**Tolerance Allocation: Balancing Quality, Cost, and Waste Through Production Rate Optimization**

Yue Wang a, Aihua Huang a, Carolyn A. Quigley b, Lei Li c, John W. Sutherland a

*a Environmental and Ecological Engineering, Purdue University, West Lafayette, IN 47906, USA*

*b School of Industrial Engineering, Purdue University, West Lafayette, IN 47906, USA*

*c School of Mechanical Engineering, Hefei University of Technology, Hefei 230009, China*

Abstract

Dimensional tolerance allocation is a very important and difficult task that traditionally seeks to balance cost/productivity and quality. Common tolerance allocation models have two shortcomings: i) they are overly reliant on models focused on minimizing cost and tend to ignore waste, and ii) they fail to connect to the root cause of many quality issues: process variation. This paper proposes a tolerance allocation model that addresses these shortcomings. The proposed model considers both product design (tolerance selection) and operation planning (or production rate selection). Relations among production rate, production cost, processing precision, and waste are considered. A gradient-based optimization method is proposed to minimize the cost and waste. A clutch assembly case study is analyzed to evaluate the method. Monte Carlo simulations are employed to validate the accuracy of the proposed cost model. The proposed method is compared with a heuristic method from the literature. The proposed method produced more satisfactory products at a lower cost while producing less waste. For the case study, it is found that when the precision of a process is high, it is not necessary from an economic standpoint to inspect the quality of individual components. For poor precision processes, inspecting the quality of individual components is the preferred approach from a cost/throughput standpoint.

Keywords:

Tolerance allocation, Cost minimization, Waste minimization.

|  |  |  |  |
| --- | --- | --- | --- |
| Nomenclature | | | |
| *β* | Product pass rate | *CS* ($) | Total scrap/recycle cost |
| *γ* | Component pass rate | *CT* ($) | Total production cost |
| *δ* | Design function sensitivity | *E* (mm) | Process related constant |
| *μ* | Process mean | *F* (mm∙min2) | Process related constant |
| *σ* | Process standard deviation | *L* | Number of scrapped components |
| *k* | Tolerance spread | *LS* | Lower specification limit |
| *r* | Production rate | *M* | Number of satisfactory products |
| *t* | Tolerance | *N* | Number of components processed |
| *x* | Characteristic value of a component | *Q* | Number of components assembled |
| *y* | Characteristic value of a product | *RN* | Remainder of Maclaurin series |
| *A* ($) | Fixed cost (set-up cost) | *SC* ($) | Scrap/recycle cost of a component |
| B ($/min) | Cost coefficient | *SN* | Maclaurin series |
| *BE* ($/min) | Electricity cost | *SP* ($) | Scrap/recycle cost of a product |
| *BL* ($/min) | Labor cost | *U* ($) | Average unit cost of a product |
| *BM* ($/min) | Machine tool cost | *US* | Upper specification limit |
| *CB* ($) | Processing cost | *W* | Number of unsatisfactory products |

# Introduction

The ability to produce high quality products with low cost and high production rate is critical for manufacturers. In addition, waste from manufacturing has become a severe environmental burden (Singh et al., 2017) (Lieder and Rashid, 2016), thus, reducing the production of waste streams and efficient use of material resources in manufacturing is ever more important from an environmental sustainability perspective (Haghighi and Li, 2018). Usually, high quality (precision), low cost, low waste, and high production rate are conflicting objectives, because excessive precision leads to excessive cost and processing time (Shin et al., 2010) (Sarkar and Saren, 2016). For example, to achieve high precision, a larger investment must be made in purchasing and maintaining highly precise machine tools and maintaining them more carefully at an associated higher cost (e.g., changing the tooling more frequently) (Leung and Hui, 2000). In addition, more precise machine tools are generally more complex and precise metrology equipment, which is more expensive. On the other hand, when inexpensive processes with low precision are applied, component-to-component variation will be large, and the resulting quality of assembled products may not meet the expectation of customers (Zhang and Huq, 1992). Also, when variation is large, components/products outside the specifications are likely rejected, and, if they cannot be reworked or recycled, they are scrapped (i.e., enter the waste stream), which is also wasteful in terms of energy consumption and resource depletion.

Owing to the ever increasing requirements of high-quality, low-cost products, and the awareness of sustainability, more and more efforts have been carried out to study the relations among product quality, cost, and waste reduction (Bhushan, 2013; Hegab et al., 2018; Mont et al., 2006; Mont, 2002; Taleizadeh et al., 2018; Wilson et al., 2014). This present research extends this effort by providing a tolerance allocation model that balances quality, time, cost, and waste through the optimization of production rate.

Tolerances are assigned for critical product/component characteristics during the engineering design process. Tolerance is the amount by which a characteristic value is allowed to deviate from the nominal value; it acknowledges that a manufacturing process cannot exactly realize a nominal value (Scholz, 1995). Most often, tolerances are selected based on product/component function considerations, as well as quality and cost. The tolerance on a component is used by manufacturing planners to select appropriate processes and their sequences (process planning) and the settings and tooling for each process (operation planning) (Cheng-Jung Lin and Hsu-Pin Wang, 1993; Kapur and Cho, 1994).

The tolerancing problem in engineering design has been widely studied by transforming it into a constrained optimization problem. The most common way of formulating such a problem is to establish allowable tolerances on a product based on functional considerations, e.g., the clearance between a shaft and hole must not be too small (this may inhibit assembly) and it must not be too large (this may not provide sufficient sealing). While the tolerances on a product are likely based on functional considerations and cost, since products are composed of components a designer must address the issue of how to allocate product tolerances to the components.

Traditional tolerance allocation methods allocate tolerances using a product/component design-oriented approach, in which almost all the focus is placed upon the tolerances of the product/components. Product engineering uses “tolerance” to communicate what is acceptable in terms of function. Manufacturing decision-makers must translate “tolerance” into their language because they think in terms of process variation or precision. The authors believe that rather than a single-minded focus on tolerance, that it is better to think in terms of process variation, since the same process settings that influence variation also determine cycle time, production cost, and amount of waste.

Few studies on tolerancing that consider the linkage between design and production have been carried out. For the few studies that have considered this linkage, attention was devoted to relating process variables and process variation for a specific situation – the connection among process variables, process variation, and tolerance was not considered in a broader sense. It is a relatively simple matter to describe how different conditions/variable settings affect the production rate for an operation, and production rate is strongly linked to the variation/precision and cost of an operation (Kim and Glock, 2018a). A planner has the latitude to explore different production rates (e.g., by adjusting process types and process variables), so long as the required tolerance is met (Kim and Glock, 2018b). For a given tolerance and associated process, if the maximum allowable production rate does not produce high enough throughput, or has too high a cost, then the tolerance will need to be loosened. This begins to illustrate the types of trade-offs that need to be made among quality, cost, and production rate (Gazzola et al., 2018; Glock, 2010; Khouja and Mehrez, 1994; Nicolaou et al., 2002).

To fill the research gap summarized above, we proposed a new tolerance allocation method. This method avoids the three problems of traditional tolerance allocation methods. One, not only cost, but also waste is minimized. This method avoids both unnecessary excessive precision and a large level of product/component scrap caused by quality or assembly issues. Two, instead of considering how to allocate a product tolerance by minimizing the sum of the costs associated with individual operations by changing process parameters, this paper optimizes the production rate for operations to achieve the product tolerance. Optimizing production rate makes the model easy to be generalized to different operations. Three, the linkage between design and production are considered. This is achieved by using statistical theory to characterize the relations among production rate, process precision, and process/component variation stack-up. A gradient-based optimization method is proposed to find optimal production rates that satisfies the tolerance requirement on a product and minimizes cost and waste. An overview of the research is shown in Fig. 1.

The remainder of the paper is structured as follows: a literature review is given in Section 2, it summarizes a few research gaps and how this research fill in the gaps. Section 3 gives a description of the problem and proposes a cost model. A method to allocate tolerance and a gradient-based optimization strategy are established in Section 4. Section 5 presents a case study, in which, the proposed method is compared to one from the literature. Finally, a Monte Carlo simulation is used to validate the accuracy of the analytical model. Section 6 concludes the paper.



Fig. 1 Overview of this research

# Literature review

Few studies on tolerance optimization consider the linkage between design and production. For the few studies that have considered this linkage, the impact of processing parameters on tolerance for specific machining processes are considered. Wang and Liang (2005) proposed a tolerance allocation method that minimizes machining cost. Based on the relation between cost and machining parameters, the tolerance of components, the machining sequences, and machining parameters are selected. Liu and Qiu (2011) integrated machining time into the tolerance optimization model. The methods that consider both design and production can help practitioners estimate how machining conditions for common machining methods affect the overall design of tolerance. Process variables, such as feed rate and cutting speed in machining or scanning speed in additive manufacturing, directly affect the cost and precision (variation) of a process (Sealy et al., 2016). A process usually has multiple variable settings (e.g., cutting speed, cutting depth, and feed rate in machining), and the impacts of each variable on the cost and process variation are different (Wang and Liang, 2005; Yan and Li, 2013). The coupling effect among different conditions/variables makes it difficult to find the optimal variable settings (for a variety of objectives), and the optimization result for one type of process are difficult to generalize to other types of processes.

