RESEARCH ARTICLE

Integration of maintenance scheduling and planning for large-scale asset fleets

Frederik Schulze Spüntrup[1] · Giancarlo Dalle Ave[2,3] · Lars Imsland[1] · Iiro Harjunkoski[2]

Abstract

Large fleets of engineering assets that are subject to ongoing degradation are posing the challenge of how and when to perform maintenance. For a given case study, this paper proposes a formulation for combined scheduling and planning of maintenance actions. A hierarchical approach and a two-stage approach (with either uniform or non-uniform time grid) are considered and compared to each other. The resulting discrete-time linear programming model follows the Resource Task Network framework. Asset deterioration is considered linearly and tackled with an enumerator-based formulation. Advantages of the model are its computational efficiency, scalability, extendability and adaptability. The results indicate that combined maintenance planning and scheduling can be solved in appropriate time and with appropriate accuracy. The decision-support that is delivered helps the choice of the specific maintenance action to perform and proposes when to conduct it. The paper makes a case for the benefits of optimally combining long-term planning and short-term scheduling in industrial-sized problems into one system.

Keywords Planning and scheduling · Maintenance · Engineering asset fleets

1 Introduction

With typical lifetimes between 30 and 50 years, industrial process plants are subject several types of degradation during their life cycle (Wintle et al. 2006). Efficient and sustainable operation during these timescales is important to

---

✉ Frederik Schulze Spüntrup
frederik@schulzespuentrup.de

1   Department of Engineering Cybernetics, Norwegian University of Science and Technology, 7491 Trondheim, Norway

2   Hitachi ABB Power Grids Research, 68309 Mannheim, Germany

3   Technical University Dortmund, 7491 Dortmund, Germany

maintain competitiveness and ensure the safety and reliability of the plants. After the construction and the commissioning of a plant, the longest phase of the life cycle starts, the operation and maintenance phase. It is desired to keep the cash flows as high as possible during this period. However, decisions affecting this happen on different time-scales.

The long-term planning of shutdowns and turnarounds is essential for companies in the process industries as they need to maintain high production levels and safe and reliable operation. At the same time, regular turnarounds are required by statutory regulations and guidelines. One of the objectives of every turnaround and maintenance action is to minimize the downtime, as production losses directly correlate to profits. In larger turnarounds there are usually multiple contractors involved, which poses a challenge from a logistics point of view. Therefore, rigorous planning in advance is required as described by Al-Turki et al. (2019).

At the same time, production losses may occur through unplanned shutdowns, caused on a short-term basis by equipment failure. This led Jardine et al. (2006) to write a review on the newest technological advancements in condition monitoring and predictive analytics that try to minimize the risk of unplanned shutdowns by detecting potential asset failures beforehand so that countermeasures in the form of maintenance actions can take place. Further advancements have been made and application in industry can be seen.

Long-term strategic planning of turnarounds or maintenance actions, but also investment projects can be handled with a lower level of detail on long time horizons. This takes into account degradation processes with slow dynamics and allows for slack for logistic between distributed production sites. On the other hand, it is for operational reasons important to schedule different operations on a detailed level to enable different stakeholders to execute their tasks on time.

In the past separate scheduling and also planning models to approach these problems have been individually developed, described in literature and applied to industrial scenarios. Planning and scheduling are from a mathematical point of view very similar problems that mostly differ in regards of the considered time horizons. Scheduling is performed for time horizons from weeks to months, while planning considers time horizons of multiple years. The integration of scheduling and planning models has not progress as much as the development of the individual scheduling or planning models.

Integration of planning and scheduling of maintenance actions is important for various reasons. The trend towards production networks makes use of economies of scale as more similar assets are aggregated in production systems to decrease the production cost per production unit. Therefore, large asset fleets are common in the process industry. While they offer the opportunity to respond to demand fluctuations, they pose substantial challenges regarding the complexity of the underlying mathematical problems.

The data perspective is also an important reason to integrate both systems. From a managerial perspective it is strategically important to be aware of the long-term plan of maintenance actions that are expensive, take a long time causing production losses or need to fulfill legal requirements. On the other hand, some decisions cannot

be taken on a long-term basis and need to therefore, be handled in the time-frame of scheduling, for example a sudden machine failure that forces a stop of production.

This paper addresses the integration of optimal short-term scheduling of maintenance actions and the long-term planning of maintenance operations to increase the operational efficiency. Sect. 2 gives an overview of the work on similar integration approaches mostly in the field of production planning and scheduling. Even though this problem is common in many subcategories of the process industry, we will focus on a case study of an offshore compressor fleet of an oil and gas company. Both the problem definition and case study will be introduced in Sect. 3.

An overview of the requirements for the scheduling and the planning model are given in Sect. 4, followed by the detailed mathematical formulation and methodology for the integration of both models. The results of the maintenance scheduling are illustrated and discussed within the different subcases in Sect. 5. Sections 6 and 7 concern the presentation of the results of the various subcases and their discussion and interpretation. The conclusion gives an overview of the findings of this work and shows possible directions for future work.

2 Background and literature review

While scheduling and planning can be done independently from each other, the interdependence of both systems is evident. The scientific community has put in the past a fair amount of effort into the integration of both systems. The review article of Maravelias and Sung (2009) summarizes the challenges and the opportunities for combined production planning and production scheduling. They introduce different modeling approaches, such as relaxed and aggregated scheduling formulations as used in Harjunkoski and Grossmann (2002), Jain and Grossmann (2001) or Wilkinson et al. (1995). Harjunkoski and Grossmann (2002) present two strategies for the decomposition of multistage scheduling problems. Jain and Grossmann (2001) combine MILP and CP techniques into one hybrid model that outperforms the techniques individually. Wilkinson et al. (1995) presented a method for producing accurate aggregate models based on a discrete-time formulation. As an alternative to relaxed and aggregated scheduling formulations, Maravelias and Grossmann (2001) name offline surrogate models (e.g. Wan et al. (2005) for the simulation-based optimization in supply chain management or Sung and Maravelias (2007) with an approach to solve production planning problems in multiproduct processes). Furthermore, an overview of different solution strategies is given, as depicted in Fig. 1. Next to iterative solution strategies which are not considered in this work, hierarchical and full-space solution strategies are explained with recent examples. In hierarchical models there is an information flow from the the master subproblem (i.e. the planning model) towards the slave problem (i.e. the scheduling model) and there is no feedback loop (e.g. McKay et al. 1995) with an aggregate formulation based on a detailed discrete-time formulation or Amaro and Barbosa-Póvoa (2008) with a supply chain case study from the pharmaceutical industry where the planning solution is used as an input for the scheduling level). Full-space models include a detailed scheduling submodel during the planning period and are hard to solve (e.g.
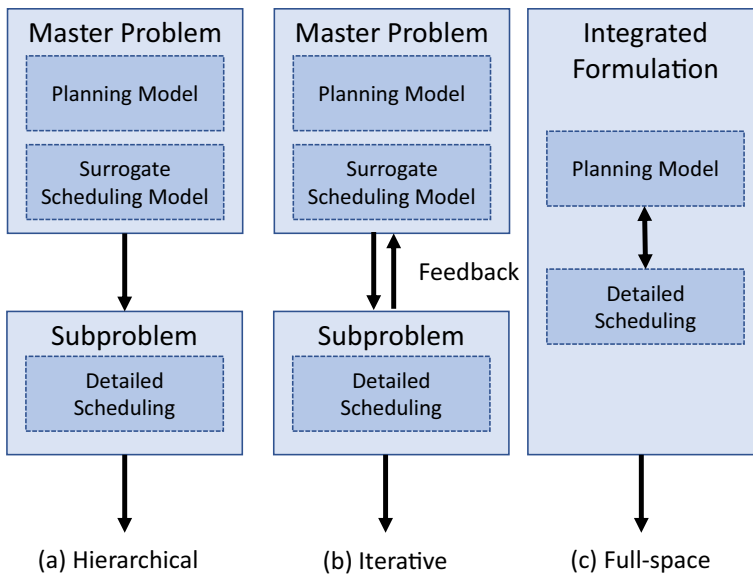
Fig. 1 Uniform discretization of the planning and scheduling horizon (Maravelias and Grossmann 2001)

Bassett et al. 1996). One way to handle this is by focusing on time-based decomposition approaches (Bassett et al. 1996) or as (Papageorgiou and Pantelides 1996) with a single-level formulation. Maravelias and Grossmann (2001) state that the biggest challenges lie in the formulation of complex process networks and that uncertainty and data integration are complicating the efficient formulation.

Similarly, Grossmann et al. (2008) give an overview of planning and scheduling for the process industries from the point of view of enterprise-wide optimization. Stated challenges are the modeling of novel mathematical programming and logic-based models that are able to capture the complexity of the reality while being simplified to a level in which the problem can be solved. If this is paired with the integration of optimal decision-making over multi-timescale, an efficient solution for industrial application can be presented. Other challenges that are mentioned are uncertainty and algorithmic challenges. Two applications are shown in case studies from batch scheduling and crude oil scheduling.

More recent work in this field is done by Zhang and Grossmann (2016). Their interest was in the intelligent management of electricity demand, also referred to as demand side management (DSM). Demand side management is used to refer to a group of actions designed to efficiently manage a site's energy consumption, this includes also planning and scheduling of production levels and similar operations. As a field for future research the aggregation of multiple DSM participants is mentioned. To tackle the growing problem size and complexity of real-world problems, as the current formulations and algorithms may not yet be able to solve these problems. Therefore, thorough handling of uncertainty still remains out of reach for large-scale problems, as it would be computationally too demanding. They present two case studies from air separation plants.

