1 Introduction

1.1 What is "modern" about modern optical tolerancing?

In a 1985 SPIE conference paper titled, "Fundamentals of Establishing an Optical Tolerance Budget," Warren J. Smith described the application of the "root sum square" (RSS) method for assigning tolerances to optical imaging systems.[1] For many decades, optical designers and engineers have successfully applied the RSS technique—and variants of it, such as the inclusion of higher-order terms—to optical tolerancing analysis.[2–7] Indeed, in the concluding paragraph of his paper, Smith wrote, "This technique has been used by the author for many years and has been applied to a wide variety of optical systems. It has been applied to fabrication quantities ranging from thousands down to lots of one or two. It has yet to fail."[1]

Challenges in modern optical product development have driven optical designers and engineers to deepen their understanding of statistics and continue developing tolerancing techniques.[8–17] Today, optical tolerancing can involve methods to "desensitize" an optical design,[18–25] consider cost,[26–30] and run Monte Carlo simulations.[31–44] With the availability of powerful modern computers, performing a Monte Carlo simulation today can be as easy as the push of a button (after an immensely tedious task of setting up physically meaningful models of possible optical errors[13]). Such simulations produce as many virtual optical systems as the computer's memory can handle.

Monte Carlo simulations allow for the study of variability in optical systems that are integrated into highly complex instruments (e.g., instruments that may comprise optics, mechanics, electronics, firmware, software, algorithms, chemistry, biological samples, and more). If the probability distributions of optical component variables are either known or assumed, one can simulate a variety of conditions for the workflow of producing, assembling, and testing a whole instrument. Such a simulation can account for virtually any interactions between and among the subsystem and system parts, and the outputs of a simulation can be analyzed statistically. Now, imagine variability in optical parameters [such as focal length and root-mean-square (RMS) spot size] combining with, say, variability in the fluorescence emitted by dyes attached to DNA molecules. This is not as crazy as it sounds. I have done it in my work—by performing Monte Carlo simulations.

1.2 What is a Monte Carlo simulation?

Suppose we were designing a commercial off-the-shelf single-element lens (i.e., a "singlet"), with no regard to aberration control (i.e., we would not be bending the lens to minimize spherical aberration). In this case, the main parameter of interest (neglecting its diameter and other ISO 10110 specifications) is this lens's effective focal length (EFL), given (in air) by[45]

$$\frac{1}{f} = (n-1)\left[\frac{1}{R_1} - \frac{1}{R_2} + \frac{t}{R_1 R_2}\left(1 - \frac{1}{n}\right)\right], \tag{1}$$

where R_1 and R_2 are the radii of curvatures of the first and second surfaces, respectively (we have assumed a sign convention that the second surface's radius of curvature is <0), t is the center thickness of the lens, and n is the lens's refractive index at some desired wavelength. In the presence of manufacturing tolerances, Eq. (1) may be expressed as

$$\begin{aligned}\frac{1}{f \pm \Delta f} = (n \mp \Delta n - 1)\Bigg[&\frac{1}{R_1 \pm \Delta R_1} - \frac{1}{R_2 \mp \Delta R_2} \\ &+ \frac{t \mp \Delta t}{(R_1 \pm \Delta R_1)(R_2 \pm \Delta R_2)}\left(1 - \frac{1}{n \mp \Delta n}\right)\Bigg],\end{aligned} \tag{2}$$

where the "Δ" symbols denote maximum changes in the magnitude of each variable in the equation. In a tolerance analysis of this lens's EFL, we may wish to know the minimum (min) and maximum (max) values $f - \Delta f$ and $f + \Delta f$, respectively, as a consequence of all possible combinations of min and max changes for all of the variables appearing in Eq. (2).

Equation (2) is not overly complicated. By performing some algebraic analysis, one can determine the min and max values for the EFL. Alternatively, if a Monte Carlo simulation is performed, all of the variables in Eq. (2) may simultaneously be assigned random values, with equal likelihood of occurrence for all possible values within permissible min and max changes. Consequently, there will result a distribution of EFL values. If a sufficiently large number of random values for the variables is generated, then among the distribution of EFL values there should exist a value very close to $f - \Delta f$ and there should exist some other value very close to $f + \Delta f$. We then regard these two values as estimates of the min and max values for this lens's EFL.

In simple terms, a Monte Carlo simulation is the act of modeling the occurrence of random events. The name "Monte Carlo" is derived from the name of the famous casino in Monaco. In slightly more formal terms, a Monte Carlo simulation may be regarded as a statistical method to evaluate definite multiple integrals.[46,47] To understand this, consider that the average value of a function of the variables x, y, and z bounded by volume V is expressed as

$$F_{\text{ave}} = \frac{1}{V}\iiint_V F(x,y,z)\,dV. \tag{3}$$

Accordingly, the multiple integrals in Eq. (3) may be expressed as

$$\iiint_V F(x,y,z)\,dV = V F_{\text{ave}}. \tag{4}$$

We then note that an estimate of the multiple integrals may be determined by the product of the quantities on the right side of Eq. (4). If we perform a Monte Carlo simulation of the variables x, y, and z within the bounds described by volume V, this would yield a distribution of values for the function $F(x, y, z)$ whose mean would be an estimate of F_{ave}. Substituting this value and the known volume V into the right side of Eq. (4) yields an estimate of the multiple integrals.

