

Cipollini F, Oneto L, Coraddu A, Murphy AJ, Anguita D.

**Condition-Based Maintenance of Naval Propulsion Systems: Data Analysis
with Minimal Feedback.**

Reliability Engineering and System Safety 2018, 117, 12-23.

Copyright:

© 2018. This manuscript version is made available under the [CC-BY-NC-ND 4.0 license](#)

DOI link to article:

<https://doi.org/10.1016/j.ress.2018.04.015>

Date deposited:

11/05/2018

Embargo release date:

19 April 2019



This work is licensed under a

[Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence](#)

Condition-Based Maintenance of Naval Propulsion Systems: Data Analysis with Minimal Feedback.

Francesca Cipollini^a, Luca Oneto^a, Andrea Coraddu^{b,*}, Alan John
Murphy^b, Davide Anguita^a

^a*DIBRIS - University of Genova, Via Opera Pia 13, I-16145 Genova, Italy*

^b*School of Marine Science and Technology
Newcastle University, Newcastle upon Tyne, NE1 7RU, UK*

Abstract

The behavior and interaction of the main components of Ship Propulsion Systems cannot be easily modeled with a priori physical knowledge, considering the large amount of variables influencing them. Data-Driven Models (DDMs), instead, exploit advanced statistical techniques to build models directly on the large amount of historical data collected by the modern on-board automation systems, without requiring any a priori knowledge. DDMs are extremely useful when it comes to continuously monitoring the propulsion equipment to avoid Preventive or Corrective Maintenance and take decisions based on the actual condition of the propulsion plant. Unfortunately, DDMs need a large amount of data to achieve satisfying performances. While sensor data are cheap and easy to collect, labeling them with the actual state of decay of a component can be quite expensive and in some cases unfeasible. In this paper, the authors investigate the problem of performing Condition-Based Maintenance through the use of DDMs. First, state-of-the-art supervised learning techniques are adopted, which require a large amount of labeled sensor data in order to be deployed. Then, an unsupervised learning approach is developed as it minimize the feedback of the operators in labeling

*Corresponding Author

Email addresses: francesca.cipollini@edu.unige.it (Francesca Cipollini), luca.oneto@unige.it (Luca Oneto), andrea.coraddu@newcastle.ac.uk (Andrea Coraddu), a.j.murphy@newcastle.ac.uk (Alan John Murphy), davide.anguita@unige.it (Davide Anguita)

the sensor data. A naval vessel, characterized by a combined diesel-electric and gas propulsion plant, has been exploited to show the effectiveness of the proposed approaches. Confidentiality constraints with the Navy require us to use a real-data validated simulator and the dataset has been published for free use through the UCI repository.

Keywords: Data Analysis, Naval Propulsion Systems, Condition-Based Maintenance, Supervised Learning, Unsupervised Learning, Minimal Feedback.

1. Introduction

Data Analysis (DA) is improving our ability to understand complex phenomena much more rapidly than a priori physical models have done in the past [1, 2, 3, 4]. Real-world systems are usually very complex and influenced by many exogenous factors, which make them hard to model only relying on the a priori knowledge of the problem [5, 3, 6]. On the contrary, Data-Driven Models (DDMs) exploit advanced statistical techniques to build models directly based on the significant amount of historical data which nowadays are produced and stored by the logging and monitoring apparatus, without requiring any a priori knowledge of the underlining physical phenomena [7, 8, 9]. These characteristics make DDMs a suitable solution in all those contexts where large amount of historical data is available, such as manufacturing [3, 10, 11], communications [12, 13, 14], finance [15, 16, 17], healthcare [18, 19, 20], social networks [21], commerce [22], and transportation [23, 24, 25].

Many of these sectors traditionally based their profit, competitiveness and strategic decisions on empirical experience, the intuition of talented people, and on simplified models built upon a priori knowledge of the specific problems [26, 3]. This approach obviously requires a significant amount of time and experience. On the contrary/ At the same time - not so clear, during the last decades, production plants and products have been equipped with many sensors for different purposes: automation, quality check, monitoring, and logging. The results of this process is the availability of a huge amount of historical and real-time data [27]. Moreover, additional heterogeneous sources such as weather information [28], product perceived quality [29], and employee satisfaction [30] can be used to enrich this data lake and improve the understanding of the business [31]. Recently, industries have realized

that these data, despite their management costs, can be considered as an excellent opportunity to improve their business. For example, they can adopt this historical information to create new services or improve the quality of their products [27]. In particular, they can leverage this huge amount of data thanks to the DMMs which can rapidly and effectively extract useful and actionable information [32, 5].

In the shipbuilding industry, one of the main objectives of ship building companies is to improve the technological quality of their products. For example, they design more efficient hull shapes and propeller geometries, study innovative propulsion systems (i.e.: Voith Schneider and Waterjet propulsor), design hybrid propulsive systems, and reduce the overall production costs [33, 34]. Recently, many ship building companies are evaluating different DA solutions for improving the quality of their products, for monitoring the equipment, and for maintenance purposes as integrative activities to their core business [26, 35]. In fact, ships are already equipped with a network of sensors that collect data for security, diagnostic and monitoring purposes, which DA can directly exploit by taking advantage of these technologies [26, 23]. DA, for example, offers the possibility to extract, from the raw sensor data, useful information about the efficiency of the ship [25], to identify operational profiles [36], to reduce the fuel consumption [37], and to improve maintenance activities [38]. These data represent strategic information for shipyards, operators, ship owners, and crews, since they can be used for advisory, control, and fault detection purposes [36]. Furthermore, DDMs allow exploiting exogenous data as well, such as weather information, which could contain some hidden information, potentially not easily representable with a conventional approach [37].

Among the different problems, maintenance is probably the most critical since it could require drydocking, and the cost of retrieving a stricken vessel offshore is non-trivial [39, 40]. Correct maintenance ensures that a ship works as it was designed, with the desired level performances, without impacting the service [41]. Maintenance policies can be divided into three main categories [42]: Corrective (CM), Preventive (PM), and Condition-Based (CBM).

In the past, the most common approach to this problem relied on the CM where maintenance is performed only after a breakdown of a component [43]. However, replacing a malfunctioning component after it has failed during service, results in exceptional costs and inevitable lower incomes. In PM, instead, a component is replaced when it reaches the end of its life cycle, which is computed through a conservative average estimation. Simi-

larly to CM, PM can bring unnecessary costs too, if the replaced component could have been used more than originally forecast. Moreover, PM does not guarantee to limit the number of faults in a fleet, since a breakdown could still happen before the replacement takes place. In PM, there is a trade-off between the number of breakdowns and the lifetime estimation of the components, which is not easy to reach since the actual ship usage can be very different from ship to ship. CBM, instead, aims at reducing the costs of CM and PM by relying on the exact decay state of each component and then by efficiently planning its maintenance. Since, in most cases, the decay state of each component cannot be tracked with a sensor, CBM requires a model able to predict it based on other sensors available. Considering the estimated state of decay, it is possible to schedule each component's replacement before failures occur, maximizing its life cycle, according to the time required for each maintenance and to the geographical location of the ship [40]. As a result, the additional costs of CM and PM can be replaced with the lower ones of equipping the propulsion system with sensors and by collecting, storing, and analyzing these data for the purpose of creating effective predictive DDMs [39, 40].

In this paper, the authors address the problem of building effective predictive models of the main components decay state in a Naval Propulsion System (NPS) for CBM purposes. While physical models have been exploited in the past, recently, the ones based on statistical analysis have shown to outperform them considerably [3]. DDMs are adopted to correlate the available sensors measurements with the decay state of the different components of an NPS. Specifically, two approaches are here proposed and compared. First, the authors build virtual sensors able to continuously estimate the state of decay of the components based on other sensors measurements which are indirectly influenced by this decay. Then the authors try to build simplified virtual sensors able to detect if the component state of decay is above or below a certain threshold. Results will show that, in the first case, a significant amount of historical data needs to be collected, together with the actual state of decay of each component. Obviously, this is not feasible in a real-world scenario where the labeling process requires the intervention of an experienced operator and, in some cases, to stop the vessel or even to put the ship in a dry dock. In the second case, instead, collecting labeled samples is an easier task that can be performed by less experienced operators since the raw information about the decay is requested and it can be retrieved without impacting the ship activities. Moreover, as the authors will show in

this paper, this task requires, in general, less labeled data or even just few labeled samples thanks to the use of state-of-the-art DA techniques.

Many examples of DA techniques applied to different CBM problems can be found in literature [44]. Among other, Support Vector Machines [45], Hidden Markov Models [46] and Kalman filter [47] are the most frequently used. Examples of DA approaches applied to the marine industry can be found in [48], where a standard Neural Network approach is used to improve monitoring of Gas Turbines, while Kernel based methods are applied in [49], and [38]. In [50] and [51] image processing techniques are adopted for hull condition assessment. In [52] the engine and propeller state is predicted adopting an Artificial Neural Network. A complete overview can be found in [53].

The novelty of our approach relies on the concurrent modeling of the four main components decay status of an NPS: Propeller, Hull, Gas Turbine and Gas Turbine Compressor. In fact, many works present in literature model just one or maximum two of these components at the same time [48, 38]. Moreover, in this work, we propose a new DDM which require a limited number of labeled samples to achieve satisfying performance. In fact, as explained before, sensor data are cheap and easy to collect, while label them with the actual state of component decay can be quite expensive and, in some cases, unfeasible. For this reason, firstly the authors performed a traditional regression analysis where the target is to estimate the actual decay state of the components described with an efficiency coefficient. The analysis has been carried out comparing different state-of-the-art methodologies such as Kernel Methods [54], Neural Network [55], Gaussian Processes [56], Similarity Based Method [57], and Ensemble Methods [58]. Then, given the massive amount of historical data required to solve the task, the authors abstracted the problem by trying to predict if the efficiency coefficient is above or below a certain threshold defined by the accepted loss in efficiency of the NPS components. A by-product of this approach is the fact that this kind of information is easier to collect with respect to the exact efficiency coefficient. The results of this abstraction lead to a series of binary classification problems, which the authors tackle with the corresponding state-of-the-art methodologies, previously adapted to the regression framework. The results show that this step remarkably reduces the number of samples required to find an accurate model which is still, unfortunately, quite large. For this reason, this second problem has been tackled with another state-of-the-art approach which, in principle, does not need any labeled sample

since it searches for novel behavior in the data through a novelty detection algorithms [59, 60]. Results show that with just a few labeled samples it is possible to fine tune this last methodology in order to achieve satisfying performances. Data coming from a Frigate characterized by a COmbined Diesel ELectric And Gas (CODLAG) propulsion plant have been exploited to show the effectiveness of the proposed approaches and to benchmark them in a realistic maritime application. These data will be available through the UCI website, but they can already be downloaded at www.cbm.smartlab.ws. Confidentiality constraints with the Navy require us to use a real-data validated complex numerical simulator of a naval vessel [38]. Nevertheless, even though the data acquired do not belong to any active unit in service, the simulator which was used to generate them, has been developed and tuned thanks to several years experience on numerical modeling of real propulsion plants [61, 62].

The paper is organized as follows. Section 2 reports a general description of the vessel, the numerical model, and the degradation phenomena. Section 3 presents a description of the dataset extracted from the numerical simulator and published through UCI. Section 4 reports the proposed DDMs. Results of the DDMs tested on our data are reported in Section 5 with conclusions in Section 6.

