Doubrovine, P. V., Steinberger, B., Torsvik, T. H. (2016): A failure to reject: Testing the correlation between large igneous provinces and deep mantle structures with EDF statistics. - *Geochemistry Geophysics Geosystems (G3)*, *17*, 3, pp. 1130–1163.

DOI: http://doi.org/10.1002/2015GC006044

AGU PUBLICATIONS

Geochemistry, Geophysics, Geosystems

RESEARCH ARTICLE
10.1002/2015GC006044

A failure to reject: Testing the correlation between large igneous provinces and deep mantle structures with EDF statistics

Pavel V. Doubrovine[1], Bernhard Steinberger[1,2], and Trond H. Torsvik[1,3,4]

[1]Centre for Earth Evolution and Dynamics, University of Oslo, Oslo, Norway, [2]Helmholtz Center Potsdam, GFZ German Research Center for Geosciences, Section 2.5 Geodynamic Modelling, Telegrafenberg, Potsdam, Germany, [3]Center for Geodynamics, Geological Survey of Norway, Trondheim, Norway, [4]School of Geosciences, University of Witwatersrand, South Africa

Abstract Absolute reconstructions of large igneous provinces (LIPs) for the past 300 Ma reveal a remarkable spatial pattern suggesting that almost all LIPs have erupted over the margins of the two large-scale structures in the Earth's lower mantle commonly referred to as the Large Low Shear-wave Velocity Provinces (LLSVPs). This correlation suggests that mantle plumes that have triggered LIP eruptions rose from the margins of LLSVPs, implying long-term stability of these structures and suggesting that they may be chemically distinct from the bulk of the mantle. Yet, some researchers consider the LLSVPs to be purely thermal upwellings, arguing that the observed distribution of LIPs can be explained by plumes randomly forming over the entire areas of LLSVPs. Here we examine the correlation between the LIPs and LLSVPs using nonparametric statistical tests, updated plate reconstructions, and a large number of alternative definitions of LLSVPs based on seismic tomography. We show that probability models assuming plume sources originating at the margins of LLSVPs adequately explain the observed distribution of reconstructed LIPs. In contrast, we find strong evidence against the models seeking to link LIPs with plumes randomly forming over the entire LLSVP areas. However, the hypothesis proposing that the correlation can be explained by plumes randomly forming over a larger area of slower-than-average shear wave velocities in the lowermost mantle cannot be ruled out formally. Our analysis suggests that there is no statistically sound reason for questioning the hypothesis that the LIPs correlate with the margins of LLSVP globally.

Key Points:
- Three geodynamic scenarios that can lead to the correlation between the eruption sites of LIPs and the LLSVPs have been tested
- The hypothesis that LIPs are predominantly formed by mantle plumes rising from the LLSVP margins is fully consistent with observations
- The hypothesis linking LIPs with plumes from the entire area of slower-than-average shear wave velocity in the lowermost mantle cannot be rejected, but results in inferior fit to the observations
- There is no statistically sound reason for questioning the hypothesis that the LIPs correlate with the margins of LLSVP globally

Supporting Information:
- Supporting Information S1

Correspondence to:
P. V. Doubrovine,
paveld@fys.uio.no

Citation:
Doubrovine, P. V., B. Steinberger, and T. H. Torsvik (2016), A failure to reject: Testing the correlation between large igneous provinces and deep mantle structures with EDF statistics, *Geochem. Geophys. Geosyst.*, *17*, 1130–1163, doi:10.1002/2015GC006044.

Received 3 AUG 2015
Accepted 25 FEB 2016
Accepted article online 25 FEB 2016
Published online 23 MAR 2016

1. Introduction

Large igneous provinces (LIPs) are bodies of predominantly mafic igneous rock (typically, basaltic flows, dikes, and sills) erupted in a short magmatic pulse (~1–5 Ma) over an extensive area (>10^5 km^2, Figure 1), and characterized by large volumes of emplaced material (>10^5 km^3) [e.g., *Eldholm and Coffin*, 2000; *Ernst*, 2014]. LIPs have formed in both continental and oceanic environments, and the geologic record of these events extends back to Archean time [*Ernst*, 2014]. LIP events are commonly attributed to thermal plumes emanating from the lower mantle and their arrival and partial melting upon reaching the lithosphere [e.g., *Richards et al.*, 1989], although nonplume mechanisms for the origin of LIPs have also been proposed [see *Ernst*, 2014, chapter 15, for a comprehensive review].

A link between deep mantle plumes and LIPs is supported by the observation that the eruption sites of LIPs for the last 300 Ma, when reconstructed paleogeographically (i.e., based on paleomagnetic data), or in mantle reference frames relying on hotspot tracks, correlate with the edges of two large, nearly antipodal regions of low shear wave velocity in the lowermost mantle, directly above the core-mantle boundary (CMB) [*Burke and Torsvik*, 2004; *Torsvik et al.*, 2006]. These lower mantle structures have been imaged in seismological studies beneath Africa and the Pacific Ocean (Figure 2), and are commonly referred to as the Large Low Shear-wave Velocity Provinces or LLSVPs [*Garnero et al.*, 2007]. The observed correlation [e.g., *Torsvik et al.*, 2006, Figures 4–7] (Figure 2) led *Torsvik et al.* [2006, 2008a, 2008b] to suggest that plumes responsible for the LIP events over the past 300 Ma have preferentially formed along the margins of the two LLSVPs, in zones characterized by the steepest shear wave velocity gradients, later dubbed as the "plume generation zones" (PGZs) by *Burke et al.* [2008]. The correlation maintained over the past 300 Ma also implies that the LLSVPs have been

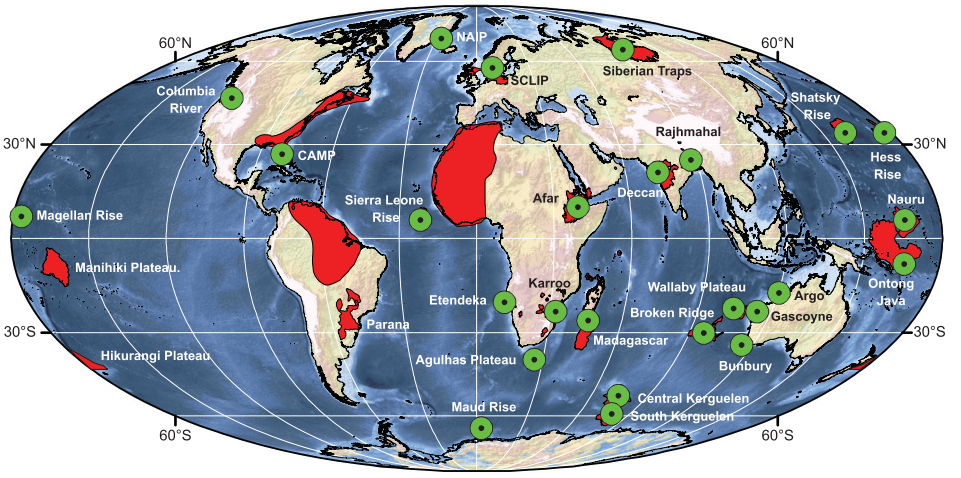

Figure 1. Twenty-six large igneous provinces (LIPs) of the past 300 Ma (15–297 Ma) and their estimated eruption centers (Table 2). Eruption centers are sometimes speculative, and LIPs can also extend over large areas that commonly do not mark the site of a plume from the deep mantle, but the location of material that has propagated horizontally, beneath the lithosphere from a plume [e.g., *Sleep*, 1997]. The location of plume eruption centers at the Earth's surface may therefore not be the places where the plume heads originally impinged the base of the lithosphere. Locating plume eruption centers with higher resolution than about 500 km may be unrealistic. The areal extent of the Central Atlantic Magmatic Province (CAMP) is liberal in the inclusion of 201 million year basalts, sills, and dykes. NAIP, North Atlantic Igneous Province. SCLIP, Skagerrak Centered LIP. Mollweide projection centered at zero meridian (updated from *Torsvik et al*. [2006]).

stable in their nearly present day positions and configurations since Permian time, and a similarly strong correlation between the reconstructed localities of Paleozoic kimberlites and the margins of LLSVPs [*Torsvik et al*., 2010] suggested that this may have been the case since the time of Pangea assembly (~320 Ma) or even before that time, back to the Early Cambrian (540 Ma) [*Torsvik et al*., 2012, 2014].

The African and Pacific LLSVPs are robust features observed in all shear wave tomography models [e.g., *Lekic et al*., 2012, Figure 2]. These structures comprise ~2% of the mantle and occupy nearly 20% of the area at the CMB [*Burke et al*., 2008; *Bower et al*., 2013]. Yet, their geodynamic nature is debated. The LLSVPs have been alternatively suggested to be "piles" of chemically distinct material compared to the bulk of the mantle (thermochemical piles, mantle anchor structures), amalgamations or clusters of thermal plumes, broad thermal "superplume" upwellings, metastable, or thermochemical superplumes [see, e.g., *Bull et al*., 2009, and references therein].

While there is a general agreement that the LLSVPs represent hotter than average regions of the lower mantle, an even hotter debate rages over the question whether their seismic characteristics are better explained by geodynamic models involving chemical heterogeneity [e.g., *Garnero et al*., 2007], or those assuming that the LLSVPs are purely thermal structures [e.g., *Davies et al*., 2012]. We will not reiterate the arguments for the thermochemical or thermal origins of LLSVPs in this paper; relevant summaries can be found in studies of *Torsvik et al*. [2006], *Garnero et al*. [2007], *Burke et al*. [2008], *Davies et al*. [2012], and *Bower et al*. [2013]. We note, however, that if the eruptions of LIPs were indeed triggered by plumes forming along the LLSVP margins, this would provide indirect support for the thermochemical case, in which the LLSVPs are composed of hotter, yet denser material, which stabilizes them on a time scale of hundreds of millions of years, as it is implied by the observed correlation.

When *Burke and Torsvik* [2004] and *Torsvik et al*. [2006] introduced the hypothesis linking LIPs with deep plumes rising from the margins of LLSVPs, the geodynamic community met it with great skepticism. One line of criticism was that it would be difficult to maintain the stability of LLSVPs over a 300 Ma time scale in global convection models [e.g., *Zhang et al*., 2010]. The second argument was that even if the LLSVP structures were stable, it would be conceptually easier to consider the entire area associated with the "hot" LLSVPs as a source region for deep mantle plumes since the mechanism by which plumes would concentrate around the LLSVP margins was not understood. The latter point is particularly relevant for the isochemical convection models that typically reproduce the LLSVPs as clusters of thermal plumes not resolved by seismic tomography [e.g., *Bull et al*., 2009]. This argument motivated several modeling studies [*Tan et al*., 2011; *Steinberger and Torsvik*, 2012; *Hassan et al*., 2015] that demonstrated that the interaction between the

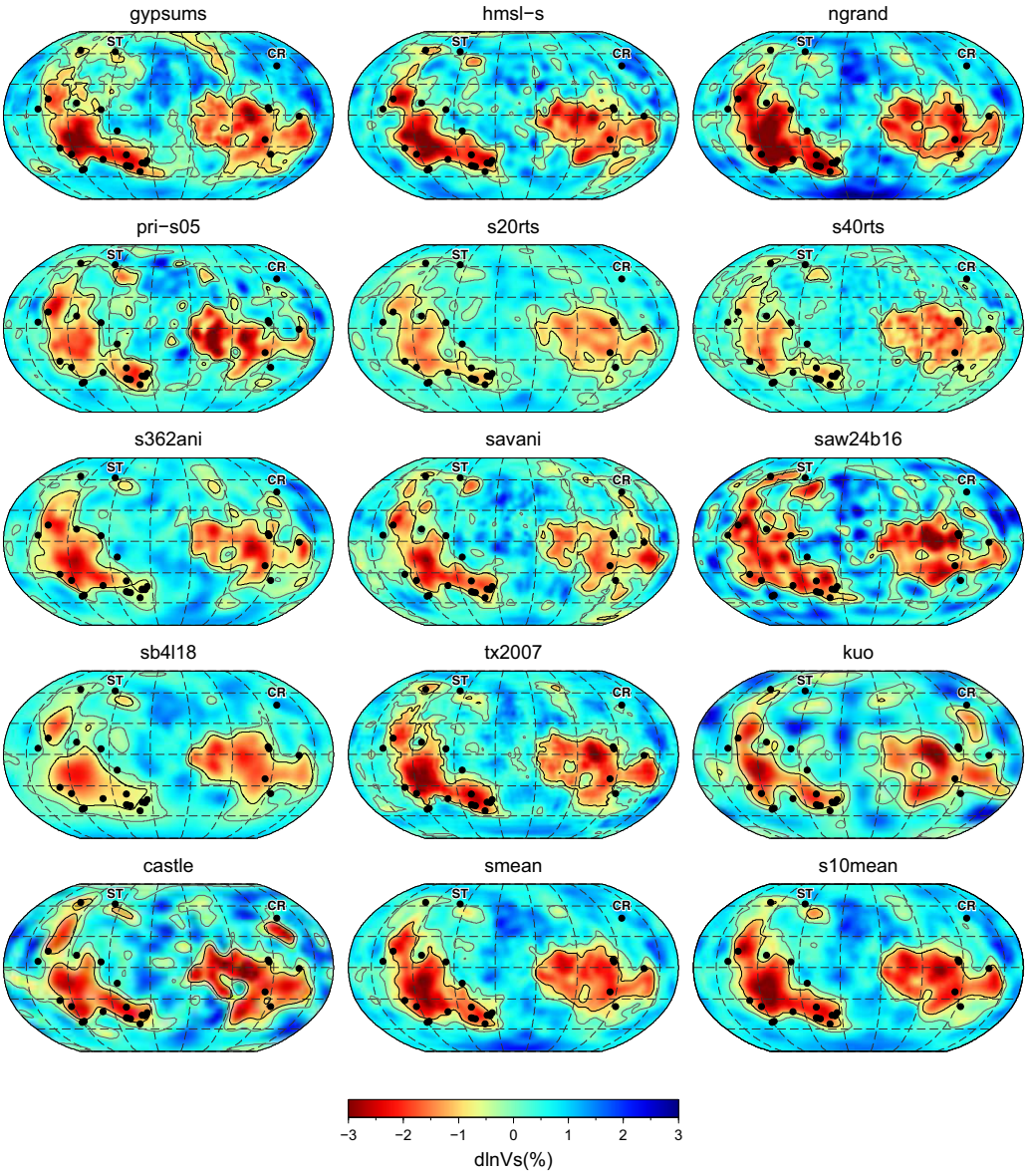

Figure 2. Variations in the shear wave velocity at 2800 km depth inferred from 11 global seismic tomography models (gypsums, hmsl-s, ngrand, pri-s05, s20rts, s40rts, s362ani, savani, saw24b16, sb4l18, tx2007), two D″ models (kuo, castle), and two mean tomography models (smean, s10mean). References for these data sets are given in Table 1. Solid black contours show the boundaries of LLSVPs and smaller-scale low shear velocity structures (Table 1). Solid gray contours outline the areas of slower-than-average velocity. Black dots are the locations of LIPs reconstructed in the DST2015 hybrid reference frame (Table 2). Maps are in the Robinson projection, centered at the 110°E meridian. CR, Columbia River Basalts. ST, Siberian Traps.

ambient mantle flow and the edges of chemically distinct piles comprising LLSVPs can naturally lead to plumes predominantly forming at the margins of these structures.

Since no consensus has yet been reached on which the two scenarios (plumes from the margins versus plumes from the entire LLSVP areas) is more plausible from the geodynamical perspective, it is sensible to take a step back and ask ourselves: if we assume that the LLSVP structures were stable over the past 300 Ma, does the observed distribution of reconstructed LIPs actually require their plume sources to be concentrated around the LLSVP margins? Two recent studies [*Austermann et al.*, 2014; *Davies et al.*, 2015] argued that from a statistical point of view, the observed correlation between the reconstructed locations of LIPs and the margins of LLSVPs may be misleading. Specifically, based on statistical analyses of the distribution of angular distance between the reconstructed LIPs and LLSVP margins, *Austermann et al.* [2014] and *Davies et al.* [2015] suggested that the correlation may be equally well (or better) explained by a distribution of plumes randomly

Table 1. Definitions of the LLSVPs and Slower-Than-Average Shear Wave Velocity Regions

Model	Contour/dlnVs Value (%)	Reference
Global S-wave tomography		
Gypsums	−0.8	*Simmons et al.* [2010]
hmsl-s	−0.9	*Houser et al.* [2008]
Ngrand	−1.1	*Grand* [2002], version of 2010
pri-s05	−0.5	*Montelli et al.* [2006]
s20rts	−0.65	*Ritsema et al.* [1999]
s40rts	−0.5	*Ritsema et al.* [2011]
s362ani	−0.6	*Kustowski et al.* [2008]
Savani	−0.8	*Auer et al.* [2014]
saw24b16	−0.9	*Megnin and Romanowicz* [2000]
sb4l18	−0.6	*Masters et al.* [2000]
tx2007	−0.8	*Simmons et al.* [2007]
D″ models		
castle (sc30lmm)	−0.96	*Castle et al.* [2000]
Kuo	−0.77	*Kuo et al.* [2000]
Mean S-wave tomography models		
Smean	−1.0	*Becker and Boschi* [2002]
s5mean	−0.9	This study
s10mean	−0.9	This study
Cluster analysis		
lekic-5	Voting contour 5	*Lekic et al.* [2012]
lekic-4	Voting contour 4	*Lekic et al.* [2012]
lekic-3	Voting contour 3	*Lekic et al.* [2012]
lekic-2	Voting contour 2	*Lekic et al.* [2012]
lekic-1	Voting contour 1	*Lekic et al.* [2012]

forming over the entire areas of LLSVPs, or the areas of slower-than-average shear wave velocity anomalies in the lowermost mantle. This implies that the observed correlation pattern, where the reconstructed LIPs are distributed along the LLSVP margins (Figure 2), may be just a chance coincidence owing to the stochastic nature of plume generation.

Austermann et al. [2014] concluded that the reconstructed locations of LIPs correlate with both the entire LLSVPs and the margins of these structures, both correlations are significant, but could not be statistically distinguished at the 95% confidence level. *Davies et al.* [2015] were more decisive, concluding that the observed correlation is better explained by a random distribution of plumes over the entire LLSVP areas, although, similar to *Austermann et al.* [2014], they were not able to confidently reject the alternative scenario (plumes from the margins). While the results of *Austermann et al.* [2014] and *Davies et al.* [2015] do support the long-term longevity of LLSVPs, the possibility that the spatial distribution of reconstructed LIPs might have been misinterpreted in terms of the correlation with the LLSVP margins raises the question whether the observed correlation is even relevant for the debate on the geodynamic nature and stability of LLSVPs [cf., *Davies et al.*, 2015].

In this study, we reexamine the proposed alternatives for the spatial correlation between the eruption sites of LIPs and the large-scale seismic structures in the deep mantle through nonparametric goodness of fit tests, using statistics based on comparisons between theoretical probability models and empirical distribution functions (EDF) [*Stephens*, 1974]. Throughout the paper, we will use the term "correlation" in a general sense, meaning that there exists a relationship between the LIPs and LLSVPs that is not expected on the basis of chance alone, as it was shown by *Burke et al.* [2008] and *Austermann et al.* [2014]. This should not be confused with the "correlation" as it is often defined in the statistical literature, meaning a measure of the strength of such association.

In our tests, we used a large number of alternative definitions of the LLSVP margins and the areas of slower-than-average shear wave velocity in order to assess the sensitivity of the results to the uncertainties arising from the limited resolution of seismic inversions; these definitions are described in section 2. In section 3, we discuss the selection of LIPs for the past 300 Ma. Section 4 describes the absolute reconstructions in a hybrid mantle reference frame, which were applied to restore the LIPs to their original eruption sites. In section 5, we present the formulations of probability models representing the proposed distributions of plumes generated (i) over the areas of slower-than-average seismic velocities in the lowermost mantle, (ii) over the areal extent of LLSVPs, and (iii) at their margins. The test procedures are outlined in this section and described in detail in supporting information S1. Section 6 illustrates the statistical analysis with representative examples and presents the result for all models we have tested. Section 7 provides a discussion of the results, comparisons with the earlier studies, and discusses caveats of our analysis. The main conclusions of the study are summarized in section 8.

