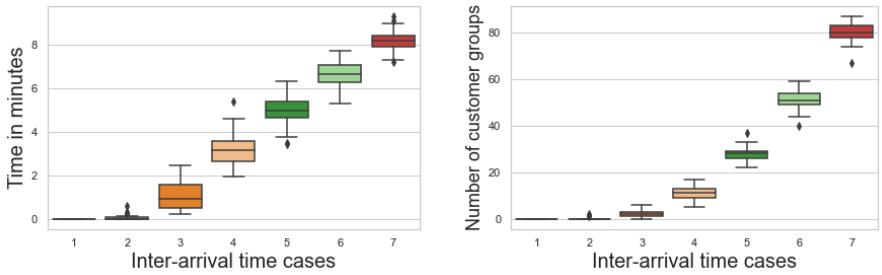
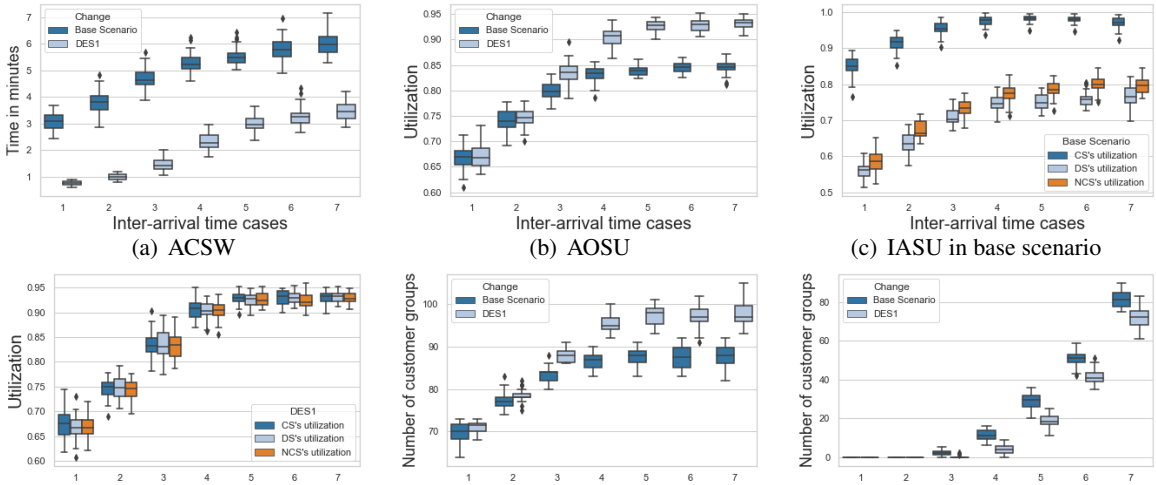


(a) ACQ (b) Number of balked customers

Figure 4: DES2 MR results show the expected increase in ACQ and balk count.



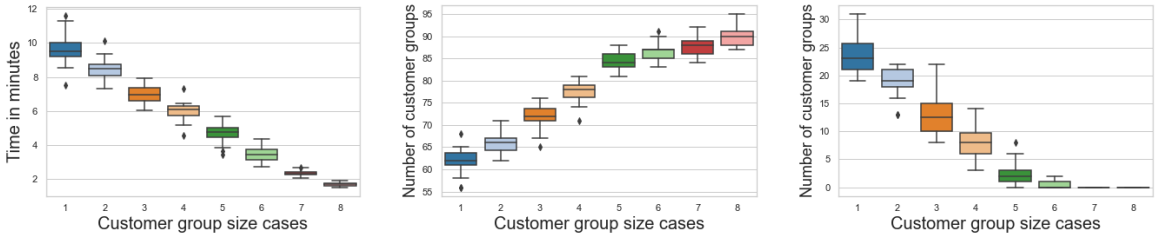
(a) ACSW (b) AOSU (c) IASU in base scenario

(d) IASU in DES1 (e) Throughput (f) Balked Customers

Figure 5: DES1 MR results on varying arrival rates show the expected changes in each output.

