sustainability

MDPI

Article

Climate-Adapted Lean Six Sigma for Maintenance Improvement in Petrochemical Operations: A Longitudinal Case Study of Cooling-Water Pumps

Abdulrahman Mohammed Albar and Walid M. Shewakh *

Department of Industrial Engineering, College of Engineering and Computer Sciences, Jazan University, Jazan 45142, Saudi Arabia
* Correspondence: wshewakh@jazanu.edu.sa

Abstract

Maintenance rules in Gulf process plants often come from manufacturer documents written for milder climates. This explanatory case study asks why such a rule persists and what a Lean Six Sigma Define–Measure–Analyse–Improve–Control (DMAIC) project needs to catch it. The case covers 36 months for four American Petroleum Institute (API) 610 cooling-water pumps at a Saudi petrochemical plant, using 47 failure events, 312 work orders, bearing temperature logs and 15 interviews. The dominant failure mode, bearing distress, traced to a quarterly re-lubrication rule fixed in 2015 and applied with a mineral-oil grease. Adopting the bearing makers' criterion that grease life halves for every 15 °C above 70 °C, an operating-envelope audit showed each summer interval aged the grease two to three times faster than at 70 °C, and eight of eleven baseline bearing events fell in June to September ($p = 0.009$). After the grade, quantity and interval basis were changed, bearing events fell more than other modes, bearing sets issued fell from nine to three, and summer corrective labour fell from 31 to 9 h. The new 2000 h interval is not supported at peak temperature, so a summer limit is added. The single-site design cannot estimate effect size.

Keywords: Lean Six Sigma; DMAIC; reliability-centred maintenance; overall equipment effectiveness; petrochemical industry; extreme heat; climate adaptation; operating envelope; occupational heat exposure; Saudi Vision 2030

1. Introduction

Industrial maintenance has moved a long way from its old position as a cost centre attached to production. In capital-intensive process plants it now shapes safety performance, regulatory standing and margin at the same time, which is why maintenance strategy has become a recurring topic in the operations literature [1,2]. Petrochemical plants make the point sharply. They run continuously, they hold large hazardous inventories, and an unplanned stop on a utility system can cost more than the equipment that caused it.

Two improvement traditions have grown up around this problem. Lean Six Sigma (LSS) offers a structured way to find and remove waste and variation, and its Define–Measure–Analyse–Improve–Control (DMAIC) cycle has been applied across manufacturing, healthcare and services [1,3]. Reliability-centred maintenance (RCM) offers a way to decide what maintenance a given asset actually needs, based on its functions and the consequences of losing them [4,5]. The two are complementary rather than competing: DMAIC supplies project discipline, while RCM supplies the engineering logic for the maintenance decision. Integrated applications exist [6,7], but empirical accounts that follow one asset

Check for updates

Academic Editor: Emmanuel Adamides

Received: 7 October 2026
Revised: 25 October 2026
Accepted: 27 October 2026
Published: 2 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

through a full cycle and report what changed are still uncommon, and accounts from the Gulf are rarer still.

Context matters more here than the generic literature suggests. Saudi Arabia's industrial base has expanded quickly: the number of operating factories rose from 7206 in 2016 to 11,549 in 2023, and the National Industrial Strategy targets roughly 36,000 by 2035 [8,9]. That growth is happening in an environment with summer ambient temperatures above 45 °C, high airborne dust, coastal salt exposure and a workforce drawn from many language backgrounds. Reviews of the oil and gas sector already treat heat and other extreme weather as a physical threat that operators must assess for their own plants [10], and regional climate projections place the Arabian Gulf among the places where summer extremes will grow most [11]. Equipment sold into that environment usually arrives with maintenance instructions developed elsewhere. Whether those instructions are revisited, and what happens when they are not, is an empirical question with direct consequences for asset life and for resource use.

This study addresses three gaps. First, most published LSS-maintenance cases report what was done and what improved, but few ask whether the rules being followed ever suited the place where they are used, and correcting such rules is rarely treated as climate adaptation for plants that already exist. Second, integration of reliability tools inside DMAIC is often described rather than measured, and where it is measured the analysis rarely acknowledges the structure of the data it rests on. Third, the moderating role of extreme operating conditions is asserted more often than it is documented, and it is seldom put in a form that can be calculated.

The aim of this study is therefore explanatory. It asks why a maintenance rule that did not fit an extreme operating climate stayed in force for years in a plant where the people doing the work knew it did not fit, and what a DMAIC project with reliability analysis built into each phase must contain to detect and correct a rule of this kind. Four objectives follow from the aim: (i) to trace, phase by phase, how the DMAIC and reliability tools led to a climate-specific root cause; (ii) to state the mismatch in a calculable form that predicts where failures should gather, and to check that prediction against the site's own records with statistics that suit a single-site design; (iii) to report the resource and heat-exposure consequences, claiming only what was measured; and (iv) to turn the case into a procedure that other Gulf plants can apply and evaluate.

The paper makes three contributions. The first is conceptual. It treats an inherited maintenance rule as an artefact with a built-in reference operating envelope, which becomes a latent failure cause once the site moves outside it, and it explains why the shop floor knew about the problem long before the procedure changed. The second is methodological, and narrower than a new model. The study adds an operating-envelope audit to the Measure and Analyse phases. Its quantitative step adopts the bearing makers' criterion that grease life halves for every 15 °C above 70 °C [12,13] and uses it to compare how hard a rule already in force works the grease in each season, from the rule, the grease type and a seasonal temperature record. Because the prediction comes from temperatures rather than failure counts, it can be checked against a baseline record that no later change could have shaped. One case shows only that the test pointed to the right failure mode and season at this site, so the audit is offered as a procedure to be tested rather than as a validated tool. The third contribution is practical: a staged framework for applying the audit across Gulf plants, with a rollout design that would give the causal test a single case cannot. The before-and-after figures are reported as a description of what followed, not as an effect.

The rest of the paper is organised as follows. Section 2 reviews LSS in maintenance, reliability analysis, the structure of maintenance management and the sustainability framing, then sets out the theoretical lens, the considerations for redesigning the maintenance process

and four propositions. Section 3 describes the case, the data and the analytical methods, and introduces a quantitative operating-envelope test. Section 4 reports the DMAIC application and the outcomes, ranks the evidence by strength and checks the prediction inside the site. Section 5 discusses the findings, alternative explanations and prior work, proposes a framework for wider application, and sets out the limitations and how future work can control them. Section 6 concludes.

2. Literature Review and Propositions

2.1. Lean Six Sigma in Maintenance Operations

LSS combines two improvement logics: waste elimination and flow from Lean, and variation reduction from Six Sigma. Systematic reviews covering the manufacturing sector describe a mature body of application work alongside persistent implementation problems [1,14]. Recent work has focused on what Industry 4.0 technologies add to each DMAIC phase. Pongboonchai-Empl et al. mapped which technologies support which phase and found the Analyse phase best served [15], and later proposed and validated a DMAIC 4.0 framework of 42 technology-enhanced tasks [16]. Skalli et al. reached similar conclusions from an integration-framework perspective [3].

Maintenance-specific applications are fewer but instructive. Trubetskaya et al. combined DMAIC with turnaround maintenance structure to shorten an annual dairy shutdown [17]. Shannon et al. used a design-for-LSS approach to build a total productive maintenance (TPM) and reliability framework in an active pharmaceutical ingredient plant, reporting large reductions in planned and corrective maintenance [6]. Gomaa applied an LSS-DMAIC framework to shutdown maintenance of a feed-water pumping station at an Egyptian petrochemical site, cutting non-value-added time by 60% and downtime by 43%, and raising shutdown efficiency from 27% to 49% [7]. Data-driven variants have also appeared: Lucantoni et al. built a framework linking multi-source plant data to TPM decisions [18], and the same group used rule-based machine learning to select improvement actions for overall equipment effectiveness (OEE) [19].

Against these successes sits a consistent list of failure factors: thin management commitment, poor project selection, weak training and, in smaller firms, limited resources and absent methodology [14,20]. Any new implementation account is more useful when it says which of these it had to manage.

2.2. Reliability Analysis and Maintenance Decision-Making

RCM originated in commercial aviation with the Nowlan and Heap report and was later codified for industry by Moubray [4,5]. Its central move is to decide maintenance from function and failure consequence rather than from equipment age. Contemporary work extends this toward digitalisation: Introna and Santolamazza merge RCM with a digitalisation score to identify opportunities that classical criticality analysis misses [2], while bibliometric work maps how maintenance performance measurement has shifted in the Industry 4.0 period [21].

Failure mode and effects analysis (FMEA) remains the workhorse for prioritising within RCM. Risk priority numbers (RPNs) are simple to compute and easy to communicate, and their weaknesses are equally well known: the multiplicative scale compresses distinct risk profiles into the same number, and the ratings depend on the experience of the team supplying them [22]. Used as a ranking device rather than an absolute risk measure, FMEA still earns its place.

Weibull analysis is the usual statistical description of a failure population, with the shape parameter β locating it on the bathtub curve. It needs a run of complete times between failures, which a short observation window does not provide. With about five functional

failures a year and 12-month windows, the first and last interval in each year would be cut by the window, so this study does not fit life distributions (Section 3.5).

2.3. Maintenance Management Structure and Where the Study Intervenes

It helps to place the project inside the maintenance system it changed. Crespo Márquez et al. describe maintenance management as a loop that runs from objectives and asset priorities, through the choice of maintenance strategy and the planning of work, to execution, assessment and improvement [23]. The loop works at three levels. At the strategic level the plant sets objectives, ranks assets by criticality and decides how its rules are governed. At the tactical level it chooses for each asset a mix of corrective, time- or usage-based preventive and condition-based tasks, and fixes their intervals, quantities and alarm limits. At the operational level technicians carry out the work orders and record what they find, which feeds the next round of assessment.

Figure 1 maps the site's maintenance structure onto these levels and marks where each countermeasure acted. Before the project the plant ran a conventional mixed programme: preventive tasks on fixed calendar intervals, including the re-lubrication rule taken from the equipment manual; monthly vibration surveys and quarterly lubricant sampling as condition checks; and corrective work raised through the computerised maintenance management system (CMMS). Planned work made up 78% of the total. Most of the project's weight fell at the tactical level, where the re-lubrication rule and the alarm limits sit. The strategic level gained a rule owner and an annual review trigger, which is where the eight-year gap had its origin. The operational level gained 5S, bilingual standard work and training, which decide whether a correct rule is carried out correctly.

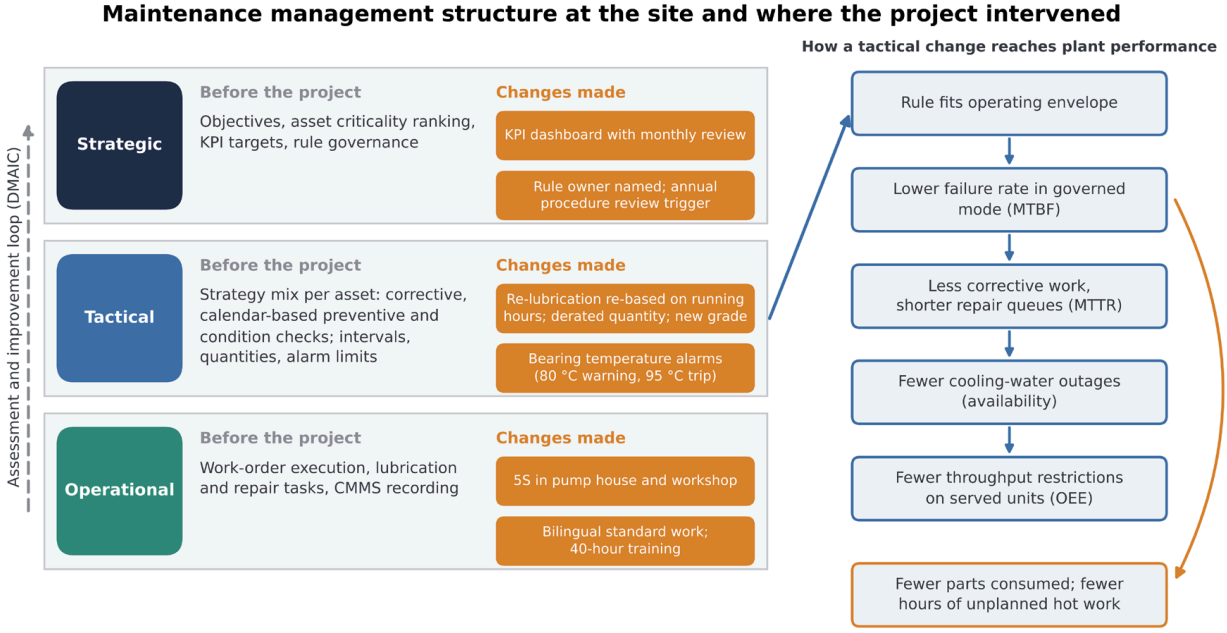


Figure 1. Maintenance management structure at the site, organised by the strategic, tactical and operational levels of the maintenance management process [23], with the changes made by the project and the chain through which a tactical change reaches plant performance. DMAIC, Define–Measure–Analyse–Improve–Control; KPI, key performance indicator; CMMS, computerised maintenance management system; MTBF, mean time between failures; MTTR, mean time to repair; OEE, overall equipment effectiveness.

The right-hand side of Figure 1 shows how a change at the tactical level reaches plant performance. A better-fitted rule lowers the failure rate of the mode it governs. Fewer

failures mean less corrective work and shorter repair queues. For a cooling-water system, fewer pump outages mean fewer throughput limits on the units the water serves. The same chain carries the resource and exposure effects, since every avoided failure is a set of parts not used and a repair not done in the heat. This chain is the reason a rule for four pumps is worth studying: its effects do not stay inside the maintenance department.

2.4. Sustainability Framing

Most of the debate on climate adaptation in industry concerns new design: hotter design ambients, larger coolers, and tougher materials. Far less is said about plants that already exist, which will run for decades on equipment bought with maintenance rules written for cooler places. Adapting those rules is cheap and quick, and it is the part of adaptation that maintenance itself controls. LSS work has begun to link improvement projects with sustainability goals. Marques et al. show how a DMAIC project in a fish processing plant can be aligned with sustainability objectives rather than treated as separate from them [24], and Hariyani and Mishra identify and rank the drivers behind the adoption of integrated sustainable, green, lean, Six Sigma and agile manufacturing systems [25]. Reviews of continuous improvement still report that the sustainability consequences of LSS projects are under-measured [26], and LSS maintenance studies have so far said little about rule correction as a form of adaptation.