2. Definitions of LLSVPs

Two large, nearly antipodal regions of slow shear-wave velocity in the deep mantle (LLSVPs) are the most prominent features observed in all shear-wave tomography models (Figure 2). Unfortunately, there is

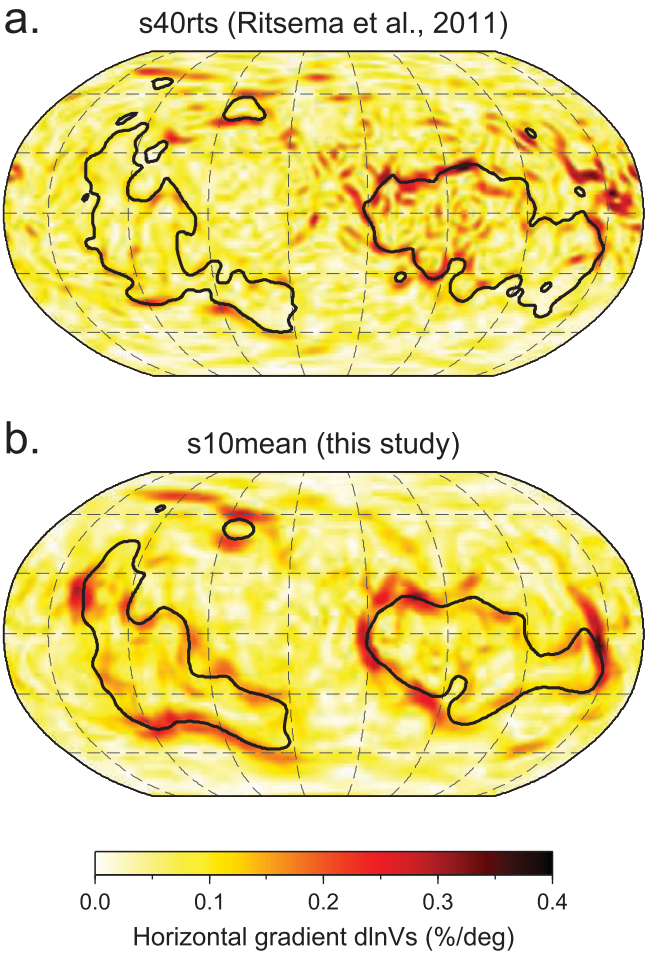

Figure 3. Horizontal gradients of shear wave velocity at 2800 km depth for the s40rts [*Ritsema et al*., 2011] and s10mean tomography models (a and b, respectively). Black lines show the contours of constant dlnVs value (−0.5% for s40rts, and −0.9% for s10mean, Table 1) that correspond to the highest gradients concentrated along at the margins of LLSVPs. Maps are in the Robinson projection, centered at the 110°E meridian.

no general agreement on which of the existing tomography models is the most reliable when it comes to the definition of the areal extent of these features. Because different research groups use different seismic data sets and inversion procedures, there may be substantial discrepancies between individual tomography models, which are most prominent for smaller-scale features that control the exact shapes of the reference seismic contours that we use to define the LLSVPs.

To address this uncertainty, we tested a number of alternative definitions, including those based on (i) eleven individual global shear-wave tomography models, (ii) two D″ models, (iii) three mean tomography models, and (iv) cluster analysis of lower mantle tomography (Table 1).

Following *Torsvik et al*. [2006], we defined the margins of LLSVPs for the individual and mean tomography models by contours of constant dlnVs corresponding to the highest horizontal gradients of shear wave velocity at the 2800 km depth level (dlnVs = dVs/Vs is the velocity perturbation relative to the mean at that depth level). This approach is based on the observation that high horizontal shear wave velocity gradients tend to concentrate along the periphery of LLSVPs (Figure 3), marking the locations of relatively steep "walls" that separate the LLSVP structures from the surrounding mantle, which may be a natural consequence of LLSVPs being composed of chemically distinct material [e.g., *Torsvik et al*., 2006; *Garnero et al*., 2007; *Burke et al*., 2008]. Accordingly, for all individual tomography models, we have examined the maps of horizontal shear wave velocity gradients at 2800 km depth and selected contours of constant dlnVs that matched the steepest velocity gradients observed at the margins of LLSVPs (e.g., Figure 3a). Similarly to *Torsvik et al*. [2006] and *Burke et al*. [2008], we interpret these contours as approximations for the boundaries within the D″ structure that correspond to the loci of plume-generation zones (PGZs).

The dlnVs values for the contours defining the PGZs are presented in Table 1. Note that prior to contouring we had subtracted the mean dlnVs values from the velocity perturbation fields at the 2800 km depth level for each model, which is equivalent to removing the mean offset relative to the 1-D reference model. Thus, the values listed in Table 1 indicate the deviation from the mean, and not from the reference model. With this convention in mind, the contours outlining the slower-than-average velocity regions always correspond to dlnVs = 0. For the D″ models of *Kuo et al*. [2000] and *Castle et al*. [2000], we adopted the contour values from *Torsvik et al*. [2006] and *Burke et al*. [2008] (Table 1).

Figures 4a and 4b show the contours defining the plume generation zones and the slower-than-average shear wave velocity regions for the individual global tomography models and D″ models. Although we observe a general similarity, the discrepancies are also apparent, especially for the contours outlining the slower-than-average velocity regions. This highlights the fact that the definitions of LLSVPs and seismically

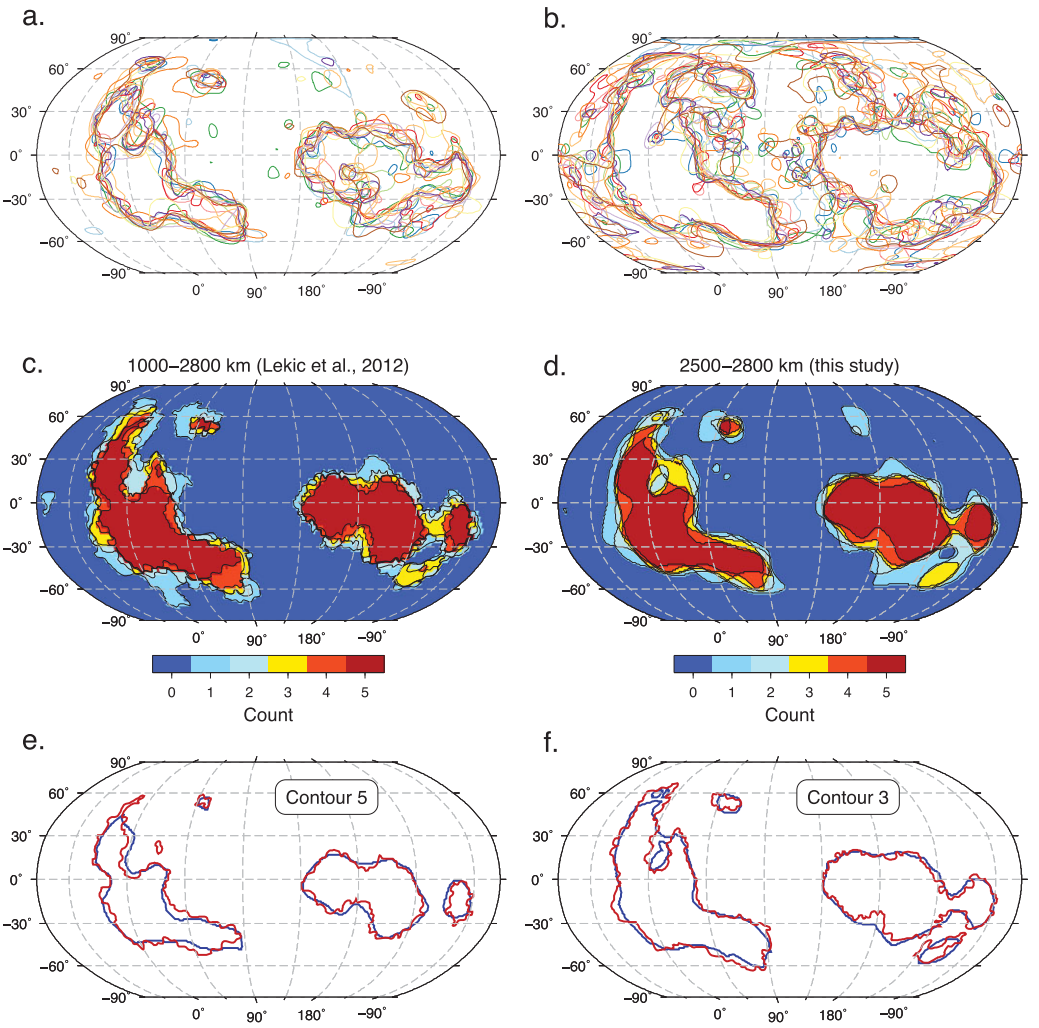

Figure 4. Contours of constant dlnVs defining (a) the boundaries of LLSVPs (Table 1) and (b) the areas of slower-than-average shear wave velocity. (c) Voting map of *Lekic et al*. [2012] based on the cluster analysis of seismic profiles for the 1000–2800 km depth range. (d) Voting map for the 2500–2800 km range. (e, f) Comparisons of voting contours of *Lekic et al*. [2012] (red lines, Figure 4c) with voting contours for the 2500–2800 km depth range (blue lines, Figure 4d).

slow areas may not be constrained by the individual models as tightly as we would like them to be. More reliable definitions can probably be obtained from identifications of velocity structures that are common across tomographic models [see *Lekic et al*., 2012, and discussion therein].

One way of increasing the "signal-to-noise" ratio is a pragmatic approach proposed by *Becker and Boschi* [2002], assuming that averaging several tomography models is likely to enhance the signal from actual velocity structures and dampen the noise (inconsistencies) arising from model artifacts. *Becker and Boschi* [2002] constructed their mean model (smean) by taking a weighted average of three global tomography models, namely, ngrand, s20rts and sb4l18 (Table 1). We followed their procedure to compute two additional mean tomography models: one that included the five S-wave models analyzed by *Lekic et al*. [2012] (gypsums, hmsl-s, s40rts, s362ani, and saw24b16, see discussion below) and all global S-wave tomography models examined in this study, except s20rts that we consider to be superseded by s40rts (Table 1). We will refer to these models as to "s5mean" and "s10mean," respectively. While this exercise may be of limited value because some of the input models are likely not independent, we still consider it useful to test mean models that average a larger number of tomography models. The examples of velocity perturbation fields in s5mean and s10mean, and the comparisons between them and the original smean model, are presented in supporting information Figures S1–S3. The spherical harmonic expansions of the s5mean and s10mean models can be obtained from the corresponding author upon request.

Similarly to *Torsvik et al*. [2006], we adopted the −1% dlnVs contour at the 2800 km depth as a definition of LLSVP boundaries for the smean model of *Becker and Boschi* [2002]. For the s5mean and s10mean models, we selected the −0.9% dlnVs contours based on comparisons with the distribution of shear wave velocity gradients at the margins of LLSVPs (Figure 3b).

Recently *Lekic et al*. [2012] suggested a novel technique for identifying large-scale velocity structures in the lower mantle through cluster analysis of seismic profiles. Cluster analysis is an objective way of grouping observations based on similarity between the group members. When applied to seismic profiles extracted from tomographic models for the depths below 1000 km, cluster analysis reveals that the profiles can be grouped into two families (clusters) characterized, respectively, by faster and slower shear velocities relative to the PREM model [*Dziewonski and Anderson*, 1981], and it is not necessary to invoke a larger number of clusters to capture the first-order structural features. Furthermore, the five shear-wave tomography models analyzed by *Lekic et al*. [2012] (gypsums, hmsl-s, s40rts, s362ani, and saw24b16) display a remarkably consistent pattern of spatial distribution for the two clusters. This can be illustrated with a voting map showing how many models agree on the classification of a geographic location as belonging to the "slow" cluster (Figure 4c). The prevalence of the areas where all models agree on the assignment (i.e., those corresponding to counts 0 and 5, as compared to the intermediate values) indicates a high level of consistency across the models, suggesting that the faster and slower-velocity regions are robustly constrained.

Although the cluster analysis of *Lekic at al*. [2012] relied on seismic profiles for the entire lower mantle (1000–2800 km depths), the areal extent of the fast and slow clusters largely reflects strong velocity anomalies in the lowermost mantle, particularly within 400 km above the core-mantle boundary. To illustrate this, we performed the cluster analysis restricting depths to the 2500–2800 km interval, which produced a voting map (Figure 4d) nearly identical to that for the entire 1000–2800 km range (Figures 4e and 4f). It should come as no surprise then that the slow cluster is spatially confined to the area above the LLSVPs. *Lekic et al*. [2012] also produced an almost identical voting map for the 1800–2800 km depth interval [*Lekic et al*., 2012, Figure 3]. These results are consistent with the strong vertical continuity of seismically slow regions of the lower mantle suggested by radial correlation functions [*Dziewonski et al*., 2010] and can be easily understood considering that the large thermal anomalies of the LLSVPs are likely to produce broad-scale upwellings extending by some 1000 km above them: the hypothesized superplumes or the "Mantle Anchor Structure" of *Dziewonski et al*. [2010].

One notable exception to the general correspondence between the slow cluster and the LLSVPs is a smaller-scale structure to the northeast of the African LLSVP, named the Perm Anomaly by *Lekic et al*. [2012] (Figures 4c and 4d). While significantly smaller, this anomaly has a similar velocity structure and is bounded by similarly steep horizontal velocity gradients as the two LLSVPs. The Perm Anomaly is also clearly identifiable in the s5mean and s10mean models (Figures 2 and supporting information Figures S1–S3).

Based on the discussion above, we argue that the voting contours of *Lekic et al*. [2012] provide more robust definitions of the LLSVP margins and areas of slower-than-average seismic velocity than those derived using individual tomography models. These definitions are especially relevant for testing the hypothesis that LIPs can be sourced from the entire area corresponding to the "footprint" of superplumes, regardless of whether LIP-sourcing plumes originate close to the core-mantle boundary or at some higher level (e.g., at the bottom of the transition zone as was envisioned in the conceptual model of *Dziewonski et al*. [2010]). In this study, we used all five voting contours of *Lekic et al*. [2012]. We note, however, that contours 5 and 4, within which all or nearly all models agree on the definition of a location as belonging to the seismically slow cluster, are arguably the most accurate representations of the LLSVP margins. In contrast, the locations outside contour 3 and within contour 1 correspond to faster-than-average velocities for the majority of models, but at least one of the models characterizes these locations as seismically slow. Hence, contours 1 and 2 may still be applicable as "most generous" definitions of slower-than-average shear wave velocity regions.

3. Present Locations of LIPs

In our selection of LIPs of the past 300 Ma, we closely followed the compilations of *Burke and Torsvik* [2004] and *Torsvik et al*. [2006, 2008a, 2008b], with only few minor revisions (Figure 1 and Table 2). In total, our sample includes 26 LIP events with the eruption ages varying from 15 to 297 Ma. All of them satisfy the definition of *Ernst* [2014] as comprising a large volume of predominantly mafic material emplaced in the

Table 2. Present and Reconstructed Locations of Large Igneous Provinces

				Present Location		Reconstructed (DST2015 Frame)		Reconstructed (S2012 Frame)	
LIP	Code	Affinity	Age (Ma)	Lat. (°N)	Lon. (°E)	Lat. (°N)	Lon. (°E)	Lat. (°N)	Lon. (°E)
Columbia River Basalts	CR	uncertain	15	46.0	241.0	47.66	244.05	46.71	245.20
Afar	AF	African	31	10.0	39.5	5.51	36.12	5.65	35.02
Iceland (NAIP)	GI	African	62	69.9	332.8	63.66	345.73	59.91	354.58
Deccan Traps	DT	African	65	21.0	73.0	−14.83	53.56	−18.64	58.42
Sierra Leone Rise	SL	African	73	6.0	338.0	5.83	327.65	2.72	334.18
Madagascar-Marion	MM	African	87	−26.0	46.0	−42.7	31.21	−40.88	39.42
Broken Ridge	BR	African	95	−30.0	96.0	−49.53	62.67	−46.02	71.26
Wallaby Plateau	WA	African	96	−22.0	104.0	−46.21	77.88	−42.39	85.35
Hess Rise	HE	Pacific	99	34.0	177.0	5.37	220.03	1.98	238.36
Central Kerguelen Plateau	CK	African	100	−52.0	74.0	−48.94	59.2	−45.35	63.32
Agulhas Plateau	AP	African	100	−39.0	26.0	−53.17	359.29	−53.16	6.76
Nauru Basalts	NA	Pacific	111	6.0	166.0	−23.31	218.02	−26.33	232.95
South Kerguelen Plateau	SK	African	114	−59.0	79.0	−48.55	57.61	−51.17	65.15
Rajhmahal Traps	RT	African	118	25.0	88.0	−37.93	60.65	−41.06	67.06
Ontong Java-Manihiki-Hikurangi	ON	Pacific	121	−6.8	167.7	−37.4	229.3	−41.66	240.78
Maud Rise	MR	African	125	−65.0	3.0	−52.15	2.33	−55.66	5.66
Bunbury Basalts	BU	African	132	−34.0	115.0	−54.63	72.87	−54.96	82.27
Parana-Etendeka	PE	African	134	−20.0	11.0	−31.68	345.84	−35.94	349.18
Gascoyne	GS	African	136	−23.0	114.0	−46.32	81.69	−45.66	89.47
Magellan Rise	MG	Pacific	145	7.0	183.0	−1.06	252.45	−21.55	267.55
Shatsky Rise	SR	Pacific	147	34.0	160.0	7.3	218.91	7.72	249.23
Argo Margin	AM	African	155	−17.0	120.0	−43.37	84.22	−40.85	91.61
Karroo	KR	African	182	−23.0	32.0	−37.55	356.84	−44.63	2.77
Central Atlantic Magmatic Province	CP	African	200	27.0	279.0	15.86	337.51	0.52	339.7
Siberian Traps	ST	Perm	251	65.0	97.0	62.42	35.39	65.15	30.56
Skagerrak Centered LIP (SCLIP)	SC	African	297	57.5	9.0	11.73	8.94	11.31	13.44

intraplate setting over a short duration pulse (<1–5 Ma), and hence likely to be sourced by a mantle plume. The locations of LIP centers listed in Table 2 (present location) were estimated from the areal extent of the preserved volcanic and intrusive rocks [e.g., *Eldholm and Coffin*, 2000; *Ernst*, 2014] and referenced to the lithospheric plates on which the LIPs (or their major fragments) presently reside.

We have assigned an older, ~62 Ma age to the Iceland/Greenland LIP (the North Atlantic Igneous Province, NAIP) because it corresponds to the starting magmatic pulse in Greenland and the British Isles [*Ganerød et al*., 2010]. We also used slightly older ages for the Madagascar-Marion and Parana-Etendeka LIPs (87 and 134 Ma, respectively) as these are suggested by newly available radiometric dates [*Torsvik and Burke*, 2015], and added the Agulhas Plateau (~100 Ma) to our compilation [e.g., *Parsiegla et al*., 2008]. Although it was suggested that the Agulhas Plateau was a conjugate to the Maud Rise in Eastern Antarctica, a ~25 Ma younger age suggests that these may represent two distinct LIP events, and we chose to treat them as separate entities. Following a recent reconstruction of *Chandler et al*. [2012], we considered the nearly contemporaneous Ontong Java, Manihiki, and Hikurangi plateaus as fragments of a single LIP (the Ontong Java Nui), which were separated from each other by seafloor spreading in Ellis Basin and at the Osbourn ridge during Cretaceous time (~120–86 Ma). Based on the initial configuration of the conjugate plateaus at ~123 Ma [*Chandler et al*., 2012], we positioned the eruption center of this LIP (referenced to the Pacific plate) at the southeastern edge of the Ontong Java Plateau (6.8°S, 167.7°E, Figure 1).

As an approximation, the LIP centers can be treated as point estimates for the geographic locations where plumes impinged the lithosphere. This does not take into account the possible effects of plume-lithosphere interaction that may lead to a mismatch between the center of eruptive activity and the respective location of the plume, especially for the LIPs that cover vast areas and have erupted into existing rift systems, as for example the Central Atlantic Magmatic Province, CAMP (Figure 1). *Torsvik et al*. [2006] noted that trying to locate the plume positions for such LIPs with the accuracy greater than few hundred kilometers might be not a very sensible thing to do. They also argued that the plume-derived magmas buoyantly rising along the base of the lithosphere toward rifts in the process of "upside-down drainage" [*Sleep*, 1997] are unlikely to travel by more than ~500 km. Several LIPs included in our sample are clearly associated with the volcanic

rifted margins (e.g., CAMP, Parana-Etendeka, NAIP), and uncertainties on the order of several hundred kilometers may be reasonably assigned to their estimated eruption centers. Other LIPs that cover large areas (e.g., Siberian Traps, Ontong Java Nui) may have similar errors associated with their centers. In our statistical analysis, we chose not to treat these uncertainties explicitly, allowing the errors to be built into the probability models. The details of this approach will be fully described in section 5.