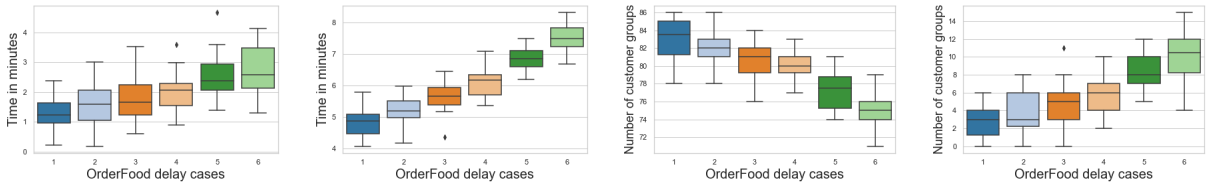
the output is reasonable; therefore, DES2 is upheld. Next, DES1 upholds the MR following a similar pattern to DES2 (Figure 5). Figure 6 shows the expected direction of change in ACSW, throughput, and balk count for DES3. DES4 alters the delay time of a single service block, and the output shows increasing/decreasing trends that correspond to the pseudo-oracles (Figure 7). Although the trend upward/downward is gradual, the delay times used for each case only differed by a minute, and thus it's reasonable. Initially the pseudo-oracle answers for DES5 were expected to exhibit no change for different delay times, but the outputs in Figure 8 show varied results, implying that either there is a bug or incorrect domain knowledge. Upon investigation, we found that combining two service blocks by adding their delay times reduces server utilization as servers spend less time moving between customers. This MR was further tested by combining other service blocks, yielding similar results as Figure 8, and was thus updated to the version in Table 1 as the issue was in domain expertise. DES6 and DES7 are tested (Figures 9 and 10), and upheld due to the expected change in output.

For ABM1, increasing the number of deciding servers shows the expected increase in overall HBC and expected drop in Individual HBC for all customer inter-arrival times. All other pseudo-oracle outputs for ABM1 also show supporting results (Figure 11). The ABM3, ABM4, and ABM5 results are in Figure 12. For ABM3, an increase in the number of tables shared by DS shows the expected increase in THT and drop in ACSW for CS's tables. For ABM4, an increase in wait time after the help time frame shows the expected drop in overall HBC & THT. For ABM5, a decrease in server walking speed shows the expected decrease in throughput and overall HBC, and expected increase in the number of balked customers and ACSW. ABM2



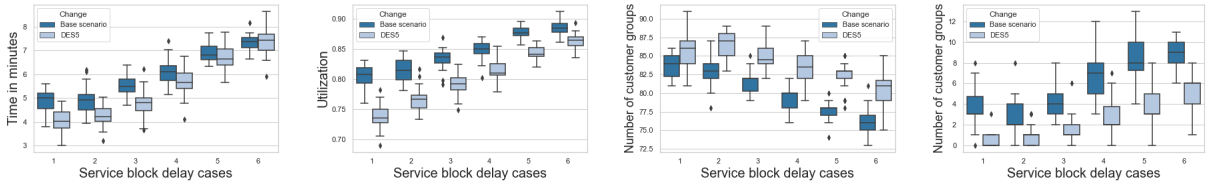
(a) ACSW (b) Throughput (c) Balked Customers

Figure 6: DES3 MR results for different customer group sizes.



(a) ACQ (b) ACSW (c) Throughput (d) Balked Customers

Figure 7: DES4 MR results for *orderFood* delay times.



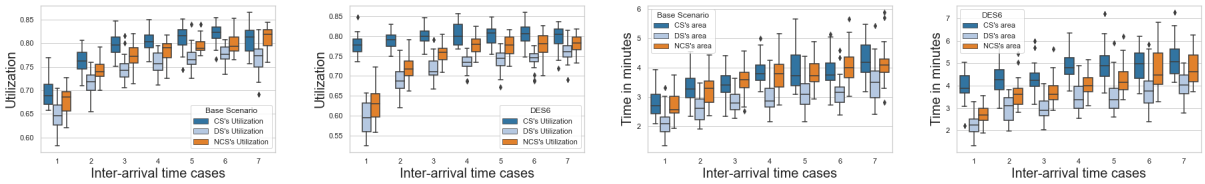
(a) ACSW (b) AOSU (c) Throughput (d) Balked Customers

Figure 8: DES5 MR results for each service block delay time.

MR results show that THT has the expected increase, but ACSW for CS's tables shows no significant change, contrary to the expected drop (Figure 13). In this experiment DS is the only helping server, only shares two of CS's tables, and only helps if CS has at least 5 tables occupied. Therefore, ACSW for CS's tables cannot be predicted only by altering the helping time frame with the default setup due to confounding factors. A different experimental setup is likely needed: DS helps with all of CS's tables, and helps with fewer tables occupied. The MR would need to be updated once the full situation is determined from additional experiments. This MR demonstrates a common step of the MT process for simulation validation.

