Revista Colombiana de Estadística
Volumen 30 No. 2. pp. 301 a 316. Diciembre 2007

Comparison of Process Capability Indices under Autocorrelated Data

Comparación de índices de capacidad de procesos con datos autocorrelacionados

Rubén Darío Guevara[1,a], José Alberto Vargas[2,b]

[1]Programa de Ciencias Básicas, Universidad de Ciencias Aplicadas y Ambientales, Bogotá, Colombia

[2]Departamento de Estadística, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia

Abstract

The process capability indices provide a measure of how a process fits within the specification limits. In calculating indices is usual to assume that the process data are independent. However, in industrial applications data are often autocorrelated. This paper deals with the indices C_p, C_{pk}, C_{pm} and C_{pmk} when data are autocorrelated. Variances for their estimators are derived and coverage probabilities of some confidence intervals are calculated.

***Key words*:** Autocorrelation, Process analysis, Estimation, Process capability indices, SPC.

Resumen

Los índices de capacidad de un proceso suministran una información numérica acerca de cómo el proceso se ajusta a unos límites de especificación. En el cálculo de estos índices se asume que las observaciones son independientes; sin embargo, en aplicaciones industriales frecuentemente los datos están autocorrelacionados. Este artículo analiza los índices C_p, C_{pk}, C_{pm} y C_{pmk} cuando los datos presentan autocorrelación, se encuentran las varianzas para sus estimadores cuando los procesos son gaussianos y se calculan los porcentajes de cobertura para algunos intervalos de confianza.

***Palabras clave*:** autocorrelación, análisis de procesos, estimación, índice de capacidad de procesos, SPC.

[a]Profesor. E-mail: r.guevara@udca.edu.co
[b]Profesor titular. E-mail: javargasn@unal.edu.co

1. Introduction

Process capability analysis is an important tool, widely used in industrial quality improvement programs. A basic assumption in process capability analysis is that process data are independent and identically normally distributed. However, in several industrial processes, data exhibit some degree of autocorrelation. Though the literature on process capability indices is voluminous, there is not significant research when data are autocorrelated, as can be observed in Kotz & Johnson (2002) and in Spiring et al. (2003). Zhang (1998) studied the indices C_p and C_{pk} for autocorrelated data. We extend Zhang's study to the capability indices C_{pm} and C_{pmk} and compare the four indices C_p, C_{pk}, C_{pm} and C_{pmk} for stationary gaussian processes. Additionally, some results have been taken from Guevara (2005) in his master's thesis. In the first part of this article we introduce the definition of the capability indices C_p, C_{pk}, C_{pm} and C_{pmk}. Subsequently, we present: 1) some relevant terminology about stationary gaussian processes, and 2) the expected values and variances of $\overline{X}$, S^2 and S in these processes. In Section 4, we calculate the variances of the estimators of C_{pm} and C_{pmk}, in stationary gaussian processes. Section 5, shows why the autocorrelation structure of process data, when it is present, should not be ignored for calculating any of the four indices. This is carried out through a simulation study with $AR(1)$ processes. In Section 6, inferential procedures are performed, by measuring the coverage probability of confidence intervals constructed for some of the indices studied. Finally, some conclusions are given.

2. Process Capability Indices

Among the several capability indices defined in the literature, the most extensively used in the industry are C_p, C_{pk}, C_{pm} and C_{pmk} , defined as follows:

$$C_p = \frac{USL - LSL}{6\sigma}$$

where USL and LSL are the upper and lower specification limits respectively and σ is the process standard deviation.

$$C_{pk} = \min\left\{\frac{USL-\mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right\} = \frac{\left[a - |\mu - b|\right]}{3\sigma} = \frac{\left[d - |2\mu - m|\right]}{6\sigma}$$

where μ is the process mean, $a = (USL - LSL)/2$, $b = (USL + LSL)/2$, $d = USL - LSL$ and $m = (USL + LSL)$.

$$C_{pm} = \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}} = \frac{C_p}{\sqrt{1+\xi^2}}$$

where T is the target value and $\xi = \frac{\mu - T}{\sigma}$.

$$C_{pmk} = \min\left\{\frac{USL-\mu}{3\sqrt{\sigma^2+(\mu-T)^2}}, \frac{\mu-LSL}{3\sqrt{\sigma^2+(\mu-T)^2}}\right\} = \frac{C_{pk}}{\sqrt{1+\xi^2}} = \frac{a-|\mu-b|}{3\sqrt{\sigma^2+(\mu-T)^2}}$$

3. Stationary Gaussian Processes

If $\{X_t\}$ is a process such that $Var(X_t) < \infty$ for each $t \in W \subset R$, where R is the set of real numbers, then the autocovariance function $\gamma_X(\cdot,\cdot)$ of $\{X_t\}$ is:

$$\gamma_X(r,s) = Cov(X_r, X_s) = E\big[(X_r - EX_r)(X_s - EX_s)\big], \quad r,s \in W$$

The time series $\{X_t, t \in Z\}$, with index set $\mathbb{Z}$, is said to be stationary if $E|X_t|^2 < \infty$, $\forall t \in \mathbb{Z}$, $E(X_t) = m$, $\forall t \in \mathbb{Z}$ and $\gamma_X(r,s) = \gamma_X(r+t, s+t)$, $\forall r,s,t \in \mathbb{Z}$; therefore $\mu_{X(t)}$ is independent of t, and $\gamma_X(t+h,t)$ is independent of t for each $h \in \mathbb{Z}$.

If $\{X_t\}$ is a stationary process, the autocovariance function (ACVF) of $\{X_t\}$ is

$$\gamma_X(h) = Cov(X_{t+h}, X_t)$$

and the autocorrelation function (ACF) is given by

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Cor(X_{t+h}, X_t)$$

(see Brockwell & Davis 1996).

$\{X_t\}$ is a process gaussian if all of its joint distributions are multivariate normal. A process is said to be stationary gaussian if it is stationary and gaussian simultaneously (see Brockwell & Davis 1996).

Let $\{X_t\}$ be a stationary gaussian process. Let $\{X_1, X_2, \ldots, X_n\}$ be a sample of size n from the process $\{X_t\}$. Let $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$, and $S^2 = \frac{\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right]}{n-1}$ be the sample mean and the sample variance respectively. Zhang (1998) gives the expected values and variances of $\overline{X}$, S^2 and S:

$$E(\overline{X}) = \mu_x$$

$$Var(\overline{X}) = \frac{\sigma_X^2}{n} g(n, \rho_i) \tag{1}$$

$$E(S^2) = \sigma_X^2 f(n, \rho_i) \tag{2}$$

$$Var(S^2) = \frac{2\sigma_X^4}{(n-1)^2} F(n, \rho_i) \tag{3}$$

$$E(S) \approx \big[E(S^2)\big]^{1/2} = \sigma_X\big[f(n, \rho_i)\big]^{1/2} \tag{4}$$

