Abbrev. Journal Name (2018) **0**, 1–32

Methods for Producing a Reliable APWP

Chenjian Fu[1⋆] Chris Rowan[2]
[1] *Department of Geology, Kent State University, 325 S Lincoln St, Kent, OH* 44242*, USA*
[2] *Department of Geology, Kent State University, 325 S Lincoln St, Kent, OH* 44242*, USA*

Received 2018 December 28; in original form 2018 November 22

SUMMARY
First, new picking/weighting methods developed here and previously published picking/weighting methods are compiled together to generate 168 different paleomagnetic APWPs. Then, the APWP similarity measuring tool is used to find which method$s$ is(are) good or bad.

Key words: Moving Average – Weighting – APWP – Paleomagnetism.

1 INTRODUCTION

APWPs are generated by combining paleomagnetic poles for a particular rigid block over the desired age range to produce a smoothed path. See the Supplementary Material for some examples how the pole datasets are constrained first.

1.1 Not All Data Are Created Equal

However, uncertainties in the age and location of paleomagnetic poles can vary greatly for different poles.

1.1.1 Age Error

Although remanent magnetizations are generally assumed to be primary, many events can cause remagnetisation (in which case the derived pole is 'younger' than the rock). If an event that has occurred since the rock's formation that should affect the magnetisation (e.g., folding, thermal overprinting due to intrusion) can be shown to have affected it, then it constrains the magnetisation to have been acquired before that event. Recognising or ruling out remagnetisations depends on these field tests, which are not always performed or possible. Even a passed field test may not be useful if field test shows magnetisation acquired prior to a folding event tens of millions of years after initial rock formation.

The most obvious characteristic we can observe from paleomagnetic data is that some poles have very large age ranges, e.g., more than 100 Myr. The magnetization age should be some time between the information of the rock and folding events. There are also others where we have similar position but the age constraint is much narrower, e.g. 10 Myr window or less. Obviously the latter kind of data is more valuable than the one with large age range.

1.1.2 Position Error

The errors of pole latitudes and longitudes are 95% confidence ellipses, which also vary greatly in magnitude. All paleomagnetic poles have some associated uncertainties due to measurement error and the nature of the geomagnetic field. More uncertainties can be added by too few samples, sampling spanning too short a time range to approximate a GAD field, failure to remove overprints during demagnetisation, etc.

1.1.3 Data Consistency

Paleomagnetic poles of a rigid plate or block should be continuous time series. For a rigid plate, two poles with similar ages shouldn't be dramatically different in location. Sometimes, this is the case. Sometimes we have further separated poles with close ages.

There are a number of possible causes for these outliers, including:

`Lithology`

For poor consistency of data, it is potentially because of different inclinations or declinations. The first thing we should consider about is their lithology. We want to check if the sample rock are igneous or sedimentary, because sediment compaction can result in anomalously shallow inclinations (Tauxe et al. 2018). In addition, we also can check if the rock are redbeds or non-redbeds. Although whether redbeds record a detrital signal or a later Chemical Remanent Magnetization (CRM) is still somewhat controversial, both sedimentary rocks and redbeds could lead to inconsistency in direction compared to igneous rocks.

`Local Rotations`

Local deformation between two paleomagnetic localities invalidates the rigid plate assumption and could lead to inconsistent VGP directions. So if discordance is due to local deformation, and we would ideally want to exclude such poles from our APWP calculation.

`Other Factors`

In most cases, mean pole age (centre of age error) has just been binned. If any of the poles have large age errors, they could be different ages from each other and sample entirely different parts of

⋆ Email: cfu3@kent.edu

the APWP. Conversely, if any of the poles have too few samples, or were not sampled over enough time to average to a GAD field, a discordant pole may be due to unreduced secular variation.

1.1.4 Data Density

As we go back in time, we have lower quality and lower density (or quantity) of data, for example, the Precambrian or Early Paleozoic paleomagnetic data are relatively fewer than Middle-Late Phanerozoic ones, and most of them are not high-quality, e.g., larger errors in both age and location. The combination of lower data quality with lower data density means that a single 'bad' pole (with large errors in age and/or location) can much more easily distort the reconstructed APWP, because there are few or no 'good' poles to counteract its influence.

Data density also varies between different plates. E.g., we have a relatively high density of paleomagnetic data for North American Craton (NAC), but few poles exist for Greenland and Arabia. Based on mean age (mean of lower and upper magnetic ages), for 120–0 Ma, the **Global Paleomagnetic Database** (GPMDB) version 4.6b (Pisarevsky 2005) has more than 130 poles for NAC, but only 17 for Greenland and 24 for Arabia.

1.1.5 Publication Year

The time when the data was published should also be considered, because magnetism measuring methodology, technology and equipments have been improved since the early 20th century. For example, stepwise demagnetisation, which is the most reliable method of detecting and removing secondary overprints, has only been in common use since the mid 1980s.

In summary, not all paleomagnetic poles are created equal, which leads to an important question: how to best combine poles of varying quality into a coherent and accurate APWP?

1.2 Existing Solutions and General Issues

Paleomagnetists have proposed a variety of methods to filter so-called "bad" data, or give lower weights to those "bad" data before generating an APWP, e.g., two widely used methods: the V90 reliability criteria (van der Voo 1990) and the BC02 selection criteria provided by Besse & Courtillot (2002). Briefly, the V90 criteria for paleomagnetic results includes seven criteria: (1) Well determined age; (2) At least 25 samples with Fisher (Fisher 1953) precision κ greater than 10 and $\alpha 95$ less than 16°; (3) Detailed demagnetisation results reported; (4) Passed field tests; (5) Tectonic coherence with continent and good structural control; (6) Identified antipodal reversals; (7) Lack of similarity with younger poles (Torsvik et al. 1992). The total criteria satisfied (0–7) is then used as a measure of a paleomagnetic result's overall reliability, which is known as Q (quality) factor (Torsvik et al. 1992). Q factor is indeed a very straightforward way to get a quantitalized reliability score. Also it then can be conveniently used in the later calculations of APWPs (Torsvik et al. 1992). But at the same time this is a fairly basic filter that lumps together criteria that may not be equally important. Compared with V90, the BC02 criteria suggests stricter filtering, e.g., using only poles with at least 6 sampling sites and 36 samples, each site having $\alpha 95$ less than 10° in the Cenozoic and 15° in the Mesozoic. B02 is also straightforward and convenient to use, but some useful data may be filtered out and wasted especially for a period where there are only limited number of data. In addition, there has been limited study of how effective these marking/filtering methods are at reconstructing a 'true' APWP, and for most studies after a basic filtering of 'low quality' poles, the remaining poles are, in fact, treated equally.

Above all, there haven't been any real attempts to study how APWP fits may be improved by filtering/weighting data. This paper is presented to address these issues.

2 METHODS

For most of Earth history, concretely for times before c. 170 Ma, the age of the oldest magnetic anomaly identification, paleomagnetism is the only accepted quantitative method for reconstructing plate motions and past paleogeographies. After about 170 Ma, multiple data sources can help constrain plate motions in more accurate ways. One of the most developed and studied plate kinematics models is the Fixed Hotspot Model (FHM) (Müller et al. 1993; Müller et al. 1999), which assumes the Atlantic and Indian hotspots are relatively fixed. Another one is the Moving Hotspot Model (MHM) (O'Neill et al. 2005), which is based on mantle convection models that indicate large motions of the Indo-Atlantic hotspots. Such a model like FHM or MHM can predict APWPs for main continents, e.g. the North America (Plate ID 101) (Fig. 1 and Fig. 2), the India (501) (Fig. 3 and Fig. 4) and the Australia (801) (Fig. 5 and Fig. 6). For about 120 Ma to the present, India drifts much faster than North America and Australia. North America has been drifting rather slow. Australia's drifting rate is between them. In addition, India drifts almost north, whereas North America east and then north, and Australia west and then north. So these three continents are representatives for three different types of plate kinematics.

2.1 Reference Path: The Hotspot Model Predicted

The oldest pole that can be predicted from the FHM is about 120 Ma. The North American 120–0 Ma APWP predicted from this rotation model and latest published spreading ridge rotations (collected data will be shared as a supplementary material) will be taken as a reference path (Fig. 1), which will be compared with paleomagnetic APWPs for the same plate or continent.

