Declustering and Poisson Tests

Brad Luen
Department of Mathematics
Reed College
Portland, OR, USA

Philip B. Stark
Department of Statistics
University of California
Berkeley, CA, 94720-3860 USA

Accepted 2011 ???. Received 2011 October ??; in original form 2011 January 29

SUMMARY
The commonly enunciated reason to decluster catalogs is so that the remaining “main” events will be consistent with a spatially inhomogeneous, temporally homogeneous Poisson process (SITHP) model. But are they? Conclusions depend on the declustering method, the catalog, the magnitude range, and the statistical test. Gardner and Knopoff’s (1974) conclusion that 1932–1971 southern California events with \( M \geq 3.8 \) are Poissonian after declustering apparently results from their use of a test with low power. That test ignores space, is insensitive to long-term rate variations, is relatively insensitive to seismicity rate fluctuations on the scale of weeks, and uses an inaccurate approximation to the null distribution of the test statistic. Better temporal tests and a novel spatio-temporal test show that SITHP does not fit \( M \geq 3.8 \) 1932–1971 or 1932–2010 Southern California Earthquake Center (SCEC) catalogs declustered using Gardner and Knopoff’s windows in a linked-window or a mainshock-window algorithm. For \( M \geq 4.0 \), SCEC catalogs declustered using the Gardner-Knopoff windows in a linked-window method are far closer to SITHP, while catalogs declustered using those windows in a mainshock-window method are inconsistent with SITHP. Reasenberg’s (1985) declustering method applied to southern California seismicity produces catalogs inconsistent with SITHP, even for events with \( M \geq 4.0 \).

If enough events are deleted from a catalog, the remainder always will be consistent with SITHP. This suggests posing declustering as an optimization problem: Delete the fewest events such that the remainder pass a particular test or suite of tests for SITHP. While that optimization problem is combinatorially complex, inexpensive suboptimal methods are surprisingly effective: Declustered catalogs can be consistent with temporal tests of SITHP at significance level 0.05 and have 50% to 80% more events than window-declustered catalogs that are inconsistent with SITHP. But tests that incorporate spatial information reject the SITHP hypothesis for those declustered catalogs, illustrating the importance of using spatial information.

Key words: Earthquake interaction, forecasting, and prediction; Probabilistic forecasting; Declustering; Statistical seismology

1 INTRODUCTION

We study the most common declustering methods, mainshock-window and linked-window declustering. There are also stochastic declustering methods, which use chance to decide whether to remove a particular event (Zhuang et al. 2002; Vere-Jones 1970); the “waveform similarity approach” (Barani et al. 2007); and others. See Davis and Frohlich (1991) and Zhuang et al. (2002) for taxonomies.

Mainshock-window methods remove the earthquakes in a space-time window around every “mainshock,” suitably defined. Mainshock-window methods can be thought of as punching a hole in the catalog after each mainshock. The hole is the window. Gardner and Knopoff’s windows (Knopoff and Gardner 1972; Gardner and Knopoff 1974) are common in mainshock-window declustering. They are larger in space and time the larger the shock is.

Linked-window methods calculate a space-time window
for every event in the catalog, not just mainshocks. In linked-window methods, an event is in a cluster if and only if it falls within the window of at least one other event in that cluster. Linked-window declustering replaces each cluster with a single event—for instance, the first, the largest, or an “equivalent event.” The most widely used linked-window method was developed by Reasenberg (1985). Reasenberg’s windows are larger in space but shorter in time the larger the shock is.

Earthquake catalogs are often declustered using window methods as a precursor to modeling the remaining events as a realization of a spatially inhomogeneous, temporally homogeneous Poisson process (SITHP). Tests of the null hypothesis that declustered catalogs are realizations of a SITHP have not rejected the null hypothesis, leading some studies to conclude that declustered catalogs are Poisson. For instance, the title of Gardner and Knopoff (1974) is “Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?“ The abstract: “Yes.”

Their claim seems to be based on a multinomial chi-square test described below in section 3.1. The assumptions of the multinomial chi-square test are false when the SITHP hypothesis is true. Moreover, not rejecting the null hypothesis does not imply that the true hypothesis is true: Failure to reject could be a Type II error, especially if the test has little power against plausible alternatives—which we show is the case.

No test has good power against every alternative, so we compare the multinomial chi-square test with several other tests of the Poisson hypothesis: conditional chi-square, Brown-Zhao, and Kolmogorov-Smirnov, described in sections 3.2, 3.3, and 3.4. Among these, the Kolmogorov-Smirnov test is most sensitive to long-term variations in the rate of seismicity. The multinomial chi-square (with $P$-value estimated by simulation rather than the chi-square approximation) is most sensitive to local departures from Poisson behavior—but Poisson behavior and constant rate are not the same thing, nor is “Poisson behavior” necessarily a good proxy for unpredictability. The conditional chi-square and Brown-Zhao tests are similar and are more sensitive to variations in the short-term rate of seismicity. An omnibus test that combines these four temporal tests using Bonferroni’s inequality rejects the hypothesis that the 1932–1971 Southern California Earthquake Center (SCSC) catalog of events is homogeneous Poisson process (SITHP). Tests of the null hypothesis that declustered catalogs are realizations of a SITHP on the spatiotemporal domain $S \times (0, T]$, the number of events in disjoint subsets of $S \times (0, T]$ are independent Poisson random variables. Within any subset, the locations of the events are independent of the times of events. The space-time rate is the product of the inhomogeneous marginal spatial rate and the uniform temporal rate.

Let $N$ denote the (random) number of events in a SITHP on $S \times (0, T]$. Denote the random locations and times of the $N$ events by $\{(X_i, Y_i, T_i)\}_{i=1}^N$. The times between successive events are marginally independent and identically distributed (iid) exponential random variables. The number of events in any subset of $S$ in disjoint subsets of $(0, T]$ are independent Poisson random variables with means proportional to the durations of the intervals; this is the basis of the multinomial chi-square test described in section 3.1. Conditional on $N = n$ (that is, given the number of events that actually occur), the times $\{T_i\}_{i=1}^n$ are (marginally) iid uniform random variables; this is the basis of the Kolmogorov-Smirnov test described in section 3.4. Since the conditional distribution of times given $N = n$ is iid uniform, the number of events in $K$ equal-length disjoint time intervals whose union is $(0, T]$ is has a multinomial conditional joint distribution with equal category probabilities. This is the basis of the conditional chi-square test and the Brown-Zhao test.

