







WHEN IS ENOUGH, ENOUGH? QUESTIONS OF SAMPLING IN VERTEBRATE ICHNOLOGY

by MATTEO BELVEDERE^{1,2} , MARCIN BUDKA¹ ,
ASHLEIGH L. A. WISEMAN³  and MATTHEW R. BENNETT¹ 

¹Institute for Studies in Landscapes & Human Evolution, Bournemouth University, Poole BH12 5BB, UK; matteo.belvedere@hotmail.com

²Dipartimento di Scienze della Terra, Università degli Studi di Firenze, via G. La Pira 4, I-50121 Firenze, Italy

³Structure & Motion Laboratory, Department of Comparative Biology, The Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield, Hertfordshire AL9 7TA, UK

Typescript received 19 September 2020; accepted in revised form 7 June 2021

Abstract: Sample size is a challenge for most field scientists determined not by the statistically ideal, but by the available. In vertebrate ichnology, track length is an important variable correlating well with the track-maker's biology. It is also key to estimating the minimum number of individuals (MNI) present on a trampled horizon. Broad assumptions on biometrics of the track-makers are often made based on a few prints without consideration for intra-trackway variability. In this study we use a simple bootstrapping algorithm to explore variance within sample size for a range of trackways with fossil and experimental examples to determine the minimum sample size required to extract linear measurements. Predictably, experimentation shows that inter-step variability changes with track-maker and substrate, but the degree of variance is not as marked as previously anticipated. Change-point modelling suggests that a

maximum sample size of 22–25 captures most of the variance present in track length at least; another threshold at 7–10 has been identified, which represents the reasonable sample size minimum. Samples of fewer than seven tracks are subject to large amounts of potential variance and are unlikely to provide reliable and consistent measurements. These sampling thresholds hold across a wide range of depositional environments and track-makers. We calculate generic standard errors for human track-makers which may assist the practitioner with small samples to estimate the likely errors, especially when making MNI estimates. The challenge is placed to the wider vertebrate ichnology to explore this issue for other track-makers and develop similar guidance.

Key words: vertebrate ichnology, measurements, reliability, confidence, standard error, sample size.

GEOLOGISTS, palaeontologists, archaeologists and bioanthropologists are pragmatic folk and have to work with what they discover even if it is never enough! This is especially true in vertebrate ichnology. Vertebrate tracks of all types and ages occur widely in the geological record from Middle Devonian (Stössel 1995; Niedźwiedzki *et al.* 2010; Stössel *et al.* 2016) to the near-present (e.g. Avanzini *et al.* 2011). Perhaps the biggest growth area in terms of discoveries has been with human tracks which were once considered a freak act of geological preservation (e.g. Leakey & Hay 1979), but a spate of recent discoveries has shown that this is far from true (e.g. Morse *et al.* 2013; Helm *et al.* 2018; Duveau *et al.* 2019; Bennett *et al.* 2020; Hatala *et al.* 2020). Human tracks allow us to make inferences about: occurrence (e.g. Morse *et al.* 2013; Bennett & Morse 2014; Altamura *et al.* 2018; Helm *et al.* 2018, 2020), biomechanics (e.g. Hatala *et al.* 2013, 2016; McClymont *et al.* 2016; Raichlen & Gordon 2017), stature and body mass (e.g. Dingwall *et al.* 2013; Domjanic *et al.* 2015) and, ultimately, behaviour and group demographics

(e.g. Roach *et al.* 2016; Hatala *et al.* 2017). This also applies more widely to any other vertebrate tracks (e.g. Thulborn 1990), and relies on three things: the relationship of track depth to plantar pressure (e.g. Bates *et al.* 2013); the accuracy with which a track outline characterizes the foot of the track-maker in terms of size and shape (e.g. Gatesy & Falkingham 2017; Marchetti *et al.* 2019, 2020; Falkingham & Gatesy 2020); and finding evidence for the contemporaneity of interacting tracks. While the latter can be achieved with one or two cross-cut tracks, the former relies on the size of the sample of tracks, the variability within that sample and its representation of the foot, biomechanics, or behaviour of the track-maker.

The degree to which an individual track represents the shape of the foot and its biomechanical function is determined by the inter-step variability in footfall, variability in substrate in the direction of travel, and track taphonomy. There is also a component in any sample of inter-step measurement precision. The biomechanics of each

step has many moving parts as demonstrated by human biomechanics (e.g. Elftman & Manter 1935; Ker *et al.* 1987; Harcourt-Smith & Aiello 2004; Caravaggi *et al.* 2009) and variation in any one may be manifest in changes in plantar pressure and therefore in the distribution of track depth and maximal shape. No footprint is identical to another and will vary within a morphological envelope (e.g. Morse *et al.* 2013). The problem for ichnologists is that they rarely sample the full range of this morphological envelope due to being limited by the number of tracks that their excavation reveals or preserves, or in some cases the extent to which they are permitted to excavate (Fig. 1). This puts an intrinsic limit on the reliability of inferences such as the biometrics or biomechanics of the track-maker. Even if the structure of the track-maker is known or easily approximable, the older literature on human ichnology is littered with dubious assertions about the characteristics of track-makers made on single tracks (e.g. Roberts *et al.* 1996; Roberts & Berger 1997). This is particularly true for stature, body mass and estimates of minimum number of track-makers (MNT or MNI). Webb *et al.* (2014) used an interesting approach in which they determine MNI based on:

$$\text{MNI} = \frac{\text{Length Range}}{\sigma \cdot \text{CI} \cdot 2} \quad (1)$$

where the Length Range is the total range in footprint lengths, σ is the standard deviation and CI represents the confidence limit being used, typically 95% which would correspond to a value of 1.96 here. They determined σ from modern analogue studies. MNI estimates are also made by reference to determining length $\pm 5\%$ of the mean; if the two tracks exceed this ‘magic’ 5% then they are deemed to belong to two individuals. In practice this should be the 95% standard error (SE) of the mean, but an ichnologist is rarely able to sample the true variability of a population due to issues of preservation or exposure. Recourse to ‘typical’ SE using both modern analogue data and data from long fossil trackways, may help to mitigate these errors and provide better guidance to the ichnologist. We aim here to explore this variability and provide such guidance. To the trained statistician this may all seem obvious, embedded in the properties of the normal distribution, but we believe it is a timely and important reminder for field scientists who usually have to work with what they have.

