Partially Observable Markov Decision Processes for Optimal Operations of Gas Transmission Networks

Michele Compare[1,2], Piero Baraldi[1], Paolo Marelli[1], Enrico Zio[1,2,3,4]

[1]Energy Department, Politecnico di Milano, Italy
[2]Aramis S.r.l., Italy
[3]MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France
[4]Eminent Scholar, Department of Nuclear Engineering, College of Engineering, Kyung Hee University, Republic of Korea

Abstract

We develop a decision-support framework based on Partially Observable Markov Decision Processes (POMDPs) for the management of Gas Transmission Networks (GTNs) operations, encoding realistic degradation state estimations provided by Prognostics and Health Management (PHM) systems, while considering demand variations and the effects of the management decisions on the GTN degradation evolution. This Operation and Maintenance (O&M) framework allows optimally operating a GTN. Furthermore, the economic impact of using PHM systems with different accuracy levels can be estimated. The approach is shown with reference to a GTN of the literature.

Keywords: Partially Observable Markov Decision Processes, Prognostics and Health Management, degradation state estimation errors, Gas Transmission Network.

1 Introduction

Gas Transmission Networks (GTNs) are Critical Infrastructures (CIs) serving multiple industrial sectors and domestic users. The management of GTNs entails making decisions about improvements of their structures (i.e., topology, technology of its elements, etc.) and their Operation and Maintenance (O&M). With respect to structural improvement, for example models to assess the vulnerability of GTNs to multiple disruptive events are developed in [1; 2], whereas a bottleneck

analysis to identify the GTN points where the gas flow reaches the maximum allowed capacity are carried out in [3]. These models support cost/benefit analyses driving the decisions about the most profitable interventions to be done on the GTN, such as adding a redundant input point or building new connections. An integrated non-linear optimization model is developed in [4] to select the location and installation schedule of major physical components in a GTN, including pipelines and compressor stations.

With respect to operations management, a major issue extensively studied in the literature is gas flow optimization. For example, dynamic programming [5; 6], gradient-based approaches [7] and genetic algorithms [8] have been proposed for optimizing gas flow in steady-state conditions.

GTN operations management concerns also the proper setting of the compressor stations to guarantee that the gas flow reaches the consumers. In this respect, optimization models are proposed in [9; 10] to find the configurations that minimize fuel consumption while guaranteeing the gas delivery to the network users, whereas settings of the compressor stations loading conditions are sought in [11] to minimize the total losses due to leakage.

As for maintenance, for example, a risk-based maintenance management approach is propounded in [12], based on the Portfolio Decision Analysis framework [13], to allocate the maintenance budget on the most risky GTN elements. Nowadays, the availability of Prognostics and Health Management (PHM) capabilities allow continuously tracking the degradation of the GTN elements and predicting their future evolution. For example, PHM systems are used for online monitoring the degradation of GTN compressor units in [14] and pipelines in [15]. Intuitively, PHM allows setting an efficient, just-in-time and just-right maintenance of GTNs: the right part is provided to the right place at the right time, handled by the right crew, with consequent benefits for profit.

However, the full potential of PHM can be larger than this: it can enable setting a dynamic management of the GTN O&M in which PHM is integrated in a decision framework that captures the intertwined relation of operations and maintenance decisions. Namely, maintenance decisions must be driven by the predictions of the degradation of the GTN elements, while giving, at the same time, due account to the changes that the equipment temporary unavailability introduce in the network topology and the expected future behavior of gas demand; these modify the capability of meeting the user needs.

The main issues to be considered for the development of the dynamic O&M management framework are:

1. When predicting the equipment degradation evolution, the typical assumption of considering its future conditions of usage constant (e.g., [16; 17; 18; 19]) or behaving according to some known exogenous stochastic process (e.g., [20; 21; 22]) cannot be applied to the O&M management framework in which, instead, they are modified by the operation decisions (e.g., setting the load of each compressor station to inject the proper amount of gas into the GTN).

2. The degradation states of the GNT elements are not precisely known, as they are estimated by PHM algorithms affected by errors. Neither are precisely known the behavior of the GTN demand, the effects of the different settings of the compressor stations on the degradation evolution and the effects of the maintenance decisions on the operability of the network.

These issues have been addressed separately. For example, a framework based on Markov Decision Processes (MDPs) is applied in [24] to a simple system, which shows the benefit with respect to the traditional scheduled maintenance of making decisions based on the predictions of both degradation state and future load of the equipment. In regards to the uncertainty in the PHM estimations, there are studies defining the metrics for its characterization (e.g., [25; 26]) and theoretical enhancements for modeling (e.g., [27]) and decision making (e.g., [28; 29]).

In this work, we address the two issues together, to develop a novel approach to decision-making in GTN equipped with PHM capabilities, in which the O&M management is framed as within the Sequential Decision Problem (SDP) paradigm in the presence of uncertainties, and we adopt the Partially Observable Markov Decision Processes (POMDPs) framework [23] for its solution. This new way of approaching the comprehensive management of GTN O&M represents the main contribution of this paper, which uses algorithms of the literature ([30; 31]) to solve POMDP.

POMDPs have already been used for managing CIs. For example, the POMDP framework is proposed in [32] to solve the issue of power supply restoration in an electrical network. More recently, POMDPs are used in [33] to estimate the Value of Information (VoI) in the management of a civil CI, where the inspections are inaccurate. In [34], the authors rely on POMDPs to investigate the flexibility of CIs like waste-water treatment plants and flood protection systems operating in an uncertain climate-changing environment. POMDPs are used in [35; 36] to optimize maintenance and inspection policies on corroded structures. However, these works are concerned with single components only and do not address the issue of managing the O&M of a GTN equipped with PHM capabilities.

The remainder of the paper is as follows: Section 2 provides a description of the problem setting and a detailed explanation of the POMDP framework; Section 3 presents the case study, which concerns a GTN of literature; Section 4 analyzes and discusses the results, whereas Section 5 concludes the work.

2 Problem statement

Consider a network delivering natural gas to its users. In all generality, this can be represented by a graph $G(N, E)$, where $N = \{1, ..., \nu\}$ is the set of nodes and $E = \{(i, j) \,|i, j \in N, i \neq j\}$ is the set of edges (i.e., pipeline stretches) connecting the nodes. Edges are assigned the identification

number $e = N + 1, ..., N + |E|$.
The network nodes can be distinguished into [37; 38; 39]:

1. Source nodes $N^s \subset N$, whose function is to inject the gas into the network, so that it can be delivered to the users.
2. User nodes $N^u \subset N$, which demand the gas and, therefore, represent sinks.
3. Support unit nodes $N^h \subset N$; these are the compressor stations guaranteeing the gas flow through the network. Since some support units consume a portion of the flowing gas, $N^h \cap N^u \neq \emptyset$ and $N^h \cap N^s \neq \emptyset$.
4. Dummy nodes $N^d \subset N$; these identify the connections of pipe stretches with different maximum capacities, which introduce discontinuities in the gas flow.

For readibility, Figure 1 shows the Euler diagram of the different types of nodes.

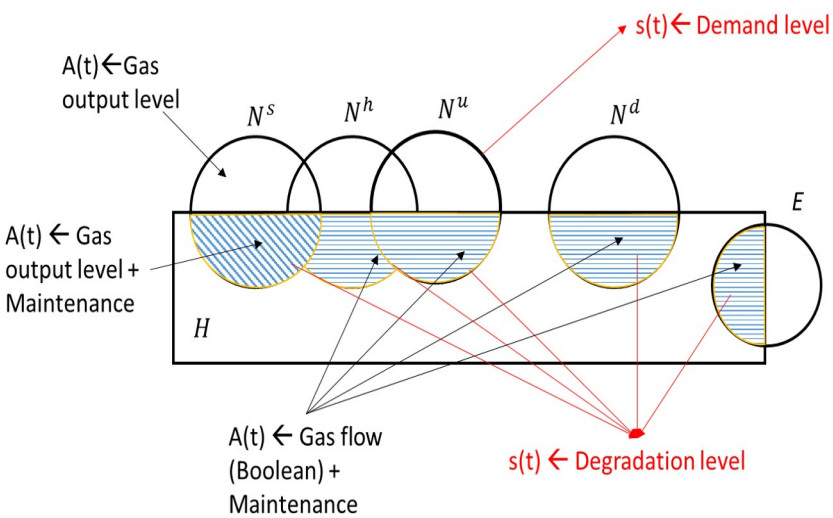


Figure 1: Euler diagram of the different types of nodes

The elements of the network are affected by degradation processes leading to disruptions. To reduce the computational burden while preserving the generality of the approach, we account for the effects of degradation only on a subset $H \subset N \cup E$ of network items that are critical with respect to the reliability metrics defined in [40]. This implies that elements in $(N \cup E) \backslash H$ are assumed to not degrade.
According to [41], the degradation processes are modeled as multi-state Markov processes [42; 43]: the degradation of every network item $i \in H$ evolves through states $\zeta^i \in \mathrm{Z}^i = \{1, ..., |Z^i|\}$ of increasing degradation level, where $\zeta^i = 1$ refers to the As Good As New (AGAN) state of the

item, whereas $\zeta^i = |Z^i|$ denotes its failure; $|\cdot|$ specifies the cardinality of its set argument.
Given the criticality of the H items for the proper network operation, they are assumed to be equipped with PHM systems for estimating their degradation states. The PHM systems, however, are not perfect: the algorithms estimating the degradation states of the items $i \in H$ based on the monitoring signals could estimate the degradation states different from the actual ones. We assume that the PHM systems performance is qualified, i.e., the probability of the misclassification errors is known and represented by a confusion matrix [38; 44; 45].
Concerning the user nodes, the gas demand of user $i \in N^u$ is indicated by $d^i \in D^i = \left\{d^i_1, ..., d^i_{|D^i|}\right\}$ and is assumed to change over time, according to a dynamics dependent on factors such as gas selling price, import/export rate, population growth, environmental concerns, climate changes [46; 47; 48]. In our framework, we model the uncertain evolution of gas demand over the months as a discrete-time Markov process, which implies assuming that the future evolution of the demand depends only on the present time value, and not on the past evolution. Although in many real-life situations this assumption is not verified, nonetheless modeling the gas demand as a Markov chain could not necessarily imply a loss of generality, as non-Markov processes can be turned into Markov processes through augmentation of the state space, i.e., by introducing additional variables [34; 49; 50]. Obviously, this results in a more computational demanding model.
The loads of the source nodes N^s need to be properly set to allow matching the users' requests, while considering possible leakages throughout the GTN. In details, the OM can select the amount of gas injected into the network $F^i \in \left\{F^i_1, ..., F^i_{L^i}\right\}$ from source node $i \in N^s$. L^i denotes the supply level at node i, such that:

$$F^i_{l-1} > F^i_l, \qquad l = 2, \ldots, L^i \tag{1}$$

Obviously, the amounts of gas inserted into the network cannot be arbitrarily set, as the continuity equation must always be verified:

$$\sum_{i \in N^s} F^i = \sum_{i \in N^u} d^i \leq F_{\max} \tag{2}$$

where F_{max} is the maximum quantity of gas that can be injected into the network by the source nodes N^s. We use the Boykov-Kolmogorov algorithm to find the maximum flow through the network ([51]).
Besides setting the loads, the network manager has to take decisions on the maintenance actions to be performed on degrading network items H, which are distinguished into preventive and corrective. The former are performed to prevent failures, whereas the latter are carried out upon failure to restore the system. For generality, we assume that in both cases the maintenance task is not always capable of restoring the item to its As Good As New (AGAN) state (i.e., imperfect maintenance

[52]). However, preventive maintenance tasks outperform the corrective tasks with respect to the repair time: preventive actions offer the possibility of properly arranging in advance the logistics to carry out the operation, thus eliminating all the related issues such as the propagation of the effects of the failure or the search for the failed component(s). This results in a faster intervention. The objective of this work is to develop a novel approach for the effective management of GTNs equipped with realistic PHM capabilities, which suffer from misclassification errors in the monitored equipment health states. To be effective, GTN management have to take into acccount that operation decisions, which mainly depend on the GTN state and gas demand variations, must be intertwined with maintenance decisions, considering their effects on the GTN degradation evolution and operations. To do this, O&M management is framed within the SDP paradigm in the presence of uncertainties, and the POMDPs framework [23; 30; 31] is used to find the optimal policy.

