

Quantifying variability in earthquake rupture models using multidimensional scaling: application to the 2011 Tohoku earthquake

Hoby N. T. Razafindrakoto,¹ P. Martin Mai,¹ Marc G. Genton,² Ling Zhang¹ and Kiran K. S. Thingbaijam¹

¹Division of Physical Sciences and Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia.
E-mail: hoby.razafindrakoto@kaust.edu.sa

²Division of Computer, Electrical and Mathematical Sciences and Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

Accepted 2015 February 19. Received 2015 February 18; in original form 2014 July 21

SUMMARY

Finite-fault earthquake source inversion is an ill-posed inverse problem leading to non-unique solutions. In addition, various fault parametrizations and input data may have been used by different researchers for the same earthquake. Such variability leads to large intra-event variability in the inferred rupture models. One way to understand this problem is to develop robust metrics to quantify model variability. We propose a Multi Dimensional Scaling (MDS) approach to compare rupture models quantitatively. We consider normalized squared and grey-scale metrics that reflect the variability in the location, intensity and geometry of the source parameters. We test the approach on two-dimensional random fields generated using a von Kármán autocorrelation function and varying its spectral parameters. The spread of points in the MDS solution indicates different levels of model variability. We observe that the normalized squared metric is insensitive to variability of spectral parameters, whereas the grey-scale metric is sensitive to small-scale changes in geometry. From this benchmark, we formulate a similarity scale to rank the rupture models. As case studies, we examine inverted models from the Source Inversion Validation (SIV) exercise and published models of the 2011 Mw 9.0 Tohoku earthquake, allowing us to test our approach for a case with a known reference model and one with an unknown true solution. The normalized squared and grey-scale metrics are respectively sensitive to the overall intensity and the extension of the three classes of slip (very large, large, and low). Additionally, we observe that a three-dimensional MDS configuration is preferable for models with large variability. We also find that the models for the Tohoku earthquake derived from tsunami data and their corresponding predictions cluster with a systematic deviation from other models. We demonstrate the stability of the MDS point-cloud using a number of realizations and jackknife tests, for both the random field and the case studies.

Key words: Instability analysis; Spatial analysis; Earthquake source observations.

1 INTRODUCTION

Spatial and temporal analysis of rupture across a fault area is a useful tool for understanding the complexity of earthquake sources and the influence of such complexity on seismic and tsunami hazard assessment. Since the 1980s, finite-fault models have been developed to characterize the kinematics of the earthquake rupture process. These models are increasingly generated in an almost routine fashion and used in subsequent seismological research. For the purpose of earthquake early warning, Minson *et al.* (2014) even proposed real-time inversions for slip models of finite faults. Soon after an earthquake occurs, source studies are now able to provide corresponding

rupture models based on different data sets (e.g. seismic waveform, GPS and/or InSAR data), alternative assumptions in the problem setup, and utilizing different inversion algorithms such as the multi-time window approach (e.g. Olson & Apsel 1982; Hartzell & Heaton 1983) or non-linear inversion with a predefined analytical source-time function (e.g. Cotton & Campillo 1995; Liu & Archuleta 2004; Tinti *et al.* 2005). Resulting rupture models often differ widely, although they typically all fit the data well. Kinematic source models of the 2011 Mw 9.0 Tohoku earthquake are prime examples of this variability.

The 2011 Tohoku event occurred off the Pacific coast of north-eastern Honshu, Japan. Data from such well-recorded earthquake

have been used in numerous source studies to capture the rupture process. These studies were based on various data sets including seismic data (e.g. Hayes 2011; Lay *et al.* 2011; Shao *et al.* 2011), geodetic data (e.g. Feng & Jónsson 2012), tsunami data (e.g. Fujii *et al.* 2011; Satake *et al.* 2013), or a combination of different types of data sets (e.g. Simons *et al.* 2011; Yue & Lay 2013). Most of the proposed rupture models suggest that the largest slip (over 50 m) was near the trench, although there are significant discrepancies regarding the spatial pattern of the rupture process. Such discrepancies eventually affect the evaluation of seismic and tsunami hazard (Goda *et al.* 2014). Without a rigorous way to compare slip models quantitatively, it is therefore difficult to assess their common and stable features, as well as the limits of their resolution.

Statistical techniques have been applied to characterize and quantify the complexity of rupture models (e.g. Somerville *et al.* 1999; Mai & Beroza 2002; Lavallée *et al.* 2006). Significant efforts have also been expended on assessing the uncertainties in finite-fault source inversions. Beresnev (2003) discussed the levels of uncertainties in source inversion. To account for uncertainties, Piatanesi *et al.* (2007), for instance, performed a statistical analysis of large sets of inferred source models, while others used Bayesian techniques (e.g. Monelli & Mai 2008; Minson *et al.* 2013; Razafindrakoto & Mai 2014). In this context, the Source Inversion Validation (SIV; Mai *et al.* 2007; Page *et al.* 2011) project seeks to understand and quantify the variability in slip-model inversions. There have also been efforts to characterize the differences and similarities between rupture models. Shao & Ji (2012), for instance, used residual analysis and spatial cross-correlation to quantify spatial heterogeneity. These standard methods, however, cannot classify and rank rupture models.

In this study, we adopt an embedding method based on multidimensional scaling (MDS) to compare rupture models quantitatively. This approach embeds the dissimilarities between all pairs of slip models in low-dimensional Euclidean space. Although the MDS technique has been widely applied in the medical, biological, and social sciences, very few studies have utilized this approach in earthquake seismology. Of these few, Dzwinel *et al.* (2005) used MDS to investigate earthquake patterns, whereas Yuen *et al.* (2009) used it in earthquake forecasting.

This paper develops a set of benchmarks and metrics that can help to assess and rank rupture models quantitatively. In doing so, we evaluate how different data sets and techniques constrain a rupture model. Our analysis is done in two steps. First, we test the performance of MDS on two-dimensional random fields generated using a von Kármán autocorrelation function, parametrized with different correlation lengths and/or Hurst parameters. In the second step, we conduct case studies on six inverted slip models from an SIV exercise and on 21 published slip models of the 2011 Mw 9.0 Tohoku earthquake. The SIV models were obtained using an identical data set and essentially identical source geometry but with different inversion techniques, and they can be compared with a known reference solution. On the other hand, the models for the Tohoku earthquake were obtained using different inversion techniques, source parametrizations, and data sets, and they include some variations in the assumed fault geometry as well. In this case, a known reference solution does not exist.

2 EMBEDDING METHOD

In this section, we present the embedding method based on MDS for comparing a set of 2D random fields. The essence of this

technique is to reduce the spatial variability of random fields, and their corresponding differences, to points in a lower dimensional space. A key step of this technique is selecting appropriate metrics that are sensitive to various spatial properties of the random field.

2.1 Metrics

Dissimilarity can be loosely defined as a quantitative measure of how close two sets of variables (random fields, denoted A , B , C) are. A dissimilarity metric needs to fulfil three requirements, consisting of reflectivity ($d(A, B) = 0$ if and only if $A = B$), symmetry ($d(A, B) = d(B, A)$), and triangle inequality ($d(A, C) \leq d(A, B) + d(B, C)$). In this study, we consider normalized squared and grey-scale metrics owing to their sensitivity to the variability in features' locations, intensities, and shapes. These two metrics detect complementary information. The normalized squared metric captures the magnitude of differences between two objects through point-by-point differences regardless of the position or intensity. It is computationally efficient and therefore commonly used for comparing images. The grey-scale metric, on the other hand, is more complicated. It requires transformation of the image to different intensity levels. It is computed based on the distance from grid points to sets of features with various intensities, instead of point-to-point distances used in the normalized squared metric. Hence, the grey-scale metric tends to detect differences in shape, as well as features with similar intensities. Despite its complexity compared with the normalized squared metric, the grey-scale metric has the advantage of gaining additional information on the spatial variability of 2D random fields. Wilson *et al.* (1997) presented particular examples for which these two metrics are able to identify specific features based on a 'letters' image. They found that the grey-scale metric is sensitive to differences such as removing the dot on the letter 'i', while the normalized squared metric is more sensitive to image translation.

For two random fields (or slip models) A and B , the normalized squared metric is defined as the square of the difference of the two random fields divided by the mean of their individual squared values. It is expressed as a percentage (Kragh & Christie 2002):

$$d_1(A, B) = 100 \frac{\sum_x [A(x) - B(x)]^2}{(\sum_x [A(x)]^2 + \sum_x [B(x)]^2)/2} \\ = 200 \frac{\sum_x [A(x) - B(x)]^2}{\sum_x [A(x)]^2 + \sum_x [B(x)]^2}, \quad (1)$$

where x denotes the grid-points on the rupture surface. This metric consists of point-by-point matching.

The grey-scale metric (Wilson *et al.* 1997), on the other hand, is an extension of the binary Baddeley metric (Baddeley 1992) and is defined, for $1 \leq p \leq \infty$, as

$$d_2(A, B) = \left\{ \frac{1}{NG} \sum_x \sum_g |\Delta[(x, g), \Gamma_A] - \Delta[(x, g), \Gamma_B]|^p \right\}^{1/p}, \quad (2)$$

where G presents the number of chosen grey levels, g , N is the number of elements, (x, g) is a point in set S (rupture surface) \times G (grey level), Γ_A and Γ_B respectively denote the subgraphs of random fields A and B , which give a set representation of the fields in different grey-scale levels, and $\Delta[(x, g), \Gamma_A]$ is a distance function that represents the shortest distance between a point $(x, g) \in S \times G$ and the subgraph of A .

In this study, we choose the intensity levels following Mai *et al.* (2005) who defined the slip heterogeneity based on a large set of finite-fault rupture models. They found that earthquake rupture tends to start close to a region they defined as a large-slip area ($\frac{1}{3}U_{\max} < U < \frac{2}{3}U_{\max}$, where U is the slip value and U_{\max} is the maximum slip). It then needs to encounter a very-large-slip area ($U \geq \frac{2}{3}U_{\max}$) within half of the rupture length to grow into a large earthquake. These characteristics of a rupture are found to be consistent with the energy balance in the dynamic rupture process. Hence, based on these findings, we consider three colour levels consisting of very-large-slip ($U \geq \frac{2}{3}U_{\max}$), large-slip ($\frac{1}{3}U_{\max} < U < \frac{2}{3}U_{\max}$), and moderate to low-slip ($U \leq \frac{1}{3}U_{\max}$) areas.

The metric d_1 is expressed as a percentage and its value is not limited within the range of 0 per cent (best similarity) to 100 per cent. The theoretical maximum is 200 per cent, in which one of the random fields contains only zeros. Hence, this metric can attain any value between 0 per cent and 200 per cent. However, d_2 is within the range of $(0, \infty)$ with 0 indicating the best similarity. To scale the two metrics similarly, we convert d_2 into a percentage as follows:

$$d_2(A, B) = 200 \frac{\left\{ \sum |\Delta[(x, g), \Gamma_A] - \Delta[(x, g), \Gamma_B]|^p \right\}^{1/p}}{\left\{ \sum \{|\Delta[(x, g), \Gamma_A]|^p\}^{1/p} + \left\{ \sum \{|\Delta[(x, g), \Gamma_B]|^p\}^{1/p} \right\}^{1/p} \right\}}. \quad (3)$$

2.2 Classical multidimensional scaling

In classical MDS, the purpose is to generate an m -dimensional configuration of n points in Euclidean space based on the (dis)similarity of objects under investigation (2D fields in our case). Accordingly, we can then visualize and examine point configurations in a lower-dimensional representation that best preserves the distances (dissimilarities). The general procedure is as follows (Borg & Groenen 2005):

- (i) Start with a matrix of metric distances, D , with elements d_{ij} containing pairwise-computed dissimilarity values between all random fields (see eqs 1 and 3).
- (ii) Construct a matrix, B , from double centring matrix D , which consists of subtracting the row and column means of a matrix from its elements and adding the grand mean,

$$b_{ij} = -\frac{1}{2} [d_{ij}^2 - d_{..}^2 - d_{..}^2 + d_{..}^2]. \quad (4)$$

The double centring is particularly important to make sure that the matrix is symmetric. It can also be obtained as follows:

$$B = -\frac{1}{2} HDH, \quad (5)$$

where $H = I - \frac{1}{n}\mathbf{1}\mathbf{1}'$, I represents the identity matrix, and $\mathbf{1}$ is a vector with the value of unity in each of its cells.

(iii) Apply Singular Value Decomposition (SVD) to the symmetric matrix, B , by eigendecomposition of B into $V\Lambda V^T$, where V is a matrix containing the eigenvectors of B and Λ is a diagonal matrix whose diagonal elements, $[\lambda_1, \dots, \lambda_n]$, represent the corresponding eigenvalues.

(iv) The coordinates of n points in m -dimensional Euclidean space are then given by $y_{ij} = V_{ij}\lambda_j^{-\frac{1}{2}}$ $i = 1, \dots, n; j = 1, \dots, m$.

The Y coordinates are constructed such that each column sums to zero (i.e. the origin of configuration Y coincides with the centroid). The construction is invariant under rotations and reflections. If a

reference model exists or can be defined, this can be used to define the origin. In this case, the point coordinates become

$$Y_s = (I - \mathbf{1}w)Y = P_w Y \quad (6)$$

in which Y_s are the new coordinates and w controls the position of the origin. For instance, $w = [0, \dots, 0, 1, 0, \dots, 0]$ defines the origin at the position of 1. Likewise, selecting the centroid as the origin can be obtained using $w = [1/n, \dots, 1/n]$. Therefore, instead of finding the SVD of B , it is computed for the matrix, B_s , defined as

$$B_s = P_w B P_w' \quad (7)$$

The dimension of the space of the derived coordinates (Y or Y_s) is chosen by selecting the first m eigenvalues. Then, the spatial patterns can be analyzed using the resulting point-clouds centred either at the centroid (mean model) or at any chosen reference model in m -dimensional space. The separation of points with respect to each other maps their percentage dissimilarity, because of the chosen normalization in eqs (1) and (3). The point-cloud therefore clusters similar random fields based on any chosen metrics. The eigenvalues typically help in determining the number of dimensions, m , that are necessary to represent the dissimilarity matrix accurately. The sum of eigenvalues, λ_j , is the total variance in the dissimilarity matrix. Hence, individual eigenvalues expressed as a proportion of the sum of the eigenvalues yield the proportion of variance explained by each axis. The purpose is to select enough dimensions to appropriately capture the data. However, for practical reasons, m is typically restricted to $m = 2$ or $m = 3$. It would be possible to use $m = 4$ by making a 3D plot and colouring the points according to their percentage value. Hair *et al.* (2010) suggest a proportion of 60 per cent as the minimum acceptable level accounted for by the approximated representation. We note that the actual physical meaning of each dimension, m , in an MDS configuration needs to be assessed based on the physical field under consideration and the chosen dissimilarity metric.

3 COMPARISON OF 2D RANDOM FIELDS

To test the performance of the MDS technique for comparing rupture models, we use 2D random fields generated using the von Kármán autocorrelation function (Goff & Jordan 1988) belonging to the Matérn family of correlation functions (Guttorp & Gneiting 2006) and defined as

$$C(r) = G(r)/G(0) \quad \text{with} \quad G(r) = r^H K_H(r) \quad (8)$$

in which $G(r)$ is the covariance, r is the distance, H is the Hurst exponent, K_H is the modified Bessel function of the second kind of order, H . The choice of this autocorrelation function is motivated by its flexibility, as it includes a range of autocorrelation functions (e.g. the exponential and Gaussian functions). Therefore, it is well suited for a variety of applications including the earthquake rupture process. It is also a commonly used class of autocorrelation functions in the spatial statistics community. In addition, by analyzing 44 published finite-source rupture models of 24 different earthquakes, Mai & Beroza (2002) found that the von Kármán autocorrelation function best describes the spatial characteristics of the rupture models.

An autocorrelation function can be converted into its power spectral density in the wave number domain and vice versa through

Table 1. Similarity scale.

Case	Excellent	Good	Fair	Poor
d_1^* (per cent)	<5	5–20	20–40	>40
d_2^\dagger (per cent)	<5	5–20	20–40	>40

* d_1 , normalized squared metric.

$\dagger d_2$, grey-scale metric.