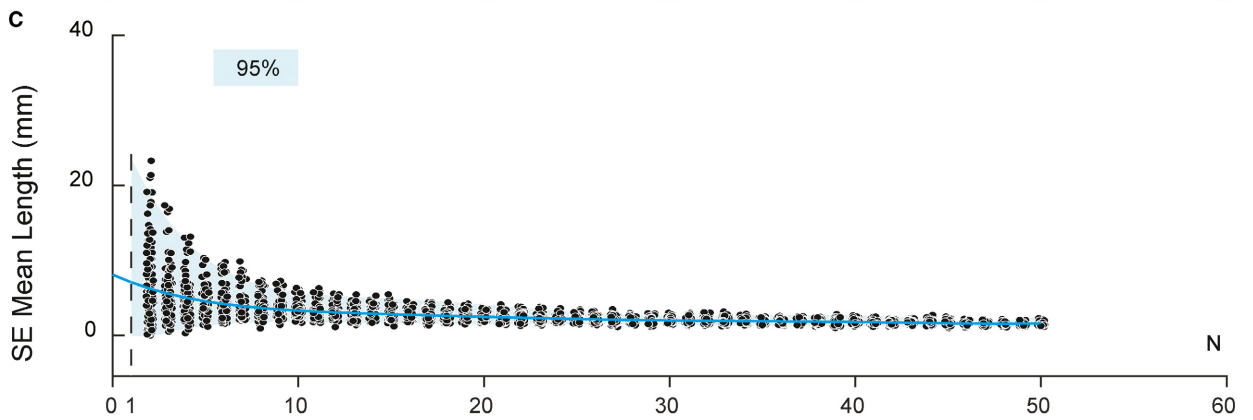
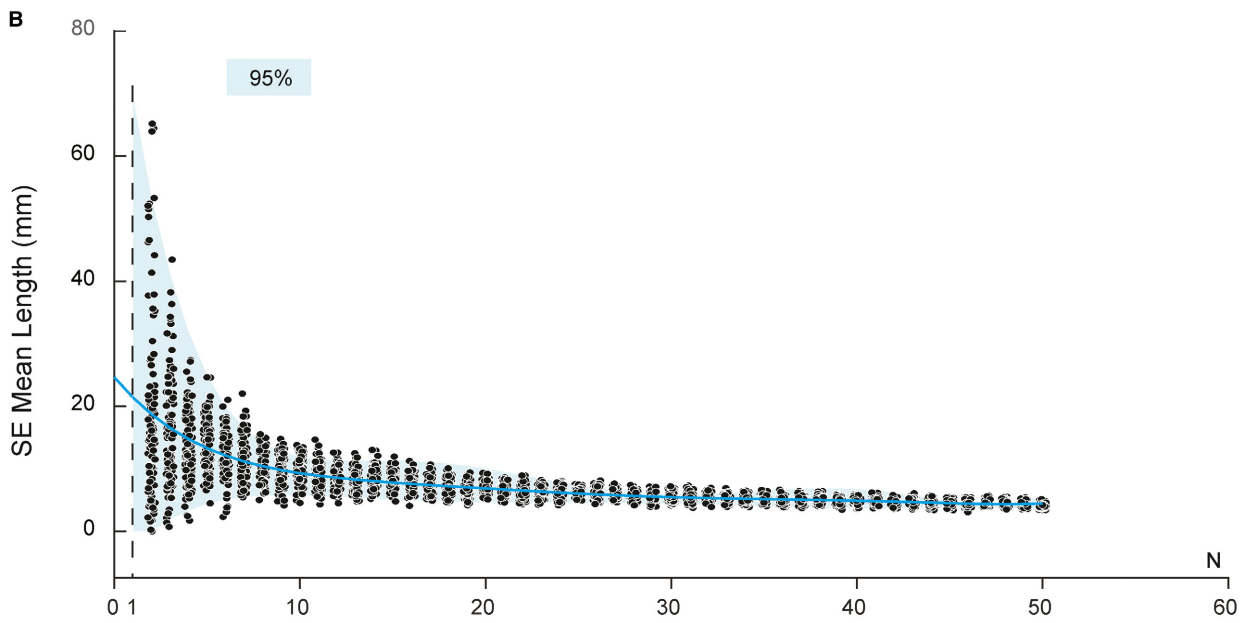
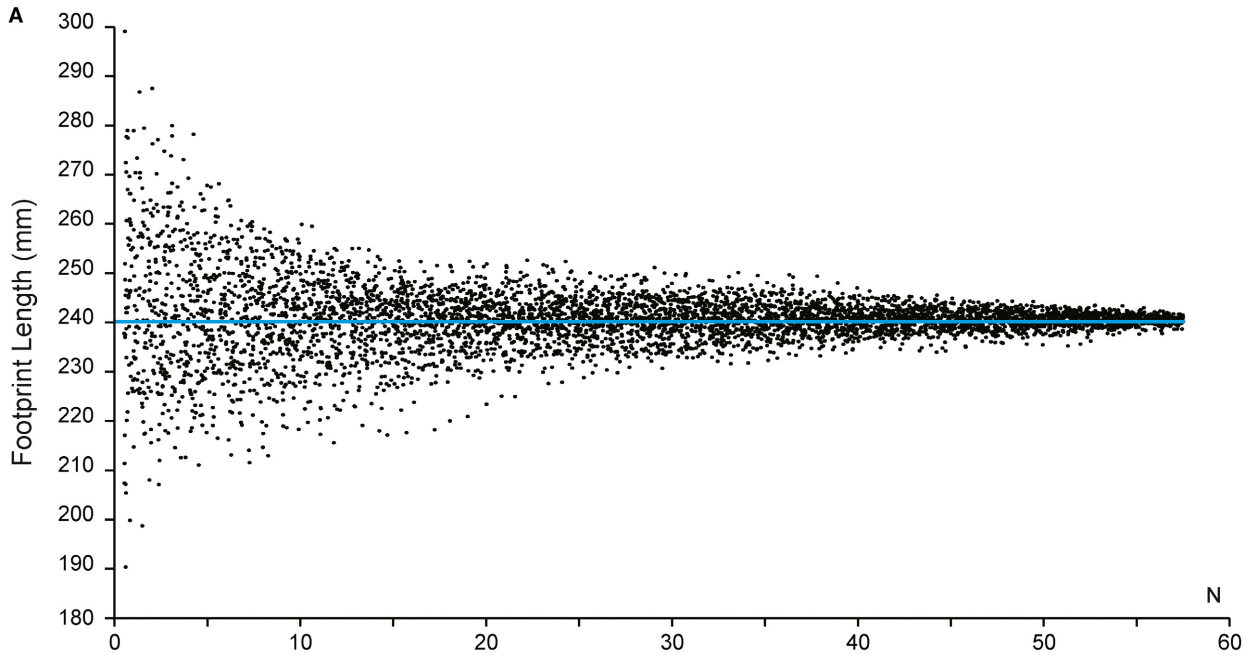
METHOD

For the purposes of this paper, we restrict ourselves to simple linear distances, such as track length (e.g. Wiseman & De Groote 2018; Wiseman *et al.* 2020), which is commonly used to infer the biometrics like stature and body mass of the track-maker or make MNI estimates (e.g. Thulborn 1990; Bennett & Morse 2014). *Median* is probably the best value to describe track variance, however most investigators measure, and above all report in publication, *mean* values. For this reason, means have been used throughout the dataset analysed here. We do, however, illustrate the difference in minimum estimates of sample size between median and mean for one of the datasets. For a long, multi-track human-trackway we estimate the SE around the mean (or median) for different sample sizes using the following analytical procedure:

1. Generate 100 bootstrap samples (with replacement) for each value of N in the range between 2 and 50. The upper bound is limited by the maximum sample size.
2. Calculate the SE for each bootstrap sample and each value of N.
3. Derive the mean and standard deviation from SE values for each value of N.
4. Fit polynomial curves to the mean and 95% confidence interval (CI) boundaries of the normal distributions calculated in Step 3 above. The CI values were clipped at 0 prior to curve fitting.
5. Estimate the SE and its CIs for all values of N.

