Quantitative Analysis of Paleomagnetic Sampling Strategies

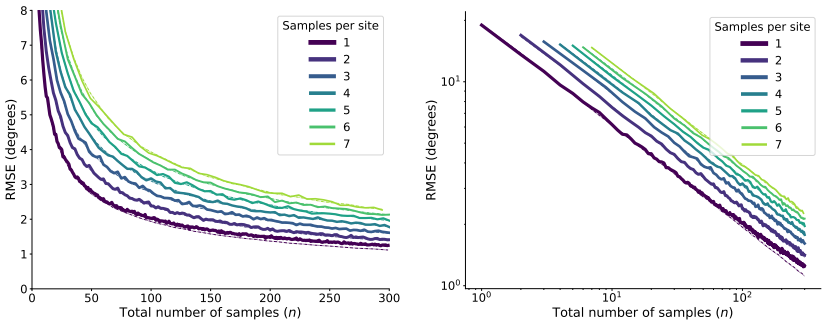
Facundo Sapienza[1], Leandro Cesar Gallo[2], Yiming Zhang[1], Bram Vaes[3], Mathew Domeier[4], and Nicholas L Swanson-Hysell[1]

[1]University of California, Berkeley
[2]The Center for Earth Evolution and Dynamics
[3]Utrecht University
[4]University of Oslo

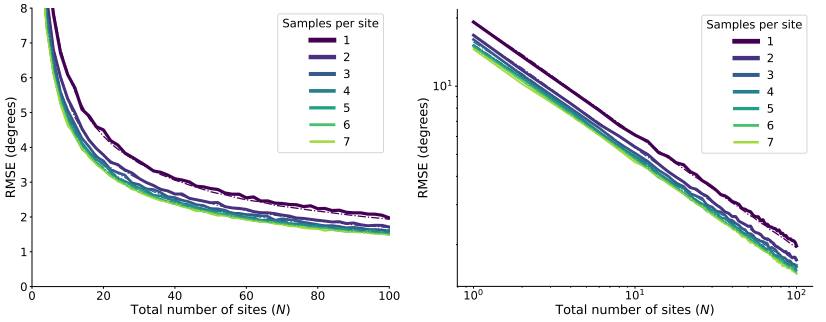
July 8, 2023

Abstract

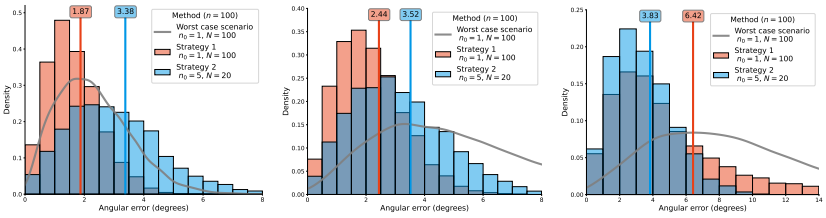
Sampling strategies used in paleomagnetic studies play a crucial role in dictating the accuracy of our estimates of properties of the ancient geomagnetic field. However, there has been little quantitative analysis of optimal paleomagnetic sampling strategies and the community has instead defaulted to traditional practices that vary between laboratories. In this paper, we quantitatively evaluate the accuracy of alternative paleomagnetic sampling strategies through numerical experiment and an associated analytical framework. Our findings demonstrate a strong correspondence between the accuracy of an estimated paleopole position and the number of sites or independent readings of the time-varying paleomagnetic field, whereas larger numbers of in-site samples have a dwindling effect. This remains true even when a large proportion of the sample directions are spurious. This approach can be readily achieved in sedimentary sequences by distributing samples stratigraphically, considering each sample as an individual reading. However, where the number of potential independent sites is inherently limited the collection of additional in-site samples can improve the accuracy of the paleopole estimate (although with diminishing returns with increasing samples per site). Where an estimate of the magnitude of paleosecular variation is sought, multiple in-site samples should be taken, but the optimal number is dependent on the expected fraction of outliers. We provide both analytical formulas and a series of interactive Jupyter notebooks allowing optimal sampling strategies to be derived from user-informed expectations.



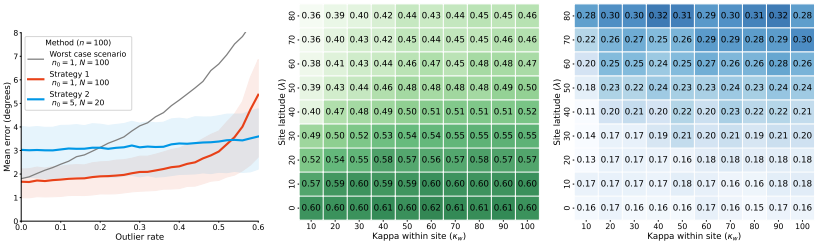
(a) (b)



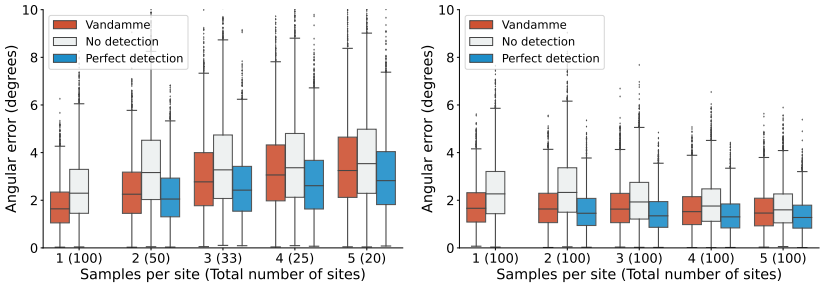
(c) (d)



(a) 10% Outliers (b) 40% Outliers (c) 60% Outliers

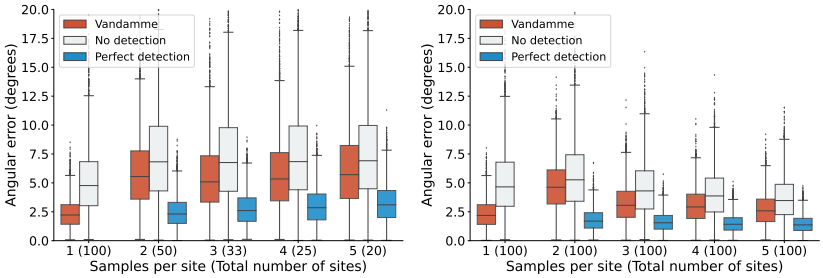


(d) Intersecting errors (e) $p_{outlier}$ critical (f) $p_{outlier}$ critical



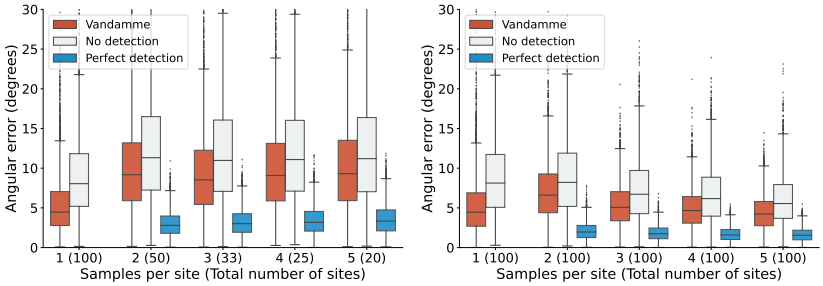
(a) $n \approx 100$, 10% Outliers

(b) $N = 100$, 10% Outliers



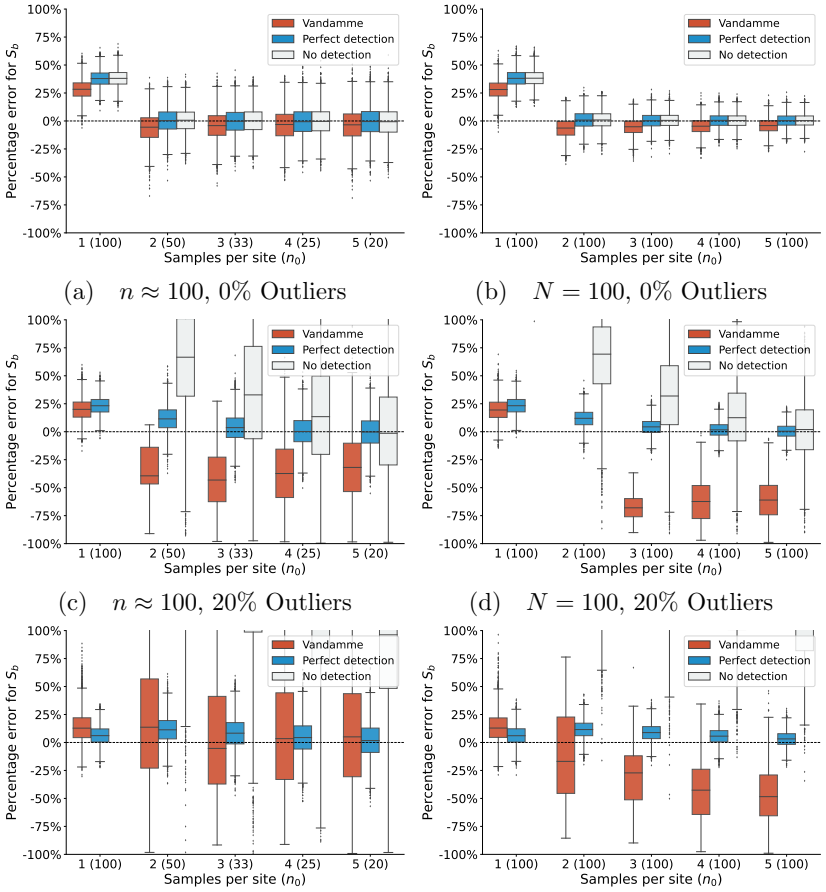
(c) $n \approx 100$, 40% Outliers

(d) $N = 100$, 40% Outliers



(e) $n \approx 100$, 60% Outliers

(f) $N = 100$, 60% Outliers



(a) $n \approx 100$, 0% Outliers

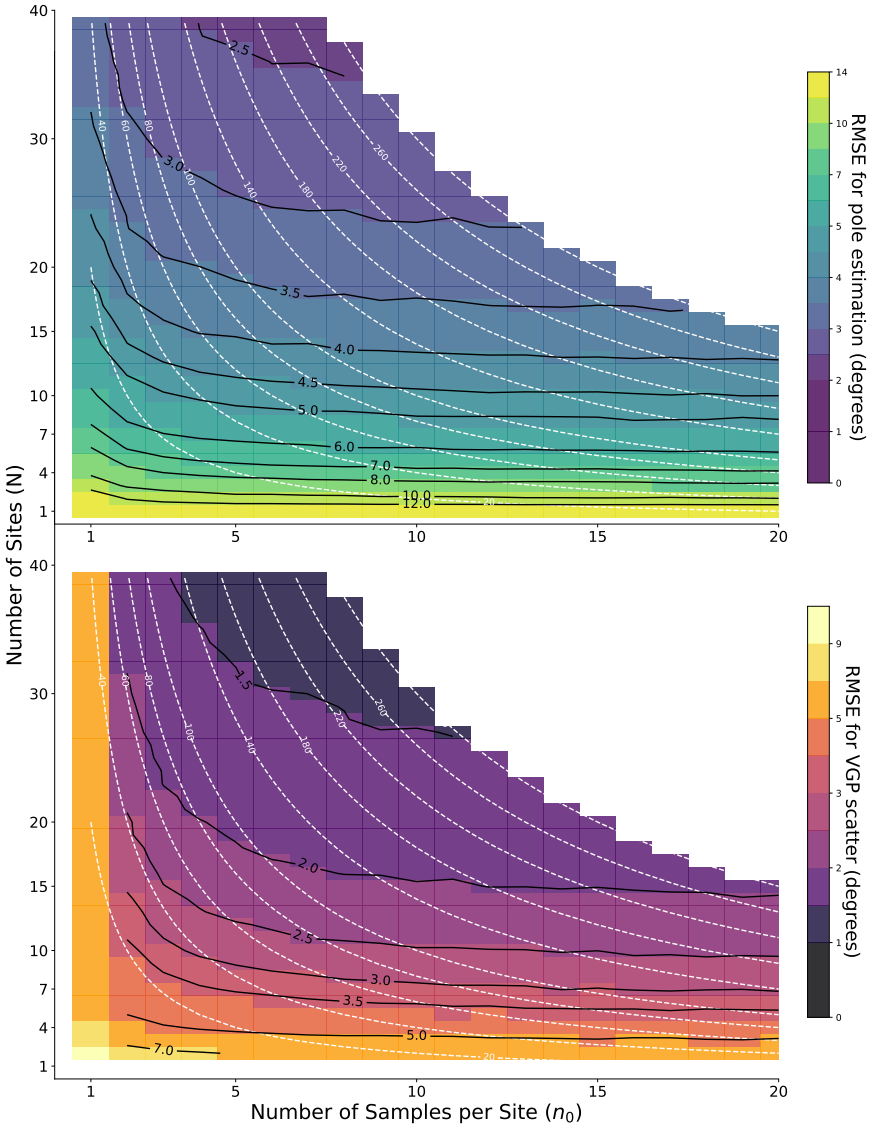
(b) $N = 100$, 0% Outliers

(c) $n \approx 100$, 20% Outliers

(d) $N = 100$, 20% Outliers

(e) $n \approx 100$, 40% Outliers

(f) $N = 100$, 40% Outliers



Quantitative Analysis of Paleomagnetic Sampling Strategies

F. Sapienza[1], L. C. Gallo[2,3], Y. Zhang[4], B. Vaes[5], M. Domeier[2,3], N. L. Swanson-Hysell[5]

[1]Department of Statistics, University of California, Berkeley, United States
[2]Centre for Earth Evolution and Dynamics, University of Oslo, Norway
[3]Centre for Planetary Habitability, University of Oslo, Norway
[4]Department of Earth and Planetary Science, University of California, Berkeley, United States
[5]Department of Earth Sciences, Utrecht University, Utrecht, The Netherlands

Key Points:

- We explore the effect of alternative sampling strategies on the accuracy of paleomagnetic parameter estimates.
- The accuracy of paleopole estimation is dominated by the number of independent site recordings instead of the number of in-site samples.
- Even in the presence of frequent outliers, or for studies of paleosecular variation, a focus on sites over in-site samples is advantageous.

