Smoothing can systematically bias small samples of one-dimensional biomechanical continua

Todd C. Pataky[a], Mark A. Robinson[b], Jos Vanrenterghem[c], John H. Challis[d]

[a]*Department of Human Health Sciences, Kyoto University, Japan*
[b]*Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, UK*
[c]*Department of Rehabilitation Sciences, KU Leuven, Belgium*
[d]*Biomechanics Laboratory, Penn State University, USA*

Abstract

The quality with which smoothing algorithms perform is often assessed in simulation by starting with a known 1D datum, adding noise, smoothing the noisy data, then quantifying the difference between the smoothed data and known datum, often using mean-square error (MSE). While effectively summarizing overall difference, MSE fails to capture localized, one-sided errors. This paper describes how smoothing noisy 1D data using a variety of algorithms can introduce systematic bias, and quantifies this bias using the false positive rate (FPR): the probability that a smoothing algorithm will yield a dataset whose 1D mean differs significantly from its true 1D datum. A simulation study was conducted involving six 1D datum continua, and four smoothing algorithms whose parameters were systematically manipulated along with sample size and noise amplitude. Approximately ten million simulation iterations were evaluated. FPRs were calculated at α=0.05, based on the calculated smoothness of the resulting datasets. Results showed that FPRs were much higher than the expected value of α, and in many cases approached 100%. FPRs were highest with aggressive smoothing parameters, large sample sizes and small noise amplitudes, irrespective of both smoothing algorithm and the 1D datum. These results suggest that smoothing 1D biomechanical data can introduce statistical bias with relatively high probability. The implications are experiment-specific because the biomechanical meaning of 1D changes can vary vastly between datasets. Smoothing-induced bias should be a cause for general concern when small 1D changes have non-trivial biomechanical consequences.

1. Introduction

Digitally smoothing noisy one-dimensional (1D) time series measurements with the goal of increasing the signal-to-noise ratio (SNR) has been studied in Biomechanics for over four decades (e.g. Winter et al., 1974; Wood, 1982; Schreven et al., 2015). Many classes and subclasses of relevant algorithms exist (Gazzani, 1994), and even common algorithms like the lowpass Butterworth filter (Butterworth, 1930) have seen a variety of algorithmic iterations (Challis, 1999; Yu et al., 1999; Erer, 2007). Collectively we refer to algorithms that process data to reduce high-frequency content and increase the SNR as 'smoothing algorithms', and these include for example: lowpass filters, cross-validated splines, and singular spectrum analysis.

November 28, 2018

The quality with which smoothing algorithms perform is often assessed in simulation by starting with a known 1D datum, adding noise, smoothing the noisy data, then quantifying the difference between the smoothed data and known datum. Quality can also be assessed by comparing smoothed data to other criterion measures (e.g. Pezzack et al., 1977; Vaughan, 1982). A variety of metrics like SNR (Yu et al., 1999), and dissimilarity metrics like absolute error (Georgakis et al., 2002) and mean squared error (MSE) (Gazzani, 1994; Challis, 1999) have been used for this purpose because all aforementioned metrics decrease (or increase) systematically as the smoothed data converge to the known datum.

While prevalent, all (dis)similarity metrics share a common weakness: they fail to capture systematic bias across a sample of measurements. 'Bias' in this case refers to statistically significant deviation from the known one-dimensional datum (Pataky, 2016). For example, consider a smoothing algorithm which always creates greater (or smaller) local maxima than exist in a true 1D datum. A dissimilarity metric like the MSE can be blind to this type of bias because (a) it averages error across the entire continuum, and can therefore miss high concentrations of error, and (b) it squares errors so exclusively positive errors (e.g. a higher local maximum) are indistinguishable from evenly distributed positive and negative errors.

Bias can nevertheless be easily captured in 1D test statistics (e.g. t and F continua) because they effectively quantify one-sided error at each continuum point, and are designed precisely to detect the local signal in noisy continuum measurements (Friston et al., 2007; Pataky, 2012). They can detect the local signal through random field theory (Adler and Taylor, 2007), which yields probability values associated with both global and local features of 1D signals. These probability values can be regarded as alternatives to the aforementioned (dis)similarity metrics: for a particular dataset, increasingly small probability values would imply that smoothing has increasingly distorted the true continuum.

When considering many datasets with many different realizations of simulated noise, it is more informative to aggregate these probabilities into a false positive rate (FPR). In the context of this study, the FPR is a single scalar which represents the likelihood that a particular smoothing procedure, when applied to an arbitrary number of noisy 1D measurements, will yield a smoothed dataset whose 1D mean is significantly different from the true 1D datum. If smoothing-induced differences are purely random (like the underlying noise), then the FPR is α by definition (conventionally $\alpha = 5\%$). The basic expectation that purely random 1D differences yield FPR=α has been extensively validated elsewhere (Pataky, 2016). In particular, both high- and low-frequency Gaussian noise tend to produce random deviations from a 1D datum, and the probability with which arbitrarily smooth Gaussian noise reaches certain deviation amplitudes can be accurately predicted using random field theory (Adler and Hasofer, 1976; Pataky, 2016). Thus, if one adds high frequency Gaussian noise to a known 1D datum, that noise by definition will produce an FPR of α. However, if one smooths the added high-frequency Gaussian noise, the resulting deviations from the known 1D datum may no longer be Gaussian (i.e. with a mean of zero), and thus may produce higher FPRs than α.

The purpose of this study was to quantify bias introduced by typical biomechanical smoothing techniques using false positive rates, and to subsequently elucidate general smoothing algorithm and dataset features associated with smoothing bias. To this end we conducted over ten million numerical simulation iterations involving a variety of datasets and smoothing algorithms. We systematically manipulated algorithm parameters like cutoff frequency, and also systematically manipulated two key dataset features: sample size and noise amplitude.

2. Methods

Analyses were conducted in Python 3.6 using Anaconda 3.5 (Anaconda, 2017) and **spm1d** (Pataky, 2012). Datasets, smoothing source code, scripts replicating key results, and all of our simulation results are available at: `https://github.com/0todd0000/smooth1d`

2.1. Datasets

Six 1D datum continua were used (Fig.1). Dataset 1 consisted of the experimentally measured vertical positions of a falling golf ball (Vaughan, 1982, p.379). Datasets 2–6 are mathematical functions, representing a range of frequency content and were borrowed from Challis (1999). Frequencies less than 2 Hz contained at least 80% of spectral power for all datasets (Supplementary Material, Fig.A.1). Following Challis (1999), Gaussian noise was added to each 1D datum (Fig.1), and was normalized across datasets using the percentage root-mean square error (RMSE):

$$\text{RMSE (\%)} = 100\% \times \frac{\sqrt{\sum (\hat{y}_q - y_q)^2}}{\sqrt{\sum y_q^2}} \tag{1}$$

Here $\hat{y}$ and y represent the noisy and datum continuum, respectively, and q indexes continuum nodes. Two different noise levels were used: 1% and 20% RMSE. Corresponding standard deviation (SD) values for the noise are indicated in Fig.1.

