# Nonconformity Root Causes Analysis Through a Pattern Identification Approach

**Michael Donauer, Paulo Peças and Américo Azevedo**

**Abstract** Controlling, maintaining, and improving quality is a central topic in manufacturing. Total Quality Management (TQM) provides several tools and techniques to deal with quality related topics, which are not always applicable. With the increased use of Information Technology (IT) in manufacturing there is a higher availability of data with great potential of further improvements. At the same time this results in higher requirements for data storage and processing with demanding, time consuming sessions for interpretation. Without suitable tools and techniques knowledge remains hidden in databases. This paper presents a methodology to help analyzing root causes of nonconformities (NCs) through a pattern identification approach. Hereby a methodology of Knowledge Discovery in Databases (KDD) is adapted and used as a quality tool. As the core element of the KDD methodology, the data mining, a well-known statistical measure from the field of economics—the Herfindahl–Hirschman Index (HHI)—is integrated. After presenting the theoretical background a new methodology is proposed and validated through an application case of the automotive industry. Results are obtained and presented in the form of patterns in matrices. They suggest that concentration indices may indicate possible root causes of NCs and invite for further investigations.

M. Donauer (✉)
Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias,
4200-465 Porto, Portugal
e-mail: michael.donauer@fe.up.pt

P. Peças
IDMEC, Instituto Superior Técnico, Technical University of Lisbon,
1049-001 Lisbon, Portugal

A. Azevedo
INESC TEC (formerly INESC Porto) and Faculdade de Engenharia,
Universidade do Porto, Rua Dr. Roberto Frias 4200-465 Porto, Portugal

A. Azevedo (ed.), *Advances in Sustainable and Competitive Manufacturing Systems*, Lecture Notes in Mechanical Engineering, DOI: 10.1007/978-3-319-00557-7_70,

## 1 Introduction

Controlling and maintaining quality is a central topic in manufacturing. Increased use of information technology in mass production entails more data availability but also demands a great deal of data processing, interpretation, and presentation. Total Quality Management (TQM) can be understood as a philosophy consisting of values, tools, and techniques to increase customer satisfaction and continuous improvement. Techniques for controlling process parameters are a central point and deviations of those controlled measures signal need for action. Nowadays information storage and processing capabilities exist but suitable tools are not always available or must be tailored for answering specific questions of interest. If suitable tools are not available knowledge remains hidden in databases [1]. Quality tools and techniques offer a variety of methods to visualize and control process data and statistics can be applied to gain certainty about cause effect relations [2]. These tools are remedies to numerous quality problems but might not always take effect. The application case of this paper's methodology for example offers too many variables to be adequate for statistical analysis and visualizations of traditional quality tools do not serve as evaluation instruments.

Occurring nonconformities (NCs) at machines within the production steps are aimed to be identified for further investigation of root causes. Hereby the traceability of products in mass production with numerous machines at several production steps is highly depending on the level of implementation of information technology. In addition to that this becomes only transparent depending whether efforts for data analysis, interpretation, and visualization are done. Knowledge Discovery in Databases (KDD) for example offers a general framework consisting of sub-elements to generate knowledge from a dataset [3]. A core element is data mining (DM), a method with the aim to identify patterns [3]. The developer who uses this method has a high degree of freedom for using the kind of method as the DM step.

In this paper a new methodology is presented, which is validated through an application case from the automotive industry. The study presented relates to a real industrial problem. Quality related data of two consecutive manufacturing process steps is evaluated and visually represented in a color highlighted matrix. These matrices may identify the source of origin that caused the NCs to emerge. This is done by including the total number of NC occurrences, measuring their concentration among the machines, and highlighting in different shades the machines with the highest incidents. The visualization takes into account production steps, production volume, and nonconformities that occur at the machines within the production steps. However, the source of origin is not identified and must be further investigated for confirmation.

Results are of interest for academia and practitioners. Different disciplines such as IT, quality, and economics are consolidated. The integration of an economics concentration measure into a KDD methodology can be used as a quality tool for quality engineers to identify possibilities to improve processes in mass production with diverse NCs.

