

# Process Performance Maximization through Input Parameters Optimization using Taylor Series Expectation Approximation

Santanu Pattanayak  
GE-Digital  
[santanu\\_pattanayak2003@yahoo.co.in](mailto:santanu_pattanayak2003@yahoo.co.in)

---

## Abstract

In several industrial settings or manufacturing units the process output is expressed as a function of several inputs and such relationships can be expressed in the form of the equation based on the prior data or known truth. It is desirable in such a setting that the output is within a specified threshold for acceptance of the quality of the product. It is not always possible to operate the industrial plant under several settings and then evaluate the one that maximizes the performance. Besides such experiments/tests can be expensive as well as time consuming. Given that the output-input relationship in the form of an equation and that the probability distribution of the inputs is known from the prior data the process performance at the current setting can be evaluated by randomly sampling inputs as per the input distributions and then checking the proportion of samples for which the output values are outside the specified threshold. If the process performance is not satisfactory there are generally two ways that one can optimize the performance. One can i) optimize the input parameters mean/mode, etc. keeping the variance of the input distributions constant (or/and) ii) optimize the variances of the inputs keeping the other parameters of the distribution constant. Generally, in such settings controlling the variances of the inputs is often difficult and harder to implement as opposed to varying the mean or other parameters that don't influence the variances. This paper attempts to illustrate an efficient technique to find the best mean of the input probability distributions (keeping the input variances, skewness and kurtosis constant) that maximize or minimize the output mean so that the proportion of output beyond the threshold is minimized.

## Background or Introduction

Any process can be expressed in the form of an equation involving process output as a function of the process inputs either through modelling the system from collected data through processes such as regression or through known literature in science. Similarly, the variation of the input data in the form of probability distribution can be captured from past collected data or known from the domain specific science.

Once the inputs-output relation equation and the input probability distributions are known generally sampling is used to generate several samples based on input probability distributions. However, if there is significant correlation between variables then that needs to be considered while sampling. An efficient approach is to first generate random samples for each variable independent of each other. The correlation matrix  $C$  can be decomposed into  $C = LL^T$  by Cholesky decomposition where  $L$  is the lower triangular matrix. Now if the uncorrelated input data vector  $x' \in \mathbb{R}^{nx1}$  where  $n$  denotes the number of inputs undergoes the required transformation  $x = Lx'$  then  $x \in \mathbb{R}^{nx1}$  will have correlated data in accordance with the covariance between input variances. If the input output equation is expressed as  $y = f(x)$  then based on the generated correlated input samples the output samples can be generated. Depending on the output acceptance criteria such as the value of output should be within certain range, one can evaluate how the process is doing based on the proportion of sample outputs satisfying the output acceptance criteria. For instance, if we generate 50000 samples out of which only 40000 samples meet the output acceptance criteria then the process in general produces 20% output with suboptimal

quality. The outputs in general corresponds to products in an industrial setting and the output value not meeting the output acceptance criteria results in the rejection of or rework on the product.

In such cases the goal is to optimize the process so that the proportion of outputs within the acceptance threshold increases. One way to do so is to maximize or minimize the output mean so that the overall output acceptance proportion increases. The output mean can be impacted by altering how the inputs vary i.e. by changing the parameters of the input distribution. Generally, two approaches are used –

- i) Optimize the input mean of the distribution keeping the variance, correlation, skewness and kurtosis of the inputs constant.
- ii) Alter the variance keeping the mean constant.

In any industrial process setting the variances of the inputs is much harder to optimize in comparison to the mean and hence mean optimization is the first resort. For example, in a setting if temperature is one of the inputs following normal distribution with mean 25 and variance 1.5 it is easier to vary the mean by increasing the temperature rather than to vary the variance. Hence mean optimization is the first resort in these kind of optimization problems. However, the mean of the inputs can't be varied at will since there would be limitations on the range of values the input mean can take. So, the mean optimization would be constrained by the range of values the input mean can assume.

