

Sequential Design of an Injection Molding Process Using a Calibrated Predictor

PO-HSU ALLEN CHEN

Battelle Memorial Institute, Columbus, OH, USA

MARÍA G. VILLARREAL-MARROQUÍN,

Centro de Investigación en Matemáticas, CONACYT, Monterrey, NL, Mexico

ANGELA M. DEAN, THOMAS J. SANTNER

Department of Statistics, The Ohio State University, Columbus, OH, USA

RACHMAT MULYANA, and JOSÉ M. CASTRO

Integrated Systems Engineering Department, The Ohio State University, Columbus, OH, USA

Abstract

This paper optimizes an injection molding process using an efficient sequential design methodology. The goal is to set the process control variables to minimize the shrinkages of a selected collection of injection molded parts. This multi-objective optimization problem is solved by finding those process control variable settings that are Pareto minimizing values, i.e., process settings for which none of the shrinkages of the parts can be decreased by an alternative process setting without increasing the shrinkages of other parts. The sequential design uses an expected improvement criterion to guide updates. The shrinkages are estimated by a calibrated predictor of the process mean shrinkage. The calibration is based on observations of the manufacturing process supplemented by computer runs of a commercial simulator code that mimics the manufacturing process.

Key Words: Advanced Manufacturing; Bayesian Analysis; Calibration; Injection Molded Plastic; Multiobjective Optimization; Pareto Set; Pareto Front; Relative shrinkage

1 Introduction

Many real-world applications require the simultaneous optimization of multiple competing objective functions. For example, Leatherman et al. (2014) developed a finite element model of the human knee in order to identify robust designs for a meniscal substitute which can provide both small mean and low variability in peak contact stress. For Titanium Nitride/Titanium multilayer tool coatings, Draguljić et al. (2015) used computer experiments for simultaneously minimizing the maximum radial stress (associated with cohesive failures) and the maximum shear stress (associated with adhesive failures). Atashkari et al. (2005) sought settings of the turbine inlet temperature, the pressure ratio and the flight Mach number to optimize specific thrust, thrust-specific fuel consumption, propulsive efficiency, and thermal efficiency in the thermodynamic cycle of ideal turbojet engines.

The goal of multi-objective optimization is to identify optimal combinations of settings of controllable variables (called *control factors* or *input variables*). A single combination of control factor settings that optimizes all responses (outputs) simultaneously may not exist, and some settings may result in better responses for one objective, while others may be better for a different objective. Consequently, control factor settings are sought which are better than all other settings *for at least one objective*. Such a set of control factor settings is called the *Pareto Set* and the corresponding set of multi-objective responses is called the *Pareto Front* (see Section 4.1).

In this paper, without loss of generality, we deal with multi-objective *minimization*. In particular, we consider the situation where a few observations of the physical process can be made, together with runs from a *computer simulator*. A simulator is a computational implementation of a mathematical model which describes the physical process. Here, we consider the situation where the simulator is deterministic (providing the same outputs at repeated runs of the same input setting), and only a limited number of observations can be taken on both the physical process and the simulator.

A common problem is that a computer simulator may provide biased output for the physical process due to the simplified physics or biology used in the mathematical model. However, when physical observations are available, it may be possible to use these data to align the simulator output to be close to the true mean response by constructing a bias-corrected predictor

(a process called *calibration*); for an example, see Williams et al. (2006).

Sacks et al. (1989) proposed a methodology for modeling deterministic simulator output as a realization of a Gaussian Process (GP); see also Santner et al. (2003). Using a GP model and a Bayesian calibrated predictor, Chen et al. (2017) developed a sequential methodology for identifying an approximate (estimated) Pareto Set and Pareto Front. In this paper, we apply the sequential methodology to minimize shrinkage in an injection molded plastic component. Section 2 describes the manufacturing process, the component, the responses, and the simulator bias. The Bayesian calibrated predictor that will be used to correct for the bias is described in Section 3. Section 4 gives an overview of the sequential procedure of Chen et al. (2017) for finding the Pareto Set and Pareto Front. This procedure is then applied in Section 5 to the injection molding optimization. Two different examples with known solutions, one with two objective functions and one with three, are used in Section 6 to show the success of the sequential procedure. In addition, Section 6 investigates the allocation of simulator runs between the initial and the sequential parts of the simulator design when the experimental budgets are fixed. Finally, Section 7 provides a summary and discussion.

2 Injection Molding Shrinkage Optimization

The multi-objective injection molding optimization problem discussed in this paper was introduced by Villarreal-Marroquín et al. (2017) to search for process settings that lead to minimum shrinkage in plastic test components. These components, depicted in Figure 1, were produced by injecting a thermoplastic polyolefin material (manufactured by LyondellBasell) into an ASTM mold (Figure 2) in a Sumitomo 180 Ton injection-molding machine.

The injection molding process starts by loading plastic pellets into a hopper at one end of the injection-molding machine. A rotating screw inside the machine drags the pellets forward, compacts and melts them. When sufficient molten plastic has accumulated, the screw head acts as a plunger to inject the melted plastic into the mold (the filling stage). After the mold is filled, more material is pushed into the mold using higher pressure (the packing stage) to account for thermal shrinkage. The temperature of the mold itself is kept below the melting point of the material so that, during packing, the material starts to cool. Once the material has cooled sufficiently and is solid, the mold is opened and the component removed.

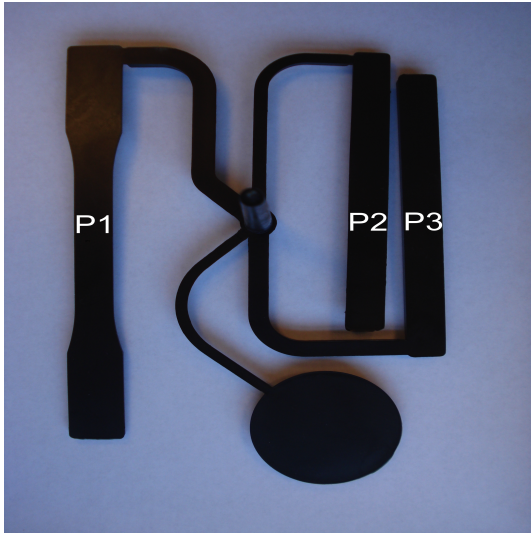


Figure 1: A plastic test component. The mid-tapered bar is P1 and the two rectangular bars are P2 and P3.



Figure 2: ASTM mold in a Sumitomo 180 Ton injection-molding machine

The goal of the study described in this paper was to identify settings of the control factors which lead to minimum shrinkage in the lengths of a mid-tapered bar and two rectangular bars, which we label P1, P2, and P3 as shown in Figure 1. The budget allowed a small number of components to be manufactured and measured and, in addition, a computer simulator of the physical process was available for providing additional shrinkage observations although, as will be seen later, its outputs were biased. The control factors, the responses, and the design of the physical and simulator experiments are described below.

2.1 Control Factors and Responses

Four control factors, whose optimum settings were sought in the manufacturing process, are

- (i) T_{melt} , melt temperature in $^{\circ}\text{C}$: temperature to which the plastic is heated before being injected into the mold.
- (ii) t_{pack} , packing time in seconds: time for which the additional material is pushed into the mold under higher pressure.
- (iii) P_{pack} , packing pressure in MPa: pressure at which the additional material is pushed into the mold.
- (iv) t_{cool} , cooling time in seconds: time for which the part is allowed to cool after packing and before the mold is opened and the component removed.

Two additional control factors, the mold temperature and filling time, were held constant at 26.7°C and 1 second, respectively. For the remainder of the paper, the term “control factor setting” refers to a value of the vector of the four control factors (T_{melt} , t_{pack} , P_{pack} , t_{cool}).

For each control factor setting, to ensure that the mold had reached thermal equilibrium, the first few molded components were discarded. The peak surface temperature of each component was monitored using a thermocouple located at the thickest bar, P3, and thermal equilibrium was assumed when three consecutive cycles showed less than 0.55°C difference. After reaching equilibrium, the next five molded components were allowed to cool for at least 48 hours to equilibrate with the ambient environment, and then the linear dimensions of parts P1, P2, and P3 were measured using a digital caliper with an accuracy of 0.01 mm. The relative shrinkage of the length, width, and thickness of P1, P2 and P3 for each component was calculated as

$$\frac{\text{Target dimension} - \text{Observed Dimension of Component}}{\text{Target dimension}}.$$

Villarreal-Marroquín et al. (2017) discussed the Pareto minimization of the relative shrinkages of the length, width and thickness of part P3 using a calibrated predictor in multiple grid searches. Here, we consider the Pareto minimization of the relative shrinkages of the *lengths* of parts P1, P2, P3 relative to their target lengths. We show that the sequential methodology of Chen et al. (2017) can find a good approximation to the Pareto set without large grid searches.

2.2 Design of the Physical Experiment

The design selected for the physical experiment consists of the 16 combinations of a 2^4 factorial design in the four control factors T_{melt} , t_{pack} , P_{pack} , t_{cool} , each observed at the high and low levels towards the ends of their reasonable ranges. Three “center points” in the variables t_{pack} , P_{pack} , t_{cool} , were added at the low, high and center values of T_{melt} . The chosen control factor settings used in the design are shown in the first four columns of Table 1. The corresponding physical observations of the average relative shrinkage lengths, labelled $L1$, $L2$, $L3$, are shown in the last three columns of Table 1; these observations are plotted as “+” in Figure 4 (and zoomed in Figure 5).

Each control factor setting was observed four times and, each time, five components were produced and measured as described above. The corresponding relative shrinkage was mea-

T_{melt}	t_{pack}	P_{pack}	t_{cool}	$L1$	$L2$	$L3$
184	14	32.14	36	.0049	.0022	.0059
184	14	32.14	50	.0048	.0020	.0057
184	14	42.67	36	.0044	.0017	.0056
184	14	42.67	50	.0043	.0016	.0053
184	21	37.41	43	.0046	.0018	.0056
184	28	32.14	36	.0049	.0022	.0058
184	28	32.14	50	.0051	.0020	.0058
184	28	42.67	36	.0041	.0016	.0049
184	28	42.67	50	.0042	.0015	.0047
200	21	37.41	43	.0041	.0018	.0050
216	14	32.14	36	.0046	.0023	.0063
216	14	32.14	50	.0046	.0022	.0060
216	14	42.67	36	.0039	.0016	.0056
216	14	42.67	50	.0041	.0015	.0053
216	21	37.41	43	.0042	.0019	.0049
216	28	32.14	36	.0044	.0022	.0049
216	28	32.14	50	.0043	.0021	.0046
216	28	42.67	36	.0037	.0015	.0044
216	28	42.67	50	.0038	.0014	.0042

Table 1: The 19 distinct combinations of control factor settings used in the physical experiment, and the corresponding observations of relative shrinkages, $L1$, $L2$, and $L3$.

sured as the average over the 20 components produced for that control factor setting. This is depicted in Figure 3.

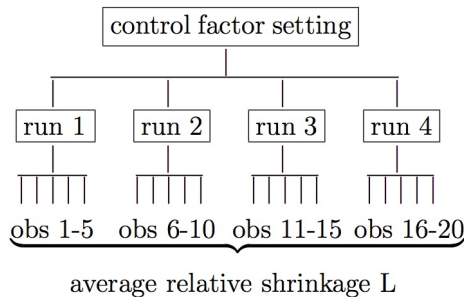


Figure 3: A diagram of the 20 components produced for a given control factor setting.