4. Reconstructions

In order to test the correlation between the LIPs and large-scale seismic structures in the deep mantle, LIP centers must be restored to their positions at eruption times in the mantle reference frame. To avoid circularity, this reference frame must not use matching LIPs (or kimberlites) as a constraint, i.e. if LIPs were used to calibrate a paleomagnetic frame in longitude so that the LIPs overlie the LLSVPs margin [*Torsvik et al*. 2014], this would not constitute a suitable reference frame for our purposes. For the last ~130 Ma, hotspot reference frames can be used to define the motions of lithospheric plates relative to the mantle, and here we used the most recent global moving hotspot reference frame (GMHRF) of *Doubrovine et al*. [2012]. This kinematic model is based on fitting plate motions that reproduce the observed geometries and age progressions along the best-documented hotspot tracks in the Atlantic, Indian, and Pacific oceans, includes the estimates of hotspot drift due to the effects of mantle convection, and utilizes updated kinematic histories of relative plate motions constrained by marine geophysical data [*Torsvik et al*., 2012].

The GMHRF defines absolute plate rotations in the Indo-Atlantic hemisphere back to 124 Ma, and back to 83.5 Ma in the Pacific region. All Indo-Atlantic LIPs younger than 124 Ma were restored to their original positions using these rotations. For the times prior to 83.5 Ma, the Pacific oceanic plates cannot be linked to those of the Indo-Atlantic realm via a suitable circuit of relative plate motion. The reconstructions of the five Pacific LIPs (99–147 Ma, Table 2) were derived by adding stage rotations of the Pacific plate relative to the Pacific hotspots [*Wessel and Kroenke*, 2008, model WK08-A] to the 83.5 Ma rotation of the Pacific plate relative to the GMHRF. We note that because the *Wessel and Kroenke* [2008] model is based on the assumption of fixed hotspots (i.e., ignores possible hotspot drift), these reconstructions may be less accurate than those in the GMHRF proper.

The remaining LIPs that are older than 124 Ma (all of them are in the Indo-Atlantic hemisphere) were reconstructed using the paleomagnetic reference frame of *Torsvik et al*. [2012], corrected for true polar wander (TPW) and offset in longitude, as explained in the following. These reconstructions are admittedly less reliable than hotspot models because of two important drawbacks that cannot be circumvented without making simplifying assumptions. First, paleomagnetism cannot define absolute longitudes. Second, plate motions are defined relative to the spin axis of the Earth (or the paleoequator), and we do not expect the spin axis to remain fixed relative to the mantle. The process commonly referred to as TPW can occur over geologic time due to redistribution of mass heterogeneities within the convecting mantle, causing a gradual shift of the spin axis relative to the solid Earth (mantle and lithosphere) manifested as migration of the geographic poles [*Goldreich and Toomre*, 1969]. (Actually, if we disregard geologically short-term variations such as precession of the equinoxes, the spin axis remains fixed in inertial space; it is the entire solid Earth that "tips" by rotating about an equatorial axis.) Previous studies [e.g., *Steinberger and Torsvik*, 2008; *Doubrovine et al*., 2012; *Torsvik et al*., 2012, 2014] identified several significant TPW episodes in the Cenozoic and Mesozoic. In order to restore the positions of LIPs in the mantle reference frame, these TPW motions need to be corrected for.

Longitudinal uncertainties of paleomagnetic reconstructions can be minimized by selecting a plate for which we expect minimal longitudinal motion and anchoring paleogeographic reconstructions to that plate. Following *Torsvik et al*. [2008a], we used Africa as the anchor plate because it occupied the central position in the Pangea configuration since 320 Ma and has been predominantly surrounded by spreading ridges since the Pangea breakup. TPW and the absolute longitudes can be easily derived for the last 120 Ma though a comparison of plate motions in the GMHRF and in the paleomagnetic frame. Our analysis shows that over the entire 120–0 Ma period Africa did not move by more than ~12° in longitude (Figure 5a), and the paleogeographic configuration at 120 Ma assuming no longitudinal motion of Africa needs to be shifted by 10.6° to the west (Figure 5b) to match the position suggested by hotspot reconstructions and the estimates of TPW since the Early Cretaceous [*Doubrovine et al*., 2012]. For the 120–300 Ma interval, where we

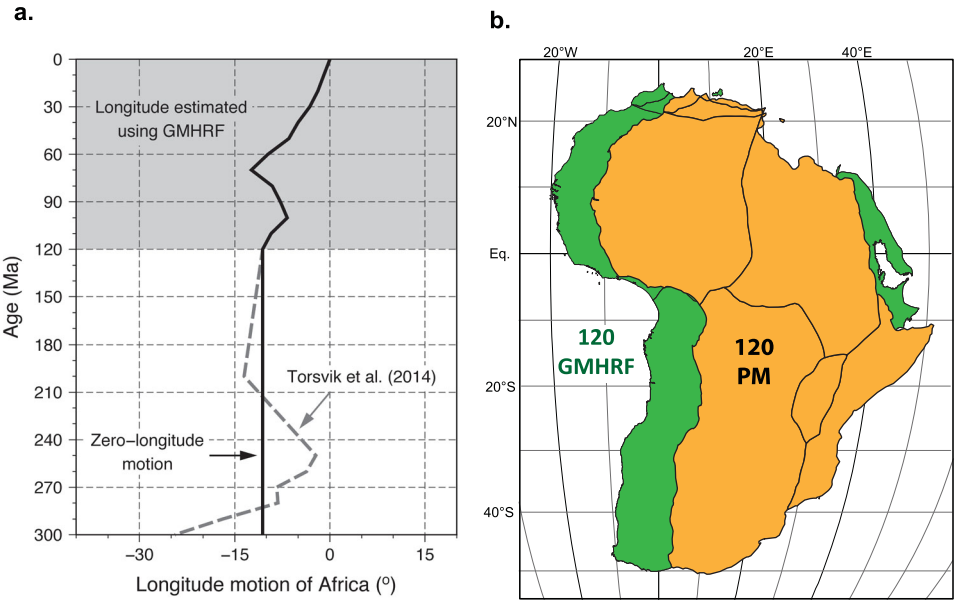

Figure 5. (a) Longitudinal motion of the African plate over the past 300 Ma adopted in this study (solid black line). For comparison, the longitudinal motion of Africa in the paleogeographic reference frame of *Torsvik et al*. [2014] is shown by the dashed, dark gray line. See text for discussion. (b) Example of reconstructing Africa at 120 Ma using the palaeomagnetic reference frame assuming zero longitudinal motion of Africa (PM) [*Torsvik et al*., 2012] and the global moving hotspot reference frame (GMHRF) [*Doubrovine et al* 2012]. The PM reconstruction is similar to the GMHRF reconstruction for Africa except for a ∼11° shift in longitude that is used to correct all PM reconstructions before 120 Ma. It is remarkable that two independent reference frames (based on very different assumptions) produce very similar reconstructions.

have no constraints from the hotspot model, we kept this longitudinal offset constant, allowing Africa to move only in latitude and change its orientation with respect to the spin axis as dictated by paleomagnetic data. In the absence of better constraints, this assumption is the *only* viable option for defining paleogeography without arbitrariness in paleolongitude. We also note that from ∼320 to ∼170 Ma all major continental blocks were assembled in the supercontinent Pangea, so that assuming zero longitudinal motion for Africa imposes minimal longitudinal displacements for other continents during this time interval as well.

To estimate the TPW corrections that needed to be applied to the paleomagnetic reconstructions for the 120–300 Ma interval, we used the method of *Steinberger and Torsvik* [2008]. This approach is based on identifications of time intervals where all continents reconstructed in the paleomagnetic frame show a coherent rotation about a nearly equatorial axis that is close to the center of mass for the entire continental assemblage. Because such rotations imply an unlikely situation when the plate motions are dominated by the toroidal flow [*Lithgow-Bertelloni and Richards*, 1998], they are interpreted as TPW, providing a straightforward way of correcting paleomagnetic reconstructions for TPW motions.

Recently, *Torsvik et al*. [2014] updated the analysis of *Steinberger and Torsvik* [2008], suggesting four prominent episodes of slow, oscillatory TPW during the 120–300 Ma time interval. In this study, the longitudinal motions of Africa in the paleomagnetic frame for the ages older than 200 Ma were adjusted to fit the reconstructed locations of LIPs and kimberlites with the margins of LLSVPs. However, the longitude differences between the model of *Torsvik et al*. [2014] and the "zero-longitude-motion" model adopted here are less than 3° for the 120–200 Ma period, and do not exceed 14° for the 200–300 Ma interval (Figure 5a). Consequently, the analysis of coherent continental rotations in the paleomagnetic frame assuming zero longitudinal motion of Africa for the entire 120–300 Ma time interval suggested TPW rotations that were nearly identical to those of *Torsvik et al*. [2014], and we found no need to modify the TPW model. Accordingly, we adopted the TPW rotations of *Torsvik et al*. [2014] and applied them to the reconstructions in the paleomagnetic frame that assumed no longitudinal motion of Africa over the 120–300 Ma interval. Our main reason for using this rotation model rather than that of *Torsvik et al*. [2014] was to avoid any circularity related to forcing the longitudes to match the LIP eruption sites to the margins of LLSVPs.

In summary, our reconstructions of LIPs are based on the global moving hotspot reference frame back to 124 Ma that was extended back to ~150 Ma for the Pacific region using the hotspot model of *Wessel and Kroenke* [2008]. In the African hemisphere for the 124–300 Ma interval, we used the paleomagnetic reference frame of *Torsvik et al.* [2012] corrected for true polar wander with the TPW estimates of *Torsvik et al.* [2014]. Adjusting the paleomagnetic reconstructions with the ages older than 120 Ma by the 10.6° longitudinal offset results in a seamless transition between the two reference frames. We will refer to this rotation model as the reconstructions in a "hybrid mantle reference frame," which will be denoted as "DST2015" in the remainder of the paper. The locations of LIPs reconstructed in the DST2015 frame are presented in Table 2.

To assess the sensitivity of our analyses to the choice of the rotation model, we also tested the reconstructions in the hybrid reference frame of *Seton et al.* [2012], which will be referred to as "S2012" (Table 2), and in the global paleomagnetic frame [*Torsvik et al.*, 2006, 2008c]. The S2012 rotation model is based on the moving Indo-Atlantic hotspot reference frame of *O'Neill et al.* [2005] for the last 100 Ma and the reconstructions in the global paleomagnetic frame of *Torsvik et al.* [2008c] for the times prior to 100 Ma, which were corrected for TPW using the estimates of *Steinberger and Torsvik* [2008]. The Pacific reconstructions in the S2012 frame differ from those in DST2015, because the former defines the motions of the Pacific plate relative to the plates of the Indo-Atlantic hemisphere back to 83.5 Ma via a plate circuit running through East and West Antarctica, whereas the GMHRF model incorporated in DST2015 uses a plate circuit through Australia and the Lord Howe Rise [see *Doubrovine et al.*, 2012, for discussion of plate circuit alternatives]. Similarly to DST2015, the S2012 model extends the Pacific plate reconstructions back to 150 Ma by adding stage rotations derived from the fixed hotspot model of *Wessel and Kroenke* [2008].

Testing the correlation between the LIPs and LLSVPs in the global paleomagnetic frame [*Torsvik et al.*, 2008c], we used the reconstructed positions of 24 LIPs published in the original studies of *Torsvik et al.* [2006, 2008b]. We note that these reconstructions may not be fully suitable for our analysis because they were not corrected for TPW, and thus record the position of LIPs in a paleogeographic reference frame (relative to the spin axis) rather than in a mantle reference frame. Furthermore, the *Torsvik et al.* [2006] reconstructions of the Pacific LIPs were based on an overly simplified rotation model that assumed no relative motion between the Pacific plate and East Antarctica prior to 90 Ma. Considering a rather complex tectonic history of the southern Pacific region during the 150–90 Ma interval, which involved spreading between several oceanic plates (the Pacific, Phoenix, Manihiki, Hikurangi, Catequil, and Chasca plates) and subduction of some of these plates beneath the Australian-Antarctic margin [*Seton et al.*, 2012], it is very unlikely that the Pacific plate has remained stationary relative to East Antarctica over the entire period. In fact, the absolute plate reconstructions using hotspot reference frames [e.g., *Doubrovine et al.*, 2012; *Seton et al.*, 2012] do not support this assumption. Hence, the positions of the Pacific LIPs reconstructed by *Torsvik et al.* [2006] are likely to be affected by errors arising from the neglect of the Pacific-Antarctic relative motion. Recognizing these limitations, we chose to include the reconstructions of *Torsvik et al.* [2006, 2008a] only for the purpose of comparison with the studies of *Austermann et al.* [2014] and *Davies et al.* [2015], who used this data set in their statistical tests.

5. Statistical Analysis

5.1. Probability Models

We have tested three alternative scenarios for the correlation between the eruption sites of LIPs and deep mantle seismic structures. In the first scenario, we assumed that plume sources of LIPs form randomly over the entire area of the slower-than-average shear wave velocity in the lowermost mantle, approximated by the 0% contour of dlnVs relative to the mean at the 2800 km depth level. The second scenario treats the LLSVPs as broad mantle upwellings (superplumes), regardless of whether these upwellings are anchored by chemically distinct bodies or have purely thermal origin, with LIP-feeding plumes being randomly generated within these regions. We approximated the spatial extent of superplume upwellings by either the areas within the margins of LLSVPs, as defined by the dlnVs contours listed in Table 1, or by voting contours of *Lekic et al.* [2012] that outline the areas belonging to the seismically slow cluster (section 2). The third alternative assumes a thermochemical nature of LLSVP and contrasts to the second scenario in that plumes are generated in the proximity of the LLSVP margins rather than over the entire area above them. We do not consider a uniform distribution over the entire CMB as a plausible model, because earlier studies have firmly

established that the observed distribution of reconstructed LIPs is not consistent with the uniform distribution [*Torsvik et al*., 2006, 2008a; *Burke et al*., 2008; *Austermann et al*., 2014].

It should be pointed out that owing to advection of plume conduits into the mantle flow ("mantle wind"), the deep plume sources are not expected not remain fixed at their initial locations [e.g., *Steinberger and O'Connell*, 1998]. In fact, *Boschi et al*. [2007] found a better correlation between dynamically modeled plume conduits and mantle tomography when the effect of advection was taken into account, compared to the cases in which plume conduits were assumed to be vertical, or anchored to their initial positions. We stress that the alternative modes of correlation tested in our study consider the initial locations of plumes forming over the LLSVPs or at their margins, and not the locations that could have been subsequently changed by the interaction with the mantle flow.

For each alternative, we generated model families of trial distributions on the sphere that exemplify the case. Model A corresponds to the first scenario, in which plumes are generated with uniform probability over the slower-than-average S-wave velocity areas. As discussed in section 3, eruption centers of LIPs may not exactly correspond to the sites where plumes impinged the lithosphere due to the process of "upside-down drainage" of the plume material along the base of the lithosphere. Another process that can modify the surface distribution of LIP centers compared to the distribution of deep sources is the lateral deflection of plumes due to the interaction with the ambient mantle flow during their initial ascent through the mantle. The third source of error is the uncertainties of reconstructions in the hybrid mantle frame (section 4). The mismatches between the LIP eruptive centers and the locations of their deep-seated plume sources corresponding to these three types of error can be of comparable magnitude, and the combined effect may result in displacements on the order of several hundreds to more than a thousand kilometers. To account for these potential biases, we imposed random Gaussian noise on the surface distribution of LIP sources, so that the surface probability density function (PDF) is given by the expression:

$$\rho(\lambda,\varphi)=C\iint\limits_{(\lambda^*,\varphi^*)\in A}\exp\left\{-\frac{\Delta^2(\lambda,\varphi,\lambda^*,\varphi^*)}{2\sigma^2}\right\}\cos\lambda^* d\lambda^* d\varphi^*, \tag{1}$$

where λ is the latitude, φ is the longitude, A is the area of slower-than-average velocity anomalies radially projected to the surface, $\Delta(\lambda, \varphi, \lambda^*, \varphi^*)$ is the angular distance between the points (λ, φ) and (λ^*, φ^*), σ is the level of Gaussian noise, and C is a normalization constant satisfying the equation

$$\int_{-\pi/2}^{\pi/2}\int_0^{2\pi}\rho(\lambda,\varphi)\cos\lambda d\lambda d\varphi=1 \tag{2}$$

To get the intuitive understanding of this model, one can think about the points (λ^*, φ^*) as surface projections of "deep sources" that are uniformly distributed over the slower-than-average velocity areas in the deep mantle. The parameter σ reflects the combined effect of the processes resulting in lateral displacements of eruptive centers with respect to geographic locations of their deep plume sources and reconstruction uncertainties. Because of these displacements and uncertainties, not only the point directly below a reconstructed LIP can be its potential source, but also the points in the vicinity, although as we move farther away from the LIP center (increasing Δ), the contribution from the remote areas decreases sharply following the Gaussian relationship. Accordingly, the probability density at the location (λ, φ) reflects a combination of all possible contributions from the area A, with much greater weights given to the closest locations. It is easy to see that in the limit of $\sigma \to 0$ (no imposed noise), the PDF becomes uniform within the area A (i.e., $\rho(\lambda,\varphi)=1/A$ for $(\lambda,\varphi)\in A$) and zero outside that area. Conversely, as $\sigma \to \infty$, the distribution approaches to the uniform distribution on the sphere.

Model B is similar to model A in design, except plumes are assumed to be randomly generated within the margins of LLSVPs, rather than within the entire area of slower-than-average velocity anomalies that includes the LLSVPs (Figure 2). The area within the margins of LLSVPs can be interpreted as the spatial extent of superplumes forming above the LLSVPs.

For the case in which plumes are assumed to form in the vicinity of LLSVP margins, we tried two alternative formulations (models C and D). Both models assume that the distribution of deep plume sources follows a Gaussian relationship:

$$\rho(\lambda,\varphi)=C\ \exp\left\{-\frac{\delta_c^2(\lambda,\varphi)}{2\sigma^2}\right\}, \tag{3}$$

where $\delta_c(\lambda,\varphi)$ is the angular distance between the point (λ,φ) and the reference contour defining the margins of LLSVPs, and σ describes the "spread" of the distribution. In other words, we do not force the plume sources to be exactly on the contour approximating the margin, but allow for the dispersion relative to this contour. This is equivalent to setting up a "Gaussian band" around the most likely location and treating this band as a plume generation zone, an approach identical to that used by *Austermann et al*. [2014] in their correlation analysis.

Model C corresponds to the simplest form of the PDF, assuming that the geodynamic biases modifying the surface distribution of eruption centers and reconstruction errors combine with the intrinsic dispersion of deep sources in such a way that the surface distribution is also described by equation (3), but with a lager value of σ, reflecting the combined uncertainty. Strictly speaking, this formulation would be valid only in the case when the LLSVP margin is a great circle on the sphere. To account for the more complex geometry of the margins, we formulate model D, defining the PDF by the expression:

$$\rho(\lambda,\varphi)=C\oint_{(\lambda^*,\varphi^*)\in L}\exp\left\{-\frac{\Delta^2(\lambda,\varphi,\lambda^*,\varphi^*)}{2\sigma^2}\right\}dL, \tag{4}$$

where L denotes the surface-projected reference contour, and dL is the element of its length. In a direct analogy with models A and B, the points (λ^*,φ^*) can be thought of as "Gaussian sources" uniformly distributed along the LLSVP margins, all contributing to the PDF. Similarly to model C, the σ parameter reflects the combination of the intrinsic dispersion of deep sources and the dispersion arising from the modification of the surface distribution by the geodynamic processes and the uncertainties of reconstruction. Because the PDF value at a specific geographic location is defined by a combination of contributions from all potential sources, higher probability densities will be observed at the locations "embayed" by the surface-projected contour.

In the remainder of the paper, we will use the same notation (σ) in describing all families of trial distributions. It should be kept in mind that the physical meaning of the σ parameter is different for the "areal" (A and B) and "marginal" models (C and D). For the same level of Gaussian noise, the value of σ will be larger in models C and D because it also includes the intrinsic dispersion of plume sources around the LLSVP margins. Since there is no way to constrain the dispersion without making arbitrary assumptions about the width of the plume generation band in models C and D, we treated σ as a free parameter, exploring the parameter space by varying σ at 1° increments from 0° to 20°. For the consistency of treatment, we used the same approach when generating trial distributions for models A and B.

We note that the use of large σ values in models A and B would be equivalent to overestimating the uncertainties of plume locations. If we assume that the mean deviation of reconstructed LIPs from their surface-projected plume sources is on the order of 500 km (~4.5°), the corresponding σ value should be set to approximately 4° (since the expected value of absolute deviation $=\sigma\sqrt{\pi/2}$ for an azimuthally symmetric bivariate Gaussian distribution, $\sigma=4.5^\circ/\sqrt{\pi/2}=3.6^\circ$). Allowing for more generous errors corresponding to a mean deviation of 1000 km will increase σ to about 7°, but the values above that would be realistic only for models C and D, in which we have an additional contribution due to the intrinsic uncertainty of plume generation zone. We also note that the use of a single σ value is an approximation because there is no reason to expect the errors of estimated locations to be identically distributed for all LIPs. However, taking into account that all probability models described here are not more than idealizations, we consider our simple formulations appropriate for the purpose of our study. Developing a more rigorous error budget will be a part of future research.

