Influence of data filters on the accuracy and precision of paleomagnetic poles: what is the optimal sampling strategy?

Gerritsen Dieke[1], Vaes Bram[1], and van Hinsbergen Douwe J.J.[1]

[1]Utrecht University

November 16, 2022

Abstract

To determine a paleopole, the paleomagnetic community commonly applies a loosely defined set of quantitative data filters that were established for studies of geomagnetic field behavior. These filters require costly and time-consuming sampling procedures, but whether they improve accuracy and precision of paleopoles has not yet been systematically analyzed. In this study, we performed a series of experiments on four datasets which consist of 73-125 lava sites with 6-7 samples per lava. The datasets are from different regions and ages, and are large enough to represent paleosecular variation, yet contain demonstrably unreliable datapoints. We show that data filters based on within-site scatter (a k-cutoff, a minimum number of samples per site, and eliminating the farthest outliers per site) cannot objectively identify unreliable directions. We find instead that excluding unreliable directions relies on the subjective interpretation of the expert, highlighting the importance of making all data available following the FAIR principles. In addition, data filters that eliminate datapoints even have an adverse effect: the accuracy as well as the precision of paleopoles decreases with the decreasing number of data. Between-site scatter far outweighs within-site scatter, and when collecting paleomagnetic poles, the extra efforts put into collecting multiple samples per site are more effectively spent on collecting more single-sample sites.

Influence of data filters on the accuracy and precision of paleomagnetic poles: what is the optimal sampling strategy?

Dieke Gerritsen[1], Bram Vaes[1], and Douwe J.J. van Hinsbergen[1]

[1]Department of Earth Sciences, Utrecht University, Princetonlaan 8A, 3584 CB Utrecht, the Netherlands.

Corresponding authors: Dieke Gerritsen (h.gerritsen1@uu.nl) and Bram Vaes (b.vaes@uu.nl)

Key Points:

- Within-site data filters (minimum k or n, or eliminating outliers per site) are unsuccessful in eliminating unreliable data.
- The precision of paleomagnetic poles is dominated by between-site scatter; within-site scatter has minimal contribution.
- For paleopoles, efforts put into collecting multiple samples per site are more effectively spent on collecting more single-sample sites.
-

Abstract

To determine a paleopole, the paleomagnetic community commonly applies a loosely defined set of quantitative data filters that were established for studies of geomagnetic field behavior. These filters require costly and time-consuming sampling procedures, but whether they improve accuracy and precision of paleopoles

has not yet been systematically analyzed. In this study, we performed a series of experiments on four datasets which consist of 73-125 lava sites with 6-7 samples per lava. The datasets are from different regions and ages, and are large enough to represent paleosecular variation, yet contain demonstrably unreliable datapoints. We show that data filters based on within-site scatter (a k-cutoff, a minimum number of samples per site, and eliminating the farthest outliers per site) cannot objectively identify unreliable directions. We find instead that excluding unreliable directions relies on the subjective interpretation of the expert, highlighting the importance of making all data available following the FAIR principles. In addition, data filters that eliminate datapoints even have an adverse effect: the accuracy as well as the precision of paleopoles decreases with the decreasing number of data. Between-site scatter far outweighs within-site scatter, and when collecting paleomagnetic poles, the extra efforts put into collecting multiple samples per site are more effectively spent on collecting more single-sample sites.

1 Introduction

Paleomagnetic poles, or paleopoles, quantify the past position of rocks relative to the geomagnetic pole and constrain tectonic reconstructions and apparent polar wander paths (APWPs) (e.g., Besse and Courtillot, 2002; Torsvik et al., 2012). The calculation of paleopoles relies on the assumption that the time-averaged geomagnetic field approximates a geocentric axial dipole (GAD), but is complicated by short-term deviations from this field (e.g., Cromwell et al., 2018; Oliveira et al., 2021) known as paleosecular variation (PSV). To obtain a paleopole, paleomagnetists therefore average virtual geomagnetic poles (VGPs), whereby every VGP is then assumed a 'spot reading': an instantaneous reading of the past geomagnetic field collected from a rock unit ('site') that represents an increment of geological time, such as a lava (Butler, 1992; Tauxe, 2010). However, not every VGP represents an accurate spot reading because artifacts may be introduced by measuring errors or remagnetization. Therefore, the paleomagnetic community commonly uses a set of data filters to acquire a set of reliable spot readings. However, these filters vary between authors and were not determined by studies aiming to constrain paleopoles.

The studies that established the data filters, investigated PSV and geomagnetic field behavior by determining the between-site scatter of a set of VGPs (e.g., Cromwell et al., 2018; de Oliveira et al., 2021; Johnson et al., 2008; Johnson & Constable, 1996; Tauxe et al., 2003). To this end, these studies aim to correct for within-site scatter induced by measuring errors and typically require a minimum number of readings per site, although this number of readings varies between authors (e.g., Biggin et al., 2008; Johnson et al., 2008; Doubrovine et al., 2019; Cromwell et al., 2018). The resulting paleomagnetic directions are then averaged to a site-mean direction which is converted to a VGP if the site passes a criterion for the minimum within-site precision value ('cutoff'), typically expressed as a Fisher (1953) precision parameter k . This value also varies between authors, e.g., $k \geq 50$ or $k \geq 100$ (e.g., Biggin et al., 2008; Cromwell et al., 2018; Johnson et al., 2008; Tauxe et al., 2003). Subsequently, similar procedures have become common for calculation of paleopoles (e.g., Butler, 1992; Lippert et al., 2014; Meert et al., 2020). However, does this time- and data-intensive procedure improve the accuracy and precision of paleopoles?

In this study, we analyze to what extent commonly applied paleomagnetic data filters established for PSV studies improve the accuracy and precision of paleopoles. To this end, we study four large paleomagnetic datasets obtained from lava sequences from the Cretaceous of Mongolia, the Permian of Norway, the Miocene of Turkey, and the Quaternary of Antarctica. These datasets are large enough to represent PSV, but contain additional between-site and within-site scatter of varying magnitude due to measurement errors, lightning-induced remagnetization, and/or tectonic deformation. We perform a series of experiments to examine the effects of applying filters on the accuracy and precision of paleopoles. We evaluate whether these filters can objectively exclude outliers and filter non-PSV induced scatter from the paleomagnetic datasets. We then assess how a given number of paleomagnetic directions is optimally distributed over a collection of paleomagnetic sites to acquire the best-constrained paleopole position. We discuss to what extent the filters used in PSV studies improve paleopole accuracy and precision. Our results aim to aid paleomagnetists to optimize their sampling and data filtering strategies.

2 Background

To evaluate the reliability of a paleopole, it is common to use filters to obtain a paleomagnetic dataset representative of PSV. Initially, Van der Voo (1990) recommended to average a minimum set of 25 samples or sites (i.e., spot readings) for a paleopole, and formulated loosely defined filters using Fisher (1953) statistics. Those statistics are used to describe a paleomagnetic dataset and can be applied to either directional data or VGPs. Three parameters are of importance in this statistical approach: the radius of the 95% confidence cone around the mean (α_{95} for directions and A_{95} for VGPs), the precision parameter (k for directions and K for VGPs), and the circular standard deviation (S) that quantifies the angular between-site dispersion of VGPs. Van der Voo (1990) suggested filters for these parameters:α_{95} or $A_{95} \leq 16°$, and k or K$\geq$10. In later studies, filters were added to determine the reliability of each VGP based on quality filters used in PSV studies (e.g., Lippert et al., 2014).

For obtaining paleopoles, paleomagnetists usually sample multiple sites on a locality, where every site likely represents a spot reading of the paleomagnetic field, e.g., a lava flow. The total scatter in a dataset then consists of within-site and between-site scatter (Biggin et al., 2008). Paleomagnetists collect multiple samples per site to test whether a paleomagnetic direction is reproducible and to average measurement and other random errors (McElhinny & McFadden, 1999). A quality check for within-site precision is a k-cutoff (Tauxe et al., 2003), where sites with a k value lower than an arbitrary value of 50 or 100 are discarded (e.g., Biggin et al., 2008; Johnson et al., 2008; Lippert et al., 2014). Furthermore, outliers are often subjectively discarded from sites based on the 'expert eye' and experience of the interpreter. How many samples should be collected per site is not widely agreed upon. Based on Monte Carlo simulations, a minimum number of five independently oriented samples was deemed necessary to estimate k reliably (Tauxe et al., 2003). Others suggest that three (Meert et al., 2020), four (Cromwell et al., 2018; Lippert et al., 2014), or six (Asefaw et al., 2021) samples are needed. But most paleomagnetic studies collected six to eight samples per lava site.

To average between-site scatter, assumed predominantly the result of PSV, multiple sites are collected, but the minimum number of sites required for a 'good' average, that is, the pole, is not well defined (Tauxe, 2010). The precision with which a paleopole is calculated increases with increasing number of underlying VGPs (Vaes et al., 2021). Meert et al. (2020) suggested eight sites are sufficient to determine a paleopole, provided that each site is constrained by at least three samples. However, based on a statistical simulation of PSV, Tauxe et al. (2003) showed that a minimum of approximately 100 sites may be required to fully sample PSV. However, this number is rarely obtained due to limited availability of sites and because resources are consumed by sampling a high number of samples for each site. Below, we analyze the effect of these different sampling and filtering strategies on the accuracy and precision of the resulting paleopole.

3 Datasets and Approach

3.1 Datasets

As basis for our analysis, we use four large paleomagnetic datasets derived from lavas from Mongolia (van Hinsbergen et al., 2008), Norway (Haldan et al., 2014), Turkey (van Hinsbergen et al., 2010), and Antarctica (Asefaw et al., 2021). The results were previously published, but for the purpose of this study, we reinterpreted the demagnetization diagrams of all samples. We did this by identifying the characteristic remanent magnetization (ChRM) by principal component analysis (Kirschvink, 1980), using the online paleomagnetic analysis platform Paleomagnetism.org (Koymans et al., 2016, 2020). We have thereby not forced interpreted components through the origin, and we have not used the remagnetization great-circle method of McFadden & McElhinny (1988), as was occasionally done in the original interpretations. All interpretable diagrams were interpreted, and no samples were discarded based on obviously outlying directions, e.g., due to lightning strikes. In other words, we deliberately kept directions that an experienced paleomagnetist would likely immediately discard as unreliable. We also did not exclude sites that were collected from tectonically deformed regions. As a result, the dataset contains larger noise and a higher between-site scatter (lower K-value) than the published values. This allows us to assess whether the objective quality criteria alone can clean the dataset from outliers, or whether this relies on a subjective 'expert eye'. The sites from Mongolia, Norway, and Turkey were sampled with a minimum of seven samples per site, and the sites from Antarctica with a minimum of six. Only sites with seven (or six, for Antarctica) interpretable directions have been used in our analysis. If sites contained more samples, only the first seven (or six, for Antarctica) samples were used. All uninterpreted data are publicly available in databases (Paleomagnetism.org (Koymans et al., 2020) or MagIC (Tauxe et al., 2016), see Data Availability Statement). Our reinterpretations are provided in the supplementary information (Table S1) to this paper, but because our interpretations are (deliberately) not better than the original interpretations, we will not upload these in the databases to avoid confusion. All paleopole positions from this study differ from the published pole positions, but still sit within the 95% confidence cone of the respective published pole (Figure 1).

After the reinterpretation, we arrived at a total of 108 sites of lower Cretaceous lavas, corresponding to the base of the Cretaceous Normal Superchron, from the NE Gobi Altai mountain range of southern Mongolia (van Hinsbergen et al., 2008; vH08). We did not include samples from the Artz Bogd area from the original dataset because these appear systematically rotated due to tectonic deformation. Furthermore, only the sites acquired by van Hinsbergen et al. (2008) are included. Our reinterpreted dataset (G21) leads to a higher scatter than the original interpretation of van Hinsbergen et al. (2008) for reasons outlined above ($A_{95,vH08} = 5.3°$ vs. $A_{95,G21} = 6.4°$, $K_{vH08} = 9.1$ vs. $K_{G21} = 5.5$). Most sites have limited within-site scatter and high k-values, as expected for lava sites.

We arrived at a total of 73 sites from lavas of Permian age from the Oslo Graben in Norway (Haldan et al., 2014; H14). This dataset contains lavas from the Permo-Carboniferous Reversed Superchron (PCRS). Interestingly, the between-site scatter is much lower than for the other three localities, as addressed by (Brandt et al., 2021; Handford et al., 2021). The within-site scatter of this dataset, however, is higher. Our reinterpretation contains more scatter than the interpretation of Haldan et al. (2014) ($A_{95,H14} = 1.9°$ vs. $A_{95,G21} = 3.1°$, $K_{H14} = 52.2$ vs. $K_{G21} = 29.4$).

A total of 125 sites of lower to middle Miocene lavas and ignimbritic tuffs come from basins in the northern Menderes Massif in western Turkey (van Hinsbergen et al., 2010; vH10). This dataset was acquired in a tectonically active area and contains basins that were interpreted to be tectonically coherent, and basins that were interpreted by van Hinsbergen et al. (2010) to be internally tectonically disturbed (later confirmed and mapped out in detail by Uzel et al. (2015, 2017)). It also contained sites acquired by preceding studies, but only the sites acquired by van Hinsbergen et al. (2010) are included in our dataset. Our re-interpreted dataset contains an equal amount of scatter as the dataset of van Hinsbergen et al. (2010) ($A_{95,vH10} = 6.7°$ vs. $A_{95,G21} = 6.4°$, $K_{vH10} = 4.6$ vs. $K_{G21} = 4.8$).