Even for the simple test components considered here, physical experimentation is costly and time-consuming; the large number of runs of the process required to achieve high quality is infeasible. However, a deterministic simulator of the physical process was available and was run to provide simulated observations to augment the physical responses. The simulator design is described in Section 2.3.

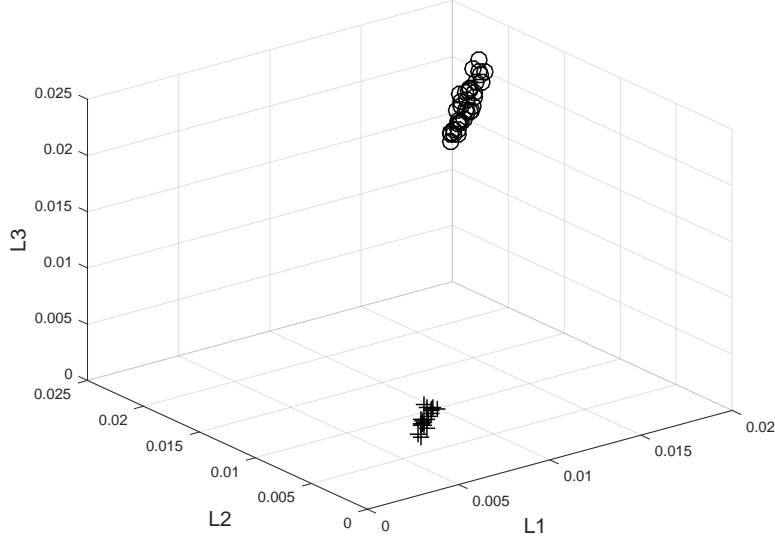


Figure 4: Scatter plot of the the 19 physical process observations (shown as “+”) of observed relative shrinkages $L1$, $L2$, $L3$ of parts P1, P2, P3 for the $2^4 + 3$ experimental design of Table 1, and the 35 Moldex3D $L1$, $L2$, $L3$ outputs (shown as “o”) from the 35-run 7-variable augmented maximin Latin hypercube design of Table 3.

2.3 Design of the Initial Simulator Experiment

The simulator, called “Moldex3D”, is commercially available software developed to simulate injection molding processes and is based on Generalized Newtonian Fluid assumptions. Moldex3D was used to simulate the non-isothermal three-dimensional flow motion of the polymer through the mold and to compute the resulting lengths of P1, P2, P3 of the simulated components (see Villarreal-Marroquín et al., 2017, for further details about the simulator). In addition to the specification of the settings of the four control factors (i)–(iv) listed in Section 2.1, the simulator requires the specification of the values of three heat transfer coefficients (HTCs) during filling, packing, and cooling. The ranges used for the four control factors and the three HTCs in the simulator experiment are shown in Table 2.

The (deterministic) Moldex3D simulator was run at settings of the seven variables (four control factors and three HTCs) defined by a 35-run augmented maximin Latin hypercube design (LHD) as listed in the left portion of Table 3. Table 3 also lists the relative shrinkages $L1$, $L2$, $L3$ in the lengths of P1, P2, P3, computed by Moldex3D. These 35 values of $L1$, $L2$, $L3$ are plotted as “o” in Figure 4. Compared with the 19 values of $L1$, $L2$, $L3$ computed from the physical observations and plotted as “+”, the bias in the simulator is readily apparent.

Consequently, to identify the control factor settings which lead to minimum relative shrinkage, calibrated predictors of the means of the relative shrinkages in the physical process were used. The exact boundary conditions of the HTCs are unknown and extremely difficult to measure in the physical injection molding process. The HTCs were used as *calibration parameters* to correct the bias in the Moldex3D simulator output.

In Section 3, a Gaussian Process (GP) model for the simulator output is described briefly. The Pareto Set and Pareto Front are defined in Section 4, together with an overview of the sequential Pareto optimization method of Chen et al. (2017). In Section 5, we will show that the sequential methodology locates an approximate Pareto Front and Pareto Set for this injection molding problem more quickly and more accurately than a fixed design with the same number of observations, even when coupled with a sizeable grid search.

3 The Simulator Model and its Calibration

As described in Section 2, the inputs to the deterministic Moldex3D simulator are the $d = 4$ control factors for the manufacturing system and the $b = 3$ heat transfer coefficients (HTCs) that will be used to calibrate the simulator. The HTCs are unknown and impossible to measure in the injection molding production environment but are required in the mathematical model that relates shrinkage to these factors. In this section, the stochastic model representing the simulator output will be described first, then that for the physical experiment, and finally the calibration model relating the two experimental outputs will be introduced.

Suppose that the deterministic simulator is calculated for n^s inputs where each input consists of d control factors and b calibration variables. Let $\mathbf{x}_i^s = (x_{i,1}^s, \dots, x_{i,d}^s)^\top$ and $\mathbf{t}_i =$

Variable	Low	High
Melt temperature (T_{melt} , °C)	180	220
Packing time (t_{pack} , sec)	10	30
Packing pressure (P_{pack} , MPa)	30	44
Cooling time (t_{cool} , sec)	25	50
HTC during filling (H_{fill} , W/m ² K)	1200	1800
HTC during packing (H_{pack} , W/m ² K)	20000	30000
HTC during cooling (H_{cool} , W/m ² K)	2000	3000

Table 2: Ranges of the four control factors for the physical and simulator experiments, and the ranges of the calibration parameters for the simulator experiment

T_{melt}	t_{pack}	P_{pack}	t_{cool}	H_{fill}	H_{pack}	H_{cool}	$L1$	$L2$	$L3$	seq
180.0	21.2	30.69	29.0	1560	27600	2940	.0168	.0182	.0218	
180.2	25.4	41.31	28.5	1728	29000	2300	.0151	.0166	.0196	–
180.8	28.9	32.44	43.7	1536	20933	2227	.0146	.0157	.0198	
181.2	23.4	37.77	48.8	1644	26400	2700	.0153	.0165	.0201	
182.4	27.4	35.67	25.3	1200	27200	2000	.0154	.0167	.0200	
183.2	12.0	31.21	50.0	1344	23800	2200	.0187	.0199	.0217	
184.0	29.6	43.80	25.0	1368	25400	2600	.0144	.016	.0191	
185.2	14.4	42.88	36.0	1788	21400	3000	.0169	.0183	.0196	
187.5	10.8	35.41	48.0	1608	24000	2920	.0183	.0196	.0209	–
188.8	20.4	33.31	39.5	1218	30000	2420	.0166	.018	.0216	
190.4	12.4	41.44	44.3	1626	29900	2120	.0173	.0187	.0197	
191.2	28.7	31.74	33.3	1800	25600	2467	.0149	.0162	.0203	
192.8	17.2	40.91	49.0	1776	23200	2060	.0165	.0178	.0199	–
194.0	20.7	36.59	30.0	1272	20500	2660	.0164	.0179	.0210	
196.0	26.0	41.96	47.5	1512	20400	2480	.0144	.0158	.0198	
198.0	13.4	38.82	37.3	1572	20300	2380	.0175	.0190	.0203	–
198.8	16.6	32.92	33.8	1710	24400	2020	.0176	.0190	.0215	
199.2	19.6	43.40	30.5	1212	23400	2160	.0161	.0175	.0196	–
202.0	11.6	34.10	43.0	1794	29600	2740	.0183	.0197	.0216	
203.6	17.6	40.65	35.0	1392	29800	2560	.0165	.0179	.0199	–
204.2	25.0	35.02	25.8	1680	21200	2900	.0158	.0175	.0214	
206.8	13.0	37.51	49.5	1314	29700	2980	.0174	.0189	.0208	
209.2	29.9	43.21	26.5	1758	20000	2260	.0144	.0161	.0196	
210.4	24.8	39.34	46.0	1296	24800	2760	.0152	.0167	.0202	–
211.8	27.8	32.66	40.3	1242	29200	2960	.0151	.0166	.0213	
212.8	14.0	30.95	46.5	1620	22000	2840	.0181	.0195	.0220	
214.9	27.1	42.57	34.0	1280	28400	2107	.0149	.0165	.0196	
215.6	29.1	40.00	40.8	1668	28000	2800	.0143	.0158	.0200	
216.0	24.4	36.20	29.5	1410	22100	2080	.0160	.0177	.0208	
216.4	11.0	34.88	39.0	1320	26000	2520	.0185	.0202	.0226	
217.3	17.7	33.66	32.3	1600	28800	2987	.0172	.0187	.0211	–
218.4	26.8	31.48	27.5	1500	25000	2620	.0159	.0177	.0218	–
218.8	15.0	42.75	45.0	1356	23300	2040	.0166	.0183	.0198	
219.2	22.4	31.08	48.3	1452	25800	2360	.0165	.0181	.0215	–
220.0	10.0	39.86	49.8	1224	20200	2860	.0177	.0193	.0226	

Table 3: The 35 combinations of input variable settings used in the simulator experiment and the corresponding relative shrinkages $L1$, $L2$, $L3$ computed from the simulator outputs. Columns 1-7 show the settings of the four control factors and the three calibration parameters at which Moldex3D was run; columns 8-10 show $L1$, $L2$, $L3$. (“–” indicates the settings that will be omitted from the initial design in the sequential procedure in Section 5)

$(t_{i,1}, \dots, t_{i,b})$ denote the control factor and calibration variable settings, respectively, where \top denotes transpose; also let $y^s(\mathbf{x}_i^s, \mathbf{t}_i)$ denotes the simulator output when run at $(\mathbf{x}_i^s, \mathbf{t}_i)$, $i = 1, \dots, n^s$. Because the simulator output is deterministic it is assumed that the $(\mathbf{x}_i^s, \mathbf{t}_i)$, for $i = 1, \dots, n^s$, are distinct.

The simulator output is modeled using a generalization of a regression model that postulates the outputs from “nearby” pairs of $(\mathbf{x}^s, \mathbf{t})$ are correlated. This model regards the $y^s(\mathbf{x}_i^s, \mathbf{t}_i)$, for $i = 1, \dots, n^s$, as realizations from

$$Y^s(\mathbf{x}^s, \mathbf{t}) = \beta_0 + Z(\mathbf{x}^s, \mathbf{t}), \quad (1)$$

where β_0 is a constant mean, and $Z(\cdot, \cdot)$ is a zero-mean, stationary Gaussian process with variance λ_Z^{-1} , and output correlation

$$\text{Cor}(Y^s(\mathbf{x}_i^s, \mathbf{t}_i), Y^s(\mathbf{x}_j^s, \mathbf{t}_j)) = R_Z((\mathbf{x}_i^s, \mathbf{t}_i), (\mathbf{x}_j^s, \mathbf{t}_j)) = \prod_{k=1}^d \rho_{x,k}^{4(x_{i,k}^s - x_{j,k}^s)^2} \prod_{e=1}^b \rho_{t,e}^{4(t_{i,e} - t_{j,e})^2}, \quad (2)$$

for $i \neq j$. The parameters in (2) must satisfy $0 \leq \rho_{x,k}, \rho_{t,e} \leq 1$, for $k = 1, \dots, d$ and $e = 1, \dots, b$; it can be shown that they control the smoothness of the $y^s(\mathbf{x}^s, \mathbf{t})$ realizations (see Section 2.3.4 of Santner et al. (2003)). The Gaussian process model is a general linear model because it can be shown that the joint distribution of any finite set of $Y^s(\mathbf{x}_i^s, \mathbf{t}_i)$ with distinct $(\mathbf{x}_i^s, \mathbf{t}_i)$ is multivariate normal with mean β_0 , variance $\sigma_Z^2 = \lambda_Z^{-1}$, and covariance matrix that has elements of the form $\lambda_Z^{-1} \times R_Z((\mathbf{x}_i^s, \mathbf{t}_i), (\mathbf{x}_j^s, \mathbf{t}_j))$. Lastly, it should be noted that the form of (2) with the constant 4 in the exponent is one of several numerical transformations that have been proposed to stabilize inverse calculations involving the correlation matrix of sets of the $Y^s(\mathbf{x}_i^s, \mathbf{t}_i)$ (see, for example, Higdon et al., 2008; MacDonald et al., 2015).