Fourier transforms. Then, the corresponding spectral representation can be written as

$$P(k) = \frac{C^2}{(1 + k^2)^{H+1}} \quad (9)$$

where k denotes the wavenumber and C is the correlation length for an isotropic random field. Seed values are used to control the phase spectrum of the randomized von Kármán autocorrelation function, and inverse Fourier transform is applied to obtain the random field distribution in the spatial domain. The seed value therefore determines the spatial locations of high- and low-slip values as well as the large-scale characteristics of the random field. On the other hand, the small-scale details of the generated random fields are sensitive to the choice of H and C . Small values of H , for instance, generate highly heterogeneous 2D fields, while values of H close to unity result in smoother distributions. H is typically within the range $0 < H < 1$, although in some examples we consider a value up to 1.5 for the purpose of extended sensitivity tests. The correlation length, C , on the other hand, scales with the source dimension. Fig. A1 presents random field realizations for various correlation lengths and Hurst parameters.

We examine the MDS configuration and its sensitivity to different parametrizations for six random fields generated on a pre-defined plane, including: (a) variable H , (b) variable C (4 to 19 km in steps of 3 km), (c) variable H and C , and (d) variable H , C , and seed-number. The motivation of this test is to define and benchmark similarity scales using the normalized squared metric, d_1 and the grey-scale metric, d_2 (see eqs 1 and 3). We define four categories or levels of similarity between the 2D random fields: ‘excellent’, ‘good’, ‘fair’, and ‘poor’, adopting quantitative comparisons that are similar to those developed by

Kristeková *et al.* (2009) for seismic signals. In Table 1, we define the values of d_1 and d_2 for each category. ‘Excellent’, for instance, indicates that two random fields have a less than 5 per cent dissimilarity, while ‘poor’ corresponds to a greater than 40 per cent dissimilarity.

3.1 MDS configuration

Figs 1 and 2 show the MDS configuration using the normalized squared and grey-scale metrics, respectively. For the normalized squared metric, each 2D representation accounts for more than 90 per cent of the variations. Hence, the choice of dimension $m = 2$ is sufficient to capture the dissimilarity between random fields. The point-cloud follows a similar pattern when we vary only the Hurst parameter or correlation length. It aligns with the axis corresponding to the largest eigenvalues (Dimension 1) and is centred at the mean model (Model 7). Fig. 1(a) presents the case for varying correlation length. Among the six random fields, Fields 1 and 6 are most dissimilar to each other, but yet with a dissimilarity (d_1) of about 3 per cent. These two random fields consist of the minimum ($C = 4$ km) and maximum ($C = 19$ km) spectral parameters, respectively. They also are most dissimilar with respect to the mean model, with d_1 of 1.5 per cent. However, Model 3 with correlation length $C = 10$ km is closest to the mean model. Fig. 1(a) also shows that for larger correlation lengths, the variability between two random fields with neighbouring C -values decreases.

As we simultaneously change the two spectral parameters (correlation length and Hurst parameter), the six points no longer lie on a line. This can be understood from the fact that, by adding an additional source of variability, we increase the dimension of the dissimilarity representation. We also find that the points are located inside a circle centred at the mean model and with a radius of 5 per cent if identical seed values are used and areas of high (low) values of the field occur at the same spatial location (Figs 1a and b). In this case, we find that the dissimilarity of each random field with respect to the mean model is less than 5 per cent. This value falls into the ‘excellent’ similarity level (see Table 1).

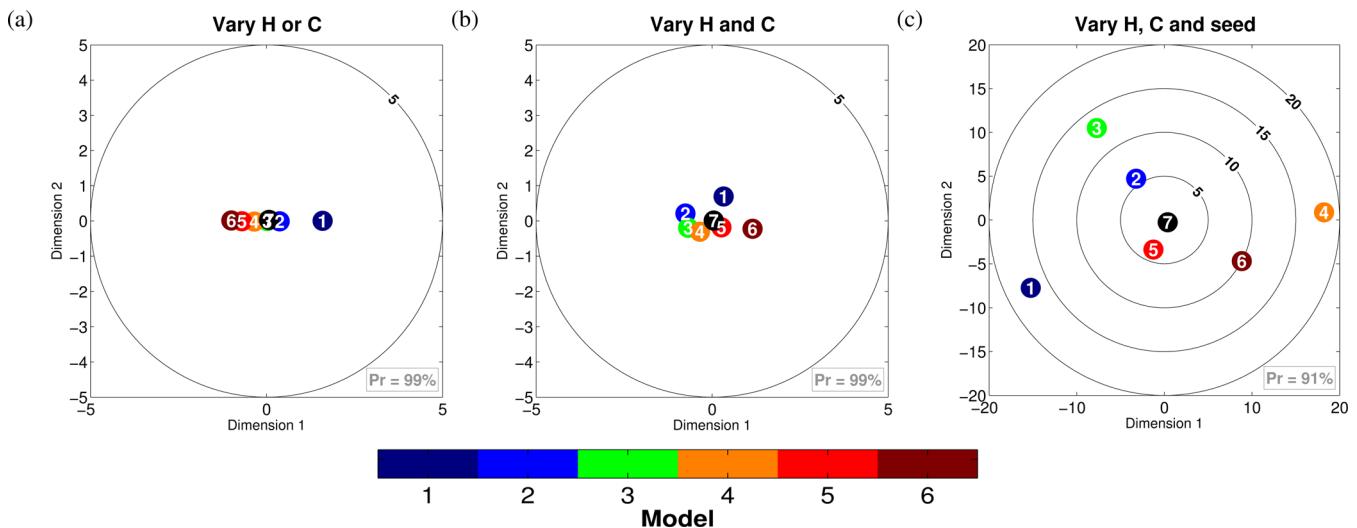


Figure 1. MDS configuration of dissimilarity between six random fields (shown in Fig. A1) using the normalized squared metric. (a) Variable Hurst parameter or correlation length; (b) variable Hurst parameter and correlation length; (c) variable Hurst parameter and correlation length with the features assumed to appear anywhere on the rectangular plane. Note the different axis scaling. Random field 7 corresponds to the mean model. The proportion of the dissimilarity explained by each configuration is listed in the bottom right of each figure.

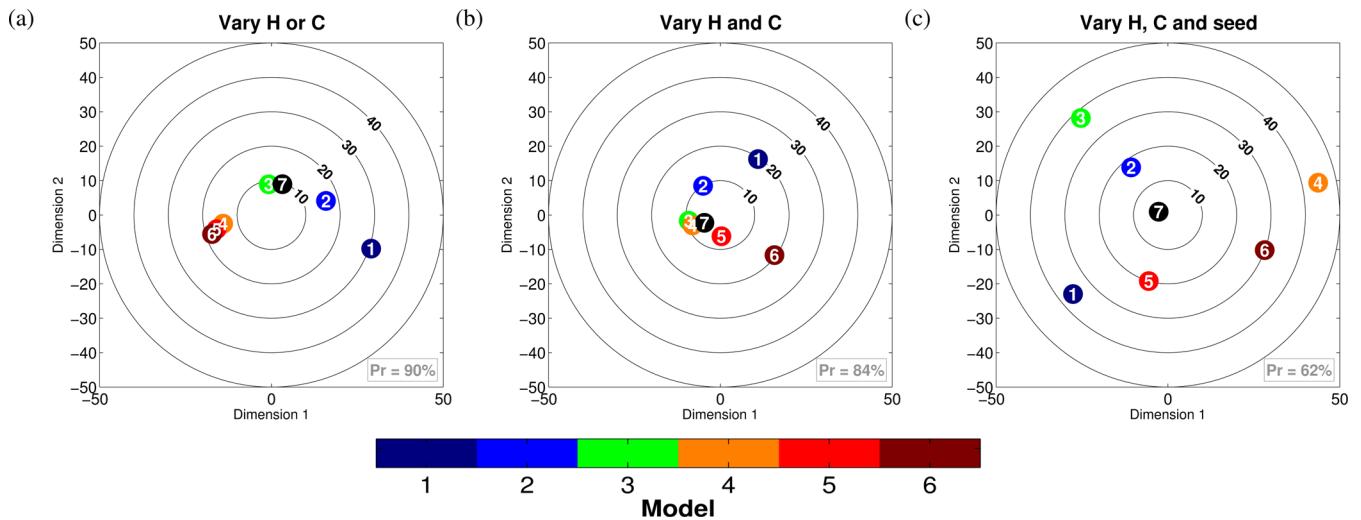


Figure 2. MDS configuration of dissimilarity between six random fields (shown in Fig. A1) using the grey-scale metric. (a) Variable Hurst parameter or correlation length; (b) variable Hurst parameter and correlation length; (c) variable Hurst parameter and correlation length with the features assumed to appear anywhere on the rectangular plane. The proportion of the dissimilarity explained by each configuration is listed in the bottom right of each figure.

However, in Fig. 1(c), we no longer constrain the seed value. The slip patch for the six random fields can thus be anywhere on the rectangular plane. This results in a widely spread point distribution that occupies a larger circle of radius 20 per cent. We define this value of d_1 (see Table 1) as the limit between similarity level ‘Good’ ($d_1 = 5 – 20$ per cent) and ‘Fair’ ($d_1 = 20 – 40$ per cent). For the normalized squared metric, the MDS configuration is strongly influenced by the patch locations, while the spread of the points shows different similarity levels. Dispersed points correspond to random fields with low similarity, whereas grouped points represent models with high similarity.

For the grey-scale metric (Fig. 2), the MDS configuration displays more distinct clusters, with larger variability along the two dimensions compared with the normalized squared metric. However, the overall point configurations are similar. Fig. 2(a) displays the MDS representation for random fields with different Hurst parameters or correlation lengths. In this case, the point configuration is characterized by a quadratic function (parabola). The axes of symmetry pass through the centroid, with the vertex at random field 3, which is closest to the mean model. The grey-scale metric is not a point-by-point distance metric; therefore the central tendency of the models is not represented by the mean model. Instead, the model with the averaged position of the structure/feature in all intensity levels defines the central tendency. The figure also shows, as in the case of the normalized squared metric, that random fields 1 and 6 are the most dissimilar with respect to each other, with dissimilarity, d_2 , of about 50 per cent. Additionally, we find that random fields 4, 5, and 6 are very close to each other with less than 5 per cent dissimilarity for both the normalized squared metric d_1 , and the grey-scale metric, d_2 . These random fields thus share common features in terms of the intensity and location of the regions of high (low) values. These values of d_1 and d_2 fall into the category ‘excellent’. Similarly to the normalized squared metric, we find in Fig. 2(c) that if we allow the slip patch to occur anywhere in the rectangular plane, the points are dispersed farther from the centroid. Random fields 1, 3, and 4, for instance, are dissimilar with respect to the centroid by more than 40 per cent. This means that the high-slip patches in these three random fields are located away from the mean patch location.

3.2 Sensitivity analysis

To capture the sensitivity of the point configurations, we generate 1000 realizations of random fields following again the four parametrizations (the variable Hurst parameter or correlation length, the variable correlation length and Hurst parameter, and finally the variable Hurst parameter, correlation length, and the seed-number controlling the patch location). We find that the point configuration follows the same pattern as for a single realization when varying only the spectral parameters.

For the normalized squared metric, Fig. 3 displays the distribution of points corresponding to the variability of the relative point location for 1000 replications. Although the variability for six random fields with different Hurst parameters (Fig. 3b) is slightly larger than for the random field with different correlation lengths (Fig. 3a), the distributions of the point configurations are similar for these two parametrizations. The variability of the point-cloud is mainly along a line, parallel to the axis of the largest eigenvalue. The variability along the second dimension (second largest eigenvalue) is very small (less than 0.5 per cent). We also notice overlapping point-clouds, implying that these models are very similar. Fig. 3(c) shows the case where both the correlation length, C , and Hurst parameter, H , are different in the six random fields. We increase C from 4 to 19 km in steps of 3 km and decrease H as follows [1.5; 1.3; 1.0; 0.8; 0.5; 0.3]. The variability of the point locations for these six random fields shows a circular shape, meaning that they are equally sensitive to the two dimensions. The variability is highest for random field 1 ($H = 1.5$; $C = 4$ km). The smallest uncertainty appears for random field 5 ($H = 0.5$; $C = 16$ km), which is closest to the mean model. These realizations also reveal that the metric, d_1 , is more sensitive to the variability of the Hurst parameter compared to that of the correlation length.

The point configuration of the 1000 realizations of the grey-scale metric (Fig. 4) also reveals an identical configuration to that of only one realization (Fig. 2c). The point-cloud for random fields with different Hurst parameters (Fig. 4b) and correlation lengths (Fig. 4a) follows a similar pattern. They both consist of a parabolic trend with the vertex around random field 3. The dissimilarities along dimensions 1 and 2 correspond to smoothness

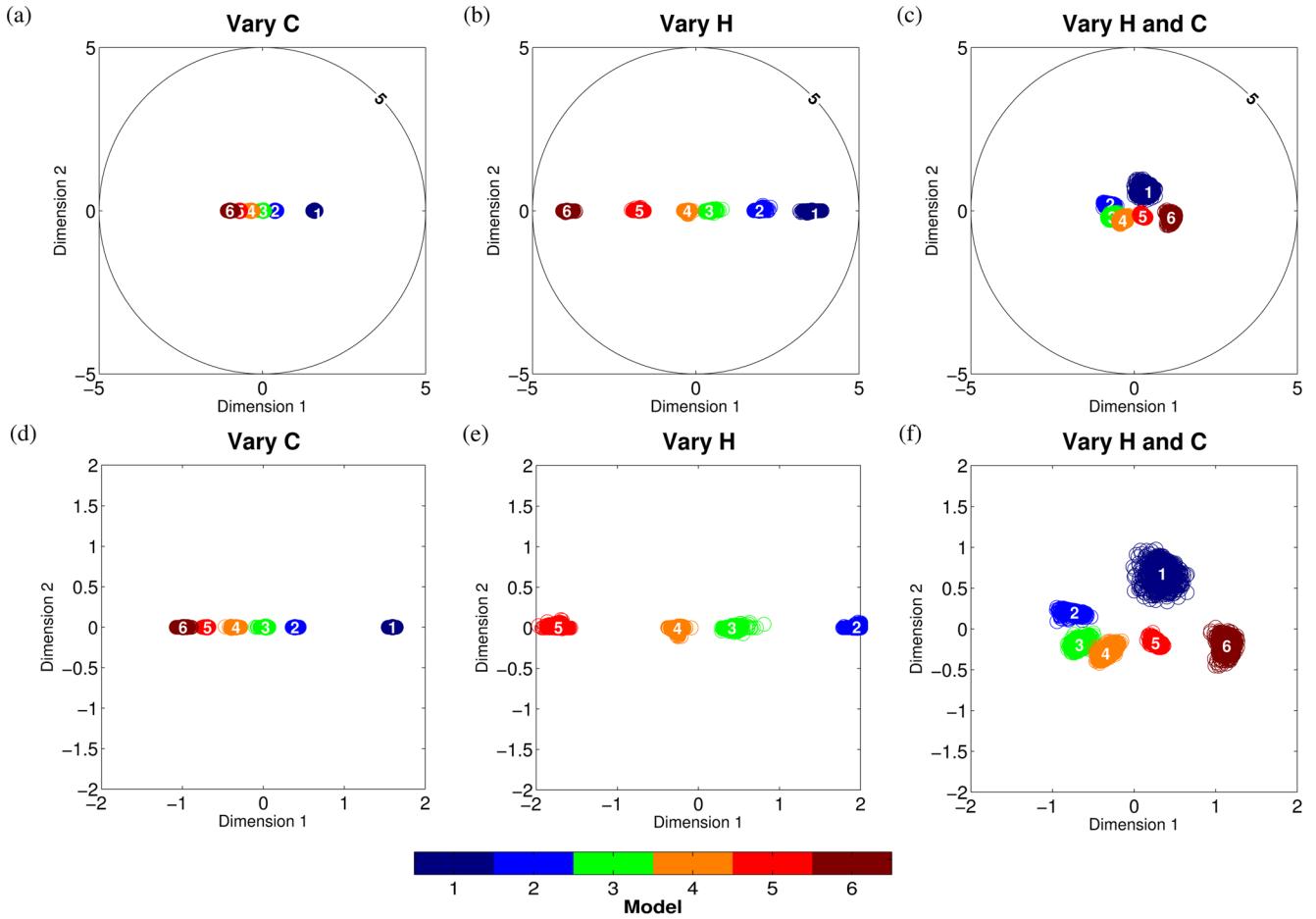


Figure 3. Sensitivity of the MDS configuration of dissimilarity between six random fields, each with 1000 realizations and using the normalized squared metric. (a) Variable Hurst parameter; (b) variable correlation length; (c) variable Hurst parameter and correlation length. The bottom panels show zoomed versions of the top panels.