5.2. Statistical Tests

Making statistical decisions on how well the proposed probability models fit the spatial distribution of reconstructed LIPs would ideally require a formal test for goodness of fit on the sphere. Because all probability models tested in this study reflect rather complex shapes of the contours defining the LLSVP margins or slower-than-average velocity areas, no simple parametric test on the sphere [e.g., *Fisher et al*., 1987] could have been sensibly used. If a similar problem arose in a one-dimensional case, that is, we wanted to decide

on whether a data sample can be drawn from a probability distribution that can be specified precisely, but does not belong to any class of simple parametric models, the standard way of solving it would be performing a goodness of fit test based on the empirical distribution function (EDF) [*Stephens*, 1974]. However, the EDF tests use cumulative distribution functions (CDF) rather than PDFs, and there is no unique way of ordering the observations in two dimensions to produce the 2-D equivalents of CDFs and EDFs. Although several studies proposed some generalizations of EDF tests for distributions on the plane [e.g., *Fasano and Franceschini*, 1987; *Justel et al*., 1997], no practically useful generalization currently exists for distributions on the sphere.

Here we adopted a simpler approach. Instead of comparing two-dimensional distributions, we analyzed the distributions of a one-dimensional variable derived from the full 2-D model. Since we are mainly concerned with discriminating the cases when LIPs correlate with the LLSVPs as a whole or with their margins, a natural choice of such variable is the angular distance to the reference contour that defines the areas of slower-than-average shear velocities in the lowermost mantle (model A) or the margins of LLSVPs (models B-D).

The empirical distribution functions were defined as follows. First, we calculated the angular distances δ_i between the reconstructed LIPs and a reference contour relevant for the probability model being tested. We considered the value of δ_i to be positive if the reconstructed location was inside the contour, and negative if it fell outside. Next, we arranged the signed angular distances in ascending order, $\delta_1 \leq \delta_2 \leq \ldots \leq \delta_n$ where n is the number of LIPs, and defined a stepwise function $S_n(\delta)$ that gives the fraction of data points to the left of a given value δ, which is the EDF.

$$S_n(\delta)=0 \text{ if } \delta < \delta_1 \tag{5.1}$$

$$S_n(\delta)=\frac{i}{n} \text{ if } \delta_i \leq \delta < \delta_{i+1} \text{ for } i=1,\ldots,n-1 \tag{5.2}$$

$$S_n(\delta)=1 \text{ if } \delta \geq \delta_n \tag{5.3}$$

The cumulative distribution function $F(\delta)$ corresponding to the proposed probability model was evaluated by integration of the model PDF over the area on the sphere satisfying the condition that the signed angular distance between the points within this area and the reference contour did not exceed a given value of δ, i.e.,

$$F(\delta)=\iint\limits_{\delta_c(\lambda,\varphi)\leq\delta} \rho(\lambda,\varphi)\cos\lambda \; d\lambda \; d\varphi, \tag{6}$$

where $\delta_c(\lambda,\varphi)$ is the angular distance between the point (λ,φ) and the reference contour.

Under the null hypothesis that the observed sample of reconstructed LIPs can be drawn from the proposed distribution, $S_n(\delta)$ is the unbiased estimator of $F(\delta)$, so that a clear mismatch between them signals that the proposed distribution does not fit the observations and the parent probability model on the sphere should be rejected.

A statistical decision on whether the discrepancies between the EDF and the proposed CDF are significant to warrant the rejection of the null hypothesis can be made by performing formal tests for the goodness of fit using different measures of the overall difference between the two distribution functions, which are commonly referred to as "EDF statistics" [*Stephens*, 1974]. Here we used five well-studied EDF statistics, namely, the Kolmogorov-Smirnov (D), Cramer-von Mises (W^2), Kuiper (V), Watson (U^2), and Anderson-Darling (A^2) statistics. The derivations of these statistics are given in supporting information S1. In all tests, we calculated the modified EDF statistics (adjusted for the sample size) and compared them with the upper critical values of their null distributions (percentage points that correspond to the selected level of statistical significance). We chose to perform all tests at a 5% significance level (the 95% confidence level). If the modified test statistic for a particular test exceeded its critical value, the test indicated a significant difference between the EDF and the proposed CDF. We also estimated probabilities of observing the values of EDF statistics as large or larger than those actually observed in our experiments using Monte Carlo simulations. The critical values of the modified test statistics and details on the estimation of probabilities are presented in supporting information S1.

It is important to note that the statistics D, V, W^2, U^2, and A^2 have different power properties [*Stephens*, 1974]. The statistics D, W^2, and A^2 are more powerful in detecting differences between the mean of the hypothesized population and the sample mean when the population and sample variances are similar. In contrast, when the population and sample means are similar, the statistics V and U^2 would detect a disagreement between variances better than the others. Monte-Carlo studies [*Stephens*, 1974, and references therein] have suggested that the statistics W^2 and A^2 tend to perform slightly better than D, and the statistic U^2 tends to be slightly more powerful than V. Taking into account fairly strong correlations between D, V, W^2, U^2, and A^2, we will adopt the following terminology when describing the results of our statistical tests in the remainder of the paper. When all five tests based on different EDF statistics did not indicate a significant discrepancy at the 5% probability level, we will term the probability model as "consistent" with observations and the fit between the model CDF and the EDF as "acceptable," meaning that the null hypothesis that the sample of reconstructed LIPs is drawn from the proposed distribution cannot be rejected with the 95% confidence. In the remaining cases, when one or more EDF tests indicated significant discrepancies, we will consider that the evidence against the model is sufficiently strong to warrant its formal rejection at the 95% confidence level.

Two-sample permutation tests for a common distribution were performed for the samples of angular distances for 19 LIPs correlated with the African LLSVP and five LIPs related to its Pacific counterpart (Table 1). The null hypothesis tested in these experiments assumes that the samples can be drawn from the same parent distribution, but the distribution itself is unspecified. The tests were based on the use of two-sample equivalents of the five EDF statistics described above: the Kolmogorov-Smirnov ($D_{n,m}$), Kuiper ($V_{n,m}$), Cramer-von Mises statistic ($W^2_{n,m}$), Watson ($U^2_{n,m}$), and Anderson-Darling ($A^2_{n,m}$) statistics [*Kuiper*, 1960; *Anderson*, 1962; *Watson*, 1962; *Pettitt*, 1976] (see supporting information S1 for derivations). The design of the two-sample test is similar to the one-sample case, except instead of comparing a EDF with a proposed theoretical model $F(\delta)$, we analyze the difference between the two empirical distribution functions, $S_n(\delta)$ and $S_m(\delta)$, defined on n observations of the first sample, and m observations of the second sample, respectively. The two-sample EDF statistics measure dissimilarity between the EDFs, and the null hypothesis can be rejected when the observed statistic is greater than the critical point of its null distributions, corresponding to the selected significance level. Because the sizes of the African and Pacific samples are not large, it was possible to conduct exact tests [e.g., *Sprent and Smeeton*, 2007], in which the pooled sample (combined data for the African and Pacific LIPs) was resampled by generating all possible combinations of n (and the remaining m) out of its $n+m$ elements, and the probability of the two-sample statistic equal to or larger than the observed value was calculated as described in supporting information S1. If the exact probability values were below 0.05, we considered that the differences between the two EDFs were significant at the 5% level, and the null hypothesis should be rejected.

Our statistical analysis has many similarities with the earlier studies [*Austermann et al*., 2014; *Davies et al*., 2015], but differs in some important aspects. Unlike *Austermann et al*. [2014], we did not reduce the observations to a single parameter (the mean angular distance), but calculated the EDFs based on the whole sample of observed distances. This allowed us to evaluate the mismatch not only around the mean, but also in the distribution tails. By using signed angular distances, we avoided the distribution folding back about the zero value, which gave us a better resolution and was especially useful for testing "categorical" models that assume that plumes are generated within certain areas, but not outside those areas (models A and B with $\sigma=0^\circ$).

Davies et al. [2015] performed "regional" Kolmogorov-Smirnov tests for the goodness of fit between the African and Pacific LIPs and their respective LLSVPs, which were completely analogous to our experiments. The Kolmogorov-Smirnov statistic is known to be less powerful than the Kuiper and Watson tests in discriminating between distributions with similar means (medians) but different spreads (variances). Furthermore, the Kolmogorov-Smirnov test is not sensitive to the presence of one or few discordant observations, those that are simply impossible under the null hypothesis. To illustrate that, consider a proposed distribution that is entirely contained within the interval $[\delta_{\min}, \delta_{\max}]$, i.e., the values of $F(\delta)$ are exactly 0 to the left of $\delta_{\min}$, and 1 to the right $\delta_{\max}$. Clearly, a single observation outside this interval would rule out such distribution. Yet, the Kolmogorov-Smirnov test (as well as the Kuiper, Cramer-von Mises, and Watson tests) would generally require many data points in the "forbidden region" before signaling a discrepancy. In contrast, the Anderson-Darling statistic becomes infinity even if only a single discordant observation is present. It is not

wise to place all bets on any particular EDF test when potentially more sensitive statistics are available. Hence, we chose to conduct all five EDF tests for each specific comparison and base our decisions on the consideration of all of them.

6. Results

6.1. Examples

Before presenting the results of our analysis for all alternative models that we have tested, we feel it would be instructive to illustrate our statistical approach with some representative examples. For these examples, we chose two global S-wave tomography models, namely, s362ani and s40rts [*Kustowski et al*., 2008; *Ritsema et al*., 2011], to define the areas of slower-than-average shear wave velocity and the boundaries of LLSVPs (section 2). We chose these models because both of them are recent and apparently well constrained, and yet the small-scale differences in the distribution of seismic velocity anomalies between these two models are large enough to result in some contrasting statistical decisions. The EDF tests described in these examples are based on the reconstructions in the DST2015 hybrid mantle reference frame (section 4, Table 2).

For the s362ani tomography model, a clear distinction between the alternative scenarios for the formation of LIP-sourcing plumes can be observed in the results of EDF tests (Figure 6). No model generating plumes within the areas of slower-than-average shear wave velocity (model A, Figures 6a and 6b), or within the area of LLSVPs (model B, Figures 6c and 6d), produced an acceptable fit to the distribution of reconstructed LIPs at the 5% significance level, regardless of the level of Gaussian noise imposed on the distribution. In contrast, fully acceptable fits were found for models generating plumes from the margins of LLSVPs (models C and D, Figures 6e–6h).

Specifically, for model A at low values of σ (Figures 6a and 6b), a mismatch between the EDF and CDF in the upper tail of the distribution (angular distance $> \sim 10°$) was sufficiently large for at least one of the five EDF tests to report a significant difference at the 5% level. This mismatch reflects a simple observation that the LIPs reconstruct to the periphery of the slower-than-average velocity areas (i.e., close to the LLSVP margins), rather than being randomly distributed over the entire area. Increasing the value of σ led to progressively higher spreads of the model distribution, and the Kuiper and Watson tests (V and U^2) detected significant discrepancies for all σ values $>3°$. Hence, model A can be rejected at the 95% confidence level.

Restricting plume generation to the area within the margins of LLSVPs (model B, Figure 6c) resulted in even larger mismatches, with the model CDFs notably shifted toward the higher values of angular distance compared to the EDF (Figure 6d). For the values of σ smaller than 18°, all five EDF tests showed a significant difference. For the higher values, the mean of the model distribution became close to the sample mean but the variance got significantly larger, and at least two out of five tests detected a significant discrepancy. Accordingly, model B can also be rejected.

Unlike the unacceptable misfits of "plumes-from-the-area" models (A and B), the models assuming that plumes form at the margins of LLSVPs (C and D) produced good fits to the observed distribution for intermediate values of σ ($\sim$6°–10°) (Figures 6e–6h). Although model D performed slightly better than model C (i.e., produced lower values of the test statistics), both of them capture the shape of the empirical distribution well, and none of the five EDF tests indicated significant discrepancies at the 5% level (Figures 6f and 6h). For this example, we can conclude that there is sufficiently strong evidence against the hypotheses assuming that the LIPs are randomly generated over the areas of slower-than-average shear wave velocity in the lowermost mantle (model A), or the areas inside the margins of LLSVPs (model B), to warrant their rejection at the 5% significance level. In contrast, there is no evidence against the alternative hypothesis assuming that LIPs are sourced by plumes from the LLSVP margins.

With the definitions of the LLSVPs and slower-than-average velocity areas based on the s40rts tomography model (Figure 7), the EDF tests for models B, C, and D were similar to those discussed in the earlier example (s362ani). Model B (Figures 7c and 7d) showed a significant misfit to the observations at the 5% level for all values of σ, suggesting that it should be rejected, whereas models C and D (Figures 7e–7h) produced fully acceptable fits for the σ values between $\sim$6° and 10°. However, in contrast to the s362ani example, model A also resulted in acceptable fits for σ ranging from 2° to 12°. Examining the overall best-fit model shown in Figures 7a and 7b ($\sigma = 6°$), it is apparent that the model distribution has a longer upper tail than the empirical distribution. This result is qualitatively similar to the comparison shown in Figure 6b, but the discrepancy

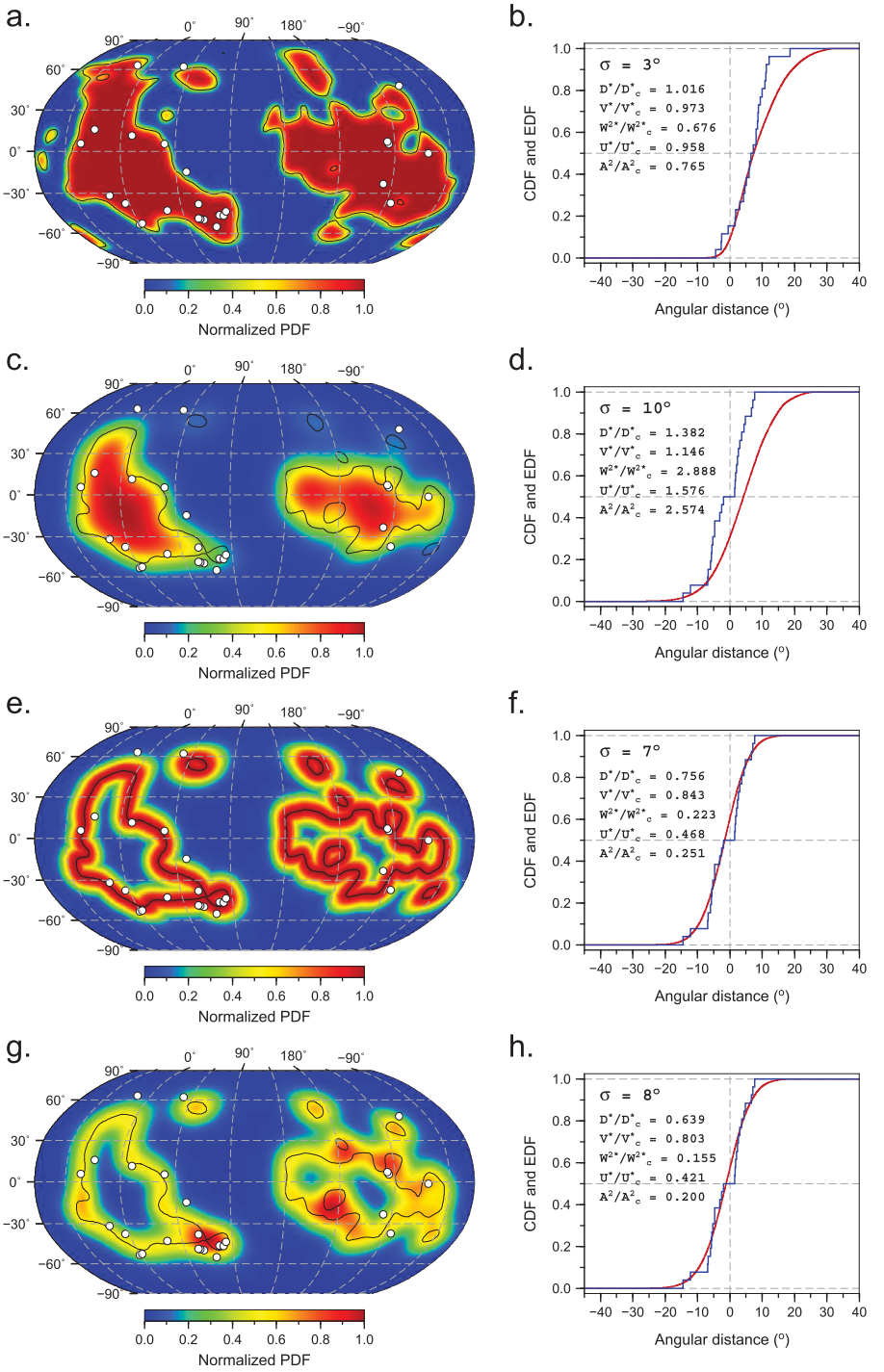

Figure 6. Examples of EDF tests for the LLSVP definitions based on the s362ani tomography model [*Kustowski et al.*, 2008]. Maps show the surface probability density for (a) model A with $\sigma=3^{\circ}$, (c) model B with $\sigma=10^{\circ}$, (e) model C with $\sigma=7^{\circ}$, and (g) model D with $\sigma=8^{\circ}$. The probability density is normalized to its maximum value for consistency of the color scheme. White circles show the reconstructed locations of LIPs in the DST2015 reference frame (Figure 2 and Table 2). Maps are in the Robinson projection, centered at the 110°E meridian. Figures 6b, 6d, 6f, and 6h show the comparisons of the EDFs (blue stepwise lines) and the CDFs (red curves) calculated using the probability densities plotted in Figures 6a, 6c, 6e, and 6g, respectively. The modified statistics for the five EDF goodness of fit tests, divided by their critical values for the 5% significance level (supporting information S1) are quoted; ratios exceeding 1 indicate a significant discrepancy.

is not large enough for any of the five tests to indicate significance at the 95% confidence level. In this example, two alternative hypotheses that pass the EDF tests (plumes from LLSVP margins and plumes from the entire seismically slow areas) have to be retained at the 95% confidence level.

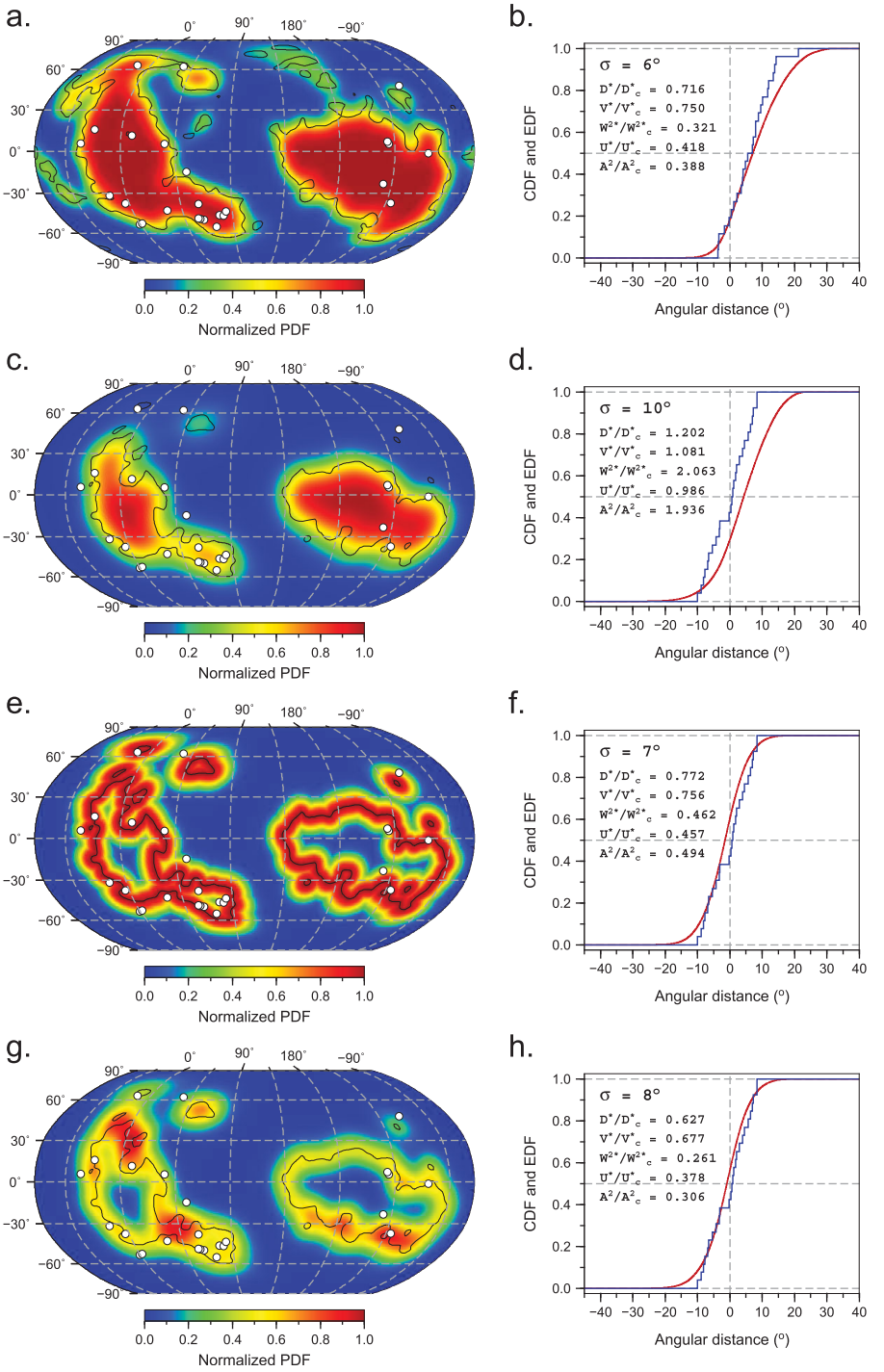

Figure 7. Examples of EDF tests for the LLSVP definitions based on the s40rts tomography model [*Ritsema et al.*, 2011]. (a, b), model A (σ=6°); (c, d) model B (σ=10°); (e, f) model C (σ=7°); and (g, h) model D (σ=8°). See the caption of Figure 6 for description.