Finally, we used 107 sites with n=6 from lavas of Plio-Pleistocene age of the Erebus volcanic province in Antarctica (Asefaw et al., 2021; A21). We did not reinterpret this dataset and used the published interpretations, but only used sites with n=6. Furthermore, we did not use the age constraint of <5 Ma and k-cutoff. Our reinterpretation does not significantly differ from the published one of Asefaw et al. (2021) ($A_{95,A21} = 5.5°$ vs. $A_{95,G21} = 5.9°$, $K_{A21} = 7.7$ vs. $K_{G21} = 6.3$).

3.2 Approach

We performed a series of experiments to study the effect on paleopole accuracy and precision of objective filters that one may use when multiple samples per site are available. These filters include the influence of (i) a k-cutoff; (ii) the number of samples per site (n), and (iii) discarding outliers of within-site dispersion. In addition, we assessed how a given number of paleomagnetic directions may be optimally distributed over a collection of paleomagnetic sites (N) to acquire the most best-constrained paleopole position. This will be addressed using Fisher (1953) parameters: the radius of the 95% confidence cone around the paleopole (A_{95}), the precision parameter (K), and the circular standard deviation (S). We performed experiments using a Python code, which we developed making extensive use of the freely available paleomagnetic software package PmagPy (Tauxe et al., 2016). We perform the calculations 1000 times, and each time different samples and/or sites will be selected from the population. We then calculate the mean and standard deviation of the different Fisher (1953) parameters. Our Python code is available on GitHub (see Data Availability Statement).

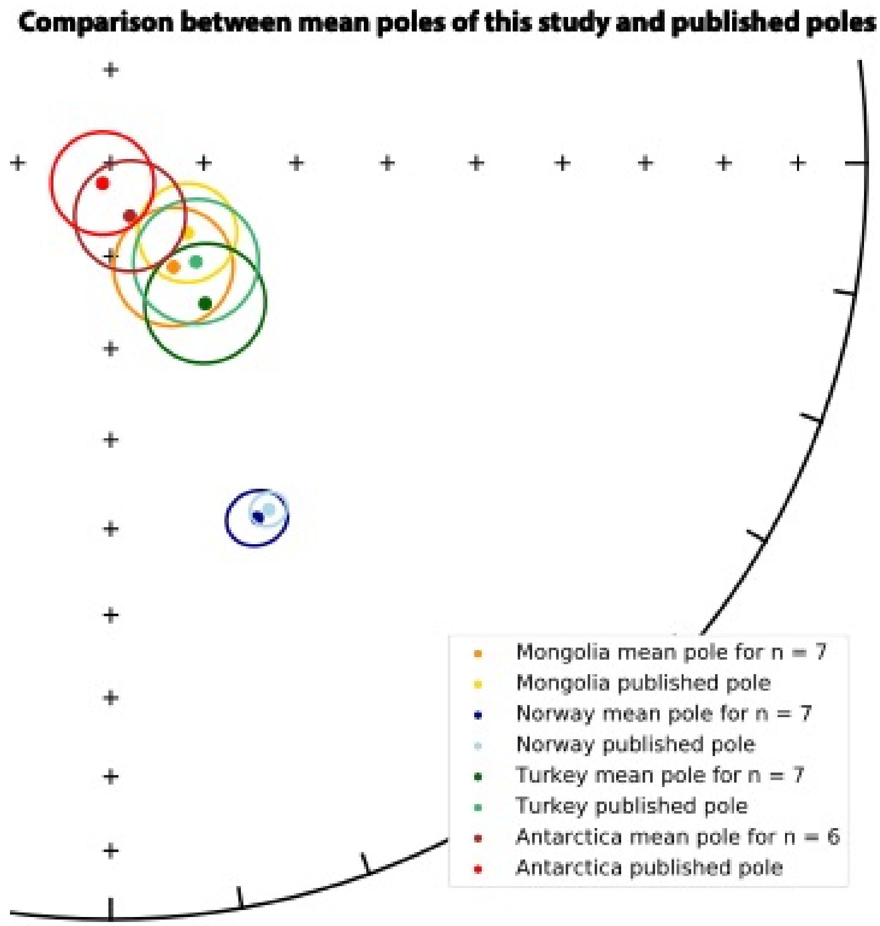

Figure 1. Stereographic projection showing the mean paleopole position and its A_{95} from this study and the published paleopole position and its A_{95} for Mongolia, Norway, Turkey, and Antarctica.

4 Results

4.1 Effect of a k-cutoff

A widely used filter to establish site-mean directions that represent spot readings of the paleomagnetic field is to check for high within-site scatter by means of a k-cutoff. Sites with low $k$values are interpreted as either remagnetized or not representative of a spot reading and are therefore discarded (Figure S1). We analyze the effect of the k-cutoff size on between-site scatter using the mean paleopole of a dataset (Figure 2) and its N, A_{95}, K, and S (Figure 3). The pole positions do not significantly change when applying a k-cutoff (Figure 2). When a low $k \geq 10$ cutoff is applied, the number of sites remaining in the dataset decreases by approximately 15-30% (Figure 3a), omitting sites with near-random direction distributions. Increasing the cutoff size beyond 20, the decay in sites decreases. The A_{95} increases with increasing k-cutoff due to a decrease in N (Figure 3b). As expected, discarding sites with high within-site scatter leads to a higher K and lower S, although the effects are small (Figure 3c, d), except in the dataset from Norway where K rises from 30 to 70. No major improvement becomes apparent in any of the variables at the commonly used cutoff sizes of 50 and 100. Importantly, although the application of a k-cutoff may subtly decrease S and increase K, the decrease in N leads to an increase of the A_{95}, and thus to a decrease in precision of the paleopole when a k-cutoff is applied, whereas the pole position remains approximately the same.

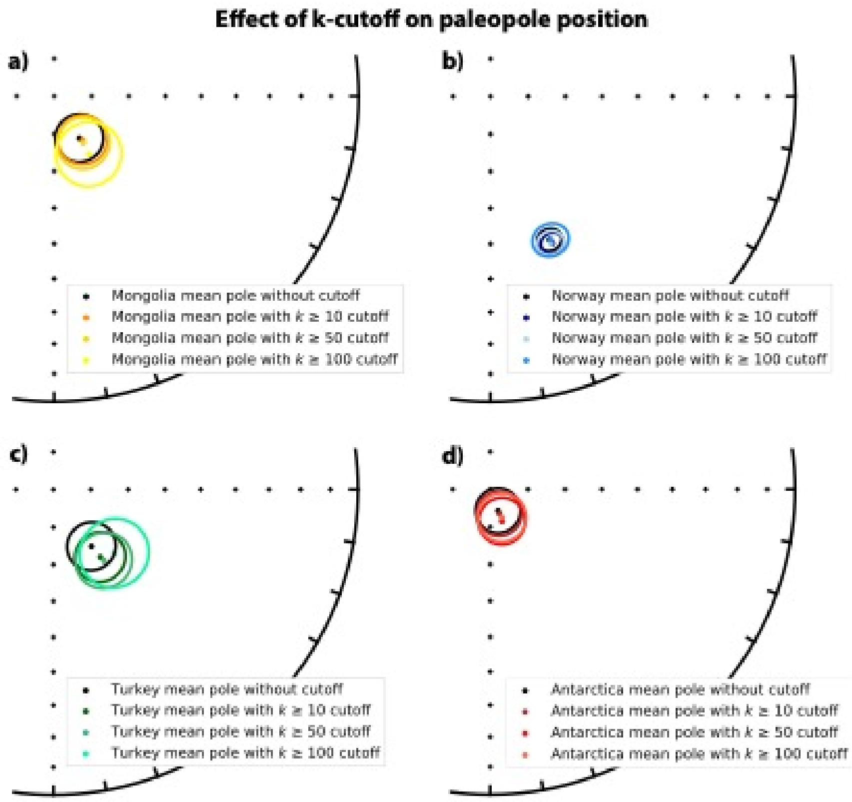

Figure 2. Stereographic projections illustrating the effect of the size of a k-cutoff on paleopole position and the A_{95} for the dataset from a) Mongolia, b) Norway, c) Turkey, and d) Antarctica.

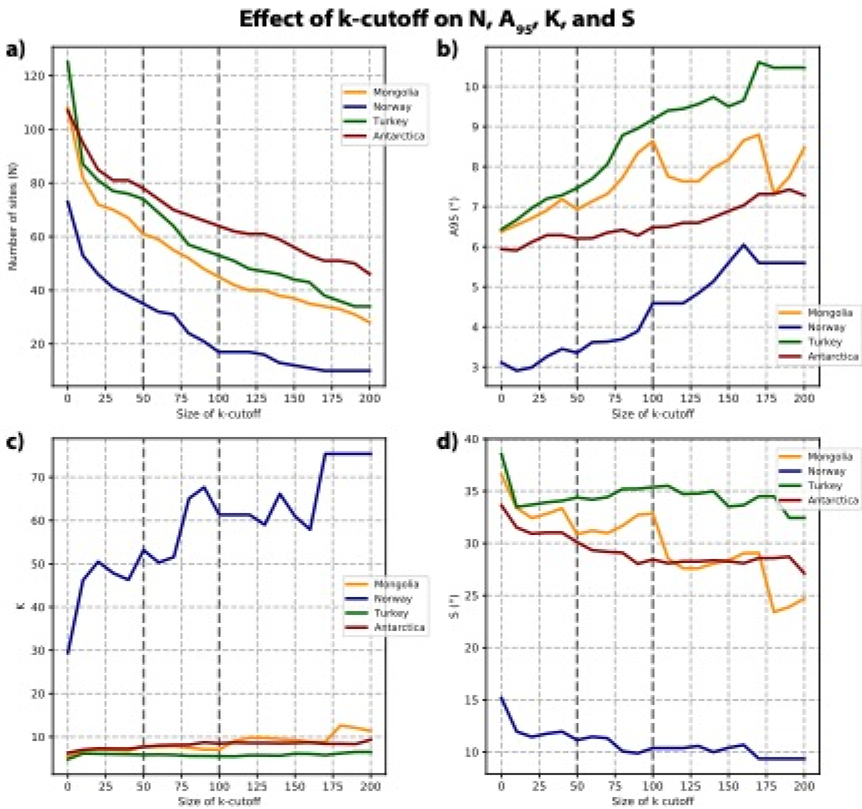

Figure 3. Effect of the size of the k-cutoff on a) the number of sites (N) remaining in the dataset, and the b) A_{95}, c) K, and d) S of the acquired paleopole. Commonly applied cutoff sizes of 50 and 100 are indicated with gray dashed lines.

4.2 Effect of the number of samples per site (n)

Next, we analyze the benefits of collecting multiple samples per site to average within-site or between-sample errors. To this end, we studied the influence of the number of samples per site (n) on the mean paleopole of a dataset (Figure 4) and its angular distance to the reference pole with n=7 (or 6, in case of Antarctica), A_{95}, K, and S (Figure 5). For each n, we performed the calculations 1000 times, each repetition collecting n different samples from the total number of samples per site. The paleopole position calculated with n=1 falls within the A_{95} of the mean pole calculated with n=7 in 100% of the calculations for the datasets from Mongolia, Turkey, and Antarctica (Figure 4). For the Norwegian dataset, this is the case for 91% of the calculations. In other words, the pole position is barely influenced by the number of samples per site (Figure 5a) and stays within the 95% confidence interval. The influence on the A_{95} is small, but there is a slight decrease in the A_{95} when within-site scatter is averaged by increasing n (Figure 5b). This varies from less than 0.5° to 1.5°, with the largest effect occurring for the dataset from Norway, whose sites have the largest within-site scatter. We find that the number of samples per site has a minor influence on K and S (Figure 5c, d). Only Norway shows an increase in K from approximately 14 to 29 and in S from 23° to 16°. This experiment shows that the effect of the number of samples per site on determining a pole position is surprisingly small, even if between-site scatter is somewhat decreasing.

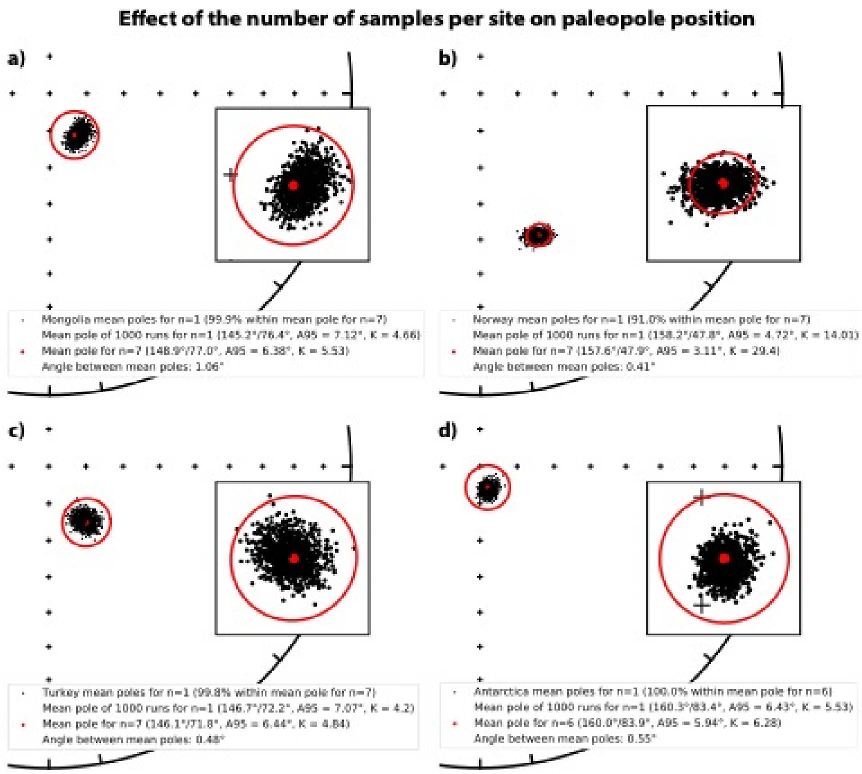

Figure 4. Stereographic projection with in black 1000 mean paleopole positions for n=1 and in red the mean paleopole position and its A_{95} for n=7 (or 6, in case of Antarctica) for **a)** Mongolia, **b)** Norway,**c)** Turkey, and **d)** Antarctica. Inset shows a zoom-in on the data.