Most of the studies on tolerance allocation consider product and process optimization separately. And most of the studies aims at minimizing total cost. Various forms of functions have been proposed to model the tolerance-cost relation (Chase et al., 1990; Chase and Greenwood, 1988; Dong and Hu, 1991; Sanz-Lobera et al., 2010; Speckhart, 1972; Sutherland and Roth, 1975). Based on these tolerance-cost models, the task of optimally allocating the tolerance is usually transformed to a constrained optimization problem. The objective of this optimization problem is to minimize cost, the constraints are tolerance requirements, and the optimization variables are tolerances of individual components.

One of the major difficulties on tolerance allocation is predicting how the variations of components stack-up into the overall variation of the product. Traditionally, this task has been solved using simple approaches such as a worst-case model or root sum square model (Wang and Liang, 2005) (Greenwood and Chase, 1988). Generally, simple models are only applicable to simple assemblies. For a complex assembly, tolerances optimized by simple models may result in high manufacturing costs due to excessively tight tolerances.

One of the most common ways to address the tolerance allocation problem for complex assemblies has been heuristic methods. Heuristic methods have the potential to find optimal or near-optimal solutions to complex problems. Singh et al. (2004) and Haq et al. (2005) used genetic algorithms to optimally allocate product tolerances. Zahara and Kao (2009) combined the Nelder-Mead simplex method with a particle swarm optimization method to minimize manufacturing cost and quality loss. Zhang et al. (2017) applied a particle swarm algorithm to satisfy tolerance requirements. While these heuristic algorithms have had success in tolerance allocation, a challenge with them is that they are very sensitive to tuning parameters, and unfortunately, such parameters are generally determined by trial and error. This limits the wide applicability of heuristic methods to industrial practitioners.

Some researchers applied optimization strategies in the tolerances allocation problem for complex assemblies. The method of Lagrange multipliers has been used to find optimal tolerances (Chase et al., 1990, 1990a; Kumar et al., 2016; Ramesh Kumar et al., 2016; Siva Kumar and Stalin, 2009). The Lagrange multiplier method is a strategy that transform a constrained optimization problem into an unconstrained problem. It helps find closed-form optimal tolerances. Tlija et al. (2019) combined the Lagrange multiplier method with a technique that evaluates the difficulties of manufacturing a given part. This combination enables designers to estimate the difficulty and cost of manufacturing the product by simulation. Another common method is the Lambert W function, which is usually used in physics. Shin et al. (2010) used the Lambert W function to find the tolerances that minimize the summation of manufacturing cost and rejection costs. (Sofiana et al., 2019) used a third-party software package to solve a tolerance optimization model. This model considered the impact of rework on the quality of product, and the impact of a profit-sharing policy, which may stimulate the commitment of suppliers in quality improvement.

Simulation-based strategies have also been widely used in predicting the stack-up of component variations. (Qureshi et al., 2012) proposed an iterative optimization procedure based on Monte Carlo simulation. Samples of given distribution is fist generated, and then the impact of component variation on the assembly is estimated by simulation. Wu et al. (2009) used Monte Carlo simulation in the statistical analysis and introduced a genetic algorithm to improve the efficiency of the estimation. (Hoffenson et al., 2015) provided survey-based tolerance allocation method that considers the economic and environmental impact of scrapping components and products. Design of experiment and Monte Carlo simulation are deployed to predict the stack-up of individual tolerance into a product. (Haghighi and Li, 2019) proposed a tolerance design method for additive manufacturing. This method estimates characteristics of population using a bootstrap statistical technique, which is based on simulation. (Huang et al., 2009) proposed a method that optimize tolerances on two stages. First, a tolerance model is built based on sampling, then a gradient-based strategy is used to optimize the model. (Rosyidi et al., 2016) and (Rosyidi et al., 2017) proposed a simulation-based method that considered a situation where the process capability of suppliers are a variable. A fuzzy quality loss function is included in the model to consider the cost related to the quality of products.

Other than the limitations summarized above, there are two major problems in the tolerance optimization methods mentioned above. First, these methods focus on minimizing cost and ignore the corresponding waste. Failure to consider waste, i.e., focusing exclusively on cost, may result in unnecessary scrapping/recycling because of excessively tight tolerances (unnecessarily rejecting components) or quality issues in assembly that result in product rejects. Second, these methods allocate tolerances using a product/component design-oriented approach, in which almost all the focus is placed upon the tolerances of the product/components, e.g., costs are modeled as a function of tolerance. Such a focus fails to capture the linkage between design and manufacturing. The advantages and disadvantages of the methods reviewed above are summarized in Table 1.

Table 1 Common tolerance optimization methods and limitations

|  |  |  |
| --- | --- | --- |
| Method | Advantage | Disadvantage |
| Optimize process parameters | Solve the problem from the root-cause, variation of processes. | Difficult to find the optimal variable settings. Results are difficult to generalize. |
| Heuristic | Can find near optimal tolerances. Do not need rigorous optimization procedures. | Results are sensitive to tuning parameters, which are determined by trial and error. Difficult to generalize. |
| Optimization strategies | Can be used in complex assemblies. No need of tuning parameters. | Statistical analysis is needed to build such a model, and the precision of the prediction depends on the model. |
| Simulation-based | No need to build statistical models to predict the accumulation of tolerances. The precision of the prediction is high. | Time-consuming. It generally needs simulation of large samples and many iterations. |
| Method of this paper | Have an overall consideration on quality, cost, and waste. The linkage between design and production are considered (considers both product design and operation planning.) | |

# Problem description and a cost model

This paper considers a tolerance allocation problem, i.e., how to best allocate the tolerance on a product assembly to the individual components. The tolerance of a component can be defined as the distance between the nominal value and the upper/lower specification limit. For a symmetric bilateral case, where the upper and lower specification limits have the same distance from the nominal value, the tolerance, *t*, is given by Eq. :

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, *LS* is the lower specification limit, *US* is the upper specification limit, *x*0 is the nominal value, *σ* is the process standard deviation, and *kσ* is the size of the tolerance (*k* will be referred to as the *tolerance spread*).

The tolerance allocation challenge can be illustrated by a shaft-hole assembly example, as shown in Fig. 2. It is assumed that only products/components with characteristic values between the lower/upper specifications, noted as satisfactory products, can be sold. Unsatisfactory products/components lying outside the specifications will be scrapped and managed as waste. The ratio between the number of satisfactory products and the total products assembled is the product pass rate, *β*. The shaft-hole clearance, i.e., characteristic value of the assembly, *y*, is the difference between the diameter of the shaft, *xs*, and the hole, *xh*. The processes for creating the shaft and hole are assumed to be normally distributed and are centered at the nominal values for the shaft and hole, *xs*0 and *xh*0 (the nominal value for the shaft/hole clearance is *y*0=*xh*0-*xs*0). The tolerance for the clearance is *ty*, which should be allocated to the tolerances on the diameters of the shaft, *ts*, and hole, *th*. Based on the allocated tolerances, manufacturing planners select appropriate processes settings and tooling to produce the shaft and hole. It is assumed that random assembly is employed, i.e., a shaft and a hole are each randomly selected and assembled. The shaft and hole each have a probability distribution associated with their size, and based on random assembly, the clearance also has a distribution (hole and shaft stack-up). In allocating the clearance tolerance to the tolerances on the shaft and hole diameters, statistical methods are needed to predict how the variations in the shaft and hole creation processes influence the clearance variation.

One strategy that might be considered by a manufacturer would be to use low precision processes to fabricate the components, which would keep manufacturing costs low. Then, a tight tolerance could be applied to the components to filter out poor quality components (this would lead to higher scrap costs/waste). This strategy would probably lead to a relatively high proportion of assembled products that satisfy the product specifications (less scrap products waste). Alternatively, process type and process parameter settings could be used to achieve precise processes, but this would come with a higher cost. However, this would likely lead to less scrap components/waste. In general, we should set tolerances on the components with an overall consideration of precision, cost, and waste.



Fig. 2 A shaft and hole tolerance allocation example

The production rate for a process depends on the type of process and the condition/variable settings for the process. The production rate in turn impacts the precision and cost of an operation (Kim and Glock, 2018b). Many studies have been carried out to study the relation between production rate, product quality, and processing cost (Glock, 2010) (Khouja and Mehrez, 1994). General relations between production rate (*r*), processing cost (*CB*), and process standard deviation (*σ*) are shown in Fig. 3(a) and (b). It may be noted that as the process standard deviation, *σ* (or variation, *σ2*) increases, the process precision erodes. Generally, the higher the production rate (usually achieved by increasing a process variable such as feed rate, step over, and cutting depth), the lower the processing cost. This is the case, since for high production rates more components are produced per unit time, and thus fixed costs are allocated across more components. However, many studies have shown that when a high production rate is applied, component-to-component variation is larger, i.e., quality and precision are decreased (Budak, 2006) (Fan et al., 2015). For example, consider a milling process; when a large feed rate is used, the quality of the machined part (evaluated by a criterion such as surface roughness or form error) will decrease (Wang et al., 2017). This work will propose a tolerance allocation method that optimizes the production rate, *r*, so as to balance quality, cost, and waste.