6.2. Correlation on a Global Scale

Here we present the results of our statistical analyses for the entire Earth, assuming that the observed angular distances relative to the reference contours defining the margins of LLSVPs (or smaller-scale structures like the Perm Anomaly, section 2), and relative to the contours delineating the areas of slower-than-average seismic velocity in the lowermost mantle, conform to the same distribution regardless of regional affinity of the reconstructed LIPs. The question of whether this assumption is statistically justified will be addressed in section 6.3.

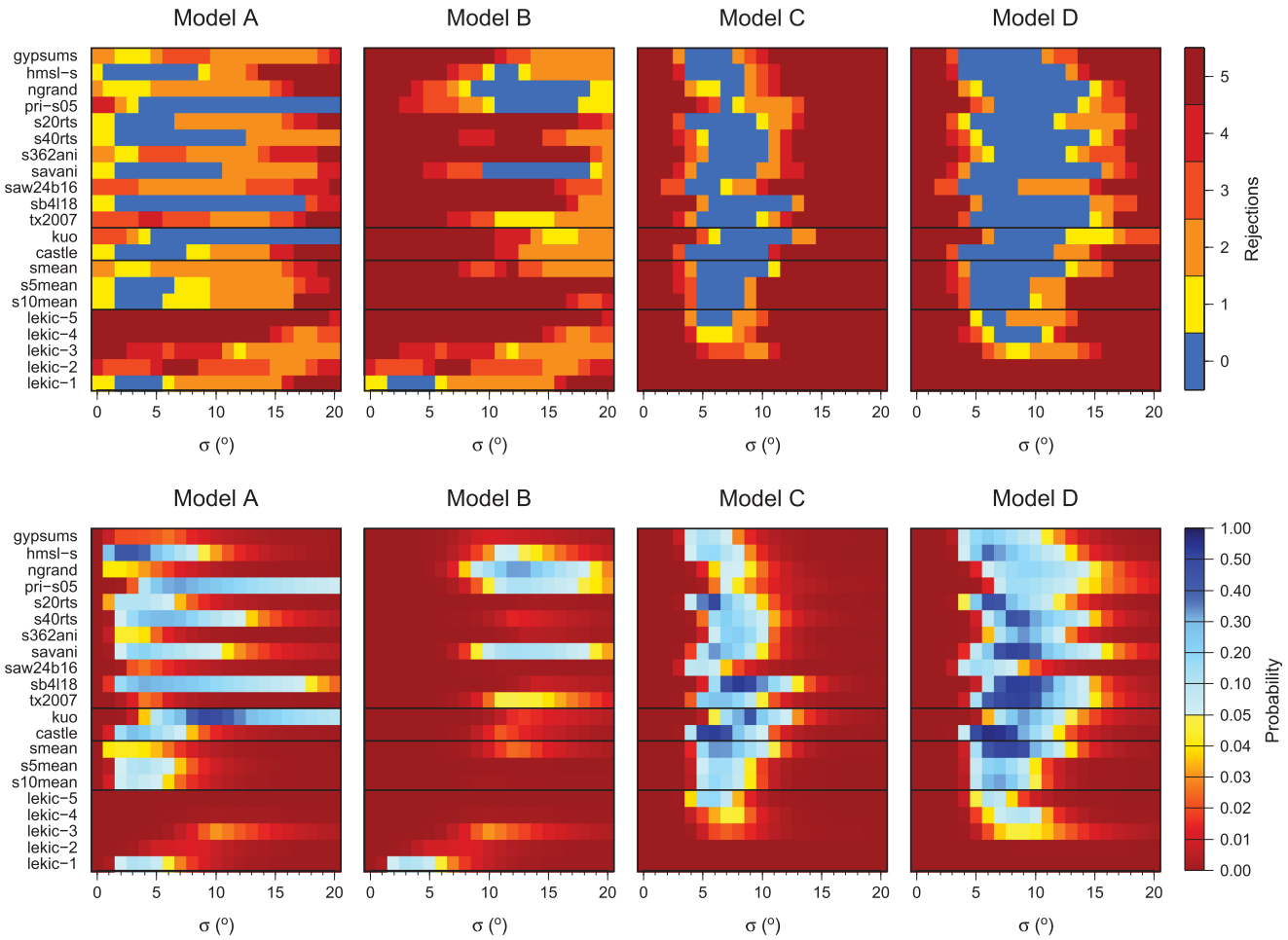

Figure 8. Summary of goodness of fit tests based on the reconstructions of LIPs in the DST2015 hybrid mantle reference frame (global correlation). (top row) The maps depicting the number of EDF tests that indicated a significant discrepancy at the 5% significance level. The blue color corresponds to acceptable fits (none of the five EDF tests signals a significance). (bottom row) The probability for one of the five EDF tests, which indicated the highest significance for the discrepancy between the EDF and the model CDF (see text).

We begin with the discussion of tests, in which the eruption centers of LIPs were reconstructed using our preferred rotation model, the DST2015 hybrid frame (section 4, Table 2). The results of these comparisons are summarized in Figure 8 as color maps showing the number of EDF tests that indicated a significant discrepancy between the model CDF and the observed distribution (EDF) at the 5% level. For each of the five statistical tests (section 5.2), we calculated the probability of observing a value as large or larger than the statistic calculated using the EDF and CDF curves, under the hypothesis that the sample of angular distances defining the EDF comes from the distribution specified by the model CDF (supporting information S1). The lowest of these five probability values (i.e., the probability for the test showing the highest significance) is also plotted in Figure 8.

In the majority of comparisons, we chose to exclude the Columbia River Basalts (CRB, labeled "CR" in Figure 2) as a discordant observation, i.e., an observation that can be judged unreasonable on the basis of the proposed probability model. The reconstructed position of this LIP projects onto the area of faster-than average shear wave velocity in the lowermost mantle for most models (Figure 2), far away from the margins of LLSVPs as they were defined in section 2. Only in the comparisons using the s40rts, s362ani, savani, kuo and castle models the CRB datum was retained. Similarly, we have excluded the Siberian Traps (ST, Figure 2) from testing the models based on the s20rts, kuo, and smean definitions of the LLSVP margins. We note that the EDF tests are generally not sensitive to the presence of one outlier in the sample, and repeating the tests with the CRB and/or ST included produced nearly identical results for all comparisons (supporting information Figure S4). The main reason for excluding outliers was to make the results of our analysis as

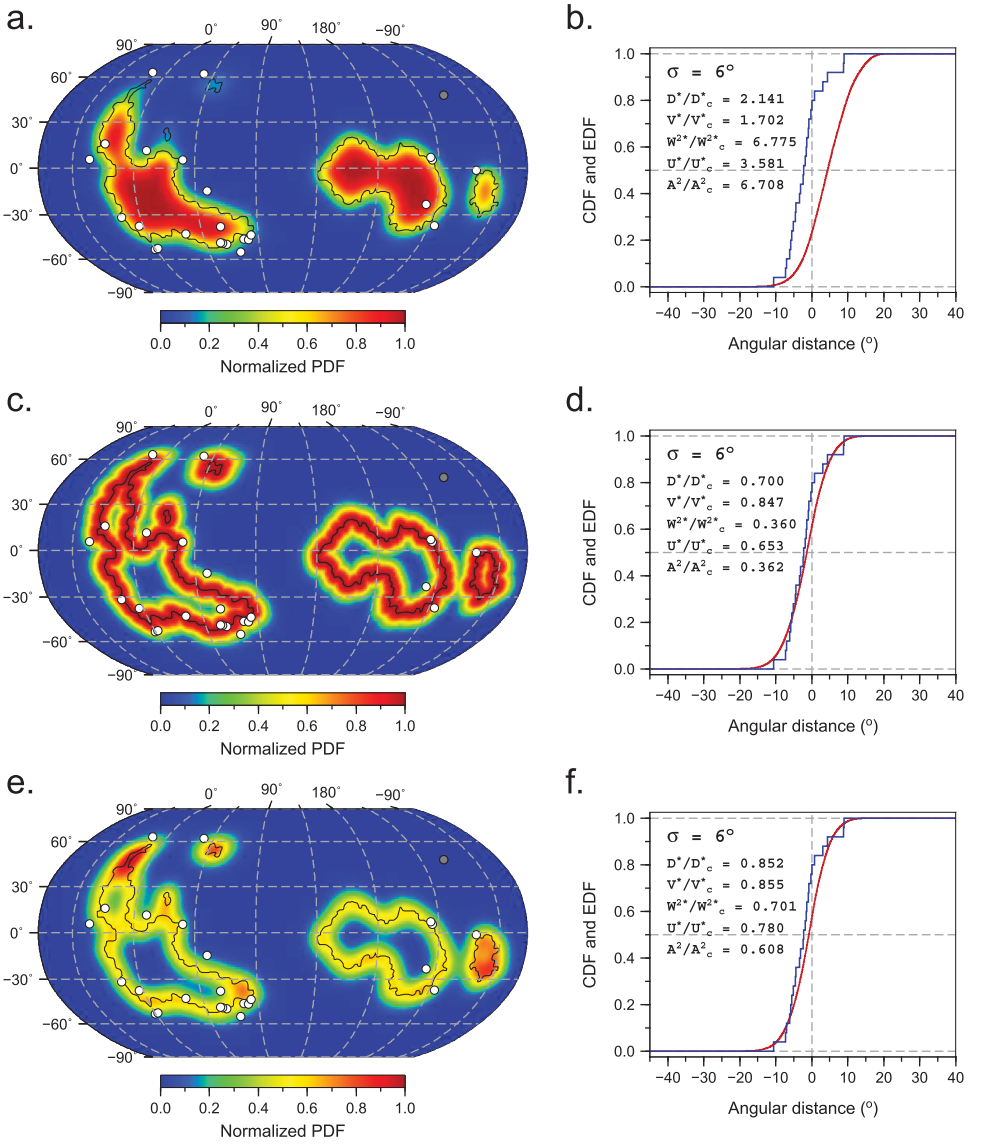

Figure 9. Examples of EDF tests for the LLSVP definitions based on voting contour 5 of *Lekic et al*. [2012]. (a, b), model A/B; (c, d) model C; (e, f) model D; σ=6° in all examples. The Columbia River Basalts (gray circle in Figures 9a, 9c, 9e) was not used to constrain the EDFs in these comparisons. See the caption of Figure 6 for description.

meaningful as possible given this known weakness of EDF tests. To illustrate this, consider the comparisons using the *Lekic et al*. [2012] voting contour 5 as a definition of LLSVP margins (Figure 9). The CRB datum (plotted as a gray dot) is 45° away from the nearest margin, whereas the remaining data (white circles) are within ±10° degrees from the margins. If we assume that the CRB comes from the same population as the remaining data, and that the population is defined by the model CDF (red curves in Figures 9b, 9d, and 9f), the probability of this LIP to be as far from its potential source region is extremely low (less than 10^{-8} with the accuracy we evaluate the CDF). If we stick to the assumption that the outlier belongs to the population, we will have to reject all models. Yet, the EDF tests alone would not be sensitive to the presence of the outlier, suggesting us that models C and D are acceptable for the σ values between 6° and 8° (supporting information Figure S4), so that the test results would be misleading. Alternatively, we can reasonably assume that the CRB datum is not related to the LLSVPs and exclude it from the LIP sample, which would result in a more meaningful comparison because the outlier no longer "disqualifies" the models (Figure 9). While it could be argued whether the outliers should or should not be excluded, we stress that the results of our EDF tests were not affected by our choice, and the conclusions we have reached would be the same even if we decided not to exclude the outliers.

For model A (plumes generated randomly over the area of slower-than-average seismic velocity), the test results were not consistent across individual tomography models used to define seismically slow areas (Figure 8). While we found acceptable fits for six tomography models (hmsl-s, pri-s05, s20rts, s40rts, savani, and sb4l18) at various levels of Gaussian noise, no fully satisfactory fits were observed for the entire range of the model parameter σ for the remaining five models (gypsums, ngrand, s362ani, saw24b16, and tx2007), with at least one of the five EDF statistics suggesting a significant departure at the 5% level. For the two D″ models [*Kuo et al*., 2000; *Castle et al*., 2000], we found the fits acceptable, but for different (although overlapping) ranges of the σ parameter. For the mean S-wave tomography, the smean model of *Becker and Boschi* [2002] did not produce acceptable fits, while the definitions based on the s5mean and s10mean models resulted in acceptable fits for low levels of Gaussian noise (σ = 2-6° and 2-5°, respectively). All models using voting contours 2 through 5 of *Lekic et al*. [2012] resulted in unacceptable misfits, whereas contour 1 produced acceptable fits for σ = 2–5°.

More conclusive results were obtained in testing model B, in which the plume generation area was restricted by the margins of LLSVPs (Figure 8). As it was already mentioned, the smaller areal extent of the LLSVPs compared to the overall extent of regions characterized by slower-than-average shear wave velocity (Figure 2) resulted in the model CDFs notably shifted to the higher values of angular distance with respect to the EDF, producing a much clearer separation between the two distribution functions (e.g., Figures 6c–6d and 7c–7d). Consequently, no acceptable fits were observed for the majority of LLSVP definitions. For four individual tomography models (hmsl-s, ngrand, pri-s05, and savani), we did find several probability models consistent with the observed distribution, but all of them required imposing high levels of Gaussian noise ($\sigma > \sim 10°$) that we do not consider realistic (section 5.1). The definitions of LLSVPs based on the D″ models (kuo and castle) and the mean tomographic models (smean, s5mean, s10mean) produced probability models that were clearly inconsistent with the observed distribution of the angular distance. This was also the case for the definitions based on voting contours 2–5 of *Lekic et al*. [2012] (e.g., Figures 9a and 9b).

For the models assuming that plumes form in the vicinity of LLSVP margins (models C and D), fully acceptable fits were observed for nearly all definitions of the boundary contours at intermediate values of the σ parameter, commonly between σ = 5° and 10° (Figure 8). Interestingly, model D, which we consider to be more realistic (section 5.1), generally preformed better than model C in fitting the observed distribution of angular distance, with somewhat wider regions of the acceptable solutions in the σ parameter space.

Among the LLSVP definitions based on the cluster analysis of *Lekic et al*. [2012], the use of voting contour 5 produced model distributions that fitted the observations well (e.g., Figures 9c–9f). For contour 4, only model D with σ = 7–10° produced an acceptable fit, whereas all models based on contours 3, 2, and 1 were not consistent with the reconstructed locations of LIPs. The reason for that can be easily understood by examining Figure 4c, which shows that as the voting count decreases, the area outlined by the contour becomes larger, so that the reconstructed LIP locations tend to be on the "inner side" of the contour rather than being nearly symmetrically distributed around it (as is the case for contour 5, Figures 9c and 9e). We stress that the area outside contour 3 corresponds to the locations for which most tomographic models suggest faster-than-average shear wave velocities. Hence, contours 1 and 2 are clearly "outboard" the margins of LLSVPs and cannot serve as their definitions; the same may perhaps apply to contour 3 as well. In contrast, contours 5 and 4 outlining the areas where all or most models agree on classification of locations within the lower mantle as seismically slow are more reliable approximations of the LLSVP boundaries, with contour 5 providing arguably the most robust definition.

The results of tests using the reconstructions in the hybrid reference frame of *Seton et al*. [2012] (S2012, Table 2) and in the global paleomagnetic reference frame [*Torsvik et al*., 2006, 2008b] are summarized in supporting information Figures S5 and S6. Similarly to our reference case (reconstructions in the DST2015 frame), models C and D consistently produced good fits to the observed distribution within the ~5–10° range of the σ parameter across the alternative definitions of the LLSVP margins, whereas the dominant majority of comparisons for model B showed misfits that were not acceptable at the 5% significance level. For model A and the reconstructions in the S2012 frame (supporting information Figure S5), we found acceptable fits for nearly all definitions of the slower-than-average velocity areas based on individual global tomography and D″ models (with the exception of tx2007), commonly for the values of σ between 4° and 7°. However, the quality of fit was notably worse compared to models C and D, as can be seen from generally low probabilities of the test statistics that barely cleared the 5% threshold for the majority of acceptable

fits. No satisfactory fits of model A were observed for the definitions based on the mean tomography models and voting contours 2–5 of *Lekic et al*. [2012]. In contrast, for the reconstructions in the global paleomagnetic reference frame (supporting information Figure S6), model A produced acceptable fits for all definitions of slower-than-average velocity regions except for voting contours 2–5 of *Lekic et al*. [2012], and the probabilities were comparable with those observed for models C and D, which is in agreement with the findings of *Austermann et al*. [2014]. However, we stress that we do not consider these results to be reliable, because the reconstructions of *Torsvik et al*. [2006, 2008b] were not corrected for TPW and used overly simplified rotation model for the Pacific plate (section 4).

In summary, we find that probability models C and D, which assume that plumes originate in the vicinity of the margins of LLSVPs and similar smaller-scale structures (e.g., the Perm Anomaly), fit the observed distribution of reconstructed LIPs well, and the goodness of fit is largely insensitive to the choices of seismic tomography model used to define the LLSVP boundaries and absolute reference frame. In contrast, the EDF-based tests of model B suggest that the hypothesis for the formation of LIP-feeding plumes restricted to the area of the LLSVPs, within the regions of postulated superplumes, is not consistent with the reconstructed positions of LIPs. Because of the sheer dominance of rejections, especially for the realistically low levels of Gaussian noise (e.g., $\sigma \leq 7°$), and the fact that the comparisons using the most robust definitions of LLSVPs (those based on the mean tomography models and cluster analysis of *Lekic et al*. [2012], voting contours 4 and 5) indicated a clearly significant misfit, we consider the evidence against this hypothesis to be sufficiently strong to warrant its formal rejection at the 5% significance level. Finally, the alternative in which plumes were assumed to be randomly generated over the entire area of slower-than-average shear wave velocities in the lowermost mantle (model A), received "mixed reviews," judging by the contrasting results that were strongly dependent on the choice of tomography model.

6.3. Correlations With the African and Pacific LLSVPs

In this section, we will present the tests conducted to address the criticism of *Davies et al*. [2015], who argued that the spatial correlation between the LIPs and the LLSVP margins is not global, being strong for the African LLSVP but weak for its Pacific counterpart, and suggested that the correlation observed on the global scale can be merely an artifact of the prevalence of the African data in the sample of reconstructed LIPs. Similarly to *Davies et al*. [2015], we analyzed the distributions of the angular distance relative to the LLSVP margins for the LIPs that can be correlated with the main LLSVP bodies, excluding the Columbia River Basalts and the Siberian Traps (19 LIPs associated with the African LLSVP, and 5 LIPs of the Pacific affinity, Table 2).

When selecting two regional subsets from the global sample of LIPs, the first question we would like to address is whether there is evidence in the data themselves that the samples of angular distance for the African and Pacific LIPs do not come from a common distribution. If it were indeed the case, this would invalidate the results of the global analysis presented in the previous section, implying distinct patterns of spatial correlation in the African and Pacific domains. To answer this question, we performed two-sample EDF tests for a common distribution (section 5, supporting information S1) for all alternative definitions of LLSVP margins (section 2). We stress that in these tests the hypothesized common distribution is unspecified; the only assumption we make is that the two samples could be drawn from the same population regardless of the actual form of its CDF.

