

# Methods for Producing a Reliable APWP

Chenjian Fu<sup>1\*</sup> Chris Rowan<sup>2</sup>

<sup>1</sup> Department of Geology, Kent State University, 325 S Lincoln St, Kent, OH 44242, USA

<sup>2</sup> 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 methods is(are) good or bad. The final results tell us that the “Age Position Picking (APP)” method is better than the “Age Mean Picking (AMP)” method for making a reliable paleomagnetic APWP and weighting is actually unnecessary.

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

## 1 INTRODUCTION

APWPs are generated by combining paleomagnetic poles, also known as paleopoles, for a particular rigid block over the desired age range to produce a smoothed path. See the Appendix A for some examples how the paleopole datasets are constrained for a particular tectonic plate during a specific time interval.

### 1.1 Not All Data Are Created Equal

However, uncertainties in the age and location of paleopoles 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 paleopoles 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

Paleopoles 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. 2019). 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 or correct (if possible) such poles from our APWP calculation.

##### Other Factors

\* Email: cfu3@kent.edu

## 2 Chenjian Fu

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 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 (McElhinny & Lock 1996; Pisarevsky 2005, updated in 2016 by the Ivar Giaever Geomagnetic Laboratory team, in collaboration with Pisarevsky) 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 paleopoles 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  $\kappa$  greater than 10 and  $\alpha_{95}$  less than  $16^\circ$ ; (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 quantified reliability score. Also it 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

**Table 1.** List of the used fields and field codes of the GPMDB.

No.	Weighting Algorithm
LOMAGAGE	Lower best estimate of the magnetic age of the magnetisation component
HIMAGAGE	Upper best estimate of the magnetic age of the magnetisation component
B	Number of sites
N	Number of samples
ED95	Radius of circle of 95% confidence about mean direction, i.e. $\alpha_{95}$
KD	Fisher precision parameter for mean direction
DP	Half-angle of confidence on the pole in the direction of paleomeridian
DM	Half-angle of confidence on the pole perpendicular to paleomeridian
K_NORM	Fisher precision parameter for Normal directions

poles with at least 6 sampling sites and 36 samples, each site having  $\alpha_{95}$  less than  $10^\circ$  in the Cenozoic and  $15^\circ$  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

### 2.1 General Approach

In this study, we use paleopoles extracted from the GPMDB to generate APWPs for the period 120–0 Ma. A range of possible APW paths for North America, India and Australia can be generated from the extracted sets of paleopoles using various binning, filtering and weighting methods (Tables 2 and 3). These paths can then be compared to synthetic APWPs independently generated from an absolute plate motion model. The three plates chosen have different attributes, both in terms of the input data set and the nature of the reference APWP.

### 2.2 Paleomagnetic Data

#### 2.2.1 Used Specific Field Codes of GPMDB

Data analysis includes a tremendous amount of manipulation of data fields/columns in the GPMDB. In the following content, several specific fields and field codes will be referred to. They are listed below for easy reference.

#### 2.2.2 Paleomagnetic Data of Three Representative Continents

Collections of paleopoles with a minimum age (LOMAGAGE)  $\leq 135$  Ma for the North American (Plate ID 101), the Indian (ID 501) and Australian (ID 801) plates, were extracted from the GPMDB. In order to include valid paleopoles from blocks that moved independently prior to 120 Ma, which therefore have different assigned plate codes in the GPMDB, the spatial join technique (Jacox & Samet 2007) was used to find all poles within the geographic region that defines the rigid plate within the period of interest (see also Appendix A):

for North America, the search region was defined by the North American (ID 101), Avalon/Acadia (ID 108) and Piedmont (ID 109) blocks, as defined by the recently published plate model of (Young et al. 2018). Following extraction, 58 poles from southwestern North America that have been affected by regional rotations since 36 Ma (McQuarrie & Wernicke 2006), were removed. The final dataset consists of 135 paleopoles (Fig. 1), with 76 (about 56.3%; average age uncertainty 14.1974 Myr) sampled from dominantly igneous sequences; 56 (about 41.5%; average age uncertainty 27.5179 Myr) sampled from mostly sedimentary sequences, including 6 from redbeds); and 3 (about 2.2%) from metamorphic sequences. The principal features of the age distribution are a larger number of young (<5 Ma) poles, and relatively fewer poles in the Late Cretaceous and Miocene.

for India, the Indian block (ID 501) as defined by (Young et al. 2018) was used, but following extraction 31 paleopoles associated with parts of the northern margin that have undergone regional rotations since the Jurassic (Gaina et al. 2015) were removed. The final dataset consists of 75 paleopoles (Fig. 2), with 39 (52%) sampled from dominantly igneous sequences and 36 (48%) sampled from mostly sedimentary sequences, including 3 from redbeds). There is a high concentration of poles from the latest Cretaceous–Early Cenozoic (c. 70–60 Ma), many of which are igneous; in younger and older intervals, there are fewer, mostly sedimentary poles.

for Australia, the Australia (ID 801), Sumba (ID 675), and Timor (ID 684) blocks as defined by (Young et al. 2018) were used, in combination with data from the Tasmania block (ID 805) younger than c. 100 Ma (with a maximum age (HIMAGAGE)  $\leq 100$  Ma), prior to which it was not fixed with respect to Australia (Young et al. 2018). The final dataset consists of 99 paleopoles (Fig. 3), with 61 (61.6%) sampled from dominantly igneous sequences and 38 (38.4%) sampled from mostly sedimentary sequences, including 9 from redbeds). The temporal distribution of poles is relatively uniform.

Compared with North American (Fig. 1) and Australian (Fig. 3) paleopoles, Indian paleopoles are in relatively lower density in general except during the period of about 70–60 Ma (Fig. 2).

### 2.3 APWP Generation

Multiple APWPs were generated using the selected poles for each of the three plates as follows:

**Picking/binning.** A moving average or moving window technique was used: paleopoles were selected for each APWP time step (initially 5 Myr step length from 0 to 120 Ma) if their age fell within a window centered on the current step age. In this study, the width of the moving window was always twice that of time step (i.e. initially 10 Myr), such that each window half-overlaps with its neighbours.

**Filtering.** Poles with characteristics thought to correspond to poor data quality, or lacking characteristics thought to correspond to good data quality, were discarded (or in some cases, corrected).

**Weighting.** Calculation of a weighted Fisher mean (Fisher 1953) of the remaining poles within each window, using weighting functions intended to increase the influence of higher quality poles relative to lower quality ones.

28 different picking and filtering algorithms were tested (Table 2), in combination with 6 different weighting algorithms (Table 3), for the three plates. The effects of changing the time step length and width of the moving window, and the reference path, were also examined.

#### 2.3.1 Picking/Binning

In studies where the moving window method is used to calculate an APWP (Torsvik & Smethurst 1999; Torsvik et al. 2008), a paleopole is generally considered to fall in the current window only if the mid-point of its age limits fall within that window. If paleopole has a large age uncertainty compared to the size of the moving window, it will not be included in the moving windows close to the beginning and end of the age range, which are arguably more likely magnetisation ages than the mid-point. To investigate this issue, we compare the performance of moving windows populated using the mid-point picking criterion, referred to hereafter as “Age Mean Picking” (AMP; even-numbered algorithms in Table 2, Fig. 9 and subsequent figures), to a less restrictive picking criterion where a paleopole is included in the current moving window if any part of its age limits falls within that window, referred to hereafter as “Age Position Picking” (APP; odd-numbered algorithms in Table 2, Fig. 9 and subsequent figures). The APP algorithm will pick more paleopoles in each moving window than the AMP algorithm (Figs. 1–3; Fig. 4).

If, for example, we have a paleopole 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 is included just in the 14–16 Ma window (for the mid-point age of 15 Ma) for AMP. For APP, this paleopole falls in the 9–11, 10–12, 11–13, 12–14 … 17–19, 18–20 and 19–21 Ma windows. So the average poles are produced from each window, and each original paleopole is represented over its entire possible acquisition age.

A shorter step and narrower window will potentially increase the clustering of the selected paleopoles, but will reduce their number. Conversely, a longer step and wider window will increase the number of poles, but potentially decrease their clustering. To investigate these trade-offs, every picking/filtering and weighting method was also used to generate APWPs with a time step and window doubled to 10 Myr and 20 Myr, respectively. Paths generated using AMP and APP with no filtering, and every weighting method, with time steps from 1 Myr to 15 Myr in 1 Myr increments and window widths from 2 Myr to 30 Myr in 2 Myr increments, were also analysed. In all cases the oldest time step was 120 Ma.

#### 2.3.2 Filtering

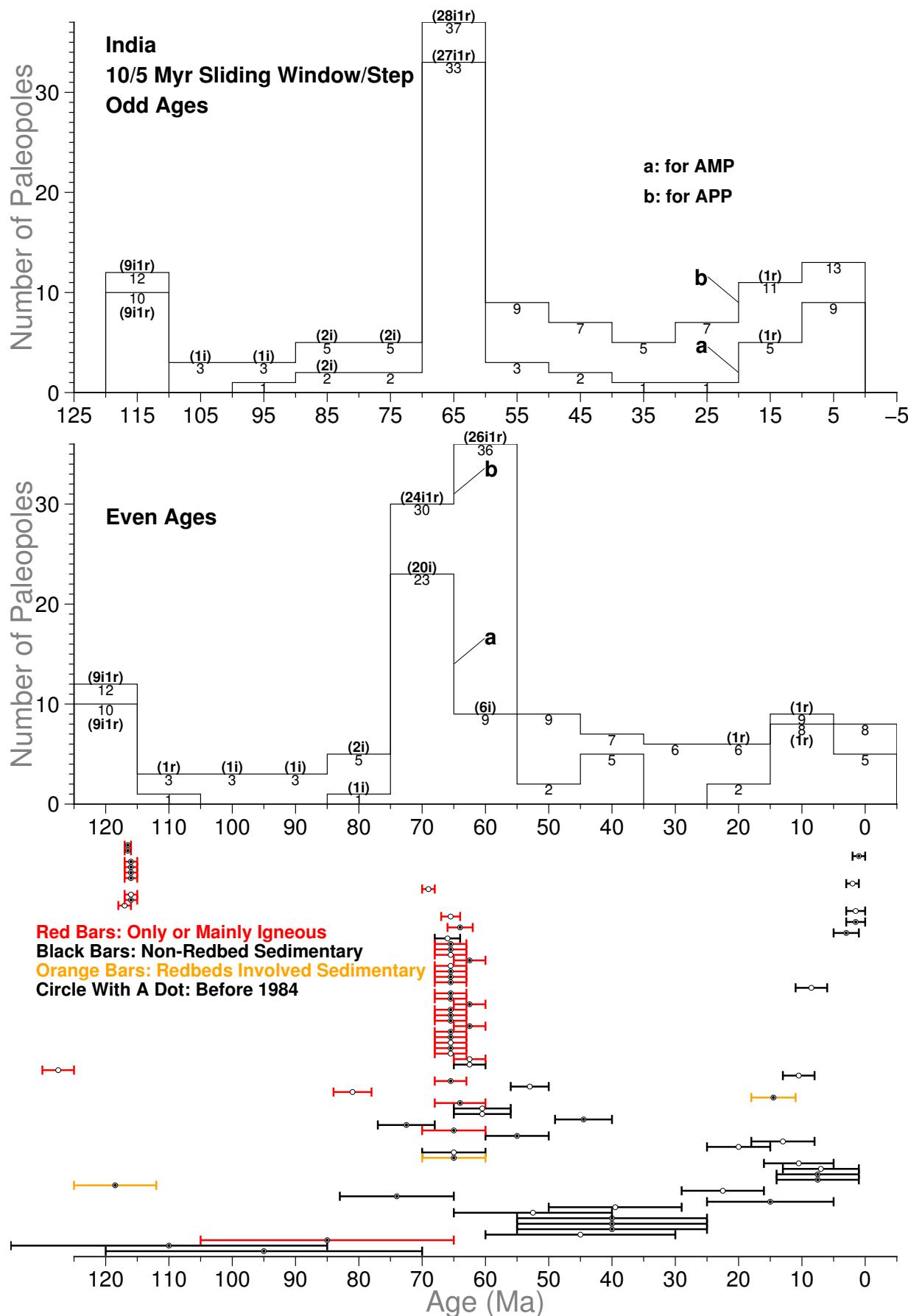
14 different filters or corrections (Table 2) were applied to both data picked using the AMP moving window method (even numbers) and data picked using the APP moving window method (odd numbers), resulting in a total of 28 unique picking algorithms. The filters or corrections can be characterised as follows:

*No modification (method 0/1).*

*Removal of poles with large spatial and temporal uncertainties (method 2/3).* Paleopoles with both large  $\alpha_{95}$  ( $ED95 > 15^\circ$ , following the BC02 threshold for the Mesozoic) and wide acquisition age limits (difference between HIMAGAGE and LOMAGAGE  $> 20$  Myr, following the V90 criteria about age within a



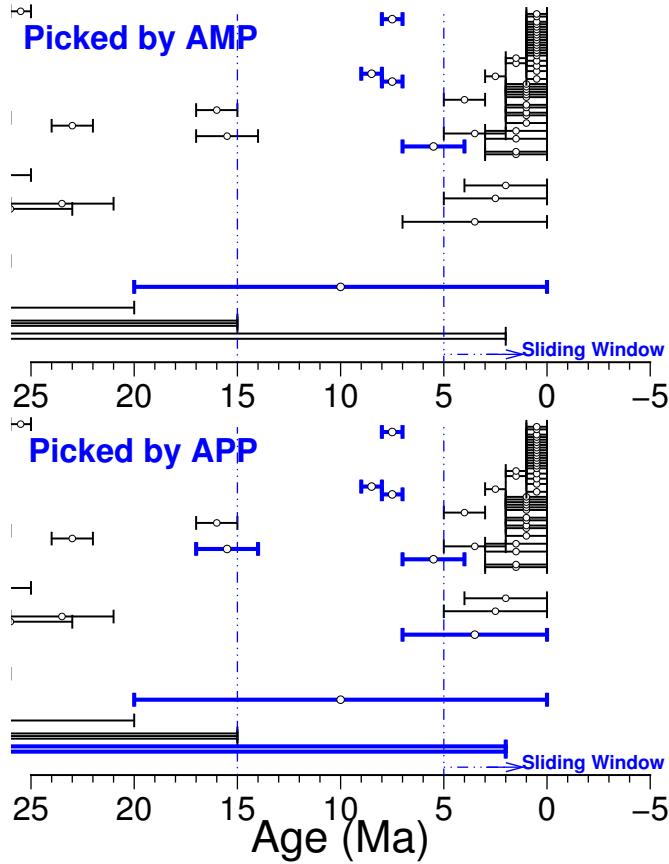
**Figure 1.** Temporal distribution of 120–0 Ma NAC (101) paleopoles in 10 Myr window length and 5 Myr step length. For distribution a, each bin only counts in the midpoints (circles) of pole error bars (not including those right at bin edges); For distribution b, as long as the bar intersects with the bin (not including those intersecting only at one of bin edges), it is counted in. Inside the parentheses, i means igneous rocks derived (red bars; only two poles, 83–77 Ma and 80–65 Ma, from igneous and also sedimentary; only one pole, 72–40 Ma, from igneous and also metamorphic), r means sedimentary rocks with redbeds involved derived (orange bars), and m means metamorphic rocks derived (blue bars); the left are non-redbed sedimentary rocks derived (black bars; only two poles, 146–65 Ma [RESULTNO 6679] and 2–0 Ma [RESULTNO 1227], are from sedimentary and also metamorphic). The data published before 1984 are shown as circles with a dot.



**Figure 2.** Temporal distribution of 120–0 Ma Indian (501) paleopoles. For red bars, only one pole, 67–64 Ma (RESULTNO 8593), is from igneous and also sedimentary. See Fig. 1 for more information.



**Figure 3.** Temporal distribution of 120–0 Ma Australian (801) paleopoles. For black bars, only four poles, 100–80 Ma (RESULTNO 1106), 10–2 Ma (RESULTNO 1208), 4–2 Ma (RESULTNO 140) and 1–0 Ma (RESULTNO 1963), are from sedimentary and also igneous. For red bars, only one pole, 65–25 Ma (RESULTNO 1872), is from igneous and also sedimentary rocks, and only one pole, 1–0 Ma (RESULTNO 1147), is from igneous and also metamorphic rocks. See Fig. 1 for more information.



**Figure 4.** An example of 10 Myr moving window and 5 Myr step in the two moving average methods, AMP and APP, based on poles of the NAC. White circles are the midpoints of low and high magnetic ages. The vertical axis has no specific meaning here. For example, for the window of 15 Ma to 5 Ma (the dashed-line bin), the AMP method calculates the Fisher mean pole (dark triangle in Fig. 6) of only 5 paleopoles, while the APP method calculates the mean pole (dark circle in Fig. 6) of 9 paleopoles.

half of a geological period; the average of the geological periods between 120 and 0 Ma [Quaternary, Neogene, Paleogene and Cretaceous] is about 20 Myr which are less likely to provide a good estimate of the actual pole position within any specific age window, were excluded.

*Prefer poles from igneous rocks (methods 4/5, 6/7).* Method 4/5 removes paleopoles potentially affected by inclination flattening by selecting only paleopoles coded as igneous or mostly igneous (ROCKTYPE starting with “intrusive” or “extrusive”). In fact, most of the paleopoles picked by method 4/5 are derived from igneous-only rocks. Method 6/7 selects paleopoles coded as containing igneous (ROCKTYPE containing “intrusive” or “extrusive”); this is a less strict filter, because the dominant rock type could potentially be another lithology. So method 6 includes also poles from method 4, and method 7 includes poles from method 5.

*Correct sedimentary poles for inclination shallowing (method 8/9).* Rather than excluding paleopoles from sedimentary rocks, paleopoles coded as sedimentary or redbeds were instead corrected for inclination flattening using the flattening function  $\tan I_o = f \tan I_f$  (King 1955), where  $I_o$  is the observed inclination,  $I_f$  is the unflattened inclination, and  $f$  is the flattening factor (also known as shallowing coefficient; 1=no flattening, 0=completely

**Table 2.** List of all Picking/Binning algorithms developed here.

No.	Picking Algorithm
0	AMP: Age Mean Picking, see Section “APWP Generation”
1	APP: Age Position Picking
2	AMP (“ $\alpha_{95}/\text{Age range}$ ” no more than “15/20”)
3	APP (“ $\alpha_{95}/\text{Age range}$ ” 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)

flattened). Here  $f = 0.6$  is used in all cases, following (Torsvik et al. 2012), unless when the rock type (ROCKTYPE field in the database) is not sedimentary dominated but contains sedimentary,  $f = 0.8$  is used instead, following the minimum anisotropy-of-thermal-remanence determined f-correction (Domeier et al. 2011a; Domeier et al. 2011b).

*Remove redbeds (method 10/11) or correct them for inclination shallowing (method 12/13).* Bias toward shallow inclinations is also observed in paleomagnetic data derived from redbeds (Tauxe & Kent 2004; Krijgsman & Tauxe 2004; Tan et al. 2007; Bilardello & Kodama 2010, e.g., in central Asia, Mediterranean region, North America, etc.). This bias can be addressed by removing the source (method 10/11; ROCKTYPE containing redbeds), or correcting for inclination flattening, setting  $f = 0.6$  as previously described (method 12/13). In the latter case, the assumption is being made that the redbeds are carrying a detrital paleomagnetic signal.

*Prefer poles from younger (methods 14/15, 24/25) or older (method 16/17) studies.* Advancements in equipment (e.g., cryogenic magnetometers) and analytical techniques (e.g., stepwise demagnetisation) mean that more recently published paleopoles are potentially more reliable than older ones. Method 14/15 removes any paleopoles published prior to 1983 (YEAR > 1983)—the mean publication date for paleopoles in the GPMDB. Method 24/25 takes a similar but less aggressive approach by excluding paleopoles that have been superseded (99 datasets) by later studies from the same sequence, which are presumed to represent a more accurate determination of the paleopole position. Conversely, removing pale-

opoles published after 1983 (method 16/17; YEAR  $\leq$  1983) should have a negative effect.

*Exclude suspected local rotations and secondary overprints (method 18/19), or correct for them where possible (method 20/21).*

Secondary remanence components and local tectonic deformation can both displace the measured pole position away from its “true” position. Such poles can be identified based on demagnetisation data, or comparison to the pre-existing APWP. Method 18/19 removes paleopoles that were identified as such in the COMMENTS field (all the paleopoles affected by local rotations are picked out by carefully going through all the datasets, including two groups with [19 datasets] and without [47 datasets] suggested corrections; the paleopoles affected by secondary overprints are extracted with the keyword “econd” contained in the COMMENTS). A subset (19 datasets) of these paleopoles have a suggested correction associated with them; method 20/21 retains these paleopoles after applying the suggested correction.

*SS05 quality criteria (methods 22/23).* As with method 2/3, SS05 (Schettino & Scotese 2005) removes paleopoles with high spatial ( $\alpha95 > 15^\circ$ ) and temporal (age range  $> 40$  Myr) uncertainty, but additionally remove paleopoles where samples had poor sampling coverage (sampling sites’ quantity [B] of  $< 4$ , samples’ quantity [N] of  $< 4$  times of the sites [B]) and were not subjected to even a blanket demagnetisation treatment (laboratory cleaning procedure code DEMAGCODE  $< 2$ ). Method 26/27 also uses these criteria, but further excludes superseded data.

### 2.3.3 Weighting

Following filtering, weights were assigned to each of the remaining paleopoles using one of the following six algorithms (Table 3), prior to calculation of a weighted Fisher mean:

*No weighting (Weighting 0).* Weighting factor=1 for all paleopoles.

*Weighting by sample and site number (Weighting 1).* Paleopoles derived from more individually oriented samples (observations; N) collected from more sampling levels/sites (B) are more likely to average out secular variation and accurately sample the GAD field (Tauxe et al. 2019; van der Voo 1990; Besse & Courtillot 2002), and are given a weighting closer to 1. Unfortunately, in the GPMDB, not all paleopoles’ B or N are given. There are datasets with only number of sampling sites (B; at least greater than 1) given, but no number of samples (N) or only one sample given, so for this case, if  $B > 1$  and  $N \leq 1$ , weight= $(1 - \frac{1}{B}) * 0.5$  (for 120–0 Ma North America, there are 8 such datasets, India 4, Australia 1). If only N (at least greater than 1) is given, and B is missing or only one, i.e.  $B \leq 1$  and  $N > 1$ , weight= $(1 - \frac{1}{N}) * 0.5$  (for 120–0 Ma North America, there are 20 such datasets, India 26, Australia 22). 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; for 120–0 Ma North America, India and Australia, there is no such dataset), weight=0.2 (No. 1 in Table 3).

*Weighting by age uncertainty (Weighting 2)* Above a maximum age range that represents a well-constrained age, defined as half of each geological period in the Phanerozoic Eon (e.g., Quaternary, Neogene; here the geological period that the middle point

**Table 3.** List of all weighting algorithms developed here.

No.	Weighting Algorithm
0	None (No weighting)
1	Larger numbers of sites (B) & observations (N), greater weight ( $w$ ): $w = \begin{cases} 0.2, & \text{if both B \& N are missing, or } B \leq 1 \text{ \& } N \leq 1; \\ (1 - \frac{1}{B}) * 0.5, & \text{if only N is missing, or } N \leq 1 \text{ \& } B > 1; \\ (1 - \frac{1}{N}) * 0.5, & \text{if only B is missing, or } B \leq 1 \text{ \& } N > 1; \\ (1 - \frac{1}{B}) * (1 - \frac{1}{N}), & \text{if } B > 1 \text{ \& } N > 1. \end{cases}$
2	Lower age uncertainty, greater weight: $\text{age\_range}=\text{HIMAGAGE-LOMAGAGE}$ $\text{age\_midpoint}=(\text{HIMAGAGE}+\text{LOMAGAGE})*0.5$ if $\text{age\_midpoint} < 2.58$ (Ma; start of the Quaternary, according to GSA Geologic Time Scale), $w = \begin{cases} 1, & \text{if } \text{age\_range} \leq 1.29 \text{ (from } \frac{2.58-0}{2}) \\ \frac{1.29}{\text{age\_range}}, & \text{if } \text{age\_range} > 1.29; \end{cases}$ if $2.58 \leq \text{age\_midpoint} < 23.03$ (Ma; Neogene), $w = \begin{cases} 1, & \text{if } \text{age\_range} \leq 10.225 \text{ (from } \frac{23.03-2.58}{2}) \\ \frac{10.225}{\text{age\_range}}, & \text{if } \text{age\_range} > 10.225; \end{cases}$ if $23.03 \leq \text{age\_midpoint} < 201.3$ (Ma; Paleogene,Cretaceous,Jurassic), $w = \begin{cases} 1, & \text{if } \text{age\_range} \leq 15 \\ \frac{15}{\text{age\_range}}, & \text{if } \text{age\_range} > 15. \end{cases}$
3	Lower $\alpha95$ , greater weight: Positive half Normal distribution with a mean and standard deviation of 0 and 10, scaled with $10\sqrt{2\pi}$ (to make the peak reach 1) $w = e^{-\frac{\alpha_{95}^2}{200}},$ where $\alpha_{95} = \begin{cases} ED95 \\ DP, \text{ if ED95 is missing} \\ \frac{140}{\sqrt{KD*N}}, \text{ if ED95 \& DP are missing} \\ \frac{140}{\sqrt{K\_NORM*N}}, \text{ if ED95, DP \& KD are missing} \\ \frac{140}{\sqrt{K\_NORM*B}}, \text{ if ED95, DP, KD \& N are missing} \\ \frac{140}{\sqrt{1.7*B}}, \text{ if ED95, DP, KD \& K\_NORM are missing,} \\ \text{using the lowest KD in GPMDB, about 1.7} \end{cases}$ finally $w=0$ if this $\alpha95$ completely overlaps with another smaller $\alpha95$ whose paleopole is exactly derived from the same place and same rock.
4	Age error Position to bin (more overlap, greater weight): $wha, window high age; wla, window low age$ $w = \begin{cases} \frac{wha-LOMAGAGE}{age\_range}, & \text{if LOMAGAGE}>wla \text{ \& HIMAGAGE}>wha \\ \frac{HIMAGAGE-wla}{age\_range}, & \text{if LOMAGAGE}<wla \text{ \& HIMAGAGE}<wha \\ \frac{wha-wla}{age\_range}, & \text{if LOMAGAGE}<wla \text{ \& HIMAGAGE}>wha \\ 1, & \text{if LOMAGAGE}>wla \text{ \& HIMAGAGE}<wha. \end{cases}$
5	Combining 3 and 4: average of the two weights from 3 and 4

of the paleopole’s age range falls within is assigned) (van der Voo 1990; Tauxe et al. 2019) or 15 Myr (the halves of the Paleogene, Cretaceous, and Jurassic periods are all at least 20 Myr, which is large for these relatively young geological periods), whichever is smaller, paleopoles are given an increasingly small weight as the age uncertainty (the high magnetic age – the low magnetic age) increases (No. 2 in Table 3).

*Weighting by spatial uncertainty (Weighting 3).* Paleopoles with a smaller  $\alpha95$  confidence ellipse are given a higher weighting than those with a larger  $\alpha95$ , using a Gaussian/Normal distribution centered on 0 with standard deviation of 10. However, not all pale-

opoles'  $\alpha_{95}$  are given in the database. If  $\alpha_{95}$  is not given, DP (the semi-axis of the confidence ellipse along the great circle path from site to pole) is assigned as  $\alpha_{95}$ . If DP is also not given,  $\alpha_{95}$  was further approximated by  $\frac{140}{KD \times N}$ , where KD is Fisher precision parameter for mean direction if this parameter is given, or Fisher precision parameter for Normal directions (K\_NORM) if only K\_NORM is given when KD is missing. If N is not given, B is used as N. If even K\_NORM is also missing, the lowest KD value 1.7 in GPMDB 4.6b is used as KD. It is also worthwhile to mention that if samples, where two paleopoles are derived, are exactly from the same place and same rock, and one  $\alpha_{95}$  is completely inside the other  $\alpha_{95}$ , a zero is assigned as the weight of the data with the larger  $\alpha_{95}$ . In fact, in the above described procedure A95 (circle of 95% confidence about mean pole) is a better alternative instead of  $\alpha_{95}$ , because A95 is directly reflecting the spatial uncertainty of the paleopoles. However, most paleopoles' A95s are not given in GPMDB 4.6b, so  $\alpha_{95}$  is used instead since  $\alpha_{95}$  is also indirectly reflecting the quality of the dataset.

*Weighting by degree of overlap between moving window and pole age uncertainty (Weighting 4).* If a large fraction of the age range for an individual paleopole falls within the current window, it is given a higher weighting than a pole where the overlap is smaller, because it is more likely to be close to the true pole position in the window interval. In other words, if window intersects with part of age range, weight = (intersecting part) / (age range width).

*Weighting by both spatial and temporal uncertainty (Weighting 5).* This weighting method is a combination of No 3 (but here the standard deviation is 15 though) and No 4. It takes the average of sums of the weights generated by weighting methods 3 and 4.

### 2.3.4 Final Scaling

The weight values obtained from the six different weight functions (Table 3) introduced above are then integrated into Fisher mean function (Fisher 1953) to calculate a weighted Fisher mean. In fact, the weights are integrated into Cartesian x, y and z components of each individual moment directions (here each individual paleopole location). Then these individual moment directions are combined through Fisher resultant vector  $R$  function (See Chapter 11 of (Taupe et al. 2019)). So the mean pole location, its spatial uncertainty  $\alpha_{95}$ , and precision parameter  $k$  are all weighted along with  $R$ .

Traditionally, weights are directly integrated by being multiplied with the variable we would like to do weighting to. For example, here weights can be directly multiplied with the Cartesian x, y and z components of each paleopole. However, this direct multiplying causes the decreasing of  $R$ , which further sensitively and extremely increases  $\alpha_{95}$  (Fig. 5), especially because  $R$  is always less than  $N$  and  $N$  is usually not that high (more than about 50 is rather rare, around 10 averagely). The consequence would be that all the  $\alpha_{95}$ s of mean poles are extremely large in size and difficult to be spatially differentiated.

