Multiscale Estimation of Event Arrival Times and Their Uncertainties in Hydroacoustic Records from Autonomous Oceanic Floats

Joel D. Simon*[1], Frederik J. Simons[1], and Guust Nolet[1,2]

ABSTRACT

We describe an algorithm to pick event onsets in noisy records, characterize their error distributions, and derive confidence intervals on their timing. Our method is based on an Akaike information criterion that identifies the partition of a time series into a noise and a signal segment that maximizes the signal-to-noise ratio. The distinctive feature of our approach lies in the timing uncertainty analysis, and in its application in the time domain and in the wavelet timescale domain. Our novel data are records collected by freely floating Mobile Earthquake Recording in Marine Areas by Independent Divers (MERMAID) instruments, midcolumn hydrophones that report triggered segments of ocean-acoustic time series.

KEY POINTS

- We develop a method to rapidly assign uncertainty estimates to wavelet-multiscale arrival-time picks.
- Our algorithm is able to identify and assess the quality of multiple phase arrivals in noisy seismograms.
- Onset picks and uncertainties made on hydroacoustic drifter records open up the oceans for global tomography.

Supplemental Material

INTRODUCTION

We wish to pick multiscale seismic arrival times and estimate their uncertainties in noisy hydroacoustic records. Our approach to onset detection centers on the evaluation of an Akaike information criterion (AIC) function at multiple scales of a wavelet-decomposed time series. In the most general sense, our procedure identifies the changepoint in the time series where the discrepancy between the segments to the left (the "noise") and the right (the "signal") is maximized. We present two changepoint estimates derived from the AIC function, two methods to estimate their uncertainties, and investigate their application in both the time and timescale domains.

Laying out preliminaries and developing a simple signal model in the General Considerations section, we define the AIC in The AIC section. Next, we formalize a general scheme for uncertainty analysis in the Formalizing the AIC Timing Uncertainty section. In the Multiscale Analysis Methodology section, we extend our procedures to the multiresolution domain via wavelet decomposition of the time series. There, we detail the mapping between the time and timescale domains and apply it in the Multiscale Analysis in Practice section to restate our general problem of changepoint estimation in both domains. By The Data Set section, we have exhausted the theory and introduce the Mobile Earthquake Recording in Marine Areas by Independent Divers (MERMAID) instrument and its data in preparation to apply our method in the Application to MERMAID Seismograms section. There, we analyze 445 publicly available MERMAID seismograms and use the algorithm we have developed to identify arrival times and estimate their uncertainties. We use a public catalog of associated earthquakes to compute their travel-time residuals with respect to the ak135 velocity model. We summarize the statistics of the multiscale residuals in the Distribution of Travel-Time Residuals section and introduce our updated events catalog in The Updated Events Catalog section, with all identified seismic phases, travel-time residuals, and timing uncertainties in the data set. In the supplemental material we remark on the general utility of our changepoint and uncertainty estimation procedures beyond the scope of this study.

The catalog available in the supplemental material to this article provides the foundation for future tomographic studies that will use MERMAID data. Our uncertainty estimates will serve as measures for data weighting in the inversion. The software developed for this work is freely and publicly available.

1. Department of Geosciences, Princeton University, Princeton, New Jersey, U.S.A.; 2. Université Côte d'Azur/CNRS/OCA/IRD, GéoAzur, Sophia Antipolis, France

*Corresponding author: jdsimon@princeton.edu

Cite this article as Simon, J. D., F. J. Simons, and G. Nolet (2020). Multiscale Estimation of Event Arrival Times and Their Uncertainties in Hydroacoustic Records from Autonomous Oceanic Floats, *Bull. Seismol. Soc. Am.* **XX**, 1–28, doi: 10.1785/0120190173

© Seismological Society of America

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

GENERAL CONSIDERATIONS

We begin by posing the task of identifying the signal from the noise as a problem of statistical inference via likelihood analysis.

A simple model of noise and signal

To start with a very simple description, we model the seismogram, $x(k)$, in which $k = [1, \ldots, N]$, as the concatenation of two separate and distinct segments joined at sample index $k_\circ$. We label the first segment noise, $n(k)$, and the second segment signal, $s(k)$:

$$x(k) = \begin{cases} n(k), & \text{if } 1 \le k \le k_\circ, \\ s(k), & \text{if } k_\circ + 1 \le k \le N. \end{cases} \tag{1}$$

A true seismogram, of course, contains some level of noise throughout the interval under consideration, although, as a starting point to describe our automatic arrival identifier, this simplified model suffices. Our goal is to find $k_\circ$, the change-point, which "best" separates the noise segment from the signal segment.

We further assume that both the noise and the signal segments are samples from Gaussian (normal) parent distributions with different population parameters. Individually, they contain independently and identically distributed (i.i.d.) data, and they are mutually uncorrelated. Aware of the oversimplification, we stick to it for the time being. We formulate the problem of finding $k_\circ$ in the context of maximum-likelihood estimation (MLE), specifically finding the sample index at which the time series is split into two segments that are most likely individually i.i.d.

Labeling the individual sample indexes in the time series as x_i, with $i = [1, \ldots, N]$, in our description the probability density of any such individual point is

$$f(x_i; \mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left[\frac{-(x_i - \mu)^2}{2\sigma^2}\right], \tag{2}$$

the expectation, μ, and variance, σ^2, of which will be identified as μ_1, μ_2, and σ_1^2, σ_2^2, depending on whether $i \le k_\circ$ or $i > k_\circ$, respectively. The population parameter sets are $\theta_1 = \{\mu_1, \sigma_1^2\}$ (for the noise) and $\theta_2 = \{\mu_2, \sigma_2^2\}$ (for the signal).

The likelihood of an initial portion of the time series $x_1, \ldots, x_k$, in which $k = [1, \ldots, N]$, being derived from the noise parent distribution is

$$\mathcal{L}_1(\mu_1, \sigma_1^2; x_1, \ldots, x_k) = \prod_{i=1}^{k} f(x_i; \mu_1, \sigma_1^2) = (2\pi\sigma_1^2)^{\frac{-k}{2}} \exp\left[\frac{-\sum_{i=1}^{k}(x_i - \mu_1)^2}{2\sigma_1^2}\right], \tag{3}$$

and the log-likelihood function is

$$\ell_1(\theta_1; x_1, \ldots, x_k) = -\frac{1}{2}\left[k\ln(2\pi) + k\ln(\sigma_1^2) + \frac{1}{\sigma_1^2}\sum_{i=1}^{k}(x_i - \mu_1)^2\right]. \tag{4}$$

The maximum-likelihood estimate (MLE), $\hat{\theta}_1 = \{\hat{\mu}_1, \hat{\sigma}_1^2\}$, solves

$$\frac{\partial \ell_1(\hat{\theta}_1; x_1, \ldots, x_k)}{\partial \theta_1} = 0, \tag{5}$$

which takes the well-known form

$$\hat{\mu}_1 = \frac{1}{k}\sum_{i=1}^{k} x_i, \tag{6}$$

$$\hat{\sigma}_1^2 = \frac{1}{k}\sum_{i=1}^{k}(x_i - \hat{\mu}_1)^2. \tag{7}$$

Substituting the expressions (6) and (7) into equation (4) yields

$$\ell_1(\hat{\theta}_1; x_1, \ldots, x_k) = -\frac{k}{2}[\ln(2\pi) + \ln(\hat{\sigma}_1^2) + 1]. \tag{8}$$

Similarly, the likelihood of the remaining portion of the time series $x_{k+1}, \ldots, x_N$, in which $k = [1, \ldots, N]$, being drawn from the signal parent distribution is

$$\mathcal{L}_2(\mu_2, \sigma_2^2; x_{k+1}, \ldots, x_N) = \prod_{i=k+1}^{N} f(x_i; \mu_2, \sigma_2^2) = (2\pi\sigma_2^2)^{\frac{-(N-k)}{2}} \exp\left[\frac{-\sum_{i=k+1}^{N}(x_i - \mu_2)^2}{2\sigma_2^2}\right], \tag{9}$$

and its log-likelihood function is

$$\ell_2(\theta_2; x_{k+1}, \ldots, x_N) = -\frac{1}{2}\left[(N-k)\ln(2\pi) + (N-k)\ln(\sigma_2^2) + \frac{1}{\sigma_2^2}\sum_{i=k+1}^{N}(x_i - \mu_2)^2\right]. \tag{10}$$

The equivalents to equations (6) and (7) are the elements of $\hat{\theta}_2$:

$$\hat{\mu}_2 = \frac{1}{(N-k)}\sum_{i=k+1}^{N} x_i, \tag{11}$$

$$\hat{\sigma}_2^2 = \frac{1}{(N-k)}\sum_{i=k+1}^{N}(x_i - \hat{\mu}_2)^2, \tag{12}$$

and substitution of equations (11) and (12) into (10) yields

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

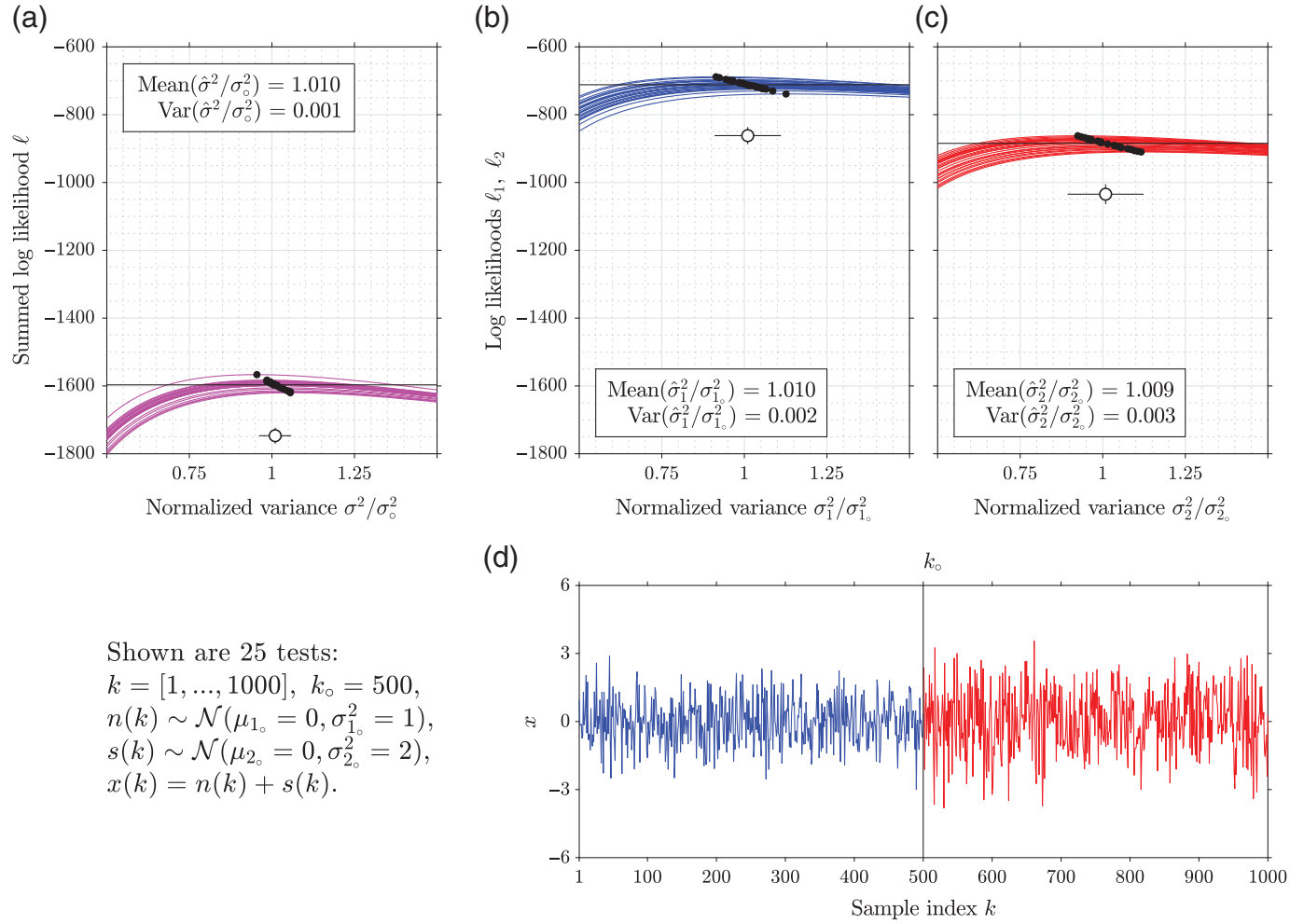


Figure 1. Maximum-likelihood estimation (MLE) of the sample variance of synthetic realizations of time series $x(k) = n(k) + s(k)$ that follow the simple model in equation (1), assuming uncorrelated Gaussian distributions with a different variance for each segment. (a) 25 examples of the sum of the log likelihoods $\ell = \ell_1 + \ell_2$ from equation (14), and (b,c) the log likelihoods ℓ_1 from equation (4), ℓ_2 from equation (10), rendered in the color corresponding to the noise (blue) or signal segment (red) of the synthetic time series, one realization of which is displayed in (d). The true variances $\sigma^2_{1_\circ} = 1$ and $\sigma^2_{2_\circ} = 2$ serve as normalization on the abscissa axes. The population expectations are $\mu_{1_\circ} = \mu_{2_\circ} = 0$. The changepoint between the noise and signal segments occurs at $k_\circ = 500$, the halfway point in each of these length $N = 1000$ simulations (which turns the color for the summed log likelihoods into magenta, the sum of an even mixture of blue and red). Black-filled circles mark the MLE of the variances $\hat{\sigma}^2_1$ and $\hat{\sigma}^2_2$, and their averages are shown by white-filled circles, offset vertically for clarity from their average log likelihoods marked by the horizontal lines, with whiskers extending two standard deviations in both directions in both coordinates.

$$\ell_2(\hat{\theta}_2; x_{k+1}, \ldots, x_N) = -\frac{(N-k)}{2}[\ln(2\pi) + \ln(\hat{\sigma}^2_2) + 1]. \quad (13)$$

The sum of the logarithmic likelihoods (4) and (10) is

$$\ell(\Theta; x; k) = \ell_1(\theta_1; x_1, \ldots, x_k) + \ell_2(\theta_2; x_{k+1}, \ldots, x_N), \quad (14)$$

denoting $\Theta = \{\theta_1, \theta_2\} = \{\mu_1, \sigma^2_1, \mu_2, \sigma^2_2\}$. Likewise, we sum the evaluated log likelihoods in equations (8) and (13) to

$$\begin{aligned}\ell(\hat{\Theta}; x; k) &= \ell_1(\hat{\theta}_1; x_1, \ldots, x_k) + \ell_2(\hat{\theta}_2; x_{k+1}, \ldots, x_N) \\ &= -\frac{1}{2}[k\ln(\hat{\sigma}^2_1) + (N-k)\ln(\hat{\sigma}^2_2) + \mathrm{C}], \quad (15)\end{aligned}$$

in which $\hat{\Theta} = \{\hat{\theta}_1, \hat{\theta}_2\} = \{\hat{\mu}_1, \hat{\sigma}^2_1, \hat{\mu}_2, \hat{\sigma}^2_2\}$, and $\mathrm{C} = N[\ln(2\pi) + 1]$.

Maximum-likelihood changepoint estimation

In Figure 1, we inspect the behavior of the log likelihoods (4), (10), and (14) evaluated at different parameter values $\Theta = \{\theta_1, \theta_2\}$, for synthetically generated data sets $x = \{x_1, \ldots, x_{k_\circ}; x_{k_\circ+1}, \ldots, x_N\}$, in other words, for $x = \{n, s\}$ drawn from known Gaussian parent distributions $\mathcal{N}(\Theta_\circ)$ with specific true population parameters $\Theta_\circ = \{\theta_{1_\circ}, \theta_{2_\circ}\}$. In this figure, we hold $k = k_\circ$ fixed and consider it to be known. Similarly, we fix the expectations to zero. Figure 1d contains one such realization, color-coded blue for the noise segment and red for the signal segment. In the same colors, Figure 1b and 1c show their log likelihoods $\ell_1(\theta_1; x_1, \ldots, x_{k_\circ})$ and $\ell_2(\theta_2; x_{k_\circ+1}, \ldots, x_N)$, respectively,

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

as a function of the normalized-variance parameters $\sigma_1^2/\sigma_{1_\circ}^2$ and $\sigma_2^2/\sigma_{2_\circ}^2$, that is, normalized by the true variances. The MLEs $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ and their log likelihoods, $\ell_1(\hat{\theta}_1; x_1, \ldots, x_{k_\circ})$ and $\ell_2(\hat{\theta}_2; x_{k_\circ+1}, \ldots, x_N)$, respectively, are marked by black-filled circles, the average of which over all the 25 trials in this example is rendered as a white-filled circle at the correct location on the abscissa axes, and offset from their average log likelihood, marked by the horizontal line, by an arbitrary amount for clarity. Whiskers extend two standard deviations in both directions on both axes. Figure 1a shows the summed log likelihoods $\ell(\Theta; x; k_\circ)$. There we use normalized-variance-sum coordinates denoted by $\sigma^2/\sigma_\circ^2$ such that the summed log-likelihood value at $\sigma^2/\sigma_\circ^2 = 1$ is the sum of the log-likelihood values of Figure 1b at $\sigma_1^2/\sigma_{1_\circ}^2 = 1$ and Figure 1c at $\sigma_2^2/\sigma_{2_\circ}^2 = 1$. Similar to Figure 1b,c, the black-filled circles correspond to the sums of the MLEs, the white-filled circle their average over the 25 trials in this experiment, and the horizontal black line the average of those summed log likelihoods. As expected, the maximum-likelihood variance estimates are on average very close to the truth; that is, given a range of parameter values to test (the normalized abscissa axes in Fig. 1a–c), the MLE, $\hat{\Theta}$, of the population parameters, Θ, converges to the truth $\Theta_\circ$.

In Figure 2, we investigate the effect on the likelihoods and their ratios of varying k, holding $k_\circ$ fixed as an unknown truth. We hold all the population parameters unchanged from the cases presented in Figure 1. What varies is the sample at which we split the time series into an assumed "noise" and an assumed "signal" segment. Figure 2e,f contains two realizations of our same process, but now we color blue the first segment, until the index k marked by the dotted line, which we may consider to be "noise," and we color red the second segment, which we may consider to be "signal." Figure 2e is an example of where our identification is tardy, $k > k_\circ$, and Figure 2f is an example of where our identification is early, $k < k_\circ$. As in Figure 1, the true noise segment contains sample indexes $k = [1, \ldots, 500]$, and the signal $k = [501, \ldots, 1000]$. Corresponding to Figure 2e, and for 25 such identical experiments, we show the log likelihoods of the misidentified "noise" segments as the blue curves in Figure 2b, and, corresponding to Figure 2f, we show the log likelihoods of the misidentified "signal" segments as the red curves in Figure 2c. We are not showing the log likelihoods of the corresponding red signal segments in Figure 2e, nor of the blue noise segments in Figure 2f, because those do not consist of wrongly mixed models, and their shapes are similar to those plotted in Figure 1c,b.

The black-filled circles on each log-likelihood curve in Figure 2a–d mark the MLEs of the variances and their corresponding log likelihoods. The horizontal black line marks the average likelihood value of the MLEs of the variances for the 25 trials shown, whereas the white-filled circle, arbitrarily offset on the ordinate axis, marks the average MLE, with whiskers extending two standard deviations in both directions on both axes.

Figure 2a contains the summed log likelihoods of the data segmentation of Figure 2e, combining the log likelihoods shown in Figure 2b with those of the trailing segment (not shown for the reason stated earlier), and again rendered in the appropriately mixed colors. Figure 2d contains the summed log likelihoods of the data segmentation plotted in Figure 2f, combining the log likelihoods of Figure 2c with those of the leading segment (not shown).

As in Figure 1, each log-likelihood curve is plotted on an abscissa axis that is normalized with respect to the true population parameters. For Figure 2a,d, as for Figure 1a, such an axis normalization amounts to summing along a 1:1 diagonal section through a 2D likelihood surface parameterized in those normalized coordinates. Such a construct is theoretical: in the Expected Behavior section, we discuss its expected behavior, and in The Summed Log Likelihood section how to sum the likelihoods appropriately in a way that is diagnostic for model identification.

