

Evaluation of Time History-Based Metrics for Validating Nonlinear Deformation Analyses of Liquefiable Geosystems

Maziar Mivehchi, S.M.ASCE¹; and Katerina Ziotopoulou, Ph.D., P.E., M.ASCE²

¹Ph.D. Student, Dept. of Civil and Environmental Engineering, Univ. of California, Davis, Davis, CA. Email: mmivehchi@ucdavis.edu

²Associate Professor, Dept. of Civil and Environmental Engineering, Univ. of California, Davis, Davis, CA. Email: kziotopoulou@ucdavis.edu

ABSTRACT

Non-linear dynamic analyses (NDAs) are widely used in engineering practice to evaluate the seismic performance of geotechnical structures affected by liquefaction or cyclic softening. The quality of results from an NDA study depends on several technical and nontechnical factors. Validation, wherein a numerical prediction is compared to its physical counterpart, can assess the ability of an NDA to capture the various metrics of the response and potentially provide guidance toward improving the prediction. This study investigates select methodologies and validation metrics commonly used in signal processing problems to assess their effectiveness in capturing discrepancies between experimental and simulation results for a specific response of interest. Three simple problems are initially evaluated to analyze the metrics' capabilities and identify necessary modifications. Then, the metrics are applied to nine sets of experimental and simulation time series, focusing on one response of interest (pore water pressure). It is found that cross-correlation successfully captures the lag in the initiation of liquefaction triggering, while Russell's error metric captures magnitude and phase discrepancies.

INTRODUCTION

The performance evaluation of liquefiable geosystems is usually achieved through nonlinear deformation analyses (NDAs). The capabilities of NDAs in considering more sophisticated system behaviors, like cyclic mobility or cyclic softening of sands and clays respectively, structural dynamic response, and complex geometries have proven to be advantageous in estimating seismic performance, especially compared to simpler analysis methods (e.g., Boulanger, 2022). However, various factors can impact the quality of an NDA evaluation. The site characterization, the selection of the constitutive model(s), the model calibration process, , and broader numerical uncertainties are some of the factors that can affect the outcomes of a numerical evaluation (Boulanger and Ziotopoulou, 2018). Validating NDAs or parts thereof with experimental measurements can help refine some of these uncertainties, detect sources of discrepancy, and improve the reliability of NDA predictions.

Oberkampf et al. (2002) described how verification and validation (V&V) can be key tools for building and assessing the reliability of numerical analyses and elaborated on the concepts and details behind these two processes. The American Society of Mechanical Engineers Standards (ASMES) defines model validation as the qualitative comparison of computational and experimental results for a response of interest and the determination whether the agreement between the experiment and simulation is acceptable or not (Sarin et al., 2010). In the case of NDAs, an acceptable agreement in a validation study for a particular geosystem does not always

guarantee a satisfactory prediction outcome in another geosystem. This is because the driving loading paths, materials, and boundary conditions may not be the same and a reasonable performance for one combination thereof cannot be extrapolated. Last but not least, the validation of NDAs has been mostly assessed visually (e.g., Ziotopoulou, 2018), in a more subjective framework, rather than rigorously quantifying the discrepancy. This qualitative approach, called viewgraph norm, can potentially hinder the unbiased objective evaluation of parametrically varied predictions and through that the improvement of the overall procedures.

Various quantitative metrics with a broad range of characteristics and behaviors have been proposed in the engineering literature that enable the comparison between experimental measurements and numerical simulation results. Their aim is commonly to capture and explain the discrepancy between the two (Schwer, 2007). According to Zhan et al. (2011), an ideal validation metric has five characteristics, namely it should: (1) produce the same result regardless of the operator, (2) be suitable for generalization, (3) satisfy the symmetrical behavior (meaning it should yield the same results even if the numerical and experimental datasets switch with each other), (4) account for data uncertainty, and (5) have a physical meaning and confirm engineering knowledge. Although a usable metric may not satisfy all these characteristics, for NDA-based analyses, it should honor the time-dependent nature of the phenomenon. Therefore, it is essential that the selected validation metric captures the discrepancy source and explains whether the error is due to magnitude, phase, or shape variations.

This paper presents an initial investigation on select methodologies and metrics that have been used in signal processing problems with the goal of evaluating their effectiveness in capturing the discrepancy between experimental and simulation results on one response of interest. The selected metrics investigated in this study are Sprague and Geers (2004), Russell's Error (1997), and Cross-Correlation (Liu et al., 2005). The experimental data selected to perform the investigation on, are those obtained from the centrifuge model test number 10 of the Liquefaction Experiments and Analysis Project, performed at the Rensselaer Polytechnic Institute (RPI) centrifuge facility (RPI-10). Respectively, numerical data were obtained from the simulation of the same test in FLAC 8.1 (Itasca, 2019), and with the constitutive model PM4Sand (Boulanger and Ziotopoulou, 2022) applied to the liquefiable sands.