2.2 120–0 Ma North American Paleomagnetic APWP

The GPMDB 4.6b (Pisarevsky 2005), data source used here, includes 9514 paleopoles for ages of 3,500 Ma to the present published from 1925 to 2016. A polygon can be drawn around a set of data, whose sampling sites we believe belong to a specific plate or rigid block. Then the *Spatial Join* technique (Jacox & Samet 2007) helps join attributes from the polygon to the paleomagnetic data based on the spatial relationship allowing data within this polygon to be extracted from the whole raw large dataset without splitting a subset just for a specific plate. That allows us to quickly select subsets of the database based on geographic constraints just as easily as for age. Of course, the boundary of this polygon must be reasonably along a tectonic boundary (see the details about data filtering for North America in the Supplementary Material). The temporal distribution of North American 120–0 Ma poles is shown in Fig. 7.

Figure 1. FHM predicted 120–0 Ma APWP for NAC through the North America–Nubia–Mantle plate circuit. Its age step is 5 Myr.

2.3 Picking Data for A Certain Time Window

2.3.1 Moving Average

The moving average method, also called "running mean" or "moving window" (Torsvik et al. 2008) method, calculates the average of values between a certain data (age in our case) range; the average is then recalculated as the limits of the bin are repeatedly incremented upwards. In addition to the traditional moving windows averaging algorithm, a newly developed moving average method is also used, referred to here as the "Age Position Picking (APP)" method. The difference of this moving average method from the one built in GMAP (Torsvik & Smethurst 1999; Torsvik et al. 2008) is that the whole magnetic age range is taken into account in each window, while GMAP only considers the mid-point of the low and high magnetic age of each pole, an algorithm referred to as the "Age Mean Picking (AMP)" method.

Normally each VGP in the paleomagnetic database is treated as a point with an age that is the mid-point between the upper and lower age limits, i.e. AMP, but this is problematic for paleomagnetic data with large age ranges (especially if they turn out to be primary magnetization that should plot at old end of age range). We are trying a method, APP, that includes a VGP in the moving average bin if any part of its specified age range falls within that bin. If, for example, we have a pole which is constrained to within 10 and 20 Ma of age, and we have a 2 Myr moving window with a 1 Myr age step, then it shouldn't just be in the 14–16 Ma bin (for the mid-point age of 15 Ma)—it should be in the 9–11,10–12,11–13,12–14 . . . 17–19,18–20, and 19–21 Ma bins. So the average poles are produced from each bin, and each original pole is represented over its entire possible acquisition age. Fig. 10 shows an example of moving average with a 10 Myr window and a 5 Myr step. So, for example, for the window of 15 Ma to 5 Ma (the light blue bin in Fig. 10), the AMP method calculates the Fisher mean pole of only 5 poles, while the APP method calculates the mean pole of 9 poles. From comparison of mean poles of the picked poles for the light blue age window with the two different algorithms (the 10 Ma mean poles in Fig. 1), the mean pole from the APP method is closer to the 10 Ma pole in the FHM predicted path.

In addition, technically we don't want to let the step length

Figure 2. MHM predicted 120–0 Ma APWP for NAC through the North America–Nubia–Mantle plate circuit. Its age step is 5 Myr. The dashed line is the FHM predicted path shown in Fig. 1.

more than the window length, in which cases we would lose data between windows.

2.3.1.1 Picking The 28 picking methods include AMP, APP and also those with filtering or corrections implemented onto the two (Table. 1).

Filtering through $\alpha 95$ and Age Range For this specific filter, the poles are picked out through setting $\alpha 95$ of $\leq 15^\circ$ and age uncertainty of ≤ 20 Myr.

Filtering Out Non-igneous Derived Poles With this filter, the poles are mainly or only from igneous rocks with extrusive or intrusive type.

Filtering Out Igneous Unrelated Poles With this filter, the poles are from rocks that contain extrusive or intrusive igneous type. In other words, the rock type could be mainly sedimentary or metamorphic.

Inclination Shallowing and Unflattening To test if unflattening possible inclination shallowing in sedimentary rocks can improve the APWP fitting outcomes, the flattening function (King 1955) $\tan I_o = f \tan I_f$ is used to unflatten assumed existing inclination shallowing in sedimentary-based or redbeds-based paleomagnetic data (No 5 and 7 in Table. 1), where I_o is the observed inclination, I_f is the unflattened inclination, and f is the flattening factor (or shallowing coefficient) ranging from unity (no flattening)

Figure 3. FHM predicted 120–0 Ma APWP for India through the India–Somalia–Nubia–Mantle plate circuit. Its age step is 5 Myr.

Table 1. List of all Picking (i.e. Binning) algorithms developed in this paper. AMP, Age Mean Picking (See Section "Moving Average"); APP, Age Position Picking.

No.	Picking Algorithm
0	AMP
1	APP
2	AMP ("α95/Age range" no more than "15/20")
3	APP ("α95/Age range" no more than "15/20")
4	AMP (mainly or only igneous)
5	APP (mainly or only igneous)
6	AMP (contain igneous and not necessarily mainly)
7	APP (contain igneous and not necessarily mainly)
8	AMP (unflatten sedimentary)
9	APP (unflatten sedimentary)
10	AMP (nonredbeds)
11	APP (nonredbeds)
12	AMP (unflatten redbeds)
13	APP (unflatten redbeds)
14	AMP (published after 1983)
15	APP (published after 1983)
16	AMP (published before 1983)
17	APP (published before 1983)
18	AMP (exclude commented local rot or secondary print)
19	APP (exclude commented local rot or secondary print)
20	AMP (exclude local rot or correct it if suggested)
21	APP (exclude local rot or correct it if suggested)
22	AMP (filtered using SS05 palaeomagnetic reliability criteria)
23	APP (filtered using SS05 palaeomagnetic reliability criteria)
24	AMP (exclude superseded data already included in other results)
25	APP (exclude superseded data already included in other results)
26	AMP (comb of 22 and 24)
27	APP (comb of 23 and 25)

Notes: SS05, (Schettino & Scotese 2005)

to 0 (completely flattened). Here $f = 0.6$ is used in our calculations, according to the previous experience (Torsvik et al. 2012).

Filtering Out Poles Related to Red-beds Bias toward shallow inclinations is also observed in paleomagnetic data derived from red-beds (Tauxe & Kent 2004). For this filter, the poles derived from red-beds are simply removed.

Filtering Out Poles Published Earlier or Later It is also worthy to see if recently published data is able to produce a more reliable APWP than relatively older data. Here 1983 is chosen as the division, because the mean of the data publication years is about 1983.

Filtering Out Poles Influenced by Local Rotations or Secondary Print Some publications of paleomagnetic data suggest the data has probably been affected by local rotations or secondary overprints. So for this filter, this type of data are removed.

Filtering Out Poles Influenced by Local Rotations or Correcting Them if Suggested Some publications suggest the data has been through local rotation and propose a solution of correction. With this filter, if there is a correction suggestion, the data is corrected; if no, the data is simply removed.

Filtering Using SS05 Liability Criteria SS05 (Schettino & Scotese 2005) provided their criteria of picking paleomagnetic data for producing their APWPs. This filter is reproducing their criteria by setting $\alpha95$ of $\leq 15^\circ$, age uncertainty of ≤ 40 Myr, sampling sites' quantity of ≥ 4, samples' quantity of at least 4 times of the sites, and laboratory analytical procedure code of at least 2.

Filtering Out Superseded Data In this filter, those superseded data already included in other newer results are excluded.

2.3.1.2 Weighting Because all data is not created equal, we want to calculate a weighted mean pole for a time interval with 'better' (more likely to be reliable) poles counting more than 'worse'. For example, a pole with small $\alpha95$ and very well constrained age is more likely to reflect APWP position at the selected age point than a pole with large $\alpha95$ and very broad age range. There are many potential ways to weight this data set which can obviously greatly influence the final result, and we want to test this.