While many approaches of integrating planning and scheduling work with discrete time representation, there are also continuous-time representation approaches. In the work of Dogan and Grossmann (2006) the problem of integration of planning and scheduling for a continuous multiproduct plant is adressed. As the problem becomes computationally very expensive, a rigorous bi-level decomposition is proposed. While the upper level determines the potential products, their production levels and inventories, the lower level solves the binary variables and a detailed sequence of the products. With integer and logic cuts the feasible search space is reduced and the gap between both solutions is tightened. For horizons of one to six months a method that is significantly faster than the full-space method was obtained, while converging with finite tolerance. The problem has less than 1000 binary variables and 6000 continuous variables. In a related work, Erdirik-Dogan et al. (2007) apply their approach to a small real-world case study from the chemical industry. Special challenges arise through sequence-dependent changeover times and two-stage production.

A recent example for the need to integrate planning and scheduling is given in Carvalho et al. (2015). The food industry, in particular the ice-cream industry, has its own challenges that affect the production process and the management of it. Specifically, the perishability of raw materials within a longer planning horizon versus a just-in-time raw material delivery policy is interesting. The Resource Task network framework from Pantelides (1994) was used considering a discrete-time representation. Even though the problem size is small, the specifics of the process require additional constraints in the MILP formulation. The results show better economic results and effects on the final product quality as the raw materials can be processed just-in-time instead of being subject to longer storage times.

Another more recent work was presented by Vieira et al. (2018) and discusses the integration and following decision support for planning and scheduling of automated assembly lines. The methodology combines mathematical programming with a discrete-event simulation model. This results in an initial production plan for a set of products, while in the scheduling step a more detailed validation of the capacity-feasible schedule is done. This work is relevant, as it highlights the needed flexibility in specific industries to enable feasibility at any time.

While there is to the knowledge of the authors no research available of the integration of maintenance scheduling and planning, the maintenance topic is common in the PSE community.

Castro et al. (2014) propose a continuous-time model for long-term scheduling of a gas power plant with parallel units. The model constraints have been provided using a generalized disjunctive programming formulation which is then transformed into MILP formulations using big-M and convex hull reformulations. A very different approach was chosen by Yang et al. (2008) to schedule the maintenance actions in a manufacturing system. By utilizing a genetic algorithm an optimization procedure identifies the most cost-effective maintenance schedule, comparing the schedule to three different maintenance strategies, namely corrective, time-based and condition-based maintenance. In this work, the time-based maintenance strategy is the reference point against which the optimized schedule is compared. Kopanos et al. (2017) include maintenance and production planning and apply it to a compressor

network in air separation plants. Similarly to the approach in this paper, a mixed integer programming model has been chosen and binary decision variables model the minimum run and shutdown times. However, the approach of Kopanos et al. (2017) used a bigger amount of binary variables and the model includes details such as specifics of the downstream process, e.g. distillation columns or inventory levels. This causes the model to be big and not suitable to model large asset fleets.

The literature shows that there is extensive research done in the field of integration of planning and scheduling. However, for the integrated planning and scheduling of maintenance tasks, especially in the context of industrial-sized problems, no solution is provided. The contribution of this work is the integration of maintenance planning and scheduling into one model. Several integration approaches are tested, including a hierarchical model where the results of the planning model set new constraints for the scheduling model and hereby tighten the solution space. Two full-space model approaches are tested, one with an uniform time grid and one with a non-uniform time grid. A performance analysis of the three variants is conducted and the application for large asset fleets is tested with a realistic case study from the oil and gas industry in order to find the integration approach with the best performance.

3 Motivating example

Processing units within the process industry are affected by process degradation. Over time, production systems are in need of maintenance. This is because specific parts in the machine are subject to stresses, fouling phenomena occur and machines degrade mechanically over time, impacting the performance. Another reason to perform maintenance is safety. Economic reasons play an important role: engineering assets (e.g. turbines or compressors) degrade and do not bring the same efficiency as they originally did (Aretakis et al. 2012).

Furthermore, there are many examples of industrial production facilities where production is distributed and/or in remote locations. A historical example for distributed production is the cotton industry (Thistlethwaite and Taylor 1953). This industry faced similar problems like other distributed production facilities face nowadays. A major example, the one considered in this paper, is the oil and gas industry with offshore production facilities or production locations in the desert. Another example is desalination plants on small islands or distributed energy generation including urban and rural wind and solar energy production facilities. Maintenance before an unplanned shutdown is in these scenarios even more critical, as the maintenance personnel are not necessarily in the proximity of the plant and production losses are higher due to the time needed for arranging the necessary logistics.

This work presents a case study from the oil and gas industry. This work focuses on the development of an integrated formulation for scheduling and planning in the case study. More detailed information about the data set can be found in the paper of Schulze et al. (2020) where the specific data structure and the parameters are explained in more detail. A general overview of the utilized case study is given below. Fleets of gas compressors are located on different offshore oil platforms (see

Fig. 2) and produce dry gas that is afterwards exported to onshore facilities. As is common in practice, all compressors are assumed to run at full capacity except when they are maintained.

Every asset is associated with the current, maximal and minimal (by operator specifications) goodness. This goodness is decreasing over time by a specific degradation factor. As in the real-world, this degradation factor can be distingushed between recoverable and non-recoverable degradation and depends strongly on the gas content and the specific compressor model. Average values are assumed here that are in the range of other literature. Kurz and Brun (2012), Lakshminarasimha and Boyce (1994) and Igie et al. (2011) describe degradation rates of about 4% per year, thereof 3% as recoverable degradation and 1% as non-recoverable degradation. Figure 3 shows the pattern of the degradation with regular maintenance activities that is most relevant for this case study. Additionally, the compressors are also deteriorating which ultimately causes an asset failure. To model this, every compressor is associated with a Remaining Useful Lifetime (RUL) value which is decreasing over time. In an application scenario, this value is assumed to be updated by external Condition Monitoring systems.

Some maintenance types are already mentioned in Fig. 3. Next to compressor inspections and online and offline washing of compressors, full maintenances that are conducted onshore are also considered. Minor machine faults, e.g. leaking bearings, may be maintained on the platform directly and in a shorter time. Every maintenance type is associated with a duration, required personnel and cost. It is common in the process industry to utilise the downtime of machinery due to the need for maintenance to perform a bigger set of maintenance actions, e.g. when performing
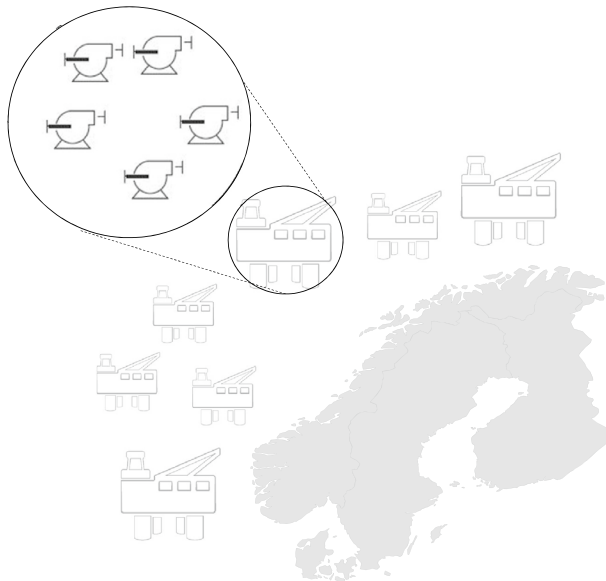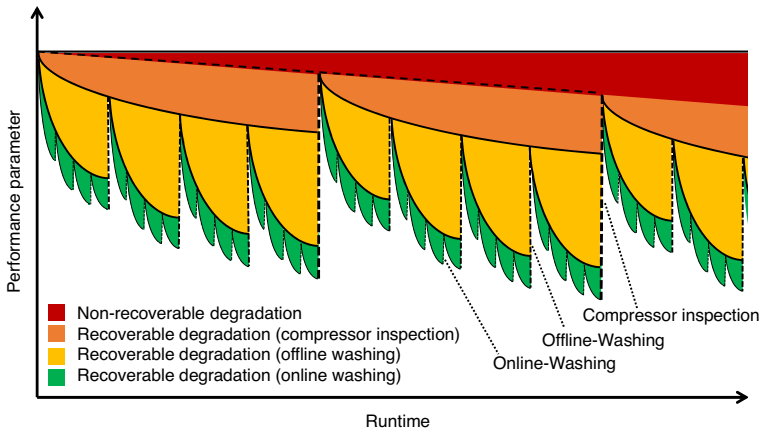

Fig. 2 A compressor network for the distributed production of gas

Fig. 3 Recoverable and non-recoverable degradation for compressors in this case study

a long maintenance task, smaller maintenance tasks are performed on the side (if possible due to accessibility and availability of spare parts) or to perform an offline washing at the same time. A detailed overview of the maintenance tasks performed in this case study are provided in Table 1. Further details can be found in the study case study description (Schulze et al. 2020).

So far, the decision to conduct maintenance has been mostly driven by legal requirements and to prevent shutdowns. The perspective of maintaining more often in order to increase the efficiency is important but makes the problem more complex.

4 Problem statement and modeling approach

With the above motivating example in mind, we define a general modeling framework for production systems that are subject to degradation with different maintenance modes as countermeasures. Our goal is to determine on both a long-term and short-term perspective, what the optimal combination of maintenance actions is in order to maximize the operational profit while fulfilling the safety and reliability requirements.

