

Methods for Producing a Reliable APWP

Chenjian Fu^{1*} Chris Rowan²

¹ Department of Geology, Kent State University, 325 S Lincoln St, Kent, OH 44242, USA

² Department of Geology, Kent State University, 325 S Lincoln St, Kent, OH 44242, USA

Received 2018 December 28; in original form 2018 November 22

SUMMARY

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

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

1 INTRODUCTION

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

1.1 Not All Data Are Created Equal

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

1.1.1 Age Error

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

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

1.1.2 Position Error

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

1.1.3 Data Consistency

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

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

Lithology

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

Local Rotations

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

Other Factors

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

* Email: cfu3@kent.edu

2 Chenjian Fu

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** (GP-MDB) version 4.6b (Pisarevsky 2005) has more than 130 poles for NAC, but only 17 for Greenland and 24 for Arabia.

1.1.5 Publication Year

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

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

1.2 Existing Solutions and General Issues

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

limited study of how effective these marking/filtering methods are at reconstructing a ‘true’ APWP, and for most studies after a basic filtering of ‘low quality’ poles, the remaining poles are, in fact, treated equally.

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

2 METHODS

For most of Earth history, concretely for times before c. 170 Ma, the age of the oldest magnetic anomaly identification, paleomagnetism is the only accepted quantitative method for reconstructing plate motions and past paleogeographies. After about 170 Ma, multiple data sources can help constrain plate motions in more accurate ways. One of the most developed and studied plate kinematics models is the Fixed Hotspot Model (FHM) (Müller et al. 1993; Müller et al. 1999), which assumes the Atlantic and Indian hotspots are relatively fixed. Another one is the Moving Hotspot Model (MHM) (O’Neill et al. 2005), which is based on mantle convection models that indicate large motions of the Indo-Atlantic hotspots.

2.1 Reference Paths: The Hotspot and Seafloor Spreading Model Predicted

Such a model like FHM or MHM can predict APWPs for main continents, e.g. the North America (Plate ID 101) (Fig. 1 and Fig. 2), the India (501) (Fig. 3 and Fig. 4) and the Australia (801) (Fig. 5 and Fig. 6), with the help of global tectonic plate motion data (i.e. plate circuits) from the ocean basins (i.e. spreading sea floors) that has been reconstructed for the last c. 180–200 Myr (although the extrapolation is required; hard to constrain uncertainties). These more accurate and totally different-source derived model-predicted APWPs can be compared to data from the paleomagnetic database. For about 120 Ma to the present, India drifts much faster than North America and Australia. North America has been drifting rather slow. Australia’s drifting rate is between North America and India. In addition, India drifts almost north, whereas North America east and then north, and Australia west and then north. So these three continents are representatives for three different types of plate kinematics.

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

2.2 120–0 Ma Paleomagnetic APWPs

The GPMDB 4.6b (Pisarevsky 2005), data source used here, includes 9514 paleopoles for ages of 3,500 Ma to the present published from 1925 to 2016. A polygon can be drawn around a set of data, whose sampling sites we believe belong to a specific plate or rigid block. Then the *Spatial Join* technique (Jacox & Samet 2007) helps join attributes from the polygon to the paleomagnetic data based on the spatial relationship allowing data within this polygon to be extracted from the whole raw large dataset without splitting a

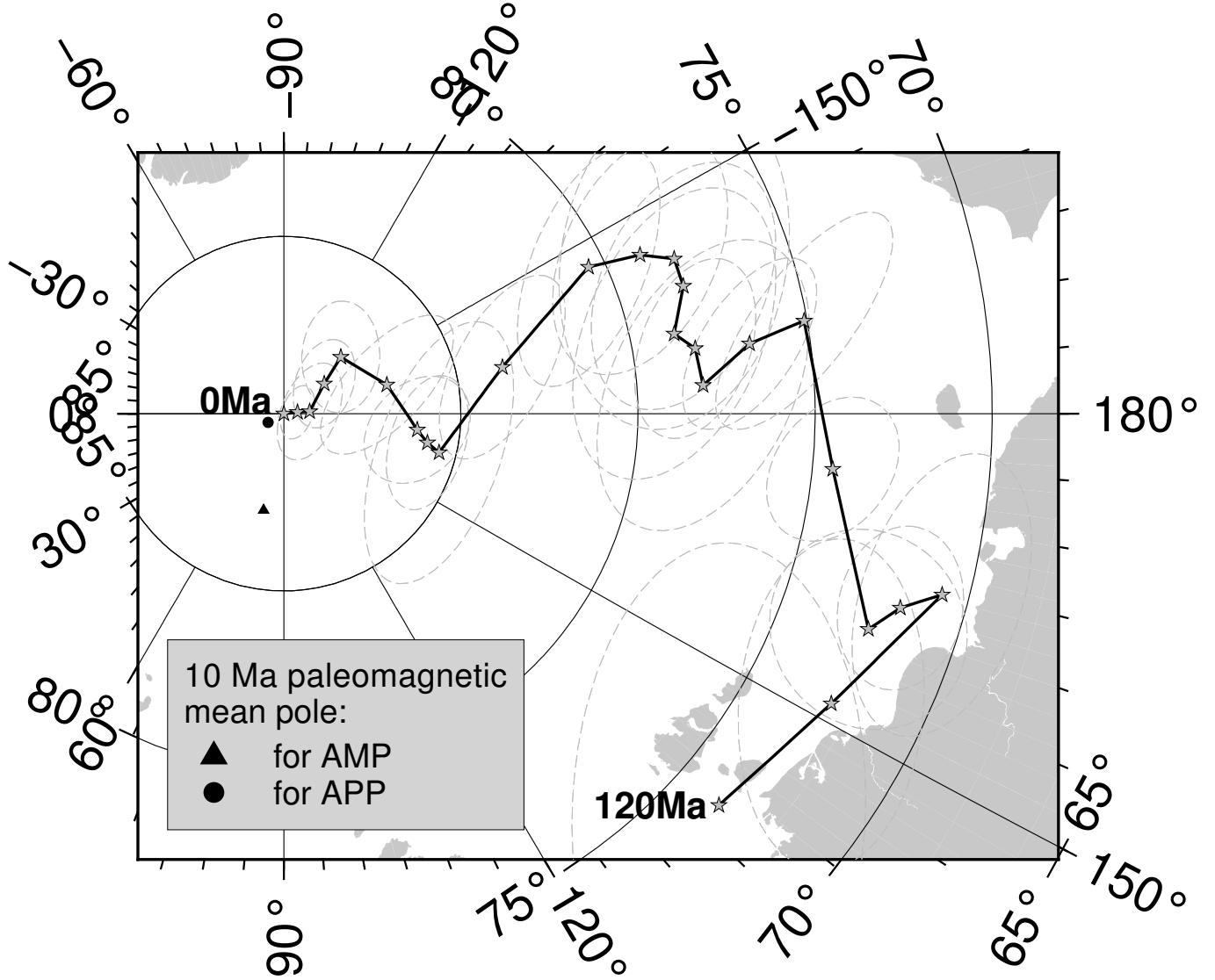


Figure 1. FHM predicted 120–0 Ma APWP for NAC through the North America–Nubia–Mantle plate circuit. Its age step is 5 Myr.

subset just for a specific plate. That allows us to quickly select subsets of the database based on geographic constraints just as easily as for age. Of course, the boundary of this polygon must be reasonably along a tectonic boundary (see the details about data filtering for North America in the Supplementary Material). The temporal distributions of North American, Indian and Australian 120–0 Ma poles are shown in Fig. 7, Fig. 8 and Fig. 9. Compared with North American and Australian paleomagnetic VGPs, Indian paleomagnetic VGPs are in relatively lower density in general except during the period of about 60–70 Ma (Fig. 8).

2.3 Picking Data for A Certain Time Window

2.3.1 Moving Average

The moving average method, also called “running mean” or “moving window” (Torsvik et al. 2008) method, calculates the average of values between a certain data (age in our case) range; the average is then recalculated as the limits of the bin are repeatedly incremented upwards. In addition to the traditional moving windows averaging algorithm, a newly developed moving average method

is also used, referred to here as the “Age Position Picking (APP)” method. The difference of this moving average method from the one built in GMAP (Torsvik & Smethurst 1999; Torsvik et al. 2008) is that the whole magnetic age range is taken into account in each window, while GMAP only considers the mid-point of the low and high magnetic age of each pole, an algorithm referred to as the “Age Mean Picking (AMP)” method.

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

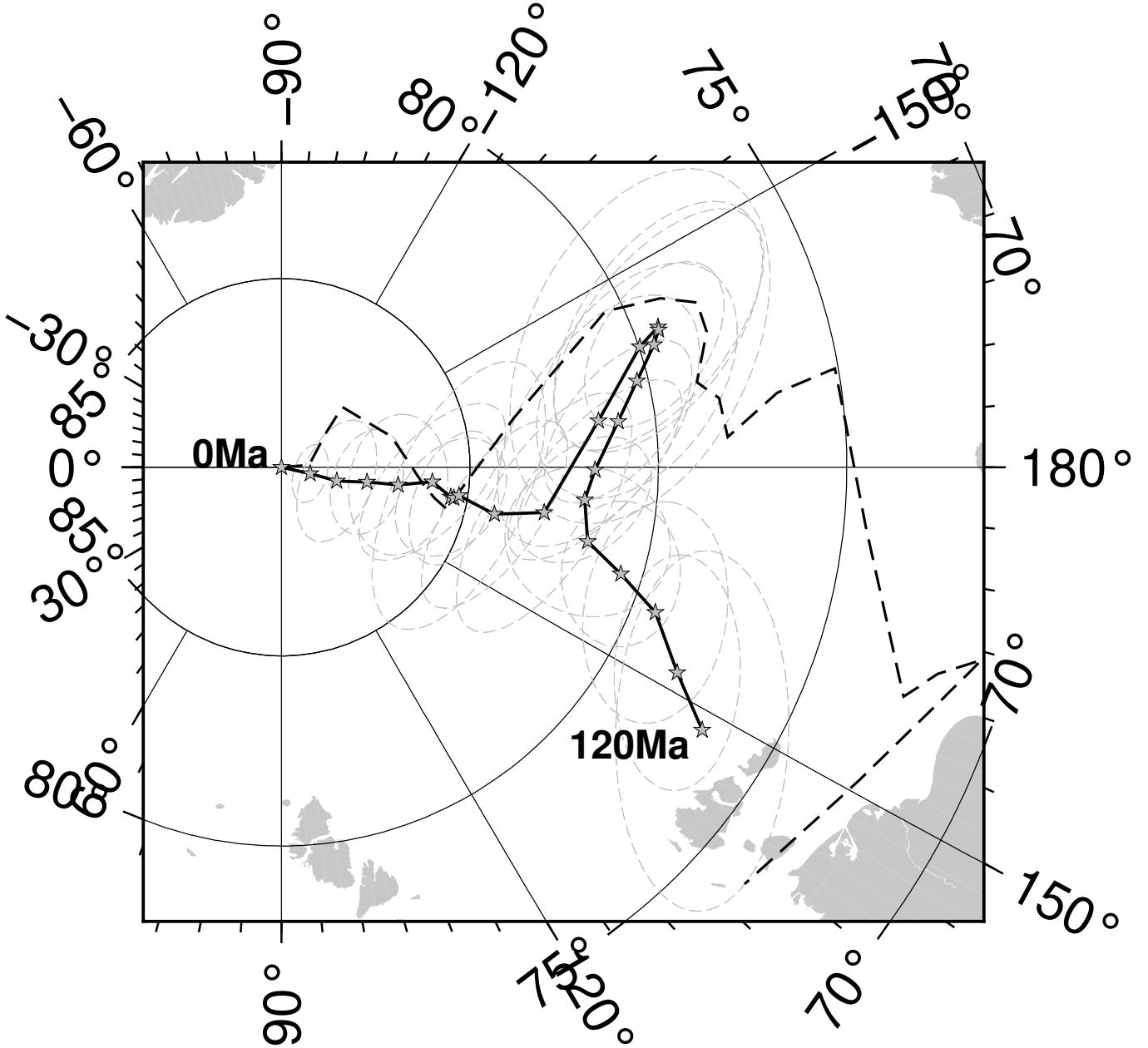


Figure 2. MHM predicted 120–0 Ma APWP for NAC through the North America–Nubia–Mantle plate circuit. Its age step is 5 Myr. The dashed line is the FHM predicted path shown in Fig. 1 without confidence ovals for clarity.

entire possible acquisition age. Fig. 10 shows an example of moving average with a 10 Myr window and a 5 Myr step. So, for example, for the window of 15 Ma to 5 Ma (the light blue bin in Fig. 10), the AMP method calculates the Fisher mean pole of only 5 poles, while the APP method calculates the mean pole of 9 poles. From comparison of mean poles of the picked poles for the light blue age window with the two different algorithms (the 10 Ma mean poles in Fig. 1), the mean pole from the APP method is closer to the 10 Ma pole in the FHM predicted path.

In addition, technically we don't want to let the step length more than the window length, in which cases we would lose data between windows.

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

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

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

Filtering Out Igneous Unrelated Poles With this filter, the poles are from rocks that contain extrusive or intrusive igneous

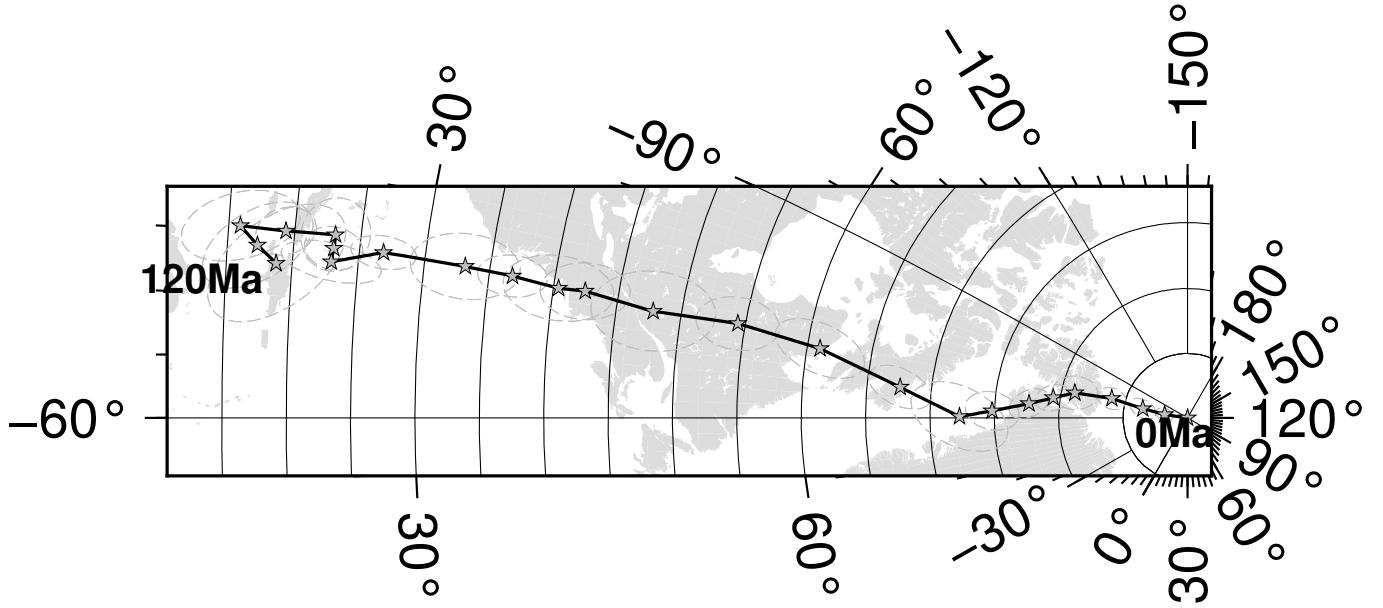


Figure 3. FHM predicted 120–0 Ma APWP for India through the India–Somalia–Nubia–Mantle plate circuit. Its age step is 5 Myr.

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

No.	Picking Algorithm
0	AMP
1	APP
2	AMP (“ α 95/Age range” no more than “15/20”)
3	APP (“ α 95/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)

type. In other words, the rock type could be mainly sedimentary or metamorphic.

Inclination Shallowing and Unflattening To test if unflattening possible inclination shallowing in sedimentary rocks can im-

prove the APWP fitting outcomes, the flattening function (King 1955) $\tan I_o = f \tan I_f$ is used to unflatten assumed existing inclination shallowing in sedimentary-based or redbeds-based paleomagnetic data (No 5 and 7 in Table. 1), where I_o is the observed inclination, I_f is the unflattened inclination, and f is the flattening factor (or shallow coefficient) ranging from unity (no flattening) to 0 (completely flattened). Here $f = 0.6$ is used in our calculations, according to the previous experience (Torsvik et al. 2012).

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

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

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

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

Filtering Using SS05 Liability Criteria SS05 (Schettino & Scotese 2005) provided their criteria of picking paleomagnetic data for producing their APWPs. This filter is reproducing their criteria by setting α 95 of $\leq 15^\circ$, age uncertainty of ≤ 40 Myr, sampling sites' quantity of ≥ 4 , samples' quantity of at least 4 times of the sites, and laboratory analytical procedure code of at least 2.

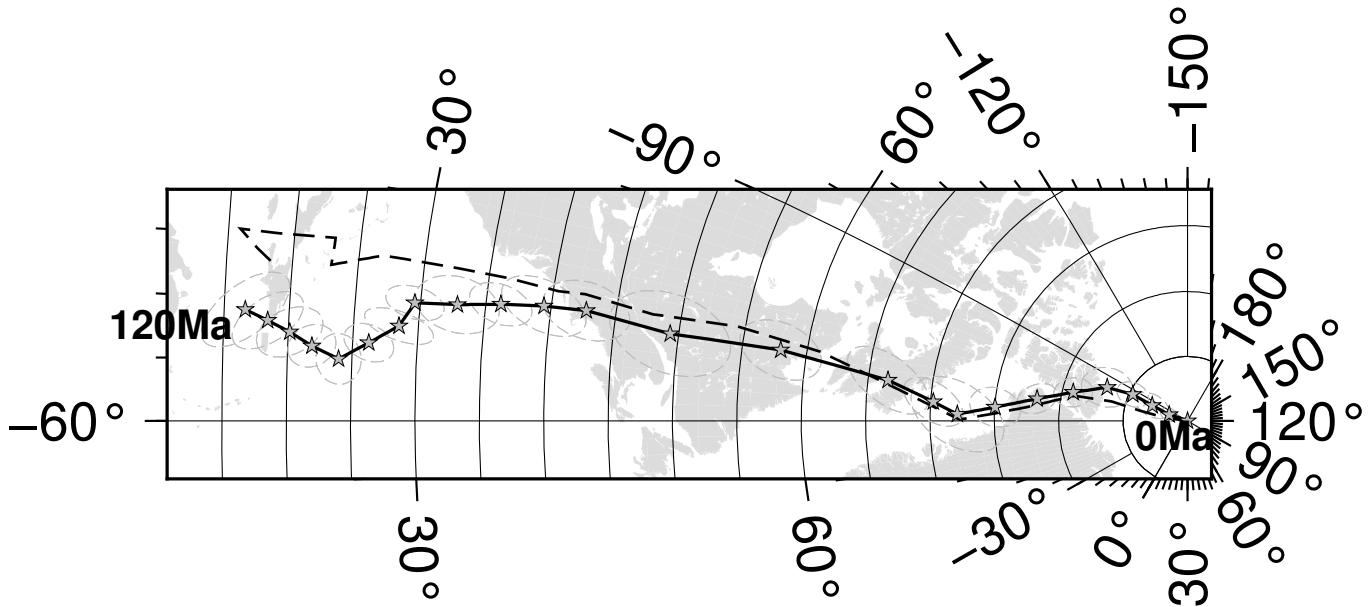


Figure 4. MHM predicted 120–0 Ma APWP for India through the India–Somalia–Nubia–Mantle plate circuit. Its age step is 5 Myr. The dashed line is the FHM predicted path shown in Fig. 3 without confidence ovals for clarity.