2. Naval Propulsion System

2.1. Vessel Description

In this work authors focus on a Frigate, characterized by a CODLAG NPS, as depicted in Figure 1. In Tables 1 and 2, the main features of the vessel are reported. The gas turbine (GT) mechanically drives the two controllable pitch propellers (CPP) through a cross-connected gearbox. Besides, each shaft has its electric propulsion motor (EPM) mounted on the two shaft-lines. Two clutches between the gearbox (GB) and the two EPM and another clutch between the GT and the GB assure the possibility of using two different type of prime movers, i.e. EPM and GT. Finally, the electric power is provided by four diesel generators (DG). In this work, only the GT operating conditions have been taken into account. The GB arrangement, depicted in Figure 1, allows the vessel to operate under different propulsive configurations to achieve the requirements of the vessel’s mission profile. The vessel is characterized by the following mission profiles: Anti-Submarine Warfare

Table 1: Propulsion System

Type	CODLAG
Gas Turbine (GE LM2500+G4) Power	1 x 32 MW
Electric Motor (Jeumont Electric) Power	2 x 2.5 MW
Diesel Generator (Isotta Fraschini VL 1716)	4 x 2.15 MW
Bow Thrusters	1 x 1 MW
Propeller type	CPP

Table 2: Vessel Features

Vessel Features	Value	Unit
Length O.A.	144.6	[m]
Beam O.A.	19.7	[m]
Draught	8.7	[m]
Displacement (full load)	6700	[ton]
Speed	29+	[knots]
Range	6700 at 15 knots	[nm]

(ASW), General-Purpose (GP) and Anti-Aircraft Warfare (AAW). In particular, for the ASW profile, the EPMS are prime movers while the GT is disconnected through the clutches. Under the GP mission profile, the GT is the prime mover while the EPMS are working as shaft generators. Finally, for the AAW mission profile both the GT and the EPM are the prime movers.

2.2. Model Description

In this work, the authors consider an NPS numerical model developed in the Matlab[®] Simulink[®] software environment within many years of research [63, 64]. The model is composed of several modules each one representing a single propulsion component (e.g. hull, main engine, propeller, rudder, gearbox, control system, etc.). For the purpose of this study this model has been updated in order to take into account the performance decay of the following components:

1. Gas Turbine (GT);
2. Gas Turbine Compressor (GTC);
3. Hull (HLL);
4. Propeller (PRP).

In particular, the GT is characterized by a two shafts arrangement with a gas generator, consisting of a compressor and a high-pressure turbine for the exhaust gas production, driving the low-pressure power turbine as reported in Figure 2. Detailed information of the GT performance can be found in [65]

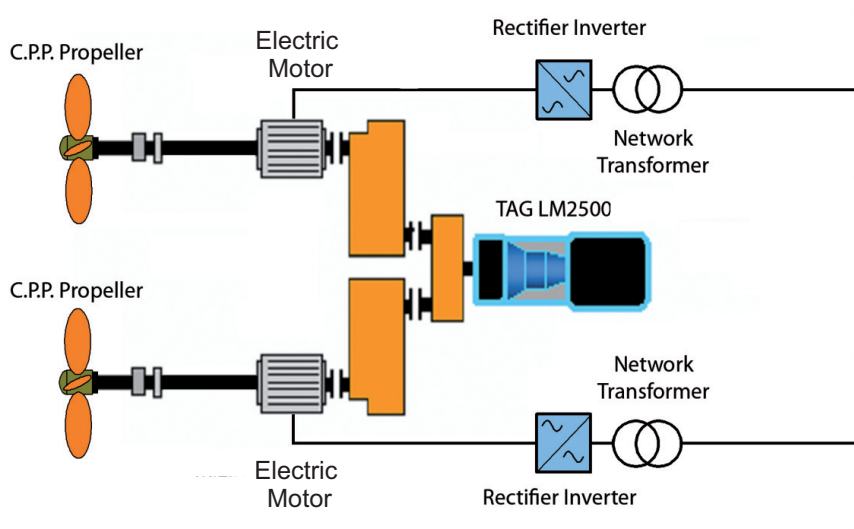


Figure 1: CODLAG propulsion system

and [62]. The GTC behavior is modeled adopting steady-state performance maps [66] representing rotational speed and isentropic efficiency as functions of the pressure ratio and the airflow rate. The HLL behavior has been modeled through the total resistance in agreement with the towing tank tests, while the PRP thrust and torque behavior has been modeled by means of the open water characteristics provided by the manufacturer.

The dynamic performance of the NPS can be described [61] solving simultaneously the following three non-linear differential equations in the time

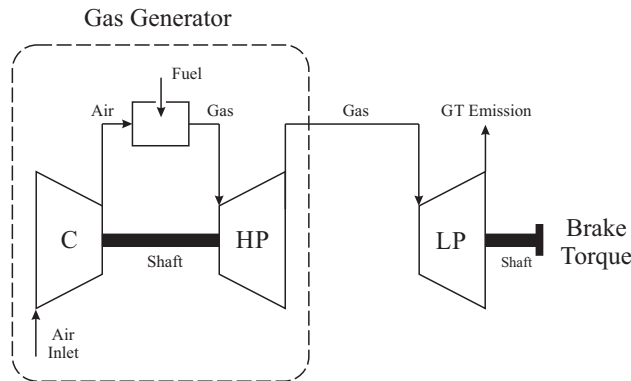


Figure 2: Gas Turbine Scheme

domain:

$$\begin{cases} M \cdot \frac{dv(t)}{dt} = p \cdot T_p(t) - R_t(t) \\ 2\pi \cdot J_p \cdot \frac{dn(t)}{dt} = i \cdot e \cdot Q_e(t) - Q_p(t) \\ \frac{dm_f(t)}{dt} = f\left(n, \frac{dn(t)}{dt}\right) \end{cases}, \quad (1)$$

where T_p is the PRP thrust, p is the number of propellers, R_t is the total ship resistance, M is the ship mass, Q_e is the GT torque, i is the gear ratio, e is the number of engines per shaft, Q_p is the PRP torque, J_p is the polar moment of the rotating masses, m_f is the fuel flow and n is the rotational speed. The PRP characteristics at different pitch and rotational speed have been modeled through the non dimensional thrust K_t and torque K_q coefficients according to the following equations:

$$Q_p = \rho_w K_q n^2 D^5, \quad (2)$$

$$T_p = \rho_w K_t n^2 D^4, \quad (3)$$

where D is the PRP diameter, n is the propeller rotational speed, and ρ_w is the sea water density. For the GT model, a General Electric LM 2500, the procedure described in [65] has been applied, and in order to validate the model, steady-state and transient results obtained by the simulator have been compared with experimental data provided by the GT manufacturer. The main differential equations to be solved in the time domain refer to both dynamic and thermodynamic engine parameters. In particular the working fluid processes are governed by the continuity 4 and energy equation 5:

$$\frac{d(\rho \cdot V) \cdot \rho}{dt} = M_i - M_o, \quad (4)$$

$$\frac{dm_f(t)}{dt} = f\left(n, \frac{dn(t)}{dt}\right). \quad (5)$$

where M_i and M_o are the inlet and outlet mass flow rates, V is the GT component volume, h and u are enthalpy and internal energy, while P and Q' represent power and heat flow respectively. The dynamics of the rotating shafts is determined by the dynamic momentum equation:

$$\frac{d\omega}{dt} = \frac{1}{J \cdot \omega} (P_m - P_{br}) \quad (6)$$

being P_m and P_{br} the motor and brake powers and ω the angular velocity. The NPS is managed by a typical combinator law, the control system has

been designed to provide a linear relationship between lever position and ship speed. The ship propulsion control system dynamics has been represented by a proportional integral controller:

$$TIC = K_p \cdot (n_{sp} - n) + K_i \cdot \int (n_{sp} - n) d\tau \quad (7)$$

where: TIC is the control signal (i.e. % of fuel flow) for the propulsion engine, n and n_{sp} are the actual and the set point propeller shaft speed respectively, while K_p and K_i represent the PI control loop constants. This controller is only an approximation of the complex controller implemented on-board, nevertheless can be considered satisfactory for the most important ship maneuvers as reported in [67].

2.3. Degradation Model

The model described in Section 2.2 has been updated to take into account the performance decay of the following components: GT, GTC, HLL, and PRP. As reported in [68] and [38], the time domain performance decay has been modeled utilizing suitable coefficients embedded in the GTC low and high power modules of the model. This, allows quantifying the variation of performance of each component and consequently of the entire propulsion system.

GTC and GT Degradation Model

As reported in [69] and [70], the fouling of the GTC is indicated as one of the leading causes of performance loss of the machine. Fouling is caused by air impurities, which, together with the exhaust gasses and oil vapors, produces a layer that covers the blades. Fouling increases both the specific fuel consumption of the GT and the temperature of the exhaust gas. The effect of the fouling is simulated by reducing the numerical values of the airflow rate M_c and of the isentropic efficiency η_c with a reduction factor kM_c and K_{η_c} , empirically derived in agreement with [69] and [71]. In Figure 3 the gas flow rate reduction factor M_c is reported. The GT performance decay in time is the effect of the blades' erosion and fouling as reported in [71] and [68]. For this reason, in agreement with the modelisation of the GTC, a reduction factor kM_T has been applied to the GT flow rate as reported in Figure 4.

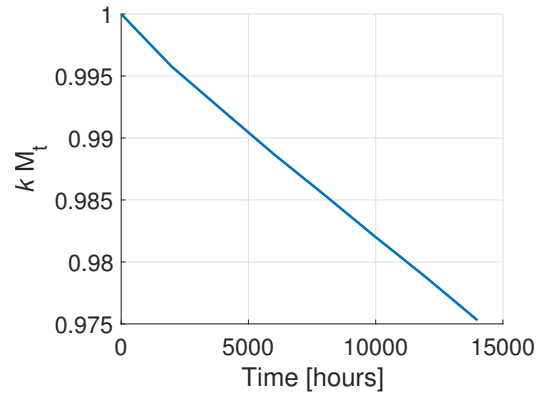


Figure 3: Gas flow rate reduction factor kM_c over service hours

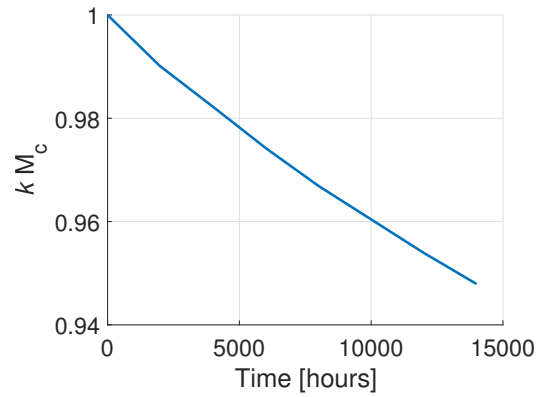


Figure 4: High-pressure and low-pressure turbines gas flow rate reduction factor kM_T over service hours

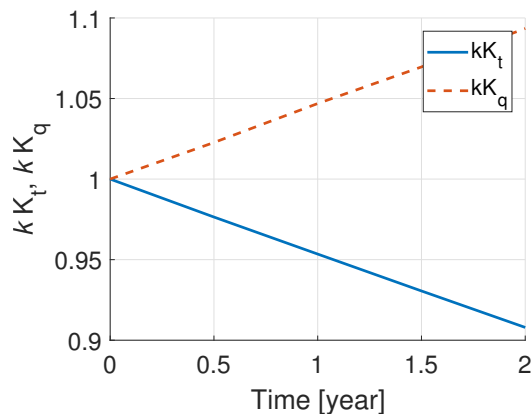


Figure 5: Correction factors for thrust reduction, kK_t , and torque increase, kK_q ,

PRP Degradation Model

The primary cause of marine PRP performance degradation is the increase in the roughness of the blade surface [72]. The performance degradation is due to the accretions of marine organisms, to the metal alloy erosion and corrosion, or to any other combination of these elements. The PRP decay status has been modeled by increasing the torque coefficient (K_q) and by reducing the thrust coefficient (K_t). The correction factors used for thrust reduction, kK_t , and torque increase, kK_q , have been derived from [73, 74] and reported in Figure 5.