Expected behavior

When the estimated changepoint is late, $k > k_\circ$, the expectation of the sample variance of the first mixed segment of x (the blue "noise" portion in Fig. 2e), is

$$\mathbb{E}[\hat{\sigma}_1^2]_{k>k_\circ} = \frac{1}{k}[k_\circ \sigma_{1_\circ}^2 + (k - k_\circ)\sigma_{2_\circ}^2]. \tag{16}$$

The expectation of the sample variance of the remaining abbreviated signal segment is the trivial identity

$$\mathbb{E}[\hat{\sigma}_2^2]_{k>k_\circ} = \sigma_{2_\circ}^2. \tag{17}$$

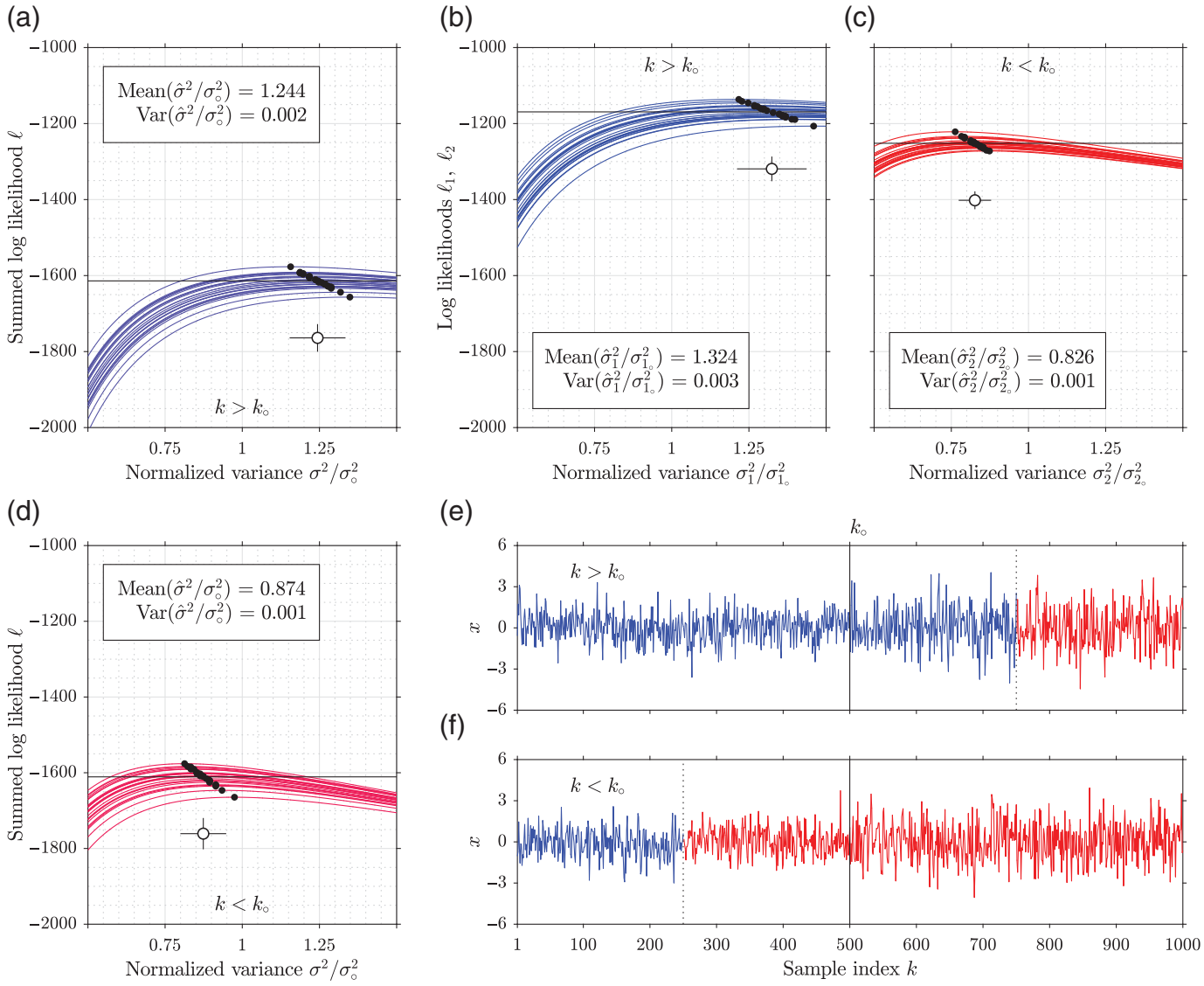
When $k > k_\circ$, the expectation of the sample variance of the mixed "noise" segment, normalized in terms of the population variance of the true noise segment, is

$$\mathbb{E}\left[\frac{\hat{\sigma}_1^2}{\sigma_{1_\circ}^2}\right]_{k>k_\circ} = \frac{1}{k}\left[k_\circ + (k - k_\circ)\frac{\sigma_{2_\circ}^2}{\sigma_{1_\circ}^2}\right], \tag{18}$$

which evaluates to $1.\overline{333}$ in the case shown in Figure 2b, in which $\sigma_{1_\circ}^2 = 1$, $\sigma_{2_\circ}^2 = 2$, and $k = 750$. The mean value over 25 tests, 1.324, is marked by the white-filled circle in Figure 2b, implying suitable convergence. The corresponding log likelihood at this incorrect candidate changepoint is given by equation (8), and is approximately equal to −1172, which is close to the value marked by the horizontal line in Figure 2b. The expectation of the sample variance of the remaining abbreviated signal segment (the red signal portion in Figure 2e, whose likelihoods are not shown) is uninteresting because it simply equals the true population variance (in which $\sigma_{2_\circ}^2 = 2$ and $\mathbb{E}[\hat{\sigma}_2^2/\sigma_{2_\circ}^2]_{k>k_\circ} = 1$). Its corresponding log likelihood is given by equation (13), and is approximately equal to −441.

Similarly, when the estimated changepoint is early, $k < k_\circ$, the expectation of the sample variance of the mixed second segment of x (the red "signal" portion in Fig. 2f), is

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet



$$\mathbb{E}[\hat{\sigma}_2^2]_{k<k_\circ} = \frac{1}{(N-k)}[(k_\circ - k)\sigma_{1_\circ}^2 + (N-k_\circ)\sigma_{2_\circ}^2]. \tag{19}$$

The expectation of the sample variance of the shortened noise segment is again trivial

$$\mathbb{E}[\hat{\sigma}_1^2]_{k<k_\circ} = \sigma_{1_\circ}^2. \tag{20}$$

When $k < k_\circ$, the expectation of the sample variance of the mixed "signal" segment, normalized in terms of the population variance of the true signal segment, is given by

$$\mathbb{E}\left[\frac{\hat{\sigma}_2^2}{\sigma_{2_\circ}^2}\right]_{k<k_\circ} = \frac{1}{(N-k)}\left[(k_\circ - k)\frac{\sigma_{1_\circ}^2}{\sigma_{2_\circ}^2} + (N-k_\circ)\right], \tag{21}$$

which evaluates to $0.8\overline{33}$ in the case shown in Figure 2c, in which $\sigma_{1_\circ}^2 = 1$, $\sigma_{2_\circ}^2 = 2$, and $k = 250$. The mean value over 25 tests, 0.826, is marked by the white-filled circle in Figure 2c,

Figure 2. MLE of the sample variance of synthetic time series that follow the same simple model as in Figure 1 but for the case of an unknown changepoint. Two realizations are shown, and the cases illustrated have $k_\circ = 500$, marked by solid vertical lines, but pertain to (e) a tardy ($k = 750$), and (f) an early ($k = 250$) assumed changepoint, marked by the dotted vertical lines. The log likelihood that applies to the blue section (which includes the full noise segment and a portion of the signal) in (e) is shown in (b), and constitutes one term in the sum, the first term in equation (14), shown in (a). The log likelihood of the corresponding red section (which only includes signal) of (e), the second term in the sum in equation (14), is not shown separately. Similarly, the log likelihood of the red section (some of the noise segment and all of the signal) of (f) is shown in (c). Again, the log likelihood of the corresponding blue section (containing only noise) of (f) is not shown. (a) Plots of the summed log likelihoods for the time series in (e), with the color revealing the relative amount of signal mixed in with "noise." (d) The equivalent summed log likelihoods for the time series in (f), with the color betraying the relative amount of noise mixed in with the "signal." As in Figure 1, MLEs are marked by black-filled circles and their average by white-filled circles with whiskers extending two standard deviations in both directions in both coordinates, offset vertically for clarity from the black horizontal lines that mark the average log likelihoods of the MLEs.

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

implying suitable convergence as in Figure 2b. The log likelihood associated with this changepoint is given by equation (13), and is approximately equal to −1256. Like the abbreviated signal section of Figure 2e, the abbreviated noise segment of Figure 2f (in blue, the likelihoods of which are not shown) contains no improperly mixed sample indexes and thus the expectation of its sample variance is equal to the corresponding true population variance ($\sigma^2_{1_\circ} = 1$; $\mathbb{E}[\hat{\sigma}^2_1/\sigma^2_{1_\circ}]_{k<k_\circ} = 1$). The corresponding log likelihood is given by equation (8) and is approximately equal to −355.

In the jointly normalized abscissa coordinates of Figures 1a, and 2a,d, the location of the maximizer along the diagonal of the summed log likelihoods lies at a linear mixture of the sample variances of the misidentified segments proportional to the length of the segments over which they are assumed to apply:

$$\frac{\hat{\sigma}^2}{\sigma^2_\circ} = \frac{1}{N}\left[\frac{1}{\sigma^2_{1_\circ}} k\hat{\sigma}^2_1 + \frac{1}{\sigma^2_{2_\circ}}(N-k)\hat{\sigma}^2_2\right]. \tag{22}$$

Substituting equations (16) and (17) into the generalized summed log-likelihood expression of equation (22) yields

$$\mathbb{E}\left[\frac{\hat{\sigma}^2}{\sigma^2_\circ}\right]_{k>k_\circ} = 1 + \frac{1}{N}\left[(k-k_\circ)\left(\frac{\sigma^2_{2_\circ}}{\sigma^2_{1_\circ}} - 1\right)\right]. \tag{23}$$

Substituting equations (19) and (20) into the generalized summed log-likelihood expression of equation (22) yields

$$\mathbb{E}\left[\frac{\hat{\sigma}^2}{\sigma^2_\circ}\right]_{k<k_\circ} = 1 + \frac{1}{N}\left[(k_\circ-k)\left(\frac{\sigma^2_{1_\circ}}{\sigma^2_{2_\circ}} - 1\right)\right]. \tag{24}$$

In Figure 2a, $k = 750$ and equation (22) evaluates to 1.250. In Figure 2d, $k = 250$ and equation (22) evaluates to 0.875. These are indeed the expected linear sums of the sample variances, normalized proportionally to the true population variances, for both Figure 2e,f, to which the experiments converged, as shown by the white-filled circles in Figure 2a,d.

The summed log likelihood

The summed log likelihoods shown in Figures 1d and 2a,d are diagonal profiles through a 2D surface in two ratio variables (they are normalized variances). No estimation method is expected to hug this diagonal unless $\sigma^2_{2_\circ}/\sigma^2_{1_\circ}$, the true signal-to-noise ratio (SNR), is unity. Thus, a general method would need to construct a 2D summed log-likelihood surface for a suitably large cross-product space of ratios in either variable, for a given candidate changepoint k, find and record the maximum of the summed log likelihood, and then repeat the process at every new candidate changepoint. See Figure 3 for a graphical illustration at three different candidate changepoints.

A procedure properly diagnostic of the true changepoint $k_\circ$ would recover both it and the variances of the noise and signal segments, $\sigma^2_{1_\circ}$ and $\sigma^2_{2_\circ}$. A profile through the 3D likelihood volume (in the parameters k, σ^2_1, and σ^2_2) does indeed peak at the correct triplet of true values $k_\circ$, $\sigma^2_{1_\circ}$, and $\sigma^2_{2_\circ}$. This behavior is borne out by Figure 4, which portends the success of a method that simply takes $\hat{\Theta}$, containing the sample means $\hat{\mu}_1$ and $\hat{\mu}_2$ and variances $\hat{\sigma}^2_1$ and $\hat{\sigma}^2_2$, of the segmentation of the time series x, calculates the summed log likelihood $\ell(\hat{\Theta}; x; k)$ according to equation (15), and defines the MLE for the changepoint as the argument that maximizes $\ell(\hat{\Theta}; x; k)$ over k.

Figure 4 plots those evaluated summed log likelihoods at the sample indexes $k = [1, \ldots, N]$, both in expectation (colored curve, substituting $\hat{\Theta} = \Theta_\circ$), and for 25 realizations of x (gray curves, when in general, $\hat{\Theta} \neq \Theta_\circ$) and their mean (black curve). As in Figures 1d and 2e,f, and for all tests here, $N = 1000$, $k_\circ = 500$, and $\Theta_\circ = \{\mu_{1_\circ}, \sigma^2_{1_\circ}, \mu_{2_\circ}, \sigma^2_{2_\circ}\} = \{0, 1, 0, 2\}$. The colored curve in Figure 4 uses a color gradient as in Figures 1 and 2, revealing the amount of incorrect mixing between the noise and signal segments at any given changepoint. Red dominates the first half of the curve, implying that at early changepoints the "noise" model includes too much signal; the second half grades to blue implying the reverse. At every changepoint, the sample variances of the noise and mixed "signal" segments ($k < k_\circ$, equations 20 and 19), or the mixed "noise" and signal segments ($k > k_\circ$, equations 16 and 17), are computed and substituted into equation (15). The filled circles highlight three sample indexes, $k = \{250, 500, 750\}$, and their respective summed log likelihoods, $\ell \approx \{-1611, -1592, -1613\}$. These correspond exactly to the maximum summed log-likelihood values obtained at the MLEs of variances in Figure 3a (changepoint early), 3b (changepoint correct), 3c (changepoint late), and to the values approached by the (slightly differently) summed log likelihoods of the similarly mixed time series underlying the curves of Figures 2d, 1a, and 2a, and drawn there as horizontal lines.

It is clear that, on average, the likelihood associated with the correct changepoint model, $k = k_\circ$, is greater than *any* of the incorrect changepoint models, $k \neq k_\circ$, wherein either the "noise" or "signal" segment contains samples from two distinct processes, and thus is not i.i.d. The gray curves in Figure 4 plot the summed log likelihood for 25 synthetically generated time series (examples of which were shown in Figs. 1d and 2e,f). Here, then, the sample variances of the "noise" and "signal" segments are given by equations (7) and (12), respectively, calculated at every changepoint and substituted into equation (15). In the limit of many trials, the average of these likelihood curves converges to the theoretical values of the colored curve. Figure 4 implies this to be the case.

Take-home message 1

From Figures 1 and 2, we learn that the expectation of the sample variance converges to the true population variance. Thus, we can drop the requirement of *a priori* knowledge of the true

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

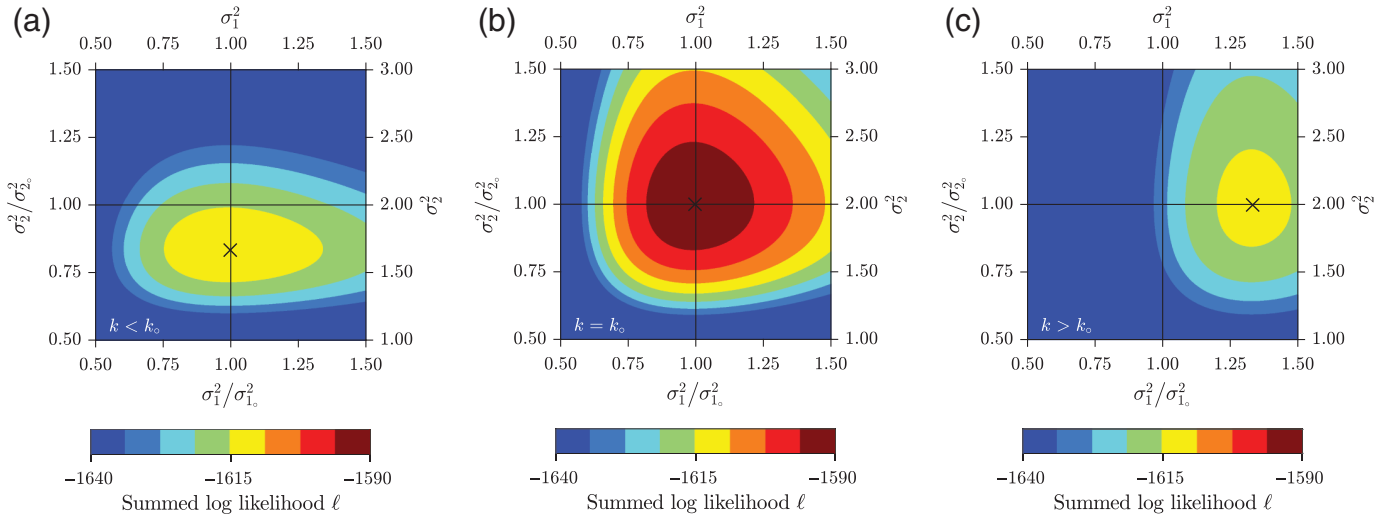


Figure 3. Summed log-likelihood plots in the two variables of interest for the cases where the changepoint estimate is early, correct, and late, respectively. In each panel, the left and lower axes are quoted in terms of the MLEs of the variances of the assumed "signal" and "noise" segmentations, one or both of which is composed entirely of the true signal or noise segment, normalized by their true variances, respectively; and the right and upper axes are in terms of the MLEs of the variances of the same segmentations not normalized by their true values. The color maps are identical, and the crosses mark the average MLEs over the 1000 tests considered here. (a) Changepoint early, (b) changepoint correct, and (c) changepoint late correspond to the segmentations of x analyzed in Figures 2d, 1a, and 2a, respectively. What differs here is that the estimate of the variances maximizes the entire summed log-likelihood surface, not the version restricted to the 1:1 diagonal shown in the earlier figures. This 2D procedure at each changepoint model k correctly identifies the normalized variances of the two segmentations, converging to the true values of $(1.000, 0.8\overline{333})$ in (a), at (1.000,1.000) in (b), and at $(1.\overline{333}, 1.000)$ in (c). Further, the summed log-likelihood value, considering all three tests shown, is maximized when the changepoint is exactly correct, as will be shown in Figure 4.

population parameters and instead rely on the statistics of many trials to converge to the truth. Further, we find that the ratio of the sample variances of both examples in Figure 2 is smaller than that in Figure 1, and indeed note the correct changepoint lies at the sample index at which the ratio of the sample variances is the largest. Figures 1a and 2a,d also show us that for three different changepoint models, the summed log likelihoods are informative, specifically telegraphing the "identicality" or "i.i.d.-ness" of the modeled "noise" and "signal" segments from which the data were most likely generated. The highest summed log likelihood identifies the split in the time series where the "noise" and "signal" segmentations are *simultaneously* individually best fit by a single-variance, i.i.d. process. The average summed log likelihoods marked by horizontal lines in Figure 2a,d, which represent two incorrect changepoint models, are each lower than in Figure 1a, the correct changepoint model.

Figures 3 and 4 cement our understanding in showing that the highest summed log likelihood among all tested changepoint models $k = [1, \ldots, N]$ corresponds to the case when the changepoint model is exactly correct, $k = k_\circ$. Therefore, a relative comparison of summed log likelihoods for every changepoint model intuitively defines a scheme for seismic arrival identification whereby the highest likelihood after testing all possible models identifies the sample index at which the seismogram is most likely split into two distinct and individually i.i.d. processes: for example, ambient noise and a seismic arrival. Our experiments show the ability of summed log likelihoods, evaluated at the maximum-likelihood variance estimate, to partition an incoming data stream into two segments that are most likely individually identifiable as being from distinct generating distributions. In the illustrations of our model, the two portions, separated at the changepoint, specifically differ only in their variance. In practice, we also estimate the mean. For every changepoint model, we assume that the first segment is composed of noise and the remainder, signal. Signal-to-noise considerations, of course, are to follow.

In preparation for defining an algorithm to apply to real data, we hereafter drop the quotes around "noise" and "signal," by which we denoted the departure of a segment from a known truth, because real data have no "true" changepoint against which to test. For the remainder of this study, we will refer to every segmentation at every model changepoint as either noise or signal, without quotes.

THE AIC

In his seminal 1973 paper (reprinted as Akaike, 1998), Akaike links MLE via evaluated likelihoods to information theory. Briefly, Akaike shows that the dual problem of parameter estimation and model testing can be solved simultaneously. At its core, the criterion that bears his name seeks to identify a best-fitting model from a set of candidate models via minimization of a loss function. The loss function used is an estimate of the Kullback and Leibler (1951) divergence (KL), which we understand as a measure of the information separation between

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

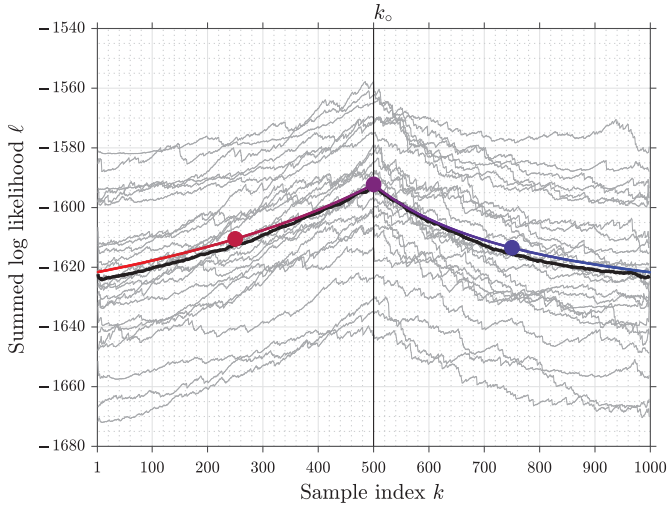


Figure 4. Summed log likelihoods (equation 15) plotted at every changepoint model, $k = [1, \ldots, N]$. The colored curve is the expected behavior, when $\hat{\Theta} = \Theta_\circ$, with the color-coding based on the amount of incorrect mixing between the (blue) noise and (red) signal sections. The three likelihood values marked by the filled circles correspond to, from left to right, the mixtures shown in Figures 2d, 1a, and 2a, respectively. Gray curves plot 25 evaluations of equation (15) for synthetically generated time series like those in Figures 1d, 2e, and 2f, and where $\hat{\Theta}$ is being estimated from those data. The black curve is the average of the 25 tests shown in gray.

two probability distributions, that is, the "distance" between "responses" due to an evaluated model and the truth. Importantly, when real data are being considered, their true generating distributions, and thus their KL divergences, can never be known and must, therefore, be estimated. Akaike's contribution was the rigorous derivation of such an estimated loss function and the proof that their relative comparisons, after accounting for model complexity via a penalty term, is useful and appropriate for model discrimination.

For our purposes, as we illustrated in The Summed Log Likelihood section, these are the ideas that allow us to make the logical leap from equation (4) to equation (8), and, similarly, from equation (10) to equation (13), and to using the likelihoods evaluated in equation (15) to discriminate between changepoint models and choose a "best" fit among them. In the examples shown there, we had access to the true generating distributions, but we showed that estimates suffice in the absence of such knowledge.

AIC-based changepoint estimation

Using the summed log likelihood of equation (15), we write the AIC as

$$\mathcal{A}(\hat{\Theta}; x; k) = -2\ell(\hat{\Theta}; x; k) + \|\hat{\Theta}\|, \tag{25}$$

whereby $\|\hat{\Theta}\|$, the length of the vector of model parameters, accounts for the degrees of freedom in the system. The first term on the right side of equation (25) is a measure of the model fit. The second term is a bias-correction term that rewards model parsimony. Likelihoods of the exponential form in our model (e.g., equations 4 and 10) are distributed as $\chi^2/2$ variates, which explains the factor of 2 in equation (25). In our case, $\|\hat{\Theta}\| = 4$, but we may ignore it altogether because it does not depend on the changepoint model k. The distributions that we consider remain unchanged between possible segmentations, and they furthermore remain in the specific Gaussian form of equation (2). Morita and Hamaguchi (1984), Maeda (1985), and Sleeman and van Eck (1999) discuss a framework by which other noise and signal models, for example, autoregressive ones, can be decorrelated into satisfying our model assumptions.