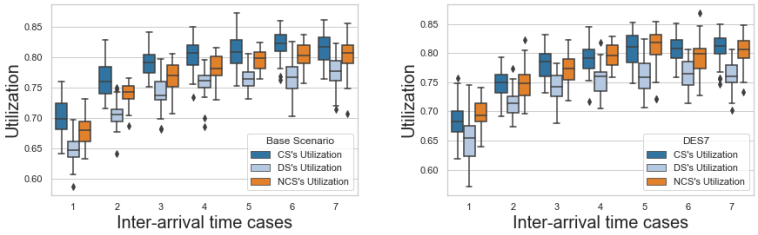
7 CONCLUSION

In this paper we presented an approach to apply metamorphic testing to validate hybrid simulation models, with a focus on ABM/DES hybrid models. We applied MT on a hybrid restaurant simulation model, where servers service their assigned customers, and under certain scenarios decide to help other servers. We systematically developed metamorphic relations following our presented framework presented, and used those MRs to run validation experiments. Our experiments applied MT for validating DES and ABM aspects of the hybrid model separately, as well as the combined overall model. By upholding of the predicted pseudo-oracle answers, we have shown an increased confidence in the validity of our hybrid model. In prior work, MT was demonstrated to be effective in validation of individual ABM or DES simulation models. Through this study, we demonstrated its effectiveness on hybrid model validation as well. In future work, we plan to utilize MT for both verification and validation of hybrid models. We also intend to apply our approach on other hybrid models using different combinations of modeling techniques and different domains.



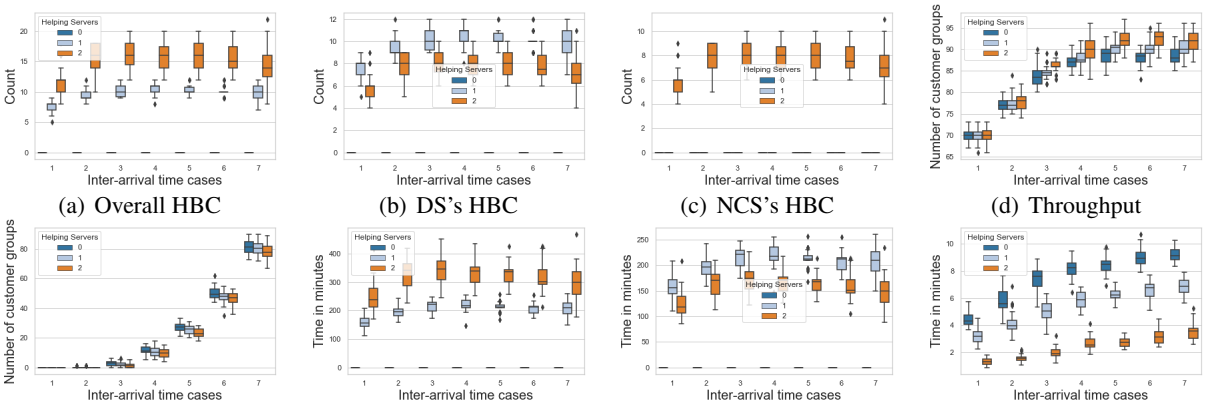
(a) AOSU in base scenario (b) AOSU in DES6 (c) ACSW in base scenario (d) ACSW in DES6

Figure 9: DES6 MR pseudo oracle outputs vs. different customer arrival rates.



(a) IASU in base scenario (b) IASU in DES7

Figure 10: DES7 MR results for each customer arrival rate.



(a) Overall HBC (b) DS's HBC (c) NCS's HBC (d) Throughput

(e) Balked Customers (f) THT (g) THT for DS (h) ACSW for CS's tables

Figure 11: ABM1 MR results for each customer inter-arrival time.

