



Technical Paper

Optimization of solidification in die casting using numerical simulations and machine learning

Shantanu Shahane*, Narayana Aluru, Placid Ferreira, Shiv G. Kapoor, Surya Pratap Vanka

Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States

ARTICLE INFO

Keywords:

Die casting

Deep neural networks

Multi-objective optimization

ABSTRACT

In this paper, we demonstrate the combination of machine learning and three dimensional numerical simulations for multi-objective optimization of low pressure die casting. The cooling of molten metal inside the mold is achieved typically by passing water through the cooling lines in the die. Depending on the cooling line location, coolant flow rate and die geometry, nonuniform temperatures are imposed on the molten metal at the mold wall. This boundary condition along with the initial molten metal temperature affect the product quality quantified in terms of micro-structure parameters and yield strength. A finite volume based numerical solver is used to determine the temperature-time history and correlate the inputs to outputs. The objective of this research is to develop and demonstrate a procedure to obtain the initial and wall temperatures so as to optimize the product quality. The non-dominated sorting genetic algorithm (NSGA-II) is used for multi-objective optimization in this work. The number of function evaluations required for NSGA-II can be of the order of millions and hence, the finite volume solver cannot be used directly for optimization. Therefore, a multilayer perceptron feed-forward neural network is first trained using the results from the numerical solution of the fluid flow and energy equations and is subsequently used as a surrogate model. As an assessment, simplified versions of the actual problem are designed to first verify results of the genetic algorithm. An innovative local sensitivity based approach is then used to rank the final Pareto optimal solutions and select a single best design.

1. Introduction

Die casting is one of the popular manufacturing processes in which liquid metal is injected into a permanent metal mold and solidified. Generally, die casting is used for parts made of aluminum and magnesium alloys with steel molds. Automotive and housing industrial sectors are common consumers of die casting. In such a complex process, there are several input parameters which affect the final product quality and process efficiency. With advances in computing hardware and software, the physics of these processes can be modeled using numerical simulation techniques. Detailed flow and temperature histories, micro-structure parameters, mechanical strength etc. can be estimated from these simulations. In today's competitive industrial world, estimating the values of input parameters for which the product quality is optimized has become highly important. There has been extensive research in numerical optimization algorithms which can be coupled with detailed numerical simulations in order to handle complex optimization problems.

Solidification in casting process has been studied by many researchers. Minaie et al. [1] have analyzed metal flow during die filling

and solidification in a two dimensional rectangular cavity. The flow pattern during the filling stage is predicted using the volume of fluid (VOF) method and enthalpy equation is used to model the phase change with convection and diffusion inside the cavity. They have studied the effect of gate location on the residual flow field after filling and the solid liquid interface during solidification. Im et al. [2] have done a combined filling and solidification analysis in a square cavity using the implicit filling algorithm with the modified VOF together with the enthalpy formulation. They studied the effect of assisting flow and opposite flow due to different gate positions on the residual flow. They found that the liquid metal solidifies faster in the opposite flow than in the assisting flow situation. Cleary et al. [3] used the smoothed particle hydrodynamics (SPH) to simulate flow and solidification in three dimensional practical geometries. They demonstrated the approach of short shots to fill and solidify the cavity partially with insufficient metal so that validation can be performed on partial castings. Plotkowski et al. [4] simulated the phase change with fluid flow in a rectangular cavity both analytically and numerically. They simplified the governing equations of the mixture model by scaling analysis followed by an analytical solution and then compared with a complete finite volume

* Corresponding author.

E-mail address: sshahan2@illinois.edu (S. Shahane).

<https://doi.org/10.1016/j.jmapro.2020.01.016>

Received 5 April 2019; Received in revised form 2 December 2019; Accepted 9 January 2020

Available online 28 January 2020

1526-6125/ © 2020 Published by Elsevier Ltd on behalf of The Society of Manufacturing Engineers.

solution.

Recently, there has been a growing interest in the numerical optimization of various engineering systems. Poloni et al. [5] applied neural network with multi-objective genetic algorithm and gradient based optimizer to the design of a sailing yacht fin. The geometry of the fin was parameterized using Bezier polynomials. The lift and drag on the fin was optimized as a function of the Bezier parameters and thus, an optimal fin geometry was designed. Elsayed and Lacor [6] performed a multi-objective optimization of a gas cyclone which is a device used as a gas–solid separator. They trained a radial basis function neural network (RBFNN) to correlate the geometric parameters like diameters and heights of the cyclone funnel to the performance efficiency and the pressure drop using the data from numerical simulations. They further used the non-dominated sorting genetic algorithm (NSGA-II) to obtain the Pareto front of the cyclone designs. Wang et al. [7] optimized the groove profile to improve hydrodynamic lubrication performance in order to reduce the coefficient of friction and temperature rise of the specimen. They coupled the genetic algorithm (GA) with the sequential quadratic programming (SQP) algorithm such that the GA solutions were provided as initial points to the SQP. Stavrakakis et al. [8] solved for window sizes for optimal thermal comfort and indoor air quality in naturally ventilated buildings. A computational fluid dynamics model was used to simulate the air flow in and around the buildings and generate data for training and testing of a RBFNN which is further used for constrained optimization using the SQP algorithm. Wei and Joshi [9] modeled the thermal resistance of a micro-channel heat exchanger for electronic cooling using a simplified thermal resistance network model. They used a genetic algorithm to obtain optimal geometry of the heat exchanger so as to minimize the thermal resistance subject to constraints of maximum pressure drop and volumetric flow rate. Husain and Kim [10] optimized the thermal resistance and pumping power of a micro-channel heat sink as a function of geometric parameters of the channel. They used a three dimensional finite volume solver to solve the fluid flow equations and generate training data for surrogate models. They used multiple surrogate models like response surface approximations, Kriging and RBFNN. They provided the solutions obtained from the NSGA-II algorithm to SQP as initial guesses. Lohan et al. [11] performed a topology optimization to maximize the heat transfer through a heat sink with dendritic geometry. They used a space colonization algorithm to generate topological patterns with a genetic algorithm for optimization. Amanifard et al. [12] solved an optimization problem to minimize the pressure drop and maximize the Nusselt number with respect to the geometric parameters and Reynolds number for micro-channels. They used a group method of data handling type neural network as a surrogate model with the NSGA-II algorithm for optimization. Esparza et al. [13] optimized the design of a gating system used for gravity filling a casting so as to minimize the gate velocity. They used a commercial program (FLOW3D) to estimate the gate velocity as a function of runner depth and tail slope and the SQP method for optimization. Patel et al. [14] modeled and optimized the wear behavior of squeeze cast products. Three different optimization methods (genetic algorithm, particle swarm optimization and desirability function approach) are combined with neural network as a surrogate model. The training data for neural network is obtained from experiments and nonlinear regression models.

In this paper, we consider the heat transfer and solidification processes in die casting of a complex model geometry. The computer program [15] solves the fluid flow and energy equations, coupled with the solid fraction–temperature relation, using a finite volume numerical method. When not significant, natural convection flow is neglected and only the energy equation is solved. The product quality is assessed using grain size and yield strength which are estimated using empirical relations. The solidification time is used to quantify the process efficiency. The molten metal and mold wall temperatures are crucial in determining the quality of die casting. The wall temperature is typically nonuniform due to the complex mold geometries and asymmetric

placement of cooling lines. This nonuniformity can be modeled by domain decomposition of the wall and assigning single temperature value to each domain. Neural networks are trained using the data generated from the simulations to correlate the initial and wall temperatures to the output parameters like solidification time, grain size and yield strength. The optimization problem formulated with these three objectives is then solved using genetic algorithm. The procedure illustrated here can be applied to any practical mold geometry with a complex distribution of wall temperatures.

2. Numerical model description

The numerical model incorporates the effects of solidification and heat transfer in die casting. Since the common die casting geometries have thin cross-sections, the solidification time is of the order of seconds and hence, the effect of natural convection has been found to be negligible. Thus, the momentum equations of the liquid metal are not solved in this work. The energy equation which can be written in terms of temperature has unsteady, diffusion and latent heat terms:

$$\rho C_p \frac{\partial T}{\partial t} = \nabla \cdot (k \nabla T) + \rho L_f \frac{\partial f_s}{\partial t} \quad (1)$$

where T is temperature, ρ is density, C_p is specific heat, k is thermal conductivity, L_f is latent heat of fusion, f_s is solid fraction and t is time. The Gulliver–Scheil equation (2) [16] relates solid fraction to temperature for a binary alloy:

$$f_s(T) = \begin{cases} 0 & \text{if } T > T_{liq} \\ 1 & \text{if } T < T_{sol} \\ 1 - \left(\frac{T - T_f}{T_{liq} - T_f} \right)^{1/(k_p - 1)} & \text{otherwise} \end{cases} \quad (2)$$

where k_p is partition coefficient, T_f is freezing temperature, T_{sol} is solidus temperature and T_{liq} is liquidus temperature.

