filled circles and secondary are empty.

Figure 6. Normalized ^{10}Be age, for the Dunai scaling (Du), for primary and secondary data sets as a function of (a) latitude, (b) elevation, and (c) independent age. Primary samples are filled circles and secondary are empty.

four scaling methods are not large enough to differentiate them in terms of merit. We discuss below, in detail, the possible causes of the misfit.

We have characterized the degree of spread of the calculated concentrations relative to the independent age in two ways: using the RMSE and the average absolute global error (i.e., the average of the absolute values of the difference between unity and the normalized ages, computed for all samples). The RMSE gives more weight to outliers and thus is always slightly larger than the average absolute error. We have calculated the RMSE weighting all samples equally. This differs from the RMSE's reported in Borchers et al. (2015), which were calculated using the median concentrations for each site. We have used the sample-based RMSE because for some nuclides there are an insufficient number of sites to calculate a meaningful site-based RMSE and we wish to be consistent in the use of the statistic. The average RMSE for the four methods was

7.2% and the average absolute error was 5.8%. These values are approaching the stated goal of the CRONUS-Earth Project of cosmogenic age accuracy and precision “at the 5% level”.

The average normalized concentration is a measure of the bias of the data set. The average values for the four scaling methods range between 1.022 and 1.038, depending on whether site or global averages are used. As illustrated in Figure 2, this fairly small bias toward high concentrations arises from the data from Promontory Point and Scotland; the New Zealand and Huancané data exhibit little bias. The contribution of overall bias to the total spread of the data is significant but not predominant, as evidenced by comparison with the average standard deviation of the ages: 0.061. We note that the age uncertainty for the New Zealand site (0.5%) is very small compared to the uncertainties on the other sites (~1 to 2.5%). As described above, this will force large adjustments in the age and concentrations of the larger-uncertainty sites. This disparity almost certainly played a role in producing the observed bias in the concentrations.

Muogenic production calibration: As described above, the muogenic production calibration is interdependent with the spallation production calibration, and the two were thus solved iteratively for each scaling method. The nuclide profiles were fit as the sum of production by stopped negative muon capture and fast muon interactions, using the mathematical description of these reactions given by Heisinger *et al.* (2002a,b). Further details on the formulation of the muogenic production for this calibration are available in Marrero *et al.* (2015a). Results for both ^{10}Be and ^{26}Al are given in Table 5.

It is important to emphasize that the muon production parameters given in Table 5 are not directly applicable within the formulations presented by Heisinger *et al.* (2002a,b). They were calibrated using the newer muon-flux equations of Lifton *et al.* (2014), and for specified values of unity for the parameters α and $\beta(Z)$, which differ from the values used by Heisinger *et al.* (2002a,b). See Marrero *et al.* (2015a) for further details. The values vary slightly for the different scaling models, not because these models contain scaling for muon fluxes, but rather because muogenic and nucleogenic production rates were solved for simultaneously, and thus differences in the nucleogenic scaling also minimally affected the muogenic production parameters.

Unlike the case for spallogenic production, statistically acceptable fits to the data were obtained for all scaling methods (Table 5 and Figure 1), but with production constants significantly lower than the values obtained by Heisinger *et al.* from muon irradiation experiments. For ^{10}Be and ^{26}Al at sea level, this calibration predicts total surface production rates by muon reactions that are 47% and 46% of the values given by Heisinger *et al.*, while confirming earlier depth-profile results (Braucher *et al.*, 2003, 2011). There is no obvious explanation for the discrepancy between the laboratory-based and natural calibrations.

The result that the muogenic production calibration yielded statistically acceptable fits whereas the spallation-production one did not is a result of several factors. One is that data were fit only at a single location and thus inadequacies in spatial scaling did not affect the fit. However, it should be noted that the magnitude of the scaling required for muogenic production is much less than that for production from spallation reactions. The two production parameters governing, respectively, slow and fast muon production, f^* and σ_o , were strongly correlated, indicating that the drill hole did not reach

sufficient depth to uniquely identify the depth dependence of the fast-muon production (Borchers et al., 2015).

Scaling Method		Lifton-Sato-Dunai Nucl-Dep (LSDn or Sa)	Lifton-Sato-Dunai Flux-Based (LSD or Sf)	Lal-Stone (St)	Lal 1991 Nish 1989 (Lm)	Lifton (Li)	Dunai (Du)	Desilets (De)
Shorthand Notation								
¹⁰ Be	f* (x 10 ⁻³)	1.89±0.18	1.89±0.18	1.87±0.18	1.87±0.18	1.86±0.18	1.85±0.18	1.84±0.17
	σ _s (x 10 ⁻³⁰ cm ²)	0.252±0.015	0.252±0.015	0.251±0.015	0.251±0.015	0.251±0.015	0.251±0.014	0.251±0.014
²⁶ Al	f* (x 10 ⁻³)	12.1±2.6	11.7±2.6	10.9±2.6	11.0±2.6	11.1±2.6	10.9±2.6	10.9±2.6
	σ _s (x 10 ⁻³⁰ cm ²)	4.03±0.73	4.10±0.73	4.22±0.73	4.22±0.73	4.19±0.73	4.21±0.73	4.21±0.73
Reduced Chi-Squared		0.94	0.92	1.26	1.25	1.13	1.34	1.51
p-value		0.998	0.999	0.958	0.961	0.987	0.929	0.819

Table 5. Muon parameters for ¹⁰Be and ²⁶Al production, from supplement to Borchers et al. (2015). These parameter values are not directly applicable in the equations of Heisinger et al. (2002a,b) (see text).

Had it done so, any inadequacies in the theoretical formulation would have been more apparent in the goodness of fit. Finally, data from only one laboratory were used. The formulation described above in this section, derived from the Jull et al. (2015) intercomparison study and intended to incorporate interlaboratory variability, probably overestimated the actual variability of the data for this experiment.

Origins of misfits between the data and the scaled production models: It would be desirable to have a statistically justified outcome to the calibration process, meaning that the combination of data sets, parameters, assigned uncertainties, and computational model would produce calculated ¹⁰Be concentrations (or ages) in agreement with the measured concentrations at the p-value greater than 0.05. Here we examine the possible origins of the misfit for the spallation calibration, with a view toward improving the results of future calibrations.

The most important possible causes of the misfit include the following: (1) unrealistically low assignment of analytical or other uncertainties, (2) sample-specific variability such as variable erosion or coverage by sand, snow, or moss that has not been adequately accounted for, (3) systematic errors such as incorrect assignment of independent ages, and (4) inadequate accuracy of the computational model, in particular the scaling.

First we address both (1) uncertainties and (2) sample-specific variability. The role of these issues can be straightforwardly evaluated by comparison of assigned analytical uncertainties with the spread of data within each site. For the four scaling methods under consideration, the average of the standard deviations of the normalized concentrations (i.e., the average coefficient of variation) at each of the sites is 0.039. This value is smaller than the average of the analytical uncertainties for all of the sites, 0.044. (Note that these are the “inflated” empirically based uncertainties used in the calibration process, not the original laboratory analytical uncertainties; see Section 2.7.3). Although the ¹⁰Be analytical uncertainty is the largest component of the uncertainties propagated into the calculated concentration uncertainties, other parameters were also given uncertainties. This comparison thus not only indicates that the assigned analytical uncertainties were conservative, it also indicates that the influence of sample-specific variability must have been small. In summary, although analytical uncertainty is an

important contributor to the overall variability of the primary ^{10}Be data set, it does not appear that its magnitude was underestimated.

Next, we will consider (3) systematic errors. This is difficult to evaluate; obviously the best estimates of site age (the primary potential contributor) were used. The independent age for the Huancané and New Zealand sites were established with multiple radiocarbon ages on organic matter in direct proximity to the sampled features and must be considered very strong. For both Promontory Point and the Scotland sites the control ages were not in direct proximity. We note that a recent study (Bromley et al., 2014) has called into question the previously accepted chronology for the recession of the Loch Lomond stade glaciers in Scotland, increasing it from 11.7 to approximately 12.4 ka. This age increase of about 6% is approximately equal to the age overestimate for the Scotland samples by the non-neutron-monitor-based scaling methods (Table 3). However, re-evaluation of the suggested age did not support a revision of the ^{10}Be production-rate calibration because: (1) a careful examination of the new deglacial age in the context of data from Skye indicated that it was not applicable to the deglaciation of Skye and nearby areas (Marrero et al., 2015d), and (2) we performed a preliminary recalibration of the primary ^{10}Be data set using the 12.4 ka age and we did not find an improvement of the fit.

Finally, we consider inadequate accuracy of the computational model, particularly the scaling (Table 4). This is clearly significant for some of the sites. As is apparent from Table 4 and Figure 2(a), the average biases of the Promontory and Scotland sites are larger than the associated one-sigma uncertainties. This is not the case for New Zealand and Huancané. When we computed residual uncertainties (i.e., complete primary-dataset normalized-age standard deviations less mean propagated age uncertainties less independent-age uncertainties, all in quadrature) using the calculated ages (for which full uncertainties were propagated), they came out very close to the average bias of the primary data set ages. For example, for the Lm scaling, the age standard deviation for the primary data set was 0.071, the mean propagated age uncertainty was 0.045, and the independent age uncertainty was 0.012, yielding a residual uncertainty of 0.054. This is very close to the bias indicated by the average normalized age: 0.058.

Although in the current calibration the misfit appears to be localized at Promontory and Scotland, it is important to recognize that this is not necessarily due to some ‘fault’ of these sites. The goodness of fit is dependent on the structure of the scaling model, which includes the physics of modulation of the cosmic-ray flux by the geomagnetic field, the particular magnetic field reconstruction employed, and any assumptions regarding the atmospheric pressure history. Inadequacies or errors in these aspects of the model will cause apparent misfit even in the absence of inadequacies in the data from the particular sites.

In summary, it appears that the misfit in excess of statistical expectations can largely be attributed to the concentration (or age) biases at the Promontory and Scotland sites. Given the elaborate replication of analyses by multiple laboratories (especially at Promontory), these biases are almost certainly not an analytical artifact (noting again that the New Zealand site was analyzed by only one laboratory). The discrepancies seem unlikely to be due to error in the independent age, or other systematic parameters, such as time-averaged elevation. Thus we consider that the biases probably are “real” (i.e., a result of physical processes that either are not included in the computational model, or are

not included accurately). However, we reemphasize that the apparent misfit for these two sites could arise from a large number of possible sources, including general structure of the model or underlying databases.

Evaluation of the secondary data set: The single most important parameter in Table 4 is the RMSE by sample, inasmuch as it provides the basis for assigning the production-parameter uncertainty. The values of the RMSE were about 10% for all four of the non-neutron-monitor-based scaling methods. In comparison to the primary calibration data set, the RMSE-by-sample is significantly larger, but the RMSE-by-site is similar. This indicates that outliers are influencing the secondary data set more strongly than was the case for the primary data set, and this is confirmed by Figure 3(a), showing numerous outliers in the Sierra Nevada, New England, and Breque data sets. The other notable feature in Table 4 is the large values for the average normalized age, both by sample and site, averaging 1.061 and 1.053, respectively. The average of the individual site standard deviations (last row in Table 4), 0.061, shows that within-site variability and site bias are contributing about equally to the average absolute error of the data set.

The significant bias in the average normalized ages is troubling. Table 4 and Figure 3(a) show that the bias arises mostly from four sites: Sierra Nevada, New England, Norway, and Breque. (We note that the average age for the Holocene Norway landslide data is biased high due to one outlier; the other samples average close to a normalized age of 1.0). Figure 5 does not reveal any consistent pattern in the variation of normalized ages with latitude, elevation, or independent age. The cause of the discrepancy in the Sierra Nevada ages has been examined by Phillips et al. (2015b). They pointed out that the samples with anomalously large normalized ages are all ones from sites with significant snow cover. ^{10}Be concentrations from sites with and without snow cover were about the same. The reason for this result has not been established, but it appears to be a study-specific anomaly.

The New England and Norway sites have in common that they are both on the shores of the North Atlantic and were close to the margins of the full-glacial ice sheets. The average of the normalized age for the four scaling methods at these two sites is about 1.06. This is very similar to the average bias of the samples from western Scotland in the primary calibration data set, which are also from a similar setting. These similarities suggest the possibility of a regional anomaly at this time period, previously noted by Balco et al. (2009) and Goehring et al. (2012b). Balco et al. (2009) evaluated various potential causes of this anomaly, including a glacially related atmospheric-pressure origin or one due to glacially induced isostatic rebound, but were unable to explain the full magnitude of the regional anomaly. The new LSD and LSDn scaling methodologies have only slightly decreased the magnitude of this regional mismatch. A reexamination of the glacial isostatic rebound issue might be fruitful.

The cause of the bias at the remaining poorly fitting site, Breque, is equivocal. The site is similar in latitude, elevation, and age to the Huanané primary calibration site, but yielded significantly higher ^{10}Be concentrations. Possible causes of the discrepancy have been discussed by Kelly et al. (2015), but without any clear resolution. Identification and sampling of additional sites at similar latitude and elevation will provide the best avenue to resolving the discrepancy.

In summary, the RMSE's for the four non-neutron-monitor-based scaling methods are about 10%. This relatively large error cannot be attributed to any one factor. Principal contributors include significant spread of the data within sampling sites (~6% coefficient of variation) and significant bias of the calculated site ages (averaging about 5%). The presence of a number of outliers increases the RMSE well above the average absolute error. In future calibrations, the spread of the data within sample sites can presumably be reduced by means of more careful sampling and more rigorous analytical methods, as evidenced by the significantly smaller average within-site coefficient of variation for the generally newer primary data set (about 4%). Part of the site age bias appears to be due to study-specific concerns.

Calibration results for other nuclides are given in Appendix B. Production rates for all nuclides are given in Table 6, from Borchers et al. (2015) and Marrero et al. (2015c) (for ^{36}Cl), but the uncertainties are original contributions of this paper, based on the evaluations of the secondary data sets, as described above.

Scaling Method	Lifton-Sato Nucl-Dep (LSDn or Sa)	Lifton-Sato Flux-Based (LSD or Sf)	Lal-Stone (St)	Lal 1991 Nish 1989 (Lm)	Lifton (Li)	Dunai (Du)	Desilets (De)
Nuclide Production Rate							
^{26}Al [atoms (g qtz) $^{-1}$ yr $^{-1}$]	28.5 $\pm$ 3.1	28.6 $\pm$ 3.3	27.9 $\pm$ 2.8	27.9 $\pm$ 2.7	28.7 $\pm$ 4.6	27.3 $\pm$ 4.4	26.3 $\pm$ 4.3
^{36}Cl [atoms (g Ca) $^{-1}$ yr $^{-1}$]	56.0 $\pm$ 4.3	56.6 $\pm$ 4.7	52.5 $\pm$ 3.6	52.0 $\pm$ 3.2	60.5 $\pm$ 12.7	55.4 $\pm$ 8.3	56.0 $\pm$ 11.9
^{36}Cl [atoms (g K) $^{-1}$ yr $^{-1}$]	157 $\pm$ 12	155 $\pm$ 13	151 $\pm$ 10	152 $\pm$ 9	143 $\pm$ 30	129 $\pm$ 19	129 $\pm$ 27
^{36}Cl [neutrons (g air) $^{-1}$ yr $^{-1}$]	704 $\pm$ 176	690 $\pm$ 181	649 $\pm$ 201	646 $\pm$ 192	721 $\pm$ 226	666 $\pm$ 204	668 $\pm$ 212
^3He [atoms (g olivine) $^{-1}$ yr $^{-1}$]	115 $\pm$ 19	119 $\pm$ 19	118 $\pm$ 18	117 $\pm$ 13	131 $\pm$ 19	123 $\pm$ 14	123 $\pm$ 17
^{14}C [atoms (g quartz) $^{-1}$ yr $^{-1}$]	12.8 $\pm$ 0	12.7 $\pm$ 0	12.2 $\pm$ 0	12.2 $\pm$ 0	13.4 $\pm$ 0.4	12.4 $\pm$ 1.2	12.5 $\pm$ 0

Table 6. Compilation of nuclide production rate constants for ^{26}Al , ^{36}Cl , ^3He , and ^{14}C , from Borchers et al. (2015) and Marrero et al. (2015c) (for ^{36}Cl), with associated uncertainties from this paper.

