Empirical Optimization of Imaging Processes by Use of Designed Experiments and Quality Loss Functions

Allan Ames*
Reading, Massachusetts

Neil Mattucci
North Weymouth, Massashusetts

Douglas Hawkins
Department of Applied Statistics, University of Minnesota, St. Paul, Minnesota

Imaging systems are multivariate, involving response functions or even sets of response functions—the red, green, and blue characteristic curves of photographic systems, for example—that relate image properties to scene properties. These functions must be simultaneously optimized to produce the best possible system. While the preferred methods for empirically optimizing the characteristics of a product or process are those of designed experimentation and response surface methodology, there is no widely accepted method that enables the application of these techniques to multivariate problems and therefore to imaging products and processes. This situation is changed with the advent of the desktop computer. We will describe a conceptually simple, though computationally intensive, method that enables application of designed experimentation and response surface methodology to multivariate systems and imaging systems. The method discussed will produce a more robust manufacturing process as well as a better product.

Journal of Imaging Science and Technology 43: 166–169 (1999)

Introduction

No product is immune to competition. Product quality must be continuously improved, even as manufacturing costs are reduced. For many manufactured products the manufacturing process is too complex, or too little understood, to have available analytical models to guide process improvement, and one must resort to empirical optimization. Almost all manufactured products must meet multiple "fit for use" criteria to sustain product quality. Thus it may come as a surprise that there is no generally accepted method for empirically optimizing a product or process that must simultaneously satisfy multiple criteria.[1,2] While several methods have been proposed, none have become widely accepted.[2,3,4,5,6,7,8]

Imaging systems are always multivariate, in that there is always at least one function, depending on the imaging device, that relates an output image to an input scene. We will call this function the image conversion function (ICF). Sturge and co-workers[9] list many imaging processes where the ICF can be identified: (1) in photography—the sensitometric curves of red, green, and blue image density as functions of exposure, (2) in electrophotograpy—the surface potential of the electrophotographic drum as a function of exposure, (3) in photolithography—the extent of mask polymerization as a function of exposure, (4) in microfabrication—the resist thickness as a function of exposure, (5) for display devices—the output brightness or intensity as a function of input electrical energy, and (6) in stereolithography—the amount of material deposited as a function of laser exposure.

The optimization problems we address in this article are those where, first, the process is reasonably well defined with a dozen or fewer continuously variable inputs. Second, there is a number, say six or more, of output variables or samples of the image conversion function at different input levels, that must be controlled to sustain product quality.

The problem that led to the methods described in this paper arose in the manufacture of Polaroid Spectra® instant photographic film.[10,11,12] A Polaroid instant picture includes three elements that are primarily chemical in nature: the negative coatings, the reagent, and the positive sheet coatings. When the color negative is manufactured, its properties may be perturbed by variations (usually due to raw materials) so that, with no subsequent adjustments, the final pictures would show unacceptable run to run variations. The cure for this problem, implemented since the days of the first Polaroid instant photographic products, was to adjust the composition of the reagent—the chemicals in the pod that are activated as the picture leaves the camera—so as to compensate for the variation in the negative. Initially,

Original manuscript received June 29, 1998

* IS&T Lifetime Member

adjustment of the chemistry was accomplished using a combination of experimentation and experience. As products evolved in complexity and improved in overall quality, the adjustment process became more complex than most technicians could execute, and in 1984 the search for more automatic methods of performing the adjustment was begun. We describe here in general terms a successful strategy to make chemical adjustment of reagent to compensate for variation in negative. Other details of the optimization are described elsewhere.[13]

Response Surface Methodology and Designed Experiments

Because the concepts are needed later in this article, we review briefly response surfaces and designed experimentation. These techniques are used for manufacturing quality control, and there is extensive literature[14,15,16,17,18] from which we cite but a few examples. Consider a chemical manufacturing step where the yield of a reaction is a function of the batch temperature and the concentration of two ingredients. Imagine we could write an equation that described the yield of the chemical reaction as a function of the batch temperature, T, and the two concentrations, C1 and C2. The yield function, with dependent units of moles, would vary with temperature, C1, and C2. Such a function is a surface embedded in the space (moles, T, C1, C2). The description of the response of a manufacturing process is called a response surface.