Corresponding author: Facundo Sapienza, fsapienza@berkeley.edu

Abstract

Sampling strategies used in paleomagnetic studies play a crucial role in dictating the accuracy of our estimates of properties of the ancient geomagnetic field. However, there has been little quantitative analysis of optimal paleomagnetic sampling strategies and the community has instead defaulted to traditional practices that vary between laboratories. In this paper, we quantitatively evaluate the accuracy of alternative paleomagnetic sampling strategies through numerical experiment and an associated analytical framework. Our findings demonstrate a strong correspondence between the accuracy of an estimated paleopole position and the number of sites or independent readings of the time-varying paleomagnetic field, whereas larger numbers of in-site samples have a dwindling effect. This remains true even when a large proportion of the sample directions are spurious. This approach can be readily achieved in sedimentary sequences by distributing samples stratigraphically, considering each sample as an individual reading. However, where the number of potential independent sites is inherently limited the collection of additional in-site samples can improve the accuracy of the paleopole estimate (although with diminishing returns with increasing samples per site). Where an estimate of the magnitude of paleosecular variation is sought, multiple in-site samples should be taken, but the optimal number is dependent on the expected fraction of outliers. We provide both analytical formulas and a series of interactive Jupyter notebooks allowing optimal sampling strategies to be derived from user-informed expectations.

Plain Language Summary

Earth's magnetic field can be preserved in rocks when they form. Through studying these magnetic records using the tools of paleomagnetism, scientists can learn about how Earth's magnetic field has changed through time and how tectonic plates have moved relative to the field. This study is about the best ways to design sampling approaches to gain these insights using statistical quantification. Traditional protocols emphasize the collection of numerous samples from units that record the field at a given instant in time. Such units are referred to as sites. Through simulating data, we develop tools for evaluating trade offs between collecting more sites and more samples per site. Our results show that strategies that maximize collecting more sites, even if fewer samples are taken at each site, leads to more accurate estimates even in the presence of spurious observation. While there is a benefit to more samples per site, particularly for studies seeking to estimate the variability of the ancient field, such sampling has diminishing returns relative to maximizing the number of sites. We provide formulas and interactive computational resources to help the community to make informed decisions about the best way to gather data.

1 Introduction

Paleomagnetism is concerned with attempting to estimate properties of the ancient geomagnetic field from magnetic records preserved in rocks. This involves laboratory measurements of paleomagnetic directions recorded by igneous and/or sedimentary rocks and statistical analyses of those directions. Two geomagnetic properties of particular interest that can be estimated from these paleomagnetic directional data are:

- The position of the time-averaged ($\gtrsim 10^4$–10^5 a) ancient geomagnetic pole (also known as a *paleopole*) that corresponds to the Earth's spin axis according to the geocentric axial dipole hypothesis (Creer et al., 1954).
- The paleosecular variation of the field, which is associated with the shorter-term ($\lesssim 10^4 - 10^5$ a) time-varying position of the geomagnetic pole.

Despite the importance of these two quantities, there has been little exploration of the best sampling practices with which to derive estimates of them. This has resulted in practices that vary according to the traditions of different laboratories; that is, the community largely relies on conventional wisdom.

In this hierarchical framework, a site should correspond to a unit of rock with a common age and direction of magnetization (McElhinny & McFadden, 2000; Tauxe et al., 2016). Note that in some contributions a site is defined more loosely as a small area from which samples are collected which is not the definition that we use here. In our preferred definition, each site is interpreted to be a *spot recording* of the time-varying geomagnetic field. In the case of an igneous rock, a site could be an individual lava flow or intrusion, whereas for a sedimentary rock, a site should ideally comprise a single depositional event. In practice, a sedimentary site typically corresponds to single stratigraphic horizon that is the height of a standard paleomagnetic sample, usually about 2.5 cm. To move up the hierarchy, a collection of paleomagnetic samples from a given site are averaged and the site-mean is transformed from a direction with an associated declination and inclination to pole space with an associated latitude and longitude, where the mean is referred to as a *virtual geomagnetic pole* (VGP). Following the definition of a site, each VGP ideally represents an independent estimate of the position of the ancient geomagnetic pole at an instant in time. Estimates of paleosecular variation of the ancient geomagnetic field prior to 10 Ma can be made from populations of VGPs by determining their angular dispersion – most typically applied to collections of igneous sites of a similar age (e.g. Model G; McFadden et al., 1988). To determine a mean paleomagnetic pole position, a group of similarly aged VGPs are averaged to a Fisher mean paleopole that is taken as the best estimate of the true position of the ancient geographic pole, relative to the observation point.

Regardless of whether we seek to discern the statistical properties of the time-averaged pole position or geomagnetic secular variation, our estimates will include error. Paleomagnetic errors come from a variety of sources which can include orientation errors both in the field and the laboratory; measurement errors; and the imperfect isolation of the magnetization of interest from secondary magnetic overprints. The frequent occurrence of imperfect magnetization acquisition or the inability to isolate primary components often results in a sample collection being contaminated by outliers. Orientation and measurement errors are generally assumed to be randomly unbiased (non-systematic) and so can be mitigated through the collection, measurement and directional averaging of multiple samples within a site. However, given finite resources, the collection of additional samples per site will come at the cost of a lower number of sites. A relevant question is thus: how should we distribute our sampling to minimize uncertainty on the property we seek to estimate? Is it better to take a few sites with many samples? Or many sites with fewer samples? How might the recommended strategy change depending on the objective (in estimating the location of the paleopole vs. the dispersion of VGPs) or the fidelity of the magnetic record?

Some notions concerning sampling have become entrenched in the paleomagnetic literature. For example, many workers seek to collect six to eight samples per site (Butler, 1992), although the rationale for this range is not entirely clear. Opdyke and Channell (1996) suggest that at least three samples per site be collected where determinations of polarity are important, whereas to reliably estimate the dispersion of sample directions within a site, a minimum of four (Cromwell et al., 2018) or five (Tauxe et al., 2003) samples per site has been deemed necessary. Having a more significant number of samples within the site provides the benefit of being able to apply data filters based on within-site scatter. However, Gerritsen et al. (2022) have found empirically that collecting and averaging multiple samples per site only results in a modest enhancement of the overall accuracy of the paleopole. Thus, where the objective is to estimate the position of a paleopole, Gerritsen et al. (2022) suggested that it is most beneficial to maximize the

number of sites, and so the collection of additional single-sample sites should be preferred over the collection of multiple samples from fewer sites. Nevertheless, a statistical and quantitative evaluation of alternative strategies has not yet been conducted.

Here we explore how the distribution of samples across sites affects the performance in the estimation of the paleopole position and the dispersion of VGPs, and how the varying influence of outliers dictates the optimal strategy to best estimate these parameters. We also derive a set of equations that can enable quantitative sampling strategy recommendations based on specified parameters informed by user expectations.

2 Mathematical Setup

Consider the problem of estimating a paleomagnetic pole μ_0 for some given interval of time, where μ_0 is a three-dimension vector contained in the unit sphere. Observations consist of a collection of a total of n samples distributed among N sites. Because the geomagnetic field is constantly varying around a mean configuration, each one of the virtual geomagnetic poles (VGP) per site, denoted by μ_i with $i = 1, 2, \ldots, N$, is going to differ from the time-averaged paleomagnetic pole μ_0. A fundamental assumption in paleomagnetic research is that this secular variation of the geomagnetic field can be effectively estimated through averaging of a sufficiently high number of independent and temporally distributed VGPs. We now seek to evaluate how our choices of n and N will affect our estimation of μ_0, as well as how we distribute the n samples among the N sites.

2.1 Data Generating Process

We define the following data generating model. First, we consider a set with a total of N VGPs sampled from a statistical model of secular variation. Examples of these models include the Gaussian process type model (Tauxe & Kent, 2004; Constable & Parker, 1988) and model G (McFadden et al., 1988). In this contribution, we use Model G which captures latitudinal variation in VGP scatter, and considers a mean geocentric axial and dipolar (GAD) field. Then, given a GAD mean direction μ_0, we sample a series of VGPs $\mu_1, \mu_2, \ldots, \mu_N$ according to

$$\mu_i \sim \mathrm{SV}(\mu_0, \kappa_b) \quad i = 1, 2, \ldots, N. \tag{1}$$

The sampling procedure depends on the mean direction μ_0 and the precision parameter κ_b that will depend on the secular variation model used. In this study, we adopt the mild assumption that VGP distributions are circularly symmetric (Tauxe & Kent, 2004) and can be sampled from a Fisher distribution (Fisher, 1953; Deenen et al., 2011), whose dispersion S_b, according to model G (McFadden et al., 1988), depends on the sampling latitude λ through the following formula

$$S_b(\lambda)^2 = a^2 + b^2\lambda^2, \tag{2}$$

with a and b two empirical coefficients, recently calculated as $a = 11.3^{\circ}{}^{+1.3^{\circ}}_{-1.1^{\circ}}$ and $b = 0.27^{+0.04}_{-0.08}$ by (Doubrovine et al., 2019). At population level, there is a one-to-one relationship between S_b and the value of κ_b we use to sample from the Fisher distribution. This relationship can be found numerically with an arbitrary level of precision. Then, VGPs can be sampled according to a Fisher distribution with mean direction μ_0 and dispersion parameter $\kappa_b(\lambda)$.

In the following, we use the supraindex $*$ to denote variables in directional space (inclination-declination). Thus, μ_i refers to any given VGP (geographic coordinates) and μ_i^* refers to its corresponding direction in inclination and declination space according to the dipole formula. Note that this transformation between pole and directional space depends on the latitude and longitude of the site.

Table 1: Parameters used for the sampling of poles

Parameter	Range	Description
N	≥ 1	Total number of sites.
n_0	≥ 1	Number of samples per site. We will assume $n_0 = n_1 = \ldots = n_N$ and denote $n = Nn_0$ the total number of samples.
κ_w	$[0, \infty)$	Precision parameter of the Fisher distribution for a given site, where $k_w = 0$ results in a uniform distribution on a sphere and $k_w \to \infty$ is a singular point.
κ_b	$[0, \infty)$	Precision parameter of the Fisher distribution between sites. For the model G, this is directly determined by λ.
λ	$[0°, 90°]$	Paleolatitude.
p_{outlier}	$[0, 1]$	Outlier rate where 0 is no outliers and 1 is all samples are outliers drawn from a uniform distribution.

Now, we assume that the ith-site has n_i individual directions that follow a Fisher distribution

$$\begin{aligned} x^*_{ij} &\sim \text{Fisher}(\mu^*_i, \kappa_i) \quad \text{with probability } 1 - p_{\text{outlier}} \text{ and} \\ x^*_{ij} &\sim \text{Unif} \quad \text{otherwise, for } j = 1, 2, \ldots, n_i, \end{aligned} \tag{3}$$

with x_{ij} the jth-direction of the ith-site; κ_i the dispersion parameters per site; and Unif represents the uniform distribution in the sphere. The parameter p_{outlier} has been added to quantify the effect of outliers in the sampling process. With probability $1 - p_{\text{outlier}}$ we are going to observe a true sample, while with probability p_{outlier} our sample will be corrupted and instead we will observe an spurious direction, modelled by a uniform distribution on the sphere where no information is provided about the true orientation of the field. For cases where we do not want to consider the effect of outliers in the sampling process, we set $p_{\text{outlier}} = 0$. Also, for cases where the number of samples and dispersion parameter are the same for all the sites, we will use n_0 and κ_w to refer to any of the n_i and κ_i, respectively. The parameters used during the model are summarized in Table 1.

2.2 Estimation of the Paleopole Direction

We can estimate the true pole location μ_0 by computing the Fisher mean of the VGPs estimated from each site, that is,

$$\hat{\mu}_0 = \frac{1}{R_0} \sum_{i=1}^{N} \hat{\mu}_i \quad R_0 = \left\| \sum_{i=1}^{N} \hat{\mu}_i \right\|, \tag{4}$$

where $\|\cdot\|$ denotes the Euclidean norm; and $\hat{\mu}_i$ is the sample mean per site, which results from transforming to pole space the estimate of the pole in directional space,

$$\hat{\mu}^*_i = \frac{1}{R_i} \sum_{j=1}^{n_i} x^*_{ij} \quad R_i = \left\| \sum_{j=1}^{n_i} x^*_{ij} \right\|. \tag{5}$$

The overall goal of this estimation procedure is to get a value for $\hat{\mu}_0$ as close as possible to the ground truth μ_0.

We assess the accuracy of the pole estimate across simulations by computing the root-mean-square error (RMSE) as

$$\mathrm{Err}_{\hat{\mu}_0} = \sqrt{\frac{1}{M}\sum_{m=1}^{M} \mathrm{angle}\left(\hat{\mu}_0^{(m)}, \mu_0\right)^2}, \tag{6}$$

where $\mathrm{angle}(\hat{\mu}_0^{(m)}, \mu_0) = (180^\circ/\pi)\cos^{-1}(\hat{\mu}_0^T \hat{\mu}_0^{(m)})$ is the angular distance in degrees between the true pole μ_0 and each one of the simulated estimations $\hat{\mu}_0^{(m)}$, where M is the total number of simulations.