Therefore, here weights are scaled before being multiplied with the Cartesian x, y and z components by

$$\text{ScaledWeight} = \frac{\text{Weight}_i * N}{\sum_i^N \text{Weight}_i},$$

where  $N$  is number of paleopoles for making a mean pole. So the scaled weight could be greater than 1 because it is actually scaled through being divided by the mean of the weights. This scaling does

not only keep the effect of weighting but also avoid dramatically changing  $R$  and indirectly and extremely changing  $\alpha_{95}$ .

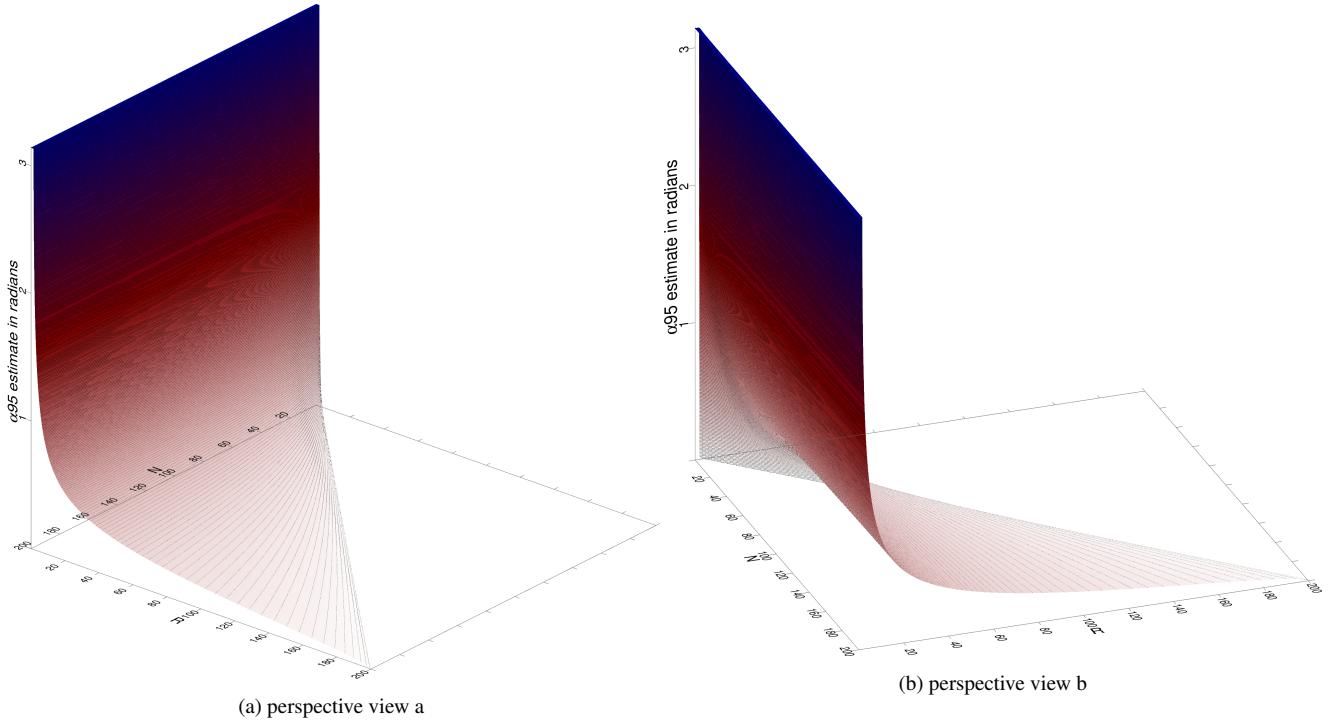
Some of the picking (Table 2) and weighting (Table 3) methods developed here are also connected with the V90 Q factors mentioned above. For example, Pt 2, 3 and Wt 2, 4, 5 are related to the V90 criteria 1; Pt 2, 3, 22, 23, 26, 27 and Wt 1, 3 are related to the V90 criteria 2; Pt 22 and 23 are related to the V90 criteria 3; The data constraining described in Appendix A is related to the V90 criteria 5; and Pt 18 and 19 are related to the V90 criteria 7.

### 2.4 Reference Paths

A prediction of the expected APWP for any plate can be generated using a plate kinematic model (e.g. the last c. 180–200 Myr plate motions reconstructed from spreading ridges in ocean basins) that is tied in to an absolute reference frame. Here, we use the rotations of (O'Neill et al. 2005), which describe motion of Nubia (plate ID 701) relative to the Indo-Atlantic hotspots. Such an absolute frame of reference based on hot spots has been extended back to 120 Ma (O'Neill et al. 2005). North America is linked to this reference frame across the Mid-Atlantic ridge, using North America-Nubia rotations from (DeMets et al. 2010) to C-Sequence chron C1no (0.78 Ma), from (Shephard et al. 2012) to chron C2An.2n (2.7 Ma), from (Müller et al. 1999) to chron C5n.1ny (9.74 Ma), from (Gaina et al. 2013) to chron C5n.2o (10.949 Ma), from (Müller et al. 1999) to chron C6ny (19.05 Ma), from (Gaina et al. 2013) to chron C6no (20.131 Ma), from (Müller et al. 1999) to chron C34ny (83.5 Ma), from C34ny to about 118.1 Ma (Shephard et al. 2012), and to closure at C34no (120.6 Ma) (Gaina et al. 2013). India is linked via the East African Rift Valley (Somalia-Nubia rotations to chron C1no [0.78 Ma] from (DeMets et al. 2017), to chron C2A.2no [3.22 Ma] from (Horner-Johnson et al. 2005), and to closure at C7.2m [25.01 Ma] from (Rowan & Rowley 2016) and C34 [85 Ma; Rowley, pers. comm.], and finally extended to 120 Ma because there was no known relative motion between Somalia and Nubia from 120 Ma to 85 Ma according to the rotations from (Müller et al. 2016)); and Australia via the East African Rift Valley, SW Indian Ridge (E Antarctica-Somalia rotations to chron C1no [0.78 Ma] from (DeMets et al. 2017), to chron C2A.2no [3.22 Ma] from (Horner-Johnson et al. 2005), to chron C5n.2no [10.95 Ma] from (Lemaux et al. 2002), to chron C13ny [33.06 Ma] from (Patriat et al. 2008), to chron C29no [64.75 Ma] from (Cande et al. 2010), to chron C34y [83 Ma; Rowley, pers. comm.], to 96 Ma from (Marks & Tikku 2001), and to closure at M0 [120.6 Ma] from (Müller et al. 2008)), and SE Indian Ridge (Australia-East Antarctica rotations to chron C1no [0.78 Ma] from (DeMets et al. 2017), to chron C6no [20.13 Ma] from (Cande & Stock 2004), to chron C8o [26 Ma] from (Granot & Dymant 2018), to chron C17n.3no [38.11 Ma] from (Cande & Stock 2004), to C34ny [83.5 Ma] from (Whittaker et al. 2013), to the Quiet Zone Boundary [96 Ma] from (Whittaker et al. 2007), to closure at 136 Ma from (Whittaker et al. 2013)). Above all, these kinematic models (see the collected rotation data for the above-mentioned plate circuits in the supplementary material) tied in to a hot spot absolute reference frame can guarantee the oldest predicted poles for the three continents back to 120 Ma.

To reconstruct a reference APWP at the required time steps for comparison with the paleomagnetic APWPs, rotations and their associated uncertainties were interpolated between constraining finite rotation poles according to the method of (Doubrovine & Tarduno 2008), assuming constant rates.

Neither the hotspot reference frame nor the paleomagnetic



**Figure 5.** Visualization of Equation 11.9 of Essentials of Paleomagnetism: Fifth Web Edition, about the estimating of the circle of 95% confidence ( $p=0.05$ ) about the mean,  $\alpha_{95}$ , from resultant vector  $R$  and number of directions (or paleopoles)  $N$ .

reference frame are truly fixed with respect to the solid Earth. In the former case, hotspots are not truly stationary in the mantle (Steinberger & O'Connell 1998); in the latter, true polar wander (TPW) may also lead to differential movements of the solid earth with respect to the spin axis (Evans 2003). In reality, it is difficult to untangle these effects. Whilst there is little clear evidence for significant TPW in the past about 120 Myr (Cottrell & Tarduno 2000; Riisager et al. 2004), modeling suggests that the effects of hotspot drift can start to become significant over 80–100 Myr timescales (O'Neill et al. 2005). Because paleomagnetic APWPs have large associated spatial errors, a synthetic APWP calculated using a fixed hotspot reference frame is unlikely to deviate significantly from the ‘true’ APWP, and most comparison experiments use a fixed hotspot model (FHM) reference path for North America (Fig. 6), India (Fig. 7) and Australia (Fig. 8). However, the full set of comparisons for the 28 picking methods and 6 weighting methods was also run for reference paths generated using the moving hotspot model (MHM) rotations of (O'Neill et al. 2005), which incorporate motions of the Indo-Atlantic hotspots relative to the mantle derived from mantle convection modeling.

When comparing the synthetic APW paths for the three plates (inset, Fig. 8), there are clear differences. The predicted mean north pole for North America at 120 Myr is still at about 75°N (Fig. 6), indicating rather slow drift with respect to the spin axis; this is due to a large component of the North American plate’s absolute motion in the past 120 Myr being to the east. In contrast, the rapid northward motion of the Indian plate in the same period, particularly prior to its collision with Asia at about 50–55 Ma (Najman et al. 2010), is reflected by the 120 Ma predicted mean north pole being located at about 20°N (Fig. 7). Australia represents an intermediate case, with north westerly plate motion from about 120–60 Myr changing to more rapid northward motion from about 60–55 Ma to the present (Whittaker et al. 2007). When comparing the FHM and

MHM tracks, differences in the oldest parts (before about 80 Ma) are apparent for India and North America.

These differences in the reference path due to different plate kinematics is another variable that may affect the performance of the different weighting algorithm for different plates, in addition to the distribution and type of the contributing mean poles used to generate the paleomagnetic APWPs.

## 2.5 Comparison Algorithm

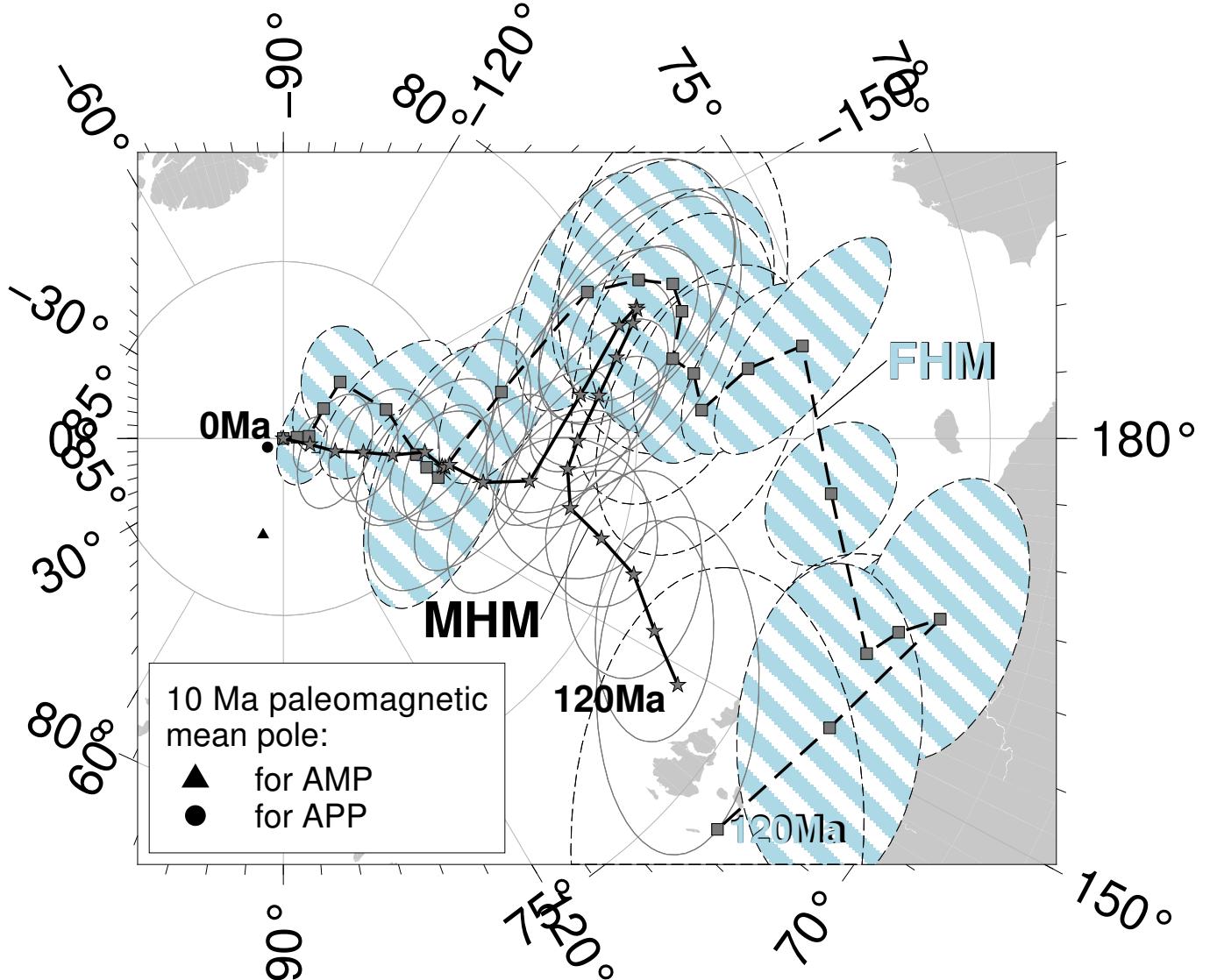
Comparisons between APWPs generated using different picking and weighting algorithms and the synthetic reference APWPs were performed using the composite path difference ( $\mathcal{CPD}$ ) algorithm described in Chapter 2, with equal weighting given to the spatial, length and angular differences (i.e.,  $W_s = W_l = W_a = \frac{1}{3}$ ).

This does not only help find the most similar paleomagnetic APWP (from the best algorithm) to the reference APWP, but also help further test and demonstrate the validity of the similarity measuring tool in practise.

## 3 RESULTS

### 3.1 Baseline results: 10 Myr window, 5 Myr step, fixed hotspot reference

Fig. 9 shows the  $\mathcal{CPD}$  scores for the APWPs generated with all 28 picking methods (AMP and APP with one of 14 separate filters applied) and one of 6 weighted mean calculations then applied, compared to the FHM reference path for North America (Fig. 9a), India (Fig. 9b) and Australia (Fig. 9c). The 27 lowest and highest of the 168 scores for each plate (values greater than 1 standard deviation from the mean) are marked in green and red, respectively. Different combinations of windowing method, filtering and weighting clearly



**Figure 6.** MHM predicted 120–0 Ma APWP (solid line) for *NAC* through the North America–Nubia–Mantle plate circuit. The FHM predicted path (dashed line with shaded uncertainties) is also shown for comparison. The age step is 5 Myr. Compared with the 10 Ma paleomagnetic mean pole calculated by the AMP method (dark triangle), the coeval mean pole derived from the APP method is closer to both FHM and MHM predicted 10 Ma poles, which indicates more data diluting the effect of outliers. See also the paleopoles that the two mean poles are composed of in Fig. 4.

affect the difference score, with  $\text{CPD}$  values ranging from 0.0023 to 0.5137. The fits for paths with low difference scores are clearly much better than for those with high ones (Fig. 10). From Fig. 9, it is clear that:

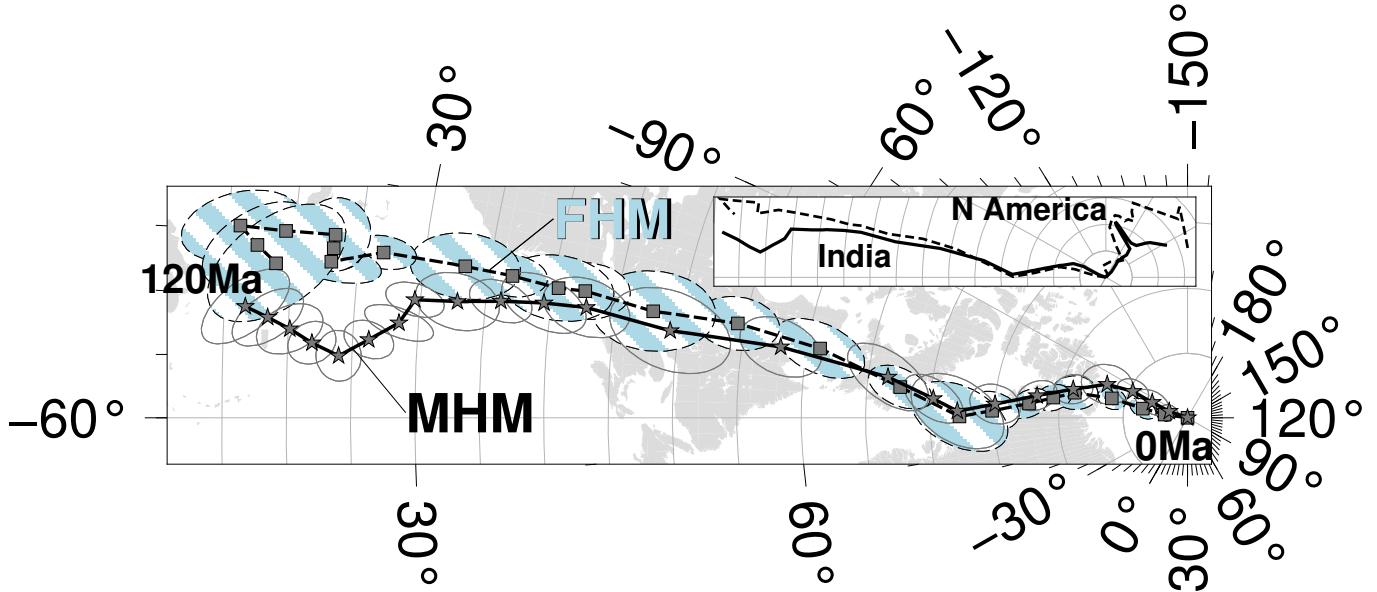
- (i) There is much more variation in scores along the horizontal axes than the vertical axes (see self-explanatory topography of bands in Fig. 9), suggesting that the choice of windowing and filtering method (Table 2) has a much greater impact than weighting (Table 3).
- (ii) Many of the highest scores (worst fits) are associated with even-numbered picking and filtering methods, i.e., those which use the AMP windowing algorithm. Even so, methods 4 and 6 are among the best methods for India.
- (iii) The magnitude and range of  $\text{CPD}$  scores for each of the three plates is different, with the North American plate having the lowest magnitudes and range (Fig. 9a), and the Indian plate having

the highest (Fig. 9b). In general, the scores of North America and Australia are relatively closer (Fig. 13).

(iv) Although there is some overlap (e.g., picking methods 19, 21 [best], and 2, 16 [worst] for all the three plates or for both India and Australia; 1, 11, 13, 19, 21, 25 [best], and 2, 14, 16, 22, 26 [worst] for both North America and Australia; 5, 7, 19, 21 [best], 2, 16, 18 [worst] for both North America and India), the best- and worst-performing picking/filtering and weighting algorithms are not exactly the same for each plate.

### 3.2 Effects of windowing method

Dividing  $\text{CPD}$  scores according to whether the AMP or APP windowing method was used (Fig. 11) confirms that whilst the lowest  $\text{CPD}$  scores for paths generated by the AMP windowing algorithm are close to the lowest scores generated using the APP method, the highest scores are much higher (Fig. 12). The mean



**Figure 7.** MHM predicted 120–0 Ma APWP (solid line) 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 for comparison. The inset shows paths for fast moving India and also much slower moving North America shown in Fig. 6.

of the  $\text{CPD}$  scores for APWPs generated using AMP is greater than the maximum APP-derived score for the Indian and Australian plates (Figs. 11b, 11c, 12c, 12c), and more than 1 standard deviation greater than the APP mean for North American APWPs (Figs. 11a, 12a).

For each of the 84 possible combinations of filter method and weighting, the AMP-derived score is typically 3–5 times higher than the equivalent APP-derived score (Fig. 12, insets). APP-generated paths yield a lower  $\text{CPD}$  score than the equivalent AMP-generated path for 82 (97.6%) of the North America scores, 72 (85.7%) of the India scores, and 84 (100%) of the Australia scores. For the North American and Indian plates, filtering that prefers igneous poles (picking methods 4/5 and 6/7) or removes poles with large various uncertainties (methods 22/23 and 26/27) is most likely to produce AMP scores close to (less than 1.5 times) or less than the APP scores (Figs. 11a, 11b). In the former case, the scores are comparable and relatively low; in the latter case, they are comparable but relatively high. For the Australian plate, only correcting for sedimentary inclination shallowing (method 8/9) produces comparable but moderate scores (Fig. 11c).

On the North American plate, the APP-derived  $\text{CPD}$  score is most likely to be significantly better (defined as more than 5 times higher) when weighting methods 0 (no weighting) and 1 (by sample and site number) have been applied (Fig. 11a). This pattern is also seen for the Australian plate, but removal of poles published after 1983 (picking method 16/17) also results in significantly better performance of the APP method (Fig. 11c). For the Indian plate, the largest difference occurs when poles with sedimentary inclination shallowing (method 8/9) are corrected, or suspected overprints or local rotations are removed (picking method 18/19, Fig. 11b).

### 3.3 Effects of filtering and weighting

When  $\text{CPD}$  scores are separated by windowing method (Fig. 11), the effects of particular filtering and weighting methods become easier to discern. In general, different filters (rows) produce larger

variations in scores than different weighting methods (columns). With the exception of AMP-derived paths for India, the  $\text{CPD}$  score for paths with no filtering (picking method 0/1) and no weighting (weighting method 0) is lower than the mean scores for that plate and windowing method. Therefore filtering and weighting at best slightly improves, and at worst significantly degrades, the APWP fit to the reference path.

#### 3.3.1 Filter aggression

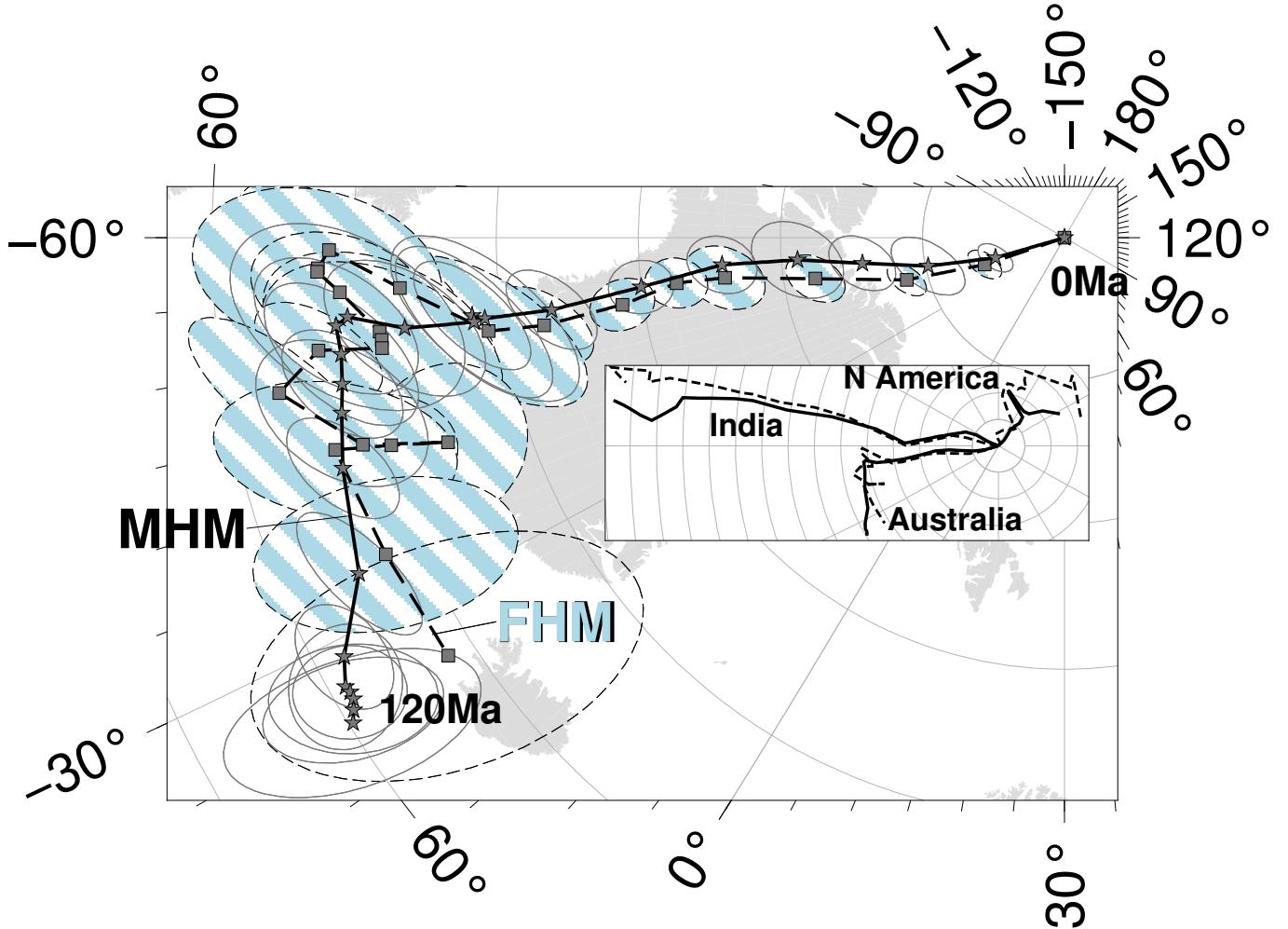
When considering the effects of filtering, it is important to consider how many poles within the data set have been removed (related to picking no 2–7, 10, 11, and 14–27) or corrected (8, 9, 12 and 13): if there is very little alteration of the data set, little change from no filtering (methods 0/1) would be expected. In terms of the numbers of poles affected, the most consequential filters are:

- (i) removal of sedimentary poles (methods 4/5 and 6/7), which removes about 40–50% of the dataset on all 3 plates, with the highest proportion being removed on the Indian plate. Although method 4/5 is more strict, it does not remove many more poles than method 6/7. Especially for North America and India, the numbers of the filtered paleopoles by method 4/5 and 6/7 are actually the same.

- (ii) correction of sedimentary poles for inclination flattening (filter 8/9), which affects 38–48% of the dataset, with the highest affected proportion on the Indian plate.

- (iii) removal of poles with large temporal and spatial uncertainty (methods 2/3, 22/23), particularly for the Australian plate, where the SS05 filtering criteria removes about 70% of the poles. Filter 26/27 combines methods 22/23 and 24/25, but no or very few (in the only case of Australia only 1 additional pole) are actually removed.

- (iv) filtering based on publication date (methods 14/15 and 16/17), with the ratio of pre/post 1983 poles varying from about 50/50 on the North American plate to about 70/30 on the Australian plate.



**Figure 8.** MHM predicted 120–0 Ma APWP (solid line) 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 for comparison. The inset shows paths for fast moving India shown in Fig. 7, much slower moving North America shown in Fig. 6, and also relatively intermediate moving Australia.

Conversely, filtering or correction for redbeds (methods 10/11 and 12/13), local rotations and overprints (methods 18/19 and 20/21 [one paleopole influenced by local rotation removed, and one corrected, for only India; about 2.7 per cent, Fig. 9b]), or superseded data (method 24/25) affected 4% or less of the poles on any plate.

### 3.3.2 Filter performance

Focussing on the filtering and weighting methods with aggressive filtering, some commonalities in the best- and worst-performing methods can be observed, although there are usually exceptions for particular plates and/or windowing methods:

(i) For all 3 plates, higher  $\text{CPD}$  scores are commonly associated with filtering based on  $\alpha_{95}$  and age range (picking methods 2/3, 22/23, 26/27), with the exception of AMP-derived paths for India, where picking methods 22 and 26 produce some of the lowest scores.

(ii) For North America and India, low scores are commonly associated with removal of non-igneous poles (methods 4/5 and 6/7), particularly for AMP-derived paths. On the Australian plate, these filters are less effective.

(iii) For North America and Australia, correction for inclina-

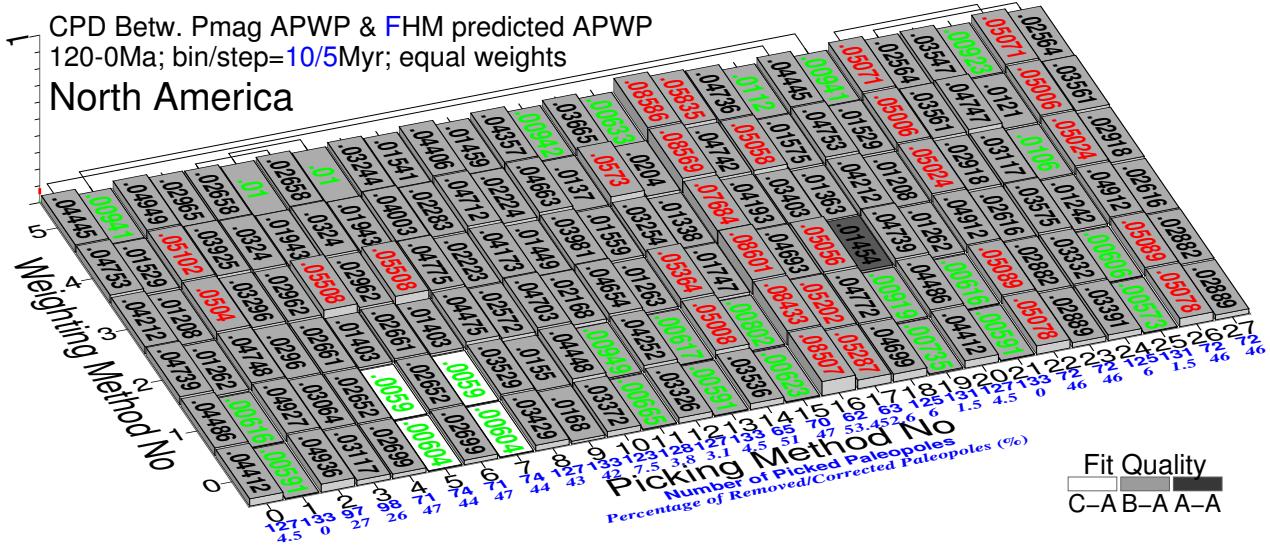
tion flattening generates  $\text{CPD}$  scores very similar to scores with no filtering for AMP-derived paths (method 8 vs. method 0), and increases scores for APP-derived paths (method 9 vs. method 1). In contrast, for India generally there is a small decrease in  $\text{CPD}$  scores compared to no filtering for both AMP- and APP-derived paths.