Now consider the model for the physical experiment observations. Suppose n^p runs of the manufacturing process are made where, at the i^{th} experimental run, the d control factors are set at $x_{i,1}^p, \dots, x_{i,d}^p$. Let $y^p(\mathbf{x}_1^p), \dots, y^p(\mathbf{x}_{n^p}^p)$ denote the outputs when run at the n^p control factor settings $\mathbf{x}_i^p = (x_{i,1}^p, \dots, x_{i,d}^p)^\top$, $i = 1, \dots, n^p$. This paper assumes the regression model

$$Y^p(\mathbf{x}_i^p) = \mu(\mathbf{x}_i^p) + \epsilon(\mathbf{x}_i^p), \quad (3)$$

where $\mu(\mathbf{x}_i^p)$ denotes the mean response of the manufacturing process at \mathbf{x}_i^p , and $\epsilon(\mathbf{x}_1^p), \dots, \epsilon(\mathbf{x}_{n^p}^p)$ are independent and identically distributed $N(0, \sigma_\epsilon^2)$ measurement errors with unknown variance $\sigma_\epsilon^2 \equiv \lambda_\epsilon^{-1}$. Unlike a linear regression model, no parametric form is assumed for $\mu(\mathbf{x}_i^p)$.

The calibration model used in this paper to link the simulator output and physical process responses was proposed by Kennedy and O’Hagan (2001). A Bayesian calibration analysis based on this model is provided by Higdon et al. (2004) and Higdon et al. (2008). The calibration model postulates that there is a true but unknown value of the calibration parameter \mathbf{t} which is denoted ϕ . The calibration model *does not assume* that the simulator runs, even when made at the true value of the calibration parameter, provide a completely accurate description of the mean of the physical responses. Instead the model allows the simulator to exhibit a non-zero bias which can be estimated. The simulator bias (the “discrepancy”) at control factor setting \mathbf{x}^p is defined to be the difference between the mean of the manufacturing process run at \mathbf{x}^p and the simulator run at control factor setting \mathbf{x}^p and calibration parameter setting ϕ , i.e.,

$$\delta(\mathbf{x}^p) \equiv \mu(\mathbf{x}^p) - y^s(\mathbf{x}^p, \phi). \quad (4)$$

As in Kennedy and O’Hagan (2001), this paper regards $\delta(\mathbf{x}^p)$ as a realization of a Gaussian process $\Delta(\mathbf{x}^p)$ which can be used to predict the bias. The $\Delta(\mathbf{x}^p)$ process is assumed to have mean zero, variance λ_δ^{-1} , and correlation function

$$R_\delta(\mathbf{x}_i^p, \mathbf{x}_j^p) = \prod_{k=1}^d \rho_{\delta,k}^{4(x_{i,k}^p - x_{j,k}^p)^2}. \quad (5)$$

for $i, j = 1, \dots, n^s$. The discrepancy $\Delta(\mathbf{x}^p)$ is assumed to be stochastically independent of $Z(\cdot, \cdot)$ and $\epsilon(\cdot)$. Under assumptions (3)-(5), the mean $\mu(\mathbf{x})$ of the manufacturing process in (3) at control factor setting \mathbf{x} is a realization of the sum of independent Gaussian processes

$$U(\mathbf{x}) \equiv Y^s(\mathbf{x}, \phi) + \Delta(\mathbf{x}). \quad (6)$$

The Bayesian analysis of the calibration model assumes that prior distributions can be stated for all model parameters. These parameters are $\Omega = (\boldsymbol{\lambda}, \boldsymbol{\rho}, \phi, \beta_0)$ where $\boldsymbol{\lambda}$ denotes

the precision vector $(\lambda_Z, \lambda_\delta, \lambda_\epsilon)^\top$, $\boldsymbol{\rho}$ denotes the vector of combined smoothness parameters $(\boldsymbol{\rho}_x^\top, \boldsymbol{\rho}_t^\top, \boldsymbol{\rho}_\delta^\top)^\top$, where $\boldsymbol{\rho}_x^\top = (\rho_{x,1}, \dots, \rho_{x,d})$, $\boldsymbol{\rho}_t^\top = (\rho_{t,1}, \dots, \rho_{t,b})$, and $\boldsymbol{\rho}_\delta^\top = (\rho_{\delta,1}, \dots, \rho_{\delta,d})$, $\boldsymbol{\phi}$ denotes the unknown calibration vector, and β_0 the unknown mean of the GP for the simulator responses. As in Higdon et al. (2004), this paper assumes mutually independent priors for the four groups of unknown parameters. The prior distribution for value of the calibration parameter $\boldsymbol{\phi}$ can often be solicited from subject-matter experts. The posterior distribution of the parameters given the physical and simulator data is formed in the usual way from the likelihood and the prior. Details can be found in Chen et al. (2017).

To predict the mean response of the physical system at a new control factor setting \mathbf{x}_0 , the following approximation is used

$$\hat{\mu}(\mathbf{x}_0) = E\{U(\mathbf{x}_0) \mid \mathcal{Y}\} \approx \frac{1}{N_{mcmc}} \sum_{q=1}^{N_{mcmc}} E\{U(\mathbf{x}_0) \mid \mathcal{Y}, \boldsymbol{\Omega}^q\} \quad (7)$$

where $\mathcal{Y} = (y^p(\mathbf{x}_1^p), \dots, y^p(\mathbf{x}_{n^p}^p), y^s(\mathbf{x}_1^s, \mathbf{t}_1), \dots, y^s(\mathbf{x}_{n^s}^s, \mathbf{t}_{n^s}))^\top$ denotes the vector of the n^p physical observations concatenated with the n^s simulator outputs, and $\boldsymbol{\Omega}^q$, $q = 1, \dots, N_{mcmc}$, is the q^{th} draw from the posterior distribution of $\boldsymbol{\Omega}$ given the data \mathcal{Y} . Chen et al. (2017) give a formula for the expectation in the right hand side of (7).

The calibrated predictor (7) was constructed separately for each output in the injection molding study of Section 2 using the $n^p = 19$ manufacturing process physical observations and the outputs from the $n^s = 35$ simulator runs. As seen in Figure 4, the simulator is biased for all three outputs $L1, L2, L3$. Figure 5 shows that, when computed for the same 19 control factor settings as used in the manufacturing experiment, the calibrated predictions of the means of $L1, L2$, and $L3$ successfully adjust for the simulator bias.

Section 4 describes the Pareto Front and Pareto Set solution to multi-objective minimization problems and a sequential design methodology for finding the Pareto solution to a given problem. Section 5 applies this methodology to optimize the injection molding process.

4 Pareto Minimization

Because conflicting objectives need not have a common minimizer, this section describes the *Pareto* approach for identifying a set of compromise “minimizing” solutions. As mentioned in

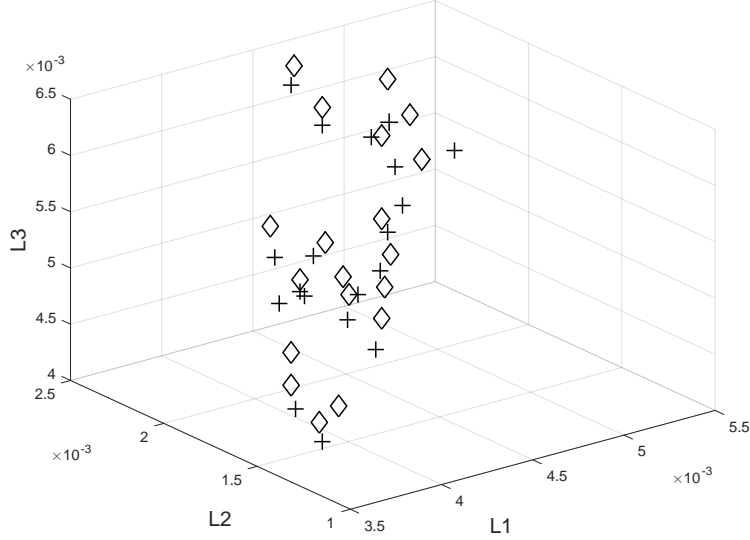


Figure 5: Scatter plot of the 19 relative shrinkages ($L1, L2, L3$) from the manufacturing process (shown as “+”) and 19 calibrated predictions (shown as “◇”) made at same control factor inputs used in the manufacturing process.

Section 1, the collection of control factor settings for these compromise solutions are termed the *Pareto Set* for the problem and the corresponding outputs constitute the *Pareto Front*. Pareto minimization of computationally expensive simulator output based on surrogate predictors is an active research area, e.g. Wilson et al. (2001), Emmerich et al. (2006), Keane (2006), Kim and deWeck (2006), Binois et al. (2014), Picheny (2015), Svenson and Santner (2016). However, none of these papers treat situations where both physical observations and simulator data are available.

4.1 Definition of Pareto Front and Set

Let $y_\ell(\mathbf{x})$, $\ell = 1, \dots, m$, denote m functions defined on a common domain \mathcal{X} . Pareto minimization seeks to identify the set of all $\mathbf{x}^* \in \mathcal{X}$ that cannot be dominated by any alternative input \mathbf{x}^a in the sense that $y_\ell(\mathbf{x}^a) \leq y_\ell(\mathbf{x}^*)$ for all $\ell = 1, \dots, m$ with $<$ for some ℓ . The Pareto Set, denoted by $\mathcal{P}_\mathcal{X}$, is then defined as the collection of all Pareto minimizing \mathbf{x}^* in \mathcal{X} . The Pareto Front, denoted by $\mathcal{P}_\mathcal{Y}$, is the set of $\mathbf{y}(\mathbf{x}^*) = (y_1(\mathbf{x}^*), \dots, y_m(\mathbf{x}^*))$ corresponding to all $\mathbf{x}^* \in \mathcal{P}_\mathcal{X}$. In words, a $\mathbf{y}(\mathbf{x})$ that is *not on* the Pareto Front can be minimized further by at least one other output *on* the Pareto Front. For example consider the collection of two-dimensional

outputs ($m = 2$) shown as filled and empty stars in Figure 6. The two points denoted by filled stars comprise the Pareto Front from this finite universe of points.

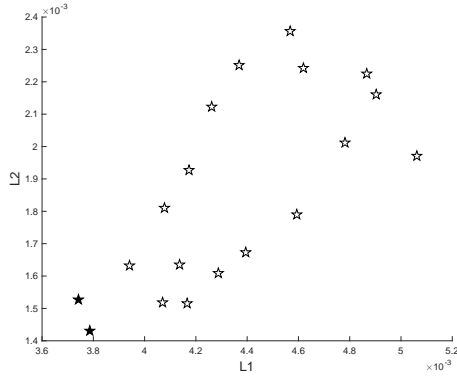


Figure 6: The two filled stars are the Pareto Front in the set consisting of the all stars (filled or hollow). Each hollow star is dominated by one or both filled stars.