Secondary dendrite arm spacing (SDAS) is a microstructure parameter which can be used to estimate the 0.2% yield strength. The cooling rate at each point in the domain is computed by numerically solving the energy equation and solid fraction temperature relation (Eqs. (1) and (2)). The following empirical relations link the cooling rate to SDAS and yield strength:

$$SDAS = \lambda_2 = A_\lambda \left(\frac{\partial T}{\partial t} \right)^{B_\lambda} \quad [\text{in } \mu\text{m}] \quad (3)$$

where $A_\lambda = 44.6$ and $B_\lambda = -0.359$ are based on the model for microstructure in aluminum alloys [17]:

$$\sigma_{0.2} = A_\sigma \lambda_2^{-1/2} + B_\sigma \quad (4)$$

where $\sigma_{0.2}$ is in MPa, λ_2 (SDAS) is in μm , $A_\sigma = 59.0$ and $B_\sigma = 120.3$ [18].

Grain size estimation is based on the work of Greer et al. [19]. The grain growth rate is given by:

$$\frac{dr}{dt} = \frac{\lambda_s^2 D_s}{2r} \quad (5)$$

where r is the grain size, D_s is the solute diffusion coefficient in the liquid and t is the time. The parameter λ_s is obtained using invariant size approximation:

$$\lambda_s = \frac{-S}{2\pi^{0.5}} + \left(\frac{S^2}{4\pi} - S \right)^{0.5} \quad (6)$$

S is given by

$$S = \frac{2(C_s - C_0)}{C_s - C_l} \quad (7)$$

where $C_l = C_0(1 - f_s)^{(k_p - 1)}$ is solute content in the liquid, $C_s = k_p C_l$ is solute content in the solid at the solid–liquid interface and C_0 is the

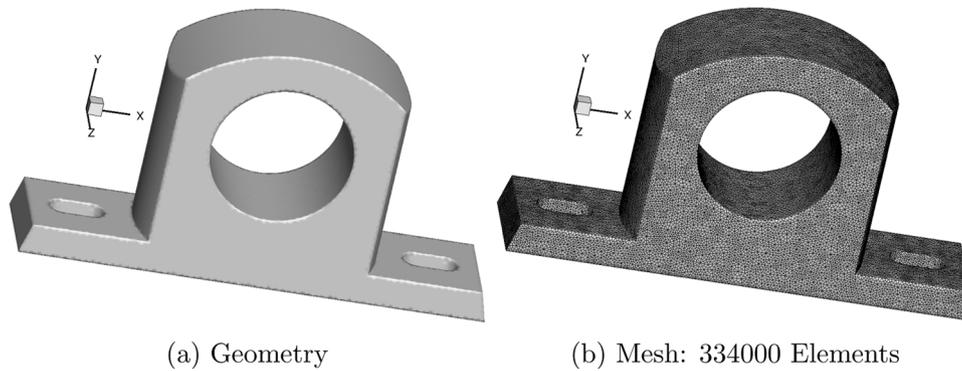


Fig. 1. Clamp: 16.5 cm × 9 cm × 3.7 cm [15].

nominal solute concentration. Hence, from the partition coefficient (k_p) and estimated solid fraction (f_s), Eqs. (5)–(6) are solved to get the final grain size.

Eqs. (1)–(5) are solved numerically using the software OpenCast [15] with a finite volume method on a collocated grid. The variations of thermal conductivity, density and specific heat due to temperature are taken into account. Most practical die casting geometries are complex and require unstructured grids. In our work, we have first generated a tetrahedral mesh using GMSH [20] and then divided into a hexahedral mesh using TETHEX [21]. The details of the numerical algorithm and verification and validation of OpenCast are discussed in previous publications [15,22]. A model geometry representing a clamp [15] has been considered to illustrate the methodology. Fig. 1 shows the clamp geometry with a mesh having 334000 hexahedral elements. It is important to assess the effects of natural convection. Hence, the clamp geometry is simulated for two cases viz. with and without natural convection. Figs. 2 and 3 plot grain size and yield strength contours with identical process conditions for both the cases. Since the solidification time is around 2.5 s, the velocities due to natural convection in the liquid metal are observed to be negligible. It is evident from the contours that there is no significant effect of natural convection and hence, it is neglected in all our further simulations.

3. Optimization

In die casting the mold cavity is filled with molten metal and solidified. The heat is extracted from the cavity walls by flowing coolants (typically water) through the cooling lines made inside the die. The quality of the finished product depends on the rate of heat extraction which in turn depends on the temperature at the cavity walls. Due to complexity in the die geometry, the wall temperature varies locally. An optimal product quality can be achieved if the temperature distribution

on the cavity walls and initial fill temperature are set properly. Thus, in this work, the following optimization problem with three objectives is proposed:

$$\begin{aligned} & \text{Minimize } \{f_1(T_{init}, T_{wall}), f_2(T_{init}, T_{wall}), f_3(T_{init}, T_{wall})\} \\ & \text{subject to } 900 \leq T_{init} \leq 1100 \text{ K and } 500 \leq T_{wall} \leq 700 \text{ K} \end{aligned} \quad (8)$$

where f_1 = solidification time, $f_2 = \max(\text{grain size})$ and $f_3 = -\min(\text{yield strength})$. Minimizing the solidification time increases productivity. Reduction in grain size reduces susceptibility to cracking [23] and improves mechanical properties of the product [24]. Thus, minimization of the maximum value of grain size over the entire geometry is set as an optimization objective. Higher yield strength is desirable as it increases the elastic limit of the material. Hence, the minimum yield strength over the entire geometry is to be maximized. For convenience, this maximization problem is converted to minimization by multiplying by minus one. This explains the third objective function f_3 . All the objectives are functions of the initial molten metal temperature (T_{init}) and mold wall temperature (T_{wall}). The initial temperature is a single value in the interval [900, 1100] K which is higher than the liquidus temperature of the alloy. As discussed before, the mold wall temperature need not be uniform in die casting due to locally varying heat transfer to the cooling lines. Thus, in this work, the wall surface is decomposed into multiple domains with each domain having a uniform temperature boundary condition which is held constant with time during the entire solidification process. If the die design with cooling line placement and coolant flow conditions are included, the thermal analysis of the die can also be done to identify these domains. Due to the lack of this information, the wall is decomposed into ten domains using the KMeans classification algorithm from Scikit Learn [25]. Fig. 4a shows the domain decomposition with 10 domain tags and Fig. 4b shows a random sample of the boundary temperature with a single temperature value assigned uniformly to each domain. Thus, the

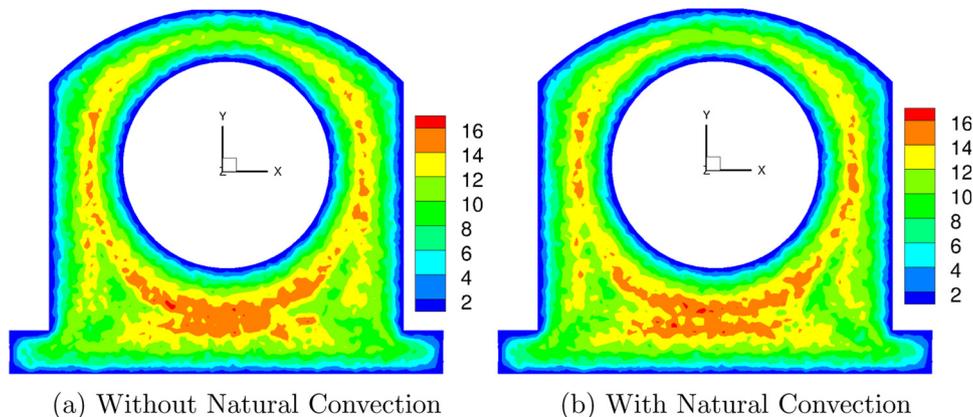


Fig. 2. Clamp: grain Size (μm).

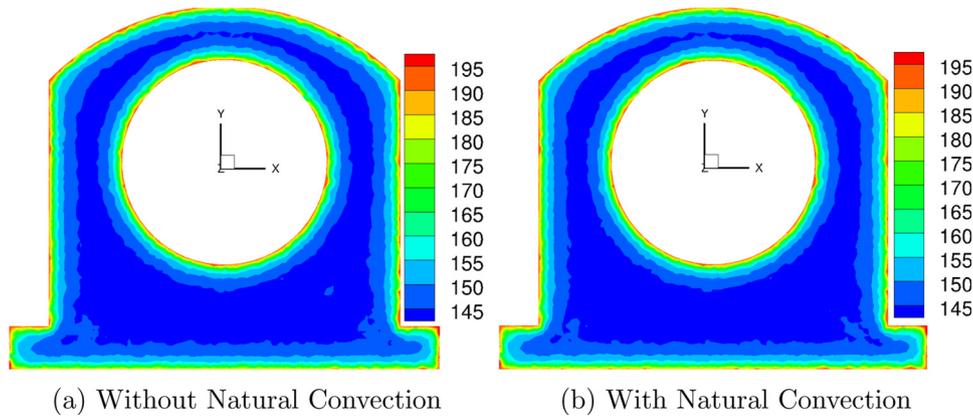


Fig. 3. Clamp: yield strength (MPa).

input wall temperature (T_{wall}) is a ten dimensional vector in the interval [500, 700] K which is lower than the solidus temperature of the alloy. Hence, this is a multi-objective optimization problem with three minimization objectives which are a function of 11 input temperatures.