(iv) Many of the highest difference scores for North America and India occur when paleopoles published after 1983 are removed (method 16/17), whilst removing paleopoles published before 1983 (method 14/15) generates  $\text{CPD}$  scores comparable to scores with no filtering (relatively much lower). In contrast, for Australia methods 14/15 produces some of the highest  $\text{CPD}$  scores, and methods 16/17 have little effect. The number of older studies (65/68; Fig. 9c) is almost 2.5 times of the number of newer studies (27/29) though.

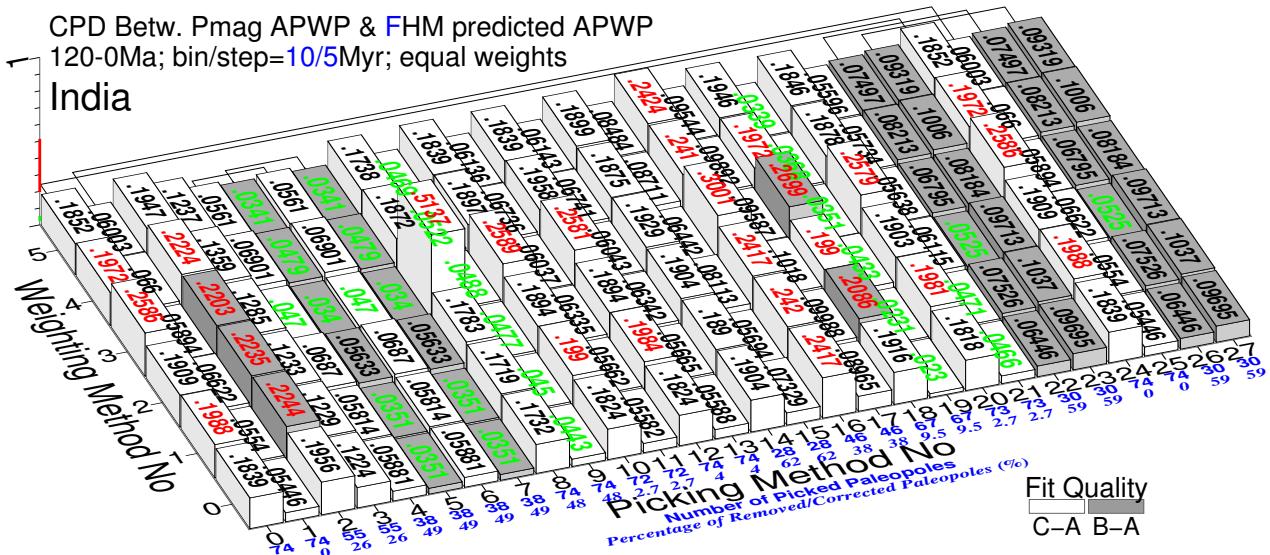
Whilst it is generally true that methods with low-aggression filters do not generate scores that differ much from the no-filtering scores, there are some exceptions:

(i) For North America and Australia, removing superseded paleopoles (picking method 24/25) produces lower  $\text{CPD}$  scores.

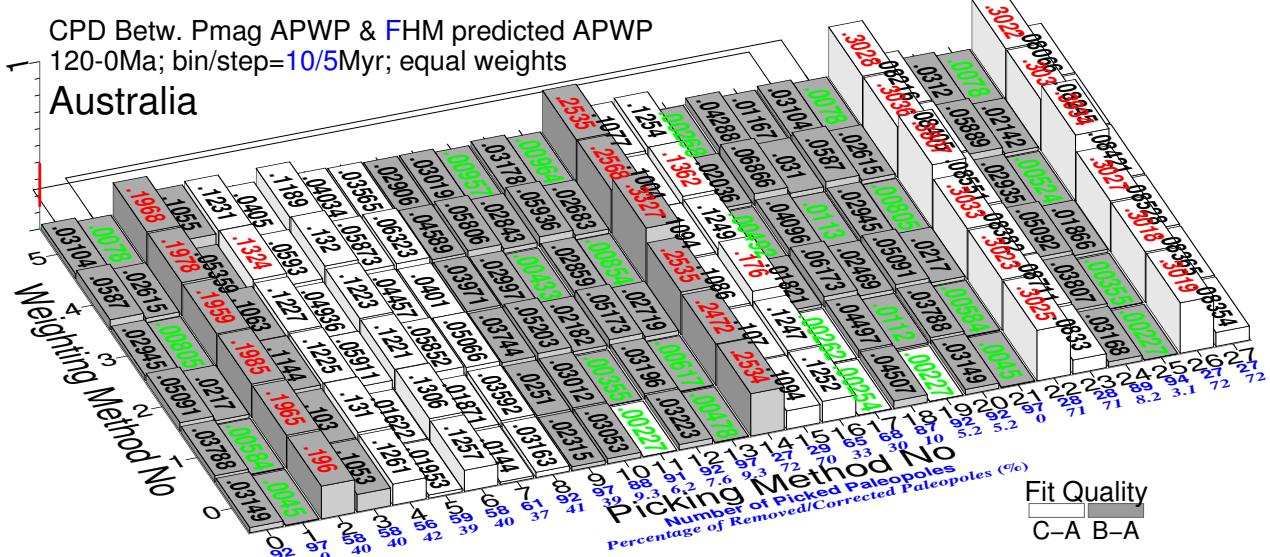
(ii) Removing paleopoles suspected to be affected by overprints or local rotations (picking method 18/19) consistently produces lower  $\text{CPD}$  scores for Indian APP-derived paths.



(a) Plate ID 101 with children: minimum 0.00573 (25(0)), maximum 0.08601 (16(2)), mean 0.032403, median 0.032395

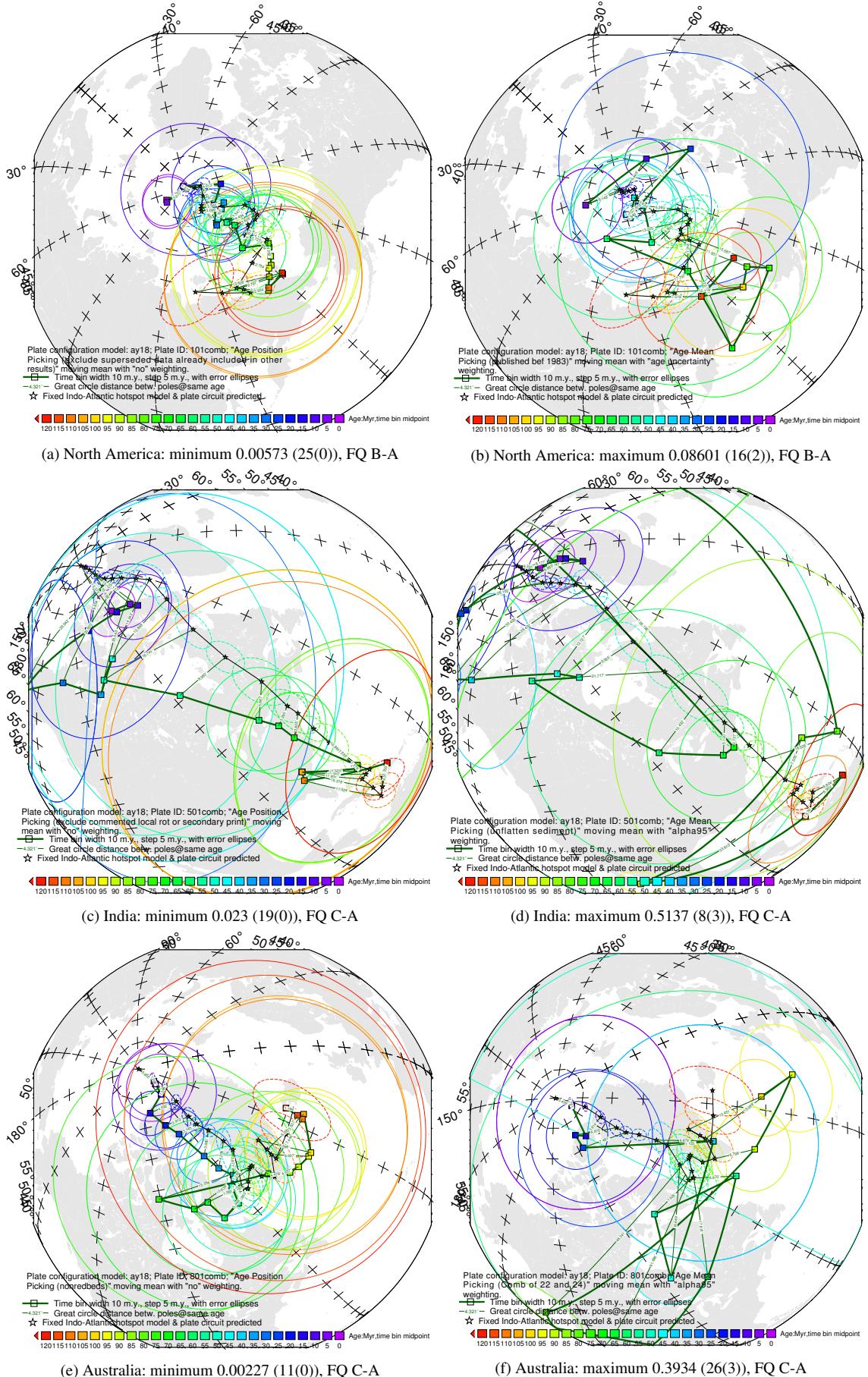


(b) Plate ID 501: minimum 0.023 (19(0)), maximum 0.5137 (8(3)), mean 0.1182, median 0.0835



(c) Plate ID 801 with children: minimum 0.00227 (11(0)), maximum 0.3934 (26(3)), mean 0.08373, median 0.05

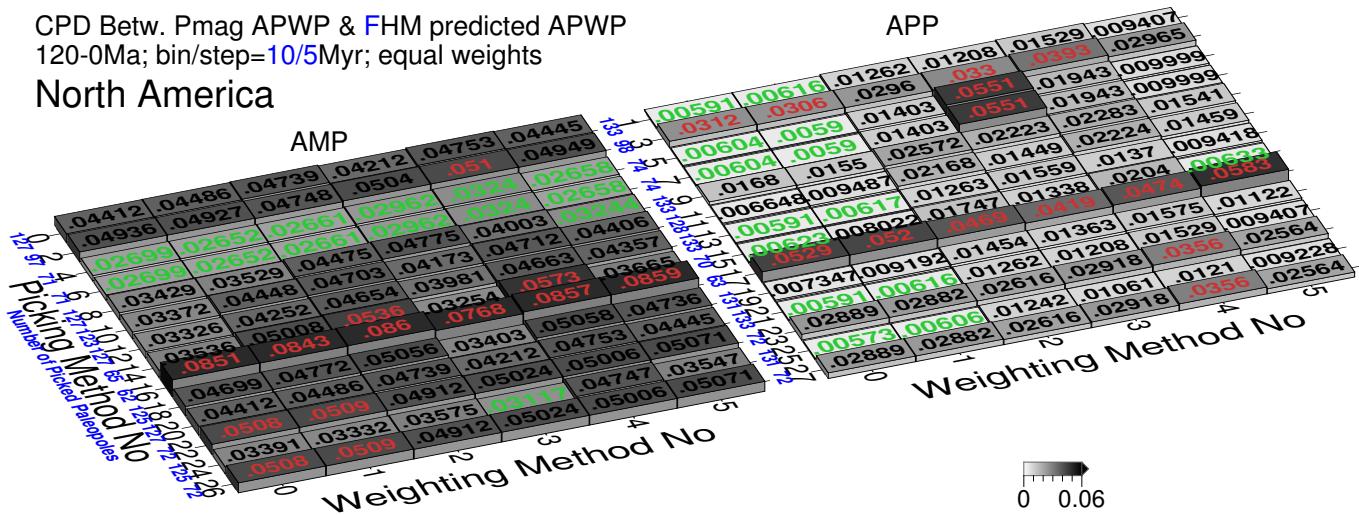
**Figure 9.** Equal-weight composite path difference ( $\mathcal{CPD}$ ) values between each continent's paleomagnetic APWPs and its predicted APWP from FHM and related plate circuits. The paths are in 10 Myr bin and 5 Myr step. The difference values less than one-standard-deviation interval of the whole 168 values (lower 15.866 per cent) are colored in green, more than one-standard-deviation interval (upper 15.866 per cent) colored in red. Exactly the same columns are connected. The percentages of removed paleopoles are derived relative to picking 1, corrected relative to each corresponding picking method (8, 9, 12, 13; 1 removed and 1 corrected by 20, 21 for India). Fit quality (FQ) for each score is color coded.



**Figure 10.** Path comparisons with best and worst difference values shown in Fig. 9. The parenthetical remarks are Picking No and Weighting No.

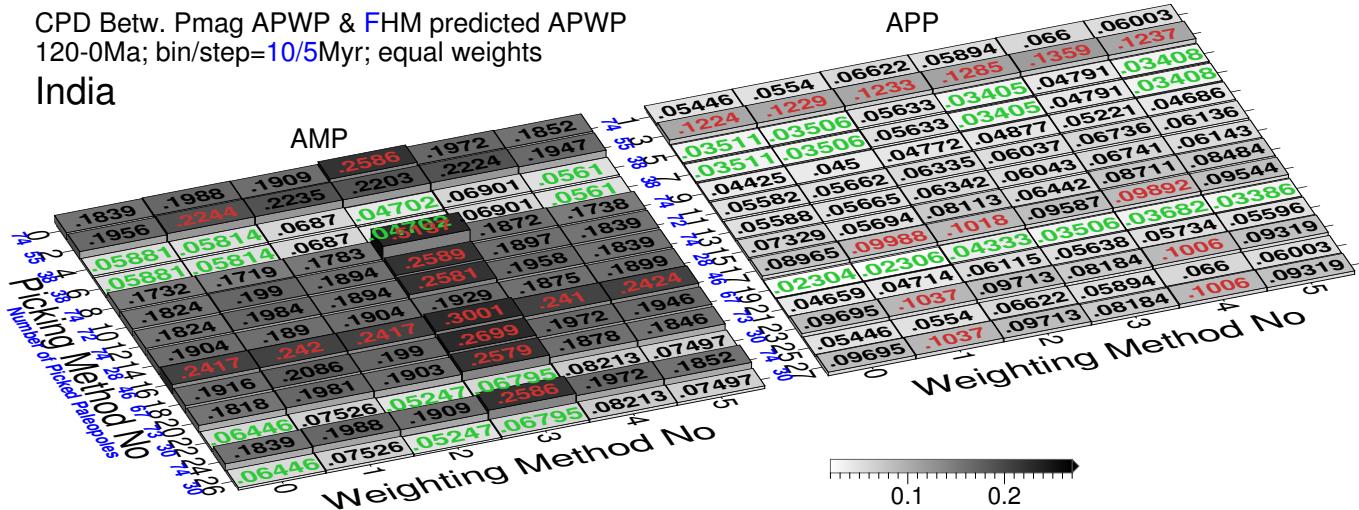
CPD Betw. Pmag APWP & FHM predicted APWP  
120-0Ma; bin/step=10/5Myr; equal weights

### North America



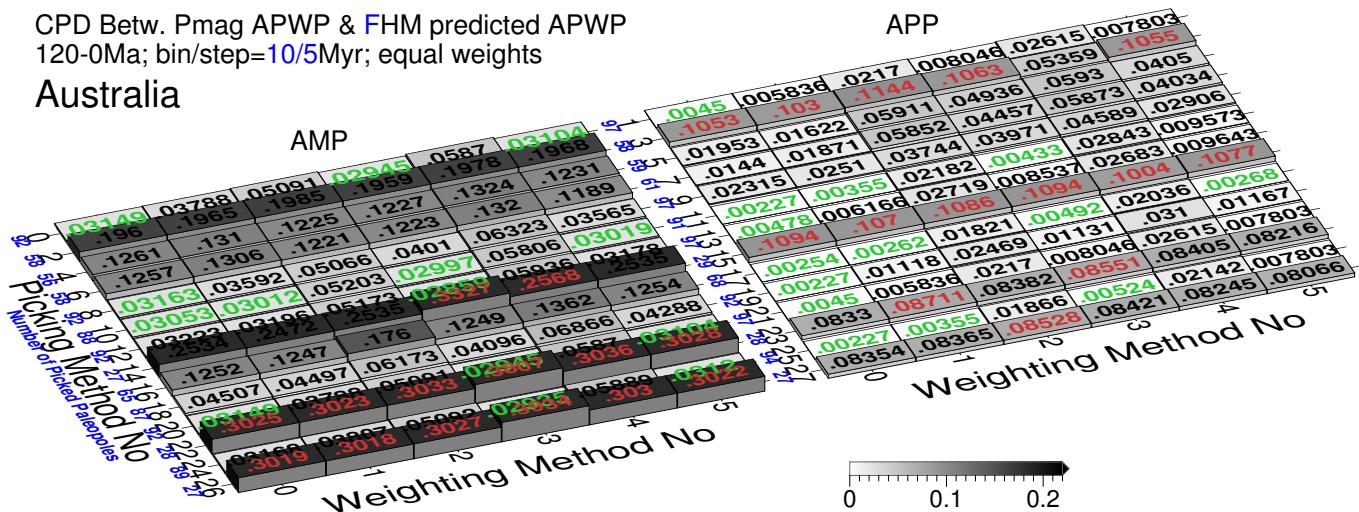
CPD Betw. Pmag APWP & FHM predicted APWP  
120-0Ma; bin/step=10/5Myr; equal weights

### India

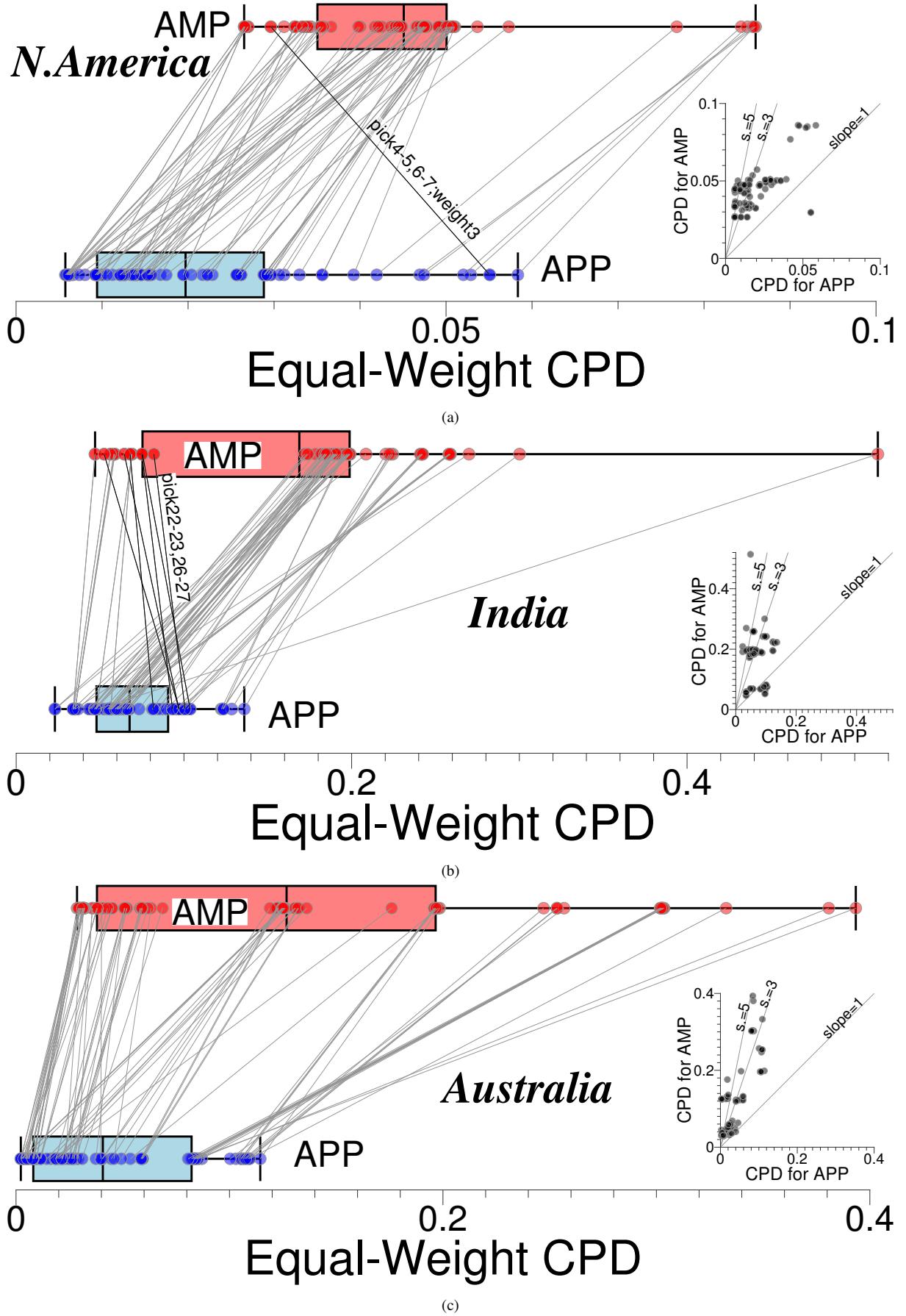


CPD Betw. Pmag APWP & FHM predicted APWP  
120-0Ma; bin/step=10/5Myr; equal weights

### Australia



**Figure 11.** Separated results from AMP and APP in Fig. 9. For each grid block (left: AMP; right: APP), the difference values less than one-standard-deviation interval of the whole 84 values are labeled in green, more than one-standard-deviation interval labeled in red.



**Figure 12.** Box-and-whisker and cross (inset) plots of Fig. 11. The  $\mathcal{CPD}$ s from same filter and weighting method (red and blue dots plotted with box-and-whisker) are connected; some special cases where  $\mathcal{CPD}$  from AMP lower than from APP are highlighted using darker connecting lines. Dot symbols are semi-transparent so a darker color indicates a greater number of data at a given  $\mathcal{CPD}$ .

### 3.3.3 Weighting performance

Compared to the variations resulting from different windowing methods and filters, Figs. 11, 14, 15 and 16 indicate that the effect of weighting the data prior to calculating a Fisher mean is generally small. Where an effect can be seen, it is negative, generating larger CPD scores.

(i) For APP-derived paths from the North American and Australian plates, weighting method 0 (no weighting) and 1 (weighting by sample and site number) appear to usually produce slightly better CPD scores than other weighting methods.

(ii) Weighting method 3 (weighting by spatial uncertainty) seems most likely to generate much higher CPD scores, particularly for AMP-derived paths, and particular for the Indian plate.

(iii) Weighting methods 0, 1 and 5 are generally producing better similarity than 2, 3 and 4.

### 3.4 Effects of window size

Fig. 17 shows the CPD scores for the APWPs generated with all 28 picking and 6 weighting methods, compared to the FHM reference paths, with the picking time window width increased from 10 to 20 Myr, and the window step increased from 5 to 10 Myr.

#### 3.4.1 Overall change

Along with the increasing sizes of moving window and step, mean, median and range values of CPD scores shrink, except for North America the mean value of CPD scores slightly increase (Fig. 9 and Fig. 17). For North America, most scores actually increase (i.e. 10-5 is better; Fig. 19a) except for methods 2, 4, 6, 16/17, 22, 26 wherein 20-10 is better ('quality filtering' is less 'bad'?). For India, most scores are lower (20-10 is better) except methods 22/26(1,2), 15(1,3), 19(0,1,3), 21(0,1) and 23/27(4,5). Biggest decreases are associated with weighting 3 (i.e., effect of weighting is being diluted). For Australia, changes are largely negative, which means 20-10 is better. Also negative change is large for methods 3, 4, 6, 15, 23, 27. For method 14(0-2,4,5), 10-5 is much better, whereas for method 15, 20-10 is much better.

#### 3.4.2 Relative performance of methods

(i) Method 19 is still among the best, and 16 still one of the worst, for all 3 plates.

(ii) Many of the biggest changes (decreases) for North America and India occur when paleopoles published after 1983 are removed and calculation is based on AMP (method 16). In contrast, for Australia the biggest changes (decreases) are associated with method 15, and methods 16/17 have little effect (Fig. 19).

(iii) Only for North America, the 10/5 Myr bin/step methods generally and unexceptionally produce better similarities than the 20/10 Myr methods do (Fig. 19).

(iv) All the APP methods with kinds of corrections, 1, 9, 13, 19, 21 and 25, show less changes for all the three plates. Methods 23 and 27 show less changes for North America and India, but more changes for Australia.

(v) APP still outperforms AMP. For North America and India, the percentage of factor of greater than 3 (about 26.19% and 3.57% respectively) is less than 10-5 (about 44.05% and 40.48% respectively), while for Australia the percentage (about 70.24%) is more than 10-5 (58.33%).

(vi) Both scores and AMP/APP difference are low for methods 4/5, 6/7 applied onto North America and India.

(vii) Both scores and AMP/APP difference are high for methods 16/17.

(viii) Scores are high but AMP/APP difference are low for methods 2/3, 22/23 on North America and India.

(ix) Scores are low but AMP/APP difference are high for methods 0/1 on North America and Australia.

(x) Comparable AMP/APP (20-10 vs 10-5) appears for 8/9 on Australian plate.

(xi) Method 1 is still a good performer.

(xii) Methods 4/5, 6/7 are still 'good' aggressive filters.

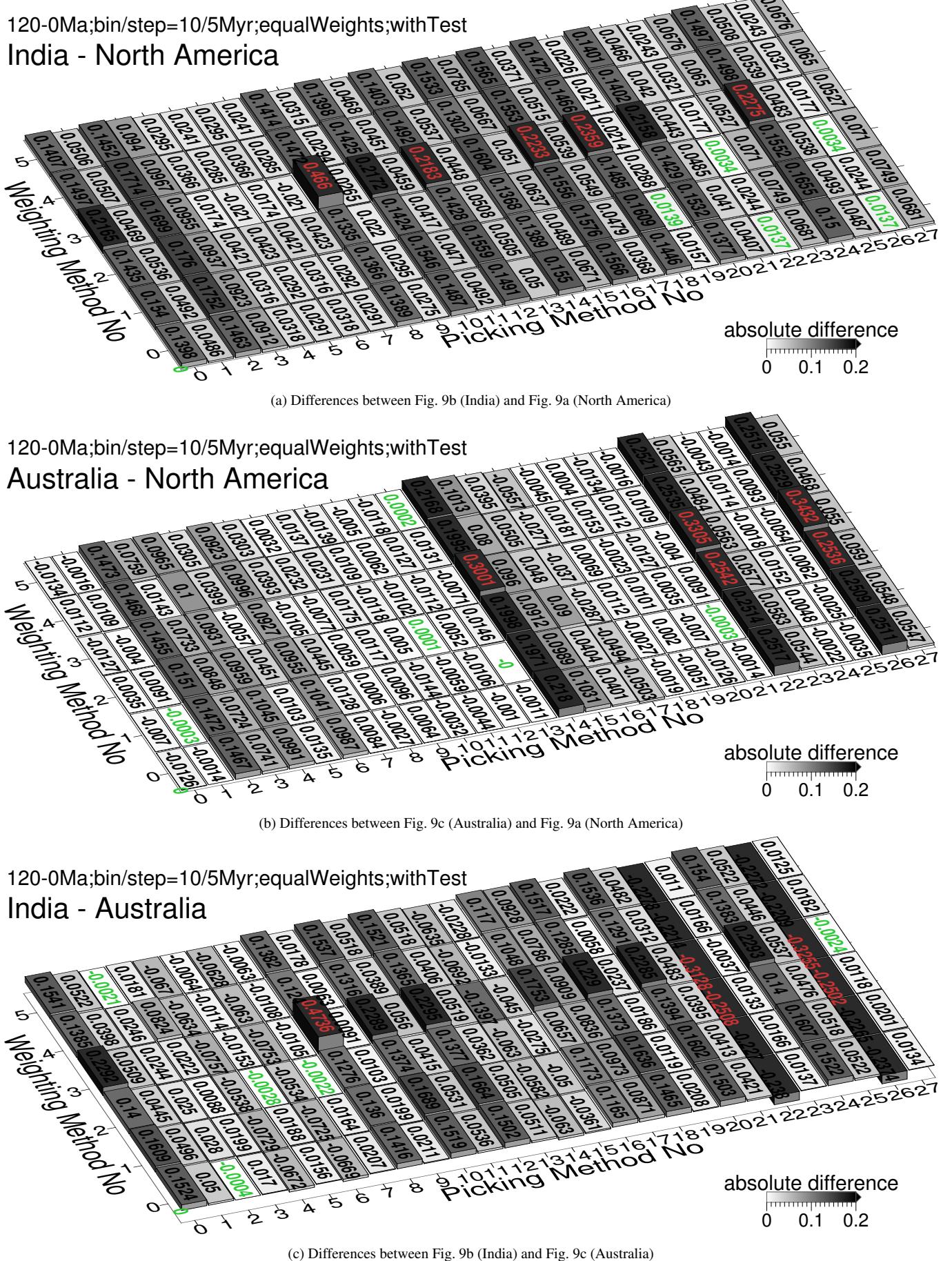
(xiii) Methods 2/3, 22/23 (at least 2 and 22) are still 'bad' aggressive filters. And the scores derived from these methods have control on overall range of scores.

