

Global Distribution of Tectonic Plates Revealed by the Cluster Analysis of Geodetic Data in the Angular Velocity Space

Tomohisa OKAZAKI^(1,2), Atsushi TAKAHASHI⁽¹⁾, Koh TAKEUCHI⁽³⁾,
Tomoharu IWATA^(2,3) and Yukitoshi FUKAHATA

(1) Graduate School of Science, Kyoto University, Japan

(2) RIKEN Center for Advanced Intelligence Project, Japan

(3) NTT Communication Science Laboratories, Japan

Synopsis

The surface of the Earth is divided into a dozen or so rigid plates which are in relative motions. In this paper, we attempted to objectively estimate the distribution of plates using a cluster analysis of geodetically observed velocity data. Taking the spherical nature of the Earth into account, we have formulated a clustering algorithm of velocity data in the angular velocity space (AVS). Each data is represented as a straight line and intersection points of lines correspond to the Euler vectors of tectonic plates in the AVS. We developed a method to estimate intersection points of lines, and then applied a hierarchical clustering algorithm to the estimated points. The proposed method was applied to observed geodetic data obtained from all over the world, and successfully reproduced the distribution of major tectonic plates previously recognized based on geological and geophysical knowledge.

Keywords: Plate tectonics, Cluster analysis, Euler vector, GNSS data

1. Introduction

The theory of plate tectonics provides a unified framework on geosciences. It states that the surface of the Earth is divided into a dozen or so rigid plates which are in relative motions, and that most tectonic events such as earthquakes (Isacks et al., 1968) and mountain building (Dewey and Bird, 1970) occur around plate boundaries. According to Euler's fixed point theorem, any rigid motion on the Earth's spherical surface is expressed as a rotation around an axis through the center of the Earth (Bullard et al., 1965), and the angular velocity of the rotation is called the Euler vector (Fig. 1). The angular velocities of plate motions have been estimated based on geological and geophysical data, such as the strike of transform faults, geomagnetic

anomalies on the sea floor and hot spot tracks (Mckenzie and Parker, 1967; Morgan, 1968; Le Pichon, 1968). These and following plate models (e.g. Minster and Jordan, 1978; DeMets et al., 1990) revolutionized the understanding of global tectonics (e.g. Molnar and Tapponnier, 1975).

Since 1980s, dense velocity data obtained by space geodetic techniques such as SLR, VLBI and GNSS have revealed finer structures of tectonic plates (e.g. Larson et al., 1997; Zhang et al., 1999; Sella et al., 2002), which leads to the increase of the number of identified crustal blocks or microplates (e.g. Thatcher, 2007, 2009). So far, in almost all studies, the distribution of tectonic boundaries has been determined based on geological and geophysical knowledge, such as major active faults. However, it is unclear whether these subjective

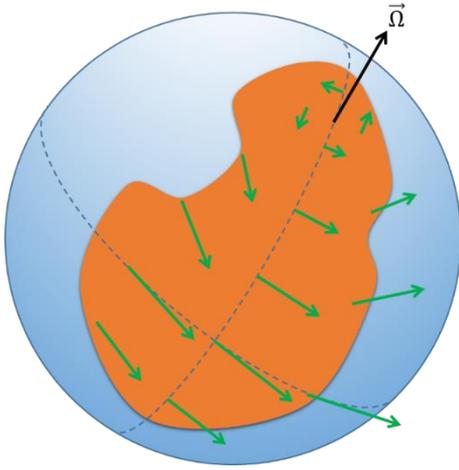


Fig. 1 Rigid plate motion on a sphere. Velocities on the same plate depend on positions, and the motion of a plate is described by the angular velocity $\vec{\Omega}$ called the Euler vector.

partitions are truly relevant. It would be preferable if we can objectively divide tectonic plates based on observed geodetic data.

In recent years, cluster analysis has been applied to GNSS velocity data to objectively reveal a crustal block structure of deformation zones (Simpson et al., 2012). Because the target area is small, Simpson et al. (2012) performed a cluster analysis of GNSS data in the velocity space by regarding the Earth as a flat plane. If a target area is not small, spherical nature of the Earth cannot be neglected because velocities on the same plate or block depend on positions (Fig. 1). In such cases, we cannot apply a cluster analysis to GNSS data in the velocity space. To cope with this problem, Savage and Simpson (2013a, b) combined a cluster analysis in the velocity space with reassignment of stations by calculating Euler vectors of estimated clusters. Savage and Wells (2015) used this method to identify a block structure in a regional scale. However, since the initial configuration of clusters is based on the clustering in the velocity space, it is difficult to apply it to global data.

In this paper, we develop a cluster analysis directly applied to the Euler vectors of tectonic plates. In Section 2, we project geodetic data in the angular velocity space (AVS) and observe that they are represented as straight lines. In Section 3, we construct a clustering algorithm of lines in the AVS. In Section 4, we apply the algorithm to observed

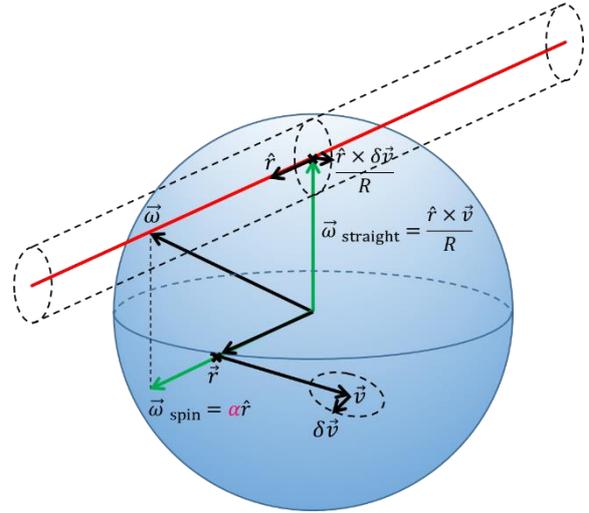


Fig. 2 Representation of velocity data in the AVS. The angular velocity $\vec{\omega}$ compatible with velocity data at a station is represented by a straight line. $\vec{\omega}$ can be linearly decomposed into the spin component $\vec{\omega}_{spin}$ and straight component $\vec{\omega}_{straight}$. $\vec{\omega}_{spin}$ cannot be constrained by velocity data, which leads to the ambiguity of velocity data in the AVS. Errors in velocity data shift $\vec{\omega}_{straight}$, which leads to parallel translation of the line.

geodetic data and confirm that the distribution of plates is reproduced in the global scale. The validity of the results is discussed in Section 5, and conclusions are given in Section 6.