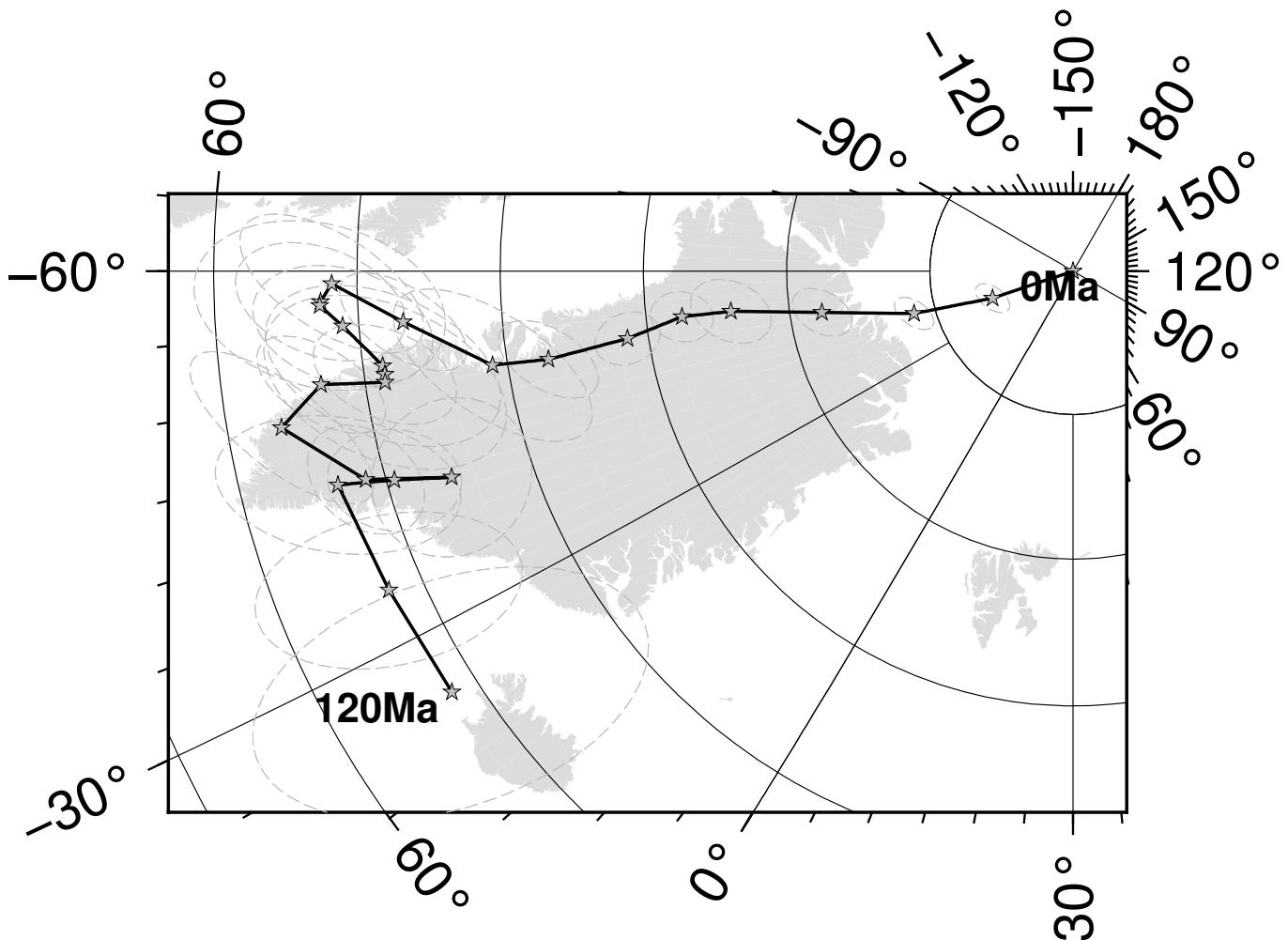


Figure 5. FHM predicted 120–0 Ma APWP for Australia through the Australia–East Antarctica–Somalia–Nubia–Mantle plate circuit. Its age step is 5 Myr.

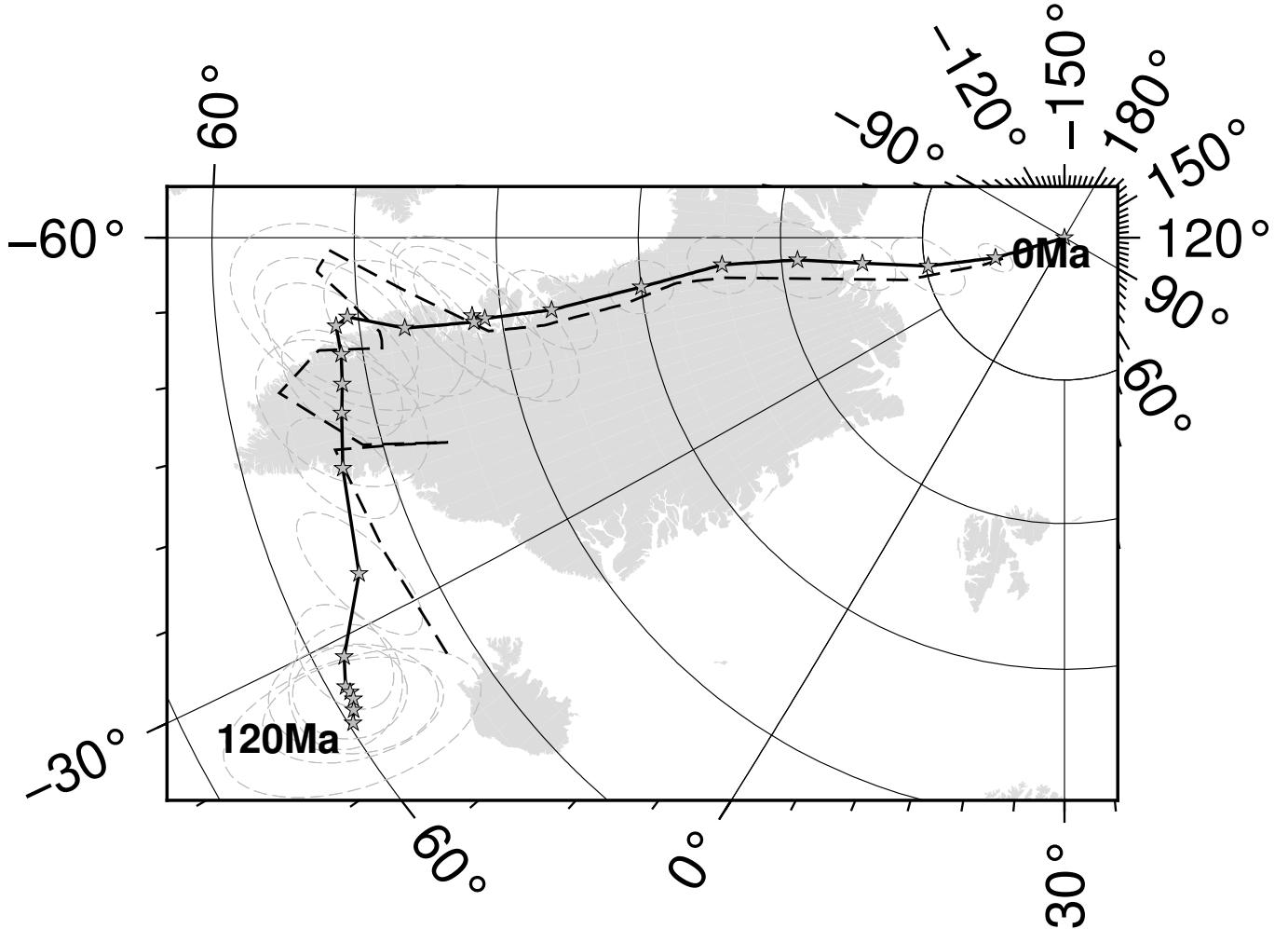


Figure 6. MHM predicted 120–0 Ma APWP for Australia through the Australia–East Antarctica–Somalia–Nubia–Mantle plate circuit. Its age step is 5 Myr. The dashed line is the FHM predicted path shown in Fig. 5 without confidence ovals for clarity.

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

2.3.1.2 Weighting Because all data is not created equal, we want to calculate a weighted mean pole for a time interval with ‘better’ (more likely to be reliable) poles counting more than ‘worse’. For example, a pole with small α_{95} and very well constrained age is more likely to reflect APWP position at the selected age point than a pole with large α_{95} and very broad age range. There are many potential ways to weight this data set which can obviously greatly influence the final result, and we want to test this.

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

In order to average errors in orientation of the samples and scatter caused by secular variation, a “sufficient” number of individually oriented samples (observations) from “enough” sites must be satisfied (Tauxe et al. 2018; van der Voo 1990; Besse & Courtillot 2002). So for the “Numbers of sites (B), Observations (N)” weighting (No 1 in Table 2), larger B and N mean stronger weighting. Through knowing the pattern of all B and N in the database, the proposed solutions are as follows. If both B and N are more than 1, $\text{weight} = (1 - \frac{1}{B}) * (1 - \frac{1}{N})$. There are data in GPMDB with only the

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

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

number of sampling sites (at least greater than 1) given, but no number of samples or only one sample given, so for this case, if $B > 1$ and $N \leq 1$, $\text{weight} = (1 - \frac{1}{B}) * 0.5$. If only the number of samples (at least greater than 1) is given, and the number of sampling sites is missing or only one, i.e. $B \leq 1$ and $N > 1$, $\text{weight} = (1 - \frac{1}{N}) * 0.5$. If $B \leq 1$ and $N \leq 1$ (there are only 23 datasets for the whole GPMDB 4.6b, including 18 with both B and N informations missing), $\text{weight} = 0.2$.

As for the “Age uncertainty” weighting (No 2 in Table 2), a well-constrained age should be known to within a half of a geological period (e.g., Quaternary, Neogene, Triassic) for Phanerozoic data (van der Voo 1990; Tauxe et al. 2018). Generally, this work follows this principle. However, for the periods of Paleogene, Cre-

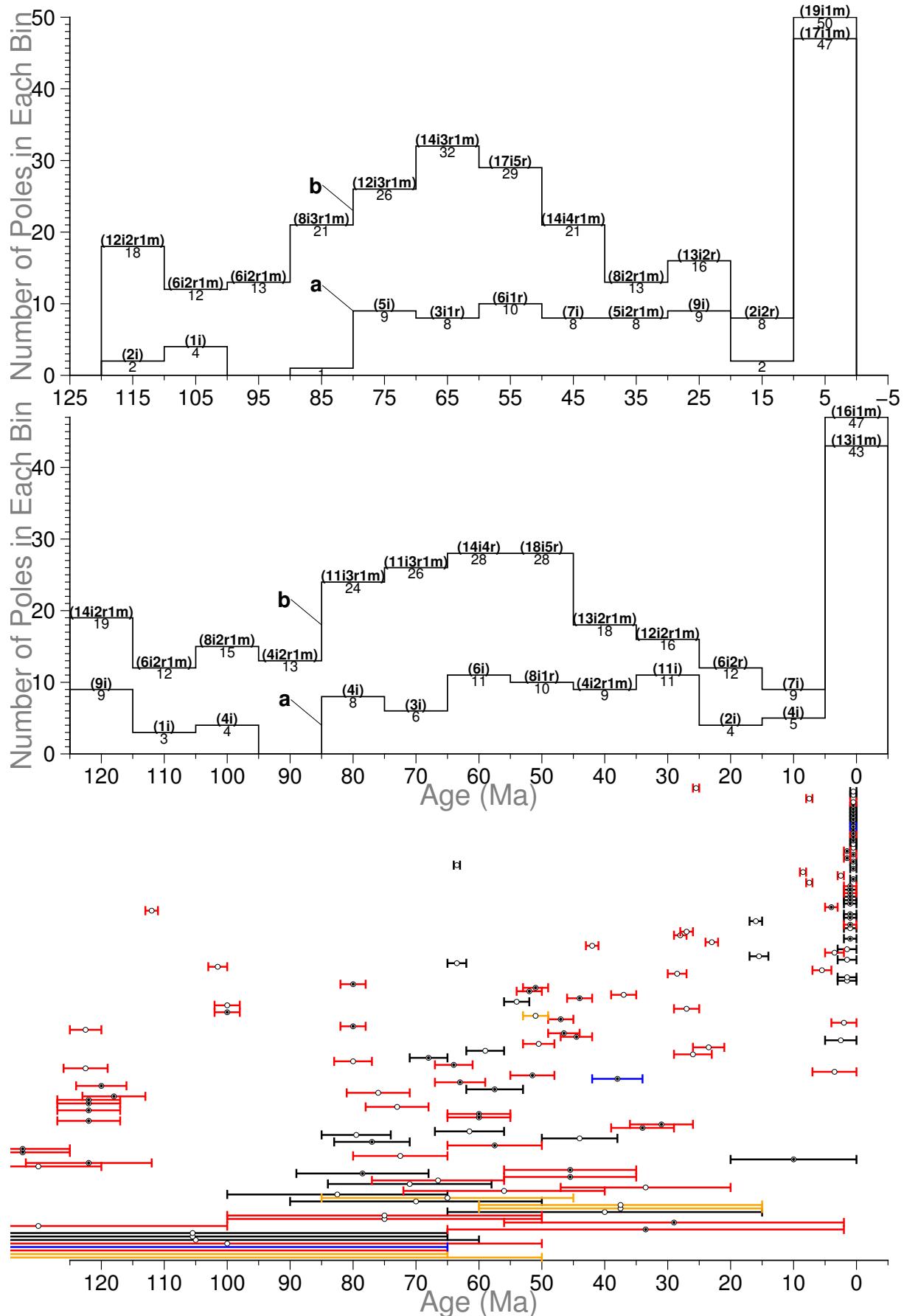


Figure 7. Temporal distribution of 120–0 Ma NAC (101) paleomagnetic poles in 10 Myr binning and 5 Myr step. For distribution a, each bin only counts in the midpoints of pole error bars (not including those right at bin edges); For distribution b, as long as the bar intersects with the bin (not including those intersecting only at one of bin edges), it is counted in. Inside the parentheses, i means igneous rocks derived (red bars), r means sedimentary rocks with redbeds involved derived (orange bars), and m means metamorphic-dominated rocks derived (blue bars); the left are pure sedimentary rocks derived (black bars). The midpoints published not later than 1983 are black-dotted.

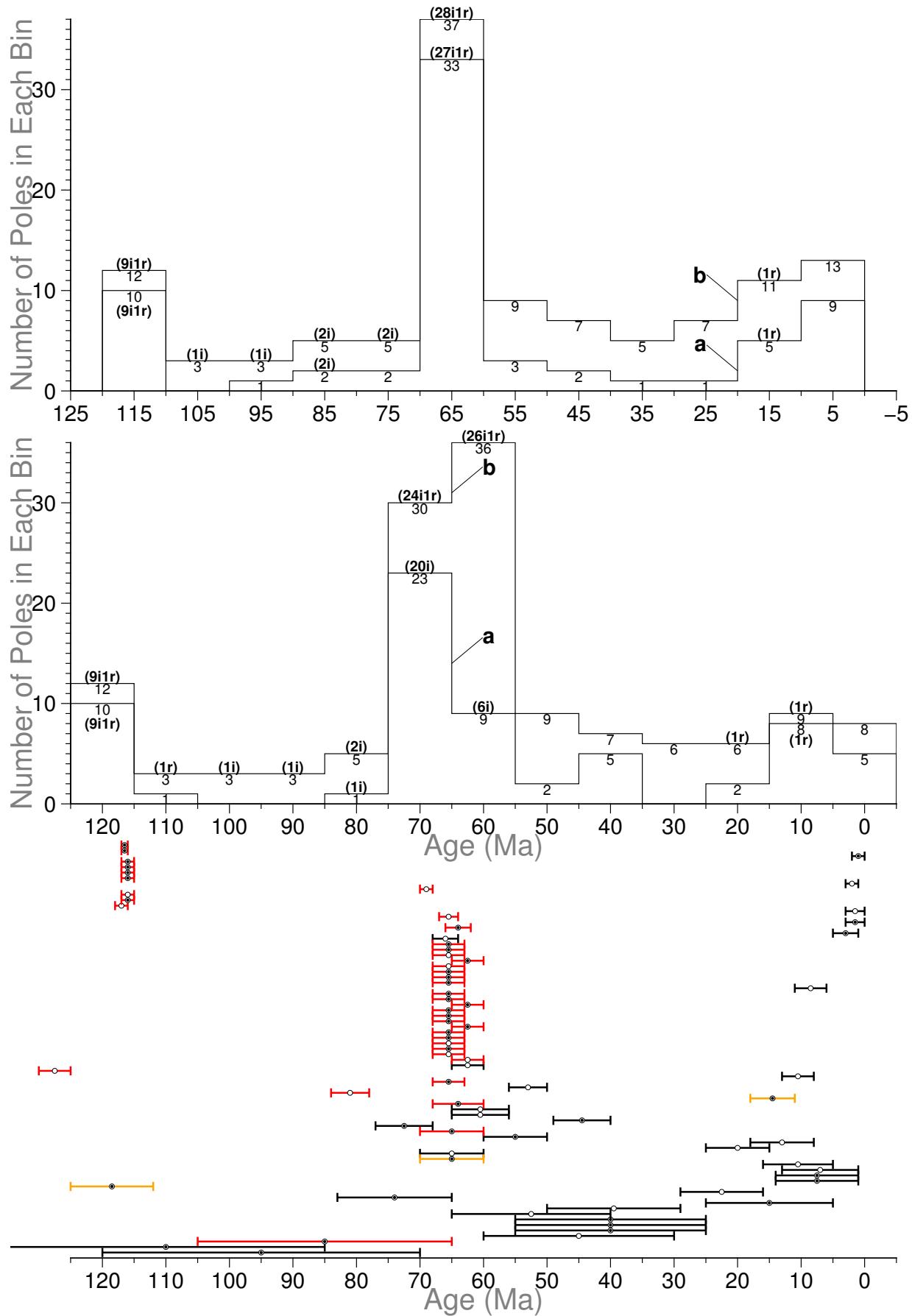


Figure 8. Temporal distribution of 120–0 Ma Indian (501) paleomagnetic poles. See Fig. 7 for more information.

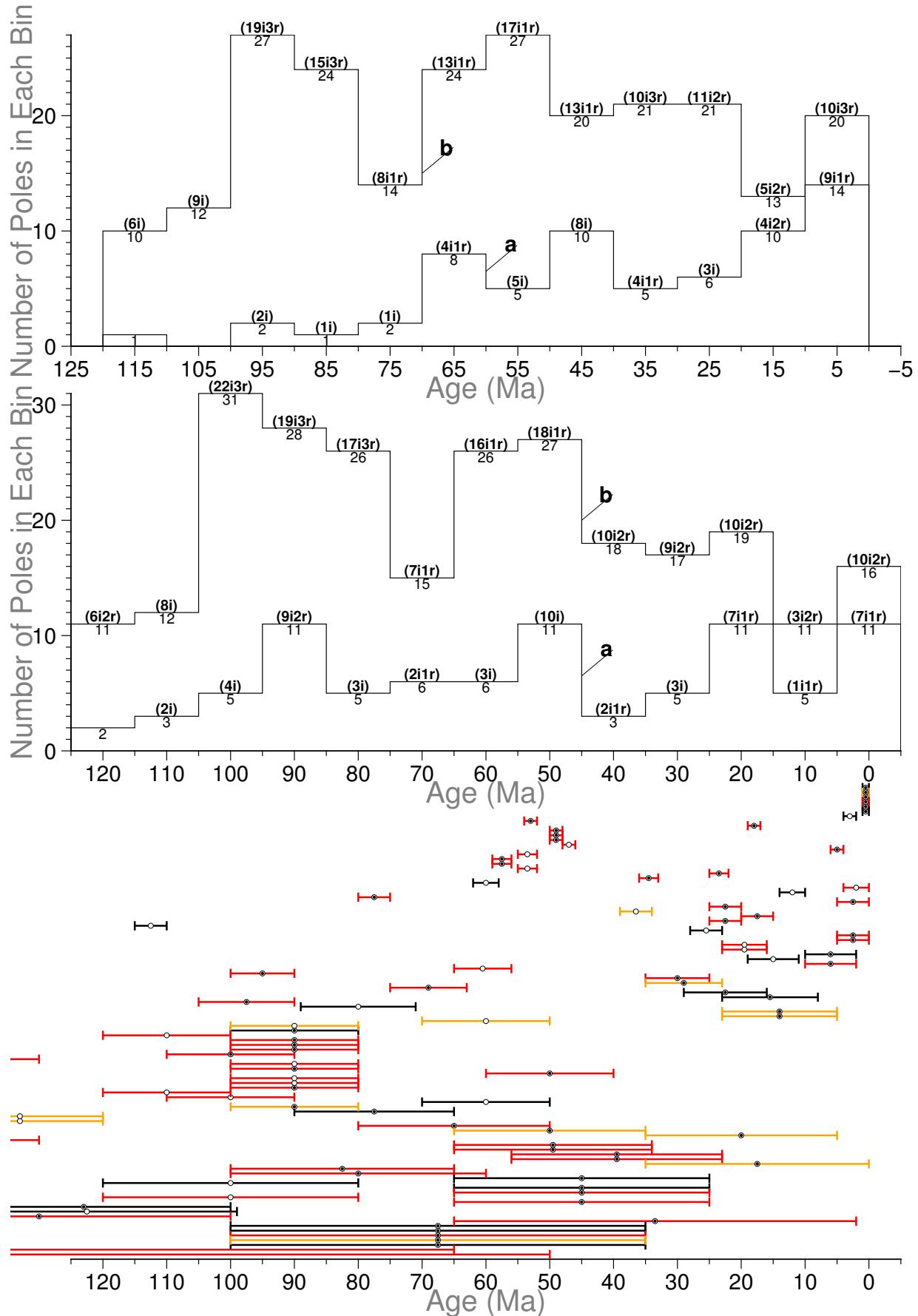


Figure 9. Temporal distribution of 120–0 Ma Australian (801) paleomagnetic poles. See Fig. 7 for more information.