This paper takes a narrow but concrete route. It claims as sustainability outcomes only quantities taken from plant records: bearing sets and seals issued from stores, grease issued, and corrective labour on the pump set in the hot season. That follows the direction of recent work on measuring the sustainability effects of maintenance with indicators tied to plant records [27]. OEE, availability and repair time are reported as maintenance outcomes, not sustainability outcomes, and energy, which was not metered, is not claimed. Read this way, the sustainability question is one of adaptation and of resource intensity. We report three plant-level indicators that a site can track year on year: parts drawn from stores per running hour, corrective labour in the hot season, and the hours a critical set runs without a spare. They point in the direction of Sustainable Development Goal (SDG) targets 8.8, 12.2 and 13.1 [28], but they do not measure those targets, and we use them as plant indicators rather than as SDG claims.

2.5. Theoretical Lens, Redesign Considerations and Propositions

Dynamic capabilities [29] offer one way to read the DMAIC sequence, as sensing, seizing and reconfiguring. That reading describes the project well enough, but it explains little about why a rule stayed wrong for eight years in a plant full of people who knew it was wrong. We therefore use a narrower lens that speaks to that question directly.

Every maintenance rule carries an assumption about where it will be used. A re-lubrication interval, a grease quantity or a grease grade is worked out for a reference operating envelope: a bearing temperature, a load, a contamination level and a duty cycle. Inside that envelope the rule is correct. Outside it, the rule is still followed to the letter while the physics it was built on no longer holds. Reason calls conditions of this kind latent: they are created by decisions made far from the point of failure, they do no visible harm on any single day, and they show up only when they combine with local conditions [30]. An unrevised rule in a hot plant fits that description closely. Nobody breaks the procedure; the procedure itself is the problem.

The lens also says something about who knows. Latent conditions are often visible to the people nearest the work long before they are formally recognised, because recognition needs a routine that can turn a shop-floor view into a changed standard. Seen this way,

DMAIC matters less as a toolkit than as a governance routine: it produces a documented root cause, a sponsor who can approve the change and a control plan that keeps it in place.

Redesigning the maintenance process under this lens rests on a small set of considerations, listed in Table 1 with the way each was handled in this case. The first two, the operating envelope and the basis of the interval, are the ones the lens puts at the centre. The rest decide whether a technically correct rule is actually carried out, noticed when it drifts and kept up to date.

Table 1. Main considerations for redesigning the maintenance process and how each was addressed in this case.

Consideration	Question Asked	How It Was Addressed Here
Operating envelope	Does the rule's reference condition (bearing temperature, contamination, duty) match what the equipment actually sees?	Summer bearing temperatures compared with the 70 °C reference that applies to the original grease; rule re-based (Section 4.1)
Interval basis	Should the task be triggered by calendar time, running hours or condition?	Moved from a quarterly calendar rule to 2000 running hours, with temperature alarms as a condition trigger; summer limit of about 1100 running hours added after the envelope test (Section 3.6)
Failure consequence and redundancy	What does loss of function cost, given the 3 + 1 arrangement?	Set ranked critical in the plant's criticality ranking; standby rotation kept
Detectability	Can degradation be seen before function is lost?	Failure mode and effects analysis (FMEA) detection rating led to alarmed bearing temperature monitoring (80 °C warning, 95 °C trip)
Lubricant selection	Is the grease grade suited to the temperature, and is the new grade compatible with the old?	Higher-temperature grade adopted; specification and changeover in Section 4.1
Workforce and documentation	Can every technician read and follow the procedure?	Arabic and English standard work with visual aids; 40 h training
Workplace organisation	Are tools, parts and lubricants clean, labelled and where they are needed?	5S in the pump house and the workshop (Section 4.1)
Governance	Who owns the rule and what triggers its review?	Reliability engineer named as owner; annual procedure review
Resource and exposure effects	What does the change do to parts, lubricant, energy and hot work?	Tracked with evidence type stated (Section 4.4)
Regulatory constraints	Which rules limit when work can be done?	Planned work can be moved outside the hottest hours; unplanned work cannot (Section 4.4)

Four propositions follow from the lens (Figure 2).

P1 (latent rule mismatch): Where the site operating envelope exceeds the reference envelope of an inherited maintenance rule, the rule acts as a latent failure cause. Its effect should be concentrated in the failure mode the rule governs and in the periods when the gap is largest. It would count against P1 if baseline bearing events were spread evenly across the year, or if bearing temperatures stayed inside the reference envelope.

P2 (governance rather than knowledge): Where such a mismatch persists, the binding constraint is the absence of a routine that authorises rule revision rather than a lack of technical knowledge. A structured improvement cycle such as DMAIC supplies that routine. It would count against P2 if shop-floor staff had not recognised the mismatch before the project, since the problem would then be one of knowledge.

P3 (mode-specific response): Correcting the mismatched rule should reduce failures in the governed mode more than in other modes, and more in the hot season than in the rest of the year. Site-wide influences such as staffing, attention or record-keeping changes should not produce that pattern. It would count against P3 if bearing and other modes fell in similar proportion after the change.

P4 (thermal gap as moderator): Across sites, the benefit of re-basing a rule should grow with the gap between the rule's reference condition and the site's actual condition. One site cannot refute P4; across sites, it would count against it if the benefit did not rise with the thermal gap defined in Section 3.6.

P1 to P3 can be examined inside one site: Section 4.3 does this for P1 and P3, and the qualitative findings in Section 4.5 bear on P2. P4 needs variation in operating conditions that one site does not provide, so it is stated in a form that multi-site work can test, with a moderator that can be measured. Section 3.6 gives the thermal gap a calculable form, which turns P1 into a prediction that can be made before the failure record is examined.

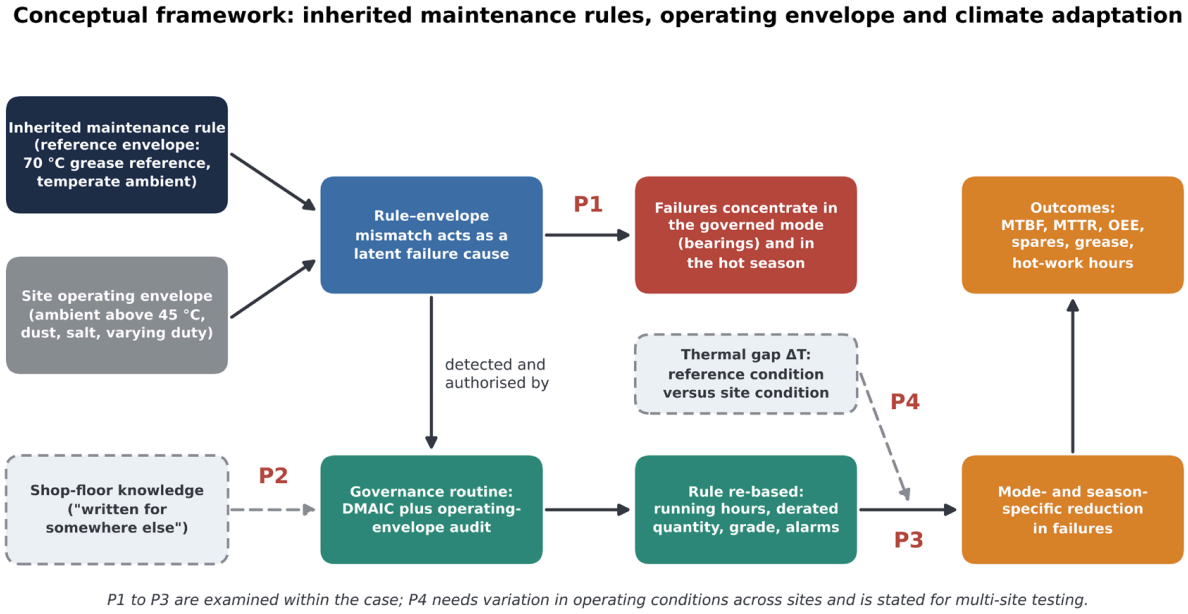


Figure 2. Conceptual framework. An inherited rule with a temperate reference envelope becomes a latent failure cause in a hotter site envelope (P1); DMAIC with an operating-envelope audit acts as the routine that detects and authorises the change (P2); the correction should reduce failures mainly in the governed mode and the hot season (P3); the thermal gap between reference and site conditions moderates the size of the benefit across sites (P4). Solid arrows show the proposed causal chain; dashed arrows and dashed boxes show conditions that enable or moderate it (shop-floor knowledge for P2, the thermal gap for P4).

3. Materials and Methods

3.1. Research Design

This is a longitudinal, mixed-methods explanatory case study. It asks how and why a mismatched rule persisted and was corrected, not how large the benefit of correcting it is. The case design follows Yin [31]: the phenomenon cannot be separated from its context (a hot, dusty, coastal plant with a multilingual workforce), which is what makes a case appropriate here. A single case of this kind supports analytic rather than statistical generalisation: it can show that a mechanism operates and under what conditions, not how often it operates across an industry [31,32]. One case can be persuasive when it shows a mechanism clearly, and much less so when it is used to estimate the size of an effect [33]. The analysis is built for the first purpose.

The design has no control group and no randomisation. Causal attribution is therefore limited, and we treat the statistical results as descriptions of the size of a shift rather than as confirmatory tests of an effect. Section 5.2 works through the alternative explanations in detail rather than listing them as a formality. To narrow the gap, the analysis adds two comparisons inside the site that the proposed mechanism and its main rivals predict differently (Section 3.5). Section 4.3 ranks the evidence by how well it resists the rival explanations.

3.2. Case and Equipment

The site is a petrochemical facility in the Eastern Province of Saudi Arabia. It was selected because it is typical of Saudi petrochemical operations, holds complete maintenance records, granted access to both data and personnel, and operates in conditions that make the climate question visible.

The focal asset is a set of four horizontal centrifugal pumps in cooling-water service, each 200 kW, built to American Petroleum Institute (API) Standard 610, arranged in a 3 + 1 configuration and supplying 850 m^3/h to process heat exchangers. Loss of cooling water restricts throughput directly, which makes the set critical in the plant's own criticality ranking. The pumps were installed in 2015 and have complete CMMS histories.

The operating environment is demanding in three separate ways. Summer ambient temperature regularly exceeds 45 °C. Airborne dust and sand load the filtration and sealing arrangements. Proximity to the sea accelerates external corrosion. None of these conditions is reflected in the equipment documentation supplied at installation.

Table 2 lists the main technical data for the pump set, and Figure 3 shows its arrangement with the bearing temperature measurement points and grease fittings. Details that would identify the manufacturer or the plant are withheld under the confidentiality agreement with the participating organisation. The data kept are the ones needed to judge the lubrication and monitoring questions discussed in Section 4. The participating organisation reviewed the technical data in Table 2 and authorised their publication.

Table 2. Technical data for the cooling-water pump set.

Item	Value
Service	Cooling-water circulation to process heat exchangers
Design standard	API 610, 11th edition (2010) [34]
Number and arrangement	Four horizontal centrifugal pumps; three duty and one standby, with the standby rotated
Design duty	850 m^3/h (reference flow for the OEE performance term)
Driver	Electric motor, 200 kW, 4-pole, on the plant 60 Hz bus; no variable-speed drive
Pump type (API 610 designation)	BB1 (axially split, between-bearings, single-stage)
Rated head and speed	52 m; 1780 r/min
Bearings (drive end/non-drive end)	DE: 6318 C3 deep-groove ball; NDE: NU 318 ECP cylindrical roller with 6318 C3 locating ball (three bearings per pump)
Bearing lubrication	Grease, manual re-lubrication through two 1/8 in National Pipe Taper (NPT) fittings on each of the two bearing housings (four fittings per pump)
Mechanical seal	Single pusher cartridge seal, API 682 Category 1, Plan 11
Coupling	Spacer-type flexible disc coupling, 200 mm spacer, rated 200 kW at 1780 r/min
Bearing temperature instrumentation	Spring-loaded dual-element Pt100 resistance temperature detector (RTD) (3-wire) in a thermowell on the DE and NDE housings, logged in the plant historian
Year installed	2015
Operating environment	Summer ambient above 45 °C; airborne dust and sand; coastal salt exposure
Criticality	Critical in the plant's ranking; loss of cooling water restricts throughput

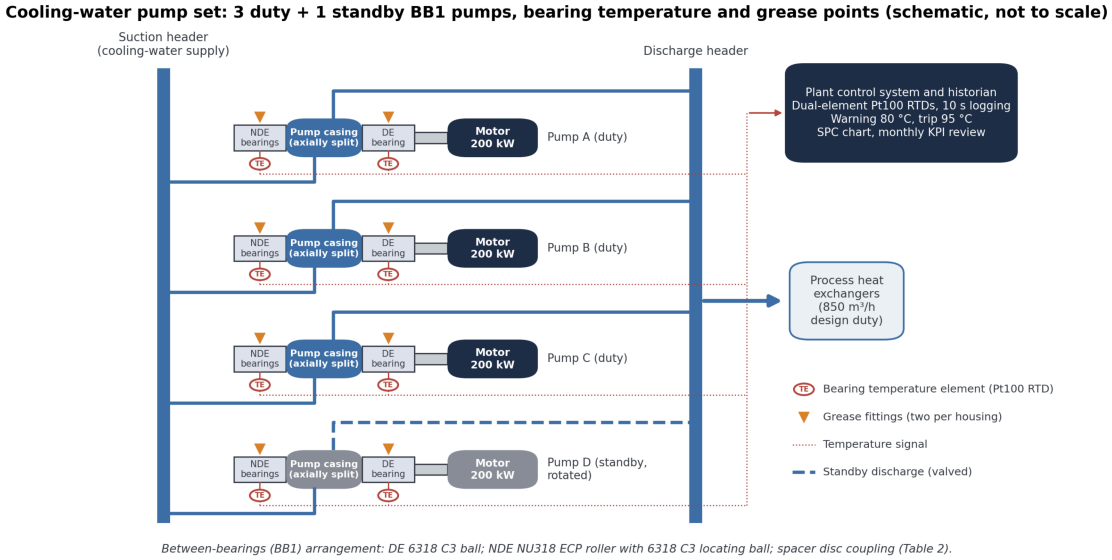


Figure 3. Arrangement of the cooling-water pump set (schematic). Each BB1 pump has Pt100 temperature elements in the drive-end and non-drive-end bearing housings and grease fittings for manual re-lubrication; temperatures are logged in the plant historian at 10 s intervals, with alarm limits added during the project.