2. Representation of geodetic data in the angular velocity space

We use 3-D vector notation with the origin at the center of the Earth. If an Euler vector $\vec{\Omega}$ is specified as the motion of a plate, the velocity at a position \vec{r} on the plate is given by

$$\vec{v} = \vec{\Omega} \times \vec{r}. \quad (1)$$

This formula states that \vec{r} and \vec{v} are orthogonal to one another and that no vertical component (parallel to \vec{r}) is generated by a motion on a sphere.

Conversely, when a velocity \vec{v} is given at a position \vec{r} , the angular velocity $\vec{\omega}$ compatible with the data is expressed as

$$\vec{\omega}(\alpha) = \frac{\hat{r} \times \vec{v}}{R} + \alpha \hat{r} \quad (-\infty < \alpha < \infty), \quad (2)$$

where R is the radius of the Earth and the hat represents the normalized vector. It can be derived as follows. With a right-handed orthonormal basis

$\{\hat{r}, \hat{v}, \hat{r} \times \hat{v}\}$, a vector $\vec{\omega}$ can be generally expressed as $\vec{\omega} = \alpha \hat{r} + \beta \hat{v} + \gamma \hat{r} \times \hat{v}$. Substituting it for $\vec{\Omega}$ in Eq. (1) and comparing the coefficients of both sides, we obtain the condition for α , β and γ as Eq. (2).

In practice, we use latitude and longitude (θ, ϕ) to express a position and the east and north components (v_E, v_N) of velocity. Rewriting Eq. (2) with them in the geocentric coordinates, we obtain

$$\vec{\omega} = \frac{1}{R} \begin{pmatrix} -v_E \sin \theta \cos \phi + v_N \sin \phi \\ -v_E \sin \theta \sin \phi - v_N \cos \phi \\ v_E \cos \theta \end{pmatrix} + \alpha \begin{pmatrix} \cos \theta \cos \phi \\ \cos \theta \sin \phi \\ \sin \theta \end{pmatrix}. \quad (3)$$

In Eq. (2), the free parameter α represents the ambiguity of the Euler vector for given velocity data. Because of this ambiguity, the Euler vector for velocity data is not represented by a point but by a straight line in the AVS (Fig. 2). The origin of this ambiguity can be understood with the aid of the decomposition of angular velocity into spin and straight components (Matsuyama and Iwamori, 2016). Eq. (2) can be linearly decomposed as $\vec{\omega} = \vec{\omega}_{\text{spin}} + \vec{\omega}_{\text{straight}}$ defined by $\vec{\omega}_{\text{spin}} = \alpha \hat{r}$ and $\vec{\omega}_{\text{straight}} = (\hat{r} \times \hat{v})/R$. Here, $\vec{\omega}_{\text{spin}}$ is a component parallel to \vec{r} generating a rotation around the axis through \vec{r} , and the parameter α controls the magnitude of rotation. This rotation does not affect the velocity at \vec{r} , so α cannot be constrained from velocity data at all. This causes the ambiguity of the Euler vector $\vec{\omega}$ in the AVS. On the other hand, $\vec{\omega}_{\text{straight}}$ is perpendicular to $\vec{\omega}_{\text{spin}}$ and uniquely determined by velocity data. Note that errors in velocity data only shift $\vec{\omega}_{\text{straight}}$, which leads to parallel translation of the line.

Let us consider the relation between the AVS and the velocity space (Fig. 3). In a small region compared to the whole Earth, positions \vec{r} are nearly identical for all stations. The lines of the Euler vector expressed by Eq. (2) are almost parallel in the AVS and hence the relative position of lines is determined by \hat{v} . This means that the representation in the AVS reduces to that in the velocity space when we consider a small region. For a larger region, on the other hand, the directions of lines in the AVS vary with positions of stations and the representation in the AVS becomes different from that in the velocity space. Then, the former captures the spherical property of the Earth while the latter does not.

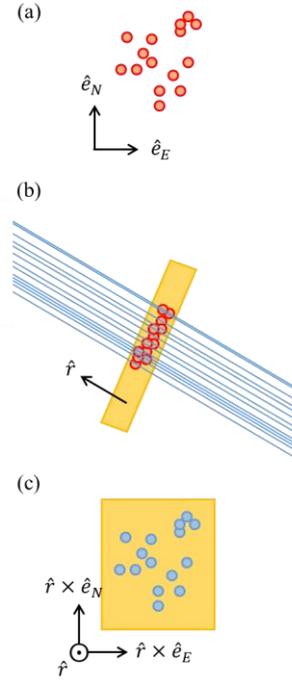


Fig. 3 Relation between the representation in the AVS and that in the velocity space when we treat stations in a small region. (a) Representation in the velocity space. (b) Representation in the AVS. Velocity data are shown by almost parallel lines. (c) A cross sectional view of the AVS. The cross section is taken to be vertical to the parallel lines. The relative positions of lines reflect the distribution of velocity data.

The location of Euler poles has been estimated using the strike of transform faults (Morgan, 1968; Le Pichon, 1968). We can consider this problem in the AVS as follows. Since the strike of a transform fault only contains the direction of velocity, its information can be represented as a straight line $\beta \hat{v}$ ($-\infty < \beta < \infty$) in the velocity space. The same derivation with Eq. (2) shows that the angular velocity compatible with the data of a transform fault is represented as a flat plane $\alpha \hat{r} + \beta \hat{r} \times \hat{v}$, which is normal to \hat{v} and passes through the origin in the AVS (Fig. 4a). If several transform faults exist on boundaries of the same pair of plates in a relative angular velocity $\vec{\Omega}$, the Euler vector is estimated as the intersection of the planes, which is represented as a straight line through $\vec{\Omega}$ and the origin (Fig. 4b). Since we cannot estimate the magnitude from these data, we can project them onto the surface of the Earth without loss of information. Then, the planes are projected to great

