Departament d'Estadística i I.O., Universitat de València.
Facultat de Matemàtiques, 46100–Burjassot, València, Spain.
Tel. 34.6.363.6048, Fax 34.6.363.6048 (direct), 34.6.386.4735 (office)
Internet: bernardo@uv.es, Web: http://www.uv.es/~bernardo/

Printed on December 9, 1996
To be published in *The Statistician*

A General Multivariate Bayesian Process Capability Index

JOSÉ M. BERNARDO* and TELBA Z. IRONY**

**Universitat de València, Spain*
***George Washington University, U.S.A.*

SUMMARY

Process capability analysis is designed to estimate the proportion of parts that do not meet engineering requirements in a *stable* production process. In this paper, we review and criticize the *capability indices* typically used in industry for this purpose, and propose a *general multivariate Bayesian capability index* which contains the conventional index as a limiting case. We further derive its analytical expression under standard assumptions, discuss numerical approximations, and illustrate the theory with some examples.

Keywords: DECISION ANALYSIS; PREDICTIVE DISTRIBUTIONS; PROCESS CAPABILITY; STATISTICAL QUALITY CONTROL; UTILITY FUNCTION.

1. INTRODUCTION

In manufacturing industry, *process capability analysis* is used to flag high values of the proportion of parts being produced that do not meet engineering requirements, in order to prevent further production of unacceptable output. This analysis assumes that the process is *stable*, that is, that any random sample of observations from the process may be regarded as a random sample from the same underlying distribution, a situation often described in the engineering literature as that of a process in *statistical control*. Capability analysis is typically performed by evaluating *capability indices* which relate the *allowable spread* of the process, defined by the engineering specifications, to the *natural spread* of the process, represented by a multiple of the standard deviation of the output. Assuming that the output is normally distributed, the expected proportion of *non-conforming* parts, *i.e.*, those which will lie outside the engineering specification limits, may be estimated from the capability index.

However, the abundance of outputs from skewed distributions and the censoring effects induced by the finite precision of actual measurements, make often rather unreasonable the normality assumption on which traditional capability indices are intuitively based. Moreover,

José M. Bernardo is Professor of Statistics at the University of Valencia; his research was partially funded with grant PB93-1204 of the DGICYT, Madrid, Spain. Dr. Telba Irony was supported by the US National Science Foundation, Grant DDM-9209334

the sampling distributions of the estimators of the capability indices are often intractable, even under normality assumptions. As a consequence, point estimators of the capability indices, with no reference to their precision, are usually quoted; this is a misleading practice, for even relatively large samples may produce rather unreliable estimators.

In this paper, we propose a Bayesian approach to evaluating process capability which, within a decision-theoretical framework, directly assesses the proportion of future parts which may be expected to lie outside the tolerance limits. This results in a *new general capability index* which (i) has a solid, decision theoretical foundation, (ii) does not require the process to be normal, (iii) may be used for multivariate observations, (iv) may accommodate measurements with error, and (v) contains the conventional index as a limiting case. The proposed capability index is a direct function of the data, whose value is sufficient to solve the relevant decision problem. Nevertheless, robustness considerations make it desirable to assess the "precision" of the capability index thus obtained; within the Bayesian framework this is achieved by deriving its posterior predictive distribution.

In Section 2, we review and criticize the indices traditionally used to evaluate process capability. In Section 3, we propose a new general capability index motivated by decision theoretical considerations and discuss some of its properties. In Section 4, we propose a general hierarchical model which describes the stochastic mechanism governing many production processes, derive the corresponding expression for the proposed capability index, and obtain its analytical form under the simplifying assumptions which are often made in industrial practice. In Section 5 the behaviour of the index proposed is evaluated with both real and simulated data. Section 6 contains some final remarks and suggests areas for additional research.

2. CONVENTIONAL CAPABILITY ANALYSIS

Let $x_1, \ldots x_n$ be the *actual values* of a certain attribute which correspond to n randomly selected items from a production process and suppose that such an attribute should lie between l_0 (lower limit) and l_1 (upper limit) in order to conform to engineering specifications. Items which lie outside (l_0, l_1) will be termed *non-conforming*. The special cases where only one specification limit is required are obtained by letting $l_0 \to -\infty$ or $l_1 \to +\infty$.

Process capability analysis is designed to monitor the proportion of items which are expected to fall outside the engineering specifications in order to prevent an excessive production of non-conforming output. This is usually done at specified rating periods, using the *measurements* $\{y_1, \ldots, y_n\}$ taken on, say, n produced items and assuming that (i) there is no measurement error, so that $y_i = x_i, i = 1, \ldots, n$, *i.e.*, the measurements are taken to be the actual values, and (ii) that the x_i's are identically distributed with, say, mean μ and standard deviation σ.

Traditional capability analysis then proceeds to evaluate *capability indices* which relate the *allowable spread* of the process $l_1 - l_0$ to its *"natural" spread*, customarily taken to be 6σ. Thus, the *capability potential* C_p is defined to be

$$C_p = \frac{l_1 - l_0}{6\sigma} \,, \tag{1}$$

i.e., the ratio of the allowable spread to the natural spread. In particular, if the x_i's are normally distributed $\mathrm{N}(x \,|\, \mu, \sigma)$, and the process is centered at the midpoint $m = (l_1 + l_0)/2$ of the specification limits, so that $E(x) = \mu = m$, then a *capable* process, defined as one for which $C_p \geq 1$, will result in, at most, 0.27% of non-conforming items, *i.e.*, 2700 non-conforming items per million.

The capability potential only takes care of the process spread. Obviously, it would be possible to have any proportion of items outside the specification limits by merely relocating the process mean; thus, C_p only quantifies the *potential* performance of the process, which will only be attained if the process is centered at the midpoint of the specification limits.

The *actual* performance of a production process is traditionally measured by the *capability index* C_{pk} defined to be

$$C_{pk} = \min\left\{\frac{l_1 - \mu}{3\sigma}, \frac{\mu - l_0}{3\sigma}\right\}$$

which is a normalized distance between the process mean and its closest specification limit. It can easily be verified that

$$C_{pk} = C_p(1 - \omega), \qquad \omega = \frac{2|m - \mu|}{l_1 - l_0}$$

so that C_{pk} modifies C_p with a standardized measure ω of non centrality of the process, and $C_{pk} = C_p$ if, and only if, the process is centered at the midpoint m of the specification limits.

