Kushtrim Bajgora

Optimizing Torque Management: Enhancing Reliability, Transparency, and Efficiency in Joint processes at Case Company X

School of Technology and Innovation management
Master's thesis
Master's Program In Industrial System Analytics

Vaasa 2026

UNIVERSITY OF VAASA

School of Technology and Innovation management

Author:	Kushtrim Bajgora
Title of the thesis:	Optimizing Torque Management: Enhancing Reliability, Transparency, and Efficiency in Joint processes at Case Company X
Degree:	Master of Science In Technology
Discipline:	Industrial System Analytics
Supervisor:	Ahm Shamsuzzoha
Year:	2026 **Pages: 114**

ABSTRACT:

This master's thesis focuses on the development and optimization of a torque tool management system for industrial production. The reliability of critical bolted joints is strongly dependent on the accuracy and condition of the torque tools used in assembly and inspection. In the target company, critical joints are ensured in production, and customer requirements for traceability and tightening reliability create a need for a more transparent, data-driven and practical tool management process. The aim of the work was to develop a real-time management system that supports daily production use and enables statistically based decisions regarding torque inspections.

The study was conducted as a constructive design study combined with quantitative empirical analysis. The new management system was implemented using two Norbar torque testers and a spreadsheet-based software solution. The system was piloted for two months in a real production environment, where torque data was collected from one production line. The collected data included time stamps, torque values, tool and user identifiers, and OK/NOK classifications based on tolerances. The data was analyzed using Measurement System Analysis (MSA), Gage R&R, Statistical Process Control (SPC), Process Performance Analysis, Tolerance Analysis, Alarm Logic Evaluation, and Measurement Interval Analysis.

The results showed that the developed measurement system is suitable for production use and that the collected data can be reliably utilized in process monitoring and parameter optimization. Based on the analyses, inspection tolerances, measurement set size, alarm logic and inspection interval suitable for the target environment could be determined. In addition, the work showed that the system enables tool deterioration and early detection and supports preventive monitoring. The master's thesis shows that the developed system improves the traceability and reliability of torque tool inspections and provides a practical basis for data-based calibration and maintenance decisions. The results support the target company in moving from a routine approach towards a more transparent, statistically justified and condition-based torque tool management process.

KEYWORDS: torque tool management, torque verification, torque measurement, measurement system analysis, calibration, Gage R&R, process capability

UNIVERSITY OF VAASA

Tekniikan ja innovaatiojohtamisen yksikkö

Tekijä: Kushtrim Bajgora

Tutkielman nimi: Optimizing Torque Management: Enhancing Reliability, Transparency, and Efficiency in Joint processes at Case Company X

Tutkinto: Diplomi-Insinööri

Oppiaine: Industrial System Analytics

Valvoja: Ahm Shamsuzzoha

Valmistusvuosi: 2026 **Pages: 114**

ABSTRACT:

Tämä diplomityö keskittyy teollisen tuotannon vääntömomenttityökalujen hallintajärjestelmän kehittämiseen ja optimointiin. Kriittisten pulttiliitosten luotettavuus riippuu vahvasti kokoonpanossa ja tarkastuksessa käytettävien vääntömomenttityökalujen tarkkuudesta ja kunnosta. Kohdeyrityksessä kriittiset liitokset varmistetaan tuotannossa, ja asiakkaiden vaatimukset jäljitettävyydelle ja kiristyksen luotettavuudelle luovat tarpeen läpinäkyvämmälle, datalähtöiselle ja käytännöllisemmälle työkalujen hallintaprosessille. Työn tavoitteena oli kehittää reaaliaikainen hallintajärjestelmä, joka tukee päivittäistä tuotantokäyttöä ja mahdollistaa tilastollisiin tietoihin perustuvat päätökset vääntömomenttitarkastuksista.

Tutkimus toteutettiin konstruktiivisena suunnittelututkimuksena yhdistettynä kvantitatiiviseen empiiriseen analyysiin. Uusi hallintajärjestelmä otettiin käyttöön käyttämällä kahta Norbarin vääntömomenttimittaria ja taulukkolaskentapohjaista ohjelmistoratkaisua. Järjestelmää pilotoitiin kahden kuukauden ajan todellisessa tuotantoympäristössä, jossa vääntömomenttidataa kerättiin yhdeltä tuotantolinjalta. Kerättyyn dataan sisältyivät aikaleimat, vääntömomenttiarvot, työkalu- ja käyttäjätunnisteet sekä toleranssien perusteella tehdyt OK/NOK-luokitukset. Dataa analysoitiin mittausjärjestelmäanalyysin (MSA), Gage R&R:n, tilastollisen prosessinohjauksen (SPC), prosessin suorituskykyanalyysin, toleranssianalyysin, hälytyslogiikan arvioinnin ja mittausvälianalyysin avulla.

Tulokset osoittivat, että kehitetty mittausjärjestelmä soveltuu tuotantokäyttöön ja että kerättyä dataa voidaan luotettavasti hyödyntää prosessin seurannassa ja parametrien optimoinnissa. Analyysien perusteella voitiin määrittää kohdeympäristöön sopivat tarkastustoleranssit, mittausjoukon koko, hälytyslogiikka ja tarkastusväli. Lisäksi työ osoitti, että järjestelmä mahdollistaa työkalujen kulumisen ja varhaisen havaitsemisen sekä tukee ennakoivaa valvontaa. Diplomityö osoittaa, että kehitetty järjestelmä parantaa momenttityökalujen tarkastusten jäljitettävyyttä ja luotettavuutta sekä tarjoaa käytännönläheisen pohjan datapohjaisille kalibrointi- ja kunnossapitopäätöksille. Tulokset tukevat kohdeyritystä siirtymisessä rutiininomaisesta lähestymistavasta kohti läpinäkyvämpää, tilastollisesti perusteltua ja kuntoon perustuvaa momenttityökalujen hallintaprosessia.

Contents

Figures

Tables

1 Introduction

Torque is one of the most important parameters in many industrial sectors, such as aerospace, automotive, mechanical, pharmaceutical, electrical and agricultural (Zhong et al., 2023). In industrial environments, torque measurements have numerous applications, such as monitoring the rotational power produced or consumed by engines, turbines, and other devices. Another application of measuring torque is found in torque tools, which are used in assembly work to ensure that fasteners are tightened correctly.

The purpose of torque measurement in assembly is to control the clamping force of joints in assembly parts or other mechanical parts (Bickford, 2018). Torque is most frequently measured by using a torque wrench. Torque wrench is a tool designed to apply a specified torque to a fastener, i.e., nut or bolt. In industrial applications, accurate measurement of torque tools is required to attain product quality and safety. Successful joint installation means that the torque values of the torque wrenches are precisely adjusted (Yang et al., 2021). Even a small deviation in the applied torque can lead to insufficient or excessive tightening, which can result in joint failure.

Failure of bolted joints can have significant consequences in production (Croccolo et al., 2011). Loose or otherwise defective joints can lead to rework, scrap, warranty claims, recalls and damage to the company's reputation. For this reason, there is a clear need for systematic and reliable torque tool management in the manufacturing industry. More accurate and transparent torque management can improve both product quality and process reliability, while reducing risks and unnecessary costs.

The purpose of this thesis is to develop an efficient and transparent torque tool management system to monitor the target company's torque tools. The system is designed to collect torque measurement data from torque tools during use. In addition to supporting daily production quality, the system aims to provide better visibility into the condition of the tools and enable earlier identification of problems, such as tool drift, maintenance needs or calibration needs.

An additional goal of the thesis is to create a data-driven torque management model and inspection procedure. Rather than following fixed routines, the collected data can be utilized in decision-making regarding tool monitoring and maintenance, such as determining inspection intervals and the timing of corrective actions. The developed system enables a move from time-based calibration, such as every 12 months, to data-driven calibration. This provides the possibility of a more proactive and dependable approach to the management of torque tools in the production environment,

1.1 Background and motivations

The purpose of this thesis is to develop and optimize a torque tool management system that improves the quality, reliability, transparency and efficiency of connections in case company's industrial processes. The system developed in this thesis responds to case company's need to create a real-time, self-managed torque tool measurement system. The system developed in the thesis enables production workers to perform regular torque measurements and at the same time the torque data of the tools at each workstation is automatically recorded, and if the measurement results deviate from the set tolerances, the alarm system reacts. The introduction presents the definition of the research problem, the limitations of the work and the structure of the thesis.

1.2 Research problem and objectives

The central research problem of the master's thesis is to develop and optimize the measurement process of torque tools. The measurement process must be developed in such a way that the installation of joints more reliable and efficient. This study aims to develop a system that not only monitors measurement results in real time, but also enables statistical analyses and optimization of the measurement process based on the data. This approach ensures that the different stages of production support each other and that the causes of errors are immediately identified and can be quickly corrected. If torque verifications are performed too infrequently and the tolerances are too loose, tool deviation may go unnoticed and this can compromise product safety. Frequent checks and overly tight tolerance limits cause frequent false alarms. False alarms reduce confidence in the new system and lead to inefficient use of calibration resources. This leads to three central research questions:

RQ1: How can we optimise the frequency and volume of torque measurements for maximal reliability?
RQ2: What tolerance thresholds best balance safety and efficiency?
RQ3: How should alerts be structured to minimise false alarms yet still detect real faults quickly?

In order to address this research problem, this thesis has four main objectives. The first objective is to develop a real-time self-service torque measurement system that is capable of measuring torque tools and that operators can use independently in their daily work. Second objective is to define statistically sound measurements intervals, sample sizes and tolerance limits, based on measurement system capability and process variation. Third, the study develops and validate an alert system, ensuring that alarms are meaningful and interpretable. Finally, the thesis establishes a data-driven approach to calibrations, where historical measurement data is systematically analyzed and used to refine calibration intervals, tool maintenance strategies and decision rules over time. Together, these objectives provide a structured framework for designing a torque measurement concept that is both robust and practical for everyday production use.

1.3 Research limitations

The system for this study was made with two Norbar torque testers, which are designed to measure the torque values produced by constant-torque machines and mechanical hand torque wrenches. Pulse machines have been excluded from the study, because the Norbar torque testers in use are not suitable for measuring pulse machines. At the case company, the pulse machines have typically been used for pre-tightening due to their lower accuracy. Joints tightened with a pulse machine are always checked afterwards with a mechanical torque wrench, the accuracy of which is monitored and ensured by the system developed in this master's thesis.

1.4 Structure of the work

The structure of the thesis is designed to cover both the theoretical and practical aspects of optimizing the measurement process of torque tools.

Chapter 1: Introduction, presents the background of the research, motivations, key research problems and objectives, research boundaries and the structure of the work.

Chapter 2: Case Company, Introduces the organization and the industrial environment where the thesis is conducted. This chapter describes the company’s business operations, its products, and the industrial context of the manufacturing environment.

Chapter 3: Literature Review, discusses the theoretical perspective, such as the analysis of measurement systems, the definition of tolerances, calibration processes and audits and checks of torque tools. The chapter also reviews previous studies, articles and theoretical models that support the basis of the research.

Chapter 4: System Design and Implementation, describes the overall architecture, technical implementation and user interface of the developed measurement system. This chapter also presents the equipment used, such as Norbar torque testers, and the spreadsheet-based software solution developed for the system.

Chapter 5: Measurement Process and Calibration, discusses the different steps of the measurement process, such as performing 10 measurements per tool, defining tolerances, and the logic of the alarm system. The chapter also analyzes statistically optimal measurement intervals.

Chapter 6: Data analysis, presents the collected measurement data is presented, statistical analysis is performed, and optimal measurement parameters are determined. This chapter provides concrete examples of how data is utilized in process optimization.

Chapter 7: Discussion, discusses the research results are in relation to previous literature and theoretical models. This chapter also presents other development suggestions and needs for further research.

Chapter 8: Conclusion and Recommendations, provides the key findings of the work, presents solutions to the research problem and presents the significance of the work in industry.

The structure of the work enables a broad and in-depth examination of the measurement processes of torque tools and provides a clear picture of how theory and practice combine to improve the quality of joints and ensure the safety and efficiency of production.

2 Case company

The case company is a global energy management company providing advanced solutions in the power management, industrial automation and sustainability sector. It operates as the Finnish subsidiary of a global power management company. The company has a diversified portfolio and is focused on innovation. Company descriptions highlight that it offers a wide range of power-related products and services and frame it, for example, as a “Global Power Management Company” that provides industry-leading electrical, aerospace, hydraulics and vehicle systems to solve customers’ power challenges. The parent company is very large and has operations in over 160 countries and approximately 94,000 employees worldwide. Case Company itself is headquartered in Finland, but benefits from this global network. The Finnish production campus is home to the production and distribution of specialized power systems. This facility focuses on three-phase uninterruptible power supply units (UPS) which are essential for maintaining operations in data centers, industrial plants and medical facilities.

The Case Company serves critical infrastructure and industry sectors, including data centers, large manufacturing facilities, utilities, and other sectors that rely on uninterrupted power. This emphasis on reliability is central to the organization’s strategic mission of enhancing operational safety and energy efficiency. With a long history of expertise in power management, the company focuses on environmental protection and social progress through advanced electrification and renewable energy technologies.

2.1 Products

The case company mainly manufactures UPSs. UPS stands for Uninterruptible Power System. It acts as an interface between the mains and sensitive applications. UPS supplies the load with continuous, high quality electrical power, regardless of the status of the mains. Digital equipment such as computers, telecom systems and high-precision instruments use microprocessors that operate at frequencies of several mega or even giga Hertz. They carry out millions of operations per second. Even a small disturbance in the electrical supply, lasting just a few seconds, can affect thousands or millions basic operations. In critical locations such as airports or hospitals, malfunctions and data loss can cause dangerous and expensive consequences. Therefore, sensitive loads require a supply that is protected against distribution system disturbances.

2.2 Rationale for torque tool management

The target company operates in a manufacturing environment where the quality and reliability of assembled electrical systems are critical. As the company's products are used in applications where uninterrupted power supply is essential, the reliability of mechanical and electrical connections directly affects the product's performance and operational safety. In such products, a faulty connection can lead to rework, reduced product reliability, field damage or, in the worst case, safety-related consequences. For this reason, the quality of the tightening processes and the condition of the torque tools used in the assembly are of paramount importance.

In the target company, critical joints are torque-checked as part of the production process. When a critical joint is tightened, for example with a pulse tool, the joint is then checked with a mechanical hand-operated torque wrench to ensure that the required torque level is achieved. In some customer cases, the requirements are even stricter. Certain customers require double-checking, which means that after the first torque check, another employee must check the same joint again with a second mechanical torque wrench. These requirements emphasize the great importance of traceability, repeatability, and confidence in the correctness of critical joints. Torque checks take time in the assembly process and reduce the efficiency of the production flow by causing interruptions to assembly work.

The target company's key development objective is to develop a method that can ensure the reliability of connections and document this information without slowing down production. In addition to monitoring connections, the company lacks a sufficiently transparent and data-based method for monitoring the condition and performance of torque tools used in production. Customers may require evidence that critical connections have been tightened to the correct torque. Such evidence can be obtained either with smart torque tools that record tightening data directly from each connection (Piontek et al., 2024). Alternatively, it can be obtained from a verification-based system that shows that the tools used in production are operating at the required torque level. The system developed in this work falls into the latter category. It provides documented evidence that the torque tools used for tightening have been inspected and found to be functional within a specified inspection interval. This thesis addresses this need by developing a real-time torque tool management system and optimizing its key operating principles based on production data and statistical analysis. The goal is to create a system that is simple and practical for production workers in their daily work, to ensure that regular torque verifications can be performed effortlessly and without special technical expertise. The system is also designed to automatically record torque verification data and provide the calibration team with reliable information on the condition of the tools. The developed system also enables the target company to move from a time-based approach to a more transparent, traceable and data-driven torque tool management process.

3 Literature review

This chapter discusses the theoretical foundations of torque tool control and measurement systems and their importance in industry. It also examines torque tolerance definition, calibration processes and measurement system analysis (MSA) methods. In addition, previous studies and theoretical models that support the research design of this master's thesis are reviewed to provide a foundation for the research design.

3.1 Torque tool verification in industrial assembly

Torque tools are used to apply a specified tightening torque to fasteners. Accurate measurement of torque is vital in industries since the degree of joint tightness and stiffness determines the safety and longevity of products (Bickford, 2018). Torque tools are essential elements in industrial assembly processes, particularly in applications where the correct screw connection preload is critical in order to ensure product reliability, safety and quality (Croccolo et al., 2011). Torque tools enable connections to be tightened precisely to specified torque requirements, avoiding both under- and over-tightening. Inaccurate torque application can cause joint failures, which can lead to serious structural damage. Therefore, proper tool selection and regular calibration are paramount to ensure consistent process quality.

The selection of torque tools is based on the requirements of each application and the characteristics of the joint (Bickford, 2018). In addition to technical features, ergonomic factors such as workstation design and work methods are play an important role. While manual torque wrenches are still standard tools for general assembly, electric and pneumatic tools are more suitable for high-volume production. Pulse tools, on the other hand, are optimized for workstations that require fast operation and ergonomic benefits, such as repetitive assembly tasks and medium to high torque applications (Lundin & Haettel, 2023). Since accuracy and repeatability vary depending on the type of tool, the selection must be tailored to the joint in question, the production environment and the level of control required.

It is essential to distinguish between laboratory-level calibration and torque verification in a production environment. Under controlled calibration conditions, torque tools can meet very strict accuracy requirements set by the manufacturer or standards. However, in production, many additional factors affect the tool, such as user technique, frequency of use, environmental conditions and mechanical wear. Therefore, practical verification of torque tools in industry requires more than just the nominal accuracy value stated on the calibration certificate. It requires a system that is able to monitor the performance of the tools within acceptable limits over the long term and detect developing deviations in time to prevent quality defects.

From the perspective of this thesis, torque tool verification is thus understood as a production yield control measure that supports the quality, traceability and predictive maintenance of joints. The purpose of verification is not only to identify already failed tools, but also to detect early signs of drift or instability before they lead to rejected joints. This closely links torque tool verification to broader quality engineering methods, such as tolerance definition, calibration management, measurement system analysis and statistical process control (SPC). These interrelationships form the theoretical basis for the data-driven torque tool management model developed in this work.

3.2 Tolerance limits and calibration of torque tools

The definition of torque tolerances is a key part of the quality assurance of joints. The purpose of the torque tool measurements performed by the system is to trigger an alarm if the torque result of the verification is above the allowed tolerance. The aim of this chapter is to answer the key question of what tolerance limit is theoretically and practically justified for torque verifications. The aim of this chapter is to justify the correct tolerance limit based on academic literature and standards and to justify the correct tolerance for our case company’s production environment. It is crucial to determine the appropriate tolerance limit for such torque verifications. If the tolerance is too tight, tools may be incorrectly considered unfit due to normal measurement variation. If the tolerance is too loose, a tool that is losing torque may go unnoticed. The correct definition of tolerances is based on both technical and statistical methods, which are used to estimate the margin of error of the measurement system and ensure product compliance. From a quality management perspective, accurate definition of tolerances can minimize variation in the production process and ensure that all joints meet the desired and agreed quality requirements.

3.2.1 Torque tool accuracy and tolerance

International standards define the basic requirements for torque tools and their calibration. ISO 6789:1 -2017 is the primary standard for hand-held torque tools. It specifies that new torque tools must operate at a specified percentage of the set torque throughout their range. Historically, ISO 6789 required an accuracy of ±6% of full scale, while for certain types of torque tools using torques above 10 Nm, the tolerance was ±4%. For example, a torque tool below 10 Nm could deviate from the set value by ±6%, while some specific tools above 10 Nm had a tolerance of ±4%. These tolerances apply in a controlled calibration environment. The 2017 revision of ISO 6789 distinguishes between Part 1 design conformity and Part 2 calibration method with uncertainty. It still essentially maintains the tolerances of ±4% and ±6%, while requiring a strict uncertainty assessment. In particular, ISO6789-2:2017 requires that calibration certificates include an uncertainty budget that takes into account factors such as repeatability, reproducibility, resolution and load point effects. In summary, ±6% of reading is a typical manufacturer’s specification and calibration tolerance for hand torque tools under ideal conditions.

The tolerances for tools specified in the ISO standard are for laboratory conditions. In our situation, we need to take into account other influencing factors that exist in practice, such as that torque testing is performed by production workers in the production environment. According to Bickford (2018) the operator accuracy is ±10%. The tolerance is in line with the bolted joint analysis performed by Atlas Copco, where a tolerance of ±10% was used for torque-controlled tightening (Scharnowski & Bozkurt, 2019). In addition, the Kalmar company standard also specifies a general tolerance of ±10% for tightening torques (Kalmar, 2018). At one of John Deere's manufacturing facilities, tool audits were performed monthly and the allowable tolerance for the nominal torque of the tools was ±7 (Enos, 2014). However, in that case, tool audits were performed by designated and trained calibrators. In our process, tool torques are audited by production workers, which is why the operator accuracy of ±10% mentioned by Bikford is more appropriate in our case. Therefore, existing research supports using a ±10% tolerance for production verification, when measurements are carried out by production workers rather than calibration technicians.

3.2.2 Torque tool calibration

The accuracy of torque tools can be lost over time by mechanical wear, repeated use and possible harm (Stewart, 2018). Regular calibration is needed to ensure the tools still have the limits of required accuracy. Calibration is performed through the comparison of the tool's torque output to a known reference torque and the adjustment of the tool accordingly. Either in-house or by an external accredited calibration laboratory can be utilized to carry out calibration, as long as the relevant equipment and expertise are available. Various alternative calibration methods, including static calibration when the load is put on in small increments and dynamic calibration that simulates real use, exist (Heyn, 2015). Calibration is to be documented, such as the date of calibration, the readings taken, reference devices used and the date of the next calibration.

3.2.3 Calibration interval and verification interval

Reliable use of torque tools requires that the tools are calibrated systematically. Systematic calibration is important, as the output deviation of the tools increases as a result of time, wear, load and handling damage. In industrial practice, the most common calibration interval is often 12 months or 5000 operations (Bangi et al., 2014). The ISO 6789-2:2017 standard also presents 12 months or 5,000 cycles as the default calibration interval. However, the standard emphasizes performing calibrations more frequently if the user's process, customer requirements or legislation require it.

It is also important to distinguish between the formal calibration interval and the production-level verification interval. Calibration refers to the regular adjustment and formal verification of the accuracy of a tool against a reference. Verification refers to more frequent torque checks performed in production, the purpose of which is to monitor whether the tool remains in acceptable condition between calibrations. This distinction is particularly relevant in this thesis, as the developed system is based on regular torque checks performed in production rather than just calibration. Calibration intervals can be controlled by performing torque checks. This creates the theoretical basis for the later empirical analysis of the measurement interval in the developed system.

3.3 Torque measurement and testing

Torque measurement and testing are critical processes in industrial assembly, as they ensure that bolted joints achieve the required preload and thus the performance required of them (Bickford, 2016). Reliable measurements are the basis for quality control and enable the identification of deviations and corrective actions.