The use of the AIC for seismological event detection is widespread (e.g., Maeda, 1985; Leonard and Kennett, 1999; Sleeman and van Eck, 1999; Zhang *et al.*, 2003, 2017; Rastin *et al.*, 2013). At its essence, the AIC approach is sensitive to changes in variance (second moments) between trial segments, which it is able to neatly separate with high temporal resolution. So-called short-term average over long-term average (STA/LTA) methods, based on comparing first-moment ratios (e.g., Allen, 1978), are most efficient at picking out segments, not points, of interest. Methods that utilize higher-order (e.g., squared-envelope, skewness, kurtosis) statistics (e.g., Baer and Kradolfer, 1987; Saragiotis *et al.*, 2002; Baillard *et al.*, 2014) base their estimates on exploiting the changing nature of the distribution over segments of fixed length, which imprints a certain time and frequency resolution to the event identification. However, powerful and performant any of these alternative approaches, our method has no tunable parameters, it obtains excellent results for our data types, and it remains usefully insensitive to perturbations in our initial model assumptions, which, additionally, we remotivate in the supplemental material.

Substituting the evaluated summed log likelihoods of equation (15) into equation (25), ignoring the constant of the former and bias-correction term of the latter, and simplifying the notation to make the dependence of the AIC value on the changepoint model k explicit, we write, for an input time series x of length N

$$\mathcal{A}(k) = k \ln(\hat{\sigma}_1^2) + (N - k) \ln(\hat{\sigma}_2^2). \tag{26}$$

The discussion in The Summed Log Likelihood section implies that a natural changepoint estimate is the sample index that maximizes equation (15), or indeed, *minimizes* the AIC in equation (26), over the set of changepoint models $k = [1, \ldots, N]$. We call this estimator

$$k_\mathrm{m} = \mathrm{argmin}[\mathcal{A}(k)]. \tag{27}$$

Furthermore, we showed there that the absolute values of summed log likelihoods are irrelevant, and instead it is their relative difference that bolsters the designation of a "best"

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

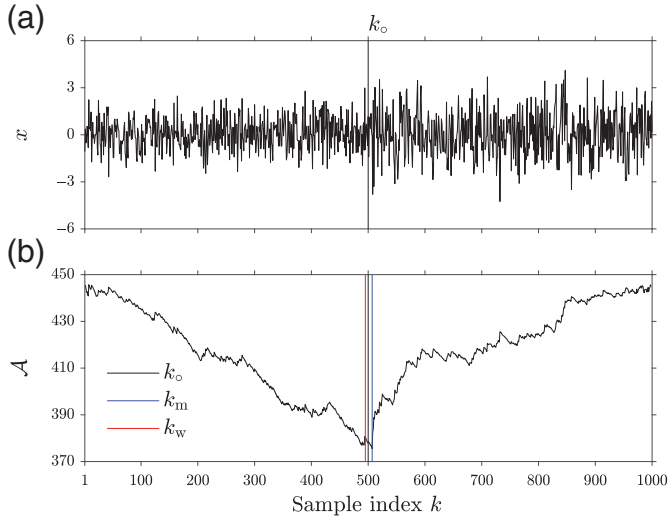


Figure 5. A model time series, its associated Akaike information criterion (AIC) curve, and the changepoint estimators k_m and k_w. (a) Synthetic time series, $x(k)$, as in Figures 1 and 2, $N = 1000$, with a true changepoint, $k_\circ$, at sample index 500, marked by a black vertical line. The noise is drawn from $\mathcal{N}(\mu = 0, \sigma^2 = 1)$, and the signal from $\mathcal{N}(\mu = 0, \sigma^2 = 2)$. (b) The AIC curve associated with the time series shown in (a), calculated from equation (26). Again, $k_\circ$ is marked in black, whereas the estimator k_m (equation 27) is marked in blue, and the estimator k_w (equation 29) in red.

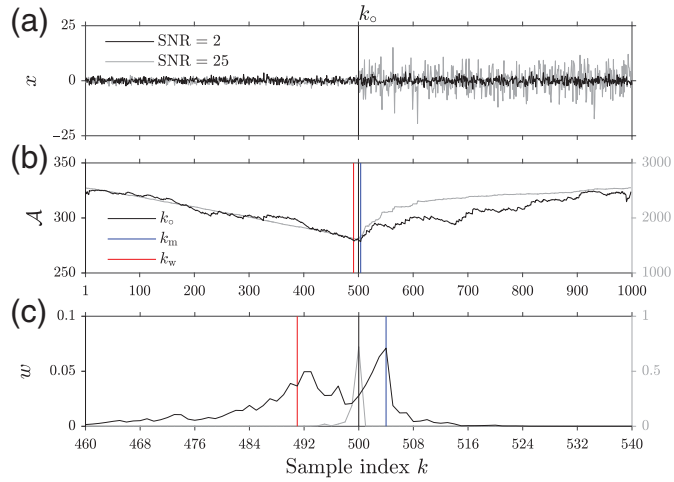


Figure 6. High-signal-to-noise ratio (SNR) and low-SNR synthetic time series, their AIC curves, their Akaike weights, and the changepoint estimates. As before, $N = 1000$ and $k_\circ = 500$ for both examples shown here, and $k_\circ$ is marked by a black vertical line. (a) High-SNR synthetic time series (SNR = 25) in gray, and a low-SNR synthetic time series (SNR = 2) in black. (b) Their associated AIC curves (equation 26), with the label on the left ordinate axis corresponding to the low-SNR example and the label on the right ordinate axis to the high-SNR example. (c) Akaike weights (equation 30) associated with both examples, again with a double ordinate axis, on a zoomed-in abscissa axis to show detail about the true changepoint.

changepoint model. Following the discussion of Burnham and Anderson (2002), we introduce the AIC difference, the distance along the ordinate axis between an AIC value obtained at a particular model k and the minimum AIC value of the ensemble

$$\Delta(k) = \mathcal{A}(k) - \mathcal{A}(k_m). \tag{28}$$

A small AIC difference implies the model in question is relatively likely given the set of all models tested, and a large AIC difference provides reason to believe the opposite is true. We leverage the AIC differences as a tool for model discrimination, by interpreting the exponential form $\exp[-\frac{1}{2}\Delta(k)]$ as a measure of the relative likelihood of the model k compared to the set. With Li *et al.* (2009), we define a second changepoint estimator, k_w

$$k_w = \sum_{k=1}^{N} k w(k), \tag{29}$$

using what are now commonly called Akaike weights

$$w(k) = \frac{\exp[-\frac{1}{2}\Delta(k)]}{\sum_{i=1}^{N} \exp[-\frac{1}{2}\Delta(i)]}. \tag{30}$$

Figure 5 shows the difference between the estimators k_m of equation (27) and k_w of equation (29) for a synthetic time series, in which $N = 1000$, and $k_\circ = 500$, as derived from the AIC curve calculated using equation (26). In this example, $k_m = 507$, which is late compared to $k_\circ$. In contrast, the second changepoint estimator is early, $k_w = 495$, preceding $k_\circ$.

The SNR

We estimate the SNR from the ratio of sample variances of the segments identified as signal and noise. For a particular changepoint index k, using equations (7) and (12), we define

$$\mathrm{SNR} = \hat{\sigma}_2^2 / \hat{\sigma}_1^2. \tag{31}$$

A time series characterized by a high SNR will be revealed by a steep AIC curve that rapidly and almost surely monotonically decreases, and then rapidly and virtually monotonically increases, after reaching an easily identified, single, global minimum, and finally, flattening asymptotically. The Akaike weights (equation 30) are near zero everywhere, except within a small sample span about the true changepoint. In these cases, k_m and k_w will generally coincide. Conversely, a low-SNR time series will have a flatter AIC curve with multiple local minima and no obvious global minimum, and the associated weights will be more broadly spread about the true changepoint. In those cases, k_w and k_w may differ greatly.

Figure 6 shows this behavior. In Figure 6a, we plot two synthetic times series: a low SNR = 2 example in black, and a high SNR = 25 example in gray. Figure 6b plots their associated AIC functions (equation 26), and Figure 6c plots their Akaike weights (equation 30). As in all previous examples $N = 1000$

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

and $k_\circ = 500$, and $k_\circ$ is marked with a black vertical line in Figure 6a–c. Both Figure 6b,c have a double ordinate axis: the left corresponds to the low-SNR time series and the right to the high-SNR time series. Note their order-of-magnitude difference in range.

In Figure 6b, the low-SNR black AIC curve has many local minima surrounding the global minimum and is much flatter compared to the gray high-SNR curve, which steeply and nearly monotonically decreases to an obvious global minimum, and then rapidly increases to an asymptote. The effect of these minima on the Akaike weights is apparent from Figure 6c, which is rescaled between sample indexes 460 and 540 to show detail about the true changepoint. Here, the weights in the high-SNR case are narrowly distributed and nearly centered on $k_\circ$, whereas the low-SNR case has a broad, multimodal distribution. In the latter case, k_w is early at $k_w = 491$ and is marked by the red vertical lines, whereas k_m is late at $k_m = 504$ and is marked by the blue vertical lines. The two changepoint estimators for the high-SNR case coincide with $k_\circ$.

FORMALIZING THE AIC TIMING UNCERTAINTY

The novelty of our work lies in the development of a statistical framework to estimate the timing uncertainty of the changepoint estimates k_m and k_w. We present two distinct methods. Both calculate error statistics using many realizations of AIC curves, the associated time series of which were generated by random sampling from distributions with known statistics, but they work in different coordinates. Method I collects the error along the sample index (or equivalently, time) axis of the AIC curve (a distance), whereas Method II performs hypothesis tests using the AIC values themselves (as percentages). We discuss the relative utility of both methods to inform the uncertainty estimation and the assignment of confidence intervals of the changepoint estimates k_m and k_w.

A baseline scenario to compare the bias and variance of the changepoint estimators k_m and k_w involves testing their performance using a time series built of noise and signal segments randomly sampled from distributions with known statistical parameters $\Theta_\circ$, concatenated at a known true changepoint $k_\circ$. As before, and for all tests in this section, we generate low-SNR synthetic time series of length $N = 1000$, for which sample indexes $k = [1, \ldots, 500]$ are drawn from $\mathcal{N}(\mu = 0, \sigma^2 = 1)$, and samples $k = [501, \ldots, 1000]$ from $\mathcal{N}(\mu = 0, \sigma^2 = 2)$. Here, then, the SNR is 2, and the true changepoint that we attempt to locate is sample index 500, the last sample index of the noise segment.

Method I: Monte Carlo resimulation

The Method I (M1) test procedure directly measures the timing error, the discrepancy between a true changepoint, $k_\circ$, and an estimated changepoint, $\hat{k}$, either k_m or k_w, calculated for a synthetic time series. We define the timing error in terms of sample indexes as the signed distance

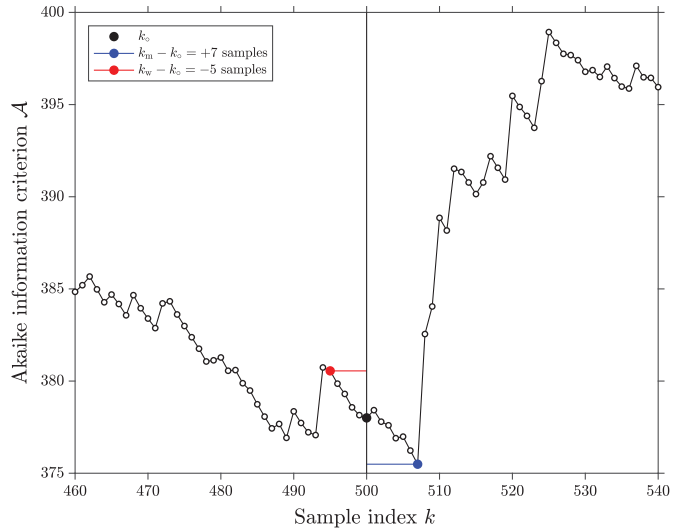


Figure 7. One realization of the Method I (M1) testing procedure, tallying the distance in sample indexes between the true changepoint and its estimates, applied to the same AIC curve in Figure 5b (SNR = 2), shown in detail about the true changepoint, $k_\circ$, marked by a black-filled circle. The estimator k_m (equation 27) is marked by a blue-filled circle, and k_w (equation 29) by a red-filled circle. The timing error is the difference between the estimate and the truth (equation 32), shown here as blue and red horizontal lines that connect k_m and k_w to the vertical line at $k_\circ$. In this example, k_m is seven sample indexes late, whereas k_w is five sample indexes early.

$$\hat{k} - k_\circ. \tag{32}$$

To evaluate the bias and variance of k_m and k_w, we generate many random realizations of the low-SNR time-series model (SNR = 2) described in the previous paragraph, calculate the corresponding AIC curve via equation (26), calculate changepoint estimates k_m and k_w with equations (27) and (29), respectively, and then tally the timing errors between these estimates and $k_\circ$ with equation (32). We summarize the error statistics after many realizations of this procedure.

Figure 7 displays one realization of the M1 testing scheme applied to the same AIC curve in Figure 5b but shown with a zoomed-in abscissa axis. Here, $k_\circ$ is marked with a black-filled circle and intersecting vertical line, and k_m and k_w are marked with blue and red-filled circles, respectively. The distance in sample indexes between each changepoint estimate and the truth is marked by a similarly color-coded horizontal bar of the appropriate length. In this example, the minimum-AIC estimator, $k_m = 507$, is late, and the weighted-average estimator, $k_w = 495$, is early.

Figure 8 summarizes the M1 error statistics after one million realizations of the procedure just described. The sample-distance errors are grouped into one-integer bins and plotted as histograms. The unfilled bars with black edges represent the distribution of the error of the estimator k_m, whereas the gray bars represent the error of the k_w estimator. Overlain are two curves, blue for k_m and red for k_w, respectively, which represent

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

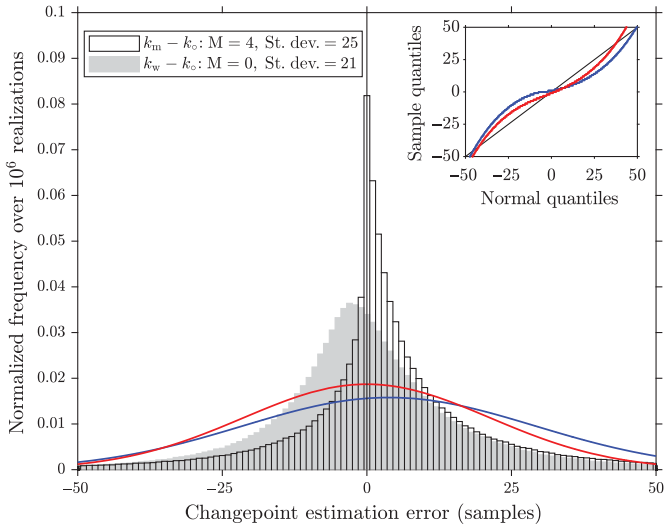


Figure 8. Histograms of estimation errors of the estimators k_m (black-unfilled bars) and k_w (gray-filled bars), after one million realizations of the Method I (M1) testing procedure, one of which is displayed in Figure 7. Overlain on both histograms are their best-fitting Gaussian probability distribution functions, blue for k_m and red for k_w, respectively, the means and standard deviations of which are quoted in the legend. Inset in the upper right is a quantile–quantile plot, again color-coded. In this low-SNR model (SNR = 2), the k_w estimator is unbiased and has a lower standard deviation than the k_m estimator, which has a positive bias revealing that, on average, it estimates the changepoint late.

their corresponding best-fitting Gaussian probability distribution functions (pdfs) given the statistics quoted in the legend and elaborated upon later. The abscissa axis is limited to ± 50 sample indexes for display purposes, but mass extends beyond these limits for both histograms. Inset into the upper right corner of Figure 8 is a quantile–quantile plot that displays the sample-distance error quantiles as a function of the quantiles of their best-fitting Gaussian pdfs, again color-coded blue for k_m and red for k_w. A black line with a slope equal to 1 is also plotted for reference.

Invoking Figure 8, we submit that k_w is a better estimator than k_m. Neither of them is Gaussian. Our test reveals that even in extremely low-SNR regimes k_w is an unbiased estimator, whereas k_m is biased, here with a mean error of 3.8 sample indexes. We also find that k_w has a narrower error distribution with a standard deviation of 21 sample indexes compared with 25 sample indexes for k_m.

The story grows somewhat more complex when the data are more fully inspected. For example, the k_m errors have their mode at 0 sample indexes and their median at 2 sample indexes, whereas the k_w errors have their mode at -3 samples and their median at -1. The peak at 0 sample indexes in the k_m error histogram shows that this estimator is about twice as likely as the k_w estimator to be exactly correct; however, the lower standard deviation of the k_w error data shows that, on average, k_w will be closer to the truth. Even though any given k_w estimate is less likely than a k_m estimate to be exactly correct, we still consider k_w to be a better estimator than k_m because it is unbiased, when k_m is not, and lower variance than k_m.

A positive bias of the minimum-AIC estimator was also documented by Leonard (2000). There, the author compares the slope of an autoregressive variant of our AIC curve immediately following k_m with analyst picks of the onset to derive an empirical relation to adjust the typically tardy estimate backward in time.

Method II: Using the shape of the AIC curve

The Method II (M2) test procedure comprises two hypothesis tests that employ a proxy for the sharpness of the AIC curve near $\hat{k}$. We know from numerical experiments that in high-SNR cases the AIC curve rapidly and monotonically descends to a single and obvious global minimum at the true changepoint, then rapidly and monotonically ascends to an asymptote. Conversely, in low-SNR cases the AIC curve is much flatter, with a lower overall range, and many local minima that amplify changepoint estimation uncertainty. The M2 test procedure proposed here seeks to quantitatively relate the shape of the AIC curve itself to a timing-confidence interval. Each test in the M2 suite asks a variation of the basic question: how steep is the AIC curve around the changepoint estimator $\hat{k}$, and how much greater is the AIC value at the true changepoint $k_\circ$ compared to the AIC value at its estimate? Like the M1 test procedure, statistics are generated via repetition of the M2 test over many randomly generated synthetic time series. Unlike M1, the confidence intervals returned are asymmetric, not measures of center and spread. Altogether, M2 produces probability curves that relate the shape of the AIC curve near $\hat{k}$ to a confidence interval on the estimate.

We define α as a percentage of the total range of the AIC curve (equation 26), and β as the sum of the αth fraction of the range and the AIC value of the changepoint estimator, $\hat{k}$, either k_m or k_w

$$\beta(\hat{k}, \alpha) = \mathcal{A}(\hat{k}) + \frac{\alpha}{100}(\max[\mathcal{A}(k)] - \min[\mathcal{A}(k)]). \tag{33}$$

With this, we define our first "unrestricted" β hypothesis test

$$\mathrm{H}_0 : \mathcal{A}(k_\circ) > \beta(\hat{k}, \alpha), \tag{34}$$

$$\mathrm{H}_1 : \mathcal{A}(k_\circ) \le \beta(\hat{k}, \alpha). \tag{35}$$

In the unrestricted β-test, the null hypothesis is rejected if the AIC value of the true changepoint is equal to or less than the AIC value of the estimated changepoint plus a variably defined percentage of the total range of the AIC curve over the segment considered. We label the first and last sample indexes whose AIC values are less than or equal to β as

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

$$k' = \min\{k : \mathcal{A}(k) \le \beta(\hat{k}, \alpha)\}, \tag{36}$$

$$k'' = \max\{k : \mathcal{A}(k) \le \beta(\hat{k}, \alpha)\}. \tag{37}$$

Last, we note a mapping exists that relates α to the corresponding maximum sample span to which this test could apply, and call it the

$$\text{unrestricted } \beta\text{-test sample span} = 1 + (k'' - k'). \tag{38}$$

Herein lies the origin of our use of the term "unrestricted," because the elements in the set $\{\mathcal{A}(k) : k = k', \ldots, k''\}$ need not all be less than or equal to β. As written in equations (34) and (35), the unrestricted β-test is applied in the ordinate direction: defined in terms of AIC values, not sample indexes (potential changepoint models), at which those AIC values are obtained. The definitions of k' and k'' were introduced to formalize the α-to-sample-span map, which is pivotal to the utility of M2.

Conversely, in defining our second hypothesis test, the "restricted" β-test, we begin by highlighting the relevant sample span of that test, which we term the

$$\text{restricted } \beta\text{-test sample span} = 1 + (k^{\dagger\dagger} - k^{\dagger}), \tag{39}$$

in which, for the relevant values of α,

$$k^{\dagger} = \max\{k < \hat{k} : \mathcal{A}(k) > \beta(\hat{k}, \alpha)\} + 1, \tag{40}$$

$$k^{\dagger\dagger} = \min\{k > \hat{k} : \mathcal{A}(k) > \beta(\hat{k}, \alpha)\} - 1. \tag{41}$$

If it is nonempty, the elements in the set $\{\mathcal{A}(k) : k = k^{\dagger}, \ldots, k^{\dagger\dagger}\}$ are all equal to or less than β. A restricted β-test asks if $k_\circ$ lies within this contiguous set

$$\mathrm{H}_0 : k_\circ \notin [k^{\dagger}, \ldots, k^{\dagger\dagger}], \tag{42}$$

$$\mathrm{H}_1 : k_\circ \in [k^{\dagger}, \ldots, k^{\dagger\dagger}]. \tag{43}$$

In graphical terms, the restricted β-test asks: (1) is $\mathcal{A}(k_\circ)$ below β, and if so, (2) is $k_\circ$ in the same trough of the AIC curve that contains $\hat{k}$? Unlike an unrestricted β-test, where the null hypothesis may be rejected even if the AIC values between $\mathcal{A}(\hat{k})$ and $\mathcal{A}(k_\circ)$ are greater than β, a restricted β-test does not allow a local maximum, where the AIC curve rises above β, to be crossed in search of $\mathcal{A}(k_\circ)$. In both unrestricted and restricted β-tests, as α increases, the sample span (number of changepoint models) under consideration increases and the null hypothesis is rejected more frequently. Given these definitions, we expect that restricted β-tests applied to k_m reject the null hypothesis at the lowest rate because they consider the smallest sample spans, and that unrestricted β-tests applied to k_w would reject the null hypothesis at the highest rate because they consider the largest sample spans.