Figure 1 show typical output for a couple of long human trackways. As one would expect, SE declines with increasing sample size. The challenge with such smooth curves is to determine point(s) at which further increases in sample size give marginal improvements in SE. Two approaches were used here. The first involved computing linear regressions and associate R^2 values for the whole sample and then progressively for $N-1$ until a minimum of $N = 5$ was reached. R^2 values improve with increasing sample size as the data tail becomes more linear and flatter. This provides an alternative way of visualizing the change in variance with sample size but does not identify any particular breaks in slope. The other approach involves using changing-point modelling. This was developed by Gallagher *et al.* (2011) to detect breaks in multivariate geochemical data within a borehole or core

FIG. 1. Trackway sampling curves. A, bootstrapped sample of track length for the White Sands National Monument (WHSA) double trackway (Bennett *et al.* 2020); as the sample size increases the variance falls. B, variation in standard error (SE) with sample size for the WHSA double trackway; note the wide variance within the 95% confidence area. C, variation in SE for a Namibian long trackway reported by Morse *et al.* (2013); the variance is much less within this trackway and demonstrates that the variance is potentially specific to each trackway.



sample and is implemented here in PAST version 4.03 (Hammer *et al.* 2001). The algorithm is Bayesian, ‘trans-dimensional’ Markov chain Monte Carlo (MCMC) and produces not a single output but a large number of simulations derived from the distribution. Used on a single curve it produces a predictable result in which there is a gradual decline in the frequency of change-points identified (PAST function: see: Model/change-point). However, by using multiple data curves, equivalent to different geochemical proxies in its intended use, with each curve given equal weight it is able to identify change points which occur slightly more frequently than others. It therefore provides a way of synthesizing common breaks in multiple curves.

Three types of input data were used in this analysis. Firstly, the last author (MRB) walked barefoot, at a similar constant speed, in four different substrate conditions exposed at low tide on the Conwy Estuary in North Wales, UK. A total of 50 footprints were photographed for each environment, rectified, and scaled in Photoshop before a simple maximum length estimate was measured in Photoshop. The second type of data was obtained by searching the human ichnological record for long trackways. Based on the Conwy dataset, a minimum trackway length of ten tracks was selected, although most are appreciably longer (Belvedere *et al.* 2021, tables 1, 2). This dataset includes tracks from a range of different sites: White Sands National Park (USA), Sefton Coast (UK), Walvis Bay (Namibia) and Engare Sero (Tanzania) and also small samples of early hominins from Ileret (*Homo erectus*, Kenya) and Laetoli (*Australopithecus aferensis*, Tanzania) (Belvedere *et al.* 2021, table 1). For the human/hominin trackways, data were either: (1) published length measurements; (2) bootstrapped from mean and standard deviations reported in the literature; or (3) measured directly from data curated by the authors. Finally, trackway data for both tridactyl (theropod) and sauropod tracks from Switzerland, South Korea, Portugal and China (Belvedere *et al.* 2021, tables 3, 4) were used. The same threshold of a minimum trackway length of ten tracks was applied for this dataset.

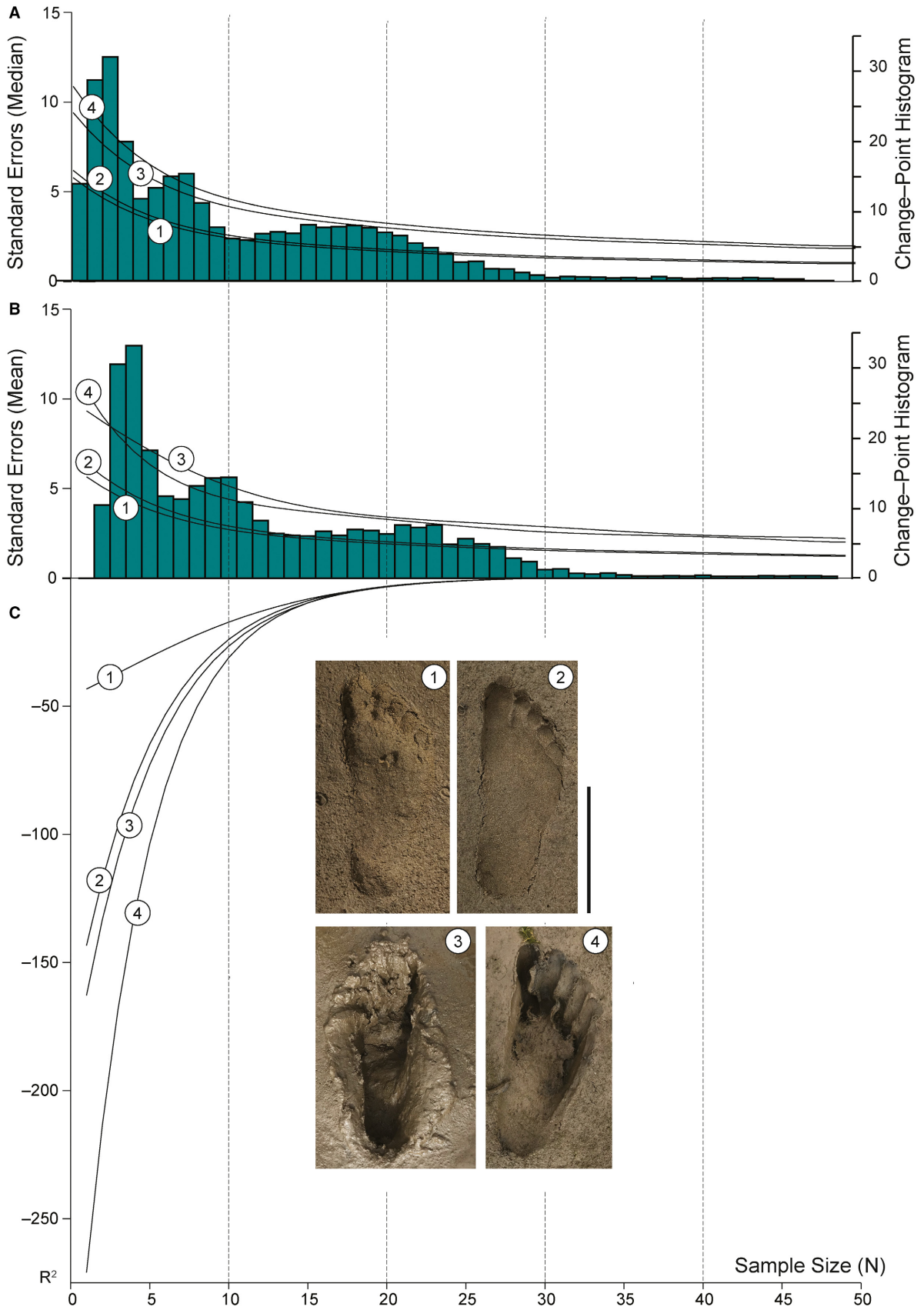
RESULTS

Figure 2 shows the output from the neoichnological estuary experiment along with typical tracks made in each of