REFERENCES

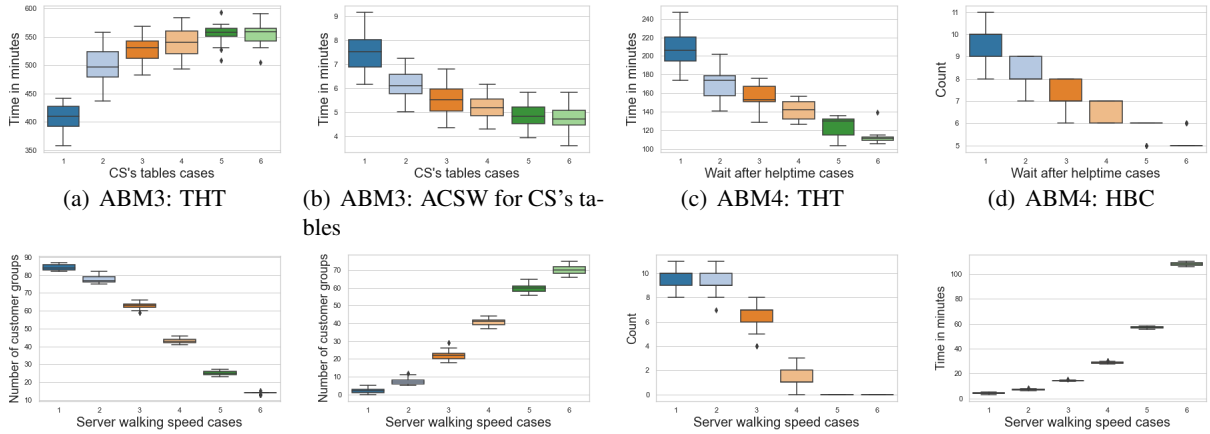
Ahlgren, J., M. E. Berezin, K. Bojarczuk, E. Dulskyte, I. Dvortsova, J. George, N. Gucevska, M. Harman, R. Lämmel, E. Meijer, S. Sapora, and J. Spahr-Summers. 2020. "WES: Agent-Based User Interaction Simulation on Real Infrastructure". In *Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops*, ICSEW'20, pp. 276–284. New York, NY, USA, ACM.

Balci, O. 2010. "Golden Rules of Verification, Validation, Testing, and Certification of Modeling and Simulation Applications". *SCS Modeling & Simulation Magazine* vol. 4.

Banks, J. 1998. *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. New York, John Wiley & Sons.

Barr, E. T., M. Harman, P. McMinn, M. Shahbaz, and S. Yoo. 2015, May. "The Oracle Problem in Software Testing: A Survey". *IEEE Transactions on Software Engineering* vol. 41 (5), pp. 507–525.

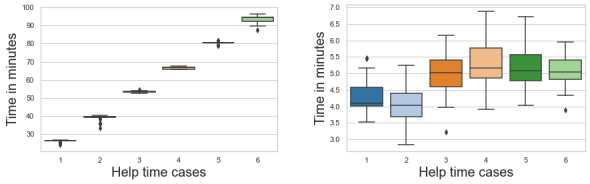
Brailsford, S. C., T. Eldabi, M. Kunc, N. Mustafee, and A. F. Osorio. 2019. "Hybrid Simulation Modelling in Operational Research: A State-of-the-Art Review". *European Journal of Operational Research* vol. 278 (3), pp. 721–737.



(a) ABM3: THT (b) ABM3: ACSW for CS's tables (c) ABM4: THT (d) ABM4: HBC

(e) ABM5: Throughput (f) ABM5: Balked Customers (g) ABM5: HBC (h) ABM5: ACSW

Figure 12: A6 ABM MR pseudo oracle outputs. All MRs are upheld.



(a) THT (b) ACSW for CS's tables

Figure 13: ABM2 MR results show the expected THT increase, and unexpected change in CS tables' ACSW.

Chen, T. Y. 2015. "Metamorphic Testing: A Simple Method for Alleviating the Test Oracle Problem". In *2015 IEEE/ACM 10th International Workshop on Automation of Software Test*.

Chen, T. Y., S. C. Cheung, and S. Yiu. 1998. "Metamorphic Testing: A New Approach for Generating Next Test Cases". Technical Report HKUST-CS98-01, Dept. of Computer Science, Hong Kong Univ. of Science and Technology.

Chen, T. Y., T. H. Tse, and Z. Q. Zhou. 2003. "Fault-Based Testing Without the Need of Oracles". *Information and Software Technology* vol. 45 (1), pp. 1–9.

De Vries, J., D. Roy, and R. De Koster. 2018. "Worth the Wait? How Restaurant Waiting Time Influences Customer Behavior and Revenue". *Journal of Operations Management* vol. 63, pp. 59–78.

Ding, J., T. Wu, D. Wu, J. Q. Lu, and X.-H. Hu. 2011. "Metamorphic Testing of a Monte Carlo Modeling Program". In *Proceedings of the 6th International Workshop on Automation of Software Test*, AST '11, pp. 1–7. New York, NY, USA, ACM.

Howden, W. E. 1978. "Theoretical and Empirical Studies of Program Testing". *IEEE Transactions on Software Engineering* vol. 4 (4), pp. 293–298.