This paper presents an efficient method for treating and visualizing data related to process quality, namely NCs that are concentrated to single machines of two consecutive production steps. The method is applied to an automotive high volume production process of similar products varying in size, composition, and shape. A literature review of relevant topics is given, a methodology suggested, and an application case presented. The definition of the production steps under analysis is firstly presented in this paper. Secondly, the identification and collection of relevant data are done. In order to identify patterns an adapted methodology from discovering knowledge in databases is used. After treating the data the core element, data mining, is introduced and results can be obtained. An application case of the automotive industry is presented, results of identifying NC root causes discussed, and conclusions drawn.

## 2 Literature Review

The following section provides the theoretical background of relevant topics. TQM and quality tools are reviewed and a method of pattern identification through knowledge discovery in databases (KDD) is presented. The applied statistical concentration measure is explained in detail.

### *2.1 TQM and Quality Tools*

The subject Total Quality Management (TQM) is extensive, diverse, and influenced on a subjective body of thoughts. There is no global definition of TQM and companies often show highly diverse interpretations and uses [4]. It is generally accepted to describe TQM as a philosophy equipped with a set of tools and techniques with the target to increase customer satisfaction and continuous improvement. Having a mindset of satisfying the needs of the internal customer is achieved through tactics for changing a company's culture and structured technical techniques [5] and [6]. TQM is also understood as a management system consisting of values, techniques, and tools, as three interdependent components [7]. Common used tools and techniques are summarized by [8] and presented in Table 1.

Tools and techniques, as portrayed in Table 1, are described to be practical methods, skills, means, or mechanisms used for a specific circumstance [2]. Their purpose when applied is to achieve positive change and improvement [2] and [9]. In case the wide spectrum of TQM tools and techniques are not applicable a new tool must be tailored to solve a specific problem. This tool generation process entails efforts for data analysis, visualization, and interpretation. Knowledge Discovery in Databases (KDD) for example offers a general framework to generate knowledge from a dataset.

**Table 1** Quality tools and techniques used in industry [8]

| The seven basic quality control tools | The seven management tools | Other tools | Techniques |
|---|---|---|---|
| Cause and effect diagram | Affinity diagram | Brainstorming | Benchmarking |
| Check sheet | Arrow diagram | Control plan | Department purpose analysis |
| Control chart | Matrix diagram | Flow chart | Design of experiments |
| Graphs | Matrix data analysis method | Force field analysis | Failure mode and effects analysis |
| Histogram | Process decision program chart | Questionnaire | Fault tree analysis |
| Pareto diagram | Relations diagram | Sampling | Poka yoke |
| Scatter diagram | Systematic diagram | | Problem solving methodology |
| | | | Quality costing |
| | | | Quality function deployment |
| | | | Quality improvement teams |
| | | | Statistical process control |

## 2.2 *Pattern Identification Through Knowledge Discovery in Databases*

Knowledge Discovery in Databases (KDD) can be described as the complete process of discovering useful knowledge from data [3]. The part of identifying patterns that is relevant for further analysis is a core element [3] and referred to as data mining, which is a specific procedure of KDD [1]. Harding et al. [10] define data mining as a concept and algorithm mix consisting of machine learning, statistics, artificial intelligence, and data management. But terminology is ambiguous and one must be aware that different communities hold different terms with same meanings. Fayyal [3] compiled across communities the following names for data mining, which is the term used in this paper: knowledge extraction, information discovery, information harvesting, data archeology, and data pattern processing.

Figure 1 portrays the process of KDD that is described by [3] and starts by selecting from a database the relevant target dataset. Preprocessing the target data is relevant to remove noise and outliers for the data to be ready for further

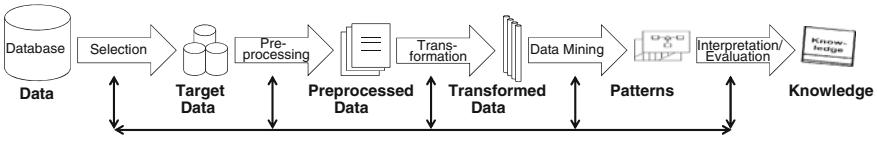


**Fig. 1** The KDD process c.f. [3]

processing. On the cleaned data the thoroughly identified or developed data mining algorithm can be performed to generate patterns. The produced patterns must be interpreted and evaluated for the knowledge to be discovered.