To optimize a process one needs a quantitative description of the various response surfaces that characterize the process in terms of inputs to the process. In many cases we do not know the exact relations, so we must approximate the response surfaces using experimental data. To this end, we assume some sort of polynomial model, run appropriate experiments, then least squares fit the data to the model. This empirical polynomial model is an approximation to the “true” response surface.

Methods devised to determine experimentally the coefficients in an assumed polynomial model with the least number of experiments are called “designed experimentation”, because the experiments are designed to capture only the needed information. Methods where response surfaces are estimated using designed experiments are called “response surface methodology”, or RSM.[2,8,14,15,19]

The Univariate Loss Function

There was once a time in American manufacturing when it was believed that, for the most important characteristics of a manufactured component, there was a band of acceptable variation. Within this tolerance band the product was “good” and outside the band the product was “bad”. A part either fit or it did not fit. The notion of failure probability was not generally understood, and it was not generally accepted that increasing departure from specification of a component of a product was associated with increased probability of failure of the product. To help change the good versus bad mindset, Taguchi[20,21] introduced the loss function as a way to model the concept that any departure from specification, however small, creates economic loss. The Taguchi loss function qualitatively describes the “economic loss to society” arising from errors in meeting a specification, including random variation from inside and outside the process as well as systematic errors in the process.

Multivariate Quality Loss Function

Assume that the economic loss associated with departures from specifications or target value of each characteristic of a product can be estimated. At least in principle, for a variety of independent uses of a product the losses can be weighted and summed to find the total loss to society for all variation of the product characteristics. Even if we cannot state exactly what the dollar loss associated with a deviation from target value is, we can make estimates of the importance of each specification for maintenance of quality. With these estimates we can combine the separate responses into a single loss function that can be used for process optimization.

For imaging processes, the loss function usually must be constructed iteratively. We start by sampling the image conversion function at several points, and for each sampled point construct a parabolic loss function. We then add the various loss functions together, starting with weight factors of unity. As experience is gained in the relative importance for overall quality of the various sampled points on the ICF, we might change the weights, or we might change the location of the sampled points themselves.

Quality Loss Functions

The Global Image Quality Loss Function. We imagine first the global image quality loss function GIQL, that describes the loss of product quality and product value as the various sampled points of the ICF depart from their specified or target values. The function we write down uses a parabolic approximation for the loss of quality due to deviation of any sampled point. We recognize that each sampled point on the ICF, P_r, will vary from the fixed target, G_r, because of random variation, systematic errors, and design compromises. We write the global image quality loss function, GIQL, as:

$$GIQL = \sum_{r=1,R} w_r \{P_r - G_r\}^2 \tag{1}$$

Here, P_r is a point on the ICF indexed on r, G_r is the target or value of P_r at optimal quality, and W_r is a weight factor. The sum is taken over all samples R in number.

Process Related Image Quality Loss Functions. We next shift consideration from outputs to inputs. To find the quality of the final product as a function of the process inputs, we need to express each of the ICF sample points in terms of the process input variables. To actually construct these functions, we use RSM or any other modeling method that can yield quantitative models for the response surfaces.

If the process inputs are $(X_1, X_2, - -)$ and the model for the ICF samples is expressed as a function of the inputs as $P_r = Y_r\,(X_1, X_2, - -)$ then we can write Eq. 2, the quality loss function referenced to process:

$$QLP = \sum_{r=1,R} w_r \{Y_r(X_1, X_2, --) + e_r - G_r\}^2 \tag{2}$$

In Eq. 2, $Y_r(X_1, X_2, - -)$ is the response surface describing each of the ICF samples P_r and $e_r = P_r - Y_r$ is any error, systematic or random, associated with the description in terms of inputs rather than in terms of responses. R is again the total number of samples, and G_r is the target.

If the errors, e_r, are small compared to the errors in hitting the targets, $(Y_r - P_r)$, we can neglect e_r and rewrite QLP as the approximate quality loss function referenced to process:

$$ALP = \sum_{r=1,R} w_r \{Y_r(X_1, X_2, --) - G_r\}^2 \tag{3}$$

Of the three loss functions above, we need only the last for process optimization. The first, Eq. 1, describes the quality loss of the imaging process in terms of the product or process requirements. The second loss function, Eq. 2, relates the quality loss of the imaging process to the process inputs by using response surfaces. Equation 3 includes only response surface polynomials and targets, and does not explicitly include random errors.