2.7.4.2 Effects of temporal range of calibration samples

The exposure ages of the primary CRONUS-Earth calibration samples were drawn from a fairly limited interval of geological time. They range from 18.4 to 9.7 ka. In some respects this is advantageous, since production rates will be minimally distorted by any errors in correction for temporal variations in the production rates. However, this relatively narrow range of calibration ages does not directly incorporate into the calibration controls on production rates for samples that may be significantly older or younger than the calibration age range.

The major temporal influence on production rates is modulation of the cosmic-ray flux by fluctuations in the geomagnetic and solar magnetic fields. The effects of these fluctuations are included in CRONUScalc by means of specified geomagnetic and solar magnetic histories, as described in Lifton et al. (2014) and Marrero et al. (2015a). These histories are used to drive models for the site production rate that are specific to each scaling formulation. The degree to which CRONUScalc produces accurate ages for samples outside the calibration age range depends on the accuracy of the magnetic histories and the models describing the modulation of the cosmic-ray flux. In general, the effects of these corrections are not large. Temporal variation of atmospheric structure over glacial/interglacial timescales may also be a significant influence, but our current state of knowledge is insufficient to evaluate the possible magnitude of this effect.

The effects of production-rate variation due to magnetic field fluctuations tend to become smoothed with exposure duration due to integration of the production rate over time, unlike, for instance, meteoric cosmogenic nuclide applications such as radiocarbon dating, for which the precise magnetic field effect must be known at a specific time. After the age reaches about 40 ka the fluctuations in the cosmogenic inventory as a function of time are greatly damped (Gosse and Phillips, 2001). After this point, any errors in the calculated age will be mostly due to any bias in the calibrated production rate, and will systematically increase as the exposure time increases.

The effects of production-rate variations due to magnetic field fluctuations are greatest for short exposure durations, as experienced in very young (late Holocene) samples. Here, the benefits of time-integration are minimal. In this age range the details of the assigned geomagnetic field history are important. Errors in the geomagnetic field history assigned during the calibration interval (18.4 to 9.7 ka), or errors in the cosmic-ray modulation model, could also influence calculated ages of young samples. Errors in the age of young (late Holocene) samples are likely to be greatest if the calibration samples are much older. Given the lack of late Holocene control samples in the primary and secondary data sets, materials in this age range should be a priority for future calibration sampling.

Given that the magnitudes of any errors in either the calibrated production rates or magnetic field histories are not known, it is not possible to quantitatively assess variation in consequent exposure ages as sample age changes. However, any such errors should be smallest for the calibration interval. Any bias in the calibrated production rate will cause increasing absolute error as age increases toward nuclide saturation and errors in the assigned magnetic field history will generally cause increasing relative error as the exposure age decreases toward the present.

2.7.4.3 Comparison of calibrations across nuclides

The CRONUS-Earth effort identified and sampled four major new calibration sites and three minor or secondary ones. The processing and analysis of the data was unprecedented in the number of laboratories participating and the level of replication of analyses. Two new, strongly physically based scaling methods were tested. Multiple nuclides were calibrated using data from the same sites. Although the results did not achieve the hoped-for goal of 5% RMSE, it did establish production rates by firm consensus of a large number of investigators and it has provided data that highlight where future improvements are likely to be most productive.

One surprise from the over-all calibration effort is the consistency of results (i.e., that the relative error of the calculated ages does not vary strongly by nuclide). The average primary-data-set RMSE's (not including the neutron-monitor-based scaling methods) varied from 7% (for ^{10}Be) to 10% (for ^{26}Al and ^3He), with ^{36}Cl not included because of insufficient number of sites. The secondary RMSE's varied from 9.3 % (for spallation ^{36}Cl) to 11.5% (for ^{26}Al), with ^3He presenting an outlier at 19.8% that probably is not representative of the method. All nuclides showed positive biases in the normalized ages, for both primary and secondary data sets. These varied from 0.3% for the secondary ^3He data to 6.1% for the ^{10}Be secondary data. The ages from the primary data sets for all nuclides (again excluding ^{36}Cl) exhibit a very consistent positive bias of 3

to 4%. This bias apparently results from an inability of all the models to scale both high elevation and low elevation Southern Hemisphere data along with mid-latitude, mid-elevation Northern Hemisphere data. The large bias calculated for the ^{10}Be secondary data set (6.1%) is probably the most representative of the actual magnitude of the problem. The relatively low bias of the secondary ^3He data set is probably a result of the limited elevation range of the secondary sites for that nuclide.

These results are somewhat surprising given the findings of the interlaboratory comparisons summarized in Tables 2 and A1. These show marked differences in analytical reproducibility among the various nuclides, ranging from 3 to 4% for ^{10}Be to almost 10% for ^{26}Al . The secondary calibration RMSE's however, show little or no relation to the actual levels of analytical reproducibility. This observation suggests that the magnitude of the RMSE's is dominated by some aspect of the nuclide-production systematics, rather than the analytical uncertainties.

2.7.4.4 Sources of variability in the calibrations

In Table 7 we have summarized the estimated magnitudes (in percent) of the sources of variability in the calibration. We have used ^{10}Be calibrated to the LSDn (Sa) scaling method as representative. Other nuclides and scalings would show somewhat different results. Calculations are shown for both the primary and secondary data sets. Generalized, the primary data set results might be viewed as an optimistic perspective on dating results now achievable and the secondary one as a more pessimistic (or 'realistic') one. Actual, high-quality, studies are likely to produce results that fall between them.

Sources of Uncertainty	Primary	Secondary
Measured RMSE	7.9	11.1
Average reported analytical uncertainty	2.5	2.6*
Estimated actual analytical uncertainty	4.0	6.0
Contribution of site bias (scaling)	6.7	9.1
Other sources	1.3	1.9
Effects of Uncertainty Reduction on RMSE		
Halve actual analytical uncertainty	7.6	10.7
Eliminate scaling bias	4.2	6.3
Both of above	3.7	5.5

*analytical uncertainties of Nishiizumi et al. (1989) Sierra samples not included.

Table 7. Estimated average magnitude of the sources of uncertainty in the ^{10}Be data sets, using LSDn scaling, with values in percent.

The table gives both the average of the actual reported analytical uncertainties and our estimate of the actual reproducibility of the analyses based on Tables 2 and A1 and the discussion in section 2.2 and Appendix A. We estimated the contribution of site bias by calculating the average normalized age for each site (given in Tables 3 and 4), dividing all of the analyses from that site by that average age (so that all sites had a mean age of unity), then recalculating the RMSE's for the entire primary and secondary data

sets (by sample) and taking the difference between the observed RMSE and the recalculated one. The RMSE's with site bias removed were close to the independently estimated actual analytical variability. "Other sources" of variability were estimated by subtracting, in quadrature, the estimated actual analytical uncertainties and contribution of site bias from the observed RMSE.

Based on this analysis, the principal contribution to the observed RMSE is the effect of site bias. Site bias may arise due to an inability to correctly scale the production rate, or due to errors in the site characteristics, especially age, or both. Actual analytical reproducibility that is significantly poorer than the laboratory-reported values makes a secondary contribution to the magnitude of the RMSE. Based on the evaluation in section 2.2, the principal contributor to the excess spread of the data for ^{10}Be is within-laboratory irreproducibility, but interlaboratory bias also plays a role. The magnitude of the remaining sources of uncertainty is small ($<2\%$). These include mainly field parameters such as elevation, sample thickness, density, and so on.

What are the potential effects of the various actions that might be taken to reduce the RMSE in future studies? In the lower half of Table 7 we recalculate the observed RMSE's for the primary and secondary data sets assuming various improvements in the methodology. It is unrealistic to think that analytical uncertainty could be eliminated, so in the second line of the lower half of the table we assume that the actual uncertainty associated with the ^{10}Be measurements could be halved (i.e., from 4.0 to 2.0% for the primary and from 6.0 to 3.0% for the secondary). This would bring the actual analytical uncertainties more-or-less in line with the currently reported laboratory analytical uncertainties. This would result in a reduction of the primary RMSE from 7.9 to 7.6%, a fairly inconsequential change. In contrast, if the scaling bias could be eliminated, the RMSE would go down to 4.2%, a very significant change. Accomplishing both of these objectives would reduce even the secondary RMSE down to 5.5%, close to the stated goal. The effect of a substantial reduction in the "other sources" of uncertainty is not shown, but it would be negligible.

Although this evaluation gives guidance for future research to improve the accuracy and precision of cosmogenic nuclide dating, the results are specific to the nuclide and scaling method used in the example. Other nuclide/scaling combinations will give somewhat different results. For example, actual analytical uncertainty plays a much more important role for ^{26}Al and its reduction would probably reduce the RMSE to an extent about equal to reducing the scaling uncertainty.

2.7.4.5 Nuclide intercalibration

One component of the CRONUS-Earth Project was an effort to measure as many nuclides as possible at each major site. The results are summarized in Table 1 of Borchers et al. (2015). The most successful in this regard was the Bonneville locality. Although two sites were sampled that were associated with ancient Lake Bonneville (quartzite at Promontory Point and basalt at Tabernacle Hill), these are close enough in time and location that they essentially constitute one site. All five nuclides were measured at Bonneville. Four were measured at the Scotland sites (not including ^3He) and three at Huanacané (not including ^3He and ^{14}C , although efforts on the latter are continuing). Unfortunately, only two nuclides (^{10}Be and ^{14}C) were measured at the

Macaulay Valley debris-flow site in New Zealand, although the site has the potential for ^{26}Al and ^{36}Cl measurements as well.

Beryllium-10 was thus co-measured with ^{26}Al , ^{36}Cl , and ^{14}C at three sites and with ^3He at one. The production rates were then calibrated for all of these nuclides using the same field parameters (e.g., elevation, erosion, snow cover) and the same model. This procedure provides very strong linkage between the ^{10}Be , ^{26}Al , and ^{36}Cl production rates, since they were all calibrated using only the five principal CRONUS-Earth primary sites. The linkage for ^{14}C is somewhat weaker, since it was calibrated using three principal sites and three additional sites. It is weakest for ^3He , calibrated using one principal site and nine additional sites.

Checking the accuracy of the intercalibration is more difficult since only two of the secondary sites (New England and Baboon Lakes) had more than one nuclide analyzed. However, at these two sites the direction of the age anomaly (too old at New England and too young at Baboon Lakes) was the same for all three nuclides and the magnitudes were also approximately the same. The basis provided by the CRONUS-Earth sampling and calibration provides a significant increase in the confidence with which ages from different nuclides can be considered equivalent.

2.7.4.6 Effects of revised model and recalibration on cosmogenic-nuclide dating

The CRONUScalc calculator was inspired by the Balco et al. (2008) calculator, but it is different in some fundamental ways, producing differences in the actual results as well as the output style of the results. The most obvious difference between the calculators is the expanded suite of applicable nuclides in CRONUScalc (^{10}Be , ^{26}Al , ^{36}Cl , ^{14}C , and ^3He). In evaluating expected differences between results of exposure-age calculations using the 2008 online calculator and the CRONUScalc calculator of Marrero et al. (2015), it is important to distinguish between (i) the effects of differences in the production rate calibration data set that was available in 2008 and the new calibration data set described by Borchers et al. (2015) and used in CRONUScalc, and (ii) the effects of differences in the calculator systems themselves, that is, the algorithms used to compute exposure ages, the numerical implementation of those algorithms, and the values of ancillary parameters such as, for example, paleomagnetic field reconstructions and particle interaction cross-sections. Simply calculating exposure ages for the same input data using the original default implementation of the 2008 and 2015 online calculators will yield a systematic difference in calculated exposure ages, such that exposure ages calculated using the default implementation of the 2008 online calculator will be, on average, 10% lower than those calculated from the same input data using the default implementation of the 2015 online calculator. For particular combinations of sample age, elevation, and location, the difference between the ages calculated by the two calculators may be substantially greater than 10%. However, if calibration data sets newer than the 2008 one are used in the calculator of Balco et al. (2008), the differences will generally be smaller. The majority of this difference is due to the different calibration data sets and not to the calculator systems themselves. Numerous studies between 2008 and the present (e.g., Balco et al., 2009; Putnam et al., 2010; Young et al., 2013) have shown that the calibration data set available in 2008 yielded inaccurate ages in many cases, and, in addition, many of the individual data in the 2008 data set have

1613 been independently found to be inaccurate or erroneous. Thus, using the 2008 online
1614 calculator with more recent calibration data, rather than only the calibration data
1615 available in 2008, eliminates nearly all the systematic bias between exposure ages
1616 computed with the 2008 and 2015 systems, leaving much smaller differences between the
1617 two.

1618 Figure 7 gives an example of these remaining differences that are attributable to
1619 the calculator systems themselves rather than the calibration data. Here we have chosen a
1620 scaling scheme ("Lm") that is common to both the 2008 and 2015 online calculators, and
1621 computed exposure ages for a synthetic set of samples that have the same ^{10}Be
1622 concentration but are located at a range of locations and elevations. To remove the effect
1623 of differences in the default calibration data sets, we have chosen a value of the reference
1624 ^{10}Be production rate used in the 2008 code that results in a zero systematic bias between
1625 the two sets of calculated exposure ages; in other words, we have used the results from
1626 the 2015 code as the calibration data set for the 2008 code. This highlights differences
1627 between the two systems that are independent of the calibration data set and are mainly
1628 due to (i) an updated paleomagnetic field model, (ii) updated muon interaction cross-
1629 sections, and (iii) differences in the numerical implementation of the scaling scheme and
1630 the depth-dependence of the production rate. This results in differences that are typically
1631 in the range of 0-3% and that vary in a complicated way with elevation, age, and position
1632 in the magnetic field. Note that because we have not repeated the fitting exercise of
1633 Borchers et al. (2015, this volume) for the 2008 online calculator code, we can not
1634 quantify whether these differences imply any corresponding difference in the
1635 performance or accuracy of the two calculators as measured by their fit to calibration
1636 data. In general, differences between exposure ages calculated using the 2008 and 2015
1637 systems will be smaller for the non-time-dependent scaling scheme and, for time-
1638 dependent scaling schemes, for locations and age ranges that are relatively insensitive to
1639 magnetic field variations, such as at latitudes $>60^\circ$, or regions with long-term quasi-
1640 stationary non-dipole field effects. Differences will be larger for time-dependent scaling
1641 schemes in ages and locations that are more sensitive to magnetic field variations.

Figure 7. Example of relative differences between exposure ages calculated for a synthetic set of samples using the 2008 and 2015 online calculators. To generate these results, we used a common scaling scheme ("Lm") available in both systems and chose a reference production rate for the 2008 system that resulted in a zero systematic bias between the two sets of results. This removes differences attributable only to a difference in production-rate calibration data sets and highlights differences caused by differences in the calculator systems themselves. The y-axis is the ratio of the exposure age computed using the 2008 system to that computed using the 2015 system. All samples had $250,000 \pm 12,500$ atoms $^{10}\text{Be/g}$ (approximately equal to that of the Promontory Point average from the CRONUS-Earth calibration study), 1 cm thickness, zero erosion, no topographic shielding, and were at 0° longitude. Data points are shown at latitudes of 90° , 75° , 65° , 55° , 50° , 45° , 40° , 35° , 25° , 15° , and 0° and for altitudes of 0, 2500, and 5000 m.