2.1 POMDP framework for GTN operation management

A POMDP is formally defined as the 7-tuple $(s, \mathbf{A}, \mathbf{T}, R, \Omega, \mathbf{b}(0))$, whose elements are described below.

Generally, the state of the managed system at time t must encode all the relevant information to make decisions about the actions to be taken. In our framework, the state space encodes both the degradation levels of items H and the demand levels of users N^u. These variables are arranged in vector $\mathbf{s}(t)$, whose k-th element is:

$$s^k = \begin{cases} \zeta^i & i = \min\{\iota \in H : |\{\mu \in H : \mu \leq \iota\}| \geq k\}\,,\ k = 1, ..., |H| \\ d^i & i = \min\{\iota \in N^u : |\{\mu \in N^u : \mu \leq \iota\}| \geq k - |H|\}\,,\ k = |H|+1, ..., |H|+|N^u| \end{cases} \tag{3}$$

Namely, the first $|H|$ entries of $\mathbf{s}(t)$ represent the degradation levels of the PHM-equipped items H, arranged in ascending identification numbers. Then, the $|N^u|$ demand levels are appended to the first $|H|$ entries, still organized in ascending node identification number.

Then, vector $\mathbf{s}$ belongs to set $S = \times_{i \in H} Z^i \times_{i \in N^u} D^i$. For convenience, we enumerate the vectors in S in arbitrary order. Thus, the generic state vector is indicated by $\mathbf{s}$, whereas the enumerated vectors by $\mathbf{s}_\sigma$, $\sigma = 1,\ ...,\ |S|$.

To satisfy the user demands of gas, the OM selects on each item $i \in N^s \cup H$ an action a^i from the corresponding set of actions Λ^i. In details, concerning nodes $i \in N^s \cap H$, Λ^i=$\{1, \ldots, L^i, L^i + 1, L^i + 2\}$, where the first L^i elements of the set specify the output level F^i of the node and can be selected only if node i is not failed (i.e., $a^i \leq L^i$ if $\zeta^i < |Z^i|$), whereas L^i+1 and L^i+2 denote the execution of preventive and corrective maintenance actions, respectively. Obviously, a^i can be set to $L^i + 1$ if $\zeta^i < |Z^i|$, whereas $a^i = L^i + 2$ whenever $\zeta^i = |Z^i|$.

For every network item (i.e., node or edge) $H \backslash N^s$, we set $L^i = 1$, whereby the set of actions

Λ^i consists of 3 elements: the first indicates that the gas flows through the item, whereas the remaining two refer to performing preventive and corrective maintenance tasks, respectively. In general, the output level associated to both maintenance activities is null:

$$F^i_{L^i+1} = F^i_{L^i+2} = 0, \quad \forall i \in H \tag{4}$$

Edges $E \backslash H$ are fully reliable pipelines, with no related action.
Finally, also nodes $N^s \backslash H$ are assumed to be fully reliable and, thus, they do not need maintenance. The set of actions for every node $i \in N^s \backslash H$ is $\Lambda^i = \{1, \ldots, L^i\}$, with each element indicating the node's gas output level F^i.
The concatenation of actions a^i yields vector $\mathbf{A} \in \mathcal{A} = \times_{i \in N^s \cup H} \Lambda^i$, whose k-th element is:

$$A^k = a^i, \quad i = \min \left\{ \iota \in N^s \cup H : |\{\mu \in N^s \cup H : \mu \leq \iota\}| \geq k \right\}, \; k = 1, ..., |N^s \cup H| \tag{5}$$

When the network OM applies action $\mathbf{A}$ to the system in state $\mathbf{s}(t)$, this undergoes a Markov transition to state $\mathbf{s}(t+1)$, according to the transition probability function $T : S \times \mathcal{A} \times S \rightarrow [0, 1]$:

$$T(\mathbf{s}(t), \mathbf{A}, \mathbf{s}(t+1)) = \Pr\left[\mathbf{s}(t+1) \,|\mathbf{s}(t), \mathbf{A}\right] \tag{6}$$

The OM incurs into a reward $R(\mathbf{s}(t), \mathbf{A}, \mathbf{s}(t+1))$ when the system ends up in state $\mathbf{s}(t+1)$ after the execution of action $\mathbf{A}$ in state $\mathbf{s}(t)$. The reward function $R : S \times \mathcal{A} \times S \rightarrow \mathbb{R}$ indicates how good the immediate effect of action $\mathbf{A}$ is (i.e., the transition between $\mathbf{s}(t)$ and $\mathbf{s}(t+1)$ for reaching the final objective of delivering gas to the users). Thus, the reward function strictly depends on the specific case study considered and it has to relate to the global behavior of the GTN.
As mentioned before, we assume that the degradation state of the condition-monitored items, H, is not exactly known by the OM, who has to take actions based on the possibly incorrect estimations provided by the PHM systems. The information on the k-th element of the state vector, s^k, is summarized by observation

$$o^k \in \Phi^i = \{1, ..., \left|\Phi^i\right|\} \quad i = \min \left\{ \iota \in H : |\{\mu \in H : \mu \leq \iota\}| \geq k \right\}, \; k = 1, ..., |H| \tag{7}$$

With respect to the current demand of gas by users N^u, we assume that this is exactly known, as the flowmeters installed in GTNs are generally very accurate [53]. Then, $o^k = s^k \in \left\{ d^i_1, ..., d^i_{|D^i|} \right\}$, $k = |H| + 1, ..., |H| + |N^u|$. This way, the observation vector $\mathbf{o}(t) = \left[o^1, \; ..., \; o^{|H|+|N^u|}\right]$ belongs to the cartesian product $\Omega = \times_{i \in H} \Phi^i \times_{i \in N^u} D^i$. Similarly to S, the vectors in Ω are enumerated in arbitrary order: $\mathbf{o}_\omega$, $\omega = 1, ..., |\Omega|$.
Based on the observation set Ω, we introduce the emission function $O : \Omega \times \mathcal{A} \times S \rightarrow [0, 1]$, which

defines the probability of having observation $\boldsymbol{o}(t+1) \in \Omega$ when the application of action $\mathbf{A} \in \mathcal{A}$ has led the system entering state $\mathbf{s}(t+1) \in S$:

$$O(\mathbf{o}(t+1), \mathbf{A}, \mathbf{s}(t+1)) = \Pr\left[\mathbf{o}(t+1)|\mathbf{s}(t+1), \mathbf{A}\right] \tag{8}$$

We use the emission function to model the probability that the PHM system provides an incorrect degradation state estimation, being in this particular case the observation at time $t+1$ independent from the action taken at time t.

In the POMDP framework, the state of the system, $\mathbf{s}(t)$, cannot be unambiguously determined at any time instant t. Therefore, the OM has to take actions based on his/her current *belief* of the system state, which depends on the sequence of actions and subsequent observations retrieved up to time t. Formally, the belief at time t is the probability distribution

$$\mathbf{b}(t) = \Pr\left[\mathbf{s}(t)|\mathbf{o}(t), \mathbf{A}(t-1), \mathbf{o}(t-1), ..., \mathbf{A}(1), \mathbf{o}(1), \mathbf{A}(0)\right] \tag{9}$$

whose σ-th element, b_σ, reads:

$$b_\sigma = \Pr\left[\mathbf{s}(t) = \mathbf{s}_\sigma|\mathbf{o}(t), \mathbf{A}(t-1), \mathbf{o}(t-1), ..., \mathbf{A}(1), \mathbf{o}(1), \mathbf{A}(0)\right], \quad \sigma = 1, ..., |S| \tag{10}$$

To conclude the definition of the POMDP 7-tuple, we define the initial belief, $\mathbf{b}(0)$, specifying the knowledge of the OM about the initial state of the system.

Whenever the OM implements an action, the current belief $\mathbf{b}(t)$ must be updated to take into account the new information retrieved from the system, i.e., the observation $\mathbf{o}(t+1)$. This process, which is known as *belief update*, is based on the following application of Bayes' rule [30; 31]:

$$b_\sigma(t+1) = \frac{\mathbf{O}\left(\mathbf{o}(t+1), \mathbf{A}, \mathbf{s}_\sigma(t+1)\right)}{\sum_{y=1}^{|S|} \mathbf{O}\left(\mathbf{o}(t+1), \mathbf{A}, \mathbf{s}_y(t+1)\right) \sum_{x=1}^{|S|} T\left(\mathbf{s}_x(t), \mathbf{A}, \mathbf{s}_y(t+1)\right) b_x(t)} \cdot \sum_{m=1}^{|S|} T\left(\mathbf{s}_m(t), \mathbf{A}, \mathbf{s}_\sigma(t+1)\right) b_m(t), \quad \sigma = 1, ..., |S| \tag{11}$$

This formula allows the OM updating the probability distributions over the states of the system, which must be associated to the optimal actions to be taken. Obviously, these do not depend on the immediately generated reward (i.e., greedy actions); rather, they must consider the expected future evolution of the system states, beliefs and the actions that will be taken correspondingly.

If we define a policy $\boldsymbol{\pi}$ as a mapping of beliefs onto actions, i.e., $\mathbf{A} = \boldsymbol{\pi}(\mathbf{b})$, then its performance is indicated by the value function $V^{\boldsymbol{\pi}}$, defining the expected return gained when the initial belief is $\mathbf{b}(0)$ and the adopted policy is $\boldsymbol{\pi}$:

$$V^{\boldsymbol{\pi}}(\mathbf{b}(0)) = E_{\boldsymbol{\pi}}\left[\sum_{t \geq 0} \gamma^t r(\mathbf{b}(t), \boldsymbol{\pi}(\mathbf{b}(t)))|\mathbf{b}(0)\right] \tag{12}$$

where $E_{\boldsymbol{\pi}}[\ \cdot\ ]$ denotes the expectation operator given that policy π is being followed, γ is a discount factor ($0 \leq \gamma < 1$) [54] and the expected future rewards read:

$$r(\mathbf{b}(t), \boldsymbol{\pi}(\mathbf{b}(t))) = \sum_{x=1}^{|S|} \left(\sum_{y=1}^{|S|} R(\mathbf{s}_x(t), \boldsymbol{\pi}(\mathbf{b}(t)), \mathbf{s}_y(t+1)) \cdot T(\mathbf{s}_x(t), \boldsymbol{\pi}(\mathbf{b}(t)), \mathbf{s}_y(t+1)) \right) \cdot b_x(t) \quad (13)$$

The objective of this work is the identification of the optimal policy $\boldsymbol{\pi}^*$, which maximizes the value function $V^{\boldsymbol{\pi}}$, starting from $\mathbf{b}(0)$. The corresponding value of $V^{\boldsymbol{\pi}^*}(\mathbf{b}(0))$ is indicated by V^*.

2.2 POMDP solution

Generally speaking, in the POMDP framework the optimal policy π^* is obtained through an iterative procedure, which ends when some convergence criterion is met. To do this, we build on the formulation developed in [30; 31]. Namely, we exploit the property that the optimal value function V^* of a policy with an infinite planning horizon can be approximated by a piecewise linear and convex function [55], which is iteratively estimated. This allows partitioning the belief space into several regions, which are associated to different mapping functions of $\mathbf{b}$ into V^* [55].

In details, according to [30; 31], we can parametrize the value function at the n-h iteration, V_n, by a set of vectors $\Xi_n = \left\{\boldsymbol{\alpha}^{n,1}, ..., \boldsymbol{\alpha}^{n,|\Xi_n|}\right\}$, such that each $\boldsymbol{\alpha}$-vector maximizes V_n in one of the regions into which the belief space has been partitioned.