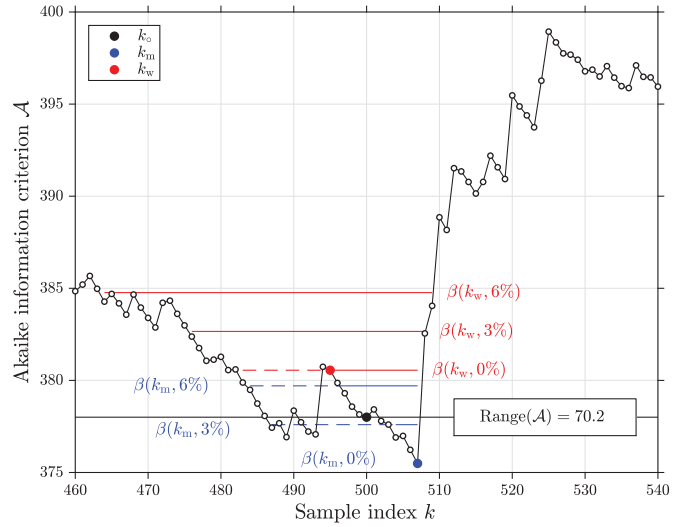


Figure 9. One realization of the Method II (M2) hypothesis test, applied to the same AIC curve shown in Figure 7. Again, $k_\circ$, k_m, and k_w are marked by black-, blue-, and red-filled circles, and $\mathcal{A}(k_\circ)$ by a black horizontal line. Three values of β (equation 33), corresponding to three percentages of the total range of the AIC curve, α, of 0%, 3%, and 6%, are shown in blue and red as related to k_m and k_w, respectively. Solid blue and red lines represent the sample spans to which a restricted β-test maps (equation 39), whereas the dashed lines represent the additive sample spans through which an unrestricted β-test maps (equation 38). For a specific α, a restricted β-test (equations 42 and 43) asks if $k_\circ$ falls in the interval below the colored solid line, whereas an unrestricted β-test (equations 34 and 35) asks whether $k_\circ$ falls under either the colored solid or dashed lines. The null hypothesis in both cases remains that $k_\circ$ is found outside those intervals. In this example, for both restricted and unrestricted β-tests the null hypothesis is only rejected for k_m when $\alpha = 6\%$, though it is rejected for all α related to k_w.

The utility of the M2 testing procedure is in the generation of probability graphs that measure the shape of the AIC curve by relating α to sample-span confidence intervals that include the truth some proportion of the time. As in M1, these curves are generated by applying M2 to many realizations of synthetic time series. With each realization, both an unrestricted and restricted β-test is performed, and their test results (null hypothesis rejection rates) and associated sample spans (equations 38 and 39) are recorded.

Figure 9 shows one such M2 test realization applied to the same section of the same AIC curve of Figure 7 (SNR = 2). The AIC value of the true changepoint, $A(k_\circ)$, is marked by a black-filled circle bisected by a black horizontal line. This represents the value that β must exceed to reject the null hypothesis in an unrestricted test, and the value at or below which both $\mathcal{A}(\hat{k})$ and $\mathcal{A}(k_\circ)$ must lie, in the same trough, to reject the null hypothesis in a restricted β-test. A blue-filled circle marks k_m and a red-filled circle marks k_w. Three β-values for both estimators are shown as horizontal lines, using the same color scheme, at α-values equal to 0%, 3%, and 6%. Solid horizontal lines

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

represent the sample spans to which the restricted β-tests map (equation 39), and dashed horizontal lines represent the sample spans to which an unrestricted β-test maps (equation 38) in addition to those already considered in a restricted β-test.

For k_m in Figure 9, $\alpha = 0\%$, $\alpha = 3\%$, and $\alpha = 6\%$ correspond to sample spans equal to 1, 4, and 11, respectively, for a restricted β-test, and 1, 21, and 24, respectively, for an unrestricted β-test. The null hypothesis is not rejected at $\alpha = 0\%$ or $\alpha = 3\%$, for either the restricted or unrestricted β-test, though it is rejected at $\alpha = 6\%$ in both tests.

In contrast, for k_w in Figure 9, $\alpha = 0\%$, $\alpha = 3\%$, and $\alpha = 6\%$ correspond to sample spans equal to 13, 33, and 46, respectively, for a restricted β-test, and 25, 33, and 46, respectively, for an unrestricted β-test. The null hypothesis is rejected at all α-values for both restricted and unrestricted β-tests.

Figure 10 displays the results of 1000 realizations of M2 applied to synthetic time series of the same model as in the Method I: Monte Carlo Resimulation section (SNR = 2), one example of which is shown in Figure 5a. Here, the test proceeds for α in the inclusive range $\alpha = [0, \ldots, 100]$, tested in $\alpha = 0.1$ increments, for both restricted and unrestricted β-tests, for both estimators k_m and k_w. The data are plotted in terms of the probability of rejecting H_0 as a function of α. As before, the probability curves are color-coded blue and red for changepoint estimators k_m and k_w, respectively. As in Figure 9, the solid curves represent the results of restricted β-tests, whereas the dashed curves show the results of unrestricted β-tests. As expected, Figure 10 shows that, in general, β-tests associated with k_m reject the null hypothesis at a lower rate than those associated with k_w, and restricted β-tests reject the null hypothesis at a lower rate than their unrestricted relatives.

Figure 11 plots the same results as in Figure 10 except here the probability of rejecting H_0 is plotted as a function of the average sample span considered under each β-test. Again, curves corresponding to k_m and k_w are blue and red, respectively, and restricted β-tests curves are solid and unrestricted β-test curves dashed. Black horizontal dashed lines originating from the ordinate axis at 0.68 and 0.95 are plotted to guide the eye to their corresponding timing confidence intervals, respectively. These confidence intervals are marked by vertical solid lines that originate from the abscissa axis and intersect the horizontal dashed lines. From left, the first pair of vertical lines marks the total spread of the average of the sample spans considered under each β-test that included the truth 68% of the time. The lower bound of this confidence interval is defined by the restricted k_m curve at 22 sample indexes, and the upper bound is defined by the unrestricted k_w curve at 37. The second pair of vertical lines marks the total spread of the average of the sample spans considered under each β-test that included the truth 95% of the time. The lower bound of this confidence interval is defined by the unrestricted k_m curve at 62 sample indexes, and the upper bound is defined by the unrestricted k_w curve at 69 sample indexes.

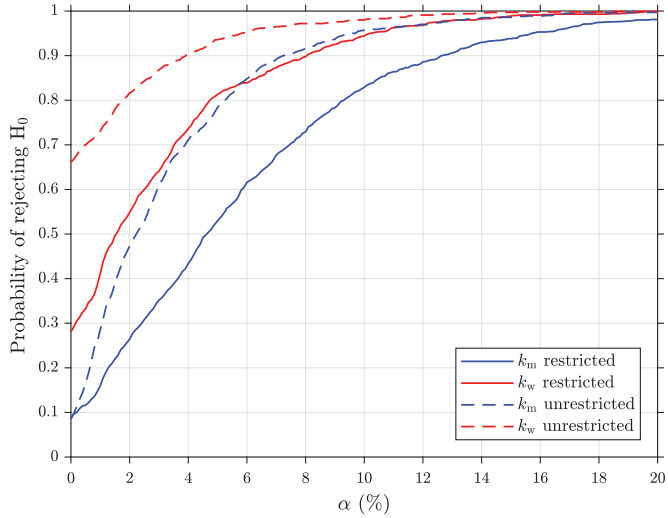


Figure 10. Probability of rejecting H_0 as a function of α for restricted and unrestricted β-tests for both changepoint estimators k_m and k_w. These curves summarize the output of M2 after many test realizations, one example of which is displayed in Figure 9. Here, restricted (equations 42 and 43) and unrestricted β-tests (equations 34 and 35) were each applied 1000 times at every α. As in Figure 9, the test curves associated with k_m and k_w are color-coded blue and red, respectively, and solid and dashed lines represent the outcomes of restricted and unrestricted β-tests, respectively. Overall, restricted β-tests corresponding to k_m generally map to the shortest sample spans (equation 39) and thus reject the null hypothesis at the lowest rate, whereas unrestricted β-tests corresponding to k_w generally have the highest rejection rate at the expense of mapping to the longest sample spans (equation 39). Figure 11 replots these results as the probability of rejecting H_0 as a function of the average sample span of each β-test.

Comparison of Methods I and II

The β-test curves of Figure 11 connect the M2 hypothesis testing procedure to the M1 Monte Carlo resimulation. In Figure 8, we find that the M1 one-standard-deviation error estimates of k_w and k_m are 21 and 25 sample indexes, respectively. From Figure 11, we see that the M2 one-standard deviation confidence interval is between 22 and 37 sample indexes, a similar though larger range than what was found in M1. However, there is a cluster of β-test curves near the lower end of the M2 one-standard-deviation confidence interval, implying that the true one-standard-deviation confidence interval is nearer 22 sample indexes than 37. The large spread in β-test sample spans between k_m restricted tests and k_w unrestricted tests is greatly diminished in higher-SNR cases, where the AIC curve is less flat and the four tests see roughly the same sample spans.

The lack of agreement between the M1 two-standard deviation error estimate between 42 and 50 sample indexes (Fig. 8) and the M2 two-standard deviation confidence interval at 62 and 69 sample indexes (Fig. 11) is likely due to a few factors. First, the asymmetric nature of the AIC curve may play a part. AIC curves tend to have a lower gradient immediately before the true changepoint compared to the gradient immediately after it (note this phenomenon in Fig. 9). What this

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

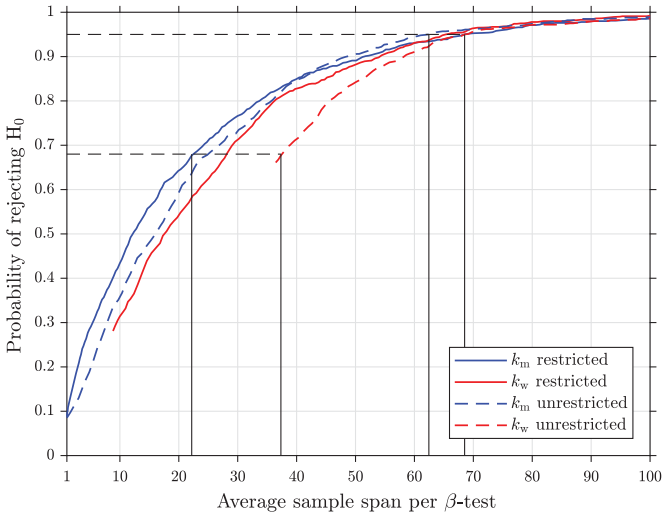


Figure 11. Probability of rejecting H_0 as a function of the average sample span considered under each β-test. These curves render the same simulation results as Figure 10 except they are now plotted in terms of the average sample span of each β-test. Vertical lines tied to the curves reaching 0.68 and 0.95 on the ordinate axis, from left to right, define the corresponding confidence intervals. The 68% confidence interval is bounded on the left at 22 sample indexes by the restricted β-test curve corresponding to k_m, and at 37 sample indexes on the right by the unrestricted β-test curve corresponding to k_w. The 95% confidence interval is bounded on the left at 62 samples by the unrestricted β-test curve corresponding to k_m, and on the right at 69 sample indexes by the unrestricted β-test curve corresponding to k_w.

means is that the sample span considered under an unrestricted β-test to the left of an incorrect changepoint estimate may be quite large, even if the estimate differs from the truth by only a single sample index. In addition, M2 confidence intervals will always be at least as large, but likely larger, than the error estimates returned in M1. This is due to the fact that M1 simply tallies the error from an estimate to a truth with no overshoot, whereas any given β-test will more than likely include more sample indexes than are required to find the truth. Stated another way: there is likely to be overshoot in the sample span of a β-test in the event that enough sample indexes are considered to reject the null hypothesis. This overshoot can be reduced by refining the discretization of α at the expense of increased computational complexity.

The utility of M2 is the generation of curves that negate the need for resimulation of synthetic data as in M1. Assuming a time series of similar SNR has previously been processed through the M2 procedure, a researcher may immediately quote any arbitrary confidence interval given a probability range of interest simply by querying the appropriate β-test curve.

Take-home message 2

The contribution of this study is a scheme for the rapid estimation of timing-error confidence intervals by inspection of the shape of the AIC curve, without the need for resimulation using synthetic data. After the generation of a library of curves detailing the probability of rejecting H_0 as a function of the average sample span of each β-test, like the one shown in Figure 11, for seismograms with various SNRs, no new synthetic time series for testing need be generated. Instead, a researcher may simply ask: "what is the probability that the truth lies within a sample span equal to (an arbitrary number of sample indexes) that also includes $\hat{k}$?" or alternatively, "what is the sample span that includes $\hat{k}$ and has a probability of (an arbitrary value) of including the truth?"

MULTISCALE ANALYSIS METHODOLOGY

The waveforms of various arriving seismic phases have distinct frequency signatures, and that information can and needs to be explicitly taken into account during inversion (e.g., Luo and Schuster, 1991; Dahlen *et al.*, 2000; Yuan and Simons, 2014). In what follows we prepare to apply the concepts of The AIC and Formalizing the AIC Timing Uncertainty sections to a very specific type of timescale analysis via the wavelet transform (Strang and Nguyen, 1997; Mallat, 1998).

Our philosophy is perhaps closest to that of Zhang *et al.* (2003), and our contribution can be understood as picking up where they left off. Although Zhang *et al.* (2003) use an AIC on wavelet-coefficient time series, they only use those picks and their consistency across neighboring scales to identify waveform segments of interest, which they then analyze with the AIC on the original time series. In contrast, here, our goal is to use AIC and wavelet analysis to determine scale-dependent seismic arrival-time picks and their associated uncertainties.

The discrete wavelet transform

We transform the time series into the multiresolution wavelet domain. Various orthogonal or biorthogonal sets of self-similar scaled (dilated) and shifted (translated) basis functions (wavelets, $\psi(k)$, scaling functions, $\phi(k)$, and their duals) can be used to decompose our time series $x(k)$. The analysis yields a set $j = [1, \ldots, J]$ of scaling coefficients $a_j(l)$, containing approximations at a certain scale j (where a high number denotes a coarse resolution of x), and wavelet coefficients, $d_j(l)$, which provide the details missing to proceed to higher-resolution (at a lower scale number). The index sets in the timescale domain are scale dependent, $l = [1, \ldots, M(j)]$, hereinafter implied where not explicitly stated, and encode the translations of the basis functions, giving them two indexes, $\psi_{j,l}(k)$ and $\phi_{j,l}(k)$. The identity

$$x(k) = \sum_{j=1}^{J} \sum_{l=1}^{M(j)} d_{j,l}\psi_{j,l}(k) + \sum_{l=1}^{M(J)} a_{J,l}\phi_{J,l}(k), \qquad (44)$$

leads to two ways of implementing our detection algorithm: either in the timescale domain, that is, directly on the coefficient sequences

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

$$d_{j,l}, \tag{45}$$

$$a_{J,l}, \tag{46}$$

or in the time domain using the timescale subspace projections,

$$x_j(k) = \sum_{l=1}^{M(j)} d_{j,l}\psi_{j,l}(k), \tag{47}$$

$$\overline{x}_j(k) = \sum_{l=1}^{M(j)} a_{j,l}\phi_{j,l}(k). \tag{48}$$

The nested structure implied by equations (44)–(48) explains why the detail coefficients $d_{j,l}$ are colloquially referred to as "differences" because they hold the information lost in the successively coarsening approximations $a_{j,l}$, or "averages." A similar procedure was followed by Anant and Dowla (1997), who called the sequences (47) and (48) "interpolated coefficients."

The timescale coefficients are inner products of the input time series with the basis functions ψ and ϕ. We utilize wavelets with compact support that have nonzero, real values only over a finite interval. As the scale number increases, the support of the wavelet and scaling functions doubles. The detail coefficients derive from a high-pass or differencing filter, whereas the approximation coefficients are the results of low-pass or moving-average filters. Our wavelet transform is nonredundant: details and approximations at increasing scales are obtained by iterating on the low-pass branch of what amounts to a filter bank (Strang and Nguyen, 1997). Various fast algorithms for wavelet analysis and synthesis are in use: here, we use the lifting scheme of Sweldens (1996).

Many wavelet bases exist and their choice depends on multiple factors including the domain of application, computational complexity, symmetry and smoothness of the basis functions, and the data themselves, as well as timescale tiling considerations. Compared to alternative decompositions as, for example, the discrete-time short-time Fourier transform, where the window length must be specified *a priori* and thus the timing resolution for all frequencies is constant, wavelet methods subdivide the timescale domain into tiles of variable size, whose duration is inversely proportional to frequency (e.g., Chakraborty and Okaya, 1995; Tibuleac and Herrin, 1999). At the lowest scale numbers, j, we thus experience excellent time resolution and are able to extract high-frequency information, whereas at the higher scale numbers the opposite is true. With the wavelet transform, the character of abrupt (e.g., the arrival of high-frequency P-wave energy) or emergent (e.g., the slow onset of a T wave) signals is revealed upon inspection (e.g., Gendron *et al.*, 2000; Simons *et al.*, 2009; Sukhovich *et al.*, 2011, 2014).

Let us return to the problem of the article. The application of the wavelet transform to the time series of interest, $x(k)$, yields two new sets of time series, sensitive to information at different scales, j. One type is a coefficient series indexed in the timescale domain, that is, the $d_{j,l}$ and $a_{j,l}$ of equation (44), the other is a regular time series containing the subspace projections, that is, the $x_j(k)$ and $\overline{x}_j(k)$ of equations (47) and (48).

Multiscale AIC-based changepoint estimation

Acting on $x(k)$, equation (26) led to the estimates k_{m} and k_{w} in equations (27) and (29). We can now substitute any of the four coefficient or time series just derived into equation (26), to return eight new changepoint estimators. Forgoing k_{m}, which we showed to be biased, we focus on the multiscale analogs to k_{w} to define the set

$$k_{\mathrm{w}_j} \text{ derived from the wavelet-space projection } x_j(k), \tag{49}$$

$$\overline{k_{\mathrm{w}_j}} \text{ derived from the scaling-space projection } \overline{x}_j(k), \tag{50}$$

$$l_{\mathrm{w}_j} \text{ derived from the wavelet coefficient series } d_{j,l}, \tag{51}$$

$$\overline{l_{\mathrm{w}_j}} \text{ derived from the scaling coefficient series } a_{j,l}. \tag{52}$$

The first two are available for immediate use in seismology. The latter two require a method to map a timescale domain coefficient index to the time domain, in practice: the range of time-domain indexes that a timescale domain index is sensitive to under the forward and inverse wavelets transforms. The mapping under the forward transform is described next, and under the inverse transform in the supplemental material, allowing for the possibility that the forward and inverse transforms need to be considered separately, as is the case for any biorthogonal transform. Readers uninterested in the mechanics of the algorithmic implementation are invited to skip forward to the Multiscale Analysis in Practice section, where we illustrate that the method works, or to the Application to MERMAID Seismograms section, where we illustrate its application to real data.

Timescale to time mapping: forward transform

A particular timescale coefficient index l at a particular scale j may be associated with a set of time-domain sample indexes, identifying those points in the time domain from which a particular coefficient in the timescale domain receives contributions under the forward wavelet transform, via the mapping

$$F_j^{'}(l) = k_{j,l}^{'\star}. \tag{53}$$

The asterisk reminds us that the output of $F_j^{'}(l)$ for a single timescale domain index l at a single scale j is a contiguous time-domain sample span and not a single time-domain index k. Such

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

a mapping exists for every index l at every scale j and is particular to the basis used. The support of a wavelet or scaling basis function used in the forward transform, at a certain scale and translation, is completely described by the corresponding set $k'^{\star}_{j,l}$.

We call out three time-domain sample indexes in each set, $k'^{\star}_{j,l}$: the "left" edge, the "middle," and the "right" boundary:

$$k'^{\perp}_{j,l} = \min(k'^{\star}_{j,l}), \tag{54}$$

$$k'^{|}_{j,l} = \left\lfloor \frac{1}{2}(k'^{\perp}_{j,l} + k'^{\top}_{j,l}) \right\rceil, \tag{55}$$

$$k'^{\top}_{j,l} = \max(k'^{\star}_{j,l}), \tag{56}$$

in which $\lfloor\rceil$ in equation (55) signifies rounding to the nearest integer. Equations (54)–(56) allow us to tack a changepoint estimate determined in the timescale domain, and thus representing a time smear, to a single time-domain sample index.

The supplemental material contains the complementary map for the inverse wavelet transform, and details our treatment of edges in both transform directions.

Multiscale changepoints to multiscale arrival times

The time-domain changepoint estimate (equation 29) by our definition is the last sample of the noise segment; thus, the reported arrival time of a specific seismic phase will be at the sample index that immediately follows, $k_{\mathrm{w}} + 1$.

Similarly, the multiscale changepoint estimates of equations (49) and (50) effortlessly map into arrival times by adding one to their time-domain sample indexes, $k_{\mathrm{w}_j} + 1$ and $\overline{k_{\mathrm{w}_j}} + 1$. We note that these may correspond to different seismic phases, or differently resolved specific arrivals.

For their part, the changepoint estimates (51) and (52) need to be mapped into the time-domain sample indexes that capture arrival times in a way that is not similarly trivial. Indeed, in this case all of our operations (AIC picking, SNR estimation, and uncertainty estimation) occur natively in the timescale domain, not after conversion to the time domain. We leverage the mapping established in the Timescale to Time Mapping: Forward Transform section to label the time-domain sample spans $k'^{\star}_{j,l_{\mathrm{w}}+1}$ and $k'^{\star}_{j,\overline{l_{\mathrm{w}}}+1}$, where we draw attention to the fact that the addition of one index occurs in the subscripted timescale domain. From there on, we can use equations (54)–(56) to report individual time-domain sample indexes.

The final conversion of course, of any and all of these estimates, will map time-domain sample indexes into time by taking into account any initial offset and the sampling rate.

Take-home message 3

The careful mapping of time-domain sample indexes of a time series of interest onto timescale domain coefficients, and vice versa, for orthogonal and biorthogonal wavelet and scaling functions, under the wavelet transform and its inverse, allows us to carry out changepoint estimation via the AIC in either domain, giving us three options: directly in the time domain, directly on the timescale domain coefficient sequences, or on partially reconstructed subspace projections in the time domain, leading to true multiscale arrival-time detection.

