Robust Dual Response Optimization*

İhsan Yanıkoğlu[‡] • Dick den Hertog[†] • Jack P.C. Kleijnen[†]
[‡]Industrial Engineering Department, Özyeğin University, İstanbul, Turkey
[†]Center for Economic Research (CentER), Tilburg University, Tilburg, The Netherlands
ihsan.yanikoglu@ozyegin.edu.tr • d.denhertog@uvt.nl • kleijnen@uvt.nl

This article presents a robust optimization reformulation of the dual response problem developed in response surface methodology. The dual response approach fits separate models for the mean and the variance, and analyzes these two models in a mathematical optimization setting. We use metamodels estimated from experiments with both controllable and environmental inputs. These experiments may be performed with either real or simulated systems; we focus on simulation experiments. For the environmental inputs, classic approaches assume known means, variances or covariances, and sometimes even a known distribution. We, however, develop a method that uses only experimental data, so it does not need a known probability distribution. Moreover, our approach yields a solution that is robust against the ambiguity in the probability distribution. We also propose an adjustable robust optimization method that enables adjusting the values of the controllable factors after observing the values of the environmental factors. We illustrate our novel methods through several numerical examples, which demonstrate their effectiveness.

Key words: robust optimization; dual response optimization; simulation optimization; phi-divergence

1. Introduction

The goal of many *experiments* is to estimate the best solution for a given practical problem. Such experiments may be conducted with a physical system, e.g., a semi-conductor manufacturing process, or a mathematical model of a physical system, e.g., a computerized simulation model of a semi-conductor manufacturing process. These experiments produce data on the outputs or responses for the given inputs or factors. Output may be univariate (a single or scalar response) or multivariate (multiple responses or a vector with responses). The number of inputs may range from a single input to 'many' inputs (e.g., thousands of inputs), but we focus on practical problems with no more than (say) twenty inputs.

This paper proposes a methodology that combines robust, simulation based and dual response optimization. This section is a rather long introduction that briefly discusses each of these methodologies. In the remainder of the introduction, we first emphasize Taguchi's (1924–2012) contribution to the literature; namely, robust parameter design. Then we discuss the shortcomings of Taguchi's approach, and we present an alternative method that

* Please cite this article as **Yanıkoğlu et al**. 2015. Robust Dual Response Optimization. ***IIE Transactions.***

is developed in response surface methodology, namely, a dual response approach. Next, we give our contribution to the associated literature, and finalize this section by giving a brief introduction on simulation based and robust optimization.

Taguchi (1986, 1987)—for an update see Myers et al. (2009, pp. 483–485)—distinguishes between the following two types of inputs: (i) controllable or decision factors (say) d_j $(j = 1, \ldots, k)$, collected in the k-dimensional vector $\mathbf{d} = (d_1, \ldots, d_k)^T$. (ii) Environmental or noise factors (say) e_g $(g = 1, \ldots, c)$, collected in the c-dimensional vector $\mathbf{e} = (e_1, \ldots, e_c)^T$. By definition, the first type of inputs is under the control of the users; e.g., in an inventory system, management controls the order quantity. The second type of inputs is not controlled by the users; e.g., demand in an inventory system.

We assume that the unknown input/output (I/O) function (say) $w = h(\mathbf{e}, \mathbf{d})$ for a single output w of a given system is approximated by the following 'incomplete' second-order polynomial regression metamodel (simulation analysts use different names for metamodels, such as response surfaces, surrogates, and emulators; see the many references in Kleijnen (2008, p. 8)):

$$y(\mathbf{e}, \mathbf{d}) = \boldsymbol{\beta_0} + \boldsymbol{\beta}^T \mathbf{d} + \mathbf{d}^T \mathbf{B} \mathbf{d} + \boldsymbol{\gamma}^T \mathbf{e} + \mathbf{d}^T \boldsymbol{\Delta} \mathbf{e} + \epsilon, \tag{1}$$

where y denotes the output of the regression metamodel of the simulation model with output w; $\boldsymbol{\beta_0}$ the intercept; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^T$ the first-order effects of the decision variables $\mathbf{d}$; $\mathbf{B}$ the $k \times k$ symmetric matrix of second-order effects of $\mathbf{d}$ with on the main diagonal the 'purely quadratic' effects B_{jj} and off this diagonal half the interactions between pairs of decision factors $B_{jj'}/2$ $(j \neq j')$; $\gamma = (\gamma_1, \ldots, \gamma_c)^T$ the first-order effects of the environmental factors $\mathbf{e}$; $\boldsymbol{\Delta} = (\delta_{jg})$ the $k \times c$ pairwise (two-factor) interactions between $\mathbf{d}$ and $\mathbf{e}$; ϵ the residual of the metamodel with $E(\epsilon) = 0$ if this model has no 'lack of fit' (so it is a valid or adequate approximation of $h(\mathbf{e}, \mathbf{d})$) and with constant variance σ_ϵ^2.

To estimate the regression parameters in (1), we need to experiment with the underlying system. Design of experiments (DoE) uses *coded*—also called *standardized* or *scaled*—values (say) x_j for the factors. So the experiment consists of (say) n combinations of the coded factors $\mathbf{x}$, which correspond to $\mathbf{d}$ and $\mathbf{e}$ in (1). Coding is further discussed in Kleijnen (2008, p. 29) and Myers et al. (2009, p. 78). We reformulate the Taguchian model (1) as the following *linear regression model*:

$$y = \boldsymbol{\zeta}^T \mathbf{x} + \epsilon \tag{2}$$

with the ℓ-dimensional vector of regression parameters or coefficients $\boldsymbol{\zeta} = (\boldsymbol{\beta_0}, \ldots, \delta_{kc})^T$ and the corresponding vector of regression explanatory variables $\mathbf{x}$ defined in the obvious way; e.g., the ℓ-th explanatory variable corresponding with the interaction effect $\beta_{(\ell-1)\ell}$ is $d_{\ell-1}d_\ell$. Then (2) leads to the *least squares* (LS) estimator:

$$\widehat{\boldsymbol{\zeta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{w}, \tag{3}$$

where $\mathbf{X}$ is the $n \times \ell$ matrix of explanatory variables with n denoting the number of scenarios (combinations of controllable and environmental factors) determined by the DoE that are actually observed (in a real or simulated system); $\mathbf{w}$ is the vector with the n observed outputs.

In this paper, we focus on experiments with simulation models of real systems. These simulation models may be either deterministic (especially in engineering) or random (stochastic, possibly with discrete events in operations research). Deterministic simulation models have noise if a parameter or input variable has an unknown value; this is called subjective or epistemic uncertainty; see Iman and Helton (2006). Random simulation models have—beside subjective—objective, aleatory or inherent uncertainty; again see Iman and Helton. We focus on random simulation, but shall also discuss deterministic simulation. Furthermore, we adopt the notation of metamodel (1) in the following sections of this paper; nevertheless, the methods proposed in the sequel are not bound to low-order polynomials. Our methods can be applied for any given metamodel type as long as the metamodel can be optimized for the controllable factors.

Taguchi emphasizes that the environmental factors $\mathbf{e}$ create variability in the output w. Consequently, the combination of controllable decision factors $\mathbf{d}$ that (say) maximizes the expected univariate output may cause the variance of that output to be much larger than a combination that 'nearly' maximizes the mean; i.e., a little sacrifice in expected output may save a lot of problems caused by the variability in the output. Samuelson (2010) also emphasizes that it may be impractical to search for the very best. The aim of *robust parameter design* (RPD), proposed by Taguchi (1986, 1987), is to find the optimal setting of the controllable factors that minimizes the response variance while keeping the mean output on a predefined target value. Taguchi has proposed using the *signal-to-noise ratios* (SNRs) that depends on $\mathrm{E}(y)/\sqrt{\mathrm{Var}(y)}$ (the standard deviation $\sqrt{\mathrm{Var}(y)}$ has the same scale as the mean $\mathrm{E}(y)$); see Myers et al. (2009, pp. 486–488). The precise definitions

of these SNRs vary with the following goals of RPD: (i) 'the smaller, the better': minimize the response. (ii) 'The larger, the better': maximize the response. (iii) 'The target is best': realize a target value (say) T for the response.

Taguchi's RPD has been criticized by *statisticians*; see the panel discussion reported in Nair et al. (1992). Their main critique concerns the statistical design and analysis in Taguchi's approach. Moreover, SNRs have been criticised because they were unable to distinguish between decision variables affecting the response mean from those affecting the variance. For details on both criticisms, we refer to Myers et al. (2009, pp. 483–495). We do not further dwell on Taguchi's SNRs, because we also think there are better formulations of the various goals of RPD.

Response surface methodology (RSM) uses a collection of statistical and mathematical techniques for developing an approximate functional relationship between a response of interest and a number of input variables. Additionally, one of the main objectives of RSM is deciding on the optimal settings of the control variables that result in a maximum (or a minimum) response over a certain region of interest. This RSM is a stepwise optimization heuristic that in the various steps uses local first-order polynomial metamodels and in the final step uses a local second-order polynomial metamodel (obviously, these polynomials are linear regression models). For additional details on RSM, we refer to Myers et al. (2009), and the review paper by Khuri and Mukhopadhyay (2010). As a remedy for the above mentioned criticisms on RPD, we may apply a *dual response* optimization approach developed in RSM. Unlike the RPD approach, the dual response approach fits separate models to the mean and the variance. The metamodel (1) implies that the regression predictor for the response mean is

$$\mathrm{E}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] = \beta_0 + \boldsymbol{\beta}^T\mathbf{d} + \mathbf{d}^T\mathbf{B}\mathbf{d} + \boldsymbol{\gamma}^T\boldsymbol{\mu}_{\mathbf{e}} + \mathbf{d}^T\boldsymbol{\Delta}\mu_{\mathbf{e}} \tag{4}$$

and the regression predictor for the response variance is

$$\mathrm{Var}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] = (\boldsymbol{\gamma}^T + \mathbf{d}\boldsymbol{\Delta})\boldsymbol{\Omega}_{\mathbf{e}}(\boldsymbol{\gamma} + \boldsymbol{\Delta}^T\mathbf{d}), \tag{5}$$

where $(\boldsymbol{\gamma} + \boldsymbol{\Delta}^T\mathbf{d}) = (\partial y/\partial e_1, \ldots, \partial y/\partial e_c)^T$ is the gradient with respect to the environmental factors, $\mathrm{E}(\mathbf{e}) = \boldsymbol{\mu}_{\mathbf{e}}$ denotes the mean vector of the environmental factors ($\mathbf{e}$), and $\mathrm{Cov}(\mathbf{e}) = \boldsymbol{\Omega}_{\mathbf{e}}$ is the covariance matrix of $\mathbf{e}$. The analysis of these 'dual' responses often involves an optimization problem (see del Castillo et al. (1997, 1999), del Castillo (2007), and Myers

and Carter (1973)). Relaxing the "mean at target" constraint of RPD, we propose the following dual response optimization problem:

$$\min_{\mathbf{d}} \operatorname{Var}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] \quad \text{s.t.} \quad \mathrm{E}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] \leq T, \tag{6}$$

which minimizes the response variance while keeping the response mean less than or equal to a predefined value T.

The dual response approach to RPD assumes that the mean and the variance—and sometimes even the probability distribution—of $\mathbf{e}$ are known; e.g., $\boldsymbol{\mu}_{\mathbf{e}}$ and $\boldsymbol{\Omega}_{\mathbf{e}}$ are assumed to be known in (4) and (5). However, the final parameter design may be sensitive to these assumptions. As del Castillo (2007, p. 246) points out "although in principle historical data can be used to estimate the mean and the covariance matrix, the possibility of errors in these estimates indicates the need for future research". Moreover, Bingham and Nair (2012) state that "the noise distributions are rarely known, and the choices are often based on convenience". In this paper, we propose a robust optimization approach that takes such distribution ambiguities of the environmental factors into account. Consequently, we consider the following *robust optimization* (RO) reformulation of (6):

$$\min_{\mathbf{d}} \max_{\mathbf{p}(\mathbf{e})\in U} \operatorname{Var}_{\mathbf{p}(\mathbf{e})}[y(\mathbf{e},\mathbf{d})] \quad \text{s.t.} \quad \mathrm{E}_{\mathbf{p}(\mathbf{e})}[y(\mathbf{e},\mathbf{d})] \leq T \quad \forall \mathbf{p}(\mathbf{e}) \in U, \tag{7}$$

where $\mathbf{p}(\mathbf{e})$ denotes the unknown true probability distribution of $\mathbf{e}$, and the uncertainty set U defines a family of distributions that is obtained via data and goodness-of-fit statistics.