Figs. 5 and 6 plot temperature and solid fraction for different time steps during solidification for the sample shown in Fig. 4b with $T_{init} = 986$ K. It can be seen that the temperature and solid fraction contours are asymmetric due to non-uniform boundary temperature. For instance, domain number 10 is held at minimum temperature and thus, region near it solidifies first. Fig. 7 plots final yield strength and grain size contours. The cooling rates and temperature gradients decrease with time as solidification progresses. Hence, the core regions which are thick take longer time to solidify. As the grains have more time to grow, the grain size is higher in the core region and correspondingly, the yield strength is lower. These trends along with the asymmetry are visible in Fig. 7.

3.1. Genetic algorithm

3.1.1. Single objective optimization

Genetic algorithms (GAs) are global search algorithms based on the mechanics of natural selection and genetics. They apply the ‘survival of the fittest’ concept on a set of artificial creatures characterized by strings. In the context of a GA, an encoded form of each input parameter is known as a gene. A complete set of genes which uniquely describe an individual (i.e., a feasible design) is known as a chromosome. The value of the objective function which is to be optimized is known as the fitness. The population of all the individuals at a given iteration is known as a generation. The overall steps in the algorithm are as follows:

1. Initialize first generation with a random population.
2. Evaluate the fitness of the population.

3. Select parent pairs based on their fitness (better fitness implies higher probability of selection).
4. Perform crossover to generate an offspring from each parent pair.
5. Mutate some genes randomly in the population.
6. Replace the current generation with the next generation.
7. If termination condition is satisfied, return the best individual of the current generation; else go back to step 2.

There are multiple strategies discussed in the literature for each of the above steps [26,27]. A brief overview of the methods used in this work is given here. The population is initialized using the Latin Hypercube Sampling strategy from the python package pyDOE [28]. Fitness evaluation is the estimation of the objective function which can be done by full scale computational model or by surrogate models. The number of objective function evaluations is typically of the order of millions and thus, step 2 becomes computationally most expensive step if a full scale model is used. Instead, it is common to use a surrogate model which is much cheaper to evaluate. In this work, a neural network based surrogate model is used, details of which are provided in Section 3.2. Tournament selection is used to choose the parent pairs to perform crossover. The idea is to choose four individuals at random and select two out of them which have better fitness. Note that since the optimization is cast as a minimization problem, lower fitness value is desired. Uniform crossover is used to recombine the genes of the parents to generate the offspring with a crossover probability of 0.9. Random mutation of the genes of the entire generation is performed with a mutation probability of 0.1. Thereafter, the old generation is replaced by this new generation. Note that the elitist version of GA is used which passes down the fittest individual of the previous generation to this generation as it is. Elitism was found helpful as it ensures that the next generation is at least as good as the previous generation.

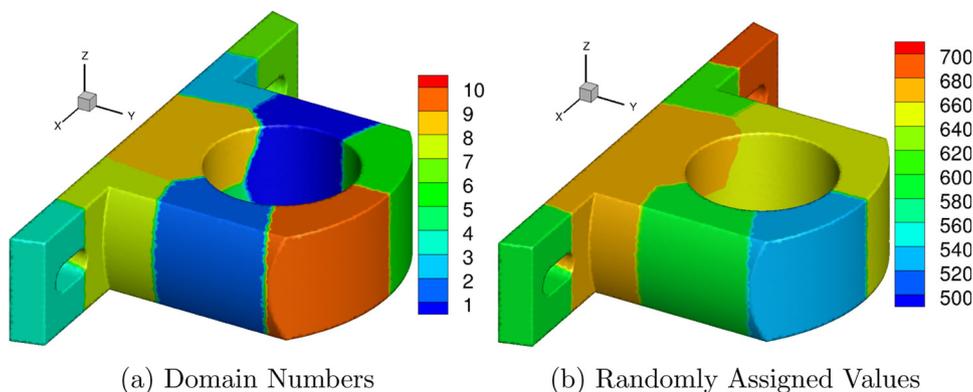


Fig. 4. Domain decomposition of the boundary and random value assignment.

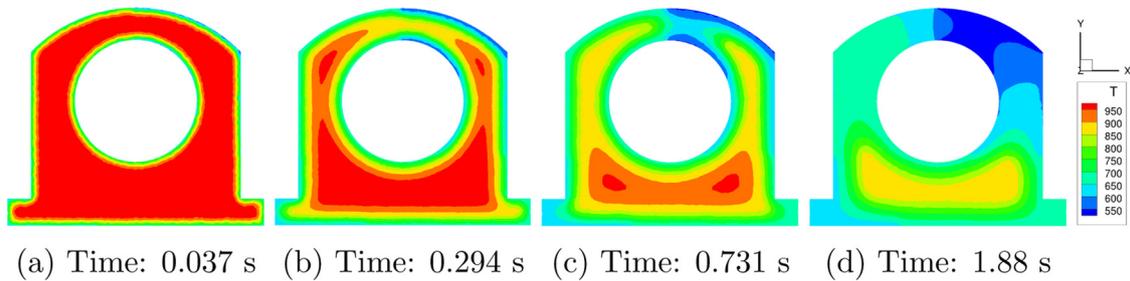


Fig. 5. Clamp temperature contours (K) (solidification time: 3.02 s).

3.1.2. Multi-objective optimization

The simultaneous optimization of multiple objectives is different than the single objective optimization problem. In a single objective problem, the best design which is usually the global optimum (minimum or maximum) is searched for. On the other hand, for multi-objective problem, there may not exist a single optimum which is the best design or global optimum with respect to all the objectives simultaneously. This happens due to the conflicting nature of objectives, i.e., improvement in one can cause deterioration of the other objectives. Thus, typically there is a set of Pareto optimal solutions which are superior to rest of the solutions in the design space which are known as dominated solutions. All the Pareto optimal solutions are equally good and none of them can be prioritized in the absence of further information. Thus, it is useful to have a knowledge of multiple non-dominated or Pareto optimal solutions so that a single solution can be chosen out of them considering other problem parameters.

One possible way of dealing with multiple objectives is to define a single objective as a weighted sum of all the objectives. Any single objective optimization algorithm can be used to obtain an optimal solution. Then the weight vector is varied to get a different optimal solution. The problem with this method is that the solution is sensitive to the weight vector and choosing the weights to get multiple Pareto optimal solutions is difficult for a practical engineering problem. Multi-objective GAs attempt to handle all the objectives simultaneously and thus, annihilating the need of choosing the weight vector. Konak et al. [29] have discussed various popular multi-objective GAs with their benefits and drawbacks. In this work, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) [30] which is a fast and elitist version of the NSGA algorithm [31] is used. The NSGA-II algorithm to march from a given generation of population size N to a next generation of same size is as follows:

1. Select parent pairs based on their rank computed before (lower rank implies higher probability of selection).
2. Perform crossover to generate an offspring from each parent pair.
3. Mutate some genes randomly in the population thus forming the offspring population.
4. Merge the parent and offspring population thus giving a set of size $2 \times N$.
5. Evaluate the fitness of the population corresponding to each

objective.

6. Divide the population into multiple non-dominated levels also known as fronts.
7. Compute the crowding distance for each individual along each front.
8. Sort the population based on front number and crowding distances and rank them.
9. Choose the best set of N individuals as next generation (i.e., N individuals with lowest ranks).
10. If termination condition is satisfied, return the best front of the current generation as an approximation of the Pareto optimal solutions; else go back to step 1.

Before iterating over the above steps, some pre-processing is required. A random population of size N is initiated and steps 5–8 are implemented to rank the initial generation. The parent selection, crossover and mutation steps are identical to the single objective GA described in Section 3.1.1. The algorithms for remainder of the steps can be found in the paper by Deb et al. [30]. Ranking the population by front levels and crowding distance enforces both elitism and diversity in the next generation.

3.2. Neural network

The fitness evaluation step of the GA requires a way to estimate the outputs corresponding to the set of given inputs. Typically, the number of generations can be of the order of thousands with several hundreds of population size per generation and thus, the total number of function evaluations can be around hundred thousand or more. It is computationally difficult to run the full scale numerical estimation software. Thus, a surrogate model is trained which is cheap to evaluate. A separate neural network is trained for each of the three optimization objectives (Eq. (8)). Hornik et al. [32] showed that under mild assumptions on the function to be approximated, a neural network can achieve any desired level of accuracy by tuning the hyper-parameters. The building block of a neural network is known as a neuron which has multiple inputs and gives out single output by performing the following operations:

1. Linear transformation: $a = \sum_{i=1}^n w_i x_i + b$; where, $\{x_1, x_2, \dots, x_n\}$ are n scalar inputs, w_i are the weights and b is a bias term.

