Process Capability Indices—A Review, 1992–2000

SAMUEL KOTZ

George Washington University, Washington, DC 20052

NORMAN L. JOHNSON

University of North Carolina, Chapel Hill, NC 27599

We provide a compact survey and brief interpretations and comments on some 170 publications on process capability indices which appeared in widely scattered sources during the years 1992 to 2000. An assessment of the most widely used process capability indices is also presented.

Personal Preamble

THE rapidity of scientific development (including statistics and operations research) towards the end of the 20^{th} century has not only been a favorite topic of commentators, it has been truly amazing. Much work done ten years ago is now often classified as "obsolete." The very limited and specific field of process capability indices (PCIs) is, in this respect, quite typical, though some early ideas and methods appear to remain important and useful. Before embarking on our review, we thought it might be desirable to summarize our attitude regarding the use and reputation of PCIs.

The entire October 1992 issue of the *Journal of Quality Technology* (*JQT*) was devoted to the topic, which was then relatively new, of PCIs. We were among the contributors. At that time we were able to locate some 50 papers on the subject since its early life in the 70's and 80's. After the *JQT* issue was published, we received a communication from then editor Peter Nelson of *JQT* informing us that he did not envision any further papers on PCIs appearing in *JQT* in the foreseeable future. It was therefore with great surprise—as well as pleasure—that we received an invitation from the present editor to provide a survey, including developments in the PCI "world" since 1992, and some comments on possible future growth.

There have, indeed, been remarkable developments in the field in the last eight years. These include four books in English (Kotz and Johnson (1993a), Bothe (1997), Kotz and Lovelace (1998), and Wheeler (1999)), and a monumental, comprehensive book in German (Rinne and Mittag (1999)). Our bibliography includes some 170 papers appearing during 1993–2000 on the topic of PCIs, from a variety of sources ranging from theoretical-mathematical journals to down-to-earth quality control publications. A considerable variety of software has been generated, geared toward implementation of some of the more recent versions of PCIs. A conservative estimate indicates the existence of about twenty variants of univariate PCIs and seven multivariate PCIs. Although the majority of contributions cited in this paper are from the U.S.A., the following countries are represented by authors of papers listed in the bibliography: Sweden, Germany, Australia, U.K., China, Netherlands, Taiwan, Canada, Spain, India, Italy, Japan, South Korea, Saudi Arabia, Russia, Israel, Romania, Brazil, and Greece. In addition, we have had communication with researchers and practitioners from France, New Zealand, and Belgium.

Despite this superficially glowing picture, we could not ignore signs of uneasiness among both practitioners and theoreticians. In particular, it was clear to us that although our book (Kotz and Johnson (1993a)) received mostly polite reviews from all quarters, it

Dr. Kotz is currently Senior Research Scholar in the Department of Engineering Management. His e-mail address is kotz@seas.gwu.edu.

Dr. Johnson is Professor Emeritus in the Department of Statistics.

was felt to be far too theoretical for practitioners. For example, Walker (1995) was of the opinion that "a person with an M.S. in a technical area should understand the results, but to fully comprehend the derivations and proofs, a Ph.D. is desirable." In our opinion, this reflects an inadequate level of training in probability theory and statistical inference in engineering education. See Annis (2001) in this connection. Furthermore, we believe that the theory underlying PCIs is essentially elementary, and that anyone who understands the structure, working formula, and usage of the t-statistic [$t = \sqrt{n}(\bar{x} - \mu)/s$ in common notation] should have few difficulties in comprehending analyses relevant to all but the most "advanced" PCI indices.

Nevertheless, we agree with sentiments expressed by Orchard (2000) and believe that parts (at least) of them deserve to be more widely known, so we quote the following sample: "there is, as far as I know, no argument about the methods used to calculate 'errors and uncertainties' but I am aware that engineers struggle to come to terms with statistical issues that are relatively trivial, and that statisticians have to learn a new vocabulary in order to talk about bias, variation and confidence intervals."

The gap between theoreticians and practitioners is, we believe and hope, closing, mainly through software (which is perhaps the medicine that cures the symptoms rather than the disease), but there still remain numerous instances of (mutual) lack of understanding of the purpose and usage of PCIs and process performance, often due to a confusion about an appropriate estimator of process variance.

Also, it seems to us that although the topic of PCIs may be used by some academicians as an opportunity for proposing new indices, regardless of their practical relevance and often for the sake of the accompanying theory, most academicians are well-aware of the kinds of problems faced by practitioners and are sincerely trying to bridge the gap.

Unfortunately, mistrust of PCIs, and especially of their estimators, when based on the meager data often available at factory level, is still not uncommon. Resistance to accompanying a single estimated PCI value by an estimate of its variability (be it confidence interval, standard deviation, or whatever) is still very pronounced, even though accepted statistics, such as sample mean and standard deviation, perform essentially the same function.

We now proceed to the survey itself, starting with an introduction which provides the background information needed for a comfortable reading of the survey proper.

Introduction

As we understand it, PCIs are intended to provide single-number assessments of ability to meet specification limits on quality characteristic(s) of interest.

We are going to attempt to describe some recent and current developments in PCIs (and related measures). They are neither purely scientific nor purely practical in nature, though elements of both are certainly present. Broadly speaking, we might term them a "social movement." We are not attempting to ascribe credit (or blame) for these developments, though we may incidentally illustrate some shortcomings.

A major feature of these developments has been the proliferation and increased variety of circumstances in which PCIs have been applied. Indeed, even within a single organization PCIs may be used in relation to products and processes of many different kinds, each with its own specific requirements and problems. As a result, it is not really possible to give a coherent "world-view" of the state of "PCI art" or its likely future course(s). However, we shall try to identify some major concepts and methods, and boldly speculate on their immediate future.

As a consequence of the varied ways in which PCIs are used, there have been two natural lines of research work: (i) studies on the properties of PCIs and their estimators in many different environments; and (ii) construction of new PCIs purporting to have better properties in certain circumstances. (Under (ii) may be included suggestions for replacing PCIs by other procedures.) We will try to relate these to an assessment of prominent features of the construction and implementation of PCIs. This will necessarily include consideration of: (a) aims of procedures, insofar as these are sufficiently clear and widely accepted; (b) the interests of users - broadly, if inaccurately, termed "engineers" ("practitioners" is a more general term); (c) technical statistical aspects of PCIs, especially in relation to presumed aims of their use; and (d) estimation of PCIs, with attention to desirable amounts of data ("sample size") on which the procedure is to be based.

Our bibliography contains over 200 references. As we have noted, most of them (about 170) are from the period 1993-2000 (i.e. subsequent to publication of Kotz and Johnson (1993a)). The large number of recent publications may attest, in part, to the importance of the subject of PCIs, in *some* quarters (though not necessarily to actual importance in the eyes of engineers).

Notation and the "Basic" PCIs

For convenience, we will denote the upper and lower specification limits by U, L respectively, rather than the more customary USL, LSL. When (as in the bulk of this survey) univariate measurements are concerned, we will denote the corresponding variate by X. The expected value and standard deviation of X will be denoted by μ and σ respectively. When multivariate measurements are involved we will use $\mathbf{X}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ represents the variance-covariance matrix of $\mathbf{X}$. We will limit ourselves to situations when μ is in the specification interval, i.e. $L \leq \mu \leq U$.

There appears to be a general acceptance of the idea that PCIs can be used only after it has been established that a process is in "statistical control" (for example, by the use of control charts). This is reasonable, if it simply required that there be no irregular changes in quality level. However, there seems to be, in some quarters, an assumption that the measured characteristic should have a normal distribution (at least, approximately), although it is difficult to see why a good industrial process *must* result in a normal distribution for *every* measured characteristic.

The commonly recognized "basic" PCIs are:

$$C_p = \frac{U-L}{6\sigma} = \frac{d}{3\sigma}, \tag{1}$$

where $d = (U-L)/2$;

$$C_{pk} = \frac{d - |\mu - M|}{3\sigma} = \frac{\min\{U-\mu, \mu - L\}}{3\sigma}, \tag{2}$$

where $M = (U+L)/2$; and

$$C_{pm} = \frac{d}{3\sqrt{\sigma^2 + (\mu - T)^2}} = \frac{d}{3\sqrt{E[(X-T)^2]}}, \tag{3}$$

where T is a "target value" and $E[\cdot]$ denotes "expected value."

Usually, $T = M$; if $T \neq M$ the situation is sometimes described as "asymmetric tolerances." (See Boyles (1994) and Vännman (1997b, 1998a)). Introduction of C_p (initially as a "capability ratio") is ascribed to Juran (1974); that of C_{pk} to Kane (1986); and that of C_{pm} for the most part to Hsiang and Taguchi (1985). The measure C_{pm} is sometimes called the "Taguchi index."

There is also the hybrid index

$$C_{pmk} = \frac{d - |\mu - M|}{3\sqrt{E[(X-T)^2]}} \tag{4}$$

(Choi and Owen (1990); Pearn et al. (1992)). Clearly $C_p \geq C_{pk} \geq C_{pmk}$ and $C_p \geq C_{pm} \geq C_{pmk}$. The relation between C_{pk} and C_{pm} is less clearcut. From Equations (1) and (2) we have

$$C_{pk} = C_p - \frac{1}{3}\left|\frac{\mu - M}{\sigma}\right|, \tag{5}$$

and from Equations (1) and (3) we have

$$C_{pm} = \frac{C_p}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}}. \tag{6}$$

In the special case when $T = M$, only, it follows from (5) and (6) that

$$\begin{aligned} \frac{C_{pk}}{C_{pm}} &= \left(1 - \frac{1}{3C_p}\left|\frac{\mu - M}{\sigma}\right|\right)\sqrt{1 + \left(\frac{\mu - M}{\sigma}\right)^2} \\ &< \left(1 - \frac{1}{3C_p}\left|\frac{\mu - M}{\sigma}\right|\right)\left\{1 + \frac{1}{2}\left(\frac{\mu - M}{\sigma}\right)^2\right\}. \end{aligned}$$

So the relation $C_{pk} < C_{pm}$ is certainly valid if

$$\begin{aligned} &\frac{1}{3C_p}\left|\frac{\mu - M}{\sigma}\right| > \frac{1}{2}\left(\frac{\mu - M}{\sigma}\right)^2 \quad \text{or} \\ &\left|\frac{\mu - M}{\sigma}\right| < \frac{2}{3C_p} \quad \text{or, equivalently,} \\ &\left|\frac{\mu - M}{d}\right| < \frac{2}{9C_p^2}. \end{aligned}$$

Parlar and Wesolowsky (1998) have noted that if $T = M$, then the three basic PCIs are connected by the relationship

$$C_{pk} = C_p - \frac{1}{3}\sqrt{\left(\frac{C_p}{C_{pm}}\right)^2 - 1}. \tag{7}$$

See also Kotz and Johnson (1999), who examine the relations between C_p, C_{pk} and C_{pm} in detail. If X has a normal distribution, then the indices C_p and C_{pk}, together, determine the expected proportion (p) of values of X falling outside the specification interval (L, U)—called "nonconforming" (NC), as exhibited in Equation (11) below. On the other hand, C_{pm} is related to a Taguchi-style loss function. Indeed,

$$\frac{1}{C_{pm}^2} = \frac{9}{d^2}E\left[(X-T)^2\right] \tag{8}$$

has been suggested as a "process incapability index" (Greenwich and Jahr-Schaffrath (1995)). This possibility will be discussed later.

However, the C_p index is clearly related to the assumption that the distribution of X is sufficiently close to normal for "practical purposes". At any rate, this appears to be the motivation for the choice of the multiplier '6' in the denominator of the second term of Equation (1).

If X does have a normal distribution, then the probability that an item is NC, i.e. its value is outside the specification limits, is

$$p = \Pr[X \notin [L, U]] = 1 - \left\{\Phi\left(\frac{U-\mu}{\sigma}\right) - \Phi\left(\frac{L-\mu}{\sigma}\right)\right\}, \quad (9)$$

where $\Phi(\cdot)$ is the unit normal cdf. If $L = \mu - 3\sigma$ and $U = \mu + 3\sigma$, so that in this case $U - L = 6\sigma$ and $C_p = 1$, this probability is rather small (0.27 percent). But note that this requires that $\mu = (L + U)/2 = M$. A value of $C_p = 1$ does not guarantee that p is 0.27 percent—in fact it does guarantee that p cannot be less than 0.27 percent. However, note also that if $C_{pk} = 1$, then [from Equation (2)] the greatest possible value for p is 0.27 percent. For $C_{pk} = 1$ means that $\min(U - \mu, \mu - L) = 3\sigma$ so that $L \leq \mu - 3\sigma$ and $U \leq \mu + 3\sigma$. It is true that $C_p = 1$ does imply that, for $L \leq \mu \leq U$, p can exceed 50 percent only slightly, even when $\mu = L$ or $\mu = U$, but this is scant practical consolation.

Keeping these features in mind, it is to be emphasized that we regard "capability" in the present context, as meaning "possibility of achieving," rather than "actually achieving"—i.e. in the terminology of Kane (1986), "process potential." Adopting this outlook, Veevers (1995, 1998, 1999) has used the term "viability" (introduced by Davis et al. (1992)) representing "capability potential," and constructed a viability index. This is of general application, not restricted to normal distributions for X, or even to univariate situations (see later). The univariate viability is the ratio of the length (w) of the interval (the "window of opportunity") for θ for which the distribution of $(X + \theta)$ would result in an expected proportion NC (p) no greater than the conventional 0.27 percent, to the length ($2d$) of the specification interval (L, U). Thus

$$V_t = \frac{w}{2d}. \quad (10)$$

For the special case of normal $N(\mu, \sigma^2)$ distributions, we have the window of opportunity (for μ)

$$M - (d - 3\sigma) \leq \mu \leq (d - 3\sigma) + M,$$

so $w = 2(d - 3\sigma)$ and

$$V_t = \frac{2(d - 3\sigma)}{2d} = 1 - \frac{1}{C_p}.$$

If V_t is less than zero, there is no possibility of attaining a value of p of 0.27 percent or lower, and the process is to be regarded as "not viable."

The above discussion is not intended to shatter totally the myth that the value of $C_p = 1$ corresponds to a satisfactory low proportion of non-conforming items in the process. This assertion indeed contains a grain of truth, but ought to be taken with a substantial grain of salt. Indeed, in many cases a value of p as small as 0.0027 is regarded as woefully inadequate.

Parenthetically, we note that C_p is sometimes called the "six sigma" index. This is, presumably, because the increasingly popular Six Sigma method of achieving good results by careful organization of quality assurance procedures has adopted PCI indices—including C_p—as a tool in some of its methods. Our critical remarks on some PCIs, and particularly on the "six sigma" C_p, are not intended as a direct criticism of the entire Six Sigma program, though they do relate to some interpretations of one of its tools.

Considerations of this kind may have played a part in motivating increases in the allowable lower bound for C_p above 1. Values of 1.33, 1.66, and "even 2.00" (Bothe (1999)) are becoming more and more common. It is noteworthy that the choices of the limits 4/3, 5/3 appear to have been based on the fact that they imply $2d = 8\sigma$ and $2d = 10\sigma$ respectively (with associated values for p (if $\mu = M$) of 0.0063 and 0.000057 percent, respectively, on the assumption of normality being valid even in extreme tails of the distribution of X).

From quite early on (early 80's, in this context), the fact that C_p does not depend on μ has been noted, together with the consequence that the value of p is not determined by that of C_p. The index C_{pk} was introduced to remedy this drawback, although p is not determined by C_{pk} alone, either, but (perhaps) because it is determined by C_p and C_{pk} together, from the formula

$$p = \Phi(-3(2C_p - C_{pk})) + \Phi(-3C_{pk}). \quad (11)$$

Of course, Equation (11) holds only under the condition that the normality condition on the distribution of X is sufficiently closely satisfied. In fact, under these conditions, C_{pk} can be regarded as a more influential index than C_p because it provides the upper bound

$$p \leq 2\Phi(-3C_{pk}), \tag{12}$$

all by itself.

It would be easy to construct a table of values of p with (C_p, C_{pk}) as entries. Vännman's (1998b) "process capability plots"—see also Deleryd and Vännman (1999) and Gabel (1990)—make it possible to derive rapid estimates of p from values of C_p and C_{pk}, without recourse to detailed tables assessing simultaneously a deviation from the target value. It might well be objected that such a table could also be constructed, even more easily, using values of $(U-\mu)/\sigma$ and $(\mu-L)/\sigma$ as entries in Equation (9). A possible objection to this suggestion, providing an interesting insight into some aspects of PCI psychology, is that "we are familiar with PCIs, but the Z-ratios $(U-\mu)/\sigma$ and $(\mu-L)/\sigma$ appear less relevant" (despite the fact that their sum is just $6C_p$).

This statement is not quoted to be an object of ridicule. It is evidently valid from the point of view of many practitioners who use PCIs regularly. From the point of view of theoreticians (such as ourselves), however, there is an unfortunate element of arbitrariness in the construction and use of PCIs. With some embarrassment, we admit that C_{pmk} [see Equation (4)], for which we are among the proposers, is especially open to criticism on this point. While C_p and C_{pk}, together at least, are related to expected proportion NC [see Equation (11)] and C_{pm} is related to a (possibly arbitrarily chosen) loss function [see Equation (8)], C_{pmk} is a mixture of the two, but not specifically related to either. Our embarrassment is not substantially reduced by the fact that several other PCIs unfortunately share the same defect.

Much recent work can be assessed in the light of reconciling the viewpoints of practitioners and theoreticians, largely on the basis of closer examination of the nature of PCIs. Another approach is the construction of novel PCIs, becoming now available in a rich, though sometimes bewildering variety ("The Avalanche," according to Kotz and Lovelace (1998, Chapter 4)). Many practitioners feel, with some reason, that such variety can lead only to confusion. Our survey concentrates on these aspects and attempts to outline current trends. As mentioned above, we limit ourselves primarily to univariate measurements, but we also include a somewhat condensed summary of multivariate PCIs later, reflecting the current relative sparseness of their use. We will first introduce some more recent newcomers among PCIs, and then discuss problems in the estimation of values of PCIs (old and new) from observed data ("samples").

We start by considering a class of PCIs directly aimed at the control of expected proportion NC (i.e., p). These appeal to both down-to-earth practitioners and theoreticians, who argue "why not use, as your indices, quantities based directly on observed or estimated proportion of NC output or on an estimate of a loss function, if such has been sufficiently well established?" Why not, indeed?

PCIs Based on Expected Proportion NC

Before embarking on studies of specific PCIs, in all their "bewildering variety," it is essential to keep in mind the basic assumption that a state of statistical control has been attained, using whatever long-established methods such as subgroup sampling, control charts, etc. are needed. This does not necessarily mean that X is assumed to have a normal distribution or, indeed, any specific distribution. It only means that the distribution does not change in the course of use of the PCI, and that observed values of X have no dependences among themselves. There have been some studies of effects of departures from these conditions—our bibliography includes references to such work by Chen and Hsu (1998a), Christofferson (1999), Shore (1997) and Zhang (1998). Also a note by Nelson (1999) includes some reflections on these matters, but we will not discuss them here.

As early as 1991, Carr (1991) suggested that one might just use the expected proportion NC as a PCI, possibly estimated by $\widehat{p}$, the observed proportion NC. Others, clearly bitten by a powerful PCI bug, even suggested transforming back to a respectable PCI-type value $(1/3)\Phi^{-1}(1-(1/2)p)$—which is the corresponding C_p value if $\mu = M$ and X has a normal distribution.

Yeh and Bhattacharya (1998) propose use of a PCI based on the ratios of expected proportion NC to actual observed or estimated proportion NC. In itself this is simply the ratio

$$\frac{p_0}{p},$$

where p_0 is the desired proportion of NC output and

p is the actual proportion. This PCI is simply estimated by

$$\frac{p_0}{\widehat{p}}.$$

(Yeh and Chen (1999) have extended this to multivariate cases.) Another PCI, suggested by Yeh and Bhattacharya (1998), distinguishes between NC items for which X is less than L, and those for which X is greater than U. Using superscripts L, U to denote values applicable to NC by reason of X being less then L, or greater than U, respectively, they suggest using the PCI

$$C_f = \min\left\{\frac{p_0^L}{p^L}, \frac{p_0^U}{p^U}\right\}. \tag{13}$$

Flaig (1992, 1996/7, 1999, 2000) strongly supports the use of "fraction conforming" [= 1 − proportion NC] as a basis for PCIs which will be suitable for any unimodal distribution for X, using the Camp-Meidell inequality (a variant of the Chebyshev inequality)

$$\Pr\left[|X-\mu| < k\sigma\right] \geq 1 - \frac{4}{9k^2}$$

or, equivalently,

$$p \leq \frac{4}{9k^2}. \tag{14}$$

Singpurwalla (1998) uses the even broader Chebyshev inequality, applicable for any distribution of X, but replaces $1-4/9k^2$ in Equation (14) by the weaker $1-1/k^2$. These approaches do not seem very efficient in the context of the very small values of p desired in many applications of PCIs. In the case of possibly asymmetric tolerances, with $L = \mu - k_1\sigma$ and $U = \mu + k_2\sigma$ for $(k_p, k_2 > 0)$ we have:

$$\begin{aligned} p &= \Pr[X < \mu - k_1\sigma] + \Pr[X > \mu + k_2\sigma] \\ &= \Pr[X - \mu < -k_1\sigma] + \Pr[X - \mu > k_2\sigma] \\ &\leq \Pr[|X-\mu| > \sigma \min(k_1, k_2)] \\ &\leq \frac{4}{9}\max\left(\frac{1}{k_1^2}, \frac{1}{k_2^2}\right). \end{aligned} \tag{15}$$

For $k_1 = k_2 = 3$, this gives $p \leq 4/81 = 4.84$ percent. This is considerably greater than the 0.27 percent from normal theory.

Flaig does not suggest using actual observed proportions of NC output, nor do most of those who suggest that p be the basis for a PCI. Rather, he and his coworkers suggest that some kind of distribution be fitted to X and the resultant tail probabilities used for estimating p. For example, Yeh and Bhattacharya (1998)—see above—suggest using the methods of Pickands (1975) and Smith (1987), prominent in extreme value distribution theory and practice, while Polansky (1998, 2000) used a popular kernel method of "smooth nonparametric" fitting.

Before leaving this topic, we note that expected proportion NC can be expressed in terms of the loss function approach by using the loss function

$$\text{Loss } = \begin{cases} 0 & \text{if } L \leq X \leq U \\ 1 & \text{otherwise}. \end{cases} \tag{16}$$

See, for example, Palmer and Tsui (1999).

Modifications of the Basic PCIs

As time has passed, assessments of usefulness and ways of interpreting PCIs have themselves developed in a sometimes bewildering variety. Informative discussions, of some generality, are to be found in Kotz and Lovelace (1998), Palmer and Tsui (1999), and Singpurwalla (1998), the last of which includes some contributions from discussants.

Notable among general impressions and those derived from some more detailed analyses are that

(a) C_{pm} is unreliable if the expected proportion NC is regarded as the most important feature (Ruczinski (1996) provides a table showing how the same value of C_{pm} can be associated with a wide range of values of the expected proportion NC);

(b) C_{pmk} is even worse in these circumstances;

(c) More attention should be paid to possible effects of non-normality of the distribution of X and ways of reducing these effects;

(d) C_{pk} seems to have the greatest degree of acceptability among the basic PCIs (indeed, from our limited information, it appears to be the one most commonly employed).

An enlightening view of relations among our basic PCIs can be obtained from studies of the "superstructure PCIs" introduced by Vännman (1995). These are defined by

$$C_p(u,v) = \frac{d - u|\mu - M|}{3\sqrt{\sigma^2 + v(\mu - T)^2}} \quad (u, v \geq 0). \tag{17}$$

The four basic PCIs are included in this class:

$$C_p \equiv C_p(0,0); \qquad C_{pk} \equiv C_p(1,0);$$
$$C_{pm} \equiv C_p(0,1); \qquad C_{pmk} \equiv C_p(1,1).$$

Vännman devotes special attention to the case $u = 0$. From detailed and ingenious numerical studies based on statistical considerations involving power of a test argument she suggests that taking $u = 0$, $v = 4$ will produce a useful PCI. To appreciate her conclusions about this particular choice of u and v an interested reader is referred to her paper. In the authors' opinion this work constitutes an important breakthrough in the theory of process capability indices with an eye on applications.