The definition of C_{pk} includes, as special cases, those processes where only one limit exists, by setting either $l_0 \to -\infty$ of $l_1 \to +\infty$, in which case it reduces to the appropriate standardized measure. Thus, if there is no lower specification limit, we obtain $C_{pk} = (l_1 - \mu)/3\sigma$ by simply letting $l_0 \to -\infty$ in the original definition. Similarly, if there is no upper specification limit, $C_{pk} = (\mu - l_0)/3\sigma$

The capability index C_{pk}, originally introduced in Japan, has become an *international standard*, especially in the car industry, to the point that it is *often specified in contractual agreements*; typical requirements are $C_{pk} \geq 1.66$ to qualify a supplier with pilot runs, and $C_{pk} \geq 1.33$ for production runs. This guarantees the practical use of good capability indices, but this also means that, to be accepted, new techniques must *clearly relate to present standard practice*.

Given a random sample $x_1, \ldots, x_n$ from the process, the capability index is typically estimated by the naïve estimator

$$\hat{C}_{pk}(\bar{x}, s) = \min\left\{\frac{l_1 - \bar{x}}{3s}, \frac{\bar{x} - l_0}{3s}\right\}, \tag{2}$$

where $n\bar{x} = \sum_i x_i$, and $(n-1)s^2 = \sum_i (x_i - \bar{x})^2$. Note that $\hat{C}_{pk}(\bar{x}, s)$ does *not* explicitely depend on the sample size n. A "capability analysis" usually consists in (i) verifying the stability of the process (ii) checking the normality assumption, and (iii) computing and reporting the value of the estimated capability index $\hat{C}_{pk}$. The sampling properties of $\hat{C}_{pk}$ under normality have been studied by Kane (1986), Chou and Owen (1989), and by Zhang *et al.* (1990). From a Bayesian point of view, Cheng and Spiring (1989) have derived the posterior distribution of C_p under normality which corresponds to a conventional "non-informative" prior; we are not aware of a similar analysis for C_{pk}. Chan *et al.* (1988) propose a variation of the capability index which uses target values instead of the process mean μ. Chan *et al.* (1991), Taam *et al.* (1993), and Wierda (1993), extend the capability index to multivariate observations. Other related references are Bisell (1990), Boyles (1991), Pearn *et al.* (1992), Kotz *et al.* (1993) and Pearn and Chen (1996). See the monograph by Kotz and Johnson (1993), and references therein, for further discussion of capability analysis.

It should be clear to the reader that the *intuitive* basis of the definition of the capability index C_{pk} *heavily* depends on an implicit assumption of normality of the output, for it is only in this

case that 6σ may be justifiable as a measure of "natural" spread; in particular, it is only under normality that a centered process will have 0.27% of its output outside the $\mu \pm 3\sigma$ interval. It is a fact, however, that production processes are *very often non normal*. Indeed,

(i) *Skewed distributions* are frequent, but quality practitioners are often not qualified to attempt possible transformations to normality. Moreover, they usually prefer to work with non-transformed values, for which they have a real experience, and which do not require awkward transformations of the customer specifications.

(ii) The precision limits of the measurement mechanisms often impose a discretization of otherwise continuous attributes yielding *discrete* distributions which only sometimes may be considered approximately normal.

Furthermore, a general capability index should be able to deal with *multivariate* observations. Indeed, several attributes are often measured on the same item. As one would expect, those measurements tend to show strong correlation, since a non-conforming item is often faulty in several related dimensions. The incorporation of such correlation into the analysis is bound to enhance the results by (i) allowing the use of smaller sample sizes, and (ii) providing useful information to identify possible causes for the observed defects.

The preceding discussion suggests the need for a *general multivariate* capability index whose *intuitive basis* and *formal definition* should both be *independent* of any particular model for the production process. Moreover, to gain acceptance, the new index should be related to the established, conventional index under appropriate conditions.

In the next section, we use a decision analysis framework to motivate such an index. As is always the case in the process capability literature, we shall assume that the process is *stable*, so that the available observations may be regarded as *exchangeable*, *i.e.*, as a random sample from some underlying distribution. It is certainly not trivial to check such an assumption in practice; the reader interested in a Bayesian formulation to such an important problem is referred to Bernardo and Smith (1994, Ch. 6) and references therein.

3. CAPABILITY ANALYSIS AS A DECISION PROBLEM

The whole purpose of capability analysis is to help *engineering decision making*. Decision problems such as to intervene or not in a production process, to accept or reject a production batch, or to review or not managerial action, need often to be made by either quality engineers or managers (see Singpurwalla, 1992, for a comprehensive view of quality engineering from a decision analysis perspective). Moreover, the consequences of these decisions often depend on the proportion of non-conforming items that are expected from the production process. Thus, a large proportion of quality-related decision problems relative to a stable production process may usefully be modelled as a decision problem with two possible actions:

$d_1 =$ *accept* (maintain the process as is)

$d_2 =$ *reject* (intervene upon the process)

with *uncertainty* about the values of the relevant characteristics $\boldsymbol{z}_N = \{\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_N\}$ of an eventual production of, say, N items, where $\boldsymbol{x}'_i$ is a vector which contains the values of the attributes corresponding to an item i yet to be produced (we will use the superindex $'$ to denote unobserved values). If $\boldsymbol{y}_i$ denotes one measurement of $\boldsymbol{x}_i$ —typically subject to error—, the *available data* D usually consist in the *measurements* $D = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$ of the characteristics of interest $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ which correspond to a random sample of, say, n items from the production process.

Basic decision theory (see, for example, Fishburn, 1981, or Bernardo and Smith, 1994, Ch. 2), suggests that coherent decision making requires:

(i) to describe with a *utility function* the preferences of the decision maker among the possible consequences of his or her decisions; here, one should specify the values $u(d_1, \boldsymbol{z}_N)$ and $u(d_2, \boldsymbol{z}_N)$ of, respectively, accepting or rejecting, as a function of the (unknown) characteristics $\boldsymbol{z}_N = \{\boldsymbol{x}_1', \ldots, \boldsymbol{x}_N'\}$ of the N items to be produced;

(ii) to describe, with a *probability distribution*, the available information about the uncertain attributes of interest; here one should specify the *predictive* distribution $p(\boldsymbol{z}_N \mid D)$ of the values of the *unknown characteristics* $\boldsymbol{z}_N$ given the available measurements D;

(iii) to choose the action d_i which *maximizes the expected utility*

$$\overline{u}(d_i \mid D) = \int_{\boldsymbol{Z}_N} u(d_i, \boldsymbol{z}_N)\, p(\boldsymbol{z}_N \mid D)\, d\boldsymbol{z}_N, \quad i = 1, 2.$$

With only two possible actions, the decision criterion takes the form

$$\text{Accept iff} \quad \int_{\boldsymbol{Z}_N} \{u(d_1, \boldsymbol{z}_N) - u(d_2, \boldsymbol{z}_N)\}\, p(\boldsymbol{z}_N \mid D)\, d\boldsymbol{z}_N \geq 0 \tag{3}$$

so that it suffices to specify the conditional difference of utilities, $u(d_1, \boldsymbol{z}_N) - u(d_2, \boldsymbol{z}_N)$.

