

# **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>
<b>4</b>	<b>How Much Data Needed to Make a Reliable APWP</b>	<b>44</b>
4.1	Reference Path . . . . .	45
4.2	Extraction Fraction . . . . .	45
4.2.1	Number of Samples . . . . .	51
4.2.2	Extreme Value Study: Suggestions on Algorithm, especially on large uncertainties . . . . .	51
4.2.3	Relationship between $d_s$ Score and Number of Paleopoles for Paleomagnetic APWP . . . . .	55
4.3	Are the rules we obtained in the last chapter are still true for less data? . . . . .	58
<b>5</b>	<b>Conclusions</b>	<b>62</b>
.0.1	Test if A Coeval Pole Pair is Distinguishable with The Bootstrap	63
.0.2	Bivariate Sampling . . . . .	64
.0.3	Synchronization . . . . .	65
.0.3.1	Equally Treated Random Weights . . . . .	65
<b>Bibliography</b>		<b>69</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, 36, 40]. 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 [29] and PALEOMAGIA [46]. . . . .	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
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. . . . .	46
4.2 Same as Fig. 4.1. Here color demonstrates the resolution of pole pairs with the path comparisons. . . . .	47
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. . . . .	48
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. . . . .	49
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. . . . .	50
4.6 Testing differences of results from different numbers of samples. See Fig. 4.1 for more details. . . . .	51
4.7 Less data, better similarity? . . . . .	52
4.8 Less data, better similarity? . . . . .	53
4.9 Less data, better similarity? . . . . .	54
4.10 Less data, better similarity? . . . . .	54
4.11 Less data, better similarity? . . . . .	55
4.12 Less data, better similarity? . . . . .	56
4.13 Less data, better similarity? . . . . .	56
4.14 . . . . .	59

4.15	60
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.	61
1 How $N_j=1$ uncertainty ellipse is simulated	66
2 Comparisons of Pairs a-g with random weights involved	67

# **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, 47]. 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 [22], which minimizes the misfit area between two isochrons. The *Hellinger* method [16] 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 [47] 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 [25].

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” [32]. 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” [24, 23], 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; [26]) and dating can be difficult (e.g. diffuse volcanic centers possibly related to large diameter plume conduits could cause the existence of time reversals; [26]). 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; [17]). 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 [9]. Last, but not least, hotspots can be susceptible to drift that may be caused by changes in sub-lithospheric mantle flow [34]. 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 [26, 35]. To overcome this source of error, the “Moving hotspot model” [26] 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 [24] with a moving hotspot frame from c. 132–100 Ma [26] (Hybrid hotspot model [32]), combining a moving hotspot frame from 100–0 Ma [26] 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) [40] from 140–100 Ma (Hybrid paleomagnetic model [32]), and combining a moving hotspot frame from 120–100 Ma [26] with a True Polar Wander (TPW) corrected paleomagnetic model [1] from 100–0 Ma (Hybrid TPW-corrected model [32]).

Recently another absolute reference frame “Subduction reference model” [43] 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

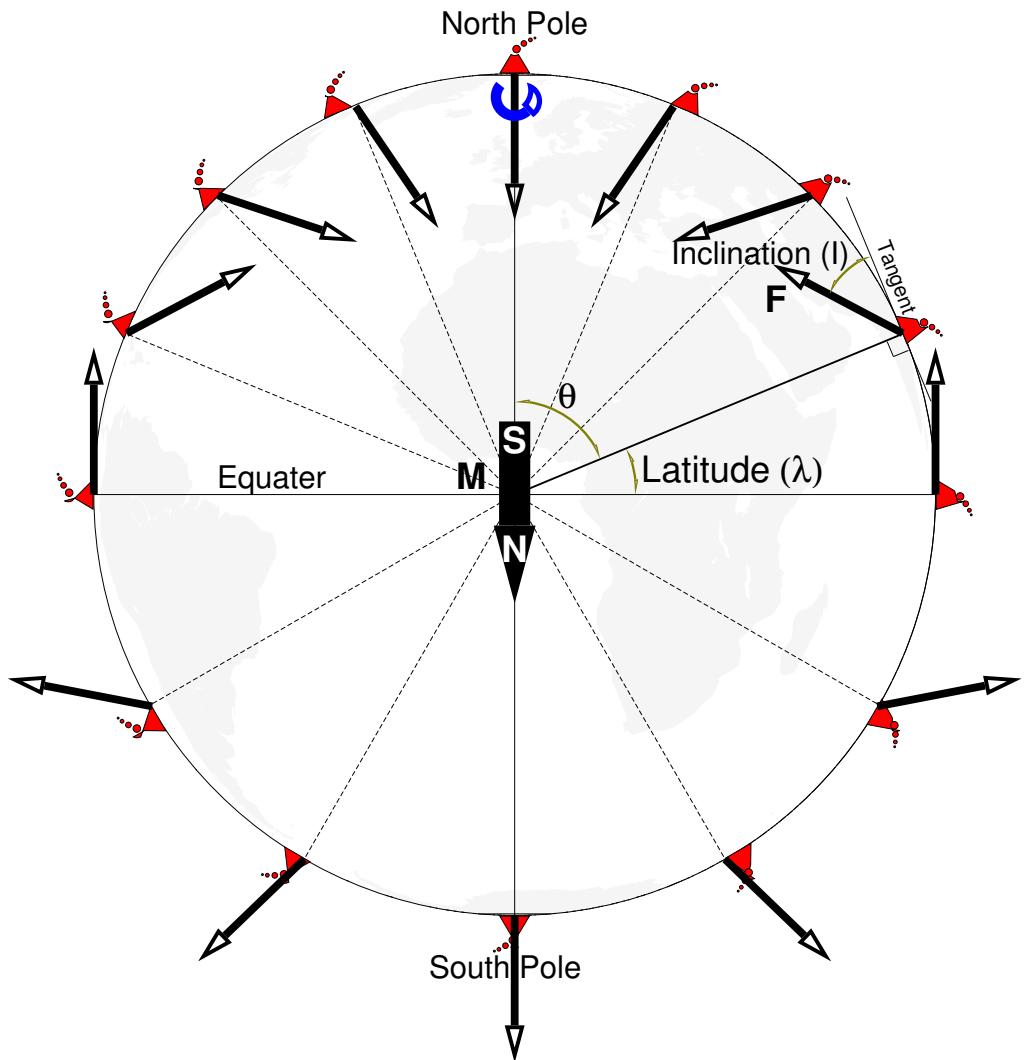
“Subduction reference model” allows for reconstructions between about 260 Ma and 140 Ma, older than the other absolute models can predict, the model is strongly dependent on the vertical subduction assumption and resolution of seismic tomography models, so its uncertainty is high. Above all, importantly, 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.

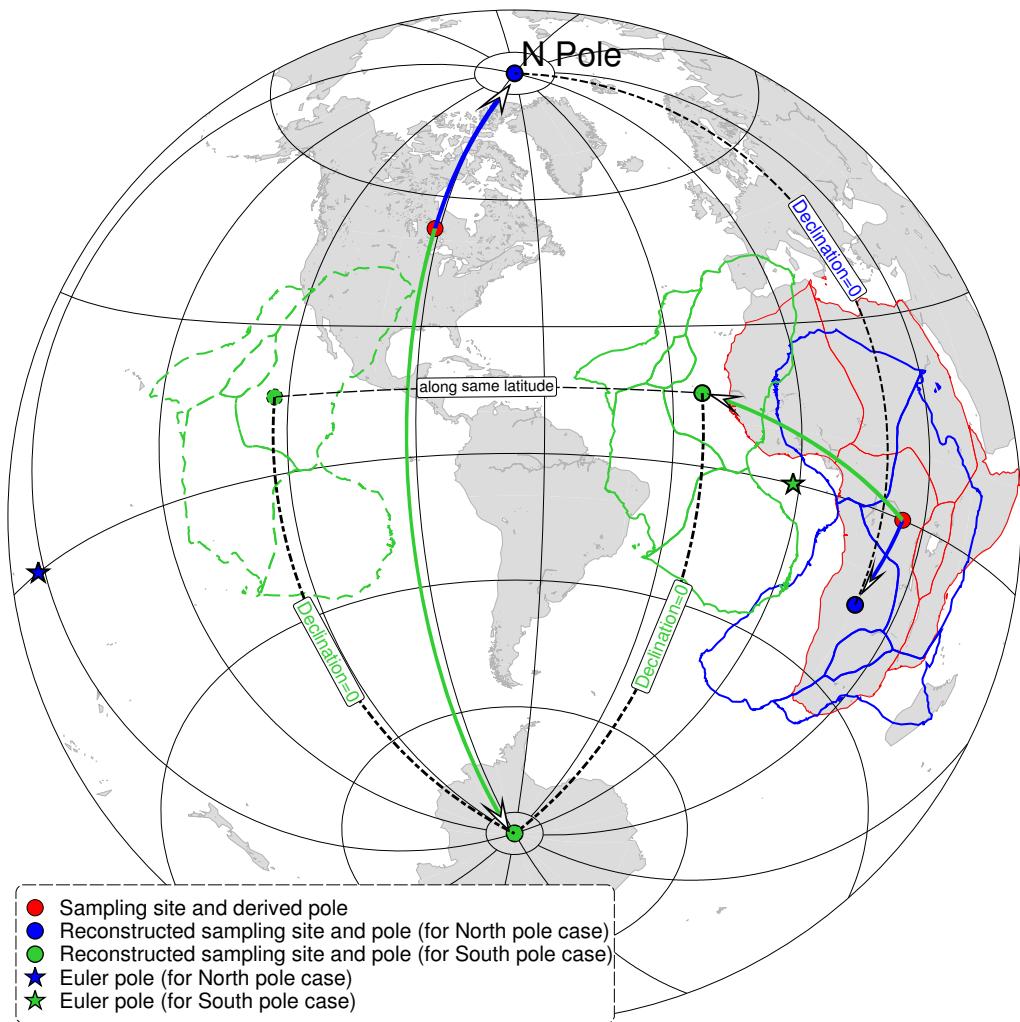
Magnetic remanence is the magnetization left behind in a ferromagnetic substance in the absence of an external magnetic field [36]. The remanent magnetisation of rocks can preserve the direction and intensity of the geomagnetic field when the rock was formed, e.g. in the process of cooling, ferromagnetic 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, 36, 40]. 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 [36]. 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 [12] 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 [40]. 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 [36]. 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).

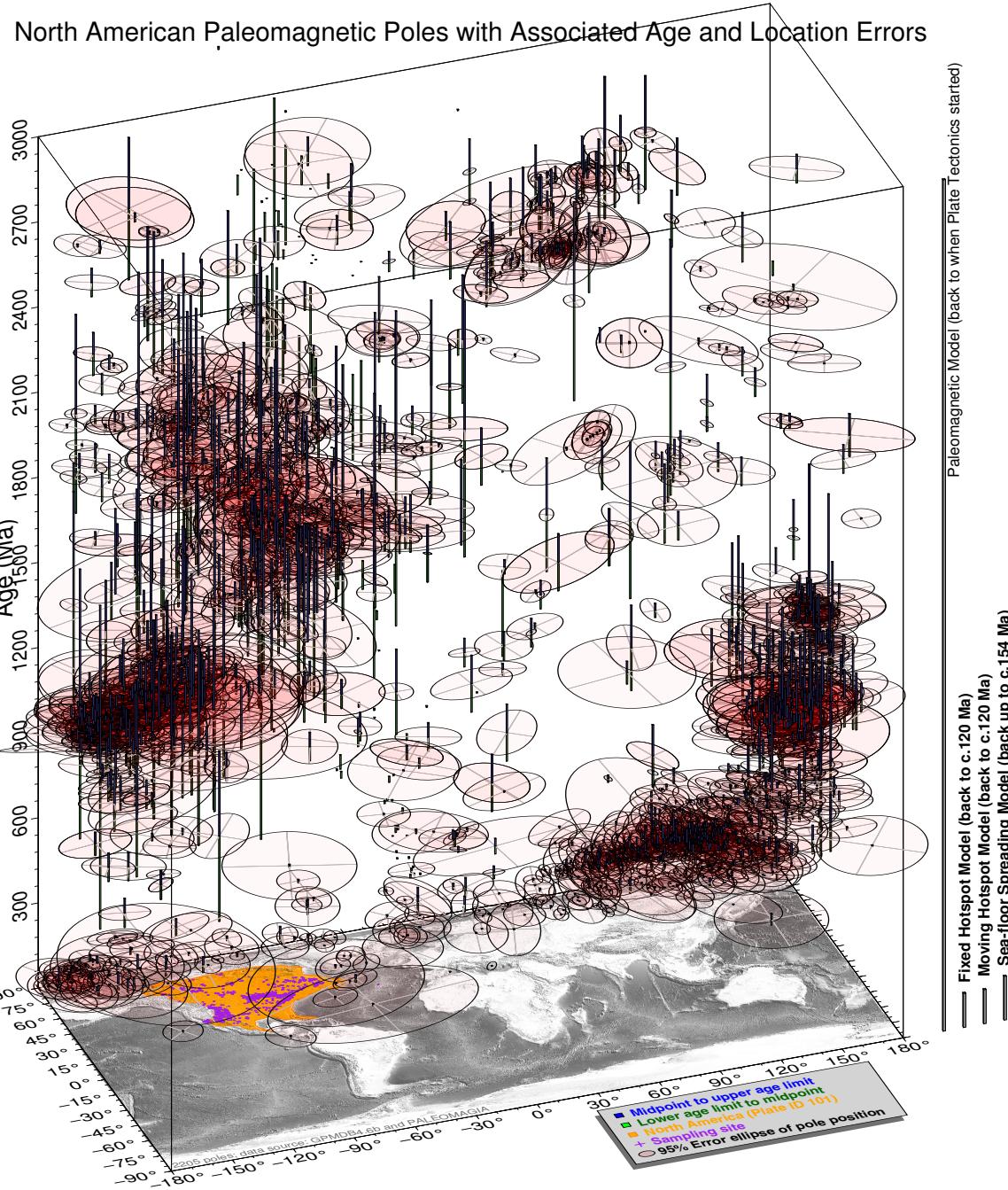
Euler rotation of pole ( $0^\circ$ ,  $1.45^\circ\text{W}$ ) and angle  $142.59^\circ$  rotates the sampling site ( $0^\circ$ ,  $25^\circ\text{E}$ ) on Africa to ( $15.6998^\circ\text{N}$ ,  $23.0121^\circ\text{W}$ ) indicating southward motion since the Late Jurassic. Second, because in a GAD field 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 [29, updated in 2016 by the Ivar Giaevers 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> [30] and CESRE's Paleomagnetism and Rock Magnetism project <https://wiki.csiro.au/display/cmfr/Palaeo> which is later updated by Ivar Giaevers 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 [40]. 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 [29] and PALEOMAGIA [46].