This representation allows us calculating the value of belief $\mathbf{b}$ at the n-th iteration as:

$$V_n(\mathbf{b}) = \max_{\boldsymbol{\alpha}^{n,\xi} \in \Xi_n} (\mathbf{b} \cdot \boldsymbol{\alpha}^{n,\xi}) \quad (14)$$

where

$$\mathbf{b} \cdot \boldsymbol{\alpha}^{n,\xi} = \sum_{\sigma=1}^{|S|} b_\sigma \alpha_\sigma^{n,\xi} \quad (15)$$

is the standard inner product in the vector space, with $\alpha_\sigma^{n,\xi}$ being the σ-th entry of vector $\boldsymbol{\alpha}^{n,\xi}$.

Methods extending the value iteration algorithms usually adopted to solve Markov Decision Processes (MDPs) have been proposed to calculate the optimal value function V^* and, thus, to find the optimal policy (e.g., [56; 57]). These, however, are computationally infeasible for large domain POMDPs [31]. To circumvent this issue, Point-Based Methods (PBMs) have been proposed [30; 66; 58; 59], which find approximate solutions of V^* in smaller computational times, at the cost of losing some optimality.

PBMs manipulate the value function updating procedure to speed up the computation of set Ξ. In fact, Eq. 14 can be re-written as follows [30; 31]:

$$V_{n+1}(\mathbf{b}) = \max_{\mathbf{A} \in \mathcal{A}} \left[\mathbf{b} \cdot P(\mathbf{s}, \mathbf{A}) + \gamma \sum_{\omega=1}^{|\Omega|} \max_{\boldsymbol{\alpha}^{n,\xi} \in \Xi_n} (\mathbf{b} \cdot \mathbf{g}^{\xi}_{\mathbf{A},\mathbf{o}_\omega}) \right] \quad (16)$$

where $\mathbf{g}^{\xi}_{\mathbf{A},\mathbf{o}_\omega}$ is a vector whose σ-th entry reads:

$$\mathbf{g}^{\xi}_{\mathbf{A},\mathbf{o}_\omega}(\sigma) = \sum_{m=1}^{|S|} \mathbf{O}\left(\mathbf{o}_\omega(t+1)|\mathbf{s}_m(t+1),\mathbf{A}\right) \cdot T\left(\mathbf{s}_m(t+1)|\mathbf{s}_\sigma(t),\mathbf{A}\right) \cdot \alpha^{n,\xi}_m, \quad \sigma = 1,...,|S| \tag{17}$$

and $\mathbf{P}(\mathbf{s},\mathbf{A})$ is a vector whose σ-th element, $\sigma = 1,...,|S|$, is:

$$\mathbf{P}(\mathbf{s}_\sigma,\mathbf{A}) = \sum_{m=1}^{|S|} R(\mathbf{s}_\sigma(t),\mathbf{A},\mathbf{s}_m(t+1)) \cdot T(\mathbf{s}_\sigma(t),\mathbf{A},\mathbf{s}_m(t+1)) \tag{18}$$

We can now apply twice the identity $\max\limits_{\boldsymbol{\alpha}^{n,\xi}\in\Xi_n}(\mathbf{b}\cdot\boldsymbol{\alpha}^{n,\xi}) = \mathbf{b}\cdot\arg\max\limits_{\boldsymbol{\alpha}^{n,\xi}\in\Xi_n}(\mathbf{b}\cdot\boldsymbol{\alpha}^{n,\xi})$ to Eq. (16) [30]; the first application yields:

$$\max_{\xi=1,...,|\Xi_n|}(\mathbf{b}\cdot\mathbf{g}^{\xi}_{\mathbf{A},\mathbf{o}_\omega}) = \mathbf{b}\cdot \arg\max_{\mathbf{g}^{\xi}_{\mathbf{A},\mathbf{o}_\omega},\xi=1,...,|\Xi_n|}(\mathbf{b}\cdot\mathbf{g}^{\xi}_{\mathbf{A},\mathbf{o}_\omega}) \tag{19}$$

so that from Eq 16:

$$V_{n+1}(\mathbf{b}) = \max_{\mathbf{A}\in\mathcal{A}}\left[\mathbf{b}\cdot\mathbf{P}(\mathbf{s},\mathbf{A}) + \gamma\sum_{\omega=1}^{|\Omega|}\mathbf{b}\cdot \arg\max_{\mathbf{g}^{\xi}_{\mathbf{A},\mathbf{o}_\omega},\xi=1,..,|\Xi_n|}(\mathbf{b}\cdot\mathbf{g}^{\xi}_{\mathbf{A},\mathbf{o}_\omega})\right] \tag{20}$$

From the second application, we obtain:

$$V_{n+1}(\mathbf{b}) = \mathbf{b}\cdot\arg\max_{\mathbf{A}\in\mathcal{A}}\left[\mathbf{b}\cdot\mathbf{P}(\mathbf{s},\mathbf{A}) + \gamma\sum_{\omega=1}^{|\Omega|}\mathbf{b}\cdot \arg\max_{\mathbf{g}^{\xi}_{\mathbf{A},\mathbf{o}_\omega},\xi=1,..,|\Xi_n|}(\mathbf{b}\cdot\mathbf{g}^{\xi}_{\mathbf{A},\mathbf{o}_\omega})\right] = \mathbf{b}\cdot\arg\max_{\mathbf{A}\in\mathcal{A}}\left[\mathbf{b}\cdot\mathbf{g}^{\boldsymbol{\alpha}}_{\mathbf{A}}\right] \tag{21}$$

where:

$$\mathbf{g}^{\boldsymbol{\alpha}}_{\mathbf{A}} = \mathbf{P}(\mathbf{s},\mathbf{A}) + \gamma\sum_{\omega=1}^{|\Omega|}\arg\max_{\mathbf{g}^{\xi}_{\mathbf{A},\mathbf{o}_\omega},\xi=1,..,|\Xi_n|}(\mathbf{b}\cdot\mathbf{g}^{\xi}_{\mathbf{A},\mathbf{o}_\omega}) \tag{22}$$

This way, the value function for belief $\boldsymbol{b}$ becomes:

$$V_{n+1}(\mathbf{b}) = \mathbf{b}\cdot \arg\max_{\mathbf{A}\in\mathcal{A},\boldsymbol{\alpha}\in\Xi_n}(\mathbf{b}\cdot\mathbf{g}^{\boldsymbol{\alpha}}_{\mathbf{A}}) \tag{23}$$

By comparing Eq. (23) with Eq. (14), we notice that:

$$\boldsymbol{\alpha}^{n+1} = \arg\max_{\mathbf{A}\in\mathcal{A},\boldsymbol{\alpha}\in\Xi_n}(\mathbf{b}\cdot\mathbf{g}^{\boldsymbol{\alpha}}_{\mathbf{A}}) \tag{24}$$

This last operation, called *backup*($\boldsymbol{b}$), allows computing the optimal $\boldsymbol{\alpha}$-vector and, thus, the optimal action, for belief $\boldsymbol{b}$. The full complexity of a point-based backup requires $O(|\mathcal{A}| \times |\Omega| \times |\Xi| \times |S|^2 + |\mathcal{A}| \times |\Omega| \times |S|)$ operations [31].

PBMs differ from techniques that aim at finding an exact solution for every belief point because they act over a finite set B of belief points. This way, the *backup* operation is performed for a fixed number of times in every iteration, leading to a significant reduction in the number of operations performed with respect to the exact algorithms [31].

The main idea of PBMs is to focus on a set of belief points that can be reached starting from $\mathbf{b}(0)$ and following an arbitrary policy. Beliefs belonging to this subset, which is much smaller than the whole belief space, can be collected by applying the belief update procedure to the starting belief $\mathbf{b}(0)$ several times. This implies that the definition of the belief set B is fundamental for the PBMs to be effective, both in terms of accuracy and convergence speed. Moreover, the optimal policy found by PBMs is conditional on the initial knowledge of $\mathbf{b}(0)$. If we change the intial belief into $\mathbf{b}'(0)$, then also the subset of belief points reached during the simulation changes, and we get the approximate value of this new belief point $V(\mathbf{b}'(0))$.

In Appendix 1, we provide a brief description and the pseudocode of the PERSEUS algorithm [30], which we have used to compute the solution to our problem. We refer to [31; 59; 60] for further details on the procedures used to determine the belief set B.

3 Case Study

We apply the proposed methodological framework to the GTN presented in [61], which can be modeled as a graph with 30 nodes and 34 edges (see Figure 2). With respect to the original network, some modifications are introduced to avoid unnecessary computational burden (Figure 3). In particular, the user nodes are partitioned into 5 groups on the basis of their location in the network and nodes in the same group are assumed to behave as a single merged node. This way, the classification of the nodes in their respective sets is as follows:

1. N^s={2,18,22};
2. N^h={2,8,10,12,14,17,19};
3. N^u={16,17,18,20,21};
4. N^d={3,4,5,6,7,9,11,13,15}.

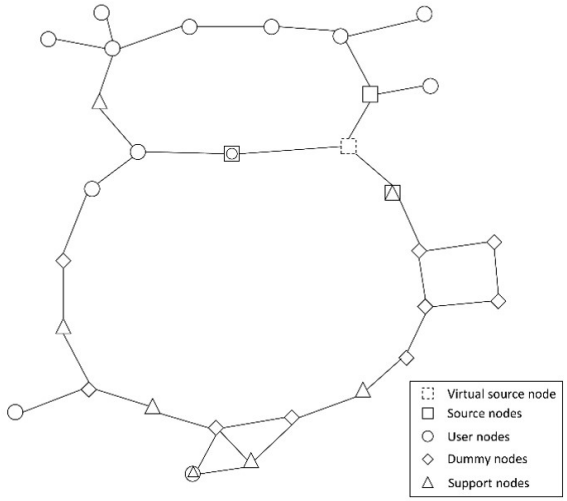
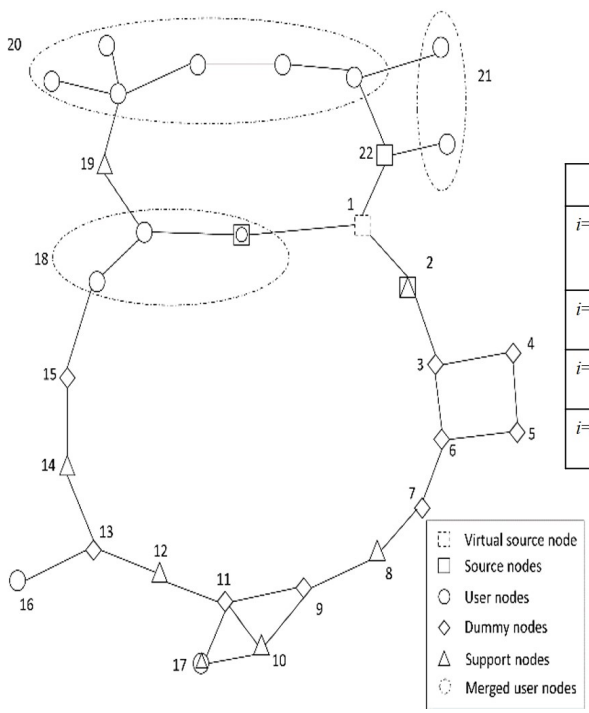


Figure 2: Original GTN [61]



s(t)=

ζ^2	ζ^{17}	d^{16}	d^{17}	d^{18}	d^{20}	d^{21}

A(t)

		$a^i = 1$	$a^i = 2$	$a^i = 3$	$a^i = 4$	$a^i = 5$
i=2	$N^s \cap H$	$F_1^2 = F_{nom}^2$	$F_2^2 = 0.8 \cdot F_{nom}^2$	$F_3^2 = 0.5 \cdot F_{nom}^2$	Preventive $F_4^2 = 0$	Corrective $F_5^2 = 0$
i=17	$N^u \cap H$	$F_1^{17} = F_{nom}^{17}$	Preventive $F_2^{17} = 0$	Corrective $F_3^{17} = 0$	-	-
i=18	N^s	$F_1^{18} = F_{nom}^{18}$	$F_2^{18} = 0$	-	-	-
i=22	N^s	$F_1^{22} = F_{nom}^{22}$	$F_2^{22} = 0$	-	-	-

Figure 3: Modified GTN

The nominal gas demand d_{nom}^i of each user node $i \in N^u$ is obtained by summing the nominal gas demands of the nodes of the original network included in group i (Table 1). We also assume that the source capacities reported in [61] represent the largest possible gas injection F_{nom}^i into the

network by sources $i \in N^s$. The time horizon we are referring to is the CI lifetime and the reference time unit is the month, although a different time scale (e.g., year) might be used. This requires converting the daily gas demands (i.e., Mm^3d^{-1}) considered in [61] into monthly gas demands (i.e., Mm^3mo^{-1}) by multiplying their values by 30. The failure of a node is assumed to reduce to 0 the capacities of the connected pipe stretches.