the four sampled environments. The data has been computed for both the median and mean to highlight the differences. The tracks made in compact sand and silty sand (Fig. 2, numbers 1 and 2) show least variance as one might expect, while the highest variance occurs in a shallow mud, in which the foot tends to skate on the sublayer (Fig. 2, number 4). This latter environment gave a few extreme values; note the steep decline in the curve with increasing values of N . The soft, wet mud (Fig. 2, number 3) also has a high variance sustained over a longer range of values of N . Here variance is due to difficulty in determining the location of maximum digit position, although the deeper mud tends to hold the foot more firmly preventing slippage. Given that the track-maker is: (1) the same individual for all trackways; (2) speed was constant; and (3) the method of measurement the same; the variance identified primarily reflects substrate. As Bates *et al.* (2013) concluded, shallow tracks (Fig. 2, number 1) often contain the ‘best data’. The regression analysis suggests that the decline in sample variance reaches a peak at $N = 25$ beyond which further sampling gives limited return. The change-point modelling reveals a similar conclusion in terms of the maximum sample size but also identifies peaks at $N = 10$ and $N = 4$. Both these thresholds show improvements in sample stability. One might tentatively conclude that across the four estuarine substrates, a track sample of $N < 4$ is going to be poor, $4 < N < 10$ better, with $10 < N < 25$ likely to be reasonable and samples where $N > 25$ ideal. If we consider the median rather than the mean, a similar four-fold division can be identified although the maximum sample size falls closer to $N > 20$, indicating that a slightly smaller sample is needed for estimates using median values. In addition, the shallower the tracks at the time of imprinting the smaller the sample that is probably needed, assuming complete preservation is achieved.

If we look at published data (Fig. 3) for both experimental (mainly sandy beaches, or sand trays) and fossil cases, as one might expect the variance is much lower for the experimental trackways where taphonomic processes are excluded (e.g. Wiseman & De Groote 2018) and the substrate more homogenous, the track-maker known and walking with a constant pace. The fossil data reveals some interesting contrasts. Two trackways, possibly the longest in the world (Bennett *et al.* 2020), from White Sands National Park show significantly more variance than the

FIG. 2. Standard errors (SE; mm) plotted against sample size for four track samples made in four different substrates found at low tide on the Conwy Estuary in July 2020 (SH 79470 77361). A, the four SE curves associated with the median values with increasing sample size superimposed on the histogram of identified change points determined from these four curves plus their 95% confidence intervals ($N = 12$). B, SE curves associated with the mean values. There are three points, corresponding to the three peaks in each histogram, at which change-points are more commonly identified by the analysis. C, R^2 values for multiple linear regressions for a succession of samples each $N-1$; R^2 values fall between infinity and 1, with the latter being a near perfect data fit. Footprint scale bar represents 150 mm.



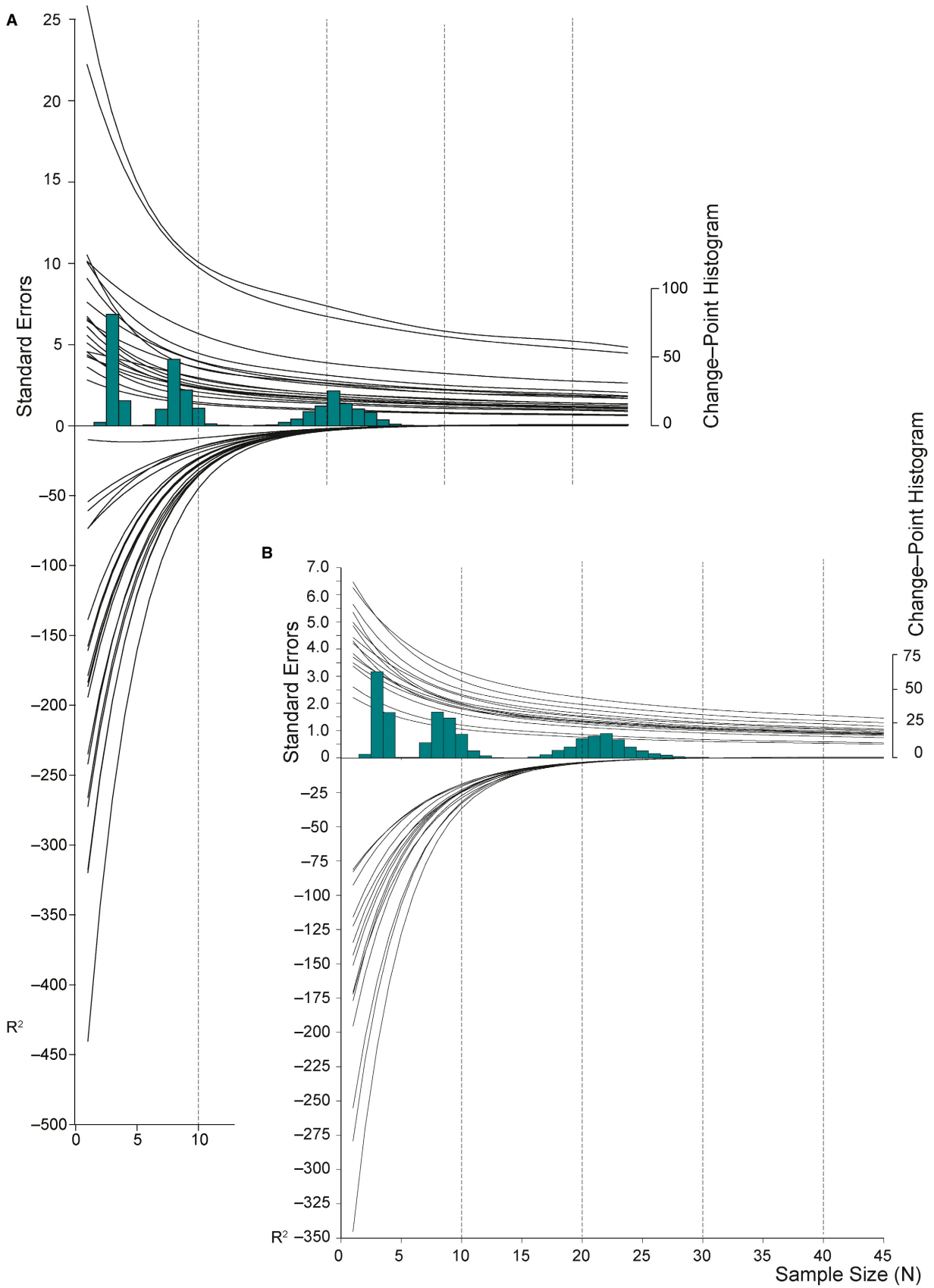


FIG. 3. Standard errors (SE; mm) plotted against sample size for a range of fossil and experimental human trackways. A, fossil trackways from Belvedere *et al.* (2021, table 1). B, experimental trackways from Belvedere *et al.* (2021, table 2). In each case, SE curves are shown above, with increasing sample size superimposed on the histogram of identified change points determined from these curves. CI curves were excluded from this analysis. There are three points at which change-points are more commonly identified by the analysis. R^2 values for multiple linear regressions for a succession of samples each $N-1$ are shown below. R^2 values fall between infinity and 1, with the latter being a near perfect data fit.

other trackways. The track-maker moved over a flat surface with a uniform substrate at a steady speed. Conditions were similar to those of substrate four in the experimental Conwy data (Fig. 2), where the track-maker skated in softer mud above a less compressible sub-layer. Broadly speaking there is a continuum in variance between the mud-rich substrates and the sandier substrates, not dissimilar to that observed in the modern analogue studies. This data more clearly establishes the three zones of change (Fig. 3) picked out by the change-point modelling in the Conwy data (Fig. 2), although the maximum sample size is closer to $N = 20$ rather than 25 and the intermediate point closer to $N = 7$.