Fayyal et al. [3] mention that the data mining algorithm can be composed of a specific mix of the model (the function of the model and the representation form), the preference criterion (some form of goodness-of-fit function of the model to the data), and the search algorithm (the specification of an algorithm for the path of finding).

Recent reviews on KDD and data mining for manufacturing exist and indicate the popular use of KDD [1, 10, 11]. Some reviews also deal with KDD and data minding surrounding the topic of quality improvement [11] such as predictive maintenance, fault detection, quality assurance, product/process quality description, predicting quality, classification of quality, and parameter optimization.

Köksal et al. [11] reported an increasing use of data mining applications for quality related tasks. In those tasks applications for predicting quality are the most widely used ones followed by classification of quality and parameter optimization. There are plenty applications or algorithms respectively to perform data mining. Some of them are maps for classification, regression, or clustering of data. Others are summaries, dependency modeling of variables, and sequence analysis [3]. Model representation reach from decision trees over linear and non-linear models to case-based reasoning and probabilistic graphical dependencies.

## *2.3 The Herfindahl–Hirschman Index*

The Herfindahl–Hirschman Index (HHI) also referred to as the Herfindahl Index is a method to measure concentration [12]. Unaware of Hirschman's published work Herfindahl developed a similar method of measuring concentration at a later date [12, 13, 14]. The equations are identical with the only difference of the square root of Hirschman's index on Herfindahl's equation [13]. Herfindahl's equation is depicted in (1).

The index is the sum of the individual market shares of the participants in a specific market. Thus one can state:

$$HHI = \sum_{i=1}^{n} a_i^2 \tag{1}$$

$$\text{with } a_i = \frac{x_i}{\sum_j^N x_j} \tag{2}$$

The index is originally used in economics to measure competition in the marked and the effects of mergers or to measure concentration of income of households [12]. It is an adopted method of the department of Justice and Federal Reserve and currently in use to analyze merger intents [12].

Transferring this sense into the realm of quality can lead to the following understanding: Each imperfect process of a production step produces output—NCs—and the concentration to single machine among the total number of producers is measured. For every production step (n-1 and n) the concentration of every single NC is measured. A high HHI is referring to a high concentration, which can be understood that the great majority of NCs is produced by (a) single machine(s). Complementing to the HHI a visualization of all machines with their NC occurrences may highlight the critical ones and might even help in identifying root causes. This has to be proven after investigating the root cause.

KDD in engineering and quality related topics is well established and known. However, data mining algorithms are plentiful and there is no strict definition for existing models. This paper integrates a well-known statistical measure from the field of economics as the data mining algorithm within the KDD methodology. When applying the suggested method on a dataset it can be used as a quality tool for fault detection in manufacturing.

## 3 Pattern Identification Methodology

In order to improve production processes and learn from data an adapted methodology for pattern identification is suggested. The methodology is alike to the KDD methodology with a concentration measuring method from the field of economics as the data mining sub-step. The resulting patterns provide the basis for interpretation and knowledge creation. Firstly, one can identify which NC occurs concentrated at individual machines. Secondly, one can identify at which individual machines specific NCs occur most. Additional knowledge serves to highlight possible origins of NCs.

To obtain results one must first gather quality related data of the manufacturing process, namely recording the NCs of the production processes of relevance. The data of relevance must include information about the machines that the product passed of all the relevant production steps and the type of NC that was identified at the inspection station. The applied methodology in Fig. 2 is an abbreviated and adapted KDD methodology as previously presented in Fig. 1.

Firstly, one must gather data over a determined period of time. In order to retrieve data in a reliable manner the format of the input data file must be defined. When having the input file the preprocessing of data can be started. This includes

