Chapter 5

Reducing costs by clustering maintenance activities for multiple critical units

Abstract. Advances in sensor technology have enabled companies to make significant progress towards achieving condition-based maintenance (CBM). CBM provides the opportunity to perform maintenance actions more effectively. However, scheduling maintenance at the unit level may imply a high maintenance frequency at the asset level, which can be costly and undesirable for safety reasons. In this paper, we consider systems consisting of multiple critical units for which a strict and conservative maintenance strategy is enforced. Although this implies that benefits cannot be obtained by delaying maintenance activities, the clustering of them can be beneficial. We consider two simple, practical systems for condition monitoring that involve either one signal (alarm) or two signals (alert, alarm). Our analysis and results provide general insights into when and how to cluster maintenance operations, with the objective of minimizing the total maintenance costs. Moreover, they show that clustering is essential for a broad range of circumstances, including those at a considered real-life case of equipment maintenance at Europe’s largest gas field.

This chapter is based on De Jonge et al. (2016):
5.1 Introduction

Recent advances in sensor technology have delivered important opportunities for condition monitoring of industrial systems and condition-based maintenance. Condition monitoring is defined as “the collection of real-time sensor information from a functioning device in order to monitor and learn about the condition of that device” (Li and Ryan, 2011). Condition-based maintenance is preventive maintenance based on condition monitoring (Han and Song, 2003). The main advantage of condition-based maintenance compared to time-based maintenance is that it takes conditions of equipment into account, and thereby leads to more effectively planned maintenance actions and improved utilization of the lifespan of a unit. For single-unit systems, the benefits are obvious and one should strictly adhere to a policy of condition-based maintenance without hesitation. However, not all business cases are as straightforward as this.

Many real-life systems contain multiple units subject to deterioration. Applying CBM for each unit separately minimizes the maintenance frequency at the unit level, but also implies many (small) maintenance operations at the asset level, which may not be cost effective. Units are often economically dependent (Cho and Parlar, 1991) since the costs of a maintenance action on one or more units consists of a fixed part (e.g. fixed costs associated with work preparation, transport, scaffolding, safety precautions and possible plant shutdowns) and a variable part (e.g. man-hours, materials and safety precautions at unit-level). If the fixed part is considerable, then clustering maintenance operations of various units may be preferable to minimize the total maintenance costs at the asset level.

As we explain in more detail in Section 5.2, existing studies on multi-component systems with condition monitoring focus on rather complex models and policies. The focus is mainly on methods and algorithms to solve these models, their computational performance, and proposed approximations. In general, only a few examples are considered, and insights on how the optimal maintenance strategy depends on the various characteristics of a situation are lacking. Instead, we focus on simple settings and policies that are easy to implement, and derive general insights on the effects of the number of units, the cost structure, and the mean times until signals.

Our approach is to consider a system that consists of a number of identical critical
units. Such systems are widespread in practice in for example the process industry, the maritime industry, and the transport industry (Bouvard et al., 2011; Camci, 2009; Yang et al., 2008), and it applies to the real-life case discussed in Section 5.7. Each unit contains a sensor that provides either one signal (alarm) or two signals (alert, alarm). These one and two-level signal systems are based on the P-F interval introduced by Moubray (1997) and are very common in industry. Some failures/systems only have a single symptom that can be detected (alarm), but for others two symptoms with different P-F intervals might occur. An example of the latter is a bearing that shows increased vibration levels about 6 months before failure, while a month prior to failure an increased heat generation could be detected (Blann, 2013).

For both systems, which we will refer to as alarm-maintain and alert-alarm-maintain, an operational guideline in place enforces a maintenance action within a fixed time period after an alarm in order to prevent an impending failure. Because of the criticality of the units benefits cannot be obtained by delaying maintenance activities. However, it can be beneficial to cluster maintenance activities. In this paper, the cost performance of two clustering policies will be evaluated and compared with the policy that does not cluster maintenance activities.

The remainder of this paper is organized as follows. In Section 5.2, we discuss the existing literature on multi-unit systems with condition monitoring. In Section 3, we formally describe the problem we consider. In Sections 5.4 and 5.5, we analyze the alarm-maintain and the alert-alarm-maintain system, respectively, under the assumption that the time until an alarm (and also until an alert in the latter system) is exponentially distributed. In Section 5.6, we relax this assumption to check the robustness of our results under different time-to-signal distributions. We discuss our results and implications for a real-life case of equipment maintenance at the Groningen gas field in Section 5.7. We end in Section 5.8 with conclusions and directions for future research.

### 5.2 Literature review

In this section we review studies that consider multi-unit systems with condition monitoring. For broader reviews on multi-component systems we refer to Cho and Parlar (1991), Dekker et al. (1997), Nicolai and Dekker (2008), and Wang (2002).
Anisimov and Güler (2003) and Gürler and Kaya (2002) consider a system with identical components with operating times consisting of a number of sequential phases with exponential durations. Anisimov and Güler (2003) propose a grouping policy that replaces all components if the fraction of components that is in the last state before failure exceeds a certain threshold. The focus is on various methods (exact and approximate) to optimize this fraction. The maintenance policy adopted by Gürler and Kaya (2002) either performs preventive or corrective maintenance on single units, or replaces the entire system if at least $N$ components are beyond a certain deterioration state $K$. Approximations are proposed to evaluate the policy. Only limited insights on the performance of the grouping policy for varying characteristics of the model are provided by these studies.

Other studies limit themselves to systems with a fixed and small number of units. Castanier et al. (2005) consider systems with two units in series that deteriorate gradually and that are monitored by non-periodic inspections. Set-up costs are saved when inspections or replacements are combined. Koochaki et al. (2012) analyze a serial production system consisting of three components, each following either a condition-based or an age-based maintenance policy. The effect of maintenance clustering on maintenance costs and line productivity are considered for a few specific system configurations. Tian and Liao (2011) consider multi-component systems with failure rates that are described as proportional hazards models. The components are economically dependent and an advanced policy is proposed that dictates the grouping of preventive replacements. Two examples, respectively with two and three components, are considered.