HLL Degradation Model

The main factors that affect hull performance are the shape, the coating used and the extent of fouling [75]. The resistance increase due to fouling has been modeled utilizing a correction factor kH based on the information reported in [76]. Using the towing tank data provided by the ship owner the resistance has been evaluated after one and two years of vessel operations in agreement with the data available in the literature [68]. In Figure 6 the behavior of the hull resistance referred to the trial condition (R_t/R_{tref}) over the time and speed is reported.

3. From Data to Condition-Based Maintenance

3.1. Dataset Creation

In this work, authors will use a real-data validated complex numerical simulator of a Navy frigate described in Section 2 in order to build a realistic

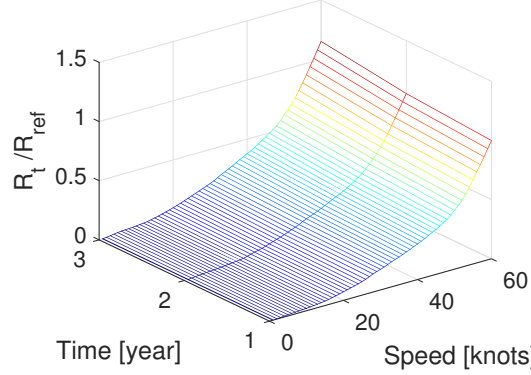


Figure 6: Hull resistance increase behavior over time and speed

set of data for designing and test purposes of DDMs. This dataset will be released for the use to the research community on the widespread well-known dataset repository of the University of California in Irvine (UCI) [77]. Currently, it can be downloaded from www.cbm.smartlab.ws.

As described in Section 2, the NPS model has six input parameters:

1. Speed: this parameter is controlled by the control lever. The latter can only assume a finite number of positions lp_i with $i \in \{0, \dots, 9\}$, which in turn correspond to a finite set of possible configurations for fuel flow and blade position. Each set point is designed to reach a desired speed v_i with $i \in \{0, \dots, 9\}$:

$$v_i = 3 * lp_i [\text{Knots}], \quad \forall i \in \{0, \dots, 9\}. \quad (8)$$

Note that, if the transients is not taken into account, lp_i and v_i are deterministically related by a linear law. In the presented analysis the transients between different speeds have been not considered.

2. As reported in Section 2.3, the PRP thrust and torque decay limit over two years of operations are:

$$kK_t \in [0.9, 1.0], \quad kK_q \in [1.0, 1.1] \quad (9)$$

kK_t and kK_q are respectively the components which define the decay of the torque and the thrust provided by the propeller in time. They are linearly correlated, since as the first decay of a certain quantity, the latter decay of the same quantity ($1 - kK_t = kK_q - 1$). For this reason

only kK_t will be analyzed, considering the linear dependency between the two variables.

3. The HLL decay has been modeled according to the available literature [76] as described in Section 2.3. The decay limit over two years of operations are:

$$kH \in [1, 1.2] \quad (10)$$

4. GTC decay:

$$kM_c \in [0.95, 1.0] \quad (11)$$

5. GT decay:

$$kM_t \in [0.975, 1.0] \quad (12)$$

It is worth underlying that the performance decay functions reported in Section 2.3 are empirically derived and are only a function of time. The real degradation behavior of the physical asset should be described through specific functions able to describe the time dependency and the real operational profile. In order to overcome this issue, the authors considered each possible combination of GTC, GT, HLL, and PRP decays. Under this assumption, the authors sampled the range of decays with a uniform grid characterized by a degree of precision sufficient to have a good granularity of representation. Given the above premises, the evolution of the system can be exhaustively explored by simulating all its possible states. The space of possible states is described via the following tuple:

$$(lp, kK_t, kK_q, kH, kM_c, kM_t)_i, \quad i \in \{1, \dots, 589223\} \quad (13)$$

since:

$$lp \in \mathcal{S}^{lp} = \{0, 3, 6, \dots, 27\}, \quad (14)$$

$$kK_t \in \mathcal{S}^{kK_t} = \{0.9, .9 + 0.1/14, .09 + 0.2/14, \dots, 1.0\}, \quad (15)$$

$$kK_q = 2 - kK_t, \quad (16)$$

$$kH \in \mathcal{S}^{kH} = \{1.0, 1.0 + 0.2/14, 1.0 + 0.4/14, \dots, 1.2\}, \quad (17)$$

$$kM_c \in \mathcal{S}^{kM_c} = \{0.95, 0.95 + 0.05/14, 0.95 + 0.1/14, \dots, 1.0\}, \quad (18)$$

$$kM_t \in \mathcal{S}^{kM_t} = \{0.975, 0.975 + 0.025/14, 0.975 + 0.05/14, \dots, 1.0\}. \quad (19)$$

Table 3: Measured values available from the continuous monitoring system

#	Variable name	Unit
1	Lever (lp)	[]
2	Vessel Speed	[knots]
3	GT shaft torque (GTT)	[kN m]
4	GT Speed (GT rpm)	[rpm]
5	Controllable Pitch Propeller Thrust stbd (CPP T stbd)	[N]
6	Controllable Pitch Propeller Thrust port (CPP T port)	[N]
7	Shaft Torque port (Q port)	[kN m]
8	Shaft rpm port (rpm port)	[rpm]
9	Shaft Torque stbd (Q stdb)	[kN m]
10	Shaft rpm stbd (rpm stbd)	[rpm]
11	HP Turbine exit temperature (T48)	[°C]
12	Generator of Gas speed (GG rpm)	[rpm]
13	Fuel flow (mf)	[kg/s]
14	ABB TIC control signal (ABB Tic)	[]
15	GT Compressor outlet air pressure (P2)	[bar]
16	CGT Compressor outlet air temperature (T2)	[°C]
17	External Pressure (Pext)	[bar]
18	HP Turbine exit pressure (P48)	[bar]
19	TCS TIC control signal (TCS Tic)	[]
20	Thrust coefficient stbd (Kt stbd)	[]
21	Propeller rps stbd (rps prop stbd)	[rps]
22	Thrust coefficient port (Kt port)	[]
23	Propeller rps port (rps prop port)	[rps]
24	Propeller Torque port (Q prop port)	[kN m]
25	Propeller Torque stbd (Q prop stbd)	[kN m]

Once these quantities are fixed, the numerical model is run until the steady state is reached. Then, the model is able to provide all the quantities reported in Table 3. These subsets of models outputs are the same quantities that the automation system installed on-board can acquire and store.

The simulator was run on a server equipped with four Intel[®] Xeon[®] CPU E5-4620 2.2 GHz, 128 GB of RAM, 120 GB SSD disk, and Matlab[®] R2016a.

3.2. Condition-Based Maintenance

This study aims at estimating the four decay variables described in the previous section, adopting different DA techniques. This section reports how the data generated can be used to create effective predictive DDMs for the CBM of an NPS.

The data described in Section 3.1 contain two sets of information: one regarding the quantities that the automation system installed on-board can

acquire and store and the other one regarding the associated state of decay (efficiency coefficient) of the different NPS components (GT, GTC, HLL, and PRP).

This problem can be straightforwardly mapped into a classical multi-output regression problem [7, 38] where the aim is to predict the actual decay coefficient based on the automation data coming from the sensors installed on-board. Unfortunately, this approach cannot be adopted in a real operational scenario. While the sensors data coming from the automation system are easy to collect, the information regarding the associated state of decay is not so easy to retrieve. In fact, to circumvent this challenge for the purpose of demonstrating DDMS authors exploited a numerical model for gathering all the information and build the dataset presented in Section 3.1. In practice, instead, retrieving the state of decay of the different NPS components requires the intervention of an experienced operator and, in some cases, to stop the vessel or even to put the ship in a dry dock. Moreover, data-driven regression models require a huge amount of historical data and therefore a long acquisition time.

Based on these considerations, authors decided to build simplified DDMS able to detect if the component state of decay is above or below a certain threshold. These thresholds represent the accepted loss in efficiency of the NPS components and the consequently sustainable costs of keeping a less performing vessel operative. This approach represents an abstraction of the problem which allows a more practical collection of the state of decay of the component. In fact, instead of requiring the precise state of decay, this approach only requires detecting if the decay state of the components is acceptable or not. Consequently, the collection of these data can be performed by less experienced operators since raw information about the decay is requested and can be retrieved without impacting the ship activities. This new problem can be straightforwardly mapped into a classical multi-output binary classification problem [7] where the aim is to predict if the decay state of an NPS component is acceptable or not based on the automation data coming from the sensors installed on-board.

The data described in Section 3.1 can be easily exploited to tackle this new problem as well. In fact, by thresholding kK_t , kH , kM_c , and kM_t the corresponding binary valued state of decay of the NPS components are obtained. In other words, if the efficiency coefficients adopted during regression analysis, are above or below a defined threshold, based on the accepted loss in efficiency of the NPS components, they will be tagged as “decayed” or

“not decayed”. Thresholds were fixed according to the least affordable value of decay of the single component. Defining these thresholds is not a trivial task. Authors approach is to define the maximum level of inefficiency that the operator or the shipowner is willing to tolerate before taking action and reestablish the efficiency of the system. The authors considered two years as a typical time frame between two important dry dock maintenance for HLL and PRP.

The HLL and PRP thresholds have been defined considering one year of operation. The proposed limits are just an example of the possible selection that is possible to setup to implement a CBM framework

$$kK_t \begin{cases} [0.9 - 0.95] & \text{decayed} \\ [0.95 - 1] & \text{not decayed} \end{cases} \quad (20)$$

$$kH \begin{cases} (1.1 - 1.2] & \text{decayed} \\ [1 - 1.1] & \text{not decayed} \end{cases} \quad (21)$$

$$(22)$$

As for GT and GTC, an effective time service of 2000 hours per year is considered as a reasonable operating time for these vessel types. In agreement with these observations the authors defined the following thresholds based on the knowledge of the time domain decay functions:

$$kM_c \begin{cases} [0.95 - 0.98] & \text{decayed} \\ [0.98 - 1] & \text{not decayed} \end{cases} \quad (23)$$

$$kM_t \begin{cases} [0.975 - 0.99] & \text{decayed} \\ [0.99 - 1] & \text{not decayed} \end{cases} \quad (24)$$

Results will show that estimating if the decay state is acceptable or not, instead of estimating its specific state, remarkably reduces the number of samples required to find accurate DDMs. However, this quantity is still too large with respect to what can be collected in a real operational scenario.