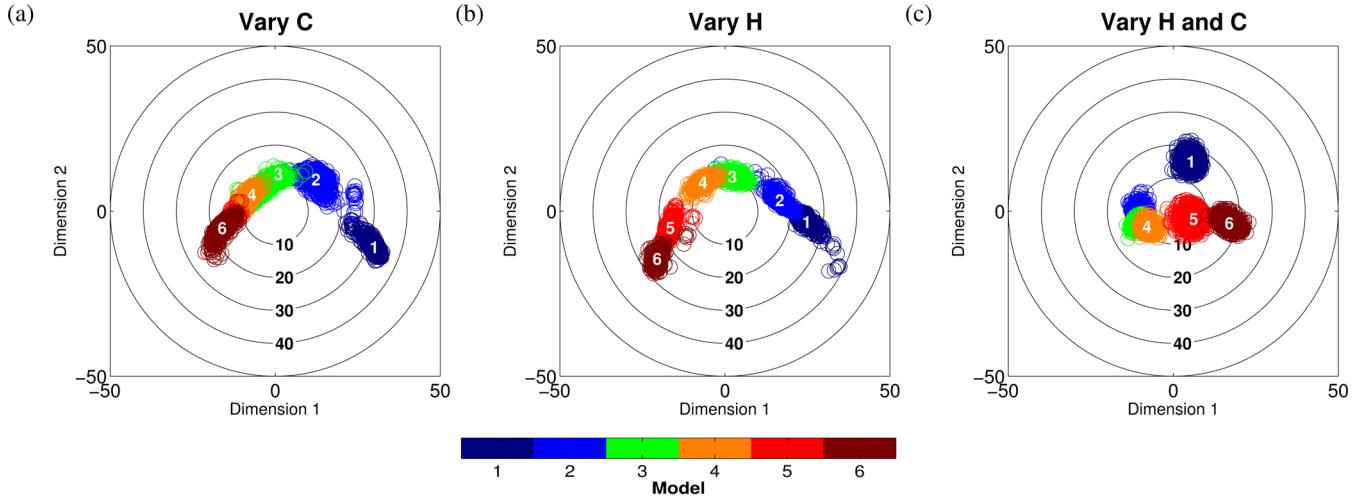


Figure 4. Sensitivity of the MDS configuration of dissimilarity between six random fields, each with 1000 realizations and using the grey-scale metric. (a) Variable Hurst parameter; (b) variable correlation length; (c) variable both Hurst parameter and correlation length.

and patch extension, respectively. As we vary both the Hurst parameter and correlation length (Fig. 4c), the point-cloud becomes circular.

When we assume that the slip patch can occur anywhere on the fault rupture plane, the point-clouds do not follow any pattern

in both the normalized squared and grey-scale metrics (Fig. 5). They spread randomly over the 2D space because, in this case, the mean model for each realization can be very different as can be the dissimilarity between each point and with respect to their corresponding centroid model. This result shows that both metrics

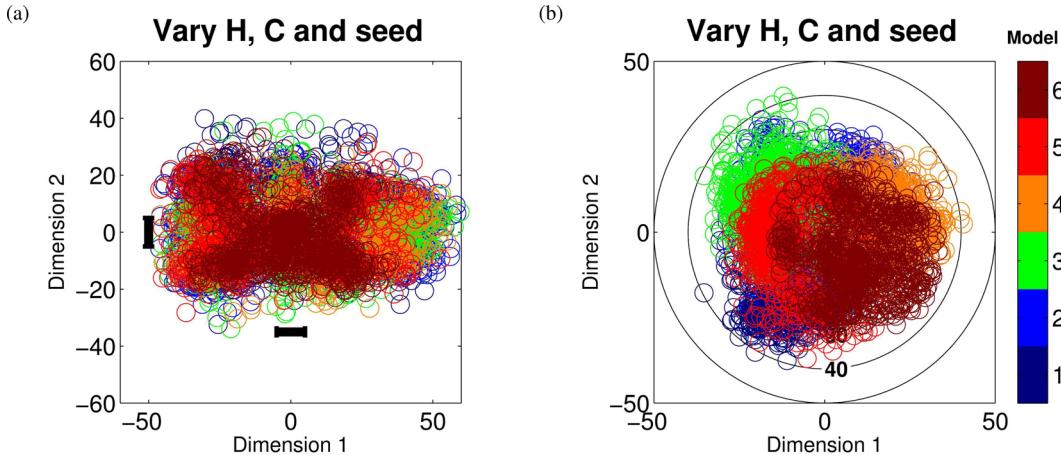


Figure 5. Sensitivity of the MDS configuration of dissimilarity between six random fields, each with 1000 realizations for the variable Hurst parameter, correlation length, and assuming that the features appear anywhere on the rectangular plane. (a) Using the normalized squared metric; the bars indicate the x and y range of 5, as used in Fig. 3. (b) Using the grey-scale metric.

are more sensitive to slip patch location variability and less sensitive to the spectral parameters.

These sensitivity tests highlight the stability of the MDS configuration and its low sensitivity to spectral parameters (the Hurst parameter and correlation length) compared to the slip location variability. These findings are also consistent with the results of Zhang *et al.* (2015) who used statistical hypothesis testing on the same data. In addition, the MDS technique can be used to rigorously rank and assess the similarity between any 2D geophysical models and eventually to rank them. In the following, we apply this technique to compare earthquake rupture models.

4 CASE STUDIES

In this section, we compare inverted slip models from the SIV exercise (accessible at <http://eqquake-rc.info/SIV/>) and for the 2011 Mw 9.0 Tohoku earthquake, respectively. These models provide cases with and without a reference model. Fig. 6 depicts the reference model and six inverted slip models from the SIV exercise. These models have approximately the same fault dimension with lengths of 30–35 km and widths of 15–20 km. However, the slip models have been generated using various grid sizes and inversion techniques. For the Tohoku earthquake, Table 2 lists 21 models obtained from the source model database (<http://eqquake-rc.info/srcmod/>; Mai & Thingbaijam 2014). These slip models were derived using different inversion techniques and different data sets, and they have different fault parametrizations (e.g. single/multiple segments, fault/subfault size).

As a preliminary analysis, we examine the variability of slip models based on a single physical value such as maximum slip and centroid location. Table 3 compares the seismic moment, maximum slip, and centroid location of the slip models from the SIV exercise. The seismic moments of Models 3 and 4 are closest to the seismic moment of the reference model. On the other hand, the maximum slips for Models 7 and 3 are closest to the reference model, while those for Models 2 and 4 are the farthest with maximum slips of 0.95 and 1.02 m, respectively. In addition, we compute a first-order estimate of the centroid location, $\mathbf{C}(c_i, c_j, c_k)$, defined as follows:

$$c_i = \frac{\sum_{x=1}^n U_i^x i_x}{\sum_{x=1}^n U_i^x}; \quad c_j = \frac{\sum_{x=1}^n U_j^x j_x}{\sum_{x=1}^n U_j^x}; \quad c_k = \frac{\sum_{x=1}^n U_k^x k_x}{\sum_{x=1}^n U_k^x}, \quad (10)$$

where U_i^x , U_j^x , and U_k^x correspond to the projection of the slip at point x of the fault along the i -, j -, and k -axis, and n represents the total number of points. This estimation is in line with McGuire (2004), who considered the temporal centroid. For the SIV slip models, we estimate this centroid location along the strike and dip directions (see Table 3). We find that the centroid location of Model 3 and the reference model are close, 0.24 km away from each other. The ranking with respect to the reference model in terms of the centroid location is as follows: Models 3, 2, 7, 1, 4, and 6. According to this first-order analysis, the best model is Model 3 in terms of centroid location and seismic moment, whereas it is Model 7 in terms of maximum slip. The worst model is Model 6 in terms of centroid location and seismic moment, whereas it is Model 2 in terms of maximum slip.

For the Tohoku earthquake, about half of the slip models have identical seismic moments ($M_0 = 3.55 \times 10^{22}$ Nm), and most of these models are from tsunami data (aside from Models 5, 9, and 21; see Table 2). The seismic moments of the remaining models vary from 4.22×10^{22} to 6.00×10^{22} Nm, with Model 13 having the largest seismic moment. In terms of maximum slip, the minimum and maximum values are observed for Models 21 and 17, respectively. The variability is large, with a mean of 46.7 m and a standard deviation of 12.6 m. We also compute the centroid location for the 21 proposed slip models for the Tohoku earthquake along latitude, longitude, and depth (see Fig. 7). The result shows that the centroid locations are in the depth range of 14–30 km. Models 9, 19 and 20 have the shallowest depths, whereas Models 8, 17, and 21 have the deepest locations. About half of the models, including Models 3, 4, 6, 7, 10, 11, 15, 16, and 18, are clustered around latitude 38°N, longitude 143.1°E, and depth 17 km. The second cluster consists of Models 1, 2, and 14 that are located to the southeast of the first group of models (latitude 37.9°N, longitude 142.8°E). The third cluster includes Models 12 and 13 located at about the same latitude and longitude as the second cluster, although the centroid location is much deeper. The remaining models do not belong to any group, as each corresponding location is isolated. According to this analysis, Models 13, 17, and 21 have extreme values in most of the cases.

These comparisons provide preliminary insight into the rupture model variability. However, they are limited, as they do not allow for comparison of the spatial distributions of slip. Thus, we use the MDS technique to compare the spatial variability of the slip models.

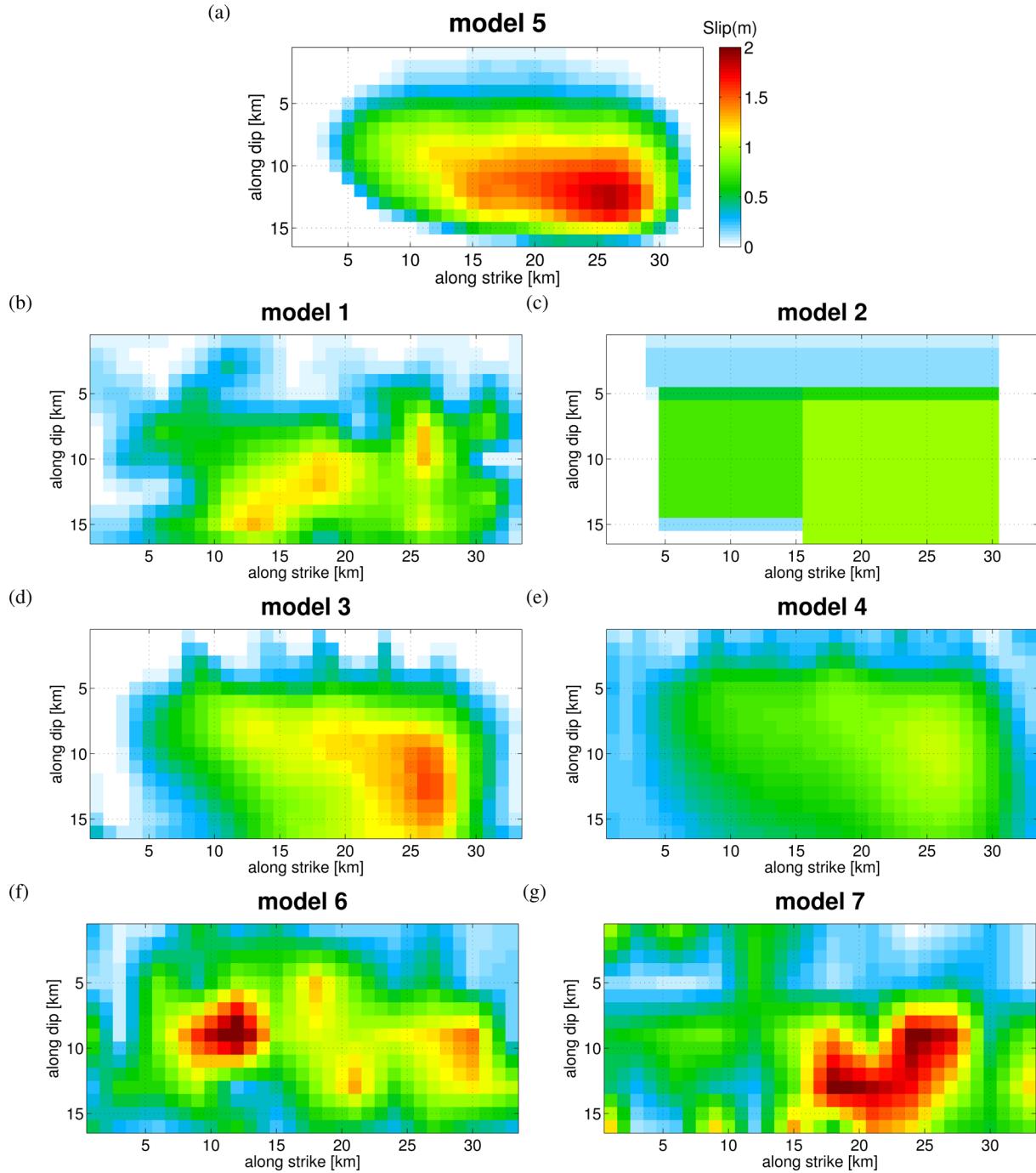


Figure 6. Selection of rupture models from the SIV exercise. (a) Reference model and (b–g) six inverted slip models.

To facilitate the slip-model comparison, we reconfigure the models such that they all have the same grid spacing. We adopt a grid spacing of 1 km and 5 km for the SIV and Tohoku models, respectively. Since the fault geometries for the SIV models are approximately identical, we do not require any additional model transformations. However, for the Tohoku slip models, we additionally adopt a single-plane representation and consider an average strike, dip, and depth of the top-edge of the fault from the surface (*htop*).

Before applying the MDS, we also need to calculate the dissimilarity between each pair of models using eqs (1) and (3) for the normalized squared metric and the grey-scale metric, respectively. For the normalized squared metric, we directly use the reconfig-

ured models. However, for the grey-scale metric, we transform the reconfigured slip models such that they consist of only three areas (very large, large, and low slip).

4.1 Source inversion validation models

Fig. 8 displays the point-cloud obtained from application of MDS to the slip models for the SIV exercise, in which Model 5 corresponds to the reference model. We examine two cases in which the point-clouds are centred at the centroid of the model ensemble and at the reference model. For each case, we use both the normalized squared

Table 2. 2011 Tohoku earthquake rupture models used in this study.

Data	Length (km)	Width (km)	Maximum slip (m)	Seismic moment (Nm)	Numbering	Author
Teleseismic	500	200	41.18	5.01×10^{22}	1	Shao <i>et al.</i> (2011)
Teleseismic	500	200	60.11	4.84×10^{22}	2	Shao <i>et al.</i> (2011)
Teleseismic	475	200	56.76	5.01×10^{22}	3	Shao <i>et al.</i> (2011)
Teleseismic	475	200	62.04	5.01×10^{22}	4	Shao <i>et al.</i> (2011)
Teleseismic	625	260	34.13	4.22×10^{22}	8	Hayes (2011)
Teleseismic	380	200	57.10	3.55×10^{22}	9	Lay <i>et al.</i> (2011)
Teleseismic	500	200	51.32	5.75×10^{22}	10	Yagi & Fukahata (2011)
Teleseismic	525	240	31.09	3.55×10^{22}	12	Wei & Sladen (2011)*
Teleseismic	445	240	30.72	3.55×10^{22}	21	Ide <i>et al.</i> (2011)
Teleseismic+GPS	625	280	30.95	6.00×10^{22}	13	Wei <i>et al.</i> (2011)†
Teleseismic+GPS	600	210	41.02	3.55×10^{22}	5	Ammon <i>et al.</i> (2011)
Teleseismic+Tsunami	340	200	67.06	3.55×10^{22}	11	Yamazaki <i>et al.</i> (2011)
Tsunami	500	200	48.57	3.55×10^{22}	6	Fujii <i>et al.</i> (2011)
Tsunami	500	200	39.11	3.55×10^{22}	7	Fujii <i>et al.</i> (2011)
Tsunami	550	200	35.09	3.55×10^{22}	18	Satake <i>et al.</i> (2013)
Tsunami	550	200	38.14	3.55×10^{22}	19	Satake <i>et al.</i> (2013)
Tsunami	550	200	44.85	3.55×10^{22}	20	Satake <i>et al.</i> (2013)
Tsunami+GPS	450	200	44.37	3.55×10^{22}	15	Gusman <i>et al.</i> (2012)
Tsunami+GPS	450	200	42.56	3.55×10^{22}	16	Gusman <i>et al.</i> (2012)
GPS+GM	525	260	48.31	5.50×10^{22}	14	Wei <i>et al.</i> (2012)
GPS+Teleseismic+Tsunami	420	240	75.72	5.92×10^{22}	17	Yue & Lay (2013)

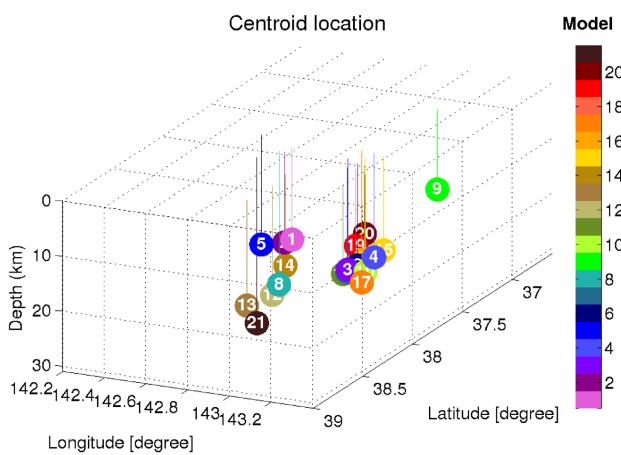
*Wei, S., Sladen, A., 2011. Preliminary Result 3/11/2011 (Mw 9.0), Tohoku-oki, Japan (accessible at http://www.tectonics.caltech.edu/slip_history/2011_tohoku-oki-tele/index.html).

