

# **Global Paleomagnetic Data Analysis:**

## **Improved Methods of Reconstructing Plate Motions Using Paleomagnetic Data**



Chenjian Fu  
Department of Geology  
Kent State University

A thesis submitted for the degree of  
*Doctor of Philosophy*  
Someday 2019

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background and Motivation . . . . .	2
1.1.1	Techniques Used in Relative and Absolute Plate Motion Studies . . . . .	2
1.1.2	Application of Paleomagnetism to Plate Tectonics . . . . .	4
1.1.3	Fact 1: Not All Regions on the Earth Surface Are Solid . . . . .	8
1.1.4	Fact 2: Not All Data Are Created Equal . . . . .	10
1.1.4.1	Age Error . . . . .	10
1.1.4.2	Position Error . . . . .	10
1.1.4.3	Data Consistency . . . . .	11
1.1.4.4	Data Density . . . . .	12
1.1.4.5	Publication Year . . . . .	13
1.2	Objectives . . . . .	13
1.2.1	Motivation and General Approach . . . . .	14
1.2.2	Research Questions or Hypotheses . . . . .	15
1.2.2.1	Question 1 . . . . .	15
1.2.2.2	Question 2 . . . . .	15
1.2.2.3	Question 3 . . . . .	15
1.2.2.4	Question 4 . . . . .	15
1.2.2.5	Question 5 . . . . .	16
1.2.2.6	Question 6 . . . . .	16
<b>2</b>	<b>Methodologies</b>	<b>17</b>
2.1	Introduction . . . . .	18
2.2	Methods . . . . .	21
2.2.1	Comparing Apparent Polar Wander Paths . . . . .	21
2.2.2	APWP Pairs Used in This Study . . . . .	21
2.2.3	Significant Spatial Difference . . . . .	24
2.2.4	Shape Difference . . . . .	24

2.2.4.1	Mean Length Difference . . . . .	24
2.2.4.2	Mean Angular Difference . . . . .	26
2.2.4.3	Significance Testing of Shape Difference . . . . .	26
2.2.5	Composite Path Difference . . . . .	28
2.2.6	Fit Quality . . . . .	28
2.3	Results and Discussion . . . . .	30
2.3.1	Discrimination of Difference Metrics . . . . .	30
2.3.1.1	$d_s$ . . . . .	30
2.3.1.2	$d_l$ . . . . .	32
2.3.1.3	$d_a$ . . . . .	32
2.3.1.4	$\mathcal{CPD}$ . . . . .	33
2.3.2	A Discussion on Weights . . . . .	34
2.3.3	Application to Real Paleomagnetic Data . . . . .	37
2.4	Conclusions . . . . .	39
<b>3</b>	<b>Finding the Way(s) to Make a Reliable APWP</b>	<b>42</b>
3.1	Introduction . . . . .	43
3.1.1	Not All Data Are Created Equal . . . . .	43
3.1.1.1	Age Error . . . . .	43
3.1.1.2	Position Error . . . . .	43
3.1.1.3	Data Consistency . . . . .	44
3.1.1.4	Data Density . . . . .	44
3.1.1.5	Publication Year . . . . .	45
3.1.2	Existing Solutions and General Issues . . . . .	45
3.2	Methods . . . . .	46
3.2.1	General Approach . . . . .	46
3.2.2	Paleomagnetic Data . . . . .	46
3.2.2.1	Used Specific Field Codes of GPMDB . . . . .	46
3.2.2.2	Paleomagnetic Data of Three Representative Continents	47
3.2.3	APWP Generation . . . . .	48
3.2.3.1	Picking/Binning . . . . .	48
3.2.3.2	Filtering . . . . .	53
3.2.3.3	Weighting . . . . .	56
3.2.3.4	Final Scaling . . . . .	59
3.2.4	Reference Paths . . . . .	60
3.2.5	Comparison Algorithm . . . . .	64

3.3	Results . . . . .	64
3.3.1	Baseline results: 10 Myr window, 5 Myr step, fixed hotspot reference . . . . .	64
3.3.2	Effects of windowing method . . . . .	67
3.3.3	Effects of filtering and weighting . . . . .	70
3.3.3.1	Filter aggression . . . . .	70
3.3.3.2	Filter performance . . . . .	71
3.3.3.3	Weighting performance . . . . .	72
3.3.4	Effects of window size . . . . .	72
3.3.4.1	Overall change . . . . .	72
3.3.4.2	Relative performance of methods . . . . .	79
3.3.4.3	How time window/step size affects results . . . . .	81
3.3.5	Different reference path . . . . .	83
3.3.5.1	Overall change . . . . .	84
3.3.5.2	Relative performance of methods . . . . .	84
3.4	Discussions . . . . .	90
3.4.0.1	Question: Why the APP methods generally produce better similarities than AMP methods do? . . . . .	91
3.4.0.2	Question: Why the AMP methods sometimes unexceptionally produce better similarities than APP methods do? . . . . .	93
3.4.0.3	Question: Why weighting is not affecting? . . . . .	95
3.4.0.4	Question: Why best and worst methods are not consistent? . . . . .	101
3.4.0.5	Question: Are there particular parts of the path that are more variable? Do different methods affect different parts of the path differently? . . . . .	102
3.5	Final Conclusions . . . . .	102
3.5.1	Universal Rules of Ways of Processing Paleomagnetic Data: . . . . .	102
3.5.2	Conditional Rules of Ways of Processing Paleomagnetic Data: . . . . .	103
3.5.3	Summary . . . . .	103
<b>4</b>	<b>How Much Data Needed to Make a Reliable APWP</b>	<b>105</b>
4.1	Reference Path . . . . .	106
4.2	Extraction Fraction . . . . .	106
4.2.1	Number of Samples . . . . .	112

4.2.2	Extreme Value Study: Suggestions on Algorithm, especially on large uncertainties . . . . .	112
4.2.3	Relationship between $d_s$ Score and Number of Paleopoles for Paleomagnetic APWP . . . . .	116
4.3	Are the rules we obtained in the last chapter are still true for less data?	119
<b>5</b>	<b>Conclusions</b>	<b>123</b>
.0.1	Test if A Coeval Pole Pair is Distinguishable with The Bootstrap	124
.0.2	Bivariate Sampling . . . . .	125
.0.3	Synchronization . . . . .	126
.0.3.1	Equally Treated Random Weights . . . . .	126
.0.4	120–0 Ma North America . . . . .	130
.0.5	120–0 Ma India . . . . .	131
.0.6	120–0 Ma Australia . . . . .	131
<b>Bibliography</b>		<b>134</b>

# List of Figures

1.1	GAD model: Inclination (angle $I = \tan^{-1}(2\tan\lambda)$ ) of the Earth's magnetic field and how it varies with latitude, redrawn from [7, 51, 56]. Magnetic dipole M is placed at the center of the Earth and aligned with the rotation axis; $\lambda$ is the geographic latitude, and $\theta$ is the colatitude. . . . .	5
1.2	Reconstruction of Africa with its c. 155 Ma paleomagnetic pole. The red polygon is today's position of Africa, while the blue and green ones shows its reconstructed position at c. 155 Ma, if the pole was North and South pole, respectively. Dashed green polygon illustrates the ambiguity of paleolongitude from paleomagnetic data alone (sites at same latitude but different longitudes record the same Declination and Inclination in a GAD field). . . . .	7
1.3	Much paleomagnetic data has been collected from the North American Craton. For younger geologic times, do we really need so much data to reconstruct accurately just like modern-day plate motions? The image shows distribution of all published paleomagnetic poles of the NAC over time, which are compiled from GPMDB 4.6b [41] and PALEOMAGIA [63]. . . . .	9
1.4	Example of AMP Moving Averaging Effects . . . . .	11
2.1	Parts of APWPs of supercontinent fragments share the same geometry	19
2.2	Examples showing GCD is a bad index of similarity . . . . .	20
2.3	Eight examples of APWP comparisons . . . . .	23
2.4	Geometric difference definition between two APWPs . . . . .	25
2.5	Testing on Geometry . . . . .	27
2.7	Mean spatial, length, angular differences . . . . .	31
2.8	Mean spatial, length, angular differences (without interpolations) . .	32
2.9	criteria of pair comparisons . . . . .	36

2.10 Reproducing Torsvik et al. 2012 Fig.13a . . . . .	38
3.1 Distribution of 120–0 Ma North American poles . . . . .	49
3.2 Distribution of 120–0 Ma Indian poles . . . . .	50
3.3 Distribution of 120–0 Ma Australian poles . . . . .	51
3.4 Moving average (MA) methods . . . . .	52
3.5 . . . . .	60
3.6 120–0 Ma MHM vs FHM predicted APWP of North America . . . . .	62
3.7 120–0 Ma MHM vs FHS predicted APWP of India . . . . .	63
3.8 120–0 Ma MHM vs FHM predicted APWP of Australia . . . . .	63
3.9 Differences of each plate’s paleomagnetic APWPs versus its FHM predicted APWP . . . . .	65
3.10 Best and worst differences (10 Myr bin, 5 Myr step) . . . . .	66
3.11 Differences of each plate’s paleomagnetic APWPs versus its FHM predicted APWP (AMP vs APP) . . . . .	68
3.13 Differences of differences of each plate’s paleomagnetic APWPs versus its FHM predicted APWP . . . . .	73
3.14 ds of each pair of poles for North American 10/5 Myr APWPs . . . . .	74
3.15 dl of each pair of segments for North American 10/5 Myr APWPs . . . . .	75
3.16 dl of each pair of segment-orientation-changes for North American 10/5 Myr APWPs . . . . .	76
3.17 Differences of each plate’s paleomagnetic APWPs versus its FHM predicted APWP . . . . .	77
3.18 Best and worst differences (20 Myr bin, 10 Myr step) . . . . .	78
3.20 Sliding window and step sizes versus CPD . . . . .	83
3.21 Differences of each plate’s paleomagnetic APWPs versus its MHM predicted APWP . . . . .	85
3.22 Best and worst differences (10 Myr bin, 5 Myr step) . . . . .	86
3.23 Differences of each plate’s paleomagnetic APWPs versus its MHM predicted APWP . . . . .	87
3.24 Best and worst differences (20 Myr bin, 10 Myr step) . . . . .	88
3.25 Differences between results from FHM and MHM . . . . .	89
3.26 APP spatially better than AMP . . . . .	91
3.27 APP spatially better than AMP . . . . .	92
3.28 Paleomagnetic APWP’s FQ score versus significant difference score . . . . .	97
3.29 Paleomagnetic APWP’s FQ score versus raw difference score . . . . .	98

3.30	Proportional changes of weighting 1-5 to 0: Paleomagnetic APWP's FQ score versus significant difference score . . . . .	99
3.31	Proportional changes of weighting 1-5 to 0: Paleomagnetic APWP's FQ score versus raw difference score . . . . .	100
3.32	Flowchart . . . . .	104
4.1	Random paleopole samplings (30 times) for the best and worst results for the 10 Myr window and 5 Myr step paleomagnetic APWPs vs FHM & plate circuit predicted APWP. The lower and upper bound lines connect the 1st and 3th quantiles ( $Q_1$ and $Q_3$ ) of the 30 samples. The bold line connects their means. The numbers in small parentheses are actual quantity of paleopoles after filtered by the corresponding picking methods for the case with no data removal. The $Q_1$ - $Q_3$ interquartile range from best method is also shown (shadowed) in the plot of the worst method for clarity. Black rings are the lowest value for each method or the lowest for the 20% case. . . . .	107
4.2	Same as Fig. 4.1. Here color demonstrates the resolution of pole pairs with the path comparisons. . . . .	108
4.3	Comparisons of results from the best and worst methods for North America (101), also applied on the other two continents (501 and 801). The $Q_1$ - $Q_3$ interquartile range from Picking No. 25 is also shown (shadowed) in the plot of Picking No. 16 for clarity. . . . .	109
4.4	Comparisons of results from the best and worst methods for India (501), also applied on the other two continents (101 and 801). The $Q_1$ - $Q_3$ interquartile range from Picking No. 19 is also shown (shadowed) in the plot of Picking No. 8 for clarity. . . . .	110
4.5	Comparisons of results from the best and worst methods for Australia (801), also applied on the other two continents (101 and 501). The $Q_1$ - $Q_3$ interquartile range from Picking No. 11 is also shown (shadowed) in the plot of Picking No. 26 for clarity. . . . .	111
4.6	Testing differences of results from different numbers of samples. See Fig. 4.1 for more details. . . . .	112
4.7	Less data, better similarity? . . . . .	113
4.8	Less data, better similarity? . . . . .	114
4.9	Less data, better similarity? . . . . .	115
4.10	Less data, better similarity? . . . . .	115

4.11	Less data, better similarity? . . . . .	116
4.12	Less data, better similarity? . . . . .	117
4.13	Less data, better similarity? . . . . .	117
4.14	. . . . .	120
4.15	. . . . .	121
4.16	Comparisons of results from Picking No. 1 (APP) and Picking No. 0 (AMP) with all the listed weighting methods for three continents. The $Q_1$ – $Q_3$ interquartile range from Picking No. 1 is also shown (shadowed) in the plot of Picking No. 0 for clarity. . . . .	122
1	How $N_j=1$ uncertainty ellipse is simulated . . . . .	127
2	Comparisons of Pairs a-g with random weights involved . . . . .	128
3	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. . . . .	131
4	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. . . . .	132
5	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. . . . .	133

# **Chapter 1**

## **Introduction**

*The first chapter introduces paleomagnetism-based paleogeographic reconstruction technique and highlights the motivation of the research conducted in the dissertation*

## 1.1 Background and Motivation

Reconstructing past paleogeographies, especially the motion of plates and their interactions through time, is a key component of understanding the Earth’s geological history, including deciphering tectonics (e.g. supercontinent reconstruction), paleoclimate history, and the evolution of life. Since the advent of plate tectonics, it has been the background for nearly all geologic events. In addition, plate reconstructions form the basis of global or regional geodynamic models.

### 1.1.1 Techniques Used in Relative and Absolute Plate Motion Studies

The earliest quantitative effort to model plate kinematics was fitting conjugate passive margins of the Atlantic [6, 64]. They showed that the Atlantic could be closed using a single Euler pole (using Euler’s theorem on rotation). Then it became fitting conjugate isochrons based on best-fitting marine magnetic anomaly and fracture zone data [32], which minimizes the misfit area between two isochrons. The *Hellinger* method [21] is a more advanced and generalised method which also fits conjugate isochrons based on best-fitting marine magnetic anomaly and fracture zone data, which however minimizes the sum of the misfits of conjugate data points that belong to a common isochron segment [64] instead. These techniques mainly through fitting conjugate lines mentioned above are relatively accurate for quantitative analysis. However they give relative, not absolute, motions between plates, because plate motions can’t be tied into absolute location on Earth’s surface, since both plates are likely moving. In addition, they are limited to survey data from the seafloor, with a maximum age of no more than c. 200 Ma [36].

Reference frames are a means of describing the motion of geologic features (e.g. tectonic plates) on the surface of the Earth, relative to a common point or “frame” [47]. An absolute reference frame is a frame that can be treated as fixed relative to the Earth’s geographic reference frame. In reality, it’s impossible to find a truly absolute reference frame, so we are actually looking for a frame that has limited (and hopefully known) motion, which approximates as “fixed” over geologically useful timescales and provides the most complete descriptions of plate motions. A commonly used absolute reference frame is the “Fixed Hotspot model” [35, 34], covering ages from about 132 Ma to present-day, which assumes that the linear volcanic chains found on most oceanic plates are artifacts of absolute plate motions over a upwelling plume from the deep mantle, which is assumed to be relatively fixed. The advantage of this “Fixed

“hotspot model” is that it is fairly straightforward if the assumption of fixed hotspots is correct. However, this model is limited to plates with well-dated volcanic hotspot chains (e.g. the Ninetyeast Ridge on the Indian Ocean floor and the Walvis Ridge in the southern Atlantic Ocean; [38]) and dating can be difficult (e.g. diffuse volcanic centers possibly related to large diameter plume conduits could cause the existence of time reversals; [38]). As for not well-dated hotspot tracks, for example, only about 5% of the seamounts (thought to be volcanic) in the Pacific are thought to be related to hotspot volcanism and radiometrically dated (39 per cent of these ages are less than 10 Ma; [22]). In addition, the fixed hotspot model is mostly confined to existing oceanic or thin continental crust because older oceanic lithosphere has been largely destroyed by subduction and old, thick continental crust mostly removed by erosion [11]. Last, but not least, hotspots can be susceptible to drift that may be caused by changes in sub-lithospheric mantle flow [49]. Generally, however, the drift rate is considered to be an order of magnitude less than the rate of plate motions, so only becomes significant over timescales of c. 50 Myr or more [38, 50]. To overcome this source of error, the “Moving hotspot model” [38] uses mantle convection modeling to predict hotspot drift. Some are apparent success, e.g. by getting motions in the Indo-Atlantic and Pacific hotspot clusters to agree with each other, but it’s very dependent on the mantle model. Hybrid models attempt to overcome the shortcomings of each reference frame by combining them, e.g. combining a fixed hotspot frame from 100 Ma to 0 Ma [35] with a moving hotspot frame from c. 132–100 Ma [38] (Hybrid hotspot model [47]), combining a moving hotspot frame from 100–0 Ma [38] with a paleomagnetic model (reflect plate motion relative to the magnetic dipole axis but cannot provide paleolongitudes because of the axial symmetry of the Earth’s magnetic dipole field) [56] from 140–100 Ma (Hybrid paleomagnetic model [47]), and combining a moving hotspot frame from 120–100 Ma [38] with a True Polar Wander (TPW) corrected paleomagnetic model [1] from 100–0 Ma (Hybrid TPW-corrected model [47]).

Recently another absolute reference frame “Subduction reference model” [60] tries to connect orogenies/sutures/subduction complexes’ on the Earth surface with their corresponding subducted slabs in the mantle. Assuming that these remnants sank vertically through the mantle, the absolute location at which they were subducted can be reconstructed. In this way, this model mainly imposes a longitude correction on the above mentioned “Hybrid TPW-corrected model”, and can theoretically give past absolute locations of plates back to about 260 Ma based on the estimated age of the oldest slab remnants that can be reliably located in the mantle. While the

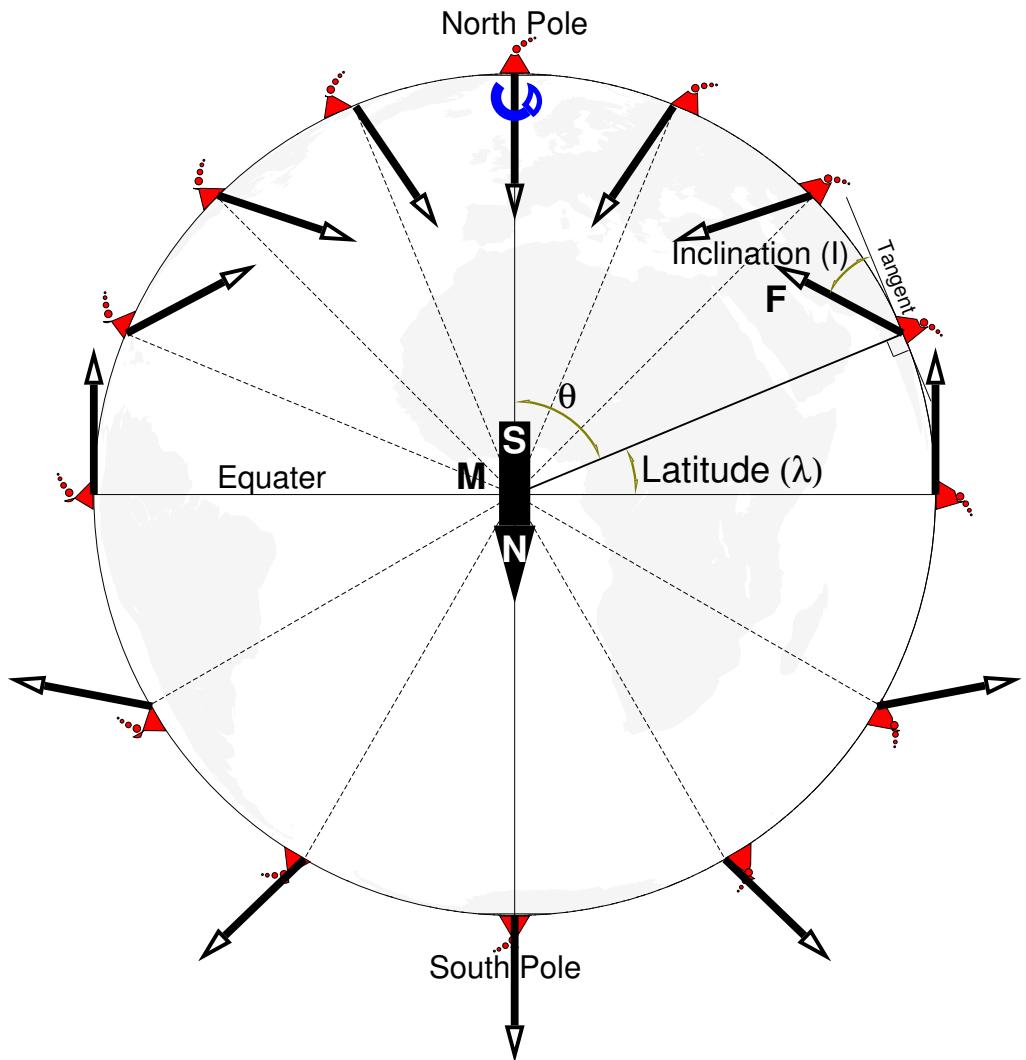
“Subduction reference model” allows for reconstructions between about 260 Ma and 140 Ma, older than the other absolute models can predict, the model is strongly dependent on the vertical subduction assumption and resolution of seismic tomography models, so its uncertainty is high. Above all, importantly, if we can describe the absolute motion of one or a few key plates, the techniques for establishing the relative plate motions described in the second paragraph above can be used to construct plate circuits that allow a full kinematic description of plate tectonics to be developed.

As we can see, all of these above reconstruction methods are limited to recent geological history. 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.

### 1.1.2 Application of Paleomagnetism to Plate Tectonics

The geomagnetic field is generated by the convective flow of a liquid iron-nickel alloy in the outer core of the Earth. It is largely dipolar and can be represented by a dipole that points from the north magnetic pole to the south pole. However, the geomagnetic field varies in strength and direction over decadal–millennial timescales due to quadropole and octopole components of the field. The most spectacular variations in direction are occasional polarity reversals (normal polarity: the same as the present direction of the field; or the opposite, i.e. reverse polarity). Over a period of a few thousand years, the magnetic axis slowly rotates/precesses around the geographic axis and the Earth’s rotation axis (secular variation), but when averaged over 10,000 year timescales, higher order components of the field are thought to largely cancel out and the position of the magnetic poles aligns with the geographic poles. This is the geocentric axial dipole (GAD) hypothesis. In a GAD field, at the north magnetic pole the inclination (angle with respect to the local horizontal plane, see Fig. 1.1) of the field is +90° (straight down), at the Equator the field inclination is 0° (horizontal) pointing north and at the south magnetic pole the inclination is -90° (straight up) (Fig. 1.1). Another direction parameter of the Earth’s magnetic field is declination. It is the angle with respect to the geographic meridian, which is 0° everywhere in a time-averaged GAD field.

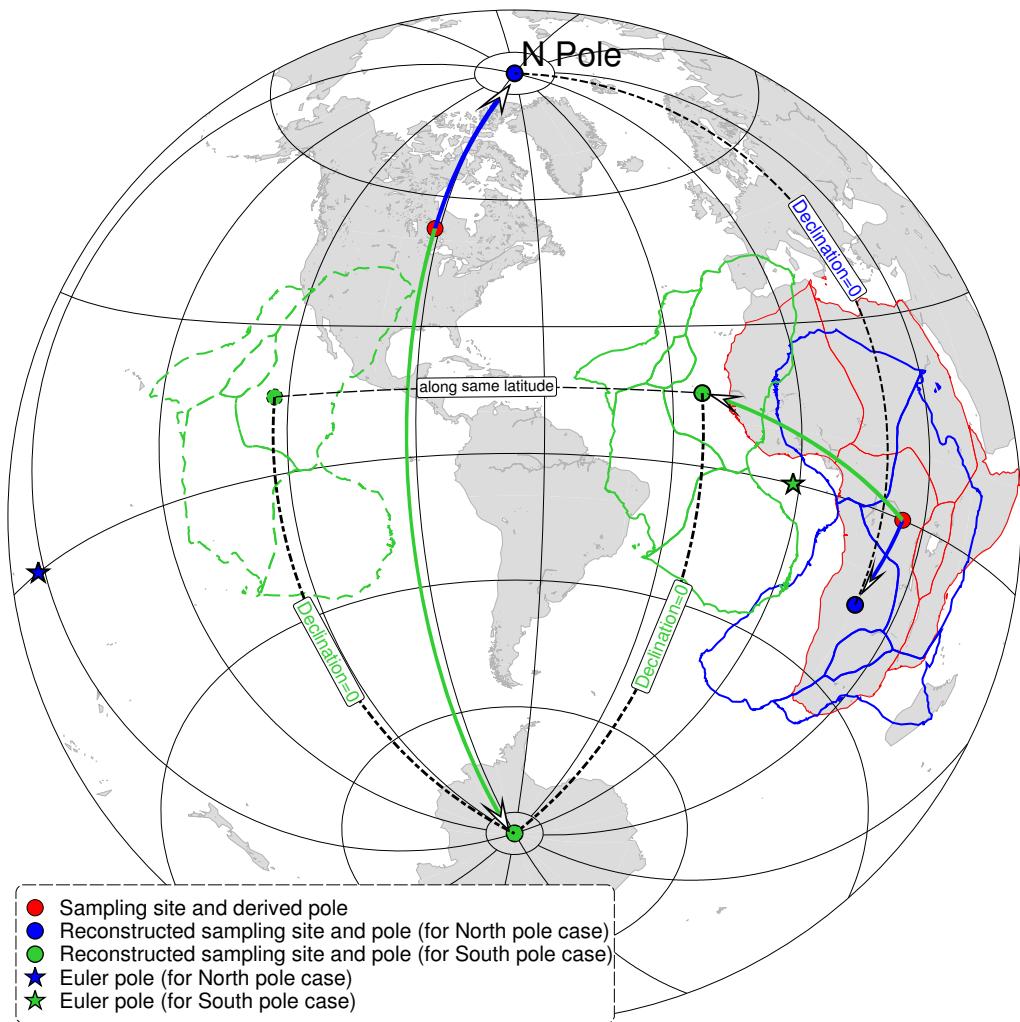
Magnetic remanence is the magnetization left behind in a ferromagnetic substance in the absence of an external magnetic field [51]. The remanent magnetisation of rocks can preserve the direction and intensity of the geomagnetic field when the rock was formed, e.g. in the process of cooling, ferromagnetic materials in the lava flow are



**Figure 1.1:** GAD model: Inclination ( $I = \tan^{-1}(2\tan\lambda)$ ) of the Earth's magnetic field and how it varies with latitude, redrawn from [7, 51, 56]. Magnetic dipole M is placed at the center of the Earth and aligned with the rotation axis;  $\lambda$  is the geographic latitude, and  $\theta$  is the colatitude.

magnetized in the direction of the Earth's magnetic field, so the local direction of the field vector is locked in solidified lava. We are often interested in whether the geomagnetic pole has changed, or whether a particular plate/terrane has rotated with respect to the geomagnetic pole [51]. By measuring the direction of the remanent magnetisation, we can calculate a virtual geomagnetic pole (VGP) to represent the geomagnetic pole of an imaginary geocentric dipole which would give rise to the observed remanent declination and inclination. Collection of VGPs (or site-mean directions) allow calculating a "paleomagnetic pole", also known as paleopole, at the formation level. Commonly a paleopole is a Fisherian mean [15] with a spatial uncertainty. A paleopole that plots away from the present geographic poles is assumed to be due to plate motions since the lava was solidified, which causes the paleopole to move with the plate [56]. Based on measurements of the remanent inclination, the ancient latitude for a plate can be calculated when the rock formed from the dipole formula  $\tan(I) = 2 * \tan(\text{latitude})$ . In addition, the remanent declination provides information about the rotation of a plate. Ideally, as a time average, a paleopole (which can be calculated from declination, inclination and the current geographic location of the sampling site) for a newly formed rock will correspond with the geographic north or south pole. To perform a reconstruction with paleopoles we therefore have to calculate the rotation (Euler) pole and angle which will bring the paleopole back to the geographic north or south pole, and then rotate the plate by the same amount of angle using the same Euler pole. This is how paleomagnetism can be used to reconstruct past positions of a plate. In our example (Fig. 1.2), a c.155 Ma paleopole (latitude=52.59°N, longitude=91.45°W) will be restored to the geographic pole by an Euler rotation of pole (0°, 178.55°E) with angle 37.41°, which rotates the sampling site from its present position of (0°, 25°E) to the Africa paleo-continent at (15.6998°S, 20.1121°E). So Africa must have drifted northwards since the Late Jurassic.

However, there are 2 problems with using paleomagnetic poles for constraining finite rotations [51]. First, if only one paleomagnetic pole is given alone without any geologic context, its polarity can be ambiguous, i.e. an upward inclination may be due to being located in the southern hemisphere during a normal polarity chron, or in the northern hemisphere during a reversed polarity chron (cf. the solid blue and solid green Africa in Fig. 1.2). In other words, we can't know if it's North pole or South pole, especially for paleomagnetic data with the Precambrian and early Paleozoic ages. Returning to the example above, if the c. 155 Ma paleomagnetic pole (52.59°N, 91.45°W) was formed during a period of reversed polarity, then it needs to be rotated to the South pole rather than the North pole. The necessary



**Figure 1.2:** Reconstruction of Africa with its c. 155 Ma paleomagnetic pole. The red polygon is today's position of Africa, while the blue and green ones shows its reconstructed position at c. 155 Ma, if the pole was North and South pole, respectively. Dashed green polygon illustrates the ambiguity of paleolongitude from paleomagnetic data alone (sites at same latitude but different longitudes record the same Declination and Inclination in a GAD field).

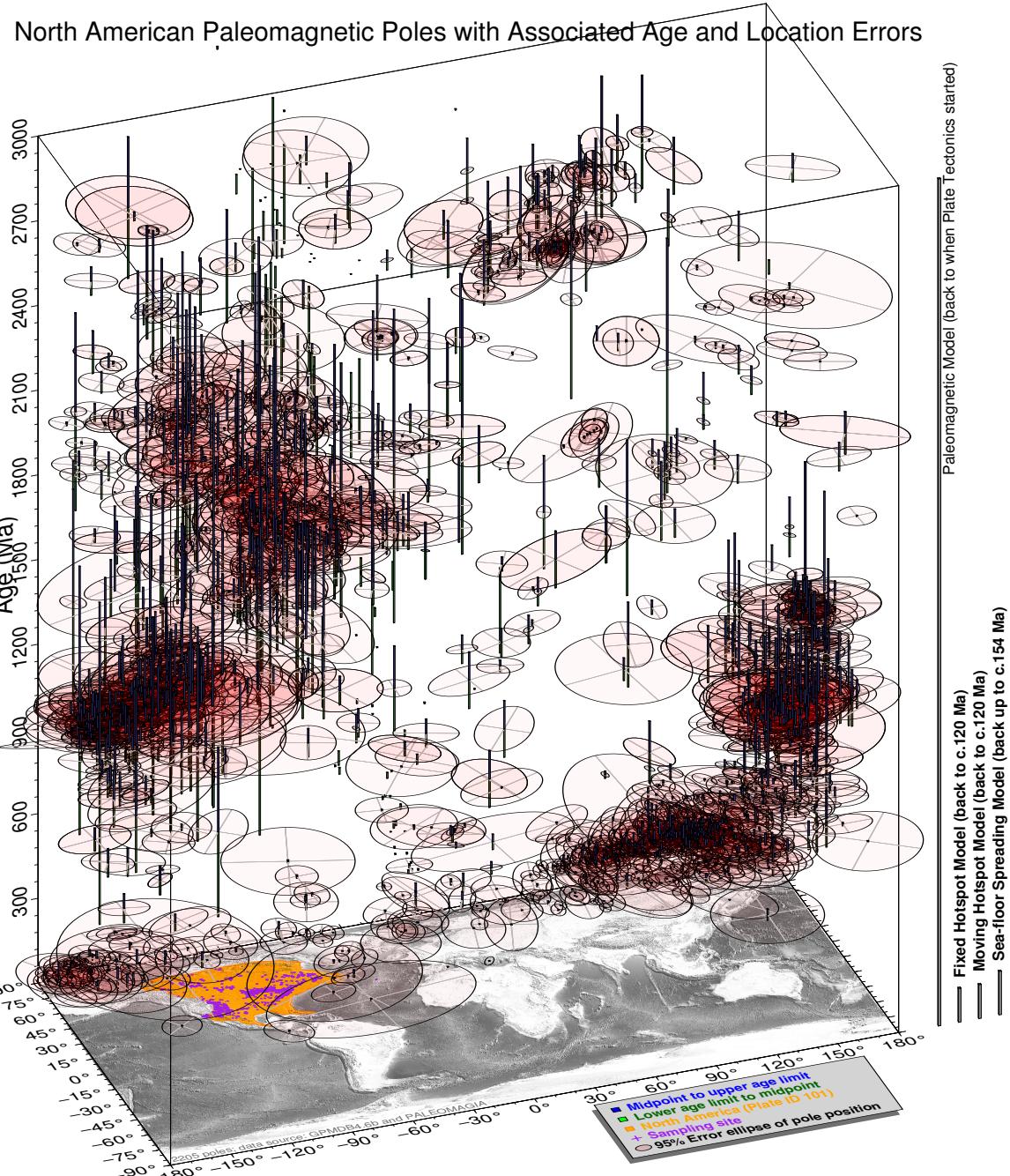
Euler rotation of pole ( $0^\circ$ ,  $1.45^\circ\text{W}$ ) and angle  $142.59^\circ$  rotates the sampling site ( $0^\circ$ ,  $25^\circ\text{E}$ ) on Africa to ( $15.6998^\circ\text{N}$ ,  $23.0121^\circ\text{W}$ ) indicating southward motion since the Late Jurassic. Second, because in a GAD field the declination equals zero everywhere (Fig. 1.2), paleomagnetic data doesn't register longitudinal motions of plates (the Euler pole for a plate moving purely to the east or west is at the geographic poles, so preserved paleomagnetic poles will experience zero rotation), which means we can position a plate at any longitude we wish subject to other geological constraints (cf. the solid and dashed green Africa in Fig. 1.2).

The data source used in this dissertation is *Global Paleomagnetic Database* (GP-MDB) Version 4.6b [41, updated in 2016 by the Ivar Giaever Geomagnetic Laboratory team, in collaboration with Pisarevsky], which includes 9514 paleopoles for ages of 3,500 Ma to the present published from 1925 to 2016. GPMDB has been published in two ways: (1) IAGA GPMDB 4.6 online query: <http://www.ngu.no/geodynamics/gpmdb/>, which is now closed; (2) Microsoft Access system in .mdb format at NOAA's National Geophysical Data Center <https://www.ngdc.noaa.gov/geomag/paleo.shtml> [42] and CESRE's Paleomagnetism and Rock Magnetism project <https://wiki.csiro.au/display/cmfr/Palaeo> which is later updated by Ivar Giaever Geomagnetic Laboratory <http://www.iggl.no/resources.html>.

An apparent polar wander path (APWP) is composed of poles of different ages from different sampling sites on the same stable (non-deforming) continent, chained together to form a record of motion relative to the fixed magnetic pole over geological time. It represents a convenient way of summarizing paleomagnetic data for a plate instead of producing paleogeographic maps at each geological period [56]. As a preliminary study, the *North American Craton* (NAC) is chosen as a research object to develop techniques we want to think about. The NAC is one of best studied cratons in paleomagnetism with the GPMD containing 2160 poles published since 1948 (Fig. 1.3). If we observe the latitudes, longitudes and age distribution of the NAC poles (Fig. 1.3), we actually can identify the general trend of its APWP. However, converting this data into a reliable, well-defined APWP can be challenging, due to the following issues:

### 1.1.3 Fact 1: Not All Regions on the Earth Surface Are Solid

If we consider the modern North America continent, the region west of the Rockies is actively deforming. Paleomagnetic data from such areas are likely to reflect local tectonic processes such as block rotation rather than rigid plate motions, and should be excluded. For example, the Rockies Mountain area was not included as my data selecting polygon (the transparent yellow area in Fig. 1.3). In order to investigate a



**Figure 1.3:** Much paleomagnetic data has been collected from the North American Craton. For younger geologic times, do we really need so much data to reconstruct accurately just like modern-day plate motions? The image shows distribution of all published paleomagnetic poles of the NAC over time, which are compiled from GPMDB 4.6b [41] and PALEOMAGIA [63].

specific craton or terrane or block's past paleogeographic motion, choosing an appropriate subregion without active tectonic activities, e.g. rotation or uplifting or rifting, to select data is often required. Such tectonics-free regions are usually called rigid. However, the difficulty of defining such tectonic boundaries makes appropriate spatial and temporal choices very difficult, particularly further in the geological past when cratonic configurations and active plate boundaries were very different to today. This leads to a question: What is the best way to constrain the data for a specific plate or block? My present solution is described in Chapter 2.

### **1.1.4 Fact 2: Not All Data Are Created Equal**

APWPs are generated by combining paleomagnetic poles for a particular rigid block over the desired age range to produce a smoothed path. However, the NAC dataset illustrates that uncertainties in the age and location of paleomagnetic poles in the GPMDB can vary greatly for different poles.

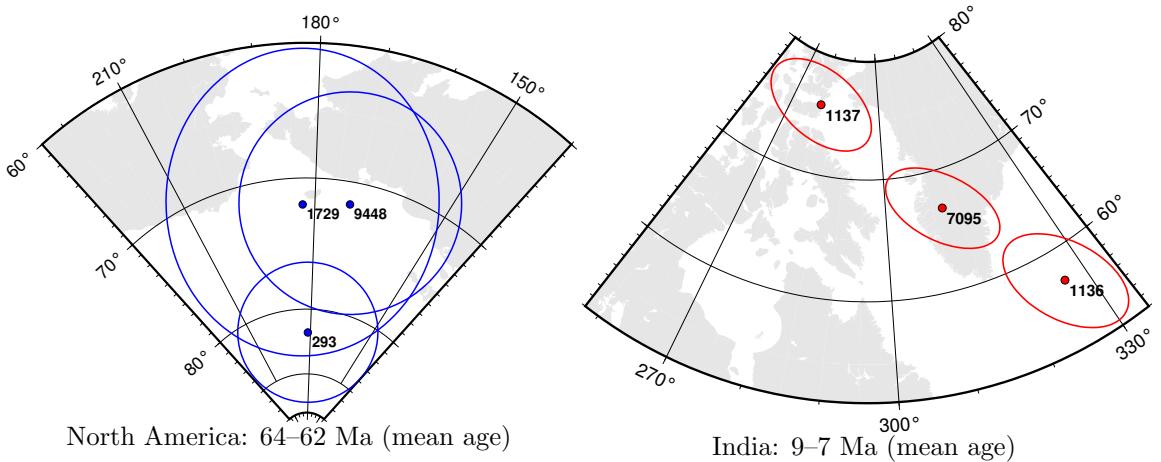
#### **1.1.4.1 Age Error**

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 NAC paleomagnetic data (Fig. 1.3) 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.

#### **1.1.4.2 Position Error**

The errors of pole latitudes and longitudes are plotted as 95% confidence ellipses (Fig. 1.3), which also vary greatly in magnitude. All paleomagnetic poles have some



**Figure 1.4:** Overlapping and further separated paleomagnetic poles of NAC and India. The oval ellipses are their 95% confidence errors. The labels are their result number given in GPMDB 4.6b.

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.

#### 1.1.4.3 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. We want to look at the consistency of NAC and India's data over smaller time periods, so the data is binned over a small time interval (e.g. 2 Myr) to see whether the paleopoles in each time interval overlap within their error ellipses, as they should. Sometimes, this is the case (Fig. 1.4a). Sometimes we have further separated poles with close ages (Fig. 1.4b).

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

**Lithology** For this poor consistency of data (Fig. 1.4b), 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 [51]. 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. For this case, all the three poles (Fig. 1.4b) are from sedimentary rocks. In addition, pole 1136 and 1137 (Result

Number in GPMDB 4.6b)'s source rocks also contain redbeds [39], although the authors did not mention about the potential inclination shallowing. For pole 7095, although the source rocks do not contain redbeds, the authors did mention about possible inclination shallowing due to haematite grains [18].

**Local Rotations** As discussed previously, local deformation between two paleomagnetic localities invalidates the rigid plate assumption and could lead to inconsistent paleopole directions. All the three poles (Fig. 1.4b) contain signals of local rotations [39, 18], e.g. pole 7095 has a signal which suggests the presence of a counter-clockwise local rotation of the Tinau Khola section [18], and therefore do not reflect motions of the whole rigid India plate in this case. So the discordance is likely due to local deformation (Fig. 1.4b), and we would ideally want to exclude or correct such poles from our APWP calculation.

**Other Factors** In Fig. 1.4, mean pole age (centre of age error) has just been binned. If any of the paleopoles 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 paleopoles 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, because in order to average errors in orientation of the samples and scatter caused by secular variation, a “sufficient” number of individually oriented samples from “enough” sites must be satisfied [51, 61, 3]. For example, pole 1136 (Fig. 1.4b) is from only 4 sampling sites, pole 1137 from only 3 sites and pole 7095's site number not even given in the GPMDB 4.6b.

#### 1.1.4.4 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 (Fig. 1.3). 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 NAC, but few poles exist for Greenland and Arabia.

Based on mean age (mean of lower and upper magnetic ages), for 100–0 Ma, GPMDB 4.6b has 198 NAC poles, but only 17 for Greenland and 24 for Arabia.

#### 1.1.4.5 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? 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 [61] and the BC02 selection criteria [3]. Briefly, the V90 criteria for paleomagnetic results includes seven criteria: (1) Well determined age; (2) At least 25 samples with Fisher [15] 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 [58]. Compared with V90, the BC02 criteria suggests stricter filtering, e.g. using only poles with at least 6 sampling sites and 36 samples, each site having  $\alpha_{95}$  less than  $10^\circ$  in the Cenozoic and  $15^\circ$  in the Mesozoic. There are many potential ways to weight the data set which could obviously greatly influence the final result, and we want to test this. But there has been limited study of how effective these filtering/weighting 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.

## 1.2 Objectives

Our overarching aims are to develop rigorous, consistent and well-documented methods of reconstructing plate motions using paleomagnetic data, and to investigate the limits of paleomagnetic data on reconstructing individual plate motions, supercontinents, and global tectonic parameters like average rate of plate motion.

### 1.2.1 Motivation and General Approach

How has plate tectonics evolved over geologic history, in terms of average plate velocities, numbers of plates and so on? The only quantitative data we have prior to about 170 Ma are paleomagnetic data. We know there are limitations, because we can't constrain the longitudes of paleo-plates very well. When we look back through geologic history, how much good paleomagnetic data do we have, and how well does it reconstruct ‘true’ plate motions? We don’t know well the effects of data quality and density, which generally degrades further back in geologic history, on producing reliable APWPs. For the past c. 130–200 Myr we have the highest density of paleomagnetic data and also independent plate motion data from reconstructions of ocean spreading combined with hotspot reference frames. These independent data sources can help constrain plate motions in more accurate ways. This allows us to ask the question: How much paleomagnetic data do we need actually to reconstruct accurately known modern-day plate motions? If we can handle that, we can go back in time. For a certain density of paleomagnetic data that we have, how reliably can we talk about what’s going on in the past given the much lower data distribution? It might turn out we don’t need very much data to say something reasonably and reliably. We can test this by looking at the last 0–120 Ma where we can compare paleomagnetically derived plate motions with other methods of paleogeographic reconstruction. This does not only include the work of developing tools and algorithms to generate those paleomagnetically derived plate motions (to use paleomagnetic data to reconstruct APWP parameters that are known from other sources like ocean basins and hotspots), but also need us to know how good these tools are or which one is the best algorithm (to compare paleomagnetic APWPs with the known data sources predicted APWP). This can give insights into how well we can ‘know’ plate motions back in the past, and what data quality and density are necessary to reliably reconstruct a ‘true’ APWP.

As a preliminary analysis, some algorithms were made to separate/calculate out so-called good paleomagnetic data (at any particular time period for a particular craton, like here from 100 Ma to the present day for NAC). We are interested in what makes ‘good’ data, how we can identify it and filter it from the database, and how sometimes ‘bad’ data is only bad in the sense that it is poorly constrained in age or position or any other parameter, in which cases it might be possible to include it by e.g., weighting. A weighted mean pole can be calculated for a time interval with ‘better’ (more likely to be reliable) poles counting more than ‘worse’. For example, a pole with small  $\alpha_{95}$  and very well constrained age is more likely to reflect APWP

position at the selected age point than a pole with large  $\alpha_{95}$  and very broad age range.

### **1.2.2 Research Questions or Hypotheses**

Questions 1–4 focus on method development, whereas 5 and 6 start using them for plate tectonic research, especially in deep times.

#### **1.2.2.1 Question 1**

What is the best way to turn a collection of individual poles, with different age constraints and uncertainties, into a smoothed APW path? This question, in fact, is about how to (1) choose a data-constraining polygon that represents a solid continent during a certain period; (2) pick (or bin) data within a certain window for Fisher statistical [15] calculation; (3) do weighting for picked data according to different uncertainties or other kinds of standards of qualifications; (4) if the derived APWP is still not smoothed enough when compared with a reference path, is further smoothing necessary? Our goal here actually is to get a reliable result, i.e. a path generated to approximate the ‘real’ APWP with appropriate uncertainties.

#### **1.2.2.2 Question 2**

Based on the consequences from the algorithms we developed, we can do research on why some algorithms are good, others bad for all plates? Why some algorithm performs well for a plate or two but not others?

#### **1.2.2.3 Question 3**

How much paleomagnetic data do we need actually to accurately reconstruct known modern known plate motions? What insights does this give us into the reliability of reconstructions from earlier in geologic history?

#### **1.2.2.4 Question 4**

Based on our analysis above, can we develop algorithms that look for matching segments of APWPs from different cratons, that might indicate they were part of the same continent or supercontinent?

### **1.2.2.5 Question 5**

What kind of dataset (in terms of data density and quality) is needed to accurately reconstruct a known APWP, or a shared APWP between two cratons? If we can establish some criteria for this, does it provide any insights into past reconstructions of plate motions (e.g., Rodinia)?

### **1.2.2.6 Question 6**

Can we develop algorithms that use APWPs from multiple continents to estimate global average plate motion rates? Can we get a good sense of how much information is lost due to lack of data on longitudinal motions? Can we use this to draw any conclusions about long term trends (or lack thereof) in the style and vigour of global plate tectonics? (Possible further question: can data on relative continental motion acquired from matching APWP curves be incorporated to improve these estimates?)

In summary, this dissertation will not be able to help answer all the above questions. However, in the end the completion of this dissertation and solving the first three questions are hoped to be helpful solving the later questions in the future.