Figure 10 shows the results of two-sample EDF tests for the LIP positions reconstructed in the DST2015 hybrid reference frame. For the majority of tests (19 out of 21 comparisons), the differences between the observed distributions for the African and Pacific LIPs were not significant at the 5% probability level. Only for the LLSVP definitions based on the hmsl-s tomography model [*Houser et al*., 2008] and D″ model of *Kuo et al*. [2000], the discrepancies were significant because in both cases four out of five Pacific LIPs reconstructed at nearly identical angular distances from the reference contours defining the LLSVP margins (e.g., Figure 10b). In 18 comparisons, the distributions were not statistically distinguishable at a 10% significance level or higher. Two-sample tests using the reconstructions in the hybrid reference frame of *Seton et al*. [2012] and in the global paleomagnetic frame [*Torsvik et al*., 2006, 2008a] produced similar results. Thus, although the uncertainties in the locations of LLSVP margins and the uncertainties of the estimated locations of LIP eruption centers were not accounted for in these tests, there appears to be no evidence against

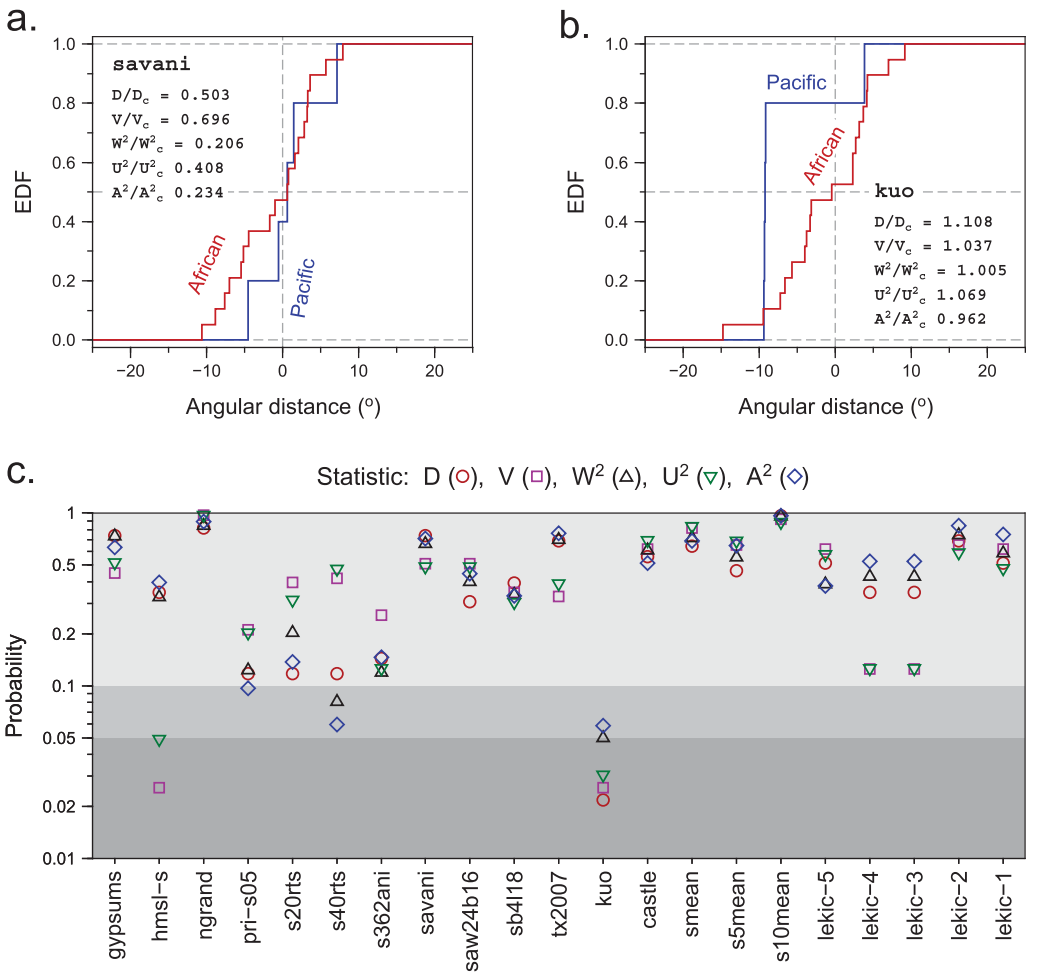

Figure 10. Two-sample EDF tests for a common distribution. (a, b) Examples of comparisons between the EDFs defined separately for the LIPs of the Pacific domain (blue lines) and the African domain (red lines), with the LLSVP definitions based on (a) the savani tomography model [*Auer et al*., 2014] and (b) the *Kuo et al*. [2000] D″ model. Two-sample EDF statistics divided by their critical values for the 5% significance level (supporting information S1) are quoted. The ratios exceeding 1 indicate a significant discrepancy. (c) Results of two-sample tests for all alternative definitions of LLSVP margins. LIPs were reconstructed in the DST2015 hybrid mantle reference frame.

the null hypothesis of a common distribution for the African and Pacific data, and hence the analysis assuming a global correlation can be applied sensibly.

However, when comparing the observed distribution to the theoretical models, examining the correlations for the African and Pacific domains separately can still be instructive. While models C and D are nearly identical when we use only the margins of the African LLSVP, the Pacific LLSVP, or a combination of the two for calculating the proposed CDFs, this is not the case for models A and B, which are more sensitive to the geometries of LLSVPs. Hence, it is important to test how the theoretical models perform when they are defined independently for the two regions and compared with the observed distributions of the African and Pacific LIPs. To address this, we calculated theoretical CDFs using the reference contours defining the boundaries and slower-than-average velocity regions corresponding to the African LLSVP alone, and removing all other contributions. Similar calculations were made for the Pacific LLSVP. These models were compared with the EDFs for the African and Pacific LIP samples, using the five EDF tests for goodness of fit described in section 5.

Figure 11 shows the results of regional goodness of fit tests for the reconstructions in the DST2015 hybrid reference frame. For the African domain, the results were similar to those of the global comparisons (section 6.2, Figure 8). This is not surprising, considering that the African LIPs dominate the combined data set. The broader regions of acceptable fits were also expected, because of a smaller number of observations constraining the EDFs and the diminishing power of EDF tests for small samples. As in the "global case," models

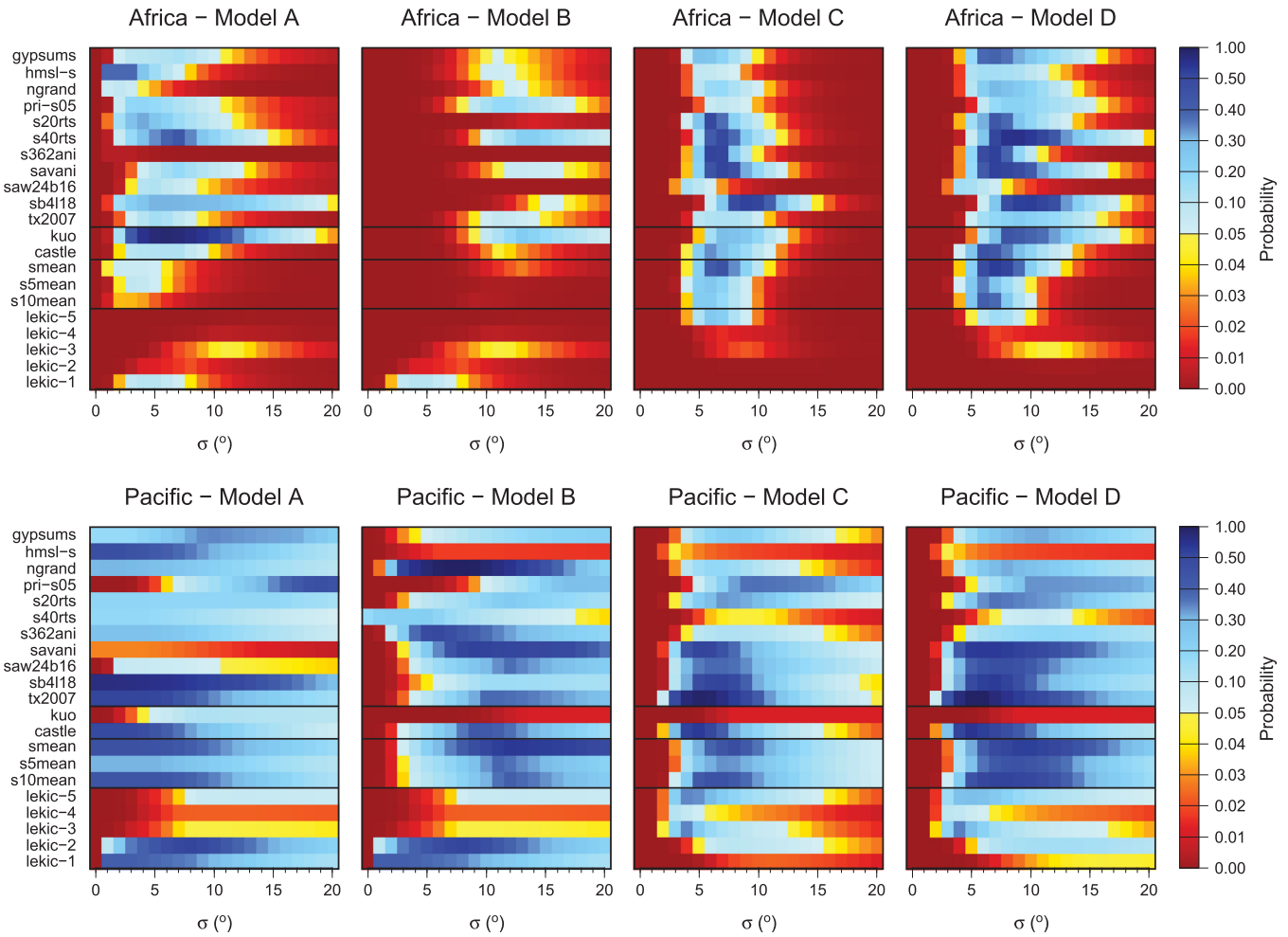

Figure 11. Summary of goodness of fit tests for regional comparisons in the (top row) African and (bottom row) Pacific domains for the reconstructions of LIPs in the DST2015 reference frame. The minimum values of probability for observing a test statistics as large or larger than the actually observed value (in one of the five EDF tests) are shown.

C and D exemplifying the plumes-from-the-margin scenario produced good fits for nearly all definitions of the LLSVP boundary. The only exceptions were the models in which the LLSVP margins were defined by *Lekic et al.* [2012] voting contours 1 through 4. Hence, we conclude that models C and D cannot be ruled out at the 5% significance level.

Model B (plumes from the entire area of the LLSVP) failed to produce acceptable fits for 11 out of 21 alternative definitions of the LLSVP margin (Figure 11), and all acceptable fits required imposing unrealistically high levels of Gaussian noise (typically, $\sigma \geq 10°$). The probabilities of test statistics for the fits acceptable at the 5% level were consistently lower than those observed for models C and D. If we chose to reject the models at a 10% significance level, the fits would be acceptable for only three definitions of the LLSVP margin (pri-s05, s40rts, and kuo), whereas all incarnations of model D, except those based on *Lekic et al.* [2012] contours 1–4, would still be acceptable. For the most reliable definitions of the LLSVP margin, those based on the mean tomographic models and voting contours 5 and 4 of *Lekic et al.* [2012], the discrepancies were clearly significant regardless of the level of imposed Gaussian noise. These observations suggest that model B could be rejected altogether for the African domain.

For model A (plumes from the area of slower-than-average velocity), unacceptable fits were observed in the tests using the s362ani and s10mean definitions, and *Lekic et al.* [2012] contours 2–5, whereas the remaining tests did not indicate a significant discrepancy. The probabilities of the test statistics were generally lower compared to models C and D, but this relationship was not as consistent as in comparisons with model B (e.g., with the hmsl-s and kuo definitions, model A produced higher probabilities than models C

and D). Thus, although in general models C and D fit the observations better, the evidence against model A is not sufficiently strong to reject it at the 5% significance level.

In contrast to the comparisons in the African domain, the results for the Pacific LLSVP were completely inconclusive: neither of the four probability models could be rejected with any reasonable confidence (Figure 11). This is, of course, a natural consequence of having only five observations in the Pacific sample. When the sample size is so small, the differences between a model CDF and the EDF must be truly drastic for the EDF tests to signal a discrepancy. The important observation though is that the models generating plumes from the margin of the Pacific LLSVP (C and D) fit the observations well for the majority of LLSVP definitions, and the estimated probabilities for the acceptable models were similar to those observed in the African comparisons. Thus, contrary to the conclusion of *Davies et al*. [2015], we do not see a statistically sound reason to prefer the models assuming that the Pacific LIPs are associated with the entire area of the Pacific LLSVP to the models assuming that these LIPs were sourced by plumes from the LLSVP margin.

The results of regional tests using the positions of LIPs restored in the *Seton et al*. [2012] hybrid frame and in the global paleomagnetic frame of *Torsvik et al*. [2006, 2008b] are summarized in supporting information Figures S7 and S8. Similarly to the tests relying on the DST2015 reconstructions, models C and D produced acceptable fits for nearly all LLSVP definitions in both the African and Pacific comparisons. Model A could not be formally rejected at the 5% significance level for the African and Pacific domains because the majority of tests did not indicate a significant discrepancy, whereas model B resulted in clearly inferior fits in the African domain, particularly for the reconstructions of *Seton et al*. [2012] (supporting information Figure S7).

7. Discussion

7.1. Competing Models

From a statistical point of view, a test for the goodness of fit is valuable only when it rejects hypotheses inconsistent with observations, thus narrowing the range of possible alternatives. In a situation as the one encountered here, when neither of the two alternative probability models can be rejected (i.e., model A versus models C and D for some of the LLSVP definitions), the researcher is left with essentially the same knowledge, or more precisely, the lack of thereof, as before the experiment. Yet, it may still be instructive to take a closer look at the competing models to see which model is in better agreement with observations, and whether there is at least a qualitative indication for giving preference to one of them.

As an illustration, Figure 12 shows representative examples of best-fitting models (corresponding to the lowest values of EDF statistics, and the highest probabilities of observing these or larger values, respectively) for the comparisons in the African domain using the reconstructions in the DST2015 reference frame. In these examples, none of the four alternatives (models A-D) can be formally rejected at the 5% significance level, with the exception of model B using the definition of the African LLSVP based on the smean tomography model (Figure 12c). Note also that the best-fitting cases for model B correspond to a high level of Gaussian noise ($\sigma = 13^\circ$ in all examples). Figures 12a–12c (comparisons using the definitions of the African LLSVP and the corresponding area of slower-than-average seismic velocity based on the s40rts, savani and smean tomography models) show typical results. It is evident that the proposed CDFs for model A have distinctly longer upper tails compared to the EDFs, a disagreement that is clearly visible at high positive values of angular distance (i.e., $> \sim 10^\circ$). This discrepancy reflects the fact that the reconstructed LIPs are located at the periphery of seismically slow areas, close to the margins of LLSVPs, rather than being randomly distributed over those areas. In contrast, models C and D that assume that the LIPs correlate with the margins of LLSVPs accurately reproduce the shape of the empirical distribution function, with no significant discrepancies either in the tails of distribution or around the mean. These models produce distinctly better fits, as can be judged from the higher probabilities of the test statistics to be equal or exceed the observed values.

The example using the *Kuo et al*. [2000] D″ model (Figure 12d) is a rare case when model A outperforms models C and D in terms of the higher probability values for the test statistics. Ironically, the CDF for model A in this particular case fits the EDF so closely that the difference cannot be explained by random variation. Note that the probability values in the Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling tests (which are 96.9%, 95.7%, and 96.9%, respectively) exceed 95%. Because these are the probabilities for the upper tails of their null distributions, the probabilities of observing statistics as small or smaller than those

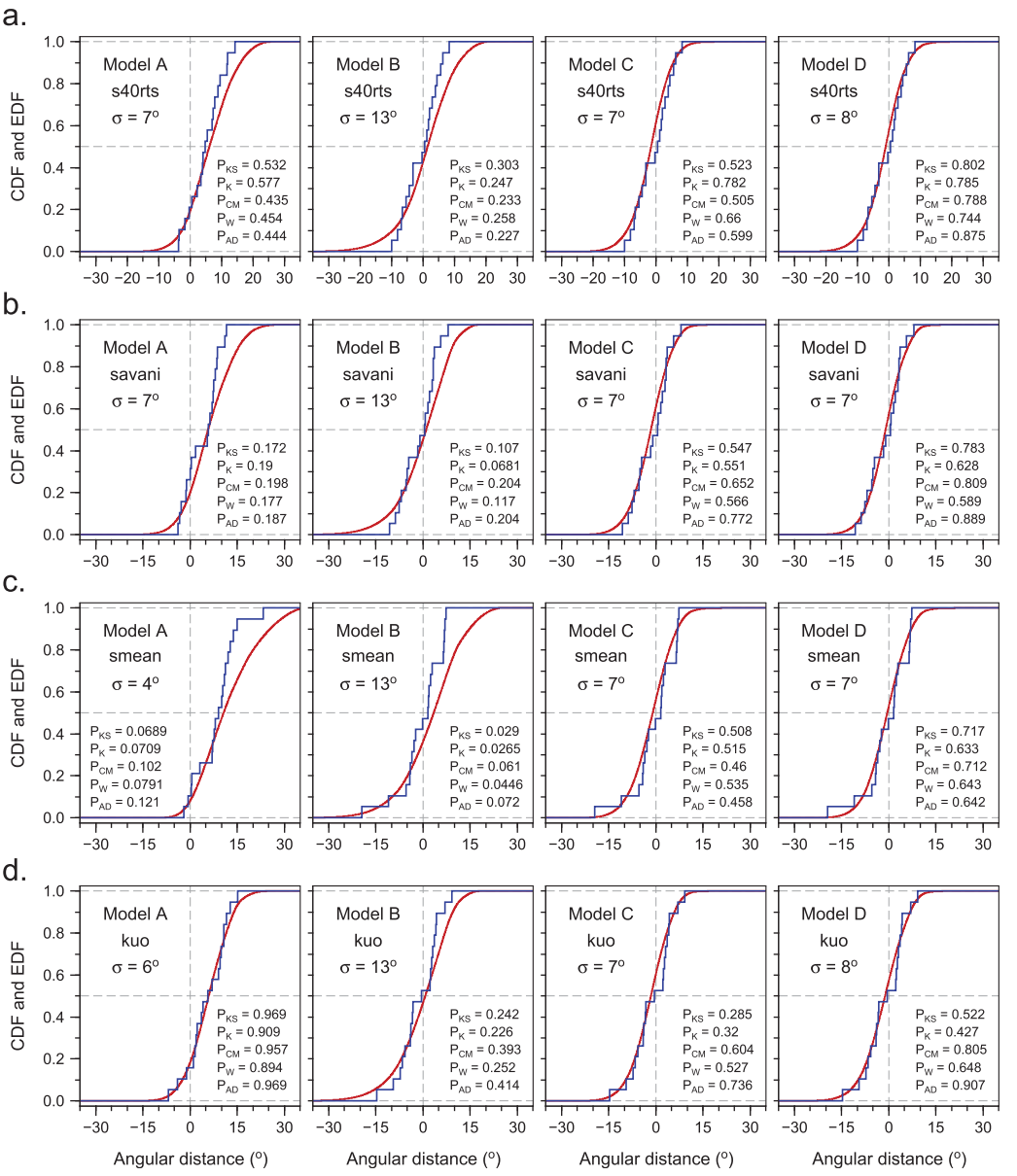

Figure 12. Representative examples of EDF tests for the African comparisons (DST2015 frame), for which none of alternative probability models can be formally rejected at the 5% significance level (except model B in Figure 12c). The values of σ for these examples correspond to overall best-fitting models (lowest values for the probabilities of test statistics, Figure 11). Red curves show the model CDFs; blue stepwise lines show the EDFs. The probabilities of observing the Kolmogorov-Smirnov (P_{KS}), Kuiper (P_K), Cramer-von Mises (P_{CM}), Watson (P_W), and Anderson-Darling (P_{AD}) statistics as large or larger than the actually observed values are quoted.

that were actually observed are less than 5%. In other words, the variation relative to the proposed CDF is subnormal, the three statistics fall into the lower 5% tails of their respective null distributions, and it would be as unlikely to obtain such small values as to observe large values somewhere in the upper 5% tail. We are perhaps not accustomed to tests against an unusual alternative of fit being "too good," but such situations do arise in statistical practice [e.g., *Pearson*, 1963], and in this particular example, the null hypothesis should be rejected at the 5% level. (Needless to say, if we impose higher or lower levels of Gaussian noise to make the fit slightly worse, the probabilities would become less than 95% (Figure 11), so that the "over-fitting" is generally not a problem to be overly concerned about.) At the same time, models C and D fit the observations almost as good as model A (Figure 12d), and no clear preference can be given to any of the three models.