†Wei, S., Sladen, A., and the ARIA group, 2011. Updated Result 3/11/2011 (Mw 9.0), Tohoku-oki, Japan (accessible at http://www.tectonics.caltech.edu/slip_history/2011_taiheiyo-oki/index.html).

Table 3. Physical description of the different slip models from the SIV exercise.

Case	Maximum slip (m)	Centroid location (km)		Seismic moment (Nm)
		Along strike	Along dip	
Model 1	1.29	18.5377	10.7413	8.037×10^{18}
Model 2	0.91	18.5950	9.89439	8.338×10^{18}
Model 3	1.54	19.2123	10.3017	9.515×10^{18}
Model 4	1.02	18.4979	9.21181	9.301×10^{18}
Model 5*	1.85	19.3622	10.1033	9.852×10^{18}
Model 6	2.35	17.6148	9.18940	1.185×10^{19}
Model 7	2.05	18.5438	10.1822	1.161×10^{19}

*Model 5, reference model.

**Figure 7.** Centroid slip locations for 21 rupture models of the 2011 Mw 9 Tohoku earthquake.

and grey-scale metrics. All configurations reveal that Models 2 and 4 are very similar to each other (less than 5 per cent dissimilarity) and cluster in one group, as their slip patches occupy roughly the same area. Model 6, on the other hand, contains inconsistent high slip patches. It is hence isolated from the other slip models.

Fig. 8 also reveals some discrepancies in the nearest neighbour models. The three nearest neighbours in the normalized squared metric (Fig. 8a) to Model 5 are Models 3, 4, and 2. However, when we centre the configuration at the reference (Fig. 8c), Models 3, 2, and 7 are closest to the centre. This is expected because there are multiple sources of spatial pattern dissimilarity (e.g. intensity, various patch locations, patch extension, and shape) for the SIV slip models. Therefore, the choice of a two-dimensional representation may no longer be sufficient to represent the model dissimilarity fully. For the normalized squared metric, the dissimilarity is less than 20 per cent with respect to the centroid for all models. Models 3 and 5 are closest to the centroid with d_1 less than 5 per cent. However, in comparison with the reference, Models 3, 7, and 2 have less than 20 per cent dissimilarity, d_1 . These models fall into the category ‘good’ in the similarity scale. The remaining models belong to the ‘fair’ similarity category, for which the dissimilarity is between 20 and 40 per cent. According to this analysis, Model 3

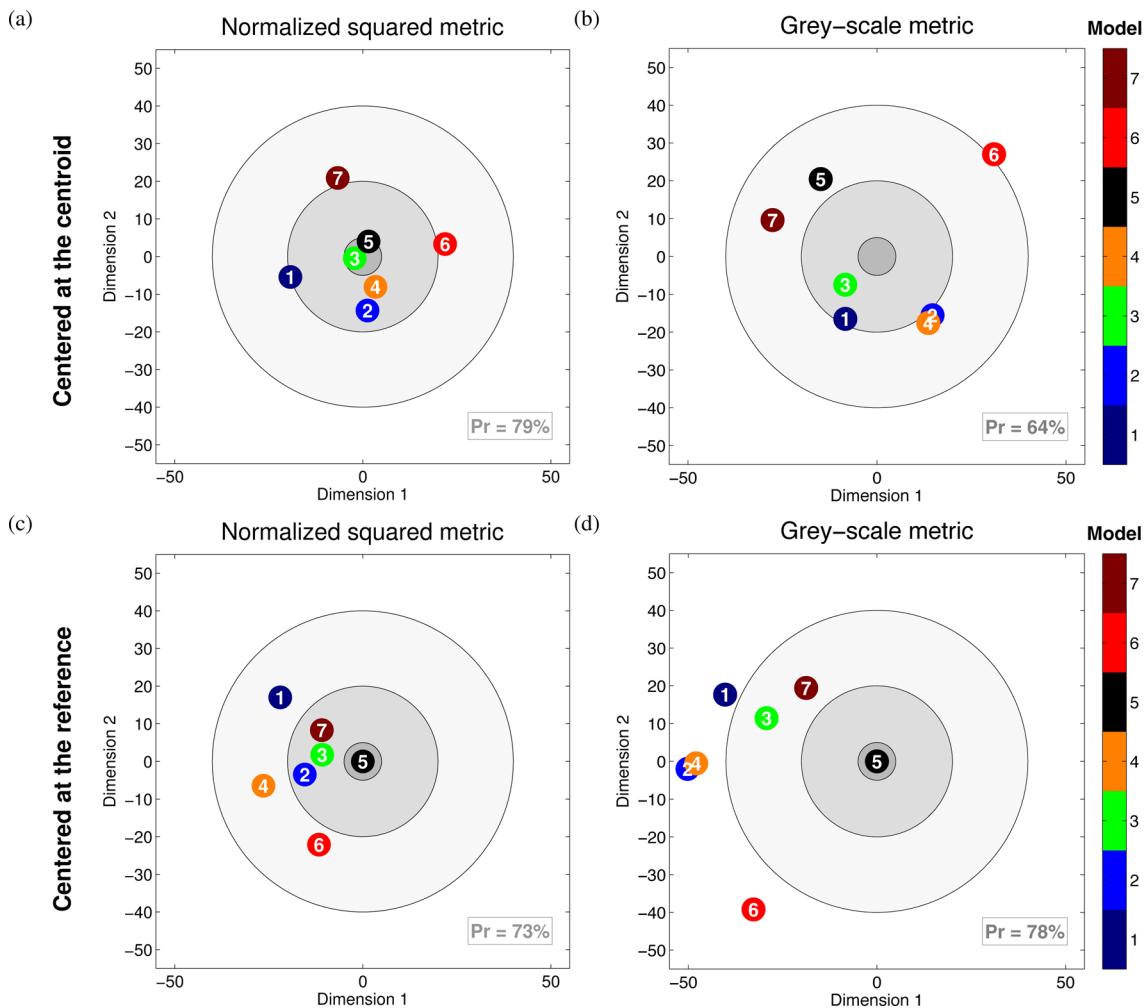


Figure 8. MDS configuration of dissimilarity between slip models in the SIV exercise. Centred at the central tendency of the model ensemble using (a) the normalized squared metric and (b) the grey-scale metric. (c) and (d) configuration with respect to reference model (Model 5). Circles limit different levels in the similarity scale, excellent (dark grey), good (grey), fair (light grey), and poor (outside the circles). The proportion of the dissimilarity explained by each configuration is quantified in the bottom right of each figure.

is the best solution to the target/reference Model 5 in the inversion exercise.

The large variability of the point-clouds for the two cases is not surprising when we use the grey-scale metric. The first case, in which the point-cloud is centred at the central tendency of the model ensemble, reveals three main clusters, Models 7 and 5, Models 3 and 1, Models 2 and 4, and one individual model, Model 6. These clusters consist of slip models that share the same intensity of slip at the same location. All these models have between 20 and 40 per cent dissimilarity with respect to the centroid, aside from Model 3, which has less than 20 per cent dissimilarity. For the second case, in which the point-cloud is centred at the reference model (Fig. 8d), Models 3 and 7 are ‘fairly similar’ (d_2 between 20 and 40 per cent) to the reference model in terms of location and extension of the three areas of slip (very large, large, and low slip). The other models underestimate the feature of the reference model, and hence have a grey-scale metric, d_2 , that is greater than 40 per cent (‘poor’) compared to the reference model.

Each of the two-dimensional representations of the slip-model variability for the SIV exercise explains about 70 per cent of the full dissimilarity. This value is considered acceptable as suggested by Hair *et al.* (2010). However, this level might change depending on

the context. Hence, to examine possible contributions to the dissimilarity measure when considering high dimensions, we analyze the three-dimensional representation of the point-clouds (Fig. 9). When we include the third dimension, the MDS representations explain about 90 per cent and 80 per cent of the full dissimilarity for the normalized squared and grey-scale metric, respectively. This representation therefore provides a more complete point distribution. By comparing the configurations centred at the centroid and reference model in 2D and 3D, we obtain more consistent results in the 3D point distribution, particularly for the nearest neighbours of each point. The three nearest neighbours to Model 5 of the normalized squared metric are Models 3, 2, and 7 for both configurations centred at the centroid and the reference. Clearly, the third dimension captures a rather detailed aspect of the spatial variability, compared to the first and second dimensions. In the normalized squared metric centred at the centroid, the third dimension consists of patch extensions. This analysis also reveals that Model 3 is the best model.

We find that the percentage of dissimilarity when the grey-scale metric is used appears larger than that for the normalized squared metric, because of their sensitivity to different properties of the slip models. The grey-scale metric is more sensitive to small-scale spatial variability. According to the three-dimensional representations

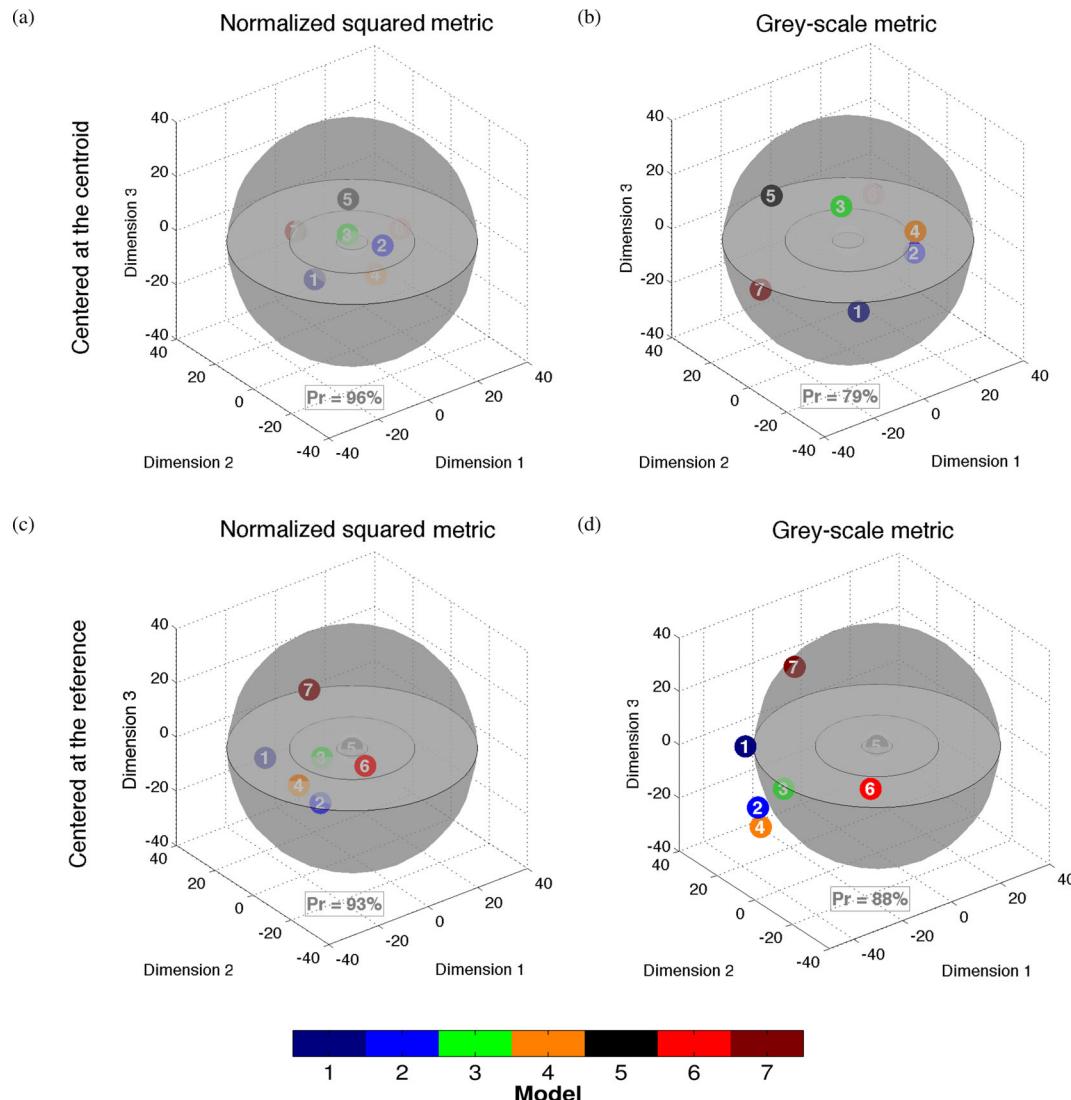


Figure 9. 3D MDS configuration of dissimilarity between slip models in the SIV exercise. Centred at the central tendency of the model ensemble using (a) the normalized squared metric and (b) the grey-scale metric. (c) and (d) configuration with respect to reference model (Model 5).

in Fig. 9 (normalized squared metric) Model 3 is the best solution to the reference model (Model 5), with between 5 and 20 per cent dissimilarity. This falls into the category ‘good’. The rest of the models are ‘fair’, with between 20 and 40 per cent dissimilarity. The rank of these models is consistent with the findings of Zhang *et al.* (2015) for squared loss functions. For the grey-scale metric, Models 3 and 7 are closest to the reference model with between 20 and 40 per cent dissimilarity. The rest of the models have greater than 40 per cent dissimilarity. We also find that Models 2, 3, and 4, which have similar spatial patterns, belong to a single cluster. This analysis illustrates the importance of a high-dimensional representation to fully capture the large variability among the models. It also reveals the ability (strength) of each metric.

4.2 2011 Tohoku earthquake models

We examine two cases of the 2011 Tohoku earthquake to compare slip models because the fault geometry of the 21 models varies significantly across the models (see Fig. 10). We consider (a) the smallest common rupture area for all models and (b) the largest fault

area that contains all models. These two cases capture the variability in terms of fault geometry and slip distribution, respectively. It is also important to note that for real earthquakes, no reference model exists. The comparison is therefore done based on a centroid model, which is adopted as the reference although it does not represent the best model.

Figs 11 and 12 display the MDS configuration of the 21 Tohoku earthquake slip models using a two-dimensional configuration. For the smallest common rupture plane, similarity appears in the overall distribution of the point-clouds of both the normalized squared and grey-scale metrics. We find that Models 15 and 16 are closest to the centroid with dissimilarity metrics, d_1 and d_2 , less than 5 per cent. Model 17, on the other hand, is farthest from the centre of the configuration, with greater than 40 per cent dissimilarity for both metrics. More than half of the slip models fall into the category ‘good’. Table 4 summarizes the similarity of each model with respect to the centroid model for these two metrics.