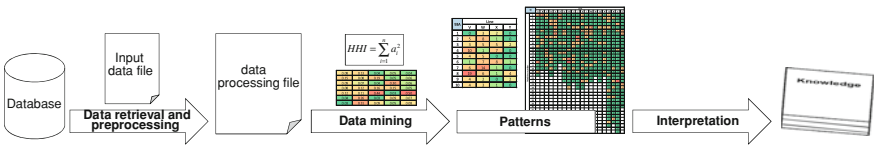


**Fig. 2** The methodology of the study

**Table 2** Retrieved data input file from database

| Step n-1 | | | | Step n | Inspection | | | |
|---|---|---|---|---|---|---|---|---|
| Barcode | Machine | Date | Time | Machine | NC type | Decision | Date | Time |
| 1***622 | V1 | 2010-12-01 | 02:32:27 | A16 | NC15 | Scrap | 2010-12-01 | 00:11 |
| 1***699 | X9 | 2010-12-01 | 00:04:53 | R12 | NC3 | Repair | 2010-12-01 | 00:32 |
| 1***244 | Y8 | 2010-12-01 | 00:03:38 | G19 | NC2 | Repair | 2010-12-01 | 00:33 |

spread sheet calculation which must be tailored or integrated to the previously obtained input data file. In this paper the Herfindahl–Hirschman index is the measure that serves as algorithm for the data mining sub-step.

The preprocessed data file provides information for every single NC: the number of incidents, the appraisal decision, and the machines the product had passed during production as presented in Table 2.

With basic calculations one can compute the occurrences of NCs according to machines on basis of the retrieved data as presented. This results in a matrix with machine number and NC type filled with the number of incidents as presented in Table 3.

Applying the statistical formula (1) and (2) to the tables one can calculate the HHI for a specific NC for one production step. After calculating for all NCs the HHI for each production step one gains information about how concentrated NCs are occurring at single machines. A specific visualization shall help in identifying the NCs with higher concentration to production machines within each production step.

The visualization in the form of patterns consists of two parts. Firstly, the concentration of a specific NC among the machines of each of the two production steps is calculated. This gives general information about whether a specific NC appears concentrated at individual machines within one production step, as one can see in Table 4. Secondly, the concentration index number of every NC for the two production steps is compared. This gives information about whether the NCs are very common and related with several machines or whether the NCs are appearing very concentrated to single machines.

## 4 The Application Case

The suggested methodology in Sect. 3 is applied to an application case of a high technology automotive parts producer. The company maintains a quality management system and is certified by quality standards such as DIN EN ISO 9000, DIN EN ISO 9004, and ISO/TS 16949.

**Table 3** Preprocessing of data to identify the number of occurrences according to machines

| Step n-1 | NC1 | NC2 | … | NCn | Step n | NC1 | NC2 | … | NCn |
|---|---|---|---|---|---|---|---|---|---|
| Machine 1 | x | y | … | z | Machine 1 | x | y | … | z |
| Machine 2 | … | … | … | … | Machine 2 | … | … | … | … |
| … | … | … | … | … | … | … | … | … | … |
| Machine n | u | v | … | w | Machine n | u | v | … | w |

**Table 4** The HHIs of production step n-1 and n according to the NCs

| | a | b | c | d | e |
|---|---|---|---|---|---|
| 1 | 0.08 | 0.13 | 0.04 | 0.05 | 0.04 |
| 2 | 0.15 | 0.08 | 0.19 | 0.05 | 0.06 |
| 3 | 0.09 | 0.07 | 0.04 | 0.30 | 0.06 |
| 4 | 0.08 | 0.12 | 0.16 | 0.15 | 0.05 |
| 5 | 0.12 | 0.11 | 0.44 | 0.03 | 0.50 |
| 6 | 0.04 | 0.26 | 0.03 | 0.09 | 0.07 |
| 7 | 0.03 | 0.31 | 0.09 | 0.05 | 0.08 |

Step n-1

| | a | b | c | d | e |
|---|---|---|---|---|---|
| 1 | 0.03 | 0.05 | 0.01 | 0.01 | 0.03 |
| 2 | 0.05 | 0.05 | 0.02 | 0.07 | 0.02 |
| 3 | 0.04 | 0.02 | 0.01 | 0.07 | 0.03 |
| 4 | 0.01 | 0.02 | 0.04 | 0.03 | 0.01 |
| 5 | 0.06 | 0.02 | 0.09 | 0.01 | 0.13 |
| 6 | 0.00 | 0.11 | 0.01 | 0.03 | 0.08 |
| 7 | 0.00 | 0.08 | 0.03 | 0.01 | 0.04 |