Spiring (1997) also defines a PCI, $C_p^{(\omega)} \equiv C_p(0,\omega)$. However, in this definition, ω is not necessarily a constant; it may be a function of $|\mu - T|/\sigma$. In principle this allows $w[(\mu-T)/\sigma]^2$ in Vännman's formula to be replaced by any function of $|\mu - T|/\sigma$. So, in effect,

$$C_p^{(\omega)} = \frac{C_p}{\sqrt{1+g\left(\frac{|\mu-T|}{\sigma}\right)}} \tag{18}$$

with a general choice of function $g(\cdot)$, though for practical purposes it should be a positive, increasing function.

In a sense this leads us back to the relation between C_{pmk} and C_{pk} given by

$$C_{pmk} = \frac{C_{pk}}{\sqrt{1+\left(\frac{\mu-T}{\sigma}\right)^2}},$$

and that between C_{pm} and C_p given by Equation (6). A further interesting relation is

$$C_{pmk} = \frac{C_{pm}C_{pk}}{C_p}. \tag{19}$$

Dealing with Non-normality of the Distribution of X

As already noted, the "6" in Equation (1) has been associated with the idea (hope?) that a normal distribution for X provides a satisfactory approximation. Of course, both practitioners and theoreticians realized that this would not always be the case, and some (at least) of the second group energetically busied themselves with the task of coming up with relevant information and suggestions. Some practitioners, on the other hand, have claimed that C_p need not be assessed on the grounds of direct relevance to properties of NC product, though it is not clear what other means of assessment are to be used.

At a relatively early date, Clements (1989), in an influential paper, suggested that "6σ" be replaced by the length of the interval between the upper and lower 0.135 percentage points of the distribution of X (this is 6σ for a normal $\mathrm{N}(\mu, \sigma^2)$ distribution).

The new PCI is then

$$C'_p = \frac{U-L}{\xi_{1-a}-\xi_a} = \frac{2d}{\xi_{1-a}-\xi_a}, \tag{20}$$

where ξ_a is defined by $\Pr[X \le \xi_a] = a$, taking $a = 0.00135$, so that ξ_{1-a}, ξ_a are the upper and lower 0.135 percentiles of the distribution of X. [For a $\mathrm{N}(\mu, \sigma^2)$ distribution $\xi_{1-a} = \mu + 3\sigma$, $\xi_a = \mu - 3\sigma$.] Note that Veevers' (1998) viability index [see Equation (10)] is equal to $1 - 1/C'_p$ for any symmetric unimodal distribution.

The corresponding definition for C'_{pk} is

$$C'_{pk} = \frac{d - |\xi_{0.5} - M|}{\frac{1}{2}(\xi_{1-a}-\xi_a)}. \tag{21}$$

The replacement of the expected value (e.g. μ) by the median, $\xi_{0.5}$, is a natural, though not essential, choice.

Clements (1989) suggested fitting a Pearson system distribution for X, in order to obtain the required ξ_a values. Applications of this kind of method, with various assumed distributional forms, have been quite numerous since 1992. References include: Rodriguez (1992), Bittanti et al. (1998), Lovera et al. (1997) - all Pearson system; Castagliola (1996) - Burr distributions; Farnum (1996/7), Polansky et al. (1998/9), Pyzdek (1992) - all Johnson system; Padgett and Sengupta (1996) - Weibull and log normal; Mukherjee and Singh (1997-8) - Weibull; Sarkar and Pal (1997) - extreme value; Somerville and Montgomery (1996/7) - t, gamma and log normal; Sundaraiyar (1996) - inverse Gaussian. As mentioned above, Polansky (1998, 2000) uses a general kernel fitting method.

A closely related approach (e.g. Rivera et al. (1996)) is to transform X to $Y = h(X)$, where $h(X)$ is a continuous monotonically increasing function of X, for which a normal distribution is (assumed to be) a satisfactory approximation, and then to use original formulas with U, L, and σ replaced by $h(U)$, $h(L)$ and the standard deviation of Y, respectively. However, this method tends to be regarded unfavorably by practitioners, because it does not relate clearly enough to the original specification limits.

Among other methods we note that Kotz and Johnson (1993b) suggest replacing "6σ" by "5.15σ," because the approximate relation

$$\Pr[\mu - 2.575\sigma \le X \le \mu + 2.575\sigma] \approx 0.99$$

varies only a little among gamma distributions (see Merrington and Pearson (1958)). This might be useful if there are insufficient data to provide sufficiently accurate fitting. On the other hand, there are two important drawbacks: the most important being the limitation to about 1 percent for the value of p when the value of the modified C_p is 1, which is too high in most cases; and the other, somewhat less formidable, drawback is the required limitation to gamma-type distributions, though this includes a wide range of distributional shapes, from exponential to normal (as a limiting case).

Wright (1995) suggested the PCI

$$C_s = \frac{d - |\mu - M|}{3\sqrt{\sigma^2 + (\mu - T)^2 + |\mu_3/\sigma|}} = \frac{(d/\sigma) - (|\mu - M|/\sigma)}{3\sqrt{1 + [(\mu - T)/\sigma]^2 + |\sqrt{\beta_1}|}} \quad (22)$$

as a "PCI sensitive to skewness" where $\mu_3 = E[(X - E(X))^3]$ and $\sqrt{\beta_1} = \mu_3/\sigma^{3/2}$ $(= \alpha_3)$ is an established measure of skewness.

Chen and Kotz (1996) suggest inserting a multiplier $\gamma > 0$ before $|\sqrt{\beta_1}|$. The value of γ may be chosen to meet a desired optimality requirement(s).

Bai and Choi (1997) have constructed PCIs for use with possibly skewed distributions of X, based on a 'weighted variance' (WV) approach. This utilizes different divisors at the upper and lower limits (U, L) of the specification interval.

With $\Pr[X \le \mu] = P$, they define the PCI corresponding to C_p as

$$C_p^w = \frac{d}{3\sigma\sqrt{2}} \left[\min\left(\frac{1}{\sqrt{P}}, \frac{1}{\sqrt{1-P}}\right)\right] = \frac{C_p}{W}, \quad (23)$$

where $W = \sqrt{1 + |1 - 2P|}$. Since $W \ge 1$, $C_p^w \le C_p$ with equality if $P = 1/2$. We also have

$$C_{pk}^w = \min\left(\frac{U - \mu}{3\sigma\sqrt{2P}}, \frac{\mu - L}{3\sigma\sqrt{2(1-P)}}\right) = \frac{C_{pk}}{W}$$

and $C_{pm}^w = C_{pm}/W_T$ where $W_T = \sqrt{1 + |1 - 2P_T|}$ with $P_T = \Pr[X \le T]$.

The "flexible PCI" of Johnson et al. (1994) also treats the two limits differently. This PCI is essentially a modification of C_{pm} and is defined as

$$C_{jkp} = \frac{1}{3\sqrt{2}} \min\left(\frac{U - T}{MSE+}, \frac{T - L}{MSE-}\right), \quad (24)$$

where $MSE+ = E[(X - T)^2 \mid X > T]\Pr[X > T]$ and $MSE- = E[(X - T)^2 \mid X < T]\Pr[X < T]$. The multiplier $\sqrt{2}$ in the denominator relates to the fact that if X has a symmetrical distribution with $E[X] = T$, then $MSE+ = MSE- = \sigma^2/2$. See also Franklin and Wasserman (1994).

We have already mentioned that Yeh and Bhattacharya (1998) propose a PCI based on comparison between allowable proportions p_0^L, p_0^U for which $X < L$, $X > U$ respectively and the corresponding actual proportions p^L, p^U. See Equation (13).

Estimation

The preceding discussion relates to properties of PCIs which depend on knowledge of the values of parameters (μ, σ, $\sqrt{\beta}$, etc.) in their formulation. Generally, values of these parameters are not known precisely, but have to be estimated from data obtained from samples of produced items. It is quite possible that the estimated values may differ substantially from the actual values, leading to serious discrepancies between estimated and actual properties (e.g. value of p). In this section we outline how some estimators might vary from the actual values being estimated.

So far as we are aware, the most prevalent methods of estimating the basic PCIs are to replace μ by $\bar{X} = n^{-1}\sum_{i=1}^n X_i$ and σ by $S = \sqrt{(n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2}$, where $X_1, X_2, \ldots, X_n$ are n independent values of X, or by an appropriate multiple of the range (greatest X_i – least X_i). Very often, estimation is based on values from a series of samples, combining the individual sample estimates into a single estimate.

On the common (though not always stated) assumption that X has a $N(\mu, \sigma^2)$ distribution, $\bar{X}$ has a $N(\mu, n^{-1}\sigma^2)$ distribution and S^2 has a $(n-1)^{-1}\sigma^2\chi^2$ distribution, and $\bar{X}$ and S are mutually independent. The resultant distributions of estimators $\widehat{C}_p$, $\widehat{C}_{pk}$, $\widehat{C}_{pm}$, $\widehat{C}_{pmk}$, of C_p, C_{pk}, C_{pm}, C_{pmk} respectively, have been thoroughly investigated (Ahmed (1998), Ahmed and Rohbar (1997), Bissell (1990), K.-S. Chen (1998b), Chou et al. (1990), Chou and Polansky (1993), Hoffman (1999), Hubele et al. (1999), Kane (1986), Kushner and Hurley (1992), Li et al. (1990), Mazzuchi (1997), Mukherjee (1995), Nagata (1995a,b), Nagata and Nagahata (1992), Pearn et al. (1992), Subbaiah and Taam (1993), Vännman (1997c), Vännman and Kotz (1995a), Wright (1998, 2000), Zimmer and Hubele (1997/8), and Zimmer et al. (2000)). References to work on asymptotic distributions (as the number of available observed values of X increases) include Chan et al. (1990), Chen (1997), Chen and Hsu (1995), and Chen and Pearn (1997).

Recent investigations of asymptotic properties of estimators (spearheaded by Hallin and Seoh (1999)) indicate the importance of determining the sample sizes n for which asymptotic results are adequate. In many instances the appropriate values of n far exceed the "folklore values" of 25 or 30 suitable in the case of asymptotic results associated with the t-distribution. To the best of our knowledge this problem has not been as yet properly addressed as far as asymptotic distributions of PCIs are concerned.

Distributions of estimators of the basic PCIs when the distribution of X is not normal have been reported by Han et al. (2000).

Unusually, Little and Harrelson (1993) suggest using $2t_{n-1,0.9986}S$, in place of $6S$, in the denominator to estimate C_p.

Use of a multiplier of the average range in a number of subsamples of small size (< 10, usually) as an estimator for σ has been prominent in quality assurance for a long time. Its susceptibility to nonnormality in the distribution of X, however, renders it an unattractive procedure for estimating PCIs, especially in regard to bias, but also in regard to variability (even if approximate normality holds). This is because the distribution of range is very sensitive to variation in the tails of the parent distribution.

Distributions of estimators of Wright's PCI, C_s, have been derived by Chen and Kotz (1996) and Sundaraiyar (1996). Distributions of estimators of Vännman's $C_p(u,v)$ PCIs are derived in Kalyanasundaram and Balamurali (1997), Vännman (1997b, 1997c, 1998b), and Vännman and Kotz (1995a, b).

Mittag (1997) has drawn attention to the fact that if measurement errors are superimposed on the values of X, producing observed values of a variable, X^* say, with a different distribution, then we are, instead of estimating C_{pk}, C_{pm}, C_{pmk}, etc., actually estimating C^*_{pk}, C^*_{pm}, C^*_{pmk}, etc.—values obtained by replacing μ, σ by new values,

$$\mu^* = \mu + \delta; \qquad \sigma^* = \sqrt{\sigma^2 + \sigma_\delta^2}, \qquad (25)$$

where δ, σ_δ^2 are the expected value and variance, respectively, of the measurement error (on the assumption that this error is independent of X). On the further assumptions that both X and the measurement error are normally distributed, the distributional results obtained for C_p, C_{pk}, C_{pm}, C_{pmk} in the articles mentioned earlier in this Section will apply to C^*_p, C^*_{pk}, C^*_{pm}, C^*_{pmk} (with μ, σ replaced by μ^*, σ^*).

Mittag's work is included in the book by Rinne and Mittag (1999). Persijn and VanNuland (1996/7) also discuss problems concerning measurement error, in a somewhat different manner and, more recently, Newton (1999) discusses them briefly. These studies are of importance, because measurement error can severely affect estimation of PCIs, with unfortunate consequences on inference regarding process capability.

Scholz and Vengel (1998) have addressed the problem of constructing confidence bounds for C_{pk}, allowing for correlation between values of X for items within the same batch. They describe the use of an "effective sample size"—less than the actual total sample size—that leads to reasonably adequate approximate results.

Multivariate PCIs

A more precise title for this Section would be "PCIs for Use When **X** is Multivariate." Many of the PCIs in this group are not, in fact, multivariate. Maybe they should be, but writers have opted for construction of univariate PCIs, based on the multivariate distributions of **X**. Nevertheless we will term them all "multivariate PCIs" (MPCIs). References with a title including the word "multivariate" or "bivariate" are: Beck and Ester (1998); Bernardo and Irony (1996); Boyles (1996b); Chan et al. (1991); Davis et al. (1992); Hellmich and Wolff (1996); Hubele et al. (1991); Karl et al. (1994); Li and Lin (1996); Mukherjee and Singh (1994); Niverthi and Dey (2000); Shariari et al. (1995); Taam et al. (1993); Tang and Barnett (1998); Veevers (1995, 1998, 1999); Wang et al. (2000); Wierda (1992, 1993, 1994a, 1998); and Yeh and Chen (1999). Multivariate situations are also discussed in the following references, that do not indicate, explicitly, in their titles that this is so: Chan et al. (1988b); Wang and Chen (1998/9); and Wang and Hubele (1999, 2001).

The univariate specification interval ($L \leq X \leq U$) is now replaced by a specification region. This may just be constructed from separate specification intervals: one for each variable X_i in **X**. The specification region is then the hyperrectangle

$$\bigcap_{i=1}^{\nu} (L_i \leq X_i \leq U_i). \qquad (26)$$

However, more complex regions may be used, reflecting perceived relations among the variables in $\mathbf{X}$. These are of the general form

$$L \le g(\mathbf{X}) \le U. \quad (27)$$

Often, L is zero. Possibly for mathematical convenience, $g(\mathbf{X})$ is often taken as a monotonic function of the joint probability density function of $\mathbf{X}$. Thus if $\mathbf{X}$ is assumed to have a multivariate normal $N_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, one might take

$$g(\mathbf{X}) = (\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \quad (28)$$

and regard an item as NC if $g(\mathbf{X}) > U$. In this way we obtain the ellipsoidal specification region

$$(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \le U. \quad (29)$$

An analogue of C_p is

$$\frac{\text{Volume of } \{(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \le U\}}{\text{Volume of } \{(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \le R\}} = \left(\frac{U}{R}\right)^\nu, \quad (30)$$

where $\Pr\left[(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \le R\right] = 1 - p$.

If the distribution of $\mathbf{X}$ is multivariate normal then $(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ has a χ^2 distribution with ν degrees of freedom and $R = \chi^2_{\nu,1-p}$ (the upper $100(1-p)\%$ point of the χ^2 or "chi-squared" distribution with ν degrees of freedom).

Chen (1994) applies this method to the case when the specification region is of the form in Equation (29). The region is defined by

$$\max_{i=1,2,\ldots,\nu} \left(\frac{|X_i - M_i|}{d_i}\right) \le 1 \quad (31)$$

with $M_i = (L_i + U_i)/2$ and $d_i = (U_i - L_i)/2$, and Chen defined MC_p as R^{-1}, where

$$\Pr\left[\max_{i=1,2,\ldots,\nu} \left(\frac{|X_i - M_i|}{d_i}\right) \le R\right] = 1 - p. \quad (32)$$

There can be many variants on these approaches. For example, the $g(\mathbf{X})$ in Equation (28) might be replaced by $(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ where $\mathbf{A}$ is a positive matrix, not necessarily the variance-covariance matrix of the distribution of $\mathbf{X}$.

Shariari et al. (1995) proposed a truly multivariate MPCI. It contains three components. The first is of the type in Equation (32). The second is the significance level of the Hotelling's T^2 statistic

$$T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}), \quad (33)$$

which is

$$\Pr\left[F_{\nu,n-\nu} > \frac{n-\nu}{\nu(n-1)}T^2\right], \quad (34)$$

where $F_{\nu,n-\nu}$ denotes a variable having the F-distribution with ν, $n-\nu$ degrees of freedom. The final component just takes values 1 or 0 according to a modified process region—defined as the smallest region similar in shape to the specification region, circumscribed about a specified probability contour (of the distribution of $\mathbf{X}$)—is or is not entirely contained in the specification region.

Wang et al. (2000) compared this 3-component MPCI with Chen's (1994) MC_p and with an index MC_{pm} proposed by Taam et al. (1993) which is also a ratio of two volumes. The volume in the denominator is the same as in Equation (30) with $R = \chi^2_{\nu,1-p}$ with $p = 0.0027$, while in the numerator we have the volume of a "modified specification region" which is the largest ellipsoid centered at the target that is within the original specification region.

Veevers' (1995) univariate viability index V_t (see Equation (10)) can be extended naturally to multivariate $\mathbf{X}$, defining

$$V_p = \frac{\text{volume of "window of opportunity"}}{\text{volume of specification region}}. \quad (35)$$

Veevers considers only the case of rectangular specification regions of type in Equation (31), and shows that, denoting the viability index for X_i by V_{ti}, we have

$$V_t = \begin{cases} \prod_{i=1}^{\nu} V_{ti} & \text{if } V_{ti} > 0, \\ & \text{for all } i = 1, \ldots, \nu \\ 1 - \prod_{V_{ti}<0} (1 - V_{ti}) & \text{otherwise.} \end{cases}$$

If any one X_i is not viable, then $V_t < 0$ and $\mathbf{X}$ is not viable.

We do not pursue the topic of multivariate PCIs any further here. There is clearly much room for inventiveness, even though we have not covered many of the indices already proposed. Our reason is that so far the variety of MPCIs outstrips their *present* amount of employment. Of course, suggestions for remedying this situation, either by increased attention to possibilities of application or by improved understanding of possible utility, are most welcome. This problem cannot be relegated to the back burner.

Concluding Remarks

Given the complexities of connections among many of the PCIs we have discussed, it is not surprising that several articles have appeared attempting to guide workers in the selection, use, and interpretation of PCIs. Study of these articles can be well worthwhile, insofar as the reader may recognize some situation(s) with which he/she is familiar, and even derive enlightening information from the text. Among our references, we draw special attention to the following articles, wherein use and interpretation of PCIs are treated in a quite general and very thoughtful manner: Cheng (1994/5), three comprehensive treatises by Deleryd (1998a, 1998b, 1999a), Kaminski, Dovich, and Burke (1998), Kotz and Lovelace (1998, Chapter 1), Palmer and Tsui (1999), Rodriguez (1992), Singpurwalla (1998), and Tsui (1997).

It is not to be expected that these authors are all in total agreement, of course. However, they do have some points in common—notably on the unfortunate non-dependence of C_p on the value of μ. There is also wide agreement on the need for establishment of a state of statistical control before PCIs are used. Of course, substantial variation in a PCI index, over time, will itself provide evidence of lack of statistical control.

We have referred to the use of various methods for estimating tails of distributions in order to obtain (in effect) estimates of p (expected proportion NC). Such indirect estimation of p—as opposed to direct observation of proportion NC—reflects, we believe, some anxiety about the sparseness of data in many applied situations. There is little, or no, comparison of accuracy of estimators from the two methods, but it seems that there should, in any case, be more appreciation of the effects of sampling variation on values of PCIs estimated from data. It is necessary to distinguish between the properties of PCIs as defined and those of the estimators of the PCIs. Also, there is obviously need for great skepticism about values of p based on extreme tails of assumed forms of distributions. For example, much analysis depends on assumptions about extreme tail probabilities of normal distributions. In spite of numerous optimistic assertions in the literature, it is *most* unlikely that a normal distribution will fit the actual distribution of X accurately as far out as four or five (even three) times σ from the value of μ. Practically always, we know that a normal distribution will not be exact everywhere, since such a distribution does not exclude negative values. Nevertheless, as G.E.P. Box reminds us in the discussion of Singpurwalla (1998, p. 33) "all models are wrong, but some models are useful."

Also, one should guard against assuming that lack of normality for a characterstic is an indicator of inadequacy of a process. Normality, in itself, is not an essential feature of good production. Lack of normality, however, serves to indicate need for caution in the use of PCIs for quality assessment.

A few of our references (Adler and Shper (1994), Bernardo and Irony (1996), and Niverthi and Dey (2000)) refer to "Bayesian" PCIs. We have not discussed these, partly because we are still of the opinion that *excessive* reliance on prior information (even, for example, just "simple normality") is a somewhat risky enterprise, and partly because these three works represent valuable but initial innovative attempts.

We finally quote from the Editorial (Nelson (1992)) in the issue of *JQT* devoted to PCIs: "in fact, it is clear from a statistical perspective that the concept of attempting to characterize a process with a single number is fundamentally flawed." See also Herman (1989). This is, of course, equally valid for *any* summarizing statistical measure used without any qualms. Nevertheless, the statement that "process capability indices are here to stay" (Kotz and Lovelace (1998, page 16)) appears, fortunately or unfortunately, to also be true.

Postscript for Practitioners

There are several papers and monographs in the references containing accounts of specific applications of PCIs. These include Bittanti et al. (1998), Boyles (1996), Cheng (1994/5), Deleryd (1998a, 1999a), El.Awady et al. (1996), Hubele et al. (1994), Kotz and Lovelace (1998), Municheka (1992), Parlar and Wesolowsky (1999), Porter and Oakland (1991), Prasad and Calis (1999), Sarkar and Pal (1997), Schneider et al. (1995/6), Zhang and Feng (1999/2000). Although these will be of interest to some readers, they are too specific for us to discuss in further detail in this survey article, wherein we have tried to deal with rather general issues. We are, naturally, aware that "general" issues are the result of synthesis of many specific examples, but adequate treatment in detail would require a book, not just an article.

We hope that these remarks may serve as a prologue in preparation for a substantial volume, to be written by some dedicated and energetic team of enthusiastic researchers and practitioners sometime during the first decade of the 21st century, which will finally bridge the gap and reduce further the lack of coordination between various directions of research and applications of the PCIs.

A good survey of scientific or technological matters ought to raise profound questions for which there may not be immediate answers. We trust that we have at least partially succeeded in this respect. Until new (less ambiguous) ways of looking at process capability will be found, we will remain, as Milton says in his *Comus*, "in the blind mazes of this tangled wood."

Acknowledgments

The authors express their sincere thanks to the Editor for his encouraging and valuable guidance and to Nozer Singpurwalla for providing thoughtful discussion. Our special gratitude goes to Kerstin Vännman and Norma Hubele, who have established prominent centers of PCI research in northern Sweden and in southwestern United States, respectively, and to the Editor, for providing us with a number of important bibliographical sources. We are also indebted to the discussants for alerting us about some controversial points and for providing useful comments and suggestions on an earlier version of our paper.

References and Bibliography

ADLER, YA. and SHPER, V. (1994). "Some Remarks on Bayesian Approach to Process Capability Estimates". *5th International Meeting on Bayesian Statistics,* Alicante, Spain.

AHMED, S. E. (1998). "Test and Confidence Intervals for Process Capability Indices in k Arbitrary Populations". Department of Mathematical Statistics, University of Regina, Saskatchewan, Canada.

AHMED, S. E. and ROHBAR, M. H. (1997). "Inference Concerning a Process Capability Index". Masters Thesis. Department of Mathematical Statistics, University of Regina, Saskatchewan, Canada.

ANNIS, C. (2001). "Review of *Statistical Distributions in Engineering* by Karl Bury". *Technometrics,* 43, pp. 101–102.

BAI, D. S. and CHOI, S. S. (1997). "Process Capability Indices for Skewed Populations". Masters Thesis. Department of Industrial Engineering, Advanced Institute of Science and Technology, Taejon, South Korea.