and

$$Var(S) \approx \frac{Var\big(S^2\big)}{4E\big(S^2\big)} = \left[\frac{\frac{2\sigma_X^4}{(n-1)^2}F(n,\rho_i)}{4\sigma_X^2 f(n,\rho_i)}\right] = \sigma_X^2 \frac{F(n,\rho_i)}{2(n-1)^2 f(n,\rho_i)}$$

where $\rho_i = \rho_X(i)$, for $i = 1, \dots, n$, is the autocorrelation of X, at lag i,

$$f(n,\rho_i) = 1 - \frac{2}{n(n-1)}\sum_{i=1}^{n-1}(n-i)\rho_i$$

$$F(n,\rho_i) = n + 2\sum_{i=1}^{n-1}(n-i)\rho_i^2 + \frac{1}{n^2}\left[n + 2\sum_{i=1}^{n-1}(n-i)\rho_i\right]^2 - \frac{2}{n}\sum_{i=0}^{n-1}\sum_{j=0}^{n-i}(n-i-j)\rho_i\rho_j$$

and

$$g(n,\rho_i) = 1 + \frac{2}{n}\sum_{i=1}^{n-1}(n-i)\rho_i$$

4. Variances of Process Capability Indices Estimators

Let $\{X_t\}$ be a stationary gaussian process. Let $\{X_1, X_2, \dots, X_n\}$ be a sample of size n from $\{X_t\}$. The usual estimators of C_p, C_{pk}, C_{pm} y C_{pmk} are:

$$\widehat{C}_p = \frac{USL - LSL}{6S}$$

$$\widehat{C}_{pk} = \min\left\{\frac{USL - \overline{X}}{3S}, \frac{\overline{X} - LSL}{3S}\right\} = \frac{\big[a - |\overline{X} - b|\big]}{3S} = \frac{\big[d - |2\overline{X} - m|\big]}{6S}$$

$$\widehat{C}_{pm} = \frac{USL - LSL}{6\sqrt{S^2 + \big(\overline{X} - T\big)^2}} = \frac{\widehat{C}_p}{\sqrt{1 + \widehat{\xi}^2}}$$

where $\widehat{\xi} = \frac{\overline{X} - T}{S}$, and

$$\widehat{C}_{pmk} = \min\left\{\frac{USL - \overline{X}}{3\sqrt{S^2 + \big(\overline{X} - T\big)^2}}, \frac{\overline{X} - LSL}{3\sqrt{S^2 + \big(\overline{X} - T\big)^2}}\right\} = \frac{\widehat{C}_{pk}}{\sqrt{1 + \widehat{\xi}^2}} = \frac{a - |\overline{X} - b|}{3\sqrt{S^2 + \big(\overline{X} - T\big)^2}}$$

Zhang (1998) found the following approximations for the variances of $\widehat{C}_p$ and $\widehat{C}_{pk}$:

$$Var\big(\widehat{C}_p\big) \approx C_p^2 \frac{F(n,\rho_i)}{2(n-1)^2 f^3(n,\rho_i)}$$

and

$$Var\left(\widehat{C}_{pk}\right) \approx \frac{C_{pk}^2}{f(n,\rho_i)}\left[\frac{g(n,\rho_i)}{9nC_{pk}^2} + \frac{F(n,\rho_i)}{2(n-1)^2 f^2(n,\rho_i)}\right]$$

For the variances of $\widehat{C}_{pm}$ and $\widehat{C}_{pmk}$, we found the following approximations (see appendix A and B):

$$Var\left(\widehat{C}_{pm}\right) \approx C_p^2\left[\frac{\frac{2F(n,\rho_i)}{(n-1)^2} + \frac{4g(n,\rho_i)\xi^2}{n}}{4[f(n,\rho_i)+\xi^2]^3}\right]$$

and

$$Var\left(\widehat{C}_{pmk}\right) \approx C_{pk}^2\left[\frac{1}{f(n,p_i)+\xi^2}\right]\times \left\{\frac{F(n,p_i)}{2(n-1)^2[f(n,p_i)+\xi^2]^2} + \frac{g(n,p_i)}{9n}\left[\frac{1}{C_{pk}} + \frac{6\xi}{2[f(n,p_i)+\xi^2]}\right]^2\right\}$$

where $\xi = \frac{\mu - T}{\sigma}$. If $\mu = T$ then

$$Var\left(\widehat{C}_{pm}\right) \approx C_p^2\frac{F(n,\rho_i)}{2(n-1)^2 f^3(n,\rho_i)}$$

which equals the variance of $\widehat{C}_p$, and

$$\begin{aligned}Var\left[\widehat{C}_{pmk}\right] &\approx \frac{C_{pk}^2}{f(n,p_i)}\left[\frac{F(n,p_i)}{2(n-1)^2 f^2(n,p_i)} + \frac{4\sigma^2 g(n,p_i)}{n}\left[\frac{1}{a-|2\mu-b|}\right]^2\right]\\ &= \frac{C_{pk}^2}{f(n,p_i)}\left[\frac{F(n,p_i)}{2(n-1)^2 f^2(n,p_i)} + \frac{g(n,p_i)}{9nC_{pk}^2}\right]\end{aligned}$$

which equals the variance of $\widehat{C}_{pk}$.

When the observations are independent, the variances reduce to:

$$Var\left[\widehat{C}_{pm}\right] \approx C_{pm}^2\left[\frac{1}{1+\xi^2}\right]\left[\frac{1}{2(n-1)} + \frac{\xi^2}{n}\right]$$

and

$$Var\left[\widehat{C}_{pmk}\right] \approx C_{pmk}^2\left\{\frac{1}{2(n-1)[1+\xi^2]^2} + \frac{1}{9n}\left[\frac{1}{C_{pk}} + \frac{6\xi}{2[1+\xi^2]}\right]^2\right\}$$

If, simultaneously the observations are independent and $\mu = T$, we have that

$$Var\left[\widehat{C}_{pm}\right] \approx C_p^2\left[\frac{\frac{2}{(n-1)}}{4}\right] = \frac{C_p^2}{2(n-1)} \approx Var\left(\widehat{C}_p\right)$$

and

$$Var\big[\widehat{C}_{pmk}\big] \approx C_{pk}^2 \left\{ \frac{1}{2(n-1)} + \frac{1}{9nC_{pk}^2} \right\} \approx \mathrm{Var}\big(\widehat{C}_{pk}\big)$$

These last two results are equal to those found by Bissel (1990).

To compare the variances of these estimators, a simulation study was carried out for a first order stationary autorregressive process with parameter ϕ. For $C_p = 1.33$, $\xi = 0, 5, 10$, $\phi = 0.25, 0.50, 0.75$ and $n = 10, 20, \ldots, 200$, Figure 1 shows great variability of $Var\big(\widehat{C}_{pm}\big)$ for $n < 100$. Fixing C_p, n and ξ as $|\phi|$ increases, $Var\big(\widehat{C}_{pm}\big)$ increases. Similar results are obtained for $Var\big(\widehat{C}_{pmk}\big)$, substituting C_{pm} for C_{pmk} and C_p for C_{pk} as can be seen in Figure 2. Now, if ϕ and ξ are fixed values and n increases, $Var\big(\widehat{C}_{pm}\big)$ and $Var\big(\widehat{C}_{pmk}\big)$ decreases. For fixed values of ϕ and n, when ξ increases, this is, the target value is far away from the mean of the process, $Var\big(\widehat{C}_{pm}\big)$ and $Var\big(\widehat{C}_{pmk}\big)$ decreases.