The consequences of quality-related decisions often solely depend on the proportion of conforming items produced. Preferences in those decision problems are often described by a linear difference of utilities of the form

$$u(d_1, \boldsymbol{z}_N) - u(d_2, \boldsymbol{z}_N) = aR - bN, \tag{4}$$

where $a > b > 0$, and R is the actual number of items which meet all the engineering specifications among the N items produced. For instance, the utility of accepting the production is often of the form

$$u(d_1, \boldsymbol{z}_N) = gR - c(N - R) - l(N),$$

where g is the unitary *gain* from accepting a conforming item, c is the unitary *cost* from accepting a non-conforming item, and $l(N)$ is the cost of producing N items; similarly, the utility from rejecting the production —and then inspecting all items and repairing or substituting those non-conforming— is often of the form

$$u(d_2, \boldsymbol{z}_N) = gN - iN - r(N - R) - l(N),$$

where $i < c$ is the unitary cost of *inspecting* an item, and $r < g$ the unitary cost of *repairing* or substituting a non-conforming item. In this case, the utility difference becomes

$$u(d_1, \boldsymbol{z}_N) - u(d_2, \boldsymbol{z}_N) = (g + c - r)R - (g + c - r - i)N,$$

which is of the linear form specified above.

If the preferences of the decision maker are well described by a utility function which satisfies (4), then the decision criterion (3) reduces to

$$\begin{aligned} \text{Accept iff} \quad \int (aR - bN)\, p(\boldsymbol{z}_N \mid D) d\boldsymbol{z}_N &= \\ &= \sum_{R=0}^{N} (aR - bN) p(R \mid N, D) = a\mathrm{E}[R \mid N, D] - bN \\ &= aN \Pr(\boldsymbol{x} \in A \mid D) - bN \geq 0, \end{aligned}$$

where A is the region of acceptable attributes, or *tolerance region* as defined by the engineering specifications. Since $a > b > 0$ and N is positive, this decision criterion may be re-expressed as

$$\text{Accept iff} \qquad \Pr(\boldsymbol{x} \in A \,|\, D) \geq p_0, \quad p_0 = \frac{a}{b}\,, \tag{5}$$

that is, if the *expected* proportion of conforming items

$$\Pr(\boldsymbol{x} \in A \,|\, D) = \int_A p(\boldsymbol{x} \,|\, D) \boldsymbol{dx}, \tag{6}$$

is larger than some threshold limit p_0, defined by the decision maker's preferences. Here, $p(\boldsymbol{x} \,|\, D)$ is the predictive distribution of $\boldsymbol{x}$ given the available data. This provides a *normative* justification for the intuitive idea, already present in Aitchison and Dunsmore (1975), or Carr (1991), that process performance should be directly evaluated in terms of the expected proportion of conforming items. Moreover, it provides a *precise* procedure to specify the lower limit of the expected proportion of conforming items required to accept the production, as an *explicit* function of the economic characteristics of the problem.

Of course, there is nothing in the general theory that *requires* the use of *linear* utility functions. If industrial economics are best described by non-linear utility functions, such functions should be used to derive the corresponding optimal decision criterion. We will now show that the criterion implied by utility functions which satisfy (4) is formally equivalent to a criterion which contains that based on the conventional index as a limiting case. It follows that, if industrial economics are *not* linear in this sense, then the use of the traditional index would be inappropriate, even in those limiting conditions.

The preceding analysis shows that attention should be centered on the expected probability (6) of an item being conforming. However, this probability will typically be *very* high, usually larger than 0.999, and most engineers are not trained to appreciate the implications of extreme probabilities. It is, therefore, convenient to use some one-to-one monotone transformation to substitute the probability scale by some other appropriate scale where large probabilities become more spaced. A natural choice would be the log-odds scale but, as mentioned before, for a new methodology to gain acceptance in practice it is important that it is easily related to established standards; as we shall immediately establish, this may easily be done using a probit scale. This motivates the following

Definition. *The **Bayes capability index** C_b (b for Bayes), is given by*

$$C_b(D) = \frac{1}{3}\Phi^{-1}\{\Pr(\boldsymbol{x} \in A \,|\, D)\}$$

where A is the tolerance region, Φ is the distribution function of the standard normal distribution, and D the available data.

We shall now prove that the corresponding criterion

$$\text{Accept iff} \qquad C_b(D) \geq c_0, \tag{7}$$

where c_0 is some threshold value, encompasses the standard criterion

$$\text{Accept iff} \qquad \hat{C}_{pk}(\bar{x}, s) \geq c_0 \tag{8}$$

as a particular limiting case. Note that criterion (7) is equivalent to criterion (5), as a trivial consequence of the fact that Φ is a one-to-one monotone function.

Note that $C_b(D)$ is a direct *function of the data*, so that *no estimation problem arises*. This is an immediate consequence of a Bayesian approach, and it is in direct contrast with other apparently related proposals, such as Wierda (1993), where the capability index is defined in terms of an unknown 'true' proportion, say θ, of conforming items, which has to be estimated.

We also note that, as a monotone function of the probability $\Pr(\boldsymbol{x} \in A \,|\, D)$ of an item being conforming, the proposed capability index is invariant under one-to-one transformations of the characteristics of interest. Indeed, if f is such a function,

$$\Pr(\boldsymbol{x} \in A \,|\, D) = \Pr(f(\boldsymbol{x}) \in f(A) \,|\, D).$$

Thus, if an engineer is interested in, say, the area within a circle, a Bayesian capability analysis may indistinctly be carried out in terms of the area itself, or in terms of its diameter. The choice may depend on which is the magnitude easier to measure, a legal requirement, or whatever, but the result will *not* depend on the particular scale chosen. This highly desirable *invariance* property is *not* shared by conventional capability indices.