BAILEY, S. P. (1995). "How to Deal With Process Capability – 'Catch-22"'. *Transactions of the ASQC Technical Conference,* pp. 1105–1116.

BARNES, R. and PIERCE, R. C. (1997). "Inference and Capability Indexes". *Proceedings of the ASA Section on Quality and Productivity,* pp. 101–102.

BECK, C. and ESTER, S. (1998). "Multicriteria Capability Indexes". *Spektrum* 24, pp. 179–187.

BERARDO, J. M. and IRONY, T. X. (1996). "A General Multivariate Bayesian Process Capability Index". *Statistician* 45, pp. 487–502.

BISSELL, A. F. (1990). "How Reliable is Your Capability Index?". *Applied Statistics* 39, pp. 331–340.

BITTANTI, S.; LOVERA, M.; and MOIRAGHI, L. (1998). "Application of Non-Normal Process Capability Indices to Semi-Conductor Quality Control". *IEEE Transactions on Semiconductor Manufacturing* 11, pp. 296–302.

BORGES, W. and HO, L. L. (1998). "A Fraction Non-Conforming Based Capability Index". Department di Estadistics, Universidede de Sao Paulo, Brazil.

BORGES, W. and HO, L. L. (2000). "On the Sampling Distribution of Clements' Capability Index". *Communications in Statistics—Simulation and Computation* 29, pp. 121–138.

BOTHE, D. R. (1992). "A Capability Study for an Entire Product". *Transactions of the ASQC Technical Conference,* pp. 171–178.

BOTHE, D. R. (1997). *Measuring Process Capability.* McGraw-Hill, New York, NY.

BOTHE, D. R. (1999). "A Capability Index for Multiple Process Streams". *Quality Engineering* 11, pp. 613–618.

BOYLES, R. A. (1991). "The Taguchi Capability Index". *Journal of Quality Technology* 23, pp. 17–26.

BOYLES, R. A. (1994). "Process Capability With Asymmetric Tolerances". *Communications in Statistics—Simulation and Computation* 23, pp. 613–643.

BOYLES, R. A. (1996a). "Exploratory Capability Analysis". *Journal of Quality Technology* 28, pp. 91–98.

BOYLES, R. A. (1996b). "Multivariate Process Analysis With Lattice Data". *Technometrics* 38, pp. 37–49.

CARR, W. E. (1991). "A New Process Capability Index: Parts per Million". *Quality Progress* 24, p. 152.

CASTAGLIOLA, P. (1996). "Evaluation of Non-Normal Process Capability Indices Using Burr's Distribution". *Quality Engineering* 8, pp. 587–593.

CHAN, L. K.; CHENG, S. W.; and SPIRING, F. A. (1988a). "The Robustness of Process Capability Index C_p to Departures from Normality" in *Statistical Theory and Data Analysis,* edited by K. Matusita. North Holland, Amsterdam, Netherlands.

CHAN, L. K.; CHENG, S. W.; and SPIRING, F. A. (1988b). "A New Measure of Process Capability, C_{pm}". *Journal of Quality Technology* 20, pp. 162–175.

CHAN, L. K.; CHENG, S. W.; and SPIRING, F. A. (1991). "A Multivariate Measure of Process Capability". *International Journal of Modeling and Simulation* 11, pp. 1–6.

CHAN, L. K. and MAK, T. K. (1993). "A Process Capability Indicator". *Total Quality Management* 4, pp. 33–45.

CHAN, L. K.; XIONG, Z.; and ZHANG, D. (1990). "On the Asymptotic Distribution of Some Process Capability Indices". *Communications in Statistics—Theory and Methods* 19, pp. 11–18.

CHEN, H. (1994). "A Multivariate Process Capability Index Over a Rectangular Solid Tolerance Zone". *Statistica Sinica* 4, pp. 749–758.

CHEN, H. F. (1997). "Asymptotic Analysis of a Class of Process Capability Indices". *Statistics* 20, pp. 149–162.

CHEN, H. F. and KOTZ, S. (1996). "An Asymptotic Distribution of Wright's Process Capability Index Sensitive to Skewness". *Journal of Statistical Computation and Simulation* 55, pp. 147–158.

CHEN, K. S. (1998a). "Incapability Index with Asymmetric Tolerances". *Statistica Sinica* 8, pp. 253–262.

CHEN, K. S. (1998b). "Estimation of the Process Incapability Index". *Communications in Statistics—Theory and Methods* 27, pp. 1263–1274.

CHEN, S. M. (1996). "Response to the 'Letter to the Editor'". *Communications in Statistics—Theory and Methods* 25, pp. 1407–1408.

CHEN, S. M. and HSU, N. F. (1995). "The Asymptotic Distribution of the Process Capability Index C_{pmk}". *Communications in Statistics—Theory and Methods* 24, pp. 1279–1291. (Erratum 25, pp. 1117–1119).

CHEN, S. M. and HSU, Y. S. (1998a). "Asymptotic Distribution of C_p(a,b) Based on Second Order Stationary Process". *Preprint.*

CHEN, S. M. and HSU, Y. S. (1998b). "Asymptotic Distribution of Estimators of the Vännman-type Indices Based on m-Dependent Stationary Processes". *Preprint.*

CHEN, S. M. and PEARN, W. L. (1997). "The Asymptotic Distribution of the Estimated Process Capability Index C_{pk}". *Communications in Statistics—Theory and Methods* 26, pp. 2489–2497.

CHENG, S. W. (1994/5). "Practical Implementation of the Process Capability Indices". *Quality Engineering* 7, pp. 239–259.

CHOI, B. C. and OWEN, D. B. (1990). "A Study of a New Process Capability Index". *Communications in Statistics—Theory and Methods* 19, pp. 1231–1245.

CHOU, Y. M. and ANDERSON, M. T. (1997). "Sequential Test of Some Process Capability Indices" in *Statistics of Quality.* Marcel-Dekker, New York, NY, pp. 31–37.

CHOU, Y. M.; OWEN, D. B.; and BORREGO, A. S. A. (1990). "Lower Confidence Limits on Process Capability Indices". *Journal of Quality Technology* 22, pp. 223–229.

CHOU, Y. M. and POLANSKY, A. M. (1993). "Power of Tests for Some Process Capability Indices". *Communications in Statistics—Simulation and Computation* 22, pp. 523–544.

CHRISTOFFERSON, J. (1999). "Process Capability Indices and Block Bootstrap for Serially Correlated Data". *Research Report 1999:4,* Division of Quality Technology and Statistics, Luleå University of Technology, Luleå, Sweden.

CLEMENTS, J. A. (1989). "Process Capability Calculations for Non-Normal Distributions". *Quality Progress* 22, pp. 95–100.

COLLINS, A. J. (1995). "Bootstrap Confidence Limits on Process Capability Indices". *The Statistician* 44, pp. 373–378.

CRAIN, C. R. (1993). "Process Capability Indices and Customer-Supplier Relations". *Proceedings of the ASA Section on Quality and Productivity,* pp. 156–158.

DAVIS, R. D.; KAMINSKY, F. C.; and SABOO, S. (1992). "Process Capability Analysis for Processes With Either a Circular or a Spherical Tolerance Zone". *Quality Engineering* 5, pp. 41–54.

DELERMO, M. (1997). "A Strategy for Mastering Variation". *Transactions of the ASQC Technical Conference,* pp. 760–768.

DELERYD, M. (1998a). "Enhancing the Industrial Use of Process Capability Studies". *Doctoral Thesis,* Luleå University of Technology, Luleå, Sweden.

DELERYD, M. (1998b). "On the Gap Between Theory and Practice of Process Capability Studies". *International Journal of Quality, Reliability and Management* 15, pp. 178–191.

DELERYD, M. (1999a). "A Pragmatic View on Process Capability Studies". *International Journal of Production Economics* 58, pp. 319–330.

DELERYD, M. (1999b). "The Effect of Skewness on Estimates of Some Process Capability Indices". *International Journal of Applied Quality Management* 2, pp. 153–186.

DELERYD, M.; DELTIN, J.; and KLEFSJÖ, B. (1999). "Critical Factors for Successful Implementation of Process Capability Studies". *Quality Management Journal* 6, pp. 40–59.

DELERYD, M. and VÄNNMAN, K. (1998). "Process Capability Studies for Short Production Runs". *International Journal of Reliability, Quality and Safety Engineering* 5, pp. 353–401.

DELERYD, M. and VÄNNMAN, K. (1999). "Process Capability Plots—A Quality Improvement Tool". *Quality and Reliability Engineering International* 15, pp. 213–227.

DUMITRESCU, M. E. and HUBELE, N. F. (1999). "An Entropic Framework for the Normal Distribution in Capability Analysis". *Communications in Statistics—Theory and Methods* 28, pp. 1361–1377.

EL.WADY, K.; SCHAPER, D.; and KAILATH, T. (1996). "Improvements in C_{pk} Using Real-Time Feedback Control". *IEEE Transactions on Semiconductor Manufacturing* 9, pp. 87–94.

FADDOUL, N. R.; ENGLISH, J. R.; and TAYLOR, C. D. (1996). "The Impact of Mixture Distributions in Classical Process Capability Analysis". *IEEE Transactions on Reliability* 28, pp. 957–966.

FARNUM, N. R. (1996/7). "Using Johnson Curves to Describe Non-Normal Process Data". *Quality Engineering* 9, pp. 329–336.

FINK, J. and WOLFF, H. (1995). "A New Approach for Describing and Controlling Process Capability". *Proceedings of the 5th International Workshop on Intelligent Statistical Quality Control,* Osaka, Japan, pp. 1–25.

FINLEY, J. C. (1992). "What is Capability?—or—What is C_p and C_{pk}?". *Transactions of the ASQC Technical Conference,* pp. 186–192.

FLAIG, J. J. (1992). "Issues Related to Capability Indices" in *Process Capability Analysis Software Manual,* Section 6, Applied Technology, San Jose, CA.

FLAIG, J. J. (1996/7). "A New Approach to Process Capability Indices". *Quality Engineering* 9, pp. 200–211.

FLAIG, J. J. (1999). "Process Capability Sensitivity Analysis". *Quality Engineering* 11, pp. 587–592.

FLAIG, J. J. (2000). "Issues Related to Capability Indices" in *Process Capability Analysis Software Analysis,* Section 6, Applied Technology, San Jose, CA.

FLAIG, J. J. and SPENCER, W. N. (2000). "Process Qualification Using Risk Analysis". *Masters Thesis,* Personal Communication.

FRANKLIN, L. A. (1995). "Improved Sample Size Determination Formula for Lower Confidence Limits for C_p and C_{pk}". *Proceedings of the ASA Section on Quality and Productivity,* pp. 55–58.

FRANKLIN, L. A.; COOLEY, B. J.; and ELROD, G. (1999). "Comparing the Importance of Variation and Mean of a Distribution". *Quality Progress* 32, pp. 90–94.

FRANKLIN, L. A. and WASSERMAN, G. S. (1992). "Bootstrap Lower Confidence Limits for Capability Indices". *Journal of Quality Technology* 24, pp. 196–210

FRANKLIN, L. A. and WASSERMAN, G. S. (1994). "Bootstrap Lower Confidence Limits Estimates for C_{jkp} (the new Flexible Capability Index)". *Pakistan Journal of Statistics* 10A, pp. 33–45.

GABEL, S. H. (1990). "Process Performance Chart". *ASQC Quality Congress Transactions,* San Francisco, CA, pp. 683–688.

GOODMAN, J. H. (1996). "Assessing Measurement Capability for Destructive and Nonrepeatable Tests". *Proceedings of the ASA Section on Physical and Engineering Sciences*, pp. 87–92.

GREENWICH, M. and JAHR-SHAFFROTH, B. (1995). "A Process Incapability Index". *International Journal of Quality, Reliability and Management* 12, pp. 58–71.

GRÜNLUND, U. (1990). "The Concept of Capability Index: Assumptions and Conclusions". *Research Report 1990:2,* Division of Quality and Technology, Luleå University of Technology, Luleå, Sweden.

GUIRGURS, G. H. and RODRIGUEZ, R. N. (1992). "Computation of Owen's Q Function Applied to Process Capability Analysis". *Journal of Quality Technology* 24, pp. 236–246.

GUNTER, B. (1989 a,b,c,d). "The Use and Abuse of C_{pk}, Parts 1-4". *Quality Progress* 22, (1) pp. 72–73; (3) pp. 108–109; (5) pp. 79–80; (7) pp. 86–87.

GUNTER, B. (1991 a,b,c,d,e). "Process Capability Studies, Parts 1-5". *Quality Progress* 24, (2) pp. 97–100; (4) pp. 102–106; (6) pp. 77–85; (8) pp. 123–132; (10) pp. 109–113.

GUPTA, A. K. and KOTZ, S. (1997). "A New Process Capability Index". *Metrika* 45, pp. 213–224.

HALLIN, M. and SEOH, M. (1999). "When Does Edgeworth Beat Berry and Essen? Numerical Evaluation of Edgeworth Expansion". *Journal of Statistical Planning and Inference* 63, pp. 19–38.

HAN, J. H.; CHO, J.-J.; and LEEM, C. S. (2000). "Bootstrap Confidence Limits for Wright's C_s". *Communications in Statistics—Theory and Methods* 29, pp. 485–505.

HELLMICH, M. and WOLFF, H. (1996). "A New Approach for Describing and Controlling Process Capability for a Multivariate Process". *Proceedings of the ASA Section on Quality and Productivity,* pp. 44–48.

HERMAN, J. T. (1989). "Capability Index—Enough for Process Industries?" *Transactions of the ASQC Quality Congress,* Toronto, pp. 670–675.

HOFFMAN, L. L. (1999). "Hypothesis Testing of C_p". Department of Statistics, University of Central Florida, Orlando, FL.

HOLMES, D. (1980). "A Quality Portfolio Management Chart". Stochas Inc., Schenectady, New York (Also in (1986). *Quality* 25, p. 67).

HSIANG, T. C. and TAGUCHI, G. (1985). "Tutorial on Quality Control and Assurance—The Taguchi Methods". *Joint Meetings of the American Statistical Association,* Las Vegas, Nevada, pp. 188.

HUANG, D.-Y. and LEE, R. F. (1995). "Selecting the Largest Capability Index from Several Quality Control Processes". *Journal of Statistical Planning and Inference* 40, pp. 335–346.

HUBELE, N. F.; BEAUMARIAGE, T.; BAWAJA, G.; HONG, S. C.; and CHU, R. (1994). "Using Experimental Design to Assess Capability of a System". *Journal of Quality Technology* 26, pp. 1–11.

HUBELE, N. F.; SHAHRIARI, H.; and CHENG, C. S. (1991). "A Bivariate Process Capability Vector" in *Statistical Process Control in Manufacturing* edited by J. B. Keats and D. C. Montgomery. Marcel Dekker, New York, NY, pp. 299–310.

HUBELE, N. F. and ZIMMER, L. S. (1994). "Average Run Lengths for a C_{pm} Control Chart". *Transactions of the ASQC Technical Conference,* pp. 577–581.

HULTING, F. L. (1993). "Process Capability Analysis with Geometric Tolerances". *Proceedings of the ASA Section on Quality and Productivity,* pp. 266–270.

JESSENBERGER, J. and WEIHS, C. (2000). "A Note on the Behavior of C_{pmk} With Asymmetric Specification Limits". *Journal of Quality Technology* 32, pp. 440–443.

JOHNSON, N. L. (1996). "Letter to the Editor". *Communications in Statistics—Theory and Methods* 25, p. 1405.

JOHNSON, N. L.; KOTZ, S.; and PEARN, W. L. (1994). "Flexible Process Capability Indices". *Pakistan Journal of Statistics* 10, pp. 23–31.

JOHNSON, T. (1992). "The Relationship of C_{pm} to Squared Error Loss". *Journal of Quality Technology* 24, pp. 211–215.

JURAN, J. M. (1974). *Juran's Quality Control Handbook,* 3rd ed. McGraw-Hill, New York, N.Y.

KALYANASUNDARAM, M. and BALAMURALI, S. (1997). "A Superstructure Capability Index for Short Run Production Processes and Its Bootstrap Lower Confidence Limits". *Economics of Quality Control* 12, pp. 85–93.

KAMINSKI, F. C.; DOVICH, R. A.; and BURKE, R. J.; (1998). "Process Capability Indices: Now and in the Future". *Quality Engineering* 10, pp. 445–453.

KANE, V. E. (1986). "Process Capability Indices". *Journal of Quality Technology* 18, pp. 41–52.

KARL, D. P.; MORISETTE, J. and TAAMS, W. (1994). "Some Applications of a Multivariate Capability Index in Geometric Dimensioning and Tolerancing". *Quality Engineering* 6, pp. 649–665.

KAZMERSKI, T. J. (1995). *Statistical Problem Solving in Quality Engineering.* McGraw-Hill, New York, N.Y.

KIRMANI, S.N.U.A.; KOCHERLAKOTA, K. and KOCHERLAKOTA, S. (1991). "Estimation of σ and the Process Capability Index Based on Sub-Samples". *Communications in Statistics—Theory and Methods* 30, pp. 275–291.

KOCHERLAKOTA, S. (1992). "Process Capability Index—Recent Developments". *Sankhyā* B 54, pp. 352–369.

KOCHERLAKOTA, S. and KOCHERLAKOTA, K. (1994). "Confidence Intervals for the Process Capability Ratio Based on Robust Estimators". *Communications in Statistics—Theory and Methods 23*, pp. 257–276.

KOTZ, S. and JOHNSON, N. L. (1993a). *Process Capability Indices*. Chapman and Hall, London, U.K.

KOTZ, S. and JOHNSON, N. L. (1993b). "Process Capability Indices for Non-Normal Populations". *International Journal of Mathematical Statistical Science* 1, pp. 35–44.

KOTZ, S. and JOHNSON, N. L. (1999). "Delicate Relations Among the Basic Process Capability Indices C_p, C_{pk}, and C_{pm}, and Their Modifications". *Communications in Statistics—Theory and Methods* 26, pp. 849–861.

KOTZ, S. and LOVELACE, C. (1998). *Introduction to Process Capability Indices*. Arnold, London, U.K.

KOTZ, S.; PEARN, W. L.; and JOHNSON, N. L. (1993). "Some Process Capability Indices are More Reliable Than One Might Think". *Applied Statistics* 42, pp. 55–62.

KUNJURANANTHRAM, S. (2001). "A Comparative Analysis of Confidence Interval Construction and Hypothesis Testing Procedures for C_{pm}". *Master's Thesis*, Department of Industrial Engineering, Arizona State University, Tempe, AZ.

KUSHLER, R. H. and HURLEY, P. (1992). "Confidence Bounds for Capability Indices". *Journal of Quality Technology* 24, pp. 188–196.

LEHRMAN, K. H. (1992/2). "Capability Analysis and Index Interval Estimates: Quantitative Characteristics (Variables)". *Quality Engineering* 4, pp. 93–129.

LEUNG, B.; YOUNG, A.; and SPIRING, F. A. (1997). "The Robustness of C_{pm} to Departures from Normality". *Technical Report No. 010397*, Department of Applied Mathematics, Hong Kong Polytechnical University, China.

LEVINSON, W. A. (1997). "Approximate Confidence Limits for C_{pk} and Confidence Limits for Non-Normal Process Capabilities". *Quality Engineering* 9, pp. 635–640.

LI, H.; OWEN, D. B.; and BORREGO, A. S. A. (1990). "Lower Confidence Limits on Process Capability Indices Based on Range". *Communications in Statistics—Simulation and Computation* 19, pp. 1–24.

LI, Y. and LIN, C. (1996). "Multivariate C_p Value". *Chinese Journal of Applied Probability and Statistics* 12, pp. 132–138.

LITTIG, S. J. and LAM, C. T. (1993). "Case Studies in Process Capability Measurement". *Transactions of the ASQC Technical Conference,* pp. 569–576.

LITTLE, T. A. and HARRELSON, C. S. (1993). "Short Run Capability Assessment Using the t-Distribution". *Transactions of the ASQC Quality Congress,* Boston, MA., pp. 181–191.

LOVERA, C.; MOIRAGHI, L.; MONTEFUSCO, A.; and VALENTINI, D. (1997). "Estimation of Industrial Process Capabilities: Some Proposals and Studies to Overcome the Obstacle of Non-Normal Distributions". *Abstract III,* National Congress, SIMAI, Pavia, Italy, pp. 627–631.

LUCEÑO, C. (1996). "A Process Capability Index with Reliable Confidence Intervals". *Communications in Statistics—Simulation and Computation* 25, pp. 235–245.

LUCEÑO, C. (1999). "Discrete Approximations to Continuous Univariate Distributions—an Alternative to Simulation". *Journal of the Royal Statistical Society* B 61, pp. 345–352.

MACKERLICH, N. A. and STEPHAN, V. (1997). "Capability Index for Nonnormal Processes". *Transactions of the ASQC Technical Conference,* pp. 769–773.

MARCUCCI, M. O. and BEAZLEY, C. C. (1988). "Capability Indices: Process/Performance Measures". *Transactions of the ASQC Quality Congress,* pp. 516–528.

MAZZUCHI, T. (1997). "Notes on C_p and C_{pk} Estimation". *Report,* School of Engineering, George Washington University, Washington, D.C.

MCCORMICK, D. W.; HURWITZ, A. M.; SPRAGUE, P. D.; and HARRIS, I. R. (1996). "Capability Indices for Non-Normal Data". *Proceedings of the ASA Section on Quality and Productivity,* pp. 108–113.

MITTAG, H. J. (1997). "Measurement Error Effects on the Performance of Process Capability Indices" in *Frontiers in Statistical Quality Control* edited by H. J. Lenz and P. Th. Wilrich, Heidelberg Physica-Verlag, pp. 195–206.

MUKHERJEE, S. P. (1995). "Process Capability Indices and Associated Inference Problems". *Proceedings of the International Conference on Statistical Method and Statistical Computation,* Seoul, South Korea pp. 243–249.

MUKHERJEE, S. P. and SINGH, N. K. (1994). "Sampling Properties of an Estimate of the Bivariate Process Capability Index". *Economic Quality Control* 9, pp. 73–78.

MUKHERJEE, S. P. and SINGH, N. K. (1997/8). "Sampling Properties of an Estimator of a New Process Capability Index for Weibull Distributed Quality Characteristics". *Quality Engineering* 10, pp. 291–294.

MUNICHIKA, M. (1992). "Studies on Process Capability in Machining Processes". *Reports of Statistical Applied Research JUSE* 39, pp. 14–29.

NAGATA, Y. (1995a). "Point Estimation of the Process Capability Index". *Okayama Economical Review* 26.

NAGATA, Y. (1995b). "On Normal Approximation to the Distribution of the Estimated Process Capability Index, C_{pk}". *Journal of the Japanese Statistical Society* 25, pp. 159–171.

NAGATA, Y. and NAGAHATA, H. (1992). "Approximation Formulas for the Confidence Intervals of Process Capability Indices". *Reports of Statistical Applied Research JUSE* 39, pp. 15–29.

NELSON, L. S. (1999). "The Calculation of Process Capability Indices". *Journal of Quality Technology* 31, pp. 249–250.

NELSON, P. R. (1992). "Editorial". *Journal of Quality Technology* 24, p. 175.

NEWTON, D. (1999). "Process Capability" in *Gower Handbook of Quality Management,* 2nd ed., edited by D. Lou. Gower. Aldershot, Hampshire, U.K.

NIVERTHI, M. and DEY, D. K. (2000). "Multivariate Process Capability: A Bayesian Perspective". *Communications in Statistics—Simulation and Computation* 29, pp. 667–687.

O'NEILL, E. E. and KRISHNAMOORTRY, K. S. (1997). "Capability Index for Non-Normal Processes". *Transactions of the ASQC Technical Conference,* pp. 769–773.

ORCHARD, T. (2000). "Quality Improvement". *Royal Statistical Society News* 27, p. 8.

OTTO, G. H. and ARNOLD, K. E. (1995). "Process Capability Regulation and the Consequences for Offshore Platform Discharge of Oil and Grease". *Proceedings of the ASA Section on Statistics and the Environment,* pp. 81–86.