5. Autocorrelation Effects on Process Capability Indices

We consider the example given in Shore (1997). A quality characteristic is normally distributed with mean 40 and standard deviation 7. The specification limits are $USL = 61$ and $LSL = 19$. Different target values are considered: 40, 41, 42, 45 and 50. We then compare two processes. A process with independent observations and a process with observations following an $AR(1)$ model, $X_t = X_{t-1} + e_t$, where $\{e_t\}$ is a series of uncorrelated errors, $e_t \sim N(0, \sigma_e^2)$ and $\sigma_e = 7$. For each process, the mean, the standard deviation and the capability indices C_p, C_{pk}, C_{pm} and C_{pmk} are calculated. We do not show the values of C_{pk} and C_{pmk} because $C_p = C_{pk}$ and $C_{pm} = C_{pmk}$ see Table 1.

We observe in Table 1 that the higher the autocorrelation level the lower the capability index value.

TABLE 1: Mean, standard deviation (STD), C_p and C_{pm} of a process not autocorrelated *vs.* a process following an $AR(1)$ model.

				C_{pm}					
$\lvert\phi\rvert$	Mean	STD	C_p	$*d = 0$	$*d = 1$	$*d = 2$	$*d = 3$	$*d = 4$	$*d = 5$
No auto	40	7.00	1.000	1.000	0.990	0.962	0.919	0.868	0.814
0.25	40	7.23	0.968	0.968	0.959	0.933	0.894	0.847	0.796
0.50	40	8.08	0.866	0.866	0.859	0.841	0.812	0.776	0.737
0.75	40	10.58	0.661	0.661	0.659	0.650	0.636	0.619	0.598

$*d = \mu - T$

Through a simulation study we analyze the effect of the autocorrelation in the expected value of the sample mean and in the expected value of the standard error. We generated 1000 samples from a no autocorrelated model and 1000 samples from an $AR(1)$ model for each of the following cases: $n = 15$, 50, 100, 200; $T = 40$, 41,

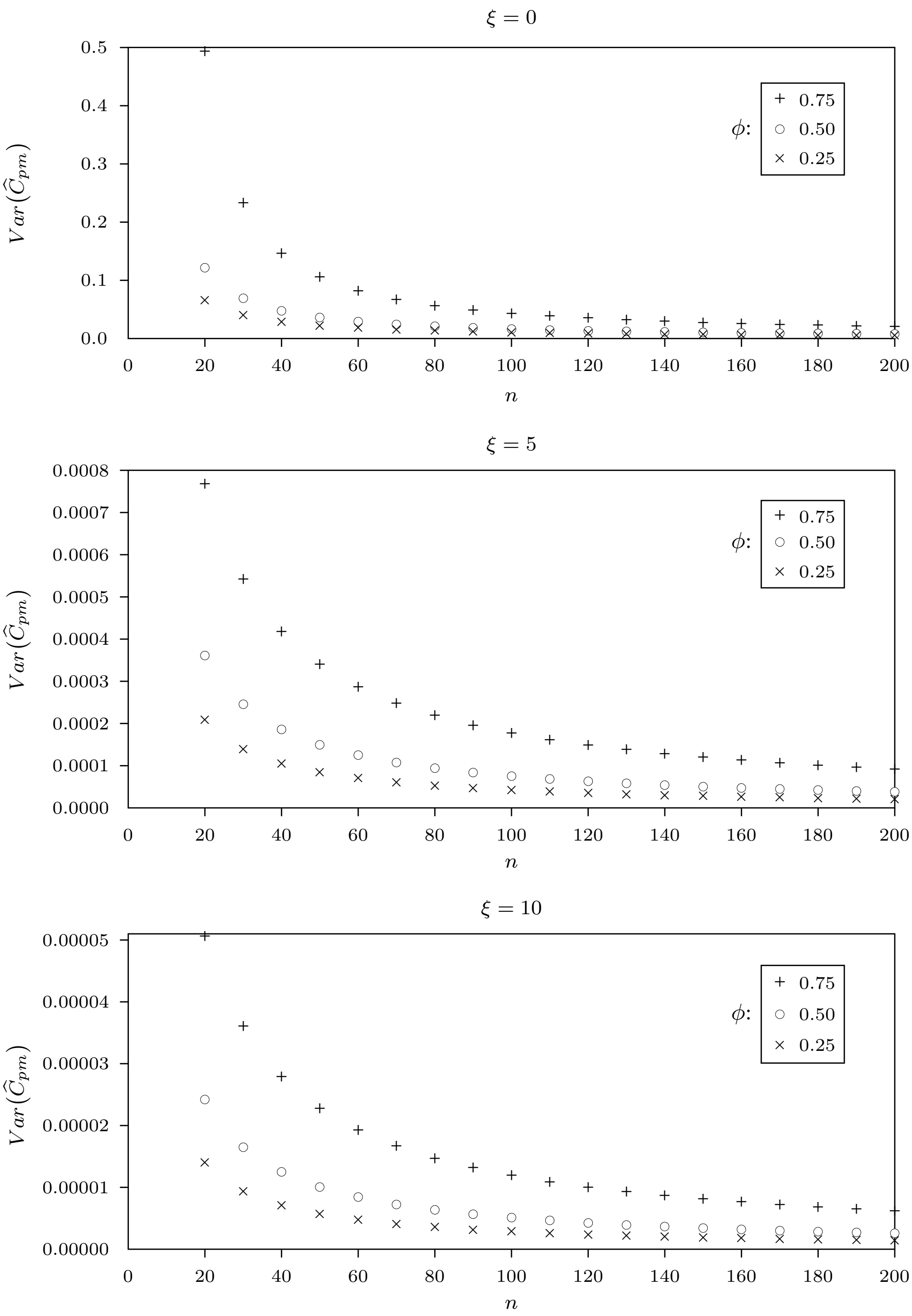

FIGURE 1: Variance of $\widehat{C}_{pm}$ in function of the sample size with $C_p = 1.33$, $\xi = 0, 5, 10$ and $\phi = 0.25, 0.50, 0.75$.