To establish if locomotory behaviour of the track-maker was a function in this analysis, data was assembled for a number tridactyl (theropod) and sauropod dinosaur tracks from a number of locations (Belvedere *et al.* 2021, tables 3, 4; Fig. 4). A total of 68 trackways are included in this analysis and show similar patterns to that found for human tracks. The level of variance is not dissimilar to that for human tracks, especially when one considers the difference in scale between some of these tracks.

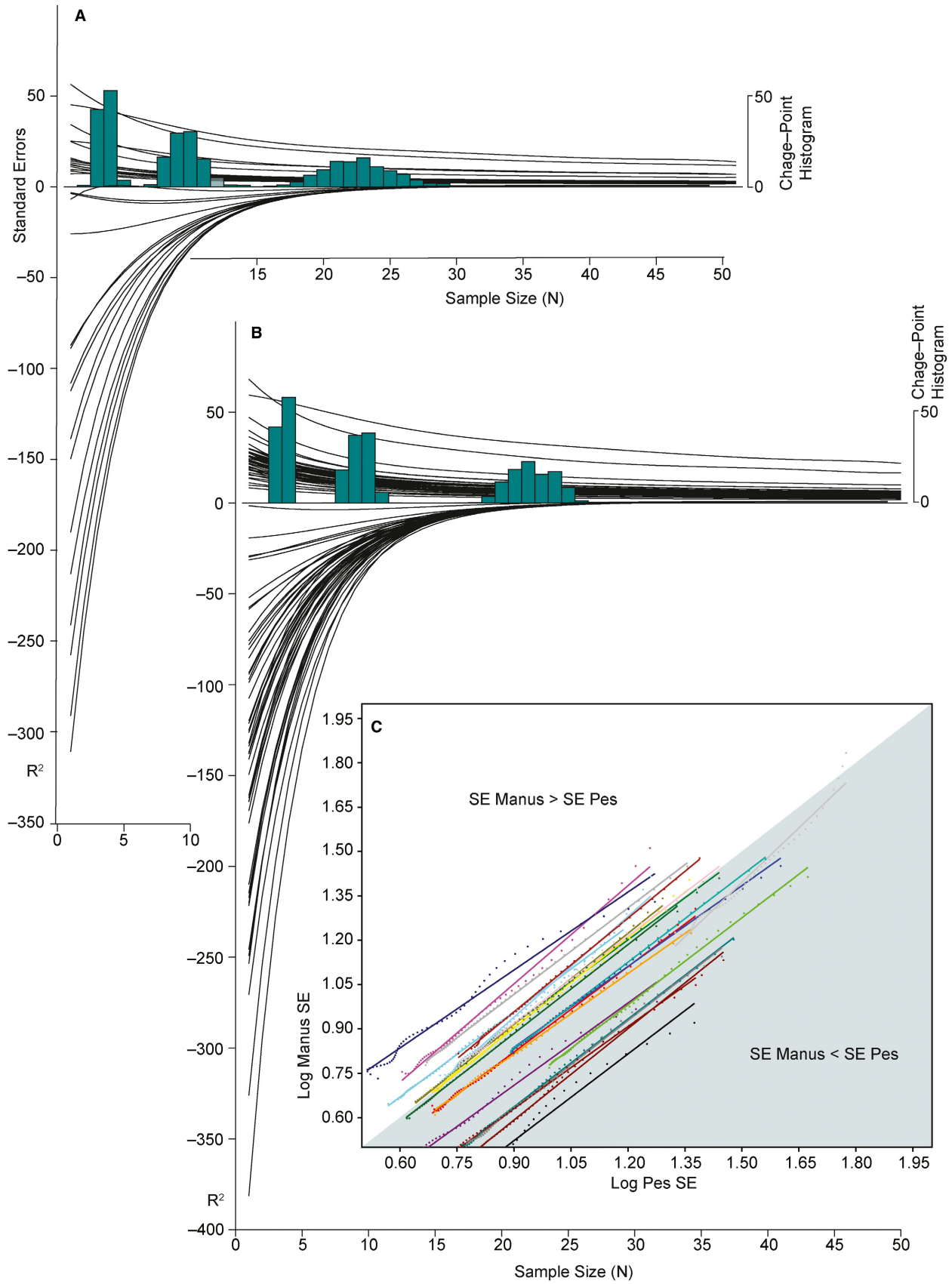
DISCUSSION

Field scientists know inherently that more is usually better in terms of a sample (e.g. Kintigh 1989; Meltzer *et al.* 1992) and their challenge typically resolves around the availability of that sample. Our results confirm this obvious point, but also place a potential constraint on sample size for statistically significant ichnological studies. The data reported here appear to suggest, across a range of track-makers, substrates and measurement systems, that a sample size in excess of 22–25 for a single individual yields little gain in terms of minimizing variance within the sample. This threshold is reduced only slightly if the median is used (*c.* $N = 20$). Variance increases with decreasing sample sizes continuously, but our analysis

suggests that it does so more significantly below a sample of seven tracks. Given that sample sizes are often small (e.g. Roberts & Berger 1997; Ashton *et al.* 2014) this is encouraging. The data also clearly show the risks of making track-length-based inferences (e.g. track-maker size, body mass) from tracks with samples of less than seven. These thresholds (Fig. 5) are to some extent artificial since the SE curves are continuous but provide a broad guide. Most field geologists have round numbers in mind when seeking samples and this information will only reinforce these natural prejudices, but every additional specimen improves the quality of the sample up to but not beyond 22–25.

The other thing that we can do with this type of data is generate generic curves for specific track-makers and/or environments by averaging the individual curves (Fig. 5). These are generated by using average standard deviations shown in Belvedere *et al.* (2021, table 1) to bootstrap between 50 and 100 length values these are then put through the SE modelling algorithm to produce average SE curves with 95% error margins. By repeating this process, a hundred times and averaging the results we obtain a stable set of ‘generic’ SE values for sample sizes up to 25 (Belvedere *et al.* 2021, table 5). There is no need for generic values where samples in excess of 25 are available and in truth there is probably no need for samples greater than 10. The average standard deviation values reported in Belvedere *et al.* (2021, table 1) can also be used for MNI calculations like that in Equation 1 (Webb *et al.* 2014). Using these estimated standard errors (SE^{est}), it is possible for a field scientist to model, at least to a first order approximation, the likely values of SE associated with a particular sample size, track-maker, and environment. Effectively Belvedere *et al.* (2021, table 5) provides a crude look-up table for potential errors when trying to estimate MNI’s or when making stature or body mass estimates. A simple $\pm 5\%$ of the mean has been argued to be problematic itself in recent years (e.g. Lin *et al.* 2013); this problem can, at least, be nuanced via a generic SE

FIG. 4. Data for long dinosaur trackways from various sites as set out in Belvedere *et al.* (2021, tables 3, 4). A, data from 16 trackways at five locations made by a bipedal tridactyl (theropod) dinosaur. B, dinosaur data from a total of 52 trackways consisting of paired manus and pes tracks (i.e. each trackway is represented by two curves) from the Canton Jura (Switzerland). C, regression plots between SE values in manus and pes tracks from the same trackway (i.e. same individual); axes are logged to improve the clarity of the plot; note that in most but not all cases variability in manus tracks is less than for pes tracks: equal variance = 27%, $SE_{manus} > SE_{pes} = 23\%$, $SE_{pes} > SE_{manus} = 50\%$.



