Proceedings Paper:

Jones, O.P.H., Oakley, J.E. and Purshouse, R.C. (2022) Simulation-based engineering design: solving parameter inference and multi-objective optimization problems on a shared simulation budget. In: Proceedings of 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 17-20 Oct 2021, Melbourne, Australia. IEEE, pp. 1399-1405. ISBN: 9781665442084. ISSN: 1062-922X. EISSN: 2577-1655.

https://doi.org/10.1109/smc52423.2021.9658645

Simulation-based engineering design: solving parameter inference and multi-objective optimization problems on a shared simulation budget*

Oliver P. H. Jones[1], Jeremy E. Oakley[2] and Robin C. Purshouse[1]

Abstract— **In recent years, the use of virtual engineering design processes has become more prevalent within industry. This increase has been facilitated by the availability of cost-effective computational machinery on which to run complex simulations of alternative candidate designs. Nevertheless it is frequently the case that, when working with complex problems, the number of simulation-based design evaluations available is limited. Within both industry and academia, it is usual for the stages of simulation model calibration and model-based optimization to be considered as separate consecutive steps rather than as a combined process. However, there is no guarantee that this approach makes the most efficient use of the available function evaluations. This work presents a new alternating methodology that aims to make more efficient use of the evaluation budget, through switching back and forth between the stages of calibration and optimization. To assess the effectiveness of the method, a new benchmark problem is introduced that contains both model parameters to be estimated and design variables to be selected. The new alternating method is found to possess improved calibration and comparable optimization performance in comparison to the sequential method on a budget of 5000 evaluations.**

I. Introduction

A. Motivation

Engineering design problems are becoming increasingly complex [1], with simulations playing a prominent role in the design process, for both the analysis and optimization of architectures, hardware components and software components [2], [3]. Simulations typically have many parameter assumptions which must be calibrated to data in order for the simulation to produce trustworthy results that can inform analysis and optimization [4]. These parameter inference processes for simulations require multiple evaluations of the simulation for different parameter settings, in the same way that optimization processes require multiple evaluations for different decision variables. Such evaluations can be very expensive in terms of computational resource for many high-fidelity simulations, requiring core hours and clock time of several hours or days. Sometimes in engineering design problems, a simulation may not have been fully calibrated prior to its use for optimization, which can lead to significant additional computational resources when issues are identified. Parameter inference and optimization problems share a need for computational resources to run simulations as part of a single overall design problem (see Fig. 1), but are rarely considered as a joint problem. In this paper, we explore the potential benefits of considering the problems in a unified way. Specifically, we address the question of whether inference should precede optimization, or whether improved multi-objective outcomes can be achieved by alternating between these activities.

B. Related works

The body of literature that simultaneously considers methods for simulation inference and methods for simulation-based optimization is very small. Numerical methods for inference [4] are similar to those used in multi-objective optimization [5], including the use of emulators or surrogate models to improve the efficiency of both processes [6], [7]. Indeed, multi-objective optimization methods have been explicitly used as the basis for simulation calibration [8], [9]. In the rarer cases when a published work considers both the calibration and optimization stages together, they are usually looking at specific case studies. For example, Gibbs and colleagues present a study of a water supply system design problem in which both stages are discussed, but treated separately [10]. Villarreal-Marroquín and colleagues also consider both stages in the context of injection moulding [11]; while a model is constructed and calibrated, the calibration is not revisited after the optimization is started. The problem is multi-objective, but the techniques being used for optimization are based on iterative grid search rather than evolutionary methods. In a study on rail diesel engine calibration and optimization, Qiang et al. consider a combined method of neural networks and adaptive network-based fussy inference system [12]. They found the approach effective, but did not frame the problem mathematically.

More generally, it is fruitful to consider methodologies used within both calibration and optimization that have the same structure or some linking element. The most obvious of these are cases in which surrogate models are used, as they should allow for direct information sharing. Kennedy & O'Hagan's calibration method uses a Gaussian process model as a surrogate [6]. This model could potentially be used a link to a surrogate-based optimization method such as ParEGO [7]. Another linking factor between these two methods is algorithm initialization. In both cases a form of Latin hypercube sampling was used. If this initial population were shared it could cut down on the number of evaluations used for initialization, releasing these for more beneficial uses. To facilitate these interactions, Jones et al. developed a

*This work was supported by Jaguar Land Rover and the EPSRC grant EPL025760/1 as part of the jointly funded Programme for Simulation Innovation. The first author acknowledges EPSRC studentship support.

