

Comparison between four quantitative goodness-of-fit criteria for acceleration time series

C. Jiang^{1,2}, C.P. Hayden³, L. Wotherspoon⁴

¹PhD candidate, Department of Civil and Environmental Engineering, The University of Auckland, Auckland, New Zealand. Tel.: +64-273530025; Email Address: zjia110@aucklanduni.ac.nz

²Geotechnical Engineer, Andy O' Sullivan Geotechnical Engineering, Auckland, New Zealand. Tel.: +64-273530025; Email Address: catherine@aosullivan.co.nz

³Lecturer, Department of Civil and Environmental Engineering, The University of Auckland, Auckland, New Zealand

⁴Associate Professor, Department of Civil and Environmental Engineering, The University of Auckland, Auckland, New Zealand

ABSTRACT

Earthquake studies often use different models to simulate acceleration time series and analyse earthquake consequences. These simulated accelerations are then generally validated against experimental data or data from recordings in the field. In order to assess the effectiveness of these simulations, a method is needed to compare between acceleration time series and quantify the similarities. Many previous studies merely qualitatively assess performance based on the match of shape and magnitude aspects. To minimize the subjective nature of these choices, various authors have proposed numerical “goodness-of-fit” criteria to quantify the similarity between two acceleration time series. This paper summarises and compares four existing goodness-of-fit criteria that have different output presentations, including parameter scores and misfit distribution over a time-frequency representation. A sensitivity analysis is used to evaluate how different factors – including amplitude, phase, and frequency – influence the criteria evaluation results. The Anderson (2004) and Olsen & Mayhew (2010) criteria both have 10 parameters and easily identify the amplitude mismatch. The Kristeková et al. (2009) criterion generates plots to visualise the envelope and phase misfit distribution over both time and frequency domains. It has two parameters that respond, independently to amplitude and phase mismatch. The Zeghal et al. (2018) criterion uses three parameters to compare the acceleration time series in the frequency domain, but the proposed metrics are not able to identify the amplitude mismatch.

Keywords: goodness-of-fit, quantitative criteria, acceleration, time-frequency representation

1 INTRODUCTION

Ground acceleration is an important parameter in evaluating the consequences of earthquakes, hence seismic analysis generally includes the study of acceleration time series. Many studies have introduced numerical modelling to simulate and analyze the ground acceleration changes over time. These models require validation against experimental data or data from recordings in the field before further investigation.

In general, researchers may compare between acceleration time series to check the accuracy of the modelling assumption or select a best-fit model. Previous studies usually qualitatively evaluate them based on the how well the overall shape and the peak ground acceleration match (Hartzell 1982; Sabetta and Pugliese 1996). However, results based on subjective judgement are generally less convincing because people may perceive different level of similarity based on the same data. Instead, a quantitative criterion is preferred to evaluate the goodness-of-fit (GOF) between acceleration time series to ensure a consistent methodological approach. Early studies suggested some straightforward methods to achieve quantitative evaluation, including using a simple difference acceleration time series defined as the difference of two acceleration time series at each time step (Aoi and Fujiwara 1999) and using the root mean square of two acceleration time series (Geller and Takeuchi 1995). However, these methods are not always appropriate, as acceleration time series are generally complicated, and it is not enough to define the similarity between two records using a single parameter. When high frequency noise is present in the data, the frequency content and spikes also need to be considered as a measure of the match between records. In addition, researchers may have different focuses when evaluating the match between acceleration time series. Some may compare based on the physical meaning, such as the measures of peak amplitudes and response spectra, while others may consider the analytical results similarity (Taborda and Bielak 2013). Therefore, the need for standard and comprehensive criteria is warranted.

In the past two decades, researchers have proposed several GOF criteria that compare acceleration time series with different output representations. The purpose of this paper is to compare four existing criteria namely: the Anderson (2004) criterion, the Olsen & Mayhew (2010) criterion, the Kristeková et al. (2009) criterion and the Zeghal et al. (2018) criterion.