# Chapter 2

## Methodologies

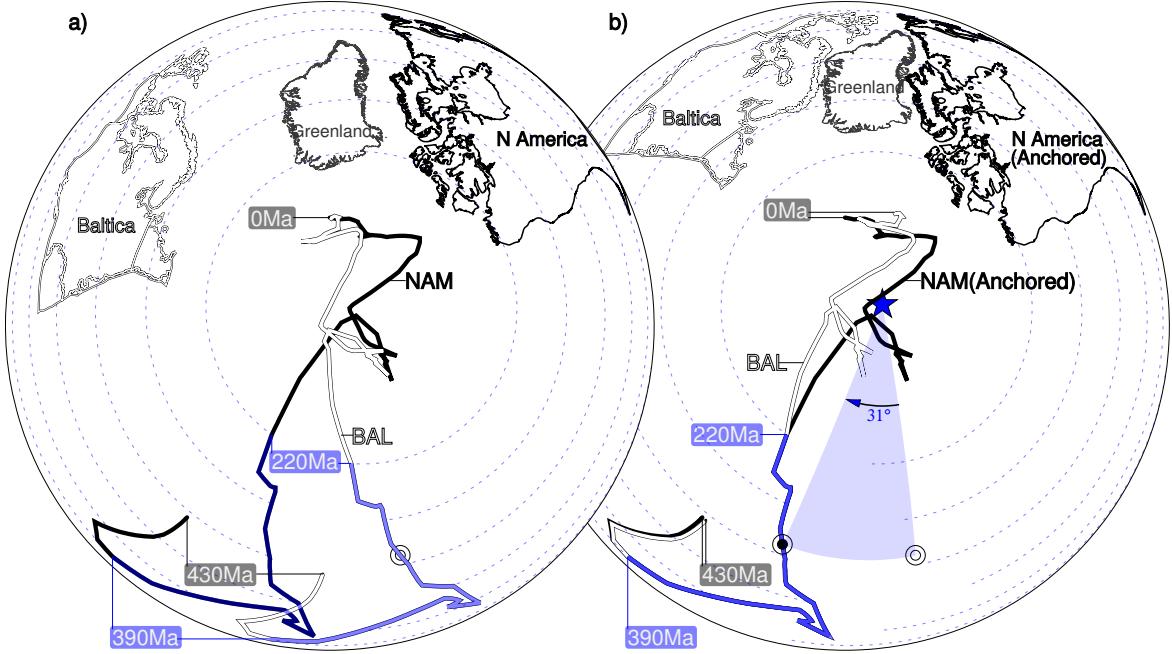
*This chapter mainly describes the development of a new APW path similarity measuring tool used throughout the dissertation. Apparent polar wander paths (APWPs) based on paleomagnetic data are the principal means of describing plate motions through most of Earth history. Comparing the spatio-temporal patterns and trends of APWPs between different tectonic plates is important for testing proposed paleogeographic reconstructions of past supercontinents. However, thus far there is no clearly defined quantitative approach to determine the degree of similarity between APWPs. This paper proposes a new method of determining the degree of similarity between two APWPs that combines three separate difference metrics that assess both spatial separation of coeval poles, and similarities in the bearing and length of coeval segments using a weighted linear summation. Bootstrap tests are used to determine whether the differences between coeval poles and segments are significant for the given spatial uncertainties in pole positions. The Fit Quality is used to discriminate between low significance scores caused by comparing poorly constrained paths with large spatial uncertainties from those caused by a close fit between well-constrained paths. The individual and combined metrics are demonstrated using tests on synthetic pairs of APWPs with varying degrees of spatial and geometric difference. In a test on real paleomagnetic data, we show that these metrics can quantify the effects of correction for inclination shallowing in sedimentary rocks on Gondwana and Laurussia's 320–0 Ma APWPs. For an APWP pair, when one APWP's three individual metrics are all greater than or equal to, or less than or equal to the other one's, weighting is dispensable because the similarity ranking order becomes straightforward; otherwise assigning equal weights is recommended, although then decision makers are allowed to arbitrarily change weights according to their preferences.*

(This chapter is also openly accessible from [https://github.com/f-i/APWP\\_similarity](https://github.com/f-i/APWP_similarity). Text: [https://github.com/f-i/APWP\\_similarity/blob/master/2.pdf](https://github.com/f-i/APWP_similarity/blob/master/2.pdf); Figures: [https://github.com/f-i/APWP\\_similarity/blob/master/2\\_figures.pdf](https://github.com/f-i/APWP_similarity/blob/master/2_figures.pdf); Supplementary: [https://github.com/f-i/APWP\\_similarity/blob/master/2Supp.pdf](https://github.com/f-i/APWP_similarity/blob/master/2Supp.pdf))

## 2.1 Introduction

Paleomagnetism is an important source of information on the past motions of the Earth’s tectonic plates. The orientation of remanent magnetisations acquired by rocks during their formation record the past position of the Earth’s magnetic poles. In older rocks, these virtual geomagnetic poles often appear to be increasingly offset from the modern day geographic poles. Because the Earth’s geomagnetic field appears to have remained largely dipolar and centered on the spin axis for at least the last 2 billion years [14], this divergence is interpreted as recording the translation and rotation of a continent by the motion of tectonic plates in the time since the rock formed. An Apparent Polar Wander Path (APWP) is a time sequence of paleomagnetic poles (or, more commonly, mean poles that average all regional paleopoles of similar age) that traces the cumulative motion of a continental fragment relative to the Earth’s spin axis.

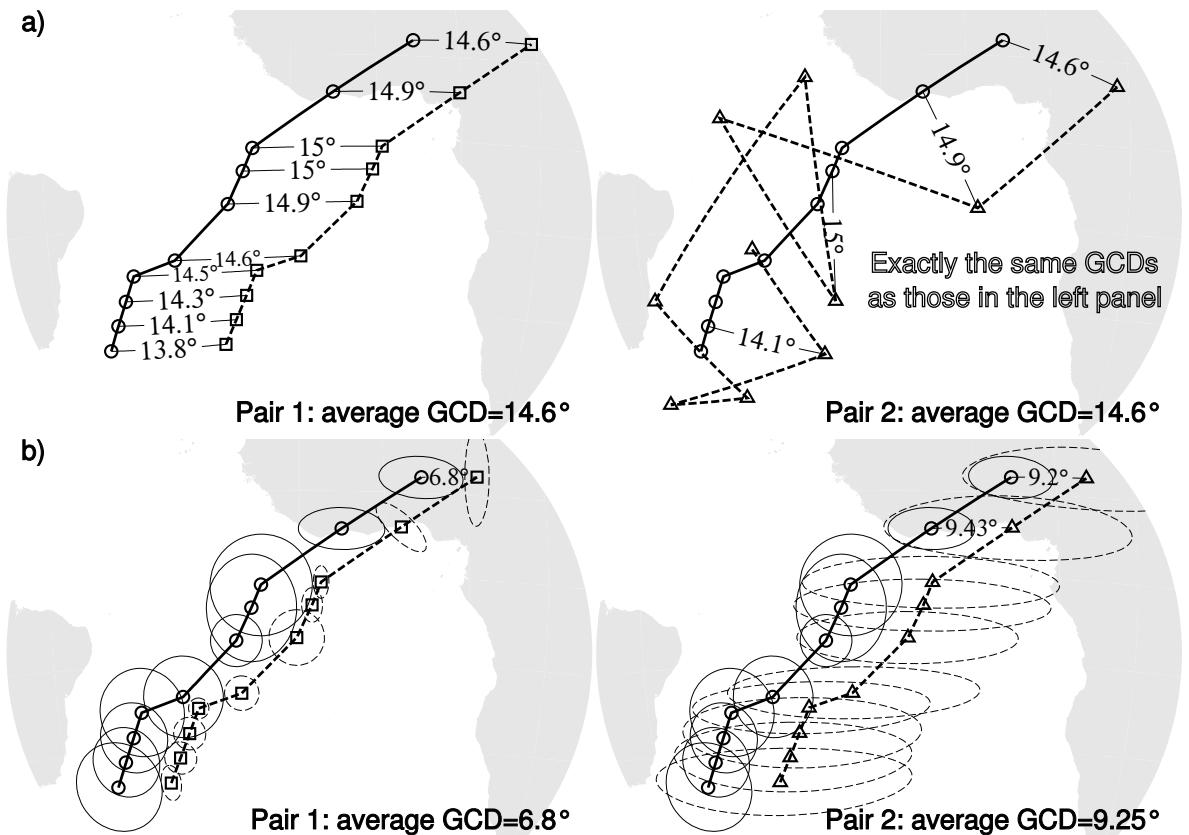
Investigations of the Earth’s past tectonic evolution and paleogeography often involve comparing APWPs. For example, if two now separated continental fragments were once part of the same supercontinent, their APWPs should share the same geometry during the interval that this supercontinent existed. If the supercontinent has been correctly reconstructed, the APWPs should also overlap during this interval (Fig. 2.1). APWP comparisons can be used to assess plate motion models generated using different datasets and/or fitting techniques [3, 2, 44, 56, 12, for example]; significant deviations from the known APWP for a continent can also be used to identify local tectonic rotations [19, 9, for example]. Despite the clear importance of measuring APWP similarity, these comparisons remain largely qualitative in nature, involving visual comparisons of specific APWP segments and checking if they have overlapping 95% confidence limits [3, 2, 19, 12, for example]. Where quantitative measures are used, the mean great circle distance (GCD) between coeval poles on the APWP pair has been commonly used as a generalised difference metric for a pair of APWPs, with a lower score indicating that they are more similar [44, 56, for example]. However, because GCD is simply a measure of spatial separation and does not incorporate geometric information about the two paths being compared, it is possible for pairs



**Figure 2.1:** (a) The APWPs for North America (black) and Baltica (grey) are spatially distinct, but their Late Paleozoic–Early Mesozoic sections are geometrically similar due to them both being part of the supercontinent Pangaea. (b) Reversing the opening of the Atlantic Ocean by rotation around a reconstruction pole (blue star) results in the overlap of these two APWPs between 390 million years ago (Ma) and 220 Ma, validating the proposed paleogeography. The effects of this rotation on Baltica and its APWP (BAL) are illustrated by the motion of the circle marker (before: blank center; after: dark center), respectively. General Perspective projection. APWPs and rotation parameters from [55].

with clearly different similarities to have similar mean GCD scores (Fig. 2.2a). Due to the inherent time variability of the geomagnetic field, uncertainties arising from the sampling and measurement of remanent magnetisations, and uncertainties in the magnetization age, the mean paleopoles that make up an APWP also have associated spatial uncertainties. The significance of a GCD score is therefore not immediately obvious. A score that indicates a relatively large difference between two paths may not be significant if the spatial uncertainties are large; a small difference could be significant if the spatial uncertainties are small (Fig. 2.2b).

We have developed an improved quantitative method of calculating the similarity between two APWPs, or coeval segments of APWPs, in the form of a composite difference score that compares both their spatial overlap and geometry. Our method incorporates statistical significance testing, allowing paths with associated spatial uncertainties to be rigorously compared to each other. The validity and effectiveness of this method, and its superior discrimination compared to a mean GCD score, are demonstrated by comparing the published APWP of the North America Plate to seven derivative paths with different degrees of spatial and geometric noise applied



**Figure 2.2:** (a) How the average GCD similarity metric ignores path geometry: *Pair1* (circles and squares, left) is clearly more similar than *Pair2* (circles and triangles, right), but for both pairs each GCD remains constant. (b) How GCD also ignores spatial uncertainties. The average GCD separation between coeval points is smaller for *Pair1* (circles and squares, left) than *Pair2* (circles and triangles, right). But if spatial uncertainties (plotted as 95% confidence ellipses) are considered, this ranking is not trustworthy: it is *Pair2* that is statistically indistinguishable from the reference path. Azimuthal Orthographic projection.

(Fig. 2.3).

We also test our algorithm on real paleomagnetic data, demonstrating that this tool can be used to quantitatively assess the effects of different corrections (in this case, bulk corrections for inclination shallowing in sediments) on the similarity between APWPs from different continents.

## 2.2 Methods

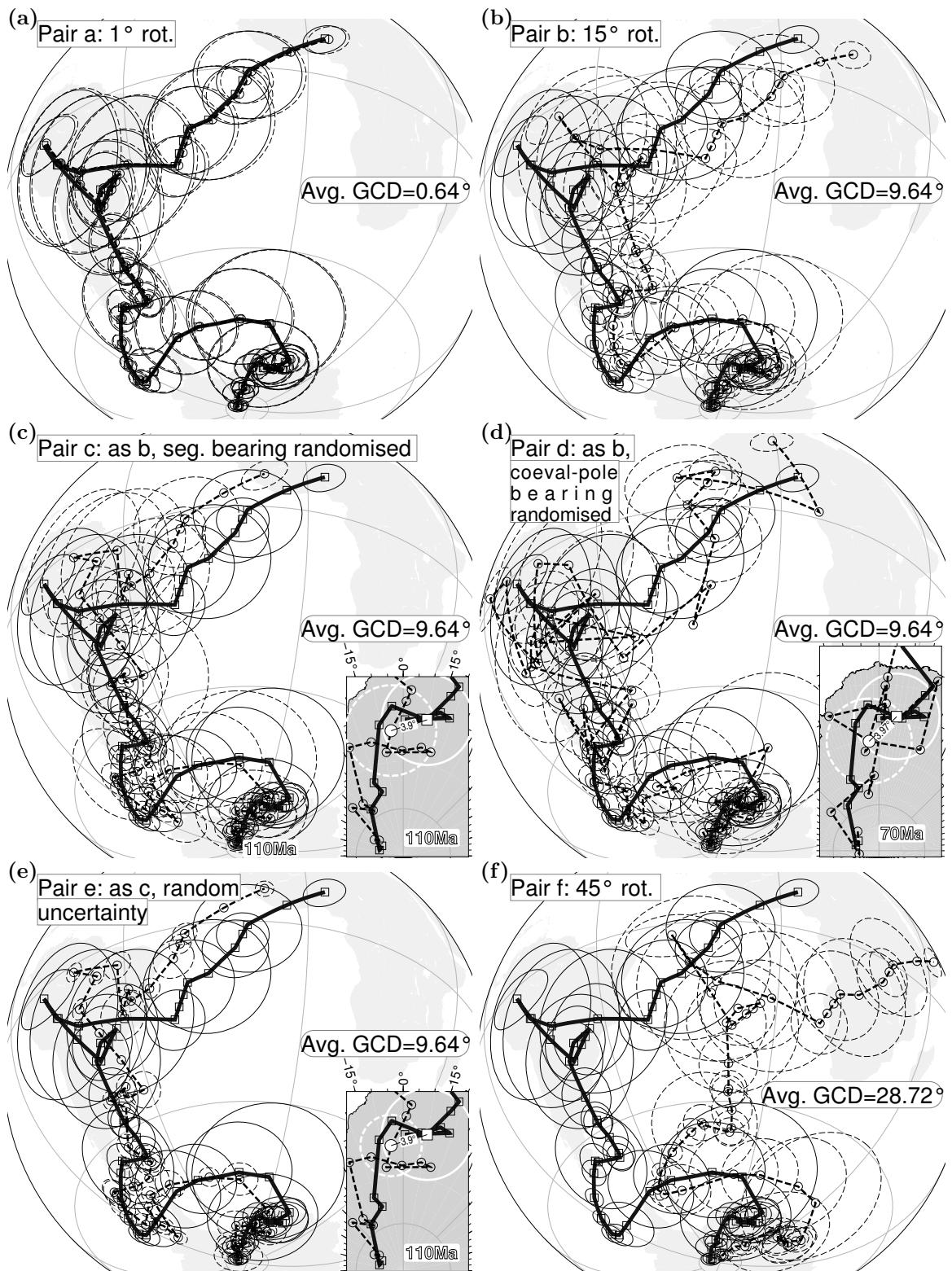
### 2.2.1 Comparing Apparent Polar Wander Paths

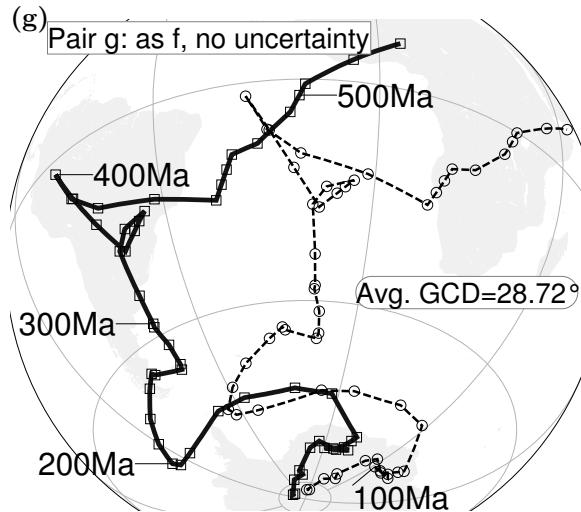
An APWP consists of a sequence of ( $\mathbf{n}$ ) mean poles,  $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_n$ , which average the published paleopoles from a particular continent for a particular time interval. Each mean pole has associated longitude ( $\phi$ ), latitude ( $\lambda$ ), and age ( $t$ ). Spatial uncertainty is represented by a 95% confidence ellipse described by semi-major axis  $\mathbf{dm}$  with azimuth  $\beta$  (angle east of north) and perpendicular semi-minor axis  $\mathbf{dp}$  (e.g. Fig. 2.3).

For two continents that were once part of a supercontinent, their APWPs for the period should perfectly overlap when rotated into a common reference frame (Fig. 2.1). However, due to (i) the spatial uncertainty associated with the mean poles that form an APWP, (ii) differences in the density and quality of data available to calculate paleopoles for different continents in coeval time periods, and (iii) uncertainties and possible errors in the rotations used to represent past plate motions, a perfect match will not be obtained. Instead, two ‘matching’ paths should share a generally similar geometry, and largely overlap with each other when rotated into a common reference frame. A quantitative measure of the spatial and geometric similarity between these two paths should ideally allow us to distinguish between non-identical paths that are similar within the associated uncertainties, and non-identical paths that are actually different, due to differential motion between the two continents or a poorly constrained reconstruction.

### 2.2.2 APWP Pairs Used in This Study

To assess the performance of the evaluation method developed here, we apply it to seven different path scenarios (Fig. 2.3) generated from transformations of the 530–0 Ma Phanerozoic APWP for Laurentia [59]. Almost exactly identical paths generated by rotating one by a degree around an Euler pole at (-55°, 88.5°) (*Pair a*, Fig. 2.3a) represent an ideal case of matching paths in the same spatial reference





**Figure 2.3:** APWP pairs used to validate new path comparison method. In each case the Phanerozoic APWP for Laurentia/North America (squares, bold line) at 10 million-year (Myr) timesteps [59, “RM” column of its Table 3], is compared to a transformed copy (circles, dashed line): (a)  $1^\circ$  finite rotation applied to all the mean poles and their 95% uncertainty ellipses around an Euler pole at  $(125^\circ\text{E}, 88.5^\circ\text{S})$ ; (b) as (a), but  $15^\circ$  rotation around same Euler pole; (c) after rotation as in (b), the orientation of each APWP segment is randomised whilst keeping their GCD length and the coeval poles’ GCD fixed; (d) after rotation as in (b), the bearing between coeval poles is randomised whilst keeping their GCD spacing fixed; (e) as (c), but with randomly varied and relatively smaller associated spatial uncertainty; (f) as (a), but  $45^\circ$  rotation around same Euler pole; (g) as (f), but with zero associated spatial uncertainty. Azimuthal Orthographic projection.

frame. Matching paths that have been rotated out of the same reference frame by small ( $15^\circ$ ; *Pair b*, Fig. 2.3b) and large ( $45^\circ$ , *Pair f*, Fig. 2.3f) amounts around the same rotation pole represent small and large reconstruction errors, respectively. Random noise added to the path (*Pair c*, Fig. 2.3c, *Pair d*, Fig. 2.3d) or the associated uncertainties (*Pair e*, Fig. 2.3e) represent differences in data source and/or quality. In the final pair (*Pair g*, Fig. 2.3g), spatial noise has been reduced by removing *Pair f*’s pole uncertainties.

These seven cases allow evaluation of the performance of any path comparison metric across a range of different spatial and geometric similarities. To be successful, such a metric must distinguish pairs with high spatial and geometric similarity (*Pair a*) from pairs with lower spatial (*Pair b*, *Pair f*, *Pair g*) or geometric (*Pair c*, *Pair d*) or both (*Pair e*) similarities.

To achieve more robust discrimination than the mean GCD, we propose combining a metric for spatial misfit (Mean Significant Spatial Difference) with metrics for geometric difference (Mean Significant Length and Angular Differences) using a weighted linear summation, as described in the following sections.

### 2.2.3 Significant Spatial Difference

As in previous quantitative comparisons [44, 56, for example], the spatial separation of two APWPs is defined by the average GCD distance between their coeval poles, but we add a filter for spatial uncertainty based on the bootstrap approach [53]. 1000 bootstrapped mean directions for each pole in a coeval pair were generated (the exact sampling method is dependent on the available information for the pole—see Supplementary Information for a full description) and their cumulative distributions in Cartesian coordinates were compared [53]. Pairs that could not be distinguished at the 95% confidence interval had their GCD separation set to 0 prior to calculation of the mean GCD distance for all pairs. This distance is then normalised by dividing by  $50^\circ$ , which is referred to the empirical fact that a 95% confidence ellipse major semi-axis of about  $25^\circ$  is considered unacceptably large by paleomagnetists [7], to obtain the significant spatial difference  $d_s$ . A  $d_s$  of zero indicates that the two paths are statistically indistinguishable from each other.

### 2.2.4 Shape Difference

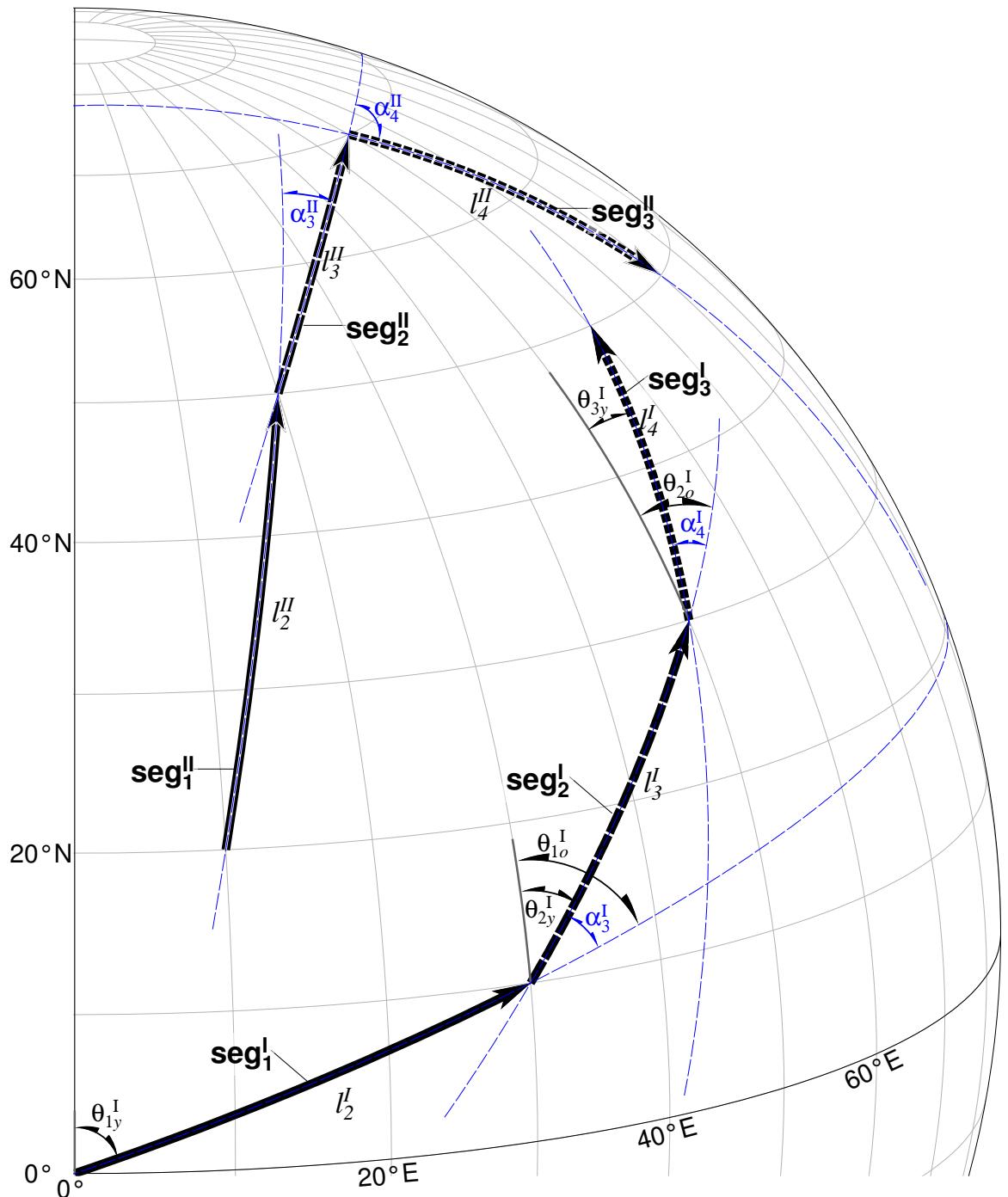
The shape of an APWP is determined by the orientations and lengths of its geodesic segments, which are related to the location of the Euler stage pole that describes plate motions, and the rotation rate about that pole, respectively. The geometric similarity of two APWPs can therefore be assessed by comparing (i) the bearings, and (ii) the lengths of their coeval segments (Fig. 2.4), with the assumption that similar geometries are generated by a common set of stage rotations.

#### 2.2.4.1 Mean Length Difference

The mean length difference between the two APWPs  $traj^I$  and  $traj^{II}$  is the absolute sum of differences between the lengths of coeval path segments (e.g.  $l_2^I$  vs  $l_2^{II}$ ,  $l_3^I$  vs  $l_3^{II}$ ,  $l_4^I$  vs  $l_4^{II}$ , Fig. 2.4), normalised by dividing by the possible maximum distance the pole could wander during the whole period, such that:

$$d_l = \frac{\sum_{k=2}^n |l_k^I - l_k^{II}|}{D_{polar} * (t_n - t_1)}, \quad \forall k \in \{2, 3, \dots, n\},$$

where  $|l_k^I - l_k^{II}|$  is the length difference of one pair of coeval segments for an APWP pair ( $traj^I$  and  $traj^{II}$ ), e.g.  $|l_2^I - l_2^{II}|$  for the beginning coeval segment pair. The normalising parameter  $D_{polar}$  is  $2.7^\circ/\text{Myr}$ , derived from estimates of magnitude of maximum plate velocity [48, 29, up to about 30 cm/year]. A  $d_l$  approaching 1 would



**Figure 2.4:** Geometric differences between coeval sections of two different APWPs ( $\text{seg}_1^I$ - $\text{seg}_2^I$ - $\text{seg}_3^I$  &  $\text{seg}_1^{II}$ - $\text{seg}_2^{II}$ - $\text{seg}_3^{II}$ ) can be described by comparing segment lengths (e.g.  $l_2^I$  vs.  $l_2^{II}$ ) or changes in bearing of coeval segments relative to their previous segment (e.g.  $\alpha_3^I$  vs.  $\alpha_3^{II}$ ). Segments are along great circles (blue dashed lines). Azimuthal Orthographic projection.

result from a comparison between a virtually stationary APWP and one associated with a rapidly moving plate.

#### 2.2.4.2 Mean Angular Difference

The mean angular difference describes the degree of consistency between the polar-wandering directions (defined as the bearing of the older pole in a segment with respect to the younger one) of two APWPs. In order to robustly compare two APWPs that have not necessarily been rotated into the same reference frame, it is more useful to define the APWP geometry relative to the path itself, rather than an external reference frame. Therefore the bearing of a segment is expressed as the change in geographic bearing with respect to the previous segment ( $\alpha_3$  and  $\alpha_4$ , Fig. 2.4). For example,  $\alpha_3^I$  is the result of subtracting the geographic azimuth  $\theta_{2y}^I$  from  $\theta_{1o}^I$ , where “y” stands for young end of segment and “o” for old end of segment. The first segment cannot record a relative bearing change: a path with  $n$  poles therefore consists of  $n-1$  segments which are described by  $n-2$  relative angles. The defined range of bearing values is set as  $-180^\circ$  to  $180^\circ$ , with clockwise (east) changes in direction defined as positive, e.g.  $\alpha_3^{II}$  and  $\alpha_4^{II}$ , and anticlockwise (west) changes defined as negative, e.g.  $\alpha_3^I$  and  $\alpha_4^I$ .

The mean angular difference  $d_a$  between two paths  $traj^I$  and  $traj^{II}$  can then be defined as

$$d_a = \frac{\sum_{k=3}^n \Delta\alpha_k}{180 * (n - 2)},$$

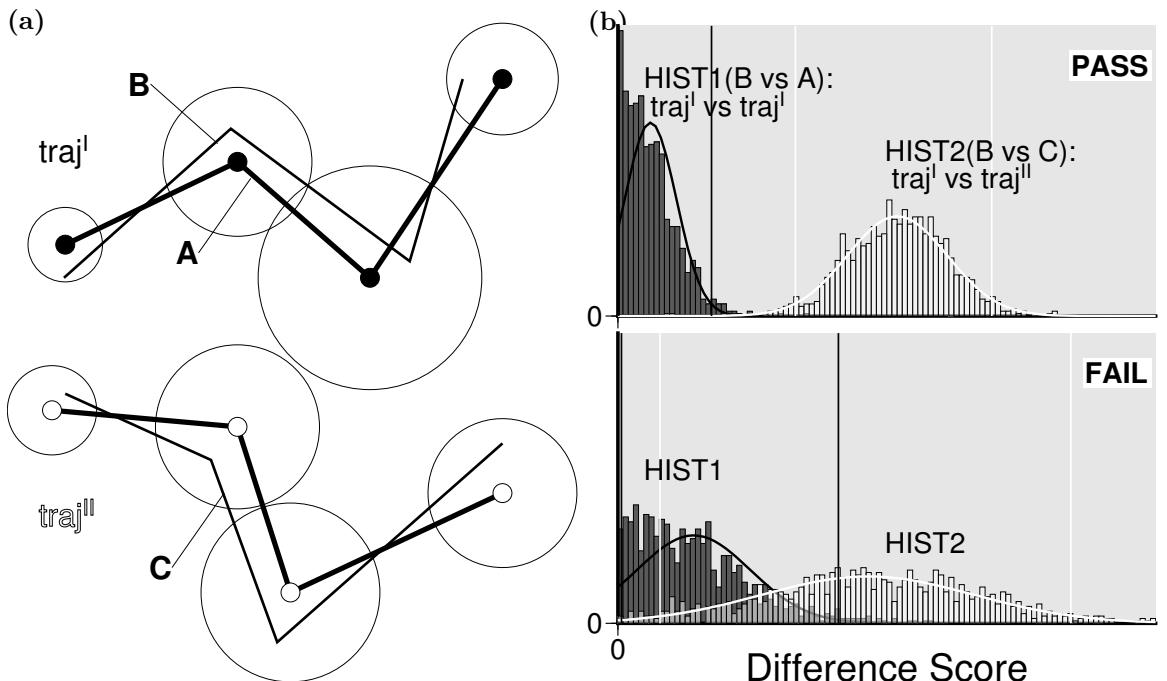
where

$$\Delta\alpha_k = \begin{cases} |\alpha_k^I - \alpha_k^{II}|, & \text{if } |\alpha_k^I - \alpha_k^{II}| \leq 180; \\ 360 - |\alpha_k^I - \alpha_k^{II}|, & \text{otherwise.} \end{cases} \quad \forall k \in \{3, 4, \dots, n\}.$$

$d_a$  is normalised by the maximum possible angular deviation of  $180^\circ$ . A score of 0 indicates exactly matching changes in the bearing of coeval segments along the length of the two paths, and a score of 1 indicates all segment bearings are antiparallel.

#### 2.2.4.3 Significance Testing of Shape Difference

Due to associated spatial uncertainty, the mean poles in an APWP trace out one possible path within a range of possible geometries (Fig. 2.5a). If the length and angular difference scores for one path fall within the range of possible scores for the other, two APWPs may not in fact be significantly different from each other. Significance testing for the shape difference scores is performed on each coeval segment pair as follows (Fig. 2.5b):



**Figure 2.5:** Significance testing for the geometric metrics,  $d_l$  and  $d_a$ . (a) Illustration of how paths  $\text{traj}^I$  and  $\text{traj}^{II}$  can be re-sampled within their uncertainty ellipses, with B being a possible trajectory of  $\text{traj}^I$  and C being a possible trajectory of  $\text{traj}^{II}$ . (b) Upper: If the 95% confidence interval (black vertical lines are its upper and lower bounds) for the distribution of difference scores HIST1, generated by comparing multiple resamplings of  $\text{traj}^I$  with the original trajectory (A vs B) does not overlap with the 95% confidence interval (bounded by white vertical lines) for the distribution of scores HIST2, generated by comparing resamplings of  $\text{traj}^I$  and  $\text{traj}^{II}$  (B vs C), then the original difference score for  $\text{traj}^I$  and  $\text{traj}^{II}$  is statistically distinguishable; Lower: If the confidence intervals overlap, then the two paths are not distinguishable.

- A bootstrapped distribution of possible geometries for each segment in a path can be created by resampling the two mean poles that define the original segment, in the same manner as described in Section 2.2.3 and the Supplementary Information.
- A histogram of statistically indistinguishable length and/or angular difference scores (HIST1, Fig. 2.5b) is created by comparing the resampled paths with the original for each  $traj^I$  segment.
- This distribution is then compared to the histogram of difference scores created by resampling the coeval segments of  $traj^I$  and  $traj^{II}$  (HIST2, Fig. 2.5b).
- If the two bootstrapped distributions HIST2 and HIST1 do not overlap at the given significance level (e.g. the upper and lower bounds of a 95% confidence intervals, Fig. 2.5b), then the difference score is interpreted to be significant. If not, then the bearings or lengths of the coeval segments are indistinguishable.

These tests allow a filter for spatial uncertainty to be added to the  $d_a$  and  $d_l$  metrics: prior to summation and normalisation, the difference score is set to zero for the coeval segments of  $traj^I$  and  $traj^{II}$  that are statistically indistinguishable.

### 2.2.5 Composite Path Difference

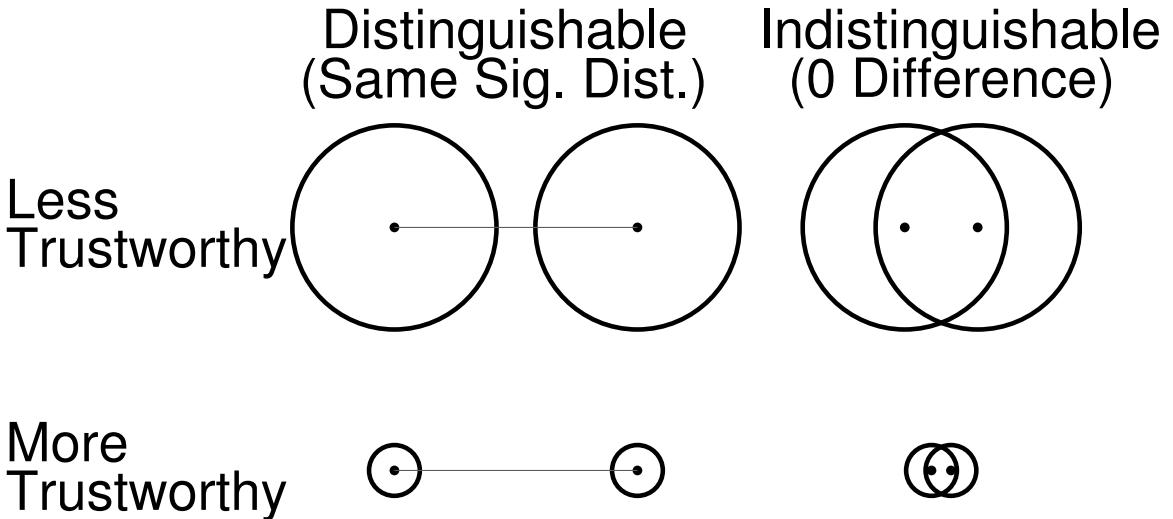
The three difference measures described above can be combined into a composite path difference ( $\mathcal{CPD}$ ) by means of a simple linear weighting rule,

$$\mathcal{CPD} = W_s \cdot d_s + W_a \cdot d_a + W_l \cdot d_l$$

for  $0 < W_s, W_a, W_l < 1$ , where  $W_s$ ,  $W_a$  and  $W_l$  are weighting coefficients that sum to 1. Different weighting values allow the relative influences of spatial and geometric (length and angular) similarity to be varied (Section 2.3.2).

### 2.2.6 Fit Quality

The three metrics are all tested to be significant based on the spatial uncertainties. However, the larger the uncertainties are, the less trustworthy the significant difference scores are (Fig. 2.6). Accordingly, we bring in a concept of “Fit Quality”, along the classification scheme of the reversal test [31]. For each mean pole of an APWP, it assigns a score based on the size of the spatial uncertainty (radius: A95, or  $(DM+DP)/2$ ): 1 if it is  $\leq 5^\circ$ , 2 if  $5^\circ < r \leq 10^\circ$ , 3 if  $10^\circ < r \leq 20^\circ$ , and 4 if it is  $> 20^\circ$ .



**Figure 2.6:** Difference scores from APWPs with large uncertainties are less trustworthy.

These values are averaged for each APWP to give a “Fit Quality” score (from 1 to 4) for the difference score. This is then converted into an A/B/C/D letter grade, A if the average is  $<1.5$ , B if  $1.5 \leq \text{avg} < 2.5$ , C if  $2.5 \leq \text{avg} < 3.5$ , and D if it is  $\geq 3.5$ , to indicate how easy it is to generate a low difference score. In other words, an A grade indicates that most poles are well-constrained and so it is fairly hard to have an indistinguishable path and a low difference score. A D grade indicates that most poles have large uncertainties so it is much easier to have a low difference score.

In addition, a short APWP segment tends to result in overlap of its two end poles’ spatial uncertainties. For example, if an APWP is generated at intervals of 10 Myr, the longest realistic segment-length would be about  $27^\circ$  [48, 29, the maximum rate of plate movement is about 30 cm/yr]. So the uncertainty size needs be less than about  $13.5^\circ$  on average to make the segment length trustworthy. Therefore, to a certain extent, the “Fit Quality” also reflects the quality of the length metric if we give each path a grade for an APWP pair, e.g. A-A. The angular metric’s quality is related to both coeval mean poles and successive mean poles, so it has already been involved in the spatial and length quality. Moreover, given the fact that usually the mean significant length and angular differences are much lower than the mean significant spatial difference (e.g., Figs. 2.8b, 2.8d, 2.8f and Figs. 2.11b, 2.11d, 2.11f), the “Fit Quality” is capable to manifest the overall quality of all the three metrics and we should trust the difference scores if we get a B-B grade at least.

For example, *Pairs a, b, c, d* and *f*’s fit quality score is all 1.809–1.809, so their fit quality is B–B, which means that the mean poles in these APWP pairs have intermediate uncertainties on average so it is relatively hard to have a low difference score. *Pair e*’s fit quality is B–A (1.809–1.085). *Pair g*’s fit quality is A–A (0–0).

## 2.3 Results and Discussion

### 2.3.1 Discrimination of Difference Metrics

The performance of the individual metrics were tested by generating and ranking scores for each pair in Fig. 2.3. Scores for comparisons of the full 530 Myr paths, and sequential 100–130 Myr subsections were calculated for path pairs with (Fig. 2.7) and without (Fig. 2.8) poles with zero spatial uncertainties at 350 Ma, 360 Ma, 380 Ma, 390 Ma, 450 Ma, 460 Ma and 520 Ma calculated using linear interpolation by [59].

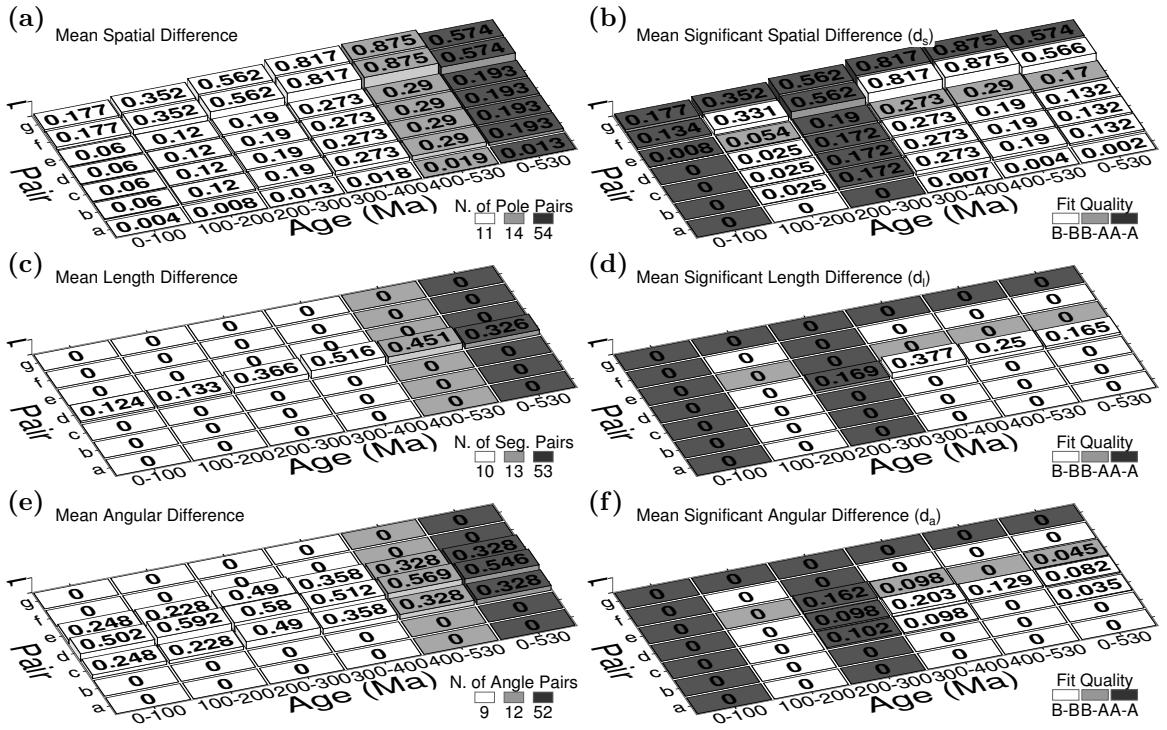
#### 2.3.1.1 $d_s$

If ordered only in terms of spatial similarity, the desired order for the seven APWP pairs (Fig. 2.3), from most similar (lowest  $d_s$ ) to least similar (highest  $d_s$ ) is

$$\textit{Pair a} < \textit{Pair b} \approx \textit{Pair c} \approx \textit{Pair d} < \textit{Pair e} < \textit{Pair f} < \textit{Pair g}. \quad (2.1)$$

This ordering is based largely on the mean GCD separations of each pair (Fig. 2.3), but also takes uncertainties into account: the relatively smaller uncertainty ellipses of *Pair e* and *Pair g* should lead to a higher  $d_s$  than *Pairs b-d* and *Pair f*, respectively. Without significance testing,  $d_s$  is directly proportional to mean GCD (Fig. 2.7a), which does not result in unique  $d_s$  for *Pair e* and *Pair g*. Significance testing reproduces the desired order (Fig. 2.7b).

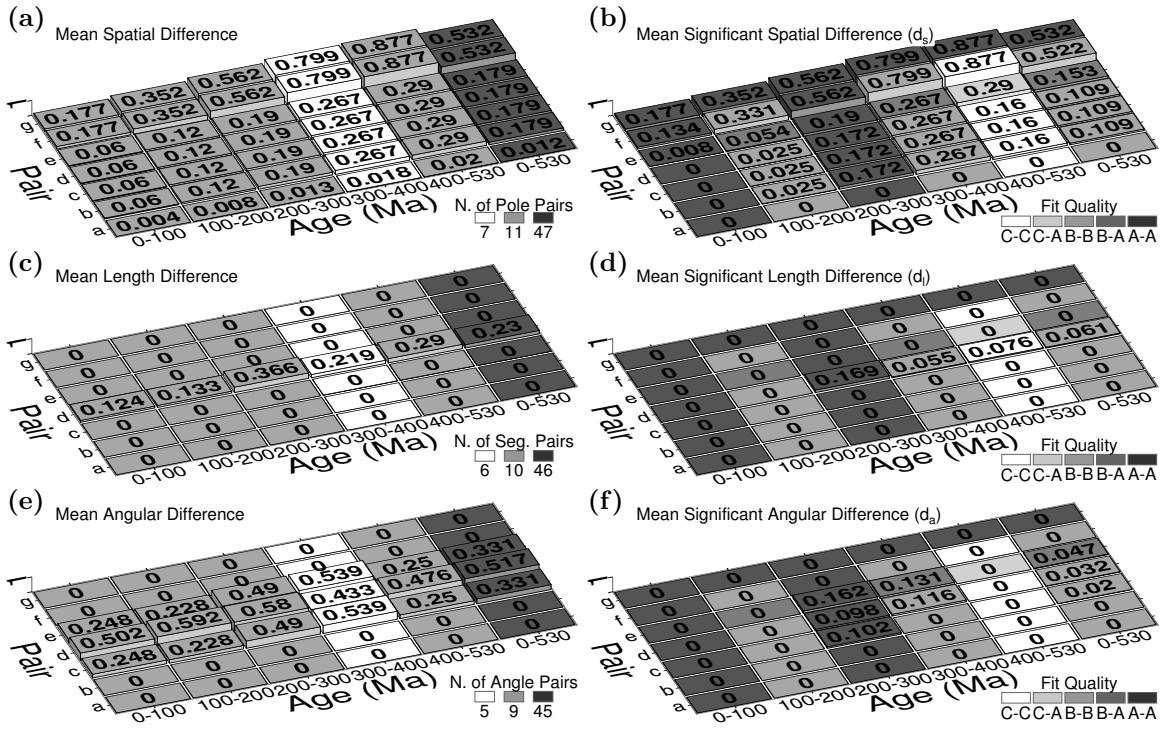
Most  $d_s$  scores are also reduced with significance testing, with the largest reductions occurring where the path separations are low and 95% confidence ellipses for coeval poles are more likely to overlap (e.g.,  $d_s$  for *Pair a* is reduced by 85%,  $d_s$  for *Pair f* is reduced by <2%).  $d_s$  that approach 0 for the 0–100 Ma and 100–200 Ma sub-paths, which are located close to the Euler pole used to separate the pairs and therefore remain in close proximity even after large rotations, also illustrate this effect. With significance testing, the 0–100 and 100–200 Ma sub-paths of *Pair f* and *Pair g* can still be distinguished (Fig. 2.8b and also Fig. 2.7b), due to no spatial uncertainty assigned to *Pair g*. In contrast, the older 300–400 Ma and 400–530 Ma sub-paths have a larger  $d_s$  than the whole path. This is because the 350 Ma, 360 Ma, 380 Ma, 390 Ma, 450 Ma, 460 Ma and 520 Ma pole coordinates are interpolated [59], and thus have no assigned spatial uncertainty on any of the test paths. Without the interpolated poles,  $d_s$  is always zero for *Pair a* and any of its sub-paths (Fig. 2.8b), which is expected.