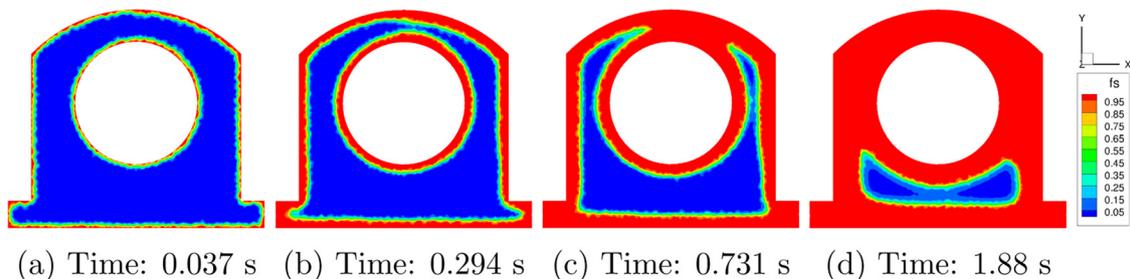


Fig. 6. Clamp solid fraction contours (solidification time: 3.02 s).

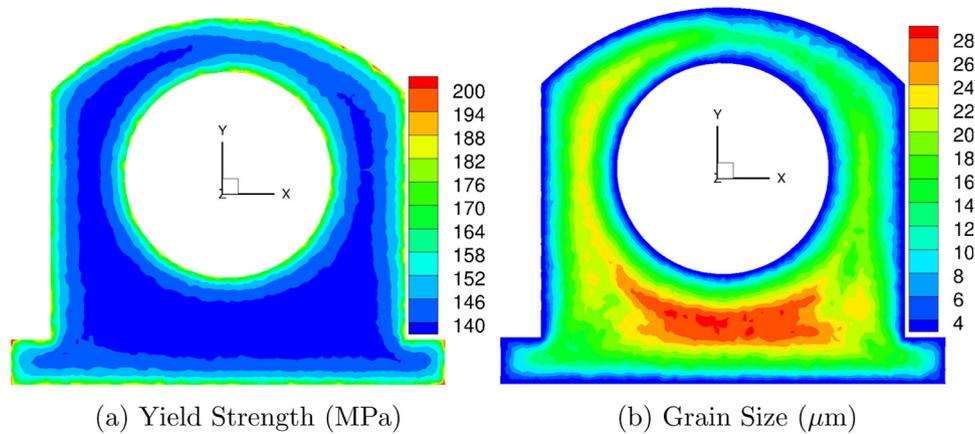


Fig. 7. Clamp microstructure parameters.

2. Nonlinear transformation applied element-wise: $y = \sigma(a)$; where, y is a scalar output and σ is the activation function.

Multiple neurons are stacked in a layer and multiple layers are connected together to form a neural network. The first and last layers are known as input and output layers respectively. Information flows from the input to output layer through the intermediate layers known as hidden layers. Each hidden layer adds nonlinearity to the network and thus, a more complex function can be approximated successfully. At the same time, having large number of neurons can cause high variance and thus, blindly increasing the depth of the network may not always help. The number of neurons in the input and output layers is defined by the problem specification. On the other hand, number of hidden layers and neurons has to be fine-tuned to have low bias and low variance. The interpolation error of the network is quantified as a loss function which is a function of the weights and bias.

A numerical optimization algorithm coupled with gradient estimation by the backpropagation algorithm [33] is used to estimate the optimal weights and bias which minimize the loss function for a given training set. This is known as training process. An approach described by Goodfellow et al. [34] is used to select the hyper-parameters and train the neural network. Mean square error is set as the loss function. A set of 500 random samples is used for training and two different sets of 200 each are used for validation and testing. The number of inputs and outputs of the problem specify the number of neurons in the input and output layers respectively. Here, there are three neural networks with 11 input neurons and one output neuron each (11 temperatures and three objectives mentioned in Eq. (8)). The number of hidden layers and hidden neurons, learning rate, optimizer, number of epochs and regularization constant are the hyper-parameters which are problem specific. They are varied in a range and then chosen so that both the training and validation errors are minimized simultaneously. The accuracy of prediction is further checked on an unseen test data. This overall approach is used for training the neural network with low bias and low variance. In this work, the neural network is implemented using the Python library Tensorflow [35] with a high level API Keras [36]. After testing various optimizers available in Keras, it is found that the Adam optimizer is the most suitable here. The parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$ are set as suggested in the original paper by Kingma and Ba [37] and the 'amsgrad' option is switched on. The activation functions used for all the hidden and output layers are ReLU and identity respectively. This choice of activation functions is generally recommended for regression problems [34]. A combination of L2 regularization with dropout is used to control the variance. The training is stopped when the loss function changes minimally after a particular epoch. This is known as the stopping criteria. After varying the hyper-parameters in a range, it is found that for the values listed in Table 1, the training, testing and validation errors (Table 2) are simultaneously

minimized. Moreover, since all the three errors are low and close to each other, it shows that the bias and variance are low. Fig. 8 plots percent relative error for 200 testing samples. It can be seen that all the three neural networks are able to predict with acceptable accuracy.

4. Assessment of genetic algorithm on simpler problems

It is difficult to visualize the variation of an output with respect to each of the eleven inputs. Hence in this section, two problems are considered which are simplified versions of the actual problem. In the first case, the boundary temperature is set to a uniform value and hence, there are only two scalar inputs: (T_{init} , T_{wall}). For the second case, the initial temperature is held constant ($T_{init} = 1000$ K) and the boundary is split into two domains instead of ten. Thus, again there are two scalar inputs: ($T_{wall}^{(1)}$, $T_{wall}^{(2)}$). $T_{wall}^{(1)}$ is assigned to domain numbers 1–5 and $T_{wall}^{(2)}$ to domain numbers 6–10 (Fig. 4a). The ranges of wall and initial temperatures are the same as before (Section 3). Such a simplified analysis gives an insight into the actual problem. Moreover, since these are problems with two inputs, the optimization can be performed by brute force parameter sweep and compared to the genetic algorithm. This helps to fine tune the parameters and assess the accuracy of the GA.

4.1. Single objective optimization

In this section, all the objectives are analyzed individually. A two dimensional mesh of size 40,000 with 200 points in each input dimension is used for this analysis. The outputs are estimated from the neural networks for each of these points. Fig. 9 plots the response surface contours for each of the three objectives with their corresponding minima. The minima are estimated from the 40,000 points. The X and Y axes are initial and wall temperatures, respectively. When the initial temperature is reduced, the total amount of internal energy in the molten metal is reduced and thus, the solidification time decreases. The amount of heat extracted is proportional to the temperature gradient at the mold wall which increases with a drop in the wall temperature. Thus, the drop in wall temperature reduces the solidification time. Hence, the minimum of solidification time is attained at the bottom left corner (Fig. 9a). The grain size is governed by the local temperature gradients and cooling rates which are weakly dependent on the initial temperature. Thus, it can be seen that the contour lines are nearly horizontal in Fig. 9b. On the other hand, as wall temperature reduces, the rate of heat extraction rises and hence, the grains get less time to grow. This causes a drop in the grain size. Thus, the minimum of the maximum grain size is on the bottom right corner (Fig. 9b). The contour values in Fig. 9c are negative as maximization objective of the minimum yield strength is converted into minimization by inverting the

Table 1
Neural network hyper-parameters.

| Name of network | No. of hidden layers | No. of neurons per hidden layer | Learning rate | No. of epochs | L2 λ | Dropout factor |
|-----------------|----------------------|---------------------------------|---------------|---------------|--------------|----------------|
| Sol. time | 4 | 50 | 0.001 | 300 | 0.004 | 0 |
| Max. grain | 6 | 75 | 0.001 | 300 | 0.005 | 0.2 |
| Min. yield | 4 | 25 | 0.003 | 400 | 0.01 | 0 |

Table 2
Average percent training, testing and validation errors.

| Name of network | Training error | Testing error | Validation error |
|-----------------|----------------|---------------|------------------|
| Sol. time | 0.90% | 1.01% | 1.08% |
| Max. grain | 1.29% | 2.04% | 1.92% |
| Min. yield | 0.25% | 0.38% | 0.43% |

sign. The minimum is at the top right corner of Fig. 9c. Fig. 10 has similar plots for the second case of constant initial temperature and split boundary temperature. Fig. 10c shows the effect of nonuniform boundary temperature. The minimum is attained at wall temperatures of 500 and 700 K since the local gradients and cooling rates vary due to the asymmetry in the geometry. This analysis shows the utility of the optimization with respect to nonuniform mold wall temperatures.