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.

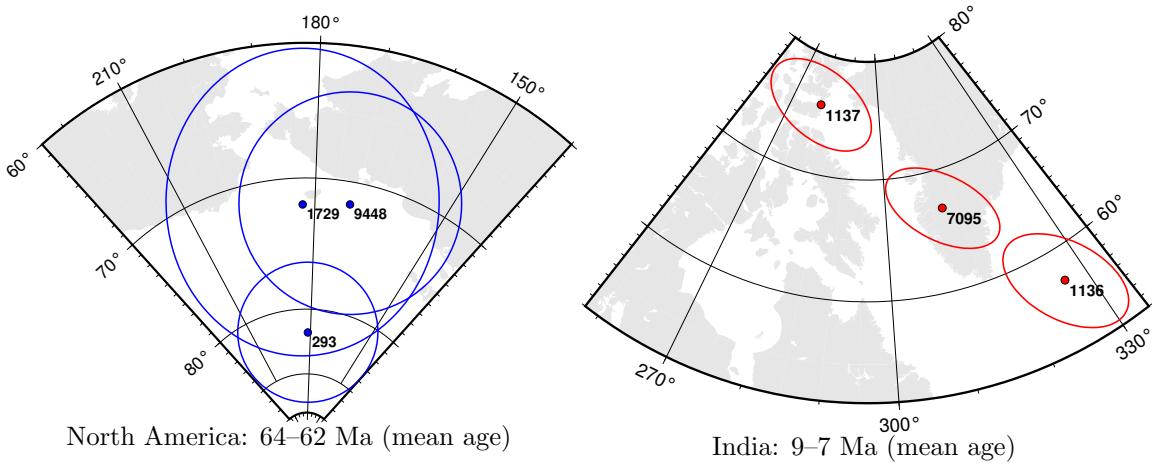
#### **1.1.4.1 Age Error**

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 [36]. 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 [27], 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 [13].

**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 [27, 13], e.g. pole 7095 has a signal which suggests the presence of a counter-clockwise local rotation of the Tinau Khola section [13], 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 [36, 44, 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 [44] 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 [12] 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 [41]. 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 [12] 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

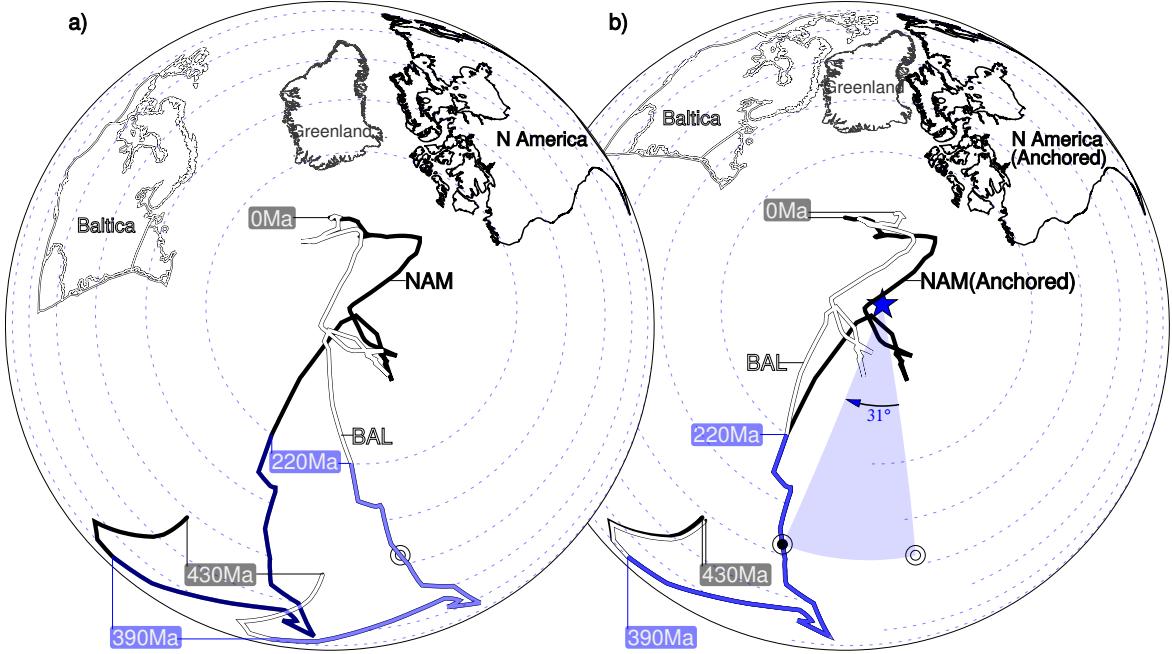
*This chapter mainly describes the development of a new APW path similarity measuring tool used throughout the dissertation. Apparent polar wander paths (APWPs) based on paleomagnetic data are the principal means of describing plate motions through most of Earth history. Comparing the spatio-temporal 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 [11], 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.

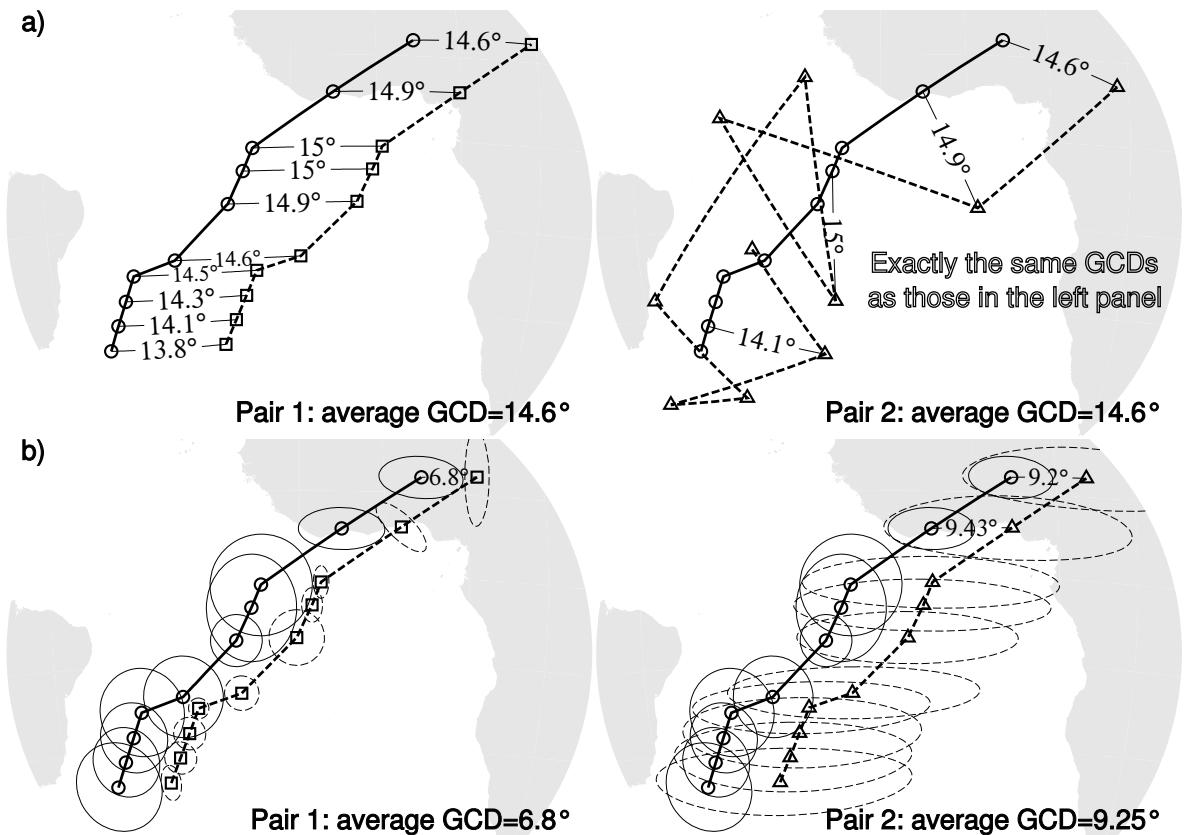
Investigations of the Earth’s past tectonic evolution and paleogeography often involve comparing APWPs. For example, if two now separated continental fragments were once part of the same supercontinent, their APWPs should share the same geometry during the interval that this supercontinent existed. If 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, 31, 40, 10, for example]; significant deviations from the known APWP for a continent can also be used to identify local tectonic rotations [14, 8, 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, 14, 10, 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 [31, 40, 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 [39].

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

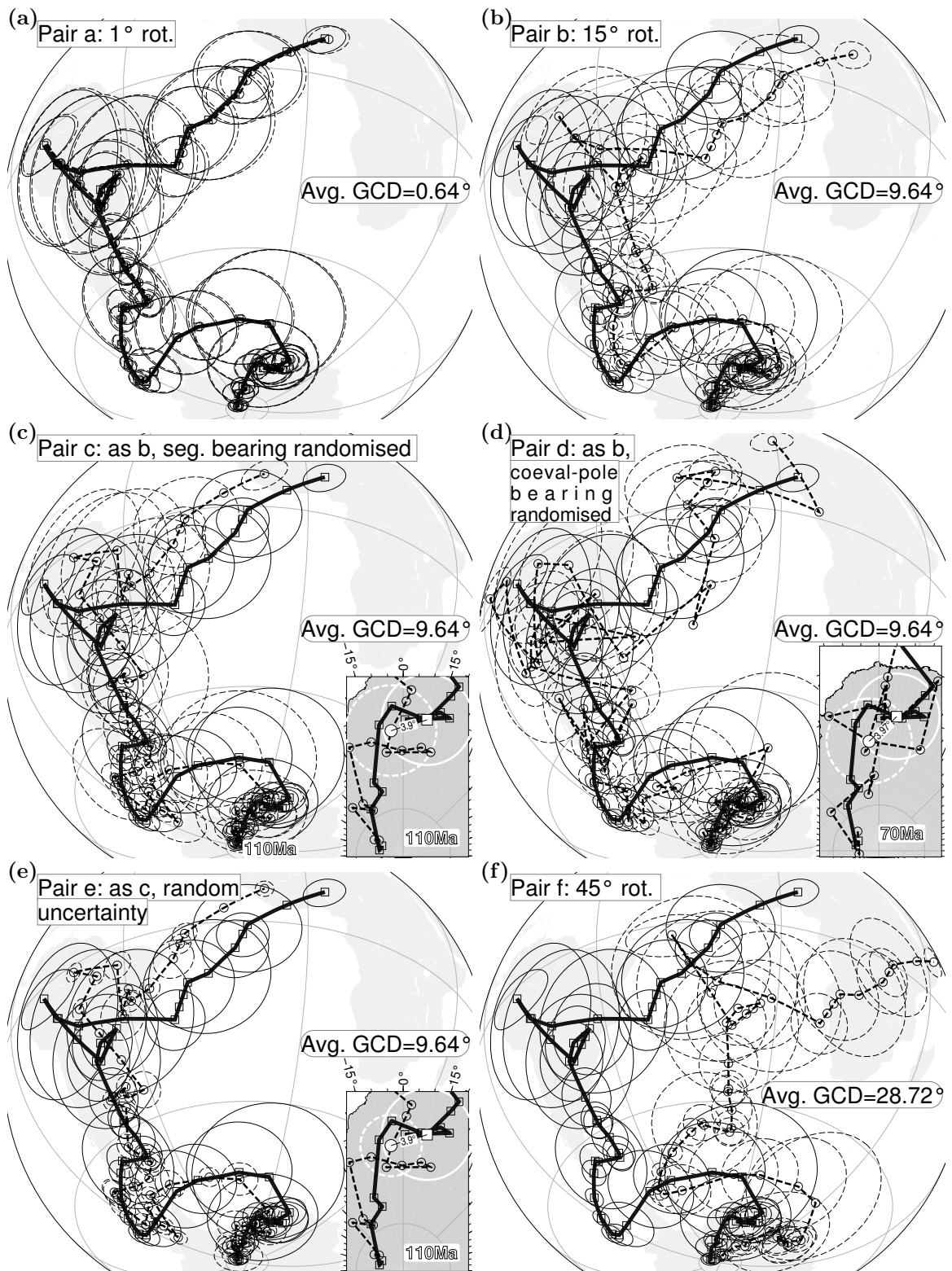
### 2.2.1 Comparing Apparent Polar Wander Paths

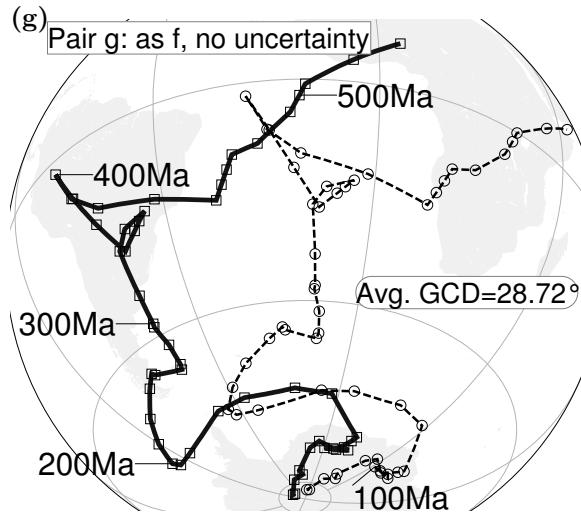
An APWP consists of a sequence of ( $\mathbf{n}$ ) mean poles,  $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_n$ , which average the published paleopoles from a particular continent for a particular time interval. Each mean pole has associated longitude ( $\phi$ ), latitude ( $\lambda$ ), and age ( $t$ ). 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 [42]. 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 [42, “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 [31, 40, 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 [37]. 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 [37]. 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