3.3.1 Torque measurement methods in a production environment

In a production environment, torque can be assessed either during tightening or after the joint has already been assembled (Meng et al., 2021). Dynamic torque refers to the torque applied during the tightening process, while static torque refers to the check torque measured on an already tightened joint. The two are not directly proportional to each other, as different factors affect the measured value depending on the measurement situation (Atlas Copco, 2015).

Static inspection of the finished joint is often used in practice to assess whether a tightening joint is acceptable. However, this approach has significant limitations. When the joint is inspected after tightening, the measured value is affected by factors such as joint settlement, internal friction, and the operator’s ability to recognize the moment at which the fastener begins to move (Archer, 2008). As a result, the value observed during inspection may not correspond to the original tightening torque or the true residual torque of the joint. Another challenge is the internal friction of the connection. The torque decreases as the bolt moves, i.e. the actual residual torque of the connection is lower than the torque needed to get it moving. This increases uncertainty and makes it difficult to validate the result, especially when tight tolerances are required.

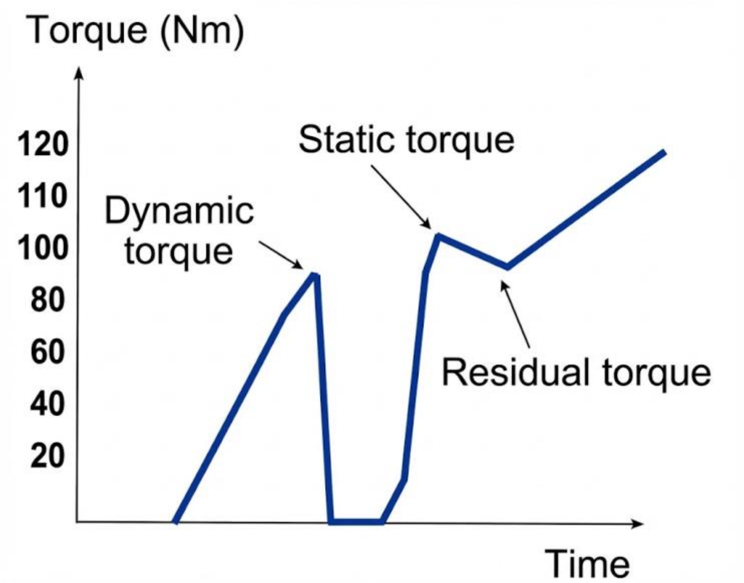


Figure 1. Example of a residual torque diagram.

Because of these limitations, the choice of measurement method is central to a torque management system. If the goal is to assess the condition and performance of the torque tool itself, direct measurement of the tool provides a more consistent basis for analysis than indirect inspection of the finished joint. This distinction is particularly relevant in this thesis, where the aim is to manage the condition and securing performance of the tool in production rather than estimating the tightening torque from individual finished joints.

3.3.2 Torque testers in production verification

The solution in this thesis for torque checks in a production environment is the use of torque testers. These separate devices are designed specifically for measuring the torque of torque tools in a production environment (Michael, 2020). Measuring torque directly from the tool with a torque tester yields more accurate results than a measurement taken from a finished joint. Therefore, torque testers provide a suitable basis for regular production level assurance of torque tools. They support both routine inspections and calibration functions by enabling repeated measurements under controlled and standardized conditions. In the context of this work, the use of torque testers is particularly important, as the aim is to monitor the condition of tools over time, detect developing deviations and produce data for statistical analysis. Thus, torque testing is not always a measurement procedure, but also a central part of the data-based torque tool management method developed in this thesis.

3.4 Measurement system analysis

Measurement system analysis (MSA) is an essential part of quality control and quality assurance in processes (Sharma et al., 2019). The goal of MSA is to assess the fitness for purpose of a measurement system by understanding the sources and magnitude of variation in the data it produces (AIAG, 2010). MSA can be used to determine how much measurement error occurs in the use of tools when a production worker measures the torque of a wrench. It is also used to determine the likelihood that measurement values vary from tolerance limits. A reliable method of measurement provides data, which are accurate, repeatable, and sensitive to detect process variation. MSA is helpful to identify and measure variation associated with the measuring system, which is beneficial for deciding to improve the measuring process and product quality.

3.4.1 Measurement uncertainty in torque verification

No measurement is perfect, and torque measurement always involves some degree of measurement uncertainty (Luca G et al., 2003). This uncertainty must be accounted when assessing the reliability of torque verification results. Measurement uncertainty is influenced by several factors, such as the calibration history and accuracy of the measuring device used, measurement conditions such as temperature and vibration, and user action such as applying force and reading. Even with high precision equipment and accurate torque tools, the presence of tolerances in both the tool and the tester means that no measured value can be considered perfectly exact.

In production environments, measurement uncertainty is particularly significant because it directly affects the reliability of pass/fail decisions. If the uncertainty of the measurement system is large relative to the allowed tolerance, normal variation may be misinterpreted as a tool defect, or true tool deviation may go undetected. The uncertainty is typically greater when torque is estimated from a finished joint, as factors such as joint alignment, friction, positioning, and operator interpretation can also affect the result. In contrast, direct verification of the torque tool with a torque tester reduces these uncertainties and provides a more consistent basis for assessing and monitoring tool condition.

From the perspective of this thesis, measurement uncertainty is not only a general metrological concept, but a practical quality issue that affects tolerance setting, alarm logic and the overall suitability of the verification system. For this reason, uncertainty must be considered together with repeatability, reproducibility and overall variability of the measurement system when assessing whether the developed torque verification process is suitable for production use.

3.4.2 Principles and objectives of measurement systems

The basic principle of MSA is to understand that the total variation in the observed measurement data consists of two main components: the true variation of the product and the variation introduced by the measurement system (Sharma et al., 2019). The primary objectives of MSA are to determine the statistical properties of the measurement system, such as accuracy, repeatability and reproducibility (AIAG, 2010). The aim is also to assess the ability of the measurement system to distinguish the true variation of the product from the noise introduced by the measurement system. The MSA system must identify the significant sources of variation in the measurement system. This also ensures that the measurement system is good enough to be acceptable for its intended use. In addition, MSA provides a basis for improving the measurement system.

Total measurement system variation can be divided into several sources (AIAG, 2010):

- Equipment: The characteristics of the measuring device (e.g., torque testers) such as design, manufacturing, wear, and calibration have effects on the measurements obtained. Systematic differences or variations may exist among devices (stability).

- User: The person performing the measurement can be a significant source of variation. Different users may have different ways of performing the measurement, which can lead to large differences in results (Appraiser variation).

- Method: The measurement process itself, including instructions, standards, and the measurement environment, can cause variation. Unclear instructions or unstable conditions can increase measurement error.

- Part to be measured: Although the system aims to measure the same parts repeatedly, internal variation within the parts can also affect the results, especially if the measurement location is not always exactly the same.

- Environment: Temperature, humidity, vibration and other environmental factors can influence the performance of the measuring device and thereby impact the measurement results.

3.4.3 Measurement system repeatability and reproducibility

In MSA, two key concepts in the assessment of measurement system variation are repeatability and reproducibility (AIAG, 2010). Repeatability, also known as equipment variation (EV). It describes the variation between measurements made repeatedly by the same operator on the same measurement device on the same part. Good repeatability means that the measurement device gives consistent results over a short period of time. Reproducibility, also known as appraiser variation (AV), describes the variation between average measurements made by different operators on the same measurement device on the same part. A good reproducibility ensures similarity of the result obtained by different operators. A combination of the two factors makes up the total variation in the measurement system (Gage Repeatability and Reproducibility, GRR).

Good repeatability but poor reproducibility can mean that one operator provides consistent results, but there is a significant variation in the results between operators (Sharma et al., 2019). Conversely, poor repeatability but good reproducibility typically means that the measuring instrument itself is unreliable, even if the average results between operators are close to each other. The overall variation (GRR) of a measurement system reflects the combined effect of repeatability and reproducibility. It is often estimated using statistical techniques like analysis of variance (ANOVA) and stated as a percentage of the process tolerance (Saikaew, 2018). A small value of GRR is preferable since it means that the majority of the variation witnessed in the measurement data is because of actual product variation, and not because of the measurement system.

3.4.4 Measurement system acceptance criteria

The acceptability of a measurement system is generally assessed based on the GRR value in relation to the process tolerance (AIAG, 2010). General guidelines suggest that a measurement system with a GRR less than 10% is considered acceptable, while values between 10% and 30% may be conditionally acceptable based on the application’s criticality. However, acceptability should not be determined solely from the total GRR value, but also from the individual contributions of repeatability and reproducibility

The inherent tolerances of torque tools and torque testers are important from the perspective of MSA. The stated tolerance or accuracy of the tool and tester determines its inherent variation (ISO 6789-1:2017). It helps to draw reliable conclusions when the tolerances are low. if the tolerance of the torque tester is large compared to the tolerance of the torques being measured, it can significantly increase the GRR value of the measurement system and thus make it difficult to draw reliable conclusions. For this reason, the total variation of the measurement system cannot be significantly smaller than the inherent tolerances of the measuring devices used.

In the developed torque management system, production workers perform the measurements, which is why user variability, or reproducibility, is particularly relevant. Different users may apply the tool differently, interpret the readings differently, or position the tool inconsistently during verification (Bickford, 2018). As a result, reproducibility may be major cause of variability, even with accurate equipment. However, according to Sennaroglu and Yurtsever (2018), the variability due to the operators can be reduced through clear instructions, standardized procedures and proper user training.

A high GRR value of a measurement system can lead either to false alarms or to true tool deviations remaining undetected, which reduces the efficiency and reliability of the entire torque management system (AIAG, 2010). Therefore, MSA is essential in the present thesis, as the verification process is carried out by production workers and the resulting data are later used for tolerance setting, alarm logic, and interval determination.

3.5 Statistical process control

Statistical process control (SPC) is defined as a process that uses statistics to monitor and control the manufacturing process (Teixeira et al., 2017). Its aim is to guarantee consistent operation and satisfactory outcomes in accordance with the set quality standards. The control chart is the principal tool utilized in SPC. It provides an opportunity to distinguish between common cause variance and special cause variance (Ostadi et al., 2020). Common cause variance is inherent in the system and can be expected, whereas special cause variance is caused by distinct reasons that operate outside the process and affect it. These identifiable factors drive the process into an out-of-control state, requiring immediate investigation and corrective action upon detection. Using SPC tools enables manufacturers to identify issues that may lead to mistakes and eliminate them before they take place

3.5.1 Control charts in torque tool monitoring

Control charts provide a practical methodology for visualizing torque data and identifying process instability in production environment (Montgomery, 2009). When data is collected in subgroups under similar conditions, $\bar{X}$ and R charts serve as the primary analytical tools. The R chart evaluates variation within the subgroups, whereas the X chart plots the mean values of the subgroups to detect the process mean shift. (Leavengood & Reeb, 2015). Subgroups must be sampled under similar conditions, so that the variation within the subgroup reflects only common causes. If the R chart remains stable, the X chart can accurately identify mean shifts. However, if the R chart indicates an out-of-control signal, it indicates that the process is unstable.

The interpretation of control charts is primarily based on control limits, which are typically set at ±3 standard deviations from the centerline. If data falls outside of these limits, it indicates that the process is affected by a special cause variation. In addition to single-points signals, non-random patterns such as sustained runs or monotonic trends can also indicate a loss of statistical control (Leavengood & Reeb, 2015). For this reason, supplementary rules such as the Western Electric or Nelson rules are often used in addition to the traditional ±3σ limits. These principles are highly effective for monitoring torque tools, since they enable the identification of both sudden changes and drift.

3.5.2 Process capability indices (C_p, C_{pk})

Process capability indices quantify how well a process can produce results within specified tolerances (Karaman & Kulahci, 2025). The process capability index C_p measures potential capability based on process spread alone, ignoring centralization. C_p is defined as:

$$C_p = \frac{\mathrm{USL}-\mathrm{LSL}}{6\sigma} \tag{1}$$

Where USL and LSL are the upper and lower specification limits and σ is the standard deviation of the process. A C_p value of 1.0 means that the natural ±3σ exactly covers the tolerance width. A C_p value greater than 1 means that the process spread is narrower than the specification limits. C_p assumes that the process mean is centered between the limits.

The process capability index C_{pk} takes into account both the spread and the location of the mean. It is defined by the one-sided indices C_{pl} and C_{pu} as follows:

$$C_{pl} = \frac{\mu-LSL}{3\sigma} \tag{2}$$

$$C_{pu} = \frac{USL-\mu}{3\sigma} \tag{3}$$

$$C_{pk} = min\left[C_{pl}, C_{pu}\right] \tag{4}$$

Here μ is the process average. C_{pk} measures the distance from the process average to the nearest specification limit in σ units (Karaman & Kulahci, 2025). In other words, C_p reflects variation only, while C_{pk} reflects both variation and centering. The relationship between these two indices reveals the alignment of the process. If the process is perfectly centered, C_{pk} and C_p are equal, whereas a C_{pk} lower than C_p indicates that the process mean has shifted toward one of the specification limits. A common industry benchmarks classify a C_{pk} exceeding 1.33 as optimal, ensuring the process remains well within tolerances. A value in the 1.0 to 1.33 range is deemed adequate and any C_{pk} below 1.0 is classified as incapable as the process mean is too close to the specification limits.

These indices are useful in monitoring torque tools, as they help to distinguish between two basic quality situations. A process can be incompetent, either because of excessive variation or because it is inherently repeatable but has systematically deviated from the target values. This distinction is important from a practical point of view. High variation may indicate instability or wear, while deviation from the target value indicates the need for calibration or centering. However, capability indices should only be interpreted when the process is stable, as an unstable process does not provide a reliable basis for assessing capability.

3.5.3 Control and specification limits

Notably, the control limits on an SPC control chart do not represent product specification limits. Control limits are based on process data and represent the “voice of the process” (Ravichandran, 2019). Meanwhile, specification limits represent the customer’s or design specification requirements and thus represent the “voice of the customer.” In other words, control charts plot process data and not pass/fail quality. Plotting a specification limit on a control chart can be misleading.

This distinction is particularly important in torque assurance. Torque tools can remain within the upper tolerance limits even if they show signs of statistical instability, such as a continuous shift or trend in the control charts. Conversely, a statistically stable process can still be poorly centered relative to the tolerance limits. For this reason, control charts and the capability index complement each other. SPC indicates whether the process is statistically under control, while the tolerance-based assessment surface determines whether the tool performance remains acceptable relative to the required tolerance.

In some applications, warning zones are also used between the centerline and the control limits to support early wear detection. These zones are used in connection with the previously mentioned Western Electric and Nelson rules. These warning limits do not replace the actual control limits, but they can be useful complementary indicators when the aim is to react before a clear loss of control or tolerance is exceeded. In the context of this thesis, this distinction forms an important theoretical basis for subsequent analyses regarding alarm logic, tolerance setting, and data-based condition monitoring of torque tools.

4 Methodology

This chapter discusses the research methods and approaches that are essential to answering the research questions and achieving the objectives. It also examines the research strategy, data collection methods, data analysis methods, and the reliability and validity of the study. In addition, the chapter describes in detail the design and implementation process of the momentum tool management system, as well as the stages of the measurement process, data collection intervals, and set tolerances.

4.1 Research strategy and research process

This thesis applies a design science research strategy combined with quantitative empirical evaluation, as the goal is to design and implement a new torque tool control system and validate its performance using measurement data collected from a real industrial environment. Design science is well suited to this research because the primary output is a practical artifact, i.e. a system, and the research systematically evaluates how well that artifact meets the requirements set for it. The research approaches the problem from a practical and engineering-oriented perspective with the goal of improving the reliability, transparency and efficiency of joining processes by utilizing real-time measurement data. The empirical part of the research is based on quantitative data generated by a spreadsheet-based system and the results are used to make reasoned recommendations for the main parameters of the system.

The research process proceeds in four main phases. First, a literature review is conducted to identify theoretically based requirements and recommendations for torque verification, such as tolerance limits and the required number of repeated measurements. The literature review provides initial target criteria and a reference point for later review. These requirements and limits are used in the piloting phase of the system. In the second phase, the developed system is demonstrated in a pilot phase, where torque measurements are performed in practice and at the same time the system's data acquisition capability and operational feasibility in the designed environment are assessed.

In the third phase, a quantitative data analysis is performed on the collected data, with the aim of evaluating the performance and measurement behavior of the system. The analysis is used to determine empirically justified values for parameters, such as tolerance limits and the required number of measurements. In the fourth phase, these empirically derived recommendations are compared with a literature-based reference level and final guidelines and recommended parameters are developed for the use of the system. This research strategy ensures that the developed solution is based on existing knowledge, but is also validated and finalized using real measurement data.

4.2 Data collection

The data was collected over a two-month pilot study that was implemented within the production environment. Measurements were taken twice a day throughout the pilot study. The pilot study focused on one production line. All torque tools on this production line were part of the data collection, excluding pulse tools. Torque verifications were performed in a series of ten measurements.

Two Norbar torque testers were used to make the measurements. The torque testers were integrated into the torque tool management system designed in the study. One Norbar torque tester had a measuring range of 1-10 Nm, while the larger Norbar had the measuring range of 10-65 Nm. The values obtained from the measurements were directly fed into the system eliminating the possibility of error in writing them down manually. The measurements in the pilot were all done by the author of the thesis.

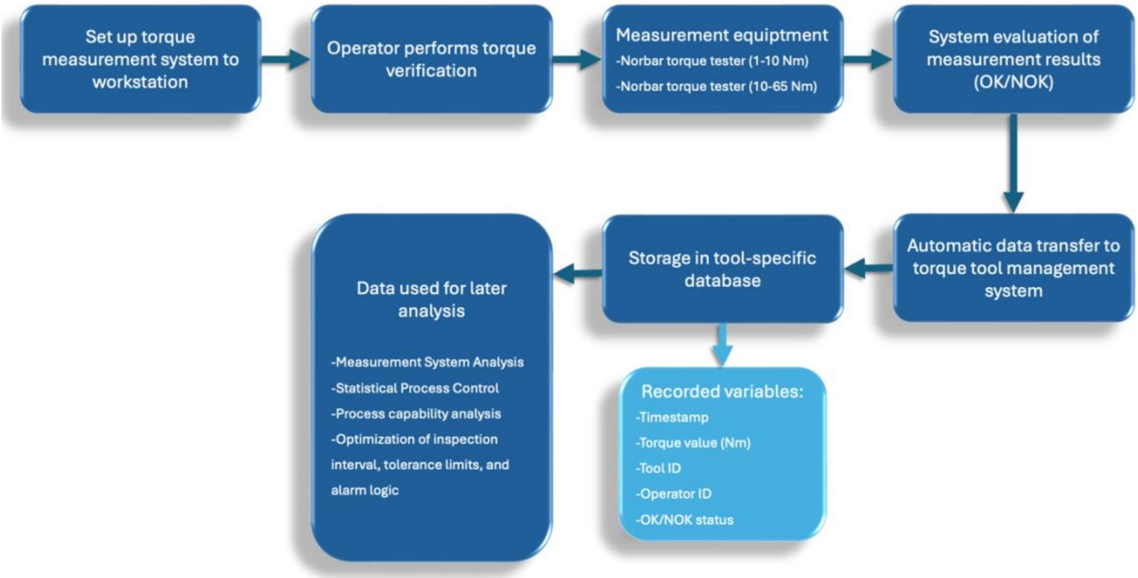
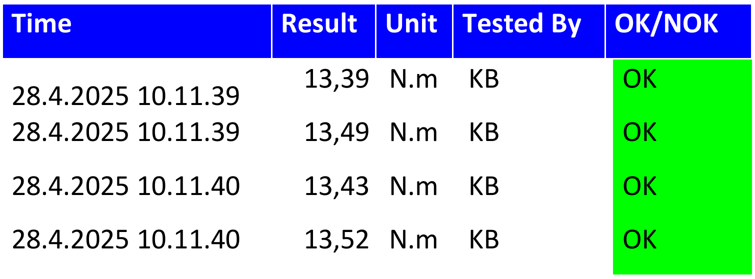


Figure 2. Data collection process of the pilot study.

Time stamps, torque values in units of Nm, tool and operator identifiers and tolerance-based OK/NOK status were recorded for every measurement. The twice-a-day measurement schedule enabled a sufficient data set to be generated in a relatively short period to discern tooling behavior as well as to answer research questions designed to develop an understanding of process conditions. Since data was relayed automatically from the testers to the system, potential errors related to manual data entry were avoided entirely. While primary data for the pilot was collected by a single operator, a separate Gage R&R analysis involving three operators was conducted as part of a measurement systems analysis. This was done to validate the measurement system and ensure that the results reflected tool performance rather than being influenced by individual operator technique.

Table 1. Sample of one torque measurement series recorded by the system.



Time	Result	Unit	Tested By	OK/NOK
28.4.2025 10.11.39	13,39	N.m	KB	OK
28.4.2025 10.11.39	13,49	N.m	KB	OK
28.4.2025 10.11.40	13,43	N.m	KB	OK
28.4.2025 10.11.40	13,52	N.m	KB	OK

28.4.2025 10.11.41	13,49	N.m	KB	OK
28.4.2025 10.11.42	13,58	N.m	KB	OK
28.4.2025 10.11.43	13,48	N.m	KB	OK
28.4.2025 10.11.43	13,58	N.m	KB	OK
28.4.2025 10.11.44	13,36	N.m	KB	OK
28.4.2025 10.11.45	13,48	N.m	KB	OK

4.3 Data analysis methods

The quantitative data collected by the developed torque tool management system is analyzed using descriptive statistics and established quality engineering methods. The analysis has two main objectives. The first is to evaluate the performance and suitability of the developed system for production use, and the second is to form empirically based recommendations for key system parameters, such as tolerance limits, alarm thresholds, measurement set size and measurement frequency.

First, the suitability of the measurement process is assessed using a measurement system analysis, more specifically a Gage R&R study. The total observed variation is divided into repeatability and reproducibility components, and these are reported in relation to the total variation and the tolerance range used. This step ensures that the measurement system is capable of supporting reliable pass/fail decisions and subsequent process monitoring.

In the second phase the behavior of the torque tool verification results as a function of time is evaluated using statistical process control. The measurement series are organized into rational groups and analyzed using Shewhart $\bar{X}$ -bar and R control charts. Standard rules (e.g. Western Electric or Nelson rules) are applied to detect process instability, trends and shifts. Based on the observed stability and change patterns, alarm limits and multi-level alarm logic are defined, which enables early detection of tool drift and supports timely maintenance actions. After this, once the process stability has been verified, the long-term ability of the tools to meet the set requirements is evaluated using process performance analysis. The performance indices C_p and C_{pk} are used to determine the variation in relation to the specification limits and to distinguish between high variation, indicated by a low C_p, and poor performance due to mean shift, indicated by a low C_{pk}. These results support tool-specific classification and prioritization of maintenance needs.

The data is analyzed using descriptive statistics such as mean, standard deviation, and coefficient of variation, as well as confidence intervals. These methods provide an overview of the data distribution, enable comparisons between different items, and create a statistical basis for determining the optimal parameters of the system. The findings of statistical process control are utilized to assess the recognizability of deviations in relation to the natural variability that occurs in the process. The goal is to find a trade-off between the fastest possible response time and the resource requirement for analysis. Finally, the results of all analyses are synthesized into a torque tool performance analysis that summarizes the tool point performance, stability, capability, and recommended actions.