The *major contributions* of our research can be summarized as follows: (i) We propose a RO methodology for a class of dual response problems where the distributional parameters are unknown but historical data on the uncertain parameters are available. (ii) We propose a *tractable* robust counterpart formulation of the dual response problem (7). Compared with other studies in the literature, our formulation can handle more general classes of uncertainty sets, and leads to a more tractable formulation. (iii) Our method is suitable for different classes of metamodels, e.g., higher-order polynomials, Kriging, and radial basis functions. (iv) We propose an adjustable robust optimization method for the dual response problem; unlike classic RO techniques, our adjustable robust reformulation is tractable, even for nonlinear decision rules. (v) Finally, we use adjustable integer decision variables in the context of RO, and propose a specific decision rule for such variables. We observe that a *limitation* of our method is that it is only suitable for low dimensional uncertainties.

Related to our research is optimization of systems being simulated—called *simulation optimization* (SO)—which is a popular research topic in discrete-event simulation; see Fu (2007, 2014), and Fu and Nelson (2003). Robust and SO approaches are also discussed in that literature. Examples are Angün (2011) combining Taguchi's worldview with RSM. Instead of Taguchi's various criteria, Angün (2011) uses the average value-at-risk (also known as conditional value-at-risk). Miranda and del Castillo (2011) perform RPD optimization through a well-known SO method; namely, simultaneous perturbation stochastic approximation, which is detailed in Spall (2003). Wiedemann (2010, p. 31) applies Taguchian RPD to an agent-based simulation, ensuring that the mean response meets the target value and the variability around that level is sufficiently small. Al-Aomar (2006) used Taguchi's signal-to-noise ratio and a quality loss function, together with a genetic algorithm with a scalar fitness measure that is a combination of the estimated mean and variance. Sanchez (2000) used Taguchi's worldview with a loss function that incorporates the system's mean and variability, and RSM. Robust approaches are discussed—albeit briefly—not only in discrete-event simulation but also in deterministic simulation if that simulation has uncertain environmental variables. A recent example is Hu et al. (2012), who propose a robust climate simulation model (called 'DICE') with input parameters that have a multivariate Gaussian distribution with uncertain or 'ambiguous' mean vector and covariance matrix. Dellino et al. (2012) investigate the well-known economic order quantity model with an uncertain demand rate. Dellino et al. use Taguchi's worldview, but replace his experimental designs (namely, orthogonal arrays) and metamodels (namely, low-order polynomials) by Latin hypercube sampling and Kriging metamodels. Many more applications can be found in engineering. For example, Chan and Ewins (2010) use RPD to manage the vibrations in bladed discs. Urval et al. (2010) use Taguchi's RPD in powder injection moulding of microsystems' components. Chen (1997) extends Taguchi's RPD approach for multiple quality characteristics. Delpiano and Sepulveda (2006) also study RPD in engineering.

RO in *mathematical programming* (MP) started with Ben-Tal and Nemirovski (1998, 1999) and El-Ghaoui and Lebret (1997). MP problems usually have uncertain coefficients in the objective function and the constraints, so the "nominal" optimal solution—i.e., the optimal solution if there would be no uncertainties—may easily violate the constraints for some realizations of the uncertain coefficients. Therefore, it is better to find a "robust"

solution that is immune to the uncertainty in a so-called uncertainty set. The robust reformulation of a given uncertain mathematical optimization problem is called the *robust counterpart* (RC) problem. The mathematical challenge in RO is to find computationally tractable RCs; see Ben-Tal et al. (2009). Inspired by Ben-Tal et al., Stinstra and den Hertog (2008) propose RO methods for three types of errors, namely, simulation-model, meta-model, and implementation errors, that may arise at different stages of a simulation-based optimization approach, considering box and ellipsoidal uncertainties for the associated errors. Bertsimas et al. (2010) on the other hand propose a RO method for unconstrained simulation-based problems with non-convex cost functions; their method is suitable for metamodels that are often used in practice such as Kriging, and it can be generalized to both implementation errors and parameter uncertainties.

The remainder of the paper is organized as follows. §2 proposes a RO approach accounting for distribution ambiguity. §3 develops adjustable RO approaches. §4 presents numerical examples. §5 summarizes conclusions.

2. Robust Optimization with Unknown Distributions

In this section, we derive a robust reformulation of the dual response optimization problem. We assume that data on the environmental factors $\mathbf{e}$ is available or can be obtained via simulation.

2.1. Uncertainty Sets

Instead of relying on the normal distribution, RO derives an *uncertainty set* for the unknown density function of the noise factors. There are several RO approaches, but we follow Ben-Tal et al. (2013) and Yanıkoğlu and den Hertog (2013), who develop RO accounting for *historical data* on the noise factors. They do not use a specific distribution for this data; instead, they use the more general concept of *phi-divergence*—as follows.

Given N historical observations on the noise factors $\mathbf{e}$, we construct m cells such that the number of observations o_i in cell i $(i = 1, \ldots, m)$ is at least five (the value five hast to do with the asymptotic distribution of the goodness-of-fit statistics; see (8) below):

$$\textstyle\sum_{i=1}^{m} o_i = N \text{ such that } \forall i : o_i \geq 5.$$

The historical data on $\mathbf{e}$ give the frequencies $\mathbf{q} = [q_1, \ldots, q_m]^T$, where q_i is the observed frequency in cell i so

$$q_i = \frac{o_i}{N}.$$

We consider $\mathbf{p} = [p_1, \ldots, p_m]^T$ as the unknown true probability vector of $\mathbf{e}$ when its support is discretized into m cells; hence $\mathbf{q}$ is an approximation of $\mathbf{p}$. The phi-divergence measure is

$$I_\phi(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{m} q_i \phi\left(\frac{p_i}{q_i}\right)$$

where $\phi(.)$ satisfies certain mathematical requirements such as $\phi(1) = 0$, $\phi(a/0) := a \lim_{t\to\infty} \phi(t)/t$ for $a > 0$, and $\phi(0/0) = 0$; for details on phi-divergence we refer to Pardo (2006). It can be proven that the test statistic

$$\frac{2N}{\phi''(1)} I_\phi(\mathbf{p}, \mathbf{q})$$

is asymptotically distributed as a chi-squared random variable with $(m-1)$ degrees of freedom. So an asymptotic $(1-\alpha)$-confidence set for $\mathbf{p}$ is

$$I_\phi(\mathbf{p}, \mathbf{q}) \le \rho \text{ with } \rho = \rho(N, m, \alpha) = \frac{\phi''(1)}{2N} \chi^2_{m-1;1-\alpha}. \tag{8}$$

Using (8), Ben-Tal et al. (2013) derive the following *uncertainty set* U for the unknown probability vector $\mathbf{p}$:

$$U = \{\mathbf{p} \in \mathbb{R}^m | \mathbf{p} \ge 0, \sum_{i=1}^{m} p_i = 1, I_\phi(\mathbf{p}, \mathbf{q}) \le \rho\}. \tag{9}$$

Table 1 Phi-Divergence Examples

Divergence	$\phi(t), t > 0$	$I_\phi(\mathbf{p}, \mathbf{q})$	$\phi^*(s)$
Kullback-Leibler	$t \log t$	$\sum_i p_i \log\left(\frac{p_i}{q_i}\right)$	e^{s-1}
Burg entropy	$-\log t$	$\sum_i q_i \log\left(\frac{p_i}{q_i}\right)$	$-1 - \log(-s), s \le 0$
χ^2-distance	$\frac{1}{t}(t-1)^2$	$\sum_i \frac{(p_i - q_i)^2}{p_i}$	$2 - 2\sqrt{1-s}, s \le 1$
Pearson χ^2-distance	$(t-1)^2$	$\sum_i \frac{(p_i - q_i)^2}{q_i}$	$s + s^2/4, s \ge -2$ $-1, s < -2$
Hellinger distance	$(1-\sqrt{t})^2$	$\sum_i (\sqrt{p_i} - \sqrt{q_i})^2$	$\frac{s}{1-s}, s \le 1$

Table 1 taken from Ben-Tal et al. (2013), presents common choices of the phi-divergence function; $\phi^*(s) := \sup_{t\ge 0}\{st - \phi(t)\}$ denotes the conjugate of a phi-divergence distance that will be used for the derivations of RCs in the next section. In various examples, we shall use the χ^2-distance in this paper.

2.2. Robust Counterparts of Mean and Variance Problem

In this section we derive the tractable robust counterpart of the robust dual response problem (7), given in §1. Slight modifications of the derivations in this section enable the derivation of the RC of any problem based on mean and variance.

We represent each cell i by its center point $\mathbf{e}^i$; e.g., e_g^i is the g^{th} $(g=1,\ldots,c)$ coordinate of cell i's center point. To apply the phi-divergence in (9), we discretize $y(\mathbf{e},\mathbf{d})$ by replacing $\mathbf{e}$ in (1) by $\mathbf{e}^i$, which gives

$$y_i(\mathbf{d}) = y(\mathbf{e}^i,\mathbf{d}) = \beta_0 + \boldsymbol{\beta}^T\mathbf{d} + \mathbf{d}^T\mathbf{B}\mathbf{d} + (\boldsymbol{\gamma}^T + \mathbf{d}^T\Delta)\mathbf{e}^i. \tag{10}$$

Consequently, the mean of $y(\mathbf{e},\mathbf{d})$ is *approximated* by

$$\mathrm{E}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] = \sum_{i\in V} y_i(\mathbf{d})p_i = \sum_{i\in V}[\beta_0 + \boldsymbol{\beta}^T\mathbf{d} + \mathbf{d}^T\mathbf{B}\mathbf{d} + (\gamma^T + \mathbf{d}^T\Delta)\mathbf{e}^i]p_i, \tag{11}$$

where p_i denotes the probability of $\mathbf{e}$ falling into cell i, and V is the set of cell indices. Note that $\mathbf{p}$ is in the uncertainty set U given by (9), and the empirical estimate $\mathbf{q}$ of $\mathbf{p}$ is obtained using the data on $\mathbf{e}$. The limitation of our approach is that it requires excessively many data points in high dimensional problems (where c is "high"). Next we define

$$\psi_i(\mathbf{d}) := (\boldsymbol{\gamma}^T + \mathbf{d}^T\Delta)\mathbf{e}^i, \tag{12}$$

so the variance of $y(\mathbf{e},\mathbf{d})$ is *approximated* by

$$\mathrm{Var}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] = \sum_{i\in V}\psi_i(\mathbf{d})^2 p_i - \left[\sum_{i\in V}\psi_i(\mathbf{d})p_i\right]^2. \tag{13}$$

Eventually, the *robust reformulation* (7) is approximated by the following *semi-infinite* optimization problem:

$$\begin{aligned}(\mathbf{SI1})\ \min_{\mathbf{d}}\ \max_{\mathbf{p}\in U} & \sum_{i\in V}\psi_i(\mathbf{d})^2 p_i - \left[\sum_{i\in V}\psi_i(\mathbf{d})p_i\right]^2 \\ \text{s.t.} & \sum_{i\in V}[\beta_0 + \boldsymbol{\beta}^T\mathbf{d} + \mathbf{d}^T\mathbf{B}\mathbf{d} + (\gamma^T + \mathbf{d}^T\Delta)\mathbf{e}^i]p_i \le T \qquad \forall \mathbf{p}\in U,\end{aligned} \tag{14}$$

where $U = \{\mathbf{p}\in\mathbb{R}^m | \mathbf{p}\ge 0, \sum_{i=1}^m p_i = 1, I_\phi(\mathbf{p},\mathbf{q})\le\rho\}$.

Notice that (SI1) is a difficult optimization problem that has infinitely many constraints (see $\forall\mathbf{p}\in U$), and includes quadratic terms in $\mathbf{p}$. Ben-Tal et al. (2009, p. 382) propose

a tractable RC of a linear optimization problem with uncertain parameters that appear quadratically, and an ellipsoidal uncertainty. The resulting formulation is a semidefinite programming (SDP) problem; see also Remark 1 below. The following theorem provides tractable RC reformulations of (SI1) for more general 'phi-divergence' uncertainty sets.