In the context of optical system tolerancing, we apply Monte Carlo simulations to assign random values to all variables of interest in an optical system, such as radii of curvatures, thicknesses, element tilts, and so on, and we examine the frequency distributions of the resulting optical system's performance metrics, such as EFL, image distortion, and modulation transfer function (MTF). From these, we may estimate manufacturing yields of the produced systems based on those performance metrics. The area under the probability distribution that is bounded by min and max values of a performance metric would be an estimate of yield for that metric. In the singlet example mentioned above, we regarded that lens's EFL as a performance metric. If EFL is the only metric for that lens, then that lens's manufacturing yield would be the area under the distribution of EFL values, bounded by the min and max EFL values.

1.3 Significance of functions of random variables

At this point, it is a good idea to make the connection between terms used in this Spotlight with language that is frequently used in statistics. In particular, let us get accustomed to the terms displayed in Fig. 1. In this figure, the terms "parameter," "function," "performance metric," and "performance criterion" are taken to mean the same. Likewise, "variables" and "factors" are used synonymously. All "Δ" quantities (including variations of parameters) shall be referred to as "perturbations," "levels," "variation," "spread," "noise," or, in much of the context of optical tolerancing, they may be called "tolerances" for respective variables and parameters.

In statistics, one often assumes that a parameter may have some unknown relationship with its factors. To study and identify any possible relationship, statisticians would perform a design of experiments (DOE), where the values of

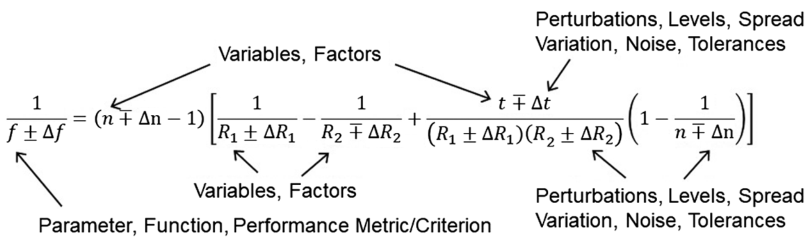

Figure 1 Terms applied in this Spotlight.

factors are either carefully assigned variations called "levels" (sometimes, levels are referred to as "treatments"), or they are randomly drawn from some distribution. When the levels are chosen, we have what is called a "fixed effects model," and when they are given a range of random values, we have what is called a "random effects model."[48] In a fixed effects model, levels are given min and max values and the resulting measured parameter values are fitted to a "least squares" curve. This action of fitting parameter data to a curve is known as regression.[48] Once a regression curve is determined, one often attempts to identify possible relationships by computing correlation coefficients (acknowledging, of course, that correlation is not necessarily causation).

In much of optical design, parameters and variables are often highly correlated. This is because most optical parameters can be expressed as deterministic functions of variables. For example, the EFL in Eq. (1) is a multivariable function. As such, it is a quantity that is deterministic in the sense that it is known how EFL will behave as its variables vary, but it can be stochastic in the sense that its variables may be subject to randomness. Under factory conditions, the EFL in Eq. (1) may be expressed by a perturbed form, as in Eq. (2). In this case, the EFL may be considered a "function of random variables." Thus for any optical parameter that is a multivariable function, one may regard that optical parameter as a function of random variables when performing optical tolerancing analyses. This has some significance to Monte Carlo simulation.

The beauty of performing Monte Carlo simulations is that one can assign arbitrary frequency distributions to variables and obtain the resulting distributions of the parameters. However, sometimes, one may wish to understand the underlying reason why an optical parameter's frequency distribution is what it is. In statistical theory, there are theorems that describe how to derive the probability distributions of functions of single- and multiple-random variables.[48–50] For example, since Eq. (2) is a function of multiple-random variables during a Monte Carlo simulation, then according to statistical theory, the EFL given by Eq. (1) has a specific probability distribution, and it need not be normal except under special conditions.[51] In Section 3, we shall explore some conditions where non-normal distributions result. (I note that I have not made the distinction between discrete and continuous random variables, nor have I made the distinction between discrete probability distributions and continuous probability density functions. In this Spotlight, we shall take the term "probability distribution" to mean either a discrete probability distribution or a continuous probability density function, or simply a "density function," depending on the context of the discussion.)

2 Prelude to Monte Carlo

2.1 Systematic versus random variability

In reality, not all of the variables in Eq. (1) should be perturbed randomly when performing a Monte Carlo simulation. In particular, the refractive index of a lens