Nepal Journal of Science and Technology 9 (2008) 41-48

An Application of Bayesian Method in Packaged Food Quality Control

R. P. Khatiwada[1] and A. B. Sthapit[2]
[1] Central Campus of Technology, Dharan
Institute of Science and Technology, Tribhuvan University
[2] Central Department of Statistics
Tribhuvan University, Kirtipur, Kathmandu
e-mail: ramktd@yahoo.com

Abstract

Conventional method of making statistical inference regarding food quality measure is absolutely based upon experimental data. It refuses to incorporate prior knowledge and historical data on parameter of interest. It is not well suited in the food quality control problems. We propose to use a Bayesian approach inferring the conformance of the data concerning quality run. This approach integrates the facts about the parameter of interest from the historical data or from the expert knowledge. The prior information are used along with the experimental data for the meaningful deduction. In this study, we used Bayesian approach to infer the weight of pouched ghee. Data are taken selecting random samples from a dairy industry. The prior information about average weight and the process standard deviation are taken from the prior knowledge of process specification and standards. Normal–Normal model is used to combine the prior and experimental data in Bayesian framework. We used user-friendly computer programmes, ‘First Bayes’ and ‘WinBUGS’ to obtain posterior distribution, estimating the process precision, credible intervals, and predictive distribution. Results are presented comparing with conventional methods. Fitting of the model is shown using kernel density and triplot of the distributions.

Key words: credible interval, kernel density, posterior distribution, predictive distribution, triplot

Introduction

Food materials are mixtures of different constituents and have multifaceted nature. Because of the wide range of variability, food sciences have high degree of uncertainty and constraints related to food are uncertain (Marten 1983). Mathematical models are essential for inferring the measurable product properties, the attributes or the variables (Hills 2001). They are widely used for life testing, hazard rates, confidence bounds, reliability measure. Montgomery (1997) and Duncan (1970) have used the mathematical models in statistical quality control and process capability measures. Hawthorn *et al.* (1984) and Bourlakis and Weightman (2004) have discussed statistical models for the consumer research and food supply chain management. Bowemen and O’Connell (1992), Poignee *et al.* (2003) and King (2000) have emphasized to pioneering quality control concepts to the business, agribusiness, and to agricultural products. Steiner (1967) strongly advised to use statistical methods for the quality control in the food industries.

The conventional statistics obtains point and confidence intervals and tests hypothesis entirely based upon the experimental data. Bayesian method uses the posterior predictive level by keeping informed the priors. Van Boekel (2003) suggests that model discrimination is more applicable than it is currently done in the food science. Besag and Higdon (1999) have discussed the application of Bayesian method in different aspects of agricultural and quality control experiments. For the quality control experiments, an uncertainty about the parameter is quantified corresponding to probabilities, and then they are updated by means of information gathered from the experiment in Bayesian approach. The uses of Bayesian approach to the statistical quality control can be found in Woodward and Naylor (1993), Wasko and Kim (2002), Colosimo and Semeraro (2002) and Zou *et al.* (2006). In view of Moe (1998), a computer based system of traceability is very important in industrial system even more significant with reference to the production of foodstuff. ‘First Bayes’ (O’Hagan 2003) and Bayesian inference using Gibbs sampling (WinBUGS) are commonly available computer programmes (Congdon 2003) dedicated to making

Bayesian calculation. WinBUGS is a complete Bayesian software based on MCMC simulation (Gilks *et al.* 1996, Smith & Roberts 1993).

In the food science world, the packaging of food-products has an important role in the business environment. The interest of the producers and customers first of all goes to the specified weight of the package or to the number of items contained in it. The producers are interested to fulfill the needs to implement necessary quality control measures at the lowest possible cost and customers to the quantity, quality and the service behaviour. In terms of quality control, the weight or the number of items in each packet is measurable variable, which is measured and predicted how it deviates from the assured. Inferences are drawn using tests of significance, maximum likelihood estimates, and confidence interval in classical approach.

On the other hand, in Bayesian approach, the principal mission is to obtain posterior distribution of the parameter of interest combining the data with prior information. A posterior predictive distribution, which makes available a complete distribution of the estimates, the Credible Intervals, Bayes factor, and Bayes risks are determined in this approach. The introduction and the method of Bayesian inference can be found in detail in Smith *et al.* (1965), Lindley (1970), Berger (1985), Lee (1997), Carlin and Louis (1996).