2.3 Estimation of the VGP Scatter

Long-term assessment of the paleomagnetic secular variation of the geomagnetic field relies on the VGPs dispersion S_b instead of their mean. The observed global dispersion S is estimated as (Cox, 1970)

$$\hat{S}^2 = \frac{1}{N-1}\sum_{i=1}^{N} \mathrm{angle}(\hat{\mu}_i, \hat{\mu}_0)^2. \tag{7}$$

The global dispersion S^2 is a combination of the dispersion *between* VGPs S_b and that arising from the dispersion among the samples *within* the site S_w (McFadden et al., 1991). We assume that the latter arises purely from random errors associated with orientation, measurement and analytical errors, whereas the former is an unknown, latitude-dependent parameter of the time-averaged geomagnetic field. In order to estimate S_b, we first need to extract the within-site dispersion from the global dispersion of the VGPs, that is

$$\hat{S}_b^2 = \hat{S}^2 - \hat{S}_w^2, \tag{8}$$

where the estimated within-site dispersion $\hat{S}_w$ is computed in directional space following McFadden et al. (1991) and Doubrovine et al. (2019)

$$\hat{S}_w^2 = \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{S}_{wi}^2}{n_i} \tag{9}$$

$$\hat{S}_{wi}^2 = 2\left(\frac{180^\circ}{\pi}\right)^2 \frac{T(\lambda)}{\hat{k}_{wi}} \tag{10}$$

$$\hat{k}_{wi} = \frac{n_i - 1}{n_i - R_i}, \tag{11}$$

with $T(\lambda) = \frac{1}{8}(5+18\sin^2\lambda+9\sin^4\lambda)$ the latitude correction introduced in Cox (1970); and R_i the resultant vector length defined in Equation 5. Notice that the within-site dispersion will lead to unrealistic estimates of the between-site dispersion in cases where n_i is small, $n_i = 1$ being the extreme case where the within-site dispersion cannot be estimated; that is, we cannot disentangle the contribution of the within-site and between-site dispersion. For cases where $n_i = 1$, we set $\hat{S}_w = 0$, that is, the within site dispersion is zero since it cannot be estimated from these series of equations.

3 Numerical Results

In this section we present the results of numerical simulations that explore how different sampling strategies affect the estimation of paleopole position μ_0 and VGP scatter S_b. These simulations implement the data generation process described in the Section 2.1 to draw samples of site directions and associated directions within a given site. For the different numerical experiments, we apply varied choices for the model parameters (Table 1) and we respectively compute the mean pole position $\hat{\mu}_0$ and VGP scatter $\hat{S}_b$. These simulations enable us to assess what differences in sampling strategy yield

estimates of the parameters of interest that are closer to the true value. We compare the results of these estimates for different choices of filters and compare them to determine which sampling strategy and method yields the highest accuracy.

3.1 Trade-off between Number of Sites and Number of Samples per Site

The top panel in Figure 1 shows the accuracy of $\hat{\mu}_0$ (Equation (6)) as a function of the number of sites N and the number of samples per site n_0 in the absence of outliers ($p_{\text{outlier}} = 0$). As the number of sites increases (moving up the y-axis), the total error reduces. The mean error is also reduced if we increase the number of samples per site while keeping the total number of sites fixed. However, in the latter case we see that the improvement afforded by increasing the number of samples per site is small relative to increasing the number of sites and saturates for small values of n_0 (see black contour lines).

In a scenario with unlimited resources to collect and analyze paleomagnetic samples, one could seek to maximize both the number of sites (N) and the number of samples per site (n_0). However, in the context of finite resources, it is interesting to consider what happens when we keep fixed the total number of samples $n = n_0 N$ but change how these samples are partitioned between number of sites (N) and number of samples per site (n_0). As visualized with the white dotted curves in Figure 1 that follow a fixed total number of samples, we see that smaller errors are associated with sampling strategies that prioritize the acquisition of additional sites over the collection of additional samples per site. The same behaviour is exposed when we plot the error as a function of the total number of samples n and for different values of n_0 (Figures 2a and 2b). For all choices of samples per site n_0, the net error decreases at rate $1/\sqrt{n}$, with the absolute value of the error being additionally affected by n_0. We quantify the improvement in accuracy due to an increase in the number of samples for different number of samples per site (Figures 2c and 2d). Even by keeping fixed the number of sites and increasing n_0 (and, consequently, increasing the total number of samples), the improvement in accuracy is minimal once $n_0 \geq 3$.

The effect of varied numbers for N and n_0 on the accuracy of estimates of VGP scatter (between-site dispersion S_b) is shown in Figure 1. As with estimating pole position, we observe similar behavior for estimating VGP scatter where, given a fixed total number of samples, there is smaller error when the number of sites is higher. However, the benefit of increasing the number of samples per site on reducing the the root mean square error between $\hat{S}_b$ and the true VGP scatter S_b is more pronounced. Notice that for $n_0 = 1$, this error is large due to the inability to estimate the within-site dispersion. However, for $n_0 \geq 3$ the error stabilizes and we observe the same behaviour as before: the acquisition of more sites over more samples per site leads to better estimation of the VGP scatter assuming $n_0 \geq 3$.

3.2 Sampling Strategy in the Presence of Outliers

From the previous section we concluded that the number of sites N is mostly what determines the accuracy of the estimated position of the paleopole. However, an argument for collecting more samples per site is the ability to detect outliers and filter spurious sample directions. A more fair comparison then is to compare two different strategies for estimating the paleopole while taking the possible occurrence of spurious data into account. When using a small number of samples per site n_0, outlier detection at the site level may be difficult, or directly impossible where $n_0 = 1$. However, it is possible to implement methods to filter VGPs that are statistically significantly apart from the mean (e.g. the paleopole) using an iterative cut-off (Vandamme, 1994). Using this filter provides a upper bound to the error that can be actually been archived by a better strategy to estimate the paleopole in the case $n_0 = 1$. We compare this first strategy

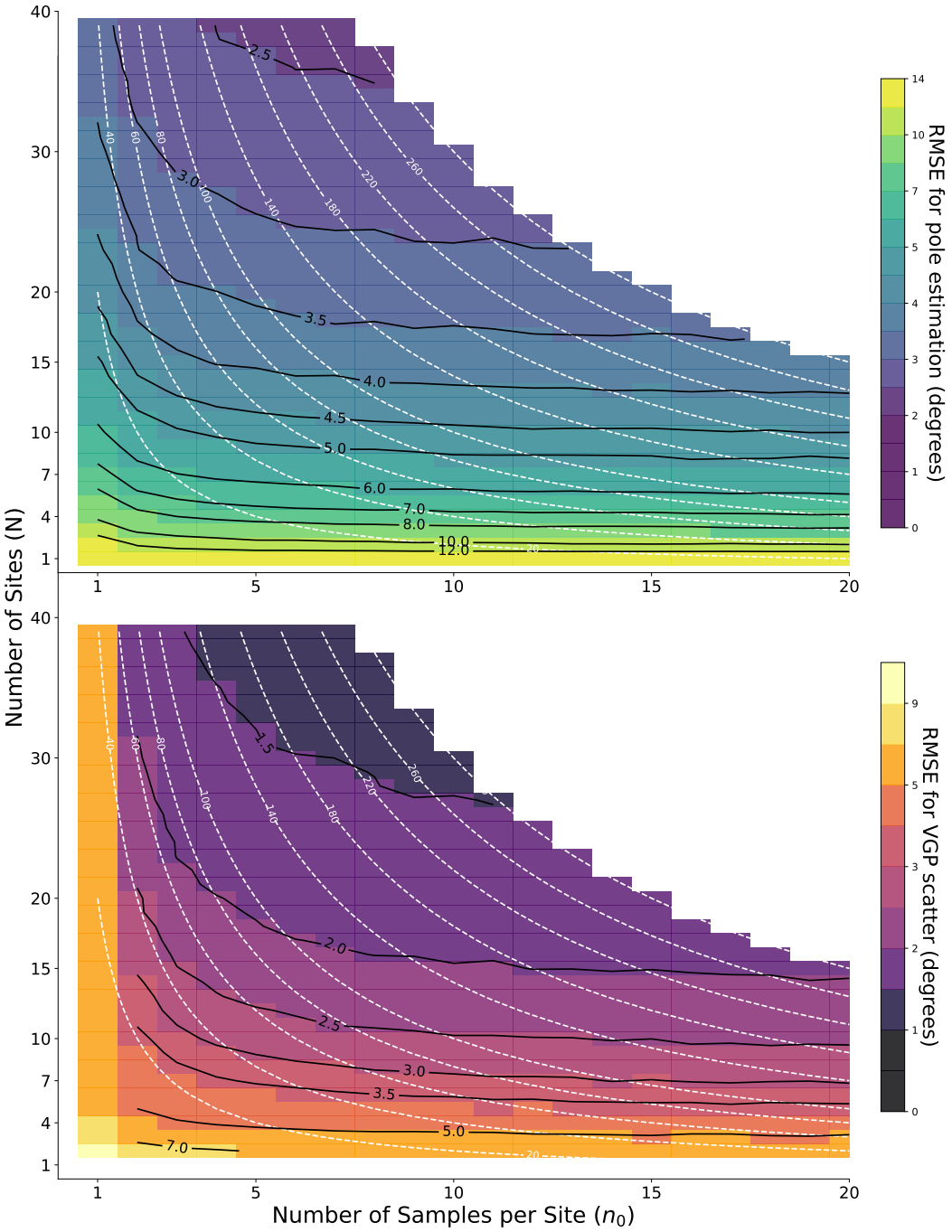


Figure 1: Root mean square error in degrees between site mean poles and the true GAD pole (top panel) and between-site VGP dispersion (bottom panel) as a function of different combinations of the total number of sites N and the number of samples per site n_0. For this diagram, we use a paleolatitude of 30° ($\kappa_b \approx 35$), $p_{\text{outlier}} = 0$, and $\kappa_w = 50$. The white dashed lines represent isolines where the total number of samples n is constant, and the black lines represent isolines with constant net mean error angle. Each point-wise estimate of the mean error (i.e. each box) is based on the results of 10,000 simulations. While these simulations represent secular variation using Model G, similar results emerge from using the TK03 model (Tauxe & Kent, 2004).

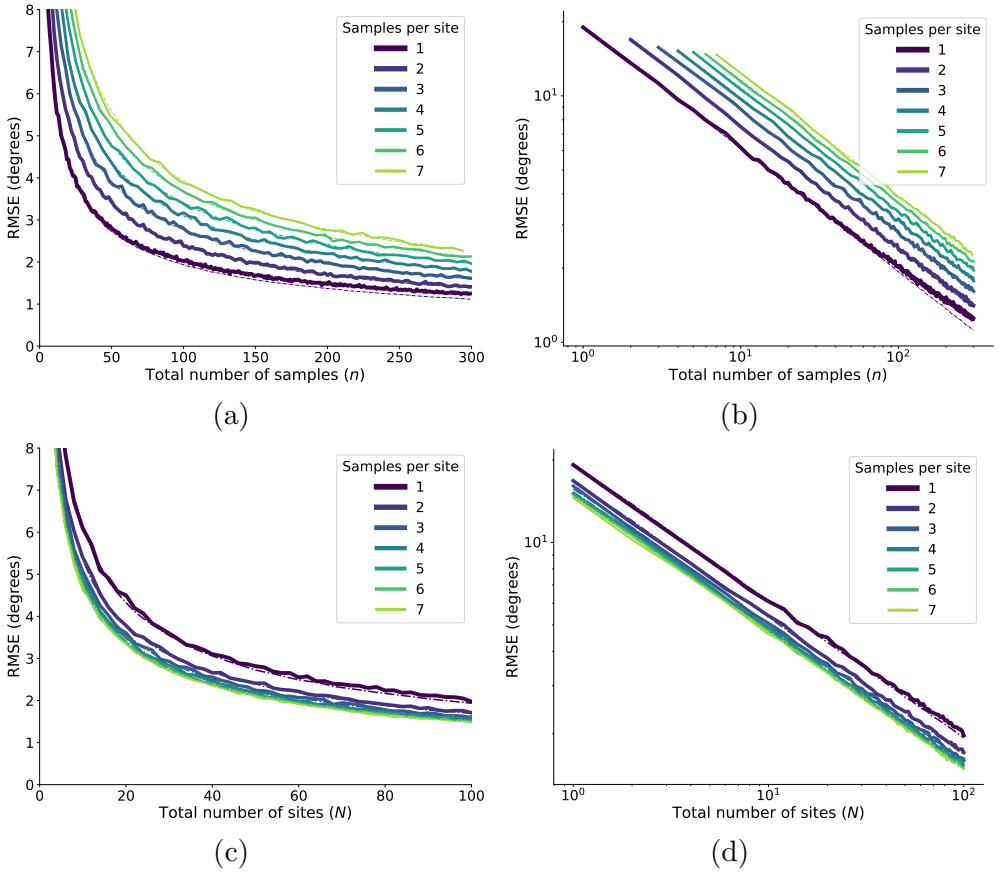
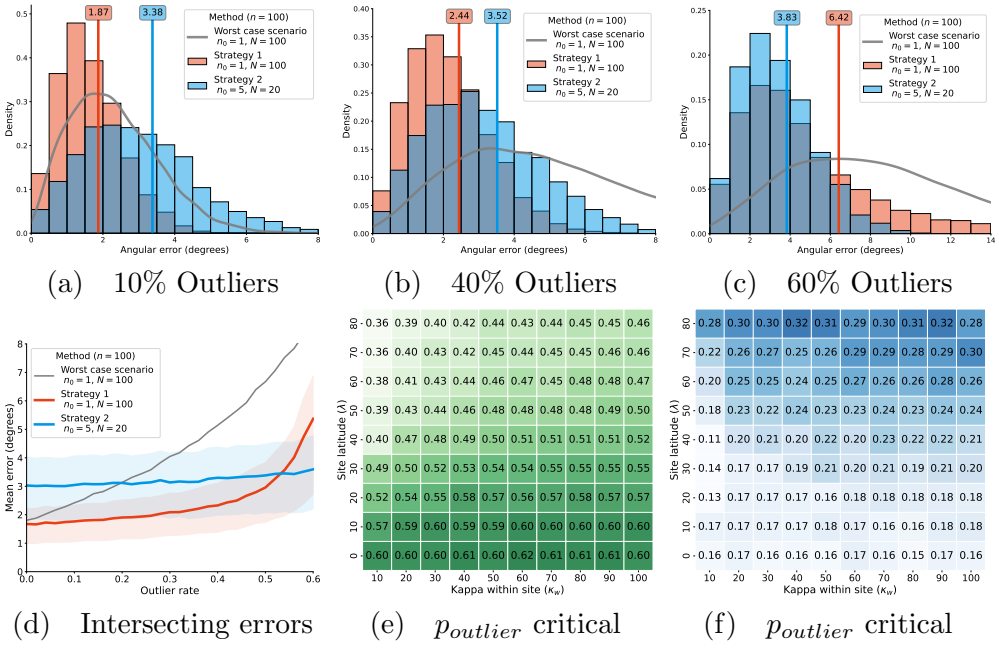


Figure 2: (a) Root mean square error (RMSE) angle of the computed mean pole as a function of the total number of samples n for different values of samples per site n_0 where an increase in samples per site results in a decrease in the number of sites. (b) Displays the same figure but in logarithmic scale, making explicit the $1/\sqrt{n}$ decay of the error, independent of the value of n_0. (c) RMSE as a function of the total number of sites N for different values of n_0, also in (d) logarithmic scale. For all the figures, we set $\lambda = 30°$, $\kappa_w = 50$, and $p_{\text{outlier}} = 0$. The dot-dashed line represents the theoretical approximation (see Section 4).