PADGETT, W. J. and SENGUPTA, A. (1996). "Performance of Process Capability Indices for Weibull and Lognormal Distributions or Autoregressive Processes". *International Journal of Reliability, Quality and Safety Engineering* 3, pp. 217–229.

PALMER, K. and TSUI, K. L. (1999). "A Review and Interpretations of Process Capability Indices". *Annals of Operations Research* 87, pp. 31–47.

PARLAR, M. and WESOLOWSKY, G. O. (1999). "Specification Limits, Capability Indices and Centering in Assembly Manufacture". *Journal of Quality Technology* 31, pp. 317–325.

PEARN, W. L. and CHANG, C. S. (1997). "The Performance of Capability Index C_s on Skewed Distributions". *Communications in Statistics—Simulation and Computation* 26, pp. 1361–1377.

PEARN, W. L. and CHEN, K. S. (1996). "A Bayesian-Like Estimator of C_{pk}". *Communications in Statistics—Simulation and Computation* 25, pp. 321–329.

PEARN, W. L. and CHEN, K. S. (1997). "A Practical Implementation of the Process Capability Index C_{pk}". *Quality Engineering* 9, pp. 721–737.

PEARN, W. L. and CHEN, K. S. (1998). "New Generalization of Process Capability Index C_{pk}". *Journal of Appplied Statistics* 25, pp. 801–810.

PEARN, W. L. and KOTZ, S. (1994/5). "Application of Clements' Method for Calculating Second- and Third-Generation Process Capability Indices for Non-Normal Pearsonian Populations". *Quality Engineering* 7, pp. 139–145.

PEARN, W. L.; KOTZ, S.; and JOHNSON, N. L. (1992). "Distributional and Inferential Properties of Process Capability Indices". *Journal of Quality Technology* 24, pp. 216–231.

PEARN, W. L.; KOTZ, S.; and WANG, Q. (1995). "Estimating the Distribution of the Flexible Process Capability Index, C_{pq}". *Preprint,* University of Maryland, College Park, MD.

PEARN, W. L.; LIN, C. H.; and CHEN, H. S. (1998). "Distributional and Inferential Properties of the Process Accuracy and Process Precision Indices". *Communications in Statistics—Theory and Methods* 27, pp. 985–1000.

PEARN, M. and XEKALAKI, E. (1998). "On a Refinement of Certain Process Capability Indices". *Technical Report No. 56,* Department of Statistics, Athens University of Economics and Business, Athens, Greece.

PERSIJN, M. and VANNULAND, Y. (1996/7). "Relation Between Measurement System Capability and Process Capability". *Quality Engineering* 9, pp. 95–98.

PHILLIPS, G. P. (1993). "Target Ratio and Loss Ratio—New Capability Indices Express Pure Target Error and Expand the Uses of C_{pm}". *Proceedings of the ASA Section on Quality and Productivity,* pp. 266–270.

PIGNIATIELLO, J. J. and RAMBERG, J. S. (1993). "Process Capability Indices—Just Say 'No"'. *Transactions of the ASQC Technical Conference,* pp. 92–104.

PIKANDS, C. (1975). "Statistical Inference Using the Extreme Order Statistics". *Annals of Statistics* 3, pp. 119–131.

PILLET, M.; ROCHON, S.; and DUCLOS, E. (1997/8). "SPC-generalization of Capability Index C_{pm}: Case of Unilateral Tolerances". *Quality Engineering* 10, pp. 171–176.

POLANSKY, A. M. (1998). "A Smooth Nonparametric Approach to Process Capability". *Quality and Reliability Engineering International* 14, pp. 43–48.

POLANSKY, A. M. (2000). "An Algorithmn for Computing a Smooth Non-Parametric Capability Estimate". *Journal of Quality Technology* 32, pp. 284–289.

POLANSKY, A. M.; CHOU, Y.-M.; and MASON, R. L. (1998/9). "Estimating Process Capability Indices for a Truncated Distribution". *Quality Engineering* 11, pp. 257–265.

PORTER, L. J. and OAKLAND, J. S. (1991). "Process Capability Indices—an Overview of Theory and Practice". *Quality and Reliability Engineering International* 7, pp. 437–448.

PYZDEK, T. (1992). "Process Capability Analysis Using Personal Computers". *Quality Engineering* 4, pp. 419–440.

RINNE, M. and MITTAG, H. J. (1999). *Prozeßfähigkeitsmessung für die industrielle Praxis (with English summary),* Carl Hanser Verlag München Wien.

RIVERA, L. A. R.; HUBELE, N. F.; and LAWRENCE, F. P. (1996). "C_{pk} Index Estimation Using Data Transformation". *Computers and Industrial Engineering* 29, pp. 55–58.

RODRIGUEZ, R. N. (1992). "Recent Developments in Process Capability Analysis". *Journal of Quality Technology* 24, pp. 176–186.

RUCZINSKI, I. (1996). "The Relation Between C_{pm} And The Degree Of Includence". *Doctoral Dissertation,* University of Würzberg, Würzberg, Germany.

RUMPH, D. L. (1996). "Process Capability Indices: Heuristics for the Smart Case of a Quality Metric". *Proceedings of the ASA Section on Physical and Engineering Sciences,* pp. 368–369.

RUNGER, G. C. (1993). "Designing Process Capability Studies". *Quality Progress* 29, pp. 31–34.

SARKAR, A. and PAL, S. (1997). "Estimation of Process Capability Index for Concentricity". *Quality Engineering* 9, pp. 665–671.

SCHNEIDER, H.; PRUETT, J.; and LAGRANGE, C. (1995/6). "Uses of Process Capability Indices in the Supplier Certification Process". *Quality Engineering* 8, pp. 225–235.

SCHOLTZ, F. and VANGEL, M. (1998). "Tolerance Bounds and C_{pk} Confidence Bounds Under Batch Effects" in *Advances in Stochastic Models for Reliability, Quality and Safety* edited by W. Kahle, Evon Collani, J. Franz, and U. Jensen. Birkhauser, Basel, Chapter 24, pp. 361–373.

SHAHRIARI, H.; HUBELE, N. F.; and LAWRENCE, F. P. (1995). "A Multivariate Process Capability Vector". *Proceedings of the 4th Industrial Engineering Research Conference,* Nashville, TN, pp. 303–308.

SHIAU, J. J. H.; CHIANG, C. T.; and HUNG, H.-N. (1999). "A Bayesian Procedure for Process Capability Assessment". *Quality and Reliability Engineering International* 15, pp. 369–378.

SHORE, H. (1997). "Process Capability Analysis When Data are Autocorrelated". *Quality Engineering* 9, pp. 615–626.

SINGPURWALLA, N. D. (1998). "The Stochastic Control of Process Capability Indices". *Test,* pp. 1–33. (Discussion: D. R. Cox, D. K. Dey, A. Fries, J. K. Ghosh, M. A. Gömez-Villegas, T. Z. Irony, W. Klieman, S. Kotz, D. V. Lindley, M.F. McGrath, and D. Pea, pp. 33-79).

SMITH, R. L. (1987). "Estimating Tails of Probability Distributions". *Annals of Statistics* 15, pp. 1174–1207.

SOMERVILLE, S. E. and MONTGOMERY, D. C. (1996/7). "Process Capability Indices and Non-Normal Distributions". *Quality Engineering* 9, pp. 305–316.

SPIRING, F. A. (1995). "Process Capability: A Total Quality Management Tool". *Total Quality Management* 6, pp. 121–133.

SPIRING, F. A. (1997). "A Unifying Approach to Process Capability Indices". *Journal of Quality Technology* 29, pp. 49–58.

STEPHENSON, M. R. and MARTZ, K. L. C. (1998). "Capability Analysis in Asymmetric Situations". *Proceedings of the ASA Section on Quality and Productivity,* pp. 65–70.

SUBBAIAH, P. and TAAM, W. (1993). "Inference on the Capability Index C_{pm}". *Communications in Statistics—Theory and Methods* 22, pp. 53–560.

SUNDARAIYAR, V. H. (1996). "Estimation of a Process Capability Index for Inverse Gaussian Distribution". *Communications in Statistics—Theory and Methods* 25, pp. 2381–2398.

TAAM, W.; SUBBAIAH, P.; and LIDDY, J. W. (1993). "A Note on Multivariate Capability Indices". *Journal of Applied Statistics* 20, pp. 339–351.

TANG, P. F. and BARNETT, N. S. (1998). "Capability Indices for Multivariate Processes". *Technical Report,* Division of Computation, Mathematics, and Science, Victoria University, Melbourne, Australia.

TSUI, K.-L. (1997). "Interpretation of Process Capability Indices and Some Alternatives". *Quality Engineering* 9, pp. 587–596.

VÄNNMAN, K. (1995). "A Unified Approach to Capability Indices". *Statistica Sinica* 5, pp. 805–820.

VÄNNMAN, K. (1997a). "Distribution and Moments in Simplified Form for a General Class of Capability Indices". *Communications in Statistics—Theory and Methods* 26, pp. 159–171.

VÄNNMAN, K. (1997b). "A General Class of Capability Indices in the Case of Asymmetric Tolerances". *Communications in Statistics—Theory and Methods* 26, pp. 2049–2072.

VÄNNMAN, K. (1997c). "Interpreting Process Capability Using Graphical Methods". *Research Report Division of Quality Technology Statistics,* Luleå University of Technology, Luleå, Sweden.

VÄNNMAN, K. (1998a). "Families of Capability Indices for One-Sided Specification Limits". *Statistics* 31, pp. 43–66.

VÄNNMAN, K. (1998b). "Process Capability Plots: A Complement to Capability Indices". Invited Paper for the *6th International Workshop on Intelligent Statistical Quality Control,* University of Würzburg, Würzburg, Germany.

VÄNNMAN, K. and KOTZ, S. (1995a). "A Superstructure of Capability Indices: Distributional Properties and Implications". *Scandinavian Journal of Statistics* 22, pp. 477–491.

VÄNNMAN, K. and KOTZ, S. (1995b). "A Superstructure of Capability Indices: Asymptotics and its Implications". *International Journal of Reliability, Quality and Safety Engineering* 2, pp. 343–360.

VEEVERS, A. (1995). "A Capability Index for Multiple Response". *CSIRO Mathematics and Statistics Report DMS-095,* Australia.

VEEVERS, A. (1998). "Viability and Capability Indices for Multiresponse Processes". *Journal of Applied Statistics* 25, pp. 545–558.

VEEVERS, A. (1999). "Capability Indices For Multiresponse Processes" in *Statistical Process Monitoring and Optimization* edited by S. H. Park and G. G. Vining, Marcel-Dekker. New York, NY pp. 241–256.

VOELCKER, M. B. and MORSE, E. P. (1997). "De-mystifying C_p and C_{pk} In Manufacturing". *The Browne and Sharp Publications of American Manufacturing,* pp. 40–42.

WALKER, E. (1995). "Review of Kotz, S. and Johnson, N. L. (1993)". *Journal of the American Statistical Association* 90, p. 1127.

WANG, C. M. and LAM, C. T. (1996). "Confidence Limits for Proportion of Conformance". *Journal of Quality Technology* 28, pp. 439–445.

WANG, F. K. and CHEN, J. C. (1998/9). "Capability Index Using Principal Components Analysis". *Quality Engineering* 11, pp. 21–27.

WANG, F. K. and HUBELE, N. F. (1999). "Quality Evaluation Using Geometric Distance Approach". *International Journal of Reliability, Quality and Safety Engineering* 6, pp. 139–153.

WANG, F. K. and HUBELE, N. F. (2001). "Quality Evaluation of Geometric Tolerance Regions in Form and Location". *Quality Engineering,* 14, pp. 203–209.

WANG, F. K.; HUBELE, N. F.; LAWRENCE, F. P.; MISKULIN, J. O.; and SHARIARI, H. (2000). "Comparison of Three Multivariate Process Capability Indices". *Journal of Quality Technology* 32, pp. 263–275.

WEIGHT, T. (1996). "Comment on Collins (1995)". *The Statistician* 45, p. 126.

WHEELER, D. J. (1999). *Beyond Capability Confusion.* SPC Press, Knoxville, TN.

WIERDA, S. J. (1992). "Multivariate Quality Control-Estimation of the Percentage of Good Product". Technical Report, University of Groningen, Groningen, Netherlands.

WIERDA, S. J. (1993). "A Multivariate Process Capability Index". *Transactions of the ASQC Quality Congress,* pp. 342–348.

WIERDA, S. J. (1994a). "Multivariate Statistical Process Control—Recent Results and Directions for Future Research". *Statistica Neerlandica* 48, pp. 147–168.

WIERDA, S. J. (1994b). "Multivariate Statistical Process Control". Thesis, Wolters-Noordhof, Groningen, Netherlands.

WRIGHT, P. A. (1995). "A Process Capability Index Sensitive to Skewness". *Journal of Statistical Computation and Simulation* 52, pp. 195–203.

WRIGHT, P. A. (1998). "The Probability Density Function of Process Capability Index C_{pmk}". *Communications in Statistics—Theory and Methods* 27, pp. 1751–1789.

WRIGHT, P. A. (2000). "The Cumulative Distribution of Process Capability Index C_{pm}". *Statistics and Probability Letters* 47, pp. 249–251.

YANG, W. (1993). "The Minimum Variance Unbiased Estimator and Best Test of the Process Capability Index C_p". *Chinese Journal of Applied Probability and Statistics* 9, pp. 82–86.

YEH, A. B. and BHATTACHARYA, S. (1998). "A Robust Capability Index". *Communications in Statistics—Simulation and Computation* 27, pp. 565–589.

YEH, A. B. and CHEN, H. (1999). "A Nonparametric Multivariate Process Capability Index". *Preprint.*

ZHANG, N. F. (1997). "Combining Process Capability Indices". *Proceedings of the ASA Section on Quality and Productivity,* pp. 84–89.

ZHANG, N. F. (1998). "Estimating Process Capability Indexes for Autocorrelated Data". *Journal of Applied Statistics* 25, pp. 559–574.

ZHANG, Y. and FENG, X. D. (1999/2000). "Target Allocation for Maximizing Wear Allowance of Running Fits Based on Process Capability". *Quality Engineering* 12, pp. 169–176.

ZIEMER, T. L. (1993). "Graphical Methods for Process Control" in *Handbook of Statistics, Vol. 9: Computational Statistics*, pp. 999-1019, Elsevier Science North Holland: Amsterdam, Netherlands.

ZIMMER, L. S. and HUBELE, N. F. (1997/8). "Quantiles of the Sampling Distribution of C_{pm}". *Quality Engineering* 10, pp. 309–329.

ZIMMER, L. S.; HUBELE, N. F.; and ZIMMER, W. J.; (2001). "Confidence Intervals and Sample Size Determination of C_{pm}". *Quality and Reliability Engineering International* 17, pp. 51–68.

Key Words: *Capability Indices, Estimation, Nonconformities.*

Discussion

NORMA F. HUBELE

Arizona State University, Tempe AZ 85287

MY thanks are to Drs. Samuel Kotz and Norman Johnson for their lead article which contained both a provocative introduction and an annotated listing of the literature. I also express my appreciation to the Editor for his generosity in inviting me to help organize and participate in this panel discussion. It is in the spirit of active exchange of ideas from many points of view that knowledge and understanding are fostered. Needless to say, process capability indices and the environment surrounding their usage are fertile ground for such an active exchange. While the discussion will undoubtedly prove to be enjoyable, the primary intent is to contribute to the proper use of statistics for the advancement of our society.

A key distinction made at the outset is that process capability analysis includes substantially more than just the computation of an index. Process capability analysis has been advocated for nearly as long as control charting and monitoring. After process control has been established, capability is assessed. An assessment is essentially the act of comparing the distribution of data, or a model, to the engineering requirements, typically in the form of engineering specifications. If the process is deemed capable, then the process will be maintained using statistical process control methods. If, on the other hand, the process is deemed not capable, i.e., it is producing an unacceptable level of non-conforming product, then the process will undergo a process improvement stage and work toward an acceptable level of capability and control.

Hence, the central issue in this discussion is not "should we perform a capability analysis," but rather, "how do we *briefly* state what we found upon completing a capability analysis?" It is the search for *brevity* that has resulted in the creation and use of process capability indices. It is the desire to create a simple, numerical, dimensionless entity (frequently found in other engineering applications, such as a Reynolds number in the field of fluid mechanics). With such a dimensionless variable, organizations can set numerical goals and find ways to summarize the performance of diverse products and processes, as noted by Drs. Kotz and Johnson. As with most things in life, the simpler the message, the better.

Superficially, and to many less practiced users, PCIs are merely computed values. They allow for the brief summary of information about the engineering requirements and the process/product behavior associated with these requirements. As pointed out in the "Concluding Remarks" of Drs. Kotz and Johnson's paper, some articles have been written to guide the user in selecting an appropriate PCI; however, they are not all in agreement. One of the major problems, even for the most educated and well intentioned user, is that there are so many types of PCIs being advocated by various organizations that there is a state of confusion and, consequently, of abandonment. Originally intended to make life simpler, we now have a conservative estimate from Drs. Kotz and Johnson that there are about twenty variants of the univariate PCIs and seven multivariate PCIs. There is no mystery in why attention is being paid to picking the "right" index and little attention is being paid to appropriately using it in conjunction with an estimate of variability. Should we stop the drive to develop yet another index and focus on convincing the practitioner to use the available indices more responsibly?

Unfortunately, I do not believe that our choices are so simple. On the one hand I believe that practitioners need to fully understand and incorporate an appreciation of variability in their products and process, as well as their performance measures. On the other hand, I do not totally agree with the statement by Drs. Kotz and Johnson that most academics are well aware of the kinds of problems faced by practitioners. While both parties are sincere in their efforts, the mutual understanding of the two worlds of theory and practice is *always* going to be a goal.

As stated at the outset, capability analysis has

Dr. Hubele is a Professor in the Department of Industrial Engineering. She is a member of ASQ. Her e-mail address is hubele@asu.edu.

almost always been a part of establishing a statistical process control system. Hence, practitioners have always had to address some of the most difficult issues in applied statistics, including measurement and sampling issues. However, they have been able to address them, in a sense, by dividing the solution between the intelligent human inspector and the rationale behind good statistical practice. In capability analysis, the experienced quality control inspector (with extensive training in and understanding of engineering drawing and design intent) can make the necessary adjustments for measurement error when certifying conformance to engineering specifications on a part-by-part, feature-by-feature basis. Then, by collectively viewing the data, they can make the go/no-go decision with respect to process capability assessment. Such an assessment does not require a quantification of a tolerance region, per se, or a model of the process behavior.

In constructing a control system, good statistical practice has been advocated and taught for over fifty years. While not all theoretical foundations have been understood by the typical practitioner, the underlying notions of sampling and parameter estimation have been at least intuitively understood and been successfully used. We can attribute this to educated engineers, extensive training materials, and software support. As a consequence, control charting has been implemented in a wide variety of processes and products.

Now, on this backdrop, superimpose the desire to incorporate both engineering and process information into a simple function reflecting process performance. Not only is it now necessary to clearly define the tolerance region mathematically, to account for the sampling strategy and the possibility of measurement error, it is also necessary to define a function mapping the engineering specification and process data to a number that is easily interpretable. Consider Figure 1, drawn to illustrate the three-dimensions of the situation facing both the practitioner and theoretician concerned with PCIs. The three dimensions are the engineering specifications, the process behavior model, and the functional relationship between the two. Symbolically, a PCI can be represented as

$$PCI = f(\textit{Engineering Specifications},\ \textit{Process Behavior Model}).$$

The avalanche (as coined by Kotz and Lovelace (1998)) is merely a reflection of the fact that, for all practical purposes, there are innumerable types of engineering specifications, giving rise to innumerable types of data distributions, i.e., process behavior models, joined together in a limitless assortment of available functions. Most indices represent one point in this three-dimensional space (not withstanding the approaches proposed by Vännman (1995) and Spiring (1997)).

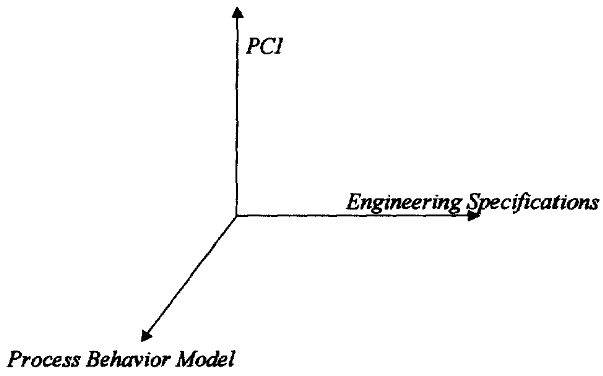

FIGURE 1. The 3-D Space Representing PCI.

Whereas the theoretician typically approaches research in PCIs from the process behavior modeling direction (e.g., assuming a normal distribution), the practitioner approaches the use of PCIs from the engineering specifications: "I have a flatness tolerance of 0.0020 inches per inch for this planar surface, which index should I use?" It is the expansive nature of engineering specifications that pose considerable problems for our engineering community when faced with reporting on process performance using a PCI.

Consider the partial list of geometric dimensioning and tolerancing (GD&T) specifications that could be attached to an engineering drawing, as shown in Table 1 (ASME (1994)). Such specification standards have been designed to create a common vocabulary to link design intent to manufacturing processes and inspection procedures, with the ultimate goal of reducing scrap. Engineering specifications necessarily influence how the part is manufactured, how each of the part features are measured, and, ultimately, how each of the processes should be described by a PCI. It is convenient, from a researcher's point of view, to approach the PCI development problem from the normal probability process model direction. However, only a subset of these specifications are typically seen to give rise to univariate, normally distributed measurements. In fact, it could be ar-

TABLE 1. Partial List of Geometric Dimensioning and Tolerancing Specifications

Straightness
Flatness
Perpendicularity
Parallelism
Circularity
Cylindricity
Symmetry

gued that the normal model for data arising from inspection of part features associated with GD&T is the *exception*, rather than the *norm*. Consequently, practitioners are burdened with the organizational edict to describe their process capability using an index, while the popular list of indices are frequently not appropriate for their situations. Even the most well-intentioned and well-educated practitioners find themselves between a rock and a hard place.

In this regard, it is my viewpoint that more needs to done in the development of capability functions (and their associate sampling distributions) that appropriately map the engineering specifications and process behavior models into brief descriptive indices. In the ideal world, there would be one functional form that captures all these discrete cases. Instead, we have a partial (and somewhat confusing) set of functions that only cover part of the space of engineering specifications and process behavior models.

Hence, the practitioner has to make some tough choices. With brevity and simplicity being important desirable characteristics of reported performance measures, the simplest forms of the PCIs are typically advocated and used (or misused). Software, the toolkit of the practitioner, is designed to support this need for brevity and simplicity. The difficult choices of the practitioner are further amplified for the software developer. Building a product suitable for a wide audience while advocating good statistical practice is the role of the software developer. Drs. Kotz and Johnson highlight the many subtleties in the computation of these process capability indices, primarily flowing from the estimation of the variance. Unless these subtleties are well-documented *and read* by the practitioner community, I am not so sure that software is a cure for the symptoms of misuse seen in the community. In the end, software can not replace understanding. And understanding flows from thoughtful reflection and education.

It should be noted that only recently has the accreditation board of engineering, ABET, created the requirement that all accredited engineering programs must have their students demonstrate "an ability to design and conduct experiments as well as to analyze and interpret data" (Engineering Accreditation Commission (1997)). Such a long-awaited feature of an engineer's education has come at a time when the technological and computer knowledge required of the graduating engineers has also increased. Consequently, the educational community is struggling to reconcile the engineering science content of the curriculum with engineering practice. While there is reason to be optimistic about the "next generation" of engineers understanding of variation in practice, there is still the issue of the necessity to tradeoff some of the science for some of the practice. Consequently, if one studies the recommendations that have been thoughtfully formulated over the past decade on what engineers need to know about probability and statistics, then one eventually realizes that the skills and methods useful for engineering problem solving are receiving the majority of the attention (e.g., Hogg (1994)). It is through the teaching of these skills and methods that we hope to pierce the veil surrounding the engineer's deterministic world. Only after a generation of engineering faculty have incorporated the notion of variability into courses *throughout* engineering education will the products of our educational system, i.e., the practitioners and industrial trainers, consistently recognize the importance of understanding and incorporating variation into their reporting systems. I postulate that this is an essential component for the cure of the disease diagnosed by Drs. Kotz and Johnson.