Kanewala, U., and J. M. Bieman. 2013. "Techniques for Testing Scientific Programs Without an Oracle". In *Proceedings of the 5th International Workshop on Software Engineering for Computational Science and Engineering*, SE-CSE '13, pp. 48–57, Institute of Electrical and Electronics Engineers, Inc.

Lindvall, M., A. Porter, G. Magnusson, and C. Schulze. 2017. "Metamorphic Model-based Testing of Autonomous Systems". In *Proceedings of the IEEE/ACM 2nd International Workshop on Metamorphic Testing (MET '17), in conjunction with ICSE*, pp. 35–41.

Liu, H., F. C. Kuo, D. Towey, and T. Y. Chen. 2014, Jan. "How Effectively Does Metamorphic Testing Alleviate the Oracle Problem?". *IEEE J SE* vol. 40 (1), pp. 4–22.

Mouha, N., M. S. Raunak, D. R. Kuhn, and R. Kacker. 2018. "Finding Bugs in Cryptographic Hash Function Implementations". *IEEE Transactions on Reliability* vol. 67 (3), pp. 870–884.

Murphy, C. 2012. *Metamorphic Testing Techniques to Detect Defects in Applications without Test Oracles*. Ph. D. thesis, Columbia University.

Mustafee, N., and J. H. Powell. 2018. "From Hybrid Simulation to Hybrid Systems Modeling". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, pp. 1430–1439. Gothenburg, Sweden, Institute of Electrical and Electronics Engineers, Inc.

Olsen, M., and M. Raunak. 2016. "Metamorphic Validation for Agent-based Simulation Models". In *Proceedings of the 2016 Summer Computer Simulation Conference*, edited by J. L. R. Martin, A. D'Ambrogio, F. D. Rango, and G. Nicolescu, pp. 234–241. Montreal, Quebec, Curran Associates, Inc.

Olsen, M., and M. S. Raunak. 2019. "Increasing Validity of Simulation Models Through Metamorphic Testing". *IEEE Transactions on Reliability* vol. 68 (1), pp. 91–108.

Pugh, S., M. S. Raunak, D. R. Kuhn, and R. Kacker. 2019. "Systematic Testing of Post-Quantum Cryptographic Implementations Using Metamorphic Testing". In *2019 IEEE/ACM 4th International Workshop on Metamorphic Testing (MET)*, pp. 2–8.

Raunak, M. S., and M. Olsen. 2014. "A Survey of Validation in Health Care Simulation Studies". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, pp. 4089–4090. Savannah, Georgia, IEEE Press.

Raunak, M. S., and M. Olsen. 2015. "Simulation Validation Using Metamorphic Testing (WIP)". In *Proceedings of the 2015 Summer Computer Simulation conference*, edited by S. Mittal, E. Syriani, and I.-C. Moon, pp. 520–525. Chicago, Illinois, Curran Associates, Inc.

Raunak, M. S., and M. Olsen. 2021. "Metamorphic Testing on the Continuum of Verification and Validation of Simulation Models". In *2021 IEEE/ACM 6th International Workshop on Metamorphic Testing (MET)*.

Raunak, M. S., L. J. Osterweil, and A. Wise. 2011. "Developing Discrete Event Simulations from Rigorous Process Definitions". In *Proceedings of the 2011 Symposium on Theory of Modeling Simulation: DEVS Integrative MS Symposium*, pp. 117–124. San Diego, CA, USA, Society for Computer Simulation International.

Sargent, R. G. 2010. "Verifying and Validating Simulation Models". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, pp. 166–183. Baltimore, Maryland, Winter Simulation Conference.

Segura, S., G. Fraser, A. Sanchez, and A. Ruiz-Cortés. 2016, Sept.. "A Survey on Metamorphic Testing". *IEEE Transactions on Software Engineering* vol. 42 (9), pp. 805–824.

Sim, K. Y., W. K. S. Pao, and C. Lin. 2005. "Metamorphic Testing Using Geometric Interrogation Technique and Its Application". In *2nd International Conference on Electicatl Engineering/Electronics, Computer, Telecommunications, and Information Technology*, pp. 91–95.

Sokolowski, and Banks. 2010. *Modeling & Simulation Fundamentals*. Wiley.

Weyuker, E. J. 1982. "On Testing Non-Testable Programs". *The Computer Journal* vol. 25 (4), pp. 465–470.

Xie, X., J. Hob, C. Muirphy, G. Kaiser, B. Xue, and T. Chen. 2011. "Testing and Validating Machine Learning Classifiers by Metamorphic Testing". *Journal of Systems and Software* vol. 84 (4), pp. 544–558.