MULTISCALE ANALYSIS IN PRACTICE

As we did in the Formalizing the AIC Timing Uncertainty section with the time-domain changepoint estimation, here in this section we test the performance of the multiscale method and its implementation in code. We stick with the Cohen–Daubechies–Feauveau wavelet basis CDF(2,4), with two primal and four dual vanishing moments (Cohen *et al.*, 1992), based on prior experience with actual data: their basis functions are short and capture the P-wave onset with just a handful of diagnostic coefficients (Simons, Dando, and Allen, 2006; Sukhovich *et al.*, 2011; Yuan and Simons 2014).

The changepoint estimators that we study in detail are: those derived by applying the AIC to the time-domain wavelet-subspace projection of the time series $x_j(k)$ at a particular scale $j = [1, \ldots, J]$, namely

$$k_{\mathrm{w}_j} \quad \text{for } j = [1, \ldots, J], \text{ from equation (49)}, \tag{57}$$

one derived from the complementary scaling-space projection $\overline{x}_J(k)$ at the coarsest scale J, namely

$$\overline{k_{\mathrm{w}_j}} \quad \text{for } j = J, \text{ from equation (50)}, \tag{58}$$

and last, the pairs of triplets that result from the application of the timescale to time-domain mappings of equations (54)–(56) to the estimates of equations (51) and (52), and which we write here as

$$k'^{\perp}_{j,l_{\mathrm{w}}} \quad \text{for } j = [1, \ldots, J] \quad \text{and} \quad k'^{\perp}_{J,\overline{l_{\mathrm{w}}}}, \tag{59}$$

$$k'^{|}_{j,l_{\mathrm{w}}} \quad \text{for } j = [1, \ldots, J] \quad \text{and} \quad k'^{|}_{J,\overline{l_{\mathrm{w}}}}, \tag{60}$$

$$k'^{\top}_{j,l_{\mathrm{w}}} \quad \text{for } j = [1, \ldots, J] \quad \text{and} \quad k'^{\top}_{J,\overline{l_{\mathrm{w}}}}. \tag{61}$$

Each of these various scale-dependent estimates will be generically referred to as $\hat{k}_j$ if they serve to estimate a certain changepoint $k_\circ$ in a time series as modeled in the General Considerations section.

Multiscale AIC-based changepoint estimators

It is important to note that we are testing the performance of measurements made at a certain scale $j \neq 0$ that estimate changepoints $k_\circ$ that, under our model espoused in the

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

A Simple Model of Noise and Signal section, are changes at scale $j = 0$. This apples-to-oranges comparison is necessary because it might correspond to how our algorithm is used in practice. The comparison of how a particular $\hat{k}_j$ is successful at estimating a particular $k_{\circ_j}$ at the same scale j is *exactly* what has been covered in the Formalizing the AIC Timing Uncertainty section: after all, our model is that of a certain sequence that changes at a certain index, regardless of scale.

Should our model be that of a time series that had a changepoint at a particular scale, even multiple different ones, the multiresolution wavelet analysis would produce coefficient sequences that, thanks to mutual orthogonality, would be uncorrelated between different scales, and the AIC-based methodology would lead to the proper identification of the changepoint at the appropriate scale. An example would be where different seismic phases with intrinsically different time–frequency signatures (e.g., *P* or *SS*) would be properly identified in time and scale (Tibuleac *et al.*, 2003).

Timing uncertainty of changepoint estimators

The uncertainty estimates of the Formalizing the AIC Timing Uncertainty section are found in the domain in which the changepoint is estimated. For both time-domain and timescale-domain cases, the procedures of the Method I: Monte Carlo Resimulation and Method II: Using the Shape of the AIC Curve sections are followed at each individual scale using the domain-specific indexes of interest: k in the time domain, and l in the timescale domain.

For uncertainty estimation of multiscale changepoint estimates in the time domain, equations (32)–(43) are computed as written, acting on k, except they are repeated at every scale, j, and each can be considered reproduced here as such with that subscript.

For uncertainty estimation associated with scale-dependent changepoint estimates $\hat{l}$ made in the timescale domain, those same equations are computed with l in place of k, again at each scale j. In that case, the M1 statistic (equation 32) becomes $\hat{l} - l_\circ$, and the index spans relevant to the two M2 hypothesis tests (equations 38 and 39) become the timescale domain coefficient index spans, $1 + [l', \ldots, l'']$, and $1 + [l^\dagger, \ldots, l^{\dagger\dagger}]$. Finally, multiplication of these statistics by the length of the appropriate span, either $k'^{\star}_{j,l_\mathrm{w}}$ or $k'^{\star}_{j,\overline{l_\mathrm{w}}}$, maps them back to the time domain.

In summary, at any given scale a unit error associated with either k_{w_j} or $\overline{k_{\mathrm{w}_j}}$ maps to one time domain sample index; whereas a unit error of l_w or $\overline{l_\mathrm{w}}$ maps to the length of the time smear associated with the changepoint estimate itself, $k'^{\star}_{j,l_\mathrm{w}}$ or $k'^{\star}_{j,\overline{l_\mathrm{w}}}$. Therefore, uncertainty estimates associated with changepoint estimates derived in the timescale domain see their uncertainty dilate in concert with the wavelets and scaling functions themselves.

Multiscale AIC-changepoint timing uncertainty

Let us examine what happens when we follow the Monte Carlo resimulation method M1 detailed in the Method I: Monte Carlo Resimulation section, using many realizations of synthetic time series of length $N = 4000$ with a known true changepoint, $k_\circ$, at time-domain sample index 2000. Each synthetic time series was generated by concatenation at $k_\circ$ of two random samples drawn from Gaussian distributions with zero expectation, the variances of which differed by a prescribed factor. As with all previous synthetic time series, the first segment, time-domain sample indexes $k = [1, \ldots, k_\circ]$, was noise drawn from the standard normal distribution, $\mathcal{N}(\mu = 0, \sigma^2 = 1)$, whereas the second segment, $k = [k_\circ + 1, \ldots, N]$, was signal drawn from a density $\mathcal{N}(\mu = 0, \sigma^2 = \mathrm{SNR})$, in which SNR was varied between 2 and 1024, increasing in powers of two. We decompose each time series to five wavelet scales.

We evaluate the performance of a generic single-scale changepoint estimate via the statistics of the signed distance $\hat{k}_j - k_\circ$, the time-domain sample-index error. Figure 12 shows the result of 1000 tests for each $\hat{k}_j$ at every SNR considered. For ease of reporting, we label results for k_{w_j} and $\overline{k_{\mathrm{w}_j}}$, and likewise for all the other pairs of changepoint estimates ultimately derived from wavelet-space or scaling-space representations, respectively, on a common axis, where we will draw a bar over the scale number to identify its scaling-space nature: for example, results for k_{w_5} will be reported at the points marked 5, whereas results for $\overline{k_{\mathrm{w}_5}}$ will be drawn at $\overline{5}$ on the abscissa axis.

In each of Figure 12a–j, the colored curves connect the observed average error at every scale, marked as a filled diamond or circle, with the ticks extending vertically in both directions from the marker indicating the width of the distribution by mirroring their standard deviations. The curves are color-coded to differentiate the changepoint estimator: those derived in the time domain, k_{w_j} and $\overline{k_{\mathrm{w}_j}}$, the average errors of which are marked with purple diamonds; and those derived in the timescale domain and mapped back to the time domain, the average errors of which are marked with filled circles, either teal for $k'^{\perp}_{j,l_\mathrm{w}}$ or $k'^{\perp}_{j,\overline{l_\mathrm{w}}}$, red for $k'^{|}_{j,l_\mathrm{w}}$ or $k'^{|}_{j,\overline{l_\mathrm{w}}}$, or green for $k'^{\top}_{j,l_\mathrm{w}}$ or $k'^{\top}_{j,\overline{l_\mathrm{w}}}$. The axes for all of the panels are equal, with the ordinate axis representing the changepoint estimation error and bounding the span of −200 and 200 time-domain sample indexes, while the abscissa axis denotes the scale at which the M1 test was performed, with the overline notation for the scaling-space as discussed earlier.

The estimator pairs $(k'^{|}_{j,l_\mathrm{w}}, k'^{|}_{j,\overline{l_\mathrm{w}}})$ and $(k_{\mathrm{w}_j}, \overline{k_{\mathrm{w}_j}})$, perform the best, the former having acceptable bias even at high scales and low SNR, and the latter being better than the others at all but the lowest-SNRs. Regardless of SNR, all changepoint estimate pairs $(k_{\mathrm{w}_j}, \overline{k_{\mathrm{w}_j}})$ display either large biases and/or large standard deviations at the highest scale, $j = 5$. Of course, the "edge" changepoint estimators, $(k'^{\perp}_{j,l_\mathrm{w}}, k'^{\perp}_{j,\overline{l_\mathrm{w}}})$ and $(k'^{\top}_{j,l_\mathrm{w}}, k'^{\top}_{j,\overline{l_\mathrm{w}}})$, perform most poorly, showing growing negative (early) and positive (late) biases regardless of SNR, as the scales increase. They represent the two end-member cases for estimates made in the timescale domain and mapped back to the time domain. As mentioned, the compromise between the two, $(k'^{|}_{j,l_\mathrm{w}}, k'^{|}_{j,\overline{l_\mathrm{w}}})$,

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

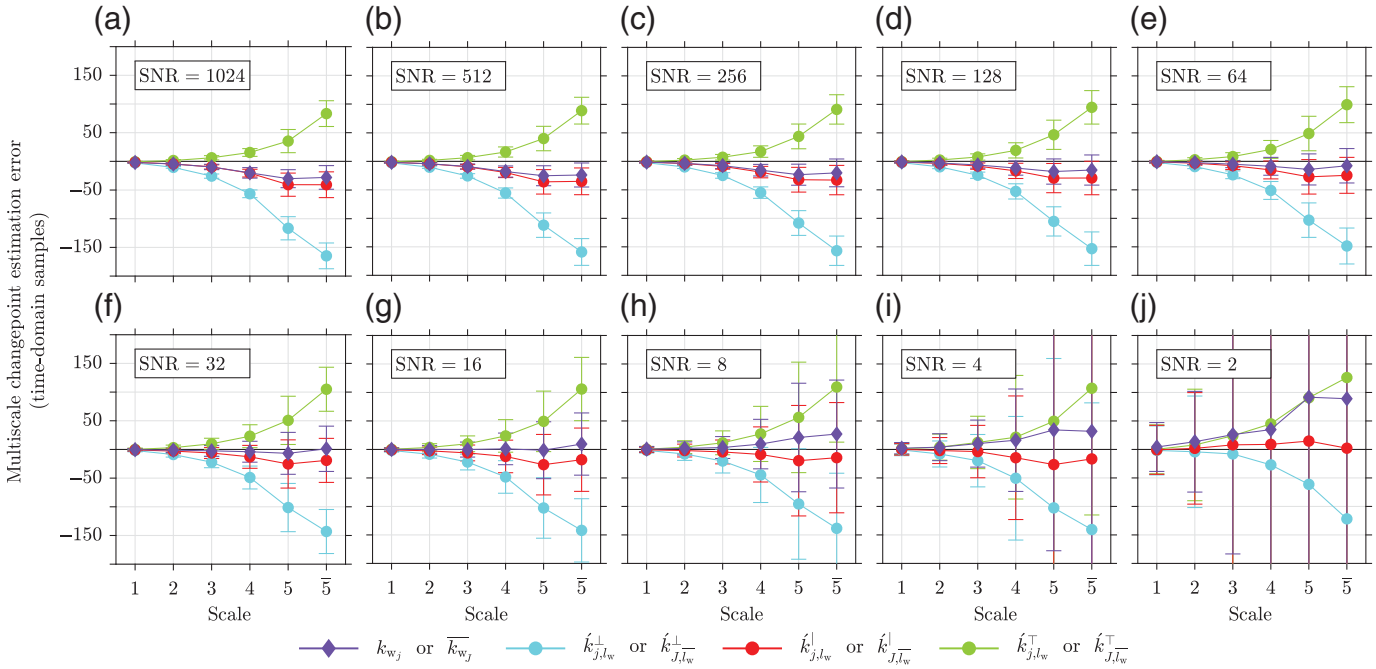


Figure 12. Multiscale changepoint estimation errors at various SNRs (equation 31) for timescale and time domain changepoint estimates at five wavelet scales, and one approximation scale, labeled $\overline{5}$. The individual panels (a–j) each have the same abscissa axis (low to high scale, or equivalently, high to low resolution) from left to right, and ordinate axis (−200 to 200 time-domain sample indexes), and are arranged from high- to low-SNR tests (noted in the upper left corner of each panel). In each of (a–j), the filled markers (circles or diamonds), connected as a curve, represent the mean changepoint estimation error (via the M1 procedure, $\hat{k}_j - k_\circ$) for 1000 test realizations in which $\hat{k}_j$ is one of: k_{w_j} or $\overline{k_{\mathrm{w}_j}}$ (purple), marked with filled diamonds; or $k'^{\perp}_{j,l_{\mathrm{w}}}$ or $k'^{\perp}_{J,\overline{l_{\mathrm{w}}}}$ (teal), $k'^{|}_{j,l_{\mathrm{w}}}$ or $k'^{|}_{J,\overline{l_{\mathrm{w}}}}$ (red), or $k'^{\top}_{j,l_{\mathrm{w}}}$ or $k'^{\top}_{J,\overline{l_{\mathrm{w}}}}$ (green), marked with filled circles. The ticks extend up and down from each mean estimate by once the standard deviation. Each test realization was performed on a synthetic time series of $N = 4000$ generated via the concatenation at $k_\circ = 2000$ of two random samples drawn from Gaussian distributions of zero expectation that differ in variance by the SNRs indicated in legends.

outperforms either of its two siblings, especially at low SNRs, though it does display bias at high SNRs.

Coherence across scales

A takeaway from Figure 12 is that all of the changepoint estimators $\hat{k}_j$ perform well at high resolution, when the scale is low, $j = 1$ or $j = 2$, regardless of SNR. Indeed, even at the lowest SNR tested, SNR = 2, all such estimators were on average within five time-domain sample indexes of the truth. The short time-domain support at the lowest scales lessens the mapped time smears of the basis functions, $k'^{\star}_{j,l}$. A shorter basis function always results in more precise changepoint estimates as long as the time window is sufficiently long to capture it.

As the scales increase timing resolution degrades because the basis functions dilate to capture longer-period features. The shape of the curves in Figure 12 reveals the twofold dilation of the basis functions at every scale. In particular for the CDF(2,4) basis that we used, the primal wavelets and scaling functions have support of three and nine time-domain sample indexes, respectively. At the highest scale, after five successive dilations, the effective support balloons to 153 for $\psi_{5,l}$ and 249 for $\phi_{5,l}$. In the frequency domain, a simple rule of thumb holds that the first scale of a wavelet-decomposed time series is approximately sensitive to the frequency band spanning 1/2–1/4 the sampling rate of the input time series. Each subsequent scale further halves the sensitivity frequency band of the previous, leading to a recursive relation approximating the frequency sensitivity of a wavelet at scale j as

$$\sim [f_s/2^j \leftrightarrow f_s/2^{(j+1)}], \tag{62}$$

in which f_s is the sampling rate of the input time series.

The behaviors in this section illustrate the coherence and the range over which it persists, between changepoint estimates made at different scales, that can be expected, and indeed, is often implicitly assumed or enforced, with the notion of there being a single onset happening at the raw sampling rate. In what we propose, our picks are single scale at multiple scales, and no interscale coherence is required for their validation.

Figure 12 may be used to intuit the accuracy of a specific changepoint method but should not be construed to represent the precision associated with each. It is presented partly to show that in extremely low-SNR cases, for example, SNR = 2, if accuracy is paramount to uncertainty estimation, one may wish to derive changepoint estimates in the timescale domain and tack their time-domain time smears to a single sample

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

index with the pairs $(k'^{|}_{j,l_w}, k'^{|}_{J,\overline{l_w}})$, instead of picking them in the time domain with the pairs $(k_{w_j}, \overline{k_{w_J}})$. However, although an estimator like $k'^{|}_{j,l_w}$ may be more accurate than k_{w_j} at low SNRs, at high scales its precision degrades.

For this reason, we adopt as our preferred arrival-time estimates k_{w_j} and $\overline{k_{w_J}}$, found at the subspace projections x_j and $\overline{x}_J$, rather than the estimates $k'^{\star}_{j,l_w}$ or $k'^{\star}_{J,\overline{l_w}}$ (or any individual sample within that span) found using the raw detail and approximation coefficients. In the Application to MERMAID Seismograms section, we apply our multiscale-AIC method in both domains and the differences in timing-uncertainty estimates will become clearer.

Uncertainty estimation in practice, with real data when $k_\circ$ is not known, is discussed next, with particular emphasis given to the problem of uncertainty estimation associated with changepoint estimates derived in the timescale domain. In that case, the assumed "truth," $l_\circ$, must exist distinctly at every scale j in the timescale domain, which differs from the error measurements of this section, which occurred after mapping l_w and $\overline{l_w}$ back to the time domain and assigning to each of them a single time domain sample index against which signed distances could be measured.

From synthetics to actual data

Until now, we have discussed changepoint estimation methods and timing uncertainty appraisal on synthetically modeled data, for which the true changepoint $k_\circ$ or $k_{\circ_j}$ is known. With actual data, which we will discuss in the Application to MERMAID Seismograms section, we replace $k_{\circ_j}$ with the changepoint estimate made on the real data at the relevant scale: one of either $(l_w, \overline{l_w})$ or $(k_{w_j}, \overline{k_{w_J}})$ depending on the domain in which it is determined.

Synthetic seismograms generated for uncertainty estimation via the M1 resimulation and M2 curvature analysis methods of the Formalizing the AIC Timing Uncertainty section are constructed in the same domain that the changepoint estimate was made, at every scale. At each scale, many randomized synthetic seismograms are constructed by concatenation at $(l_w, \overline{l_w})$ or $(k_{w_j}, \overline{k_{w_J}})$ of two samples drawn from two different and distinct Gaussian distributions: $\mathcal{N}(\hat{\mu}_1, \hat{\sigma}_1^2)$ and $\mathcal{N}(\hat{\mu}_2, \hat{\sigma}_2^2)$, the parameters of which are computed from the noise and signal segments, respectively, of the seismogram at that scale. The means and variances $\hat{\mu}$ and $\hat{\sigma}^2$ pertain to the coefficients $(d_{j,l}, a_{J,l})$ in the timescale domain, or to the projections $(x_j, \overline{x}_J)$ in the time domain, as segmented into noise and signal by the changepoint estimators.

THE DATA SET

The parameters of many of our examples shown thus far to illustrate our procedures and methodology have been appropriate for a very specific data set collected by a very specific type of seismological instrument. In this section, we turn to describing MERMAID before applying our techniques to seismograms collected at sea in the Application to MERMAID Seismograms section.

MERMAID is an autonomous ocean-going diver with a hydrophone that continuously records and processes the ambient acoustic wavefield at midcolumn depths. Its primary goal is to monitor worldwide earthquake activity, and specifically, to provide arrival times of teleseismic waves suitable for global seismic tomography (Simons, Nolet, *et al.*, 2006; Hello *et al.*, 2011). A subset of the data collected, all of the data that we analyze in this study, are publicly available in Seismic Analysis Code and miniSEED formats (see Data and Resources).

The instrument

While at depth, the acoustic data: a hydroacoustic time series, hereafter the "seismogram" (Joubert, Nolet, Sukhovich *et al.*, 2015), are filtered between [0.10–10] Hz and digitized in real time at a sampling rate of 40 Hz. The digitized data are then immediately processed by an STA/LTA algorithm (Allen, 1978) to identify segments of interest (Simons *et al.*, 2009) where a possible signal rises above the level of the noise. When a predetermined STA/LTA trigger threshold is exceeded, MERMAID passes a windowed segment containing pre- and post-trigger data into the wavelet-based (Simons, Nolet, *et al.*, 2006; Sukhovich *et al.*, 2011) detection algorithm of Sukhovich *et al.* (2014). This discrimination procedure inspects the energy partitioning between different wavelet scales and assigns a criterion value to the hypothesis that the waveform includes a *P*-wave arrival generated by a teleseismic earthquake, and not some other energy generated by a nonseismic source (e.g., ship propellers or ocean storms, which distribute energy differently over various scales). The on-board algorithm decomposes the 40 Hz data to six wavelet scales via the same lifting algorithm (Sweldens, 1996) and CDF(2,4) wavelet basis (Cohen *et al.*, 1992) as in this study. If the signal is deemed to be a teleseismic arrival, a 200–250-second-long seismogram, containing the STA/LTA trigger at about 100 s, is returned to shore via the Iridium satellite constellation at MERMAID's next surfacing. Because MERMAID freely drifts with the ocean currents its location at the time of recording the seismogram must be interpolated from multiple Global Positioning System locations fixed at the surface. The exact details of this procedure are described by Joubert, Nolet, Bonnieux, *et al.* (2015).

The seismograms

By default, MERMAID transmits the raw detail and approximation coefficient series at scales 2–6, omitting scale 1 to save data-transfer cost, and because scale 1 is not particularly useful for teleseismic *P*-wave analysis (at $f_s = 40$ Hz, it covers the frequency band spanning roughly [10–20] Hz). The seismogram is then reconstructed onshore via the inverse wavelet transform at an effective sampling rate of $f_s = 20$ Hz. In this study, we decompose these 20 Hz seismograms to five scales so as to

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

analyze the data at the same resolution as the on-board detection algorithm. In some cases, only three out of six wavelet coefficient sets are being returned, which leaves us with $f_s = 5$ Hz seismograms.

Table 1 lists the approximate frequency bands to which each wavelet is sensitive at the five scales in our numbering scheme for seismograms sampled at 20 Hz and 5 Hz. Later, we compute arrival time estimates and their uncertainties, and travel-time residuals considering the entire public MERMAID catalog. So as to compare arrival-time picks made at the same resolutions, scale 1 for 5 Hz seismograms will be analyzed with scale 3 for 20 Hz seismograms, scale 2 for 5 Hz seismograms will be analyzed with scale 4 for 20 Hz seismograms, and so on.