Additional References

ASME (1994). *Dimensioning and Tolerancing,* The American Society of Mechanical Engineers, New York.

ENGINEERING ACCREDITATION COMMISSION (1997). "ABET Engineering Criteria 2000". Accreditation Board for Engineering and Technology Inc., Baltimore, MD, 3rd Edition.

HOGG, R. V. (1994). "A Core in Statistics for Engineering Students". *The American Statistician* 48, pp. 285–293.

Discussion

FRED SPIRING

Pollard Banknote Ltd, Winnipeg, Manitoba, Canada R3T 1L7

SMILEY CHENG and ANTHONY YEUNG

The University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2

BARTHOLOMEW LEUNG

The Hong Kong Polytechnic University, Kowloon, Hong Kong

We would like to congratulate the *Journal of Quality Technology* and, in particular, its recent editors on their attitudes towards process capability and their willingness to promote reviews of subject matter areas that are not limited to bibliographies. We would also like to thank the Editor for the opportunity to participate in this forum and his encouragement to complete the project.

Much like the authors (Kotz and Johnson), we too were surprised and dismayed by the editorial comments contained in Nelson (1992). Our excitement regarding the current paper was, however, dampened somewhat upon review of the original manuscript. We felt the manuscript was neither an adequate bibliography of current work (as there were and continue to be many missing references) nor a reasonable review of philosophies surrounding process capability and process capability indices (PCIs). In particular, there appear to be "holes" in the reference list around 1991-1993, possibly attributable to the timing of Rodriguez (1992), and since 1998. An incomplete list of additional published references is included (see Additional Bibliography).

Dr. Spiring is an Adjunct Professor in the Department of Statistics, The University of Manitoba and Director of Quality at Pollard Banknote Ltd. He is a Senior Member of ASQ. His email address is spiring@cc.manitoba.ca.

Dr. Cheng is a Professor and Head of the Department of Statistics. He is a Senior Member of ASQ.

Mr. Yeung a Doctoral student in the Department of Statistics and an Analyst with Statistics Canada. He is a Member of ASQ.

Dr. Leung is a Lecturer in the Department of Applied Mathematics. He is a Member of ASQ.

Although Nelson (1992) tended to dampen general research in the area of process capability, several areas have continued to move ahead. As discussed by the authors, the research effort devoted to finding a better index for assessing capability has continued. The addition of the C_p alphabet appears to have had little impact on practical use; C_p and C_{pk} (including C_{pl} and C_{pu}) continue to be the most heavily used indices in practice, with C_{pm} occurring occasionally. Much of the development in the area of new PCIs has gone unused for many reasons, including lack of interpretation, software support, and dissemination. A second area of research that has continued to progress is that concerning the stochastic behavior/properties of the estimated PCIs. This is a positive development, since early research efforts in the area tended to focus on the PCIs with little statistical theory. There have been, however, statistical developments from authors that lack background knowledge of the use and interpretation of PCIs.

Interpreting Process Capability Measures

Traditionally, process capability measures have been used to provide insights into the number (or proportion) of non-conforming product. Practitioners cite a C_p value of one as representing 2700 parts per million (ppm) non-conforming, while 1.33 represents 63 ppm; 1.66 corresponds to .6 ppm; and 2 indicates .1 ppm. The interpretation of C_{pk} is similar, with a C_{pk} of 1.33 representing a maximum of 63 ppm non-conforming. A process with a C_p greater than or equal to one has traditionally been deemed capable. A C_p of less than one indicates that the process is producing more than 2700 ppm non-conforming, and it is used as an indication that the process is not capable of meeting customer requirements. In

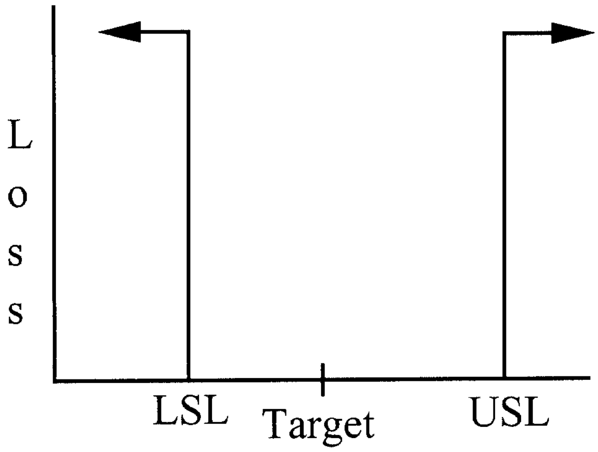

FIGURE 1. Square-Well Loss Function.

the case of C_{pk}, the auto industry frequently uses 1.33 as a benchmark in assessing the capability of a process.

Inherent in the translation of a numerical PCI into a number non-conforming is the assumption that product produced just inside the specification limit is of equal quality to that produced at the target. This is equivalent to assuming a square-well loss function (Figure 1) for the quality variable.

In practice, the magnitudes of C_p, C_{pl}, C_{pu}, and C_{pk} are interpreted as a measure of non-conforming, and therefore can be represented by the square-well loss function. Any change in the magnitude of these indices (holding the customer requirements constant) is due entirely to changes in the distance between the specification limits and the process mean. As Boyles (1991) points out, "C_{pk} does not in itself say anything about the distance between μ and T" and "is essentially a measure of process yield only." By design, C_p, C_{pl}, C_{pu}, and C_{pk} are used to identify changes in the amount of product beyond the specification limits (not proximity to the target) and are, therefore, consistent with the square-well loss function.

Taguchi used the quadratic loss function to motivate the idea that a product imparts no loss only if that product is produced at its target. He maintained that even small deviations from the target result in a loss of quality, and that as the product increasingly deviates from its target there are larger and larger losses in quality. This approach to quality and quality assessment is different from the traditional approach, where no loss in quality is assumed until the product deviates beyond its upper or lower specification limit (i.e., square-well loss function). Taguchi's philosophy highlights the need to have low variability around the target. Clearly, in this context, the most capable process will be one that produces its entire product at the target, with the next best being the process with the smallest variability around the target.

The motivation for C_{pm} does not arise from examining the number of non-conforming product in a process, but rather from requiring the ability of the process to be in the neighborhood of the target. This motivation has little to do with the number of non-conforming parts, although upper bounds on the number of non-conforming parts can be determined for numerical values of C_{pm} (Spiring (1991)). As discussed in Johnson (1992), C_{pm} is related to the quadratic loss function and is, thus, consistent with the Taguchi approach to quality.

Note that C_{pk} and C_{pm} have different functional forms, are represented by different loss functions, and have different relationships with C_p as the process drifts from the target (Spiring (1997)). Hence, although C_{pm} and C_{pk} are lumped together as second generation measures, they are very different in their development and assessment of process capability. A discussion surrounding the relationships of capability indices (C_p, C_{pk}, and C_{pm}) linked to specifications at the supplier level and specifications to the assembly level are cited in Parlar and Wesolowsky (1999). These differences manifest themselves in several areas associated with PCIs, two of which (robustness and interpretation) will be discussed here.

Effects of Non-normality and "Dealing with Non-normality of the Distribution of X"

If process measurements do not arise from a normal distribution, none of the traditional indices provide valid measures of the number of parts non-conforming. As many authors have pointed out, standard deviation has become synonymous with the term dispersion, but its physical meaning need not be the same for different families of distributions, or, for that matter, within a family of distributions. Therefore, as long as 6σ carries some practical interpretation when assessing process capability (i.e., is translated into ppm non-conforming), none of the indices should be used if the distribution of the characteristic under investigation is not normal.

Regardless of how robust an estimator may be, if its associated parameter is not stable, then any robustness claims carry little meaning. Similarly, developing actual and approximate confidence intervals for capability indices when the process characteristics arise from non-normal distributions is an academic pursuit with no application. For those capability indices that attempt to assess the ability of the process to cluster around the target, the robustness of the estimator is a valid concern.

Extensive studies have been conducted to determine the effects of non-normality on the various capability indices since Gunter (1989a-d) bemoaned the many flaws of C_{pk} in particular. A number of methods for handling non-normal data have been suggested. They can be classified in the categories of data transformation (Clements (1989) and Page (1994)), empirical percentile method (Clements (1989) and McCormark et al. (2000)), and Monte Carlo simulation (English and Taylor (1993) and Somerville and Montgomery (1996)). However, for those capability indices that attempt to assess the ability of the process to cluster around the target, the robustness of the estimator is an important question.

As discussed, C_{pm} is used to provide an assessment of the ability of the process to be clustered around the target. As C_{pm} is not traditionally used to provide insights into the number of parts non-conforming, it does not require 6σ to reflect a precise number of non-conforming. As a result, unlike other capability indices including C_p, C_{pu}, C_{pl}, and C_{pk}, the C_{pm} parameter is not distributionally sensitive.

Assuming that process capability assessments are studies of the ability of the process to produce product around the target, then C_{pm} will provide practitioners with an assessment of capability regardless of the distribution associated with the measurements. Clustering around the target, rather than a measure of non-conforming parts releases the physical meaning attached to 6σ. The denominator of C_{pm} then provides a measure of the clustering around the target and compares this with customer tolerance.

Eliminating the physical meaning allows C_{pm} to be used to compare the capability of various processes (or processes over time) regardless of the underlying distribution. The underlying distribution will impact the inferences that we can make from samples gathered from the population; however, the population parameter is no longer distributionally sensitive. The effects of non-normality on the stochastic estimator of C_{pm} and related properties have been examined in Leung (1999).

Loss Functions and Process Capability Indices—"PCIs based on Expected Loss"

The use of loss functions in quality assurance settings has grown with the introduction of Taguchi's philosophy. Theoretical statisticians and economists have for many years used the squared error loss function when making decisions or evaluating decision rules. With the increasing importance of clustering around the target, rather than conforming to specification limits, and the understanding of loss functions there appears to be an alternative to PCIs. Rather than numbers or percentage non-conforming, economic/production costs or losses may provide improved opportunities to assess, monitor, and compare process capability.

Johnson (1992) provided insights into the inferential properties of C_{pm}, outlining its relationship with expected relative loss for a process. English and Taylor (1993) investigated the loss imparted to society by examining expected loss for process measurements arising from non-normal populations. Gupta and Kotz (1997) related relative loss to a modified C_{pm} index they refer to as C_{pq}. However, for the most part there has been little research effort devoted to the area of loss and loss functions as methods for assessing process capability. This may be due in part to several criticisms of quadratic/squared error loss.

Criticism of the quadratic loss function includes that by statistical decision analysts (see Box and Tiao (1992) and Berger (1985)) and quality assurance practitioners and researchers (Leon and Wu (1992) and Tribus and Szonyi (1989)), for reasons that include its failure to provide a quantifiable maximum loss (i.e., unbounded loss) and because the magnitude of losses are much too severe for extreme deviations from the target. Pearn, Kotz, and Johnson (1992) also point out that the squared error loss function is almost always chosen because of the simplified mathematical derivations and not for its ability in depicting actual process losses.

Similar to Johnson's (1992) development for C_{pm}, C_{pw} can be expressed as a function of the expected squared deviation from the target. Defining $L(x)$ to be the loss associated with a characteristic X not produced at its target, loss can be depicted as a weighted squared error loss function

$$L(x) = w\,(x-T)^2\,,$$

where w is a non-stochastic weight function and T is the target value. This implies that the loss is zero when the process is on target and positive for any deviation from target. The expected loss becomes

$$\begin{aligned} E[L(X)] &= E\left[w\,(X-T)^2\right] \\ &= E\left[w\,(X-\mu+\mu-T)^2\right] \\ &= w\sigma^2 + w\,(\mu-T)^2 \\ &= (w-1)\sigma^2 + \sigma^2 + w\,(\mu-T)^2\,, \end{aligned}$$

allowing C_{pw} to be written in terms of $E\left[L\left(X\right)\right]$;

$$C_{pw} = \frac{\text{USL}-\text{LSL}}{6\sqrt{E\left[L\left(X\right)\right]}},$$

or, alternatively, as a function of C_{pw};

$$E\left[L(X)\right] = \frac{\left[\text{USL}-\text{LSL}\right]^2}{36C_{pw}^2}.$$

The link between $C_p(u,v)$ and the weighted quadratic loss function follows directly from the relationship between C_{pw} and $C_p(u,v)$ as described by Kotz and Johnson. Details of the properties, estimators, and inferences associated with the relationship between loss and C_{pw} can be found in Leung (1999).

The general PCI relationship with expected loss and the expanding research effort in the area of more applicable loss functions offers both practical and research opportunities for developing improved assessment, monitoring, and comparison methods in the area of process capability. Spiring (1993), Sun, Laramee, and Ramberg (1996), and Spiring and Yeung (1998) have developed a class of loss functions that provide practitioners with a wide range of loss functions that can be used in depicting loss due to departures from the process target. Spiring and Leung (2001) have studied the properties of this class of loss functions. Research efforts relating PCIs and loss would appear to offer opportunities that could potentially address practitioners', managers', and researchers' concerns and differences in the area of process capability.

Additional References

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis,* 2nd ed. Springer Verlag, New York.

BOX, G. E. P. and TIAO, G. C. (1992). *Bayesian Inference in Statistical Analysis*. John Wiley, New York, NY.

ENGLISH, J. R. and TAYLOR, G. D. (1993). "Process Capability Analysis—A Robustness Study". *International Journal of Production Research* 31, pp. 1621–1635.

LEUNG, B. P. K. (1999). "Contributions to Industrial Statistics", *Ph.D. Thesis,* University of Manitoba, Canada.

LEON, R. V. and WU, C. F. J. (1992). "A Theory of Performance Measures in Parameter Design". *Statistica Sinica* 2, pp. 335–357.

MCCORMARK, D. W. JR.; HARRIS, I. R.; HURWITZ, A. M.; and SPAGON, P. D. (2000). "Capability Indices for Non-Normal Data". *Quality Engineering* 12, pp. 489–495.

PAGE, M. (1994). "Analysis of Non-Normal Process Distributions". *Semiconductor International* 10, pp. 88–96.

SPIRING, F. A. (1991). "A New Measure of Process Capability". *Quality Progress* Feb., pp. 57–61.

SPIRING, F. A. (1993). "The Reflected Normal Loss Function". *The Canadian Journal of Statistics* 21, pp 321–330.

SPIRING, F. A. and YEUNG, A. (1998). "A General Class of Loss Functions with Industrial Applications". *Journal of Quality Technology* 30, pp. 152–162.

SPIRING F. A. and LEUNG, B. P. K. (2001). "Some Properties of the Inverted Probability Loss Functions" (to appear).

SUN, F.; LARAMEE, J. and RAMBERG, J. (1996). "On Spiring's Inverted Normal Loss Function". *The Canadian Journal of Statistics* 24, pp. 241–249.

Additional Bibliography

BOYLES, R. A. (1991). "Corrigenda". *Journal of Quality Technology* 23, pp. 382–383.

CHAN, L. K.; CHENG, S. W.; and SPIRING, F. A. (1988). "A Graphical Technique for Process Capability". *ASQC Quality Congress Transactions,* pp. 268–275.

CHENG, S. W. (1992). "Is the Process Capable? Tables and Graphs in Assessing C_{pm}". *Quality Engineering* 4, pp. 563–576.

CHOU, Y-M. and OWEN, D. B. (1989). "On the Distributions of the Estimated Process Capability Indices". *Communications in Statistics–Theory and Methods* 18, pp. 4549–4560.

CLIFFORD, P. C. (1971). "A Process Capability Study Using Control Charts". *Journal of Quality Technology* 3, pp. 107–111.

CONSTABLE, G. K. and HOBBS, J. R. (1992). "Small Samples and Non-Normal Capability". *American Society for Quality Control, Technical Conference, pp. 37–43.*

ENGLISH, J. R. and TAYLOR, G. D. (1993). "Process Capability Analysis—A Robustness Study". *International Journal of Production Research* 31, pp. 1621-1635.

HUBELE, N. F. (1999). "Using Graphical Hypothesis Testing for Assessing Process Capability With C_{pm}". *43rd Annual Fall Technical Conference.*

KANE, V. E. (1986). "Corrigenda". *Journal of Quality Technology* 18, p. 265.

KOCHERLAKOTA, S.; KOCHERLAKOTA, K.; and KIRMANI, S. N. U. A. (1992). "Process Capability Indices Under Nonnormality". *International Journal of Mathematical and Statistical Sciences* 1, pp. 175–210.

SHORE, H. (1998). "A New Approach to Analyzing Non-Normal Quality Data With Application to Process Capability Analysis". *International Journal of Production Research* 36, pp. 1917–1933.

SPIRING, F. A. (2000). "Assessing Process Capability with Indices" in *Statistical Process Monitoring and Optimization,* edited by S. H. Park and G. G. Vining, Marcel Dekker, New York, pp. 269–290.

Von Collani, E. (1994). A Note on Process Capability Indices". *Proceedings of the Third Umeå-Würzburg Conference on Statistics.*

Zabecki, D. T. (1986). "Process Percent Nonconforming Curves for Process Capability Ratio and Targeting Shift". *Quality Progress,* July, pp. 35–38.

Zhang, N. F.; Stenback, G. A.; and Wardrop, D. M. (1990). "Interval Estimation of Process Capability Index C_{pk}". *Communications in Statistics–Theory and Methods* 19, pp. 4455–4470.

Discussion

ROBERT N. RODRIGUEZ

SAS Institute Inc., Cary, NC 27513

I WOULD like to begin by thanking Professors Kotz and Johnson for providing a valuable update on developments in the area of process capability indices. In the tradition of their lengthy and distinguished collaboration, this survey is both comprehensive and clearly organized, and it provides researchers with a definitive starting point for further work.

While this paper focuses almost entirely on theoretical results, the authors do express concern about whether or not newer capability indices are finding their way into practice, and whether or not capability indices are being used more effectively in industry. I will comment on the "gap between theoreticians and practitioners," which I believe has widened since 1992.

Usage and Trends

Despite the large number of PCIs that have been proposed during the past decade, the indices C_p and C_{pk} are still the mainstay of process capability analysis in manufacturing environments. Deleryd (1998) and Kotz and Lovelace (1998) report the results of a survey of Swedish manufacturers involving 97 respondents who indicated that they used the indices C_p, C_{pk}, and (to a much lesser extent) C_{pm}, but not the newer indices C_{pmk} and $C_p(u, v)$. This is a universal pattern, as indicated by the following observations:

- Companies who use capability indices to measure process improvement or to compare the processes of vendors and internal suppliers continue to rely heavily on C_p and C_{pk}. Rarely are other indices mentioned in corporate or industry standards for statistical process control or quality management systems.
- C_p and C_{pk} are standard tools in Six Sigma training programs. Discussion of newer indices is missing even in the more comprehensive statistical references on Six Sigma practice, such as Breyfogle (1999).
- There are dozens of websites for quality management firms who offer consulting and training services which cover the basic indices C_p and C_{pk}. Web searches for information on newer indices yield almost no results, except for a few academic sites.

Since 1992, I have observed two encouraging trends in the application of basic PCIs. The first is an increased commitment to reporting confidence limits for C_p, C_{pl}, C_{pu}, C_{pk}, and C_{pm}, stemming from the recognition that point estimates are subject to variability and will change over time even when the process remains stable. This trend has benefited from computational support for confidence limits in modern statistical software for process control (see, for example, SAS Institute Inc. (1999)). The second trend is a greater awareness that checking for normality of the data is essential to the interpretability of PCIs and the validity of confidence limits. Again, software has helped by providing high-quality goodness-of-fit tests for normality, graphical displays, and other diagnostic tools (see Rodriguez (1992)).

At the same time, process capability applications have been plagued by three forms of conceptual confusion, which often prevail over good statistical practice.

Confusion Between Process Capability Indices and Process Performance Indices

In a number of companies, sound statistical thinking has been undermined by the practice of reserving the term "process capability index" and the notation C_p and C_{pk} for indices that are computed using a within-subgroup plug-in estimate for the process standard deviation σ (see the section on "Notation and the Basic PCIs"). Typically, this estimate is computed as

$$\bar{R}/d_2, \tag{1}$$

Dr. Rodriguez is Director of Linear Models and Statistical Quality Improvement R&D. He is a Member of ASQ. His email address is bob.rodriguez@sas.com.

where $\bar{R}$ is the average range for a set of consecutive subgroup samples and d_2 is the control chart constant used to determine control limits for a range chart. The estimate in Equation (1) is then denoted as $\widehat{\sigma}_{\text{within}}$ or as $\widehat{\sigma}_{\text{st}}$ to indicate that it measures within-subgroup or so-called "short-term" variation. When the indices C_p and C_{pk} are computed using a plug-in estimate for σ based on individual measurements, such as

$$\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad (2)$$

they are denoted as P_p and P_{pk}, and they are designated as "process performance indices." The estimate in Equation (2) (divided by the control chart constant c_4) is denoted as $\widehat{\sigma}_{\text{lt}}$ to indicate that it measures "long-term" variation (see, for example, Bothe (1997)).

Theoretically, this dichotomy is artificial, to say the least, because both estimates (1) and (2) are valid estimates of σ, provided that the process is in control (and assuming that σ^2 is the only component of variation in the process). Furthermore, estimate (2) is statistically more efficient than estimate (1), which is why it was originally adopted as the plug-in estimate for the basic PCIs, and why it has been the estimate of choice in the vast majority of the literature surveyed by this paper. Conversely, if the process is not in control, then neither estimate (1) nor (2) is a valid estimate of σ since the notion of a process distribution is meaningless.

In practice, the choice of estimate for the in-control situation is not entirely clear-cut because an in-control process is never perfectly stable. Nelson (1999) illuminates this in a *JQT* "Technical Aids" article which should be read by anyone who is perplexed by this issue. He explains that "a control chart will reveal instability only when it is beyond a threshold level," and consequently estimate (2) will appropriately "take account of variations from subgroup to subgroup that the control chart does not signal." Because estimate (1) does not accomplish this, Nelson (1999) concludes that it should not be used to compute capability indices (of course, estimates of within-subgroup variation such as (1) are still recommended for computing control chart limits).

The dichotomy between capability indices and performance indices may have started out as a well-intentioned attempt to force PCI users to check for stability with a control chart, but in many situations it has had the opposite effect. One unfortunate consequence of introducing dual notation is that P_p and P_{pk} are seemingly legitimized, and so they tend to be accepted as valid process measures for the out-of-control situation. For instance, one misconception I have encountered is the recommendation that "P_{pk} is a poorer estimate of process capability, so you need a P_{pk} of 1.67 but only 1.33 for C_{pk}."

Another consequence is that there is much less analysis of individual measurements from in-control processes using tools which support and complement the use of PCIs. These methods include goodness-of-fit tests, distributional models, robust estimates of location and scale, nonparametric density estimates, comparative histograms, comparative boxplots, and other graphical displays (see Rodriguez (1992) for further discussion).

In many situations, users must choose blindly between C_{pk} and P_{pk} in output from badly designed software that always displays both quantities without offering qualification or guidance. Even in in-control situations, users of capability indices based on Equation (1) are often unaware that this estimate, and hence C_p and C_{pk}, will depend on how the data are subgrouped. Typically, this type of software selects the subgroups automatically according to a default rule (such as "take groups of five") and does not provide a way for the user to select the subgroups. Consequently, the problem of subgrouping, which is one of the most poorly understood aspects of control chart analysis, now lurks in many capability analysis applications.

Until this dichotomy is eliminated or resolved, it will remain a barrier to good statistical practice based on process capability analysis of individual measurements.

Confusion Over How to Assess Process Capability When the Data Are Not Normally Distributed

There have been many proposals for solving this problem, as summarized in the section of this paper titled "Dealing with Non-normality of the Distribution of X."

In practice, none of these methods have succeeded in providing a broadly accepted general-purpose capability index. The main reason for this is that the root problem involves modeling the process distribution, or at least capturing its tail behavior in a precise fashion, and this is simply not feasible with a single statistical measure.