In describing the tests below, we assume that we are given a declustered catalog with $n$ events. The longitude, latitude, and time of the $i$th event are $(x_i, y_i, t_i)$, $i = 1, \ldots, n$. Events are not necessarily in chronological order. We do not consider earthquake depths. We study the null hypothesis that the points $\{(x_i, y_i, t_i)\}_{i=1}^n$ are a realization of a SITHP.

## 2 THE POISSON NULL HYPOTHESIS

Consider a fixed spatial domain $S$ and time interval $(0, T]$. In a spatially inhomogeneous, temporally heterogeneous Poisson process (SITHP) on the spatiotemporal domain $S \times (0, T]$, the number of events in disjoint subsets of $S \times (0, T]$ are independent Poisson random variables. Within any subset, the locations of the events are independent of the times of events. The space-time rate is the product of the inhomogeneous marginal spatial rate and the uniform temporal rate.

Let $N$ denote the (random) number of events in a SITHP on $S \times (0, T]$. Denote the random locations and times of the $N$ events by $\{(X_i, Y_i, T_i)\}_{i=1}^N$. The times between successive events are marginally independent and identically distributed (iid) exponential random variables. The number of events in any subset of $S$ in disjoint subsets of $(0, T]$ are independent Poisson random variables with means proportional to the durations of the intervals; this is the basis of the multinomial chi-square test described in section 3.1. Conditional on $N = n$ (that is, given the number of events that actually occur), the times $\{T_i\}_{i=1}^n$ are (marginally) iid uniform random variables; this is the basis of the Kolmogorov-Smirnov test described in section 3.4. Since the conditional distribution of times given $N = n$ is iid uniform, the number of events in $K$ equal-length disjoint time intervals whose union is $(0, T]$ is has a multinomial conditional joint distribution with equal category probabilities. This is the basis of the conditional chi-square test and the Brown-Zhao test.

In describing the tests below, we assume that we are given a declustered catalog with $n$ events. The longitude, latitude, and time of the $i$th event are $(x_i, y_i, t_i)$, $i = 1, \ldots, n$. Events are not necessarily in chronological order. We do not consider earthquake depths. We study the null hypothesis that the points $\{(x_i, y_i, t_i)\}_{i=1}^n$ are a realization of a SITHP.

## 3 TEMPORAL TESTS

### 3.1 The multinomial chi-square test (MC)

We believe that the chi-square test of the hypothesis that declustered catalogs are realizations of a homogeneous temporal Poisson process used by Gardner and Knopoff (1974) and Barani et al. (2007) was a multinomial chi-square test (Brown and Zhao 2002). It works as follows:

(i) Pick $K \geq 1$. Partition the study period into $K$ disjoint time intervals of length $T/K$. Count the events in each interval:

$$N_k \equiv \#\{i : t_i \in ((k-1)T/K, kT/K]\}, \quad k \in \{1, \ldots, K\}. \quad (1)$$

(ii) Estimate the theoretical rate of events per interval. We believe Gardner and Knopoff (1974) and Barani et al. (2007) used the estimate

$$\hat{\lambda} = n/K. \quad (2)$$
(iii) Pick \( C \geq 2 \), the number of “categories,” such that the expected number of intervals that fall into each category, assuming that events follow a Poisson process with rate \( \lambda \), is at least 5. That is, choose the smallest integer \( C \) such that
\[
E_c \geq 5 \quad \forall c \in \{0, \ldots, C - 1\},
\]
where
\[
E_c \equiv \begin{cases} 
K e^{-\lambda} \frac{\lambda^c}{c!}, & c = 0, 1, \ldots, C - 2 \\
K - \sum_{j=0}^{C-2} E_j, & c = C - 1. 
\end{cases} 
\]

For \( k \in \{1, \ldots, K\} \), interval \( k \) is in category \( c \in \{0, \ldots, C - 2\} \) if it contains \( c \) events; interval \( k \) is in category \( C - 1 \) if it contains \( C - 1 \) or more events. Let \( O_c \) be the number of intervals observed to be in category \( c \).

(iv) Calculate the chi-square statistic:
\[
\chi^2_m \equiv \sum_{c=0}^{C-1} \frac{(O_c - E_c)^2}{E_c}.
\]

Take the nominal \( P \)-value to be
\[
P \equiv \Pr \{ X \geq \chi^2_m \},
\]
where \( X \) is a random variable with a chi-square distribution with \( d \) degrees of freedom. We believe that Gardner and Knopoff (1974) used \( d = C - 2 \).

The nominal and true \( P \)-values depend on arbitrary choices: \( K, C, d \), and the method of estimating \( \lambda \). Moreover, the true \( P \)-value depends on whether these choices are made before or after looking at the data.

Let \( I_k = c \) if the number of events in the \( k \)th interval is in category \( c \). In the basic chi-square test for goodness of fit, \( C \) is fixed before observing the data, and the null hypothesis is that (i) \( \{\Pr \{ I_k = c \}\}_{c=0}^{C-1} \) are known and do not depend on \( k \), and (ii) \( \{I_k\}_{k=1}^{K} \) are independent (Lehmann 2005). The numbers of intervals in the \( C \) categories, \( \{O_c\}_{c=0}^{C-1} \), then have a multinomial joint distribution. The null distribution of the chi-square statistic converges to a chi-square distribution with \( C - 1 \) degrees of freedom as the number \( K \) of data increases—but the finite-sample distribution is only approximately chi-squared.

In testing whether declustered catalogs are Poisson, \( C \) generally is not fixed ahead of time: It is chosen after looking at the data to estimate \( \lambda \), for instance, so that the expected number of intervals in each category exceeds some minimum, such as 5. Moreover, neither (i) nor (ii) is true in testing whether declustered catalogs are Poisson. (i) is false because

the hypothesis that declustered seismicity is Poisson does not completely specify the category seismicity is Poisson does not completely specify the category probabilities \( \{\Pr \{ I_k = c \}\}_{c=0}^{C-1} \). Instead, those probabilities are estimated from an estimate \( \hat{\lambda} \) of the marginal temporal rate \( \lambda \) of the Poisson process. Estimating \( \{\Pr \{ I_k = c \}\}_{c=0}^{C-1} \) from the data changes the distribution of the chi-square statistic; moreover, the theoretical value of those probabilities conditional on the observed rate is different from the values used in practice, which are (estimated) unconditional probabilities.