2 FOUR EXISTING GOODNESS-OF-FIT CRITERIA

The criteria mentioned in the previous section are slightly modified in this paper as the outputs of the four existing criteria are not directly comparable. In particular, Kristeková et al. (2009) generates plots while the other three calculate several parameter scores. This study proposes and modifies the score equations so that all the criteria output an overall score ranging from 0 (poor match) to 10 (excellent match) for later analysis. A scale is defined as: a score over 8 is an excellent fit, a score between 6-8 is a good fit, a score between 4-6 is a fair fit, and a score below 4 is a poor fit (Anderson 2004).

2.1 The Anderson (2004) criterion

Anderson (2004) proposed ten GOF parameters: Arias duration, Energy duration, Arias intensity, Energy integral, Peak acceleration, Peak velocity, Peak displacement, Response spectra, Fourier Spectra, and Cross correlation. Most of the parameters are widely used in earthquake studies, and they comprehensively compare a pair of acceleration time series, including comparison of spectral and magnitude aspects. A generic score equation – a negative exponential function (1) – is defined to evaluate the parameter scores:

$$S(p_1, p_2) = 10 \exp \left\{ - \left[\frac{(p_1 - p_2)}{\min(p_1, p_2)} \right]^2 \right\} \quad (1)$$

Where p_1 & p_2 are the parameter values of two records and S is the parameter score. Acceleration time series are evaluated by these parameters in ten frequency bands (nine narrow bands and one overall broad band) in an approximate logarithmic spacing: 0.05 – 0.1 Hz, 0.1 – 0.2 Hz, 0.2 – 0.5 Hz, 0.5 – 1.0 Hz, 1.0 – 2.0 Hz, 2.0 – 5.0 Hz, 5.0 – 10.0 Hz, 10.0 – 20.0 Hz, 20.0 – 50.0 Hz, and 0.05 – 50.0 Hz. The author recommended arithmetically averaging the parameter scores over the ten frequency bands to increase the weight of low frequencies, as they are generally more closely related to the waveform fitting.

To propose an overall score to summarize the 10 parameters, the current study proposes weighting parameters as they have different engineering significance, with Table 1 summarizing the weight allocation. Although the Fourier spectra and Response spectra are generally analysed simultaneously (Bolisetti 2015; Dang and Liu 2020), researchers usually consider the structural response more than the frequency content. Therefore, the response spectrum receives the highest weight (20%), and the Fourier spectrum receives 15%. The measures of peak amplitudes occupy a large weight (35% in total), as they are the commonly used and straightforward properties of an acceleration time series. Various studies have indicated that PGV is better correlated to earthquake destructive consequences (Orense 2005; Yaghmaei-Sabegh et al. 2011), so the weights are allocated accordingly (15% for PGV and 10% for PGA and PGD). Arias intensity and Energy integral are two pairs of parameters that highly correlate to each other, so each of them takes a relatively low weight (5%). Cross-correlation usually relates to the phase shifts, and it receives an average weight of 10%.

Table 1. The parameters and their weight allocation of Anderson (2004) and Olsen & Mayhew (2010)

Parameter number	Parameters of Anderson 2004 criterion	Weights (%)	Parameters of Olsen & Mayhew 2010 criterion	Weights (%)
1	Fourier spectrum (FS)	15	FS	15
2	Response spectrum (RS)	20	RS	10
3	Arias duration (DURA)	5	SA16	10
4	Arias intensity (AI)	5	IE ratio	10
5	Energy duration (DURE)	5	DURE	5
6	Energy integral (EI)	5	EI	5
7	Peak acceleration (PGA)	10	PGA	10
8	Peak velocity (PGV)	15	PGV	15
9	Peak displacement (PGD)	10	PVD	10
10	Cross correlation (Xcor)	10	Xcor	10

2.2 The Olsen & Mayhew (2010) criterion

Olsen & Mayhew (2010) also defined the criterion with ten parameters, and the current study uses a weighted average as the overall score with a similar weight distribution (details shown in Table 1). Most of the parameters are the same as the previous criterion, but the authors replaced DURA and AI with SA16 and IE ratio, respectively. The SA16 is the response spectral acceleration at 16 individual periods used by recent NGA relations and IE ratio is the ratio between inelastic and elastic response spectra. The weights are modified accordingly. The SA16 and RS are both response spectra with different input periods, so they share the 20% weight. The IE ratio is of interest to structural engineers and receives 10 % weight. It then defines the error function (2) as its score equation:

$$S = A \times \operatorname{erfc} \left[\frac{2|p_1 - p_2|}{p_1 + p_2} \right] \quad (2)$$