There is a variety of commercial software which attempts to fit the data distribution with a broad system of parametric distributions (such as the Pearson system or the Johnson system) and then computes generalized indices as in Equations (20) and (21) from Kotz and Johnson's review. The statistical reliability of such indices should not be taken for granted because such indices can vary greatly depending on the system of distributions, the estimation technique, and the presence of outliers. These issues are masked by black-box software implementations, adding to the confusion.

This approach has a much greater potential for success in situations where the process distribution is modeled carefully, and where a relatively narrow *family* of distributions is found to provide an adequate model for the process distribution. For example, I have noticed that measurements in geometric tolerancing problems (see Minnick (2001)) can often be modeled with a lognormal distribution. This simplifies the estimation problem, and it allows the use of goodness-of-fit tests to validate the model. Alternatively, one can apply a log transformation to the data to achieve normality. Even so, it may not be clear how to interpret and compare the resulting PCIs (see the next section of these comments). Thus the real benefit of this type of analysis lies in using the model to predict a variety of quantities, such as percentiles and probabilities of nonconformance, that are much easier to interpret.

The difficulties of distributional modeling are circumvented by a number of PCIs which attempt to adjust for the skewness of the distribution. Among these PCIs, which are surveyed in "Dealing with Non-normality of the Distribution of X," several are quite straightforward to compute, including the index C_s due to Wright (1995), the modification of C_s due to Chen and Kotz (1996), and the index C_{jkp} due to Johnson et al. (1994). However, clear guidelines for using these proposals are scarce in the research literature, and there have not been any follow-up papers which compare their advantages in specific applications.

Confusion About the Practical and Statistical Interpretation of PCIs

One reason for the enduring popularity of C_p and C_{pk} is that these basic indices are easy for busy managers to understand. In practice, they are accompanied by guidelines such as

> "$C_p > 1.67$ means the process is highly capable,"

and

> "C_{pk} between 1 and 1.33 indicates the process is barely capable."

While there is nothing magical about these numbers from the standpoint of statistical theory, they have been established through considerable shared context and experience, and they facilitate communication.

As far as I know, similar guidelines have not been developed for any of the newer PCIs. To illustrate the confusion that might arise in practice, consider the following options available to an enlightened user whose measurements are lognormally distributed:

(a) compute C_{pk} from the log transformation of the data and the specifications,
(b) compute C'_{pk} as in Equation (21) from Kotz and Johnson using a fitted lognormal distribution for the data,
(c) compute Wright's C_s as in Equation (23) using an estimate of skewness.

Can this user apply the cutoff values of 1.33 and 1.67 to (a) and (b)? What would be an "acceptable" value for (c)? Can any of these indices be compared on the same scale when computed for two or more samples from the same process but with different lognormal distributions?

Aids to the *practical* interpretation of PCIs, including baselines, scaling, and calibration, should not be overlooked by researchers who propose new indices, and they should be considered carefully in selecting PCIs for implementation in process control applications. In fact, these aids are essential to *any* index that is intended for monitoring a complex process (it is revealing to see how guidelines are provided and used for economic indices and stock market indices).

By comparison, the *statistical* interpretation of newer PCIs has received a great deal of attention in the literature, and it is a focal point of ongoing debate concerning the value of PCIs and the need for alternative methods (see, for example, Post (2000)). The performance of PCIs, including C_{pm}, C_{pmk}, and others based on the concept of the "Taguchi" loss function, continues to be evaluated using the expected proportion nonconforming as the criterion. Ironically, this reinforces the view, held by many, that the expected proportion NC is still the most natural—if not the most feasible—way to measure process capability.

Additional References

BREYFOGLE, F. W., III (1999). *Implementing Six Sigma: Smarter Solutions Using Statistical Methods*. John Wiley and Sons, New York, Ny.

MINNICK, T. (2001). "C_{pu} Calculations for True Position with MMC Modifier". *2001 Quality and Productivity Research Conference,* Austin, Texas.

POST, R. (2000). "An Approach to a Simple Statistically Valid Alternative to Process Capability Indices". *44th Annual Fall Technical Conference,* American Society for Quality and American Statistical Association.

SAS INSTITUTE INC. (1999). *SAS/QC User's Guide, Version 8*. SAS Institute Inc., Cary, NC.

Discussion

DAVIS R. BOTHE

International Quality Institute, Cedarburg, WI 53012-2910

I WOULD like to thank the authors for sharing their extensive compilation of articles and books related to capability indices. Being greatly interested in this subject for many years, I thought I was aware of just about everything written on the topic, but I am familiar with only about 70 percent of the works listed in their References and Bibliography section.

Having worked with capability indices in industry for the past 25 years, I approach this topic from more of an applied, rather than theoretical, perspective. I thank the Editor for this opportunity to present a few of my thoughts on this subject.

One Size Does Not Fit All

I wholeheartedly agree with Nelson's (1992) contention that process capability cannot be adequately characterized by a single number. Based on my experience, C_{pk} is by far the most popular capability index in use today and is the only one reported in many companies. Sadly, no one index, including C_{pk}, is appropriate for all processes.

For example, suppose that a part feature has a specification of 65, plus 5, minus 3. With a lower specification limit (LSL) of 62 and an upper specification limit (USL) of 70, the midpoint of the tolerance, M, is equal to 66. Assuming a normal distribution for the process output, the C_{pk} index will achieve its highest value when the mean, μ, is located at 66. Therefore, in order to maximize the reported C_{pk} for this process, shop floor personnel will strive to center the output at 66. Unfortunately, optimal product performance occurs when μ is positioned at T, the target average of 65.

Because a capability index should reflect customer satisfaction, C_{pk} is inappropriate for product features with asymmetric tolerances, i.e., when $T \neq M$. An index, like C^*_{pk} (Bothe (2001)), that achieves its highest value when μ equals T should be used instead.

Mr. Bothe is the Director of Quality Improvement. He is a Fellow of ASQ. His e-mail address is drbothe@sprynet.com.

C_{pk} Can Be Misleading

Companies often rank processes by their C_{pk} indices to sort those with high quality from those with low quality. Even when all the usual assumptions are met, however, a higher C_{pk} doesn't necessarily mean a higher level of quality for the customer.

In the top portion of Figure 1, process A has 4 percent of its output above the USL and 0 percent below the LSL. With a total of 4 percent nonconforming parts, there are 96 percent conforming parts. Process B has 3 percent above the USL and another 3 percent below the LSL. With a total of 6 percent nonconforming, B is producing only 94 percent conforming parts. So which is the better process given their current performances?

When the C_{pk} indices are calculated for each with the authors' Equation (2), A ($\mu_A = 3$, $\sigma_A = 1.143$) has a rating of 0.58 while B ($\mu_B = 0$, $\sigma_B = 2.660$) has a rating of 0.63. This method of assessing process capability would suggest that B has better quality than A, even though A has a higher percentage of conforming parts, 96 versus 94. In addition, because shifting the average of a process usually requires less effort than reducing its variation, increasing the percentage of conforming parts will probably be easier for process A than for B. Yet, the C_{pk} index rates B as the better process.

In this example, both processes have low C_{pk} values, namely 0.58 and 0.63. Knowing that capability is lacking, how should we improve process performance? Given only that the C_{pk} index is low, we don't know whether we should shift the average, μ, or reduce the standard deviation, σ. For process A, we need to shift μ so it is centered at M. For process B, efforts must concentrate on reducing σ.

Problems with Using Just p

As noted in the review, several authors recommend reporting just the percentage of nonconforming product, p, as an indication of capability. Although p provides a simple and concise summary of process

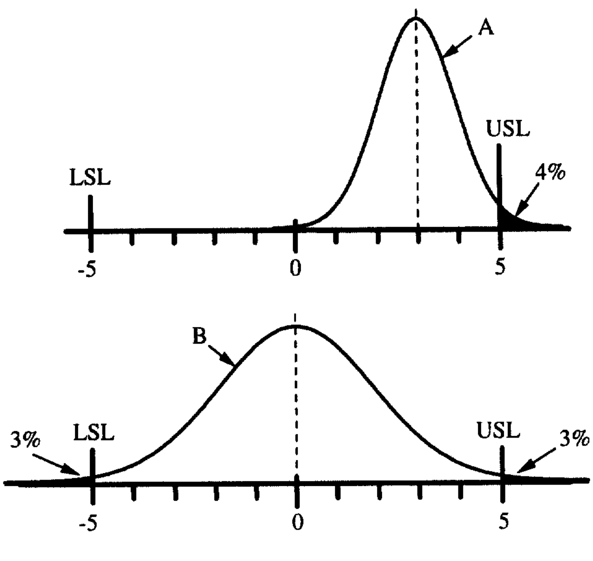

FIGURE 1. Two processes with different nonconformance rates.

quality that is easy for customers and top-level managers to understand, it is of limited use for practitioners who are working to improve the process.

Given that p is 8 percent for a process, how would one improve quality? If 4 percent of the nonconforming parts are below the LSL and the other 4 percent are above the USL, the process standard deviation must be reduced. If all 8 percent of the nonconforming parts are below the LSL, the process average must be raised. If the nonconforming parts are all above the USL, then efforts must be made to lower μ. If the split is 6 percent below the LSL and 2 percent above the USL, μ should be shifted slightly higher and σ reduced. Based on just p, which course of action should one pursue to improve process performance? The capability index p provides few answers to these vital questions.

In addition to the above issues, suppose a part below the LSL must be scrapped at a cost of $5 while a part above the USL can be reworked for $1. Now it becomes very important to know what percentage is below the LSL and what is above the USL. With a process lacking in capability, we would temporarily shift μ a little higher than M until efforts to reduce σ are successful. Doing so will increase the percentage of rework but lower the percentage of the costlier scrap, thus minimizing operating costs.

Strength in Numbers

Peter Drucker (1972) admonished managers to "never look at any one measure alone in any business; look at multiple measures." I believe the same advice holds true for capability indices. An engineering drawing for a complex part displays many different views (front, top, side) to help visualize what the finished part should look like. In a similar manner, several capability measures are required to fully describe the ability of the process to manufacture such a part.

Suppose a critical product characteristic has a bilateral specification and its output is stable and close to following a normal distribution. If the indices C_p, C_{pk}, p_{LSL} (percentage nonconforming below the LSL), and p_{USL} (percentage nonconforming above the USL) are reported, we will have a very good idea of what is happening regarding the process output and what actions are necessary to improve it.

Obviously, the desired state is to have both C_p and C_{pk} fairly "large," as this means p_{LSL} and p_{USL} will be "small."

Whenever C_p is "large" and C_{pk} is "small," then μ is not centered at the middle of the tolerance. If p_{LSL} is less than p_{USL}, the practitioner knows μ should be shifted lower. Conversely, when p_{LSL} is greater than p_{USL}, μ should be moved higher.

In situations where both C_p and C_{pk} are "small," μ is centered near the middle of the tolerance but the process spread is too wide. With both p_{LSL} and p_{USL} being "large," improvement efforts must focus on reducing σ.

By supplying a more detailed understanding of a process's current capability, this group of four indices would help shop floor personnel make better decisions on how to improve its future performance. In addition, managers and customers can still be given p ($p = p_{\text{LSL}} + p_{\text{USL}}$) as a measure of overall process yield.

Since there is no one "perfect" measure to encapsulate every important facet of a process's output, I believe it is better to rely on a family of slightly imperfect ones. There is indeed strength in numbers; what one index lacks, another can furnish.

Universal Capability Goal

Most large companies mandate an identical capability goal for every characteristic of every product,

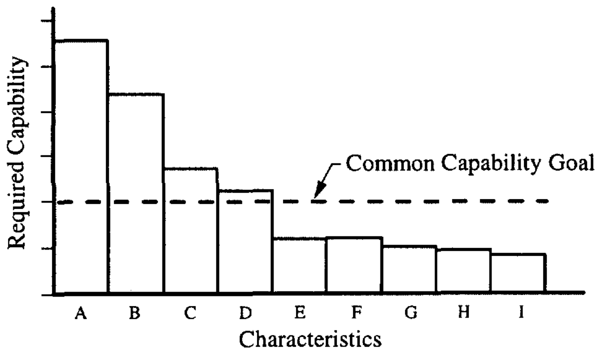

FIGURE 2. A ranking of required capability levels.

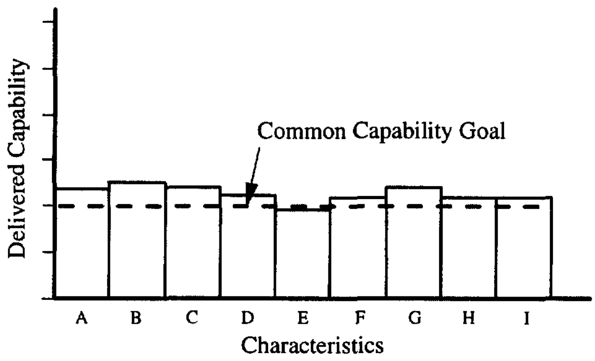

FIGURE 3. Capability delievered by working to a common goal.

including those supplied from completely different industries. This is frequently done because these companies don't wish to invest the necessary time scrutinizing each product to determine its critical characteristics and their true capability needs. However, rarely are all characteristics of equal importance to the customer and, therefore, would not have the same capability requirements. A ranking of the capability levels needed for the characteristics of a hypothetical product is presented in Figure 2.

Although a simple method, the common-goal policy is very wasteful because it leads to an improper allocation of resources. Time and money are spent improving all product characteristics whose current capability falls short of the universal goal. Those that don't require a capability as high as the common goal are worked on with the same intensity as those requiring a capability higher than this goal. Efforts expended on improving features whose true capability needs are less than the common one would be better invested in those whose capability needs exceed the general goal.

Making matters worse, those features truly needing higher capability are improved only until they reach the common goal. At this point, improvement halts as attention and resources are redirected to features whose capability is currently less than the universal goal. The net result of this inefficient goal-setting policy is a product with every feature having about the same level of capability, as is depicted in Figure 3.

Influence of Corporate Cultures

The authors note that the index C_{pm} does not directly relate to the percentage of nonconforming product, p. This is true, but if p is regarded as the most important quality aspect of the process, this is definitely the *wrong* capability index to use. Note that C_{pm} is based on the Taguchi philosophy of minimizing process variation around the target average, which, for followers of this philosophy, is more important than the amount of nonconforming product. If p is of paramount importance, then C_p and C_{pk} should be used to assess a process since they are more closely associated with the percentage of nonconforming products.

For example, suppose that the output of a process initially looks like distribution A shown in Figure 4, with an average of 12 and a standard deviation of 2. Given these process parameters, the C_{pk} index is 1.00, as is the C_{pm} index:

$$\begin{aligned} C_{pk} &= \min\left\{\frac{\mu - \text{LSL}}{3\sigma}, \frac{\text{USL} - \mu}{3\sigma}\right\} \\ &= \min\left\{\frac{12-6}{3(2)}, \frac{18-12}{3(2)}\right\} \\ &= 1.00; \end{aligned}$$

$$\begin{aligned} C_{pm} &= \frac{\text{USL} - \text{LSL}}{6\sqrt{\sigma^2 + (\mu - T)^2}} \\ &= \frac{18 - 6}{6\sqrt{(2)^2 + (12-12)^2}} \\ &= 1.00. \end{aligned}$$

After a modification is made to this process, the output changes to distribution B, with an average of 15 and a standard deviation of .667. There are now fewer nonconforming parts, but also fewer parts produced at the target average. Did this change improve process capability? The C_{pk} index certainly indicates so, jumping from 1.00 to 1.50. However, the C_{pm} index indicates otherwise, reporting a seri-

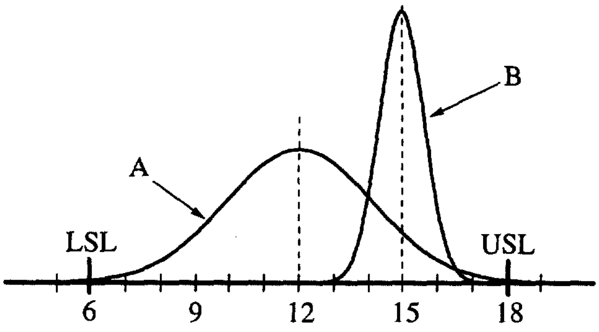

FIGURE 4. A change in process output from A to B.

ous deterioration in quality by dropping from 1.00 to only 0.65:

$$C_{pk} = \min\left\{\frac{15-6}{3(.667)}, \frac{18-15}{3(.667)}\right\} = 1.50;$$

$$C_{pm} = \frac{18-6}{6\sqrt{(.667)^2 + (15-12)^2}} = 0.65.$$

Was the change helpful or harmful to the quality level of this process? The answer depends on a particular corporation's quality philosophy. Those companies concerned mainly with making parts to print and reducing p would claim that B's performance is better than A's. Those whose primary concern is making every part on target will believe that B is worse than A.

In Defense of the C_{pmk} Index

The authors correctly note that C_{pmk} is certainly worse than C_{pk} for being associated with a certain percentage of nonconforming product, but again, one should not choose this index if p is the main interest. C_{pmk} (and usually C_{pm}) is much more sensitive than other capability indices to movements in the process average relative to M. As seen in Figure 5, when μ is equal to M, C_{pmk} is equal to C_{pk}. If μ moves away from M, however, C_{pmk} decreases more rapidly than does C_{pk} (although both are 0 when μ equals one of the specification limits). Conversely, when μ is brought closer to M, C_{pmk} increases much faster than does C_{pk}.

In addition to the above advantage, C_{pmk} reveals the most information about the location of the process average. Given a C_{pk} index of 1.0, all a practitioner can say about μ is that it is somewhere between the LSL and the USL, i.e., $M-d < \mu < M+d$, where d equals (USL − LSL)/2.

With the C_{pm} index, it can be shown (Bothe

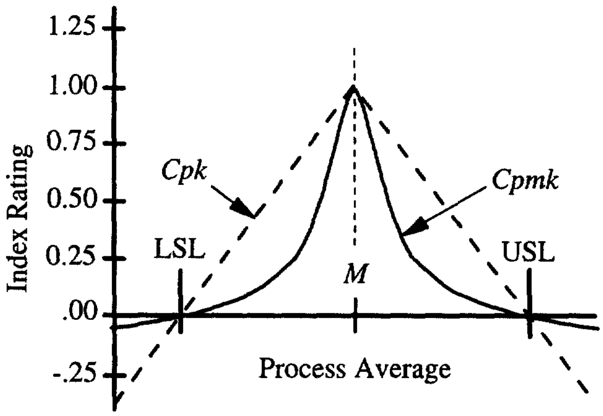

FIGURE 5. Changes in the C_{pk} and C_{pmk} versus changes in μ.

(2001)) that the distance between μ and M must be less than $d(3C_{pm})^{-1}$. Therefore, given a C_{pm} index of 1.0, we know that $M - d/3 < \mu < M + d/3$. This is a much smaller interval than the one for C_{pk} equal to 1.0.

For the C_{pmk} index, it can be shown that the distance between μ and M is less than $d(3C_{pmk}+1)^{-1}$. For a C_{pmk} index of 1.0, one knows that $M - d/4 < \mu < M + d/4$, which is even a smaller interval than the one for C_{pm}.

Although C_{pk} offers the most information about p, it provides the least insight about the location of μ. On the other hand, C_{pmk} provides the most information about the location of μ and the least about p. Again, a corporation's quality philosophy will dictate which index is more appropriate for a given situation.

Non-Normal Capability Indices

Equation (21) of the authors' review applies best to symmetrical non-normal distributions because it assumes the process median should be centered at M, the midpoint of the tolerance. In addition, this formula does not agree with the one suggested by the ISO Technical Committee 69 on applications for statistical methods when estimating capability for a process whose output has a non-normal distribution (document N13 by Working Group 6 of Subcommittee 4). The "ISO" formula is stated as follows, where $x_{a\times100}$ is equivalent to the authors' symbol ξ_a:

$$C_{pk} = \min\left\{\frac{x_{50} - \text{LSL}}{x_{50} - x_{.135}}, \frac{\text{USL} - x_{50}}{x_{99.865} - x_{50}}\right\}.$$

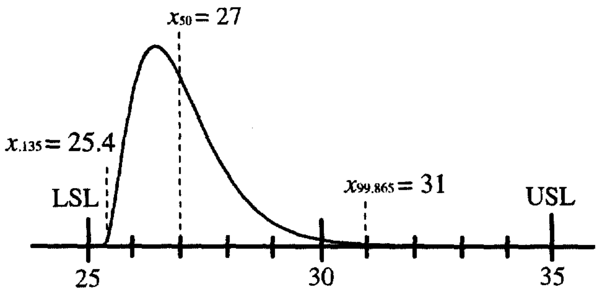

FIGURE 6. A non-normally distributed process output.

There is a substantial difference between the results generated by these two formulas. For the output distribution displayed in Figure 6, $\xi_{0.00135} = x_{.135} = 25.4$, $\xi_{0.50} = x_{50} = 27$, and $\xi_{0.99865} = x_{99.865} = 31$. With a LSL of 25 and an USL of 35, meaning that M equals 30, the index calculated with the authors' formula would be 0.71 since we have

$$\begin{aligned} C'_{pk} &= \frac{.5(\text{USL} - \text{LSL}) - \mid \xi_{0.50} - M \mid}{.5(\xi_{0.99865} - \xi_{0.00135})} \\ &= \frac{.5(35 - 25) - \mid 27 - 30 \mid}{.5(31 - 25.4)} \\ &= 0.71. \end{aligned}$$

A conventional C_{pk} index that is less than 1.00 implies the process is producing more than .135 percent of its output beyond at least one of the specification limits. In this case, the .135 percentile point of 25.4 is above the LSL of 25 (meaning that less than .135 percent of the output is below the LSL) and the 99.865 percentile of 31 is well below the USL of 35, meaning less than .135 percent is above the USL.

With the "ISO" formula, the capability index for this process is 1.25. A C_{pk} index greater than 1.00 implies less than .135 percent of the process output is outside either specification limit, which is indeed true for this process. The detailed calculation is

$$\begin{aligned} C_{pk} &= \min\left\{\frac{27 - 25}{27 - 25.4}, \frac{35 - 27}{31 - 27}\right\} \\ &= \min(1.25, 2.00) \\ &= 1.25. \end{aligned}$$

The "ISO" formulation of C_{pk} also reveals the direction in which the median should be moved to increase process capability. Because 1.25 is less than 2.00 in the equation above, moving the median higher will improve quality and increase this index.

Notational Inconsistencies

Traditionally, practitioners estimate C_p and C_{pk} with $\widehat{\sigma}_{\overline{R}}$ (AIAG (1995)). For example,

$$\widehat{C}_p = \frac{\text{USL} - \text{LSL}}{6\widehat{\sigma}_{\overline{R}}} \quad \text{where} \quad \widehat{\sigma}_{\overline{R}} = \frac{\overline{R}}{d_2}.$$

Because of its statistical properties, or lack thereof, statisticians don't care much for $\widehat{\sigma}_{\overline{R}}$, preferring to use $\widehat{\sigma}_s$ when estimating C_p and C_{pk} (Kotz and Lovelace (1998)):

$$\widehat{C}_p = \frac{\text{USL} - \text{LSL}}{6\widehat{\sigma}_s} \quad \text{where} \quad \widehat{\sigma}_s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}.$$

Unfortunately, many practitioners, especially those in the automotive industry, use $\widehat{\sigma}_s$ to estimate what is called the P_p and P_{pk} indices (AIAG (1995)). Most SPC software programs currently on the market also utilize this notation:

$$\widehat{P}_p = \frac{\text{USL} - \text{LSL}}{6\widehat{\sigma}_s}.$$

To lessen confusion on the part of practitioners, there is a definite need for consensus on a standardized notation. Because the vast majority of practitioners are already using the "AIAG" notation, it would be easier for statisticians to adopt it as well.

Areas of Concern

Far too many practitioners estimate capability without verifying process stability, and don't take the time to determine the shape of the process distribution. Even if these checks are performed, very few compute a confidence interval for the capability index. In addition, indices need to be created for reliably measuring the capability of processes having: inherent tool wear; variation in setup between runs; limited data due to short production runs; autocorrelation; and features with geometric dimensioning and tolerancing. These are situations daily faced by practitioners.

Capability indices are very powerful, but, like many powerful tools, can inflict heavy damage if used incorrectly. Properly calculated, they provide a wealth of vital information concerning how the current output of a process satisfies customer requirements. Incorrectly applied and/or interpreted, these indices can generate an abundance of misinformation that will confuse practitioners, waste resources, and lead to incorrect decision making.