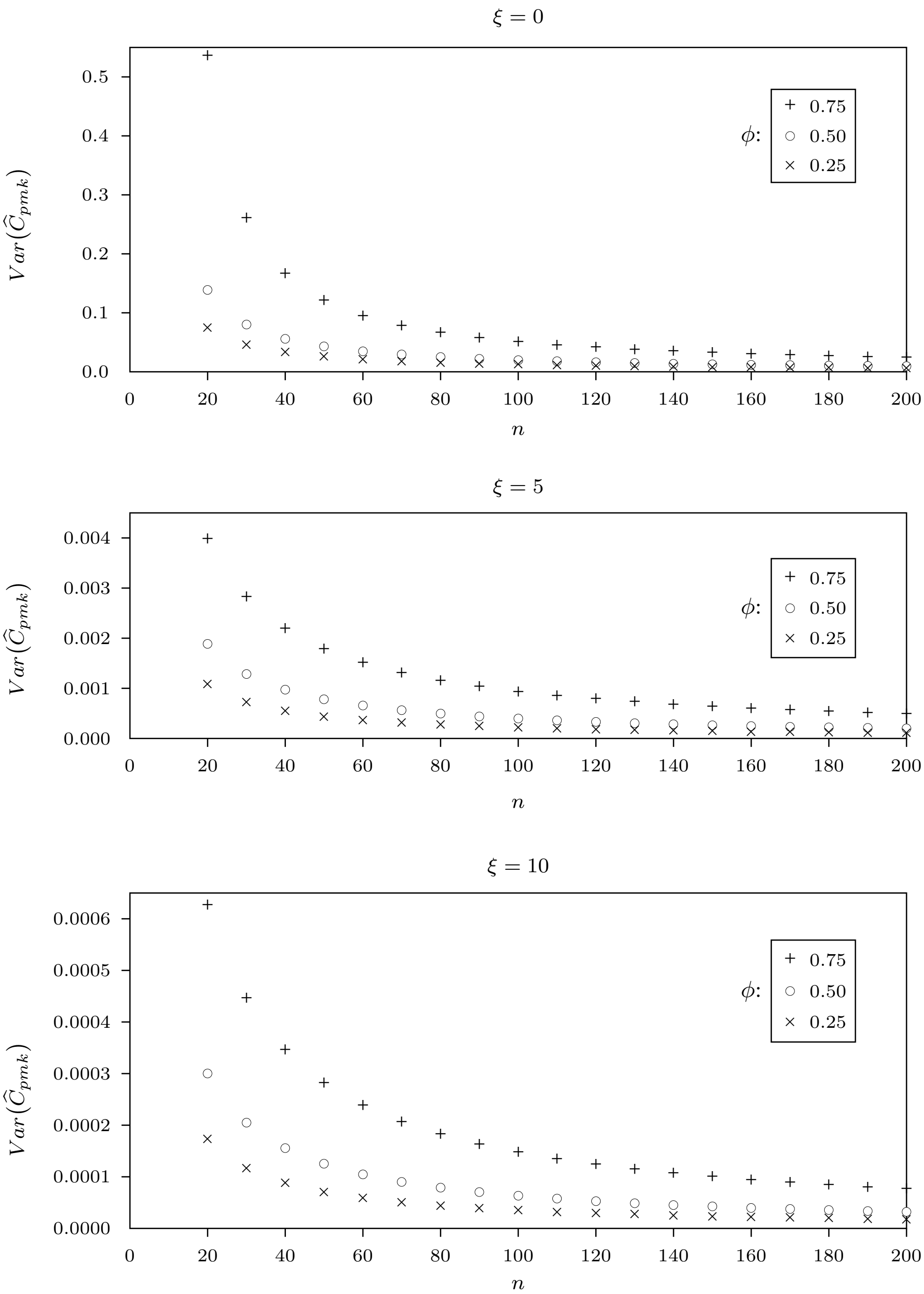

FIGURE 2: Variance of $\widehat{C}_{pmk}$ in function of the sample size with $C_{pk} = 1.33$, $\xi = 0, 5, 10$ and $\phi = 0.25, 0.50, 0.75$.

42, 43, 44, 45 and $\phi = -0.75, -0.50, -0.25, 0.25, 0.50, 0.75$. Table 2 shows partial results of these simulations.

TABLE 2: Expected values and standard errors of the sample mean and sample standard deviation for processes no autocorrelated and processes following an $AR(1)$ model.

			Mean		Std	
n	ϕ		No Auto	Auto	No Auto	Auto
15	−0.75	Average	39.90	39.98	6.88	10.38
		Std Error	1.76	1.15	1.32	2.03
	−0.50	Average	40.08	39.99	6.85	7.95
		Std Error	1.84	1.24	1.28	1.54
	−0.25	Average	40.15	40.02	6.88	7.13
		Std Error	1.86	1.44	1.33	1.34
	0.25	Average	40.03	39.96	6.79	7.05
		Std Error	1.82	2.41	1.34	1.36
	0.50	Average	39.98	40.14	6.94	7.95
		Std Error	1.82	3.52	1.32	1.53
	0.75	Average	40.06	40.36	6.79	10.53
		Std Error	1.85	6.33	1.26	1.99
50	−0.75	Average	39.98	40.01	6.96	10.51
		Std Error	0.99	0.59	0.71	1.09
	−0.50	Average	40.02	40.03	7.00	8.00
		Std Error	0.98	0.66	0.69	0.82
	−0.25	Average	40.01	40.00	6.97	7.18
		Std Error	0.96	0.80	0.72	0.72
	0.25	Average	39.99	40.04	6.98	7.18
		Std Error	0.99	1.32	0.70	0.71
	0.50	Average	40.02	39.97	6.99	8.08
		Std Error	1.00	1.93	0.70	0.84
	0.75	Average	40.02	40.19	6.95	10.49
		Std Error	1.00	3.73	0.70	1.04
100	−0.75	Average	40.00	40.00	7.01	10.54
		Std Error	0.68	0.41	0.50	0.76
	−0.50	Average	40.00	40.00	6.97	8.02
		Std Error	0.71	0.48	0.50	0.56
	−0.25	Average	40.02	40.03	6.97	7.23
		Std Error	0.69	0.55	0.48	0.52
	0.25	Average	40.00	39.97	6.99	7.23
		Std Error	0.70	0.95	0.51	0.51
	0.50	Average	39.99	40.01	6.98	8.06
		Std Error	0.70	1.34	0.50	0.58
	0.75	Average	39.99	39.97	6.97	10.57
		Std Error	0.69	2.77	0.48	0.75

Table 2 shows that the autocorrelation does not affect the expected value of the sample mean while a different situation occurs with the expected value of the standard error. Let us remember that $Var(X_t) = \frac{\sigma_e^2}{1-\phi^2}$ where σ_e^2 is the white noise variance. For example, for $n = 15$ and $\phi = -0.25$ the estimated expected value of the standard error is 7.13 in autocorrelated processes, for $\phi = -0.5$ is 7.95 and for $\phi = -0.75$ is 10.38. For independent observations the values are 6.88, 6.85 and

6.88 respectively. As n increases, the estimated expected value of the standard error increases slightly for autocorrelated data. For example, for $\phi = 0.25$ the estimated expected values for $n = 15, 50, 100$ are 7.05, 7.18 and 7.23 respectively.