One of the frequently used techniques is to perform sampling at various combination of mean values for the inputs based on their mean ranges specified and then choose the combination of input means that provides the minimum proportion of output values beyond the user specified acceptance threshold. The most popular technique to do so is to use grid-search or random search on the mean ranges of the input variables.

Grid search divides the parameter (mean here) ranges into  $n$  equally spaced mean values for each input variable and then computes all possible input mean combination. If there are two input variables, then there would be  $n^2$  input mean combinations to evaluate the process at. With the increase in the number of input variables the number of input mean combinations increases exponentially.

Another important technique that's popularly used in Random Search where not all exhaustive set of mean combination setting is considered but the algorithm tries a specified number of random mean combination setting and chooses whatever optimal point is reached at the end of the number of iterations. The algorithm starts with a mean combination setting  $x$  and then randomly samples a new mean setting  $y$  from a hypersphere of given radius within the mean ranges and checks if process performance at  $y$  is better than that in  $x$ . If the performance is better in the new setting, then the new setting is accepted and then the whole process is repeated.

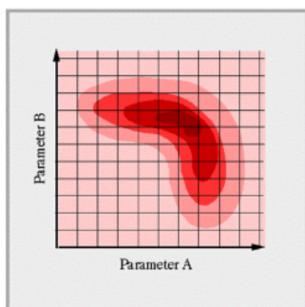


Figure -1: Grid for a two-input grid search

Although technically in random search at less number of settings the process performance needs to be evaluated, for better results the number of iterations must be increased. At each setting, for both grid-search and random search, the process performance is evaluated by randomly generating samples and then checking the output proportion meeting output acceptance criteria. This itself is computationally intensive and when clubbed with several settings in which the process performance needs to be evaluated the overall computational cost and complexity for the optimization problem increases. Also since both grid-search and random search evaluates the process at some discrete mean settings for the inputs the results obtained might lack the desired level of accuracy.

### **Proposed Solution**

To overcome the shortcomings of the sampling method through grid search and random search in the optimization problem, this solution proposes converting the inputs-output equation into an equation where the output mean is expressed as a function of the input means through a 4<sup>th</sup> order Taylor series expansion and then optimizing the output mean based on the constraints on the range of the input means. Such Taylor Series expansion around the mean is possible since in industrial settings the input variable values tend to remain near the mean with small variance or operate within a small range of values. This also leads to simple probability distributions for the inputs such as Normal, Weibull, Triangular and Uniform distribution. The advantage with these distributions is that the variance, covariance, skewness and the kurtosis can be held constant while altering the mean of these distribution. Such liberties are not possible when working with distributions such as Poisson distribution where the mean and the variance of the distribution are same.

Given a function  $y = f(x)$  through Taylor Series 4<sup>th</sup> order expansion it can be expanded as below –

$$f(x + h) = f(x) + \sum_{i=1}^n \frac{\partial f}{\partial x_i} h_i + \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i^2} h_i h_j + \sum_{i,j,k=1}^n \frac{\partial^3 f}{\partial x_i^3} h_i h_j h_k + \sum_{i,j,k,l=1}^n \frac{\partial^4 f}{\partial x_i^4} h_i h_j h_k h_l$$

where  $x, h \in \mathbb{R}^{nx1}$  and  $x$  is the input vector of variables while  $h$  is the change in the input vector.

This solution proposes a 4<sup>th</sup> order Taylor series expansion around the mean of the inputs. Let the input vector of variables be represented by  $x = [x_1 \ x_2 \ \dots \ x_n]^T \in \mathbb{R}^{nx1}$  and the mean of the input vector be represented by  $\mu = E[x] = [\mu_1 \ \mu_2 \ \dots \ \mu_n]^T \in \mathbb{R}^{nx1}$ . The function can be expanded around its mean as below –

$$\begin{aligned} f(x) &= f(\mu + x - u) \\ &= f(u) + \sum_{i=1}^n \frac{\partial f}{\partial x_i} (x_i - \mu_i) + \frac{1}{2!} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} (x_i - \mu_i)(x_j - \mu_j) \\ &\quad + \frac{1}{3!} \sum_{i,j,k=1}^n \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k} (x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k) \\ &\quad + \frac{1}{4!} \sum_{i,j,k,l=1}^n \frac{\partial^4 f}{\partial x_i \partial x_j \partial x_k \partial x_l} (x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)(x_l - \mu_l) \end{aligned}$$

All the partial derivatives are evaluated at  $x = \mu$  and to keep the deductions neat the same has not been explicitly mentioned in the equations.