Table 1: Nominal gas demands of user nodes

User Node	Nominal gas demand, d^i_{nom} $[Mm^3mo^{-1}]$	Importance wt^i
16	82.2	0.046
17	1094.7	0.616
18	105	0.059
20	135	0.076
21	360	0.203

3.1 State

We set $H = \{2,17\}$. Source node i=2 represents a Liquefied Natural Gas (LNG) terminal ([61]). Its pumping system is critical because it supplies the largest quantity of gas. User node i=17 represents a compressor station ([61]), which is critical because its demand covers more than 50% of the GTN total demand [61]. These two nodes are assumed to be equipped with PHM systems and their degradation levels have been discretized into 3 states, $Z^2 = Z^{17} = \{1, 2, 3\}$, with 1 indicating the As Good As New condition and 3 the failure of the node. Notice that the degradation mechanisms affecting the GTNs typically take some years to reach the failure states. This is coherent with the assumption of considering the month as time unit.

The demand levels of nodes i=16, 18, 20 are set to $D^i = \{0, d^i_{nom}\}$ due to their small gas request (i.e., less than 15% of the total demand of the network): the null demand level indicates that gas is not requested, whereas d^i_{nom} denotes that the i-th node requires its nominal supply of gas. On the other hand, the demand levels of user nodes i=17, 21 are set to $D^i = \{0.95d^i_{nom}, d^i_{nom},\ 1.05d^i_{nom}\}$, which refer to 95% of the total nominal gas demand, 100% of the total nominal gas demand and 105% of this value, respectively.

This way, the state $\mathbf{s}$ of the network is represented by a vector with $|S| = 7$ entries, the first two denoting the degradation levels of nodes H and the remaining five indicating the gas demands of the users (see Figure 3).

3.2 Actions

To meet the gas request of the users, the OM must select for each node $i \in H \cup N^s = \{2, 17, 18, 22\}$ an action from the corresponding set Λ^i (see Figure 3). Specifically, we assume that for source node $i = 2$, the OM can choose the output level among 3 different values of gas supply. Thus, $\Lambda^2 = \{1, 2, 3, 4, 5\}$: the first action, $a^2 = 1$, sets the output level of the source exactly to its nominal value (i.e., $F_1^2 = F_{nom}^2$); the second action, a^2, to 80% of its nominal value (i.e., $F_2^2 = 0.8F_{nom}^2$); the third, $a^2 =$, to 50% of its nominal value (i.e., $F_3^2 = 0.5F_{nom}^2$). The remaining two actions, $a^2 = 4$ and $a^2 = 5$, indicate the implementation of preventive and corrective maintenance tasks, respectively, with an associated null output level (i.e., $F_4^2 = 0$, $F_5^2 = 0$).
With respect to node $i = 17 \in H \backslash N^s$, the network OM can only decide whether to allow the arriving gas flowing through the node, a^{17}=1, or perform a preventive maintenance, a^{17}=2. In case the node is failed, OM is forced to make corrective maintenance, a^{17}=3. Notice that this node cannot receive any gas when it is undergoing maintenance. Finally, the two minor sources (nodes i=18 and i=22) can be operated at two distinct levels (i.e., $\Lambda^{18} = \Lambda^{22} = \{1, 2\}$), which indicate the source switch on and switch off, respectively. The $MATLAB^{®}$ function maxflow is used to evaluate the gas supply to the user nodes N^u, for every state $\mathbf{s}(t+1)$ of the network reached upon the execution of action $\mathbf{A}(t)$.

3.3 Transition Matrices

We assume that both the degradation processes affecting the nodes H and the demand levels of users N^u do not influence each other, although the degradation processes affecting nodes $i = 2$ and $i = 17$ evolve depending on the actions implemented on the network elements. Thus, the transition probability function reads:

$$T(\mathbf{s}(t), \mathbf{A}(t), \mathbf{s}(t+1)) = \prod_{k=1,\ldots,|H|+|N^u|} \Pr\left[s^k(t+1) \left| s^k(t), \mathbf{A}(t)\right.\right] \tag{25}$$

In details, the evolution of the degradation mechanism on source node $i = 2$ is faster when it supplies larger amounts of gas. Therefore, the smallest probability of failure of the node pertains to action $a^2 = 3$, whereas the largest probability to action $a^2 = 1$. On the other hand, maintenance activities are taken to restore the node to healthier conditions. Then, we formalize the transition probability function for the actions available at node $i = 2$ through $|Z^2| \times |Z^2| = 3 \times 3$ matrices $\mathbf{T}_{a^2}^{i=2}$, whose (v,w) entry denotes the probability that the state of node $i = 2$ switches from v to w according to action $a^2 \in \Lambda^2$. The 3×3 matrices corresponding to the actions in Λ^2 can be found in Appendix 2. Notice that we have arbitrarily set these transition probability values, as they are

not defined in [61].
Furthermore, we assume that the degradation of node i=17 depends only on the action taken on the node itself and on those performed on the sources. Evidently, this is a simplifying assumption, as the resulting degradation model does not encode all the functional dependencies (e.g., gas flows, demand level, current degradation state and so forth). However, the development of a refined degradation model of the network is beyond the purpose of this paper. We characterize the evolution of the degradation mechanism on node i=17 with 6 matrices, corresponding to the different possible network configurations generated by the combinations of actions taken. Namely, the first transition matrix $\mathbf{T}_1^{i=17}$ refers to the situation when source $i = 2$ operates at 80% or 100% of its nominal output level ($a^2 = 1$ or $a^2 = 2$) and, thus, it is able to satisfy the demand of user i=17, if $a^{17} = 1$. A different transition matrix $\mathbf{T}_2^{i=17}$ is assigned to node $i = 17$ when node $i = 2$ supplies 50% of its nominal output level and node $i = 17$ is demanding gas ($a^2 = 3$ and $a^{17} = 1$): in this configuration, the amount of gas provided by node i=2 is not sufficient to fully satisfy the request of the user node and, thus, a supply from the minor source nodes i=18 and i=22 is required. If node i=2 is not injecting gas, node i=17 relies on the supply of gas provided by the two minor sources. Thus, we provide a different transition matrix $\mathbf{T}_3^{i=17}$ describing the degradation evolution for these action settings ($a^{18} = 1$, $a^{22} = 1$ and $a^{17} = 11$; $a^{18} = 1$, $a^{22} = 2$ and $a^{17} = 1$; $a^{18} = 2$, $a^{22} = 1$ and $a^{17} = 1$; $a^{18} = 2$, $a^{22} = 2$ and $a^{17} = 1$). Finally, when no gas flows in the network, the node does not degrade and the transition probability function is described by an identity matrix $\mathbf{T}_4^{i=17} = I$.
Maintenance tasks on node i=17 are modeled by two transition matrices, $\mathbf{T}_5^{i=17}$ and $\mathbf{T}_6^{i=17}$, related to preventive and corrective actions, respectively. These are reported in Appendix 2.
The transition matrices relative to the gas demand variability of the users can be found in Appendix 2.
Notice that the entries of the transition matrices can be estimated from the available degradation data through known statistical techniques (e.g., [62; 63]). In general, this may require the availability of large datasets, which can be difficult to find in practice. In these cases, expert judgement and physics-based modeling approaches can be used to relate the actions taken to the degradation behaviors over time of the network items.

3.4 Rewards

To specify the reward function R, we consider the importance of user node $i \in N^u$ towards the global income, which is quantified by its weight:

$$wt^i = \frac{d^i}{\sum_{i \in N^u} d^i} \tag{26}$$

The larger the weight of the user node, the larger its importance (Table 1). Then, the income/loss obtained by supplying a quantity of gas su^i to node $i \in N^u$ is:

$$\begin{cases} r^i = c \cdot su^i & \text{if } su^i = d^i \\ r^i = -c \cdot (d^i - su^i) - p \cdot wt^i & \text{if } su^i < d^i \end{cases} \tag{27}$$

where we assume a price of gas $c = 0.2M€/Mm^3$ and a penalty $p = 9M€$is incurred by the gas operator for not delivering gas to the users, which corresponds to collecting a penalty of $0.3M€$ for 30 days.

Notice that the reward function can encode the costs related to the fuel consumption and leakages throughout the network and a more sophisticated algorithm can be embedded into the model to optimize the network model for minimizing the cost. For simplicity, these have not been explicitly considered.

Consequently, the reward gathered at the network level is defined as:

$$r = \sum_{i \in N^u} r^i + C_{PM} \cdot (\#PM\ actions) + C_{CM} \cdot (\#CM\ actions) \tag{28}$$

where $C_{PM} = -2.5M€$ and $C_{CM} - 5M€$ are the cost of preventive and corrective maintenance actions, respectively. The values for c, p, C_{PM} and C_{CM} have been arbitrarily set by the authors. The discount factor used to weigh the rewards obtained in the future is $\gamma = 0.9$.

Notice that the meaning of γ relates to decision maker objectives rather than to economic considerations (e.g., [54]): the smaller the discount factor is, the more greedy is the solution sought. The resulting state-action values of the optimal policy depend on this perspective.

3.5 Observations

We assume that the PHM systems on nodes H provide estimations of the degradation states independently on each other, which is typical of the PHM algorithms not working on fleet data. Yet, we assume that for every action A^k we know the probability that the OM observes $o^k = w$ when the k-th element of the state vector enters state $s^k = v$, $k \leq |H|$, which is not influenced by the observations on the other nodes. This is equivalent to assuming to have a qualified PHM system, with known and validated classification performances. For every A^k, these probabilities can be arranged in a $|Z^i| \times |\Phi^i|$ matrix $\mathbf{O}^k$, whose (v, w) entry reads:

$$\mathbf{O}^k(v, w; A^k) = \Pr\left[o^k = w | s^k = v, A^k\right] \tag{29}$$

For $k > |H|$, we can define $\mathbf{O}^k$ as $|D^i| \times |D^i|$ identity matrices, meaning that there is no uncertainty on the measured demand levels of the users N^u, whichever vector $\mathbf{A}$ is.

Based on these assumptions, the emission function for the network reads:

$$O(\mathbf{o}(t+1), \mathbf{A}(t), \mathbf{s}(t+1)) = \prod_{k=1,\ldots,|H|} \Pr\left[o^k(t+1) \,\middle|\, s^k(t+1), A^k(t)\right] \tag{30}$$

The probabilities that the PHM system misclassifies the degradation states of nodes H, nodes i=2 and i=17, are summarized by the observations matrices $\mathbf{O}^{i=2}$ and $\mathbf{O}^{i=17}$. In our work, $\Phi^i \equiv Z^i$, as we assume that PHM systems classify the degradation states of the nodes in one out of the three possible alternatives. This entails that the size of the observation matrices $\mathbf{O}^{i=2}$ and $\mathbf{O}^{i=17}$ is $|Z^i| \times |\Phi^i| = |Z^i| \times |Z^i| = 3 \times 3$ (see Appendix 2). We also assume that uncertainties affect the outcome of the PHM devices only when the monitored nodes are in an operational state (i.e., $\zeta^2 = \zeta^{17} = 1$ or $\zeta^2 = \zeta^{17} = 2$) and that the state estimation is provided by a classification algorithm independently from the implemented action. Yet, we assume that the network OM knows exactly when the nodes have reached a failed state [27].