(a) 10% Outliers (b) 40% Outliers (c) 60% Outliers

(d) Intersecting errors (e) $p_{outlier}$ critical (f) $p_{outlier}$ critical

Figure 3: Comparison between sampling strategies for two different sampling strategies in the presence of outliers for a fixed number of total samples ($n = 100$). The red histograms and curve are strategy 1 where we have one sample per site ($n_0 = 1$), on hundred site ($N = 100$) and we use the Vandamme filter. The blue histograms and curve are strategy 2 where $n_0 = 5$, ($N = 20$) and we filter all the outliers (perfect detection algorithm) for (a) $p_{\text{outlier}} = 0.10$; (b) $p_{\text{outlier}} = 0.4$; and (c) $p_{\text{outlier}} = 0.6$. Here $\kappa_w = 66$ is such that the angular dispersion within site is 10°, and $\lambda = 30°$. The gray line denotes the case in which we sample for $n_0 = 1$ but we do not use any outlier detection method. (d) As we increase the number of outliers p_{outlier}, the error increases differently depending on whether we can detect and filter outliers or not. The intersection of the two errors corresponds to the value of p_{outlier} whereupon there is a crossover in the efficacy of the two methods. The shaded envelopes around the solid lines correspond to the $(0.25, 0.75)$ percentile bands. (e) Value of the intersection between the mean errors for strategies 1 and 2 (panel d) for different values of latitude λ and dispersion within-site k_w. (f) Same as in (e) but comparing $n_0 = 5$ with the worse case scenario with no outlier detection.

($n_0 = 1$ with Vandamme's iterative cut-off applied on the estimated population of VGPs) with the optimistic case where we collect more samples per site and are able to identify and filter all the outliers directly at the site level. The latter case provides a lower bound on the most optimistic error when using any outlier detection criteria at site level. For this second strategy, no outliers are included in the calculation of the final estimated pole $\hat{\mu}_0$. This means that the effective number of samples used to estimate μ_0 will be less than n, but since the samples removed are spurious directions, we expect the estimate of the paleopole will be more accurate than if we included all the samples in the calculation. We also show the results of the first method without using any outlier filter whatsoever.

Histograms in Figures 3a, 3b and 3c show the distribution of the angles between μ_0 (true GAD pole) and $\hat{\mu}_0$ (estimated pole) for the two sampling strategies and with 10%, 40% and 60% outlier rate, respectively. Even in the presence of outliers, using $n_0 = 1$ gives lower angular errors than when using $n_0 = 5$ until the proportion of outliers p_{outlier}

increases by a significant amount. We illustrate this by showing in Figure 3d the mean of these two errors as a function of the outlier rate p_{outlier}. Until the proportion of outliers reaches a critical point of approximately 55%, having $n_0 = 1$ but being able to sample more sites N still out-performs the case where $n_0 = 5$ and all outliers are removed. Figure 3e shows this critical value of p_{outlier} for different site latitudes and within-site dispersion, showing that we need to have more than 40% outliers before the second strategy out-performs the $n_0 = 1$ strategy. Panel 3f further shows this critical value in the case where no filter is used for $n_0 = 1$. It is noteworthy that despite the small variance, this critical value of p_{outlier} grows as a function of site latitude (increasing S_b) and remains almost the same as a function of within-site dispersion.

A wider comparison of these methods for a range of samples per site n_0 is provided in Figure 4. Here again we can observe that for a fixed number of total samples the scenario with $n_0 = 1$ leads to better estimation of the true pole until the proportion of outliers becomes very high. On the right side of the panel we can also observe the improvement in accuracy when we fix the number of sites N and we increase the number of samples per site and thus the total number of samples. In agreement with Figure 2, we observe that the improvement due to an increase in the number of samples per site n_0 by keeping N fixed is small compared to a change in the overall sampling strategy.

We conducted the same analysis for estimating the VGP scatter S_b and its associated error. Figure 5 shows the signed percentage error $100\% \cdot (\hat{S}_b - S_b)/S_b$ for different choices of n_0. When $n_0 = 1$, all methods overestimate the real VGP scatter due to the lack of estimates of the within site dispersion S_w^2 (Equation (9)). On the other hand, S_b tends to be underestimated when we use the Vandamme (1994) filter, since the cut-off of outliers reduces the total dispersion of the VGPs (Equation (7)). As we increase the number of outliers, we observe a significant deterioration of the VGP scatter estimation due to the inability to filter outliers. This behaviour is rather different to what we observed for paleopole estimation, where the estimation is more robust to outliers. However, after reaching a minimum required value of samples per site (around $n_0 = 3$), the accuracy only minimally improves by adding more samples per site. In the case where no outliers are present, we are back to the case in Figure 1 where we observed that, for the same budget of total samples n, a larger value of sites N leads to more accurate estimates as long as $n_0 \geq 3$.

4 Theoretical Results

We can quantify the trade-offs between the different model parameters introduced in the previous section by theoretically deriving approximations for the dispersion parameter of the distribution of the estimated pole $\hat{\mu}_0$. This procedure works by finding the effective precision parameter κ_{eff} of a Fisher distribution that minimizes the Kullback-Leibler divergence with respect to the actual dispersion of $\hat{\mu}_0$ (Kurz et al., 2016; Heslop & Roberts, 2020). As is it derived in Kurz et al. (2016), this is equivalent to finding the mean direction and dispersion parameter that matches the resultant vector length of the target distribution. In Appendix A, we have provided the essential definitions and theoretical derivations used in our analysis. Using this method, we can derive the following approximation for the dispersion of the estimated $\hat{\mu}_0$:

$$\hat{\mu}_0 \approx \text{Fisher}(\mu_0, \kappa_{\text{eff}}), \qquad \kappa_{\text{eff}} = \frac{N\kappa_b}{1 + \frac{\kappa_b}{n_0\,(1-p_{\text{outlier}})\,\kappa_w\,T(\lambda)}}. \tag{12}$$

The effective dispersion parameter κ_{eff} is a function of all the parameters in the model. Under the assumptions of model G (McFadden et al., 1988), we have $\kappa_b = \kappa_b(\lambda)$ is a function of the paleolatitude according to Equation (2). However, this results holds for other choices of κ_b where the Fisher approximation of the VGP scatter is appropriate.

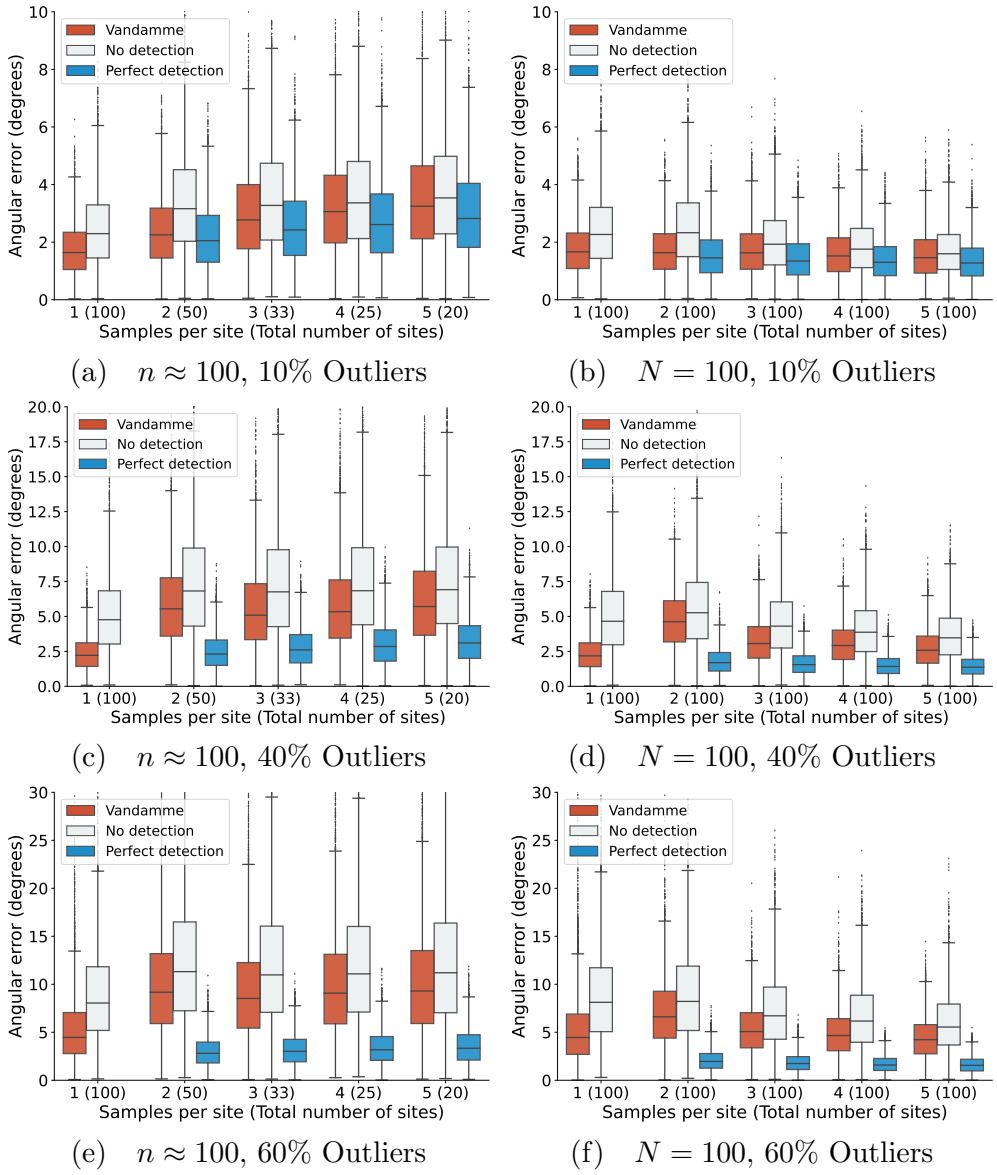
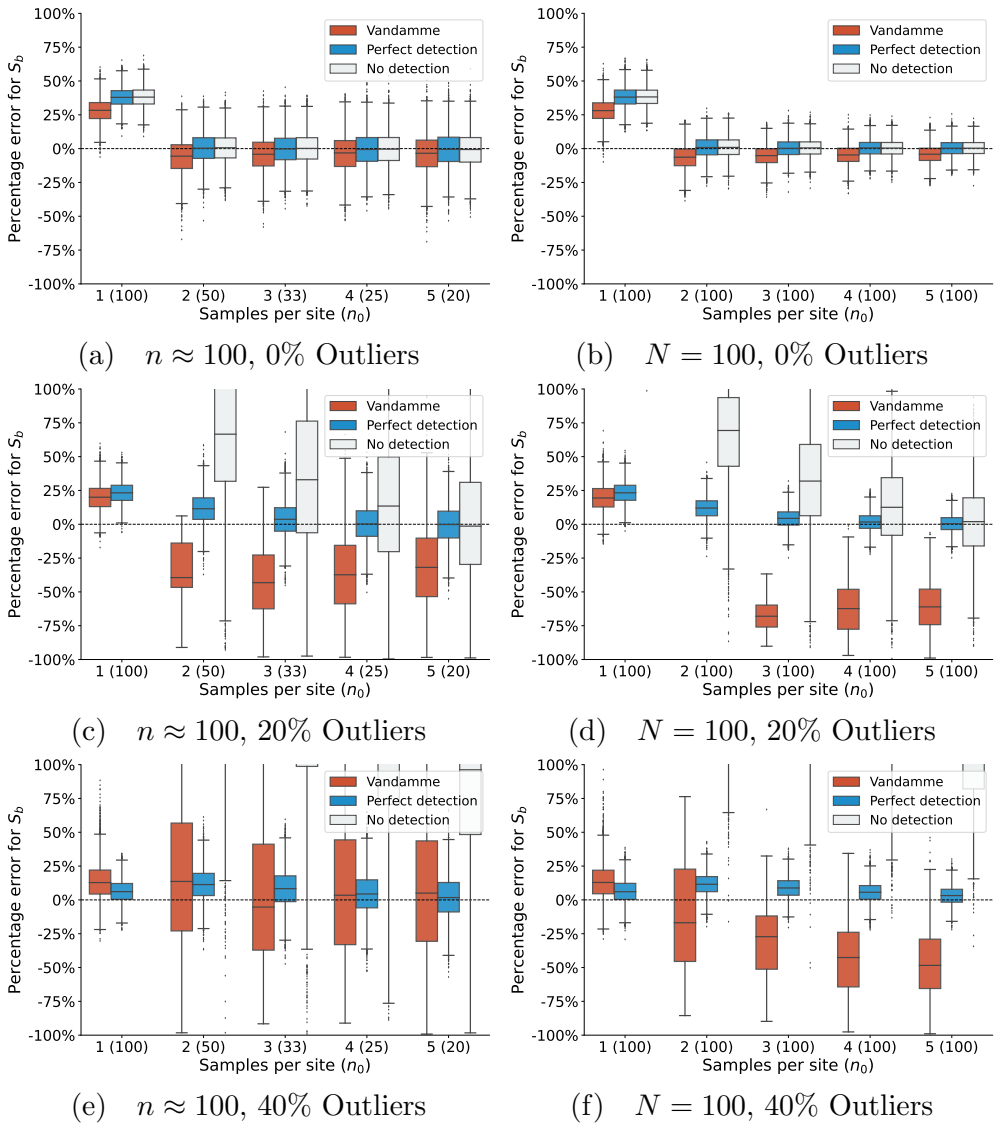