**Figure 2.7:** Mean spatial, length and angular differences between two paths of the seven APWP pairs shown in Fig. 2.3. Left column: results without significance testing imposed in the metric; Right column: results with significance testing. Note that the spatial difference results of *Pairs b, c* and **d** are always the same for both the untested case (a) and the tested case (b). In addition, for those segments that do not begin from 0 Ma, their beginning segments are different from the 0–100 Ma sub-path’s and the full path’s. For example, for the 200–300 Ma sub-path, its beginning segment is the 200–210 Ma one.

Even with significance testing,  $d_s$  for *Pairs b-d* are the same (Fig. 2.7a and Fig. 2.7b) despite their different geometries (Fig. 2.3b, Fig. 2.3c and Fig. 2.3d), because GCDs between coeval poles and their uncertainties are the same. This result emphasises that a spatial difference metric alone cannot discriminate these pairs from each other. The comparison of *Pairs c* and **e** indicates that well-constrained mean poles with lower uncertainties make it relatively harder to have an indistinguishable APWP and a low difference score.

In summary, as Fig. 2.8b and also Fig. 2.7b illustrate,  $d_s$ , scaling with mean significant GCD, reproduces the expected order of spatial similarity (Order (2.1)) for the full path. It also compensates for the deficiency of the algorithm without statistical test (Fig. 2.8a and Fig. 2.7a) in differentiating *Pair f* and *Pair g*. Although our algorithm also works for APWPs with interpolations (e.g., Fig. 2.7; see how we do significance testing on the interpolated mean poles in the Supplementary Information), a meaningful and valid analysis should be based on the results with uninterpolated paleomagnetic APWPs (e.g., Fig. 2.8).



**Figure 2.8:** Mean spatial, length and angular differences between two paths of the seven APWP pairs with no interpolated poles shown in Fig. 2.3. Left column: results without significance testing imposed in the metric; Right column: results with significance testing. See explanation of Fig. 2.7.

### 2.3.1.2 $d_l$

When ordered only according to the length similarity  $d_l$  the expected order is

$$0 = \text{Pair } \mathbf{a} = \text{Pair } \mathbf{b} = \text{Pair } \mathbf{c} = \text{Pair } \mathbf{e} = \text{Pair } \mathbf{f} = \text{Pair } \mathbf{g} < \text{Pair } \mathbf{d} \quad (2.2)$$

Because only the path generated for *Pair d* allowed the length of coeval segments to vary, it is expected that other five pairs of APWPs have zero  $d_l$  for both the full-path and the five specified sub-paths even prior to significance testing (Fig. 2.8c and Fig. 2.8d), and this expected order is trivially reproduced. The effect of significance testing (Fig. 2.8d) is to substantially reduce  $d_l$ . Many segment length differences do not pass the significance test because the angular uncertainties of the poles that define individual segments are large compared to the length of those segments.

### 2.3.1.3 $d_a$

If ordered only according to angular similarity  $d_a$ , the expected order is

$$0 = \text{Pair } \mathbf{a} = \text{Pair } \mathbf{b} = \text{Pair } \mathbf{f} = \text{Pair } \mathbf{g} \leq \text{Pair } \mathbf{c} \leq \text{Pair } \mathbf{d} < \text{Pair } \mathbf{e} \quad (2.3)$$

Because path geometries are not altered by a simple Euler rotation, only *Pairs c*, *d* and *e* are expected to have a non-zero  $d_a$ .  $d_a$  for *Pair d* and *Pair e* should be

larger than *Pair c*'s due to more geometric variation and lower spatial uncertainties, respectively, although the expected ordering of *Pairs d* and *e* is less immediately obvious from visual inspection.

Without significance testing, non-zero  $d_a$  for *Pairs c, d* and *e* are consistently generated for both the full path and sub-paths (Fig. 2.8e).  $d_a$  for *Pair d* is usually higher (Fig. 2.8e), but there is no discrimination between *Pairs c* and *e*, which have the same score because geometrically they are identical. When significance testing is applied  $d_a$  is markedly reduced (Fig. 2.8f vs Fig. 2.8e), and is actually reduced to 0 for the two youngest and oldest sub-paths in all cases. This is somewhat expected because the segment lengths of the APWPs being tested are often of the same order as the angular uncertainty in their spatial position. As a result, a large range of different path geometries are possible within the specified uncertainty bounds, and the bearing of coeval segments has to be very large for the difference to be significant.

For the full paths and the 200–300 and 300–400 Ma sub-paths where  $d_a$  after significance testing is non-zero for *Pairs d, e* and (usually) *c*, *Pair e* can now be discriminated from *Pair c*, and consistently has the highest  $d_a$  of the 3 pairs.

In summary, our angular difference algorithm with statistical test (Fig. 2.8f) reproduces the expected order of angular similarity (Order (2.3)).

#### 2.3.1.4 CPD

When the seven different APWP pairs (Fig. 2.3) are rank-ordered in terms of the three criteria combined, their expected order is

$$\text{Pair a} < \text{Pair b} \leq \text{Pair c} < \text{Pair d} ? \text{Pair e} ? \text{Pair f} < \text{Pair g} \quad (2.4)$$

In order to be useful, a path difference metric needs to reproduce this order. Note that a question mark is put on top of the “less than” symbols between *Pairs d* and *e*, and *Pairs e* and *f* because when comparing pairs with different spatial separation, geometric difference, and relative spatial uncertainty, it can be hard to objectively define which “should” have the highest similarity, and the ordering will depend on the relative weighting of  $d_s$ ,  $d_l$  and  $d_a$ . If the weightings are equal (i.e.,  $W_s = W_l = W_a = \frac{1}{3}$ ), significant CPD scores for paths without interpolated poles (i.e., using scores from Fig. 2.8) reproduce the expected order:

$$\mathbf{a}(0) < \mathbf{b}(0.036) < \mathbf{c}(0.043) < \mathbf{d}(0.067) \approx \mathbf{e}(0.067) < \mathbf{f}(0.174) < \mathbf{g}(0.177),$$

however *Pairs d* and *e* have almost identical scores and are not discriminated. However, their fit quality (B-B for *Pair d* and B-A for *Pair e*; Fig. 2.8b, Fig. 2.8d and

Fig. 2.8f) indicates that *Pair e*'s  $\mathcal{CPD}$  is relatively more trustworthy. This order might also not be preserved with different applied weights. The impact of weighting will be discussed in the following section.

### 2.3.2 A Discussion on Weights

Although  $W_s$ ,  $W_a$  and  $W_l$  can be defined by user, this is subjective.

However, we do explicitly know that: when comparing two APWP pairs, our aim is to find the one whose similarity ranks higher. A simple subtraction between unsolved (because of unknown weights)  $\mathcal{CPDs}$  can help determine which pair's similarity ranks higher. If a positive difference is obtained, no matter what  $W_s$ ,  $W_a$  and  $W_l$  values are assigned, the subtrahend pair's similarity ranks higher; if the difference is always negative, the minuend pair's similarity is always higher. In addition, the difference could be always zero. Interpretation is straightforward in these three scenarios. However, for some pairs, a positive, zero or negative CPD difference could result depending on the chosen weightings.

For example, for the full (i.e., 0–530 Ma) path with no interpolated poles, the mean significant spatial, length and angular differences  $d_s$ ,  $d_l$  and  $d_a$  are known (Fig. 2.8b, Fig. 2.8d and Fig. 2.8f). Also we know  $W_l = 1 - W_s - W_a$ . Then we do subtractions

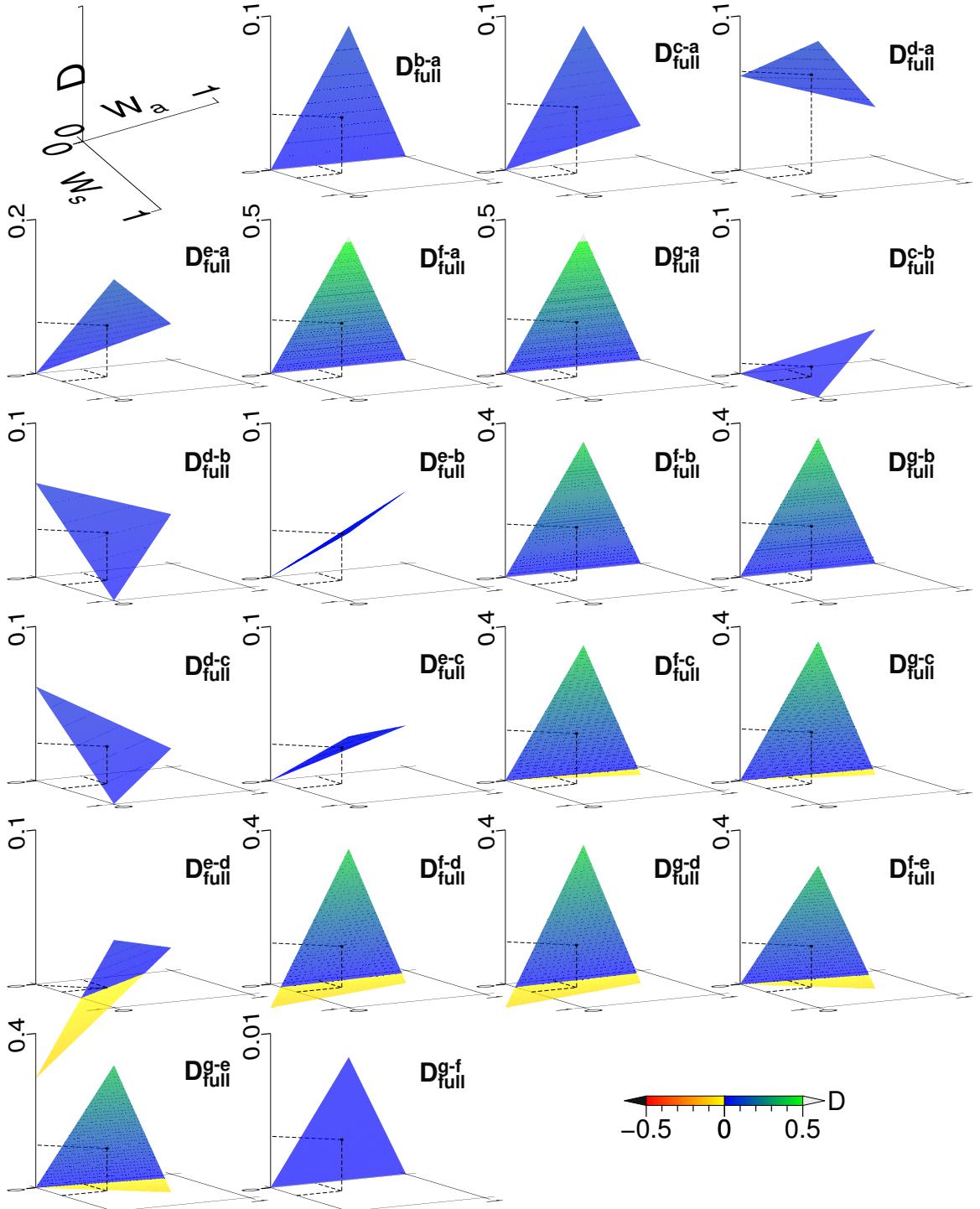
of  $\mathcal{CPD}s$  from each two APWP pairs:

$$\left\{ \begin{array}{l} D_{full}^{b-a} = \mathcal{CPD}_{full}^b - \mathcal{CPD}_{full}^a = 0.109W_s \\ D_{full}^{c-a} = \mathcal{CPD}_{full}^c - \mathcal{CPD}_{full}^a = 0.109W_s + 0.02W_a \\ D_{full}^{d-a} = \mathcal{CPD}_{full}^d - \mathcal{CPD}_{full}^a = 0.038W_s - 0.029W_a + 0.061 \\ D_{full}^{e-a} = \mathcal{CPD}_{full}^e - \mathcal{CPD}_{full}^a = 0.153W_s + 0.047W_a \\ D_{full}^{f-a} = \mathcal{CPD}_{full}^f - \mathcal{CPD}_{full}^a = 0.522W_s \\ D_{full}^{g-a} = \mathcal{CPD}_{full}^g - \mathcal{CPD}_{full}^a = 0.532W_s \\ D_{full}^{c-b} = \mathcal{CPD}_{full}^c - \mathcal{CPD}_{full}^b = 0.02W_a \\ D_{full}^{d-b} = \mathcal{CPD}_{full}^d - \mathcal{CPD}_{full}^b = -0.061W_s - 0.029W_a + 0.061 \\ D_{full}^{e-b} = \mathcal{CPD}_{full}^e - \mathcal{CPD}_{full}^b = 0.044W_s + 0.047W_a \\ D_{full}^{f-b} = \mathcal{CPD}_{full}^f - \mathcal{CPD}_{full}^b = 0.413W_s \\ D_{full}^{g-b} = \mathcal{CPD}_{full}^g - \mathcal{CPD}_{full}^b = 0.423W_s \\ D_{full}^{d-c} = \mathcal{CPD}_{full}^d - \mathcal{CPD}_{full}^c = -0.061W_s - 0.049W_a + 0.061 \\ D_{full}^{e-c} = \mathcal{CPD}_{full}^e - \mathcal{CPD}_{full}^c = 0.044W_s + 0.027W_a \\ D_{full}^{f-c} = \mathcal{CPD}_{full}^f - \mathcal{CPD}_{full}^c = 0.413W_s - 0.02W_a \\ D_{full}^{g-c} = \mathcal{CPD}_{full}^g - \mathcal{CPD}_{full}^c = 0.423W_s - 0.02W_a \\ D_{full}^{e-d} = \mathcal{CPD}_{full}^e - \mathcal{CPD}_{full}^d = 0.105W_s + 0.076W_a - 0.061 \\ D_{full}^{f-d} = \mathcal{CPD}_{full}^f - \mathcal{CPD}_{full}^d = 0.474W_s + 0.029W_a - 0.061 \\ D_{full}^{g-d} = \mathcal{CPD}_{full}^g - \mathcal{CPD}_{full}^d = 0.484W_s + 0.029W_a - 0.061 \\ D_{full}^{f-e} = \mathcal{CPD}_{full}^f - \mathcal{CPD}_{full}^e = 0.369W_s - 0.047W_a \\ D_{full}^{g-e} = \mathcal{CPD}_{full}^g - \mathcal{CPD}_{full}^e = 0.379W_s - 0.047W_a \\ D_{full}^{g-f} = \mathcal{CPD}_{full}^g - \mathcal{CPD}_{full}^f = 0.01W_s, \end{array} \right. \quad (2.1)$$

and we also have the following constraints of feasible regions

$$\left\{ \begin{array}{l} 0 < W_s < 1 \\ 0 < W_a < 1 \\ 0 < W_s + W_a < 1. \end{array} \right. \quad (2.2)$$

The linear equations (2.1) subject to (2.2) can be graphed in the three-variable ( $W_s$ - $W_a$ - $D$ ) coordinate system (Fig. 2.9). For all possible combinations of  $W_s$  and  $W_a$ , there is a consistent ordering of CPD scores such that  $\mathbf{a} < \mathbf{b} < \mathbf{c} < \mathbf{d}$ ,  $\mathbf{a} < \mathbf{b} < \mathbf{c} < \mathbf{e}$ , and  $\mathbf{a} < \mathbf{b} < \mathbf{f} < \mathbf{g}$  (Fig. 2.9). However, the ranking for *Pairs f, g* and *Pairs c, d, e*, or *Pair d* and *Pair e* has multiple possibilities, because their differences can be positive, negative or zero (Fig. 2.9). For this situation, assigning equal weights is recommended (giving centroid of all possible  $D$ s, Fig. 2.9; see also Supplementary Information for testing equally likely random weights) for deciding the rank order. With equal weights used, the order from most similar pair to least similar pair is  $\mathbf{a} < \mathbf{b} < \mathbf{c} < \mathbf{d} \approx \mathbf{e} < \mathbf{f} < \mathbf{g}$ . These conclusions do not contradict the expected Order (2.4).



**Figure 2.9:** Graphical depiction of  $\mathcal{CPD}$  differences ( $D$ ) between the seven APWP pairs for full-path (0–530 Ma) comparisons. If the planes derived from the equations intersect the  $D = 0$  plane at a point or in a straight line, that point or the infinite number of points (i.e., sets of  $W_s$ ,  $W_a$  values) on the line of intersection represent that the similarities of the minuend pair and the subtrahend pair are equal to each other. If  $D > 0$  or  $D < 0$  on the planes of the equations, the subtrahend pair or the minuend pair respectively owns higher similarities. The square dot locates the result when  $W_s = W_a = W_l = \frac{1}{3}$ .

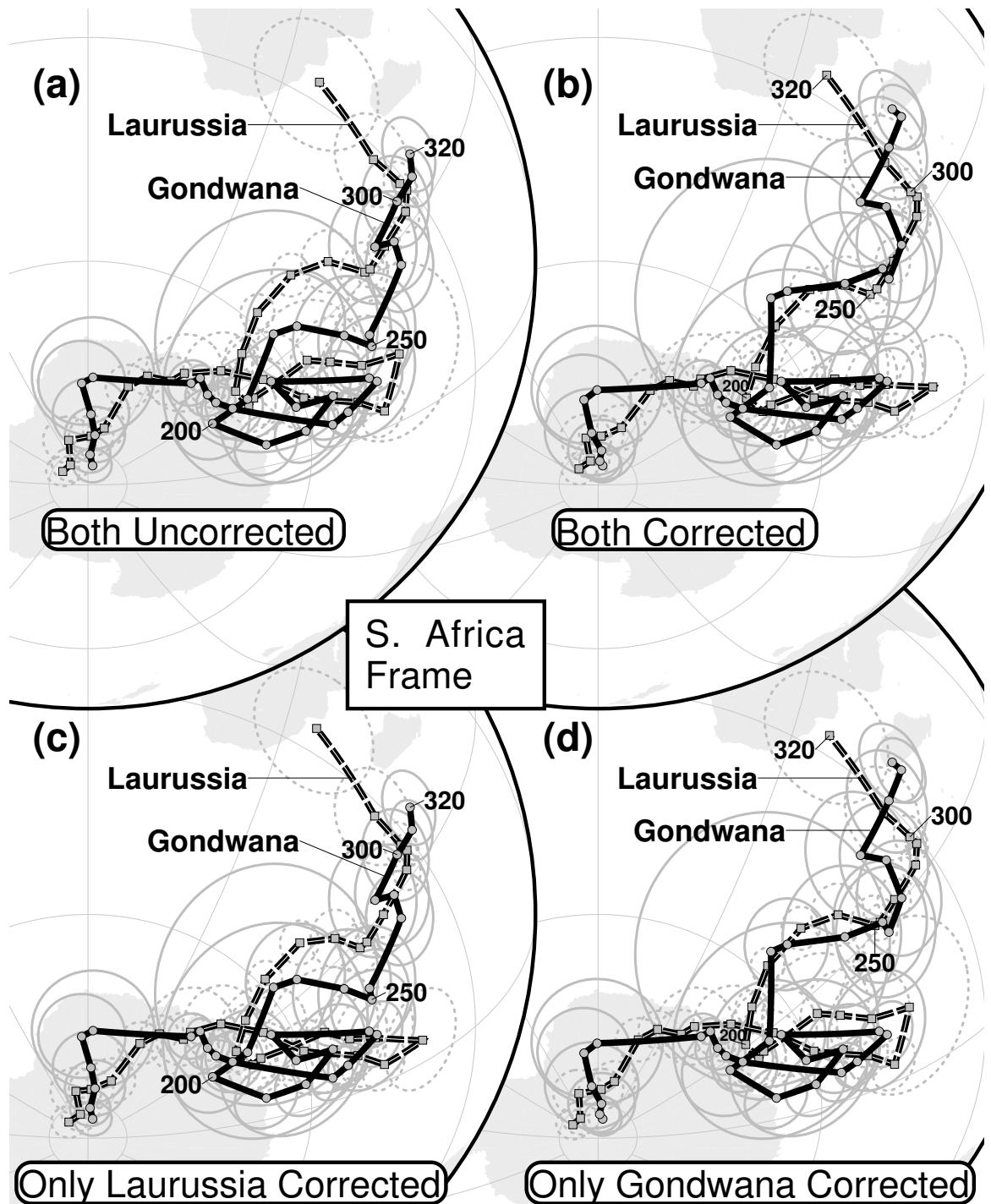
In summary, as Fig. 2.3 illustrates, mean GCD has trouble discriminating between *Pairs f* and *g*, and *Pairs c* and *e*, and also between intermediate similarities where the differences are mainly in path geometry (*Pairs b, c, d*). Our algorithm provides an improved solution for this problem. Obtaining similarity order can be straightforward, such as for *Pairs a, b, c, and d*, *Pairs a, b, c, and e*, or *Pairs a, b, f, and g*. In other words, when one APWP's three individual metrics are all greater than or equal to, or less than or equal to the other one's, weighting is irrelevant. However, when the ranking of individual metrics for a pair are not consistent (e.g., *Pair f* and *Pair c*; Fig. 2.9), obtaining similarity order is less straightforward. When this occurs, equally weighting is recommended for concluding the final rank order.

### 2.3.3 Application to Real Paleomagnetic Data

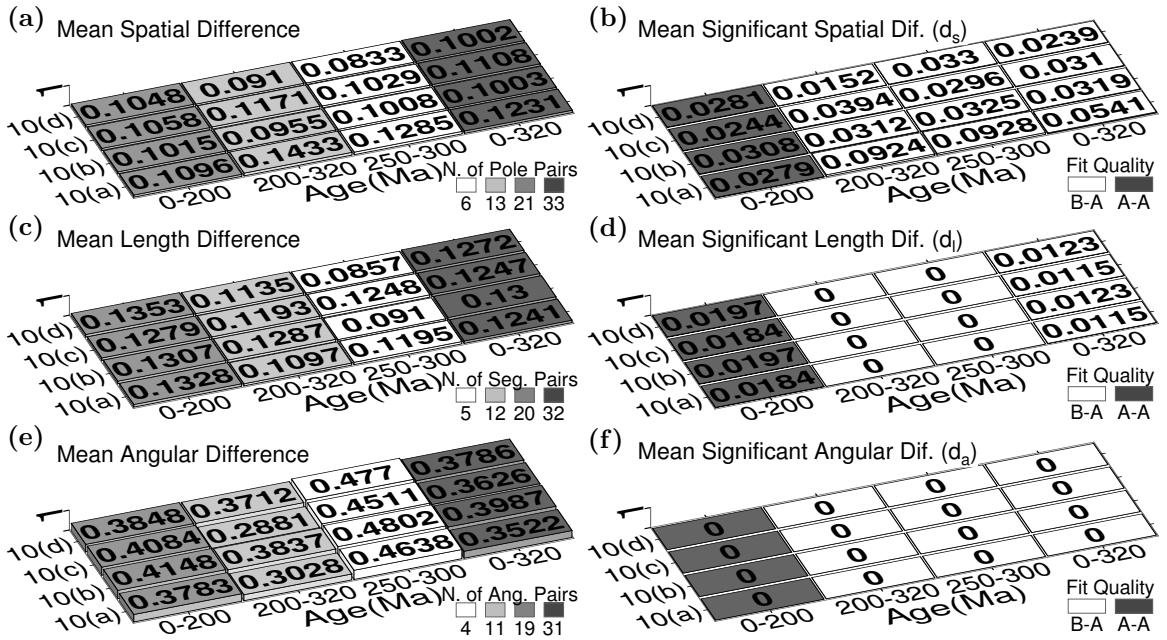
To illustrate how these metrics might be useful when applied to real paleomagnetic data, we compare 320–0 Ma APWPs for Gondwana and Laurussia calculated using a running mean method by [59], with paleopoles from sedimentary rocks both uncorrected (Fig. 2.10a) and bulk corrected for inclination shallowing ( $f=0.6$ ; Fig. 2.10b).

When comparing the 320–0 Ma paths [59, in their Fig. 13a] observed that a bulk correction for inclination shallowing applied to poles from sedimentary rocks reduced the mean GCD separation between poles, particularly in the Permian section of the path.

The full-path (320–0 Ma)  $d_s$  scores of *Pair 2.10a* and *Pair 2.10b* (Fig. 2.11) confirm that the corrected Gondwana and Laurussia APWPs (*Pair 2.10b*) are more similar than the uncorrected pair (*Pair 2.10a*). This difference is significant (Fig. 2.11b), and it is principally the result of an improved fit (lower  $d_s$ ) in the Permian (300–250 Ma) and Carboniferous to Triassic (320–200 Ma) sections. When geometry is considered,  $d_l$  and  $d_a$  scores without significance testing (Fig. 2.11c and Fig. 2.11e) are actually higher for corrected (*Pair 2.10b*) than uncorrected pair (*Pair 2.10a*), particularly in the same Permo-Triassic segment; however, none of these differences are statistically significant (Fig. 2.11d and Fig. 2.11f). Therefore, the equally weighted CPD score is actually worse for corrected pair (*Pair 2.10b*) (0.213 vs 0.1998 for *Pair 2.10a*), whereas with significance testing applied the CPD of the corrected pair is a clear improvement (0.0147 vs 0.0218). This emphasises the importance of significance testing for the geometric scores, particularly where the spatial uncertainties are relatively large compared to the step length, as is indicated by the fit quality score (B-A, Gondwana-Laurussia).



**Figure 2.10:** (a) 320–0 Ma APWPs (10 Myr step) for Gondwana and Laurussia [59, rotated to Southern Africa frame using the rotations from]; (b) as (a). but both paths corrected for inclination shallowing; (c) and (d) as (a). but only Laurussia path and only Gondwana path respectively corrected for inclination shallowing. Note that all the paleomagnetic APWPs are reproduced using the same moving average method and same paleopoles for the APWPs in Figure 13(a) of [59]. Azimuthal Orthographic projection.



**Figure 2.11:** Mean spatial, length and angular differences between two paths of the four APWP pairs shown in Fig. 2.10. Left column: results without significance testing imposed in the metric; Right column: results with testing. Note that Pair 2.10b with both APWPs corrected for inclination shallowing is not the most similar pair according to both the untested (left column) and tested (right column) results.

Furthermore, if this analysis is extended to compare an Laurussia APWP corrected for inclination flattening with a Gondwana APWP that has not (*Pair 2.10c*), and vice versa (*Pair 2.10d*), removing the blanket correction from Gondwana poles has only a minor effect on the overall  $d_s$ , whilst removing it from the Laurussia poles actually improves  $d_s$  after significance testing (Fig. 2.11a and Fig. 2.11b). Comparison of the changes in sub-path scores for *Pairs 2.10b, 2.10c and 2.10d*) suggests that the effect of the bulk flattening correction is sometimes positive and sometimes negative for different time periods, supporting arguments that flattening corrections need to be more judiciously applied [5, 4, for example]. A more detailed study might further constrain the regions, continents and/or time periods where a correction is appropriate, and those where it is not. But for the purposes of this paper, this overview is sufficient to demonstrate the potential usefulness of our difference metrics when considering the effect of different techniques and corrections used to generate an APWP.

## 2.4 Conclusions

A new synthetic evaluation method is proposed in this paper to serve as a numerical tool for the purpose of quantitatively matching paleomagnetic APWPs. Multidimensional information tested by bootstrapping, such as overlap of coeval poles and shape

of paths, are taken into account in the algorithm. This method can also be utilized to detect APWP subsections' degree of similarity by changing trajectory beginning and end poles. As an example of how this method can be applied, we confirm a previously published suggestion that applying a blanket correction for inclination shallowing in sedimentary rocks does significantly improve the fit between Carboniferous to Recent APWPs for Gondwana and Laurussia. However a more detailed analysis also indicates that such blanket inclination corrections are unlikely to produce the best possible fit.

## Acknowledgments

The similarity measuring tool depends on the open-source softwares PmagPy [54] and GMT [65]. All images are produced using GMT [65]. We thank 3 anonymous reviewers for suggestions that greatly improved a previous version of this manuscript.

## Computer Code Availability

**Name of Code** Spherical Path Comparison (spComparison)

**Developer** Chenjian Fu and Christopher J. Rowan

**Contact Address** 221 McGilvrey Hall, 325 S Lincoln St, Kent, OH 44242 USA

**Telephone Number** 4408479166

**E-mail** cfu3@kent.edu

**Year First Available** 2019

**Hardware Required** Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz or higher; 8 GB DDR3 RAM or higher

**Software Required** GMT5 or higher; Python 3.6 or higher; Bash 4.4.23 or higher; Linux as the best platform, macOS also fine, for Windows further setup needed

**Program Language** Python 3 and Shell Scripting (Bash)

**Program Size** 75 KB

**Details on How to Access the Source Code** The source code can be accessed from [https://github.com/f-i/Spherical\\_Path\\_Comparison](https://github.com/f-i/Spherical_Path_Comparison). Please use the provided Jupyter Notebook file “demo.ipynb” to reproduce some calculations shown in the paper.

# Chapter 3

## Finding the Way(s) to Make a Reliable APWP

*This chapter mainly describes how to generate paleomagnetic APWPs using 168 different methods, and then the application of the new APW path similarity measuring tool used in finding the best APWP generating method(s). 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.*

(Chapter 3 is also openly accessible from [https://github.com/f-i/making\\_of\\_reliable\\_APWPs](https://github.com/f-i/making_of_reliable_APWPs).)

## 3.1 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.

### 3.1.1 Not All Data Are Created Equal

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

#### 3.1.1.1 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.

#### 3.1.1.2 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.

### 3.1.1.3 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 [51]. 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

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.

### 3.1.1.4 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 [30, 41, 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.

### 3.1.1.5 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?

## 3.1.2 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 [61] and the BC02 selection criteria [3]. Briefly, the V90 criteria for paleomagnetic results includes seven criteria: (1) Well determined age; (2) At least 25 samples with Fisher [15] 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 [58]. 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 [58]. Q factor is indeed a very straightforward way to get a quantalized reliability score. Also it then can be conveniently used in the later calculations of APWPs [58]. But at the same time this is a fairly basic filter that lumps together criteria that may not be equally important. Compared with V90, the BC02 criteria suggests stricter filtering, e.g., using only poles with at least 6 sampling sites and 36 samples, each site having  $\alpha_{95}$  less than  $10^\circ$  in the Cenozoic and  $15^\circ$  in the Mesozoic. B02 is also straightforward and convenient to use, but some

**Table 3.1:** 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. $\alpha_{95}$
EP95	Radius of circle of 95% confidence about paleopole position
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

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.

## 3.2 Methods

### 3.2.1 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. 3.2 and 3.3). 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.

### 3.2.2 Paleomagnetic Data

#### 3.2.2.1 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.

### 3.2.2.2 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 [25] 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 [66]. Following extraction, 58 poles from southwestern North America that have been affected by regional rotations since 36 Ma [33], were removed. The final dataset consists of 135 paleopoles (Fig. 3.1), with 76 of them (about 56.3%; average age uncertainty 14.1974 Myr, average EP95 9.2627°) sampled from dominantly igneous sequences, 56 (about 41.5%; average age uncertainty 27.5179 Myr, average EP95 10.5027°) 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 [66] was used, but following extraction 31 paleopoles associated with parts of the northern margin that have undergone regional rotations since the Jurassic [17] were removed. The final dataset consists of 75 paleopoles (Fig. 3.2), with 39 of them (52%; average age uncertainty 5.2308 Myr, average EP95 7.7079°) sampled from dominantly igneous sequences and 36 (48%; average age uncertainty 14.1944 Myr, average EP95 7.0222°) 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 [66] 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 [66]. The final dataset consists of 99 paleopoles (Fig. 3.3), with 61 of them (61.6%; average age uncertainty 23.4918

Myr, average EP95 10.9158°) sampled from dominantly igneous sequences, and 38 (38.4%; average age uncertainty 23.2632 Myr, average EP95 9.4097°) sampled from mostly sedimentary sequences, including 9 from redbeds. The temporal distribution of poles is relatively uniform.

Compared with North American (Fig. 3.1) and Australian (Fig. 3.3) paleopoles, Indian paleopoles are in relatively lower density in general except during the period of about 70–60 Ma (Fig. 3.2).

### 3.2.3 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.

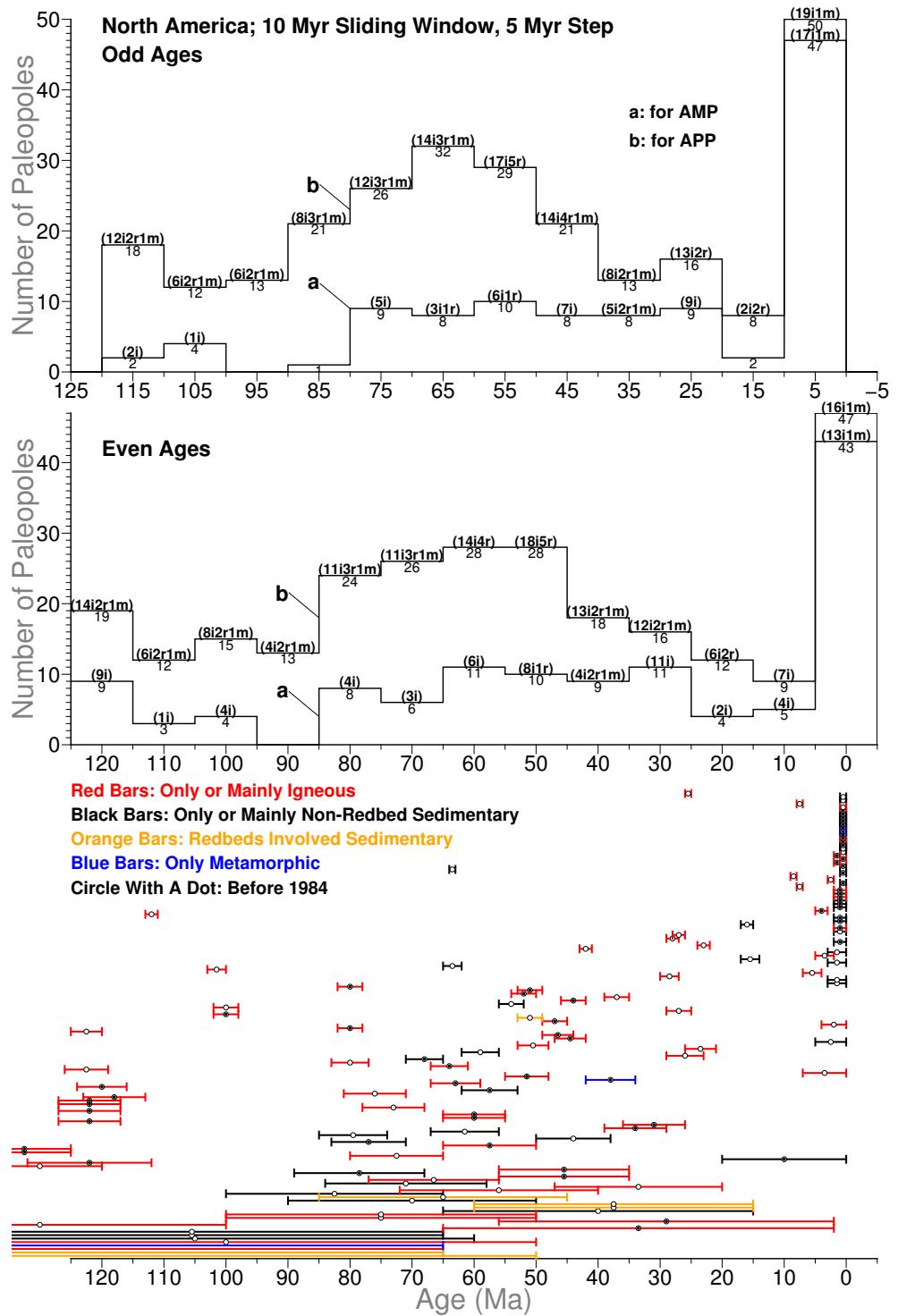
**Filtering.** Poles with characteristics thought to correspond to poor data quality, or lacking characteristics thought to correspond to good data quality, were discarded (or in some cases, corrected).

**Weighting.** Calculation of a weighted Fisher mean [15] of the remaining poles within each window, using weighting functions intended to increase the influence of higher quality poles relative to lower quality ones.

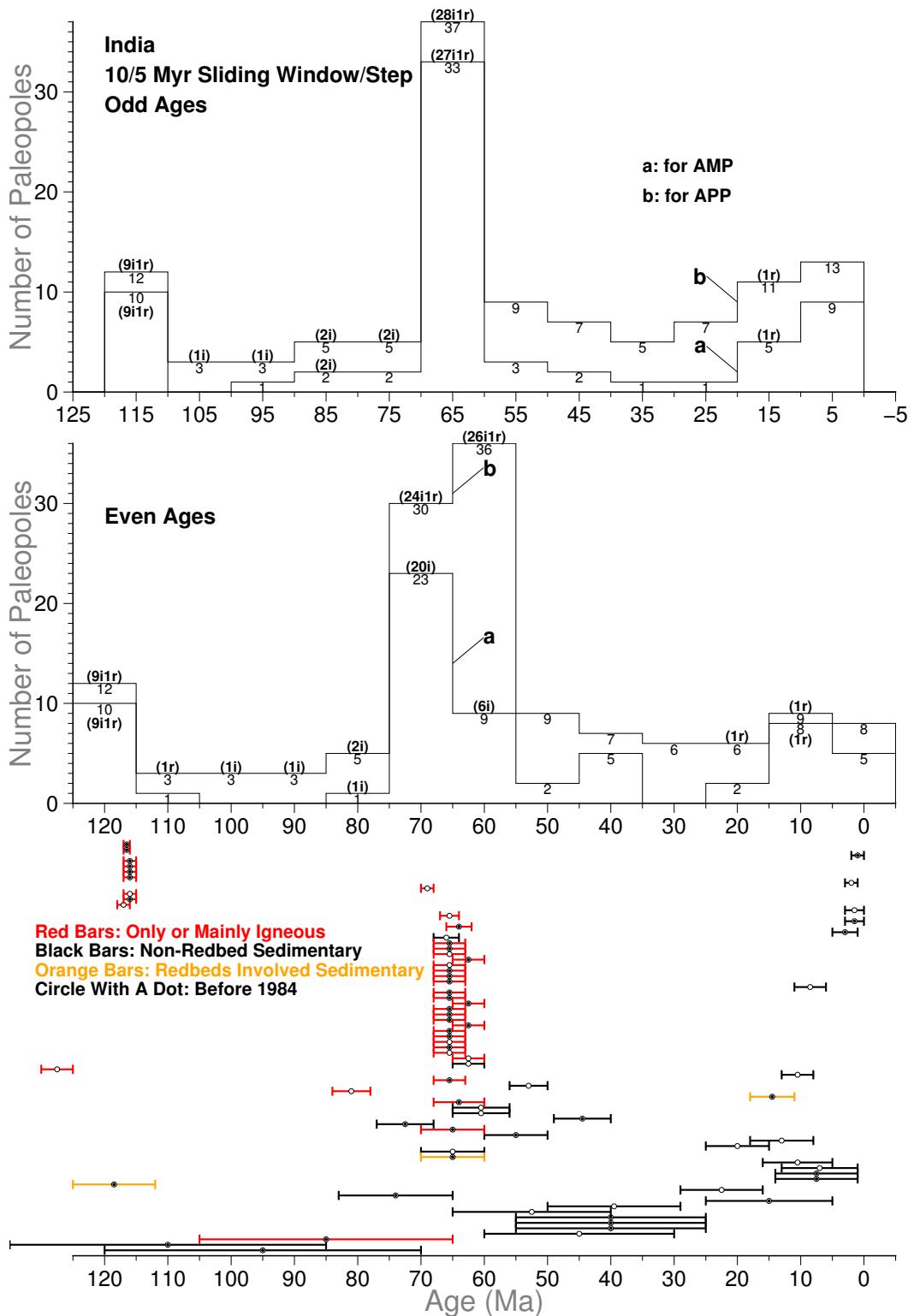
28 different picking and filtering algorithms were tested (Table 3.2), in combination with 6 different weighting algorithms (Table 3.3), for the three plates. The effects of changing the time step length and width of the moving window, and the reference path, were also examined.

#### 3.2.3.1 Picking/Binning

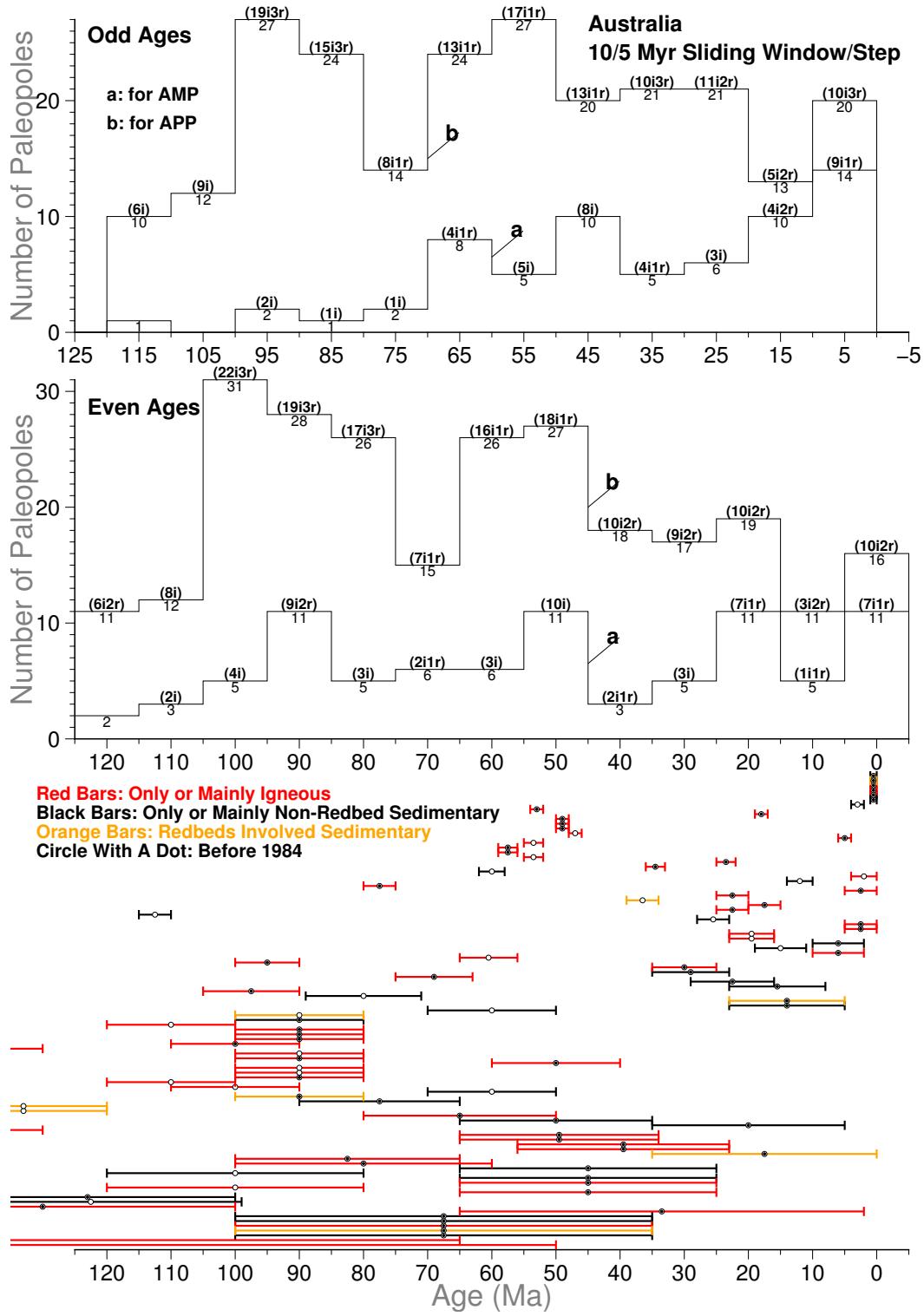
In studies where the moving window method is used to calculate an APWP [57, 56], a paleopole is generally considered to fall in the current window only if the mid-point of its age limits fall within that window. If paleopole has a large age uncertainty compared to the size of the moving window, it will not be included in the moving windows close to the beginning and end of the age range, which are arguably more



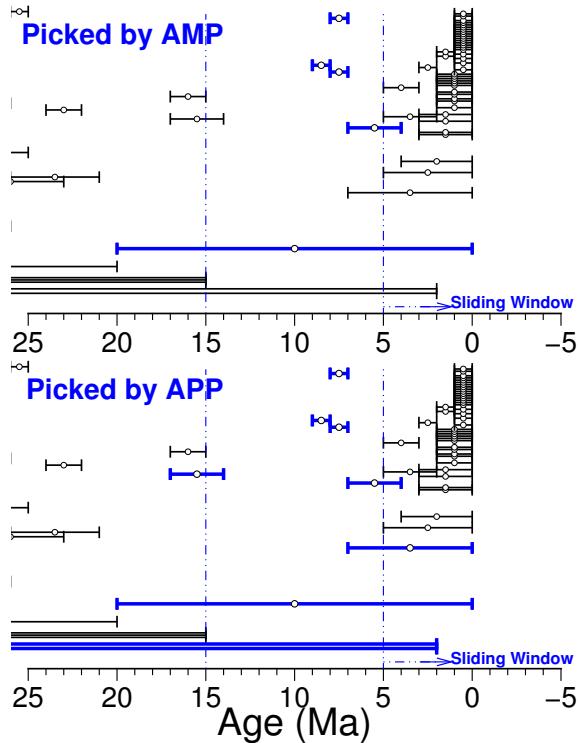
**Figure 3.1:** Temporal distribution of 120–0 Ma NAC (101) paleopoles in 10 Myr window length and 5 Myr step length. For distribution a, each bin only counts in the midpoints (circles) of pole error bars (not including those right at bin edges); For distribution b, as long as the bar intersects with 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 3.2:** 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. 3.1 for more information.



**Figure 3.3:** 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. 3.1 for more information.



**Figure 3.4:** 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. 3.6) of only 5 paleopoles, while the APP method calculates the mean pole (dark circle in Fig. 3.6) of 9 paleopoles.

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 3.2, Fig. 3.9 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 3.2, Fig. 3.9 and subsequent figures). The APP algorithm will pick more paleopoles in each moving window than the AMP algorithm (Figs. 3.1-3.3; Fig. 3.4).

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.

### 3.2.3.2 Filtering

14 different filters or corrections (Table 3.2) 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 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.

**Table 3.2:** List of all Picking/Binning algorithms developed here.

No.	Picking Algorithm	Notes:
0	AMP: Age Mean Picking, see Section “APWP Generation”	
1	APP: Age Position Picking	
2	AMP (“ $\alpha_{95}$ /Age range” no more than “15/20”)	
3	APP (“ $\alpha_{95}$ /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)	
	SS05, [45]	

**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  $\tan I_o = f \tan I_f$  [27], 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=完全ly flattened). Here  $f = 0.6$  is used in all cases, following [59], 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 [12, 13].

**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 red-beds [52, 28, 50, 5, 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 paleopoles published after 1983 (method 16/17; YEAR ≤ 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 [45] removes paleopoles with high spatial ( $\alpha_{95} > 15^\circ$ ) 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.