3.6 Initial state

The OM exactly knows the starting state of source node $i = 2$ and user node $i = 17$, together with the demands of the users. In particular, we assume that both nodes $i = 2, 17$ are in the AGAN state (i.e., $\zeta^2 = \zeta^{17} = 1$) and that the demand of each user node $i \in N^u$ is equal to d^i_{nom}. This way, we define the initial state of the system, $\mathbf{b}(0)$.

In this framework, our purpose is to find the optimal management policy, which meets the gas demands of the network users at the minimum supply losses.

With respect to the computational effort needed to complete a point-based backup operation (see Section 2.2), we observe that $|\mathcal{A}| = 60$, $|S| = |\Omega| = 648$. Assuming an average size of the set of $\boldsymbol{\alpha}$-vectors equal to $|\Xi| = 150$, the complexity of the point-based backup requires $O(2.44 \cdot 10^{12})$ operations, which can be handled by an average personal computer. Notice, however, that more complex problems would require more sophisticated machines or different algorithms. To appreciate the practical consequences of the definition of the POMDP 7-tuple on the computational burden, we report in Table 2 some examples of different possible settings. From this, it can be easily seen that considering more complex case studies can lead to increments of orders of magnitude in the computational time.

Finally, Table 3 summarizes the pieces of information required to develop the POMDP setting and the corresponding sources to obtain those estimates from a real system.

Difference with $\mathbf{s}(t)$	Number of operations
$\lvert Z^2\rvert$ or $\lvert Z^{17}\rvert$ or $\lvert D^{17}\rvert$ or $\lvert D^{21}\rvert = 4$	$O(5.80 \cdot 10^{12})$
$\lvert D^{16}\rvert$ or $\lvert D^{18}\rvert$ or $\lvert D^{20}\rvert = 3$	$O(8.26 \cdot 10^{12})$
$\lvert \Lambda^2\rvert = 6$	$O(2.93 \cdot 10^{12})$
$\lvert \Lambda^{17}\rvert = 4$	$O(3.26 \cdot 10^{12})$
$\lvert \Lambda^{18}\rvert$ or$\lvert \Lambda^{22}\rvert = 3$	$O(3.67 \cdot 10^{12})$
$\lvert D^{21}\rvert = 4$ and $\lvert D^{16}\rvert = 3$	$O(1.96 \cdot 10^{13})$
$\lvert D^{21}\rvert = 4$ and $\lvert D^{16}\rvert = 3$ and $\lvert \Lambda^{18}\rvert = 3$	$O(2.93 \cdot 10^{13})$

Table 2: Number of operations for backup in different settings

Variable	Source
H	Expert judgment
$\lvert H\rvert$ and $\lvert Z^i\rvert, i \in H$	Trade off: node relevance vs computational times
$D^i, i \in N^u$	Past History
$\lvert D^i\rvert, i \in N^u$	Trade off: node relevance vs computational times
$\lvert \Lambda^i\rvert, i \in H \cup N^s$	Trade off: node relevance vs computational times
$T(\mathbf{s}(t), \mathbf{A}, \mathbf{s}(t+1))$	Available datasets, expert judgment and physics-based models
r	GTN contracts obligations
$O(\mathbf{o}(t+1), \mathbf{A}, \mathbf{s}(t+1))$	PHM validated classification performance
$\mathbf{b}(0)$	DM initial belief

Table 3: Information to build the POMDP setting

4 Results and Comments

To assess the enhancement of the proposed framework, we need to define a reference O&M policy, which the strategy found by the POMDP can be compared to. For this, we consider the case in which no PHM systems are installed in the GTN, which is obtained by reducing the number of states of both nodes $i = 2$ and $i = 17$ from 3 to 2, the last two states being lumped together. The profitability of the GTN operation with no PHM is evaluated by solving a MDP, in which the optimal policy is found through the value iteration method [54], with tolerance for convergence set to 10^{-5} [54]. The solution found is that maximizing the profit by properly setting the source node gas levels while applying corrective maintenance to the degrading nodes.

The same MDP setting is then applied to find the optimal policy obtained in case of perfect PHM, which is then compared with those obtained from the POMDP considering increasing probabilities of degradation state misclassification by the PHM systems. This allows isolating the effect of the

uncertainty due to PHM system underperformance on the profitability of the GTN operations. In particular, the following settings are considered:

(a) the accuracy of the PHM systems is 0.99;

(b) the accuracy of the PHM systems is 0.95;

(c) the accuracy of the PHM systems is 0.9.

Table 4 reports the expected discounted returns of the policies obtained for the considered PHM system configurations, starting from state $\mathbf{s}_0$ or $\mathbf{b}(0)$, in case of MDP or POMDP settings, respectively.

Table 4: Expected discounted returns (in M€) obtained using different configurations of PHMs.

PHM accuracy	100%	99% (setting a)	95% (setting b)	90% (setting c)	No PHM
Expected Discounted Return	877	874	872	869	795

As expected, Table 4 shows that when the knowledge of the degradation states of the nodes is perfect, the network OM reaches the largest possible income from the infrastructure, being always able to act optimally, thus obtaining the largest possible income from the infrastructure. When the PHM system accuracy is reduced, instead, the expected discounted return slightly decreases. Furthermore, it can be seen that the expected return gained when critical nodes i=2 and i=17 are not equipped with a PHM is remarkably smaller than in the setting where the nodes are equipped with PHM systems, even if with reduced accuracy.

Notice that the differences of the the expected discounted returns of the MDP policy (first column in Table 4) and those of the POMDP settings (following columns in Table 4) provide the Values of Perfect Information (VoPI). VoPI represents the possible increase of the expected value of the POMDP setting if the maintenance strategy were decided based on perfect information about the systems health states [64]. For example, the VoPI for POMDP setting c is 877-869= 8M€. This quantity is the maximum value that the OM is willing to pay, to provide perfect PHM systems to the critical nodes.

To further investigate the results, we select from the set B built by the PERSEUS algorithm the subset of beliefs corresponding to the states $\mathbf{s} = [\zeta^2, \zeta^{17}, d^{16}_{nom}, d^{17}_{nom}, d^{18}_{nom}, d^{20}_{nom}, d^{21}_{nom}]$ (i.e., the state in which the gas demand of each user node $i \in U$ is equal to d^i_{nom}). Figures 4, 5 and 6 show in a parallel coordinates [65] the optimal action corresponding to these states, in which, for illustration, we lumped the preventive and corrective maintenance actions into one action, indicated by M. Parallel coordinates enable representing high-dimensional data by drawing vertical, parallel

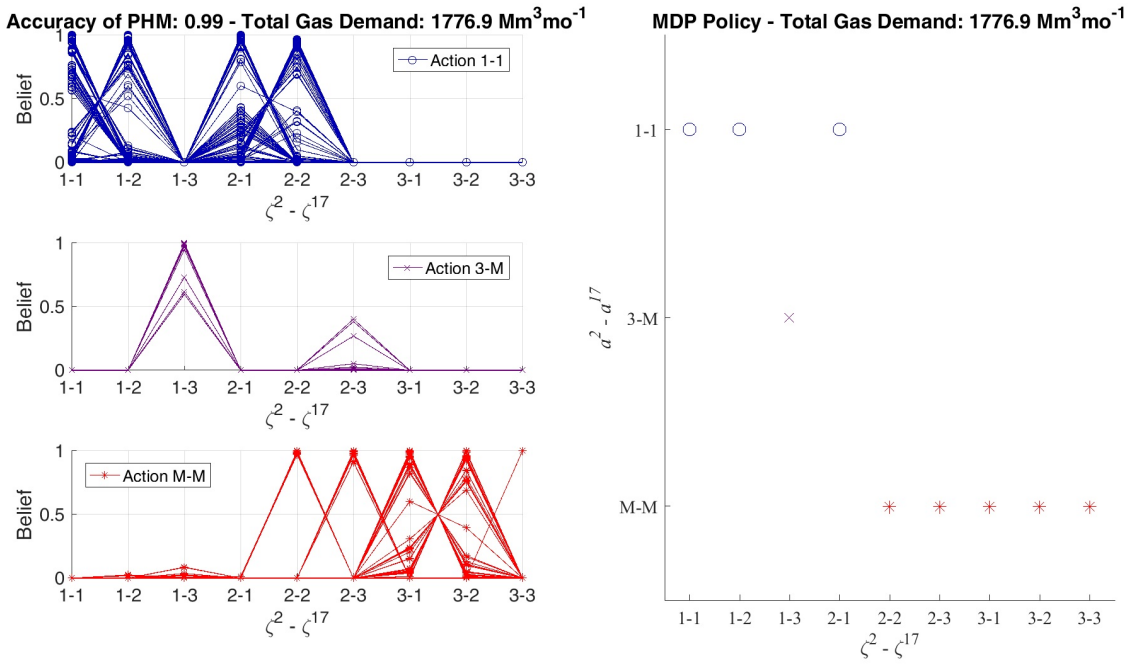


Figure 4: Optimal action as a function of the belief of the states of nodes i=2 and i=17 (1=AGAN; 2=partially degraded; 3=failed), when the accuracy of the PHM is 0.99 (setting a) and the demand level is nominal for all user nodes. Top refers to the actions with $a^2 = 1$ and $a^{17} = 1$ (the output of the source node is set to the nominal value and the user node lets the gas flow); middle to the actions with: $a^2 = 3$ and $a^{17} = M$ (output of the source node set to 50% of the nominal value and user node under maintenance); bottom to the actions with: maintenance on both critical nodes. Right: MDP policy for the selected demand level

and equidistant lines (axes) in number equal to the dimension of the state space. Then, a point in the state space is identified by a polygonal line whose vertices, representing the coordinates of the point, lie on the parallel axes. In our case, the demand level is fixed; then, the parallel coordinates require 9 axes, each corresponding to one out of the combination of the degradation states of the critical nodes. Notice that the number of beliefs contained in the chosen subset is similar for the three considered cases, although it might look different due to the superposition of the lines.

On the left of each Figure, we show the optimal action provided by the PERSEUS algorithm, whereas on the right we show the corresponding optimal actions provided by value-state iteration algorithm in case of perfect PHM systems (i.e., MDP). Notice that the assumption that the unknown degradation states of the nodes are those characterized by the maximum belief lead the optimal actions provided by the MDP and the three POMDPs settings to always coincide, except in few cases in which the states of the two critical nodes are $\zeta^2 = 2$ (partially degraded) and $\zeta^{17} = 3$