Before presenting the detailed mathematical formulation, we will describe the basic principle that is followed in the given case study.

The generalized case is a continuous production plant with a set $r \in R$ of producing assets (e.g. compressors, reactors, furnaces). Each unit is available for production if maintenance is currently not carried out. The possible maintenance tasks $i \in I$ are associated with data for cost, duration, personnel needs and resulting asset improvement.

The mathematical model that is used in this work comprises both planning and scheduling decisions. Both are realized as discrete-time models and follow the Resource-Task Network (RTN) framework (Pantelides 1994). In the past, the RTN framework has been proven as both generic and simple, which has resulted in the

Table 1 Overview of different maintenance types available in the case study

Maintenance type	Duration	Cost (in USD)	Manpower required	Recovery goodness (%)	Recovery RUL	Comments
Short maintenance	3 days	10,000	2	0	Yes	Performed off-shore
Compressor inspection	5 days	15,000	2	80	No	Performed off-shore
Long maintenance	10 days	80,000	3	100	Yes	Performed on-shore
Online washing	1 h	400	0	20	No	No production stop required
Offline washing	12 h	800	1	35	No	Production stop required

successful handling of industrial case studies, e.g. in Castro et al. (2013) where the demand side management of a steel plant is modeled as an RTN. As both models follow the RTN framework, their formulation structure is very similar, which allows for easier integration. The main differences of this work are the time length of the grids and the consideration of different maintenance types. This will be explained for both the planning and scheduling model individually, followed by an overview of the formulation itself. The Sects. 4.1 and 4.2 focus hereby on the 'why and what' of the methodology, while the following sections dive into the 'how' of the methodology. In order to avoid duplicate information, the specific formulations for scheduling and planning are not introduced.

4.1 Planning model

The planning model is responsible for the long-term planning of the maintenance actions. On this time-scale the model is able to take slow degradation processes into account and to propose that a pre-emptive maintenance action should be performed because of degradation rather than because of machine failure. However, if a maintenance operation is pre-determined, e.g. by the scheduling model, the long-term planning is updated by fixing the degradation during this overhaul. The planning model does not include maintenance modes such as online or offline washing, as these are more frequent and short-term and thus do not match with the more coarse planning time grid.

4.2 Scheduling model

The scheduling model is responsible for the short-term planning of the maintenance actions. On this time-scale the slow degradation processes (e.g. fouling) have a lower influence, while real-time information from condition monitoring systems can give valuable input about the status of specific production assets, e.g. when an asset is predicted to break down within the next 30 days, maintenance needs to be scheduled before this moment. From an operational point of view it is important to allow this more detailed scheduling, as maintenance personnel need to be available and at the right location to perform the maintenance. While the planning model has a very coarse time-grid, the scheduling model also allows for timing those long-ahead planned maintenance types to the most optimal time point.

4.3 Mathematical formulation

A number of sets, parameters and variables are defined and presented in the nomenclature Table 2. Index i refers to a specific task and the index t refers to a specific time point. The following sets are defined: Let I be the set of all maintenance tasks. It has five subsets, since there are five types of maintenance as explained in the problem statement: The short maintenance I_{SM} which is performed off-shore, the long maintenance I_{LM} which is performed on-shore, the online washing I_{OnW} which has a short duration and is non-invasive for the unit

Table 2 Notation for the planning and scheduling formulation

	Description
Index/set	
i	Maintenance task
t	Time interval
R	Resource (all)
R_c	Resource (asset)
R_n	Resource (maintenance personnel)
R^{max}	Maximum available amount of resource
T	Time interval
T_S	Time interval (scheduling)
T_P	Time interval (planning)
I	Maintenance tasks
I_r	Maintenance tasks performed on specific resource
I_{SM}	Short maintenance task
I_{LM}	Long maintenance task
I_{Insp}	Inspection maintenance task
I_{OffW}	Offline washing maintenance task
I_{OnW}	Online washing maintenance task
Continuous variable	
$E_{r,t}^{RUL}$	Enumerator for remaining useful life
$E_{r,t}^{GR}$	Enumerator for recoverable goodness
$E_{r,t}^{GN}$	Enumerator for non-recoverable goodness
$E_{r,t}^{Insp}$	Enumerator for compressor inspection
$E_{r,t}^{OffW}$	Enumerator for offline washing
$E_{r,t}^{OnW}$	Enumerator for online washing
$G_{r,t}$	Goodness of unit
$G_{r,t}^{max}$	Maximum goodness of specific unit
$G_{r,t}^{min}$	Minimum goodness of specific unit
U_r	Remaining useful life of specific unit
U_r^{init}	Remaining useful life of specific unit
Binary variable	
$N_{i,t}$	Maintenance task i starts at t
Parameters	
D_r^{GN}	Degradation factor for non-recoverable degradation
D_r^{GR}	Degradation factor for recoverable degradation (long maintenance)
D_r^{Insp}	Degradation factor for recoverable degradation (compressor inspection)
D_r^{OffW}	Degradation factor for recoverable degradation (offline washing)
D_r^{OnW}	Degradation factor for recoverable degradation (Online Washing)
K_{profit}	Production profit per asset
K_i	Cost for maintenance action i
τ_i	Duration for maintenance action i
H_P	Planning horizon

Table 2 (continued)

	Description
H_S	Scheduling horizon
L_P	Planning slot length
L_S	Scheduling slot length
$\mu_{i,t-\theta}$	Discrete interaction
π	Length of time interval
π_S	Length of scheduling time interval
π_P	Length of planning time interval
M	Big M

itself, which means that it can be done without removing the unit from the system, the offline washing I_{OffW} which requires to shut down the compressor and the inspection I_{Insp} which is shorter than a long maintenance action, but also does not restore the goodness as much as the long maintenance does. In time slots where maintenance is happening every type of maintenance (except online washing) stops the production, but the duration τ_i is individual for every compressor and type of maintenance. Another simplification has been done regarding the duration of maintenance tasks: The possible production losses due to a ramp-up and ramp-down of the production before and after the maintenance itself are included in the maintenance duration in order to account for all production losses due to the maintenance. The set of resources R has two subsets: compressors R_C and maintenance personnel R_n. Each compressor is part of a compressor train on a specific platform. For the example case study, each platform has exactly one compressor train with five compressors in series. In real-world there exists platforms with more than one compressor train. This may affect the cost for maintenance, if maintenance actions are performed for two trains on the same platform, if conducted simultaneously or consecutively. However, this paper refers to the simpler case described beforehand.

The amount of resources available is limited for each available resource type between zero and the maximum available resources of the specific type:

$$0 \leq R_{r,t} \leq R_r^{max} \quad \forall r \in R, t \in T \tag{1}$$

For each compressor further information is needed to plan the maintenance. For each compressor R_C and at each time t we track the Remaining Useful Life (RUL) $U_{r,t}$ and the goodness $G_{r,t}$ of the unit. Predictions about the Remaining Useful Life are available only up to a certain point in time, when the measurement of the condition of the equipment indicates that a failure may happen soon. If the condition monitoring does not suggest that a machine failure is upcoming, the RUL is basically endless. However, in this formulation, a mixed approach between Time Based Maintenance and Condition Based Maintenance is applied and therefore the RUL is never infinite, but set to a reoccurring maintenance interval. The binary variable $N_{i,t}$ takes the value 1 if a maintenance task i starts at time slot t and remains zero if the maintenance task does not start in that time slot.

4.4 Time grids

The problem described in this paper tackles an existing challenge in the process industry: How to cope with long-term planning of maintenance, which includes planned turnarounds that allow opportunistic maintenance and takes degradation processes with slow dynamics into account, but also unforeseen machine failures and the inputs from condition monitoring (which is able to sense first indicators for upcoming machine failure about three months in advance)? We define two non-overlapping time intervals: T_S defines the set of time intervals of the short-term scheduling horizon, while T_P defines the long-term planning time intervals, where the following holds true:

$$T_S \cup T_P \subseteq T \tag{2}$$

When choosing the size of the time-grid, two cases are possible: If the planning time-grid has the same level of detail as the scheduling time-grid, then the time-grids are uniform, otherwise they are non-uniform (see Fig. 4). More information about the non-uniform grid approach applied to an example case study about demand side management of an steel plant can be found in (Dalle Ave et al. 2019). Even though the time scales in this work are much longer than in the aforementioned paper, the time-grid ratio between the near and the far future shows similarly elongated time intervals. While the model captures the finer aspects of the scheduling problem for the near future, the further future is represented by longer time intervals.
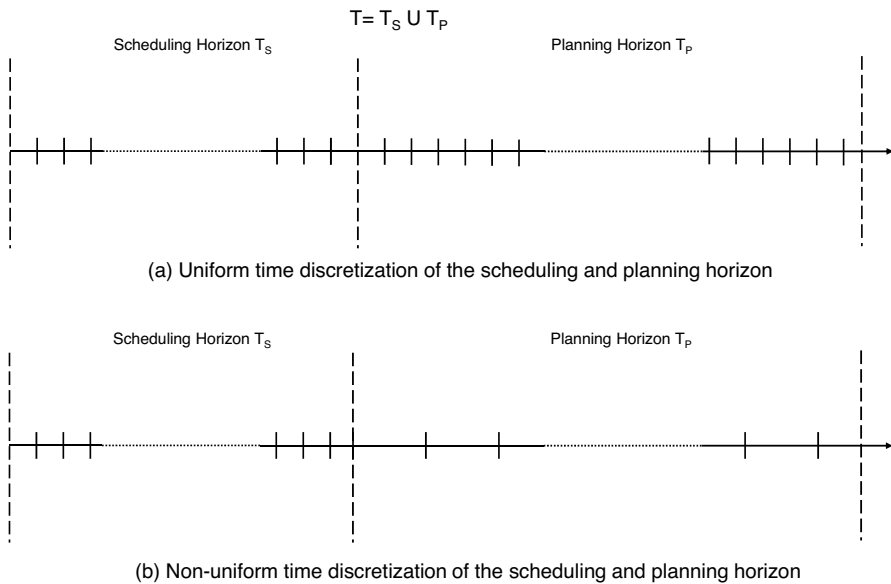