(xiv) Methods 16/17 are still 'bad' for North America and India. Method 16 is still 'bad' and method 17 is still 'good' for Australia.

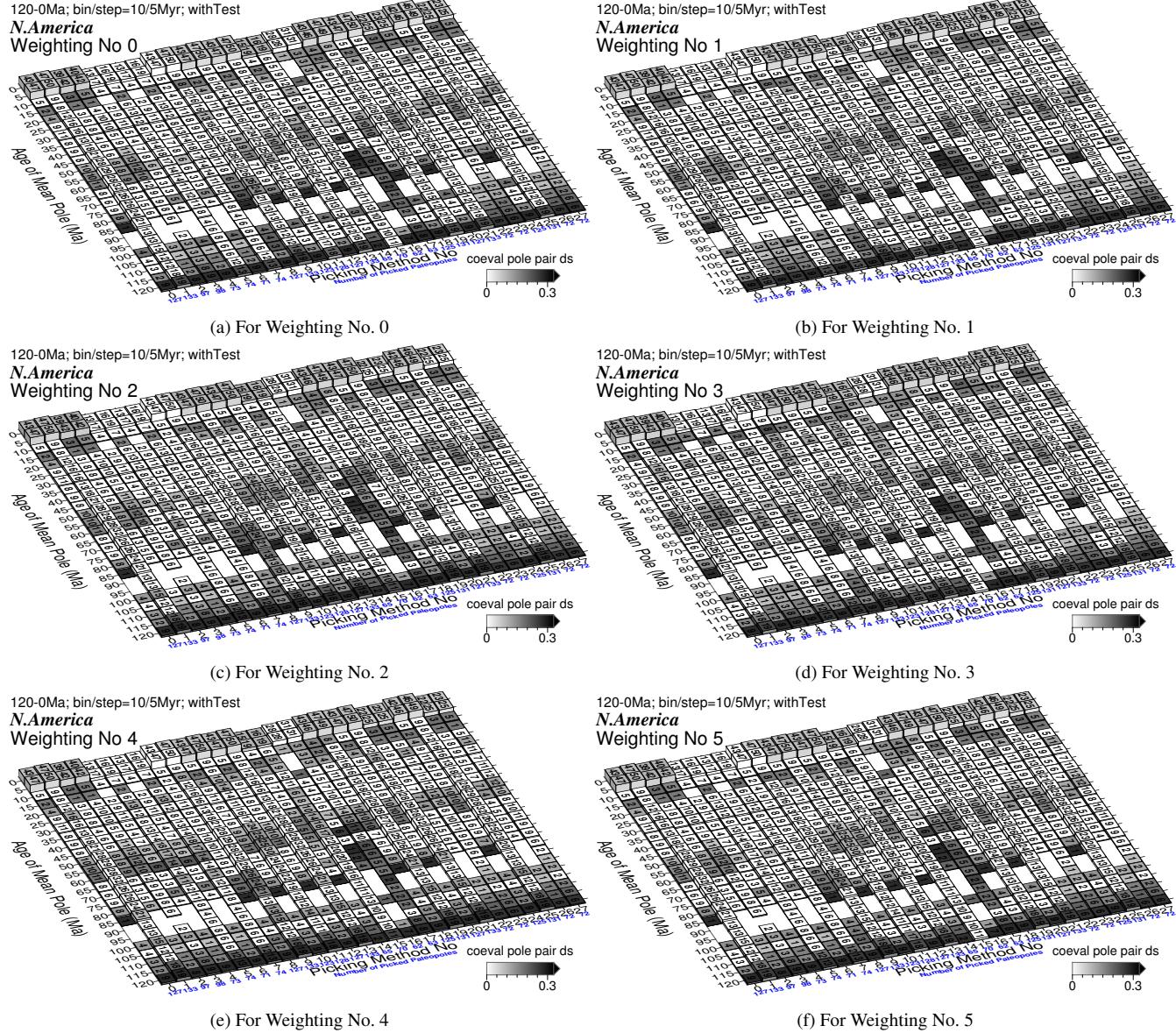
#### 3.4.3 How time window/step size affects results

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. 18e is improved a bit compared to Fig. 10e). 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. 18 and even 30/15 Myr window and step; Table 5 and Fig. 20) 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. 2 Myr window 1 Myr step; Table 5 and Fig. 20). 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 in Chapter 4). A variety of ways of binning the data (here 30–2 Myr window size and half of the size as step) are being tested to see which one produces the better and more appropriately smoothed fit.

Note that there are 135, 75 and 99 paleopoles that compose of 120–0 Ma APWPs of North America, India and Australia respectively. Does the reason of 10/5 Myr generally better than 20/10 could be the relatively larger number of paleopoles for North America? Since theoretically for each sliding window, the more "bad" paleopoles it contains, the worse similarity we should obtain. In the contrary, the less paleopoles the window contains, the weaker the effect of averaging out "bad" poles' influence would be. So is there a threshold number of paleopoles for making an paleomagnetic APWP? For example, for making a 120–0 Ma APWP, do the results indicate the best number of paleopoles we need should be some value between 99 and 135? Here a test is implemented as follows. With the results from the 10/5 and 20/10 bin/step together, 2/1, 4/2, 6/3, 8/4, 12/6, 14/7, 16/8, 20/10, 24/12 and 30/15 Myr bin/step are also used to generate paleomagnetic APWPs for North America, India and Australia to see which bin/step size would make paleomagnetic APWP closest to reference path. Will the similarities they generate be generally worse than those the 10/5 Myr bin/step generates? Or will they be better first and then worse than those the 10/5 Myr bin/step generates when the bin/step sizes increase up to 20/10 Myr? For the best results (Table 5), as expected, AMP needs wider sliding window and step to get closer to the reference path while APP does not (Fig. 20). Even the best sizes of sliding window and step are assigned for AMP, the results from APP are still much better than those from AMP. Picking methods (directly re-



**Figure 13.** Differences between grids in Fig. 9. 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.



**Figure 14.** Tested spatial difference ( $d_s$ ) values (color shaded) between North American paleomagnetic APWPs and its predicted APWP from the FHM and related plate circuits. The paths are in 10 Myr bin and 5 Myr step. The number labels on the grids (including grid heights) are the numbers of site mean poles that are contributing to make each mean path pole.

lated to N) are still the key influence factor of choosing a better sliding window size and step size of moving averaging, although weighting methods are also important.

**3.4.3.1 What to expect is** the difference values for larger window/step size should be generally lower than those for smaller window/step size, which further could result in more best methods and less worst methods.

**3.4.3.2 The results** are summarised in Table 4, Table 5 and Fig. 20.

### 3.4.3.3 Conclusions:

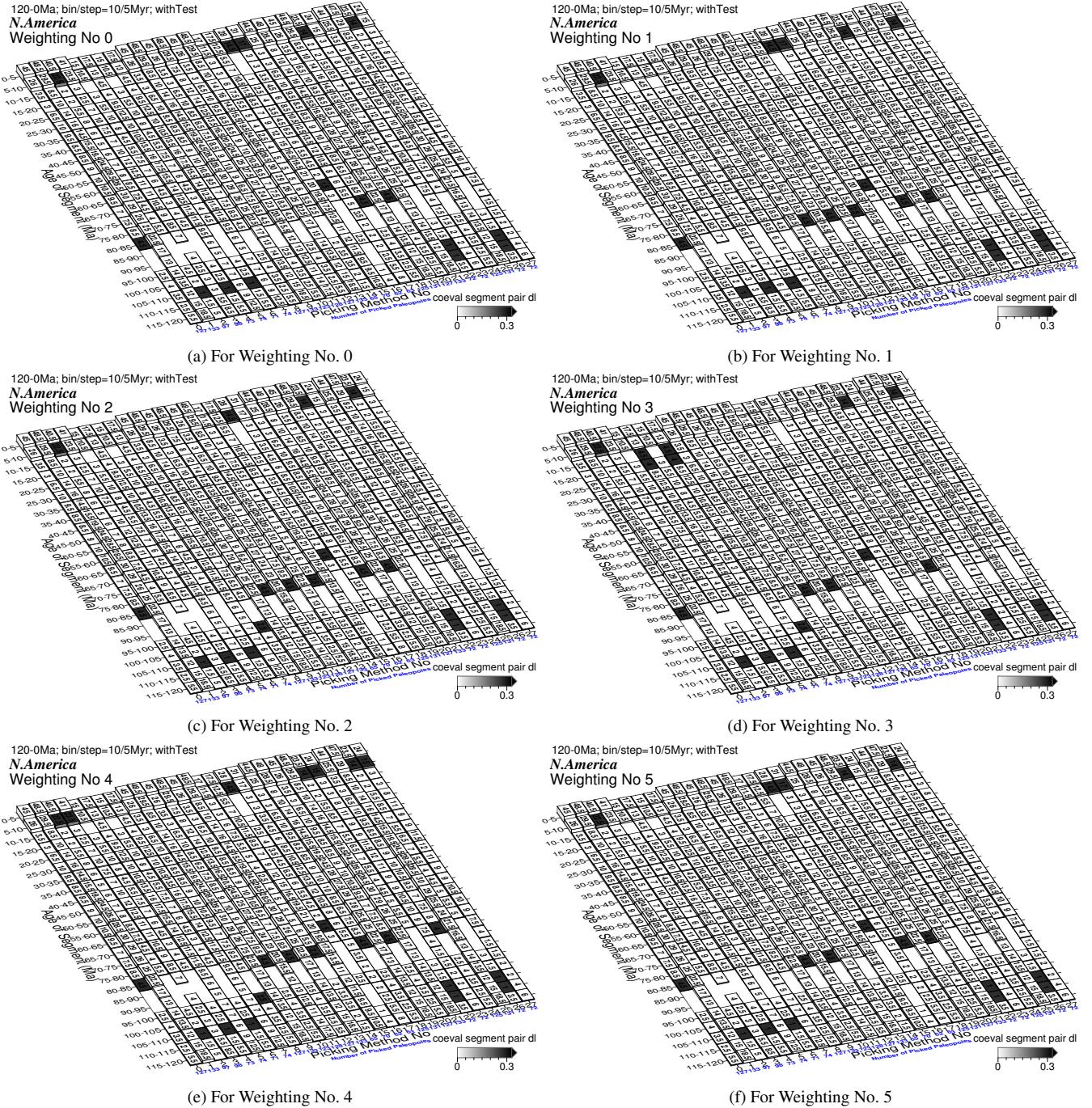
If AMP has to be used, better results can be obtained through using large sizes of sliding window and step, commonly more than

24/12 Myr. In addition, we should be cautious when weighting 3 is used with AMP.

APP is still recommended, not only because the temporal uncertainty is incorporated into the algorithm but also the results from APP are not as sensitive as AMP to the changes of sliding window and step sizes. In fact, for APP the results from different window and step sizes are much more stable than those from AMP (Fig. 20). This means we actually do not need to worry about what sizes should be chosen for the sliding window and step when we use APP method.

### 3.5 Different reference path

Fig. 21 and Fig. 23 show the CPD scores for the APWPs generated with all 28 picking and 6 weighting methods, compared to the



**Figure 15.** Tested length difference ( $d_l$ ) values (color shaded) between North American paleomagnetic APWPs and its predicted APWP from FHM and related plate circuits. The paths are in 10 Myr bin and 5 Myr step. The labeled numbers on the grids are the averaged numbers of site mean poles that are contributing to each segment's two mean path poles.

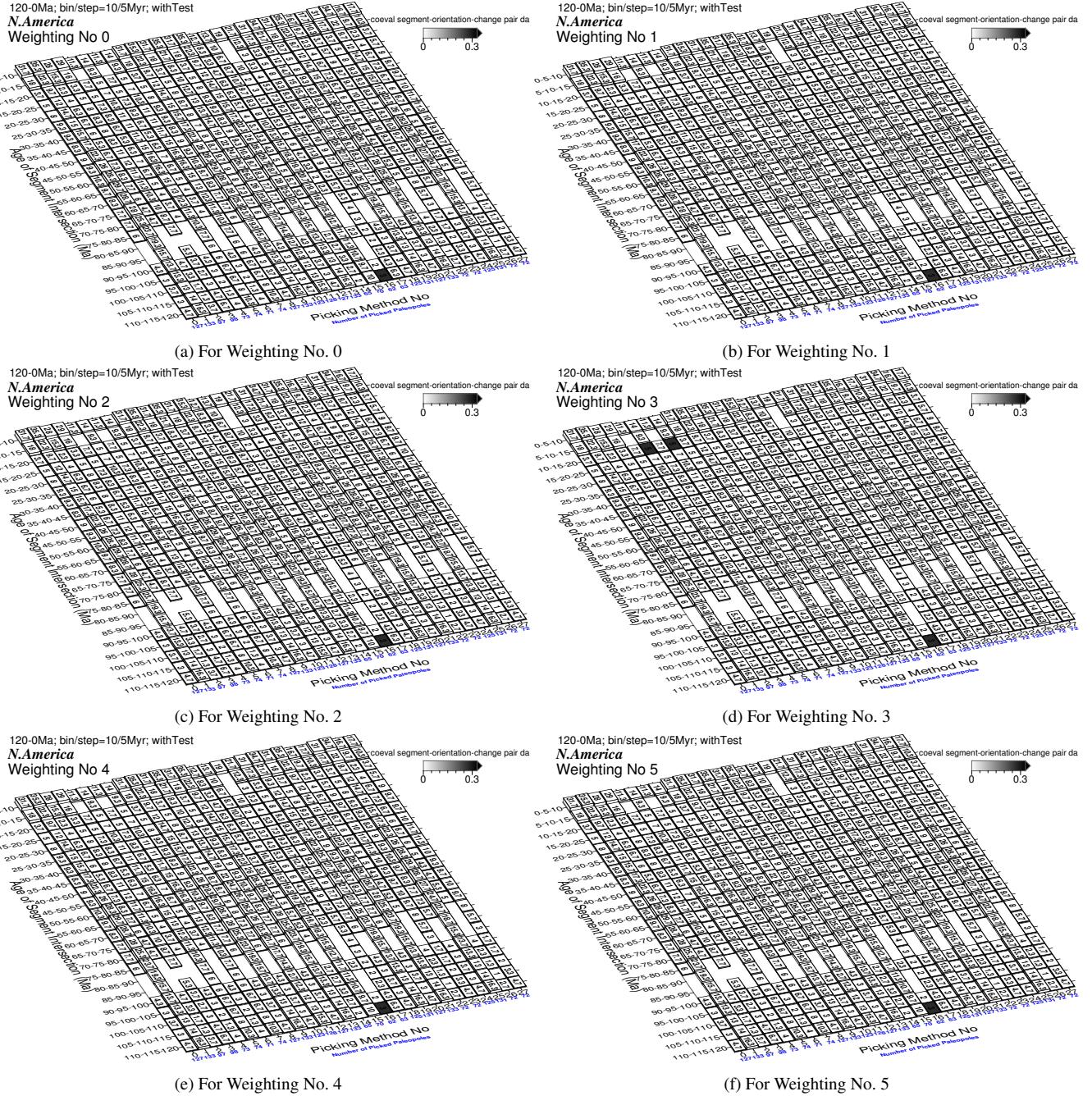
MHM reference paths instead, still with the same picking time window width 10 and 20 Myr, and the same window step 5 and 10 Myr respectively.

### 3.5.1 Overall change

Although the reference path is changed to MHM, mean, median and range values of  $\mathcal{CPD}$  scores are almost unchanged and comparable (Fig. 9 vs. Fig. 21, and Fig. 17 vs. Fig. 23).

For 10/5 Myr bin/step, the absolute differences (Fig. 25) are

all lower than 0.066, and actually most are less than 0.01. This indicates that for comparing with paleomagnetic APWPs choosing fixed or moving hotspot model for generating a reference path is not quite different. Therefore selecting fixed or moving model for having a reference path is not a priority. However, based on the signs of the differences between the scores from FHM and MHM (Fig. 25), for North America (Fig. 25a), FHM derived path is a slightly better reference in general, while for both India (Fig. 25b) and Australia (Fig. 25c), generally MHM derived path is a slightly better choice.



**Figure 16.** Tested angular difference ( $d_a$ ) values (color shaded) between North American paleomagnetic APWPs and its predicted APWP from FHM and related plate circuits. The paths are in 10 Myr bin and 5 Myr step. The labeled numbers on the grids are the averaged numbers of site mean poles that are contributing to each segment-orientation-change's three mean path poles.

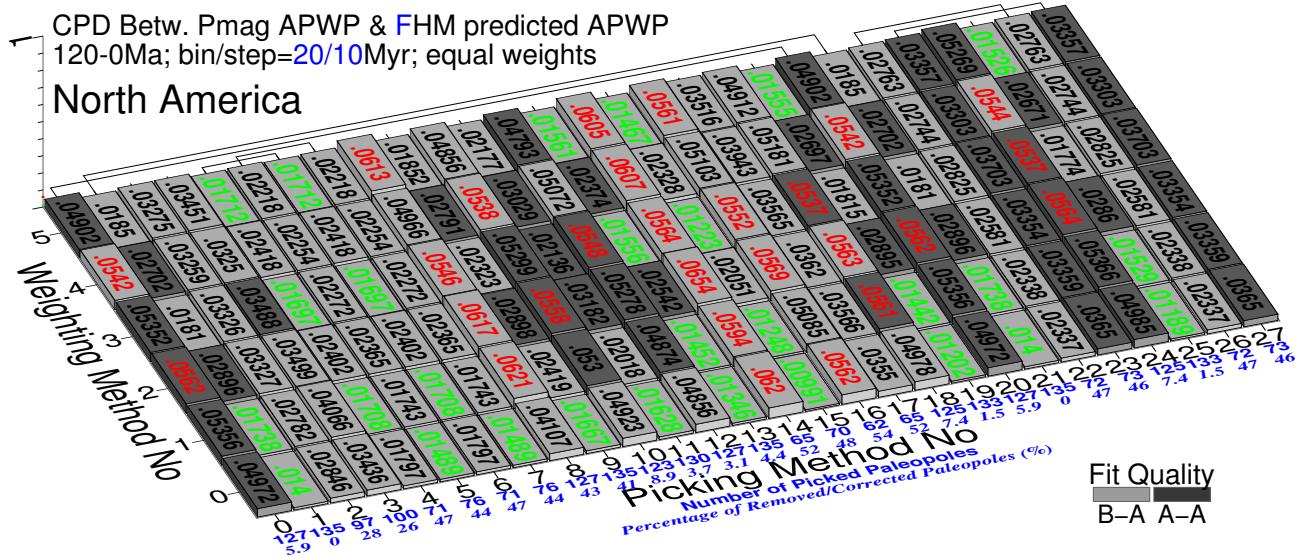
For North America, large changes seem favored by weighting method 3, and small changes seem favored by weighting method 5 or picking method 15 (Fig. 25). Method 3 brings minor changes to both North America and India. Methods 22 and 26 show large changes and methods 17, 19, 21 and 25 show minor changes for both India and Australia.

### 3.5.2 Relative performance of methods

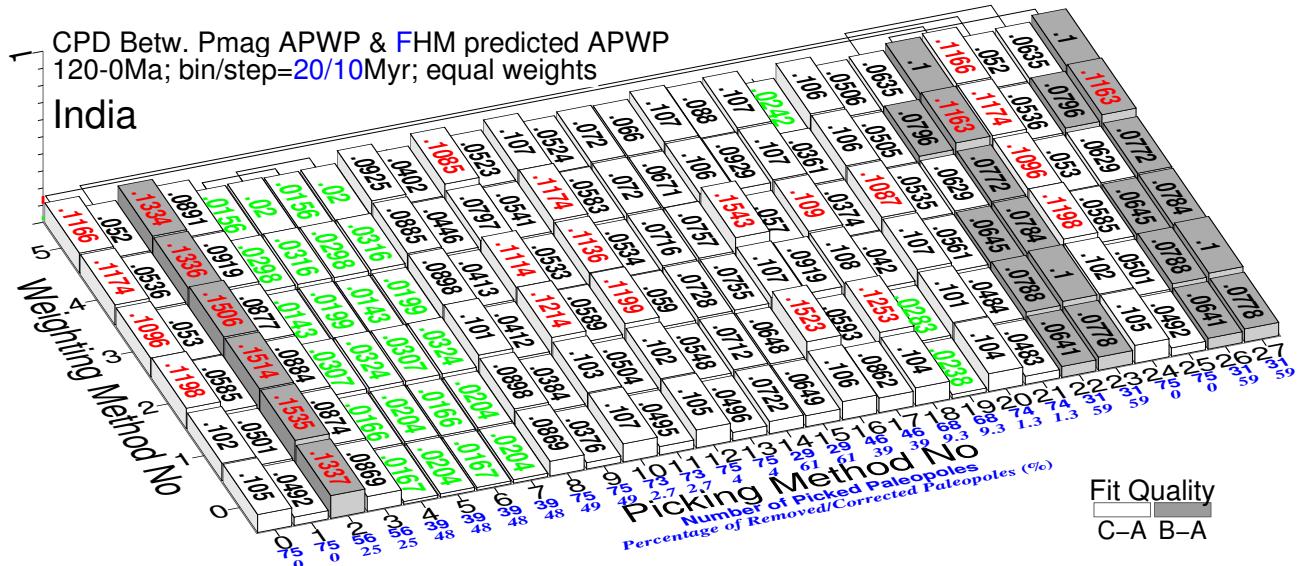
(i) When 10/5 Myr bin/step is applied, method 19 is still among the best, and 16 still one of the worst, for all 3 plates. Even when

20/10 Myr bin/step is applied, 19 is still relatively a “good” method, and 16 a “bad” one.

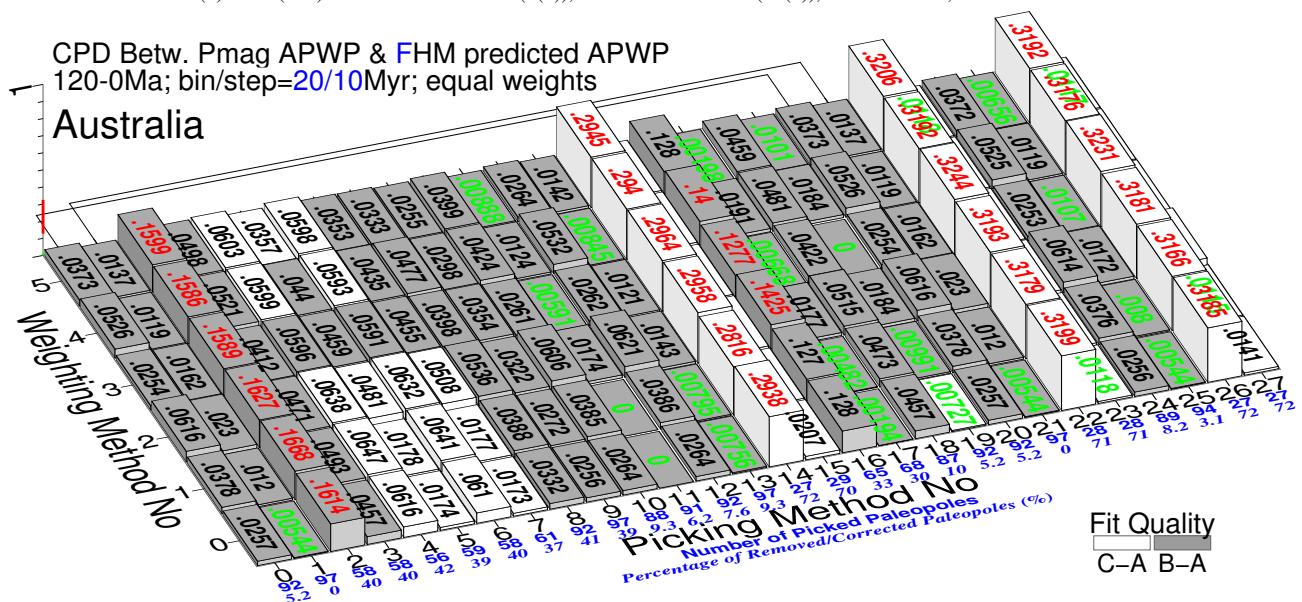
(ii) APP still outperforms AMP. For 10/5 Myr bin/step, the percentage of factor of greater than 3 (about 19.05%, 38.1% and 48.81% for North America, India and Australia respectively) is less than FHM (about 40.5%, 39.3% and 50%). For 20/10 Myr bin/step, the percentage of factor of greater than 3 (about 3.6% and 46.43% for India and Australia) is still less than FHM (about 3.8% and 71.4%) whilst for North America the percentage (about 44.05%) is more than FHM (28.6%).



(a) North America (101 with children): minimum 0.00991 (15(0)), maximum 0.0654 (14(2)), mean 0.0339, median 0.0296371

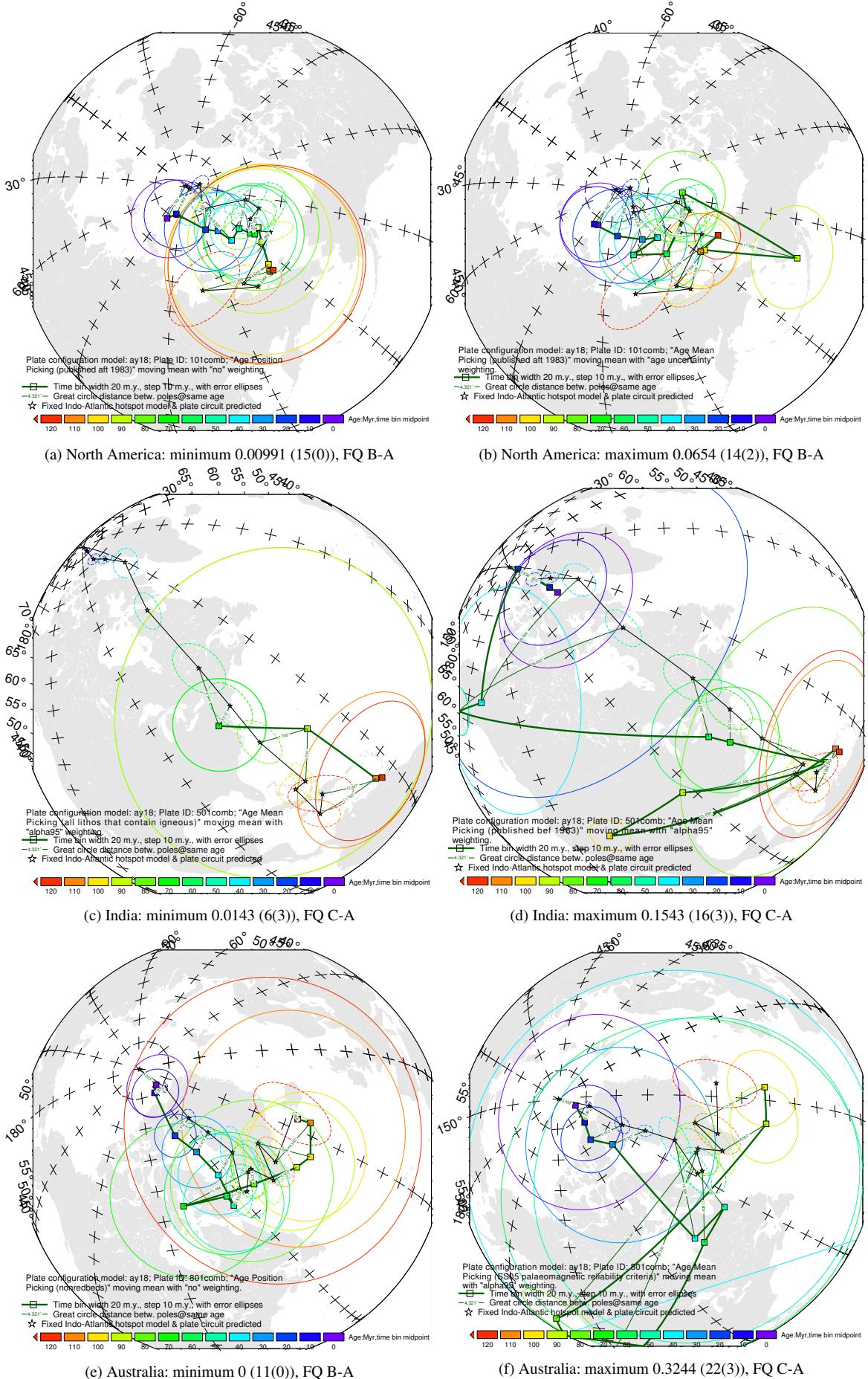


(b) India (501): minimum 0.0142822 (6(3)), maximum 0.154278 (16(3)), mean 0.07384, median 0.072088



(c) Australia (801 with children): minimum 0 (11(0,1),19(3)), maximum 0.324438 (22(3)), mean 0.0682, median 0.036432

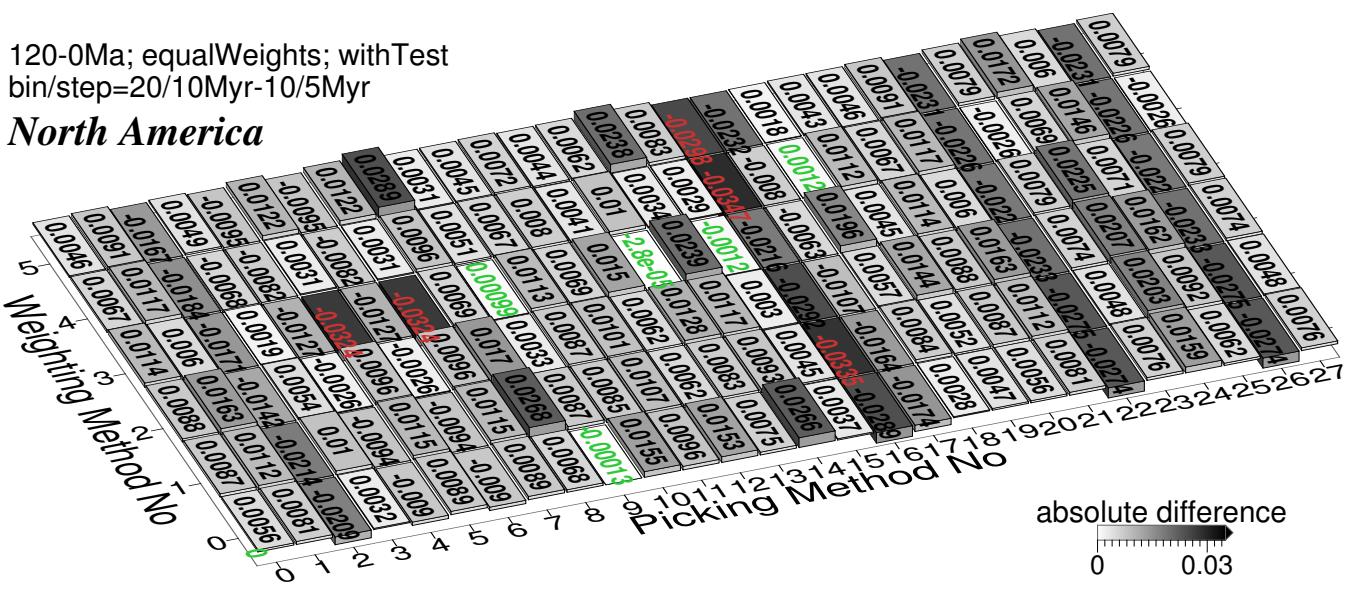
**Figure 17.** As Fig. 9, here the paths are generated in 20 Myr bin and 10 Myr step. The difference values less than one-standard-deviation interval of the whole 168 values are colored in green, more than one-standard-deviation interval colored in red. Compare the numbers of picked paleopoles with those in Fig. 9.