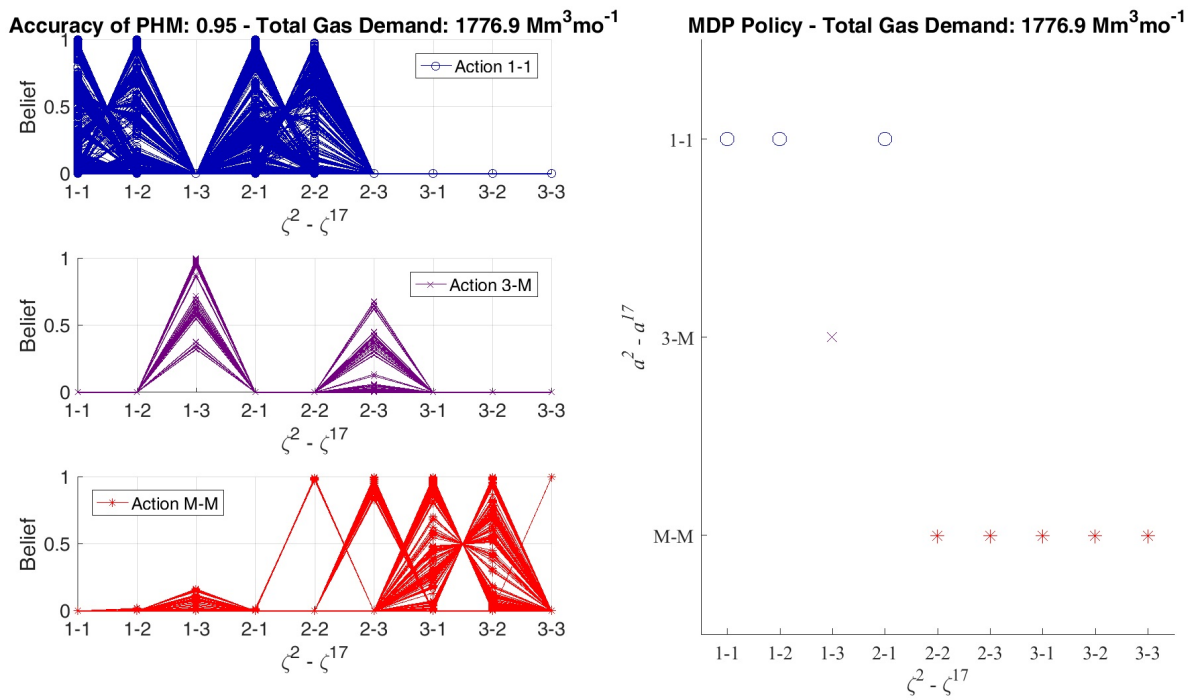


Figure 5: Optimal action as a function of the belief of the states of nodes i=2 and i=17 (1=AGAN; 2=partially degraded; 3=failed) when the accuracy of the PHM is 0.95 (setting b) and the demand level is nominal for all user nodes. Top refers to the actions with: $a^2 = 1$ and $a^{17} = 1$ (the output of the source node is set to the nominal value and the user node lets the gas flow); middle to the actions with: $a^2 = 3$ and $a^{17} = M$ (output of the source node set to 50% of the nominal value and user node under maintenance); bottom to the actions with: maintenance on both critical nodes. Right: MDP policy for the selected demand level

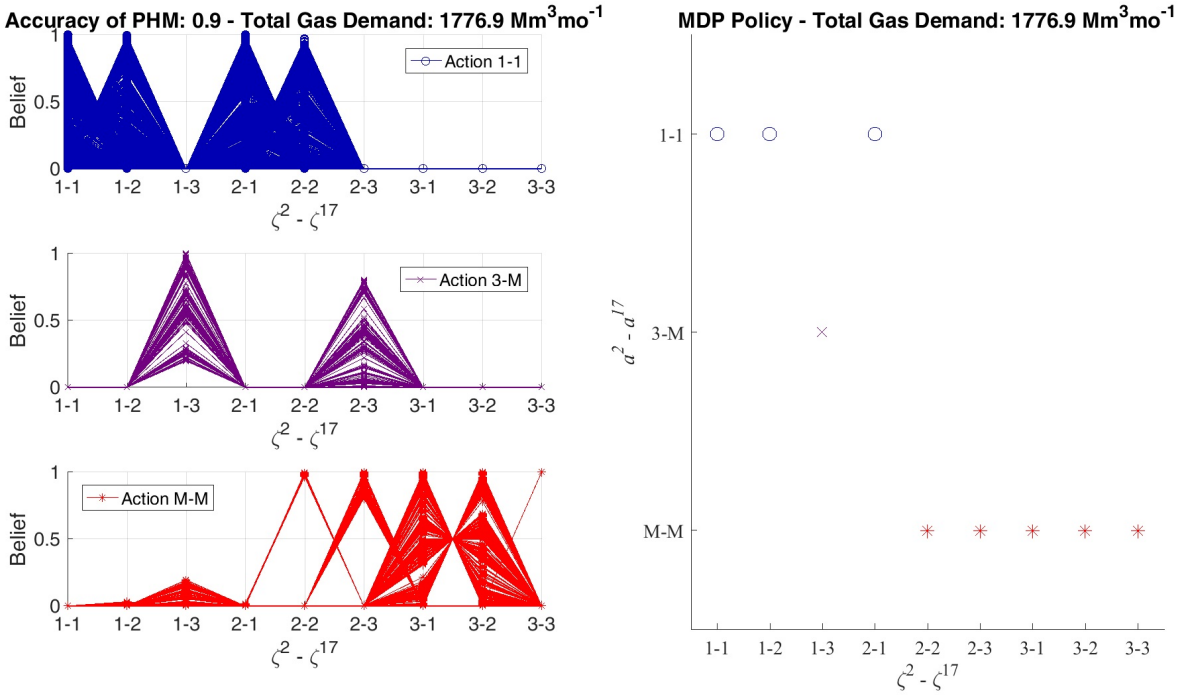


Figure 6: Optimal action as a function of the belief of the states of nodes i=2 and i=17 (1=AGAN; 2= partially degraded; 3= failed) when the accuracy of the PHM is 0.9 (setting c) and the demand level is nominal for all user nodes. Top refers to the actions with: $a^2 = 1$ and $a^{17} = 1$ (the output of the source node is set to the nominal value and the user node lets the gas flow); middle to the actions with: $a^2 = 3$ and $a^{17} = M$ (output of the source node set to 50% of the nominal value and user node under maintenance); bottom to the actions with: maintenance on both critical nodes. Right: MDP policy for the selected demand level

(failed). In these cases, the optimal action provided by the MDP is to perform maintenance on both nodes (Figures 5, 6 and 7 right), whereas the POMDP settings provide output load $= 0.5F^i_{nom}$ to node i=2 and maintenance to node i=17 in a restricted set of beliefs (middle left graphs of Figures 5, 6 and 7). These states are characterized by the presence of some uncertainty on the degradation state of node i=2 (beliefs smaller than 0.9). A justification for this result can be found in the fact that the losses deriving from the impossibility of supplying gas when maintenance is implemented on node i=2 are expected to be larger than the benefits obtained by running the node in order to leverage on the small chances of the node being in the AGAN condition. In a similar way, we can apply this cost-benefit analysis to explain the different optimal actions provided by the MDP and the three POMDPs for the states $\zeta^2 = \zeta^{17} = 2$ (both nodes are partially degraded).
The good match between the actions found in the MDP and POMDP settings justifies the similarities in the results of the Monte Carlo simulations: despite the uncertainties affecting the PHM systems, the OM is able to obtain an expected return that is close to that obtained with a perfect PHM.
The expected return largely depends on two quantities: the average amount of gas successfully supplied to the users of the network, G_s, and the quantity of required gas that remains undelivered G_u. These can be estimated through Monte Carlo simulation.
Consider the discounted economic value of the amount of gas successfully delivered to the GTN users, for the different settings. The differences of this value in case of perfect PHM, $I^{PHM}_{G_s}$, and those of the other settings, $I^a_{G_s}$, $I^b_{G_s}$, $I^c_{G_s}$ and $I^{NoPHM}_{G_s}$, can be summed over the time horizon. The cumulative differences are reported in dot-dashed lines in the subplots of Figure 7. For example, in case of no PHM Figure 7, top-left shows:

$$\sum_{\tau=1}^{t} \left(I^{PHM}_{G_s}(\tau) - I^{noPHM}_{G_s}(\tau) \right) \qquad t = 1, ...[months] \tag{31}$$

Similarly, the differences in the cumulative discounted losses due to the amount of gas that remains undelivered are reported in Figure 7 in dashed line. These differences are larger than 0 when the perfect PHM outperforms the other settings in saving money losses.

In details, when the GTN is not equipped with PHM systems, the loss is mainly due to the fact that a smaller amount of gas is successfully supplied to the users (thus representing a decrease in the gain), which in turn leads also to the payment of significant penalties. Notice that this subplot has a different scale on the ordinate axis, due to the different order of magnitude of the losses, and that the sum of the cumulative differences in supply ($33M$€) and undelivered amount ($47M$€) almost equals the difference in the values of the initial state in Table 4 for the two considered MDP settings ($877 - 795 = 83M$€). This confirms that these two parameters almost entirely justify the differences among the expected returns of the different PHM settings.

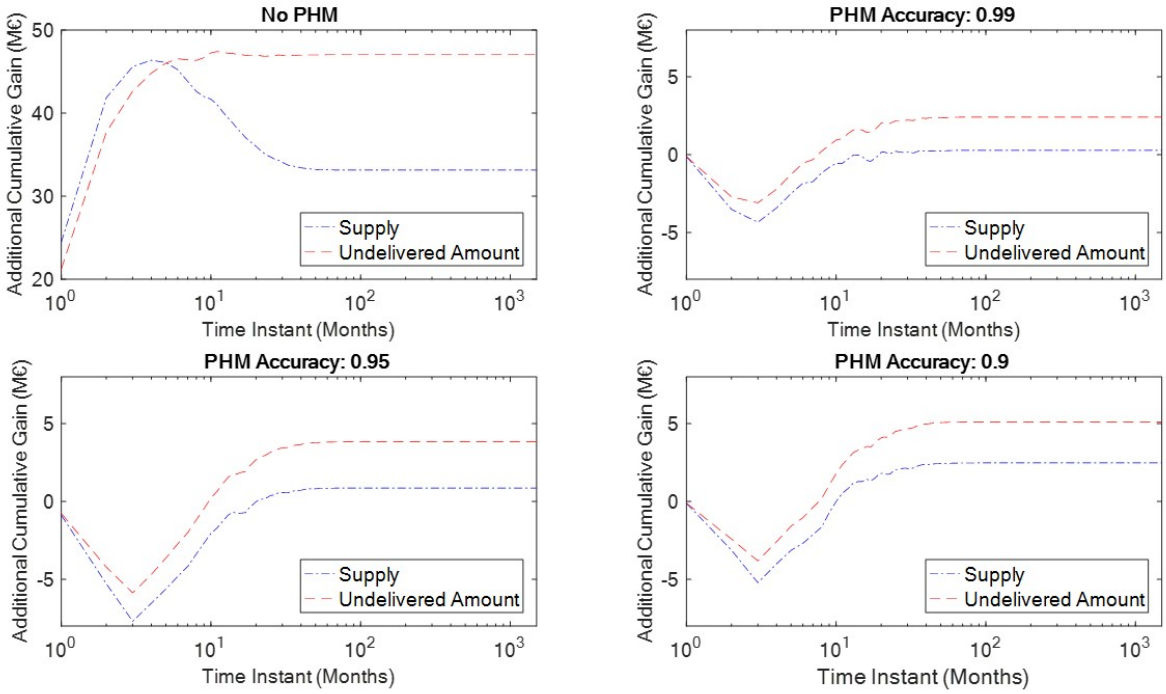


Figure 7: Additional cumulative gain obtained with perfect knowledge of the degradation states. Comparison with: no PHM(top left); PHM accuracy: 0.99(top right); PHM accuracy: 0.95(bottom left); PHM accuracy: 0.9(bottom right)

Concerning accuracies 0.99 and 0.95, we can see that the OM is able to successfully satisfy the requests of the users almost in the same measure as when the degradation states are perfectly known. Yet, the OM incurs into losses because the amount of undelivered gas is larger. Finally, when the accuracy is set to 0.9, the losses are mostly due to the differences in the undelivered amount of gas, although differences in the successful supply of gas provide an important contribution.
Furthermore, we notice that when the critical nodes are equipped with a PHM system with accuracy smaller than 1, the OM is able to gather an income larger than that in the case of a perfect PHM system at the beginning of the operation of the GTN: in the first 2-3 months, the difference in the cumulated gain is negative; nonetheless, perfect PHM systems enable a better management strategy, as in the following time instants they yield a larger discouned expected return (i.e., the cumulative differences of both supply and undeliverable amounts become positive). This means that the optimal policy yields the largest return on the long run, thus taking into account the future effects of present actions. The effect of the classification errors in imperfect PHMs is to lead the OM to act so as to maximize the immediate return (i.e., greedy actions), which prevents achieving future gains.