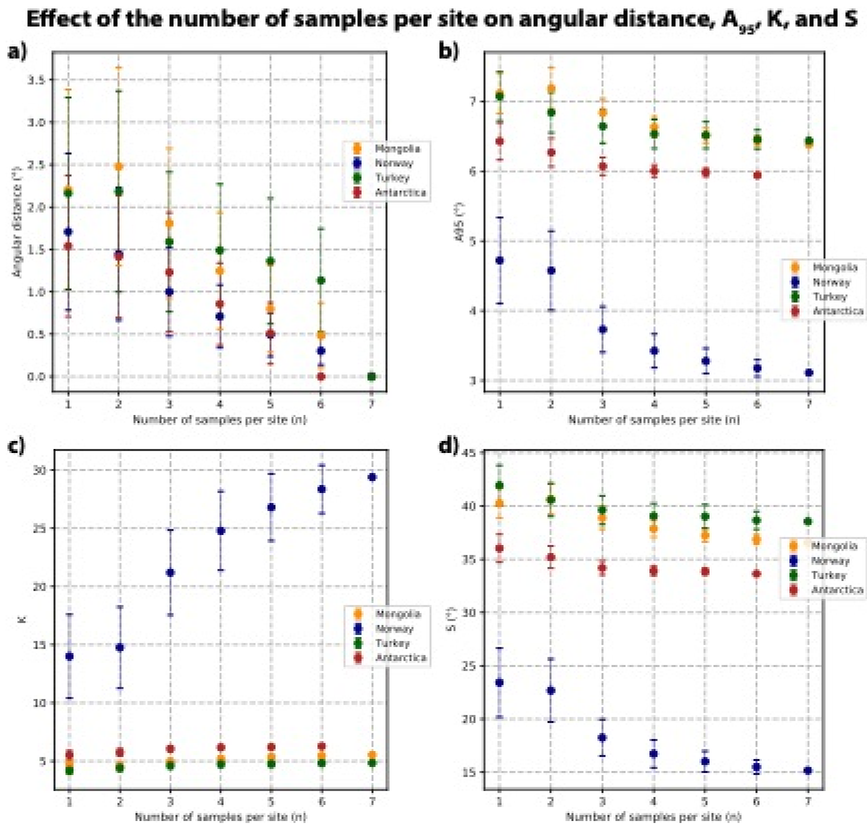

Figure 5. Effect of the number of samples per paleomagnetic site (n) on the **a)** angular distance of the acquired paleopole to the reference pole with n=7 (or 6, in case of Antarctica), and the **b)** A_{95},**c)** K, and **d)** S of the acquired paleopole. Every datapoint represents 1000 simulations, with error bars indicating the 95% confidence interval.

4.3 Effect of discarding outliers on site level

One potential benefit of having multiple samples per site is to discard outliers, commonly done in the paleomagnetic community based on the 'expert eye' of the interpreter. Here, we objectively studied the effect of discarding outliers on site level, by eliminating the farthest outlying sample(s) from the site. We did this three times, each time discarding the sample furthest from the recalculated site mean. Interestingly, we find generally no effect of eliminating within-site outliers on mean paleopole position and its A_{95}, K, and S (Figure 6), and for the Norway dataset, the between-site scatter even increases. This is because VGP positions do not systematically shift towards or away from the paleopole (Figures S2-S5). As a result, decreasing the within-site scatter by removing outlying directions has little effect on the between-site dispersion of VGPs.

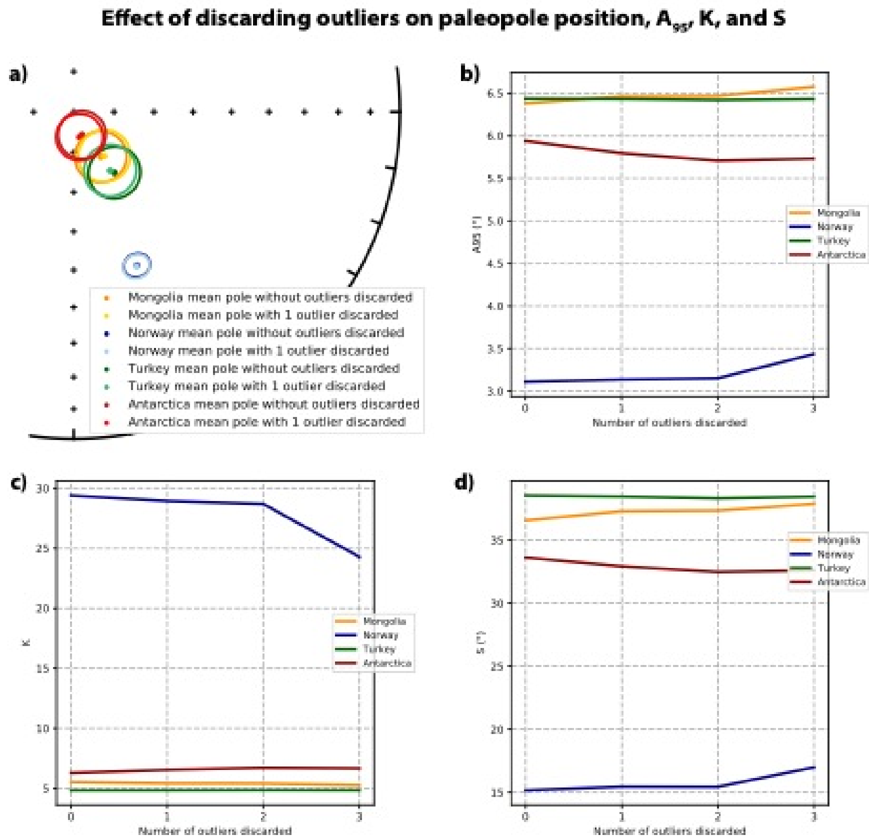

Figure 6. Effect of the number of samples discarded per paleomagnetic site on a) paleopole position in stereographic projection, shown without outliers discarded and with one outlier discarded, and the b) A_{95}, c) K, and d) S of the acquired paleopole.

4.4 Effect of sample distribution over sites

To average between-site scatter, it is common to collect multiple sites at a locality. The minimum number of sites recommended for determining a paleopole varies between authors (e.g., Meert et al., 2020; Van der Voo, 1990), but as shown by Vaes et al. (2021), poles with larger N tend to plot closer to a best estimate of the time-averaged pole. We performed experiments varying the number of samples per site and the number of sites, with combinations chosen such that the total amount of paleomagnetic samples remained approximately 100. We studied the effect on angular distance to the reference pole, A_{95}, K, and S (Figure 7) of the acquired paleopole. The means were calculated from 1000 runs of the specific number of samples per site and number of sites, where the number of samples and sites was randomly selected from the total population during each run. The paleopole position is strongly influenced by the distribution of samples over sites in case of Mongolia, Turkey, and Antarctica, and lies further from the reference pole if the samples are distributed over fewer sites (Figure 7a). There also is a major effect on the A_{95}. For all datasets, the A_{95} is by far the smallest when it is constrained by 100 sites of $n = 1$, and the highest when it is constrained by 15 sites of $n=7$ (Figure 7b). The effect on K and S is small, except in the dataset from Norway (Figure 7c, d). In that dataset, K is highest and S is lowest for 15 sites with $n=7$. Nonetheless, the paleopole position is far better determined when taking many sites with few samples.

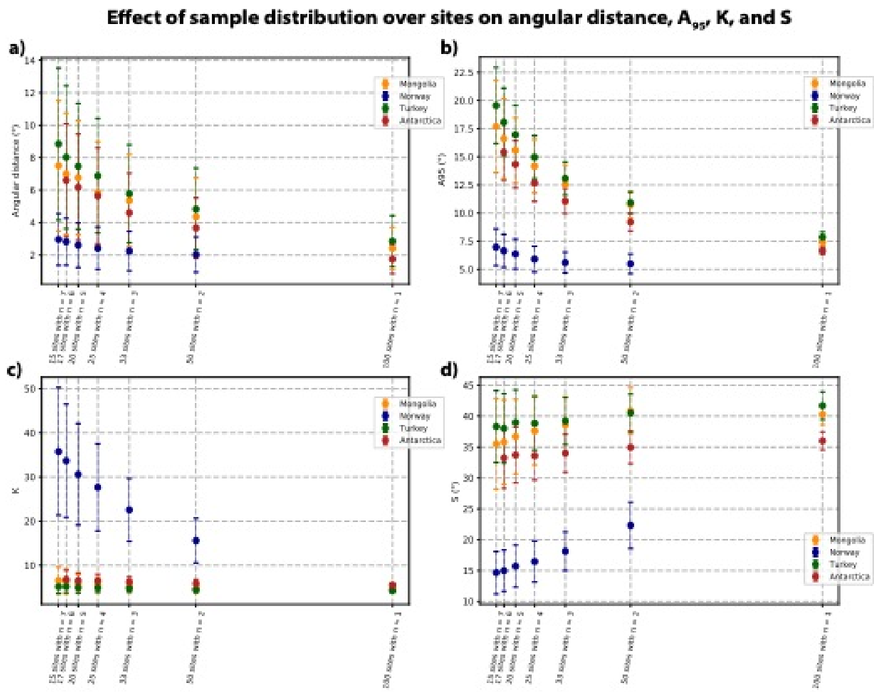

Figure 7. Effect of distributing 100 paleomagnetic samples over a varying number of paleomagnetic sites on the a) angular distance of the acquired paleopole to the reference pole with all sites with n=7 (or 6, in case of Antarctica), and the b) A_{95}, c) K, and d) S of the acquired paleopole. Every datapoint represents 1000 runs of a distribution, with error bars indicating the 95% confidence interval of these 1000 runs.

5 Discussion and conclusions

We first evaluate whether measuring multiple samples per site to constrain within-site scatter allows to objectively filter unreliable VGPs. In our reinterpretation, we closed our 'expert eye' and included sites that were excluded in the original studies because of subjective criteria: for instance, our dataset contains sites with erratic demagnetization behavior or very high intensities that were originally interpreted as lightning struck. Our analysis illustrates that objective statistical filters, such as a k-cutoff, a minimum number of samples per site, or eliminating the farthest outliers per site, are incapable of filtering these unreliable VGPs. The subjective judgement of the expert thus determines which samples and sites are included in the calculation of the paleopole (and makes it important that all data are made available following the FAIR principles (Wilkinson et al., 2016)).

Our analysis next shows that for the three datasets with high between-site scatter and low within-site scatter (i.e., Turkey, Mongolia, and Antarctica), applying the within-site scatter-based criteria leads to marginal increase in data clustering (i.e., an increase in K and decrease in S). For the Norwegian dataset, which has the rare combination of high within-site scatter and low between-site scatter, these criteria even lead to significantly higher clustering of the data. This lends support to using these data filters for PSV studies.

But applying these filters when calculating paleopoles in all cases leads to a decrease in pole precision (i.e., an increase in A_{95}). This is the result of the decrease in N that is caused by applying data filters. Our experiments thus show that the effect of between-site scatter on pole precision is far dominant over within-site scatter. And because the accuracy of paleopoles also decreases with decreasing N (Vaes et al., 2021), applying within-site scatter filters systematically decreases pole accuracy as well as precision (see also Figure 7).

Averaging multiple samples per site, without applying filters, does not negatively influence pole accuracy or precision. In case only few sites are available, or there is reason to question data quality, collecting multiple samples per site, as recommended by Meert et al. (2021), may still be a useful strategy. It may also be useful to demonstrate for a selection of lava sites in a study that within-site precision (k) is high, as a field test equivalent to a fold, reversal, or conglomerate test. But, our analysis shows that within-site scatter does not significantly influence paleopole accuracy or precision. The extra efforts put into collecting multiple samples per site are thus more effectively spent collecting more single-sample sites.

Acknowledgements

BV and DJJvH acknowledge NWO Vici grant 865.17.001 to DJJvH.

Data Availability Statement

All uninterpreted paleomagnetic data are publicly available on Paleomagnetism.org for the Mongolia, Norway, and Turkey dataset and in the MagIC database for the Antarctica dataset: https://earthref.org/MagIC/17076. The Python code used for our analysis will be made public on GitHub upon acceptance of this manuscript.