(a) Uniform time discretization of the scheduling and planning horizon

(b) Non-uniform time discretization of the scheduling and planning horizon

Fig. 4 Uniform (**a**) and non-uniform (**b**) discretization of the planning and scheduling horizon

The structural parameters for the length of the time horizons, H_P and H_S, as well as the length of the slots, L_P and L_S, are used to set up the the amount of slots that are needed to represent the time horizon with the required granularity. For the ease of simplicity, they are not explicitly used in the mathematical equations in this chapter. While the horizon length is solely used to initialize the number of time slots in the specific model, the slot length is also used to scale other parameters, e.g. the degradation per slot.

The interface between these two time grids will be discussed in Sect. 4.12 about model integration.

4.5 Limits

Both the goodness and the remaining useful life have limits:

$$U_r^{min} \leq U_{r,t} \leq U_r^{max} \quad \forall r \in R_c, t \in T \tag{3}$$

The remaining useful life should not fall below a certain threshold U_r^{min}. When this boundary is reached, a maintenance task needs to be performed. There are different options to choose U_r^{min}. When the value for the remaining useful life reaches zero, the unit will break and cannot be used any further. Setting $U_r^{min} = 0$ is equivalent to a run-to-failure-strategy. Since predictions from a condition-monitoring system only have a certain accuracy, in reality U_r^{min} will be set to a value at which the risk of unit failure is acceptable. The higher this value is, the smaller is the risk of a machine failure, but also, the greater the risk, that machines might be maintained earlier than necessary. Similarly for the goodness:

$$G_r^{min} \leq G_{r,t} \leq G_r^{max} \quad \forall r \in R_c, t \in T \tag{4}$$

While the goodness is decreasing over time, we can define a value G_r^{min} which is the minimum goodness we allow the compressor to operate with (While $G_r^{min} \geq 0$ must hold true). Furthermore, the maximum goodness (i.e. the maximum performance that can be achieved) G_r^{max} of each compressor is different and the initial goodness G_r^{init} (with $G_{r,t}^{max} \geq G_{r,t}^{init}$) also needs to be defined, since the scheduling and planning model is applied to an existent asset fleet and not all of them operate in this moment with their maximum performance.

Both the goodness and the remaining useful life are decreasing over time. Each of the three maintenance types has an impact on these two properties. This is displayed in Fig. 3.

4.6 Enumerator formulation

The concept of the degradation of the goodness and the decreasing remaining useful life is based on the time since the last maintenance that affects these properties has been performed. Since there are different types of maintenance available and they all do not have the same effect on goodness and RUL, we introduce one counter for each type of maintenance and one for the remaining useful life (as it is affected by

the multiple maintenance types). For reasons of brevity, only the enumerator used for the remaining useful life $E_{r,t}^{RUL}$ is introduced:

$$E_{r,t}^{RUL} \geq E_{r,t-1}^{RUL} + 1 - M \sum_{i \in I_r \cap (I_{SM} \cup I_{LM})} N_{i,t} \quad , \forall r \in R_c, t \in T \tag{5}$$

$$E_{r,t}^{RUL} \leq E_{r,t-1}^{RUL} + 1 \quad , \forall r \in R_c, t \in T \tag{6}$$

$$E_{r,t}^{RUL} \leq 1 + M(1 - \sum_{i \in I_r \cap (I_{SM} \cup I_{LM})} N_{i,t}) \quad , \forall r \in R_c, t \in T \tag{7}$$

$$E_{r,t}^{RUL} \geq 1 \quad , \forall r \in R_c, t \in T \tag{8}$$

The enumerator counts up by 1 for every time interval without maintenance that affects the RUL, otherwise it is reset. The constraint in Eq. 5 is active when there is no maintenance performed and it increments the enumerator. If there is maintenance it is relaxed by the big-M term. Equation 6 ensures that the enumerator is increasing by maximum 1 in each time interval. Equation 7 also resets the enumerator to one. When there is maintenance, the upper bound for the enumerator is 1, otherwise the big-M term relaxes the upper bound. The lower bound of the counter is always 1 (Eq. 8). An overview The relationship between the different enumerators and the associated maintenance tasks is given in Table 3.

4.7 Remaining useful lifetime

As mentioned before, the formulation for the remaining useful lifetime should allow the model to be used in two settings: Running in Time Based Maintenance (Overhaul of the asset after a pre-defined time) or Condition Based Maintenance (Maintenance performed on the basis of additional input information that predict when a fault will happen and when a maintenance needs to be performed in order to prevent this). Time Based Maintenance was the industry standard for a long time, but nowadays information about the condition are are a crucial part for making decisions about maintenance. For the Condition Based Maintenance there

Table 3 Relationships between the maintenance actions and the enumerators they reset

Maintenance type		Impact on enumerators					
Index	Description	E^{RUL}	E^{Insp}	E^{GN}	E^{GR}	E^{OffW}	E^{OnW}
I_{LM}	Long maintenance	✓	✓	✓	✓	✓	✓
I_{SM}	Short maintenance	✓	✓				
I_{Insp}	Inspection		✓		✓	✓	✓
I_{OffW}	Offline washing					✓	✓
I_{OnW}	Online washing						✓

is a maximum lifetime given before a maintenance needs to be performed. With every time interval since the last maintenance this remaining useful lifetime is decreasing:

$$U_{r,t} \leq U_r^{max} - E_{r,t}^{RUL}\pi \quad \forall r \in R_c, t \in T \tag{9}$$

Equation 9 holds true for the formulation in the fullspace model with uniform time-grid ($\pi_P = \pi_S$). For the hierarchical model and the fullspace-model with nonuniform time grid, the equation needs to be adapted to be feasible in the time horizon of the scheduling and the planning time horizon with their corresponding length for each discrete time-slot in the scheduling (π_S) or planning (π_P) time horizon.

$$U_{r,t} \leq U_r^{max} - E_{r,t}^{RUL}\pi_S \quad \forall r \in R_c, t \in T_S \tag{10}$$

$$U_{r,t} \leq U_r^{max} - E_{r,t}^{RUL}\pi_P \quad \forall r \in R_c, t \in T_P \tag{11}$$

It is assumed that Time-Based Maintenance is always performed. The set value gives the time at which the equipment is by latest maintained. This is reflecting legal requirements in many companies that defines the maximum interval between two maintenance actions. If this value is set very high, it will rarely be activated. Condition monitoring will suggest maintenance actions to be performed at an earlier time. This is done as the value for $U_{r,t}$ is updated with an prediction from condition monitoring systems. In the practical implementation of this paper, there exists no real-time condition monitoring system and therefore the initial case study contains a pre-defined RUL for every asset.

4.8 Goodness calculation

With the counters that have been defined before, it is possible to formulate the goodness $G_{r,t}$ for each compressor at each time t.

$$\begin{aligned}
G_{r,t} = G_r^{max} &- D_r^{GR}\pi(E_{r,t}^{GR} - 1) - D_r^{GN}\pi(E_{r,t}^{GN} - 1) \\
&- D_r^{Insp}\pi(E_{r,t}^{Insp} - 1) - D_r^{OffW}\pi(C_{r,t}^{OffW} - 1) \\
&- D_r^{OnW}\pi(E_{r,t}^{OnW} - 1) \\
&\forall r \in R_c, t \in T
\end{aligned} \tag{12}$$

The notation of the different parameters and variables is specified in Table 2. Every term of Equation 12 refers to one specific enumerator and subtracts, based on when the enumerator for a specific maintenance type was reset the last time, a certain amount of goodness. This deterioration is dependent on the degradation factor for the specific degradation type and the length of the time intervals. Respectively, for the hierarchical model and the full-space model with non-uniform time grid this mathematical formulation must be separated into the two different horizons:

$$G_{r,t} = G_r^{max} - D_r^{GR}\pi_S(E_{r,t}^{GR} - 1) - D_r^{GN}\pi_S(E_{r,t}^{GN} - 1)$$
$$- D_r^{Insp}\pi_S(E_{r,t}^{Insp} - 1) - D_r^{OffW}\pi_S(E_{r,t}^{OffW} - 1)$$
$$- D_r^{OnW}\pi_S(E_{r,t}^{OnW} - 1)$$
$$\forall r \in R_c, t \in T_S \qquad (13)$$

$$G_{r,t} = G_r^{max} - D_r^{GR}\pi_P(E_{r,t}^{GR} - 1) - D_r^{GN}\pi_P(E_{r,t}^{GN} - 1)$$
$$- D_r^{Insp}\pi_P(E_{r,t}^{Insp} - 1) - D_r^{OffW}\pi_P(E_{r,t}^{OffW} - 1)$$
$$- D_r^{OnW}\pi_P(E_{r,t}^{OnW} - 1)$$
$$\forall r \in R_c, t \in T_P \qquad (14)$$

While the dynamics of the considered degradation are rather slow, they have a larger impact on the time-horizon of the planning model. However, they are still considered in the scheduling model. The maintenance types that are considered in this shorter time-horizon are economically beneficial in regards to the degradation. The actual maintenance tasks are planning in the scheduling horizon, the results of the planning model are just estimated predictions.