VALIDATION METRICS

The three metrics chosen herein to evaluate the discrepancy between simulation and experimental data are Sprague and Geers (S&G), Russell's Error, and Cross Correlation. S&G (2004) proposed Equations 1 and 2 for calculating the magnitude and phase errors, respectively, between the computed (simulation) and measured (experiment) data.

$$M_{S\&G} = \sqrt{\frac{\vartheta_{cc}}{\vartheta_{mm}}} - 1 \quad (1)$$

$$P_{S\&G} = \frac{1}{\pi} \cos^{-1} \left(\frac{\vartheta_{cm}}{\sqrt{\vartheta_{cc}\vartheta_{mm}}} \right) \quad (2)$$

where

$$\vartheta_{cc} = \frac{\sum_{i=1}^N a_i^2}{N}, \quad \vartheta_{mm} = \frac{\sum_{i=1}^N b_i^2}{N}, \quad \vartheta_{cm} = \frac{\sum_{i=1}^N a_i b_i}{N} \quad (3)$$

In the above equations, ϑ_{cc} is the computed vector, ϑ_{mm} is the measured vector, and ϑ_{cm} is the computed-measured vector. These vectors are defined in Equation 3 based on the computed values (a_i) and measured values (b_i) at each time increment, and the total number of recorded values (N). Russell's Error and S&G have a similar phase error calculation but differ in the magnitude error calculation. The Russell's Error equation for calculating the magnitude error (Equation 4) is defined in a way to produce values in a similar scale as the phase error defined in Equation 2. Cross-correlation, or sliding dot product, represented by Equation 5, is a modification to the concept of coefficient of correlation, used to measure phase lag between two time histories (Sarin et al., 2010).

$$M_R = \text{sign}(\vartheta_{cc} - \vartheta_{mm}) \log^{10} \left(1 + \left| \frac{\vartheta_{cc} - \vartheta_{mm}}{\sqrt{\vartheta_{cc} \vartheta_{mm}}} \right| \right) \quad (4)$$

$$\rho(n) = \frac{(N-n) \sum_{i=1}^{N-n} a_i b_{i+n} - \sum_{i=1}^{N-n} a_i \sum_{i=1}^{N-n} b_{i+n}}{\sqrt{(N-n) \sum_{i=1}^{N-n} a_i^2 - (\sum_{i=1}^{N-n} a_i)^2} \sqrt{(N-n) \sum_{i=1}^{N-n} b_{i+n}^2 - (\sum_{i=1}^{N-n} b_{i+n})^2}} \quad (5)$$

where $n = 0, 1, 2, \dots, N-1$. The maximum value of $\rho(n^*)$ will be used to shift the simulation relative to the experiment by n^* step in order to measure the phase lag. Cross-correlation results in values between -1 and 1. Higher absolute values indicate stronger agreement, while values closer to 0 indicate greater phase difference between the simulation and experiment.

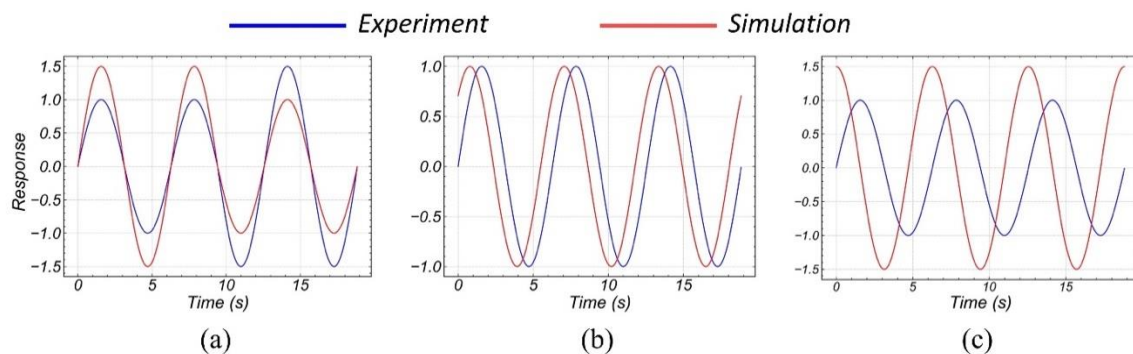


Figure 1. Three simple problems defined with having (a) only magnitude error, (b) only phase error, and (c) phase and magnitude errors.

APPLICATION OF METRICS IN SIMPLE PROBLEMS

Before evaluating the metrics for the LEAP model test, the ability of the metrics for capturing different sources of discrepancies is analyzed on three simple known problems. These problems are illustrated in Figure 1 and are all defined with simple harmonic waves for both the experimental and simulation datasets. One example problem illustrates magnitude error, one phase error, and one the combination of magnitude and phase errors. In the first problem, simulation produces a 50% higher magnitude than the experiment for the first half of the time domain, then the experiment has 50% higher magnitude compared to the simulation. In the second problem there is no magnitude difference, but the simulation is 45° ahead of the experiment. In the third problem, the simulation has 50% higher magnitudes with a 90° phase lag

compared to the experiment throughout. The goal of this simple problem analysis is to identify any necessary modification to the metrics to better capture and distinguish the errors due to magnitude and phase differences. Once the modifications are applied, the validation is performed and analyzed on the experiment and simulation pore water pressure (pwp) dataset at all 9 locations using the modified metric.

Results for Simple Problems. S&G and Russell's Error metrics are used to assess these problems in "Delta-T" and "Cumulative" formats. Delta-T plots show the calculations at individual time intervals, whereas the cumulative plot shows the accumulated calculations over the time history.