Figure 4: Boxplot of the angular error between estimated and true GAD pole for different sampling strategies (number of samples per site, and total number of sites in parenthesis) for (a,b) $p_{\text{outlier}} = 0.10$, (c,d) $p_{\text{outlier}} = 0.40$ and (e,f) $p_{\text{outlier}} = 0.60$. The left column corresponds to the case where the total number of samples is fixed around $n \approx 100$, while the right column is the case with fixed $N = 100$. Following the convention in Figure 3, the red diagrams correspond to $n_0 = 1$ using the Vandamme filter; the blue to $n_0 = 5$ with perfect outlier detection algorithm; and the grey boxes correspond to $n_0 = 1$ with no outlier detection been applied. Here $k_w = 50$ and $\lambda = 30°$.



(a) $n \approx 100$, 0% Outliers (b) $N = 100$, 0% Outliers

(c) $n \approx 100$, 20% Outliers (d) $N = 100$, 20% Outliers

(e) $n \approx 100$, 40% Outliers (f) $N = 100$, 40% Outliers

Figure 5: Boxplot of the relative error when estimating the between-site dispersion S_b, that is, $100\%(\hat{S}_b - S_b)/S_b$, where $\hat{S}_b$ is estimated as it was explained in Section 2.3, and S_b is the true VGP scatter. Parameters, color references and panel arrangements are the same than in Figure 4, while here the choice of outliers rates is (a,b) $p_{\text{outlier}} = 0$, (c,d) $p_{\text{outlier}} = 0.20$ and (e,f) $p_{\text{outlier}} = 0.40$.

In the case where no outliers are included ($p_{\text{outlier}} = 0$), based on the approximated relationship between angular dispersion S and κ we can approximate the angular error $\text{Err}_{\hat{\mu}_0}$ introduced in Equation (6) as

$$\text{Err}_{\hat{\mu}_0} \approx \frac{81^\circ}{\sqrt{N}} \sqrt{\frac{1}{\kappa_b} + \frac{1}{n_0 \kappa_1 T(\lambda)}}. \tag{13}$$

This equation allow us to quantify the amount of error associated with different choices of n_0. Comparing this theoretical approximation with the simulations (Figure 1 and 2) reveals relative error of around 1% between simulation and theory.

From the theoretical expression for $\text{Err}_{\hat{\mu}_0}$ we can see that as n_0 increases, the improvement in accuracy to the final error becomes rather minimal since the coefficient $1/n_0\kappa_1 T(\lambda)$ is dominated by $1/\kappa_b$. Surprisingly, this limit is reached for very small values of n_0, which shows the small amount of improvement that increasing n_0 adds to the final error, especially when we compare this with the decay of the error given by the factor $1/\sqrt{N}$. No matter the choice of n_0, the error goes to zero as N increases. On the other hand, no matter how large n_0 becomes, the overall error will never be lower than $81^\circ/\sqrt{N\kappa_b}$, N being the quantity that controls the overall error most.

The approximation with outliers is accurate for values of which $n_0(1-p_{\text{outlier}})$ is strictly larger than one. For the case of $n_0 = 1$, a more accurate approximation is given by

$$\rho^{-1}\left((1 - p_{\text{outlier}}) \frac{N\kappa_b}{1 + \frac{\kappa_b}{n_0\, \kappa_w\, T(\lambda)}}\right), \tag{14}$$

where $\rho(\kappa) = 1/\tanh(\kappa) - 1/\kappa$ is the expected length of a Fisher distribution with precision parameter κ and ρ^{-1} its inverse. When using a perfect outlier algorithm with $(1-p_{\text{outlier}})n_0 \geq 2$, the approximation in Equation (13) is still appropriate. Further investigation is needed to estimate the final error when using the iterative cut-off method (Vandamme, 1994).

Notice that the theoretical expression for the final dispersion can be used to define confidence intervals around the true pole for a specific study case. Effectively, given a sampling procedure with prescribed N and n_0, we can estimate the dispersion parameters κ_w and κ_b and then, by plugging these into Equations (12) and (13), we obtain a confidence region around the sample estimated pole. This procedure will take into account the hierarchical nature of paleomagnetic samples at the moment of quantifying uncertainty.

5 Recommendations

When the goal is to estimate the position of a paleopole, our results show that the total number of sites N has a far larger impact on accuracy than the number of samples per site n_0. We therefore recommend the following rule of thumb for sample collection where the objective is paleopole estimation: the more samples the better, but efforts to maximize the number of independent sites will have a greater effect on improving accuracy than more samples per site. In particular, the benefit of collecting more samples per site is small for $n_0 \geq 3$ and diminishes at $n_0 \geq 5$. Analyzing more samples than this per site is inadvisable if it will result in fewer overall sites in a given study. As it was concluded in Gerritsen et al. (2022), for the purpose of computing a paleopole and for a fixed total number of samples, it is always better to collect these samples from different sites than to collect more samples per sites. For paleopole estimates, filters based on populations of VGPs can aid in the detection of outliers (e.g. Vandamme, 1994). If there is an appreciable outlier rate, such filtering schemes are necessary when $n_0 = 1$ given that outliers cannot be detected through within site consistency. We recommend

the use of Equation (13) to get an estimate of the net error as a function of the expected parameters present in the sampling.

An important caveat concerning the use of directional filters is that while the mean may be relatively insensitive to their effects, they can significantly distort the shape of the true directional distribution and should therefore be avoided where the latter is a parameter of interest (e.g. paleosecular variation studies). Indeed, the presence of outliers has a major impact on the estimation of the dispersion, and thus the VGP scatter S_b. Increasing the number of samples per site n_0 is beneficial as long as this helps us to detect outliers more accurately. However, this is not always straightforward using conventional data filters and cutoffs, which leads to a reliance on the expert's subjective interpretation (Gerritsen et al., 2022). There is a greater improvement in the accuracy of estimates of VGP scatter through increasing the number of samples per site, even in the absence of outliers, than there is for estimating the mean pole position. However, the improvement in the estimate of the VGP scatter progressively diminishes for increasing samples per site. When outliers can be detected efficiently, and for a minimum of three or four samples per site, the same trade-offs as noted above for paleopole estimation again apply: the preferential collection of more sites over more samples per site leads to more accurate estimates of the VGP scatter. And again, the most optimal sampling scheme given any suite of expected parameters can be determined from the results presented herein.

For general calculations of pole and VGP scatter accuracy, we recommend the interested reader to run their own experiments directly from the source code, which can be executed directly from the cloud using the provided Binder link in the Code Availability section (Project Jupyter et al., 2018).

6 Conclusions and Future Directions

The hierarchical nature of sampling in paleomagnetic investigations is a long-standing practice, but the community's specific default sampling strategies have largely relied upon conventional wisdom. Here we quantitatively explored, both numerically and analytically, the impact of different sampling strategies on the accuracy of estimates of paleopole position and VGP scatter. Our results demonstrate that when the objective is to estimate the position of the time-averaged magnetic pole, a strategy that maximizes the number of sites is always the most favorable. Thus, given an infinite number of possible sites, it would be advantageous to collect as many single-sample sites as possible.

Where an estimate of VGP scatter is sought, the situation changes and the collection of single-sample sites hinders the estimation and exclusion of within-site directional scatter. Here the optimal sampling strategy is more nuanced and the ideal number of samples per site depends on the expected proportion of outliers. However, the same general rule of thumb still applies: beyond some minimum number of samples per site the collection of additional sites should be prioritized over the collection of additional within-site samples.

We also emphasize that beyond these general rules of thumb, we herein provided tools enabling quantitative sampling recommendations to be generated from user-provided expectations. We hope that these may free the community from the adoption of default sampling practices, and rather move towards statistically-informed strategies.

In this work we asserted the accuracy of poleopole estimation and VGP scatter following the estimates introduced in Section 2.2 and 2.3, either under the influence of filters or not. However, the hierarchical nature of the data and the way this estimates are calculated open the floor to other methods that can lead to more accurate and robust estimations.

7 Open Research

All the code used for this work can be found in the GitHub repository `https://github.com/PolarWandering/PaleoSampling` under release `v0.0.1`. The repository includes a series of Jupyter notebooks and a Python package that allows to run all the data simulations and analysis included in this paper. We also provided reproducible support by including a Binder (Project Jupyter et al., 2018) link to execute all the code in the cloud here `https://mybinder.org/v2/gh/PolarWandering/PaleoSampling/HEAD` and a JupyterBook (Community, 2020) link here `https://polarwandering.github.io/PaleoSampling/`. We benefit from the use of PmagPy (Tauxe et al., 2016) for calculations and Dask for parallel computing (Dask Development Team, 2016).

Appendix A Mathematical derivation of approximated dispersion for Fisher distribution

In this section we consider a series of theoretical derivation of the expected dispersion for the estimate of the true pole $\hat{\mu}_0$. These calculations will allow us to approximate the final distribution of the estimated $\hat{\mu}_0$ as a Fisher distribution with precision parameter that will depend of the parameters listed in Table 1. The building blocks that lead to that final results in Equations (12) and (13) consist in finding approximate Fisher distributions for the following procedures:

1. Mean of Fisher distributions (Section A1)
2. Hierarchical sample of two nested Fisher distributions (Section A2)
3. Superposition of Fisher and uniform distributions (Section A3).

Just as we assumed before, we randomly sample a total of N VGPs μ_i in latitude-longitude space from a Fisher distribution with mean μ_0 and concentration parameter κ_b. Then, we sample site measurements x^*_{ij} in directional space from a Fisher distribution with mean μ^*_i and concentration parameter κ_w, where $j = 1, 2, \ldots, n_i$ (see Section 2.1). We are going to use $\rho(\cdot)$ to refer to the function

$$\rho(\kappa) = \frac{1}{\tanh(\kappa)} - \frac{1}{\kappa}, \tag{A1}$$

where κ will refer to the precision parameter of Fisher distributions. It is easy to see that $\rho(\kappa)$ is the expected length of a Fisher distribution with concentration parameter κ (Mardia et al., 2000).

The method for approximating Fisher distributions follows the moment matching procedure also used in Heslop and Roberts (2020). If $p(x)$ represents the probability density function of some random estimate with support in the unit-sphere given by $\mathcal{S}^2 = \{x \in \mathbb{R}^3 : \|x\| = 1\}$), then we aim to find the parameters $\mu \in \mathbb{R}^3$ ($\|\mu\| = 1$) and κ of the Fisher probability density function $q(x; \mu, \kappa)$,

$$q(x; \mu, \kappa) = \frac{\kappa}{4\pi \sinh(\kappa)} e^{\kappa \mu^T x}, \tag{A2}$$

such that they minimized the Kullback–Leibler divergence $D_{KL}(p|q)$ given by

$$\min_{\mu,\kappa} D_{KL}(p|q) = \int_{\mathcal{S}^2} p(x) \log \frac{p(x)}{q(x; \mu, \kappa)} dx. \tag{A3}$$

As it was found in Kurz et al. (2016), this is equivalent to finding a Fisher distribution $q(x; \mu, \kappa)$ with same mean direction and mean vector length, where the mean vector (both direction and length) is computed as

$$\int_{\mathcal{S}^2} x\, p(x)\, dx. \tag{A4}$$

The technique then consists in estimating the mean resultant length of the estimated paleopole $\hat{\mu}_0$ and match it with the corresponding Fisher distribution $q(x; \mu, \kappa)$ with same mean resultant length.

A1 Mean of Fisher Distributions

Let us begin with a result about the distribution of the mean of a total of n Fisher distributions with same dispersion parameter κ. The case $n = 1$ is excluded since it leads to a trivial result.

Proposition 1 (Mean of Fisher Distributions)**.** *Consider a sample of $n \geq 2$ independent Fisher distributions x_i, $i = 1, 2, \ldots, n$, with mean μ_0 and precision parameter κ. Then the Fisher mean*

$$\hat{\mu} = \frac{1}{nR} \sum_{i=1}^{n} x_i \qquad R = \left\| \frac{1}{nR} \sum_{i=1}^{n} x_i \right\| \tag{A5}$$

is approximately Fisher distributed with mean direction μ_0 and precision parameter $\kappa n \rho(\kappa)$.