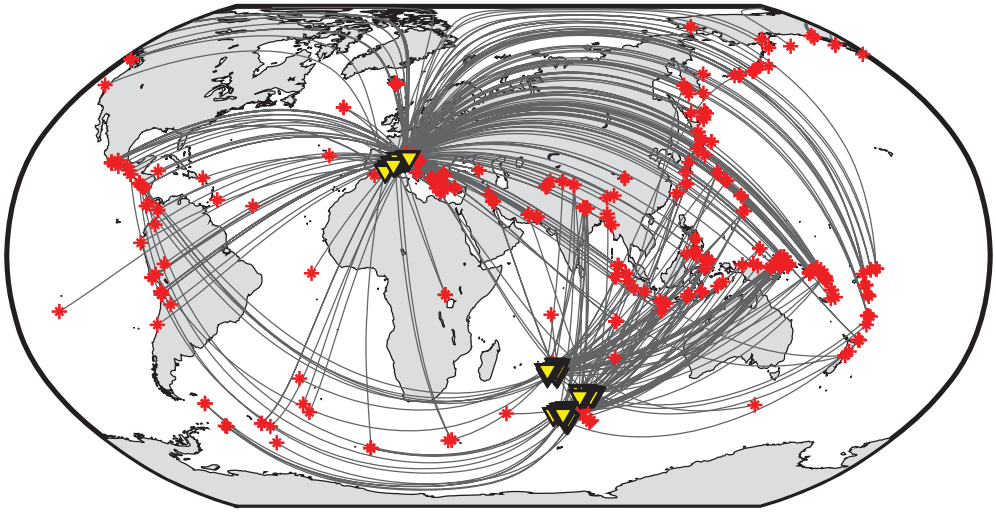


Figure 13. Global map of all events and stations used in this study. Great-circle paths connect the event locations (red asterisks) to the Mobile Earthquake Recording in Marine Areas by Independent Divers (MERMAID) positions (yellow triangles) at the time the seismogram was recorded.

The initial events catalog

Prior work (Joubert, Nolet, Bonnieux, *et al.*, 2015; Sukhovich *et al.*, 2015; Nolet *et al.*, 2019) has resulted in a catalog of "identified" events (and a complement with "unidentified" seismograms). The classification indicates whether the seismograms were matched to known seismic events, that is, earthquakes, as determined by querying published seismic catalogs available at the time. In this study, we focus our attention on 445 identified MERMAID seismograms. These seismograms represent global earthquakes recorded at disparate times between December 2012 and February 2018 and various (time-variable) locations both in the Indian Ocean and Mediterranean Sea.

We maintained the event identifications but updated their details to the most up-to-date information available from the Incorporated Research Institutions for Seismology (IRIS). Because the original event identification consisted of an epicentral location, origin time, and magnitude, we maintained the match by querying the latest catalogs within a buffer of 30 s and 1° in time and location, and of no more than one magnitude unit less than the original. In the few rare instances where our match was nonunique using these criteria, we selected the one provided by the International Seismological Centre (2016) online bulletin. Only a handful of cases required manual intervention to complete our revised event catalog.

Figure 13 displays the updated locations of our events (red asterisks) and the interpolated positions of the MERMAID instruments that recorded them (yellow triangles). Great circles connect event–station pairs to give a sense of the geographic areas sampled most by our data set. Although only 13 individual MERMAID instruments are present, there are 445 unique receiver locations plotted, illustrating their passive drift with the ocean currents.

We computed arrival times of various seismic phases in the ak135 velocity model of Kennett *et al.* (1995). We first computed the theoretical travel times with MatTaup, specifically using the taupTime method. These codes are dated November 2002 but were distributed without a version number. They are based on TauP, described by Crotwell *et al.* (1999). The arrival time was obtained as the updated event time plus the theoretical travel time, minus the time at the first sample of the seismogram.

Inspired by our experience with our own data sets (e.g., Simons *et al.*, 2009; Sukhovich *et al.*, 2015; Nolet *et al.*, 2019),

TABLE 1
Approximate Frequency Sensitivity of Wavelets

Scale	$f_s = 20$ Hz	$f_s = 5$ Hz
1	[5.00–10.0] Hz	[1.25–2.50] Hz
2	[2.50–5.00] Hz	[0.62–1.25] Hz
3	[1.25–2.50] Hz	[0.31–0.62] Hz
$\overline{3}$	—	[0.1*–0.31] Hz
4	[0.62–1.25] Hz	—
5	[0.31–0.62] Hz	—
$\overline{5}$	[0.1*–0.31] Hz	—

The approximate frequency bands (equation 62) to which individual wavelets are sensitive to, based on the sampling rate (f_s) in Hz of the Mobile Earthquake Recording in Marine Areas by Independent Divers (MERMAID) seismogram.
*The approximation, noted here with an overline at scale 3 or 5 depending on the sampling rate of the seismogram, in theory approximates a low-pass filter (with sensitivity down to 0 Hz), but in practice only has sensitivity to 0.1 Hz because MERMAID data are filtered onboard between [0.10–10] Hz before digitization (see The Instrument section).

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

and by work conducted by other researchers on hydroacoustic time series elsewhere (e.g., Bohnenstiehl *et al.*, 2002; Smith *et al.*, 2002; Dziak *et al.*, 2004; McGuire *et al.*, 2012), we consider the following phases likely to be present in MERMAID seismograms: *p*, *P*, *pP*, *PP*, *Pn*, *Pg*, *PcP*, *Pdiff*, *PKP*, *PKiKP*, *PKIKP*, *s*, *S*, *Sn*, *Sg* (using the phase-naming convention of Crotwell *et al.*, 1999). We allowed the computation of arrival times of extraordinary phases (e.g., *SKiKP* for seismograms m31.20140910T053727.sac and m33.20150916T142424.sac) only if they were listed in the events.txt file distributed with the public MERMAID catalog. Such instances were rare, and we reserve the discussion on possible phase ambiguity for the Computing Travel-Time Residuals section.

APPLICATION TO MERMAID SEISMOGRAMS

We apply our AIC-based multiscale methods to the MERMAID data set described. We detrended the seismograms and trimmed the last sample index to render all of them of even length. We used the CDF(2,4) wavelet transform to five scales for the $f_s = 20$ Hz seismograms and to three scales for the $f_s = 5$ Hz set (see Table 1).

Multiscale AIC measurements and their uncertainty

Figures 14 and 15 provide the first illustrations of the determination of an independent set of multiscale arrival-time estimates and their uncertainties from which to compute residuals with the theoretical arrival times in the catalog. For these two example earthquakes in the Bali Sea and the Tyrrhenian Sea, which yielded a low-SNR and a high-SNR seismogram, respectively, the top set of panels, labeled (a), illustrate changepoint estimates made in the time domain (on the subspace projections), whereas the bottom set, labeled (b), illustrate changepoint estimates made in the timescale domain (on the coefficient series) as explained in the Multiscale AIC-Based Changepoint Estimation section.

In both figures, for both sets (a) and (b), the topmost panel plots the MERMAID seismogram, x, in blue, normalized between −1 and 1. Earthquake magnitude, epicentral distance (in degrees) to the recording MERMAID, and event depth are listed in the legends. For the set (a), the panels below the raw seismogram are the subspace projections after wavelet decomposition $(x_j, \overline{x}_J)$, in gray, normalized between −1 and 1. For the set (b), the panels below the first one show the absolute values of the detail and approximation coefficients, $(d_{j,l}, a_{J,l})$, in gray, smeared over the time spans to which they are sensitive, $k'^{\star}_{j,l}$, and normalized between 0 and 1. The abscissa axes, in seconds offset from the start of the seismogram ($t = 0$ at $k = 1$), are unchanged between subplots. The subspace projection series in (a) are at least as truncated as the raw coefficient series in (b), but usually more so, due to the increased influence of the edges in the time domain compared to the timescale domain, as explained in the supplemental material.

Overlain in black are the AIC functions, normalized per panel, as time-domain traces for the set (a), and as black

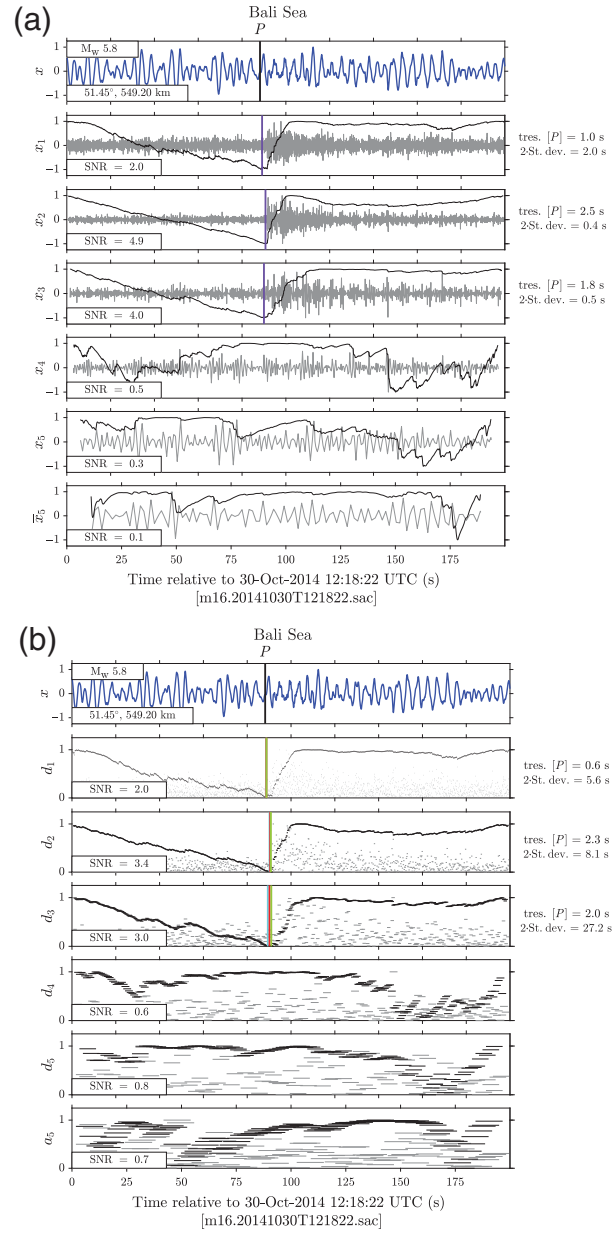


Figure 14. Multiscale arrival-time estimation of a seismogram from the MERMAID data set in the low-SNR regime: a seismogram detected in the Mauritius–Réunion region corresponding to an earthquake in the Bali Sea. In both (a,b), the top panel plots the same raw seismogram (normalized between −1 and 1) in blue with the same theoretical arrival time of the *P* wave marked with a black vertical line. Legends provide magnitude, great-circle distance between the epicenter and MERMAID, and the depth of the corresponding earthquake. The panels below the first one show, for the set (a), subspace projections at varying wavelet scales in gray (normalized between −1 and 1), with their AIC curves overlain in black. In the set (b), below the seismogram, the panels show the absolute values of the wavelet and scaling coefficient time series (normalized between 0 and 1), and their AIC function, rendered as the corresponding time smears. Event detections were made at scales 1–3, in the time domain in set (a) where they are marked in purple, and in the timescale domain in set (b) using teal, red, and green to mark the beginning, middle, and end of the arrival time smears. The two standard deviations of the error distributions, obtained by Monte Carlo resimulation, are listed to the right of the ordinate axis. Scales 4 and 5 each had SNRs less than or equal to 1, and thus no arrival time is reported.

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

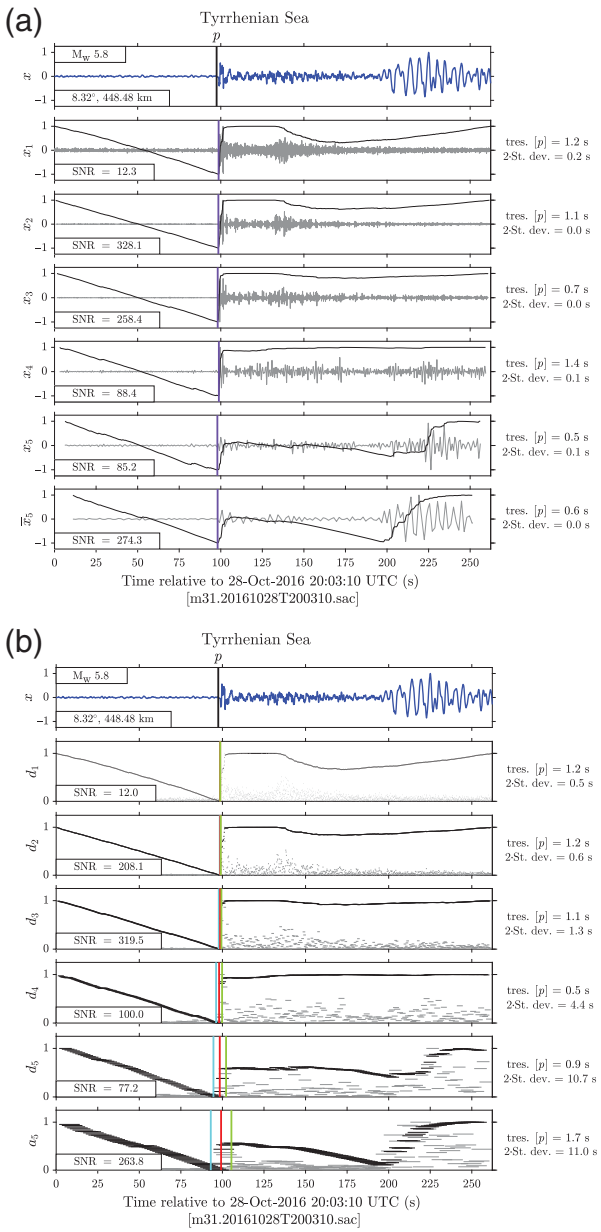


Figure 15. Multiscale arrival-time estimation of a seismogram from the MERMAID data set in the high-SNR regime: a seismogram detected near the Balearic Islands, Spain, corresponding to an earthquake in the Tyrrhenian Sea. Arrangement and labeling are as in Figure 14. Here, we see clear *P*-wave detection coherently across all scales. The increase in low-frequency energy around 200 s is the arrival of an *S* wave. This arrival is especially apparent at scale 5, where the AIC function dips to nearly the same low value (high likelihood of a second changepoint) as obtained at the *P*-wave pick, hinting at the prospect of an extended utility of our method by recursive implementation; that is, reapplication of our AIC-based arrival-time detector on a shortened $(x_j, \bar{x}_J)$ or $(d_{j,l}, a_{J,l})$ series that begins immediately after the initial *P*-wave detection. Also note here that although the ak135 travel-time residuals in (a,b) are similar at every scale, their estimated uncertainties are not, (b) with those made in the timescale domain much greater than (a) those made in the time domain. This is due to the dilation of the wavelets themselves at increasing scales, which manifests as stretched time smears, and which is well illustrated in (b) by the increased separation between the teal and green vertical bars that mark the beginning and end of the timescale coefficient smears of the estimated arrivals.

horizontal dashes over the corresponding time smears per scale in (b). The vertical lines mark the changepoint estimates: purple for the subspace-projection estimates in (a), and teal, red, and green for the coefficient series estimates in (b). The latter triplet of colors marks the beginning, middle, and end of the time smears, $(k'^{\star}_{j,l_w}, k'^{\star}_{J,\overline{l_w}})$, of the estimated arrival-time detail or approximation coefficients, as detailed in the Timescale to Time Mapping: Forward Transform section. These are the same colors used in Figure 12, but here we mark the estimated arrival times, and there we analyze the statistics of the estimated changepoints (the mapping between the two is discussed in the Multiscale Changepoints to Multiscale Arrival Times section). The dilation of the wavelet with increasing scales is easily seen when plotted in this manner, especially in Figures 15b, 16b, and 17b, where arrival-time estimates were made at every scale (all SNR > 1), and the time smear of the basis function is seen to lengthen at every scale to over 12 s at scale $\overline{5}$.

The corresponding SNR at each scale is labeled in the lower-left corner legend of each panel, and when it is smaller than or equal to one no arrival-time estimate is reported. The two-standard deviation of the error distribution (after 1000 test realizations at each scale) is to the right of the ordinate axis, in seconds, as determined using the M1 test of the Method I: Monte Carlo Resimulation section.

Figures 16 and 17 contain additional examples, presented in the same layout and with legends and labeling as in Figures 14 and 15. Figure 16 successfully separates core phases *PKIKP* and *PKP*, and Figure 17 separates *P* from *PP*. One more example is given in the supplemental material.

Computing travel-time residuals

We make the assumption that the theoretical phase arrival nearest in time to our AIC-based arrival-time estimate corresponds to the true seismic phase identified by our method, which gives us a scale-dependent travel-time residual against the ak135 velocity model. In Figures 14–17, the travel-time residuals and their associated phases are quoted to the right of the ordinate axes, at each scale. There will be ambiguities: for example, in Figure 17a at wavelet scale 5, where the arrival time is tagged as a *pP* wave instead of the preceding *P* wave, which is arguably more likely, and which one might disambiguate by taking amplitude information, or other attributes, into account. Similarly, the arrivals in scales 1–4 in Figure 17b might derive from either of two *P* waves, from the same earthquake, with predicted arrival times very near each other.

As explained in the Take-Home Message 1 section, the MLE procedure that underpins our AIC-based method selects the index in the time series for which the ratio of variances of the signal and noise segmentations is the largest; it maximizes the SNR after testing all possible combinations. Without iteration, our method only allows for the identification of a single seismic phase at each scale. This is generally appropriate in our

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

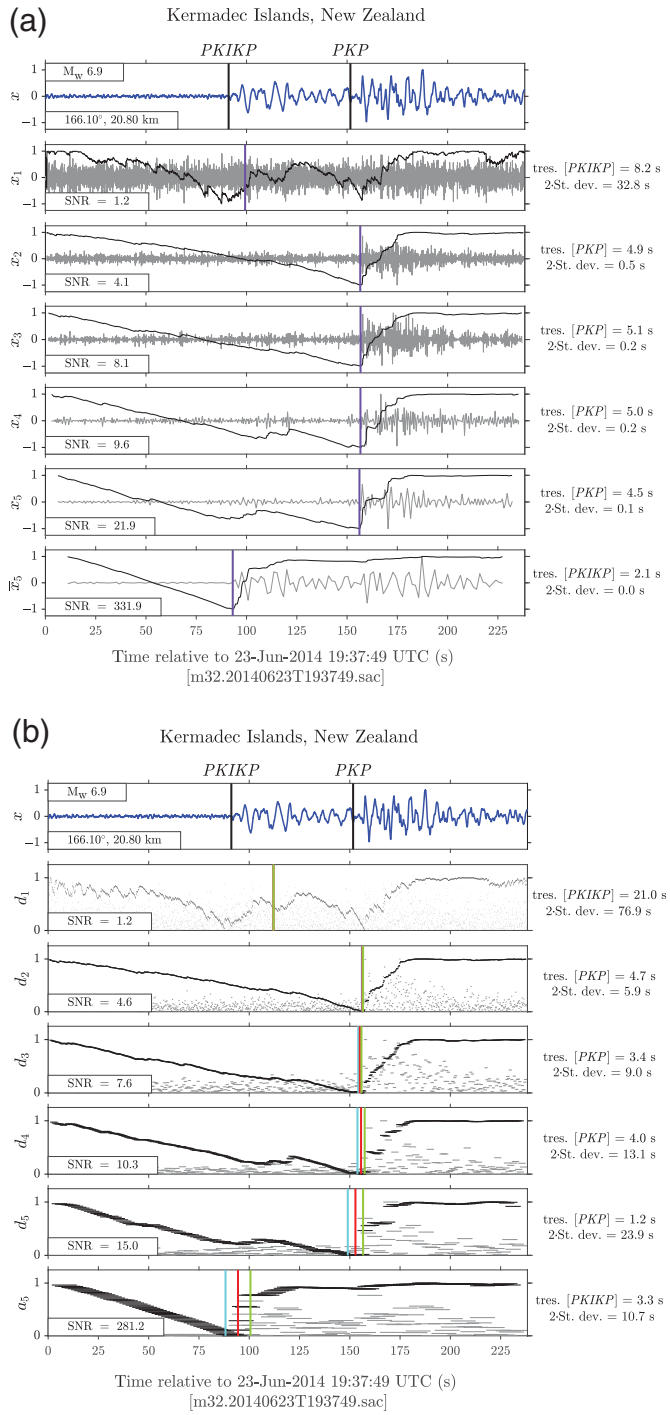


Figure 16. A mid-SNR seismogram detected near the south coast of France corresponding to an earthquake in the Kermadec Islands, New Zealand. This is the first illustration of our procedure resulting in multiphase detection of two distinct core phases at separate scales: *PKP*-wave detections at scales 2–5 (at the resolution of the details) and *PKIKP*-wave detection at scale $\overline{5}$ (at the resolution of the approximation). There is an arrival pick at scale 1, though its SNR is low and thus its uncertainty is high. Such a pick is not considered high quality and is not included in Figure 18 or Table 2.

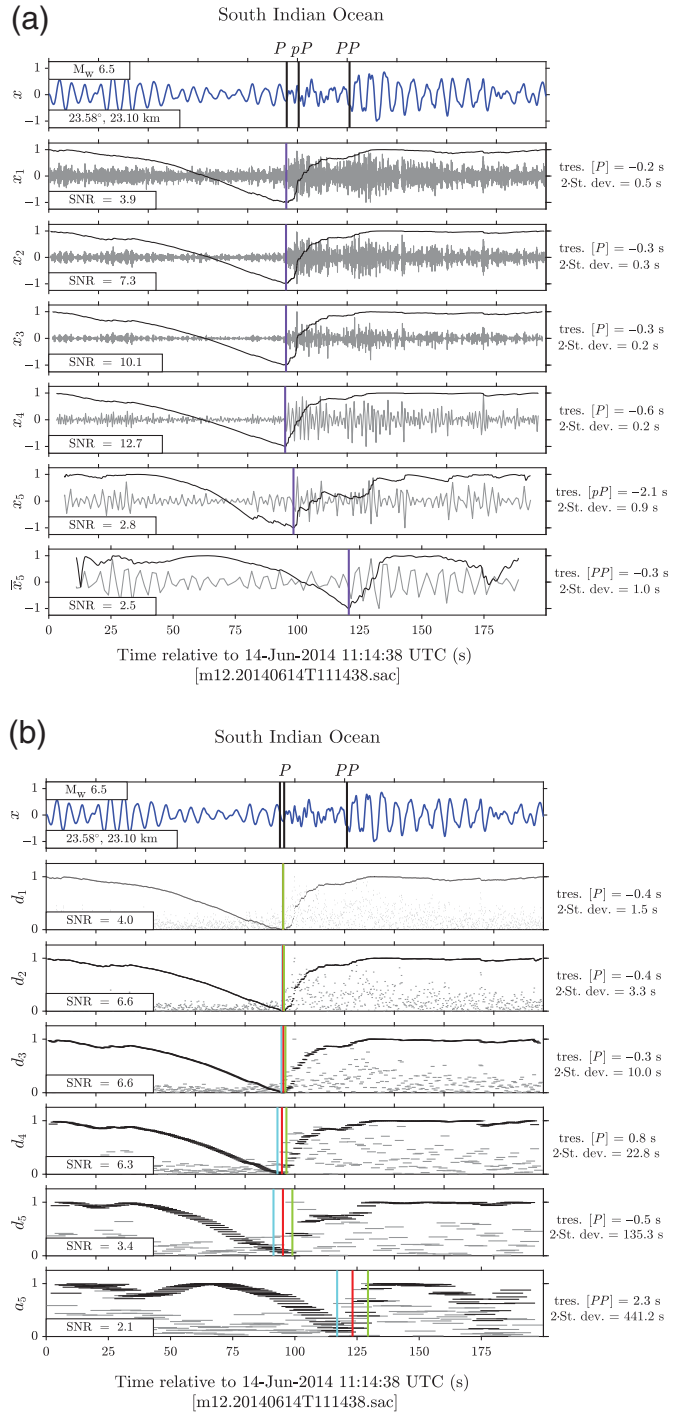


Figure 17. A low-SNR seismogram detected over Broken Ridge corresponding to an earthquake in the southern Indian Ocean. Here, we see the matching of different phases at scale 5 between arrival-time picks made in the two domains: (a) a *pP* wave in time, and (b) a *P* wave in timescale. Furthermore, in (b) there are two distinct *P* waves with ak135 arrivals very near in time to each other, one or the other of which minimizes the travel-time residual at one of the scales 1–5. Unresolved ambiguities of this sort are discussed in the Computing Travel-Time Residuals section.