It is important to note that we do not know in advance the meaning of each dimension in terms of the physical parameter of the fields. In fact, it can be interpreted as a physical parameter that appears to order the models in the configuration. For the smallest

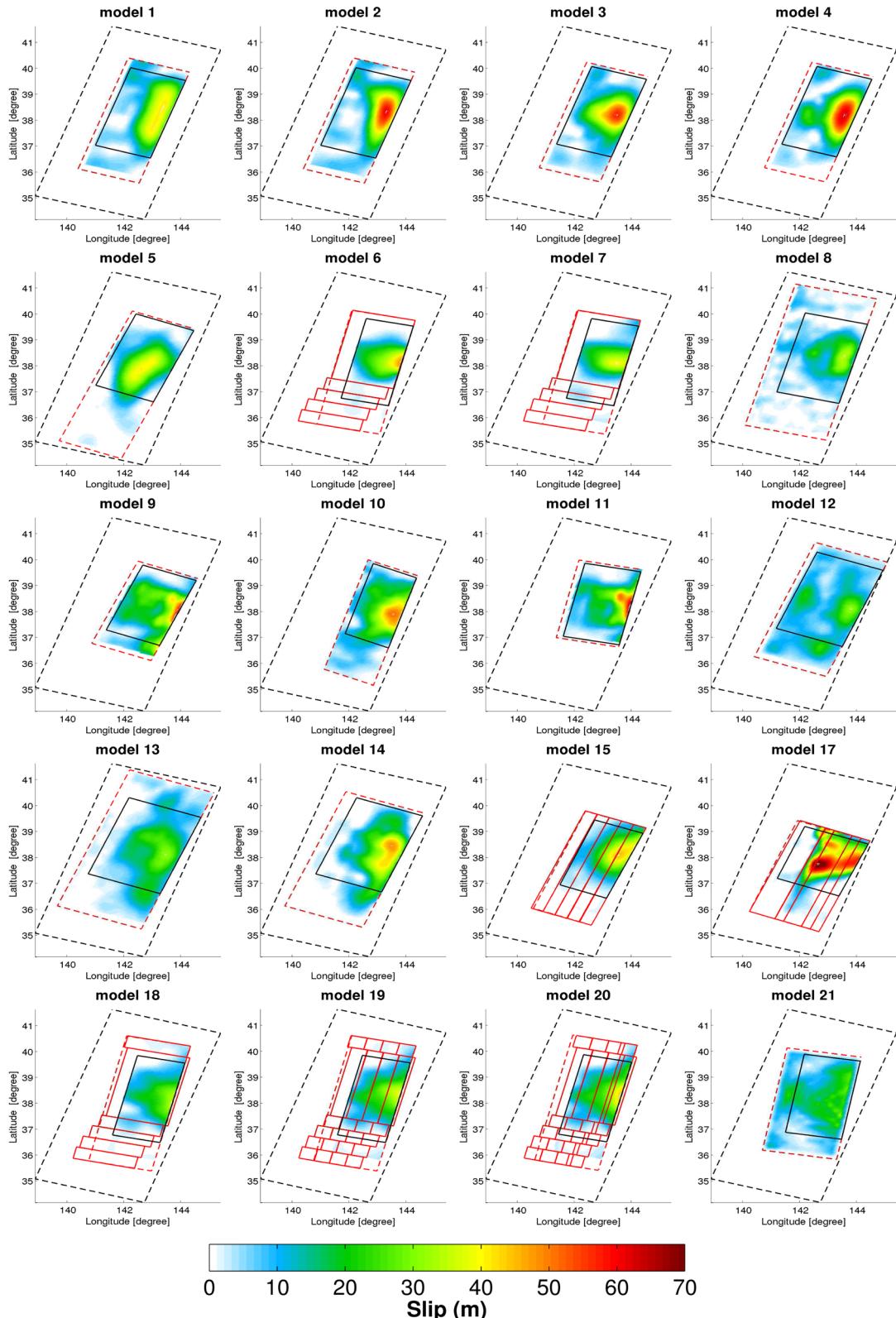


Figure 10. Twenty slip models for the 2011 Tohoku earthquake (excluding Model 16 that has a similar pattern to Model 15). The original fault geometry is indicated with red lines. The red dashed lines outline single plane representations of the corresponding model. The smallest and largest areas used in the slip-model comparison are denoted by black lines and black dashed lines, respectively.

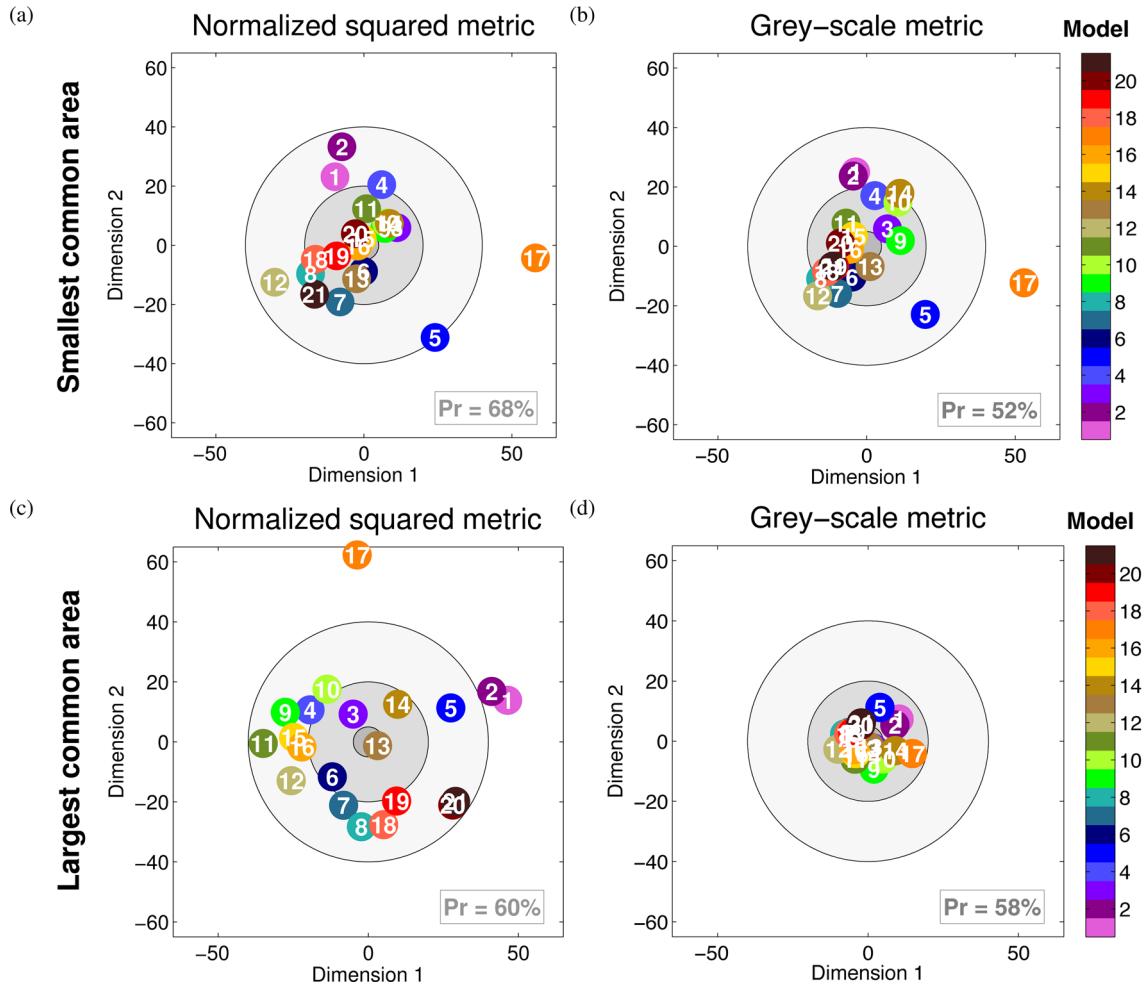


Figure 11. MDS point-cloud considering the smallest and largest common area of the 2011 Tohoku slip models and using the normalized squared and grey-scale metrics. The proportion of the dissimilarity explained in each configuration is specified in the bottom right of each figure.

common rupture plane (Figs 11 and 12), Models 17 and 12 are quite dissimilar on dimension 1, but rather similar on dimension 2. Hence, dimension 1 relates to the overall intensity (or magnitude) level of the slip. Dimension 2, on the other hand, corresponds to the extent of the slip patches. We also find that most of the models are aligned along the diagonal, aside from Models 1, 2, 5, and 17 which are separated. This diagonal alignment shows that the variability of the slip models encompasses both the intensity and the slip patch extensions.

In the Appendix (see Fig. B6), we present the 3D MDS configuration. The third dimension contributes about 9 per cent and 20 per cent of the slip-model variability in the grey-scale and normalized squared metrics, respectively. For the grey-scale metric, the alignment of Models 9 and 3, as well as 6 and 13, shows that the third dimension is related to the variability of the slip patch extension along the strike direction. However, the contribution of this component is small. For the normalized squared metric, the alignment of Models 1, 2, 19, and 7 suggests that this dimension indicates the compactness of the slip patches. It also can be interpreted as the change in the patch locations along the dip direction. Fig. B6 particularly illustrates that the 3D-visualization becomes more challenging as we have more models. Additionally, with the eigenvalues ordered from the largest to the smallest, the first dimension contributes most to the dissimilarity, followed by the second dimension and so forth. The physical interpretation of these dimensions

becomes more difficult (non-unique) as the relative contribution of the dimension becomes insignificant. We therefore restrict our analysis to two-dimensional representation for the following cases.

To validate our result from the 2011 Tohoku slip models, we compared the MDS results with the results obtained from the Spatial Prediction Comparison Test (SPCT; Hering & Genton (2011)). SPCT is a statistical test that consists of comparing loss functions between competing forecasts. It was developed for general spatial fields and applied to wind speed (Hering & Genton 2011), precipitation fields (Gilleland 2013), and earthquake slip models (Zhang *et al.* 2015). These studies extensively describe the technique. In Fig. 13, we present the SPCT result using the squared-error (SE) loss function when considering the mean model (see Appendix) as a reference model. For further details, we refer to Zhang *et al.* (2015) who note that negative values (blue) indicate that the case named in the corresponding row is the better model in terms of SE loss functions, and the location with letter ‘a’ indicates that the corresponding two models differ significantly from each other at the 5 per cent confidence level.

Fig. 13 shows that Model 17 differs significantly from all other models at the 5 per cent level. This model is thus the most dissimilar to the mean model. This model is in the category ‘poor’ for MDS. Models 1, 2, 5, 7, and 12, on the other hand, are significantly different at the 5 per cent level from more than four models. These models fall into ‘fair’ category in the MDS. The rest of the models are either

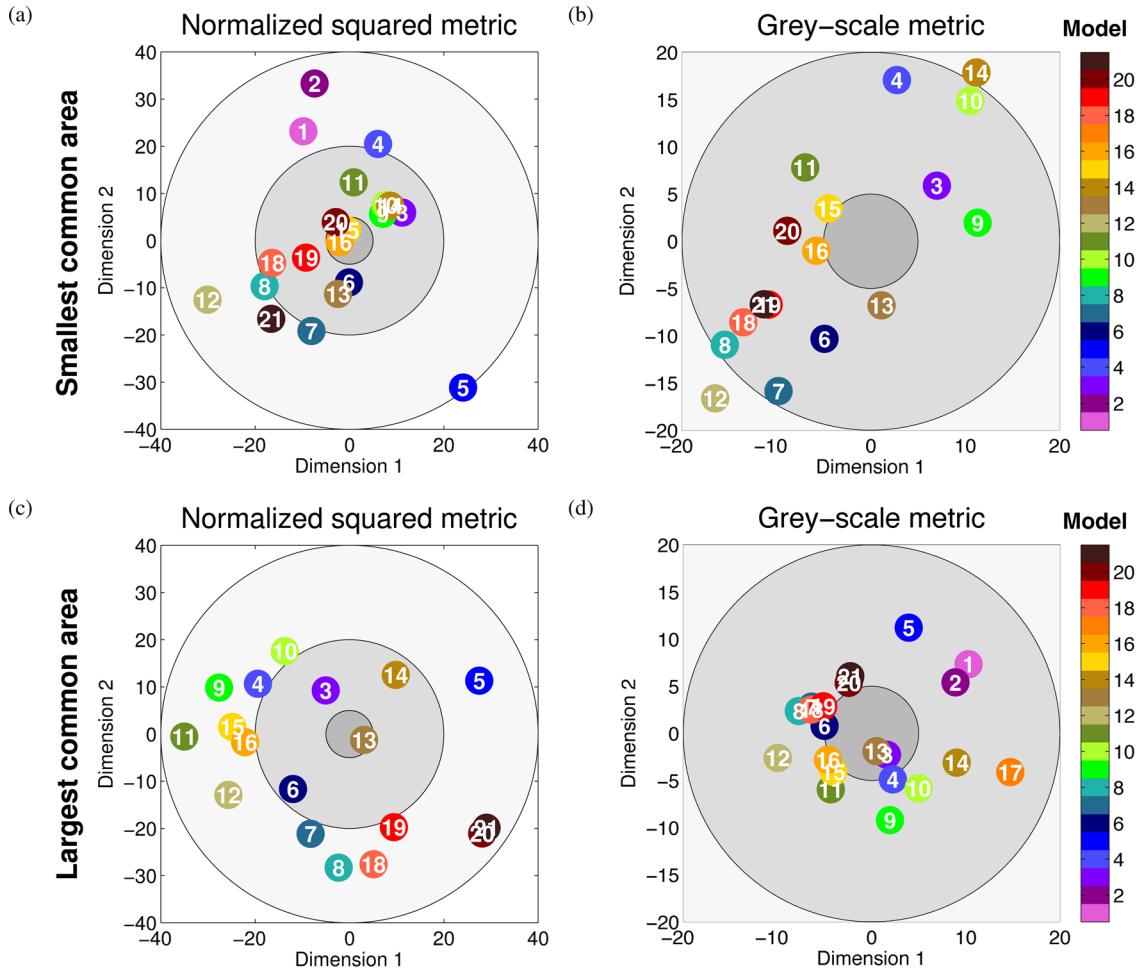


Figure 12. Zoomed version of Fig. 11. Note the different axis scaling.

Table 4. Tohoku slip-model similarity compared to the mean model (smallest common area).

Case	Excellent	Good	Fair	Poor
d_1^*	15,16,20	3,4,6,8,9,10,11,13,14,18,19	1,2,5,7,12,21	17
d_2^\dagger	15,16	3,4,6,7,8,9,10,11,13,14,18,19,20,21	1,2,5,12	17

* d_1 , normalized squared metric.

$\dagger d_2$, grey-scale metric.

‘good’, ‘excellent’, or along the boundary between ‘fair’ and ‘good’. They are significantly different from fewer than three models. Both techniques also reveal that Models 15 and 16 best represent the mean model (see Fig. B1 in the Appendix). This validation test shows that the MDS results are consistent with the SPCT method.

The patterns of the MDS configurations for the largest common rupture plane (Figs 11 and 12) are not similar between the normalized squared metric and grey-scale metric. The metrics have different sensitivities. The variability among the slip models with the normalized squared metric, for instance, is large. The points appear to be more scattered and occupy a wider area (greater than 40 per cent). This illustrates the large variability in the geometry, particularly the strike. Despite the scattered points, we still observe that Models 6, 7, 18, and 19 are clustered. These models are all obtained by inverting the tsunami data. As we unify the fault geometry (strike, dip, and htop), the slip models become more similar. With the grey-scale metric, all models appear similar, with less than

20 per cent dissimilarity with respect to the centroid. This is due to numerous zero-valued grid points over the enlarged area of the slip models (see white area between red and black dashed line in Fig. 10). We also examine the alignment of the model along each axis and find that for the normalized squared metrics, the two principle sources of variability correspond to the compactness of the patch (dimension 1) and the intensity (dimension 2). With the grey-scale metrics, on the other hand, these two dimensions correspond to the slip patch extension and the shape of the features. The summary of classification of the 21 Tohoku slip models with respect to the centroid considering the largest common rupture area is shown in Table 5.

4.2.1 MDS sensitivity

To assess the stability and accuracy of the MDS solution, we use a jackknife approach, which is a resampling method that is applied

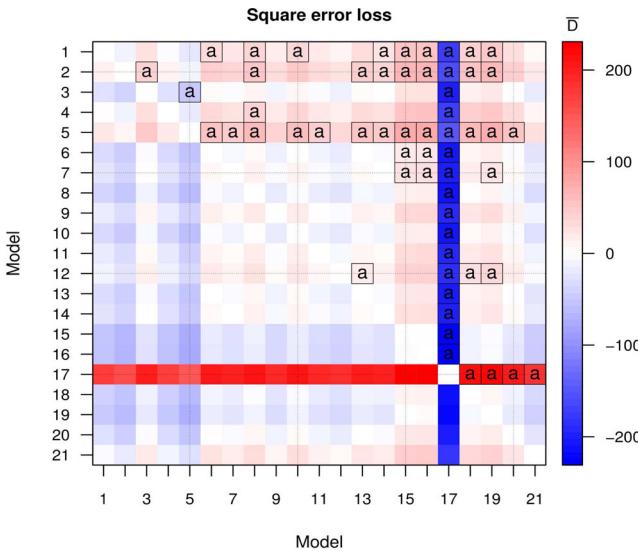


Figure 13. Mean loss differentials for the squared loss function with the hypothesis test results from the spatial prediction comparison test. Locations with letter ‘a’ indicate that the corresponding two models differ significantly from each other at the 5 per cent level. Negative values (blue) indicate that the case named in the corresponding row is the better model.

to statistical inference. It is widely used due to its simplicity and efficiency. In particular, we use the delete-one jackknife (Efron 1982) that works as follows:

- Omit one slip model and generate the dissimilarity matrix using the remaining slip models.
- Carry out MDS.
- Apply the similarity transformation (Gower & Dijksterhuis 2004) with respect to the original model.