Theorem 1 *The vector* $\mathbf{d}$ *solves (SI1) if and only if* $\mathbf{d}, \boldsymbol{\lambda}, \boldsymbol{\eta}$*, and* z *solve the following RC problem:*

$$(\mathbf{RC1}) \min_{\mathbf{d},\boldsymbol{\lambda},z,\boldsymbol{\eta}\geq\mathbf{0}} \lambda_2 + \rho\eta_2 + \eta_2 \sum_{i\in V} q_i \phi^* \left(\frac{(\psi_i(\mathbf{d}) + z)^2 - \lambda_2}{\eta_2} \right)$$
$$s.t.\ \beta_0 + \boldsymbol{\beta}^T \mathbf{d} + \mathbf{d}^T \mathbf{B}\mathbf{d} + \lambda_1 + \rho\eta_1 + \eta_1 \sum_{i\in V} q_i \phi^* \left(\frac{\psi_i(\mathbf{d}) - \lambda_1}{\eta_1} \right) \leq T, \qquad (15)$$

where ρ *is given by (8),* $\phi^*(s) := \sup_{t\geq 0}\{st - \phi(t)\}$ *denotes the conjugate of* $\phi(.)$*,* $V = \{1, ..., m\}$ *is the set of cell indices in the uncertainty set* U*,* q_i *is the data frequency in cell* $i \in V$ *using the historical data on* $\mathbf{e}$*, and* $\boldsymbol{\lambda}$*,* $\boldsymbol{\eta}$*, and* z *are additional variables.*

Proof. Using Yanıkoğlu and den Hertog (2013, Theorem 1), we can easily derive the explicit RC of the expectation constraint (14) that is linear in $\mathbf{p} \in U$. Next we consider the 'more difficult' objective function of (SI1), which is quadratic in $\mathbf{p}$. In the following parts of the proof we use some of the results in Ben-Tal et al. (2014) to account for the nonlinear uncertainty in the constraint. Using a linear transformation, we reformulate the inner maximization at (SI1)'s objective as

$$\max_{\mathbf{a}\in\widehat{U}} \hat{g}(\mathbf{a}), \qquad (16)$$

where $\widehat{U} := \{\mathbf{a} : \mathbf{a} = \mathbf{A}\mathbf{p}, \mathbf{p} \in U\}$, $\mathbf{A}^T = [\psi^2(\mathbf{d}), \psi(\mathbf{d})]$, and $\hat{g}(\mathbf{a}) = a_1 - a_2^2$. Using the indicator function

$$\delta(\mathbf{a}|\widehat{U}) := \begin{cases} 0, & \mathbf{a} \in \widehat{U} \\ +\infty, \text{ elsewhere,} \end{cases}$$

we reformulate (16) as

$$\max_{\mathbf{a}\in\mathbb{R}^2} \{\hat{g}(\mathbf{a}) - \delta(\mathbf{a}|\widehat{U})\}. \qquad (17)$$

The Fenchel dual of (17)—for details see Rockafellar (1970, pp. 327–341)— is equivalent to

$$\min_{\mathbf{v}\in\mathbb{R}^2} \{\delta^*(\mathbf{v}|\widehat{U}) - \hat{g}_*(\mathbf{v})\} \qquad (18)$$

where $\mathbf{v}$ denotes the dual variable, and $\delta^*(\mathbf{v}|\widehat{U}) := \sup_{\mathbf{p}\in U}\{\mathbf{a}^T\mathbf{v}|\mathbf{a}=\mathbf{Ap}\}$ and $\hat{g}_*(\mathbf{v}) := \inf_{\mathbf{a}\in\mathbb{R}^2}\{\mathbf{a}^T\mathbf{v}-\hat{g}(\mathbf{a})\}$ denote the convex and concave conjugates of the functions δ and $\hat{g}$, respectively. Going from (17) to (18) is justified since the intersection of the relative interiors of the domains of $\hat{g}(.)$ and $\delta(.|\widehat{U})$ is non-empty, since $\mathbf{a}=\mathbf{Aq}$ is always in the relative interiors of both domains. Moreover, it is easy to show that $\delta^*(\mathbf{v}|\widehat{U}) = \delta^*(\mathbf{A}^T\mathbf{v}|U)$. Then (18) becomes

$$\min_{\mathbf{v}\in\mathbb{R}^2}\{\delta^*(\mathbf{A}^T\mathbf{v}|U) - \hat{g}_*(\mathbf{v})\}.$$

Now we derive the complete formulas of the conjugate functions δ^* and $\hat{g}_*$. If $\mathbf{v}^T=[w,z]$, then the concave conjugate of $\hat{g}$ is equivalent to

$$\hat{g}_*(\mathbf{v}) = \inf_{\mathbf{a}\in\mathbb{R}^2}\{a_1w+a_2z-\hat{g}(\mathbf{a})\} = \begin{cases} -z^2/4, & w=1 \\ -\infty, & \text{elsewhere.} \end{cases}$$

The convex conjugate of δ is equivalent to

$$\delta^*(\mathbf{A}^T\mathbf{v}|U) = \inf_{\lambda,\eta_2\geq 0}\left\{\rho\eta_2+\lambda+\eta_2\sum_{i\in V}q_i\phi^*\left(\frac{\psi_i^2(\mathbf{d})+\psi_i(\mathbf{d})z-\lambda}{\eta_2}\right)\right\}.$$

but we can delete the infimum since the result follows from the Lagrangian dual problem as in the expectation constraint. Thus the final RC reformulation of the objective becomes

$$\min_{\mathbf{d},\boldsymbol{\lambda},z,\eta_2\geq 0}\lambda+\rho\eta_2+\eta_2\sum_{i\in V}q_i\phi^*\left(\frac{\psi_i^2(\mathbf{d})+\psi_i(\mathbf{d})z-\lambda}{\eta_2}\right)+\frac{z^2}{4}. \tag{19}$$

Substituting $\lambda_2=\lambda+z^2/4$ into (19) gives the final RC reformulation (RC1) ■

Remark 1 *An ellipsoidal uncertainty set is a special case of the phi-divergence uncertainty set (9) when the Pearson chi-squared distance is used as the phi-divergence. Moreover, we can reformulate (RC1) as a second order cone problem (SOCP) for the associated distance measure. Notice that SOCP is an 'easier' formulation of the problem than the SDP by Ben-Tal et al. (2009).*

Remark 2 *Ben-Tal et al. (2014) also propose an RC for the variance uncertainty, however the associated RC introduces additional non-convexity. We overcome this difficulty by using the substitution $\lambda_2=\lambda+z^2/4$ in the proof of Theorem 1.*

We now discuss the 'general' computational tractability of (RC1). First, $\phi^*(h(\mathbf{d},\lambda_2,z))$ is convex, since the convex conjugate $\phi^*(s) := \sup_{t\geq 0}\{st-\phi(t)\}$ is non-decreasing in s, and $h(\mathbf{d},\lambda_2,z) = (\psi_i(\mathbf{d})+z)^2-\lambda_2$ is convex in $\mathbf{d}, z$, and λ_2. It is easy to show that $\eta_2\phi^*(\cdot/\eta_2)$ is convex, since the perspective of a convex function is always convex. Eventually, the convexity of the perspective function implies that the objective function of (RC1) is convex. On the other hand, (15) is not necessarily convex, since y is non-convex in $\mathbf{d}$ unless $\mathbf{B}$ is a positive semidefinite (PSD) matrix. Nevertheless, (RC1) does not introduce additional non-convexity into the general optimization problem (6) (i.e., (RC1) is a convex optimization problem, if $\mathbf{B}$ is a PSD matrix).

2.3. Alternative Metamodels and Risk Measures

In this subsection, we focus on extensions of our method. §2.3.1 presents a generalization of our method for other metamodels besides low-order polynomials, and §2.3.2 presents an extension of our method to SNRs. Finally, §§2.3.3 and 2.3.4 show how to apply our method for different risk measures.

2.3.1. Alternative Metamodels In this paper we focus on the low-order polynomial (1), since most of the literature and real-life applications use low-order polynomials to approximate the I/O function of the underlying simulation or physical experiment. However, our methodology can also be used to other metamodel types such as higher-order polynomials, Kriging, and radial basis functions. More precisely, consider

$$y(\mathbf{d},\mathbf{e}) = f(\mathbf{d}) + \psi(\mathbf{d},\mathbf{e})$$

where $f(\mathbf{d})$ is the part that affects only the mean of the response (e.g., it is $f(\mathbf{d}) = \beta_0 + \beta^T\mathbf{d} + \mathbf{d}^T\mathbf{B}\mathbf{d}$ in (1)), and $\psi(\mathbf{d},\mathbf{e})$ is the part that affects the response variance (e.g., it is $\psi(\mathbf{d},\mathbf{e}) = \boldsymbol{\gamma}^T\mathbf{e} + \mathbf{d}^T\boldsymbol{\Delta}\mathbf{e}$ in (1)). We then reformulate the RC in Theorem 1 as

$$\begin{aligned}\min_{\mathbf{d},\boldsymbol{\lambda},z,\boldsymbol{\eta}\geq\mathbf{0}}\ & \lambda_2 + \rho\eta_2 + \eta_2\sum_{i\in V} q_i\phi^*\left(\frac{(\psi_i(\mathbf{d})+z)^2-\lambda_2}{\eta_2}\right)\\ \text{s.t. }& f(\mathbf{d}) + \lambda_1 + \rho\eta_1 + \eta_1\sum_{i\in V} q_i\phi^*\left(\frac{\psi_i(\mathbf{d})-\lambda_1}{\eta_1}\right) \leq T.\end{aligned}$$

The complexity of $f(\mathbf{d})$ and $\psi_i(\mathbf{d})$ $(=\psi(\mathbf{d},\mathbf{e}^{\mathbf{i}}))$ determines the complexity of the RC; the problem can be non-convex depending on $f(\mathbf{d})$ and $\psi_i(\mathbf{d})$. Again our robust reformulation does introduces additional variables, but does not introduce additional non-convexity to $y(\mathbf{e},\mathbf{d})$.

2.3.2. Signal-to-Noise Ratios SNRs are performance criteria used in many areas including engineering, chemistry, and physics. Taguchi (1986, 1987) focuses on three performance measures in his SNRs. Our method can be applied for other optimization problems including functions of type $(y(\mathbf{e},\mathbf{d})-0)^2$ and $(1/y(\mathbf{e},\mathbf{d}))^2$. Therefore, the robust reformulations of such problems are special cases of Theorem 1, when we have no constraint on the variance. A true SNR is formulated as:

$$\max_{\mathbf{d}} \frac{\mathrm{E}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})]}{\sqrt{\mathrm{Var}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})]}}, \tag{20}$$

which we reformulate as

$$\max_{\mathbf{d},w} \; w \quad \text{s.t.} \; -\mathrm{E}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] + w\sqrt{\mathrm{Var}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})]} \le 0. \tag{21}$$

Like in §2.3.1, we define the response model through $y(\mathbf{e},\mathbf{d}) = f(\mathbf{d}) + \psi(\mathbf{e},\mathbf{d})$. Using the associated notation, the general RC of (21) after discretization is given by

$$(\mathbf{SI2}) \max_{\mathbf{d},\lambda,\eta,w,\mathbf{v}} \; w$$
$$\text{s.t.} \; -f(\mathbf{d}) - \sum_{i\in V}\psi(\mathbf{d})p_i + w\sqrt{\sum_{i\in V}\psi_i(\mathbf{d})^2 p_i - \left[\sum_{i\in V}\psi_i(\mathbf{d})p_i\right]^2} \le 0 \qquad \forall \mathbf{p}\in U.$$

The left-hand side of the inequality in (SI2) is concave in $\mathbf{p}$ when $w \ge 0$. The next theorem provides the tractable RC of (SI2).