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

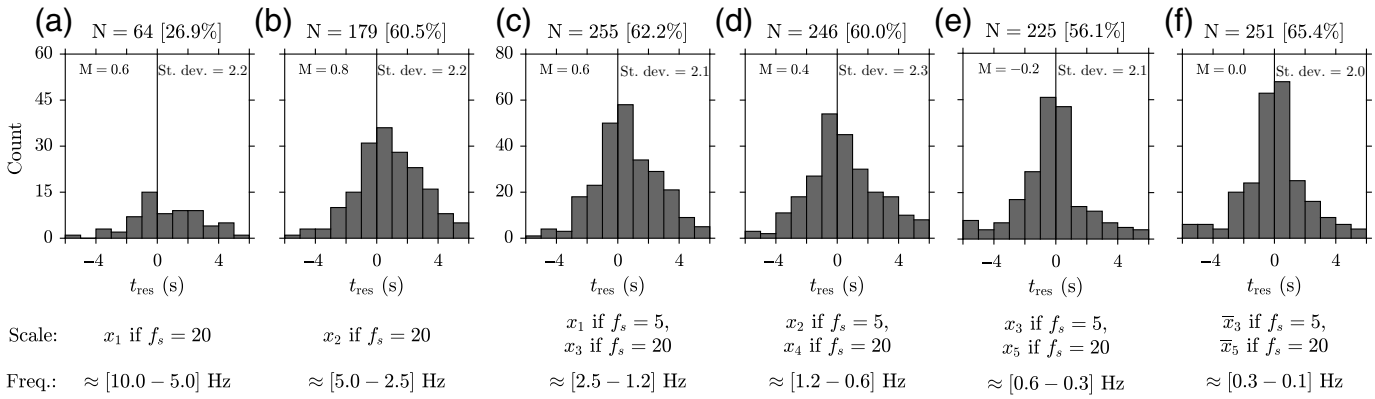


Figure 18. High-quality travel-time residuals in the public MERMAID catalog. (a–f) Each is a histogram of residuals from high to low resolutions. Each histogram displays only travel-time residuals for which the absolute values are less than or equal to 6 s and for which the two-standard-deviation error by the M1 method of the Method I: Monte Carlo Resimulation section is smaller than 1 s. In each panel, the mean and standard deviation of the residuals at that resolution are listed inside, the total number of residuals and their percentage relative to all arrival-time estimates (SNR > 1) at that resolution are listed above, and the scales considered and their approximate frequency sensitivities (Table 1) at that resolution are listed below.

case, given that a MERMAID seismogram most typically includes a single *P*-wave arrival, as in Figure 14, or a single picked *P*-wave arrival, as in Figure 15. Furthermore, the frequency bands of our wavelet decomposition are sufficiently narrow to successfully partition the arrival of distinctly pickable seismic phases over separate scales, as exemplified by Figures 16 and 17.

Picking up additional arrivals, should they contain substantial energy within the same scale, can be accomplished, if further decomposition to higher scales is not an option, by iterating our method on the identified signal segment. A good candidate is shown in Figure 15, where the energy increase near 200 s is an *S*-wave arrival drowned out at wavelet scale 5, and in the shadow of the coda of the preceding *P*-wave arrival at approximation scale $\overline{5}$. Our algorithm could be run recursively at those scales, on a truncated time window beginning just after the *P*-wave arrival so as to bracket only the segment containing the *S*-wave arrival.

Distribution of travel-time residuals

We apply the procedure illustrated in Figures 14–17 to the complete data set of 445 identified MERMAID seismograms. We work in the time domain using the subspace projections $(x_j, \overline{x}_J)$ per the timing considerations discussed in the Multiscale Analysis in Practice section. For every seismogram at each scale, we compute: (1) an AIC-based arrival-time estimate; (2) the uncertainty associated with that arrival time using 1000 realizations of method M1; (3) the travel-time residual by considering all phases that might arrive in the time window of the seismogram, retaining the minimum travel-time residual and its associated phase.

Figure 18 shows the multiscale distribution of travel-time residuals in our data set. We only show those that we deem to be of "high quality," falling within 6 s of a theoretical phase arrival, and for which the two-standard deviation uncertainties per the M1 method are smaller than 1 s. These quality criteria resolve some of the possible issues with phase ambiguity, guarantee sufficient SNR to reduce the likelihood of falsely triggering on spurious energies not related to a phase arrival, and, last, they produce histograms that are in line with expectations for mantle *P* waves (e.g., Simmons *et al.*, 2012) without generating long tails beyond the 6 s cutoff.

Figure 18a–f plots histograms of travel-time residuals at scales 1–5 for 20 Hz seismograms. Figure 18c–f also contains the residuals at scales 1–3 for 5 Hz data, so as to compare residuals made at the same resolution. In each panel, the mean of the residuals is quoted in the upper left, and their standard deviation in the upper right. The scale(s) that contributed data to the histogram, and the approximate frequency bands spanned by these scales (see Table 1), are quoted also. Table 2 is a breakdown of the phases included in the histograms of Figure 18.

Although we are confident that, taken as a whole, the residuals presented in this study and summarized here faithfully record the signal of the Earth, we are aware that false triggers and mismatched phases may exist in our catalog, as is true for all automated arrival-time identification and phase-picking methods. Future work will necessarily include waveform modeling to better tack our AIC-based arrival times to their associated phases in MERMAID seismograms.

Of the 445 seismograms considered, 339 were sampled at 20 Hz and 106 at 5 Hz. The limits of the ordinate axes in Figure 18a–f are adjusted to reflect the fact that roughly 25% more data are available for consideration in Figure 18c–f than in Figure 18a,b. The total numbers of arrival-time estimates (SNR > 1) and residuals computed at each resolution of Figure 18a–f are 238, 296, 410, 410, 401, and 384, with the numbers of high-quality residuals actually plotted displayed above each plot. The bracketed numbers are the percentages of high-quality residuals contained in the histogram, relative to the total

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

```
Filename m18.20130417T032249.sac
IRIS Event ID 9316474              SOUTHWESTERN PAKISTAN
Last updated 2019/03/22

             Date        Time      Latitude Longitude  Depth Distance Magnitude
Initial: 2013/04/17 03:15:52.30  28.1233   62.3378  59.90  47.4858   5.7
Updated: 2013/04/17 03:15:52.30  28.1233   62.3378  59.90  47.4858   5.7  MW

GeoAzur phase pick: P

JDS multiscale phase picks: 3 scales at 5 Hz
Phase  Time    Tres      SNR        Mu  2Sigma
    P  89.40   -2.32 1.310E+01  -0.01   0.64
    P  90.00   -1.72 6.053E+00  -0.02   1.21
    P  90.80   -0.92 4.540E+00   0.01   1.71
    P  38.00  -53.72 2.353E+00   0.86   9.46
_____________________________________________________________________

Filename m17.20130417T090915.sac
IRIS Event ID 9316859     NEAR S. COAST OF HONSHU, JAPAN
Last updated 2019/03/22

             Date        Time      Latitude Longitude  Depth Distance Magnitude
Initial: 2013/04/17 08:57:34.76  34.0355  139.3186  14.60  91.9820   5.7
Updated: 2013/04/17 08:57:34.76  34.0355  139.3186  14.60  91.9820   5.7  MW

GeoAzur phase pick: P

JDS multiscale phase picks: 3 scales at 5 Hz
Phase  Time    Tres      SNR        Mu  2Sigma
    P  87.00   -0.63 1.732E+00   0.47  15.26
   pP  93.80    1.28 2.768E+00   0.03   3.56
   pP  90.80   -1.72 3.023E+00  -0.02   3.19
    P  37.40  -50.23 1.716E+01   0.01   0.55
_____________________________________________________________________
```

Figure 19. Two event blocks in the catalog that accompanies this study in the supplemental material. The upper portion of each block (ending with "Updated") lists event metadata, including the originally reported event information recorded in the Seismic Analysis Code (SAC) header, and the updated event information per the last query to Incorporated Research Institutions for Seismology (IRIS), noted as the "Last updated" date. The lower portion of each block lists phase information, beginning with the phase associated with this seismogram as initially reported without an arrival time. Next we list, from the lowest to the highest scale: the phase name; arrival time; travel-time residual; SNR; and the mean and two standard deviation of the error estimate derived from 1000 realizations of the M1 method, identified by our arrival-time estimation procedure. Considering both events shown, only the estimated travel-time residual (−2.32 s) and phase (*P*) associated with the arrival-time estimate made at scale 1 for the top event is included in Figure 18 and listed in Table 2. All other travel-time residuals and/or the two-standard-deviation errors of the uncertainty estimates are too great to be considered of "high quality."

available at each resolution. We immediately see that the lowest percentage of high-quality residuals is computed at scale 1 for 20 Hz data. Otherwise, the percentage of high-quality residuals hovers near or just above 60% at all other resolutions. Computing these travel-time residuals is the first step in the inverse problem of tomography, and their bias alerts us to velocity perturbations encountered along the ray paths. From the discussion in the Method I: Monte Carlo Resimulation section, we know that our AIC-based changepoint estimator is unbiased. Thus, these residuals are not an artifact of our arrival-time estimation procedure but rather a direct measure of velocity anomalies in the Earth.

The updated events catalog

We have compiled an updated events catalog of all of the seismograms in our data set, without applying any quality criteria. Figure 19 shows two entries of the updated catalog available in the supplemental material. The upper half of each block (ending at the line beginning "Updated") lists general event information, including the date that the IRIS database was last queried for metadata related to that event, and the "Initial" (which contained no magnitude type) and "Updated" event parameters as returned on the date of the last query. The initial and updated event information are often (but not always) the same, as shown in the examples here.

The lower half of each block in Figure 19 lists phase-arrival information starting with the name of the phase initially associated with this seismogram as reported in the events.txt file distributed by GéoAzur, but reported there without an arrival time. The lines that follow list the multiscale arrival picks, travel-time residuals, and uncertainties found in this study. The first alerts the reader to the number of scales used in the wavelet decomposition: always either five if the sampling rate was 20 Hz or three if the sampling rate was 5 Hz. Next, a header line describes the forthcoming columns: the phase name associated with this travel-time residual; the observed arrival time in seconds into the seismogram, computed by our changepoint estimation procedure; the time residual in seconds between our arrival-time pick and the theoretical arrival time of the listed phase; the SNR of this arrival; and the mean and two standard deviation of the estimated error distribution in seconds, per the M1 procedure of the Method I: Monte Carlo Resimulation section after 1000 realizations. Each line corresponds to one scale starting at scale 1 and ending at the scale corresponding to the resolution of the approximation, either scale $\overline{5}$, or scale $\overline{3}$ as shown here.

TABLE 2
Phases in Figure 18

Phase	(a)	(b)	(c)	(d)	(e)	(f)	Total
P	44	106	134	129	100	110	623
pP	6	26	48	46	43	33	202
PcP	8	14	25	22	27	29	125
PKIKP	0	10	15	18	20	33	96
PKP	1	10	15	16	13	6	61
PKiKP	0	7	10	7	8	6	38
PP	1	2	2	2	5	9	21
Pn	2	1	1	1	0	6	11
p	1	1	2	1	1	2	8
Pdiff	0	0	0	0	0	1	1
S	1	1	2	3	7	14	28
Sn	0	1	1	1	1	1	5
SKiKP	0	0	0	0	0	1	1
Total	64	179	255	246	225	251	

Identified phases associated with the high-quality travel-time residuals of Figure 18. The columns correspond to the histograms in Figure 18a–f, listed from fine to coarse resolution. The last column sums the number of high-quality identifications across all scales for the phase specified. The bottom row totals the number of high-quality travel-time residuals collected at each scale and is the same number listed above each histogram in Figure 18. We did not actually observe an arrival associated with the phase *SKiKP* (nor *PKiKP*, see e.g., in the supplemental material) but kept the entry as an example of the issue of phase ambiguity associated with our phase-matching scheme. See the discussion in The Initial Events Catalog and Computing Travel-Time Residuals sections.

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

For the event described in the first block in Figure 19, only the travel-time residual and two-standard deviation of the uncertainty estimate made at scale 1 are smaller than 6 and 1 s, respectively. Thus, that is the only travel-time residual included in Figure 18, with its associated phase, *P*, listed in Table 2. The estimates made at all other resolutions are characterized by a large travel-time residual and/or a high two-standard-deviation uncertainty, and thus are not plotted in Figure 18 or listed in Table 2. Similarly, none of the estimated arrival times or their associated phases in the second event block in Figure 19 are plotted in Figure 18 or listed in Table 2, because the two-standard-deviation uncertainty at scales 1–3 are too large to be considered likely arrivals, and the travel-time residual at scale $\overline{3}$ is too great to truly match the assumed *P*-wave arrival listed there.

CONCLUSION

We have developed a method to detect arrival times of seismic energy at multiple scales in noisy seismograms and estimate their uncertainties. Our procedure centers on the wavelet-multiscale application of an AIC-based changepoint-detection scheme, which we have applied to the problem of computing travel-time residuals in low-SNR hydroacoustic records. Our uncertainty estimation procedure provides a quantitative metric for weighting residuals during tomographic inversions.

We have defined two changepoint estimators useful for arrival-time identification: the minimum, k_m, and the weighted average k_w, respectively, of the AIC function. We investigated two methods to compute the uncertainty associated with these estimates, one using brute-force Monte Carlo resimulation, and another through the analysis of the shape of the AIC curve itself. The former gives an estimated error distribution, whereas the latter assigns confidence intervals to the estimates. We showed that the weighted-average k_w estimator is unbiased, unlike the oft-used k_m estimator. For this reason, we suggest the adoption of the former. We discussed the nuances of applying our procedure in the time and time-scale domains and recommend the former to ensure the lowest uncertainties.

We apply our preferred method to 445 MERMAID seismograms to identify seismic phase arrivals and compute their travel-time residuals. Our multiscale AIC-based method is able to identify seismic phases with low uncertainty. The majority of the events in our MERMAID catalog corresponds to *P* waves, but we are also able to identify a few *S* waves as well as core phases.

We discussed the multiscale distributions of high-quality travel-time residuals in the MERMAID data set. In the supplemental material, we provide an updated events catalog that details those multiscale phase identifications, their travel-time residuals, and their uncertainty estimates. The software that automates these tasks is available for immediate deployment by other scientists. We hope it will be of use to the community.

DATA AND RESOURCES

We use Mobile Earthquake Recording in Marine Areas by Independent Divers (MERMAID) seismograms with identified events available from geoazur.unice.fr/ftp/mermaid/ (last accessed March 2019). We rely on irisFetch.m v.2.0.10, available from Incorporated Research Institutions for Seismology (IRIS), to query seismic catalogs available through the International Federation of Digital Seismograph Networks (FDSN). Manual event matching, where required (as described in The Initial Events Catalog section), was performed using the International Seismological Centre (2016) online bulletin. We use MatTaup, written in MATLAB by Qin Li while at the University of Washington in 2002, to compute theoretical travel times in the ak135 velocity model of Kennett *et al.* (1995). We maintain all of those codes, with minor modifications, at github.com/joelsimon/omnia (last accessed January 2020), which furthermore contains all of our software developed for this study. The supplemental material contains our catalog of MERMAID arrival times and their uncertainties, examples of our method applied to non-Gaussian synthetic and more seismic data, and details on our treatment of the edges of the seismogram during the forward and inverse wavelet transform.

ACKNOWLEDGMENTS

Part of this work was supported by the National Science Foundation (DGE-1656466 to J. D. S. and EAR-1550732, EAR-1736046, and OCE-1917058 to F. J. S.). Mobile Earthquake Recording in Marine Areas by Independent Divers (MERMAID) data were obtained with financing from the European Research Council (Advanced Grant 226837 to G. N.). Constructive comments by Associate Editor Eric Chael and an anonymous reviewer were greatly appreciated.

REFERENCES

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle, in *Selected Papers of Hirotugu Akaike*, E. Parzen, K. Tanabe, and G. Kitagawa (Editors), Springer, New York, New York, 199–213, doi: 10.1007/978-1-4612-1694-0_15.

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces, *Bull. Seismol. Soc. Am.* **68,** no. 5, 1521–1532.

Anant, K. S., and F. U. Dowla (1997). Wavelet transform methods for phase identification in three-component seismograms, *Bull. Seismol. Soc. Am.* **87,** no. 6, 1598–1612.

Baer, M., and U. Kradolfer (1987). An automatic phase picker for local and teleseismic events, *Bull. Seismol. Soc. Am.* **77,** no. 4, 1437–1445.

Baillard, C., W. C. Crawford, V. Ballu, C. Hibert, and A. Mangeney (2014). An automatic kurtosis-based *P*- and *S*-phase picker designed for local seismic networks, *Bull. Seismol. Soc. Am.* **104,** no. 1, 394–409, doi: 10.1785/0120120347.

Bohnenstiehl, D. R., M. Tolstoy, R. P. Dziak, C. G. Fox, and D. K. Smith (2002). Aftershock sequences in the mid-ocean ridge environment: An analysis using hydroacoustic data, *Tectonophysics* **354,** 49–70, doi: 10.1016/S0040-1951(02)00289-5.

Burnham, K. P., and D. R. Anderson (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Second Ed., Springer, New York, New York.

Chakraborty, A., and D. Okaya (1995). Frequency-time decomposition of seismic data using wavelet-based methods, *Geophysics* **60,** no. 6, 1906–1916, doi: 10.1190/1.1443922.

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

Cohen, A., I. Daubechies, and J. Feauveau (1992). Biorthogonal bases of compactly supported wavelets, *Comm. Pure Appl. Math.* **45,** 485–560, doi: 10.1002/cpa.3160450502.

Crotwell, H. P., T. J. Owens, and J. Ritsema (1999). The TauP toolkit: Flexible seismic travel-time and ray-path utilities, *Seismol. Res. Lett.* **70,** no. 2, 154–160, doi: 10.1785/gssrl.70.2.154.

Dahlen, F. A., S.-H. Hung, and G. Nolet (2000). Fréchet kernels for finite-frequency traveltimes—I. Theory, *Geophys. J. Int.* **141,** no. 1, 157–174, doi: 10.1046/j.1365-246X.2000.00070.x.

Dziak, R. P., D. R. Bohnenstiehl, H. Matsumoto, C. G. Fox, D. K. Smith, M. Tolstoy, T.-K. Lau, J. H. Haxel, and M. J. Fowler (2004). *P*- and *T*-wave detection thresholds, *Pn* velocity estimate, and detection of lower mantle and core *P*-waves on ocean sound-channel hydrophones at the Mid-Atlantic Ridge, *Bull. Seismol. Soc. Am.* **94,** no. 2, 665–677.

Gendron, P., J. Ebel, and D. Manolakis (2000). Rapid joint detection and classification with wavelet bases via Bayes theorem, *Bull. Seismol. Soc. Am.* **90,** no. 3, 764–774, doi: 10.1785/0119990103.

Hello, Y., A. Ogé, A. Sukhovich, and G. Nolet (2011). Modern mermaids: New floats image the deep Earth, *Eos Trans. AGU* **92,** no. 40, 337–338, doi: 10.1029/2011EO400001.

International Seismological Centre (2016). *On-line Bulletin*, Internatl. Seismol. Cent., Thatcham, United Kingdom, available at http://www.isc.ac.uk (last accessed March 2019).

Joubert, C., G. Nolet, S. Bonnieux, A. Deschamps, J.-X. Dessa, and Y. Hello (2015). P-delays from floating seismometers (MERMAID), part I: Data processing, *Seismol. Res. Lett.* **87,** no. 1, 73–80, doi: 10.1785/0220150111.

Joubert, C., G. Nolet, A. Sukhovich, A. Ogé, J.-F. Argentino, and Y. Hello (2015). Hydrophone calibration at very low frequencies, *Bull. Seismol. Soc. Am.* **105,** no. 3, 1797–1802, doi: 10.1785/0120140265.

Kennett, B. L. N., E. R. Engdahl, and R. Buland (1995). Constraints on seismic velocities in the Earth from travel-times, *Geophys. J. Int.* **122,** no. 1, 108–124.

Kullback, S., and R. A. Leibler (1951). On information and sufficiency, *Ann. Math. Stat.* **22,** no. 1, 79–86, doi: 10.1214/aoms/1177729694.

Leonard, M. (2000). Comparison of manual and automatic onset time picking, *Bull. Seismol. Soc. Am.* **90,** no. 6, 1384–1390, doi: 10.1785/0120000026.

Leonard, M., and B. L. N. Kennett (1999). Multi-component autoregressive techniques for the analysis of seismograms, *Phys. Earth Planet. In.* **113,** nos. 1/4, 247–263, doi: 10.1016/S0031-9201(99)00054-0.