Figures 2 and 3 show the performance of S&G and Russell's Error metrics on the three simple problems. In the first problem (only magnitude error), both S&G and Russell's Error correctly capture a zero-phase difference between the simulation and magnitude. The magnitude error captured by S&G metric in the second half of the problem 1 is different than the first half, while Russell's Error remains consistent. This issue comes from the fact that S&G does not consider a symmetrical behavior for calculating the magnitude error. Therefore, it produces a different result when the simulation and experiment time series switch their places in this problem. The second problem shows evolving magnitude errors in S&G despite being designed with zero error. Besides, S&G calculates the magnitude error on a larger scale than phase errors, unlike Russell's Error metric. The greater magnitude scale in S&G metric reduces the significance of the phase error, as can be seen by comparing the second and third problems in Figures 2 and 3. Problem three has the same magnitude difference as problem one. However, both S&G and Russell's Error show a different magnitude error trend compared to problem one due to the influence of phase error. Therefore, minimizing the phase difference is essential to measure the magnitude error more accurately.

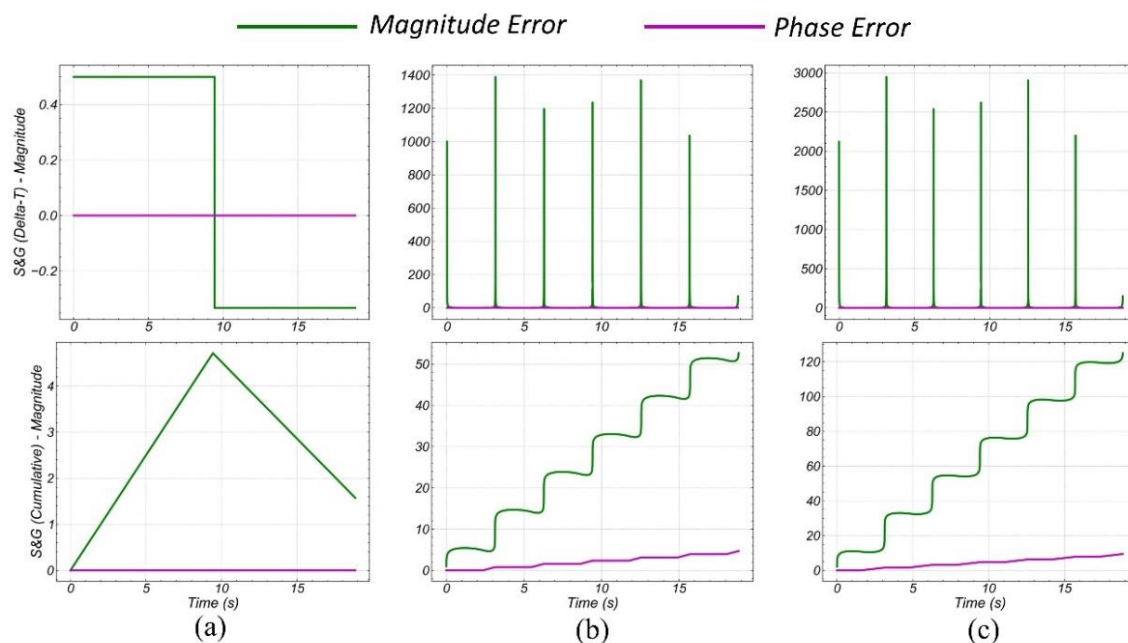


Figure 2. The performance of S&G metric on problems with (a) only magnitude error, (b) only phase error, and (c) phase and magnitude errors in both 'Delta-T' (1st row) and 'Cumulative' (2nd row) format.