Due to the longer calculation time necessary to provide the most accurate model in the new framework, we made some changes to the style of the output. While the calculator runs multiple samples in a single submission, each sample only runs a single

selected scaling model. It is time-prohibitive to run all of the scaling methods, especially now that there are seven models. The ability to email results to the user decreases the dependence on a reliable internet connection and reduces the chances of browser timeout while the calculator is running. Emails and all sample data, except usage statistics on general location and nuclide, are erased every 24 hours from the server. Results, original inputs, and all data necessary to recreate the output graphs, are provided to the user in a spreadsheet in addition to being displayed on the screen.

A disadvantage of the new calculator is that computation times are substantially longer than with the old one. This is a necessary consequence of basing the calculator on a single model capable of simulating the production of all nuclides, and of incorporating more physics into the nuclide-production model. The capability of emailing results was added to minimize any internet connection problems related to the longer turnaround times.

3. Conclusions

3.1 Status of analytical capability

The CRONUS-Earth Project conducted extensive intercomparisons of reference materials to attempt to quantify analytical reproducibility at the community level and to understand how that reproducibility related to the analytical uncertainties commonly reported by the analytical laboratories. Numerous intercomparison materials (ICM's) were evaluated in sections 2.1 and 2.2. Although these ICM's are now publically available, analyses reported have been fairly limited. Continued use will improve statistics and provide laboratories with materials necessary for internal and external checks on quality in a way that was not previously possible.

The fundamental finding of these intercomparison studies was that stated analytical uncertainties nearly always underestimate the actual degree of variability, as quantified by the over-all coefficient of variation of the data. The average amount by which the actual coefficient of variation exceeded the analytical uncertainty was a factor of two (100%). The magnitude of this excess ranged from about 15% for ^{10}Be and ^{26}Al measurements on the CRONUS 'N' material to over 300% for ^{14}C on the CRONUS 'A' and ^{21}Ne on the CREU-1 materials. *These results make it clear that uncertainties calculated for individual geological samples based on the propagation of stated laboratory analytical uncertainties, plus uncertainties on associated sample parameters (e.g., elevation, thickness, chemistry), will systematically underestimate actual age uncertainties.*

One bright spot was that, for materials having comparable nuclide concentrations, the results of intercomparisons based on field samples yielded coefficients of variation very similar to those based on laboratory ICM's. This indicates that additional sources of variability peculiar to field samples (e.g., variable erosion, cover by snow or sand, unaccounted-for chemical heterogeneities) were insignificant, at least for suites of field samples carefully selected for uniformity.

Furthermore, our uncertainty analyses use nuclide concentrations that were in many instances measured more than 10 years ago. While it is difficult to demonstrate without exhaustive reexamination of laboratory and AMS facility data, refinements in

mineral separation and target production (Mifsud et al., 2013; Binnie et al., 2015), and improvements in facilities and mass spectrometry (Granger et al., in press; Rood et al., 2013; Zimmerman et al., 2013) will have decreased some of the error that contributed to the total uncertainties.

The nuclide exhibiting the smallest coefficients of variation was ^{10}Be , which varied from 3 to 4% for all the intercomparison materials. If only laboratories that measured three or more aliquots are included, the CoV goes down to 1.5%. These results show that for results from most laboratories, ^{10}Be is already at least marginally capable of producing ages reproducible at the desired 5% level, and some laboratories have significantly surpassed the needed level of reproducibility. Helium-3 was also reproducible at the 3% level on the ‘hot’ CRONUS-P material, but it remains to be demonstrated whether similar reproducibility is possible on materials with concentrations closer to most natural samples. Even the lower-concentration ICM CRONUS-N is higher than most natural samples. ICM exercises on still-lower concentration materials, or smaller aliquots of the higher-concentration ones, will need to be performed to adequately quantify the analytical reproducibility of typical field samples.

The results are not as positive for the other nuclides. Chlorine-36, ^{21}Ne , and ^{14}C all exhibited CoV’s in the range of 6 to 8%. Aluminum-26 was the least reproducible, showing CoV’s ranging from 5 to 11%. Detailed evaluation of the outcomes of the intercomparison studies indicated that the sources of variability included both random, within-laboratory scatter and presumably systematic between-laboratory bias. The relative importance of these two sources varied from nuclide to nuclide. The analytical reproducibility of these nuclides will have to improve by a factor of 2 to 4 in order to hope to produce ages reproducible at the 5% level. This makes it clear that if age accuracy and reproducibility are to approach the 5% level, a systematic program is needed to address issues of analytical reproducibility.

3.2 Issues of scaling and nuclide-production systematics

Traditionally, cosmogenic-nuclide production models have been formulated by taking two data sets, measurements of phenomena proportional in some fashion to the cosmic-ray flux and measurements of cosmogenic nuclides at sites of known age, and inverting both of these to arrive at the spatial distribution of production rates. This approach is primarily empirical.

The CRONUS-Earth Project has addressed this limitation in two ways. The first is the development of the LSD and LSDn models by Lifton et al. (2014). These models avoid the fundamental difficulty of a fully physics-based approach, which is unreasonable computational demands, by employing analytical approximations to the full numerical modeling results. The analytical approximations were originally developed by Sato et al. (2008). This enables the dependence of the energy of the cosmic-ray cascade on atmospheric depth and cutoff rigidity to be incorporated at a feasible computational demand.

The second way the limitation has been addressed is the measurement of the critical energy-dependent neutron-reaction cross sections in laboratory experiments (Caffee et al., 2013; Nishiizumi et al., 2009). When complete, this will provide a much-improved physical basis for the nuclide-dependent production calculations.

One important outcome of this effort, combined with the geological calibration effort, has been the ability to discriminate between the two competing families of scaling models: those based on the original model by Lal (1991), which was founded mainly on counting spallation events in photo-emulsions, and the models based on data from the neutron-monitor network. The fits to the calibration and secondary data for the physically based models of Lifton et al. (2014) are much more similar to the Lal models than the neutron-monitor-based ones. The data do confirm that the neutron-monitor based scalings are lacking at low latitude and high elevation. However, the similarity between the fits for the Lal and LSD scalings should not be interpreted as implying that differences between these models are negligible. Examined in detail, at low latitude/low elevation and at high latitude/high elevation the Lal-based models differ from the LSD models by ~20% (Lifton et al., 2014). Lack of calibration samples from these latitude/elevation combinations (see Figure 4) limits the ability of the CRONUS calibration to discriminate between these scaling models. Such calibration sites should be a high-priority target for future calibration efforts.

What is the origin of the remaining mismatch between the data and the model? One possibility is that the remaining discrepancies between all models and the data set originate from unexplored phenomena, rather than inadequacy in the physics of the models or the data on which they are based. However, given the quite limited number of high-quality calibration sites and their restricted distribution in latitude and altitude (Figure 4), the global distribution of production is still not strongly constrained. In terms of assessing time-variation in production rates, our dataset is even more limited. It is perhaps more likely that either an inadequate distribution of sample sites, or an inadequate control on factors such as interlaboratory bias or geological/geophysical phenomena such as snow cover, solar modulation, long-term atmospheric pressure history, or geomagnetic variation history, are distorting the calibration, even if the underlying physical model is correct. Additional calibration sites at low latitude from both high and low altitude are probably the most effective way to test these alternatives, along with reproducing data from sites measured by only one laboratory (e.g., New Zealand).

3.3 Status of geological calibration

The CRONUS-Earth Project investigated and sampled four major calibration sites and three secondary sites. Major sites used in the primary calibration data sets were Huanacané (Peru), Bonneville shoreline features (Utah, USA), the western Highlands of Scotland, and the Macaulay Valley debris flow, New Zealand. Sites included in the secondary data set were the New England moraines, glacial erratics in the Puget Lowland, and moraine boulders at the Baboon Lakes, all in the USA.

In general, data from these sites produced very good results. For example, the reduced chi-squared value for the ^{10}Be calibration (about 1.5) (Borchers et al., 2015) was about half that from about half that of the best fit to the calibration data available in 2008, which was about 2.8 (Balco et al., 2008). However, the only nuclide that produced fits consistent with the data uncertainties was ^{26}Al . This is probably largely due to the quite large analytical uncertainties assigned to the ^{26}Al data, averaging 10.2%, based on the laboratory intercomparison results. The acceptable fits may, however, have also been

positively influenced by the small number of geological sites and laboratories compared to ^{10}Be .

Reduced chi-squared values for other nuclides (for the primary calibration data sets) ranged from 1.5 for ^{10}Be to 4.5 for ^3He (Borchers et al., 2015). RMSE's ranged from 7% (for ^{10}Be) to 10% (for ^{26}Al and ^3He), which is a small range of variation considering the large differences in actual analytical variability between nuclides. Bias, which was on average always positive, ranged from 4% for ^{10}Be to 3% for ^3He . For ^{10}Be , most of the “excess variability” (i.e., in excess of that predicted based on the input uncertainties) was due to bias in the ages (or concentrations) from certain localities, with random excess spread of the data playing a secondary role. For other nuclides, especially ^{26}Al , random analytical scatter may affect the over-all spread at a magnitude similar to that of the site bias. Within the CRONUS calibration data set, the site bias appears to have some geographical continuity, being prominent in sites around the North Atlantic (Scotland, Norway, New England) and interior North America (Bonneville). However, two considerations dictate caution when evaluating this observation. (1) The entire calibration is interdependent and thus misfits in one part may force apparent errors in another. (2) Speculation regarding a regional bias should be evaluated in the context of a comprehensive evaluation of the numerous existing cosmogenic data sets from the region, and this has not been performed. If it is real, this apparent localization could be attributed to a regional inadequacy of the scaling model, but could also be due to some other regional effect, such as unaccounted-for post-glacial rebound. The extremely small independent age uncertainty attributed to the New Zealand site may also have played a significant role in the difficulty encountered in achieving a satisfactory fit.

Inasmuch as in general the data did not yield statistically justified models, the secondary data sets were used to provide an empirical estimate of the production-parameter uncertainty, based on varying the production uncertainty until the average age uncertainty equaled the RMSE of the secondary data set ages. The secondary RMSE's varied from 9.3% (for spallation ^{36}Cl) to 11.5% (for ^{26}Al), with ^3He presenting an outlier at 19.8% that probably is not representative of the method. The RMSE for samples dominated by production of ^{36}Cl due to low-energy neutron absorption by Cl is also large (about 29%), in this case due to the environmental dependence of the low-energy neutron flux. It is important to note that these production-parameter uncertainties account for multiple sources of uncertainty, including random uncertainty due to laboratory analytical scatter that is larger than stated analytical uncertainty, systematic error manifested as interlaboratory bias, systematic error apparently deriving from spatially coherent inadequacies in all of the scaling models, and possibly additional error due to factors such as incorrect independent ages for calibration or secondary sites. Of these, the error in the scaling models (including factors such as geomagnetic and atmospheric pressure histories) appears in most cases to be the largest contributor to the calculated production-parameter uncertainties.

Phillips et al. (2015a; this volume) propose recommendations for future improvements in cosmogenic-nuclide dating based on these conclusions.

Acknowledgements

The principal support for the CRONUS-Earth Project came through U.S. National Science Foundation grant EAR-0345949 to Phillips. Additional funding was provided through grants EAR-0345150 (Lifton), EAR-0345932 (Finkel), EAR-0345820 (Caffee), EAR-0345949 (Phillips), EAR-0345574 (Stone), EAR-0345835 (Schaefer), EAR-0345817 (Nishiizumi), and EAR-0345949 to Argento. Funding for CRONUS-EU was provided by the European Union's sixth framework program, grant MC-RTN- 511927. Members of both projects gratefully acknowledge the contributions of the late Dr. Devendra Lal.

References

- Ackert, J., Robert P., Singer, B.S., Guillou, H., Kaplan, M.R., Kurz, M.D., 2003. Long-term cosmogenic ^3He production rates from $^{40}\text{Ar}/^{39}\text{Ar}$ and K-Ar dated Patagonian lava flows at 47°S. *Earth and Planetary Science Letters* 210, 119-136.
- Amidon, W.H., 2011. Cosmogenic ^3He production rates in apatite, zircon, and pyroxene inferred from Bonneville flood erosional surfaces. *Quaternary Geochronology* 6, 10-21.
- Antevs, E., 1922. The recession of the Last Ice Sheet in New England. Research Series No. 11, American Geographical Society, 176 p.
- Argento, D.C., Stone, J.O., Reedy, R.C., O'Brien, K., 2015a. Physics-based modeling of cosmogenic nuclides, Part II: Key aspects of *in-situ* cosmogenic nuclide production. *Quaternary Geochronology* 26, 44-55.
- Argento, D.C., Stone, J.O., Reedy, R.C., O'Brien, K., 2015b. Physics-based modeling of cosmogenic nuclides, Part I: Radiation transport methods and new insights. *Quaternary Geochronology* 26, 29-43.
- Arnold, J.R., Libby, W.F., 1949. Age determinations by radiocarbon content : Checks with samples of known age. *Science* 110, 678-680.
- Aster, R.C., Borchers, B., Thurber, C.H., 2012. Parameter Estimation and Inverse Problems, 2nd Ed. ed. Academic Press, Amsterdam.
- Balco, G., Briner, J., Finkel, R.C., Rayburn, J.A., Ridge, J.C., Schaefer, J.M., 2009. Regional beryllium-10 production rate calibration for late-glacial northeastern North America. *Quaternary Geochronology* 4, 93-107.
- Balco, G., Stone, J.O., Lifton, N.A., Dunai, T.J., 2008. A complete and easily accessible means of calculating surface exposure ages or erosion rates from ^{10}Be and ^{26}Al measurements. *Quaternary Geochronology* 3, 174-195.
- Bevington, P.R., Robinson, D.K., 1992. Data Reduction and Error Analysis for the Physical Sciences, 2nd Ed. ed. McGraw-Hill, New York.
- Binnie, S. A., Dunai, T. J., Voronina, E., Goral, T., Heinze, S., & Dewald, A. (2015). Separation of Be and Al for AMS using single-step column chromatography. *Nuclear Instruments and Methods in Physics Research Section B*, in press .
- Blard, P.-H., Balco, G., Burnard, P.G., Farley, K.A., Fenton, C.R., Friedrich, R., Jull, A.J.T., Niedermann, S., Pik, R., Schaefer, J.M., Scott, E.M., Shuster, D.L., Stuart, F.M., Stute, M., Tibari, B., Winckler, G., Zimmerman, L., 2015. An inter-laboratory comparison of cosmogenic ^3He and radiogenic ^4He in the CRONUS-P pyroxene standard. *Quaternary Geochronology* 26, 11-19.

1889 Blard, P.H., Braucher, R., Lavé, J., Bourlès, D., 2013a. Cosmogenic ^{10}Be production rate
1890 calibrated against ^3He in the high Tropical Andes (3800–4900 m, 20–22° S). *Earth*
1891 *and Planetary Science Letters* 382, 140-149.

1892 Blard, P.H., Lavé, J., Sylvestre, F., Placzek, C.J., Claude, C., Galy, V., Condom, T.,
1893 Tibari, B., 2013b. Cosmogenic ^3He production rate in the high tropical Andes
1894 (3800 m, 20°S): Implications for the local last glacial maximum. *Earth and*
1895 *Planetary Science Letters* 377-378, 260-275.