Many in industry use inspection-oriented approaches to identify and remove out of specification component to ensure quality. The tolerance and nominal value together serve to define the specification limits. During inspection, components within the specification limits are deemed satisfactory, and can be assembled. Otherwise, the components are rejected and scrapped (where they enter the waste stream or are recycled). Both the tolerance and process precision affect the component pass rate, *γ*, which is the ratio between the number of satisfactory components and the total number of manufactured components, as shown in Fig. 3 (b) and (c). The precision of the process depends on the production rate, with higher rates generally leading to reduced precision. The often utilized ±3*σ* tolerance band may also be modified to adjust the pass rate; herein we consider values other than “3” for *k* (tolerance spread). *k* may be adjusted to avoid unnecessary scrap or to avoid passing too many low-quality components, as shown in Fig. 3(c).



Fig. 3 The relation among production rate (*r*), processing cost (*CB*), precision (*σ*), and tolerances spread (*k*)

To find a trade-off among quality, cost, and waste, an optimal tolerance allocation method has to consider three factors: (1) the impact of production rate on the cost, variance of components, and waste caused by unsatisfactory components/products, (2) the stack-up of component variations for an assembled product, and (3) different quality management strategies, and their impact on cost and the quality of products/components. The average unit cost of a satisfactory product assembled, *U*, will be used to evaluate the economic and environmental performance (since the cost of waste is considered) of the production system. This work builds a cost model to calculate *U*, with these three factors considered.

To understand the cost model, it is necessary to define the relations between the number of products and components, under different quality inspection strategies. Consider an assembly process, in which *m* types of components are assembled into a product. For each component type, *Q* components are provided for assembly. All the products are assembled and are then inspected. The number of unsatisfactory products is *W*, and the number of satisfactory products is *M*. The summation of *W* and *M* equals to *Q*. For random assembly, a product that is assembled from satisfactory components may still have a characteristic value that falls outside its tolerance.

Three common component inspection strategies are considered: i) no inspection, ii) 100% inspection, and iii) acceptance sampling. If no inspection is carried out, all components are assembled. If 100% inspection is carried out, every component is inspected, and only the satisfactory components will be assembled, and the unsatisfactory components will be scrapped/recycled (Farooq et al., 2017). For acceptance sampling, inspection is carried out on a small subset of the components. If the qualities of the sampled components are acceptable, then all the components will be considered acceptable, and will be assembled. But if the subset of components is deemed to have poor quality, then 100% inspection is performed on all the components.

For both the case of no inspection and the case of acceptance sampling (when the sample of components pass inspection), the vast majority of components are not inspected. These two cases will serve as scenario one: “no component inspection.” For the case of 100% inspection and the case of acceptance sampling (when the sample of components fails inspection), every component is inspected. These two cases will serve as scenario two: “100% inspection.”

The assembly processes for both scenarios are shown in Fig. 4, which is illustrated by a product assembled from two types of components. In scenario one, there will be no waste associated with scrapping/recycling components. For each component type, the total number of components that are processed is equal to *Q*. In scenario two, every component is inspected. Only satisfactory components will be assembled, while unsatisfactory components are scrapped. For the ith component type, *Ni* components are manufactured, with *Li* not meeting the specifications and *Q* meeting the specifications. For both scenarios, the quality of every product is inspected. The numbers of satisfactory product and unsatisfactory products are *M* and *N*.



Fig. 4 Assembly problem

The average unit cost, *U*, of a satisfactory product, can be calculated as:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, *M* is the number of satisfactory products assembled. *CT* is the total cost, which includes the costs incurred in manufacturing and assembling all the components and managing the scrap. *CT* is defined as:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, *CB* is the total processing cost of all the components, and *CS* is the scrap/recycle cost of all the unsatisfactory products and unsatisfactory components.

The total processing cost, *CB*, is the summation of processing cost of all *m* types of components assembled into the product, and is given by the following equation:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, *i* is the component type index, *Ni* is the number of components of type *i* that are manufactured, and *CBi* is the processing cost per unit of component *i*.

The total scrap cost, *CS*, is

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, *SP* is the cost to scrap an unsatisfactory product, *W* is the number of unsatisfactory products, *SCi*  is the cost to scrap the *i*th component, and *Li* is the number of unsatisfactory components of type *i*.

# Methodology

In this section, a tolerance allocation method that minimizes the total cost by optimizing production rate is proposed. The influence of production rate on process precision and cost is considered. This method avoids both unnecessarily high and low process precision. Two assumptions are made for the tolerance allocation problem. First, the process is under statistical control (i.e., the process mean and variation are assumed stable), and the characteristic value of a component can be modeled as a random variable that follows a normal distribution, with the mean, *μ*, being equal to the design-specified nominal value, *x*0 (Devor et al., 2007; Otsuka and Nagata, 2017). Second, there are no constraints on the production time, i.e., for a specified production rate there is sufficient time to produce the required number of components (Geetha et al., 2015).

The component processing cost, *CB*, can be modeled as a function of production rate, *ri*:

|  |  |  |
| --- | --- | --- |
|  |  |  |

Similarly, the value of *σ* for a component can be modeled as a function of *r*:

|  |  |  |
| --- | --- | --- |
|  |  |  |

Equations and provide general functional forms for the cost and precision of a process. Some studies have provided general forms of the equations (Khouja and Mehrez, 1994). In Sec. 5.2, expressions are presented for these two equations.

Figure 4 summarizes how the production rate, *r*, and the tolerance spread, *k*, affect the average unit cost, *U*, through a computation graph. The relations among these variables are given in Eqs. -. For scenario one, no inspection is carried out for the components. Thus, tolerances are not needed for the components and the tolerance spread need not be defined; rather, the production rate (and thus precision, *σ*) for the component processes is used to control the precision. For scenario two, both production rate and tolerance spread, *k*, are controlled variables.

As shown in Fig. 5, for scenario one, the processing cost, *CBi*, and the variance, *σxi2*, of individual components is affected by the production rate for that process, *ri*. The variations of individual components *σxi2* will stack-up into the variations of a product, *σy2*, which affects the product pass rate *β* (when the products are inspected relative to the specifications). Both the scrap cost of products, *SP* and the number of satisfactory products, *M*, are determined by *β*. For scenario two, the pass rate, *γ*, for a given component is determined by *σxi2* and *k*; the product pass rate, *β*, determined by *σy2*; and the product tolerance, *ty* (a fixed value by design). The tolerance spread, *k*, affects both *σxi2* and *σy2*(more discussion on this is provided in Sec. 4.1.2). The pass rates for the components and the product determine the total scrap cost associated with the components and product.



Fig. 5 Logic flow of the cost model including key model parameters

## Statistical analysis of variation stack-up

For a product assembled from *m* types of components, its characteristic value, *y*, is determined by the characteristic values of components assembled, noted as *x*1, *x*2, …, and *xm*, through the design function *y*=*f*(*x*1, *x*2, …, *xm*). Usually, different types of components are processed independently, so variables *x*1, *x*2, …, and *xm* can be modeled as independent variables. Herein, we assume that *xi* is normally distributed. Let us also assume that *y* is a normally distributed random variable (because random assembly is deployed, also, this assumption is verified in the case study).

The number of satisfactory products, *M*, can be estimated by:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, *β*(*r*1, *r*2,… *rm*) is the pass rate of products, which is a function of the production rate of each process, *ri*. Since the product distribution is unbiased (i.e., the product distribution is centered at the nominal value, *y*0) and a bilateral tolerance is used, the function *β*(*r*1, *r*2,… *rm*) can be evaluated with Eq. (9).

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, *erf*(*x*) is the Gauss error function, and *ty* is the tolerance of the product.

The number of unsatisfactory products, *W*, can be computed given the pass rate of the product, as shown below:

|  |  |  |
| --- | --- | --- |
|  |  |  |

Two options are available to predict how the variations in individual components stack-up in the product: variation simulation by Monte Carol simulation and statistical theory-based approach. Monte Carlo simulation predicts the distribution associated with an assembly as parts are randomly drawn from distributions associated with each individual component and virtually assembled. Monte Carlo simulation may lead to excessive computation time. The statistical theory-based approach relates the component variances *σxi*2 to the product variance *σy*2 using a first order approximation of the design function. This is shown in Eq. (Kawlra and Hancock, 1996):

|  |  |  |
| --- | --- | --- |
|  |  |  |

where *δi* is the partial derivative of *y* with respect to *xi*(the design function sensitivity withrespect to *xi*):

|  |  |  |
| --- | --- | --- |
|  |  |  |

In Eq. it is assumed that the design function *f*(*x*1, *x*2,… *xm*) is differentiable. If *f*(*x*1, *x*2,… *xm*) is not differentiable, the *δi* values may be estimated using a numerical approximation.