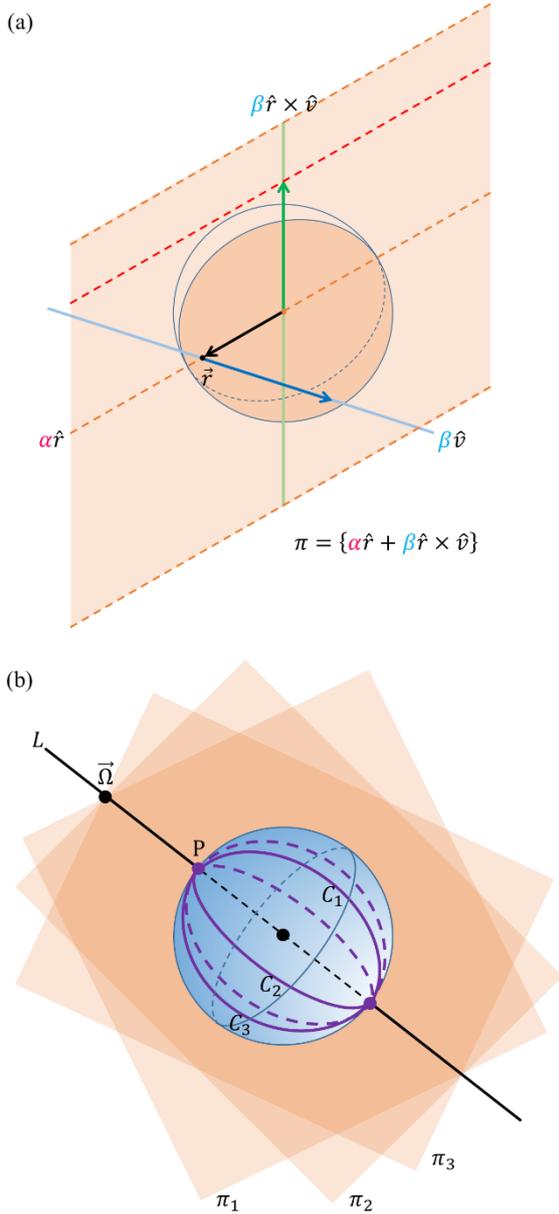


Fig. 4 Estimation of the Euler pole from the strike of transform faults in the AVS. (a) The strike of a transform fault shows the direction of velocity, so its information is expressed as a plane through the origin in the AVS. (b) If several transform faults exist on boundaries of the same pair of plates, these planes intersect at a line through the origin and the Euler vector $\vec{\Omega}$. If we neglect the magnitude, the planes are projected to great circles and the line to a pair of antipodal points.

circles and the line to a pair of antipodal points, which coincides with the construction of the Euler pole by estimating an intersection point of great circles normal to transform faults (Morgan, 1968; Le Pichon, 1968).

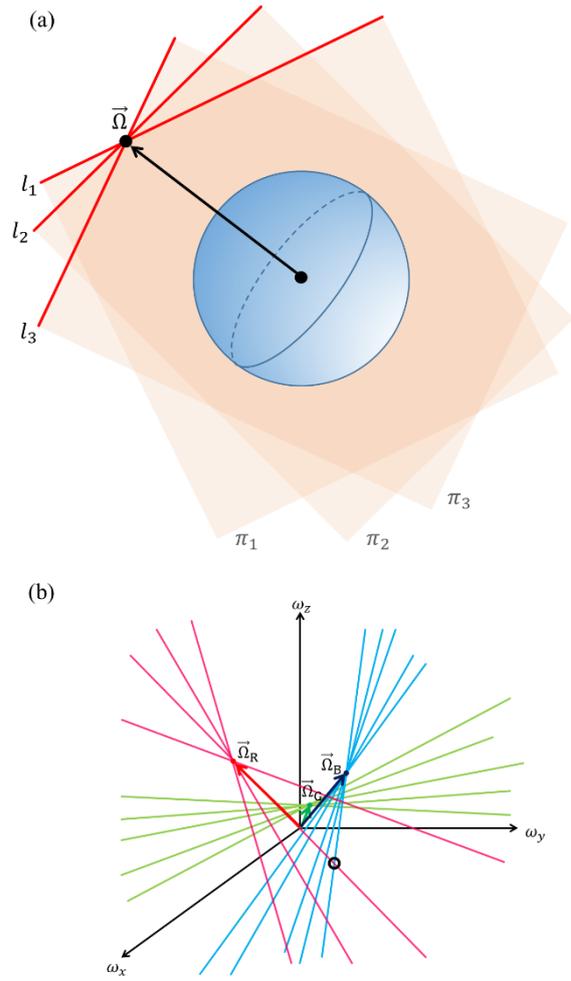


Fig. 5 Estimation of the Euler vector from velocity data in the AVS. (a) The Euler vector of a rigid plate is expressed as an intersection point of lines. (b) If we have data on several plates, lines intersect at different points corresponding to Euler vectors of the plates. Even if a cyan and a magenta lines intersect at the black circle, they should be classified to different plates; it shows the difficulty in estimating intersection points in Section 3.

On the other hand, velocity data are represented as lines in the AVS. If velocity data on the same plate are given, the corresponding lines in the AVS intersect at a point corresponding to the Euler vector $\vec{\Omega}$ of the plate (Fig. 5a). In practice, lines do not exactly intersect at a point due to errors in velocity data and internal deformation in a plate, so the intersection $\vec{\Omega}$ is estimated by the least-square method. The solution is analytically expressed as

$$\mathbf{\Omega} = \mathbf{G}^{-1}\mathbf{L}, \quad (4)$$

with

$$\mathbf{G} = \sum_i (\mathbf{I} - \hat{\mathbf{r}}_i \hat{\mathbf{r}}_i^T), \quad \mathbf{L} = \sum_i \boldsymbol{\omega}_i(0). \quad (5)$$

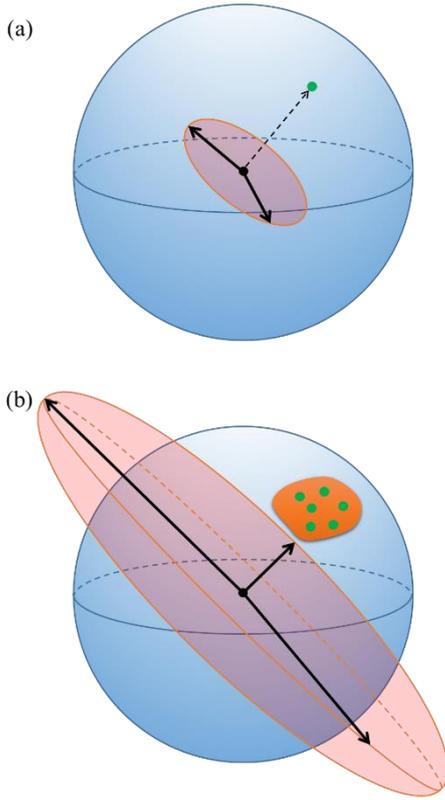


Fig. 6 Precision of the estimated Euler vector. Red ellipsoids represent precision (the net number of data) \mathbf{G} in each direction. (a) Data on a station has no information on its direction and full information on the perpendicular directions (an allipsoid flattens out in a circle). (b) If stations localize in a small region, the precision is low in its direction.