1896 Blard, P.H., Pik, R., Lave, J., Bourles, D., Burnard, P.G., Yokochi, R., Marty, B.,
1897 Trusdell, F., 2006. Cosmogenic ^3He production rates revisited from evidences of
1898 grain size dependent release of matrix-sited helium. *Earth and Planetary Science*
1899 *Letters* 247, 222-234.

1900 Borchers, B., Marrero, S., Balco, G., Caffee, M., Goehring, B., Gosse, J., Lifton, N.,
1901 Nishiizumi, K., Phillips, F., Schaefer, J., Stone, J.O.H., 2015. Geological
1902 calibration of spallation production rates for terrestrial cosmogenic nuclides in the
1903 CRONUS-Earth Project. *Quaternary Geochronology*, this issue.

1904 Bourlès, D., Raisbeck, G.M., Yiou, F., Loiseaux, J.M., Lieuvain, M., Klein, J., Middleton,
1905 R., 1984. Investigation of the possible association of ^{10}Be and ^{26}Al with biogenic
1906 matter in the marine environment. *Nuclear Instruments and Methods in Physics*
1907 *Research, Section B* 5, 365-370.

1908 Braucher, R., Brown, E.T., Bourlès, D.L., Colin, F., 2003. In situ produced ^{10}Be
1909 measurements at great depths: implications for production rates by fast muons.
1910 *Earth and Planetary Science Letters* 211, 251-258.

1911 Braucher, R., Merchel, S., Borgomano, J., Bourlès, D.L., 2011. Production of
1912 cosmogenic nuclides at great depth: A multi element approach. *Earth and*
1913 *Planetary Science Letters* 309, 1-9.

1914 Bromley, G.R.M., Putnam, A.E., Rademaker, K.M., Lowell, T.V., Schaefer, J.M., Hall,
1915 B., Winckler, G., Birkel, S.D., Borns, H.W., 2014. Younger Dryas deglaciation of
1916 Scotland driven by warming summers. *Proceedings of the National Academy of*
1917 *Science USA* 111, 6215-6219.

1918 Caffee, M.W., Nishiizumi, K., Sisterson, J.M., Ullmann, J., Welten, K.C., 2013. Cross
1919 section measurements at neutron energies 71 and 112 MeV and energy integrated
1920 cross section measurements ($0.1 < E_n < 750$ MeV) for the neutron induced
1921 reactions $\text{O}(n,x)^{10}\text{Be}$, $\text{Si}(n,x)^{10}\text{Be}$, and $\text{Si}(n,x)^{26}\text{Al}$. *Nuclear Instruments and*
1922 *Methods in Physics Research Section B* 294, 479-483.

1923 Carlson, P., 2012. A century of cosmic rays. *Physics Today* 65, 30-36.

1924 Cerling, T.E., Craig, H., 1994. Cosmogenic ^3He production rates from 39° N to 46° N
1925 latitude, western U.S.A. and France. *Geochimica et Cosmochimica Acta* 58, 249-
1926 255.

1927 Charalambus, S., 1971. Nuclear transmutation by negative stopped muons and the
1928 activity induced by the cosmic-ray muons. *Nuclear Physics A* 166, 145-161.

1929 Clark, D.H., Gillespie, A.R., 1997. Timing and significance of late-glacial and Holocene
1930 cirque glaciation in the Sierra Nevada, California. *Quaternary International* 38/39,
1931 21-38.

1932 Clem, J.M., De Angelis, G., Goldhagen, P., Wilson, J.W., 2004. New calculations of the
1933 atmospheric cosmic radiation field - Results for neutron spectra. *Radiation*
1934 *Protection and Dosimetry* 110, 423-428.

- 1935 Craig, H., Poreda, R.J., 1986. Cosmogenic ^3He in terrestrial rocks : The summit lavas of
1936 Maui. Proceedings of the National Academy of Science USA 83, 1970-1974.
- 1937 CRONUS-Earth Project, 2005. Sampling Issues and Protocols. Cosmic-Ray Produced
1938 Nuclide Systematics on Earth Project, Coeur d'Alene, Idaho, Workshop 26-28
1939 May 2005, p. 16.
- 1940 Davis, R.J., Schaeffer, O.A., 1955. Chlorine-36 in nature. Annals of the New York
1941 Academy of Sciences 62, 105-122.
- 1942 Desilets, D., Zreda, M., 2001. On scaling cosmogenic nuclide production rates for
1943 altitude and latitude using cosmic-ray measurements. Earth and Planetary Science
1944 Letters 193, 213-225.
- 1945 Desilets, D., Zreda, M., 2003. Spatial and temporal distribution of secondary cosmic-ray
1946 nucleon intensities and applications to in situ cosmogenic dating. Earth and
1947 Planetary Science Letters 206, 21-42.
- 1948 Desilets, D., Zreda, M., Prabu, T., 2006. Extended scaling factors for in situ cosmogenic
1949 nuclides: New measurements at low latitude Earth and Planetary Science Letters
1950 246, 265-276.
- 1951 Dunai, T.J., 2000. Scaling factors for production rates of in situ produced cosmogenic
1952 nuclides: a critical reevaluation. Earth and Planetary Science Letters 176, 157-
1953 169.
- 1954 Dunai, T.J., 2001. Influence of secular variation of the geomagnetic field on production
1955 rates of in situ produced cosmogenic nuclides. Earth and Planetary Science
1956 Letters 193, 197-212.
- 1957 Dunai, T.J., Binnie, S.A., Hein, A.S., Paling, S.M., 2014. The effects of a hydrogen-rich
1958 ground cover on cosmogenic thermal neutrons: Implications for exposure dating.
1959 Quaternary Geochronology 22, 183-191.
- 1960 Dunai, T.J., Wijbrans, J.R., 2000. Long-term cosmogenic ^3He production rates (152 ka-
1961 1.35 Ma) from $^{40}\text{Ar}/^{39}\text{Ar}$ dated basalt flows at 29°N latitude. Earth and Planetary
1962 Science Letters 176, 147-156.
- 1963 Elmore, D., Fulton, B.R., Clover, M.R., Marsden, J.R., Gove, H.E., Naylor, H., Purser,
1964 K.H., Kilius, L.R., Beukens, R.P. and Litherland, A.E. (1979) Analysis of ^{36}Cl in
1965 environmental water samples using an electrostatic accelerator. Nature 277, 22-
1966 25.
- 1967 Elmore, D., Conard, N., Kubik, P.W., Fabryka-Martin, J., 1984. Computer controlled
1968 isotope ratio measurements and data analysis. Nuclear Instruments and Methods
1969 in Physics Research, Section B5, 233-237.
- 1970 Evans, J.M., 2001. Calibration of the production rates of cosmogenic ^{36}Cl from
1971 potassium. Ph.D. dissertation, Australia National University, Canberra, 142 p.
- 1972 Evans, J.M., Stone, J.O.H., Fifield, L.K., Cresswell, R.G., 1997. Cosmogenic chlorine-36
1973 production in K-feldspar. Nuclear Instruments and Methods in Physics Research,
1974 Section B123, 334-340.
- 1975 Farber, D.L., Hancock, G.S., Finkel, R.C., Rodbell, D.T., 2005. The age and extent of
1976 tropical alpine glaciation in the Cordillera Blanca, Peru. Journal of Quaternary
1977 Science 20, 759-776.
- 1978 Fenton, C.R., Hermanns, R.L., Blikra, L.H., Kubik, P.W., Bryant, C., Niedermann, S.,
1979 Meixner, A., Goethals, M.M., 2011. Regional ^{10}Be production rate calibration for

1980 the past 12+ka deduced from the radiocarbon-dated Grøtlandsura and Russenes
1981 rock avalanches at 69° N, Norway. *Quaternary Geochronology* 6, 437-452.

1982 Fenton, C.R., Mark, D.F., Barfod, D.N., Niedermann, S., Goethals, M.M., Stuart, F.M.,
1983 2013. $^{40}\text{Ar}/^{39}\text{Ar}$ dating of the SP and Bar Ten lava flows AZ, USA: Laying the
1984 foundation for the SPICE cosmogenic nuclide production-rate calibration project.
1985 *Quaternary Geochronology* 18, 158-172.

1986 Fifield, L.K., Tims, S.G., Gladkis, L.G., Morton, C.R., 2007. ^{26}Al measurements with
1987 ^{10}Be counting statistics. *Nuclear Instruments and Methods in Physics Research*,
1988 Section B 259, 178-183.

1989 Foeken, J.P.T., Stuart, F.M., Mark, D.F., 2012. Long-term low latitude cosmogenic ^3He
1990 production rate determined from a 126 ka basalt from Fogo, Cape Verdes. *Earth*
1991 *and Planetary Science Letters* 359/360, 14-25.

1992 Galbraith, R.F., Laslett, G.M., 1993. Statistical models for mixed fission track ages.
1993 *Nuclear Tracks and Radiation Measurement* 21, 459-470.

1994 Galbraith, R.F., Roberts, R.G., Laslett, G.M., Yoshida, H., Olley, J.M., 1999. Optical
1995 dating of single and multiple grains of quartz from Jinmium Rock Shelter,
1996 Northern Australia: Part 1, Experimental design and statistical models.
1997 *Archeometry* 41, 339-364.

1998 Gheorghiu, D.M., Fabel, D., Hansom, J.D., Xu, S., 2012. Lateglacial surface exposure
1999 dating in the Monadhliath Mountains, Central Highlands, Scotland. *Quaternary*
2000 *Science Reviews* 41, 132-146.

2001 Goehring, B.M., Kurz, M.D., Balco, G., Schaefer, J.M., Licciardi, J., Lifton, N., 2010. A
2002 reevaluation of in situ cosmogenic ^3He production rates. *Quaternary*
2003 *Geochronology* 5, 410-418.

2004 Goehring, B.M., Lohne, Ø.S., Mangerud, J., Svendsen, J.-I., Gyllencreutz, R., Schaefer,
2005 J., Finkel, R., 2012a. Erratum: Late Glacial and Holocene Beryllium-10
2006 production rates from western Norway. *Journal of Quaternary Science* 27, 544.

2007 Goehring, B.M., Lohne, Ø.S., Mangerud, J., Svendsen, J.-I., Gyllencreutz, R., Schaefer,
2008 J., Finkel, R., 2012b. Late Glacial and Holocene beryllium-10 production rates
2009 from western Norway. *Journal of Quaternary Science* 27, 89-96, doi:
2010 10.1002/jqs.1517.

2011 Goldhagen, P., Clem, J.M., Wilson, J.W., 2004. The energy spectrum of cosmic-ray
2012 induced neutrons measured on an airplane over a wide range of altitude and
2013 latitude. *Radiation Protection Dosimetry* 110, 387-392.

2014 Gordon, M.S., Goldhagen, P., Rodbell, K.P., Zabel, T.H., Tang, H.H.K., Clem, J.M.,
2015 Bailey, P., 2004. Measurement of the flux and energy spectrum of cosmic-ray
2016 induced neutrons on the ground. *IEEE Transactions on Nuclear Science* 51, 3427-
2017 3434.

2018 Gosse, J.C., Evenson, E.B., Klein, J., Lawn, B., Middleton, R., 1995. Precise cosmogenic
2019 ^{10}Be measurements in western North America: Support for a global Younger
2020 Dryas cooling event. *Geology* 23, 877-880.

2021 Gosse, J.C., Phillips, F.M., 2001. Terrestrial in-situ cosmogenic nuclides: Theory and
2022 application. *Quaternary Science Reviews* 20, 1475-1560.

2023 Granger, D. E., Gibbon, R. J., Kuman, K., Clarke, R. J., Bruxelles, L., & Caffee, M. W.
2024 (2015). New cosmogenic burial ages for Sterkfontein Member 2 Australopithecus
2025 and Member 5 Oldowan. *Nature* 522, 85-88.

2026 Heidebreder, E., Pinkau, K., Reppin, C., Schönfelder, V., 1971. Measurement of the
 2027 distribution in energy and angle of high-energy neutrons in the lower atmosphere.
 2028 Journal of Geophysical Research 76, 2905-2916.
 2029 Heisinger, B., Lal, D., Jull, A.J.T., Kubik, P., Ivy-Ochs, S., Knie, K., Nolte, E., 2002a.
 2030 Production of selected cosmogenic radionuclides by muons: 2. Capture of
 2031 negative muons. Earth and Planetary Science Letters 200, 357-369.
 2032 Heisinger, B., Lal, D., Jull, A.J.T., Kubik, P., Ivy-Ochs, S., Neumaier, S., Knie, K.,
 2033 Lazarev, V., Nolte, E., 2002b. Production of selected cosmogenic radionuclides
 2034 by muons 1. Fast muons. Earth and Planetary Science Letters 200, 345-355.
 2035 Jull, A.J.T., Scott, E.M., Bierman, P., 2015. The CRONUS-Earth inter-comparison for
 2036 cosmogenic isotope analysis. Quaternary Geochronology 26, 3-10.
 2037 Kaplan, M.R., Strelin, J.A., Schaefer, J.M., Denton, G.H., Finkel, R.C., Schwartz, R.,
 2038 Putnam, A.E., Vandergoes, M.J., Goehring, B.M., Travis, S.G., 2011. In-situ
 2039 cosmogenic ^{10}Be production rate at Lago Argentino, Patagonia: Implications for
 2040 late-glacial climate chronology. Earth and Planetary Science Letters 309, 21-32.
 2041 Kelly, M.A., Lowell, T.V., Applegate, P.J., Phillips, F.M., Schaefer, J., Smith, C.A., Kim,
 2042 H., Leonard, K.C., Hudson, A.M., 2015. A locally calibrated, late glacial ^{10}Be
 2043 production rate from a low-latitude, high-altitude site in the Peruvian Andes.
 2044 Quaternary Geochronology 26, 70-85.
 2045 Kelly, M.A., Lowell, T.V., Applegate, P.J., Smith, C.A., Phillips, F.M., Hudson, A.M.,
 2046 2012. Late glacial fluctuations of Quelccaya Ice Cap, southeastern Peru. Geology
 2047 40, 991-994.
 2048 Kim, K.J., Englert, P.A.J., 2004. Profiles of in situ ^{10}Be and ^{26}Al at great depths at the
 2049 Macraes Flat, East Otago, New Zealand. Earth and Planetary Science Letters 223,
 2050 113-126.
 2051 Kohl, C.P., Nishiizumi, K., 1992. Chemical isolation of quartz for measurement of in-
 2052 situ-produced cosmogenic nuclides. Geochimica et Cosmochimica Acta 56, 3583-
 2053 3588.
 2054 Kowatari, M., Nagaoka, K., Satoh, S., Ohta, Y., Abukawa, J., Tachimori, S. and
 2055 Nakamura, T., 2005. Evaluation of the altitude variation of the cosmic-ray
 2056 induced environmental neutrons in the Mt. Fuji area. Journal of Nuclear Science
 2057 and Technology, 42, 495-502.
 2058 Kubik, P.W., Korschinek, G., Nolte, E., 1984. Accelerator mass spectrometry with
 2059 completely stripped chlorine-36 ions at the Munich postaccelerator. Nuclear
 2060 Instruments and Methods in Physics Research, Section B 229, 51-59.
 2061 Kurz, M.D., 1986. Cosmogenic helium in a terrestrial rock. Nature 320, 435-439.
 2062 Lal, D., 1958. Investigations of nuclear interactions produced by cosmic rays. Ph.D.
 2063 dissertation, Tata Inst. of Fundam. Res., Bombay, India.
 2064 Lal, D., 1991. Cosmic-ray labeling of erosion surfaces : *in situ* production rates and
 2065 erosion models. Earth and Planetary Science Letters 104, 424-439.
 2066 Libby, W.F., 1946. Atmospheric helium three and radiocarbon from cosmic radiation.
 2067 Physics Reviews 69, 671-673.
 2068 Licciardi, J.M., Kurz, M.D., Clark, P.U., Brook, E.J., 1999. Calibration of cosmogenic
 2069 ^3He production rates from Holocene lava flows in Oregon, USA, and effects of
 2070 the Earth's magnetic field. Earth and Planetary Science Letters 172, 261-271.