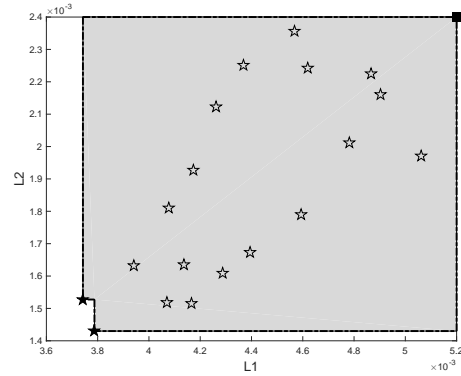


Figure 7: The gray area is the hypervolume indicator for the PF of Figure 6 when the reference point, \mathbf{r} , is the filled square.

4.2 Hypervolume Indicator of Pareto Minimization

An important issue in multi-objective optimization is the evaluation of the quality of an estimated Pareto Front, especially when $m \geq 4$. The *hypervolume indicator* is a popular real-valued measure that can be used to compare competing approximate Pareto Fronts. While the hypervolume indicator can be defined analytically, for this paper we show only the following graphical interpretation. Let $\mathbf{r} = (r_1, \dots, r_m)$ be a fixed point whose components are upper bounds for each output function $y_1(\mathbf{x}), \dots, y_m(\mathbf{x})$, i.e., $r_\ell \geq y_\ell(\mathbf{x})$ for all $\ell = 1, \dots, m$ and all $\mathbf{x} \in \mathcal{X}$. The hypervolume indicator of an approximate Pareto Front \mathcal{A} is the volume of the set of points in output space that both dominate \mathbf{r} and are dominated by at least one point of \mathcal{A} . The shaded area in Figure 7 is the hypervolume indicator for the Pareto Front of Figure 6 (see Zitzler and Thiele, 1999).

4.3 A Sequential Pareto Minimization Procedure

This paper shows the effectiveness of the sequential design and analysis procedure of Chen et al. (2017) for Pareto minimization to optimize the manufacturing process described in Section 2. It also illustrates the procedure's efficiency using two analytic examples.

In brief, the Chen et al. (2017) sequential design procedure is based on Bayesian calibrated predictors and, in the spirit of the Jones et al. (1998) EGO global optimization procedure, uses an expected improvement function, called the `minimax fitness function` (mMFF), to efficiently guide the search for the new settings of the control input variables close to the Pareto Set.

We describe the mMFF in the simplest possible setting of a known mean function $\boldsymbol{\mu}(\mathbf{x}) = (\mu_1(\mathbf{x}), \dots, \mu_m(\mathbf{x}))$ of an m -output physical process where $\mathbf{x} \in \mathcal{X}$ and \mathcal{X} is the space of possible control factor settings. The mMFF of $\boldsymbol{\mu}(\cdot)$ at $\mathbf{x} \in \mathcal{X}$ is defined to be

$$I_{\mathcal{F}}(\boldsymbol{\mu}(\mathbf{x})) \equiv \min_{\mathbf{x}_i \in \mathcal{P}_{\mathcal{X}}^{n^p+n^s}} \max_{\ell=1, \dots, m} (\mu_{\ell}(\mathbf{x}_i) - \mu_{\ell}(\mathbf{x})) \times \mathbb{1}_{\left[\min_{\mathbf{x}_i \in \mathcal{P}_{\mathcal{X}}^{n^p+n^s}} \max_{\ell=1, \dots, m} (\mu_{\ell}(\mathbf{x}_i) - \mu_{\ell}(\mathbf{x})) > 0 \right]}, \quad (8)$$

where $\mathbb{1}_E$ is an indicator function having value 1 or 0 as the event E is true or not and $\mathcal{P}_{\mathcal{X}}^{n^p+n^s}$ is the Pareto Set among the $n^p + n^s$ control factor settings based on the n^p physical observations and the control portions of the n^s simulation runs. The mMFF is non-negative, and it can be shown that positive values can be viewed as the “improvements” to the current Pareto Front. The indicator function $I_{\mathcal{F}}(\boldsymbol{\mu}(\mathbf{x})) = 0$ if and only if \mathbf{x} is dominated by some \mathbf{x}^* in $\mathcal{P}_{\mathcal{X}}^n$ while $I_{\mathcal{F}}(\boldsymbol{\mu}(\mathbf{x})) > 0$ if and only if \mathbf{x} is not dominated by any \mathbf{x}^* in $\mathcal{P}_{\mathcal{X}}^n$. Thus, an intuitive choice is to select the next control factor setting as that $\mathbf{x} \in \mathcal{X}$ which maximizes $I_{\mathcal{F}}(\boldsymbol{\mu}(\mathbf{x}))$.

Example 4.1 This example illustrates how the mMFF (8) can be used to construct sequential designs for Pareto minimization. Consider the *MOP2* function $\boldsymbol{\mu}(\mathbf{x}) = (\mu_1(\mathbf{x}), \mu_2(\mathbf{x}))$ where

$$\mu_{\ell}(\mathbf{x}) = 1 - \exp \left\{ - \sum_{k=1}^2 \left(x_k + \frac{(-1)^{\ell}}{\sqrt{2}} \right)^2 \right\}, \quad \text{for } \ell = 1, 2. \quad (9)$$

which has $m = 2$ outputs and $d = 2$ inputs with the input space $\mathcal{X} \in [-2, 2]^2$ (Fonseca and Fleming, 1995). The function $\mu_1(\mathbf{x})$ has global minimum equal to zero when $\mathbf{x} = (1/\sqrt{2}, 1/\sqrt{2})$ while $\mu_2(\mathbf{x})$ has global minimum equal to zero when $\mathbf{x} = (-1/\sqrt{2}, -1/\sqrt{2})$. It can be shown that the Pareto Set of $\boldsymbol{\mu}(\mathbf{x})$ is

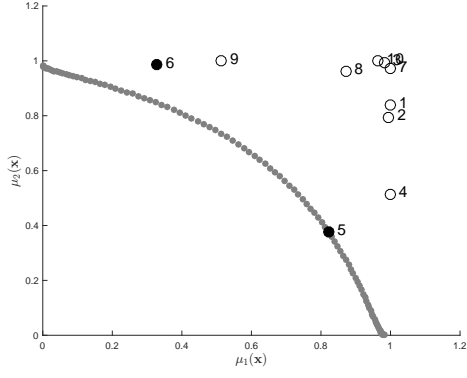
$$\mathcal{P}_{\mathcal{X}} = \left\{ \mathbf{x} : x_1 = x_2 \text{ and } -\frac{1}{\sqrt{2}} \leq x_1, x_2 \leq \frac{1}{\sqrt{2}} \right\}. \quad (10)$$

Points on the Pareto Front are plotted as the grey curve in Figures 8(a), (c) and (d).

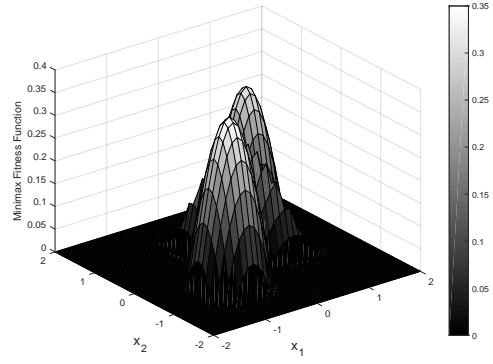
Suppose that, initially, 10 observations of $\boldsymbol{\mu}(\mathbf{x})$ have been made at a set of space-filling control factor settings \mathbf{x} in $[-2, 2]^2$ and that these $\boldsymbol{\mu}(\mathbf{x})$ are plotted as circles (some filled and others hollow) in Figure 8(a). Two of these 10 points, plotted as filled circles, are not dominated by any other points and form the current approximate Pareto Front. The plot of $I_{\mathcal{F}}(\boldsymbol{\mu}(\mathbf{x}))$ corresponding to these 10 points is shown in Figure 8(b), where $\mathcal{P}_{\mathcal{X}}^{n^p+n^s}$ is taken to be the 2-point Pareto Set identified above. The maximum value of $I_{\mathcal{F}}(\boldsymbol{\mu}(\mathbf{x}))$ occurs at $\mathbf{x} = (-0.7, -0.7)$. The point $\boldsymbol{\mu}(-0.7, -0.7)$ is added to Figure 8(a) and is labeled “11” in Figure 8(c), where the approximate Pareto Front (filled circles) has been updated and now contains 3 points. The maximum of $I_{\mathcal{F}}(\boldsymbol{\mu}(\mathbf{x}))$ based on the 11 observations (not shown) occurs at $\mathbf{x} = (0.7, 0.7)$. Figure 8(d) adds $\boldsymbol{\mu}(0.7, 0.7)$ to the plot of the previous points, and updates the approximate Pareto Front again. The effectiveness of the mMFF function in guiding the sequential search in this idealized, known output setting is seen from the fact that the two added points give $\boldsymbol{\mu}(\mathbf{x})$ on, or close to, the true Pareto Front. \square

Example 4.1 illustrates the use of the mMFF function in the selection of control factor settings at which to run the simulation. This forms the basis of a sequential design procedure for finding the Pareto Front and Set of a *known* $\boldsymbol{\mu}(\mathbf{x})$. As in many other applications, the optimization of the injection molding process described in this paper is to find the Pareto Front and Set for the vector of *unknown* means of a manufacturing process. The unknown mean vector is denoted $\boldsymbol{\mu}(\mathbf{x})$ in model (3). This paper illustrates the sequential procedure for solving this problem where the data are a *fixed* number of physical observations and an initial set of simulator runs to which a researcher can add simulator runs. Chen et al. (2017) also provides theoretical procedures for the case of supplemental physical observations, and for the choice of adding either simulator runs or physical observations to an initial set of runs.

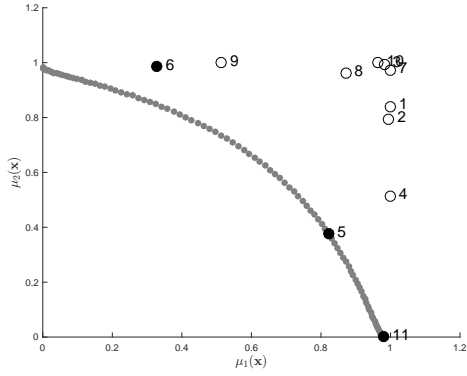
Assume that *initial data* consists of observations from the physical system based on a fixed control factor design $\mathbf{X}_{n^p} = (\mathbf{x}_1^p, \dots, \mathbf{x}_{n^p}^p)^\top$ and runs of the simulator based on an initial control factor design $\mathbf{X}_{n^s} = (\mathbf{x}_1^s, \dots, \mathbf{x}_{n^s}^s)^\top$ with associated calibration variables. The sequential approach estimates $\boldsymbol{\mu}(\mathbf{x})$ using the calibrated predictors and replaces the exact MmFF calculations by its expectation. These features are seen in the sequential Pareto minimization procedure below.



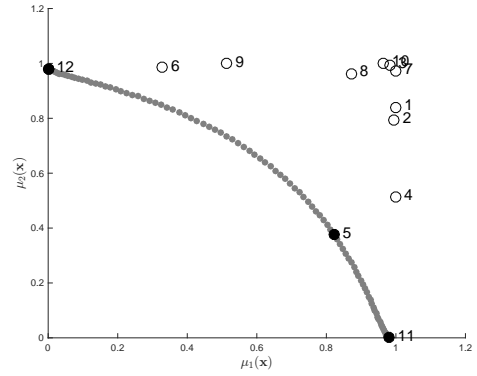
(a) Ten initial $\mu(\mathbf{x})$ observations for the MOP2 function. The $\mu(\mathbf{x})$ shown as filled circles are not dominated.