(ii) is false too: Conditional on the estimated temporal rate, the random variables \( \{I_k\}_{k=1}^{K} \) are not independent because they are related through the total number of earthquakes—an ingredient in estimating the rate. For instance, if \( n \geq C - 1 \) and \( I_k = 0, k = 1, \ldots, K - 1 \), we would know that \( I_K = C - 1 \). The joint distribution of \( \{O_c\}_{c=0}^{C-1} \) is not multinomial when the Poisson null hypothesis is true, and the chi-squared statistic, as calculated to test declustered catalogs, may not have even approximately a chi-square distribution.

Hence, we calibrate the \( P \)-value for the multinomial chi-square test using a simulation that takes into account estimating the rate of events from the data, choosing \( C \) on the basis of that estimate, and calculating the category probabilities in an inconsistent way (using the observed number of events to estimate the probabilities, but ignoring that conditioning in computing the probabilities). The simulation conditions on the observed number of events, and takes the times of the events to be independent, identically distributed (iid) uniform random variables.

3.2 The conditional chi-square test (CC)

The MC test described in the previous subsection assesses whether the numbers of intervals with various numbers of events agree well with the numbers expected for iid uniformly distributed event times. MC uses \( C \) categories of possible values of the number of events per interval. An interval with \( C - 1 \) events is in the same category as one with \( C + 10 \) events. For this and other reasons, MC is not as sensitive to overdispersion—apparent fluctuations in the rate of seismicity—as some other tests.

The conditional chi-square test (or Poisson dispersion test) uses the fact that, conditional on the total number of events, the joint distribution of the numbers of events in the windows is multinomial with equal category probabilities. The test statistic is
\[
\chi^2_c \equiv \sum_{k=1}^{K} \frac{(N_k - \hat{\lambda})^2}{\hat{\lambda}}.
\]

This is proportional to the variance of the counts across windows. If the Poisson hypothesis is true, the distribution of \( \chi^2_c \) is approximately chi-square with \( K - 1 \) degrees of freedom. The conditional chi-square test involves choosing the number of intervals \( K \) but not \( C \), and, unlike the multinomial chi-square test, it uses the information \( N = n \) in a consistent way. While the multinomial chi-square test tries to look at the detailed distribution of the number of events per interval, the conditional chi-square test looks only at the variability of the observed number of events across intervals. High variability—overdispersion—is a sign that the process is not a homogeneous Poisson process.
3.3 The Brown-Zhao test (BZ)

Brown and Zhao (2002) proposed an alternative test. Let $Y_k \equiv \sqrt{N_k + 3/8}$ and $\bar{Y} \equiv \sum Y_k / K$. The Brown-Zhao (BZ) test statistic is

$$\chi^2_{BZ} \equiv 4 \sum_{k=1}^{K} (Y_k - \bar{Y})^2.$$  

Under the Poisson hypothesis, the statistic $\chi^2_{BZ}$ has a distribution that is approximately chi-square with $K - 1$ degrees of freedom. The chi-square approximation to the null distribution of $\chi^2_{BZ}$ tends to be better than the chi-square approximation to the null distribution of $\chi^2$ or $\chi^2_m$ (Brown and Zhao 2002). Like CC, the BZ test requires choosing $K$ but not $C$, and uses the information $N = n$ in a consistent way. Also like CC, the BZ test rejects when there is high variability of the observed numbers of events in different intervals—i.e., when the counts are dispersed.

3.4 The Kolmogorov-Smirnov test (KS)

The Kolmogorov-Smirnov test compares the empirical cumulative distribution function (cdf) $\hat{F}_n(x)$ of a random variable to a fixed reference cumulative distribution function $F(x)$ (Lehmann 2005). The KS test rejects when

$$D_n \equiv \sup_{x} |\hat{F}_n(x) - F(x)| \geq C(n, \alpha).$$  

In seismology, the KS test has been used to assess the uniformity of declustered earthquake sequences preceding mainshocks (Matthews and Reasenberg 1988; Reasenberg and Matthews 1988).

If declustered earthquakes follow a SITHP, then conditional on $N = n$, the times $\{T_i\}_{i=1}^{n}$ are iid uniform on $(0, T]$. Their common cumulative distribution function is $F(x) = t/T$. Hence, conditional on $N = n$,

$$D_n = \sup_{t} \left| \frac{1}{n} \sum_{i=1}^{n} 1_{t_i \leq t} - t/T \right|.$$  

Unlike the chi-square tests, the KS test has no ad hoc choices analogous to $K$, $C$, and $d$. The KS test has asymptotic power 1 against the alternative that the data are iid with any fixed distribution $G \neq F$.

3.5 Power

MC, CC, and BZ ignore the order of the $K$ intervals. This causes them to have low power against some kinds of clustering that violate the Poisson hypothesis, such as long-term variations in the rate of seismicity.

To see why, consider any particular catalog. Divide the study period into $K$ disjoint equal-length windows. Now, rearrange the windows so that the first has the most events, the second has the next most events, and so on, to form a new catalog. This catalog has a monotonically decreasing rate of seismicity, which would be very unlikely if declustered seismicity followed a homogeneous Poisson process. However, the test statistics for MC, CC, and BZ would have the same values for this rearranged catalog as they did for the original catalog.

In contrast, KS would tend to reject the null hypothesis for the new data: The empirical cdf would be far above $t/T$ in the early part of the rearranged catalog. KS is more sensitive to long-term rate variations than the other tests are, but less sensitive to short-term variations. KS, CC, and BZ are sensitive to whether the rate varies with time: to clustering. For instance, if events were equispaced in time, none of those tests would reject the null hypothesis. MC would reject the Poisson hypothesis if events were equispaced, given enough data.

None of these tests uses spatial information; in section 5 we propose a test that does. A process can be non-Poisson in space-time yet have a homogeneous Poisson marginal temporal distribution; see, e.g., Luen (2010, Chapter 3), so using spatial information can increase power, as we show empirically.