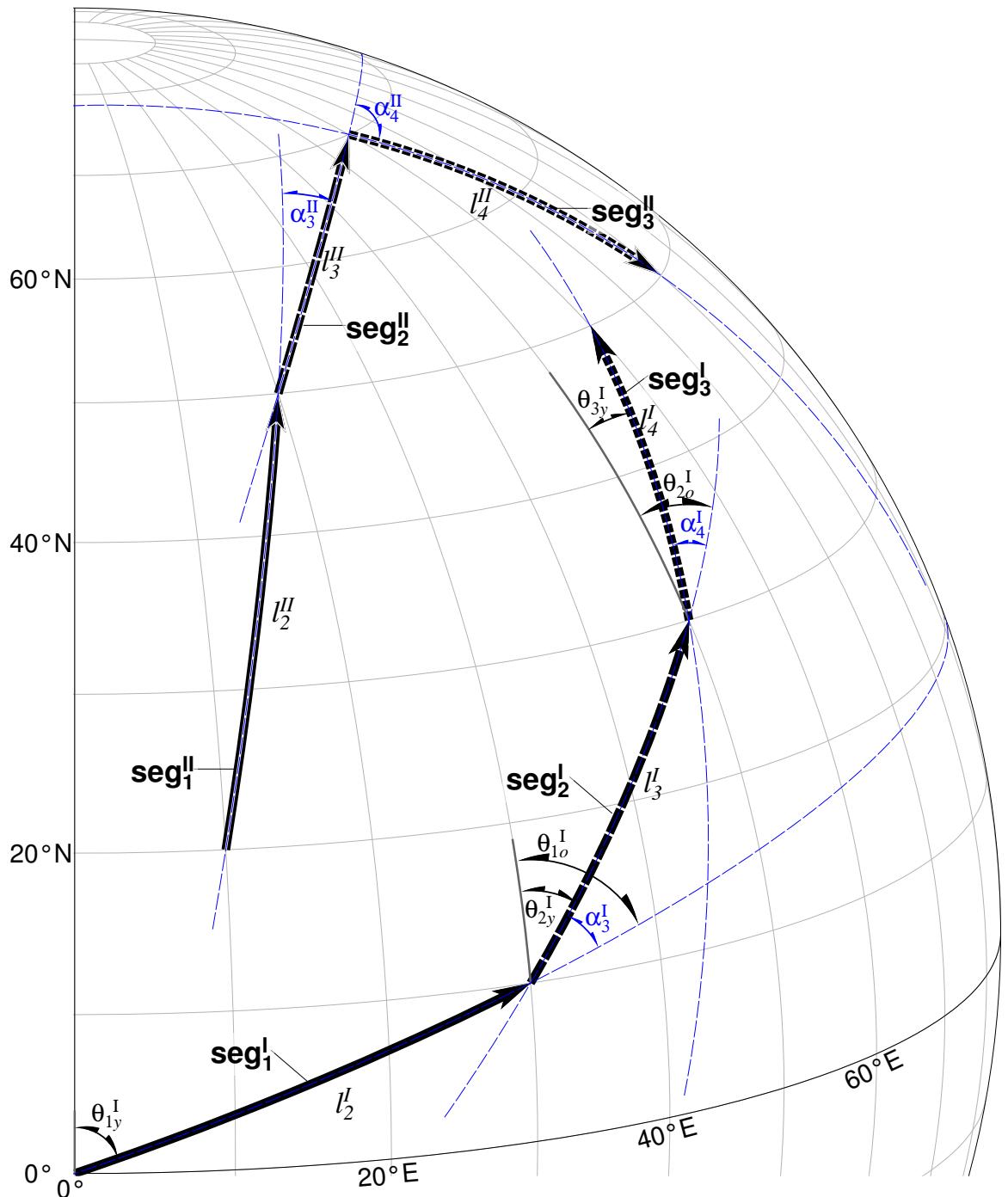
The shape of an APWP is determined by the orientations and lengths of its geodesic segments, which are related to the location of the Euler stage pole that describes plate motions, and the rotation rate about that pole, respectively. The geometric similarity of two APWPs can therefore be assessed by 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 [33, 20, 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

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

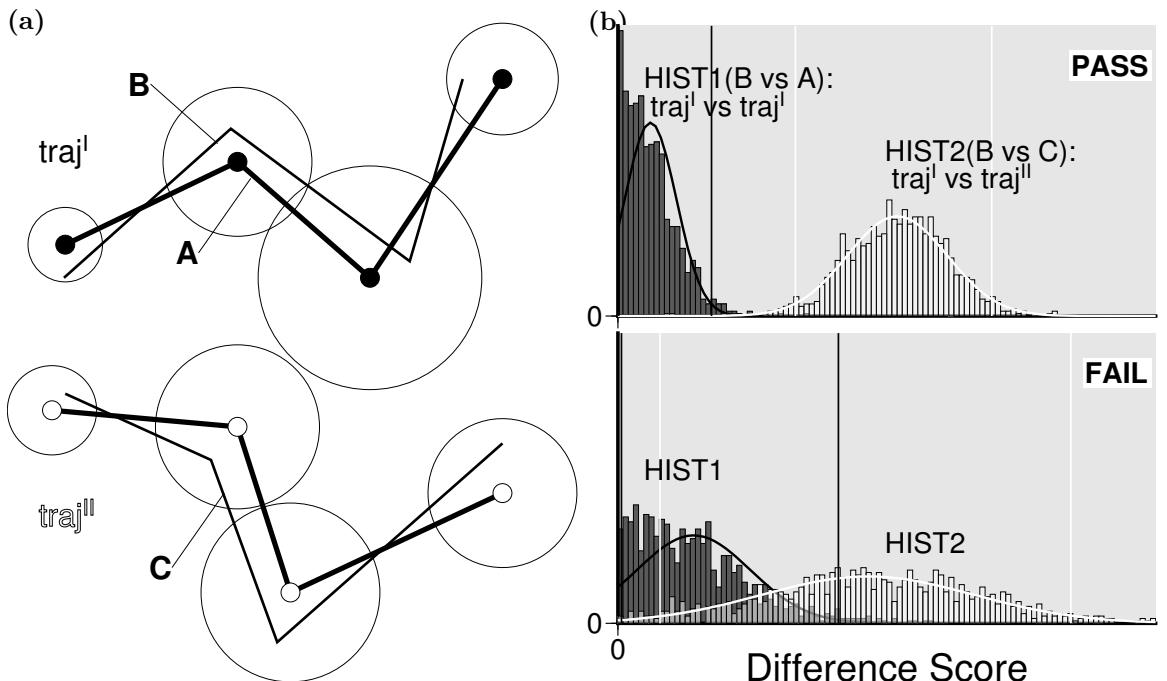
where

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

$d_a$  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

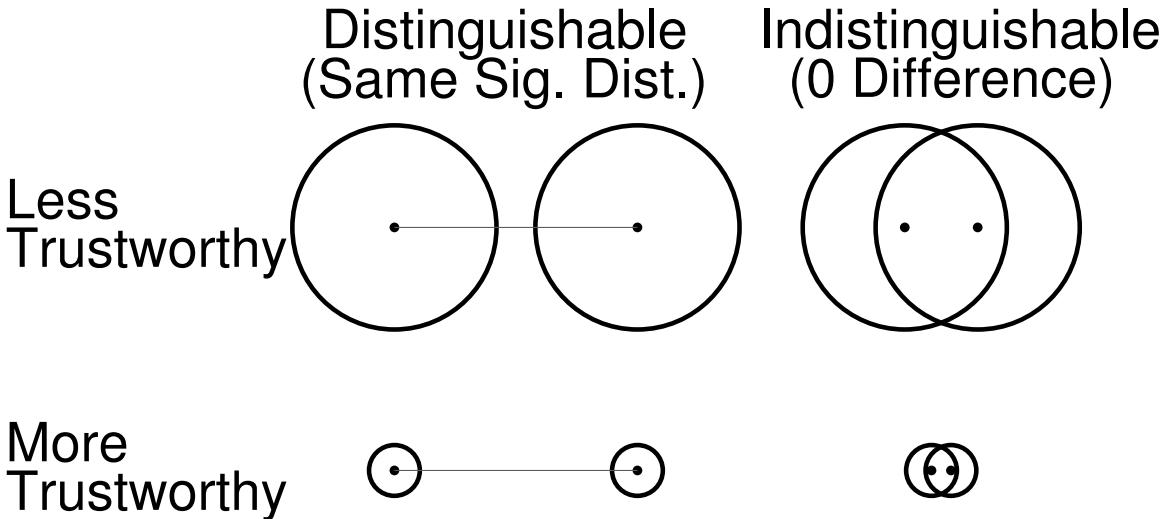
The three difference measures described above can be combined into a composite path difference ( $\mathcal{CPD}$ ) by means of a simple linear weighting rule,

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

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 [21]. 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  $27^\circ$  [33, 20, 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 [42].

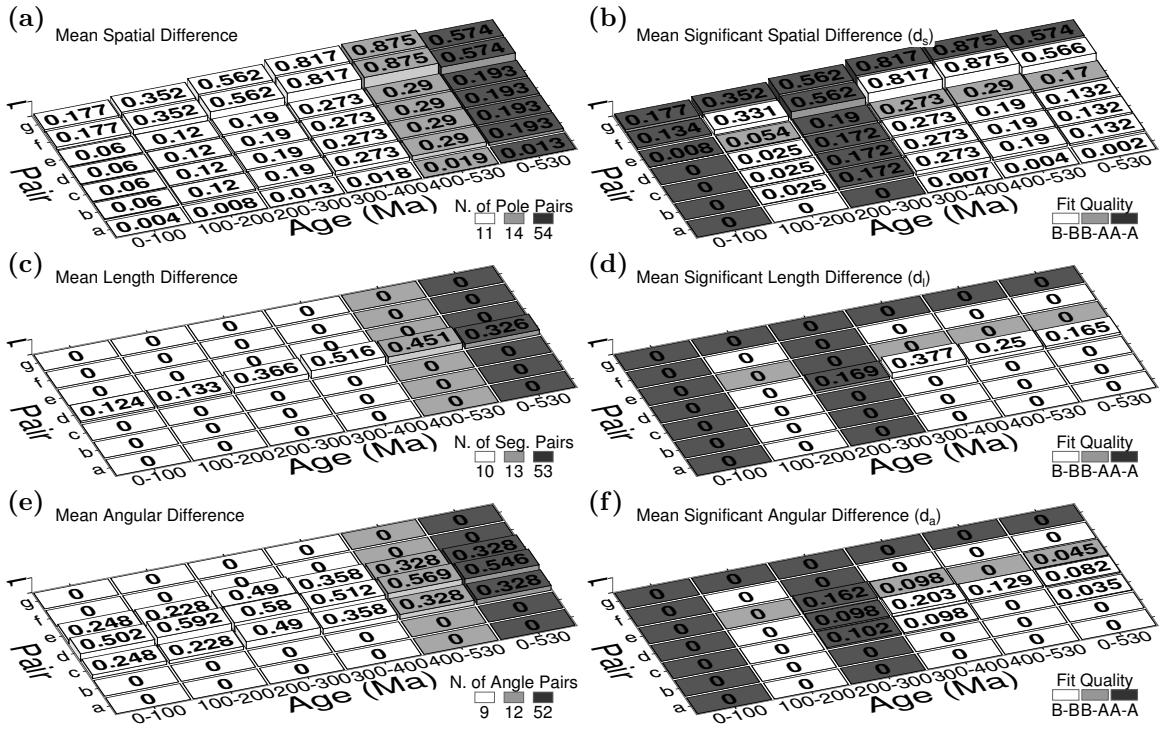
#### 2.3.1.1 $d_s$

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).

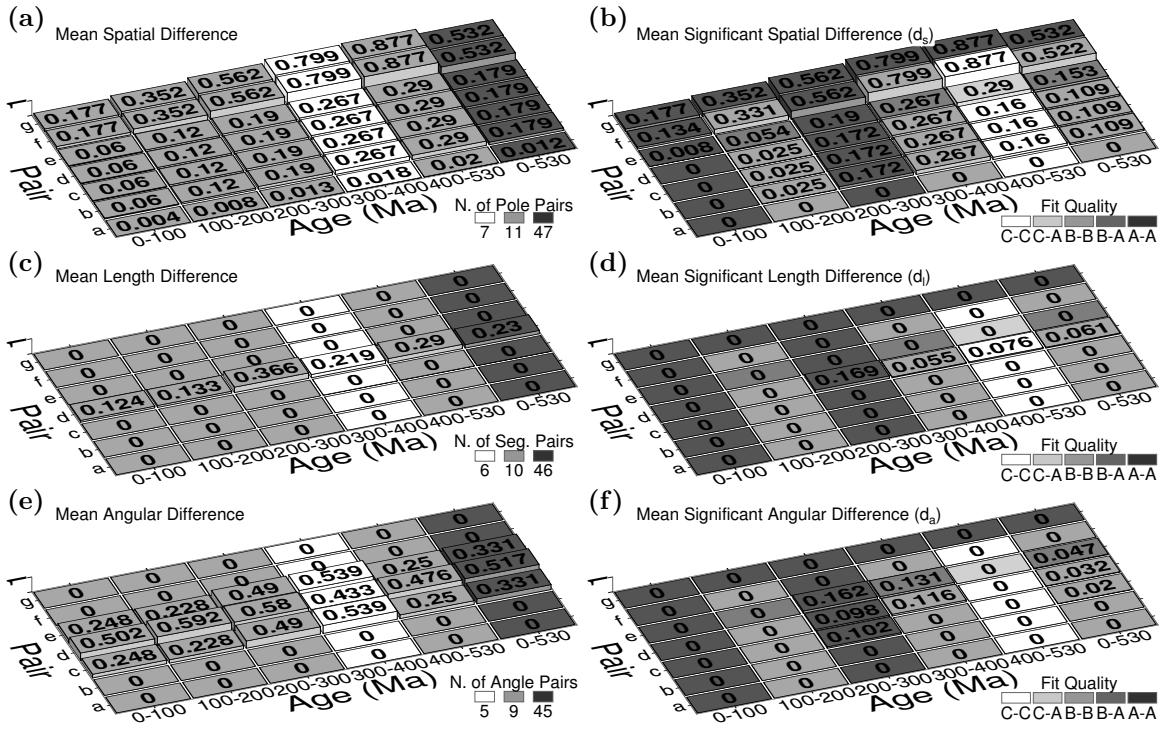
Most  $d_s$  scores are also reduced with significance testing, with the largest reductions occurring where the path separations are low and 95% confidence ellipses for coeval poles are more likely to overlap (e.g.,  $d_s$  for *Pair a* is reduced by 85%,  $d_s$  for *Pair f* is reduced by <2%).  $d_s$  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 [42], 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