[1]Oliver Jones and Robin Purshouse are with the Department of Automatic Control and Systems Engineering, University of Sheffield, UK r.purshouse@sheffield.ac.uk

[2]Jeremy Oakley is with the School of Mathematics and Statistics, University of Sheffield, UK j.oakley@sheffield.ac.uk

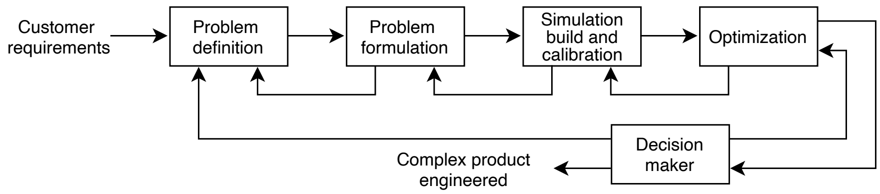


Fig. 1: The roles of simulation parameter inference and optimization in virtual engineering design workflows

unified mathematical framework for laying out the combined problem of model calibration and optimization [13], but did not fully realise methods for solving the combined problem.

C. Overview of the paper

The remainder of the paper is organised as follows. In Section II, the classical and new alternating approaches to the combined calibration–optimization problem are introduced, together with a description of the empirical approach used to compare the performance of the methods. A new benchmark problem for combined calibration–optimization processes—the first of its kind—is provided in Section III. Findings for both calibration and optimization performance are presented and discussed in Section IV. Conclusions and directions for future research are offered in Section V.

II. Methodology

This paper considers two methods for combining calibration and optimization, drawing on concepts from [13]. The first is the classical sequential approach of performing calibration followed by optimization; the second is an approach that alternates between phases of calibration and optimization.

A. Classical series approach

A high level schematic of the two processes of model calibration and optimization is displayed in Fig. 2. For the classical approach, the dashed connection is not present. Within model calibration there are three main components: *calibrator*, *model* and *expert*. The expert component represents an entity that knows how the system should perform for a given set of control inputs. In this study the expert is represented by a test function for which the predefined 'true' parameter set is used. The model takes both a set of control inputs and the current estimated values of the model parameters and returns an output. This work considers both cases where the structure of the model exactly matches that of the expert function and when there is a modelling error present. The modelling error is incorporated as an alteration in the test function. The calibrator represents the use of maximum likelihood estimation, in which Markov chain Monte Carlo (MCMC) is used to explore the parameter space [14]. The calibrator produces a set of potential parameters using an initial set of control inputs, along with corresponding true system outputs and the model output for a previous assumption of the parameters.

The second half of the schematic shows the optimization process. There are two main components: (1) the *model*, which performs the same function as the one present within model calibration; and (2) the *optimizer*. The optimizer used in this study is the MOEA/D multi-objective optimization algorithm [15], which determines new estimates for the optimal sets of inputs.

Once the evaluation budget allocated for calibration has been exhausted, the calibrator passes the selected parameter set to the optimization process. The parameter set is selected based on maximum likelihood. When the optimization stage has expended its evaluation budget, the results are passed to the decision maker for consideration.

B. Alternating approach

The second method is the alternating approach which considers moving back and forth between the stages of model calibration and optimization in order to make better use of the available function evaluation budget. The motivation for the alternating approach is that calibrated parameter sets that minimize the model prediction error for inputs in the region of the global optimum may be easier to obtain than parameter sets calibrated using non-optimal reference inputs or global consideration of the input space (e.g. through Latin hypercube sampling) — particularly in cases where the calibration search space is multi-modal.

Considering again the schematic in Fig. 2, the alternating method extends the structure of the classical approach. The difference between the two setups is the connection which can pass new control inputs back to the model calibration (the dotted arrow shown). This connection enables iteration between the stages of calibration and optimization. In the present study, the switching condition between the two stages is when a predefined portion of the available evaluation budget has been used.