In order to solve this issue, authors tried to look at the same problem from another perspective. Specifically, it is reasonable to state that, for the vast majority of the time, NPS components of the ships operate in an acceptable state of decay. Consequently, most of the sensor data collected by the automation system represent ordinary operating conditions corresponding to a reasonable decay state of the NPS components (GT, GTC, HLL,

and PRP). Just very few times during the ship lifetime it happens that it has to operate with over-decayed components. If, for some reasons, one or more NPS components decay too fast, the corresponding automation data measurements will deviate from their expected behavior. This new problem can be straightforwardly mapped into a classical outlier (novelty) detection problem [78, 54, 60] where the aim is to detect unexpected behavior in the sensor data collected by the automation system which may correspond to an over-decayed state of an NPS components. This method does not require to know either the actual state of decay of the components, as in the regression framework, or the less detailed information about “decayed” or “not decayed”, as in the binary classification framework. In this case, the method just needs the sensor data collected by the automation system (see Table 3) without any supervision or feedback from the operator. These kinds of DDMs try to build a model of the “usual” operational profile of the ship and automatically detect if the sensor data collected by the automation system are deviating too much from the established behavior.

As for the binary classification framework, the data described in Section 3.1 can be easily exploited to tackle this problem as well. In fact, it is just necessary to keep the data corresponding to an acceptable decay state with respect to kK_t , kH , kM_c , and kM_t and in accordance with Eqns. (20), (21), (23), and (24). Finally, for testing and tuning the DDMs, it is possible to use just a few samples of the dataset corresponding to an unacceptable decay state. Note that these are the only samples which are costly to retrieve since they are the only ones that require the intervention of expert operators. Results will show that with just very few samples (≈ 10) of decayed state of the vessel, it is possible to obtain effective DDMs for CBM of NPS.

Since in authors’ analysis four different components of an NPS are taken into account (GT, GTC, HLL, and PRP), they decided to perform an incremental analysis by breaking down the problem into simpler ones. In particular, first, only one NPS decayed component at the time is considered, then authors consider the possible combination of two NPS decayed components and so on until finally all the four NPS components are contemporarily considered. Consequently, since four NPS components are considered, $4 = \binom{4}{1}$ problems have to be solved when just one NPS decayed component at the time is considered, then $6 = \binom{4}{2}$ problems have to be solved when two NPS decayed components at the time are considered and so on. Finally when all the four NPS components are contemporarily considered $1 = \binom{4}{4}$ problem

Table 4: The 15 sub-problems corresponding to considering different decayed component at the time

Prob. #	Decayed Component Name	Non Decayed Component Name	Decayed Component	Non Decayed Component
1	GT	GTC,HLL,PRP	$kM_t \in \mathcal{S}^{kM_t}$	$kM_c=1, kH=1, kK_t=1$
2	GTC	GT,HLL,PRP	$kM_c \in \mathcal{S}^{kM_c}$	$kM_t=1, kH=1, kK_t=1$
3	HLL	GT,GTC,PRP	$kH \in \mathcal{S}^{kH}$	$kM_c=1, kM_t=1, kK_t=1$
4	PRP	GT,GTC,HLL	$kK_t \in \mathcal{S}^{kK_t}$	$kM_c=1, kM_t=1, kH=1$
5	GT,GTC	HLL,PRP	$kM_t \in \mathcal{S}^{kM_t}, kM_c \in \mathcal{S}^{kM_c}$	$kH=1, kK_t=1$
6	GT,HLL	GTC,PRP	$kM_t \in \mathcal{S}^{kM_t}, kH \in \mathcal{S}^{kH}$	$kM_c=1, kK_t=1$
7	GT,PRP	GTC,HLL	$kM_t \in \mathcal{S}^{kM_t}, kK_t \in \mathcal{S}^{kK_t}$	$kM_c=1, kH=1$
8	GTC,HLL	GT,PRP	$kM_c \in \mathcal{S}^{kM_c}, kH \in \mathcal{S}^{kH}$	$kM_t=1, kK_t=1$
9	GTC,PRP	GT,HLL	$kM_c \in \mathcal{S}^{kM_c}, kK_t \in \mathcal{S}^{kK_t}$	$kM_t=1, kH=1$
10	HLL,PRP	GT,GTC	$kH \in \mathcal{S}^{kH}, kK_t \in \mathcal{S}^{kK_t}$	$kM_t=1, kM_c=1$
11	GT,GTC,HLL	PRP	$kM_t \in \mathcal{S}^{kM_t}, kM_c \in \mathcal{S}^{kM_c}, kH \in \mathcal{S}^{kH}$	$kK_t=1$
12	GT,GTC,PRP	HLL	$kM_t \in \mathcal{S}^{kM_t}, kM_c \in \mathcal{S}^{kM_c}, kK_t \in \mathcal{S}^{kK_t}$	$kH=1$
13	GT,HLL,PRP	GTC	$kM_t \in \mathcal{S}^{kM_t}, kH \in \mathcal{S}^{kH}, kK_t \in \mathcal{S}^{kK_t}$	$kM_c=1$
14	GTC,HLL,PRP	GT	$kM_c \in \mathcal{S}^{kM_c}, kH \in \mathcal{S}^{kH}, kK_t \in \mathcal{S}^{kK_t}$	$kM_t=1$
15	GT,GTC,HLL,PRP		$kM_t \in \mathcal{S}^{kM_t}, kM_c \in \mathcal{S}^{kM_c}, kH \in \mathcal{S}^{kH}, kK_t \in \mathcal{S}^{kK_t}$	

has to be solved, for a total of $15 = \sum_{i=1}^4 \binom{4}{i}$ problems. Therefore, from the dataset described in Section 3.1, 15 sub-datasets corresponding to the cases mentioned above have been extracted. For the sake of clarity in Table 4 the 15 problems with the corresponding decay values are reported, not that in all the problems $lp \in \mathcal{S}^{lp}$.

As a final remark, authors would like to recall that each navy frigate is characterized by different mission profiles (AAW, ASW, and GP) as described in Section 2.1. Each mission profile is characterized by a particular use of the ship in terms of speed. CBM DDMs for NPS do not need to estimate the state of decay of the NPS components for all the possible mission profiles. In fact, the vessel operates at a cruise speed (which is approximately ≈ 15 knots) while the time spent by the vessel at different speeds is negligible. For this reason, we conduct the same analysis described in the previous paragraph by setting $lp = 15$. Results will show that this simplification will further reduce the amount of historical data needed to build effective CBM DDMs for NPS.

4. Machine Learning Techniques

In this section, the authors will present the Machine Learning (ML) techniques adopted in order to build the CBM DDMs for NPS described in Section 2, based on the data outlined in Section 3.

Let authors consider an input space $\mathcal{X} \subseteq \mathbb{R}^d$ and an output space \mathcal{Y} . Note that, for what concerns this paper, \mathcal{X} takes into account the different sensors measurements, also called features, reported in Table 3, while the output space \mathcal{Y} depends on the particular problem identified in Section 3.2. ML techniques aim at estimating the unknown rule $\mu : \mathcal{X} \rightarrow \mathcal{Y}$ which associates an element $y \in \mathcal{Y}$ to an element $\mathbf{x} \in \mathcal{X}$. Note that, in general, μ can be non-deterministic. An ML technique estimates μ through a learning algorithm $\mathcal{A}_{\mathcal{H}} : \mathcal{D}_n \times \mathcal{F} \rightarrow h$, characterized by its set of hyperparameters \mathcal{H} , which maps a series of examples of the input/output relation contained in a datasets of n samples $\mathcal{D}_n : \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ into a function $f : \mathcal{X} \rightarrow \mathcal{Y}$ chosen in a set of possible ones \mathcal{F} .

When both \mathbf{x}_i and y_i with $i \in \{1, \dots, n\}$ are available, the problems is named supervised and consequently supervised ML technique are adopted [7]. Classification and regression are the most popular examples of supervised ML problems [54]. In classification, the output space is composed of a finite set of c possibilities $\mathcal{Y} \in \{C_1, \dots, C_c\}$. Binary classification is a particular example of classification problem where $\mathcal{Y} \in \{\pm 1\}$. In regression, instead, $\mathcal{Y} \in \mathbb{R}$.

When just \mathbf{x}_i with $i \in \{1, \dots, n\}$ are available, which means that we do not know explicitly the associated element of the output space y_i with $i \in \{1, \dots, n\}$, it has to be assumed that “similar” \mathbf{x}_i are associated with “similar” y_i where the concept of similarity is something that needs to be defined based on μ . In this last case, the ML problems are called unsupervised, and consequently, unsupervised ML techniques need to be adopted [79]. Anomaly (novelty, outlier) detection is a common example of unsupervised learning problem where the unknown $y \in \mathcal{Y}$ can assume only two possible values: -1 for “non-anomaly” and $+1$ for “anomaly” [54].

The error that f commits in approximating μ is measured with reference to a loss function $\ell : \mathcal{X} \times \mathcal{Y} \times \mathcal{F} \rightarrow [0, \infty)$. Obviously, the error that f commits over \mathcal{D}_n , is optimistically biased since \mathcal{D}_n has been used, together with \mathcal{F} , for building f itself. For this reason, another set of fresh data, composed of m samples and called test set $\mathcal{T}_m = \{(\mathbf{x}_1^t, y_1^t), \dots, (\mathbf{x}_m^t, y_m^t)\}$, needs to be exploited. Note that, $\mathbf{x}_i^t \in \mathcal{X}$ and $y_i^t \in \mathcal{Y}$ with $i \in \{1, \dots, m\}$, and the association of y_i^t to \mathbf{x}_i^t is again made based on μ . Moreover, both for supervised and unsupervised problems \mathcal{T}_m must contain both $\mathbf{x}_i^t \in \mathcal{X}$ and $y_i^t \in \mathcal{Y}$ with $i \in \{1, \dots, m\}$ to estimate the error of f , while, for unsupervised learning problems, y_i with $i \in \{1, \dots, n\}$ in \mathcal{D}_n is unknown.

4.1. Measuring the Error

In this work, many state-of-the-art ML techniques will be tested and their performances will be compared to understand what is the most suited solution for building CBM DDMs for NPS.

In order to perform this analysis, authors have to define different measures of error, also called indexes of performance, able to well characterize the quality of the different CBM DDMs for NPS. Once f has been chosen based on \mathcal{D}_n it is possible to use the fresh set of data \mathcal{T}_m in order to compute its error based on different losses. The choice of the loss strongly depends on the problem under examination [80].