Table 3 lists the optima for the single objective problems estimated from parameter sweep and GA. Note that since the 200 K range is divided into 200 divisions, the resolution of the parameter sweep estimation is 1 K. For all the six cases, the outputs and corresponding inputs show that the GA estimates are accurate. The GA parameters are varied in the following ranges: [25–100] generations, [10–50] population size, [0.75–0.9] crossover probability and [0.05–0.2] mutation probability. Similar ranges have been used in the literature [38–40]. Parameter values of 50 generations with population size of 25 and crossover and mutation probability of 0.8 and 0.1 respectively are found to give accurate estimates as shown in Table 3. The elitist version of GA is used which passes on the best individual from previous generation to the next generation.

4.2. Bi-objective optimization

In this section, two objectives are taken at a time for each of the two simplified problems defined in Section 4. As before, a two dimensional mesh of size 40,000 with 200 points in each input dimension is used for this analysis. The outputs are estimated from the neural networks for each of these points. The feasible region is the set of all the attainable designs in the output space which can be estimated by the parameter sweep. For a minimization problem, a design d_1 is said to dominate another design d_2 if all the objective function values of d_1 are less than or equal to d_2 . The design space can be divided in two disjoint sets S_p and S_d such that S_p contains all the designs which do not dominate each

other and at least one design in S_p dominates any design in S_d [41]. S_p is called as the Pareto optimal or non-dominated set whereas, S_d is called as the non-Pareto optimal or dominated set. Since the designs in Pareto optimal set are non-dominated with respect to each other, they all are equally good and some additional information regarding the problem is required to make a unique choice out of them. Thus, it is useful to have a list of multiple Pareto optimal solutions. Another way to interpret the Pareto optimal solutions is that any improvement in one objective will worsen at least one other objective thus, resulting in a trade-off [42].

The blue region in the left parts of Figs. 11–14 indicates the feasible region. Using a pairwise comparison of the designs in the feasible region obtained by parameter sweep, the Pareto front is estimated which is plotted in red. The right side plots of Figs. 11–14 show the Pareto fronts obtained using NSGA-II. The NSGA parameters are varied in the following ranges: [25–100] generations, [500–1500] population size, [0.75–0.9] crossover probability and [0.05–0.2] mutation probability. The population size used in this work is kept higher than the literature [43] to get a good resolution of the Pareto front. It can be seen that both the estimates match which implies that the NSGA-II implementation is accurate. A population size of 1000 is evolved over 50 generations with crossover and mutation probability of 0.8 and 0.1 respectively. Existence of multiple designs in the Pareto set implies that the objectives are conflicting. This can be confirmed from the single objective analysis. For instance, consider the two objectives solidification time and minimum yield strength in the uniform boundary temperature case. From Figs. 9a and c it can be seen that individual minima are attained at different corners. Moreover, the directions of descent are different for each objective and thus, at some points, improvement in one objective can worsen other. This effect is visible on the corresponding Pareto front plot in Fig. 11.

5. Results of multi-objective optimization problem with 11 inputs

After verification of the NSGA-II implementation on simplified problems, multi-objective design optimization with 11 inputs is solved. As discussed before, some additional problem information is required to choose a single design from all the Pareto optimal designs. In die casting, there is a lot of stochastic variation in the wall and initial temperatures. Shahane et al. [15] have performed parameter uncertainty propagation and global sensitivity analysis and found that the die casting outputs are sensitive to the input uncertainty. Thus, from a

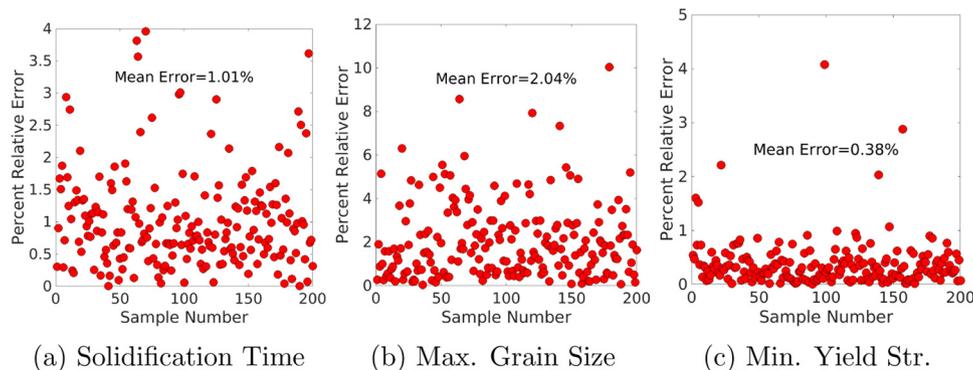


Fig. 8. Neural networks: error estimates for 200 testing samples.

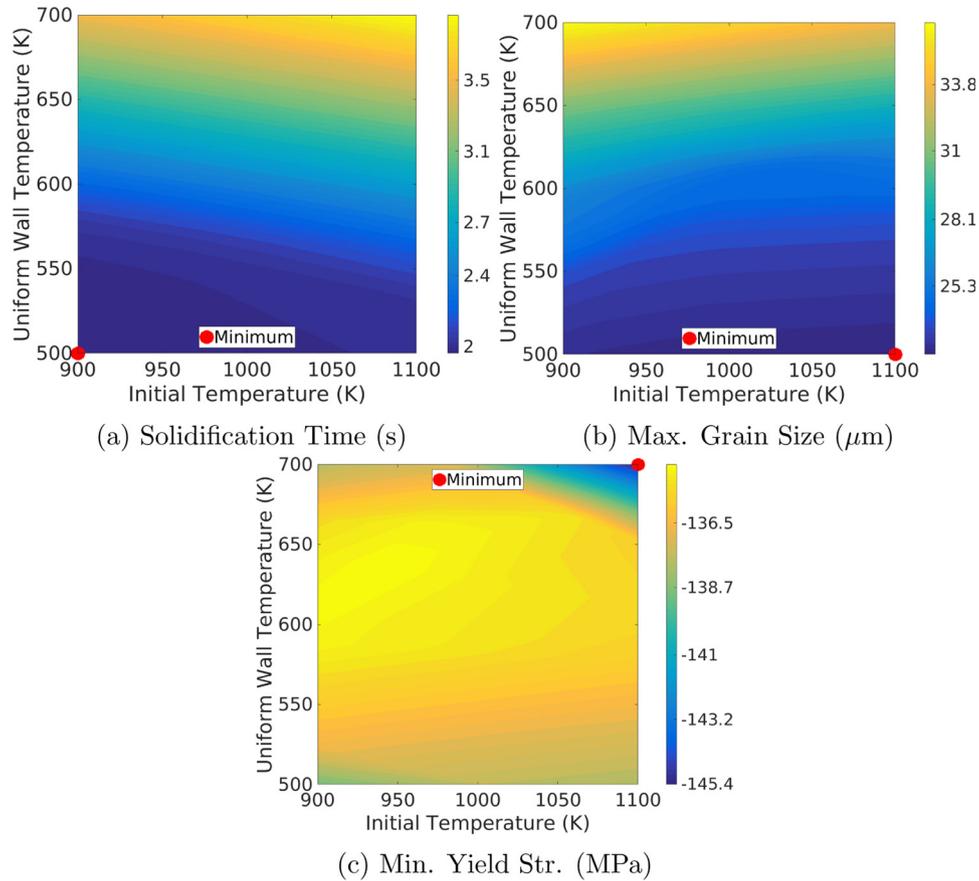


Fig. 9. Parameter sweep: uniform boundary temperature.

practical point of view, it is sensible to choose a Pareto optimal design which is least sensitive to the inputs. In this work, such an optimal point is known as a ‘stable’ optimum since any stochastic variation in the inputs has minimal effect on the outputs. A local sensitivity analysis is performed to quantify the sensitivity of outputs towards each input for all the Pareto optimal designs. For a function $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ which takes input $\mathbf{x} \in \mathbb{R}^n$ and produces output $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^m$, the $m \times n$ Jacobian matrix is defined as:

$$\mathbb{J}_f[i, j] = \frac{\partial f_i}{\partial x_j} \quad \forall \quad 1 \leq i \leq m, \quad 1 \leq j \leq n \quad (9)$$

At a given point \mathbf{x}_0 , the local sensitivity of \mathbf{f} with respect to each input can be defined as the Jacobian evaluated at that point: $\mathbb{J}_f(\mathbf{x}_0)$ [44]. Here, there are eleven inputs and two outputs. Thus, the 2×11 Jacobian is estimated at all the Pareto optimal solutions evaluated using the neural networks with a central difference method. Then, the L_1 norm of the Jacobian given by the sum of absolute values of all its components is defined as a single scalar metric to quantify the local sensitivity.