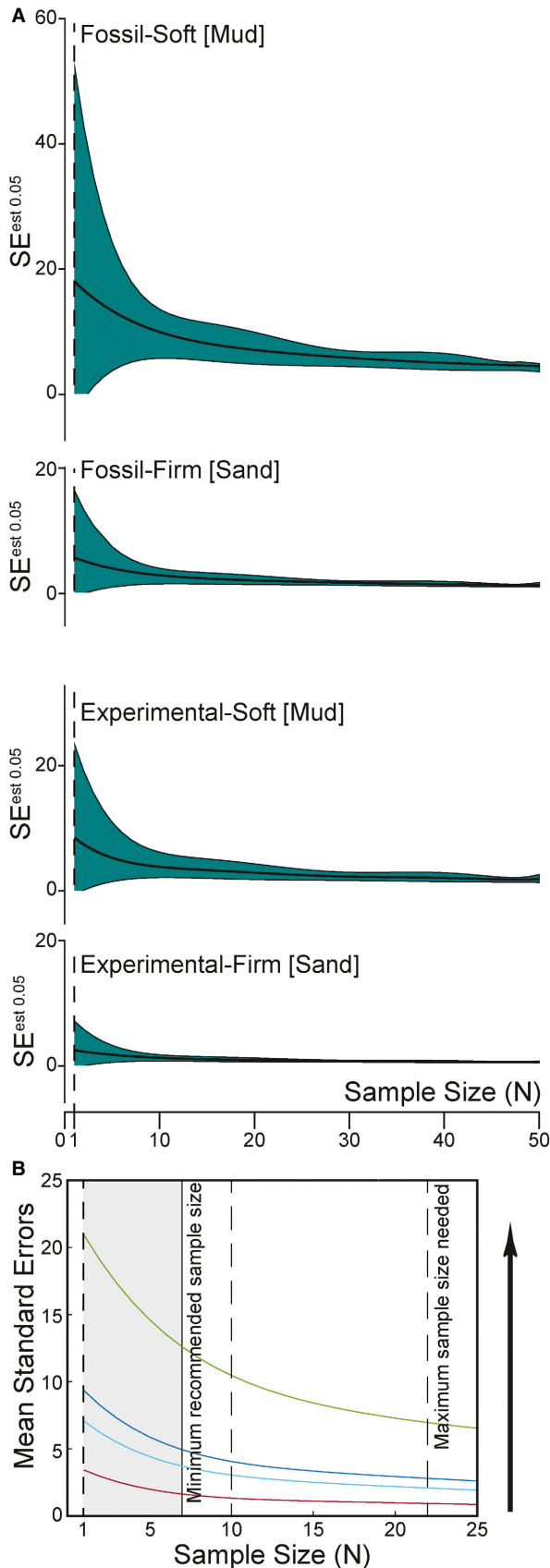


FIG. 5. A, generic curves and estimated standard errors with CI of 95% produced by averaging the data in Belvedere *et al.* (2021, table 5). B, schematic illustration of recommended sample sizes; the arrow indicates increasing variance induced by different factors: substrate, track-maker, taphonomy and measurements; curves refer to possible trackways.

with the potential SE for the sample size. Take for example the artificial sample in Figure 6 which shows a collection of tracks tentatively grouped into four categories. Table 1 shows the basic foot length data for this scenario and, based on the pair-wise comparison of the mean size differences in light of the estimated SE values in Belvedere *et al.* (2021, table 5), for a given sample size we can conclude tentatively that the minimum number of individuals is 3. Using the method of Webb *et al.* (2014), set out in Equation 1, the estimate is two individuals (specifically 2.34). The generic SE values provide a means of estimating potential SE and making conservative estimates on this basis. It is important to note that we have simply focused on the use of foot length for MNI estimates and it may be possible to develop superior measures using multi-dimensional properties.

We have chosen to only provide a SE^{est} for human tracks here, since the dinosaur data used is from a limited number of sites and environments. Long dinosaur trackways are relatively common in the literature compared to human ones but the raw data are not always reported and they are often ichnotaxon-specific, making such data tables harder to compile. Collecting and presenting multiple dimensional measurements from long trackways is something we would encourage the community to focus on in future so that tables of SE^{est} values can be compiled. An illustration of this point with respect to human tracks is the dataset in Hatala *et al.* (2020), which discussed several long trackways but despite documenting the total number of tracks, actually measured only a few. We would encourage the community to sample and report all track measurements to improve our understanding of natural intra-trackway variability.

CONCLUSION

Palaeontological and archaeological ichnological records can be fickle and rarely produce the ideal (i.e. large) statistical sample that one might hope for. Variability in tracks of the same individual adds to uncertainty when estimating the minimum number of individuals present on a trampled horizon or when making biometric inferences. The associated SE falls with increasing sample size in a predictable way, and is remarkably consistent across different track-makers, environments and measurement

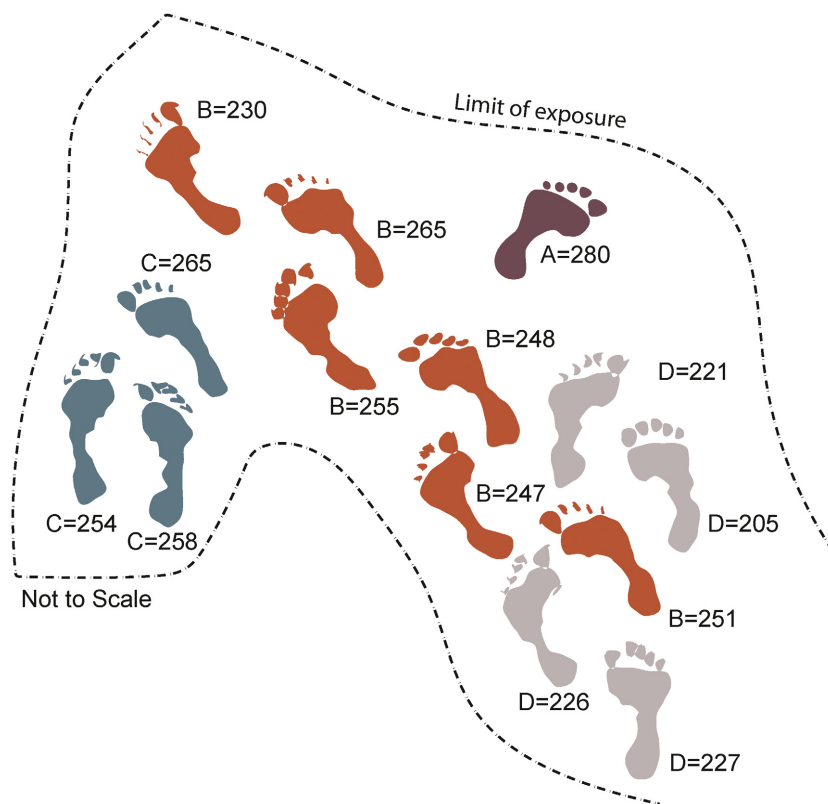


FIG. 6. Artificial scenario with a series of tracks. See Table 1 and the text for explanation.