It has to be noted, that Eq. 12 is calculating the degradation of the goodness in a linear fashion. Figure 3 shows however, that the degradation is not necessarily linear. The simplification made here is done to keep the mathematical problem in the MILP domain as solving a MINLP of the same problem size would not be feasible in industrially reasonable time.

4.9 Allocation of tasks to units

In order to allocate a maintenance task to a compressor and to set the starting time of the maintenance, the binary variable $N_{i,t}$ is introduced. It takes the value 1 if the maintenance task i on unit r starts at time t. At any time t only one maintenance task i can be performed on each compressor R_c. The previous constraint can be expressed by:

$$\sum_{i \in I_r} N_{i,t} \leq 1 \quad \forall t \in T, r \in R_c \qquad (15)$$

4.10 Resource balance

The resource balance describes the fact that in each active time interval that every task $N_{i,t}$ consumes a specific amount of resources $\mu_{r,i,t}$ from the initial amount of resources R_{r0} that have been available at the beginning of the time interval.

$$R_{r,t} = R_{r,t-1} + R_{r,0} - \sum_i \sum_\theta N_{i,t-\theta}\mu_{i,r,t-\theta} \quad \forall r \in R, t \in T \qquad (16)$$

In contrast to other RTN resource balances this resource balance does not consider the raw materials and products. It is assumed, that these are homogeneous commodities. Removing this helps to decrease the model size, which is often a limiting factor for complex real-life models. The included resources concern only the maintenance personnel and the assets themselves. Figure 5 displays an example of the resources within the RTN framework. The five tasks *I* can be performed on any of the specific resources, i.e. the different compressors. Whenever such a maintenance task is conducted, also the maintenance personnel resource is consumed for the duration of the maintenance task.

4.11 Objective function

The target of the optimization is to maximize the profit, i.e. the sum of the production profit from each compressor minus the cost for maintenance.

The objective functions of the planning and scheduling models are identical. They consist of two different terms. First, the profit from the production must be summed up. Every platform has a specific production capacity. As long as a compressor is not in maintenance, it is operated at maximum capacity as defined by its goodness value. As soon as the compressors are maintained, the production needs to be shut down for the time period of the maintenance task. During the periods when the production is running, the maximum production capacity of a train needs to be multiplied with the product of the goodness of each compressor in the train. The second term takes the cost of every maintenance into account, which in addition to the production loss is a negative driver.

The objective function is formulated as follows:

$$\max z = \sum_{r \in R_c} (K_{profit} G_{r,t} - K_i \sum_{i}^{I_r} \sum_{t}^{T} N_{i,t}) \tag{17}$$

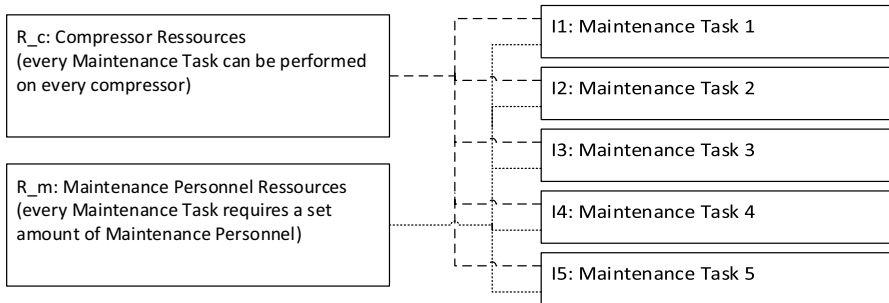| R_c: Compressor Ressources (every Maintenance Task can be performed on every compressor) | I1: Maintenance Task 1 |
	I2: Maintenance Task 2
	I3: Maintenance Task 3
R_m: Maintenance Personnel Ressources (every Maintenance Task requires a set amount of Maintenance Personnel)	I4: Maintenance Task 4
	I5: Maintenance Task 5

Fig. 5 Example representation of the resources within the RTN formulation

4.12 Model integration

As described in the background section, there are different ways to integrate the planning and the scheduling models. In this work three different approaches are considered: A full-space model with a uniform time-grid, a full-space model with a non-uniform time-grid and a hierarchical model with a non-uniform time grid.

While the full-space models can be realized easier, the hierarchical model requires a more sophisticated formulation. In the first step the planning model is run with a coarse time-grid. This will determine when the relevant maintenance types need to be performed. In a top-down approach those maintenance actions that are overlapping with the scheduling horizon are assigned to the scheduling model. This is realized by specific matching constraints (see Fig. 6), that leave the freedom to pick the most detailed scheduling time-slots that happen to be within the time span of the corresponding planning time-slot that is overlapping with the scheduling horizon. The scheduling model is then solved and finds solutions for the fixed maintenance types but also for those maintenance types that were not considered in the planning model. Figure 6 illustrates how the integration works from a structural point of view. The figure shows in a simplified example that the first two planning slots are within the scheduling horizon. Each of the planning slots spans over five scheduling slots. If the planning model suggests a maintenance in one of the planning slots, the sum of maintenance actions in the five corresponding scheduling slots also needs to be one. Vice versa, if no maintenance is planned in one of the two planning slots, the sum of maintenance actions in the corresponding scheduling slots must be equal to zero.

Fig. 6 Overview of the structural integration in an hierarchical integration of planning and scheduling

5 Example results

In this section, several cases from the described motivating example are run. In order to perform a performance analysis, the fleet size, the granularity of the time grid and the length of the time horizon are varied. Furthermore, the different options for the integration of planning and scheduling, as explained in the methodology Section 4, are tested.

The different cases were solved using GAMS 24.8.4 with CPLEX 12.7.1.0 on an Intel(R) Core(TM) i7-6700HQ at 2.6 GHz and 16 GB RAM.

In order to compare between the different formulations (hierarchical model as well as full-space model with uniform/non-uniform time grid), the improvement of operational profit of the different cases with optimized maintenance schedules is calculated. The base of comparison is the same case study, but with a time-based maintenance plan for the long maintenance action that is performed on-shore. In this base-case every compressor is maintained every second year, independent from its actual degradation level. Short maintenance actions to avoid machine failure are still allowed and incorporated into the case studies. The improvement is calculated as the ratio of operational profit with optimized maintenance schedules to the operational profit for the same case, but with a time-based maintenance approach.

The computation is in all cases terminated after a total of 20,000 CPU-s. In the case of the hierarchical model where technically two sequential optimization problems are solved, 10,000 CPU-s are dedicated to the planning problem and the remaining 10,000 CPU-s are used to solve the scheduling problem. Furthermore, the calculation is terminated once the MIP-gap is below the threshold of 0.01%. This optimality gap is acceptable and a quick solution is preferred over decreasing the gap further.

Table 4 show the size of the mathematical problems in terms of constraints, binary variables and variables for various fleet sizes. In fact, the final problem size is affected by all the model parameters that are changed by individual case studies. The amount of compressors has a major effect on the problem size, next to the number of considered maintenance types and the granularity and length of the time grid. The table shows the example of the full-space model with non-uniform time grid for a

Table 4 Overview of the model statistics for the smaller case studies (full-space model with non-uniform time grid for one compressor and varying time grid parameters) [horizon in days and slot length in hours]

Planning		Scheduling		Constraints	Binary variables	Variables
Horizon	Slot length	Horizon	Slot length			
60	24	30	3	3134	5524	1210
60	24	30	6	1572	2757	610
120	24	30	3	3160	5570	1220
120	24	30	6	1596	2797	620
180	24	30	3	3186	5616	1230
180	24	30	6	1620	2837	630

planning horizon of 8 years with a granularity of months and a scheduling horizon of 1 month with a granularity of 3 h. Precise values for every case that was run would go beyond the scope of this paper, but the table provides an overview of the ballpark of the number of constraints, variables and binary variables that have been subject of this work. This highlights the complexity of the mathematical problem.

5.1 Hierarchical Model

A few much smaller case studies were run to give a better comparison to the two other formulation approaches, as the larger case studies cannot be solved for all modeling approaches. The results can be found in Table 5. For these cases, just one compressor is considered and the planning horizon is either 2, 4 or 6 months with one-month timeslots and in the scheduling horizon of one month the timeslots are either 3 or 6 h.

5.2 Fullspace model—uniform time grid

Table 6 shows the application of the full-space model formulation with a uniform time grid in which different cases are run. The fleetsize is varying between 1, 5 and 10 compressors. For the length of the horizon, either two, four or six months were chosen, while the granularity is either 2, 3 or 6 h. A few cases with a time horizon of two or four years were run in order to compare the formulation against the hierarchical model, but as the formulation already struggles with smaller problem size and is not able to solve over such long time horizons, these are not included in the table.

5.3 Fullspace model—non-uniform time grid

Similar as for the hierarchical model, some much smaller cases were run to facilitate comparison with the uniform model. These cases have a planning horizon of 2, 4 or 6 months with timeslots of 1 month and a scheduling horizon of 1 month with time slots in 3 or 6 h. The results are presented in Table 7.