Theorem 2 *The vector* $\mathbf{d}$ *and* w *solve (SI2) if and only if* $\mathbf{d},\lambda,\eta,w,$ *and* $\mathbf{v}=[v_1,v_2]$ *solve the following optimization problem:*

$$(\mathbf{RC2}) \max_{\mathbf{d},\lambda,\mathbf{v},\eta\ge 0,w\ge 0} \; w$$
$$s.t. \; -f(\mathbf{d}) + \lambda + \rho\eta + \frac{w^2}{4v_1} + \eta\sum_{i\in V} q_i \phi^*\left(\eta^{-1}\left[\left(\psi_i(\mathbf{d})\sqrt{v_1}+v_2\right)^2 - \psi_i(\mathbf{d}) - \lambda_1\right]\right) \le 0.$$

Proof. We reformulate the semi-infinite problem (SI2) as

$$\max_{\mathbf{p}\in U}\left\{-f(\mathbf{d}) - \sum_{i\in V}\psi(\mathbf{d})p_i + w\sqrt{\sum_{i\in V}\psi_i^2(\mathbf{d})p_i - \left[\sum_{i\in V}\psi_i(\mathbf{d})p_i\right]^2}\right\} \le T. \tag{22}$$

Analogous to the proof of Theorem 1, we reformulate (22) as

$$-f(\mathbf{d}) + \max_{\mathbf{a}\in\widehat{U}} g(\mathbf{a}) \le T, \tag{23}$$

where $\widehat{U} := \{\mathbf{a} : \mathbf{a} = \mathbf{Ap}, \mathbf{p} \in U\}$, $\mathbf{A}^T = [\psi^2(\mathbf{d}), \psi(\mathbf{d})]$ and $g(\mathbf{a}) = -a_2 + w\sqrt{a_1 - a_2^2}$. Using the indicator function

$$\delta(\mathbf{a}|\widehat{U}) := \begin{cases} 0, & \mathbf{a} \in \widehat{U} \\ +\infty, \text{ elsewhere,} \end{cases}$$

we reformulate (23) as

$$-f(\mathbf{d}) + \max_{\mathbf{a}\in\mathbb{R}^2}\{g(\mathbf{a}) - \delta(\mathbf{a}|\widehat{U})\} \leq T.$$

Deleting the minimization in the Fenchel dual of $\max_{\mathbf{a}\in\mathbb{R}^2}\{g(\mathbf{a}) - \delta(\mathbf{a}|\widehat{U})\}$ as in Theorem 1, the RC is equivalent to

$$-f(\mathbf{d}) + \delta^*(\mathbf{v}|\widehat{U}) - g_*(\mathbf{v}) \leq T \tag{24}$$

where the concave conjugate is

$$g_*(\mathbf{v}) = \inf_{\mathbf{a}\in\mathbb{R}^2}\{a_1v_1 + a_2v_2' - g(\mathbf{a})\} = \begin{cases} -[w^2 + (1+v_2')^2]/4v_1, & v_1 > 0 \\ -\infty, & \text{elsewhere,} \end{cases}$$

and $\mathbf{v} = [v_1; v_2']$ denotes the additional dual variables. The convex conjugate of δ is equivalent to

$$\delta^*(\mathbf{A}^T\mathbf{v}|U) = \inf_{\lambda',\eta\geq 0}\left\{\rho\eta + \lambda' + \eta\sum_{i\in V} q_i\phi^*\left(\frac{\psi_i^2(\mathbf{d})v_1 + \psi_i(\mathbf{d})v_2' - \lambda'}{\eta}\right)\right\}$$

where η and λ' are the additional Lagrangian dual variables. Thus (24) becomes

$$-f(\mathbf{d}) + \rho\eta + \lambda' + \eta\sum_{i\in V} q_i\phi^*\left(\frac{\psi_i^2(\mathbf{d})v_1 + \psi_i(\mathbf{d})v_2' - \lambda'}{\eta}\right) + \frac{w^2 + (1+v_2')^2}{4v_1} \leq T$$

$$\eta \geq 0, w \geq 0.$$

Using the substitutions $\lambda = \lambda' + (1+v_2)^2/4v_1$ and $v_2 = v_2' + 1/2\sqrt{v_1}$, the final RC becomes (RC2) ■

Notice that when v_1 is fixed, (RC2) does not introduce extra non-convexity. We can find the optimal v_1 by solving the problem for various values of v_1.

2.3.3. Standard Deviation as Risk Measure Standard deviation is used as a risk measure in finance and engineering to quantify the worst-case risk, e.g., Norman and Smith (2006) optimizes such a measure to find optimal facility design that is subject to material handling uncertainties. In this subsection, we examine the robust reformulation of such a constraint:

$$\mathrm{E}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] + k\sqrt{\mathrm{Var}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})]} \leq T$$

where $k \geq 0$. Analogous to the previous subsection, the general RC is given by

$$(\mathbf{SI3})\ f(\mathbf{d}) + \sum_{i \in V} \psi(\mathbf{d}) p_i + k \sqrt{\sum_{i \in V} \psi_i(\mathbf{d})^2 p_i - \left[\sum_{i \in V} \psi_i(\mathbf{d}) p_i\right]^2} \leq T \qquad \forall \mathbf{p} \in U.$$

Corollary 1 *The vector* $\mathbf{d}$ *satisfies constraint (SI3) if and only if* $\mathbf{d}, \lambda, \eta$, *and* $\mathbf{v} = [v_1, v_2]$ *satisfy the following RC constraints:*

$$f(\mathbf{d}) + \lambda + \rho\eta + \frac{k^2}{4v_1} + \eta \sum_{i \in V} q_i \phi^* \left((\psi_i(\mathbf{d}) \sqrt{v_1} + v_2)^2 + \psi_i(\mathbf{d}) - \lambda_1 \right) \leq T,$$

$$\eta \geq 0.$$

Proof. The proof follows from Theorem 2, when $g(\mathbf{a}) = a_2 + k\sqrt{a_1 - a_2^2}$ ■

2.3.4. Downside and Upside Risk Measures The variance considers upper and lower deviations from the mean; however, the variance may not be a good risk measure when one-sided deviations (e.g., semivariance) are considered as risk. For example, in portfolio optimization, having more return than an expected target (T) cannot be considered as risk, in other words, risk is one-sided. In this subsection, we examine the robust reformulation of one-sided risk measures, namely, downside and upside. The downside risk is formulated as

$$\min_{\mathbf{d}} \mathrm{E}_{\mathbf{e}} \left[\min\{y(\mathbf{e}, \mathbf{d}) - T, 0\}^2 \right]$$

and the robust reformulation of the downside risk is given as

$$(\mathbf{SI4}) \quad \min_{\mathbf{d}, \mathbf{z}} \max_{\mathbf{p} \in U} \sum_{i \in V} z_i^2 p_i \quad \text{s.t.} \quad z_i \leq y_i(\mathbf{d}) - T \ \ \forall i \in V, \ \mathbf{z} \leq \mathbf{0}.$$

Corollary 2 *The vectors* $\mathbf{d}$ *and* $\mathbf{z}$ *solve (SI4) if and only if* $\mathbf{d}, \mathbf{z}, \lambda$, *and* η *solve the following RC problem:*

$$\min_{\mathbf{d}, \mathbf{z} \leq \mathbf{0}, \eta \geq 0, \lambda} \ \lambda + \rho\eta + \eta \sum_{i \in V} q_i \phi^* \left(\frac{-\lambda + z_i^2}{\eta} \right)$$

$$s.t.\ z_i \leq y_i(\mathbf{d}) - T \ \ \forall i \in V.$$

Proof. Since the objective function in (SI4) is linear in $\mathbf{p}$, the proof easily follows from Yanıkoğlu and den Hertog (2013) and Theorem 1 ■

Robust reformulation of the upside risk $[\min_{\mathbf{d}} \mathrm{E}_{\mathbf{e}} [\min\{T - y(\mathbf{e}, \mathbf{d}), 0\}^2]]$ can be similarly derived when constraints $[z_i \leq y_i(\mathbf{d}) - T \ \ \forall i \in V]$ are replaced by $[z_i \leq T - y_i(\mathbf{d}) \ \ \forall i \in V]$ in Corollary 2.

3. Adjustable Robust Optimization

In the preceding sections the controllable factors $\mathbf{d}$ are 'here and now' decisions; i.e., decisions on $\mathbf{d}$ must be made before $\mathbf{e}$ is realized, and hence $\mathbf{d}$ does not depend on the actual values of $\mathbf{e}$. In practice, a part of the controllable factors can often be adjusted after observing the actual values of $\mathbf{e}$. For example, in a multi-stage inventory system with uncertain demand, the decisions on the replenishment orders are made one-at-a-time, and the replenishment order t is placed when the actual demands in periods 1 through $t-1$ are known. Hence it is realistic to allow the replenishment order for period t to be adjusted according to the demands in the preceding periods, even though the upcoming demands remain uncertain. The adjustable factors are called 'wait and see' decisions. Often here-and-now and wait-and-see decisions appear together in the same problem setting.

To model this situation, Ben-Tal et al. (2009, Chapter 14) reformulate adjustable factors as functions of the uncertain parameters as follows:

$$d_j = D_j(\mathbf{x_j}, \mathbf{P_j e}) \quad j \in \{1, \ldots, n\}, \tag{25}$$

where $D_j(.)$ are the so-called decision rules that define the class of functions (e.g., affine functions), $\mathbf{x_j}$ is the vector of coefficient variables to be optimized for the associated function class, and $\mathbf{P_j}$ is the information-base matrix; e.g., if $\mathbf{P_j}$ is a zero matrix, then $\mathbf{d}$ and $\mathbf{e}$ become functionally independent and we are back to here-and-now decisions; if $\mathbf{P_j}$ is a unit matrix, then we allow d_j to depend on all components of $\mathbf{e}$. In addition, d_j can depend on a portion of the observed data on $\mathbf{e}$; e.g., in the multi-stage inventory example, $\mathbf{P_j}$ has the value 1 in the first $j-1$ diagonal elements and zero elsewhere.

To obtain the *adjustable robust counterpart* (ARC) of the mean-variance problem (SI1), we replace $\mathbf{d}$ by $\mathbf{D}(\mathbf{X}, \mathbf{Pe}) = [D_1(\mathbf{x_1}, \mathbf{P_1 e}), \ldots, D_n(\mathbf{x_n}, \mathbf{P_n e})]^T$ in the general RC:

$$\min_{\mathbf{X}} \max_{\mathbf{p} \in U} \sum_{i \in V} y_i \left(\mathbf{D}(\mathbf{X}, \mathbf{Pe^i})\right)^2 p_i - \left[\sum_{i \in V} y_i \left(\mathbf{D}(\mathbf{X}, \mathbf{Pe^i})\right) p_i\right]^2$$

$$\text{s.t.} \sum_{i \in V} [\beta_0 + \boldsymbol{\beta}^T \mathbf{D}(\mathbf{X}, \mathbf{Pe^i}) + \mathbf{X}, \mathbf{Pe^i})^T \mathbf{B}\, \mathbf{D}(\mathbf{X}, \mathbf{Pe^i}) + (\boldsymbol{\gamma}^T + \mathbf{D}(\mathbf{X}, \mathbf{Pe^i})^T \Delta) \mathbf{e}^i] p_i \leq T \quad \forall \mathbf{p} \in U.$$

The tractable ARC of the above problem results from Theorem 1 in §2.2. The adjustable reformulation has the following advantages as we shall detail below. First, the optimal solution of ARC is less conservative than that of the non-adjustable RC, since it is flexible in adjusting the robust optimal decisions according to revealed data. Second, our approach

is tractable even for nonlinear decision rules because we translate the uncertainty from $\mathbf{e}$ to $\mathbf{p}$. Notice that, in classic *adjustable robust optimization* (ARO) tractability is generally scarce for non-linear decision rules; Ben-Tal et al. (2009, pp. 382-387) show that the explicit ARC for a quadratic decision rule is derivable only for an ellipsoidal uncertainty set, and the resulting ARC is an SDP problem. Third, we propose a new decision rule that enables us to model integer adjustable decision variables.