In this study, we concentrate on the weight of pouched ghee, assuming that the weight of pouched product is the customers' primary interest for the assurance and to measure consistency of the producer's claim. We use Bayesian approach for the inference of mean weight of a lot having known process variability, by estimating credible intervals and process control limits using 'First Bayes'. We display the suitability of the model iterating through MCMC simulation using WinBUGS. Comparing proposed method with classical method we interpret the results and form our conclusions.

Materials and Methods

Posterior density and predictive posterior model

The processed food item, ghee, is filled in pouch (packet) using a very precise computer controlled machine. Let, X be the weight of ghee in a packet, x_{ij} denotes the weight of jth packet of ith sample from a lot of size N; (i =1,2,……n) (j =1,2,……..k_i). The total number of samples observed is n. The mean weight (θ) of the packet is parameter of our interest. For a well mechanized filling process, and the lots of thousands item, the weights assumed to be distributed normally. Our assumption is the data (X) follows Gaussian distribution with parameter θ and σ

$$X \sim N(\theta, \sigma^2) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(1).}$$

θ and σ are the mean and standard deviation (given that σ known). The distribution of X for given θ is the *likelihood* of θ..

We come up to the question of the variation in mean pouch weight in different samples having a well preset process. The mean weight *(θ)* is considered not only a fixed unknown quantity as in conventional classical method; it is assumed as an uncertain quantity, having property of a random variable, defined in some real line, $\theta \in \Theta$.

We assume, θ follows Gaussian *prior distribution* with mean θ_0 and standard deviation σ_0

$$\theta \sim N(\theta_0, \sigma_0^{\,2}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2).}$$

Then, the *posterior distribution* of parameter of interest (mean) given the data is also Gaussian (For detail and proof see, Lindley 1970 and Berger 1985).

$$\theta | X \sim N(\theta_1, \sigma_1^{\,2}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3).}$$

θ_1 and σ_1^2 are the posterior mean and variance, where

$$\theta_1 = \sigma_1^{\,2}\left(\frac{\theta_0}{\sigma_0^{\,2}} + \frac{n\cdot\hat{\theta}}{\sigma^2}\right) \text{ and } \sigma_1^{\,2} = \left(\frac{1}{\sigma_0^{\,2}} + \frac{n}{\sigma^2}\right)^{-1}$$

And, n is the total number of sample observed; $\hat{\theta}$ is the estimated value (from data) of θ.

$$\hat{\theta} = \bar{\bar{x}} = \frac{\sum_{i=1}^{n}\bar{x}_i}{n}, \text{where } \bar{x}_i = \frac{\sum_{j=1}^{k_i} x_{ij}}{k_i},$$ k is the size of each sample.

The *predictive distribution* of the new sample after obtaining the posterior density of the first n samples is given by $(X_{n+1} | X_n \sim N(\theta_p, \sigma_p^{\,2}))$.

$(X_{n+1} | X_n) \sim N(\theta_1, (\sigma^2 + \sigma_1^2))$....... ....(4)

where, $\theta_p = \theta_1$, and $\sigma_p^2 = \sigma^2 + \sigma_1^2$ (see, Lindley 1970 and Lee 1997)

Credible interval

Bayesian confidence interval is well-known as the credible interval. It is the probability that a parameter of interest shall occur in that interval. If probability density concentrated around the posterior mean it is called the highest probability density (HPD). The shortest Bayesian confidence region is the region of HPD, and is called highest density region (HDR) or highest density interval (HDI) (Lee 1997; Carlin & Louis 1996).

For an unknown parameter of interest, θ, we assume within a real line it is a member of Θ, i.e., $\theta \in \Theta$. $\hat{\theta}$ is the estimate of θ, $f(\hat{\theta}|\theta)$ is the probability density function (pdf) of $\hat{\theta}$ given $\theta \in \Theta$, and $f(\theta|\hat{\theta})$ is the pdf of θ given $\hat{\theta}$. Our interest is to obtain the probability $\gamma = p|(\theta' < (\theta \in \Theta) < \theta'')|\hat{\theta}|$............(5) for all θ, θ' and θ'' real numbers; usually γ is a symbol of $(1-\alpha)100\%$.

The value of θ, θ' and θ'' together with our knowledge and $f(\hat{\theta}|\theta)$ allows us to calculate the probability γ, the credible interval, within which the parameter of interest $\theta \in \Theta$ lie. In (5), the value of γ is alike to the posterior density, therefore, the calculation of the confidence interval after the posterior upgrading is momentous (D'Agostini 2006).