2.2. Smoothing algorithms

Five smoothing methods were used (Table 1). The first ('None') employed no smoothing and instead routed noisy data directly to statistical testing. The second was a lowpass Butterworth filter with parameters: cutoff and order. The third ('Autocorr') optimized the Butterworth lowpass cutoff frequency by minimizing the sum of the absolute autocorrelation sequence of the filtered data. The fourth ('GCVSPL') was the generalized cross-validatory spline method described by Craven and Wahba (1979); implemented using the C translation of Twisk (1994) of the original Fortran code (Woltring, 1986) for use in Python. The last method ('SSA') was the singular spectrum analysis method (Golyandina et al., 2001; Alonso et al., 2005).

Table 1: smoothing algorithms. The listed parameter values were iterated in simulation. See also Supplementary Material, Fig.A.2.

Algorithm	**Reference(s)**	**Parameters**
None	—	—
Butterworth[1]	Butterworth (1930)	Cutoff: {2, 3, . . . , 10} Order: {2, 3, 4, 5}
Autocorr	Challis (1999)	Order: {2, 3, 4, 5}
GCVSPL	Craven and Wahba (1979)	Spline degree: {3, 5, 7, 9}
SSA	Golyandina et al. (2001)	Window: {5, 10, 15, 20, 25} Components: {2, 3, 4, 5}

[1]Applied in both forward and reverse directions to eliminate phase lag. Post-second pass cutoff values are listed.

2.3. Simulations

Numerical simulations were conducted in which unique combinations of dataset, noise level, sample size and smoothing algorithm were set, then repeated 1000 times each as summarized in Fig.A.2 (Supplementary Material). For each simulation iteration, novel Gaussian noise was realized using a random number generator and was added to the datum as depicted in Fig.1. Sample sizes ranging from J=5 to J=50 were employed to represent the small-to-moderate sample sizes typical of Biomechanics research (Knudson, 2017). In total, 10.14 million simulation iterations were conducted.

2.4. False positive rate calculations

Each simulation iteration yielded a noisy 1D sample which was compared to its known 1D datum using a continuum-level one-sample t test (Pataky, 2012) as depicted in Fig.2. Briefly, the average temporal smoothness of the 1D residuals (Fig.2e-g) was estimated according to Kiebel et al. (1999), and this smoothness parameter, together with sample size, was used to calculate a two-tailed critical threshold at α=0.05 (Fig.2h-j) using Gaussian random field theory (Adler and Taylor, 2007; Friston et al., 2007). This critical threshold represents the absolute t value above which only 5% of t statistic continua (generated by identically smooth Gaussian continua) would traverse in an infinite number of experiments. Traversing the threshold implies a false positive, or equivalently a data sample whose 1D mean continuum differs significantly from its known 1D datum. This approach to false positive calculation has been extensively validated for 1D continua (Pataky, 2016). We note that an unbiased algorithm would produce a false positive rate of α.

3. Results

Unsmoothed data exhibited false positive rates (FPRs) close to α=0.05, irrespective of sample size and noise level (Fig.3). Increasing the number of simulation iterations would yield closer convergence (Pataky, 2016). Since these results converge to the expected value of α=0.05, they validate the employed FPR calculation approach.

In contrast, all smoothing procedures were found to yield FPRs greater than α in many cases (Figs.4 to 7). The observed FPRs approached 100% in many cases, but the actual FPR value depended largely on all investigated factors: dataset, noise level, sample size, and smoothing parameters. The most consistently observed trend was that FPRs increased with sample size; this was generally consistent across all datasets and smoothing algorithms. A second consistent trend was that the smaller noise (1%) was generally associated with higher FPRs than moderate noise (20%) (e.g. Fig.5a; Fig.6e).

A third general observation was that settings that generated greater smoothing induced higher false positive rates. For example, both smaller Butterworth filter cutoff frequencies (Fig.4d), and broader SSA windows (Fig.7a), tended to yield higher FPRs.

A final general observation was that no dataset was immune from false positives, and some which yielded lower FPRs than other datasets in one case (e.g. Fig.6a, noise=20%) tended to yield higher FPRs than other datasets in another closely related case (e.g. Fig.6a, noise=1%).

4. Discussion

4.1. Main implications

This study quantified smoothing bias using false positive rates (FPRs) and found that many smoothing algorithms commonly employed in the Biomechanics literature can yield systematically biased 1D data in many situations. The clearest problem identified was that over-smoothing can yield high FPRs, thus confirming and extending the long-known understanding of systematic distortion of data by over-smoothing (Pezzack et al., 1977; Bisseling and Hof, 2006).

This finding is not new, and selecting appropriate smoothing parameters has indeed been the focus of many Biomechanics papers (e.g. Winter et al., 1974; Challis, 1999; Yu et al., 1999; Erer, 2007; Schreven et al., 2015). The fact that FPR successfully identified the expected consequences of overly-aggressive smoothing suggests that it is likely valid for detecting bias more generally.

Similar to other commonly employed metrics like RMSE (Eqn.1), the FPR is a single scalar which can be used to quantify smoothing performance. Unlike the RMSE and other metrics, the FPR operates on multiple and not on single 1D measurements, making it somewhat more relevant to experimental Biomechanics which simultaneously analyses multiple 1D measurements. Also unlike other metrics, the FPR is grounded in probability theory (Adler and Taylor, 2007; Friston et al., 2007), so its experiment-relevant meaning is clear. An FPR of 0.4, for example, suggests that a particular smoothing algorithm is expected to produce a 1D dataset whose mean significantly differs from the true datum in 40% of an infinite number of identical experiments. The FPR thus offers complimentary information to the RMSE, which quantifies only absolute deviation from the true datum. Since most experimental Biomechanics studies ultimately report probability results (Knudson, 2017), reporting the probability-relevant FPR alongside RMSE is recommended for future smoothing studies.

Using the FPR, this study identified other factors that can exacerbate smoothing bias. In particular, larger sample sizes and smaller noise amplitudes were both associated with greater bias, albeit in both a dataset- and algorithm-dependent manner. Both of these trends were somewhat unexpected. The sample size trend was unexpected because increasing sample size is expected to yield tighter convergence to the true 1D population mean (Pataky, 2016). The fact that FPRs increased with sample size therefore provides strong evidence that smoothing-induced bias is real. The noise amplitude trend was also unexpected because less noise is intuitively associated with a truer representation of the mean. Nevertheless, with white noise, which occupies all frequencies, one can expect that increasingly amplified noise would increasingly drown out the true mean signal, and therefore increasingly behave like true noise and null signal, for which the FPR is α by definition. A third factor identified was that FPRs can vary substantially with dataset, even when the datasets have similar frequency content and a constant smoothing procedure is used (e.g. Fig.4b vs. Fig.4e).

We surmise that false positives, which represent smoothing-induced systematic deviation from a known 1D datum, are most generally caused by the competing interests of frequency-based signal processing and amplitude-based statistical analysis. smoothing algorithms are designed primarily to control the frequency content of 1D measurements, and in particular to attenuate frequencies believed to be primarily noise/error. Statistical analysis in Biomechanics is nevertheless amplitude-based; systematic changes in the amplitudes and frequency content of physical variables including forces and positions can have important biomechanical consequences (Bisseling and Hof, 2006). smoothing procedures stem from communications theory, where the primary goal is to maintain the true signal's frequency component ratios, and where amplitude changes can be compensated for with volume changes. In Biomechanics, frequency's amplitudes themselves are important. This study's results suggest that the interaction between frequency-based processing and amplitude-based biomechanical inference warrants further investigation.