### Scenario one: no inspection of components

In this subsection, scenario one is considered: none of the components are inspected (or, for the case of acceptance sampling, very few). The number of components produced, *Ni,* is equal to *Q*. The number of scrap components of type *i*, *Li*, is equal to 0. Thus, the scrap cost of unsatisfactory components is 0, and the total scrap cost can be simplified to:

|  |  |  |
| --- | --- | --- |
|  |  |  |

By placing Eqs. , , and into Eq. , the following expression is obtained for the average unit cost, *U*:

|  |  |  |
| --- | --- | --- |
|  |  |  |

The production rate, *ri*, impacts the average unit cost by affecting the processing cost and the component variation stack-up (evaluated by *σy*).

### Scenario two: 100% inspection of components

In this subsection, scenario two is considered: every component is inspected and judged as satisfactory/not satisfactory. The number of components of the *i*th type that must be processed to produce Q satisfactory components is *Ni*, and can be calculated using:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where *γi* is the pass rate for component type *i*, which is affected by both the variance (precision), *σi*2, of the process and the tolerance spread, *ki*, of the component. Similar to computing the pass rate of a product, *β*, the pass rate of a component, *γi*, can be evaluated using a normal distribution.Since the process is unbiased (mean of process, *μi*, is equal to the nominal value *xi*0) and a symmetrical bilateral tolerance is used, the component pass rate can be evaluated with Eq. :

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, *kiσi* is the component specification (tolerance) (“No Title,” n.d.).

The number of components scrapped, *Li*, is

|  |  |  |
| --- | --- | --- |
|  |  |  |

By placing Eqs. - into Eq. , the average unit cost, *U*, can be represented as a function of *ri* and *ki*, is defined as:

|  |  |
| --- | --- |
|  |  |

Through inspection unsatisfactory components will be removed, and the resulting distribution of “passed” components will follow a truncated normal distribution. The standard deviation of the truncated normal distribution, *σxi*′, is smaller than the standard deviation of the distribution before truncation, *σxi*. The value of *σxi*′ can be computed from *σxi* using the following expression (Devor et al., 2007):

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, *h*(*ki*) is a function of *ki*, and is given by:

|  |  |  |
| --- | --- | --- |
|  |  |  |

and *T*(*ki*) and *φ*(*ki*) are functions of *ki*, and can be expressed as:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  |  | |
|  |  | | |  |

The function *h*(*ki*) is shown in Fig. 6. When *ki* (tolerance spread) increases, meaning that a larger tolerance is assigned and fewer components are scrapped, then *σxi*′ also increases and approaches *σxi*. When *ki* decreases, a tighter tolerance is applied and more components are scrapped, and thus the value of *σxi*′ decreases.

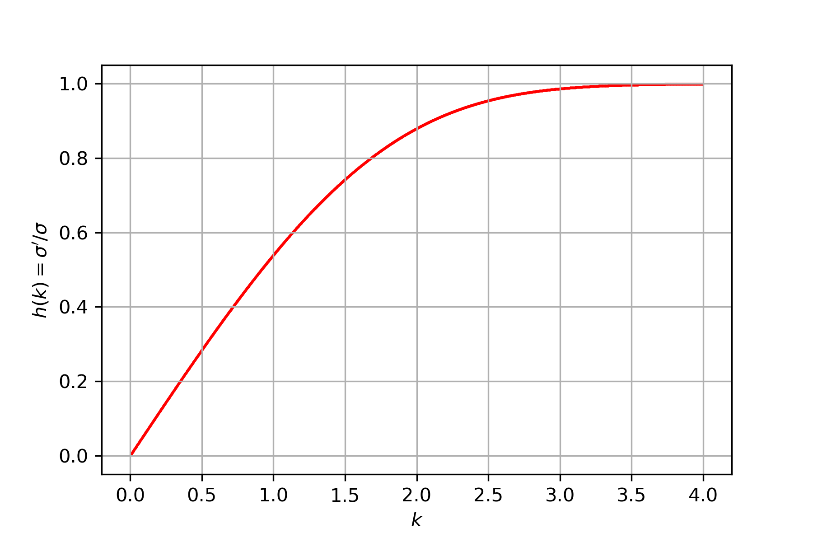


Fig. 6 Behavior of *h*(*k*), ratio of standard deviation of truncated normal distribution to standard deviation of starting distribution

Figure 5 shows how the values of *ki* for the individual components affects their respective standard deviations. The collective effect of these *ki* values on the standard deviation of the assembled products is given by the following function:

|  |  |  |
| --- | --- | --- |
|  |  |  |

Furthermore, let us assume that the shape of the distribution of the characteristic value of the product, *y*, is not too dramatically affected by truncations in the distributions of component values. This assumption will be verified through the use of Monte Carlo simulations in Sec. 5.3.3.3.

## Optimization

The average unit cost, *U*, can be minimized by optimizing manufacturing rates of components and the tolerance spread. As a reminder, for a given process, the production rate affects the cost and the precision of the process. This is an unconstrained multivariate optimization problem. The problem can be solved using a variety of optimization methods, including gradient based optimization algorithms (e.g., method of moving asymptotes) (Svanberg, 2002). Compare to optimization methods such as Monte Carlo method and Heuristic method, gradient-based methods have advantages such as easy to implement, low storage requirement, and easy to generalize (no need of parameter tuning). A gradient-based optimization procedure was used in the present research.

The procedure of the gradient-based optimization method is straightforward. We first find the partial derivatives of the average unit cost, *U*, with respect to all variables (for scenario one, only production rates; for scenario two, both production rates and tolerance spreads). Then the partial derivatives with respect to intermediate variables and approximations for non-elementary functions are needed. The details are given in the sections below.

### Preparing for optimization of scenario one

For scenario one, *U* only depends on the production rates, *ri*, and is defined by Eq. . For the gradient based optimization procedure, we must be able to evaluate the partial derivative of *U* with respect to *ri*, with *σi* being the intermediate variable. The derivative of *U* with respect to *ri* is:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where z is given by:

|  |  |  |
| --- | --- | --- |
|  |  |  |

As the standard deviation of the product, *σy*, is dependent on the production rates of individual components, the partial derivative of *σy* with respect to *ri* is:

|  |  |  |
| --- | --- | --- |
|  |  |  |

With Eqs. (24)- in place, the values of *ri* may be optimized to minimize *U*.

### Preparing for optimization of scenario two

For scenario two, *U* depends on both *ri* and *ki*, and is defined by Eq. . Again, the gradient based optimization procedure requires values for the partial derivative of *U* with respect to *ri* and *ki*, with *σi* being the intermediate variable. The derivative of *U*, with respect to *ri* is:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, z is:

|  |  |  |
| --- | --- | --- |
|  |  |  |

The partial derivative of *σy*, with respect to *ri* is given by:

|  |  |  |
| --- | --- | --- |
|  |  |  |

and the derivative of *U* with respect to *ki* is:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where

|  |  |  |
| --- | --- | --- |
|  |  |  |

### Approximation of the Gauss error function and its derivative

The Gauss error function, erf(*x*), is a non-elementary function, its value can be approximated using an nth order Maclaurin series:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, S*N*(x) is:

|  |  |  |
| --- | --- | --- |
|  |  |  |

and *RN* (*x*) is the remainder. By using the Leibniz Criterion, the remainder satisfies the following inequality equation:

|  |  |  |
| --- | --- | --- |
|  |  |  |

Similarly, the derivative of the Gauss error function is:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, *DN*(*x*) is the remainder of the derivative of the Gauss error function, which satisfies the following inequality equation:

|  |  |  |
| --- | --- | --- |
|  |  |  |

The upper bound of approximation errors (remainders) *RN* and *DN* can be estimated by Eqs. and . The larger the value of *n*, the smaller the approximation error. For any given maximum allowable error, there exists a value of *n* that will satisfy this requirement.

# Case study

To validate the proposed method, an overrunning clutch product is considered (Feng and Kusiak, 1997). This is an example commonly used in the literature on tolerance allocation. Using this example makes it easier to compare different methods. Some traditional optimization methods have been applied in solving the problem, for example, worst case tolerance analysis (Greenwood and Chase, 1988) and brute-force search (Choi et al., 2000). Because the simple optimization models could not find optimal/near optimal solutions, these methods tend to generate tight tolerance. Some heuristic based algorithms have also been used, such as particle swarm optimization (Zahara and Kao, 2009) (Haq and Saravanan, 2006) and genetic algorithm (Singh et al., 2004), these algorithms usually requires the tuning of optimization parameters, which are difficult to be compared. Thus, the proposed method is compared with a heuristic method from the literature that does not need parameter tuning. The advantages of the proposed methods are analyzed in the comparison.