In the regression framework, there are two losses that are mainly used for estimating the quality of a regressor: the absolute loss $\ell_1(f(\mathbf{x}), y) = |f(\mathbf{x}) - y|$ and the squared loss $\ell_2(f(\mathbf{x}), y) = (f(\mathbf{x}) - y)^2$. Based on these losses it is possible to define different indexes of performance, which differently weight the distance between y_i^t and $f(\mathbf{x}_i^t)$ [81]:

- the Mean Squared Error (MSE) is computed by taking the mean square loss value of f over \mathcal{T}_m : $\text{MSE} = 1/m \sum_{i=1}^m \ell_2(f(\mathbf{x}_i^t), y_i^t)$;
- the Normalized Mean Square Error (NMSE) is similar to the MSE but a normalization term composed of the mean value of the difference between the true values and their mean is applied: $\text{NMSE} = \sum_{i=1}^m \ell_2(f(\mathbf{x}_i^t), y_i^t) / \Delta$;
- the Relative Error Percentage (REP) is similar to NMSE but the normalization term is composed of the sum of the squared true values. Then the result is square rooted and reported in percentage: $\text{REP} = 100 \sqrt{\sum_{i=1}^m \ell_2(f(\mathbf{x}_i^t), y_i^t) / \sum_{i=1}^m (y_i^t)^2}$;
- the Mean Absolute Error (MAE) is similar to MSE but the absolute loss is used instead: $\text{MAE} = 1/m \sum_{i=1}^m \ell_1(f(\mathbf{x}_i^t), y_i^t)$;
- the Mean Absolute Percentage Error (MAPE) can be described as the MAE expressed in percentage: $\text{MAPE} = 100/m \sum_{i=1}^m \ell_1(f(\mathbf{x}_i^t), y_i^t) / y_i^t$;
- the Pearson Product-Moment Correlation Coefficient (PPMCC) is a measure of the linear dependency between $f(\mathbf{x}_i^t)$ and y_i^t with $i \in \{1, \dots, m\}$: $\text{PPMCC} = \sum_{i=1}^m (y_i^t - \bar{y})(f(\mathbf{x}_i^t) - \hat{y}) / \sqrt{\sum_{i=1}^m (y_i^t - \bar{y})^2} \sqrt{\sum_{i=1}^m (f(\mathbf{x}_i^t) - \hat{y})^2}$,

where authors have exploited the following notation

$$\bar{y} = \frac{1}{m} \sum_{i=1}^m y_i^t, \quad \Delta = \frac{1}{m} \sum_{i=1}^m (y_i^t - \bar{y})^2, \quad \hat{y} = \frac{1}{m} \sum_{i=1}^m f(\mathbf{x}_i^t). \quad (25)$$

In the classification framework, the previously depicted indexes of performance cannot be adopted, since the concept of distance between the $y \in \mathcal{Y}$

cannot be defined. In this case, the most natural choice as loss function is the Hard loss one, which counts the number of misclassified samples $\ell_H(f(\mathbf{x}), y) = [f(\mathbf{x}) \neq y]$. Note that the Iverson bracket notation is exploited. In this work, only binary classification problems are investigated, then the Hard loss function can be expressed as $\ell_H(f(\mathbf{x}), y) = |1 - yf(\mathbf{x})|/2$. Moreover, this measure will be also used for the anomaly detection problems since, also in this case, a binary output is considered (non-anomaly or anomaly).

Based on the Hard loss it is possible to define different indexes of performance [82]:

- the Average Misclassifications Rate (AMR) is the mean number of misclassified samples: $\text{AMR} = \frac{1}{m} \sum_{i=1}^m \ell_H(f(\mathbf{x}_i^t), y_i^t)$;
- the Confusion Matrix, which measures four different quantities:
 - $\text{TN} = 100/m \sum_{i=1}^m [f(\mathbf{x}_i^t) = y_i^t \wedge y_i^t = -1]$ which is the percentage of true negative;
 - $\text{TP} = 100/m \sum_{i=1}^m [f(\mathbf{x}_i^t) = y_i^t \wedge y_i^t = +1]$ which is the percentage of true positive;
 - $\text{FN} = 100/m \sum_{i=1}^m [f(\mathbf{x}_i^t) \neq y_i^t \wedge y_i^t = -1]$ which is the percentage of false negative;
 - $\text{FP} = 100/m \sum_{i=1}^m [f(\mathbf{x}_i^t) \neq y_i^t \wedge y_i^t = +1]$ which is the percentage of false positive.

4.2. Machine Learning Techniques

In this section, authors will present the supervised and unsupervised learning algorithms exploited in this paper for building CBM DDMs for NPS. Moreover, authors will show how to tune their performances by tuning their hyperparameters during the so-called Model Selection MS phase [83, 84, 85]. Finally, authors will also check for possible spurious correlation in the data by performing the Feature Selection (FS) phase [86, 87, 88, 89, 90]. In fact, once f is built based on the different learning algorithm and has been confirmed to be a sufficiently accurate representation of μ , it can be interesting to investigate how the model f is affected by the different features that have been exploited to build f itself during the feature ranking procedure [86]. As authors will describe later, for some algorithms, the feature ranking procedure is a by-product of the learning process itself and allows to simply check the physical plausibility of f .

4.2.1. Supervised Learning Algorithms for Regression and Classification

Supervised ML techniques can be grouped into different families, according to the space of function \mathcal{F} from which the learning algorithm chooses the particular f , the approximation of μ , based on the available data \mathcal{D}_n . In fact, techniques belonging to the same family, share an affine \mathcal{F} . Among the several possible ML families, authors choose the state-of-the-art ones which are commonly adopted in real-world application, and, in each family, the best performing techniques are selected. In particular, Neural Networks (NNs), Kernel Methods (KMs), Ensemble Methods (EMs), Bayesian Methods (BMs), and Lazy Methods (LMs) are adopted.

NNs are ML techniques which combine together many simple models of a human brain neuron, called perceptrons [55], in order to build a complex network. The neurons are organized in stacked layers connected together by weights that are learned based on the available data via backpropagation [91]. The hyperparameters of an NN \mathcal{H}^{NN} are the number of layers h_1 and the number of neurons for each layer $h_{2,i}$ with $i \in \{1, \dots, h_1\}$. Note that it is assumed that NN with only one hidden layer has $h_1 = 1$. If the architecture of the NN consists of only one hidden layer, it is called shallow (SNN) [92, 93], while, if multiple layers are stacked together, the architecture is defined as deep (DNN) [94, 95, 96]. Extreme Learning Machines (ELMs) are a particular kind of SNN, where the weights of the first layer are randomly chosen while the ones of the output layers are computed according to the Regularized Least Squares (RLS) principle [97, 98, 99]. The hyperparameters of the ELM \mathcal{H}^{ELM} are the number of neurons of the hidden layer, h_1 , and the RLS regularization hyperparameter h_2 [100]. NNs can be adopted in both regression and classification problems.

KMs are a family of ML techniques which exploits the “Kernel trick” for distances in order to extend linear techniques to the solution of non-linear problems [101, 102]. In the case of regression and classification, KMs select f as the function which minimizes the tradeoff between the sum of the accuracy over the data, namely the empirical error, and the complexity of the solution, namely the regularization term [103, 101, 104]. The most known and effective KM techniques are: Kernelized Regularized Least Squares (KRLS), Support Vector Regression (SVR), and Support Vector Machines (SVMs). KRLS is both used in regression and binary classification problems. The hyperparameters of the KRLS $\mathcal{H}^{\text{KRLS}}$ are: the kernel, which is usually fixed and in this paper author chose the Gaussian Kernel for the reasons described

in [105, 106], its hyperparameter h_1 and the regularization hyperparameter h_2 . SVM, instead, is a classification method, which roots in the Statistical Learning Theory [7] and differs from the KRLS mainly because of its particular loss function which has been proved to be the best one for classification purposes [80]. The hyperparameters of the SVM are the same as the one of the KRLS. SVR, finally, is the counterpart of SVM for regression problems which introduces an ϵ insensitivity value for the errors [54]. The hyperparameters of the SVR \mathcal{H}^{SVR} are the same as the ones of SVM plus the insensitivity value h_3 .

EMs ML techniques relies on the simple fact that combining the output of several classifiers results in a much better performance than using any one of them alone [107, 58]. Random Forest (RF) [58] and Random Rotation Ensembles (RRF) [108], two popular state-of-the-art and widely adopted methods, combine many decision trees in order to obtain a effective predictors which have limited hyperparameter sensitivity and high numerical robustness [109, 110]. Both RF and RRF have hidden hyperparameters which we consider fixed in this work because of their limited effect [111]. RF and RRF can be adopted in both regression and classification problems.

BMs are ML techniques where, instead of choosing a particular $f \in \mathcal{F}$ a distribution for choosing $f \in \mathcal{F}$ is defined [112]. Gaussian Processes (GP) learning algorithm is a popular BM [56] which employs a collection of Gaussians in order to compute the posterior distribution of the $f(\mathbf{x})$. In fact, this algorithm defines the probability distribution of the output values as a sum of Gaussians whose variance is fixed according to the training data. The hyperparameter of the GP \mathcal{H}^{GP} is the parameter which governs the Gaussians width h_1 . GP can be adopted in both regression and classification problems.

LMs ML techniques are learning method in which the definition of f is delayed until $f(\mathbf{x})$ needs to be computed [57]. LMs approximate μ locally with respect to \mathbf{x} . K-Nearest Neighbors (KNN) is one of the most popular LM due to its implementation simplicity and effectiveness [113]. The hyperparameter of the KNN \mathcal{H}^{KNN} is the number of neighbors of \mathbf{x} to be considered h_1 . KNN can be adopted in both regression and classification problems.

4.2.2. *Unsupervised Learning Algorithms for Anomaly Detection*

Similarly to their supervised counterpart, also unsupervised ML methods can be divided into different families. Since this work deals only with anomaly detection problems, authors will recall the most known and effective techniques for solving these problems according to [60]. In particular

[60] shows that two anomaly detection methods based on SVM and KNN respectively, are the top choices in this context.

In particular One-Class SVM (OCSVM) is a boundary-based anomaly detection method, inspired by SVM, which enclose the inlier class in a minimum volume hypersphere by minimizing a Tikhonov regularization problem, similar to the one reported for SVR and SVM frameworks. Like traditional SVMs, OCSVM can also be extended to non-linearly transformed spaces using the “Kernel trick” for distances. The hyperparameters OCSVM $\mathcal{H}^{\text{OCSVM}}$ are the same as the ones of SVM.

The Global KNN (GKNN), inspired by the KNN, has been originally introduced as an unsupervised distance-based outlier detection method [114, 60]. The hyperparameter GKNN $\mathcal{H}^{\text{GKNN}}$ is the same as the one of KNN.

4.2.3. Model Selection

MS deals with the problem of tuning the hyperparameters of each learning algorithm [85]. Several methods exist for MS purpose: resampling methods, like the well-known k -Fold Cross Validation (KCV) [83] or the nonparametric Bootstrap (BTS) approach [115, 116] approaches, which represent the state-of-the-art MS approaches when targeting real-world applications. Resampling methods rely on a simple idea: the original dataset \mathcal{D}_n is resampled once or many (n_r) times, with or without replacement, to build two independent datasets called training, and validation sets, respectively \mathcal{L}_i^r and \mathcal{V}_v^r , with $r \in \{1, \dots, n_r\}$. Note that $\mathcal{L}_i^r \cap \mathcal{V}_v^r = \emptyset$, $\mathcal{L}_i^r \cup \mathcal{V}_v^r = \mathcal{D}_n$. Then, in order to select the best combination the hyperparameters \mathcal{H} in a set of possible ones $\mathfrak{H} = \{\mathcal{H}_1, \mathcal{H}_2, \dots\}$ for the algorithm $\mathcal{A}_{\mathcal{H}}$ or, in other words, to perform the MS phase, the following procedure has to be applied:

$$\mathcal{H}^* : \min_{\mathcal{H} \in \mathfrak{H}} \frac{1}{n_r} \sum_{r=1}^{n_r} \frac{1}{v} \sum_{(\mathbf{x}_i, y_i) \in \mathcal{V}_v^r} \ell(\mathcal{A}_{\mathcal{H}, \mathcal{L}_i^r}(\mathbf{x}_i), y_i), \quad (26)$$

where $\mathcal{A}_{\mathcal{H}, \mathcal{L}_i^r}$ is a model built with the algorithm \mathcal{A} with its set of hyperparameters \mathcal{H} and with the data \mathcal{L}_i^r . Since the data in \mathcal{L}_i^r are independent from the ones in \mathcal{V}_v^r , the idea is that \mathcal{H}^* should be the set of hyperparameters which allows to achieve a small error on a data set that is independent from the training set.