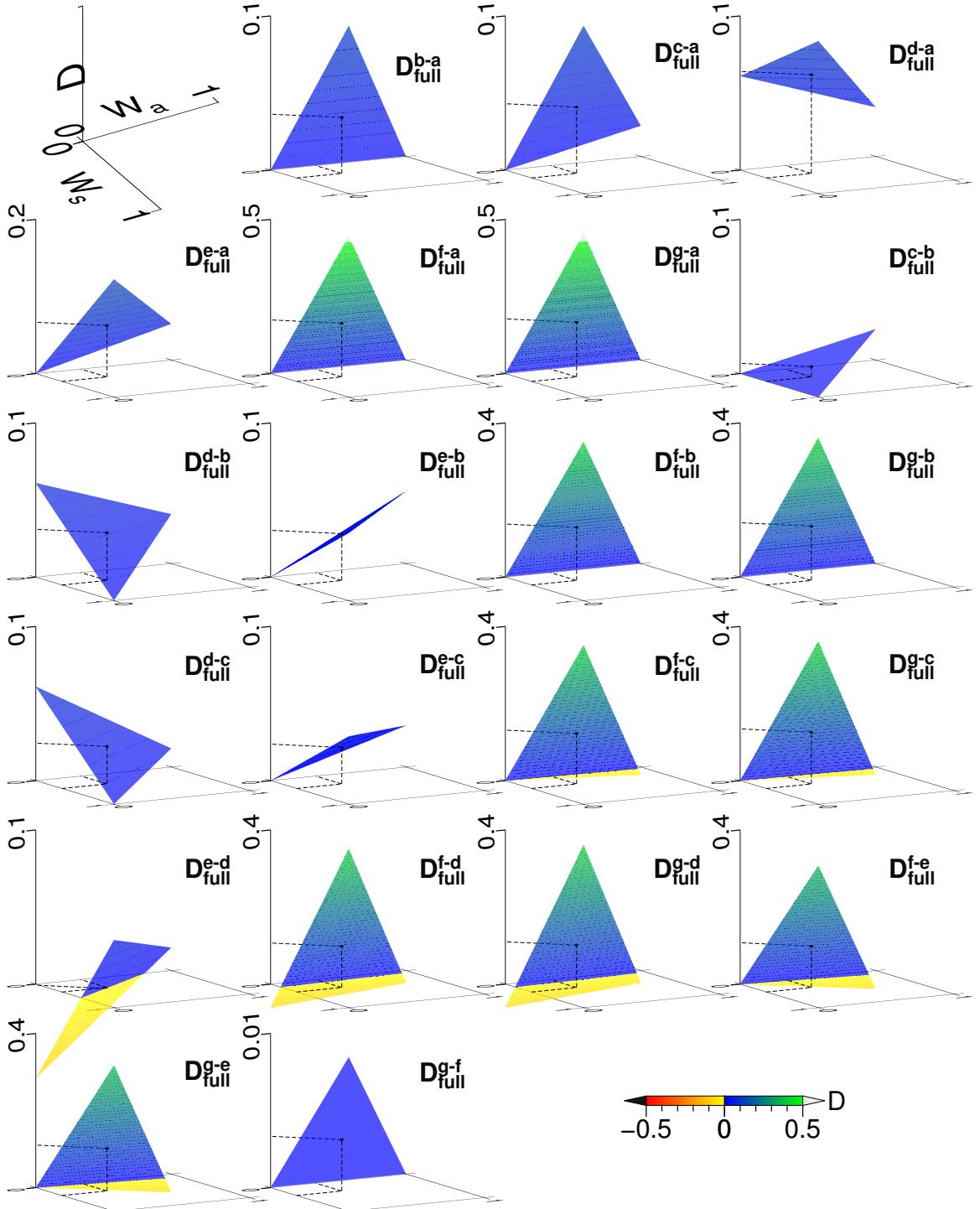
of  $\mathcal{CPD}s$  from each two APWP pairs:

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

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  $W_s = W_a = W_l = \frac{1}{3}$ .

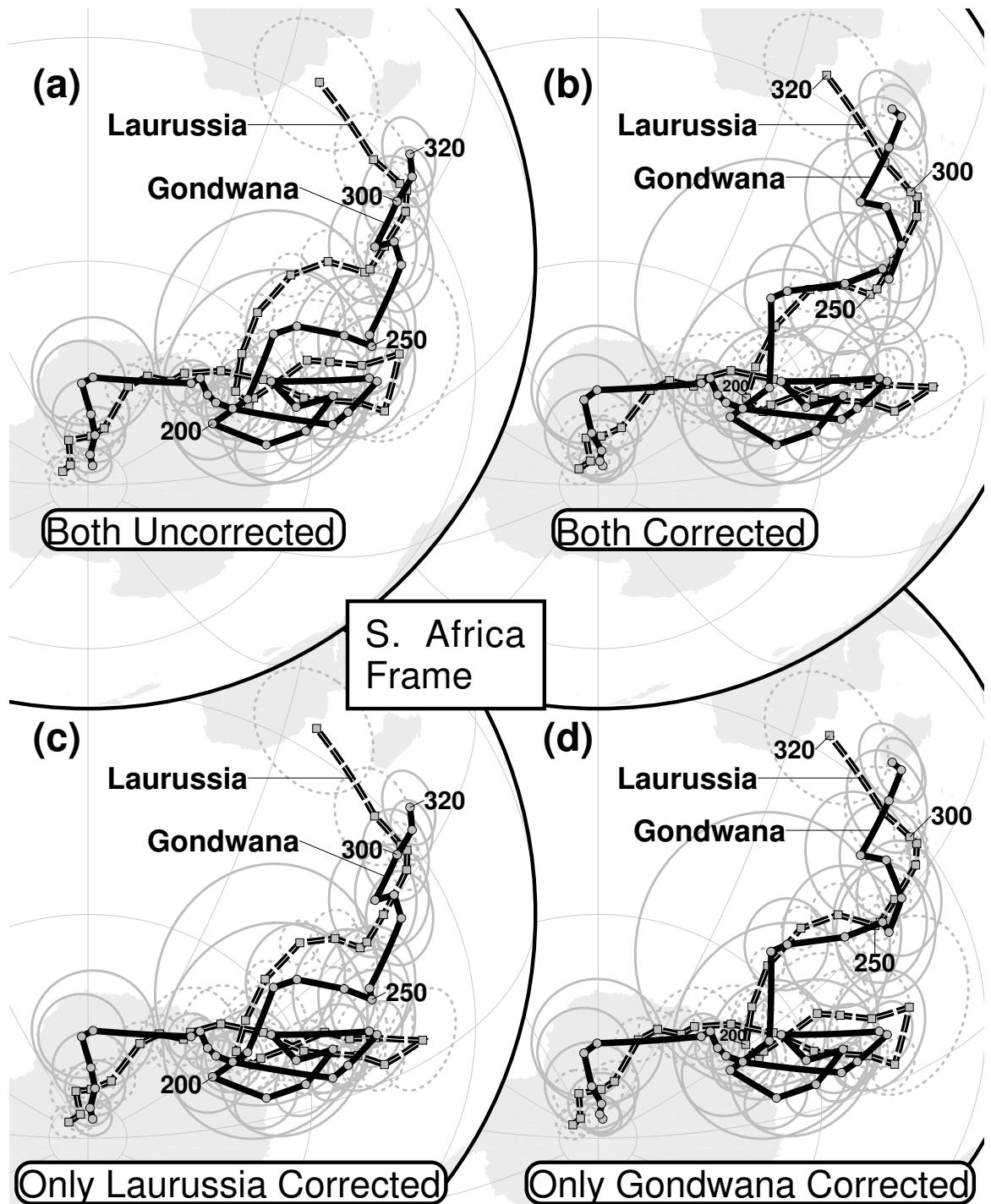
In summary, as Fig. 2.3 illustrates, mean GCD has trouble discriminating between *Pairs f* and *g*, and *Pairs c* and *e*, and also between intermediate similarities where the differences are mainly in path geometry (*Pairs b, c, d*). Our algorithm provides an improved solution for this problem. Obtaining 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

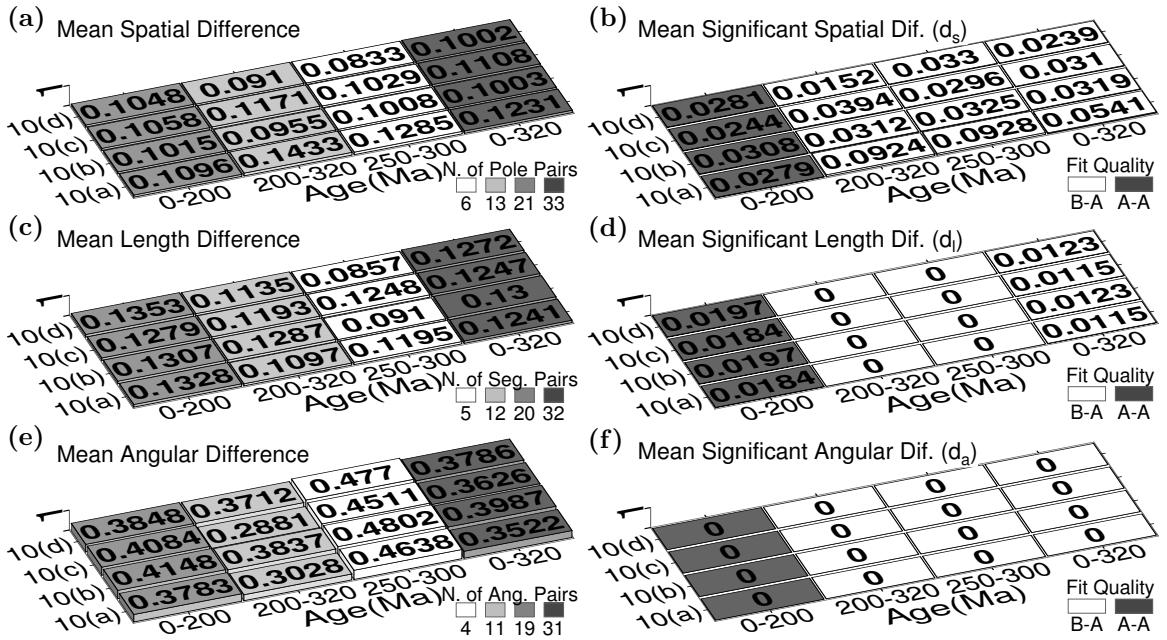
To illustrate how these metrics might be useful when applied to real paleomagnetic data, we compare 320–0 Ma APWPs for Gondwana and Laurussia calculated using a running mean method by [42], with paleopoles from sedimentary rocks both uncorrected (Fig. 2.10a) and bulk corrected for inclination shallowing ( $f=0.6$ ; Fig. 2.10b).

When comparing the 320–0 Ma paths [42, 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 [42, 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 [42]. 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 [38] and GMT [48]. All images are produced using GMT [48]. 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 in different ways, and then the application of the new APW path similarity measuring tool used in finding the best APWP generating method(s)*

(This chapter is appended right after the paper draft for Chapter 2; 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).)

# 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

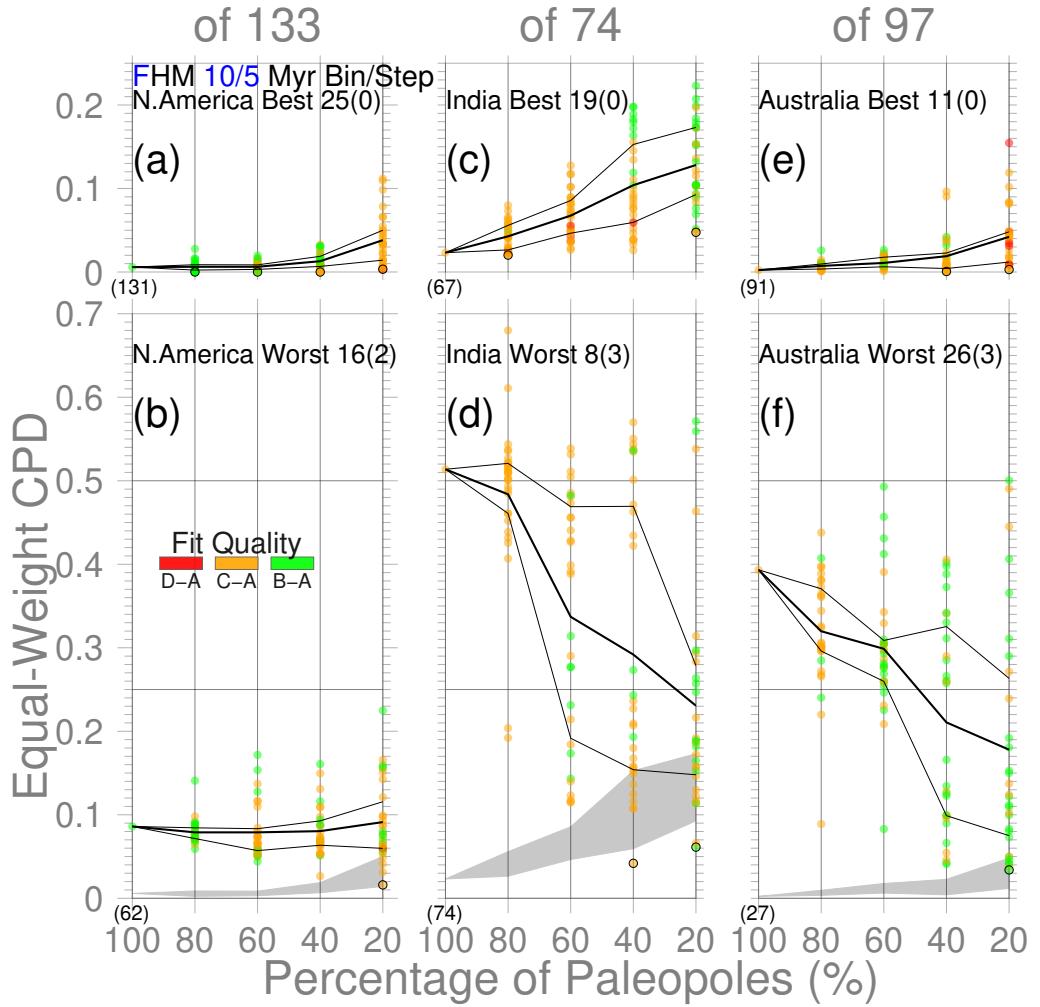
The fixed hotspot model (FHM) and related plate circuit predicted APWPs are used as references. In fact, as mentioned in the last chapter, whether we choose FHM or whether we pick MHM as reference, they do not make much difference at all.