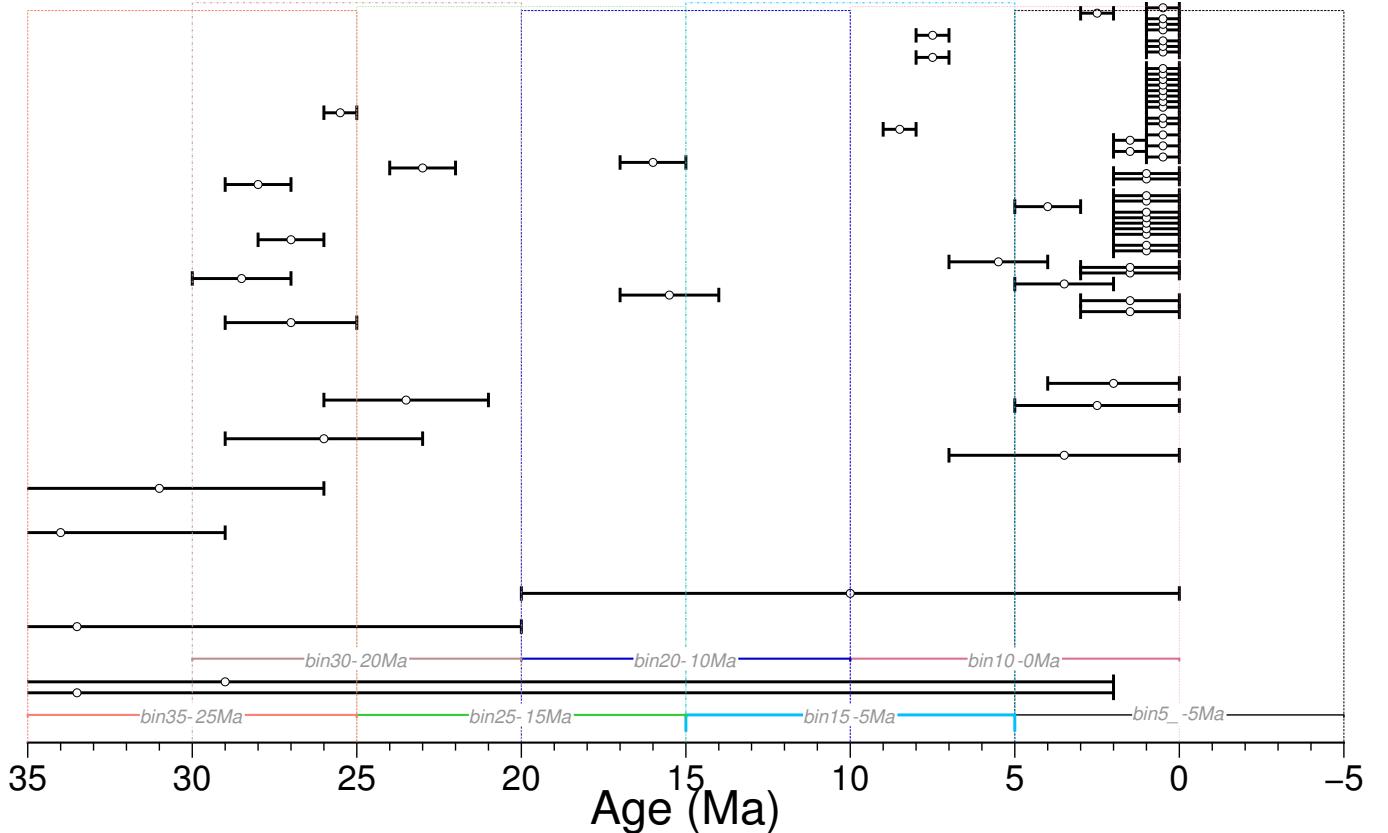


Figure 10. An example of 10 Myr moving window and 5 Myr step in the moving average method, based on poles of the NAC. Every age window has a different color. Red points are the midpoints of low and high magnetic ages. The vertical axis has no specific meaning here.

taceous, and Jurassic, their halves are all beyond a time span of at least 20 Myr, which is relatively large for these relatively young geologic periods. So for these three periods, a tighter age constraint is set using age uncertainties of ≤ 15 Myr. So, for example, for NAC's Neogene (23.03–2.58 Ma according to GSA Geologic Time Scale) data, if age uncertainty (the high magnetic age – the low magnetic age) ≤ 10.225 (from $0.5^*(23.03–2.58)$) Myr, its weight = 1; if age uncertainty > 10.225 Myr, its weight = $10.225 / (\text{high magnetic age} - \text{low magnetic age})$. For the periods spanning Jurassic to Paleogene, if age uncertainty ≤ 15 Myr, it gets its weight of 1; if age uncertainty > 15 Myr, a weight of $15 / (\text{high magnetic age} - \text{low magnetic age})$ is assigned instead.

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

For the “Age error Position to window” weighting (No 4 in Table. 2), if window intersects with young/old end of age bracket or whole window overlaps with a part of age range, weight= (overlapping part) / (age range width); if whole age range is within window, weight= (window width) / (age range width) (note that when weight > 1 , it is set back to 1).

The “Age error Position to window, and $\alpha95$ ” weighting (No 5 in Table. 2), is a combination of No 3 (but here the standard deviation is 15 though) and No 4.

Some of the picking (Table. 1) and weighting (Table. 2) 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 the Supplementary Material is related to the V90 criteria 5; and Pt 18 and 19 are related to the V90 criteria 7.

2.4 Path Comparison Method

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

First, the definition of the significant spatial difference d_s is changed into the fraction of n coeval pole pairs that are statistically distinguishable from each other, as determined by a test for a common mean direction. Then for the other two shape metrics d_a and d_l , there is no significance testing on them. The way of combining them three into the final composite path difference is still the same.

It will be interesting to see if there is any difference between the results given by these two methods. So we will have them both tested in the following content.

3 RESULTS AND DISCUSSIONS

3.1 Results

Is there any common pattern of the similarities for all the three continents? First, the best and worst methods need to be determined. Here, the difference values less than the one-standard-deviation interval (containing about 68.269% of the data values) are picked out as the “best” ones (lower about 15.866% of the data values), more than the one-standard-deviation interval as the “worst” ones (upper about 15.866%). Then we will see if there is one method or several methods labeled as “best” or “worst” for all the three continents.

3.1.1 When FHM and Global Plate Tectonic Circuits Predicted APWP as Reference

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

3.1.1.1 10 Myr Binning and 5 Myr Stepping

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

(i) the groups of picking-method-no 19 (APP with commented local-rotation or secondary-print studies excluded) and 21 (APP with local rotation excluded or corrected as suggested in the original sources) (Table. 1) are among the best ones for all the three continents, while the groups of picking-method-no 2 (AMP with “ α 95/Age range” no more than “15°/20 Myr”) and 16 (AMP with only earlier-than-1983 studies), (Table. 1) are among the worst for all them three.

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

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

(iv) For North America and India, newer studies (later than 1983) give better results, whereas for Australia older studies (before 1983) give better results, mainly because the number of older studies (65; Fig. 11c) is much greater than the number of newer studies (27).

(v) Weighting influences more to APP than AMP, especially for the cases when there are plenty of VGPs (e.g., for North America and Australia; Fig. 14). Wt 0 and 1 generally work well with APP.

(vi) Lithology related picking methods (Pt 4–13) are generally improving the fits, compared with other picking methods. However, sometimes, Wt 3 does not work well with these lithology-related methods (Fig. 11 and Fig. 14).

Measure 2: Only Space Tested According to the results (Fig. 15a, Fig. 15b and Fig. 15c), we can observe:

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

3.1.1.2 20 Myr Binning and 10 Myr Stepping

Measure 1: Both Space and Shape Tested According to the results (Fig. 17), we can observe:

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

Measure 2: Only Space Tested According to the results (Fig. 19), we can observe:

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

3.1.2 When MHM and Global Plate Tectonic Circuits Predicted APWP as Reference

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

3.1.2.1 10 Myr Binning and 5 Myr Stepping

Measure 1: Both Space and Shape Tested According to the results (Fig. 21a, Fig. 21b and Fig. 21c), we can observe:

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

Measure 2: Only Space Tested According to the results (Fig. 23a, Fig. 23b and Fig. 23c), we can observe:

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

3.1.2.2 20 Myr Binning and 10 Myr Stepping

Measure 1: Both Space and Shape Tested According to the results (Fig. 25), we can observe:

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

Measure 2: Only Space Tested According to the results (Fig. 27), we can observe:

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

3.1.3 Summary of Results

According to all the above results (Fig. 11, Fig. 15, Fig. 17, Fig. 19), we can observe:

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

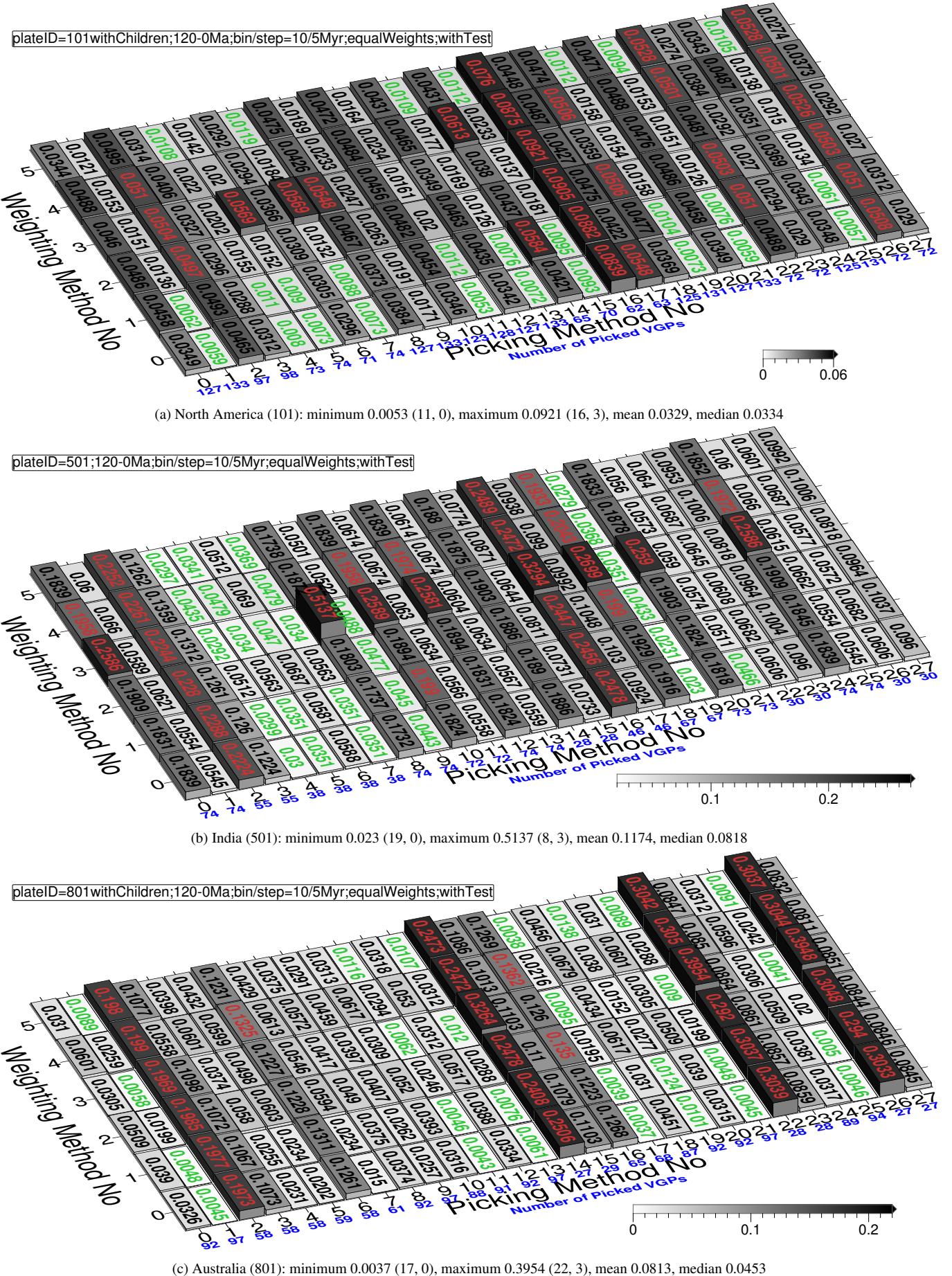


Figure 11. Equal-weight composite path difference (CPD) values with test between each continent's paleomagnetic APWPs and its predicted APWP from FHM and related plate circuits. The paths are in 10 Myr bin and 5 Myr step. The difference values less than one-standard-deviation interval of the whole 168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

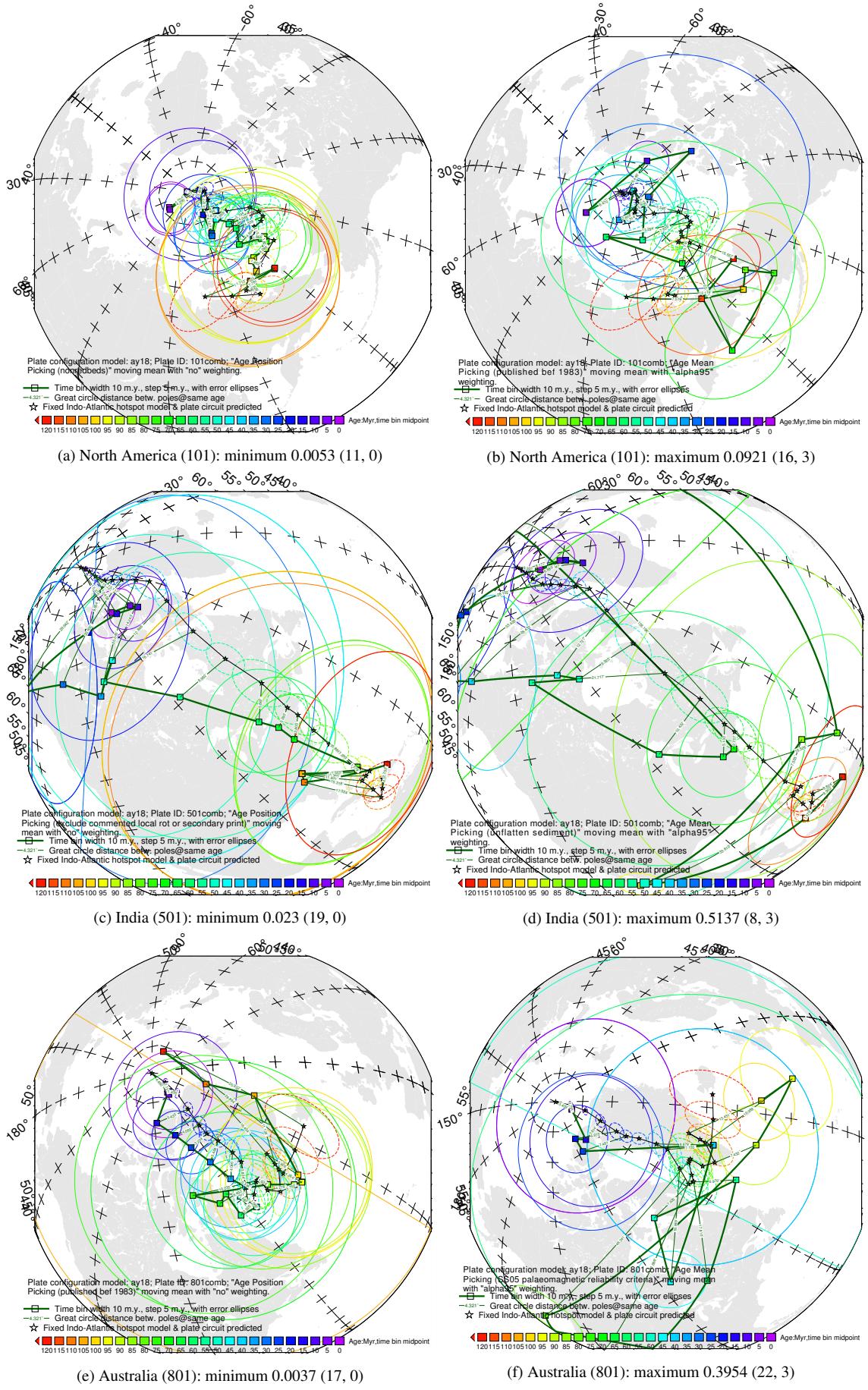


Figure 12. Path comparisons with best and worst difference values shown in Fig. 11.

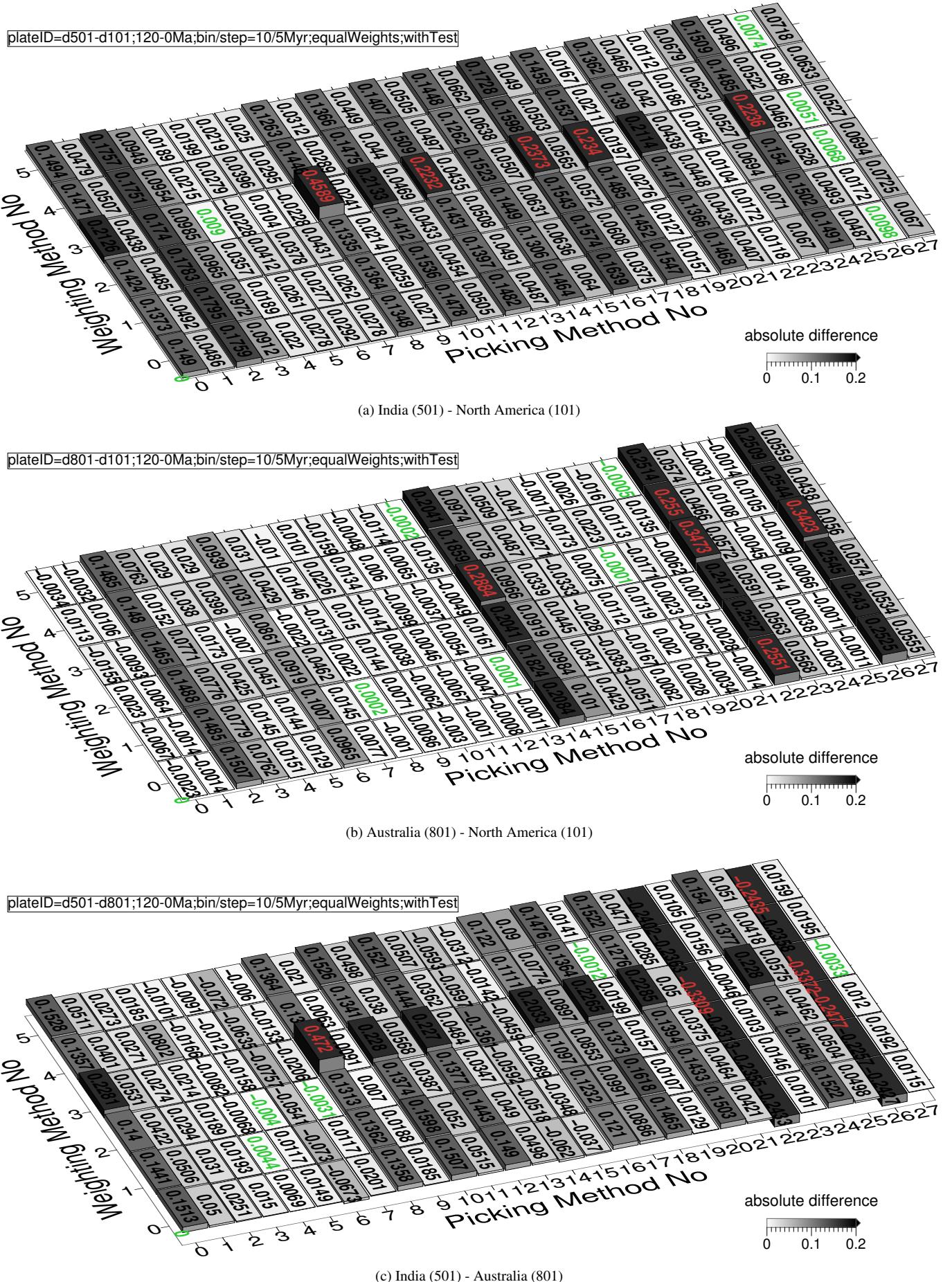


Figure 13. Differences between grids in Fig. 11. The absolute difference values less than 1.96-standard-deviation interval of the whole 168 values are labeled in green, more than 1.96-standard-deviation interval labeled in red.

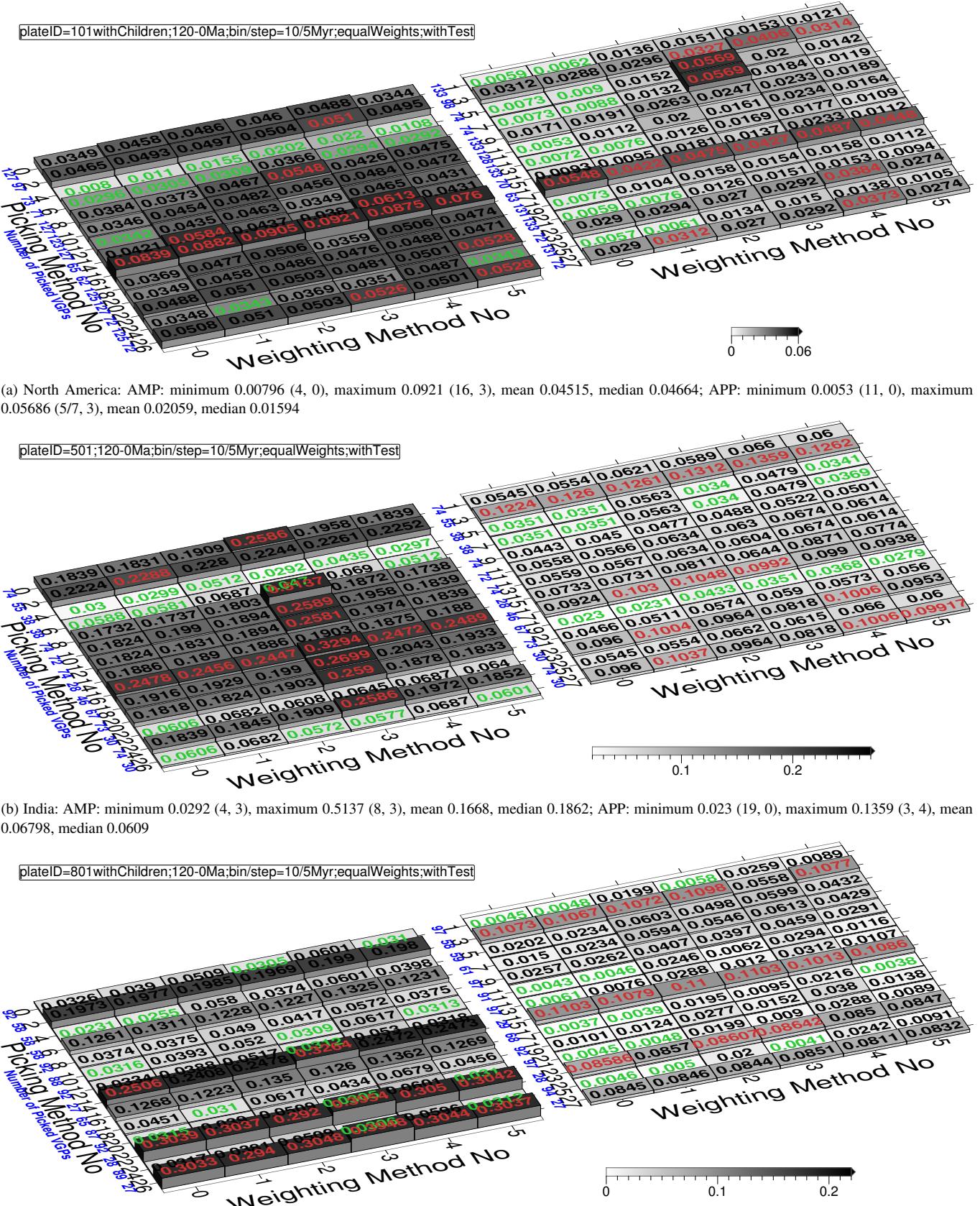


Figure 14. Separated results from AMP and APP in Fig. 11. For each grid block (left: AMP; right: APP), the difference values less than one-standard-deviation interval of the whole 84 values are labeled in green, more than one-standard-deviation interval labeled in red.