Through other simulation study we analyze the performance of the capability indices estimators. Comparing the estimated expected values of the capability indices estimators shown in Table 3 with the theoretical values shown in Table 1, it can be observed that for autocorrelated processes the estimators are slightly biased, bias that decreases as n increases. For example, for $\phi = -0.75$ and $n = 15$, 50, 100 the expected values of $\widehat{C}_p$ are 0.703, 0.673 and 0.668 respectively, while the true value is 0.661. For $\phi = 0.25$ and $n = 15$, 50 y 100 the expected values of $\widehat{C}_{pk}$ are 0.937, 0.935 and 0.938 while the true value is 0.968. For $n = 15$ and $\phi = 0.50$, the expected values of $\widehat{C}_{pm}$ are 0.84, 0.84, 0.83 and 0.81 when $\mu - T = 0, 1, 2$ and 3 respectively, while the true values are 0.866, 0.859, 0.841 and 0.812.

6. Confidence Intervals for the Capability Indices under Stationary Gaussian Processes

Let us assume that $\mu = 50$, $USL = 3$ and $LSL = -3$. For sample sizes $n = 25$, 50, 100 and standard deviation values $\sigma = 0.5$, 1.0, 2.0, $AR(1)$ processes with normally distributed white noise are generated with $\phi = -0.75$, -0.50, -0.25, 0.25, 0.50 and 0.75. For each combination of n, ϕ and σ, 5000 random samples from a normal distribution were generated. For each sample, the values of $\overline{X}$, S, $\widehat{C}_p$, $\widehat{C}_{pk}$, $\widehat{C}_{pm}$, $\widehat{C}_{pmk}$, $\widehat{\sigma}_{c_p}$, $\widehat{\sigma}_{c_{pk}}$, $\widehat{\sigma}_{c_{pm}}$ and $\widehat{\sigma}_{c_{pmk}}$ are obtained. We calculated the proportion of times that the true index is contained in the interval

$$\widehat{capability\ index} \pm k(\widehat{\sigma}_{capability\ index})$$

The simulations showed a coverage probability for C_p of about 95% for $k = 2$ and 99% for $k = 3$. For C_{pk} the coverage probability was of around 90% for $k = 2$. Tables 4 and 5 show the coverage probabilities of the intervals for C_p and C_{pk}. The coverage probabilities for C_{pm} and C_{pmk} are not shown here because these were very low.

In Tables 4 and 5 we observe that as n increases, the coverage probabilities increase. In Table 5 it can be observed that coverage percentage for C_{pk} decreases when σ increases, being more evident for large values of ϕ. We observe that for $\phi = -0.75$ the coverage probabilities are very similar for different values of σ.

We also calculate through simulations the coverage probabilities of the confidence interval for C_{pm}, proposed by Boyles (1991). This interval is defined as follows:

$$\left(\widehat{C}_{pm}\sqrt{\chi^2_{\widehat{f},\alpha/2}\Big/\widehat{f}}, \widehat{C}_{pm}\sqrt{\chi^2_{\widehat{f},1-\alpha/2}\Big/\widehat{f}}\right)$$

where $\chi^2_{\widehat{f},\alpha}$ denotes the $100\alpha\%$ percentile of a chi-square distribution with $\widehat{f}$ degrees of freedom, for $\widehat{f} = n\big(1 + \widehat{\delta}\big)^2/\big(1 + 2\widehat{\delta}\big)$, $\widehat{\delta} = \big(\overline{X} - T\big)^2/\widehat{\sigma}^2$ and $\widehat{\sigma}^2 =$

$S^2(n-1)/n$. We did not find reliable results for C_{pm}, this is the coverage probabilities obtained are low.

TABLE 3: Effect of the autocorrelation in the expected values and standard errors of the capability indices for processes following an $AR(1)$ model.

			C_p		C_{pk}		C_{pm}									
							$\mu - T = 0$		$\mu - T = 1$		$\mu - T = 2$		$\mu - T = 3$		$\mu - T = 5$	
n	ϕ		No Auto	Auto	No Auto	Auto	No Auto	Auto	No Auto	Auto	No Auto	Auto	No Auto	Auto	No Auto	Auto
15	−0.75	Av	1.057	0.703	0.987	0.673	1.02	0.70	1.01	0.69	0.98	0.68	0.94	0.67	0.83	0.62
		St	0.200	0.213	0.206	0.206	0.19	0.21	0.19	0.21	0.18	0.20	0.17	0.18	0.14	0.15
	−0.50	Av	1.059	0.916	0.986	0.873	1.02	0.90	1.01	0.90	0.99	0.88	0.95	0.84	0.84	0.76
		St	0.200	0.236	0.206	0.228	0.19	0.23	0.19	0.22	0.18	0.21	0.17	0.19	0.14	0.15
	−0.25	Av	1.060	1.019	0.985	0.963	1.02	1.00	1.01	0.99	0.99	0.96	0.95	0.92	0.84	0.81
		St	0.200	0.241	0.206	0.232	0.19	0.23	0.19	0.22	0.18	0.21	0.17	0.19	0.15	0.14
	0.25	Av	1.074	1.031	1.000	0.937	1.03	0.98	1.02	0.97	1.00	0.95	0.95	0.91	0.84	0.82
		St	0.203	0.245	0.208	0.230	0.19	0.22	0.19	0.21	0.18	0.20	0.17	0.19	0.14	0.15
	0.50	Av	1.048	0.915	0.976	0.794	1.01	0.84	1.00	0.84	0.97	0.83	0.93	0.81	0.83	0.75
		St	0.198	0.237	0.204	0.218	0.19	0.20	0.19	0.20	0.18	0.19	0.17	0.18	0.14	0.16
	0.75	Av	1.070	0.690	0.995	0.526	1.03	0.61	1.02	0.60	0.99	0.60	0.95	0.60	0.84	0.58
		St	0.202	0.204	0.207	0.180	0.19	0.16	0.19	0.16	0.18	0.16	0.17	0.15	0.14	0.14
50	−0.75	Av	1.016	0.673	0.978	0.658	1.01	0.67	1.00	0.67	0.97	0.66	0.92	0.65	0.82	0.60
		St	0.103	0.138	0.110	0.135	0.10	0.14	0.10	0.14	0.10	0.13	0.09	0.12	0.08	0.10
	−0.50	Av	1.009	0.885	0.972	0.863	1.00	0.88	0.99	0.88	0.96	0.86	0.92	0.83	0.82	0.75
		St	0.102	0.137	0.109	0.135	0.10	0.14	0.10	0.13	0.10	0.13	0.09	0.11	0.08	0.09
	−0.25	Av	1.015	0.985	0.978	0.955	1.00	0.98	1.00	0.97	0.97	0.94	0.93	0.90	0.82	0.80
		St	0.102	0.132	0.110	0.130	0.10	0.13	0.10	0.13	0.10	0.12	0.09	0.10	0.08	0.08
	0.25	Av	1.013	0.985	0.975	0.935	1.00	0.97	0.99	0.96	0.96	0.94	0.92	0.90	0.82	0.80
		St	0.102	0.131	0.109	0.129	0.10	0.13	0.10	0.12	0.10	0.12	0.09	0.11	0.08	0.08
	0.50	Av	1.012	0.876	0.973	0.812	1.00	0.85	0.99	0.85	0.96	0.83	0.92	0.80	0.82	0.74
		St	0.102	0.137	0.109	0.134	0.10	0.13	0.10	0.13	0.10	0.12	0.09	0.11	0.08	0.09
	0.75	Av	1.018	0.674	0.979	0.578	1.01	0.64	1.00	0.64	0.97	0.63	0.93	0.62	0.82	0.59
		St	0.103	0.137	0.110	0.132	0.10	0.12	0.10	0.12	0.10	0.12	0.09	0.12	0.08	0.11
100	−0.75	Av	1.004	0.668	0.978	0.658	1.00	0.67	0.99	0.66	0.96	0.66	0.92	0.64	0.81	0.60
		St	0.071	0.104	0.077	0.103	0.07	0.10	0.07	0.10	0.07	0.10	0.06	0.09	0.05	0.08
	−0.50	Av	1.009	0.877	0.982	0.860	1.00	0.88	0.99	0.87	0.97	0.85	0.92	0.82	0.82	0.74
		St	0.072	0.100	0.077	0.098	0.07	0.10	0.07	0.10	0.07	0.09	0.06	0.08	0.05	0.06
	−0.25	Av	1.009	0.973	0.982	0.952	1.00	0.97	0.99	0.96	0.97	0.94	0.92	0.90	0.82	0.80
		St	0.072	0.094	0.077	0.093	0.07	0.09	0.07	0.09	0.07	0.08	0.06	0.08	0.05	0.05
	0.25	Av	1.007	0.973	0.980	0.938	1.00	0.97	0.99	0.96	0.96	0.93	0.92	0.89	0.82	0.80
		St	0.072	0.093	0.077	0.093	0.07	0.09	0.07	0.09	0.07	0.08	0.06	0.08	0.05	0.06
	0.50	Av	1.009	0.873	0.982	0.829	1.00	0.86	0.99	0.86	0.96	0.84	0.92	0.81	0.82	0.74
		St	0.072	0.098	0.077	0.099	0.07	0.10	0.07	0.09	0.07	0.09	0.06	0.08	0.05	0.07
	0.75	Av	1.009	0.666	0.983	0.597	1.00	0.65	0.99	0.64	0.97	0.64	0.92	0.62	0.82	0.59
		St	0.072	0.103	0.077	0.103	0.07	0.10	0.07	0.10	0.07	0.09	0.06	0.09	0.05	0.08