To assess the accuracy of the proposed statistical approach, as described in Eqs. -, a Monte Carlo simulation is employed. Analyses are carried out to compare the two scenarios (no component inspection and 100% component components).

## Problem description

An overrunning clutch is assembled from three types of components, i.e., hub, roller, and cage. For proper functioning, the contact angle (product characteristic), *y*, of an overrunning clutch should be within ±0.035 rad from the nominal value of *y*0=0.122 rad (*ty* =0.035). The value of *y* for an assemblage of components is given by the following design function:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, *x*1 is the diameter of the hub, *x*2 is the diameter of the roller, and *x*3 is the inner diameter of the cage (please refer to Fig. 7). The nominal values of *x*1, *x*2, and *x*3 are 55.29 mm, 22.86 mm, and 101.69 mm. The tolerance of the clutch must be allocated to the three components.



Fig. 7 An overrunning clutch (Feng and Kusiak, 1997)

## Rate-cost and rate-sigma relationships

In the literature, production rates have been in such terms as material removal rate (Yan and Li, 2013) and units per unit time (Glock, 2010). In this case study, the production rate, *r*, is expressed in terms of the number of components produced per minute. For a single component, the processing cost, *CB*, is:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, *A* is the fixed cost per part such as set-up cost, and *B/r* is a rate-dependent cost. The values of *A* for the three components are shown in Table 2.

Turning attention to the production rate-dependent term in the processing cost, its coefficient, *B*, consists of the following:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where, *BM* is the cost of the machine tool (includes machine depreciation and operating cost), *BL* is the cost of labor, and *BE* is the cost of electricity. These three coefficients are estimated using typical cost rates for use of machine tools, labor, and electricity (all expressed in $/min), as shown in Table 3. The relation between *r* and *CB* is shown in Fig. 8.

Table 2 Values of the fixed cost, *A*

|  |  |
| --- | --- |
| Component | *A* ($) |
| Hub | 0.98 |
| Roller | 0.52 |
| Cage | 1.22 |

Table 3 Values of *B*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Component | *BM* ($/min) | *BL* ($/min) | *BE* ($/min) | *B* ($/min) |
| Hub | 0.142 | 0.500 | 1.494 | 2.136 |
| Roller | 0.150 | 0.500 | 1.150 | 1.800 |
| Cage | 0.368 | 0.500 | 1.700 | 2.568 |



Fig. 8 The relation between *r* and *CB*

Let us assume that machining processes are used to produce the components. In machining, feed rate is a major factor that affects production rate. It also affects the precision of the process, as it is related to measures such as cutting forces, which influence deflection (Budak and Altintas, 1995). We propose that the precision (*σ*) is linearly related to the square of feed rate. This hypothesis is supported by several studies from the literature. For example, Yeh and Hsu showed that the tolerance value of a chord error, *G*, is linearly related to the square of thefeed rate of a CNC machine (Yeh and Hsu, 2002). Boothroyd and Knight demonstrated that the roughness (*Ra*) of a process is linearly related to the square of feed rate (Boothroyd and Knight, 1989). Lim and Meng expressed the cutting force as a second degree polynomial function of the feed rate (Lim and Meng, 1997). Based on these relations from the literature,the following expression is proposed to describe the value of *σ*:

|  |  |  |
| --- | --- | --- |
|  |  |  |

where *Ei* and *Fi* are coefficients corresponding to the process used to fabricate the ith component type. *Ei* provides a lower bound for *σ* (the best precision the process can achieve), and *Fi* affects the shape of the curve. Fig. 9 shows the assumed behavior of *σ* as a function of *r* for the given case (adapted from (Choi et al., 2000)). The values used for *Ei* and *Fi* in the current study are given in Table 4.

Table 4 Values of coefficients for *Ei* and *Fi*

|  |  |  |
| --- | --- | --- |
| Component | *Ei* (mm) | *Fi* (mm∙min2) |
| Hub | 0.0320 | 4×10-4 |
| Roller | 0.0214 | 2×10-4 |
| Cage | 0.0534 | 6×10-4 |



Fig. 9 The relation between *r* and *σ*

## Results and analysis

For the given design function, Eq. , the derivatives, *δ*1, *δ*2, and *δ*3, may be determined using the general expression of Eq. . As has been noted, these derivatives are needed to compute *σy* in Eq. . For the specified design function (Eq. ), the derivative relationships are:

|  |  |  |
| --- | --- | --- |
|  |  |  |

|  |  |  |
| --- | --- | --- |
|  |  |  |

|  |  |  |
| --- | --- | --- |
|  |  |  |

For the ith component type, to optimize the production rate values, *ri*, across all values for *i*, the derivative of *Cpi* and *σi*, with respect to *ri* must be determined. Based on the relation between *r* and *Cp*, as given by Eq. , and the relation between *r* and *σ*, as given by Eq. , the sensitivity of the processing cost, *CB* and process variance, *σ*2, to changes in the production rate for each process are:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

The open source optimization library, NLopt (“NLopt,” 2020), was employed to solve the cost optimization problem.

### Scenario one: no inspection of components

First, scenario one, as described in Sec. 4.1.1, was considered. Recall that this scenario either does not inspect individual components or uses acceptance sampling (and based on the sample, the lot is judged to be satisfactory). The components are assembled to create a product, and then these products are inspected. An additional cost is incurred if a product is deemed to be unsatisfactory. This cost, *SP*, is associated with the additional expense of scrapping a product. The value of *SP* was assumed to be ten percent of the summation of the fixed costs, *A*, of all components that are assembled into the product. The average unit cost of a satisfactory product, *U*, is defined by Eq. . The heuristic algorithm-based tolerance allocation method proposed by Wang et al. (Wang et al., 2019) was compared with the method proposed in this paper. The heuristic method (Wang et al., 2019) allocated product tolerances by optimizing the *σ* of the processes. The number of each type of component that was processed and assembled, *Q*, was set equal to 10,000.

The results of the Wang et al. heuristic method (Wang et al., 2019) and the approach proposed in this paper are shown in Table 5. Using the method proposed in this paper, a lower average unit cost per satisfactory product is achieved (a $150 difference for 10,000 parts). Furthermore, 235 more satisfactory products can be produced per 10,000 units of components processed (≈2%).

Table 5 Results of scenario one

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Wang et al. heuristic  method (Wang et al., 2019) |  | Method of this paper | |
| Component | *σ* |  | *r* | *σ* |
| Hub | 0.077 |  | 10.909 | 0.080 |
| Roller | 0.046 |  | 10.149 | 0.042 |
| Cage | 0.118 |  | 9.284 | 0.105 |
| *U* ($) | 3.5002 |  | 3.485 | |
| *M* | 9410 |  | 9645 | |

### Scenario two: 100% inspection of components

This section considers scenario two that was described in Sec. 4.1.2. Recall that this scenario inspects every component prior to assembly. The cost to scrap/recycle unsatisfactory components, *SC*, is assumed to be ten percent of the fixed cost, *A*, of the process used to fabricate the component. As in Sec. 4.3.1, the cost to scrap/recycle unsatisfactory products, *SP*, is assumed to be ten percent of the summation of the fixed costs, *A*, of all components that are assembled into the product. The settings of the Monte Carlo simulation are the same as in Sec. 5.3.1. The results using the method of this research are again compared with the heuristic algorithm-based tolerance allocation method (Wang et al., 2019). The results of the two methods are shown in Table 6.

Compared to a heuristic method from the literature (Wang et al., 2019), the method proposed in this paper achieves a smaller average unit cost per satisfactory product (a $160 difference for 10,000 parts), and assembled 254 more satisfactory products per 10,000 satisfactory products produced (≈2.5%). The two methods have similar component pass rates. In the heuristic method, the production rate (derived from *σ*) and tolerance spread, *k*, had to be optimized separately, while in the proposed method of this research, the production rate and tolerance spread were optimized together. Thus, a better solution was reached in terms of cost, quality, and waste reduction (material efficiency) using the method of this research.

Table 6 Results of scenario two

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Wang et al. heuristic method  (Wang et al., 2019) | | |  | Method of this paper | | | |
| Component | *σ* | *k* | *N* |  | *r* | *σ* | *k* | *N* |
| Hub | 0.077 | 3.484 | 10004 |  | 10.9 | 0.080 | 4.071 | 10000 |
| Roller | 0.046 | 3.785 | 10001 |  | 10.2 | 0.042 | 3.030 | 10024 |
| Cage | 0.118 | 3.723 | 10003 |  | 9.3 | 0.105 | 3.535 | 10003 |
| *U* ($) | 3.500 | | |  | 3.484 | | | |
| *M* | 9401 | | |  | 9655 | | | |

Several observations associated with the two scenarios, as provided in Tables 4 and 5, merit attention. The results of both methods for both scenarios produce a relatively high percentage of unsatisfactory assembled products (about 3%). Across all cases, the objective of minimizing the average unit cost of a satisfactory product, *U*, is pursued. The obtained solutions are optimal from a cost perspective given the capabilities (precisions) and costs of the processes. These results can help practitioners adjust the production process. For example, if the current scrap rate is unacceptable, then this may indicate that the capabilities of the processes need to be improved or that the scrap cost is not large enough in the model.