### 3.2.3.3 Weighting

Following filtering, weights were assigned to each of the remaining paleopoles using one of the following six algorithms (Table 3.3), 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 [51, 61, 3], 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

**Table 3.3:** 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:

age\_range=HIMAGAGE-LOMAGAGE

age\_midpoint=(HIMAGAGE+LOMAGAGE)\*0.5

if age\_midpoint<2.58 (Ma; start of the Quaternary, according to GSA Geologic Time Scale),

$$w = \begin{cases} 1, & \text{if age\_range} \leq 1.29 \text{ (from } \frac{2.58-0}{2} \text{)} \\ \frac{1.29}{age\_range}, & \text{if age\_range} > 1.29; \end{cases}$$

if  $2.58 \leq \text{age\_midpoint} < 23.03$  (Ma; Neogene),

$$w = \begin{cases} 1, & \text{if age\_range} \leq 10.225 \text{ (from } \frac{23.03-2.58}{2} \text{)} \\ \frac{10.225}{age\_range}, & \text{if age\_range} > 10.225; \end{cases}$$

if  $23.03 \leq \text{age\_midpoint} < 201.3$  (Ma; Paleogene,Cretaceous,Jurassic),

$$w = \begin{cases} 1, & \text{if age\_range} \leq 15 \\ \frac{15}{age\_range}, & \text{if age\_range} > 15. \end{cases}$$

3 Lower  $\alpha_{95}$ , 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

$$\alpha_{95} = \begin{cases} ED95 \\ DP, \text{ if ED95 is missing} \\ \frac{140}{\sqrt{KD*N}}, \text{ if ED95 \& DP are missing} \\ \frac{140}{\sqrt{K\_NORM*N}}, \text{ if ED95, DP \& KD are missing} \\ \frac{140}{\sqrt{K\_NORM*B}}, \text{ if ED95, DP, KD \& N are missing} \\ \frac{140}{\sqrt{1.7*B}}, \text{ if ED95, DP, KD \& K\_NORM are missing,} \\ \quad \text{using the lowest KD in GPMDB, about 1.7} \end{cases}$$

finally  $w=0$  if this  $\alpha_{95}$  completely overlaps with another smaller  $\alpha_{95}$  whose paleopole is exactly derived from the same place and same rock.

4 Age error Position to bin (more overlap, greater weight):

North America, India and Australia, there is no such dataset), weight=0.2 (No. 1 in Table 3.3).

**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 of the paleopole's age range falls within is assigned) [61, 51] 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 geologic 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 3.3).

**Weighting by spatial uncertainty (Weighting 3).** Paleopoles with a smaller  $\alpha_{95}$  confidence ellipse are given a higher weighting than those with a larger  $\alpha_{95}$ , using a Gaussian/Normal distribution centered on 0 with standard deviation of 10. However, not all paleopoles'  $\alpha_{95}$  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  $\frac{140}{KD*N}$ , 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.

#### **3.2.3.4 Final Scaling**

The weight values obtained from the six different weight functions (Table 3.3) introduced above are then integrated into Fisher mean function [15] 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 [51]). So the mean pole location, its spatial uncertainty  $\alpha 95$ , and precision parameter  $k$  are all weighted along with  $R$ .

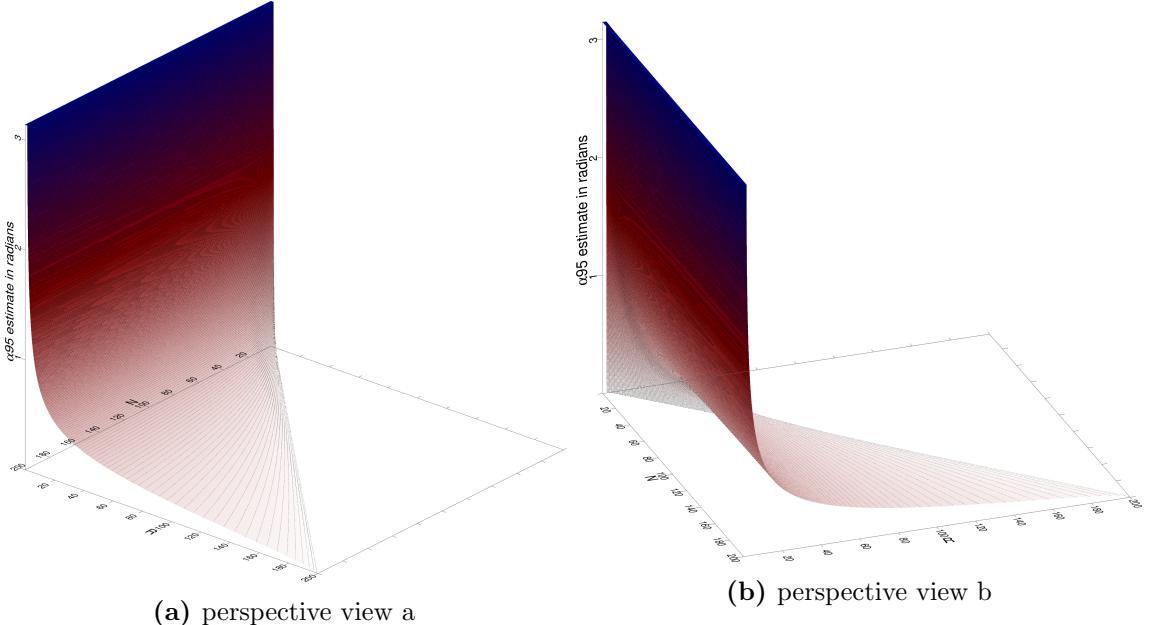
Traditionally, weights are directly integrated by being multiplied with the variable we would like to do weighting to. For example, here weights can be directly multiplied with the Cartesian x, y and z components of each paleopole. However, this direct multiplying causes the decreasing of  $R$ , which further sensitively and extremely increases  $\alpha 95$  (Fig. 3.5), 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

$$ScaledWeight = \frac{Weight_i * N}{\sum_i^N Weight_i},$$

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 3.2) and weighting (Table 3.3) 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.



**Figure 3.5:** 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)  $N$ .

### 3.2.4 Reference Paths

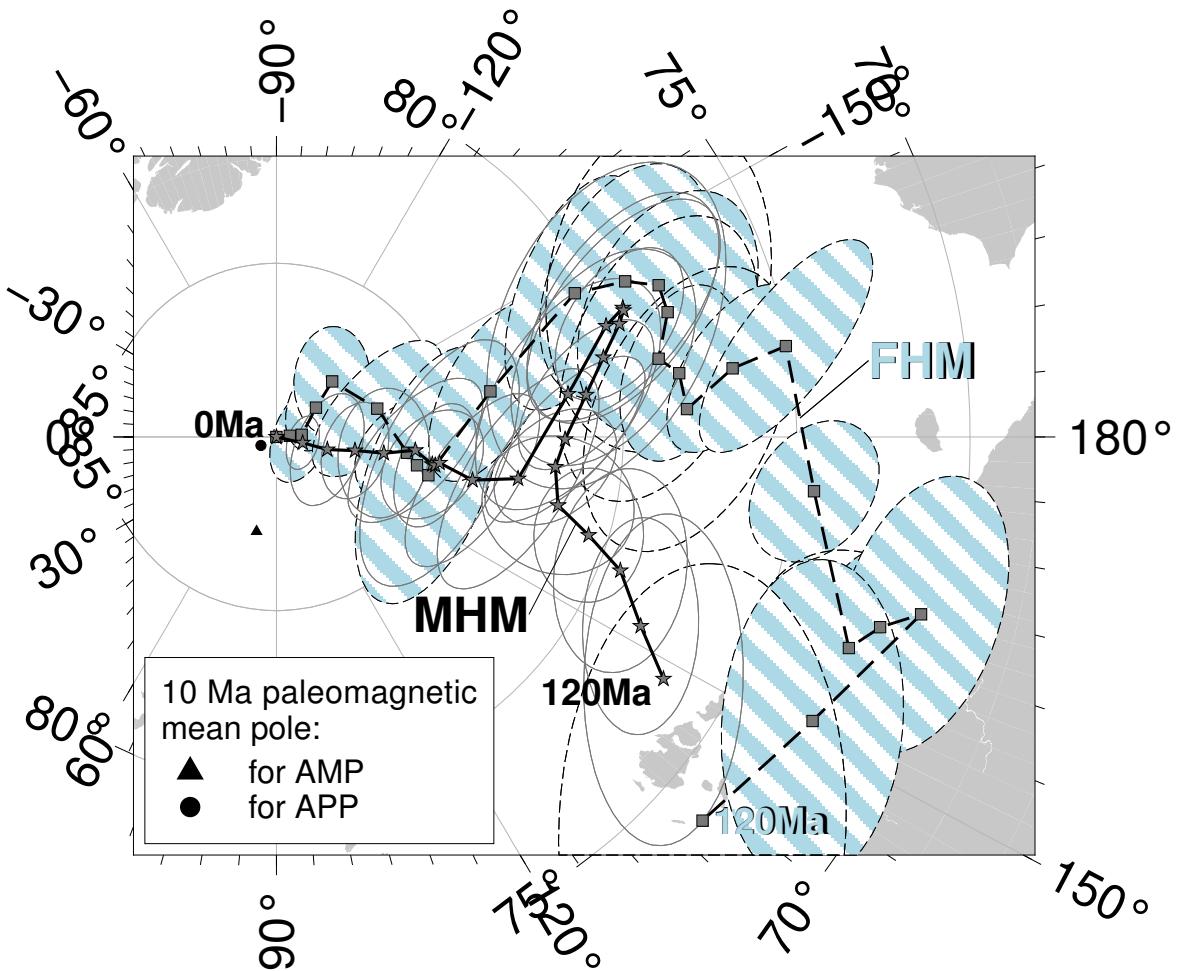
A prediction of the expected APWP for any plate can be generated using a plate kinematic model (e.g. the last c. 180–200 Myr plate motions reconstructed from spreading ridges in ocean basins) that is tied in to an absolute reference frame. Here, we use the rotations of [38], which describe motion of 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 [38]. North America is linked to this reference frame across the Mid-Atlantic ridge, using North America-Nubia rotations from [8] to C-Sequence chron C1no (0.78 Ma), from [46] to chron C2An.2n (2.7 Ma), from [34] to chron C5n.1ny (9.74 Ma), from [16] to chron C5n.2o (10.949 Ma), from [34] to chron C6ny (19.05 Ma), from [16] to chron C6no (20.131 Ma), from [34] to chron C34ny (83.5 Ma), from C34ny to about 118.1 Ma [46], and to closure at C34no (120.6 Ma) [16]. India is linked via the East African Rift Valley (Somalia-Nubia rotations to chron C1no [0.78 Ma] from [?], to chron C2A.2no [3.22 Ma] from [24], and to closure at C7.2m [25.01 Ma] from [43] 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 [37]); and Australia via the East African Rift Valley, SW Indian Ridge (E Antarctica-Somalia rotations to chron C1no [0.78 Ma] from [?], to chron C2A.2no [3.22 Ma] from [24], to chron C5n.2no [10.95 Ma] from [?], to chron C13ny [33.06 Ma] from [?], to chron C29no

[64.75 Ma] from [?], to chron C34y [83 Ma; Rowley, pers. comm.], to 96 Ma from [?], and to closure at M0 [120.6 Ma] from [36]), and SE Indian Ridge (Australia-East Antarctica rotations to chron C1no [0.78 Ma] from [?], to chron C6no [20.13 Ma] from [?], to chron C8o [26 Ma] from [?], to chron C17n.3no [38.11 Ma] from [?], to C34ny [83.5 Ma] from [?], to the Quiet Zone Boundary [96 Ma] from [64], to closure at 136 Ma from [?]). 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 [?], assuming constant rates.

Neither the hotspot reference frame nor the paleomagnetic reference frame are truly fixed with respect to the solid Earth. In the former case, hotspots are not truly stationary in the mantle [?]; in the latter, true polar wander (TPW) may also lead to differential movements of the solid earth with respect to the spin axis [?]. In reality, it is difficult to untangle these effects. Whilst there is little clear evidence for significant TPW in the past about 120 Myr [?, ?], modeling suggests that the effects of hotspot drift can start to become significant over 80–100 Myr timescales [38]. 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. 3.6), India (Fig. 3.7) and Australia (Fig. 3.8). 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 [38], which incorporate motions of the Indo-Atlantic hotspots relative to the mantle derived from mantle convection modeling.

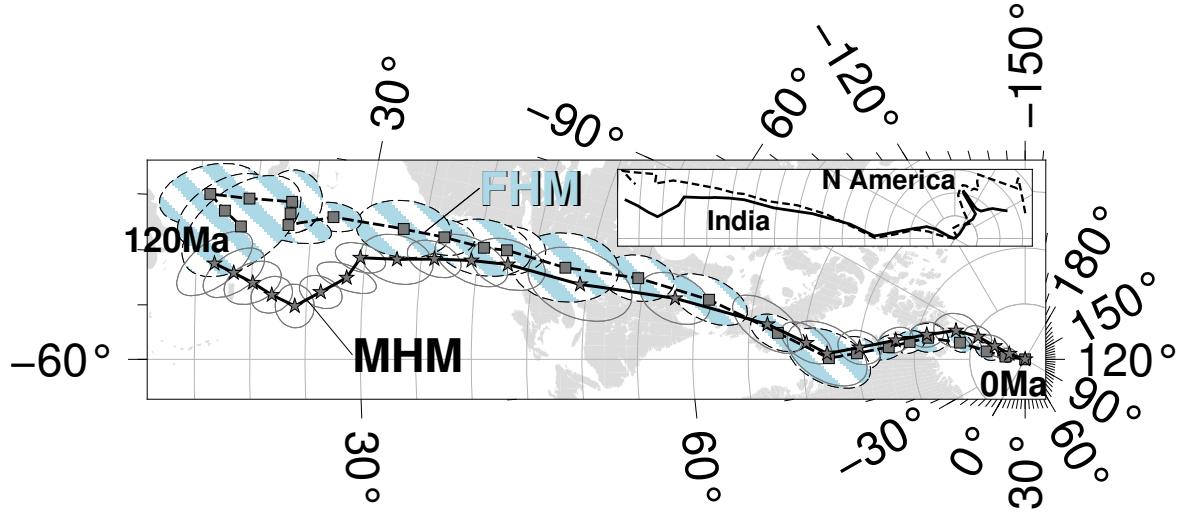
When comparing the synthetic APW paths for the three plates (inset, Fig. 3.8), there are clear differences. The predicted mean north pole for North America at 120 Myr is still at about 75°N (Fig. 3.6), indicating rather slow drift with respect to the spin axis; this is due to a large component of the 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 [?], is reflected by the 120 Ma predicted mean north pole being located at about 20°N (Fig. 3.7). Australia represents an intermediate case,



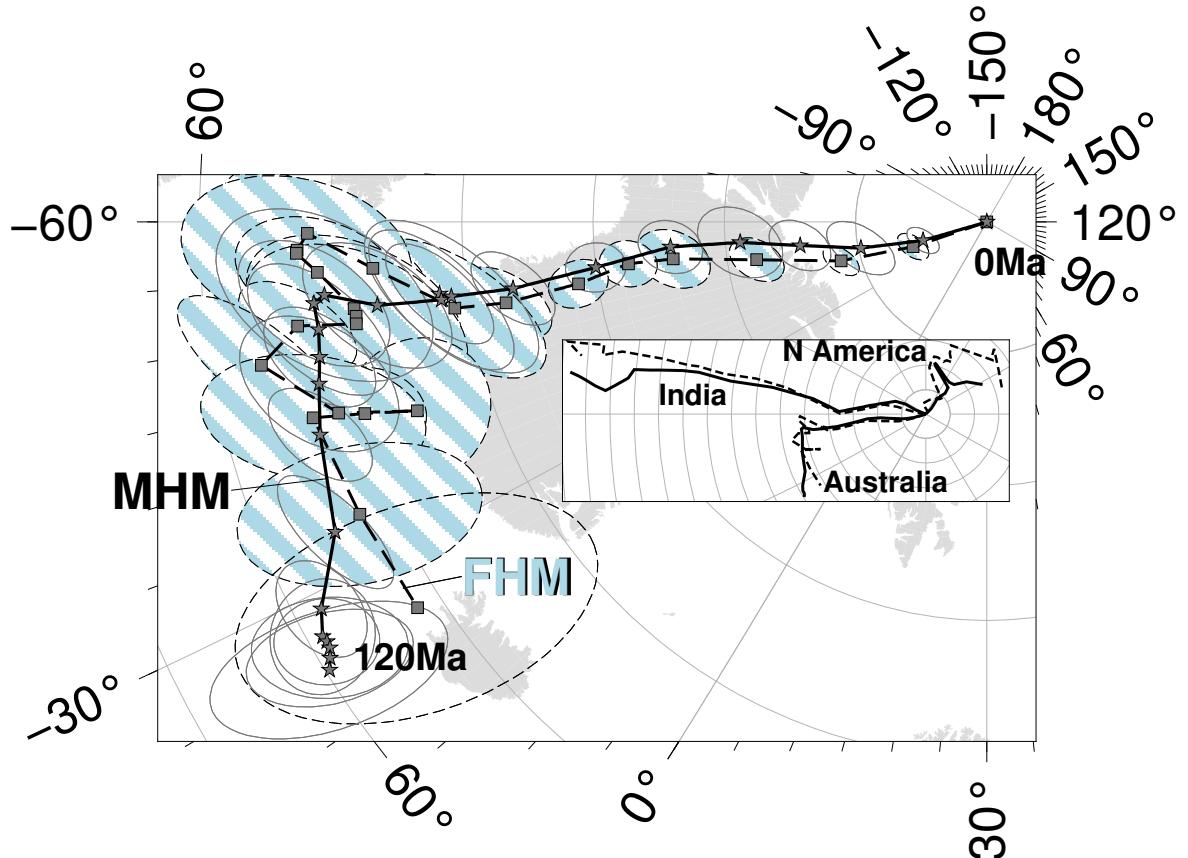
**Figure 3.6:** MHM predicted 120–0 Ma APWP (solid line) for NAC through the North America–Nubia–Mantle plate circuit. The FHM predicted path (dashed line with shaded uncertainties) is also shown for comparison. The age step is 5 Myr. Compared with the 10 Ma paleomagnetic mean pole calculated by the 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. 3.4.

with north westerly plate motion from about 120–60 Myr changing to more rapid northward motion from about 60–55 Ma to the present [64]. 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.



**Figure 3.7:** 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. 3.6.



**Figure 3.8:** 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. 3.7, much slower moving North America shown in Fig. 3.6, and also relatively intermediate moving Australia.

### 3.2.5 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.,  $W_s = W_l = W_a = \frac{1}{3}$ ).

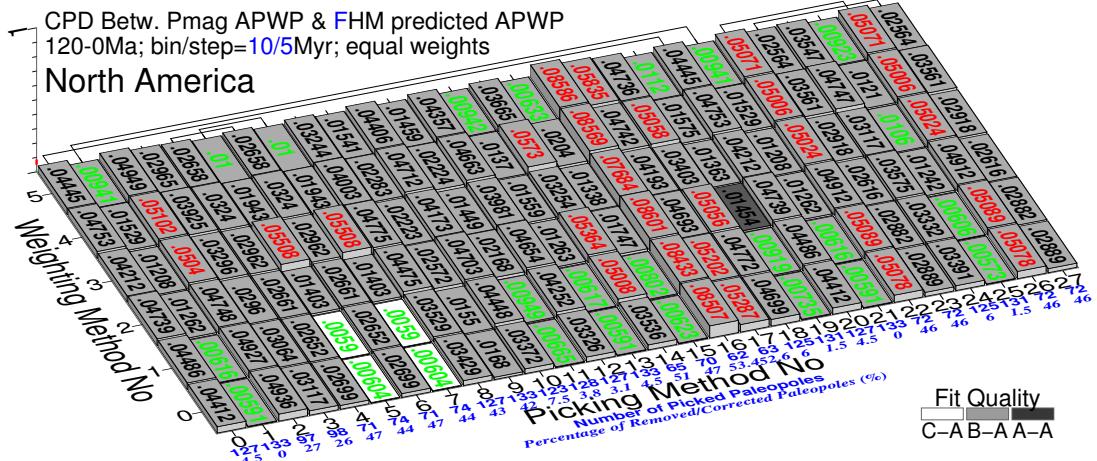
This does not only help find the most similar paleomagnetic APWP (from the best algorithm) to the reference APWP, but also help further test and demonstrate the validness of the similarity measuring tool in practise.

## 3.3 Results

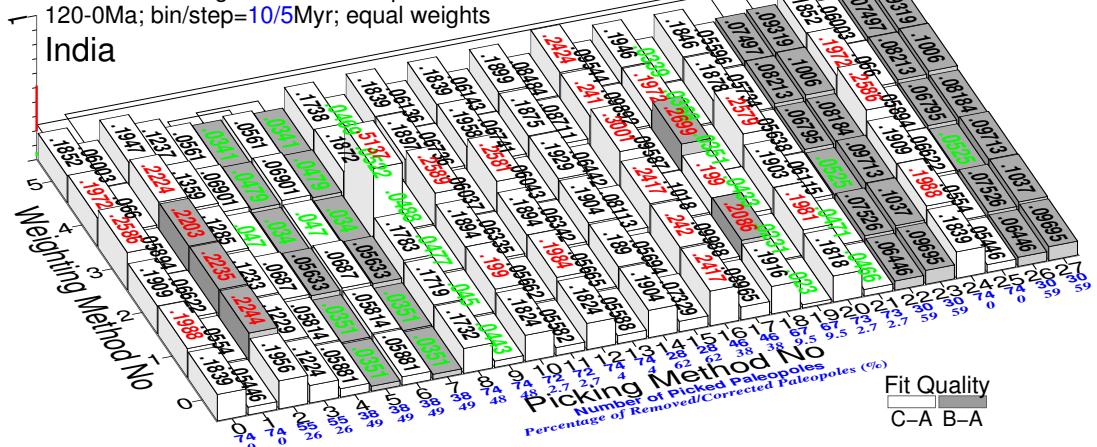
### 3.3.1 Baseline results: 10 Myr window, 5 Myr step, fixed hotspot reference

Fig. 3.9 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. 3.9a), India (Fig. 3.9b) and Australia (Fig. 3.9c). 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 affect the difference score, with  $\mathcal{CPD}$  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. 3.10). From Fig. 3.9, it is clear that:

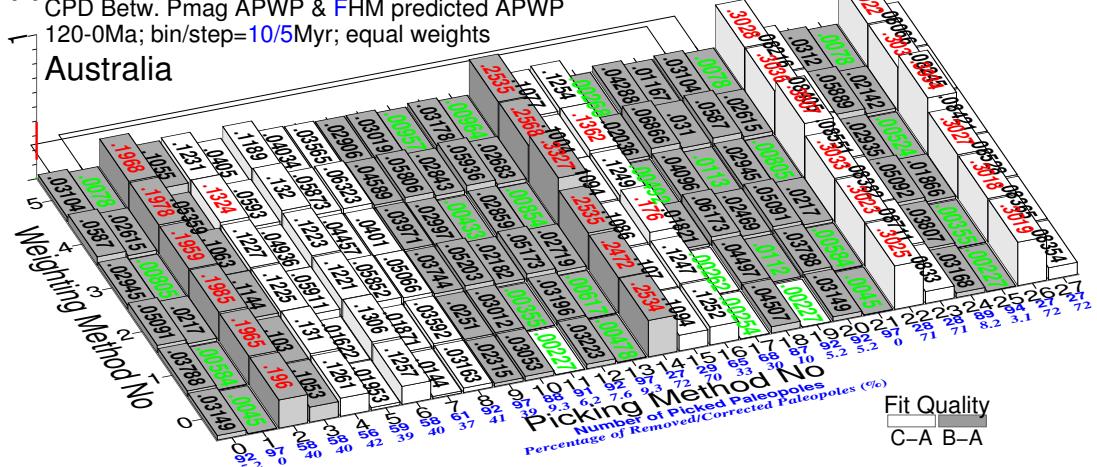
1. There is much more variation in scores along the horizontal axes than the vertical axes (see self-explanatory topography of bands in Fig. 3.9), suggesting that the choice of windowing and filtering method (Table 3.2) has a much greater impact than weighting (Table 3.3).
2. Many of the highest scores (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.
3. The magnitude and range of  $\mathcal{CPD}$  scores for each of the three plates is different, with the North American plate having the lowest magnitudes and range (Fig. 3.9a), and the Indian plate having the highest (Fig. 3.9b). In general, the scores of North America and Australia are relatively closer (Fig. 3.13).



(a) Plate ID 101 with children: minimum 0.00573 (25(0)), maximum 0.08601 (16(2)), mean 0.0CPD Betw. Pmag APWP & FHM predicted APWP  
120-0Ma; bin/step=10/5Myr; equal weights

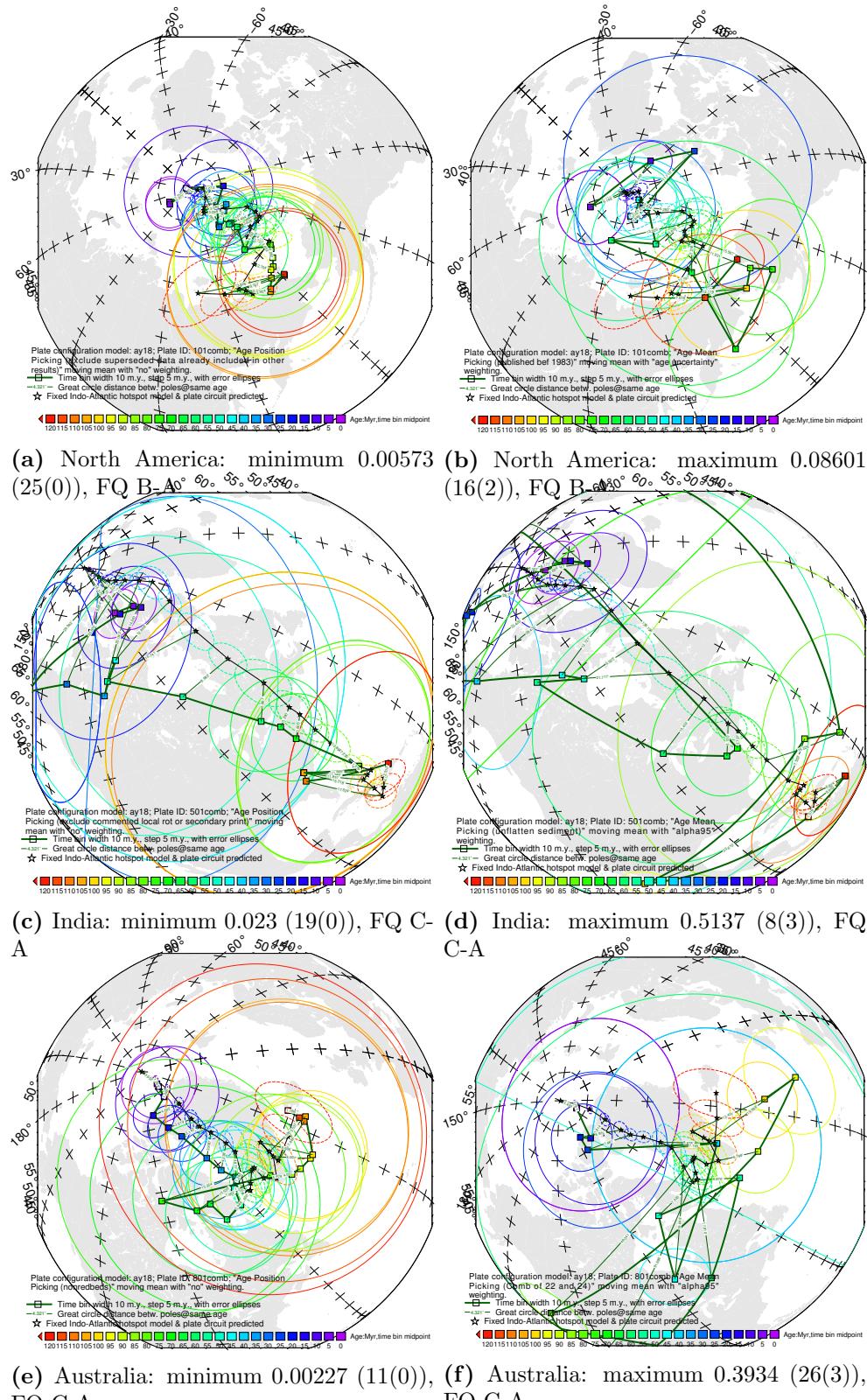


(b) Plate ID 501: minimum 0.023 (19(0)), maximum 0.5137 (8(3)), mean 0.1182, median 0.0835  
CPD Betw. Pmag APWP & FHM predicted APWP  
120-0Ma; bin/step=10/5Myr; equal weights



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

**Figure 3.9:** Equal-Weight composite path difference ( $\text{CPD}$ ) 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 3.10:** Path comparisons with best and worst difference values shown in Fig. 3.9. The parenthetical remarks are Picking No and Weighting No.

4. 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.

### 3.3.2 Effects of windowing method

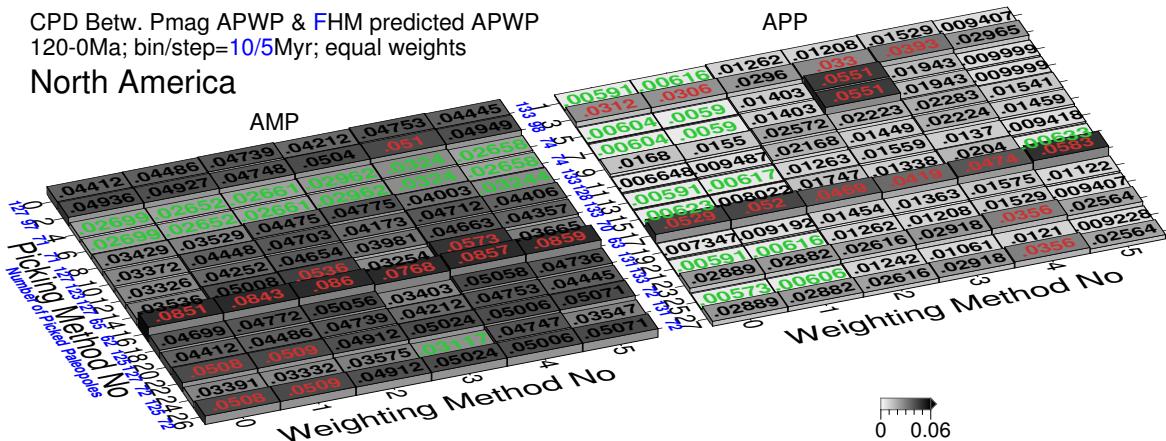
Dividing  $\mathcal{CPD}$  scores according to whether the AMP or APP windowing method was used (Fig. 3.11) confirms that whilst the lowest  $\mathcal{CPD}$  scores for paths generated by the AMP windowing algorithm are close to the lowest scores generated using the APP method, the highest scores are much higher (Fig. 3.12). The mean of the  $\mathcal{CPD}$  scores for APWPs generated using AMP is greater than the maximum APP-derived score for the Indian and Australian plates (Figs. 3.11b, 3.11c, 3.12c, 3.12c), and more than 1 standard deviation greater than the APP mean for North American APWPs (Figs. 3.11a, 3.12a).

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. 3.12, insets). APP-generated paths yield a lower  $\mathcal{CPD}$  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. 3.11a, 3.11b). 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. 3.11c).

On the North American plate, the APP-derived  $\mathcal{CPD}$  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. 3.11a). 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. 3.11c). 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. 3.11b).

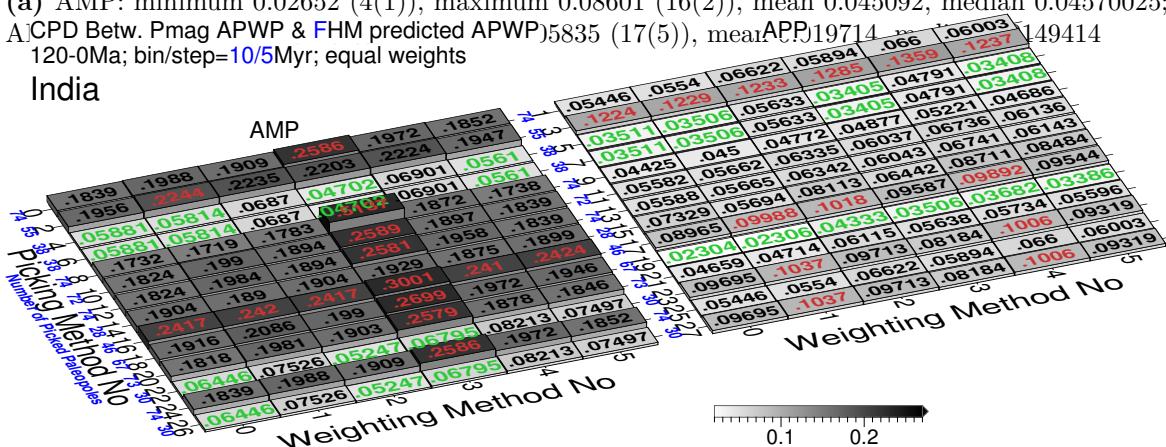
CPD Betw. Pmag APWP & FHM predicted APWP  
120-0Ma; bin/step=10/5Myr; equal weights

### North America



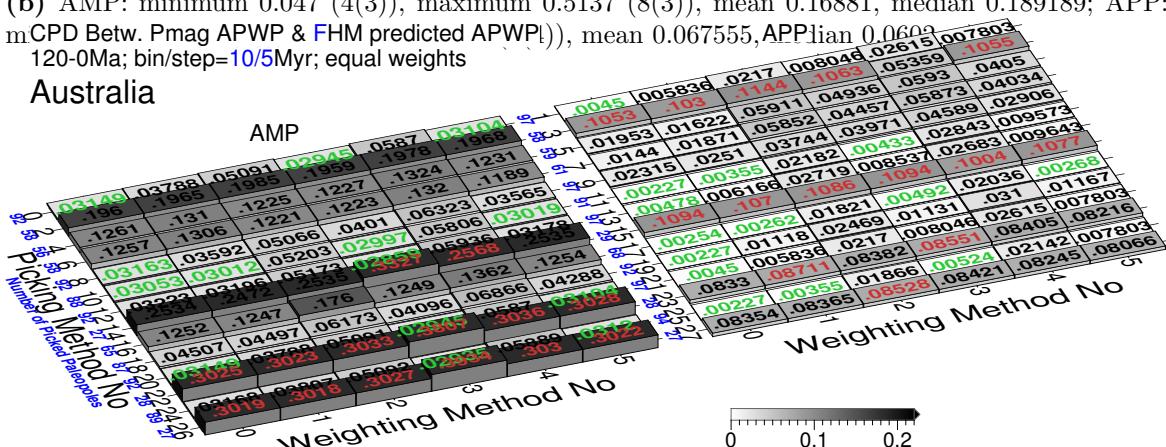
(a) AMP: minimum 0.02652 (4(1)), maximum 0.08601 (16(2)), mean 0.045092, median 0.04570025;  
mCPD Betw. Pmag APWP & FHM predicted APWP)5835 (17(5)), meanAPP19714 149414  
120-0Ma; bin/step=10/5Myr; equal weights

### India



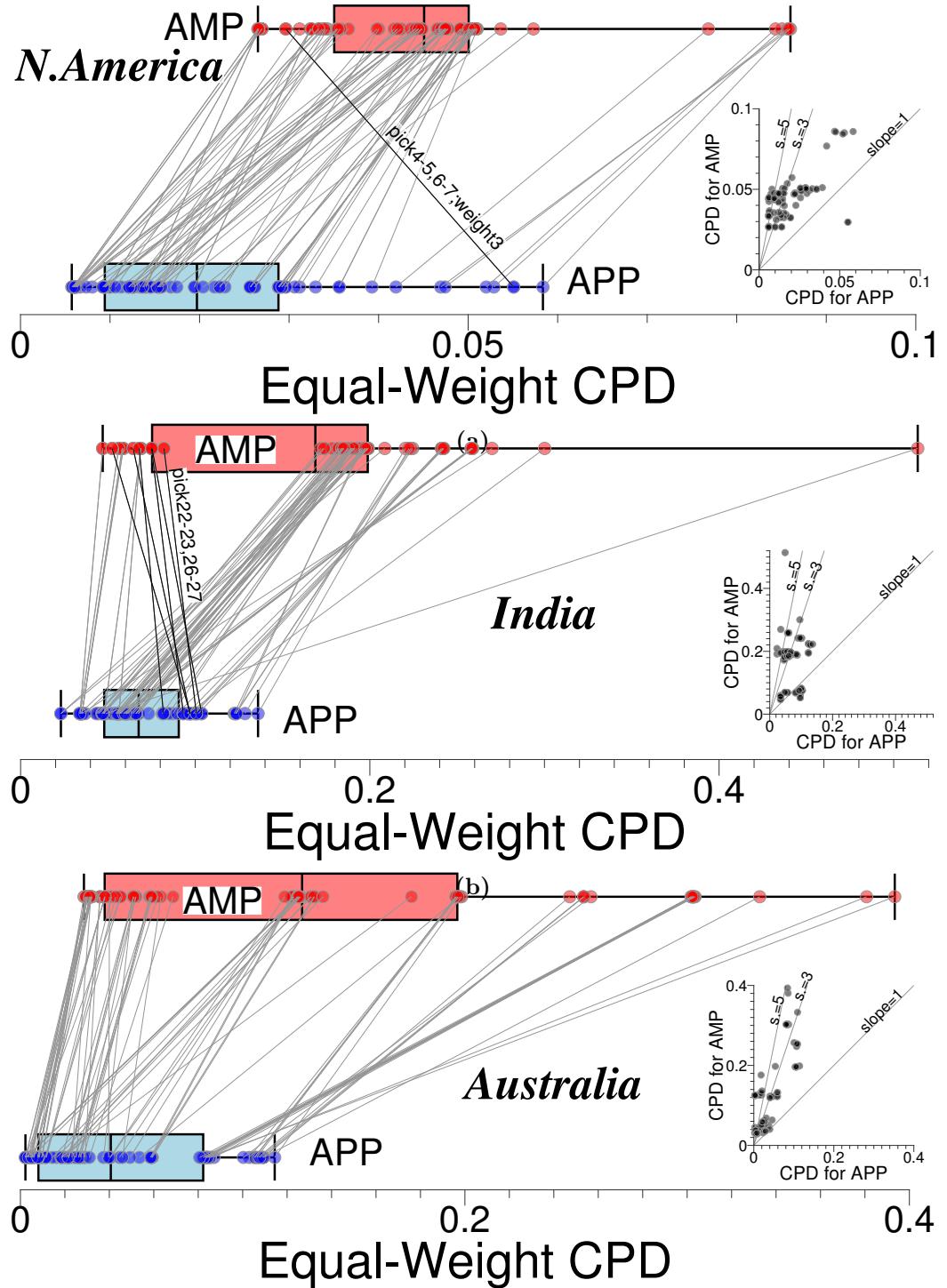
(b) AMP: minimum 0.047 (4(3)), maximum 0.5137 (8(3)), mean 0.16881, median 0.189189; APP:  
mCPD Betw. Pmag APWP & FHM predicted APWP)1), mean 0.067555, APPlian 0.0602  
120-0Ma; bin/step=10/5Myr; equal weights

### Australia



(c) AMP: minimum 0.0285868 (3(0)), maximum 0.3934 (26(3)), mean 0.12675, median 0.0938; APP:  
minimum 0.00227 (11(0)), maximum 0.11445 (3(2)), mean 0.04071, median 0.025625

**Figure 3.11:** Separated results from AMP and APP in Fig. 3.9. 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 3.12:** Box-and-whisker and cross (inset) plots of Fig. 3.11. 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 Effects of filtering and weighting

When  $\mathcal{CPD}$  scores are separated by windowing method (Fig. 3.11), 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  $\mathcal{CPD}$  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.

#### 3.3.3.1 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:

1. 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.
2. 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.
3. 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.
4. 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.

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. 3.9b]), or superseded data (method 24/25) affected 4% or less of the poles on any plate.

### 3.3.3.2 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:

1. For all 3 plates, higher  $\mathcal{CPD}$  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.
2. 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.
3. For North America and Australia, correction for inclination flattening generates  $\mathcal{CPD}$  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  $\mathcal{CPD}$  scores compared to no filtering for both AMP- and APP-derived paths.
4. 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  $\mathcal{CPD}$  scores comparable to scores with no filtering (relatively much lower). In contrast, for Australia methods 14/15 produces some of the highest  $\mathcal{CPD}$  scores, and methods 16/17 have little effect. The number of older studies (65/68; Fig. 3.9c) 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:

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

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

### 3.3.3.3 Weighting performance

Compared to the variations resulting from different windowing methods and filters, Figs. 3.11, 3.14, 3.15 and 3.16 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.

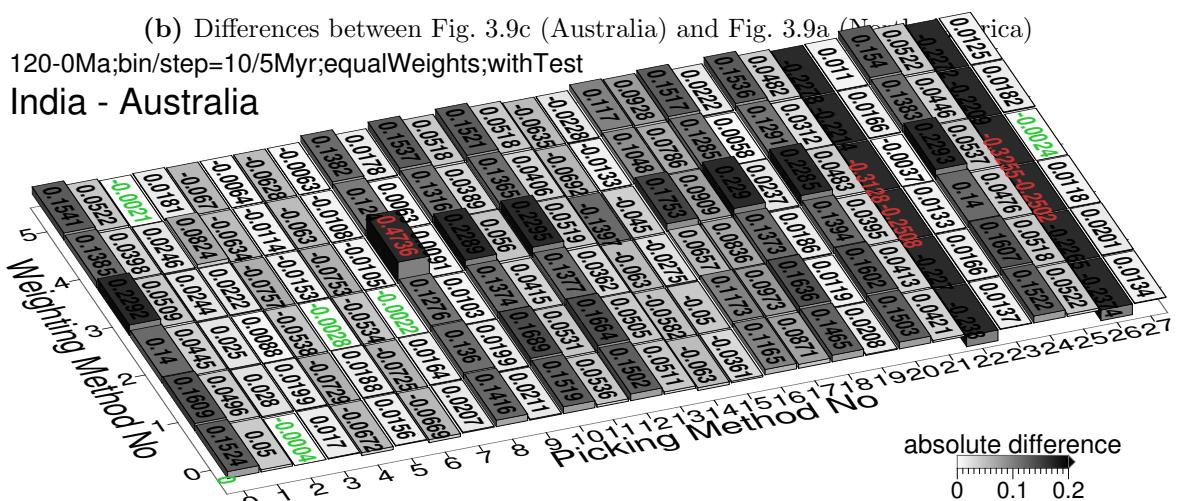
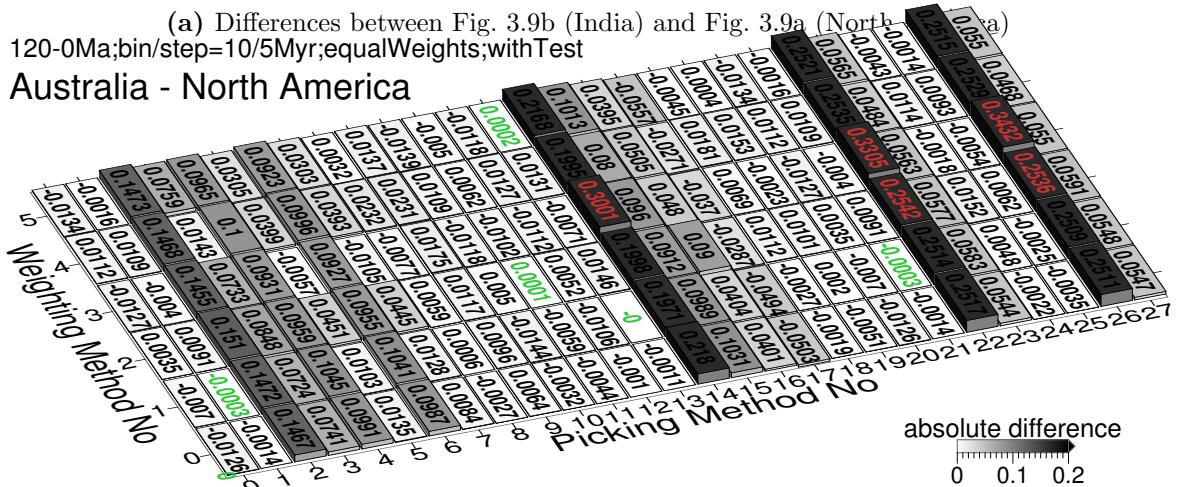
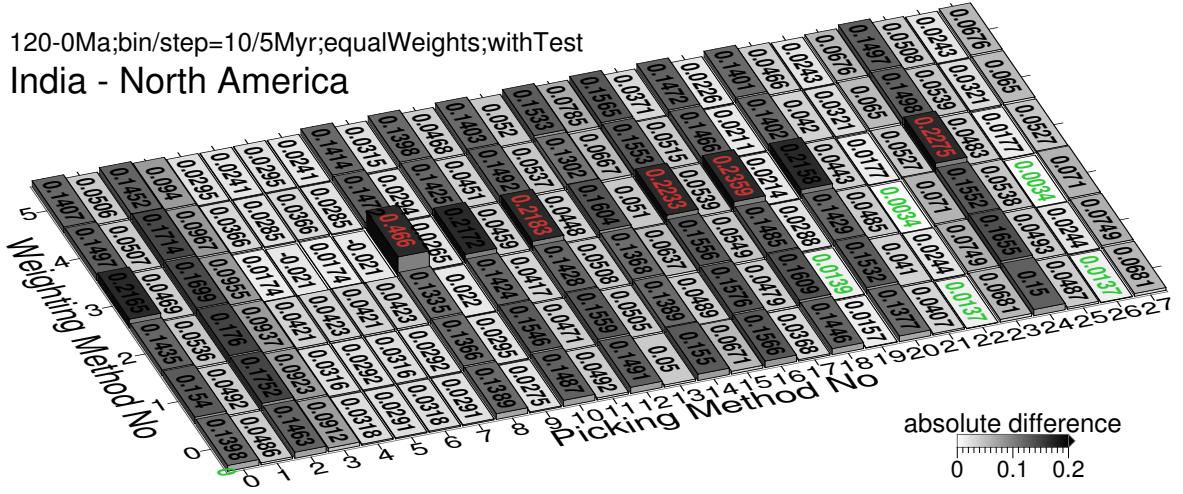
1. For APP-derived paths from the North American and Australian plates, weighting method 0 (no weighting) and 1 (weighting by sample and site number) appear to usually produce slightly better  $\mathcal{CPD}$  scores than other weighting methods.
2. Weighting method 3 (weighting by spatial uncertainty) seems most likely to generate much higher  $\mathcal{CPD}$  scores, particular for AMP-derived paths, and particular for the Indian plate.
3. Weighting methods 0, 1 and 5 are generally producing better similarity than 2, 3 and 4.