**Figure 18.** Path comparisons with best and worst difference values shown in Fig. 17. The parenthetical remarks are Picking No and Weighting No.

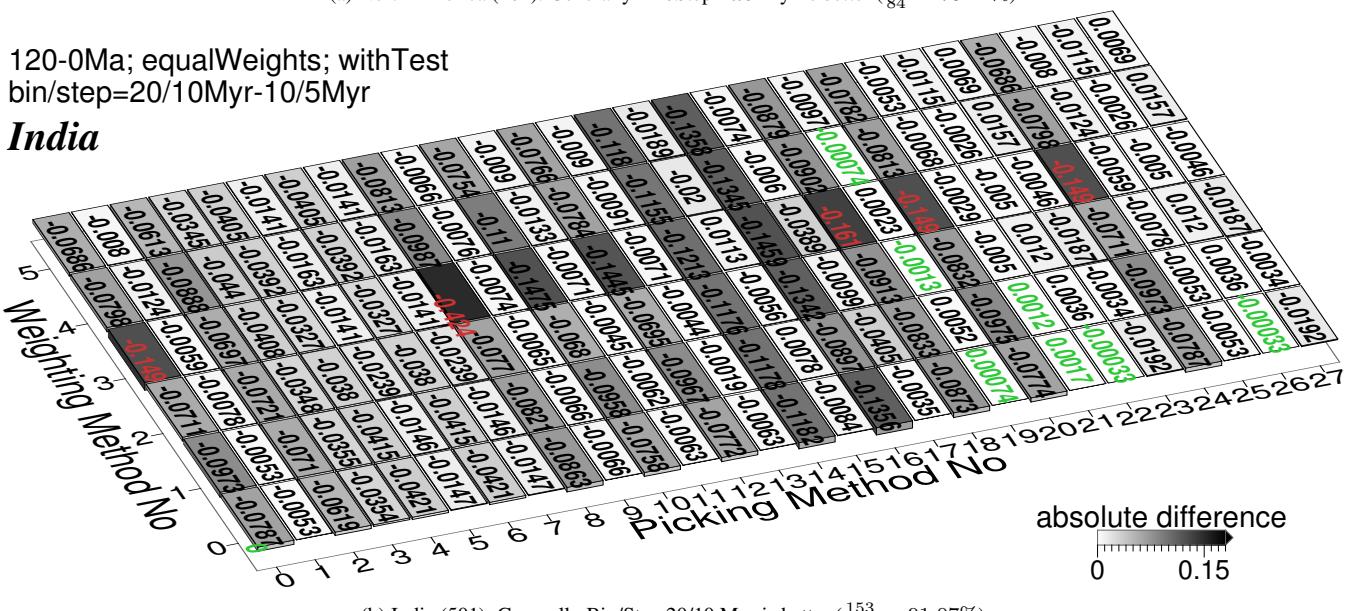
120-0Ma; equalWeights; withTest  
bin/step=20/10Myr-10/5Myr

### North America



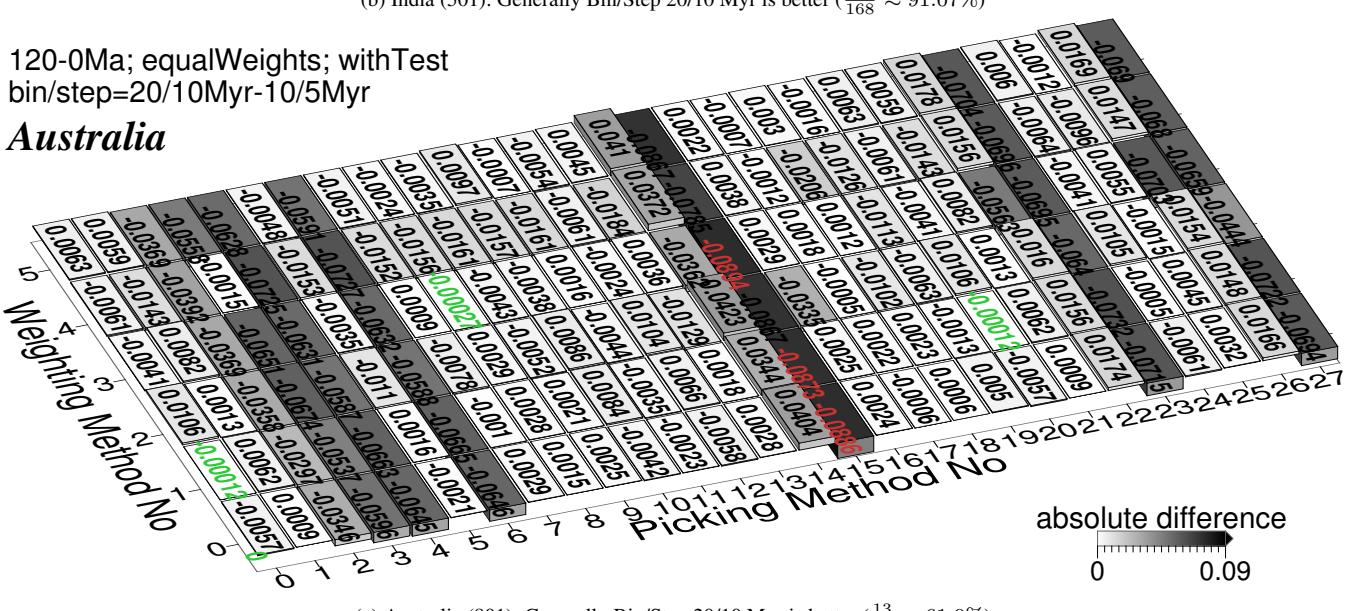
120-0Ma; equalWeights; withTest  
bin/step=20/10Myr-10/5Myr

### India



120-0Ma; equalWeights; withTest  
bin/step=20/10Myr-10/5Myr

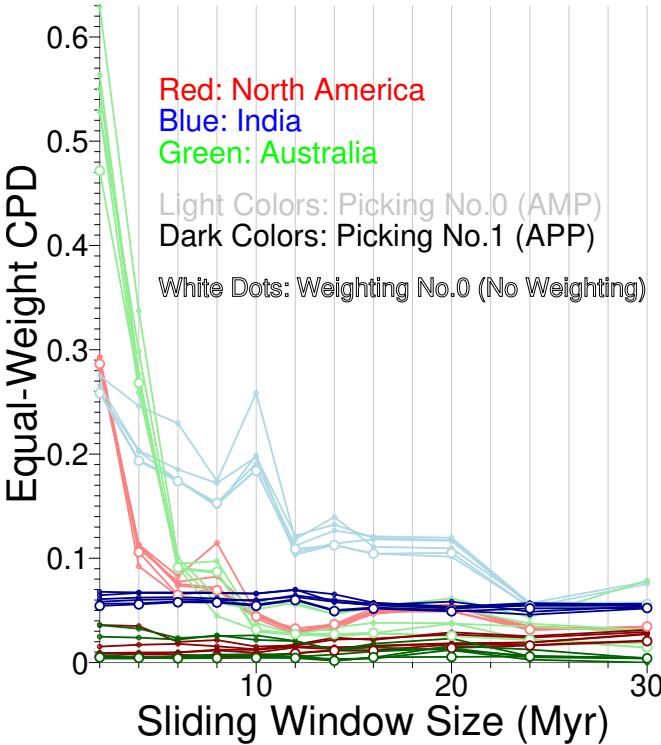
### Australia



**Figure 19.** Differences between grids in Fig. 9 (10 Myr bin, 5 Myr step) and Fig. 17 (20 Myr bin, 10 Myr step). 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.

**Table 4.** Consistency check on comparisons of picking methods' performance between 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 Myr 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. 9a	Fig. 17a	Y	3 more best: Picking no. 4, 6 and 9 only for 20/10 (E)	Same: Picking no. 1, 5, 7, 11, 13, 15, 19, 21 and 25	N	Picking no. 2, 5, 7, 17, 22 and 26 for 10/5 (E); 0, 8, 10, 12, 20 and 24 for 20/10 (UE)	Same: Picking no. 14, 16 and 18	N	Y	Y	N	N	Positive values in Fig. 19a			
Fig. 9b	Fig. 17b	Y	4 more best: Picking no. 9, 21, 22 and 26 only for 10/5 (UE)	Same: Picking no. 4, 5, 6, 7 and 19	N	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	N	Positive values in Fig. 19b				
Fig. 9c	Fig. 17c	Y	2 more best: Picking no. 23 and 27 only for 20/10 (E)	Same: Picking no. 1, 11, 13, 17, 19, 21 and 25	Y	1 more worst: Picking no. 4 only for 10/5 (E)	Same: Picking no. 2, 14, 16, 22 and 26	Y	Y	Y	N	Positive values in Fig. 19c				
MHM																
Fig. 21a	Fig. 23a	N	Picking no. 1, 9, 11, 13, 19, 21 and 25 for 10/5 (UE); 22 and 26 for 20/10 (E)	Same: Picking no. 5, 7 and 15	N	Picking no. 16 and 17 only for 10/5 (E); 8 only for 20/10 (UE)	Same: Picking no. 0, 10, 12, 14, 18, 20 and 24	N	Y	N	N	N	(0,8,10,11,12,14,18,20,24)(0-5) (1,13,19,21)(0,1,5) 3(0,3,5) (5,7)(0-2,4,5) 15(0-4) (9,23,27)(0,1,3,5); account for 60.71%			
Fig. 21b	Fig. 23b	N	Picking no. 19 and 21 only for 10/5 (UE); 4 and 6 only for 20/10 (E)	Same: Picking no. 5, 7, 22 and 26	Y	3 more worst: Picking no. 8, 14 and 20 for 10/5 (E)	Same: Picking no. 0, 2, 10, 12, 16, 18 and 24	Y	Y	Y	Y	N	(1,25)(0,3,5) 15(3) (19,23,27)(0-5) 21(0,1,3-5) (22,26)(4); account for 19.05%			
Fig. 21c	Fig. 23c	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	(0,20)5 (1,11,13,16,18,19,21,25)(0,1,3,5) (5,7)3 (8,17)(0-3,5) 9(0,1) 10(1,5) (12,2) 14(2,4,5) (22,26)(0-2,4,5) 24(1,2,5); account for 39.88%			



**Figure 20.** Plot of the equal-weight CPD scores collected in Table 5. Note that here the step size is always half of the sliding window size and the reference path is the FHM derived.

(iii) Both scores and AMP/APP difference are still low for methods 4/5, 6/7.

(iv) Both scores and AMP/APP difference are still high for methods 16/17 (method 17 gives low scores for Australia).

(v) Still scores are high but AMP/APP difference are low for methods 2/3.

(vi) Still scores are low but AMP/APP difference are high for 0/1 on Australia.

(vii) Comparable AMP/APP (20-10 vs. 10-5, or FHM vs. MHM) still appears for 8/9 on Australian plate.

(viii) Method 1 is still a good performer.

(ix) Methods 4/5, 6/7 are still ‘good’ aggressive filters.

(x) Methods 2/3 (at least 2) are still ‘bad’ aggressive filters.

(xi) For 10/5 Myr bin/step, method 22 is still better than 23 for India. For 20/10 Myr bin/step, method 22 is still better than 23 for both North America and India.

(xii) Methods 16/17 are still ‘bad’ for North America and India.

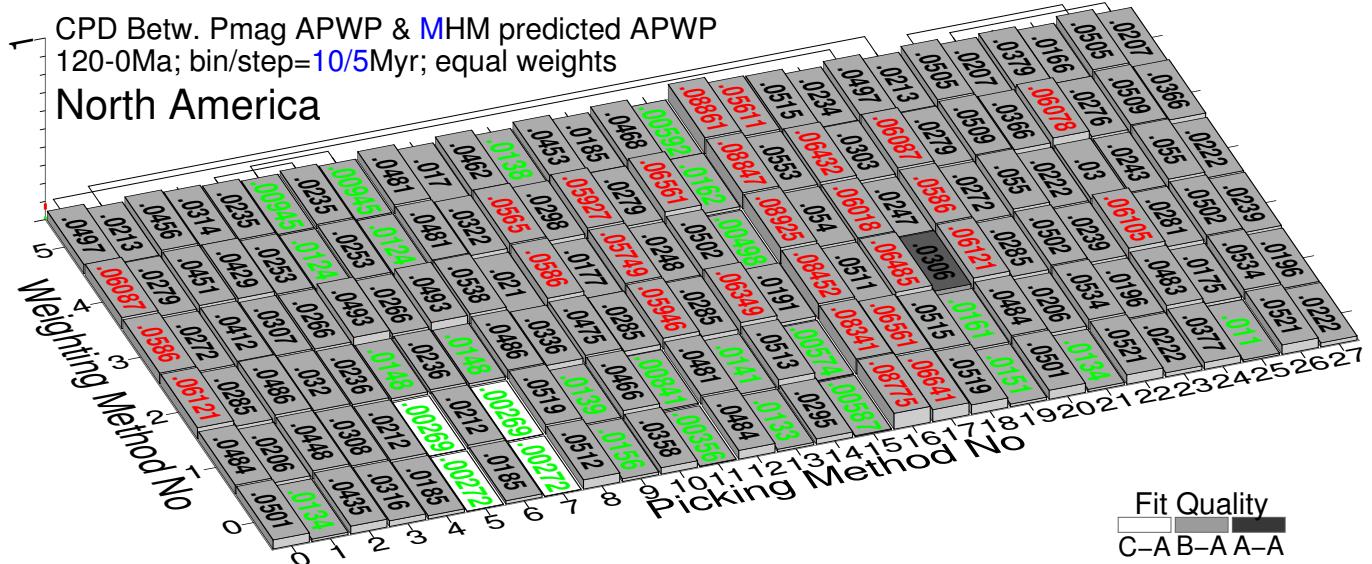
## 4 DISCUSSIONS

The following discussions will be in Q&A style.

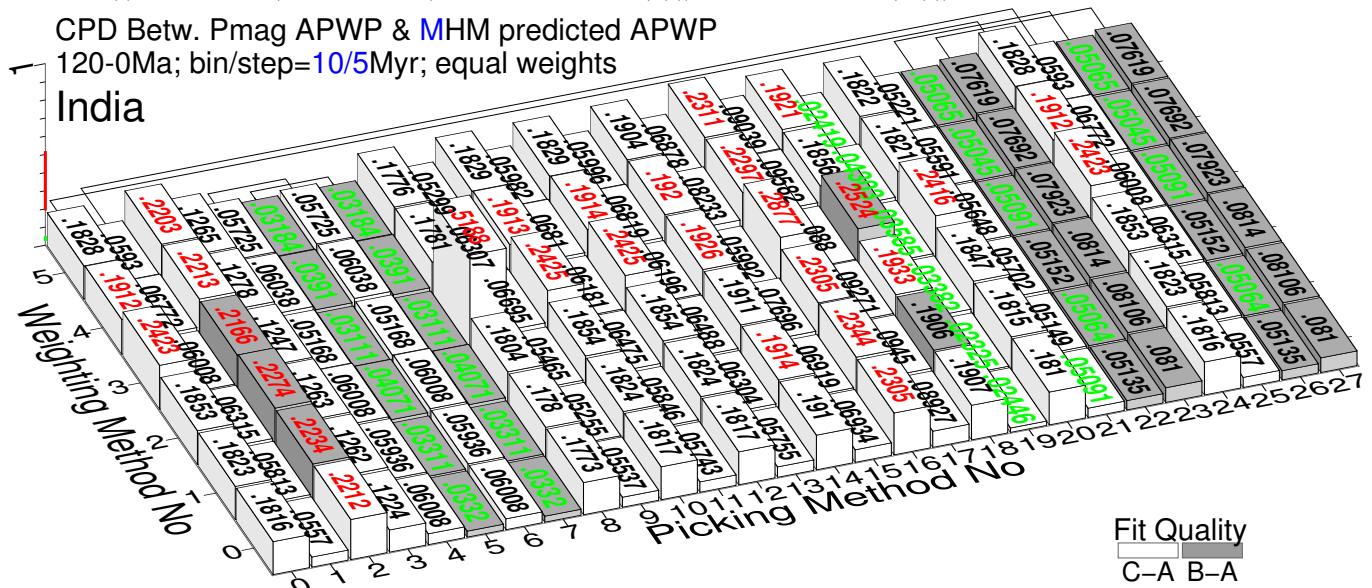
### 4.0.1 Question: Why the APP methods generally produce better similarities than AMP methods do?

Paleomagnetic (Mean) A95 represents precision (how well constrained calculated poles are), and (mean) coeval poles' GCD represents accuracy (how close calculated poles are to the reference path; Fig. 26a and Fig. 26b). Compared with AMP, APP usually improves both and generates paths with higher accuracy and also higher precision (generally increasing N).

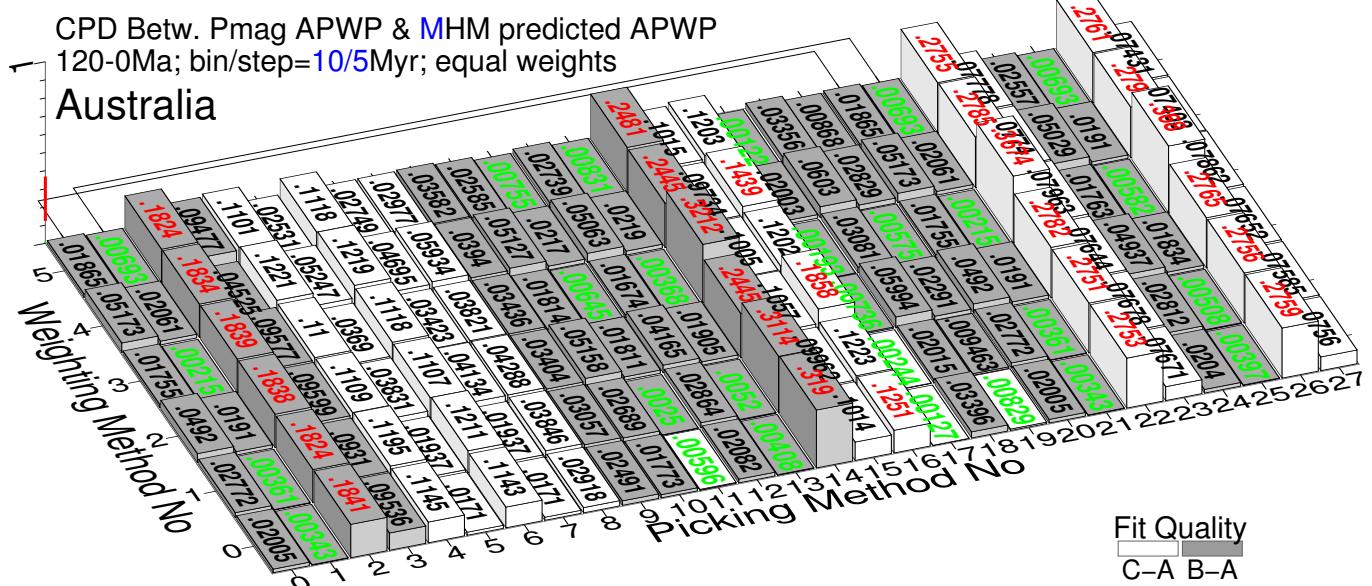
The fact that APP increases the number of paleopoles (N) in each sliding window would potentially average out some “bad” (i.e. inaccurate) poles and improves the fit between the paleomagnetic APWPs and the model-predicted APWPs. The general effects that APP brings include the decreases in paleomagnetic A95s, or/and distances between compared coeval poles of paleomagnetic APWP and reference APWP (Fig. 26 and Fig. 27). However, if the added paleopoles were all or mostly “bad”, the improvement of fit would not occur. So the improvement of fit is not only because of the increase in N, but also because the majority of the additional poles



(a) North America (101 with children): minimum 0.00268588 (5(1)), maximum 0.0892467 (16(3)), mean 0.03674, median 0.03177075

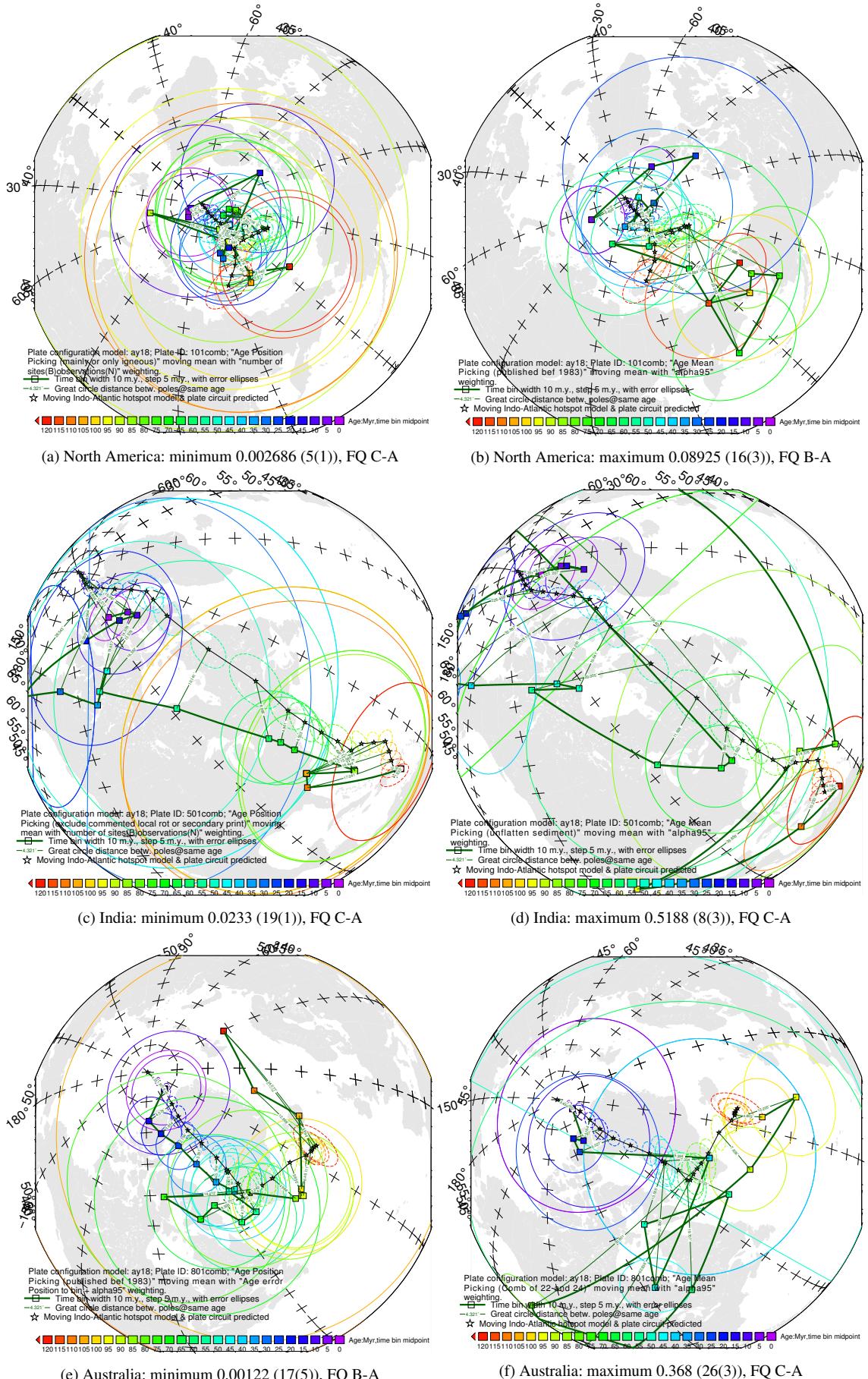


(b) India (501): minimum 0.0232517 (19(1)), maximum 0.51876 (8(3)), mean 0.11364, median 0.076556

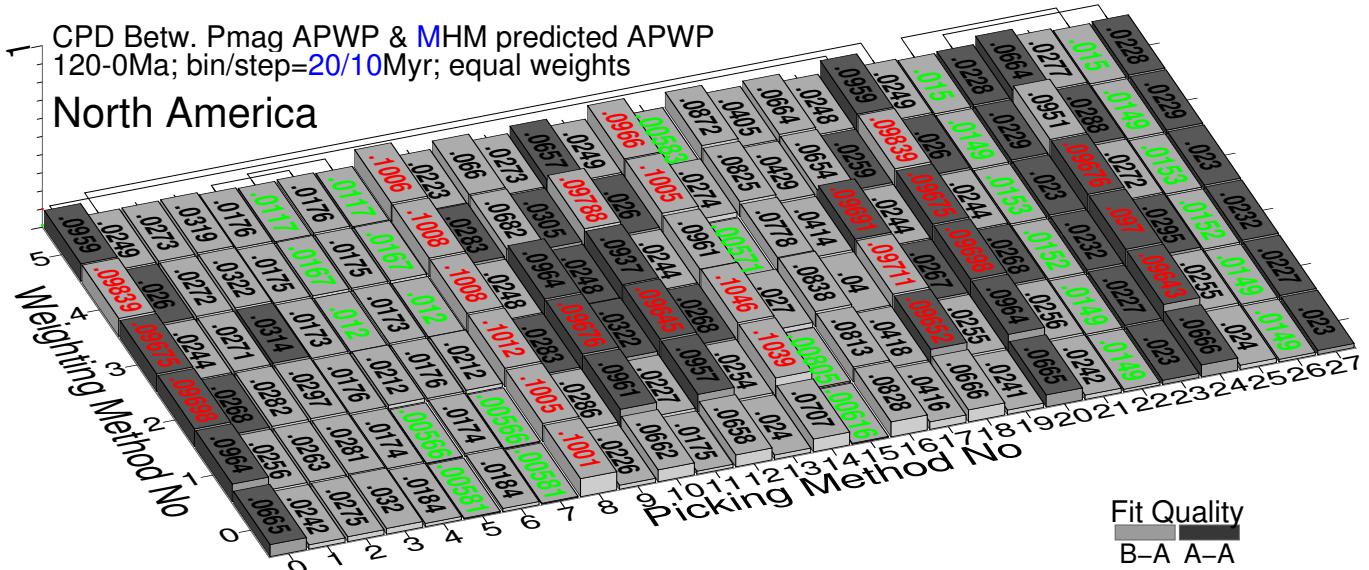


(c) Australia (801 with children): minimum 0.00122074 (17(5)), maximum 0.367952 (26(3)), mean 0.077, median 0.03893

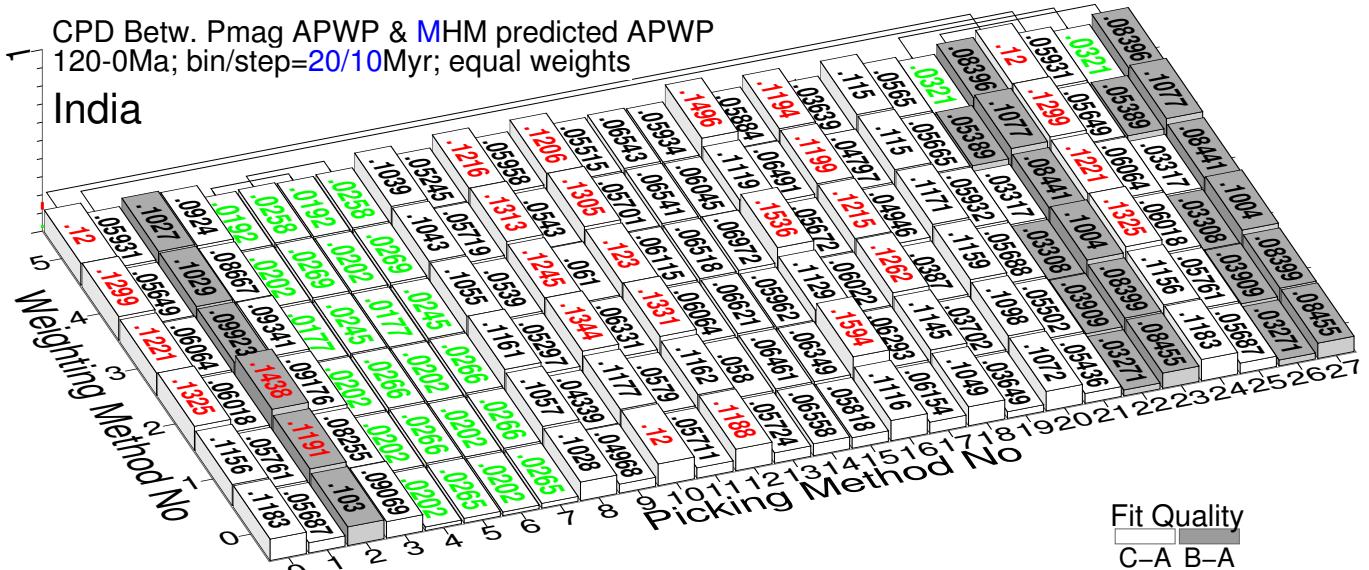
**Figure 21.** As Fig. 9, here the reference path is predicted from MHM. See the numbers of the picked paleopoles for methods in Fig. 9.