Where p_1 & p_2 are the parameter value of two records, S is the parameter score, and A is the multiplier. The original study used $A = 100$, while the current study uses $A = 10$ for consistency. Figure 1 shows the score distributions of the two criteria, with normalized scores plotted against parameter ratios. A steep slope between parameter ratios 1 – 1.4 indicates that this criterion is more sensitive to small misfits, and the score changes less when the parameter ratio is greater than 2. In contrast, Anderson (2004) has a more uniform distribution as scores change almost linearly for parameter ratio values between 1.3 – 2.1.

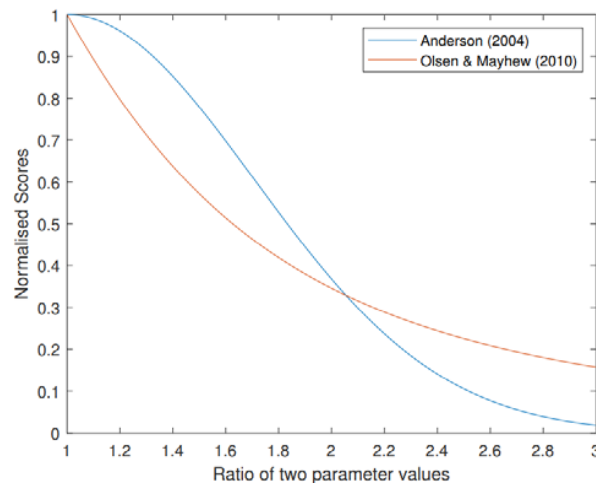


Figure 1. The plot of normalized score distribution against parameter ratio for the Anderson and the Olsen & Mayhew criterion.

2.3 The Kristeková et al. (2009) criterion

Kristeková et al. introduced a misfit criterion based on the time-frequency representations of two acceleration time series (Kristeková et al. 2006), then improved and extended the criterion (Kristeková et al. 2009). The main feature of this criterion is the visualisation of envelope and phase misfit distribution over the time-frequency (TF) representation. The authors defined a series of parameters to compare the signals. For example, the TF envelope misfit (TFEM) and TF phase misfit (TFPM) evaluate the envelope and phase differences on both time and frequency domains, and the respective GOF scores (TFEG and TFPG) are derived from them. All these parameters are matrices measuring the misfits or GOF over the time-frequency ranges, so that plots can be generated to visualise the misfit distribution. According to the definition, the parameter matrices generally have a large proportion of small misfits or high GOF, and it is difficult to summarize the parameter. Therefore, the current study modifies TFEG equation (3) and TFPG equation (4) so that they have a broader score range for later analysis and defines the overall parameter as the mean of the matrix.

$$\text{TFEG}(t, f) = A \exp\{-|\text{TFEM}(t, f) * 10|^k\} \quad (3)$$

$$\text{TFPG}(t, f) = A \exp\{-|\text{TFPM}(t, f) * 10|^k\} \quad (4)$$

Where factor A quantifies the agreement between two signals with lower misfits leading to higher GOF (e.g., when TFEM is 0, TFEG is A), and k determines the sensitivity of the GOF with respect to misfit. The following analysis uses A = 10 and k = 1.

2.4 The Zeghal et al. (2018) criterion

Zeghal et al. investigated the acceleration time series in the frequency domain by firstly converting the time series to Fourier series, and then compares them based on three aspects: phase, shape, and frequency shift. The authors defined the overall parameter d_{ij} as (5) ranging from 1 (out of phase) to 0 (in phase):

$$d_{ij} = d_{ij}^{\text{phase}} + d_{ij}^{\text{shape}} + d_{ij}^{\text{Fshift}} \quad (5)$$