4 DECLUSTERING METHODS

We consider the following five window-based declustering algorithms:

- **GKl** (Gardner-Knopoff linked) Remove every event that is in the window of some other event (Gardner and Knopoff 1974).
- **GKlb** (Gardner-Knopoff linked, biggest) Divide the catalog into clusters as follows: An event is in a given cluster if and only if it is in the window of at least one other event in the cluster. In every cluster, remove all events except the largest (Gardner and Knopoff 1974).
- **GKm** (Gardner-Knopoff mainshock) Consider the events in chronological order. If the $i$th event is in the window of a preceding larger shock that has not already been deleted, delete it. If a larger shock is in the window of the $i$th event, delete the $i$th event. Otherwise, retain the $i$th event (Knopoff and Gardner 1972).
- **Rl** (Reasenberg linked) Reasenberg’s method (Reasenberg 1985).
- **dT deTest**, described below.

GKI, GKlb, and Rl are linked-window methods. Gardner and Knopoff (1974) found that GKI and GKlb gave similar results for 1932–1971 Southern California seismicity. GKm is a mainshock-window method.

deTest is not a window method. It has no physical basis, not even a heuristic one. It is offered as a “straw man” to show two things:

1. A declustered catalog can have rather more events than window-based declustering methods leave, and still pass a test for temporally homogeneous Poisson behavior.
2. Using spatial and temporal data by testing for conditional exchangeability of times given the locations (described below) can be more powerful than testing only for temporal homogeneity.

We assume that $K$ is given. Declustering a catalog to make the result pass the MC, CC, or BZ test is constrained by the number of intervals among the $K$ in the original catalog that have no events, since declustering can delete events but not add them. The number of intervals with no events gives an implicit estimate of the rate of a Poisson process that the declustered catalog can be coerced to fit well: If
seismicity followed a homogenous Poisson process with the theoretical rate $\lambda$ events per interval, the chance that an interval would contain no events is $e^{-\lambda}$. So,

$$\hat{\lambda} \equiv -\log \left( \frac{\# \text{ intervals with no events}}{K} \right)$$

(11)

is a natural estimate of the rate of events per interval. deTest tries to construct a catalog with about this rate that passes all four tests described above (MC, CC, BZ, and KS). If seismicity followed a homogenous Poisson process with the theoretical rate per interval $\lambda$, then the expected number of intervals in the declustered catalog with at least $c$ events would be

$$G_c \equiv \sum_{i=c}^{\infty} K \times \frac{\hat{\lambda}^i e^{-\hat{\lambda}}}{i!}.$$  

(12)

deTest constructs a catalog in which $[G_c]$ intervals contain $c$ or more events, where $[x]$ denotes the integer closest to $x$.

deTest is defined by the following algorithm, which starts with an empty catalog and adds events from the original catalog until the result has approximately the correct expected number of intervals with each number of events (ensuring it will pass the first three tests), then removes events until the catalog passes the KS test.

(i) Count the events in the raw catalog in each of the $K$ intervals.

(ii) Define $\hat{\lambda}$ by equation 11 and $G_c$ by equation 12.

(iii) Let $c = 1$. From each interval in the raw catalog that has at least one event, include one event selected at random from that interval.

(iv) Let $c \leftarrow c + 1$. If $[G_c] = 0$, go to step (vi). Otherwise, go to step (v).

(v) This step adds events to the declustered catalog until $[G_c]$ intervals have at least $c$ events, while trying to keep the KS statistic small. Let $N_t$ be the number of events in the current declustered catalog that have occurred by time $t$. Find the element $t_m$ of the set $t \in \{T/K, 2T/K, \ldots, T\}$ at which $N_t/N_T-1/t/T$ is minimized. Adding an event before time $t_m$ will tend to reduce the KS statistic. Find the set of intervals that

a. contain $c-1$ events in the current declustered catalog;
b. contain at least $c$ events in the raw catalog.

From this set, select the interval prior to $t_m$ but closest to $t_m$ (if no interval is prior to $t_m$, choose the first interval in the set.) Choose an event at random from the events in the selected interval that have not yet been added to the declustered catalog, and add that event to the declustered catalog. Repeat this step until $[G_c]$ intervals contain $c$ events. Return to step (iv).

(vi) Find the KS $P$-value. If it is above the target significance level, find a time $t$ at which the empirical cdf differs maximally from the uniform cdf. Either $t$ is infinitesimally before an event or $t$ is the time of one or more events. If $t$ is just before an event, $t/T$ is larger than the empirical cdf. In that case, delete the event after time $t$ at which the empirical cdf minus the uniform cdf is largest. If $t$ is the time of an event, the empirical cdf at $t$ is larger than $t/T$. In that case, delete an event at time $t$. Repeat this step until the KS $P$-value is below 0.05.

5 SPATIO-TEMPORAL TESTS

5.1 A weaker null hypothesis: conditionally exchangeable times

The marginal distribution of event times for a SITHP is Poisson, so if the hypothesis that declustered event times follow a Poisson distribution is rejected, so is the hypothesis that event times and locations follow a SITHP. Moreover, SITHPs can have events arbitrarily close together. But catalogs declustered with window methods have a minimum spacing between events: If a catalog contains two events very close in space and time, the later event will fall within the window of the former, and one or both of them will be deleted. However, catalogs declustered using window methods may still have some properties of SITHPs.

To try to salvage part of the SITHP hypothesis, we develop a test of a weaker condition implied by SITHP: the hypothesis that times are conditionally exchangeable given event locations. Let $\Pi$ be the set of all $n!$ permutations of $\{1,\ldots,n\}$. We say a process has conditionally exchangeable times if, conditional on the locations,

$$\{T_1,\ldots,T_n\} \overset{d}{=} \{T_{\pi(1)},\ldots,T_{\pi(n)}\}$$

(13)

for all permutations $\pi \in \Pi$. (The notation $\overset{d}{=} \pi X$ means “has the same probability distribution as.”) If event times are conditionally iid given event locations, they are conditionally exchangeable given event locations. Since event times in SITHPs are conditionally iid uniform given event locations, SITHPs have conditionally exchangeable times given event locations.