C. Experimental setup

The calibration and optimization algorithms used for the alternating approach match those of the classical approach to allow for fair comparison. Both the classical and alternating methods use a total of 5000 function evaluations and 10 expert evaluations. The function evaluations are split, with 3000 of them being used for model calibration and 2000 being used for optimization. All the expert points are used at the start of the classical method while they are spread out over the run of the alternating approach. MOEA/D is

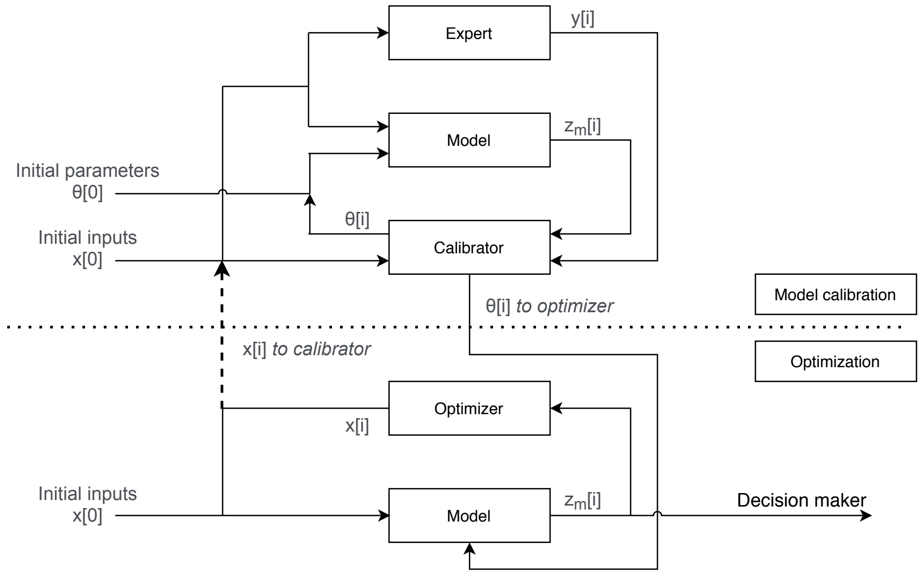


Fig. 2: Schematic of the classical and new alternating approach for tackling model calibration and optimization. The additional dotted connection is only present for the alternating approach

configured with 10 reference directions and a neighbourhood size of 3. Each approach is run for 21 replications to assess if the results are consistent. Additionally, when testing the classical and alternating methods, a predefined set of random seeds is used to help ensure any difference is due to the methods rather than starting conditions. The hypervolume indicator is used to measure the performance of the optimizer [16], using the worst case achievable in each objective as the reference point.

III. Benchmark Problem: DTLZ1$_\theta$

In order to assess the performance of the combined model calibration and optimization approaches it is necessary to develop suitable test problems that combine both activities. For the present study, an existing multi-objective benchmark problem is adapted for this purpose. The main structures of the objective functions are kept the same, while constants are converted into model parameters that now require identification. To reproduce the notion of an imperfect model of a real-world system, a second version of the problem is introduced in which the equations of the objective functions are altered to introduce structural model error.

Here, a modified version of a less rugged instance of the DTLZ1 benchmark problem is used [7], [17]—referred to as DTLZ1$_\theta$. A larger variety of combined calibration-optimization problems, along with a new component for the WFG framework of benchmark problems [18], can be found in [19].

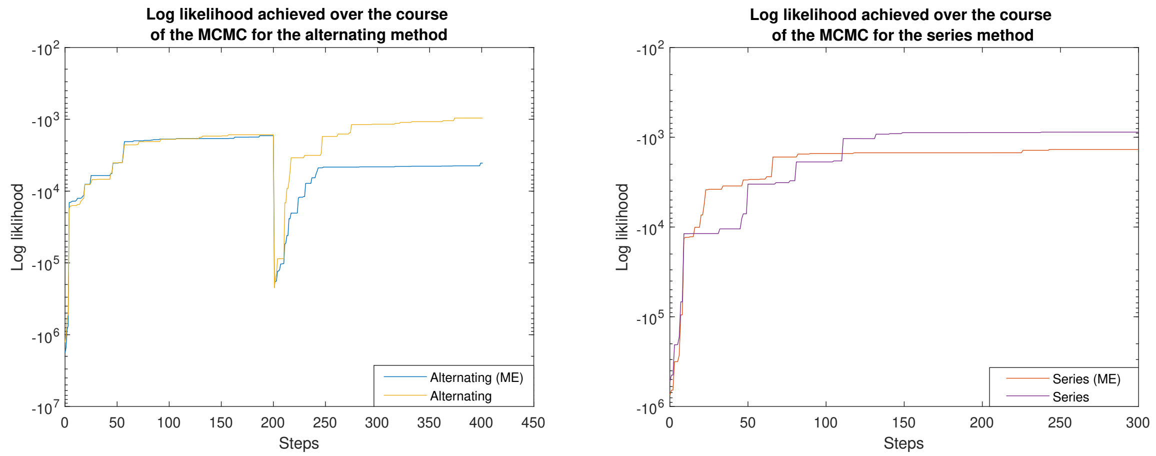
DTLZ1$_\theta$ has 6 control inputs and 5 parameters. When no error is present, it has an affine Pareto front between [0.5 0] and [0 0.5] and is given by the equations:

$$\begin{aligned}
&\text{Min } f_1 = \theta_1 x_1 (1+g)\\
&\text{Min } f_2 = \theta_2 (1-x_1)(1+g)\\
&\qquad g = 10\Big[\theta_3 + \sum_{i\in\{2,\dots,6\}} (x_i+\theta_4)^2 - cos(2\pi(x_i+\theta_5))\Big]\\
&\text{where } x_i \in [0,1], i \in \{1,\dots,n\}, n = 6.
\end{aligned} \tag{1}$$

When model error is included the second objective becomes:

$$\text{Min } f_2 = \theta_2(1-x_1^2)(1+g) \tag{2}$$

To obtain the Pareto front, the inputs should be set to $x_i = 0.5$ for all but x_1 which should be in the range of $[0,1]$. The parameters that correspond to the 'true' model are $\theta = (+0.5, +0.5, +5.0, -0.5, -0.5)$. The model error that has been incorporated causes the shape of the front to become concave. Each of the parameters impact the model in different ways: θ_1 and θ_2 work to scale the objectives, θ_3 represents an amount that is required to cancel out the effects of inputs $x_2 : x_n$. These first three parameters will not have any effect on the inputs required to achieve the 'true' Pareto front—they simply alter the output of the model. θ_4 has a varying impact which depends on the difference between the current input and the value of θ_4—when this difference is small it has a negligible impact on what constitutes a good input. θ_5 has the largest impact on which inputs will produce the 'true' Pareto front. When θ_4 is a reasonable approximation of its true value, the inputs ($x_2 : x_n$) that will



(a) Alternating

(b) Series

Fig. 3: The progress of the log-likelihood obtained by the MCMC over the course of the calibration. The alternating method, (a), and classical (series) method, (b), are depicted for both when modelling error (ME) is present and absent

realise an 'apparent' Pareto front are $x_i = -\theta_5$.

IV. Results

A. Quality of calibration

The first aspect to examine is the progression of the calibrator over the course of a representative run. Fig. 3 shows the log-likelihood achieved at each step of MCMC. The four cases depicted here are when the alternating and series (classical) approach were run for both the cases when modelling error was present and absent. When considering these results, it is important to recognise that the expert points used to calculate the likelihood differ between the two methods, so performing a direct comparison of the maximum values obtained can be misleading.

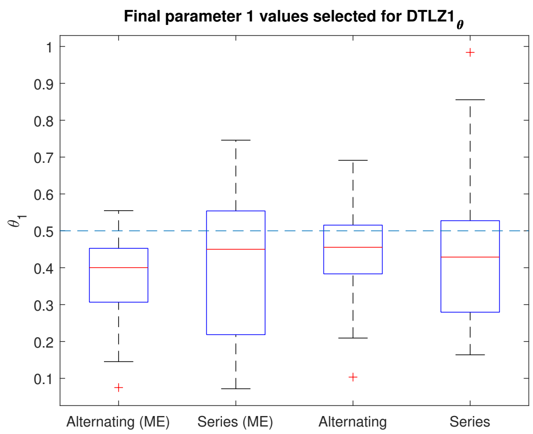
During the series runs, fast initial improvement is seen as better parameter sets are discovered. By around 150 steps, however, the rate of improvement is greatly reduced with little change for when model error is either present or absent. Considering the alternating method, a similar rate of improvement is observed during the first calibration stage. After the 200th step—when the new expert points are gained by the alternating method—a drop in the log-likelihood can be seen, due to the algorithm gaining a better knowledge of the system. Once the alternating algorithm commences its second calibration stage, rapid improvement is again observed. The large difference between the final obtained log-likelihood for the alternating method when model error is present and absent is likely due to the lack of a corrective term in the calibrator to estimate and remove the modelling error.