Stochastic dependence between the components is considered in two studies. Hong et al. (2014) investigate the influence of dependent stochastic degradation of multiple components on the optimal maintenance decisions. Emphasis is on the copula that are used to model the dependent stochastic deterioration of the components. Song et al. (2014) consider systems where each individual component may fail due to two competing statistically dependent failure modes, and the failure processes among the components are also statistically dependent. Reliability analysis is performed for two specific series systems and one specific parallel system.

Some studies mainly focus on the computational performance of their algorithms. Barata et al. (2002) consider a rather comprehensive model to optimize maintenance decisions for multi-component systems. Emphasis is on the optimization procedure
Reducing costs by clustering maintenance activities for multiple critical units

and the required computing times. System downtime for a two-component series system is considered for two settings of the parameters. Zhou et al. (2013) develop an algorithm for maintenance optimization of series-parallel systems with multi-state economically dependent components. Stochastic ordering theory is used to reduce the search space and improve the computational efficiency. A single system configuration is considered, with the main focus on the computational performance of the algorithm.

More advanced algorithms are proposed by the following authors. Van Horenbeek and Pintelon (2013) consider a system with components that are economically and structurally interdependent. The entire system has to be stopped and a setup cost has to be paid if preventive or corrective maintenance is performed on one or more components. A rather complex policy is proposed and applied to a single numerical example. Bouvard et al. (2011) aim to group maintenance actions for components of a system to reduce set-up costs. An advanced algorithm is proposed that consists of various steps. A few numerical experiments are performed, but these do not provide general insights. Marseguerra et al. (2002) consider a continuously monitored multi-component system and use a Genetic Algorithm for determining the optimal degradation level beyond which preventive maintenance has to be performed. Clustering of maintenance activities is not taken into account.

Wijnmalen and Hontelez (1997) consider a rather complicated system with various component types. The repair cost depends on the condition of the component. The system setup cost is reduced when components of various types are repaired simultaneously; the type setup cost is reduced when components of the same type are repaired simultaneously. A heuristic approach based on a decomposition of the multi-component problem into several single-component Markov decision problems is presented. The focus is mainly on the performance of the heuristics, general insights are not provided.

There are also studies that take a broader perspective than just maintenance. Hsu (1991) consider a serial production system consisting of a number of stations, where parts must proceed through the stations. Each station is maintained when it breaks down or when a certain number of parts has been processed. The joint performance of the stations is studied, but clustering of maintenance actions is not considered. Xu et al. (2012) jointly consider multi-component safety related systems and the production systems that are being protected. An extensive model is proposed and a specific
illustrative example of a high integrity pressure protection system is analyzed.

Finally, Van Der Duyn Schouten and Vanneste (1993) consider the grouping of maintenance actions for a set of identical components based on their age. Instead of the exact ages of the components, three age groups are considered. In an approximated model that uses exponential distributions for the durations that a component is in a certain group, analytical results are derived. These are compared with the outcome of a simulation that uses the exact lifetime distribution of the components. Although age-based maintenance is considered, the approximated model with exponential durations has similarities with the condition-based maintenance problem with a finite number of deterioration states that we consider. The study is mainly about the difference between the outcomes of the approximated model and of the simulation. The influence of model parameters is not considered, and no optimization of the decision variables is performed. Furthermore, only one grouping rule is considered, namely to replace the entire system.

Rather than analyzing complex policies that rely on detailed (continuous) condition information, we consider systems with just one (alarm) or two (alert, alarm) signals. Studying these simple, practical policies allows us to find explicit solutions. Instead of optimizing a few specific cases, we define the problem in a generic sense. The structured analysis of the sensitivity of the results for changes in the relevant parameters (costs, number of units, failure rate and response time) yields useful and general insights that can provide decision support in a wide range of applications. We illustrate for a real-life case that the benefits of applying the proposed multi-unit clustering CBM policies can be considerable.

5.3 General problem description

We consider a system that consists of \( n \) identical critical units with continuous degradation processes that are operated in the same way. Such systems are widespread in various industries and provide a natural starting point. As for example in Barlow and Hunter (1960) and McKone and Weiss (2002), the objective is to minimize the expected maintenance cost per unit time. We assume that maintenance activities require a negligible amount of time and make the units as-good-as-new.

We analyze two versions of the system that differ in its deterioration signaling. The first is called the alarm-maintain system and is based on the P-F curve intro-
Reducing costs by clustering maintenance activities for multiple critical units

duced by Moubray (1997), see Figure 5.1. When the sensor records that the measured parameter exceeds a certain threshold value, it gives an alarm signal (P in the P-F curve). The time between the moment a potential failure becomes detectable and the moment where it deteriorates into a functional failure (F) is called the P-F interval. As Mobley points out, the P-F interval is in general not constant. In Figure 5.1, \( F_1 \) and \( F_2 \) respectively denote the earliest and the latest point in time at which a functional failure can be expected. In our systems, maintenance should be performed within a period with constant length \( D \) after the alarm signal, and because of the criticality of the units, \( D \) is chosen smaller than the shortest P-F interval (i.e. smaller than \( F_1 - P \)).

![Figure 5.1: P-F curve and interval (Moubray, 1997).](image)

Our second version of the system is the alert-alarm-maintain system. In this system, an additional deterioration state is introduced. The sensors now use two threshold values for the measured parameter (combination). At the lowest level, it produces an alert signal, and at the next level it raises an alarm. Again, maintenance should be performed within a period with constant length \( D \) after the alarm signal. In both versions of the system the alarm level should be chosen such that the time \( D \) between an alarm and maintenance is larger than the planning time required to perform a maintenance activity.