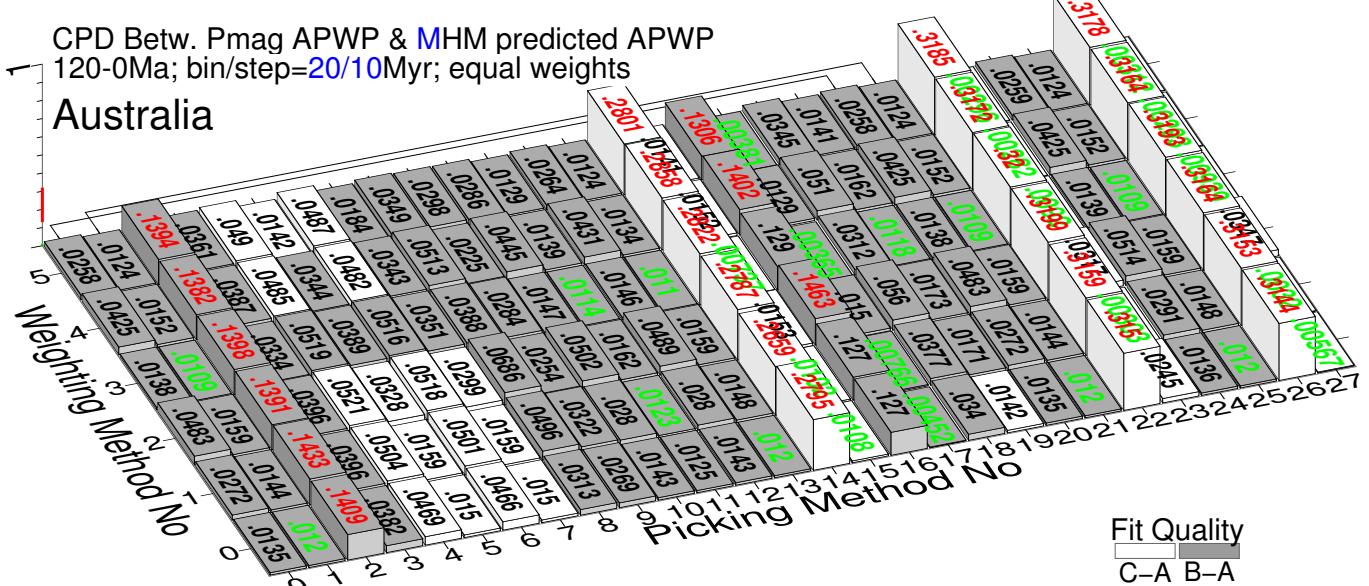
**Figure 22.** Path comparisons with best and worst difference values shown in Fig. 21. The parenthetical remarks are Picking No with Weighting No.



(a) North America (101 with children): minimum 0.00565784 (5(1)), maximum 0.104618 (14(2)), mean 0.043777, median 0.02679955

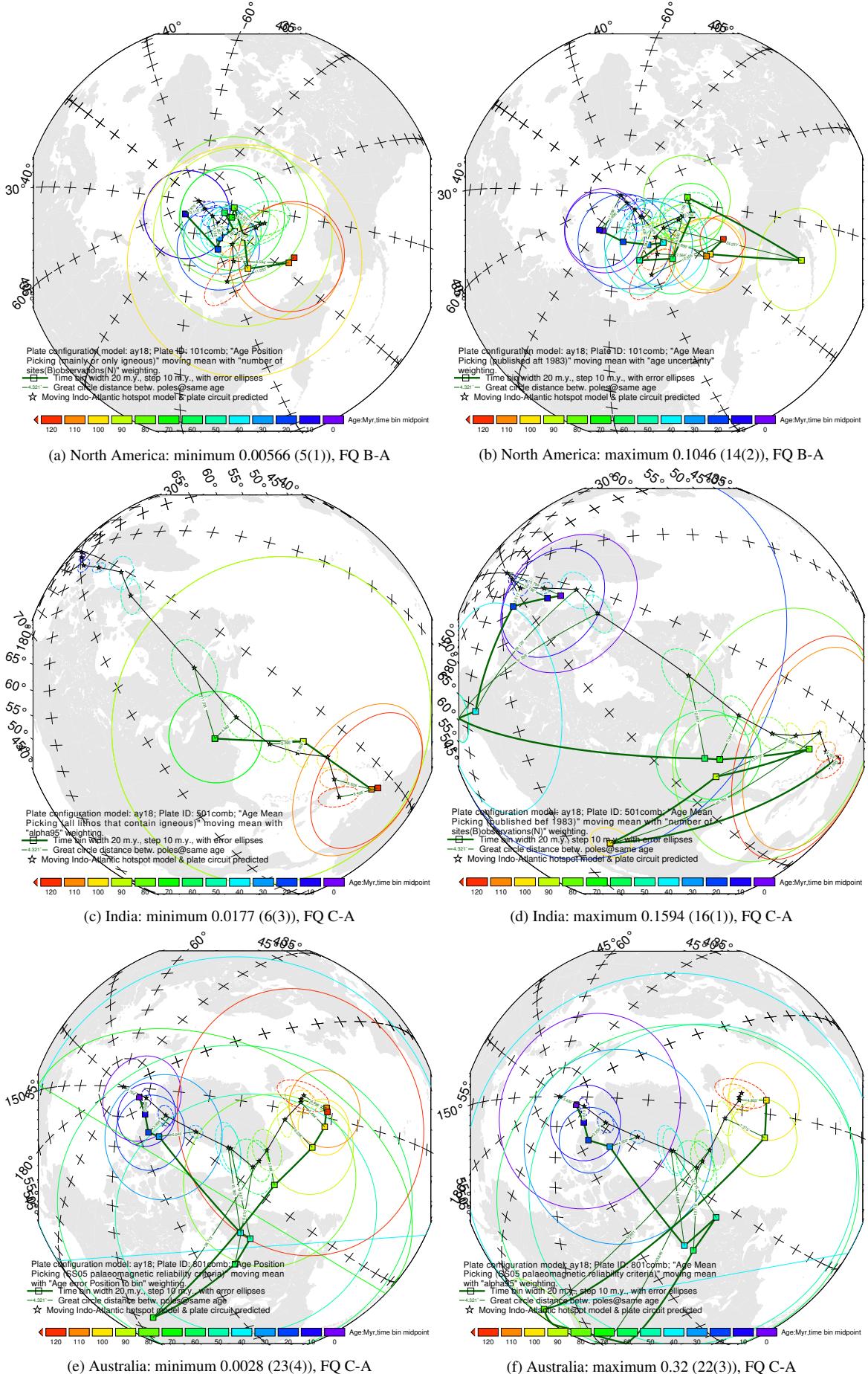


(b) India (501): minimum 0.0177 (6(3)), maximum 0.15937 (16(1)), mean 0.0745, median 0.061346

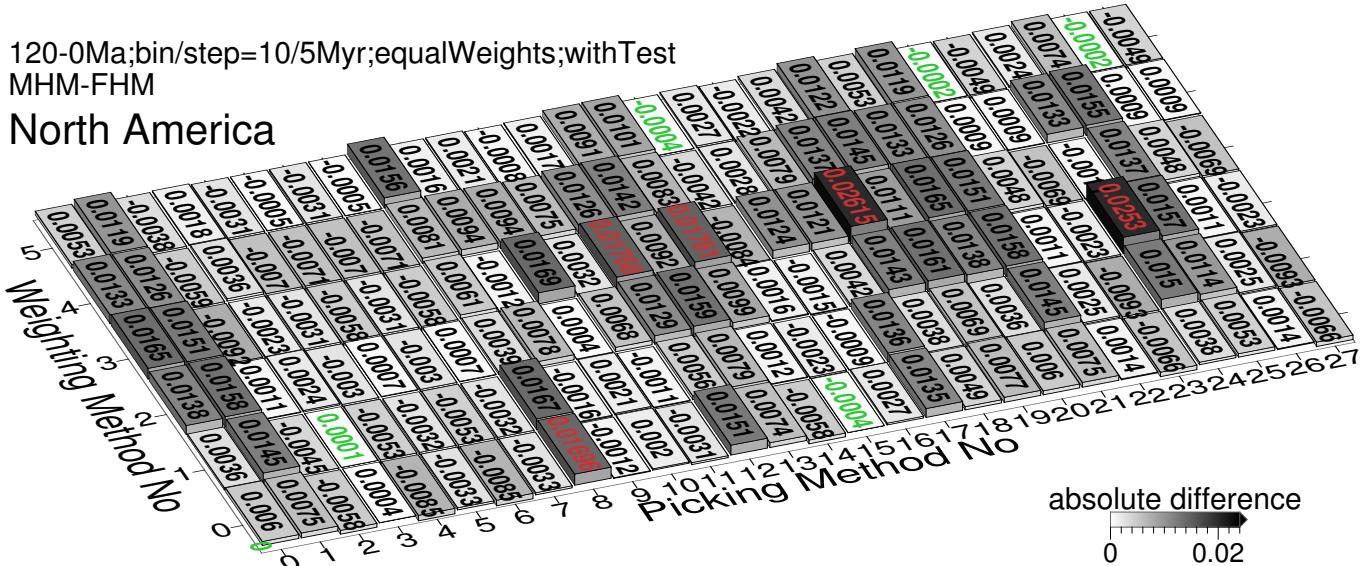


(c) Australia (801 with children): minimum 0.00282 (23(4)), maximum 0.31998 (22(3)), mean 0.062766, median 0.02803

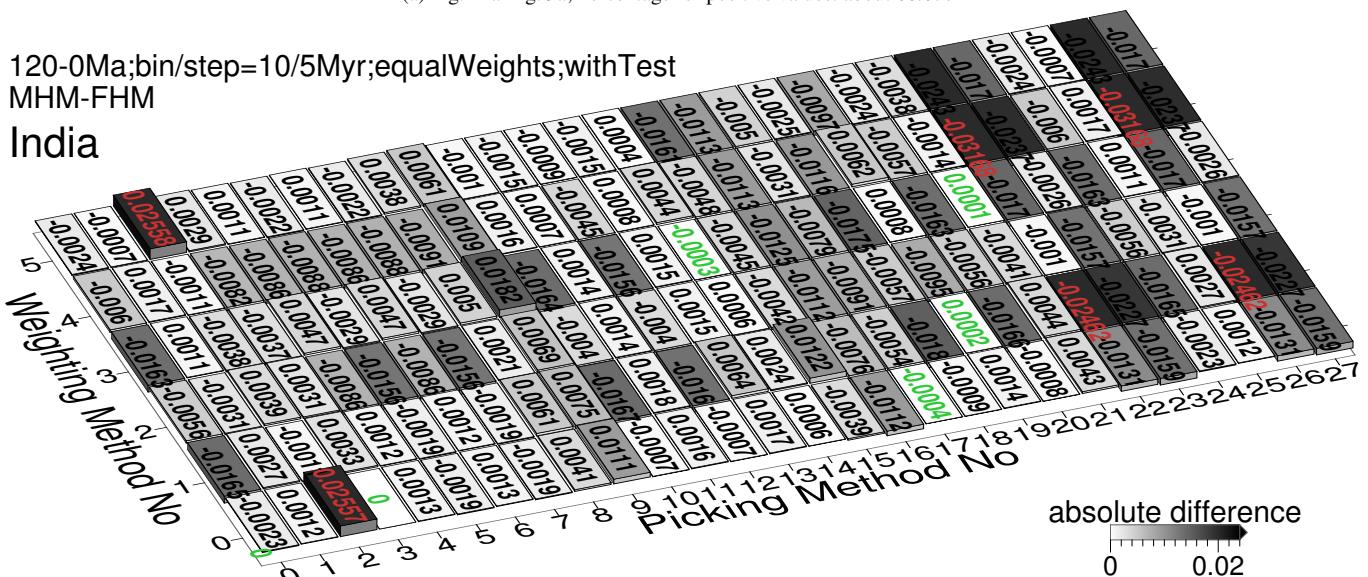
**Figure 23.** As Fig. 17, here the reference path is predicted from MHM. See the numbers of picked paleopoles in Fig. 9.



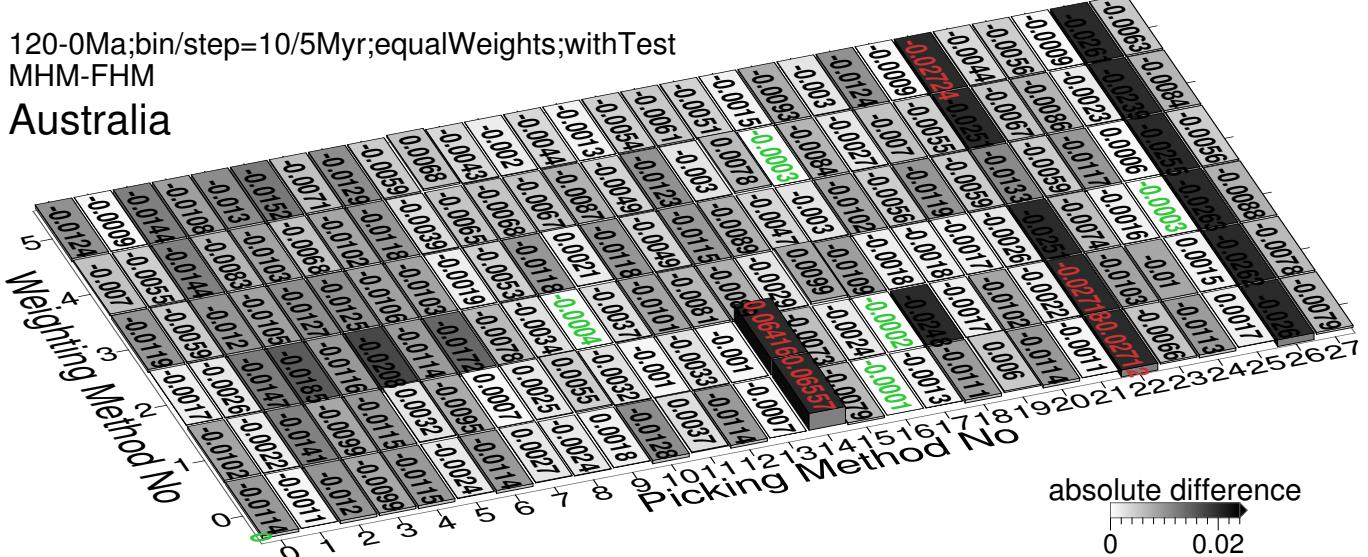
**Figure 24.** Path comparisons with best and worst difference values shown in Fig. 23. The parenthetical remarks are Picking No and Weighting No.



(a) Fig. 21a-Fig. 9a; Percentage for positive values: about 66.67%



(b) Fig. 21b-Fig. 9b; Percentage for positive values: about 34.52%



(c) Fig. 21c-Fig. 9c; Percentage for positive values: about 19.62%

**Figure 25.** Differences between results from two different reference paths, FHM (Fig. 9) and MHM (Fig. 21) derived. 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.

**Table 5.** Equal-weight 120–0 Ma CPDs for the three representative continents' paleomagnetic APWPs compared with their FHM predicted APWPs. The best are in dark green and underlined, second best in green and third in light green.

Window, Step (size in Myr)	N America Pk 0						N America Pk 1					
	Wt 0	Wt 1	Wt 2	Wt 3	Wt 4	Wt 5	Wt 0	Wt 1	Wt 2	Wt 3	Wt 4	Wt 5
2, 1	0.2864	0.2925	0.28685	0.280944	0.287	0.29324	<b>0.005759</b>	0.0091586	<b>0.01528</b>	<b>0.0090683</b>	0.03601	<b>0.009034</b>
4, 2	0.10584	0.09228	0.10938	0.11301	0.1131	0.11263	<b>0.00525</b>	0.007665	0.01701	<b>0.009612</b>	0.03504	<b>0.007863</b>
6, 3	0.06486	0.064078	0.077426	0.073621	0.082867	0.075522	<b>0.005418</b>	<b>0.007522</b>	0.017936	<b>0.0097317</b>	<b>0.01983</b>	0.0095782
8, 4	0.06934	0.07168	0.082474	0.0707846	0.11473	0.071686	0.006	<b>0.006217</b>	<b>0.01556</b>	0.01242	0.02163	0.011483
10, 5	0.04412	0.04486	0.04739	0.04212	0.04753	0.04445	0.00591	<b>0.00616</b>	<b>0.01262</b>	0.01208	<b>0.01529</b>	<b>0.00941</b>
12, 6	<b>0.03271</b>	<b>0.02987</b>	<b>0.03024</b>	<b>0.0276143</b>	<b>0.03012</b>	<b>0.02988</b>	0.008736	0.00902	0.01715	0.015665	<b>0.01642</b>	0.01453
14, 7 (119-0 Ma path)	0.0367	0.03695	0.03498	0.0347	<b>0.0319</b>	0.0352	0.0115	0.01198	0.023922	0.01477	0.02205	0.012985
16, 8	0.052386	0.04923	0.04964	0.04653	0.047712	0.048523	0.01138	0.0117	0.02177	0.015409	0.023068	0.01616
20, 10	0.04972	0.05356	0.0562	0.05352	0.0542	0.04902	0.014	0.01728	0.02896	0.0181	0.02702	0.0185
24, 12	<b>0.031642</b>	<b>0.0342</b>	<b>0.03273</b>	<b>0.03469</b>	<b>0.0314765</b>	<b>0.031253</b>	0.0164	0.016737	0.02521	0.01907	0.02397	0.01924
30, 15	<b>0.0345</b>	<b>0.0298</b>	<b>0.0317</b>	<b>0.0307</b>	0.03402	<b>0.0341</b>	0.0206	0.0211	0.0313	0.0272	0.0293	0.0277
Window, Step (size in Myr)	India Pk 0						India Pk 1					
	Wt 0	Wt 1	Wt 2	Wt 3	Wt 4	Wt 5	Wt 0	Wt 1	Wt 2	Wt 3	Wt 4	Wt 5
2, 1	0.25838	0.258757	0.258701	0.275184	0.265442	0.258692	0.05446	0.05639	0.06456	0.06075	0.0677596	0.0583844
4, 2	0.19345	0.20276	0.19344	0.24597	0.203264	0.19423	0.05588	0.056691	0.066803	0.062283	0.067216	0.060121
6, 3	0.17396	0.17476	0.173975	0.229325	0.18514	0.175174	0.057887	0.05882	0.0664754	0.0627534	0.06725	0.06083
8, 4	0.15309	0.14966	0.15321	0.17412	0.172154	0.15124	0.05761	0.05848	0.06671	0.06139	0.0669	0.05941
10, 5	0.1839	0.1845	0.1909	0.2586	0.1972	0.1852	0.0545	0.0554	0.0662	0.0589	0.066	0.06
12, 6	0.108924	0.1035	<b>0.11205</b>	0.11831	0.121497	<b>0.10587</b>	0.059897	0.06068	0.06993	0.064499	0.06951	0.062945
14, 7 (119-0 Ma path)	0.112537	0.112885	0.126554	0.139516	0.132359	0.114195	<b>0.04942</b>	<b>0.0502588</b>	0.0579931	0.060018	0.0654112	0.0582519
16, 8	<b>0.104461</b>	0.104463	0.121002	0.110942	0.119599	0.118336	<b>0.051735</b>	0.052813	<b>0.055188</b>	<b>0.056389</b>	0.0574883	0.055042
20, 10	0.1052	<b>0.1015</b>	0.1198	<b>0.1096</b>	<b>0.1174</b>	0.1166	<b>0.0492</b>	<b>0.0501</b>	0.0585	<b>0.053</b>	<b>0.0536</b>	<b>0.052</b>
24, 12	<b>0.053143</b>	<b>0.05356</b>	<b>0.056986</b>	<b>0.05747</b>	<b>0.05558</b>	<b>0.0553047</b>	0.051926	<b>0.045995</b>	<b>0.04868</b>	<b>0.0557455</b>	<b>0.05698</b>	<b>0.05456</b>
30, 15	<b>0.05617</b>	<b>0.0754578</b>	<b>0.0775947</b>	<b>0.0575459</b>	<b>0.0565421</b>	<b>0.056635</b>	0.0523614	0.0519862	<b>0.054158</b>	0.0563985	<b>0.0555998</b>	<b>0.0543501</b>
Window, Step (size in Myr)	Australia Pk 0						Australia Pk 1					
	Wt 0	Wt 1	Wt 2	Wt 3	Wt 4	Wt 5	Wt 0	Wt 1	Wt 2	Wt 3	Wt 4	Wt 5
2, 1	0.471554	0.529417	0.52834	0.563622	0.628921	0.55071	0.00498465	0.0047458	0.0247455	0.0072776	0.035936	<b>0.0039155</b>
4, 2	0.26822	0.29862	0.25881	0.276464	0.33741	0.268965	0.004543	0.004578	0.023895	0.005417	0.032511	<b>0.0037674</b>
6, 3	0.090985	0.0947676	0.09448	0.08977	0.10083	0.091376	<b>0.0040955</b>	0.004847	0.02199	<b>0.0046074</b>	0.023489	0.0064495
8, 4	0.0870445	0.097201	0.057284	0.086078	0.06245	0.04463	0.004265	<b>0.004191</b>	0.026156	0.0067075	0.02533	0.00757145
10, 5	0.0315	0.039	<b>0.0509</b>	0.0305	0.0601	0.031	0.0045	0.0048	0.0199	0.0058	0.0288	0.0089
12, 6	0.027348	<b>0.026942</b>	0.057592	0.026687	0.055426	<b>0.0270495</b>	0.004341	0.0049061	0.020529	0.007813	0.020667	0.0054794
14, 7 (119-0 Ma path)	0.02725	0.033687	<b>0.046242</b>	<b>0.0252737</b>	<b>0.04651</b>	<b>0.027222</b>	<b>0.0017226</b>	<b>0.00354029</b>	<b>0.0119085</b>	<b>0.00157378</b>	0.0121774	0.00765843
16, 8	0.028261	0.037993	0.0536856	0.0278993	0.05263	<b>0.0282745</b>	0.0050338	<b>0.0037223</b>	0.018136	0.0047548	0.014537	0.01061
20, 10	<b>0.0257</b>	0.0378	0.0616	0.0254	0.0526	0.0373	0.00544	0.012	0.023	0.0162	<b>0.0119</b>	0.0137
24, 12	<b>0.0219624</b>	<b>0.0214394</b>	<b>0.043667</b>	<b>0.0216832</b>	<b>0.0372685</b>	0.0345906	0.0058473	0.0063788	<b>0.0059937</b>	0.0125858	<b>0.002742</b>	0.0051643
30, 15	<b>0.014614</b>	<b>0.0183819</b>	0.0786527	<b>0.014254</b>	<b>0.031029</b>	0.0293416	<b>0.00412276</b>	0.00444694	<b>0.00448627</b>	<b>0.00397289</b>	<b>0.0054747</b>	<b>0.00387337</b>

are “good”. AMP only regards the time uncertainty of each pole as one mid-point. Then this mid-point is treated as the most likely age of that mean pole. This is actually incorrect. The age uncertainty of paleopole is not obtained from a probability density function derived from an observed frequency distribution. As defined, the time uncertainty's lower (older) limit is a stratigraphic age, and its upper (younger) limit could be also a stratigraphic age or be constrained by a tectonic event using the field tests (e.g. fold/tilt test and conglomerate test). So the true age of the pole could be any one that is not older than the lower limit and also not younger than the upper limit. In other words, the mid-point could be the true age of the pole, but it is not known as the most likely age of that pole. If the mid-point is the most likely age of a pole, AMP should gen-

erate a path that is closer to the reference. However, mostly APP generates better similarities (See the high proportions of APP better than AMP in Table 6). Most reasonably, the mid-point should be regarded as one possibility of all uniformly (not necessarily normally bell shaped, or U shaped, or left or right skewed) distributed ages between the two time limits.

So APP remains the effect of a paleopole borne on the mean poles during all the period of its age uncertainty, and use the increased number of paleopoles (N) to average out the negative effect of those “bad” poles, including the paleopoles that should not be included at that age for mean pole.

**Table 6.** 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. 9a	1, 5, 7, 11, 13, 15, <b>19, 21, 25</b>	0, 1, 3, 5	2, 5, 7, 14, <b>16, 17</b> , 18, 22, 26	0, 1, 2, 3, 4, 5	41/42	<b>9</b>	4	4	7	0	5	Y/Y
	Fig. 9b	4, 5, 6, 7, 9, <b>19</b> , <b>21, 22, 26</b>	0, 1, 2, 3, 4, 5	0, 2, 8, 10, 12, <b>16, 18</b> , 20, 24	0, 1, 2, 3, 4, 5	18/21	<b>15</b>	2	2	6	2	1	Y/Y
	Fig. 9c	1, 11, 13, 17, <b>19</b> , <b>21, 25</b>	0, 1, 3, 5	2, 4, 14, <b>16, 22, 26</b>	0, 1, 2, 3, 4, 5	1	<b>10</b>	5	1	7	2	3	N/N
	Fig. 17a	1, 4, 5, 6, 7, 9, 11, 13, 15, <b>19, 21, 25</b>	0, 1, 3, 5	0, 8, 10, 12, 14, <b>16</b> , 18, 20, 24	0, 1, 2, 3, 4, 5	61/84	<b>14</b>	2	0	3	3	6	N/Y
	Fig. 17b	4, 5, 6, 7, <b>19</b>	0, 1, 2, 3, 4, 5	0, 2, 10, 12, <b>16, 18</b> , 20, 23, 24, 27	0, 1, 2, 3, 4, 5	29/42	<b>11</b>	6	0	9	1	1	Y/4y2n
	Fig. 17c	1, 11, 13, 17, <b>19</b> , <b>21, 23, 25, 27</b>	0, 1, 3, 4, 5	2, 14, <b>16, 22, 26</b>	0, 1, 2, 3, 4, 5	1	9	6	0	<b>10</b>	1	2	N/N
	Fig. 21a	1, 5, 7, 9, 11, 13, <b>15, 19, 21, 25</b>	0, 1, 2, 3, 4, 5	0, 10, 12, 14, <b>16</b> , 17, 18, 20, 24	0, 1, 2, 3, 4, 5	41/42	<b>10</b>	9	3	4	0	2	Y/Y
MHM	Fig. 21b	5, 7, <b>19, 21</b> , 22, 26	0, 1, 2, 3, 4, 5	0, 2, 8, 10, 12, 14, <b>16</b> , 18, 20, 24	0, 1, 2, 3, 4, 5	6/7	<b>12</b>	2	0	7	4	3	Y/Y
	Fig. 21c	1, 11, 13, 17, <b>19</b> , <b>21, 25</b>	0, 1, 2, 3, 5	2, 14, <b>16, 22, 26</b>	0, 1, 2, 3, 4, 5	83/84	6	4	2	<b>10</b>	4	2	N/N
	Fig. 23a	5, 7, 15, 22, 26	0, 1, 2, 3, 4, 5	0, 8, 10, 12, 14, 18, 20, 24	0, 1, 2, 3, 4, 5	16/21	<b>10</b>	8	1	4	1	4	5n1y/Y
	Fig. 23b	4, 5, 6, 7, 22, 26	0, 1, 2, 3, 4, 5	0, 2, 10, 12, <b>16, 18</b> , 24	0, 1, 2, 3, 4, 5	59/84	6	7	0	6	3	6	Y/3n3y
	Fig. 23c	1, 11, 13, 15, 17, <b>19, 21</b> , 23, 25, 27	0, 1, 2, 3, 4, 5	2, 14, <b>16, 22, 26</b>	0, 1, 2, 3, 4, 5	1	<b>12</b>	1	0	10	4	1	N/N

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

Because of small number of paleopoles (not necessarily “bad”) involved in each sliding window, the produced mean poles by AMP should be relatively far from its contemporary model-predicted pole. In other words, AMP intends to give fairly small change in accuracy. This also could potentially bring more distinguishable  $d_s$  for AMP if the corresponding A95 is not large enough. For example, for Fig. 9a, there are only two special (of 84 APP vs. AMP comparisons) cases picking/weighting 4/3, 6/3 better than 5/3, 7/3 respectively. Compared with the picking/weighting 4/3 APWP, although most of the mean paleopoles are closer to the FHM predicted APWP and also the number of the significant pole pairs is two less for the APP derived path (i.e. 5/3), the A95s are smaller and most importantly there are one more significant  $d_a$  orientation-change pair and one more significant  $d_l$  segment pair (Table 7). If we observe carefully, it is because of the much smaller 15 Ma A95 for 5/3. The similar phenomenon occurs to the case of 6/3 vs 7/3, a relatively much smaller paleomagnetic A95 causes more distinguishable  $d_a$  and  $d_l$  for the APP results, and they offset the improvement of spatial similarity  $d_s$  APP brings.

For 2/0 vs 3/0 for 20/10 Myr window/step North America, all their  $d_a$  and  $d_l$  are indistinguishable. Compared with the results from AMP, although the coeval pole GCDs are generally unchanged or decreased or even increased (but not too much) for APP, this spatial improvement is not able to offset the negative effects of also generally unchanged or decreased or even increased (but not too much) paleomagnetic A95s, which potentially brings more statistically distinguishable coeval poles (e.g. the 20 Ma and 110 Ma poles for picking 3 and weighting 0; Table 8). This further causes greater distinguishable mean  $d_s$  from the APP methods. The similar phenomenon occurs to Fig. 17a picking 2 vs 3 with weightings 2, 3 and 5, picking 4 vs 5 with weightings 1, 3 and 5, and 6 vs 7 with weightings 1, 3 and 5, and so on.

