Geophysical Journal International

Geophys. J. Int. (2015) **200,** 185–198
GJI Seismology

doi: 10.1093/gji/ggu383

Analysing earthquake slip models with the spatial prediction comparison test

Ling Zhang,[1] P. Martin Mai,[1] Kiran K.S. Thingbaijam,[1] Hoby N.T. Razafindrakoto[1] and Marc G. Genton[2]

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

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

Accepted 2014 October 1. Received 2014 September 30; in original form 2014 April 3

SUMMARY

Earthquake rupture models inferred from inversions of geophysical and/or geodetic data exhibit remarkable variability due to uncertainties in modelling assumptions, the use of different inversion algorithms, or variations in data selection and data processing. A robust statistical comparison of different rupture models obtained for a single earthquake is needed to quantify the intra-event variability, both for benchmark exercises and for real earthquakes. The same approach may be useful to characterize (dis-)similarities in events that are typically grouped into a common class of events (e.g. moderate-size crustal strike-slip earthquakes or tsunamigenic large subduction earthquakes). For this purpose, we examine the performance of the spatial prediction comparison test (SPCT), a statistical test developed to compare spatial (random) fields by means of a chosen loss function that describes an error relation between a 2-D field ('model') and a reference model. We implement and calibrate the SPCT approach for a suite of synthetic 2-D slip distributions, generated as spatial random fields with various characteristics, and then apply the method to results of a benchmark inversion exercise with known solution. We find the SPCT to be sensitive to different spatial correlations lengths, and different heterogeneity levels of the slip distributions. The SPCT approach proves to be a simple and effective tool for ranking the slip models with respect to a reference model.

Key words: Inverse theory; Spatial analysis; Earthquake source observations.

1 INTRODUCTION

In recent years, finite-fault source inversions have become a standard tool to image the kinematic space–time evolution of the earthquake rupture process. Due to increased and often real-time availability of seismic and geodetic data, and advancements in inversion techniques and computing facilities, such source inversions are conducted routinely for large events, and solutions are often presented hours after a large earthquake (United States Geological Survey's 'finite-fault models', http://earthquake.usgs.gov/earthquakes/eqinthenews/, and California University of Technology's 'Source Models of Large Earthquakes', http://www.tectonics.caltech.edu/slip_history/index.html). Often, refined source inversions are then subsequently published to examine and study the details of the rupture kinematics (e.g. Hayes *et al.* 2010; Wei *et al.* 2012; Fielding *et al.* 2013; Yue & Lay 2013).

Such earthquake rupture models constitute an important resource for subsequent studies on fault mechanics, seismotectonic interpretation, stress modelling and ground-motion simulations. However, there are several impeding issues regarding the reliability of the inverted models arising from the ill-posed nature of the inverse problem, limited and non-uniform data coverage, differences in selection and processing of the available data, incompletely known Earth structure, and variations in *a priori* assumptions on the fault-geometry (Beresnev 2003; Mai *et al.* 2007; Shao & Ji 2012). Hence, earthquake source inversions come with considerable uncertainty, which however is only rarely investigated in as much detail as by Hartzell *et al.* (1991, 2007), Custodio *et al.* (2005), Monelli & Mai (2008), Monelli *et al.* 2009 and Razafindrakoto & Mai (2014).

Often, several rupture models have been published for the same earthquake, but source parameters (e.g. slip on the fault) visually appear vastly different from each other. Results from a 'blind test' of earthquake source inversions (Mai *et al.* 2007) indicate (1) that a good-fit to the data does not necessarily yield a correct source model and (2) that accuracy is not ensured even with near-fault data, known Earth structure and fault geometry. The performance of source inversion codes, and the resulting uncertainties in kinematic source parameters are still not well understood. We thus strive to gain further insight into source-model uncertainties, to understand the factors that cause the model variability, and to identify the robust

features of these source models. To this end, the source inversion validation (SIV, http://equake-rc.info/siv) project (Page *et al.* 2011; Mai 2013) provides benchmark exercises and corresponding datasets to facilitate verification of inversion (and forward-modelling) codes, as well as the comparison and validation of the inferred rupture models.

Kinematic finite-fault models employ several parameters to describe the spatiotemporal rupture evolution, namely the final slip on the fault, the rise-time (duration of slip at each point of the fault), the rupture onset time, and the rake (angle of slip direction). Typically, the parametrization involves an elementary source time function defined for the entire rupture plane, which can be repeated to build so-called multitime-window inversions (Olson & Apsel 1982; Cohee & Beroza 1994; Wald & Heaton 1994). Simple source model parametrizations employ regular grids over a single plane, while more complex ones may comprise multiple fault segments to capture complex faulting geometry. First order earthquake source parameters like hypocentre location and local magnitude are generally independently derived, while seismic moment is an immediate outcome of the inversion.

This study focuses on evaluating the intra-event differences of distributed slip on a single fault plane (henceforth called slip models or simply 'models') by means of a rigorous quantitative analysis of the observed variability. Our analysis considers various cases for spatial distributions of final slip. To develop a quantitative understanding and statistically robust approach, we adopt the spatial prediction comparison test (SPCT) proposed by Hering & Genton (2011). The SPCT was developed for general spatial (random) fields, and is here applied for comparing characteristics of earthquake slip on a fault. Our analyses consider two specific cases: (1) synthetic slip models generated as stochastic random fields whose spatial complexities are controlled by spatial correlation lengths and heterogeneity levels and (2) inverted source models from a series of SIV benchmark exercises.

We first present a brief review on the testing procedure of the SPCT, which is then applied to a set of synthetic slip models with varying characteristics in random field parametrization. In this case, the reference model is computed as the average of the tested slip models. Subsequently, we perform the SPCT-procedure on inverted rupture models that are obtained for a benchmark exercise with known ('true') solution that serves as the reference model. We conclude with discussing the results in terms of further applicability of the SPCT, and present limitations and possible extensions of the approach.

2 SPATIAL PREDICTION COMPARISON TEST

The testing procedure of the SPCT is an extension of the time series test introduced by Diebold & Mariano (1995) to spatial random fields. The SPCT yields a statistical test for testing the null hypothesis that there is no significant difference, on average, between two competing spatial predictions. The results of the SPCT can thus be used to determine which of two competing models better matches a reference model.

The major advantage of the SPCT method is that it does not require any assumption on the distribution of the prediction errors computed from a loss function; they can be Gaussian or non-Gaussian, and may or may not have zero-mean. In addition, any loss function can be used in the SPCT that provides statistical tests that account for intensity or/and location errors. For example, the square and absolute error loss functions inform about intensity errors, while the correlation skill measures location errors. Hence, SPCT results have to be interpreted in the context of the chosen loss function.

Assume a spatial process $\{Z(\mathbf{s}) \in \mathbb{R} : \mathbf{s} \in \Omega \subset \mathbb{R}^2\}$ has been observed at discrete locations (possibly on a grid) denoted by $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$, where $\mathbf{s}_i$ is the spatial location and Ω is the entire domain of the process. This spatial process is predicted by two sets of spatial prediction models denoted by $\hat{Z}_1(\mathbf{s})$ and $\hat{Z}_2(\mathbf{s})$. A general loss function $g[Z(\mathbf{s}_i), \hat{Z}_P(\mathbf{s}_i)]$ is then introduced between a particular realization $[Z(\mathbf{s}_i)]_{i=1}^n$ of the spatial process, and an arbitrary prediction model $[\hat{Z}_P(\mathbf{s}_i)]_{i=1}^n$; this loss function quantifies the difference between the prediction model and a given realization. Among the many possible loss functions, the following ones have proven to be particularly effective (Hering & Genton 2011; Gilleland 2013a):

Square error (SE) loss at location $\mathbf{s}_i$

$$g[Z(\mathbf{s}_i), \hat{Z}_P(\mathbf{s}_i)] = [Z(\mathbf{s}_i) - \hat{Z}_P(\mathbf{s}_i)]^2. \tag{1}$$

Absolute error (AE) loss at location $\mathbf{s}_i$

$$g[Z(\mathbf{s}_i), \hat{Z}_P(\mathbf{s}_i)] = |Z(\mathbf{s}_i) - \hat{Z}_P(\mathbf{s}_i)|. \tag{2}$$

Correlation loss (or correlation skill)

$$g[Z(\mathbf{s}_i), \hat{Z}_P(\mathbf{s}_i)] = \frac{n}{(n-1)\hat{\sigma}_Z\hat{\sigma}_P}[Z(\mathbf{s}_i) - \bar{Z}][\hat{Z}_P(\mathbf{s}_i) - \bar{Z}_P]. \tag{3}$$

In eq. (3), $\bar{Z}$ and $\bar{Z}_P$ are the estimated means for the observed and predicted values, respectively, and $\hat{\sigma}_Z$ and $\hat{\sigma}_P$ are their estimated standard deviations.

In these three loss functions, both the SE and AE loss functions account for intensity errors, and the correlation skill measures location errors. While smaller values resulting from the SE and AE loss indicate better prediction performance, the opposite applies for the correlation loss for which larger values imply better prediction performance.