### Further analysis

#### Comparison of the two scenarios

In an ideal world, it would be desirable for a process to be sufficiently precise so that no inspection of components is needed. However, when the precision of the process is limited, or when it is expensive to increase the precision of the process, inspection may be a cost-effective strategy to ensure that the components/products have acceptable quality. Comparison of the cost for the two scenarios will help practitioners make a decision as to whether to carry out inspection. As is evident from a comparison of the results of Tables 4 and 5, the inspection of components, does not dramatically affect either *U* or *M*. However, this result may be dependent on the precision (*σ*) of the processes. For processes with poorer precision (bigger *σ*), the difference may be larger.

In this section, the lower bound for *σ* of processes is varied, and the average unit cost per satisfactory product of the two scenarios are compared. For scenario two, two cases are considered. For the first case of scenario two, the value of *k* is fixed at 3 (i.e., specifications are ±3*σ*), and for the second case of scenario two, the value of *k* is optimized.

In the production rate-*σ* model given by Eq. , the value of *E* is the lower bound on the standard deviation for the process. The precision of the process was varied by multiplying the value of *E* given in Table 3 by a “precision scaling constant.” The value of the precision scaling constant was varied from 0.5 to 3.0 (a value of 0.5 reduces the lower bound on *σ* by ½ and a value of 3.0 increases the low bound on *σ* by a factor of 3). The value of *F* was fixed at the value given in Table 3. For both scenarios, the cost-production rate relationship was the same as used previously, and the values of the constants *A* and *B* in Eq. were the same as in Table 2. Because the cost to inspect the component is independent of production rate, and the component scrap cost is negligible (very few components are scrapped), the inspection cost and component scrap cost were both assumed to be 0. The minimum values of *U* for scenario one and scenario two (two cases for scenario two) for different precision scaling coefficients are shown in Fig. 10. The figure shows that for processes with high precision, the difference among the three cases is small, so the economic benefit of inspection is negligible. As the precision of the process decreases (*σ* increases), the economic benefits of inspection increase. Optimizing the value of *k* can reduce *U* – the average unit cost per satisfactory product.



Fig. 10 Comparison of two scenarios

#### Influence of precision on tolerance spread and pass rate

This section analyzes the influence of process precision on the allocated tolerance of components. As with Sec. 5.3.3.1, the precision of the process was varied by multiplying the value of *E* given in Table 3 by a precision scaling constant, which was varied from 0.5 to 3.0. The value of *F* was fixed at the value listed in Table 3. For all three cases, the values of the constants *A* and *B* in Eq. are the same as given in Table 2. The values of *γ*, *β*, and *k* for minimum *U* are shown in Fig. 11. A general trend for the components is that when the process precision decreases (*σ* increases), the value of *k* also decreases, which means a tighter tolerance is applied to the components.

The value of *k* and the pass rate of the roller are the smallest among the three components, which means the tolerance of the roller is the tightest, even though the process for the roller has the highest precision. Since the process for the roller has the lowest processing cost, tightening the tolerance of the roller reduces the loss caused by assembling expensive hub and cage components that may be scrapped if the product fails inspection due to large variation stack-up. The optimization seems to be telling us that driving down the stack-up variation by focusing on the inexpensive rollers is the least expensive way to affect the average unit product cost, *U*.



Fig. 11 Relation among process precision, tolerance spread, and pass rate

#### Accuracy of the cost model

In the previous subsections, the optimization results, such as the average unit product cost, *U*, and the number of satisfactory products *M*, were calculated by the proposed analytical model. This section reports on the Monte Carlo simulations that were used to validate the accuracy of the analytical model. Model validation was achieved by comparing the relative error between the average unit product cost estimated by the Monte Carlo simulations and the average unit product cost computed by the proposed cost model (scenario one: Eq. , scenario two: Eq. ).

To review, the proposed optimization method finds the optimal production rates and the minimum average unit cost of a satisfactory product. For the Monte Carlo simulations, the production rates found via the proposed method were used to generate components of each type. The dimension of each component type was assumed to follow a normal distribution (with mean equal to the nominal value, and standard deviation functionally dependent on the production rate).

Both scenario one and scenario two were considered. For scenario one, 10,000 components of each type were generated. These components were randomly assembled, the number of satisfactory products was counted, and the average unit cost of a satisfactory product was computed. For scenario two, components of each type were again generated, each component was inspected, and unsatisfactory components were scrapped. This process continued until 10,000 satisfactory components of each type were produced. The 10,000 satisfactory components of each type were randomly assembled, the number of satisfactory products was counted, and the average unit product cost was computed.

To validate that the accuracy of the model does not change dramatically with the change of process precision, the Monte Carlo simulations were carried out multiple times with varying precision (multiplying the value of *E* given in Table 3 by a precision scaling constant, which was varied from 0.5 to 3.0). The conditions for the simulations were the same as those given in Secs. 5.3.3.1 and 5.3.3.2. The relative errors, (Monte Carlo – analytical model)/Monte Carlo, are shown in Fig. 12. For both cases, the relative errors were all positive, which means the cost predicted by Monte Carlo simulations was larger than the theoretical model. The error being all positive means the product variance estimated from component variances is slightly smaller than the real value. The relative errors in both cases are modest (average of about 1%) and may be compensated by slightly increase the estimated product variance in Eq. .

|  |  |
| --- | --- |
|  |  |

Fig. 12 Errors between Monte Carlo simulations and analytical model: scenario one (left) and scenario two (right)

#### Managerial insights

The case study has shown the advantages of jointly considering product design and operation design in allocating tolerance. Based on the analysis from the above three subsections, some managerial insights on inspection strategy, cost reduction, and waste reduction are provided.

The results in Table 6 shows that tightening the tolerance of components that is less expensive to process (inexpensive components) and loosening the tolerance of components that is more expensive to process (expensive components) is an effective way to reduce the total cost. Though a tighter tolerance may initially look to increase the cost, this strategy reduces the loss caused by assembling expensive components into a product that fails inspection. Especially when the failing is caused by large variation stack-up from the variation of inexpensive components. It should also be noted that this strategy in saving cost should not be exploited because unnecessary tight tolerance may cause unnecessary precise process to be used, which increases the energy consumption (low production rate and long production time) and increases waste (more components are scrapped).

Quality inspection is generally time consuming and costly. Practitioners must make decisions on whether to carry out inspection based on the precision of the process and the cost of inspection. The comparison in 5.3.3.1 only considered the costs that are directly affected by tolerance and production rate, so the result reveals how the precision of the process affect the cost. The comparison gives practitioners some guidance on the level of process precision at which inspection of components becomes economically desirable. Practitioners should evaluate cost related to inspection, such as investment on equipment and labor, to make decisions.

Costs to scrap products and component are generally low (Omachonu et al., 2004). If the solution to tolerance allocation is driven only by high quality and low cost, a manufacturer may take a strategy that uses low precision processes to produce the components, which would keep manufacturing costs low. Then, a tight tolerance could be applied to the components and products to filter out poor quality assemblies. This strategy would lead to a higher waste. If the manufacturer would like to reduce waste in addition to cost, they can add a “punish” factor into the cost model. The “punish” factor can be represented by the ratio between product scrap cost *SP*, and fixed cost, *A* (or similarly, the ratio between component scrap cost, *SC*, and *A*). A high ratio has a higher punish on waste. We compared three cases, with three ratios being 10%, 50%, and 100%. The results are shown in Table 7. It is evident that when the punish factor is increased, less products and components go into the waste stream. Practitioners can adjust this factor based on requirement.

Table 7 Impact of scrap cost on cost and waste

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *N* | | | *U* ($) | *M* |
| Hub | Roller | Cage |
| 10% | 10000 | 10000 | 10000 | 3.484 | 9655 |
| 50% | 10024 | 10009 | 10000 | 3.514 | 9701 |
| 100% | 10003 | 10000 | 10000 | 3.544 | 9746 |

# Summary and conclusions

This paper introduced a new tolerance allocation model that considered production cost, quality, and waste simultaneously. This model for the first time, jointly considers product design and operation design. For product design, a statistical approach was used to predict how component variations contributed to the variation of an assembled product. For operation design, the relations among production rate, processing precision, processing cost, and waste were characterized.