To begin with, two pairs of objectives are chosen: {solidification time, minimum yield strength} and {maximum grain size, minimum yield strength}. For both of these cases, a population size of 500 with 5000 generations is set. Fig. 15 plots the Pareto fronts colored by the value of Jacobian norm at each design. It can be seen that the norm varies significantly and thus, ranking the designs based on the sensitivity is useful. The design with minimum norm is chosen and marked on the Pareto fronts as a stable optimum. Note that minimum norm is observed at the end of the Pareto front. However, the norm is low on the entire left vertical side of the Pareto front. Hence, it may be a good idea to choose the design near shown ‘knee’ region which has similar value of the objective on the X-axis but much lower value of the objective on the Y-axis.

The next step is to perform the complete multi-objective optimization analysis as mentioned in Eq. (8). NSGA-II is used with a population size of 2000 evolved over 250 generations. Fig. 16 plots the Pareto optimal designs in a three dimensional objective space colored by the value of Jacobian norm at each design. Since it is difficult to visualize the colors on the three dimensional Pareto front, a histogram of norm of the Jacobian at each of these designs is also plotted in Fig. 17. It can be seen that the norm varies from 0.95 to 11.1. The histogram is skewed towards the left which implies that multiple designs are stable. The stable optimum is:

$$\begin{aligned} \text{Inputs: } & T_{\text{init}} = 1015.8 \text{ K} \\ & T_{\text{wall}} = \{500.7, 502.8, 500.0, 501.5, 500.5, \\ & \quad 503.6, 643.6, 508.8, 502.3, 500.7\} \text{ K} \\ \text{Outputs: } & \text{Solidification Time} = 1.99 \text{ s} \\ & \text{MaxGrainSize} = 22.39 \text{ } \mu\text{m} \\ & \text{MinYieldStrength} = 137.95 \text{ MPa} \end{aligned} \quad (10)$$

6. Conclusions

This paper presents an application of multi-objective optimization of the solidification process during die casting. Although the procedure is illustrated for a model clamp geometry, the process is general and can be applied to any practical geometry. Practically, it is not possible to hold the entire mold wall at a uniform temperature. The final product quality in terms of strength and micro-structure and process productivity in terms of solidification time depends directly on the rate and direction of heat extraction during solidification. Heat extraction in turn depends on the placement of coolant lines and coolant flow rates thus, being a crucial part of die design. In this work, the product quality is assessed as a function of initial molten metal and boundary

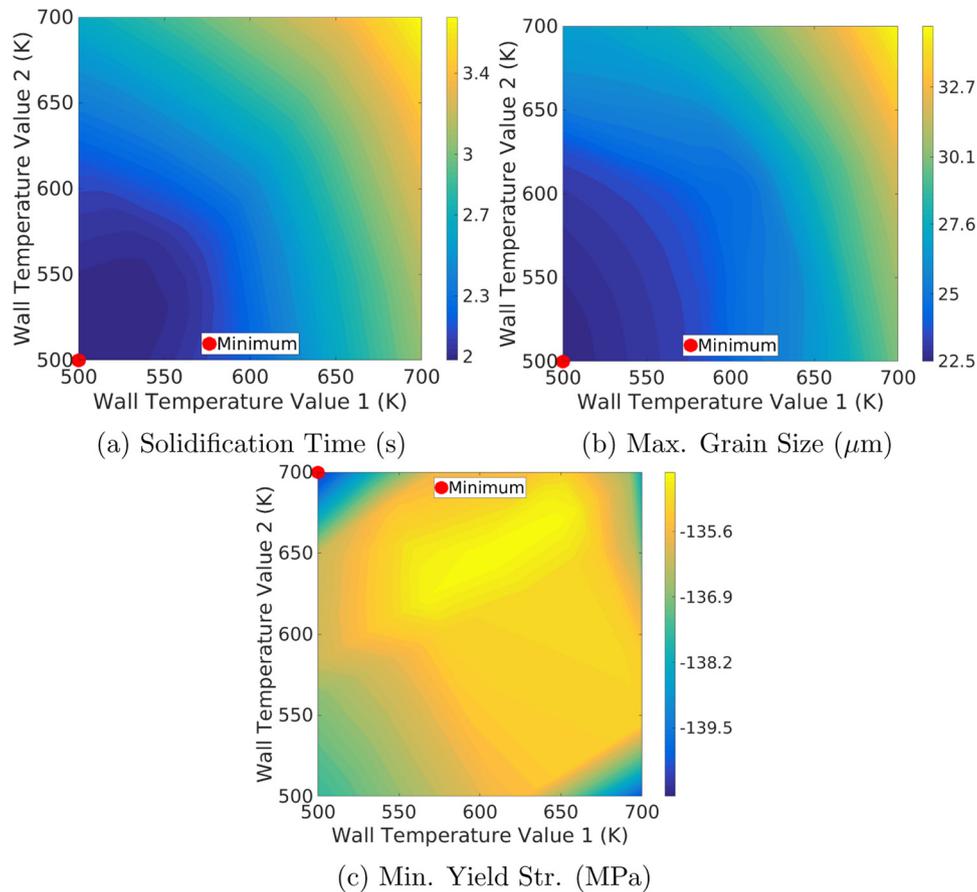


Fig. 10. Parameter sweep: split boundary temperature with $T_{init} = 1000$ K.

temperatures. Knowledge of boundary temperature distribution in order to optimize the product quality can be useful in die design and process planning. NSGA-II, which is a popular multi-objective genetic algorithm, was used for the optimization process. Since the number of function evaluations required for a GA is extremely high, a deep neural network was used as a surrogate to the full computational fluid dynamics simulation. The training and testing of the neural network was completed with less than thousand full scale finite volume simulations. The run time per simulation using OpenCast was about 20 min on a single processor, i.e., around 333 compute hours for 1000 simulations. All the simulations were independent and embarrassingly parallel. Thus, a multi-core CPU was used to speed up the process without any additional programming effort for parallelization. Computationally, this was the most expensive part of the process. Subsequent training and testing of the neural network took a few minutes. Implementation of GA is computationally cheap since the evaluation of a neural network is a sequence of matrix products and thus, was completed in few minutes. Hence, it can be seen that the strategy of coupling the GA and

neural network with finite volume simulations is computationally beneficial.

In this work, the wall is divided into ten domains. Together with the initial temperature, this is an optimization problem with 11 inputs. Both single and multi-objective genetic algorithms were programmed and verified with parameter sweep estimation for simplified versions of the problem. The single objective response surfaces were used to get an insight regarding the conflicting nature of the objectives since the individual optimal solutions were completely different from each other. Moreover, the solidification time, maximum grain size and minimum yield strength varied in the ranges [2, 3.5] s, [22, 34] μm and [134, 145] MPa respectively for the given inputs. This showed the utility of the simultaneous optimization of all the objectives since there was a significant scope for improvement. After estimating multiple Pareto optimal solutions, a common question is to choose a single design. The strategy of choosing the design with minimum local sensitivity towards the inputs was found to be practically useful due to the stochastic variations in the input process parameters. Overall, although

Table 3

Single objective optimization: genetic algorithm estimates compared with parameter sweep values for two input problems.

| Problem type | Optim. objective | Optim. type | Param. sweep | | GA estimates | |
|---------------------|------------------------------|-------------|--------------|--------|---------------|--------|
| | | | Inputs (K) | Output | Inputs (K) | Output |
| Uniform B.C. | Solid. time (s) | Min. | 900, 500 | 1.962 | 900.2, 500.2 | 1.962 |
| | Max. grain (μm) | Min. | 1100, 500 | 22.41 | 1099.1, 500.3 | 22.42 |
| | Min. yield (MPa) | Max. | 1100, 700 | 145.4 | 1099.9, 699.9 | 145.4 |
| $T_{init} = 1000$ K | Solid. time (s) | Min. | 500, 500 | 1.982 | 500.6, 500.3 | 1.982 |
| | Max. grain (μm) | Min. | 500, 500 | 22.49 | 500.1, 500.1 | 22.50 |
| | Min. yield (MPa) | Max. | 500, 700 | 140.9 | 500.1, 699.9 | 140.8 |

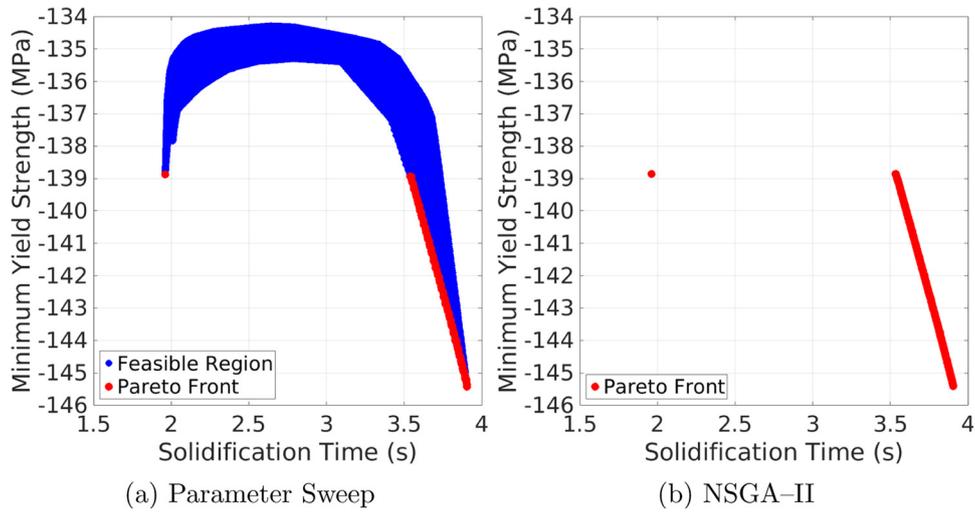


Fig. 11. Uniform boundary temperature: solidification time vs min. yield strength. (For interpretation of the references to color in this figure citation, the reader is referred to the web version of this article.)