The posterior density itself is a set of estimates. The highest density region (HDR) or the credible interval, for a given α, is obtained using $HDR = \theta_1 \pm Z_{\alpha/2}\sigma_1^2$ ....(6), where θ_1 and σ_1 are posterior mean and standard deviation. These estimates are used to make inference in quality control problems too. The calculations of complex integrals are easily solved and a complete analysis is done using First Bayes and WinBUGS.

Sample and Data

Processed liquid ghee is filled in the 1 litre pouch. Because of using well computerized filling machine, the distribution of the weight of pouched ghee is assumed to be normally distributed. Random samples of the pouches were taken from the finished product to measure the weights. The mean weight of the pouch (the parameter of interest) is assumed to be distributed normally with unknown mean and known standard deviation. 25 samples of size 5 were taken in different time period and the average weight measured. The specified lower limit of the average weight is 920, the target value is 930.

Data: (average weights in g):

900	*905*	*914*	*913*	*927*	*915*	*900*	*918*	*908*	*916*
918	*924*	*925*	*934*	*929*	*920*	*925*	*930*	*930*	*930*
924	*930*	*934*	*922*	*934*					

Data analysis

A classical method of calculating point estimate and the confidence interval is used for the estimation, initially. Assuming normal conjugate prior we obtained the posterior density for the parameter of interest, subsequently. Posterior and predictive distributions, precisions, Credible Intervals, 3σ tolerance intervals are computed using First Bayes. The prior and posterior distributions along with likelihood (experimental data) are shown using triplots. Further, the inferences are interpreted using WinBUGS through simulation. The results obtained using Bayesian methods are compared with the results of the conventional method.

Results and Discussion

From the mechanization of the industry the target value (the weight of the individual pouch) is set as 930g with tolerance limits of ± 30g from the target value. This indicates the process spread $\mu \pm 3\sigma$ is 930 ± 30 and the process standard deviation (σ) for the weight of the individual pouch is 10. To estimate the average weight of the pouch, the size of a sample (k) is taken as 5. So, the variance of the sample average weight of the pouch is $(\sigma_{\bar{x}}^2 = \sigma^2 / k)$ is expected as 20g. This information is assumed as the prior information and the Bayesian framework is used to obtain estimates and probabilities.

Table 1. Summary of the estimated parameters and confidence intervals using classical method

Source	Mean	Standard Deviation	Confidence limits (for, σ = 10)			lower specification limit = 920 Target value =930	
			Central 50%	95%	3σ	P(<920)	P(≥930)
	930	10 (known)	923.25, 936.75	910.4, 949.6	900.00, 960.00	0.1587	0.5
Data, x_i x_i|~N(921,100)	921	10.169 (computed)	914.25, 927.75	901.4, 940.6	890.86, 951.14	0.4602	.184
For mean ($\overline{x}$)	921	2 *	919.65, 922.35	917.1, 924.92	914.97, 927.03	0.3085	0.000

*The total number of sample (n) is 25, processsd =10 and $SE(\bar{x}) = \sigma / \sqrt{n} = 2$

Table 2. Summary of the prior density, likelihood and the posterior density with precisions

Density of X	Prior density $\pi(\theta)$	prior precision	likelihood of data $f(X\|\theta)$	Data precision	Posterior density $p(\theta\|X)$	Posterior precision
$X\sim N(\theta,\sigma^2)$	$\theta\sim N(\theta_0,\sigma_0^2)$		$x\|\theta\sim N(\hat{\theta},\sigma^2)$		$\theta\|X\sim N(\theta_1,\sigma_1^2)$	
$N(\theta, 10^2)$	$N(930, 20)$	0.05	N(921.00, 100), n =25	0.25	N(922.5, 3.333)	0.30

Table 3. Summary of the posterior distribution

Posterior density θ|X	50%HDR(Q_1, Q_3)	95%HDR	3σ limits for mean	P(<920)	P(≥930)
N(922.5, 3.3333)	921.27, 923.73	918.92, 926.08	917.00, 928.00	0.0855	0.0000

Table 4. Summary of the predictive distribution

Predictive density $X_{i+1}\|X_i$	50% HDR (Q_1,Q_3)	95% HDR	3σ limits for X_{i+1}	P(<920)	P(≥930)	1-P(900 ≤X≤960)
N(922.5, 103.33)	915.64, 929.36	902.57, 942.43	891.87, 953.13	0.4028	0.2303	0.0136