Step n

## *4.1 Problem Description*

The production process is composed of several production steps that do require dominating well different scientific fields. Mixing of raw materials, assembling subassemblies, and an injection akin process characterize the production steps. Each step consists of numerous machines and every product passes exactly one machine at every step. Barcodes are attached to the product and every machine equipped with a barcode scanner saves the product machine relationship to a database. Thus the database offers information about the history of the path of the production steps and the individual machines that the products took. An inspection station is installed at the end of the manufacturing line and humans inspect manually the products upon conformance to requirements. Conforming products are forwarded to be shipped to the customer. The type of NC is added to the information system and a decision of the recoverability through rework or the product to be scrapped is made. Due to the complex nature of the product and its processes causes of nonconformities are manifold and attributable to process failures, machine stoppages, incorrect composition, quality of raw materials, or human error. In addition NCs are often only detectable at the finished product. They vary from minor cosmetic to severe imperfections that may not be recoverable. The company uses brain storming approaches and is focused on reducing scrap rates of specific NCs and form multi-departmental teams to initiate investigations. But root cause identification has proven to be difficult, which is why the tool in this paper was developed.

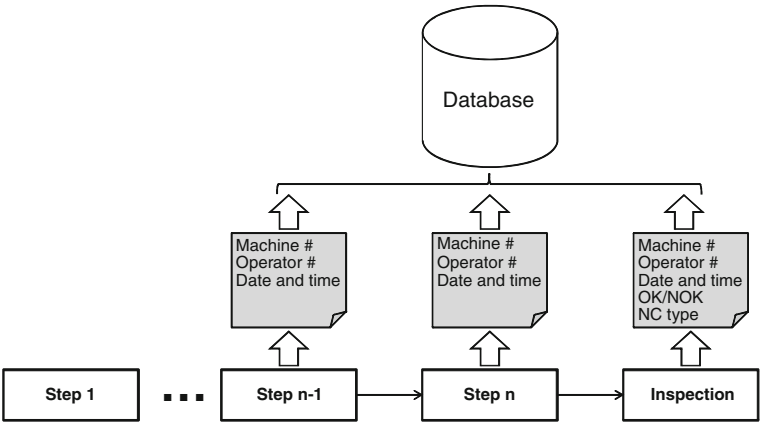


**Fig. 3** The production flow and data input to the database

Figure 3 illustrates the production steps and the input of information into the database. At the last two production steps before the inspection station corresponding data is input into the database. This data contains information about the specific production machine, the involved operator as well as time and date.

## *4.2 Overview of the NC Concentration*

After applying the tool's methodology one can build the tables illustrated below. Table 4 presents the concentration of all NCs for the two production steps. The numbers in the cells are the HHI results for each NC at the two production steps. Results for step n-1 are depicted on the left and for step n on the right side of Table 4. Each field is the concentration of NC incident of one specific NC of a production step.

Table 5 provides information about which NC corresponds to the HHI presented in the fields of Table 4 by comparing the horizontal and vertical index numbers and letters. NC18 in field '4c' for example has an HHI of 0.16 in step n-1

**Table 5** Corresponding NCs for HHI in Table 4 for step n-1 and n

| | a | b | c | d | e |
|---|---|---|---|---|---|
| 1 | NC1 | NC8 | NC15 | NC22 | NC29 |
| 2 | NC2 | NC9 | NC16 | NC23 | NC30 |
| 3 | NC3 | NC10 | NC17 | NC24 | NC31 |
| 4 | NC4 | NC11 | NC18 | NC25 | NC32 |
| 5 | NC5 | NC12 | NC19 | NC26 | NC33 |
| 6 | NC6 | NC13 | NC20 | NC27 | NC34 |
| 7 | NC7 | NC14 | NC21 | NC28 | NC35 |

and an HHI of 0.04 at step n. Both numbers are not comparable with each other because the numbers of machines are different for the two production steps. However, within one production step they do become comparable with each other. NC33 and 19 (Field 5e and 5c) show the highest HHIs for step n-1. NC33 and 13 (Field 5e and 6b) show the highest HHIs for step n.

## *4.3 Result Tables of the Production Steps*

Production step n-1 consists of fewer machines than the ones in production step n. The machines at each production step operate in parallel and exactly one machine at each production step is passed for the product to be produced. The route that the product takes from one production step to another depends on the set-up configuration of the machines. Different configurations allow producing products varying in a size, composition, and shape.