The maintenance costs in our systems consist of a fixed and a variable part. We denote the fixed cost of performing a maintenance action by \( C \), and the variable cost per unit on which maintenance is performed by \( c \). Furthermore, we let \( \bar{C} = C/c \) denote the relative fixed cost of performing maintenance. Obviously, clustering be-
comes more attractive as $C$ (and thereby $\bar{C}$) increases. We consider two clustering policies. The Alarm (Alert) Clustering policy maintains all units that have given an Alarm (Alert) signal if some unit must be maintained (because it reaches the end of the period with length $D$). These are compared with the policy without coordination of maintenance activities for the different units, which we will call No Clustering. We remark that it is appropriate to use these simple clustering rules as all units are identical. For systems with non-identical units, more complex rules with unit-dependent clustering thresholds that are possibly dependent on the unit that initiates the maintenance action should be considered.

We assume that a condition monitoring technique is available, e.g. vibration monitoring or oil debris monitoring, that provides the alert and alarm signals based on predefined thresholds. The random distributions that determine the time that a component is in a certain stage could be estimated based on monitoring data or on opinions of experts, but this is not within the scope of our study. Instead, we start our investigation in Sections 5.4 and 5.5 with exponentially distributed times until the alarm (alert) signals, i.e., with constant alarm (alert) rates. This will allow us to obtain analytical insights into the benefits of clustering. We remark that although the individual rates until alert and alarm signals are constant in this setting, the total time to failure bears the common property of an increasing failure rate. To test the robustness of the insights obtained in Sections 5.4 and 5.5, we study more general gamma durations in Section 5.6. Future studies could base the distributions of the time until a signal on theoretical or empirical deterioration processes, and eventually optimize over the threshold levels.

5.4 Alarm-Maintain with constant alarm rate

In our study of the alarm-maintain system we let $\lambda$ denote the alarm rate (for each of the $n$ units). Under the cost structure described in the previous section, we will compare the policies No Clustering and Alarm Clustering. These policies were also described in the previous section.

**No Clustering.** The cost of each maintenance action is $C + c$ and the mean time between two maintenance actions on a specific unit equals $\lambda^{-1} + D$. This implies that the mean cost per unit time per unit equals $(C + c)/(\lambda^{-1} + D)$. Because the $n$
units are independent from each other, this gives a total mean cost per unit time of

\[ MC_N = n \cdot \frac{C + c}{\lambda^{-1} + D}. \]

**Alarm Clustering.** Since we assumed that alarm rates are constant, all units are as-good-as-new after each maintenance operation. The time between such a renewal and the next alarm signal is the minimum of \( n \) independent exponential distributions with parameter \( \lambda \), and is therefore exponentially distributed with parameter \( n\lambda \). This implies that the mean time between maintenance actions is \((n\lambda)^{-1} + D\). Maintenance is performed on the first unit that gives an alarm signal and on all other units that give an alarm signal during the \( D \) time units from then on, which happens with probability \( 1 - e^{-\lambda D} \) for each of the \( n - 1 \) units. So the mean cost per unit time equals

\[ MC_A = \frac{C + c + (n - 1)(1 - e^{-\lambda D})c}{(n\lambda)^{-1} + D}. \]

A comparison of \( MC_N \) and \( MC_A \) shows that the cost ratio of the No Clustering and Alarm Clustering policies depends on the combined factors \( R = \lambda D \) and \( \bar{C} = C/c \) rather than the separate parameters \( \lambda, D, C \) and \( c \). Note that \( R \) can be interpreted as the length of \( D \) relative to the average time \( 1/\lambda \) until a unit gives an alarm. Another notable result is that the policy that is optimal is not influenced by the number of units \( n \). This is the case because \( n \) cancels out when solving \( MC_A < MC_N \). Alarm Clustering is optimal, i.e. \( MC_A < MC_N \), if

\[ C > \frac{1}{R} - e^{-R} \left( \frac{1}{R} + 1 \right). \] (5.1)

The relative fixed maintenance cost \( \bar{C} \) for which Alarm Clustering starts to be more beneficial than No Clustering corresponds to the point where the cost saving \( C \) of a maintenance action of a unit that is included in the maintenance action of another unit that has to be maintained, starts to outweigh the additional cost of the wasted remaining useful life of this unit. The relative cost savings (that can be negative) from clustering are

\[ \frac{MC_N - MC_A}{MC_N} = 1 - \frac{(\bar{C} + n)(1 + R)}{(\bar{C} + 1)(1 + nR)} + e^{-R} \cdot \frac{(n - 1)(1 + R)}{(\bar{C} + 1)(1 + nR)}. \]
Figure 5.2 depicts these savings graphically for various values of the number \( n \) of units, \( R \), and the relative fixed cost \( \bar{C} \) of performing a maintenance action. The relative cost saving does depend on the number of units \( n \). More units results in a higher degree of clustering and, provided that \( \bar{C} \) is large enough, in a larger relative cost saving. Also, a higher relative fixed maintenance cost \( \bar{C} \) makes clustering more cost effective. Furthermore, for higher values of \( \bar{C} \), higher values of \( R \) result in lower costs. This is expected since a higher fixed cost desires a higher level of clustering through higher values of \( R \). Moreover, unless the fixed cost is very small compared to the variable maintenance cost, clustering is preferred.

Figure 5.2: Percentage decrease in cost Alarm Clustering compared with No Clustering.
It also appears from Figure 5.2 that the cost savings from clustering are less sensitive to \( R \) if \( n \) is larger. This is explained by the fact that if the number of units is large, there are ample opportunities for clustering even if units have to be maintained relatively shortly after they give an alarm.