### 3.3.4 Effects of window size

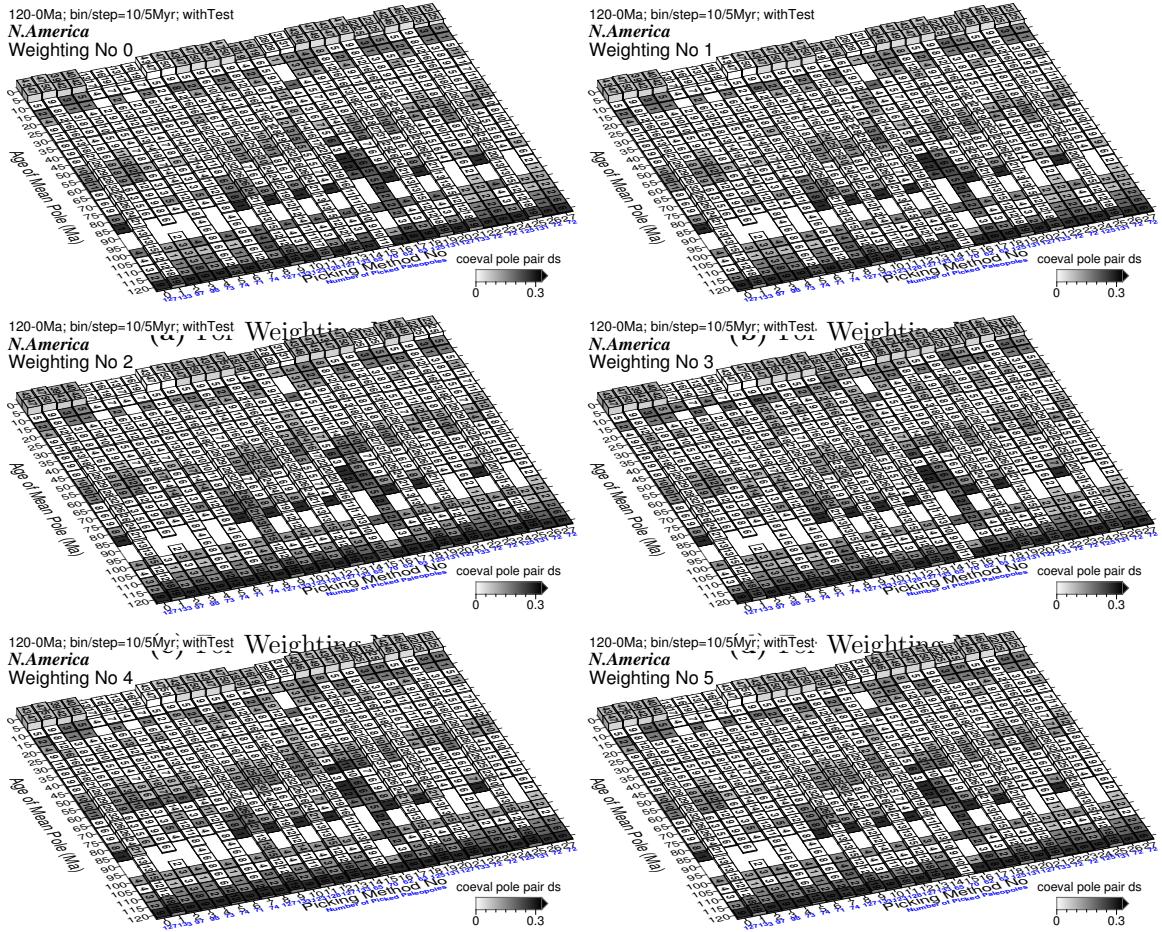
Fig. 3.17 shows the  $\mathcal{CPD}$  scores for the APWPs generated with all 28 picking and 6 weighting methods, compared to the FHM reference paths, with the picking time window width increased from 10 to 20 Myr, and the window step increased from 5 to 10 Myr.

#### 3.3.4.1 Overall change

Along with the increasing sizes of moving window and step, mean, median and range values of  $\mathcal{CPD}$  scores shrink, except for North America the mean value of  $\mathcal{CPD}$  scores slightly increase (Fig. 3.9 and Fig. 3.17). For North America, most scores actually increase (i.e. 10-5 is better; Fig. 3.19a) 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

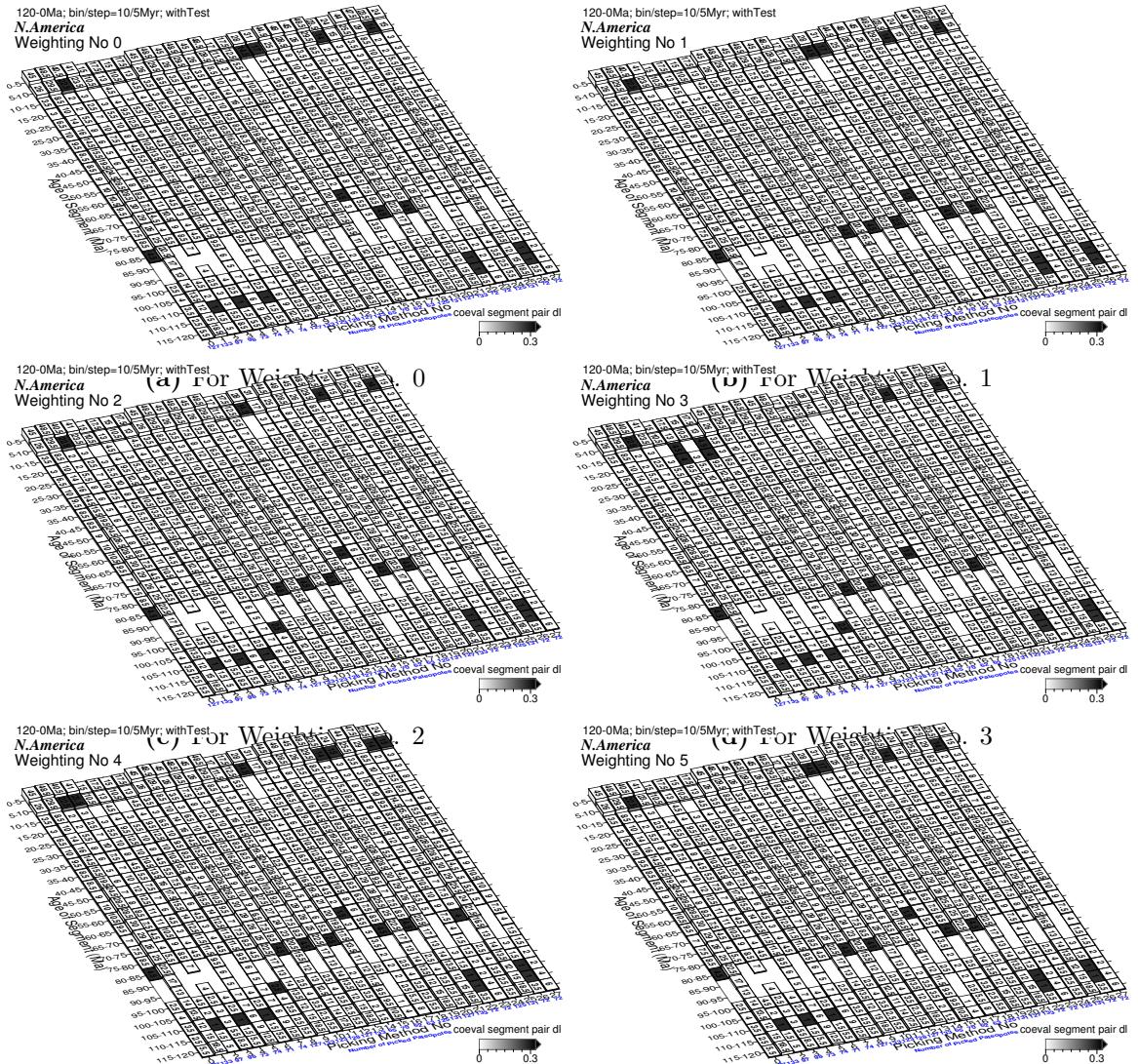


**Figure 3.13:** (a) Differences between grids in Fig. 3.9b (India) and Fig. 3.9a (Australia). The absolute difference values less than 1.96-standard-deviation interval of the whole 168 values are labeled in green, more than 1.96-standard-deviation labeled in red.

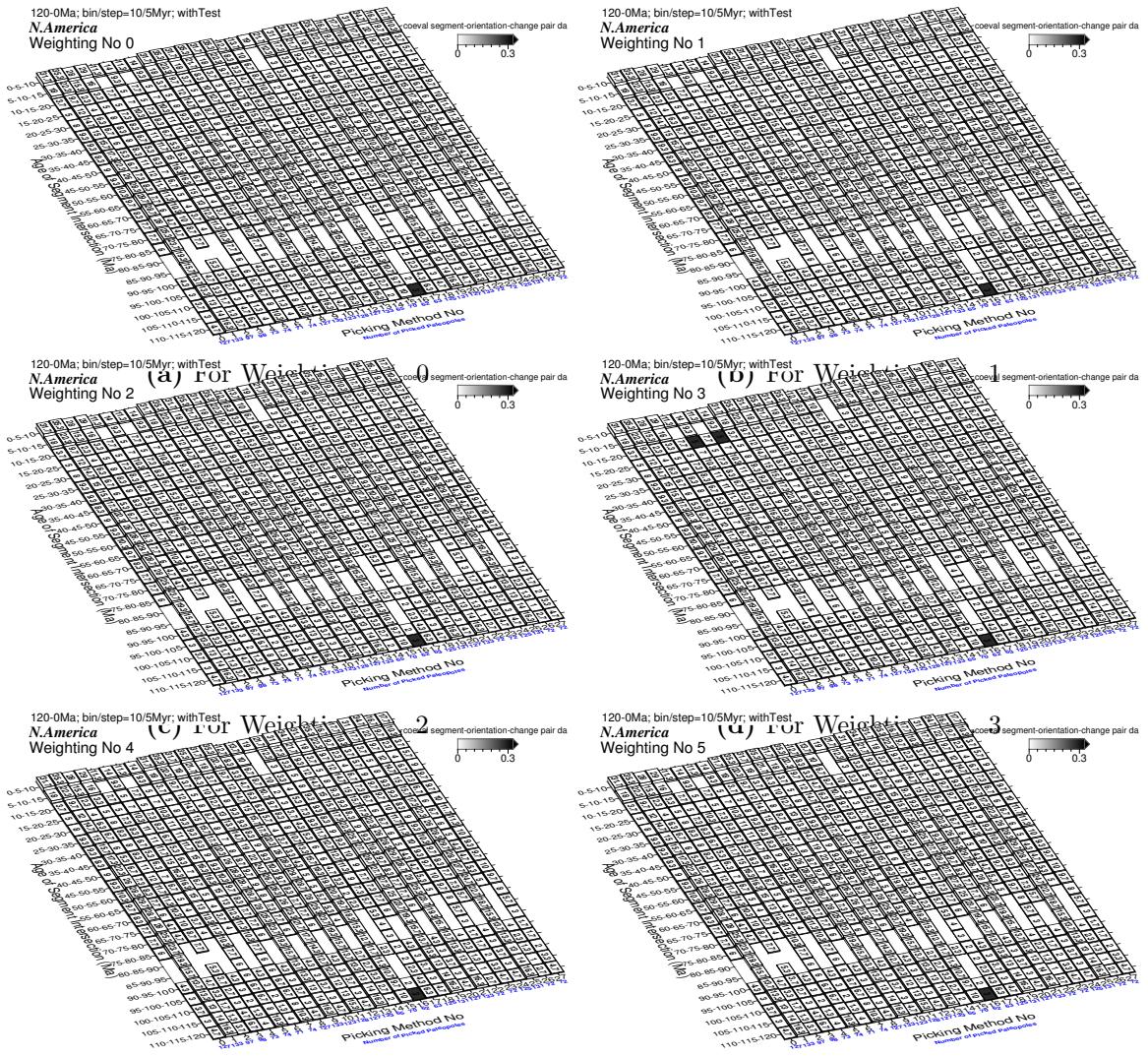


**Figure 3.14(e)** For Weighting No. 4

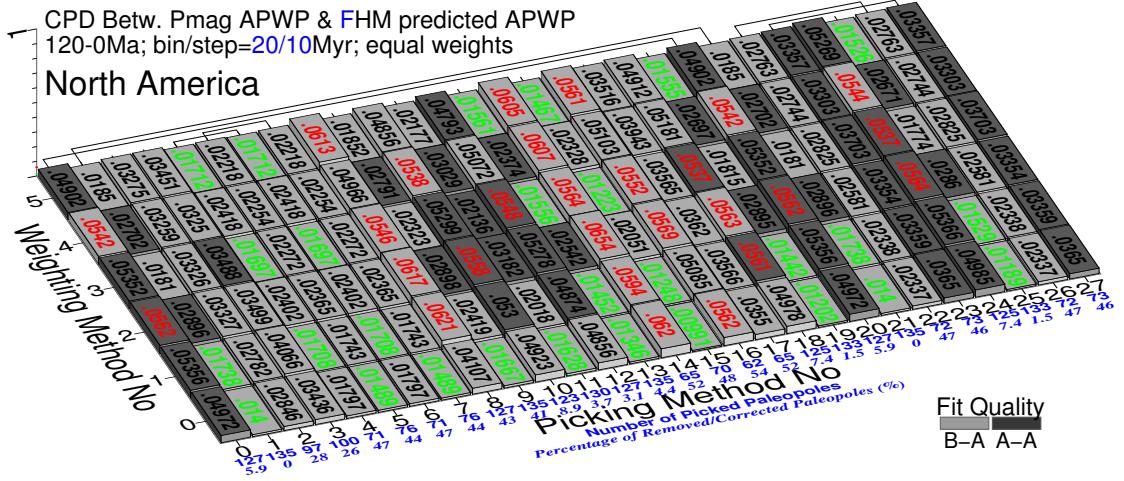
Tested spatial difference ( $d_s$ ) values (color shaded) between North American paleomagnetic APWPs and its predicted APWP from the FHM and related plate circuits. The paths are in 10 Myr bin and 5 Myr step. The number labels on the grids (including grid heights) are the numbers of site mean poles that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



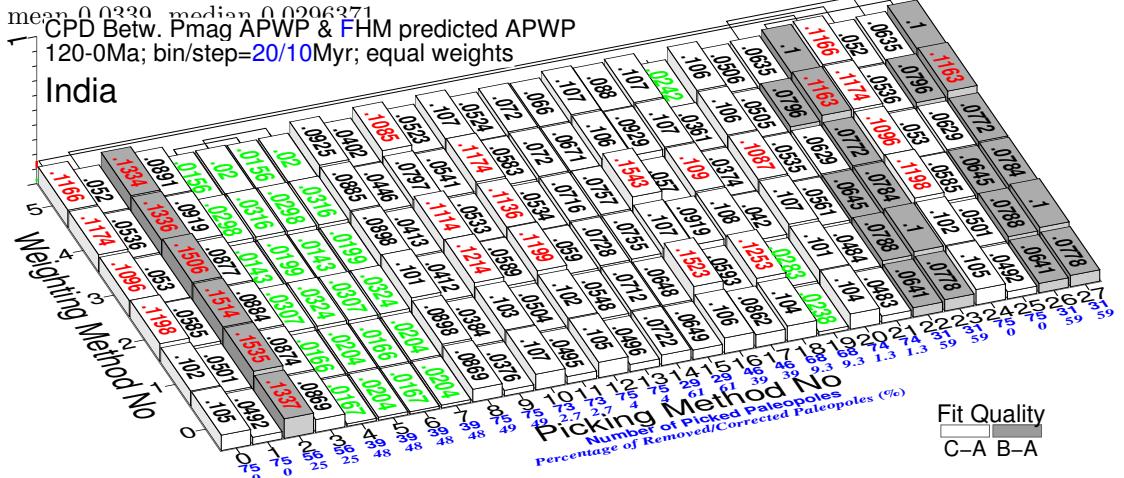
**Figure 3.15(e) For Weighting No 4 (f) For Weighting No 5**  
**Figure 3.15!** Tested length difference ( $d_l$ ) values (color shaded) between North American paleomagnetic APWPs and its predicted APWP from FHM and related plate circuits. The paths are in 10 Myr bin and 5 Myr step. The labeled numbers on the grids are the averaged numbers of site mean poles that are contributing to each segment's two mean path poles.



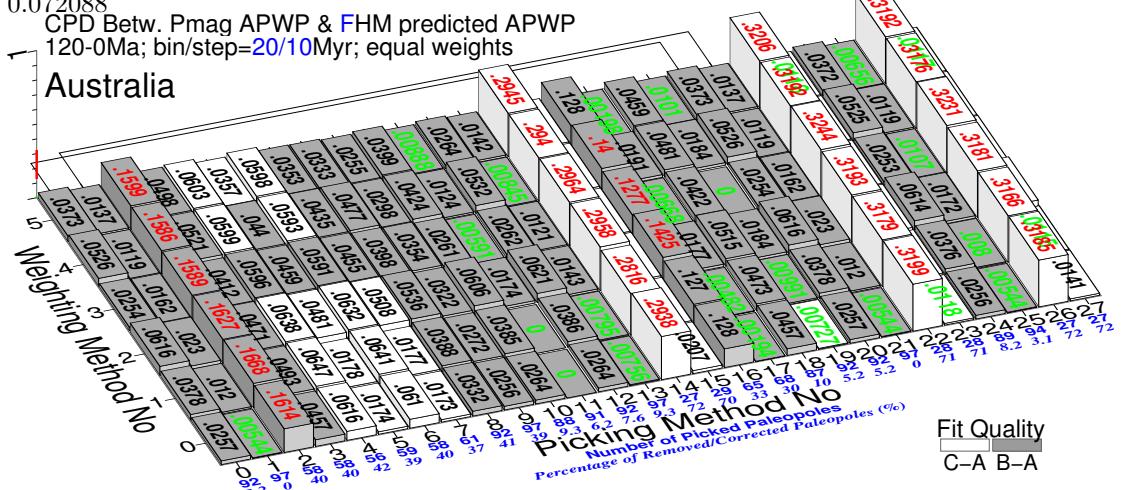
**Figure 3.16(e) For Weighting No 4 (f) For Weighting No 5** Tested angular difference ( $d_a$ ) values (color shaded) between North American paleomagnetic APWPs and its predicted APWP from FHM and related plate circuits. The paths are in 10 Myr bin and 5 Myr step. The labeled numbers on the grids are the averaged numbers of site mean poles that are contributing to each segment-orientation-change's three mean path poles.



(a) North America (101 with children): minimum 0.00991 (15(0)), maximum 0.0654 (14(2)), mean 0.0330, median 0.0206371

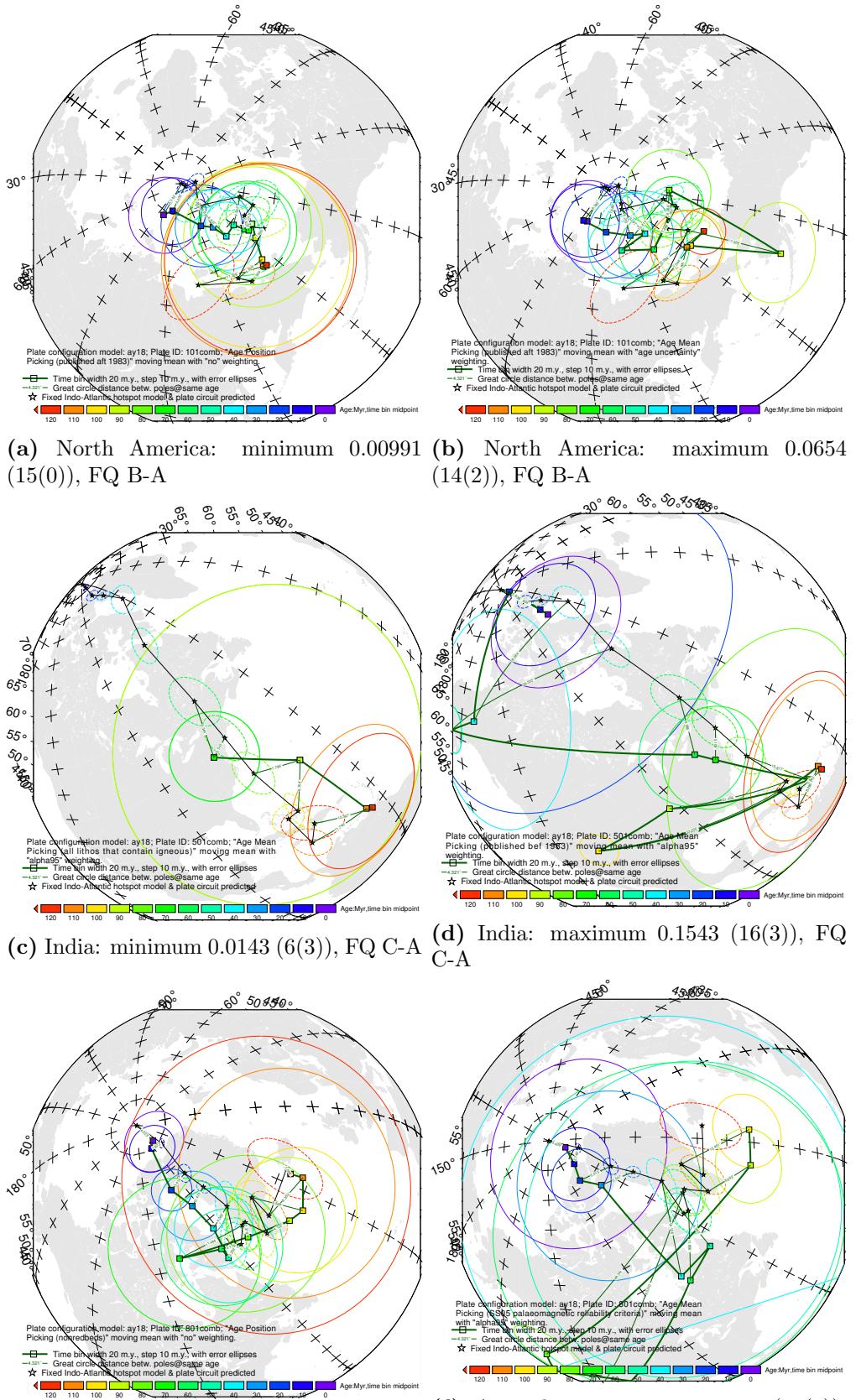


(b) India (501): minimum 0.0142822 (6(3)), maximum 0.154278 (16(3)), mean 0.07384, median 0.072088



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

**Figure 3.17:** As Fig. 3.9, 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. 3.9.



**Figure 3.18:** Path comparisons with best and worst difference values shown in Fig. 3.17. The parenthetical remarks are Picking No and Weighting No.

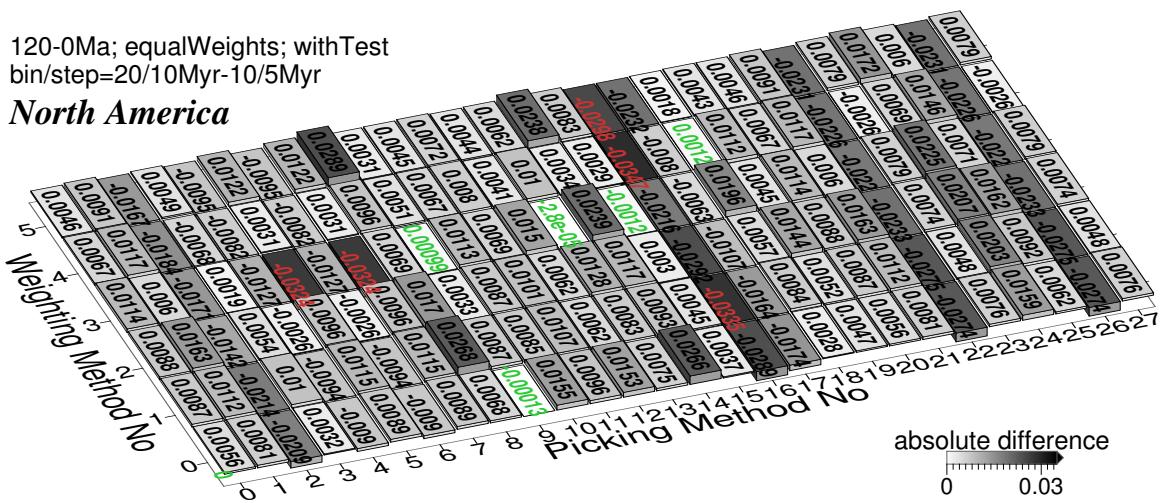
is being diluted). For Australia, changes are largely negative, which means 20-10 is better. Also negative change is large for methods 3, 4, 6, 15, 23, 27. For method 14(0-2,4,5), 10-5 is much better, whereas for method 15, 20-10 is much better.

### 3.3.4.2 Relative performance of methods

1. Method 19 is still among the best, and 16 still one of the worst, for all 3 plates.
2. 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. 3.19).
3. 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. 3.19).
4. 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.
5. 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%).
6. Both scores and AMP/APP difference are low for methods 4/5, 6/7 applied onto North America and India.
7. Both scores and AMP/APP difference are high for methods 16/17.
8. Scores are high but AMP/APP difference are low for methods 2/3, 22/23 on North America and India.
9. Scores are low but AMP/APP difference are high for methods 0/1 on North America and Australia.
10. Comparable AMP/APP (20-10 vs 10-5) appears for 8/9 on Australian plate.
11. Method 1 is still a good performer.
12. Methods 4/5, 6/7 are still ‘good’ aggressive filters.

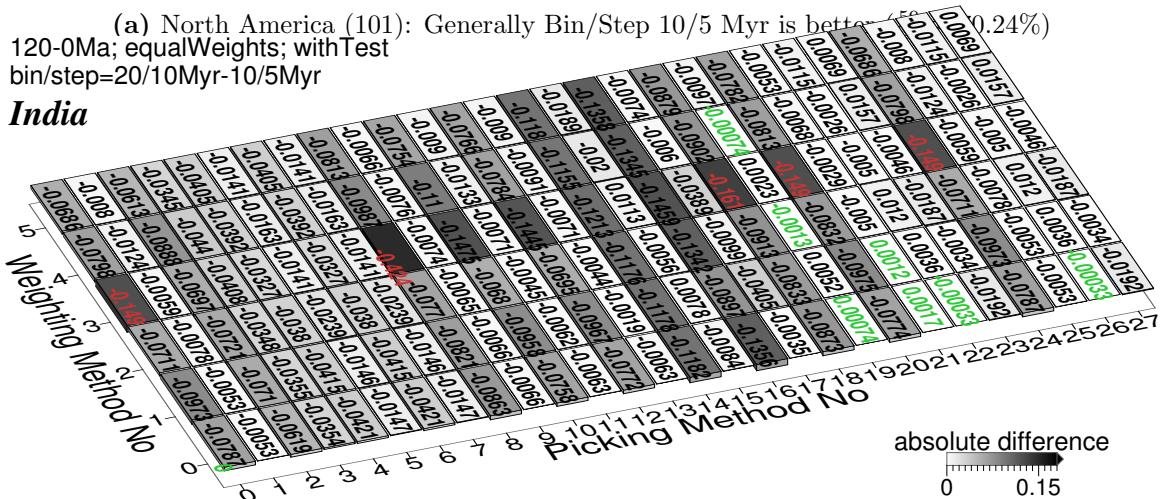
120-0Ma; equalWeights; withTest  
bin/step=20/10Myr-10/5Myr

## *North America*



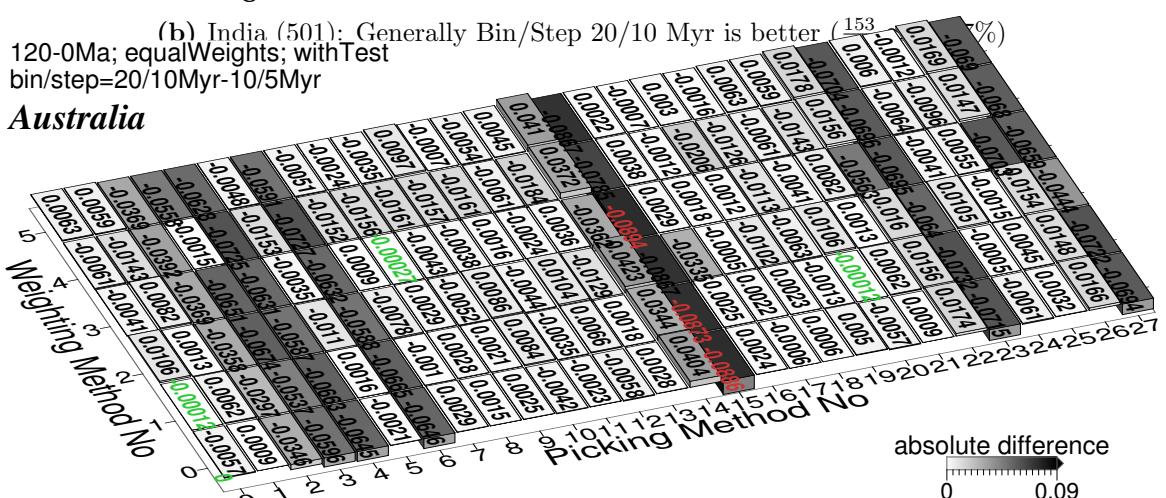
(a) North America (101): Generally Bin/Step 10/5 Myr is better (0.24%)  
 120-0Ma; equalWeights; withTest  
 bin/step=20/10Myr-10/5Myr

India



(b) India (501): Generally Bin/Step 20/10 Myr is better ( $\frac{153}{120}$ %)  
 120-0Ma; equalWeights; withTest  
 bin/step=20/10Myr-10/5Myr

Australia



**Figure 3.19** Australia (801). Differences between grids in Fig. 3.9 (10 Myr bin, 5 Myr step) and Fig. 3.17 (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 3.4:** 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. 3.9a	Fig. 3.17a   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	N	Picking no. 2, 5, 7, 17, 22 and 26 for 10/5 (E); 0, 8, 10, 12, 20	Picking no. 10/10 (3 UE): 23 and 27 for	Same: Picking no. 0, 2,	Y	Y	Y	Y	N	N	Positive values in Fig. 3.19a
Fig. 3.9b	Fig. 3.17b   Y	4 more best:	Picking no. 9, 21, 22 and 26	Same: Picking no. 4, 5, 6,	N	Picking no. 10/10 (3 UE): 23 and 27 for	Picking no. 10/10 (3 UE): 23 and 27 for	Same: Picking no. 0, 2,	Y	Y	Y	Y	N	N	Positive values in Fig. 3.19b
Fig. 3.9c	Fig. 3.17c   Y	only 20/10	Picking no. 23 and 27 only for 10/5	Same: Picking no. 1, 11, (E)	Sand: 19	Y	20/10/20/25	Same: Picking no. 4 only for 10/5 (E); 8 only for 20/10	Y	Y	Y	Y	N	N	Positive values in Fig. 3.19c
Fig. 3.21a	Fig. 3.23a   N	Picking no. 1, 9, 11, 13, 19, 21 and 25 for 10/5 (UE); 22 and 26 for	Same: Picking no. 5, 7 and 15	Sand: 25	N	Picking no. 16 and 17 only for 10/5 (E); 8 only for 20/10	Picking no. 10/10 (3 UE): 8, 14 and 20 for 10/5 (E)	Same: Picking no. 0, 10, 12, 14, 18, 20 and 24	N	Y	N	N	N	(0.8,10,11,12,14,18,20,24,25)[0-5)	(1.13,19,21)(0,1,5) 3(0,3,5) (5,7)(0-2,4,5) 15(0,4)
Fig. 3.21b	Fig. 3.23b   N	Picking no. 19 and 21 only for 10/5 (UE); 4 and 20/10 (E)	Same: Picking no. 5, 7, 22 and 26	Picking no. 19 and 21	Same: Picking no. 8, 14 and 20 for 10/5 (E)	Picking no. 10/10 (3 UE): 8, 14 and 20 for 10/5 (E)	Picking no. 0, 2, 10, 12, 16,	Picking no. 0, 2, 10, 12, 16,	Y	Y	Y	Y	N	N	(0.23,20,24,5,3,5); 4(5,3)mt for 6(0,7,13,27)(0-5) 21(0,1,3-5)
Fig. 3.21c	Fig. 3.23c   Y	6 more before 20/10 (E) no. 15, 23 and 27 only for 20/10 (E)	Same: Picking no. 1, 11, 13, 17, 19, 21 and 25	Same: Picking no. 19 and 21	Sand: 24	Y	Picking no. 10/10 (3 UE): 15, 23 and 27 only for 20/10 (E)	Picking no. 2, 14, 16, 22 and 26	Picking no. 2, 14, 16, 22 and 26	Y	Y	Y	N	N	(0.22,0,4); account for 4(0,1,1,1,1,1,1,1,1,21,25)(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

- 13. 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.
- 14. Methods 16/17 are still ‘bad’ for North America and India. Method 16 is still ‘bad’ and method 17 is still ‘good’ for Australia.

### 3.3.4.3 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. 3.18e is improved a bit compared to Fig. 3.10e). 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. 3.18 and even 30/15 Myr window and step; Table 3.5 and Fig. 3.20) 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 3.5 and Fig. 3.20). 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.

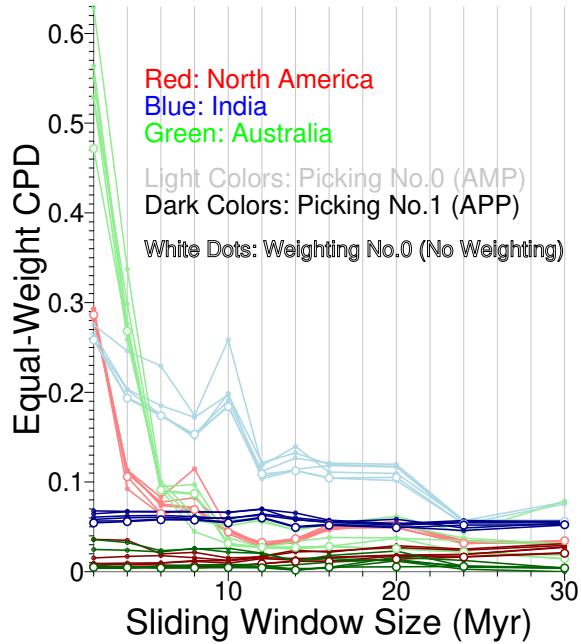
Note that there are 135, 75 and 99 paleopoles that compose of 120–0 Ma APWPs of North America, India and Australia respectively. Does the reason of 10/5 Myr generally better than 20/10 could be the relatively larger number of paleopoles for

**Table 3.5:** 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.

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 3.5), as expected, AMP needs wider sliding window and step to get closer to the reference path while APP does not (Fig. 3.20). 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 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.

**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.

The results are summarised in Table 3.4, Table 3.5 and Fig. 3.20.



**Figure 3.20:** Plot of the equal-weight  $\mathcal{CPD}$  scores collected in Table 3.5. Note that here the step size is always half of the sliding window size and the reference path is the FHM derived.

#### 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. 3.20). 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.

#### 3.3.5 Different reference path

Fig. 3.21 and Fig. 3.23 show the  $\mathcal{CPD}$  scores for the APWPs generated with all 28 picking and 6 weighting methods, compared to the 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.

### 3.3.5.1 Overall change

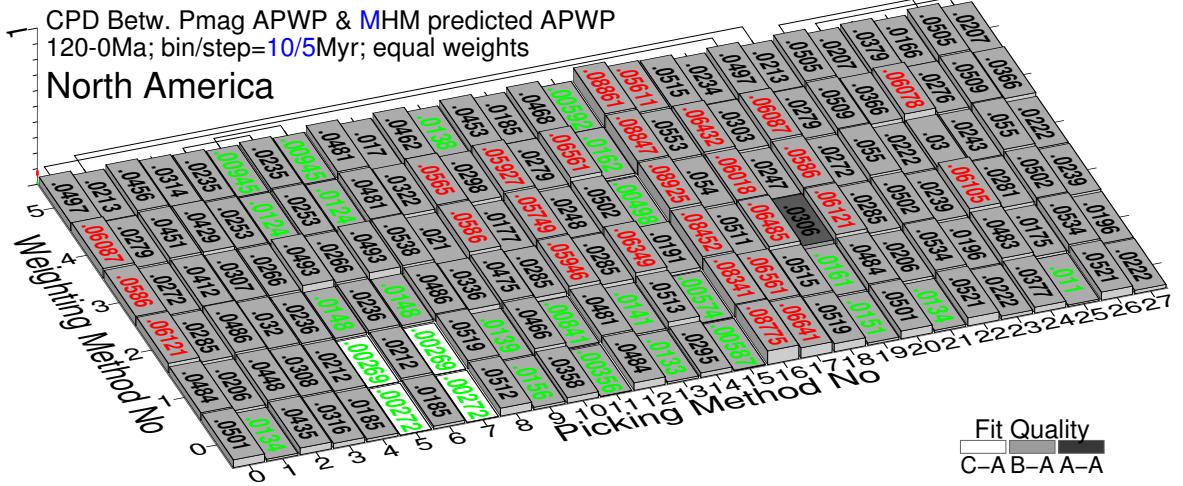
Although the reference path is changed to MHM, mean, median and range values of  $\mathcal{CPD}$  scores are almost unchanged and comparable (Fig. 3.9 vs. Fig. 3.21, and Fig. 3.17 vs. Fig. 3.23).

For 10/5 Myr bin/step, the absolute differences (Fig. 3.25) 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. 3.25), for North America (Fig. 3.25a), FHM derived path is a slightly better reference in general, while for both India (Fig. 3.25b) and Australia (Fig. 3.25c), generally MHM derived path is a slightly better choice.

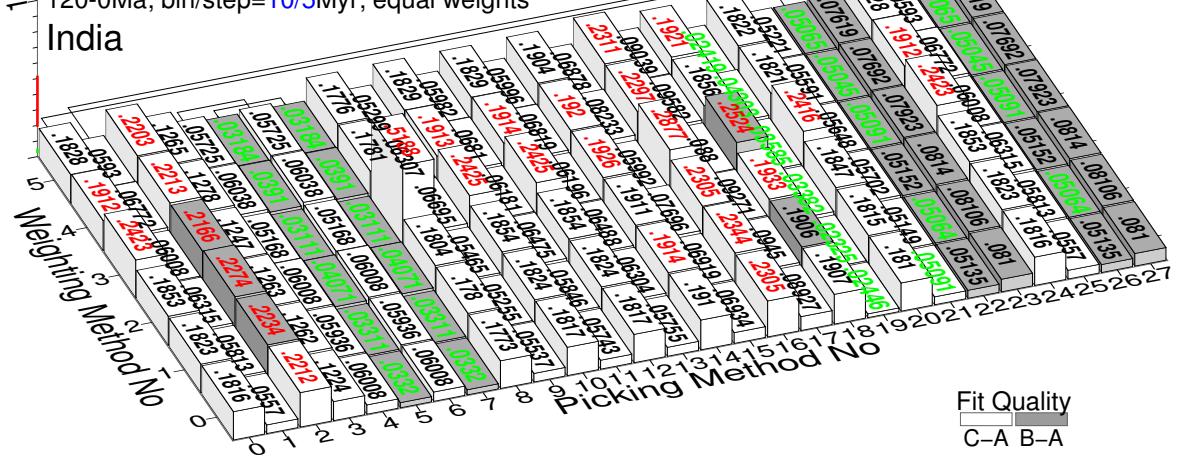
For North America, large changes seem favored by weighting method 3, and small changes seem favored by weighting method 5 or picking method 15 (Fig. 3.25). Method 3 brings minor changes to both North America and India. Methods 22 and 26 show large changes and methods 17, 19, 21 and 25 show minor changes for both India and Australia.

### 3.3.5.2 Relative performance of methods

1. 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.
2. 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%).
3. Both scores and AMP/APP difference are still low for methods 4/5, 6/7.
4. Both scores and AMP/APP difference are still high for methods 16/17 (method 17 gives low scores for Australia).
5. Still scores are high but AMP/APP difference are low for methods 2/3.

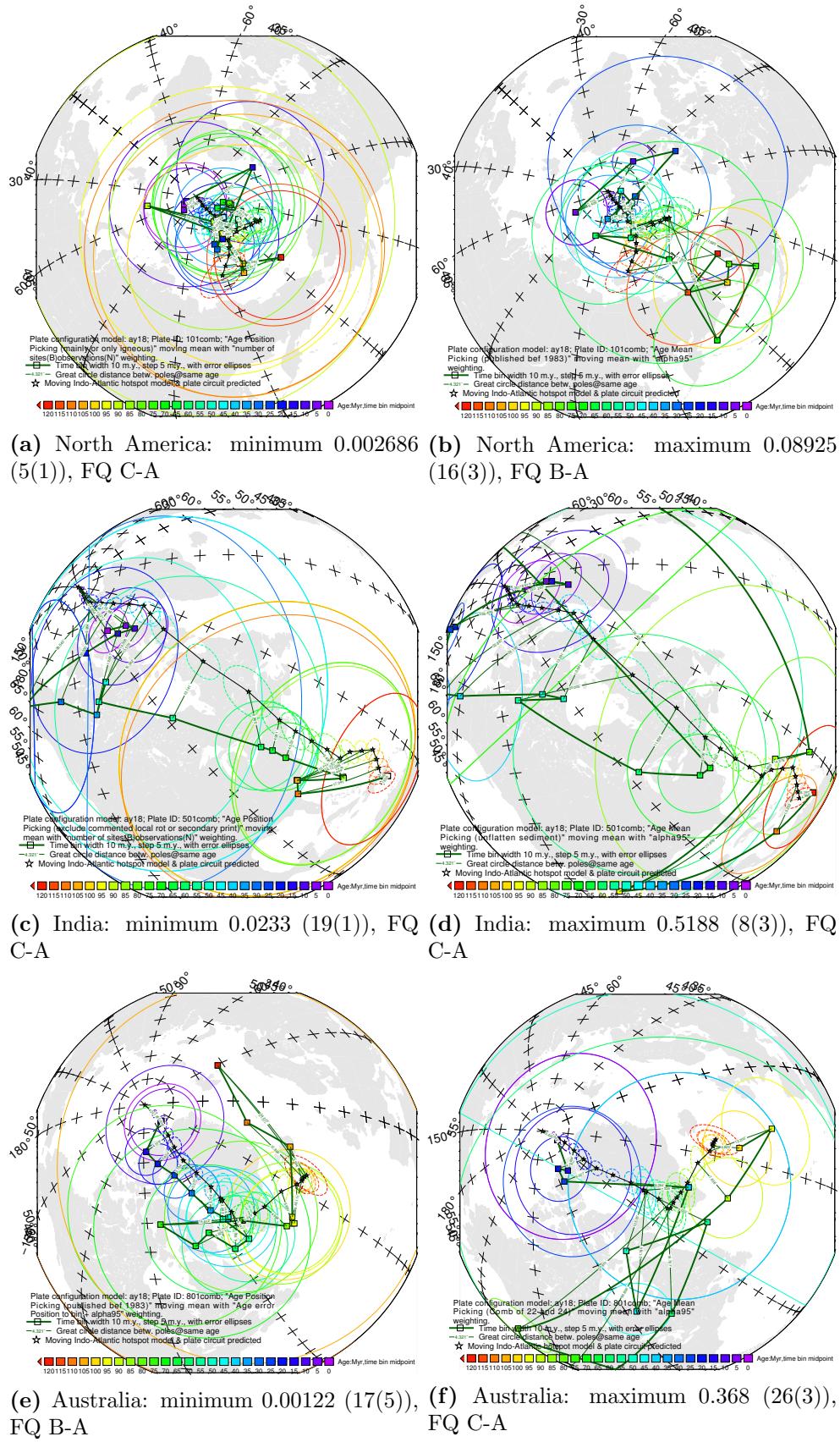


(a) North America (101 with children): minimum 0.00268588 (5(1)), maximum 0.0892467 (16(3)),  
me:CPD Betw. Pmag APWP & MHM predicted APWP  
120-0Ma; bin/step=10/5Myr; equal weights

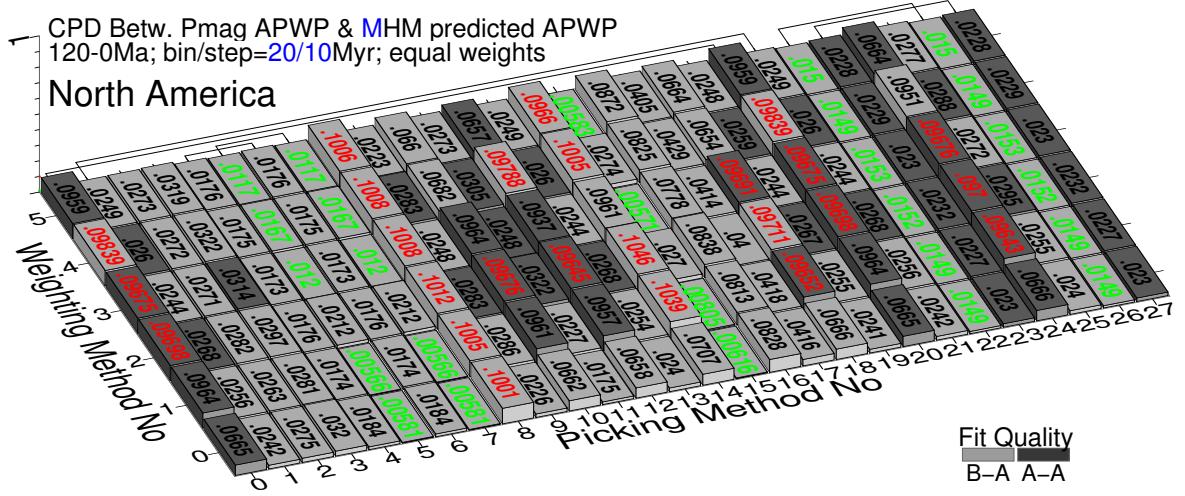


(b) India (501): minimum 0.0232517 (19(1)), maximum 0.51876 (8(3)), mean 0.11364, median 0.076556  
CPD Betw. Pmag APWP & MHM predicted APWP  
120-0Ma; bin/step=10/5Myr; equal weights



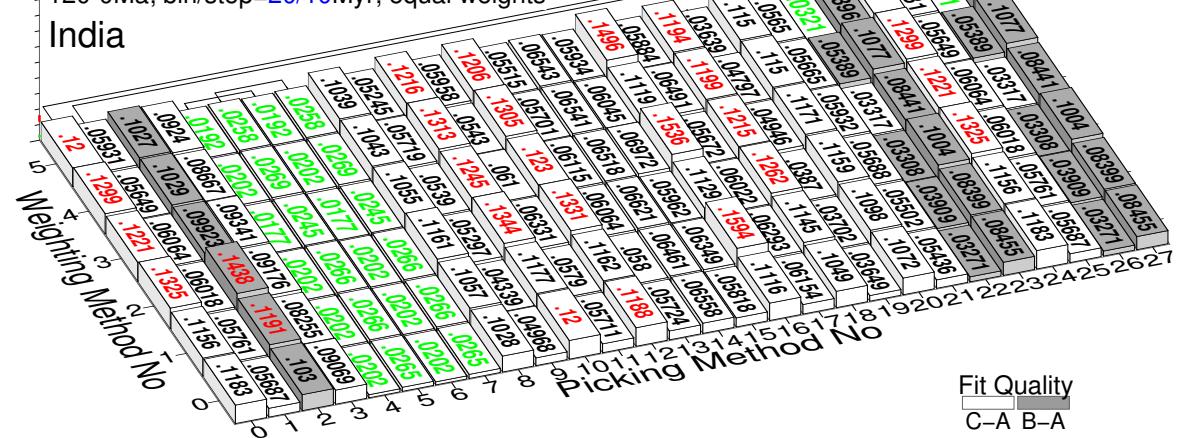


**Figure 3.22:** Path comparisons with best and worst difference values shown in Fig. 3.21. The parenthetical remarks are Picking No with Weighting No.