5 Developed torque tool management system

This chapter describes the developed torque tool management system in detail. The system was developed as a spreadsheet-based software solution using a programmable scripting environment. The main objective of the developed system is to optimize the management of torque tools in the case company's production process. It includes automatically recording the torque measurement performed by production workers, checking the measurements against tolerances, and generating an alarm notification when the tolerance limit is exceeded. The company's torque tools have traditionally been managed manually. This has brought with it challenges related to data scattered in many locations, human errors in the process, and even late responses to challenges. These challenges were intended to be addressed with a system that provides a platform on which torque measurements are made. It also addresses the challenges with a system that automatically inputs the torque measurement data performed by production workers, thus avoiding errors related to human actions. In addition, the system uses automatic alarm notification, which is directed to the calibration team when measurement values deviate from preset tolerance limits.

5.1 System architecture and components

The developed torque tool management system has been implemented as a spreadsheet-based software solution with built-in automation. The system structure is tool-centric, with each torque tool having its own separate spreadsheet. This enables a clear separation of measurement histories and allows tool-level tracking of status, trends, and verification outcomes over time.

The system consists of a measurement user interface that guides the user through the torque verification task. QR-based tool recognition and user identification input ensure that each measurement set is linked to the correct tool and user. The measurement results are stored in tool-specific spreadsheets that contain all measurement-related information. The core functionality is implemented with VBA-based calculation and decision-making logic that compares measured values to defined target values and tolerance limits, according to which it determines the resulting OK/NOK result. The system also includes a summary view that combines the status of all tools in the system for the calibration team. In addition, alarm and reporting functions are integrated into the system and are described in detail in the following sections. The system performs basic checks during torque checks, evaluating each set of measurements against defined limits and updating the tool status.

5.2 User interface and measurement process

The interface is designed to be straightforward and accessible for everyone. In addition to being easy to use, the user interface is also designed to enable production workers to perform torque tool checks quickly and consistently without any special technical expertise. The user interface provides step-by-step instructions and clear visual feedback. This, along with the automatic entry of measurements, reduces the likelihood of input errors and makes the check procedure a routine part of the production process.

After tool and user identification, the user is instructed to perform ten consecutive torque measurements with the torque tool. The required number of measurements is later estimated and justified again based on the data collected in the data analysis phase. The user interface provides real-time feedback during the measurement series, showing the value of each measurement and its classification in relation to the tolerance values (OK/NOK). This allows the user to see the progress of the series without additional calculations

Once the torque tool has been checked and the series is complete, the user interface allows the user to review the values before final recording, in case an error has occurred during the measurements. The error does not refer to a NOK measurement result, but rather an accidental measurement or incorrect execution. The user can delete the measurement in question and replace it with a new measurement. The purpose of this review option is to improve data quality while maintaining the principle that correctly performed measurements are recorded to ensure traceability.

5.3 Saving and managing measurement results

The system's saving and data management functions are designed to ensure data integrity, traceability and long-term preservation of measurement history. Measurement results are automatically saved after the user has confirmed and saved the measurement set. Each saved measurement set, in addition to the measured torque value, contains a time stamp, tool ID, and user ID. This enables full traceability and supports time-based analysis of tool behavior. User identification and storage of the ID is important, as it allows the tightenings performed to be assigned to a specific user if necessary and allows the measurement history to be viewed at a personal level.

The results are stored in a structured format so that each torque tool maintains its own measurement history. Measurement series are stored and grouped as a set of consecutive values together with other related information. In addition to the measured torque values, the system also stores OK/NOK statuses derived from the tolerance limits of the measurements. This structure supports both operational decision-making, such as identifying tools that require attention, and subsequent statistical analyses, including SPC, capability analysis and tolerance optimization.

To maintain the reliability of the data set and prevent manipulation of the results, the system is designed to ensure that correctly made measurements are retained in the measurement history. The operator can edit the measurements only before final saving, allowing clearly incorrect measurements or annotations to be deleted and replaced. This approach balances practical usability with the need for consistent and reliable measurement data.

5.4 Alarm system

The management system of the developed torque tool management system is a key part of its functionality. The primary objective of the alarm system is to enable a rapid response to deviations in the performance of torque tools and thus prevent potential quality problems in the production process. Several alarm criteria are integrated into the system, which activate an automatic notification to the calibration team when predefined limits are exceeded. This rapid response approach allows for the immediate initiation of corrective actions, minimizing the use of defective tools and the costs they may incur.

The alarm logic is based on two criteria categories in the pilot phase. One criterion category is failed measurements (NOK) detected during a single measurement session. The system generates an alarm if one or more measurement results exceed or fall below the set tolerance limits. Such an alarm may indicate that the tool requires calibration, or has a mechanical fault. The second alarm criterion is related to the calibration history of the torque tool. The system continuously monitors the last calibration time of each tool. If more than six months have passed since the previous calibration, the system automatically sets the tool status to NOK on the summary page. This reminds the calibration team of the expiration of the calibration interval and ensures that torque tools remain regularly calibrated to maintain measurement accuracy.

In addition to these criteria, the system supports SPC-based monitoring. This allows the system to detect early signs of drift or instability before the actual NOK result is issued. SPC-related thresholds, such as continuous shifts or trends, and driving rules are parameterized based on the pilot data. These are implemented and refined in the data analysis phase (Chapter 6). The pilot phase of the work focuses on collecting measurement data and monitoring NOK observations using specification limits. SPC-based limits are adjusted to balance sensitivity and false alarm rate according to the process behavior.

When the alarm criteria are met, the system automatically generates and sends an alarm email to the calibration team. Through the alarm message, the calibration team receives information about the torque tool that caused the alarm, the time of the alarm, and information about the tool’s location. This clear and informative alarm enables the calibration team to react quickly and take the right measures.

The designing of the alarm system aims to achieve a balance between the sensitivity of the system and workload. A system that is very sensitive may trigger alarms that can be a burden to the calibration team, whereas insufficient sensitivity may fail to indicate malfunctions when they occur. The initial pilot settings had tolerance limits of ±10%. The final tolerance limits, alert thresholds and possible warning rules (SPC) are defined and justified using empirical data and qualitative analyses, which are presented in Chapter 6.

5.5 Reporting and summary view

The reporting and summary view of the developed torque tool management system is designed to provide a real-time, concise, and easy-to-read snapshot of the status of all the torque tools in the system. This overall view allows various stakeholders within the firm, such as production management and the calibration team, to quickly assess the overall condition of the tools. The system also allows for increased insight into the performance of an individual tool. By combining a real-time overview with access to detailed measurement history, the system improves transparency and supports predictive calibration planning

The summary page includes direct links to the individual torque wrench measurement data pages. This allows for easy and quick navigation to view the individual measurement history and performance trends of an individual tool. Each torque tool's own page displays all measurements performed for that tool in chronological order, including measurement times, individual measurement results and their statuses

(OK/NOK). In addition, the tool page has an updated graphical representation of the measurement results over time. This visual element is valuable and useful in analyzing the tool's performance. This visualization supports identification of longer-term behavior such as gradual drift, increasing variability, or other deviations. The chart also helps interpret measurement history in relation to defined tolerance limits, making it easier to detect potential problems without manual data processing.

Overall, the reporting and summary view of the fully developed torque tool management system provides an effective tool for torque tool management. The real-time overview combined with the ability to view detailed measurement history and visual presentations enables predictive maintenance, supports and improves decision-making, and thus promotes the reliability and efficiency of joining processes in the production environment of the case company.

6 Results and analysis

This chapter presents the analysis of the measurement data collected during a two month of operation of the developed torque tool management system. A total of 50 tools were measured during the monitoring period, and 25 566 measurements were recorded. In the first monitoring period of approximately four weeks, the system was used to record measurements without activation of real-time alarms. This four-week monitoring period allowed to understand the normal operation of a torque wrench as part of the daily production process. It allowed to map the natural variation of the measurement results within the preset tolerance limits of ±10 percent. The first phase was implemented without real-time alerts to provide a solid and empirical foundation for future optimization work.

During the second month, real-time alarms were activated to evaluate the system performance in a real-world situation where alarms are responded to. In parallel, we investigated the effects of calibration on tool performance: whether the inherent variability per tool remains stable after calibration, how the C_p and C_{pk} indices change before and after calibration, and whether the process can be obtained and maintained in statistical control when calibrations are applied. Analyses were performed at both the individual tool level and at aggregate level using standardized quality engineering methods, including measurement system analysis, statistical process control and control charts, and process capability assessment.

The collected data is used to derive data driven recommendations for key system parameters and to compare these findings against the literature-based baseline established earlier in the thesis. In particular, the aim is to determine appropriate tolerance and alarm limits, the required number of measurements per verification series and the recommended measurement interval for periodic tool checks. The results of the analyses support the evidence-based development of the system and guide its practical implementation in the case company.

6.1 Gage repeatability and reproducibility (Gage R&R)

This chapter evaluates the measurement system used to verify production torque tools by means of a Gage R&R (Measurement Repeatability and Reproducibility) analysis. The analysis followed the AIAG (Automotive Industry Action Group) measurement systems analysis, 4th edition, and ANOVA analysis was performed to determine measurement variation. The objective of the analysis is to determine how much of the observed variation in measured torque values is due to the measurement system itself. The analysis provides estimates of repeatability (Equipment variation, EV) and reproducibility (appraiser variation, AV), which together form the overall Gage R&R. The results are evaluated with the principles of the AIAG, 4^{TH} edition guideline, using the AIAG acceptance criteria for %GRR values (<10% = acceptable, 10-30% = inadequate, >30% = unacceptable). Based on the study, conclusions are drawn about the suitability of the measurement system for production use.

6.1.1 Measurement system description and methodology

Ten torque tools (P1-P10) were selected to cover the relevant operating range. The 10 tools in the analysis represent different torque levels covering the range of 1 Nm to 25 Nm. Three operators measured each tool five times, resulting in total of $3 \times 10 \times 5 = 150$ observations. The measurements were performed under controlled and consistent conditions. No special causes or deviations were observed during data collection, therefore the observed variation is assumed to be due to the measurement system and natural differences between the tools. The objective is to assess whether the torque testers and the operators can measure the torque of each tool with sufficient consistency and accuracy for production needs. The measurement system is considered acceptable for production if the variation it causes is only a small fraction of the observed total variation or tolerance according to the AIAG guidelines. In addition, the system should be able to sufficiently distinguish between the variations between parts (number of distinct classes, ndc ≥ 5).

The analysis first applied the average and range method as an initial consistency check. The X-R analysis led to the same overall conclusion regarding acceptability. However, the main analysis and results are based on the two-way ANOVA approach, as it provides a more comprehensive variance decomposition for a grossed Gage R&R design, estimating the interaction component between operators and tools.

The ANOVA identifies four sources of variation: Part-to-Part, Appraiser, Part-to-Appraiser and Repeatability. The variance components were calculated from the ANOVA table´s mean square values and expressed in standard deviation for easier interpretation. Reproducibility is specified in its subcomponents and as a combined reproducibility term. Furthermore, the number of distinct categories was calculated to assess the ability of the system to differentiate the levels of the torque.

6.1.2 ANOVA analysis and the results

Using a two-way analysis of variance, we obtain a more detailed separation of measurement variation, including the interaction term between appraiser-by-part. Table 2 reports the ANOVA results with degrees of freedom (DF), sum of squares (SS), mean squares (MS), F-ratios, and p-values. The variance components are then extracted from the ANOVA to calculate the standard deviation and their contributions from each source. The part-to-part effect is very large and significant ($F \approx 2890$, $p<0.001$), confirming that the different parts have significantly different torque values. The main effect of appraiser in the comparison is negligible ($F=2.93$) and is not statistically significant at the $\alpha = 0.05$ level ($p = 0.079$). This indicates that there is no consistent overall bias across the three appraisers. However, the interaction effect between part x appraiser is significant ($F \approx 16.31$, $p<0.001$). This means that the consistency of appraisers’ performance varies from part to part, i.e., certain items were measured differently by different operators, even though their overall means were similar. Repeatability error (within-trial) captures the residual variation caused by trial-to-trial variation.

Table 2. Two-way ANOVA for the Gage R&R study.

Source	DF	SS	MS	F	p-value
Part-to-Part	9	8604,93	956,103	2890,31	<0,001
Appraiser	2	1,936	0,968	2,93	0,0794
Part × Appraiser	18	5,956	0,331	16,31	<0,001
Repeat (Error)	120	2,434	0,0203	–	–
Total	**149**	**8615,25**	–	–	–

The root mean square variance components were estimated for each source:

- Repeatability Variance: $\alpha^2_{repeat} = MS_{error} = 0{,}02302$
- Part-to-Appraiser interaction variance: $\alpha^2_{Part\times App} = \frac{MS_{Part\times App} - MS_{Error}}{n} =$ $\frac{0.331-0{,}0203}{5} = 0{,}0621$
- Appraiser. Variance: $\alpha^2_{App} = \frac{MS_{Appraiser} - MS_{Part\times App}}{p\cdot n} = \frac{0.968-0{,}331}{10\cdot 5} = 0.0127$
- Part-to-Part Variance: $\alpha^2_{Part} = \frac{MS_{Part} - MS_{Part\times App}}{p\cdot n} = \frac{0956{,}103-0{,}331}{3\cdot 5} = 63.718$

These variance components are summarized in Table 3, together with the corresponding percentage contributions to the standard deviations (σ), %contribution (each component's contribution to the total variance) and the contributions to the %Study variation, (each component's contribution to the total standard deviation). The number of distinct categories has also been calculated from the ANOVA results.

Table 3. Variance components from ANOVA, with standard deviations and contribution percentages.

Source of Variation	Variance (σ2)	Std. dev (σ)	%Contribution	% Study Var.
Rebeatability (EV)	0,0203	0,1424	0,03 %	1,8 %
Reproducibility (AV) – Appraiser	0,0127	0,1129	0,02 %	1,4 %

Reproducibility (AV) – Part×App.	0,0621	0,2493	0,10 %	3,1 %
Total reproducibility (AV)	**0,0748**	**0,2737**	**0,12 %**	**3,4 %**
Gage R&R (GRR)	**0,0951**	**0,3085**	**0,15 %**	**3,9 %**
Part-to-Part (PV)	63,7180	7,9849	99,85 %	99,9 %
Total Variation	**~63,813**	**7,994**	**100 %**	**100 %**
Number of distinct Categories (ndc)		**~36**		

In the table3, Reproducibility is divided into Appraiser and PartXAppraiser components for clarity. Total Reproducibility is the sum of these components. %Contribution is based on variance, while % Study variation is based on standard deviation (total variation is taken to be 7.994). According to the ANOVA method, the standard deviation of the repeatability of the measuring device is σ_E = 0.1424 and the standard deviation of the total repeatability (including inter-measurer differences and user x part inconsistency) is σ_A_total = 0.2737. Combining these gives GRR=0.3085. The standard deviation for part-to-part is σ_P =7.985, which completely controls the total variation. As a result, the %Study variation of the measuring device is only about 3.9% of the total. The ndc calculated based on the ANOVA is approximately 36, which is clearly above the adequacy threshold. This confirms that the measurement system is capable of distinguishing dozens of different component values.

6.1.3 Gage R&R analysis results

It is important to consider the application requirements. For torque checks in production, each tool has an acceptable tolerance range, which is ±10% of the nominal torque. The ANOVA-based Gage R&R results show that the measurement system variation is small relative to the observed tool-to-tool variation. The total Gage R&R standard deviation is less than 4%. This means that the torque measurements are highly repeatable and reproducible for the intended use. In practical terms, the observed differences are primarily driven by real differences between the torque tools rather than measurement noise. Thus, the system can reliably identify whether a torque tool is within specifications or requires adjustment.

According to AIAG (2010), when the overall %GRR is less than 10%, the variation of the measurement system is considered acceptable for production use and strict process control. According to AIAG criteria, less than 10% is acceptable, 10-30% is avoidable, and more than 30% is rejected and should not be used. In this case, the system is well below the good range. From a practical point of view, the measurement system is more than adequate. With a %GRR of approximately 3,9%, the measurement system falls well within the acceptable range. This confirms the system's reliability for routine production verification and its capability for effective process monitoring. In addition, the ANOVA-based number of distinct categories is approximately 36, which is well above the recommended minimum of 5. This indicates that the system can reliably distinguish differences between tools across the studied torque range and provides sufficient resolution for identifying meaningful deviations in tool performance.

6.2 Statistical process control and out-of-control detection

The proper use of statistical process control methods requires the division of data into rational subgroups (Leavengood & Reeb, 2015). In the developed torque tool system, each tool was measured with 10 consecutive torque measurements during a single inspection. Therefore, it is methodologically justified to choose a subgroup size of $n=10$. This choice more accurately reflects process conditions where short-term variation caused by tool repeatability and operator's instantaneous performance is isolated within one subgroup, while long-term variation caused by calibration drift and wear is observed between the mean values of different subgroups over time. Since the collected data is continuous variable data, two Shewhart control charts are needed. One chart for the process average level ($\bar{X}$ Chart) and the other for monitoring process variation (R Chart).

In this study, Shewhart's X̄ and R control charts were applied to analyze continuous variable data of torque measurements. The X̄ chart tracks the stability of the process average level over time, while the R chart examines the predictability of the process's internal dispersion. The selected subgroup size was n=10. This size provides sufficient sensitivity to small shifts in the process center while benefiting from the normalizing effect of the Central Limit Theorem, thereby ensuring the reliability of the results even in potentially non-normally distributed data.

6.2.1 Calculation of control limits

The control limits are derived from the entire historical data distribution, using two months of measurement data to assess stability. The calculation used standardized coefficients that depend on the subgroup size (n=10), as shown in Table 4. The process standard deviation estimate is based on the average range of the subgroups (R)

Table 4. Control chart constans for subgroup size n=10.

Constant	*Description*	*Value (n=10)*
A_2	Factor for X̄ UCL/LCL	0.308
D_3	Factor for R LCL	0.223
D_4	Factor for R UCL	1.777
d_2	Factor for a	3.078

The baseline control limits are calculated as follows.

A. Calculation of process averages (X̄ and R)

The mean (X̄$_j$) and range (R_j) are calculated for each subgroup (j, where n=10). After this, the overall mean (X̄) and the average range (R) for all k subgroups are calculated:

Overall average:

$$\bar{X} = \frac{\sum_{j=1}^{k} \bar{X}_j}{k} \tag{5}$$

Average range:

$$\bar{R} = \frac{\sum_{j=1}^{k} \bar{R}_j}{k} \qquad (6)$$

B. R-chart control limits

The R chart monitors the level of process variation. If the R chart is unstable, the X chart limits will also be unreliable:

Centerline (CL_R):

$$\mathrm{C}L_{\bar{R}} = \bar{R} \qquad (7)$$

Upper Control Limit (UCL_R):

$$\mathrm{UC}L_{\bar{R}} = D_4\bar{R} \qquad (8)$$

Lower control limit (LCL_R):

$$\mathrm{LC}L_{\bar{R}} = D_3\bar{R} \qquad (9)$$

It is worth noting that in the case of n=10, D3 is greater than zero, and LCL_R is therefore positive. A point that falls below LCL_R is a signal of a special variation, for example when the operator starts measuring very carefully, or the measurement device locks up.

C. X̄ chart Control limits

The X̄ chart monitors the average level shifts of the process.

Centerline (CL_X):

$$\mathrm{C}L_{\bar{X}} = \bar{X} \qquad (10)$$

Upper Control Limit (UCLx):

$$\mathrm{UCL}_{\bar{X}} = \bar{X} + A_2 A_2 \bar{R} \qquad (11)$$

Lower Control Limit (LCLx):

$$LCL_{\bar{X}} = \bar{X} - A_2 \bar{R} \qquad (12)$$

6.2.2 Torque tool population stability summary (n=10)

Control chart analysis was performed for all 50 torque tools using a subset size of n=10. As mentioned earlier, the deviation between the calculated mean (X̄) and the nominal torque (T) indicates systematic bias, while a large average range ($\bar{R}$) indicates poor repeatability. The table 5 below summarizes the selected tools and their stability status, considering the analysis parameters based on n=10.

Table 5. Summary of process status for selected torque tools.

Tool ID	**Nominal Torque (T) (Nm)**	**Calculated Mean (X̄) (Nm)**	**Average range ($\bar{R}$)(Nm)**	**OOC points (X̄)**	**OOC Points ($\bar{R}$)**	**Stability Status**
28	13.5	13.560	0.751	11	0	Stable
37	4.5	4.672	0.376	38	6	Unstable
8	24.0	24.879	0.991	11	0	Stable but bias

In data processing, the measurement results of each tool were arranged chronologically according to the time stamp. The data were then segmented into consecutive subgroups of size n = 10, allowing the average (X̄) and range ($\bar{R}$) of each subgroup to be calculated. The process medians were defined based on the data collected over the entire two-month period:

1. Overall Mean X̄: The average of the averages of all subgroups. This represents the historical average of the process, which serves as a reference for X̄ map control.
2. Average Range $\bar{R}$: The average of the R values for all subgroups. This is a measure of the variation within the process, from which the control limits and the estimated short-term standard deviation of the process (σ) are derived.

Tools with a high $\bar{R}$ value relative to the nominal torque such as tool ID 37, indicate high process variation, which may be an indication of mechanical wear or operator inconsistency in measurements. Such a large dispersion is critical and should be investigated in the Gage R&R analysis. Tools with a significant difference between X̄ and the nominal torque, such as tool ID 8 (24.00 Nm vs 24.879 Nm), require recalibration. SPC information can be directly used in maintenance planning: High R indicates a need for mechanical maintenance due to increased variation, whereas large difference between nominal torque and X̄ indicates a need for calibration due to bias.

6.2.3 Detailed case study: Torque tool performance

To illustrate the analysis method, selected torque tools representing different stability profiles are examined, based on an n=10 analysis.

6.2.3.1 Case study 1: Tool ID 28

The nominal torque of the tool ID 28 is 13.5 Nm. Based on the statistical process control analysis performed, its calculated average X̄ was 13.560 Nm, which indicates only a very small bias compared to the nominal value. The average variation range R was 0.751, and most importantly, no out-of-control points (OOC) were observed in the R chart, confirming the smooth variation of the process. Although a few OOC points were observed in the X̄ chart, it can be seen from the figure 3, that most of these points are really close to the control limits. The largest swing in the average was observed in subgroups 49-51. Despite small momentary deviations, the tool ID 28 operates within its target values and is statistically controlled overall, which indicates good process stability and proximity to the ideal state.