(b) The $I_{\mathcal{F}}(\mu(\mathbf{x}))$ function based on the 10 observations in Panel (a).



(c) The 10 + 1 $\mu(\mathbf{x})$ observations where the added $\mu(\mathbf{x})$ is labeled “11”; $\mu(\mathbf{x})$ shown as filled circles are not dominated.



(d) 10 + 2 $\mu(\mathbf{x})$ observations where the added $\mu(\mathbf{x})$ is labeled “12”; $\mu(\mathbf{x})$ shown as filled circles are not dominated.

Figure 8: For the MOP2 function, sequentially added points obtained by using an update formula based on the mMFF.

Step 1 Fit the Bayesian calibrated model (6) to each output function ($\ell = 1, \dots, m$) independently.

Step 2 Predict the vector of physical process means at \mathbf{X}_{n^p} and \mathbf{X}_{n^s} using the Bayesian calibrated predictor (7).

Step 3 Based on the predictions in Step 2, determine the current approximate Pareto Set $\mathcal{P}_x^{n^p+n^s}$ among the observed $n^p + n^s$ control factor settings.

Step 4 Select the next control factor setting \mathbf{x}^* to maximize the posterior expected minimax

fitness function given the current data \mathcal{Y}^c , i.e.,

$$\mathbf{x}^* \in \operatorname{argmax}_{\mathbf{x} \in \mathcal{X}} E\{I_{\mathcal{F}}(\mathbf{U}(\mathbf{x})) \mid \mathcal{Y}^c\}, \quad (11)$$

where $I_{\mathcal{F}}(\cdot)$ is defined in (8). Chen et al. (2017) derived a nearly closed-form expression for (11) and calculated it using Markov chain Monte Carlo.

Step 5 Construct the vector of calibration parameter settings \mathbf{t}^* to minimize the sum of the posterior mean squared prediction errors (MSPE) given the current data \mathcal{Y}^c and the \mathbf{x}^* selected from (11), i.e.

$$\mathbf{t}^* \in \operatorname{argmin}_{\mathbf{t} \in \mathcal{T}} \sum_{\ell=1}^m E\{[\hat{U}_{\ell}(\mathbf{x}^*) - U_{\ell}(\mathbf{x}^*)]^2 \mid \mathcal{Y}^c\}, \quad (12)$$

where $\hat{U}_{\ell}(\mathbf{x}^*)$ is the mean of $[U_{\ell}(\mathbf{x}^*) \mid \mathcal{Y}_{\ell}, Y_{\ell}^s(\mathbf{x}^*, \mathbf{t})]$. The mean $\hat{U}_{\ell}(\mathbf{x}^*)$ is a function of the augmented data $Y_{\ell}^s(\mathbf{x}^*, \mathbf{t})$. The expectation of $\hat{U}_{\ell}(\mathbf{x}^*)$ given \mathcal{Y}^c can be expressed in closed form; the details can be found in Chen et al. (2017).

Step 6 Run the simulator to evaluate $\mathbf{y}^s(\mathbf{x}^*, \mathbf{t}^*)$ and add it to the vector of simulator outputs.

Step 7 Set $n^s = n^s + 1$, and go to Step 1 unless the sampling budget has been exhausted or a stopping criterion has been met. (A possibility is to stop when the posterior expected minimax fitness function fails to increase by more than a given value ϵ for a specified number of consecutive iterations).

5 Optimizing the Injection Molding Process

The goal of the injection molding experiment described in Section 2 was to obtain the Pareto Set of optimizing conditions for the four control variables (melt temperature, packing time, packing pressure, and cooling time) corresponding to the Pareto Front of the relative shrinkages $L1, L2, L3$ in the lengths of the parts P1, P2, P3, where all three shrinkages were to be minimized.

In Table 4, the second and third groups of columns list the points in the approximate Pareto Set and Pareto Front for three approximating methods. The first of these methods uses the 19 physical observations in Table 1 and the 35 runs of Moldex3D listed in Table 3

Method	Approximate Pareto Set				Approximate Pareto Front			Hypervol. Indicator	
	T_{melt}	t_{pack}	P_{pack}	t_{cool}	$10^3 \times L1$	$10^3 \times L2$	$10^3 \times L3$		
19 physical obs	216	28	42.67	36	3.7802	1.5535	4.4033	◆	0.0365
+ 35 LHD runs	216	28	42.67	50	3.8477	1.4810	4.1759		
19 physical obs	200	24.5	42	43	2.9884	0.6481	3.3808	■	0.1427
+ 35 LHD runs	208	24.5	42	43	3.0275	0.7843	3.3645		
+ 625 grid preds	200	28	42	43	2.9054	0.6495	3.4485		
	208	28	42	43	2.9537	0.7918	3.4193		
19 physical obs	203.42	24.17	43.25	43.07	2.8094	0.8206	3.2010	●	0.1457
+ 25 space filling	203.56	26.27	42.09	43.66	2.7658	0.9016	3.3088		
+ 10 sequen. runs	202.12	22.50	43.80	43.96	2.9032	0.8061	3.2224		

Table 4: Approximate Pareto Set and Pareto Front, and value of the hypervolume indicator from each of three methods based on three designs. Each design consisted of the same 19 physical control factor settings, together with the following designs for the simulator control factor settings (i) a 35-run augmented maximin LHD (with predictions at the control factor settings of these 54 runs), (ii) a 35-run augmented maximin LHD (with predictions at the control factors with values on a grid of 625 control factor runs), (iii) a 25-run space filling design from among the settings of the 35-run augmented maximin LHD, plus 10 additional sequentially selected settings (with predictions at the control factor settings for these 54 runs). The plotting symbols shown in the penultimate column are used in Figures 9–10.

to build a calibrated predictor (Section 3) and predicts the true mean values of $L1, L2, L3$ at these $19+35=54$ control factor settings. Two of the predicted means form the Pareto Front among this set of 54 predictions and these are listed in the first two rows in Table 4, (and also shown as the filled diamonds in Figure 10). The control factor settings in the approximate Pareto Set differ from each other only in the value of cooling time. The longer cooling time is better for minimizing shrinkage in the lengths of P2 and P3, while the shorter time is slightly better for that of P1.

The second method was based on the following strategy to improve the approximation to the Pareto Front. Using the same $19+35$ runs as for the first method, we predicted the true mean values of $L1, L2, L3$ over a grid of $5^4 = 625$ input points, where there were 5 equally spaced values over the ranges of each of the four control factors. Projections of the grid points and the predictions over the grid are shown as open diamonds in Figure 9. The approximate Pareto Set from among the 625 points consists of four control factor settings. These and the points on the approximate Pareto Front are listed as the second group of four rows in Table 4 and are denoted by filled squares in Figures 9 and 10. The approximate Pareto Front based on the 625 predictions is a great improvement over that obtained from predictions at the 54 design points —the relative shrinkages in length of all three parts are smaller, and

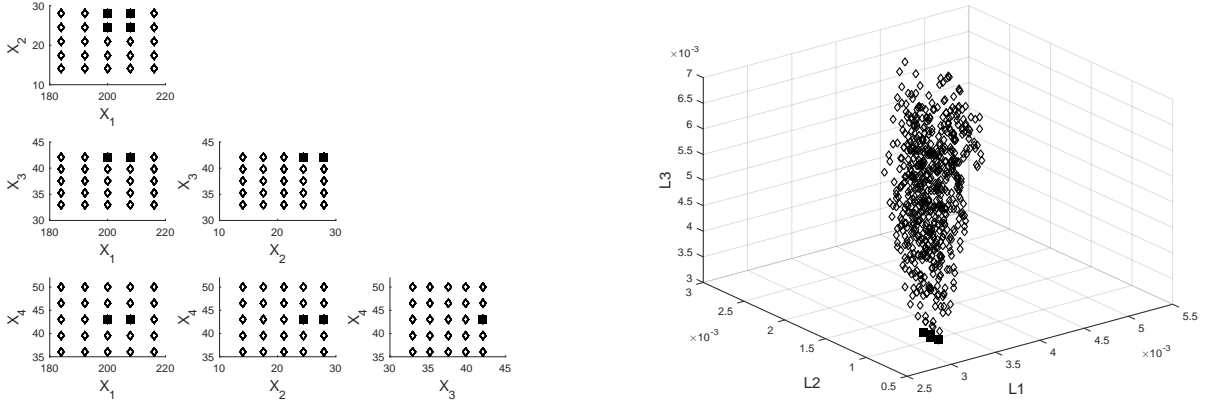


Figure 9: Left hand side shows the two-dimensional projections of the 625 grid points at which calibrated predictions were made. The variables x_1, \dots, x_4 are, respectively, the control factors T_{melt} , t_{pack} , P_{pack} , t_{cool} . The predictions of the means of the relative shrinkages $L1, L2, L3$ are shown on right hand side as open diamonds. The Pareto Front and Set are indicated by filled squares.

the hypervolume indicator increased three-fold (calculated relative to the same reference point $\mathbf{r} = (.0051, .0024, .0063)$). Also, the optimizing settings for the melt temperature are lower.

The third method of approximating the Pareto Set and Front uses the sequential approach of Chen et al. (2017) described in Section 4.3. We show that this method provides further improvements to the hypervolume indicator of the approximate Pareto Front (relative to the same \mathbf{r}). Using the same 19 physical observations as before (Table 1), a space-filling initial simulator design was selected which consisted of the subset of 25 points from among the 35 points of Table 3 which resulted in the largest value of the minimum interpoint distance. The omitted points are indicated by “-”. The sequential procedure was then used to provide an additional 10 input points. The approximate Pareto Set and Front obtained from this method are shown in the last three rows of Table 4 and plotted as filled circles in Figure 10. The hypervolume indicator is slightly larger than that obtained from the second method (0.1457 as compared with 0.1427) with relative shrinkages $L1$ and $L3$ being a little smaller (although $L2$ is a little larger).