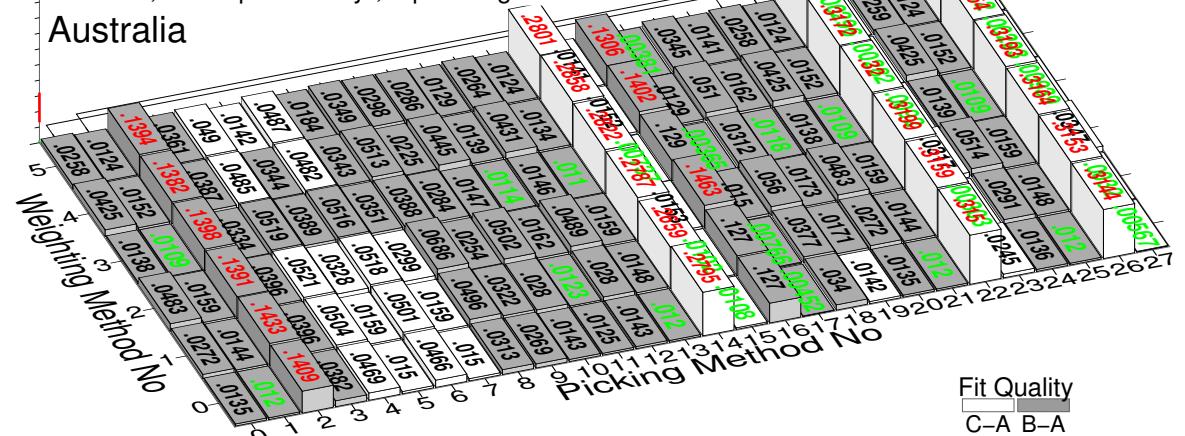
(a) North America (101 with children): minimum 0.00565784 (5(1)), maximum 0.104618 (14(2)), mean 0.039561, median 0.03717

CPD Betw. Pmag APWP & MHM predicted APWP  
120-0Ma; bin/step=20/10Myr; equal weights



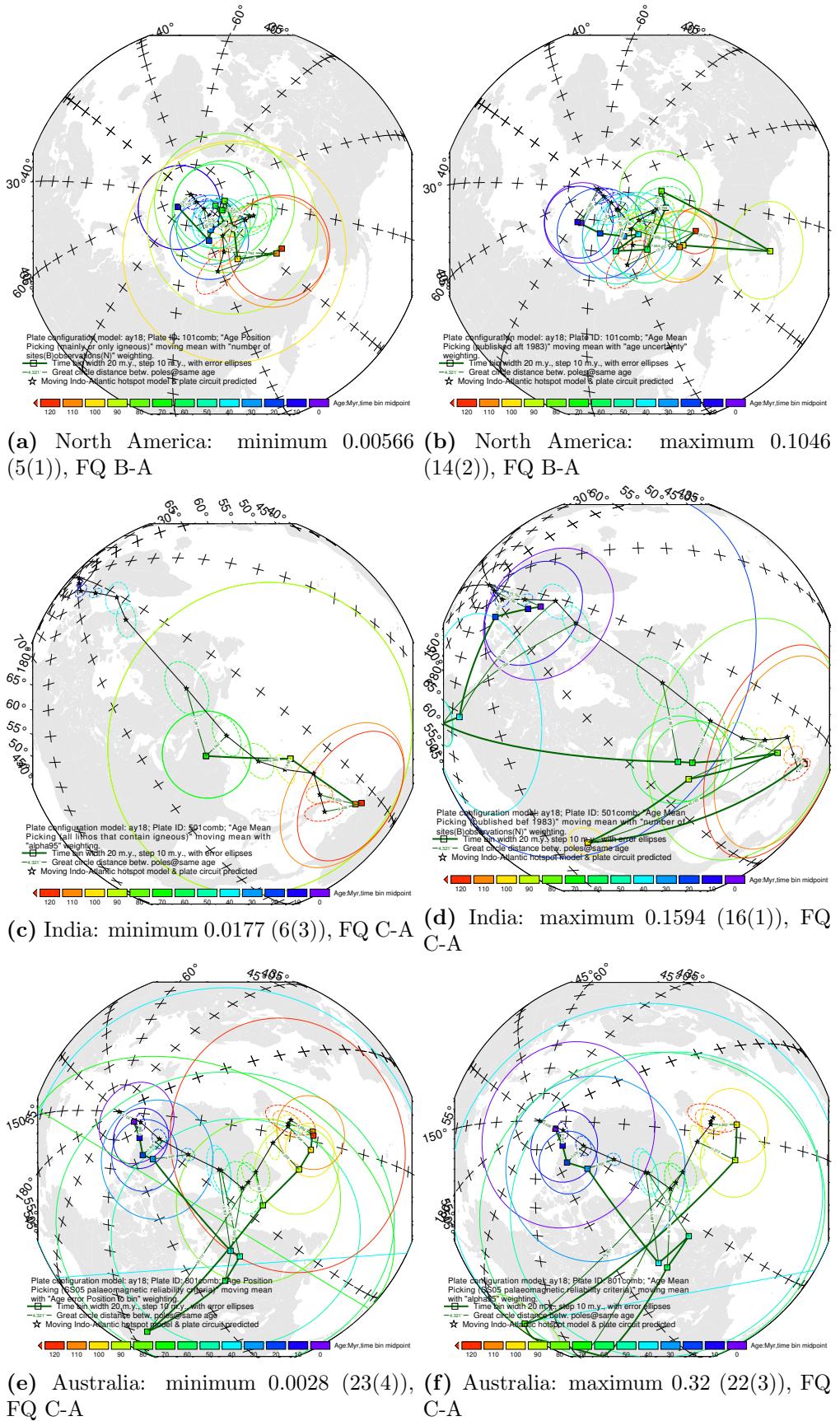
(b) India (501): minimum 0.0177 (6(3)), maximum 0.15937 (16(1)), mean 0.061346, median 0.061346

CPD Betw. Pmag APWP & MHM predicted APWP  
120-0Ma; bin/step=20/10Myr; equal weights



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

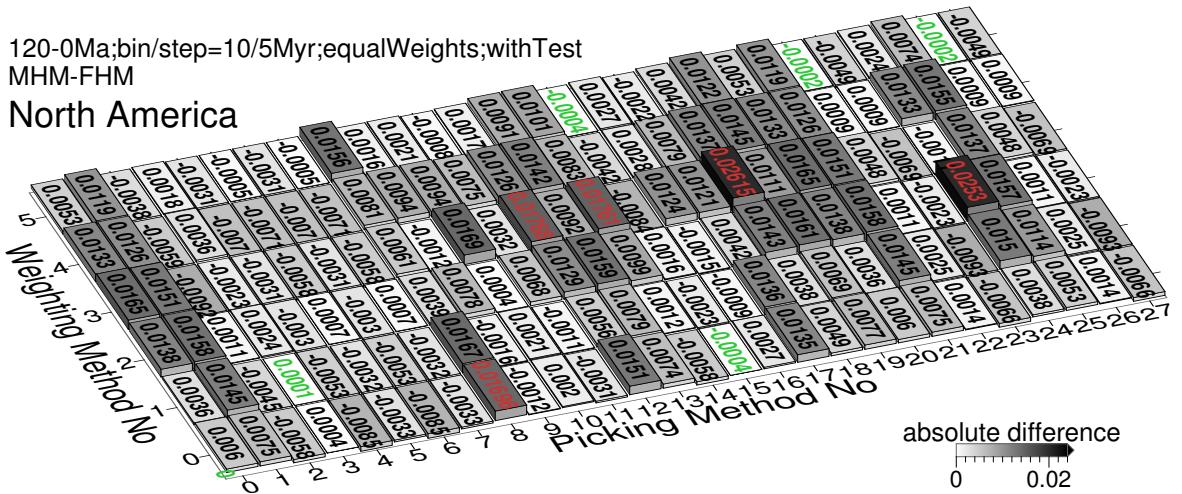
**Figure 3.23:** As Fig. 3.17, here the reference path is predicted from MHM. See the numbers of picked paleopoles in Fig. 3.9.



**Figure 3.24:** Path comparisons with best and worst difference values shown in Fig. 3.23. The parenthetical remarks are Picking No and Weighting No.

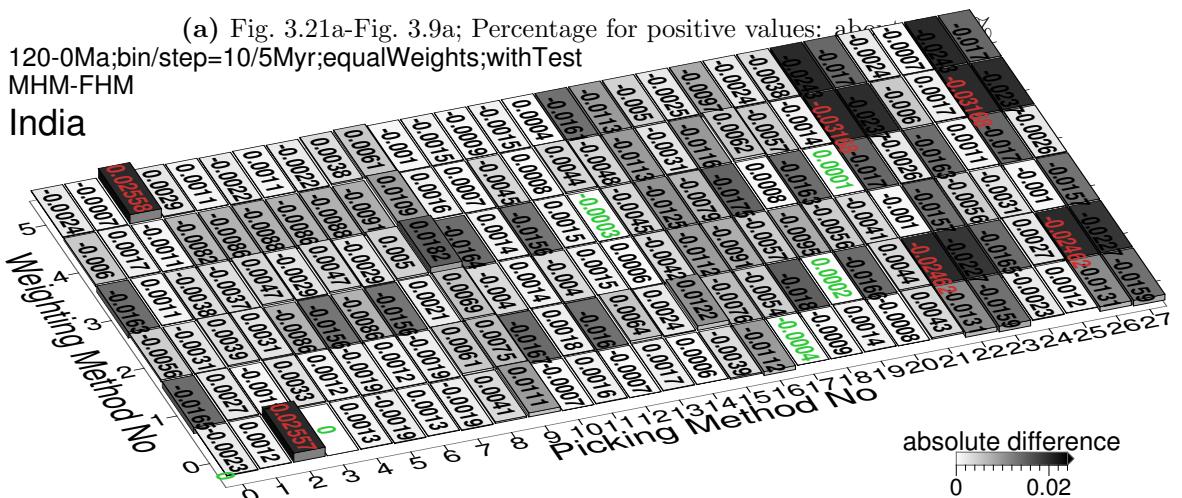
120-0Ma;bin/step=10/5Myr;equalWeights;withTest  
MHM-FHM

### North America



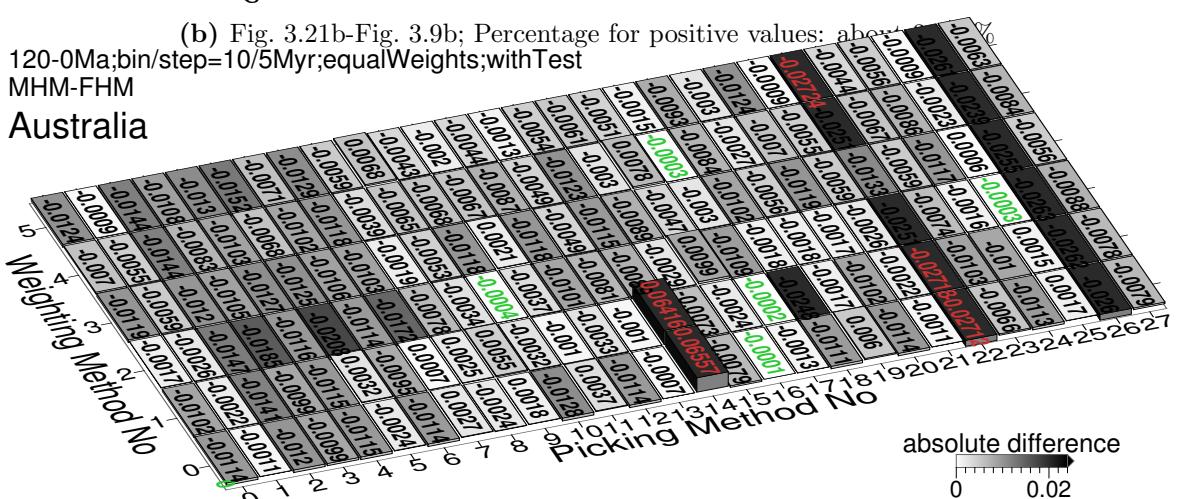
(a) Fig. 3.21a-Fig. 3.9a; Percentage for positive values: about 19.62%  
120-0Ma;bin/step=10/5Myr;equalWeights;withTest  
MHM-FHM

### India



(b) Fig. 3.21b-Fig. 3.9b; Percentage for positive values: about 19.62%  
120-0Ma;bin/step=10/5Myr;equalWeights;withTest  
MHM-FHM

### Australia



**Figure 3.25:** (c) Fig. 3.21c-Fig. 3.9c; Percentage for positive values: about 19.62%  
and MHM (Fig. 3.21) 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 3.6:** Performance statistics of all the picking and weighting methods.

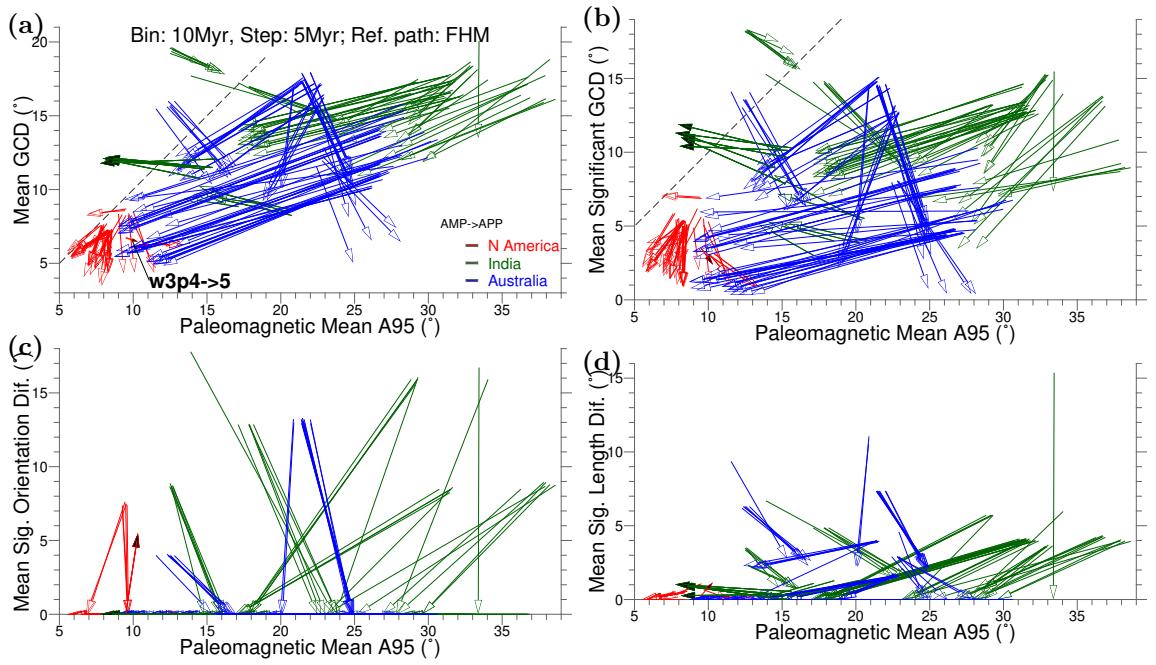
Grid	Best No.		Worst No.		Proportion Better	For All 28 Picking Methods, Count of Occurrences							Picking 14/15 of APP (Studies After 1983)	
	Picking	Weighting	Picking	Weighting		Fig	Best	4	7	0	5	Y/N	16/17 (Older)	
FHM	Fig. 3.9a	1, 5, 7, 11, 13, 15, <b>19</b> , 3, 5	0, 1, <b>16</b> , 17, 18, 2, 3,	2, 5, 7, 14, <b>16</b> , 17, 18, 2, 3,	Fig 4/APP AMP	9	1	4	7	0	5	Y/N	16/17	
	Fig. 3.9b	<b>21</b> , 5, 25, 6, 7, 9, <b>19</b> , <b>21</b> , 2, 3,	0, 1, 12, <b>16</b> , 18, 2, 3,	<b>02</b> , <b>26</b> , 8, 10, 12, 16, 18, 2, 3,	18/21	15	2	2	6	2	1	Y/Y		
	Fig. 3.9c	<b>22</b> , <b>26</b> , 13, 17, <b>19</b> , <b>21</b> , 3, 5	0, 51, 22, 26	<b>20</b> , <b>24</b> , 14, <b>16</b> , 17, 26	1	10	5	1	7	2	3	N/N		
	Fig. 3.17a	25, 4, 5, 6, 7, 9, 11, 3, 5	0, 1, 14, <b>16</b> , 18, 2, 3,	0, 8, 10, 12, 14, 16, 18, 2, 3,	61/84	14	2	0	3	3	6	N/Y		
	Fig. 3.17b	43, 5, 56, <b>19</b> , <b>20</b> , 25	0, 1, 2, 3,	00, <b>22</b> , 10, 12, <b>16</b> , 18, 20, 2, 3,	29/42	11	6	0	9	1	1	Y/4y2n		
	Fig. 3.17c	1, 11, 13, 17, <b>19</b> , <b>21</b> , 3, 4,	0, 51, 22, 26	<b>23</b> , <b>24</b> , <b>26</b> , 12, 26	1	9	6	0	<b>10</b>	1	2	N/N		
MHM	Fig. 3.21a	23, 32, 57, 29, 11, 13, 15, 2, 3,	0, 1, 14, <b>16</b> , 17, 2, 3,	0, 10, 12, 14, 16, 17, 2, 3,	41/42	10	9	3	4	0	2	Y/Y		
	Fig. 3.21b	<b>19</b> , <b>21</b> , <b>20</b> , 21, 22, 26	0, 51, 2, 3,	08, <b>20</b> , <b>24</b> , 10, 12, 14, <b>16</b> , 2, 3,	6/7	12	2	0	7	4	3	Y/Y		
	Fig. 3.21c	1, 11, 13, 17, <b>19</b> , <b>21</b> , 2, 3,	0, 51, 26	<b>28</b> , <b>29</b> , <b>16</b> , 22, 14, 18, 20, 24, 2, 3,	83/84	6	4	2	<b>10</b>	4	2	N/N		
	Fig. 3.23a	35, 7, 15, 22, 26	0, 1, 2, 3,	0, 8, 10, 12, 14, 18, 20, 24, 2, 3,	16/21	10	8	1	4	1	4	5n1y/Y		
	Fig. 3.23b	4, 5, 6, 7, 22, 26	0, 51, 2, 3,	0, 2, 10, 12, <b>16</b> , 18, 24 2, 3,	59/84	6	7	0	6	3	6	Y/3n3y		
	Fig. 3.23c	1, 11, 13, 15, 17, <b>19</b> , 21, 23, 25, 4, 5	0, 51, 22, 26 4, 5	2, 14, <b>16</b> , 22, 26 4, 5	1	12	1	0	10	4	1	N/N		

27

6. Still scores are low but AMP/APP difference are high for 0/1 on Australia.
7. Comparable AMP/APP (20-10 vs. 10-5, or FHM vs. MHM) still appears for 8/9 on Australian plate.
8. Method 1 is still a good performer.
9. Methods 4/5, 6/7 are still ‘good’ aggressive filters.
10. Methods 2/3 (at least 2) are still ‘bad’ aggressive filters.
11. 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.
12. Methods 16/17 are still ‘bad’ for North America and India.

## 3.4 Discussions

The following discussions will be in Q&A style.

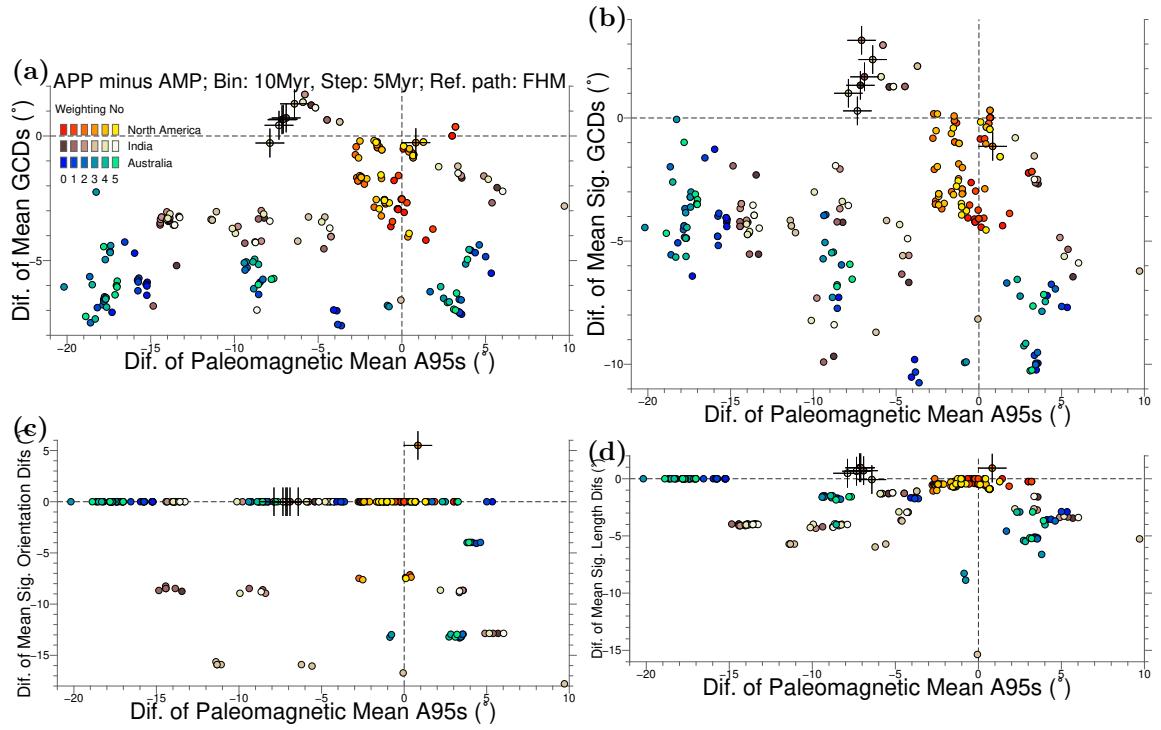


**Figure 3.26:** Paleomagnetic APWP's mean A95 versus (a) “mean GCD”, (b) “mean significant GCD”, (c) “mean significant orientation difference”, and (d) “mean significant length difference” between paleomagnetic APWP and its corresponding FHM-and-plate-circuit predicted APWP. Arrowtails are the results from AMP, while arrowheads are from APP. Black color filled arrowheads are the small number of special cases of AMP derived equal-weight  $\mathcal{CPD}$ s better than APP (see details in Fig. 3.9 and Table 3.6).

### 3.4.0.1 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. 3.26a and Fig. 3.26b). 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. 3.26 and Fig. 3.27). 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 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



**Figure 3.27:** Differences of APP and AMP coordinates shown in Fig. 3.26. Crosses locates the small minority cases of AMP derived equal-weight  $\mathcal{CPD}$ s better than APP (see details in Fig. 3.9 and Table 3.6).

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 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 3.6). 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.

### **3.4.0.2 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. 3.9a, 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 3.7). 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 3.8). This further causes greater distinguishable mean  $d_s$  from the APP methods. The similar phenomenon occurs to Fig. 3.17a 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 3.9) 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.

**Table 3.7:** 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					
		ds		dl		ds		da			
Age (Ma)	A95 (°)	Age (Ma)	Pmag A95 (°)	Dist (°)	Age (Ma)	Diff (°)	Age (Ma)	Pmag A95 (°)	Dist (°)	Age (Ma)	Diff (°)
10	1.44607/0.793714	10	14.876819	9.937	105-110	5.91855	15	2.0857	11.805	10-15-20	126.59
15	1.2875/0.816514						25	6.3358	6.873	dl	
25	2.48031/1.10915									Age (Ma)	Diff (°)
55	3.58782/2.14032	55	4.6347	5.372			60	6.5922	6.215	10-15	13.52
60	4.85938/3.17602						65	8.6632	7.6	15-20	14.68
65	3.68984/2.30014										
75	2.6435/1.54052	75	9.0812	8.836							
100	2.8983/2.68346	100	8.892	8.455							
105	2.32328/1.74639	105	5.3	5.03							
110	4.13015/2.25964	110	3.8	9.8064							
115	4.63512/2.58006	115	19.6676	9.3345							
120	7.34408/4.06043	120	3.515	17.35							
125	2.73408/2.06043	125	1.515	17.35							

**Table 3.8:** 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	
	A95 (°)	Pmag A95 (°)	Dist (°)	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>

Such cases include Fig. 3.9b 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 3.6).

Other Type 1 (e.g. Table 3.7) cases: Fig. 3.17a picking/weighting 2/1 vs 3/1. Fig. 3.21a 4/3 vs 5/3, 6/3 vs 7/3.

Other Type 2 (e.g. Table 3.8) cases: Fig. 3.17a 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. 3.17b 4/0-5 vs 5/0-5, 6/0-5 vs

**Table 3.9:** 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 ds			picking 23 + weighting 2 ds			Age
		Pmag A95 (°)	Dist (°)	N	Pmag A95 (°)	Dist (°)	N	
Age (Ma)	A95 (°)							
0	0	<b>6.28</b>	<b>12.72</b>	<b>2</b>	<b>23.54</b>	<b>18.14</b>	<b>3</b>	<b>80-8</b>
10	1.12124/0.673225	5.4/3.1	29.9	1	5.4/3.1	29.9	1	<b>110-</b>
<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	

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. 3.17c 4/2 vs 5/2. Fig. 3.21c 8/5 vs 9/5. Fig. 3.23a 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. 3.23b 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. 3.17a picking/weighting 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5. Fig. 3.17b picking/weighting 22/1,4,5 vs 23/1,4,5, 26/1,4,5 vs 27/1,4,5. Fig. 3.23b 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5.

Other Type 3 (e.g. Table 3.9) cases: Fig. 3.21a 4/2–5 vs 5/2–5. Fig. 3.21b 22/0–5 vs 23/0–5, 26/0–5 vs 27/0–5.

### 3.4.0.3 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. 3.9, Fig. 3.17, Fig. 3.21, Fig. 3.23). This closeness is also generally observed in the form of clusters in Fig. 3.28 and Fig. 3.29. In addition, from the general statistics of performance of the six weighting methods shown in Table 3.6, 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 3.5), 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.

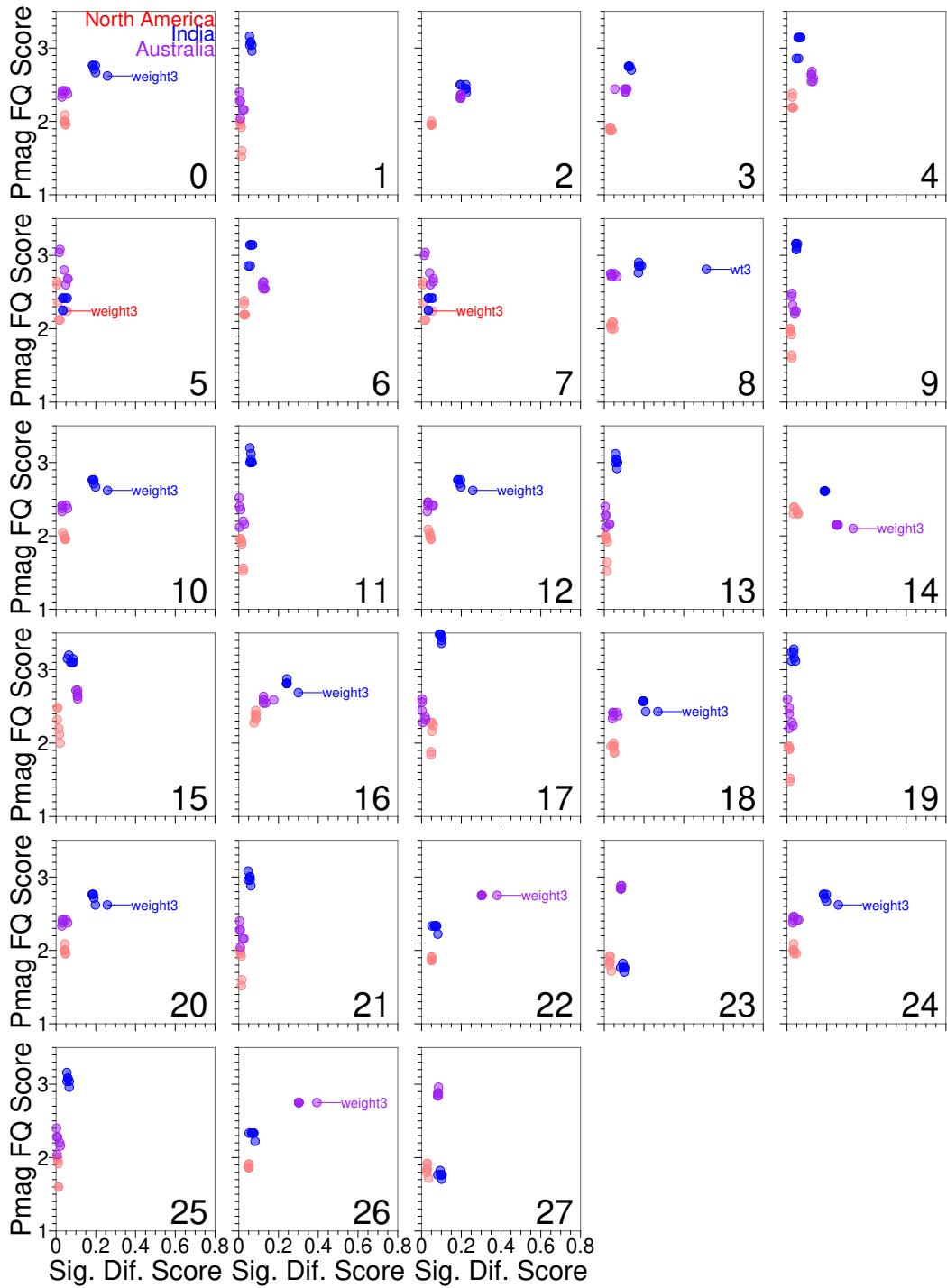
Therefore, at the very basic level, for example, a lower score is the result of one of, or combination of:

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).

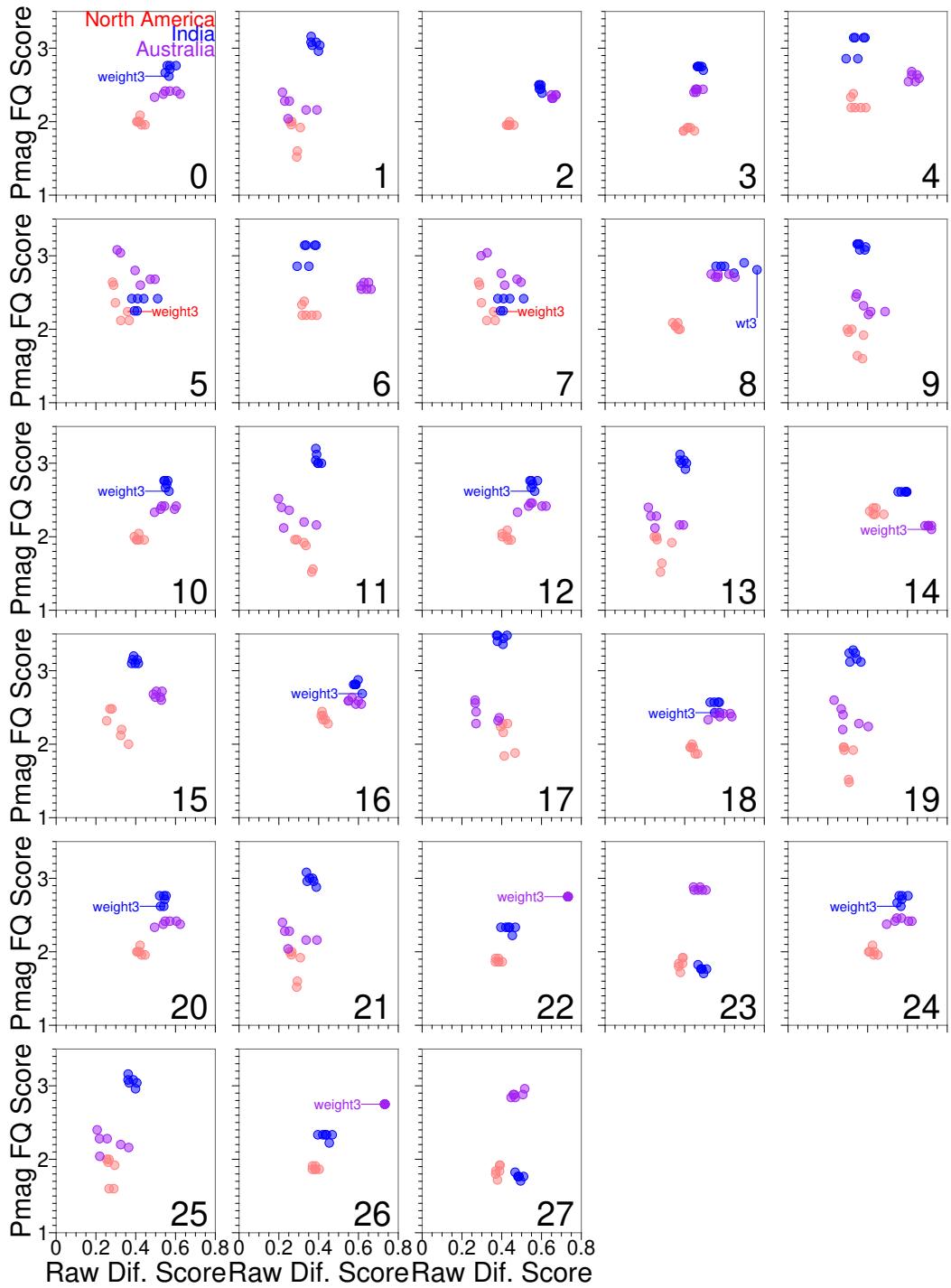
Fig. 3.30 and Fig. 3.31 show that the proportions of results from Weighting No 1–5 to result from 0 are generally in the second quadrant of the coordinate plane, where proportional change in difference score is positive whereas proportional change in paleomagnetic FQ score is negative. This means the 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. 3.9). From the labeled dots of Wt 3 in Fig. 3.28, Fig. 3.29, Fig. 3.30 and Fig. 3.31, 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.

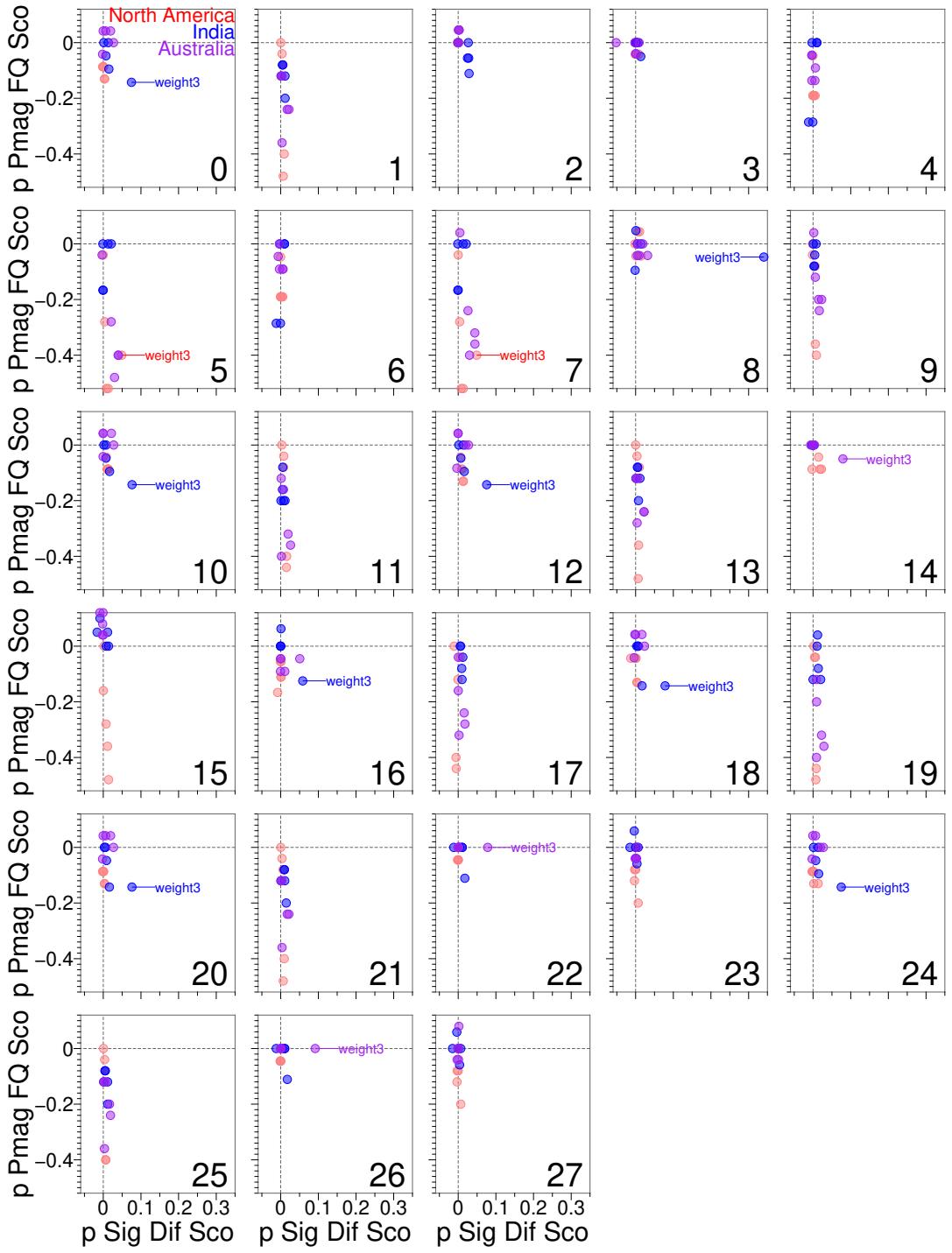
In addition, for Wt 3, a small size of  $\alpha_{95}$  (high precision) could be caused by those sampled directions not covering enough long period (thought to be at least about  $10^4$



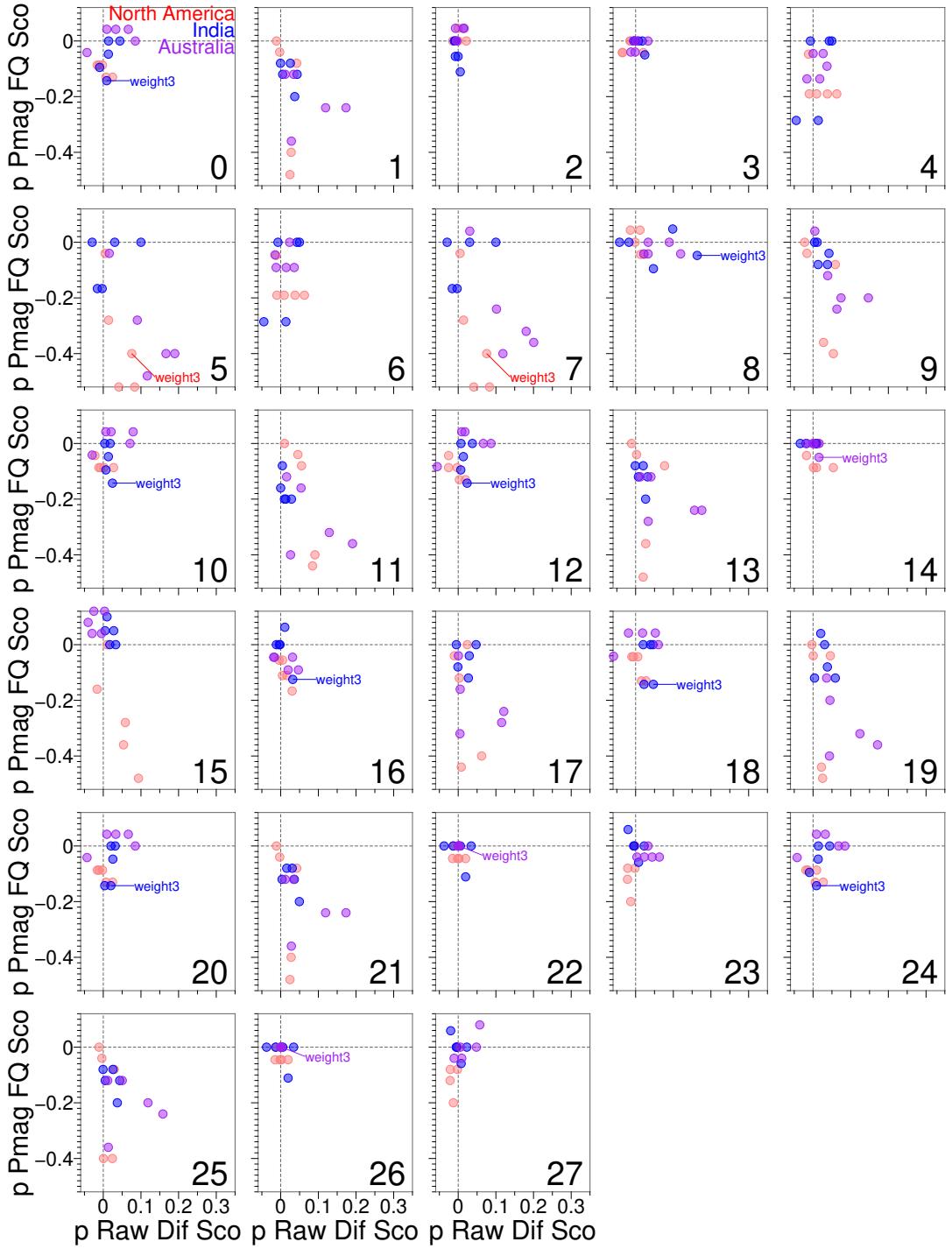
**Figure 3.28:** 10/5 Myr bin/step paleomagnetic APWP's FQ score (different from FQ, see the definitions of FQ and FQ score in Chapter 2) versus significant  $\mathcal{CPD}$  score (reference path: FHM predicted) for the 28 different picking methods. See the significant  $\mathcal{CPD}$  scores and Ppath-Rpath FQ in Fig. 3.9. Only those results dramatically worsened by weighting 3 are labeled.



**Figure 3.29:** 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. 3.28. Only those results dramatically worsened by weighting 3 are labeled.



**Figure 3.30:** 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 Ppath-Rpath FQ in Fig. 3.9. Only those results dramatically worsened by weighting 3 are labeled.



**Figure 3.31:** 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. 3.30. Only those results dramatically worsened by weighting 3 are labeled.

years) to “average out” secular variation for giving a paleopole. That is to say, the smallest  $\alpha_{95s}$  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 3.6), 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 3.6). However, Wt 3 is not recommended for North America and India.

#### **3.4.0.4 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. 3.17b, 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.

### **3.4.0.5 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. 3.14, Fig. 3.15 and Fig. 3.16).

## **3.5 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.

### **3.5.1 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. 3.9) 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.

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 3.6). 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 3.6).

Actually weighting is not improving the fit but improving precision generally. For quite many of the methods, no weighting is the best performer (Table 3.6). 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.

### **3.5.2 Conditional Rules of Ways of Processing Paleomagnetic Data:**

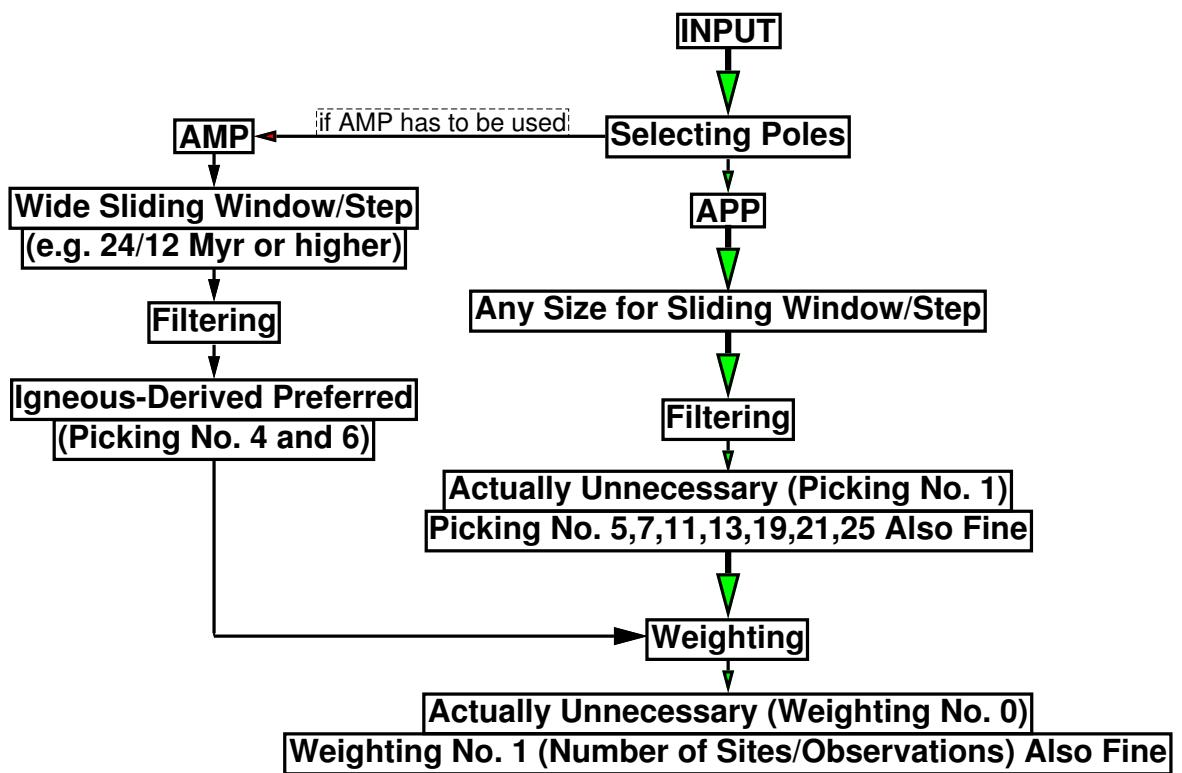
Picking methods no. 16 (AMP with data from old studies) and 18 (AMP without data affected by rotations and secondary print) are not recommended for generating a paleomagnetic APWP.

### **3.5.3 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 3.5 and Fig. 3.20). 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. 3.32).

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. 3.32).

All images are produced using GMT [65]. Thanks to Ohio Supercomputer Center for their remote HPC resources.



**Figure 3.32:** Flowchart for recommended procedure of processing paleomagnetic data.

# Chapter 4

## How Much Data Needed to Make a Reliable APWP

*This chapter mainly describes how the mean poles with their original paleopoles at random reduced densities can make a reliable APWP. Further we will see how much data (raw paleopoles) are needed on earth to make a reliable APWP, and how the “bad” paleopoles influence the final result when we have less data. Are we be able to make a final determination of best number of paleopoles in each sliding window in average for moving-averaging out an APWP? (No, different situations for different continents.)*

In the past, especially in deep time, the density and quality of paleomagnetic data are lower, compared with younger geological times. Reducing the data density for relatively younger geological times (for example, about 120–0 Ma) can help see if our methodology is still able to reliably give reasonable results from data aged in deep times (for example, geological times that are older than about 120 Ma, even to the Precambrian times or older).