In most practical problems, one is interested in assessing the *robustness* of the optimal decision with respect to the particular sample obtained. Obviously, this depends on the uncertainty associated with the value of capability index which corresponds to that sample. Within the Bayesian framework, this is described by its posterior predictive distribution $p(C_b \,|\, D)$ given by

$$\begin{aligned} p(C_b \,|\, D) &= p(C_b \,|\, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) \\ &= \int \ldots \int p(C_b \,|\, \boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_n)\, p(\boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_n \,|\, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)\, d\boldsymbol{y}'_1, \ldots, d\boldsymbol{y}'_n, \end{aligned} \tag{9}$$

where $p(\boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_n \,|\, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ is the joint posterior predictive distribution of a future sample $\boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_n$ of the same size.

Often, the integral (9) will not be analytically feasible. If a collection of, say m exchangeable data sets of size n, $\{(\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{in}), i = 1, \ldots, m\}$, were available, then $p(C_b \,|\, D)$ could be approximated by Monte Carlo to

$$p(C_b \,|\, D) \approx \frac{1}{m} \sum_{i=1}^{m} p(C_b \,|\, \boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{in}). \tag{10}$$

Otherwise, resampling m times with replication from the original sample $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$, and using this as a proxy for $\{(\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{in}), i = 1, \ldots, m\}$ in (10), would provide a *bootstrap* (Efron, 1982) approximation for the desired posterior predictive distribution (9).

We conclude this section by deriving the precise relationship between the general Bayesian index $C_b(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ and the traditional univariate index $\hat{C}_{pk}(x_1, \ldots, x_n)$. To do this, we specialize to the *univariate normal* case and further assume that (i) there are no observation errors, —so that $y_i = x_i$, $i = 1, \ldots, n$, and the data consist of a random sample $D = \{x_1, \ldots, x_n\}$ of measurements from a normal process $\mathrm{N}(x_i \,|\, \mu, \sigma)$—, and that (ii) n is sufficiently large for C_{pk} to be well approximated by its estimator $\hat{C}_{pk}$.

In those conditions, the predictive distribution $p(x|x_1, \ldots, x_n)$ will converge to the normal distribution $\mathrm{N}(x|\mu, \sigma)$ and, hence, the decision criterion (7) will reduce to

$$\text{Accept iff} \qquad \frac{1}{3}\Phi^{-1}\left\{\Phi\left(\frac{l_1 - \mu}{\sigma}\right) - \Phi\left(\frac{l_0 - \mu}{\sigma}\right)\right\} \geq c_0$$

If the process is *non-centered and potentially capable*, *i.e.*, if μ is appreciably different from $m = (l_0 + l_1)/2$ *and* its "allowable" spread is sufficiently large, then either

(i) $(l_1 - \mu) >> (\mu - l_0)$ and $(l_1 - \mu) >> \sigma$, in which case $\Phi((l_1 - \mu)/\sigma)) \simeq 1$, and the criterion reduces to accepting iff

$$\frac{1}{3}\Phi^{-1}\left\{1 - \Phi\left(\frac{l_0 - \mu}{\sigma}\right)\right\} = \frac{1}{3}\Phi^{-1}\left\{\Phi\left(\frac{\mu - l_0}{\sigma}\right)\right\} = C_{pk} \geq c_0, \quad \text{or}$$

(ii) $(\mu - l_0) >> (l_1 - \mu)$ and $(\mu - l_0) >> \sigma$, in which case $\Phi\{(l_0 - \mu)/\sigma)\} \simeq 0$, and the criterion reduces to accepting iff

$$\frac{1}{3}\Phi^{-1}\left\{\Phi\left(\frac{l_1 - \mu}{\sigma}\right)\right\} = C_{pk} \geq c_0$$

It follows that, *under univariate normal conditions, with no observation errors, and for non-centered potentially capable processes*, the Bayesian criterion (7) is *asymptotically* equivalent to the traditional criterion (8). Note that if *either* $|\mu - m|$ or $(l_1 - l_0)$ are not large enough compared to σ, then *both* specification limits become relevant and the Bayesian index will (appropriately) *not* reduce to the conventional index. This demonstrates one advantage of the Bayesian index, for it automatically takes into account the possibility of having non-conforming units in both sides of the interval defined by the specification limits, whereas $\hat{C}_{pk}$ will only consider that side in which μ is closer to the specification limit, however small the difference between both sides might be. This is the reason why, in practice, quality engineers insist on a two stages strategy, whereby the potential capability C_p of the process is established before the C_{pk} index is used. This precaution is unnecessary if the Bayesian index is used.

We should stress that even with non-centered potentially capable normal process with no observation errors, the equivalence between Bayesian and traditional capability indices is *only asymptotic*. Thus, even for relatively large samples, numerical differences will remain; this is a good example of the known inadequacies of classical prediction techniques, which, by simply substituting the unknown parameters by their estimates, totally ignore the uncertainty thereby added to the intended prediction.

In the next section, we propose a hierarchical model intended to describe standard production processes. As a simple, particular case, we derive the exact analytic form of the Bayesian capability index under the assumptions of normality and no observation errors usually made in industrial practice.

4. A HIERARCHICAL MODEL FOR PROCESS OUTPUT

As mentioned before, standard capability analysis typically assumes that the *observed* measurements $\boldsymbol{y}_j$ are identical to the actual values $\boldsymbol{x}_j$, thereby ignoring any limitations of the measurement mechanism. In this section, we formulate a hierarchical model which describes the stochastic behaviour of many stable output processes. The *actual* values, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, of the vector of interest in a random sample of production items may be assumed to be exchangeable, *i.e.*, a random sample from some fixed underlying distribution, and the *observed* measurements $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are assumed to be an approximation, with finite precision, to the actual values $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.