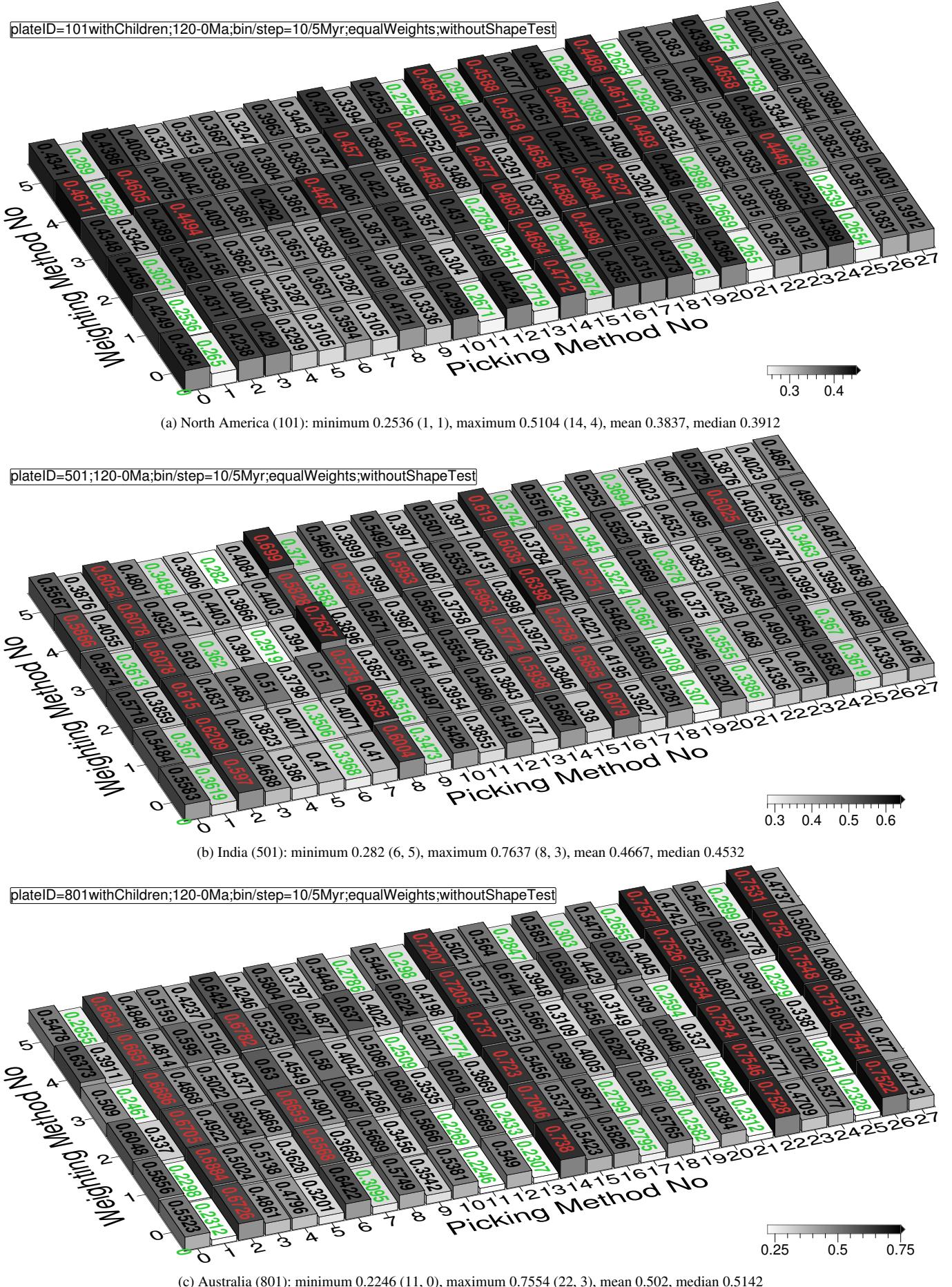
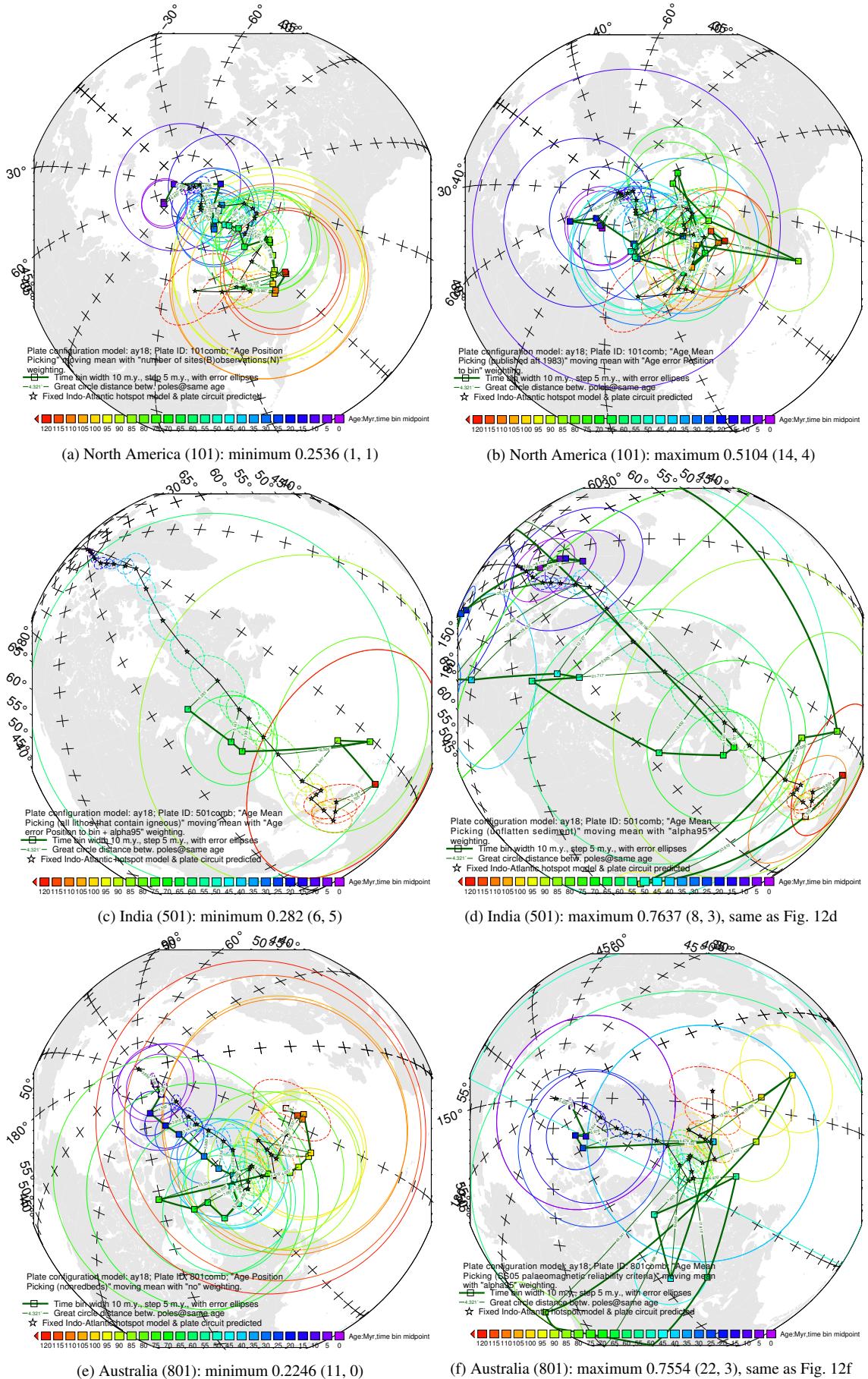


Figure 15. Difference values without shape test between each continent's paleomagnetic APWPs and its predicted APWP from FHM and related plate circuits. The paths are in 10 Myr bin and 5 Myr step. The difference values less than one-standard-deviation interval of the whole 168 values are labeled in green, more than one-standard-deviation interval labeled in red. See the numbers of picked VGPs in Fig. 11.

**Figure 16.** Path comparisons with best and worst difference values shown in Fig. 15.

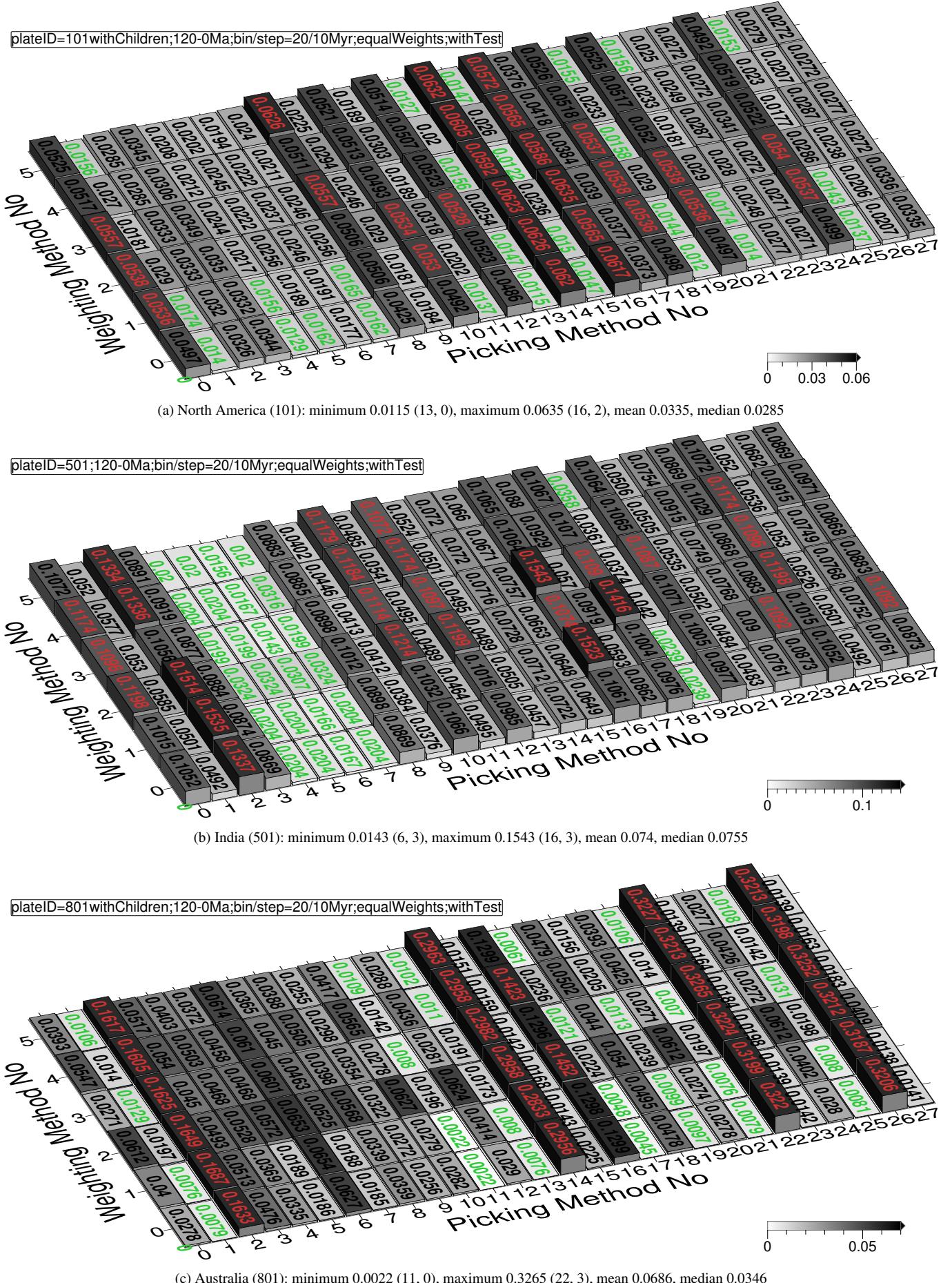
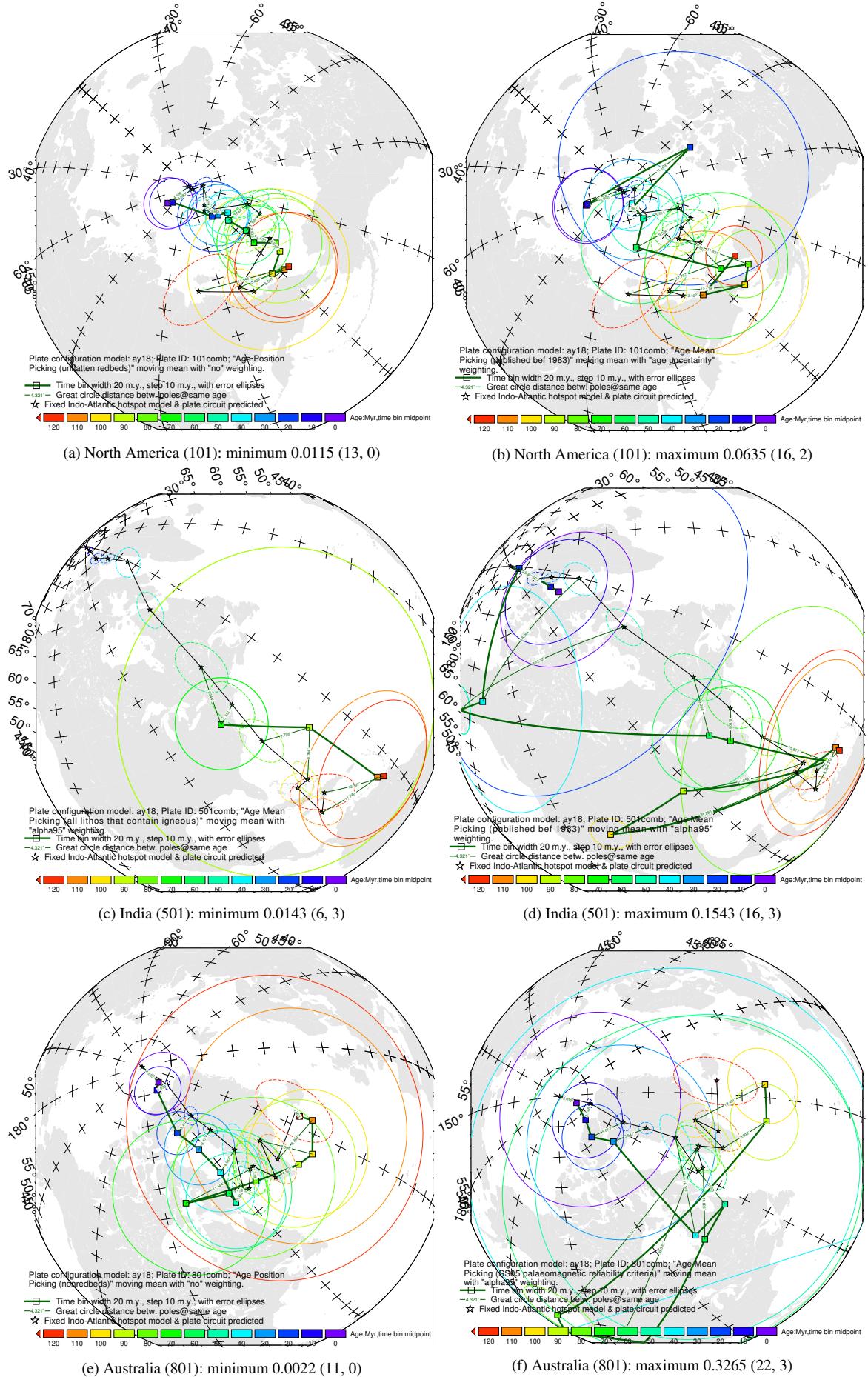
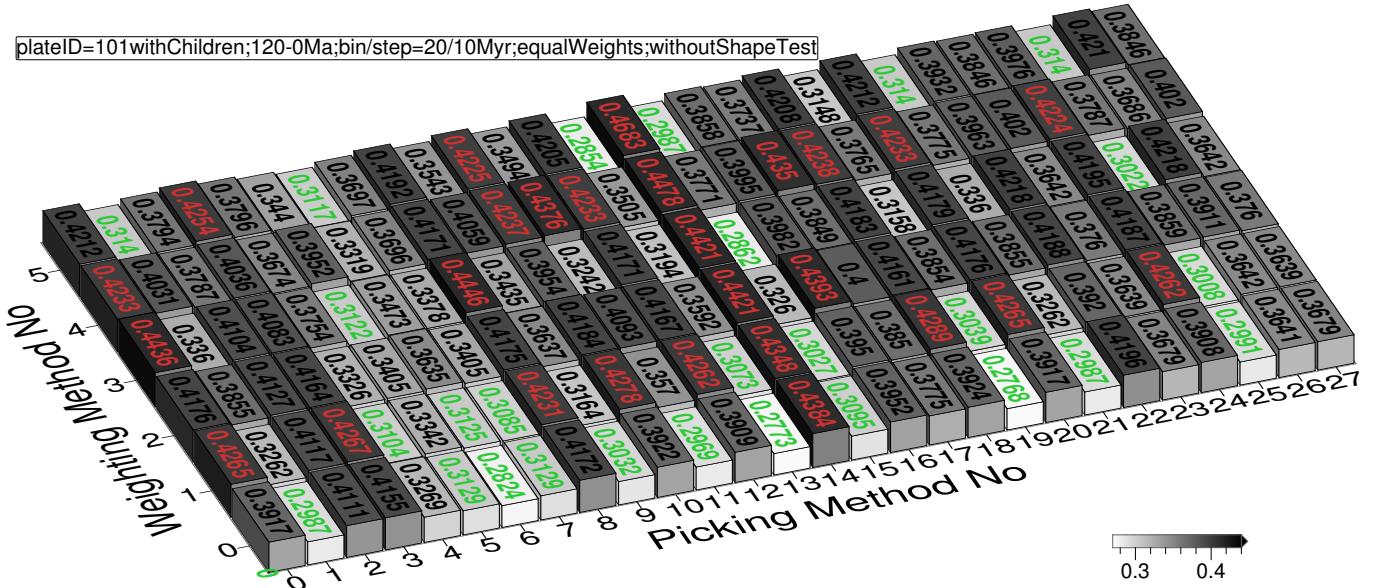
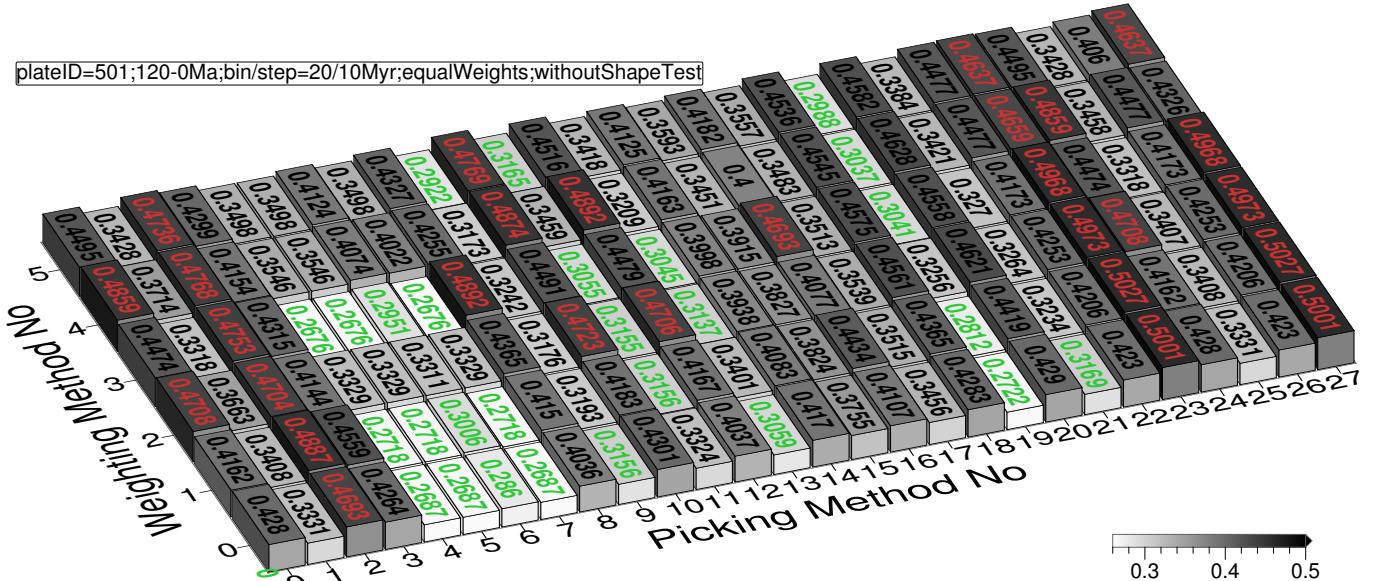


Figure 17. Same as Fig. 11. The only difference is here the paths are in 20 Myr bin and 10 Myr step. The difference values less than one-standard-deviation interval of the whole 168 values are labeled in green, more than one-standard-deviation interval labeled in red. See the numbers of picked VGPs in Fig. 11.