In fact, the spatial process of interest is then the loss differential, $D(\mathbf{s})$, given as

$$D(\mathbf{s}) = g[Z(\mathbf{s}), \hat{Z}_1(\mathbf{s})] - g[Z(\mathbf{s}), \hat{Z}_2(\mathbf{s})] = f(\mathbf{s}) + \delta(\mathbf{s}), \tag{4}$$

where $f(\mathbf{s})$ is the mean trend, and $\delta(\mathbf{s})$ is a mean-zero stationary process with unknown covariance function $C(\mathbf{h}) = \mathrm{cov}[\delta(\mathbf{s}), \delta(\mathbf{s}+\mathbf{h})]$. That is, we are not comparing the loss function of a prediction model and a particular realization with each other, but the differences in loss functions for two prediction models, $\hat{Z}_1(\mathbf{s})$ and $\hat{Z}_2(\mathbf{s})$ with respect to a common realization or reference model, $Z(\mathbf{s})$. The test procedure of the SPCT using the loss differential therefore tests the null hypothesis that $\hat{Z}_1(\mathbf{s})$ and $\hat{Z}_2(\mathbf{s})$ have equal predictive ability on average with respect to $Z(\mathbf{s})$, that is

$$H_0 : \frac{1}{|\Omega|}\int_\Omega E[D(\mathbf{s})]\mathrm{d}\mathbf{s} = 0, \tag{5}$$

where $|\Omega|$ is the area of the domain Ω and E is expectation. If the mean trend is constant in space, that is, $f(\mathbf{s}) = \mu$, then the null hypothesis becomes $H_0 : \mu = 0$. Under increasing domain asymptotics, the mean loss differential, $\bar{D} = \frac{1}{n}\sum_{i=1}^n D(\mathbf{s}_i)$, asymptotically approaches a normal distribution

$$\frac{\bar{D} - \mu}{\sqrt{\mathrm{var}(\bar{D})}} \to N(0, 1), \tag{6}$$

as the number of points n approaches infinity (Park *et al.* 2009), and its variance is given by

$$\mathrm{var}(\bar{D}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} C(h_{ij}). \tag{7}$$

In eq. (7), $C(h_{ij})$ is the covariance function that describes the structure of spatial dependencies in the loss differential associated with $\delta(\mathbf{s})$ (h_{ij} is the distance between points $\mathbf{s}_i$ and $\mathbf{s}_j$).

Following Hering & Genton (2011), to estimate the unknown covariance function, we choose to first estimate the semivariogram $\gamma(h_{ij})$ given by the following equation:

$$\hat{\gamma}(h_{ij}) = \frac{1}{2|N(h_{ij})|}\sum_{N(h_{ij})}[D(\mathbf{s}_i) - D(\mathbf{s}_j)]^2, \tag{8}$$

where $N(h_{ij})$ is the set of all pairs of points that are separated by h_{ij}, and $|N(h_{ij})|$ is the total number of these points. Because the empirical semivariogram $\hat{\gamma}(h_{ij})$ cannot be computed at all lag distances but only at distances h_{ij} separating observations, and due to variability in $\hat{\gamma}(h_{ij})$-values (in particular at larger h_{ij}; see Fig. 11 for example), it is not ensured that eq. (8) returns a valid semivariogram (positive definite function) for subsequent (geo-)statistical analysis. Therefore, empirical semivariograms are often approximated by selected model functions, so-called parametric semivariograms. In this case, a parametric semivariogram, $\hat{\gamma}(h_{ij}\,|\,\boldsymbol{\theta})$, is fit to the empirical semivariogram by applying a weighted least-squares technique to estimate model parameters $\boldsymbol{\theta}$; see Cressie (1993). The covariance function $C(h_{ij})$ is then given by the relationship $C(h_{ij}) = \gamma(\infty) - \gamma(h_{ij})$.

According to eqs (6) and (7), the test statistic proposed by Hering & Genton (2011), under the assumption of constant trend, is then obtained as

$$S_V = \frac{\bar{D}}{\sqrt{\dfrac{1}{n^2}\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n}[\hat{\gamma}(\infty|\hat{\boldsymbol{\theta}}) - \hat{\gamma}(h_{ij}|\hat{\boldsymbol{\theta}})]}}. \tag{9}$$

where $\hat{\boldsymbol{\theta}}$ is the estimated value of the (unknown true) parameter $\boldsymbol{\theta}$. We analyse the resulting test statistic in terms of p-values. The null hypothesis (eq. 5) is not rejected (that is, we have no evidence that two competing models do not have—on average—equal predictive ability with respect to the reference model), if the p-value is greater than a chosen statistical significance level (e.g. 5 per cent). Otherwise, the null hypothesis is rejected indicating that the two competing models have significantly different predictive ability on average. We report our findings for the most common significance levels in statistical analysis, that is 10, 5 and 1 per cent (Fisher 1925). Lower significance levels result in more conservative statistical tests.

The test above assumes that the mean trend $f(\mathbf{s})$ is constant over space. However, if a trend is known or suspected, it can be estimated using ordinary least squares (OLS), and then removed from $D(\mathbf{s})$, so that the test S_V can be computed for $D(\mathbf{s}) - \hat{f}(\mathbf{s})$ instead (Hering & Genton 2011; Gilleland 2013a).

Since the SPCT procedure is only related to the distance between two spatial locations, it can be used on gridded and non-gridded spatial fields. Generally, the fields are assumed to be isotropic when fitting the semivariogram models to the data. Directional semivariograms can be used to examine whether isotropy is present. If the process is found to be anisotropic, the SPCT can be performed based on an anisotropic semivariogram model (Cressie 1993).

In this study, all analyses are conducted assuming isotropic parametric semivariogram models to fit the empirical semivariogram. The widely used exponential model is our first choice, given as

$$\gamma(h|a,b) = a[1 - \exp(-h/b)]. \tag{10}$$

In case a good-fitting semivariogram cannot be obtained with this simple model, we apply the hole-effect model that may better characterize the spatial fields used in this study. It is given by (Cressie & Wikle 2011)

$$\gamma(h\,|\,a,b,c) = \begin{cases} 0 & h = 0, \\ c - a\sin(h/b)/(h/b) & h > 0, \end{cases} \tag{11}$$

where h is distance, and a, b and c are positive parameters to be estimated in eqs (10) and (11).

In order to decide whether or not to reject the null hypothesis, the 5 per cent level is our first choice. If differentiating competing models becomes difficult using the 5 per cent level, we also test at the 10 per cent level.

Below, we summarize the main steps of test procedure of the SPCT:

(1) Choose the loss function and compute the loss differential (eq. 4).
(2) Estimate the empirical semivariogram (eq. 8).
(3) Choose an appropriate parametric semivariogram (eqs 10 or 11) to fit the empirical semivariogram and apply weighted least-squares technique to estimate the parameters.
(4) Compute the test statistic (eq. 9).
(5) Calculate the p-value.
(6) Examine whether or not to reject the null hypothesis at a chosen significance level, and conclude whether or not two competing models have equal predictive ability on average with respect to the reference model.

3 SYNTHETIC SLIP MODELS

To test the SPCT methodology and gain an understanding of its capabilities and limitations, we start our analysis with a controlled experiment using synthetic slip models. The goals is to also examine which loss function(s) may be best suited to characterize and quantify the differences in slip models with varying degree of spatial heterogeneity. Synthetic slip models are generated using the approach of Mai & Beroza (2002) in which earthquake slip distributions are modelled based on the von Kármán autocorrelation function. Its power spectral density in wave number domain is given by

$$P(k) = \frac{a_x a_z}{(1 + k^2)^{H+1}}, \tag{12}$$

where H is the Hurst exponent, k is wave number, and a_x and a_z are the correlation lengths in along-strike and downdip direction, respectively. Mai & Beroza (2002) show that the correlation lengths scale with earthquake size (magnitude), while H remains constant ($H \sim 0.7$). In the context of this study, we fix earthquake size, and then systematically vary parameters a_x, a_z and H to generate slip models with different levels of spatial complexities. Applying the SlipReal software package (http://equake-rc.info/CERS-software/rupgen/), the slip distribution is first prescribed in the Fourier domain in terms of its power-spectral density (eq. 12) with the chosen heterogeneity parameters. To this, uniformly distributed random phase angles are applied to complete the complex spectrum, with the additional condition of