Table 5 Results of the hierarchical model for shorter time horizons (horizon in days, Slot length in hours)

Compres-sors	Planning		Scheduling		Improve-ment (%)	CPU-s		MIP-Gap	
	Horizon	Slot length	Horizon	Slot length		Plan.	Sched.	Plan. (%)	Sched. (%)
1	60	24	30	3	2.52	4.19	5.95	0.00	0.01
1	60	24	30	6	1.25	3.09	5.83	0.00	0.01
1	120	24	30	3	1.48	3.56	5.92	0.00	0.01
1	120	24	30	6	0.53	3.73	5.95	0.00	0.01
1	180	24	30	3	5.99	3.59	6.19	0.00	0.01
1	180	24	30	6	4.92	3.52	5.81	0.00	0.01

Table 6 Results of the fullspace model with uniform time grid (horizon in days, slot length in hours)

Compressors	Horizon	Slot length	Improvement (%)	CPU-s	MIP-Gap (%)
1	60	6	1.30	51.02	0.01
1	60	3	1.30	987.3	0.01
1	60	2	1.07	8680.3	0.01
1	120	6	0.56	692.22	0.01
1	120	3	0.95	20002.09	0.03
1	120	2	1.71	20001.34	6.95
1	180	6	4.19	2891.11	0.01
5	60	6	4.48	5878.61	0.01
5	60	3	3.67	20008.31	4.55
5	120	6	0.20	20001.73	0.54
5	120	3	2.56	20001.57	5.29
5	180	6	0.88	20000.99	1.30
10	60	6	0.72	20001.8	0.33

Table 7 Results of the fullspace model with non-uniform time grid for shorter time horizons (horizon in days, slot length in hours)

Compressors	Planning		Scheduling		Improvement (%)	CPU-s	MIP-Gap (%)
	Horizon	Slot length	Horizon	Slot length			
1	60	24	30	3	8.75	19.23	0.09
1	60	24	30	6	6.03	10.39	0.06
1	120	24	30	3	5.11	21.89	0.09
1	120	24	30	6	4.48	10.5	0.09
1	180	24	30	3	6.60	27.47	0.10
1	180	24	30	6	5.60	10.97	0.06

5.4 Larger case studies—hierarchical model and non-uniform fullspace model

While the results in the previous sections aimed to show the different performances of the modeling approaches, in this section the case studies are scaled up in order to illustrate the ability of the proposed methods to handle case studies of industrial-sized problems.

Table 8 shows the application of the hierarchical model formulation to each of the different cases run. The fleet size is varies between 1, 5, 10, 25, 50 or 100 compressors. While the planning model is running over a horizon of either 4 or 8 years with timeslots being one month, the scheduling model is run over a horizon of one month with the timeslots being either 3 h or 6 h.

Table 9 shows the application of the full-space model formulation with a non-uniform time grid in which different cases are run. The fleet size is varied between 1, 5 and 10 compressors. For the length of the planning horizon either four or eight

Table 8 Results of the hierarchical model (horizon in days, slot length in hours)

Compressors	Planning		Scheduling		Improvement (%)	CPU-s		MIP-Gap	
	Horizon	Slot length	Horizon	Slot length		Plan.	Sched.	Plan. (%)	Sched. (%)
1	1440	24	30	3	14.97	3.42	5.55	0.01	0.01
1	1440	24	30	6	12.17	2.06	2.17	0.01	0.01
1	2880	24	30	3	10.80	7.61	5.59	0.00	0.01
1	2880	24	30	6	9.84	7.97	5.97	0.00	0.01
5	1440	24	30	3	7.88	24.58	25.92	0.01	0.01
5	1440	24	30	6	6.83	19.73	31.28	0.01	0.01
5	2880	24	30	6	6.37	1011.74	36.13	2.29	0.01
5	2880	24	30	3	6.18	10,006.00	6.16	2.16	0.00
10	1440	24	30	6	7.49	541.25	36.30	0.01	0.01
10	1440	24	30	3	6.89	551.00	14.91	0.01	0.01
10	2880	24	30	6	6.19	10000.95	40.28	3.09	0.01
10	2880	24	30	3	5.64	10000.89	10.27	3.09	0.01
25	1440	24	30	3	3.56	1258.58	844.08	0.01	0.01
25	1440	24	30	6	3.76	1311.16	29.34	0.01	0.01
25	2880	24	30	3	2.92	10,002.47	956.58	4.33	0.01
25	2880	24	30	6	2.64	10,002.39	32.88	4.33	0.01
50	1440	24	30	3	2.45	10,003.89	10,002.42	1.08	0.01
50	1440	24	30	6	2.32	10,005.69	96.13	1.09	0.01
50	2880	24	30	3	1.79	10,002.01	10,001.34	4.22	0.01
50	2880	24	30	6	1.71	10,001.26	104.33	4.42	0.01
100	1440	24	30	3	4.68	1515.42	10,005.55	0.01	–
100	1440	24	30	6	4.35	1742.16	427.48	0.01	0.01
100	2880	24	30	3	2.41	10,001.16	10,005.88	6.04	–

Table 8 (continued)

| Compressors | Planning | | Scheduling | | Improvement (%) | CPU-s | | MIP-Gap | |
	Horizon	Slot length	Horizon	Slot length		Plan.	Sched.	Plan. (%)	Sched. (%)
100	2880	24	30	6	2.22	10,001.19	512.23	6.05	0.01

Table 9 Results of the fullspace model with non-uniform time grid (horizon in days, slot length in hours)

Compressors	Planning		Scheduling		Improvement (%)	CPU-s	MIP-Gap (%)
	Horizon	Slot length	Horizon	Slot length			
1	1440	24	30	3	5.70	456.23	0.01
1	1440	24	30	6	4.63	245.39	0.01
1	2880	24	30	3	5.66	648.89	0.01
1	2880	24	30	6	5.02	341.05	0.01
5	1440	24	30	3	1.43	20,001.83	5.05
5	1440	24	30	6	1.37	20,001.14	3.29
5	2880	24	30	3	1.23	20,000.78	10.92
5	2880	24	30	6	1.06	20,001.34	6.95
10	1440	24	30	3	1.31	20,002.41	15.70
10	1440	24	30	6	0.87	20,002.19	13.98

years were chosen, with a scheduling horizon of one month. The granularity of the planning time-grid is one month, while the scheduling time-grid is divided into slots of either 3 or 6 h.

For the fullspace model with uniform time grid, no cases bigger than the ones in the previous subsection could be solved.

6 Discussion

In this section the results that were presented in the previous section will be discussed. First, the results for the individual formulation approaches are discussed, followed by a comparison of the performance of the different formulations, also regarding the ability to solve case studies of different sizes.

6.1 Hierarchical model

The obtained improvements for the hierarchical model range between 1 and 15%. If cases with the same fleet size are compared, it can be observed that in those cases where the ratio between scheduling horizon and planning horizon is the largest, the biggest improvements are achieved. This can be explained due to the increased efficiency that can be obtained by applying short-term maintenance modes such as washing. They restore a substantial part of the performance, while lasting very briefly. It can also be observed that the case studies including just one single compressor have very high improvement rates. This might also be due to the specific case study that incorporates a single asset, as this may have an high initial degradation that can be resolved by maintenance. For cases with more compressors, this is evened out, as the properties of each asset are uniformly distributed between given boundaries.

Another factor that can be seen are the limitations imposed by the computation time limits. While the time limits for both planning and scheduling models are set to 10,000 CPU-s, this value is often reached for the planning horizon, especially by all case studies that have larger fleet sizes or a longer time horizon and/or a finer discretization.

However, the authors want to emphasize that the chosen limits of computation times are adequate. While calculations for planning horizons of several years could also take up more computation time (e.g. up to one week), the finer scheduling of maintenance tasks over a horizon of one or two months should happen around the timeframe of 10,000 CPU-s. Furthermore, the optimality gap in the planning model is assumed to be uncritical, as in a real-world implementation a rolling horizon approach would be chosen and therefore, it will be re-evaluated if the maintenance actions with a long duration should be performed within the next months (during the scheduling horizon). Such a rolling-horizon approach would be able to re-schedule as new information about the RUL would affect existing schedules and even without approaching asset-failure steadily new schedules need to be provided to the maintenance personnel.

For some large problems, the solution time was actually not the only limitation, but also the available memory to process the optimization. In these cases, the optimization was terminated prematurely and the improvement is in the lower range of the spectrum. Also, the largest MIP-gaps can be seen if the memory-limit was reached. This happened for the two cases with a fleet size of 100 assets and a scheduling time slot duration of 3 h. The MIP-Gap for the scheduling problem was therefore not indicated in the result table.

While the (time or memory) resources limit the planning model to achieve bigger improvements and create large MIP-gaps (ranging up to 6%), the MIP-gap in the scheduling model can be closed for most case studies (with the treshold set to 0.01%). For two cases, the scheduling model did not find any solution and the MIP-gap could not be calculated. In both cases, the time/memory resources were exhausted. This happened for the two cases with a fleet size of 100 assets and the finer discretization of the scheduling model. This means, that the case is too large to be solved with the given hardware resources in the given amount of time.

Summarizing, the hierarchical model shows the expected behaviour and the sensitivity analysis with alternating some parameters to influence the case study size proves that the formulation is efficient for already quite big case studies, however the improvement benefits from more computational power and bigger time resource availability.

6.2 Fullspace model—uniform time grid

As mentioned in the results section, a few cases were run over longer time horizons. However, these cases demonstrated the size of the problem and the computational resources that are required. Even with just one asset, a long time horizon resulted in a problem size that is solvable, but with large MIP-gap and small improvement rates (ranging below 2%).