The subscripts i refer to the labels of stations, and the matrix notation is used for convenience (\mathbf{I} is the unit matrix). The derivation is given in Appendix. If we suppose a system of unit point masses at $\hat{\mathbf{r}}_i$ with velocity \mathbf{v}_i , its inertial moment and total angular momentum coincide with \mathbf{G} and \mathbf{L} , respectively. Then $\boldsymbol{\Omega}$ corresponds to the angular velocity of the system. It should be noted that the precision of estimation heavily depends on the distribution of stations on a plate. As inferred from Fig. 3, if stations localize in a small region, the directions of lines in the AVS are almost parallel. Then, the component of the Euler vector in this direction can be hardly constrained. The precision can be quantified by \mathbf{G} (e.g. Menke, 1989). Each term $(\mathbf{I} - \hat{\mathbf{r}}_i \hat{\mathbf{r}}_i^T)$ has principal values 0 to the direction $\hat{\mathbf{r}}_i$ and 1 to the perpendicular directions, that is, it has no information on the spin component parallel to $\hat{\mathbf{r}}_i$ while full information on the straight

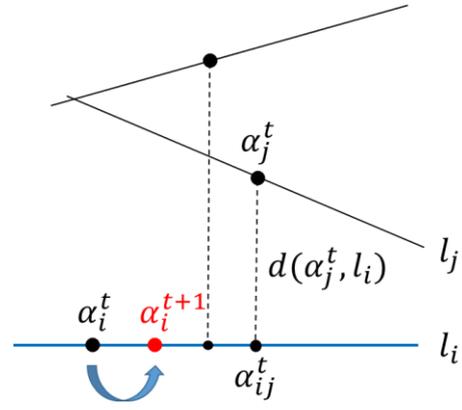


Fig. 7 The algorithm for estimation of intersection points of lines. An estimated point α_i^t at step t is updated to the center of mass α_i^{t+1} of points α_{ij}^t weighted according to the distance.

component (Fig. 6a). It implies that \mathbf{G} expresses the net number of data in each direction. Localized distribution of stations gives a small value of \mathbf{G} in its direction (Fig. 6b).

When we have data on several plates, lines intersect at different points corresponding to the Euler vectors of these plates (Fig. 5b). It suggests that we can reveal the distribution of plates by identifying intersection points of lines in the AVS, and then we can estimate the Euler vector of each plate using Eq. (4).

3. Clustering method

In this section, in order to estimate Euler vectors of plate motions, we propose a clustering method of velocity data in the AVS. The procedure has two steps: estimation of a tentative intersection point on each line and clustering of the estimated points.

As we saw in the previous section, velocity data are projected to straight lines in the AVS, and lines of data in the same plate are expected to intersect at a point. We suppose that each line l_i given by Eq. (2) has a true angular velocity $\vec{\omega}_i(\alpha_i^*)$ which is close to (ideally identical to) the Euler vector of a plate. Here, we estimate the parameter α_i^* to specify an intersection point of the line. The difficulty of the problem is that lines of two stations in different plates may intersect coincidentally. Therefore, we need to find a point where several lines converge (Fig. 5b). For this purpose, we developed the following procedure.

Suppose that we have velocity data at N stations. We set initial values as $\alpha_i^0 = 0$ considering that the smallest magnitude of angular velocity (i.e. no spin motion) would be natural to explain velocity data at one station. We then repeatedly update the parameter from α_i^t to α_i^{t+1} for T times by the relation

$$\alpha_i^{t+1} = \frac{\sum_{j=1}^N w_{ij}^t \alpha_{ij}^t}{\sum_{j=1}^N w_{ij}^t}, \quad w_{ij}^t = \exp[-\lambda^t d(\alpha_j^t, l_i)]. \quad (6)$$

Here α_{ij}^t is obtained by orthogonally projecting α_j^t onto the line l_i , and $d(\alpha_j^t, l_i)$ is the distance between a point α_j^t and a line l_i in the AVS (Fig. 7). We use α_i^T as the estimation of α_i^* .

Eq. (6) is a weighted mean of all points α_j^t projected onto l_i , so it locates where many lines accumulate. The weight w_{ij}^t suppresses the effect of far points. The length scale is controlled by $1/\lambda^t$ which is specified by hand for each step. λ^t is initially set small to count points in vast space, and gradually increases to focus on the neighborhood of the estimated point. It plays a similar role with temperature in simulated annealing (Kirkpatrick et al., 1983; Černý, 1985). The choice of T and λ^t in this study will be described in the next section.

Once data reduce to points in the AVS, the hierarchical agglomerative clustering (HAC) algorithm (e.g. Simpson et al., 2012) can be applied. In the beginning, each point constitutes a cluster, so the number of clusters K is equal to N . We select the closest two clusters in the AVS, and replace them for a cluster at the center of mass (the mass is set to the number of points contained in the cluster); then K decreases by 1. This procedure is repeated until all points are combined to one cluster ($K = 1$). Following the procedure in reverse, the hierarchical structure of clusters can be visualized as a dendrogram (see Fig. 9b).

For each K , the accuracy of fitting can be measured by the root mean squares (RMS):

$$\left[\frac{1}{N} \sum_{i=1}^N d(\vec{\Omega}_{c(i)}, l_i)^2 \right]^{1/2}. \quad (7)$$

Here $c(i)$ is the label of the cluster to which station i belongs and $\vec{\Omega}_k$ is the best-fit Euler vector of the cluster k calculated with Eq. (4). RMS misfit monotonically decreases with K , and we adopt a number where RMS settles down to an almost constant value as the optimal number of clusters.

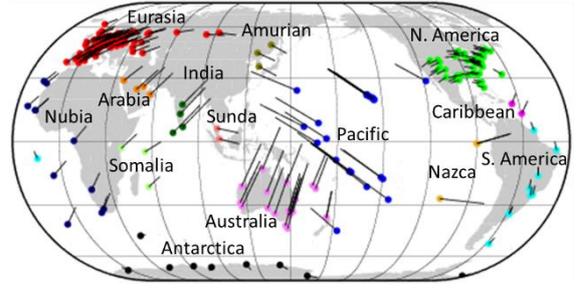


Fig. 8 Geodetically observed velocity data and their classification to 14 plates (Altamimi et al., 2012). Points and arrows indicate the locations and velocities of stations, respectively. Data on different plates are distinguished by colors.