Av: Average
St: Std Error

TABLE 4: Estimation average coverage rate of C_p for $AR(1)$ processes using intervals $\widehat{C}_p \pm 2\sigma_{\widehat{C}_p}$.

	$\phi = 0.75$				$\phi = 0.25$				$\phi = -0.75$			
$n \setminus \sigma$	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
25	0.921	0.921	0.920	0.918	0.987	0.988	0.983	0.989	0.960	0.957	0.954	0.954
50	0.949	0.949	0.944	0.944	0.991	0.988	0.988	0.988	0.961	0.969	0.967	0.966
100	0.970	0.967	0.966	0.963	0.987	0.989	0.987	0.987	0.975	0.978	0.977	0.982

Table 5: Estimation average coverage rate of C_{pk} for $AR(1)$ processes using intervals $\widehat{C}_{pk} \pm 2\sigma_{\widehat{C}_{pk}}$.

	$\phi = 0.75$				$\phi = 0.25$				$\phi = -0.75$			
$n \setminus \sigma$	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
25	0.947	0.903	0.820	0.759	0.984	0.964	0.940	0.895	0.960	0.957	0.954	0.954
50	0.963	0.920	0.854	0.794	0.984	0.975	0.946	0.915	0.961	0.969	0.967	0.966
100	0.976	0.940	0.879	0.815	0.986	0.975	0.949	0.920	0.976	0.979	0.977	0.981

Finally, we constructed confidence intervals of the form

$$\left(\widehat{C}_{pm} - k_1\widehat{\sigma}_{C_{pm}}, \widehat{C}_{pm} + k_2\widehat{\sigma}_{C_{pm}}\right)$$

for C_{pm} and

$$\left(\widehat{C}_{pmk} - k_1\widehat{\sigma}_{C_{pmk}}, \widehat{C}_{pmk} + k_2\widehat{\sigma}_{C_{pmk}}\right)$$

for C_{pmk} for different values of k_1 and k_2. Table 6 presents some limited values of k_1 and k_2, which offer a coverage between 74% and 85% for the index C_{pm} for $n = 50$, $\phi = 0.50$, $\sigma = 1.5$ and $|\mu - T| = 5$ when $2 \leq k_1 \leq 3$ and $2 \leq k_2 \leq 3$. This coverage is greater or equal than 90% for $k_1 \geq 2.7$ and $k_2 \geq 3.7$. We did not find reliable results for C_{pmk}, this is, the coverage probabilities obtained are low, for example the coverage is between 33% and 48% for $n = 50$, $\phi = 0.50$, $\sigma = 1.5$ and $|\mu - T| = 5$ when $2 \leq k_1 \leq 3$ and $2 \leq k_2 \leq 3$, see Table 7. We believe that the low coverage probabilities are due to the bias of the estimators.

Table 6: Interval estimation average coverage rate of C_{pm} for $AR(1)$ processes with $\phi = 0.50$, $\sigma = 1.5$, $n = 50$.

		k_2																				
		2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
	2.0	74	74	75	75	77	78	78	80	80	80	82	84	84	85	86	86	86	87	88	88	89
	2.1	75	75	76	76	78	79	79	81	81	81	83	85	85	86	87	87	87	88	89	89	90
	2.2	75	75	76	76	78	79	79	81	81	81	83	85	85	86	87	87	87	88	89	89	90
	2.3	75	75	76	76	78	79	79	81	81	81	83	85	85	86	87	87	87	88	89	89	90
	2.4	75	75	76	76	78	79	79	81	81	81	83	85	85	86	87	87	87	88	89	89	90
	2.5	76	76	77	77	79	80	80	82	82	82	84	86	86	87	88	88	88	89	90	90	91
	2.6	76	76	77	77	79	80	80	82	82	82	84	86	86	87	88	88	88	89	90	90	91
	2.7	77	77	78	78	80	81	81	83	83	83	85	87	87	88	89	89	89	90	91	91	92
	2.8	77	77	78	78	80	81	81	83	83	83	85	87	87	88	89	89	89	90	91	91	92
	2.9	77	77	78	78	80	81	81	83	83	83	85	87	87	88	89	89	89	90	91	91	92
k_1	3.0	77	77	78	78	80	81	81	83	83	83	85	87	87	88	89	89	89	90	91	91	92
	3.1	77	77	78	78	80	81	81	83	83	83	85	87	87	88	89	89	89	90	91	91	92
	3.2	77	77	78	78	80	81	81	83	83	83	85	87	87	88	89	89	89	90	91	91	92
	3.3	77	77	78	78	80	81	81	83	83	83	85	87	87	88	89	89	89	90	91	91	92
	3.4	77	77	78	78	80	81	81	83	83	83	85	87	87	88	89	89	89	90	91	91	92
	3.5	77	77	78	78	80	81	81	83	83	83	85	87	87	88	89	89	89	90	91	91	92
	3.6	77	77	78	78	80	81	81	83	83	83	85	87	87	88	89	89	89	90	91	91	92
	3.7	78	78	79	79	81	82	82	84	84	84	86	88	88	89	90	90	90	91	92	92	93
	3.8	78	78	79	79	81	82	82	84	84	84	86	88	88	89	90	90	90	91	92	92	93
	3.9	78	78	79	79	81	82	82	84	84	84	86	88	88	89	90	90	90	91	92	92	93
	4.0	78	78	79	79	81	82	82	84	84	84	86	88	88	89	90	90	90	91	92	92	93