We repeat this procedure for each slip model. Ultimately, we have $n - 1$ jackknifed coordinates for each slip model, in which n is the number of slip models. Fig. 14 depicts the variability of the point-cloud considering the largest common area of the Tohoku slip models and using the normalized squared metric. We chose this metric because it shows the largest dissimilarities among models. The jackknifed point-cloud shows similar pattern as in the original configuration. Only Model 17, which is far away from the average model, has a significant effect in changing the MDS configuration. We also find that the point-clouds of very similar models such as Models 15 and 16 as well as 6 and 7 overlap.

4.2.2 Predicted displacement comparison

The MDS technique has been applied to some slip models of the Tohoku earthquake that required processing in order to map them onto the same physical space. An alternative approach is to compute an independent and hence unbiased physical predictor quantity from each model and to use that predictor for the MDS analysis. In this section, we use MDS to compare the three components of seafloor

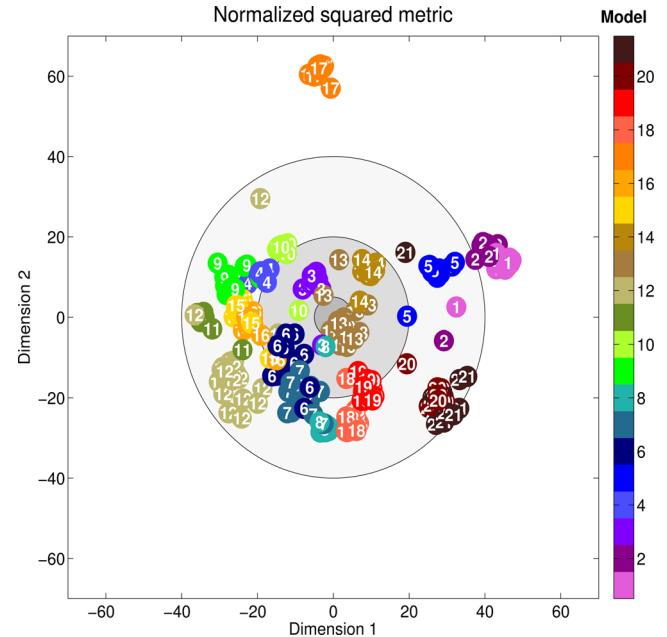


Figure 14. A jackknifed MDS point-cloud for the largest common area of the Tohoku slip models using the normalized squared metric.

displacement computed from the 21 slip models. The displacements are computed using Okada (1985) and are presented in the Appendix (Figs B2, B3, and B4). They all predict displacements over 20m and 8m on the horizontal and vertical components, respectively. However, significant variability appears in their spatial patterns that will affect the prediction of tsunami properties (Goda *et al.* 2014; Tappin *et al.* 2014).

Fig. 15 displays the MDS point-cloud of the predicted seafloor displacements using the normalized squared metric. We observe that dissimilarities among the horizontal components are less pronounced than those among the vertical. For the EW component (Fig. 15a), for instance, the dissimilarities of the predicted displacements with respect to the mean displacement are all less than 40 per cent, aside from the predicted displacements from Models 9, 19, and 20, which have greater than 40 per cent dissimilarity. For the NS component (Fig. 15b), the dissimilarity is less than 40 per cent for all models, aside from the displacement corresponding to Models 9, 11, 13, 19, and 21. However dissimilarities among the vertical displacements (Fig. 15c) are greater than 40 per cent for most of the models. Hence, more variability appears in the vertical component of the seafloor displacements that in fact largely govern the tsunami generation process. Hence, we conjecture that the predictive ability to match the tsunami observations from all these models is very different. We also identify clusters of points for the vertical component, such as points 5, 6, 7, 15, 16, 18, 19, 20, and 21; points 3, 10, 13, and 14; and points 1, 2, and 17. The remaining models are isolated. The first cluster consists mostly of predictions from slip models generated from tsunami data. Figs B4 and 15(c) also

Table 5. Tohoku slip-model similarity compared to the mean model (largest common area).

Case	Excellent	Good	Fair	Poor
d_1^*	13	3, 6, 14	4, 5, 7, 8, 9, 10, 11, 12, 15, 16, 18, 19, 20, 21	1, 2, 17
d_2^\dagger	3, 4, 6, 13, 15, 16	1, 2, 5, 7, 8, 9, 10, 11, 12, 14, 17, 18, 19, 20, 21		

* d_1 , normalized squared metric.

† d_2 , grey-scale metric.

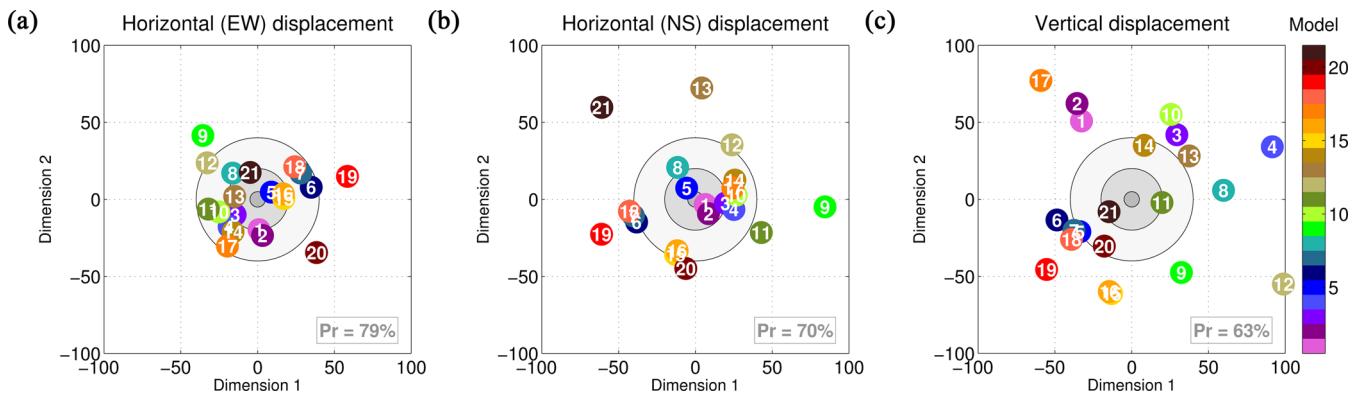


Figure 15. MDS configurations of dissimilarity between predicted seafloor displacement from the 21 Tohoku slip models in Fig. 10. (a) EW component; (b) NS component; and (c) Vertical component (see Appendix, Figs B2–B4).

suggest that the first and second axes correspond to variability in the location and in the significance of the uplifted area, respectively.

We observe that the computed seafloor displacements from Model 9 appear to be isolated for all components. This model has a compact and shallow slip patch. It is not very different from the other models in terms of the normalized squared metric. However, the 3D cloud points of the slip dissimilarity for the grey-scale metric (see Appendix, Fig. B6) show that this model is isolated and has between 20 and 40 per cent dissimilarity ('fair') compared to the central tendency of the model. The corresponding EW seafloor displacement is low (around 10 m), and the NS seafloor displacement is large (around 20 m), compared to the predictions near the trench from other models.

5 DISCUSSION

In this study, we apply MDS to investigate the characteristics of the dissimilarity between slip models. This technique augments standard residual analysis through its ability to cluster the models with common intensity and patch geometry. This ability eventually allows for ranking the models based on a reference or central tendency of the model ensemble. Additionally, we investigate the effect of various physical assumptions and data sets. For example, Fig. 11 shows that slip models from tsunami data form a single cluster. This suggests that the inverted slip models from tsunami data vary systematically from the mean, which consists of underestimation of slip values compared to the mean model. An alternative explanation is that the differences are due to the fault geometry, as they use an average strike of 192° to 194° , compared to those of 198° to 202° for slip models obtained from seismic and/or geodetic data.

We also examine the MDS results based on grey-scale and normalized squared metrics. These metrics are sensitive to different spatial characteristics of the slip. The normalized squared metric is more sensitive to the overall intensity, whereas the grey-scale metric particularly detects the patch geometry and the location of the areas of large and low slip. Therefore, models that are very similar using both metrics share common aspects in terms of patch intensity and geometry. Models with different smoothness and correlation lengths, for instance, show that they are very similar in terms of the normalized squared metric, with less than 5 per cent dissimilarity (Fig. 1a). However, the variability using the grey-scale metric is larger (Fig. 2), since the change in smoothness or correlation length particularly affects the geometry of the patches.

In this study, we compared only the slip values. However, the proposed approach can be easily extended and applied to examining

the variability of the full spatio-temporal parameters of the rupture process. Also, the results we obtained are based on two metrics relevant to capturing the spatial variability of the slip. However, additional metrics are proposed that could be utilized and tested to quantify the dissimilarity of the slip models. Hence, this study could be extended to examine alternative metrics such as correlation and warping loss functions (Gilleland 2013), which could potentially contribute to additional information on the source of the discrepancies. We could additionally mix and weight the different metrics, although the MDS point-cloud would be more complex as it would be a mixture of different properties of the spatial pattern.

There is no rigorous statistical method to evaluate the quality and reliability of a representation produced by MDS. Here, we assess the accuracy based on the variability of the configuration using a number of realizations. For the case studies, we obtain the variability from a jackknife test. An MDS configuration also provides a richer interpretation compared to a simple residual analysis, as it reveals classes of slip models sharing common features such as slip patch intensity, shape, and extension.

5.1 Insights for source inversion

As an application of MDS to a real earthquake, we compared the slip models from the 2011 Tohoku earthquake. We could, however, consider this analysis as a benchmark to assess the accuracy of any earthquake for which multiple finite-fault source models have been published. We can also use this tool to investigate possible systematic classification depending on earthquake magnitudes and tectonic regimes. Additionally, the sensitivity of data sets in resolving the rupture process can be extensively explored using this comparison tool.

We find that the grey-scale metric complements the normalized squared metric for comparing slip models. The grey-scale metric thus appears to be a useful Euclidean norm in kinematic source inversion to statistically discriminate between various proposed solutions. To the best of our knowledge, the use of this metric to examine the resolution of inverted slip models has never been described before.

5.2 Accuracy and limitation

We considered an approximate representation using a dimension-reduction technique with two or three dimensions. However, two- or three-dimensional representations may be insufficient to represent the full dissimilarity between slip models. In Fig. 1(a), for instance,

the variability comes from one source, the Hurst parameter, H , or the correlation length, C . Therefore, even one dimension suffices to represent the dissimilarity between the random fields. In this case, the interpretation of the dimension axes is straightforward: it is the C and H values. As we increase the complexity and the source of variability in the models, the dimensionality of the dissimilarity becomes larger. Therefore, the 2D representation is just an approximation to facilitate the visualization and the interpretation of the different classes, although it still constitutes a limited representation of the dissimilarity. A comparison of the slip models for the SIV exercise illustrates this limitation and shows the importance of higher dimensional-representations. Therefore, it is necessary to assess percentages of dissimilarity accounted for by each dimension. For the SIV case, for instance, the two-dimensional representation accounts for 79 per cent of the full dissimilarity. As we incorporate the third dimension, they account for 95 per cent. The physical meaning of each dimension also changes depending on the main source of variability among the model ensemble. The factors involve patch location, intensity, dimension, extension, shape, or their alternative combinations.

One limitation of our technique is that we do not have a reference model for a real earthquake. Therefore, we make comparisons with respect to the central tendency of the models. However, this mean model does not, in general, represent the best model in the sense of, for instance, its capability to fit observation data. Consequently, this approach does not rely on the best and optimal model, but it helps to identify those models that share common features and hence may be able to achieve similar predictions of data or physical quantities.

6 CONCLUSIONS

The objective of our study is to quantify both differences and similarities between rupture models. We find that the MDS technique efficiently identifies slip models sharing common spatial pattern characteristics in terms of the slip patch intensity and geometry. We also propose a similarity scale for the rupture models based on this technique. The scale allows for ranking the models with respect to the centroid or reference model. We find that the generated MDS point-clouds change depending on the choice of metric. The normalized squared metric is insensitive to the spectral parameters of the random field but very sensitive to the slip patch locations. The grey-scale metric, on the other hand, is sensitive to the patch geometry and hence to small-scale variability in the model. The case studies also reveal the importance of higher-dimensional representations for rupture models with large sources of variability. A natural extension of this work would be to assess the accuracy of inter-event rupture models and consider both spatial and temporal patterns.

ACKNOWLEDGEMENTS

We would like to thank Guangcui Feng for his help in computing seafloor displacements. We also thank the reviewers for constructive comments that helped to improve this study. The work reported here was supported by the King Abdullah University of Science and Technology (KAUST).

REFERENCES

- Ammon, C.J., Lay, T., Kanamori, H. & Cleveland, M., 2011. A rupture model of the 2011 off the Pacific coast of Tohoku earthquake, *Earth planet. Space*, **63**(7), 693–696.
- Baddeley, A.J., 1992. An error metric for binary images, *Robust Computer Vision*, Wichmann, Karlsruhe, pp. 59–78.
- Beresnev, I.A., 2003. Uncertainties in finite-fault slip inversions: to what extent to believe? (a critical review), *Bull. seism. Soc. Am.*, **93**(6), 2445–2458.
- Borg, I. & Groenen, P.J.F., 2005. *Modern Multidimensional Scaling*, 2nd edition, Springer.
- Cotton, F. & Campillo, M., 1995. Frequency domain inversion of strong motions: application to the 1992 Landers earthquake, *J. geophys. Res.*, **100**, 3961–3975.
- Dzwirniewicz, W., Yuen, D.A., Boryczko, K., Ben-Zion, Y., Yoshioka, S. & Ito, T., 2005. Cluster analysis, data-mining, multi-dimensional visualization of earthquakes over space, time and feature space, *Nonlinear Proc. Geophys.*, **12**, 117–128.
- Efron, B., 1982. *The Jackknife, the Bootstrap, and Other Resampling Plans*, Society for Industrial and Applied Mathematics, doi: <http://dx.doi.org/10.1137/1.9781611970319>.
- Feng, G. & Jónsson, S., 2012. Shortcomings of InSAR for studying megathrust earthquakes: the case of the Mw 9.0 Tohoku-Oki earthquake, *Geophys. Res. Lett.*, **39**, L10305, doi:10.1029/2012GL051628.
- Fujii, Y., Satake, K., Sakai, S., Shinohara, M. & Kanazawa, T., 2011. Tsunami source of the 2011 off the Pacific coast of Tohoku earthquake, *Earth planet. Space*, **63**(7), 815–820.
- Gilleland, E., 2013. Testing competing precipitation forecasts accurately and efficiently: the spatial prediction comparison test, *Mon. Wea. Rev.*, **141**(1), 340–355.
- Goda, K., Mai, P.M., Yasuda, T. & Mori, N., 2014. Sensitivity of tsunami wave profile and inundation simulations to earthquake slip and fault geometry for the 2011 Tohoku earthquake, *Earth planet. Space*, **66**(1), 1–20.
- Goff, J.A. & Jordan, T.H., 1988. Stochastic modeling of seafloor morphology: inversion of sea beam data for second-order statistics. *J. Geophys. Res.*, **93**, 13 589–13 608.
- Gower, J.C. & Dijksterhuis, G.B., 2004. *Procrustean Problems*, Oxford.
- Gusman, A.R., Tanaka, Y., Sakai, S. & Tsuchihama, H., 2012. Source model of the great 2011 Tohoku earthquake estimated from tsunami waveforms and crustal deformation data, *Earth planet. Sci. Lett.*, **341**, 234–242.
- Guttorp, P. & Gneiting, T., 2006. Studies in the history of probability and statistics XLIX on the Matérn correlation family, *Biometrika*, **93**(4), 989–995.
- Hair, J.F., Anderson, T.E., Tatham, R.L. & Black, W.C., 2010. *Multivariate Data Analysis*, Prentice Hall, 7th edn.
- Hartzell, S. & Heaton, T.H., 1983. Inversion of strong ground motion and teleseismic waveform data for the fault rupture history of the 1979 Imperial Valley, California, earthquake, *Bull. seism. Soc. Am.*, **73**, 1553–1583.
- Hayes, G., 2011. Rapid source characterization of the 03-11-2011 Mw 9.0 off the Pacific coast of Tohoku earthquake, *Earth planet. Space*, **63**(7), 529–534.
- Hering, A. & Genton, M.G., 2011. Comparing spatial predictions, *Technometrics*, **53**, 414–425.
- Ide, S., Baltay, A. & Beroza, G.C., 2011. Shallow dynamic overshoot and energetic deep rupture in the 2011 Mw 9.0 Tohoku-Oki earthquake, *Science*, **332**, 1426–1429.
- Kragh, E. & Christie, P., 2002. Seismic repeatability, normalized rms, and predictability, *The Leading Edge*, **21**, 640–647.
- Kristeková, M., Kristek, J. & Moczo, P., 2009. Time-frequency misfit and goodness-of-fit criteria for quantitative comparison of time signals, *Geophys. J. Int.*, **178**, 813–825.
- Lavallée, D., Liu, P. & Archuleta, R.J., 2006. Stochastic model of heterogeneity in earthquake slip spatial distributions, *Geophys. J. Int.*, **165**, 622–640.
- Lay, T., Ammon, C.J., Kanamori, H., Xue, L. & Kim, M.J., 2011. Possible large near-trench slip during the 2011 Mw 9.0 off the Pacific coast of Tohoku earthquake, *Earth planet. Space*, **63**(7), 687–692.
- Liu, P. & Archuleta, R.J., 2004. A new nonlinear finite fault inversion with three-dimensional Green's functions: application to the 1989 Loma Prieta, California, earthquake, *J. geophys. Res.*, **109**, B02318, doi:10.1029/2003JB002625.