Additional References

AIAG (1995). *Statistical Process Control Reference Manual.* Automotive Industry Action Group, Southfield, MI. p. 80.

BOTHE, D. R. (2001). *Measuring Process Capability.* Landmark Publishing, Cedarburg, WI.

DRUCKER, P. F. (1972). *Technology, Management, and Society.* Harper and Row, New York, NY.

KOTZ, S. and LOVELACE, C. R. (1998). *Process Capability Indices in Theory and Practice.* Arnold, London. p. 16.

Discussion

MING-WEI LU and RICHARD J. RUDY

DaimlerChrysler Corporation, Auburn Hills, MI 48326

THERE are many process capability indices available for use in industry. However, our comments are directed toward the use of PCIs in the automotive industry. The Automotive Industry Action Group (AIAG) has established P_p/P_{pk} and C_p/C_{pk} as the four basic PCIs used in the automotive industry. The definitions of P_p/P_{pk} (C_p/C_{pk}) are given below:

$P_p = (\text{USL} - \text{LSL})/6s$

$P_{pk} = \text{minimum of } (\text{USL} - \bar{X})/3s,\ (\bar{X} - \text{LSL})/3s$

$C_p = (\text{USL} - \text{LSL})/6s'$

$$C_{pk} = \min\left\{ \left(\text{USL} - \bar{\bar{X}}\right)/3s', \left(\bar{\bar{X}} - \text{LSL}\right)/3s' \right\},$$

where

USL = the upper specification limit,

LSL = the lower specification limit,

s = the sample standard deviation,

$\bar{X}$ = the sample average, taken from the average of all data in the sample,

s' = the process standard deviation, taken from $\bar{R}/d_2$, where $\bar{R}$ is the average range of a series of subgroups of constant size and d_2 is a divisor of $\bar{R}$ used to estimate the process standard deviation, and

$\bar{\bar{X}}$ = the process average, taken from the average of subgroups.

In this case, P_p and P_{pk} are used prior to production and help to provide a preliminary indication of process potential and capability; P_p/P_{pk} can be computed from a small sample without any assumption regarding process stability. It is also recognized that not all the variation is included in the calculation since it is only a "snapshot" in time. Thus, the required P_p and P_{pk} values are more stringent than C_p/C_{pk}. We use C_p/C_{pk} during volume production, and their use requires that the process be statistically stable. A larger sample size collected over a longer period is also required. Typically, a minimum of 25 subgroups or 125 individuals is required to determine process stability.

One major issue with all capability indices is the assumption that the underlying process distribution is unimodal and approximately normal, yet, in many industrial situations, the normal distribution does not provide an adequate approximate model. Examples of non-normal distributed quality characteristics include flatness, roundness, and diameter. Lack of normality may provide a misleading interpretation of the result. For example, if a population distribution is uniformly distributed over the interval from 1 to 2 with USL = 2, LSL = 1, and nominal = 1.5, then the mean is $\mu = 1.5$ and $\sigma = 0.29$. Hence, $C_{pk} = (2 - 1.5)/[3(0.29)] = 0.57$. That is, we have good parts (all in the spec limits) with a bad C_{pk} value. An important issue facing our industry today with respect to PCIs is how to deal with such non-normal data. The following steps are recommended in our company for calculating P_p/P_{pk} (C_p/C_{pk}).

Step 1 – Is data normal?

One can use a normal probability plot or a goodness-of-fit test to check whether the data are normally distributed. If the data are normal, then compute the "Percent out of spec" (i.e., proportion NC) and P_p/P_{pk} (C_p/C_{pk}) values. If data are not normal, go to Step 2.

Step 2 – Can data be transformed to be normal?

One can usually transform the original data in such a way that the transformed data will meet the normal distribution assumption. The Box-Cox transformation can be used to transform the original non-normal data to normal data. If the data can be transformed to be normal, then compute the "Percent out of spec" and P_p/P_{pk} (C_p/C_{pk}) values by using the transformed data. If not, proceed to Step 3.

Dr. Lu is Senior Quality and Reliability Specialist, Product and Process Integrity. His e-mail address is MWL6@Daimler Chrysler.Com.

Mr. Rudy is Senior Manager, Product and Process Integrity. He is a Senior Member of ASQ. His e-mail address is RJR11@DaimlerChrysler.Com.

Step 3 – Find the best-fit distribution

If the original data can not be transformed to normal data, we can fit the actual original sample data with the appropriate non-normal distribution. If we find the best-fit distribution, then we report the expected "Percent out of spec" and the "equivalent" P_{pk} (C_{pk}) values.

Another issue industry faces is the rigid mentality of achieving PCI requirements at the expense of rational decision-making. The dimensional, material, functional, and appearance quality acceptance requirements of all parts are set at $P_{pk} \geq 1.67$ (based on 30 samples) and long term $C_{pk} \geq 1.33$. A single capability index, however, can not replace a detailed functional review of the part. For example, consider non-rigid parts, such as sheet metal and trim panels, which, due to noise variables such as gage error, mechanical property variables, flimsiness of sheet metal, and assembly process, have great difficulty in satisfying the $P_{pk} \geq 1.67$ requirement at all checking points. Because the parts conform to the shape of their rigid mating parts, it may not be important to meet rigid PCI requirements. Suppliers request relief for certain checkpoints, but payment for the parts is based on 100% compliance to PCI requirements. If the requirements are unrealistic and require excessive cost to achieve with little or no effect on the customer, then a more rational approach should be taken. An engineering analysis should be conducted on the impact to the downstream processes and, ultimately, the build objective and impact on the customer.

Finally, the limitations of an index that tries to simultaneously describe the location of the mean and the size of variation within a single number are readily apparent. Our challenge to academia is to develop a simple-to-use method to accurately predict process capability. Reporting the expected "Percent out of spec" may be a more practical way for product quality to be assessed. Whatever method is developed must be computationally simple and easy for an engineer to implement.

Discussion

KERSTIN VÄNNMAN

Luleå University of Technology, SE-971 87 Luleå, Sweden

I WANT to thank the authors for writing this eagerly awaited review paper on capability indices. It gives a broad overview of this area, many interesting perspectives on process capability indices, and shows where the research stands today. It contains a large number of recent references to both theoretical and more practically oriented papers on capability indices, which makes this paper very valuable. The authors have done an excellent job in reviewing this area. I am also grateful to the Editor, who has given me this opportunity to participate in the discussion.

Let me first comment on the issue mentioned by the authors regarding the gap between theory and practice. Software, of course, can be helpful in some respect in closing this gap, but much more is needed to bridge the gap with regard to the proper use and interpretation of capability indices. Furthermore, capability indices ought to be considered in their context, as a part of what Kotz and Lovelace (1998, chapter 8) call a capability analysis process. This capability analysis process starts with a focus on process definition and ends with documented results. In between there are several important steps, where one is to estimate process capability. See also Deleryd (1997), who describes capability studies as part of a strategy to master variation and suggests a procedure with similarities to Deming's PDSA-cycle for improvements. My feeling is that more research needs to be done in order to better understand why the *whole* capability analysis process is difficult to implement in practice. This is an area where practitioners, theoretical statisticians, and social scientists could cooperate to contribute in producing new knowledge. Contributions have been made in this area by Deleryd (1998b). He surveyed Swedish industries to try to identify and quantify the use and misuse of process capability studies, but much more needs to be done. I share the opinion of the authors that "process capability indices are here to stay;" however, they need to be used properly, and both theoreticians and practitioners have a responsibility to cooperate and put both time and effort into this area of capability studies.

As the authors mention, process capability indices are intented to provide a *single-number* assessment of ability to meet specification limits. When, furthermore, target values are used, capability indices combine information about closeness to target and process spread and express the capability with a single number. For management, it may be convenient to have a single-number summary; however, in some cases it may also be held as one of the drawbacks of the indices. If, for instance, the process is found non-capable, then the engineer is interested in knowing whether this non-capability is caused by the fact that the process output is off target, that the process spread is too large, or that the result is a combination of these two factors. This information cannot instantly be found from one index value. To circumvent this drawback, I suggest that a process capability plot be used as a complement to the capability index. Process capability plots are described in Deleryd and Vännman (1999) and in Vännman (1998b). The second reference is now published in Vännman (2000).

A process capability plot is principally a plot of spread against deviation from target, and in this plot the capability index is shown by using a contour line. The plot also contains either a confidence region or a critical value from a hypothesis test in order to assess the capability of the process. Using such process capability plots, it is possible, with a single plot, to instantly get visual information about the location and spread of the quality characteristic as well as information about the capability of the process. The authors mention process capability plots, but in a context where they do not really belong. Process capability plots are not primarily intended to be used to derive estimates of p from values of C_p and C_{pk}. Instead, these plots are simple graphical tools, used in combination with capability indices to make decisions about the capability of the process and at the

Dr. Vännman is a Professor in the Division of Quality Technology & Statistics. Her e-mail address is kerstin.vannman@ies.luth.se.

same time to give deeper insight into the cause of a non-capable process. Furthermore, the above mentioned process capability plots are invariant to the values of the specification limits. Hence, in the same plot, several characteristics of a process can be monitored and information on the location and spread of the process can be presented. In this way, more information is obtained on how to improve the process compared to the use of traditional capability indices alone.

In general, I think that there is a need for simple graphical tools to bridge some of the gap between practice and theory in capability studies. It is also well known that the visual impact of a plot is more effective than numbers, such as estimates or confidence limits. I believe that much more can be done with regard to innovative simple graphical tools to be used in capability studies.

In the beginning of the development of capability indices, the probability of non-conformance or expected proportion of non-conforming items was considered of great importance. According to today's modern quality improvement theories, the focus should not be on the probability of non-conformance only. It is also very important to use target values and to keep the process on target. These theories are based, among other things, on Taguchi's quality philosophy, in which reduction of variation from the target value is the guiding principle. According to this philosophy, the specification limits, in connection with manufacturing, must not be interpreted as permission to be anywhere within the tolerance interval. Instead, target values should be used and attention should focus on meeting the target instead of meeting the tolerances. See, for example, Bergman and Klefsjö (1994). Hence, if μ is far away from the target T, then the process should not be considered capable even if σ is so small that the probability of non-conformance is small. Examples illustrating these ideas are found in Sullivan (1984). Taking such ideas into account, it seems that more emphasis should be given to research on the question of what is an "acceptable deviation from target" combined with an "acceptably small spread" when discussing process capability. This can, of course, be related to the probability of non-conformance, but cost, not only with regard to spread but also with regard to deviation from target, has to be taken into account. This is also a research area which needs close cooperation between practitioners and theoreticians to produce fruitful results. Use of the simple graphical tool suggested above may help to focus on the two aspects, deviation from target and spread, and combine them in convenient ways.

I would like to stress the conclusion in the last section, where the authors say that "it is necessary to distinguish between the properties of PCIs as defined and those of the estimators of the PCIs." This is an important statement. I believe that some of the confusion about capability indices found among practitioners has its origin in not being aware of this distinction. As a statistician, it seems to be a simple task to distingush between a parameter, its estimator, and their respective properties. However, for an engineer with little practice in statistical thinking this is not so easy. After many years of experience in teaching basic statistics to engineering students, I have realized that statistical concepts like an estimator and its distribution can be quite difficult to grasp. I do not agree with the authors when they claim that "anyone who understands the structure, working formula, and usage of the t-statistic should have few difficulties in comprehending analyses relevant to all but the most 'advanced' PCI indices so far proposed." This, of course, depends on what is meant by "understand." But, in general, I think that engineers with a basic course, only, in statistics will have to acquire more statistical knowledge to really understand the whole statistical picture involved in a capability analysis process. I believe that it is the responsibility of theoreticians to help to convey such useful knowledge to the practitioners via short courses, continuing education, or journals like *JQT*. In my opinion, this review article is an important contribution in this respect, since it contains so many valuable references. These references can form a broad basis for many different short courses on capability studies.

As a final comment, I would like to mention some further thoughts about what the authors refer to as "the lack of coordination between various directions of research and applications of the PCIs." I agree with the authors that the gap has to be bridged and that the lack of coordination has to be reduced. By now there is a quite sound theoretical base for process capability indices, as can be seen from this excellent review paper. I think it is now time for theoreticians, in a helpful way, to look more closely at how practitioners apply the theory of capability indices and how capability studies and the whole capability analysis process are used in practice. It may not be the "most efficient or optimal way" according to the theory, but it may provide useful contributions to the field. Then, questions such as the following can be asked: if practitioners do it their own way, how will this affect the decision procedures and final conclusions about the capability of the process? Will it differ much compared to "the optimal way?" Can guidelines be developed to support practitioners in the avoidance of the most troublesome pitfalls? Moving in this direction may be another way to further reduce lack of coordination.

I again thank the Editor for an opportunity to comment on this paper and congratulate Professors Kotz and Johnson for their important contributions to the field of process capability indices.

Additional References

BERGMAN, B. and KLEFSJÖ, B. (1994). *Quality from Customer Needs to Customer Satisfaction*. McGraw-Hill, New York and Studentlitteratur, Lund.

SULLIVAN, L. P. (1984). "Reducing Variability: A New Approach to Quality". *Quality Progress* 17, pp. 15–21.

VÄNNMAN, K. (2000). "A Graphical Method to Control Process Capability" in *Frontiers in Statistical Control* 6, edited by H. J. Lenz and P. Th. Wilrich. Physica-Verlag, Heidelberg, New York, pp. 290–311.

Discussion

RUSSELL A. BOYLES

Westview Analytics, Lake Oswego, OR 97034

THIS paper surveys a decade or so of research on process capability indices. Kotz and Johnson (K&J) review the definitions of many basic and not-so-basic indices, review mathematical relationships between indices, summarize what is known about interpreting some of these indices, and give many references to research on estimation of indices. K&J observe the proliferation of different indices and the gap between PCI researchers and practitioners. They promise to "identify some major concepts and methods, and boldly speculate on their immediate future." In the end, K&J fulfill the first objective but not the second. The paper is primarily bibliographic in nature, and doesn't offer practitioners much advice on what to do. Perhaps this just reflects a lack of consensus in the literature surveyed.

General Comments

Regarding the researcher-practitioner gap, it is true that many practitioners would have difficulty following the statistical calculations in Kotz and Johnson (1993). I agree with K&J that "statisticians have to learn a new vocabulary in order to talk about bias, variation and confidence intervals." Engineering programs typically do not provide adequate training in probability and statistics, and statisticians tend to write books and articles for each other. What practitioners really need, and don't often get in a comprehensible form, is guidance on which statistical procedure they should use and how to interpret the results.

I agree with K&J that "the topic of PCIs may be used by some academicians as an excuse for proposing new indices, regardless of their practical relevance, and mainly for the sake of the accompanying theory." This might explain why "the volume of recent publications may attest...to the importance...of PCIs...though not necessarily...in the eyes of engineers." In part, this is just the problem of statisticians writing books and articles for each other. It also prompts my observation that engineers usually think of capability indices as a nuisance imposed by management. Engineers are usually more interested in actual measurements, with actual units, and summary statistics based thereon.

Managers like PCIs because unitless indices reduce the complexity of the information with which they must grapple. The important issue here is the appropriate formulation of indices. In fact, organizations should design their own process capability indices (i.e., performance metrics) based on their own requirements. Statisticians are uniquely skilled to add value to this undertaking by making sure indices really mean what engineers and managers want them to mean, and by providing correct methods of inference for indices based on process data.

Regarding inferences from data, I must differ with K&J over the statement that "it is essential to keep in mind the basic assumption that a state of statistical control has been attained...and that observed values of X have no dependencies among themselves." Of course, to speak of process capability, one must first have a process. To me, this implies that we think we are done developing, and want to move into manufacturing, or have already done so. Sometimes, these distinctions are not so clear. In any case, we should have settled on materials, procedures, equipment, metrology, etc. Does this necessarily mean that our measurements resemble the output of a random number generator? I think not. This definition of "statistical control" is far too narrow. What it does mean is that process outputs should form a statistical distribution that is stable enough over long enough periods of time that we can sensibly try to characterize it. There are always assignable causes at work, so this distribution will in reality always be a mixture of some kind. Also, we cannot rule out autocorrelation in our measurements—for example, this would rule out virtually all high-tech manufacturing. If we restrict applications of PCIs to processes in the narrowly-defined state of statistical control, we are essentially saying that PCIs are never applicable. In

Dr. Boyles is Principal Consultant. He is a Member of ASQ. His email is rboyles@concentric.net.

my experience, one of the most positive uses of PCIs is to set goals and track progess as assignable causes are discovered and eliminated. Of course, using PCIs in this way requires that we base them on long-term rather than short-term variability. PCI values based on short-term variation can be used to set goals.

The assumption that process data comprises a simple random sample is virtually universal in the existing literature on PCIs. It is simplistic, misleading and detrimental to practitioners to portray the important area of process capability analysis as the study of simple random samples. Instead of proposing new indices, statisticians should be working on how to obtain confidence intervals for general indices when processes follow random effects or time series models.

Another deficiency in the existing literature is the preoccupation with finding sampling distributions of PCI estimators. For each new index, it seems we must go back to the drawing board and do new research to figure out how to get confidence intervals. Given the edifice of existing knowledge about inference for "regular" statistical models, this situation is bizarre and unsettling. In the regular case, any index imaginable is a parameter $\chi = C(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter vector of fixed dimension. All of the indices and distributions mentioned in this survey fall under this heading. It should, by now, be a routine matter to compute confidence limits for χ based on the likelihood function for $\boldsymbol{\theta}$. It should, by now, be well known that such intervals calibrate closely to nominal coverage probabilities, even with moderate sample sizes.

Particular Comments

In the first paragraph following Equation (12), K&J speculate that practitioners may not appreciate the relevance of Z-ratios, but they seem to be overlooking the fact that the PCIs $C_{pu} = (U - \mu)/3\sigma$ and $C_{pl} = (\mu - L)/3\sigma$ are themselves Z-ratios divided by 3. These are important because $C_p = (1/2)(C_{pu} + C_{pl})$ and $C_{pk} = \min(C_{pu}, C_{pl})$.

In the second paragraph following Equation (12), and once again in a later section, K&J state that C_{pmk}, a.k.a. C_{pn} (Choi and Owen (1990)), is not related to the expected proportion nonconforming. This is incorrect, at least for normally distributed processes. To see this, note that $C_{pmk} \leq C_{pk}$. Therefore, a given value of C_{pmk} implies at least as large a value of C_{pk}, which gives an upper bound on the expected proportion nonconforming.

In connection with Equation (24), it should be noted that C_{jkp} was studied in Boyles (1994) in relation to asymmetric tolerances, and that quantities similar to *MSE*+ and *MSE*− were used there to generalize C_{pm} to asymmetric tolerances.

K&J discuss, in connection with Equation (25), the role of measurement error in estimating PCIs. What bothers me about this discussion is the implication that there exist situations where there is *no* measurement error. It should be common knowledge that any assessment of process capability includes a component attributable to the measurement system.

In connection with Equation (27), the authors mention that the specification region for a vector characteristic $\mathbf{X}$ might have the general form

$$L \leq g(\mathbf{X}) \leq U$$

where $g(\cdot)$ might depend on the distribution of $\mathbf{X}$. I guess this could happen if specifications were based on the distribution of a baseline set of $\mathbf{X}$ values. But do we keep changing the definition of $g(\cdot)$ as we improve the $\mathbf{X}$ process? If we did, the index would serve no purpose. In any case, specifications should be performance-based whenever possible, and in this case $g(\cdot)$ cannot depend on the distribution of $\mathbf{X}$. Instead, it makes sense to talk about dependence on a vector $\mathbf{T}$ of target values. For example, bivariate alignment measurements (X, Y) occur in wafer fabrication and various types of post-processing of integrated circuits. In many cases there is a circular specification of the form

$$0 \leq g(X, Y) = (X - T_x)^2 + (Y - T_y)^2 \leq U.$$

The most common approach to this type of situation is to model the distribution of the random variable $G \equiv g(X, Y)$. Often, G is well modeled by Weibull or a lognormal distribution. Then we can apply univariate capability analysis using specifications $[0, U]$ for G.

Discussion

JOHN S. RAMBERG

University of Arizona, Tucson AZ 85721

It is an honor to contribute to a discussion of the Kotz and Johnson review. My students and I owe a great debt to Sam and Norman for their many outstanding contributions to statistics. Their work has provided a foundation for our research.

It is time to achieve closure on process capability issues. I hope that the review and discussions facilitate and accelerate this process. The authors' choice of the term "avalanche" is enlightening. Experienced mountaineers recognize the inherent danger: quality practitioners should also beware, lest they be "snowed" by the numerous works and miss the important contributions.

The authors have provided detailed commentary on the literature. The challenge to the discussants is to enhance these results for presentation to the *JQT* readership. It is important to understand the boundaries of the review. It addresses statistical advances in capability indices, in an enumerative statistics framework—random sampling from a population. I will begin my discussion in this framework, and then will go on to discuss the role of Deming's analytic statistics framework, followed by a brief commentary on engineering, management, and organizational issues.

My discussion is driven by my assessment of the potential impact of these papers on quality practice. I hope that it will enable practitioners to focus on the important developments, and that it will provide direction to researchers who have not had the opportunity to participate in quality initiatives. I conclude with a proposal for consensus.

State of Knowledge, October 1992

The first generation capability index C_p, and its predecessor $(1/C_p)$, describe a process using only process variability and the specifications. Both of these dimensionless ratios facilitate comparisons of quality measures of different metrics. The reciprocal represents the proportion of a specification length required due to the traditional representation of the process as $[6\sigma = (\mu + 3\sigma) - (\mu - 3\sigma)]$.

Second generation indices introduced the process mean and target into a single number index, but in quite different ways; C_{pk} was the result of a misguided attempt to introduce the process mean into an index. Its widespread acceptance has been a major setback to the quality profession. Unfortunately, no amount of tinkering with C_{pk} will make it into a rational measure of process capability. While C_{pm} also combines the mean deviation and standard deviation into a single number, it does so in a rational manner that engineers and statisticians have used for decades, the "root mean square error."

The state of knowledge as of 1992 concerning these indices is crisply summarized in the authors' introduction through their definitions of C_p, C_{pk}, C_{pm}, and two fundamental relationships between these indices. I submit another that relates C_{pk} to C_{pm}, using the "standardized mean deviation from target," $\beta = (\mu - T)/\sigma$. Combining Equations (5) and (6) from Kotz and Johnson, one obtains

$$C_{pk} = \frac{-\beta}{3} + \sqrt{1+\beta^2}C_{pm}.$$

I have not joined the stampede to summarize process capability by a single number. Rather, I prefer the use of two numbers, say $C_p = (U - L)/6\sigma$ and $D_p = (\mu - T)/6\sigma$. The standard deviation to the specification width is indexed by C_p, while D_p (which includes a sign) indexes the mean deviation from target to the same base (Ramberg (1989)). The former measures process variability; the latter measures location with respect to target. This pair provides better information for decision-making than a single number, and confidence intervals can be calculated for each.

An alternative to capability indices, average loss, had also been introduced by 1992. Taguchi (1986) popularized this loss function concept in the quality context. See Pignatiello and Ramberg (1991) for a

Dr. Ramberg is a Professor in the Department of Systems and Industrial Engineering. He is a Fellow of ASQ. His e-mail address is ramberg@sie.arizona.edu.

summary. Loss functions provide a natural foundation for the assessment of capability indices, as well as a worthy alternative. Taguchi demonstrated the superiority of quadratic loss to "goalpost" loss. Since the goalpost philosophy had been the "darling" of the quality profession, this result was of critical importance. Taguchi led the quality profession out of the "conformance domain," which had served as an obstruction to quality improvement, and toward the modern improvement philosophy.