4.2. Limitations

The primary limitation of this study is that the FPR has well-defined statistical meaning but effectively undefined biomechanical meaning. In particular, the amplitudes of the deviations FPR identifies might be small and biomechanically irrelevant. For example, one may regard the deviations observed in Fig.2b to be mechanically small/insignificant, despite being associated with relatively large test statistic values (Fig.2h). If these data represented joint reaction forces during clinical gait analysis, for example, changing the force continuum in the manner shown may induce very little practical consequence on walking kinematics and thus on biomechanical interpretations. On the other hand, if these data represented joint rotations during elite sports performance, the apparently small reduction in range of motion may have large functional consequences. The proposed FPR metric therefore has no consistent biomechanical meaning. Biomechanical meaning, and the interpretive consequences of smoothing-induced bias must instead be ascertained on an experiment-by-experiment basis, in the context of the measured data, the measurement's noise characteristics, and most importantly the hypothesis which drives the experiment. Only through a biomechanically meaningful hypothesis can the disparity between statistical and biomechanical meaning be resolved. In a separate paper we discuss power-based computational tools that can facilitate the process of unifying statistical and biomechanical meaning (Pataky et al., 2018)

Another key limitation of this study is that it was not designed to elucidate the mathematical or computational mechanisms through which smoothing induces bias. A qualitative appreciation for those mechanisms may nevertheless be garnered through a consideration of general smoothing effects. smoothing tends to simultaneously induce two statistically important changes: (1) increased test statistic values, and (2) decreased critical threshold. It achieves the former by reducing noise and consequently increasing the signal-to-noise ratio (SNR); the SNR is conceptually equivalent to the t statistic (ratio of the effect to measurement variance). smoothing decreases the critical threshold by yielding increasingly smooth data, which effectively embody fewer independent processes (Friston et al., 2007). In other words, a high threshold is required to control α across many independent processes, but the threshold lowers as those processes become increasingly correlated (Adler and Taylor, 2007). Further exploration of the interaction between dataset characteristics and smoothing procedures is warranted.

A third limitation is that we only considered high-frequency noise (Fig.1). In some cases real biomechanical noise may be lower-frequency. For example, given identical calcaneus kinematics, low-frequency noise could be expected in 1D ground reaction force data due to the viscoelastic properties, and corresponding hysteresis, of the calcaneal soft tissue. We leave the relative importance of high- vs. low-frequency noise for future work.

A final limitation is that this study's results do not directly relate to strategies one may use to check for and/or avoid smoothing-induced bias. At this point we can only suggest conducting sensitivity analyses. Repeating analyses with a range of smoothing parameters from highly conservative to highly aggressive, and potentially also with a variety of different smoothing algorithms, would provide data which represent the consequential span of smoothing particulars on one's ultimate results. If the ultimate results are sensitive to smoothing parameter changes this should be reported.

4.3. Summary

This study introduced the use of the false positive rate (FPR) to quantify smoothing-induced bias for 1D biomechanical continua. A large-scale simulation study was conducted using a variety of datasets, sample sizes, noise characteristics and smoothing algorithms, and FPRs were recorded. Results show that FPRs are controlled at a rate of α for unsmoothed data, but that smoothing can induce much higher FPRs, indicating significant phantom signal, especially for large sample sizes, small noise amplitudes and aggressive smoothing approaches. The biomechanical meaning of this bias must be resolved on a case-by-case basis in the context of one's hypotheses. smoothing-induced bias can be ascertained and avoided through simulation studies like this one or through sensitivity analysis.

Acknowledgments

This work was supported in part by Kiban B Grant 17H02151 from the Japan Society for the Promotion of Science.

Conflict of Interest Statement

The authors report no conflict of interest, financial or otherwise.

References

Adler, R. J., and Hasofer, A. M., 1976. Level crossings for random fields. *The Annals of Probability 4*, 1–12.

Adler, R. J., and Taylor, J. E., 2007. *Random Fields and Geometry*. Springer-Verlag.

Alonso, F. J., Castillo, J. M. D., and Pintado, P., 2005. Application of singular spectrum analysis to the smoothing of raw kinematic signals. *Journal of Biomechanics 38*, 1085–1092.

Anaconda, 2017. Anaconda Software Distribution version 3-5.1. URL: `https://anaconda.com`.

Bisseling, R. W., and Hof, A. L., 2006. Handling of impact forces in inverse dynamics. *Journal of Biomechanics 39*, 2438–2444.

Butterworth, S., 1930. On the theory of filter amplifiers. *Wireless Engineer 7*, 536–541.

Challis, J. H., 1999. A procedure for the automatic determination of filter cutoff frequency for the processing of biomechanical data. *Journal of Applied Biomechanics 15*, 303–317.

Craven, P., and Wahba, G., 1979. Smoothing noisy data with splines functions. *Numerische Mathematik 31*, 377–403.

Erer, K. S., 2007. Adaptive usage of the Butterworth digital filter. *Journal of Biomechanics 40*, 2934–2943.

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., and Penny, W. D., 2007. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. London: Elsevier.

Gazzani, F., 1994. Comparative assessment of some algorithms for differentiating noisy biomechanical data. *International Journal of Bio-Medical computing 37*, 57–76.

Georgakis, A., Stergioulas, L. K., and Giakas, G., 2002. Automatic algorithm for filtering kinematic signals with impacts in the Wigner representation. *Medical and Biological Engineering and Computing 40*, 625–633.

Golyandina, N., Nekrutkin, V., and Zhigljavsky, A., 2001. *Analisys of Time Series Structure - SSA and Related Techniques*. Chapman & Hall/CR.

Kiebel, S., Poline, J., Friston, K., Holmes, A., and Worsley, K., 1999. Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. *NeuroImage 10*, 756–766.

Knudson, D., 2017. Confidence crisis of results in biomechanics research. *Sports Biomechanics* (pp. 1–9).

Pataky, T. C., 2012. One-dimensional statistical parametric mapping in Python. *Computer Methods in Biomechanics and Biomedical Engineering 15*, 295–301.

Pataky, T. C., 2016. rft1d: Smooth one-dimensional random field upcrossing probabilities in Python. *Journal of Statistical Software 71*, 1–22.

Pataky, T. C., Robinson, M. A., and Vanrenterghem, J., 2018. A computational framework for estimating statistical power and planning hypothesis-driven experiments involving one-dimensional biomechanical continua. *Journal of Biomechanics in press*.

Pezzack, J. C., Norman, R. W., and Winter, D. A., 1977. An assessment of derivative determining techniques used for motion analysis. *Journal of Biomechanics 10*, 377–382.

Schreven, S., Beek, P. J., and Smeets, J. B. J., 2015. Optimising filtering parameters for a 3D motion analysis system. *Journal of Electromyography and Kinesiology 25*, 808–814.

Twisk, D., 1994. GCVSPL in C. URL: `https://isbweb.org/resources/software-resources/137-signal-processing-software/497-gcvspl-in-c-d-twisk`.

Vaughan, C. L., 1982. Smoothing and differentiation of displacement-time data: an application of splines and digital filtering. *International Journal of Bio-Medical Computing 13*, 375–386.

Winter, D. A., Sidwall, H. G., and Hobson, D. A., 1974. Measurement and reduction of noise in kinematics of locomotion. *Journal of Biomechanics 7*, 157–159.

Woltring, H. J., 1986. A Fortran package for generalized, cross-validatory spline smoothing and differentiation. *Advances in Engineering Software 8*, 104–113.