Thus, we assume that, for any produced item,

$$\boldsymbol{y}_j = \boldsymbol{x}_j + \boldsymbol{\delta}_j$$

so that the actual attributes of interest, $\boldsymbol{x}_j$, are observed as $\boldsymbol{y}_j$ with a measurement error $\boldsymbol{\delta}_j$; we will further assume that the measurement errors are independently distributed according

to some probability distribution $p(\boldsymbol{\delta}_j \mid \boldsymbol{\alpha})$, indexed by a hyperparameter $\boldsymbol{\alpha}$ associated to the measurement procedure, and such that $E[\boldsymbol{\delta}_j \mid \boldsymbol{\alpha}] = 0$. Moreover, we assume that

$$\boldsymbol{x}_j = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_j,$$

so that the production process is centered at $\boldsymbol{\mu}$, and the $\boldsymbol{\varepsilon}_j$'s describe the "common" causes of variation (see, *e.g.*, Deming, 1986); we further assume that the $\boldsymbol{\varepsilon}_j$'s are independently distributed according to some probability distribution $p(\boldsymbol{\varepsilon}_j \mid \boldsymbol{\beta})$, indexed by a hyperparameter $\boldsymbol{\beta}$ and such that $E[\boldsymbol{\varepsilon}_j \mid \boldsymbol{\beta}] = 0$. Finally, to complete the specification of the model, and hence to be able to compute the capability index, we have to specify a joint prior probability distribution $p(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ describing the information available on the values of the hyperparameters. Very often, however, either no reliable information is available on those values, or the engineers prefer not to use such information, so that the capability index only depends on the observed data; indeed, the supplier is often bound by contractual agreement to "pass" some "test" which *only* depends on the collected sample. The situation of no relevant prior may be mathematically described by the appropriate *reference prior distribution* (Bernardo, 1979, Berger and Bernardo, 1992, Bernardo and Smith, 1994, Ch. 5), which will be denoted by $\pi(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta})$

In terms of the established notation, the Bayesian capability index, $C_b(D)$, of the process may then be written as

$$C_b(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) = \frac{1}{3}\Phi^{-1}\left\{\int_A p(\boldsymbol{x} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)\, d\boldsymbol{x}\right\}, \tag{11}$$

where A is the *tolerance region*, *i.e.*, the region of acceptable attributes, as defined by the engineering specifications, and where

$$p(\boldsymbol{x} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) = \int\int p(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\beta})\; p(\boldsymbol{\mu}, \boldsymbol{\beta} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)\; d\boldsymbol{\mu} d\boldsymbol{\beta}$$

$$p(\boldsymbol{\mu}, \boldsymbol{\beta} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) = \int p(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)\; d\boldsymbol{\alpha}$$

$$p(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) \propto \prod_{j=1}^{n} p(\boldsymbol{y}_j \mid \boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta})\; \pi(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta})$$

$$p(\boldsymbol{y}_j \mid \boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \int p(\boldsymbol{y}_j \mid \boldsymbol{x}_j, \boldsymbol{\alpha})\; p(\boldsymbol{x}_j \mid \boldsymbol{\mu}, \boldsymbol{\beta})\; d\boldsymbol{x}_j$$

The equations above outline the evaluation of the Bayesian capability index (11) under very general conditions. Under simplifying assumptions, the integrations involved may be solved in closed form, and the Bayesian capability index may then be analytically obtained. We conclude this section by deriving an analytic expression for (11) when, as typically assumed in industrial practice, (i) *observation errors are negligible* compared to data variation, and (ii), observations may plausibly be regarded as a *random sample from a normal multivariate distribution*.

If there are no appreciable observation errors, so that $\boldsymbol{y}_j \approx \boldsymbol{x}_j$, the data $D = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ will consist of a random sample from a multivariate normal distribution of, say, dimension k, so that $p(\boldsymbol{x}_j \mid \boldsymbol{\mu}, \boldsymbol{\beta}) = \mathrm{N}_k(\boldsymbol{x}_j \mid \boldsymbol{\mu}, \Sigma)$. If no prior information about $\boldsymbol{\mu}$ and Σ is available, then the predictive distribution of a future observation which corresponds to the conventional "non-informative" default prior

$$\pi(\boldsymbol{\mu}, \Sigma) \propto |\Sigma|^{-(k+1)/2}$$

is given (see *e.g.*, Geisser, 1993, p. 192) by the k-variate Student density with $n-k$ degrees of freedom

$$p(\boldsymbol{x} \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \propto \left[1 + \frac{1}{n-1}\frac{n}{n+1}(\boldsymbol{x} - \overline{\boldsymbol{x}})^t S^{-1} (\boldsymbol{x} - \overline{\boldsymbol{x}})\right]^{-n/2}, \tag{12}$$

where, $n\overline{\boldsymbol{x}} = \sum_i \boldsymbol{x}_i$, and $(n-1)S = \sum_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^t$. Therefore, the Bayesian capability index $C_b(D)$ is given by

$$C_b(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \frac{1}{3}\Phi^{-1}\{\Pr(\boldsymbol{x} \in A \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)\}$$

where A is the tolerance region, which is often —but not always— of the form

$$\{\boldsymbol{x};\ l_{0i} \le x_i \le l_{1i}, \quad i = 1, \ldots, k\}.$$

The desired predictive probability $\Pr(\boldsymbol{x} \in A \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ may then be obtained by numerical integration with the Student density (12).

In particular, with univariate observations, the predictive distribution (12) reduces to the univariate Student density with $n-1$ degrees of freedom

$$p(x \,|\, x_1, \ldots, x_n) \propto \left\{1 + \frac{1}{n-1}\frac{n}{n+1}\left(\frac{x - \bar{x}}{s}\right)^2\right\}^{-n/2}, \tag{13}$$

where, $n\bar{x} = \sum_i x_i$, and $(n-1)s^2 = \sum_i (x_i - \bar{x})^2$. We have thus proved the following result:

Theorem. *With no observation errors, conditionally independent univariate normal data* $\mathrm{N}(x \,|\, \mu, \sigma)$*, and default prior* $\pi(\mu, \sigma) \propto 1/\sigma$*, the Bayesian capability index is*

$$C_b(x_1, \ldots, x_n) = C_b(n, \bar{x}, s) = \frac{1}{3}\Phi^{-1}\left\{F_{n-1}\left(\frac{l_1 - \bar{x}}{s\sqrt{\frac{n+1}{n}}}\right) - F_{n-1}\left(\frac{l_0 - \bar{x}}{s\sqrt{\frac{n+1}{n}}}\right)\right\} \tag{14}$$

where F_{n-1} *is the distribution function of a (univariate) standard Student* t *density with* $n-1$ *degrees of freedom,* $\bar{x} = \sum x_i/n$ *and* $s^2 = \sum (x_i - \bar{x})^2/(n-1)$.