To present the associated decision rules we use the following illustrative example:

$$\min_{\mathbf{d}} \mathrm{E}_{\mathbf{e}} \left[(1 + 5d_1 + 5d_2 + e_1 - e_2)^2 + (1 + 5d_1 + 10d_2 + e_1 + e_2)^2 \right], \tag{26}$$

where all factors are coded such that $-1 \leq \mathbf{d} \leq 1$ and $-1 \leq \mathbf{e} \leq 1$. We divide the support of $\mathbf{e}$—namely, the unit box in two dimension $[-1, 1]^2$—into four cells of equal size. Hence the center points of the cells are $\{\mathbf{e^1}, \mathbf{e^2}, \mathbf{e^3}, \mathbf{e^4}\}$ ={(-0.5, -0.5), (-0.5, 0.5), (0.5, -0.5), (0.5, 0.5)}). So the nominal problem after discretization is

$$\min_{\mathbf{d}} \sum_{i=1}^{4} \left[\left(1 + 5d_1 + 5d_2 + e_1^i - e_2^i\right)^2 + \left(1 + 5d_1 + 10d_2 + e_1^i + e_2^i\right)^2 \right] q_i. \tag{27}$$

Suppose the observed data are $\mathbf{q} = [0.4, 0.3, 0.2, 0.1]^T$. It is easy to derive the optimal solution of the corresponding *nominal* problem: (d_1, d_2) =(-0.08, -0.08). The uncertainty set for the RC is given by

$$\mathcal{P} := \left\{ \mathbf{p} = (p_1, p_2, p_3, p_4) \in \mathbb{R}^4 \middle| p_1 + p_2 + p_3 + p_4 = 1, \sum_{i=1}^{4} \frac{(p_i - q_i)^2}{p_i} \leq 0.5, \mathbf{p} \geq 0 \right\}.$$

Using this uncertainty set, we derive that the worst-case objective value for the nominal solution is 1.2; moreover, the non-adjustable robust counterpart of (26) is

$$\min_{\mathbf{d}} \max_{\mathbf{p} \in \mathcal{P}} \sum_{i=1}^{4} \left[\left(1 + 5d_1 + 5d_2 + e_1^i - e_2^i\right)^2 + \left(1 + 5d_1 + 10d_2 + e_1^i + e_2^i\right)^2 \right] p_i,$$

and its robust optimal solution is (-0.2, 0) with objective value 1.0, which is lower than the worst-case nominal objective value 1.2.

Next we shall examine three adjustable formulations; namely, linear, nonlinear, and cell-based decision rules.

Linear Decision Rule Now we assume that the functions of the adjustable controllable factors $\mathbf{d}$ are linear in the observed values of the environmental factors $\mathbf{e}$. The *fully adjustable* linear decision rule is

$$d_j = D_j(\mathbf{x_j}, \mathrm{I}_2\mathbf{e}) := x_{j0} + x_{j1}e_1 + x_{j2}e_2 \quad \forall j \in \{1,2\} \tag{28}$$

where $\mathrm{I}_2 = [1,0;0,1]$. When both controllable factors are fully adjustable, all decisions are made only after the uncertainty has been revealed. Consequently, the ARC of (26) with linear decision rule is

$$\min_{\mathbf{X}} \max_{\mathbf{p}\in\mathcal{P}} \sum_{i=1}^{4} \Bigg[\left(1 + 5D_1(\mathbf{x_1}, \mathbf{e^i}) + 5D_2(\mathbf{x_2}, \mathbf{e^i}) + e_1^i - e_2^i\right)^2 + \left(1 + 5D_1(\mathbf{x_1}, \mathbf{e^i}) + 10D_2(\mathbf{x_2}, \mathbf{e^i}) + e_1^i + e_2^i\right)^2 \Bigg] p_i.$$

The optimal solution is $\mathbf{x_1}^*=(-0.2,-0.2,0.6)^T$ and $\mathbf{x_2}^*=(0,0,-0.4)^T$; i.e., the linear decision rules are $D_1(\mathbf{x_1},\mathbf{e})=-(1+e_1-3e_2)/5$ and $D_2(\mathbf{x_2},\mathbf{e})=-2e_2/5$, so we have

$$\begin{aligned} 5d_1 + 5d_2 &= 5D_1(\mathbf{x_1},\mathbf{e}) + 5D_2(\mathbf{x_2},\mathbf{e}) = -1 - e_1 + e_2 \\ 5d_1 + 10d_2 &= 5D_1(\mathbf{x_1},\mathbf{e}) + 10D_2(\mathbf{x_2},\mathbf{e}) = -1 - e_1 - e_2. \end{aligned}$$

Therefore, the ARC yields the lowest possible optimal objective value; namely, zero for the problem, whereas the non-adjustable RC yields one.

More interesting cases have wait-and-see and here-and-now decisions together or at least one of the controllable factors is not fully adjustable. Table 2 presents the numerical results for all possible combinations of linear decision rules. The first column gives the possible linear decision rules for the two adjustable factors d_1 and d_2, where 'na' denotes *non-adjustable*, 'e1' denotes a factor that is adjustable on e_1; similarly, 'e2' denotes adjustability on e_2, and 'e[1,2]' denotes a *fully adjustable* factor. The second and third columns are the optimal coefficients (variables) $\mathbf{x_1}$ and $\mathbf{x_2}$ of the decision rules $D_1(.)$ and $D_2(.)$, where $(-)$ denotes a variable that vanishes in the associated decision rule. The final column (Obj.) presents the robust optimal objective value for the associated decision rule. Altogether, the numerical results show that when one of the factors is non-adjustable and the other is adjustable on e_2—see row (na, e2) or (e2, na) in Table 2—the optimal objective value of the ARC is the same as that of the non-adjustable RC. In all other cases the optimal objective

Table 2 Linear Decision Rules (LDR) and Objective Values

LDR		$D_1(.)$	$D_2(.)$	Obj.
d_1	d_2	(x_{10}, x_{11}, x_{12})	(x_{20}, x_{21}, x_{22})	
na	na	(-0.2,−,−)	(0,−,−)	1.00
na	e1	(-0.189,−,−)	(-0.009,-0.103,−)	0.66
na	e2	(-0.2,−,−)	(0,−,0)	1.00
e1	na	(-0.2,-0.2,−)	(0,−,−)	0.50
e2	na	(-0.2,−,0)	(0,−,−)	1.00
na	e[1,2]	(-0.186,−,−)	(-0.008,-0.12,-0.04)	0.62
e[1,2]	na	(-0.2,-0.2,0)	(0,−,−)	0.50
e1	e1	(-0.2,-0.2,−)	(0,−,−)	0.50
e1	e2	(-0.2,-0.2,0)	(0,−,-0.04)	0.45
e2	e1	(-0.21,−,0.04)	(-0.001,-0.09,−)	0.68
e2	e2	(-0.2,−,0.6)	(0,−,-0.4)	0.50
e[1,2]	e1	(-0.2,-0.2,0)	(0,−,−)	0.50
e[1,2]	e2	(-0.2,-0,2,0.6)	(−,−,0.4)	0.00
e1,	e[1,2]	(-0.2,-0.2,−)	(0,0,-0.04)	0.45
e2,	e[1,2]	(-0.2,−,0.6)	(0,-0.12,-0.4)	0.05
e[1,2]	e[1,2]	(-0.2,-0.2,0.6)	(0,0,-0.4)	0.00

value of the non-adjustable RC improves with at least 32% (see row (e2, e1)) for the ARC, and the highest improvement (100%) is attained when the first factor is fully adjustable and the second one is non-adjustable; see row (e[1,2], e2). Another interesting conclusion is that introducing an adjustable factor into the problem may change the optimal decision for the non-adjustable factor; i.e., an optimal here-and-now factor can have different values in the ARC and RC. For example, if d_1 is adjustable on e_1 and d_2 is non-adjustable, then the optimal d_2 is -0.189 in the ARC, but it is -0.2 in the RC; see (na, na) and (na, e1).

Nonlinear Decision Rule Table 3 shows the following nonlinear (quadratic) decision rule:

$$D_j(\mathbf{x_j}, \mathrm{I}_2\mathbf{e}) := x_{j0} + x_{j1}e_1 + x_{j2}e_2 + x_{11}^{(j)}e_1^2 + x_{22}^{(j)}e_2^2 + x_{12}^{(j)}e_1e_2, \tag{29}$$

where $\mathbf{x_j} = [x_{j0}, x_{j1}, x_{j2}, x_{11}^{(j)}, x_{12}^{(j)}, x_{22}^{(j)}]$. For example, if the controllable factors are only adjustable in e_1, then the decision rule is

$$D_j(\mathbf{x_j}, P_j\mathbf{e}) := x_{j0} + x_{j1}e_1 + x_{11}^{(j)}e_1^2,$$

where $P_j = [1, 0; 0, 0]$. Obviously, the nonlinear decision rule is more general than the linear rule, so it is at least as good as the linear decision rule used in Table 2.

Table 3 Nonlinear Decision Rules (NDR) and Objective Values

NDR		$D_1(.)$	$D_2(.)$	Obj.
d_1	d_2	$(x_{10}, x_{11}, x_{12}, x_{11}^{(1)}, x_{22}^{(1)}, x_{12}^{(1)})$	$(x_{20}, x_{21}, x_{22}, x_{11}^{(2)}, x_{22}^{(2)}, x_{12}^{(2)})$	
na	na	(-0.2,-,-,-,-,-)	(0,-,-,-,-,-)	1.00
na	e1	(-0.196,-,-,-,-,-)	(0.014,0.093,-,-0.079,-,-)	0.65*
na	e2	(-0.2,-,-,-,-,-)	(-0.33,-,0,-,1.33,-)	1.00
e1	na	(-0.195,-0.2,-,-0.02,-,-)	(0,-,-,-,-,-)	0.50
e2	na	(-0.192,-,0,-,-,-)	(-,-0.03,-,-,-,-)	1.00
na	e[1,2]	(-0.188,-,-,-,-,-)	(0.062,-0.106,-0.053,-0.244,0.015,0.05)	0.58*
e[1,2]	na	(-0.178,-0.2,0,-0.044,-0.044,0)	(0,-,-,-,-,-)	0.50
e1	e1	(-0.213,-0.2,-,0.05,-,-)	(-0.002,0,-,0.008,-,-)	0.50
e1	e2	(-0.533,-0.2,-,1.33,-,-)	(-0.005,-,-0.04,-,0.02,-)	0.45
e2	e1	(-0.19,-,0.044,-,-0.047,-)	(-0.045,-0.063,-,0.197,-,-)	0.65*
e2	e2	(-0.223,-,0.6,-,0.094,-)	(-0.002,-,-0.4,-0.008,-)	0.50
e[1,2]	e1	(-0.186,-0.2,0,-0.028,0.028,0)	(-0.006,-,-,0.023,-,-)	0.50
e[1,2]	e2	(-0.253,-0.2,0.6,0.107,0.107,0)	(-0.002,-,-0.4,-,0.008,-)	0.00
e1	e[1,2]	(-0.146,-0.199,-,-0.214,-,-)	(-0.052,-0.003,-0.047,0.148,0.057,-0.001)	0.44*
e2	e[1,2]	(-0.233,0,0.6,-,0.133,0)	(-0.151,-0.12,-0.4,0.105,0.5,-)	0.05
e[1,2]	e[1,2]	(-0.214,-0.2,0.6,0.028,0.028,0)	(0.013,0,-0.4,-0.066,0.014,0)	0.00

(∗) denotes an improved optimal objective value compared with that in Table 2

We denote the cases where the nonlinear decision rule performs better than the linear by (*) in the last column of Table 3. The highest improvement compared with Table 2 is obtained when the first factor is non-adjustable and the second factor is fully adjustable;

compare (na, e[1,2]) in Table 3, where the optimal objective value is 0.584, with the objective 0.621 for the same situation in Table 2. Quantifying the *value of information* is an important topic in both adjustable robust decision making and in general decision making; Table 3 shows that having extra information on e_2 but not on e_1 for one of the controllable factors has no added value in the adjustable decision making; i.e., the non-adjustable and the adjustable RCs have the same optimal objective at (na, e2) and (e2, na). On the other hand, having information on e_1 for one of the controllable factors yields improvement in the objective; see (na, e1) and (e1, na) in Table 3. Moreover, if d_1 responds to both environmental factors, and d_2 uses information on e_2 only, then we obtain the lowest possible optimal objective value (namely, zero); see (e[1,2], e2) in Table 3.

Cell-Based Decision Rule Now we propose a new type of decision rule that we call the *cell-based* decision rule:

$$D_j(\mathbf{x_j}, \mathbf{e}) := x_{ji} \text{ if } \mathbf{e} \in \text{cell}(i), i \in V_j, \tag{30}$$

where $\mathbf{x_j}$ is the decision vector for the jth adjustable variable (x_{ji} being the decision for the ith cell), cell(i) is the region determined by the ith cell, and V_j is the set of cell indices for the associated information-base.