Wood, G. A., 1982. Data smoothing and differentiation procedures in biomechanics. *Exercise and Sport Sciences 10*, 308–362.

Yu, B., Gabriel, D., Noble, L., Applied, K. A. J. o., and 1999, 1999. Estimate of the optimum cutoff frequency for the Butterworth low-pass digital filter. *journals.humankinetics.com 15*, 318–329.

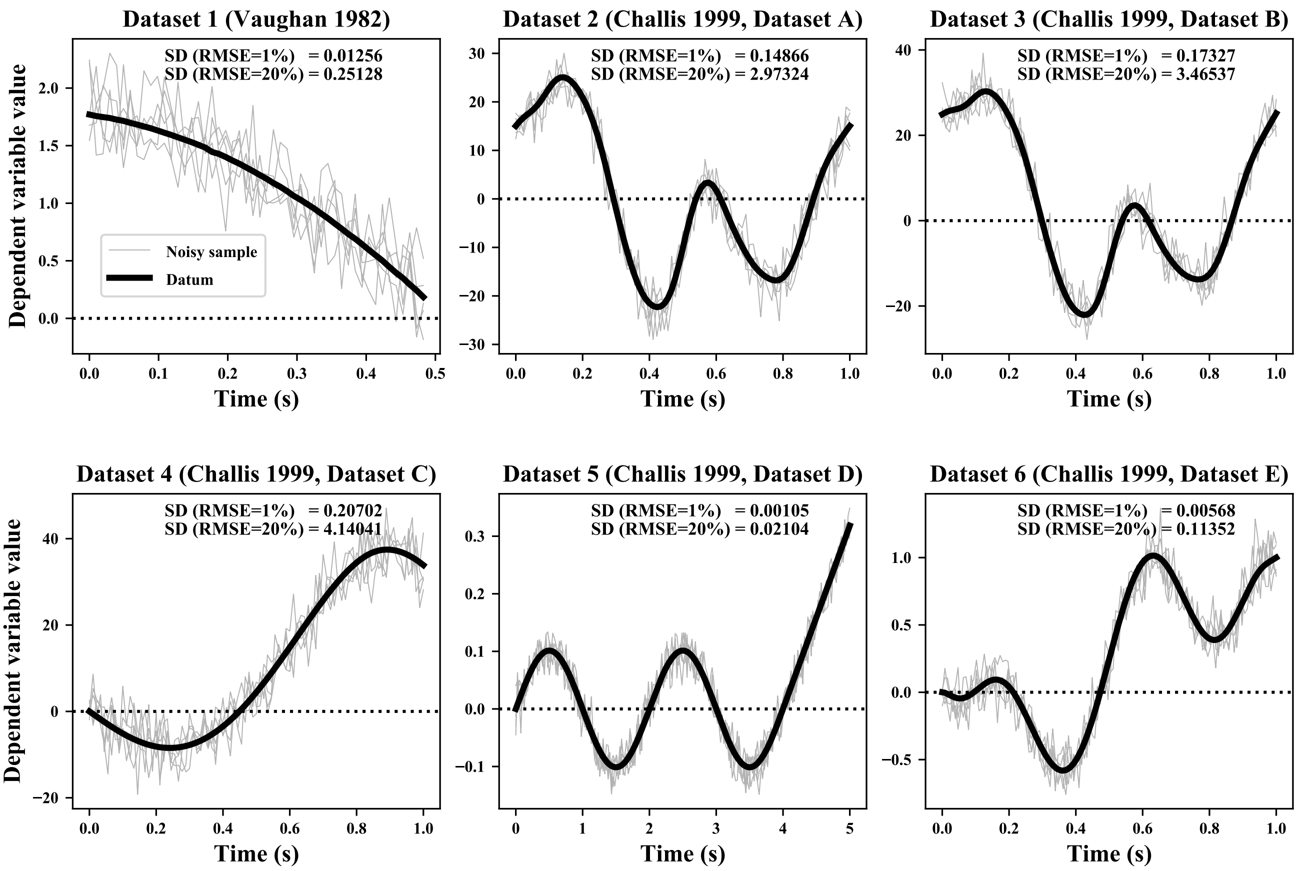

Figure 1. Datasets. Thick and thin lines represent the 1D datum and added Gaussian noise, respectively. Noise amplitude is characterized using standard deviation (SD); SD values necessary to produce 1% and 20% root mean square error (RMSE) values are listed. Each panel depicts 20% RMSE noise and a sample size of J=5 (i.e. five 1D continua per datum).

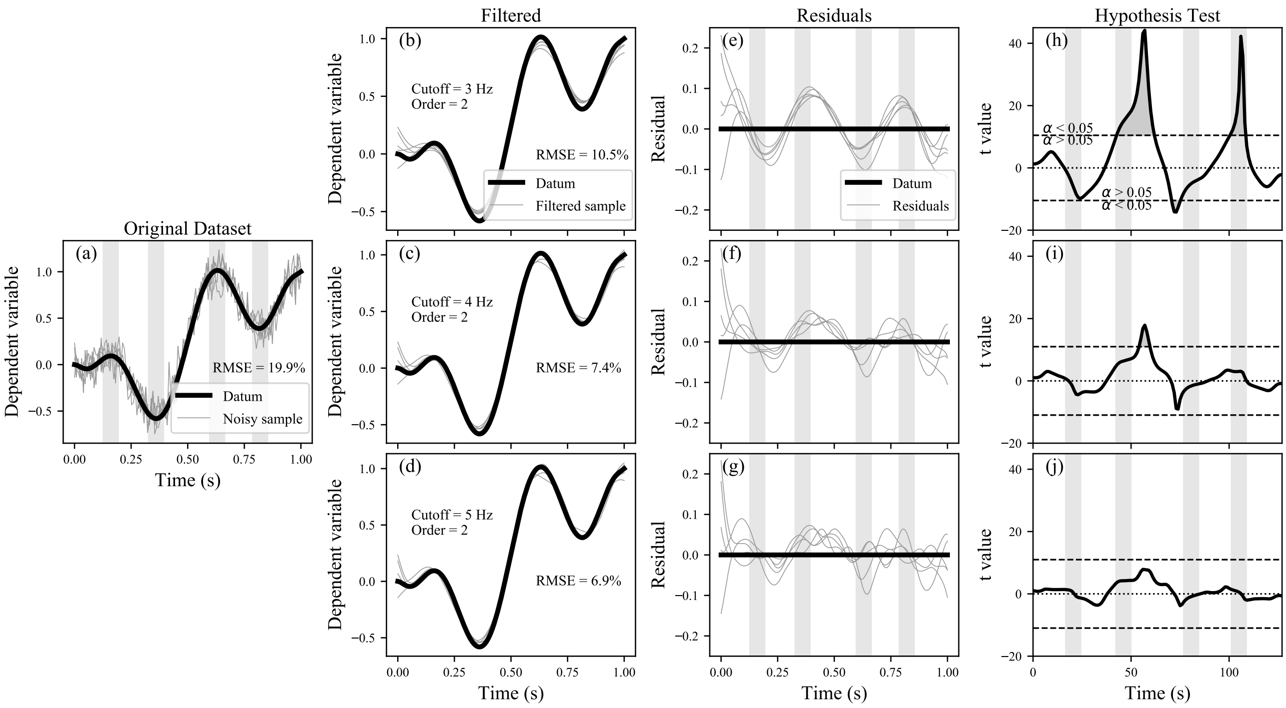

Figure 2. Example smoothing results. (a) Dataset 6 (see Fig.1) with added noise, theoretically at 20% RMSE (percentage root mean square error); the actual PRMSE for this noise realization was 19.9%. Light vertical bars highlight positions of local extrema. (b-d) Lowpass filtering results with cutoffs of 3, 4 and 5 Hz, respectively. (e-f) Residuals with respect to the known datum. (h-j) One-sample t tests comparing the smoothed data to the known datum. Horizontal dotted lines represent the critical two-tailed threshold at alpha=0.05.