Under the hypothesis of conditionally exchangeable times, conditional on the set of locations $\{(x_i, y_i)\}_{i=1}^n$ and, separately, on the set of times $\{t_i\}_{i=1}^n$, all one-to-one assignments of times to locations have the same chance. If events close in space tend to be close in time—the kind of clustering real seismicity exhibits—times are not conditionally exchangeable. If events close in space tend to be distant in time—which can result from deleting events in windows—times are not conditionally exchangeable.

We test the hypothesis that times are conditionally exchangeable by adapting abstract methodology of Romano (1988, 1989). Let $P_{\pi}$ be the empirical distribution of the times and locations of the $n$ observed events: $P_{\pi}$ assigns probability $1/n$ to each observed (time, location) pair $(x_i, y_i, t_i)$. For each permutation $\pi$ of $\{1,\ldots,n\}$, let $P_{\pi n}$ be the distribution that assigns probability $1/n$ to each pair $(x_{\pi(i)}, y_{\pi(i)}, t_{\pi(i)})$.

If the null hypothesis of conditionally exchangeable times holds, then conditional on the times and (separately) the locations, the empirical distribution was just as likely to have been $P_{\pi n}$ as it was to be $P_n$. Consider a test statistic $\phi$ that can be computed from the empirical distribution of the data. Such a statistic is called a functional statistic. If the

---

\[ \text{\textsuperscript{†}} \] To ensure that there are at least $[G_c]$ intervals that contain $c$ events in principle require modifying the rule for deciding which intervals to include at each stage, so that things “telescope” correctly. In practice, we have not found it necessary to complicate the algorithm in that way.
null hypothesis is true, then conditional on the times and the locations, all values of $\phi(P_{e,n})$ as $\pi$ varies over the $n!$ permutations of $\{1, \ldots, n\}$ were equally likely. We can test hypotheses (conditional on the times and the locations) by determining whether $\phi(P_{e,n})$ is surprisingly large compared to those $n!$ values. If $\phi(P_{\pi,n}) \geq \phi(P_{e,n})$ for a fraction $P$ of the $n!$ permutations, then the $P$-value of the null hypothesis is $P$.

We now define the functional test statistic $\phi$ we will use. It measures the “distance” between the empirical distribution $P_\tau$ and a transformation $\tau P_\pi$ of the empirical distribution that satisfies the null hypothesis by construction. In particular, we take $\tau P_\pi$ to be the distribution that assigns probability $1/n^2$ to the $n^2$ pairs $((x_i, y_j), t_j)_{i,j=1}^n$. For $\tau P_\pi$, times and locations are independent, and hence times are conditionally exchangeable. Thus $\tau P_\pi$ satisfies the null hypothesis.

We now define the measure of “distance” that $\phi$ uses. A set $V \subset R^3$ is a lower-left quadrant if, for some $(x_0, y_0, t_0)$, it is of the form:

$$V = \{(x, y, t) \in R^3 : x \leq x_0 \text{ and } y \leq y_0 \text{ and } t \leq t_0\}.\ (14)$$

Let $V$ be the set of all lower-left quadrants. The test statistic $\phi$ is the supremum (over all lower-left quadrants $V \in \mathcal{V}$) of the difference between the probability $P_\tau$ assigns to $V$ and the probability that $\tau P_\pi$ assigns to $V$:

$$\phi(P_\tau) \equiv \sup_{V \in \mathcal{V}} |P_\tau(V) - (\tau P_\pi)(V)|.\ (15)$$

In principle, one could enumerate all $n!$ distributions $P_{\pi,n}$, calculate the $n!$ values of $\phi(P_{\pi,n})$, and find the fraction of such values that are greater than or equal to $\phi(P_{\tau})$ to determine the $P$-value. But since $n!$ is enormous, it is more practical to estimate the $P$-value by comparing $\phi(P_\pi)$ to the values of $\phi(P_{\pi,n})$ for a large random sample of permutations $\pi$. The larger the number of random permutations, the smaller the variability of the estimated $P$-value. Alternatively, it is still conservative simply to redefine the test to examine only a smaller, pre-determined subset of permutations, since in any such subset all the permutations are equally likely, conditional on the times and on the locations. The $P$-value for that test is the fraction of that subset of permutations for which the test statistic is at least as large as $\phi(P_\tau)$ in $n$.


## 6 DATA AND RESULTS

We assessed whether four subsets of the SCEC catalog‡ (1932–1971 and 1932–2010, each for $M \geq 3.8$ and $M \geq 4.0$), declustered using the five methods in section 4, are consistent with SITHP. The 1923–1971, $M \geq 3.8$ subset was chosen to be as similar as possible to the catalog used by Gardner and Knopoff (1974), for comparison with their results.

Gardner and Knopoff (1974) performed multinomial chi-square tests on a number of catalogs declustered using GKI. Among other things, they report results for a catalog of earthquakes with $M \geq 3.8$ occurring in the “Southern California Local Area” from 1932–1971. That raw catalog had 1,751 events; the declustered catalog had 503 events. They divided the forty-year period into ten-day intervals, found $O_c$ and estimated $E_c$ for some range of $c$, and found nominal $P$-values using the chi-square distribution with 2 degrees of freedom. They did not state $C$, how they estimated $\lambda$, nor whether they used $d = C - 1$ or $d = C - 2$ in their tests. They found a $P$-value of 0.0599, and hence did not reject the hypothesis that declustered catalogs are Poisson at significance level 0.05.

The SCEC catalog contains 1,556 events with magnitude at least 3.8 between 1932 and 1971. We declustered that catalog using GKI, GKIb, and GKM with the Gardner and Knopoff (1974) windows and using RI and dT. (We used Stefan Wiemer’s ZMAP package for MATLAB® to apply Reasenberg’s method.) The declustered catalogs contained 437, 424, 985, and 608 events, respectively. Figure 1 maps the events in the original catalog and the events that remain after declustering using each of the methods. Figure 2 shows the 1932–2010 SCEC catalog of 3,368 events of magnitude 3.8 and above, declustered using the five methods. Those declustered catalogs contained 913, 892, 1,120, 2,046, and 1,615 events, respectively.