The final obtained values for the five parameters can be seen in Fig. 4, which shows the distribution of results achieved across all 21 replications. The first thing to notice is that the calibration runs better when no model error is present. The reason why both methods struggle to identify the correct parameter values when model error is present is that the model error changes the parameters required for an evaluated point to match the expert value for a given set of inputs. This results in values close to the true parameters producing a worse likelihood than those in a different location and so being rejected by the MCMC.

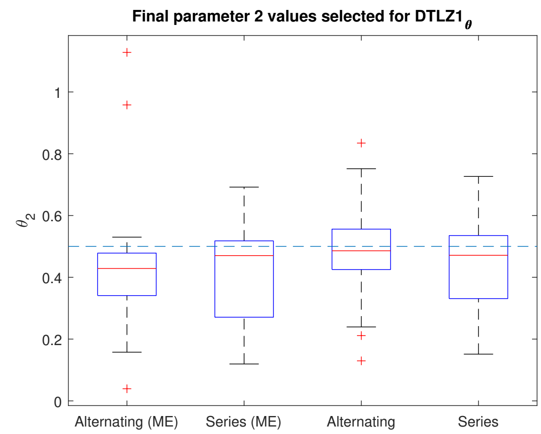
The parameters each have a different impact on the Pareto front, with some having greater importance for optimization than others. Parameters θ_1 and θ_2 both work to scale the objective values. While incorrect selection of these parameters may affect the convergence speed of the optimizer, it is still possible to obtain the optimal values for the decision variables. Parameters θ_3, θ_4 and θ_5 have a more direct interaction with the decision variables being selected. In the case of θ_3 and θ_4, it is possible to find combinations that can make it look like the problem is correctly calibrated when it is not. It is therefore important to ensure these latter parameters are well calibrated. For these two parameters, the alternating method achieves a better calibration than the series method. It should also be noted that for all five of the parameters, the final distributions achieved by the alternating method are tighter than those seen for the series method, indicative of more robust performance.

B. Quality of optimization

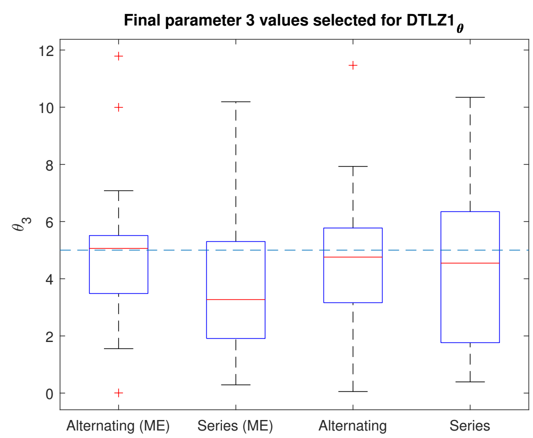
Optimization results for the hypervolume indicator are shown in Fig. 5. The *apparent* hypervolume examined within these results is obtained by using the maximum likelihood estimates for the parameters, rather than their true values. In Fig. 5a there are many cases in which the achieved hypervolume exceeds 100% of its maximum value. These scenarios can occur due to incorrect calibration of the model. The alternating approach appears to perform more consistently well when compared to the series method, both when model error is present as well as when it is not. It should be noted that there are still outlying cases in which the alternating method performs badly. The presence of model error seems to have had a larger impact on the