References

Asefaw, H., Tauxe, L., Koppers, A. A. P., & Staudigel, H. (2021). Four-Dimensional Paleomagnetic Dataset: Plio-Pleistocene Paleodirection and Paleointensity Results From the Erebus Volcanic Province, Antarctica. *Journal of Geophysical Research: Solid Earth* ,*126* (2), 1–22. https://doi.org/10.1029/2020JB020834

Biggin, A. J., van Hinsbergen, D. J. J., Langereis, C. G., Straathof, G. B., & Deenen, M. H. L. (2008). Geomagnetic secular variation in the Cretaceous Normal Superchron and in the Jurassic. *Physics of the Earth and Planetary Interiors* , *169* (1–4), 3–19. https://doi.org/10.1016/j.pepi.2008.07.004

Brandt, D., Ernesto, M., & Constable, C. (2021). Consistent and Contrasting Aspects of the Geomagnetic Field Across Epochs With Distinct Reversal Frequencies Revealed by Modeling the Kiaman Superchron.*Geochemistry, Geophysics, Geosystems* , *22* (7), 1–24. https://doi.org/10.1029/2021GC009866

Butler, R. F. (1992). *Paleomagnetism: Magnetic domains to geologic terranes* (Vol. 319). Boston: Blackwell Scientific Publications.

Cromwell, G., Johnson, C. L., Tauxe, L., Constable, C. G., & Jarboe, N. A. (2018). PSV10: A Global Data Set for 0–10 Ma Time-Averaged Field and Paleosecular Variation Studies. *Geochemistry, Geophysics, Geosystems* , *19* (5), 1533–1558. https://doi.org/10.1002/2017GC007318

Fisher, R. A. (1953). Dispersion on a sphere. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* ,*217* (1130), 295–305. https://doi.org/10.1098/ rspa.1953.0064

Haldan, M. M., Meijers, M. J. M., Langereis, C. G., Larsen, B. T., & Heyer, H. (2014). New palaeomagnetic results from the Oslo Graben, a Permian Superchron lava province. *Geophysical Journal International* , *199* (3), 1554–1571. https://doi.org/10.1093/gji/ggu351

Handford, B. T., Biggin, A. J., Haldan, M. M., & Langereis, C. G. (2021). Analysing Triassic and Permian Geomagnetic Palaeosecular Variation and the Implications for Ancient Field Morphology.*Geochemistry, Geophysics, Geosystems* , *In press* . https://doi.org/10.1029/2021GC009930

van Hinsbergen, D. J. J., Straathof, G. B., Kuiper, K. F., Cunningham, W. D., & Wijbrans, J. (2008). No vertical axis rotations during Neogene transpressional orogeny in the NE Gobi Altai: Coinciding Mongolian and Eurasian early Cretaceous apparent polar wander paths. *Geophysical Journal International* , *173* (1), 105–126. https://doi.org/10.1111/j.1365-246X.2007.03712.x

van Hinsbergen, D. J. J., Dekkers, M. J., Bozkurt, E., & Koopman, M. (2010). Exhumation with a twist: Paleomagnetic constraints on the evolution of the Menderes metamorphic core complex, western Turkey.*Tectonics* , *29* (3), 1–33. https://doi.org/10.1029/2009TC002596 van Hinsbergen, D. J. J., De Groot, L. V., Van Schaik, S. J., Spakman, W., Bijl, P. K., Sluijs, A., et al. (2015). A paleolatitude calculator for paleoclimate studies. *PLoS ONE* , *10* (6), 1–21. https://doi.org/10.1371/journal.pone.0126946

Johnson, C. L., & Constable, C. G. (1996). Palaeosecular variation recorded by lava flows over the past five million years.*Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* , *354* (1704), 89–141. https://doi.org/10.1098/rsta.1996.0004

Johnson, C. L., Constable, C. G., Tauxe, L., Barendregt, R., Brown, L. L., Coe, R. S., et al. (2008). Recent investigations of the 0-5 Ma geomagnetic field recorded by lava flows. *Geochemistry, Geophysics, Geosystems* , *9* (4). https://doi.org/10.1029/2007GC001696

Kirschvink, J. L. (1980). The least-squares line and plane and the analysis of palaeomagnetic data. *Geophysical Journal of the Royal Astronomical Society* , *62* (3), 699–718. https://doi.org/10.1111/j.1365-246X.1980.tb02601.x

Koymans, M. R., Langereis, C. G., Pastor-Galán, D., & van Hinsbergen, D. J. J. (2016). Paleomagnetism.org: An online multi-platform open source environment for paleomagnetic data analysis. *Computers and Geosciences* , *93* , 127–137. https://doi.org/10.1016/j.cageo.2016.05.007

Koymans, M. R., van Hinsbergen, D. J. J., Pastor-Galán, D., Vaes, B., & Langereis, C. G. (2020). Towards FAIR Paleomagnetic Data Management Through Paleomagnetism.org 2.0. *Geochemistry, Geophysics, Geosystems* , *21* (2), 1–7. https://doi.org/10.1029/2019GC008838

Lippert, P. C., van Hinsbergen, D. J. J., & Dupont-Nivet, G. (2014). Early Cretaceous to present latitude of the central proto-Tibetan Plateau: A paleomagnetic synthesis with implications for Cenozoic tectonics, paleogeography, and climate of Asia. *Special Paper of the Geological Society of America* , *507* , 1–21. https://doi.org/10.1130/2014.2507(01)

McElhinny, M. W., & McFadden, P. L. (1999). *Paleomagnetism: continents and oceans* . Elsevier.

McFadden, P. L., & McElhinny, M. W. (1988). The combined analysis of remagnetization circles and direct observations in palaeomagnetism.*Earth and Planetary Science Letters* , *87* (1–2), 161–172. https://doi.org/10.1016/0012-821X(88)90072-6

Meert, J. G., Pivarunas, A. F., Evans, D. A. D., Pisarevsky, S. A., Pesonen, L. J., Li, Z. X., et al. (2020). The magnificent seven: A proposal for modest revision of the Van der Voo (1990) quality index.*Tectonophysics* , *790* (March), 228549. https://doi.org/10.1016/j.tecto.2020.228549

Merrill, R. T., & McFadden, P. L. (2003). The geomagnetic axial dipole field assumption. *Physics of the Earth and Planetary Interiors* ,*139* (3–4), 171–185. https://doi.org/10.1016/j.pepi.2003.07.016

de Oliveira, W. P., Hartmann, G. A., Terra-Nova, F., Brandt, D., Biggin, A. J., Engbers, Y. A., et al. (2021). Paleosecular variation and the time-averaged geomagnetic field since 10 Ma. *Geochemistry, Geophysics, Geosystems* . https://doi.org/10.1029/2021gc010063

Tauxe, L. (2010). *Essentials of paleomagnetism* . University of California Press.

Tauxe, L., Constable, C., Johnson, C. L., Koppers, A. A. P., Miller, W. R., & Staudigel, H. (2003). Paleomagnetism of the southwestern U.S.A. recorded by 0-5 Ma igneous rocks. *Geochemistry, Geophysics, Geosystems* , *4* (4). https://doi.org/10.1029/2002GC000343

Tauxe, L., Shaar, R., Jonestrask, L., Swanson-Hysell, N. L., Minnett, R., Koppers, A. A. P., et al. (2016). PmagPy: Software package for paleomagnetic data analysis and a bridge to the Magnetics Information Consortium (MagIC) Database. *Geochemistry, Geophysics, Geosystems* , *17* (6), 2450–2463. https://doi.org/10.1002/2016GC006307

Torsvik, T. H., Van der Voo, R., Preeden, U., Mac Niocaill, C., Steinberger, B., Doubrovine, P. V., et al. (2012). Phanerozoic Polar Wander, Palaeogeography and Dynamics. *Earth-Science Reviews* ,*114* (3–4), 325–368. https://doi.org/10.1016/j.earscirev.2012.06.007

Uzel, B., Langereis, C. G., Kaymakci, N., Sozbilir, H., Ozkaymak, C., & Ozkaptan, M. (2015). Paleomagnetic evidence for an inverse rotation history of Western Anatolia during the exhumation of Menderes core complex. *Earth and Planetary Science Letters* , *414* , 108–125. https://doi.org/10.1016/j.epsl.2015.01.008

Uzel, B., Sumer, O., Ozkaptan, M., Ozkaymak, C., Kuiper, K., Sozbilir, H., et al. (2017). Palaeomagnetic and geochronological evidence for a major middle miocene unconformity in Soke Basin (western Anatolia) and its tectonic implications for the Aegean region. *Journal of the Geological Society* , *174* (4), 721–740. https://doi.org/10.1144/jgs2016-006

Vaes, B., Gallo, L., and van Hinsbergen, D.J.J. (2021). On pole position: causes of dispersion of the paleomagnetic poles behind apparent polar wander paths, *Journal of Geophysical Research, submitted.*

van der Voo, R. (1990). The reliability of paleomagnetic data. *Tectonophysics* , *184* (1), 1–9. https://doi.org/10.1016/0040-1951(90)90116-P

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., et al. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* ,*3* , 1–9. https://doi.org/10.1038/sdata.2016.18

Hosted file

`supporting information.docx` available at https://authorea.com/users/524610/articles/595513-influence-of-data-filters-on-the-accuracy-and-precision-of-paleomagnetic-poles-what-is-the-optimal-sampling-strategy

Influence of data filters on the accuracy and precision of paleomagnetic poles: what is the optimal sampling strategy?

Dieke Gerritsen[1], Bram Vaes[1], and Douwe J.J. van Hinsbergen[1]

[1]Department of Earth Sciences, Utrecht University, Princetonlaan 8A, 3584 CB Utrecht, the Netherlands.

Corresponding authors: Dieke Gerritsen (h.gerritsen1@uu.nl) and Bram Vaes (b.vaes@uu.nl)

Key Points:

- Within-site data filters (minimum k or n, or eliminating outliers per site) are unsuccessful in eliminating unreliable data.
- The precision of paleomagnetic poles is dominated by between-site scatter; within-site scatter has minimal contribution.
- For paleopoles, efforts put into collecting multiple samples per site are more effectively spent on collecting more single-sample sites.
-

Abstract

To determine a paleopole, the paleomagnetic community commonly applies a loosely defined set of quantitative data filters that were established for studies of geomagnetic field behavior. These filters require costly and time-consuming sampling procedures, but whether they improve accuracy and precision of paleopoles has not yet been systematically analyzed. In this study, we performed a series of experiments on four datasets which consist of 73-125 lava sites with 6-7 samples per lava. The datasets are from different regions and ages, and are large enough to represent paleosecular variation, yet contain demonstrably unreliable datapoints. We show that data filters based on within-site scatter (a k-cutoff, a minimum number of samples per site, and eliminating the farthest outliers per site) cannot objectively identify unreliable directions. We find instead that excluding unreliable directions relies on the subjective interpretation of the expert, highlighting the importance of making all data available following the FAIR principles. In addition, data filters that eliminate datapoints even have an adverse effect: the accuracy as well as the precision of paleopoles decreases with the decreasing number of data. Between-site scatter far outweighs within-site scatter, and when collecting paleomagnetic poles, the extra efforts put into collecting multiple samples per site are more effectively spent on collecting more single-sample sites.

1 Introduction

Paleomagnetic poles, or paleopoles, quantify the past position of rocks relative to the geomagnetic pole and constrain tectonic reconstructions and apparent polar wander paths (APWPs) (e.g., Besse and Courtillot, 2002; Torsvik et al., 2012). The calculation of paleopoles relies on the assumption that the time-averaged geomagnetic field approximates a geocentric axial dipole (GAD), but is complicated by short-term deviations from this field (e.g., Cromwell et al., 2018; Oliveira et al., 2021) known as paleosecular variation (PSV). To obtain a paleopole, paleomagnetists therefore average virtual geomagnetic poles (VGPs), whereby every VGP is then assumed a 'spot reading': an instantaneous reading of the past geomagnetic field collected from a rock unit ('site') that represents an increment of geological time, such as a lava (Butler, 1992; Tauxe, 2010). However, not every VGP represents an accurate spot reading because artifacts may be introduced by measuring errors or remagnetization. Therefore, the paleomagnetic community commonly uses a set of data filters to acquire a set of reliable spot readings. However, these filters vary between authors and were not determined by studies aiming to constrain paleopoles.

The studies that established the data filters, investigated PSV and geomagnetic field behavior by determining the between-site scatter of a set of VGPs (e.g., Cromwell et al., 2018; de Oliveira et al., 2021; Johnson et al., 2008; Johnson & Constable, 1996; Tauxe et al., 2003). To this end, these studies aim to correct for within-site scatter induced by measuring errors and typically require a minimum number of readings per site, although this number of readings varies between authors (e.g., Biggin et al., 2008; Johnson et al., 2008; Doubrovine et al., 2019; Cromwell et al., 2018). The resulting paleomagnetic directions are then averaged to a site-mean direction which is converted to a VGP if the site passes a criterion for the minimum within-site precision value ('cutoff'), typically expressed as a Fisher (1953) precision parameter k. This value also varies between authors, e.g., $k \geq 50$ or $k \geq 100$ (e.g., Biggin et al., 2008; Cromwell et al., 2018; Johnson et al., 2008; Tauxe et al., 2003). Subsequently, similar procedures have become common for calculation of paleopoles (e.g., Butler, 1992; Lippert et al., 2014; Meert et al., 2020). However, does this time- and data-intensive procedure improve the accuracy and precision of paleopoles?