Process Optimization with ALP. As long as the random effects in the process are independent of the process inputs, or alternatively so long as the dominant loss of quality arises from systematic errors in the ICF, then the minima of ALP occur at particular values of the inputs $(X_1, X_2, \cdots)$. These minima define the process operating conditions that result in the least amount of quality loss for the process or product, and are therefore the best operating conditions for the product or process.

Robustness Characteristics. If a process is being operated under the conditions where overall quality loss is at a minimum, a second advantage accrues in addition to highest product quality. At a minimum the rate of change of the function with respect to change in the inputs is 0. This raises a point of some significance: the set of process inputs that produces the highest overall quality also produces the greatest stability or robustness against variations in quality due to input variation. Lucas[22] discusses this subject in greater detail.

Notes on Construction and Use of Loss Functions

Choice of Weight Factors. One of the difficulties of the method is the lack of an obvious objective basis for choosing the weight factors, the W_r. One can seldom estimate the "loss to society" of a departure from specification. One strategy is to find factors that will eliminate the units of measure of the squared error, that is, scale so that the "usual error" will have a value of unity, then subsequently assess the relative importance of each of these nondimensional squared error. Another strategy might be to scale the squared error by its expected value based on independent criteria, perhaps experimental error, then choose a multiplier of the scaled value that expresses the contribution to quality loss. Other strategies could compare deviations from target to control limit ranges or other process control variables.

For many imaging systems it is feasible to do testing of computer generated images with panels of viewers to find the comparative importance of various imaging errors. Whatever strategy is employed, keep in mind that the weight factors are the means to the end of finding the best process settings, and insofar as the targets can be hit, the weight factors are irrelevant because the contribution to quality loss of a parameter that meets specification is zero.

Less Than or Greater Than Criteria. For constructing loss functions to represent a response that must be less than or greater than a particular value, we have successfully used the half parabola method of Tribus and Szonyi.[23] This method maintains continuous first derivatives, that can be useful for finding the optima numerically. Broad regions of acceptability that are not well described with a single parabola can be described with split parabolas.

Need for Higher Order Designs. As optimization of a product under development proceeds and as more outputs and more inputs are considered, the likelihood of finding important nonlinear interactions within designed experiments increases. Similarly, as a process under development becomes more capable and errors get smaller, the probability increases that interactions between the inputs will be observed as significant. Assessment of nonlinearities requires higher order designs, that is, experimental designs that can provide estimates of quadratic (or higher) coefficients within the response surface polynomial. The important point is that for multivariable optimizations, the user should anticipate non-linear effects and interactions which must be supported by the basic experimental designs. For fitting second-degree polynomials we have found the central composite designs[15] to be particularly useful.

Some Differences Between Optimizations with RSM and with Loss Functions. It is conventional in RSM to subject each coefficient in the polynomial describing the surface to a test for significance, and thus to justify the use of the coefficient in the final regression. For use within quality loss functions, our experience is that best results are obtained when there is no selection of the coefficients. While there is no firm theoretical justification, the most successful strategy has been to use the largest number of coefficients that can be supported by whatever experimental design has been chosen.

Minimizing ALP on the Computer. Finding the minima of Eq. 3 with respect to the process inputs requires the use of a computer in most practical cases. Possible methods range from a simple "steepest ascent" method to elaborate forms of Newton's method.[24,25,26] For smaller problems, one can use the solvers within spread sheet programs. All work discussed in the article was performed on minicomputers or workstations programmed in Fortran. The earliest minimizations were done with steepest ascent methods because they could be more reliably coded. As confidence in the technique grew, work shifted to Newton's method that proved much faster in most cases, though not all. Either minimization method, or even others, may prove useful for particular problems.

Fabricated Example

To show how the method works, we discuss a realistic, but largely fabricated example abstracting data from an old text. In Fig. 1 we show an example of a characteristic curve of a photographic negative—the ICF for some particular product. We identify four points on the curve by sampling at exposures of 5, 10, 15, and 20 steps.