Plots using First Bayes

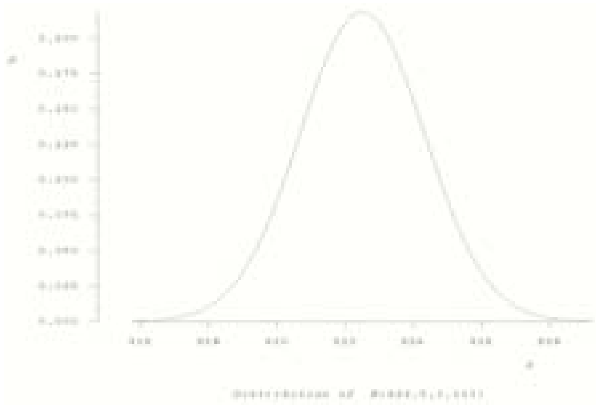

Fig. 1. Graph of the posterior density of average weight

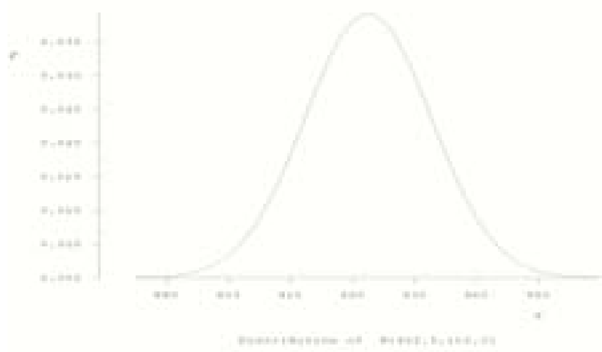

Fig. 3. Graph of the predictive density of the weight (X)

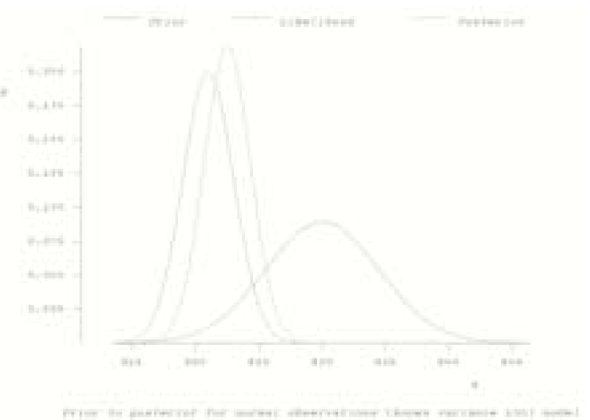

Fig. 2. Triplot of the prior density (solid line), posterior density (dotted line) and likelihood (dashed line)

Using classical method (Table 1), we have computed process mean (= 921) less than the target value (930), with almost equal to the given process variance (100). The estimated 3σ limit for mean is within the natural tolerance limits (900,960), but, a sample falls below the lower specification limit (920), 46 out of 100 chances. Also, the 95% confidence interval is within the natural tolerance limit, having lower limit below the lower specification. The precision of the estimate is 0.25.

Table 2 shows the prior density, likelihood of the data and the posterior density. The posterior mean and variances are 922.5 and 3.33 respectively, obtained using First Bayes. The posterior densities describe the distribution of the estimate. The posterior precision ($1/\sigma_1^2 = 0.3$) is equal to the sum of the prior precision and data precision.

Table 3 shows the posterior probabilities occurring between the different intervals. The probability of occurring average weight beyond the lower specification (120) is 12.7/1000. No part of the average weight appears more than the target value (930). There is almost sure probability that the average weight occurring between 917 and 928.

Table 4 shows the predictive distribution of the weight (X). If we wish to draw a random sample, given the posterior distribution, the probability that the sample value will occur beyond the lower specification is 0.4028. The chance of being a newly drawn random sample above the lower specification is 0.5972. For all new draw, given the posterior distribution, the probability of occurring average weight exceeding 930 is 0.2303. Also, the probability of the weight of a sample drawn given the posterior, will lie beyond the natural tolerance limit, is 0.0136.

The model $X \sim N(930, 100)$, allowed that the chance of happening weight below 920gm (lower specification) not to be more than 15.9%. Data information shows the part of the weight below specification is 46% (Table 1). From the classical method, we find 30.9% chance that a sample average weight falling below the specification. Using Bayesian method, we obtained the probability of the average weight occurring below specification is only 8.5%. And, if a new sample will be drawn, the chance that it will below specification is 40.3%. There is almost zero probability of occurring an average weight more than the target value, and that of the newly drawn sample is 0.23. The lower capability index (Mitra 2001) (for the lower half) for the posterior distribution is 0.25, where that of the data is 0.1.