In this study, we analyze to what extent commonly applied paleomagnetic data filters established for PSV studies improve the accuracy and precision of paleopoles. To this end, we study four large paleomagnetic datasets obtained from lava sequences from the Cretaceous of Mongolia, the Permian of Norway, the Miocene of Turkey, and the Quaternary of Antarctica. These datasets are large enough to represent PSV, but contain additional between-site and within-site scatter of varying magnitude due to measurement errors, lightning-induced remagnetization, and/or tectonic deformation. We perform a series of experiments to examine the effects of applying filters on the accuracy and precision of paleopoles. We evaluate whether these filters can objectively exclude outliers and filter non-PSV induced scatter from the paleomagnetic datasets. We then assess how a given number of paleomagnetic directions is optimally distributed over a collection of paleomagnetic sites to acquire the best-constrained paleopole position. We discuss to what extent the filters used in PSV studies improve paleopole accuracy and precision. Our results aim to aid paleomagnetists to optimize their sampling and data filtering strategies.

2 Background

To evaluate the reliability of a paleopole, it is common to use filters to obtain a paleomagnetic dataset representative of PSV. Initially, Van der Voo (1990) recommended to average a minimum set of 25 samples or sites (i.e., spot readings) for a paleopole, and formulated loosely defined filters using Fisher (1953) statistics. Those statistics are used to describe a paleomagnetic dataset and can be applied to either directional data or VGPs. Three parameters are of importance in this statistical approach: the radius of the 95% confidence cone around the mean ($_{\mathbf{95}}$ for directions and A_{95} for VGPs), the precision parameter (k for directions and K for VGPs), and the circular standard deviation (S) that quantifies the angular between-site dispersion of VGPs. Van der Voo (1990) suggested filters for these parameters: $_{\mathbf{95}}$ or $A_{95} \leq 16°$, and k or $K \geq 10$. In later studies, filters were added to determine the reliability of each VGP based on quality filters used in PSV studies (e.g., Lippert et al., 2014).

For obtaining paleopoles, paleomagnetists usually sample multiple sites on a locality, where every site likely represents a spot reading of the paleomagnetic field, e.g., a lava flow. The total scatter in a dataset then consists of within-site and between-site scatter (Biggin et al., 2008). Paleomagnetists collect multiple samples per site to test whether a paleomagnetic direction is reproducible and to average measurement and other random errors (McElhinny & McFadden, 1999). A quality check for within-site precision is a k-cutoff (Tauxe et al., 2003), where sites with a *k* value lower than an arbitrary value of 50 or 100 are discarded (e.g., Biggin et al., 2008; Johnson et al., 2008; Lippert et al., 2014). Furthermore, outliers are often subjectively discarded from sites based on the 'expert eye' and experience of the interpreter. How many samples should be collected per site is not widely agreed upon. Based on Monte Carlo simulations, a minimum number of five independently oriented samples was deemed necessary to estimate *k* reliably (Tauxe et al., 2003). Others suggest that three (Meert et al., 2020), four (Cromwell et al., 2018; Lippert et al., 2014), or six (Asefaw et al., 2021) samples are needed. But most paleomagnetic studies collected six to eight samples per lava site.

To average between-site scatter, assumed predominantly the result of PSV, multiple sites are collected, but the minimum number of sites required for a 'good' average, that is, the pole, is not well defined (Tauxe, 2010). The precision with which a paleopole is calculated increases with increasing number of underlying VGPs (Vaes et al., 2021). Meert et al. (2020) suggested eight sites are sufficient to determine a paleopole, provided that each site is constrained by at least three samples. However, based on a statistical simulation of PSV, Tauxe et al. (2003) showed that a minimum of approximately 100 sites may be required to fully sample PSV. However, this number is rarely obtained due to limited availability of sites and because resources are consumed by sampling a high number of samples for each site. Below, we analyze the effect of these different sampling and filtering strategies on the accuracy and precision of the resulting paleopole.

3 Datasets and Approach

3.1 Datasets

As basis for our analysis, we use four large paleomagnetic datasets derived from lavas from Mongolia (van Hinsbergen et al., 2008), Norway (Haldan et al., 2014), Turkey (van Hinsbergen et al., 2010), and Antarctica (Asefaw et al., 2021). The results were previously published, but for the purpose of this study, we reinterpreted the demagnetization diagrams of all samples. We did this by identifying the characteristic remanent magnetization (ChRM) by principal component analysis (Kirschvink, 1980), using the online paleomagnetic analysis platform Paleomagnetism.org (Koymans et al., 2016, 2020). We have thereby not forced interpreted components through the origin, and we have not used the remagnetization great-circle method of McFadden & McElhinny (1988), as was occasionally done in the original interpretations. All interpretable diagrams were interpreted, and no samples were discarded based on obviously outlying directions, e.g., due to lightning strikes. In other words, we deliberately kept directions that an experienced paleomagnetist would likely immediately discard as unreliable. We also did not exclude sites that were collected from tectonically deformed regions. As a result, the dataset contains larger noise and a higher between-site scatter (lower K-value) than the published values. This allows us to assess whether the objective quality criteria alone can clean the dataset from outliers, or whether this relies on a subjective 'expert eye'. The sites from Mongolia, Norway, and Turkey were sampled with a minimum of seven samples per site, and the sites from Antarctica with a minimum of six. Only sites with seven (or six, for Antarctica) interpretable directions have been used in our analysis. If sites contained more samples, only the first seven (or six, for Antarctica) samples were used. All uninterpreted data are publicly available in databases (Paleomagnetism.org (Koymans et al., 2020) or MagIC (Tauxe et al., 2016), see Data Availability Statement). Our reinterpretations are provided in the supplementary information (Table S1) to this paper, but because our interpretations are (deliberately) not better than the original interpretations, we will not upload these in the databases to avoid confusion. All paleopole positions from this study differ from the published pole positions, but still sit within the 95% confidence cone of the respective published pole (Figure 1).

After the reinterpretation, we arrived at a total of 108 sites of lower Cretaceous lavas, corresponding to the base of the Cretaceous Normal Superchron, from the NE Gobi Altai mountain range of southern Mongolia (van Hinsbergen et al., 2008; vH08). We did not include samples from the Artz Bogd area from the original dataset because these appear systematically rotated due to tectonic deformation. Furthermore, only the sites acquired by van Hinsbergen et al. (2008) are included. Our reinterpreted dataset (G21) leads to a higher scatter than the original interpretation of van Hinsbergen et al. (2008) for reasons outlined above ($A_{95,vH08} = 5.3°$ vs. $A_{95,G21} = 6.4°$, $K_{vH08} = 9.1$ vs. $K_{G21} = 5.5$). Most sites have limited within-site scatter and high k-values, as expected for lava sites.

We arrived at a total of 73 sites from lavas of Permian age from the Oslo Graben in Norway (Haldan et al., 2014; H14). This dataset contains lavas from the Permo-Carboniferous Reversed Superchron (PCRS). Interestingly, the between-site scatter is much lower than for the other three localities, as addressed by (Brandt et al., 2021; Handford et al., 2021). The within-site scatter of this dataset, however, is higher. Our reinterpretation contains more scatter than the interpretation of Haldan et al. (2014) ($A_{95,H14} = 1.9°$ vs. $A_{95,G21} = 3.1°$, $K_{H14} = 52.2$ vs. $K_{G21} = 29.4$).

A total of 125 sites of lower to middle Miocene lavas and ignimbritic tuffs come from basins in the northern Menderes Massif in western Turkey (van Hinsbergen et al., 2010; vH10). This dataset was acquired in a tectonically active area and contains basins that were interpreted to be tectonically coherent, and basins that were interpreted by van Hinsbergen et al. (2010) to be internally tectonically disturbed (later confirmed and mapped out in detail by Uzel et al. (2015, 2017)). It also contained sites acquired by preceding studies, but only the sites acquired by van Hinsbergen et al. (2010) are included in our dataset. Our re-interpreted dataset contains an equal amount of scatter as the dataset of van Hinsbergen et al. (2010) ($A_{95,vH10} = 6.7°$ vs. $A_{95,G21} = 6.4°$, $K_{vH10} = 4.6$ vs. $K_{G21} = 4.8$).

Finally, we used 107 sites with n=6 from lavas of Plio-Pleistocene age of the Erebus volcanic province in Antarctica (Asefaw et al., 2021; A21). We did not reinterpret this dataset and used the published interpretations, but only used sites with n=6. Furthermore, we did not use the age constraint of <5 Ma and k-cutoff. Our reinterpretation does not significantly differ from the published one of Asefaw et al. (2021) ($A_{95,A21} = 5.5°$ vs. $A_{95,G21} = 5.9°$, $K_{A21} = 7.7$ vs. $K_{G21} = 6.3$).

3.2 Approach

We performed a series of experiments to study the effect on paleopole accuracy and precision of objective filters that one may use when multiple samples per site are available. These filters include the influence of (i) a k-cutoff; (ii) the number of samples per site (n), and (iii) discarding outliers of within-site dispersion. In addition, we assessed how a given number of paleomagnetic directions may be optimally distributed over a collection of paleomagnetic sites (N) to acquire the most best-constrained paleopole position. This will be addressed using Fisher (1953) parameters: the radius of the 95% confidence cone around the paleopole (A_{95}), the precision parameter (K), and the circular standard deviation (S). We performed experiments using a Python code, which we developed making extensive use of the freely available paleomagnetic software package PmagPy (Tauxe et al., 2016). We perform the calculations 1000 times, and each time different samples and/or sites will be selected from the population. We then calculate the mean and standard deviation of the different Fisher (1953) parameters. Our Python code is available on GitHub (see Data Availability Statement).

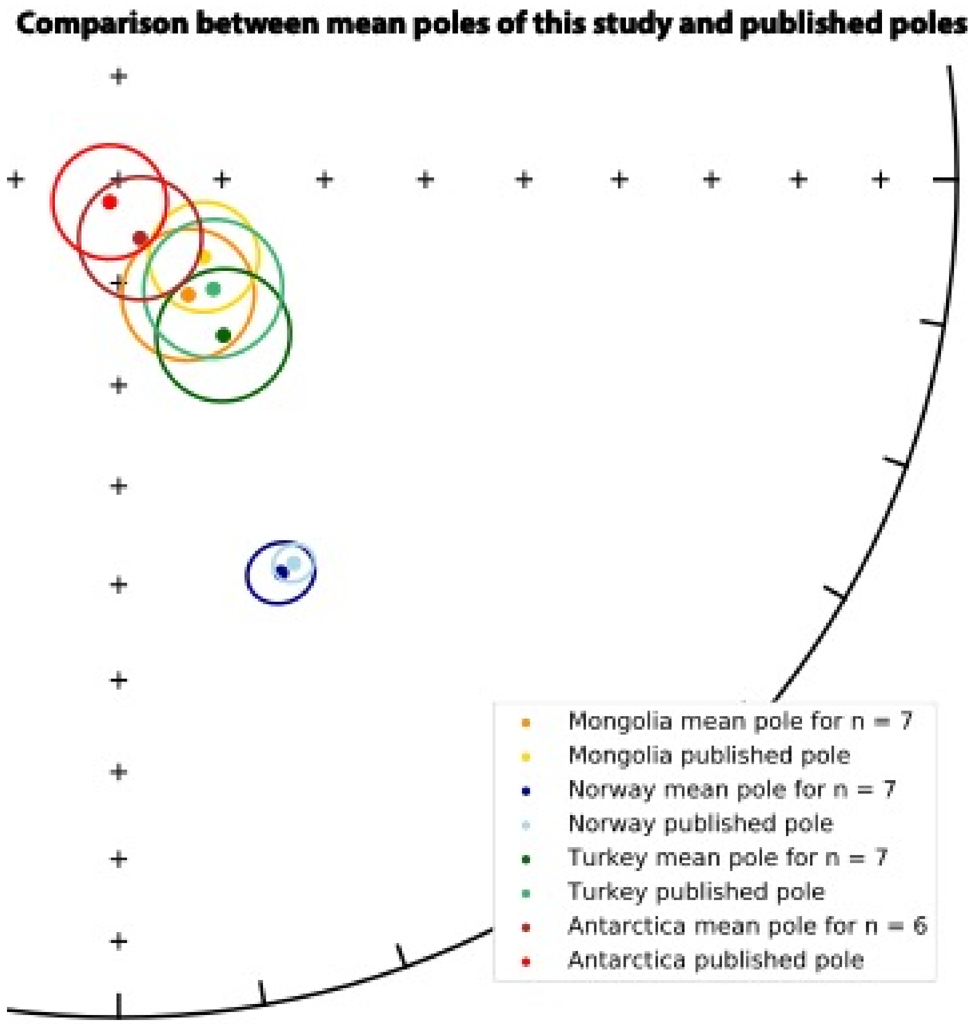

Figure 1. Stereographic projection showing the mean paleopole position and its A_{95} from this study and the published paleopole position and its A_{95} for Mongolia, Norway, Turkey, and Antarctica.