Note that, for the anomaly detection problem, the algorithms do not need any label in \mathcal{L}_i^r , consequently authors just need the labeled data for \mathcal{V}_v^r .

If $r = 1$, if l and v are aprioristically set such that $n = l + v$, and if the resample procedure is performed without replacement, the hold out method is obtained [85]. For implementing the complete k -fold cross validation, instead, it is needed to set $r \leq \binom{n}{k} \binom{n-\frac{n}{k}}{k}$, $l = (k-2)\frac{n}{k}$, $v = \frac{n}{k}$, and $t = \frac{n}{k}$ and the resampling must be done without replacement [83, 117, 85]. Finally, for implementing the bootstrap, $l = n$ and \mathcal{L}_i^r must be sampled with replacement from \mathcal{D}_n , while \mathcal{V}_v^r and \mathcal{T}_t^r are sampled without replacement from the sample of \mathcal{D}_n that have not been sampled in \mathcal{L}_i^r [115, 85]. Note that for the bootstrap procedure $r \leq \binom{2n-1}{n}$. In this paper the BTS is exploited because it represents the state-of-the-art approach [115, 85].

4.3. Feature Selection

Once the CBM NPS models are built and have been confirmed to be sufficiently accurate representation of the real decays of the components, it can be interesting to investigate how these models are affected by the different features used in the model identification phase (see Table 3).

In DA this procedure is called FS or Feature Ranking [86, 87, 88, 89, 90]. This process allows detecting if the importance of those features, that are known to be relevant from a physical perspective, is appropriately described by the different CBM NPS models. The failure of the statistical model to properly account for the relevant features might indicate poor quality in the measurements or spurious correlations. FS therefore represents an important step of model verification, since it should generate consistent results with the available knowledge of the physical system under exam.

In addition to its use for regression and classification purposes, the EMs can also be used to perform a very stable FS procedure. The procedure is a combination of EMs, together with the permutation test [118], in order to perform the selection and the ranking of the features. In details, for every tree, two quantities are computed: the first one is the error on the out-of-bag samples as they are used during prediction, while the second one is the error on the out-of-bag samples after a random permutation of the values of variable j . The errors are measured with the AMR in classification and the MSE in regression. These two values are then subtracted and the average of the result over all the trees in the ensemble is the raw importance score for variable j (mean decrease in accuracy). This procedure was adopted since it can be easily carried out during the main prediction process inexpensively.

5. Experimental results

In this section, the authors report the results obtained by the different methods applied to the CBM of the main components of an NPS, in regression, classification, and novelty detection frameworks, as described in Section 2, based on the data described in Section 3.

As described in Section 4, the first problem is a regression one where the actual value of the decay parameters needs to be estimated (REG-PROB). In order to approach this problem a complete dataset must be available. In this work, thanks to the model described in Section 2, this dataset has been collected, although it cannot be collected in a real-world scenario. In facts, vessels are rarely available in the harbor for maintenance, and the decay parameters cannot be easily acquired when the ship is operative. For this reason a more realistic approach was attempted by modeling the original problem into a classification one, taking only a discrete label of the components decay instead of their real value (CLASS-PROB). This label indicates if the decay components have overcome a particular pre-defined threshold or not. In this case, the dataset label can be obtained more easily, once it results evident for a vessel to need maintenance or not. In fact, this label is easier to collect but still not feasible in particular cases (e.g. when it requires to stop the vessel in a dry dock). For this reason another approach has been tested. In particular, authors attempted to solve the problem in an unsupervised fashion by modeling the problem as a novelty detection one in order to further reduce the necessity of labeled data (ANOMALY-PROB).

The different datasets considered in Section 3.2, corresponding to the 15 problems of Table 4 $\{P1, \dots, P15\}$, were divided into training and test set, respectively \mathcal{D}_n and \mathcal{T}_m , as reported in Section 4. Moreover, different dimensions of the training set $n \in \{10, 24, 55, 130, 307, 722, 1700, 4000\}$ were considered.

For each supervised learning task (regression and classification) and each ML technique, an MS procedure was performed, as described in Section 4.2.3. Here-below, the list of hyperparameters tested during the MS, with their respective intervals, is reported:

1. DNN: the set of hyperparameters is $\mathcal{H}^{DNN} = \{h_1, h_{2,1}, \dots, h_{2,h_1}\}$ and we chose it in $\mathfrak{H}^{DNN} = \{1, 3, 5, 7, 10\} \times \{10, 10^{1.2} \dots, 10^3\} \times \dots \times \{10, 10^{1.2} \dots, 10^3\}$;
2. SNN: the set of hyperparameters is $\mathcal{H}^{SNN} = \{h_1\}$ and we chose it in $\mathfrak{H}^{SNN} = \{1, 3, 5, 7, 10\}$;

3. ELM: the set of hyperparameters is $\mathcal{H}^{ELM} = \{h_1, h_2\}$ and we chose it in $\mathfrak{H}^{ELM} = \{10, 10^{1.2}, \dots, 10^3\} \times \{10^{-2}, 10^{-1.5}, \dots, 10^2\}$;
4. SVM: the set of hyperparameters is $\mathcal{H}^{SVM} = \{h_1, h_2\}$ and we chose it in $\mathfrak{H}^{SVM} = \{10^{-2}, 10^{-1.4}, \dots, 10^3\} \times \{10^{-2}, 10^{-1.4}, \dots, 10^3\}$;
5. SVR: the set of hyperparameters is $\mathcal{H}^{SVR} = \{h_1, h_2, h_3\}$ and we chose it in $\mathfrak{H}^{SVR} = \{10^{-6}, 10^{-5}, \dots, 10^3\} \times \{10^{-2}, 10^{-1.4}, \dots, 10^3\} \times \{10^{-2}, 10^{-1.4}, \dots, 10^3\}$;
6. KRLS: the set of hyperparameters is $\mathcal{H}^{KRLS} = \{h_1, h_2\}$ and we chose it in $\mathfrak{H}^{KRLS} = \{10^{-2}, 10^{-1.4}, \dots, 10^3\} \times \{10^{-2}, 10^{-1.4}, \dots, 10^3\}$;
7. KNN: the set of hyperparameters is $\mathcal{H}^{KNN} = \{h_1\}$ and we chose it in $\mathfrak{H}^{KNN} = \{1, 3, 7, 13, 27, 51\}$;
8. GP: the set of hyperparameters is $\mathcal{H}^{GP} = \{h_1\}$ and we chose it in $\mathfrak{H}^{GP} = \{10^0, 10^{0.3}, \dots, 10^3\}$;

When RF is exploited also the FS phase is performed in order to understand how the data-driven model combines the different features in order to predict the decay state of each component.

Similarly to the supervised learning task, in the unsupervised case different dimensions of the training set were considered $n \in \{1500, 2000, 3000, 4000\}$ and the MS procedure was performed as follows:

1. OCSVM: the set of hyperparameters is $\mathcal{H}^{OCSVM} = \{h_1, h_2\}$ and we chose it in $\mathfrak{H}^{OCSVM} = \{10^{-4}, 10^{-3.7}, \dots, 10^3\} \times \{10^{-4}, 10^{-3.8}, \dots, 10^{-1.0}\}$;
2. GKNN: the set of hyperparameters is $\mathcal{H}^{GKNN} = \{h_1\}$ and we chose it in $\mathfrak{H}^{GKNN} = \{1, 3, 7, 13, 27, 51\}$;

The \mathcal{V}_v^r cardinality was varied $v \in \{10, 20, 30\}$ with linear step, in order to test the possibility of building an efficient model with a small number of labeled samples.

The performances of each model are measured according to the metrics described in Section 4.1. Each experiment was performed 10 times in order to obtain statistical relevant result, and the t-student 95% confidence interval is reported when space in the table was available without compromising their readability.

For SNN and DNN the Python Keras library [119] has been exploited. For ELM, SVM, SVR, KRLS, KNN, and GKNN a custom R implementation has been developed. For RF the R package of [120] has been exploited. For RFE the implementation of [108] has been exploited. For GP the R package of [121] has been exploited. For OCSVM the R package of [122] has been exploited.

5.1. REG-PROB

In this section, the results on the REG-PROB are reported. In Tables 5, 6, 7, and 8 it is reported, respectively for PRP, HLL, GTC, and GT, the MAPE of the models learned with the different algorithms (DNN, SNN, ELM, SVR, KRLS, KNN, and GP) when varying n for the different problems $P1, \dots, P15$. Note that, based on the problem under exam, just a subset of the components may decay. In order to check which components decay in each problem it is possible to refer to Table 4. Moreover, since our dataset cardinality is limited, it is possible that a particular combination of problem and cardinality of the training set cannot be tested. For example, when just one component decays, nearly hundred samples are available (see Section 3) and then a maximum of $n = 55$. In Tables 9, 10, 11, and 12, instead, respectively for PRP, HLL, GTC, and GT, the different indexes of performances (MAE, MSE, NMSE, REP, PPMCC, and MAPE) of the models learned with the different algorithms are reported, when n is the largest possible for the different problems under exam. Note that, in a real-world scenario, the only useful problem is $P15$, namely when all the components decay contemporarily. In Figures 7 we report the MAPE of the models learned with the different algorithms when varying n for $P15$ and the four NPS components. In Figures 8 we report the MAPE of the DNN (the best performing model) when varying n for the different problems under examination and the four NPS components.

From the different tables and figures it is possible to observe that:

- as expected the larger is n the better performances are achieved by the learned models (see Tables 5, 6, 7, and 8 and Figures 8 and 7);
- the models learned with ELM, SNN, and especially DNN generally show the best performances (see Tables 9, 10, 11, and 12 and Figure 7);
- the larger is the number of decaying components authors consider the lower performances are achieved by the learned models (see Tables 5, 6, 7, 8, 9, 10, 11, and 12 and Figure 8);
- the most complicate decay to predict is the one of the HLL, in fact the problems where the HLL decays are the ones which show lower accuracies (see Figure 8);
- as expected, in order to achieve a reasonable MAPE a large number of samples is needed making this regression-based approach not feasible in practice.

Finally in Figure 9 and 10 the FS phase is reported, performed with RF, for $P1$, $P2$, $P3$, and $P4$ (see Figure 9) and for $P15$ (see Figure 10). In particular, for each problem and each feature (see Table 3) the mean decrease in accuracy is reported as described in Section 4.3.