The WinBUGS result

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
mu (theta)	922.5	1.825	0.02374	918.9	922.5	926.1	1001	5000
mu (theta)	922.5	1.817	0.01043	918.9	922.5	926.1	1	30000

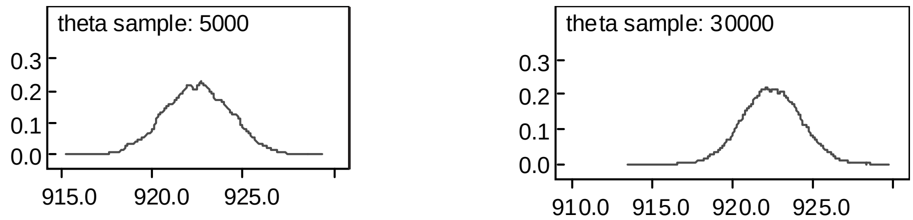

Fig. 4. Kernel density plots of the posterior distribution through MCMC using WinBUGS

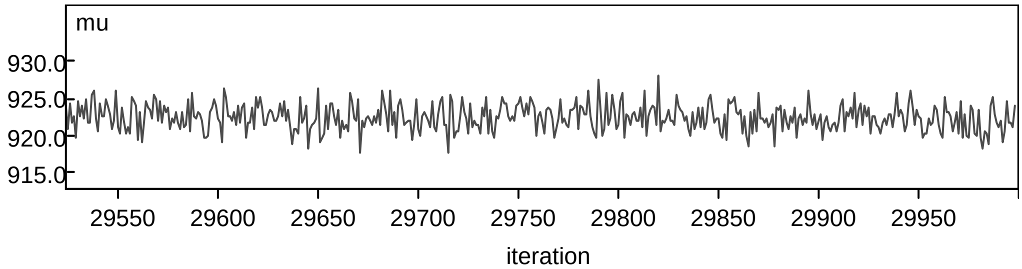

Fig. 5. Trace of the sample posterior distribution through MCMC using WinBUGS

The WinBUGS result visualizes that the mean of the distribution converges to 922.5 for a moderate and extended iteration with sd of 1.8. The Kernel density plot (figure 4) attempts to reproduce the mean and underlying distribution and shows the increasing smoothness of the curve as increase in iteration. The trace (figure 5) shows the consistency in sample values for a cross section of the iteration.

The statistical method of data analysis in quality control is increasing concern of the food scientists. Their effort goes to maintain, producers' assertion and satisfaction with the quality, quantity and other assurances of the customers, by means of reliable bases of statistics. However, there is a lack of the use of statistical methods in this field, because of assuming it as a tedious job. In this study we have proposed a latest, straightforward, remarkable and attractive method, which may lend a hand to interested people. The Bayesian method is robust for estimation of the quality characteristics and the process variations. The results can be strengthened to predict the process capability and optimize customers' concern to facilitate as their requirements. The uncertainties about process or product parameters and quality characteristics can be estimated in very appealing way using this approach.

Our study is simply an opening of the use of Bayesian method in food quality concern. In this study, we have used the Normal-Normal model, which is just an example of using conjugate normal prior with known variance. There is a broad spectrum to use different powerful and practical models in Bayesian approach. In food science world, food quality, safety and consumers' concern is increasing day by day; so we hope, our attempt opens a door of the potentiality of the Bayesian inference for this field.

Acknowledgement

This paper was prepared during the Ph. D. research work of the first author, partially supported by a grant from Nepal Academy of Science and Technology.

References

Berger, J. O. 1985. *Statistical decision theory and Bayesian analysis*. Springer series in Statistics, Springer-Verlag, New York.

Besag, J. and D. Higdon. 1999. Bayesian inference for agricultural field experiments. *Journal of the Royal Statistical Society, B*, (61): 691-746.

Bourlakis, M. and P. Weightman. 2004. *Food Supply Chain Management.* Blackwell Publication, UK.

Bowemen, B. I. and R. T. O'Connell. 1992. Integrating quality control concepts into the introductory business statistics course. *The American Statistician*, **46** (2): 124-128.

Carlin, B.P. and T.A. Louis. 1996. *Bayes and Empirical Bayes Method for Data Analysis.* Chapmen & Hall, London.