Remark 3 *Cells used in the decision rule are non-intersecting—excluding the boundaries—squares in two dimensions, cubes in three dimensions, and hypercubes in higher dimensional uncertainty spaces. For the sake of simplicity, we assume that—given the dimension—all cells have the same volume.*

If in the illustrative example (26), d_1 is fully adjustable, then the decision rule is

$$D_1(\mathbf{x_1}, \mathbf{e}) := \begin{cases} x_{11}, \ \mathbf{e} \in \text{cell}(1) \\ x_{12}, \ \mathbf{e} \in \text{cell}(2) \\ x_{13}, \ \mathbf{e} \in \text{cell}(3) \\ x_{14}, \ \mathbf{e} \in \text{cell}(4), \end{cases} \tag{31}$$

where $\text{cell}(i) := \{(e_1, e_2) \in \mathbb{R}^2 : \ell_{1i} \leq e_1 \leq u_{1i}, \ell_{2i} \leq e_2 \leq u_{2i}\}$ $(i \in \{1,2,3,4\})$, $\boldsymbol{\ell_1} = [0, -1, -1, 0]$, $\boldsymbol{u_1} = [1, 0, 0, 1]$, $\boldsymbol{\ell_2} = [0, 0, -1, -1]$, $\boldsymbol{u_2} = [1, 1, 0, 0]$, and $V_1 = \{1, 2, 3, 4\}$; cells are represented by their center points in (27). To show the difference between full and

partial information, we assume that d_1 is adjustable on e_1 but not on e_2. The associated decision then becomes

$$D_1(\mathbf{x_1}, e_1) := \begin{cases} x_{11}, \ e_1 \in \text{cell}(1) \\ x_{12}, \ e_1 \in \text{cell}(2), \end{cases} \tag{32}$$

where $\text{cell}(1) := \{e_1 \in \mathbb{R} : 0 \leq e_1 \leq 1\}$ and $\text{cell}(2) := \{e_1 \in \mathbb{R} : -1 \leq e_1 \leq 0\}$, and $V_1 = \{1, 2\}$). It is easy to see that (32) implies that when e_2 is extracted from the information-base, the new cells are projections from the cells in the two-dimensional space in (31) onto the one-dimensional space on e_1 in (32). The disadvantage of the cell-based decision rule is that this rule often has more variables compared with the linear and nonlinear decision rules, especially when the number of cells is high. Nevertheless the numerical results for the example show that the new decision rule is better than the linear, and is 'almost' as good as the nonlinear decision rule—even when the total number of cells is only four; see Table 4.

To the best of our knowledge, parametric decision rules in the RO literature cannot handle adjustable *integer* variables, since the adjustable decision is a function of the uncertain parameter $\mathbf{e}$, and the function does not necessarily take integer values for all $\mathbf{e}$; see (28) and (29). However, our cell-based decision rule can handle such variables. As we can see from (30), the adjustable decision x_{ji} can take integer values since the cell-based decision rule relates $\mathbf{e}$ and x_{ij} through an 'if' statement. Therefore, if we make x_{ij} an integer variable, then the cell-based decision rule gives integer decisions. Note that decision rules for discrete variables have also been proposed by Bertsimas and Caramanis (2010) and Vayanos et al. (2011) but they propose different decision rules for different settings. Now using the illustrative example, we show the viability of our approach for adjustable integer variables. We modify the old example in the following way: We assume d_1 and d_2 are adjustable, and they take values that are multiples of 1/4.

Table 5 presents optimal decision rules and resulting objective values for all possible combinations of adjustability. These numerical results show that we obtain important improvements for the non-adjustable optimal objective value by using the cell-based decision rule. As may be anticipated, the integer formulation yields higher (worse) optimal objective values. For example, we can no longer get a zero objective in Table 5. Moreover, in contrast to the continuous case, we can no longer improve the optimal objective of the non-adjustable RC at decision rule combinations: (na, e1), (na, e[1,2]), and (e2, e1); see the corresponding rows of Tables 4 and 5.

Table 4 Cell-Based Decision Rules (CDR) and Objective Values

CDR		$D_1(.)$	$D_2(.)$	Obj.
d_1	d_2	$(x_{11}, x_{12}, x_{13}, x_{14})$	$(x_{21}, x_{22}, x_{23}, x_{24})$	
na	na	(-0.2,−,−,−)	(0,−,−,−)	1.00
na	e1	(-0.19,−,−,−)	(-0.06,0.04,−,−)	0.66
na	e2	(-0.2,−,−,−)	(0,0,−,−)	1.00
e1	na	(-0.3,-0.1,−,−)	(0,−,−,−)	0.50
e2	na	(-0.2,-0.2,−,−)	(0,−,−,−)	1.00
na	e[1,2]	(-0.19,−,−,−)	(-0.09,0.03,0.07,-0.05)	0.62
e[1,2]	na	(-0.3,-0.1,-0.1,-0.3)	(0,−,−,−)	0.50
e1	e1	(-0.3,-0.1,−,−)	(0,0,−,−)	0.50
e1	e2	(-0.3,-0.1,−,−)	(-0.02,0.02,−,−)	0.45
e2	e1	(-0.18,-0.2,−,−)	(-0.06,0.04,−,−)	0.65
e2	e2	(0.1,-0.5,−,−)	(-0.2,0.2,−,−)	0.50
e[1,2]	e1	(-0.3,-0.1,-0.1,-0.3)	(0,0,−,−)	0.50
e[1,2]	e2	(0,0.2,-0.4,-0.6)	(-0.2,0.2,−,−)	0.00
e1	e[1,2]	(-0.3,-0.1,−,−)	(-0.02,-0.02,0.02,0.02)	0.45
e2	e[1,2]	(0.1,-0.5,−,−)	(-0.26,-0.14,0.26,0.14)	0.05
e[1,2]	e[1,2]	(0,0.2,-0.4,-0.6)	(-0.2,-0.2,0.2,0.2)	0.00

Computational Complexity of Decision Rules In adjustable robust dual response optimization, the uncertainty is transferred from the uncontrollable factors ($\mathbf{e}$) to probability vector ($\mathbf{p}$) via discretization. This is why, unlike the classic decision rule approach in ARO (Ben-Tal et al. 2009, Chapter 14), switching to nonlinear decision rules from linear decision rules does not affect the complexity of the optimization problem structure. Nevertheless, as we mentioned before, the overall computational complexity of the decision rule approach increases exponentially in the number of uncertain parameters.

4. Realistic Examples

In this section, we present realistic examples to demonstrate the effectiveness of our methods. We use a 64-bit Windows PC with a 2.2 GHz Intel Core i7 processor, and 8 GB of

Table 5 Integer Cell-Based Decision Rules and Objective Values

CDR		$D_1(.)$	$D_2(.)$	Obj.
d_1	d_2	$(x_{11}, x_{12}, x_{13}, x_{14})$	$(x_{21}, x_{22}, x_{23}, x_{24})$	
na	na	(-0.25,−, −, −)	(0,−, −, −)	1.41
na	e1	(-0.25,−, −, −)	(0,0,−, −)	1.41
na	e2	(-0.25,−, −, −)	(0,0,−, −)	1.41
e1	na	(-0.25,0,−, −)	(0,−, −, −)	0.92
e2	na	(-0.25,-0.25,−, −)	(0,−, −, −)	1.41
na	e[1,2]	(-0.25,−, −, −)	(0,0,0,0)	1.41
e[1,2]	na	(-0.25,0,0,-0.25)	(0,−, −, −)	0.92
e1	e1	(-0.25,0,−, −)	(0,0,−, −)	0.92
e1	e2	(-0.25,0,−, −)	(0,0,−, −)	0.92
e2	e1	(-0.25,-0.25,−, −)	(0,0,−, −)	1.41
e2	e2	(0.25,-0.5,−, −)	(-0.25,0.25,−, −)	1.29
e[1,2]	e1	(-0.25,0,0,-0.25)	(0,0,−, −)	0.92
e[1,2]	e2	(0,0.25,-0.50,-0.75)	(-0.25,0.25,−, −)	0.26
e1	e[1,2]	(-0.25,0,−, −)	(0,0,0,0)	0.92
e2	e[1,2]	(0,-0.50,−, −)	(-0.25,0,0.25,0.25)	0.99
e[1,2]	e[1,2]	(0,0.25,-0.50,-0.75)	(-0.25,-0.25,0.25,0.25)	0.26

RAM. To solve the mathematical optimization problems, we use KNITRO 8.0 embedded in MATLAB (2012a) and AIMMS 3.12.

4.1. Television Images

In color televisions the quality of signals is determined by the power signal-to-noise-ratios in the image transmitted. We take the response function $y(\mathbf{d}, \mathbf{e})$ from Myers et al. (2009, p. 512), where the response y measures the quality of transmitted signals in decibels. The controllable factors are the number of tabs in a filter d_1, and the sampling frequency d_2; the environmental factors are the number of bits in an image e_1, and the voltage applied e_2. The least-square estimate of the metamodel is

$$\hat{y}(\mathbf{d}, \mathbf{e}) = 33.389 - 4.175d_1 + 3.748d_2 + 3.348d_1d_2 - 2.328d_1^2 - 1.867d_2^2$$

$$-4.076e_1+2.985e_2-2.324d_1e_1+1.932d_1e_2+3.268d_2e_1-2.073d_2e_2,$$

where all factors are coded; for details on the DoE we refer to Myers et al. (2009, pp. 511–515).

We find the optimal design settings of d_1 and d_2 for the optimization problem:

$$\begin{aligned} &\max_{\mathbf{d}} \ \mathrm{E}_{\mathbf{e}}[\hat{y}(\mathbf{d},\mathbf{e})] \\ &\text{s.t. } \mathrm{Var}_{\mathbf{e}}[\hat{y}(\mathbf{d},\mathbf{e})] \le T. \end{aligned} \tag{33}$$

The robust counterpart of (33) can be derived similar to Theorem 1 in §2.2. To estimate $\mathbf{q}$, we use the "historical" data in Figure 1; because we have no real data, we have randomly created these data. The sample size is $N=350$, and the support of $\mathbf{e}$ is divided into 25 cells of the same volume so $V=\{1,\ldots,25\}$, $\mathbf{q}=[q_1,\ldots,q_{25}]^T$ and $\rho=\chi^2_{0.999,24}/350$; see (8). We use the same data in our two realistic examples so the data do not favor our method.

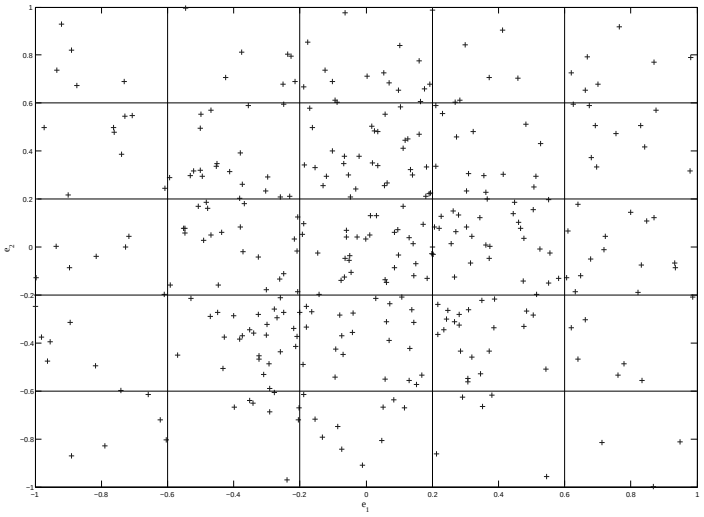

Figure 1 Historical Data on e

The goal of these examples is to compare the optimal solutions of the nominal and robust counterpart problems. In §4.1.1 we compare the worst-case and average performances of these two solutions via the objective value and the constraint violation. In §4.1.2 we compare the confidence level covarage probabilities of the nominal and robust optimal solutions.

4.1.1. Robust versus Nominal Solutions We vary the right-hand side value T of the variance constraint from 0.1 to 0.8; see column one in Table 6. We solve the nominal and the RC problems for these T values, and compare the worst-case performances of the nominal and robust optimal solutions.