Figure 11 illustrates the steps in the sequential method. First, the calibrated predictions of the means of $L1, L2, L3$ at the 19 + 25 initial design settings of the four control factors are plotted in Figure 11(a). The filled diamond is the approximate Pareto Front at this stage. The next setting of the four control factors and three calibration parameters was obtained using the MATLAB program `MultiOpt.m` which was developed by Chen (2016) for the sequential

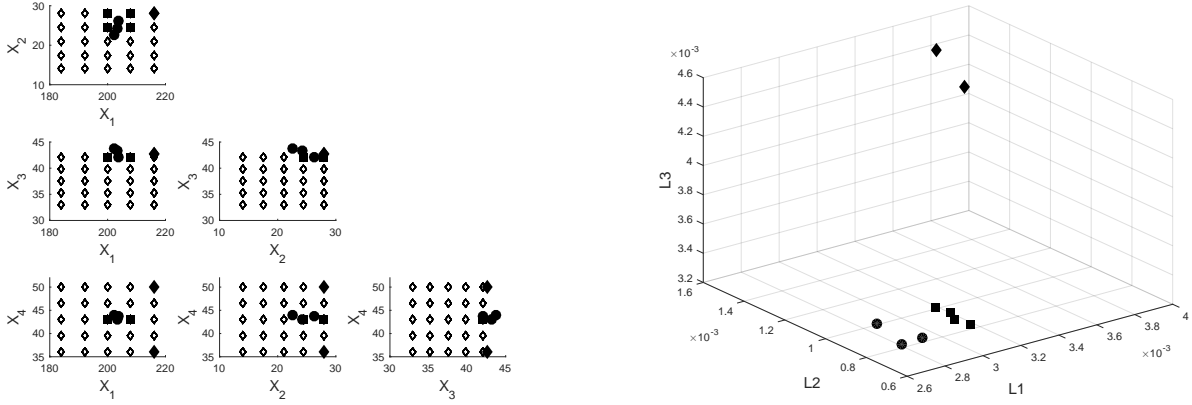


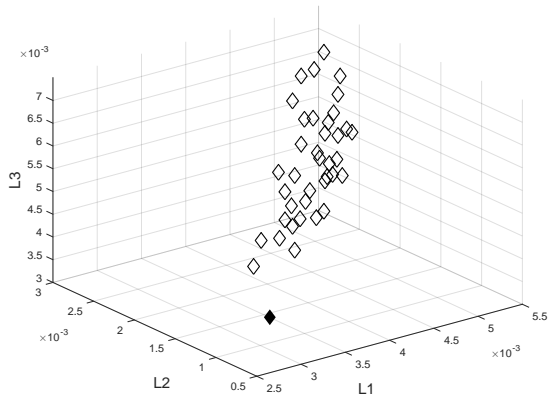
Figure 10: Approximate Pareto Front (right panel) and corresponding approximate Pareto Set (left panel) using the 19 physical observations and (i) the fixed 35-run maximin LHD (two filled diamonds), (ii) the 35-run maximin LHD and predictions over a 5^4 grid (four filled squares), and (iii) the sequential approach with a 25-run space-filling design and 10 added points (three filled circles).

procedure of Section 4.3. The filled circle in Figure 11(b) is the calibrated prediction at the first sequentially added point. Since this minimizes all three output variables, it becomes the approximate Pareto Front and the previous point is removed.

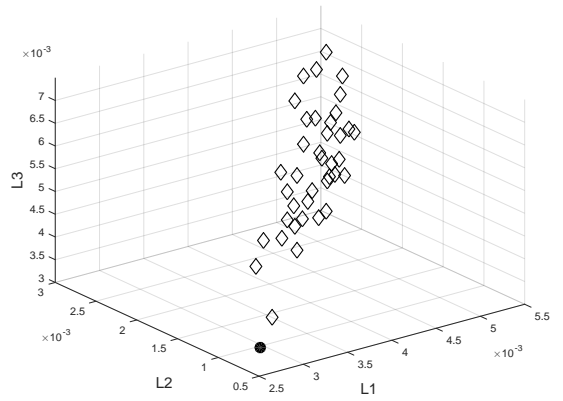
Figure 11(c) shows, as open and filled circles, the calibrated predictions for each of the 10 additional points identified by the sequential design. Notice that every one of these gives smaller predicted L_1, L_2, L_3 than the predictions at almost all of the points in the initial design, and the three filled circles indicate the final approximate Pareto Front. Figure 11(d) shows the values of the hypervolume indicator of the approximate Pareto Front at each stage, calculated relative to reference point $\mathbf{r} = (.0051, .0024, .0063)$. There appears to be a large improvement in the approximation when the first sequential point is added to the 25-run space-filling design. After that, there is a small improvement in the approximation at stages 5 and 10.

6 Properties of the Sequential Pareto Minimization Procedure

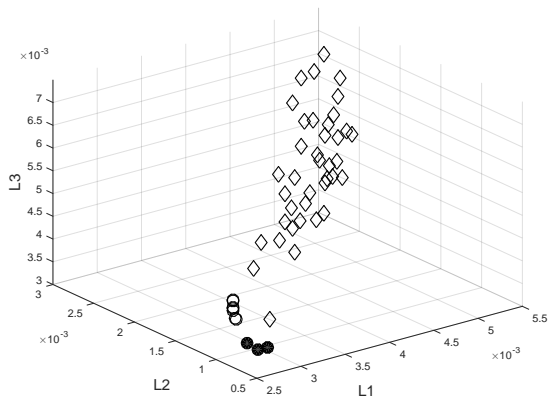
In this section, the performance of the sequential Pareto minimization methodology of Chen et al. (2017), described in Section 4.3, is examined for two known functions in terms of its ability to predict points in, or close to, the true Pareto Set.



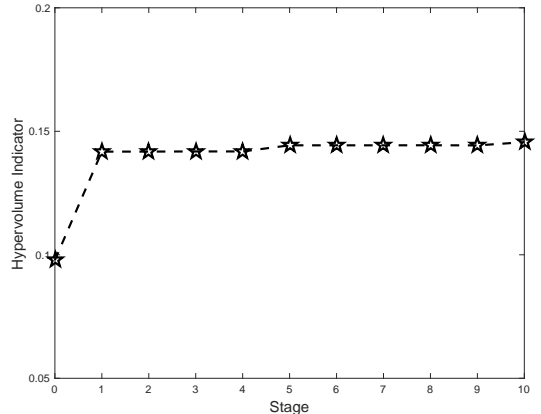
(a)



(b)



(c)



(d)

Figure 11: Calibrated predictions of the means of $L1, L2, L3$ for the sequential procedure at (a) the 19 + 25 initial design points, (b) after adding the first sequential point, (c) after adding 10 sequential points. In each case the filled symbols indicate the current approximate Pareto Front. The hypervolume indicator at each stage of the sequential procedure is shown in (d).

6.1 Accuracy of the Approximate Pareto Set and Front

6.1.1 MOP2 Function

The *MOP2* function was described in Section 4.3. We assume that the two output functions in (9) form the means of a “physical process” at control factor setting $\mathbf{x} = (x_1, x_2)$ with $(x_1, x_2) \in [-2, 2]^2$. We generate the individual physical observations as realizations of

$$Y_\ell^p(\mathbf{x}) = \mu_\ell(\mathbf{x}) + \epsilon_\ell, \ell = 1, 2, \quad (13)$$

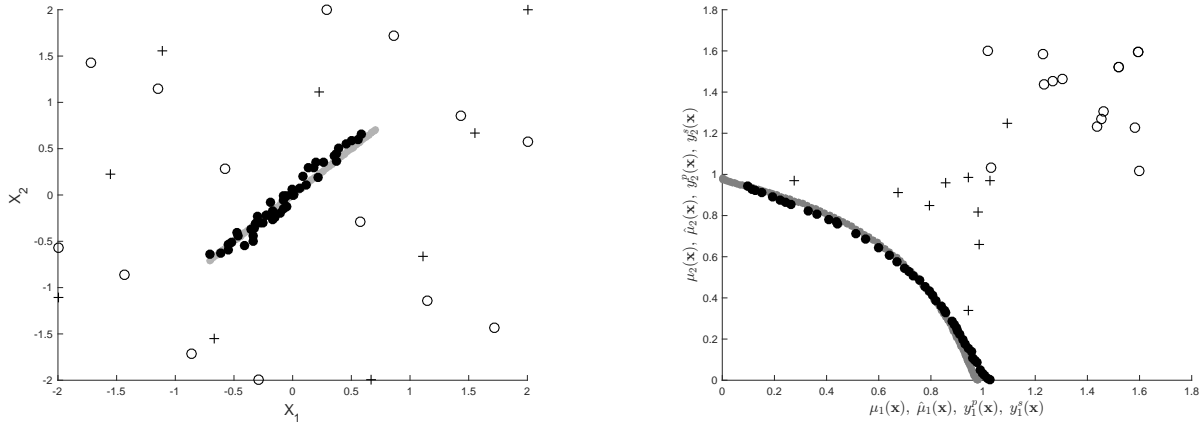


Figure 12: The *MOP2* example: Left Panel: (x_1, x_2) for the initial physical control factor settings (“+”), simulator control factor settings (“o”), the true Pareto Set (grey circles), and the approximate (estimated) Pareto Set after adding 50 additional simulator runs identified by the sequential procedure (filled circles). Right Panel: The corresponding initial observations (“+” and “o”), true Pareto Front (grey circles), and approximate Pareto Front (filled circles).

with means as in (9) and with independent error variables $\epsilon_\ell \sim N(0, 0.1^2)$. The simulator outputs are generated as

$$y_\ell^s(\mathbf{x}, t) = \mu_\ell(\mathbf{x}) + |t|/10 + 0.4 + 0.3(x_1 x_2 / 4)^2, \quad (14)$$

for $\ell \in \{1, 2\}$; there is a single calibration input $t \in [-2, 2]$ for each simulator.

The initial designs for the physical and computer experiments are each chosen to be a maximin LHD with 5 runs per input dimension. Therefore the LHD for the physical process is 10×2 and that for the simulator is 15×3 . Both designs are obtained from the website of van Dam et al. (2013). Assuming a fixed physical experiment budget, the simulator experiment design is augmented using the sequential update procedure of Section 4.3.

The left panel of Figure 12 plots the design points for the physical experiment (“+”) and those for the initial simulator experiment (“o”). The true Pareto Set of the *MOP2* function, equation (10), is shown as the grey line; the approximate (estimated) Pareto Set after applying the sequential update formula and selecting 50 additional simulator input settings is shown as filled circles. The right panel of Figure 12 plots the corresponding physical observations (“+”) generated from (13), the initial observations (“o”) generated from the simulator in (14), the Pareto Front of the *MOP2* function (grey circles), and the predicted means (filled circles) of the *MOP2* function at the control factor settings of the approximate Pareto Set in the left panel.

This example indicates that the sequential procedure has led to very good approximations to the true Pareto Set and Pareto Front using the inputs identified by the sequential design.

6.1.2 *DTLZ2* Function

Introduced by Deb et al. (2005), the version of the *DTLZ2* functions used here has $d = 3$ inputs and $m = 3$ outputs defined as

$$\begin{aligned}\mu_1(\mathbf{x}) &= (1 + g(\mathbf{x})) \cos\left(\frac{\pi x_1}{2}\right) \cos\left(\frac{\pi x_2}{2}\right), \\ \mu_2(\mathbf{x}) &= (1 + g(\mathbf{x})) \cos\left(\frac{\pi x_1}{2}\right) \sin\left(\frac{\pi x_2}{2}\right), \\ \mu_3(\mathbf{x}) &= (1 + g(\mathbf{x})) \sin\left(\frac{\pi x_1}{2}\right),\end{aligned}\tag{15}$$

where $g(\mathbf{x}) = (x_3 - 0.5)^2$ is independent of x_1 and x_2 . Similar to the example of the *MOP2* function in Section 6.1.1, the physical observations of *DTLZ2* functions are generated as the means (15) plus independent errors $\epsilon_\ell \sim N(0, 0.1^2)$. The simulator outputs $y^s(\mathbf{x}, t)$ are taken to have a single calibration input, $t \in [0, 1]$. Each $y_\ell^s(\mathbf{x}, t)$ was obtained by adding $(t - 0.5)^2$ to a modified $\mu_\ell(\mathbf{x})$ in which $g(\mathbf{x})$ is replaced by $g^s(\mathbf{x}) = x_1 + (x_3 - 0.5)^2$, leading to a complicated bias function, even under the “true” value $\phi = 0.5$ of the calibration parameter.

As for the *MOP2* example, the initial designs were chosen to be maximin LHDs with 5 runs per input dimension, so there are 15 design points for the physical experiment and, initially, 20 design points for the computer simulator. These initial design points are augmented by sequentially adding 50 new simulator design points using the procedure of Section 4.3.