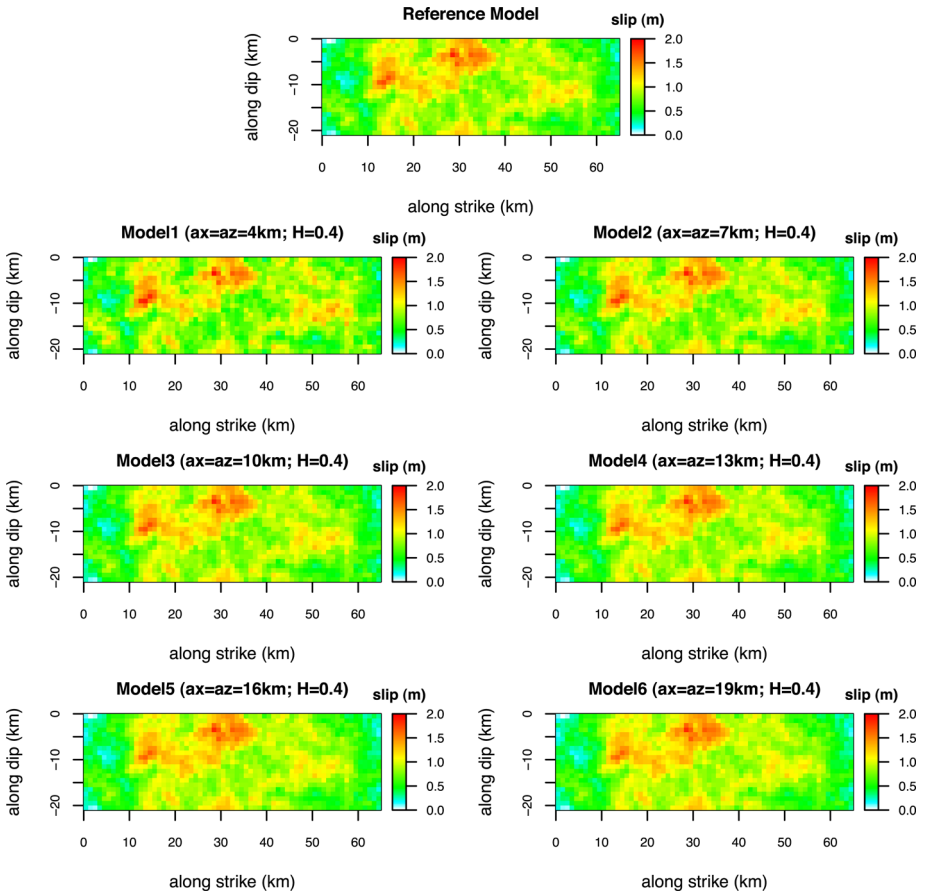

Figure 1. Reference slip model (top) and six different synthetic slip models with variable correlation lengths, but identical H-value ($H = 0.4$) and identical random-seed.

Hermitian symmetry. This spectral characterization of the slip distribution is then transformed into the spatial domain through inverse Fourier transformation. Since the distribution of random phase angles can be controlled through the choice of specific random seed values, we can regenerate slip distributions that only differ in terms of the scale lengths of the spatial heterogeneity patterns, but maintain their overall characteristics (Fig. 1).

We generate synthetic slip distributions on a predefined plane of 65×21 km^2, with sub-fault size of 1×1 km^2, and scaled uniformly to moment magnitude M_w 7.0. We then compare various classes of slip models using the SPCT to determine whether significant differences exist between these models and to help assess which model—on average—is best with respect to the reference case. Classes of slip models are generated based on variations in the parameters of the von Kármán autocorrelation function (eq. 12). These variations include the different ranges of values (from small to large). Testing a wider range of parameter is possible, but in the context of examining heterogeneity of earthquake slip, we restrict our analysis to a series of test models that cover the expected physical range, while a large number of test models with smaller differences between them will not provide any additional information.

3.1 Variable correlation lengths and fixed seed

Applying the von Kármán model, we generate six synthetic slip models by systematically changing the correlation lengths according to $a_x = a_z = [4 + 3(i - 1)]$km, $i = 1, \ldots, 6$, but fixing $H = 0.4$ (Fig. 1). In this case, we fix the random seed, so that high and low slip zones appear in the same general region of fault plane. Fig. 1 reveals that these six competing models are visually almost indistinguishable. Denoting these as Model-1 to Model-6, we compute the reference model (or 'mean model') as the average of these six synthetic distributions. Depending on the particular study, the reference model can be an observed (slip) distribution, a known distribution (from a controlled experiment), or any other sensible reference distribution. Due to the choice of the average model as reference, all six slip distributions visually appear to be close to the reference model.

Applying the SPCT to these six models with respect to the reference model, we cannot reject the null hypothesis (eq. 5) for all models, even at the 10 per cent level, considering the correlation skill. Hence, statistically there are no significant differences on average between these models in terms of location error. This fact is actually a verification of what we would expect to observe from Fig. 1, so this means that the finding of the SPCT is consistent with expectations. However, using the SE and AE loss functions, the null hypothesis (eq. 5) is rejected for each pair of models even at the 1 per cent level, that is, the differences between each pair of models from the reference model become significantly different from each other on average. Hence, in this case the test remains inconclusive to assess which model best matches the reference model.

Next, we compare these six synthetic models in terms of the mean loss differentials $\bar{D}$ (eq. 4) for different loss functions (Fig. 2). Smaller square or absolute errors mean that a particular slip model better matches the reference solution. In contrast, higher correlation

Figure 2. Mean loss differentials $\bar{D}$ for different loss functions. Negative values (blue) indicate that the case named in the corresponding row is the better model in terms of the SE and AE loss functions, and positive values (red) indicate that the case named in the corresponding row is better based on the correlation skill.

Table 1. Percentage of null hypotheses rejected in 1000 replications. The tests for the SE and AE loss functions are reported at the 1 per cent level, and for the correlation skill at the 10 per cent level.

	Model-2	Model-3	Model-4	Model-5	Model-6
SE loss					
Model-1	97.6	98.9	98.9	98.0	96.2
Model-2		99.4	78.7	96.8	99.7
Model-3			99.2	99.7	99.8
Model-4				99.7	99.8
Model-5					99.8
AE loss					
Model-1	99.9	100.0	100.0	100.0	99.8
Model-2		100.0	72.8	97.4	100.0
Model-3			99.8	100.0	100.0
Model-4				100.0	100.0
Model-5					100.0
Correlation skill					
Model-1	0.7	0.0	0.0	0.0	0.0
Model-2		0.0	0.0	0.0	0.0
Model-3			0.0	0.0	0.0
Model-4				0.1	0.7
Model-5					4.0

skill value implies a better agreement between models. Thus, we can differentiate two competing models based on the sign of their mean loss differential. For example, in Fig. 2 negative values (blue) indicate that the case named in the corresponding row is the better model in terms of the SE and AE loss functions, and accordingly based on the correlation skill positive values (red) indicate that the case named in the corresponding row is better. Fig. 2 then reveals that Model-3 best matches the reference model (and hence is called the 'best model'), because $\bar{D}$ is negative in the row for the SE and AE loss functions, and positive in the row for the correlation skill. Similarly, Model-4, Model-2, Model-5, Model-6 and Model-1 come out as being ranked second, third, fourth, fifth and sixth, respectively.

To examine the robustness of the SPCT approach and investigate its results for many realizations of statistically equivalent slip distributions, we repeat this experiment 1000 times by using different random seeds. Table 1 reports correspondingly the percentage of cases for which the null hypothesis has been rejected for different loss functions. The ensemble results are almost identical to the initial example: based on the SE and AE loss functions, the null hypothesis is mostly rejected for each pair of models even at the 1 per cent level, but for the correlation skill we almost always cannot reject the null hypothesis for all models, even at the 10 per cent level. This statistical experiment reveals that the SPCT approach is stable and that the test result is insensitive to the change in correlation length for fixed Hurst number and fixed random seed.

3.2 Variable correlation lengths and random seeds

Building on the previous case, we now consider different random-seed values to generate the slip models (Fig. 3). Locations of high and low slip on the fault plane vary, and the size of the slipping patches increase. However, the small-scale roughness, controlled by H, is identical for all six slip realizations. The mean distribution of these six distributions is considered as reference model (Fig. 3). By construction, these six competing models are quite different from each other, while each shares some common features with the reference model. It is difficult to visually assess the differences in these models, but some underlying information can be obtained by applying the SPCT.