Proof. Following Fisher (1953), the estimated mean Fisher distribution $\hat{\mu}$ can be approximated with a Fisher distribution with mean μ and concentration parameter $\kappa n R$. This can be derived by the fact that the conditional probability of $\hat{\mu}|R = r$ has distribution Fisher$(\mu_0, n\kappa r)$ (Mardia, 1975). Taking then expectation over R we obtain an approximate value of the effective concentration parameter of $\hat{\mu}$. Now, we need to find the expected value of the vector length of the mean estimate R. For $n > 1$, it is easy to see that this last quantity coincides in expectation with the expected length of the Fisher distribution with parameter κ, that is $\mathbb{E}[R] = \rho(\kappa)$ (Mardia et al., 2000; Heslop & Roberts, 2020). On the other side, for $n = 1$ we simply have $R = 1$. Even when this may be seem clear, let us derive a secondary proof that will be useful in the later section. For distributions in the sphere, the following relationship holds ((Mardia et al., 2000), equation 9.2.13)

$$\mathbb{E}\left[R^2\right] = \mathbb{E}[R]^2 + \frac{1}{n}(1 - \mathbb{E}[R]^2), \tag{A6}$$

or equivalently

$$\mathbb{E}[R]^2 = \frac{n\,\mathbb{E}\left[R^2\right] - 1}{n - 1}. \tag{A7}$$

This last equation is useful because it allow us to compute the expected value of R as a function of the expected value of R^2, which is mathematically easier to manipulate. Now,

$$R^2 = \frac{1}{n^2} \sum_{i,j=1}^{n} x_i^T x_j = \frac{1}{n^2} \sum_{i=1}^{n} \|x_i\|^2 + \frac{1}{n^2} \sum_{i \neq j} x_i^T x_j. \tag{A8}$$

Now, taking expectation and using $\|x_i\| = 1$ and the independence of the x_i we have

$$\mathbb{E}\left[R^2\right] = \frac{1}{n} + \frac{n-1}{n}\,\mathbb{E}\left[x_1^T x_2\right]. \tag{A9}$$

The only thing that remains to be calculated is the expectation of the cosine of the angle between independent Fisher distributed vectors $x_1^T x_2$. However, notice

$$\mathbb{E}\left[x_1^T x_2\right] = \rho(\kappa)\,\mathbb{E}\left[\mu_0^T x_2\right] = \rho(\kappa)^2 \tag{A10}$$

which then leads to

$$\mathbb{E}\left[R^2\right] = \frac{1}{n} + \frac{n-1}{n}\rho(\kappa)^2 \tag{A11}$$

and $\mathbb{E}[R] = \rho(\kappa)$. Finally, we have that $\hat{\mu}$ is approximately Fisher distributed with mean μ and expected concentration parameter equal to $\kappa n \rho(\kappa)$. □

A2 Hierarchical Sampling of Fisher Distributions

Now, let us consider the case where we hierarchically sample Fisher distribution with random mean directions. This emulates the hierarchical computation of mean directions used to estimate paleopole directions.

Proposition 2 (Hierarchical Sampling on Fisher Distributions)**.** *Consider the following hierarchical sampling of Fisher distributed random variables.*

$$\begin{aligned} \mu_1 &\sim \mathit{Fisher}(\mu_0, \kappa_0) \\ x &\sim \mathit{Fisher}(\mu_1, \kappa_1) \end{aligned} \tag{A12}$$

Then the full distribution of x can be approximated by a Fisher distribution with mean μ_0 and precision parameter κ^* equal to

$$\kappa^* = \frac{\kappa_0 \kappa_1}{\kappa_0 + \kappa_1}. \tag{A13}$$

Proof. We can write the full probability density function of x by integrating the product of conditional densities over all the possible values of μ_1 in the sphere, that is

$$\begin{aligned} p(x) &= \int_{\mathcal{S}^2} p(x|\mu_1) p(\mu_1|\mu_0) d\mu_1 \\ &= \frac{\kappa_0 \kappa_1}{(4\pi)^2 \sinh(\kappa_0)\sinh(\kappa_1)} \int \exp\left\{\kappa_0 \mu_0^T \mu_1 + \kappa_1 x^T \mu_1\right\} d\mu_1 \\ &= \frac{\kappa_0 \kappa_1}{4\pi \sinh(\kappa_0)\sinh(\kappa_1)} \frac{\sinh(\|\kappa_0\mu_0 + \kappa_1 x\|)}{\|\kappa_0\mu_0 + \kappa_1 x\|}. \end{aligned} \tag{A14}$$

Without lost of generality, we can assign $\mu_0 = (0, 0, 1)$ and then

$$\|\kappa_0\mu_0 + \kappa_1 x\| = \sqrt{\kappa_1^2 x^2 + \kappa_1 y^2 + (\kappa_1 z + \kappa_0)^2} = \sqrt{\kappa_1^2 + \kappa_0^2 + 2\kappa_0\kappa_1 z}. \tag{A15}$$

Now, we need to find the first moment of the previous distribution in order to compute the mean length, which implies solving the integral

$$\int_{\mathcal{S}^2} \frac{\sinh(\sqrt{\kappa_1^2 + \kappa_0^2 + 2\kappa_0\kappa_1 z})}{\sqrt{\kappa_1^2 + \kappa_0^2 + 2\kappa_0\kappa_1 z}} z d\Omega = 2\pi \int_{-1}^{1} \frac{\sinh(\sqrt{\kappa_1^2 + \kappa_0^2 + 2\kappa_0\kappa_1 z})}{\sqrt{\kappa_1^2 + \kappa_0^2 + 2\kappa_0\kappa_1 z}} z dz. \tag{A16}$$

Now, this last integral can be solved analytically as

$$\begin{aligned} &2\pi \int_{-1}^{1} \frac{\sinh(\sqrt{\kappa_1^2 + \kappa_0^2 + 2\kappa_0\kappa_1 z})}{\sqrt{\kappa_1^2 + \kappa_0^2 + 2\kappa_0\kappa_1 z}} z dz \\ &\quad= \frac{2\pi}{\kappa_0^2 \kappa_1^2} \Bigg[(\kappa_0\kappa_1 z + 1) \cosh(\sqrt{\kappa_0 + \kappa_1^2 + 2\kappa_0\kappa_1 z}) \\ &\quad- \sqrt{\kappa_0 + \kappa_1^2 + 2\kappa_0\kappa_1 z} \sinh(\sqrt{\kappa_0 + \kappa_1^2 + 2\kappa_0\kappa_1 z}) \Bigg]_{z=-1}^{z=1} \\ &\quad= \frac{2\pi}{\kappa_0^2 \kappa_1^2} \Bigg((\kappa_0\kappa_1 + 1)\cosh(\kappa_0 + \kappa_1) - (\kappa_0 + \kappa_1)\sinh(\kappa_0 + \kappa_1) \\ &\quad- (\kappa_0\kappa_1 - 1)\cosh(|\kappa_1 - \kappa_0|) + |\kappa_1 - \kappa_0| \sinh(|\kappa_1 - \kappa_0|) \Bigg), \end{aligned} \tag{A17}$$

which lead to the fact that the expected length of the vector x is

$$\begin{aligned} &\frac{\kappa_0\kappa_1}{2\kappa_0\kappa_1 \sinh(\kappa_0)\sinh(\kappa_1)} \Bigg((\kappa_0\kappa_1 + 1)\cosh(\kappa_0 + \kappa_1) - (\kappa_0 + \kappa_1)\sinh(\kappa_0 + \kappa_1) \\ &\quad -(\kappa_0\kappa_1 - 1)\cosh(|\kappa_1 - \kappa_0|) + |\kappa_1 - \kappa_0| \sinh(|\kappa_1 - \kappa_0|) \Bigg) \\ &\quad\asymp 1 - \frac{\kappa_0 + \kappa_1}{\kappa_0\kappa_1}, \end{aligned} \tag{A18}$$

where we use $\sinh(\kappa) \asymp \cosh(\kappa) \asymp e^{\kappa}/2$ for κ large enough. Comparing with the equivalent vector length $\rho(\cdot)$, we obtain that the equivalent dispersion parameter κ^* for the superposition of two Fisher distribution is given by

$$\kappa^* = \frac{\kappa_0 \kappa_1}{\kappa_0 + \kappa_1}, \tag{A19}$$

as we wanted to prove. □

Notice that under the approximation that the dispersion coefficient S can be approximated as

$$S^2 \approx 2\left(\frac{180}{\pi}\right)^2 \frac{1}{\kappa}, \tag{A20}$$

we can then derive that the dispersion S_*^2 associated to κ^* can be approximated as

$$S_*^2 \approx 2\left(\frac{180}{\pi}\right)^2 \frac{1}{\kappa} = 2\left(\frac{180}{\pi}\right)^2 \frac{\kappa_1 + \kappa_2}{\kappa_1 \kappa_2} \approx S_1^2 + S_2^2, \tag{A21}$$

where S_1 and S_2 are the dispersion associated to Fisher distribution with precision parameters κ_1 and κ_2, respectively.

A3 Ensemble of Fisher and Uniform Distributions

We will now consider how to approximate a superposition of Fisher and uniform distribution. This approximation is going to be much more limited than the other ones, due to the fact than a superposition of Fisher and uniform does not have a shape similar to a Fisher distribution. However, when many samples are consider and we are computing the mean of samples coming form this ensemble, this approximation is quite accurate.

Proposition 3 (Superposition of Fisher with Uniform distributions)**.** *Consider the model where we sample a total of n samples x_i, $i = 1, 2, \ldots, n$, from a Fisher distribution with some probability $1 - p_{outlier}$ and with uniform distribution with probability $p_{outlier}$:*

$$\begin{aligned} &x_i \sim \textit{Fisher}(\mu, \kappa) \quad \textit{with probability } 1 - p_{outlier} \textit{ and} \\ &x_i \sim \textit{Unif} \quad \textit{otherwise, for } i = 1, 2, \ldots, n, \end{aligned} \tag{A22}$$

Then the Fisher mean $\hat{\mu}$ of the n samples can be approximated with a Fisher distribution with mean μ and precision parameter equal to $n(1 - p_{outlier})\kappa\rho(\kappa)$.

Proof. In order to compute the dispersion parameter, we need to compute the approximated vector length resulting from adding. Given a total number of $n_0 \leq n$ points that are not outliers, a similar calculation as in the derivation of Equation (A8) leads to

$$\mathbb{E}\left[R^2 | n_0\right] = \frac{1}{n^2}\left(n + n_0(n_0 - 1)\rho(\kappa)^2\right). \tag{A23}$$

Now, using that n_0 has Binomial distribution with success probability $1 - p_{\text{outlier}}$ and a total of N samples, taking expectation over n_0 and noticing that $\mathbb{E}[n_0] = n(1 - p_{\text{outlier}})$ and $\mathbb{E}\left[n_0^2\right] = np_{\text{outlier}}(1 - p_{\text{outlier}}) + n^2(1 - p_{\text{outlier}})^2$, we obtain

$$\mathbb{E}\left[R^2\right] = \frac{1}{n} + \frac{n-1}{n}(1 - p_{\text{outlier}})^2 \rho(\kappa)^2. \tag{A24}$$

This leads to $\mathbb{E}[R] = (1 - p_{\text{outlier}})\rho(\kappa)$ for $n \geq 2$. □

A4 General Fisherian approximation of the pole mean

These last three results allow us to approximate a hierarchical sample of Fisher distributions with a very good level of accuracy. In order to compute the final dispersion of the pole, notice that each estimated VPG $\hat{\mu}_i^*$ in directional space can be approximated as a sample from a Fisher distribution with dispersion parameter

$$n_i \kappa_i (1 - p_{\text{outlier}}) \rho_{n_i}(\kappa_i), \tag{A25}$$

where $\rho_n(\kappa) = \rho(\kappa)$ for $n \geq 2$ and $\rho_1(\kappa) = 1$ (Propositions 1 and 3). We have introduced this extra notation in order to include both the $n_i = 1$ and $n_i \geq 2$ cases in the same expression. Now, since the Fisher mean of the VGPs is computed in directional space, we need to include the latitude correction factor $T(\lambda)$ when we convert these to VGP space (Cox, 1970). This then implies that we can approximate

$$\hat{\mu}_i \sim \text{Fisher}\Big(\mu_i, n_i \kappa_w (1 - p_{\text{outlier}}) \rho_{n_i}(\kappa_w) T(\lambda)\Big). \tag{A26}$$

Finally, since μ_i (the mean direction for $\hat{\mu}_i$) is also Fisher distributed with mean μ_0 and precision parameter κ_b, using Proposition 2 we have that the final pole $\hat{\mu}_0$ will have dispersion parameter equal to

$$\frac{\kappa_b}{1 + \frac{\kappa_w}{\kappa_b (1 - p_{\text{outlier}}) n_i \rho_{n_i}(\kappa_w) T(\lambda)}} \tag{A27}$$

Now, if $n_i = n_0$ are all the same, we can average all the $\hat{\mu}_i$ to came up with the final pole dispersion parameter

$$\hat{\mu}_0 \sim \text{Fisher}\left(\mu_0, \frac{N \kappa_b \rho_N(\kappa_b)}{1 + \frac{\kappa_b}{\kappa_w n_i (1 - p_{\text{outlier}}) \rho_{n_0}(\kappa_w) T(\lambda)}}\right). \tag{A28}$$

Assuming $\rho(\kappa_i) \approx 1$, we then obtain that the final estimate $\hat{\mu}_0$ has a concentration parameter κ^* approximately equal to

$$\frac{N \kappa_b}{1 + \frac{\kappa_b}{\kappa_w (1 - p_{\text{outlier}}) n_0 T(\lambda)}},$$

which is the same expression as in Equation 12. In order to derive Equation (13), we rely again in the approximation of the dispersion given in Equation (A20).