3.3. Timeline and Data Sources

The study covers 36 months in three consecutive blocks (Figure 4): a 12-month baseline observation period from January to December 2022; a 12-month implementation year in 2023, within which the DMAIC project ran for 28 weeks followed by a sustaining period; and 12 months of post-implementation monitoring from January to December 2024. All pre/post comparisons in Section 4 use the two 12-month observation periods. The 36-month failure record used for the Pareto and FMEA work spans the whole window.

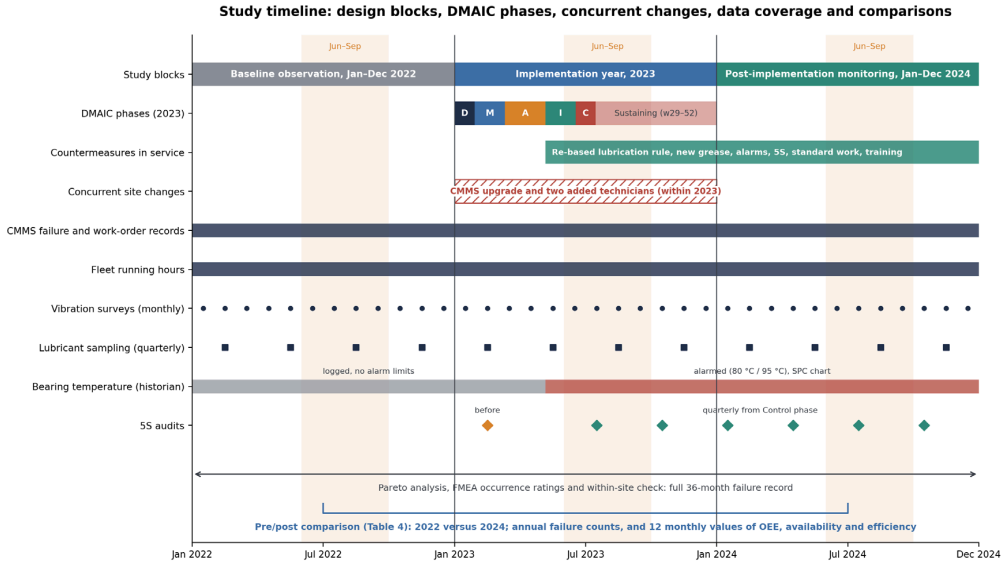


Figure 4. Study timeline: the three 12-month blocks, the DMAIC phases in 2023 by week, the period in which the countermeasures were in service, concurrent site changes, the coverage of each data stream and the two comparisons used in the analysis. Shaded bands mark June to September in each year. D, M, A, I and C are the DMAIC phases. The hatched bar marks the concurrent site changes. Circles mark monthly vibration surveys and squares quarterly lubricant sampling; diamonds mark 5S audits (orange, before implementation; green, after). The bearing temperature bar is grey while the record was logged without alarm limits and red once alarms and the SPC chart were in place.

Figure 4 sets out the design on one calendar axis: the three 12-month blocks, the DMAIC phases inside 2023, the period in which the countermeasures were in service, the two concurrent site changes that compete with the intervention as explanations (Section 5.2), and the coverage of each data stream. The post-implementation year begins about six months after the re-lubrication change took effect.

Four data sources were used, which allows triangulation between what the records say and what the people who maintain the pumps say:

Operational records: Extracted from the CMMS: 47 recorded failure events, 312 maintenance work orders, equipment running hours and spare parts issues (bearing sets, mechanical seals and grease drawn from stores).

Condition monitoring: Vibration surveys at monthly intervals, bearing temperature from the plant historian, and quarterly lubricant sampling. The way bearing temperature was measured and used is described with the countermeasures in Section 4.1. Bearing housing temperatures were logged at 10 s intervals throughout the 36 months, but that record was not released for this study; the analysis therefore uses the typical and peak summer values reported in Section 3.6.

Qualitative data: Semi-structured interviews with 8 maintenance technicians, 3 supervisors and 4 operations staff ($n = 15$), plus direct observation of maintenance work and review of procedures and training materials.

Process data: Value stream observations of the work-order process, 5S audit sheets and FMEA worksheets.

3.4. Definitions of Performance Measures

Ambiguity in maintenance metrics is a common source of non-comparable results, so the definitions used here are stated explicitly and follow ISO 14224 terminology where applicable [35].

A *failure event* is any recorded corrective work order raised against one of the four pumps. This is the count used for Pareto analysis ($n = 47$ over 36 months) and includes both loss of required function and degraded operation detected before function was lost. For the within-site check, events are grouped by the component group recorded on the work order; seal, coupling and other events together form the non-bearing group.

Mean time between failures (MTBF) is fleet running hours divided by the number of functional failures, that is, events in which a pump could no longer perform its required duty. Because degraded-mode events are excluded, the MTBF count is smaller than the 47 recorded events; the two figures are not interchangeable and are kept separate throughout. MTBF is reported once per observation year, as 26,280 h divided by the number of functional failures in that calendar year, together with the count itself. Because both the failures and the running hours are counted inside the same 12-month window, annual MTBF is a rate estimate. A failure just before or just after a window leaves out its exposure as well as its count, and each window covers a full seasonal cycle, so the window edges do not bias the rate. What the edges do cut is the time between failures, which is why no life distribution is fitted (Section 3.5).

Mean time to repair (MTTR) is the mean active repair time per functional failure, from work order start to return to service. It is reported as the mean over the functional failures in each observation year, together with their number. Repairs last hours, so the window edges do not cut them.

Operational availability is uptime divided by the sum of uptime and total downtime, including planned maintenance downtime. It is computed for each pump and averaged over the four pumps, in line with the per-pump performance term of the OEE. It is not the

inherent availability MTBF/(MTBF + MTTR), which for this equipment exceeds 99% in both periods and is therefore uninformative.

OEE is availability × performance × quality, where performance is delivered flow against the 850 m^3/h design duty and quality is the share of running hours within the specified cooling-water supply band.

Maintenance process efficiency is value-added time divided by total lead time for the maintenance work-order process, measured from the value stream observations.

Fleet running hours are taken as 26,280 h per year, corresponding to three of the four pumps running continuously with rotation of the standby unit. This figure is the numerator of the annual MTBF in both years. With rotation, each pump runs about 6570 h a year, or 75% of the time, and this per-pump duty is the basis for converting running hours into calendar days (2000 running hours ≈ 111 days; a calendar quarter ≈ 1640 running hours). The 26,280 h figure is the scheduled duty of the set rather than a metered total. It is consistent with the availability figures because availability is defined per pump and the standby covers a duty pump while it is down: four pumps at 89.2% availability give about 31,260 available pump-hours against 26,280 required. As a check, Section 4.2 also reports the annual MTBF with running hours capped at availability × 26,280 h, the most conservative reading of the availability data.

Annual rather than monthly reliability measures. With about five functional failures a year, most calendar months contain none, so a monthly MTBF or MTTR is undefined in most months and a month-by-month series would have to be constructed rather than observed. Both measures are therefore reported once per observation year, with the number of failures they rest on. OEE, operational availability and maintenance process efficiency are defined in every month and are summarised from their 12 monthly values.

Summer is taken as June to September, the hottest part of the year in the Eastern Province and the months that overlap the national midday outdoor work ban. The seasonal comparisons in Sections 4.3 and 4.4 use this definition.

3.5. Analytical Methods

Pre/post comparison: Functional failures are counted per year and, because fleet running hours were the same in 2022 and 2024, compared as Poisson rates with an exact conditional test and an exact 95% interval for the rate ratio; annual MTBF follows from the counts. MTTR is compared with a Welch *t*-test on the individual repair times. OEE, operational availability and maintenance process efficiency were aggregated monthly (12 observations per period) and are summarised by the mean difference, with Cohen's d as description [36]. The monthly series are autocorrelated, since consecutive months share equipment state and staff, so each mean difference is given with intervals recomputed for a lag-1 autocorrelation ρ of 0.3, 0.5 and 0.7, using the variance inflation $(1 + \rho)/(1 - \rho)$ and the effective sample size $n(1 - \rho)/(1 + \rho)$ [37]. A range is used because ρ estimated from 12 points is imprecise and biased toward zero [37]. A block bootstrap was set aside because each period would hold only three or four blocks. An interrupted time-series model [38] would describe the shape of the change better, but it would not settle attribution, because the CMMS upgrade and the staffing change fell in the same year as the intervention. All tests are therefore descriptive. Statistical calculations and the Monte Carlo simulation in Section 4.4 were run in Python 3.13 (Python Software Foundation, Wilmington, DE, USA) using SciPy 1.18.1 and NumPy 2.5.3.

Power of the failure-count comparison: The exact test on functional failures has very little power at these counts. With five failures in the baseline year and equal running hours, a follow-up year with no failures at all would give a two-sided *p* of 0.0625; six baseline failures would have been needed before even a perfect result could reach the 5% level. The

comparison of functional failures therefore could not show an effect under any outcome. It is reported in Section 4.2 as a record of what happened, not as a test of the change, and the same applies to annual MTBF, which is computed from the same counts.

Life distributions: Weibull models are not fitted. With four or five functional failures in a 12-month window, the first and last time between failures in each year cross the window boundary and are censored at one end, and a fit to so few intervals, several of them censored, would describe the window more than the pumps. None of the conclusions depends on a fitted life distribution.

Within-site mechanism check: Because there is no untreated comparison group, two comparisons inside the site were added that the proposed mechanism and its main rivals predict differently [39,40]. The first treats the non-bearing failure modes (seal, coupling and other events) as non-equivalent dependent variables. The lubrication changes act on bearings. Staffing, the CMMS upgrade, attention effects and general learning act on all modes alike. If the lubrication mechanism carried the improvement, recorded bearing events should fall by clearly more than non-bearing events between 2022 and 2024; if the site-wide influences carried it, the two should fall in similar proportion. The second comparison uses season. Bearing temperatures peak in summer, so the mechanism predicts that baseline bearing events cluster between June and September and that the post-period reduction is largest there. Both comparisons use the recorded failure events defined in Section 3.4, which give larger counts than the functional-failure series. Because fleet running hours were the same in both years, the odds ratio of the 2 × 2 table of mode by year equals the ratio of the bearing and non-bearing rate ratios, and Fisher's exact test was applied to that table. With counts this small, the result is read as a pattern rather than a precise estimate. The 2023 counts are also examined as a trend indicator, since a decline that starts before a change is fully in service is the usual sign of regression to the mean when an asset has been selected for its problems [41].

Qualitative analysis: Interviews were conducted in Arabic or English according to the participant's preference, audio-recorded where consent was given and otherwise recorded in contemporaneous notes, then transcribed and translated into English where needed. Coding followed a thematic approach: open coding by one researcher, review of the code list by the second, and consolidation into themes retained only where they appeared across more than one role group. Illustrative quotations were checked back with the speaker.

3.6. A Quantitative Operating-Envelope Test

The operating-envelope audit has a qualitative side, writing down the conditions a rule assumes, and a quantitative side that can be computed before anything is changed. For grease-lubricated rolling bearings, the quantitative side follows from public bearing-maker guidance, which sets re-lubrication intervals for a reference bearing temperature of 70 °C with good-quality lithium-soap greases on mineral base oil, and halves them for every 15 °C above it [12,13]. Greases with better high-temperature performance start the halving at a higher temperature; in the grease performance classes of SKF (Gothenburg, Sweden), each step up shifts the temperature scale by 15 °C [42].

Let T_{ref} be the temperature at which the halving starts for the grease the rule is applied with (70 °C for a mineral-oil lithium grease) and T_{site} the bearing temperature the equipment actually reaches in a given season, taken from historian or handheld records. The thermal gap is $\Delta T = T_{\mathrm{site}} - T_{\mathrm{ref}}$, and the envelope correction factor is

$$f = 2^{-\Delta T/15} \text{ for } \Delta T > 0, \text{ and } f = 1 \text{ otherwise.} \quad (1)$$

The factor f is not a new model. It restates the halving rule in [12,13] and is used here as an adopted engineering criterion. What the audit adds is the way it is applied. A

fixed rule with interval t_{rule}, in running hours, ages the grease in a season with factor f as much as an interval of t_{rule}/f would at the reference temperature, so $1/f$ says how much harder the same rule works the grease in that season than at 70 °C. Read this way, the test needs neither the absolute base interval, which depends on bearing size, speed, load and contamination [12], nor the reference condition the equipment maker had in mind. It needs the rule, the grease type and a seasonal temperature record, and it predicts where failures in the governed mode should gather before any outcome data are examined.

For this site, the equipment documentation supplied at purchase cannot be reproduced under the confidentiality agreement, so the test is built not to depend on it. The original grease was a lithium-soap grease on mineral base oil (Section 4.1), the type for which the public 70 °C reference applies, so $T_{ref} = 70$ °C comes from the grease rather than from the manual. Under the 3 + 1 rotation each pump ran about 75% of the time (Section 3.4), so the quarterly rule meant about 1640 running hours between re-lubrications. Summer bearing housing temperatures recorded in 2022 were typically 85 °C with a peak of 93 °C, giving ΔT of 15 °C and 23 °C and f of 0.50 and 0.35. Each summer interval therefore aged the grease 2.0 to 2.9 times as much as the same interval at 70 °C, before any allowance for dust. The statement is relative. If the quarterly rule was adequate at 70 °C, which the low cool-season bearing rate in 2022 is consistent with (3 events in eight months against 8 in four), then it was not adequate in summer. The test predicts that bearing events should cluster in June to September, and Section 4.3 checks this against the failure record. The factor is a screening tool rather than a design calculation: it leaves out load, speed and contamination, which a full re-lubrication calculation would include [12].