Li, C., L. Huang, N. Duric, H. Zhang, and C. Rowe (2009). An improved automatic time-of-flight picker for medical ultrasound tomography, *Ultrasonics* **49,** no. 1, 61–72, doi: 10.1016/j.ultras.2008.05.005.

Luo, Y., and G. T. Schuster (1991). Wave-equation traveltime inversion, *Geophysics* **56,** no. 5, 654–663, doi: 10.1190/1.1443081.

Maeda, N. (1985). A method for reading and checking phase time in auto-processing system of seismic wave data, *Zisin* **38,** no. 3, 365–379, doi: 10.4294/zisin1948.38.3_365.

Mallat, S. (1998). *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, California.

McGuire, J. J., J. A. Collins, P. Gouédard, E. Roland, D. Lizarralde, M. S. Boettcher, M. D. Behn, and R. D. van der Hilst (2012). Variations in earthquake rupture properties along the Gofar transform fault, East Pacific Rise, *Nature Geosci.* **5,** no. 5, 336–341, doi: 10.1038/ngeo1454.

Morita, Y., and H. Hamaguchi (1984). Automatic detection of onset time of seismic waves and its confidence interval using the autoregressive model fitting, *Zisin* **37,** no. 2, 281–293, doi: 10.4294/zisin1948.37.2_281.

Nolet, G., Y. Hello, S. van der Lee, S. Bonnieux, M. C. Ruiz, N. A. Pazmino, A. Deschamps, M. M. Regnier, Y. Font, Y. J. Chen, *et al.* (2019). Imaging the Galápagos mantle plume with an unconventional application of floating seismometers, *Sci. Rep.* **9,** 1326, doi: 10.1038/s41598-018-36835-w.

Rastin, S. J., C. P. Unsworth, R. Benites, and K. R. Gledhill (2013). Using real and synthetic waveforms of the Matata Swarm to assess the performance of New Zealand GeoNet phase pickers, *Bull. Seismol. Soc. Am.* **103,** no. 4, 2173–2187, doi: 10.1785/0120120059.

Saragiotis, C. D., L. J. Hadjileontiadis, and S. M. Panas (2002). PAI-S/K: A robust automatic seismic P phase arrival identification scheme, *IEEE Trans. Geosci. Remote Sens.* **40,** no. 6, 1395–1404, doi: 10.1109/TGRS.2002.800438.

Simmons, N. A., S. C. Myers, G. Johannesson, and E. Matzel (2012). LLNL-G3Dv3: Global *P* wave tomography model for improved regional and teleseismic travel time prediction, *J. Geophys. Res.* **117,** no. B10302, doi: 10.1029/2012JB009525.

Simons, F. J., B. D. E. Dando, and R. M. Allen (2006). Automatic detection and rapid determination of earthquake magnitude by wavelet multiscale analysis of the primary arrival, *Earth Planet. Sci. Lett.* **250,** nos. 1/2, 214–223, doi: 10.1016/j.epsl.2006.07.039.

Simons, F. J., G. Nolet, J. M. Babcock, R. E. Davis, and J. A. Orcutt (2006). A future for drifting seismic networks, *Eos Trans. AGU* **87,** no. 31, 305–307, doi: 10.1029/2006EO310002.

Simons, F. J., G. Nolet, P. Georgief, J. M. Babcock, L. A. Regier, and R. E. Davis (2009). On the potential of recording earthquakes for global seismic tomography by low-cost autonomous instruments in the oceans, *J. Geophys. Res.* **114,** no. B05307, doi: 10.1029/2008JB006088.

Sleeman, R., and T. van Eck (1999). Robust automatic P-phase picking: An on-line implementation in the analysis of broadband seismogram recordings, *Phys. Earth Planet. In.* **113,** nos. 1/4, 265–275, doi: 10.1016/S0031-9201(99)00007-2.

Smith, D. K., M. Tolstoy, C. G. Fox, D. R. Bohnenstiehl, H. Matsumoto, and M. J. Fowler (2002). Hydroacoustic monitoring of seismicity at the slow spreading Mid-Atlantic Ridge, *Geophys. Res. Lett.* **29,** no. 11, 1518, doi: 10.1029/2001GL013912.

Strang, G., and T. Nguyen (1997). *Wavelets and Filter Banks*, Second Ed., Wellesley-Cambridge Press, Wellesley, Massachusetts.

Sukhovich, A., S. Bonnieux, Y. Hello, J.-O. Irisson, F. J. Simons, and G. Nolet (2015). Seismic monitoring in the oceans by autonomous floats, *Nat. Comm.* **6,** Article Number 8027, doi: 10.1038/ncomms9027.

Sukhovich, A., J.-O. Irisson, J. Perrot, and G. Nolet (2014). Automatic recognition of T and teleseismic P waves by statistical analysis of their spectra: An application to continuous records of moored hydrophones, *J. Geophys. Res.* **119,** no. 8, 6469–6485, doi: 10.1002/2013JB010936.

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

Sukhovich, A., J.-O. Irisson, F. J. Simons, A. Ogé, Y. M. Hello, A. Deschamps, and G. Nolet (2011). Automatic discrimination of underwater acoustic signals generated by teleseismic *P*-waves: A probabilistic approach, *Geophys. Res. Lett.* **38,** L18605, doi: 10.1029/2011GL048474.

Sweldens, W. (1996). The lifting scheme: A custom-design construction of biorthogonal wavelets, *Appl. Comput. Harmon. Anal.* **3,** no. 2, 186–200, doi: 10.1006/acha.1996.0015.

Tibuleac, I. M., and E. T. Herrin (1999). An automatic method for determination of Lg arrival times using wavelet transforms, *Seismol. Res. Lett.* **70,** no. 5, 577–595, doi: 10.1785/gssrl.70.5.577.

Tibuleac, I. M., E. T. Herrin, J. M. Britton, R. Shumway, and A. C. Rosca (2003). Automatic determination of secondary seismic phase arrival times using wavelet transforms, *Seismol. Res. Lett.* **74,** no. 6, 884–892, doi: 10.1785/gssrl.74.6.884.

Yuan, Y. O., and F. J. Simons (2014). Multiscale adjoint waveform-difference tomography using wavelets, *Geophysics* **79,** no. 3, WA79–WA95, doi: 10.1190/GEO2013-0383.1.

Zhang, H., C. Thurber, and C. Rowe (2003). Automatic *P*-wave arrival detection and picking with multiscale wavelet analysis for single-component recordings, *Bull. Seismol. Soc. Am.* **93,** no. 5, 1904–1912, doi: 10.1785/0120020241.

Zhang, J., Y. Tang, and H. Li (2017). STA/LTA fractal dimension algorithm of detecting the *P*-wave arrival, *Bull. Seismol. Soc. Am.* **108,** no. 1, 230–237, doi: 10.1785/0120170099.

Manuscript received 27 June 2019
Published online 17 March 2020

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/doi/10.1785/0120190173/4967090/bssa-2019173.1.pdf
by guustnolet

Supplemental Material:

Multiscale estimation of event arrival times and their uncertainties in hydroacoustic records from autonomous oceanic floats

Joel D. Simon, Frederik J. Simons and Guust Nolet

Description

This supplement contains details concerning the algorithmic implementation of our method, as well as more examples of it applied to synthetic and real data. Specifically it contains, in order: a mapping similar to that described in section *Timescale to time mapping: forward transform*, but described here for the inverse transform; the treatment of the edges by our multiscale method; our method applied to non-Gaussian synthetic time series; and our method applied to real data containing a supposed *PKiKP*-wave arrival. The latter is provided both to display an interesting seismogram as well as to illustrate the issue of phase ambiguity with our automated method.

Finally, as supplementary data to this study we also provide a separate plain text catalog of MERMAID arrival times and their uncertainties, the format of which is detailed in section *The updated events catalog* of the main text.

Timescale to time mapping: inverse transform

In preparing to describe our scheme to handle the edges of the time series we first define the inverse map which complements that described in section *Timescale to time mapping: forward transform*. The mapping between a timescale domain coefficient index l at scale j, and the time-domain sample-span to which l is sensitive under the inverse wavelet transform, is

$$\grave{F}_j(l) = \grave{k}^{\star}_{j,l}. \tag{S1}$$

As in the case of $\acute{F}_j$, the output of $\grave{F}_j$ is a time-domain sample span. Similarly to $\acute{k}^{\star}_{j,l}$, the support of a wavelet or scaling basis function of the inverse wavelet transform is completely described by $\grave{k}^{\star}_{j,l}$.

Similar to equations (54) and (56) we define the left and right boundaries by, respectively,

$$\grave{k}^{\perp}_{j,l} = \min(\grave{k}^{\star}_{j,l}), \tag{S2}$$

$$\grave{k}^{\top}_{j,l} = \max(\grave{k}^{\star}_{j,l}). \tag{S3}$$

Handling edges

In the main text we ignored the edges of any times series considered. Rather than building them into the construction of the transform itself, we opt to remove spurious wavelet values, those influenced by the edges, before changepoint estimation.

Edge sensitivity in the timescale domain

At every scale, the last, and first, timescale domain coefficient indices, which sense the left, and right, edges of x during forward wavelet transformation are

$$\acute{l}_{\mathrm{L}} = \max\{l : 1 \in \acute{k}^{\star}_{j,l}\}, \tag{S4}$$

$$\acute{l}_{\mathrm{R}} = \min\{l : N \in \acute{k}^{\star}_{j,l}\}, \tag{S5}$$

respectively. Upon mapping to the time domain the first and last sample indices which are not influenced by the edges after forward wavelet transformation are

$$\acute{k}_{\mathrm{L}} = \min(\acute{k}^{\perp}_{j,\acute{l}_{\mathrm{L}}+1}), \tag{S6}$$

$$\acute{k}_{\mathrm{R}} = \max(\acute{k}^{\top}_{j,\acute{l}_{\mathrm{R}}-1}), \tag{S7}$$

respectively. In the forward wavelet transformation every timescale domain coefficient is independent and unaware of the edge unless it receives contributions directly from the edge.

At every scale, before applying the multiscale analog of equation (26) directly on the timescale domain coefficient series $d_{j,l}$ and $a_{j,l}$, en route to finding the estimators l_{w_j} and $\overline{l_{\mathrm{w}_j}}$ of equations (51) and (52), we remove from consideration the edge-sensitive timescale domain coefficients.

Edge sensitivity in the time domain

Similarly, at every scale, the last and first timescale coefficient indices which sense the left and right edges, respectively, of x during inverse wavelet transformation are

$$\grave{l}_{\mathrm{L}} = \max\{l : 1 \in \grave{k}^{\star}_{j,l}\}, \tag{S8}$$

$$\grave{l}_{\mathrm{R}} = \min\{l : N \in \grave{k}^{\star}_{j,l}\}. \tag{S9}$$

During reconstruction via the inverse wavelet transformation, a temporally overlapping wavelets potentially propagate edge-effects deeper into the partially-reconstructed time series, x_j or $\overline{x}_j$, than in the case presented in section Edge sensitivity in the timescale domain. Therefore, during reconstruction we cannot simply augment l by ± 1 to locate the time-domain sample indices which are surely not influenced by the edge, as we did in equations (S6) and (S7). Indeed, this overlap implies that

$$\max(\acute{k}^{\perp}_{j,\acute{l}_{\mathrm{L}}}) \;\geq\; \min(\acute{k}^{\perp}_{j,\acute{l}_{\mathrm{L}}+1}), \tag{S10}$$

$$\max(\grave{k}^{\perp}_{j,\grave{l}_{\mathrm{L}}}) \;\geq\; \min(\grave{k}^{\perp}_{j,\grave{l}_{\mathrm{L}}+1}), \tag{S11}$$

$$\min(\acute{k}^{\top}_{j,\acute{l}_{\mathrm{R}}}) \;\leq\; \max(\acute{k}^{\top}_{j,\acute{l}_{\mathrm{R}}-1}), \tag{S12}$$

$$\min(\grave{k}^{\top}_{j,\grave{l}_{\mathrm{R}}}) \;\leq\; \max(\grave{k}^{\top}_{j,\grave{l}_{\mathrm{R}}-1}). \tag{S13}$$

The left-hand sides in equations (S10)–(S13) are time-domain sample indices that are assured to be free of edge influence: the first and last being given respectively by equations (S10) and (S12) after forward wavelet transformation, and equations (S11) and (S13) after inverse wavelet transformation. To be conservative we keep the larger and smaller of these as the first and last time domain sample indices, respectively, which are not influenced by the edges of x, defining

$$\grave{k}_{\mathrm{L}} = \max\{\acute{k}^{\perp}_{j,\acute{l}_{\mathrm{L}}}\ ,\ \grave{k}^{\perp}_{j,\grave{l}_{\mathrm{L}}}\} + 1, \tag{S14}$$

$$\grave{k}_{\mathrm{R}} = \min\{\acute{k}^{\top}_{j,\acute{l}_{\mathrm{R}}}\ ,\ \grave{k}^{\top}_{j,\grave{l}_{\mathrm{R}}}\} - 1. \tag{S15}$$

At every scale, before applying the multiscale analog of equation (26) on the time-domain subspace projection series x_j and $\overline{x}_j$, en route to finding the estimators k_{w_j} and $\overline{k_{\mathrm{w}_j}}$ of equations (49) and (50), we remove the edge-sensitive time-domain sample indices from consideration.

Non-Gaussian models

We have intimated that our AIC-based event detection method will enjoy broad application beyond the scope of this study. Indeed, our multiscale changepoint estimation procedure is completely agnostic of seismology. The general form of equation (26) in terms of natural logarithms multiplied by the sample variances of the segments implies that it can be readily applied to any time series that can be modeled as concatenated samples from distributions in the larger exponential family. We forgo showing such examples here.

Future applications that may opt for entirely different synthetic model formulations may simply rederive the appropriate AIC function and deploy it as part of our workflow. On the other hand, our AIC formulation quite simply compares ratios of variances and thus it is likely to remain useful even when applied to time series whose generating distributions are not in the exponential form. Figures S1 and S2 recreate Figure 5, assuming that x is drawn from two concatenated Student *t*-distributions, or from two *F*-distributions, respectively, with different variances. Our changepoint detection method remains on target, although the longer tails of these generating distributions will lead to broader error distributions than is the case of the assumed Gaussian models of this study.

A final data example

Presented in the same layout as Figures 14–17, Figure S3 discriminates the (reported) inner-core reflection *PKiKP* wave from the reflected mantle phase *PP*. Note that the former is more likely a *PKIKP* wave that bottoms just inside the inner core, as this phase also appears around 114° in the ak135 velocity model for an earthquake at 607 km depth.

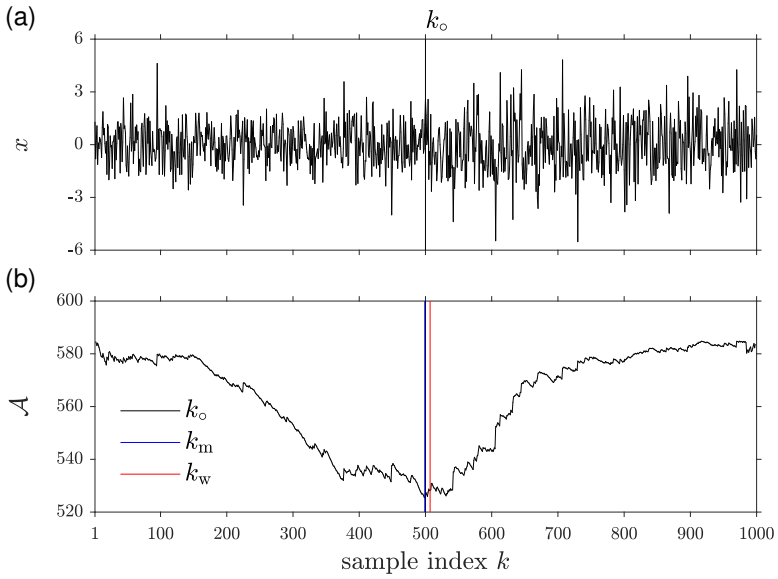


Figure S1. A recreation of Figure 5 using a non-Gaussian generating distribution. Here x is a sample of length $N = 1000$ drawn from a Student *t*-distribution with 10 degrees of freedom. Sample indices $k = [501, \ldots, 1000]$ are multiplied by $\sqrt{2}$, therefore generating a time series with SNR ≈ 2 and a true changepoint at sample index 500. In this example $k_\mathrm{m} = 499$ and $k_\mathrm{w} = 507$.

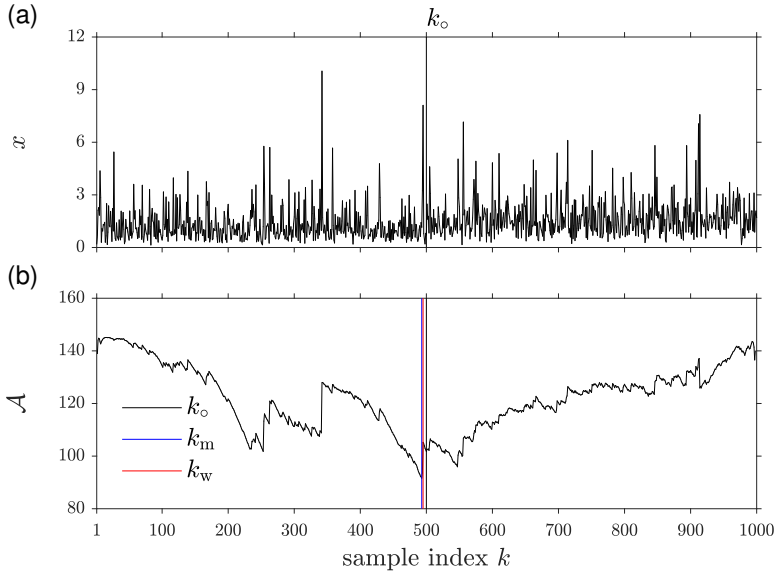


Figure S2. A recreation of Figure S1, this time assuming that the time series under consideration x is drawn from the *F*-distribution with parameters $(10, 10)$. Like Figure S1, sample indices $k = [501, \ldots, 1000]$ are multiplied by $\sqrt{2}$ to yield SNR ≈ 2 and $k_\circ = 500$. In this example the estimated changepoints are $k_\mathrm{m} = 493$ and $k_\mathrm{w} = 495$, illustrating again that our procedure remains broadly valid, while uncertainties and confidence intervals will require suitable adaptation.

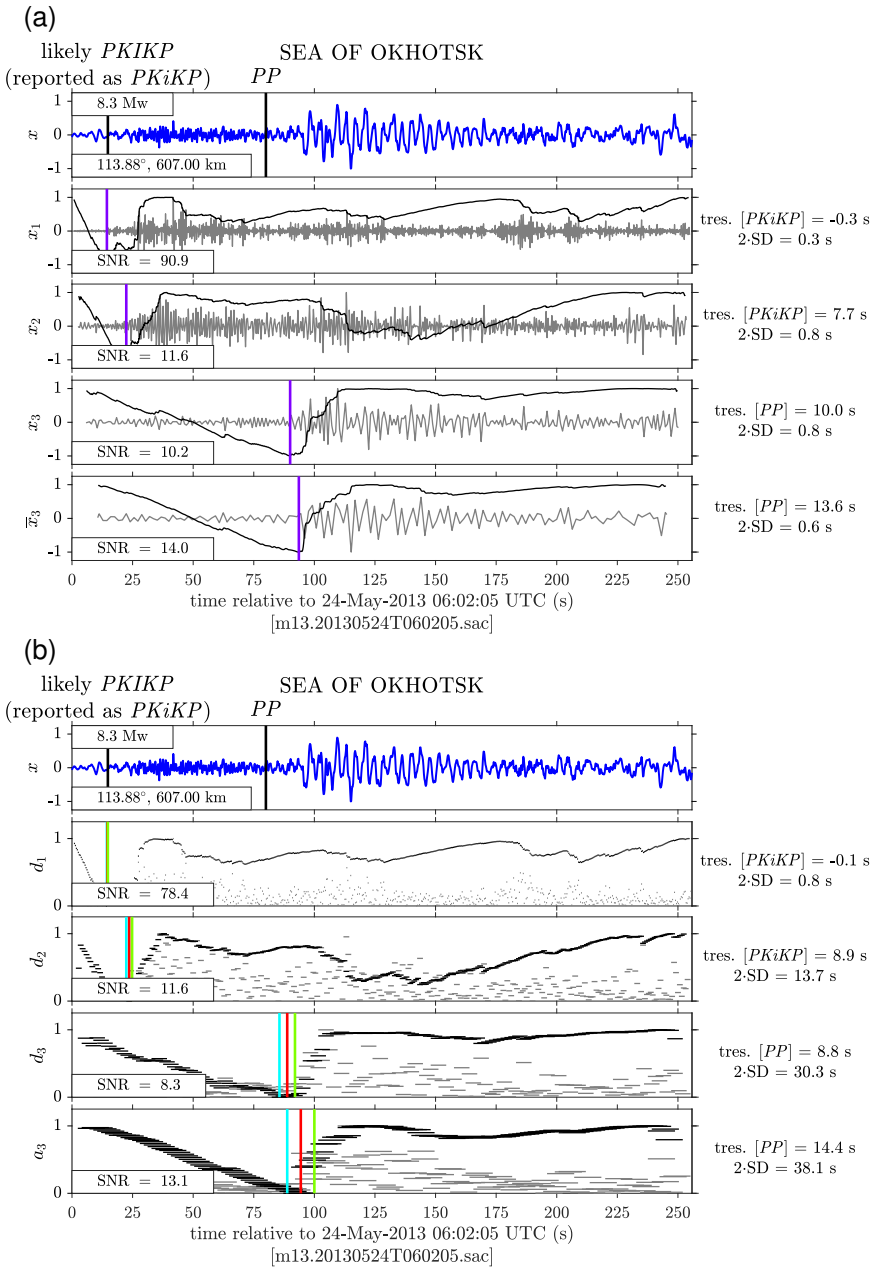


Figure S3. A mid-SNR seismogram detected in the southern Indian Ocean corresponding to an earthquake in the Sea of Okhotsk. Here our procedure picks a reported *PKiKP* wave (but more likely a *PKIKP* wave; see section *Computing travel-time residuals*) at scales 1 and 2 and a *PP* wave at scale 3, in both domains. Note that the sampling rate of this seismogram is only 5 Hz, and therefore was decomposed only to three scales (see Table 1).