As we mentioned before, Proposition 3 will fail when the number of samples per site n_0 is small and the number of outliers p_{outlier} is large. For those cases, a better approximation is given by Equation (14). This arises from computing the expected vector length without outliers and then multiply the expected vector length by the factor $(1 - p_{\text{outlier}})$, which gives an approximated vector length for this case. We then find the corresponding κ for such resultant length by computationally inverting the function $\rho(\kappa)$.

Acknowledgments

All the authors benefited from a grant from the Peder Sather Center for Advanced Study which is dedicated to promoting research collaborations between UC Berkeley and Norway. FS benefits from the Jupyter meets the Earth project supported by the NSF Earth Cube Program under awards 1928406, 1928374. LCG has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 101025975. LCG and MD received support from the Research Council of Norway (RCN) through its Centres of Excellence funding scheme, project 223272 (Centre for Earth Evolution and Dynamics) and project 332523 (Centre for Planetary Habitability). YZ and NLSH received support through NSF Grant EAR-2148719.

References

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

Community, E. B. (2020, February). *Jupyter book.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.4539666` doi: 10.5281/zenodo.4539666

Constable, C., & Parker, R. (1988). Statistics of the geomagnetic secular variation for the past 5 my. *Journal of Geophysical Research: Solid Earth*, *93*(B10), 11569–11581.

Cox, A. (1970). Latitude dependence of the angular dispersion of the geomagnetic field. *Geophysical Journal International*, *20*(3), 253–269.

Creer, K., Irving, E., & Runcorn, S. (1954). The direction of the geomagnetic field in remote epochs in Great Britain. *Journal of geomagnetism and geoelectricity*, *6*(4), 163–168.

Cromwell, G., Johnson, C., Tauxe, L., Constable, C., & Jarboe, N. (2018). Psv10: A global data set for 0–10 ma time-averaged field and paleosecular variation studies. *Geochemistry, Geophysics, Geosystems*, *19*(5), 1533–1558.

Dask Development Team. (2016). Dask: Library for dynamic task scheduling [Computer software manual]. Retrieved from `https://dask.org`

Deenen, M. H., Langereis, C. G., van Hinsbergen, D. J., & Biggin, A. J. (2011). Geomagnetic secular variation and the statistics of palaeomagnetic directions. *Geophysical Journal International*, *186*(2), 509–520.

Doubrovine, P. V., Veikkolainen, T., Pesonen, L. J., Piispa, E., Ots, S., Smirnov, A. V., ... Biggin, A. J. (2019). Latitude dependence of geomagnetic paleosecular variation and its relation to the frequency of magnetic reversals: observations from the cretaceous and jurassic. *Geochemistry, Geophysics, Geosystems*, *20*(3), 1240–1279.

Fisher, R. A. (1953). Dispersion on a sphere. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *217*, 295 - 305.

Gerritsen, D., Vaes, B., & van Hinsbergen, D. J. (2022). Influence of data filters on the position and precision of paleomagnetic poles: what is the optimal sampling strategy? *Geochemistry, Geophysics, Geosystems*, *23*(4), e2021GC010269.

Heslop, D., & Roberts, A. (2020). Uncertainty propagation in hierarchical paleomagnetic reconstructions. *Journal of Geophysical Research: Solid Earth*, *125*(6), e2020JB019488.

Kurz, G., Pfaff, F., & Hanebeck, U. D. (2016). Kullback-leibler divergence and moment matching for hyperspherical probability distributions. In *2016 19th international conference on information fusion (fusion)* (pp. 2087–2094).

Mardia, K. V. (1975). Distribution theory for the von mises-fisher distribution and its application. In G. P. Patil, S. Kotz, & J. K. Ord (Eds.), *A modern course on statistical distributions in scientific work* (pp. 113–130). Dordrecht: Springer Netherlands.

Mardia, K. V., Jupp, P. E., & Mardia, K. (2000). *Directional statistics* (Vol. 2). Wiley Online Library.

McElhinny, M., & McFadden, P. (2000). *Paleomagnetism: Continents and oceans* (Vol. 73). San Diego: Academic Press.

McFadden, P., Merrill, R., & McElhinny, M. (1988). Dipole/quadrupole family modeling of paleosecular variation. *Journal of Geophysical Research: Solid Earth*, *93*(B10), 11583–11588.

McFadden, P., Merrill, R., McElhinny, M., & Lee, S. (1991). Reversals of the earth's magnetic field and temporal variations of the dynamo families. *Journal of Geophysical Research: Solid Earth*, *96*(B3), 3923–3933.

Opdyke, M. D., & Channell, J. E. (1996). *Magnetic stratigraphy.* Academic press.

Project Jupyter, Matthias Bussonnier, Jessica Forde, Jeremy Freeman, Brian Granger, Tim Head, ... Carol Willing (2018). Binder 2.0 - Reproducible,

interactive, sharable environments for science at scale. In Fatih Akici, David Lippa, Dillon Niederhut, & M. Pacer (Eds.), *Proceedings of the 17th Python in Science Conference* (p. 113 - 120). doi: 10.25080/Majora-4af1f417-011

Tauxe, L., Constable, C., Johnson, C. L., Koppers, A. A., Miller, W. R., & Staudigel, H. (2003). Paleomagnetism of the southwestern usa recorded by 0–5 ma igneous rocks. *Geochemistry, Geophysics, Geosystems*, *4*(4).

Tauxe, L., & Kent, D. V. (2004). Timescales of the paleomagnetic field. *Geophysical Monograph Series*, 101–115. doi: 10.1029/145gm08

Tauxe, L., Shaar, R., Jonestrask, L., Swanson-Hysell, N., Minnett, R., Koppers, A., ... Fairchild, L. (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.

Vandamme, D. (1994). A new method to determine paleosecular variation. *Physics of the Earth and Planetary Interiors*, *85*(1-2), 131–142.

Figure 1.

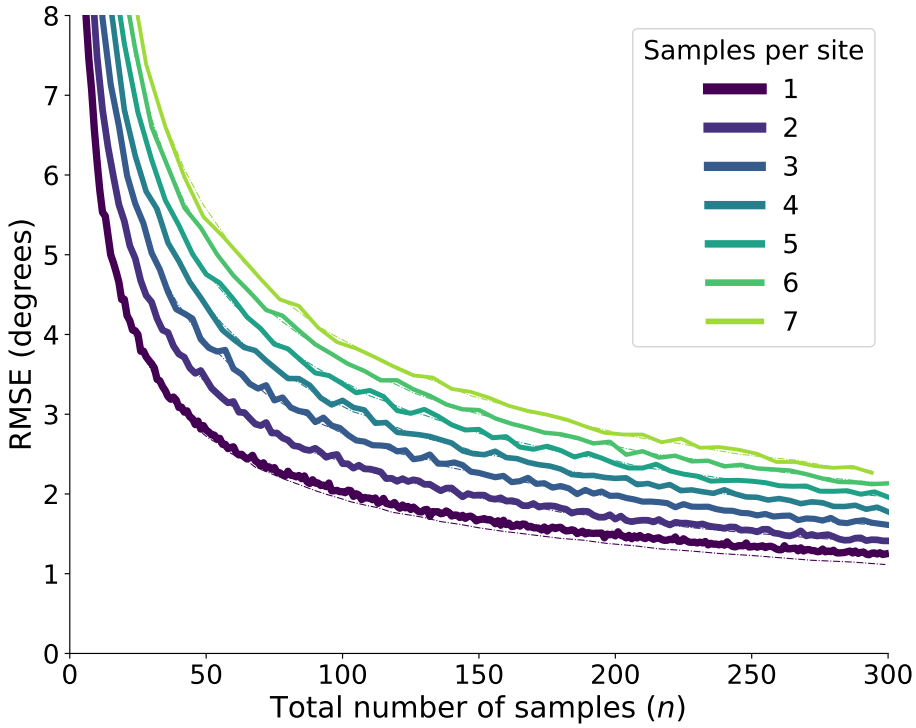
Number of Sites (N)

Number of Samples per Site (n_0)

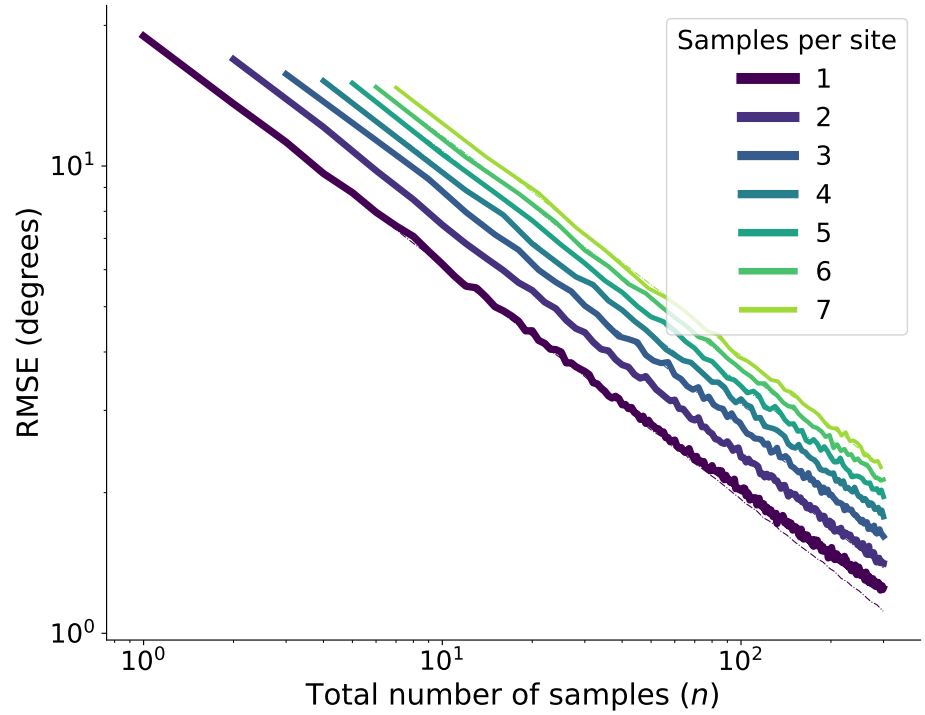
RMSE for pole estimation (degrees)

RMSE for VGP scatter (degrees)

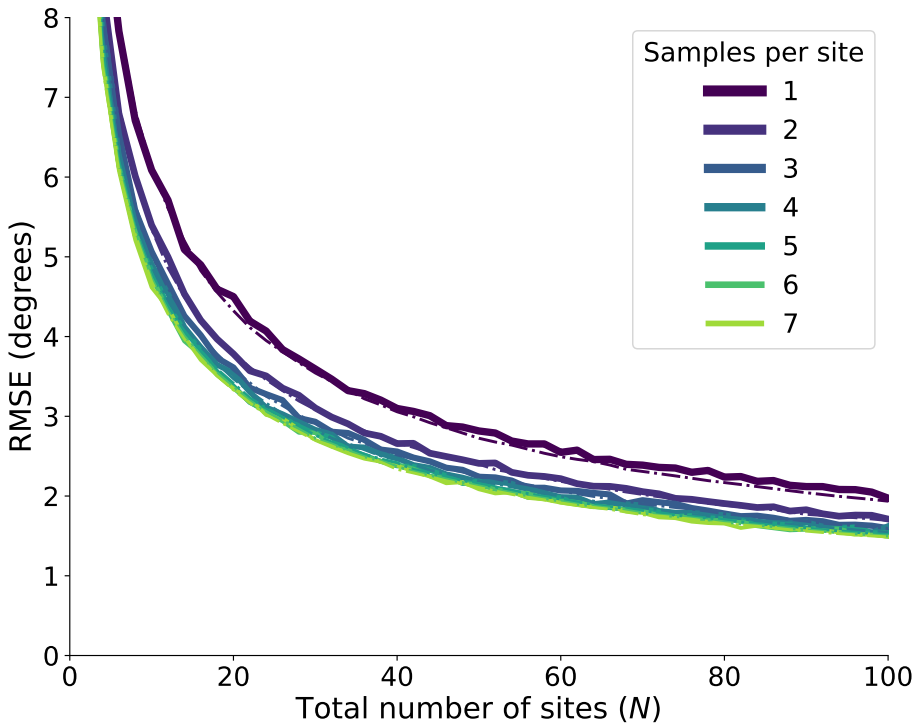
Figure 2.