- Mai, P.M. & Beroza, G.C., 2002. A spatial random field model to characterize complexity in earthquake slip, *J. geophys. Res.*, **107**(B11), 2308–2329.
- Mai, P.M. & Thingbaijam, K.K.S., 2014. SRCMOD: an online database of finite-fault rupture models, *Seism. Res. Lett.*, **85**(6), doi:10.1785/0220140092.
- Mai, P.M., Spudich, P. & Boatwright, J., 2005. Hypocenter locations in finite-source rupture models, *Bull. seism. Soc. Am.*, **95**(3), 965–980.
- Mai, P.M., Burjanek, J., Delouis, B., Festa, G., Holden, C., Monelli, D. & Uchide, T., 2007. Earthquake source inversion blindtest: initial results and further developments, *EOS, Trans. Am. geophys. Un.*, **88**, 855–877.
- McGuire, J.J., 2004. Estimating finite source properties of small earthquake ruptures, *Bull. seism. Soc. Am.*, **94**(2), 377–393.
- Minson, S.E., Simons, M. & Beck, J.L., 2013. Bayesian inversion for finite fault earthquake source models I—theory and algorithm, *Geophys. J. Int.*, **194**, 1701–1726.
- Minson, S.E., Murray, J.R., Langbein, J.O. & Gomberg, J.S., 2014. Real-time inversions for finite fault slip models and rupture geometry based on high-rate GPS data, *J. geophys. Res.*, **119**(4), 3201–3231.
- Monelli, D. & Mai, P.M., 2008. Bayesian inference of kinematic earthquake rupture parameters through fitting of strong motion data, *Geophys. J. Int.*, **173**(1), 220–232.
- Okada, R., 1985. Surface deformation due to shear and tensile faults in a half-space, *Bull. seism. Soc. Am.*, **75**(4), 1135–1154.
- Olson, A.H. & Apsel, R.J., 1982. Finite faults and inverse theory with applications to the 1979 Imperial Valley earthquake, *Bull. seism. Soc. Am.*, **72**, 1969–2001.
- Page, M., Mai, P.M. & Schorlemmer, D., 2011. Testing earthquake source inversion methodologies, *EOS, Trans. Am. geophys. Un.*, **92**(9), 75–75.
- Piatanesi, A., Cirella, A., Spudich, P. & Cocco, M., 2007. A global search inversion for earthquake kinematic rupture history: application to the 2000 Western Tottori, Japan earthquake, *J. geophys. Res.*, **112**, B07314, doi:10.1029/2006JB004821.
- Razafindrakoto, H.N.T. & Mai, P.M., 2014. Uncertainty in earthquake source imaging due to variations in source time function and Earth structure in source time function and earth structure, *Bull. seism. Soc. Am.*, **104**(2), 855–874.
- Satake, K., Fujii, Y., Harada, T. & Namegaya, Y., 2013. Time and space distribution of coseismic slip of the 2011 Tohoku earthquake as inferred from tsunami waveform data, *Bull. seism. Soc. Am.*, **103**(2B), 1473–1492.
- Shao, G. & Ji, C., 2012. What the exercise of the SPICE source inversion validation blindtest1 did not tell you, *Geophys. J. Int.*, **189**, 569–590.
- Shao, G., Li, X., Ji, C. & Maeda, T., 2011. Focal mechanism and slip history of the 2011 Mw 9.1 off the Pacific coast of Tohoku earthquake, constrained with teleseismic body and surface waves, *Earth planet. Space*, **63**(7), 559–564.
- Simons, M. et al., 2011. The 2011 magnitude 9.0 Tohoku-Oki earthquake: mosaicking the megathrust from seconds to centuries, *Science*, **332**(6036), 1421–1425.
- Somerville, P.G. et al., 1999. Characterizing crustal earthquake slip models for the prediction of strong ground motion, *Seis. Res. Lett.*, **70**(1), 59–80.
- Tappin, D.R. et al., 2014. Did a submarine landslide contribute to the 2011 Tohoku tsunami, *Marine Geology*, **357**, 344–361.
- Tinti, E., Spudich, P. & Cocco, M., 2005. Earthquake fracture energy inferred from kinematic rupture models on extended faults, *J. geophys. Res.*, **101**, 1–25.
- Wei, S., Graves, R.W., Helmberger, D., Avouac, J.P. & Jiang, J., 2012. Sources of shaking and flooding during the Tohoku-Oki earthquake: a mixture of rupture styles, *Earth planet. Sci. Lett.*, **333**, 91–100.
- Wilson, D.L., Baddeley, A.J. & Owens, R.A., 1997. A new metric for grey-scale image comparison, *Int. J. Comput. Vis.*, **24**(1), 5–17.
- Yagi, Y. & Fukuhata, Y., 2011. Rupture process of the 2011 Tohoku-oki earthquake and absolute elastic strain release, *Geophys. Res. Lett.*, **38**, L19307, doi:10.1029/2011GL048701.
- Yamazaki, Y., Lay, T., Cheung, K.F., Yue, H. & Kanamori, H., 2011. Modeling near-field tsunami observations to improve finite-fault slip models for the 11 March 2011 Tohoku earthquake, *Geophys. Res. Lett.*, **38**, L00G15, doi:10.1029/2011GL049130.
- Yue, H. & Lay, T., 2013. Source rupture models for the Mw 9.0 2011 Tohoku earthquake from joint inversions of high-rate geodetic and seismic data, *Bull. seism. Soc. Am.*, **103**(2B), 1242–1255.
- Yuen, D., Dzwinel, W., Ben-Zion, Y. & Kadlec, B., 2009. *Visualization of Earthquake Clusters Over Multi-dimensional Space*, Springer, Encyclopedia of complexity and system science edn.
- Zhang, L., Mai, P.M., Thingbaijam, K.K.S., Razafindrakoto, H.N.T. & Genton, M.G., 2015. Comparing earthquake slip models with the spatial prediction comparison test, *Geophys. J. Int.*, **200**(1), 185–198.

APPENDIX A: REALIZATIONS OF RANDOM FIELDS

Fig. A1 shows realizations of random fields using different parametrizations. At the top of the figure, we plot the mean of the random field ensemble. The generated random fields locate the slip patches at the same position (same seed value) while varying the Hurst parameter, H , and the correlation length, C .

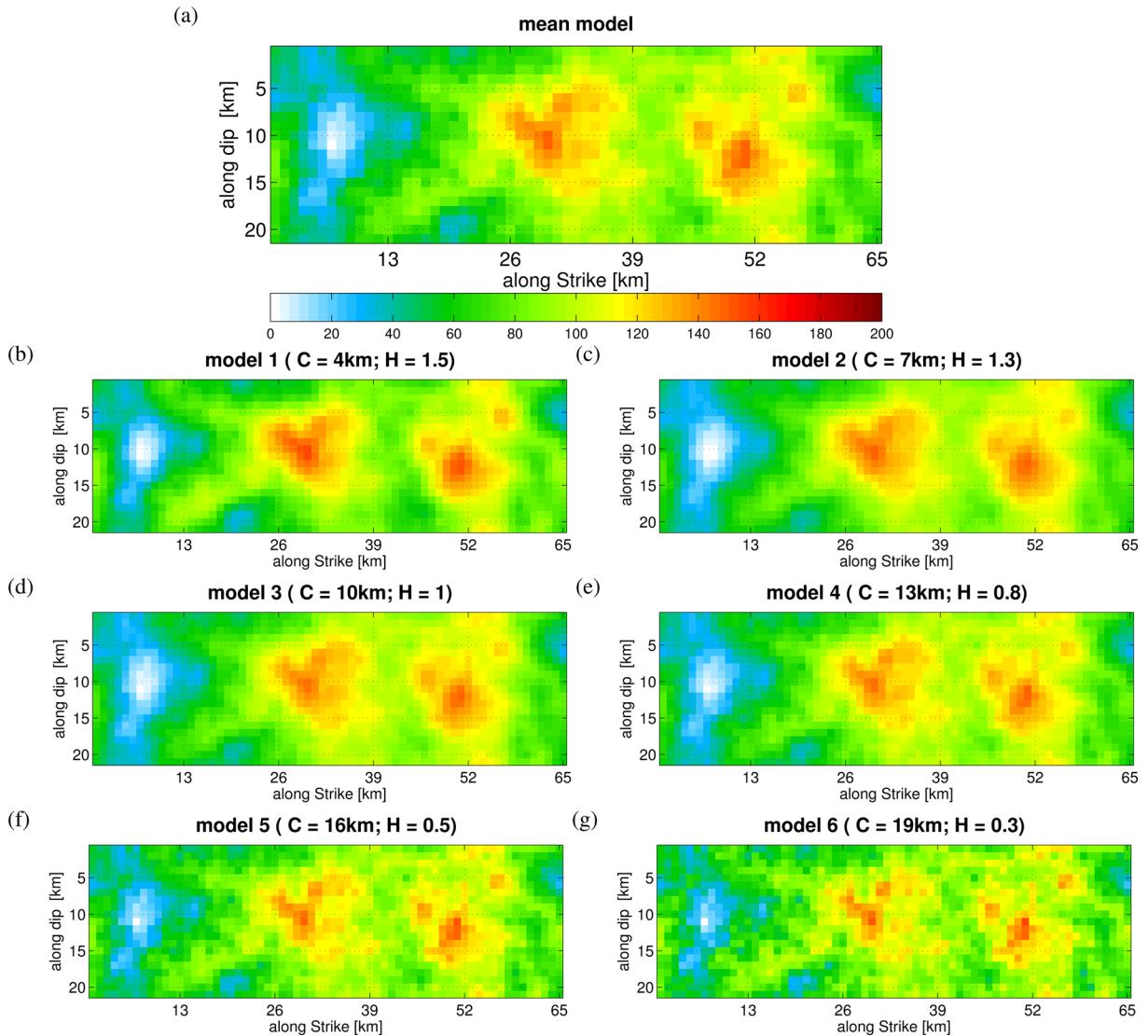


Figure A1. Random field example of different Hurst parameters, H , and correlation lengths, C .

APPENDIX B: PREDICTED SEAFLOOR DISPLACEMENT AND 3D CONFIGURATION OF THE TOHOKU SLIP MODELS

The following figures show the mean model and the predicted displacements from the 21 inverted slip models for the 2011 Tohoku earthquake. Fig. B1 presents the mean model. Figs B2, B3 and B4, on the other hand, illustrate the computed EW, NS, and vertical seafloor displacement, respectively.

We also show in Figs B5 and B6 the 3D MDS configuration for the smallest common rupture plane of the 2011 Tohoku slip models using normalized squared and grey-scale metrics.

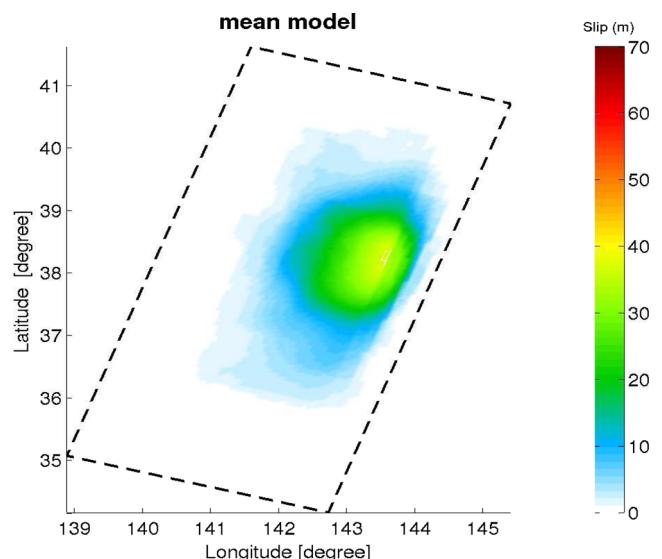


Figure B1. Mean model of the 21 inverted source models for the 2011 Tohoku earthquake.

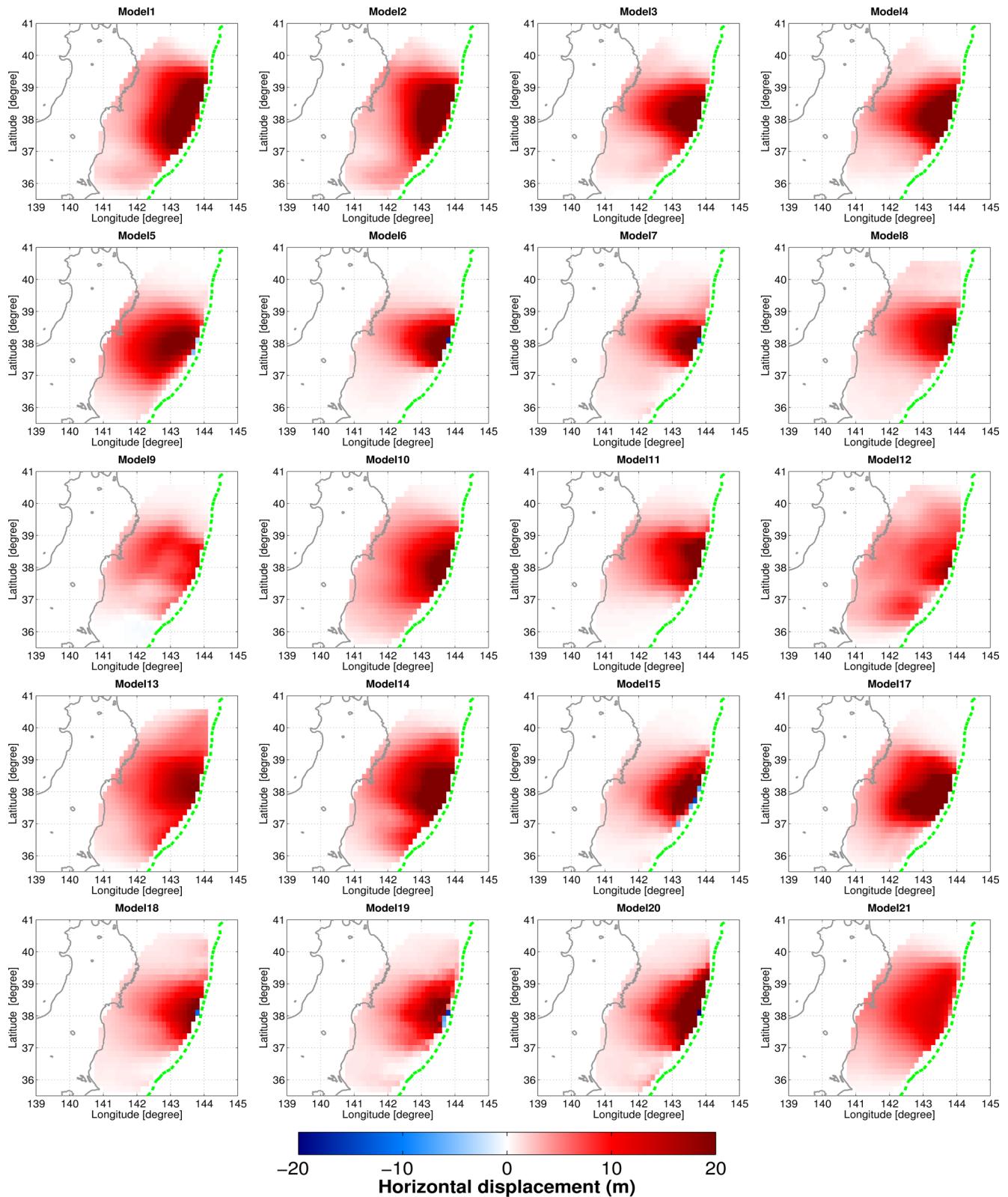


Figure B2. East–west horizontal seafloor displacement from 20 inverted slip models for the 2011 Tohoku earthquake (excluding the prediction from Model 16 that has a similar pattern as the one from Model 15). The green dashed line indicates the Japan trench subduction zone, while the grey curve indicates the coast line of central and northern Honshu, Japan.