5.5 Alert-Alarm-Maintain with constant alert and alarm rates

We let \( \lambda_1 \) and \( \lambda_2 \) denote the alert and alarm (after an alert) rates. We will first analyze the system analytically for 2 units, after which we will perform a simulation study for the case with more than two units. We note that the alert-alarm-maintain system is equivalent to the alarm-maintain system if we set \( \lambda_1 \) (or \( \lambda_2 \)) equal to \( \infty \).

5.5.1 Systems with 2 units: Analysis

No Clustering. Because maintenance is performed on single units, the cost of each maintenance action is \( C + c \). For both units, the mean time between maintenance actions is \( \lambda_1^{-1} + \lambda_2^{-1} + D \) and so the total mean cost per unit time is

\[
2 \cdot \frac{C + c}{\lambda_1^{-1} + \lambda_2^{-1} + D}.
\]

Alarm Clustering or Alert Clustering. We next derive some results that apply to both Alarm Clustering and Alert Clustering, after which these two policies will be analyzed separately. Let us denote the condition of a unit by 1 if it is as-good-as-new, 2 if it has given an alert but not an alarm signal, 3\(^1\) if it has given an alarm signal (but maintenance is not yet needed), and 3\(^2\) when maintenance is needed. Let us denote the system state by \((a, b)\) if one of the units has condition \( a \) and the other unit has condition \( b \). We remark that state \((a, b)\) is equivalent to state \((b, a)\), and that symmetrical states are therefore left out for representational ease. Figure 5.3 shows all possible system states, distributions of the times between arriving in a state and a transition to another state (either exponential or degenerate), and the transition probabilities \((p_i s, q_i s, \text{ and } r_i s)\).

The following expressions for the transition probabilities are derived in Appen-
Chapter 5

5.3: State space of the alert-alarm-maintain system with 2 units.

\[ p_1 = \frac{\lambda_2}{\lambda_1 + \lambda_2}, \quad p_2 = \frac{\lambda_1}{\lambda_1 + \lambda_2}, \]

\[ q_1 = e^{-\lambda_1 D}, \quad q_2 = \frac{\lambda_1}{\lambda_1 - \lambda_2} \left( e^{-\lambda_2 D} - e^{-\lambda_1 D} \right), \]

\[ q_3 = 1 - \frac{1}{\lambda_1 - \lambda_2} \left( \lambda_1 e^{-\lambda_2 D} - \lambda_2 e^{-\lambda_1 D} \right), \]

\[ r_1 = e^{-\lambda_2 D}, \quad r_2 = 1 - e^{-\lambda_2 D}. \]

Obviously, maintenance is performed when one of the states at the bottom of Figure 5.3 is reached. What units are maintained depends on the clustering policy.

**Alarm Clustering.** The cost of performing maintenance on the unit that needs to be maintained is \( C + c \). If the system reaches one of the states labeled as \( (3^2, 3^1) \), maintenance is also performed on the second unit with additional cost equal to \( c \). Because the system always returns to state \( (1, 1) \) or \( (2, 1) \) after a maintenance action, the probability that one of the states \( (3^2, 3^1) \) is reached equals \( p_1 q_3 + p_2 r_2 \). It follows that the mean cost per maintenance action equals

\[ C + c + (p_1 q_3 + p_2 r_2) c = C + 2c - \frac{1}{\lambda_1^2 - \lambda_2^2} \left( \lambda_1^2 e^{-\lambda_2 D} - \lambda_2^2 e^{-\lambda_1 D} \right) c. \]

If the state \( (3^2, 1) \) or one of the states \( (3^2, 3^1) \) is reached, the system returns to state \( (1, 1) \), otherwise it returns to state \( (2, 1) \). This implies that the system returns to state \( (1, 1) \) with probability \( p_1 q_1 + p_1 q_3 + p_2 r_2 \). With the same probability, the exponentially
Reducing costs by clustering maintenance activities for multiple critical units

distributed time with parameter $2\lambda_1$ between state $(1, 1)$ and $(2, 1)$ is part of the time between two maintenance actions. Furthermore, the time between maintenance actions always includes an exponentially distributed time with parameter $\lambda_1 + \lambda_2$ and a constant time $D$. With probability $p_2$, the time between two maintenance actions also includes an exponentially distributed time with parameter $2\lambda_2$. Thus, the mean time between two maintenance actions equals

$$
\frac{(p_1q_1 + p_1q_3 + p_2r_2)}{2\lambda_1} + \frac{1}{\lambda_1 + \lambda_2} + D + \frac{p_2}{2\lambda_2} = \frac{1}{2(\lambda_1^2 - \lambda_2^2)} \left( \lambda_2 e^{-\lambda_1 D} - \lambda_1 e^{-\lambda_2 D} \right) + \frac{\lambda_1 + \lambda_2}{2\lambda_1 \lambda_2} + \frac{1}{2(\lambda_1 + \lambda_2)} + D.
$$

The exponentially distributed time with parameter $2\lambda_1$ is included in the time until the next maintenance action with a fixed probability $p_1q_1 + p_1q_3 + p_2r_2$, independent of the preceding time until maintenance. The times between maintenance actions are therefore i.i.d. distributed. Furthermore, the rewards (i.e. the maintenance costs) are also i.i.d. distributed. It follows therefore from the renewal reward theorem that the mean cost per unit time is

$$
C + 2c - \frac{1}{\lambda_1^2 - \lambda_2^2} \left( \lambda_1^2 e^{-\lambda_2 D} - \lambda_2^2 e^{-\lambda_1 D} \right) c
$$

$$
\frac{1}{2(\lambda_1^2 - \lambda_2^2)} \left( \lambda_2 e^{-\lambda_1 D} - \lambda_1 e^{-\lambda_2 D} \right) + \frac{\lambda_1 + \lambda_2}{2\lambda_1 \lambda_2} + \frac{1}{2(\lambda_1 + \lambda_2)} + D.
$$