2071 Licciardi, J.M., Kurz, M.D., Curtice, J.M., 2006. Cosmogenic ^3He production rates from
 2072 Holocene lava flows in Iceland. *Earth and Planetary Science Letters* 246, 251-
 2073 264.
 2074 Lifton, N., Caffee, M., Finkel, R., Theriault, B.D., Marrero, S., Gosse, J., Nishiizumi, K.,
 2075 Phillips, F.M., Stone, J., Schaefer, J., Jull, A.J.T., Fifield, N.K., Goehring, B.,
 2076 2015. In situ cosmogenic nuclide production rate calibration for the CRONUS-
 2077 Earth Project from Lake Bonneville, Utah, shoreline features. *Quaternary*
 2078 *Geochronology* 26, 56-69.
 2079 Lifton, N., Sato, T., Dunai, T., 2014. Scaling *in situ* cosmogenic nuclide production rates
 2080 using analytical approximations to atmospheric cosmic-ray fluxes. *Earth and*
 2081 *Planetary Science Letters* 386, 149-160.
 2082 Lifton, N.A., Bieber, J.W., Clem, J.M., Duldig, M.L., Evenson, P., Humble, J.E., Pyle,
 2083 R., 2005. Addressing solar modulation and long-term uncertainties in scaling *in*
 2084 *situ* cosmogenic nuclide production rates. *Earth and Planetary Science Letters*
 2085 239, 140-161.
 2086 Marrero, S., Phillips, F.M., Borchers, B., Lifton, N., Aumer, R., 2015a. Cosmogenic
 2087 nuclide systematics and the CRONUScalc program. *Quaternary Geochronology*,
 2088 this issue.
 2089 Marrero, S., Phillips, F.M., Caffee, M., Stone, J.O.H., Swanson, T.W., Hinz, M., 2015b.
 2090 A reevaluation of the Puget Lowland cosmogenic ^{36}Cl calibration site. *Quaternary*
 2091 *Geochronology*, this issue.
 2092 Marrero, S., Phillips, F.M., Hinz, M., Caffee, M., Gosse, J., 2015c. CRONUS-Earth
 2093 cosmogenic ^{36}Cl calibration. *Quaternary Geochronology*, this issue.
 2094 Marrero, S.M., Caffee, M.W., Lifton, N., Phillips, F.M., Stone, J.O.H., Theriault,
 2095 B., Owen, L.A., Ballantyne, C. and Fifield, K., 2015d. Cosmogenic nuclides from
 2096 CRONUS-Earth geologic calibration sites in Scotland. *Quaternary*
 2097 *Geochronology*, this issue.
 2098 Masarik, J., 2002. Numerical simulation of in-situ production of cosmogenic nuclides.
 2099 *Geochimica et Cosmochimica Acta* (Abstr. 12th Annual V.M. Goldschmidt Conf.,
 2100 Davos, Switzerland, August 18-23, 2002) 66, A491.
 2101 Masarik, J., Reedy, R.C., 1995. Terrestrial cosmogenic-nuclide production systematics
 2102 calculated from numerical simulations. *Earth and Planetary Science Letters* 136,
 2103 381-395.
 2104 Mercer, J.H., Palacios, O.M., 1977. Radiocarbon dating of the last glaciation in Peru.
 2105 *Geology* 5, 600-604.
 2106 Mifsud, C., Fujioka, T., & Fink, D. (2013). Extraction and purification of quartz in rock
 2107 using hot phosphoric acid for in situ cosmogenic exposure dating. *Nuclear*
 2108 *Instruments and Methods in Physics Research Section B*.294, 203-207.
 2109 Muller, R.J., Stephenson, E.J., Mast, T.S., 1978. Radioisotope measurement with an
 2110 accelerator: A blind measurement. *Science* 201, 347-348.
 2111 Nakamura, T., Nunomiya, T., Terunuma, K., Suzuki, H., 2005. Sequential measurements
 2112 of cosmic-ray neutron spectrum and dose rate at sea level in Sendai, Japan.
 2113 *Journal of Nuclear Science and Technology* 42, 843-853.
 2114 Nishiizumi, K., Lal, D., Klein, J., Middleton, R., Arnold, J.R., 1986. Production of ^{10}Be
 2115 and ^{26}Al by cosmic rays in terrestrial quartz in-situ and implications for erosion
 2116 rates. *Nature* 319, 134-136.

2117 Nishiizumi, K., Welten, K.C., Matsumura, H., Caffee, M.W., Ninomiya, K., Omoto, T.,
 2118 Nakagaki, R., Shima, T., Takahashi, N., Sekimoto, S., Yashima, H., Shibata, S.,
 2119 Bajo, K., Nagao, K., Kinoshita, N., Imamura, M., Sisterson, J.M., Shinohara, A.,
 2120 2009. Measurements of high-energy neutron cross sections for accurate
 2121 cosmogenic nuclide production rates. *Geochimica et Cosmochimica Acta* 73,
 2122 A945.
 2123 Nishiizumi, K., Arnold, J.R., Kohl, C.P., Caffee, M.W., Masarik, J., Reedy, R.C., 2009b,
 2124 Solar cosmic ray records in lunar rock 64455. *Geochimica et Cosmochimica Acta*
 2125 73, 2163-2176.
 2126 Nishiizumi, K., Winterer, E.L., Kohl, C.P., Lal, D., Arnold, J.R., Klein, J., Middleton, R.,
 2127 1989. Cosmic ray production rates of ^{10}Be and ^{26}Al in quartz from glacially
 2128 polished rocks. *Journal of Geophysical Research* 94, 17907-17915.
 2129 O'Brien, K., 2008. Limitations of the use of the vertical cut-off to calculate cosmic-ray
 2130 propagation in the Earth's atmosphere. *Radiation Protection and Dosimetry* 128,
 2131 259-260.
 2132 Palonen, V., Tikkanen, P., 2007. A shot at a Bayesian model for data analysis in AMS
 2133 measurements. *Nuclear Instruments and Methods in Physics Research, Section B*
 2134 259, 154-157.
 2135 Phillips, F.M., 2015. Constraints on cosmogenic nuclide production rates by samples
 2136 from the Sierra Nevada, California: I. Late Pleistocene glacial chronology.
 2137 *Quaternary Geochronology*, this issue.
 2138 Phillips, F.M., Argento, D.C., Caffee, M.W., Dunai, T.J., Finkel, R., Goehring, B., Gosse,
 2139 J.C., Hudson, A., Jull, A.J.T., Kelly, M., Lifton, N., Marrero, S.M., Nishiizumi,
 2140 K., Reedy, R., Stone, J.O.H., 2015a. Where now? Reflections on future directions
 2141 for cosmogenic nuclide research from the CRONUS Projects. *Quaternary*
 2142 *Geochronology*, this issue.
 2143 Phillips, F.M., Hinz, M., Marrero, S.M., Nishiizumi, K., 2015b. Constraints on
 2144 cosmogenic nuclide production rates by samples from the Sierra Nevada,
 2145 California: II. Sample sites and evaluation. *Quaternary Geochronology*, this
 2146 issue.
 2147 Phillips, F.M., Kelly, M.A., Hudson, A.M., Stone, J.O.H., Schaefer, J., Marrero, S.M.,
 2148 Fifield, N.K., Finkel, R., Lowell, T.V., 2015c. CRONUS-Earth calibration
 2149 samples from the Huancané II moraines, Quelccaya Ice Cap, Peru. *Quaternary*
 2150 *Geochronology*, this issue.
 2151 Phillips, F.M., Leavy, B.D., Jannik, N.O., Elmore, D., Kubik, P.W., 1986. The
 2152 accumulation of cosmogenic chlorine-36 in rocks: A method for surface exposure
 2153 dating. *Science* 231, 41-43.
 2154 Phillips, F.M., Stone, W.D., Fabryka-Martin, J.T., 2001. An improved approach to
 2155 calculating low-energy cosmic-ray neutron fluxes near the land/atmosphere
 2156 interface. *Chemical Geology* 175, 689-701.
 2157 Phillips, F.M., Zreda, M., Plummer, M.A., Elmore, D., Clark, D.H., 2009. Glacial
 2158 geology and chronology of Bishop Creek and vicinity, eastern Sierra Nevada,
 2159 California. *Geological Society of America Bulletin* 121, 1013-1033.
 2160 Poreda, R.J., Cerling, T.E., 1992. Cosmogenic neon in recent lavas from the western
 2161 United States. *Geophysical Research Letters* 19, 1863-1866.

2162 Putnam, A.E., Schaefer, J.M., Barrell, D.J.A., Vandergoes, M., Denton, G.H., Kaplan,
2163 M.R., Finkel, R.C., Schwartz, R., Goehring, B.M., Kelley, S.E., 2010. In situ
2164 cosmogenic ^{10}Be production-rate calibration from the Southern Alps, New
2165 Zealand. *Quaternary Geochronology* 5, 392-409.

2166 Raisbeck, G.M., Yiou, F., Fruneau, M. and Loiseaux, J.M. (1978a) Beryllium-10 mass
2167 spectrometry with a cyclotron. *Science* 202, 215-217.

2168 Raisbeck, G.M., Yiou, F., Fruneau, M., Lieuvin, M., Loiseaux, J.M.(1978b)
2169 Measurement of beryllium-10 in 1,000-and 5,000-year-old Antarctic ice. *Nature*
2170 257, 731-733.

2171 Raisbeck, G.M., Yiou, F., Klein, J., Middleton, R., 1983. Accelerator mass spectrometry
2172 measurement of cosmogenic ^{26}Al in terrestrial and extraterrestrial matter. *Nature*
2173 301, 690-692.

2174 Reedy, R.C., 2013. Cosmogenic nuclide production rates: Reaction cross section update.
2175 *Nuclear Instruments and Methods in Physics Research, Section B* 294, 470-474.

2176 Ridge, J.C., Balco, G., Bayless, R.L., Beck, C.C., Carter, L.B., Dean, J.L., Voytek, E.B.,
2177 Wei, J.H., 2012. The new North American Varve Chronology: A precise record of
2178 southeastern Laurentide Ice Sheet deglaciation and climate, 18.2-12.5 kyr BP, and
2179 correlations with Greenland ice core records. *American Journal of Science* 312,
2180 685-722.

2181 Rood, D. H., Brown, T. A., Finkel, R. C., & Guilderson, T. P. (2013). Poisson and non-
2182 Poisson uncertainty estimations of $^{10}\text{Be}/^9\text{Be}$ measurements at LLNL-CAMS.
2183 *Nuclear Instruments and Methods in Physics Research Section B* 294, 426-429.

2184 Sato, T., Niita, K., 2006. Analytical functions to predict cosmic-ray neutron spectra in the
2185 atmosphere. *Radiation Research* 166, 544-555.

2186 Sato, T., Yasuda, H., Niita, K., Endo, A., Sihver, L., 2008. Development of PARMA:
2187 PHITS-based analytical radiation model in the atmosphere. *Radiation Research*
2188 170, 244-259.

2189 Schaefer, J.M., Winckler, G., Blard, P.-H., Balco, G., Shuster, D.L., Friedrich, R., Jull,
2190 A.J.T., Wieler, R., Schluechter, C., 2015. Performance of CRONUS-P – a
2191 pyroxene reference material for helium isotope analysis. *Quaternary*
2192 *Geochronology*, this issue.

2193 Schimmelpfennig, I., Benedetti, L., Garreta, V., Pik, R., Blard, P.-H., Burnard, P.,
2194 Boulès, D., Finkel, R., Ammon, K., Dunai, T., 2011a. Calibration of cosmogenic
2195 ^{36}Cl production rates from Ca and K spallation in lava flows from Mt. Etna (38°N,
2196 Italy) and Payun Matru (36°S, Argentina). *Geochimica et Cosmochimica Acta* 75,
2197 2611-2632.

2198 Schimmelpfennig, I., Williams, A., Pik, R., Burnard, P., Niedermann, S., Finkel, R.,
2199 Schneider, B., Benedetti, L., 2011b. Inter-comparison of cosmogenic in-situ ^3He ,
2200 ^{21}Ne and ^{36}Cl at low latitude along an altitude transect on the SE slope of
2201 Kilimanjaro volcano (3°S, Tanzania). *Quaternary Geochronology* 6, 425-436.

2202 Schimmelpfennig, I., Schaefer, J.M., Goehring, B.M., Lifton, N., Putnam, A.E., Barrell,
2203 D.J.A., 2012. Calibration of the in situ cosmogenic ^{14}C production rate in New
2204 Zealand's Southern Alps. *Journal of Quaternary Science* 27, 671-674.

2205 Stone, J.O., 2000. Air pressure and cosmogenic isotope production. *Journal of*
2206 *Geophysical Research* 105, 23753-23759.

- Stone, J.O., 2005. Terrestrial chlorine-36 production from spallation of iron [abstr.], The 10th International Conference on Accelerator Mass Spectrometry, Berkeley, CA.
- Stone, J.O.H., Evans, J.M., Fifield, L.K., Allan, G.L., Cresswell, R.G., 1998. Cosmogenic chlorine-36 production in calcite by muons. *Geochimica et Cosmochimica Acta* 62, 433-454.
- Stuart, F.M., Dunai, T.J., 2009. Advances in cosmogenic isotope research from CRONUS-EU. *Quaternary Geochronology* 4, 435-436.
- Swanson, T.W., Caffee, M.L., 2001. Determination of ^{36}Cl production rates derived from the well-dated deglaciation surfaces of Whidbey and Fidalgo Islands, Washington. *Quaternary Research* 56, 366-382.
- Usoskin, I.G., Kovaltsov, G.A., 2006. Cosmic ray induced ionization in the atmosphere: Full modeling and practical applications. *Journal of Geophysical Research* 111, 1-9.
- Vermeesch, P., 2009. RadialPlotter: a Java application for fission track, luminescence and other radial plots. *Radiation Measurements* 44, 409-410.
- Vermeesch, P., Balco, G., Blard, P.-H., Dunai, T.J., Kober, F., Niedermann, S., Shuster, D.L., Strasky, S., Stuart, F.M., Wieler, R., Zimmerman, L., 2015. Interlaboratory comparison of cosmogenic ^{21}Ne in quartz. *Quaternary Geochronology* 26, 20-28.
- Zimmermann, L., Blard, P. H., Burnard, P., Medynski, S., Pik, R., and Puchol, N. (2012). A new single vacuum furnace design for cosmogenic ^3He dating. *Geostandards and Geoanalytical Research* 36, 121-129.
- Zweck, C., Zreda, M., Desilets, D., 2013. Snow shielding factors for cosmogenic nuclide dating inferred from Monte Carlo neutron transport simulations. *Earth and Planetary Science Letters* 379, 64-71.

Appendix A.