Table 6 Worst-Case Analysis for TV Image Example

T	Robust			Nominal		
	(d_1^*, d_2^*)	y^*	Var.(%)	(d_1^*, d_2^*)	W-C($\mathbb{E}$)	W-C(Var)(%)
0.1	(-0.43,0.83)	35.27	0.10(0%)	(-0.4472,0.7755)	35.27	0.19(88%)
0.2	(-0.44,0.79)	35.28	0.15(-25%)	(-0.4625,0.6853)	35.26	0.37(87%)
0.3	(-0.44,0.79)	35.28	0.15(-50%)	(-0.4763,0.6152)	35.23	0.56(86%)
0.4	(-0.44,0.79)	35.28	0.15(-62%)	(-0.4867,0.5648)	35.20	0.71(79%)
0.5	(-0.44,0.79)	35.28	0.15(-70%)	(-0.4867,0.5648)	35.20	0.71(43%)
0.6	(-0.44,0.79)	35.28	0.15(-75%)	(-0.4867,0.5648)	35.20	0.71(19%)
0.7	(-0.44,0.79)	35.28	0.15(-78%)	(-0.4867,0.5648)	35.20	0.71(2%)
0.8	(-0.44,0.79)	35.28	0.15(-81%)	(-0.4867,0.5648)	35.20	0.71(-11%)

Columns two and three are the robust optimal solution (d_1^*, d_2^*) and its objective value (y^*). Column four (Var.) is the robust variance of the response. Column five is the nominal optimum solution. Column six presents the mean ($\mathbb{E}_{\mathbf{e}}[y(\mathbf{e}, \mathbf{d})]$) when the decision factors are fixed to the nominal optimum solution and the worst-case probability vector that minimizes the expectation is realized. Column seven (W.-C.(Var)) gives the worst-case value of the variance for the nominal solution; now we consider the probability vector that maximizes the variance, as the worst-case. Notice that we also report the percentages of the worst-case constraint violations for the nominal and robust optimal solutions; these values are shown within brackets (%) in the columns four and seven. We use the formula $[(\text{W.-C.(Var)-}T) \times 100/T]$ to calculate the constraint violation percentage of the nominal solution when the worst-case uncertain parameters are realized. If the violation percentage is negative, then the constraint is satisfied; 0% means the constraint is binding (see column four, row T=0.1); a positive percentage means constraint violation at column seven.

Table 6 reveals that the robust reformulation of the variance constraint becomes redundant when $T \geq 0.2$; e.g., the left-hand side value of the constraint for the robust optimal

solution (-0.4439,0.7988) when $T = 0.2$ is 0.149. Therefore, we have the same robust optimal solution for $T \geq 0.2$. The nominal variance constraint (not the worst-case) also becomes redundant when $T \geq 0.389$, and therefore we have the same nominal optimum solution when $T \geq 0.4$. The robust optimal objective values and the worst-case objective values of the nominal optimum solutions are within 1% of each other. However, the constraint violation percentages favor our robust approach; e.g., the percentages in column seven show that when $T \leq 0.4$, then on average, the nominal optimal solution violates the constraint 85% in the worst-case. When $T \geq 0.715$ the nominal optimum solution no longer violates the constraint in the worst-case, but it is closer to be binding than the robust solution. Altogether, in this example our robust optimization method creates immunity to the worst-case uncertainty, without being penalized by the objective.

Now we analyze the average mean ($\text{Avg}(\mathbb{E}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})])$) and variance ($\text{Avg}(\text{Var}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})])$) of the response for the nominal and robust optimal solutions, randomly sampling 1,000 probability vectors $\mathbf{p}$ from the uncertainty set that is used in the RC; see (9). First we uniformly sample 25 individual probabilities ($\hat{p}_i$). Then to guarantee that the probabilities sum up to one, we use the adjustment: $p_i = \hat{p}_i / \sum_{j=1}^{25} \hat{p}_j$. Finally, if the probability vector $\mathbf{p}$ is within the uncertainty set, we keep $\mathbf{p}$ in the sample; else we discard $\mathbf{p}$ and sample again. We repeat this procedure until we have 1,000 accepted vectors that are within the set. The confidence level of the uncertainty set (9) is 99.9%; i.e., α in (9) is 0.001. From the $\mathbf{p}$ in the sample we estimate the mean in (11) and the variance in (13) for a given solution, and then take the averages; see Table 7.

Table 7 Average (Avg) Mean ($\mathbb{E}$) and Variance (Var)

T	Robust		Nominal	
	Avg($\mathbb{E}$)	Avg(Var)	Avg($\mathbb{E}$)	Avg(Var)
0.1	35.37	0.064	35.41	0.12
0.2	35.40	0.095	35.45	0.24
0.3	35.40	0.095	35.46	0.35
0.4	35.40	0.095	35.47	0.46

Table 7 shows that the average expected responses for the nominal solutions are less than 1% better than that of the robust solutions (we maximize $\mathbb{E}(y)$; see (26)). Nevertheless, the average response variances of the robust optimal solutions are more than 100% better (smaller) than that of the nominal solutions when $T \leq 0.2$, and the improvement becomes more than 350% when $T \geq 0.3$. Moreover, on average the nominal solutions violate 5% of the variance constraint, whereas the robust solutions are not at all binding. The effectiveness of our approach is further analyzed in the following subsection.

4.1.2. Confidence Level The uncertainty set (9) coincides with the $(1-\alpha)$-confidence set for the unknown probability vector $\mathbf{p}$ centered at the empirical estimate $\mathbf{q}$. In the numerical examples we use $\alpha = 0.001$, which means that the robust optimal solution is immune to uncertainty with at least 99.9% probability. These uncertainties are ignored in the nominal problem; nevertheless, the nominal solution may be robust for some part of the uncertainty set. To find the confidence level of a nominal solution, we calculate the tightest uncertainty set for which the nominal solution is robust; i.e., we adjust the ρ value in (9); see Table 8.

Table 8 Confidence Levels (1-α) of Nominal Solutions

T	≤ 0.4	0.5	0.6	0.7	≥ 0.8
$(1-\alpha)$	0%	2%	70%	98%	99.9%

Table 8 shows that when $T \leq 0.4$ the nominal solutions have no immunity to uncertainty at all; i.e., a 'small' change in the given empirical estimate $\mathbf{q}$ results in infeasibility in the nominal variance constraint. We may anticipate this result, since the nominal solutions are binding for the associated cases. On the other hand, for $T = 0.5$ to $T = 0.7$ we see important improvement in the immunity to uncertainty for the nominal solution. In addition, when $T \geq 0.8$, the immunity of the nominal optimum solution is as good as the robust optimum solution—even though the nominal optimum solution is closer to be binding in the constraint than the robust solution. Concerning the objective, we have already shown in the last row of Table 6 that the robust and the 'worst-case' nominal optimum objective values are almost the same; i.e., the robust solutions are less than 1% better than the nominal solutions. In conclusion, the robust reformulation improves the immunity to uncertainty when $T \leq 0.8$ and the improvement is even better when $T \leq 0.4$. Additionally, we prefer the robust solution when $T \geq 0.8$, since it performs better in the worst-case.

4.2. Distribution Center

In this example we focus on the simulation of a cross-docking distribution center (CDDC) developed by Shi et al. (2014). The CDDC is used by a Chinese automobile manufacturer that needs to improve the physical flow in its supply chain. The objective of the company is steady production. The challenge in such production is the supply uncertainty due to environmental factors, discussed above (34). To model the associated supply chain system, Shi proposes a hybrid approach that combines discrete-event simulation, RSM, and Taguchi's world view. We focus on Shi's linear regression metamodel for the total throughput with five controllable factors; namely, the number of receiving doors d_1, the number of shipping doors d_2, the number of forklifts d_3, the number of conveyors d_4, and the supply chain threshold in storage d_5. The model has two environmental factors; namely, the variability in less-than-truckload supply shipments e_1, and the production interruption or delay probabilities of the suppliers e_2. Using the simulation's I/O data, Shi approximates the unknown I/O function of the total throughput by the following least-squares estimate $\hat{y}$:

$$\begin{aligned}\hat{y}_{\mathrm{TT}}(\mathbf{e},\mathbf{d}) =& 10^4 \times \Big[47.97 + 3.982d_1 + 2.025d_2 - 0.031d_3 + 0.734d_4 + 0.034d_5 \\ &+ 0.789d_1d_2 + 0.012d_1d_3 + 0.003d_1d_4 - 0.002d_1d_5 + 0.0007d_2d_3 \\ &- 0.065d_2d_4 - 0.1131d_2d_5 - 0.078d_3d_4 - 0.041d_3d_5 + 0.11d_4d_5 \\ &- 3.406d_1^2 - 1.781d_2^2 + 0.011d_3^2 - 1.033d_4^2 + 0.111d_5^2 \\ &+ (16.66 + 1.511d_1 + 2.374d_2 - 0.059d_3 + 0.824d_4 - 0.093d_5)e_1 \\ &- (0.005 + 0.27d_1 + 0.661d_2 - 0.086d_3 + 0.335d_4 - 0.005d_5)e_2\Big],\end{aligned} \tag{34}$$

where all factors are coded such that $-1 \leq \mathbf{d} \leq 1$, and $-1 \leq \mathbf{e} \leq 1$. More precisely, the coded controllable factors are between -1 and 1 because of the physical restrictions of the production facility. Shi et al.'s analysis of variance (ANOVA) shows that the metamodel $\hat{y}_{\mathrm{TT}}$ have non-significant lack-of-fit; and for the estimated individual parameters the level-of-significance is 0.05. Using Shi's response model (34), we focus on the robust reformulation of the following optimization problem:

$$\begin{aligned}\min_{\mathbf{1}\leq\mathbf{d}\leq\mathbf{1}} \ & \mathrm{Var}_{\mathbf{e}}[\hat{y}(\mathbf{d},\mathbf{e})] \\ \text{s.t. } & E_{\mathbf{e}}[\hat{y}[\mathbf{d},\mathbf{e})] \geq T.\end{aligned} \tag{35}$$

To estimate the frequencies $\mathbf{q}$ used by the nominal and RC problems, we use the same historical data as in Figure 1.

Table 9 Worst-Case Comparison for CDDC Example

T	Robust			Nominal (Worst-Case)		
	$\sigma_{\mathbf{e}}(\hat{y})$	$\mathbb{E}_{\mathbf{e}}[\hat{y}]$	$\mathbb{E}_{\mathbf{e}}[\hat{y}] \pm 3\hat{\sigma}_{\mathbf{e}}$	$\sigma_{\mathbf{e}}(\hat{y})$	$\mathbb{E}_{\mathbf{e}}[\hat{y}]$	$\mathbb{E}_{\mathbf{e}}[\hat{y}] \pm 3\hat{\sigma}_{\mathbf{e}}$
$\leq$339	67	339	(136, 541)	67	339	(136, 541)
340	67	340	(137, 542)	67	339	(136, 541)
370	70	370	(158, 581)	68	348	(143, 552)
400	74	400	(176, 623)	71	377	(163, 590)
420	79	420	(182, 657)	73	396	(174, 617)
430	81	430	(184, 675)	75	405	(178, 632)
440	85	440	(184, 695)	77	414	(181, 648)
450	88	450	(183, 716)	80	424	(183, 664)

$***$ all entries should be multiplied by 1,000

Table 9 compares the worst-case performances of the nominal and robust optimal solutions. Besides the worst-case mean and standard deviation, we construct the worst-case confidence interval for $\hat{y}$; namely, $\mathbb{E}_{\mathbf{e}}[\hat{y}] \pm 3\hat{\sigma}_{\mathbf{e}}$. Because the probability distribution of the environmental factors is unknown, we cannot say much about the coverage probability of the confidence interval. The numerical results show that in the worst-case the nominal solutions are on average 6% lower than the target T for the expected total throughput; i.e., the average violation of the constraint is 6%. The worst-case standard deviations of the total throughput for the nominal solutions are on average 8% lower than that of the robust ones. Nevertheless the confidence intervals for the robust optimal solutions are always shifted to the right compared with the nominal solutions, which is in favor of the robust approach since a higher total throughput is better. Altogether the numerical results favor the robust approach when $T >$339,000; when $T \leq$ 339,000 both methods yield the same outcome.