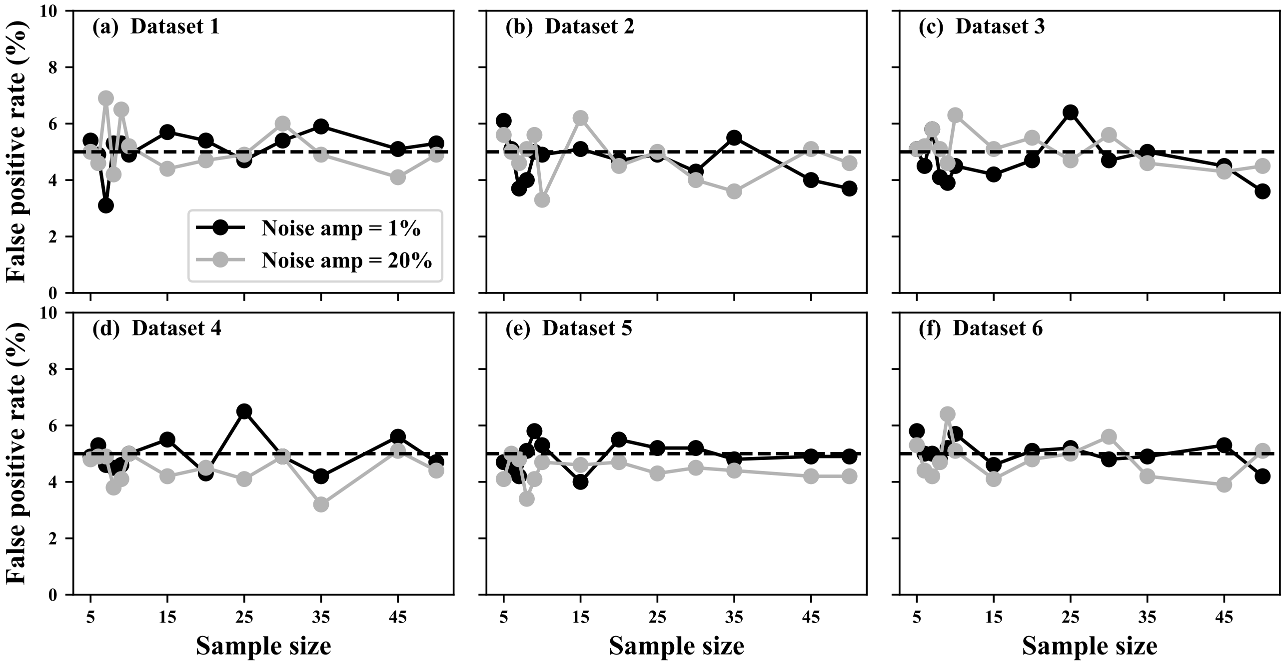

Figure 3. False positive results, unsmoothed data. Horizontal dashed lines indicate alpha=5%. Numerical false positive rates are expected to converge to alpha=5% as the number of simulation iterations increases. Note that the vertical scale is set at 0-10%. The remaining figures (Figs.4-7) use a scale of 0-100%.

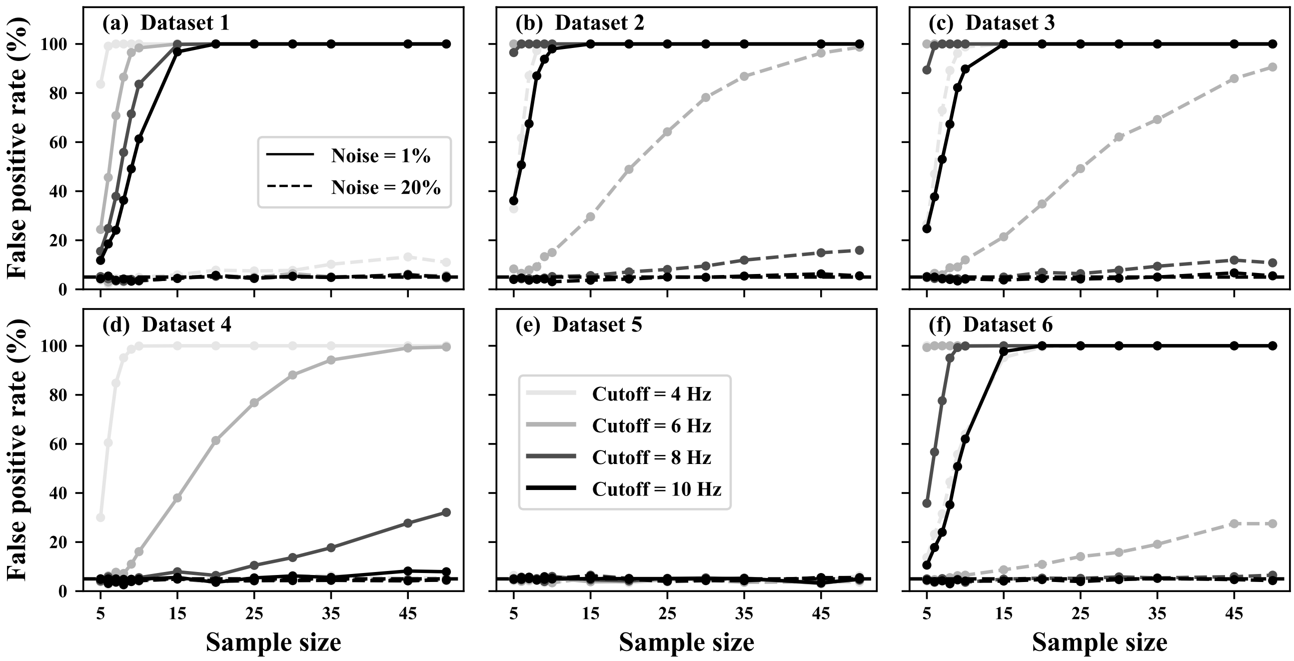

Figure 4. False positive results, Butterworth filtering method. Results are presented only for filtering order = 2; filtering order did not qualitatively affect results (see Fig.A3, Supplementary Material).