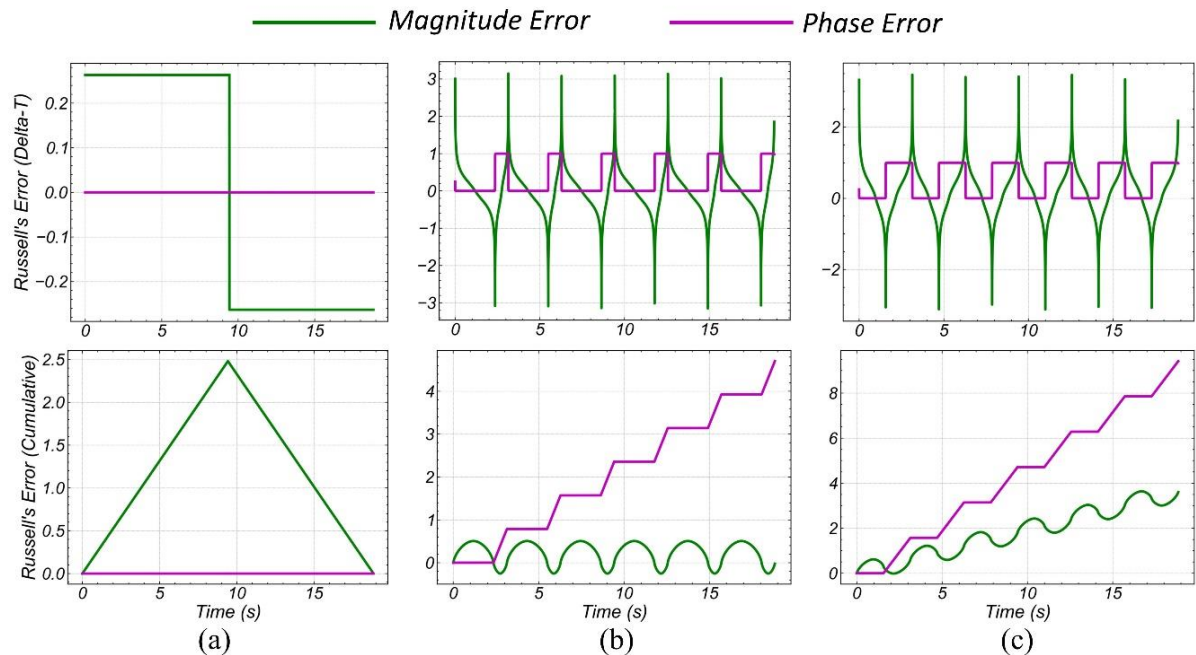


Figure 3. The performance of Russell’s Error metric on problems with (a) only magnitude error, (b) only phase error, and (c) phase and magnitude errors in both ‘Delta-T’ (1st row) and ‘Cumulative’ (2nd row) format.

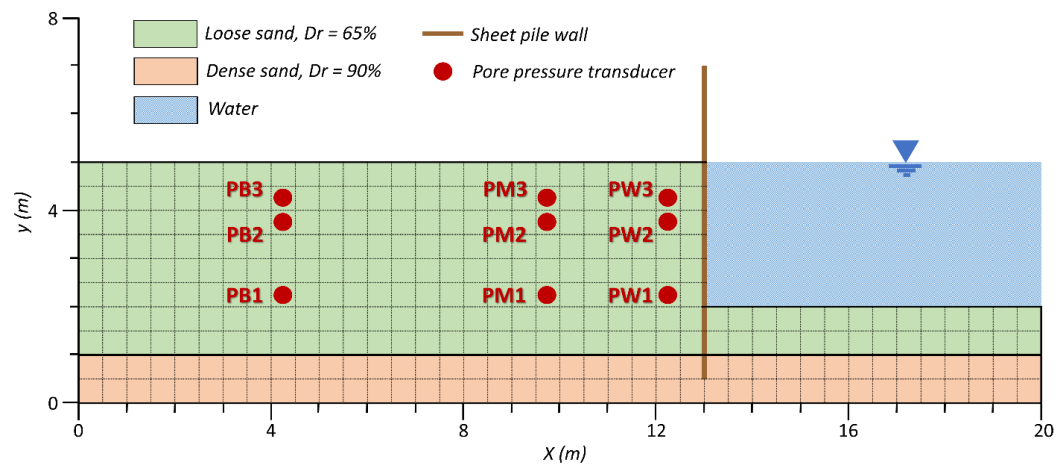


Figure 4. Schematic cross-section of the LEAP-2020 centrifuge model test sheet-pile-wall geosystem (after Basu et al., 2022).

APPLICATION OF METRICS IN LEAP-2020

The focus of the LEAP-2020 exercise was a soil-sheet-pile wall system which was studied experimentally (centrifuge) and numerically. Eleven experiments were conducted at six different centrifuge facilities, all following the identical geosystem illustrated in Figure 4. A deposit of 65% relative density (D_R) liquefiable sand was overlying a 90% D_R sand (both were Ottawa F-65

sand). A sheet-pile wall with specified properties was retaining the liquefiable sand named ‘backfill’. This centrifuge model was spun up to the target centrifugal acceleration, shaken with a predetermined input motion, and different system responses (pore pressures, accelerations, displacements, and settlements) were recorded at multiple locations. Figure 4 shows the locations of the pore pressure transducers on the ‘backfill’ side of the wall. More information about the centrifuge experiments as part of the LEAP-2020 exercise can be found in Basu et al. (2022) and Perez et al. (2023). Herein, the measured pwp from the RPI-10 centrifuge test, at the nine different locations shown in the Figure 4 are selected for the validation study. This geosystem was modeled using the numerical platform FLAC 8.1 and the responses were simulated using the nonlinear constitutive model PM4Sand Version 3.2. The calibration of the model was performed through single-element simulations of undrained cyclic stress-controlled direct simple shear (DSS) tests. All details can be found in Basu et al. (2022). Figure 5 shows the experimental recordings along with the simulation results in terms of the excess pore water pressure measured at nine different locations of the RPI-10 centrifuge model test.