Evaluation of Sources of Excessive Scatter in Intercomparison Data Sets

Table 2 in the main text shows widespread overdispersion. The purpose of this appendix is to evaluate, by means of a detailed statistical analysis, the possible causes of this ‘excess’ variability. The causes can be grouped as either random, mostly within-laboratory sources, or systematic, mostly between-laboratory, sources. ‘Geological’ sources of variability uniquely affecting field data sets were not included, based on the evaluation in section 2.1 that they contributed little additional variability. In order to investigate the sources of variability, summary statistics by laboratory and between laboratories have been listed in Table A1. If the overdispersion is mainly due to simple underestimation of random uncertainties, then the coefficients of variation of the individual laboratory data sets should be large compared to the average laboratory analytical uncertainties and the average absolute percent deviation from the mean of the data sets should be small in comparison. The average overdispersion of the laboratory data sets should be large compared to the overdispersion of the laboratory means. If, on the other hand, the overdispersion has interlaboratory systematic origins, the average absolute percent deviation from the mean of the laboratory data sets should be large compared to the average of the analytical uncertainties. Correspondingly, the

overdispersion of the laboratory means should be large compared to the overdispersion within the laboratory data sets.

For the evaluation to be meaningful, the data set must contain both an adequate number of laboratories participating and an adequate number of replicates by each laboratory. A minimum of three was accepted for both replicates and laboratories. Some data sets, such as CRONUS-N, did not have sufficient replication to be worth evaluating. For field data sets there exists the possibility that intra-laboratory variability could be due to heterogeneity between field samples, rather than true analytical variability. Therefore in Table A1 the field data sets all have statistics by sample shown first, in order to allow evaluation of the sample-to-sample variability. The various data sets must be evaluated individually.

A.1 ^{10}Be – CRONUS-A

The results from analysis of the CRONUS-A material (Jull et al., 2015) in Table 2 show a moderate degree of overdispersion (1.9%, compared to an average analytical uncertainty of 1.7%). The results were evaluated on a laboratory basis in Table A1. Only laboratories that returned at least three replicate analyses were included (5 out of 13). For these laboratories, the average coefficient of variation (within laboratory; 1.5%) was comparable to the average analytical uncertainty (1.8%) and was larger than the average absolute deviation from the mean of all these laboratories (0.9%). These results suggest that there is no overdispersion, and this is confirmed by comparing the average overdispersion by laboratory (0.2%) with the overdispersion of the laboratory means (0%).

This evaluation indicates that for the five laboratories that returned more than three analyses the stated analytical uncertainties provide a good estimate of the actual spread of the data. The overdispersion calculated for the entire CRONUS-A dataset

Lab or Sample	Number Analyses	Mean Conc (a/g) or (a)	Mean Conc Unc (a/g) or (a)	Mean Analytical Unc (%)	Standard Deviation (a/g) or (a)	Coefficient Variation (%)	Percent Deviation from Mean	Coeff Var Minus Mean Analytical Uncertainty (%)	Overdispersion (%)
CRONUS-A (10Be)									
1	3	3.46E+07	9.90E+05	2.86	6.03E+05	1.74	0.63	-1.12	0.0
4	3	3.45E+07	8.11E+05	2.35	1.00E+05	0.29	0.43	-2.06	0.0
8	5	3.47E+07	3.97E+05	1.14	1.95E+05	0.56	1.13	-0.58	0.0
10	6	3.43E+07	2.61E+05	0.76	6.50E+05	1.90	-0.29	1.14	1.4
11	4	3.37E+07	5.75E+05	1.71	1.01E+06	3.00	-1.90	1.29	0.0
Mean		3.44E+07	6.07E+05	1.77	5.12E+05	1.50	0.88	-0.27	0.2
Overdispersion of lab means (%)									
Pramontary Point, by sample (10Be)									
OSPPT-01	6	2.58E+05	5.27E+03	2.04	8.87E+03	3.44	0.30	1.40	1.9
OSPPT-02	6	2.50E+05	5.16E+03	2.07	9.75E+03	3.90	-2.92	1.83	3.2
OSPPT-03	7	2.55E+05	5.48E+03	2.15	1.04E+04	4.07	-1.13	1.92	3.0
OSPPT-04	7	2.59E+05	5.29E+03	2.04	6.79E+03	2.62	0.65	0.58	0.7
OSPPT-05	7	2.63E+05	5.64E+03	2.15	1.23E+04	4.68	1.97	2.53	3.6
OSPPT-08	6	2.60E+05	5.61E+03	2.15	1.12E+04	4.31	1.13	2.15	3.1
Mean		2.57E+05	5.41E+03	2.10	9.88E+03	3.84	1.35	1.74	2.6
Overdispersion of sample means (%)									
Pramontary Point, by laboratory (10Be)									
UW	8	2.59E+05	3.88E+03	1.50	1.00E+04	3.86	0.15	2.37	2.9
LEDO	19	2.55E+05	5.64E+03	2.21	1.02E+04	4.02	-1.56	1.80	3.2
CAMS	6	2.54E+05	4.93E+03	1.95	2.13E+03	0.84	-1.99	-1.11	0.0
UCB	6	2.68E+05	6.88E+03	2.57	1.11E+04	4.13	3.39	1.56	2.8
Mean		2.59E+05	5.33E+03	2.06	8.36E+03	3.21	1.77	1.16	2.2
Overdispersion of lab means (%)									
Huancane II, by sample (10Be)									
HU08-01	4	5.14E+05	1.04E+04	2.02	2.35E+04	4.58	0.54	2.56	3.5
HU08-02	4	5.14E+05	1.14E+04	2.23	2.14E+04	4.16	0.42	1.93	2.8
HU08-04	3	5.06E+05	8.44E+03	1.67	1.24E+04	2.44	-1.06	0.77	0.5
HU08-06	3	5.06E+05	8.93E+03	1.76	4.76E+03	0.94	-1.00	-0.82	0.0
HU08-10	3	5.17E+05	1.04E+04	2.02	1.60E+04	3.09	1.09	1.07	2.0
Mean		5.11E+05	9.92E+03	1.94	1.56E+04	3.04	0.82	1.10	1.8
Overdispersion of sample means (%)									
Huancane II, by laboratory (10Be)									
DM	9	5.25E+05	9.73E+03	1.85	1.24E+04	2.36	3.02	0.51	1.4
LEDO	6	4.99E+05	9.80E+03	1.96	5.85E+03	1.17	-2.20	-0.79	0.0
UW	10	5.06E+05	9.55E+03	1.89	1.20E+04	2.36	-0.82	0.48	1.3
Mean		5.10E+05	9.69E+03	1.90	1.01E+04	1.97	2.02	0.07	0.9
Overdispersion of lab means (%)									
Pramontary Point, by sample (26Al)									
OSPPT01	4	1.63E+06	8.20E+04	5.03	1.19E+05	7.31	-5.04	2.28	4.7
OSPPT02	4	1.63E+06	7.54E+04	4.61	2.28E+05	14.0	-4.78	9.35	12
OSPPT03	5	1.70E+06	7.38E+04	4.34	2.64E+05	15.5	-0.94	11.18	14
OSPPT04	4	1.67E+06	1.41E+05	8.44	1.82E+05	10.9	-2.44	2.41	3.6
OSPPT05	5	1.82E+06	7.18E+04	3.93	9.36E+04	5.13	6.32	1.20	0
OSPPT08	3	1.83E+06	7.97E+04	4.35	1.33E+05	7.28	6.88	2.93	3.1
Mean		1.72E+06	8.73E+04	5.12	1.70E+05	10.0	4.40	4.89	6.2
Overdispersion of sample means									
Pramontary Point, by laboratory (26Al)									
LEDO	5	1.48E+06	6.15E+04	4.15	1.91E+05	12.9	-11.7	8.8	11
UCB	12	1.76E+06	1.12E+05	6.39	1.57E+05	8.94	5.01	2.6	4.3
UW	8	1.79E+06	6.25E+04	3.50	4.07E+04	2.28	6.66	-1.2	0
Mean		1.68E+06	7.88E+04	4.68	1.30E+05	8.05	7.77	3.4	5
Overdispersion of laboratory means									
Tabernacle, by sample (36Cl)									
OSTA801	4	2.10E+04	7.93E+02	3.77	1.57E+03	7.46	6.14	3.68	5.7
OSTA802	1	1.95E+04	1.01E+03	5.18			-1.38		
OSTA803	2	1.97E+04	1.83E+03	9.28			-0.59		
OSTA804	2	1.87E+04	6.96E+02	3.73			-5.81		
OSTA805	5	1.95E+04	1.09E+03	5.60	1.91E+03	9.81	-1.42	4.21	4.9
OSTA806	5	1.99E+04	8.48E+02	4.26	1.83E+03	9.19	0.47	4.93	6.0
OSTA807	5	2.03E+04	1.15E+03	5.66	1.87E+03	9.19	2.60	3.53	4.1
Mean		1.98E+04	1.06E+03	5.35	1.79E+03	8.91	2.63	3.56	5.2
Overdispersion of sample means									
Tabernacle, by laboratory (36Cl)									
NMT	11	2.14E+04	9.98E+02	4.66	6.84E+02	3.19	7.07	-1.47	0
DAL	6	1.95E+04	1.20E+03	6.13	5.82E+02	2.98	-2.48	-3.15	0
PRIME	6	1.80E+04	8.66E+02	4.81	9.66E+02	5.37	-10.1	0.56	1.5
Mean		2.00E+04	1.02E+03	5.11	7.44E+02	3.85	6.54	-1.26	0.5
Overdispersion of lab means									
CRONUS-A (14C)									
1	7	7.17E+05	1.67E+04	2.33	3.59E+04	5.00	3.42	2.68	3.8
3	4	6.69E+05	1.25E+04	1.87	1.05E+04	1.57	-3.45	-0.29	0
4	8	6.93E+05	6.75E+03	0.97	4.00E+04	5.77	0.03	4.80	5.3
Mean		6.93E+05	1.20E+04	1.72	2.88E+04	4.12	2.30	2.39	3
Overdispersion of lab means									
CREU-1 (3He)									
Caltech	9	4.87E+09	1.37E+08	2.81	1.38E+08	2.83	-2.20	0.03	1.4
GFZ	7	4.93E+09	1.29E+08	2.61	1.15E+08	2.34	-1.10	-0.27	0.4
CRPG	10	4.93E+09	9.40E+07	1.91	5.79E+07	1.17	-1.01	-0.73	0.0
SUERC	17	5.02E+09	1.60E+08	3.19	1.77E+08	3.53	0.76	0.34	1.3
BGC	5	4.88E+09	1.60E+08	3.28	1.47E+08	3.01	-2.10	-0.27	0.0
LEDO	14	5.26E+09	9.00E+07	1.71	9.77E+07	1.86	5.65	0.15	0.4
Mean		4.98E+09	1.28E+08	2.58	1.22E+08	2.46	1.43	-0.13	0.6
Overdispersion of lab means (%)									
CREU-1 (21Ne) Coarse Fraction									
ETH	7	3.47E+08	3.85E+06	1.11	5.81E+06	1.67	0.90	0.56	0.0
BCG	9	3.68E+08	6.36E+06	1.73	7.60E+06	2.07	6.92	0.34	0.0
SUERC	14	3.42E+08	9.91E+06	2.90	1.92E+07	5.61	-0.58	2.72	5.1
GFZ	7	3.41E+08	8.00E+06	2.35	1.10E+07	3.22	-1.02	0.87	0.0
CRPG	2	3.23E+08	1.16E+07	3.60	6.80E+06	2.11	-6.22	-1.50	0.0
Mean		3.44E+08	7.95E+06	2.34	1.01E+07	2.94	3.13	0.60	1.0
Overdispersion of lab means (%)									

Table A1. Summary statistics grouped by laboratory (and by sample for field data sets). Note that the mean of the "Percent Deviation from Mean" is the mean of the absolute percent deviation. Standard deviations and derivative statistics are one sigma. Data are from same sources as Table 2. Laboratory ID's, in order of appearance, are as follows: UW =University of Washington, LDEO=Lamont-Doherty Earth Observatory, CAMS=Center for Accelerator Mass Spectrometry, Lawrence Livermore National Laboratory, UCB=University of California at Berkeley, DM= Dartmouth University, NMT=New Mexico Tech, Socorro, PRIME=Purdue Rare Isotope Measurement Laboratory, Caltech=California Institute of Technology, GFZ= Deutsches GeoForschungsZentrum, Potsdam, CRPG=Centre de Recherches Pétrographiques et Géochimiques, Nancy, SUERC=Scottish Universities Environmental Research Centre, Glasgow, BGC=Berkeley Geochronology Center, ETH=Eidgenössische Technische Hochschule, Zürich.

therefore arises from the data from the remaining eight laboratories. This result is perhaps not unexpected. Probably, the laboratories that returned multiple replicates of CRONUS-A were larger, better equipped, and more experienced than those that only returned one or two analyses. This inference emphasizes the importance of laboratory technique in obtaining reproducible data, but does not help in identifying the source of the overdispersion.

A.2 ^{10}Be – Promontory Point

The data from Promontory Point (Lifton et al., 2015) in Table 2 show significant overdispersion (3.0%) relative to the average analytical uncertainty (2.1%). Since Promontory Point is a field site, by-sample statistics are given first in Table A1. They show that the average absolute percent deviation from the mean, by sample, is relatively small (1.35%) compared to both the average analytical uncertainty (2.1%) and the average by-sample coefficient of variation (3.2%). The overdispersion of the sample means (0%) is consistent with this observation. We therefore find no evidence that the observed overdispersion of the entire data set is due to sample-to-sample heterogeneity.

When evaluated by laboratory, the average coefficient of variation is relatively large (3.2%) compared to the average analytical uncertainty (2.1%), and is also large compared to the average absolute deviation from the mean (1.8%). This implies that the overdispersion of the whole data set is predominately due to within-laboratory variation, presumably of a random nature. This inference is confirmed by comparison of the average within-laboratory overdispersion (2.2%) with the overdispersion of the laboratory means (0%).

A.3 ^{10}Be – Huancané

The overall data set from the Huancané II moraines in Peru (Phillips et al., 2015c) in Table 2 shows overdispersion (2.2%) comparable to the average analytical uncertainty (1.9%). Similar to the Bonneville by-sample statistics, the overdispersion of the by-sample mean concentrations is 0% and the average percent deviation from the mean is

small compared to the measures of within-sample spread of the data, indicating that the samples are homogeneous.

The result for the by-laboratory evaluation, however, is somewhat different than for Bonneville. For Huancané, the average absolute deviation from the mean (2.0%) is comparable to both the average by-laboratory coefficient of variation (2.0%) and the average analytical uncertainty (1.9%). The average overdispersion by laboratory (0.9%) is also similar to the overdispersion of the laboratory means (1.1%). We therefore conclude that the overdispersion in this dataset should be attributed both to random within-laboratory spread and to probably systematic between-laboratory bias. This is consistent with the evaluation of the site data by Phillips et al. (2015c).

A.4 ^{26}Al – Promontory Point

The overall ^{26}Al data set from Promontory Point (Lifton et al., 2015) in Table 2 exhibits large overdispersion (10%) compared to the average analytical uncertainty (5.0%). The by-sample statistics in Table A1 show a fairly large average absolute deviation from the mean (4.4%) compared to the average analytical uncertainty, but the average overdispersion by sample (6.2%) is significantly larger than the overdispersion of the sample means (2.2%). The strongest argument that the samples are homogeneous as an ensemble is the by-sample data reviewed above in section A.2 for ^{10}Be , which is much more precise than the ^{26}Al data. It seems quite unlikely that aliquots of the same samples are homogeneous in ^{10}Be but heterogeneous in ^{26}Al .

The by-laboratory data from Promontory Point (Table A1) show very large average absolute deviations from the mean (7.8%) and also average coefficient of variation (8.05%) compared to the average reported analytical uncertainty (4.7%). The average of the by-laboratory overdispersions is 5%, which is smaller than the overdispersion of the laboratory means, 7.6%. We conclude that the entire-data-set overdispersion results from a combination of within-laboratory excessive spread and between-laboratory bias, of roughly equal magnitude. At least part of the difference in the ^{26}Al and ^{10}Be overdispersion results from the relatively lower current (up to one order of magnitude) for Al than Be achieved by AMS facilities at the time these experiments were conducted.