We want to find the equation for the mean of the output with respect to the mean of the input variables. Taking expectation on both sides we get –

$$\begin{aligned}
E[f(x)] &= E[f(u)] + E\left[\sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_i - \mu_i)\right] + \frac{1}{2!} E\left[\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x_i - \mu_i)(x_j - \mu_j)\right] \\
&\quad + \frac{1}{3!} E\left[\sum_{i,j,k=1}^n \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)\right] \\
&\quad + \frac{1}{4!} E\left[\sum_{i,j,k,l=1}^n \frac{\partial^4 f}{\partial x_i \partial x_j \partial x_k \partial x_l}(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)(x_l - \mu_l)\right]
\end{aligned}
\tag{a}$$

Mean of the 1<sup>st</sup> order terms will vanish while the 2<sup>nd</sup> order terms will yield variances and covariances.

$$E\left[\sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_i - \mu_i)\right] = \sum_{i=1}^n \frac{\partial f}{\partial x_i} E[(x_i - \mu_i)] = 0 \tag{i}$$

$$\begin{aligned}
\frac{1}{2!} E\left[\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x_i - \mu_i)(x_j - \mu_j)\right] &= \frac{1}{2!} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} E[(x_i - \mu_i)(x_j - \mu_j)] = \frac{1}{2!} \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2} \sigma x_i^2 + \\
\frac{1}{2!} \sum_{i,j=1, i=j, i \neq j}^n \frac{\partial^2 f}{\partial x_i \partial x_j} \text{cov}(x_i, x_j) &\tag{ii}
\end{aligned}$$

where  $\sigma x_i^2$  is the variance of the  $i$ th input variable and  $\text{cov}(x_i, x_j)$  is the co-variance between the  $i$ th and the  $j$ th input variables.

the 3<sup>rd</sup> order term involves skewness terms when  $i = j = k$  and co-skewness terms otherwise. Similarly, the 4<sup>th</sup> order terms involve kurtosis when  $i = j = k = l$  and co-kurtosis terms otherwise. In process performance models in industrial settings the input variables taken are as much independent as possible and hence the co-skewness and the co-kurtosis terms can be ignored since they add complexity to the model and are difficult to estimate from given data. Taking all these into consideration, for the 3<sup>rd</sup> and 4<sup>th</sup> order terms the below approximations hold –

$$\begin{aligned}
\frac{1}{3!} E\left[\sum_{i,j,k=1}^n \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)\right] &= \frac{1}{3!} \sum_{i,j,k=1}^n \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k} E[(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)] \approx \\
\frac{1}{3!} \sum_{i=1}^n \frac{\partial^3 f}{\partial x_i^3} E[(x_i - \mu_i)^3] &= \frac{1}{3!} \sum_{i=1}^n \frac{\partial^3 f}{\partial x_i^3} \gamma(x_i) \tag{iii}
\end{aligned}$$

where  $\gamma(x_i)$  is the un-normalized skewness of the input variable  $x_i$

$$\begin{aligned}
\frac{1}{4!} E\left[\sum_{i,j,k,l=1}^n \frac{\partial^4 f}{\partial x_i \partial x_j \partial x_k \partial x_l}(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)(x_l - \mu_l)\right] &= \frac{1}{4!} \sum_{i,j,k,l=1}^n \frac{\partial^4 f}{\partial x_i \partial x_j \partial x_k \partial x_l} E[(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)(x_l - \mu_l)] \approx \\
\frac{1}{4!} \sum_{i=1}^n \frac{\partial^4 f}{\partial x_i^4} E[(x_i - \mu_i)^4] &= \frac{1}{4!} \sum_{i=1}^n \frac{\partial^4 f}{\partial x_i^4} \kappa(x_i) \tag{iv}
\end{aligned}$$