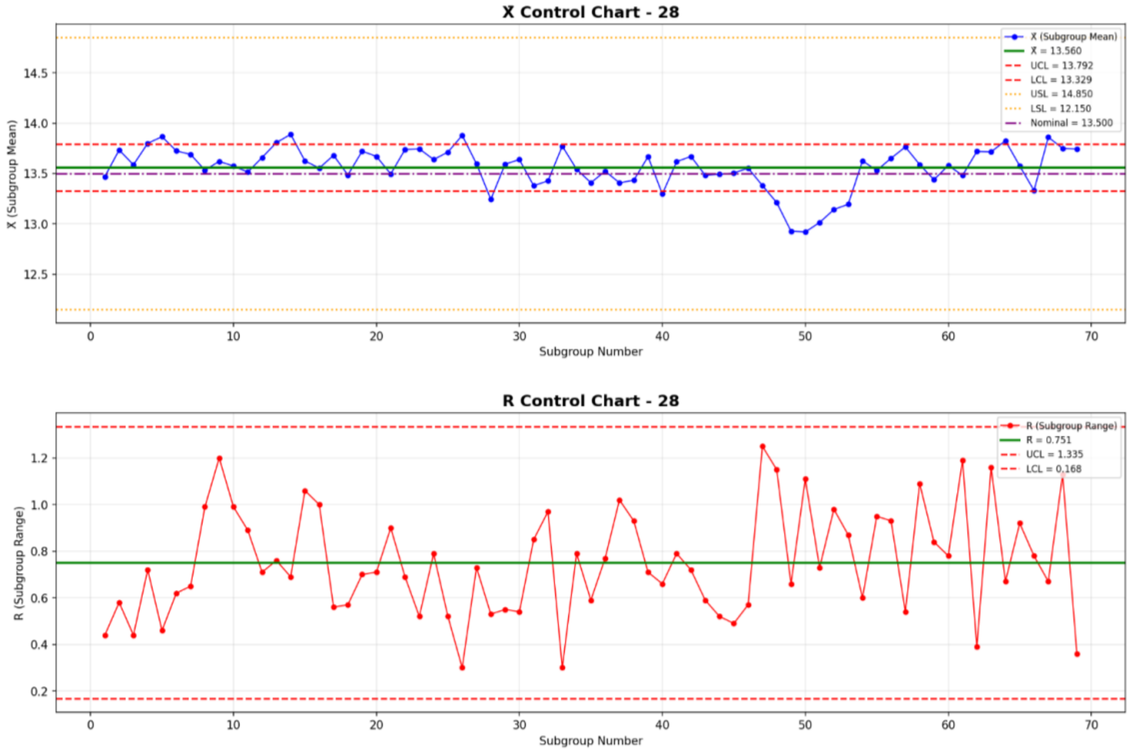


Figure 3. Tool ID 28 X̄ and R control charts.

6.2.3.2 Case study 2: Tool ID 37

As can be seen in Figure 4, tool ID 37, with a nominal torque of 4.5 Nm, represents the most critical stability state of the analyzed tools. Statistical process control indicates an unstable state requiring immediate action for this tool. Based on the calculated data, the tool had as many as 38 OOC points in the X̄ chart out of a total of 58 measurement points, in addition to which the nominal torque is clearly outside the control limits, indicating a significant positive bias. 8 OOC points were also observed in the R chart, which means that the repeatability of the tool has deteriorated and may indicate, for example, mechanical wear or damage. Since both SPC charts give strong instability signals at the same time, the process must be considered completely statistically out of control, and tool ID 37 must be immediately taken out of service for root cause analysis and thorough maintenance.

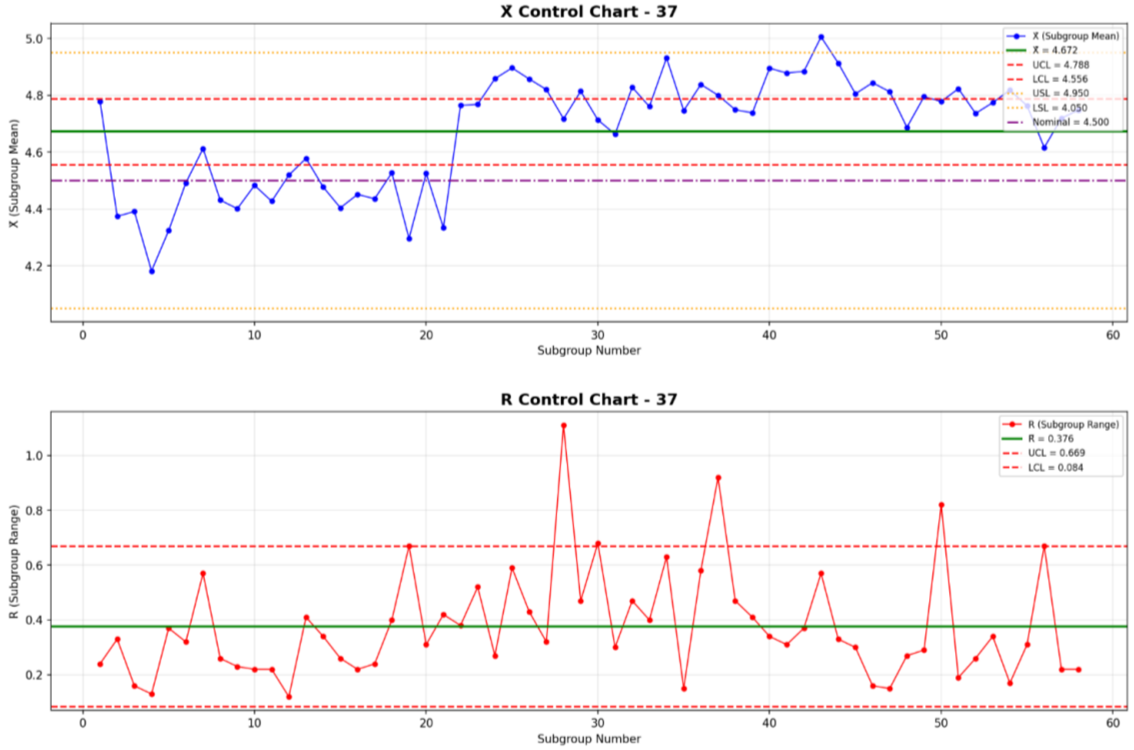


Figure 4. Tool ID 37 X̄ and R control charts.

6.2.3.3 Case study 3: Tool ID 8

Tool ID 8, with a nominal torque of 24 Nm, which does not have a single OOC point in the R chart, still appears problematic. The tool is noticeably mis centered, which is visible as a shift in the average (X̄ = 24.879Nm). A few OOC points were observed in the X̄ chart. Although no OOC points were observed in the R chart, which would indicate a momentary good repeatability, the continuous location of the average clearly above the set nominal torque is a clear indication that the tool has failed calibration or has been subjected to a systematic shift. Figure 5 shows that the nominal torque is not even within the control limits. The tool is therefore solid and stable, but it is mis centered and has a positive bias.

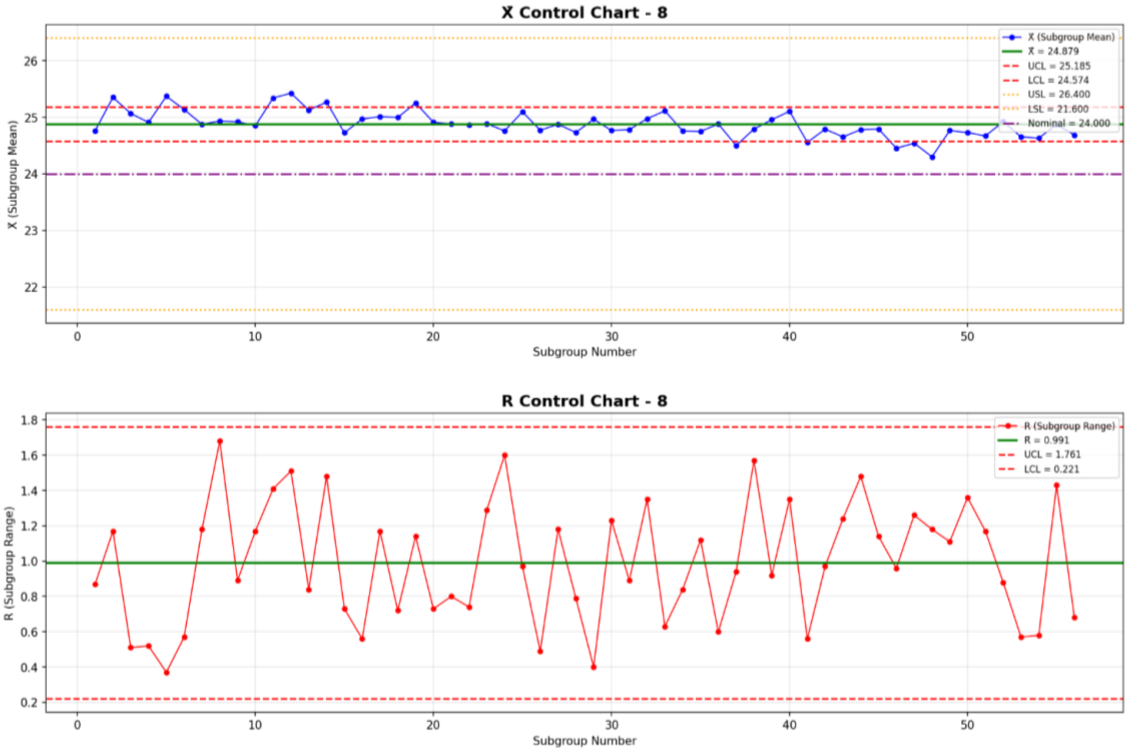


Figure 5. Tool ID 8 X̄ and R control charts.

6.2.4 Identifying special variation using Nelson's rules

Process stability assessment is not limited to exceeding the 3σ control limits (rule 1) (Cristian Pisarciuc, 2016). Standard industry practices, such as Western Electric or Nelson rules, require the identification of unnatural patterns. Applying these rules improves the sensitivity of the chart to detect subtle changes in the process at an early stage.

The chart is divided into three zones (A, B, C) on either side of the centerline (CL), using the process standard deviation as the base unit.

- Zone C: At a distance of ±1σ from the center line.
- Zone B: 1σ and 2σ distance from the center line.
- Zone A: 2σ and 3σ distance from the center line.

Standardized Nelson/Western Electric rules were applied to all 50 tool diagrams:

1. Rule 2 (Movement): Nine consecutive points on the same side of the centerline.
2. Rule 3 (Trend): Six consecutive points continuously rising or falling

3. Rule 5 (Approaching the Limit): Two of three consecutive points fall between 2σ and 3σ on the same side of the centerline
4. Rule 6 (Movement): Four of five consecutive points fall beyond 1σ on the same side.

Analysis using these rules revealed several sources of specific variation that would not have been apparent by monitoring the 3σ limit alone. This more sensitive detection would have allowed for proactive intervention. For example, in a case such as tool ID 8 ($\bar{X}$=24.879 Nm, T = 24.0 Nm), which did not have OOC points within the normal limits of the R-chart, Nelson's rule 2 and rule 6 would have triggered an alarm. This early signal would have indicated that although the process appeared to be stable externally, it was in fact systematically drifting or targeting the wrong target value, thus providing an opportunity to correct the problem and restore statistical control before the value rose above the 3σ limit or approached the specification limits.

The signals in the control charts must be linked to information contained in the measurement data identifying the operator who made the measurement. This allows for a precise separation of the source of the specific variation. If the $\bar{X}$ chart repeatedly shows a shift, it is likely a matter of tool calibration or wear. If the R chart shows instability, the source of the variation may be more complex. If a large dispersion R is consistently associated with a particular operator, the specific variation is likely due to operator inconsistency, e.g. a weakness in the measurement method. This separation allows for the cost-effective targeting of corrective actions.

6.2.5 Conclusion on process stabilization

The analysis presented in this section, based on the use of a sample size of n=10, provided a more refined statistical picture of the stability of the torque tool population. A larger sample size improves the ability of the $\bar{X}$ chart to detect even smaller shifts in the process mean. Although many tools show good short-term repeatability, as indicated by the R being chart under control, a significant proportion of tools suffer from systematic bias and/or instability, requiring immediate calibration or mechanical maintenance. Trends and transitions identified using Nelson's rules enable a shift from reactive maintenance to predictive maintenance. For each torque tool that has been statistically proven to be stable, a reliable process baseline and associated 3σ control limits have now been determined. This establishment of a stability baseline is essential and serves as statistical input for future data analysis steps, such as setting alarm limits and process capability analysis.

6.3 Process capability analysis (C_p, C_{pk})

The performance analysis was conducted on all the tools using a tolerance of ± 10% of the nominal torque (T), calculated as $LSL = 0.9T$ and $USL = 1.1T$. Capability was assessed for each tool using the indices C_p and C_{pk} . C_p describes how wide the process variation is relative to the specification width, while C_{pk} describes how well the process is centered within the specification limits. C_p and C_{pk} were computed using the standard normal-theory formulas. The analysis used the common benchmarks regarding C_p and C_{pk} measures in which a high-performing process was reflected by a C_{pk} or C_p of 1.33 or higher, a C_p or C_{pk} between 1.00 and 1.33 reflected a marginal process that had to be closely monitored, and a C_p or C_{pk} of less than 1.00 reflected a poor process that had to be improved. As shown in Figure 6, A total of 26% of the tools have a high-performing capability in relation to the benchmark in which $C_{pk} \geq 1.33$. A total of 28% of the tools have capability values between 1.00 and 1.33, while the remaining tools have a poor or unacceptable capability with a C_p or C_{pk} values below 1.00.

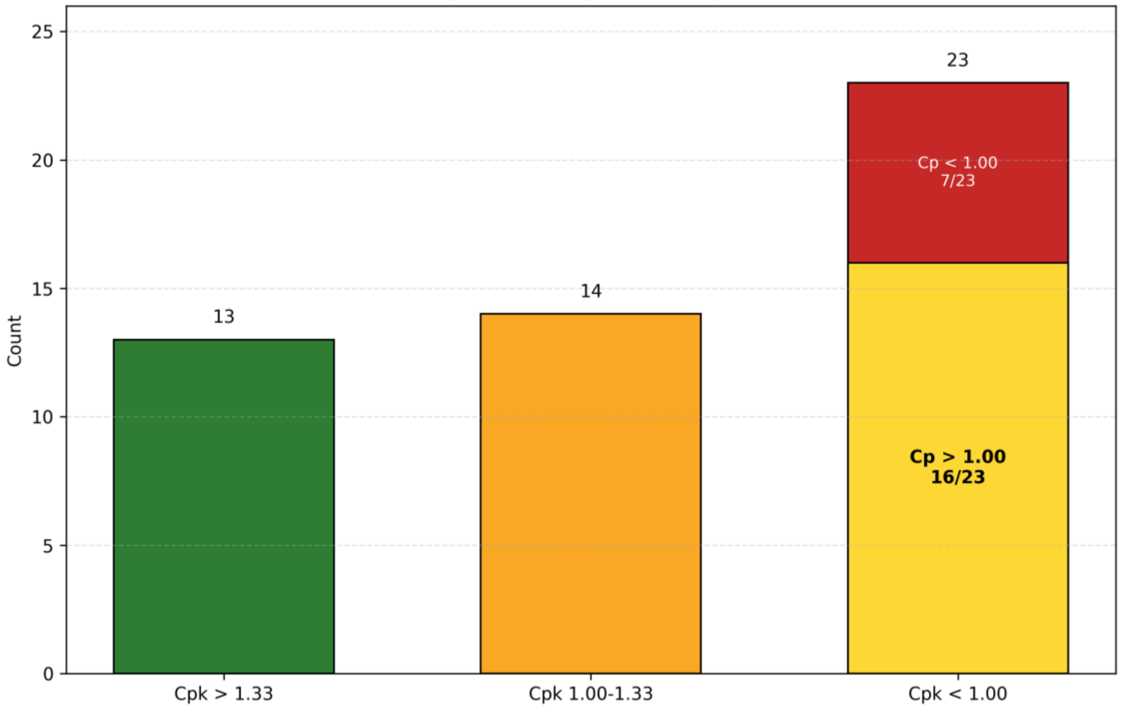
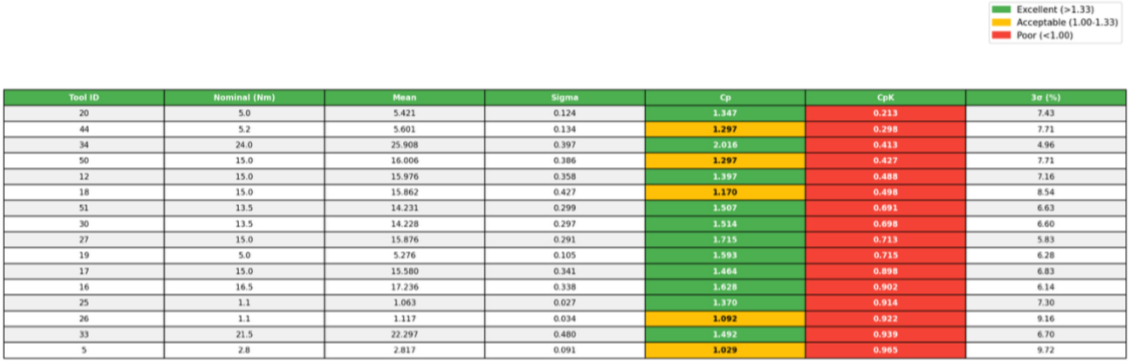


Figure 6. Torque tool capability distribution by C_{pk} category.

Tools with C_{pk} < 1.0 represent the most critical group, indicating that the tool does not have the ability to control torque within ±10% of tolerance on a consistent basis. Of 50 tools analyzed, 23 had a C_{pk} value below 1.0, indicating that they are inadequate at producing a result within the defined ±10% tolerance range and should be improved. Importantly, however, within this subset were 16 tools showing a C_p value above 1.0, indicating the natural variability of the tool is less than the specified tolerance limits. This means that, these 16 tools were not excessively variable. The natural variability of these tools would be acceptable if the process were properly centered.

The significance of having a $C_p > 1$ while the $C_{pk} < 1$ implies that the main source of the problem is not the random variation, but the systematic error in the process. The main effect of the mentioned tools is the production of torques with values that are systematically too high or too low from the nominal torques, with only the tail of the distribution outside the tolerance. This is a critical distinction. It means that the majority of the non-capable tools are not inherently defective, but rather improperly adjusted. For these 16 tools, calibration is expected to return the process mean to the center of the specification range and thereby would also increase the C_{pk} significantly, in many cases to above 1.0.



🟩 Excellent (>1.33)
🟨 Acceptable (1.00-1.33)
🟥 Poor (<1.00)

Tool ID	Nominal (Nm)	Mean	Sigma	Cp	CpK	3σ (%)
20	5.0	5.421	0.124	1.347	0.213	7.43
44	5.2	5.601	0.134	1.297	0.298	7.71
34	24.0	25.908	0.397	2.016	0.413	4.96
50	15.0	16.006	0.386	1.297	0.427	7.71
12	15.0	15.976	0.358	1.397	0.488	7.16
18	15.0	15.862	0.427	1.170	0.498	8.54
51	13.5	14.231	0.299	1.507	0.691	6.63
30	13.5	14.228	0.297	1.514	0.698	6.60
27	15.0	15.876	0.291	1.715	0.713	5.83
19	5.0	5.276	0.105	1.593	0.715	6.28
17	15.0	15.580	0.341	1.464	0.898	6.83
16	16.5	17.236	0.338	1.628	0.902	6.14
25	1.1	1.063	0.027	1.370	0.914	7.30
26	1.1	1.117	0.034	1.092	0.922	9.16
33	21.5	22.297	0.480	1.492	0.939	6.70
5	2.8	2.817	0.091	1.029	0.965	9.72

Figure 7. Tools with $C_{pk} < 1$ and $C_p \geq 1$.

The figure 7 shows a table listing tools with a process capability of less than 1 ($C_{pk} < 1$) and a C_p of at least 1. The last column of the table shows the 3σ results of the wrenches as a percentage. It can be seen from the table that all of the tools in question have a 3σ value of less than 10%. This observation supports the previous conclusion that calibrating and properly centering the tools could improve their C_{pk} values, to a level at which all of these tools would meet the acceptance criteria. These tools would potentially have the ability to maintain torque consistently within a 10% tolerance.

Only the remaining 8 tools, with both C_p and C_{pk} values below 1.0, appear to suffer from real excessive variation, indicating wear, mechanical instability, or inconsistent use. Thus, the data show that the majority of tools that had low performance in this system were not due to uncontrolled dispersion but to improper adjustment.

6.3.1 High-capability tools ($C_{pk} \geq 1.33$)

The ten best performing tools with the highest C_{pk} values are shown in Table 6. All these tools had a C_{pk} above 1.33, meaning that they are classified as high performers. Their values ranged from 1.34 to 1.96, which is an indicator that the tools performed with a very high degree of focus. These tools produce tightening results that not only meet ±10%, but do so by a considerable margin. In general, the best performers also have high C_p values, often well above 1.5, meaning that the inherent variation (3σ-deviation) of their output is small relative to the set tolerances. For example, tool 40, with a nominal torque of 14.9Nm, achieved a C_p value of 2.26 and a C_{pk} value of 1.91. This reflects the very low variance and average of the torque output. For most of these top tools, C_p and C_{pk} are nearly equal, suggesting that the processes are well centered with minimal deviation from the target. This suggests that calibration is quite effective for these tools: Eight of the top ten results correspond to "post-calibration" data, while the two top performers had no calibration during the data collection. The high performance of these tools indicates that tool variation is well controlled and their averages are adjusted close to nominal values. As a result, they operate with a comfortable buffer from the specification limits, reducing the risk of erroneous readings.

Table 6. Comparison of the ten highest-performing tools.

TOOL ID	Nominal (Nm)	Cp	C_{pk}
31	**14.9**	**2.328**	**1.966**
40	**14.9**	**2.261**	**1.911**
47	**13.5**	**1.904**	**1.774**
32	**13.5**	**1.991**	**1.719**
11	**5.0**	**1.702**	**1.612**
36	**13.5**	**1.665**	**1.607**
24	**3.0**	**1.556**	**1.434**
48	**16.5**	**1.821**	**1.422**
49	**14.9**	**2.021**	**1.363**
21	**5.0**	**1.564**	**1.343**

6.3.2 Low-capability tools (C_{pk} <1.0)

In contrast, Tabel 7 presents the ten worst-performing tools in terms of performance. Each of these tools has a C_{pk} well below 1.0 and is far from the acceptable threshold. Such low C_{pk} indices clearly indicate that these processes are not capable of producing within the 10% tolerance. These tools have either excessive variation, significant mean bias, or, as in most cases, both. For example, tool 20 produces a torque with a C_{pk} value of 0.21, the lowest of all the tools. Its mean torque was very close to the upper specification limit and had significant standard deviation, leading to frequent tolerance overruns. The tool produced torque values above the target value. Although the tool was calibrated during the pilot, it drifted towards the upper specification limit again. Based on these observations, It was concluded that this tool is defective and requires either repair or removal from production. Another example is tool 34, which has a nominal torque of 24.0 Nm. This tool had a C_{pk} value of 0.38 despite a relatively high potential C_p of 1.76. In this case, the process dispersion is acceptable, but the average torque has shifted up, too close to the upper limit, meaning that the tool is accurate but inaccurately calibrated. This was discussed in the previous section. This is a classic case of a tool that needs calibration. A high C_p indicates that the variation could be kept within tolerance if it were centered, but a low C_{pk} reveals that the process is out of balance.

Table 7. Comparison of ten tools with the lowest C_{pk} values.