Figure 4. MHM predicted 120–0 Ma APWP for India through the India–Somalia–Nubia–Mantle plate circuit. Its age step is 5 Myr. The dashed line is the FHM predicted path shown in Fig. 3.

Figure 5. FHM predicted 120–0 Ma APWP for Australia through the Australia–East Antarctica–Somalia–Nubia–Mantle plate circuit. Its age step is 5 Myr.

Figure 6. MHM predicted 120–0 Ma APWP for Australia through the Australia–East Antarctica–Somalia–Nubia–Mantle plate circuit. Its age step is 5 Myr. The dashed line is the FHM predicted path shown in Fig. 5.

Table 2. List of all weighting algorithms developed in this paper.

No.	Weighting Algorithm
0	None
1	Numbers of sites (B), Observations (N)
2	Age uncertainty
3	$\alpha 95$
4	Age error Position to bin
5	comb of 3 and 4

Six weighting algorithms (Table. 2) have currently been developed or reproduced according to published work to give different weights to data with different qualities.

In order to average errors in orientation of the samples and scatter caused by secular variation, a "sufficient" number of individually oriented samples (observations) from "enough" sites must be satisfied (Tauxe et al. 2018; van der Voo 1990; Besse & Courtillot 2002). So for the "Numbers of sites (B), Observations (N)" weighting (No 1 in Table. 2), larger B and N mean stronger weighting. Through knowing the pattern of all B and N in the database, the proposed solutions are as follows. If both B and N are more than 1, weight=$(1-\frac{1}{B})*(1-\frac{1}{N})$. There are data in GPMDB with only the number of sampling sites (at least greater than 1) given, but no number of samples or only one sample given, so for this case, if B>1 and N $\leq$ 1, weight=$(1-\frac{1}{B})*0.5$. If only the number of samples (at least greater than 1) is given, and the number of sampling sites is missing or only one, i.e. B$\leq$ 1 and N>1, weight=$(1-\frac{1}{N})*0.5$. If B $\leq$ 1 and N $\leq$ 1 (there are only 23 datasets for the whole GPMDB 4.6b, including 18 with both B and N informations missing), weight=0.2.

As for the "Age uncertainty" weighting (No 2 in Table. 2), a well-constrained age should be known to within a half of a geological period (e.g., Quaternary, Neogene, Triassic) for Phanerozoic data (van der Voo 1990; Tauxe et al. 2018). Generally, this work follows this principle. However, for the periods of Paleogene, Cretaceous, and Jurassic, their halves are all beyond a time span of at least 20 Myr, which is relatively large for these relatively young geologic periods. So for these three periods, a tighter age constraint is set using age uncertainties of $\leq$ 15 Myr. So, for example, for NAC's Neogene (23.03–2.58 Ma according to GSA Geologic Time Scale) data, if age uncertainty (the high magnetic age − the low magnetic age) $\leq$ 10.225 (from 0.5* (23.03−2.58)) Myr, its weight = 1; if age uncertainty > 10.225 Myr, its weight = 10.225 / (high magnetic age − low magnetic age). For the periods spanning Jurassic to Paleogene, if age uncertainty $\leq$ 15 Myr, it get its weight of

Figure 7. Temporal distribution of 120–0 Ma NAC (101) paleomagnetic poles. For distribution a, each bin (spanning 10 Myr here) only counts in the midpoints of pole error bars; For distribution b, as long as the bar intersects with the bin, it is counted in.

1; if age uncertainty $>$ 15 Myr, a weight of 15 / (high magnetic age $-$ low magnetic age) is assigned instead.

For the "α95" weighting (No 3 in Table. 2), smaller radius of circle of 95% confidence about mean remanence direction means less error, so should get larger weight. Here, weight is from a Gaussian distribution centered on 0 with standard deviation of 10, i.e., when $\alpha95 \leq 10$, weight=1; when $\alpha95>10$, weight$<$1. It is also worthwhile to mention that if samples, where two poles are derived, are exactly from the same place and same rock, and one α95 is completely inside the other α95, a zero is assigned as the weight of the data with the larger α95. Here, the same procedure can be applied on A95 (circle of 95% confidence about mean pole) instead of α95.

For the "Age error Position to window" weighting (No 4 in Table. 2), if window intersects with young/old end of age bracket or whole window overlaps with a part of age range, weight= (overlap-

Figure 8. Temporal distribution of 120–0 Ma Indian (501) paleomagnetic poles. For distribution a, each bin (spanning 10 Myr here) only counts in the midpoints of pole error bars; For distribution b, as long as the bar intersects with the bin, it is counted in.

ping part) / (age range width); if whole age range is within window, weight= (window width) / (age range width) (note that when weight > 1, it is set back to 1).

The "Age error Position to window, and $\alpha 95$" weighting (No 5 in Table. 2), is a combination of No 3 and No 4.

2.4 Path Comparison Method

Except the path comparison method (referred to as Measure 1 in the following text) that has been detailedly described in Chapter 2, here is another comparison method (referred to as Measure 2 in the following text) developed originally.

First, the definition of the significant spatial difference d_s is changed into the fraction of $\mathbf{n}$ coeval pole pairs that are statistically distinguishable from each other, as determined by a test for a com-

Figure 9. Temporal distribution of 120–0 Ma Australian (801) paleomagnetic poles. For distribution a, each bin (spanning 10 Myr here) only counts in the midpoints of pole error bars; For distribution b, as long as the bar intersects with the bin, it is counted in.

mon mean direction. Then for the other two shape metrics d_a and d_l, there is no significance testing on them. The way of combining them three into the final composite path difference is still the same.

3 RESULTS AND DISCUSSIONS

3.1 Results

Is there any common pattern of the similarities for all the three continents? First, the best and worst methods need to be determined. Here, the difference values less than the one-standard-deviation interval (containing about 68.269% of the data values) are picked out as the "best" ones (lower about 15.866% of the data values), more

Figure 10. An example of 10 Myr moving window and 5 Myr step in the moving average method, based on poles of the NAC. Every age window has a different color. Red points are the midpoints of low and high magnetic ages. The vertical axis has no specific meaning here.

than the one-standard-deviation interval as the "worst" ones (upper about 15.866%). Then we will see if there is one method or several methods labeled as "best" or "worst" for all the three continents.

3.1.1 When FHM and Plate Circuit Predicted APWP as Reference

First, we focus on analysing the results with FHM and plate circuit predicted APWP (Fig. 1, Fig. 3, Fig. 5) as the reference path.

3.1.1.1 10 Myr Binning and 5 Myr Stepping

Measure 1: Both Space and Shape Tested According to the results (Fig. 11 and Fig. 13), we can observe:

(i) the groups of picking-method-no 19 (APP with commented local-rotation or secondary-print studies excluded) and 21 (APP with local rotation excluded or corrected as suggested in the original sources) (Table. 1) are among the best ones for all the three continents, while the groups of picking-method-no 2 (AMP with "α95/Age range" no more than "15°/20 Myr") and 16 (AMP with only earlier-than-1983 studies), (Table. 1) are among the worst for all them three.

(ii) for both North America (101) and Australia (801), the groups of picking-method-no 1, 11, 13, 19, 21 and 25 (Table. 1) are the best, and 2, 14, 16, 22 and 26 the worst. For both North America (101) and India (501), the groups of picking-method-no 4, 5, 7, 19 and 21 are the best, and 2, 8, 16 and 18 the worst. For both India (501) and Australia (801), the best 19, 21 and the worst 2, 16 are the same as the ones for all the three continents as above-mentioned. These results also further indicate that APP methods generally produce better similarity than AMP methods, however, the picking-method-no 4 is special, which is one of the AMP methods but also one of the best for both North America and India.

(iii) the results of North America (101) and Australia (801) are closer (Fig. 13).

Measure 2: Only Space Tested According to the results (Fig. 14a, Fig. 14b and Fig. 14c), we can observe:

(i) the groups of picking-method-no 1, 19, 21 and 25 are the best, and 2 and 14 the worst, for all the three continents.

3.1.1.2 20 Myr Binning and 10 Myr Stepping

Measure 1: Both Space and Shape Tested According to the results (Fig. 16), we can observe:

(i) the groups of picking-method-no 19 is the best, and 16 the worst, for all the three continents.