From Figure 9 and 10 it is possible to observe that:

- as expected, when just one component decays (see Figure 9), few predictors have strong predictive power while when all the components decay (see Figure 10) many predictors need to be considered in order to achieve satisfying accuracies;
- from Figure 9, it is possible to note that the RF model can adequately accounts for the relevant features as the outcome is consistent with the available knowledge (note that even if DNN has higher predictive capabilities with respect to RF the latter is still competitive, see Tables 9, 10, 11, and 12 and Figure 7). In fact, for the $P1$ and $P2$ (PRP and HLL decays) the features 22 and 24 (Thrust coefficient stbd and port) have strong predictive power. Moreover, for the $P3$ (GTC decay) the features describing the thermodynamic process, 17, 19, and 20 (GT Compressor outlet air pressure, External Pressure, and HP Turbine exit pressure) have the highest predictive power. Finally, several features are necessary for the GT decay prediction in $P4$. These features are in agreement with Equations (4), (5) and (6).

5.2. CLASS-PROB

In this section, the results on the CLASS-PROB are reported. In Table 13, respectively for PRP, HLL, GTC, and GT, the MAPE of the models learned with the different algorithms (DNN, SNN, ELM, SVM, KRLS, KNN, and GP) is reported, when varying n for $P15$. Note that problems $P1, \dots, P14$ are not reported because of space constraints and because in a real-world scenario the only useful problem is $P15$, namely when all the components decay contemporarily. In Table 14, instead, respectively for PRP, HLL, GTC, and GT, the different indexes of performances (AMR, TP, TN, FP, and FN) of the models learned with the different algorithms are reported, when n is the largest possible for $P15$. In Figures 11 the AMR of the models learned with the different algorithms is reported, when varying n for $P15$ and for the four main NPS components. In Figures 12 the AMR of the DNN (the best performing model) is reported, when varying n for $P15$ and for the four main NPS components.

From the different tables and figures it is possible to observe that:

	n	P4	P7	P9	P10	P12	P13	P14	P15
		PRP	GT, PRP	GTC, PRP	HLL, PRP	GT, GTC, PRP	GT, HLL, PRP	GTC, HLL, PRP	GT, GTC, HLL, PRP
KRLS	10	2.63±0.54	2.73±0.84	2.71±0.74	2.69±0.54	2.88±0.72	2.81±0.22	2.89±0.39	2.86±0.35
	24	2.15±0.81	2.22±0.82	2.29±0.55	2.36±0.67	2.43±0.39	2.59±0.36	2.62±0.34	2.61±0.28
	55	1.33±0.71	1.31±0.46	1.68±0.55	1.75±0.76	1.73±0.45	1.95±0.40	1.87±0.32	2.11±0.60
	130	—	0.75±0.51	1.04±0.49	0.95±0.47	1.07±0.23	1.06±0.09	1.21±0.15	1.19±0.22
	307	—	0.52±0.48	0.76±0.48	0.60±0.39	0.71±0.16	0.69±0.07	0.76±0.15	0.84±0.13
	722	—	0.36±0.34	0.46±0.40	0.36±0.23	0.52±0.09	0.45±0.10	0.46±0.06	0.59±0.07
	4000	—	0.17±0.19	0.28±0.21	0.16±0.11	0.35±0.08	0.27±0.04	0.27±0.04	0.39±0.06
ELM	10	3.35±2.10	2.88±0.75	2.91±0.92	2.90±1.07	2.76±0.37	3.00±0.56	2.88±0.19	2.99±0.47
	24	1.43±1.12	1.76±1.35	2.09±1.40	2.11±1.49	1.92±1.32	2.35±1.23	1.89±1.07	2.00±1.38
	55	0.47±0.58	0.88±0.94	0.87±0.77	0.65±0.57	1.18±0.78	1.05±0.63	1.13±0.40	1.13±0.45
	130	—	0.31±0.41	0.36±0.57	0.18±0.13	0.49±0.18	0.38±0.14	0.48±0.22	0.54±0.19
	307	—	0.18±0.19	0.16±0.26	0.12±0.09	0.25±0.07	0.22±0.05	0.23±0.07	0.28±0.03
	722	—	0.14±0.13	0.12±0.14	0.08±0.08	0.16±0.03	0.16±0.03	0.15±0.02	0.19±0.02
	4000	—	0.10±0.08	0.09±0.11	0.06±0.05	0.12±0.02	0.12±0.02	0.11±0.01	0.15±0.02
KNN	10	3.61±1.26	3.43±1.08	3.25±1.24	3.05±0.37	3.39±0.80	3.27±0.74	3.16±0.75	3.11±0.70
	24	3.18±0.84	2.99±1.27	2.92±0.70	2.98±1.05	2.94±0.40	2.95±0.45	2.99±0.43	3.01±0.47
	55	1.88±1.10	1.60±0.69	2.44±0.85	2.13±1.00	2.07±0.77	2.51±0.78	2.47±0.77	2.70±0.45
	130	—	1.12±0.49	1.64±0.62	1.53±0.61	1.51±0.20	1.62±0.10	1.84±0.30	1.85±0.35
	307	—	0.80±0.71	1.22±0.52	1.14±0.45	1.25±0.29	1.25±0.09	1.42±0.16	1.55±0.29
	722	—	0.62±0.53	1.05±0.53	0.84±0.43	1.06±0.18	0.95±0.09	1.12±0.21	1.31±0.13
	4000	—	0.35±0.57	0.58±0.34	0.48±0.38	0.84±0.24	0.70±0.07	0.80±0.07	1.13±0.16
RF	10	2.94±1.16	2.59±0.53	2.46±0.47	2.82±0.73	2.73±0.46	3.00±0.47	2.96±0.31	2.93±0.36
	24	2.24±0.88	1.99±0.72	2.03±0.73	2.55±0.74	2.18±0.37	2.65±0.24	2.73±0.18	2.74±0.20
	55	1.43±0.57	1.44±0.61	1.39±0.53	1.92±0.38	1.61±0.28	2.25±0.21	2.33±0.15	2.43±0.20
	130	—	0.83±0.56	0.80±0.55	1.50±0.47	0.91±0.17	1.65±0.14	1.72±0.20	1.80±0.16
	307	—	0.38±0.30	0.41±0.32	0.93±0.53	0.48±0.09	1.08±0.11	1.15±0.11	1.22±0.09
	722	—	0.19±0.17	0.19±0.17	0.51±0.14	0.23±0.02	0.67±0.04	0.71±0.07	0.77±0.05
	4000	—	0.10±0.10	0.10±0.09	0.29±0.07	0.12±0.02	0.38±0.04	0.41±0.04	0.46±0.03
GP	10	2.65±0.56	2.72±0.80	2.68±0.64	2.67±0.57	2.78±0.41	2.78±0.22	2.79±0.30	2.79±0.15
	24	2.36±0.67	2.41±0.82	2.41±0.41	2.44±0.50	2.50±0.22	2.58±0.17	2.56±0.16	2.59±0.17
	55	1.83±0.51	1.85±0.54	1.99±0.48	2.05±0.59	2.03±0.24	2.23±0.19	2.18±0.23	2.31±0.31
	130	—	1.23±0.59	1.49±0.31	1.48±0.48	1.56±0.16	1.64±0.13	1.72±0.11	1.74±0.13
	307	—	0.86±0.49	1.13±0.44	1.05±0.39	1.14±0.16	1.18±0.11	1.28±0.12	1.32±0.11
	722	—	0.66±0.52	0.87±0.45	0.79±0.36	0.88±0.11	0.87±0.09	0.93±0.07	1.02±0.10
	4000	—	0.49±0.37	0.63±0.31	0.54±0.32	0.69±0.11	0.65±0.07	0.69±0.04	0.79±0.08
SVR	10	2.51±0.54	2.55±0.83	2.60±0.74	2.59±0.51	2.68±0.71	2.78±0.20	2.67±0.35	2.71±0.34
	24	2.02±0.76	2.07±0.79	2.18±0.50	2.27±0.61	2.36±0.38	2.57±0.34	2.48±0.31	2.37±0.26
	55	1.20±0.64	1.18±0.43	1.57±0.50	1.73±0.71	1.70±0.43	1.90±0.40	1.81±0.32	2.07±0.60
	130	—	0.71±0.48	1.01±0.45	0.88±0.44	0.97±0.23	1.03±0.09	1.14±0.14	1.15±0.20
	307	—	0.49±0.44	0.70±0.47	0.58±0.36	0.66±0.15	0.68±0.07	0.71±0.14	0.82±0.13
	722	—	0.34±0.31	0.43±0.39	0.33±0.21	0.49±0.08	0.44±0.09	0.46±0.06	0.53±0.07
	4000	—	0.16±0.18	0.28±0.19	0.15±0.11	0.34±0.08	0.26±0.04	0.27±0.04	0.38±0.06
DNN	10	3.01±0.96	2.67±0.35	2.54±0.41	2.68±0.50	2.51±0.17	2.80±0.25	2.51±0.09	2.66±0.21
	24	1.33±0.49	1.64±0.63	1.92±0.64	1.89±0.69	1.79±0.58	2.21±0.54	1.79±0.49	1.79±0.63
	55	0.43±0.26	0.82±0.40	0.77±0.36	0.55±0.25	1.12±0.37	0.96±0.29	0.97±0.17	1.01±0.21
	130	—	0.29±0.18	0.31±0.26	0.16±0.06	0.47±0.08	0.35±0.06	0.45±0.10	0.49±0.09
	307	—	0.17±0.08	0.14±0.11	0.10±0.04	0.22±0.03	0.20±0.02	0.21±0.03	0.26±0.01
	722	—	0.13±0.06	0.10±0.06	0.08±0.04	0.14±0.01	0.14±0.01	0.14±0.01	0.17±0.01
	4000	—	0.09±0.04	0.08±0.05	0.06±0.02	0.11±0.01	0.10±0.01	0.10±0.01	0.13±0.01
RFR	10	2.41±0.48	2.24±0.22	2.13±0.19	2.53±0.32	2.41±0.18	2.63±0.20	2.45±0.14	2.49±0.15
	24	1.80±0.35	1.79±0.29	1.81±0.30	2.16±0.32	1.91±0.15	2.20±0.10	2.44±0.07	2.32±0.08
	55	1.23±0.25	1.20±0.25	1.23±0.22	1.57±0.16	1.29±0.12	1.86±0.09	1.92±0.07	2.05±0.09
	130	—	0.68±0.23	0.68±0.24	1.27±0.19	0.79±0.07	1.37±0.06	1.43±0.09	1.47±0.07
	307	—	0.31±0.12	0.35±0.13	0.78±0.23	0.43±0.04	0.93±0.05	0.96±0.04	1.07±0.04
	722	—	0.17±0.07	0.16±0.07	0.45±0.06	0.20±0.01	0.58±0.02	0.59±0.03	0.62±0.02
	4000	—	0.08±0.04	0.08±0.04	0.23±0.03	0.10±0.01	0.32±0.02	0.36±0.02	0.41±0.01
SNN	10	3.24±1.02	2.75±0.35	2.77±0.44	2.77±0.53	2.58±0.18	2.99±0.27	2.75±0.09	2.90±0.22
	24	1.30±0.54	1.66±0.62	2.07±0.68	2.03±0.69	1.90±0.61	2.18±0.56	1.79±0.49	1.98±0.69
	55	0.44±0.29	0.85±0.43	0.80±0.36	0.62±0.26	1.17±0.38	0.99±0.29	1.05±0.20	1.12±0.22
	130	—	0.30±0.19	0.33±0.28	0.17±0.06	0.48±0.08	0.37±0.07	0.47±0.10	0.51±0.09
	307	—	0.17±0.09	0.15±0.12	0.11±0.04	0.24±0.03	0.21±0.02	0.22±0.03	0.27±0.01
	722	—	0.13±0.06	0.12±0.07	0.08±0.04	0.15±0.01	0.16±0.01	0.14±0.01	0.19±0.01
	4000	—	0.09±0.04	0.09±0.05	0.06±0.02	0.12±0.01	0.11±0.01	0.11±0.01	0.14±0.01