In addition, compared with AMP, APP potentially could generate more mean poles, because sometimes for some sliding window there is no paleopole involved at all for AMP while there

is(are) paleopole(s) involved for APP. For APP, the mean poles at all ages should be composed of more paleopoles than it is for AMP, which should generally decrease both coeval pole distance and paleomagnetic A95. However, sometimes a rare case (e.g. the 0 Ma comparison shown in Table 9) happens. It is sometimes that an additional very “bad” paleopole gets included by APP and this increases both coeval pole distance and paleomagnetic A95 even though N increases. Such cases include Fig. 9b picking 22 vs 23 (actually exactly the same as picking 26 vs 27) with all the six types of weightings.

So generally as we discussed in the last section APP decreases the distances between paleomagnetic APWPs and the hotspot and ocean-floor spreading model predicted APWP, and also the uncertainties of paleomagnetic APWPs. However, as we described in this section, special cases like decreased A95 potentially intends to make coeval poles differentiated if the coeval poles’ distance is not decreased effectively or even increased, or very “bad” paleopoles got involved in some sliding windows, occurs. In summary, when the negative effect from these types of rare cases is beyond the positive effect the generally improved mean poles contribute, the composite difference score would increase. However, this phenomenon seldom occur (Table 6).

Other Type 1 (e.g. Table 7) cases: Fig. 17a picking/weighting 2/1 vs 3/1, Fig. 21a 4/3 vs 5/3, 6/3 vs 7/3.

Other Type 2 (e.g. Table 8) cases: Fig. 17a picking/weighting 2/0, 2, 3, 5 vs 3/0, 2, 3, 5, 4/1, 3, 5 vs 5/1, 3, 5, 6/1, 3, 5 vs 7/1, 3, 5. Fig. 17b 4/0–5 vs 5/0–5, 6/0–5 vs 7/0–5, 14/2, 3 vs 15/2, 3, 22/0, 2, 3 vs 23/0, 2, 3, 26/0, 2, 3 vs 27/0, 2, 3. Fig. 17c 4/2 vs 5/2. Fig. 21c 8/5 vs 9/5. Fig. 23a 2/0–5 vs 3/0–5, 4/2 vs 5/2, 6/2 vs 7/2, 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5. Fig. 23b 4/0–5 vs 5/0–5, 6/0–5 vs 7/0–5, 14/3 vs 15/3.

Combined Type 1 and 2 cases: Fig. 17a picking/weighting 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5. Fig. 17b picking/weighting 22/1, 4, 5 vs 23/1, 4, 5, 26/1, 4, 5 vs 27/1, 4, 5. Fig. 23b 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5.

Other Type 3 (e.g. Table 9) cases: Fig. 21a 4/2–5 vs 5/2–5. Fig. 21b 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5.

**Table 7.** One example of the Type 1 rare cases where AMP gives better similarity result than APP does from North America (101; Window size: 10 Myr, step size: 5 Myr). Only statistically significant values are listed here.

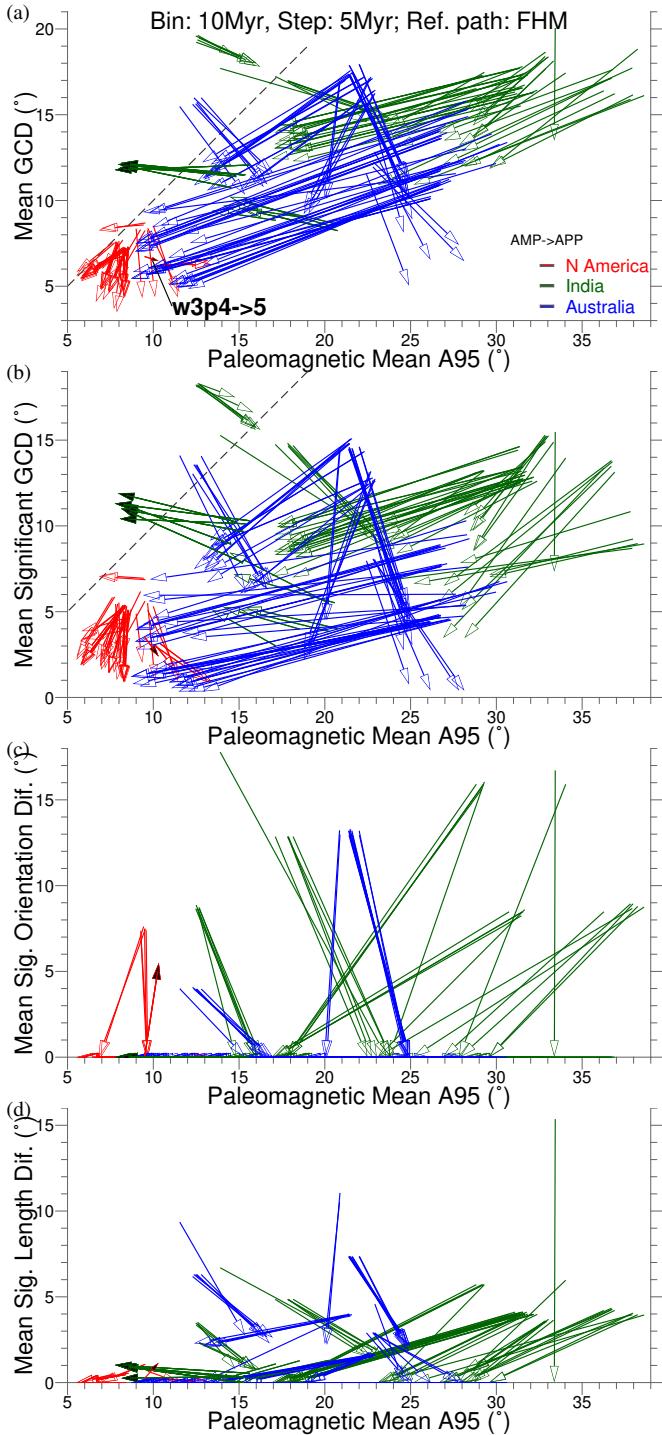
FHM predicted		picking 4 + weighting 3						picking 5 + weighting 3							
Age (Ma)	A95 (°)	Age (Ma)	Pmag A95 (°)	Dist (°)	Age (Ma)	Diff (°)	Age (Ma)	Pmag A95 (°)	Dist (°)	Age (Ma)	Diff (°)	ds	dl	ds	da
10	1.44607/0.793714	10	14.876819	<b>9.937</b>	105-110	5.91855								<b>10-15-20</b>	<b>126.59</b>
15	1.2875/0.816514						15	2.0857	<b>11.805</b>						
25	2.48031/1.10915						25	6.3358	<b>6.873</b>						
55	3.58782/2.14032	55	4.6347	<b>5.372</b>										Age (Ma)	Diff (°)
60	4.85938/3.17602						60	6.5922	6.215	<b>10-15</b>	<b>13.52</b>				
65	3.68984/2.30014						65	8.6632	<b>7.6</b>	<b>15-20</b>	<b>14.68</b>				
75	2.6435/1.54052	75	9.0812	<b>8.836</b>											
100	2.8983/2.68346	100	8.892	<b>8.455</b>											
105	2.32328/1.74639	105	5.3	<b>5.03</b>											
110	4.13015/2.25964	110	3.8	<b>9.8064</b>											
115	4.63512/2.58006	115	19.6676	<b>9.3345</b>											
120	7.34408/4.06043	120	3.515	<b>17.35</b>											

**Table 8.** One example of the Type 2 rare cases where AMP gives better similarity result than APP does from North America (101; Window size: 20 Myr, step size: 10 Myr). Only statistically significant values are listed here.

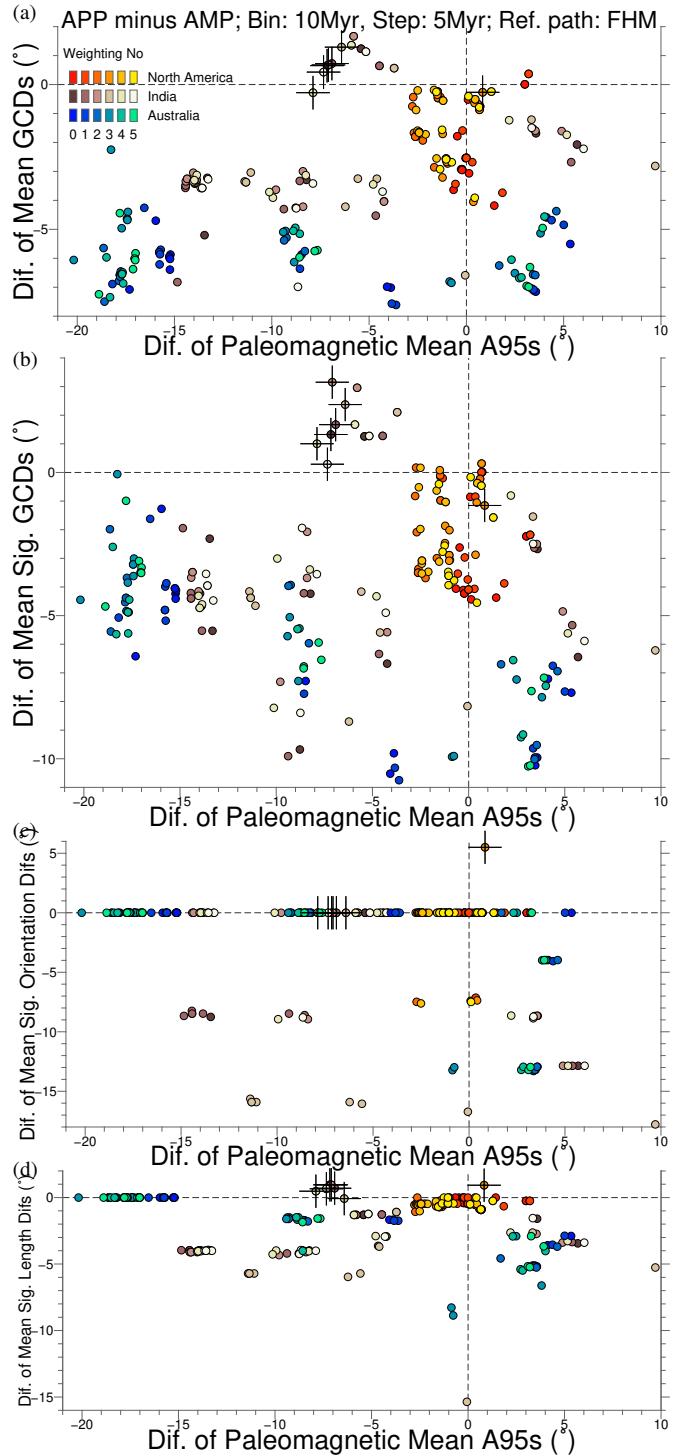
Age (Ma)	FHM predicted	ds			
		picking 2 + weighting 0	picking 3 + weighting 0	Pmag A95 (°)	Dist (°)
0	0	3.97	<b>5.714</b>	3.97	<b>5.714</b>
10	1.44607/0.793714	3.879	<b>6.034</b>	3.879	<b>6.034</b>
20	1.58039/1.10047			6.771	<b>6.934</b>
50	3.57782/1.61328	3.8644	<b>7.304</b>	4.03	<b>8.6</b>
60	4.85938/3.17602	5.716	<b>8.457</b>	5.55	<b>7.367</b>
100	2.8983/2.68346	10.769	<b>7.308</b>	10.769	<b>7.308</b>
110	4.13015/2.25964			3.29	<b>8.311</b>
120	7.34408/4.06043	3.38	<b>16.41</b>	3.083	<b>16.728</b>

**Table 9.** One example of the Type 3 rare cases where AMP gives better similarity result than APP does from India (501; Window size: 10 Myr, step size: 5 Myr). Only statistically significant values are listed here. Note that for the bold-number ages, there is no mean poles at all for the “picking 22 (AMP) + weighting 2” case.

FHM predicted		picking 22 + weighting 2				picking 23 + weighting 2				ds	dl
Age (Ma)	A95 (°)	Pmag A95 (°)	Dist (°)	N	Pmag A95 (°)	Dist (°)	N	Age (Ma)	Diff (°)		
0	0	<b>6.28</b>	<b>12.72</b>	2	<b>23.54</b>	<b>18.14</b>	3	<b>80-85</b>	6.286		
10	1.12124/0.673225	5.4/3.1	29.9	1	5.4/3.1	29.9	1	<b>110-115</b>	16.684		
<b>15</b>	1.1347/0.8127				5.4/3.1	28.28	1				
<b>60</b>	4.79687/3.07133				8.817	8.28	20				
<b>70</b>	4.26508/2.48783				3.26	4.464	20				
<b>75</b>	2.6777/1.57975				5	4.477	1				
<b>80</b>	4.20828/2.50294				5	3.358	1				
85	2.50744/1.24746	5	7.632	1	5	7.632	1				
<b>90</b>	3.88998/1.43423				5	10.884	1				
<b>95</b>	2.23389/1.6247				5	11.099	1				
<b>100</b>	2.8062/2.59819				5	11.4155	1				
<b>105</b>	2.32328/1.74639				5	14.908	1				
<b>110</b>	4.55519/2.49218				6.8/4.9	13.962	1				
115	4.63512/2.58006	10.73	10.508	5	10.73	10.508	5				
120	6.02639/3.3319	10.73	10.508	5	10.73	10.508	5				



**Figure 26.** Paleomagnetic APWP's 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. Arrowtails are the results from AMP, while arrowheads are from APP. Black color filled arrowheads are the small number of special cases of AMP derived equal-weight CPDs better than APP (see details in Fig. 9 and Table 6).



**Figure 27.** Differences of APP and AMP coordinates shown in Fig. 26. Crosses locates the small minority cases of AMP derived equal-weight CPDs better than APP (see details in Fig. 9 and Table 6).

#### 4.0.3 Question: Why weighting is not affecting?

Generally, weighting does not affect the similarities dramatically, because the six results from the six weighting methods are mostly very close to each other (Fig. 9, Fig. 17, Fig. 21, Fig. 23). This closeness is also generally observed in the form of clusters in Fig. 28 and Fig. 29. In addition, from the general statistics of performance of the six weighting methods shown in Table 6, Weighting

No. 0 mostly performs the best or at least the second best, which means no weighting works better in general.

When the above-mentioned question, about why APP generally produces better fits than AMP, is tackled, we already find that both accuracy (how closely do the pairs/segments/angles match) and precision (how large are the uncertainties on the pairs/segments/angles, i.e. how difficult do they distinguish) can be the factors that finally determine the difference score. Although another factor, resolution (how many pairs/segments/angles are actually being compared) can also influence the difference score (e.g. Table 5), here this factor is not relevant to comparisons between different weightings for a certain picking method, because the numbers of picked paleopoles are the same for the six weighting methods.

Therefore, at the very basic level, for example, a lower score is the result of one of, or combination of:

1. A reduction in the difference scores of significantly different pairs/segments, straightforwardly interpreted as a better fit (improved accuracy).

2. Previously significantly different pairs/segments becoming insignificant. This can occur either because the fit is better (improved accuracy), or because of an increase in the uncertainty of the mean poles (they become less distinguishable - decreased precision).

Fig. 30 and Fig. 31 show that the proportions of results from Weighting No 1–5 to result from 0 are generally in the second quadrant of the coordinate plane, where proportional change in difference score is positive whereas proportional change in paleomagnetic FQ score is negative. This means the five weightings (Wt 1–5) do make effects, not obviously in accuracy but mainly in improving precision, which could potentially expose more pairs of distinguishable poles/segments/angle-changes. Or even accuracy is improved in a small amount, improved precision intends to cancel out the effects from improved accuracy.

There are a few special cases that one or two of the six weighting methods gives a result with a dramatically worsened difference score, e.g. for weighting method (Wt) 3 (Fig. 9). From the labeled dots of Wt 3 in Fig. 28, Fig. 29, Fig. 30 and Fig. 31, we can get a general impression that weighting 3 is indeed improving precision but not accuracy (at least not enough to offset the effects from improved precision) so that this precision improvement is also the culprit that worsens the final score. Highly improved precision without corresponding improved accuracy would potentially bring more significant differences in shape metrics. Wt 3 is exactly performing in this form.

In addition, for Wt 3, a small size of  $\alpha_{95}$  (high precision) could be caused by those sampled directions not covering enough long period (thought to be at least about  $10^4$  years) to “average out” secular variation for giving a paleopole. That is to say, the smallest  $\alpha_{95}s$  could get the greatest weights that they should not deserve.

Generally, weighting is affecting because different weighting functions give obviously different results. However, interestingly weighting does not improve fit and generally no weighting (Wt 0) is giving the best fit, although in most cases weighting does improve precision. Wt 2 or 4 is not recommended, because they never have generated the best similarities (Table 6), compared with other weighting methods. There is no general pattern about which weighting (of Wt 1–5) is better or worse. So weighting, for making a paleomagnetic APWP, is not absolutely necessary. However, there are some patterns about which weighting is better or worse for some specific continent or some specific picking methods. For

example, Wt 3 works generally fine with Australian data (Table 6). However, Wt 3 is not recommended for North America and India.

#### 4.0.4 Question: Why best and worst methods are not consistent?

For all the three continents, North America, India and Australia, picking methods 19 (APP with local rotation or secondary print excluded) and 21 (APP with local rotation and secondary print corrected) are consistently the best or at least the relatively better (for example, in Fig. 17b, picking 21 results are not colored in green, but ranging from 0.0483–0.0561 that is still much less than the mean, 0.074 and also the median, 0.072); whereas picking methods 16 (AMP with publications before 1983) and 18 (AMP with local rotation or secondary print excluded) are consistently the worst or at least the relatively worse.

Nevertheless, for each single continent, they have their own consistently best and worst picking methods that are not the best or worst for other continents. For example, picking method 15 (APP with publications after 1983) works well with North America. We know that 70 North American paleopoles (about 53%) have contributed to picking method 15 for 120–0 Ma. However, only 28 Indian paleopoles (about 38%) and 29 Australian paleopoles (about 30%) have contributed to method 15. In contrast, the picking method 17 (APP with publications before 1983) works well with only Australia. This also tells that number of paleopoles is a key factor that affects the fit score. The fact that picking method 22 (AMP with SS05 criteria) doesn't work well with only Australia also reflects the importance of this factor.

In summary, the reason why some best and worst methods are not consistent is generally that different continents have their unique data sets and get their own paleomagnetic studies in varying degrees.

#### 4.0.5 Question: Are there particular parts of the path that are more variable? Do different methods affect different parts of the path differently?

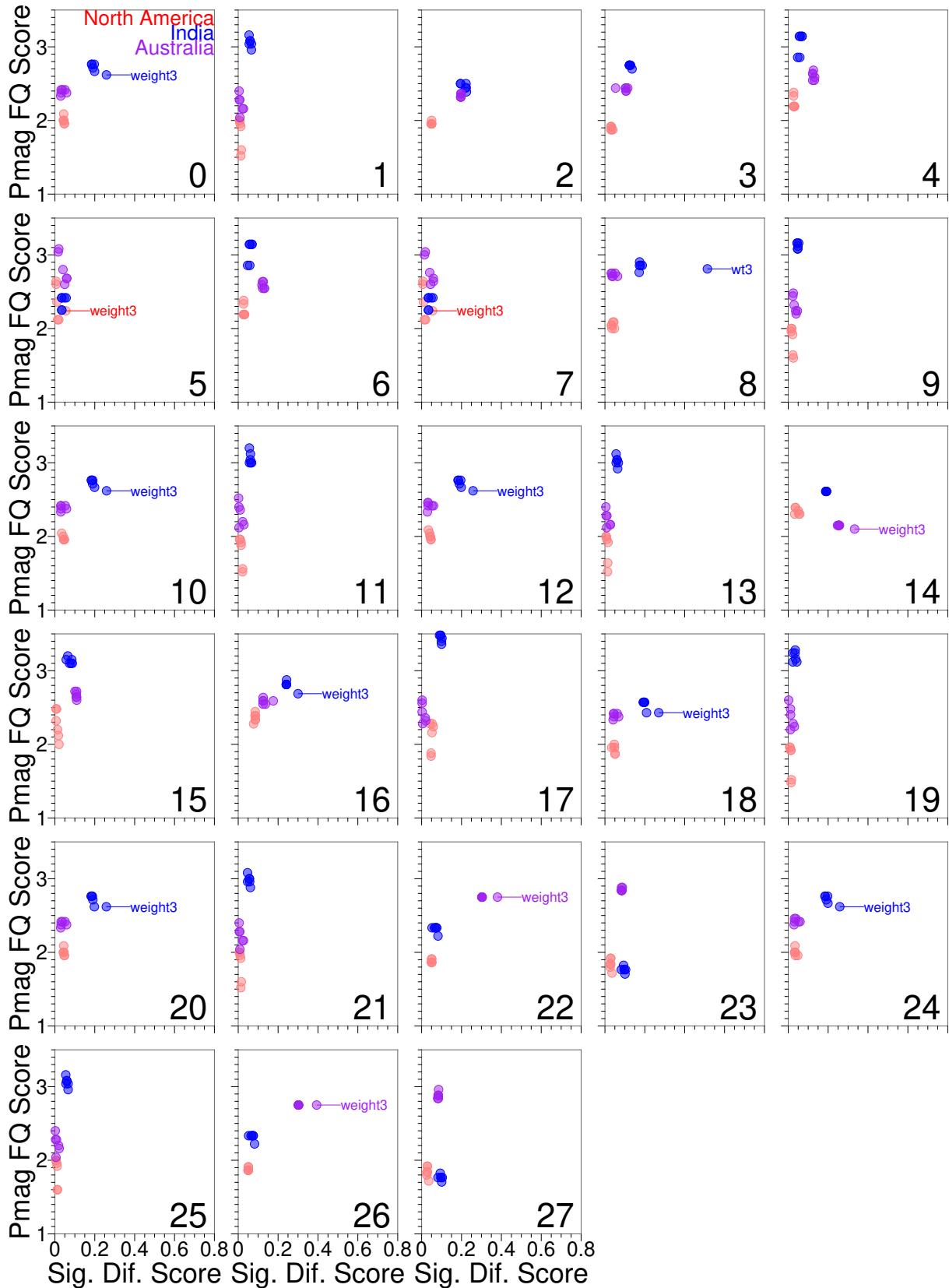
The results may highlight the trade-off between more data diluting the effect of outliers, and fewer but ‘better’ data being more easily affected by a bad point that gets through the filters (Fig. 14, Fig. 15 and Fig. 16).

## 5 FINAL CONCLUSIONS

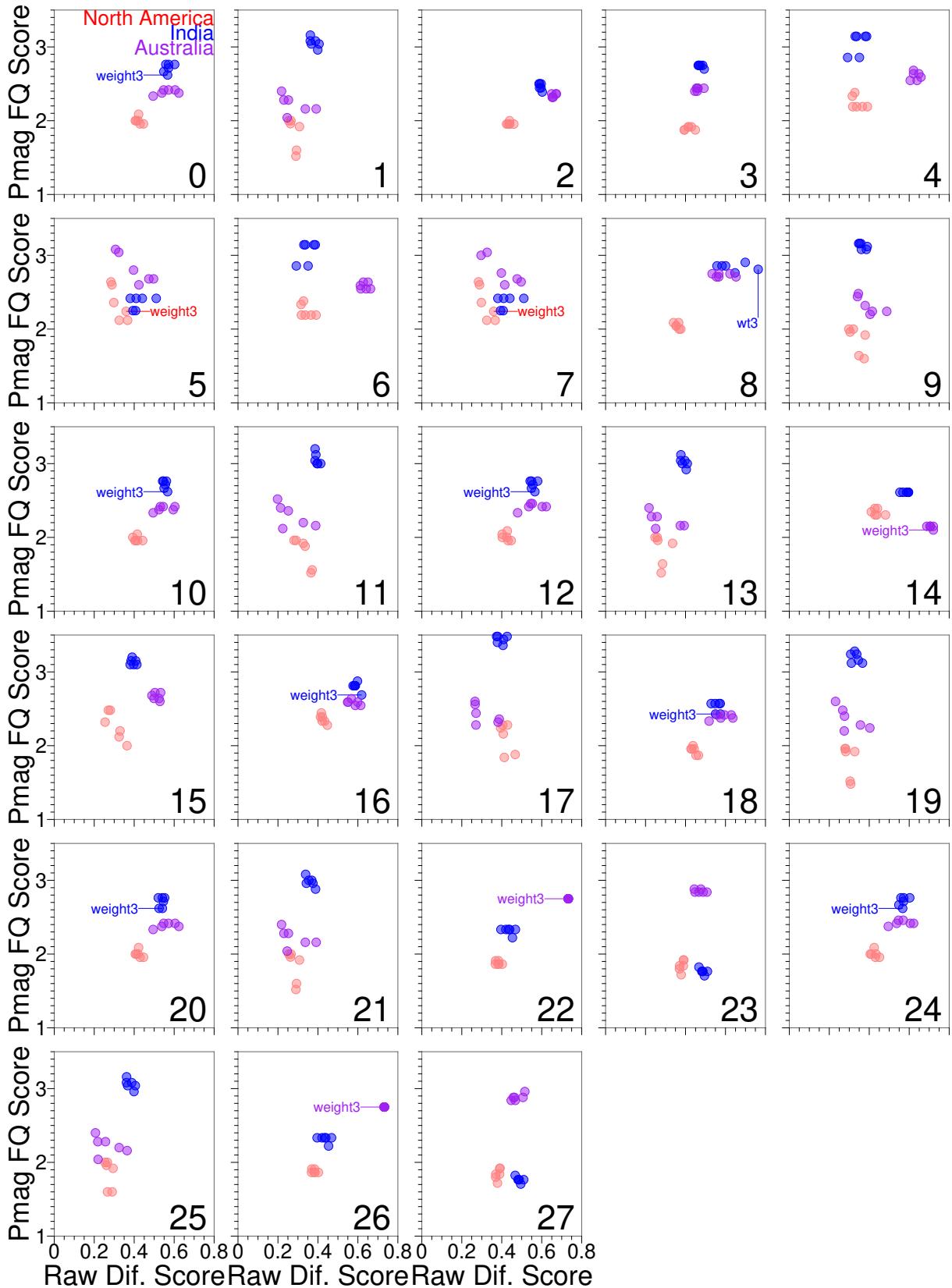
From the perspective of the general similarities between those paleomagnetic APWPs and the hot spot model and ocean-floor spreading model predicted APWPs, GAD hypothesis is proved valid for at least the last 120 Myr.

### 5.1 Universal Rules of Ways of Processing Paleomagnetic Data:

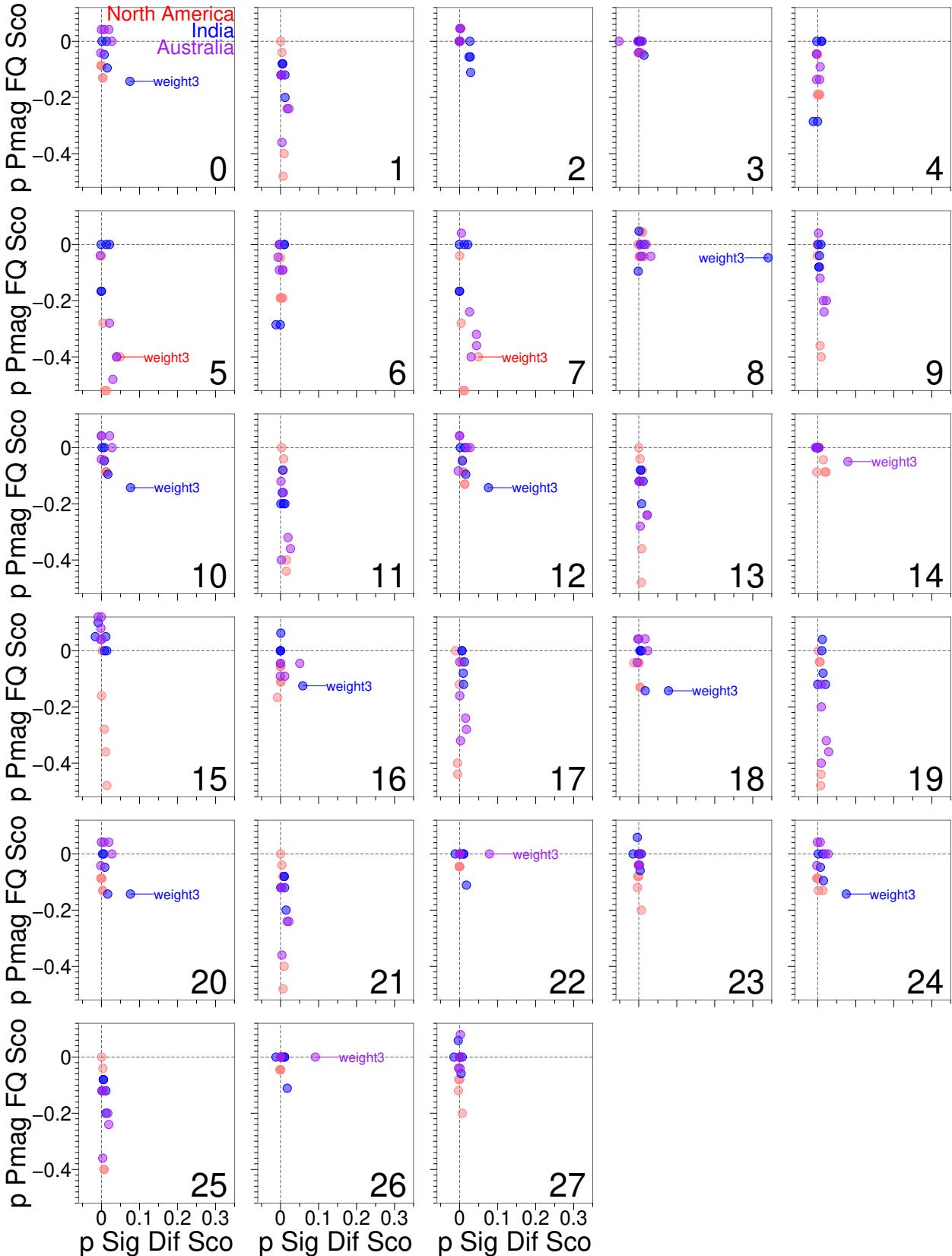
Although effects of filters (all the picking methods where number of paleopoles shrinks compared to Pt 0 and Pt 1; see Fig. 9) have a marginal change in reducing N (precision potentially going down), some filters do improve the similarity score, for example, Pt 25 (APP without superseded data) is always giving better scores than Pt 1 for all the three continents. However, Pt 0 and Pt 1 (no filtering and corrections applied) still generally perform well.