Measure 2: Only Space Tested According to the results (Fig. 18), we can observe:

(i) the groups of picking-method-no 11, 13, 19 and 21 are the best, and 16 the worst, for all the three continents.

(a) North America (101): minimum 0.0053 (11, 0), maximum 0.0921 (16, 3), mean 0.0329, median 0.0334

(b) India (501): minimum 0.023 (19, 0), maximum 0.5137 (8, 3), mean 0.1174, median 0.0818

(c) Australia (801): minimum 0.0037 (17, 0), maximum 0.3954 (22, 3), mean 0.0813, median 0.0453

Figure 11. Equal-weight composite path difference ($\mathcal{CPD}$) values with test between each continent's paleomagnetic APWPs and its predicted APWP from FHM and related plate circuit. The paths are in 10 Myr bin and 5 Myr step. The difference values less than one-standard-deviation interval of the whole 168 values are labeled in green, more than one-standard-deviation interval labeled in red.

(a) North America (101): minimum 0.0053 (11, 0)

(b) North America (101): maximum 0.0921 (16, 3)

(c) India (501): minimum 0.023 (19, 0)

(d) India (501): maximum 0.5137 (8, 3)

(e) Australia (801): minimum 0.0037 (17, 0)

(f) Australia (801): maximum 0.3954 (22, 3)

Figure 12. Path comparisons with best and worst difference values shown in Fig. 11.

(a) India (501) - North America (101)

(b) Australia (801) - North America (101)

(c) India (501) - Australia (801)

Figure 13. Differences between grids in Fig. 11. The absolute difference values less than 1.96-standard-deviation interval of the whole 168 values are labeled in green, more than 1.96-standard-deviation interval labeled in red.

(a) North America (101): minimum 0.2536 (1, 1), maximum 0.5104 (14, 4), mean 0.3837, median 0.3912

(b) India (501): minimum 0.282 (6, 5), maximum 0.7637 (8, 3), mean 0.4667, median 0.4532

(c) Australia (801): minimum 0.2246 (11, 0), maximum 0.7554 (22, 3), mean 0.502, median 0.5142

Figure 14. Difference values without shape test between each continent's paleomagnetic APWPs and its predicted APWP from FHM and related plate circuit. The paths are in 10 Myr bin and 5 Myr step. The difference values less than one-standard-deviation interval of the whole 168 values are labeled in green, more than one-standard-deviation interval labeled in red.

(a) North America (101): minimum 0.2536 (1, 1)

(b) North America (101): maximum 0.5104 (14, 4)

(c) India (501): minimum 0.282 (6, 5)

(d) India (501): maximum 0.7637 (8, 3), same as Fig. 12d

(e) Australia (801): minimum 0.2246 (11, 0)

(f) Australia (801): maximum 0.7554 (22, 3), same as Fig. 12f

Figure 15. Path comparisons with best and worst difference values shown in Fig. 14.

(a) North America (101): minimum 0.0115 (13, 0), maximum 0.0635 (16, 2), mean 0.0335, median 0.0285

(b) India (501): minimum 0.0143 (6, 3), maximum 0.1543 (16, 3), mean 0.074, median 0.0755

(c) Australia (801): minimum 0.0022 (11, 0), maximum 0.3265 (22, 3), mean 0.0686, median 0.0346

Figure 16. Same as Fig. 11. The only difference is here the paths are in 20 Myr bin and 10 Myr step. The difference values less than one-standard-deviation interval of the whole 168 values are labeled in green, more than one-standard-deviation interval labeled in red.

(a) North America (101): minimum 0.0115 (13, 0)

(b) North America (101): maximum 0.0635 (16, 2)

(c) India (501): minimum 0.0143 (6, 3)

(d) India (501): maximum 0.1543 (16, 3)

(e) Australia (801): minimum 0.0022 (11, 0)

(f) Australia (801): maximum 0.3265 (22, 3)

Figure 17. Path comparisons with best and worst difference values shown in Fig. 16.

(a) North America (101): minimum 0.2768 (19, 0), maximum 0.4683 (14, 5), mean 0.3773, median 0.3855

(b) India (501): minimum 0.2676 (7, 3), maximum 0.5027 (23, 1; 27, 1), mean 0.3914, median 0.408

(c) Australia (801): minimum 0.1849 (11, 1), maximum 0.5772 (22, 2), mean 0.3882, median 0.3892

Figure 18. Same as Fig. 14. The only difference is here the paths are in 20 Myr bin and 10 Myr step. The difference values less than one-standard-deviation interval of the whole 168 values are labeled in green, more than one-standard-deviation interval labeled in red.

(a) North America (101): minimum 0.2768 (19, 0)

(b) North America (101): maximum 0.4683 (14, 5)

(c) India (501): minimum 0.2676 (7, 3)

(d) India (501): maximum 0.5027 (23, 1; 27, 1)

(e) Australia (801): minimum 0.1849 (11, 1)

(f) Australia (801): maximum 0.5772 (22, 2)

Figure 19. Path comparisons with best and worst difference values shown in Fig. 18.

3.1.2 When MHM and Plate Circuit Predicted APWP as Reference

Then, we focus on analysing the results with MHM and plate circuit predicted APWP (Fig. 2, Fig. 4, Fig. 6) as the reference path.

3.1.2.1 10 Myr Binning and 5 Myr Stepping

Measure 1: Both Space and Shape Tested According to the results (Fig. 20a, Fig. 20b and Fig. 20c), we can observe:

(i) there is no best picking method, whereas no. 16 is the worst, for all the three continents.

Measure 2: Only Space Tested According to the results (Fig. 22a, Fig. 22b and Fig. 22c), we can observe:

(i) the groups of picking-method-no 1, 19 and 21 are the best, and 8, 10 and 14 the worst, for all the three continents.

3.1.2.2 20 Myr Binning and 10 Myr Stepping

Measure 1: Both Space and Shape Tested According to the results (Fig. 24), we can observe:

(i) there is no best picking method, whereas no. 16 is the worst, for all the three continents.

Measure 2: Only Space Tested According to the results (Fig. 26), we can observe:

(i) the groups of picking-method-no 15 and 19 are the best, and 8 the worst, for all the three continents.

3.1.3 Summary of Results

According to all the above results (Fig. 11, Fig. 14, Fig. 16, Fig. 18), we can observe:

(i) the self-explanatory topography of bands indicates that the picking methods (Table. 1) influence the similarity more than the weighting methods (Table. 2) do.

(ii) weighting is not always making similarities better. In fact, for quite many of the methods, no weighting is the best performer (Table. 3).

(iii) generally the APP methods (adding data to a time window with overlapping age selection criterion) produce better similarities than the AMP methods (Table. 3).

(iv) the picking method no 19 is always the best when the FHM predicted APWP is the reference.

(v) the weighting method no 0, 1 and 5 are generally producing better similarity than 2, 3 and 4.

(vi) North America (101) owns better similarity results than Australia (801) and India (501), because its worst and mean composite differences are always less than the other two continents'.

(vii) for both North America (101) and India (501), more recent studies generally give better results than (or results close to) older studies. However, this is not true for Australia (801) (Fig. 11c and Fig. 14c).

3.2 Discussions

The following discussions will be in Q&A style.

3.2.1 Question: Why the APP methods generally produce better similarities than AMP methods do?

3.2.2 Question: Why the AMP methods sometimes unexceptionally produce better similarities than APP methods do?

3.2.2.1 Measure 1: For example, for Fig. 11a, there are only three special (of 84 APP vs. AMP comparisons) cases picking/weighting 4/3, 4/5, 6/3 better than 5/3, 5/5, 7/3 respectively. For the cases of 4/3 vs 5/3 and 6/3 vs 7/3, there are more distinguishable d_a and d_l from the APP methods (Situation 1). For 4/5 vs 5/5, all d_a and d_l are indistinguishable, and the paleomagnetic APWP error ellipses are generally much smaller from the APP methods than from the AMP methods, whereas the coeval pole GCDs are generally not very different for APP and AMP methods. This could further cause more distinguishable d_s or greater distinguishable mean d_s from the APP methods (Situation 2). In addition, APP potentially could generate more mean poles than AMP, because sometimes for some sliding window there is no VGP at all for AMP. If so, because of small number of VGPs, the newly produced mean poles by APP could be far from its contemporary model-predicted pole. This also could potentially bring more distinguishable d_s, d_a and d_l for APP. Admittedly, at the same time, the mean poles at other ages also should be derived from more VGPs from APP, but the extended mean poles from small number of VGPs should be extremely far from its contemporary model-predicted pole, so when this negative effect is beyond the positive effect the other mean poles contribute, the composite difference score would increase (Situation 3). However this phenomenon seldom occur (Table. 3).