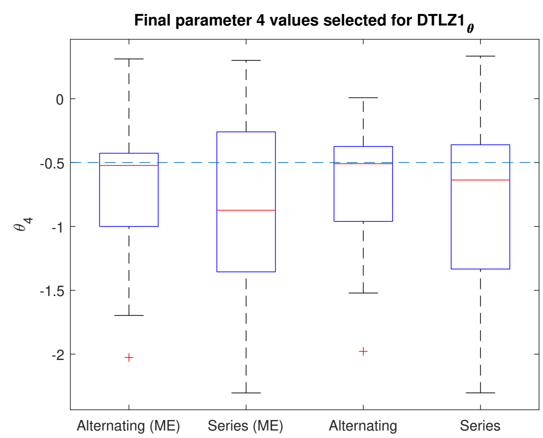
(a) Parameter θ_1



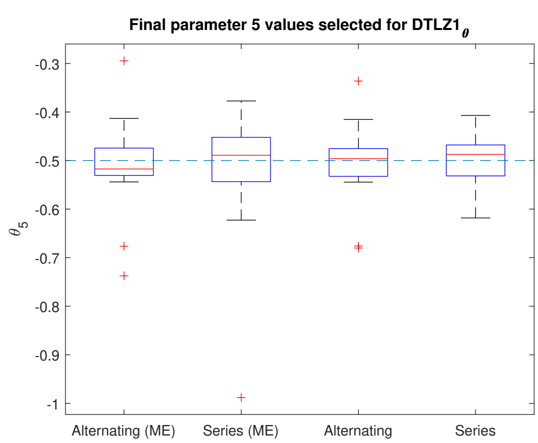
(b) Parameter θ_2



(c) Parameter θ_3



(d) Parameter θ_4

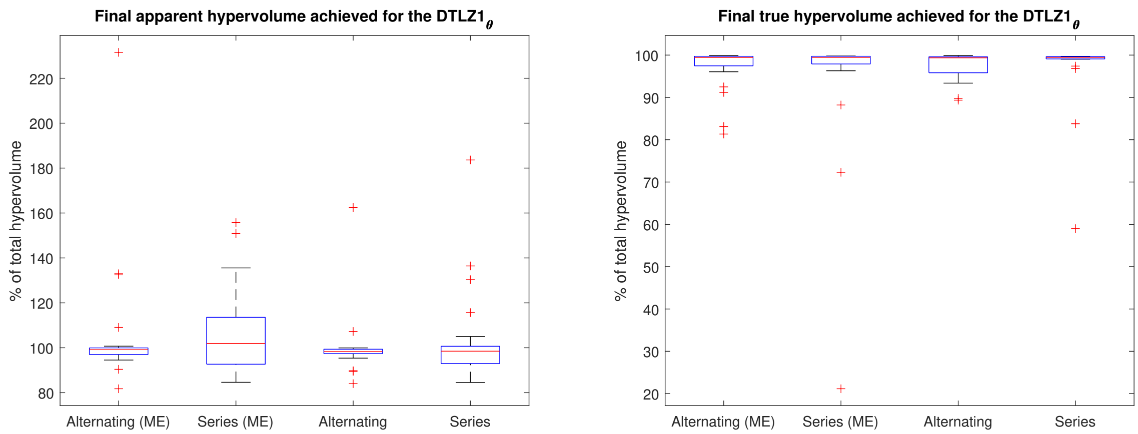


(e) Parameter θ_5

Fig. 4: Box plots for final parameter values achieved by alternating and series methods for the cases when modelling error (ME) is present and absent when calibrating DTLZ1$_\theta$. True parameter values are shown by the dashed horizontal line.

series method causing the results to be much more spread out. The alternating method tends not to overestimate the value of the hypervolume with the 75th percentile of runs coming in at under 100%, indicating it is likely to be a better representation of the true performance. In comparison the classical method regularly overestimates the hypervolume, especially when there is model error present.

The results presented in Fig. 5b show the same points as



(a) Apparent hypervolume

(b) True hypervolume

Fig. 5: Box plot comparisons of (a) apparent and (b) true hypervolume achieved by alternating and series methods, for cases when modelling error (ME) is present and absent. The vertical axis is the percentage of the optimal hypervolume achieved.

were seen for the apparent hypervolume in Fig. 5a but they are now being evaluated on the true system. The performance of the alternating and classical methods is very similar when modelling error is present. When no modelling error is present, the series method has the tighter distribution—although its performance is not statistically significantly better than the alternating method (p=0.80). These findings suggest that, even if the determined parameters are suboptimal, it is still possible for the optimizer to identify good values for the decision variables.