Where  $\kappa(x_i)$  is the unnormalized kurtosis of the input variable  $x_i$

Using (i), (ii), (iii) and (iv) in (a):

$$E[f(x)] = g(\mu) = f(u) + \frac{1}{2!} \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2} \sigma x_i^2 + \frac{1}{2!} \sum_{i,j=1, i \neq j}^n \frac{\partial^2 f}{\partial x_i \partial x_j} cov(x_i, x_j) + \frac{1}{3!} \sum_{i=1}^n \frac{\partial^3 f}{\partial x_i^3} \gamma(x_i) + \frac{1}{4!} \sum_{i=1}^n \frac{\partial^4 f}{\partial x_i^4} \kappa(x_i)$$

For each of the input variables there would be constraints on the mean range within which the input mean can be optimized. For the  $i$ th variable  $x_i$  let the input mean  $\mu_i$  range within which it can be optimized be given by –

$$\mu_l^{(i)} \leq \mu_i \leq \mu_h^{(i)} \quad \forall i \in \{1, 2, 3, \dots, n\}$$

where  $\mu_l^{(i)}$  and  $\mu_h^{(i)}$  denote the lower and upper bound on the mean range for the  $i$ th variable. So, there would be  $n$  constraints corresponding to  $n$  input variables.

Based on the objective or the evaluated process performance at existing setting the output mean may either be maximized or minimized to ensure the proportion of output values beyond the user specified threshold is reduced.

For output mean maximization, the optimization problem would be as below –

$$\begin{aligned} \mu^* &= \underbrace{\arg\max}_{\mu} g(\mu) \\ &= \underbrace{\arg\max}_{\mu} f(u) + \frac{1}{2!} \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2} \sigma x_i^2 + \frac{1}{2!} \sum_{i,j=1, i \neq j}^n \frac{\partial^2 f}{\partial x_i \partial x_j} cov(x_i, x_j) + \frac{1}{3!} \sum_{i=1}^n \frac{\partial^3 f}{\partial x_i^3} \gamma(x_i) \\ &\quad + \frac{1}{4!} \sum_{i=1}^n \frac{\partial^4 f}{\partial x_i^4} \kappa(x_i) \end{aligned}$$

$$\text{subject to } \mu_l^{(i)} \leq \mu_i \leq \mu_h^{(i)} \quad \forall i \in \{1, 2, 3, \dots, n\}$$

Similarly, for output mean minimization, the optimization problem would be as below –

$$\begin{aligned} \mu^* &= \underbrace{\arg\min}_{\mu} g(\mu) \\ &= \underbrace{\arg\min}_{\mu} f(u) + \frac{1}{2!} \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2} \sigma x_i^2 + \frac{1}{2!} \sum_{i,j=1, i \neq j}^n \frac{\partial^2 f}{\partial x_i \partial x_j} cov(x_i, x_j) + \frac{1}{3!} \sum_{i=1}^n \frac{\partial^3 f}{\partial x_i^3} \gamma(x_i) \\ &\quad + \frac{1}{4!} \sum_{i=1}^n \frac{\partial^4 f}{\partial x_i^4} \kappa(x_i) \end{aligned}$$

$$\text{subject to } \mu_l^{(i)} \leq \mu_i \leq \mu_h^{(i)} \quad \forall i \in \{1, 2, 3, \dots, n\}$$

The above optimization problems can be solved through nonlinear constrained optimizers such as SLSQP i.e. Sequential Least Square Programming which uses the Han–Powell quasi–Newton method with a BFGS update of the B–matrix and an L1–test function in the step–length algorithm.

The  $\mu^*$  is the optimized mean vector maximizing or minimizing the output mean based on the optimization problem. The process should be operated at  $\mu^*$  as the mean vector for the input variables to achieve maximum process performance given the constraints on the input means.