Applied to the revised rule, the same test gives a less comfortable answer. The new limit of 2000 running hours is about 22% longer than the 1640 it replaced (111 against 91 calendar days), so the correction lay in the grade, the quantity and the basis of the interval, not in a shorter interval. With the original mineral grease, a 2000 h interval would age the grease 2.4 (85 °C) to 3.5 (93 °C) times as much as the old rule did in the cool season. The new lithium-complex polyalphaolefin (PAO) grade has a 160 °C upper limit; if it is credited with a 15 °C shift in the start of halving, one step in SKF's performance classes [42], the figures fall to 1.2 and 1.8. The credit is an assumption, since no bearing maker has classed this grade. The 2000 h limit therefore cannot be derived from f, and at peak temperature it lies outside the range f supports even with the credit. With the credit, holding summer grease ageing at the old cool-season level would need about 1640 running hours at 85 °C and 1140 at 93 °C, so the control plan now adds a summer limit of about 1100 running hours from June to September, or re-lubrication at the 80 °C warning, whichever comes first (Section 4.1). One summer with one bearing event cannot show that 2000 h is safe at peak temperature.

4. Results

4.1. DMAIC Application

Define (weeks 1 to 4): A cross-functional team was formed: the maintenance supervisor, two senior technicians, a reliability engineer, an operations representative and a process engineer. Members received 16 h of training covering DMAIC, basic statistical tools and reliability concepts. Targets were set against the baseline: raise MTBF by at least 20%, cut MTTR by at least 25%, raise the planned maintenance ratio above 90%, raise OEE above 80%, and cut maintenance cost per running hour by at least 20%. Section 4.2 reports performance against each of these, including the one that was not met.

Measure (weeks 5 to 10): Baseline key performance indicators (KPIs) were established from the 2022 records: 5 functional failures (annual fleet MTBF 5256 h), MTTR 8.5 h, planned maintenance ratio 78%, OEE 68.5% and operational availability 89.2%. Value stream

observation of the maintenance work-order process gave a total lead time of 495 min, of which 289 min (58.4%) were value-adding (Figure 5). The largest single pool of non-value-added time sat in the repair step, where waiting for parts and for permits dominated.

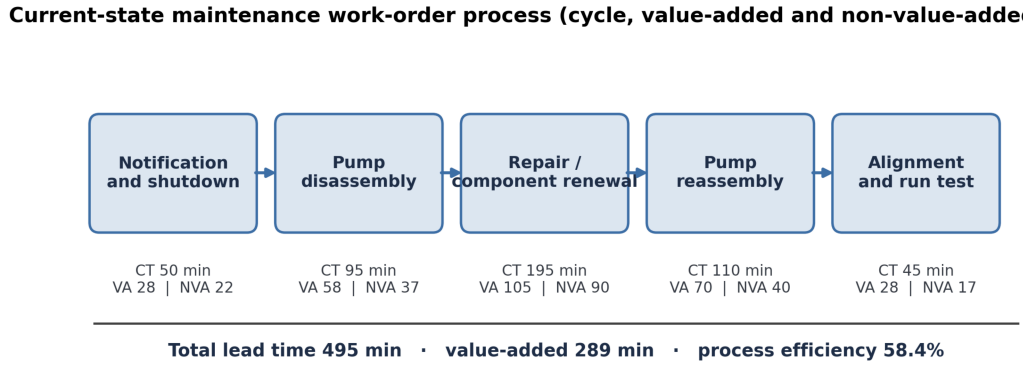


Figure 5. Current-state map of the maintenance work-order process. CT is cycle time, VA value-added time and NVA non-value-added time. Flow runs left to right.

Analyse (weeks 11 to 18): Pareto analysis of the 47 recorded failure events over 36 months identified three dominant modes accounting for 83% of events: bearing failure (20 events, 43%), mechanical seal leakage (13 events, 28%) and coupling failure (6 events, 13%) (Figure 6).

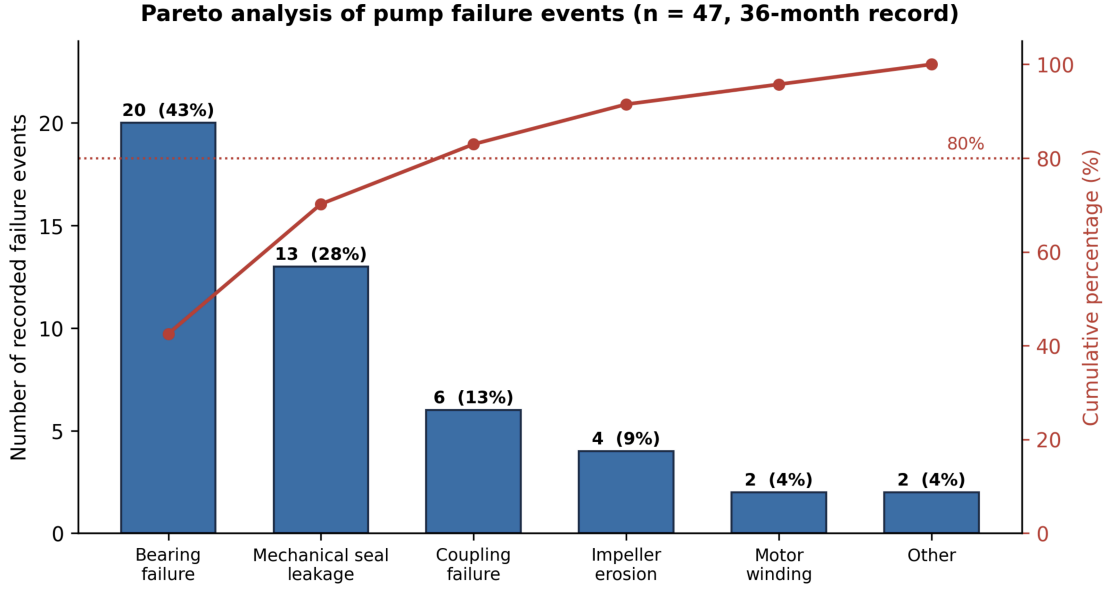


Figure 6. Pareto analysis of pump failure events, 36-month record ($n = 47$). The top three modes account for 83% of recorded events.

A 5-Whys exercise on the bearing mode produced the finding that shaped the whole project. Re-lubrication was governed by a fixed calendar rule set at installation in 2015 and never reviewed. Grease quantity had been taken from the equipment documentation supplied at purchase, which gave a single quantity and interval for every season. Neither the interval basis nor the quantity reflected the bearing temperatures actually reached in summer. The gap between installation and review was eight years, and no procedure existed that would have triggered a review.

The team then did what we now call an operating-envelope audit (Sections 3.6 and 5.5): it set the conditions the rule assumed beside the conditions the bearings actually saw. The plant's rule, as held in its CMMS since installation, was a fixed calendar interval of three

months with 8 g of grease per fitting, taken from the equipment documentation supplied at purchase. That documentation is not reproduced or quoted here, and the audit does not rely on any reference temperature it may contain. The audit took its reference from the grease instead: the 70 °C reference in public bearing-maker guidance applies to the lithium-soap mineral-oil grade the plant was using [12,13,42], and API 610 sets the design requirements for the pump while leaving the lubrication schedule to the vendor's instructions [34]. Applying the test in Section 3.6 to the summer bearing temperatures recorded in 2022 (typical 85 °C, peak 93 °C) gave a thermal gap of about 15 °C in a typical summer and 23 °C at the peak, and correction factors of 0.50 and about 0.35, before any allowance for dust. At about 1640 running hours per interval, each summer interval therefore aged the grease 2.0 to 2.9 times as much as at 70 °C (Section 3.6).

FMEA across the five main component groups ranked bearing overheating highest by a clear margin (RPN 280), ahead of mechanical seal leakage (168), coupling element failure (150), impeller erosion (120) and motor issues (96) (Figure 7). Severity, occurrence and detection were each rated on a 1 to 10 scale by the six-person team working to consensus, using plant history for occurrence and the existing monitoring arrangement for detection. The RPN thresholds used for the risk bands are a working convention adopted by the team, not a standard.

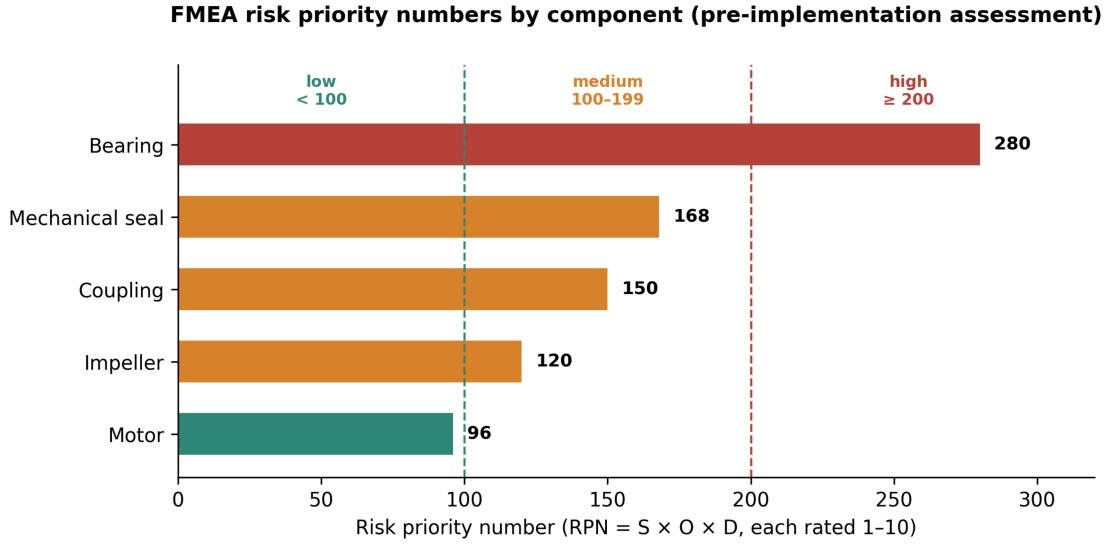


Figure 7. FMEA risk priority numbers by component group, pre-implementation assessment. Band boundaries are a working convention adopted by the project team.

Improve (weeks 19 to 24): Five countermeasures were deployed.

First, and most consequentially, re-lubrication moved from a fixed calendar rule to a running-hours rule of 2000 h, with grease quantity recalculated from bearing geometry and derated for operating temperature according to the bearing manufacturer's high-temperature guidance, and with a change to a higher-temperature grease (Table 3). It is worth being precise about what did and did not change here, because the arithmetic is easy to misread. A quarter is about 2190 h of elapsed time, which at the 75% duty of each pump under the 3 + 1 rotation is about 1640 running hours; 2000 running hours at that duty is roughly 111 calendar days. The new interval is therefore about 22% longer than the old one, not shorter. Section 3.6 shows that it is not supported by the temperature correction at peak summer temperature, and the control plan now adds a summer limit (Table 3). The substantive change was in the basis of the interval, which now tracks actual usage rather than the calendar, and in the quantity and grade of grease applied at each

event. Over-greasing carries its own penalty through churning losses and higher bearing temperature, so the quantity was calculated rather than increased by judgement.

Table 3. Lubrication and bearing monitoring practice before and after the project.

Item	Before (2015 to Mid-2023)	After (from the Improve Phase, 2023)
Interval basis	Calendar, quarterly (about 1640 running hours per pump at 75% duty)	Running hours, 2000 h (about 111 calendar days per pump at 75% duty; about 22% longer than the old rule in running hours)
Summer grease ageing relative to the old rule at 70 °C ($1/f$, Section 3.6)	2.0 (85 °C) to 2.9 (93 °C)	1.2 (85 °C) to 1.8 (93 °C) if the new grade is credited with a 15 °C shift; 2.4 to 3.5 if not
Summer limit (control plan, revised after the envelope test)	None	About 1100 running hours from June to September, or re-lubrication at the 80 °C warning, whichever comes first
Grease quantity per event	Fixed value from the equipment documentation supplied at purchase: 8 g per fitting × 4 fittings = 32 g per pump	Calculated from bearing geometry (SKF G_p = 0.002·D·B) and derated for operating temperature: 18 g per bearing × 3 bearings = 54 g per pump
Grease product	Lithium-soap thickener, mineral base oil International Organization for Standardization viscosity grade (ISO VG) 150, National Lubricating Grease Institute (NLGI) grade 2, upper operating temperature 120 °C	Lithium-complex thickener, polyalphaolefin (PAO) synthetic base oil ISO VG 220, NLGI 2, upper operating temperature 160 °C
Changeover	Not applicable	Compatibility check against the lithium-complex data sheet (compatible thickener family), then purge through the housing drain until the new grease appeared at the drain port; old grease quarantined and labelled
Bearing temperature	Logged in the historian; no alarm limits; not reviewed routinely	Warning 80 °C, trip 95 °C; statistical process control (SPC) chart; monthly key performance indicator (KPI) review
Rule review	No trigger; unchanged since 2015	Annual procedure review; rule owned by the reliability engineer

The grease change raises a fair question: would a high-temperature grease alone have been enough? It would not, because grease life in a rolling bearing shortens as temperature rises for every grade, so the interval still has to follow the operating temperature [12,13]. The grade, the quantity and the basis of the interval changed together, so the case cannot separate their effects, and Section 5.2 treats this as a limit on attribution. The full reasoning, the grease quantity calculation and the changeover procedure are given in Supplementary Material S1, and Table 3 sets the old and new practice side by side.

Second, 5S was implemented in the pump house, where re-lubrication is done, and in the maintenance workshop, where repairs are staged. The project records give three reasons: time lost finding parts and tools once they were released, the risk of mixing two grease grades during the changeover [13], and sand carried into bearings from dirty grease guns, fittings and containers. The full rationale is given in Supplementary Material S1. Third, standard work instructions were written in Arabic and English with visual aids, reflecting the language mix of the maintenance workforce.

Fourth, bearing temperature monitoring changed from a logged value to an alarmed and reviewed one. The Pt100 elements listed in Table 2 had always been logged, but before the project they had no alarm limits and were not reviewed routinely, so a bearing could run hot through a summer without anyone acting on it. The project set a warning at 80 °C

and a trip at 95 °C, added bearing temperature to the monthly KPI review and placed it on a statistical process control chart (Control phase); details of the instrumentation and the alarm logic are in Supplementary Material S1. Fifth, a 40 h technical training programme covered lubrication practice, interpretation of condition-monitoring data and root cause analysis.