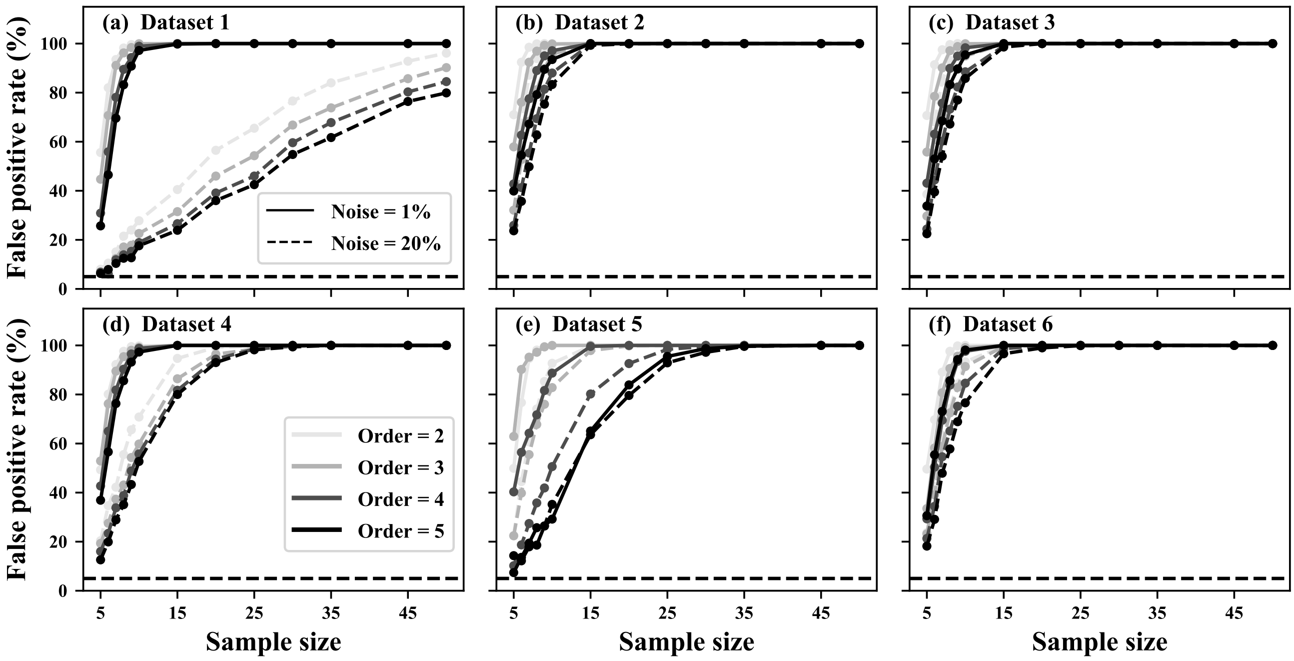

Figure 5. False positive results, Autocorr filtering method.

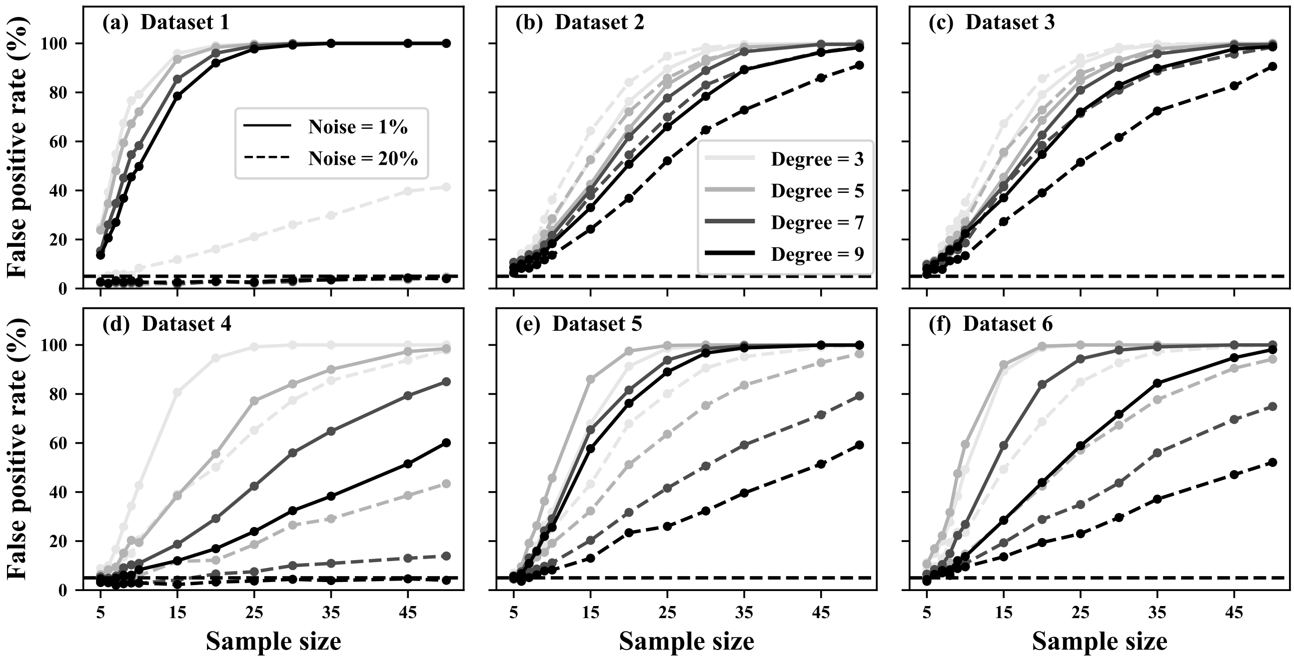

Figure 6. False positive results, GCVSPL filtering method.

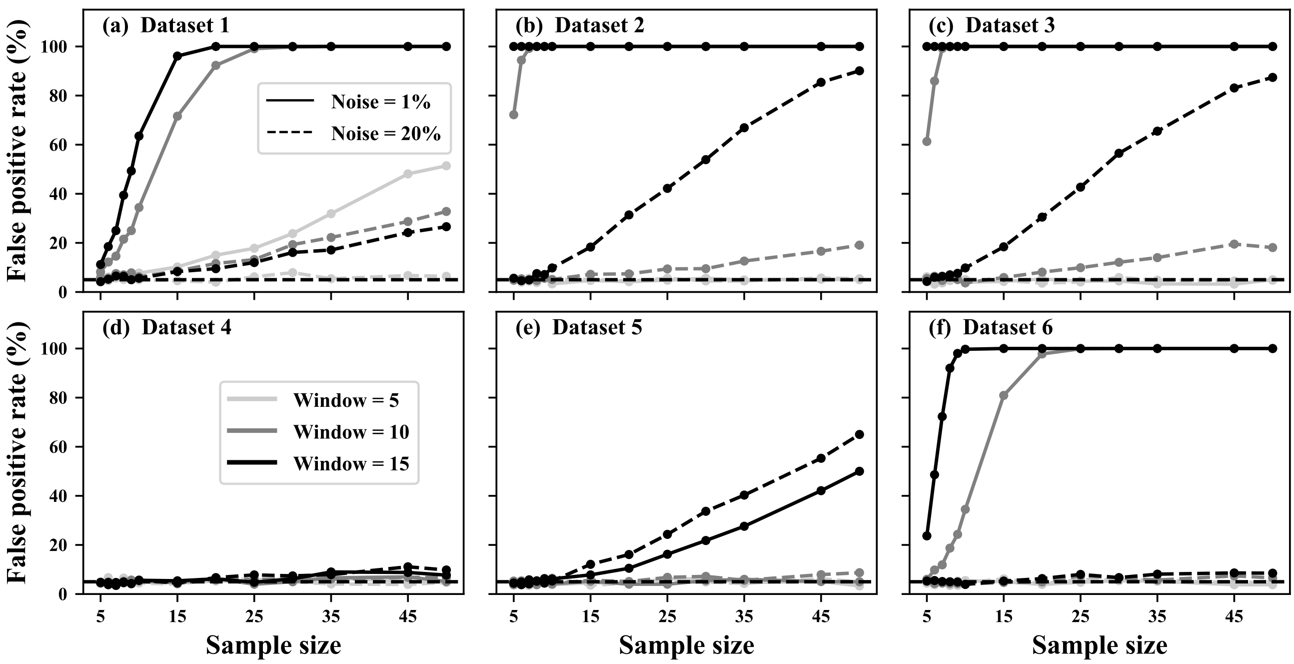

Figure 7. False positive results, SSA filtering method. Typical results, three components only; results were highly sensitive to the number of components selected (see Fig.A4 and A5, Supplementary Material).