Colosimo, B. M. and Q. Semeraro. 2002. A Bayesian control chart for service quality control. A paper presented at the *Quality and Productivity Research Conference* (June 5-7, 2002), Available in AVL proceedings; Tempe, Arizona.

Congdon, P. 2001. *Bayesian statistical modelling.* Wiley series in Probability and Statistics, Wiley, Chichester.

D'Agostini, G. 2006. About the proof of the so called exact classical confidence intervals. Where is the trick? A paper presented *at Xiv:Physics/0605140v1, 16 May2006, Version:http://www.roma1.infn.it/"dagos.*

Duncan, A. J. 1970. *Quality control and industrial statistics*. Richard D. Irwin Inc., Homewood, Illinois.

Gilks, W. R., S. Richardson and D. J. Spiegelhalter. 1996. *Markov Chain Monte Carlo in practice*. Chapman & Hall, London.

Hawthorn, J., B. Schweigert, G. Stewart, M. Jr. Gacula and J. Singh. 1984. *Statistical methods in food and consumer research*. Academic press, USA. version:http://books.elsevier.com/bookscat/links/details.asp?isbn =0122720504, July 2006

Hills, B. P. 2001. The power and pitfalls of deductive modelling. In: *Food Process Modelling* (Eds. L.L.M. Tijskens, T. M. Hertog, *et al*.). Woodhead Publishing House, England.

King, R. P. 2000. *Supply chain design for identity preserved agricultural products. Version: http://www.sls.wau.nl/mi/Activities/papers/0002113pdf*.

Lee, P.M. 1997. *Bayesian statistics: An introduction*. Arnold, London. *version:* http://wwwusers.york.ac.uk/~pml1/bayes/book.htm. July, 2006.

Lindley, D.V. 1970. *Introduction to probability and statistics from a Bayesian view point*, part 2. Cambridge University press, London.

Martens, H. 1983. Understanding food research data. In: *Food Research and Data Analysis*. (Eds. H. Martens, and H. Jr. Russwurm). Norwegian Food Research Institute As, Applied Science Publisher, New York.

Mitra, A. 2001. *Fundamentals of quality control and improvement* (2nd edition), Indian reprint. Pearson Education, India.

Moe, T. 1998. Perspectives on traceability in food manufacture. *Trends in Food Science & Technology* **9** (5): 211-214.

Montgomery, D. C. 1997. *Introduction to statistical quality control* **(**Third edition), John Wiley and Sons, New York.

O'Hagan, A. 2003. Bayesian statistics: Principles and benefits. In: *Bayesian Statistics And Quality Modeling In The Agro- Food Production Chain* (Eds. M. A. J. S. Van Boekel, A. Stein *et al.*). Kluwer Academic Publication, The Netherlands.

Poignee O., T. Hannus and G. Schiefer. 2003. Vertical integrated quality management in agribusiness: A case study. *A paper presented at EFITA 2003 Conference* (5-9 July), Debrecan, Hungary.

Smith, A. F. M., and G. D. Roberts. 1993. Bayesian computation via the Gibbs Sampler and related Markov Chain Monte Carlo Method. *Journal of the Royal Statistical Society, B*, **55**(1): 3-33.

Smith, A.F.M., A. M. Skene, I. E. H. Shaw, J. C. Naylor and M. Dransfield. 1965. The Implementation of the Bayesian paradigm. *Communication in Statistics, A, (Theory and Methods)*, **14**: 1079-1102.

Steiner, E. H. 1967. Statistical methods in quality control. In: *Quality control in the food industry: A Series of Monographs* (ed. S. M. Herschdoerfer), Vol. 1. Academic Press New York, pp. 121-243.

Van Boekel, M. A. J. S. 2003. Bayesian solutions for food-science problems. In: *Bayesian Statistics and Quality Modelling in the Agro- Food Production Chain* (Eds. M. A. J. S. Van Boekel, A. Stein *et al*.). Kluwer Academic Publication, The Netherlands.

Wasko, J. A., D. B. Kim. 2002. An almost natural application of Bayesian statistics in packaging quality control. A paper presented at *United State Military Academy, West point, NY 10996. Version: dvipdf m0.132c.©1998by Mark A Wicks*, text output 2003/1/31.

Woodward, P. W. and J. C. Naylor. 1993. An application of Bayesian method in SPC. *The Statistician* **48**: 461-449.

Zou, C., Y. Zhang and Z. Wang. 2006. A control chart based on a change point model for monitoring linear profiles. *IIE Transactions* **38**: 1093-1103.