4 Results

4.1 Effect of a k-cutoff

A widely used filter to establish site-mean directions that represent spot readings of the paleomagnetic field is to check for high within-site scatter by means of a k-cutoff. Sites with low k values are interpreted as either remagnetized or not representative of a spot reading and are therefore discarded (Figure S1). We analyze the effect of the k-cutoff size on between-site scatter using the mean paleopole of a dataset (Figure 2) and its N, A_{95}, K, and S (Figure 3). The pole positions do not significantly change when applying a k-cutoff (Figure 2). When a low $k \geq 10$ cutoff is applied, the number of sites remaining in the dataset decreases by approximately 15-30% (Figure 3a), omitting sites with near-random direction distributions. Increasing the cutoff size beyond 20, the decay in sites decreases. The A_{95} increases with increasing k-cutoff due to a decrease in N (Figure 3b). As expected, discarding sites with high within-site scatter leads to a higher K and lower S, although the effects are small (Figure 3c, d), except in the dataset from Norway where K rises from 30 to 70. No major improvement becomes apparent in any of the variables at the commonly used cutoff sizes of 50 and 100. Importantly, although the application of a k-cutoff may subtly decrease S and increase K, the decrease in N leads to an increase of the A_{95}, and thus to a decrease in precision of the paleopole when a k-cutoff is applied, whereas the pole position remains approximately the same.

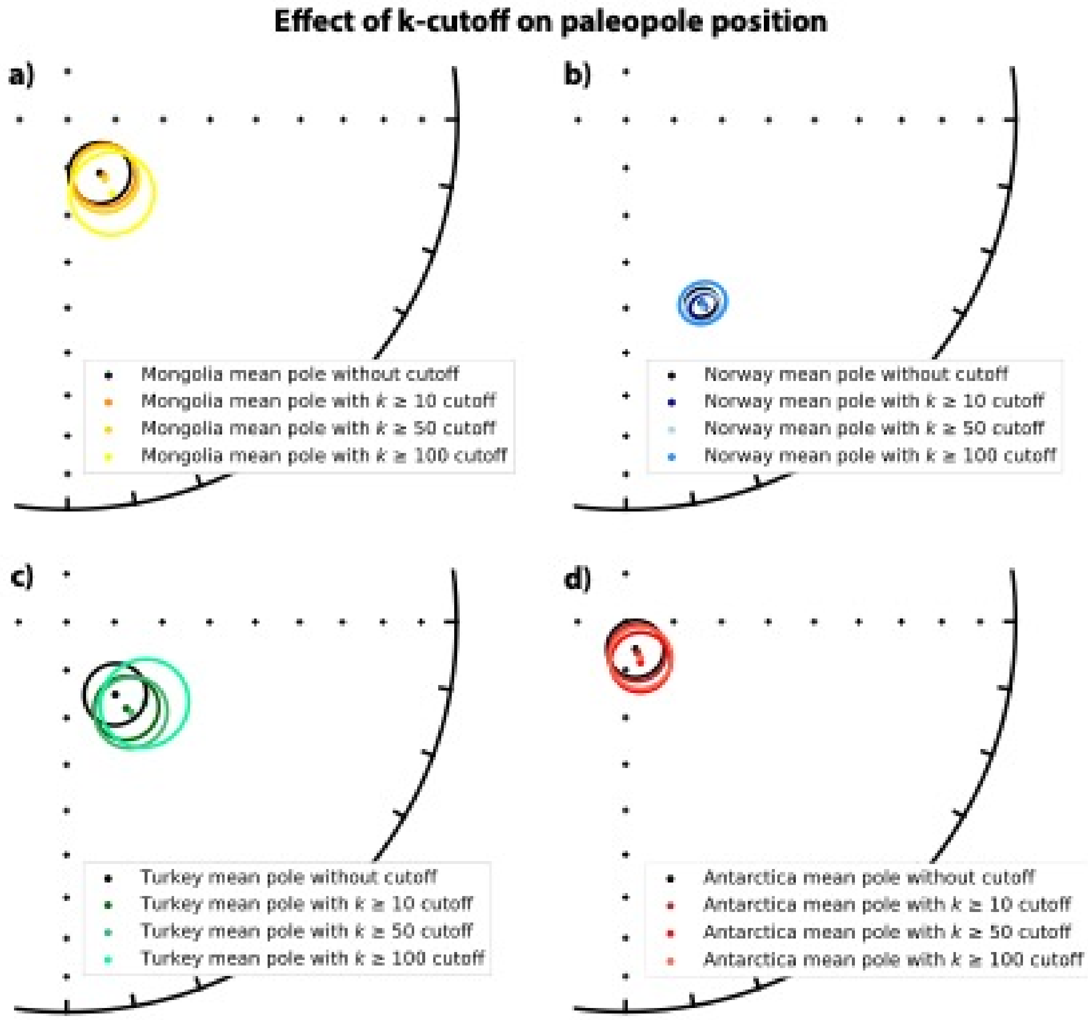

Figure 2. Stereographic projections illustrating the effect of the size of a k-cutoff on paleopole position and the A_{95} for the dataset from a) Mongolia, b) Norway, c) Turkey, and d) Antarctica.

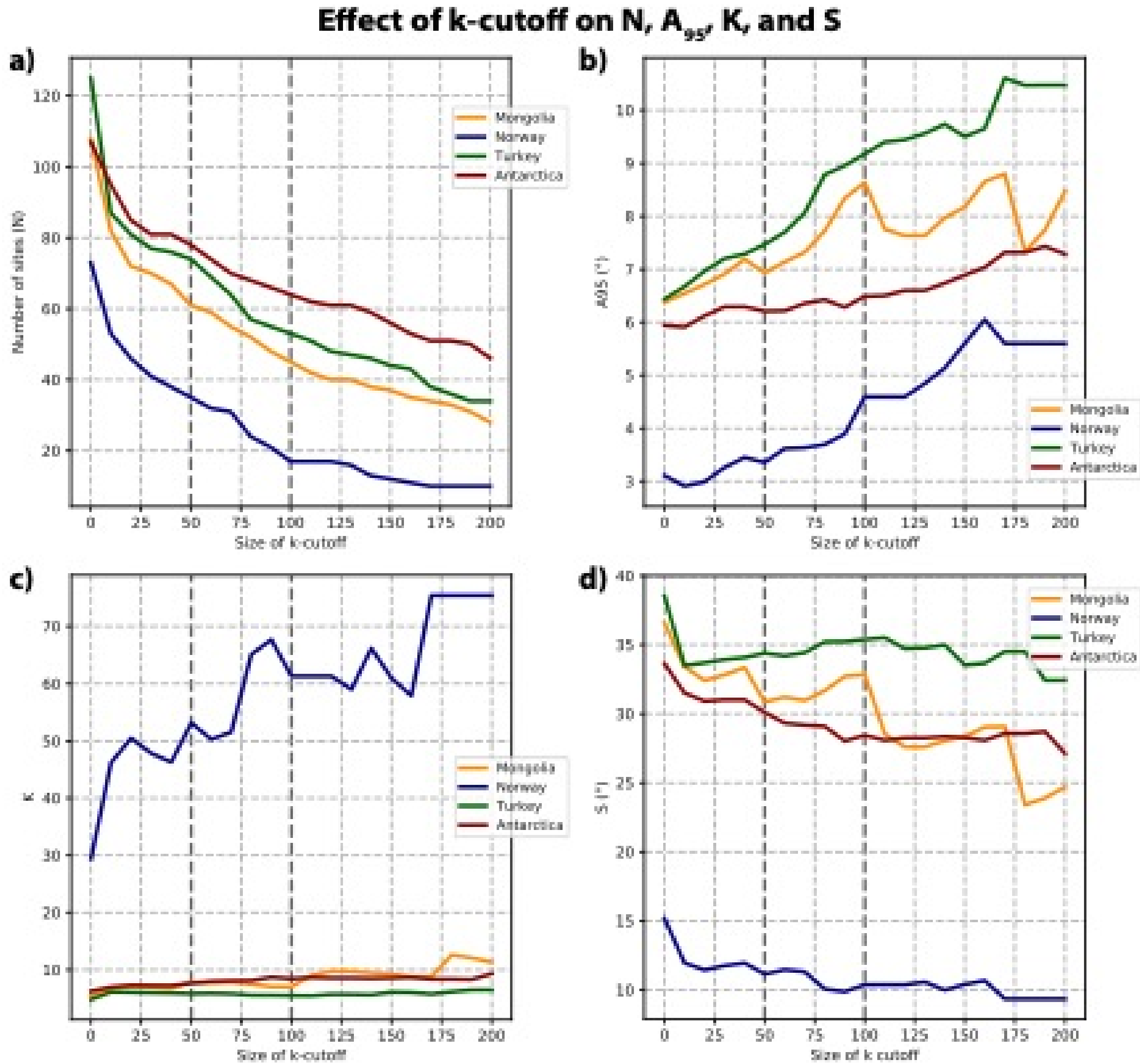

Figure 3. Effect of the size of the k-cutoff on a) the number of sites (N) remaining in the dataset, and the b) A_{95}, c) K, and d) S of the acquired paleopole. Commonly applied cutoff sizes of 50 and 100 are indicated with gray dashed lines.

4.2 Effect of the number of samples per site (n)

Next, we analyze the benefits of collecting multiple samples per site to average within-site or between-sample errors. To this end, we studied the influence of the number of samples per site (n) on the mean paleopole of a dataset (Figure 4) and its angular distance to the reference pole with n=7 (or 6, in case of Antarctica),

A_{95}, K, and S (Figure 5). For each n, we performed the calculations 1000 times, each repetition collecting n different samples from the total number of samples per site. The paleopole position calculated with n=1 falls within the A_{95} of the mean pole calculated with n=7 in 100% of the calculations for the datasets from Mongolia, Turkey, and Antarctica (Figure 4). For the Norwegian dataset, this is the case for 91% of the calculations. In other words, the pole position is barely influenced by the number of samples per site (Figure 5a) and stays within the 95% confidence interval. The influence on the A_{95} is small, but there is a slight decrease in the A_{95} when within-site scatter is averaged by increasing n (Figure 5b). This varies from less than 0.5° to 1.5°, with the largest effect occurring for the dataset from Norway, whose sites have the largest within-site scatter. We find that the number of samples per site has a minor influence on K and S (Figure 5c, d). Only Norway shows an increase in K from approximately 14 to 29 and in S from 23° to 16°. This experiment shows that the effect of the number of samples per site on determining a pole position is surprisingly small, even if between-site scatter is somewhat decreasing.

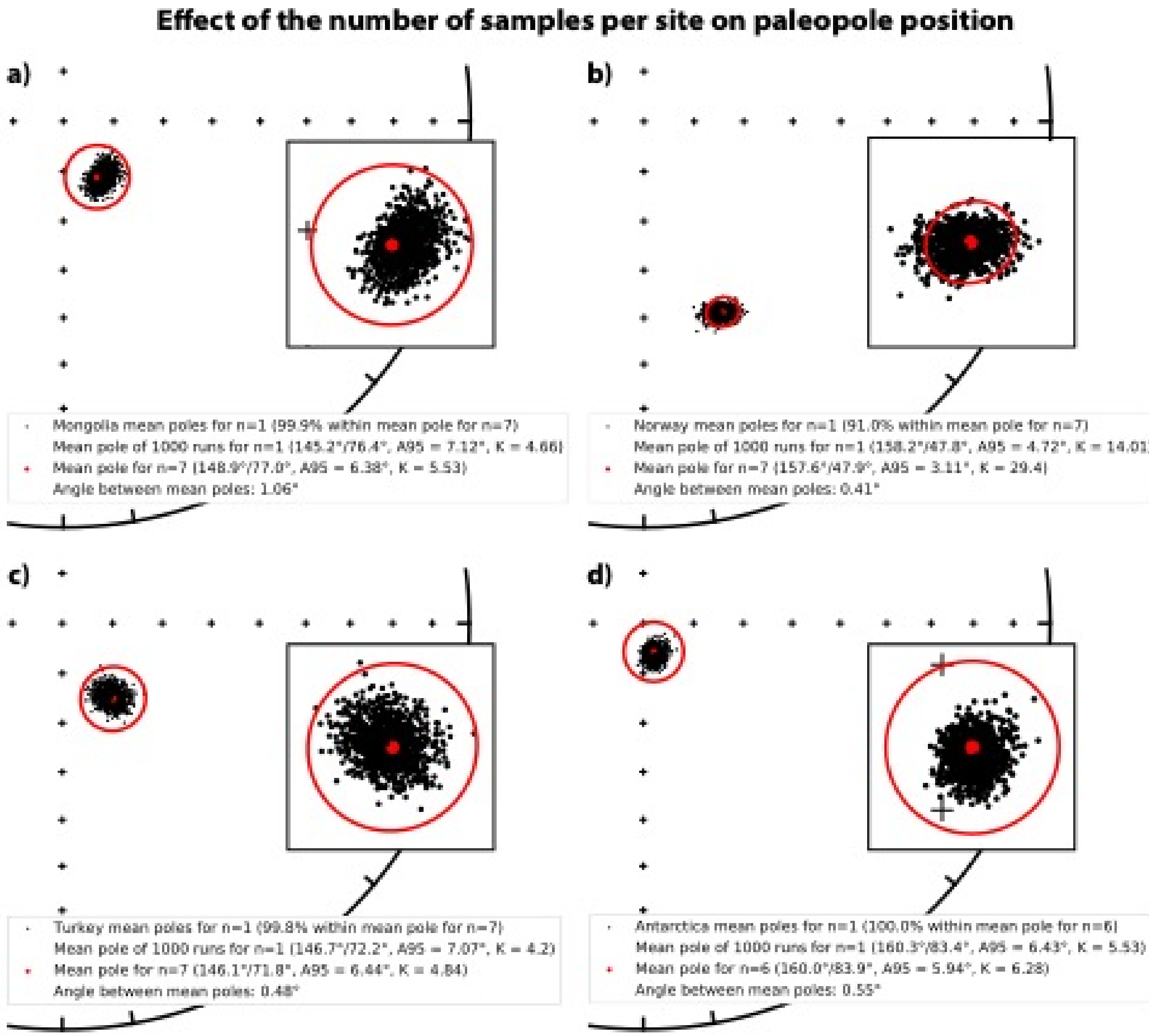

Figure 4. Stereographic projection with in black 1000 mean paleopole positions for n=1 and in red the mean paleopole position and its A_{95} for n=7 (or 6, in case of Antarctica) for **a)** Mongolia, **b)** Norway, **c)** Turkey, and **d)** Antarctica. Inset shows a zoom-in on the data.