Fig. 4 summarizes the findings of the SPCT and the mean loss differentials. Based on the SE and AE loss functions, Model-6 is most similar to the reference model. It is significantly different from all other models (except maybe Model-4) at the 5 per cent level, and $\bar{D}$ is negative in the corresponding row. Model-2 is the worst (positive $\bar{D}$ in the corresponding row) at the 5 per cent significance level. Based on the correlation skill, the SPCT only finds that

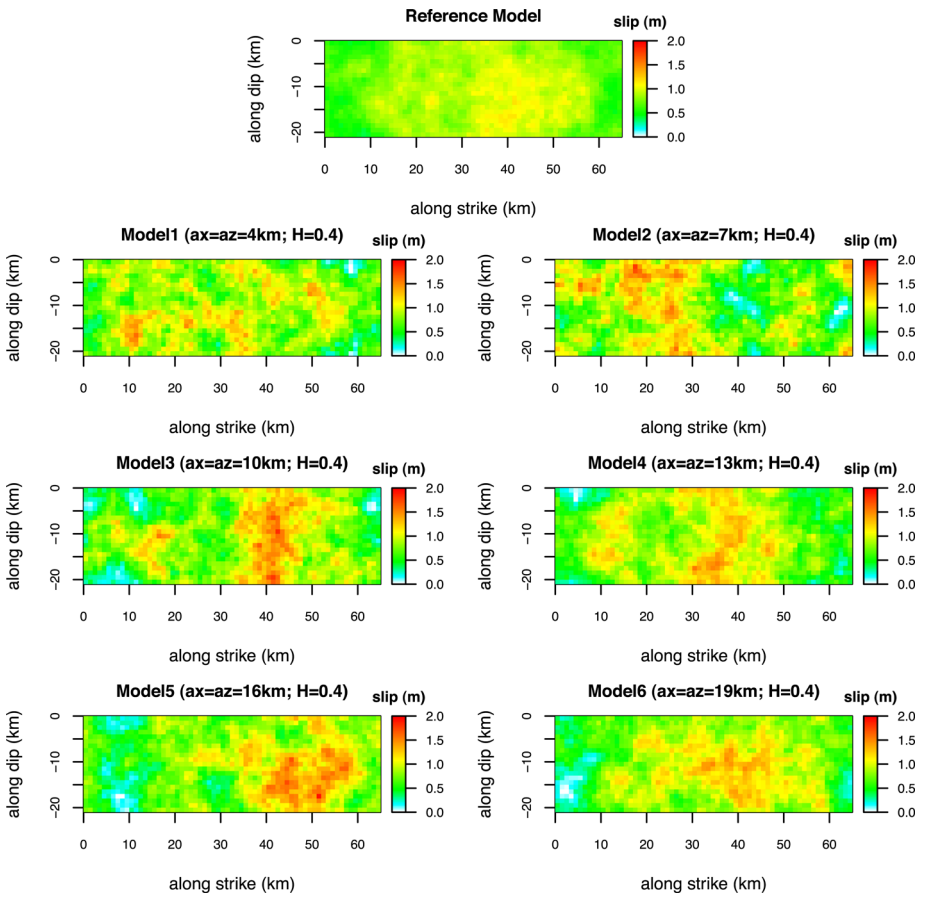

Figure 3. Reference slip model (top) and six different synthetic slip models with variable correlation length and identical H-value ($H = 0.4$). Here, random-seed values change, and hence strong spatial variations in the locations of high/low slip areas are present.

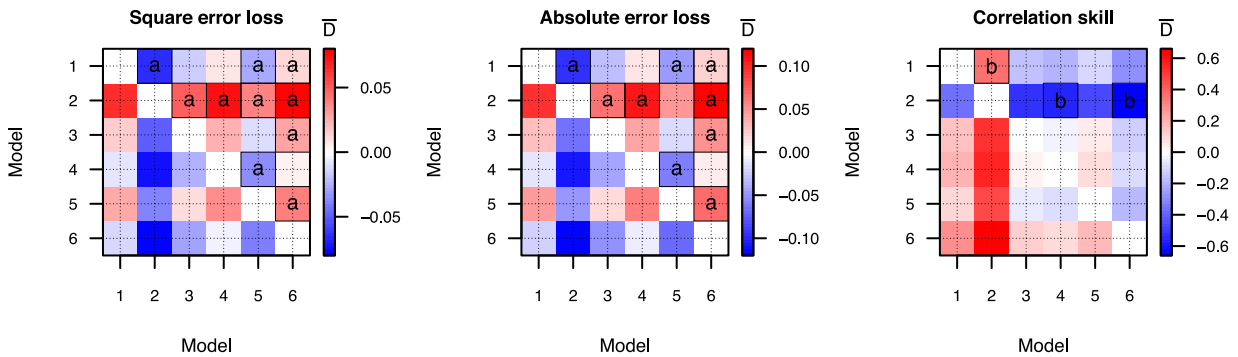

Figure 4. Mean loss differentials $\bar{D}$ for different loss functions with the hypothesis test results from the SPCT. Location with letter 'a' or 'b' indicates that the corresponding two models differ significantly from each other at the 5 or 10 per cent level, respectively.

Model-2 is significantly different from Model-1, Model-4 and Model-6 at the 10 per cent level. According to the mean loss differentials, the values of Model-6 are all positive in its row, indicating that Model-6 is best, and Model-2 is worst (because of the negative values in its row).

Performing again 1000 replications of this experiment, we aggregate the results in Table 2 that shows the percentage of cases for which the null hypothesis is rejected for different loss functions. For all loss functions, this percentage is found to be less than 30 per cent. Note that the random seed determines the phase spectrum of the randomized von Kármán distribution, and hence controls the spatial locations of high and low slip values. Combined with our analysis in the previous subsection, we observe that the test statistic exactly reflects these spatial differences, and is therefore sensitive to changes of the random seed value.

3.3 Variable correlation lengths, Hurst numbers and fixed seed

Modifying the first set of slip distributions, this class of models has identical correlation lengths and random-seed values as in Section 3.1 (Fig. 1), but now we change the Hurst exponent from Model-1 to Model-6, taking on values of 1.5, 1.3, 1.0, 0.8, 0.5 and 0.3, respectively (Fig. 5). The reference model is computed again as the mean of all six slip models. Fig. 5 illustrates that the

Table 2. Percentage of null hypotheses rejected in 1000 replications. The tests for the SE and AE loss functions are reported at the 5 per cent level, and for the correlation skill at the 10 per cent level.

	Model-2	Model-3	Model-4	Model-5	Model-6
SE loss					
Model-1	23.4	15.3	20.4	14.3	18.8
Model-2		24.4	16.4	21.9	16.4
Model-3			25.2	14.6	20.9
Model-4				22.7	14.9
Model-5					21.8
AE loss					
Model-1	22.7	14.2	20.0	12.7	18.3
Model-2		23.2	15.8	22.4	15.8
Model-3			25.1	12.6	19.9
Model-4				22.3	14.4
Model-5					20.2
Correlation skill					
Model-1	15.1	11.1	18.0	7.9	14.0
Model-2		14.6	9.2	10.9	6.6
Model-3			12.7	9.3	12.3
Model-4				12.8	8.3
Model-5					11.7

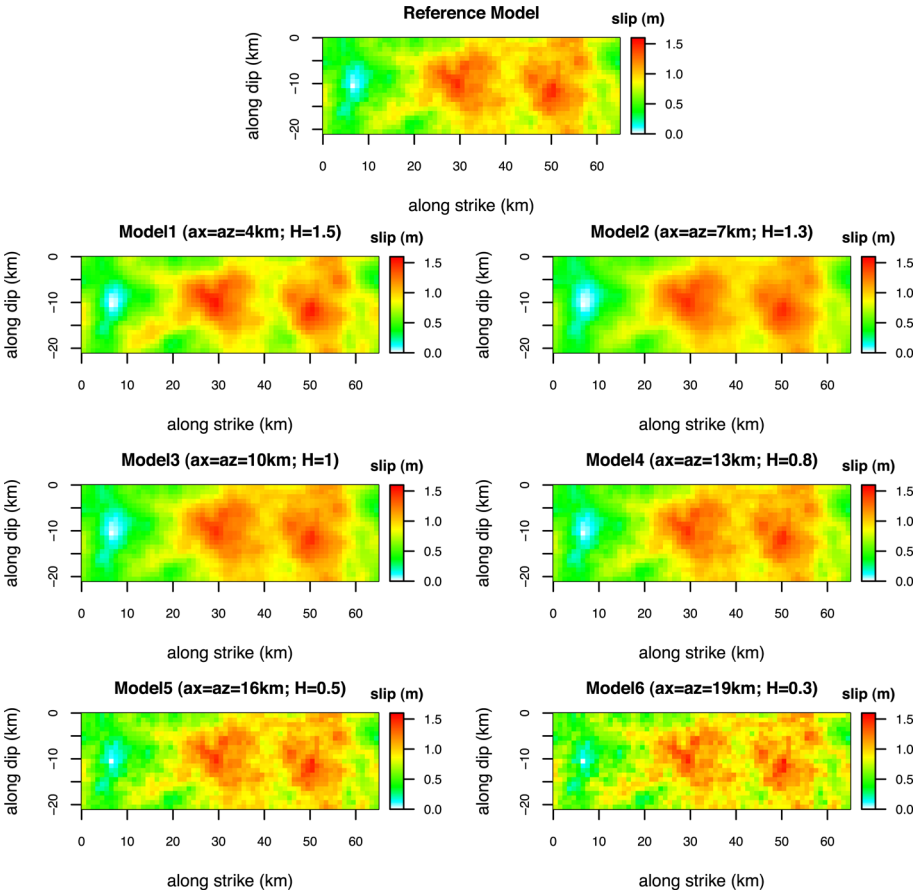

Figure 5. Reference slip model (top) and six different synthetic slip models with variable correlation lengths and variable H-values, but identical random-seed.

visual differences between these models are not very obvious, aside maybe from the variations in small-scale roughness through the changing Hurst exponent. However, the SPCT approach allows for discriminating these models.