## 4.2 Extraction Fraction

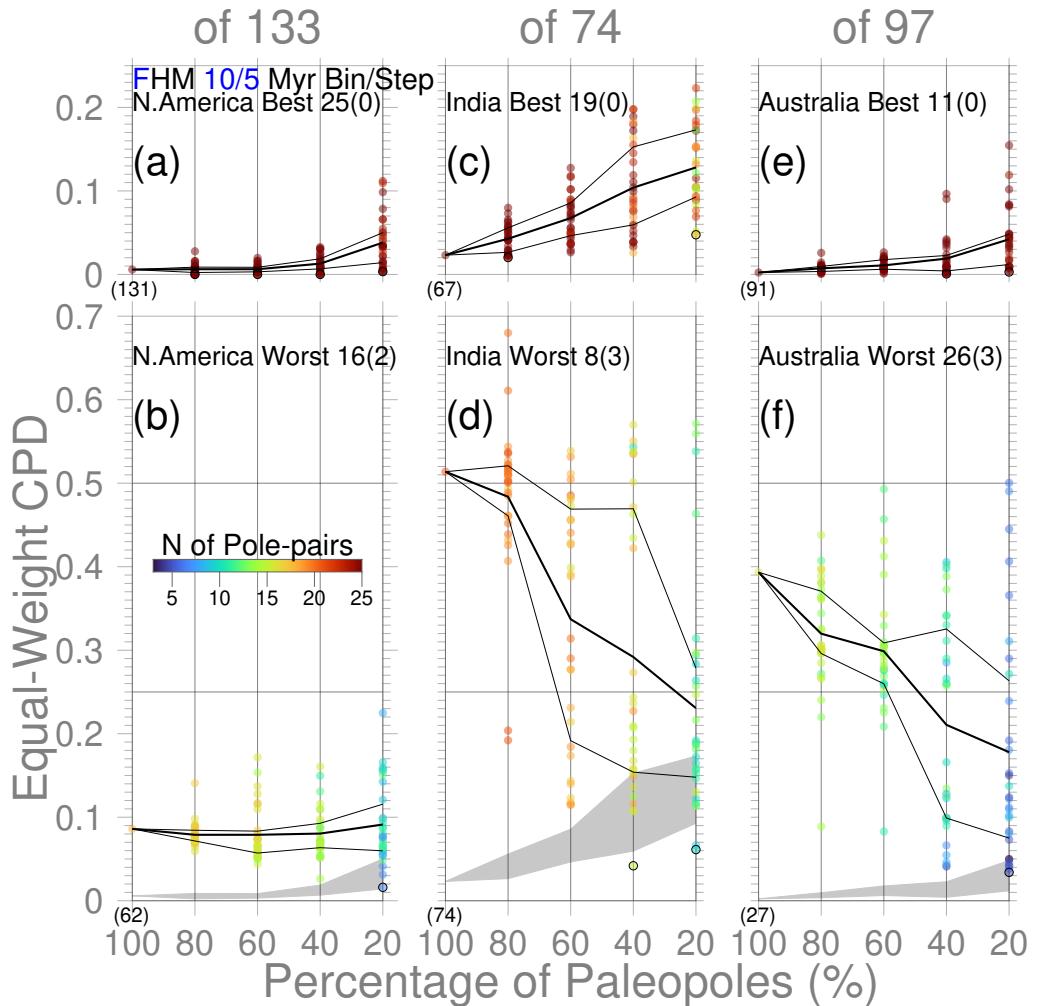
Sub-sampling (extracting part of raw paleopoles) is implemented before moving-averaging with filtering/correcting and weighting at four discrete percentages, 80%, 60%, 40% and 20%, which mean respectively 20, 40, 60 and 80 per cent of raw paleopoles are removed. This means not all sub-samples at, for 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.

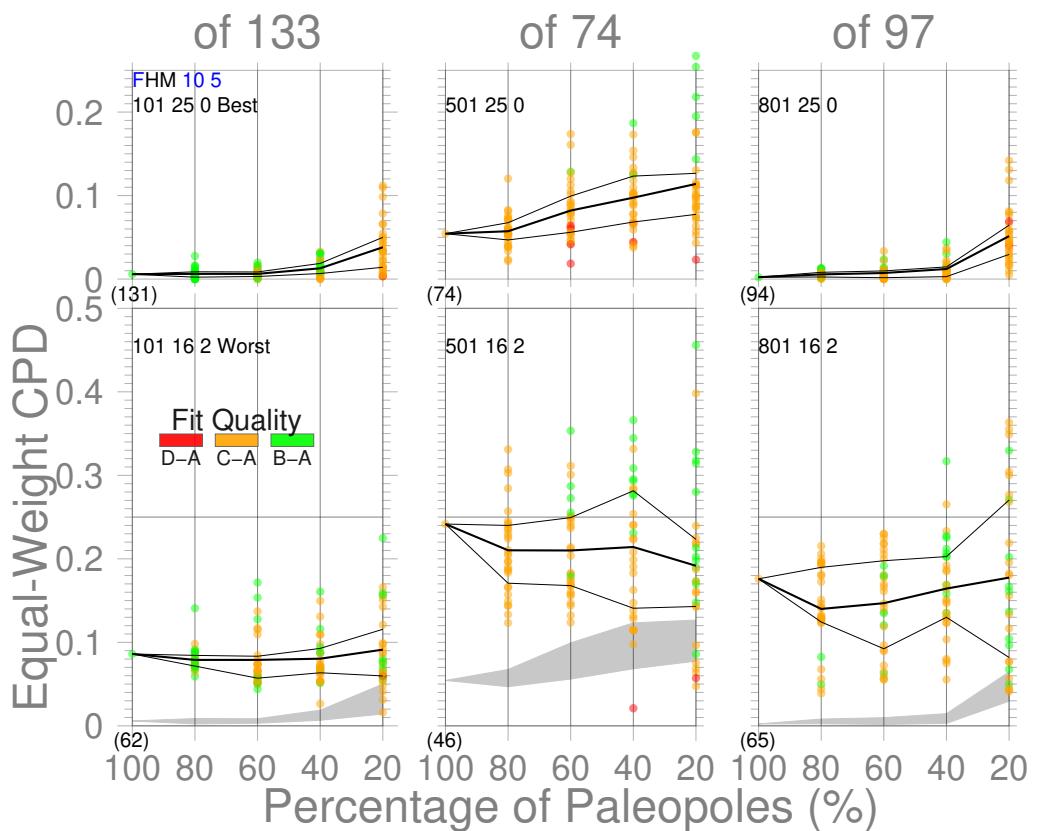
For the worst methods applied onto Indian and Australian data, the equal-weight  $\mathcal{CPD}$  surprisingly decreases when the percentage of extracted data decreases (Fig. 4.1). This is because after the data density is reduced the left data are not always enough to cover all the time range of 120–0 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).



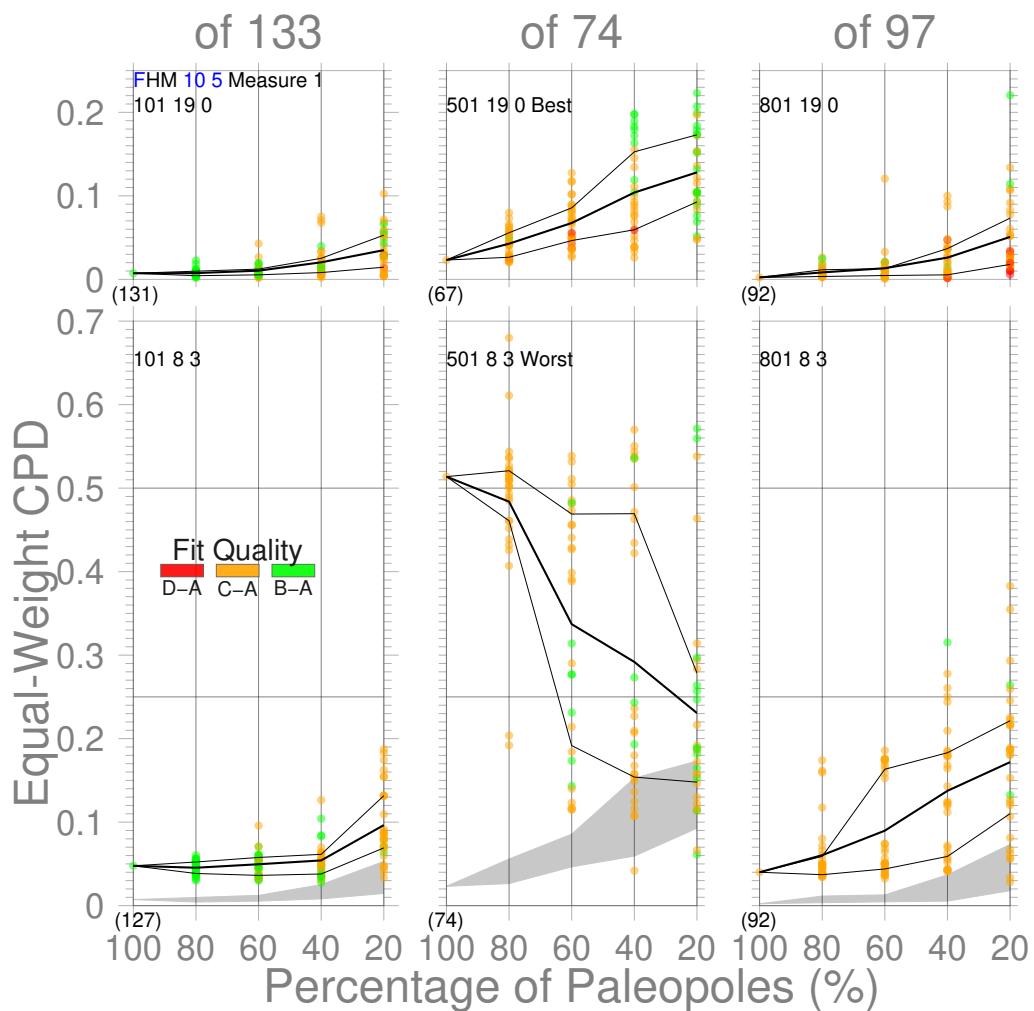
**Figure 4.1:** Random paleopole samplings (30 times) for the best and worst results for the 10 Myr window and 5 Myr step paleomagnetic APWPs vs FHM & plate circuit predicted APWP. The lower and upper bound lines connect the 1st and 3rd quantiles ( $Q_1$  and  $Q_3$ ) of the 30 samples. The bold line 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.



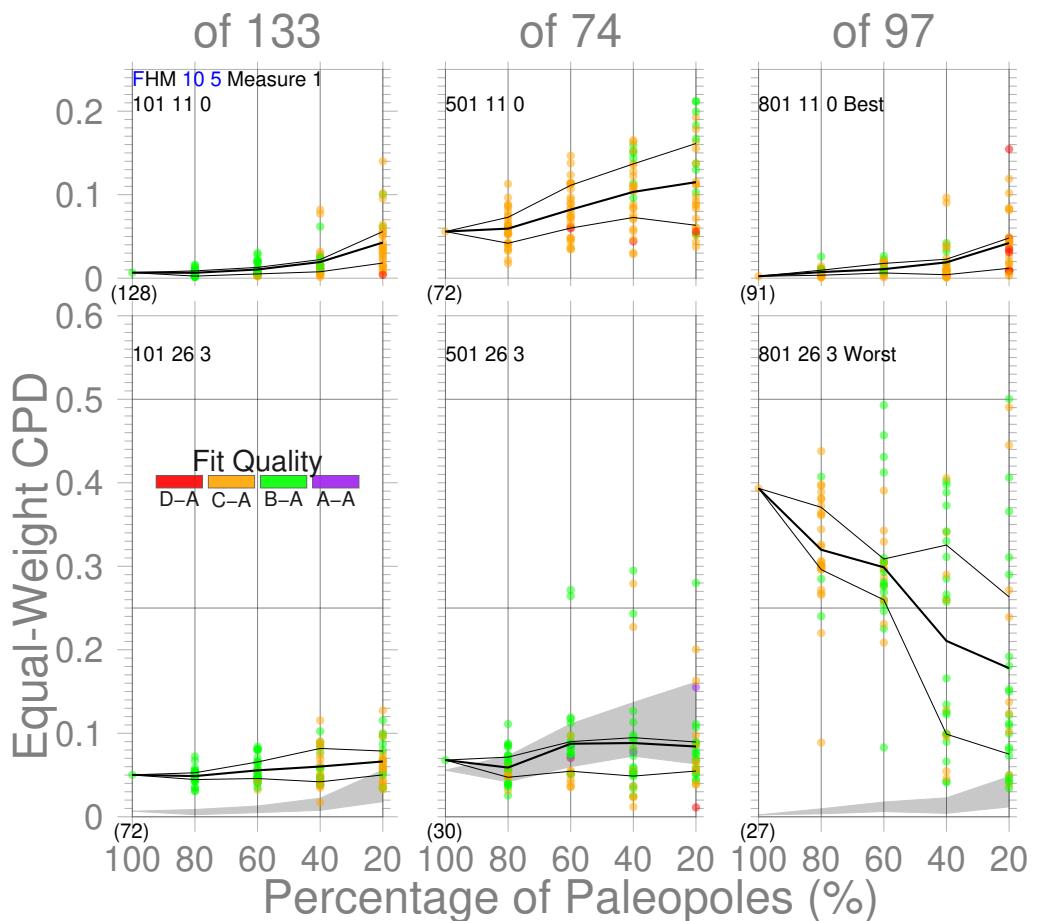
**Figure 4.2:** Same as Fig. 4.1. Here color demonstrates the resolution of pole pairs with the path comparisons.