TABLE 1. Worked example of an MNI estimate using the scenario shown in Figure 6.

	A	B	C	D	Min	Max	Mean	N	SE ^{est}
A		31	21	61	280	280	280	1	16.52
B	31		10	30	230	265	249	6	5.82
C	21	10		40	254	265	259	3	10.66
D	61	30	40		205	227	219	4	8.62

Errors are taken from the look-up table of generic errors in Belvedere *et al.* (2021, table 5) based on all categories (fossil/experimental soft/firm substrates) and in all but one case (comparison of tracks B and C) the difference between the means exceeds the maximum estimated SE^{est}, approximated by the upper value of the 95% CI (CIU). The tentative conclusion is that there is evidence for only three track-makers.

protocols. Change-point modelling suggests that along this continuum one can identify three thresholds, which, although perhaps slightly artificial, can provide some guidance to the field scientist. The improvements in SE beyond a sample of 25 appear to be relatively minor. This threshold represents a reasonable maximum sample size if such a number is required for sampling permissions or conservation statements. The results presented here also suggest that a sample greater than 7–10 also gives better results than a smaller one, and that samples less than 7 are to be avoided if at all possible. In practice, however,

small samples are often the norm where exposure and environment limit preservation. To assist with this, we provide a simple ‘look-up’ table for human tracks which provides estimated SE for small samples and may help the ichnologist frustrated by a small sample, make appropriately conservative estimates. We suggest that a similar look-up table could be developed for other vertebrates. One final word of caution, this outcome only applies to footprint length. More complex measures, involving multiple dimensions or the whole plantar surface of a track, for example, will involve more degrees of freedom and with them the minimum sample size is likely to increase substantially.

Acknowledgements. We wish to thank Dr Zarah Goshi (AstraZeneca PLC) for suggesting the statistical approach in the context of another project. We wish to thank the editor and the two reviewers (J. Lallensack and M. Buchwitz) for the comments and suggestions that helped to improve this paper.

DATA ARCHIVING STATEMENT

Datasets for this study are available in the Dryad Digital Repository: <https://doi.org/10.5061/dryad.83bk3j9qr>. Matlab script is available in Github: <https://github.com/bosmart/enough-is-enough-paper>.

Editor. Lorenzo Marchetti

REFERENCES

- ALTAMURA, F., BENNETT, M. R., D'AOÛT, K., GAUDZINSKI-WINDHEUSER, S., MELIS, R. T., REYNOLDS, S. C. and MUSSI, M. 2018. Archaeology and ichnology at Gombore II-2, Melka Kunture, Ethiopia: everyday life of a mixed-age hominin group 700,000 years ago. *Scientific Reports*, **8**, 2815.
- ASHTON, N., LEWIS, S. G., DE GROOTE, I., DUFFY, S. M., BATES, M., BATES, R., HOARE, P., LEWIS, M., PARFITT, S. A., PEGLAR, S., WILLIAMS, C. and STRINGER, C. 2014. Hominin footprints from Early Pleistocene deposits at Happisburgh, UK. *PLoS One*, **9**, e88329.
- AVANZINI, M., BERNARDI, M. and PETTI, F. M. 2011. Soldier tracks in a First World War fort (Valmorbiawerk, Trento, Italy). *Ichnos*, **18**, 72–78.
- BATES, K. T., SAVAGE, R., PATAKY, T. C., MORSE, S. A., WEBSTER, E., FALKINGHAM, P. L., REN, L., QIAN, Z., COLLINS, D., BENNETT, M. R., McCLYMONT, J. and CROMPTON, R. H. 2013. Does footprint depth correlate with foot motion and pressure? *Journal of the Royal Society Interface*, **10**, 20130009.
- BELVEDERE, M., BUDKA, M., WISEMAN, A. L. A. and BENNETT, M. R. 2021. Data from: When is enough, enough? Questions of sampling in vertebrate ichnology. *Dryad Digital Repository*. <https://doi.org/10.5061/dryad.83bk3j9qr>
- BENNETT, M. R. and MORSE, S. A. 2014. *Human footprints: Fossilised locomotion?* Springer International Publishing.
- BUSTOS, D., ODESS, D., URBAN, T. M., LALLENSACK, J., BUDKA, M., SANTUCCI, V. L., MARTINEZ, P., WISEMAN, A. L. A. and REYNOLDS, S. C. 2020. Walking in mud: remarkable Pleistocene human trackways from White Sands National Park (New Mexico). *Quaternary Science Reviews*, **249**, 106610.
- CARAVAGGI, P., PATAKY, T., GOULERMAS, J. Y., SAVAGE, R. and CROMPTON, R. 2009. A dynamic model of the windlass mechanism of the foot: evidence for early stance phase preloading of the plantar aponeurosis. *Journal of Experimental Biology*, **212**, 2491–2499.
- DINGWALL, H. L., HATALA, K. G., WUNDERLICH, R. E. and RICHMOND, B. G. 2013. Hominin stature, body mass, and walking speed estimates based on 1.5 million-year-old fossil footprints at Ileret, Kenya. *Journal of Human Evolution*, **64**, 556–568.
- DOMJANIC, J., SEIDLER, H. and MITTEROECKER, P. 2015. A combined morphometric analysis of foot form and its association with sex, stature, and body mass. *American Journal of Physical Anthropology*, **157**, 582–591.
- DUVEAU, J., BERILLON, G., VERNA, C., LAISNÉ, G. and CLIQUET, D. 2019. The composition of a Neandertal social group revealed by the hominin footprints at Le Rozel (Normandy, France). *Proceedings of the National Academy of Sciences*, **116**, 19409–19414.
- ELFTMAN, H. and MANTER, J. 1935. Chimpanzee and human feet in bipedal walking. *American Journal of Physical Anthropology*, **20**, 69–79.
- FALKINGHAM, P. L. and GATESY, S. M. 2020. Discussion: Defining the morphological quality of fossil footprints. Problems and principles of preservation in tetrapod ichnology with examples from the Palaeozoic to the present by Lorenzo Marchetti et al. *Earth-Science Reviews*, **208**, 103320.
- GALLAGHER, K., BODIN, T., SAMBRIDGE, M., WEISS, D., KYLANDER, M. and LARGE, D. 2011. Inference of abrupt changes in noisy geochemical records using transdimensional changepoint models. *Earth & Planetary Science Letters*, **311**, 182–194.
- GATESY, S. M. and FALKINGHAM, P. L. 2017. Neither bones nor feet: track morphological variation and 'preservation quality'. *Journal of Vertebrate Paleontology*, **37**, e1314298.
- HAMMER, D. A. T., RYAN, P. D., HAMMER, Ø. and HARPER, D. A. T. 2001. PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, **4** (1), 4. <https://www.nhm.uio.no/english/research/infrastructure/past/>
- HARCOURT-SMITH, W. E. H. and AIELLO, L. C. 2004. Fossils, feet and the evolution of human bipedal locomotion. *Journal of Anatomy*, **204**, 403–416.
- HATALA, K. G., DINGWALL, H. L., WUNDERLICH, R. E. and RICHMOND, B. G. 2013. The relationship between plantar pressure and footprint shape. *Journal of Human Evolution*, **65**, 21–28.
- DEMES, B. and RICHMOND, B. G. 2016. Laetoli footprints reveal bipedal gait biomechanics different from those of modern humans and chimpanzees. *Proceedings of the Royal Society B*, **283**, 20160235.
- ROACH, N. T., OSTROFSKY, K. R., WUNDERLICH, R. E., DINGWALL, H. L., VILLMOARE, B. A., GREEN, D. J., BRAUN, D. R., HARRIS, J. W. K., BEHRENSMEYER, A. K. and RICHMOND, B. G. 2017. Hominin track assemblages from Okote Member deposits near Ileret, Kenya, and their implications for understanding fossil hominin paleobiology at 1.5 Ma. *Journal of Human Evolution*, **112**, 93–104.
- HARCOURT-SMITH, W. E. H., GORDON, A. D., ZIMMER, B. W., RICHMOND, B. G., POBINER, B. L., GREEN, D. J., METALLO, A., ROSSI, V. and LIUTKUS-PIERCE, C. M. 2020. Snapshots of human anatomy, locomotion, and behavior from Late Pleistocene footprints at Engare Sero, Tanzania. *Scientific Reports*, **10**, 7740.
- HELM, C. W., CAWTHRA, H. C., COWLING, R. M., DE VYNCK, J. C., LOCKLEY, M. G., MAREAN, C. W., THESEN, G. H. H. and VENTER, J. A. 2020. Pleistocene vertebrate tracksites on the Cape south coast of South Africa and their potential palaeoecological implications. *Quaternary Science Reviews*, **235**, 105857.
- McCREA, R. T., CAWTHRA, H. C., LOCKLEY, M. G., COWLING, R. M., MAREAN, C. W., THESEN, G. H. H., PIGEON, T. S. and HATTINGH, S. 2018. A new Pleistocene hominin tracksite from the Cape South Coast, South Africa. *Scientific Reports*, **8**, 3772.
- KER, R. F., BENNETT, M. B., BIBBY, S. R., KESTER, R. C. and ALEXANDER, R. M. 1987. The spring in the arch of the human foot. *Nature*, **325**, 147–149.
- KINTIGH, K. 1989. Sample size, significance, and measures of diversity. 25–36. In LEONARD, R. D. and JONES, G. T.