The situation when models C and D fit the observed distribution better than model A is typical for the global comparisons based on the reconstructions of LIPs in the hybrid mantle frames (DST2015, S2012). In

tests using the DST2015 frame (Figure 8), model A produced higher probabilities compared to models C and D in only four cases (hmsl-s, pri-s05, kuo, and lekic-1), whereas in the remaining 17 comparisons, the probabilities were either lower than those for models C and D (7 cases when the acceptable fits were observed for the s20rts, s40rts, savani, sb4l18, castle, s5mean, and s10mean definitions), or the model could be formally rejected at the 5% significance level. For the reconstructions in S2012 frame (supporting information Figure S5), the probabilities of test statistics for model A were higher than those for models C or D for only two definitions of seismically slow areas (sb14l18 and lekic-1). Better performance of models C and D compared to model A is also evident in regional tests for the African region (Figures 11 and supporting information Figure S7), but for the Pacific domain, the small number of LIPs used to constrain the EDFs makes the qualitative comparisons discussed above meaningless. We note, however, that the probabilities for models A, C, and D in the Pacific tests were comparable, and the margin models (C and D) did not result in a markedly worse (or better) fit than that observed for model A.

When testing model B, the best fit was generally achieved at such values of the parameter σ where the mean of the proposed distribution (CDF) was close to the sample mean (Figure 12). This required imposing high levels of Gaussian noise (typically, $\sigma > 10^{\circ}$, Figures 8 and 11) that increased the spread (variance) of the model CDFs compared to the sample variance. Consequently, the largest differences between the model CDF and the EDF were observed in the tails of the distribution. For the majority of LLSVP definitions, for the African domain or globally, this discrepancy was large enough to reject the null hypothesis at the 5% significance level (e.g., Figure 12c). In the remaining cases, the probabilities of the test statistics in the tests of model B were consistently lower than those for models C and D (e.g., Figures 8, 11, 12a–12b and 12d). Because the levels of Gaussian noise greater than 10° are likely to overestimate the data uncertainties (section 5.1), and because no acceptable fits were found for the models using the most reliable definitions of LLSVP margins (section 6), we consider the evidence against model B to be sufficiently strong to reject it as a whole.

Overall, the models assuming that plumes responsible for the eruptions of LIPs rise from the margins of LLSVPs (C and D), are in notably better agreement with the distribution of LIPs reconstructed in the mantle reference frame than the alternatives postulating that plumes are randomly generated over the entire areas of LLSVPs or the areas of slower-than-average shear wave velocity in the lowermost mantle (models A and B). Good fits were consistently observed for models C and D at the intermediate values of the parameter σ (typically between 6° and 10°) across the alternative definitions of LLSVPs, in both the regional and global comparisons. We conclude that even though models A or B cannot be formally rejected at the 5% significance level in some comparisons, it would be reasonable to prefer the hypothesis proposing that LIPs are sourced by plumes from the LLSVP margins.

7.2. Comparison With the Results of Earlier Studies

The results of our global correlation analysis based on updated absolute plate reconstructions (section 6.2) partially support the findings of *Austermann et al*. [2014], who compared the distribution of reconstructed LIPs with three global tomography models (ngrand, s20rts, and sb4l18) and two D″ models [*Kuo et al*., 2000; *Castle et al*., 2000], and concluded that the LIPs correlated with both the margins of LLSVPs and the areas of slower-than-average shear wave velocity, but these two correlations were not statistically distinguishable at the 5% significance level. Indeed, considering the results for the reconstructions of LIPs in the DST2015 hybrid mantle frame (Figure 8), we found the fits of model A acceptable at the 5% level for four of these models (with the exception of ngrand), and so were the fits of models C and D. Hence, based only on these comparisons, it is indeed not possible to give preference to either of the two correlations. In our analysis we used a wider array of seismic models, and for 11 out of 21 alternative definitions of the seismically slow areas, the discrepancies between the CDFs for model A and the observed distributions of angular distance were significant. We also note that all models that assumed a uniform distribution over the areas of slower-than-average seismic velocity but did not impose the Gaussian noise on the surface distribution of eruptive centers ($\sigma = 0^{\circ}$, as in *Austermann et al*. [2014]) were not consistent with the observed distribution of reconstructed LIPs. In fairness, a conservative approach to the interpretation of our results is to retain model A since it produced acceptable fits in nearly a half of comparisons. However, given a ∼50% "rejection frequency" of this model in our tests, we will also have to conclude that the statistical evidence against it is stronger than the evidence against models C and D.

The disparity of the results for model A, when nearly a half of global comparisons suggested that it should be rejected, while the remaining tests did not indicate a significant discrepancy at the 5% level, can be qualitatively understood by examining Figure 4b. It is evident that the inferred geometries of slower-than-average velocity regions are very sensitive to the choice of tomography model, which results in large discrepancies between the alternative definitions. In contrast, high horizontal gradients of the shear wave velocity field at the margins of LLSVPs (section 2) make their definitions less susceptible to small velocity variations (Figure 4a). Thus, the geometries of LLSVP margins are more accurately constrained by the available seismic data than the areal extent of the regions characterized by slower-than-average shear wave velocities. It comes as no surprise then that the tests of models B, C, and D consistently suggested a rejection of model B, and did not show evidence against models C and D, implying that these results are robust.

Davies et al. [2015] performed regional tests for the African and Pacific domains using the Kolmogorov-Smirnov statistic to evaluate the goodness of fit of areal and marginal distributions defined in a similar, but not identical way compared to the probability models used in our analysis. They tried three alternative definitions of LLSVP margins, and three values defining the width of the plume generation zone around the margin, totaling in nine probability models for each scenario. For one of the LLSVP definitions, which *Davies et al*. [2015] arbitrarily assigned to be their reference case without providing a clear justification why this contour should be preferred, the discrepancies between the observed (EDF) and the model (CDF) distributions were significant at the 5% level for all three models assuming that the Pacific LIPs were sourced by plumes from the margin of the Pacific LLSVP. In the remaining comparisons for the Pacific region, only one of six tests showed significance, whereas the remaining five tests did not indicate a significant discrepancy between the observed distribution and the proposed marginal models. In contrast, nearly all comparisons for the African region (eight out of nine tests) showed no significant discrepancy. These results led *Davies et al*. [2015] to conclude that the correlation between the reconstructed LIPs and LLSVP margins is not global, being strong for the African LLSVP, but weak for the Pacific. They also concluded that the observed distribution of LIPs is better explained by plumes emanating from the entire areas of the African and Pacific LLSVPs.

The results of our regional tests (section 6.3) indicate that this is not the case. But before diving into the discussion why our results differ from those of *Davies et al*. [2015], we have to address one misconception that plagued the latter study. Namely, we would like to make it clear that EDF analysis does not really tell us how strong the correlation is.

Testing the correlation between the LIPs and LLSVP margins has the following logic. First we establish that the correlation exists; this has been done in studies of *Burke et al*. [2008] and *Austermann et al*. [2014]. Then we think what possible process could explain the correlation and propose a probability model that describes it; this becomes our statistical (null) hypothesis. Next we perform EDF tests, which tell us what is the probability of the data being a random draw from the proposed distribution. If the probability is low, e.g., less than 5%, we conclude that the model does not provide an adequate statistical description of the data and reject the null hypothesis. A rejection can be viewed as evidence against the proposed mechanism, but can also be due to the limitations of the model itself, in the case when the probability model does not fully capture the complexity of the hypothesized process. However, the fact that the hypothesis has or has not been rejected, or the probability value we obtained, has no relevance for the question whether the correlation exists (for this was our starting point) or how strong it is (because that was not what we tested).

It may be tempting to say that a better fit (higher probability) is indicative of a stronger correlation, but this is not correct because the goodness of fit depends on the parameters used to set up the model while the strength of the correlation does not. For example, if someone was to argue that the correlation between the African LIPs and the margin of the African LLSVP is weak or not significant because our model C fails the EDF tests for the values of $\sigma \leq 5^\circ$ (Figure 11), while an opponent would say that it is strong because the probabilities for the range of σ from 6° to 8° are generally high, this would be a silly argument since the underlying correlation is exactly the same regardless of σ. The goodness of fit is not a measure of the correlation strength and we strongly caution against its use in this role as that can result in conflicting and misleading inferences.

As we discussed in section 6.3, our results indicate that both classes of models, those assuming plumes randomly forming over the entire area of slower-than-average shear wave velocities associated with LLSVPs

(model A), and those assuming plumes rising from the LLSVP margins (models C and D), resulted in acceptable fits to the observations in both the African and Pacific regions, and no formal preference can be given to any of these models based on the results of EDF tests. For the Pacific comparisons, none of the four models (A–D) can be rejected at the 5% significance level, and the probabilities obtained in the EDF tests for models A, C, and D are comparable with those observed in the African comparisons (Figure 11). The difference between our results and those of *Davies et al*. [2015] cannot be ascribed to a different reconstruction of LIPs used in our analysis, since the reconstructions in the global paleomagnetic frame used by *Davies et al*. [2015] produced similar results for the African and Pacific regional tests (supporting information Figure S8). The main reason why our findings contrast with those of *Davies et al*. [2015] appears to be the different reference contours used to define the margins of LLSVPs. Noting that the unacceptable fits of the plume-from-the-margin models were consistently observed only for one definition of the Pacific LLSVP margin (the aforementioned reference case), while the tests using two alternative definitions did not indicate a significant discrepancy, we consider the findings of *Davies et al*. [2015] inconclusive, and do not see a reason for rejecting the hypothesis that the Pacific LIPs were sourced by plumes from the margin of LLSVP, even if we were to rely exclusively on the results of *Davies et al*. [2015]. Furthermore, for nearly all definitions of LLSVP margins used in our study, the distributions of angular distance for the African and Pacific LIPs are not distinguishable at the 5% significance level (Figure 10), and there is no statistically sound reason for questioning a global correlation between the reconstructed positions of LIPs and the margins of LLSVPs. Because *Davies et al*. [2015] did not perform a test for the strength of correlation between the LIPs and LLSVP margins in the African and Pacific domains that was independent of the probability models, we consider their conclusion about a weak correlation for the Pacific LLSVP and a strong correlation for the African LLSVP unfounded.

7.3. Caveats

Here we will briefly discuss the limitations of our statistical analysis, which are equally applicable to the earlier studies. First, we would like to stress that simple probability models we have tested, and those considered in the studies of *Austermann et al*. [2014] and *Davies et al*. [2015], are ultimately ad hoc formulations that might not fully capture the true complexity of the hypothesized modes of plume generation at the margins of LLSVPs or over the entire areas associated with these structures. For instance, when testing the hypothesis assuming a purely thermal origin of plumes from the areas of slower-than-average velocity anomalies (models A and B), it would be more realistic to assign higher probabilities to the regions characterized by the largest negative values of dlnVs (the "hottest" parts at the centers of LLSVPs) rather than to assume a flat, uniform distribution. If we chose to do so, the proposed CDFs for models A and B would be shifted to higher positive values of angular distance, making the discrepancies between the model CDF and the EDF more pronounced than those in the case of a uniform distribution, and leading to more confident rejections of these models. For the models assuming that plumes are formed at the LLSVP margins, the proposed distributions either simply disregard the possible variations of probability density along these boundaries (our model C, and the margin models of *Austermann et al*. [2014] and *Davies et al*. [2015]) or account only for their geometries (model D), but not for the other potential sources that may lead to the clustering of LIPs at some "preferred" locations along the boundaries (Figure 2). For example, in the model of *Steinberger and Torsvik* [2012] plumes form preferentially at sharp corners along the LLSVP margins. This issue, however, cannot be addressed by the analysis of the angular distance alone.

The definitions of the areal extent of slower-than-average shear wave velocities in the lowermost mantle are prone to large uncertainties owing to the discrepancies between the individual tomography models. Moreover, the Pacific LLSVP is less well imaged than the African LLSVP by the available seismic data [*Bower et al*., 2013]. Until more robust definitions are obtained by the continuing efforts to resolve these seismic structures with better accuracy, a statistical analysis of the correlation between the LIPs and LLSVPs has to rely on identifications of common correlation patterns for a large number of individual tomographic models, as it was implemented in our study.

The reconstruction uncertainties further complicate the analysis. Specifically, the reconstructed locations of the Pacific LIPs are subject to larger uncertainties than those in the African domain, because here we have to rely on fixed hotspot models in order to extent the reconstructions beyond the Late Cretaceous (section 4). Numerous studies have shown that the kinematic models assuming hotspot fixity are less reliable than those incorporating the estimates of hotspot motion [e.g., *Doubrovine et al*., 2012, and references therein]. It

is truly remarkable that despite this limitation the reconstructed locations of the Pacific LIPs closely fit the margin of the Pacific LLSVP (Figure 2). The assumption of zero-longitude Africa motion for Permian to Early Cretaceous time (120–300 Ma) is another source of uncertainty. Yet, this is currently the only tenable way of constraining the absolute longitudes for this time interval without invoking the spatial correlation between the LIPs and LLSVP margins [*Steinberger and Torsvik*, 2008]. We also suggest that paleogeographic reconstructions that have not been corrected for TPW should no longer be used for testing the correlation between the LIPs and LLSVPs, since they may introduce unrecognized biases leading to erroneous inferences.

It should be pointed out that not all LIPs might be sourced by deep mantle plumes associated with the LLSVPs or similar smaller-scale structures of the lowermost mantle. As it was already noted in the original studies of *Burke and Torsvik* [2004], *Torsvik et al*. [2006, 2008a], and *Burke et al*. [2008], the Siberian Traps and the Columbia River Basalts (CRB) do not appear to be related to the main LLSVP bodies. Now we can confidently correlate the Siberian Traps with the margin of the Perm Anomaly identified by the cluster analysis of *Lekic et al*. [2012]. We suggest that at the time when the Siberian Traps erupted (∼251 Ma), the Perm Anomaly may have been a contiguous promontory of the African LLSVP that later become separated from the main LLSVP body by the Tethyan subduction [e.g., *Van der Voo et al*., 1999]. However, the origin of the Columbia River Basalts remains uncertain.

The CRB is the smallest among 26 LIPs considered in this study (both volumetrically and by its areal extent), and is unusual in that it is located over the region that has been dominated by subduction of oceanic lithosphere (the Farallon and Kula plates) underneath the western North American margin since at least Jurassic time [e.g., *Doubrovine and Tarduno*, 2008]. While seismic images of the upper mantle below the CRB and Yellowstone hotspot track seem to be best explained by a model in which a deep plume has "burned through" the subducted Farallon slab [e.g., *Obrebski et al*., 2010], the tail of the plume is not resolved below ∼500 km depth [*Yuan and Dueker*, 2005] or 660 km depth [*Smith et al*., 2009], and an alternative explanation relating the CRB volcanism to the fragmentation of the subducted plate has been put forward [*James et al*., 2011]. No plumes in the CRB region or in nearby offshore areas were imaged to be situated above low-velocity regions at the CMB in a recent study by *French and Romanowicz* [2015]. It is intriguing that some recent tomography models (e.g., s362ani and savani), as well as the D″ models of *Kuo et al*. [2000] and *Castle et al*. [2000], suggest a north-eastern promontory of the Pacific LLSVP, or a separate structure comparable in size with the Perm Anomaly, and the reconstructed CRB province lies near the margin of this structure (Figure 2). Whether this lower mantle anomaly is real or not is the subject of future geophysical studies; yet, the deep plume origin of the CRB cannot be ruled out.

8. Conclusions and Implications

We have tested whether the correlation between the eruption sites of large igneous provinces and the low shear wave velocity structures in the lowermost mantle (LLSVPs) can be explained by three alternative geodynamic scenarios: (i) deep mantle plumes emanating from the areas of slower-than-average seismic velocity, (ii) plumes forming within the area restricted by the LLSVP margins, and (iii) plumes rising from the margins themselves. We used a large number of seismic tomography models to avoid arbitrariness in defining the areal extent of these structures, and based our analysis on updated absolute plate reconstructions and nonparametric goodness of fit tests for the distribution of the angular distance relative to the LLSVP margins and to the reference contours defining the areas of slower-than-average seismic velocity.

Our statistical analysis shows that the probability models exemplifying the causal link between the LIPs and plumes from the LLSVP margins consistently produced acceptable fits to the observed distribution of angular distance across the alternative definitions of LLSVP boundaries. In contrast, we found ample evidence against the models assuming that the distribution of LIPs reflects deep plume sources randomly generated over the entire areas of LLSVPs or slower-than-average shear wave velocity anomalies at large. The evidence against the hypothesis assuming plume generation within the boundaries of LLSVPs was sufficiently strong to warrant its formal rejection at the 5% significance level. The tests of the alternative proposing that the LIP-sourcing plumes randomly form over the entire area of slower-than-average seismic velocities were less conclusive, with nearly a half of comparisons suggesting a rejection of the null hypothesis and the remaining tests not indicating a significant discrepancy. Given the current knowledge, both hypotheses linking the

LIPs to the plumes from the LLSVP margins and to the areas of slower-than-average velocity should be considered as viable options. However, the qualitative comparison of the goodness of fit indicates that the probability models seeking to explain the observed distribution by plumes originating from the margins of LLSVPs provide a more accurate description for the distribution of reconstructed LIPs.

It cannot be ruled out that the correlation between the LIP eruption sites and the LLSVP margins is global. Two-sample EDF tests demonstrate that the distribution of the angular distance for the Pacific LIPs relative to the boundary of the Pacific LLSVP is not distinguishable from that for their African counterparts. There is no reason to invoke different correlation patterns in the two regions, and the same process of plume generation at the margins of LLSVPs adequately explains the observed distributions both in the African and Pacific domains.

Our findings have direct implications for the debate on the long-term stability of LLSVPs originally proposed by *Burke and Torsvik* [2004]. The reconstruction spans a ~300 Ma interval, and the correlation between the reconstructed LIPs and the margins of LLSVPs as they are imaged today would not be possible if the LLSVPs were mobile or transient features, prone to changing their positions, sizes and shapes over a shorter time scale. It is true that this time interval is not sampled evenly by the LIP ages. As *Davies et al*. [2015] pointed out, only five LIPs in our sample are older than 150 Ma, and none of them are of Pacific origin (Table 2). Yet, there is no justifiable reason to ignore these observations.

An independent data set comprising more than a thousand kimberlites erupted in the cratonic parts of Africa, South America, Australia and India over the past 320 Ma shows a correlation with the margin of the African LLSVP, which is remarkably similar to the distribution of reconstructed African LIPs [*Torsvik et al*., 2010]. The observation that the kimberlite localities restored in a hybrid mantle reference frame [*Torsvik et al*., 2008c; *Steinberger and Torsvik*, 2008] concentrate along the margin of the African LLSVP led *Torsvik et al*. [2010] to suggest that these igneous rocks are likely to be related to the mantle plumes that have risen from the margin, implying that the African LLSVP was already in place at the time of Pangea assembly and has remained in a nearly present day configuration since that time. A recent study of *Torsvik and Burke* [2015] demonstrated that binning the kimberlite data into the 100–200 Ma and 200–300 Ma intervals (513 and 47 kimberlite localities, respectively), produces equally strong correlations between the kimberlites and the margin of the African LLSVP.

It is more difficult to argue for the stability of the Pacific LLSVP prior to late Jurassic time, because no LIPs older than 150 Ma have been preserved in the Pacific Ocean, and kimberlites do not form in oceanic domains. In the absence of direct constraints from the reconstructions of LIPs and kimberlites, one has to rely on numerical models of the temporal evolution of LLSVPs, but the results of the studies that have been attempted so far are controversial.

The geodynamic model of *Zhang et al*. [2010] suggested that the Pacific LLSVP has existed since as early as Ordovician time (~460 Ma), although it was twice larger and its shape was different from what is observed today, whereas the African LLSVP came into existence only in the late Triassic or early Jurassic (after ~200 Ma) when it became separated from the Pacific "megapile" by subducted slab material. Even in the early Cretaceous (120 Ma), the modeled African LLSVP does not resemble the present day structure and is offset by ~60° to the east relative to the position imaged by seismic tomography. In contrast, four recent numerical studies [*Steinberger and Torsvik*, 2012; *Bower et al*., 2013; *Bull et al*., 2014; *Hassan et al*., 2015], have demonstrated that two distinct, nearly antipodal LLSVPs can remain in their original positions over hundreds of millions of years (230–410 Ma depending on the model), although in all these studies, the shapes and the areal extent of LLSVP structures evolve considerably. None of the numerical models cited here arrives at the present day lower mantle structure that closely resembles that imaged by seismic tomography, especially for the Pacific LLSVP. Noting this limited success in reproducing the present day structure, a strong dependence on the choice of model parameters that in most cases are poorly constrained, and sensitivity to the plate motion histories, which are subject to large uncertainties for Mesozoic and especially Paleozoic time, we suggest that the LLSVPs are less mobile than what has been obtained in the numerical models so far. This could be the case if these structures are either stiffer than what has been assumed, or act as the "anchoring" mechanism that may have controlled where downward flow and subduction have occurred for the past 300 Ma [cf. *Dziewonski et al*., 2010]. An indication of this is that the net characteristics of plate tectonics reveal that active mantle upwellings have been stable for the past 250 Ma [*Conrad et al*., 2013],

whereas regions of strongest downwellings have been more mobile. The reconstructions of LIPs and kimberlites are another observational data set we have to test this hypothesis.