We applied MC (using both the $\chi^2$ approximation and simulation to approximate the null distribution), CC, BZ, and KS to the declustered catalogs. We combined these four tests (using the simulation $P$-value rather than the $\chi^2$ approximation) to obtain a composite level 0.05 temporal test of the SITHP hypothesis, using Bonferroni’s equality: We rejected the SITHP hypothesis if any of these four tests gave a $P$-value less than 0.0125. If the null hypothesis is true, the chance of a Type I error is no greater than $4 \times 0.0125 = 0.05$.

Results, reported in Table 1, varied. For 1932–1971, $M \geq 3.8$, the catalog most similar to that studied by Gardner and Knopoff (1974), none of the window-declustered catalogs appears to be Poisson, contradicting Gardner and Knopoff (1974). For the four window declustering methods, the KS test rejects the Poisson hypothesis at level 0.0125; the KS test does not reject the Poisson hypothesis for deTest. The other tests reject the Poisson hypothesis at level 0.0125 for GKM and RI. For 1932–1971 and 1932–2010 $M \geq 4.0$, the Poisson hypothesis is rejected for GKM and RI. For 1932–2010 $M \geq 3.8$, the Poisson hypothesis is rejected for all methods except deTest.

Table 1 also gives results for the permutation test of the hypothesis that event times are conditionally exchangeable given event locations. As discussed above, this hypothesis is weaker than SITHP; nonetheless, incorporating spatial information can lead to more power to reject the SITHP hypothesis when that hypothesis is in fact false. This is evident in the results for deTest. deTest only tries to pass the temporal tests, which it succeeds in doing for all four cat-

---


**Figure 1.** (a): 1932–1971 SCEC catalog of 1,556 events of magnitude 3.8 or greater in Southern California. (b): The 437 events that remain after declustering using GKI. (c): The 424 events that remain after declustering using GKIb. (d): The 544 events that remain after declustering using GKn. (e): The 985 events that remain after declustering using RI. (f): The 608 events that remain after declustering using dT.

**Figure 2.** (a): 1932–2010 SCEC catalog of 3,368 events of magnitude 3.8 or greater in Southern California. (b): The 913 events that remain after declustering using GKI. (c): The 892 events that remain after declustering using GKIb. (d): The 1,120 events that remain after declustering using GKn. (e): The 2,046 events that remain after declustering using RI. (f): The 1,615 events that remain after declustering using dT.
alog analysis, despite the fact that it retains more events than all the other methods but Reasenberg’s. Unsurprisingly, it fails the spatio-temporal test for all four catalogs: The spatio-
temporal behavior of catalogs declustered by deTest is not
consistent with the hypothesis that times are conditionally
exchangeable. Of course, one could devise an analog of de-
Test to produce declustered catalogs that pass the permuta-
tion test; we have not tried.

7 DISCUSSION

Conclusions about whether declustered catalogs are consis-
tent with the SITHP hypothesis depend not only on the
declustering method but also on the catalog—and on the sta-
tistical test. The multinomial chi-square test commonly used
to assess whether declustered catalogs have Poisson tempo-
ral behavior relies on ad hoc tuning constants, lacks theoret-
ical justification, can have a significance level larger than its
nominal significance level, and has low power against many
plausible alternatives. Comparing the nominal \( P \)-value with
\( P \)-values (conditional on the number of events) estimated by
simulation using SCEC data (Table 1) shows that the \( \chi^2 \)
approximation to the \( P \)-value can be too low by at least 2.4% in
seismological applications. The multinomial chi-square test
is sensitive to departures from Poisson behavior within in-
tervals, but not to clustering per se. In particular, the multi-
nominal chi-square test discards information about the time
order of the intervals, which reduces its power to detect long-
term rate variations.

Compared with the multinomial chi-square test, the con-
titional chi-square test and the Brown-Zhao test have better theoretical justification and fewer ad hoc tuning con-
stants, but they still require an arbitrary choice of the num-
ber of intervals into which to divide the study period. Both
tests condition on the number of events; they test whether
the conditional distribution of times is iid uniform. Both are
sensitive to variation of the observed rate of events across
intervals—to clustering on the scale of the intervals. Al-
though it might be more powerful against some alternatives,
for the data we studied, the Brown-Zhao test never gave
a smaller \( P \)-value than the conditional chi-square test, to
which it is closely related. The primary advantage of the
Brown-Zhao test over the conditional chi-square test in this
application seems to be that the chi-square approximation
to the distribution of the test statistic is more accurate than
it is for the conditional chi-square test.

The Kolmogorov-Smirnov test of the Poisson hypothe-
sis also conditions on the number of events and tests whether
times are conditionally iid uniform. In contrast to the other
three temporal tests, it has no ad hoc tuning constants and—
because keeps the entire observation period intact—
has more power against long-term rate variations than the
other three tests, which divide the study period into shorter
intervals and ignore the temporal order of the shorter in-
tervals. This is evident in Table 1: \( P \)-values for the Kolmogorov-
Smirnov test are most often the smallest. However, the
tests are to some extent complementary: the chi-square tests
sometimes give small \( P \)-values when the Kolmogorov-
Smirnov test does not. The conditional chi-square test seems
preferable to the multinomial chi-square test in that it has
a firmer theoretical foundation and requires fewer ad hoc
choices, although it does not have power against some of
the same alternatives, for instance, periodic or nearly pe-
riodic seismicity. (Given enough data, the multinomial chi-
square test will reject the null hypothesis if events are nearly
equispaced, but the conditional chi-square test, the Brown-
Zhao test, and the Kolmogorov-Smirnov test will not.) Us-
ing the Kolmogorov-Smirnov test in conjunction with the
conditional chi-square test and combining the results using
Bonferroni’s inequality seems like a good compromise. That
is, if one wishes to test at significance level \( \alpha \), reject the null
hypothesis if either test has a \( P \)-value less than \( \alpha/2 \).

Such a composite test shows that 1932–1971 SCEC seis-
micity with \( M \geq 3.8 \), declustered using standard window
methods, is not consistent with the Poisson hypothesis. The
opposite conclusion by Gardner and Knopoff (1974) seems
to have resulted from their choice of tests: the multinomial
chi-square. It is surprising that the 1932–2010 SCEC data
declustered using Gardner-Knopoff windows is more consis-
tent with the Poisson hypothesis, since the Gardner-Knopoff
method was derived for the earlier data.