Quadratic loss is defined as

$$L_Q(y,T) = k_Q(y-T)^2,$$

and goalpost loss is

$$L_{GP}(y,T) = \begin{cases} 0 & L \le y \le U \\ k_{GP} & \text{otherwise} \end{cases}.$$

The corresponding expected losses are:

$$\text{EQL} = E[k_Q(Y-T)^2] = k_Q[\sigma^2 + (\mu - T)^2]$$

$$E_F[L_{GP}(Y,T)] = k_{GP}[1 - (F(U) - F(L))].$$

The authors show that EQL is inversely proportional to C_{pm}^2. A slightly different version of their result, where $k_Q = (U-L)^2/3$, is

$$E_F[L_Q(Y,T)] = \frac{k_Q}{(C_{pm})^2}.$$

This important characterization explains several of the authors' comments concerning C_{pm} and the probability of non-conformance, P[NC].

First and second generation indices, as well as EQL, depend upon the probability distribution representing the process only through the distribution moments, and not a specific form, such as the normal distribution. This fact is often overlooked by researchers and practitioners. Estimators of these capability indices and expected quadratic loss are obtained by substituting sample estimators for the mean and standard deviation. These results are well known for the indices. The corresponding result for EQL is

$$\hat{\text{EQL}} = k_Q\left[(\bar{Y} - T)^2 + \frac{N}{N-1}S^2\right].$$

The standard errors were also well known in 1992; they can be found in the texts and standards cited, and they have been implemented in most professional quality software. These standard errors are asymptotic approximations, albeit good ones.

Development of capability standards had begun by 1992, and continues (Bigelow (1992)). These influential works deserve attention. They formalize procedures for calculation of capability, suggest new avenues of research, and their development requires statistical expertise.

New Generations, P[NC], and Normality

The authors discuss subsequent generations of indices, new indices based on P[NC], and methods for addressing non-normality. My conclusions, based on their review, are

1) that third and subsequent generations provide little for practitioners;
2) that P[NC] based indices, despite their appeal should not be used, nor should related results;
3) that procedures that address non-normality are suspect, and should be discarded, unless a standard error is given (I am not aware of any that do);
4) and that C_{pm} is a rational index for those requiring a single number summary.

To support these statements, let us first consider P[NC]. It does seem, on first thought, to be an excellent basis for approaching these problems, and many of us have pursued this path of thinking, including Carr (1991), who first proposed it.

The authors establish the basis for discarding the P[NC] approach. It follows from Palmer and Tsui's (1999) informative characterization of P[NC] as the expected value of "goalpost loss." Hence, its inappropriateness as an index follows from Taguchi's thesis. Specifically, a deviation of a performance measure from target results in a loss to society. Restricting observation to yes or no, if a measure is within or outside of specifications, is not sound practice.

Quality professionals should leave goalposts (and P[NC]) to their sports counterparts: goalposts to football, and goals to soccer and hockey, where "any shot within specifications" counts equally. Another significant drawback to P[NC] is that it is extremely sensitive to the underlying probability distribution. Sommerville and Montgomery (1996/97) exhibited this sensitivity by posing alternative distributions. Ramberg and Shetty (2001) elaborate further in a Six-Sigma framework.

Methodologies that address non-normality through systems of distributions are subject to the same problem, in that their use requires selection of a member from the family. These include the Johnson and the Pearson systems. It also applies to Ramberg and Schmeiser's (1978) lambda family, because four parameters must be estimated. I first observed the instability created by distribution selection when I was queried about a radical change in a capability index value from one report to the next. The reason was a switch in family member by the algorithm imbedded in the software: a switch based on only a few new observations.

My criticisms also apply to the procedure of Clement (1989) and to all kernel estimates (e.g., Polansky (1998, 2000)). They also undermine the conclusions of Flaig (1992, 1996/97, 1999, and 2000).

At the other end of the P[NC] spectrum are the "crude" bounds of Chebychev and Camp-Meidel, which do not depend upon the process distribution. They are of little value, and suggesting them seems ill-advised. Practitioners are better advised to direct their effort toward seeking an appropriate process distribution and ignore the bounds. Third and later generations of indices are of little value to practitioners. Superstructures (Vännman (1997b)) do provide insightful statistical explanations and support my conclusions.

In summary, I conclude that the normal distribution is representative of a majority of process situations. Augmenting it with the log normal distribution enlarges coverage substantially. The central limit theorem and its multiplicative counterpart provide the rational for this conclusion. Distinguishing between these two distributions is also straightforward. I recommend the use of these two distributions unless there is overwhelming evidence that neither is appropriate. See Pyzdek (1992) for some examples where such evidence does exist.

Six-Sigma

The reviewers link the Six-Sigma initiative to C_p, which could mislead some. The technical elaboration of Six-Sigma is often accomplished through the use of a normal distribution and capability indices. There is no doubt that the creators of Six-Sigma employed C_p, because it was a standard quality measure at that time. Early in the development of Six-Sigma as a methodology for improving business performance (quality, cost, and timeliness, or cycle time) proponents embraced the contributions of Taguchi, especially quadratic loss and robust design.

Six-Sigma was created in recognition of the increasing complexity of products, and the observed failure of products in achieving their predicted performance. It has evolved into a disciplined program for improvement of business performance, one that embraces leadership, infrastructure, and theoretical tools and methods. Ramberg (2000) discusses programmatic and technical issues. The computation of 3.4 parts per million (ppm) (Six-Sigma specs, 1.5 sigma mean deviation, normal random sampling) can be criticized for the same reason that I stated for P[NC]. The estimation of probabilities on the order of 3 ppm is a Herculean task. However, I should not be misunderstood here: I am not damning the philosophy that extremely low P[NC]s are required to produce complex products. Rather, I am stating that other metrics, such as mean square error, are necessary for achieving this and for measuring progress.

Education, Training, and Practice

I concur with the authors' comment about the inadequacy of engineering "education" in probability theory and statistical inference. A 1989 ABET sponsored conference spurred interest in both engineering statistics and probability, but this interest was transient in nature. Subsequently, the number of semester credits in engineering curricula was trimmed to a maximum of 120, a reasonable action. As a direct result, probability and statistics courses were eliminated from most curricula, with the exception of electrical, industrial, and systems engineering. In other curricula, this material, if it is covered, appears as part of a laboratory course, as one of many topics.

The authors' working "knowledge of the t-statistic" measure is an excellent one. I would appreciate a further elaboration on their interpretation of this term. I doubt that 25% of the students completing an engineering-level statistics course could answer this to the satisfaction of the authors. The problem in engineering education is even more fundamental; I think that only a slightly higher percentage of engineering faculty would do any better.

Seminar training, while valuable, is also inadequate: too little time, too little education. What causes us to think that we can teach all of quality engineering and management in a four week course (one without exams or graded homework)? Deming was among the first to point out the fallacy of the approach. Paraphrasing him, if differential calculus had been invented in the 22^{nd} century, would every industrial firm send their engineers to a three day course to become experts in the topic?

Many of the publications cited in the review illustrate a different type of inadequacy in education. They suggest a "background check," and a different type of "working knowledge" measure for quality educators and researchers.

How many statistics teachers could pass Sir Ronald Fisher's (1938) working knowledge background check? "I want to insist on the important moral that the responsibility for the teaching of statistical methods in our universities must be entrusted, certainly to highly trained mathematicians, but only to such mathematicians as have had sufficient prolonged experience of practical research, and of responsibility for drawing conclusions from actual data, upon which practical action is to be taken. Mathematical acuteness alone is not enough." How many have working knowledge of the leverage principle underlying variation reduction? How many have limited their discussion of statistics to enumerative statistics and ignored analytic statistics?

I do not think that the authors' "hope" that the knowledge gap between quality engineers and statisticians is closing is actually happening. Some of these papers make clear one noteworthy contribution of C_{pk}; to busy statisticians "who were fresh out of good ideas." It has been the equivalent of a full-employment act for statisticians over the last decade. I shudder to think about what the "participants" will teach the next generation of quality professionals if they do not gain some engineering and manufacturing background.

Issues in Practice

Capability indices are the most common summaries used by corporations to communicate quality issues. Most firms collect data from their processes and summarize these results in monthly reports; some provide them to their customers, often in fulfillment of a contractual agreement. Some conduct process capability studies on a regular basis.

Seldom are these performance measures given priorities, even in vague statements. Few reports state the portion of index that is due to variability versus mean deviation from target, a very important piece of information for quality improvement. Finally, the variability of the capability estimates is hardly ever addressed, in the report or otherwise.

The variability issue can be addressed by stating the estimate, its standard error, and the effective sample size. I use the following notation; [Parameter Estimator, Standard Error of Estimator]. For example, the general form for a capability index is

$$\left[\widehat{\theta}, \frac{\widehat{\theta}}{\sqrt{2N^*}} \cdot Cf\right].$$

Comparable results for the expected quadratic loss (Ramberg and Shetty (2001)) are also given. Each standard error estimator is a function of the parameter estimator, which can be traced to a common element of each index, the standard deviation. These standard errors are inversely proportional to the square root of the effective sample size, N^* (or, more formally, the degrees of freedom).

An artifact of the form of the estimator is that the variability of each parameter estimator increases with the value of the estimator. Hence, it takes a larger sample size to estimate an index of a process, within a given percentage of its true value, for a process that is capable versus one that is not. Deming introduced the concept of analytical statistics for the study of processes, rather than populations. Observations collected in subgroups over time should not be assumed to constitute a random sample. Pignatiello and Ramberg (1993) drew attention to improper uses of indices as well as the inadequacy of the enumerative framework, using the phrase "Just Say No" in their title. We did not think that such a trite action was sufficient. Rather, we employed the term in response to the "statistical terrorism" being perpetrated on the quality profession, per Burke, Davis and Kaminsky (1991). Calculation of an index should be resisted if it does not accurately represent the process. Pignatiello and Ramberg (1993, 1996) provide guidelines for collecting and analyzing process data using control charts (for stability assessment and data editing). See Hahn and Meeker (1991) for an interesting commentary on the use of enumerative (random sample) statistics in analytical situations.

In the following, I assume that "m" sub-samples of size "n" have been collected and analyzed, that the process has been judged to be stable, and that an edited data set which can be regarded as representative of the process has been selected (for individual observations, $m = N$ and $n = 1$.) Process capability standards (Bigelow (1992)) specify two different procedures for estimating the "process standard deviation" if the process is judged to be stable.

Procedure 1: Calculate the sample standard deviation over the edited data set, N' being the number of observations remaining. The effective sample size is $N^* = N' - 1$ (here, we are acting as though the edited data set constitutes a random sample.)

Procedure 2: Calculate subgroup variances, and estimate the process standard deviation as the square root of the weighted (by degrees of freedom) average (here, we are acting as though the sample is representative of the process, but a model more complex than simple random sampling is appropriate.) The effective sample size is then slightly more complicated, depending on the number of observations deleted from each sub-sample. Let m^* be the number of sub-samples remaining, and n_i^* be the number of observations in sub-sample i. Then we have

$$N^* = \sum_{i=1}^{m^*}(n_i^* - 1).$$

If the standard deviations are averaged, N^* is further reduced, although the estimator is more robust. If ranges are employed, N^* is still further reduced, and hence ranges should not be used.

These two procedures estimate potentially different standard deviations, unless the simple random sampling model is correct. In more complicated situations, which are the norm, there are multiple sources of variation (i.e., between sample and within sample, multiple filling heads, multiple molds, or multiple process lines, etc).

In these situations, Procedure 1 estimates the overall standard deviation, and Procedure 2 estimates the sub-sample standard deviation. The latter can be, and often is, smaller than the former. Standards permit the use of this potentially smaller, within sub-sample standard deviation. Thus, Procedure 2 actually estimates the process potential rather than the capability.

Experienced process professionals understand the advantages and disadvantages of each procedure. Selection of procedures is not simply a matter of which one yields the largest number of degrees of freedom. In support of the standards committee, I think that it is clear that their goal was a simple method that was clear to all, even lawyers and judges. A more sophisticated procedure might have been better received by statisticians.

Motorola's 1.5 sigma mean deviation was created to serve a similar purpose. Many have criticized this representation by commenting that a properly controlled process (monitored by a control chart) would not deviate this far, except for a short time. My experience is that deviations of this magnitude are common. The reason is, apparently, organizational in nature. Either the process is not properly controlled, or the ramifications of these deviations are not recognized. In one setting, the performance measure, which was important in itself, was also the basis for dimensional control at subsequent steps of the process. Yet the process mean deviated well over 1.5 sigma from target, and this information seemed of little concern to the organization.

A Proposal

1. Assess the importance of each performance measure by agreeing on a value for its conversion constant, k, in the (quadratic) loss function. (Create a loss function for the situation, especially if it is a prominent one.)
2. Estimate the average quadratic loss in monetary units by using estimators of the mean and standard deviation. (Two procedures were given for the latter.)
3. Estimate the standard error of the reported value, using the results given in Table 1. (Choose the effective sample size, N^*, based on the discussion under procedures.)
4. Calculate the percentage of the value of the reported value that is due to variability. (This is easily done by setting the mean equal to the target value and taking the ratio of this result to the original.)

If one can not depart from the comfortable world of dimensionless indices, then C_{pm} can be used instead of average loss, together with a "D_p" value. If one is being terrorized by a C_{pk} wielding customer, then C_{pk} accompanied by "C_p and D_p" values can be used.

Acknowledgment

I gratefully acknowledge an NSF grant (Transformations to Quality Organizations, now Innovation and Organizational Change Program) that provided me the opportunity to observe quality practices throughout the US, as well as the contributions of several graduate students: Salvador Arreola, Venu Dubagunta, Vinay Kulkarni, Manoj Sharma, and Nihar Senapati.

TABLE 1. Estimators and Standard Error Estimators

Parameter Estimator $\widehat{\theta}$	Standard Error Estimator $\widehat{\sigma}_{\widehat{\theta}}$	Correction Factor Cf
S	$\frac{S}{\sqrt{2N^*}}$	1
$\widehat{C}_p$	$\frac{\widehat{C}_p}{\sqrt{2N^*}}$	1
$\widehat{C}_{pk}$	$\frac{\widehat{C}_{pk}}{\sqrt{2N^*}} \cdot Cf$	$\sqrt{1+\frac{2}{9\widehat{C}_{pk}}} \leq \sqrt{\frac{11}{9}}$ when $\widehat{C}_{pk} \geq 1$
$\widehat{C}_{pm}$	$\frac{\widehat{C}_{pm}}{\sqrt{2N^*}} \cdot Cf$	$\left[\frac{\sqrt{1+2\widehat{\beta}^2}}{(1+\widehat{\beta}^2)}\right] \leq 1$
$\widehat{EQL}$	$\frac{\widehat{EQL}}{\sqrt{2N^*}} \cdot Cf$	$\frac{2\sqrt{1+2\widehat{\beta}^2}}{(1+\widehat{\beta}^2)} \leq 2$
$\overline{X}$	$\frac{S}{\sqrt{N^*}}$	1

Additional References

BIGELOW, J. S. (1992). *Standard Method for Calculating Process Capability and Performance Measures*, Draft 0.3, ANSI-ZI Committee on Quality Assurance.

BURKE, R. J.; DAVIS, R. D.; and KAMINSKY, F. C. (1991). "Resorting to Statistical Terrorism in Quality Control". *47th Annual Conference of American Society for Quality*, Rochester Section.

FISHER, R. A. (1938). "Presidential Address, First Indian Statistical Conference". *Sankhyā* 4, pp. 14–17.

HAHN, G. J. and MEEKER, W. Q. (1991). *Statistical Intervals: A Guide for Practitioners*. John Wiley and Sons, New York, NY.

KOTZ, S.; KOZUBOWSKI, T. J.; and PODGORSKI, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhauser, Boston, MA.

PIGNATIELLO J. J, JR. and RAMBERG, J. S. (1991). "Top Ten Triumphs and Tragedies of Genichi Taguchi". *Quality Engineering* 4, pp. 211–226.

PIGNATIELLO J. J, JR. and RAMBERG, J. S. (1996). "Process Capability: Engineering and Statistical Issues" in *Quality and Reliability Series 47, Statistical Applications in Process Control* edited by J. Bert Keats and Douglas C. Montgomery, Marcel Dekker, New York, NY, pp. 407–434.

RAMBERG, J. S. (1989). "Research Directions in Quality". *Proceedings of the IIE Integrated Systems Conference on the Resurgence of Excellence*, Nov. 1989, pp. 368–374.

RAMBERG, J. S.(2000). "Six Sigma: Fad or Fundamental", *Quality Digest*, May, p. 28-30.

RAMBERG, J. S.; TADIKAMALLA, P. R.; DUDEWICZ, E. J.; and MYKYTKA, E. F. (1979). "A Probability Distribution for Fitting Empirical Data, Which Simplifies Applications". *Technometrics* 21, pp. 201–214.

RAMBERG, J. S. and SHETTY, S. (2001). "Six Sigma, Quadratic Loss, and the Normal Distribution", *Systems Science*, (to appear).

TAGUCHI, G. (1986). *Introduction to Quality Engineering*. Tokyo, Asian Productivity Organization.

Response

SAMUEL KOTZ

George Washington University, Washington, DC 20052

NORMAN L. JOHNSON

University of North Carolina, Chapel Hill, NC 27599

We appreciate the time, thought, and effort that the discussants have devoted to our review. Their contributions are wide-ranging in content, and reveal a remarkable variation of attitudes toward our subject matter. Consequently, the discussions received from these experts have, in our opinion, thrown considerably greater light on many aspects of PCIs than we could provide in a single paper.

In a subject with so broad a field of applications, affecting many different kinds of users, it is inevitable that there are disagreements among us. Indeed, reactions of the discussants to our review ranged from approval to disagreement and disappointment. The latter contribute to giving us a deeper understanding of the place and possibilities of PCIs in current circumstances. It also reflects the situation described in our "Postscript for Practitioners," and explains, incidentally, why we felt it to be a 'bold' undertaking, in the sense of 'risky,' to speculate on even the immediate future of the field. Our use of the word 'bold' seems to have caused some semantic misunderstandings, references to which will appear in the following individual remarks on each commentary.

Hubele

Dr. Hubele writes sensibly regarding the choice and use of PCIs. We enthusiastically endorse her advice that "software cannot replace understanding."

Spiring, Chang, Yeung, and Leung

These authors are disappointed that our bibliography is inadequate in regard to current work, and that there is no "reasonable review" of philosophies regarding process capability and PCIs.

The main contribution of their discussion appears to be in supporting the use of Taguchi-type index(es), based on an 'expected loss' function, and in drawing attention to recent work on developing suitable loss functions. We realize, providing it is possible to express expected loss as a function of departure(s) from target(s) with sufficient accuracy, that it is logical to use such formulas. It would also, possibly, be logical to use the expected loss itself as the index, rather than clothing it in a way analogous to that used in constructing PCIs.

We would, of course, be happy to see accounts of ways in which sufficiently accurate estimates of expected loss can be obtained in different kinds of practical situations.

Rodriguez

Dr. Rodriguez devotes his discussion to the consideration of the "gap between theoreticians and practitioners." Like some other discussants, he disagrees with our (perhaps over-optimistic) opinion. His assessment of the current usage of available PCIs is in agreement with that of Spiring et al.: that C_p and C_{pk} are by far the most commonly used PCIs, that C_{pm} is a much less commonly used runner-up, and that the newer indices are used very seldomly. His discussion of three current sources of "confusion" in the application of PCIs is, we think, a reasonable and clear description of the present situation by an author who is familiar with both the theory and application of PCIs.

Bothe

Mr. Bothe provides several insightful remarks on the use (and abuse) of PCIs. In particular, he advocates the use of 'multiple measures,' which we also support. Mr. Bothe provides useful details about a desirable set of measures.

He shares Dr. Ramberg's dislike of C_{pk}, and provides explicit details of situations where a comparison of C_{pk} values (alone) can be misleading. He also makes some comments defending the use of C_{pm} and even C_{pmk}. We do not support his general conclusions, but find that his arguments are clear and challenging. The comments on the use of an estimate of p as an index are also well-expressed. The final paragraph of Bothe's discussion should be read, and memorized, if possible, by all existing or potential users of PCIs.

Lu and Rudy

These authors address the uses of performance PCIs (P_p,P_{pk}) and process PCIs (C_p, C_{pk}), with particular emphasis on practice in the automotive industry. Their discussion provides a useful insight into the natures of the two classes of indices, and may assist in resolving the confusion between them described by Dr. Rodriguez. There is some lack of consistency between the statement that the Box-Cox transformation can be used to transform data to normality and the immediately following passage describing what to do if they cannot be transformed.

We agree with the authors' comments on difficulties arising from "rigid mentality" in the use of PCIs, and also, of course, with their final remarks on the intrinsic limitations of trying to use a single index to describe a distribution.

Vännman

Dr. Vännman's discussion starts by making suggestions for reducing the "gap between theory and practice," which we have already discussed above. She proceeds to support the contention that too much should not be expected from the value of a single index. Her proposal for the use of a process capability plot is certainly worthy of consideration. She next presents Taguchi-type indices and the underlying concept of loss-function as an alternative to the expected proportion of nonconforming items. We are grateful for Dr. Vännman's support of our opinion on the need for carefully distinguishing between the properties of PCIs and those of their estimators.

Boyles

Dr. Boyles' comments have brought to light some potential misunderstandings that we would like to address. In particular, we do not understand why 'speculation' (however "bold") must necessarily lead to advising practitioners what to do. As we have noted above, our use of the word "bold" refers to our feelings about attempting to assess future movements in a sea of apparently chaotic motion, and our intentions are not to tell practitioners "what to do." It is interesting that Dr. Boyles regards our review as being "primarily bibliographic," while Spiring et al. refer to the inadequacy of our bibliography of recent research. We also note that, while we agree that "statisticians have to learn a new vocabulary ...", the quotation is, in fact, from Orchard (2000).

A possible misunderstanding also arises from our discussion of "statistical control." Dr. Boyles regards this as requiring that measurements "resemble the output of a random number generator," and notes that this is often not the case. Our statement on the matter does, unfortunately, refer to the lack of dependency among values of X. We should have limited ourselves to the stability of distribution. The latter is not affected even if the distribution is a mixture.

Dr. Boyles states that we have "overlooked" the fact that Z-ratios are related to PCIs. This appears to refer to our remarks on some practitioners' lack of appreciation for the relevance of Z-ratios, despite the fact that, as we remark, the two ratios add up to C_p. Our intended point was that Z-ratios need explicit reference to a distribution for interpretation, while PCIs are regarded as having an independent status.

There again appears to be a misunderstanding, possibly due to insufficient emphasis in our paper, of our argument that C_{pmk} is not *specifically* related to p, because we cannot calculate p from C_{pmk}, as we can from C_p and C_{pk} *together*. Although C_{pmk} does provide an upper bound for p, it can be quite a generous one if the Taguchi component is sizeable. Incidentally, we regret our omission of the fact that C_{jkp} was discussed in Boyles (1994).

Finally, some of Dr. Boyles' comments appear to imply that by considering measurement error, we suggest that there are situations wherein there is no measurement error. We are puzzled by this comment, and certainly did not deliberately imply such a dogmatic position. We did not even explicitly state that there might be situations in which the effect of measurement error could be negligible (although we do think that this is so).

Ramberg

This discussion contains a number of positions and statements with which we disagree. In particular, there is an almost unqualified faith in the relevance of quadratic loss functions, in true Taguchi fashion. In the section titled "Process Summary Proposal," the first item assumes that a quadratic loss function *is* appropriate; the only problem is deciding on the value of the multiplier ("conversion constant").

We are somewhat surprised to learn that we "establish the basis for discarding the P[NC] approach" (in the "New Generations ..." section). We thought that we were just commenting on considerations (mostly statistical) relevant to the use of this approach. We did note the existence of distribution-free bounds, but we did not attempt to hide their crudeness.

It seems to us that a PCI that corresponds to a specific form of expected loss function different from the Taguchi quadratic loss function is an insufficient basis for its immediate dismissal; this seems to imply that we always *ought* to have a loss function in mind, and that only the quadratic form endorsed by Taguchi is acceptable. We disagree, therefore, with the position implied in Ramberg's comment that "its inappropriateness as an index follows from Taguchi's thesis."

It is a pleasure to express our appreciation of Dr. Ramberg's equation connecting C_{pk} and C_{pm}, obtained by eliminating C_p from our Equations (5) and (6). This is probably more instructive than the equations we present. However, it should be noted that: (a) they are valid only if $T = M$; and (b) β should be $|\mu - T|/\sigma$.

In conclusion, let us not, so to speak, throw out the baby with the bathwater.