The 5S audit scores rose from an average of 35% to 85% across the five pillars (Figure 8). The lowest post-implementation score was Sustain at 78%, which is consistent with the general pattern that the fifth pillar is the hardest to hold.

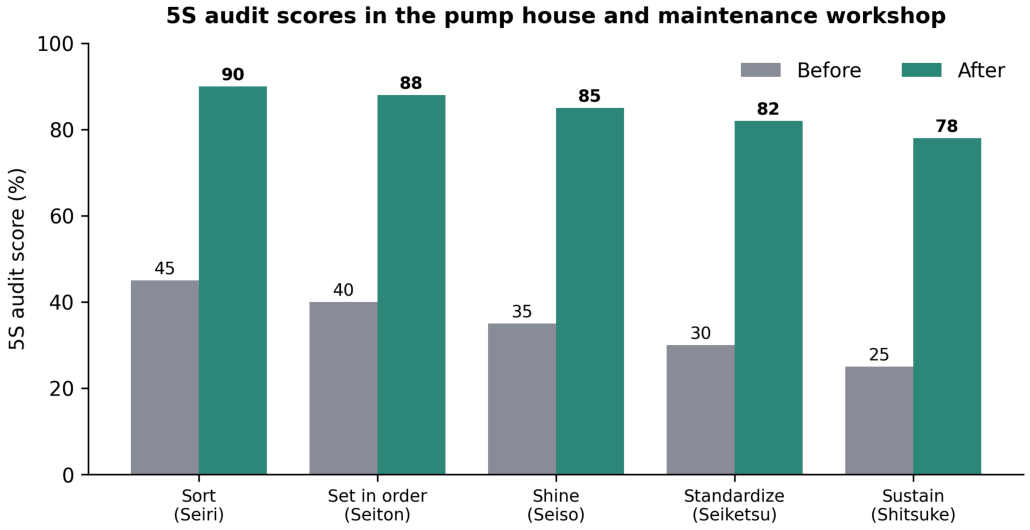


Figure 8. 5S audit scores before and after implementation, by pillar.

Control (weeks 25 to 28, then sustaining): A KPI dashboard with monthly management review was established, along with statistical process control charts on bearing temperature and work-order cycle time, quarterly 5S audits against a standard checklist, and an annual procedure review protocol. The last of these is the direct organisational answer to the eight-year gap found in the Analyse phase. Process ownership sits with the reliability engineer, with the maintenance manager as sponsor.

4.2. Operational Performance

Table 4 compares the two observation periods. MTBF and MTTR are annual values taken from the failure records (Panel A); OEE, operational availability and maintenance process efficiency are summarised from 12 monthly values (Panel B). These figures describe what followed the change; for the reasons given in Section 3.5, they are not estimates of its effect.

Table 4. Performance comparison between the baseline (2022) and post-implementation (2024) observation periods: annual reliability measures from the failure records (Panel A) and monthly measures (Panel B).

KPI	Baseline 2022	Post 2024	Change	Test Statistic	*p*	Effect Size [95% CI]
A. Annual measures from the failure records (exact conditional Poisson test for the failure counts; Welch *t*-test on individual repair times)						
Functional failures (count)	5	4	−20.0%	Exact Poisson	1.00	RR 0.80 [0.16, 3.72]
MTBF (h), 26,280 h ÷ failures	5256	6570	+25.0%	From counts	–	–
MTBF (h), availability-bounded hours ÷ failures	4688	6294	+34.2%	Exact Poisson	0.75	RR 0.75 [0.15, 3.46]
MTTR (h), mean (SD) of repair times	8.5 (1.20), $n = 5$	5.2 (0.91), $n = 4$	−38.8%	−4.70 (7.0)	0.002	–

Table 4. *Cont.*

KPI	Baseline 2022	Post 2024	Change	Test Statistic	*p*	Effect Size [95% CI]
B. Monthly measures, mean (SD) of 12 monthly values (mean difference Δ in percentage points with 95% intervals assuming independence and under lag-1 autocorrelation ρ = 0.5 and 0.7; Cohen's *d* as description only; no *p*-values, see Section 3.5)						
OEE (%)	68.5 (3.8)	82.3 (3.2)	+20.1%	Δ = 13.8	–	ρ = 0: [10.8, 16.8]; ρ = 0.5: [7.7, 19.9]; ρ = 0.7: [0.2, 27.4]; d = 3.93
Availability (%) [a]	89.2 (2.1)	95.8 (1.8)	+7.4%	Δ = 6.6	–	ρ = 0: [4.9, 8.3]; ρ = 0.5: [3.2, 10.0]; ρ = 0.7: [−1.0, 14.2]; d = 3.37
Process efficiency (%) [b]	58.4 (4.2)	74.2 (3.5)	+27.1%	Δ = 15.8	–	ρ = 0: [12.5, 19.1]; ρ = 0.5: [9.1, 22.5]; ρ = 0.7: [0.8, 30.8]; d = 4.09
Planned maint. ratio (%)	78.0	91.0	+16.7%	not tested	–	–

Notes: Panel A: MTBF is fleet running hours (26,280 h) divided by the number of functional failures in the year; RR is the 2024 to 2022 failure-rate ratio with an exact 95% confidence interval from the exact conditional Poisson test; MTTR is compared with a Welch *t*-test on the individual repair times. Panel B: [a] operational availability, including planned downtime. [b] maintenance work-order process efficiency. SD is standard deviation; CI is confidence interval. Panel headings A and B are set in bold. Percentage change is relative to the baseline mean. Cohen's d uses the pooled standard deviation and is given only as a description of the shift [36]. The planned maintenance ratio was recorded as a period total rather than as a monthly series and is therefore reported without a test. Effect sizes of this magnitude are unusual in organisational research and reflect a large shift against small month-to-month dispersion in a single controlled asset set; Section 3.5 explains why they should not be read as confirmatory evidence. The second MTBF row caps running hours at availability × 26,280 h (23,442 h in 2022 and 25,176 h in 2024), the most conservative reading of the availability figures; its exact test uses the unequal exposure. In Panel B, Δ is the post minus baseline mean; intervals under $\rho > 0$ use the effective sample size $n(1-\rho)/(1+\rho)$ [37]. Welch *t*-tests and their *p*-values are not reported for Panel B because they assume independent months.

Against the Define-phase targets, annual MTBF rose 25.0% (target: at least +20%, met by a difference of one failure) and MTTR fell 38.8% (target: at least −25%); OEE passed the 80% threshold, and the planned maintenance ratio reached 91% against a target of above 90%. The fifth target, a 20% reduction in maintenance cost per running hour, could not be evaluated. Cost data were held in a separate financial system on a different cost-centre structure, and the project team was not able to reconcile the two within the study window. This is a real gap rather than a presentational one, and it is carried into the limitations.

Monthly MTBF and MTTR are not computed, because most months contain no failure. The annual comparison is the weakest evidence in the paper. One more failure in 2024 would have left MTBF where it was, the exact test cannot separate the change from chance ($p = 1.00$; rate ratio 0.80, 95% confidence interval (CI) 0.16 to 3.72), and capping running hours at the availability figures changes the size of the change (+34.2%) but not this reading ($p = 0.75$). MTTR is on firmer ground: all four 2024 repairs were shorter than the shortest 2022 repair (Welch t = −4.70, df 7.0, $p = 0.002$), though part of that gain may reflect the added technicians (Section 5.2). Neither measure bears on the mechanism, which rests on the evidence in Section 4.3.

Figure 9 shows the three monthly shifts in their own units, with intervals that widen as the assumed lag-1 autocorrelation rises. Under independence all three intervals sit well clear of zero. At ρ = 0.5 they still do, though they are about twice as wide. At ρ = 0.7, a level that consecutive months sharing equipment state and staff could reach, the effective sample size falls to about two months per period; the intervals for OEE and process efficiency then barely clear zero, and the interval for availability includes it. The monthly shifts are large relative to month-to-month scatter, but how precisely they are known depends on an autocorrelation the data cannot pin down. They describe what followed the change, not its effect.

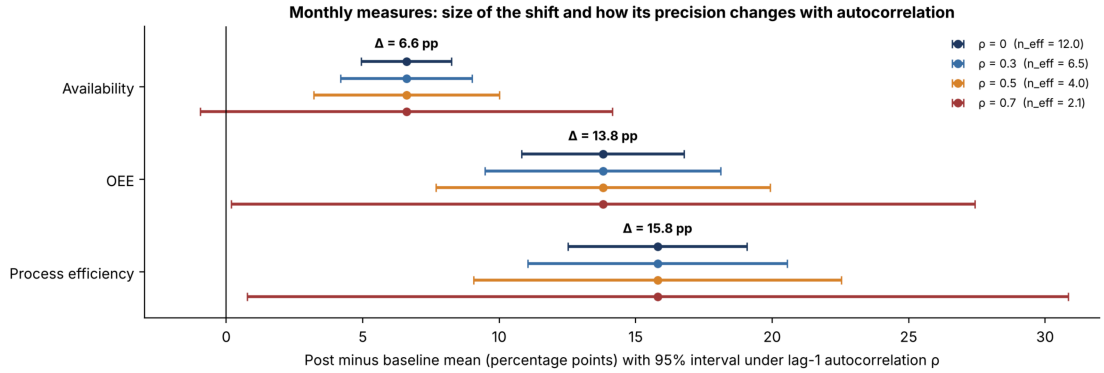


Figure 9. Post minus baseline mean difference for the three monthly measures, with 95% intervals assuming independence ($\rho = 0$) and under lag-1 autocorrelation $\rho = 0.3$, 0.5 and 0.7. n_eff is the effective number of independent months per period.

4.3. Within-Site Check of the Lubrication Mechanism

Table 5 ranks the evidence by how well it resists the rival explanations in Section 5.2, chiefly the CMMS upgrade and the two technicians added in 2023, rather than by statistical significance.

Table 5. Evidence on the lubrication mechanism, ranked by how well it resists rival explanations.

Evidence	Result	What It Can and Cannot Show
1. Envelope test (Section 3.6): summer bearing temperatures against the 70 °C reference	ΔT of 15 °C typical and 23 °C peak; f = 0.50 and 0.35; each summer interval aged the grease 2.0 to 2.9 times as much as at 70 °C	Rests on measured temperatures, the grease type and public guidance, not on failure counts, the 2023 changes or the equipment manual. Shows that the same rule worked the grease two to three times harder in summer than at 70 °C. Does not show that the interval was too long at 70 °C, or what the mismatch cost.
2. Seasonal pattern in the baseline year	8 of 11 bearing events in June to September; exact binomial $p = 0.009$	All events occurred in 2022, before the project, the CMMS upgrade and the staffing change, which therefore cannot explain it. Matches the prediction of the envelope test; chance has no reason to favour summer.
3. Mode-specific change, 2022 to 2024	Bearing events −72.7%, other modes −20.0%; ratio of rate ratios 0.34; Fisher's exact $p = 0.27$	Site-wide changes act on all modes, so a bearing-specific fall points to the lubrication change. Small counts: consistent with the mechanism but not conclusive, and a bad bearing year in 2022 drifting back to normal is not excluded.
4. Before-and-after performance measures	Functional failures 5 to 4; MTTR, OEE and availability improved	Open to every rival explanation in Section 5.2; the failure-count test could not reach significance under any outcome (Section 3.5). Reported as description only.

Notes: Tiers are ordered by how well the evidence resists the rival explanations in Section 5.2, not by statistical significance. Counts are from Table 6.

Table 6 splits the 47 recorded events by mode, year and season. The logic of the check was set out in Section 3.5: the lubrication changes act on bearings, while staffing, the CMMS upgrade, attention and learning act on every mode. Recorded bearing events went from 11 in 2022 to 3 in 2024 (−72.7%), and non-bearing events from 10 to 8 (−20.0%). The ratio of the two rate ratios is 0.34 (Fisher's exact test, $p = 0.27$). A ratio well below one is the pattern P3 predicts. A ratio near one would mean that the site-wide influences did most of the work. With counts this small, the mode comparison is consistent with the mechanism but not conclusive.

The seasonal split bears on P1. In the baseline year, 8 of the 11 bearing events fell between June and September, against the one-third that an even spread across the year

would give (exact binomial test, $p = 0.009$). In 2024 the figure was 1 of 3, and summer bearing events fell from 8 to 1 across the two observation years. The seasonal concentration is the clearest single result in the check, because unlike the mode comparison it does not depend on a difference between two small counts. It is also the pattern the envelope test in Section 3.6 predicts from temperature data alone, and because all 11 events occurred in 2022, the changes made in 2023 cannot explain it.

Table 6. Recorded failure events by mode, year and season ($n = 47$).

Failure Mode	2022	2023	2024	36-Month Total	Change 2022 to 2024
Bearing	11	6	3	20	−72.7%
Mechanical seal	5	4	4	13	−20.0%
Coupling	2	2	2	6	0%
Other	3	3	2	8	−33.3%
Non-bearing total	10	9	8	27	−20.0%
All modes	21	15	11	47	−47.6%
Bearing events, June to September	8	3	1	12	−87.5%

Notes: Fleet running hours were the same in each year (Section 3.4), so counts are directly comparable. Ratio of rate ratios (bearing to non-bearing, 2024 against 2022) = 0.34; Fisher's exact test on the 2 × 2 table of mode by year (bearing 11 and 3, non-bearing 10 and 8), two-sided $p = 0.27$. Share of 2022 bearing events in June to September tested against one-third with an exact binomial test, $p = 0.009$. Row totals match the Pareto counts in Figure 6. The countermeasures entered service around mid-2023 (Figure 4), so the 2023 column is a transition year.

Two cautions apply. The counts are small, so the test has little power, and a non-significant result would not show that the effect is absent. The check also guards only against site-wide explanations; it does not rule out a rival that acts on bearings alone, such as an unusually bad bearing year in 2022 drifting back toward normal. Section 5.2 returns to this point.