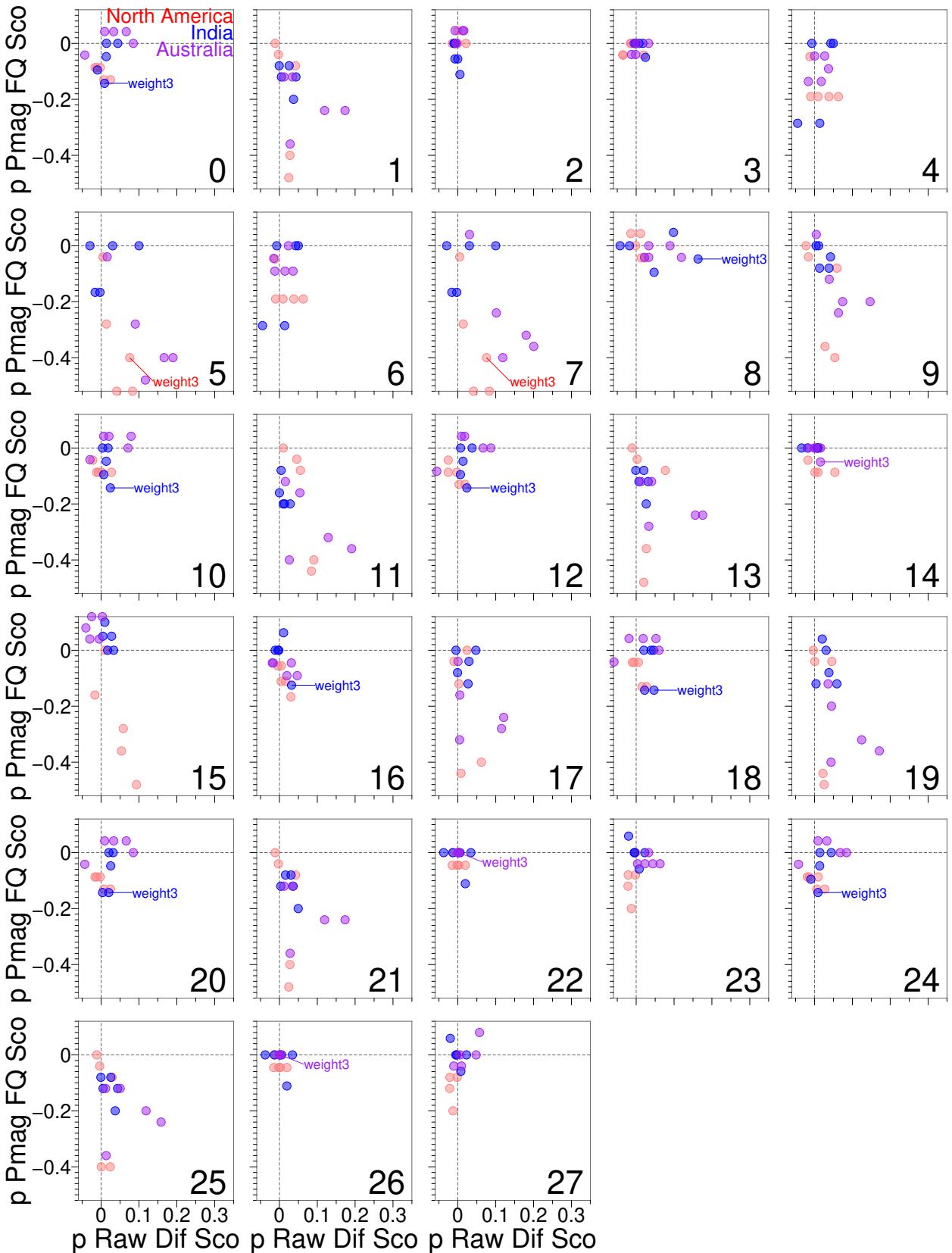
**Figure 28.** 10/5 Myr bin/step paleomagnetic APWP's FQ score (different from FQ, see the definitions of FQ and FQ score in Chapter 2) versus significant  $\mathcal{CPD}$  score (reference path: FHM predicted) for the 28 different picking methods. See the significant  $\mathcal{CPD}$  scores and Ppath-Rpath FQ in Fig. 9. Only those results dramatically worsened by weighting 3 are labeled.



**Figure 29.** 10/5 Myr bin/step paleomagnetic APWP's FQ score (different from FQ, see the definitions of FQ and FQ score in Chapter 2) versus raw difference score (reference path: FHM predicted) for the 28 different picking methods. See FQ vs significant difference score in Fig. 28. Only those results dramatically worsened by weighting 3 are labeled.



**Figure 30.** Proportion of weighting 1–5 to 0: Proportional change in 10/5 Myr bin/step paleomagnetic APWP's FQ score (different from FQ, see the definitions of FQ and FQ score in Chapter 2) versus proportional change in significant  $\mathcal{CPD}$  score (reference path: FHM predicted) for the 28 different picking methods. See the significant  $\mathcal{CPD}$  scores and  $P_{path-R_{path}}$  FQ in Fig. 9. Only those results dramatically worsened by weighting 3 are labeled.



**Figure 31.** Proportion of weighting 1–5 to 0: Proportional change in 10/5 Myr bin/step paleomagnetic APWP's FQ score (different from FQ, see the definitions of FQ and FQ score in Chapter 2) versus proportional change in raw difference score (reference path: FHM predicted) for the 28 different picking methods. See FQ vs significant difference score in Fig. 30. Only those results dramatically worsened by weighting 3 are labeled.

APP (adding data to a time window with overlapping age selection criterion) is better than AMP for making paleomagnetic APWPs, for both kinds of situations when there are lots of data (APP even better, e.g. for North America and Australia) and not much data (APP still a better option, e.g. for India; Table 6). APP with most filters/corrections (Pt 3, 5, 7, 9, 11, ..., 27) are generally giving worse scores than APP without any filter/correction (Pt 1).

In most cases the APP methods produce better similarities than the AMP methods (Table 6).

Actually weighting is not improving the fit but improving precision generally. For quite many of the methods, no weighting is the best performer (Table 6). For example, score is likely worse for the combined methods of weight method 3 and AMP.

APP itself helps incorporate temporal uncertainty into the algorithm. With the bootstraps test helping incorporate spatial uncertainty into the algorithm together, both spatial and temporal uncertainties are successfully considered in APP methods.

## 5.2 Conditional Rules of Ways of Processing Paleomagnetic Data:

Picking methods no. 16 (AMP with data from old studies) and 18 (AMP without data affected by rotations and secondary print) are not recommended for generating a paleomagnetic APWP.

## 5.3 Summary

According to the results we have from the three continents, North America, India and Australia, using the similarity measuring tool developed in Chapter 2, it is recommended that APP should be used to select the input paleopoles. According to all the paleomagnetic data we have from the three continents, the results from any size for sliding window and step are interestingly and extremely close to each other when the APP method is used, compared with the results from the AMP method (Table 5 and Fig. 20). So any size for binning and stepping is ok when APP is used. Then filtering is actually not necessary. However, some filtering methods (e.g. Picking No. 5, 7 (igneous-derived), 11, 13 (nonredbeds or corrected redbeds derived), 19, 21 (non-local-rotation/reprinted or corrected-rotation derived) and 25 (non-superseded data derived)) are fine too and will not give worst or worse results than the other filtering methods (i.e. the left Picking methods). With APP used, weighting is actually not necessary either. If a weighting has to be used, Weighting No. 1 (related to the number of paleomagnetic sampling sites and observations) is generally better than the other four given weighting methods (Fig. 32).

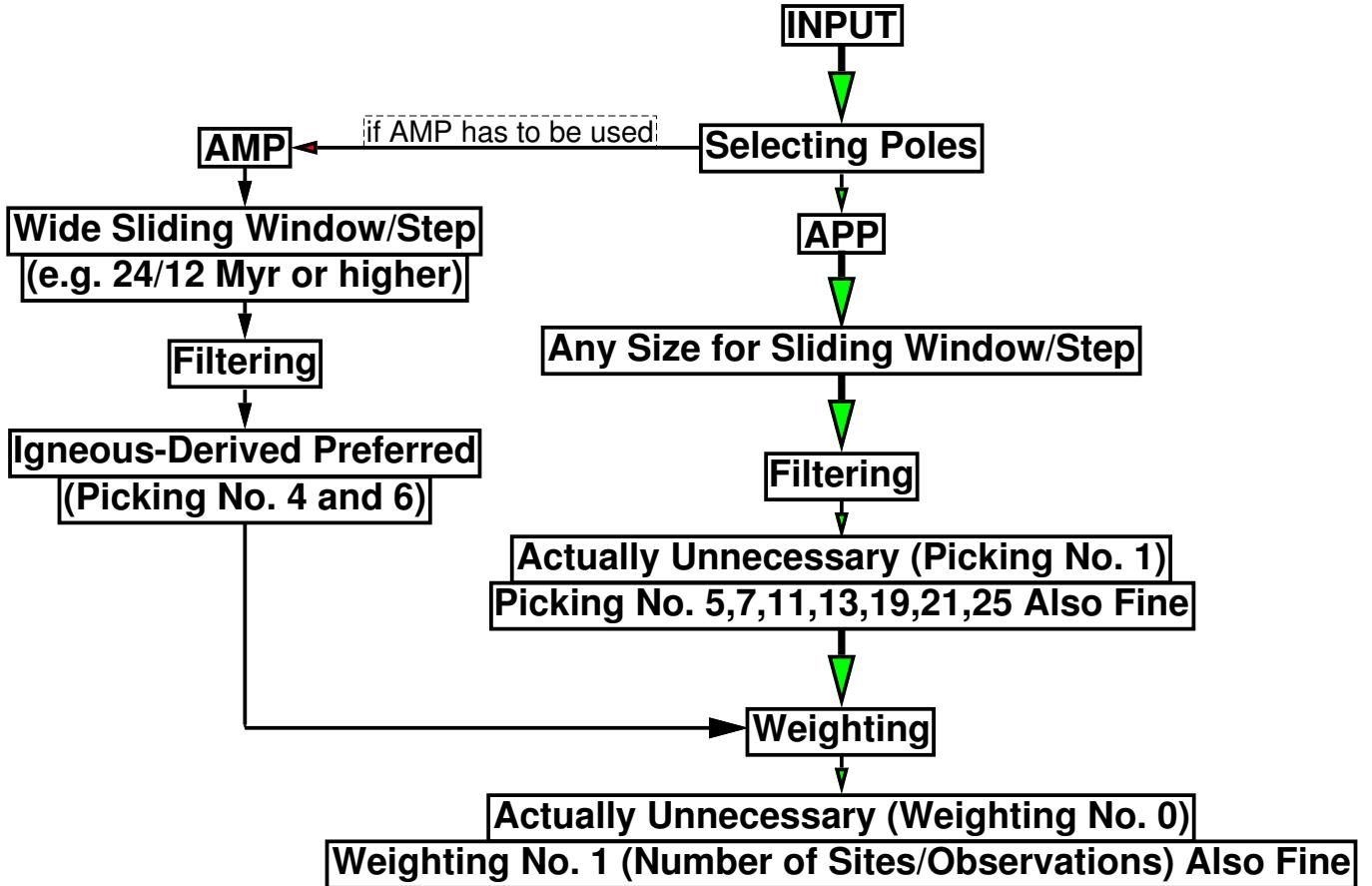
If AMP has to be used, relatively wide sliding window and step are needed. According to our tests, more than 20/10 Myr is recommended. In addition, AMP works relatively better with igneous-derived data (i.e. Picking No. 4 and 6), which indicates that if we have fewer data, these data need to be better in quality (Fig. 32).

## 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.
- Bilardello, D. & Kodama, K. P., 2010. Palaeomagnetism and magnetic anisotropy of Carboniferous red beds from the Maritime Provinces of Canada: evidence for shallow palaeomagnetic inclinations and implications for North American apparent polar wander, *Geophys. J Int.*, **180**, 1013–1029.
- Cottrell, R. D. & Tarduno, J. A., 2000. Late Cretaceous True Polar Wander: Not So Fast, *Science*, **288**, 2283.
- Cande, S. C. & Stock, J. M., 2004. Pacific–Antarctic–Australia motion and the formation of the Macquarie Plate, *Geophys. J Int.*, **157**, 399–414.
- Cande, S. C., Patriat, P. & Dymnt, J., 2010. Motion between the Indian, Antarctic and African plates in the early Cenozoic, *Geophys. J Int.*, **183**, 127–149.
- Christeson, G. L., Van Avendonk, H. J. A., Norton, I. O., Snedden, J. W., Eddy, D. R., Karner, G. D. & Johnson, C. A., 2014. Deep crustal structure in the eastern Gulf of Mexico, *J. Geophys. Res. Solid Earth*, **401**, 183–195.
- Doubrovine, P. V. & Tarduno, J. A., 2008. Linking the Late Cretaceous to Paleogene Pacific plate and the Atlantic bordering continents using plate circuits and paleomagnetic data, *J. Geophys. Res. Solid Earth*, **113**, B07104.
- DeMets, C., Gordon, R. G. & Argus, D. F., 2010. Geologically current plate motions, *Geophys. J Int.*, **181**, 1–80.
- Domeier, M., van der Voo, R. & Denny, F. B., 2011. Widespread inclination shallowing in Permian and Triassic paleomagnetic data from Laurentia: Support from new paleomagnetic data from Middle Permian shallow intrusions in southern Illinois (USA) and virtual geomagnetic pole distributions, *Tectonophysics*, **511**, 38–52.
- Domeier, M., Van der Voo, R., Tomezzoli, R. N., Tohver, E., Hendriks, B. W. H., Torsvik, T. H., Vizan, H. & Dominguez, A., 2011. Support for an “A-type” Pangea reconstruction from high-fidelity Late Permian and Early to Middle Triassic paleomagnetic data from Argentina, *Journal of Geophysical Research: Solid Earth*, **116**, B12114.
- DeMets, C., Calais, E. & Merkouriev, S., 2017. Reconciling geodetic and geological estimates of recent plate motion across the Southwest Indian Ridge, *Geophys. J Int.*, **208**, 118–133.
- Evans, D. A. D., 2003. True polar wander and supercontinents, *Tectonophysics*, **362**, 303–320.
- Fisher, R. A., 1953. Dispersion on a sphere, *Proc. Roy. Soc. London Ser. A.*, **217**, 295–305.
- Gaina, C., Torsvik, T. H., van Hinsbergen D. J. J., Medvedev, S., Werner, S. C. & Labails, C., 2013. The African Plate: A history of oceanic crust accretion and subduction since the Jurassic, *Tectonophysics*, **604**, 4–25.
- Gaina, C., van Hinsbergen D. J. J. & Spakman, W., 2015. Tectonic interactions between India and Arabia since the Jurassic reconstructed from marine geophysics, ophiolite geology, and seismic tomography, *Tectonics*, **34**, 875–906.
- Granot, R. & Dymnt, J., 2018. Late Cenozoic unification of East and West Antarctica, *Nature Communications*, **9**, 3189.
- Horner-Johnson, B. C., Gordon, R. G., Cowles, S. M. & Argus, D. F., 2005. The angular velocity of Nubia relative to Somalia and the location of the Nubia–Somalia–Antarctica triple junction, *Geophysical Journal International*, **162**, 221–238.
- 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.
- Krijgsman, W. & Tauxe, L., 2004. Shallow bias in Mediterranean paleomagnetic directions caused by inclination error, *Earth Planet. Sci. Lett.*, **222**, 685–695.
- Lemaux, J., II, Gordon, R. G. & Royer, J-Y, 2002. Location of the Nubia–Somalia boundary along the Southwest Indian Ridge, *Geology*, **30**, 339–342.
- Müller, R. D., Royer, J. Y. & Lawver, L. A., 1993. Revised plate motions rel-



**Figure 32.** Flowchart for recommended procedure of processing paleomagnetic data.

- ative to the hotspots from combined Atlantic and Indian-Ocean hotspot tracks, *Geology*, **21**, 275–278.
- McElhinny, M. W. & Lock, J., 1996. IAGA Paleomagnetic Databases with access, *Surveys in Geophysics*, **17**, 575–591.
- Müller, R. D., Roest, W. R., Royer, J. Y., Gahagan, L. M. & Sclater, J. G., 1997. Digital isochrons of the world's ocean floor, *J Geophys Res Solid Earth*, **102**, 3211–3214.
- 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.
- Marks, K. M. & Tikku, A. A., 2001. Cretaceous reconstructions of East Antarctica, Africa and Madagascar, *Earth and Planetary Science Letters*, **186**, 479–495.
- McQuarrie, N. & Wernicke, B. P., 2006. An animated tectonic reconstruction of southwestern North America since 36 Ma, *Geosphere*, **1**, 147–172.
- Müller, R. D., Sdrolias, M., Gaina, C. & Roest, W. R., 2008. Age, spreading rates, and spreading asymmetry of the world's ocean crust, *Geochemistry, Geophysics, Geosystems*, **9**, Q04006.
- Müller, R. D., Seton, M., Zahirovic, S., Williams, S. E., Matthews, K. J., Wright, N. M., Shephard, G. E., Maloney, K., Barnett-Moore, N., Hosseinpour, M., Bower, D. J. & Cannon, J., 2016. Ocean basin evolution and global-scale plate reorganization events since Pangea breakup, *Annu Rev Earth Planet Sci*, **44**, 107–138.
- Najman, Y., Appel, E., Boudagher-Fadel, M., Bown, P., Carter, A., Garzanti, E., Godin, L., Han, J., Liebke, U., Oliver, G., Parrish, R., Vezzoli, G., 2010. Timing of India-Asia collision: Geological, biostratigraphic, and palaeomagnetic constraints, *J. Geophys. Res. Solid Earth*, **115**, B12416.
- 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.
- Patriat, P., Sloan, H. & Sauter, D., 2008. From slow to ultraslow: A previously undetected event at the Southwest Indian Ridge at ca. 24 Ma, *Geology*, **36**, 207–210.
- Riisager, P., Hall, S., Antretter, M. & Zhao, X., 2004. Early Cretaceous Pacific palaeomagnetic pole from Ontong Java Plateau basement rocks, in *Origin and Evolution of the Ongong Java Plateau*, **229**, pp. 31–44, eds Fitton, J. G., Mahoney, J. J., Wallace, P. J. & Saunders, A. D., Geological Society of London Special Publications.
- Rowan, C. J. & Rowley, D. B., 2016. Preserved history of global mean spreading rate: 83 Ma to present, *Geophys. J Int.*, **208**, 1173–1183.
- Steinberger, B. & O'Connell, R. J., 1998. Advection of plumes in mantle flow: implications for hotspot motion, mantle viscosity and plume distribution, *Geophys. J Int.*, **132**, 412–434.
- 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.
- Shephard, G. E., Bunge, H.-P., Schuberth, B. S. A., Müller, R. D., Talsma, A. S., Moder, C. & Landgrebe, T. C. W., 2012. Testing absolute plate reference frames and the implications for the generation of geodynamic mantle heterogeneity structure, *Earth Planet. Sci. Lett.*, **317–318**, 204–217.
- 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

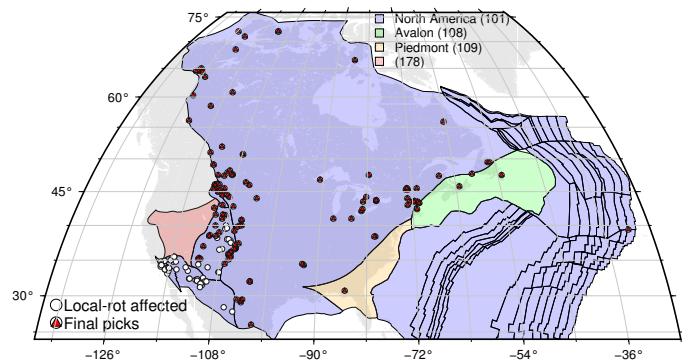
- reality with GMAP, *Comput. Geosci.*, **25**, 395–402.
- Tauxe, L. & Kent, D. V., 2004. A simplified statistical model for the geomagnetic field and the detection of shallow bias in paleomagnetic inclinations: Was the ancient magnetic field dipolar? *Geophys. Monogr. AGU*, **145**, 101–115.
- Tan, X., Kodama, K. P., Gilder, S., & Courtillot, V., 2007. Rock magnetic evidence for inclination shallowing in the Passaic Formation red beds from the Newark basin and a systematic bias of the Late Triassic apparent polar wander path for North America, *Earth Planet. Sci. Lett.*, **254**, 345–357.
- 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., 2019. *Essentials of Paleomagnetism*, 5th web edn, Available on line.
- van der Voo, R., 1990. The reliability of paleomagnetic data, *Tectonophysics*, **184**, 1–9.
- Whittaker, J. M., Müller, R. D., Leitchenkov, G., Stagg, H., Sdrolias, C. G. & Goncharov, A., 2007. Major Australian-Antarctic Plate Reorganization at Hawaiian-Emperor Bend Time, *Science*, **318**, 83–86.
- 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.
- Whittaker, J. M., Williams, S. E. & Müller, R. D., 2013. Revised tectonic evolution of the Eastern Indian Ocean, *Geochem. Geophys. Geosyst.*, **14**, 1891–1909.
- Young, A., Flament, N., Maloney, K., Williams, S., Matthews, K., Zahirovic, S. & Müller, D., 2019. Global kinematics of tectonic plates and subduction zones since the late Paleozoic Era, *Geosci. Front.*, **10**, 989–1013.

## APPENDIX A: CONSTRAIN PALEOPOLES FOR A CERTAIN TECTONIC PLATE

A polygon can be drawn around a set of paleomagnetic 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. Regions like those close to the plate boundaries are usually tectonically active (e.g. local rotations), so we should also be careful when we deal with the paleopoles derived from this type of locations.

### A1 120–0 Ma North America

The data-constraining polygons are from the recently published plate model (Young et al. 2018) (Fig. A1). Plate ID 101 polygon in the recently published Plate Model (Young et al. 2018), including its children 108 (Avalon/Acadia block) and 109 (Piedmont block) polygons for 120–0 Ma, is used to select the sampling sites of the paleopoles for North America. According to the plate model rotation data (Young et al. 2018), 108 is fixed to 101 during the geologic period from Cretaceous to the present day. 109 is also fixed to 101 since about 300 Ma (Christeson et al. 2014). Then in order to be



**Figure A1.** The final filtered datasets (red triangle-inside-circles) for later analysis on 120–0 Ma North America. Those poles that had been influenced by local tectonic rotations are shown as white circles.

compared with the FHM (120–0 Ma) (Müller et al. 1993; Müller et al. 1999), the paleopoles with age ranging 120–0 Ma are further selected through constraining the lower magnetic age “LOMAGAGE  $\leq 135$ ” (here it is not 120 but 135, because for the lower resolution case when the window length is 30 Myr, the Age Position Picking method will include those data with their lower magnetic age between 120 Ma and 135 Ma). In addition, the RESULTNO=6007 dataset should also be included according to a published plate kinematic model (McQuarrie & Wernicke 2006) with a relatively higher resolution of polygons and rotations, although the dataset is in the PlateID=178 polygon. In the end, 193 datasets in total are extracted (both white circles and red triangle-inside-circles in Fig. A1).

Also based on this model of southwestern North America since 36 Ma (McQuarrie & Wernicke 2006), part of the paleopoles constrained by the four small western terranes whose Plate IDs are also 101 (white circles in Fig. A1) in fact had gone through regional rotations and here are removed. However, the poles with age younger than 10 Ma located within the largest western 101 terrane (on the south of the smallest western 101 terrane; corresponding to the RANGE\_ID=74 polygon in the model (McQuarrie & Wernicke 2006)) should be included. So finally 135 of the 193 datasets remain (Fig. A1). Spatially North American paleomagnetic data are mainly from the western and eastern margins of the plate.

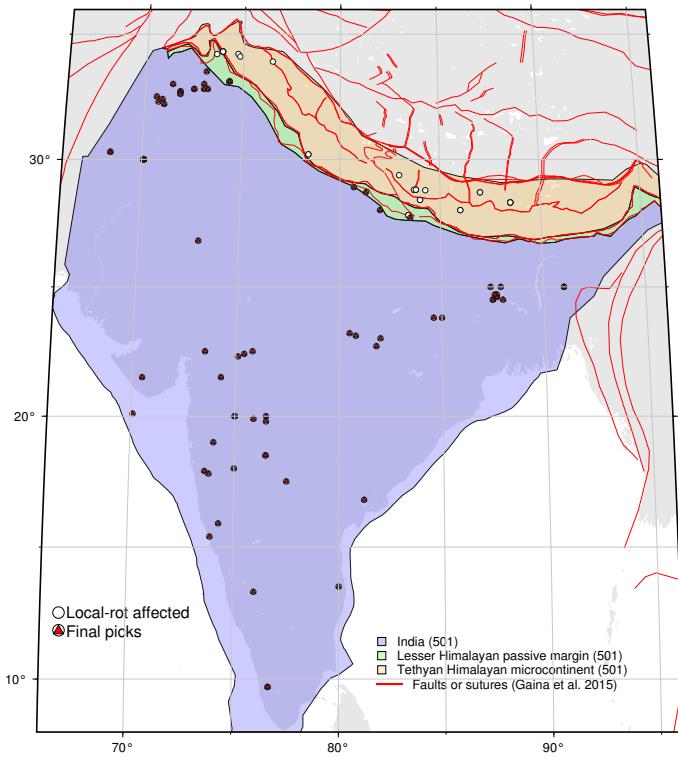
### A2 120–0 Ma India

Plate ID 501 polygons in the recently published Plate Model (Young et al. 2018) also include the two small polygons of the northern “Lesser Himalayan passive margin of Greater Indian Basin” and “Tethyan Himalayan microcontinent of Greater India” (Fig. A2). The polygons are used to select the sampling sites of the paleopoles for India (Fig. A2).

Based on the model of the tectonic interactions between India, Arabia and Asia since the Jurassic (Gaina et al. 2015) (Fig. A2), part of the paleopoles constrained by the north two small terranes whose Plate IDs are also 501 in fact had gone through regional rotations and here are removed. So finally 75 datasets are left (Fig. A2). Spatially Indian paleomagnetic data are more evenly distributed on the India plate than North American and Australian poles.

### A3 120–0 Ma Australia

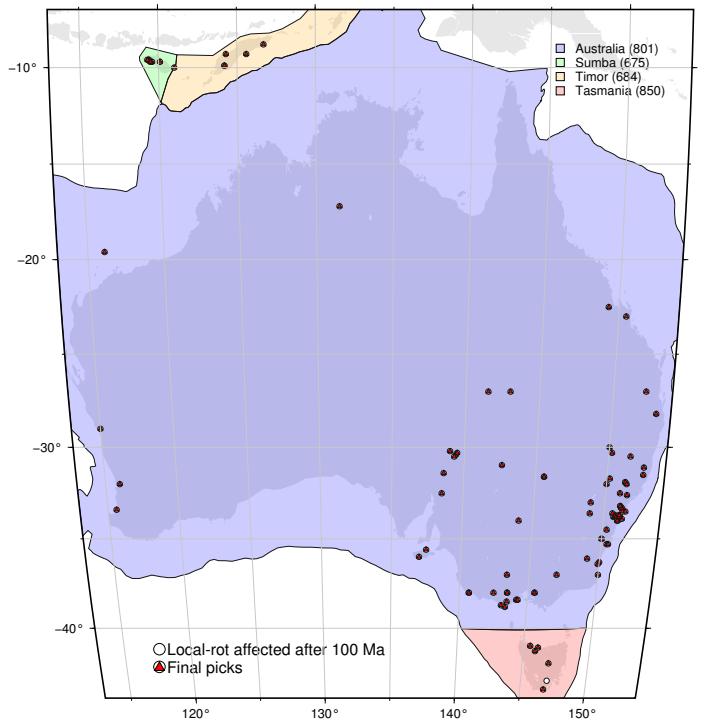
Plate ID 801 polygon in the recently published Plate Model (Young et al. 2018), including its children 675 (Sumba block) and 684



**Figure A2.** The final filtered datasets (red triangle-inside-circles) for later analysis on 120–0 Ma India. Those poles that had been influenced by local tectonic rotations are shown as white circles. The rifts, faults and detachments (red lines) around India are used to filter out those data that are influenced by local tectonic rotations.

(Timor block) polygons for 120–0 Ma (Fig. A3), is used to select the sampling sites of the paleopoles for Australia. According to the plate model rotation data (Young et al. 2018), 675 and 684 are fixed to 801 during the geologic period from c.145 Ma to the present.

On the southeast of the main Australia plate (the blue polygon in Fig. A3), there is a triangle-shaped small polygon 850 (Tasmania block) which is fixed to 801 since c.100 Ma according to the (Young et al. 2018) rotation data. With that attribute, 805 contributes more data younger than c.100Ma for the later analysis. Ultimately the final 99 extracted datasets is shown in Fig. A3.



**Figure A3.** The final filtered datasets (red triangle-inside-circles) for later analysis on 120–0 Ma Australia. Those poles that had been influenced by local tectonic rotations are shown as white circles. The Plate ID 850 helps increase the amount of qualified datasets for 100–0 Ma.