Under the SPCT, Model-4 and Model-5 are the same on average, but significantly different from all other models at the 5 per cent level (Fig. 6) according to the SE and AE loss functions. Comparing the mean loss differentials, we find that Model-5 is closer to the reference model than Model-4. Model-6 is not significantly different from Model-1 on average, but is significantly different from all other models at the 5 per cent level and has positive $\bar{D}$ in the corresponding row. Therefore, Model-6 is most different from the mean model.

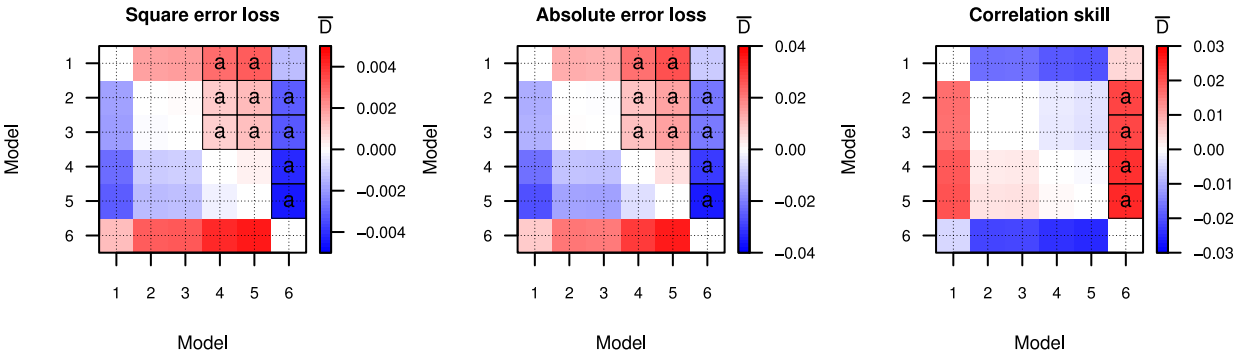

Figure 6. Mean loss differentials $\bar{D}$ for different loss functions with the hypothesis test results from the SPCT. Location with letter 'a' indicates that the corresponding two models differ significantly from each other at the 5 per cent level.

Table 3. Percentage of null hypotheses rejected in 1000 replications. The tests for the SE and AE loss functions are reported at the 5 per cent level, and for the correlation skill at the 10 per cent level.

	Model-2	Model-3	Model-4	Model-5	Model-6
SE loss					
Model-1	29.0	43.7	84.7	79.3	1.4
Model-2		6.0	43.2	91.1	95.5
Model-3			88.0	95.5	86.6
Model-4				26.6	86.9
Model-5					95.0
AE loss					
Model-1	25.3	37.1	97.1	87.8	0.6
Model-2		7.4	18.0	99.6	100.0
Model-3			90.3	97.1	97.5
Model-4				36.2	96.9
Model-5					100.0
Correlation skill					
Model-1	0.0	0.0	0.0	0.2	0.0
Model-2		0.0	0.0	0.0	0.1
Model-3			0.1	0.0	0.6
Model-4				0.0	6.4
Model-5					51.8

In terms of the correlation skill, we find from the SPCT that Model-6 is significantly different from all other models except Model-1 at the 5 per cent level. Checking also the values of $\bar{D}$, we conclude that Model-6 is worst and Model-5 is best in terms of how close they are to the reference model.

Table 3 reports the test results of the SPCT for 1000 replications. When using the SE and AE loss functions, some percentages are greater than 79 per cent, while others are lower than 45 per cent. According to the correlation skill, all values are very small except when comparing Model-5 and Model-6. Overall, this test case generates similar results as reported in Section 3.1, with only low sensitivity to changes in correlation length and Hurst number for fixed random seed.

3.4 Variable correlation lengths, Hurst numbers and random seeds

Now we consider the most complex, but perhaps also the most realistic case. Correlation lengths and Hurst number are variable (chosen as in the previous class of slip realizations), and different random-seed values are chosen for the six models. Thus, the variability in these slip realization is immediately visually evident. The reference model is again computed as the mean of all six slip realizations (Fig. 7). Because of different random-seed values, and variations in the random-field parameters, there are no obvious similarities between these six models and they all differ from the reference model. Hence, an intuitive or visual judgment which of the six models best matches the reference case is impossible. However, the SPCT approach provides this information.

The results are obtained by comparing these six models with the SPCT (Fig. 8). Based on the SE and AE loss functions, Model-2 is statistically indistinguishable from Model-5 and Model-6, on average, and is significantly different from all other models at the 5 per cent level. However, Model-5 and Model-6 differ significantly from each other at the 5 per cent level, and they reveal significant difference on average from all other models at the 5 per cent level. According to the mean loss differentials, we conclude that Model-5 is best, and Model-2 and Model-6 are ranked second and third respectively. In addition, Model-4 is worst because of the positive $\bar{D}$ in the corresponding row.

Using the correlation skill, the SPCT only finds that Model-6 is significantly different from Model-4 and Model-5, on average, at the 10 per cent level. Comparing the values of the mean loss differentials, we obtain that Model-2, Model-5, Model-3, Model-1, Model-6 and Model-4 come out as being ranked first, second, third, fourth, fifth and sixth, respectively. This result is slightly different from that found using the SE and AE loss functions.

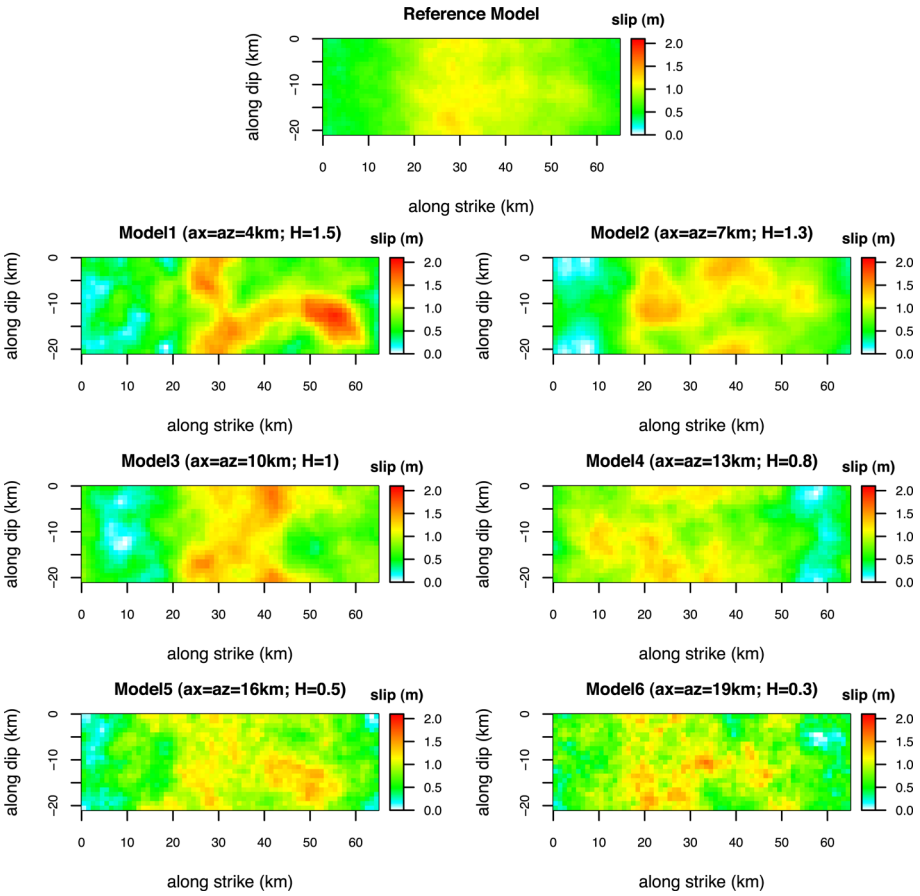

Figure 7. Reference slip model (top) and six different synthetic slip models with variable correlation length and variable H-values. Here, random-seed values change, and hence strong spatial variations in the locations of high/low slip areas are present.