The true Pareto Set is $\mathcal{P}_\mathcal{X} = \{\mathbf{x} : x_3 = 0.5\}$; the left panel of Figure 13 plots 1,000 points from $\mathcal{P}_\mathcal{X}$ in grey. The right hand panel shows the corresponding outputs on the true Pareto Front, again in grey. Figure 13 also displays the design points for the initial observations (“+” for the physical observations and “o” for the simulator outputs) and the estimated Pareto Set and Front (filled circles) after sequentially selecting the 50 additional simulator settings. The estimated Pareto Set is a good approximation to the true Pareto Set with most points in Figure 13 close to the plane with $x_3 = 0.5$ and spread out over the entire (x_1, x_2) space; only a few points have x_3 away from 0.5.

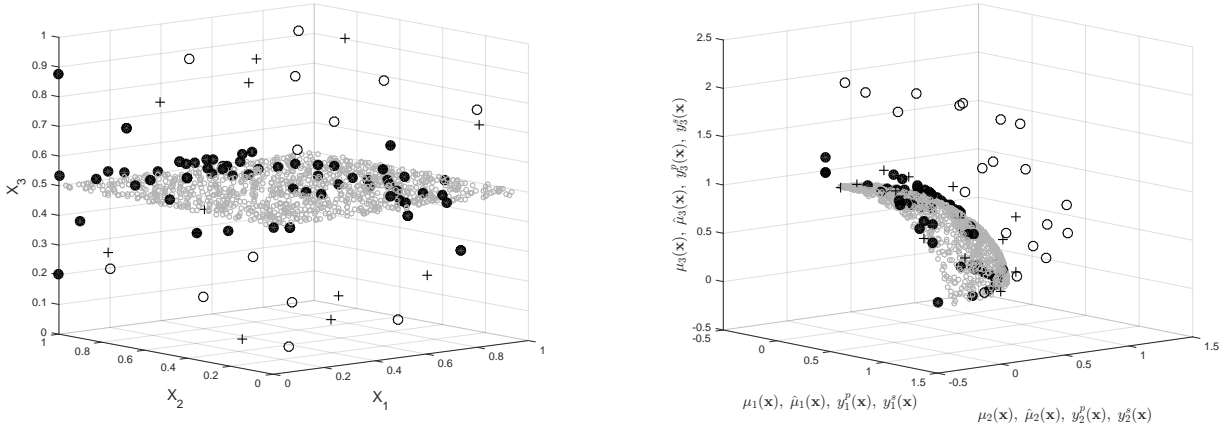


Figure 13: The *DTLZ2* example: Left Panel: (x_1, x_2, x_3) for physical control factor settings (“+”), initial simulator control factor settings (“o”), the true Pareto Set (grey circles), and the approximate (estimated) Pareto Set after adding 50 additional simulator runs identified by the sequential procedure (filled circles). Right Panel: The corresponding initial observations (“+” and “o”), true Pareto Front (grey circles), and approximate Pareto Front (filled circles).

6.1.3 The Hypervolume Indicator

The value of the hypervolume indicator was calculated after each additional point was sequentially added to the initial design. These values are plotted for the 50 added points in Figure 14 for both the *MOP2* function and the *DTLZ2* function. In each case the value of the hypervolume indicator of the true Pareto Front is shown as a horizontal line. The speed of approach of the hypervolume indicator of the approximate PFs identified by the sequential procedure to the true value depends upon the sizes of the bias and the experimental error. Figure 14 shows that the sequential procedure improves the approximation to the true PF at most stages, and that approach is fairly rapid for the two-dimensional *MOP2* function, and a little slower for the *DTLZ2* function.

6.2 Allocation of Simulator Runs

Now consider the allocation problem in which the number of observations of the physical process and the number of runs of the simulator are each fixed by the budget. The division, however, of the number of simulator runs between the initial design and the sequential stages still needs to be determined. We return to the *MOP2* function to address this problem. For example, suppose that the maximum number of 10 physical observations has been made and that a total number n^s has been allocated for simulator runs. In this empirical study, we take

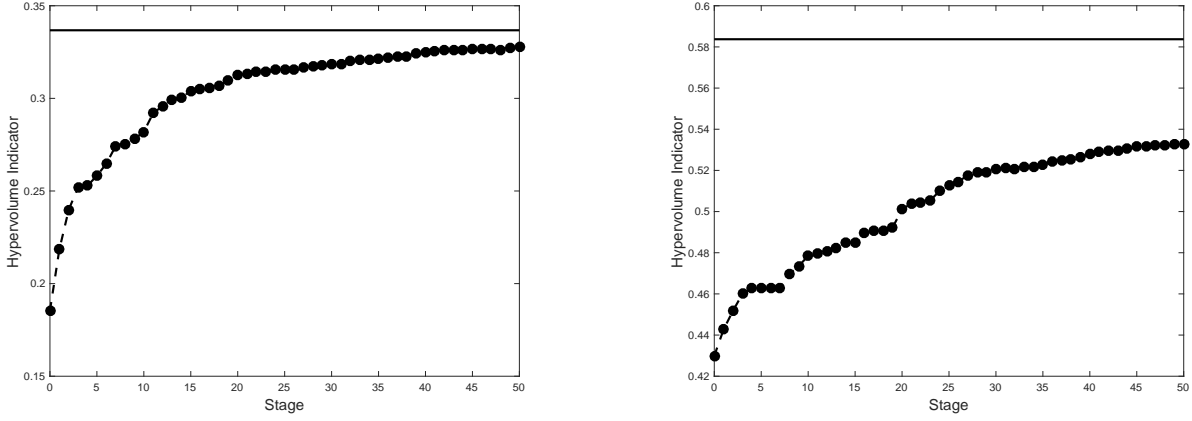


Figure 14: The hypervolume indicator (HI) of the approximate Pareto Front at each stage of the sequential Pareto minimization procedure adding 50 simulator runs. Left panel: the MOP2 function (with reference point $\mathbf{r} = (1.0, 1.0)$); Right panel: and DTLZ2 function (with reference point $\mathbf{r} = (1.1, 1.1, 1.1)$) The horizontal line shows the hypervolume indicator of the true Pareto Front.

$n^s = 15$ and allocate n_I^s runs (in the range $4, 5, \dots, 15$) to the initial simulator design, with the remaining $15 - n_I^s$ runs to be added sequentially. All initial designs are selected as maximin LHDs obtained from the website of van Dam et al. (2013). We investigate the effect of choice of n_I^s on (i) the hypervolume indicator of the final identified approximate PF, and (ii) the mean squared prediction error of outputs on the true Pareto Front.

The hypervolume indicator of the approximate Pareto Front identified using the calibrated predictor based on the n_I^s initial and the $15 - n_I^s$ sequentially added simulator outputs, and the 10 physical observations, is shown in Figure 15 for $n_I^s = 4, 5, \dots, 15$. The results suggest that the size, n_I^s , of the initial simulator design should be about 25% to 75% of the total number of runs, and perhaps towards the lower end of this range.

To examine the accuracy of prediction of the Pareto Front, the mean squared prediction error,

$$\text{MSPE} = \frac{\sum_{\ell=1}^2 \sum_{i=1}^{100} (\hat{\mu}_{\ell}(\mathbf{x}_i) - \mu_{\ell}(\mathbf{x}_i))^2}{100}, \quad (16)$$

is used where $\hat{\mu}_{\ell}(\mathbf{x}_i)$ is the calibrated prediction of the true mean $\mu_{\ell}(\mathbf{x})$ of the ℓ th function. Figure 16 shows the MSPE calculated over 100 points on the true Pareto Front of the MOP2 function, when the initial simulator design has $n_I^s (= 4, \dots, 15)$ points, and when the physical observation error and the bias functions are as described in Section 6.1.1. Combining the

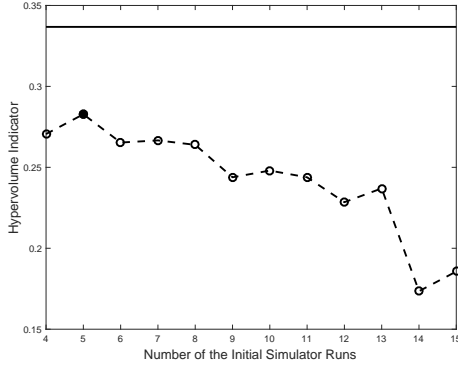


Figure 15: The hypervolume indicators of the approximate Pareto Fronts identified for the *MOP2* function of Section 6.1.1 using the 10 physical observations, the n_I^s initial and the $15 - n_I^s$ sequentially added simulator outputs, for each of $n_I^s = 4, 5, \dots, 15$.

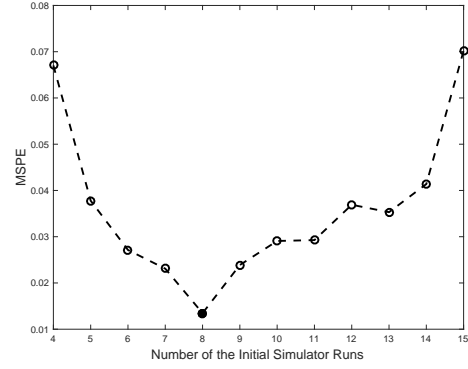


Figure 16: The MSPEs of the prediction of the 100 points on the true Pareto Front for the *MOP2* function of Section 6.1.1 for each of $n_I^s = 4, 5, \dots, 15$.

results of Figures 15 and 16, the initial simulator design for this example should be in the range $n_I^s = 6, \dots, 11$; i.e. around 40% to 75% of the $n^s = 15$ total number of runs.

It is not surprising that the above recommendation is in the center of the range of possible values for n_I^s since a small initial design does not allow for an accurate initial calibrated predictor to be built, and a large initial design whose points happen to be far from the true Pareto Front gives few opportunities for additional points to home in to the goal. In general, the optimal fraction of observations allocated to the initial design is likely to depend not only upon the complexity and dimension of the output functions, but also the size of the bias. Further studies are needed in this direction.

7 Summary and Discussion

This paper presents a method of efficiently designing a manufacturing process for injection molding by determining the Pareto optimal Set of control factor settings; here these are the values of the melt temperature, packing time, packing pressure, and cooling time of the molding machine. The objectives are the minimization of the relative shrinkages of a set of test components. A final choice among the Pareto Set of control factor settings can be made on the basis of additional criteria, such as cost, ease of manufacture, etc. The method presented uses a sequential design guided by a minimax improvement function to rank the ability of candidate

control factor settings to improve on the Pareto Set and Pareto Front of the current data. A calibrated predictor that combines manufacturing process observations and simulator runs is used to estimate the mean of the manufacturing process. The methodology can be used in any other setting where there are conflicting outputs for which Pareto optimization provides meaningful solutions.

In general the points added by the sequential design will not be space filling over the entire input region because the Pareto Set is the target rather than good overall prediction. However, the added points will typically be space filling over the region around the Pareto Set. This can be seen in the two studies of analytic functions presented in Section 6; the spread of added points across the Pareto Set for the first example is apparent in the left hand panel of Figure 12 and, to a lesser extent, in the left hand panel of Figure 13 for the second example. For the injection molding experiment, the location of the true Pareto Set is unknown, but the ten sequentially added points are clustered around the estimated Pareto Set identified in the last three lines of Table 4.