4. Application to global geodetic data

Altamimi et al. (2012) constructed a plate motion model (ITRF2008 plate model) using global velocity data measured by GNSS, SLR, VLBI and DORIS (Altamimi et al., 2011). The distribution of plates was determined beforehand based on geological and geophysical knowledge, and their Euler vectors were estimated from velocity data at 206 sites (Fig. 8). We apply the clustering method to these velocity data to identify plate structures without referring to any other information.

We first estimate intersection points of lines in the AVS. For a given total step T , we set the parameter at step t as $\lambda^t = 10t/T$ (rad/Gyr) $^{-1}$. This means that the length scale $1/\lambda^t$ of neighborhood is infinite at first, and gradually decreases to 0.1 (rad/Gyr) at the last step. Varying the values of T every digit from 100 to 1,000,000, we confirmed that estimated points converge to almost identical positions for $T \geq 10,000$. We use the result of $T = 1,000,000$ (Fig. 9a) in the following analysis.

We apply the HAC algorithm to the obtained angular velocity points. The hierarchical structure is shown as the dendrogram in Fig. 9(b). The large RMS for $K = 2$ indicates that the motion of the Pacific plate is quite different from the others. For $K = 3$, the Arabia, India and Australia plates are separated as an independent block. It breaks for $K = 5$; although the Arabia plate was a part of Eurasia plate, and the India and Australia plates

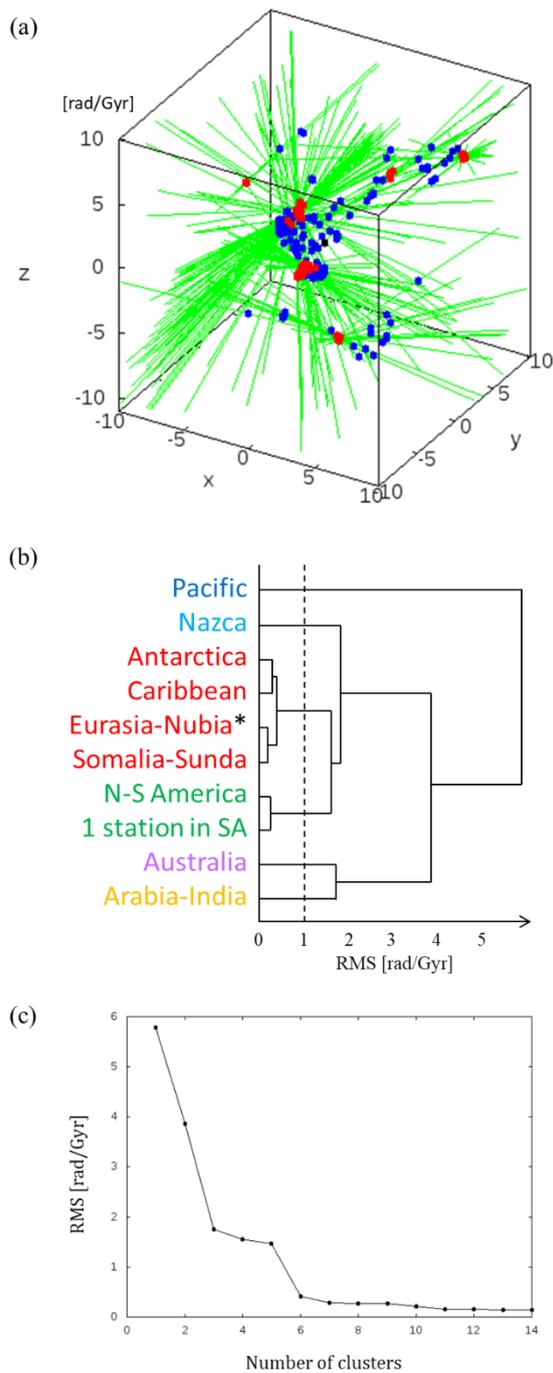


Fig. 9 (a) Estimated intersection points (red) on lines corresponding to observed velocity data in the AVS. Blue points show the initial locations of intersection points for iteration. (b) The hierarchical structure of the cluster analysis. The dendrogram is drawn until 10 clusters. The symbol * notifies that some stations in the Nubia plate belong to the Somalia-Sunda cluster. (c) The plot of RMS misfit for different numbers of clusters. The value settles down for $K \geq 6$, so we take 6 as the optimal number of major plates. N-S, North-South; SA, South America.

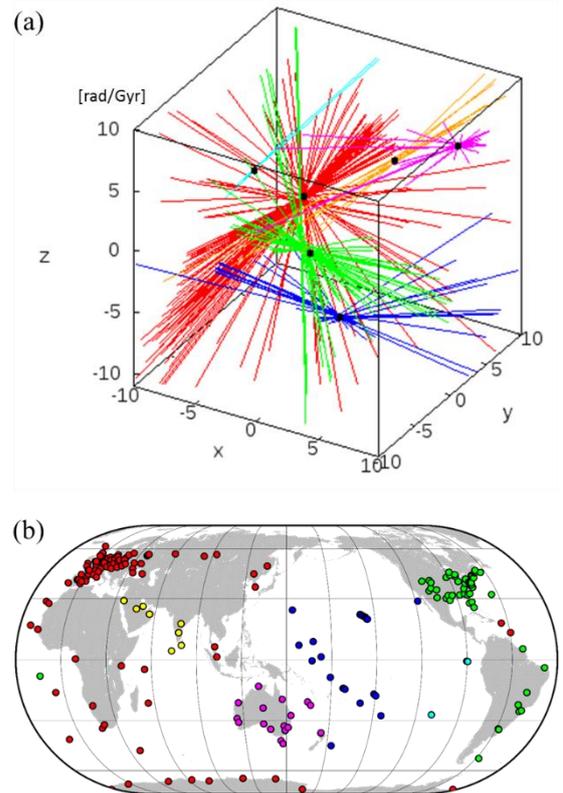


Fig. 10 The distribution of clusters (a) in the AVS and (b) in the geographical map for $K = 6$. (a) Lines of different clusters are well separated in the AVS. Black dots indicate estimated Euler vectors of each cluster. (b) The assignment of stations is consistent with the ITRF2008 model (Altamimi et al., 2012).

were regarded as a single one in the early years of plate tectonics, the clustering result shows that the Arabia and India plates have similar motions, and the Australia plate forms a separate cluster. The Nazca plate and the North-South America plate are also separated from the Eurasia plate for $K = 4, 6$, respectively. Note that the Caribbean plate belongs not to the America plate but to the Eurasia plate. For $K = 7, 8$, the Antarctica and Caribbean plates are separated from the Eurasia plate. There is no false assignment until $K = 8$. Erroneous partitions appear for more clusters: one station in the South America plate constitutes an isolated cluster for $K = 9$, and the Somalia, Sunda and a part of Nubia plates are separated for $K = 10$.