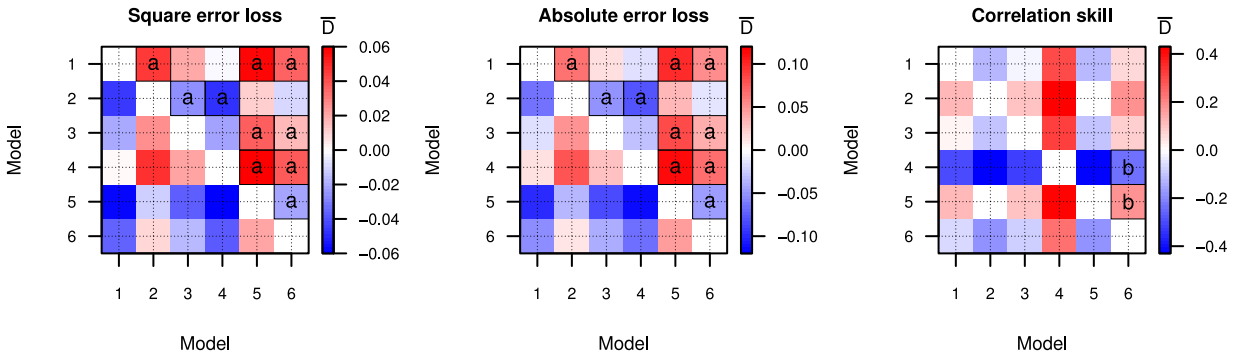

Figure 8. Mean loss differentials $\bar{D}$ for different loss functions with the hypothesis test results from the SPCT. Location with letter 'a' or 'b' indicates that the corresponding two models differ significantly from each other at the 5 or 10 per cent level, respectively.

The results for 1000 replications (Table 4) are similar to those of Section 3.2: the percentage of cases for which the null hypothesis is rejected remains small and is not much different from each other. This illustrates more evidently that the test result is very sensitive to changes of the random seed value.

4 INVERTED SLIP MODELS FROM A BENCHMARK EXERCISE

The four case studies in Section 3 indicate that the SPCT based on the loss differential is well suited to test which of two competing slip models better matches a common reference model, and to determine whether a significant difference exists on average. We also find that results may be different for different loss functions that account for intensity or location errors, implying that one has to choose the loss function based on the desired criteria or characteristics of the slip distribution that one wants to compare.

In this section, we perform a comparative analysis on slip models obtained in the context of benchmark exercises for the source inversion validation (SIV) project. The SIV efforts aim at better understanding the capabilities and limitations of finite-fault earthquake source inversion in order to robustly quantify the corresponding uncertainties. To this end, a series of benchmark exercises are constructed in which simulated rupture models for hypothetical earthquakes are used to generate synthetics datasets

Table 4. Percentage of null hypotheses rejected in 1000 replications. The tests for the SE and AE loss functions are reported at the 5 per cent level, and for the correlation skill at the 10 per cent level.

	Model-2	Model-3	Model-4	Model-5	Model-6
SE loss					
Model-1	27.7	18.6	27.3	16.5	23.9
Model-2		30.2	20.1	27.2	21.4
Model-3			34.0	22.3	30.7
Model-4				35.3	20.9
Model-5					33.8
AE loss					
Model-1	25.6	16.6	25.1	15.5	23.9
Model-2		29.8	14.9	27.2	21.6
Model-3			32.5	20.3	31.9
Model-4				33.7	22.9
Model-5					35.7
Correlation skill					
Model-1	12.3	5.2	14.4	5.0	9.5
Model-2		12.5	6.5	13.7	3.4
Model-3			14.0	7.2	12.3
Model-4				16.2	6.5
Model-5					14.5

(seismic and geodetic). The input rupture model is initially concealed from the SIV-participants who therefore invert the given datasets in a blind-test mode to retrieve kinematic source parameters. Further details on the benchmark exercise and corresponding results can be found at http://equake-rc.info/SIV/sivtools/list_solutions_for_benchmark/inv1.

Here we make use of the SIV benchmark in which the input source model is constructed from a spontaneous dynamic rupture calculation. Given 40 distributed observation points, the modelling teams inverted the corresponding synthetic waveforms to find the best-fitting rupture model. We investigate how their distributions of final slip compare with the known input slip distribution. Hence, in the context of Section 3, the reference model is given as an independent slip distribution that is not computed from the tested models themselves. Fig. 9 depicts the reference model, and the six different models obtained by different participants using different inversion procedures. All models cover the domain of 33×16 km^2, sampled by 1×1 km^2 cells. Obviously, these six slip distributions are visually very different from each other, in particular in terms of spatial resolution. The goal is to quantify which of these six proposed solutions best matches the known but independent reference model.

To compare these six slip models with the reference model, many methods can be applied. First, we calculate the differences of mean slip value and seismic moment between different slip models and the reference model (Table 5). It can be seen that Model-3 and Model-4 have small absolute differences, and hence are close to the reference model in terms of these overall macroscopic source parameters. In addition, we compute effective source dimensions and effective mean slip (Mai & Beroza 2000), shown for the reference model in Fig. 10. Table 6 compares these measurements using the ratio between the original dimensions ($L = 33$ km, $W = 16$ km) and the corresponding effective dimensions, that is, $\Delta L = L_{\text{eff}}/L$ and $\Delta W = W_{\text{eff}}/W$ and effective mean slips. Results depend on the chosen quantity, for example, Model-3 is closest to the reference model according to the effective length, but Model-1 is best by comparing the effective width. Using the effective mean slip Model-6 is nearest to the reference model. Even when combining these results, it remains difficult to determine which model best matches the reference model. We thus apply the SPCT to compare these models to obtain additional insight into (dis-)similarities of these models.

In applying the SPCT, we first examine whether the parametric semivariogram model fits the empirical semivariogram and whether isotropy is a reasonable assumption. Fig. 11 illustrates this process for Model-1 and Model-4. These two slip distributions are representative of the general behaviour of slip-inversion results, including their empirical and fitted hole-effect semivariograms and the corresponding empirical directional semivariogram for different loss functions. We find that the hole-effect model well describes the empirical semivariogram (Fig. 11, left column). We can check for isotropy from the empirical semivariogram by direction. If isotropy is perfectly satisfied, covariance values should be equal at identical distances in different directions, that is, the right column plots of Fig. 11 would show perfect semicircles extending outwardly from the origin (at coordinate [0,0]). Here we find there are some departure from isotropy, but none that strongly violates the isotropy assumption. Therefore, we only consider isotropic semivariograms in this analysis.

The findings of the SPCT and the mean loss differentials are summarized in Fig. 12. Based on the SE loss function, Model-3 is significantly different from all other models at the 5 per cent level, and has negative in the corresponding row, while Model-2, Model-4 and Model-5 differ significantly from each other at the 5 per cent level. Hence, Model-3 is most similar to the reference model. When considering the mean loss differentials, we are able to rank Model-2, Model-4, Model-1, Model-5 and Model-6 as second, third, fourth, fifth and sixth, respectively.

Based on the AE loss function, Model-3 is best (negative $\bar{D}$ in its row) with the statistical significance at the 5 per cent level, and Model-4 is significantly different from Model-2 and Model-5 on average at the 5 per cent level, and Model-1 and Model-2 differ significantly from each other on average at the 5 per cent level. On the basis of the mean loss differentials, we therefore rank Model-2, Model-1, Model-4, Model-6 and Model-5 as second, third, fourth, fifth and sixth, respectively.

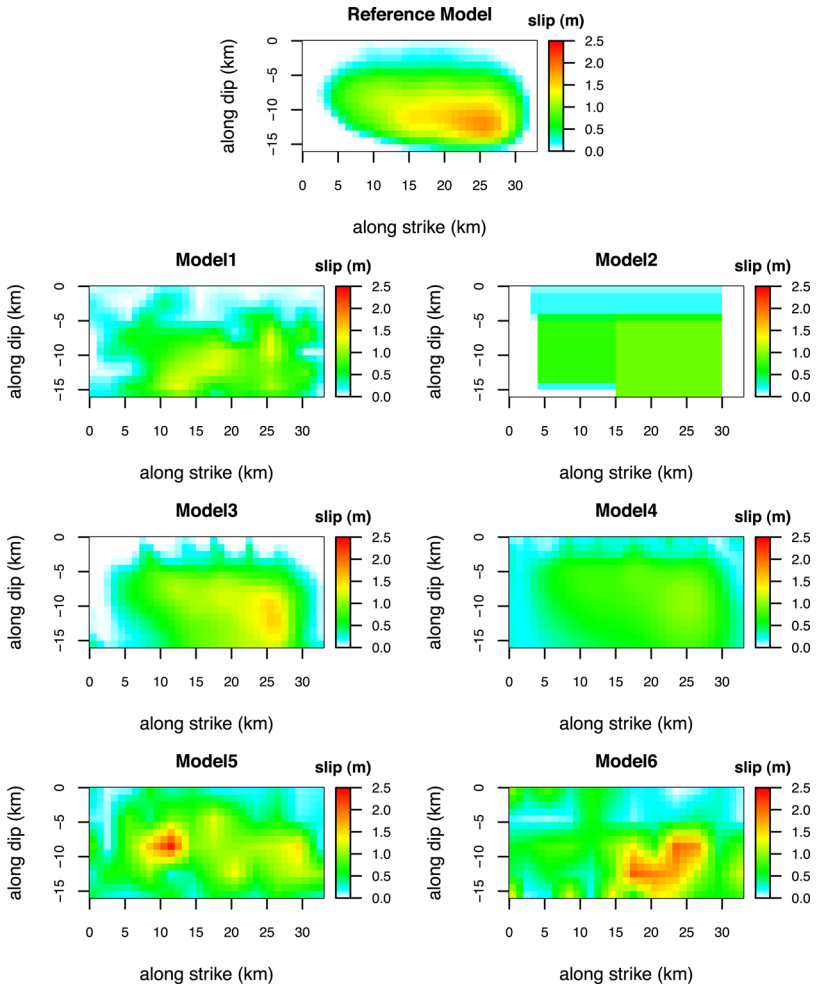

Figure 9. Reference slip model (top) and six different inverted slip models obtained in the context of benchmark exercises for the source inversion validation (SIV) project. The reference model is constructed from a spontaneous dynamic rupture calculation, while six different models obtained by different participants using different inversion procedures. These six slip distributions are visually very different from each other, in particular in terms of spatial resolution.

