Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 33 (2015) 203 – 208

www.elsevier.com/locate/procedia

9th CIRP Conference on Intelligent Computation in Manufacturing Engineering - CIRP ICME '14

Statistical modeling of industrial process parameters

Francesco Aggogeri[a], Giulio Barbato[b], Gianfranco Genta[b,]*, Raffaello Levi[b]

[a]DIMI, Università degli Studi di Brescia, Via Branze 38, 25123 Brescia, Italy
[b]DIGEP, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

* Corresponding author. Tel.: +390110907295; fax: +390110907299. *E-mail address:* gianfranco.genta@polito.it

Abstract

Identification of models of process parameters provides a way to clarify some hitherto unexplained patterns of deviation from design values, leading to enhanced opportunities of quality improvement. While most standard procedures are based upon normal distribution hypothesis, the latter sometimes is liable to fail to accommodate actual data even to a first approximation. Skew, bounded, multimodal data sets call for reasonably close description if meaningful inferences are to be drawn. Graphic representation may pose challenges, the aspect of grouped data being materially affected by a more or less arbitrary choice among several options. Issues in modeling are discussed in the light of an actual case, concerning a critical bore realization on an automotive component.

Selection and peer-review under responsibility of the International Scientific Committee of “9th CIRP ICME Conference”

Keywords: Process Modelling; Boring; Multimodal Distribution

1. Introduction

A stiff manufacturing schedule and tight specifications turned a bore finishing operation on a component into a foreman’s nightmare, as performance of a complex manufacturing system was marred by scatter well beyond target capability, owing to a broad range of factors. A peculiar pattern of deviations from nominal diameter was observed, exhibiting inter alia a bimodal shape, as well as outliers affecting mainly one tail. In the quest for identification of main sources of trouble, statistical process modeling was resorted to, uncovering some problems concerning empirical distributions approximating those underlying data at hand.

Exploratory data analysis pointed to associations among process parameters and deviations from nominal diameter, leading to identification of steps susceptible to ensure process improvement. Machining of a cast iron component on a flexible manufacturing system was dictated by processing constraints, leading to problems linked to inherent system’s complexity, compounded by a scheduling strategy dictated by tight requirements concerning production rate. Multiple fixtures were involved as well as different spindles and associated tooling, entailing additional sources of variation.

The case concerns a SME, tier one supplier of automotive powertrain components. Substantial investments were made in innovation technologies and human resources in order to increase the product portfolio, supplying special products for different requirements and applications. A surge in production volume with a downfall of increasing scrapped items suggested application of advanced statistical tools, in a drive to identify main factors affecting performances in current processes.

Quality issues surfaced concerning finish machining of a bore on a cast iron component, with tight specifications concerning diameter. The manufacturing system includes an interlinked set of CNC units, performing a range of machining operations including drilling, boring and grinding. A detailed process mapping was performed in order to identify, among the following list of potentially relevant factors, those requiring further investigation.

- Material: rough castings were provided by two different suppliers, chemical analysis being performed on incoming parts to check conformance with specifications before machining.
- Machine: the manufacturing system included an interlinked set of CNC machining centers earmarked for the specific

Selection and peer-review under responsibility of the International Scientific Committee of “9th CIRP ICME Conference”
doi:10.1016/j.procir.2015.06.037

operation at hand, whose parameters were mapped and investigated.

- Method: a set of measurements were collected to monitor production quality and yield, on a sample of pieces checked on a CMM at every shift. Bore finishing was performed either by boring or reaming, selected according to availability and set-up team criteria.
- Man: an operator loaded components on fixtures set on pallets shifted among machining units according to availability, with a dedicated set-up team on duty taking care of possible issues. Since production was carried out in two shifts, systematic differences were not unlikely.

Statistical analysis of production data indicated a rather poor fit to either Student or normal distribution; however the theoretical appeal of the latter, provided by the Central Limit Theorem (CLT), justified adoption of a mixture of normal distributions to provide an empirical model.

2. Modeling empirical data distribution

A set of over 600 parts machined in a pilot run exhibited a peculiar pattern of deviations from the reference value of the diameter of a critical bore, as shown by dot plot in Fig. 1.

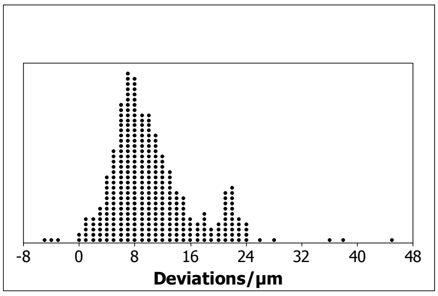

Fig. 1. Dot plot of deviations from reference value of diameter of a critical bore. Each dot may represent up to 2 observations.

A bimodal shape may be observed; furthermore, a few discrepancies appear on the left tail and some outliers may be identified on the right tail.

Outlier detection methods may be resorted to in order to identify discordant observations, a major shortcoming of most methods being the underlying hypothesis of normality, or even the requirement of knowing the underlying statistical distribution [1,2]. Given such knowledge, the problem of getting robust information from a reasonable number of data may be readily solved. When a few data only are available, difficulties are compounded by the fact that the main points of interest are on the tails, where data quality is inevitably poorer. Confidence or outlier identification intervals depend on probability concerning tails, and the difficulty of working in these regions appears evident. In fact some two centuries elapsed, since the groundbreaking work of Abraham de Moivre [3] on normal distribution, before a solution was provided to some practical tail problems by William Sealy Gosset [4] with his Student distribution.

Sound identification of statistical distribution on a purely empirical basis requiring a fairly large number of data, such an approach is ruled out in a number of instances. In the case at hand, the problem of outlier identification may not be approached in terms of the more common exclusion principles, as they are based on normal data distribution.

In the present work, an alternative method for outlier detection is proposed, based on an approximation of the experimental distribution with sound theoretical foundations. Some methods of exploratory data analysis are considered to model the empirical distribution.

At a preliminary level, histograms may offer a better representation of the empirical distribution of experimental data than dot plots, as bin width may be selected in order to highlight the most important aspects, a process entailing obviously individual appreciation. While there's no such thing as the "correct" bin width, some empirical rules provide a rough guidance, usually in terms of sample size n. Thus according to Sturges' rule [5] data range R is split into k equally spaced bins h wide, with

$$k \approx 1 + \log_2 n \qquad (1)$$

A common rule in software packages, e.g. *Minitab*, requires:

$$k \approx n^{1/2} \qquad (2)$$

In the case at hand, the number of classes is k=11 according to Sturges' rule and k=26 taking the square root of n, with corresponding bin widths of about 4.9 μm and 2.1 μm. Among a number of more or less similar rules, some take into account also measures of spread besides range (see e.g. [6,7]).

A good connection with the real situation can be given considering the concept of resolution, as described by VIM [8] in clause 4.14 :

> ***"resolution:***
> *smallest change in a quantity being measured that causes a perceptible change in the corresponding indication"*

Indeed resolution is a variability interval within which CLT works correctly, therefore it can be represented by a small normal distribution. This gives an indication justified by conceptual composition, even if its direct application is not easy: in fact resolution, as defined by VIM, depends on the measurement contest. The concept of reading resolution, also defined by VIM in clause 4.15, provides an easier approach:

> ***"resolution of a displaying device:***
> *smallest difference between displayed indications that can be meaningfully distinguished"*

Reading resolution is a well-defined, readily known characteristic of the measuring instrument concerned. As real variability is also affected by other factors, direct use of reading resolution as bin width would lead to an over-detailed description. A practical approach connecting such a readily