As suggested by a referee, it may be proved that, for *any* sample, the value of $C_b(n, \bar{x}, s)$ given above is *smaller* than the value of $\hat{C}_{pk}(\bar{x}, s)$ given by (2). In fact, it may be established that, for fixed $\bar{x}$ and s, $C_b(n, \bar{x}, s)$ increases with n and, as $n \to \infty$, either tends to $\hat{C}_{pk}(\bar{x}, s)$, if the process is non-centered and capable, or to some quantity smaller than $\hat{C}_{pk}(\bar{x}, s)$ if it is not. This is a mathematical consequence of the fact that Student densities have lighter tails as their degrees of freedom increase and tend to normal tails as their degrees of freedom tend to infinity. This provides an interesting example of the way in which Bayesian prediction automatically takes into account the residual uncertainty about the parameters left after any sample. Indeed, while two samples of different sizes but the same $\bar{x}$ and s would produce identical $\hat{C}_{pk}$ values, (a rather dubious property!), the corresponding Bayesian indices will be different, with the smaller index corresponding to the smaller sample, reflecting a smaller confidence in producing conforming items.

We claim that in the frequent applications where the simplifying assumptions discussed above may apply (stable univariate normal process with no appreciable measurement errors), one should routinely replace (2) by (14). This simultaneously corrects the two more obvious pitfalls of the conventional index, by (i) automatically taking into account both extremes when necessary, and (ii) appropriately accounting for the actual (finite!) sample size.

We have derived an analytical expression for $C_b(D)$ under some simplifying assumptions. In more complex problems, however, the value of (11) cannot be analytically obtained without strong, often unwarranted, distributional assumptions. Today's ubiquitous numerical approximations then become necessary. The hierarchical structure of the model suggested facilitates the implementation of a Markov chain Monte Carlo algorithm.

5. EXAMPLES

To study the behaviour of the proposed methodology under different sets of assumptions, we simulated 1000 samples of size 100 from a normal distribution $N(x \mid 10, 1)$, which we have systematically used through the examples.

5.1. *Normal data, capable process*

The engineering specifications are assumed to be $l_0 = 5$ and $l_1 = 13$, so that the tolerance region is $A = \{x; 5 \leq x \leq 13\}$, with a negligible probability of non-conforming items in the left tail. The conditional C_{pk} value is

$$C_{pk}(\mu, \sigma) = \min\left\{\frac{10-5}{3}, \frac{13-10}{3}\right\} = 1.000,$$

and the conditional probability of an item being conforming is

$$P(A \mid \mu, \sigma) = \int_5^{13} \mathrm{N}(x \mid 10, 1)\, dx = 0.99865.$$

The results in Section 3 indicate that, in those conditions, one would expect an approximate numerical agreement between the conventional and the Bayesian indices.

Table 1. *1000 samples of size 100 from* $N(x \mid 10, 1)$*;* $l_0 = 5$ *and* $l_1 = 13$*.*

	$\hat{C}_{pk}(D)$	$P(x \in A \mid D)$	$C_b(D)$
Mean	0.98367	0.99754	0.95543
StDev.	0.07297	0.00149	0.06792

Table 1 reproduces the mean and standard deviation of the C_{pk}'s, the predictive probabilities of conforming items, and the Bayesian indices, which correspond to the 1000 simulated samples. We note that, although the results are not qualitatively different, the Bayesian index is appreciably smaller; as pointed out by a referee, a frequentist statistician may like to check that $C_b(n, \bar{x}, s)$ will typically be "significantly" smaller that $\hat{C}_{pk}(\bar{x}, s)$. As mentioned before, this is due to the thicker tails of the Student densities, and reflects the consequences on inferences about the distribution tails of the residual uncertainty about the parameters.

Table 2. *1000 samples of size 10 from* $N(x \mid 10, 1)$*;* $l_0 = 5$ *and* $l_1 = 13$*.*

	$\hat{C}_{pk}(D)$	$P(x \in A \mid D)$	$C_b(D)$
Mean	1.09220	0.98137	0.77691
StDev.	0.43232	0.02396	0.17385

As expected, this tail effect is larger as the sample decreases. Table 2 shows the results with 1000 subsamples of size 10; those show $\hat{C}_{pk}$ values which are, on average, 40% higher than what we claim they should be.

5.2. *Normal data, non-capable process*

The tolerance region is assumed to be $\{x; 7 \leq x \leq 13\}$, so that the probability of non-conforming items is now the same in both tails. The conditional $C_{pk}(\mu, \sigma)$ value is still 1.000, for the conventional index can only take one tail into account, but the conditional probability of an item being conforming is now $P(A \mid \mu, \sigma) = 0.99730$. Here, the conventional approach, by ignoring one of the tails, is bound to give too large a value.

Table 3. *1000 samples of size 100 from* $N(x \mid 10, 1)$*;* $l_0 = 7$ *and* $l_1 = 13$*.*

	$\hat{C}_{pk}(D)$	$P(x \in A \mid D)$	$C_b(D)$
Mean	0.96809	0.99554	0.88747
StDev.	0.06761	0.00251	0.06521

Table 3 reproduces the results which correspond to the 1000 simulated samples of size 100. We note that the $\hat{C}_{pk}$'s are, on average, 9% higher than what we claim they should be. Again, the difference would dramatically increase as the sample decreases.

5.3. *Log-Normal data, two-sided tolerance region*

To study the behaviour of the proposed methodology with skewed distributions, we use again the 1000 samples of size 100 from a normal distribution $N(x \mid 10, 1)$ and made the exponential transformation $z = \exp(x/2)$, to obtain 1000 samples of size 100 from a log-normal distribution with mean $E[z] = 168.17$ and standard deviation $D[z] = 89.62$ (see *e.g.*, Johnson and Kotz, 1970, Ch. 14). We further assume the engineering specifications to be $l_0 = \exp(5/2) = 12.182$ and $l_1 = \exp(13/2) = 665.14$ *i.e.*, we chose as tolerance region the image by $z = f(x) = \exp(x/2)$ of the tolerance region used in Example 5.1. The conditional value of the (untransformed) conventional index is now

$$C_{pk}(\mu, \sigma) = \min \left\{ \frac{168.17 - 12.182}{3 * 89.62} , \frac{665.14 - 168.17}{3 * 89.62} \right\} = 0.58018.$$

However the conditional probability of an item being conforming remains invariant:

$$P(f(A) \mid \mu, \sigma) = \int_{f(A)} \text{LogN}(z \mid 10, 1)\, dz = \int_5^{13} \text{N}(x \mid 10, 1)\, dx = 0.99865. \qquad (15)$$

and, more generally, the Bayesian analysis will remain precisely the same.