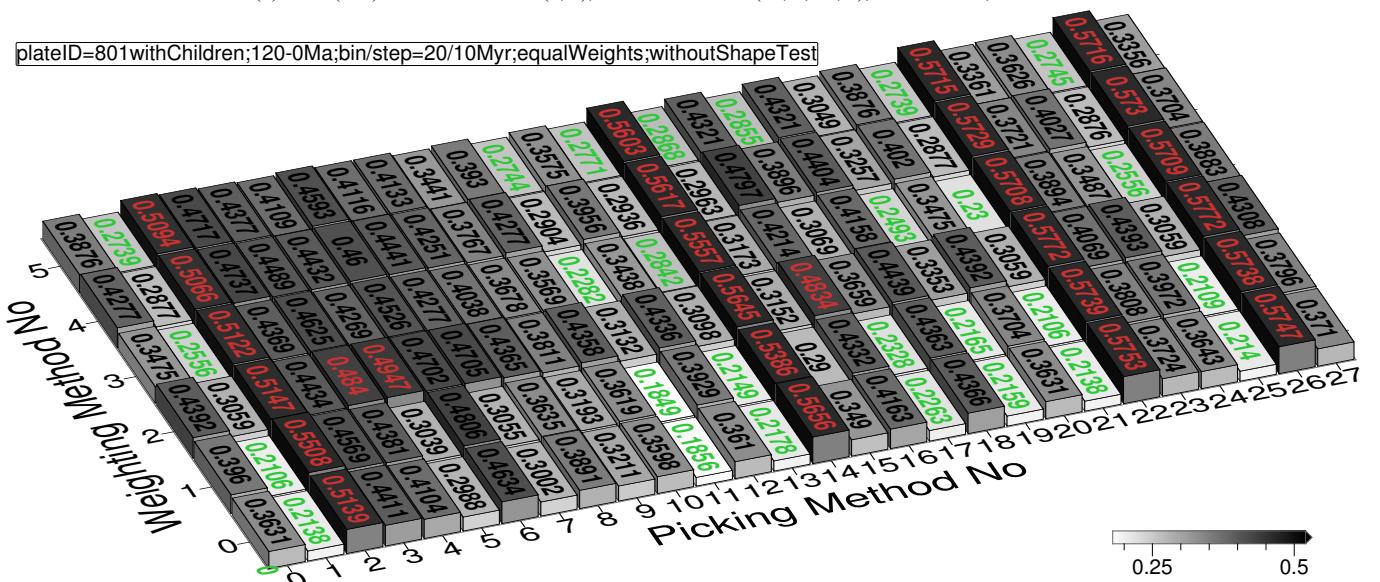
**Figure 18.** Path comparisons with best and worst difference values shown in Fig. 17.



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

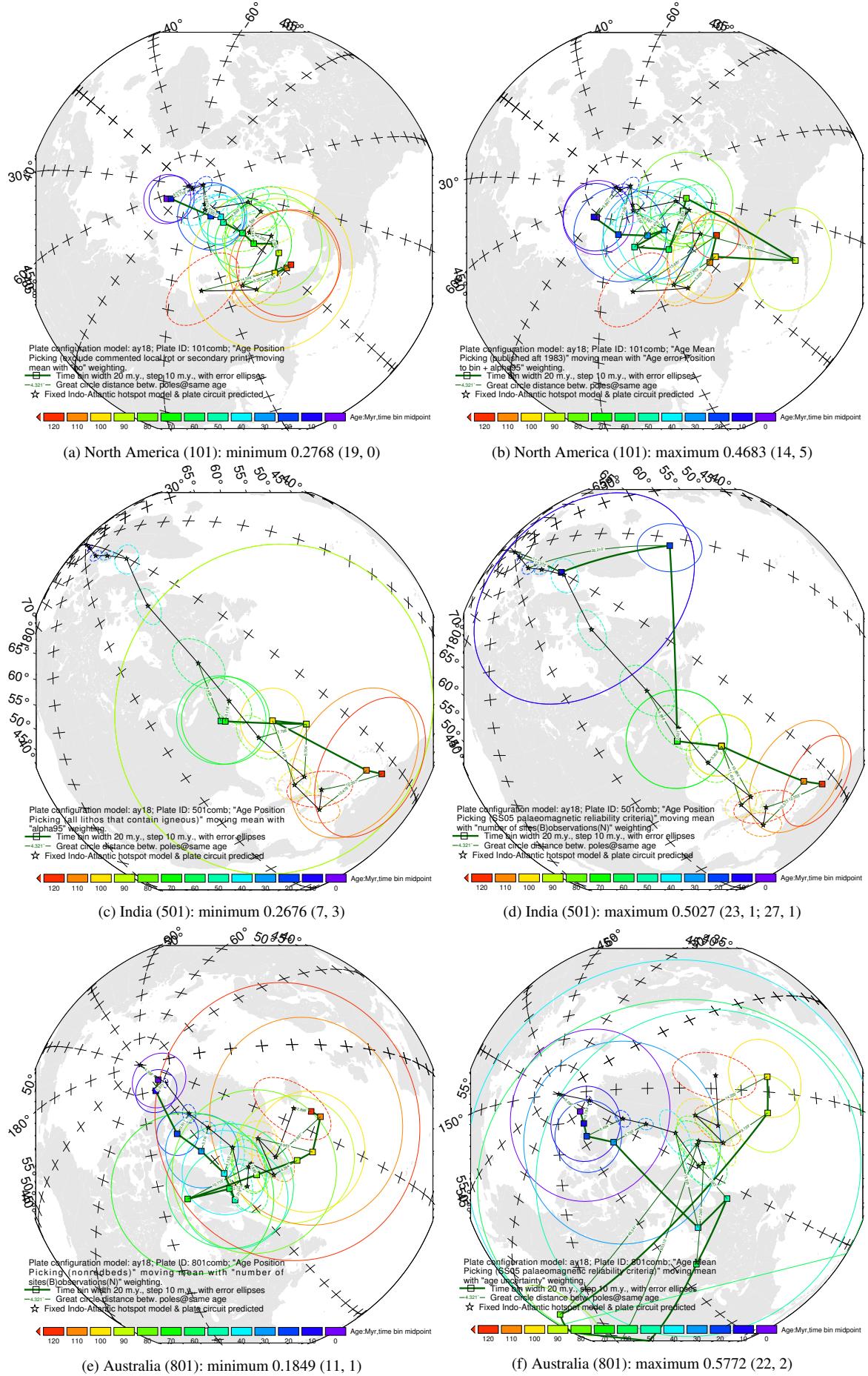


(b) India (501): minimum 0.2676 (7, 3), maximum 0.5027 (23, 1; 27, 1), mean 0.3914, median 0.408



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

Figure 19. Same as Fig. 15. The only difference is here the paths are in 20 Myr bin and 10 Myr step. The difference values less than one-standard-deviation interval of the whole 168 values are labeled in green, more than one-standard-deviation interval labeled in red. See the numbers of picked VGPs in Fig. 11.

**Figure 20.** Path comparisons with best and worst difference values shown in Fig. 19.

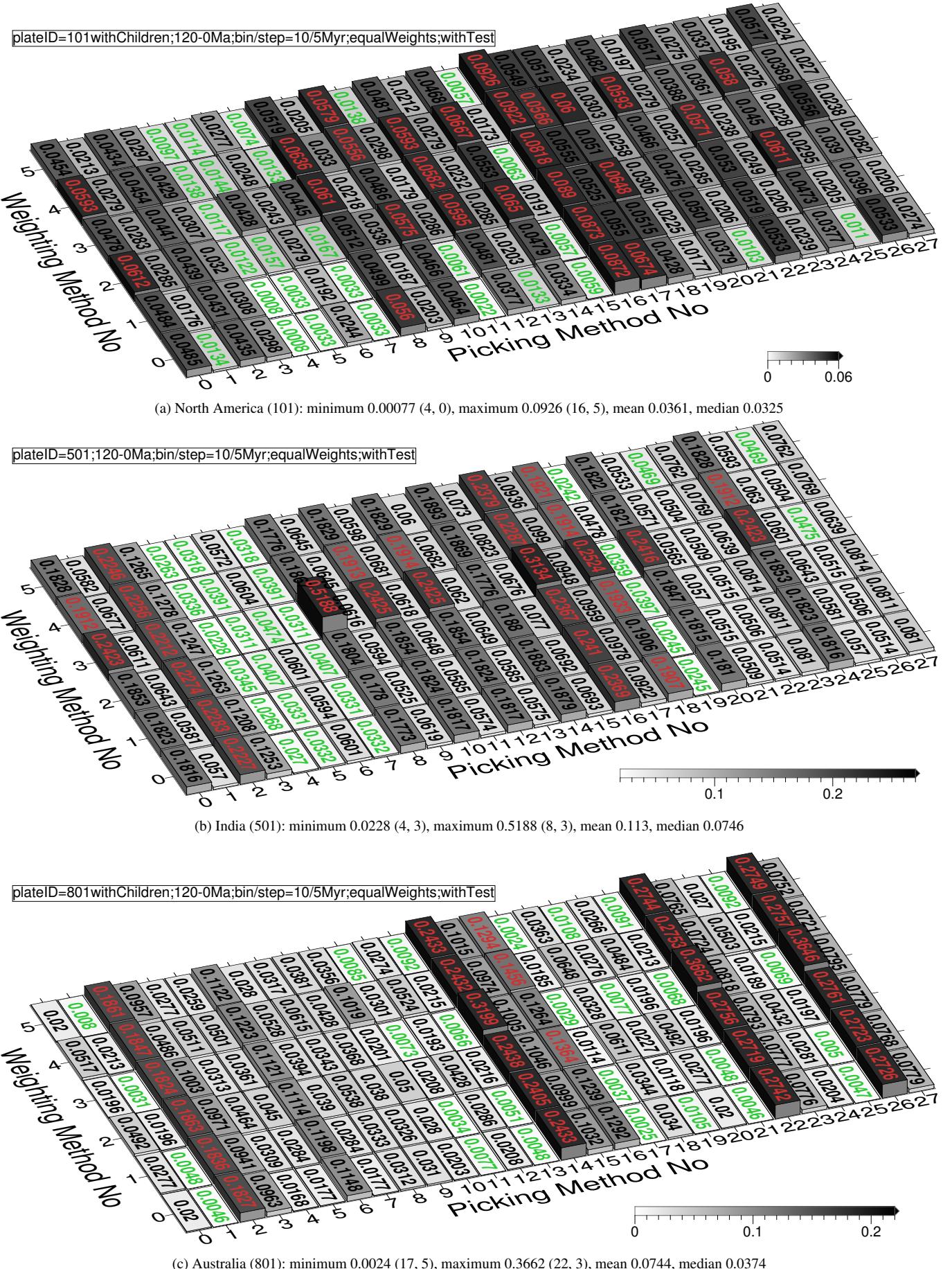


Figure 21. Same as Fig. 11 except that the reference path is predicted from MHM here. See the numbers of picked VGPs in Fig. 11.

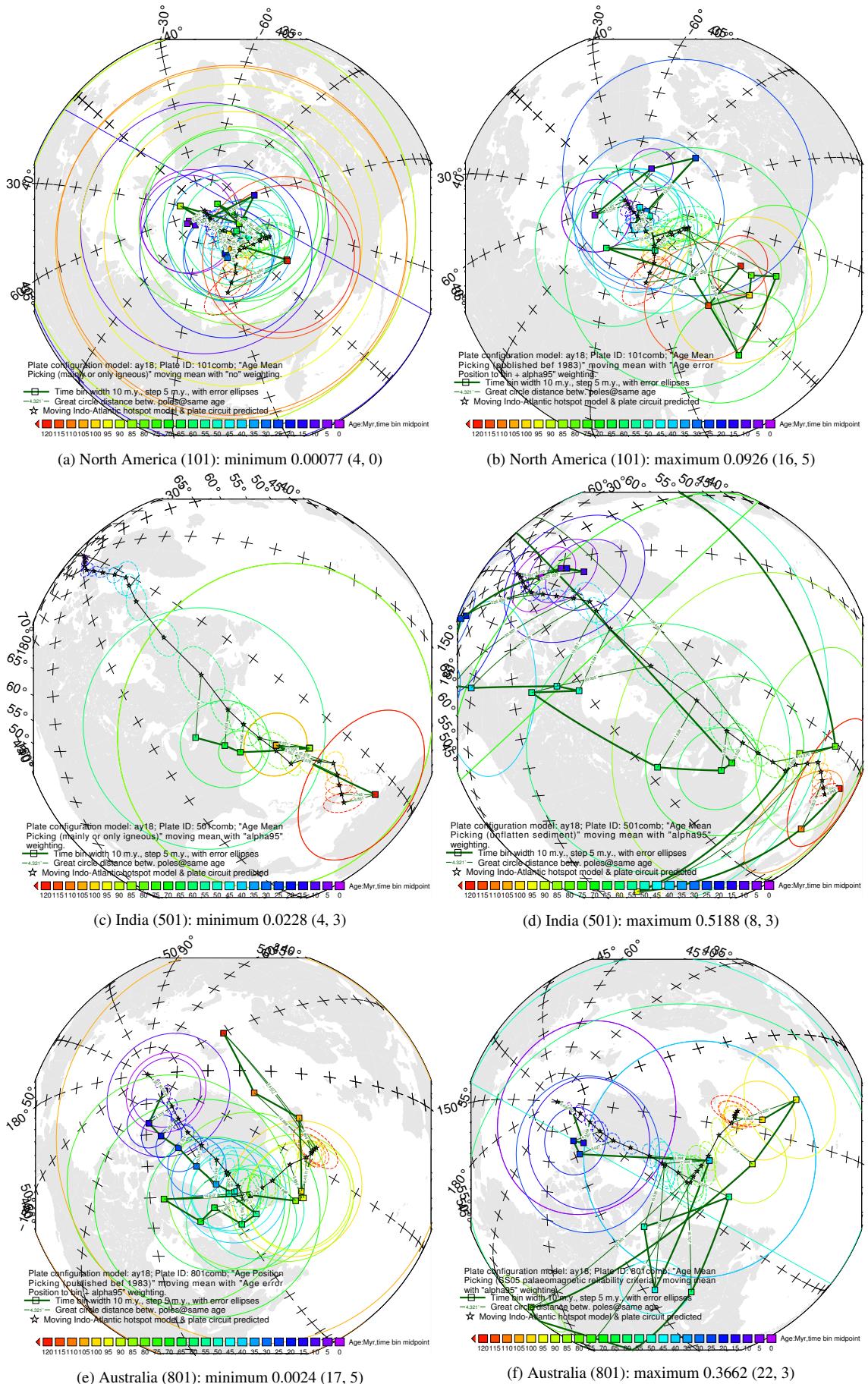


Figure 22. Path comparisons with best and worst difference values shown in Fig. 21.

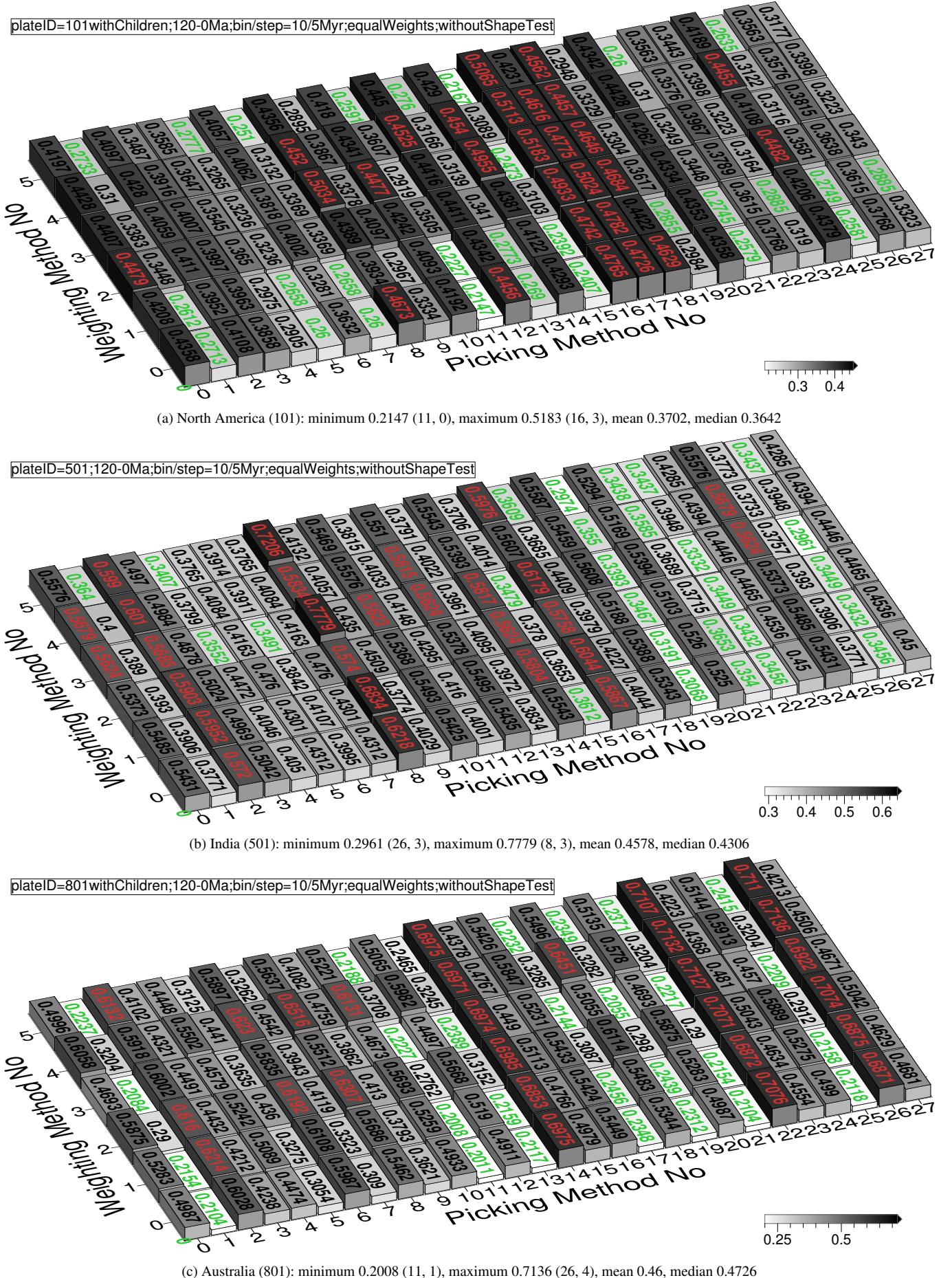
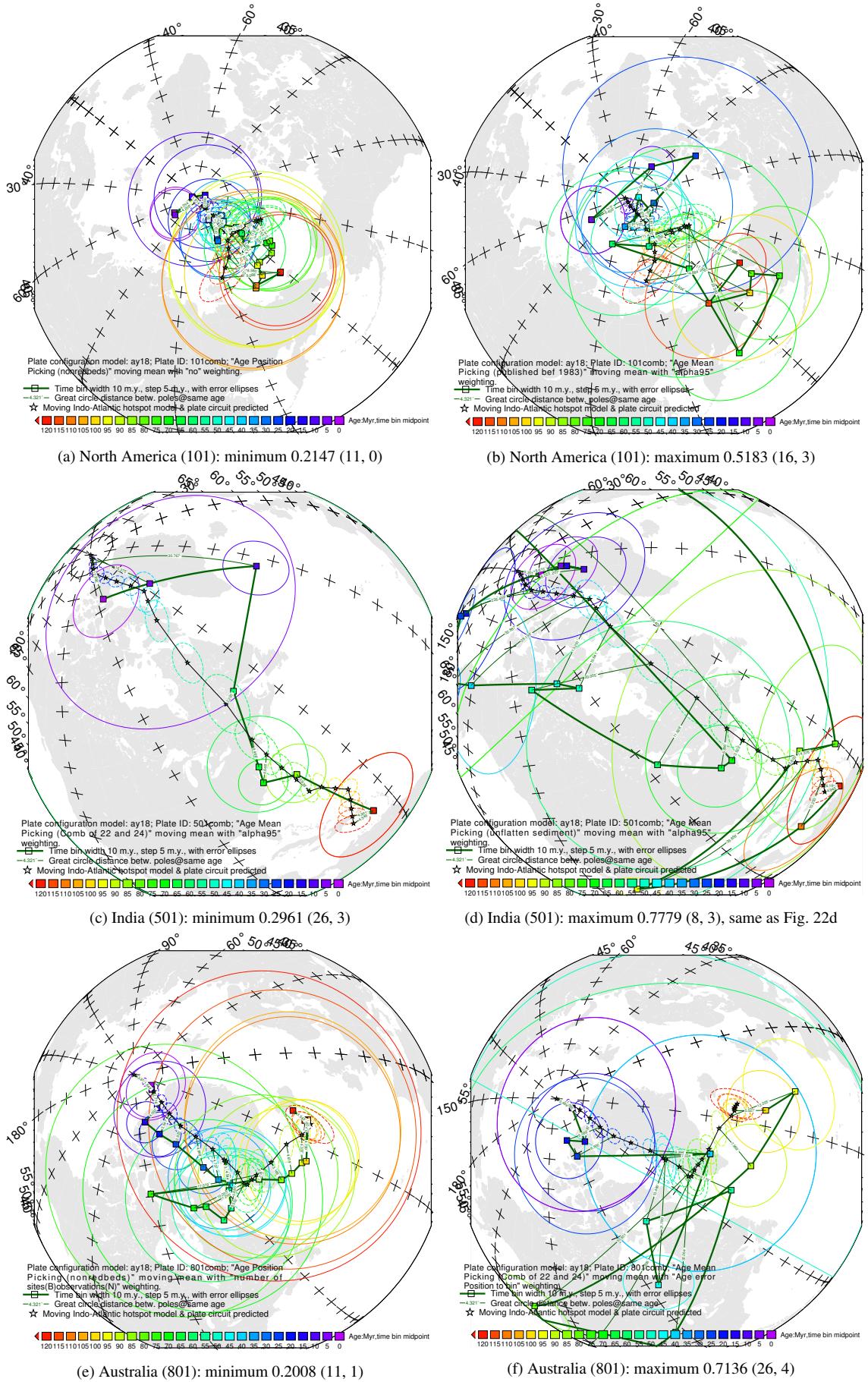


Figure 23. Same as Fig. 15 except that the reference path is predicted from MHM here. See the numbers of picked VGPs in Fig. 11.