Situation 1: Fig. 16a picking/weighting 22/3 vs 23/3. Fig. 20a 6/3 vs 7/3.

Situation 2: Fig. 11b picking/weighting 4/0–5 vs 5/0–5, 22/3 vs 23/3, 26/3 vs 27/3; Fig. 11c picking/weighting 4/2–3 vs 5/2–3, 4/5 vs 5/5. Fig. 16a picking/weighting 2/0–5 vs 3/0–5, 4/0–2 vs 5/0–2, 4/4 vs 5/4, 6/5 vs 7/5, 6/2–3 vs 7/2–3, 22/4–5 vs 23/4–5, 22/1 vs 23/1, 26/4 vs 27/4, 26/2 vs 27/2. Fig. 16b 14/3 vs 15/3. Fig. 16c 4/2 vs 5/2. Fig. 20a 4/0–1 vs 5/0–1. Fig. 20b 4/0–4 vs 5/0–4, 22/0–1 vs 23/0–1, 22/3–4 vs 23/3–4, Fig. 20c 4/3 vs 5/3, 4/0 vs 5/0, 8/5 vs 9/5, 8/3 vs 9/3. Fig. 24b 14/2–3 vs 15/2–3. Fig. 24c 4/4 vs 5/4.

Combination of both Situation 1 and Situation 2: Fig. 11b picking/weighting 22/0–2 vs 23/0–2, 22/4–5 vs 23/4–5, 26/0–2 vs 27/0–2, 26/4–5 vs 27/4–5, Fig. 16a picking/weighting 26/0–1 vs 27/0–1. Fig. 20a 4/3 vs 5/3. Fig. 20b 22/2 vs 23/2, 22/5 vs 23/5, 26/0–5 vs 27/0–5.

Situation 3: Fig. 16b picking/weighting 6/0–5 vs 7/0–5, 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5. Fig. 20a 4/2–5 vs 5/2–5. Fig. 24a 2/0–5 vs 3/0–5, 4/2 vs 5/2, 4/4 vs 5/4, 6/2 vs 7/2, 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5. Fig. 24b 6/0–5 vs 7/0–5, 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5.

3.2.2.2 Measure 2: For example, for Fig. 14a, there are 13 special (of 84 APP vs. AMP comparisons) cases. For the case of 2/3 vs 3/3, coeval segment angular differences d_a are greater from APP method than from AMP method (Situation 1). For the case of 2/0 vs 3/0, 4/5 vs 5/5, 16/2 vs 17/2, 22/0–4 vs 23/0–4, 26/0–3 vs 27/0–3, fraction of distinguishable coeval poles are more from APP method than from AMP method (Situation 2). In addition, segment angular differences d_l could be greater from APP method than from AMP method (Situation 3).

Situation 1: Fig. 18a 10/4 vs 11/4.

Situation 2: Fig. 14b picking/weighting 4/0–5 vs 5/0–5, 6/0–5

(a) North America (101): minimum 0.00077 (4, 0), maximum 0.0926 (16, 5), mean 0.0361, median 0.0325

(b) India (501): minimum 0.0228 (4, 3), maximum 0.5188 (8, 3), mean 0.113, median 0.0746

(c) Australia (801): minimum 0.0024 (17, 5), maximum 0.3662 (22, 3), mean 0.0744, median 0.0374

Figure 20. Same as Fig. 11 except that the reference path is predicted from MHM here.

(a) North America (101): minimum 0.00077 (4, 0)

(b) North America (101): maximum 0.0926 (16, 5)

(c) India (501): minimum 0.0228 (4, 3)

(d) India (501): maximum 0.5188 (8, 3)

(e) Australia (801): minimum 0.0024 (17, 5)

(f) Australia (801): maximum 0.3662 (22, 3)

Figure 21. Path comparisons with best and worst difference values shown in Fig. 20.

(a) North America (101): minimum 0.2147 (11, 0), maximum 0.5183 (16, 3), mean 0.3702, median 0.3642

(b) India (501): minimum 0.2961 (26, 3), maximum 0.7779 (8, 3), mean 0.4578, median 0.4306

(c) Australia (801): minimum 0.2008 (11, 1), maximum 0.7136 (26, 4), mean 0.46, median 0.4726

Figure 22. Same as Fig. 14 except that the reference path is predicted from MHM here.

(a) North America (101): minimum 0.2147 (11, 0)

(b) North America (101): maximum 0.5183 (16, 3)

(c) India (501): minimum 0.2961 (26, 3)

(d) India (501): maximum 0.7779 (8, 3), same as Fig. 21d

(e) Australia (801): minimum 0.2008 (11, 1)

(f) Australia (801): maximum 0.7136 (26, 4)

Figure 23. Path comparisons with best and worst difference values shown in Fig. 22.

(a) North America (101): minimum 0.0058 (15, 5), maximum 0.1041 (8, 4), mean 0.0437, median 0.0267

(b) India (501): minimum 0.0177 (6, 3), maximum 0.162 (16, 3), mean 0.0759, median 0.0621

(c) Australia (801): minimum 0.0047 (23, 4), maximum 0.3158 (26, 3), mean 0.0638, median 0.0307

Figure 24. Same as Fig. 16 except that the reference path is predicted from MHM here.

(a) North America (101): minimum 0.0058 (15, 5)

(b) North America (101): maximum 0.1041 (8, 4)

(c) India (501): minimum 0.0177 (6, 3)

(d) India (501): maximum 0.162 (16, 3)

(e) Australia (801): minimum 0.0047 (23, 4)

(f) Australia (801): maximum 0.3158 (26, 3)

Figure 25. Path comparisons with best and worst difference values shown in Fig. 24.

(a) North America (101): minimum 0.1923 (4, 0), maximum 0.4794 (16, 2), mean 0.3384, median 0.3318

(b) India (501): minimum 0.2462 (6, 0), maximum 0.5417 (8, 3), mean 0.3858, median 0.3659

(c) Australia (801): minimum 0.2091 (5, 5), maximum 0.5777 (14, 0), mean 0.3694, median 0.3676

Figure 26. Same as Fig. 18 except that the reference path is predicted from MHM here.

(a) North America (101): minimum 0.1923 (4, 0)

(b) North America (101): maximum 0.4794 (16, 2)

(c) India (501): minimum 0.2462 (6, 0)

(d) India (501): maximum 0.5417 (8, 3)

(e) Australia (801): minimum 0.2091 (5, 5)

(f) Australia (801): maximum 0.5777 (14, 0)

Figure 27. Path comparisons with best and worst difference values shown in Fig. 26.

Table 3. Performance statistics of all the picking and weighting methods.