While the computational budget in this application prevented additional simulator runs from being made, a further small improvement in the Pareto Set and Front may still be possible. One candidate strategy to find improvement is as follows. The *predictor* based on the data collected after the sequentially added observations could be used to search in a grid around the current approximate Pareto Set. Table 4 suggests searching the hyper-rectangle bounded by $200 \leq T_{melt} \leq 208$, $22 \leq t_{pack} \leq 28$, $42 \leq P_{pack} \leq 44$, and $43 \leq t_{cool} \leq 44$. Of course, validation runs should also be run when relying on the calibrated predictor.

Additional physical observations are likely to help improve the calibration. In this study, the limited availability of experimental material prevented this option.

Viewed as a frequentist model (by omitting priors), the likelihood portion of the Kennedy and O’Hagan (2001) (KOH) calibration model need not be identifiable as pointed out by Wynn (2001) in his discussion of KOH. Even with the Bayesian addition of priors, the KOH model can also be non-identifiable, see Gramacy et al. (2015); Tuo and Wu (2015); Plumlee (2017). For example, non-identifiability would be indicated when there are multiple posterior modes for the calibration parameters that provide comparable explanations of a given data set.

Priors for the simulator output, $y^s(\mathbf{x}, \mathbf{t})$, can be determined using the techniques described

in Oakley (2002). However, forming priors for the calibration parameters and the bias function is more difficult. The usual advice is to rely on expert opinion and experience with similar data to suggest, marginally, priors for the values of the calibration parameters and the magnitude and shape of the bias function. However, the calibration parameters and the bias function are correlated: a large adjustment in the value of the calibration parameters may allow for a smaller bias. Relying on expert knowledge to formulate priors for calibration parameters may not lead to correct inference because the simulator is an approximation to the physical reality and thus the calibration parameters may perform a different role in the simulator mathematics than in the physical system. For this reason several recent papers (e.g., Tuo and Wu, 2015; Plumlee, 2017) have advocated selecting the value of the calibration parameters mathematically to minimize a specified deviation metric that measures the difference between mean of the physical system and the simulator. Because the bias function depends on the selected value of the calibration parameters, Plumlee (2017) suggests choosing the prior distribution of the bias to depend on the calibration parameters in such a way that the draws of the bias function are stochastically orthogonal to the gradient of the deviation metric evaluated at the calibration parameters.

The initial design of each of the examples in Section 6 used two independent maximin LHDs: one for the physical observations and another for the simulator runs. Similarly, two independent initial designs were used for the physical and simulator experiments in the injection molding optimization study. In terms of prediction of outputs over the entire surface, a different pair of initial designs might possibly perform better. For example, Kennedy and O’Hagan (2001) suggested that at least some parts of the design for the computer simulator should be “close” to the physical design points in order to better estimate the simulator bias. A specific implementation of this concept is given in Leatherman et al. (2017) who proposed use of *maximin augmented nested Latin hypercube designs* (MmANLHD) for combined physical/simulator experiments, and provided an algorithm for their construction. Such designs build on the notion of *nested Latin hypercube designs* discussed in Qian (2009) and Rennen et al. (2010) but allow different numbers of factors for the two designs as would be needed in the case of a physical experiment having only control factors and a simulator which also requires calibration parameter settings. Leatherman et al. (2017) showed that, for prediction over the

test surfaces studied, the MmANLHD performed well, as did designs that were composed of a maximin Latin hypercube design for the simulator experiment coupled with an I -optimal design for the physical experiment. The performance of linked designs for the objective of predicting the Pareto Front, as in this paper, has not yet been studied

Lastly, we note that many multi-factor physical experiments are not able to be run using a completely randomized design but involve blocking and randomization restrictions. For example, in the physical experiment for the injection molding study of this paper, we could only do a restricted randomization of the control factor settings due to the nature of the manufacturing equipment and the length of time it took to heat up the mold to equilibrium and to cool it down again. To help overcome any biases, the order of observation of control factor settings was arranged so as to be trend free across days and during the day so that the estimated effects of the control factor settings would be free of any day or time effects. An open area of research is how to design efficient sequential update rules that account for such complicated physical models when the goal is Pareto optimization.

Acknowledgments

This research was partially supported by CONACYT-Mexico and by the National Science Foundation under Agreements DMS-0806134 and DMS-1310294 (The Ohio State University). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Atashkari, K., Nariman-Zadeh, N., Pilechi, A., Jamali, A. and Yao, X. (2005). “Thermodynamic Pareto Optimization of Turbojet Engines Using Multi-objective Genetic Algorithms”. *International Journal of Thermal Sciences* 44, pp. 1061–1071.
- Binois, M., Ginsbourger, D. and Roustant, O. (2014). “Quantifying Uncertainty on Pareto Fronts with Gaussian Process Conditional Simulations”. *European Journal of Operational Research* 243, pp. 386–394.

- Chen, P.-H. (2016). *Modeling Multivariate Simulator Outputs with Applications to Prediction and Sequential Pareto Minimization*. PhD thesis. The Ohio State University.
- Chen, P., Santner, T. and Dean, A. (2017). “Sequential Pareto Minimization of Physical Systems Using Calibrated Computer Simulators”. *Statistica Sinica*, in press.
- Deb, K., Thiele, L., Laumanns, M. and Zitzler, E. (2005). “Scalable Test Problems for Evolutionary Multi-Objective Optimization”. In *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications* (A. Abraham, R. Jain and R. Goldberg (eds)), chapter 6, pp. 105–145, Springer.
- Draguljić, D., Nekkanty, S., Santner, T. J., Dean, A. M. and Shivpuri, R. (2015). “Optimizing Thin Film Tool Coatings Using a Finite Element Computer Simulator”. *Quality Engineering* 27, pp. 461–472.
- Emmerich, M., Giannakoglou, K. C. and Naujoks, B. (2006). “Single- and Multiobjective Evolutionary Optimization Assisted by Gaussian Random Field Metamodels”. *IEEE Transactions on Evolutionary Computation* 10, pp. 421–439.
- Fonseca, C. M. and Fleming, P. J. (1995). “Multiobjective Genetic Algorithms Made Easy: Selection Sharing and Mating Restriction”. *Genetic Algorithms in Engineering Systems: Innocations and Applications, 1995. GALEZIA. First International Conference on (Conf. Publ. No. 414)*, pp. 45–52.
- Gramacy, R. B., Bingham, D., Holloway, J. P., Grosskopf, M. J., Kuranz, C. C., Rutter, E., Trantham, M., Drake, R. P. et al. (2015). “Calibrating a Large Computer Experiment Simulating Radiative Shock Hydrodynamics”. *The Annals of Applied Statistics* 9, pp. 1141–1168.
- Higdon, D., Gattiker, J., Williams, B. and Rightley, M. (2008). “Computer Model Calibration Using High Dimensional Output”. *Journal of the American Statistical Association* 103, pp. 570–583.
- Higdon, D., Kennedy, M., Cavendish, J., Cafo, J. and Ryne, R. (2004). “Combining Field Data and Computer Simulations for Calibration and Prediction”. *SIAM Journal of Scientific Computing* 26, pp. 448–466.

- Jones, D. R., Schonlau, M. and Welch, W. J. (1998). “Efficient Global Optimization of Expensive Black-Box Functions”. *Journal of Global Optimization* 13, pp. 455–492.
- Keane, A. J. (2006). “Statistical Improvement Criteria for Use in Multiobjective Design Optimization”. *AIAA Journal* 44, pp. 879–891.
- Kennedy, M. C. and O’Hagan, A. (2001). “Bayesian Calibration of Computer Models (with discussion)”. *Journal of the Royal Statistical Society B* 63, pp. 425–464.
- Kim, I. Y. and deWeck, O. L. (2006). “Adaptive Weighted Sum Method for Multiobjective Optimization: A New Method for Pareto Front Generation”. *Structural and Multidisciplinary Optimization* 31, pp. 105–116.
- Leatherman, E. R., Dean, A. M. and Santner, T. J. (2017). “Designing Combined Physical and Computer Experiments to Maximize Prediction Accuracy”. *Computational Statistics and Data Analysis* 113, pp. 346–362.
- Leatherman, E. R., Guo, H., Gilbert, S. L., Hutchinson, I. D., Maher, S. A. and Santner, T. J. (2014). “Using a Statistically Calibrated Biphasic Finite Element Model of the Human Knee Joint to Identify Robust Designs for a Meniscal Substitute”. *Journal of Biomechanical Engineering* 136, pp. 071007.
- MacDonald, B., Ranjan, P. and Chipman, H. (2015). “GPfit: An R Package for Gaussian Process Model Fitting Using a New Optimization Algorithm”. *Journal of Statistical Software* 64, pp. 1–23.
- Oakley, J. E. (2002). “Eliciting Gaussian Process Priors for Complex Computer Codes”. *The Statistician* 51, pp. 81–97.
- Picheny, V. (2015). “Multiobjective Optimization Using Gaussian Process Emulators via Stepwise Uncertainty Reduction”. *Statistics and Computing* 25, pp. 265–280.
- Plumlee, M. (2017). “Bayesian Calibration of Inexact Computer Models”. *Journal of the American Statistical Association* 112, pp. 1274–1285.
- Qian, P. Z. G. (2009). “Nested Latin Hypercube Designs”. *Biometrika* 96, pp. 957–970.

- Rennen, G., Husslage, B., Van Dam, E. R. and Dick, D. H. (2010). “Nested Maximin Latin Hypercube Designs”. *Structural and Multidisciplinary Optimization* 41, pp. 371–395.
- Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989). “Design and Analysis of Computer Experiments”. *Statistical Science* 4, pp. 409–423.
- Santner, T. J., Williams, B. J. and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer Verlag, New York.
- Svenson, J. and Santner, T. J. (2016). “Multiobjective Optimization of Expensive-To-Evaluate Deterministic Computer Simulator Models”. *Computational Statistics & Data Analysis* 94, pp. 250–264.
- Tuo, R. and Wu, C. F. J. (2015). “Efficient Calibration for Imperfect Computer Models”. *The Annals of Statistics* 43, pp. 2331–2352.
- van Dam, E., den Hertog, D., Husslage, B. and Rennen, G. (2013). “Space-Filling Designs”. <https://spacefillingdesigns.nl>, accessed: July 2015 -June 2016.
- Villarreal-Marroquín, M., Chen, P.-H., Mulyana, R., Santner, T. J., Dean, A. M. and Castro, J. M. (2017). “Multiobjective Optimization of Injection Molding Using a Calibrated Predictor Based on Physical and Simulated Data”. *Polymer Engineering and Science* 57, pp. 248–257.
- Williams, B., Higdon, D., Gattiker, J., Moore, L., McKay, M., Keller-McNulty, S. et al. (2006). “Combining Experimental Data and Computer Simulations, with an Application to Flyer Plate Experiments”. *Bayesian Analysis* 1, pp. 765–792.
- Wilson, B., Cappelleri, D., Simpson, T. W. and Frecker, M. (2001). “Effect Pareto Frontier Exploration Using Surrogate Approximations”. *Optimization and Engineering* 2, pp. 31–50.
- Wynn, H. P. (2001). “Discussion of Bayesian Calibration of Computer Models by M. C. Kennedy and A. O’Hagan”. *Journal of the Royal Statistical Society B* 63, pp. 450–451.
- Zitzler, E. and Thiele, L. (1999). “Multiobjective Optimization using Evolutionary Algorithms - A Comparative Case Study”. *Evolutionary Computation, IEEE Transactions on* 3, pp. 251–271.