Table 5. Differences of mean slip value and seismic moment between different models and reference model.

Model	Mean slip difference (m)	Seismic moment difference (Nm)
Model-1	−0.104	-1.815×10^{18}
Model-2	−0.086	-1.514×10^{18}
Model-3	−0.019	-0.337×10^{18}
Model-4	−0.031	-0.551×10^{18}
Model-5	0.115	1.998×10^{18}
Model-6	0.101	1.758×10^{18}

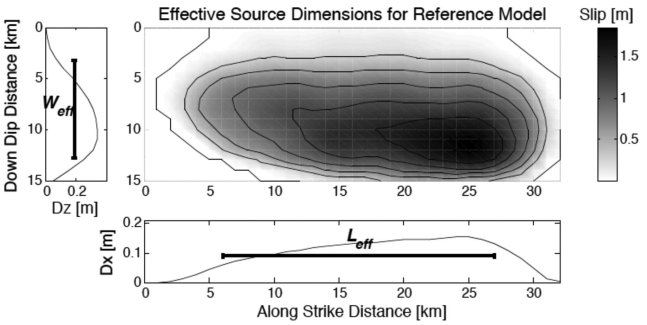

Figure 10. Slip distribution of the reference model (Fig. 9, top). Slip values in each subfault are shown in grey scale, contoured at 0.3-m intervals. Side boxes show effective length L_{eff} and effective width W_{eff} in down-dip (D_{x}) and along-strike direction (D_{z}), respectively.

According to the correlation skill, the SPCT finds that Model-3 differs significantly from Model-1, Model-5 and Model-6 on average at the 5 per cent level. The mean loss differentials reveal that Model-3, Model-4, Model-2, Model-1, Model-5 and Model-6 rank as first, second, third, fourth, fifth and sixth, respectively. Only the ranking of Model-4 is different from that found with the SE loss function.

Combining the results from these three loss functions, we conclude that Model-3 is the model that best reproduces the known (true) reference model, while Model-5 and Model-6 are far from the reference model.

5 DISCUSSION

The SPCT is applied here to quantitatively compare earthquake slip models, which is typically difficult to achieve when targeting the spatial characteristic of the slip distributions. The SPCT based on the loss differential provides a novel testing procedure that can be applied for any loss function. It can also be used to test which of

Table 6. Differences of effective length, width ratio and effective mean slip between different models and reference model.

Model	Effective length ratio difference	Effective width ratio difference	Effective mean slip difference (m)
Model-1	0.072	0.024	−0.439
Model-2	0.011	0.058	−0.359
Model-3	0.008	0.040	−0.157
Model-4	0.102	0.151	−0.520
Model-5	0.132	0.103	−0.223
Model-6	0.098	0.058	−0.103

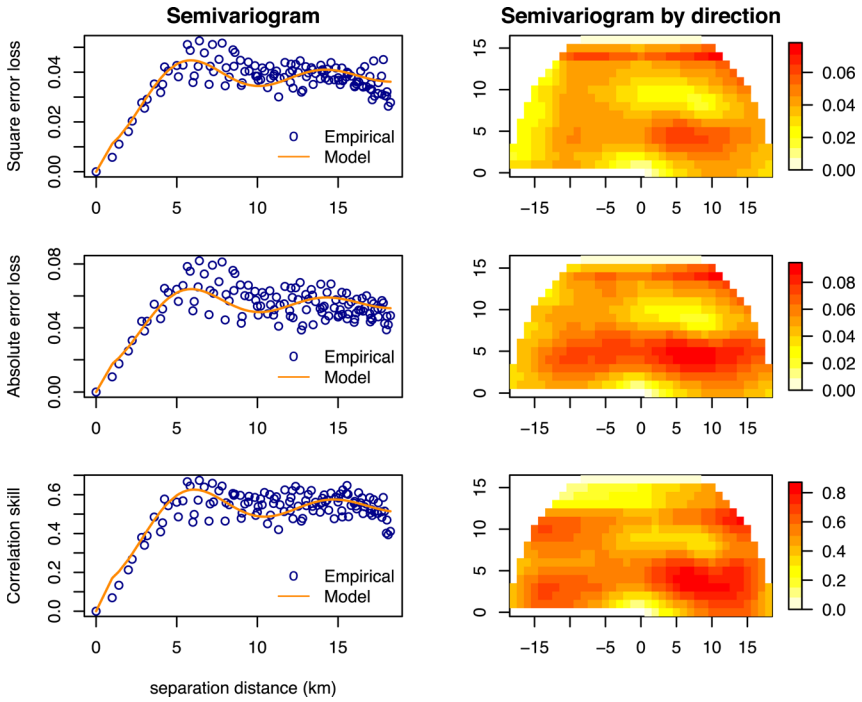

Figure 11. Semivariogram plots for comparing Model-1 and Model-4. Empirical and fitted hole-effect semivariograms (left) and the corresponding empirical directional semivariograms (right) for the different loss functions.

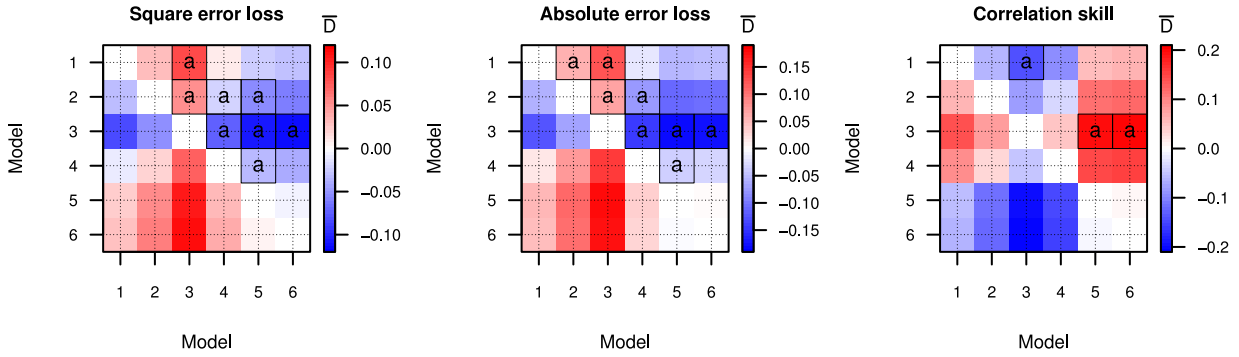

Figure 12. Mean loss differentials $\bar{D}$ for different loss functions with the hypothesis test results from the SPCT. Location with letter 'a' indicates that the corresponding two models differ significantly from each other at the 5 per cent level.

two competing slip models better matches a reference model, and to determine whether a significant difference exists on average.

We find that for different loss functions, the results of the SPCT are generally different. Since both the SE and AE loss functions account for intensity errors, their results are similar. The correlation skill measures location errors, and hence the corresponding results are slightly different from the findings with the SE and AE loss functions. We also find that for the examples presented, the SE and AE loss functions may be better suited to discriminate the different slip models than the correlation skill, potentially due to the relative small location errors among different slip models. Therefore, these data are best analysed in terms of the effective loss functions for different data.

Advantages of the SPCT are, for instance, that no distributional assumptions are imposed, that it can be applied efficiently to either gridded or non-gridded spatial fields, and that the procedure is both fast and straightforward to apply. The limitations of the SPCT are related to whether or not a parametric semivariogram model may well fit the empirical semivariogram, which in turn impacts strongly the final results. So if valid parametric semivariograms cannot be

found for the empirical semivariograms, then the SPCT approach cannot be applied. Of course, the majority of cases will not suffer from this drawback.

For earthquake rupture models, slip distributions from different inversion procedures may not be given over the same area, or may even be specified over entirely different geometries. To quantitatively compare these vastly different slip models, the SPCT can be applied only after some manipulation of the given slip distribution. For instance, we could define an enlarged area that encompasses all slip models, and then interpolate slip values onto a common grid. This approach may suffer from the fact that many zero-values over large areas of slip models will negatively affect the SPCT results. Furthermore, earthquake source inversions estimate the distribution of rupture time, rise time, rake angles, or even the entire slip-rate history on points of the fault. An application of SPCT to these 2-D fields is straight-forward, the extension of using SPCT on time-dependent quantities possible, but has not been attempted yet.