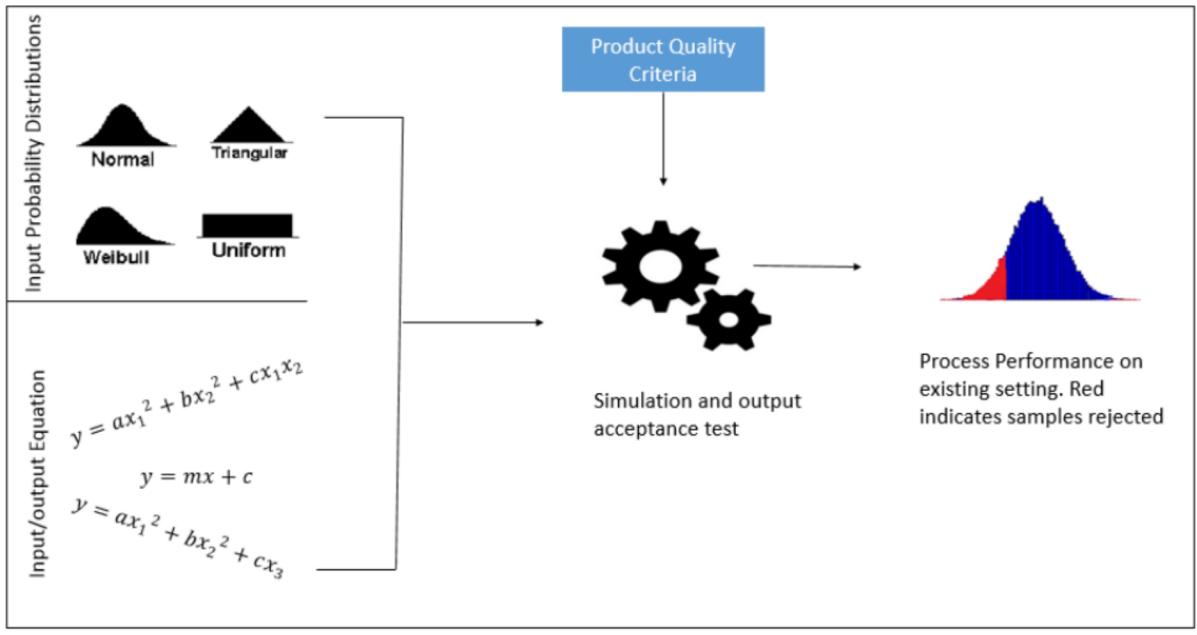
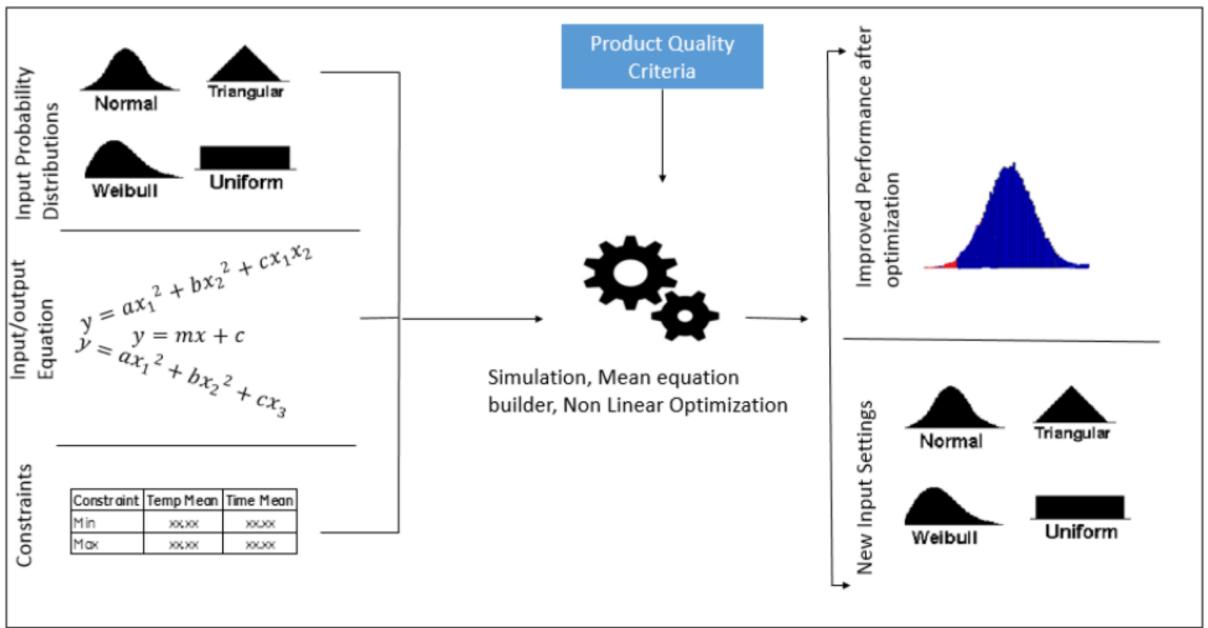