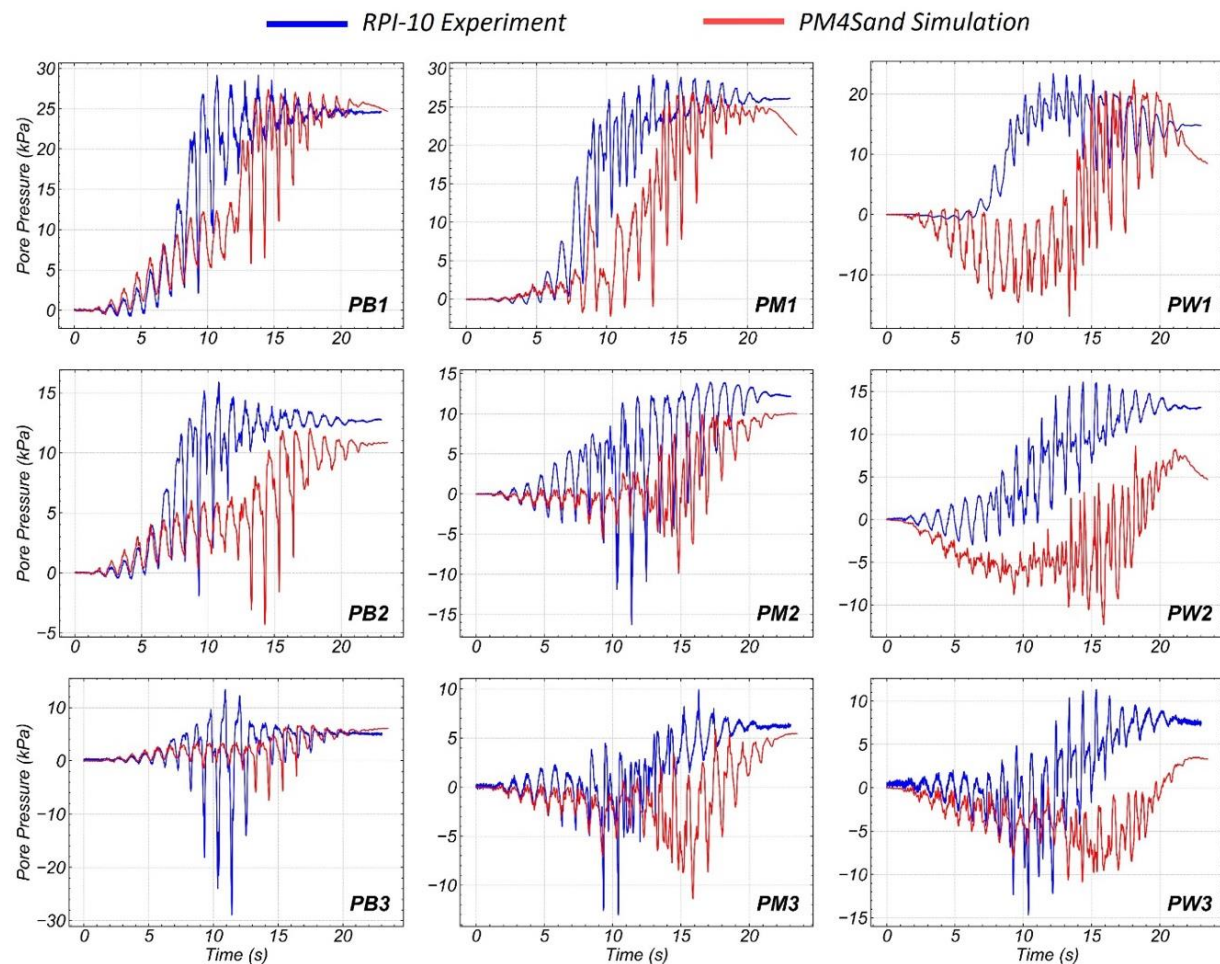


Figure 5. Experimental versus baseline simulation results of porewater pressures at locations of Figure 4. Variable y-axis scales are used in order to allow the most comfortable viewing of the results.

RESULTS AND DISCUSSION

After evaluating the metrics on three simple problems, Russell's Error is chosen for LEAP-2020 measurements. Cross correlation is first used for measuring the phase lag between the experiment and simulation, so that it can be minimized before implementing the metric. Figure 6 displays the cross-correlation results and corresponding $\rho(n^*)$ values for each location. According to the $\rho(n^*)$ values, agreement between the simulation and experiment time series weakens at shallower depths (PB3, PM3, PW3) and closer to the retaining wall (PW1, PW2, PW3), as $\rho(n^*)$ decreases. The calculated n^* ranges from 3s to 7s for all locations, representing the time difference in liquefaction triggering initiation. This time of initiation strongly depends on the soil cyclic strength that was used in the model calibration process. Higher cyclic strength delays the triggering. Like $\rho(n^*)$, n^* increases near the retaining wall, indicating lower cyclic strength at those locations. Figure 7 shows the time series after shifting the simulation by n^* steps from cross correlation. Better matching occurs in the liquefaction triggering phase, but the simulation has greater pre-liquefaction magnitudes compared to Figure 5. The level of agreement between the simulation and experiment at different locations in Figure 7 correlates with the calculated $\rho(n^*)$ values shown in Figure 6. The comparisons at PB and PM locations show a better agreement (higher $\rho(n^*)$) than at PW locations (lower $\rho(n^*)$).