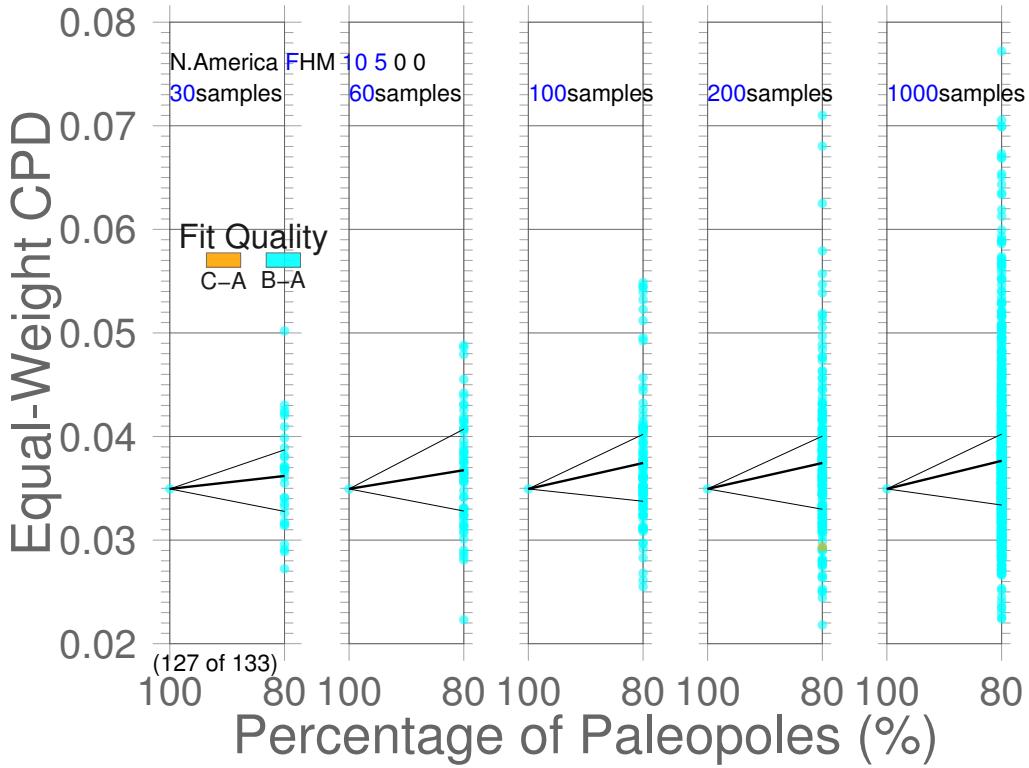
**Figure 4.3:** Comparisons of results from the best and worst methods for North America (101), also applied on the other two continents (501 and 801). The  $Q_1 - Q_3$  interquartile range from Picking No. 25 is also shown (shadowed) in the plot of Picking No. 16 for clarity.



**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.



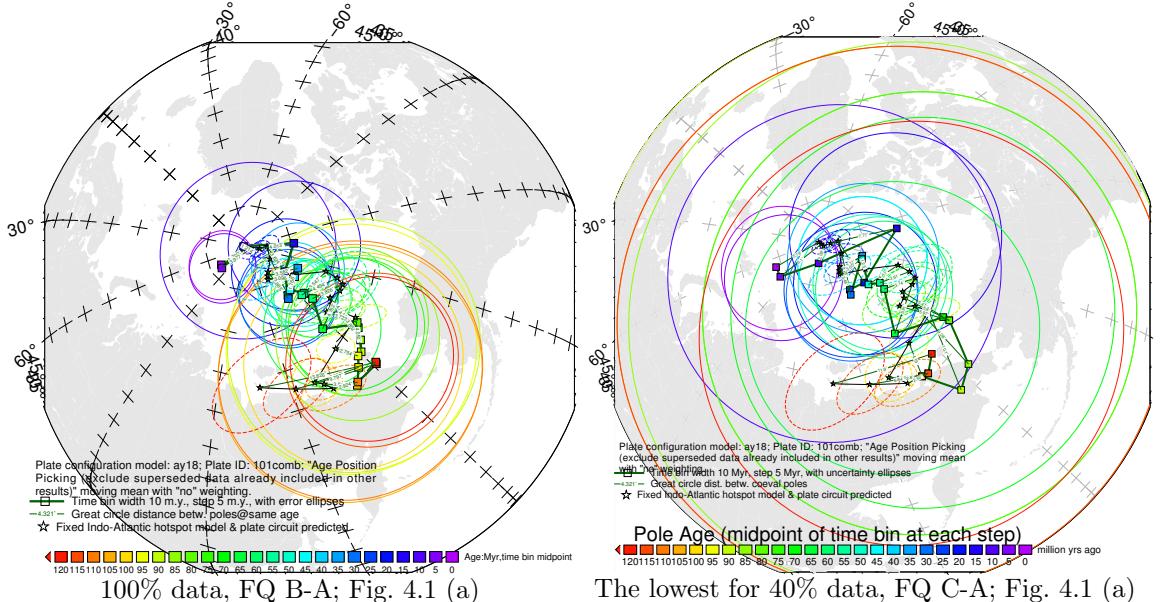
**Figure 4.6:** Testing differences of results from different numbers of samples. See Fig. 4.1 for more details.

#### 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. [18] 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

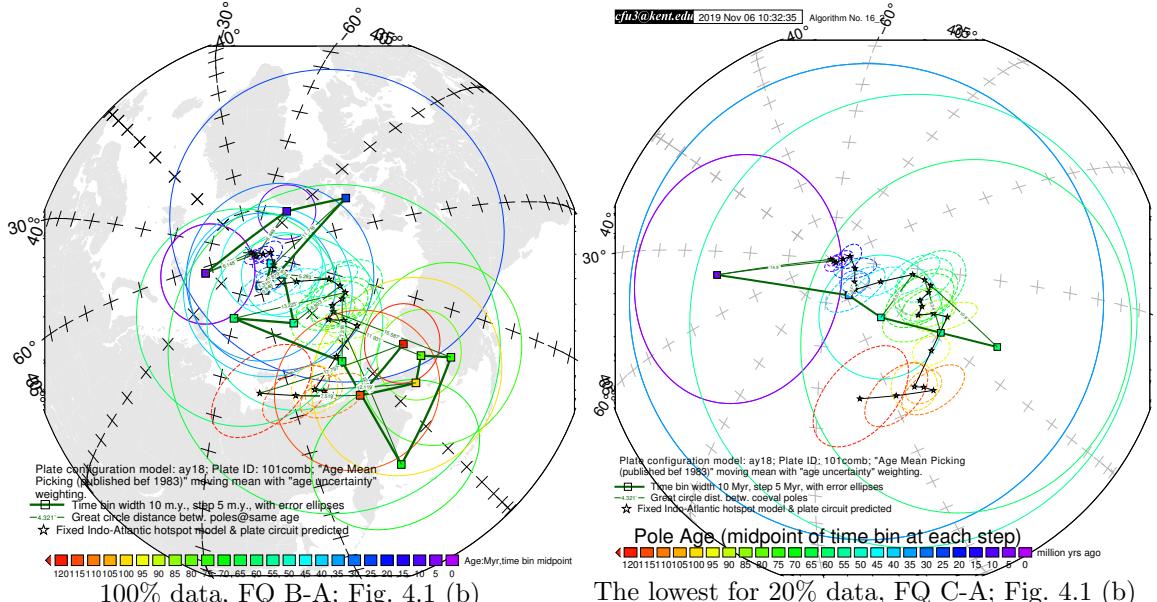
It seems rational for us to agree on that less paleomagnetic data tends to bring poorer similarity between paleomagnetic APWPs and the reference path. However, it is noticeable that even though the data density is tremendously reduced (e.g. by 80%), it seems still possible to have a better similarity 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)).



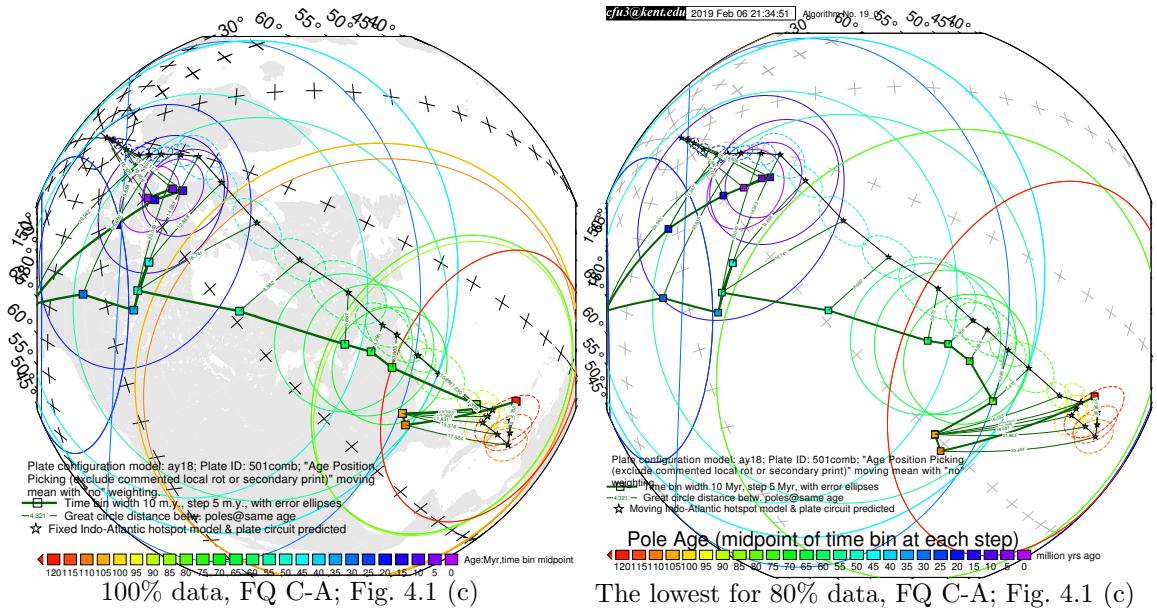
**Figure 4.8:** Comparing the 100% North American 120–0 Ma paleomagnetic data derived result with the best of the only 20% data (giving even better similarity) derived results (the yellow dot in Fig. 4.1 (b)).

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.

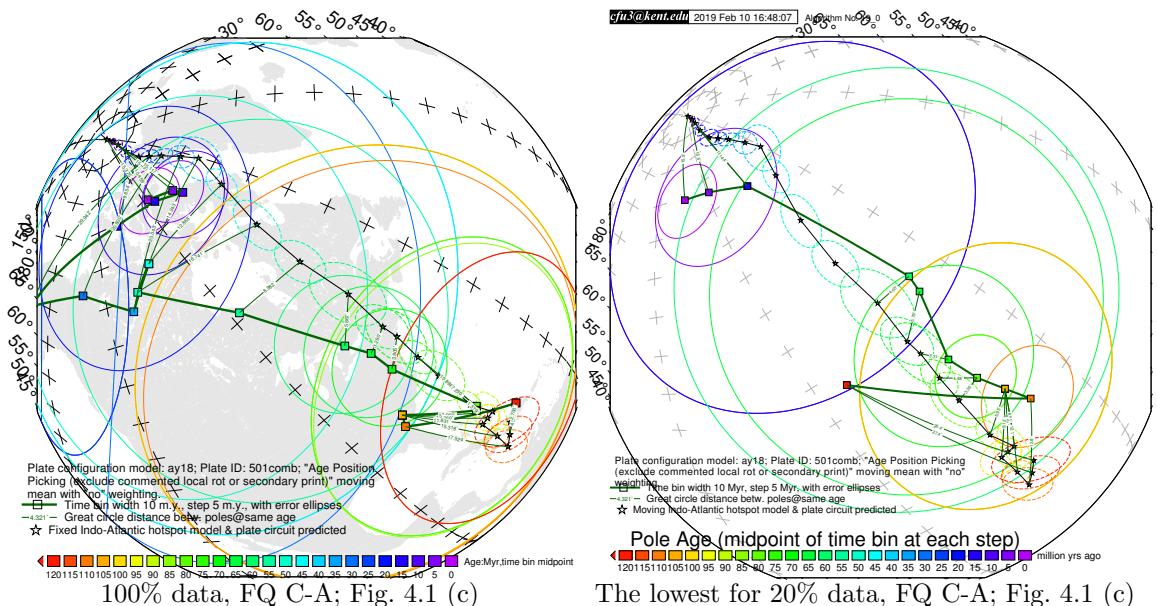
Still for the case (c) (Fig. 4.1), of the 30 samples for the 20% data, none is able to give better result compared with the 100% data, but the lowest difference value we can get from these 30 samples (20% in Fig. 4.1 (c)) indicates that even 20% data is still able to give good (not dramatically different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

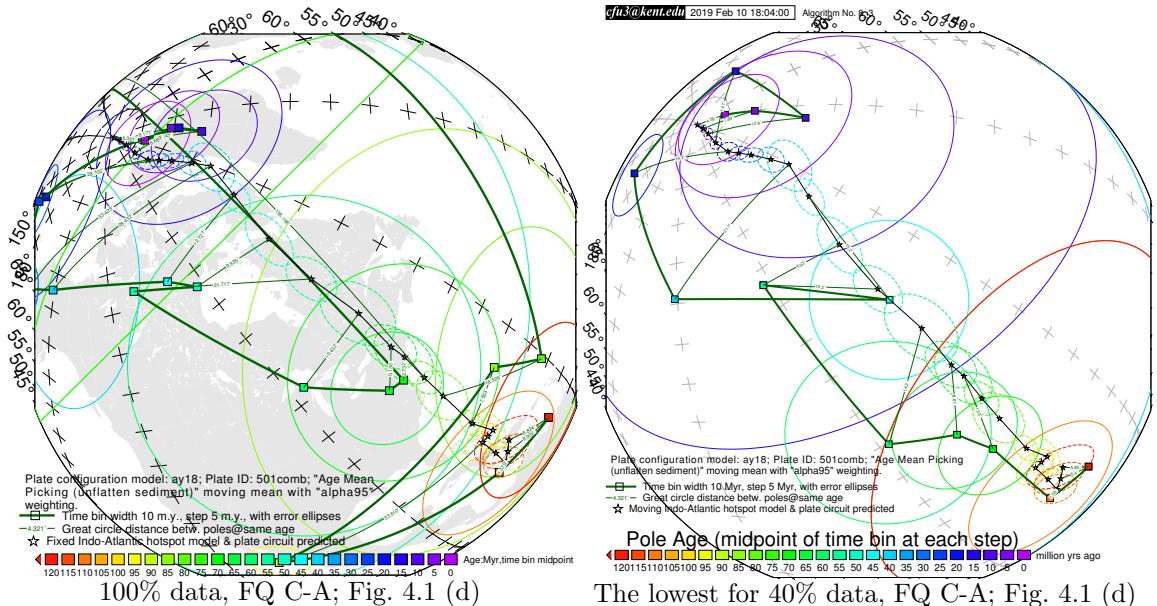
For the case (e) (Fig. 4.1), although the 20% data derived paleomagnetic paths are not closer to the reference path than the 100% data derived one, the closest one (the lowest yellow dot along the 20% grid line in Fig. 4.1 (e)) still performs quite well (Fig. 4.12b). This is mainly because the number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