Moreover, it is hard to explain why increasing the
threshold magnitude from 3.8 to 4.0 makes as much dif-
ference as it does. It would be expected to increase \( P \)
values somewhat simply because it reduces sample size,
but that does not appear to be all that is at play: The
Kolmogorov-Smirnov test seems to reject the Poisson hy-
pothesis because the rate of small events is too low in the
earlier part of the catalog. This might be explained by cat-
alog incompleteness—that events of magnitude 3.8–4.0 are
more often missing from the earlier catalog—but according
to Hutton et al. (2010) the SCEC catalog has been essen-
tially complete above magnitude 3.25 from its earliest days.
The accuracy of magnitude and location estimates in the
early part of the catalog might contribute to the difference.

All four of these tests—multinomial chi-square, con-
tditional chi-square, Brown-Zhao, and Kolmogorov-Smirnov—
condition on the total number of events. None uses spatial
information, and it is the spatio-temporal distribution of
seismicity that matters. A test that uses spatial information
could be much more powerful against some alternatives. In
a spatially inhomogeneous, temporally homogeneous Pois-
son process (SITHP), two events may occur arbitrarily close
to one another with strictly positive probability. Catalogs
declustered using window methods can never have events
very close in space and time.

But declustered catalogs may still have properties in
common with SITHP. For instance, the times might be condi-
tionally exchangeable given the locations. As a special
case, knowing the location of an event might give no in-
formation about the time of the event. A novel permutation
test can be used to assess whether event times are condi-
tionally exchangeable given event locations. The power of
incorporating spatial information is evident in the fact that
catalogs declustered using deTest pass all the temporal tests,
but fail the spatio-temporal test for exchangeability.

“Ok, so why do you decluster the catalog?” asks the
online FAQ for the Earthquake Probability Mapping Appli-
Declustering and Poisson Tests

Table 1. P-values for tests of the null hypotheses that subsets of the 1932–2010 SCEC catalog of events, declustered using GK1, GKlb, GKm, Rl, and dT, have a homogeneous Poisson distribution in time or have a temporally homogeneous, spatially heterogeneous distribution in space and time. Column 1 gives the catalog year range. Column 2 is the magnitude threshold and the number of events before declustering. Column 3 is the declustering method. The number of events that remain after declustering is n. “χ²” is the nominal P-value for a multinomial chi-square test using the chi-square approximation to the null distribution of the test statistic. “Sim” is the P-value for a multinomial chi-square test estimated by simulation that includes conditioning on the observed number of events to estimate the rate of the process and to define the categories. “CC” is the P-value for the conditional chi-square test. “BZ” is the P-value for the Brown and Zhao (2002) test. Values in columns “Sim”, “CC” and “BZ” are estimated using 10⁵ simulated catalogs; sampling error in those estimated P-values is on the order of 0.16%. “KS” is the P-value for the Kolmogorov-Smirnov test that event times are iid uniform given the number of events. “Romano” is the permutation test for conditional exchangeability of times of events given their locations. Romano P is the P-value estimated from 1,000–10,000 simulations. Romano CI are confidence intervals for the P-value based on the number of simulations performed in each case. “Reject” is “Yes” for “Time” if the simulation P-value for any of the four temporal tests is less than 0.0125 (using the simulation P-value rather than the χ² P-value for the MC test). “Reject” is “Yes” for “Space-time” if the P-value for the Romano test is less than 0.05.

<table>
<thead>
<tr>
<th>Years (events)</th>
<th>Mag (events)</th>
<th>Meth</th>
<th>n</th>
<th>MC</th>
<th>CC</th>
<th>BZ</th>
<th>KS</th>
<th>Romano</th>
<th>CI</th>
<th>Time</th>
<th>Space-time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1932–1971</td>
<td>GK1</td>
<td>437</td>
<td>0.087</td>
<td>0.089</td>
<td>0.069</td>
<td>0.096</td>
<td>0.011</td>
<td>0.005</td>
<td>[0.003, 0.007]</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>GKlb</td>
<td>424</td>
<td>0.636</td>
<td>0.656</td>
<td>0.064</td>
<td>0.108</td>
<td>0.006</td>
<td>0.000</td>
<td>[0.000, 0.001]</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>3.8</td>
<td>GKm</td>
<td>544</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.021</td>
<td>0.069</td>
<td>[0.063, 0.076]</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>(1,556)</td>
<td>Rl</td>
<td>985</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>[0.000, 0.001]</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>1932–2010</td>
<td>dT</td>
<td>608</td>
<td>0.351</td>
<td>0.353</td>
<td>0.482</td>
<td>0.618</td>
<td>0.054</td>
<td>0.001</td>
<td>[0.000, 0.006]</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>GK1</td>
<td>296</td>
<td>0.809</td>
<td>0.824</td>
<td>0.304</td>
<td>0.344</td>
<td>0.562</td>
<td>0.348</td>
<td>[0.318, 0.378]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>GKlb</td>
<td>286</td>
<td>0.903</td>
<td>0.927</td>
<td>0.364</td>
<td>0.385</td>
<td>0.470</td>
<td>0.452</td>
<td>[0.421, 0.483]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4.0</td>
<td>GKm</td>
<td>369</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.000</td>
<td>0.540</td>
<td>0.504</td>
<td>0.504</td>
<td>[0.473, 0.535]</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>(1,047)</td>
<td>Rl</td>
<td>659</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>[0.000, 0.004]</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>dT</td>
<td>417</td>
<td>0.138</td>
<td>0.134</td>
<td>0.248</td>
<td>0.402</td>
<td>0.051</td>
<td>0.000</td>
<td>[0.000, 0.004]</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>GK1</td>
<td>913</td>
<td>0.815</td>
<td>0.817</td>
<td>0.080</td>
<td>0.197</td>
<td>0.111</td>
<td>0.214</td>
<td>[0.189, 0.241]</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>GKlb</td>
<td>892</td>
<td>0.855</td>
<td>0.855</td>
<td>0.141</td>
<td>0.204</td>
<td>0.005</td>
<td>0.256</td>
<td>[0.229, 0.284]</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>3.8</td>
<td>GKm</td>
<td>1120</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.032</td>
<td>0.006</td>
<td>0.006</td>
<td>[0.002, 0.013]</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>(3,368)</td>
<td>Rl</td>
<td>2046</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>[0.000, 0.004]</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>dT</td>
<td>1615</td>
<td>0.999</td>
<td>1.000</td>
<td>0.463</td>
<td>0.466</td>
<td>0.439</td>
<td>0.000</td>
<td>[0.000, 0.004]</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>GK1</td>
<td>606</td>
<td>0.419</td>
<td>0.421</td>
<td>0.347</td>
<td>0.529</td>
<td>0.528</td>
<td>0.247</td>
<td>[0.221, 0.275]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>GKlb</td>
<td>592</td>
<td>0.758</td>
<td>0.768</td>
<td>0.442</td>
<td>0.500</td>
<td>0.137</td>
<td>0.251</td>
<td>[0.224, 0.279]</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4.0</td>
<td>GKm</td>
<td>739</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.252</td>
<td>0.023</td>
<td>0.023</td>
<td>[0.015, 0.034]</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>(2,169)</td>
<td>Rl</td>
<td>1333</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>[0.000, 0.004]</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>dT</td>
<td>1049</td>
<td>0.995</td>
<td>0.999</td>
<td>0.463</td>
<td>0.465</td>
<td>0.340</td>
<td>0.001</td>
<td>[0.000, 0.006]</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The answers: “to get the best possible estimate of the rate of mainshocks,” and “the methodology of the Earthquake Probability Mapping Application] requires a catalog of independent events (Poisson model), and declustering helps to achieve independence.” The evidence presented here suggests that declustered catalogs generally do not consist solely of “independent events” that follow a Poisson model.