In Table 10 we compare the average performance of robust and nominal solutions via Monte Carlo simulation. First, using the given historical data in Figure 1 as the nominal value ($\mathbf{q}$) of the uncertainty set, we generate 1,000 probability vectors, as we did for the TV images example. Then we calculate the expected response and variance for the nominal and robust solutions at each probability vector, and take the averages. Table 10 shows higher means and lower variances than Table 9 in the worst-case. Table 9 is based on the

Table 10 Average Comparison for CDDC Example (Simulation Results)***

T	Robust			Nominal		
	$\bar{\sigma}_{\mathbf{e}}(\hat{y})$	$\bar{\mathbb{E}}_{\mathbf{e}}[\hat{y}]$	$\bar{\mathbb{E}}_{\mathbf{e}}[\hat{y}] \pm 3\bar{\sigma}_{\mathbf{e}}$	$\bar{\sigma}_{\mathbf{e}}(\hat{y})$	$\bar{\mathbb{E}}_{\mathbf{e}}[\hat{y}]$	$\bar{\mathbb{E}}_{\mathbf{e}}[\hat{y}] \pm 3\bar{\sigma}_{\mathbf{e}}$
$\leq$339	58	361	(186, 536)	58	360	(185, 535)
340	58	361	(186, 536)	58	360	(185, 535)
370	61	392	(209, 575)	59	369	(192, 547)
400	64	423	(229, 618)	61	399	(214, 585)
420	68	445	(238, 652)	64	419	(227, 612)
430	71	456	(241, 670)	65	429	(232, 627)
440	74	467	(244, 690)	67	439	(236, 643)
450	77	478	(245, 711)	69	449	(239, 659)

$***$ all entries should be multiplied by 1,000

worst-case scenario; i.e., we maximize the variance and minimize the expectation; Table 10 is based on the average performance. Table 10 shows that the confidence intervals for the robust solutions are always shifted to the right compared with the nominal solutions, so we prefer the robust approach, however, the robust solutions have slightly higher variations (from 0% to 10%) than the nominal solutions.

Adjustable Robust Optimization The number of shipping doors d_2 is an adjustable decision in the CDDC. A moderate number of shipping doors may increase the inventory in the temporary storage area, whereas an excessive number causes a low utilization of doors. We now assume that d_2 is adjustable according to the uncertain parameter e_1—namely, the variability in less-than-truckload supply shipments. More precisely, we assume that the number of shipping doors can be adjusted after the variability in supply shipments has been observed. We use the cell-based decision rule introduced in §3; the data and the 25 cells have already been presented in Figure 1. Notice that the domain of e_1 is divided into five equal intervals; for each interval we have a different decision.

In Table 11, columns two through four present the actual 'worst-case' performance of the ARC, and columns five through eight present the average performance of adjustable robust solutions via Monte Carlo simulation as in Table 10. We use the same probability vector sample as in the television images example. These numerical results show that

Table 11 ARO Results for CDDC Example*

	Worst-Case			Average		
T	$\sigma_{\mathbf{e}}(\hat{y})$	$\mathbb{E}_{\mathbf{e}}[\hat{y}]$	$\mathbb{E}_{\mathbf{e}}[\hat{y}] \pm 3\sigma_{\mathbf{e}}$	$\bar{\sigma}_{\mathbf{e}}(\hat{y})$	$\bar{\mathbb{E}}_{\mathbf{e}}[\hat{y}]$	$\bar{\mathbb{E}}_{\mathbf{e}}[\hat{y}] \pm 3\bar{\sigma}_{\mathbf{e}}$
$\leq$339	67 (0%)	339	(136, 541)	58	367	(192, 541)
340	67 (0%)	340	(137, 542)	58	367	(193, 542)
370	68 (2%)	370	(167, 572)	59	389	(194, 543)
400	71 (3%)	400	(188, 611)	61	418	(212, 566)
420	73 (6%)	420	(201, 639)	63	425	(235, 615)
430	74 (7%)	430	(209, 650)	64	448	(246, 630)
440	75 (10%)	440	(215, 665)	65	449	(254, 644)
450	78 (10%)	450	(216, 684)	69	456	(249, 661)

$***$ all entries should be multiplied by 1,000

the worst-case confidence intervals of the ARC are tighter than those of the general RC in Table 9. This is because of the improved response variances of the adjustable robust solutions, the improvement percentages (%) are reported in column two within brackets; e.g., it is as high as 10% when T =450000. As we anticipated, the simulation results show that the average performance of the adjustable robust solutions is better than the worst-case; i.e., the average of the response mean $\bar{\mathbb{E}}_{\mathbf{e}}[\hat{y}]$ is higher than the worst-case mean $\mathbb{E}_{\mathbf{e}}[\hat{y}]$, and the average variance $\bar{\sigma}^2_{\mathbf{e}}(\hat{y})$ is lower than the worst-case variance $\sigma^2_{\mathbf{e}}(\hat{y})$. Comparing the average performances of the non-adjustable and adjustable robust solutions in Tables 10 and 11 shows that the ARC yields tighter confidence intervals that are subintervals of the confidence intervals in the general RC. Therefore, the ARC reduces the response uncertainty compared with the general RC. Finally, additional experimentation showed that making 'only' d_3, d_4 or d_5 adjustable on e_1 has an improvement less than 1% in the objective; i.e., in those cases the non-adjustable solutions are as good as the adjustable solutions. This shows that the quantitative value of information may significantly change for different parameters in ARO.

5. Conclusions

In this paper, we propose a RO method for the dual response optimization problem. Our methodology uses an uncertainty set based on historical data on the environmental vari-

ables; this data may be collected from either real or simulated systems. Adding RO to the dual response approach has the following advantages: (i) we do not make any distributional assumptions on the uncertain parameters. (ii) We do account for the ambiguity caused by the lack of knowledge about distributions, using the so-called phi-divergence uncertainty sets based on historical data. (iii) Both our RO and ARO methods are computationally tractable; ARO is tractable even for nonlinear decision rules. (iv) Our ARO approach can be used for modeling adjustable integer decision variables; to the best of our knowledge, this paper is one of the two publications that models such variables (see, Bertsimas and Caramanis (2010) and Vayanos et al. (2011)).

References

Al-Aomar, R. 2006. Incorporating robustness into genetic algorithm search of stochastic simulation outputs. *Simulation Model. Practice and Theory* **14**(3) 201–223.

Angün, E. 2011. A risk-averse approach to simulation optimization with multiple responses. *Simulation Model. Practice and Theory* **19**(3) 911–923.

Ben-Tal, A., A. Nemirovski. 1998. Robust convex optimization. *Math. Oper. Res.* **23**(4) 769–805.

Ben-Tal, A., A. Nemirovski. 1999. Robust solutions of uncertain linear programs. *Oper. Res. Lett.* **25**(1) 1–13.

Ben-Tal, A., D. den Hertog, J.-P. Vial. 2015. Deriving robust counterparts of nonlinear uncertain inequalities. *Math. Prog.* **149**(1–2) 265–299.

Ben-Tal, A., L. El Ghaoui, A. Nemirovski. 2009. *Robust Optimization*. Princeton Press, Princeton, NJ.

Ben-Tal, A., D. den Hertog, A. De Waegenaere, B. Melenberg, G. Rennen. 2013. Robust solutions of optimization problems affected by uncertain probabilities. *Management Sci.* **59**(2) 341–357.

Bertsimas, D., C. Caramanis. 2010. Finite adaptability in multistage linear optimization *IEEE Trans. Automatic Cont.* **55**(12) 2751–2766.

Bertsimas, D., O. Nohadani, K. M. Theo. 2010. Robust optimization for unconstrained simulation-based problems. *Oper. Res.* **58**(1) 161–178.

Bingham, D., V. N. Nair. 2012. Noise variable settings in robust design experiments. *Technometrics* **54**(4) 388–397.

Chan, Y.-J., D. J. Ewins. 2010. Management of the variability of vibration response levels in mistuned bladed discs using robust design concepts. *Mech. Systems and Signal Processing Part 1: Parameter Design* **24**(8) 2777–2791.

Chen, L. H. (1997). Designing robust products with multiple quality characteristics. *Computers & Oper. Res.* **24**(10) 937–944.

del Castillo, E. 2007. *Process Optimization: A Statistical Approach*. Springer.

del Castillo, E., S.-K. Fan, J. Semple. 1999. Optimization of dual response systems: a comprehensive procedure for degenerate and nondegenerate problems. *European J. Oper. Res.* **112**(1) 174–186

del Castillo, E., S.-K. Fan, J. Semple. 1999. The computation of global optima in dual response systems. *J. Qual. Tech.* **29**(3) 347–353.

Dellino, G., J. P. C. Kleijnen, C. Meloni. 2012. Robust optimization in simulation: Taguchi and Krige combined *INFORMS J. Comput.* **24**(3) 471–484.

Delpiano, J., M. Sepulveda. 2006. Combining iterative heuristic optimization and uncertainty analysis methods for robust parameter design. *Engineering Optim.* **38**(7) 821–831.

El-Ghaoui, L., H. Lebret. 1997. Robust solutions to least-squares problems with uncertain data. *SIAM. J. Matrix Anal. & Appl.* **18**(4) 1035–1064.

Fu, M. C. 2007. Are we there yet? The marriage between simulation & optimization. *OR/MS Today* **34**(3) 16–17.

Fu, M. C. (Ed.) (2014). *Handbook of Simulation Optimization*. International Series in Operations Research & Management Science (Vol. 216), Springer, Chicago, IL.

Fu, M., B. Nelson. 2003. Simulation optimization. *ACM Trans. Model. and Simulation* **13**(2) 105–107.

Hu, Zhaolin, J. Cao, L. J. Hong. 2012. Robust simulation of global warming policies using the DICE model. *Management Sci.* **58**(13) 2190–2206.

Iman, R. L., J. C. Helton. 2006. An investigation of uncertainty and sensitivity analysis techniques for computer models. *Risk Anal.* **8**(1) 71–90.

Khuri, A. I., S. Mukhopadhyay. 2010. Response surface methodology. *Wiley Interdisciplinary Reviews: Computational Statistics* **2**(2) 128–149.

Kleijnen, J. P. C. 2008. *Design and Analysis of Simulation Experiments*. Springer-Verlag, Heidelberg, Germany.

Miranda, A. K., E. del Castillo. 2011. Robust parameter design optimization of simulation experiments using stochastic perturbation methods. *J. Oper. Res. Soc.* **62**(1) 198–205.

Myers, R. H, W. H. Carter. 1973. Response surface techniques for dual response systems. *Technometrics* **15**(2) 301–317.

Myers, R. H., D. C. Montgomery, C. M. Anderson-Cook. 2009. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 3rd ed. John Wiley & Sons, Hoboken, NJ.

Nair, V. N., B. Abraham, J. MacKay, G. Box, R. N. Kacker, T. J. Lorenzen, J. M. Lucas. 1992. Taguchi's parameter design: a panel discussion. *Technometrics* **34**(2) 127–161.

Norman, B. A., Smith, A. E. 2006. A continuous approach to considering uncertainty in facility design. *Computers & Oper. Res.* **33**(6) 1760–1775.

Pardo, L. 2006. *Statistical Inference Based on Divergence Measures*. Chapman & Hall/CRC, Boca Raton, FL.

Rockafellar, R. T. 1970. *Convex Analysis*. Princeton University Press, Princeton, NJ.

Samuelson, D. A. 2010. When close is better than optimal. *ORMS-Today* **37**(6).

Sanchez, S. M. 2000. Robust design: seeking the best of all possible worlds. *Proceedings of the 2000 Winter Simulation Conference* (edited by J.A. Joines, R.R. Barton, K Kang, and P.A. Fishwick) **1** 69–76.

Shi, W., Shang, J., Liu, Z., Zuo, X. (2014). Design of pre-enhanced cross-docking distribution center under supply uncertainty: RSM robust optimization method. *European J. Oper. Res.* **236**(2) 664–676.

Spall, J. 2003. *Introduction to Stochastic Search and Optimization*. Wiley-Interscience, New York, NY.

Stinstra, E., D. den Hertog. 2008. Robust optimization using computer experiments. *European J. Oper. Res.* **191**(3) 816–837.

Taguchi, G. 1986. *Introduction to Quality Engineering*. Asian Productivity Organization, UNIPUB, White Plains, NY.

Taguchi, G. 1987. *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost*. UNIPUB/Kraus International, White Plains, NY.

Urval, R., S. Lee, S. V. Atre, S.-J. Park, R. M. German. 2010. Optimisation of process conditions in powder injection moulding of microsystem components using robust design method. *Powder Metallurgy Part 2: Secondary Design Parameters* **53**(1) 71–81.

Vayanos, P., D. Kuhn, B. Rustem. 2011. Decision rules for information discovery in mutli-stage stochastic programming. *Proc. 50th IEEE Conf. on Decision and Control and European Control Conf. (CDC-ECC)*, Orlando, FL, 7368–7373.

Wiedemann, M. 2010. Robust Parameter Design for Agent-Based Simulation Models With Application in a Cultural Geography Model. Master's Thesis, Naval Postgraduate School, Monterey, CA.

Yanıkoğlu, İ., D. den Hertog. 2013. Safe approximations of ambiguous chance constraints using historical data. *INFORMS J. Comput.* **25**(4) 666–681.