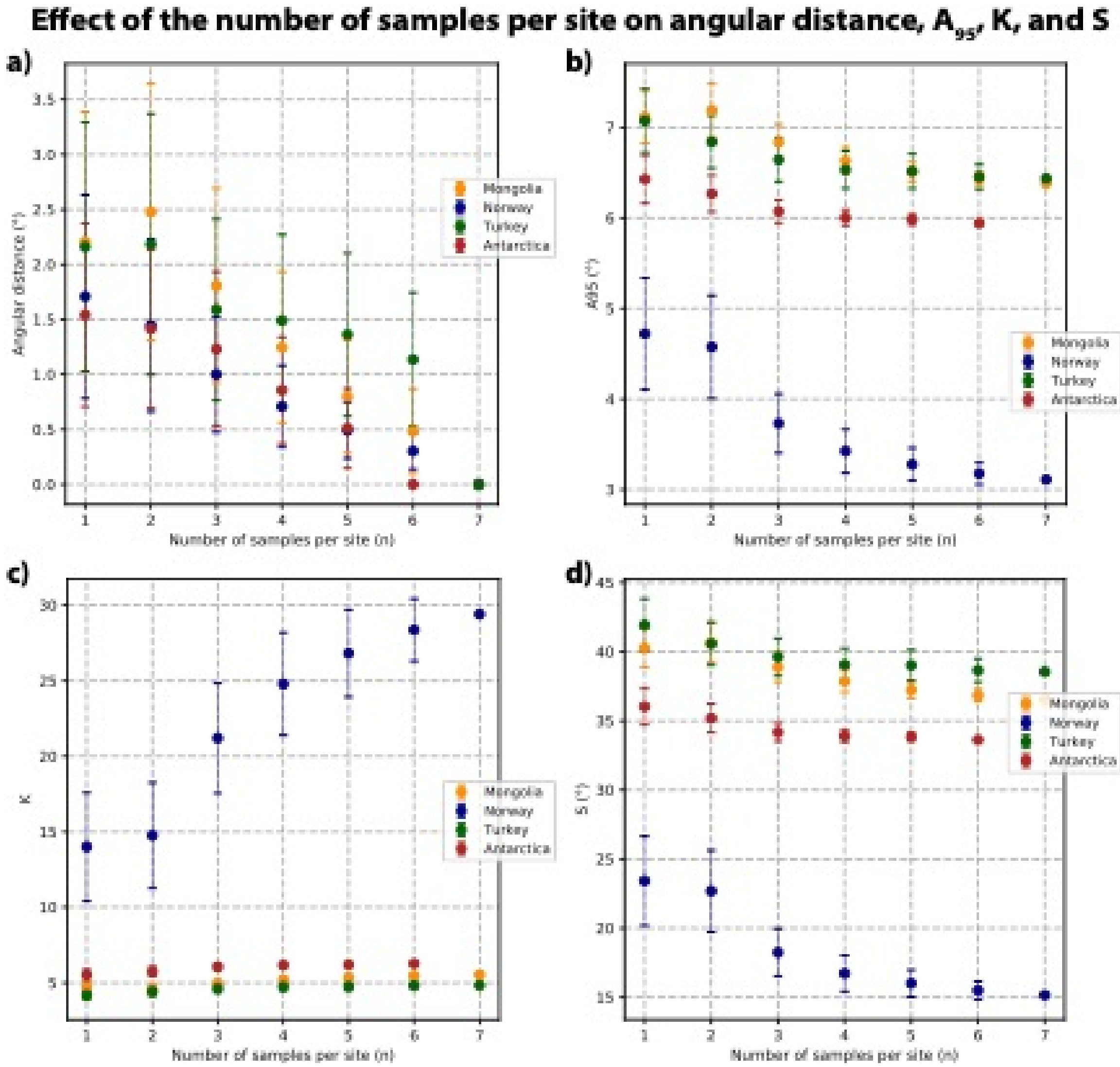

Figure 5. Effect of the number of samples per paleomagnetic site (n) on the **a)** angular distance of the acquired paleopole to the reference pole with n=7 (or 6, in case of Antarctica), and the **b)** A_{95}, **c)** K, and **d)** S of the acquired paleopole. Every datapoint represents 1000 simulations, with error bars indicating the 95% confidence interval.

4.3 Effect of discarding outliers on site level

One potential benefit of having multiple samples per site is to discard outliers, commonly done in the paleomagnetic community based on the 'expert eye' of the interpreter. Here, we objectively studied the effect of discarding outliers on site level, by eliminating the farthest outlying sample(s) from the site. We did this three times, each time discarding the sample furthest from the recalculated site mean. Interestingly, we find generally no effect of eliminating within-site outliers on mean paleopole position and its A_{95}, K, and S (Figure 6), and for the Norway dataset, the between-site scatter even increases. This is because VGP positions do not systematically shift towards or away from the paleopole (Figures S2-S5). As a result, decreasing the within-site scatter by removing outlying directions has little effect on the between-site dispersion of VGPs.

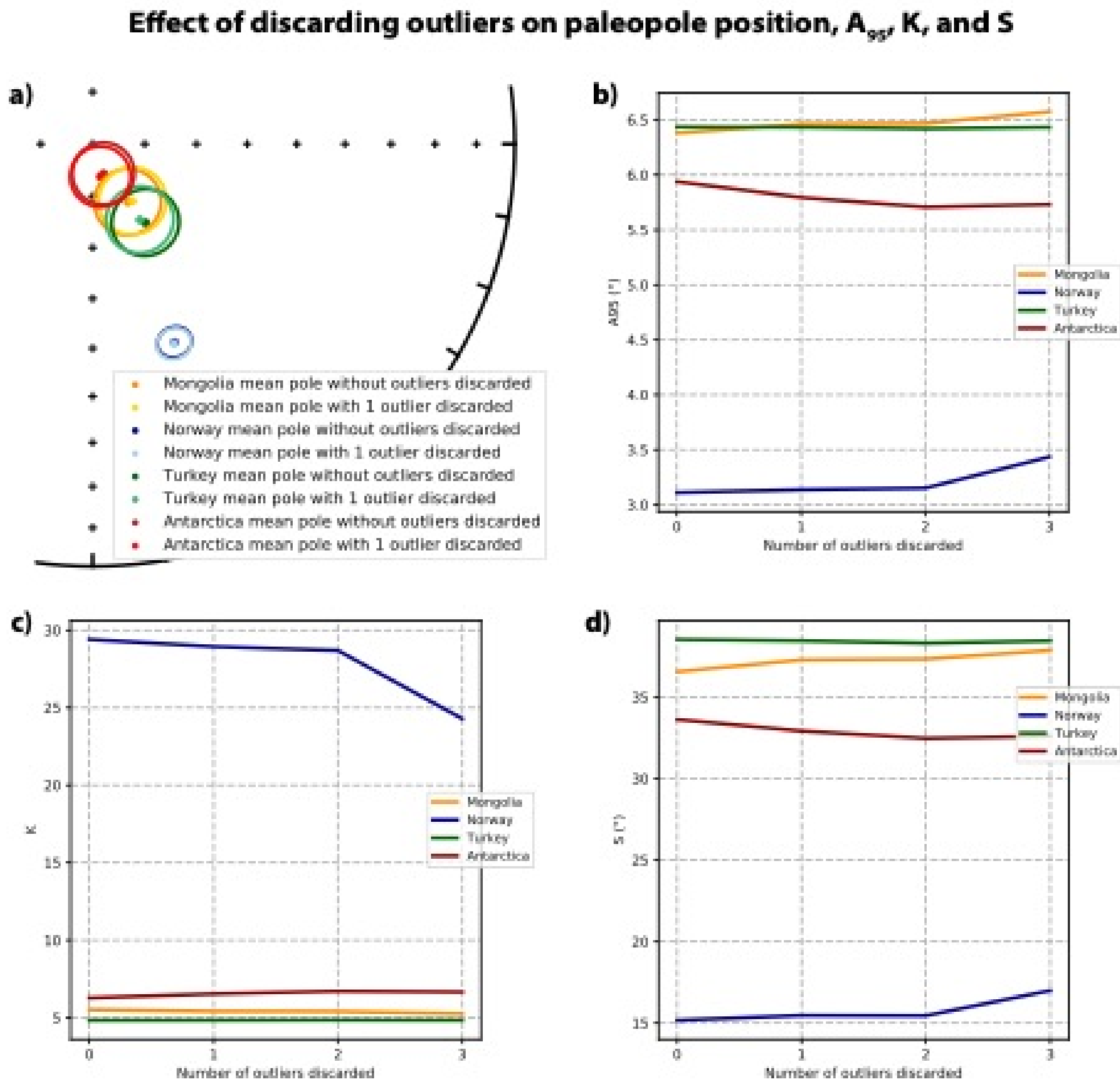

Figure 6. Effect of the number of samples discarded per paleomagnetic site on a) paleopole position in stereographic projection, shown without outliers discarded and with one outlier discarded, and the b) A_{95}, c) K, and d) S of the acquired paleopole.

4.4 Effect of sample distribution over sites

To average between-site scatter, it is common to collect multiple sites at a locality. The minimum number of sites recommended for determining a paleopole varies between authors (e.g., Meert et al., 2020; Van der Voo, 1990), but as shown by Vaes et al. (2021), poles with larger N tend to plot closer to a best estimate of the time-averaged pole. We performed experiments varying the number of samples per site and the number of sites, with combinations chosen such that the total amount of paleomagnetic samples remained approximately 100. We studied the effect on angular distance to the reference pole, A_{95}, K, and S (Figure 7) of the acquired paleopole. The means were calculated from 1000 runs of the specific number of samples per site and number of sites, where the number of samples and sites was randomly selected from the total population during each run.

The paleopole position is strongly influenced by the distribution of samples over sites in case of Mongolia, Turkey, and Antarctica, and lies further from the reference pole if the samples are distributed over fewer sites (Figure 7a). There also is a major effect on the A_{95}. For all datasets, the A_{95} is by far the smallest when it is constrained by 100 sites of $n = 1$, and the highest when it is constrained by 15 sites of n=7 (Figure 7b). The effect on K and S is small, except in the dataset from Norway (Figure 7c, d). In that dataset, K is highest and S is lowest for 15 sites with n=7. Nonetheless, the paleopole position is far better determined when taking many sites with few samples.

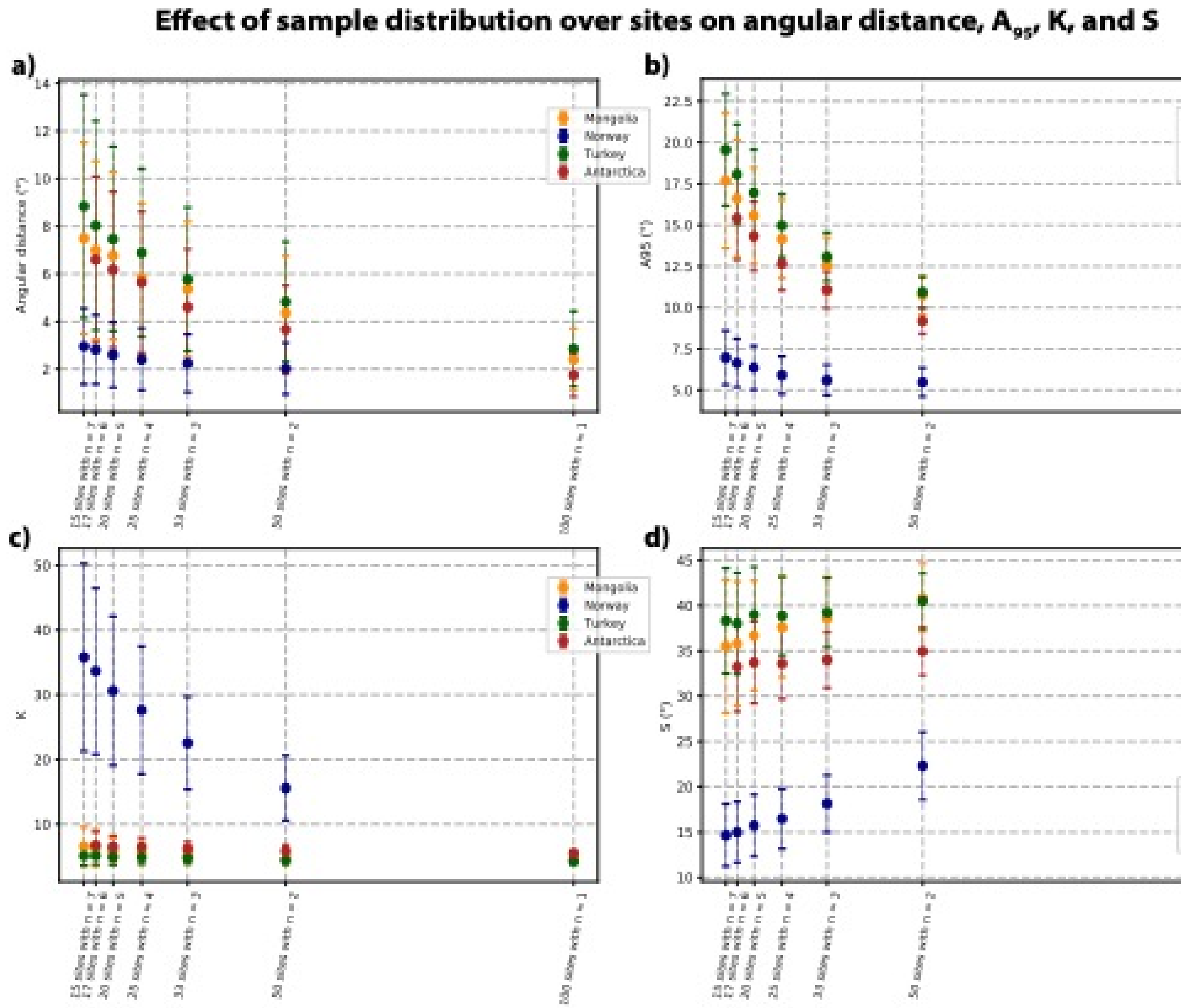

Figure 7. Effect of distributing 100 paleomagnetic samples over a varying number of paleomagnetic sites on the a) angular distance of the acquired paleopole to the reference pole with all sites with n=7 (or 6, in case of Antarctica), and the b) A_{95}, c) K, and d) S of the acquired paleopole. Every datapoint represents 1000 runs of a distribution, with error bars indicating the 95% confidence interval of these 1000 runs.