The plot of RMS misfit (Fig. 9c) indicates that partitions into $K \leq 5$ require a large misfit, and partitions into $K \geq 7$ does not dramatically improve the fitting. We hence adopt $K = 6$ as the

Table 1 Estimated Euler vectors of each cluster.

Clus ter	Sta tions	Euler vector (rad/Gyr)			RMS (rad/Gyr)	Principal values of \mathbf{G}			Principal axis (P3)	
		x	y	z		P1	P2	P3	Latitude	Longitude
1	99	-0.48	-2.55	3.65	0.527	89.35	84.22	24.43	50.93°	7.93°
2	23	-1.86	5.13	-10.57	0.124	20.93	20.33	4.74	3.14°	-167.23°
3	54	-0.04	-2.51	-0.97	0.335	52.24	45.88	9.89	34.26°	-80.39°
4	8	5.72	-0.35	7.21	0.230	7.97	7.33	0.70	18.51°	63.25°
5	3	-1.67	-7.56	7.93	0.023	3.00	2.80	0.20	9.23°	84.33°
6	19	7.40	5.67	5.92	0.088	18.68	17.74	1.58	30.99°	-35.09°

optimal number of major plates. Shown in Fig. 10 is the distribution of 6 clusters expressed in the AVS and in the geographical map. Compared with the ITRF2008 plate model, Cluster 1 (red) consists of the Amurian, Antarctica, Caribbean, Eurasia, Nubia, Somalia and Sunda plates, Cluster 2 (blue) Pacific, Cluster 3 (green) North and South America, Cluster 4 (yellow) Arabia and India, Cluster 5 (cyan) Nazca, and Cluster 6 (magenta) Australia plate, without any false assignment of stations. The classification of major plates is successfully reproduced in the global scale only using velocity data.

5. Discussion

5.1 The accuracy of estimated Euler vectors

The Euler vector (Eq. 4) and RMS misfit (Eq. 7) of each cluster are calculated for $K = 6$ in Table 1. RMS misfit is the largest in Cluster 1. This is natural because the Antarctica and Caribbean plates are separated for $K = 7, 8$; if they are removed (there remain 88 stations), RMS decreases to 0.308. The second largest is Cluster 3. This turns out to be caused by a poor estimation of intersection points for this cluster. As shown in Fig. 11, lines of the North (green) and South (cyan) America intersect in different points in the AVS, but the proposed algorithm cannot distinguish them. The estimated points in the South America (blue) are pulled by a large number of points in the North America (red). That is why a station in the South America is separated for $K = 9$ (Fig. 9b). It indicates that the algorithm is incomplete and need to be improved.

As mentioned in Section 2, the precision of the Euler vector highly depends on its components. The principal values $P1 > P2 > P3$ of \mathbf{G} express the net number of data in the corresponding directions.

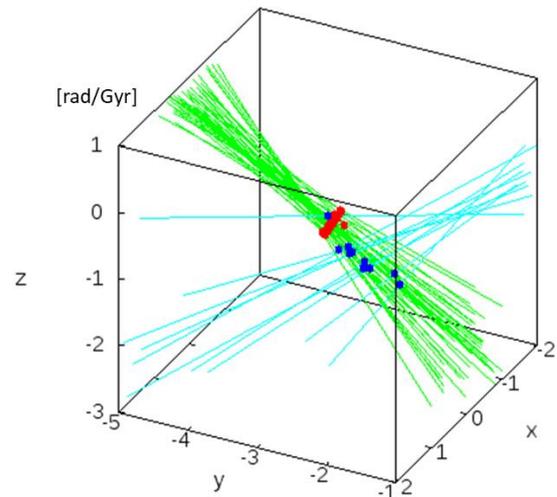


Fig. 11 Estimation of intersection points for the lines of the North (green) and South (cyan) America in the AVS. It can be seen that these lines intersect at different points. Estimated points of the North America (red) capture the intersection point, but those of the South America (blue) are scattered.

$P1$ and $P2$ are close to the actual number of stations. In contrast, $P3$ is much smaller than the others: it is less than unity for Clusters 4 and 5, and less than two even for Cluster 6 (Australia) with 19 stations. The axis shows a center of the distribution of stations (Fig. 6b). We must keep in mind that the Euler vector is hardly determined in this direction when we use regional data.

5.2 Comparison with other clustering methods

Let us evaluate the performance of the proposed clustering algorithm by comparing it with other simple methods. We apply the HAC algorithm and K-means method (e.g. MacQueen, 1967) modified to line data, and their combination (we call it K+HAC) for $K \leq 14$. We measure the performance

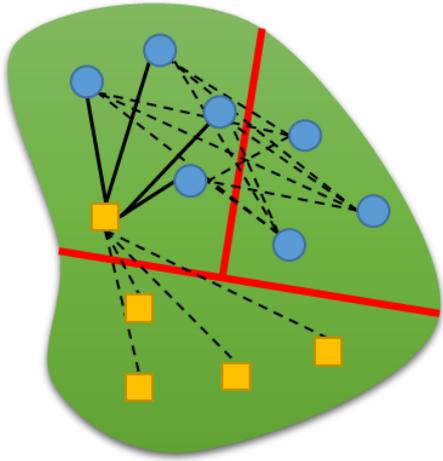


Fig. 12 Schematic illustration of the index to evaluate the performance of clustering results. Red lines show the boundary of plates, and two types of marks indicate stations of two clusters. Dotted lines (pairs in different plates but in the same cluster) should not be penalized (especially lines connecting blue points) because the number of clusters are less than the number of plates. Solid lines (pairs in the same plate but in different clusters) are penalized because they indicate false divisions of plates.