The parameter d_{ij} is independent of d_{ij}^{shape} in magnitude, as d_{ij}^{shape} is deducted in d_{ij}^{Fshift} by its definition. Therefore, the current study considers d_{ij} and d_{ij}^{shape} separately. To allow a consistent scale with the other criteria, D is defined as (6) with a score range of 0 (out of phase) – 10 (in phase):

$$D = 10 \times (1 - d_{ij}) \quad (6)$$

3 SENSITIVITY ANALYSIS

To assess each criterion, a Gabor signal function (7) is used, which is a harmonic carrier with a Gaussian envelope:

$$\text{Signal} = a * \exp[-0.3(t - t_0)^2] \times \cos[2\pi f(t - t_0) + \varphi\pi] \quad (7)$$

The Reference signal has the following parameters: $a = 2$, $t_0 = 5$, $f = 1$, $\varphi = 0$. To evaluate how amplitude, phase, and frequency influence the criteria evaluation, three signals are defined as: $a = 4$ for the Amplitude signal, $\varphi = 0.4$ for the Phase signal, and $f = 1.5$ for the Frequency signal, while all the other parameters remain unchanged. Figure 2 presents the four signals, with each criteria used to evaluate the GOF of each modified signal against the Reference signal.

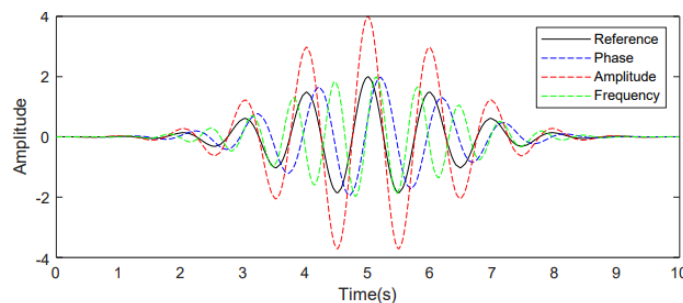


Figure 2. Test signal plots for the sensitivity analysis

Table 2 summarises the results of the four criteria and their evaluation of the GOF of the Amplitude, Phase and Frequency signals compared with the Reference signal. The first two criteria each have one overall score based on the weighted average of 10 parameters, and the last two criteria each have two parameters evaluating different aspects. Compared with a score of 9.3 for the Phase pair, Anderson (2004) gives lower scores for the Amplitude and Frequency pairs of 4.8 and 4.1, respectively. This criterion is highly sensitive to amplitude and frequency differences. Olsen & Mayhew (2010) also evaluates low scores for the Amplitude and Frequency pairs (4.8 and 4.5) compared with the Phase pair (9.1). By contrast, both parameters of Kristeková et al. (2009) give high scores for all cases (> 7). TFEG responds more to the Amplitude and Frequency pairs (8.0 and 7.5), while TFPG is more sensitive to the Phase and Frequency pairs (8.7 and 8.7). These high scores indicate that the overall parameter values are less sensitive to these changes, and it is necessary to analyze the output plots to add more information to the criterion evaluation. Figure 3 displays the TFEG and TFPG plots for the Frequency pair, with both showing a poor fit between 0.5 – 2 Hz throughout the time range, which indicates that the two signals have large misfit within the specified range and reasonable misfit outside the range. The two parameters of the Zeghal et al. (2018) perform differently, with D sensitive to the Phase (6.6) and

Frequency (5.0) pairs compared with the Amplitude pair (9.0), while D^{shape} remains almost unchanged for all factor changes. This criterion cannot identify amplitude mismatches and D^{shape} plays a relatively unimportant role in most earthquake studies.

Table 2. Sensitivity analysis summary table

Criteria name	Score name	Score values of different cases		
		Phase	Amplitude	Frequency
Anderson (2004)	Weighted score	9.3	4.8	4.1
Olsen & Mayhew (2010)	Weighted score	9.1	4.8	4.5
Kristeková et al. (2009)	TFEG	10.0	8.0	7.5
	TFPG	8.7	10.0	8.7
Zeghal et al. (2018)	D	6.6	9.0	5.0
	D^{shape}	10.0	9.9	10.0