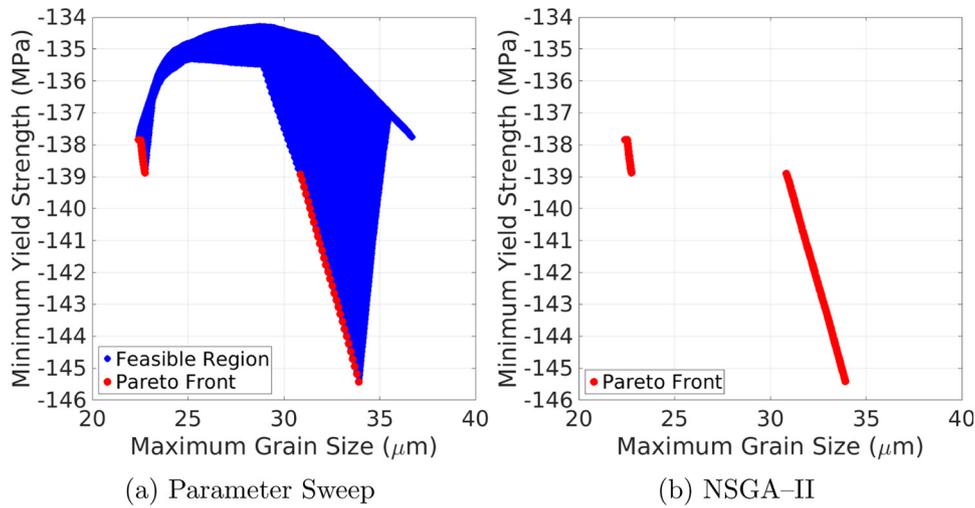


Fig. 12. Uniform boundary temperature: max. grain size vs min. yield strength. (For interpretation of the references to color in this figure citation, the reader is referred to the web version of this article.)

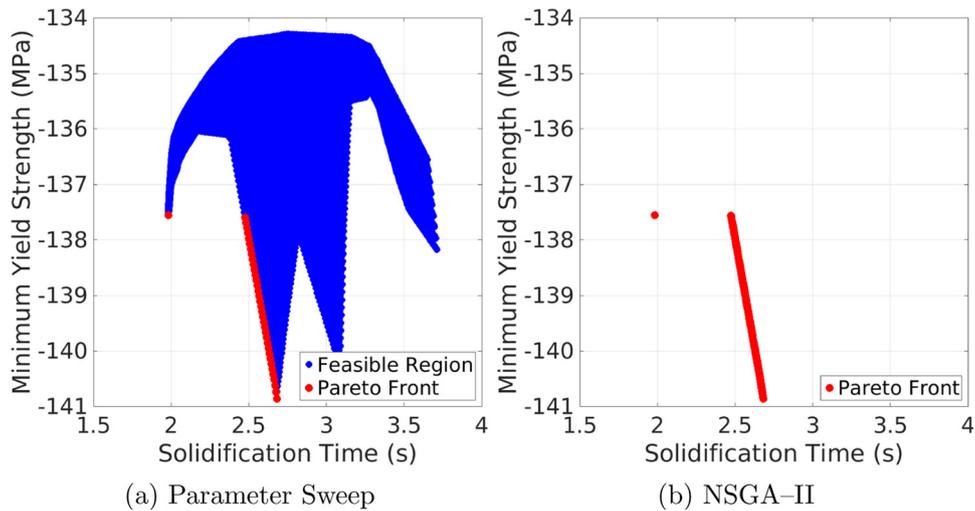


Fig. 13. Split boundary temperature with $T_{init} = 1000$ K: solidification time v/s min. yield str. (For interpretation of the references to color in this figure citation, the reader is referred to the web version of this article.)

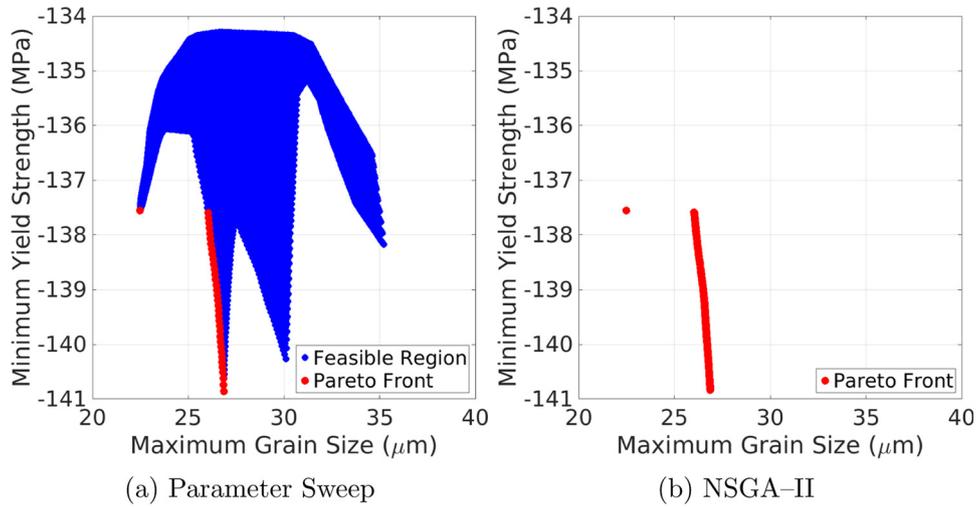


Fig. 14. Split boundary temperature with $T_{init} = 1000$ K: max. grain size v/s min. yield str. (For interpretation of the references to color in this figure citation, the reader is referred to the web version of this article.)

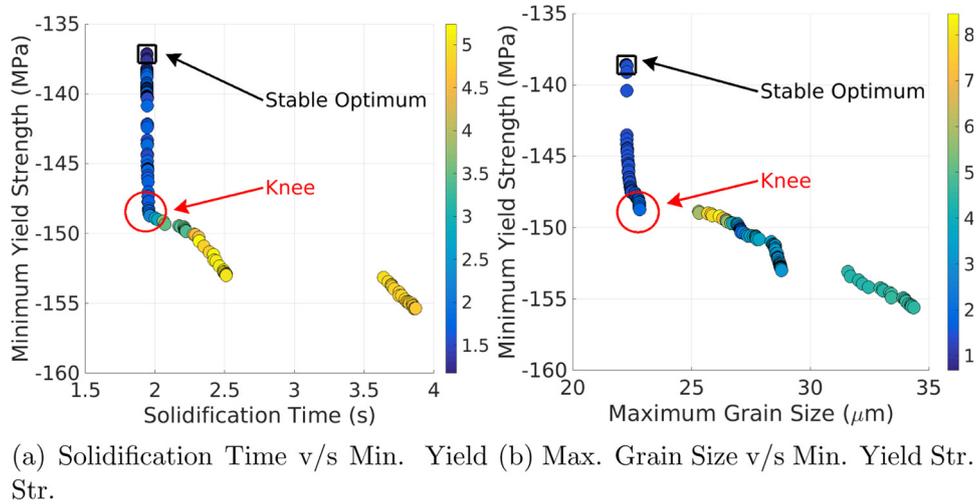


Fig. 15. Pareto front for two objectives (colored by Jacobian norm). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

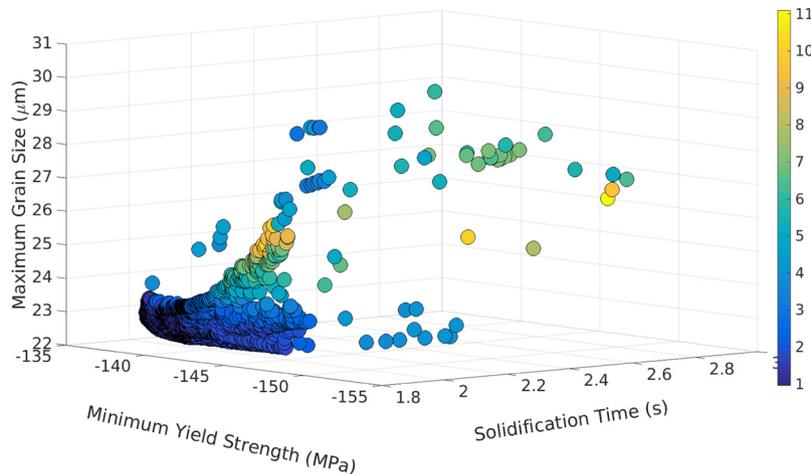


Fig. 16. Pareto front for three objectives (colored by Jacobian norm). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