## 4.1 Reference Path

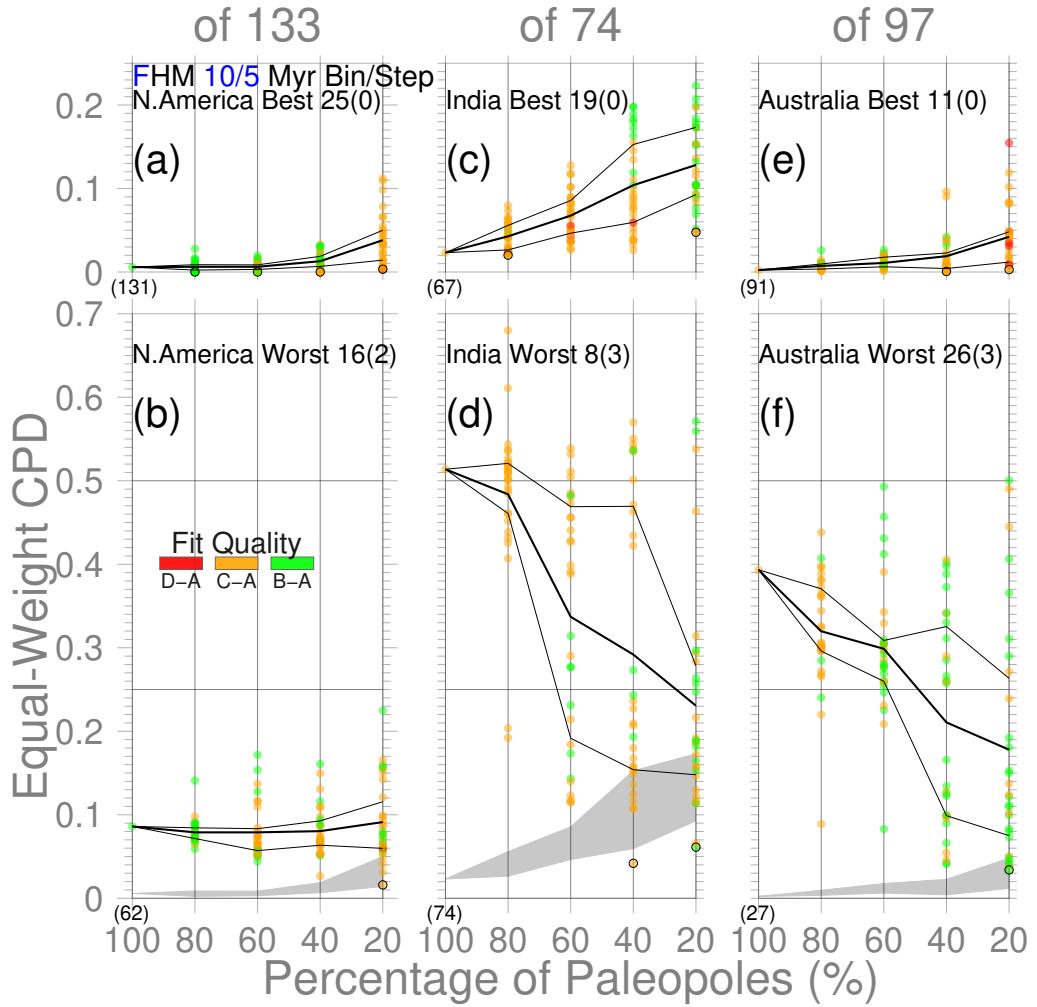
The fixed hotspot model (FHM) and related plate circuit predicted APWPs are used as references. In fact, as mentioned in the last chapter, whether we choose FHM or whether we pick MHM as reference, they do not make much difference at all.

## 4.2 Extraction Fraction

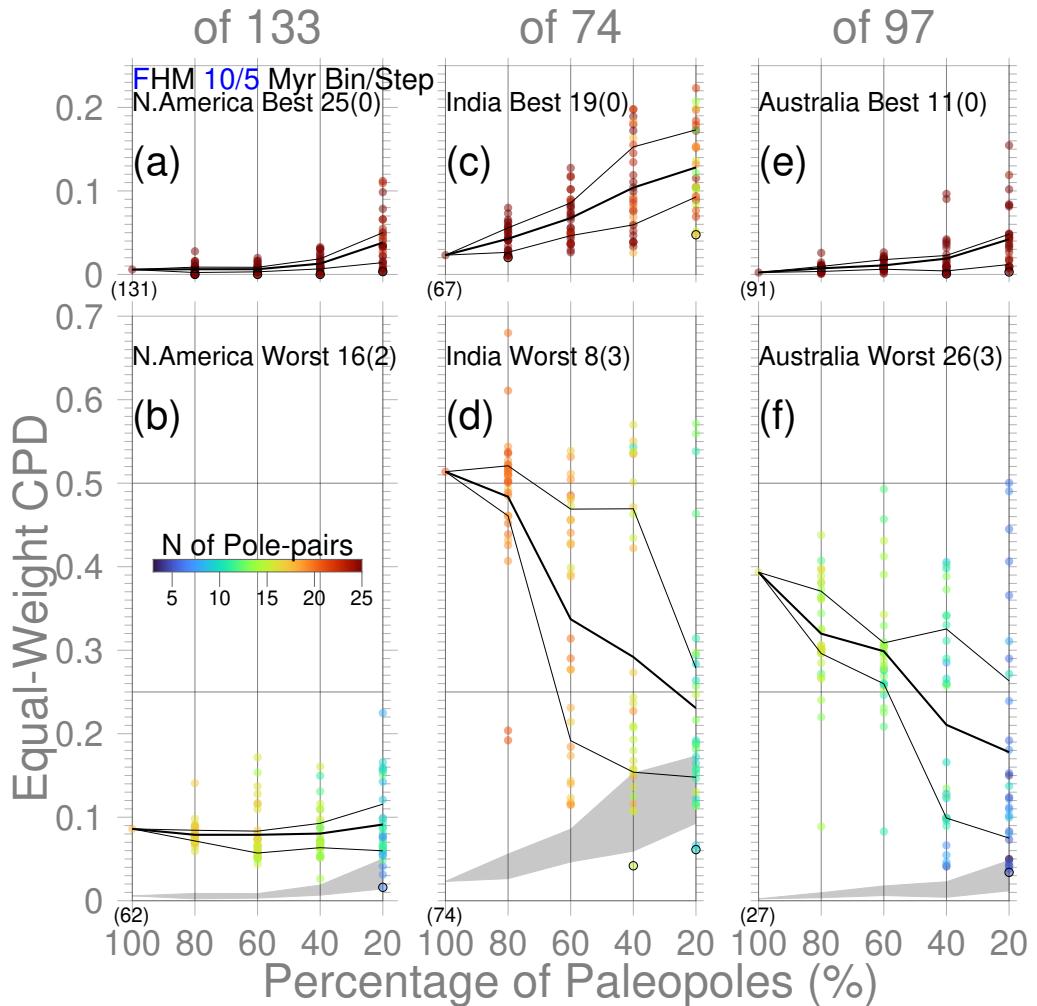
Sub-sampling (extracting part of raw paleopoles) is implemented before moving-averaging with filtering/correcting and weighting at four discrete percentages, 80%, 60%, 40% and 20%, which mean respectively 20, 40, 60 and 80 per cent of raw paleopoles are removed. This means not all sub-samples at, for example, 80% are going to be used to generate a path from the same number of paleopoles after filtering. In some cases a large number might be removed, in others much less, depending on the properties of the sub-sampled population. This is definitely an additional factor that would affect the difference score.

We can see that the best picking and weighting methods are statistically always better than the worst ones even if only 20 percent of the 120–0 Ma paleopoles are used to compose the APWPs (Fig. 4.1 and Fig. 4.2) for the three continents, North America, India and Australia.

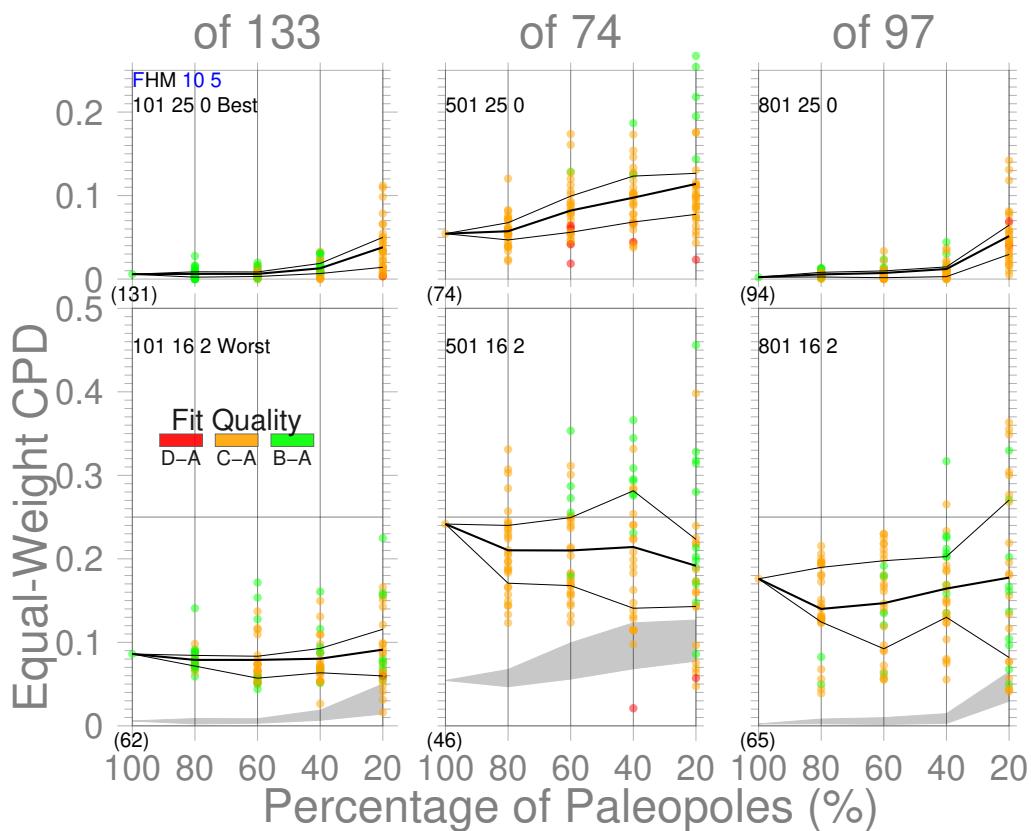
For the worst methods applied onto Indian and Australian data, the equal-weight  $\mathcal{CPD}$  surprisingly decreases when the percentage of extracted data decreases (Fig. 4.1). This is because after the data density is reduced the left data are not always enough to cover all the time range of 120–0 Ma but only part, or even though the 120 and 0 Ma mean poles (two ends) exist, the number of intermediate mean poles between 120 Ma and 0 Ma is much less than the APWP from data without reducing density (Fig. 4.2).



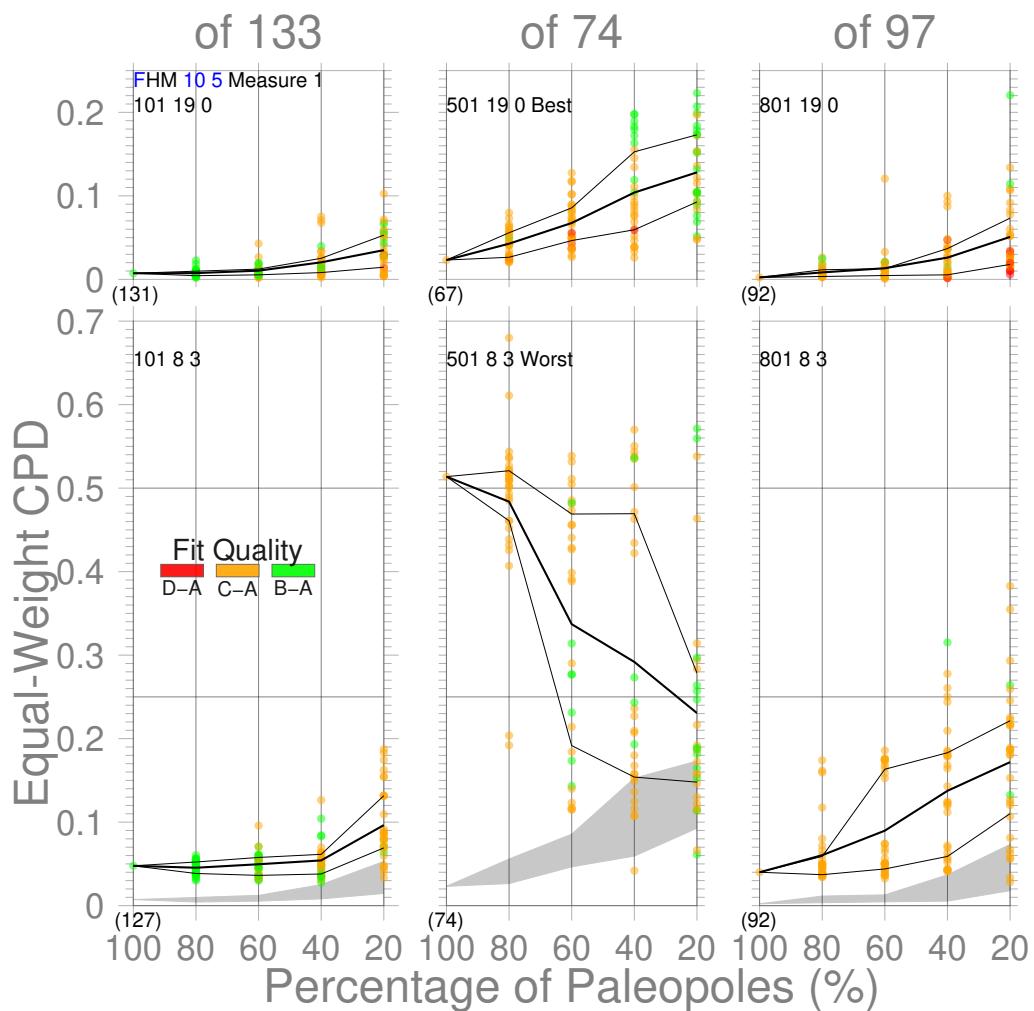
**Figure 4.1:** Random paleopole samplings (30 times) for the best and worst results for the 10 Myr window and 5 Myr step paleomagnetic APWPs vs FHM & plate circuit predicted APWP. The lower and upper bound lines connect the 1st and 3rd quantiles ( $Q_1$  and  $Q_3$ ) of the 30 samples. The bold line connects their means. The numbers in small parentheses are actual quantity of paleopoles after filtered by the corresponding picking methods for the case with no data removal. The  $Q_1$ - $Q_3$  interquartile range from best method is also shown (shadowed) in the plot of the worst method for clarity. Black rings are the lowest value for each method or the lowest for the 20% case.



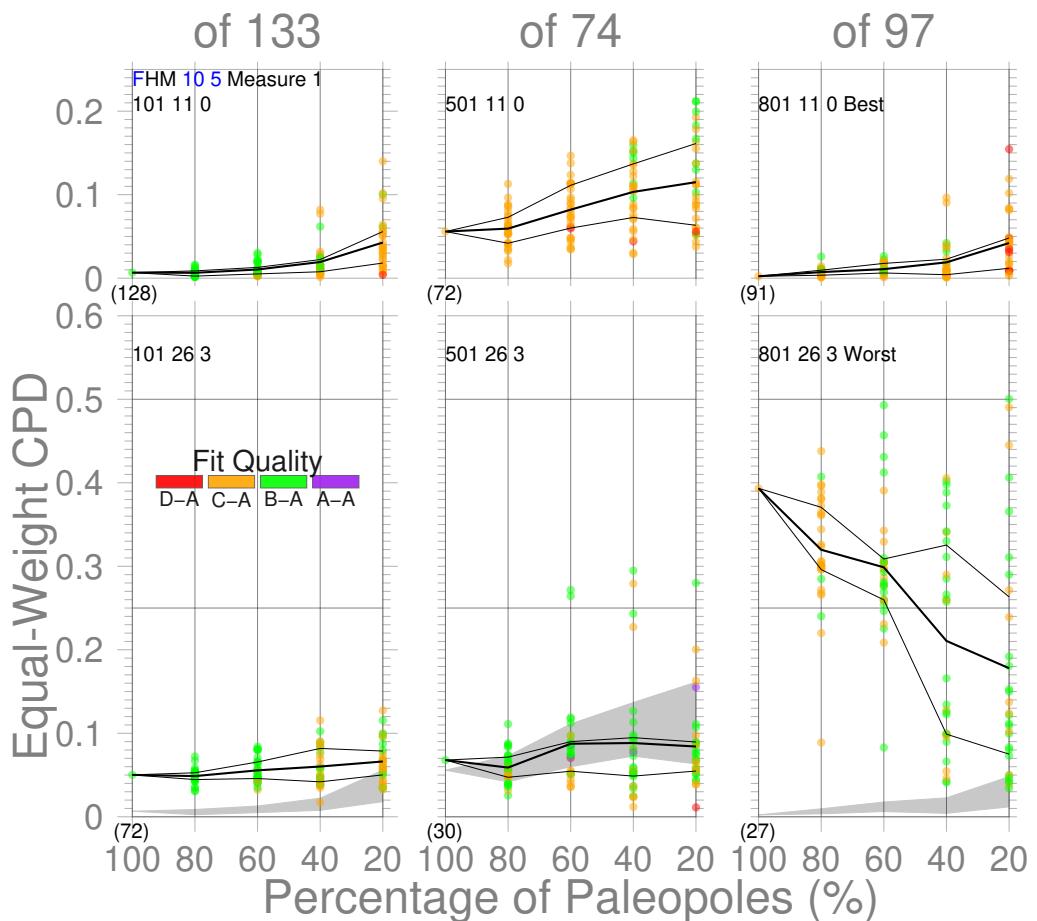
**Figure 4.2:** Same as Fig. 4.1. Here color demonstrates the resolution of pole pairs with the path comparisons.



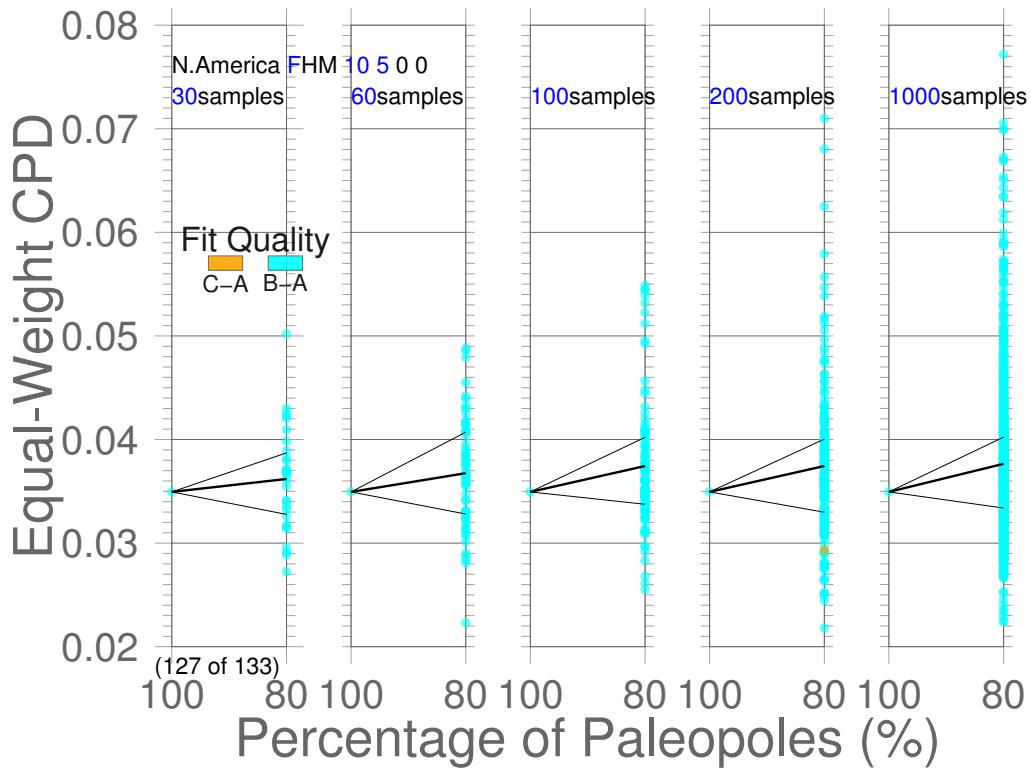
**Figure 4.3:** Comparisons of results from the best and worst methods for North America (101), also applied on the other two continents (501 and 801). The  $Q_1$ – $Q_3$  interquartile range from Picking No. 25 is also shown (shadowed) in the plot of Picking No. 16 for clarity.



**Figure 4.4:** Comparisons of results from the best and worst methods for India (501), also applied on the other two continents (101 and 801). The  $Q_1$ - $Q_3$  interquartile range from Picking No. 19 is also shown (shadowed) in the plot of Picking No. 8 for clarity.



**Figure 4.5:** Comparisons of results from the best and worst methods for Australia (801), also applied on the other two continents (101 and 501). The  $Q_1$ - $Q_3$  interquartile range from Picking No. 11 is also shown (shadowed) in the plot of Picking No. 26 for clarity.



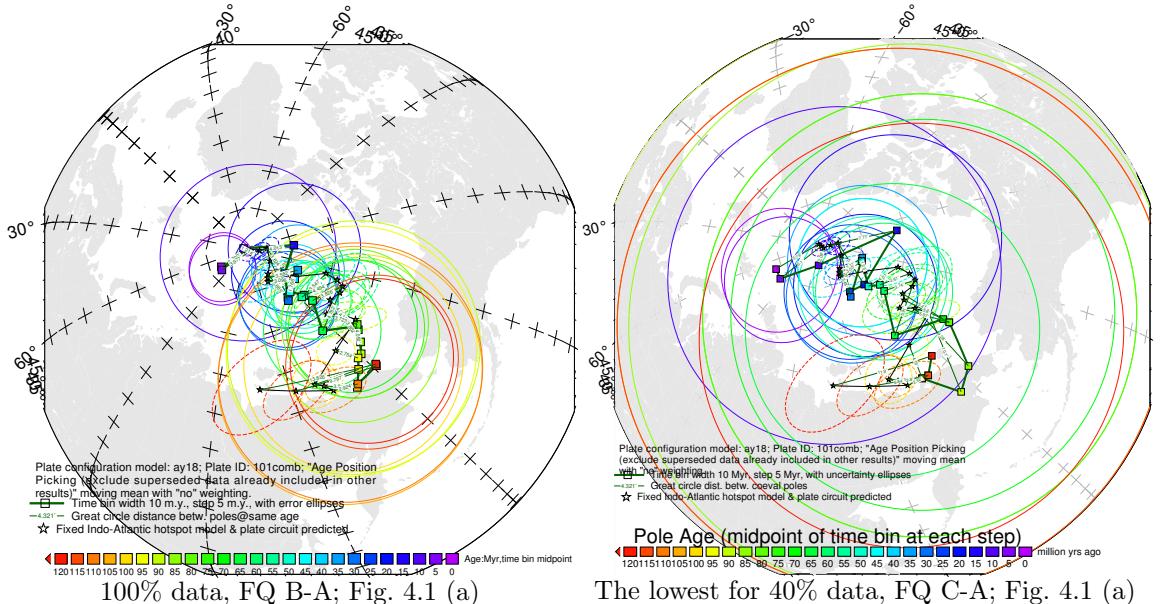
**Figure 4.6:** Testing differences of results from different numbers of samples. See Fig. 4.1 for more details.

#### 4.2.1 Number of Samples

Here because the thousands of times of testing for each representative percentage of data removal and also even further for each picking and weighting method is rather expensive, only 30 samples (a rule of thumb; e.g. [23] says “greater than 25 or 30”) are obtained for each percentage and method. In fact, the 25th percentiles ( $Q_1$ ), 75th percentiles ( $Q_3$ ) and the means of 30, 60, 100, 200 and 1000 samples are not quite different (Fig. 4.6), although 200 seems a better and relatively cheaper option.

#### 4.2.2 Extreme Value Study: Suggestions on Algorithm, especially on large uncertainties

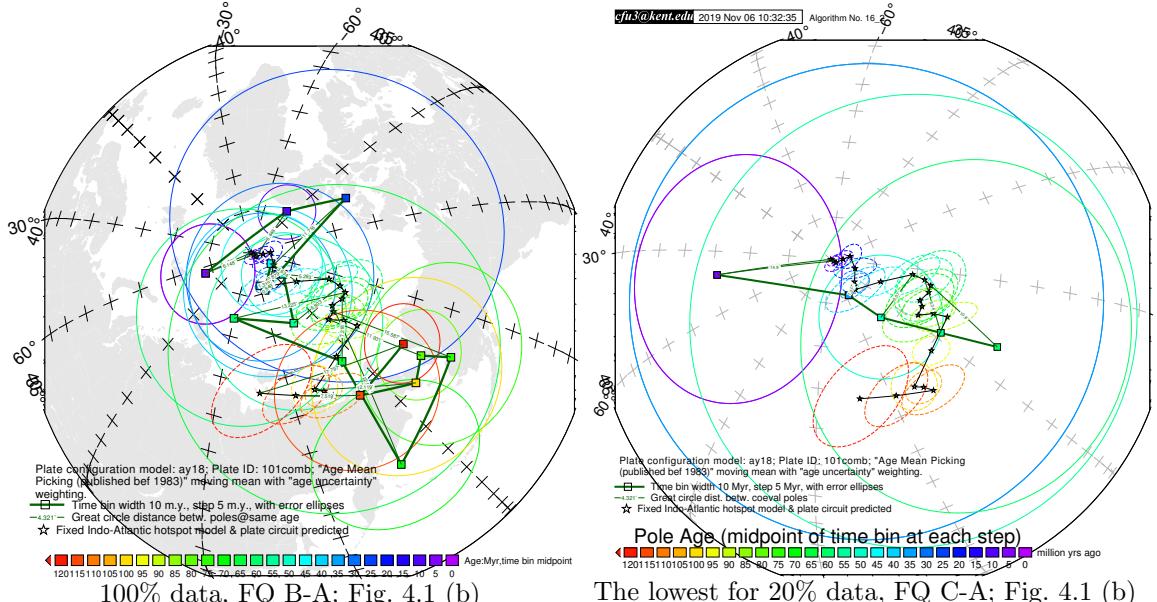
It seems rational for us to agree on that less paleomagnetic data tends to bring poorer similarity between paleomagnetic APWPs and the reference path. However, it is noticeable that even though the data density is tremendously reduced (e.g. by 80%), it seems still possible to have a better similarity score for paleomagnetic APWPs and the reference, sometimes even better than for the paleomagnetic APWP with all original number of datasets (e.g. the black-outlined rings in Fig. 4.1). For example, for the case (a) (Fig. 4.1), even though 60% of the paleopoles are removed, we still



**Figure 4.7:** Comparing the 100% North American 120–0 Ma paleomagnetic data derived result with the best of the only 40% data (giving even better similarity) derived results (the lowest yellow dot in Fig. 4.1 (a)).

can get a better similarity score (here, 0) using the paleomagnetic APWP composed of the left 40% of the paleopoles than the original. Although this 40% data generated paleomagnetic APWP owns the same number of mean poles as the 100% data generated paleomagnetic APWP, the average number of paleopoles for each mean pole is obviously much less. The main reason why this 40% data generated paleomagnetic APWP is more similar to the reference path is that this APWP's spatial uncertainties (FQ: C-A) are much larger than those (FQ: B-A) of the 100% data generated paleomagnetic APWP (Fig. 4.7). Even only 20% of the paleopoles could also give a better similarity (the lowest yellow dot in Fig. 4.1 (a)). Unfortunately, the reason why this 20% data generated paleomagnetic APWP is more similar to the reference path is the same as for the 40% data generated path: extremely large spatial uncertainties (FQ: C-A). So we need to be cautious about the similarity score when we do not have enough paleomagnetic data for making an APWP, which tends to generate large spatial uncertainties for mean poles. The situation is the same for the lowest difference given by the 20% North American paleomagnetic data with the worst method 16(2) applied (the lowest yellow dot in Fig. 4.1(b)).

Further for the case (b) (Fig. 4.1), the main reason why the only 20% data could give a better result than the 100% data does is that not only, for example, for the lowest yellow dot the 20% data gives less mean poles, but also a few extremely large spatial uncertainties appear for this 20% case (Fig. 4.8). The same situation happens to the green dot case (Fig. 4.1 (f)).



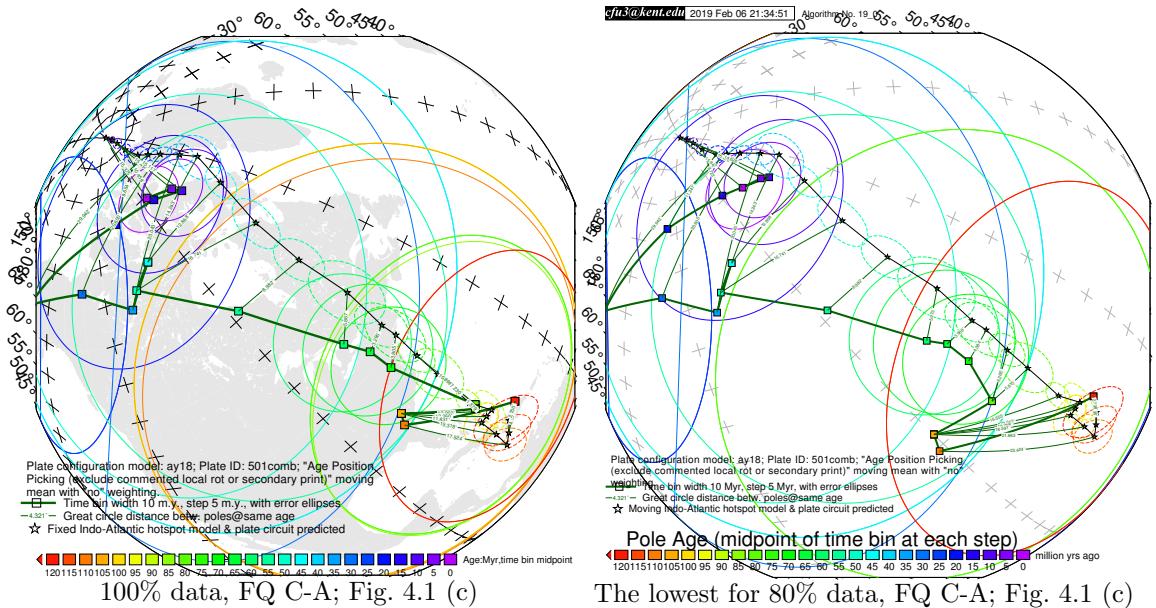
**Figure 4.8:** Comparing the 100% North American 120–0 Ma paleomagnetic data derived result with the best of the only 20% data (giving even better similarity) derived results (the yellow dot in Fig. 4.1 (b)).

For the case (c) (Fig. 4.1), the reason why the 80% data is able to give a better result than the 100% data does is that the 10 Ma mean pole of the 80% data derived path is a bit closer to reference, because both the 10 Ma pole pair in Fig. 4.9 are distinguishable. Although the 80% data derived paleomagnetic APWP (Fig. 4.9b) generally owns relatively larger spatial uncertainties, the related pole pairs are distinguishable for both path pairs in Fig. 4.9.

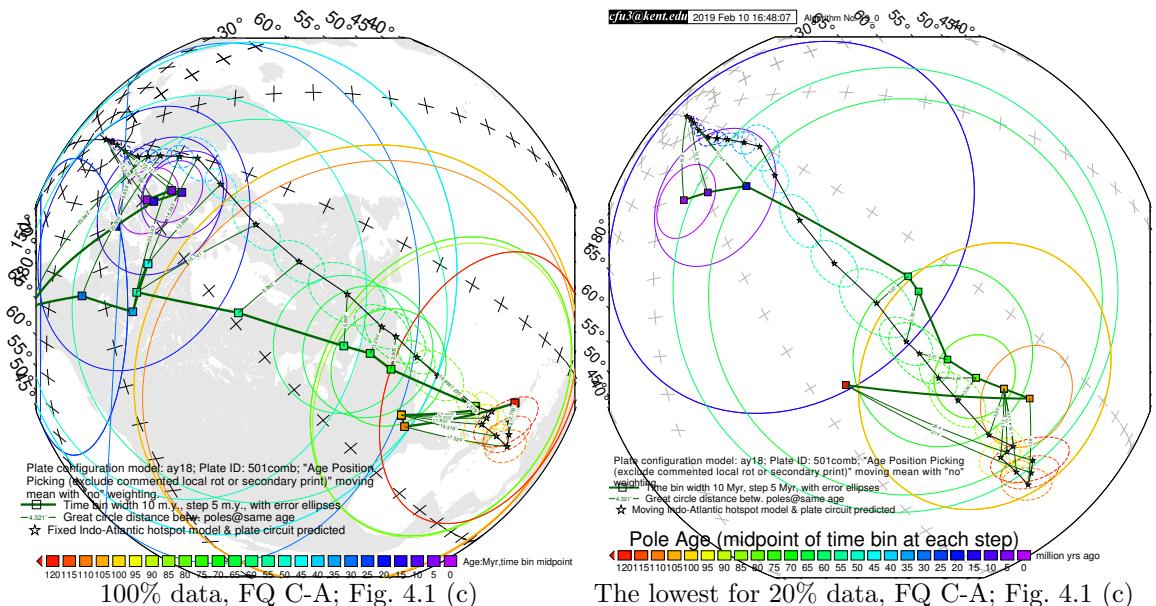
Still for the case (c) (Fig. 4.1), of the 30 samples for the 20% data, none is able to give better result compared with the 100% data, but the lowest difference value we can get from these 30 samples (20% in Fig. 4.1 (c)) indicates that even 20% data is still able to give good (not dramatically different from that 100% data gives) similarity (Fig. 4.10b).

For the case (d) (Fig. 4.1), the reason why the only 40% data could give a better result than the 100% data does is that for the yellow dot (Fig. 4.1 (d)) not only the 40% data gives less mean poles (but two ends 120 Ma and 0 Ma still exist), but also the 40% data does not contain some “bad” paleopoles that are far away from the reference path (Fig. 4.11b). It’s the same for the lowest difference given by the 20% data samples (the green dot in Fig. 4.1 (d)).

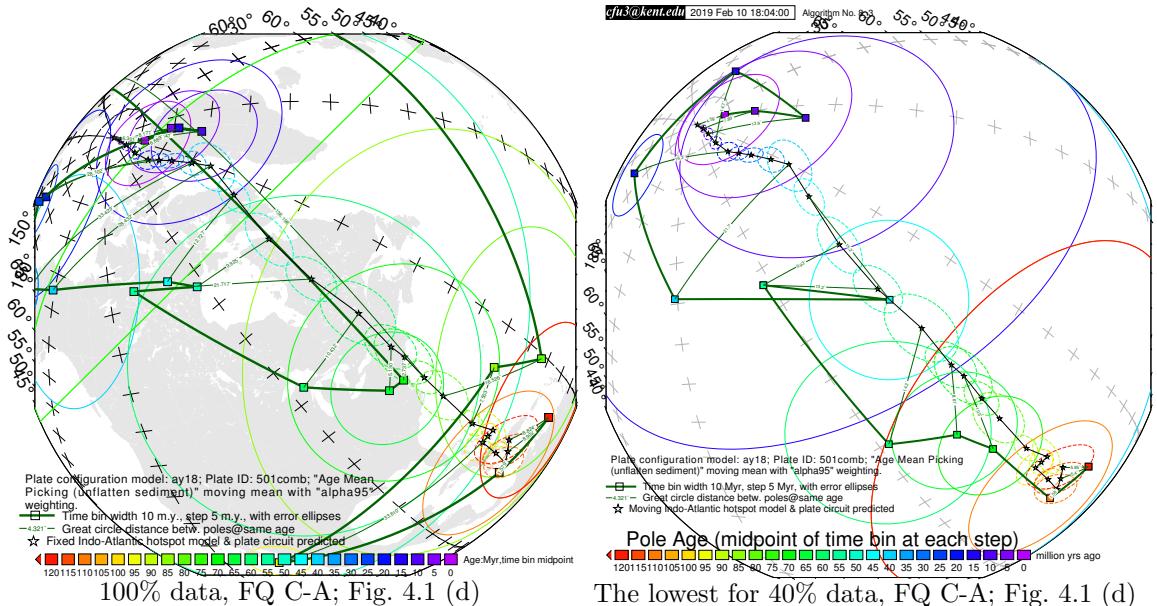
For the case (e) (Fig. 4.1), although the 20% data derived paleomagnetic paths are not closer to the reference path than the 100% data derived one, the closest one (the lowest yellow dot along the 20% grid line in Fig. 4.1 (e)) still performs quite well (Fig. 4.12b). This is mainly because the number of mean poles is still the same as



**Figure 4.9:** Comparing the 100% Indian 120–0 Ma paleomagnetic data derived result with the best of the only 80% data (giving even better similarity) derived results (the green dot in Fig. 4.1 (c)).



**Figure 4.10:** Comparing the 100% Indian 120–0 Ma paleomagnetic data derived result with the best of the only 20% data (giving even better similarity) derived results (the dark green dot in Fig. 4.1 (c)).



**Figure 4.11:** Comparing the 100% Indian 120–0 Ma paleomagnetic data derived result with the best of the only 40% data (giving even better similarity) derived results (the green dot in Fig. 4.1 (d)).

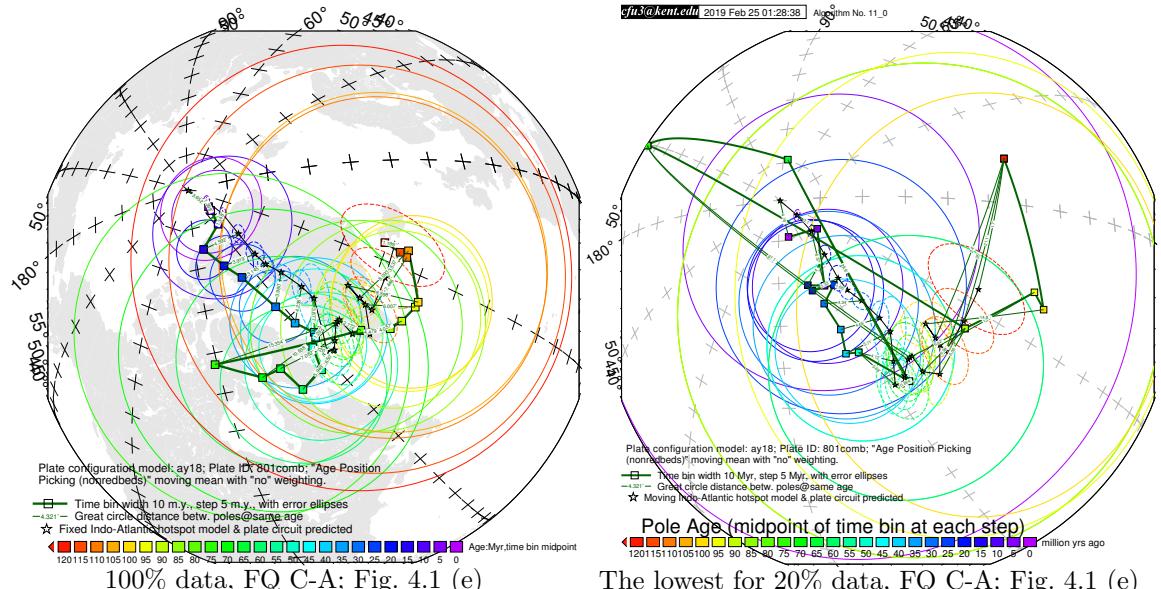
the 100% data generated path when there are only 20% of the paleopoles. However, for 120, 115, 110 and 105 Ma, where the mean poles, actually the same pole, are relatively far from the reference path for the 20% data (Fig. 4.12a), the uncertainties are also rather large (Fig. 4.1 (e)).

For the case (f) (Fig. 4.1), most of the 20% data derived paleomagnetic paths are closer to the reference path than the 100% data derived one, especially for the bottom green dot case in Fig. 4.1 (f) (Fig. 4.13b). This is mainly because the number of mean poles becomes much less when there are only 20% of the paleopoles, especially two end mean poles missing.

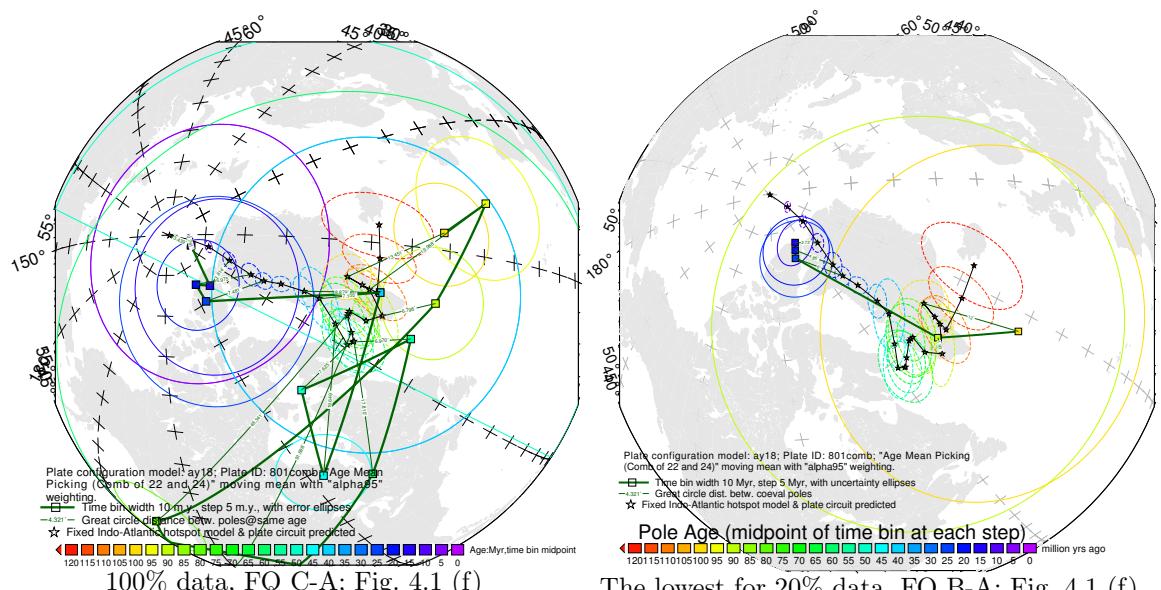
#### 4.2.3 Relationship between $d_s$ Score and Number of Paleopoles for Paleomagnetic APWP

From Table 4.1, it is noticeable that all the  $\mathcal{CPD}$ s contain contribution from  $d_s$ , and most of the  $\mathcal{CPD}$ s are actually only from  $d_s$ . This indicates the key role of  $d_s$  in calculating the difference score of paleomagnetic APWP from reference path.

Then the relationship between number of paleopoles for each mean pole and  $d_s$  from all the reduced-data-density experiments is investigated according to different tectonic plates (Fig. 4.14) and different amounts of density reducing (Fig. 4.15). Based on the trends of the means of the  $d_s$ s, after the number of paleopoles is more than 10, all the means of  $d_s$ s intend to be stable no matter what plates (Fig. 4.14) or how



**Figure 4.12:** Comparing the 100% Australian 120–0 Ma paleomagnetic data derived result with the best of the only 20% data derived results (the bottom yellow dot on the 20% gridline in Fig. 4.1 (e)).



**Figure 4.13:** Comparing the 100% Australian 120–0 Ma paleomagnetic data derived result with the best of the only 20% data derived results (the bottom green dot in Fig. 4.1 (f)).

**Table 4.1:** How many there are in the 168 CPDs where spatial difference ( $d_s$ ) contributes

Contributing Factors to CPD	FHM						MHM					
	101	501	801	101	501	801	101	501	801	101	501	801
Only $d_s$ (?/168[?%])	90[53.6]	60[35.7]	103[61.3]	102[60.7]	87[51.8]	138[82.1]	82[48.8]	67[39.9]	103[61.3]	114[67.9]	94[56]	138[82.1]
Only $d_t$ (?/168)	0	0	0	0	0	0	0	0	0	0	0	0
Only $d_a$ (?/168)	0	0	0	0	0	0	0	0	0	0	0	0
Only $d_s, d_t$ (?/168)	70	78	47	66	72	30	66	77	51	13	60	17
Only $d_s, d_a$ (?/168)	0	0	0	0	0	0	0	0	0	0	1	0
Only $d_t, d_a$ (?/168)	0	0	0	0	0	0	0	0	0	0	0	0
$d_s, d_t, d_a$ (?/168)	8	30	18	0	9	0	20	24	14	41	13	13

<sup>1</sup> See Figure 9 in Chapter 3

<sup>2</sup> See Figure 17 in Chapter 3

<sup>3</sup> See Figure 21 in Chapter 3

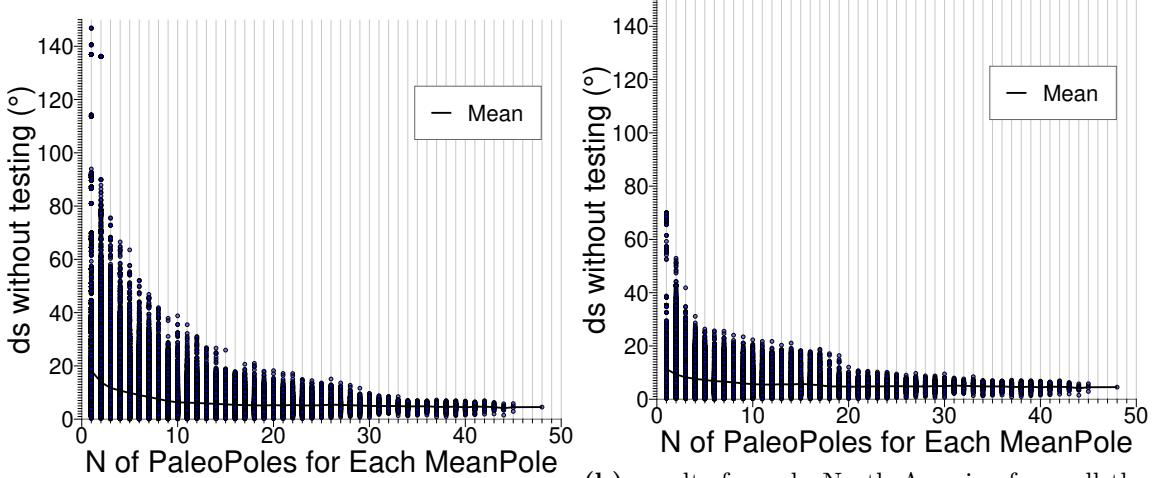
<sup>4</sup> See Figure 23 in Chapter 3

much data density is reduced (Fig. 4.15), although more than 10 paleopoles make the mean less oscillatory (when enough experiments are implemented). So it seems that at least 10 paleopoles are needed for establishing a mean pole that is reasonably close to “true” pole.