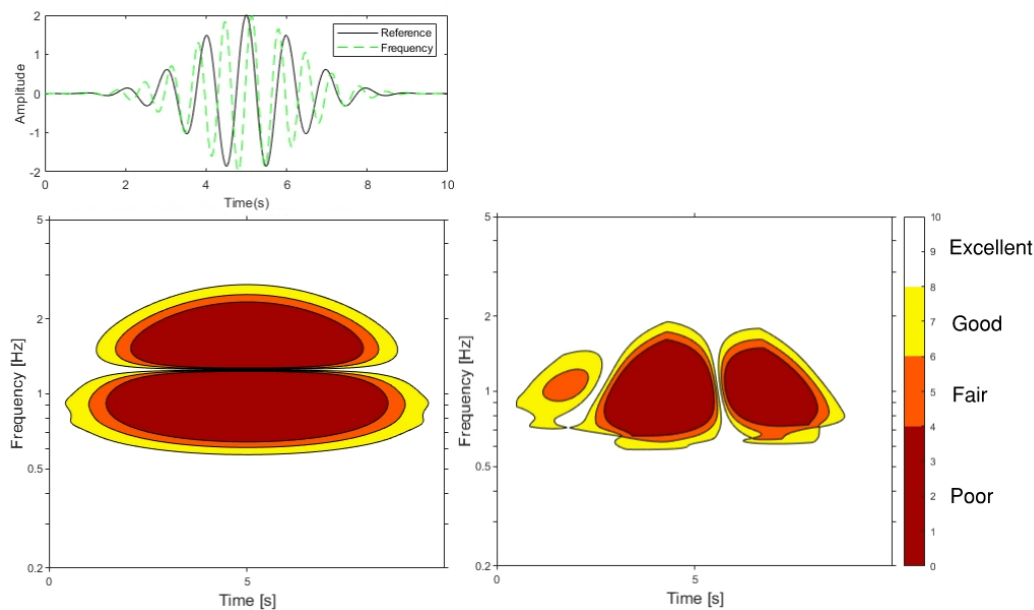


Figure 3. The GOF plots of the Kristeková et al. (2009) criterion comparing the Frequency signal with the Reference signal in the sensitivity analysis. (a) The Reference signal and the Frequency signal; (b) TFEF plot and (c) TFPG plot with scales.

Overall, Anderson (2004) defines frequency filtering of the acceleration time series as the first step to minimizing the influence of high and low-frequency motions which contribute less to the earthquake consequences. The parameter scores are closely related to frequency and amplitude differences, which most earthquake studies generally focus on. However, the parameter scores cannot indicate whether the simulation overestimates or underestimates the actual data and are less sensitive to small differences than significant differences as shown in Figure 1. Therefore, although this criterion is desirable to most applications, it may benefit from further improvements, such as adding signs to the scores to identify underestimation or overestimation and providing extra score equation choices according to different requirements.

Olsen & Mayhew (2010) has a higher resolution for small differences as shown in Figure 1. The sensitivity analysis shows that it responds slightly less to frequency changes as it does not include frequency filtering, and the updates to the Anderson criteria have a limited influence on the scores. Even so, structural engineers may appreciate the addition of the new parameter IE ratio to evaluate the building behaviors. This criterion is a reasonable choice for many studies, but it could be further improved by combining with the advantages of Anderson (2004).

Kristeková et al. (2009) generates plots to enable a detailed analysis of the acceleration time series comparison, as researchers can separate envelope and phase misfits and clearly visualize the misfit distribution over a time-frequency representation. However, earthquake studies usually need to compare numerous acceleration time series, so analyzing each plot individually is less efficient

compared with other criteria. As the criterion outputs matrices of scores, this study proposes an overall criterion score to summarize the performance differences between a pair of acceleration time series. In general, this criterion is desirable if it is used for specific case studies or small-scale comparisons and is helpful to improve a certain simulation as it presents the most detail, while it is less effective when used in a large scale.

Zeghal et al. (2018) compares acceleration time series based on the frequency domain. It is less applicable to earthquake studies compared with the other three criteria, as the parameters are less effective when evaluating factor changes. In the sensitivity analysis, the overall parameter responds more to phase difference compared with amplitude difference, and the shape parameter gives stable and high scores.