Finally, with respect to the computational times, to find the MDP optimal policies, an 8GB RAM machine running an Intel Core i-7 processor of 2.20 GHz took:

1. 82 seconds in case no PHM system is installed in the GTN (setting 1).
2. 225 seconds in both cases in which only one PHM system is used (settings 2 and 3).
3. 511 seconds in case both nodes are equipped with perfect PHM systems (setting 4).

The PBM PERSEUS algorithm [30] is run with a set B containing 10^5 reachable beliefs, with a tolerance for convergence $\varepsilon = 10^{-5}$. The time required to find the optimal policy is:

1. Almost 24 hours in case of PHM accuracy equal to 0.99.
2. Almost 36 hours in case of PHM accuracy equal to 0.95.
3. Almost 48 hours in case of PHM accuracy equal to 0.9.

Due to the large computational times, no sensitivity analysis has been performed on the algorithm and model parameters.

5 Conclusion

The management of the operations of GTNs is a complex task due to the many and diverse aspects it involves. On the one hand, the necessity of satisfying the demand of the users requires the proper setting of the load on the compressor stations to not incur into losses; on the other hand, the need of preserving the health conditions of the network elements drives the management decisions towards the implementation of maintenance activities, which can reduce gas supply. Management strategies are affected also by the errors in the PHM systems used to track the states of the network elements: an incorrect estimation of the GTN element degradation state may lead the OM to take unprofitable and unsafe decisions, with potential major consequences (e.g., reducing the amount of delivered gas).

In this work, we relied on the POMDP framework to provide a tool aimed at supporting the decision making for the operation of a GTN. The framework is such to enable the comprehensive management of the GTN by taking into account the diverse aspects that are involved in the selection of a strategy, i.e., the uncertain estimations of the degradation states of the critical network elements, the variability of the gas demand and the effects of the maintenance tasks.

The application to a realistic GTN proved the capability of the framework. Furthermore, the method provides the OM with a solution, which can be easily analyzed to evaluate the factors most affecting GTN. Finally, the modeling approach has allowed estimating the VoPI of PHM sys-

tems with different accuracy levels, which is a useful support to make decisions about investments in PHM.

Some criticalities emerged from this study. Namely, the computational times are long when the state space is large. The consequences are that sensitivity analyses cannot be performed in reasonable time and that a compromise solution needs to be found between the level of detail of the model and the computational burden. Future research work will focus on other methods for belief function approximation, to investigate whether and to which extent this issue can be overcome.

Moreover, one of the underlying assumptions of the work is that both the equipment degradation process and the accuracy of PHM do not depend on the amount of gas circulating through the network (see Appendix 2). Future research work will focus on overcoming this limitation, which entails solving a significantly more complex optimization problem where the algorithm calculating the maximum flow through the network has to encode the uncertainty on the equipment state.

Authors' contribution

All the authors contributed equally to this work.

References

[1] Praks, P., Kopustinskas, V., Masera, M., 2015, *Probabilistic modelling of security of supply in gas networks and evaluation of new infrastructure*, Reliability Engineering & System Safety, Vol. 144, 2015, pp. 254-264.

[2] Praks, P., Kopustinskas, V., Masera, M., 2017, *Monte-Carlo-based reliability and vulnerability assessment of a natural gas transmission system due to random network component failures*, Sustainable and Resilient Infrastructure.

[3] Kopustinskas, V., Praks, P., 2015, *Bottleneck analysis of the gas transmission network using ProGasNet*, Safety and Reliability of Complex Engineered Systems, pp. 1259-1266.

[4] Kabirian, A, Hemmati, M.R., 2007, *A strategic planning model for natural gas transmission networks*, Energy Policy, Vol. 35, Issue 11, pp. 5656–5670.

[5] Peretti, A., Toth, P., 1982, *Optimization of a pipe-line for the natural gas transportation*, European Journal of Operational Research, Vol. 11, Issue 3, pp. 247-254.

[6] Ríos-Mercado, R.Z., 2002, *Natural Gas Pipeline Optimization*, Handbook of Applied Optimization, Chapter: 18.8.3, pp. 813-825.

[7] Wu, S., Ríos-Mercado, R.Z., Boyd, E.A., Scott, L.R., 2000, *Model relaxations for the fuel cost minimization of steady-state gas pipeline networks*, Mathematical and Computer Modelling, Vol. 31, Issue 2, pp. 197-220.

[8] Goldberg, D.E., 1987, *Computer-aided pipeline operation using genetic algorithms and rule learning. PART I: Genetic algorithms in pipeline optimization*, Engineering with Computers, Vol.3, Issue 1, pp. 35-45.

[9] Zheng, Q.P., Rebennack, S., Iliadis, N.A., Pardalos, P.M., 2010, *Optimization Models in the Natural Gas Industry*, Handbook of Power Systems I, pp. 121-148, Springer Berlin Heidelberg.

[10] Ríos-Mercado, R.Z., Borraz-Sánchez, C., 2015, *Optimization problems in natural gas transportation systems: A state-of-the-art review*, Applied Energy, Vol. 147, pp. 536-555.

[11] Lo, K.L., 1984, *Optimisation and analysis of gas supply networks*, Transactions of the Institute of Measurement and Control, Vol. 6, Issue 4, pp. 253-260.

[12] Sacco, T., Compare, M., Zio, E., Sansavini, G., 2017, *Portfolio Decision Analysis for Risk-Based Maintenance of Gas Networks.submitted for publication*

[13] Salo A., Keisler J., Morton A., Eds. *Portfolio Decision Analysis – Improved Methods for Resource Allocation*, International Series in Operations Research & Management Science, Vol. 162, Springer-Verlag, 2011.

[14] Suomilammi, A., & Leppäkoski, J., 2006, *Gas compressor unit performance monitoring using fuzzy clustering*, 23rd World Gas Conference.

[15] Mironov, A., Doronkin, P., Priklonsky, A., Kabashkin I., 2015, *Condition Monitoring Of Operating Pipelines With Operational Modal Analysis Application*, Transport and Telecommunication Journal, Vol. 16, Issue 4, pp. 305-319.

[16] Sankararaman, S., Ling, Y., Shantz, C., Mahadevan, S., 2011, *Uncertainty quantification in fatigue crack growth prognosis*, International Journal of Prognostics and Health Management, Vol. 2, Issue 1.

[17] Cadini, F., Zio, E., Avram, D., 2009, *Monte Carlo-based filtering for fatigue crack growth estimation*, Probabilistic Engineering Mechanics, Vol. 24, Issue 3, pp. 367-373.

[18] Cadini, F., Sbarufatti, C., Corbetta, M., Giglio, M., 2017, *A particle filter-based model selection algorithm for fatigue damage identification on aeronautical structures*, Structural Control and Health Monitoring, Vol. 24, Issue 11, art. number e2002.

[19] Leser, P.E., Hochhalter, J.D., Warner, J.E., Newman, J.A., Leser, W.P., Wawrzynek, P.A., Yuan, F.-G., 2017, *Probabilistic fatigue damage prognosis using surrogate models trained via three-dimensional finite element analysis*, Structural Health Monitoring, Vol. 16, Issue 3, pp. 291-308.

[20] Ling, Y., Mahadevan, S., 2012, *Integration of structural health monitoring and fatigue damage prognosis*, Mechanical Systems and Signal Processing, Vol. 28, pp.89-104.

[21] Sankararaman, S., 2015, *Significance, interpretation, and quantification of uncertainty in prognostics and remaining useful life prediction*, Mechanical Systems and Signal Processing, Vol. 52-53, Issue 1, pp. 228-247.

[22] Ding, F., Tian, Z., Zhao, F., Xu, H., 2018, *An integrated approach for wind turbine gearbox fatigue life prediction considering instantaneously varying load conditions*, Renewable Energy, Vol. 129, pp. 260-270.

[23] Cassandra, A., 1998, *Exact and Approximate Algorithms for Partially Observable Markov Decision Processes*, Ph.D. Thesis, Brown University.

[24] Compare, M., Marelli, P., Baraldi, P., Zio, E., 2018, *A Markov decision process framework for optimal operation of monitored multi-state systems*, Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, Vol. 232, Issue 6, pp. 677-689.

[25] Saxena, A., Celaya, J., Saha, B., Saha, S., Goebel, K., 2010, *Evaluating prognostics performance for algorithms incorporating uncertainty estimates*, IEEE Aerospace Conference Proceedings, art. number 5446828.

[26] Saxena, A., Celaya, J., Saha, B., Saha, S., Goebel, K., 2012, *Metrics for offline evaluation of prognostic performance*, International Journal of Prognostics and Health Management, Vol. 1, Issue, 1.

[27] Moghaddass, R., Zuo, M.J., 2014 *An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process*, Reliability Engineering and System Safety, Vol. 124, pp. 92-104.

[28] Compare, M., Bellani, L., Zio, E., 2017, *Reliability model of a component equipped with PHM capabilities*, Reliability Engineering and System Safety, Vol. 168, pp. 4-11.

[29] Compare, M., Bellani, L., Zio, E., 2017, *Availability Model of a PHM-Equipped Component*, IEEE Transactions on Reliability, Vol. 66, Issue 2, pp. 487-501.

[30] Spaan, M., Vlassis, N., 2005, *PERSEUS Randomized point-based value iteration for POMDPs*, Journal of Artificial Intelligence Research, Vol. 24, pp. 195–220.

[31] Shani, G., Pineau, J., Kaplow, R., 2013, *A survey of point-based POMDP solvers,* Autonomous Agents and Multi-Agent Systems, Vol. 1, Issue 1, pp. 1-51.

[32] Thiébaux, S., Cordier, M., Jehl, O., Krivine, J., 1996, *Supply Restoration in Power Distribution Systems - A Case Study in Integrating Model-Based Diagnosis and Repair Planning*, Proceedings UAI, pp. 525–532.

[33] Memarzadeh, M., Pozzi, M., 2016, *Value of information in sequential decision making: Component inspection, permanent monitoring and system-level scheduling*, Reliability Engineering and System Safety, Vol. 154, pp. 37–151.

[34] Špačková, O., Straub, D., 2017, *Long-term adaption decisions via fully and partially observable Markov decision processes*, Sustainable and Resilient Infrastructure, Vol. 2, Issue 1, pp. 37-58.

[35] Papakostantinou, K.G., Shinozua, M., 2014, *Planning optimal inspection and maintenance policies via dynamic programming and Markov processes. Part II: POMDP implementation*, Reliability Engineering and System Safety, Vol. 130, pp. 214-224.

[36] Papakostantinou, K.G., Shinozua, M., 2014, *Optimum inspection and maintenance policies for corroded structures using partially observable Markov decision processes and stochastic, physically based models*, Probabilistic Engineering Mechanics, Vol. 37, pp. 93-108.

[37] Fedorowicz, R., Kołodziński, E., Solarz L., 2002, *Flow Modeling in Gas Transmission Networks. Part I: Mathematical Model*, Journal of Theoretical and Applied Mechanics, Vol. 40, Issue 4, pp. 873-894.

[38] Munoz, J., Jimenez-Redondo, N., Perez-Ruiz, J., Barquin, J., 2003, *Natural gas network modeling for power systems reliability studies*, IEEE Bologna Power Tech Conference Proceedings, Vol. 4.

[39] Üster, H., Dilaveroğlu, Ş., 2014, *Optimization for design and operation of natural gas transmission networks*, Applied Energy, Vol. 133, pp. 56-69.

[40] Han, F., Zio, E., Kopustinskas, V., Praks, P., 2017, *Quantifying the importance of elements of a gas transmission network from topological, reliability and controllability perspectives, considering capacity constraints*, Risk, Reliability and Safety: Innovating Theory and Practice - Proceedings of the 26th European Safety and Reliability Conference, ESREL 2016.

[41] Su, H., Zhang, J., Zio, E., Yang, N., Li, X., Zhan, Z., *An integrated systemic method for supply reliability assessment of natural gas pipeline networks*, 2018, Applied Energy, Vol. 209, pp.489 - 501.