An analytical cost model was proposed. The cost model considered processing cost and scrap cost. Two scenarios were studied: i) no component inspection, and ii) 100% inspection of components (assembled products were always inspected). The tolerance on the product characteristic of interest was allocated to individual components by optimizing the production rate for each component (the production rate affects the processing cost and precision). For the scenario where components were inspected, the tolerance spread, *k* (tolerances are ±*kσ*), was also optimized. Since component inspection may change the distribution of characteristic values, an adaptation function was introduced to appropriately adjust the standard deviation of components. A gradient-based optimization method was used to minimize the cost. Tolerance allocation for a clutch assembly was used to demonstrate the proposed method. The obtained results were compared with the results of a heuristic-based method. It is shown that the proposed method leads to settings (production rate and tolerance spread) that produce more satisfactory products at a lower cost and produce less waste. In addition, model validation was conducted by comparing the relative error between the average cost computed from Monte Carlo simulations and the average cost computed by the proposed theoretical cost model.

Some conclusions and practical guidance may be drawn from this work as follows.

* Monte Carlo simulations demonstrated the accuracy of the analytical cost model (average of about 1% error rate).
* The case study showed that the proposed method can optimize tolerance by balancing cost, quality, and waste. Compared to a heuristic method from the literature, the proposed method produces more satisfactory products at a lower average unit product cost and lower waste (fewer scrapped/recycled components/products).
* When the precision of a process is high, it is more economical not to inspect the quality of individual components. For poor precision processes (large *σ*), inspecting the quality of individual components is the preferred approach from cost/throughput standpoint. In the long term, continuous quality improvement should be pursued to improve process precision and minimize waste.
* The cost model and the optimization method were developed based on general functional forms for the rate-cost and rate-sigma relationships. Two specific forms of these relationships were considered in the case study. Manufacturers can establish these relationships that were suitable to their situation of interest.

The ability to produce high quality products with low cost and high throughput rate is critical for manufacturers. This research has demonstrated how allocating product tolerance by jointly considering product design and operation design can help improve this ability while minimizing scrap. In moving forward, it should be noted that continuous improvement approaches (a topic not addressed in this paper) that reduce process variation while also reducing costs are likely to add to the benefits of this research, and produce even smaller levels of scrap/waste.

Acknowledgement

The authors gratefully acknowledge the support from the Discovery Park Big Idea Challenge Grant (F.00038152.06.005) at Purdue University, and the Indiana Next Generation Manufacturing Competitiveness Center (IN-Mac) (F.00139086.02.004). Author Sutherland substantially benefited from discussions with the late Professor Tsong-how Chang for his thoughtful insights into tolerancing.

References

Bhushan, R.K., 2013. Optimization of cutting parameters for minimizing power consumption and maximizing tool life during machining of Al alloy SiC particle composites. J. Clean. Prod. 39, 242–254. https://doi.org/10.1016/j.jclepro.2012.08.008

Boothroyd, G., Knight, W.A., 1989. Fundamentals of Metal Machining, 2nd ed. Marcel Dekker, New York.

Budak, E., 2006. Analytical models for high performance milling. Part I: Cutting forces, structural deformations and tolerance integrity. Int. J. Mach. Tools Manuf. 46, 1478–1488. https://doi.org/10.1016/j.ijmachtools.2005.09.009

Budak, E., Altintas, Y., 1995. Modeling and avoidance of static form errors in peripheral milling of plates. Int. J. Mach. Tools Manuf. 35, 459–476. https://doi.org/10.1007/BF01654345

Chase, K., Greenwood, W.H., Loosli, B.G., Hauglund, L.F., 1990a. Least cost tolerance allocation for mechanical assemblies with automated process selection. Manuf. Rev.

Chase, K., Greenwood, W.H., Loosli, B.G., Hauglund, L.F., 1990b. Least cost tolerance allocation for mechanical assemblies with automated process selection. Manuf. Rev. 49–59.

Chase, K.W., Greenwood, W.H., 1988. Design issues in mechanical tolerance analysis. Manuf. Rev. 1, 50–59. https://doi.org/10.1.1.65.2404

Cheng-Jung Lin, Hsu-Pin Wang, 1993. Optimal operation planning and sequencing: Minimization of tool changeovers. Int. J. Prod. Res. 31, 311–324. https://doi.org/10.1080/00207549308956727

Choi, H.-G.R., Park, M., Salisbury, E., 2000. Optimal Tolerance Allocation With Loss Functions. J. Manuf. Sci. Eng. 122, 529. https://doi.org/10.1115/1.1285918

Devor, R.E., Chang, T., Sutherland, J.W., 2007. Statistical Quality Design and Control, 2nd ed. Prentice Hall, New Jersey.

Dong, Z., Hu, W., 1991. Optimal process sequence identification and optimal process tolerance assignment in computer-aided process planning. Comput. Ind. 17, 19–32. https://doi.org/10.1016/0166-3615(91)90100-N

Fan, W., Lee, C.H., Chen, J.H., 2015. A realtime curvature-smooth interpolation scheme and motion planning for CNC machining of short line segments. Int. J. Mach. Tools Manuf. 96, 27–46. https://doi.org/10.1016/j.ijmachtools.2015.04.009

Farooq, M.A., Kirchain, R., Novoa, H., Araujo, A., 2017. Cost of quality: Evaluating cost-quality trade-offs for inspection strategies of manufacturing processes. Int. J. Prod. Econ. 188, 156–166. https://doi.org/10.1016/j.ijpe.2017.03.019

Feng, C.X., Kusiak, A., 1997. Robust tolerance design with the integer programming approach. J. Manuf. Sci. Eng. Trans. ASME 119, 603–610. https://doi.org/10.1115/1.2831193

Gazzola, G., Choi, J., Kwak, D.S., Kim, B., Kim, D.M., Tong, S.H., Jeong, M.K., 2018. Integrated Variable Importance Assessment in Multi-Stage Processes. IEEE Trans. Semicond. Manuf. 31, 343–355. https://doi.org/10.1109/TSM.2018.2853586

Geetha, K., Ravindran, D., Kumar, M.S., Islam, M.N., 2015. Concurrent tolerance allocation and scheduling for complex assemblies. Robot. Comput. Integr. Manuf. 35, 84–95. https://doi.org/10.1016/J.RCIM.2015.03.001

Glock, C.H., 2010. Batch sizing with controllable production rates. Int. J. Prod. Res. 48, 5925–5942. https://doi.org/10.1080/00207540903170906

Greenwood, W.H., Chase, K.W., 1988. Worst case tolerance analysis with nonlinear problems. J. Manuf. Sci. Eng. 110, 232–235.

Haghighi, A., Li, L., 2019. Joint Asymmetric Tolerance Design and Manufacturing Decision-Making for Additive Manufacturing Processes. IEEE Trans. Autom. Sci. Eng. 16, 1259–1270. https://doi.org/10.1109/TASE.2018.2879719

Haghighi, A., Li, L., 2018. Study of the relationship between dimensional performance and manufacturing cost in fused deposition modeling. Rapid Prototyp. J. 24, 395–408. https://doi.org/10.1108/RPJ-11-2016-0177

Haq, A.N., Saravanan, K.S.R., 2006. Particle swarm optimization ( PSO ) algorithm for optimal machining allocation 865–869. https://doi.org/10.1007/s00170-004-2274-5

Haq, A.N., Sivakumar, K., Saravanan, R., Muthiah, V., 2005. Tolerance design optimization of machine elements using genetic algorithm. Int. J. Adv. Manuf. Technol. 25, 385–391. https://doi.org/10.1007/s00170-003-1855-z

Hegab, H.A., Darras, B., Kishawy, H.A., 2018. Towards sustainability assessment of machining processes. J. Clean. Prod. 170, 694–703. https://doi.org/10.1016/j.jclepro.2017.09.197

Hoffenson, S., Dagman, A., Söderberg, R., 2015. Visual quality and sustainability considerations in tolerance optimization: A market-based approach. Int. J. Prod. Econ. 168, 167–180. https://doi.org/10.1016/j.ijpe.2015.06.023

Huang, W., Phoomboplab, T., Ceglarek, D., 2009. Process capability surrogate model-based tolerance synthesis for multi-station manufacturing systems. IIE Trans. (Institute Ind. Eng. 41, 309–322. https://doi.org/10.1080/07408170802510408

Kapur, K.C., Cho, B.R., 1994. Economic design and development of specifications. Qual. Eng. 6, 401–417. https://doi.org/10.1080/08982119408918736

Kawlra, R., Hancock, W., 1996. Tolerance allocation methodology for manufacturing. SAE Tech. Pap. 105, 912–919. https://doi.org/10.4271/961274

Khouja, M., Mehrez, A., 1994. Economic Production Lot Size Model with Variable Production Rate and Imperfect Quality. J. Oper. Res. Soc. 45, 1405–1417.