TOOL ID	Nominal (Nm)	Cp	C_{pk}
20	**5.0**	**1.299**	**0.208**
44	**5.2**	**0.936**	**0.259**
2	**3.0**	**0.495**	**0.275**
37	**4.5**	**0.597**	**0.357**
9	**5.0**	**0.770**	**0.359**
1	**3.0**	**0.712**	**0.362**
34	**24.0**	**1.759**	**0.382**

50	**15.0**	**1.250**	**0.422**
29	**3.0**	**0.814**	**0.423**
3	**1.1**	**0.767**	**0.447**

6.3.3 Summary and implications

In summary, the C_p and C_{pk} analysis in the current 10% tolerance scenario reveals a significant performance gap in the tool set. While a subset of tools demonstrates high capability ($C_{pk} \geq 1.33$) and a further group meets the minimum capability level ($C_{pk} > 1.00$), a segment of tools does not achieve the required $C_{pk} < 1.00$ threshold. The tools that fall into this category have a higher risk of generating measurements outside of the specification limits and therefore require corrective actions.

It is important to note that many tools with low C_{pk} values still exhibit $C_p \geq 1$, which means that the problem is not excessive random variation but rather mis centering. The existence of tools with high C_p but low C_{pk} emphasizes that centering is as critical as variation. Even a relatively stable tool can create quality problems if it produces a systematically too high or too low. For these tools, calibration is expected to significantly enhance C_{pk} by shifting the mean closer to the target value. In contrast, tools with $C_{pk} < 1.00$ show genuinely excessive variation, typically caused by mechanical instability or wear. These units might require more than a simple calibration, such as maintenance, repair or removal from service.

The results from this analysis support the practical importance of regular torque verification and timely calibration. The strong association between recent calibration and high C_{pk} suggests that calibration practices are effective in reducing variation and correcting bias when applied. By implementing a structured verification schedule helps prevent drift and maintain processes closer to nominal value. Furthermore, utilizing SPC and capability metrics enables early detection of tool degradation, allowing corrective actions before deviations lead to out-of-tolerance outcomes.

6.4 Optimization of tolerance limits

This chapter examines what is an appropriate tolerance setting for the developed torque tool verification system for the target company's practical conditions. During the pilot, a ±10% tolerance limit was used as a default acceptance criterion in order to perform a consistent comparison between different tools and to collect sufficient baseline data on the natural behavior of the tool stock in production. The aim of this chapter is to determine, using the collected data and the theoretical basis of the presented literature review, what is the most appropriate tolerance limit that provides a realistic and statistically justified balance between the sensitivity of detecting real tool problems and the feasibility of routine production verification.

As mentioned in the literature review, it is important to distinguish between laboratory accuracy requirements and the reality of production use when interpreting tolerance limits. Although standards define the accuracy requirements for torque tools under certain standardized conditions, additional sources of variation that affect verification results in production must be taken into account. The tolerance limits used in a production verification system must be wide enough to cover the total variation expected under real-world conditions, but at the same time tight enough to reliably detect significant deviations and prevent continued use of tools that require adjustment or maintenance.

6.4.1 Analysis method

The validation is based on a comparison between the total variation of each tool and the ±10% tolerance limit. For each individual tool, all verification measurements recorded during the pilot period were aggregated to determine the standard deviation and the mean. The key metric used for comparison between tool variations and tool tolerances was the three-sigma range (3σ), representing three times the standard deviation of the tool’s measurement results. In quality engineering, the $\pm 3\sigma$ range is widely used and accepted as the “natural spread” of a stable process. The $\pm 3\sigma$ range provides a practical reference level for relating process dispersion to specification limits.

The use of a 3σ as the comparison metric aligns with established SPC standards. These limits are widely used to balance sensitivity to meaningful changes in the process and resistance to excessive false alarms. For a normally distributed process that is also statistically stable, it can be expected that 99.73% of the process data points will dall within the range of $\pm 3\sigma$ of the process mean. Evaluating the measured spread of a torque tool against the selected tolerance window provides a practical capability-oriented interpretation of whether the tool variation is reasonably small relative to the allowed tolerance.

To facilitate an equitable comparison between tools across varying torque ranges, the 3σ values were normalized relative to the nominal torque, thereby expressing the dispersion as a percentage. The normalized 3σ spread of each tool was compared against the ±10% tolerance limit. In the comparison, we classified the performance of each torque tool into four variation ranges according to how much selected tolerance window is consumed by the tool's 3σ spread:

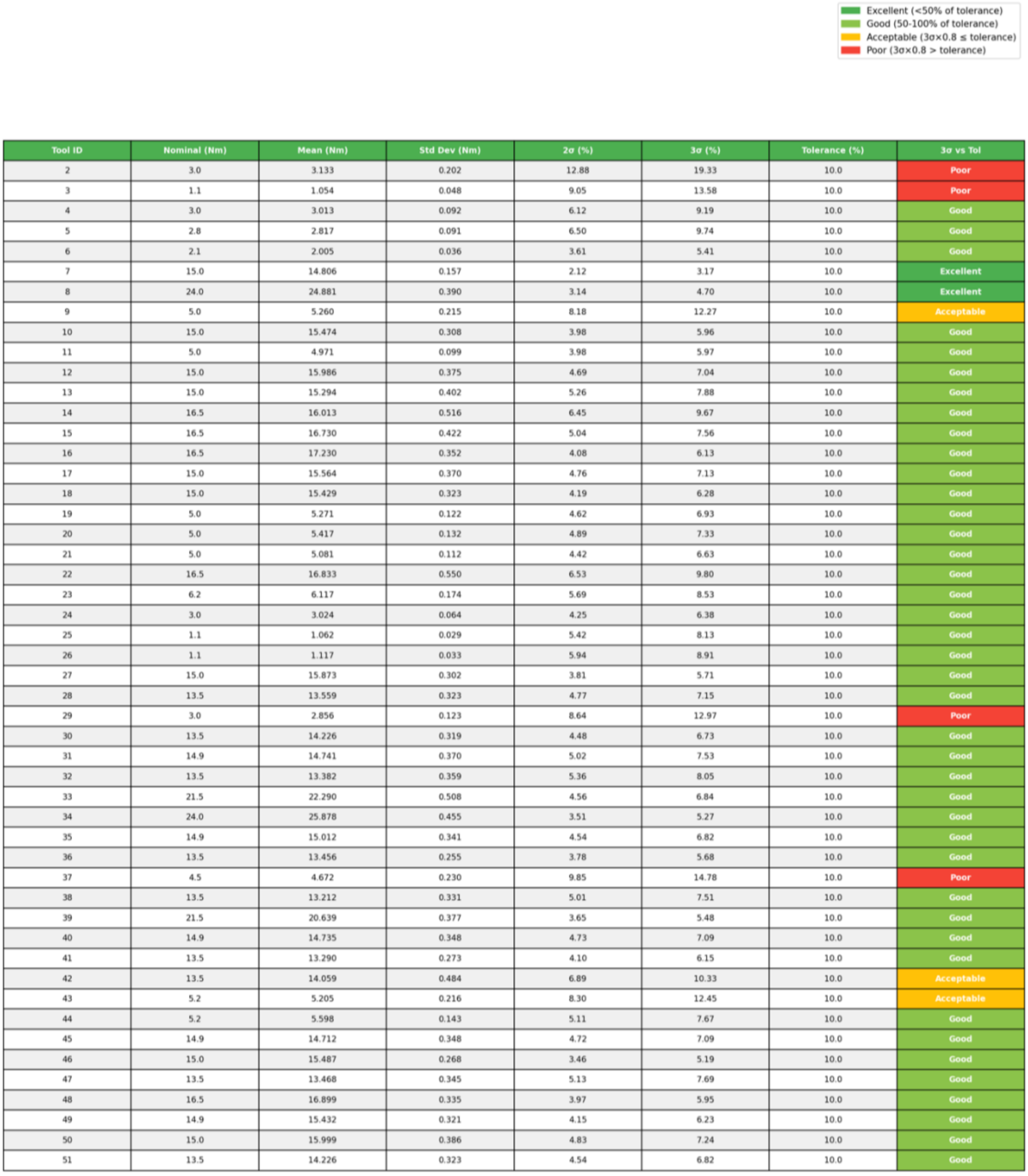
- Excellent: $3\sigma<50\%$ from the set tolerance
- Good: 3σ on 50-100% from the set tolerance
- Acceptable:$3\sigma < 125\%$ from the set tolerance ($3\sigma \times 0.8 <$ tolerance)
- Poor: $3\sigma > 125\%$ from the set tolerance

These ranges help compare and quantify the amount of variation of each tool against the requirements. In the Excellent category, the 3σ spread is less than 50% of the tolerance limit. This reflects very low variation relative to the allowed limit and a clear capability margin. In the Good category, the 3σ spread ranges between 50% and 100% of the tolerance, meaning that the tool variation remains within the tolerance window and the tool is generally capable under stable and well-centered conditions. The Acceptable category, where the 3σ spread is 100-125% of the tolerance, represents a marginal operating zone where the tool's inherent process variation slightly exceeds the tolerance limits. According to Industry guideline VDI/VDE 2645-2, it is recommended to set the tightening tolerance at approximately 80% of the tools $\pm 3\sigma$ capability, which conversely means the tool's 3σ spread is 125% of the tolerance. For this reason, tools in this category may still be considered usable in production under controlled conditions. In the Poor category, the 3σ spread is greater than 125% of the tolerance. This indicates that the tool's inherent variation is significantly greater than the tolerance limit, indicating that the process is statistically incapable.

This categorization provides a practical framework for screening tools based on their observed variation relative to the selected tolerance window. It is used in this analysis to support the determination of an appropriate tolerance limit for the developed system.

6.4.2 Results and interpretation

The analysis demonstrates that a ±10% tolerance window is technically compatible with the variability observed for the pool population. As shown in Figure 8, the majority of tools show a calculated 3σ spread that accounts for a relatively small percentage of the available tolerance budget, suggesting that ±10% is not overly restrictive under realistic operating conditions. Tools that are well performing typically exhibit a small 3σ percentage relative to nominal torque, implying that their natural variation remains comfortably within the tolerance window. On the contrary, tools that consume a large share of the tolerance budget and possibly exceed it, are easily identified as statistical outliers. The outliers can then be targeted for corrective actions such as recalibration, maintenance or removal from service.



🟩 Excellent (<50% of tolerance)
🟩 Good (50-100% of tolerance)
🟨 Acceptable (3σ×0.8 ≤ tolerance)
🟥 Poor (3σ×0.8 > tolerance)

Tool ID	Nominal (Nm)	Mean (Nm)	Std Dev (Nm)	2σ (%)	3σ (%)	Tolerance (%)	3σ vs Tol
2	3.0	3.133	0.202	12.88	19.33	10.0	Poor
3	1.1	1.054	0.048	9.05	13.58	10.0	Poor
4	3.0	3.013	0.092	6.12	9.19	10.0	Good
5	2.8	2.817	0.091	6.50	9.74	10.0	Good
6	2.1	2.005	0.036	3.61	5.41	10.0	Good
7	15.0	14.806	0.157	2.12	3.17	10.0	Excellent
8	24.0	24.881	0.390	3.14	4.70	10.0	Excellent
9	5.0	5.260	0.215	8.18	12.27	10.0	Acceptable
10	15.0	15.474	0.308	3.98	5.96	10.0	Good
11	5.0	4.971	0.099	3.98	5.97	10.0	Good
12	15.0	15.986	0.375	4.69	7.04	10.0	Good
13	15.0	15.294	0.402	5.26	7.88	10.0	Good
14	16.5	16.013	0.516	6.45	9.67	10.0	Good
15	16.5	16.730	0.422	5.04	7.56	10.0	Good
16	16.5	17.230	0.352	4.08	6.13	10.0	Good
17	15.0	15.564	0.370	4.76	7.13	10.0	Good
18	15.0	15.429	0.323	4.19	6.28	10.0	Good
19	5.0	5.271	0.122	4.62	6.93	10.0	Good
20	5.0	5.417	0.132	4.89	7.33	10.0	Good
21	5.0	5.081	0.112	4.42	6.63	10.0	Good
22	16.5	16.833	0.550	6.53	9.80	10.0	Good
23	6.2	6.117	0.174	5.69	8.53	10.0	Good
24	3.0	3.024	0.064	4.25	6.38	10.0	Good
25	1.1	1.062	0.029	5.42	8.13	10.0	Good
26	1.1	1.117	0.033	5.94	8.91	10.0	Good
27	15.0	15.873	0.302	3.81	5.71	10.0	Good
28	13.5	13.559	0.323	4.77	7.15	10.0	Good
29	3.0	2.856	0.123	8.64	12.97	10.0	Poor
30	13.5	14.226	0.319	4.48	6.73	10.0	Good
31	14.9	14.741	0.370	5.02	7.53	10.0	Good
32	13.5	13.382	0.359	5.36	8.05	10.0	Good
33	21.5	22.290	0.508	4.56	6.84	10.0	Good
34	24.0	25.878	0.455	3.51	5.27	10.0	Good
35	14.9	15.012	0.341	4.54	6.82	10.0	Good
36	13.5	13.456	0.255	3.78	5.68	10.0	Good
37	4.5	4.672	0.230	9.85	14.78	10.0	Poor
38	13.5	13.212	0.331	5.01	7.51	10.0	Good
39	21.5	20.639	0.377	3.65	5.48	10.0	Good
40	14.9	14.735	0.348	4.73	7.09	10.0	Good
41	13.5	13.290	0.273	4.10	6.15	10.0	Good
42	13.5	14.059	0.484	6.89	10.33	10.0	Acceptable
43	5.2	5.205	0.216	8.30	12.45	10.0	Acceptable
44	5.2	5.598	0.143	5.11	7.67	10.0	Good
45	14.9	14.712	0.348	4.72	7.09	10.0	Good
46	15.0	15.487	0.268	3.46	5.19	10.0	Good
47	13.5	13.468	0.345	5.13	7.69	10.0	Good
48	16.5	16.899	0.335	3.97	5.95	10.0	Good
49	14.9	15.432	0.321	4.15	6.23	10.0	Good
50	15.0	15.999	0.386	4.83	7.24	10.0	Good
51	13.5	14.226	0.323	4.54	6.82	10.0	Good

Figure 8. Torque tool standard deviation analysis.

This type of behaviour supports the practical role of ±10% as an acceptance criterion in a production verification context. A tolerance window that is too narrow relative to natural variation would result in a high rate of false negatives, where tools would be flagged due to inherent statistical dispersion rather than actual degradation. This would lead to excessive alarms and unnecessary workload. Conversely, if the tolerance were excessively wide, the system would fail to detect systematic shifts and degradation in time. The pilot data indicates that ±10% threshold provides an optimal balance, it is wide enough to accommodate normal production variation, while remaining sensitive enough to identify tools that deviate significantly from expected behaviour.

This interpretation is also supported by the consistency of findings across analyses. As shown in Table 7 and Figure 8, tool ID 29 and tool ID 37, which were identified in the SPC analysis as not performing at the required quality level, are also in the present tolerance variation analysis classified in the Poor category. This means that their observed variation significantly exceeds the defined limits, confirming that their performance is inadequate relative to quality requirements. This alignment between SPC- based instability detection and tolerance-based variation screening provides practical evidence that the selected ±10% tolerance is functional and sufficiently sensitive for the developed system.

6.4.3 Recommended optimal torque tolerance limit for Case company.

Based on the pilot data and the above comparison between observed variation and the tolerance window, this study recommends a tolerance limit of ±10% of nominal torque for the case company’s production verification process. This is based on three primary considerations: production feasibility, statistical justification and alignment with the theoretical baseline.

In terms of production feasibility, a ±10% tolerance serves as a practical threshold that supports routine verification without generating excessive false alarms due to normal variation. Empirical evidence from the pilot study shows that most tools operate with sufficient margin within this window, while tools with performance issues remain clearly distinguishable and can be prioritized for corrective actions.

From a statistical justification perspective, when we compare tool dispersion against the tolerance window, we can see that the tolerance is not systematically smaller than the observed natural variability of the tool population. In other words, the ±10% limit is broad enough to accommodate realistic dispersion in production while still enabling reliable detection of highly variable tools as well as mis centered tools when combined with the mean-level evaluation and capability indices reported earlier.

Third, the proposed ±10% tolerance aligns with the theoretical and practical framework, as presented in the literature review. While some of the tools might achieve tighter accuracy under controlled calibration conditions, the torque measurements are done in active production environment, and the measurements are affected by additional factors such as conditions and user-induced variation. This user-related contribution was discussed as operator accuracy in the literature review (Bickford, 2018). Therefore, the tolerance of torque checks performed in production must reflect the production accuracy rather than laboratory accuracy alone. For this reason, the ±10% operator-related accuracy level identified by Bickford is more suitable for the present application than laboratory-oriented limits such as ±4% and ±6% margins found in ISO-based calibration standards.

To conclude, the empirical variation observed during the pilot phase validates a ±10% tolerance limit as a functional and technically justified threshold for the developed verification system within the case company’s production environment. This tolerance limit serves as the foundational baseline for the alarm limit design in the next section and for the final recommended operating procedure of the system.

6.5 Setting alarm limits

We will move ahead to the most crucial stage of the torque tool management system, where the observations of the statistical process control limits are to be merged with the quality demands specification limits to form a proactive warning system. The implementation of the concept of statistical process control within the torque tool management system demands an explicit multitier limits definition. The current chapter describes the terms concerning specification limits and control limits and explains the rationale behind the proposed three-tier warning map: OK, WARN, NOK, based on these two categories of limits that aim to provide a predictive calibration to avoid the tool to move beyond specification limits before an error occurs.

6.5.1 Philosophy of the dual border system

As discussed in the literature review, it is necessary to distinguish between specification limits and control limits. According to Montgomery (2009), control limits should not be derived from specification limits, because there is no mathematical or statistical relationship between them. If control limits are set too tightly within specification limits, the process is over-adjusted, which actually increases process variation and reduces quality. If control limits are set too loosely relative to specification limits, process changes can be ignored, causing natural process variation to increase unnecessarily. Both errors lead to higher costs and lower quality. Therefore, control limits should reflect actual process performance.

6.5.2 Setting specification limits and control limit zones

Based on the tolerance optimization in Section 6.4, the specification limits for torque verification are set at ±10% of the nominal value. In addition to specification limits, the alarm logic uses monitoring zones from control limits. These zones are based on the statistical mean and standard deviation calculated from the measurement history, and the monitoring zones are defined at $\pm 1\sigma$, $\pm 2\sigma$ and $\pm 3\sigma$ around the process center line. These zones allow the application of statistical process control rules, such as variations of the Western Electric Rules or Nelson Rules, to detect subtle signs of a process drifting into an unstable state. By identifying these transitions early, the monitoring zone provides the basis for the WARN and NOK logic.

As an Example, for torque tool ID 41 (Nominal torque = 13.5 Nm) the specification limits are calculated as follows:

$$USL = 13.5\,Nm \times 1.10 = 14.85\,Nm \qquad (13)$$

$$LCL = 13.5\ Nm \times 0.90 = 12.15\ Nm \qquad (14)$$

In the analysis of torque wrench ID 41, the statistical parameters of the process were determined as follows:

Process average ($\bar{X}$): 13.290 Nm (center line)

Process standard deviation (σ): 0.273 Nm

Control limits are calculated as follows:

$$UCL = \bar{X} + 3\sigma = 13.290\ Nm + 3(0.273\ Nm) = 14.109\ Nm \qquad (15)$$

$$LCL = \bar{X} - 3\sigma = 13.290\ Nm - 3(0.273\ Nm) = 12.471\ Nm \qquad (16)$$

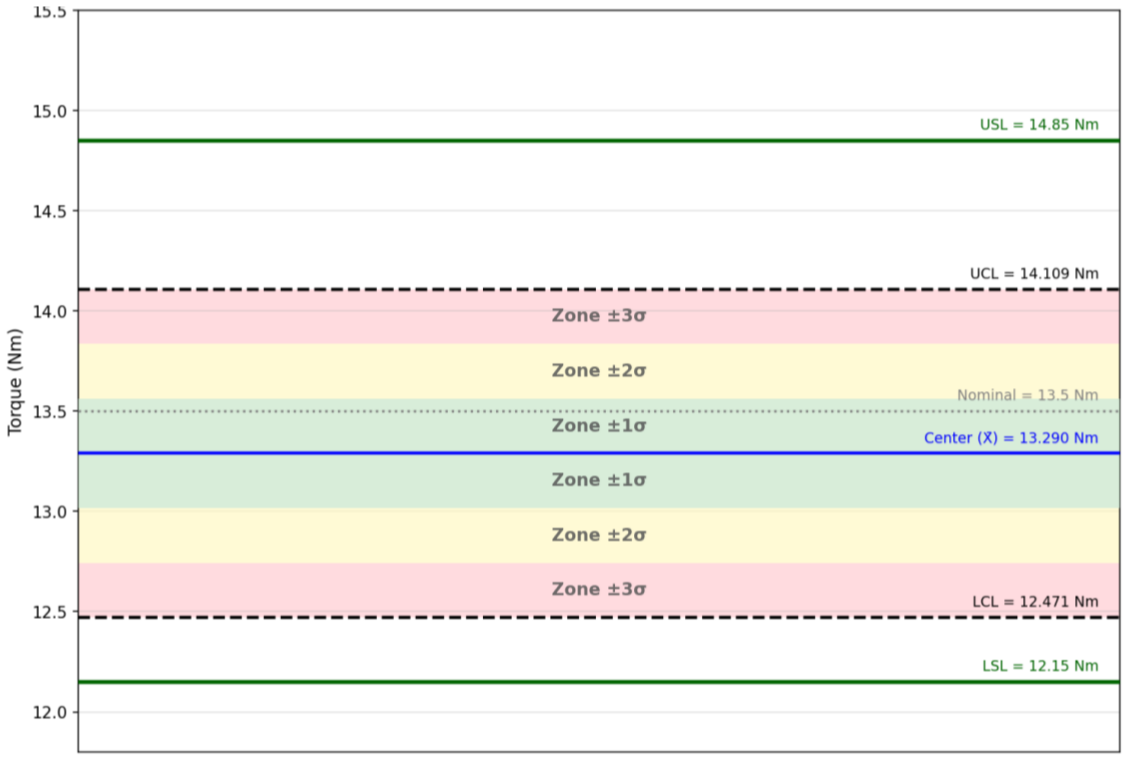


Figure 9. Control and specification limits for tool ID 41.

In the case of torque tool ID 41, the process is seen to be statistically in control, though slightly misaligned ($\bar{X}$ = 13.290 Nm and T_{nom} = 13.5Nm). As shown in Figure 9, the comparison between the control limits to specification limits reveals that the natural process range (12.471 Nm - 14.109 Nm) fully resides inside the specification limits (12.150Nm - 14.850Nm). However, the process center has drifted closer to the lower specification limit, indicating that even though the process is statistically in control, the tool is not performing on its target value. This result further identifies the importance of the need for the predictive monitoring. A tool may remain within specification while already showing drift or bias that can be corrected before the tool exceeds the specification limits.

6.5.3 Multi-sigma warning zones usage criteria

Torque tool performance degradation is most often a slow, systematic average shift in the process mean. To detect this degradation before a critical specification limit is breached, an intermediate-level statistical process control alarm system is implemented.