Table 4. *1000 samples of size 100 from LogN*$(x \mid 10, 1)$*;* $l_0 = \exp(5/2)$ *and* $l_1 = \exp(13/2)$.

	$\hat{C}_{pk}(D)$	$P(x \in A \mid D)$	$C_b(D)$
Mean	0.58544	0.99754	0.95543
StDev.	0.04673	0.00149	0.06792

Table 4 reproduces the results which correspond to the 1000 simulated samples of size 100. The average $\hat{C}_{pk}$ is 61% smaller than the average C_b; this illustrates the fact that if the conventional analysis is routinely applied to skewed distributions *without a previous normalizing transformation*, one may conclude that the process is far worse than it really is. We now show that the reverse may also happen.

5.4. *Log-Normal data, one-sided tolerance region*

We now assume the engineering specifications to be $l_0 = 0$ and $l_1 = \exp(13/2) = 665.14$, *i.e.*, we chose as tolerance region the image by $z = f(x) = \exp(x/2)$ of the one-sided tolerance region $A = \{x; x < 13\}$. The conditional value of the conventional index is now

$$C_{pk}(\mu, \sigma) = \left\{ \frac{665.14 - 168.17}{3 * 89.62} \right\} = 1.84843,$$

while conditional probability of an item being conforming will be

$$P(z \in f(A) \mid \mu, \sigma) = \int_{-\infty}^{13} \mathrm{N}(x \mid 10, 1)\, dx = 0.99865,$$

numerically equal to (15).

Table 5. *1000 samples of size 100 from LogN*$(x \mid 10, 1)$*;* $l_0 = 0$ *and* $l_1 = \exp(13/2)$.

	$\hat{C}_{pk}(D)$	$P(z \in f(A) \mid D)$	$C_b(D)$
Mean	1.81185	0.99754	0.95543
StDev.	0.21506	0.00149	0.06792

Table 5 reproduces the results which correspond to the 1000 simulated samples of size 100. The average $\hat{C}_{pk}$ is now 89% larger than the average C_b. This further illustrates the fact that conventional analysis routinely applied to *non-transformed* skewed distributions may be grossly misleading.

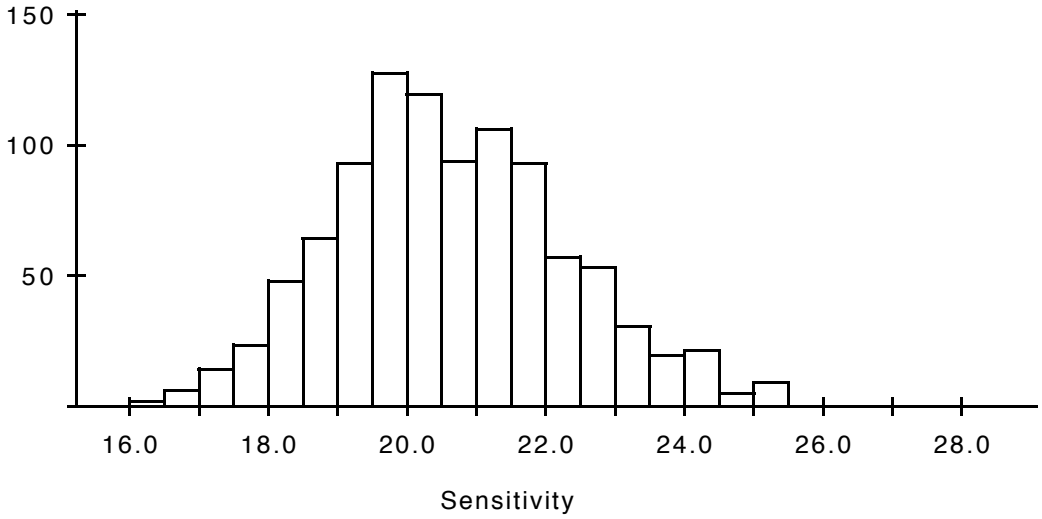

Figure 1. *Histogram of 1,000 actual sensitivity measures.*

5.5. *A real data example*

A supplier of safety parts for a motor company had a contract which specified that the $\hat{C}_{pk}$ of a certain sensitivity measure should be computed for samples of size $n = 1,000$, taken periodically from the production process, and that the capability index should be larger than 1.33.

Figure 1 is a histogram of one of those samples. The mean is 20.65 and the standard deviation is 1.72. The upper specification limit was established at 28 and there was no lower limit. The corresponding estimated $\hat{C}_{pk}$ was found to be 1.42, well above 1.33. However, the distribution is moderately skewed. We verified that the sample could better be treated as a random sample from a lognormal distribution; specifically, $y = 10 * \log(x) + 100$ is approximately normal $\mathrm{N}(y \,|\, 130.27, 0.82)$. Using (14) with the transformed tolerance region, the corresponding Bayesian capability index is easily found to be 1.24, below the required 1.33 level.

Thus, if the motor company *really* requires the tiny probability of non-conforming items which corresponds to a C_{pk} value of 1.33 for normal samples, then they should not accept that batch. Alternatively, they could directly specify the proportion of non-conforming items they require, possibly by assessing the relevant utility functions, and leave a competent statistician verify whether or not this is achieved by their suppliers.

Table 6. *100 bootstrap resamples form the original sample.* $l_0 = -\infty$ *and* $l_1 = 28$.

	$\hat{C}_{pk}(D)$	$P(x \in A \,|\, D)$	$C_b(D)$
Mean	1.42298	0.99989	1.23540
StDev.	0.03573	0.00004	0.02948

To verify the robustness of those numbers, we obtained 100 bootstrap resamples from the original sample. The results, contained in Table 6, establish that, indeed, when the distributional form is taken into account, the 1.33 capability level is *not* attained.

6. DISCUSSION

We have proposed a new capability index, C_b, that (i) has solid decision-theoretical foundations, (ii) is defined for any type of, possibly multivariate, output process, (iii) may accommodate appropriately modelled observation errors and (iv) is invariant under monotone transformations of the quantities of interest.