Appendix A Supplementary figures

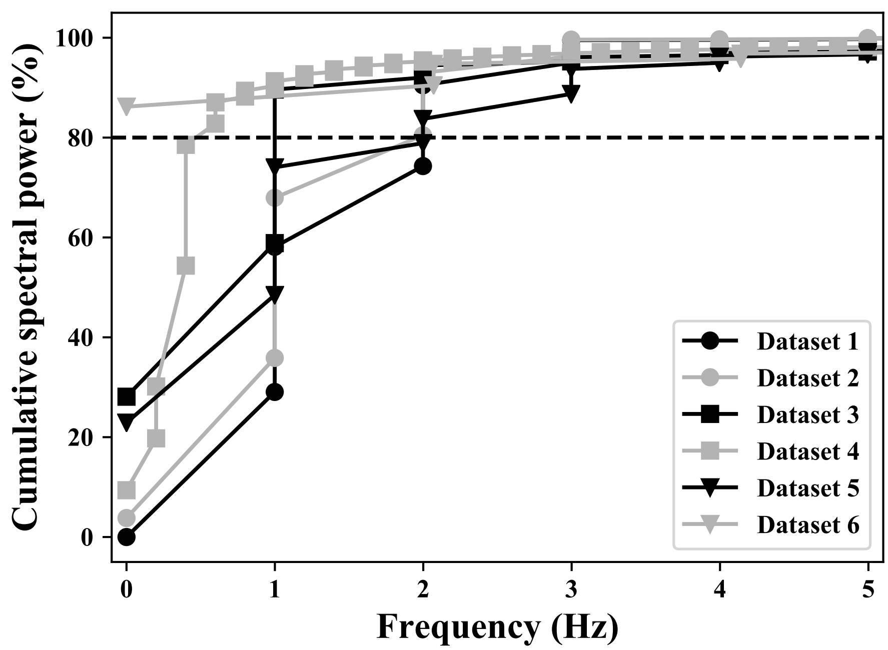

Figure A.1: Cumulative spectral power for each dataset. Frequencies less than 2 Hz contained at least 80% of cumulative power in all datasets.

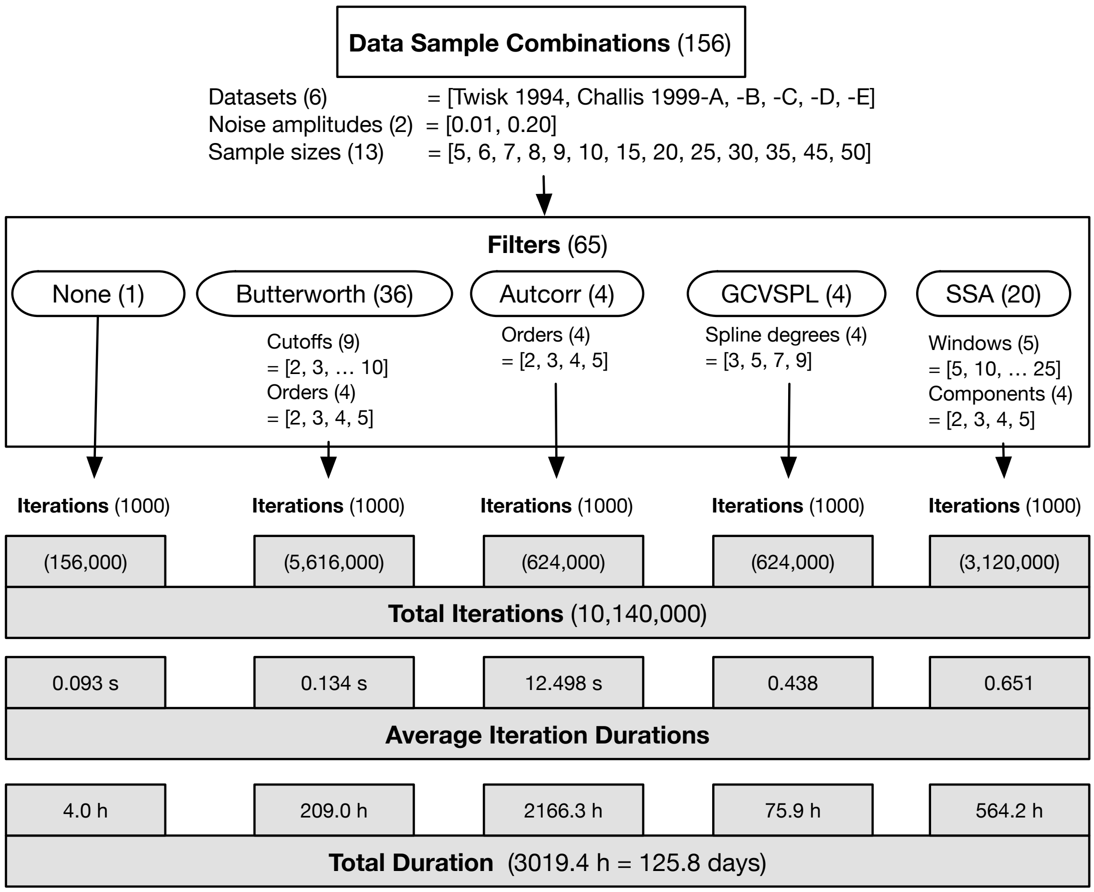

Figure A.2: Simulation summary. Numbers in parentheses indicate counts. Square brackets indicate parameters tested. A total of 156 data sample combinations and five filtering algorithms were used. Each filtering algorithms parameters were systematically varied as indicated, yielding a total of 65 uniquely parameterized filters. One-thousand iterations were conducted for each data-sample, filter combination in which Gaussian noise was added to the datasets 1D datum. Noise amplitude was normalized to each dataset so that the post-noise root mean square error with the 1D datum was either 1% or 20%. Total core computational duration was approximately 125 days.