Grid		Best No.		Worst No.		Proportion of APP Better Than AMP	For All 28 Picking Methods, Count of Occurrences of Each Weighting No. Being Best						Picking 14/15 (Studies After 1983) Better Than 16/17 (Older)
		Picking	Weighting	Picking	Weighting		0	1	2	3	4	5	
FHM	Fig. 11a	1, 4, 5, 7, 11, 13, 15, 19, 21, 25	0, 1, 5	2, 5, 7, 8, 14, 16, 17, 18, 22, 26	0, 1, 2, 3, 4, 5	27/28	15	4	2	3	1	4	Y/Y
	Fig. 14a	1, 11, 13, 15, 19, 21, 25	0, 1, 2, 4, 5	0, 2, 3, 8, 10, 12, 14, 16, 17, 18, 20, 24	0, 1, 2, 3, 4, 5	71/84	9	13	0	1	1	4	N/Y
	Fig. 11b	4, 5, 6, 7, 9, 19, 21	0, 1, 2, 3, 4, 5	0, 2, 8, 10, 12, 16, 18, 20, 24	0, 1, 2, 3, 4, 5	11/14	17	1	2	7	1	0	Y/Y
	Fig. 14b	1, 4, 6, 9, 17, 19, 21, 25, 26	0, 1, 2, 3, 4, 5	0, 2, 8, 10, 12, 14, 16, 18, 24	0, 1, 2, 3, 4, 5	5/7	13	1	4	5	0	5	Y(4y2n)/Y(4y2n)
	Fig. 11c	1, 11, 13, 17, 19, 21, 25	0, 1, 3, 5	2, 6, 14, 16, 22, 26	0, 1, 2, 3, 4, 5	27/28	11	4	1	8	3	1	N/N
	Fig. 14c	1, 7, 11, 13, 17, 19, 21, 25	0, 1, 3, 5	2, 6, 14, 22, 26	0, 1, 2, 3, 4, 5	1	9	6	2	7	2	2	N/N
	Fig. 16a	1, 4, 5, 7, 11, 13, 15, 19, 21, 25	0, 1, 3, 5	2, 6, 14, 22, 26	0, 1, 2, 3, 4, 5	3/4	18	6	0	2	4	3	N(5n1y)/Y
	Fig. 18a	1, 4, 5, 6, 7, 9, 11, 13, 15, 19, 21, 25	0, 1, 3, 5	0, 3, 8, 10, 11, 12, 14, 16, 17, 18, 20, 24	0, 1, 2, 3, 4, 5	17/21	15	6	0	2	3	2	N/Y
	Fig. 16b	4, 5, 6, 7, 19	0, 1, 2, 3, 4, 5	0, 2, 10, 12, 16, 18, 20, 23, 24, 27	0, 1, 2, 3, 4, 5	59/84	12	6	0	9	0	1	Y/Y(4y2n)
	Fig. 18b	4, 5, 6, 7, 9, 11, 13, 19, 21	0, 1, 2, 3, 4, 5	0, 2, 8, 10, 12, 16, 23, 24, 27	0, 1, 2, 3, 4, 5	11/14	9	3	2	8	3	3	Y(4y2n)/N(5n1y)
	Fig. 16c	1, 11, 13, 17, 19, 21, 25	0, 1, 3, 4, 5	2, 14, 16, 22, 26	0, 1, 2, 3, 4, 5	41/42	7	9	0	11	1	2	N/N(4n2y)
	Fig. 18c	1, 11, 13, 15, 17, 19, 21, 25	0, 1, 3, 5	2, 4, 5, 14, 16, 22, 26	0, 1, 2, 3, 4, 5	41/42	6	8	0	10	1	3	N/N(4n2y)
MHM	Fig. 20a	1, 4, 5, 7, 11, 13, 15, 21, 25	0, 1, 2, 3, 4, 5	0, 8, 10, 12, 14, 16, 17, 18, 20, 22, 24	0, 1, 2, 3, 4, 5	11/12	14	6	1	1	2	4	Y/Y
	Fig. 22a	1, 5, 7, 11, 13, 15, 19, 21, 23, 25, 27	0, 1, 3, 5	0, 8, 10, 12, 14, 16, 17, 18, 24	0, 1, 2, 3, 4, 5	27/28	6	12	0	2	0	8	Y/Y
	Fig. 20b	4, 5, 6, 7, 19, 22, 26	0, 1, 2, 3, 4, 5	0, 2, 8, 10, 12, 16, 18, 20, 24	0, 1, 2, 3, 4, 5	11/14	12	3	0	9	1	3	Y/Y
	Fig. 22b	1, 4, 6, 15, 17, 19, 21, 22, 26	0, 1, 2, 3, 4, 5	0, 2, 8, 10, 12, 14, 16, 24	0, 1, 2, 3, 4, 5	61/84	0	1	7	6	3	11	Y/Y(4y2n)
	Fig. 20c	1, 11, 13, 17, 19, 21, 25	0, 1, 3, 5	2, 14, 16, 22, 26	0, 1, 2, 3, 4, 5	20/21	8	5	1	9	4	1	N/N
	Fig. 22c	1, 11, 13, 17, 19, 21, 25	0, 1, 3, 5	2, 6, 8, 10, 14, 18, 22, 26	0, 1, 2, 3, 4, 5	1	8	3	0	11	1	5	N/N
	Fig. 24a	4, 5, 7, 11, 15, 22, 26	0, 1, 2, 3, 4, 5	0, 8, 10, 12, 14, 16, 18, 20, 24	0, 1, 2, 3, 4, 5	3/4	15	6	1	0	1	5	N(5n1y)/Y
	Fig. 26a	1, 4, 5, 7, 11, 13, 15, 19, 21, 25	0, 1, 3, 4, 5	0, 8, 10, 12, 14, 16, 18, 20, 24	0, 1, 2, 3, 4, 5	6/7	8	6	2	6	3	3	Y(5y1n)/Y(5y1n)
	Fig. 24b	4, 5, 6, 7, 19	0, 1, 2, 3, 4, 5	0, 2, 10, 12, 16, 18, 23, 24, 27	0, 1, 2, 3, 4, 5	29/42	8	8	2	5	3	2	Y/N(5n1y)
	Fig. 26b	6, 9, 15, 17, 19, 22, 23, 26, 27	0, 1, 2, 3, 4, 5	0, 2, 3, 8, 10, 12, 16, 18, 24	1, 2, 3, 4, 5	61/84	5	4	3	5	3	8	Y/N(5n1y)
	Fig. 24c	1, 11, 13, 15, 17, 19, 21, 23, 25, 27	0, 1, 2, 3, 4, 5	2, 14, 16, 22, 26	0, 1, 2, 3, 4, 5	41/42	9	5	0	8	5	1	N/Y(4y2n)
	Fig. 26c	1, 5, 7, 11, 13, 15, 19, 21, 23, 25	0, 1, 3, 4, 5	2, 6, 8, 14, 20, 22, 26	0, 1, 2, 3, 4, 5	1	7	4	0	12	4	1	N/Y(4y2n)

vs 7/0–5, 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5. Fig. 18a 2/4–5 vs 3/4–5, 2/0–2 vs 3/0–2, 4/1–2 vs 5/1–2, 4/4 vs 5/4, 6/0 vs 7/0, 6/4–5 vs 7/4–5, 22/4 vs 23/4. 26/4 vs 27/4, 26/0 vs 27/0. Fig. 22b 6/0–1 vs 7/0–1, 6/3–4 vs 7/3–4. Fig. 26a 22/1 vs 23/1, 22/3–4 vs 23/3–4, 26/0–1 vs 27/0–1, 26/3–5 vs 27/3–5.

Combination of Situation 1 and Situation 2: Fig. 18a 16/4 vs 17/4. Fig. 18b 22/0–5 vs 23/0–5. 26/0–3 vs 27/0–3, 26/5 vs 27/5. Fig. 18c 4/2 vs 5/2, 6/2 vs 7/2. Fig. 22a 2/1 vs 3/1, 16/1–2 vs 17/1–2. Fig. 22b 4/0–5 vs 5/0–5, 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5. Fig. 26a 4/4 vs 5/4, 4/2 vs 5/2

Situation 3: Fig. 18b 6/2 vs 7/2.

Combination of Situation 2 and Situation 3: Fig. 22b 6/2 vs 7/2.

Combination of Situation 1 and Situation 3: Fig. 26a 4/3 vs 5/3, 4/0 vs 5/0

Combination of Situations 1, 2 and 3: Fig. 26b 6/0–3 vs 7/0–3, 14/2 vs 15/2, 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5.

3.2.3 Question: Why weighting is not affecting?

For Fig. 11a Picking No (Pk) 0, Weighting No (Wt) 1–4 cause 80–85 Ma coeval segments distinguishable (d_a and d_l) and Wt 2 and 4 cause one more pair of distinguishable mean poles (d_s) than other weighting methods. However, compared with Wt 0, Wt 5 does affect and bring a better similarity, because Wt 5 decreases the GCDs of the 20, 80 and especially 100 Ma distinguishable pole pairs (by more than 2 degrees).