A.5 ^{36}Cl – Whole Rock - Tabernacle Hill

As noted above, the reported target chemical compositions of the Tabernacle Hill samples, which were processed as whole-rock material with secondary minerals (principally pedogenic calcite and zeolites) removed, are somewhat heterogeneous. This was presumably due to differences in sample processing between laboratories. Inasmuch as ^{36}Cl is produced by three major reactions, ^{36}Cl concentrations alone cannot be used to compare the consistency of the data. In order to perform a comparison analogous to the other nuclides, the measured ^{36}Cl concentrations were divided by the calculated production rate. This yields nominal accumulation times in units of years, but these do not correspond to ages. To avoid using the value of $P_f(0)$ calculated from the Tabernacle Hill whole-rock samples, the average of the three previous $P_f(0)$ calibrations, by Evans et

al. (2001), Phillips et al. (2001), and Swanson and Caffee (2001) was used, giving a mean value of $709 \text{ n (g air)}^{-1} \text{ yr}^{-1}$. To explore the sensitivity of the variability of the normalized concentrations to the assumed value for $P_f(0)$, it was varied by $\pm 20\%$ around this mean value. The coefficient of variation of the reduced concentrations for all of the data, using the mean $P_f(0)$, was 8.0%; increasing or reducing $P_f(0)$ by 20% caused this coefficient of variation to decrease or increase by approximately 0.9%. The variability of the normalized concentrations is thus not particularly sensitive to the assumed value for $P_f(0)$.

Lifton et al. (2014) have noted that the aliquots from the Tabernacle Hill samples show a much larger range of ^{36}Cl and Cl concentrations than would be expected for homogenized materials, especially when compared between laboratories. The coefficient of variation of the non-normalized concentrations was 14% and laboratory means varied by as much as 24%. Dividing the concentrations by the sample-specific production rates reduces the coefficient of variation to 8.0% and the maximum interlaboratory difference to about 17%. This reduction is because ^{36}Cl and the Cl concentrations are covariant, causing the production rate to increase as the ^{36}Cl concentration increases. This result strongly suggests that at least part of the interlaboratory variation is caused by laboratory procedures that cause Cl-rich grains to be retained or lost in a fashion that varies systematically from laboratory to laboratory. This hypothesis has not yet been tested.

The overall normalized whole-rock ^{36}Cl data set from the Tabernacle Hill basalt flow (Lifton et al., 2015) in Table 2 exhibits overdispersion (6.2%) that is slightly larger than the average analytical uncertainty (5.1%). The by-sample evaluation in Table A1 shows an average absolute deviation from the mean of 2.6%, which is considerably smaller than either the coefficient of variation by sample (8.9%) or the average ‘age’ uncertainty. The average overdispersion by sample (5.2%) is larger than the overdispersion of the sample means (1.5%). All of these statistics indicate that the Tabernacle Hill samples were homogeneous as an ensemble, but that interlaboratory bias was the cause of most of the overdispersion. The homogeneity of the samples is strongly supported by the consistency of the ^3He data from olivines in the same samples (Goehring et al., 2010).

The average absolute deviation from the mean of the by-laboratory data was 6.5%, larger than either the coefficient of variation by laboratory (3.8%) or the average normalized concentration uncertainty (5.1%). The average overdispersion of the by-laboratory data was 0.5%, but that of the laboratory means was 5.7%. As above, we interpret these statistics to indicate that the main source of excess spread in the ^{36}Cl ages was interlaboratory bias. As described above, this bias probably originates from laboratory processing procedures that cause variable enrichment or loss of mineral phases containing either chlorine or other target elements.

A.6 ^{14}C – CRONUS-A

The overall ^{14}C data set on the CRONUS-A material (Jull et al., 2015) in Table 2 shows overdispersion (7.1%) that is much larger than the average reported analytical uncertainty (1.82%). The by-laboratory evaluation in Table A1 contains both a coefficient of variation (4.12%) and an average absolute deviation from the mean (2.3%) that are much larger than the average analytical uncertainty. The average overdispersion

of the by-laboratory data (3%) is larger than the overdispersion of the laboratory means (0%). In spite of the 0% overdispersion of the laboratory means, the large overdispersion of the entire data set appears to predominantly result from underestimation of within-laboratory random dispersion, but to also have a significant contribution from interlaboratory bias.

A.7 ^3He – CRONUS-P

The overall ^3He data set on the CRONUS-P ICM (Blard et al., 2015; Schaefer et al., 2015) in Table 2 shows overdispersion (3.1%) that is somewhat larger than the average analytical uncertainty (2.5%). The by-laboratory evaluation in Table A1 shows an average coefficient of variation (2.5%) that is comparable to the average by-laboratory analytical uncertainty (2.6%), but has in addition an average absolute deviation from the mean of 1.4%. The average of the by-laboratory overdispersion is 0.6% while the overdispersion of the laboratory means is 2.2%. This evaluation indicates that the major source of overdispersion is interlaboratory bias.

A.8 ^{21}Ne – CREU-1

The overall ^{21}Ne data set on the CRONUS-P ICM (Vermeesch et al., 2015) in Table 2 shows overdispersion (7.1%) that is large compared to the average analytical uncertainty (2.3%). The by-laboratory evaluation in Table A1 shows an average coefficient of variation (2.9%) that is comparable to the average by-laboratory analytical uncertainty (2.3%), but is smaller than the average absolute deviation from the mean of 3.1%. The average of the by-laboratory overdispersion (1.0%) is smaller than the overdispersion of the laboratory means (3.2%). We conclude that the overdispersion results from both within-laboratory random dispersion and inter-laboratory bias, but that the interlaboratory bias predominates.

APPENDIX B

Evaluation of Calibration Results for Nuclides Other than ^{10}Be

B.1 ^{26}Al calibration results

Aluminum-26 was calibrated using 53 samples from three primary sites: Huancané, Promontory Point, and Scotland. The calibration was tested against a secondary data set consisting of 30 samples from four sites: New England moraines, the ‘Sierra Nevada moraines’ data set, and the Baboon Lakes. The procedure for assigning analytical uncertainty was similar to that for ^{10}Be and was based on the analytical intercomparison study of Jull et al. (2015). Assigned uncertainties were interpolated

between 10.1% at 1.06×10^6 atoms/g and 4.9% at 145×10^6 atoms/g. The results are illustrated in Figure B-1 and statistically summarized in Table B1.

The ^{26}Al calibration results in Table B1 exhibit the same bimodal distribution of goodness of fit that the ^{10}Be data did. All of the neutron-monitor-based scaling results yielded reduced chi-squared values well above unity and p-values much smaller than 0.05, whereas the results from the other scaling methods show much better fits. The reasons for this distribution that were discussed in section 2.7.4.1 for ^{10}Be are equally applicable to these ^{26}Al results.

A major difference from the ^{10}Be calibration results is that all of the non-neutron-monitor-based methods gave goodness-of-fit test results that are statistically acceptable. As noted by Borchers et al. (2015), this is in part because of the smaller number of calibration sites (three sites instead of four) and the smaller number of laboratories contributing, which could lead to better precision at each site. However, the main factor is probably the much larger assigned analytical uncertainty (average of 10.2% compared to 4.2% for the primary ^{10}Be data set). (Note that these “assigned” analytical uncertainties are significantly larger than the analytical uncertainties provided by the AMS laboratories, as described above). This average uncertainty is slightly larger than the average of the standard deviations of ages from the four scaling models: 9.6%, in contrast to the case for ^{10}Be , where it was significantly smaller than the standard deviations. Both the standard deviation of the ages and the RMSE for the best-fitting model (LSD) are significantly smaller than the average analytical uncertainty assigned, suggesting that the analytical uncertainty may have been slightly overestimated.

Like the ^{10}Be data sets, both the primary and secondary ^{26}Al data sets exhibit a bias toward old ages. The bias averages 3% for the primary data set and 4% for the secondary one, which is similar to that from the primary ^{10}Be data set, but smaller than that from the ^{10}Be secondary data set. The reasons for the bias are presumably similar to those for the ^{10}Be data set, discussed above in the main text.

The RMSE’s of the ages from the secondary data set, averaging 11.5%, are generally consistent with the assigned uncertainties for the primary data set and are significantly larger than the average of the actual laboratory analytical uncertainties of the data set: 5.0%. These RMSE’s were empirically used to estimate the uncertainties for the ^{26}Al production parameters. The RMSE’s are similar to those from the ^{10}Be secondary data set (which averaged 10.8%), but are somewhat larger. The difference appears to be mainly attributable to poorer analytical precision of the ^{26}Al analyses.

B.2 ^{36}Cl calibration results

The calibration of ^{36}Cl production rates poses special challenges. Chlorine-36 has three major production reactions: spallation of Ca and K and low-energy neutron absorption by Cl. In addition, it is also produced in smaller amounts by spallation of Ti and Fe and by muon-induced reactions with Ca, K, and Cl. The CRONUS-Earth calibration of ^{36}Cl production is detailed in Marrero et al. (2015c).

for the minor production reactions involving Ti and Fe. Data from previously published depth profiles were used to calibrate the muogenic production reactions.

Due to the large number of production reactions, the primary data set was insufficient to adequately test the production systematics on a global basis. The burden of proof therefore had to be laid on the secondary data set. Fortunately, a large number of samples were available for this purpose. The secondary data set contained 85 samples, making it the largest of the CRONUS-Earth secondary data sets. The localities included the Littleton-Bethlehem moraines in New England, USA, the Baboon Lakes, basalt flows from the Craters of the Moon National Monument in Idaho, USA, Elsmere Island, Canada, a basalt flow in New Mexico, USA, the Sierra Nevada, USA, and Scotland. These are described in more detail by Marrero et al. (2015c).

In the calibration (Borchers et al., 2015; Marrero et al., 2015c), production by spallation of Ca was constrained principally by Ca-feldspar separates from the Tabernacle Hill basalt flow, that by spallation of K by K-feldspar separates from Huanacané and Scotland, and the production parameter for low-energy neutron absorption by biotite separates from Baboon Lakes and whole-rock samples from the Tabernacle Hill basalt flow. The results are given in Table B2 and Figure B-2. The calibration produced reduced chi-squared values close to unity and p-values well above 0.05 (Borchers et al., 2015). In spite of this favorable result, the calibrated parameters cannot be considered to have been statistically validated because the number of sample sites (one for Ca spallation and two for the other two reactions) is too small to provide an adequate test of the model.

Figure B-2. Normalized ^{36}Cl ages of the primary and secondary data sets as a function of percent of ^{36}Cl production from low-energy neutron absorption by Cl.

Scaling	Ca at ^{36}Cl (g Ca) $^{-1}$ yr $^{-1}$		K at ^{36}Cl (g K) $^{-1}$ yr $^{-1}$		Pf(0) n (g air) $^{-1}$ yr $^{-1}$	
LSDn (Sa)	56.0	4.3	157	12	704	176
LSD (Sf)	56.6	4.7	155	13	690	181
St	52.5	3.6	151	10	649	201
Lm	52.0	3.2	152	9	646	192
Li	60.5	12.7	143	30	721	226
Du	55.4	8.3	129	19	666	204
De	56.0	11.9	129	27	668	212

Table B2. Production parameters and associated uncertainties for the major ^{36}Cl production reactions. Production rates are from Marrero et al. (2015c), uncertainties from this study.

Element	Ca		K	
Production Parameter	σ_o	f*	σ_o	f*
Scaling	(10^{-30} cm 2) $^{\checkmark}$	(10^2)	(10^{-30} cm 2) $^{\checkmark}$	(10^2)
LSDn (Sa)	8.23	1.34	9.33	5.76
LSD (Sf)	8.25	1.34	935.00	5.75
St	8.24	1.34	9.20	5.85
Lm	8.35	1.40	8.81	6.16
Li	8.27	1.35	10.06	5.24
Du	8.32	1.38	9.89	5.37
De	8.25	1.34	9.97	5.30

Table B3. Muogenic production parameters for ^{36}Cl , from Marrero et al. (2015c).

Scaling Method	Lifton-Sato Nucl-Dep (LSDn or Sa)	Lifton-Sato Flux-Based (LSD or Sf)	Lal-Stone (St)	Lal 1991 Nish 1989 (Lm)	Lifton (Li)	Dunai (Du)	Desilets (De)
Shorthand Notation							
Secondary Data Set							
Root Mean Square Error (by sample; %)	9.4 (17.8)	9.9 (18.2)	9.0 (20.2)	8.5 (20.4)	21.3 (22.2)	16.0 (21.3)	22.8 (24.5)
Ave Absolute Age Error (by sample)	0.078 (0.133)	0.088 (0.137)	0.078 (0.147)	0.070 (0.145)	0.162 (0.168)	0.102 (0.151)	0.182 (0.180)
Average Global Norm Age (by sample)	1.032 (0.991)	1.036 (0.992)	1.014 (1.023)	1.012 (1.016)	1.148 (1.048)	0.982 (1.010)	1.178 (1.068)
St Dev Global Norm Age (by sample)	0.092 (0.180)	0.095 (0.184)	0.090 (0.204)	0.086 (0.206)	0.151 (0.220)	0.164 (0.215)	0.152 (0.238)
Average Norm Analytical Unc	0.054 (0.085)	0.054 (0.085)	0.054 (0.085)	0.054 (0.085)	0.054 (0.085)	0.054 (0.085)	0.054 (0.085)

Table B4. Statistical summary for the secondary ^{36}Cl data set. Values are in terms of normalized age and are unitless. Values in each column without parentheses are for samples with >90% of production from spallation reactions. Those in parentheses are for the entire data set.

Muogenic production was incorporated into our calibration (Marrero et al., 2015c) using data from the previously published depth-profile studies of Stone et al. (1998) in limestone and Evans (2001) in granite. The calibration produced good fits to the measured profiles. As was the case for the ^{10}Be and ^{26}Al muogenic production calibrations, the field-calibrated parameters for production of ^{36}Cl off of Ca were smaller by about 50% than those published by Heisinger et al. (2002a and b), based on laboratory experiments. The reason for the discrepancy has not been established, but we prefer the field-based values. The production parameters are given in Table B3.

Spallation of Ti and Fe produces small amounts of ^{36}Cl in many rocks (the production from these reactions represented 1-2% of the total ^{36}Cl production for the secondary data set). The SLHL value of $1.9 \text{ atoms (g Fe)}^{-1} \text{ yr}^{-1}$ from Stone (2005) was accepted for production from iron. This value was multiplied by the theoretically modeled Ti/Fe production ratio of 2 from Masarik (2002) to obtain the Ti-spallation production rate: $3.8 \text{ atoms (g Ti)}^{-1} \text{ yr}^{-1}$.

Table B4 presents the statistical evaluation of the results from the ^{36}Cl secondary data set. We do not give the results from the primary data set (which can be found in Borchers et al., 2015) because, as explained above, the numbers of sites are too small for the statistics to be meaningful. In Table B4 statistics are given for both the subset of the samples having >90% of the production by spallation and for the entire data set (in parentheses), which includes samples having up to 100% of the production by low-energy neutron absorption by Cl. Consistent with the other nuclides discussed above, the statistical fits for the neutron-monitor-based scaling methods are much worse than those for the remaining four scaling methods and the discussion below will be restricted to these four methods.