**Alert Clustering.** Again, the maintenance cost of the unit that needs to be maintained equals $C + c$. If one of the states $(3^2, 2)$ or one of the states $(3^2, 3^1)$ is reached, i.e., only if a state different from state $(3^2, 1)$ is reached, the second unit is also maintained, implying an additional maintenance cost of $c$. Because the system always returns to state $(1, 1)$ after a maintenance action, this occurs with probability $1 - p_1q_1$. The mean cost per maintenance action is therefore equal to

$$
C + c + (1 - p_1q_1)c = C + 2c - \frac{\lambda_2}{\lambda_1 + \lambda_2} e^{-\lambda_1 D} c.
$$

The time between two maintenance actions always includes an exponentially distributed time length with parameter $2\lambda_1$, an exponentially distributed time length with parameter $\lambda_1 + \lambda_2$, and a constant time $D$. Furthermore, an additional exponentially distributed time length with parameter $2\lambda_2$ is included with probability $p_2$. 
It follows that the mean time between two maintenance actions equals
\[
\frac{1}{2\lambda_1} + \frac{1}{\lambda_1 + \lambda_2} + D + p_2 \frac{1}{2\lambda_2} = \frac{\lambda_1 + \lambda_2}{2\lambda_1 \lambda_2} + \frac{1}{2(\lambda_1 + \lambda_2)} + D.
\]

Dividing the mean cost per maintenance action by the mean time between two maintenance actions results in the mean cost per unit time under Alert Clustering being equal to
\[
C + 2c - \frac{\lambda_2}{\lambda_1 + \lambda_2} e^{-\lambda_1 D} c = \frac{\lambda_1 + \lambda_2}{2\lambda_1 \lambda_2} + \frac{1}{2(\lambda_1 + \lambda_2)} + D.
\]

This completes the analysis of the considered policies for the alert-alarm system with 2 units and constant alert and alarm rates. Next, we explore the relative performance of these policies.

### 5.5.2 Systems with 2 units: Results

Figure 5.4 shows the percentage decrease in cost under Alarm Clustering and Alert Clustering compared with No Clustering as a function of the relative fixed maintenance cost \( \bar{C} \) of performing a maintenance action and for different values of the parameters \( \lambda_1 \), \( \lambda_2 \), and \( D \). Figure 5.4 (a) shows these percentage decreases for the ‘base’ values \( \bar{\lambda}_1 = 3 \), \( \bar{\lambda}_2 = 2 \), and \( D = 1 \). In the respective parts (b), (c) and (d) of this figure, the values of these parameters are changed one by one. We remark that other base values produce similar results.

A comparison of Figures 5.4 and 5.2 shows that the results for the alarm-maintain system largely carry over to the more sophisticated alert-alarm-maintain system. Clustering offers a larger benefit if the fixed maintenance cost increases. Furthermore, for a sufficiently high fixed maintenance cost, the benefit also becomes larger for higher values of \( D \) as this leads to more clustering opportunities.

For all parameter settings in Figure 5.4, there is a threshold for the fixed maintenance cost after which Alert Clustering becomes preferable to Alarm Clustering. This is expected, since Alert Clustering has a higher degree of clustering. Moreover, the threshold appears to be decreasing in the average time \( 1/\lambda_1 \) until an alarm, increasing in the average time \( 1/\lambda_2 \) between an alert and alarm, and increasing in the time \( D \) between alarm and maintenance. If \( 1/\lambda_1 \) increases or \( D \) decreases, Alarm
Clustering becomes less effective and alert clustering more ‘desired’. An increase in $1/\lambda_2$ implies more clustering under the Alert Clustering policy, and a higher fixed maintenance cost is needed to justify the larger amount of unused service life.

![Graphs showing percentage decrease in cost for different values of $n$](image)

Figure 5.4: Percentage decrease in cost Alarm Clustering and Alert Clustering compared with No Clustering ($n = 2$).

### 5.5.3 Systems with $n$ units

For a number of units $n$ greater than 2, we perform simulations to compare the policies. Figure 5.5 shows the percentage decrease in cost under Alarm Clustering and Alert Clustering compared with No Clustering for different values of $n$. The
values $\lambda_1^{-1} = 3$, $\lambda_2^{-1} = 2$, and $D = 1$ are used in this simulation, and we run each simulation until 100,000 maintenance actions have been performed. The case with $n = 2$ units is included for validation purposes and Figure 5.5 (a) is indeed virtually identical to Figure 5.4 (a).

In accordance with (5.1) for the Alarm-Maintain system, the values of the relative fixed maintenance cost $\bar{C}$ for which the respective clustering policies start to outweigh No Clustering seem not to depend on the number of units $n$. Again, the profits that can be obtained from clustering increase with the number of units. A higher number of units implies that per maintenance action the mean number of units on which maintenance is performed is also higher. Therefore, the fixed cost $C$ of performing a maintenance action can be divided over more units, resulting in a larger relative cost decrease.

Another important observation is that the threshold relative fixed cost $\bar{C}$ per maintenance action for which Alert Clustering becomes more beneficial than Alarm Clustering, increases if the number of units $n$ increases. An increase in the number of units (all else unchanged) implies more clustering opportunities for Alarm Clustering and, apparently, this can make it more cost effective than Alert Clustering by extending the average lifespan of units.