Table 2 The number of false divisions.

K	HAC	K-means	K+HAC	Proposed
2	0	400	0	0
3	0	483	0	0
4	52	17	0	0
5	54	17	0	0
6	60	14	9	0
7	90	10	9	0
8	90	10	9	0
9	108	34	9	9
10	133	34	297	37
11	235	108	297	51
12	287	13	297	57
13	287	78	297	57
14	293	400	361	61

by the consistency with the ITRF2008 model, and quantify as the number of false divisions, that is, pairs in the same plate of the ITRF2008 model but in different clusters (Fig. 12). This is motivated by the Rand index (Rand, 1971). The value 0 indicates a perfect consistency with the ITRF2008 model, and larger values indicate worse consistency with

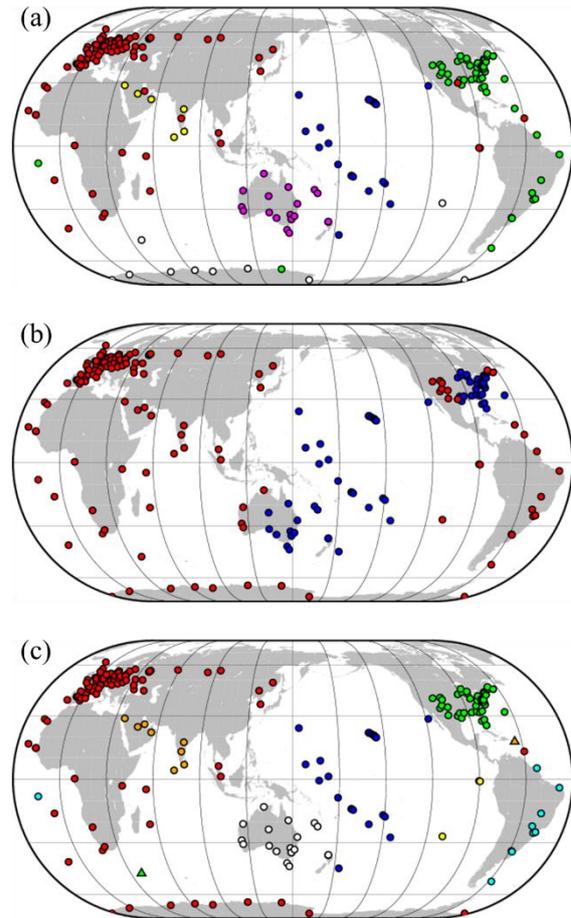


Fig. 13 Comparison of the results of different clustering methods. (a) The result of the HAC algorithm for $K = 6$. The distribution of stations is geographically discontinuous for some clusters. (b) The result of the K-means method for $K = 2$. The Pacific and parts of the North America and Australia plates form a cluster. (c) The result of the K+HAC method for $K = 9$. The North and South America are correctly divided, although two stations constitute isolated clusters (triangles).

it. Those for different clustering methods are presented in Table 2, and we discuss their implications in the following.

First, the HAC algorithm works well only for small K , and some clusters become geographically discontinuous (Fig. 13a). This is caused by the reason mentioned in Section 3: if two lines of velocity data on different plates happen to cross one another, the intersection point is mistakenly picked up as the Euler vector of them. A more robust clustering algorithm is necessary to avoid such misclassifications. Next, the K-means method does

not work well for $K \leq 3$, caused by the tendency of this algorithm to generate similar sizes of clusters (Fig. 13b). The performance improves for $K > 3$. Another drawback is that sudden changes of cluster boundary sometimes occur because the K-means method is not hierarchical. Finally, we combined these two methods to cover the shortcomings of one another (K+HAC). We first perform the K-means method for a relatively large number of clusters (set at 20 in this study), and then apply the HAC algorithm to the obtained cluster centers. It is hierarchical, robust (the HAC is applied to point data) and free from the tendency of similar sizes (the HAC generates size differences). This method works better than the above two methods, and it is advantageous that the North and South America plates are correctly divided for $K = 9$ (Fig. 13c).

The proposed algorithm performs best among the other simple methods. The results show perfect consistencies until $K = 8$ and relatively small false divisions for larger K . This algorithm avoids picking up intersection of few lines by widely looking for intersection points at first (i.e. λ^f is initially small). However, this property also leads to the wrong unification of different intersection points (Fig. 11). In fact, some false divisions for $K > 8$ are due to fragmentation of the South America plate. It is difficult to construct an algorithm satisfying these conflicting requirements.

6. Conclusions

We have developed a methodology to explore a global plate distribution from geodetically observed velocity data. Data at a single station are mapped to a straight line in the AVS reflecting one-parameter ambiguity of the Euler vector. Block structures can be identified by finding intersection points of lines in the AVS. Advantages of this representation are that it naturally takes account of the spherical property of the Earth, and it is independent of coordinates. Moreover, different kinds of data such as GNSS velocities and strikes of transform faults can be expressed in a unified manner, so a joint analysis can be performed in the AVS.

To objectively reveal the global distribution of plates, we have constructed a clustering method of straight lines. It was applied to global geodetic data

and reproduced the distribution of major plates consistent with the ITRF2008 plate model by only using velocity data. This worked best among other simple methods, but it could not distinguish different intersection points of the North and South America. There is still room for improvement in the estimation of intersection points.

Although we focused on the block structures of the crust in this paper, internal deformation is observed in continental deformation zones. It would be interesting to investigate how such phenomena can be visualized and organized in the AVS.

References

- Altamimi, Z., Collilieux, X. and Métivier, L. (2011): ITRF2008: An improved solution of the International Terrestrial Reference Frame, *J. Geod.*, 85(8), pp. 457-473.
- Altamimi, Z., Métivier, L. and Collilieux, X. (2012): ITRF2008 plate motion model, *J. Geophys. Res.*, 117, B07402.
- Bullard, E. C., Everett, J. E. and Smith, A. G. (1965): The fit of the continents around the Atlantic, in *A Symposium on Continental Drift*, ed. By Blackett, P. M. S., Bullard, E. and Runcorn S. K., *Phi. Trans. Roy. Soc. London, A*, 1088, pp. 41-51.
- Černý, V. (1985): Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm, *J. Optimization Theory and Applications* 45, pp. 41-51.
- DeMets, C. R., Gordon, R. G., Argus, D. and Stein, S. (1990): Current plate motions. *Geophys. J. Inter.*, 101, pp. 425-478.
- Dewey, J. F. and Bird, J. M. (1970): Mountain Belts and the New Global Tectonics, *J. Geophys. Res.*, 75, 14, pp. 2625-2645.
- Isacks, B., Oliver, J. and Sykes, L. R. (1968): Seismology and the new global tectonics, *J. Geophys. Res.*, 73, pp. 5855-5899.
- Kirkpatrick, S., Gelatt Jr, C. D. Vecchi, M. P. (1983): Optimization by Simulated Annealing, *Science*, 220, 4598, pp. 671-680.
- Larson, K. M., Freymueller, J. T., Philipson, S. (1997): Global plate velocities from the Global Positioning System, *J. Geophys. Res.*, 102, pp. 9961-9981.