The 2023 column should be read as a trend indicator. Bearing and seal events fell in 2023, before the countermeasures had been in service for a full year, and non-bearing events fell steadily from ten to nine to eight. An early decline in an asset chosen for its problems is the usual sign of regression to the mean [41], so the −20% non-bearing fall is not a clean background level. Two features limit how much of the bearing fall regression can carry. If the long-run bearing rate were near the 2023 level of six a year, a year with 11 or more events would occur about once in 23 years (Poisson $p = 0.043$); near the 2022 cool-season rate of about 4.5 a year, about once in 150 years ($p = 0.007$). Also, only three of the 2022 bearing events fell outside June to September, so January to May held at most three, and the cool-season rate (0.38 a month) was already close to the 2024 rate (0.25 a month). The excess of 2022 was a summer excess, which regression to the mean has no reason to produce.

4.4. Resource and Heat-Exposure Consequences: Measured and Derived

Table 7 lists each outcome with the kind of evidence behind it. The pump set had 5 functional failures in 2022 and 4 in 2024, with 42.5 and 20.8 h of corrective repair on them; stores and work-order records add parts, grease and hot-season labour. Energy was not metered and is not claimed.

Three consequences follow, and we state them at the level the evidence supports.

Material consumption: Each avoided bearing failure avoids a bearing set and, in this plant's practice, the seal renewal done at the same opportunity. Stores issued nine bearing sets and seven seals to the pump set in 2022 against three and five in 2024 (Table 7), or 0.34 and 0.27 per 1000 running hours against 0.11 and 0.19. Using catalogue masses for the three bearings in each set (about 15.3 kg [43,44]), the bearing steel replaced fell from about 138 to 46 kg a year, about 176 kg CO_2 a year less at the world crude-steel average of 1.92 t CO_2 per

tonne [45]. This is a screening figure built from counted parts and a published factor, not a life-cycle result. The counterweight is grease. Issued grease rose from 3.4 to 5.8 kg, but the larger shot at a similar frequency accounts for only about 0.2 kg more applied (about 0.51 kg a year under the old 32 g quarterly rule against 0.71 kg at 54 g per 2000 running hours on the three duty pumps); the rest is a standing reserve of the new grade replenished in early 2024. We do not net these flows into one figure, since mass is a poor common unit for steel, elastomer and oil, and a proper balance would need life-cycle data we do not have.

Table 7. Resource and heat-exposure outcomes by type of evidence.

Outcome	Type of Evidence	2022	2024	Source
Functional failures	Counted (failure records)	5	4	CMMS work orders
Corrective repair hours on functional failures	Measured	42.5	20.8	CMMS work orders
Bearing sets issued	Measured	9	3	Stores records
Mechanical seals issued	Measured	7	5	Stores records
Grease issued (kg)	Measured	3.4	5.8	Stores records
Corrective labour hours, June to September	Measured	31	9	Work orders
Bearing steel replaced (kg)	Derived (stores count × catalogue mass)	138	46	Stores records; [43,44]
CO_2 from crude steel for replaced bearings (kg)	Derived (screening factor)	264	88	World average 1.92 t CO_2/t [45]
Specific energy (kWh/m^3)	Not measured	–	–	No sub-metering; not claimed

Worker heat exposure: Corrective work on a failed cooling-water pump cannot be deferred, and most baseline bearing failures fell in the hottest months. Corrective labour on the pump set between June and September went from 31 h in 2022 to 9 h in 2024 (Table 7). Saudi Arabia's midday outdoor work ban, enforced between 15 June and 15 September, reflects the same risk [46,47], but earlier announcements exempted oil and gas operations [48], so for plants like this one the burden falls on plant-level scheduling; field measurements in Al-Ahsa, in the same province, found excessive heat stress over much of the summer work day [49]. Planned work can be moved into cooler hours; unplanned work cannot. The operational risk during repairs, by contrast, was small. While a failed duty pump is under repair the standby is in service, and a second failure would restrict cooling water. At the observed failure rates and repair times, the expected number of such overlaps was about 0.02 a year in 2022 and 0.01 in 2024, so the 3 + 1 arrangement absorbed the failures without making the other pumps run longer than normal.

Direct cost: The cost target could not be evaluated because the cost system could not be reconciled with the asset records (Section 4.2), but a rough estimate narrows the gap. Table 8 combines the measured changes in Table 7 with public list prices where they exist and stated ranges where they do not, carried through a Monte Carlo simulation (200,000 draws). The direct saving is about US$27,000 a year (90% interval 19,000 to 35,000), or about US$1000 per 1000 running hours, almost all of it from bearings and seals. As a check, one avoided functional failure at Bloch's fully burdened average refinery pump repair cost of about US$11,000 [50] accounts for about 40% of that figure; the rest fits the 9 bearing sets drawn in 2022 against 5 functional failures, since many renewals followed degraded-mode events. The estimate leaves out the cost of the project, production effects and changes in planned work, so it cannot be turned into a percentage of maintenance cost per running hour.

Table 8. Rough estimate of the annual direct cost change, 2022 to 2024 (US$).

Item	Measured Change	Unit Price Range	Basis of Price	Annual Effect, Median [90% Interval]
Bearing sets (2 × 6318 C3, 1 × NU 318 ECP)	−6 sets	1377–2754 per set	50–100% of public list price [43,44]	−12,400 [−16,100, −8700]
Mechanical seals (API 682 Category 1 cartridge)	−2	3000–10,000 each	Assumed; no public list price found	−13,000 [−19,300, −6700]
Corrective repair labour	−21.7 h	50–100 per hour	Assumed burdened rate	−1600 [−2100, −1100]
Grease issued from stores	+2.4 kg	15–60 per kg	Assumed, synthetic grade	+90 [+40, +140]
Net direct change	–	–	Monte Carlo, 200,000 draws	−26,900 [−34,600, −19,200]
Net change per 1000 running hours	–	–	26,280 running hours a year	−1020 [−1320, −730]
Top-down check: one avoided functional failure	−1 failure	about 11,000	Average refinery pump repair, 2005 US$ [50]	−11,000
Not included	–	–	Project cost, training time, grease stock premium, production effects	–

Notes: Negative values are savings. Ranges are uniform. Grease is costed on the amount issued from stores, which overstates the cost of the rule itself, since only about 0.2 kg more was applied; costing the applied amount changes the net figure by less than US$100. Plant contract prices would replace these ranges in a full evaluation.

4.5. Qualitative Findings

Three themes recurred across role groups.

Climate as a design assumption nobody had checked: Technicians consistently described the original re-lubrication rule as written for somewhere else. One senior technician put it plainly: the quarterly schedule was designed for a temperate climate, and in practice the bearings were running with inadequate lubrication through the hottest months. What is notable is that this was widely understood on the shop floor well before it was written into a procedure. The project's contribution here was less discovery than authorisation. In the terms of Section 2.5, the knowledge was present; the routine to act on it was not.

Training changed what people looked at: The reliability engineer described technicians becoming more willing to act on early indications from condition-monitoring data rather than waiting for a failure to declare itself. Several technicians linked this specifically to the interpretation module in the training rather than to the monitoring hardware.

Visibility sustained attention: Multiple informants attributed continued resourcing to the monthly review of the KPI dashboard. Their account was pragmatic: what management sees monthly keeps its budget.

5. Discussion

5.1. What the Case Supports

The four propositions fare differently, and it is worth being direct about which the evidence carries.

P1 is supported within the limits of one site. The 5-Whys and the FMEA traced the dominant mode to a rule with a known reference envelope, the site envelope lay well outside it in summer, and baseline bearing events clustered in the hot months (Table 6). It also fits the prediction of the envelope test in Section 3.6, which rests on temperatures rather than on failure counts. Because the 70 °C reference is taken from the grease in use rather than from the equipment manual, this support does not depend on documentation the study cannot publish. A direct count of exceedance hours from the historian was not available (Section 5.7). What the case cannot show is how common such mismatches are; that is the purpose of the plant audit proposed in Section 5.8.

P2 is supported by the qualitative evidence more than by any number. Technicians knew the rule did not suit the climate well before it changed (Section 4.5), and no procedure existed that would have forced a review. The eight-year gap was not a gap in knowledge. It was a gap in the routine that turns knowledge into a changed standard, and DMAIC filled it with a documented root cause, a sponsor and a control plan. The project's contribution was therefore less discovery than authorisation, and we think that is the finding: in adapting maintenance to climate, the scarce resource may be the authority to change an inherited rule rather than the technical insight behind the change.

P3 is the proposition the design tests most directly, because staffing, attention, record-keeping and learning have no reason to make bearing events fall faster than other modes. They did fall faster (Table 6), but with 32 events across the two years the contrast shows direction rather than size ($p = 0.27$), and part of it may be regression to the mean (Section 4.3). The seasonal result is firmer (Table 5, tier 2).

P4 is not tested. There is no variation in operating context inside one site. What the revised framing adds is a moderator that can be measured, the thermal gap between a rule's reference condition and the site condition, so that a multi-site study can test P4 rather than illustrate it (Section 5.5). For this site the gap was about 15 °C at typical summer bearing temperatures and 23 °C at the peak (Section 4.1), which gives such a study a first reference point.

5.2. Alternative Explanations

Several explanations compete with the framework account, and none can be fully excluded.

Concurrent organisational change: During the implementation year the plant upgraded its CMMS and added two maintenance technicians. Both matter. Additional technicians reduce MTTR mechanically, independent of any methodology, and part of the 38.8% reduction in MTTR is plausibly attributable to staffing rather than to standard work. The CMMS upgrade is the more serious threat, because a change in how events are classified would alter the very counts the comparison rests on. The project team confirmed that the failure classification scheme was carried across unchanged and that the 2022 records were migrated rather than re-coded, which reduces but does not remove the concern.

Hawthorne effects: The pump set received sustained attention for a year. Some improvement from observation alone is likely. The within-site check in Section 4.3 is the better guard here, because attention paid to the whole pump set gives no reason for bearing failures to fall faster than seal and coupling failures.

Regression to the mean: The asset was selected partly because it had documented reliability problems, so some reversion toward typical performance would be expected. This is the selection effect and the regression effect working together, and it argues for treating the reported magnitudes as an upper bound. Section 4.3 sets out how far it can go. Regression can plausibly carry part of the fall in non-bearing events and part of the bearing fall, since bearing and seal events were already declining in 2023, but it does not explain why the baseline excess sat in summer [41]. For the same reason the −20% fall in non-bearing events should not be read as a clean background. Bearings, the mode for which the asset drew attention, would be expected to regress further than the other modes, so the bearing-specific contrast in Table 6 is best read as an upper bound on what the lubrication change did.

Learning: Twelve months separated the observation periods. Technicians improve with experience regardless of formal training.

Chance: With five functional failures in 2022 and four in 2024, the annual MTBF comparison rests on nine events, and the exact test ($p = 1.00$; rate ratio 0.80, 95% CI 0.16 to 3.72) cannot separate the change from ordinary year-to-year variation. As Section 3.5 shows, the test could not have separated them even with no failures in 2024. The stronger evidence on

the mechanism is the envelope test and the baseline seasonal pattern (Table 5), neither of which depends on the changes made in 2023.

Bundled countermeasures: The grease grade, the grease quantity and the interval basis changed together, alongside monitoring, 5S, standard work and training. The within-site check can separate bearing-specific from site-wide influences, but it cannot separate the lubrication elements from one another, or from the temperature alarms, which also act on bearings. The improvement belongs to the package. We do not attempt to apportion the change among them.

The check in Section 4.3 bears on these rivals unequally. Staffing, the CMMS upgrade, attention and learning act on all modes, so they cannot by themselves produce a bearing-specific fall, and the check is a real test of them. Regression to the mean is harder to exclude. If 2022 was an unusually bad bearing year, bearing events would fall faster than other modes with no intervention at all. The seasonal pattern helps a little, since chance has no reason to favour summer, and the low cool-season bearing rate of 2022 helps further (Section 4.3), but only a longer baseline or a comparison site could settle it.

Taken together, these suggest the observed magnitudes overstate what a controlled study would find. They do not undermine the mechanism, which has an independent physical basis: re-lubrication intervals and grease quantities that ignore operating temperature will shorten bearing life, and correcting them will lengthen it.

5.3. Comparison with Prior Work

Table 9 sets the case beside published maintenance improvement studies. Definitions differ, so the table is orientation rather than benchmarking. The one figure that can be compared fairly directly is process efficiency, where our gain (58.4% to 74.2%) is smaller in relative terms than in Gomaa's petrochemical shutdown case [7], as a higher starting point would lead one to expect.

Table 9. Comparison with published maintenance improvement cases.

Study	Setting	Approach	Reported Outcomes	Main Source of Improvement
Gomaa [7]	Feed-water pumping station shutdown, petrochemical plant, Egypt	LSS-DMAIC	Non-value-added time −60%; downtime −43%; shutdown efficiency 27% to 49%	Shutdown planning and preparation
Shannon et al. [6]	Active pharmaceutical ingredient plant, Ireland	Design for LSS; TPM and reliability	Planned maintenance −33%; corrective −70%; centrifuge OEE +20%	TPM routines and reliability engineering
Trubetskaya et al. [17]	Dairy processing site, Ireland	Hybrid DMAIC and turnaround model	Overhaul duration −36%	Shutdown scheduling and execution
Köse et al. [51]	Textile plant, Turkey	Axiomatic design of lean autonomous maintenance	Downtime −69.2%; time between failures +65.1%	Operator-based routines
Lucantoni et al. [18]	Manufacturing case	Data-driven TPM framework	OEE +14% (24 kaizen actions)	Data-driven choice of TPM actions
Lucantoni et al. [19]	Automotive company	Rule-based machine learning with plan–do–check–act (PDCA)	Average OEE gain of 5.4% per action (7 actions)	Data-driven choice of improvement actions
This study	Cooling-water pumps, petrochemical plant, Saudi Arabia	LSS-DMAIC with reliability analysis and envelope audit	Functional failures 5 to 4 (MTBF +25.0%); MTTR −38.8%; OEE +20.1%	Correction of a climate-mismatched lubrication rule

Two points stand out from Table 9. The first is size. Köse et al. report a 65.1% rise in the time between failures from a lean autonomous maintenance system in a textile plant [51], and Shannon et al. a 70% cut in corrective maintenance [6]; our 25.0% rise in annual MTBF, a difference of one failure, sits at the modest end and cannot be told apart from chance (Section 3.5); a modest change is plausible for an asset set that already had a working preventive programme. The second concerns the mechanism. The seasonal concentration of bearing failures, and their fall after a change in grease grade, are what bearing tribology would lead one to expect: grease ageing is a thermally driven process that roughly doubles in rate for each 15 °C rise within a grease's working range, and greases built for higher temperature move that range upward [13,42]. The finding is less a surprise than a reminder that a known physical rule had never been set against the plant's procedure. In the petrochemical literature heat is mostly discussed as a threat to plant integrity during extreme events [10]; this case suggests that the same heat also acts quietly, through routine maintenance rules, every summer. With summer extremes around the Gulf projected to grow [11], the gap between inherited rules and site conditions is more likely to widen than to close.