In addition, for a real earthquake, there is no reference model or known true solution. However, a reference model could be defined, for instance using a statistical measure based on the available solutions (e.g. the mean or median model). Applying then the SPCT to these available solutions (or models) and comparing them to the defined reference model helps to analyse and understand the relative uncertainties in these solutions. Without defining a reference model, comparisons can still be made between pairs of models, to test whether or not these two models differ significantly from each other on average.

6 CONCLUSION

We have compared earthquake slip models by applying the SPCT that provides a statistical procedure to test which of two (or any number of) competing models better matches a given reference model. The reference model can be computed from the test models themselves, but ideally should be an independently available slip distribution. The SPCT extracts underlying information when comparing two predictions with the reference field. In this study, we first consider synthetic slip distributions, generated using a von Kármán autocorrelation function whose three parameters are varied in a systematic manner. The insight from the synthetic test helps us to understand the subsequent application of the SPCT to inverted slip models of a benchmark exercise. We use three different loss functions to compare different models, applying suitable parametric semivariogram models to fit the empirical semivariograms. This approach allows for determining the most similar or most dissimilar model with respect to the reference model, and to rank all competing models. Thus, the SPCT approach helps to quantify the variability in earthquake rupture models by comparing different inverted slip models.

To assess the stability and sensitivity of the SPCT, we apply the method to 1000 stochastic realizations to generate statistically identical random slip distributions for which the location of high- and low-slip regions on the fault vary. The results indicate that the SPCT is rather insensitive to variations in random-field parameters (correlation length, Hurst number) if the random-phase information is not changed. Hence, the SPCT does not detect variations and short spatial scales. In contrast, it reliably discriminates slip distributions with variability in the locations of regions with high- and low-slip. We therefore conjecture that the SPCT can be applied, potentially with additional refinements and improvements, to automatically discriminate, quantify and rank similarities in slip distributions, and any other (geophysical) 2-D scalar field. Extensions to three dimensions and time-dependent fields are possible, but beyond the scope of this study.

Future work will focus on considering new and stronger loss functions, for instance, the warping loss function (Gilleland *et al.* 2010), which accounts for both intensity and location errors. In addition, we can apply the SPCT to other rupture parameters, for instance rise time or slip duration. Besides, the current SPCT approach lacks the ability to estimate the magnitude and location of a statistically significant spatial signal. For this purpose, local methods like the false discovery rate (FDR; Benjamini & Heller 2007) or enhanced FDR (Shen *et al.* 2002) can be used. We plan to implement such local methods into our testing procedure and to compare them to the SPCT results for better quantifying the variability of earthquake slip models, and to finally rank them with respect to a chosen reference model.

ACKNOWLEDGEMENTS

We would like to thank R Development Core Team (2012) for the R programming language and environment, and specifically the SpatialVx package (Gilleland 2013b) and the fields package (Nychka *et al.* 2013). We thank two anonymous reviewers and the Editor Eiichi Fukuyama for their thoughtful comments and constructive criticism. This research was supported by King Abdullah University of Science and Technology (KAUST).

REFERENCES

Benjamini, Y. & Heller, R., 2007. False discovery rates for spatial signals, *J. Am. Stat. Assoc.,* **102,** 1272–1281.

Beresnev, I.A., 2003. Uncertainties in finite-fault slip inversions: to what extent to believe? *Bull. seism. Soc. Am.,* **93,** 2445–2458.

Cohee, B. & Beroza, G., 1994. A comparison of two methods for earthquake source inversion using strong motion seismograms, *Ann. Geophys.,* **37,** 1515–1538.

Cressie, N., 1993. *Statistics for Spatial Data,* revised edition, Wiley.

Cressie, N. & Wikle, C.K., 2011. *Statistics for Spatio-Temporal Data,* Wiley.

Custodio, S., Liu, P. & Archuleta, R., 2005. The 2004 Mw 6.0 Parkfield, California, earthquake: inversion of near-source ground motion using multiple data sets, *Geophys. Res. Lett.,* **32,** L23312, doi:10.1029/2005GL024417.

Diebold, F.X. & Mariano, R.S., 1995. Comparing predictive accuracy, *J. Buss. Econ. Stat.,* **13,** 253–263.

Fielding, E.J., Sladen, A., Li, Z.H, Avouac, J.P., Burgmann, R. & Ryder, I., 2013. Kinematic fault slip evolution source models of the 2008 M7.9 Wenchuan earthquake in China from SAR interferometry, GPS and teleseismic analysis and implications for Longmen Shan tectonics, *Geophys. J. Int.,* **194,** 1138–1166.

Fisher, R.A., 1925. *Statistical Methods for Research Workers,* Oliver and Boyd.

Gilleland, E., 2013a. Testing competing precipitation forecasts accurately and efficiently: the spatial prediction comparison test, *Mon. Wea. Rev.,* **141,** 340–355.

Gilleland, E., 2013b. *SpatialVx: Spatial Forecast Verification,* R package version 0.1–5, Available at: http://cran.r-project.org/web/packages/SpatialVx/ (last accessed 1 August 2013).

Gilleland, E., Lindstrom, J. & Lindgren, F., 2010. Analyzing the image warp forecast verification method on precipitation fields from the ICP, *Wea. Forecast.,* **25,** 1249–1262.

Hartzell, S., Stewart, G. & Mendoza, C., 1991. Comparison of L1 and L2 norms in a teleseismic waveform inversion for the slip history of the Loma Prieta, California, earthquake, *Bull. seism. Soc. Am.,* **81,** 1518–1539.

Hartzell, S., Liu, P., Mendoza, C., Ji, C. & Larson, K., 2007. Stability and uncertainty of finite-fault slip inversions: application to the 2004 Parkfield, California, earthquake, *Bull. seism. Soc. Am.,* **97,** 1911–1934.

Hayes, G.P. *et al.*, 2010. Complex rupture during the 12 January 2010 Haiti earthquake, *Nat. Geosci.,* **3,** 800–805.

Hering, A.S. & Genton, M.G., 2011. Comparing spatial predictions, *Technometrics,* **53,** 414–425.

Mai, P.M., 2013. Uncertainty quantification in earthquake source inversions: the source inversion validation (SIV) project, *Proceedings of the EGU General Assembly Conference,* Abstracts EGU2013–3596, Vol. 15, p. 3596.

Mai, P.M. & Beroza, G.C., 2000. Source-scaling properties from finite-fault rupture models, *Bull. seismol. Soc. Am.,* **90,** 604–615.

Mai, P.M. & Beroza, G.C., 2002. A spatial random field model to characterize complexity in earthquake slip, *J. geophys. Res.,* **107**(B11), ESE 10-1-ESE 10-21.

Mai, P.M., Burjanek, J., Delouis, B., Festa, G., Francois-Holden, C., Monelli, D., Uchide, T. & Zahradnik, J., 2007. Earthquake source inversion blindtest: initial results and further developments, *EOS, Trans. Am. geophys. Un.,* **88**(52), Abstract S53C-08.

Monelli, D. & Mai, P.M., 2008. Bayesian inference of kinematic earthquake rupture parameters through fitting of strong motion data, *Geophys. J. Int.,* **173,** 220–232.

Monelli, D., Mai, P.M., Jónsson, S. & Giardini, D., 2009. Bayesian imaging of the 2000 Western Tottori (Japan) earthquake through fitting of strong motion and GPS data, *Geophys. J. Int.,* **176,** 135–150.

Nychka, D.W., Furrer, R. & Sain, S., 2013. *Fields: Tools for Spatial Data,* R package version 6.8, Available at: http://cran.r-project.org/web/packages/fields (last accessed 1 August 2013).

Olson, A. & Apsel, R., 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.

Park, B.U., Kim, T.Y., Park, J.S. & Hwang, S.Y., 2009. Practically applicable central limit theorem for spatial statistics, *Math. Geosci.,* **41,** 555–569.

R Development Core Team, 2012. *R: A Language and Environment for Statistical Computing,* R Foundation for Statistical Computing.

Razafindrakoto, H.N.T. & Mai, P.M., 2014. Uncertainty in earthquake source imaging due to variations in source time function and Earth structure, *Bull. seism. Soc. Am.,* **104,** 855–874.

Shao, G.F. & Ji, C., 2012. What the exercise of the SPICE source inversion validation BlindTest 1 did not tell you, *Geophys. J. Int.,* **189,** 569–590.

Shen, X., Huang, H.C. & Cressie, N., 2002. Nonparametric hypothesis testing for a spatial signal, *J. Am. Stat. Assoc.,* **97,** 1122–1140.

Wald, D.J. & Heaton, T.H., 1994. Spatial and temporal distribution of slip for the 1992 Landers, California earthquake, *Bull. seism. Soc. Am.,* **84,** 668–691.

Wei, S.J., Graves, R.W., Avouac, J.P. & Jiang, J.L., 2012. Sources of shaking and flooding during the Tohoku-Oki earthquake: a mixture of rupture styles, *Earth. planet. Sci. Lett.,* **333,** 91–100.

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,** 1242–1255.