For Fig. 11a Pk 1, Wt 2–5 cause at least 3 more pairs of distinguishable mean poles (d_s) than other weighting methods. Wt 1 increases the GCDs of all the distinguishable pole pairs (0, 5 and 120 Ma).

For Fig. 11a Pk 2, Wt 1–5 cause at least 1 more pair of distinguishable mean poles (d_s) than Wt 0 (60 Ma). Wt 4 increases another pair of distinguishable mean poles (d_s) (80 Ma).

3.2.4 Question: Why best and worst methods are not consistent?

3.2.4.1 Question: Why the picking method 21 is not among the best for 20 Myr binning and 10 Myr stepping with both

Figure 28. Paleomagnetic APWPs' mean A95 versus (a) "mean GCD", (b) "mean significant GCD", (c) "mean significant orientation difference", and (d) "mean significant length difference" between paleomagnetic APWP and its corresponding FHM-and-plate-circuit predicted APWP. Starting points of the arrows are results from AMP, while ending points are from APP. Black color filled arrow heads are the small number of special cases of AMP derived equal-weight CPDs better than APP (see details in Fig. 11).

space and shape tested? As shown in Fig. 11, Fig. 14 and Fig. 18, both the picking methods 19 and 21 are among the best. However, the picking method 21 is not one of the best any more for 20 Myr binning and 10 Myr stepping with both space and shape tested (Fig. 16). Further, in fact, in Fig. 16, we can see the picking method 21 is still one of the best for North America (101) and Australia (801), but just not for India (501). However, even for India (501), the difference values (ranging 0.0483–0.0535) produced by the picking method 21 are still closer to the left bound of the one-standard-deviation interval 0.0359–0.1072 and relatively farther from the mean 0.074, which means the picking method 21 is still a relatively better one.

3.2.4.2 Question: Why the picking method 16 is not among the worst for 10 Myr binning and 5 Myr stepping with only space tested? As shown in Fig. 11, Fig. 16 and Fig. 18, the picking method 16 is always one of the worst. However, the picking method 16 is not among the worst any more for 10 Myr binning and 5 Myr stepping with only space tested (Fig. 14). Further, in fact, in Fig. 14, we can see the picking method 16 is still one of the worst for North America (101) and India (501), but just not for Australia (801). However, even for Australia (801), the difference values (ranging 0.564–0.6144) produced by the picking method 16 are still closer to the right bound of the one-standard-deviation interval 0.3102–0.6548 and relatively farther from the mean 0.502, which means the picking method 16 is still a relatively worse one.

3.2.5 Question: Why the 10/5 bin/step methods sometimes unexceptionally produce better similarities than the 20/10 methods do?

3.2.6 Question: Do time window size and step affect the results?

A balance needs to be made between having windows that are too wide and steps that are too long which will smooth the data so much we miss actual details in the APWP (e.g. those 20 Myr window 10 Myr step paleomagnetic paths in Fig 17 and Fig 19) and windows that are too narrow and steps that are too short which introduces noise by having too few poles in each window (e.g. those 10 Myr window 5 Myr step paleomagnetic paths in Fig 12 and Fig 15). There is a dependence here on data density: higher density allows smaller windows/steps (this is one of the things we want to test with selective data removal mentioned in Chapter 4). Fitting curves by moving averaging change with different time window lengths and time increment lengths (i.e. steps) (e.g., the similarity of the pair in Fig 17f is improved a bit compared to Fig 12f). A variety of ways of binning the data (here both 20 Myr window 10 Myr step and 10 Myr window 5 Myr step) are being tested to see which one produces the better and more appropriately smoothed fit.

3.2.6.1 What to expect is the difference values for 20/10 window/step should be generally lower than those for 10/5 window/step, which further could result in more best methods and less worst methods.

3.2.6.2 The results are summarised in Table. 4.

3.2.6.3 Conclusions:

The APP methods (except with weighting no. 3) is recommended.

Table 4. Consistency check on comparisons of 20/10 window/step and 10/5 window/step. Notes: E: expected; UE: unexpected.

Comparisons		Consistency of Best			Consistency of Worst			If Difference Values for 20/10 Bin/Step Are Lower (Y/N)						
10/5	20/10	Y/N	Special Case(s)	Notes	Y/N	Special Case(s)	Notes	Mean	Median	Maximum	Minimum	All	If No, Unexpected Case (s)	Notes
						FHM								
Fig. 11a	Fig. 16a	Y	–	Same: Picking no. 1, 4, 5, 7, 11, 13, 15, 19, 21 and 25	N	Picking no. 2, 5, 7, 22 and 26 for 10/5 (E); 0, 10, 12, 20 and 24 for 20/10 (UE)	Same: Picking no. 8, 14, 16 and 18	N	Y	Y	N	N	(0,1,8,10,11,12,18,19,20,21, 24,25)(0-5) (3,4,14),(0-3,5) (5,7,13,15),(0-2,4,5) 9,(0,2,4,5) 23,(2,3) 27,(0-2); account for 29/42	
Fig. 14a	Fig. 18a	Y	5 more best: Picking no. 4, 5, 6, 7 and 9 only for 20/10 (E)	Same: Picking no. 1, 11, 13, 15, 19, 21 and 25	N (almost Y)	Picking no. 2 for 10/5 (E); 11 for 20/10 (UE)	Same: Picking no. 0, 3, 8, 10, 12, 14, 16, 17, 18, 20 and 24	N	Y	Y	N	N	0,(1,3) (1,21),(0-5) 3,(1,2,5) 4,5 5,(0,1,4) 7,(0,2,5) (8,11,13,25),(0-2,4,5) (9,27),(4,5) (10,12,20,24),1 15,(0,1,5) 17,4 19,(1,2,4,5) 22,(0-3,5) 26,(2,3,5); account for 17/42	
Fig. 11b	Fig. 16b	Y	2 more best: Picking no. 9 and 21 only for 10/5 (UE)	Same: Picking no. 4, 5, 6, 7 and 19	N (almost Y)	Picking no. 8 for 10/5 (E); 23 and 27 for 20/10 (UE)	Same: Picking no. 0, 2, 10, 12, 16, 18, 20 and 24	Y	Y	Y	Y	N	15,3 19,(0,1,3,5) 21,0 (22,26),(0-5) 23,(1,3,4) 27,(1,3); account for 23/168	
Fig. 14b	Fig. 18b	N (partially Y)	Picking no. 1, 17, 25 and 26 for 10/5 (UE); 5, 7, 11 and 13 for 20/10 (E)	Same: Picking no. 4, 6, 9, 19 and 21	N (almost Y)	Picking no. 14 and 18 for 10/5 (E); 23 and 27 for 20/10 (UE)	Same: Picking no. 0, 2, 8, 10, 12, 16 and 24	Y	Y	Y	Y	N	4,5 6,(3-5) 15,3 22,(3,5) 23,(0-3) 26,(2,3,5) 27,(0,2,3); account for 17/168	
Fig. 11c	Fig. 16c	Y	–	Same: Picking no. 1, 11, 13, 17, 19, 21 and 25	Y	1 more worst: Picking no. 6 only for 10/5 (E)	Same: Picking no. 2, 14, 16, 22 and 26	Y	Y	Y	Y	N	0,(1,2,5) (1,4,18,25),(0,1,3,5) 7,0 (8,10,20),(2,5) 9,1 11,3 (12,24),(1,2) 13,(0,1,3) (14,22,26),(0-2,4,5) (16,17),(0-5) 19,5 21,(0,1,5); account for 33/84	
Fig. 14c	Fig. 18c	N (almost Y)	Picking no. 7 for 10/5 (UE); 15 for 20/10 (E)	Same: Picking no. 1, 11, 13, 17, 19, 21 and 25	N	Picking no. 6 for 10/5 (E); 4, 5 and 16 for 20/10 (UE)	Same: Picking no. 2, 14, 22 and 26	Y	Y	Y	Y	N	(1,25),(3,5) 5,2 13,3 17,5 19,5 21,5; account for 9/168	
						MHM								
Fig. 20a	Fig. 24a	N	Picking no. 1, 13, 21 and 25 for 10/5 (UE); 22 and 26 for 20/10 (E)	Same: Picking no. 4, 5, 7, 11, and 15	Y	2 more worst: Picking no. 17, 22 only for 10/5 (E)	Same: Picking no. 0, 8, 10, 12, 14, 16, 18, 20 and 24	N	Y	N	N	N	(0,4,8,10,11,12,14,15,18, 20,24,25),(0-5) (1,19,21),(0,1,5) 3,5 (5,7),(0-2,4,5) 6,1 (9,13),(0,1,3,5) 23,(1,3) 27,(1,3,5); account for 53/84	
Fig. 22a	Fig. 26a	N (almost Y)	Picking no. 23 and 27 for 10/5 (UE); 4 for 20/10 (E)	Same: Picking no. 1, 5, 7, 11, 13, 15, 19, 21 and 25	N (almost Y)	Picking no. 17 for 10/5 (E); 20 for 20/10 (UE)	Same: Picking no. 0, 8, 10, 12, 14, 16, 18 and 24	Y	Y	Y	Y	N	(0,3,21),5 (2,8,14),1 9,(1,3,5) 10,(2,4) 11,(0,1,5) 15,(2-5) 20,2 23,(1,3) 25,4 27,(0,1,3,5); account for 13/84	
Fig. 20b	Fig. 24b	Y	2 more best: Picking no. 22 and 26 only for 10/5 (UE)	Same: Picking no. 4, 5, 6, 7 and 19	N	Picking no. 8, 20 for 10/5 (E); 23 and 27 for 20/10 (UE)	Same: Picking no. 0, 2, 10, 12, 16, 18 and 24	Y	Y	Y	Y	N	1,5 (4,15),3 (19,23,27),(0-5) 21,(0,1,35) 22,(1,4) 25,(3,5) 26,4; account for 5/28	
Fig. 22b	Fig. 26b	N (partially Y)	Picking no. 1, 4 and 21 for 10/5 (UE); 9, 23 and 27 for 20/10 (E)	Same: Picking no. 6, 15, 17, 19, 22 and 26	N (almost Y)	Picking no. 14 for 10/5 (E); 3 and 18 for 20/10 (UE)	Same: Picking no. 0, 2, 8, 10, 12, 16 and 24	Y	Y	Y	Y	N	4,5 15,(1-3) 19,2 21,(4,5); account for 7/168	
Fig. 20c	Fig. 24c	Y	3 more best: Picking no. 15, 23 and 27 only for 20/10 (E)	Same: Picking no. 1, 11, 13, 17, 19, 21 and 25	Y	–	Same: Picking no. 2, 14, 16, 22 and 26	Y	Y	Y	N	N	(0,24),(1,2,5) (1,11,13,18,19,21,25),(0,1,3,5) 4,(0,3,5) (5,7),3 (8,17),(0-3,5) 12,(1-3,5) 20,(1,2,4,5) 10,(1,2) (14,22,26),(0-2,4,5) 16,(0-5); account for 10/21	
Fig. 22c	Fig. 26c	N (almost Y)	Picking no. 17 for 10/5 (UE); 5, 7, 15, 23 for 20/10 (E)	Same: Picking no. 1, 11, 13, 19, 21 and 25	N	Picking no. 10, 18 for 10/5 (E); 20 for 20/10 (UE)	Same: Picking no. 2, 6, 8, 14, 22 and 26	Y	Y	Y	N	N	(1,11,19,21,25),(0,1,3,5) 13,(0,1) 17,(0-3,5); account for 9/56	