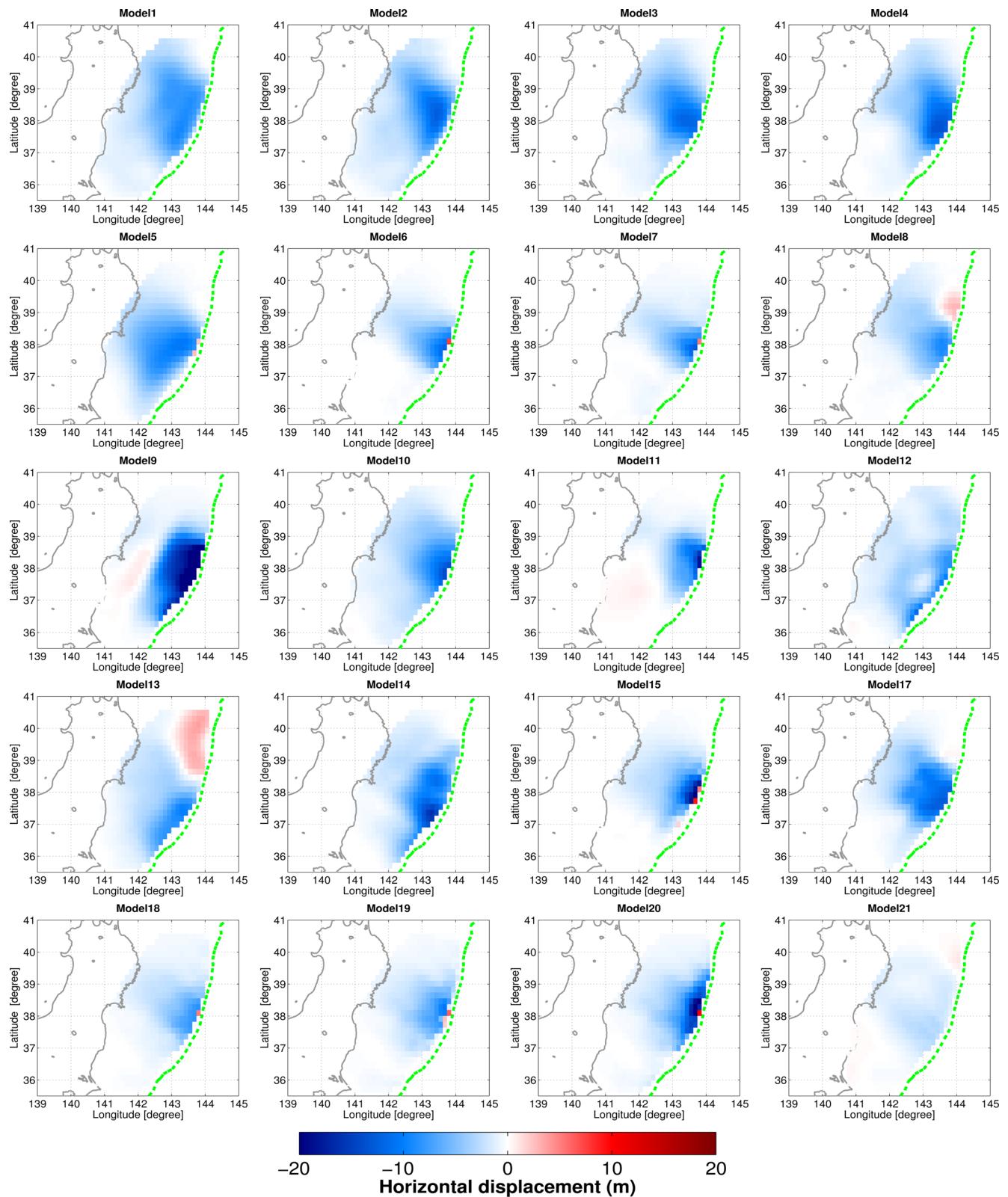


Figure B3. North-south horizontal seafloor displacement from 20 inverted slip models for the 2011 Tohoku earthquake (excluding the prediction from Model 16 that has a similar pattern as the one from Model 15). The green dashed line indicates the Japan trench subduction zone, while the grey curve indicates the coast line of central and northern Honshu, Japan.

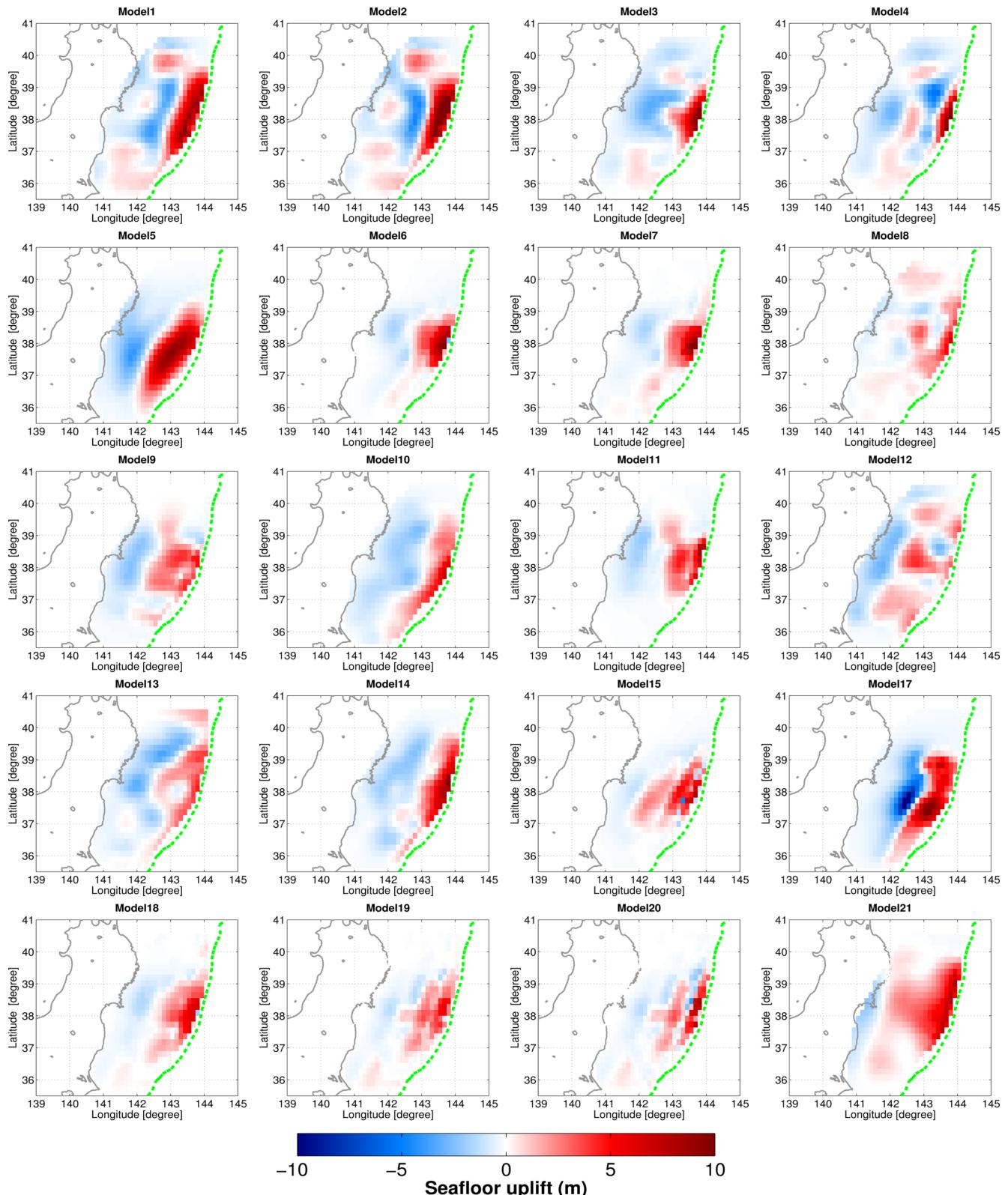


Figure B4. Vertical seafloor displacement from 20 inverted slip models for the 2011 Tohoku earthquake (excluding the prediction from Model 16 that has a similar pattern as the one from Model 15). The green dashed line indicates the Japan trench subduction zone, while the grey curve indicates the coast line of central and northern Honshu, Japan.

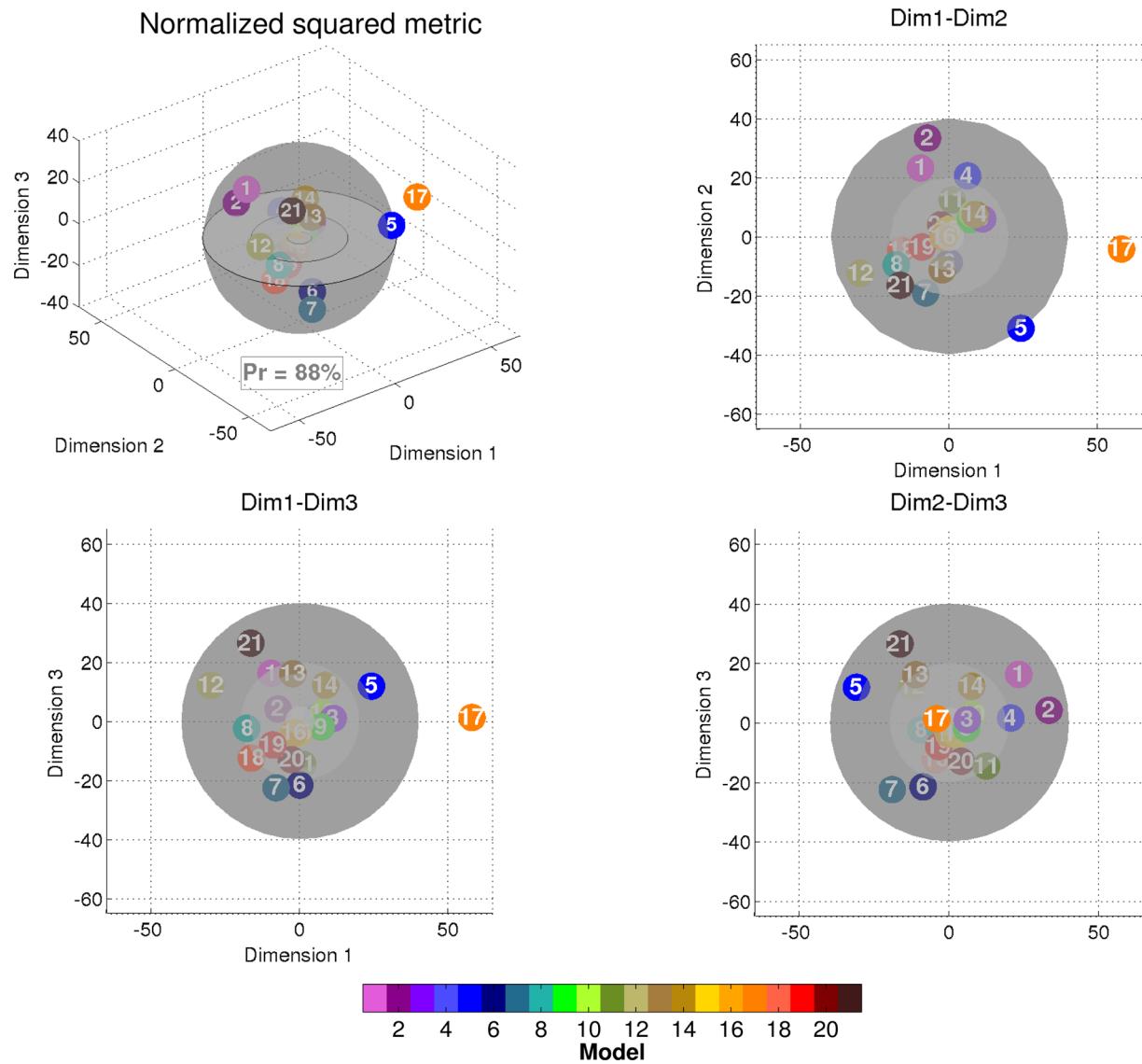


Figure B5. 3D MDS point-cloud considering the smallest common area of the 2011 Tohoku slip models and using the normalized squared metric.

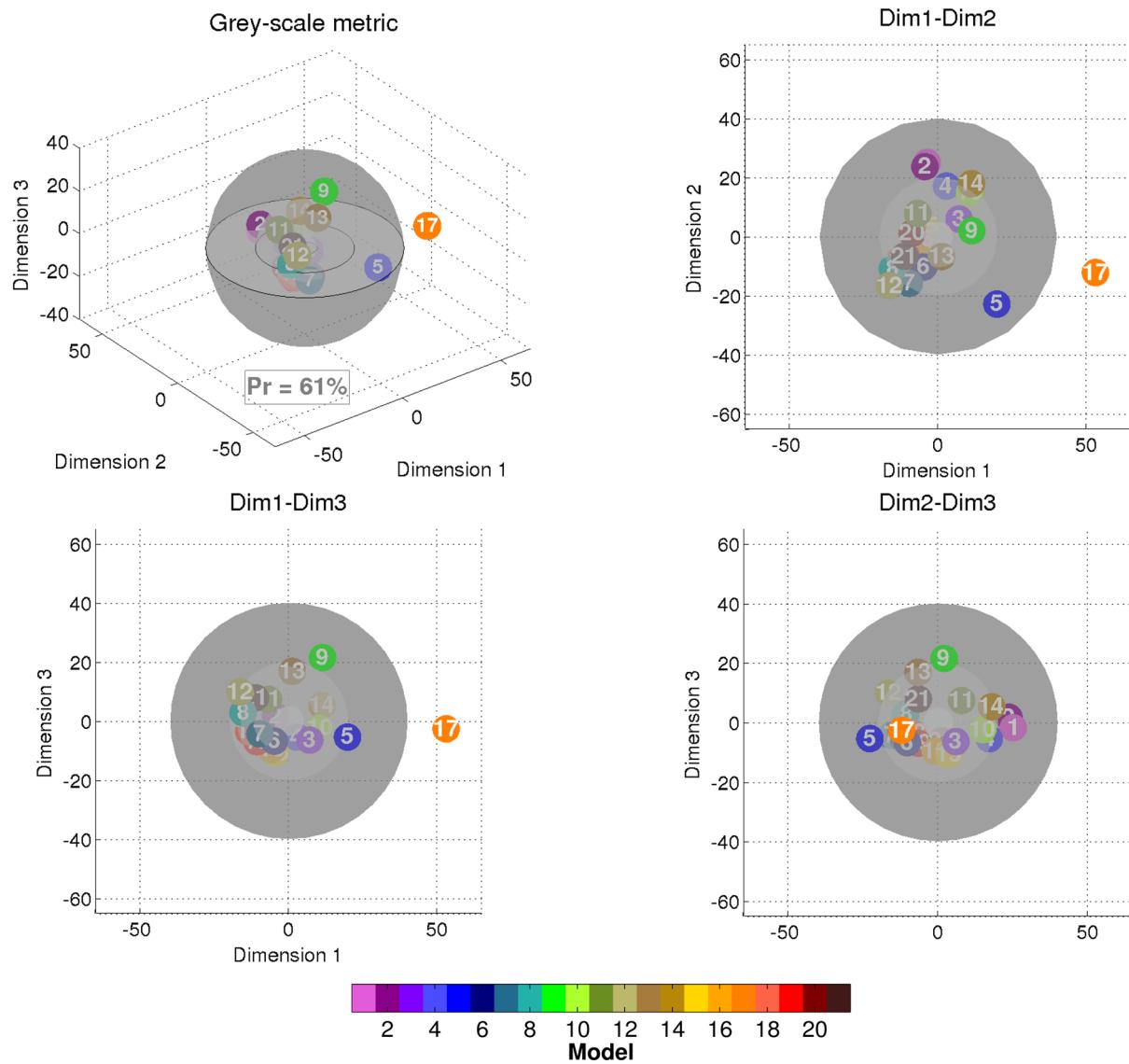


Figure B6. 3D MDS point-cloud considering the smallest common area of the 2011 Tohoku slip models and using the grey-scale metric.