5.6 Non-constant alert and alarm rates

So far, we assumed that the alarm (and alert) rates were constant, i.e., that degradation phases had exponential durations. Doing so allowed us to obtain analytical insights into the cost performance of different clustering policies. In this section we test whether our findings are robust for settings with non-constant alert and alarm rates. More specifically, we assume that the respective durations follow gamma distributions with shape parameter $k$ and scale parameter $\theta$. We keep the means $k\theta$ of the gamma distributions constant, and let the coefficients of variation $CV = 1/\sqrt{k}$ vary. Note that the exponential distribution is a special case of the gamma distribution with $CV = 1$. A gamma distribution with $CV < 1$ implies an increasing failure rate (IFR), whereas $CV > 1$ implies a decreasing failure rate (DFR). We remark that a complete lifetime can have an IFR even if the durations of some stages do not.

The results in this section are also obtained by performing simulations, similar to those in Subsection 5.5.3. We again run the simulations until 100,000 maintenance
Reducing costs by clustering maintenance activities for multiple critical units

Figure 5.5: Percentage decrease in cost Alarm Clustering and Alert Clustering compared with No Clustering ($1/\lambda_1 = 3$, $1/\lambda_2 = 2$, $D = 1$).

actions have been performed.

5.6.1 Alarm-Maintain

We start our analysis with the alarm-maintain system with a gamma distributed time until the alarm signal. We keep the mean of this gamma distribution constant at 5, and we let the coefficient of variation vary. We also set $C$, $c$ and $D$ all equal to 1 and remark that other settings lead to similar (sensitivity) results as those reported below.

Figure 5.6 shows the percentage decrease in cost of Alarm Clustering compared
with No Clustering for \( n = 2 \) and for \( n = 5 \) units. It appears that the cost reduction is quite robust for changes in the CV. Situations with an IFR are most realistic and the savings for such situations are slightly larger than for the case with a constant failure rate (\( CV = 1 \)). So, the closed form expressions derived for that case in previous sections provide conservative but accurate estimates of the clustering benefits. A very low coefficient of variation implies that maintenance can always be clustered for all units, leading to much higher savings. Indeed, it easily follows that, for \( C = c \), the maximum savings are 25% for \( n = 2 \) and 40% for \( n = 5 \).

![Graphs showing percentage decrease in cost for different values of CV](image)

Figure 5.6: Percentage decrease in cost Alarm Clustering compared with No Clustering in the alarm-maintain system with gamma distributed time until the alarm signal (\( k\theta = 5, D = 1, \bar{C} = 1 \)).

### 5.6.2 Alert-Alarm-Maintain

In our analysis of the alert-alarm-maintain system, we first consider a gamma distributed time until the alert signal (with parameters \( k_1 \) and \( \theta_1 \)), then a gamma distributed time between the alert and alarm signal with parameters \( k_2 \) and \( \theta_2 \), and finally the case in which both durations follow a gamma distribution. In all these cases, we fix the mean time until the alert signal to 3 and the mean time between the alert and alarm signal to 2. A maintenance action should again be performed within a period with length \( D = 1 \) after the alarm signal and the values of the cost parameters of performing maintenance are again \( C = c = 1 \). The sensitivity results for variations in the CV that we will present below are illustrative for other cases.
that we considered. The percentage decrease in cost under Alarm Clustering and Alert Clustering compared with No Clustering for gamma durations until the alert signal are shown in Figure 5.7; those for gamma durations between the alert and the alarm signal in Figure 5.8; and lastly those for the case where both durations follow a gamma distribution (with the same $CV$) in Figure 5.9. Systems with $n = 2$ and with $n = 5$ units are again considered.

![Graphs showing percentage decrease in cost under Alarm and Alert Clustering](image)

Figure 5.7: Percentage decrease in cost Alarm Clustering and Alert Clustering compared with No Clustering in the alert-alarm-maintain system with gamma distributed time until the alert signal ($k_1 \theta_1 = 3$) and exponential distributed time between the alert and alarm signal ($\lambda_2^{-1} = 2, D = 1, \bar{C} = 1$).

For both clustering policies and for all extensions considered, the cost savings compared to No Clustering decrease in the $CV$s implying that the analytical results for exponential durations ($CV = 1$) again provide a lower bound for the most realistic situation with IFR durations. As we observed for the alarm-maintain system, the $CV$s only have a minor influence on the savings that can be obtained under Alarm Clustering (implying that assuming $CV = 1$ provides accurate estimates) unless the $CV$ of both the time until the alert signal and the time between the alert and alarm signal is so small that maintenance can (almost) always be clustered for all units. The Alert Clustering savings are more affected by the $CV$s, especially by that of the durations between the alert and the alarm signal. Alert Clustering becomes more attractive, also relative to Alarm Clustering, for realistic scenarios with IFR durations until the alert signal. An intuitive explanation for this result is that Alert Clustering is more effective if the remaining lifetime after an alert signal is more
Figure 5.8: Percentage decrease in cost Alarm Clustering and Alert Clustering compared with No Clustering in the alert-alarm-maintain system with exponential distributed time until the alert signal $($\(\lambda_1^{-1} = 3\)\) and gamma distributed time between the alert and alarm signal $($\(k_2\theta_2 = 2, D = 1, \bar{C} = 1\)$).

5.7 Implications for the Groningen gas field

In this section we will apply the alert-alarm-maintain system to a real-life case of equipment maintenance at the Groningen gas field in the north of The Netherlands. The field is operated by the Nederlandse Aardolie Maatschappij (NAM), a joint venture between Royal Dutch Shell (Shell) and ExxonMobil. NAM is managed as an operating company of Shell. The actual maintenance activities on the Groningen field are largely outsourced to GLT-PLUS, a consortium of contractors and strategic equipment suppliers.

The use of condition-based maintenance is promoted within Shell and NAM, for several important reasons, including the avoidance of (i) safety risks associated with breakdowns and unplanned stoppages and (ii) operational risks (stable processes ensuring reliable gas production are of the essence). The condition of several (major) units is monitored in order to optimize the operational performance and maintenance costs. This is done in accordance with the Shell Maintenance and Integrity Management Standard EP2005-5038 (Shell International Exploration and Produc-
The standard describes three condition stages of the monitored equipment: ‘alert’, ‘alarm’, and ‘failure’.