The different values for 20/10 window/step are indeed generally lower than those for 10/5 window/step, excluding the unexpected case of North America with both space and shape tested (Fig. 11a and Fig. 16a).

4 FINAL CONCLUSIONS

Picking method no. 16 (AMP with data from old studies) is not recommended, e.g. before 1983) for generating a paleomagnetic APWP.

ACKNOWLEDGMENTS

All images are produced using GMT (Wessel et al. 2013). Thanks to Ohio Supercomputer Center for their remote HPC resources.

REFERENCES

Besse, J. & Courtillot, V., 2002, Apparent and true polar wander and the geometry of the geomagnetic field over the last 200 Myr, *J. Geophys. Res.*, **107**, 2300.

Fisher, R. A., 1953. Dispersion on a sphere, *Proc. Roy. Soc. London Ser. A.*, **217**, 295–305.

Jacox, E. H. & Samet, H., 2007. Spatial Join Techniques, *ACM Trans. Database Syst.*. **32**, 7.

King, R. F., 1955. The remanent magnetism of artificially deposited sediments, *Geophys. Suppl. Mon. Not. Roy. Astron. Soc. Lett.*, **7**, 115–134.

Müller, R. D., Royer, J. Y. & Lawver, L. A., 1993. Revised plate motions relative to the hotspots from combined Atlantic and Indian-Ocean hotspot tracks, *Geology*, **21**, 275–278.

Müller, R. D., Royer, J. Y., Cande, S. C., Roest, W. R. & Maschenkov, S., 1999. New constraints on the Late Cretaceous/Tertiary plate tectonic evolution of the Caribbean, *Sedimentary Basins of the World*. **4**, 33–59.

McQuarrie, N. & Wernicke, B. P., 2006. An animated tectonic reconstruction of southwestern North America since 36 Ma, *Geosphere*, **1**, 147–172.

O'Neill, C., Müller, R. D. & Steinberger, B., 2005. On the uncertainties in hot spot reconstructions and the significance of moving hot spot reference frames, *Geochem. Geophys. Geosyst.*, **6**, Q04003.

Pisarevsky, S. A., 2005. New edition of the Global Paleomagnetic Database, *Eos. Trans. AGU*, **86**, 170.

Schettino, A. & Scotese, C. R., 2005. Apparent polar wander paths for the major continents (200 Ma to the present day): a palaeomagnetic reference frame for global plate tectonic reconstructions, *Geophys. J Int.*, **163**, 727–759.

Torsvik, T. H., Smethurst, M. A., van der Voo, R., Trench, A., Abrahamsen, N. & Halvorsen, E., 1992. Baltica. A synopsis of Vendian-Permian palaeomagnetic data and their palaeotectonic implications, *Earth Sci. Rev.*, **33**, 133–152.

Torsvik, T. H. & Smethurst, M. A., 1999. Plate tectonic modelling: virtual reality with GMAP, *Comput. Geosci.*, **25**, 395–402.

Tauxe, L. & Kent, D. V., 2004. A simplified statistical model for the geomagnetic field and the detection of shallow bias in paleomagnetic inclinations: Was the ancient magnetic field dipolar? *Geophys. Monogr. AGU*, **145**, 101–115.

Torsvik, T. H., Müller, R. D., van der Voo, R., Steinberger, B., & Gaina, C., 2008. Global plate motion frames: Toward a unified model, *Rev. Geophys.*, **46**, RG3004.

Torsvik, T. H., van der Voo, R., Preeden, U., Mac Niocaill, C., Steinberger, B., Doubrovine, P. V., van Hinsbergen, D. J. J., Domeier, M., Gaina, C., Tohver, E., Meert, J. G., McCausland, P. J. A. & Cocks, L. R. M., 2012. Phanerozoic polar wander, palaeogeography and dynamics, *Earth Sci. Rev.*, **114**, 325–368.

Tauxe L., Banerjee S.K., Butler R.F. & van der Voo R., 2018. *Essentials of Paleomagnetism*, 5th web edn, Available on line

van der Voo, R., 1990. The reliability of paleomagnetic data, *Tectonophysics*, **184**, 1–9.

Wessel, P., Smith, W. H. F., Scharroo, R., Luis, J. & Wobbe, F.,2013. Generic Mapping Tools: Improved version released, *Eos. Trans. AGU*, **94**, 409–410.

Young, A., Flament, N., Maloney, K., Williams, S., Matthews, K., Zahirovic, S. & Mller, D., 2018. Global kinematics of tectonic plates and subduction zones since the late Paleozoic Era, *Geosci. Front.*, **in press**, 000–000.