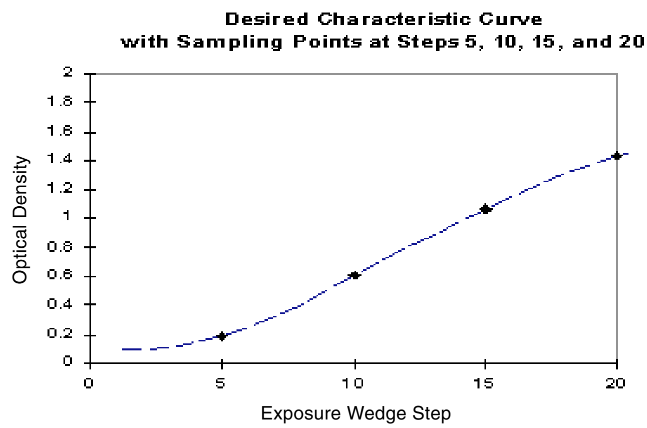

Figure 1.

TABLE I. Two Factor Central Composite Experimental Design for Black and White Development

	Experimental Inputs		Scaled Inputs	
Exp. #	Dev. Time (minutes)	Temperature (Deg. C.)	X_1	X_2
1	4	18	-1	-1
2	9	18	1	-1
3	4	22	-1	1
4	9	22	1	1
5	2.965	20	-1.414	0
6	10.035	20	1.414	0
7	6.5	17.172	0	-1.414
8	6.5	22.828	0	1.414
9	6.5	20	0	0
10	6.5	20	0	0

TABLE II. Values for the Polynomial Coefficients

Polynomial Term		ICF AT STEP 5	ICF AT STEP 10	ICF AT STEP 15	ICF AT STEP 20
Zero level	(B_0)	.23000	.82400	1.35400	1.73800
Development time	(B_1)	.06201	.22580	.29949	.31089
Development temp.	(B_2)	.00352	.01259	.01679	.01729
Time squared	(B_{11})	.01213	.00269	–.01438	–.01969
Temp. squared	(B_{22})	.00112	.00219	.00138	.00131
Time * temp.	(B_{12})	.00150	–.00025	–.00175	–.00225

The chemical development of a photographic negative depends on, among other things, the temperature and the development time. We will construct an example with these two variables.

A central composite experimental design in two factors[15] (CCD2) permits representing a process response Y_r in terms of the scaled input variables X_1 and X_2 with the following general polynomial:

$$Y_r\,(X_1, X_2\,) = B_0 + B_1 X_1 + B_2\,X_2 + B_{11}\,X_1{}^2 + B_{22}\,X_2{}^2 + B_{12}\,X_1\,X_2 \qquad (4)$$

To estimate the coefficients in Eq. 4 for the variables X_1 (development time) and X_2 (development temperature) we need to perform experiments. The experiments in which the variables are varied consistent with the CCD2 sampling requirements are listed in Table I.

For each of the four exposures where the ICF is sampled, there is a corresponding polynomial. Representative values for the coefficients in Eq. 4 are listed in Table II.

The initial trial value for the approximate quality loss function ALP is given in Eq. 5.

$$ALP = (Y_{\mathrm{STEP\_5}}\,(X_1, X_2\,) - .18)^2 + (Y_{\mathrm{STEP\_10}}\,(X_1, X_2) - .62)^2 + (Y_{\mathrm{STEP\_15}}(X_1, X_2\,) - 1.07)^2 + (Y_{\mathrm{STEP\_20}}(X_1, X_2\,) - 1.44)^2 \qquad (5)$$

For the various numbers given in Tables I and II, the ALP is minimized, and target values (0.184, 0.621, 1.070, 1.440) achieved for X_1 = -0.91 (scaled) corresponding to 4.23 minutes and X_2 = -0.02 (scaled) corresponding to 20°C.

Summary

We have presented a novel and powerful method for dealing with a problem frequently encountered in the manufacture and development of imaging products, that of process tuning to achieve highest product quality. The basic strategy is to treat sampled points on the image conversion function as response surfaces, which can be described with experimentally derived polynomials. The individual response surfaces are then combined into a single loss function using known specifications or desired targets. Minimizing the loss function with respect to process inputs locates operating conditions which produce the product of highest quality and most robustness against process input variation.

Acknowledgments. The authors are indebted to S. MacDonald, S. Baistos and L. Clemente for their work on networks and process automation, to Geza Szonyi for his help in designed experimentation, to Polaroid Corporation for permission to publish this work, and to R. Eckert, E. Lindholm and other supporters of this project over the years.