Table 7: Interval estimation average coverage rate of C_{pmk} for $AR(1)$ processes with $\phi = 0.50$, $\sigma = 1.5$, $n = 50$.

k_1 \ k_2	2.0	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	3.0	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9	4.0
2.0	33	33	35	36	37	38	39	41	41	42	44	44	44	45	46	46	46	46	46	49	49
2,1	34	34	36	37	38	39	40	42	42	43	45	45	45	46	47	47	47	47	47	50	50
2,2	34	34	36	37	38	39	40	42	42	43	45	45	45	46	47	47	47	47	47	50	50
2,3	34	34	36	37	38	39	40	42	42	43	45	45	45	46	47	47	47	47	47	50	50
2,4	35	35	37	38	39	40	41	43	43	44	46	46	46	47	48	48	48	48	48	51	51
2,5	35	35	37	38	39	40	41	43	43	44	46	46	46	47	48	48	48	48	48	51	51
2,6	35	35	37	38	39	40	41	43	43	44	46	46	46	47	48	48	48	48	48	51	51
2,7	36	36	38	39	40	41	42	44	44	45	47	47	47	48	49	49	49	49	49	52	52
2,8	37	37	39	40	41	42	43	45	45	46	48	48	48	49	50	50	50	50	50	53	53
2,9	37	37	39	40	41	42	43	45	45	46	48	48	48	49	50	50	50	50	50	53	53
3.0	37	37	39	40	41	42	43	45	45	46	48	48	48	49	50	50	50	50	50	53	53
3,1	38	38	40	41	42	43	44	46	46	47	49	49	49	50	51	51	51	51	51	54	54
3,2	38	38	40	41	42	43	44	46	46	47	49	49	49	50	51	51	51	51	51	54	54
3,3	38	38	40	41	42	43	44	46	46	47	49	49	49	50	51	51	51	51	51	54	54
3,4	38	38	40	41	42	43	44	46	46	47	49	49	49	50	51	51	51	51	51	54	54
3,5	40	40	42	43	44	45	46	48	48	49	51	51	51	52	53	53	53	53	53	56	56
3,6	40	40	42	43	44	45	46	48	48	49	51	51	51	52	53	53	53	53	53	56	56
3,7	41	41	43	44	45	46	47	49	49	50	52	52	52	53	54	54	54	54	54	57	57
3,8	41	41	43	44	45	46	47	49	49	50	52	52	52	53	54	54	54	54	54	57	57
3,9	42	42	44	45	46	47	48	50	50	51	53	53	53	54	55	55	55	55	55	58	58
4.0	42	42	44	45	46	47	48	50	50	51	53	53	53	54	55	55	55	55	55	58	58

7. Conclusions

We found approximations of the variances of $\widehat{C}_{pm}$ and $\widehat{C}_{pmk}$ for stationary gaussian processes. In particular, we show expressions for the variance of these estimators when the observations are independent, for $\mu = T$ and $\mu \neq T$.

Through a simulation study, we show that the higher the autocorrelation level the lower the capability index value. We also observed that for autocorrelated processes the estimators are slightly biased, bias that decreases as n increases.

The autocorrelation does not affect the expected value of the sample mean of the capability indices estimators but affect the estimated expected value of the standard error, that increases slightly for autocorrelated data when n increases.

Recibido: julio de 2007
Aceptado: octubre de 2007

References

Bissel, A. F. (1990), 'How Reliable is your Capability Index?', *Applied Statistics* **39**, 331–340.

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

Brockwell, P. J. & Davis, R. A. (1996), *Introduction to Time Series and Forecasting*, Springer, New York.

Guevara, R. D. (2005), Comparación de los índices de capacidad c_p, c_{pk}, c_{pm} y c_{pmk} con datos autocorrelacionados, Tesis de Maestría, Estadística, Universidad Nacional de Colombia, Facultad de Ciencias, Departamento de Estadística, Bogotá.

Kotz, S. & Johnson, N. L. (2002), 'Process Capability Indices – A Review, 1992-2000', *Journal of Quality Technology* **34**, 2–19.

Mood, Graybill & Boes (1974), *Introduction to the Theory of Statistics*, McGraw-Hill, New York.

Shore, H. (1997), 'Process Capability Analysis when Data are Autocorrelated', *Quality Engineering* **9**(4), 615–626.

Spiring, F. A., Leung, B., Cheng, S. & Yeung, A. (2003), 'A Bibliography of Process Capability Papers', *Quality and Reliability Engineering International* **19**, 445–460.

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

Appendix A.

In this appendix, we derive an approximation of the variance of $\widehat{C}_{pm}$.

Let $\{X_t\}$ be a stationary gaussian process. Using the estimator of C_{pm}:

$$\widehat{C}_{pm} = \frac{USL - LSL}{6\sqrt{S^2 + \left(\overline{X} - T\right)^2}}$$

the variance of $\widehat{C}_{pm}$, $Var\big(\widehat{C}_{pm}\big)$ is

$$Var\big(\widehat{C}_{pm}\big) = \left(\frac{USL - LSL}{6}\right)^2 Var\left\{\left[S^2 + \left(\overline{X} - T\right)^2\right]^{-\frac{1}{2}}\right\}$$

Using the approximation

$$Var\big[h(X,Y)\big] \approx Var[X]\left\{\frac{\partial}{\partial x}h(x,y)\Big|_{\mu_x,\mu_y}\right\}^2 +$$
$$Var[Y]\left\{\frac{\partial}{\partial y}h(x,y)\Big|_{\mu_x,\mu_y}\right\}^2 + 2Cov[X,Y]\left\{\frac{\partial}{\partial x}h(x,y)\Big|_{\mu_x,\mu_y}, \frac{\partial}{\partial y}h(x,y)\Big|_{\mu_x,\mu_y}\right\}$$

(see Mood et al. 1974):