Kim, T., Glock, C.H., 2018a. Production planning for a two-stage production system with multiple parallel machines and variable production rates. Int. J. Prod. Econ. 196, 284–292. https://doi.org/10.1016/j.ijpe.2017.11.018

Kim, T., Glock, C.H., 2018b. Production planning for a two-stage production system with multiple parallel machines and variable production rates. Int. J. Prod. Econ. 196, 284–292. https://doi.org/10.1016/j.ijpe.2017.11.018

Kumar, L.R., Padmanaban, K., Balamurugan, C., 2016. Least cost-tolerance allocation based on Lagrange multiplier. Concurr. Eng. Res. Appl. 24, 164–177. https://doi.org/10.1177/1063293X15625722

Leung, L.C., Hui, Y. V., 2000. Dynamic management of cutting tools for flexible and quality machining. Int. J. Prod. Res. 38, 3385–3401. https://doi.org/10.1080/002075400418306

Lim, E.M., Meng, C.-H., 1997. Integrated planning for precision machining of complex surfaces. Part 1: cutting-path and feedrate optimization. Int. J. Mach. Tools Manuf. 37, 61–75.

Liu, S., Qiu, J., 2011. Tolerance design considering the resources and environment characters during manufacturing process. Comput. Integr. Manuf. Syst. 7.

Mont, O., Dalhammar, C., Jacobsson, N., 2006. A new business model for baby prams based on leasing and product remanufacturing. J. Clean. Prod. 14, 1509–1518. https://doi.org/10.1016/j.jclepro.2006.01.024

Mont, O.K., 2002. Clarifying the concept of product–service system. J. Clean. Prod. 10, 237–245. https://doi.org/10.1109/ACIIDS.2009.18

Nicolaou, P., Thurston, D.L., Carnahan, J. V., 2002. Machining quality and cost: Estimation and tradeoffs. J. Manuf. Sci. Eng. Trans. ASME 124, 840–851. https://doi.org/10.1115/1.1511169

NLopt [WWW Document], 2020. URL https://nlopt.readthedocs.io/en/latest/

No Title [WWW Document], n.d. URL https://stats.stackexchange.com/questions/260860/normal-distribution-with-uniform-mean

Omachonu, V.K., Suthummanon, S., Einspruch, N.G., 2004. The relationship between quality and quality cost for a manufacturing company. Int. J. Qual. Reliab. Manag. 21, 277–290. https://doi.org/10.1108/02656710410522720

Otsuka, A., Nagata, F., 2017. Design method of Cpm-index based on product performance and manufacturing cost. Comput. Ind. Eng. 113, 921–927. https://doi.org/10.1016/J.CIE.2017.04.025

Qureshi, A.J., Dantan, J.-Y., Sabri, V., Beaucaire, P., Gayton, N., 2012. A statistical tolerance analysis approach for over-constrained mechanism based on optimization and Monte Carlo simulation. Comput. Des. 44, 132–142. https://doi.org/10.1016/J.CAD.2011.10.004

Ramesh Kumar, L., Padmanaban, K.P., Balamurugan, C., 2016. Least cost-tolerance allocation based on Lagrange multiplier. Concurr. Eng. Res. Appl. 24, 164–177. https://doi.org/10.1177/1063293X15625722

Rosyidi, C.N., Murtisari, R., Jauhari, W., 2016. A concurrent optimization model for suppliers selection, tolerance and component allocation with fuzzy quality loss. Cogent Eng. 3. https://doi.org/10.1080/23311916.2016.1222043

Rosyidi, C.N., Murtisari, R., Jauhari, W.A., 2017. A concurrent optimization model for supplier selection with fuzzy quality loss. J. Ind. Eng. Manag. 10, 98–110. https://doi.org/10.3926/jiem.800

Sanz-Lobera, A., Sebastián, M.A., Pérez, J.M., 2010. New cost-tolerance model for mechanical part design. Int. J. Adv. Manuf. Technol. 51, 421–430. https://doi.org/10.1007/s00170-010-2661-z

Sarkar, B., Saren, S., 2016. Product inspection policy for an imperfect production system with inspection errors and warranty cost. Eur. J. Oper. Res. 248, 263–271. https://doi.org/10.1016/j.ejor.2015.06.021

Scholz, F., 1995. Tolerance Stack Analysis Methods A Critical Review, Research and technology boeing information & support services.

Sealy, M.P., Liu, Z.Y., Zhang, D., Guo, Y.B., Liu, Z.Q., 2016. Energy consumption and modeling in precision hard milling. J. Clean. Prod. 135, 1591–1601. https://doi.org/10.1016/j.jclepro.2015.10.094

Shin, S., Kongsuwon, P., Cho, B.R., 2010. Development of the parametric tolerance modeling and optimization schemes and cost-effective solutions. Eur. J. Oper. Res. 207, 1728–1741. https://doi.org/10.1016/J.EJOR.2010.07.009

Singh, P.K., Jain, P.K., Jain, S.C., 2004. A genetic algorithm-based solution to optimal tolerance synthesis of mechanical assemblies with alternative manufacturing processes: Focus on complex tolerancing problems, International Journal of Production Research. https://doi.org/10.1080/00207540410001733931

Siva Kumar, M., Stalin, B., 2009. Optimum tolerance synthesis for complex assembly with alternative process selection using Lagrange multiplier method. Int. J. Adv. Manuf. Technol. 44, 405–411. https://doi.org/10.1007/s00170-008-1866-x

Sofiana, A., Rosyidi, C.N., Pujiyanto, E., 2019. Product quality improvement model considering quality investment in rework policies and supply chain profit sharing. J. Ind. Eng. Int. 15, 637–649. https://doi.org/10.1007/s40092-019-0309-7

Speckhart, F.H., 1972. Calculation of tolerance based on a minimum cost approach. J. Manuf. Sci. Eng. 94, 447–453.

Sutherland, G.H., Roth, B., 1975. Mechanism design: accounting for manufacturing tolerances and costs in function generating problems. J. Manuf. Sci. Eng. 97, 283–286. https://doi.org/10.1115/1.3438551

Svanberg, K., 2002. Methods Based on Conservative Convex Separable. SIAM J. Optim. 12, 555–573.

Taleizadeh, A.A., Moshtagh, M.S., Moon, I., 2018. Pricing, product quality, and collection optimization in a decentralized closed-loop supply chain with different channel structures: Game theoretical approach. J. Clean. Prod. 189, 406–431. https://doi.org/10.1016/j.jclepro.2018.02.209

Tlija, M., Ghali, M., Aifaoui, N., 2019. Integrated CAD tolerancing model based on difficulty coefficient evaluation and Lagrange multiplier. Int. J. Adv. Manuf. Technol. 101, 2519–2532. https://doi.org/10.1007/s00170-018-3140-1

Wang, P., Liang, M., 2005. An integrated approach to tolerance synthesis, process selection and machining parameter optimization problems. Int. J. Prod. Res. 43, 2237–2262. https://doi.org/10.1080/00207540500050063

Wang, Y., Li, L., Hartman, N.W., Sutherland, J.W., 2019. Allocation of assembly tolerances to minimize costs. CIRP Ann. 68, 13–16. https://doi.org/10.1016/j.cirp.2019.04.027

Wang, Y., Yan, C., Yang, J., Lee, C.-H., 2017. Tool path generation algorithm based on covariant field theory and cost functional optimization and its applications in blade machining. Int. J. Adv. Manuf. Technol. 90, 927–943. https://doi.org/10.1007/s00170-016-9337-2

Wilson, J.M., Piya, C., Shin, Y.C., Zhao, F., Ramani, K., 2014. Remanufacturing of turbine blades by laser direct deposition with its energy and environmental impact analysis. J. Clean. Prod. 80, 170–178. https://doi.org/10.1016/j.jclepro.2014.05.084

Wu, F., Dantan, J.-Y., Etienne, A., Siadat, A., Martin, P., 2009. Improved algorithm for tolerance allocation based on Monte Carlo simulation and discrete optimization. Comput. Ind. Eng. 56, 1402–1413. https://doi.org/10.1016/J.CIE.2008.09.005

Yan, J., Li, L., 2013. Multi-objective optimization of milling parameters-the trade-offs between energy, production rate and cutting quality. J. Clean. Prod. 52, 462–471. https://doi.org/10.1016/j.jclepro.2013.02.030

Yeh, S.S., Hsu, P.L., 2002. Adaptive-feedrate interpolation for parametric curves with a confined chord error. CAD Comput. Aided Des. 34, 229–237. https://doi.org/10.1016/S0010-4485(01)00082-3

Zahara, E., Kao, Y.T., 2009. A hybridized approach to optimal tolerance synthesis of clutch assembly. Int. J. Adv. Manuf. Technol. 40, 1118–1124. https://doi.org/10.1007/s00170-008-1418-4

Zhang, H.C., Huq, M.E., 1992. Tolerancing techniques: The state-of-the-art. Int. J. Prod. Res. 30, 2111–2135. https://doi.org/10.1080/00207549208948140

Zhang, Y., Zhang, G., Qu, T., Liu, Y., Zhong, R.Y., 2017. Analytical target cascading for optimal configuration of cloud manufacturing services. J. Clean. Prod. 151, 330–343. https://doi.org/10.1016/j.jclepro.2017.03.027