The distinctive element here is not the size of the improvement but its source. In most published cases the improvement comes from better scheduling, better preparation or better use of data. In this case the change centred on the correction of a maintenance rule that had been wrong for the environment since installation. That is a different kind of finding, and a cheaper one to act on. It also comes with a caution that van der Poel and van Waveren make for petrochemical pump fleets: a higher MTBF is not the same thing as a lower total cost of ownership, and pushing MTBF up can cost more than it saves [52]. The rough estimate in Section 4.4 suggests that the parts saved outweigh the grease and labour added by about US$27,000 a year, but without the cost of the project and the plant's own cost data we cannot say where this plant sits on that curve. Bloch's reminder that the true cost of a pump repair includes labour, overhead and administration as well as parts [50] is the reason the figure is given as a range rather than as a saving.

5.4. Implications for Practice

What transfers from one plant is not the size of the improvement. It is a question, a way to answer it and a finding about why it goes unasked. The question can be put to any fixed time-based rule on equipment whose wear depends on temperature: what condition was the rule written for, and how far beyond it does the equipment run in the hottest months? The answer needs only the rule, the lubricant or material it governs and a seasonal temperature record, and it gives a prediction that can be checked against the plant's own failure history before anything is changed. The finding is that the obstacle may be governance rather than knowledge. None of this is specific to pumps or to the Gulf: fans, motors, compressors and gearboxes carry rules of the same kind, and so does any plant that runs hotter or dirtier than its equipment documents assume. Whether the test predicts failures reliably elsewhere is a question this case raises but cannot settle.

Four points seem transferable.

First, audit maintenance rules against local conditions before investing in monitoring technology. The re-lubrication review cost almost nothing and preceded every other gain in this project. Plants in hot climates that have never revisited manufacturer intervals should assume they have a similar finding waiting.

Second, tie the interval basis to usage rather than the calendar where duty cycles vary. The change described in Section 4.1 was as much about the basis as about the number. Usage alone is not enough in a hot climate, though. A running-hours rule still needs a

seasonal limit or a temperature trigger, because the grease ages faster per running hour in summer; the 2000 h rule here had none until the envelope test was applied to it (Section 3.6).

Third, write an expiry date into procedures. The eight-year gap here was not a failure of competence; it was the absence of any mechanism that would prompt a review. The annual procedure review protocol added in the Control phase is the cheapest control in the whole project.

Fourth, treat documentation language as an engineering variable in multilingual workforces. We cannot quantify the effect of bilingual work instructions from this study, and we do not claim one, but technicians consistently raised it unprompted.

5.5. A Framework for Applying the Approach Across Gulf Plants

A single case cannot show how often the finding recurs, but it can be written up as a procedure that other plants can run and that a multi-site study can evaluate. Figure 10 and Table 10 set out that procedure in six stages. It keeps DMAIC as the governance routine and adds the operating-envelope audit as an explicit step, so the question this plant answered almost by accident becomes one that every plant asks on purpose. Stages 1 and 2 use the test in Section 3.6, so a plant can rank its inherited rules by the correction factor before any intervention.

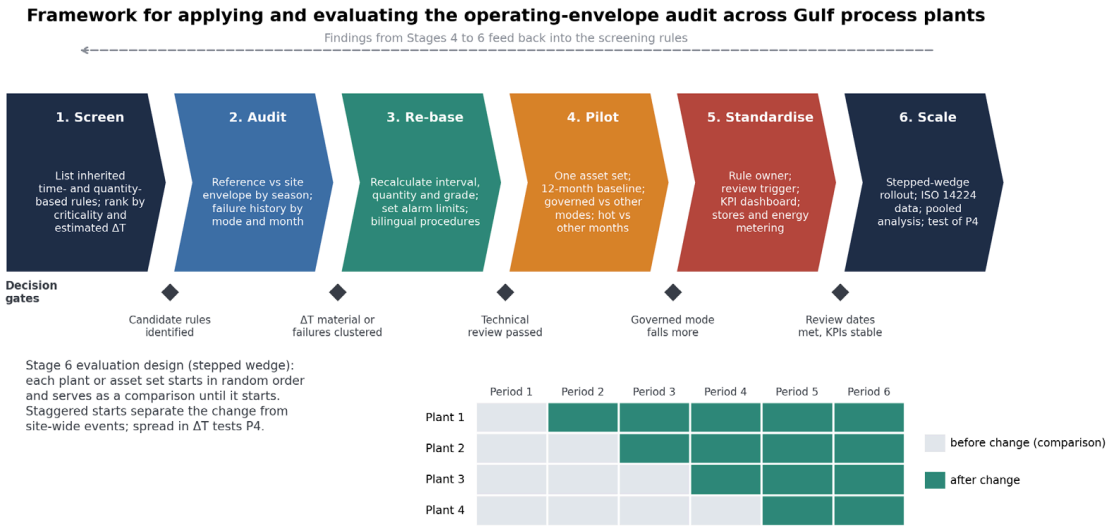


Figure 10. Six-stage framework for applying the operating-envelope audit across Gulf process plants, with decision gates between stages and a stepped-wedge design for the multi-site evaluation in Stage 6.

Table 10 gives the question, inputs, output and decision gate for each stage. Two points need adding to it. Stage 4 uses the within-site check from Section 4.3 as its decision gate, and exceedance hours by season, from the historian or a temperature logger, are a mandatory input at that stage, because they test the mechanism directly rather than through failure counts. Stage 6 pools records with the ISO 14224 taxonomy so that data from different plants can be compared [35].

The scaling stage is also where the causal question can finally be answered. If a group of plants, or several pump sets within a large site, adopt the change in a randomised order, each unit serves as a comparison for the others until its own start date. This is the stepped-wedge design widely used in health services research [53]. It suits industrial rollouts because no unit is denied the change, only asked to wait its turn, and because staggered starts separate the intervention from site-wide events that happen on a single date. With plants that differ in climate, the same data allow a direct test of P4: whether the benefit grows with the thermal gap.

Table 10. Stages of the transfer framework.

Stage	Key Question	Inputs and Tools	Output	Decision Gate
1. Screen	Which inherited rules could be out of envelope?	Rule register from the CMMS; criticality ranking; historian temperatures	Ranked list of candidate rules	Proceed for rules on critical assets with a positive estimated thermal gap
2. Audit	How far does the site envelope depart from the rule's reference envelope?	Manufacturer manuals; bearing and lubricant guidance; seasonal temperature data; failure history by mode and month	Envelope audit sheet per rule; estimate of ΔT	Proceed if ΔT is material or failures cluster in the governed mode or hot season
3. Re-base	What should the rule be under site conditions?	DMAIC Analyse and Improve; re-lubrication calculation; FMEA; envelope test of the revised rule, including a seasonal limit	Revised interval basis, quantity and grade; alarm limits; bilingual standard work	Technical review by the reliability engineer and sponsor
4. Pilot	Did the change act where the mechanism says it should?	At least 12 months of baseline and follow-up; within-site check (governed vs. other modes, hot vs. other months); mandatory: bearing-temperature exceedance hours above the grease reference temperature and above the warning limit, by season, for baseline and follow-up	Pilot report with measured and derived outcomes kept apart	Scale if the governed-mode reduction clearly exceeds that of other modes
5. Standardise	Will the new rule stay correct?	Rule owner; annual review; KPI dashboard; stores and energy metering	Controlled procedure; measured resource flows	Review dates met and KPIs stable
6. Scale	Does the benefit hold across plants, and does it grow with ΔT?	Stepped-wedge rollout; ISO 14224 data; pooled analysis	Multi-site estimate of the effect; test of P4	Continue, revise or stop the procedure

5.6. Sustainability Implications

The sustainability contribution rests on three plant-level indicators tied to plant records, in the direction Franciosi et al. set for measuring the effects of maintenance [27]. Material intensity fell from 0.34 to 0.11 bearing sets and from 0.27 to 0.19 seals per 1000 running hours, about 92 kg of bearing steel and 176 kg CO_2 a year at the crude-steel average [45]. Hot-season corrective labour fell from 31 to 9 h, in a province where outdoor heat stress is well documented [49]. The third, the expected overlap of a repair with a second failure, stayed near zero in both years (Section 4.4). All three can be reported by another plant in the same form. The case is best read as adaptation of existing plants, which will run for decades on rules written for cooler places, more so as Gulf summers grow hotter [11]. The link to SDG targets 8.8, 12.2 and 13.1 [28] is one of direction, not measurement.

Two qualifications remain. The rise in grease use is a real counterweight, and a net material or carbon balance would need life-cycle data for bearings, seals and lubricant that we do not have. Energy was not metered, so no energy saving is claimed; the framework in Section 5.5 builds sub-metering into Stage 5 so that future studies can measure it. The absolute quantities are small for one pump set. Their weight comes from repetition across the many similar sets in a petrochemical plant, which a multi-site study would have to measure rather than assume.

5.7. Limitations

Design: One pump set at one plant, with no comparison group and no randomisation, and the countermeasures changed together. The alternatives in Section 5.2 cannot be excluded, the magnitudes should be read as an upper bound, and the separate effects of grade, quantity, interval, alarms, 5S and training cannot be estimated.

Statistics: Annual counts rest on four or five functional failures, the 12-month windows cut the intervals between failures, so no life distribution is fitted, and the monthly measures are autocorrelated. All tests are descriptive; at $\rho = 0.7$ the interval for availability includes zero (Figure 9).

Cost and environmental flows: The cost estimate in Section 4.4 uses price ranges and leaves out project costs; energy was not metered and no life-cycle balance was attempted.

Interval not derived from the temperature correction: The revised 2000 h interval is longer than the old one and is not supported at peak summer temperature even if the new grease is credited with a 15 °C shift (Section 3.6). The improvement cannot be attributed to a shorter interval, and the summer limit added to the control plan has not yet been observed in service.

Equipment documentation: The equipment manual cannot be reproduced under the confidentiality agreement. The analysis takes its 70 °C reference from the grease, which is documented in Table 3, but what the original rule was meant to assume remains unknown.

No direct exceedance count: The plant historian holds bearing temperatures at 10 s intervals for all 36 months, but that record was not released for this study. Support for the mechanism therefore rests on the envelope test, built from the typical and peak summer temperatures, and on the seasonal pattern of the baseline year, not on counts of hours above 70 °C or 80 °C. A direct count would test P1 and P3 more sharply than the failure counts can, and Stage 4 of the framework now requires it as an input (Table 10).

Twelve-month follow-up: Improvement held for a year. Whether it holds for five is unknown, and the 5S Sustain score of 78%, the lowest of the five pillars, is a reminder that decay is the default.

5.8. Future Research and How to Control the Present Limitations

Each limitation in Section 5.7 points to a specific design change, and Table 11 pairs them. The common thread is that most weaknesses of this study come from studying one asset set after the fact, and most can be fixed by planning the next study before the intervention starts.

Table 11. Limitations of the present study and how future work can control them.

Limitation	Why It Matters	How Future Work Can Control It
Single site and single asset class	No basis for estimating how common the finding is	Multi-site audit of inherited rules; replication on fans, compressors and motors
No comparison group	Site-wide changes cannot be fully excluded	Stepped-wedge rollout; untouched comparison asset sets at the same site
Few functional failures per year	Annual counts rest on a handful of failures, and 12-month windows cut the intervals between failures; low power in the within-site check	Pool data across years and plants with the ISO 14224 taxonomy; count-based rate ratios with exact intervals; hierarchical reliability models; fit life distributions only to multi-year records
Autocorrelated monthly series	Monthly tests overstate precision	At least 24 months per segment; interrupted time-series with autocorrelated errors
Bundled countermeasures	Effects of grade, interval and alarms cannot be separated	Introduce each element on a different pump set or at a different time

Table 11. *Cont.*

Limitation	Why It Matters	How Future Work Can Control It
Interval not derived from the temperature correction	Seasonal adequacy of the 2000 h rule is unknown	Run the summer limit for at least two seasons; log exceedance hours and grease condition at each re-lubrication
No direct exceedance count	Mechanism inferred from typical and peak temperatures and the timing of failures	Record exceedance hours by season and year before and after the change, from the historian or a temperature logger
Incomplete cost data	Economic case and total cost of ownership unknown	Map cost centres to asset tags before the project; track cost of ownership alongside MTBF
Energy not metered	Energy effect remains an expectation	Sub-meter the pump set; log flow and head to compute specific energy
Twelve-month follow-up	Decay of the gain is unknown	Follow up for three years or more; track 5S Sustain scores and review compliance

Of the pairings in Table 11, a stepped-wedge rollout across plants with different thermal environments matters most, because it would supply the missing comparison and test P4. A systematic audit of how often Gulf plants revisit manufacturer maintenance rules would show whether this case is unusual or typical; the interviews suggest typical, but that is a hypothesis rather than a result.

6. Conclusions

This study asked why a maintenance rule that did not fit an extreme climate stayed in force for eight years, and what a DMAIC project needs in order to catch such a rule. One cooling-water pump set supports three conclusions. The dominant failure mode, bearing distress, traced to a quarterly re-lubrication rule applied with a mineral-oil grease that, by the bearing makers' own criterion, aged two to three times as fast in summer as at its 70 °C reference, and eight of eleven baseline bearing events fell in June to September ($p = 0.009$). The rule persisted for want of a routine that could authorise its revision, not for want of knowledge, and DMAIC supplied that routine. After the grade, quantity and interval basis were changed, bearing events fell further than other modes, a pattern consistent with the mechanism but resting on small counts.