- (eds). *Quantifying diversity in archaeology*. Cambridge University Press.
- LEAKEY, M. D. and HAY, R. L. 1979. Pliocene footprints in the Laetoli Beds at Laetoli, northern Tanzania. *Nature*, **278**, 317–323.
- LIN, M., LUCAS, H. C. and SHMUELI, G. 2013. Research Commentary—too big to fail: large samples and the *p*-value problem. *Information Systems Research*, **24**, 906–917.
- MARCHETTI, L., BELVEDERE, M., VOIGT, S., KLEIN, H., CASTANERA, D., DÍ AZ-MARTÍ NEZ, I., MARTY, D., XING, L., FEOLA, S., MELCHOR, R. N. and FARLOW, J. O. 2019. Defining the morphological quality of fossil footprints. Problems and principles of preservation in tetrapod ichnology with examples from the Palaeozoic to the present. *Earth-Science Reviews*, **193**, 109–145.
- 2020. Reply to discussion of “Defining the morphological quality of fossil footprints. Problems and principles of preservation in tetrapod ichnology with examples from the Palaeozoic to the present” by Marchetti et al. (2019). *Earth-Science Reviews*, **208**, 103319.
- McClymont, J., PATAKY, T. C., CROMPTON, R. H., SAVAGE, R. and BATES, K. T. 2016. The nature of functional variability in plantar pressure during a range of controlled walking speeds. *Royal Society Open Science*, **3**, 160369.
- MELTZER, D. J., LEONARD, R. D. and STRATTON, S. K. 1992. The relationship between sample size and diversity in archaeological assemblages. *Journal of Archaeological Science*, **19**, 375–387.
- MORSE, S. A., BENNETT, M. R., LIUTKUS-PIERCE, C., THACKERAY, F., McClymont, J., SAVAGE, R. and CROMPTON, R. H. 2013. Holocene footprints in Namibia: the influence of substrate on footprint variability. *American Journal of Physical Anthropology*, **151**, 265–279.
- NIEDŹWIEDZKI, G., SZREK, P., NARKIEWICZ, K., NARKIEWICZ, M. and AHLBERG, P. E. 2010. Tetrapod trackways from the early Middle Devonian period of Poland. *Nature*, **463**, 43–48.
- RAICHLLEN, D. A. and GORDON, A. D. 2017. Interpretation of footprints from Site S confirms human-like bipedal biomechanics in Laetoli hominins. *Journal of Human Evolution*, **107**, 134–138.
- ROACH, N. T., HATALA, K. G., OSTROFSKY, K. R., VILMOARE, B., REEVES, J. S., DU, A., BRAUN, D. R., HARRIS, J. W. K., BEHRENSMEYER, A. K. and RICHMOND, B. G. 2016. Pleistocene footprints show intensive use of lake margin habitats by *Homo erectus* groups. *Scientific Reports*, **6**, 26374.
- ROBERTS, L. and BERGER, D. 1997. Last interglacial (c. 117 kyr) human footprints from South Africa. *South African Journal of Science*, **93**, 349–350.
- ROBERTS, G., GONZALEZ, S. and HUDDART, D. 1996. Intertidal Holocene footprints and their archaeological significance. *Antiquity*, **70**, 647–651.
- STÖSSEL, I. 1995. The discovery of a new Devonian tetrapod trackway in SW Ireland. *Journal of the Geological Society*, **152**, 407–413.
- WILLIAMS, E. A. and HIGGS, K. T. 2016. Ichnology and depositional environment of the Middle Devonian Valentia Island tetrapod trackways, south-west Ireland. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **462**, 16–40.
- THULBORN, T. 1990. *Dinosaur tracks*. Chapman & Hall.
- WEBB, D., ROBU, M., MOLDOVAN, O., CONSTANTIN, S., TOMUS, B. and NEAG, I. 2014. Ancient human footprints in Ciur-Izbuc Cave, Romania. *American Journal of Physical Anthropology*, **155**, 128–135.
- WISEMAN, A. L. A. and DE GROOTE, I. 2018. A three-dimensional geometric morphometric study of the effects of erosion on the morphologies of modern and prehistoric footprints. *Journal of Archaeological Science: Reports*, **17**, 93–102.
- STRINGER, C. B., ASHTON, N., BENNETT, M. R., HATALA, K. G., DUFFY, S., O’BRIEN, T. and DE GROOTE, I. 2020. The morphological affinity of the Early Pleistocene footprints from Happisburgh, England, with other footprints of Pliocene, Pleistocene, and Holocene age. *Journal of Human Evolution*, **144**, 102776.