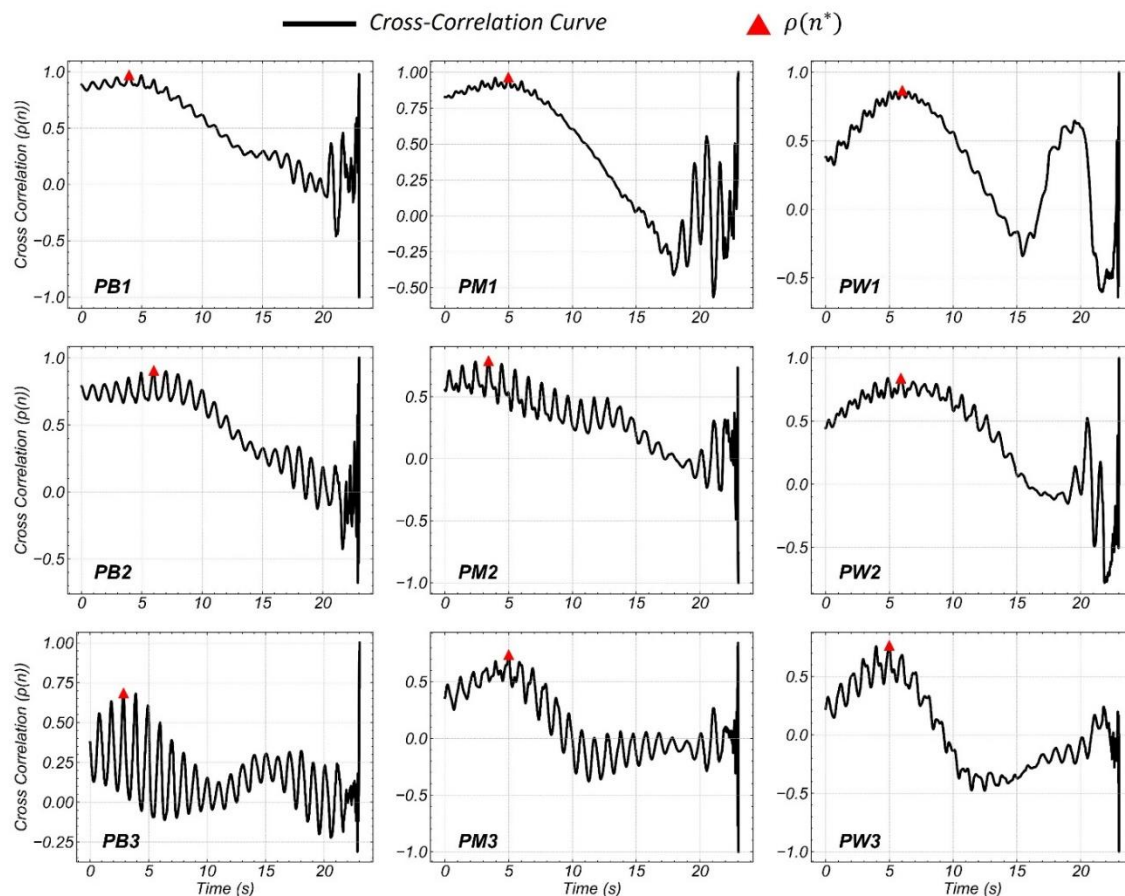


Figure 6. Cross-correlation result with time at all nine locations in Figure 4. The location of the $p(n^*)$ is marked with a red triangle in all figures.