Following the suggestions of Table 4 the highest NC occurrences shows NC33 for both production steps. The NC occurrences at the machines are delineated in Fig. 4. Each field of the matrices represents one specific machine. The machines are located in lines and are numbered. Step n-1 consists of four lines (V, W, X, and

| NC 33 | Line | | | |
|---|---|---|---|---|
| | V | W | X | Y |
| 1 | 2 | 0 | 1 | 0 |
| 2 | 1 | 3 | 3 | 99 |
| 3 | 3 | 0 | 2 | 1 |
| 4 | 1 | 0 | 0 | 0 |
| 5 | 1 | 0 | 0 | 0 |
| 6 | 4 | 0 | 1 | 0 |
| 7 | 1 | 0 | 1 | 2 |
| 8 | 0 | 0 | 1 | 0 |
| 9 | 0 | 1 | 3 | 2 |
| 10 | 4 | 4 | 0 | 0 |

| NC33 | Line | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Machine number | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 15 | 0 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 47 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| 10 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 6 | 0 | 0 |
| 14 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 0 | 0 |
| 15 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 16 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17 | 0 | 0 | 0 | 2 | 0 | | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 19 | 0 | | | 0 | | | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| 20 | 0 | | | 0 | | | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 21 | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | 7 | | 0 | 0 | 0 | 0 |
| 22 | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | 18 | | 0 | 0 | 0 | 0 |
| 23 | | | | | | | | 0 | | 0 | 0 | | | | | | 0 | 0 | 0 | 1 |
| 24 | | | | | | | | 0 | | 0 | 0 | | | | | | 0 | 0 | 0 | 0 |
| 25 | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| 26 | | | | | | | | | | | | | | | | | 0 | 1 | 0 | 0 |
| 27 | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| 28 | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| 29 | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| 30 | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| 31 | | | | | | | | | | | | | | | | | 0 | 0 | 0 | |
| 32 | | | | | | | | | | | | | | | | | 0 | 0 | 0 | |
| 33 | | | | | | | | | | | | | | | | | 0 | 1 | 0 | |
| 34 | | | | | | | | | | | | | | | | | 0 | 0 | 0 | |
| 35 | | | | | | | | | | | | | | | | | 0 | 0 | 0 | |
| 36 | | | | | | | | | | | | | | | | | 1 | | 0 | |
| 37 | | | | | | | | | | | | | | | | | | | 0 | |
| 38 | | | | | | | | | | | | | | | | | | | 0 | |

**Fig. 4** Result presentation of process step n-1 (*left*) and n (*right*)

Y) each with 10 machines. Step n consists of 20 lines (A, B, …, T) each equipped with machines varying in number between 16 and 38.

The left matrix in Fig. 4 presents the number of occurrences of NC33 for every machine of production step n-1. As one can see machine number Y2 is related with 99 NCs among a total number of 141 NCs. All other machines of this production step show numbers of occurrences between zero and four. The right matrix in Fig. 4 presents the number of occurrences for every machine of production step n. In comparison to step n-1 the NCs occur more fragmented. Machine number R8 has the highest number of NCs (47). Machine number R7, R14, and O22 show number of occurrences of 15, 18, and 13. All other machines do not produce NCs or only few.

The above obtained results highlight the high concentration of NCs to machine Y2 of production step n-1. Thus, the number of occurrences is very concentrated to one single production machine and steps for further analysis of root causes of the NC must be done. A possible tool for doing that can be the cause and effect diagram from Table 1. This lists all possible factors of contribution such as operator, machine, method, or material and one can observe and investigate with this structured help the root cause if there is one to find.

A first observation may indicate that this machine (Y2) is the main contributor of the NC and that the root cause might be found when further analyzing this machine. However, this information shall be taken as a direction and invite for further investigation and must be considered cautiously. Additional information is required to gain higher certainty of this assumption. The methodology indeed does take into account the total number of a specific NC. But it does not consider the total number of the products based on the set-up configuration of the production machines. This means the method does only take into account NC types regardless of further product features, such as size, shape, or composition.

Furthermore, when comparing the two matrices in Fig. 4 there is a mismatch in total numbers of NCs. While step n-1 has a sum of 141 occurrences of NC33, step n shows 155 occurrences. Theoretically both numbers should match since every product passes exactly one machine at each step. This inconsistency has to do with incomplete datasets, which are attributable to technical defects of scanned bar-codes or neglect of data entry by operators, among others.