Comparing the absolute difference between the apparent and true hypervolume achieved is useful as it gives an indication of how much the user would be able to trust the outcomes from the simulation-based design process. It is important to remember that the decision maker would realistically not possess the true system outputs unless some of the expert evaluations have been set aside for this purpose, which would reduce the amount available for calibration and optimization. The absolute difference observed is much smaller for the alternating methods (Fig. 6a) due to the better calibration. For the series method, when model error is present, the majority of results presented by the solver would be misleading until evaluated on the true system.

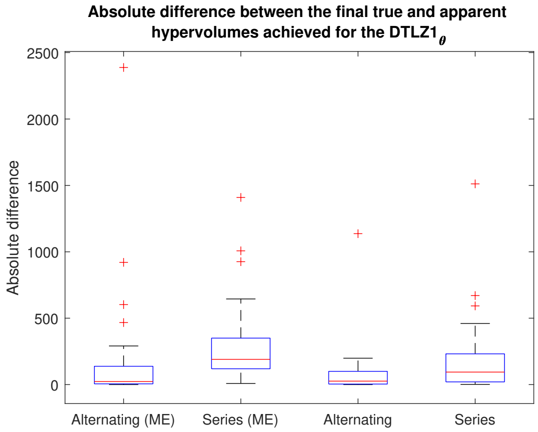
To understand how these results translate to the objective space, the final output populations achieved by the alternating and series methods when model error was present are shown in Fig. 6b. The results relate to the median hypervolume outcome across the alternating runs, together with the series run possessing the same seed. This means that, apart from the expert population, both setups had the same initial information. The true output populations of the methods can be seen to perform equivalently. Considering the apparent outputs achieved, the series method outputs are projected further from the front than those of the alternating method. It is interesting that the series method overestimates the front location while the alternating method underestimates it.

V. Conclusion

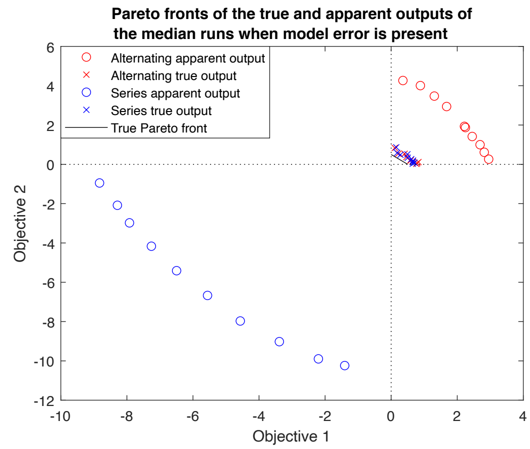
This paper has presented a new alternating methodology employing a budget of 5000 functions evaluations and compares it to the classic approach of performing calibration followed by subsequent optimization. The new approach employs alternating stages of model calibration and optimization, aiming to more efficiently use the available function evaluations. Additionally, the new approach evaluates interim points on the true system between alternations, rather than fully expending them in the initial calibration. The paper also contributes a new benchmark problem that explicitly features calibration parameters and model error—these are key features of real-world simulation-based problems that have been missing from the benchmarking community to date. However, it is questionable the extent to which the underlying benchmark problem, DTLZ1, that was adapted for this purpose is itself representative of real-world problems [20]. This is a general issue in the optimization benchmarking community, but it is hoped that more realistic benchmarks—including features such as model error—will become more prevalent as benchmarking methodology begins to better address real-world problem features [21]. The generalisability of findings from the present study are clearly limited in that the methods have not been demonstrated on a real-world simulation-based engineering design case study.

Accepting this limitation, the performance of the proposed alternating method has been evaluated within a robust test framework based on both final parameters obtained and the final hypervolume achieved by each run of the methods, both when model error is present and absent. The performance of the alternating approach is found to offer benefits over the series approach, particularly in terms of the reliability of the objective space estimates obtained in relation to their true values—which is crucially important for decision making.

Looking towards future research, there are two priority areas for investigation. Firstly, developing and assessing



(a) Absolute differences



(b) Pareto fronts

Fig. 6: (a) Comparison of absolute differences between apparent and true hypervolumes achieved by alternating and series methods, for cases when modelling error (ME) is present and absent. (b) True and apparent final optimizer populations for the alternating and series methods when modelling error is present. The populations represent the median-performing alternating run along with its corresponding series run.

different criteria for determining when to switch between calibration and optimization stages could aid in improving efficiency. Secondly, further testing across a larger variety of problems—including benchmark engineering design problems—will aid in achieving a better understanding of how the new approach operates. Once the new approach has been better explored on benchmark problems, it will be vital to assess its performance on a real-world complex engineering design problem.