The fact that a uniform time grid requires too many discrete time slots when it is not relevant (i.e. in the far future, where detailed scheduling is not actually required) and therefore the problem size becomes too big, disqualifies this formulation actually for the usage in the aforementioned case study with large asset fleets. However, from a scientific point, the evaluation of the results still remains interesting and might be solved in the future with progression of computational power or with further research in order to develop better algorithms and supporting heuristics.

Firstly, the results show an improvement of up to about 4.5% compared to the base case. Within groups of cases with the same fleet size a trend can be seen that a finer granularity leads to more improvement. For the reported cases it can be seen that a longer time horizon decreases the improvement. However, this is a fallacy, as the compared cases have very short time horizons and the way in which the comparison with the base case is constructed, it may happen that shorter time horizons appear to be better.

Secondly, even though the cases range over shorter time horizons compared to the planning horizons of the other approraches, the level of discretization is higher and therefore the problem contains more discrete time slots. This causes high computational effort, so that the entire given time resources are required in most cases.

The MIP-gaps are actually rather small for most cases and do not exceed 7%.

Summarizing, the fullspace model with a uniform time grid is not able to provide an appropriate solution for the case study with longer time horizons. Cases with a larger amount of assets cannot be solved in appropriate time. Only for very limited case studies good results can be achieved.

6.3 Fullspace model—non-uniform time grid

The behaviour of the fullspace model with a non-uniform time grid as described in the results, was expected. The formulation is able to solve case studies that are larger than the formulation with an uniform time grid. With the reduced amount of time slots this results is as expected.

However, for larger case studies, the problem size is still too big. The two problems of scheduling and planning, if combined into one problem, are substantially harder to solve than both models separately in a sequence. The case studies with one compressor were solved very quickly (below 30 s) and obtained improvements of up to 9%. The case studies with 5 or 10 compressors could be solved, however the time limit was reached and the remaining MIP-Gaps were between 3% and 15% with improvements lower than 1.5%.

Cases with a more detailed time grid (3 h slots) for the scheduling problem produce slightly higher improvements, while the computational time increases. For small problem sizes of one compressor over short time horizons, we can again observe the biggest improvements for very short case studies, which is however due to the ratio of the time span in which maintenance modes like washing are considered and therefore higher improvements can be achieved.

6.4 Comparison of the formulations

When the hierarchical formulation is compared with the full-space models, both with uniform and non-uniform time grid, the most prevalent difference is the performance, especially for larger case studies. The hierarchical model is able to solve case studies for up to 100 assets, while the full-space model with non-uniform time grid can be applied for up to about 5 assets and the full-space model with uniform time grid just for one asset, except when the time horizon is reduced to a level which is not suitable for industrial application of maintenance planning. The level of detail and the length of the time horizon is however suitable for maintenance scheduling.

The improvement rates of all models are, if the MIP-gap is closed up to the predefined threshold, in a comparable range. While the way the base case is calculated influences the cases with a shorter time horizon, it is obvious that the improvement rates also in the longer cases are substantial enough for industrial application.

All three formulations result in solutions that are feasible. The degradation patterns that can be observed are as expected by the linear formulation and also a good representation of the real-world behaviour. An example of the efficiency degradation is shown in Fig. 7. This is one example that is picked out of the different case studies. The overall goodness of the compressor is deteriorating from an initial efficiency of above 65% down to about 62%, with a number of longer and shorter maintenance actions like a complete overhaul or compressor washing in between. As soon as a maintenance action is performed, the goodness of the compressor will be (partially) reset in the following time slot. At the border between the scheduling and the planning model it can be seen, that specific maintenance types such as the compressor washing are not considered anymore. The cyclical pattern is occurring as expected. With enough resources (i.e. maintenance personnel available) this pattern will not change during regular operation (under the condition, that the degradation can be predicted and is uniform). Just in case of a suddenly reduced RUL of an asset or a bottle-neck of maintenance personnel, perhaps due to uncertainty in the length of a maintenance action, this pattern will change.

In order to visualize the results and to show how the optimized maintenance plan can achieve improvements in the sense of decision-support to operators, a base case schedule for long maintenance actions is compared with the proposed schedule from the hierarchical model. Figure 8 shows that in the optimized schedule all long maintenance actions for each platform happen in parallel and stacked. The base case suggests to perform each long maintenance subsequently instead of parallelizing all available maintenance personnel. Another room for improvement are additional maintenance actions that are available, in this case the online washing of compressors, which is not considered in the base case.

7 Summary and conclusions

This paper presented different formulations for integrating planning and scheduling of maintenance actions for compressor fleets in the oil and gas industry in order to maximize the operational profit while considering asset degradation. An
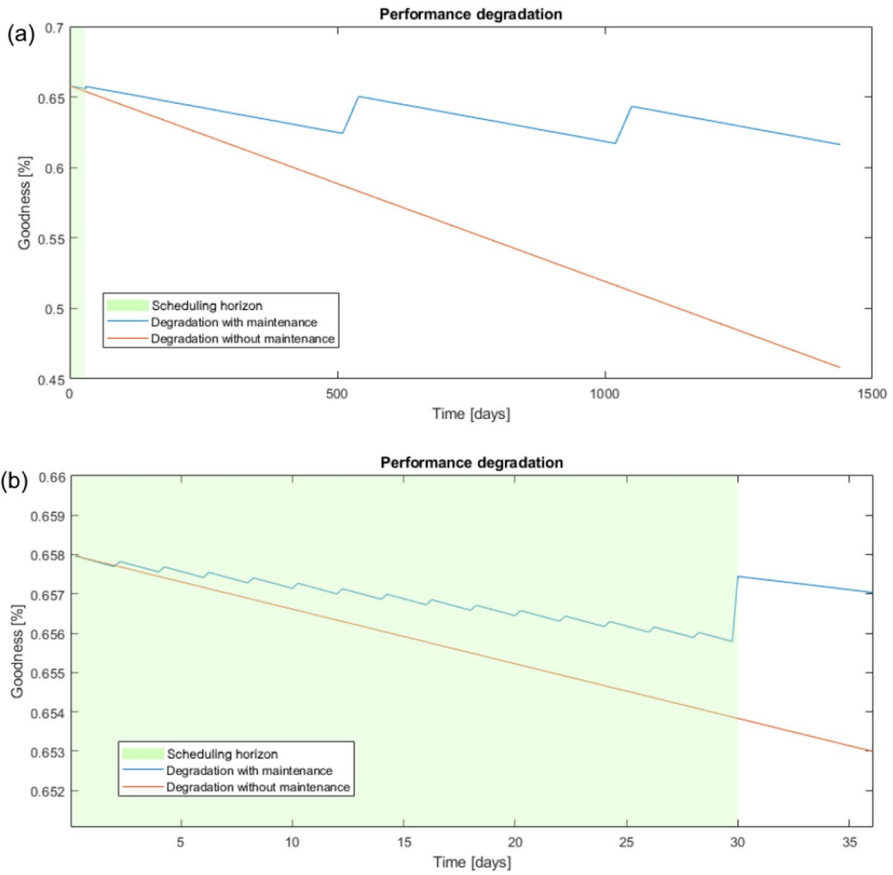
Fig. 7 Degradation pattern for one compressor for a time horizon of 8 years for the planning model (**a**) and 1 month for the scheduling model (**b**)

integral part of both the planning and scheduling models are the enumerator formulations that help to represent the degradation of assets. The considered degradation can be classified as a non-factor based model that is a function purely based on time and defined through the case study itself. Considered effects range from different types of performance degradation (e.g. fouling) to the remaining useful lifetime that gives insights into the operability of the asset and they are strictly linear with time. A set of different maintenance modes is included with individual pricing, duration and effect on the asset degradation.

Different formulations to address the problem have been introduced: A hierarchical model and a full-space model with either non-uniform or uniform time-grid. Both incorporate planning and scheduling with an MILP formulation following the Resource-Task-Network approach. Maintenance tasks are scheduled on a long-term and short-term horizon.
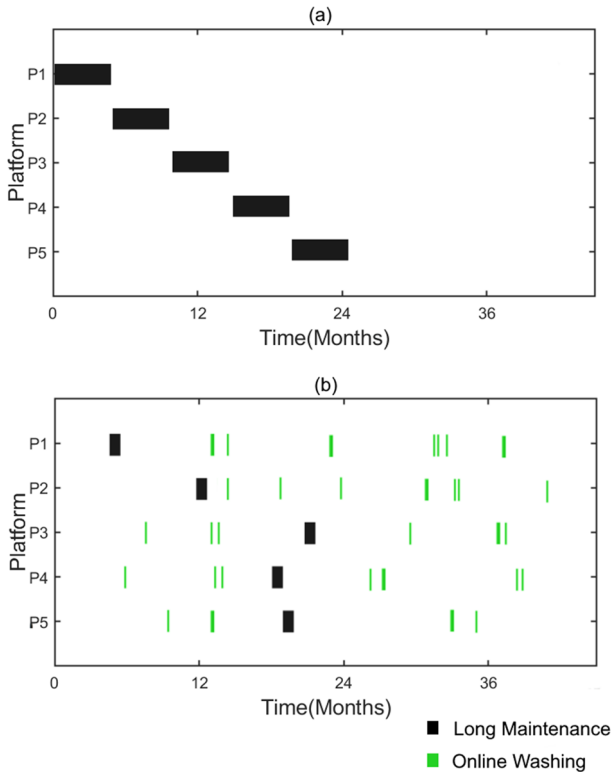
Fig. 8 Gantt chart of the maintenance schedule over a period of 48 months for a fleet of 25 compressors. **a** shows the base case (just long maintenance in a 4-year interval) and **b** shows the optimized schedule including long maintenance and online washing

The results of applying the formulation to a reality-inspired case study from the oil and gas industry shows that the formulations are able to improve the operational profit. With a performance analysis, the impact of adjusting model parameters such as the length of the time horizons of scheduling and planning model and the granularity of the time-grid have been tested. While improvements of up to 10% for the given case study can be reported, the required computational resources are, as expected, still very high. However, commercial solvers are able to solve the mathematical problem in a time and accuracy that is appropriate for industrial application. The presented case studies highlighted the effectiveness of the proposed formulation to tackle the integration of maintenance planning and scheduling.