**Figure 24.** Path comparisons with best and worst difference values shown in Fig. 23.

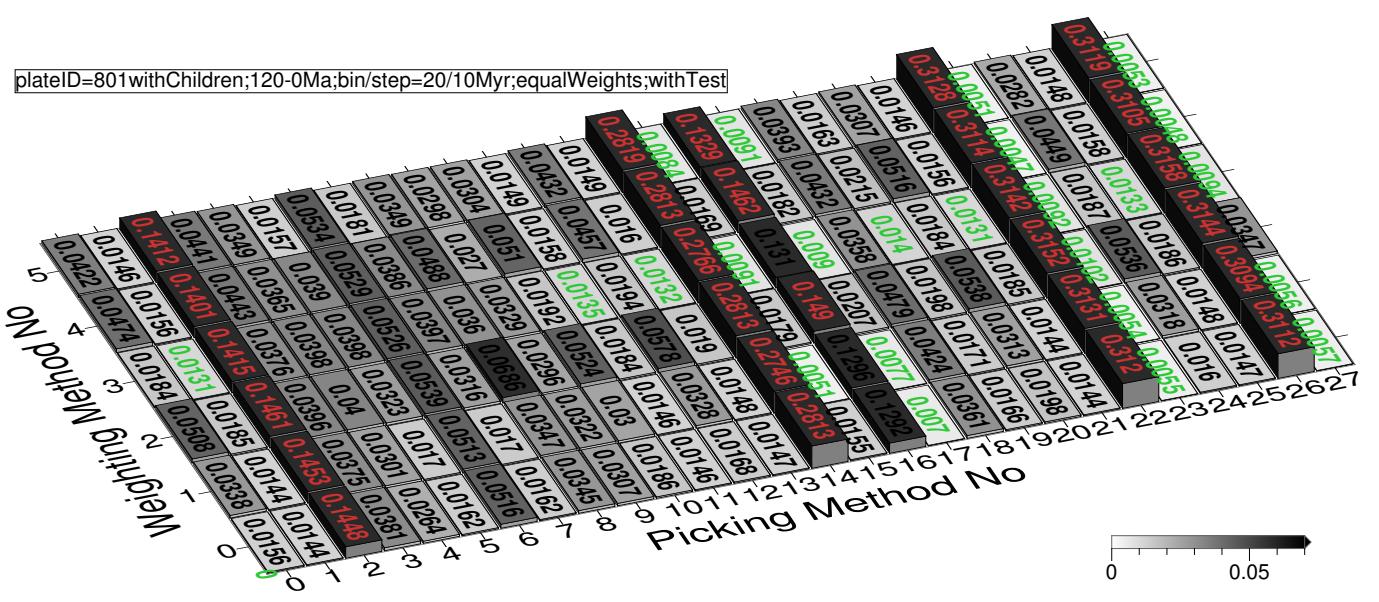
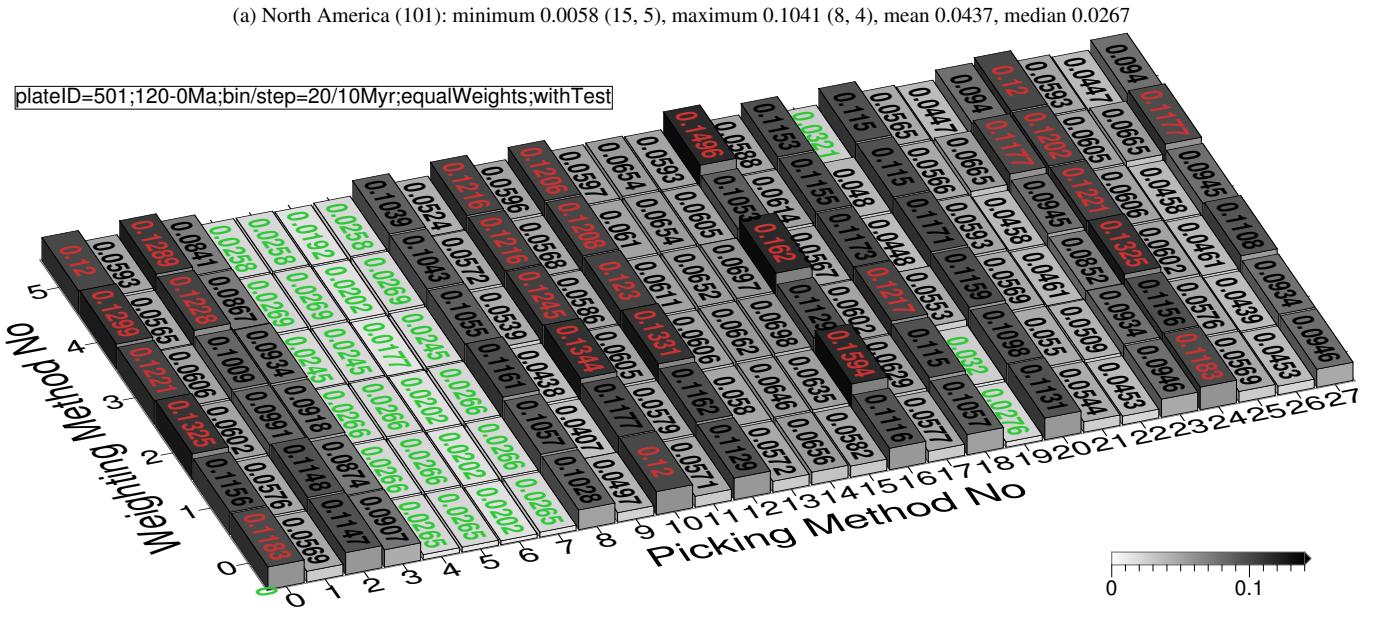
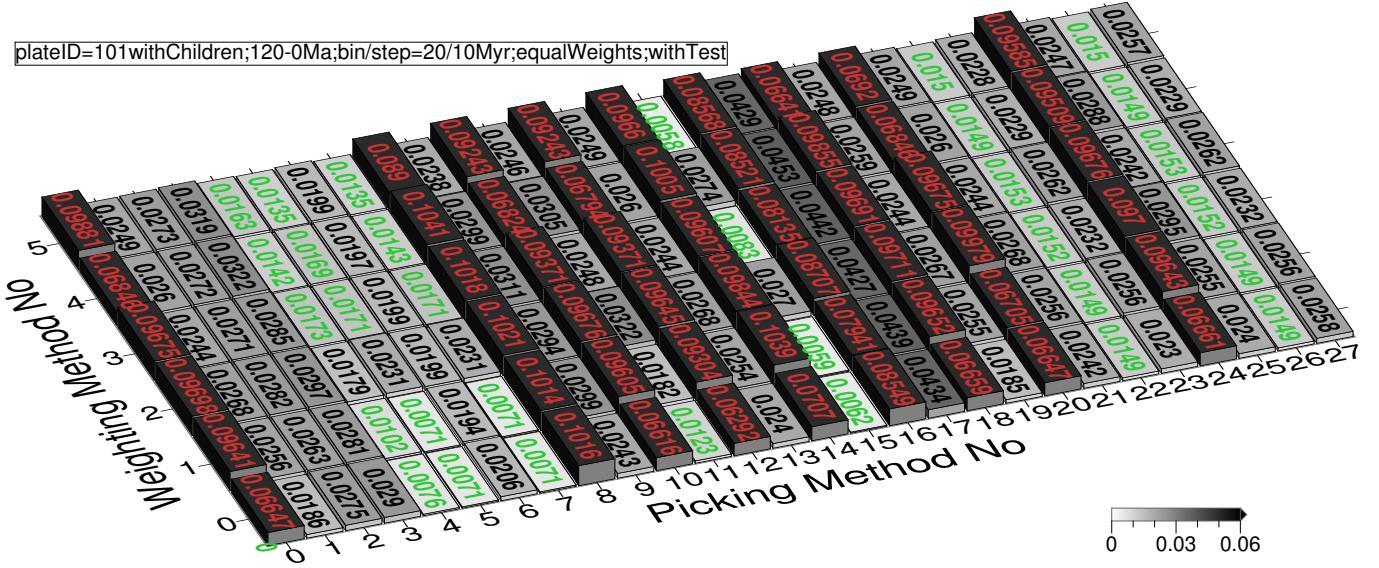


Figure 25. Same as Fig. 17 except that the reference path is predicted from MHM here. See the numbers of picked VGPs in Fig. 11.

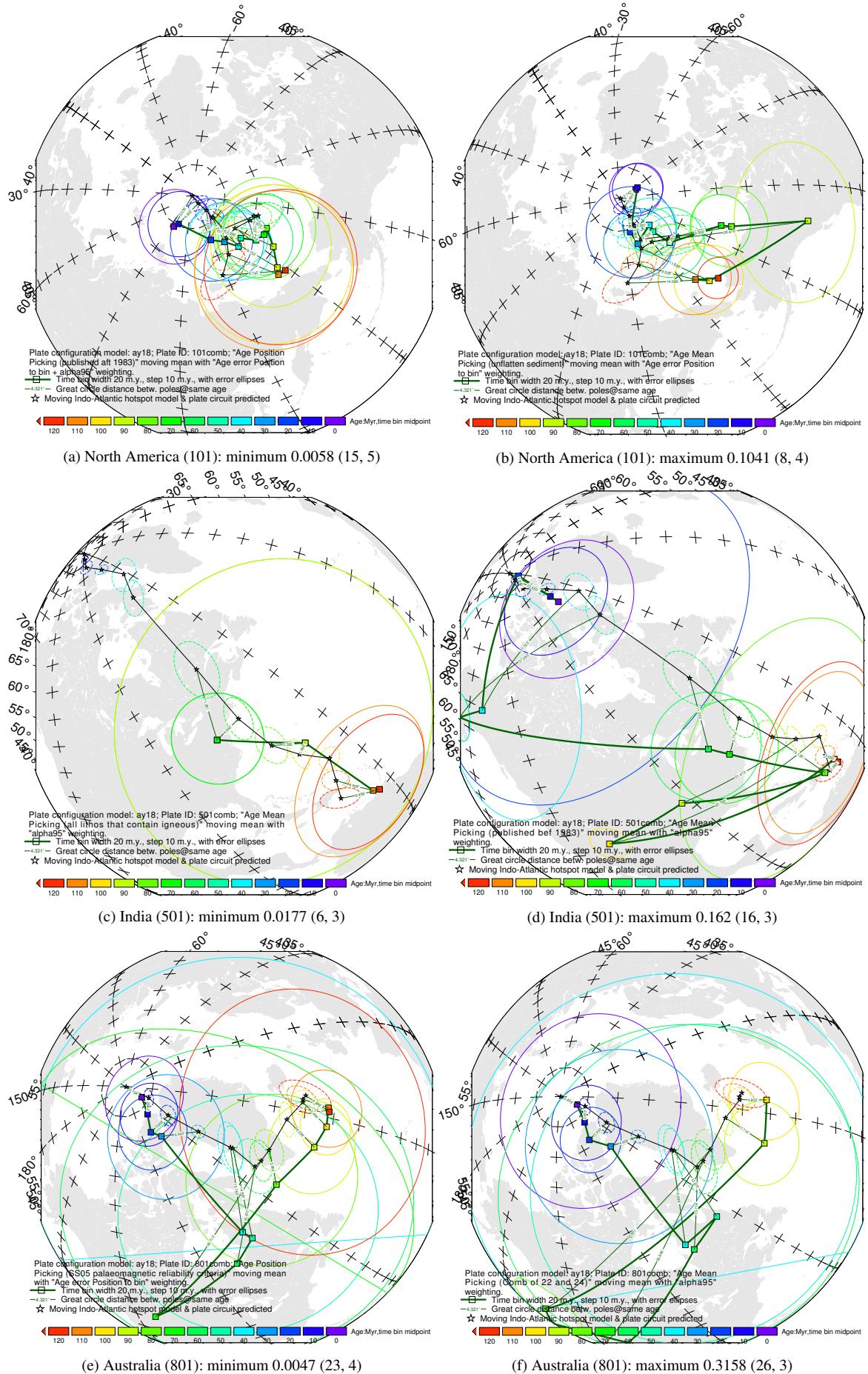


Figure 26. Path comparisons with best and worst difference values shown in Fig. 25.

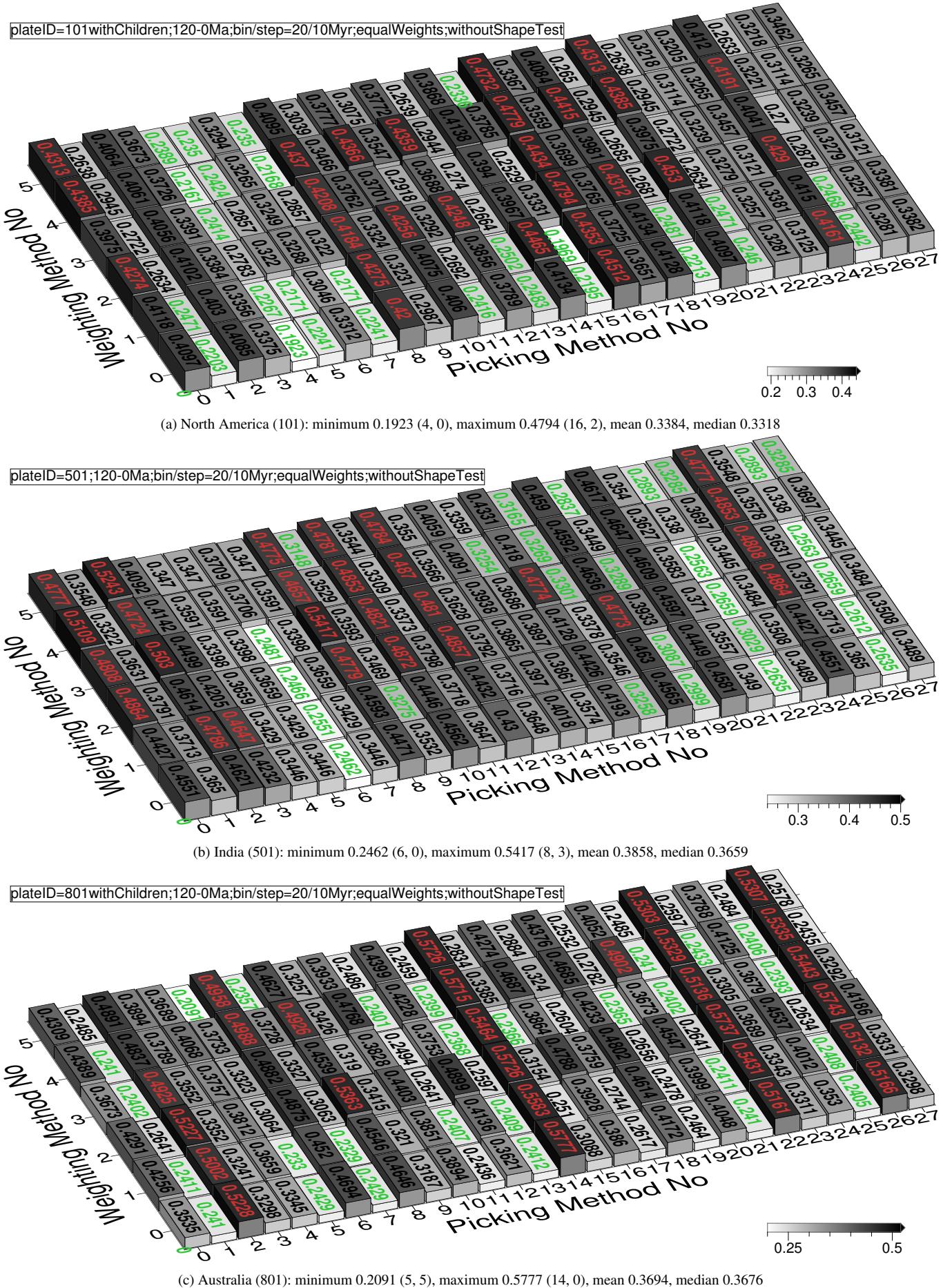


Figure 27. Same as Fig. 19 except that the reference path is predicted from MHM here. See the numbers of picked VGPs in Fig. 11.

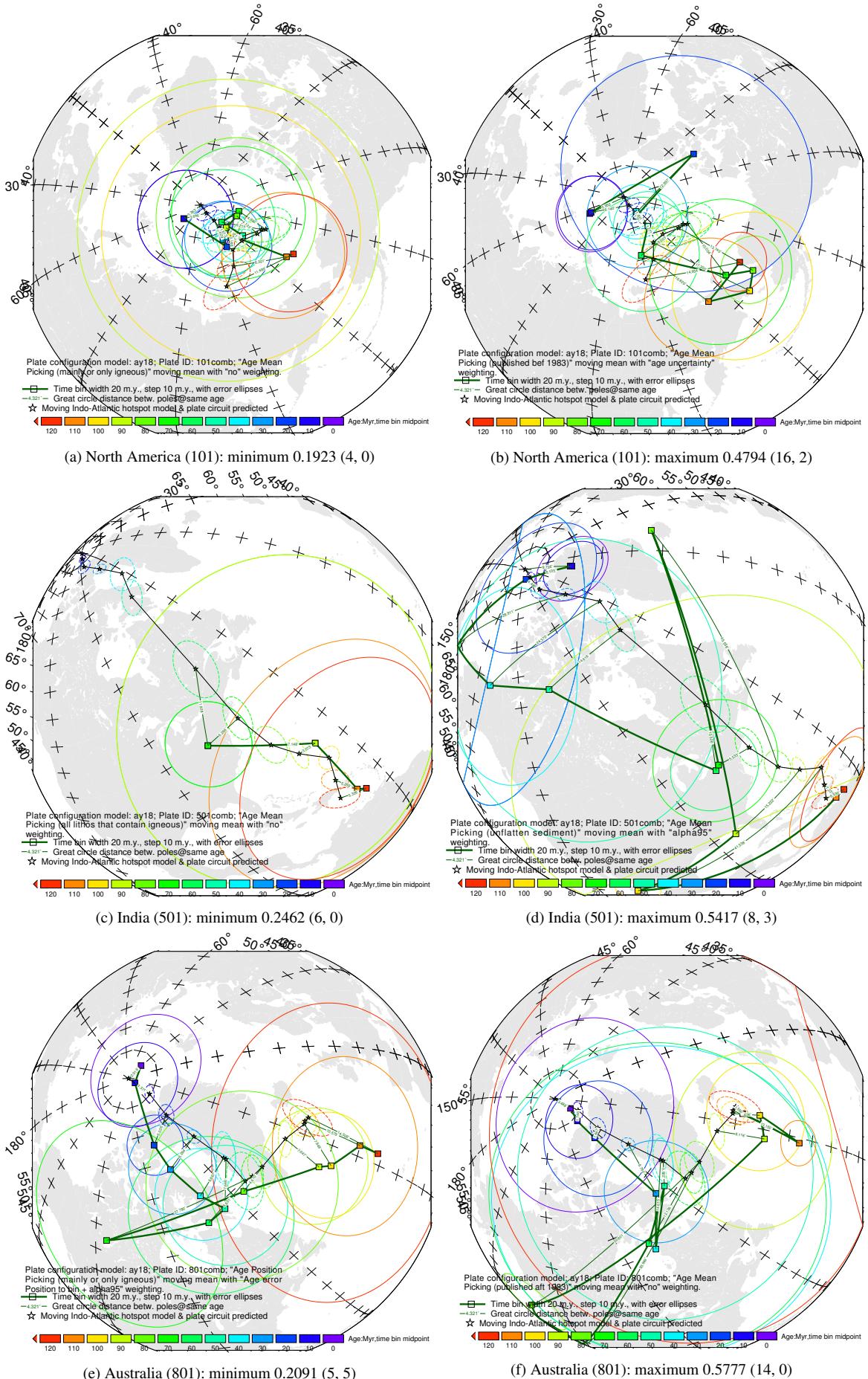
**Figure 28.** Path comparisons with best and worst difference values shown in Fig. 27.