Let $h\big(S^2,\overline{X}\big)=\Big[S^2+\big(\overline{X}-T\big)^2\Big]^{-\frac{1}{2}}$, then

$$\begin{aligned}
Var\big\{h\big(S^2,\overline{X}\big)\big\} &= Var\left\{\Big[S^2+\big(\overline{X}-T\big)^2\Big]^{-\frac{1}{2}}\right\}\\
&\approx Var(S^2)\left\{-\frac{1}{2}\Big(S^2+\big(\overline{X}-T\big)^2\Big)^{-\frac{3}{2}}\Big|_{E(S^2),E(\overline{X})}\right\}^2\\
&\quad+Var\big(\overline{X}\big)\left\{-\frac{1}{2}\Big(S^2+\big(\overline{X}-T\big)^2\Big)^{-\frac{3}{2}}2\big(\overline{X}-T\big)\Big|_{E(S^2),E(\overline{X})}\right\}^2\\
&\quad+2Cov\big[S^2,\overline{X}\big]\left\{-\frac{1}{2}\Big(S^2+\big(\overline{X}-T\big)^2\Big)^{-\frac{3}{2}}\Big|_{E(S^2),E(\overline{X})}\right\}\\
&\quad*\left\{-\frac{1}{2}\Big(S^2+\big(\overline{X}-T\big)^2\Big)^{-\frac{3}{2}}2\big(\overline{X}-T\big)\Big|_{E(S^2),E(\overline{X})}\right\}
\end{aligned}$$

where $Cov[S^2,\overline{X}]=0$ (see Zhang 1998). Then

$$\begin{aligned}
Var\big\{h\big(S^2,\overline{X}\big)\big\} &\approx Var\big(S^2\big)\left\{-\frac{1}{2}\Big(E\big(S^2\big)+\big[E\big(\overline{X}\big)-T\big]^2\Big)^{-\frac{3}{2}}\right\}^2\\
&\quad+Var\big(\overline{X}\big)\left\{-\frac{1}{2}\Big[E\big(S^2\big)+\big(E\big(\overline{X}\big)-T\big)^2\Big]^{-\frac{3}{2}}2\big[E\big(\overline{X}\big)-T\big]\right\}^2
\end{aligned}$$

$$\begin{aligned}
&\approx\frac{1}{4}\Big[E\big(S^2\big)+\big(E\big(\overline{X}\big)-T\big)^2\Big]^{-3}\Big\{Var\big(S^2\big)+4Var\big(\overline{X}\big)\big[E\big(\overline{X}\big)-T\big]^2\Big\}\\
&\approx\frac{Var\big(S^2\big)+4Var\big(\overline{X}\big)\big[E\big(\overline{X}\big)-T\big]^2}{4\Big[E\big(S^2\big)+\big(E\big(\overline{X}\big)-T\big)^2\Big]^3}
\end{aligned}$$

Using (1), (2), (3) and (4)

$$\begin{aligned}
Var\big\{h\big(S^2,\overline{X}\big)\big\} &\approx\frac{\frac{2\sigma^4}{(n-1)^2}F(n,\rho_i)+4\frac{\sigma^2}{n}g(n,\rho_i)[\mu-T]^2}{4[\sigma^2f(n,\rho_i)+(\mu-T)^2]^3}\\
&\approx\frac{\sigma^2\big\{\frac{2\sigma^2}{(n-1)^2}F(n,\rho_i)+\frac{4}{n}g(n,\rho_i)[\mu-T]^2\big\}}{4[\sigma^2f(n,\rho_i)+(\mu-T)^2]^3}
\end{aligned}$$

Therefore, the variance of $\widehat{C}_{pm}$ is

$$
\begin{aligned}
Var\big(\widehat{C}_{pm}\big) &\approx \left(\frac{USL-LSL}{6}\right)^2 \left[\frac{\sigma^2\big\{\frac{2\sigma^2}{(n-1)^2}F(n,\rho_i)+\frac{4}{n}g(n,\rho_i)[\mu-T]^2\big\}}{4[\sigma^2 f(n,\rho_i)+(\mu-T)^2]^3}\right] \\
&\approx \left(\frac{USL-LSL}{6\sigma}\right)^2 \left[\frac{\sigma^4\big\{\frac{2\sigma^2}{(n-1)^2}F(n,\rho_i)+\frac{4}{n}g(n,\rho_i)[\mu-T]^2\big\}}{4[\sigma^2 f(n,\rho_i)+(\mu-T)^2]^3}\right] \\
&\approx C_p^2 \left[\frac{\sigma^4\big\{\frac{2\sigma^2}{(n-1)^2}F(n,\rho_i)+\frac{4}{n}g(n,\rho_i)[\mu-T]^2\big\}}{4[\sigma^2 f(n,\rho_i)+(\mu-T)^2]^3}\right] \\
&\approx C_p^2 \left[\frac{\sigma^6\Big\{\frac{2}{(n-1)^2}F(n,\rho_i)+\frac{4}{n}g(n,\rho_i)\big[\frac{\mu-T}{\sigma}\big]^2\Big\}}{4\sigma^6\Big[f(n,\rho_i)+\big(\frac{\mu-T}{\sigma}\big)^2\Big]^3}\right] \\
&\approx C_p^2 \left[\frac{\Big\{\frac{2}{(n-1)^2}F(n,\rho_i)+\frac{4}{n}g(n,\rho_i)\big[\frac{\mu-T}{\sigma}\big]^2\Big\}}{4\Big[f(n,\rho_i)+\big(\frac{\mu-T}{\sigma}\big)^2\Big]^3}\right]
\end{aligned}
$$

Let $\xi = \frac{\mu-T}{\sigma}$,

$$
Var\big(\widehat{C}_{pm}\big) \approx C_p^2 \left[\frac{\frac{2F(n,\rho_i)}{(n-1)^2}+\frac{4g(n,\rho_i)\xi^2}{n}}{4[f(n,\rho_i)+\xi^2]^3}\right]
$$

Appendix B.

In this appendix we derive an approximation of the variance of $\widehat{C}_{pmk}$. Let $\{X_t\}$ be a stationary gaussian process. Writing the estimator of C_{pmk} as function of $\overline{X}$ and S^2,

$$
\widehat{C}_{pmk} = \frac{a-\big|2\overline{X}-b\big|}{6\Big[S^2+\big(\overline{X}-T\big)^2\Big]^{\frac{1}{2}}} = h\big(\overline{X},S^2\big)
$$

we have,

$$
Var\big[\widehat{C}_{pmk}\big] \approx Var\big[S^2\big]\left\{\frac{\partial h(S^2,\overline{x})}{\partial S^2}\Big|_{\mu_{S^2},\mu_{\overline{X}}}\right\}^2 + Var\big(\overline{X}\big)\left\{\frac{\partial h(S^2,\overline{x})}{\partial \overline{x}}\Big|_{\mu_{S^2},\mu_{\overline{x}}}\right\}^2
$$

where $\frac{\partial h(\overline{x},S^2)}{\partial S^2} = \frac{a-|2\overline{x}-b|}{6}\big[-\frac{1}{2}\big]\big[S^2+(\overline{x}-T)^2\big]^{-\frac{3}{2}}$ after some algebra, we obtain

$$
\begin{aligned}
Var\big(\widehat{C_{pmk}}\big) \approx C_{pk}^2 \left[\frac{1}{f(n,p_i)+\xi^2}\right] \times \\
\left\{\frac{F(n,p_i)}{2(n-1)^2[f(n,p_i)+\xi^2]^2}+\frac{g(n,p_i)}{9n}\left[\frac{1}{C_{pk}}+\frac{6\xi}{2[f(n,p_i)+\xi^2]}\right]^2\right\}
\end{aligned}
$$