To estimate the rate of mainshocks presumes an unambiguous definition of “mainshock.” Often, mainshocks are taken to be the events that remain after a catalog is declustered—essentially a circular definition. And different methods will produce different declustered catalogs. Rather than try to identify mainshocks, it might be better to model all large events.

The most popular such model is the epidemic-type aftershock model (ETAS) (Ogata 1988, 1993; Ogata and Zhuang 2006), which includes mainshocks and aftershocks, which may themselves have aftershocks. ETAS models estimated from catalog data are often explosively non-stationary (Helmstetter and Sornette 2002). Some properties of real catalogs, such as the distribution of inter-event times, are inconsistent with fitted ETAS models (Luen 2010, Chapter 4). To our knowledge, ETAS modeling has not led to improved forecasts. We therefore question the utility of ETAS for modeling seismic risk, despite its intuitive appeal.

Regardless, the goal of declustering remains unclear to us. Is it to remove the “predictable” component of catalogs and leave only the “unpredictable”? Is it to make the residual catalog appear to be Poisson? Is it simply to thin clusters? Is it to make some algorithm perform better? These are different goals, none of which seems particularly well tied to the underlying physics. As the USGS FAQ also notes, large foreshocks and aftershocks can do just as much damage as mainshocks. Removing them from the catalog will neither avert nor repair the damage.

ACKNOWLEDGMENTS

We note with sadness that Leon Knopoff, a giant in geophysics who pioneered the methods assessed here, passed away this year. We are grateful to Steven N. Evans and Pe-


For a different approach, see Zaliapin et al. (2008).
ter M. Shearer for helpful conversations, and to anonymous referees for helpful comments.

APPENDIX A: ALGORITHM TO TEST THE HYPOTHESIS THAT TIMES ARE CONDITIONALLY EXCHANGEABLE

R code that implements the algorithm to test whether times are conditionally exchangeable is available at http://statistics.berkeley.edu/~stark/Code/Quake/permutest.r. The algorithm has the following steps:

(i) Sort the catalog of longitudes, latitudes, and times in time order. Label the sorted points \( \{x_i, y_i, t_i\} \) for \( i \in \{1, \ldots, n\} \). Find the longitude and latitude ranks of every event.

(ii) Find the empirical spatial measure of all lower-left quadrants in \( \mathbb{R}^2 \) with corners

\[
(y_i, x_j), \quad 1 \leq i, j \leq n.
\]

In the online R code, this spatial distribution is stored in the matrix \( xy.upper \). The entry indexed by \( (i, j) \) is

\[
\frac{1}{n} \sum_{i=1}^{n} 1(y_i \leq y, x_j \leq x).
\]

This is the number of events in the catalog with latitude less than the latitude of the \( i \)th event in the catalog and with longitude less than the longitude of the \( j \)th event in the catalog.

(iii) Find the absolute differences between the empirical measure and the empirical null measure for the \( n^3 \) lower-left quadrants with corners

\[
(x_j, y_i, t_k), \quad i, j, k \in \{1, \ldots, n\}.
\]

Find the maximum value of all these differences; this is the test statistic \( \phi \). To reduce storage requirements, the code finds the distances for quadrants with corners \( (x_j, y_i, t_k) \) for every value of \( k \) successively; that is, it finds

\[
\phi = \max_k \left[ \max_{j,i} \left( P(V(j, i, k)) - (\tau \hat{P})(V(j, i, k)) \right) \right],
\]

where \( V(j, i, k) \) is the lower-left quadrant with corner \( (x_j, y_i, t_k) \).

(iv) Set the iteration counter \( h \) to 0.

(v) Increment \( h \). Create a random permutation of \( \{1, \ldots, n\} \). Apply this permutation to the locations, leaving times fixed. (One could permute the times instead of the locations, but that would require the re-sorting the catalog into temporal order after each permutation.) The spatial measure has not changed, but its indexing has; apply the permutation to both the rows and the columns of \( xy.upper \).

(vi) As in step (iii), find the absolute differences between the empirical measure and the empirical null measure of the \( n^3 \) lower-left quadrants. Let \( \phi_h \) be the maximum value of all these distances.

(vii) Determine whether to stop or to return to step (v). (We might simply stop when \( h = 10,000 \), or we might apply Wald’s sequential probability ratio test (Wald 1945) to determine whether, on the basis of the random permutations taken so far, it is possible to conclude whether \( P \leq \alpha \).) If the algorithm stops, estimate the \( P \)-value as

\[
P = \frac{1}{H} \# \{h : \phi_h \geq \phi\},
\]

where \( H \) is the total number of iterations.

REFERENCES