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

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

- (ii) the self-explanatory topography of bands indicates that the picking methods (Table. 1) influence the similarity more than the weighting methods (Table. 2) do.
- (iii) filtering (picking no 2–7, 10, 11, and 14–27) and correcting (picking no 8, 9, 12 and 13) has limited effectiveness.
- (iv) weighting is not always making similarities better. In fact, for quite many of the methods, no weighting is the best performer (Table. 3).
- (v) the picking method no 19 is always the best when the FHM predicted APWP is the reference.
- (vi) the weighting method no 0, 1 and 5 are generally producing better similarity than 2, 3 and 4.
- (vii) North America (101) owns better similarity results than Australia (801) and India (501), because its worst and mean composite differences are always less than the other two continents'.
- (viii) for both North America (101) and India (501), more recent studies generally give better results than (or results close to) older studies. However, this is not true for Australia (801) (Fig. 11c and Fig. 15c).

3.2 Discussions

The following discussions will be in Q&A style.

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

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. 29a and Fig. 29b). 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 VGPs (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. 29 and Fig. 30). However, if the added VGPs were all or mostly “bad”, the improvement of fit would not

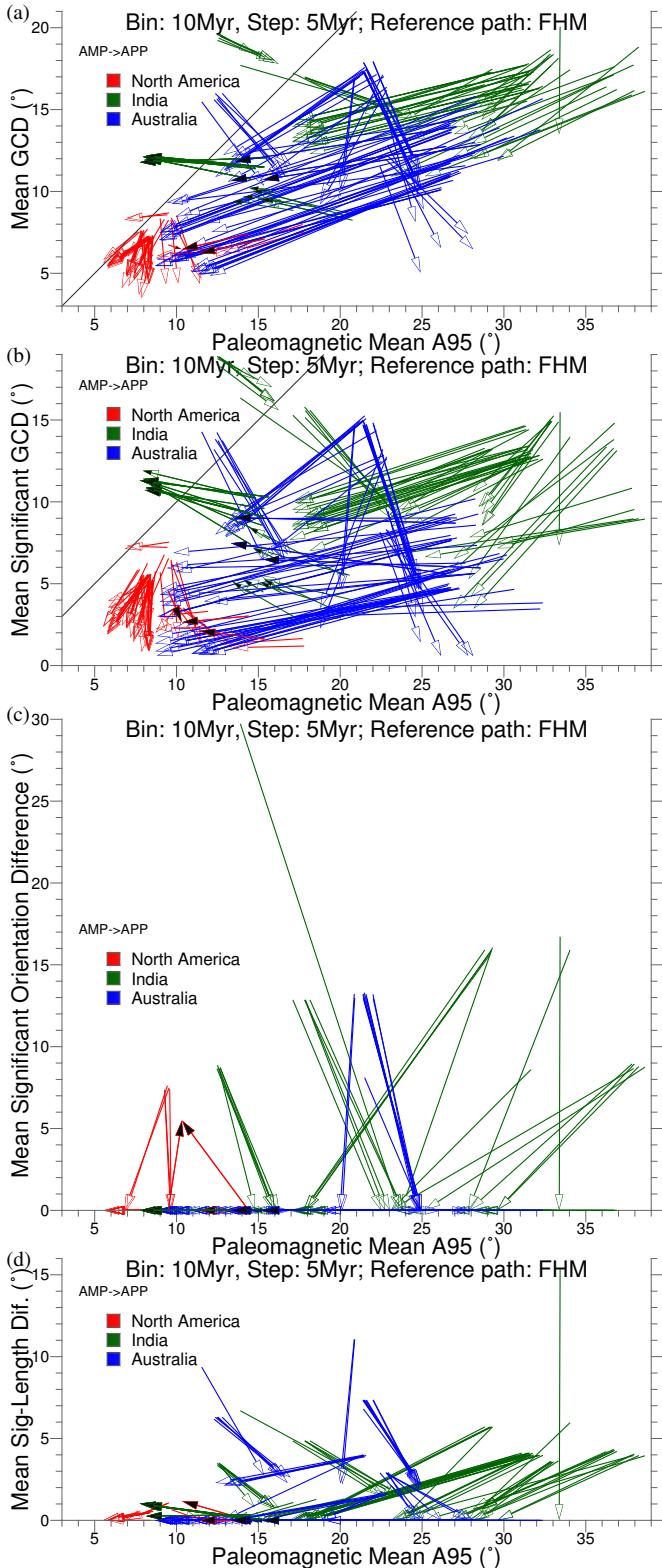


Figure 29. Paleomagnetic APWP's mean A95 versus (a) “mean GCD”, (b) “mean significant GCD”, (c) “mean significant orientation difference”, and (d) “mean significant length difference” between paleomagnetic APWP and its corresponding FHM-and-plate-circuit predicted APWP. Starting points of the arrows are results from AMP, while ending points are from APP. Black color filled arrow heads are the small number of special cases of AMP derived equal-weight CPDs better than APP (see details in Fig. 11).

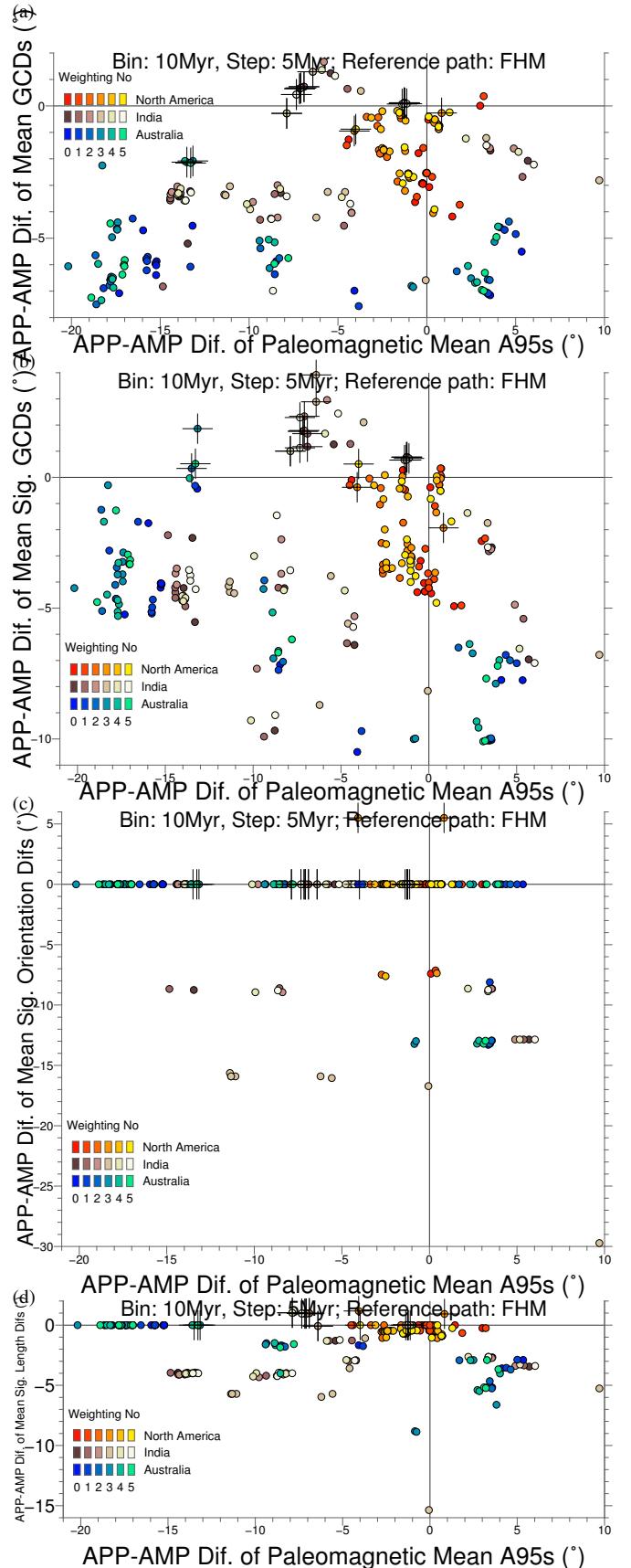


Figure 30. Differences of APP and AMP coordinates shown in Fig. 29. Crosses locates the small minority cases of AMP derived equal-weight CPDs better than APP (see details in Fig. 11).

occur. So the improvement of fit is not only because of the increase in N, but also because the majority of the additional poles 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 paleomagnetic pole 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 generate 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. 3). 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 VGP borne on the mean poles during all the period of its age uncertainty, and use the increased number of VGPs (N) to average out the negative effect of those “bad” poles, including the VGPs that should not be included at that age for mean pole.

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

3.2.2.1 Measure 1: Because of small number of VGPs (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. 11a, there are only three special (of 84 APP vs. AMP comparisons) cases picking/weighting 4/3, 4/5, 6/3 better than 5/3, 5/5, 7/3 respectively. Compared with the picking/weighting 4/3 APWP, although most of the mean paleomagnetic poles are closer to the FHM predicted APWP and also the number of the significant pole pairs is one 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 two more significant d_l segment pair (Table. 4). 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 spacial similarity d_s APP brings.

For 4/5 vs 5/5, all d_a and d_l are indistinguishable. Compared with the results from AMP, although the coeval pole GCDs are all decreased for APP, this spatial improvement is not able to offset the negative effects of also generally decreased paleomagnetic A95s, which potentially brings more statistically distinguishable coeval poles (e.g. the 15 Ma and 30 Ma poles for picking 5 and weighting 5; Table. 5). This further causes greater distinguishable mean d_s from the APP methods. The similar phenomenon occurs to Fig. 11b picking 4 vs 5 with all the six types of weightings, Fig. 11c picking/weighting 4/2–3 vs 5/2–3, 4/5 vs 5/5.

In addition, compared with AMP, APP potentially could gen-

erate more mean poles, because sometimes for some sliding window there is no VGP involved at all for AMP while there is(are) VGP(s) involved for APP. For APP, the mean poles at all ages should be composed of more VGPs 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. 6) happens. It is sometimes that an additional very “bad” VGP gets included by APP and this increases both coeval pole distance and paleomagnetic A95 even though N increases. Such cases include Fig. 11b 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 sea-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” VGPs 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. 3).

Other Type 1 (e.g. Table. 4) cases: Fig. 17a picking/weighting 22/3 vs 23/3. Fig. 21a 6/3 vs 7/3.

Other Type 2 (e.g. Table. 5) cases: Fig. 17a picking/weighting 2/0–5 vs 3/0–5, 4/0–2 vs 5/0–2, 4/4 vs 5/4, 6/5 vs 7/5, 6/2–3 vs 7/2–3, 22/4–5 vs 23/4–5, 22/1 vs 23/1, 26/4 vs 27/4, 26/2 vs 27/2. Fig. 17b 14/3 vs 15/3. Fig. 17c 4/2 vs 5/2. Fig. 21a 4/0–1 vs 5/0–1. Fig. 21b 4/0–4 vs 5/0–4, 22/0–1 vs 23/0–1, 22/3–4 vs 23/3–4, Fig. 21c 4/3 vs 5/3, 4/0 vs 5/0, 8/5 vs 9/5, 8/3 vs 9/3. Fig. 25b 14/2–3 vs 15/2–3. Fig. 25c 4/4 vs 5/4.

Combination of both Type 1 and 2 cases: Fig. 17a picking/weighting 26/0–1 vs 27/0–1. Fig. 21a 4/3 vs 5/3. Fig. 21b 22/2 vs 23/2, 22/5 vs 23/5, 26/0–5 vs 27/0–5.

Other Type 3 (e.g. Table. 6) cases: Fig. 17b picking/weighting 6/0–5 vs 7/0–5, 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5. Fig. 21a 4/2–5 vs 5/2–5. Fig. 25a 2/0–5 vs 3/0–5, 4/2 vs 5/2, 4/4 vs 5/4, 6/2 vs 7/2, 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5. Fig. 25b 6/0–5 vs 7/0–5, 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5.

3.2.2.2 Measure 2: For example, for Fig. 15a, there are 13 special (of 84 APP vs. AMP comparisons) cases. For the case of 2/3 vs 3/3, coeval segment angular differences d_a are greater from APP method than from AMP method (Type 4). For the case of 2/0 vs 3/0, 4/5 vs 5/5, 16/2 vs 17/2, 22/0–4 vs 23/0–4, 26/0–3 vs 27/0–3, fraction of distinguishable coeval poles are more from APP method than from AMP method (Type 5). In addition, segment angular differences d_l could be greater from APP method than from AMP method (Type 6).

Other Type 4 cases: Fig. 19a 10/4 vs 11/4.

Other Type 5 cases: Fig. 15b picking/weighting 4/0–5 vs 5/0–5, 6/0–5 vs 7/0–5, 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5. Fig. 19a 2/4–5 vs 3/4–5, 2/0–2 vs 3/0–2, 4/1–2 vs 5/1–2, 4/4 vs 5/4, 6/0 vs 7/0, 6/4–5 vs 7/4–5, 22/4 vs 23/4. 26/4 vs 27/4, 26/0 vs 27/0. Fig. 23b 6/0–1 vs 7/0–1, 6/3–4 vs 7/3–4. Fig. 27a 22/1 vs 23/1, 22/3–4 vs 23/3–4, 26/0–1 vs 27/0–1, 26/3–5 vs 27/3–5.

Combination of Type 4 and 5 cases: Fig. 19a 16/4 vs 17/4. Fig. 19b 22/0–5 vs 23/0–5, 26/0–3 vs 27/0–3, 26/5 vs 27/5. Fig. 19c 4/2 vs 5/2, 6/2 vs 7/2. Fig. 23a 2/1 vs 3/1, 16/1–2 vs 17/1–2. Fig. 23b 4/0–5 vs 5/0–5, 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5. Fig. 27a 4/4 vs 5/4, 4/2 vs 5/2

Table 4. One example of the Type 1 rare cases where AMP gives better similarity result than APP does from North America (101). 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)	Pmag A95 (°)	Dist (°)	Age (Ma)	Diff (°)
0	0	0	8.3	4.29	0	7.6	3.885	10-15-20	126.59
5	1.56039/0.87367	5	7	4.853	5	6.8	4.3453		
10	2.89214/1.58743	10	12	9.91	10	8.6	5.79	dl	
15	2.575/1.63303	15	58/49	6.72	15	2.0857	11.8	Age (Ma)	Diff (°)
20	3.16077/2.20094	20	12.43	8.58				10-15	13.52
25	4.96061/2.2183	25	6.76	6.96	25	6.3358	6.873	15-20	14.68
30	3.39692/2.57114	30	6.68	6.46	30	6.68	6.4583		
115	9.27023/5.16012	115	8.6	12.2535	115	8.5	11.7		
120	14.6882/8.12086	120	7.76	15.5744	120	7.728	15.258		

Table 5. One example of the Type 2 rare cases where AMP gives better similarity result than APP does from North America (101). Only statistically significant values are listed here. Note that the number of the ages of the significant differentiated mean poles are the same for both paleomagnetic APWPs in this type of situation.

Age (Ma)	FHM predicted A95 (°)	ds			
		picking 4 + weighting 5		picking 5 + weighting 5	
		Pmag A95 (°)	Dist (°)	Pmag A95 (°)	Dist (°)
0	0	8.3526	3.79	7.6664	3.391
5	1.56039/0.87367	7.5215	4.17	7.29	3.74
10	2.89214/1.58743	21.351	4.76	14.598	2.87
15	2.575/1.63303			10.153	10.71
30	3.39692/2.57114			6.9664	5.7245
115	9.27023/5.16012	9.635	11.935	9.452	11.276
120	14.6882/8.12086	8.0223	15.8183	7.943	15.511

Table 6. One example of the Type 3 rare cases where AMP gives better similarity result than APP does from India (501). 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 0” case.

FHM predicted		picking 22 + weighting 0				picking 23 + weighting 0			
Age (Ma)	A95 (°)	Pmag A95 (°)	Dist (°)	N	Pmag A95 (°)	Dist (°)	N	Age (Ma)	Diff (°)
0	0	6.85	12.973	2	23.6214	17.698	3	80-85	6.286
5	1.415/0.965	23.6214	15.941	3	23.6214	15.941	3	110-115	16.684
10	2.2425/1.34645	5.4/3.1	29.897	1	5.4/3.1	29.897	1		
15	2.2694/1.62543				5.4/3.1	28.28	1		
70	8.53016/4.97567	4.0164	4.864	16	3.246	4.436	20		
75	5.3554/3.1595				5	4.477	1		
80	8.41657/5.00588				5	3.358	1		
85	5.01489/2.49492	5	7.632	1	5	7.632	1		
90	7.77997/2.86845				5	10.884	1		
95	4.46779/3.24941				5	11.099	1		
100	5.6124/5.19639				5	11.4155	1		
105	4.64657/3.49277				5	14.908	1		
110	9.11039/4.98436				6.8/4.9	13.962	1		
115	9.27023/5.16012	10.73	10.508	5	10.73	10.508	5		

Other Type 6 cases: Fig. 19b 6/2 vs 7/2.
 Combination of Type 5 and 6 case: Fig. 23b 6/2 vs 7/2.
 Combination of Type 4 and 6 cases: Fig. 27a 4/3 vs 5/3, 4/0 vs 5/0
 Combination of Type 4, 5 and 6: Fig. 27b 6/0–3 vs 7/0–3, 14/2 vs 15/2, 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5.

3.2.3 Question: Why weighting is not affecting?

Generally, weighting does not affect the similarities dramatically, because the six results from the six weighting methods are mostly very close. There are a few special cases that one of them dramatically changes similarity (e.g. for picking method (Pk) 5 and weighting method (Wt) 3, etc.; Fig. 11a). As follows, those normal cases with close results are discussed first. Those rare cases are examined later.

For Pk 0 in the North America (101) example (Fig. 11a), the reason of why Wt 1–4 do not produce better similarity scores than Wt 0 (i.e. no weighting) does is examined here. Take the comparison between Wt 0 and Wt 1 (Table. 7) as an example, although APP indeed decreases the values of most distances from the reference and most A95s, APP also brings one more significant differentiated segment-length difference (80–85 Ma coeval segments; Fig. 31) because of the slightly larger A95 and shorter coeval pole GCD for 80 Ma Pk 0 Wt 1. Although they are distinguishable, the overlapping of their distributions is quite small (Fig. 31b). Wt 2 and 4 cause one more pair of distinguishable mean poles (d_s) than other weighting methods. However, compared with Wt 0, Wt 5 (i.e. combination of α_{95} and age error position to bin) does affect and bring the best similarity, because Wt 5 decreases the GCDs of the 20, 80 and especially 100 Ma distinguishable pole pairs (by more than 2 degrees; Table. 7).

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

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

As the above examples shown, although no weighting is often better than weighting, the results are actually rather close to each other and also APP still improves precision and accuracy of most poles for the cases with larger difference scores, although APP brings extra significant differences in shape metrics which is also the main cause of their larger difference scores (Table. 7). However, APP does not always improve precision or accuracy, for example, for the Pt 5 (mainly or only igneous derived) and Wt 3 (according to α_{95}) results (Table. 8). For Pt 5, the result with no weighting is still the best, and the results from Wt 0–2, 4, 5 are close and APP generally improves precision and accuracy, just like the above mentioned general cases. Only for Wt 3, APP worsens precision and improves accuracy of most poles at the same time. This makes the difference score dramatically larger than others. In addition, Wt 3 brings more significant differences in shape metrics, which causes the difference score even larger (Table. 8). That Wt 3 sometimes gives dramatically worse similarity could be related to the way we do the weighting according to the size of α_{95} , i.e. the smaller the α_{95} is, the larger the assigned weight is (see more details about Weighting No. 3 in the Weighting paragraph). However, small size of α_{95} (high accuracy) could be becausei of those sampled directions not covering enough long period (thought to be at least about 10^4 years) to “av-

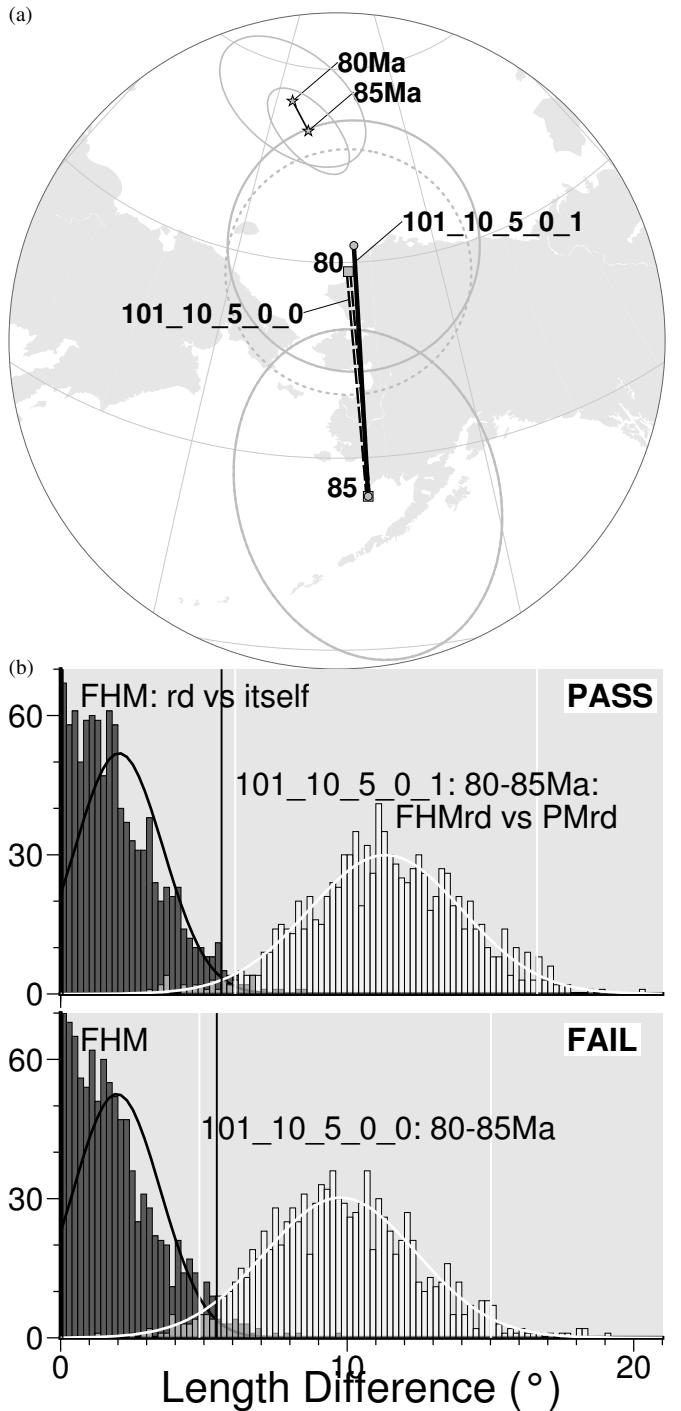


Figure 31. Significance testing on the 80–85 segment length difference between North American paleomagnetic APWPs derived from picking 0 and weightings 0 and 1 and the FHM predicted APWP (Fig. 1; Table. 7). (a) The thin segment through stars is from FHM predicted path, and the bold solid and dashed are from paleomagnetic paths.

erage out” secular variation for giving a paleomagnetic pole. That is to say, the smallest α_{95} s could get the greatest weights that they should not deserve.