5 Discussion and conclusions

We first evaluate whether measuring multiple samples per site to constrain within-site scatter allows to objectively filter unreliable VGPs. In our reinterpretation, we closed our 'expert eye' and included sites that were excluded in the original studies because of subjective criteria: for instance, our dataset contains sites with erratic demagnetization behavior or very high intensities that were originally interpreted as lightning struck. Our analysis illustrates that objective statistical filters, such as a k-cutoff, a minimum number of samples per site, or eliminating the farthest outliers per site, are incapable of filtering these unreliable VGPs. The subjective judgement of the expert thus determines which samples and sites are included in the calculation of the paleopole (and makes it important that all data are made available following the FAIR principles (Wilkinson et al., 2016)).

Our analysis next shows that for the three datasets with high between-site scatter and low within-site scatter (i.e., Turkey, Mongolia, and Antarctica), applying the within-site scatter-based criteria leads to marginal increase in data clustering (i.e., an increase in K and decrease in S). For the Norwegian dataset, which has the rare combination of high within-site scatter and low between-site scatter, these criteria even lead to significantly higher clustering of the data. This lends support to using these data filters for PSV studies.

But applying these filters when calculating paleopoles in all cases leads to a decrease in pole precision (i.e., an increase in A_{95}). This is the result of the decrease in N that is caused by applying data filters. Our experiments thus show that the effect of between-site scatter on pole precision is far dominant over within-site scatter. And because the accuracy of paleopoles also decreases with decreasing N (Vaes et al., 2021), applying within-site scatter filters systematically decreases pole accuracy as well as precision (see also Figure 7).

Averaging multiple samples per site, without applying filters, does not negatively influence pole accuracy or precision. In case only few sites are available, or there is reason to question data quality, collecting multiple samples per site, as recommended by Meert et al. (2021), may still be a useful strategy. It may also be useful to demonstrate for a selection of lava sites in a study that within-site precision (k) is high, as a field test equivalent to a fold, reversal, or conglomerate test. But, our analysis shows that within-site scatter does not significantly influence paleopole accuracy or precision. The extra efforts put into collecting multiple samples per site are thus more effectively spent collecting more single-sample sites.

Acknowledgements

BV and DJJvH acknowledge NWO Vici grant 865.17.001 to DJJvH.

Data Availability Statement

All uninterpreted paleomagnetic data are publicly available on Paleomagnetism.org for the Mongolia, Norway, and Turkey dataset and in the MagIC database for the Antarctica dataset: https://earthref.org/MagIC/17076. The Python code used for our analysis will be made public on GitHub upon acceptance of this manuscript.

References

Asefaw, H., Tauxe, L., Koppers, A. A. P., & Staudigel, H. (2021). Four-Dimensional Paleomagnetic Dataset: Plio-Pleistocene Paleodirection and Paleointensity Results From the Erebus Volcanic Province, Antarctica. *Journal of Geophysical Research: Solid Earth*, *126*(2), 1–22. https://doi.org/10.1029/2020JB020834Biggin, A. J., van Hinsbergen, D. J. J., Langereis, C. G., Straathof, G. B., & Deenen, M. H. L. (2008). Geomagnetic secular variation in the Cretaceous Normal Superchron and in the Jurassic. *Physics of the Earth and Planetary Interiors*, *169*(1–4), 3–19. https://doi.org/10.1016/j.pepi.2008.07.004Brandt, D., Ernesto, M., & Constable, C. (2021). Consistent and Contrasting Aspects of the Geomagnetic Field Across Epochs With Distinct Reversal Frequencies Revealed by Modeling the Kiaman Superchron. *Geochemistry, Geophysics, Geosystems*, *22*(7), 1–24. https://doi.org/10.1029/2021GC009866Butler, R. F. (1992). *Paleomagnetism: Magnetic domains to geologic terranes* (Vol. 319). Boston: Blackwell Scientific Publications.Cromwell, G., Johnson, C. L., Tauxe, L., Constable, C. G., & Jarboe, N. A. (2018). PSV10: A Global Data Set for 0–10 Ma Time-Averaged Field and Paleosecular Variation Studies. *Geochemistry, Geophysics, Geosystems*, *19*(5), 1533–1558. https://doi.org/10.1002/2017GC007318Fisher, R. A. (1953). Dispersion on a sphere. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *217*(1130), 295–305. https://doi.org/10.1098/rspa.1953.0064Haldan, M. M., Meijers, M. J. M., Langereis, C. G., Larsen, B. T., & Heyer, H. (2014). New palaeomagnetic results from the Oslo Graben, a Permian Superchron lava province. *Geophysical Journal International*, *199*(3), 1554–1571. https://doi.org/10.1093/gji/ggu351Handford, B. T., Biggin, A. J., Haldan, M. M., & Langereis, C. G. (2021). Analysing Triassic and Permian Geomagnetic Palaeosecular Variation and the Implications for Ancient Field Morphology. *Geochemistry, Geophysics, Geosystems*, *In press*. https://doi.org/10.1029/2021GC009930van Hinsbergen, D. J. J., Straathof, G. B., Kuiper, K. F., Cunningham, W. D., & Wijbrans, J. (2008). No vertical axis rotations during Neogene transpressional orogeny in the NE Gobi Altai: Coinciding Mongolian and Eurasian early Cretaceous apparent polar wander paths. *Geophysical Journal International*, *173*(1), 105–126. https://doi.org/10.1111/j.1365-246X.2007.03712.xvan Hinsbergen, D. J. J., Dekkers, M. J., Bozkurt, E., & Koopman, M. (2010). Exhumation with a twist: Paleomagnetic constraints on the evolution of the Menderes metamorphic core complex, western Turkey. *Tectonics*, *29*(3), 1–33. https://doi.org/10.1029/2009TC002596van Hinsbergen, D. J. J., De Groot, L. V., Van Schaik, S. J., Spakman, W., Bijl, P. K., Sluijs, A., et al. (2015). A paleolatitude calculator for paleoclimate studies. *PLoS ONE*, *10*(6), 1–21. https://doi.org/10.1371/journal.pone.0126946Johnson, C. L., & Constable, C. G. (1996). Palaeosecular variation recorded by lava flows over the past five million years. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *354*(1704), 89–141.

https://doi.org/10.1098/rsta.1996.0004Johnson, C. L., Constable, C. G., Tauxe, L., Barendregt, R., Brown, L. L., Coe, R. S., et al. (2008). Recent investigations of the 0-5 Ma geomagnetic field recorded by lava flows. *Geochemistry, Geophysics, Geosystems*, *9*(4). https://doi.org/10.1029/2007GC001696Kirschvink, J. L. (1980). The least-squares line and plane and the analysis of palaeomagnetic data. *Geophysical Journal of the Royal Astronomical Society*, *62*(3), 699–718. https://doi.org/10.1111/j.1365-246X.1980.tb02601.xKoymans, M. R., Langereis, C. G., Pastor-Galán, D., & van Hinsbergen, D. J. J. (2016). Paleomagnetism.org: An online multi-platform open source environment for paleomagnetic data analysis. *Computers and Geosciences*, *93*, 127–137. https://doi.org/10.1016/j.cageo.2016.05.007Koymans, M. R., van Hinsbergen, D. J. J., Pastor-Galán, D., Vaes, B., & Langereis, C. G. (2020). Towards FAIR Paleomagnetic Data Management Through Paleomagnetism.org 2.0. *Geochemistry, Geophysics, Geosystems*, *21*(2), 1–7. https://doi.org/10.1029/2019GC008838Lippert, P. C., van Hinsbergen, D. J. J., & Dupont-Nivet, G. (2014). Early Cretaceous to present latitude of the central proto-Tibetan Plateau: A paleomagnetic synthesis with implications for Cenozoic tectonics, paleogeography, and climate of Asia. *Special Paper of the Geological Society of America*, *507*, 1–21. https://doi.org/10.1130/2014.2507(01)McElhinny, M. W., & McFadden, P. L. (1999). *Paleomagnetism: continents and oceans*. Elsevier.McFadden, P. L., & McElhinny, M. W. (1988). The combined analysis of remagnetization circles and direct observations in palaeomagnetism. *Earth and Planetary Science Letters*, *87*(1–2), 161–172. https://doi.org/10.1016/0012-821X(88)90072-6Meert, J. G., Pivarunas, A. F., Evans, D. A. D., Pisarevsky, S. A., Pesonen, L. J., Li, Z. X., et al. (2020). The magnificent seven: A proposal for modest revision of the Van der Voo (1990) quality index. *Tectonophysics*, *790*(March), 228549. https://doi.org/10.1016/j.tecto.2020.228549Merrill, R. T., & McFadden, P. L. (2003). The geomagnetic axial dipole field assumption. *Physics of the Earth and Planetary Interiors*, *139*(3–4), 171–185. https://doi.org/10.1016/j.pepi.2003.07.016de Oliveira, W. P., Hartmann, G. A., Terra-Nova, F., Brandt, D., Biggin, A. J., Engbers, Y. A., et al. (2021). Paleosecular variation and the time-averaged geomagnetic field since 10 Ma. *Geochemistry, Geophysics, Geosystems*. https://doi.org/10.1029/2021gc010063Tauxe, L. (2010). *Essentials of paleomagnetism*. University of California Press.Tauxe, L., Constable, C., Johnson, C. L., Koppers, A. A. P., Miller, W. R., & Staudigel, H. (2003). Paleomagnetism of the southwestern U.S.A. recorded by 0-5 Ma igneous rocks. *Geochemistry, Geophysics, Geosystems*, *4*(4). https://doi.org/10.1029/2002GC000343Tauxe, L., Shaar, R., Jonestrask, L., Swanson-Hysell, N. L., Minnett, R., Koppers, A. A. P., et al. (2016). PmagPy: Software package for paleomagnetic data analysis and a bridge to the Magnetics Information Consortium (MagIC) Database. *Geochemistry, Geophysics, Geosystems*, *17*(6), 2450–2463. https://doi.org/10.1002/2016GC006307Torsvik, T. H., Van der Voo, R., Preeden, U., Mac Niocaill, C., Steinberger, B., Doubrovine, P. V., et al. (2012). Phanerozoic Polar Wander, Palaeogeography and Dynamics. *Earth-Science Re-*

views, *114*(3–4), 325–368. https://doi.org/10.1016/j.earscirev.2012.06.007Uzel, B., Langereis, C. G., Kaymakci, N., Sözbilir, H., Özkaymak, Ç., & Özkaptan, M. (2015). Paleomagnetic evidence for an inverse rotation history of Western Anatolia during the exhumation of Menderes core complex. *Earth and Planetary Science Letters*, *414*, 108–125. https://doi.org/10.1016/j.epsl.2015.01.008Uzel, B., Sümer, Ö., Özkaptan, M., Özkaymak, Ç., Kuiper, K., Sözbilir, H., et al. (2017). Palaeomagnetic and geochronological evidence for a major middle miocene unconformity in Söke Basin (western Anatolia) and its tectonic implications for the Aegean region. *Journal of the Geological Society*, *174*(4), 721–740. https://doi.org/10.1144/jgs2016-006Vaes, B., Gallo, L., and van Hinsbergen, D.J.J. (2021). On pole position: causes of dispersion of the paleomagnetic poles behind apparent polar wander paths, *Journal of Geophysical Research, submitted.*van der Voo, R. (1990). The reliability of paleomagnetic data. *Tectonophysics*, *184*(1), 1–9. https://doi.org/10.1016/0040-1951(90)90116-PWilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., et al. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 1–9. https://doi.org/10.1038/sdata.2016.18