4 CONCLUSIONS

This paper compared four GOF criteria that quantify the similarity between two acceleration time series. The sensitivity analysis indicates that Anderson (2004) and Olsen & Mayhew (2010) are sensitive to amplitude difference. They can easily summarize the similarities between time series and suggest the one with better performance, so they are appropriate for most of earthquake studies, such as selecting a best-fit models and justifying modelling assumptions. Kristeková et al. (2009) has two parameters that measure amplitude and phase differences separately and generates several misfit distribution plots. This criterion is more suitable for specific case studies or to improve a certain model. Zeghal et al. (2018) is not very sensitive to either amplitude or frequency differences, which means the criterion is less applicable to general earthquake studies. These criteria focus on different aspects and have distinct features, so researchers can choose the most suitable one based on their requirements.

5 ACKNOWLEDGEMENTS

The authors gratefully acknowledge Earthquake Commission (EQC) for the financial support and Andy O'Sullivan Geotechnical Engineering for the technical support.

REFERENCES

- Anderson, J. G. (2004). "Quantitative measure of the goodness-of-fit of synthetic seismograms." Proc., Proceedings of the 13th World Conference on Earthquake Engineering, International Association for Earthquake Engineering, 243.
- Aoi, S., and Fujiwara, H. (1999). "3D Finite-Difference Method using Discontinuous Grids." *Bulletin of the Seismological Society of America*, 89(4), 918-930.
- Bolisetti, C. (2015). *Site Response, Soil-Structure Interaction and Structure-Soil-Structure Interaction for Performance Assessment of Buildings and Nuclear Structures*, State University of New York at Buffalo.
- Dang, P., and Liu, Q. (2020). "Stochastic Finite-Fault Ground Motion Simulation for the Mw 6.7 Earthquake in Lushan, China." *Nat. Hazards*, 100(3), 1215-1241.
- Geller, R. J., and Takeuchi, N. (1995). "A New Method for Computing Highly Accurate DSM Synthetic Seismograms." *Geophysical Journal International*, 123(2), 449-470.
- Hartzell, S. (1982). "Simulation of Ground Accelerations for the May 1980 Mammoth Lakes, California, Earthquakes." *Bulletin of the Seismological Society of America*, 72(6A), 2381-2387.
- Kristeková, M., Kristek, J., Moczo, P. (2009). "Time-Frequency Misfit and Goodness-of-Fit Criteria for Quantitative Comparison of Time Signals." *Geophysical Journal International*, 178(2), 813-825.
- Kristeková, M., Kristek, J., Moczo, P., Day, S. M. (2006). "Misfit Criteria for Quantitative Comparison of Seismograms." *Bulletin of the Seismological Society of America*, 96(5), 1836-1850.
- Olsen, K. B., and Mayhew, J. E. (2010). "Goodness-of-Fit Criteria for Broadband Synthetic Seismograms, with Application to the 2008 Mw 5.4 Chino Hills, California, Earthquake." *Seismol. Res. Lett.*, 81(5), 715-723.
- Orense, R. P. (2005). "Assessment of Liquefaction Potential Based on Peak Ground Motion Parameters." *Soil Dyn. Earthquake Eng.*, 25(3), 225-240.
- Sabetta, F., and Pugliese, A. (1996). "Estimation of Response Spectra and Simulation of Nonstationary Earthquake Ground Motions." *Bulletin of the Seismological Society of America*, 86(2), 337-352.
- Taborda, R., and Bielak, J. (2013). "Ground-motion Simulation and Validation of the 2008 Chino Hills, California, Earthquake." *Bulletin of the Seismological Society of America*, 103(1), 131-156.
- Yaghmaei-Sabegh, S., Tsang, H., Lam, N. T. (2011). "Conversion between Peak Ground Motion Parameters and Modified Mercalli Intensity Values." *J. Earthquake Eng.*, 15(7), 1138-1155.
- Zeghal, M., Goswami, N., Manzari, M., Kutter, B. (2018). "Discrepancy Metrics and Sensitivity Analysis of Dynamic Soil Response." *Geotechnical Earthquake Engineering and Soil Dynamics V: Numerical Modeling and Soil Structure Interaction*, American Society of Civil Engineers Reston, VA, 115-122.