[42] Lisnianski, A., Levitin, G., 2003, *Multi-State System Reliability: Assessment, Optimization and Applications*, Chapter 1, pp.15-50, World Scientific, New Jersey, London, Singapore, Hong Kong.

[43] Lisnianski, A., Laredo, D., Haim, H.B., 2016, *Multi-state Markov Model for Reliability Analysis of a Combined Cycle Gas Turbine Power Plant*, 2016 Second International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations Management (SMRLO), Beer Sheva, pp. 131-135.

[44] Fink, O., Zio, E., Weidmann U., 2014, *Quantifying the reliability of fault classifiers*, Information Sciences, Vol. 266, pp. 65-74.

[45] Cannarile, F., Compare, M., Baraldi, P., Di Maio, F., Zio, E., 2018, *Homogeneous continuous-time, finite-state hidden semi-markov modeling for enhancing empirical classification system diagnostics of industrial components*, Machines, Vol. 6, Issue 3, art. number=34.

[46] Bianco, V., Scarpa, F., Tagliafico, L.A., 2014, *Scenario analysis of nonresidential natural gas consumption in Italy*, Applied Energy, Vol. 113, pp. 392-403.

[47] Gorucu F.B., 2004, *Evaluation and Forecasting of Gas Consumption by Statistical Analysis*, Energy Sources, Vol. 26, Issue 3, pp. 267-276.

[48] Holditch, S., Chianelli, R., 2008, *Factors That Will Influence Oil and Gas Supply and Demand in the 21st Century*, MRS Bulletin, Vol. 33, Issue 4, pp. 317-323.

[49] Papakostantinou, K.G., Shinozua, M., 2014, *Planning optimal inspection and maintenance policies via dynamic programming and Markov processes. Part I: Theory*, Reliability Engineering and System Safety, Vol. 130, pp. 202-213.

[50] Rachelson, E., Quesnel, G., Garcia F., Fabiani, P., 2008, *A simulation-based approach for solving generalized semi-markov decision processes*, Proceedings of the 2008 Conference on ECAI 2008: 18th European Conference on Artificial Intelligence, pp. 583–587.

[51] Boykov, Y., Kolmogorov, V., 2004, *An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No. 9, pp. 1124-1137, Sept. 2004

[52] Pham, H., Wang, H., 1996, *Imperfect maintenance*, European Journal of Operational Research, Vol. 94, Issue 3, pp. 425-438.

[53] Dopheide, D., Taux, G., Krey, E.-A., 1990, *Development of a New Fundamental Measuring Technique for the Accurate Measurement of Gas Flowrates by Means of Laser Doppler Anemometry*, Metrologia, Vol. 27, Issue 5, pp. 65-73.

[54] Sutton, R.S., Barto, A.G., 1998, *Introduction to Reinforcement Learning (1st ed.)*. MIT Press, Cambridge, MA, USA.

[55] Sondik, E.J., 1978, *The Optimal Control of Partially Observable Markov Processes over the Infinite Horizon: Discounted Costs*, Operations Research, Vol. 26, Issue 2, pp. 282-304.

[56] Cassandra, A., Kaelbling, L.P., Littman M.L., 1994, *Acting Optimally in Partially Observable Stochastic Domains,* Technical Report. Brown University, Providence, RI, USA.

[57] Cassandra, A., Kaelbling, L.P., Zhang, N.L., 1997, *Incremental Pruning: A Simple, Fast, Exact Method for Partially Observable Markov Decision Processes*, Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence, pp. 54-61.

[58] Lovejoy, W., 1991, *Computationally Feasible Bounds for Partially Observed Markov Decision Processes*, Operations Research, Vol. 39, Issue 1, pp. 162-175.

[59] Pineau, J, Gordon, G, Thrun, S., 2003, *Point-based value iteration: An anytime algorithm for POMDPs*, International Joint Conference on Artificial Intelligence (IJCAI), pp. 1025-1032.

[60] Hauskrecht M., 2000, *Value-Function Approximations for Partially Observable Markov Decision Processes*, Journal of Artificial Intelligence Research, Vol. 13, pp. 33-94.

[61] Praks, P., Kopustinskas, V., 2014, *Monte-Carlo based reliability modelling of a gas network using graph theory approach*, 2014 Ninth International Conference on Availability, Reliability and Security, Fribourg, pp. 380-386.

[62] Sherlaw-Johnson, C., Gallivan, S., Burridge, J., 1995, *Estimating a Markov Transition Matrix from Observational Data*, The Journal of the Operational Research Society, Vol. 46, Issue 3, pp. 405–410.

[63] Zio E., 2009, *Computational Methods for Reliability and Risk Analysis*, Series on Quality, Reliability and Engineering Statistics: Volume 14, world Scientific, Singapore.

[64] Vilkkumaa, E., Liesiö, J., Salo, A., 2014, *Optimal strategies for selecting project portfolios using uncertain value estimates*, European Journal of Operational Research, Vol. 233, Issue 3, pp. 772–783.

[65] Inselberg, A., & Dimsdale, B., 1990, *Parallel coordinates: a tool for visualizing multi-dimensional geometry*, Proceedings of the 1st conference on Visualization '90 (VIS '90), Arie Kaufman (Ed.). IEEE Computer Society Press, Los Alamitos, CA, USA, pp. 361-378.

[66] Kurniawati, H., Hsu, D., Lee, W.S., 2008, *SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces*, Proc. Robotics: Science and Systems.

6 Appendix 1

The first step in PERSEUS builds the belief set B through random exploration: several trajectories through the belief space are generated by randomly sampling an action and an associated observation at each time step, so that they can be used for the belief update procedure. Once the belief set B is built, it remains fixed throughout the execution of the algorithm.
The value function $V(0)$ is initialized to a single vector with all components equal to $\frac{1}{1-\gamma}\min_{s,A}\rho(s,A)$, which corresponds to collecting the minimal possible reward in every step. This way, we are ensured that the value function is initialized to a lower bound that can always be improved [9].
After this, PERSEUS implements the *backup* procedure as follows:

1. Set $\Xi_{n+1}=\emptyset$, i.e., initialize the set of non-improved (by the *backup* operation) belief points $\tilde{B}$ to B.
2. Sample a random belief point uniformly from $\tilde{B}$ and compute $\boldsymbol{\alpha}^{n+1}$=*backup*$(\mathbf{b})$.
   (a) If $\mathbf{b}\cdot\boldsymbol{\alpha}^{n+1}\geq V_n(\mathbf{b})$, add $\boldsymbol{\alpha}^{n+1}$ to Ξ_{n+1}.
   (b) Otherwise, add $\alpha'=\arg\max_{\Xi_n}(\mathbf{b}\cdot\boldsymbol{\alpha}^n)$.
3. Remove from $\tilde{B}$ all points whose value is improved by the newly calculated $\boldsymbol{\alpha}$-vector: $\tilde{B}=\{\mathbf{b}\in B: V_{n+1}(\mathbf{b})<V_n(\mathbf{b})\}$
4. If $\tilde{B}=\emptyset$, stop, else go to 2.

This backup stages are repeated until a convergence criterion is met; in our work, we stopped the iterations when the relative difference between the value functions estimated on all the belief set B between two consecutive iterations is below a tolerance ε:

$$\frac{\sum_{\mathbf{b}\in B}V_{n+1}(\mathbf{b})-\sum_{\mathbf{b}\in B}V_n(\mathbf{b})}{\sum_{\mathbf{b}\in B}V_n(\mathbf{b})}<\varepsilon$$

7 Appendix 2

Here we report the transition and observation matrices that have been used for the calculations in the case study. Notice that a dash ('-') is used to indicate that the associated action is not available when the state of the node is the selected one.

1. Transition matrices for node 2 with no PHM:

$$\mathbf{T}_1^{i=2} = \begin{bmatrix} 0.7953 & 0.2047 \\ - & - \end{bmatrix} \quad \mathbf{T}_2^{i=2} = \begin{bmatrix} 0.8442 & 0.1558 \\ - & - \end{bmatrix}$$

$$\mathbf{T}_3^{i=2} = \begin{bmatrix} 0.8926 & 0.1074 \\ - & - \end{bmatrix} \quad \mathbf{T}_4^{i=2} = \begin{bmatrix} 1 & 0 \\ 0.5988 & 0.4012 \end{bmatrix}$$

2. Transition matrices for node 17 with no PHM:

$$\mathbf{T}_1^{i=17} = \begin{bmatrix} 0.8504 & 0.1496 \\ - & - \end{bmatrix} \quad \mathbf{T}_2^{i=17} = \begin{bmatrix} 0.9248 & 0.0752 \\ - & - \end{bmatrix}$$

$$\mathbf{T}_3^{i=17} = \begin{bmatrix} 0.9688 & 0.0312 \\ - & - \end{bmatrix} \quad \mathbf{T}_4^{i=17} = \begin{bmatrix} 1 & 0 \\ 0.6990 & 0.3010 \end{bmatrix}$$

3. Transition matrices for node 2 with PHM:

$$\mathbf{T}_1^{i=2} = \begin{bmatrix} 0.55 & 0.3 & 0.15 \\ 0 & 0.75 & 0.25 \\ - & - & - \end{bmatrix} \quad \mathbf{T}_2^{i=2} = \begin{bmatrix} 0.65 & 0.25 & 0.1 \\ 0 & 0.8 & 0.2 \\ - & - & - \end{bmatrix} \quad \mathbf{T}_3^{i=2} = \begin{bmatrix} 0.75 & 0.2 & 0.05 \\ 0 & 0.85 & 0.15 \\ - & - & - \end{bmatrix}$$

$$\mathbf{T}_4^{i=2} = \begin{bmatrix} 1 & 0 & 0 \\ 0.9 & 0.1 & 0 \\ - & - & - \end{bmatrix} \quad \mathbf{T}_5^{i=2} = \begin{bmatrix} - & - & - \\ - & - & - \\ 0.4 & 0.2 & 0.4 \end{bmatrix}$$

4. Transition matrices for node 17 with PHM:

$$\mathbf{T}_1^{i=17} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0 & 0.7 & 0.3 \\ - & - & - \end{bmatrix} \quad \mathbf{T}_2^{i=17} = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0 & 0.85 & 0.15 \\ - & - & - \end{bmatrix}$$

$$\mathbf{T}_3^{i=17} = \begin{bmatrix} 0.95 & 0.03 & 0.02 \\ 0 & 0.95 & 0.05 \\ - & - & - \end{bmatrix} \quad \mathbf{T}_5^{i=17} = \begin{bmatrix} 1 & 0 & 0 \\ 0.8 & 0.2 & 0 \\ - & - & - \end{bmatrix}$$

$$\mathbf{T}_6^{i=17} = \begin{bmatrix} - & - & - \\ - & - & - \\ 0.4 & 0.3 & 0.3 \end{bmatrix}$$

5. Transition matrix relative to the gas demand variability of users with 2 levels of gas demand:

$$\mathbf{T} = \begin{bmatrix} 0.1 & 0.9 \\ 0.1 & 0.9 \end{bmatrix}$$

6. Transition matrix relative to the gas demand variability of users with 2 levels of gas demand:

$$\mathbf{T} = \begin{bmatrix} 0 & 0.75 & 0.25 \\ 0.1 & 0.8 & 0.1 \\ 0.05 & 0.2 & 0.75 \end{bmatrix}]$$

7. Observation matrices for both node 2 and node 17 with PHM with accuracy 0.99:

$$\mathbf{O}^{i=2} = \mathbf{O}^{i=17} = \begin{bmatrix} 0.99 & 0.01 & 0 \\ 0.01 & 0.99 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

8. Observation matrices for both node 2 and node 17 with PHM with accuracy 0.95:

$$\mathbf{O}^{i=2} = \mathbf{O}^{i=17} = \begin{bmatrix} 0.95 & 0.05 & 0 \\ 0.05 & 0.95 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

9. Observation matrices for both node 2 and node 17 with PHM with accuracy 0.9:

$$\mathbf{O}^{i=2} = \mathbf{O}^{i=17} = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.1 & 0.9 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$