Similar matrices as presented in Fig. 4 are obtainable for all other NCs presented in Table 5 and ready for interpretation to discover knowledge. But presenting these figures would exceed the frame of this paper.

## 5 Conclusion

This paper proposes a methodology to help analyzing root causes of NCs. Visually represented discovered knowledge supports to identify possible root causes in mass production. An economic concentration measure (HHI) is integrated as the data-mining element of the KDD method. The proposed methodology can be used

as a quality tool and is validated by an application case from the automotive industry. Data tables are generated with different cell shadings according to the concentration of specific incidents. An incident in this context is an occurrence of a specific NC. These tables may help in disguising main contributors of NCs exposing them to the user to be further investigated.

Results indicate that with the applied visualization technique it is possible to identify single machines that are highly related with specific NCs and may be the originator. Further investigation of the likelihood of being the originator of NCs is required as the next step.

The methodology integrates several disciplines namely IT, quality, and economics. A well-known and established economical concept finds in quality an additional field of application. Quality engineers of industrial companies may find interest in using the tool to identify root causes in high volume production with numerous machines and diverse NCs. According to the presented results the visual representation of the data helps to quickly understand which NCs show the highest concentrations to machines at different production steps. The highly visual results ease the interpretation and further analysis to constantly improve production quality.

While initial findings are promising, further research is necessary. As a start, the success rate of being able to identify the root cause of an NC after having highlighted a possible contributor must be identified to further validate this tool. This tool is currently developed to be used offline. With further development and integration to the installed IT system of a company it can turn into an online tool. Additional development can even automatically alert responsible persons when a critical value of concentration is exceeded and further investigations of root causes become attractive.

As this paper demonstrates combining knowledge of different disciplines can result in new emerging methods, tools, and knowledge. The authors highly encourage cross- and interdisciplinary research.

**Acknowledgments** The authors would like to acknowledge the support of Fundação para a Ciência e a Tecnologia (FCT) under the PhD grant SFRH/BD/33791/2009.

Special gratitude is directed to the anonymous company with their managers and employees involved for providing data and being availabilities for questions and validation sessions.

## References

1. Choudhary AK, Harding JA, Tiwari MK (2009) Data mining in manufacturing: a review based on the kind of knowledge. J Intell Manuf 20(5):501–521
2. McQuater RE, Scurr CH, Dale BG, Hillman PG (1995) Using quality tools and techniques successfully. TQM Mag 7(6):37–42
3. Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) The KDD process for extracting useful knowledge from volumes of data. Commun ACM 39(11):27–34
4. Bounds GM (1994) Beyond total quality management. Toward the emerging paradigm. McGraw-Hill, New York
5. Hradesky JL (1995) Total quality management handbook. McGraw-Hill, New York

6. Rampey J, Roberts H (1992) Perspectives on total quality. In: Proceedings of total quality forum 4, Cincinnati, Ohio
7. Hellsten U, Klefsjö B (2000) TQM as a management system consisting of values, techniques and tools. TQM Mag 12(4):238–244
8. Dale BG, McQuater R (2009) Managing business improvement and quality: implementing key tools and techniques. Wiley (Business Series), London
9. Donauer M, Azevedo A, Peças P (2012) Evaluating the effects of soft TQM tools on quality costs by means of simulation: a case study from the automotive industry. In: Proceedings of 20th international conference on flexible automation and intelligent manufacturing (FAIM), Helsinki, pp 413–421
10. Harding JA, Shahbaz M, Srinivas, Kusiak A (2006) Data Mining in manufacturing: a review. J Manuf Sci Eng 128(4):969–976
11. Köksal G, Batmaz İ, Testik MC (2011) A review of data mining applications for quality improvement in manufacturing industry. Expert Syst Appl 38(10):13448–13467
12. Rhoades SA (1993) The Herfindahl-Hirschman index (cover story). Fed Reserve Bull 79(3):188
13. Hirschman AO (1964) The paternity of an index. Am Econ Rev 54:761–762
14. Hirschman AO (1969) National power and the structure of foreign trade. University of California Press, California