The gas is produced by unmanned facilities that each contain 20 identical pump units. In this paper we will consider the maintenance actions of the pump units within a single facility. Relative (normalized) cost data for a typical maintenance action of such pump units is presented in Table 5.1. The values are based on estimates obtained from the reliability engineers, who are tasked with the design and optimization of maintenance policies at the Groningen gas field. The table distinguishes between fixed cost per work order and variable cost per pump unit on which maintenance is performed. The following activities are involved in the execution of a maintenance work order: (i) Work preparation: preparing the work order by prescribing the scope, supplying drawings and work instructions and organizing work permits; (ii) Finance/ procurement: estimating, budgeting, approval of the work order and procurement or call-off of materials and/or services involved; (iii) Transport/ travel contractor: transport of the contractor to the maintenance location; (iv) Safety procedure per unit: a procedure at unit-level to ensure that all necessary safety precautions are taken before the work can commence (this includes the involvement of NAM-staff, if and when contractor staff enter the site of a gas facility, NAM staff will be present); (v) Installing scaffolding and other maintenance aids; (vi) Man hours,
the direct work involved with the maintenance action; (vii) Any direct materials involved; and (viii) Aftercare: evaluation of the work order, payment, closure of the work order.

Table 5.1: Relative maintenance cost for a typical set of identical pump units on a single facility at the Groningen gas field.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Fixed cost per work order (%)</th>
<th>Variable cost per pump</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work preparation</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Finance/administration</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>Transport</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Safety procedure unit</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Scaffolding/aids</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Man hours</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>Materials</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>Aftercare</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>50</strong></td>
<td><strong>50</strong></td>
</tr>
</tbody>
</table>

The table indicates that, for this pump unit, the fixed and variable cost of maintenance are in balance and, in modeling terms, that it is reasonable to choose $C$ equal to $c$ (implying that the relative fixed maintenance cost $\bar{C}$ equals 1).

Failures modes that are encountered for the pump units are bearing failures, impeller wear (e.g. due to cavitation) and shaft failures (mainly due to misalignment between motor and pump). All these failure modes can be detected by vibration monitoring, as they all lead to higher vibration levels. The alarm level is chosen such that a sufficient amount of preparation time is available after the alarm signal. According to the reliability engineers, realistic rough estimates for the mean times until an alert signal, between an alert and an alarm signal, and between an alarm signal and a maintenance action are “a couple of years”, “a couple of months”, and “a couple of weeks”, respectively. Based on a simulation study, we evaluate the cost benefits of clustering for a set of 20 identical pump units with a mean time until an alert signal of 2 years (24 months), a mean time between an alert and an alarm signal of 3 months, and a fixed period of 4 weeks (1 month) between an alarm signal and a maintenance action. The reliability engineers confirm that a period of 4 weeks is sufficient for the planning of a maintenance activity on the pump units. We know from the results in previous sections and in particular from those in Figures 5.4 (a) and (b)
that Alert Clustering is more cost effective than Alarm Clustering even for relatively low values of $\bar{C}$ if the mean time until an alert signal is relatively large compared to the other durations, which is the case in our setting. On the other hand, as Figure 5.5 shows, Alarm Clustering is often more beneficial if the number of units is large. It turns out that the first effect prevails in our setting. The percentage decrease in cost under Alert Clustering compared with No Clustering is about 28%, while this is about 20% under Alarm Clustering. These results are based on constant alert and alarm rates. However, we have learned in Section 5.6 that the benefit of Alert Clustering compared to Alarm Clustering becomes larger if the alert and alarm rates are increasing, implying that Alert Clustering is the preferred policy under a wide variety of conditions.

In practice, there is some flexibility in setting the alert level. As the time between an alert and an alarm (with mean $1/\lambda_2$) increases, Alert Clustering reduces the maintenance frequency but also the unit lifespan. The latter disadvantage starts to outweigh the former advantage, compared to Alarm Clustering, for some threshold mean duration between an alert and alarm. This is shown in Figure 5.10, where this duration is varied whilst the mean time to imminent failure is kept constant. It appears that the threshold is about 8 months and that having an alert signal about 3 months on average before the alarm signal is indeed nearly optimal.

![Figure 5.10: Percentage decrease in cost Alarm Clustering and Alert Clustering compared with No Clustering for a typical set of 20 pump units at the Groningen gas field. The mean time $1/\lambda_2$ between alert and alarm is varied while the total mean time until alarm $1/\lambda_1 + 1/\lambda_2$ is kept constant at 27 months ($D = 1$ month, $\bar{C} = 1$).](image-url)
5.8 Conclusions and future extensions

We have investigated the cost benefits from clustering maintenance operations for systems with multiple critical units and condition monitoring. Earlier studies that combine condition monitoring and clustering focus on rather complex systems that offer limited general insights. Our study considers monitoring systems with one (alarm) or two (alert-alarm) deterioration states/signals that do offer such insights. These systems are easy to implement and used in various industries. We restrict our attention to situations with identical units (a ‘fleet’ of e.g. pumps or ships) as this provides a natural and realistic starting point.

For systems with constant alert and alarm rates, analytical results were derived on the cost savings that result from using clustering and on the optimal degree of clustering, i.e. after an alarm or after an alert. It transpired that clustering offers considerable benefits for a wide range of realistic settings and of about 28% for a specific real-life case that we considered. Furthermore, if the fixed maintenance cost is above a certain threshold, then the higher degree of clustering (after an alert) is preferred. A particularly insightful result is that this threshold increases with the number of units. This is explained by the fact that more units imply more clustering opportunities for highly deteriorated units, and so there is less additional benefit of opportunistically maintaining somewhat deteriorated units as well, relative to the reduction in lifespan that results.