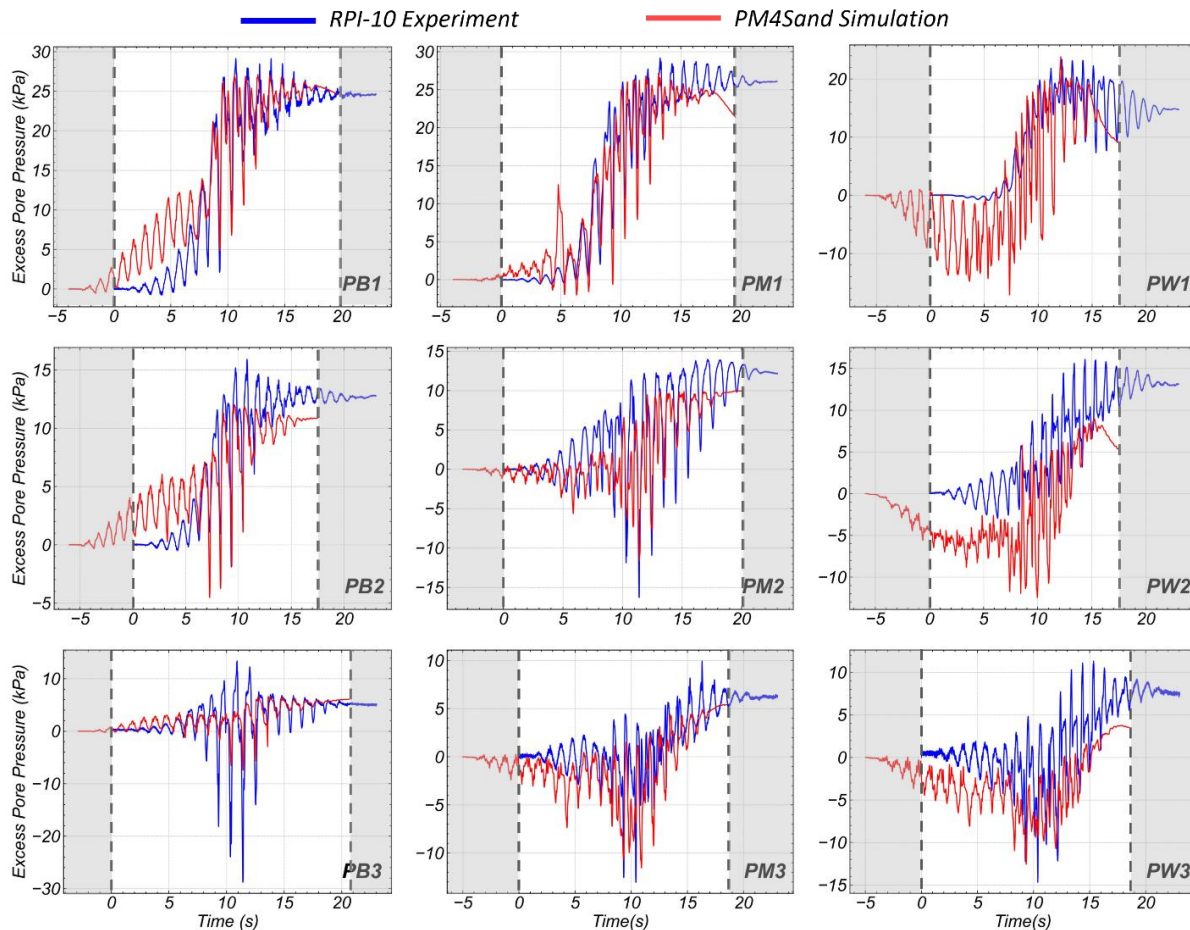


Figure 7. The simulation versus experiment results of pwp at all nine locations, after adjusting the simulation results with respect to the $p(n^*)$ in Figure 6 (non-gray area shows the overlapping time window between the two signals).

After the phase error adjustment, Russell's Error metric was applied to assess its effectiveness in capturing discrepancies during the overlapping time window of simulation and experiment (see Figure 8). The phase error captured by the metric is higher in locations with lower $\rho(n^*)$, which indicates greater remaining phase error after minimizing the lag. As the presence of phase error affects the calculated magnitude error, the captured magnitude errors are more accurate at locations with lower phase errors. In Figure 8, magnitude errors are evolving with a positive, a negative, or a zero slope. The magnitude error solely focuses on the relative difference in signal amplitudes between the experiment and simulation, irrespective of their absolute values. A positive slope indicates that the simulation produces higher amplitudes than the experiment, a negative slope suggests the opposite, and a zero slope signifies the same amplitude for both the experiment and the simulation. At all locations, the magnitude error starts with a positive slope, confirming the greater simulation amplitudes in the pre-liquefaction phase that was observed in Figure 7.

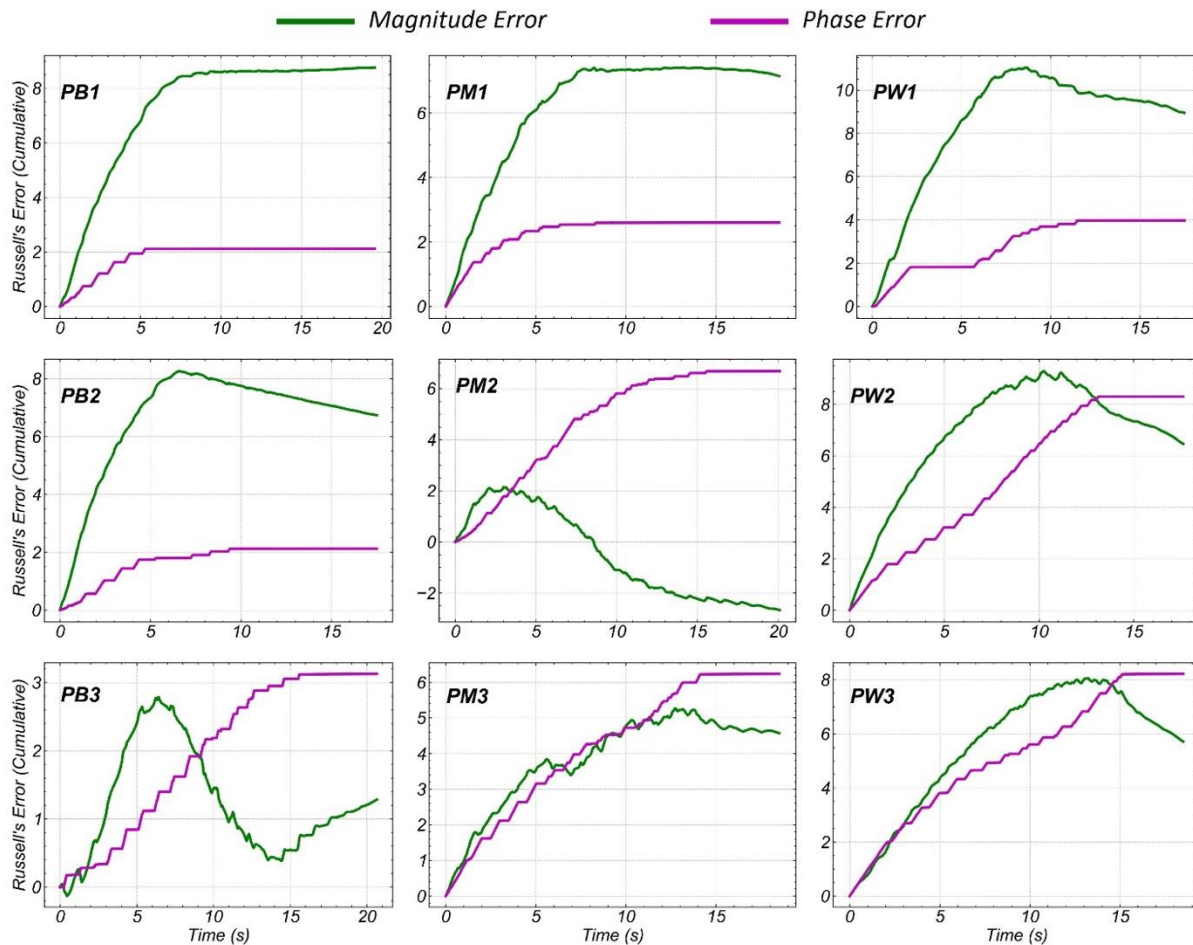


Figure 8. The captured cumulative magnitude and phase errors from Russell's Error metric after minimizing the phase lag on pwp experiment and simulation datasets at 9 locations.