The three integration approaches showed in the direct comparison results in just minor differences regarding the improvement rate when using similar hyperparameters (planning/scheduling horizon and granularity of the time grids). However, there was a clear difference regarding the performance. While the hierarchical model was able to solve the maintenance planning and scheduling problem for large asset fleets over long time horizons, both full-space models resulted in

too large optimization problems. The non-uniform time grid reduced the problem space by a lower level of detailedness during the planning horizon, but the resulting problem was still intractable.

The proposed formulations are able to solve the problem with a certain level of detail. However, for large cases, the maximum time resources are used up and the MIP-gap remains rather big. Different approaches can be part of future work, starting by using hardware with better computational power or by increasing the time resources for the optimization. If the proposed formulations are implemented in a real-world scenario, this would be suggested, as computation times for a planning over several years is adequate to take several days of computational time. The scheduling formulation on the other hand requires with 10,000 CPU-s already a lot of time to compute, with respect to the scheduling horizon. This is especially the case for case studies with many assets. Without performance improvements the inclusion of stochastic approaches will be impossible. Another approach to solve this issue is the distributed optimization, where the asset fleet is segregated into smaller groups for which each a smaller problem is solved.

Availability of data and material The data set used for this manuscript is described and available via (Schulze et al. 2020).

Code availability The code is available upon request to the corresponding author.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

References

Al-Turki U, Duffuaa S, Bendaya M (2019) Trends in turnaround maintenance planning: literature review. J Qual Maint Eng 25:253–271. https://doi.org/10.1108/JQME-10-2017-0074

Amaro A, Barbosa-Póvoa A (2008) Planning and scheduling of industrial supply chains with reverse flows: a real pharmaceutical case study. Comput Chem Eng 32:2606–2625. https://doi.org/10.1016/j.compchemeng.2008.03.006

Aretakis N, Roumeliotis I, Doumouras G, Mathioudakis K (2012) Compressor washing economic analysis and optimization for power generation. Appl Energy 95:77–86. https://doi.org/10.1016/j.apenergy.2012.02.016

Bassett MH, Pekny JF, Reklaitis GV (1996) Decomposition techniques for the solution of large-scale scheduling problems. AIChE J 42:3373–3387. https://doi.org/10.1002/aic.690421209

Carvalho M, Pinto-Varela T, Barbosa-Póvoa AP, Amorim P, Almada-Lobo B (2015) Optimization of production planning and scheduling in the ice cream industry, In: Computer aided chemical engineering, vol 37. Elsevier, pp 2231–2236. https://doi.org/10.1016/B978-0-444-63576-1.50066-2

Castro PM, Grossmann IE, Veldhuizen P, Esplin D (2014) Optimal maintenance scheduling of a gas engine power plant using generalized disjunctive programming. AIChE J 60: 2083–2097. https://doi.org/10.1002/aic.14412, arXiv:0201037v1

Castro PM, Sun L, Harjunkoski I (2013) Resource-task network formulations for industrial demand side management of a steel plant. Indus Eng Chem Res 52:13046–13058. https://doi.org/10.1021/ie401044q

Dalle Ave G, Harjunkoski I, Engell S (2019) A non-uniform grid approach for scheduling considering electricity load tracking and future load prediction. Comput Chem Eng 1–56 (Paper submitted)

Dogan ME, Grossmann IE (2006) A decomposition method for the simultaneous planning and scheduling of single-stage continuous multiproduct plants. Indus Eng Chem Res 45:299–315. https://doi.org/10.1021/ie050778z

Erdirik-Dogan M, Grossmann EI, Wassick J (2007) A bi-level decomposition scheme for the integration of planning and scheduling in parallel multi-product batch reactors. Comput Aided Chem Eng 24:625–630. https://doi.org/10.1016/S1570-7946(07)80127-1

Grossmann I, Erdirik-Dogan M, Karuppiah R (2008) Overview of planning and scheduling for enterprise-wide optimization of process industries (Übersicht über Planungs- und Feinplanungssysteme für die unternehmensweite Optimierung in der verfahrenstechnischen Industrie. Automatisierungstechnik 56:64–79. https://doi.org/10.1524/auto.2008.0690

Harjunkoski I, Grossmann IE (2002) Decomposition techniques for multistage scheduling problems using mixed-integer and constraint programming methods. Comput Chem Eng 26:1533–1552. https://doi.org/10.1016/S0098-1354(02)00100-X

Igie U, Pilidis P, Fouflias D, Ramsden K, Lambart P (2011) On-line compressor cascade washing for gas turbine performance investigation, in: ASME 2011 turbo expo: turbine technical conference and exposition, ASME, Vancouver, pp 221–231. https://doi.org/10.1115/GT2011-46210

Jain V, Grossmann IE (2001) Algorithms for hybrid MILP/CP models for a class of optimization problems. INFORMS J Comput 13:258–276. https://doi.org/10.1287/ijoc.13.4.258.9733

Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Sig Process 20: 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012, arXiv:0208024

Kopanos GM, Xenos DP, Cicciotti M, Thornhill NF (2017) Operational and maintenance planning of compressors networks in air separation plants, In: Advances in energy systems engineering. Springer International Publishing, Cham, pp 565–600. https://doi.org/10.1007/978-3-319-42803-1_19

Kurz R, Brun K (2012) Fouling mechanisms in axial compressors. J Eng Gas Turb Power 134:1–9. https://doi.org/10.1115/1.4004403

Lakshminarasimha AN, Boyce MP (1994) Modelling and analysis of gas turbine performance deterioration. J Eng Gas Turb Power 116:46–52

Maravelias CT, Grossmann IE (2001) Simultaneous planning for new product development and batch manufacturing facilities. Indus Eng Chem Res 40:6147–6164. https://doi.org/10.1021/ie010301x

Maravelias CT, Sung C (2009) Integration of production planning and scheduling: overview, challenges and opportunities. Comput Chem Eng 33:1919–1930. https://doi.org/10.1016/j.compchemeng.2009.06.007

McKay KN, Safayeni FR, Buzacott JA (1995) A review of hierarchical production planning and its applicability for modern manufacturing. Prod Plan Control 6:384–394. https://doi.org/10.1080/09537289508930295

Pantelides CC (1994) Unified framework for optimal process planning and scheduling, In: Proceedings of second conference on foundations of computer aided operations, pp 253–274

Papageorgiou LG, Pantelides CC (1996) Optimal campaign planning scheduling of multipurpose batch semicontinuous plants. 1. Mathematical formulation. Indus Eng Chem Res 35:488–509. https://doi.org/10.1021/ie9500811

Schulze Spüntrup F, Dalle Ave G, Imsland L, Harjunkoski I (2020) Optimal maintenance for degrading assets in the context of asset fleets—a case study. Front Appl Math Stat 6. https://doi.org/10.3389/fams.2020.528181

Sung C, Maravelias CT (2007) An attainable region approach for production planning of multiproduct processes. AIChE J 53:1298–1315. https://doi.org/10.1002/aic.11167. arXiv:1402.6991

Thistlethwaite F, Taylor GR (1953) The transportation revolution, 1815–1860. Econ His Rev 5:417. https://doi.org/10.2307/2591821

Vieira M, Barbosa-Povoa AP, Moniz S, Pinto-Varela T (2018) Simulation-optimization approach for the decision-support on the planning and scheduling of automated assembly lines. In: 2018 13th APCA international conference on control and soft computing (CONTROLO), IEEE, pp 265–269. https://doi.org/10.1109/CONTROLO.2018.8514297

Wan X, Pekny JF, Reklaitis GV (2005) Simulation-based optimization with surrogate models—application to supply chain management. Comput Chem Eng 29:1317–1328. https://doi.org/10.1016/j.compchemeng.2005.02.018

Wilkinson SJ, Shah N, Pantelides CC (1995) Aggregate modelling of multipurpose plant operation. Comput Chem Eng 19:583–588. https://doi.org/10.1016/0098-1354(95)87098-9

Wintle J, Moore P, Smalley S (2006) Plant ageing—management of equipment containing hazardous fluids or pressure. Technical report, Health and Safety Executive, Norwich

Yang Z, Djurdjanovic D, Ni J (2008) Maintenance scheduling in manufacturing systems based on predicted machine degradation. J Intell Manuf 19:87–98. https://doi.org/10.1007/s10845-007-0047-3

Zhang Q, Grossmann IE (2016) Planning and scheduling for industrial demand side management: advances and challenges. In: Alternative energy sources and technologies. Springer International Publishing, Cham, pp 383–414. https://doi.org/10.1007/978-3-319-28752-2_14

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.