These results are robust for more general systems with varying alert and alarm rates. For the most realistic situation with increasing alert and alarm rates, the closed-form results for constant alert and alarm rates provide conservative estimates of cost savings from clustering and can thus be used as a quick check to establish whether clustering activities (after an alarm or an alert) is sufficiently profitable. Furthermore, for alarm clustering these estimates are also very accurate. (Additional) savings from alert clustering are more sensitive to the variation in alert and alarm rates. Alert clustering is less beneficial if the alert signal is less indicative of the remaining life of a unit.

In line with our modeling, there are a number of avenues to extend this research. First, systems with non-identical units or units that are not operated in the same way can be analyzed. Second, the deterioration process can be modeled in more detail by using more than two deterioration states/signals, although this will also complicate
the implementation of both condition monitoring and maintenance clustering planning. Third, imperfect signaling (i.e., false positives and false negatives) could be added to the model (Berrade et al., 2013b; Flage, 2014). Fourth, the time until maintenance can be modeled as a decision variable, although this will make the analysis more difficult as it implies that maintenance actions may not be performed in time and failure risks and costs have to be modeled as well. Furthermore, the focus will then be more on the balance between preventive and corrective maintenance.

A final direction for future research is to consider the simultaneous optimization of both the deterioration levels that trigger a signal and the type of clustering policy used, as both affect the degree of clustering that turns out to be a critical factor in our results. However, as already mentioned in Section 5.3, this extension requires the deterioration process to be modeled explicitly. This is already a challenge in itself as it is often not possible to obtain the required data. Relevant data is commonly distributed across various systems that are difficult to link and the data that is available is regularly of an insufficient quality (Braaksma et al., 2013). This also emerged in the case study discussed in Section 5.7 where we had to rely on estimates of reliability engineers. A prerequisite for the development and analysis of more sophisticated models is the availability of more reliable data.
5.A Derivation of conditional transition probabilities

Consider Figure 5.3. Starting at state $(1, 1)$ in which none of the units has given an alert signal, the time until the first alert signal is exponentially distributed with parameter $2\lambda_1$. Thereafter, the time until the next signal is exponentially distributed with parameter $\lambda_1 + \lambda_2$. This signal is an alarm signal of the unit that has already given an alert signal with probability

$$p_1 = \frac{\lambda_2}{\lambda_1 + \lambda_2},$$

implying that the system goes to state $(3^1, 1)$, and is an alert signal of the other unit with probability

$$p_2 = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$  

In the latter case, a transition to state $(2, 2)$ takes place. From the moment that the state $(3^1, 1)$ is reached, it takes $D$ time units until maintenance will be performed. During this period, we may distinguish three cases for the other unit. With probability

$$q_1 = e^{-\lambda_1 D}$$

it does not give an alert signal. The probabilities $q_2$ and $q_3$ depend on the sum $X_1 + X_2$ of two exponentially distributed random variables $X_1 \sim \exp(\lambda_1)$ and $X_2 \sim \exp(\lambda_2)$. 


The distribution function $F_{X_1+X_2}$ of $X_1 + X_2$ equals
\[
F_{X_1+X_2}(y) = P(X_1 + X_2 \leq y)
= \int_0^y \left(1 - e^{-\lambda_1(y-x_2)}\right) \lambda_2 e^{-\lambda_2 x_2} \, dx_2
= \int_0^y \left(\lambda_2 e^{-\lambda_2 x_2} - \lambda_2 e^{-\lambda_1 y + (\lambda_1 - \lambda_2) x_2}\right) \, dx_2.
\]

For the case that $\lambda_1 = \lambda_2$ this gives
\[
F_{X_1+X_2}(y) = 1 - e^{-\lambda_2 y}(1 + \lambda y),
\]

For the more realistic case where $\lambda_1$ and $\lambda_2$ have different values, to which we shall restrict ourselves in what remains, we get
\[
F_{X_1+X_2}(y) = \left. \left(-e^{-\lambda_2 x_2} - \frac{\lambda_2}{\lambda_1 - \lambda_2} e^{-\lambda_1 y + (\lambda_1 - \lambda_2) x_2}\right) \right|_{x_2=0}^y
= 1 - \frac{1}{\lambda_1 - \lambda_2} \left(\lambda_1 e^{-\lambda_2 y} - \lambda_2 e^{-\lambda_1 y}\right).
\]

It follows that the probability $q_3$ that the second unit gives both an alert signal and an alarm signal during the period $D$ after the alarm signal of the first unit equals
\[
q_3 = F_{X_1+X_2}(D) = 1 - \frac{1}{\lambda_1 - \lambda_2} \left(\lambda_1 e^{-\lambda_2 D} - \lambda_2 e^{-\lambda_1 D}\right).
\]

This implies that the second unit only gives an alert signal during this period with probability
\[
q_2 = 1 - q_1 - q_3 = \frac{\lambda_1}{\lambda_1 - \lambda_2} \left(e^{-\lambda_2 D} - e^{-\lambda_1 D}\right).
\]

If the system is in state $(2, 2)$, the time until one of the units gives an alarm signal and the system goes to state $(3^1, 2)$ is exponentially distributed with parameter $2\lambda_2$.

Once this state has been reached, it takes a deterministic time $D$ until maintenance will be performed. During this period, the other unit does not given an alarm signal with probability
\[
r_1 = e^{-\lambda_2 D},
\]
implying that a transition to state \((3^2, 2)\) takes place. With probability

\[
r_2 = 1 - r_1 = 1 - e^{-\lambda_2 D}
\]

the other unit gives an alarm signal during the period with length \(D\), and the system goes to state \((3^3, 3^2)\).