We will discuss the samples in which production is dominated by spallation reactions (>90%) first, then those in which low-energy neutron absorption is important. The age RMSE for the spallation-dominated samples varies from 8.5 to 9.9%. This compares favorably to the RMSE's for the ^{10}Be secondary data set, by sample (10.5 to 11.1%). The average normalized age (by sample) ranges from 1.014 to 1.032. The magnitude of bias is much smaller (by a factor of two to three) than for the ^{10}Be data set. One possible reason for this is that the ^{36}Cl primary data set did not contain samples from New Zealand, as did the ^{10}Be primary data set, and if there is a systematic anomaly involving late-glacial age samples from the North Atlantic area, compared to ones from New Zealand and South America, perhaps lack of this locality in the primary data set enabled a better scaling for the locations in the ^{36}Cl secondary data set. The standard deviation of the normalized sample ages ranges from 0.086 to 0.095. This is very similar to the range of RMSE's discussed above, indicating that most of the spread of the data is due to random variability. This inference is consistent with the small magnitude of the bias. However, the magnitude of the standard deviation of the ages is significantly larger than the average analytical uncertainty, which is 5.4%. Data from the field intercomparison at Tabernacle Hill, discussed above in Appendix A.5, indicate that interlaboratory bias was a significant contributor to the spread of the data, thus the difference between the analytical uncertainty and the standard deviation of the ages can most likely be attributed to this source.

The flux of high-energy neutrons responsible for spallation reactions is very largely insensitive to environmental conditions. This is not the case for the thermal and

epithermal neutron fluxes responsible for production of ^{36}Cl by low-energy neutron absorption. These are highly sensitive to the amount of hydrogen in the environment (Gosse and Phillips, 2001). Although the mean transport distance of low-energy neutrons in rocks is generally less than a meter, in the air it is hundreds of meters (Zweck et al., 2013). Thus vegetation, moss, or bodies of water or snow hundreds of meters distant can affect the thermal-neutron flux passing through a rock sample. Snow, in particular, can cause major perturbations in the low-energy neutron flux by neutron reflection off of the hydrogen-rich layer (Dunai et al., 2014). These effects have been both experimentally observed and modeled to produce order-of-magnitude variations in the low-energy neutron fluxes (Dunai et al., 2014; Zweck et al., 2013). An empirical examination of the sensitivity of ^{36}Cl ages to Cl content is thus highly desirable.

In Figure B-2 the normalized ^{36}Cl ages are plotted as a function of the percent of the production that is from low-energy neutron absorption by Cl (i.e., not from spallation). The sample with close to 100% of production by low-energy neutron absorption clearly show significantly greater spread than do those with near-zero production from low-energy absorption, but the amount of spread does not increase dramatically as the low-energy production increases. In Figure B-3 the RMSE of the normalized ages is plotted as a function of the proportion of low-energy production. The RMSE increases in a fairly linear fashion from 9.4% at 5.1% low-energy production to 29.8% at 98.9% low-energy production, equivalent to a factor of about three over the entire production range. Although this is obviously quite significant, it does indicate that time-integrated ^{36}Cl production from absorption of low-energy neutrons varies much less than the order-of-magnitude, or more, short-time fluctuations seen in experiments and models.

Figure B-3. RMSE of normalized ^{36}Cl ages from the secondary data set as a function of the percent of ^{36}Cl production from low-energy neutron absorption by Cl. The dotted lines are the limits of the percent-production bins. The data points are plotted at the mean percent production for the bin. The numbers in parentheses are numbers of samples in each bin.

The statistics of the entire secondary data set, presented in Table B4, are similar in pattern to those for the low-chlorine samples, except that the random component of variability is much larger (the RMSE's vary from 17.8 to 20.4%). These entire-data-set RMSE's are strongly influenced by the large scatter of the ~100% Cl-production samples that were produced by leaching feldspars. If these are excluded and the RMSE is computed only using samples with 0 to 85% production from Cl, the average for the four non-neutron-monitor-based scalings is 15.3%. The bias evidenced by the average normalized age is even smaller than for just the low-Cl samples. We attribute the larger measures of spread (RMSE, average absolute age error, standard deviation) to variability in the production of ^{36}Cl by low-energy neutron absorption caused by variability in the hydrogen content of the surrounding environment. We therefore extrapolated the regression line in Figure B-3 to 0% production from Cl and to 100% production from Cl and used the first value to estimate the uncertainty associated with the spallation-production parameters and the second one to estimate the uncertainty associated with the Pf(0) parameter. The relative magnitudes of these came out about 7% and about 25%, respectively. These uncertainties will be reflected in the final ages in proportion to the relative amounts of spallation and low-energy absorption production.

One implication of these results is that it is clearly desirable to seek out low-chlorine samples in order to minimize the age uncertainties, usually accomplished by processing mineral separates. However, the advantage in this selection should not be exaggerated. The data in the secondary data set having percentages of production from Cl greater than 90% were produced by finely grinding and leaching minerals, specifically for the purpose of calibrating the Pf(0) parameter. Such samples would not normally be processed for a geochronological study. If these samples are excluded, the average percent production from Cl for the data set is 30%. This is probably fairly typical for whole-rock ^{36}Cl samples in general. The age uncertainty for a whole-rock sample having 30% of production from low-energy neutron absorption by Cl would be about 13%, compared to about 9% for a mineral separate having 0% production from Cl. For certain geochronological applications the difference between 13% and 9% may be critical, but for many others it may not.

B.3 ^3He calibration results

Sites selected for the primary data set were from Iceland by Licciardi et al. (2006), from Argentina by Ackert et al. (2003), and from Oregon and Idaho in the western United States from Cerling and Craig (1994). Twenty-one of the 62 primary calibration samples were from Tabernacle Hill, published in three separate studies. The calibration was tested against a secondary data set consisting of 40 samples from 13 sites, including the Canary Islands, Mount Etna in Italy, Hawaii, and several lava flows in the western United States. Stated laboratory uncertainties were used for the calibration

because at the time that the calibration was performed the CRONUS-P intercomparison results were not available. The results are illustrated in Figure B-4 and statistically summarized in Table B5.

Like the other nuclides, with the exception of ^{26}Al , the calibration did not produce a statistically justified result, yielding reduced chi-squared values of about four and p-values below 10^{-14} . The reason for this result appears to mainly be the discrepancy between the stated laboratory uncertainties (average of 4.1%) and the RMSE of the calibrated ^3He ages (averaging about 10%). Although the ages exhibit a moderate degree of bias (about 3%), the large standard deviation of the normalized ages (9 to 10%) indicates that random variability of the data is the most important contributor to the RMSE. We note that, although the CRONUS-P intercomparison results (discussed above in 2.2.1) from Blard et al. (2015) were not available, their inclusion would have made little difference to the statistical validity of the outcome. The average reported analytical uncertainty for that material was about 2.5% and the intercomparison showed additional spread of about 3% that was attributed mainly to interlaboratory bias. Even so, increasing the analytical uncertainty from the 4% average used in the calibration to the ~5% indicated by the intercomparison would still have left it well below the standard deviation of the ^3He ages of 9 to 10% and thus would still not have produced a statistically consistent calibration.

Figure B-4. Normalized ages for the primary and secondary ^3He data sets.

<i>Scaling Method</i>	Lifton-Sato Nucl-Dep (LSDn or Sa)	Lifton-Sato Flux-Based (LSD or Sf)	Lal-Stone (St)	Lal 1991 Nish 1989 (Lm)	Lifton (Li)	Dunai (Du)	Desilets (De)
<i>Shorthand Notation</i>							
Production Rate (atoms g ⁻¹ yr ⁻¹ at SLHL)	115	119	118	117	131	123	123
Production Rate Uncertainty	19	19	18	13	19	14	17
<i>Primary Data Set</i>							
Reduced Chi-Squared	4.37	4.38	4.58	4.39	4.27	4.09	4.29
p-value	<1 x 10 ⁻¹⁴	<1 x 10 ⁻¹⁴	<1 x 10 ⁻¹⁴	<1 x 10 ⁻¹⁴	<1 x 10 ⁻¹⁴	<1 x 10 ⁻¹⁴	<1 x 10 ⁻¹⁴
Root Mean Square Error (by sample; %)	10.1	10.2	10.5	9.9	9.7	9.6	9.8
Ave Absolute Age Error (by sample)	0.080	0.084	0.080	0.076	0.076	0.074	0.076
Average Global Norm Age (by sample)	1.033	1.031	1.028	1.027	1.029	1.027	1.030
St Dev Global Norm Age (by sample)	0.095	0.097	0.101	0.095	0.092	0.091	0.092
Average Fractional Analytical Unc	0.041	0.041	0.041	0.041	0.041	0.041	0.041
<i>Secondary Data Set</i>							
Root Mean Square Error (by sample; %)	22.9	23.0	18.7	16.7	19.9	17.9	19.3
Ave Absolute Age Error (by sample)	0.157	0.156	0.152	0.138	0.147	0.146	0.145
Average Global Norm Age (by sample)	1.039	1.056	0.948	0.966	1.014	0.996	1.003
St Dev Global Norm Age (by sample)	0.223	0.220	0.178	0.162	0.196	0.177	0.191
Average Fractional Analytical Unc	0.075	0.075	0.075	0.075	0.075	0.075	0.075

Table B5. Helium-3 production rates and statistical evaluation of primary and secondary data sets. Production rates are from Borchers et al. (2015) and uncertainties from this study.

The most notable result of the ³He calibration was that, unlike all of the other nuclides discussed above, the neutron-monitor-based scaling methods and those not based on neutron-monitor data both produced about the same match to the independent ages (reduced chi-squared values of about four, RMSE's of about 10%, and biases of about 3%). The cause of this difference cannot uniquely be determined, but it likely is related to the elevation distribution of the primary-calibration samples. These range between 27 and 1461 m. In comparison, the ¹⁰Be data ranged between 86 and 4871 m. The ²⁶Al and ³⁶Cl calibration data had similar elevation ranges. Additionally, the highest elevation samples in the ¹⁰Be, ²⁶Al, and ³⁶Cl data sets were from low latitude where the cutoff rigidity is high, whereas the higher elevation ³He samples were from mid-latitude. The pronounced difference between the neutron-monitor-based scaling methods and the others is at high elevation and low latitude where the effects of multiplicity are large. Within the relatively narrow and low elevation range of these samples, it is not surprising that both methods give similar fits to the data (cf. Lifton et al., 2014).

Another systematic difference between the ³He calibration and those of the other nuclides was the much larger RMSE calculated from the ages of the secondary data set. For the other nuclides the RMSE of the secondary data set averaged 1 to 2% larger than that of the primary data set and was in the range of 9 to 12%. For ³He, in contrast, the RMSE of the secondary data set is 7 to 13% larger than that of the primary data set and averages about 20%. This difference cannot be attributed to bias in the secondary data set; the average bias of both the primary and secondary is about 1%. Rather, the difference is principally due simply to increased scatter in the secondary data set (see Figure B-4). We feel that the anomalously large RMSE of the secondary data set is probably not representative of the capability of the method. This is evidenced by the results from the very large (60 sample) primary data set, which yielded an RMSE similar to that of the other nuclides. Rather, it can probably be attributed to less-than-optimal strategy in the selection of the primary and secondary data sets.

B.4 ¹⁴C calibration results

Sites selected for the primary ^{14}C data set were Promontory Point (Lifton et al., 2015), Western Scotland (Marrero et al., 2015d), and the Macaulay Valley in New Zealand (Putnam et al., 2010). Data from sites with saturated samples from the western United States, northern Chile, and Antarctica were also included (Borchers et al., 2015). No secondary data set was available due to lack of samples. The results are shown in Table B6 and are graphed in Figure B-5.

As described above, production rates were estimated based on matching calculated to measured ^{14}C concentrations. For this purpose the saturated samples could be used. However, production-rate uncertainties were estimated based on matching the RMSE's of the calculated ages to the average age uncertainty. For this, the saturated samples could not be used because calculated age uncertainties 'blow up' as saturation is approached. In addition, four analyses from the Promontory Point data set (the third replicate of sample 01, the first replicate of sample 03, the second replicate of 05, and 08; see Figure B-5) were omitted from the production-rate uncertainty estimation because, for samples of this age (~3 half lives), slightly higher ^{14}C concentrations for these samples (by about 15%) caused the age uncertainties to increase in a nonlinear fashion and strongly biased the uncertainty estimation.

<i>Scaling Method</i>	Lifton-Sato Nucl-Dep (LSDn or Sa)	Lifton-Sato Flux-Based (LSD or Sf)	Lal-Stone (St)	Lal 1991 Nish 1989 (Lm)	Lifton (Li)	Dunai (Du)	Desilets (De)
<i>Shorthand Notation</i>							
Production Rate (atoms $\text{g}^{-1} \text{yr}^{-1}$ at SLHL)	12.76	12.72	12.24	12.22	13.42	12.44	12.49
Production Rate Uncertainty	± 0	± 0	± 0	± 0	± 0.4	± 1.2	± 0
<i>Primary Data Set</i>							
Reduced Chi-Squared	2.13	2.07	2.14	2.14	2.18	2.25	2.31
p-value	3.7×10^{-9}	1.4×10^{-8}	2.7×10^{-9}	2.6×10^{-9}	1.0×10^{-9}	1.9×10^{-10}	3.8×10^{-11}
Root Mean Square Error (by sample; %)	19.7	19.4	18.8	18.4	24.6	29.5	18.5
Ave Absolute Age Error (by sample)	0.169	0.165	0.162	0.159	0.183	0.202	0.153
Average Global Norm Age (by sample)	0.873	0.890	0.909	0.889	0.923	0.913	0.958
St Dev Global Norm Age (by sample)	0.168	0.173	0.210	0.180	0.252	0.304	0.194
Average Norm Analytical Unc	0.076	0.102	0.102	0.102	0.102	0.102	0.102

Table B6. Carbon-14 production rates and statistical evaluation of primary data set. Production rates, chi-squared, and p-values are from Borchers et al. (2015), uncertainties are from this study.

The production rates have very low uncertainties (zero, or near-zero) because the ages of the calibration data set were all relatively old (two or three half-lives of ^{14}C). The age dispersion increases nonlinearly with increasing age (i.e., increases more slowly than linear), and the age distribution of the calibration samples calculated using only the propagated uncertainties of the analyses and field parameters matched the observed spread of ages. These production-rate uncertainties may underestimate the actual uncertainty of samples that have exposure ages of one half-life of ^{14}C , or less.

Figure B-5. Normalized ages for the primary ^{14}C data set. The four samples with very large uncertainties were not used in estimating the production-rate uncertainty.

Like the other nuclides, ^{14}C did not achieve a statistically valid calibration. This is largely due to the spread of the concentrations or ages (RMSE's of about 14% for concentration and 19% for age), which is much greater than the average analytical uncertainty of about 9%. The lack of valid calibration is unlikely to be due to bias in the calculated concentrations, since the average normalized concentration was close to unity (1.009). The average normalized age in Table B6 does show significant young bias, but this is due to the deletion of the four samples with large uncertainty; with them included the average normalized age is also close to unity.

Like ^3He , and unlike the remaining nuclides, the ^{14}C gave fits to the neutron-monitor-bases scaling methods that were nearly as good as it did to the others. We attribute this to the same cause as the ^3He primary data set: a limited range of elevations. The elevations of the samples with finite ages ranged between 1800 and 100 m.

Lacking a secondary data set, the uncertainty estimation had to be performed using the primary calibration data set. This may cause the production-rate uncertainty to be slightly underestimated. The uncertainties of the secondary data sets for the other nuclides, except ^3He , were 1 to 2% greater than for the primary ones. The lack of an adequate secondary data set is due to the relative 'youth' of the technique; as it matures and additional data are obtained, the estimates of both the production rate and its uncertainty can be refined.