References

1. R. H. Myers and D. C. Montgomery, *Response Surface Methodology; Process and Product Optimization Using Designed Experiments*, John Wiley and Sons, Inc., New York, NY, 1995.
2. A. I. Khuri and J. A. Cornell, *Response Surfaces,* 2nd ed., Marcel Dekker, Inc., New York, NY, 1996.
3. R. H. Myers and W. H. Carter, Jr., Response Surface Techniques for Dual Response Systems, *Technometrics* **15,** 301–317 (1973)
4. G. G. Vining and R. H. Myers, Combining Taguchi and Response Surface Philosophies: A Dual Response Approach, *J. Quality Technol.* **22,** 38–45 (1990).
5. E. Del Castillo and D. C. Montgomery, A Nonlinear Programming Solution to the Dual Response Problem, *J. Quality Technol.* **25,** 199–204 (1993).
6. G. C. Derringer, A Balancing Act: Optimizing a Product's Properties, *Quality Progress* **27,** 51–58 (1994).
7. G. Derringer and R. Suich, Simultaneous Optimization of Several Response Variables, *J. Quality Technol.* **12,** 214–219 (1980).
8. R. H. Myers, A. I. Khuri and W. H. Carter, Jr., Response Surface Methodology, 1966-1988"., *Technometrics* **31,** 137–157 (1989).
9. J. M. Sturge, V. Walworth and A. Shepp, Eds., *Imaging Processes and Materials, Neblette's Eighth Edition*, Van Nostrand Reinhold, New York, NY, 1989.
10. F. Meneghini, A New Instant Photography: A Chemists's View, *J. Imaging Tech.* **15,** 114–119 (1989).
11. R. J. Lambert, A Technical Description of Polaroid Spectra Film, *J. Imaging Tech.* **15,** 108–113 (1989)
12. E. S. McCaskill and S. R. Herchen, Electrochemical Properties and Mechanisms of Action of Developers in Spectra Film, *J. Imaging Tech.* **15,** 103–107 (1989).
13. A. E. Ames, N. Mattucci, S. MacDonald, G Szonyi, and D. Hawkins, Quality Loss Functions for Optimization Across Multiple Response Surfaces, *J. Quality Technol.* **29,** 339–346 (1997)
14. G. E. P. Box and N. R. Draper, *Empirical Model Building and Response Surfaces*, John Wiley and Sons, Inc., New York, 1987.
15. W. Cochran and G. M. Cox, *Experimental Designs* , 2nd ed., John Wiley and Sons, New York, NY, 1957.
16. G. E. P. Box and D. W. Behnken, Some New Three Level Designs for the Study of Quantitative Variables, *Technometrics* **2,** 455–475 (1960).
17. M. S. Phadke, Quality Engineering Using Design of Experiments, *Proc.of the Section on Statistical Education*, *American Statistical Association,* 11–22 (1982).
18. G. J. Hahn, Experimental Design in a Complex World, *Technometrics* **26,** 19–31 (1984).
19. D. C. Montgomery, *Design and Analysis of Experiment,* 4th ed. John Wiley and Sons, Inc., New York, NY, 1997.
20. G. Taguchi and Y. Wu, *Off-line Quality Control,* Central Japan Quality Control Association, 1979.(Available from American Supplier Institute, Dearborn, MI).
21. P. J. Ross, *Taguchi Techniques for Quality Engineering*, McGraw-Hill Book Company, New York, 1988.
22. J. M. Lucas, How to Achieve a Robust Process Using Response Surface Methodology, *J. Quality Technol.* **26,** 248–260 (1994).
23. M. Tribus and G. Szonyi, An Alternative View of the Taguchi Approach, *Quality Progress* **22,** 46–52 (1989).
24. A. V. Fiacco and G. P. McCormack, *Nonlinear Programming, Sequential Unconstrained Minimization Techniques*, SIAM, Philadelphia, PA, 1990.
25. *Nonlinear Optimization, Theory and Algorithms*, L. C. W. Dixon, E. Spedicato and G. B. Szego, Eds., Birkhauser, Boston, MA, 1980.
26. J.-P. Aubin *Explicit Methods of Optimization.* Gauthier-Villars (c)BORDAS, Paris, 1984.
27. H. N. Todd and R. D. Zakia *Photographic Sensitometry, The Study of Tone Reproduction.* Morgan and Morgan, Inc., Hastings-on-Hudson, NY, 1969.