Samples per site
1
2
3
4
5
6
7
RMSE (degrees)
Total number of samples (n)
0
50
100
150
200
250
300
0
1
2
3
4
5
6
7
8

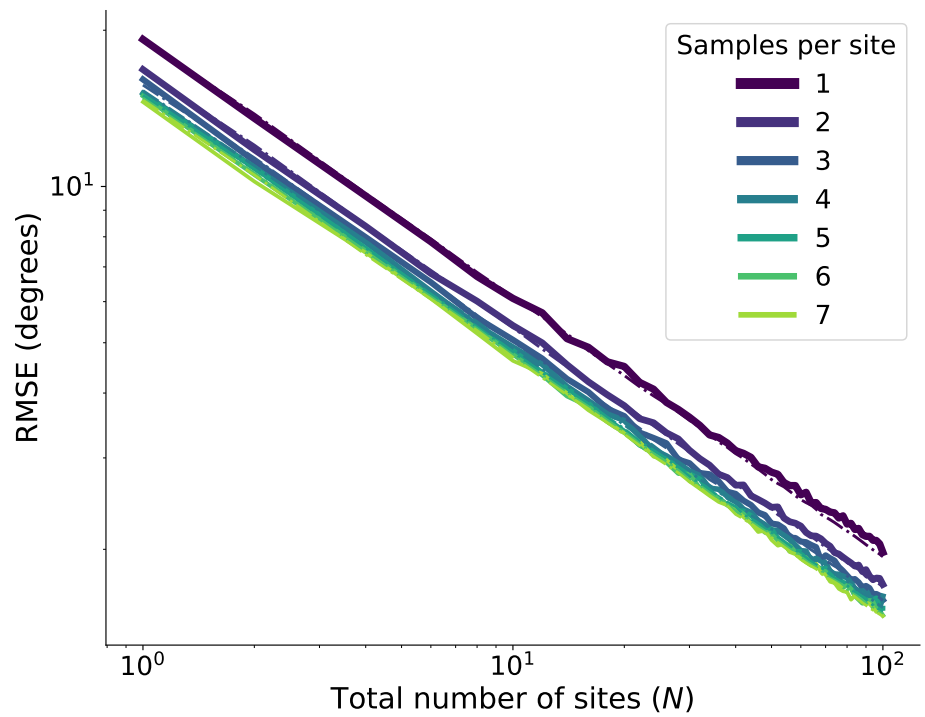
(a)


Samples per site
1
2
3
4
5
6
7
RMSE (degrees)
Total number of samples (n)
10^0
10^1
10^2

(b)

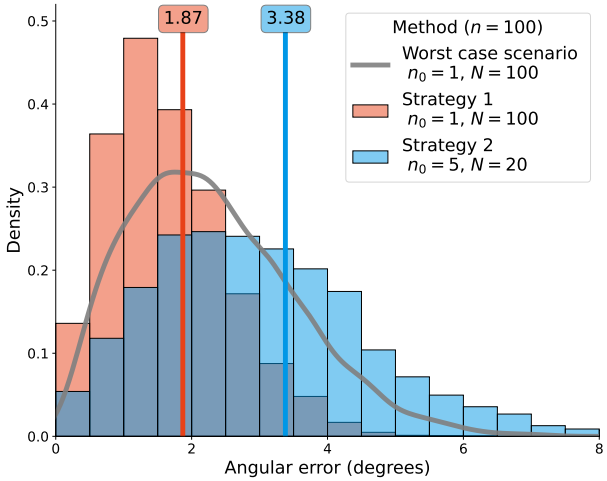

Samples per site
1
2
3
4
5
6
7
RMSE (degrees)
Total number of sites (N)
0
20
40
60
80
100
0
1
2
3
4
5
6
7
8

(c)

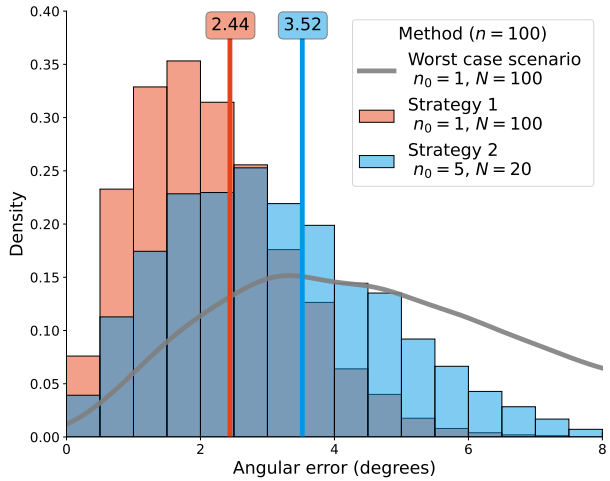

Samples per site
1
2
3
4
5
6
7
RMSE (degrees)
Total number of sites (N)
10^0
10^1
10^2

(d)

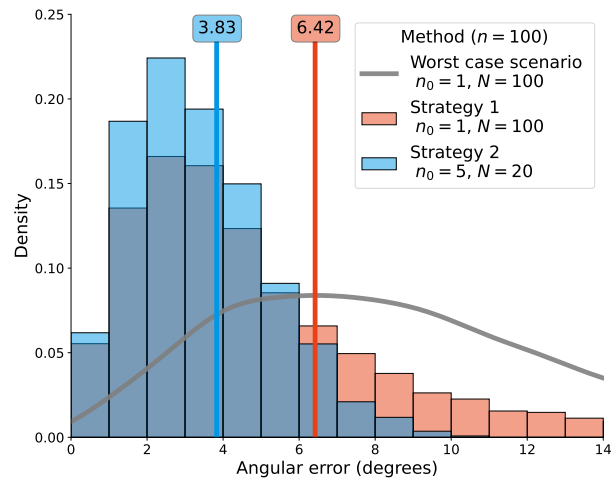
Figure 3.



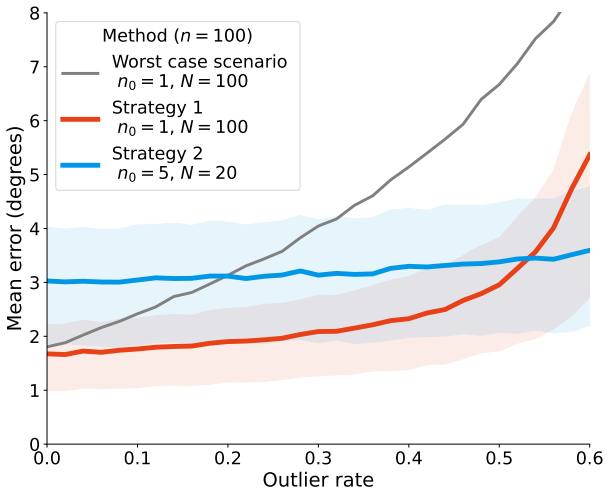
(a) 10% Outliers



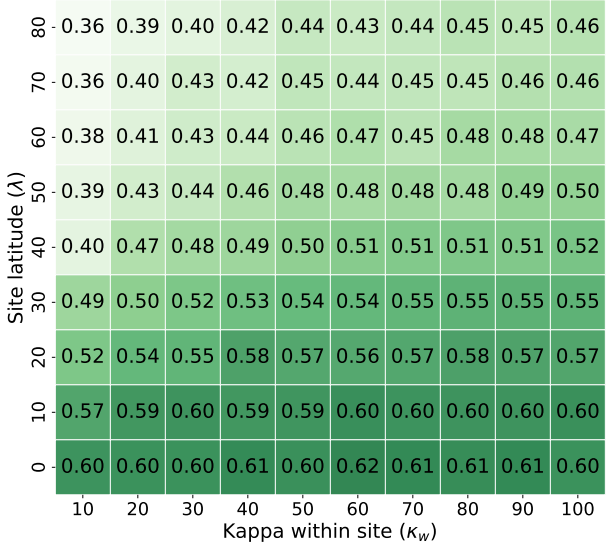
(b) 40% Outliers



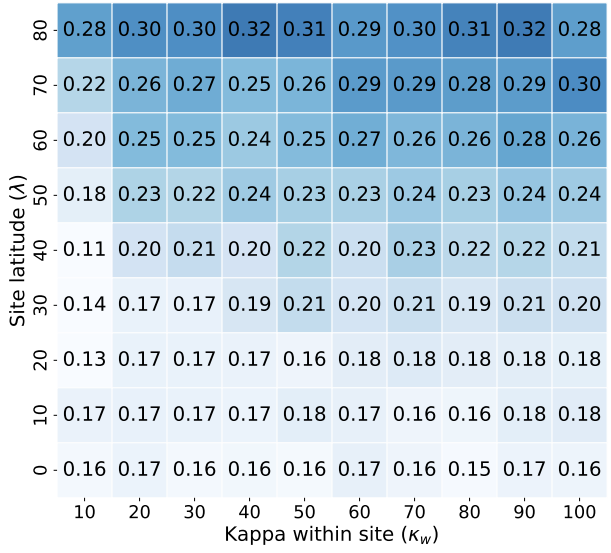
(c) 60% Outliers



(d) Intersecting errors

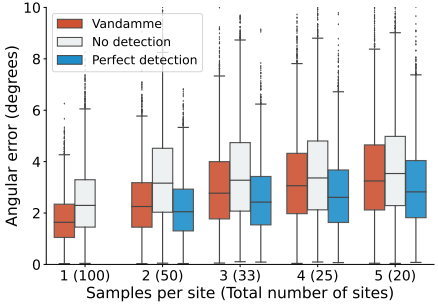


(e) $p_{outlier}$ critical

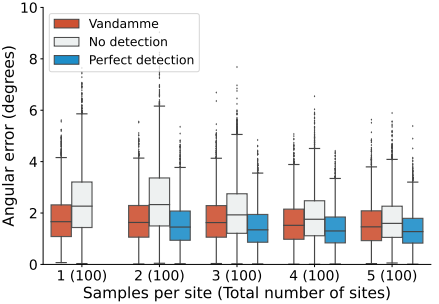


(f) $p_{outlier}$ critical

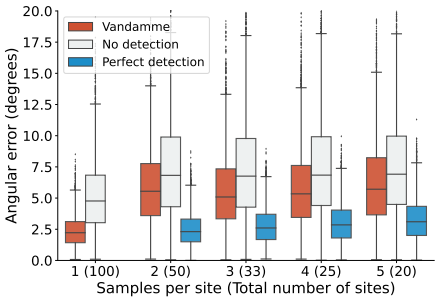
Figure 4.



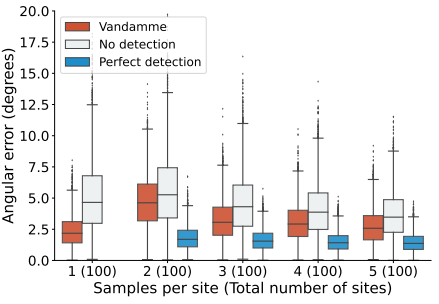
(a) $n \approx 100$, 10% Outliers



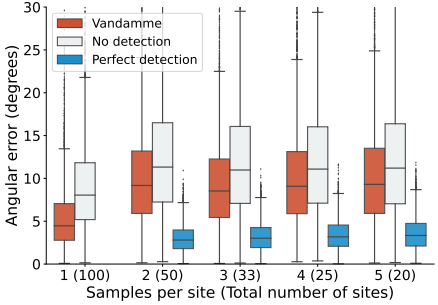
(b) $N = 100$, 10% Outliers



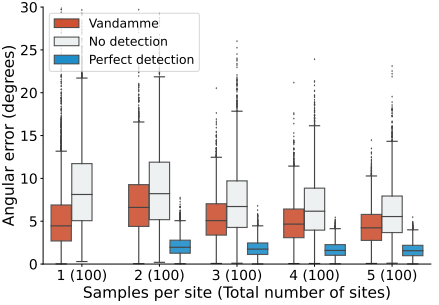
(c) $n \approx 100$, 40% Outliers



(d) $N = 100$, 40% Outliers

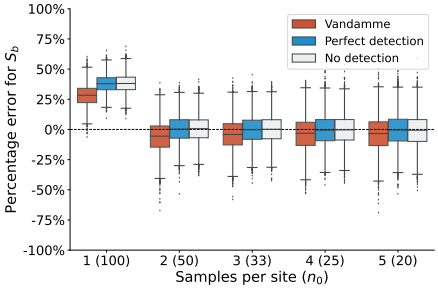


(e) $n \approx 100$, 60% Outliers

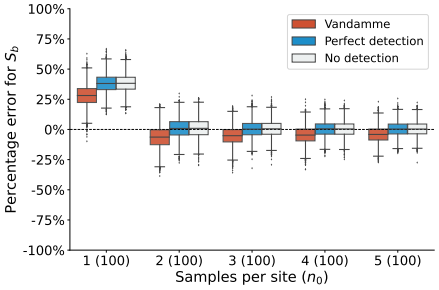


(f) $N = 100$, 60% Outliers

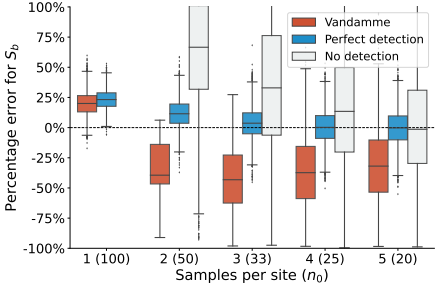
Figure 5.



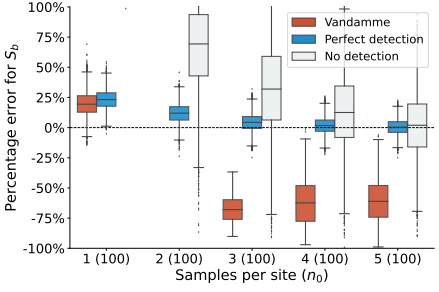
(a) $n \approx 100$, 0% Outliers



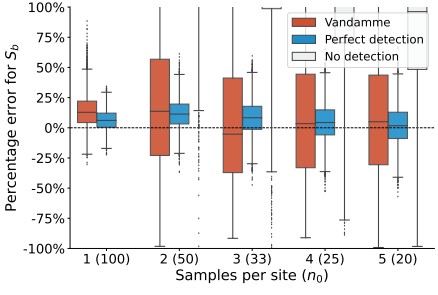
(b) $N = 100$, 0% Outliers



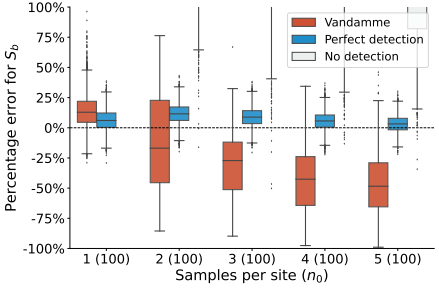
(c) $n \approx 100$, 20% Outliers



(d) $N = 100$, 20% Outliers



(e) $n \approx 100$, 40% Outliers



(f) $N = 100$, 40% Outliers