CONCLUSIONS AND RECOMMENDATIONS

This study performed an initial investigation on three discrepancy metrics: Sprague and Geers (S&G), Russell's Error, and Cross Correlation. The aim was to (1) evaluate their potential for use in the validation of geosystems' simulations, and through that (2) provide a more unbiased approach to discussing the agreement between experimental and numerical data than the standard viewgraph norm of visual inspection. Through analysis of the metrics on three simple problems with known answers, it was observed that S&G exhibits asymmetrical behavior when calculating magnitude error and generating magnitude and phase errors on different scales. Additionally, both S&G and Russell's Error metrics are influenced by the presence of phase error in magnitude error calculations.

Russell's Error metric was selected for evaluating the agreement of LEAP-2020 porewater pressure recorded and simulated time histories. Cross-correlation was used to capture and minimize phase lag prior to calculating the magnitude error. The $\rho(n^*)$ obtained through cross-correlation addressed the existing lag between simulation and experiment, attributed to the

existing difference between the experimentally realized and numerically predicted cyclic strengths. Additionally, it explained variations in phase lag with changing measurement location. Cross-correlation effectively minimized phase lag in locations with higher agreement (higher $\rho(n^*)$). Russell's Error magnitude and phase errors explanation aligned with cross-correlation and visual comparison, reporting higher simulation amplitude in the pre-liquefaction phase and an improved match in the liquefaction triggering phase after phase lag minimization.

Validation metrics offer insights into the quality of agreement between numerical and experimental results. As such, they can form a solid and objective basis for the implementation of targeted improvements for either the simulations or the execution of experiments. When embarking upon such comparisons, it is important to understand that both the experiment and the simulation can be responsible for discrepancies in the results. Future research should focus on enhancing the performance of the presented metrics through modifications and evaluating additional metrics. These evaluations will be further improved by considering other responses of interest (beyond porewater pressure time histories), introducing numerical and experimental uncertainties, and exploring different geosystems.

REFERENCES

- Basu, D., Pretell, R., Montgomery, J., and Ziotopoulou, K. (2022). Investigation of key parameters and issues in simulating centrifuge model tests of a sheet-pile wall retaining a liquefiable soil deposit. *Soil Dynamics and Earthquake Engineering* 156, 107243.
- Boulanger, R. W. (2022). Nonlinear dynamic analyses involving soil liquefaction or cyclic softening, in: *20th ICSMGE-State of the Art and Invited Lectures*, Sydney, Australia.
- Boulanger, R. W., and Ziotopoulou, K. (2022). *PM4Sand (Version 3.2): A Sand Plasticity Model for Earthquake Engineering Applications*. Center for Geotechnical Modeling, Davis, California.
- Boulanger, R. W., and Ziotopoulou, K. (2018). On NDA Practices for Evaluating Liquefaction Effects, in: *Geotechnical Earthquake Engineering and Soil Dynamics V, Presented at the Geotechnical Earthquake Engineering and Soil Dynamics V*, American Society of Civil Engineers, Austin, Texas, pp. 1–20.
- Itasca. (2019). *Fast Lagrangian Analysis of Continua*.
- Liu, X., Chen, W., and Paas, M. (2005). Automated Occupant Model Evaluation and Correlation, in: *Engineering/Technology Management, Presented at the ASME 2005 International Mechanical Engineering Congress and Exposition, ASMEDC*, Orlando, Florida, USA, pp. 353–358.
- Perez, K., Reyes, A., and Taiebat, M. (2023). Roles of pre- and post-liquefaction stages in dynamic system response of liquefiable sand retained by a sheet-pile wall. *Soil Dynamics and Earthquake Engineering* 171, 107937.
- Russell, D. (1997). Error Measures for Comparing Transient Data: Part I : Development of a Comprehensive Error Measure Part II : Error Measures Case Study, in: *68th Shock and Vibration Symposium*, Hunt Valley, MD.
- Sarin, H., Kokkolaras, M., Hulbert, G., Papalambros, P., Barbat, S., and Yang, R.-J. (2010). Comparing Time Histories for Validation of Simulation Models: Error Measures and Metrics. *Journal of Dynamic Systems, Measurement, and Control* 132, 061401.
- Schwer, L. E. (2007). Validation metrics for response histories: perspectives and case studies. *Engineering with Computers* 23, 295–309.

- Sprague, M. A., and Geers, T. L. (2004). A spectral-element method for modelling cavitation in transient fluid–structure interaction. *Int. J. Numer. Meth. Engng.* 60, 2467–2499.
- Zhan, Z., Fu, Y., Yang, R.-J., and Peng, Y. (2011). An Enhanced Bayesian Based Model Validation Method for Dynamic Systems. *Journal of Mechanical Design* 133, 041005.
- Ziotopoulou, K. (2018). Seismic response of liquefiable sloping ground: Class A and C numerical predictions of centrifuge model responses. *Soil Dynamics and Earthquake Engineering* 113, 744–757.