Le Pichon, X. (1968): Sea-floor spreading and continental drift, *J. Geophys. Res.*, 73, pp. 3661-3695.

MacQueen, J. B. (1967): Some methods for classification and analysis of multivariate observations, *Proc. 5th Berkeley Sympo. Math. Stat. Prob.*, Univ. California Press, pp. 281-297.

Matsuyama, T. and Iwamori, H. (2016): Analysis of plate spin motion and its implications for strength of plate boundary., *Earth Planet Sp.*, 68:36.

McKenzie, D. P. and Parker, R. L. (1967): The north Pacific: An example of tectonics on a sphere, *Nature*, 216, pp. 1276-1280.

Menke, W. (1989): *Geophysical data analysis: discrete inverse theory*, Academic Press.

Minster, J. B. and Jordan, T.H. (1978): Present-day plate motions, *J. Geophys. Res.*, 83, pp. 5331-5354.

Molnar, P. and Tapponnier, P. (1975): Cenozoic tectonics of Asia: Effects of a continental collision, *Science*, 189, pp. 419-426.

Morgan, W. J. (1968): Rises, trenches, great faults, and crustal blocks, *J. Geophys. Res.*, 73, pp. 1959-1981.

Rand, W. M. (1971): Objective criteria for the evaluation of clustering methods, *J. American Stat. Assoc.*, 66, 336, pp. 846-850.

Savage, J. C., and Simpson, R. W. (2013a): Clustering of GPS velocities in the Mojave Block, southeastern California, *J. Geophys. Res. Solid Earth*, 118, pp. 1747-1759.

Savage, J. C., and Simpson, R. W. (2013b): Clustering of velocities in a GPS network spanning the Sierra Nevada Block, the Northern Walker Lane Belt, and the Central Nevada Seismic Belt, California-Nevada, *J. Geophys. Res. Solid Earth.*, 118, pp. 4937-4947.

Savage, J. C., and Wells, R. E. (2015): Identifying block structure in the Pacific northwest, USA. *J. Geophys. Res. Solid Earth*, 120, pp. 7905–7916.

Sella, G. F., Dixon, T. H. and Mao, A (2002): REVEL: A model for Recent plate velocities from space geodesy, *J. Geophys. Res.*, 107, B4.

Simpson, R. W., Thatcher, W. and Savage, J. C. (2012): Using cluster analysis to organize and explore regional GPS velocities, *Geophys. Res. Lett.*, 39, L18307.

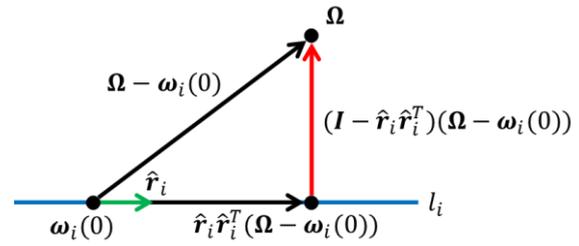


Fig. A1 Sketch for the calculation of the distance $d(\mathbf{\Omega}, l_i)$ between a point $\mathbf{\Omega}$ and a line l_i .

Thatcher, W. (2007): Microplate model for the present-day deformation of Tibet, *J. Geophys. Res.*, 112, B01401.

Thatcher, W. (2009): How the continents deform: The evidence from tectonic geodesy, *Annu. Rev. Earth Planet. Sci.*, 37, pp. 237-262.

Zhang, Q., Zhu, W. and Xiong, Y. (1999): Global plate motion models incorporating the velocity field of ITRF96, *Geophys. Res. Lett.*, 26, pp. 2813-2816.

Appendix

Here, we derive Eqs. (4) and (5) in Section 2. The squared sum $D(\mathbf{\Omega})$ of the distances between a point $\mathbf{\Omega}$ and lines l_i is calculated as (see Fig. A1)

$$D(\mathbf{\Omega}) = \sum_i d(\mathbf{\Omega}, l_i)^2 = \sum_i \|(I - \hat{r}_i \hat{r}_i^T)(\mathbf{\Omega} - \boldsymbol{\omega}_i(0))\|^2 = \sum_i (\mathbf{\Omega} - \boldsymbol{\omega}_i(0))^T (I - \hat{r}_i \hat{r}_i^T) (\mathbf{\Omega} - \boldsymbol{\omega}_i(0)), \quad (\text{A1})$$

where $\boldsymbol{\omega}_i(0)$ is defined by Eq. (2) and we used $(I - \hat{r}_i \hat{r}_i^T)^T (I - \hat{r}_i \hat{r}_i^T) = I - \hat{r}_i \hat{r}_i^T$. Eq. (A1) is a quadratic function of $\mathbf{\Omega}$ and has a unique minimum given by the root of its derivative

$$D'(\mathbf{\Omega}) = 2 \sum_i (I - \hat{r}_i \hat{r}_i^T) (\mathbf{\Omega} - \boldsymbol{\omega}_i(0)) = 2(\mathbf{G}\mathbf{\Omega} - \mathbf{L}), \quad (\text{A2})$$

where we used the orthogonality $\hat{r}_i^T \boldsymbol{\omega}_i(0) = 0$ and the notation in Eq. (5). We obtain Eq. (4) by solving $D'(\mathbf{\Omega}) = 0$. By completing the square, Eq. (A1) is rewritten as

$$D(\mathbf{\Omega}) = D(\mathbf{\Omega}_0) + (\mathbf{\Omega} - \mathbf{\Omega}_0)^T \mathbf{G} (\mathbf{\Omega} - \mathbf{\Omega}_0), \quad (\text{A3})$$

where $\mathbf{\Omega}_0$ is the root of Eq. (A2). The Hesse matrix \mathbf{G} expresses the precision of estimation which is independent of velocity data.

(Received June 13, 2018)