By basing process monitoring on the more sensitive 2σ and 3σ warning zones and identifying non-random patterns, such as. trends or accumulation, we move from reactive to proactive monitoring. These more subtle signals indicate a systematic change that requires action before the change approaches critical limits. For example, the detection of a continuous upward or downward trend indicates a slowly progressing problem in the torque tool, such as wear. When such a transition is detected already at the warning level, it provides time for predictive maintenance and optimal timing of calibration. As a result, NOK conditions and the resulting problems that would arise if the specification limits were exceeded can be avoided.

6.5.4 Development of multi-layer alarm matrix (OK, WARN, NOK)

The new torque tool management system utilizes three classification levels (OK, WARN, NOK) by combining the quality requirements suggested by specification limits and the statistical process control rules of control limits. The purpose is to separate immediate quality-critical situations from early signs of progress degradation.

6.5.4.1 Immediate out-of-specification and critical out-of-control alarm

A measurement event is classified as NOK when any of the following conditions is met:

1. Out of Specification Limits (OOS)
   A singel measurement exceeds the USL or falls below the LSL.
   Specification limit violation is a direct quality requirement violation. It requires immediate action.

2. A single point exceeds the 3σ Control Limit (OOC)
   This is the traditional rule of statistical process control, Rule 1, which refers to a sudden and statistically unlikely deviation caused by a special cause.

When a NOK measurement is detected, the system generates an NOK email notification and logs the event in the NOK monitoring records. A NOK measurement requires immediate corrective actions.

6.5.5 Setting the predictive warning level alarm

The warning level is designed to detect slow-moving signs of process degradation, which are essential for predictive maintenance of torque tools. WARN conditions require immediate attention, but not immediate action, unlike NOK conditions. A measurement is classified as WARN when any of the following conditions is met:

1. Two consecutive points in the 2σ - 3σ zone

Two consecutive measurement points are located between 2σ and 3σ on the same side of the centerline. This rule strongly indicates a non-random process drift or increase in variability.

2. Continuous process bias

Seven consecutive points on the same side of the centerline. This rule is very effective in identifying slow process mean drift, which is a typical torque wrench wear condition. This is a non-random pattern that provides early warning of the need to calibrate the tool before it reaches critical limits.

3. Systematic trend

A trend of six consecutive points increasing or decreasing is statistically unlikely in a controlled process and is a typical sign of systematic wear or loss of control. This alarm provides critical early warning of the need for calibration.

When a WARN measurement is detected, the system sends a WARN email notification. WARN status is designed to trigger closer monitoring and planning of corrective actions, as opposing to removing the tool immediately.

6.5.6 Operational protocols and triggering of predictive calibration

This hybrid alert system supports torque tool maintenance by moving from time-based maintenance to an incident-based, data-driven predictive model. Using statistical process control rules to monitor torque tools allows the system to distinguish random variation caused by common causes from true variation caused by special causes, which is an important step towards predictive rather than corrective maintenance.

WARN signals are considered major triggers in the system's predictive calibration process. These signals determine when there is a slow drift or variation in the tool's average. Since all of these are identified by being well within specification limits, calibration is done proactively before the tool fails into a NOK status. In contrast to a NOK scenario, which requires immediate action to ensure quality integrity, WARN signals give warnings and alert calibration team members of an early deterioration status so that an examination or preparation of spare parts begins by the calibration unit before the next calibration process.

6.6 Determining the required number of measurements per measurement series

In this section, we determine how many repeated measurements are required in each measurement series to accurately estimate the true output of a torque tool. The objective is to define an optimal sample size for routine verification that is statistically justified and operationally feasible in production. The aim is to ensure that the sample mean stays within ±2% of the tool's nominal torque relative to the true mean, at a 95% confidence level. Using the statistical confidence interval principle, we derive a general formula for the required sample size from confidence interval theory and then apply it to the torque tool dataset. Finally, we interpret the results and justify the recommended sample size for routine use.

6.6.1 Sample size formula based on confidence intervals

To estimate the population mean with a specified degree of accuracy and margin of error, the standard confidence interval approach for the mean is used. Assuming a normal distribution for the sampling distribution of the mean, the margin of error E, is defined as:

$$E = Z\frac{\sigma}{\sqrt{n}} \qquad (17)$$

where

- $\bar{X}$ is the sample mean
- σ is the standard deviation
- n is the amsple size, and
- Z is the critical value of the standard normal distribution for the selected confidence level.

For the 95% confidence level, the corresponding value for Z is 1.96. Solving the margin-of-error equation for n gives the required sample size formula:

$$n = (\frac{Z\sigma}{E})^2 \qquad (18)$$

In this study, the acceptable estimation error is set to ±2% of nominal torque ($E = 0.02 \times T_{nom}$). The required sample size for each tool is therefore calculated as:

$$n = (\frac{Z\sigma}{0.02 \times T_{nom}})^2 \qquad (19)$$

The standard deviation σ is estimated from the historical measurement data specific to each tool. As the derived formula establishes a minimum threshold for statistical validity, the calculated sample size is rounded upward to the next integer. The use of the Z -value is justified by the fixed confidence level and the standard deviation is estimated from an extensive pilot dataset. This is a standard and practical approach in sample-size planning for repeated measurements.

6.6.2 Analysis of required sample sizes for 50 tools

Using the formula above, the required sample size n was calculated for 50 torque tools. For each tool, the calculation incorporated the standard deviation (σ) of the tool and the allowable estimation error, defined as ±2% of the nominal torque, at a 95% confidence level ($Z = 1.96$). The results showed a considerable variation in n between tools, depending on the consistency of each tool in relation to its nominal torque. A summary of the distribution of required sample sizes for the 50 tools analyzed resulted in a median of 5 measurements. This median indicates that the tools would achieve a ±2% accuracy with approximately 5 measurements. In practice, a typical torque tool in our sample would require approximately 5 readings for the desired confidence and accuracy. As shown in Figure 10, the distribution of the calculated sample size requirements demonstrates that the majority of the tools required $n \leq 10$ measurements to meet the selected criterion. For a small minority of tools, the required n was more than 10. The required sample size fell between 11 and 17, reflecting a higher level of observed variation in relation to the specified estimation error. However, this applied only to six tools and the overrun was relatively moderate.

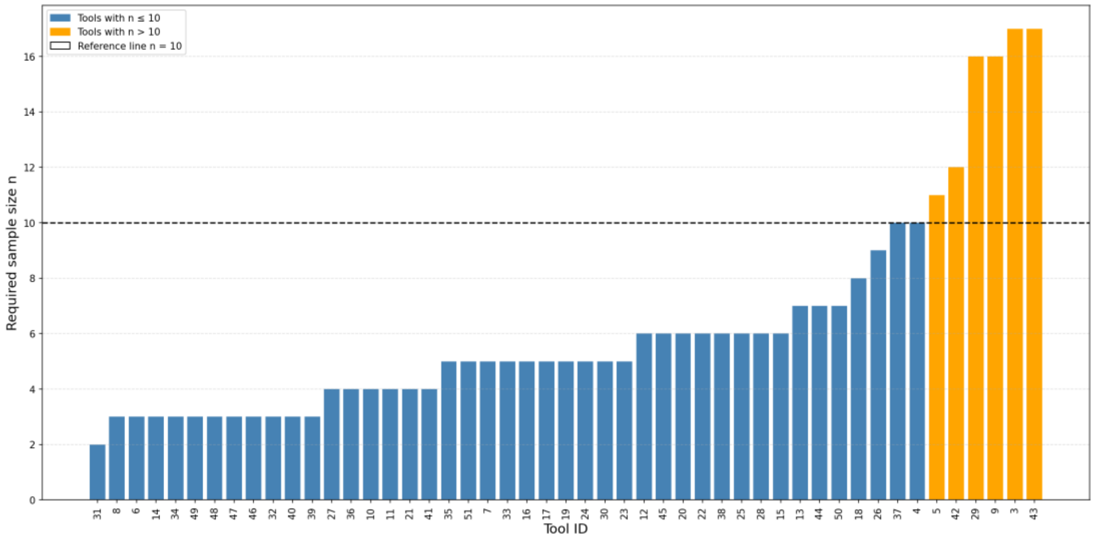


Figure 10. Calculated sample size requirements per tool ID relative to the n = 10 threshold.

One unit emerged as a distinct outlier within the population. The tool ID 2 required a sample size of $n = 44$, indicating exceptionally high variation relative to the selected error criterion. This case was excluded from Figure 10. This finding is consistent with the earlier capability analysis, where the same tool demonstrated poor performance metrics (C_{pk} = 0.495, C_p = 0.275), indicating that the tool is not capable of performing at the required quality level under current conditions. This case is examined further in the torque tool performance analysis in Section 6.8.

6.6.3 Recommended sample size and justification.

Based on the analysis, the recommended number of measurements per measurement series for torque tool verification is n = 10. Although the median required sample size, calculated from the data on the 50 tools, is 5 measurements, the recommendation is intentionally set higher to ensure robust performance across the broader tool population under practical production conditions. The results show that a majority of tools of the tools required no more than 10 measurements to meet the selected criterion of a 95% confidence level with a ±2% margin of error. For the six tools that required more than 10 measurements, the required sample size was only a moderately above the proposed standard value.

The choice of n = 10 is justifiable from both statistical an operational perspective. Statistically, this sample size provides a clear safety margin above the median requirement, ensuring that the target confidence and precision is maintained across the majority of tools. From an operational perspective, using a fixed series size for the measurements is easy to implement, fosters consistent operator practices, and minimizes procedure complexity compared with a tool-specific sample size rule. The process also aligns with the subgroup size used in the SPC analysis (Section 6.2), which improves consistency between the monitoring methodology and the practical measurement procedure. Furthermore, increasing the sample size beyond n = 10 would provide diminishing returns in estimation accuracy. The standard error decreases proportionally to $\frac{1}{\sqrt{n}}$, therefor increasing the sample size beyond 10 will increase the workload of the production process more than it will improve the accuracy.

Overall, a measurement sample size of 10 readings per tool provides a practical and statistically justified balance between confidence, accuracy, and production feasibility. This also aligns with the literature review, which identifies ten repeated measurements as an effective sample size for assessing tool repeatability. It is therefore recommended as the standard sample size for the developed torque tool verification system.

6.7 Determining the measurement interval

The purpose of this section is to determine a reasonable verification interval for the developed torque tool management system. In this study, the interval is not determined solely on the basis of calibration practices or assumed long-term drift, but by examining the progression of tool behavior from normal conditions towards warning limits and non-conforming conditions. In statistical process control, the goal of monitoring is to detect significant process changes at a sufficient frequency so that corrective action can be taken before unacceptable quality deviations arise. Control chart methods are specifically intended to distinguish normal process variation from signals that indicate variation due to specific causes or process degradation.

Since the pilot phase was intentionally intensive and multiple torque checks were performed on the same tool during the same calendar day, the time-to-event analysis was performed using a session-based time-series. In this case, one torque check measurement series was treated as one monitoring phase, regardless of whether two or more sessions occurred during the same day. This approach is more appropriate than calendar time, as the purpose of the analysis is to determine after how many torque verifications the first warning or rejection condition typically occurs. The resulting time-to-event analysis thus reflects the number of verifications from baseline to the first statistically significant signal.

6.7.1 Analytical basis for interval determination

A suitable measurement interval must meet two requirements. First, it must be short enough to detect tool degradation before it reaches a critical state for quality. Second, it must be operationally feasible and not impose an unnecessary inspection burden. In the current system, this balance is assessed by comparing the occurrence of two different types of events: WARN, which indicates early signs of performance degradation, and NOK, which indicates exceeding specification limits or critical loss of control.

The logic of this approach follows the principles of statistical process control. A point or a non-random series outside the expected control range, such as a continuous run or trend, indicates that the process may no longer be operating as a stable system. At the same time, the acceptability of the result must still be assessed against externally defined tolerance limits. In this study, the determination of the inspection interval is based on the practical relationship between early statistical warning signals and actual NOK events.

6.7.2 Time-to-first-WARN and time-to-first-NOK analysis

The measurement sessions for each torque tool were identified from the timestamp structure of the data, and each session consisted of one verification set of 10 measurements. In the time-interval analysis, one session was treated as one monitoring phase. The first verification session classified as OK was defined as the baseline for each tool, after which the number of sessions to the first WARN and first NOK event/session was calculated.

WARN and NOK classifications were based on the system logic defined earlier in this study. A WARN signal represented an early statistical reference group for a process change, while a NOK event represented either a direct specification exceedance or a critical loss of control signal. In addition, right censoring was used for those tools that did not reach a WARN or NOK event during the pilot follow-up period. Since censored observations are common in time-to-event analysis, Kaplan-Meier estimation was used in addition to standard descriptive statistics to obtain censored median times to events. Kaplan-Meier estimation is a standard nonparametric method for right-censored data, and is particularly useful when some units do not experience an event within the follow-up period.

The analysis included 50 tools. Of these, 42 tools reached a WARN event during the monitoring period, while 8 tools remained WARN-censored. For the NOK event, 24 tools reached the event and 26 tools were censored. These results in themselves indicate an important difference between the two event types, with early warning signals being significantly more common than actual NOK events.

For those tools that detected a WARN event, the average time to first warning was 11.40 sessions and the median was 9.50 sessions. The 25th and 75th percentiles were 3.00 and 14.00 sessions, while the 10th and 90th percentiles were 1.10 and 23.80 sessions. These results indicate that the first warning signal often occurred relatively early, although there was still considerable dispersion between tools.

Table 8. Statistical summary of time-to-event data and Kaplan-Meier estimates.

Metric	**WARN**	**NOK**
Tools analyzed	50	50
Observed	42	24
Censored	8	26
Mean (observed only)	11.40	14.80
Median (observed only)	9.50	7
10th percentile (observed only)	1.10	1.00
25th percentile (observed only)	3.00	2.00
75th percentile (observed only)	14.00	19.75
90th percentile (observed only)	23.80	45.10
Kaplan -Meier median (Censored)	11	46

For those tools that did experience NOK events, the average time to first rejection was 14.80 sessions, while the median observed was 7.00 sessions. The 25th and 75th percentiles were 2.00 and 19.75 sessions, and the 10th and 90th percentiles were 1.00 and 45.10 sessions. These values indicate that among the tools that eventually reached NOK status, some reached that point at an early stage, while others remained acceptable for considerably longer.

The Kaplan-Meier results provide a more informative interpretation because they also take into account tools that never reached an event during the monitoring period. The Kaplan-Meier median for WARN events was 11 sessions, while the Kaplan-Meier for NOK events was 46 sessions. The difference between these two values is very significant. It shows that at the population level, the first warning signal typically appears much earlier than the first NOK condition. In other words, the warning logic seems to be working as intended and identifies deterioration well before most tools reach a critical condition. This is precisely the role that a predictive warning system plays in predictive calibration and process monitoring.

The apparent difference between the median time to observed NOK events (7 sessions) and the Kaplan-Meier median (46 sessions) must be interpreted with caution. The observed median is calculated only for those tools that entered the NOK state during the study period, whereas the Kaplan-Meier estimate also includes tools that remained event-free throughout the follow-up period. Thus, the summary of observed events reflects the behavior of the failure-prone subgroup, whereas the Kaplan-Meier estimate better reflects the entire tool population. This distinction is important and justifies the use of lifetime analysis methods in this study.

Overall, the results showed that the developed monitoring system creates an early warning margin between the first statistical indication of deterioration and the first unacceptable tool condition. This margin is critical for the selection of the inspection interval. The verification interval should be short enough to ensure that most tools are inspected after the occurrence of the warning condition, but before the probability of a NOK event increases significantly.

6.7.3 Implications for interval selection

The results support two key observations. First, the fixed calendar time assumption alone does not adequately describe the performance degradation of torque tools, as tool behavior is non-uniform and some tools degrade faster than others. Second, the warning logic provides useful predictive information, since WARN events occur significantly earlier than NOK events after censoring.

In practical terms, the median time to first warning of approximately 11 sessions indicates that early process changes typically occur after a reasonable number of verifications. In contrast, the significantly longer Kaplan-Meier median time to first rejection (46 sessions) indicates that for most tools, the actual loss of acceptability occurs significantly later. This suggests that the inspection interval should be set so that tools are not inspected until close to the likely NOK moment, but sufficiently in advance to allow for intervention based on the warning.

At the same time, the determination of the inspection interval should not be based solely on the most stable tools. A significant proportion of tools raised warning signals relatively early, and the lowest percentiles show that some tools reached WARN or NOK status after just a few sessions. Therefore, the inspection interval recommendation must be conservative enough to take into account the most vulnerable tools but still remain practical for production use.

The decision on the inspection interval must also support the broader goal of condition-based calibration. Sufficiently frequent verification is needed not only to detect defects, but also to accumulate a sufficiently frequent series of observations to identify persistent variations, trends, and gradual drift patterns. Such pattern-based detection is one of the main advantages of SPC-based monitoring, and it would be impaired if the time between verification cycles were too long.

6.7.4 Recommended measurement interval

The recommended operational verification interval for the system developed based on the results of the session-based lifetime analysis is one week. This recommendation is not based on the assumption that all tools would necessarily reach a warning state within a week, nor on a direct computational conversion of session counts into calendar days. Instead, it is based on a practical interpretation of the session-based results in conjunction with the target company's production environment. The analysis shows that warning signals typically appear well before most NOK events, which means that the system is capable of supporting proactive actions. The weekly interval transforms this early warning into a practical monitoring routine that is frequent enough to detect degradation in time but still remains feasible in a manufacturing environment.

a longer interval would likely reduce the workload, but it would also reduce the responsiveness of the system and reduce the amount of data available for trend-based analysis. Since one of the main objectives of this work is to move from time-based calibration to data-based calibration, the verification interval must support the accumulation of a sufficiently informative performance history for each tool. Weekly verification provides a balance where it is frequent enough to maintain visibility into the tool's behavior over time, but not so frequent that it becomes impractical for routine use.

In summary, session-based time-to-event analysis shows that the first WARN event typically occurs significantly earlier than the first NOK event, and this warning margin can be used to support proactive calibration decisions. Based on this, a weekly verification interval is recommended for the torque tool management system developed. This interval is a practical compromise between early detection capability, quality assurance, and the long-term goal of data-driven calibration interval management.

6.8 Torque tool performance analysis

This section examines a performance comparison based on torque tools from various brands and models over 50 units during the pilot daily measurement schedule. However, the performance comparison is based on the measurement consistency, accuracy, and stability for the torque tools. The performance analysis revealed that that the tool population it is quite efficient and stable, as only about 4.8% of the measured values exceeded the limit tolerance limit of ±10%. Generally, all the tools measured torques close to their nominal's settings, indicating a satisfactory level of overall performance, but differences lie in their brand and models. Some tool groups, such as the AC and AA series, function exceptionally and show very minimal instances beyond the tolerance value, whereas the rest, such as brand X and Y, showed clearly weaker performance.

6.8.1 Evaluation criteria and comparative analysis basis

All torque tools were categorized according to model such as X/1 and AA/3, for comparison purposes. Some key torque-related metrics were determined for each torque tool: average/median torque, standard deviation, min/nominal torques, max torques, and the percentage NOK%, or measurements exceeding a tolerance of ±10%. The deviation or drift of each torque tool from nominal torques was determined by subtracting the nominal torques from the average torques measured and expressed as a percentage of nominal torques. Furthermore, torque drift, or trends over time, were investigated for each torque tool by comparing early measurements with late measurements. Torque tools showing an increasing/decreasing torques trend were identified as potentially out of calibration.

This data driven approach follows best practices in torque tool quality control: regular in process checks can spot potential tooling issues before they occur due to wear or drift. During production verification, any tool exceeding ±10% of the nominal torque is effectively a failed result that requires adjustment or calibration. The proactive monitoring discussed here aligns with the recommendations of the industry to use torque testers for alerting users about drift or inaccuracies and determining the tools that may need immediate attention between official calibration intervals. Below is an analysis comparing the performance of each make/model category, describing both high performing tools as well as those tools, that do not meet desirable stability and accuracy.

6.8.2 High-performing tool groups

Most tool groups showed strong overall performance. These groups were characterized by low variability, good centering, and stable behavior on control charts, with little or no out-of-control signals.

The Z/1 group (inline electric screwdriver) represented the strongest single performance in the data. The only tool in this category showed very high capability, with both C_p and C_{pk} clearly at an acceptable level and close to each other, indicating both low variability and good centering. As shown in Figure 11, the corresponding control chart showed a very stable process with no significant signs of instability. Based on the available data, this tool can be considered an example of a well-performing torque tool in a control system.

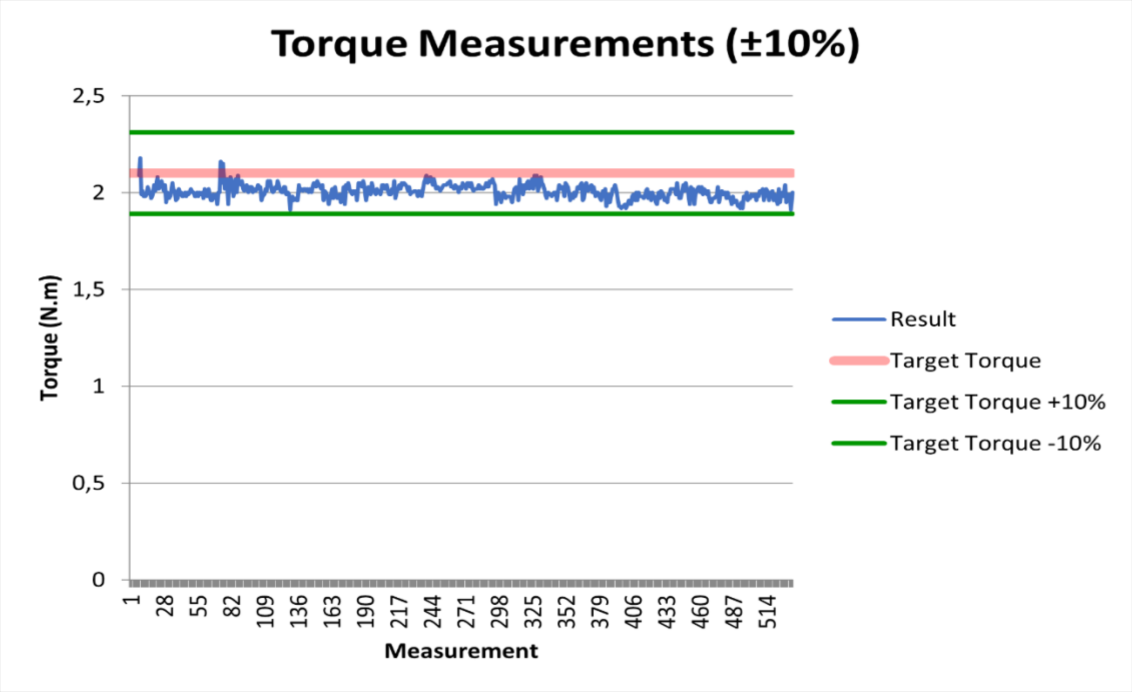


Figure 11. Tool ID 6 (Z/1) torque measurement control chart.