The correlation between the LIPs and the LLSVP margins has equally important, albeit indirect implications for the debate on whether the LLSVPs are predominantly thermal or thermochemical structures. Numerical models of mantle convection assuming that the mantle is isochemical (i.e., a dominantly thermal nature of LLSVPs), and utilizing subduction histories inferred from absolute plate reconstructions, commonly reproduce two clusters of thermal plumes, or sometimes linear, interconnected ridges of hot, buoyant material, with plumes off-shooting from their crests, which are distributed over the regions of the African and Pacific LLSVPs [e.g., *Bull et al*., 2009; *Davies et al*., 2012]. *Bull et al*. [2009] showed that separate plumes might not be resolved in seismic tomography models, so that the plume clusters would appear in tomographic images as continuous structures. Hence, it may not be possible to decide whether the LLSVPs are amalgamations of narrow thermal plumes and/or ridges or coherent large-scale bodies based on seismic inversions alone.

However blurred seismic images might be, if we accept the plume cluster explanation of LLSVP structures, we would expect the plumes to form over the entire areas of LLSVPs, as in the model of *Bull et al*. [2009]. This is not supported by the results of our statistical analysis. *Davies et al*. [2012] noted that in their thermal convection model, the majority of plumes rose from the margins of the large-scale thermal structures beneath Africa and the Pacific Ocean. However, since *Davies et al*. [2012] did not plot the present distribution of modeled plumes and did not document the temporal evolution of plume clusters, it is not possible to test whether their model is in qualitative agreement with the observed distribution of reconstructed LIPs. In contrast, the studies of *Steinberger and Torsvik* [2012] and *Hassan et al*. [2015], who modeled the generation of plumes related to chemically distinct LLSVPs, demonstrated that (i) plumes form at the margins of LLSVPs and (ii) the locations where plumes reach the lithosphere after their transit through the mantle (corresponding to the LIP eruption sites) align with the outlines of LLSVPs vertically projected onto the Earth surface. These results are fully consistent with observations.

The numerical models of plumes generated at the margins of LLSVPs are still in their infancy and are likely to be improved in the future, building on the techniques that have been implemented already. Whether or not the models of isochemical convection will be able to reproduce the long-term stability of LLSVPs and plumes predominantly originating from their margins is an open question. Given the current state, we suggest that the observed spatial distribution of reconstructed LIPs and kimberlites is most satisfactorily explained by plumes rising from the margins of stable LLSVPs that are chemically distinct from the bulk of the mantle.

Acknowledgments

The LIP reconstruction data used in this study are available in Table 2. The seismic tomography data are properly cited in Table 1 and referred to in the reference list. Additional mean shear-wave tomography models (s5mean and s10mean) may be obtained from PVD upon request (e-mail: paveld@fys.uio.no). We would like to thank D. Rosa d'Alvadorez for helpful comments, Thorsten Becker for making available the savani tomography model, and Christine Houser and an anonymous referee for reviews that have considerably improved the paper. The Research Council of Norway, through its Centers for Excellence funding scheme, and the European Research Council, under the European Union's Seventh Framework Programme (FP7/2007-2013), ERC Advance grant agreement 267631 (Beyond Plate Tectonics), are acknowledged for financial support.

References

Anderson, T. W. (1962), On the distribution of the two-sample Cramer-von Mises criterion, *Ann. Math. Stat.*, *33*(3), 1148–1159.

Auer, L., L. Boschi, T. W. Becker, T. Nissen-Meyer, and D. Giardini (2014), Savani: A variable resolution whole-mantle model of anisotropic shear velocity variations based on multiple data sets, *J. Geophys. Res. Solid Earth*, *119*, 3006–3034, doi:10.1002/2013JB010773.

Austermann, J., B. T. Kaye, J. X. Mitrovica, and P. Huybers (2014), A statistical analysis of the correlation between large igneous provinces and lower mantle seismic structure, *Geophys. J. Int.*, *197*(1), 1–9, doi:10.1093/gji/ggt500.

Becker, T., and L. Boschi (2002), A comparison of tomographic and geodynamic mantle models, *Geochem. Geophys. Geosyst.*, *3*(1), 1003, doi:10.1029/2001GC000168.

Boschi, L., T. W. Becker, and B. Steinberger (2007), Mantle plumes: Dynamic models and seismic images, *Geochem. Geophys. Geosyst.*, *8*, Q10006, doi:10.1029/2007GC001733.

Bower, D. J., M. Gurnis, and M. Seton (2013), Lower mantle structure from paleogeographically constrained dynamic Earth models, *Geochem. Geophys. Geosyst.*, *14*, 44–63, doi:10.1029/2012GC004267.

Bull, A. L., A. K. McNamara, and J. Ritsema (2009), Synthetic tomography of plume clusters and thermochemical piles, *Earth Planet. Sci. Lett.*, *278*, 152–162, doi:10.1016/j.epsl.2008.11.018.

Bull, A. L., M. Domeier, and T. H. Torsvik (2014), The effect of plate motion history on the longevity of deep mantle heterogeneities, *Earth Planet. Sci. Lett.*, *401*, 172–182, doi:10.1016/j.epsl.2014.06.008.

Burke, K., and T. H. Torsvik (2004), Derivation of large igneous provinces of the past 200 million years from long-term heterogeneities in the deep mantle, *Earth Planet. Sci. Lett.*, *227*, 531–538, doi:10.1016/j.epsl.2004.09.015.

Burke, K., B. Steinberger, T. H. Torsvik, and M. A. Smethurst (2008), Plume generation zones at the margins of large low shear velocity provinces on the core-mantle boundary, *Earth Planet. Sci. Lett.*, *265*, 49–60, doi:10.1016/j.epsl.2007.09.042.

Castle, J. C., K. C. Creager, J. P. Winchester, and R. D. van der Hilst (2000), Shear wave speeds at the base of the mantle, *J. Geophys. Res. 105*, 21,543–21,558, doi:10.1029/2000JB900193.

Chandler, M. T., P. Wessel, B. Taylor, M. Seton, S.-S. Kim, and K. Hyeong (2012), Reconstructing Ontong Java Nui: Implications for Pacific absolute plate motion, hot spot drift and true polar wander, *Earth Planet. Sci. Lett.*, *331–332*, 140–151, doi:10.1016/j.epsl.2012.03.017.

Conrad, C. P., B. Steinberger, and T. H. Torsvik (2013), Stability of active mantle upweling revealed by net characteristics of plate tectonics, *Nature*, *498*, 479–482, doi:10.1038/nature12203.

Davies, D. R., S. Goes, J. H. Davies, B. S. A. Schuberth, H.-P. Bunge, and J. Ritsema (2012), Reconciling dynamic and seismic models of Earth's lower mantle: The dominant role of thermal heterogeneity, *Earth Planet. Sci. Lett.*, *353–354*, 253–269, doi:10.1016/j.epsl.2012.08.016.
Davies, D. R., S. Goes, and M. Sambridge (2015), On the relationship between volcanic hotspot locations, the reconstructed eruption sites of large igneous provinces and deep mantle seismic structures, *Earth Planet. Sci. Lett.*, *411*, 121–130, doi:10.1016/j.epsl.2014.11.052.
Doubrovine, P. V., and J. A. Tarduno (2008), A revised kinematic model for the relative motion between Pacific oceanic plates and North America since the Late Cretaceous, *J. Geophys. Res.*, *113*, B12101, doi:10.1029/2008JB005585.
Doubrovine, P. V., B. Steinberger, and T. H. Torsvik (2012), Absolute plate motions in a reference frame defined by moving hot spots in the Pacific, Atlantic, and Indian oceans, *J. Geophys. Res.*, *117*, B09101, doi:10.1029/2011JB009072.
Dziewonski, A. M., and D. L. Anderson (1981), Preliminary reference earth model, *Phys. Earth Planet. Inter.*, *25*, 297–356.
Dziewonski, A. M., V. Lekic, and B. Romanowicz (2010), Mantle anchor structure: An argument for bottom up tectonics, *Earth Planet. Sci. Lett.*, *299*, 69–79, doi:10.1016/j.epsl.2010.08.013.
Eldholm, O., and M. F. Coffin (2000), Large igneous provinces and plate tectonics, in *The History and Dynamics of Global Plate Motions*, edited by M. A. Richards, R. G. Gordon, and R. D. van der Hilst, pp. 309–326, AGU, Washington, D. C.
Ernst, R. E. (2014), *Large Igneous Provinces*, 653 pp., Cambridge Univ. Press, Cambridge, U. K.
Fasano, G., and A. Franceschini (1987), A multidimensional version of the Kolmogorov-Smirnov test, *Mon. Not. R. Astron. Soc.*, *225*, 155–170.
Fisher, N. I., T. Lewis, and B. J. J. Embleton (1987), *Statistical Analysis of Spherical Data*, 329 pp., Cambridge Univ. Press, N. Y.
French, S. W., and B. Romanowicz (2015), Broad plumes rooted at the base of the Earth's mantle beneath major hotspots, *Nature*, *525*, 95–99.
Ganerød, M., M. A. Smethurst, T. H. Torsvik, T. Prestvik, S. Rousse, C. McKenna, D. J. J. van Hinsbergen, and B. W. H. Hendriks (2010), The North Atlantic Igneous Province reconstructed and its relation to the Plume Generation Zone: The Antrim Lava Group revisited, *Geophys. J. Int.*, *182*, 183–202, doi:10.1111/j.1365-246X.2010.04620.x.
Garnero, E. J., T. Lay, and A. McNamara (2007), Implications of lower mantle structural heterogeneity for existence and nature of whole mantle plumes, in *Plates, Plumes and Planetary Processes, Geol. Soc. Am. Spec. Pap.*, vol. 430, edited by G. R. Foulger and D. M. Jurdy, pp. 79–101, Geol. Soc. Amer., Boulder, Colo., doi:10.1130/2007.2430(05).
Goldreich, P., and A. Toomre (1969), Some remarks on polar wandering, *J. Geophys. Res.*, *74*, 2555–2567, doi:10.1029/JB074i010p02555.
Grand, S. P. (2002), Mantle shear–wave tomography and the fate of subducted slabs, *Philos. Trans. R. Soc. London A*, *360*, 2475–2491, doi:10.1098/rsta.2002.1077.
Hassan, R., N. Flament, M. Gurnis, D. J. Bower, and R. D. Müller (2015), Provenance of plumes in global convection models, *Geochem. Geophys. Geosyst.*, *16*, 1465–1489, doi:10.1002/2015GC005751.
Houser, C., G. Masters, P. Shearer, and G. Laske (2008), Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, *Geophys. J. Int.*, *174*, 195–212, doi:10.1111/j.1365-246X.2008.03763.x.
James, D. E., M. J. Fouch, R. W. Carlson, and J. B. Roth (2011), Slab fragmentation, edge flow and the origin of the Yellowstone hotspot track, *Earth Planet. Sci. Lett.*, *311*, 124–135, doi:10.1016/j.epsl.2011.09.007.
Justel, A., D. Pena, and R. Zamar (1997), A multivariate Kolmogorov-Smirnov test of goodness of fit, *Stat. Prob. Lett.*, *35*, 251–259.
Kuiper, N. H. (1960), Tests concerning random points on a circle, *Ned. Acad. Wet. Proc. Ser. A*, *63*, 38–47.
Kuo, B.-Y., E. J. Garnero, and T. Lay (2000), Tomographic inversion of S-SKS times for shear velocity heterogeneity in D″: Degree 12 and hybrid models, *J. Geophys. Res.*, *105*, 28,139–28,157, doi:10.1029/2000JB900334.
Kustowski, B., G. Ekstrom, and A. M. Dziewonski (2008), Anisotropic shear-wave velocity structure of the Earth's mantle: A global model, *J. Geophys. Res.*, *113*, B06306, doi:10.1029/2007JB005169.
Lekic, V., S. Cottaar, A. Dziewonski, and B. Romanowicz (2012), Cluster analysis of global mantle tomography: A new class od structure and implications for chemical heterogeneity, *Earth Planet. Sci. Lett.*, *357–359*, 68–77, doi:10.1016/j.epsl.2012.09.014.
Lithgow-Bertelloni, C., and M. A. Richards (1998), The dynamics of Cenozoic and Mesozoic plate motions, *Rev. Geophys.*, *36*, 27–78, doi:10.1029/97RG02282.
Masters, G., G. Laske, H. Bolton, and A. M. Dziewonski (2000), The relative behavior of shear velocity, bulk sound speed, and compressional velocity in the mantle: Implications for chemical and thermal structure, in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale, Geophys. Monogr. Ser.*, vol. 117, edited by S Karato et al., pp. 63–87, AGU, Washington, D. C.
Megnin, C., and B. Romanowicz (2000), The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveform tomography, *Gephys. J. Int.*, *143*, 709–728, doi:10.1046/j.1365-246X.2000.00298.x.
Montelli, R., G. Nolet, F. Dahlen, and G. Masters (2006), A catalogue of deep mantle plumes: New results from finite-frequency tomography, *Geochem. Geophys. Geosyst.*, *7*, Q11007, doi:10.1029/2006GC001248.
Obrebski, M., R. M. Allen, M. Xue, and S. Hung (2010), Slab-plume interaction beneath the Pacific Northwest, *Geophys. Res. Lett.*, *37*, L14305, doi:10.1029/2010GL043489.
O'Neill, C., R. D. Müller, and B. Steinberger (2005), On the uncertainties in hot spot reconstructions and the significance of moving hot spot reference frames, *Geochem. Geophys. Geosyst.*, *6*, Q04003, doi:10.1029/2004GC000784.
Parsiegla, N., K. Gohl, and G. Uenzelmann-Neben (2008), The Agulhas Plateau: Structure and evolution of a Large Igneous Province, *Geophys. J. Int.*, *174*, 336–350, doi:10.1111/j.1365-246X.2008.03808.x.
Pearson, E. S. (1963), Comparisons of tests for randomness of points on the line, *Biometrica*, *50*(3-4), 315–325.
Pettitt, A. N. (1976), A two-sample Anderson-Darling rank statistic, *Biometrica*, *63*(1), 161–168.
Richards, M. A., R. A. Duncan, and V. E. Courtillot (1989), Flood basalts and hot-spot tracks: Plume heads and tails, *Science*, *246*, 103–107.
Ritsema, J., J. H. van Heijst, and J. H. Woodhouse (1999), Complex shear velocity structure beneath Africa and Iceland, *Science*, *286*, 1925–1928.
Ritsema, J., A. Deuss, J. H. van Heijst, and J. H. Woodhouse (2011), S40RTS: A degree-40 shear-velocity model for the mantle from the Rayleigh wave dispersion, teleseismic travel time and normal-mode splitting function measurements, *Geophys. J. Int.*, *184*, 1223–1236, doi:10.1111/j.1365-246X.2010.04884.x.
Seton, M., et al. (2012), Global continental and ocean basin reconstructions since 200 Ma, *Earth Sci. Rev.*, *113*, 212–270, doi:10.1016/j.earscirev.2012.03.002.
Simmons, N. A., A. M. Forte, and S. P. Grand (2007), Thermochemical structure and dynamics of the African superplume, *Gephys. Res. Lett.*, *34*, L02301, doi:10.1029/2006GL028009.
Simmons, N. A., A. Forte, L. Boschi, and S. Grand (2010), GyPSuM: A joint tomographic model of mantle density and seismic wave speeds, *J. Geophys. Res.*, *115*, B12310, doi:10.1029/2010JB007631.
Sleep, N. H. (1997), Lateral flow and ponding of starting plume material, *J. Geophys. Res.*, *102*, 10,001–10,012, doi:10.1029/97JB00551.
Smith, B., M. Jordan, B. Steinberger, C. M. Puskas, J. Farrell, G. P. Waite, S. Husen, W.-L. Chang, and R. O'Connell (2009), Geodynamics of the Yellowstone hotspot and mantle plume: Seismic and GPS imaging, kinematics and mantle flow, *J. Volcanol. Geotherm. Res.*, *188*, 26–56, doi:10.1016/j.jvolgeores.2009.08.020.

Sprent, P., and N. C. Smeeton (2007), *Applied Nonparametric Statistical Methods*, 4th ed., 530 pp., Chapman and Hall, Boca Raton, Fla.
Steinberger, B., and R. J. O'Connell (1998), Advection of plumes in mantle flow: Implications for hotspot motion, mantle viscosity and plume distribution, *Geophys. J. Int.*, *132*, 412–434.
Steinberger, B., and T. H. Torsvik (2008), Absolute plate motions and True Polar Wander in the absence of hotspot tracks, *Nature*, *452*, 620–623.
Steinberger, B., and T. H. Torsvik (2012), A geodynamic model of plumes from the margins of Large Low Shear Velocity Provinces, *Geochem. Geophys. Geosyst.*, *13*, Q01W09, doi:10.1029/2011GC003808.
Stephens, M. A. (1974), EDF statistics for goodness of fit and some comparisons, *J. Am. Stat. Assoc.*, *69*(347), 730–737.
Tan, E., W. Leng, S. Zhong, and M. Gurnis (2011), On the location of plumes and lateral movement of thermochemical structures with high bulk modulus in the 3-D compressible mantle, *Geochem. Geophys. Geosyst.*, *12*, Q07005, doi:10.1029/2011GC003665.
Torsvik, T. H., and K. Burke (2015), Large igneous province locations and their connections with the core-mantle boundary, in *Volcanism and Global Environmental Change*, edited by A. Schmidt, A., et al., pp. 30–46, Cambridge Univ. Press, U. K.
Torsvik, T. H., M. A. Smethurst, K. Burke, and B. Steinberger (2006), Large igneous provinces generated from the margins of the large low-velocity provinces in the deep mantle, *Geophys. J. Int.*, *167*, 1447–1460, doi:10.1111/j.1365-246X.2006.03158.x.
Torsvik, T. H., B. Steinberger, L. R. M. Cocks, and K. Burke (2008a), Longitude: Linking Earth's ancient surface to its deep interior, *Earth Planet. Sci. Lett.*, *276*, 273–282, doi:10.1016/j.epsl.2008.09.026.
Torsvik, T. H., M. A. Smethurst, K. Burke, and B. Steinberger (2008b), Long term stability in deep mantle structure: Evidence from the ~300 Ma Skagerrak-centered large igneous province (the SCLIP), *Earth Planet. Sci. Lett.*, *267*, 444–452, doi:10.1016/j.epsl.2007.12.004.
Torsvik, T. H., R. D. Müller, R. Van der Voo, B. Steinberger, and C. Gaina (2008c), Global plate motion frames: Toward a unified model, *Rev. Geophys.*, *46*, RG3004, doi:10.1029/2007RG000227.
Torsvik, T. H., K. Burke, B. Steinberger, S. J. Webb, and L. D. Ashwal (2010), Diamonds sampled by plumes from the core-mantle boundary, *Nature*, *466*, 352–355.
Torsvik, T. H., et al. (2012), Phanerozoic polar wander and palaeogeography and dynamics, *Earth Sci. Rev.*, *114*, 325–368, doi:10.1016/j.earscirev.2012.06.007.
Torsvik, T. H., R. Van der Voo, P. V. Doubrovine, K. Burke, B. Steinberger, L. D. Ashwal, R. G. Trønnes, S. J. Webb, and L. A. Bull (2014), Deep mantle structure as a reference frame for movements in and on the Earth, *Proc. Nat. Acad. Sci. U. S. A.*, *111*(24), 8735–8740, doi:10.1073/pnas.1318135111.
Van der Voo, R., W. Spakman, and H. Bijwaard (1999), Tethyan subducted slabs under India, *Earth Planet. Sci. Lett.*, *171*, 7–20.
Watson, G. S. (1962), Goodness-of-fit tests on a circle, II, *Biometrica*, *49*(1–2), 57–63.
Wessel, P., and L. Kroenke (2008), Pacific absolute plate motion since 145 Ma: An assessment of the fixed hot spot hypothesis, *J. Geophys. Res.*, *113*, B06101, doi:10.1029/2007JB005499.
Yuan, H., and K. Dueker (2005), Teleseismic P-wave tomogram of the Yellowstone plume, *Geophys. Res. Lett.*, *32*, L07304, doi:10.1029/2004GL022056.
Zhang, N., S. Zhong, W. Leng, and Z.-X. Li (2010), A model for the evolution of the Earth's mantle structure since the Early Paleozoic, *J. Geophys. Res.*, *115*, B06401, doi:10.1029/2009JB006896.