We have established that the new index, $C_b(n, \bar{x}, s)$, contains the traditional $\hat{C}_{pk}(\bar{x}, s)$ estimated index as a limiting case when (i) data are univariate normal, (ii) there are no observation errors, (iii) the production process is non-centered and potentially capable, and (iv) the sample size is large. However, we claim that these conditions are often not met in practice and hence, that the use of $\hat{C}_{pk}$ as an indicator of the performance of the process with respect to the engineering specifications may be misleading. When, as it is often the case, data are indeed univariate normal and observation errors are negligible compared with the standard deviation of the process, then equation (14) provides a useful replacement for the conventional $\hat{C}_{pk}(\bar{x}, s)$, and may safely be used with small samples.

We have used numerical examples to illustrate the limitations of the conventional approach. In particular, we have illustrated the fact that the tail behaviour of the production process is very important in this problem. In a sense, Examples 5.3 to 5.5 may be regarded as unfair to $\hat{C}_{pk}$ in that they compare the Bayesian index with the untransformed conventional index; if one calculates the $\hat{C}_{pk}$'s for the transformed data, or uses one of the *ad hoc* "modified" forms of the conventional index, one reproduces the situation discussed in Examples 5.1 and 5.2. However, it is a fact that, with histograms such as that of Figure 1, many practitioners would be tempted to go ahead and quote the conventional, untransformed, $\hat{C}_{pk}$; Examples 5.3 to 5.5 are intended to warn against such temptation. Indeed, by the form it is defined, directly in terms of the sample mean and the sample standard deviation, $\hat{C}_{pk}(\bar{x}, s)$ invites routine, often misleading use. By comparison, the *definition* of the Bayesian index forces the analysis of the tail behaviour — numerically or analytically— and is, therefore, less likely to be misused. In fact, one should ideally perform some model sensitivity analysis to check the robustness of the results to changes in the tail behaviour. However, this will require the derivation of the exact expression of $C_b(D)$ for the set of models entertained, which is not a trivial task.

The analytic results (12), (13) and (14) have been obtained using the conventional "non-informative" prior $\pi(\boldsymbol{\mu}, \Sigma) \propto |\Sigma|^{-(k+1)/2}$, which reduces to $\pi(\mu, \sigma) \propto \sigma^{-1}$ in the one-dimensional case. This prior is not necessarily the more appropriate 'default' prior if the parameter of interest is of the form

$$\theta = \Phi\left(\frac{l_1 - \mu}{\sigma}\right) - \Phi\left(\frac{l_0 - \mu}{\sigma}\right)$$

as in the present case. Preliminary results show that the corresponding *reference* prior (Bernardo, 1979; Berger and Bernardo, 1992), $\pi_\theta(\mu, \sigma)$, which should be used obtain the reference posterior distribution of θ and, presumably, the reference posterior predictive distribution of any function of the data whose sampling distribution only depends on the parameters through θ, is *not* the conventional $\pi(\mu, \sigma) \propto 1/\sigma$. The consequences of such modification of the prior, which will only be appreciable for small samples, are presently being analyzed.

A number of other important problems remain to be studied. In particular, (i) exact expressions for the Bayesian index $C_b(D)$ could be obtained for non-normal standard models, incorporating —when appropriate— relevant assumptions about the measurement errors; (ii) density estimation techniques specially oriented towards tail behaviour analysis could be used to provide numerical approximations to the Bayesian index; and (iii) using dynamic modelling,

it should be possible to generalize the ideas presented to non stable processes. All this is left for the near future.

ACKNOWLEDGEMENT

The authors would like to thank Professor Dennis Lindley and two anonymous referees for their constructive comments.

REFERENCES

Aitchison, J. and Dunsmore, I. R. (1975). *Statistical Prediction Analysis*. Cambridge: University Press.

Berger, J. O. and Bernardo, J. M. (1992). On the development of the reference prior method. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 35–60 (with discussion).

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. B* **41**, 113–147 (with discussion).

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. Chichester: Wiley

Bisell, A. F. (1990). How reliable is your capability index? *Appl. Statist.* **39**, 331-340.

Boyles, R. A. (1991). The Taguchi capability index. *J. Qual. Technol.* **23**, 17–26.

Carr, W. E. (1991). A new process capability index: parts per million. *Quality Progress* **24**, 152.

Chang, L. K., Cheng, S. W. and Spiring, F. A. (1988). A new measure of process capability: C_{pm}. *J. Quality Technology* **20**, 162–175.

Chang, L. K., Cheng, S. W. and Spiring, F. A. (1991). A multivariate measure of process capability. *Internat. J. Modelling and Simulation* **11**, 1–6.

Cheng, S. W. and Spiring, F. A. (1989). Assessing process capability: a Bayesian approach. *IIE Transactions* **21**, 97–98.

Chou, Y. M. and Owen, D. B. (1989). On the distribution of the estimated process capability indices. *Comm. Statist. Theory and Methods* **18**, 4549–4560.

Deming, E. W. (1986). *Out of the Crisis*. Cambridge, MA: The MIT Press.

Fishburn, P. C. (1981). Subjective expected utilities: a review of normative theories. *Theory and Decision* **13**, 139–199.

Efron, B. (1982). *The Jacknife, the Bootstrap and other Resampling Plans*. Philadelphia, PA: SIAM.

Geisser, S.(1993). *Predictive Inference: An Introduction*. London: Chapman and Hall.

Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions*. New York: Wiley.

Kane, V. E. (1986). Process capability indices. *J. Quality Technology* **18**, 41–52.

Kotz, S. and Johnson, N. L. (1993). *Process Capability Indices*. London: Chapman and Hall.

Kotz, S., Pearn, W. L. and Johnson, N. L. (1993). Some process capability indices are more reliable than one may think. *Appl. Statist.* **42**, 55–62.

Pearn, W. L. and Chen K. S. (1996). A Bayesian-like estimator of C_{pk}. *Comm. Statist. Theory and Methods* , (to appear).

Pearn, W. L., Kotz, S. and Johnson, N. L. (1992). Distributional and inferential properties of process capability indices. *J. Qual. Technol.* **24**, 216–233.

Singpurwalla, N. D. (1992). A Bayesian perspective on Taguchi's approach to quality engineering and tolerance design. *IIE Transactions* **24**, 18–32 (with discussion).

Taam, W, Subbaiah, P, and Liddy, J. W. (1993). A note on multivariate capability indices. *J. Applied Statistics* **20**, 339–351.

Wierda, S. J. (1993). A multivariate process capability index. *Trans ASQC Quality Congress*, 342–348.

Zhang, N. F., Stenback, G. A. and Wardrop, D. M. (1990). Interval estimation of the process capability index C_{pk}. *Comm. Statist. Theory and Methods* **19**, 4455–4470.