The AA/1 and AA/2 groups (mechanical torque wrench) also performed strongly overall. In these groups, natural process variation was small relative to the ±10% tolerance range, and several tools performed with a comfortable margin of capability. As shown in Figure 12, the AA/2 tools in particular showed high C_p and C_{pk} values and stable control chart behavior, suggesting that these tools were both repeatable and well centered. In cases where there was a small mean shift, the main limitation appeared to be centering rather than instability. These tools can thus be considered inherently capable, and any lower C_{pk} values were mainly due to calibration-related bias rather than excessive dispersion.

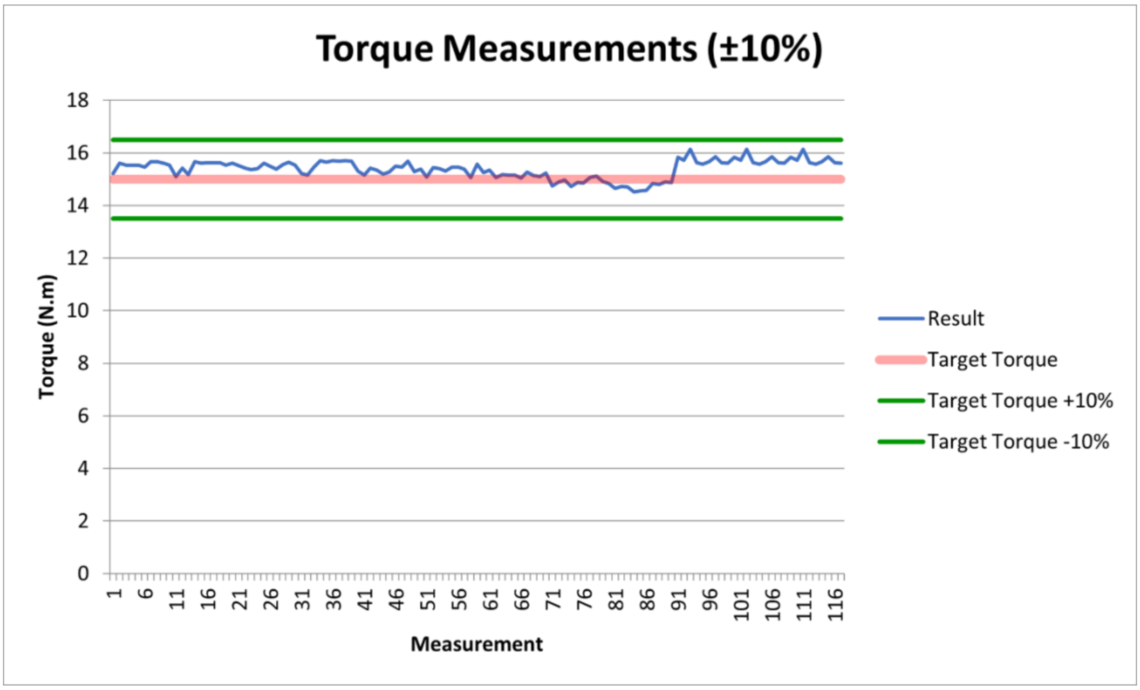


Figure 12. Tool ID 7 (AA/1) torque measurement control chart.

The strongest AA/3 tools (mechanical torque wrench) also performed well. Although the AA/3 group as a whole was inconsistent, some individual tools in this category showed good repeatability, acceptable capability, and very few control chart violations. These higher-performing tools indicate that the model family itself is not inherently problematic. The variation observed within the group suggests that individual tool quality plays a significant role.

The C/1 group (battery powered pistol-grip torque tool) also performed well in terms of repeatability (Figure 13). These tools had low process variation and relatively good C_p values, indicating that their natural variation in torque output was well within the allowable tolerance. However, in some cases a small but systematically negative deviation reduced the C_{pk} value. The main limitation of this group was therefore not instability or excessive variation, but centering accuracy. This suggests that many C/1 tools could perform well after recalibration.

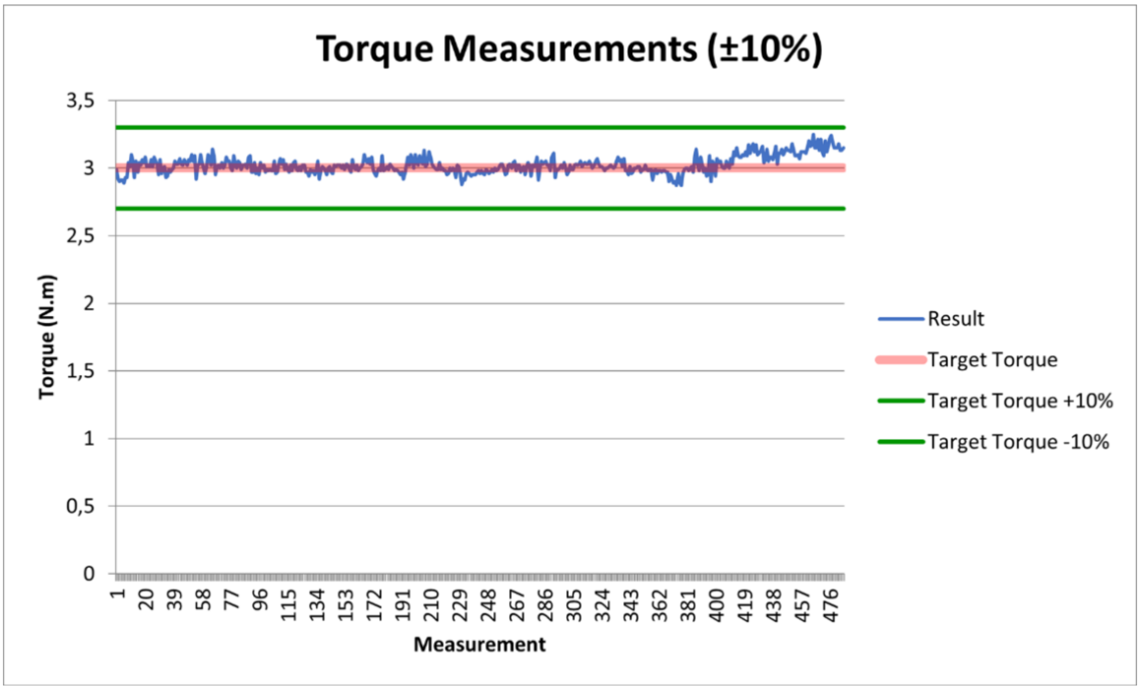


Figure 13. Tool ID 24 (C/1) torque measurement control chart.

Overall, the higher-performing groups had a common feature. Their process variability was low, their control charts were generally stable, and any remaining weaknesses were more often related to centralization than to uncontrolled dispersion. This is a favorable result, as centralization problems are typically easier to correct than inherently unstable tool behavior.

6.8.3 Problematic tool groups and dominant failure modes

The weakest tightening performance was observed in groups X/1 and X/2 (inline electric screwdrivers). These tools showed low capability and repeated problems on the control charts, indicating both instability and insufficient process performance in relation to tolerance limits. The X/1 group was clearly the most problematic. As shown in Figure 14, the tool in this category showed poor C_p and C_{pk} values, indicating that both the process dispersion and the centering were unacceptable. The control chart showed that the tool behavior indicated repeated instability and a continuous shift towards overtightening. It is important to note that this problem remained evident even after calibration, indicating that recalibration alone was not enough to restore performance to an acceptable level. In practice, such a result indicates that the tool should be thoroughly serviced or removed from production use.

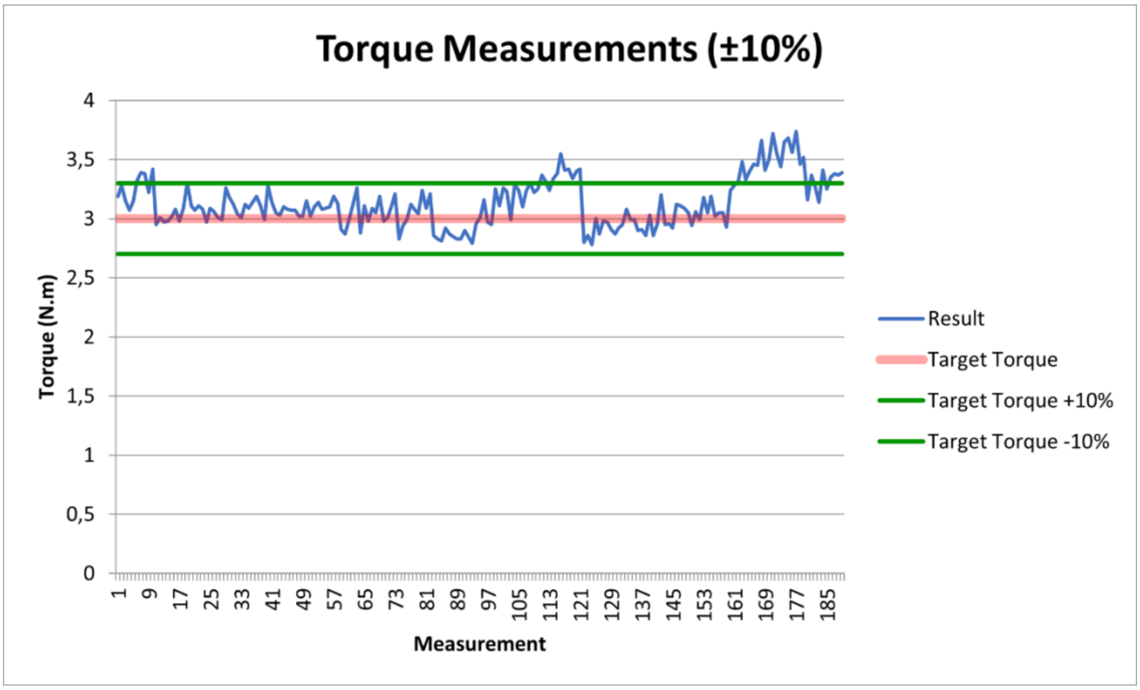


Figure 14. Tool ID 2 (X/1) torque measurement control chart.

The X/2 group performed slightly better, but still at an unsatisfactory level. The main problem in this category was a combination of moderate variability and poor centering. The control charts indicated that these tools were not completely out of control, but they also did not operate as stable and well-centered processes. The results showed that calibration alone may not be sufficient in all cases, and that some tools require maintenance in addition to centering adjustment.

The results for the Y/1 group (inline electric screwdriver) were mixed. One tool in this category performed at an acceptable level and remained stable throughout the pilot period, while the other exhibited a gradual downward drift and repeated signs of performance degradation. As shown in figure 15, the downward drift of the tool continued even after calibration. This means that the model cannot be consistently evaluated based on the pilot data. Instead, the results suggest that the problematic behavior was tool-specific and not clearly model-wide.

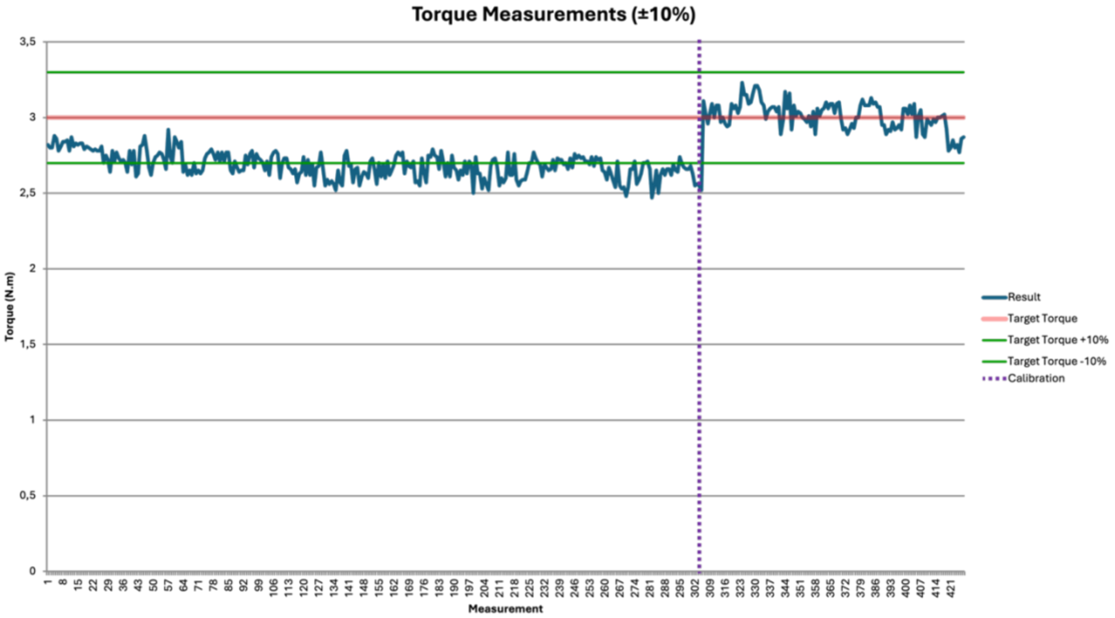


Figure 15. Tool ID 4 (Y/1) torque measurement control chart.

In the case of the AA/4 (mechanical torque wrench) group, the tool under consideration had very low variation and thus a high C_p value, but the average torque was miscentered, resulting in a poor C_{pk} value. The control chart showed a stable but miscentered process. This means that the tool was consistently delivering incorrect torque values. For these tools calibration could solve the problem. Figure 16 shows that the underlying repeatability of these tools was good, but the output was centered too close to the upper tolerance limit.

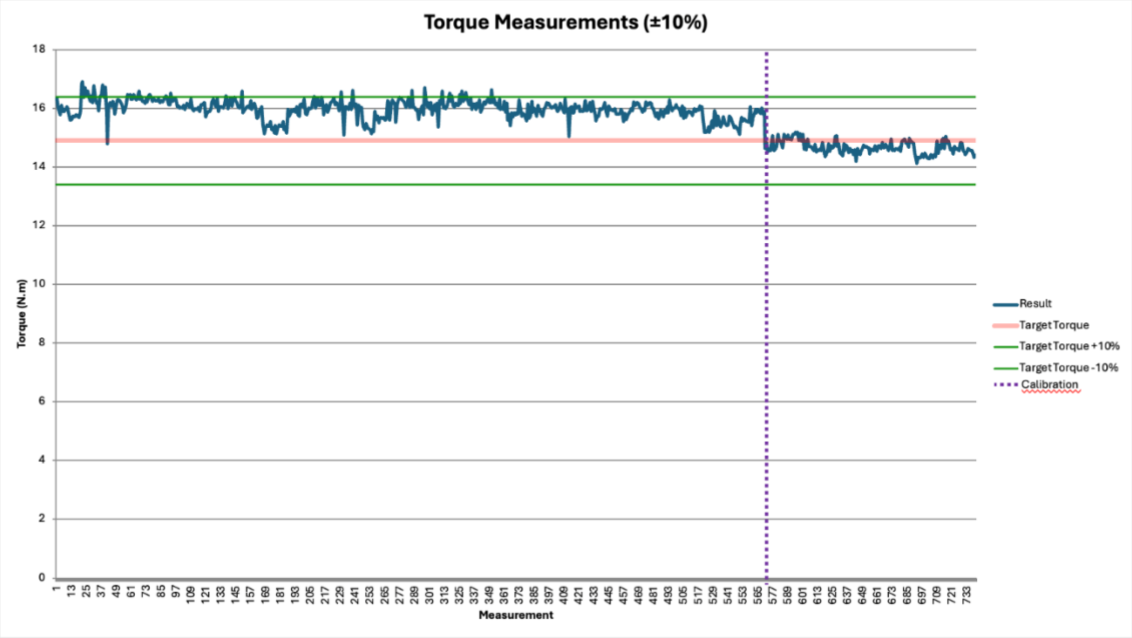


Figure 16. Tool ID 40 (AA/4) torque measurement control chart.

The poor-performing C/1 tools followed a similar logic. As can be seen in Figure 17, their natural variation was typically small, but some of them had settled below the nominal torque. This lowered the C_{pk} value. These tools are therefore not fundamentally unstable, but instead require calibration for centering adjustment to achieve better overall capability.

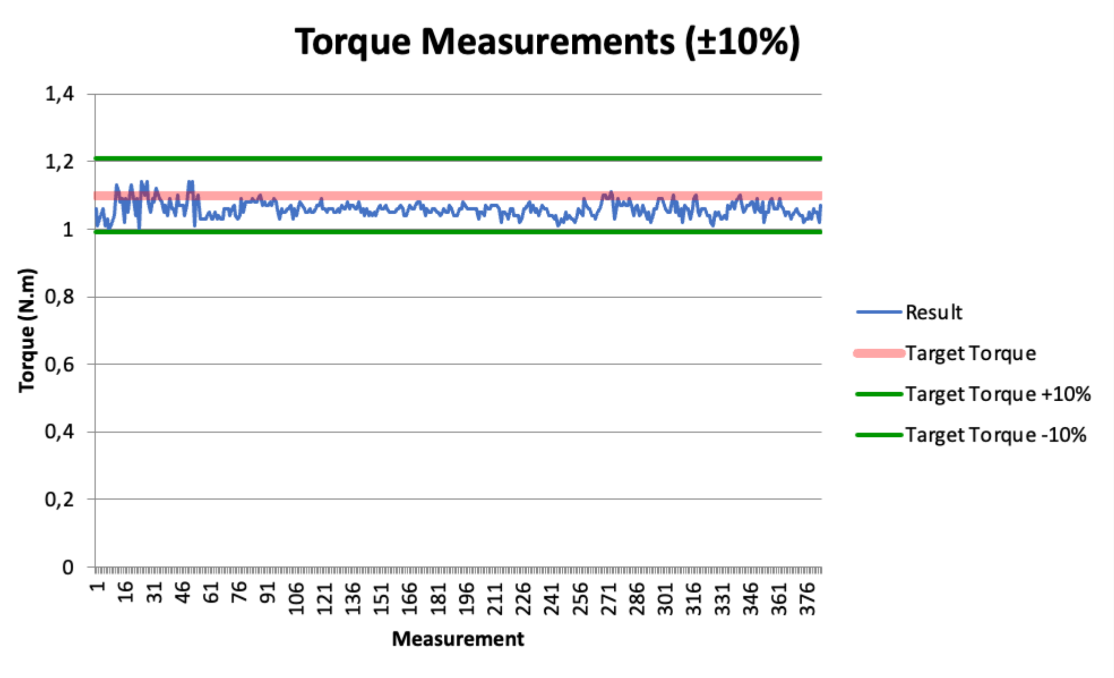


Figure 17. Tool ID 25 (C/1) torque measurement control chart.

Among the weakest-performing tool groups, two dominant types of failure can be distinguished. The first is excessive variation, where both C_p and C_{pk} are low and the process is unstable. The second is systematic deviation, where C_p remains acceptable, but C_{pk} is weak because the tool mean has moved away from the nominal value. This distinction is central to the maintenance strategy recommended in this work.

6.8.4 Comparative summary of brand and model performance

When comparing tool groups on a broader level, several conclusions can be drawn. First, the AA tool family showed the strongest overall performance in the data, although there was clear variation between model variants and individual tools. AA/1 and AA/2 performed well overall, while AA/3 contained both capable tools and weaker performing individuals. AA/4 showed good repeatability but poor centering in the case under review. Overall, AA tools appear to have good natural capability, but some individual tools require calibration more frequently than others.

Second, the X family of tools performed poorly overall. Both the X/1 and X/2 exhibited poor capability and recurring SPC issues, and the performance deficiencies were not limited to a single isolated point. Compared to other brands, these models appear to have the most difficulty maintaining acceptable torque performance over time.

Third, the Z/1 result was excellent. The tool was stable and had capable performance, but as it was represented by only one tool, the result must be interpreted with caution and does not justify broader model-level generalization.

Fourth, the C/1 and Y/1 groups demonstrated that relatively good repeatability does not necessarily guarantee satisfactory process performance if the average torque is not centered correctly, reinforcing the importance of evaluating both C_p and C_{pk} rather than relying solely on variation.

In summary, the comparative analysis shows that the key practical value of performance analysis is not just to rank brands or models, but also to identify the underlying cause of poor performance. Some tools require calibration primarily because their torque output is consistently shifted or misaligned. Other tools require more significant corrective action because their phasing is inherently too large or their behavior is unstable over time. The findings in this section were consistent with previous SPC capability analyses: high-performing tools are characterized by both low variability and good centering, while problematic tools typically exhibit either instability or persistent process deviation.

From the perspective of the developed management system, this result is useful. It shows that the system is able to identify not only NOK events, but also different types of tool wear patterns and support subsequent maintenance decisions. This confirms the practical value of the system as a basis for predictive calibration, targeted maintenance and long-term tool stock management.

6.9 Summary of results

Chapter 6 analyzed the performance of the developed torque tool management system and the measurement data collected during the pilot project from several complementary perspectives. The results showed that the developed system provides a viable basis for the verification of production torque tools and for the data-driven management of calibration and maintenance activities. Overall, the analyses show that the measurement system is capable of reliable monitoring, the collected data is suitable for statistical process control, and the system can be used for empirical justification of key operating parameters instead of just fixed routines or assumptions.

Gage R&R analysis showed that the measurement system dispersion was sufficiently small for production use, supporting the reliability of the collected torque verification data. Statistical process control demonstrated that the behavior of torque tools can be effectively monitored using control charts and that both stable and unstable tools can be identified from production data. Process capability analysis revealed clear differences between tools, with some tools operating with good performance and centered accuracy, while others were limited by either excessive dispersion or systematic bias. This distinction is important from a practical perspective, as it helps to distinguish tools that require calibration from those that require maintenance or possible removal from service.

The analyses related to tolerance settings, alarm logic, measurement series size and measurement interval produced the main practical recommendations of the study. The results supported the use of a ±10% production verification tolerance as a realistic and justified limit in the target environment. The alarm analysis showed that a multi-level alarm logic based on statistical warning signals and specifications can support earlier detection of tool deterioration. The sample size analysis supported a series size of 10 measurements as a practical compromise between statistical reliability and production feasibility. Furthermore, the time interval analysis showed that warning signals usually occur well before NOK events, which supports regular verification as a preventive control method. Based on these findings, a weekly inspection interval was selected as the most suitable practical solution for the developed system. This time frame offers a balanced compromise between early detection capability, production smoothness, and sufficient data for condition-based monitoring and decision-making.

Finally, the comparison of torque tool performance showed that the system is also useful in identifying differences between tool groups and detecting recurring performance patterns at both tool and model level. Overall, the results showed that the developed management system can improve traceability, support quality assurance and provide a practical basis for condition-based torque tool management in the target company. Thus, the chapter meets the main analytical objectives of the work and lays the foundation for the final conclusions and recommendations presented in the next chapter.

7 Discussion

The designed real-time torque tool management system has been successful in addressing issues related to tightening quality and traceability and has been able to take preventive measures with respect to safety and efficiency. The system was able to monitor torque measurements for 50 torque tools for a span of two months and to detect deviation from normal torque measurement behavior. Only 4.8% of all measurements were beyond the tolerance limit of ±10%. The system has been able to perform the required job of maintaining the integrity of the joint without halting the production process. Finally, in this chapter, we explore the correlation between the results obtained from this research study and how they relate to research questions and objectives, determine the functionality and usability of our proposed system, compare our results from this research study to similar research work available in literature and related standards, and determine their practical applications to our target industry.