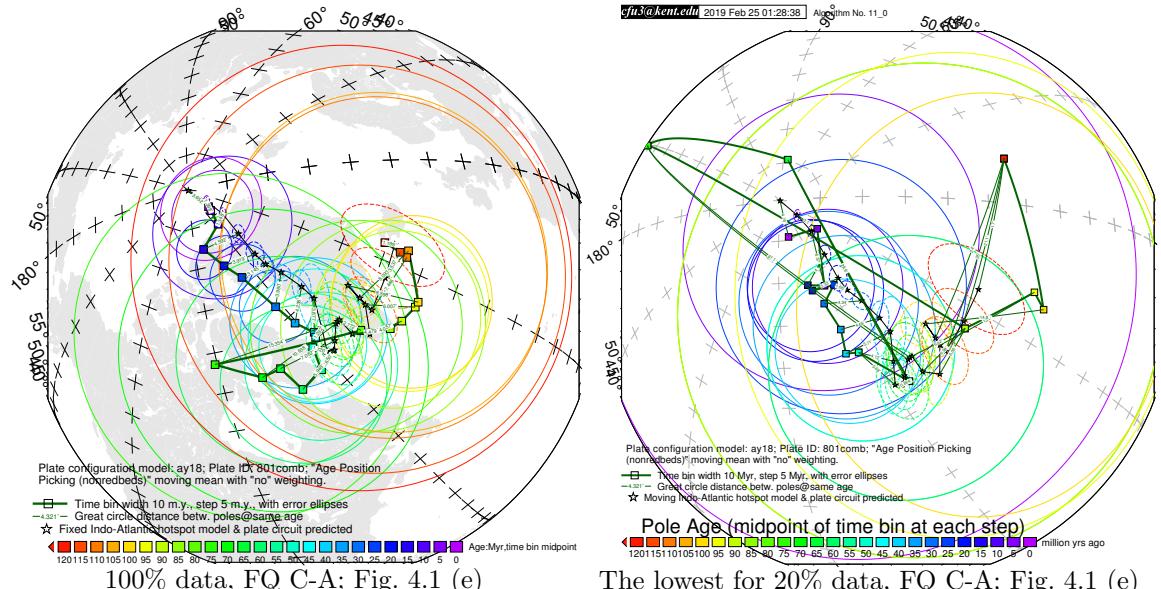
the 100% data generated path when there are only 20% of the paleopoles. However, for 120, 115, 110 and 105 Ma, where the mean poles, actually the same pole, are relatively far from the reference path for the 20% data (Fig. 4.12a), the uncertainties are also rather large (Fig. 4.1 (e)).

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.

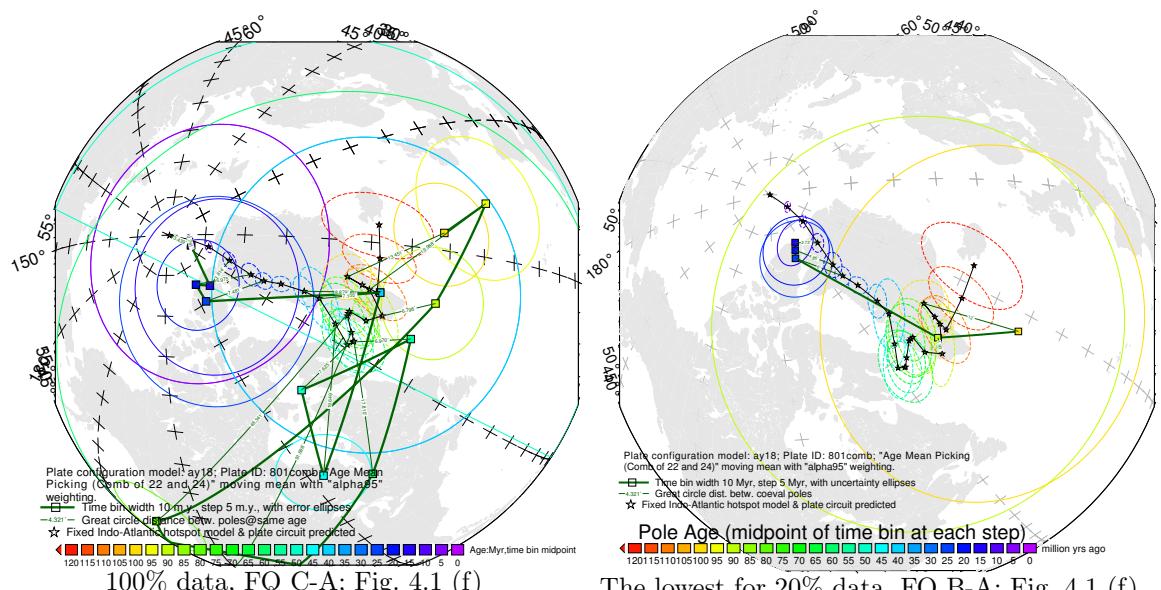
#### 4.2.3 Relationship between $d_s$ Score and Number of Paleopoles for Paleomagnetic APWP

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

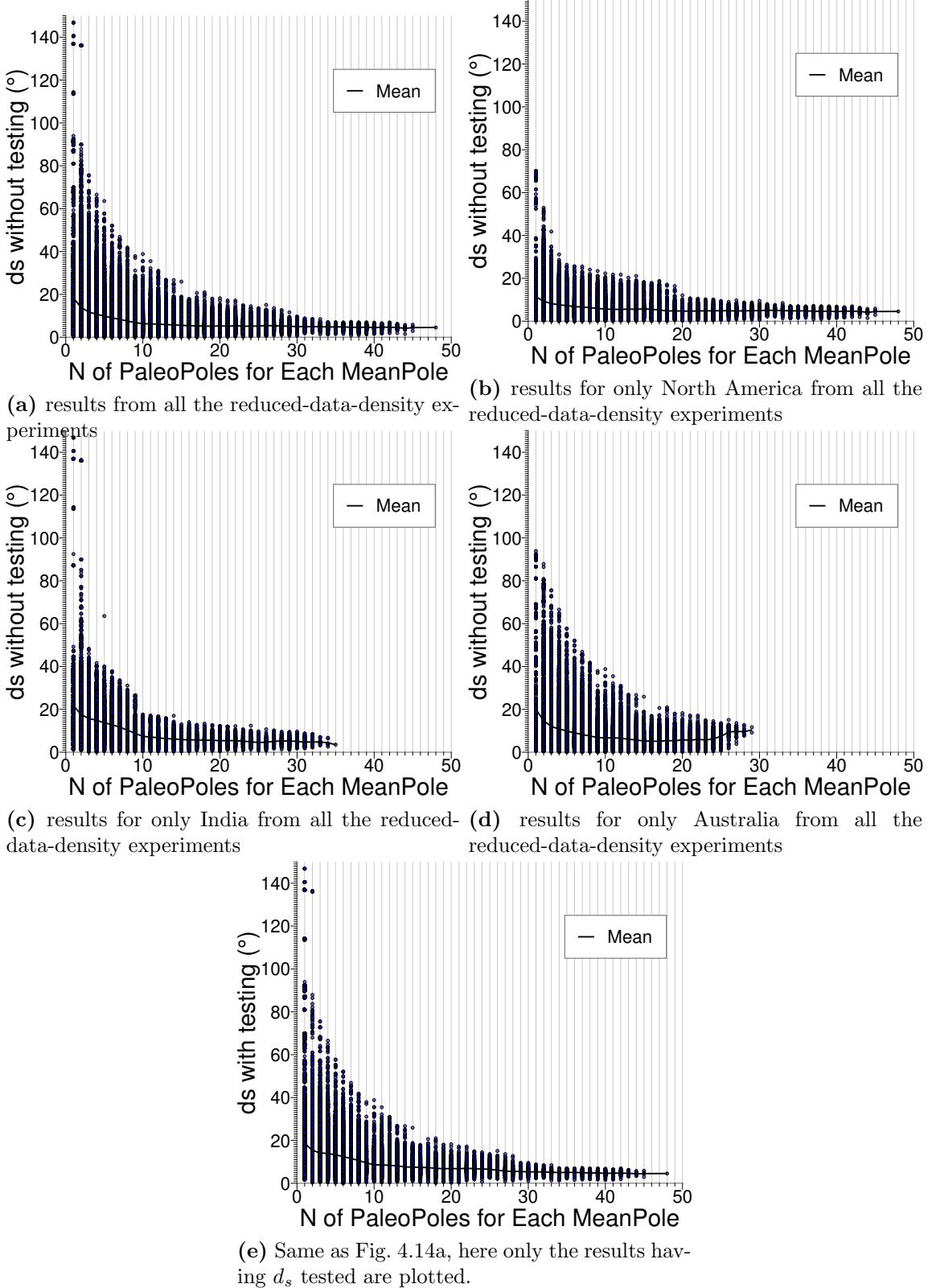
<sup>4</sup> See Figure 23 in Chapter 3

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.

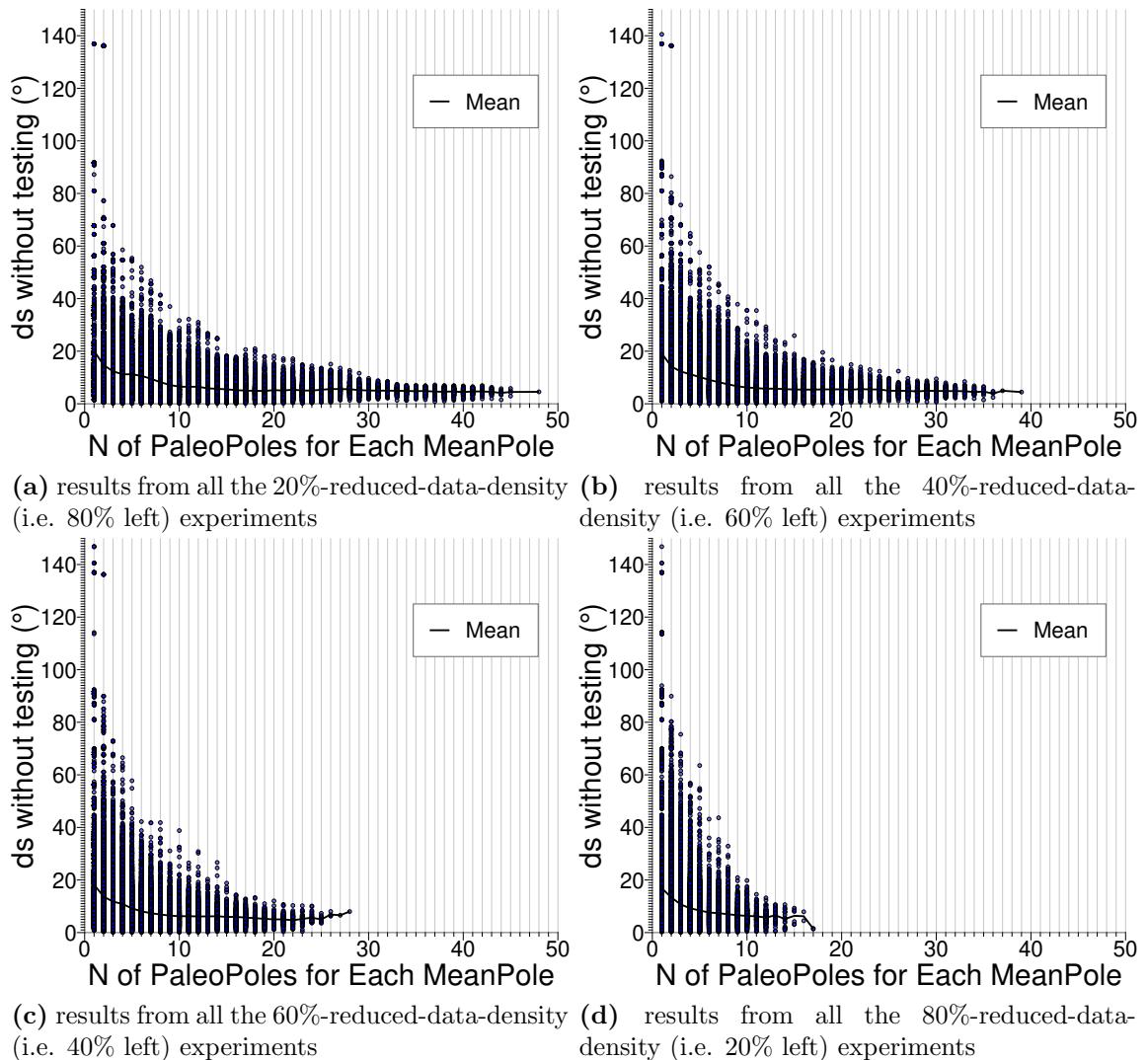
### 4.3 Are the rules we obtained in the last chapter are still true for less data?