Figure -2. Evaluating the process at existing setting.



*Figure -3. Proposed approach of Process Performance Maximization*

## Results

A use case requiring Chips Packet Seal Strength optimization was evaluated for accuracy and computational efficiency using the proposed method and Grid Search method.

Chips packet seal strength depends on the temperature of the sealing device and the sealing time. Suppose the relationship between the seal strength and the inputs follows the below relation.

$$\text{Strength} = 19.128 + 0.006 * \text{Temp} + 8.0042 * \text{Time} + 0.00029 * \text{Temp} * \text{Time}$$

At the current manufacturing unit setting, the chips packages are sealed at an average temperature of 120 degrees Celsius, with a standard deviation of 25.34. The mean sealing time is 1.5 seconds, with a standard deviation of 0.5. Both the inputs temperature and time of sealing follow a normal distribution. Both the inputs i.e. temperature and time of sealed are uncorrelated. The package seal strength should be greater than 27 Newtons/sq-mm for the chips packet to pass the quality acceptance test. Also, the mean of temperature can be optimized in the range from 100 to 200 degree Celsius while the mean time of sealing can be optimized between 1 and 2 seconds.

Through random sampling around 50000 input samples were generated based on the existing setting for the inputs and the process performance was evaluated. Around 11.004 % output samples corresponding to the input samples generated were rejected because the seal strength for those samples were less than 27 Newtons/sq-mm.

The proposed method of optimization was applied and the new setting the optimizer came up with was as below –

The output of the Performance Optimizer shows that by changing the mean of Temp and Time to 200 degree Celsius and 2 seconds respectively the percentage of output samples beyond the threshold of 27 Newtons/sq-mm could be brought down to 0.988 % which is a huge win. Once the new optimum setting

was determined by the proposed model, around 50000 random samples were generated at the new settings to confirm that only around 0.988 % samples were below the seal strength of 27 and the results agreed.

The same optimization task was performed through grid search in a  $10 \times 10$  grid with around only 5000 samples in each grid-point.

## Conclusion

- Grid search processing took 20 times more time in comparison to the proposed method. So, we see a processing time difference of the order of 20 only for two input variables. The processing time difference increases exponentially for sampling based techniques with the increase in the number of input variables whereas with the proposed optimization method the computational time is quite robust to increase in input variables.
- The optimal setting for the mean proposed by the grid-search was 193 degree Celsius for Temp and 1.8 seconds for time of sealing which is a not the best optimal setting. With the proposed approach optimized mean values are as close to the theoretical optimized values as possible by the precision of the computer. For sampling the precision depends on the input parameters space explored and hence the parameters (i.e. mean of inputs for this problem) obtained might not be the most optimal one.

## Authors Contribution

The proposed solution is designed and implemented by author Santanu Pattanayak

## References

- *Mathematical Statistics by Irwin Miller / Marylees Miller*
- *Introduction to Mathematical Statistics by Robert V. Hogg/Joseph McKean/ Allen T. Craig*