The study cannot show how large the benefit is or how often such rules go unrevised. What it offers other plants is a question to put to any fixed time-based rule on temperature-sensitive equipment, and a way to answer it from data most plants already hold, with a prediction that can be checked against the plant's own failure history before anything is changed. Whether that prediction holds elsewhere is untested; the framework in Section 5.5 is the way to test it.

For petrochemical operations, the gains were in parts and hot work rather than in process capacity. Bearing sets drawn from stores fell from 9 to 3 and summer corrective labour from 31 to 9 h, while the 3 + 1 arrangement kept the risk to cooling-water supply small in both years. The larger value lies in the many grease-lubricated machines a plant runs on rules written elsewhere, and in services with thinner redundancy.

The revised 2000 h interval is not supported at peak summer temperature, so a summer limit has been added and still needs to be observed in service.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/su182111159/s1: Supplementary Material S1: Detailed account of the DMAIC countermeasures.

Author Contributions: Conceptualization, W.M.S. and A.M.A.; methodology, W.M.S.; formal analysis, W.M.S.; investigation, A.M.A. and W.M.S.; resources, A.M.A.; data curation, A.M.A.; writing—original draft preparation, W.M.S.; writing—review and editing, A.M.A. and W.M.S.; visualization, W.M.S.; supervision, W.M.S.; project administration, W.M.S.; funding acquisition, A.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deanship of Graduate Studies and Scientific Research, Jazan University, Saudi Arabia, under project number JU-RSP2026.

Institutional Review Board Statement: Ethical review and approval were not required for this study. The study analyses equipment maintenance records, and its only human contact was voluntary, anonymised interviews with adult employees about maintenance procedures and work practices; no health information was collected, and names and other identifying details were removed before analysis. Such work falls outside the definition of research in the Law of Ethics of Research on Living Creatures (Royal Decree No. M/59, Kingdom of Saudi Arabia).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The participating organisation authorised publication of the data and technical details reported in this article. Aggregated data supporting the results are available from the corresponding author on reasonable request, subject to the confidentiality agreement with the participating organisation. Raw CMMS extracts, interview transcripts and the manufacturer's documentation cannot be shared.

Acknowledgments: The authors thank the participating organisation for access to its records and staff, and for authorising publication of the data reported here.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kumar, S.; Swarnakar, V.; Phanden, R.K.; Khanduja, D.; Chakraborty, A. Role of Lean Six Sigma in manufacturing setting: A systematic literature review and agenda for future research. *TQM J*. **2024**, *36*, 1996–2047. [CrossRef]
2. Introna, V.; Santolamazza, A. Strategic maintenance planning in the digital era: A hybrid approach merging Reliability-Centered Maintenance with digitalization opportunities. *Oper. Manag. Res*. **2024**, *17*, 1397–1420. [CrossRef]
3. Skalli, D.; Charkaoui, A.; Cherrafi, A.; Garza-Reyes, J.A.; Antony, J.; Shokri, A. Industry 4.0 and Lean Six Sigma integration in manufacturing: A literature review, an integrated framework and proposed research perspectives. *Qual. Manag. J*. **2023**, *30*, 16–40. [CrossRef]
4. Nowlan, F.S.; Heap, H.F. *Reliability-Centered Maintenance*; Report AD-A066579; United Air Lines/Office of the Assistant Secretary of Defense: Washington, DC, USA, 1978.
5. Moubray, J. *Reliability-Centered Maintenance*, 2nd ed.; Industrial Press: New York, NY, USA, 1997.
6. Shannon, N.; Trubetskaya, A.; Iqbal, J.; McDermott, O. A total productive maintenance and reliability framework for an active pharmaceutical ingredient plant utilising design for Lean Six Sigma. *Heliyon* **2023**, *9*, e20516. [CrossRef] [PubMed]
7. Gomaa, A.H. Enhancing shutdown maintenance performance using Lean Six Sigma: A case study. *Int. J. Lean Six Sigma* **2026**, *17*, 743–781. [CrossRef]
8. Saudi Press Agency. Ministry of Industry and Mineral Resources: Number of Factories Increased by 10% in 2023. 26 February 2024. Available online: https://www.spa.gov.sa/en/N2054779 (accessed on 4 September 2026).
9. Arab News. Saudi Vision 2030 Boosts Industrial Facilities by 60%, Official Figures Reveal. 27 August 2024. Available online: https://www.arabnews.com/node/2569152/business-economy (accessed on 4 September 2026).
10. Cruz, A.M.; Krausmann, E. Vulnerability of the oil and gas sector to climate change and extreme weather events. *Clim. Change* **2013**, *121*, 41–53. [CrossRef]
11. Pal, J.S.; Eltahir, E.A.B. Future temperature in southwest Asia projected to exceed a threshold for human adaptability. *Nat. Clim. Change* **2016**, *6*, 197–200. [CrossRef]
12. SKF Group. *Rolling Bearings*; PUB BU/P1 17000/1 EN; SKF: Gothenburg, Sweden, 2018.
13. Lugt, P.M. A review on grease lubrication in rolling bearings. *Tribol. Trans*. **2009**, *52*, 470–480. [CrossRef]
14. Albliwi, S.; Antony, J.; Abdul Halim Lim, S.; van der Wiele, T. Critical failure factors of Lean Six Sigma: A systematic literature review. *Int. J. Qual. Reliab. Manag*. **2014**, *31*, 1012–1030. [CrossRef]
15. Pongboonchai-Empl, T.; Antony, J.; Garza-Reyes, J.A.; Komkowski, T.; Tortorella, G.L. Integration of Industry 4.0 technologies into Lean Six Sigma DMAIC: A systematic review. *Prod. Plan. Control* **2024**, *35*, 1403–1428. [CrossRef]

16. Pongboonchai-Empl, T.; Antony, J.; Garza-Reyes, J.A.; Tortorella, G.L.; Komkowski, T.; Stemann, D. DMAIC 4.0: Innovating the Lean Six Sigma methodology with Industry 4.0 technologies. *Prod. Plan. Control* **2026**, *37*, 170–191. [CrossRef]
17. Trubetskaya, A.; Ryan, A.; Powell, D.J.; Moore, C. Utilising a hybrid DMAIC/TAM model to optimise annual maintenance shutdown performance in the dairy industry: A case study. *Int. J. Lean Six Sigma* **2024**, *15*, 70–92. [CrossRef]
18. Lucantoni, L.; Antomarioni, S.; Ciarapica, F.E.; Bevilacqua, M. A data-driven framework for supporting the total productive maintenance strategy. *Expert Syst. Appl.* **2025**, *268*, 126283. [CrossRef]
19. Lucantoni, L.; Antomarioni, S.; Ciarapica, F.E.; Bevilacqua, M. A rule-based machine learning methodology for the proactive improvement of OEE: A real case study. *Int. J. Qual. Reliab. Manag.* **2024**, *41*, 1356–1376. [CrossRef]
20. Ibikunle, A.K.; Rajemi, M.F.; Mohd Zahari, F.; Faqera, A.F.O.; Ali, G.A. Factors hindering lean manufacturing and six sigma implementation within manufacturing SMEs: Prisma-based approach for future research directions. *J. Asia Bus. Stud.* **2025**, *19*, 511–526. [CrossRef]
21. Werbińska-Wojciechowska, S.; Winiarska, K. Maintenance performance in the age of Industry 4.0: A bibliometric performance analysis and a systematic literature review. *Sensors* **2023**, *23*, 1409. [CrossRef] [PubMed]
22. Carlson, C.S. *Effective FMEAs: Achieving Safe, Reliable, and Economical Products and Processes Using Failure Mode and Effects Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2012. [CrossRef]
23. Crespo Márquez, A.; Moreu de León, P.; Gómez Fernández, J.F.; Parra Márquez, C.; López Campos, M. The maintenance management framework: A practical view to maintenance management. *J. Qual. Maint. Eng.* **2009**, *15*, 167–178. [CrossRef]
24. Marques, P.; Conceição, L.; Carvalho, A.M.; Reis, J. Driving sustainable operations: Aligning Lean Six Sigma practices with sustainability goals. *Sustainability* **2025**, *17*, 8898. [CrossRef]
25. Hariyani, D.; Mishra, S. An analysis of drivers for the adoption of integrated sustainable-green-lean-six sigma-agile manufacturing system (ISGLSAMS) in Indian manufacturing industries. *Benchmarking Int. J.* **2023**, *30*, 1073–1109. [CrossRef]
26. Sakib, M.N.; Kawsar, M.; Bithee, M.M. Continuous improvement through Lean Six Sigma: A systematic literature review and bibliometric analysis. *Int. J. Lean Six Sigma* **2025**, *16*, 1252–1275. [CrossRef]
27. Franciosi, C.; Voisin, A.; Miranda, S.; Riemma, S.; Iung, B. Measuring maintenance impacts on sustainability of manufacturing industries: From a systematic literature review to a framework proposal. *J. Clean. Prod.* **2020**, *260*, 121065. [CrossRef]
28. United Nations, Department of Economic and Social Affairs. The 17 Goals (Goal 8, Target 8.8; Goal 12, Target 12.2; Goal 13, Target 13.1). Available online: https://sdgs.un.org/goals (accessed on 7 October 2026).
29. Teece, D.J.; Pisano, G.; Shuen, A. Dynamic capabilities and strategic management. *Strateg. Manag. J.* **1997**, *18*, 509–533. [CrossRef]
30. Reason, J. *Managing the Risks of Organizational Accidents*; Ashgate: Aldershot, UK, 1997.
31. Yin, R.K. *Case Study Research and Applications: Design and Methods*, 6th ed.; Sage Publications: Los Angeles, CA, USA, 2018.
32. Flyvbjerg, B. Five misunderstandings about case-study research. *Qual. Inq.* **2006**, *12*, 219–245. [CrossRef]
33. Siggelkow, N. Persuasion with case studies. *Acad. Manag. J.* **2007**, *50*, 20–24. [CrossRef]
34. API Standard 610. *Centrifugal Pumps for Petroleum, Petrochemical and Natural Gas Industries*, 11th ed.; American Petroleum Institute: Washington, DC, USA, 2010.
35. *ISO 14224:2016*; Petroleum, Petrochemical and Natural Gas Industries: Collection and Exchange of Reliability and Maintenance Data for Equipment, 3rd ed. International Organization for Standardization: Geneva, Switzerland, 2016. Available online: https://www.iso.org/standard/64076.html (accessed on 4 September 2026).
36. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988.
37. Zwiers, F.W.; von Storch, H. Taking serial correlation into account in tests of the mean. *J. Clim.* **1995**, *8*, 336–351. [CrossRef]
38. Bernal, J.L.; Cummins, S.; Gasparrini, A. Interrupted time series regression for the evaluation of public health interventions: A tutorial. *Int. J. Epidemiol.* **2017**, *46*, 348–355. [CrossRef] [PubMed]
39. Shadish, W.R.; Cook, T.D.; Campbell, D.T. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*; Houghton Mifflin: Boston, MA, USA, 2002.
40. Coryn, C.L.S.; Hobson, K.A. Using nonequivalent dependent variables to reduce internal validity threats in quasi-experiments: Rationale, history, and examples from practice. *New Dir. Eval.* **2011**, *2011*, 31–39. [CrossRef]
41. Barnett, A.G.; van der Pols, J.C.; Dobson, A.J. Regression to the mean: What it is and how to deal with it. *Int. J. Epidemiol.* **2005**, *34*, 215–220. [CrossRef] [PubMed]
42. SKF Evolution. Grease Life in Lubricated-for-Life Deep Groove Ball Bearings. 26 May 2004. Available online: https://evolution.skf.com/grease-life-in-lubricated-for-life-deep-groove-ball-bearings/ (accessed on 24 October 2026).
43. Motion Industries. SKF 6318/C3 Radial/Deep Groove Ball Bearing: Specifications and List Price. Available online: https://www.motion.com/products/sku/00058091 (accessed on 24 October 2026).
44. Motion Industries. SKF NU 318 ECP Cylindrical Roller Bearing: Specifications and List Price. Available online: https://www.motion.com/products/sku/00141733 (accessed on 24 October 2026).
45. World Steel Association. *World Steel in Figures 2026*; worldsteel: Brussels, Belgium, 2026. Available online: https://worldsteel.org/wp-content/uploads/World-Steel-in-Figures-2026.pdf (accessed on 24 October 2026).

46. Saudi Press Agency. Human Resources Ministry to Enforce Midday Outdoor Work Ban Starting June 15. 9 June 2025. Available online: https://www.spa.gov.sa/en/N2337578 (accessed on 4 September 2026).
47. Saudi Press Agency. Midday Outdoor Work Ban Begins June 15. 15 June 2026. Available online: https://www.spa.gov.sa/en/N2612867 (accessed on 4 September 2026).
48. Argaam. Midday Work Break to Begin in Saudi Arabia from June 15. 11 June 2019. Available online: https://www.argaam.com/en/article/articledetail/id/613169 (accessed on 21 September 2026).
49. Al-Bouwarthan, M.; Quinn, M.M.; Kriebel, D.; Wegman, D.H. Assessment of heat stress exposure among construction workers in the hot desert climate of Saudi Arabia. *Ann. Work Expo. Health* **2019**, *63*, 505–520. [CrossRef] [PubMed]
50. Bloch, H.P. Pump Life Expectancy: Myths, Math, and Meaningful Metrics. *Reliability Magazine*. Available online: https://reliamag.com/articles/pump-life-expectancy/ (accessed on 24 October 2026).
51. Köse, Y.; Muftuoglu, S.; Cevikcan, E.; Durmusoglu, M.B. Axiomatic design for lean autonomous maintenance system: An application from textile industry. *Int. J. Lean Six Sigma* **2023**, *14*, 555–587. [CrossRef]
52. van der Poel, J.; van Waveren, C.C. Managing centrifugal pump reliability in the context of total cost of ownership. *S. Afr. J. Ind. Eng.* **2007**, *18*, 57–70. [CrossRef]
53. Hemming, K.; Haines, T.P.; Chilton, P.J.; Girling, A.J.; Lilford, R.J. The stepped wedge cluster randomised trial: Rationale, design, analysis, and reporting. *BMJ* **2015**, *350*, h391. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.