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

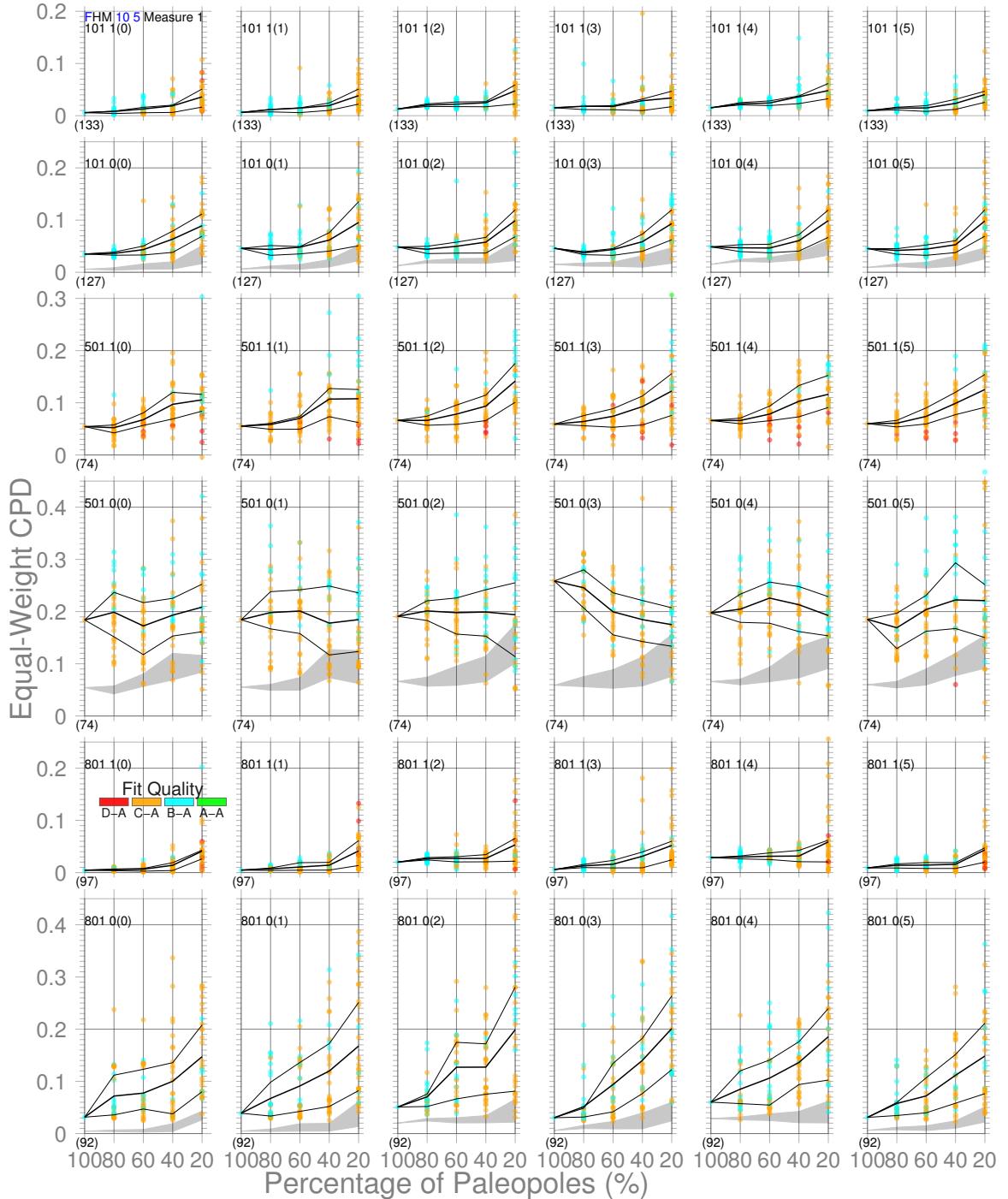
From the perspective of checking if the  $Q_1$ - $Q_3$  intervals overlap, APP is indeed still better than AMP and weighting is indeed still not affecting for more than 15 paleopoles making a 10/5 Myr bin/step APWP (ideally composed of 25 mean poles for 120–0 Ma). For the 20% Indian data case, which 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?



**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 [37]. 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 [37]. 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 [37] 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 [37] 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 [19]. 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” [45]. 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 [15] (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 [28, 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

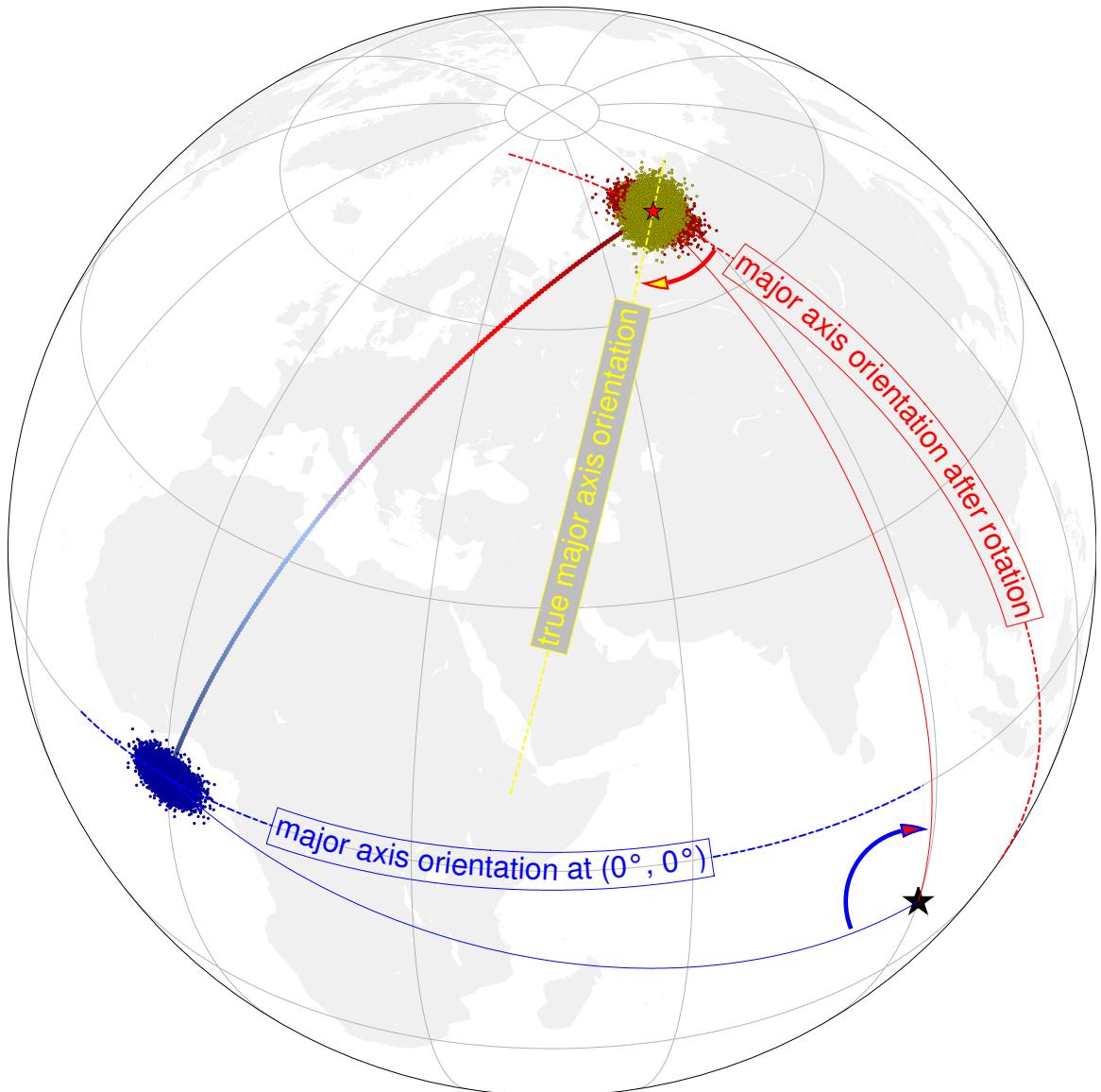
This algorithm is developed for comparing time-synchronized APWPs. In other words, the compared APWPs should have the same timestamps. If the number of their timestamps are different, the unpaired *pole(s)* would be removed to make the timestamps the same before the comparison. APWPs with a pole interpolated 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

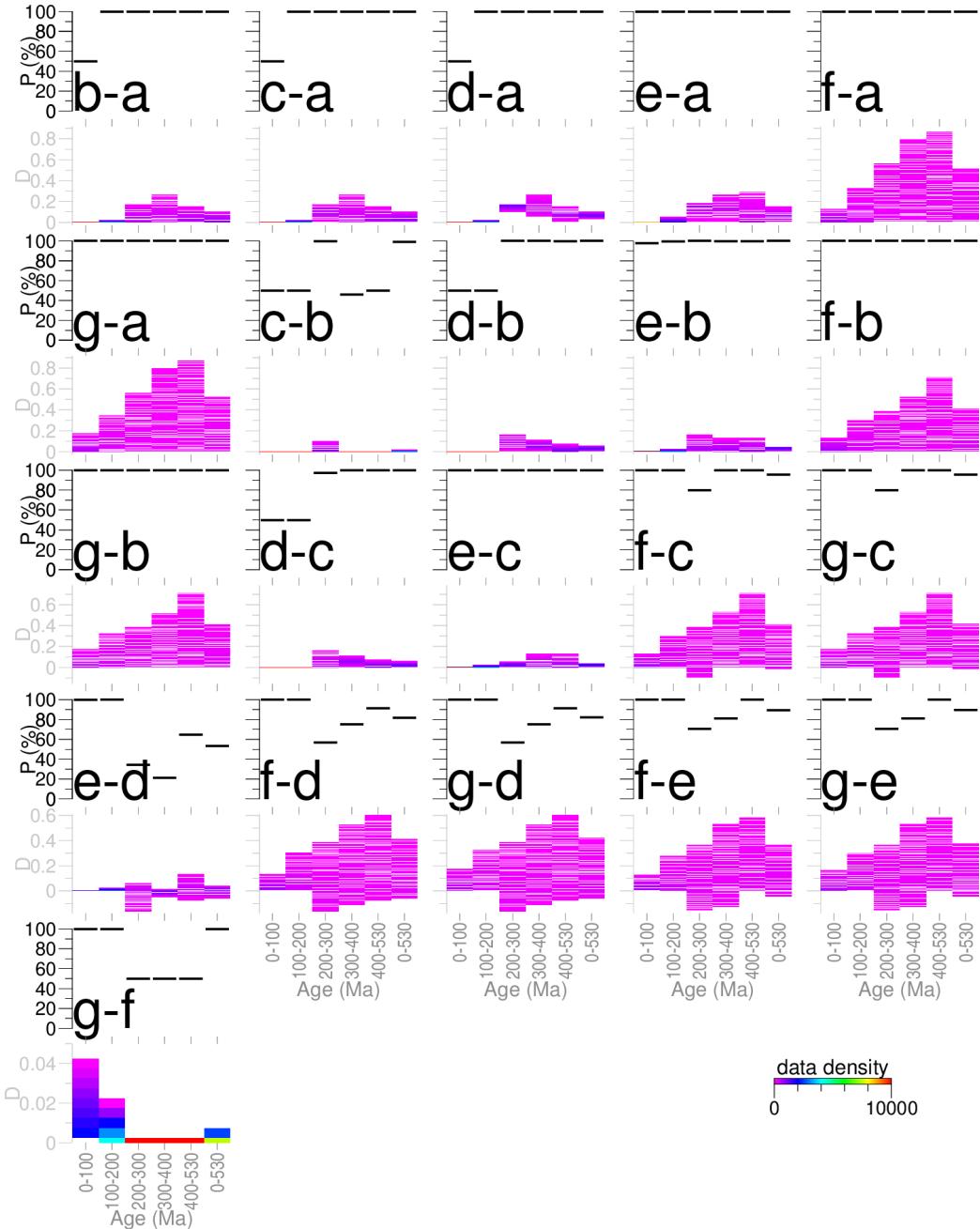
Assigning equally likely (not necessarily equal in value) random values to  $W_s, W_a, W_l$  is also tested. Three uniformly distributed random numbers with a given sum 1 are generated for, for example, 10 000 times here, and then are substituted into the *CPD* formula for deriving the seven APWP pairs’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

# 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] 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.
- [9] 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.

- [10] 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.
- [11] David A. D. Evans. Proterozoic low orbital obliquity and axial-dipolar geomagnetic field from evaporite palaeolatitudes. *Nature*, 444(7115):51–55, 2006.
- [12] R A Fisher. Dispersion on a sphere. *Proc Roy Soc London Ser A*, 217:295–305, 1953.
- [13] 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.
- [14] 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.
- [15] B Greiner. Euler rotations in plate-tectonic reconstructions. *Comput Geosci*, 25(3):209–216, 1999.
- [16] S J Hellinger. The uncertainties of finite rotations in plate-tectonics. *J Geophys Res*, 86(NB10):9312–9318, 1981.
- [17] J. K. Hillier. Pacific seamount volcanism in space and time. *Geophysical Journal International*, 168(2):877–889, 02 2007.
- [18] R V Hogg, Tanis E, and Zimmerman D. *Probability and Statistical Inference*. Pearson, New York, NY, 10 edition, 2019.
- [19] J T Kent. The Fisher-Bingham distribution on the sphere. *J Roy Stat Soc B*, 44(1):71–80, 1982.
- [20] 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.

- [21] P L McFadden and M W McElhinny. Classification of the reversal test in palaeomagnetism. *Geophysical Journal International*, 103(3):725–729, 12 1990.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] S A Pisarevsky. New edition of the Global Paleomagnetic Database. *Eos Trans AGU*, 86(17):170, 2005.
- [30] S A Pisarevsky and M W McElhinny. Global Paleomagnetic Data Base developed into its visual form. *Eos Trans AGU*, 84(20):192, 2003.

- [31] 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.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] J A Tarduno. On the motion of Hawaii and other mantle plumes. *Chem Geol*, 241(3-4):234–247, 2007.
- [36] 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.
- [37] L Tauxe, N Kylstra, and C Constable. Bootstrap statistics for paleomagnetic data. *J Geophys Res Solid Earth*, 96(B7):11723–11740, 1991.
- [38] 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.
- [39] T H Torsvik and L R M Cocks. *Earth History and Palaeogeography*. Cambridge University Press, Cambridge, 2016.
- [40] 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.
- [41] 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.

- [42] 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.
- [43] 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.
- [44] R van der Voo. The reliability of paleomagnetic data. *Tectonophysics*, 184(1):1–9, 1990.
- [45] 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.
- [46] 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.
- [47] P Wessel, R D Müller, and G Schubert. *Plate Tectonics*, book section 6.02, pages 49–98. Elsevier, Amsterdam, 2007.
- [48] 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.