References

[1] J. Luo and K. L. Wood, "The growing complexity in invention process," *Research in Engineering Design*, vol. 28, no. 4, pp. 421–435, 2017.

[2] R. Sinha, C. J. Paredis, V.-C. Liang, and P. K. Khosla, "Modeling and simulation methods for design of engineering systems," *J. Comput. Inf. Sci. Eng.*, vol. 1, no. 1, pp. 84–91, 2001.

[3] M. Karlberg, M. Löfstrand, S. Sandberg, and M. Lundin, "State of the art in simulation-driven design," *International Journal of Product Development*, vol. 18, no. 1, pp. 68–87, 2013.

[4] K. Cranmer, J. Brehmer, and G. Louppe, "The frontier of simulation-based inference," *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30 055–30 062, 2020.

[5] A. Díaz-Manríquez, G. Toscano, J. H. Barron-Zambrano, and E. Tello-Leal, "A review of surrogate assisted multiobjective evolutionary algorithms," *Computational intelligence and neuroscience*, vol. 2016, 2016.

[6] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001.

[7] J. Knowles, "ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 1, pp. 50–66, Feb. 2006.

[8] Z. Zhang, T. Wagener, P. Reed, and R. Bhushan, "Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization: predictions in ungauged basins," *Water Resources Research*, vol. 44, no. 12, Dec. 2008.

[9] X. Li, D. E. Weller, and T. E. Jordan, "Watershed model calibration using multi-objective optimization and multi-site averaging," *Journal of Hydrology*, vol. 380, no. 3-4, pp. 277–288, Jan. 2010.

[10] M. S. Gibbs, G. C. Dandy, and H. R. Maier, "Calibration and optimization of the pumping and disinfection of a real water supply system," *Journal of Water Resources Planning and Management*, vol. 136, no. 4, pp. 493–501, Jul. 2010.

[11] M. G. Villarreal-Marroquín, P.-H. Chen, R. Mulyana, T. J. Santner, A. M. Dean, and J. M. Castro, "Multiobjective optimization of injection molding using a calibrated predictor based on physical and simulated data," *Polymer Engineering & Science*, vol. 57, no. 3, pp. 248–257, Mar. 2017.

[12] H. Qiang, Y. Fuyuan, Z. Ming, and O. Minggao, "Study on modeling method for common rail diesel engine calibration and optimization," SAE Technical Paper, Tech. Rep., 2004.

[13] O. P. Jones, J. E. Oakley, and R. C. Purshouse, "Toward a unified framework for model calibration and optimisation in virtual engineering workflows," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, 2019, pp. 3148–3153.

[14] J. S. Liu, *Monte Carlo strategies in scientific computing.* New York, NY: Springer, 2001.

[15] Qingfu Zhang and Hui Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, Dec. 2007.

[16] C. M. Fonseca, L. Paquete, and M. López-Ibánez, "An improved dimension-sweep algorithm for the hypervolume indicator," in *2006 IEEE international conference on evolutionary computation*. IEEE, 2006, pp. 1157–1163.

[17] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable test problems for evolutionary multiobjective optimization," *Evolutionary Multiobjective Optimization. Theoretical Advances and Applications*, pp. 105–145, 2005.

[18] S. Huband, P. Hingston, L. Barone, and L. While, "A review of multiobjective test problems and a scalable test problem toolkit," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 5, pp. 477–506, 2006.

[19] O. P. Jones, "A framework for combining model calibration with model-based optimization in virtual engineering design workflows," Ph.D. dissertation, The University of Sheffield, 2021.

[20] H. Ishibuchi, L. He, and K. Shang, "Regular Pareto front shape is not realistic," in *2019 IEEE Congress on Evolutionary Computation (CEC)*, 2019, pp. 2034–2041.

[21] K. van der Blom, T. M. Deist, T. Tušar, M. Marchi, Y. Nojima, A. Oyama, V. Volz, and B. Naujoks, "Towards realistic optimization benchmarks: a questionnaire on the properties of real-world problems," in *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*, 2020, pp. 293–294.