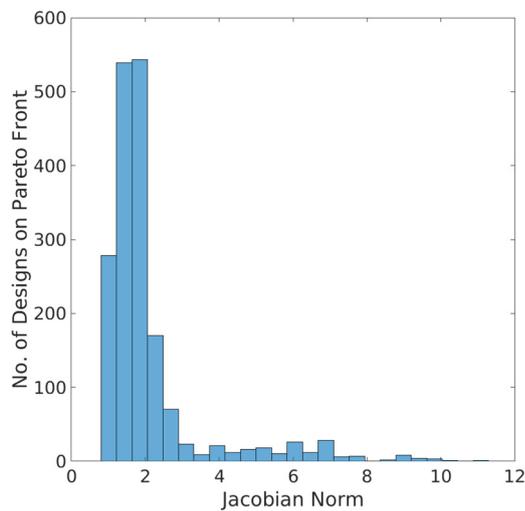


Fig. 17. Histogram of local sensitivity of designs on the Pareto front for 3 objectives.

die casting was used as an example for demonstration, this approach can be used for process optimization of other manufacturing processes like sand casting, additive manufacturing, welding, etc.

Conflict of interest

None declared.

Acknowledgments

This work was funded by the Digital Manufacturing and Design Innovation Institute with support in part by the U.S. Department of the Army. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of the Army. The authors would like to thank Beau Glim of North American Die Casting Association (NADCA) and Alex Monroe of Mercury Castings for their insightful suggestions.

References

- Minaie B, Stelson K, Voller V. Analysis of flow patterns and solidification phenomena in the die casting process. *J Eng Mater Technol* 1991;113:296–302.
- Im I, Kim W, Lee K. A unified analysis of filling and solidification in casting with natural convection. *Int J Heat Mass Transfer* 2001;44:1507–15.
- Cleary P, Ha J, Prakash M, Nguyen T. Short shots and industrial case studies: understanding fluid flow and solidification in high pressure die casting. *Appl Math Modell* 2010;34:2018–33.
- Plotkowski A, Fezi K, Krane M. Estimation of transient heat transfer and fluid flow for alloy solidification in a rectangular cavity with an isothermal sidewall. *J Fluid Mech* 2015;779:53–86.
- Poloni C, Giurgevich A, Onesti L, Pediroda V. Hybridization of a multi-objective genetic algorithm, a neural network and a classical optimizer for a complex design problem in fluid dynamics. *Comput Methods Appl Mech Eng* 2000;186:403–20.
- Elsayed K, Lacor C. Cfd modeling and multi-objective optimization of cyclone geometry using desirability function, artificial neural networks and genetic algorithms. *Appl Math Modell* 2013;37:5680–704.
- Wang W, He Y, Zhao J, Mao J, Hu Y, Luo J. Optimization of groove texture profile to improve hydrodynamic lubrication performance: theory and experiments. *Friction* 2018:1–12.
- Stavarakakis G, Karadimou D, Zervas P, Sarimveis H, Markatos N. Selection of window sizes for optimizing occupational comfort and hygiene based on computational fluid dynamics and neural networks. *Build Environ* 2011;46:298–314.
- Wei X, Joshi Y. Optimization study of stacked micro-channel heat sinks for micro-electronic cooling. *IEEE Trans Comp Pack Technol* 2003;26:55–61.
- Husain A, Kim K. Enhanced multi-objective optimization of a microchannel heat sink through evolutionary algorithm coupled with multiple surrogate models. *Appl Thermal Eng* 2010;30:1683–91.
- Lohan D, Dede E, Allison J. Topology optimization for heat conduction using generative design algorithms. *Struct Multidisc Optim* 2017;55:1063–77.
- Amanifard N, Nariman-Zadeh N, Borji M, Khalkhali A, Habibdoust A. Modelling and pareto optimization of heat transfer and flow coefficients in microchannels using GMDH type neural networks and genetic algorithms. *Energy Convers Manag* 2008;49:311–25.
- Esparza C, Guerrero-Mata M, Ríos-Mercado R. Optimal design of gating systems by gradient search methods. *Comput Mater Sci* 2006;36:457–67.
- Patel M, Shettigar A, Parappagoudar M. A systematic approach to model and optimize wear behaviour of castings produced by squeeze casting process. *J Manuf Process* 2018;32:199–212.
- Shahane S, Aluru N, Ferreira P, Kapoor S, Vanka S. Finite volume simulation framework for die casting with uncertainty quantification. *Appl Math Modell* 2019;74:132–50.
- Dantzig J, Tucker C. Modeling in materials processing. Cambridge University Press; 2001.
- Backer G, Wang Q. Microporosity simulation in aluminum castings using an integrated pore growth and interdendritic flow model. *Metall Mater Trans B* 2007;38:533–40.
- Okayasu M, Takeuchi S, Yamamoto M, Ohfuji H, Ochi T. Precise analysis of microstructural effects on mechanical properties of cast ADC12 aluminum alloy. *Metall Mater Trans A* 2015;46:1597–609.
- Greer A, Bunn A, Tronche A, Evans P, Bristow D. Modelling of inoculation of metallic melts: application to grain refinement of aluminium by Al-Ti-B. *Acta Mater* 2000;48:2823–35.
- Geuzaine C, Remacle J. GMSH: a 3-d finite element mesh generator with built-in pre-and post-processing facilities. *Int J Numer Methods Eng* 2009;79:1309–31.
- Artemiev M. TETHEX: conversion from tetrahedra to hexahedra. 2015 <https://www.github.com/martemyev/tethex>.
- Shahane S, Mujumdar S, Kim N, Priya P, Aluru N, Ferreira P, et al. Simulations of die casting with uncertainty quantification. *J Manuf Sci Eng* 2019;141:041003.
- Kimura R, Hatayama H, Shinozaki K, Murashima I, Asada J, Yoshida M. Effect of grain refiner and grain size on the susceptibility of Al-Mg die casting alloy to cracking during solidification. *J Mater Process Technol* 2009;209:210–9.
- Camicia G, Timelli G. Grain refinement of gravity die cast secondary AlSi7Cu3Mg alloys for automotive cylinder heads. *Trans Nonferr Metals Soc China* 2016;26:1211–21.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikitlearn: machine learning in Python. *J Mach Learn Res* 2011;12:2825–30.
- Goldberg D. Genetic algorithms in search, optimization and machine learning. Addison-Wesley Publishing Company, Inc.; 1989.
- Coley D. An introduction to genetic algorithms for scientists and engineers. World Scientific Publishing Company; 1999.
- pyDOE: The experimental design package for python. <https://www.pythonhosted.org/pyDOE/>.
- Konak A, Coit D, Smith A. Multi-objective optimization using genetic algorithms: a tutorial. *Reliab Eng Syst Saf* 2006;91:992–1007.
- Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput* 2002;6:182–97.
- Srinivas M, Deb K. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evol Comput* 1994;2:221–48.
- Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Netw* 1989;2:359–66.
- Chauvin Y, Rumelhart D. Backpropagation: theory, architectures, and applications. Psychology Press; 2013.
- Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016 <http://www.deeplearningbook.org>.
- Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: large-scale machine learning on heterogeneous systems. 2015 Software available from <https://www.tensorflow.org/>.
- Chollet F. Keras. 2015 <https://keras.io>.
- Kingma D, Ba J. Adam: a method for stochastic optimization. 2014 arXiv:1412.6980.
- Kittur J, Patel M, Parappagoudar M. Modeling of pressure die casting process: an artificial intelligence approach. *Int J Metalcast* 2016;10:70–87.
- Patel M, Krishna P, Parappagoudar M. An intelligent system for squeeze casting process-soft computing based approach. *Int J Adv Manuf Technol* 2016;86:3051–65.
- Patel M, Shettigar A, Krishna P, Parappagoudar M. Back propagation genetic and recurrent neural network applications in modelling and analysis of squeeze casting process. *Appl Soft Comput* 2017;59:418–37.
- Deb K. Multi-objective optimization using evolutionary algorithms. John Wiley and Sons; 2001.
- Messac A. Optimization in practice with MATLAB®: for engineering students and professionals. Cambridge University Press; 2015.
- Vundavilli P, Kumar J, Priyatham C, Parappagoudar M. Neural network-based expert system for modeling of tube spinning process. *Neural Comput Appl* 2015;26:1481–93.
- Smith R. Uncertainty quantification: theory, implementation, and applications. 2013.