7.1 Summary of key results

The results presented in Chapter 6 show that the developed torque checking concept can be parameterized in a practically feasible and statistically justified manner. Based on empirical analyses, the most suitable operating parameters of the system turned out to be a weekly verification interval, a measurement series of 10 measurements per tool, and a tolerance of ±10% of the nominal torque for torque checks performed in production. Together, these settings form a balanced compromise between statistical reliability, production feasibility, and early detection capability. The selected measurement series size proved to be sufficient to assess the tool performance with an acceptable confidence interval, while the selected tolerance was loose enough to reflect normal production conditions, but still tight enough to identify significant deviations in the tool's operation.

The overall performance of the analyzed tool sample was acceptable, although clear differences were observed between individual tools and groups of tools. During the two-month pilot period, most tools remained close to their nominal torque, and only a small proportion of all measurements fell outside the selected tolerance range. At the same time, SPC analysis showed that the behavior of the tools was not uniform. Some tools were statistically stable and well centered, some were stable but systematically miscentered, and a small proportion showed clear instability or repeated out-of-control behavior. This distinction is important in practice, as it shows that not all tools that perform well have failed for the same reason. In some cases, the main issue was in the centering, in which case calibration was the most appropriate corrective action, while in other cases the dispersion itself was excessive, indicating the need for maintenance, repair, or possible removal from service.

The measurement system analysis further demonstrated that the developed verification system is suitable for production use. The Gage R&R results showed that the variation of the measurement system was small in relation to both the observed process variation and the applied tolerance range. This means that the variation seen in the collected data can be interpreted primarily as the actual behavior of the tool and not as measurement noise. This was a critical result for subsequent analyses, as it confirmed that the verification data provided a reliable basis for setting tolerances, alarm logic, performance evaluation and determining the inspection interval.

The process performance analysis also revealed clear differences between the tools. Some of the tools performed at a clearly capable level, while others were only marginally capable, and a small group fell outside the acceptable threshold. It is noteworthy that the limitation of many of the weaker tools was not primarily excessive random variation, but rather poor centering of the process mean. This suggests that a significant proportion of the problematic tools could be improved by calibration. In contrast, tools with both poor performance and high instability represent the most critical cases, as their behavior indicates more fundamental performance problems.

Overall, the results show that the developed system is capable of supporting a more systematic and data-driven approach to torque tool management. The analyses not only identified suitable parameter values for the verification process, but also demonstrated that the system is able to distinguish between stable, miscentered and unstable tools, thereby supporting more accurate maintenance and calibration decisions.

7.2 Evaluation of the developed system

The developed torque tool management system corresponds well to the objectives set at the beginning of the study. Its main purpose was to create a practical and transparent solution for torque verification in production while at the same time producing reliable data for statistical analysis and decision-making. Based on the pilot results, the system can be considered successful in achieving these objectives. It enabled regular torque verifications to be performed directly in the production environment, supported traceability through automatic data collection, and provided a structural basis for long-term monitoring of tool condition.

From an operational perspective, the system was designed to be used by production workers as part of their normal routine. The user interface guides the operator through the torque verification procedure using user identification and tool-specific selection, after which the system records the measurement sequence to be performed from the connected Norbar torque testers. The process is therefore standardized and does not rely on manual data entry or operator memory. Measurement results are automatically stored with time stamps, tool and user IDs, and OK/NOK ratings. In addition, the calibration team has access to a summary view of the tool status, as well as access to tool-specific historical data and charts, which improves visibility into the condition of the tool population and speeds up response to abnormal results.

The system can also be assessed on the basis of its technical and methodological consistency. The use of direct torque verification for tool assessment reduces uncertainty compared to residual torque measurements on finished joints, which is in line with the principles presented in the literature review. Similarly, the chosen verification concept is supported by established quality engineering methods. The use of repeated measurement series, Gage R&R analysis, SPC charts and process performance analysis ensures that both the quality of the measurement system and the performance of the tools are assessed using established statistical approaches. In this sense, the developed system is not only practically functional, but also methodologically sound.

Overall, the system provides a functional foundation for production-level torque tool management. It supports daily verification operations without requiring extensive specialist knowledge from operators, while also generating the necessary data for more advanced monitoring, alarm handling and calibration planning. This combination of operational usability and analytical capabilities is one of the key strengths of the developed solution.

7.3 Comparison with literature and standards

The empirical results of this study are largely in line with the literature review's theoretical and normative framework. The literature review highlighted the difference between laboratory-based requirements and practical production conditions. Standards such as ISO6789 define basic accuracy requirements for torque tools under controlled calibration conditions. Typically, the tolerance is ±6%, and for certain more accurate tools it is ±4%. These values provide an important reference for tool calibration, but they do not define an appropriate inspection tolerance for torque checks performed in production. The empirical results of this study supported the use of a wider production inspection tolerance, which is consistent with the literature that torque verification performed in production must also take into account user and process-related sources of variation.

Results presented further confirm the notion of treating calibration and in-production torque verification as two different operations. The literature and standards suggest formal calibration intervals, such as 12 months or 5000 cycles, while, at the same time, acknowledging the need to conduct more frequent inspections depending on the criticality of the process and customer requirements. This statement also corresponds to the results obtained during this study. Although the observed deviation behavior suggested that some of the tools could remain at an acceptable level for longer periods, the time interval analysis and SPC results showed that a shorter verification interval is justified when the goal is not only to ensure compliance, but also to detect developing deterioration in time enough for preventive actions. In this sense, the recommended weekly verification interval reflects a cautious but well-founded production control approach, rather than a non-standard practice.

The conclusions drawn from the performance analysis of the process further validate another commonly acknowledged finding in the literature. Poor process performance is not always due to excessive variation alone. In many cases, what limits the performance of poorly performing tools was their poor centering, and not their variance. This finding makes an important difference since it has implications for the required corrective actions. These findings support the literature-based view that performance analysis should be interpreted in conjunction with SPC and measurement system analysis, rather than as a separate metric.

In general, the developed torque verification concept is consistent with both the literature and relevant standards. At the same time, the study extends this theoretical foundation by showing how these principles can be translated into a practical torque tool management system. Thus, the comparison confirms that the selected tolerance, verification logic and control method are not only empirically supported in the target environment, but also in line with established quality and calibration principles.

7.4 Managerial and operational implications

The torque tool management system developed in this study and the results of this research have several practical applications for the company. One of the biggest advantages of the created system is that it allows the company to manage the tool portfolio in a more structured and informed way. Instead of only performing calibrations at regular intervals, the company can use torque inspection data to manage the condition of the tools and detect poorly performing tools. This allows the company to make better decisions about quality control, calibration scheduling and maintenance. The system also improves traceability by linking each torque inspection result to a specific tool, user and inspection date. This is useful in situations where it is necessary to ensure a critical joint and where customer requirements require documentation of the reliability of the tightening process.

The system also has implications for resource allocation in functions such as maintenance and calibration teams. The data from the developed system enables the identification and separation of consistently drifting and clearly unstable tools, which allows calibration and maintenance resources to be allocated more efficiently. Systematically miscentered tools can be prioritized for calibration, while tools with excessive variability or repeated instability can be targeted for repair or replacement. Thus, the system supports the transition from uniform routine operations to condition-based management of the tool base.

The implementation of the developed system also places certain demands on the company. Achieving the benefits of the system depends on its consistent application and the commitment of the personnel to follow the procedures. The reliability of the system depends on the consistency and quality of its use. Poor measurement techniques and inconsistent use weaken the reliability of the data collected. Proper implementation requires comprehensive training and clear work instructions. In addition, management support is needed to make the system part of everyday production practice.

The study also highlights opportunities for the target company's digital development. Although the current application was found to be effective in a pilot environment as a spreadsheet-based software solution with programmed automation, broader implementation and support for a wider variety of torque tools may require a more advanced technical platform. From a managerial perspective, the implemented system can be seen as both a local process optimization measure and a foundation for broader digitalization in torque management.

8 Conclusion and recommendations

8.1 Answers to the research questions

The study answered all three research questions formulated in Chapter 1.2 with the developed torque tool management system. The system was piloted and empirically evaluated to answer all research questions. The first research question focused on how often the torque of the tools should be checked and how many measurements should be included in the measurement series in order to obtain reliable information on the tool performance without causing excessive load on the production process. Based on the pilot test and data analysis, a weekly inspection cycle and ten repeated measurements per tool proved to be the most suitable solution. The determined parameters provide an optimal balance between the practicality of the production process and the reliability of the results. The results also showed that WARN signals occurred before the more critical NOK signals, which supports the use of regular verification as a preventive measure. Sample size analysis showed that a series of 10 measurements is sufficient to evaluate tool performance with acceptable accuracy and is feasible in routine production.

The second research question addressed the definition of a tolerance limit that would support both production safety and operational efficiency. Based on the literature review and empirical observations, a tolerance of ±10% was considered the most appropriate choice for the case environment. Although this limit was wider than the tolerances typically used for calibration purposes in laboratories, it better reflects real production use, where operator technique, practical measurement conditions, and normal tool behavior introduce additional variation. The selected tolerance level was found to be tight enough to detect and identify tool deviations without causing excessive false alarms. In addition, the Gage R&R results indicated that the variation introduced by the measurement system was small relative to the selected tolerance, supporting the conclusion that the verification results primarily reflect the actual performance of the tool, rather than measurement error associated with the system.

The third research question asked how the alarm system should be designed to detect actual performance issues at an early stage and minimize false alarms. The results supported the use of a multi-level alarm logic that combines OK/NOK conditions based on tolerance limits and statistically determined control limits, i.e. warning signals. In this model, the NOK condition is activated either by direct tolerance violations or by statistical control limits being exceeded. In addition, the system utilizes statistical process instruction-based warning terms that identify early trends and instabilities even before actual failure occurs. This approach was well suited for the developed system, as it improves sensitivity to significant performance degradation while avoiding excessive reactions to individual random fluctuations. The combination of specification limits and SPC-based warning logic thus provides a more practical and justified basis for production-level tool monitoring.

Overall, the answers to the research questions indicate that the developed torque tool management system can be operated using a weekly verification interval, a 10-measurement verification set, a production tolerance of ±10% and a multi-level alarm structure. These parameters were not chosen solely on theoretical grounds, but were supported by empirical evidence from the pilot study. As a result, the study provides a practical and knowledge-based framework for improving torque tool verification, traceability and calibration decision-making in the target environment.

8.2 Main contributions of the thesis

The key research contributions are both theoretical and practical. From a theoretical point of view, the current research integrates torque tool theories, the phenomenon of bolted joint, measurement system analysis (MSA) and statistical process control (SPC) into a comprehensive methodology for torque control during the production process. Torque tool specifications, uncertainties, and human variability are connected to the concept of confidence intervals, natural variation, C_p and C_{pk}, and Gage R&R. Based on this, the current study manages to extract recommendations for the key parameters of the torque control method, like the uniform 10% process tolerance, the verification frequency every week, and the sampling size of 10 measurements for each series. Another key area covered within this thesis is the inclusion of control charts and the Western Electric rules into alarm logic, which includes hard limits and early warning indications, along with the use of past measurements for the transition from time-based to condition-based calibration.

As a practical deliverable, the project provides the target company with a working torque tool management system using a spreadsheet-based software solution combined with Norbar torque testers. The system provides a user-friendly interface using the user ID and the tools QR code, automatic data acquisition, storage, and a summary dashboard where the overall status of the tool can be assessed instantly, including access to the overall histories and the respective formulas. This adds to the early identification of trouble-generating tools, as well as improving the overall tightening quality process.

8.3 Limitations of the study

Although the study produced useful and practically significant results, several limitations must be taken into account when interpreting the results. Although the empirical evaluation focused on a two-month pilot conducted on one line, the results provide a comprehensive basis for the development of the entire production, as similar tools and methods are used on other lines. Although there are still line-specific differences, the pilot provides a reliable and generalizable picture of the functionality of the method on different production lines.

Second, the pilot focused on torque tools that could be verified with selected Norbar torque testers. This meant that the study primarily covered mechanical hand-operated torque wrenches and direct-drive power tools. Impulse tools were excluded from the empirical analysis. This limits the applicability of the developed system to all tool types used by the target company. The current results should therefore be considered valid specifically for the tool categories included in the pilot, and not for all production tightening tools.

An additional limitation is related to the participation of operators in the pilot phase. The primary pilot data was collected by a single operator, ensuring consistency of the collected data and minimizing the impact of user-specific variation on the initial results. Variation between users was addressed through a separate MSA analysis, but it was not possible to validate the long-term use of the system across the entire operator population within the pilot period. This should be taken into account especially in the initial phase of the system implementation, when the impact of different user practices may have a strong effect on the results before routines become established.

The statistical analyses of the study also rely on general quality engineering assumptions, such as approximate process stability, rational sub-grouping, and the practical utility of methods based on normal distributions, including 3σ logic, confidence intervals, and process performance indices. These results are well-established and justified in this type of analysis, but they do not always describe real production behavior perfectly. In particular, if the tool behavior deviates strongly from normality, is time-invariant, or contains sudden transitions, the calculated C_p and C_{pk} values and time interval estimates may give an incorrect picture of the real situation.

8.4 Recommendations for the case company

Based on the results of this thesis, several practical recommendations can be made for the case company. First, the introduction of the torque tool management system has to be integrated into the standard operation of the assembly line. Instructions with regard to the procedure of the weekly inspection, e.g. identifying the tool via a QR code, executing 10 subsequent measurements of a tool, handling OK, WARN, and NOK indications, must be established. Roles and responsibilities should also be defined. Production operators should handle routine weekly checks, while the calibration team manages WARN/NOK alerts, master data, and corrective actions. In addition, it is important that all involved parties receive comprehensive training regarding both technical aspects and the importance of torque verification.

Second, it is important to determine KPIs for assessing the effectiveness of the system. These indicators may include the number of WARN and NOK incidents, the number of tools verified on time and late, the C_p and C_{pk} values of the tool base, the frequency of calibration or corrective actions, and response time for NOK incidents. Analysis of these KPIs will help the company to detect troublesome tools, evaluate the adequacy of chosen parameters, and develop an efficient strategy for further calibration work.

Third, torque tool management should be more integrated to the company's broader quality management and maintenance processes. The data generated by the system should be utilized as part of calibration planning, internal quality monitoring and audit documentation. This way, torque checking would no longer function as a separate inspection procedure, but as part of a broader condition-based management model. Tools that give repeated warnings, have increasing dispersion or have poor performance should be prioritized for calibration, maintenance or surveillance based on actual performance data, rather than based solely on fixed routine inspections.

Fourth, the developed system should be expanded to other production lines. Since the pilot phase on one line demonstrated the functionality of the concept, the system is ready to be deployed more widely on different lines and under varying operating conditions. The expansion enables the scalability of the concept to be verified and provides valuable benchmarking information on how the tools behave in different production environments.

8.5 Suggestions for future work

Based on this thesis, several further research areas can be identified. First the developed torque tool management system could be integrated with the company's other digital infrastructure, for example quality systems. The connectivity would allow for critical data, such as NOK events, to automatically trigger maintenance actions, quality notifications or other predefined workflows. Additionally, linking tool data directly to broader production and quality information can improve traceability and support compliance. This would strengthen the practical value of the system and support the broader digitalization of quality control.

Second, the current fixed assurance concept could be extended towards a more dynamic and condition-based inspection interval management. This study recommended a common inspection interval for all tools for practical implementation. In contrast, a more complex system would be able to take into consideration the SPC history, past performance, usage frequency, and criticality level of the joint where the tool operates for every tool separately, allowing the inspections period to vary for each of them according to their individual characteristics. This type of approach allows for the resource optimization by increasing the frequency for unreliable tools and extending inspection intervals for reliable tools.

Third, the concept should be extended to include other tool types, especially impulse tools, which were excluded in this study because they could not be measured with the selected verification devices. Since impulse tools are widely used in industrial assembly, further studies should investigate suitable measurement and verification methods that can include these tools in the same torque control framework. This would broaden the applicability of the system and increase its practical significance for the target company.

References

AIAG. (2010). *Automotive Industry Action Group. Measurement systems analysis (MSA) (4th ed.).* https://www.aiag.org

Archer, D. (2008). *Torque audits are a good way to check joints*. *51*(10). https://www.researchgate.net/publication/296527785_Torque_audits_are_a_good_way_to_check_joints

Atlas Copco. (2015). *Pocket Guide to Tightening Technique*. https://www.atlascopco.com/content/dam/atlas-copco/industrial-technique/general/documents/pocketguides/9833864801_L.pdf

Bangi, J. O., Maranga, S. M., Nganga, S. P., & Mutuli, S. M. (2014). *Torque Wrench Calibration and Uncertainty of Measurement*. Academia.edu. https://www.academia.edu/64189457/Torque_Wrench_Calibration_and_Uncertainty_of_Measurement

Bickford, J. (2018). *An Introduction to the Design and Behavior of Bolted Joints, Third Edition, Revised and Expanded*. Routledge.

Cristian Pisarciuc. (2016). The use of statistical process control to improve the accuracy of turning. *IOP Conference Series Materials Science and Engineering*, *161*(1), 012011. https://doi.org/10.1088/1757-899X/161/1/012011

Croccolo, D., De Agostinis, M., & Vincenzi, N. (2011). Failure analysis of bolted joints: Effect of friction coefficients in torque–preloading relationship. *Engineering Failure Analysis*, *18*(1), 364–373. https://doi.org/10.1016/j.engfailanal.2010.09.015

Enos, D. (2014). *“Clicker-Type” Torque Tooling - Calibration Study*. UNI ScholarWorks. https://scholarworks.uni.edu/grp/3896/

International Organization for Standardization. (2012). *ISO 11148-6:2012*. ISO. https://www.iso.org/standard/61681.html

Kalmar. (2018). *Åtdragningsmoment för skruvförband Tightening torques for screw joints*. https://www.kalmarglobal.com/498129/globalassets/media/268837/268837_20060.0001---Tightening-torques-for-screw-joints.pdf.pdf

Karaman, M. O., & Kulahci, M. (2025). Process capability indices: The allure and the perils of summarizing process performance in a single index. *Quality Engineering*, 1–10. https://doi.org/10.1080/08982112.2025.2482203

Leavengood, S., & Reeb, J. E. (2015, May 15). *Performance Excellence in the Wood Products Industry: Statistical Process Control (SPC) Part 7: Variables Control Charts*. https://doi.org/10.13140/RG.2.1.2999.6967

Luca G, Bochese, & Ferrero, C. (2003). *Uncertainty of Measurement in Torque Wrench Calibration*. ResearchGate; unknown. https://www.researchgate.net/publication/283899913_Uncertainty_of_Measurement_in_Torque_Wrench_Calibration

Lundin, O., & Haettel, R. (2023). Evaluation and Damping of High-Frequency Vibrations on a Tightening Tool. *ICHAV 2023*, 18. https://doi.org/10.3390/proceedings2023086018

Meng, F., Jiao, J. P., Zhang, Z. M., & Zhang, D. L. (2021, December). *Various torque tools calibration at NIM*. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S2665917421001409

Michael. (2020, July 13). *Delta Regis Tools Inc.* Delta Regis Tools Inc. https://deltaregis.com/what-to-know-before-getting-a-torque-tester/

Montgomery, D. C. (2009). *Introduction to statistical quality control. 6. ed*. Wiley.

Ostadi, B., Taghizadeh-Yazdi, M., & Mohammadi-Balani, A. (2020). *Process Capability Studies in an Automated Flexible Assembly Process: A Case Study in an Automotive Industry*. *14*(1), 2008–7055. https://doi.org/10.22059/ijms.2020.222198.672415

Piontek, S., Gäbler, M., & Hermann Lödding. (2024). Integration of smart hand tools and digital assistance systems. *Procedia CIRP*, *130*, 427–435. https://doi.org/10.1016/j.procir.2024.10.110

Ravichandran, J. (2019). A review of specification limits and control limits from the perspective of Six Sigma quality processes. *International Journal of Six Sigma and Competitive Advantage*, *11*(1), 58. https://doi.org/10.1504/ijssca.2019.098725

Ren, X., Wang, Z., Li, H., & Li, X. (2024). Research on calibration parameters of hydraulic impulse tools. *Measurement: Sensors*, 101388. https://doi.org/10.1016/j.measen.2024.101388

Saikaew, C. (2018). An implementation of measurement system analysis for assessment of machine and part variations in turning operation. *Measurement*, *118*, 246–252. https://doi.org/10.1016/j.measurement.2018.01.008

Scharnowski, D., & Bozkurt, S. (2019). *Auslegung und Weiterzugsmomentmessungen Schraubfall Anschlusskopfmontage Seite 1 von 30 Seiten Prüfbericht Auslegung und Weiterzugsmoment- messungen am Schraubfall Anschlusskopfmontage*. https://www.atlascopco.com/content/dam/atlas-copco/local-countries/germany/documents-industrial-technique/Atlas-Copco-Schraubfallanalyse.pdf

Sennaroglu, B., & Yurtsever, O. (2018). *Evaluating Measurement System by Gauge Repeatability and Reproducibility*. https://ieomsociety.org/paris2018/papers/308.pdf

Sharma, M., Sahni, S. P., & Sharma, S. (2019, November 22). *Validating a destructive measurement system using Gauge R&R — A case study*. ResearchGate; unknown. https://www.researchgate.net/publication/354573323_Validating_a_destructive_measurement_system_using_Gauge_RR_-_A_case_study

Teixeira, E. L. S., Dias, T. S. D., Andrade, M. de O., & Oliveira, A. B. S. (2017). *STATISTICAL PROCESS CONTROL APPLICATION IN AUTOMOTIVE INDUSTRY*. ResearchGate. https://www.researchgate.net/publication/323209874_STATISTICAL_PROCESS_CONTROL_APPLICATION_IN_AUTOMOTIVE_INDUSTRY

Umer, M., Mahesh, B., Hanson, L., Khabbazi, M. R., & Onori, M. (2018). *Smart Power Tools: An Industrial Event-Driven Architecture Implementation*. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S2212827118301604

VDI/VDE 2645-2. (2014). *VDI/VDE 2645 Blatt 2 - Capability test for fastening technology - Machine capability test - MCT*. Www.vdi.de. https://www.vdi.de/en/home/vdi-standards/details/vdivde-2645-blatt-2-capability-test-for-fastening-technology-machine-capability-test-mct

Yang, G., Yang, L., Chen, J., Xiao, S., & Jiang, S. (2021). Competitive Failure of Bolt Loosening and Fatigue under Different Preloads. *Chinese Journal of Mechanical Engineering*, *34*(1). https://doi.org/10.1186/s10033-021-00663-3

Zhong, S., Chen, L., Liang, W., Nsengiyumva, W., Yu, Y., Li, T., Zhang, Q., Lin, J., Zhong, J., & Li, J. (2023). Contactless torque sensors based on optical methods: A review. *Optics and Lasers in Engineering*, *173*, 107832. https://doi.org/10.1016/j.optlaseng.2023.107832