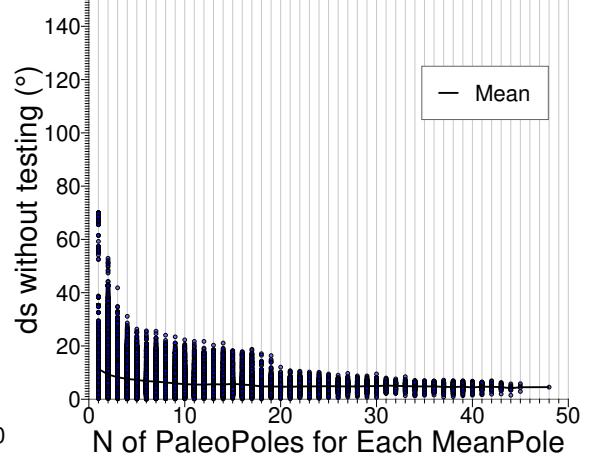
### 4.3 Are the rules we obtained in the last chapter are still true for less data?

First, that if APP is still better, and weighting is still not affecting for less dense paleomagnetic data, is needed to be tested.

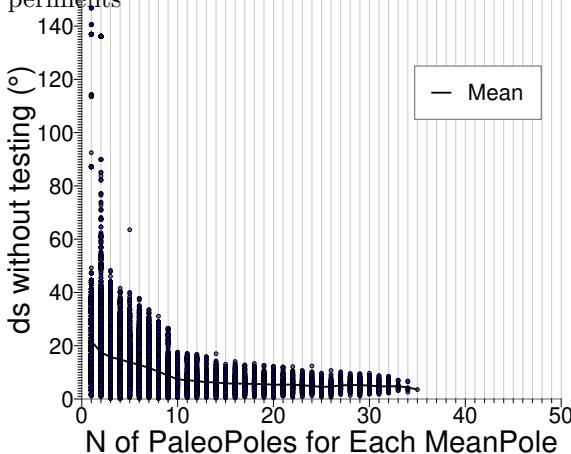
From the perspective of checking if the  $Q_1$ - $Q_3$  intervals overlap, APP is indeed still better than AMP and weighting is indeed still not affecting for more than 15 paleopoles making a 10/5 Myr bin/step APWP (ideally composed of 25 mean poles for 120–0 Ma). For the 20% Indian data case, which contains not more than 15 paleopoles making a 120–0 Ma APWP, there is overlapping between APP’s Mean- $Q_3$  interval and AMP’s  $Q_1$ -Mean interval for weighted cases (i.e. for Weighting No. 1–5). Even so, APP’s means are still lower than AMP’s for this no-more-than-15-paleopoles case (Fig. 4.16). So here is the question: is 15 the threshold?



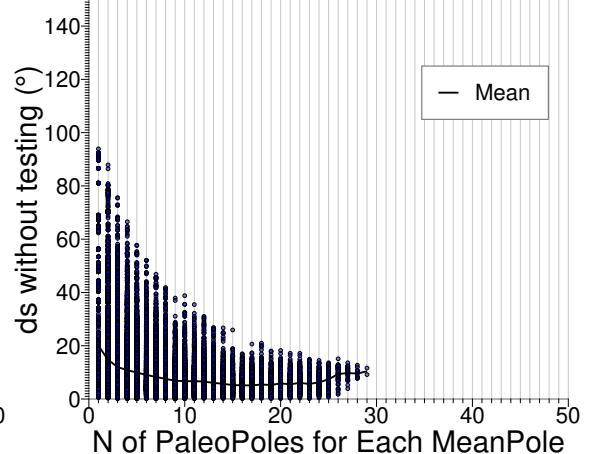
(a) results from all the reduced-data-density experiments



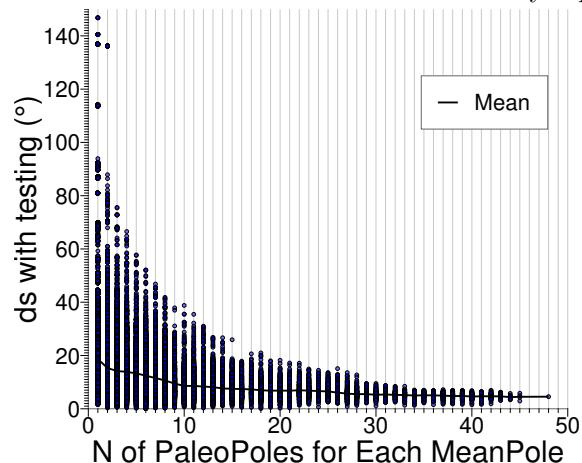
(b) results for only North America from all the reduced-data-density experiments



(c) results for only India from all the reduced-data-density experiments

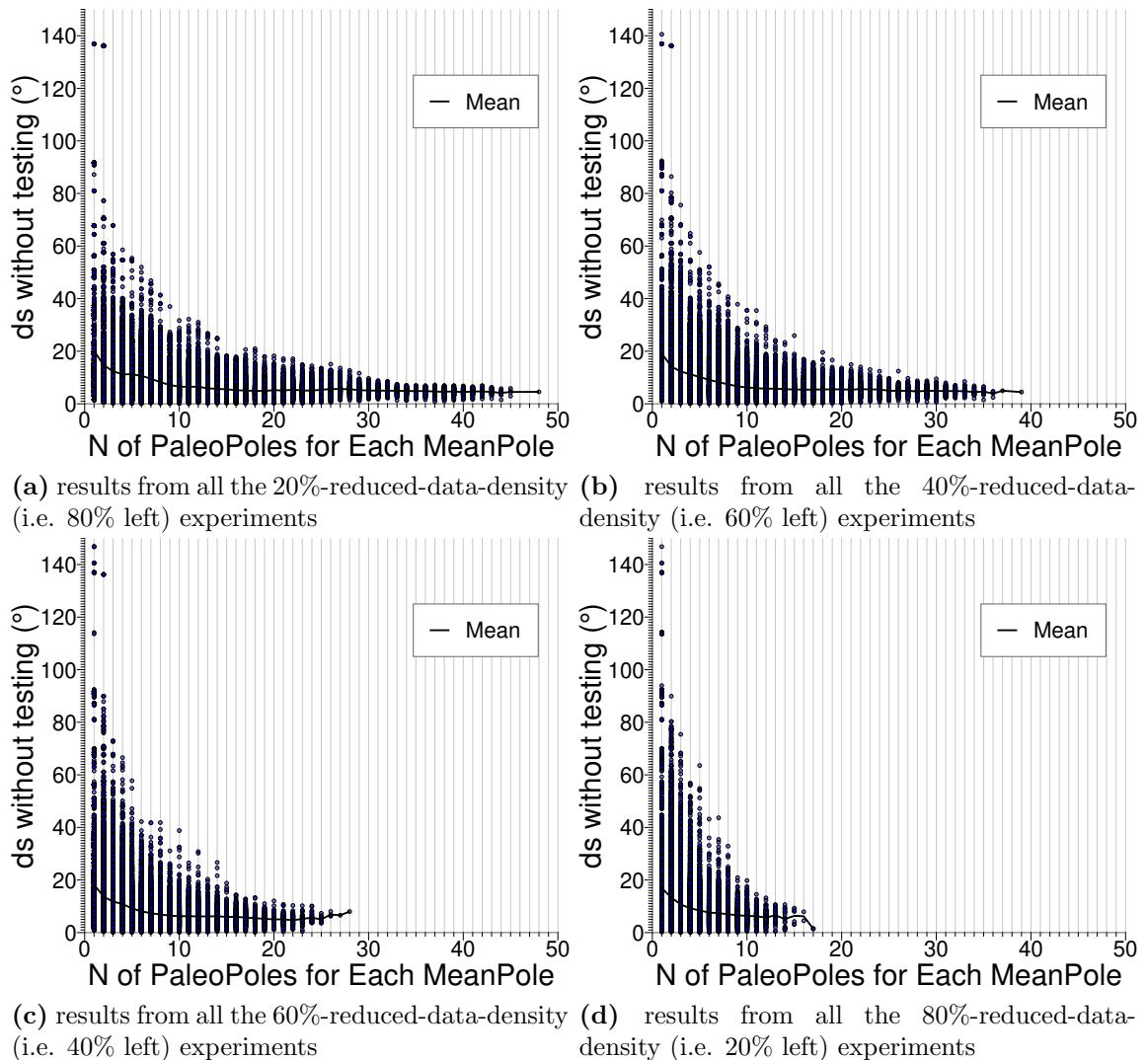


(d) results for only Australia from all the reduced-data-density experiments

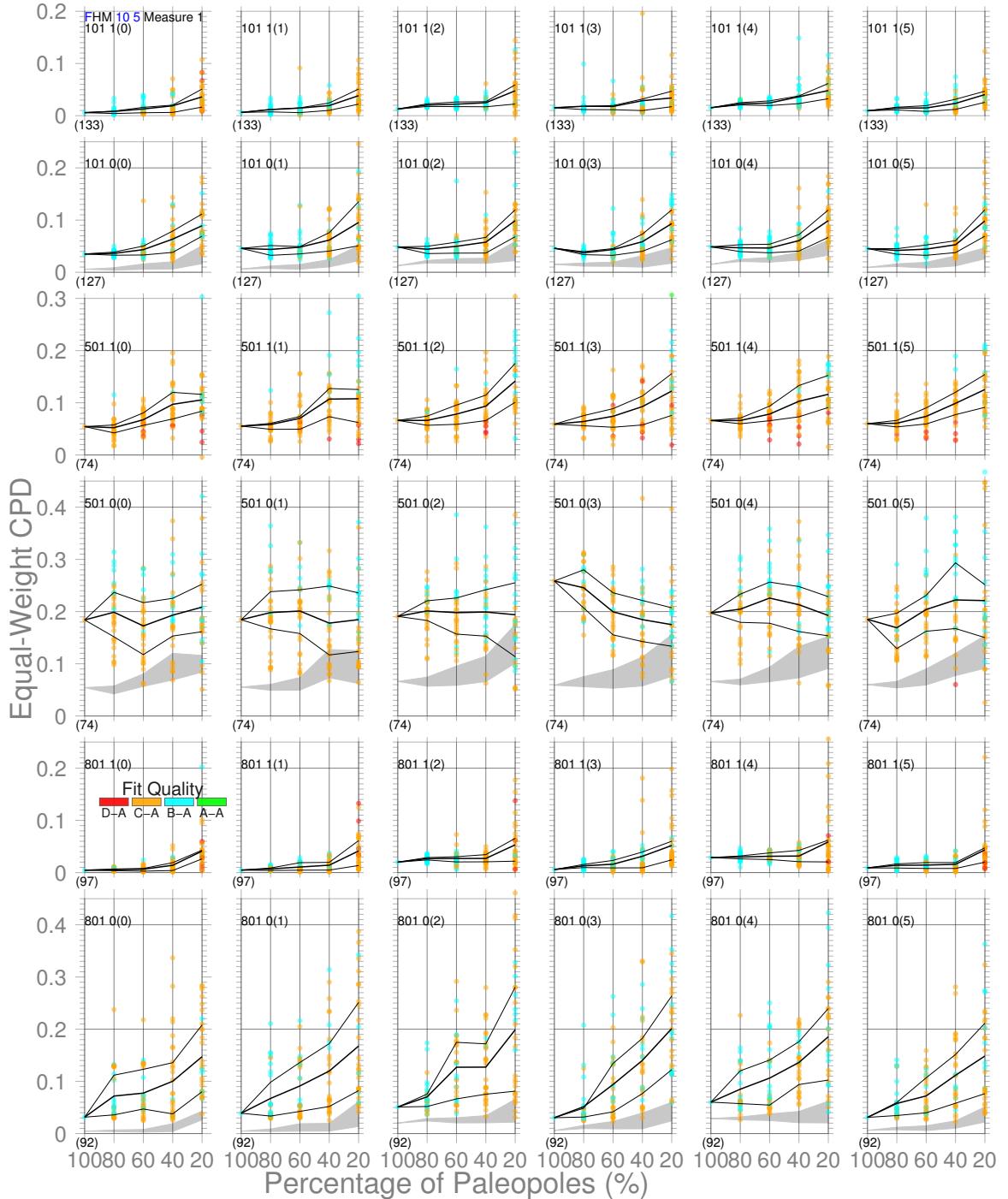


(e) Same as Fig. 4.14a, here only the results having  $d_s$  tested are plotted.

**Figure 4.14:** Relationship between spatial difference ( $d_s$ ) and number of paleopoles for each mean pole from all the above reduced-data-density experiments.



**Figure 4.15:** Relationship between spatial difference ( $d_s$ ) and number of paleopoles for each mean pole from all the above reduced-data-density experiments.



**Figure 4.16:** Comparisons of results from Picking No. 1 (APP) and Picking No. 0 (AMP) with all the listed weighting methods for three continents. The  $Q_1$ - $Q_3$  interquartile range from Picking No. 1 is also shown (shadowed) in the plot of Picking No. 0 for clarity.

# **Chapter 5**

## **Conclusions**

*3–5 pages of summary of results, significance and future directions/work.*

## Supplementary Materials for Chapter 2

### .0.1 Test if A Coeval Pole Pair is Distinguishable with The Bootstrap

Whether a pair of coeval mean poles are statistically distinguishable from each other is investigated, as it can be determined by checking if the confidence intervals of their bootstrap means (based on two poles' uncertainty attributes) overlap [53]. The 95% confidence bounds of the Cartesian coordinates of the bootstrap means are determined and compared. If the poles are distinguishable, the confidence bounds along at least one coordinate axis do not overlap. Otherwise, if the confidence bounds along all the three coordinate axes overlap, the poles are indistinguishable [53]. The actual test used is dependent on the number of paleopoles ( $N$ ) used to calculate the mean pole in the APWP:

- when  $N > 25$  a simple bootstrap [53] generates a pseudo-mean pole from  $N$  directions drawn randomly from the original set of paleopoles. 1000 such simple Bootstraps are implemented here.
- when  $1 < N \leq 25$  a parametric bootstrap [53] generates a pseudo-mean from  $N$  directions drawn from a Fisher distribution with the same  $K$  and  $N$  as the mean pole. 1000 such parametric Bootstraps are implemented here.
- when  $N = 1$ , a pseudo-mean is drawn from a bivariate normal distribution, defined by the properties of the associated A95 uncertainty circle or dm/dp ellipse (see the following Section .0.2). Here 1000 samples are drawn from such a normal distribution.
- if  $N$  is not given, because for example sometimes the pole could be an interpolated result, a negligible A95 like  $0.1^\circ$  or  $0^\circ$  is assigned and the same sampling way as used for the  $N = 1$  case is applied here. This is for the situation when only one of the coeval poles is interpolated, and one would like to keep this pair of poles. Note that if the coeval poles are both interpolated, we suggest directly removing this pair of poles.

**Special cases** Sometimes, like in the cases in Fig. 3 and Fig. 10, we have complete access to the parameters of the mean poles, e.g.  $N$  and precision parameter  $K$ , and also the paleopoles. However this is not necessarily true. If, for instance, we only have access to the path with only its mean poles and spatial uncertainties, we can

keep the way of doing bootstrap sampling consistent for all the mean poles, and just draw bootstrapped means from a bivariate normal distribution based on each spatial uncertainty’s geometry. This is implemented through arbitrarily setting  $N=1$ . The consistency of bootstrap sampling makes it independent of the state of knowledge of the underlying dataset and even the underlying method used to calculate the uncertainty. This means the method can be more generalisable beyond APWPs, because the metrics and the significance testing procedure are more broadly applicable to comparison of other trajectories with associated spatial uncertainties, such as hurricane tracks and bird migration routes.

The final results for each coeval pole pair of all the seven APWP pairs (Fig. 3), are given in the sub-folder “0.result\_tables”, which is contained in the main “data” folder. The results for length and angular differences are listed starting from the rows for the second and third poles respectively, simply because one pole can not compose a APWP segment and at least three poles could constitute an APWP orientation change.

## 0.2 Bivariate Sampling

For some specific poles of the APWP, e.g., only one paleomagnetic pole makes up that “mean”, i.e.,  $N = 1$ , or even there is no paleomagnetic pole in that specific bin (i.e.,  $N = 0$ ) but an interpolated pole that might be given by authors at that specific age, the bivariate normal distribution is used to generate random samples based on its uncertainty ellipse’s semi axes and the major axis’ azimuth, then we use the cumulative distributions of Cartesian coordinates of those random samples to see if the confidence intervals overlap.

However, here the scenario is not a two dimensional (2D) domain, but rather a spherical surface. Directly simulating random points for an ellipse on a sphere is a complicated problem [26]. An analogue approach is proposed here as follows. First, random points of a 2D bivariate normal distribution are generated with NumPy’s random sampling routine “multivariate\_normal” [62]. The lengths of the uncertainty ellipse’s semi-major and semi-minor axes are used as about 1.96 standard deviations of the bivariate normal distribution. The center of the ellipse is located at the intersection of the equator ( $0^\circ$  latitude) and the prime meridian ( $0^\circ$  longitude) with its major axis lying equator-ward (blue point cloud in Fig. 1). Then according to the actual pole coordinates (red star in Fig. 1), an Euler rotation [20] (black star and blue angle arc) can be calculated along the great circle (progressing from blue to red) from the location ( $0^\circ, 0^\circ$ ) to the actual pole location. After those random

points (blue points) are rotated using the same Euler rotation to the new locations (red point cloud in Fig. 1), this elliptical cloud (red point cloud) then is adjusted to its actual azimuth (i.e., the major-axis azimuth of the pole’s uncertainty ellipse; the red dashed line rotated to the yellow dashed line using the red star as the Euler pole shown in Fig. 1).

Note that directly using NumPy’s “random.multivariate\_normal” or “random.normal” routine (2D calculations) and spherical trigonometry to draw random points for an elliptical uncertainty distorts the point cloud out of a bivariate normal distribution, especially at high-latitude areas [40, see the examples given by their Figure 7] and makes the simulation inaccurate. This analogue approach avoids producing declination and inclination vectors beforehand and directly generates random pole vectors, which saves the transformation from declination and inclination to pole and further helps keep us away from the distortion.

### .0.3 Synchronization

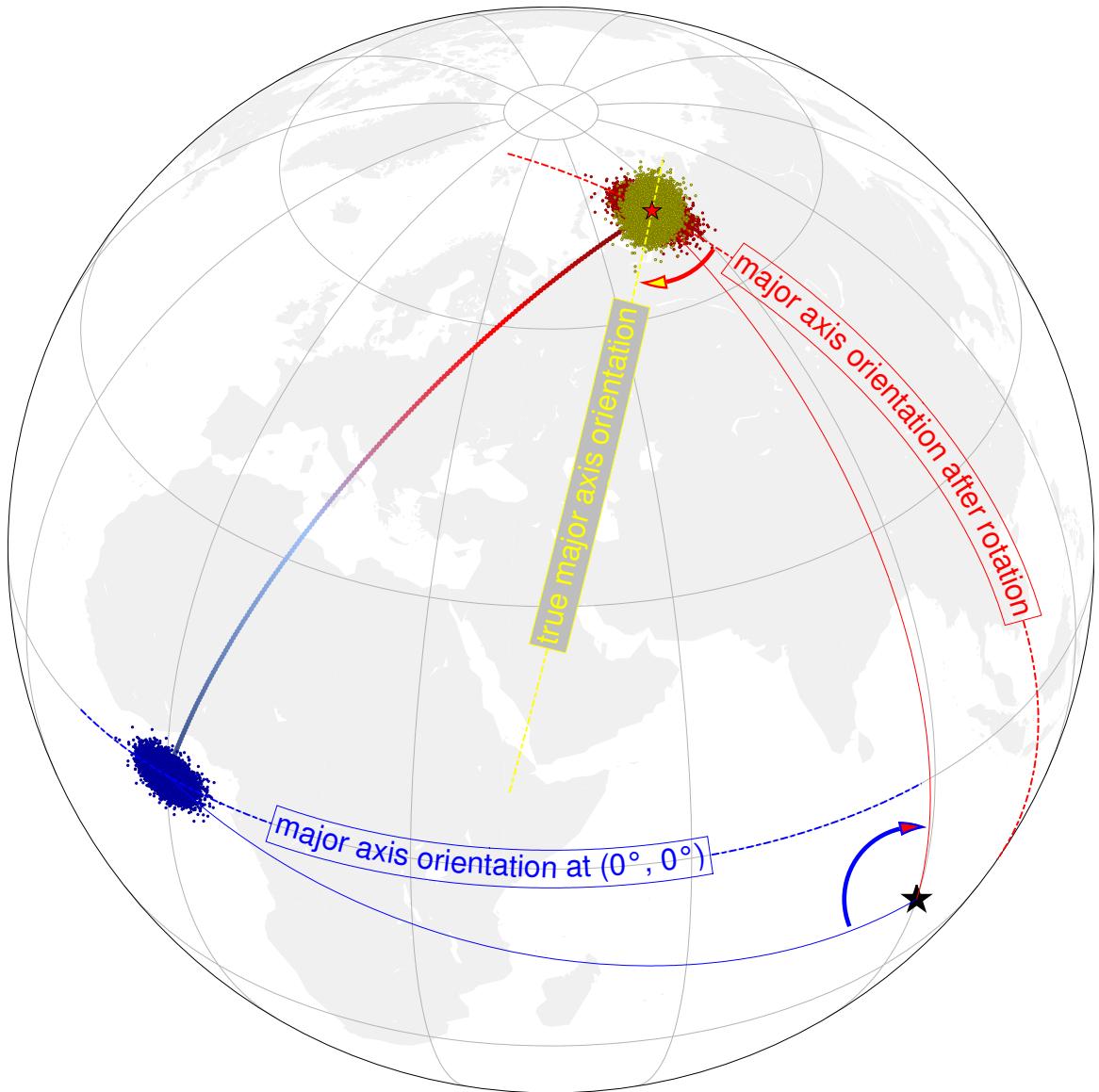
This algorithm is developed for comparing time-synchronized APWPs. In other words, the compared APWPs should have the same timestamps. If the number of their timestamps are different, the unpaired *pole(s)* would be removed to make the timestamps the same before the comparison. APWPs with a pole interpolated for pairing an unpaired pole can be processed by our tool, as we noted earlier, but it is not recommended for a valid analysis. For example, for paleomagnetic APWPs, sometimes there are no paleopoles for a given time window ( $N = 0$ ); sometimes a mean pole is an interpolated result.

#### .0.3.1 Equally Treated Random Weights

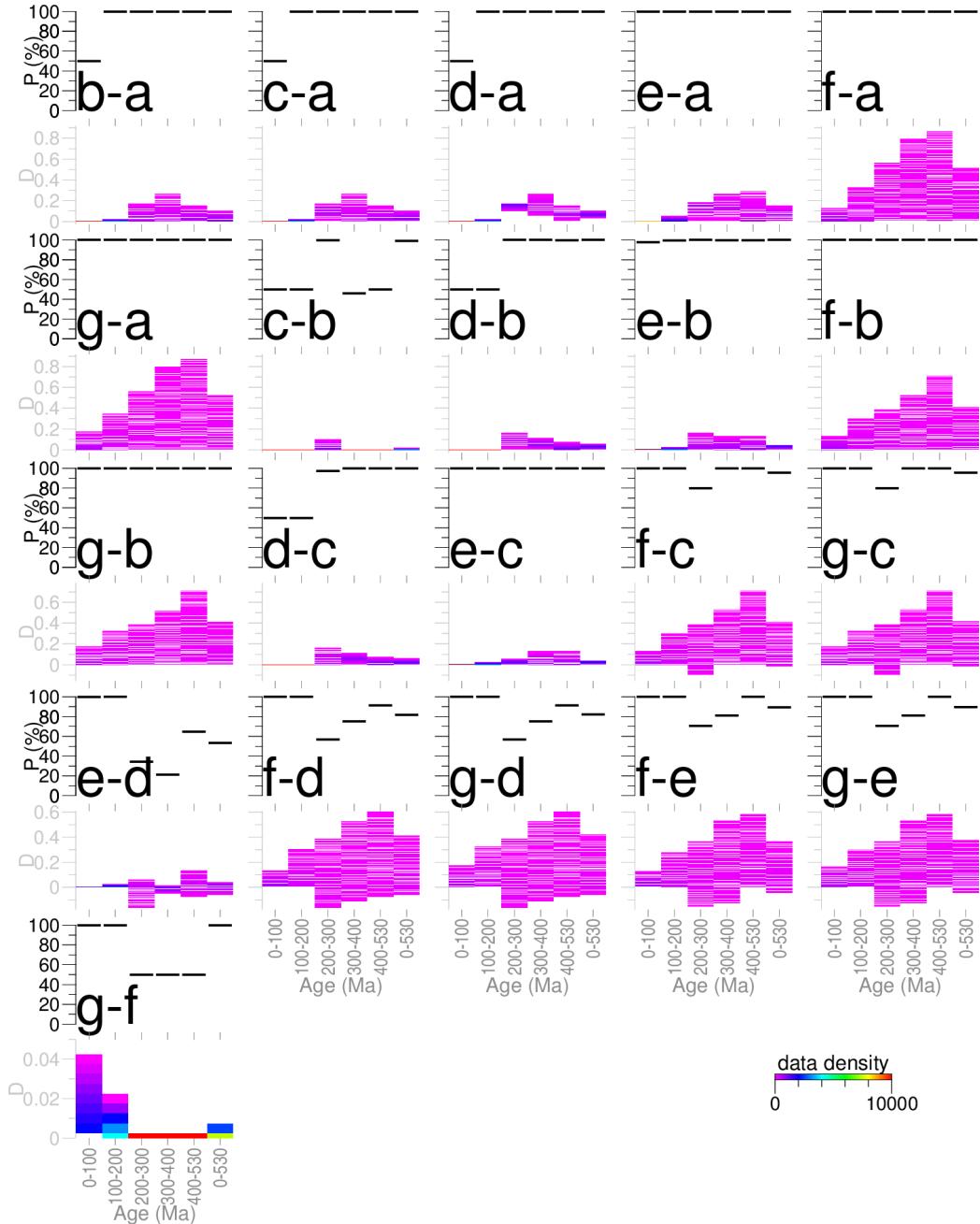
Assigning equally likely (not necessarily equal in value) random values to  $W_s, W_a, W_l$  is also tested. Three uniformly distributed random numbers with a given sum 1 are generated for, for example, 10 000 times here, and then are substituted into the *CPD* formula for deriving the seven APWP pairs’  $D_{full}, D_{0-100Ma}$  etc. to check the possibility that “one pair is superior to the other pair” (Fig. 2).

The full-path results (Fig. 2) again re-verify Order (4) and the results shown in Fig. 8. Although the possibility that *Pair d* is more similar than *Pair e* is not significant (around 50%), the possibility that *Pairs f,g* are more similar than *Pairs c,d,e* is significant (more than 95%),

All the sub-path results (Fig. 2) are explicable using the results shown in Fig. 8. For example, for 0–100 Ma, both *Pair a* and *Pair b* are assigned values of zero for



**Figure 1:** Example of modeling random points for an ellipse uncertainty on the Earth's surface. Sample points (blue) from a bivariate normal distribution centered at the intersection of the equator and the prime meridian are rotated to their new locations (red points) together with the uncertainty ellipse center (i.e. the  $0^\circ$  longitude  $0^\circ$  latitude point prior to the rotation) exactly rotated to its actual pole coordinate (red star), then adjusted to the true orientation (yellow dashed line).



**Figure 2:** Differences of CPDs between *Pair a*, *Pair b*, *Pair c*, *Pair d*, *Pair e*, *Pair f* and *Pair g*, when 10 000 sets of three uniformly random weights (with their sum 1) are applied. If the difference  $D$  is positive, the subtrahend pair ranks higher in similarity, and if it is negative, the minuend pair ranks higher. The  $y$  axis in each upper plot is for the percentage  $P$  that the subtrahend pair owns higher similarity.

all the three metrics  $d_s^{0-100Ma}$ ,  $d_l^{0-100Ma}$  and  $d_a^{0-100Ma}$  (Figures 8b, 8d and 8f), which means they are always undifferentiated.

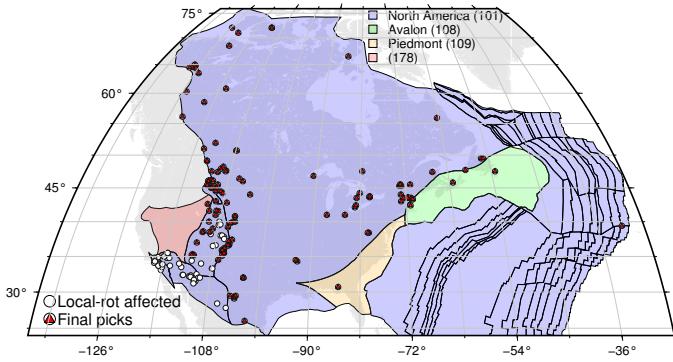
## Supplementary Materials for Chapter 3: Constrain Paleopoles For A Certain Tectonic Plate

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 [25] 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.

### .0.4 120–0 Ma North America

The data-constraining polygons are from the recently published plate model [66] (Fig. 3). Plate ID 101 polygon in the recently published Plate Model [66], 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 [66], 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 [10]. Then in order to be compared with the FHM (120–0 Ma) [35, 34], 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 [33] 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. 3).

Also based on this model of southwestern North America since 36 Ma [33], part of the paleopoles constrained by the four small western terranes whose Plate IDs are also 101 (white circles in Fig. 3) 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 [33]) should be included. So finally 135



**Figure 3:** 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.

of the 193 datasets remain (Fig. 3). Spatially North American paleomagnetic data are mainly from the western and eastern margins of the plate.

### .0.5 120–0 Ma India

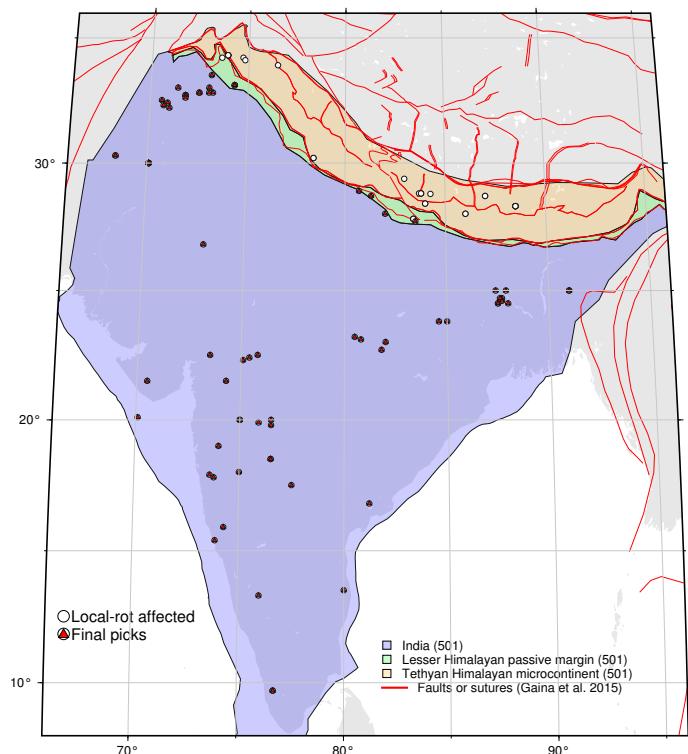
Plate ID 501 polygons in the recently published Plate Model [66] also include the two small polygons of the northern “Lesser Himalayan passive margin of Greater Indian Basin” and “Tethyan Himalayan microcontinent of Greater India” (Fig. 4). The polygons are used to select the sampling sites of the paleopoles for India (Fig. 4).

Based on the model of the tectonic interactions between India, Arabia and Asia since the Jurassic [17] (Fig. 4), 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. 4). Spatially Indian paleomagnetic data are more evenly distributed on the India plate than North American and Australian poles.

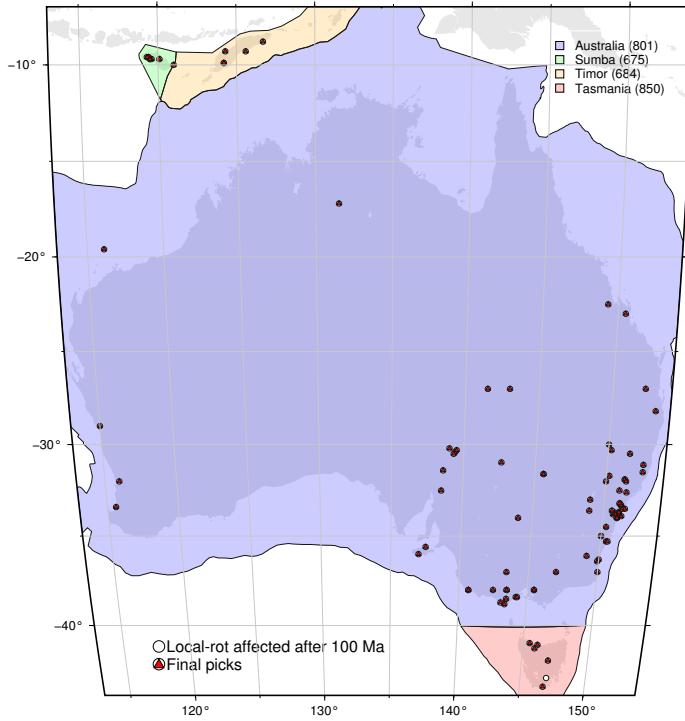
### .0.6 120–0 Ma Australia

Plate ID 801 polygon in the recently published Plate Model [66], including its children 675 (Sumba block) and 684 (Timor block) polygons for 120–0 Ma (Fig. 5), is used to select the sampling sites of the paleopoles for Australia. According to the plate model rotation data [66], 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. 5), there is a triangle-shaped small polygon 850 (Tasmania block) which is fixed to 801 since c.100 Ma according to the [66] rotation data. With that attribute, 805 contributes more



**Figure 4:** 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.



**Figure 5:** 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.

data younger than c.100Ma for the later analysis. Ultimately the final 99 extracted datasets is shown in Fig. 5.

# Bibliography

- [1] Steinberger B and Torsvik T H. Absolute plate motions and true polar wander in the absence of hotspot tracks. *Nature*, 452(7187):620–623, 2008.
- [2] M Beaman, W W Sager, G D Acton, L Lanci, and J Pares. Improved late Cretaceous and early Cenozoic paleomagnetic apparent polar wander path for the Pacific plate. *Earth Planet Sci Lett*, 262(1-2):1–20, 2007.
- [3] J Besse and V Courtillot. Apparent and true polar wander and the geometry of the geomagnetic field over the last 200 myr. *J Geophys Res*, 107(B11):2300, 2002.
- [4] D Bilardello. The do’s and don’ts of inclination shallowing corrections. *Institute for Rock Magnetism Quarterly*, 26(3):1,9–11, 2016.
- [5] D Bilardello and K P Kodama. Rock magnetic evidence for inclination shallowing in the early Carboniferous Deer Lake Group red beds of western Newfoundland. *Geophys J Int*, 181(1):275–289, 2010.
- [6] E Bullard, F.R.S., J E Everett, and G Smith. The fit of the continents around the Atlantic. *Phil Trans Roy Soc Lond Math Phys Sci*, 258:41–51, 1965.
- [7] R F Butler. *Paleomagnetism: Magnetic Domains to Geologic Terranes*. Blackwell Scientific Publications, Malden, Massachusetts, electronic edition, 1992.
- [8] DeMets C, Gordon R G, and Argus D F. Geologically current plate motions. *Geophys J Int*, 181(1):1–80, 2010.
- [9] M T Chandler, P Wessel, and W W Sager. Analysis of Ontong Java Plateau palaeolatitudes: evidence for large-scale rotation since 123 Ma? *Geophys J Int*, 194(1):18–29, 2013.

- [10] G L Christeson, H J A Van Avendonk, I O Norton, J W Snedden, D R Eddy, G D Karner, and C A Johnson. Deep crustal structure in the eastern Gulf of Mexico. *J Geophys Res Solid Earth*, 119(9):6782–6801, 2014.
- [11] R Chu, W Leng, D V Helmberger, and M Gurnis. Hidden hotspot track beneath the eastern United States. *Nature Geosci*, 6(11):963–966, 2013.
- [12] M Domeier, R van der Voo, and F B Denny. 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(1-2):38–52, 2011.
- [13] Mathew Domeier, Rob Van der Voo, Renata N. Tomezzoli, Eric Tohver, Bart W. H. Hendriks, Trond H. Torsvik, Haroldo Vizan, and Ada Dominguez. 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):1–26, 2011.
- [14] David A. D. Evans. Proterozoic low orbital obliquity and axial-dipolar geomagnetic field from evaporite palaeolatitudes. *Nature*, 444(7115):51–55, 2006.
- [15] R A Fisher. Dispersion on a sphere. *Proc Roy Soc London Ser A*, 217:295–305, 1953.
- [16] C Gaina, T H Torsvik, D J J van Hinsbergen, S Medvedev, S C Werner, and C Labails. The African Plate: A history of oceanic crust accretion and subduction since the Jurassic. *Tectonophysics*, 604:4–25, 2013. Progress in understanding the South Atlantic margins.
- [17] C Gaina, D J J van Hinsbergen, and W Spakman. Tectonic interactions between India and Arabia since the Jurassic reconstructed from marine geophysics, ophiolite geology, and seismic tomography. *Tectonics*, 34(5):875–906, 2015.
- [18] Pitambar Gautam and Erwin Appel. Magnetic-polarity stratigraphy of Siwalik Group sediments of Tinau Khola section in west central Nepal, revisited. *Geophysical Journal International*, 117(1):223–234, 04 1994.

- [19] S E Geuna, L D Escosteguy, and C O Limarino. Paleomagnetism of the Carboniferous-Permian Patquia Formation, Paganzo basin, Argentina: implications for the apparent polar wander path for South America and Gondwana during the Late Palaeozoic. *Geol Acta*, 8(4):373–397, 2010.
- [20] B Greiner. Euler rotations in plate-tectonic reconstructions. *Comput Geosci*, 25(3):209–216, 1999.
- [21] S J Hellinger. The uncertainties of finite rotations in plate-tectonics. *J Geophys Res*, 86(NB10):9312–9318, 1981.
- [22] J. K. Hillier. Pacific seamount volcanism in space and time. *Geophysical Journal International*, 168(2):877–889, 02 2007.
- [23] R V Hogg, Tanis E, and Zimmerman D. *Probability and Statistical Inference*. Pearson, New York, NY, 10 edition, 2019.
- [24] Benjamin C. Horner-Johnson, Richard G. Gordon, Sara M. Cowles, and Donald F. Argus. The angular velocity of Nubia relative to Somalia and the location of the Nubia–Somalia–Antarctica triple junction. *Geophysical Journal International*, 162(1):221–238, 07 2005.
- [25] E H Jacox and H Samet. Spatial join techniques. *ACM Trans. Database Syst.*, 32:7, 2007.
- [26] J T Kent. The Fisher-Bingham distribution on the sphere. *J Roy Stat Soc B*, 44(1):71–80, 1982.
- [27] R F King. The remanent magnetism of artificially deposited sediments. *Geophys Suppl Mon Not Roy Astron Soc Lett*, 7(3):115–134, 1955.
- [28] Wout Krijgsman and Lisa Tauxe. Shallow bias in mediterranean paleomagnetic directions caused by inclination error. *Earth and Planetary Science Letters*, 222(2):685–695, 2004.
- [29] E V Kulakov, A V Smirnov, and J F Diehl. Paleomagnetism of the ∼1.1 Ga Coldwell Complex (Ontario, Canada): Implications for Proterozoic geomagnetic field morphology and plate velocities. *J Geophys Res Solid Earth*, 119(12):8633–8654, 2014.

- [30] M W McElhinny and J Lock. Iaga paleomagnetic databases with access. *Surveys in Geophysics*, 17(5):575–591, 1996.
- [31] P L McFadden and M W McElhinny. Classification of the reversal test in palaeomagnetism. *Geophysical Journal International*, 103(3):725–729, 12 1990.
- [32] Dan McKenzie and John G. Sclater. The evolution of the indian ocean since the late cretaceous. *Geophysical Journal of the Royal Astronomical Society*, 24(5):437–528, 1971.
- [33] N McQuarrie and B P Wernicke. An animated tectonic reconstruction of south-western North America since 36 Ma. *Geosphere*, 1(3):147–172, 2006.
- [34] R D Müller, J Y Royer, S C Cande, W R Roest, and S Maschenkov. *New constraints on the Late Cretaceous/Tertiary plate tectonic evolution of the Caribbean*, volume 4, Chap. 2, pages 33–59. Elsevier, 1999.
- [35] R D Müller, J Y Royer, and L A Lawver. Revised plate motions relative to the hotspots from combined Atlantic and Indian-Ocean hotspot tracks. *Geology*, 21(3):275–278, 1993.
- [36] R D Müller, M Sdrolias, C Gaina, and W R Roest. Age, spreading rates, and spreading asymmetry of the world’s ocean crust. *Geochem Geophys Geosyst*, 9(4):Q04006, 2008.
- [37] R D Müller, M Seton, S Zahirovic, S E Williams, K J Matthews, N M Wright, G E Shephard, K Maloney, N Barnett-Moore, M Hosseinpour, D J Bower, and J Cannon. Ocean basin evolution and global-scale plate reorganization events since Pangea breakup. *Annu Rev Earth Planet Sci*, 44(1):107–138, 2016.
- [38] C O’Neill, R D Müller, and B Steinberger. On the uncertainties in hot spot reconstructions and the significance of moving hot spot reference frames. *Geochem Geophys Geosyst*, 6(4):Q04003, 2005.
- [39] N.D. Opdyke, N.M. Johnson, G.D. Johnson, E.H. Lindsay, and R.A.K. Tahirkheli. Paleomagnetism of the middle siwalik formations of northern pakistan and rotation of the salt range decollement. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 37(1):1–15, 1982. The geochronology and biochronology of the Siwalik Group, Pakistan.

- [40] E. Pakyuz-Charrier, M. Lindsay, V. Ogarko, J. Giraud, and M. Jessell. Monte Carlo simulation for uncertainty estimation on structural data in implicit 3-D geological modeling, a guide for disturbance distribution selection and parameterization. *Solid Earth*, 9(2):385–402, 2018.
- [41] S A Pisarevsky. New edition of the Global Paleomagnetic Database. *Eos Trans AGU*, 86(17):170, 2005.
- [42] S A Pisarevsky and M W McElhinny. Global Paleomagnetic Data Base developed into its visual form. *Eos Trans AGU*, 84(20):192, 2003.
- [43] C J Rowan and D B Rowley. Preserved history of global mean spreading rate: 83 Ma to present. *Geophysical Journal International*, 208(2):1173–1183, 07 2016.
- [44] W W Sager. Divergence between paleomagnetic and hotspot-model-predicted polar wander for the Pacific plate with implications for hotspot fixity. *GSA Spec Paper*, 430:335–357, 2007.
- [45] A Schettino and C R Scotese. 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(2):727–759, 2005.
- [46] M Seton, R D Müller, S Zahirovic, C Gaina, T Torsvik, G Shephard, A Talsma, M Gurnis, M Turner, S Maus, and M Chandler. Global continental and ocean basin reconstructions since 200 Ma. *Earth Sci Rev*, 113(3-4):212–270, 2012.
- [47] G E Shephard, H P Bunge, B S A Schuberth, R D Müller, A S Talsma, C Moder, and T C W Landgrebe. Testing absolute plate reference frames and the implications for the generation of geodynamic mantle heterogeneity structure. *Earth Planet Sci Lett*, 317-318(0):204–217, 2012.
- [48] N L Swanson-Hysell, A C Maloof, B P Weiss, and D A D Evans. No asymmetry in geomagnetic reversals recorded by 1.1-billion-year-old Keweenawan basalts. *Nature Geosci*, 2(10):713–717, 2009.
- [49] J Tarduno, H-P Bunge, N Sleep, and U Hansen. The bent Hawaiian-Emperor hotspot track: Inheriting the mantle wind. *Science*, 324(5923):50–53, 2009.
- [50] J A Tarduno. On the motion of Hawaii and other mantle plumes. *Chem Geol*, 241(3-4):234–247, 2007.

- [51] L Tauxe, S K Banerjee, R F Butler, and R van der Voo. Essentials of paleomagnetism: Fifth web edition. <http://earthref.org/MAGIC/books/Tauxe/Essentials/>, 2019.
- [52] L Tauxe and D V Kent. *A Simplified Statistical Model for the Geomagnetic Field and the Detection of Shallow Bias in Paleomagnetic Inclinations: was the Ancient Magnetic Field Dipolar?*, pages 101–115. American Geophysical Union (AGU), 2004.
- [53] L Tauxe, N Kylstra, and C Constable. Bootstrap statistics for paleomagnetic data. *J Geophys Res Solid Earth*, 96(B7):11723–11740, 1991.
- [54] L. Tauxe, R. Shaar, L. Jonestrask, N. L. Swanson-Hysell, R. Minnett, A. A. P. Koppers, C. G. Constable, N. Jarboe, K. Gaastra, and L. Fairchild. Pmagpy: Software package for paleomagnetic data analysis and a bridge to the magnetics information consortium (magic) database. *Geochemistry, Geophysics, Geosystems*, 17(6):2450–2463, 2016.
- [55] T H Torsvik and L R M Cocks. *Earth History and Palaeogeography*. Cambridge University Press, Cambridge, 2016.
- [56] T H Torsvik, R D Müller, R van der Voo, B Steinberger, and C Gaina. Global plate motion frames: Toward a unified model. *Rev Geophys*, 46(3):RG3004, 2008.
- [57] T H Torsvik and M A Smethurst. Plate tectonic modelling: virtual reality with GMAP. *Comput Geosci*, 25(4):395–402, 1999.
- [58] T H Torsvik, M A Smethurst, R van der Voo, A Trench, N Abrahamsen, and E Halvorsen. Baltica. A synopsis of Vendian-Permian palaeomagnetic data and their palaeotectonic implications. *Earth Sci Rev*, 33(2):133–152, 1992.
- [59] T H Torsvik, R van der Voo, U Preeden, C Mac Niocaill, B Steinberger, P V Doubrovine, D J J van Hinsbergen, M Domeier, C Gaina, E Tohver, J G Meert, P J A McCausland, and L R M Cocks. Phanerozoic polar wander, palaeogeography and dynamics. *Earth Sci Rev*, 114(3-4):325–368, 2012.
- [60] D G van der Meer, W Spakman, D J J van Hinsbergen, M L Amaru, and T H Torsvik. Towards absolute plate motions constrained by lower-mantle slab remnants. *Nat Geosci*, 3(1):36–40, 2010.

- [61] R van der Voo. The reliability of paleomagnetic data. *Tectonophysics*, 184(1):1–9, 1990.
- [62] S van der Walt, S C Colbert, and G Varoquaux. The NumPy array: A structure for efficient numerical computation. *Comput Sci Eng*, 13(2):22–30, 2011.
- [63] T Veikkolainen, L Pesonen, and D D Evans. PALEOMAGIA: A PHP/MYSQL database of the Precambrian paleomagnetic data. *Studia Geophysica et Geodaetica*, pages 1–17, 2014.
- [64] P Wessel, R D Müller, and G Schubert. *Plate Tectonics*, book section 6.02, pages 49–98. Elsevier, Amsterdam, 2007.
- [65] P Wessel, W H F Smith, R Scharroo, J Luis, and F Wobbe. Generic Mapping Tools: Improved version released. *Eos Trans AGU*, 94(45):409–410, 2013.
- [66] A Young, N Flament, K Maloney, S Williams, K Matthews, S Zahirovic, and R D Müller. Global kinematics of tectonic plates and subduction zones since the late Paleozoic Era. *Geoscience Frontiers*, 10(3):989–1013, 2019. Special Issue: Advances in Himalayan Tectonics.