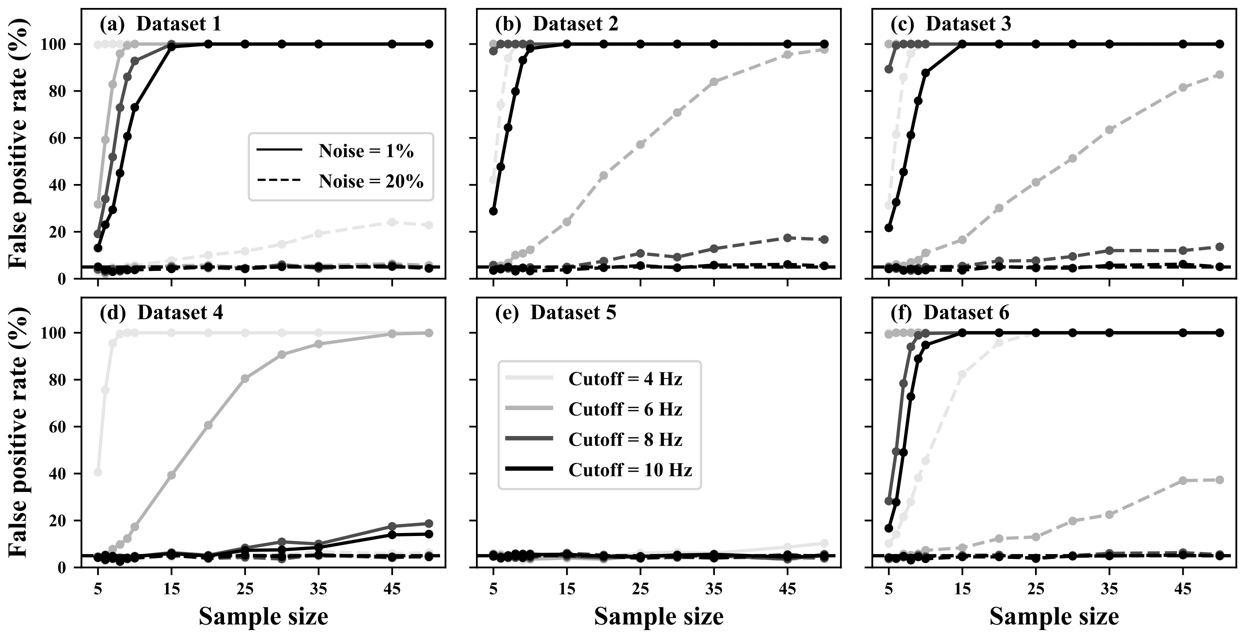

Figure A.3: False positive results, Butterworth filtering method (order=5). Results are similar to the order = 2 results presented in Fig.4, main manuscript.

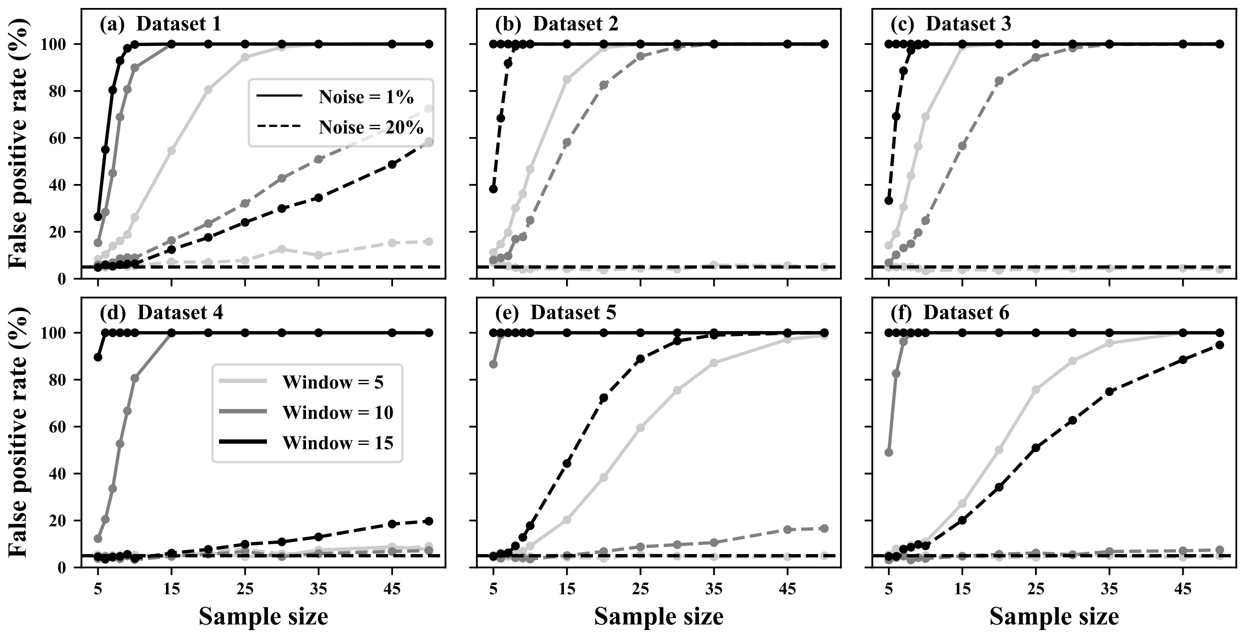

Figure A.4: False positive results, SSA filtering method (n=2, where n is the number of components). The n parameter greatly affected false positive results, with larger n yielding fewer false positives; results for n=3 and n=4 are presented in Fig.7 (main manuscript) and Fig.A.5, respectively.

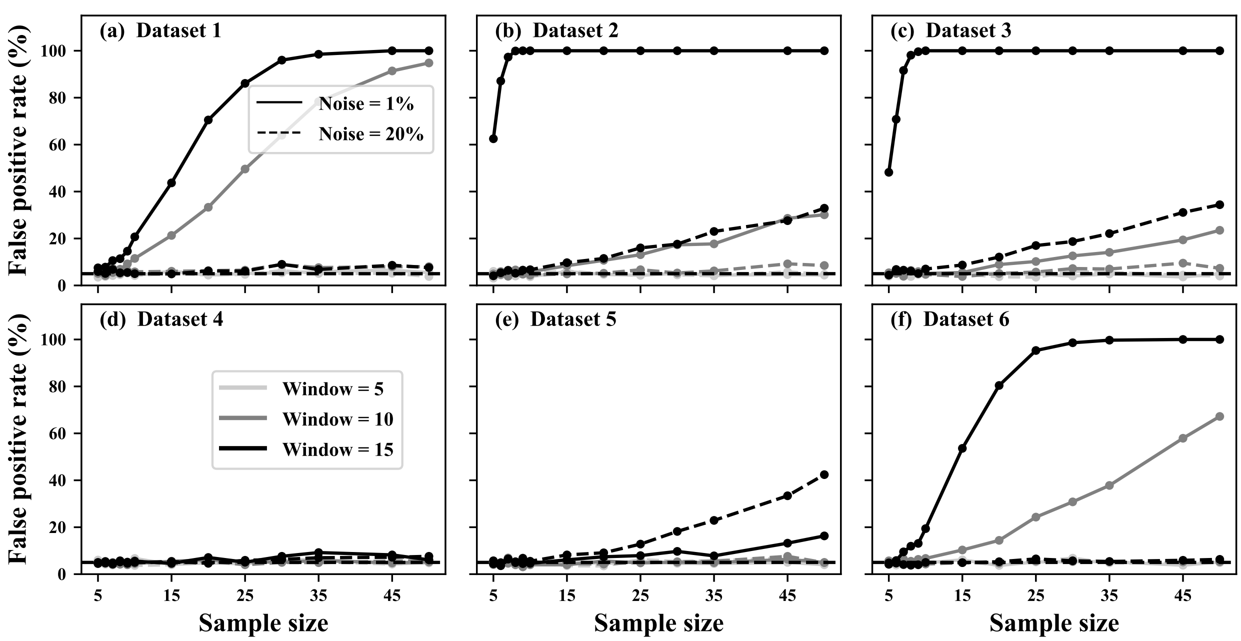

Figure A.5: False positive results, SSA filtering method (n=4), see Fig.A.4 caption.