Generally, weighting does not improve fit. In other words, Wt 0 is generally the best. Wt 2 or 4 is not recommended, because they never have generated the best similarities (Table. 3), compared with other weighting methods. There is no general pattern about which

Table 7. My caption

FHM predicted for 101	Pk 0 + Wt 0				Pk 0 + Wt 1				Pk 0 + Wt 5	
	ds		ds		dl		ds			
Age (Ma)	A95 (°)	Pmag A95 (°)	Dist (°)	Pmag A95 (°)	Dist (°)	Age (Ma)	Diff (°)	Pmag A95 (°)	Dist (°)	
0	0	4.27286602	5.01	3.950661	5.05647	80.85	11.103	4.143	5.356	
5	1.56039/0.87367	4.22350537	5.146	3.936601	5.1286			4.0534	5.407	
10	2.89214/1.58743	20.9920176	3.076	19.826829	3.22624			19.868	3.36	
15	2.575/1.63303	13.8698147	10.3	13.827757	10.34			13.85	10.2753	
20	3.16077/2.20094	8.36501201	7.2162	8.463096	6.9973			8.413	6.7906	
50	7.15565/3.22656	4.3991229	6.22	4.651517	6.273			4.3326	6.3563	
55	7.17564/4.28065	5.6991191	8.647	5.670719	9.6724			5.52	8.53	
60	9.71876/6.35204	7.71537555	9.498	6.889233	8.607			7.77	9.8237	
80	8.76515/5.14459	6.29356332	9.26	6.452368	8.098			6.033	8.369	
85	5.54221/2.65419	8.7/6.7	18.995	8.7/6.7	18.995			8.7/6.7	18.995	
100	5.79659/5.36693	10.2720657	10.75658	9.286878	9.035			9.045	8.68	
115	9.27023/5.16012	19.767437	9.074	18.483547	10.054			19.652	9.3547	
120	14.6882/8.12086	3.56957955	17.3331	3.060561	17.062			3.606	17.47	

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. For example, Wt 3 prefers Australia (Table. 3). Wt 3 works fine with India. However, Wt 3 is not recommended for North America.

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

3.2.4.1 Question: Why the picking method 21 is not among the best for 20 Myr binning and 10 Myr stepping with both space and shape tested? As shown in Fig. 11, Fig. 15 and Fig. 19, both the picking methods 19 and 21 are among the best. However, the picking method 21 is not one of the best any more for 20 Myr binning and 10 Myr stepping with both space and shape tested (Fig. 17). Further, in fact, in Fig. 17, we can see the picking method 21 is still one of the best for North America (101) and Australia (801), but just not for India (501). However, even for India (501), the difference values (ranging 0.0483–0.0535) produced by the picking method 21 are still closer to the left bound of the one-standard-deviation interval 0.0359–0.1072 and relatively farther from the mean 0.074, which means the picking method 21 is still a relatively better one.

3.2.4.2 Question: Why the picking method 16 is not among the worst for 10 Myr binning and 5 Myr stepping with only space tested? As shown in Fig. 11, Fig. 17 and Fig. 19, the picking method 16 is always one of the worst. However, the picking method 16 is not among the worst any more for 10 Myr binning and 5 Myr stepping with only space tested (Fig. 15). Further, in fact, in Fig. 15, we can see the picking method 16 is still one of the worst for North America (101) and India (501), but just not for Australia (801). However, even for Australia (801), the difference values (ranging 0.564–0.6144) produced by the picking method 16 are still closer to the right bound of the one-standard-deviation interval 0.3102–0.6548 and relatively farther from the mean 0.502, which means the picking method 16 is still a relatively worse one.

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

A balance needs to be made between having windows that are too wide and steps that are too long which will smooth the data so much we miss actual details in the APWP (e.g. those 20 Myr window 10 Myr step paleomagnetic paths in Fig. 18 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. those 10 Myr window 5 Myr step paleomagnetic paths in Fig. 12 and Fig. 16). There is a dependence here on data density: higher density allows smaller windows/steps (this is one of the things we want to test with selective data removal mentioned in Chapter 4). Fitting curves by moving averaging change with different time window lengths and time increment lengths (i.e. steps) (e.g., the similarity of the pair in Fig. 18f is improved a bit compared to Fig. 12f). A variety of ways of binning the data (here both 20 Myr window 10 Myr step and 10 Myr window 5 Myr step) are being tested to see which one produces the better and more appropriately smoothed fit.

For Pk 2 (AMP with “ α 95/Age range” no more than “ $15^\circ/20^\circ$ ”), the 20/10 Myr bin/step methods always generate better similarities than the 10/5 Myr ones.

Interestingly, 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. 32), which mainly depends on the picking methods. Note that as mentioned in the Supplementary materials and also Chapter 4, there are 135, 75 and 99 VGP s that compose of 120–0 Ma APWPs of North America, India and Australia respectively. Does the reason could be because of the relatively larger number of VGPs for North America? Since theoretically for each sliding window, the more “bad” VGPs it contains, the worse similarity we should obtain. In the contrary, the less VGPs the window contains, the weaker the effect of averaging out “bad” poles’ influence would be. So is there a threshold number of VGPs for making an paleomagnetic APWP? For example, for making a 120–0 Ma APWP, do the results indicate the best number of VGPs we need should be some value between 99 and 135? We did the test in Chapter 2 on the 530–0 Ma paleomagnetic APWP using the AMP method, and we did find that larger windows and steps bring the paleomagnetic APWPs closer to the reference path. Here another test will be 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 will be used to generate

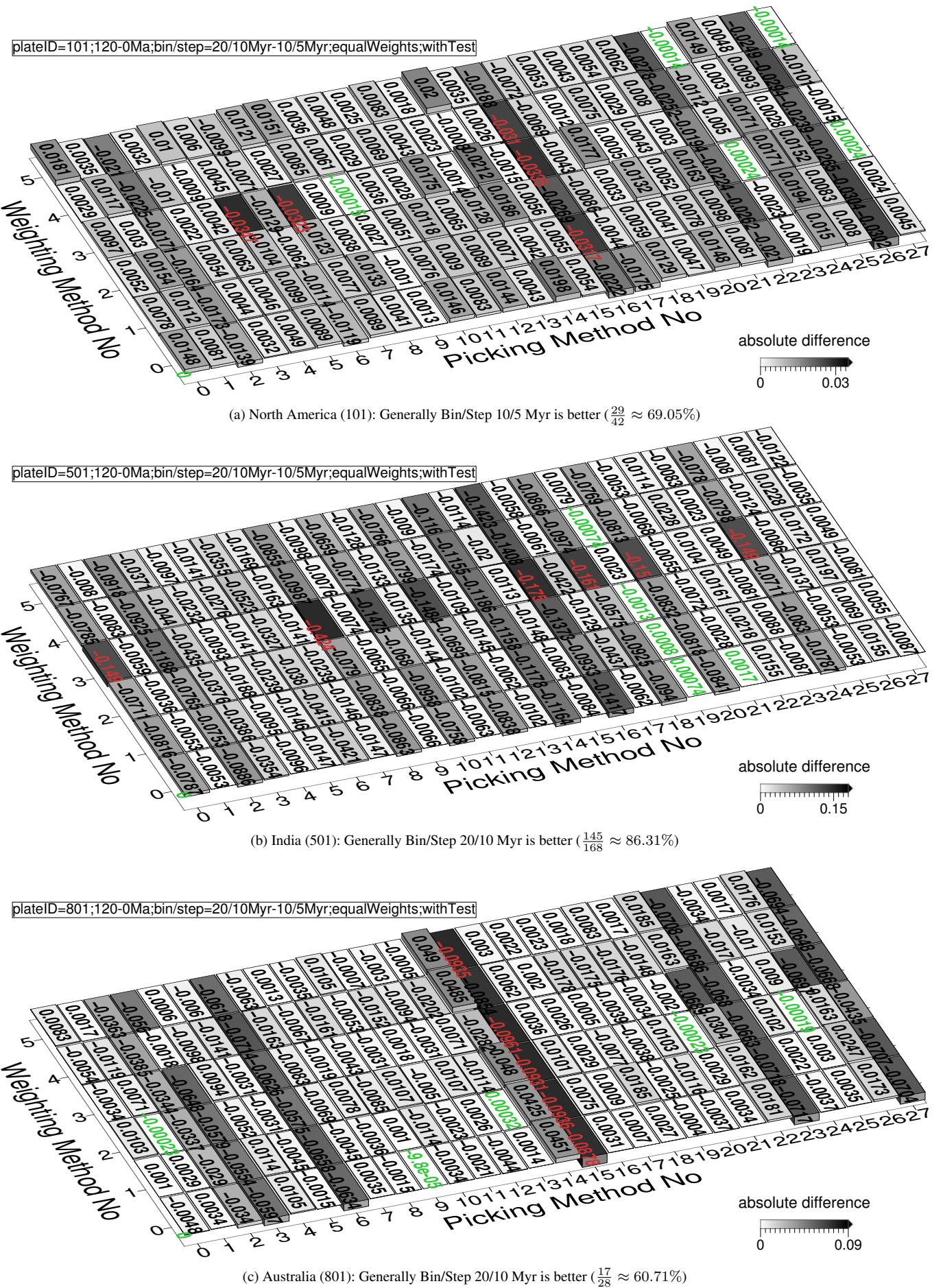


Figure 32. Differences between grids in Fig. 11 (10 Myr bin, 5 Myr step) and Fig. 17 (20 Myr bin, 10 Myr step). The absolute difference values less than 1.96-standard-deviation interval of the whole 168 values are labeled in green, more than 1.96-standard-deviation interval labeled in red.

Table 8. Highest and lowest values for the same variable are highlighted in red and green respectively.

FHM predicted for 101	Pk 5 + Wt 0			Pk 5 + Wt 1			Pk 5 + Wt 2			Pk 5 + Wt 3			Pk 5 + Wt 4			Pk 0 + Wt 5		
	ds	ds	ds	ds	ds	ds	ds	da	ds	ds	ds	ds	ds	ds	ds	ds	ds	
Age (Ma)	A95 (°)	Pmag A95 (°)	Dist (°)	Pmag A95 (°)	Dist (°)	Pmag A95 (°)	Dist (°)	Pmag A95 (°)	Dist (°)	Age (Ma)	Diff (°)	Pmag A95 (°)	Dist (°)	Pmag A95 (°)	Dist (°)	Pmag A95 (°)	Dist (°)	
0	0	7.458	2.058	7.4575	2.387	8.027	3.539	7.598	3.885	10-15-20	126.5907	7.7351	3.624	7.67	3.3909			
5	1.56039/0.87367	7.3814	2.624	7.3814	2.995	7.887	3.876	6.8	4.3453			7.515	3.9475	7.29	3.74			
10	2.89214/1.58743					15.208	3.402	8.602	5.79	dl		16.783	4.2726	14.598	2.87			
15	2.575/1.63303	12.421	9.2	12.4213	9.077	12.384	9.4	2.0857	11.805	Age (Ma)	Diff (°)	12.3843	9.4	10.153	10.71			
20	3.16077/2.20094									10-15	13.52							
25	4.96061/2.2183			6.463	6.2097			6.336	6.873	15-20	14.68	6.435	6.68					
30	3.39692/2.37114							6.678	6.458					6.97	5.724			
50	7.15565/3.22656												3.34	4.51				
55	7.17564/4.28065												5.44	6.2034				
65	7.37969/4.60029						7.6917	7.214										
100	5.79659/5.36693												8.275	7.013				
115	9.27023/5.16012						5.1	12.92	8.5	11.704			5.92	12.355	9.452	11.276		
120	14.68828/12086	11.4266	13.435	11.4266	13.0664	4.7143	16.543	7.728	15.258				4.509	17.112	7.943	15.511		

paleomagnetic APWPs for North America to see which one would make the paleomagnetic APWP closest to the 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. 9), as expected, AMP needs wider sliding window and step to get close to the reference while APP does not. 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 related 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.

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

3.2.5.2 The results are summarised in Table. 10.

3.2.5.3 Conclusions:

The sizes of sliding window and step can be optimised.

APP is recommended. We should be cautious when Wt 3 is used with AMP.

The different values for 20/10 Myr window/step are indeed generally lower than those for 10/5 window/step, excluding the unexpected case of North America with both space and shape tested (Fig. 11a and Fig. 17a).

3.2.6 Question: Is there any difference between choosing FHM and MHM as the reference path?

The differences (Fig. 33) are all lower than 0.041, and most are less than 0.01. That indicates the uncertainties of paleomagnetic mean poles are relatively large enough to make uses of Fixed and Moving Hotspot Models as references not quite differentiated. So far choosing Fixed or Moving model is not a priority.

However, through looking at the signs of these differences (Fig. 33), for North America (Fig. 33a), FHM and the North America–Nubia–Mantle plate circuit derived path is a better reference option in general, while for both India (Fig. 33b) and Australia

(Fig. 33c), MHM and the related plate circuit derived path is a generally better choice.

4 FINAL CONCLUSIONS

4.1 Universal Rules of Ways of Processing Paleomagnetic Data:

Although effects of filters (all the picking methods except Pk 1) have marginal change in reducing N (precision going down), some filters do improve the similarity score generally, for example, Pt 11 (APP without redbed-derived data) and 13 (APP with redbed-derived data corrected) are better than Pt 0 for all the three continents.

APP 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. 3). APP with filters (Pt 3, 5, 7, 9, 11, ..., 27) is generally worse than APP without any filter (Pt 1).

Weighting is actually not affecting, and in some cases makes it much worse, for example, likely worse for the combined methods of Wt 3 and AMP.

4.2 Conditional Rules of Ways of Processing Paleomagnetic Data:

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

(Add a flowchart of processing paleomagnetic data)

ACKNOWLEDGMENTS

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

REFERENCES

- Besse, J. & Courtillot, V., 2002, Apparent and true polar wander and the geometry of the geomagnetic field over the last 200 Myr, *J. Geophys. Res.*, **107**, 2300.
- Fisher, R. A., 1953, Dispersion on a sphere, *Proc. Roy. Soc. London Ser. A.*, **217**, 295–305.

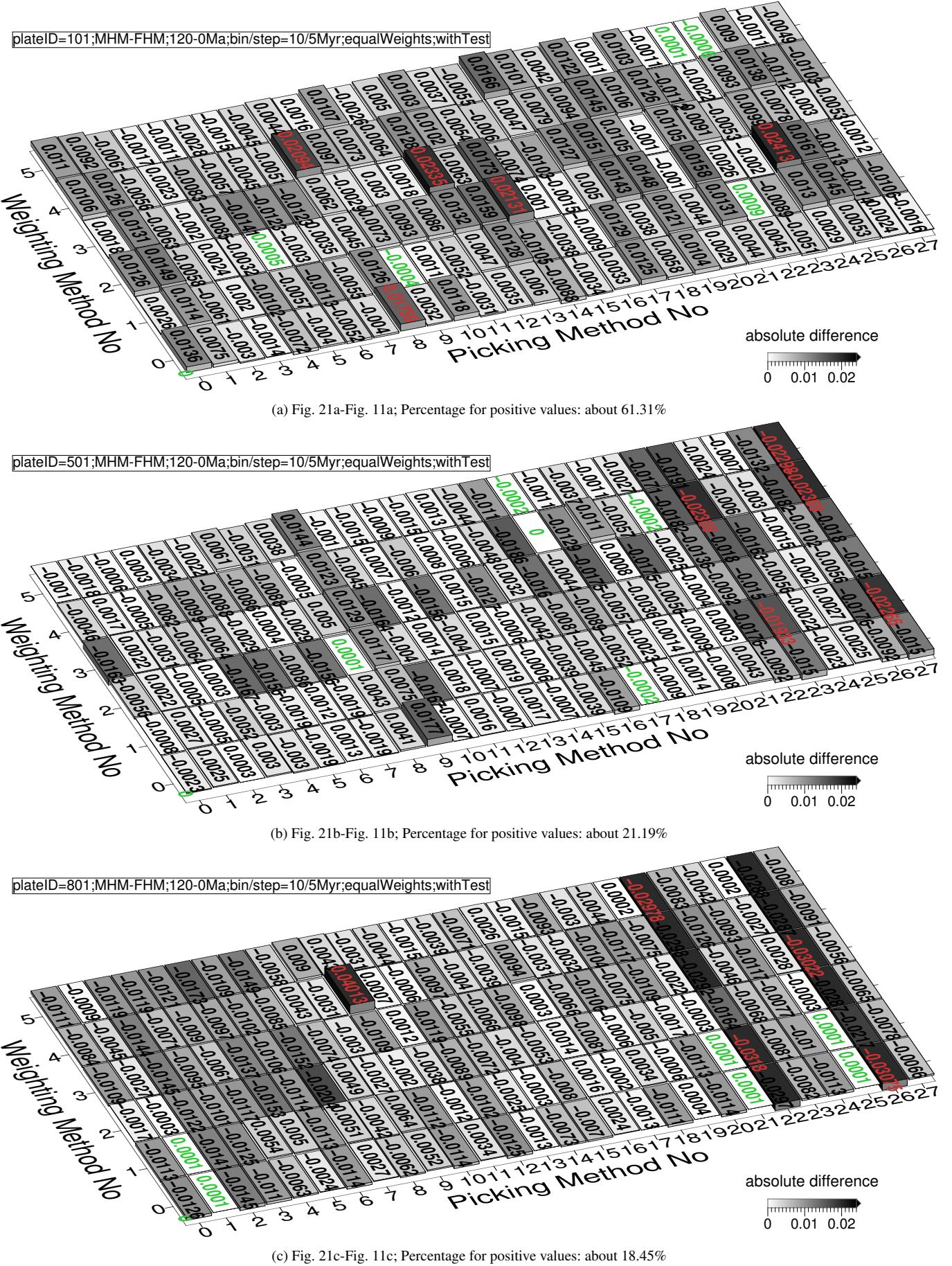


Figure 33. Differences between results from FHM (Fig. 11) and MHM (Fig. 21) as reference paths.

Table 9. Equal-weight 120–0 Ma CPDs for North American paleomagnetic APWPs compared with the FHM predicted APWP. The best are in dark green and underlined, second best in green and third in light green.

Window, Step (size in Myr)	Pk 0						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	0.26	0.2877	0.2801	0.2632	0.2619	0.00747	0.00776	0.01688	0.01283	0.0372	0.0097
4, 2	0.08205	0.08422	0.1034	0.09653	0.10295	0.09863	0.0064145	0.00711	0.0182	0.011423	0.03606	0.00909
6, 3	0.06657	0.06817	0.06788	0.06229	0.08557	0.07634	0.00627	0.007797	0.01754	0.01254	0.02113	0.008955
8, 4	0.0614	0.0772	0.0653	0.06214	0.0903	0.0646	0.006	0.01099	0.01576	0.01399	0.02027	0.01271
10, 5	0.0349	0.0458	0.0486	0.046	0.0488	0.0344	0.0059	0.0062	0.0136	0.0151	0.0153	0.0121
12, 6	0.0318	0.0316	0.0325	0.0298	0.0323	0.0299	0.0087	0.009	0.017	0.0126	0.0191	0.0145
14, 7 (119–0 Ma path)	0.0367	0.0348	0.0353	0.0369	0.0319	0.0352	0.00996	0.01198	0.023922	0.013	0.0203	0.0113
16, 8	0.0493	0.0492	0.0496	0.0486	0.0477	0.0485	0.0114	0.0117	0.023902	0.0123	0.0207	0.0129
20, 10	0.0497	0.0536	0.0538	0.0557	0.0517	0.0526	0.014	0.0174	0.029	0.0181	0.027	0.0156
24, 12	0.0274	0.0304	0.0327	0.0324	0.0315	0.0313	0.0138	0.0143	0.0221	0.0191	0.0203	0.0192
30, 15	0.0345	0.0298	0.0317	0.0307	0.03402	0.0307	0.0174	0.01797	0.0276	0.02402	0.0252	0.02414

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Wessel, P., Smith, W. H. F., Scharroo, R., Luis, J. & Wobbe, F., 2013.

Generic Mapping Tools: Improved version released, *Eos. Trans. AGU*, **94**, 409–410.

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

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

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