Table 5: REG-PROB: MAPE of models learned with the different algorithms (DNN, SNN, ELM, SVR, KRLS, KNN, and GP) when varying n for the different problems $P1, \dots, P15$ and for the PRP NPS component.

	n	P3	P6	P8	P10	P11	P13	P14	P15
		HLL	GT, HLL	GTC, HLL	HLL, PRP	GT, GTC, HLL	GT, HLL, PRP	GTC, HLL, PRP	GT, GTC, HLL, PRP
KRLS	10	4.53±1.13	4.39±1.29	4.49±1.20	4.52±1.06	4.84±0.92	5.20±1.06	4.79±0.26	4.85±0.34
	24	3.57±1.54	3.67±1.16	3.86±1.02	4.21±1.21	4.08±0.48	4.54±0.95	4.52±0.73	4.37±0.46
	55	1.66±1.28	2.34±1.02	2.70±1.16	3.22±1.06	2.72±0.59	3.33±0.63	3.35±0.86	3.36±0.52
	130	—	1.27±0.98	1.43±0.88	1.69±1.02	1.46±0.26	1.90±0.31	1.97±0.29	1.97±0.35
	307	—	0.78±0.56	0.89±0.51	0.88±0.57	0.96±0.18	1.02±0.17	1.11±0.16	1.21±0.19
	722	—	0.38±0.24	0.51±0.34	0.43±0.27	0.58±0.10	0.56±0.09	0.63±0.06	0.71±0.12
	1700	—	0.17±0.10	0.24±0.25	0.15±0.11	0.30±0.06	0.28±0.05	0.34±0.05	0.40±0.08
4000	—	—	—	—	0.15±0.03	0.14±0.02	0.16±0.02	0.21±0.03	
ELM	10	5.17±1.51	5.38±2.27	5.52±1.91	5.53±2.13	5.25±1.67	5.28±1.23	5.06±0.83	5.12±0.85
	24	2.50±1.08	2.98±1.89	2.87±1.72	3.99±2.38	4.30±1.95	4.00±2.06	4.03±2.33	3.71±1.33
	55	0.83±0.99	1.12±1.24	1.39±2.15	1.62±1.78	1.54±0.95	1.28±1.15	1.46±0.96	1.69±1.51
	130	—	0.43±0.31	0.49±0.32	0.45±0.37	0.64±0.29	0.54±0.25	0.62±0.32	0.62±0.16
	307	—	0.27±0.16	0.26±0.16	0.27±0.28	0.30±0.07	0.30±0.07	0.31±0.05	0.35±0.09
	722	—	0.15±0.08	0.15±0.09	0.16±0.12	0.18±0.04	0.19±0.04	0.19±0.02	0.21±0.04
	1700	—	0.10±0.05	0.09±0.03	0.10±0.05	0.11±0.01	0.13±0.03	0.12±0.01	0.14±0.02
4000	—	—	—	—	0.07±0.01	0.09±0.01	0.09±0.02	0.10±0.01	
KNN	10	5.84±1.42	5.63±2.03	5.70±1.94	5.30±1.56	5.72±1.52	5.64±1.05	5.19±0.73	5.43±1.23
	24	5.44±2.26	4.92±1.29	5.11±1.65	5.18±1.78	5.08±0.67	5.30±0.76	5.11±0.39	5.08±0.76
	55	2.45±1.22	3.02±1.04	3.60±1.29	4.07±1.01	3.56±1.52	4.46±1.22	4.57±1.42	4.23±1.17
	130	—	1.99±1.18	2.34±1.19	2.81±1.08	2.33±0.34	2.90±0.27	3.01±0.27	3.06±0.16
	307	—	1.53±0.76	1.88±1.03	2.09±0.83	1.89±0.32	2.14±0.18	2.38±0.18	2.50±0.30
	722	—	1.01±0.82	1.20±0.66	1.47±0.59	1.47±0.17	1.60±0.16	1.79±0.15	2.00±0.19
	1700	—	0.61±0.39	0.60±0.45	1.16±0.54	1.01±0.10	1.18±0.13	1.38±0.10	1.56±0.17
4000	—	—	—	—	0.70±0.08	0.85±0.13	1.03±0.11	1.25±0.12	
RF	10	4.85±1.65	4.95±1.19	4.63±1.55	5.04±1.56	4.82±1.13	5.09±1.29	5.19±1.14	4.88±0.53
	24	3.81±1.50	3.86±1.18	3.83±1.16	4.45±1.14	3.81±0.61	4.11±0.69	4.07±0.57	4.24±0.70
	55	1.93±0.71	2.90±1.21	2.69±0.98	3.57±1.11	2.69±0.33	3.37±0.47	3.45±0.49	3.27±0.43
	130	—	1.65±0.71	1.65±0.67	2.56±1.09	1.69±0.27	2.52±0.36	2.58±0.25	2.50±0.29
	307	—	0.78±0.42	0.83±0.38	1.77±0.70	0.91±0.21	1.70±0.17	1.79±0.24	1.83±0.14
	722	—	0.31±0.15	0.32±0.18	1.07±0.37	0.44±0.09	1.06±0.09	1.16±0.11	1.18±0.11
	1700	—	0.14±0.12	0.17±0.16	0.60±0.19	0.20±0.04	0.65±0.04	0.70±0.05	0.74±0.08
4000	—	—	—	—	0.11±0.02	0.37±0.03	0.40±0.03	0.43±0.04	
GP	10	4.52±1.18	4.32±1.38	4.40±1.44	4.50±1.23	4.72±0.38	4.88±0.34	4.77±0.20	4.82±0.33
	24	3.81±1.29	3.72±1.09	3.88±1.12	4.20±1.07	4.15±0.26	4.46±0.30	4.44±0.35	4.43±0.31
	55	2.66±1.01	2.87±0.88	3.11±0.90	3.59±0.85	3.33±0.35	3.78±0.42	3.72±0.38	3.76±0.38
	130	—	2.07±0.91	2.27±0.95	2.68±1.06	2.31±0.17	2.88±0.30	2.89±0.25	2.96±0.40
	307	—	1.43±0.71	1.64±0.76	1.77±0.89	1.65±0.11	1.95±0.15	2.04±0.18	2.12±0.19
	722	—	1.07±0.48	1.19±0.53	1.22±0.61	1.25±0.13	1.34±0.13	1.42±0.11	1.45±0.16
	1700	—	0.69±0.31	0.79±0.54	0.76±0.47	0.90±0.07	0.94±0.08	1.03±0.09	1.06±0.12
4000	—	—	—	—	0.61±0.04	0.61±0.05	0.67±0.05	0.74±0.09	
SVR	10	4.38±1.09	4.09±1.17	4.29±1.20	4.36±0.98	4.82±0.86	5.18±1.03	4.75±0.26	4.43±0.33
	24	3.51±1.50	3.42±1.15	3.59±0.94	3.85±1.15	3.99±0.48	4.25±0.87	4.35±0.73	3.96±0.45
	55	1.54±1.18	2.16±0.99	2.63±1.16	3.17±0.97	2.60±0.56	3.00±0.59	3.05±0.85	3.30±0.48
	130	—	1.24±0.89	1.33±0.85	1.63±0.94	1.35±0.25	1.73±0.30	1.87±0.27	1.93±0.33
	307	—	0.72±0.54	0.83±0.49	0.80±0.56	0.90±0.18	0.97±0.16	1.10±0.15	1.17±0.18
	722	—	0.36±0.23	0.46±0.33	0.39±0.25	0.52±0.09	0.54±0.08	0.58±0.06	0.67±0.11
	1700	—	0.17±0.10	0.23±0.24	0.14±0.10	0.30±0.06	0.28±0.04	0.33±0.04	0.38±0.07
4000	—	—	—	—	0.14±0.03	0.14±0.02	0.15±0.02	0.20±0.03	
DNN	10	4.70±0.67	4.74±1.06	4.81±0.89	4.91±0.91	4.95±0.71	4.84±0.58	4.45±0.37	4.75±0.39
	24	2.33±0.49	2.59±0.90	2.68±0.81	3.69±1.05	3.92±0.90	3.62±0.90	3.65±1.02	3.20±0.60
	55	0.76±0.44	1.05±0.56	1.22±1.02	1.47±0.79	1.33±0.41	1.19±0.54	1.30±0.45	1.55±0.66
	130	—	0.40±0.14	0.42±0.15	0.41±0.16	0.56±0.13	0.48±0.11	0.53±0.14	0.58±0.07
	307	—	0.24±0.07	0.23±0.07	0.24±0.13	0.26±0.03	0.26±0.03	0.29±0.02	0.33±0.04
	722	—	0.13±0.03	0.14±0.04	0.14±0.05	0.16±0.02	0.17±0.02	0.17±0.01	0.19±0.02
	1700	—	0.09±0.02	0.09±0.01	0.08±0.02	0.10±0.01	0.12±0.01	0.11±0.01	0.13±0.01
4000	—	—	—	—	0.06±0.01	0.08±0.01	0.08±0.01	0.09±0.00	
RFR	10	4.37±0.69	4.05±0.50	4.16±0.69	4.05±0.67	4.18±0.47	4.08±0.56	4.22±0.47	4.08±0.23
	24	2.74±0.34	3.26±0.51	3.13±0.52	3.76±0.47	3.28±0.27	3.61±0.28	3.56±0.23	3.43±0.29
	55	1.55±0.30	2.56±0.52	2.39±0.42	3.06±0.48	2.23±0.14	2.78±0.20	2.79±0.21	2.75±0.19
	130	—	1.49±0.30	1.47±0.27	2.17±0.47	1.40±0.12	2.05±0.15	2.13±0.10	2.13±0.12
	307	—	0.66±0.18	0.68±0.16	1.58±0.31	0.79±0.09	1.40±0.08	1.48±0.11	1.49±0.06
	722	—	0.26±0.06	0.26±0.08	0.91±0.15	0.38±0.04	0.94±0.04	0.97±0.05	1.01±0.04
	1700	—	0.11±0.05	0.14±0.07	0.51±0.08	0.18±0.02	0.58±0.02	0.58±0.02	0.65±0.03
4000	—	—	—	—	0.10±0.01	0.31±0.01	0.33±0.01	0.37±0.02	
SNN	10	4.90±0.74	5.06±1.04	5.44±0.94	5.44±1.03	5.11±0.82	4.85±0.57	4.69±0.38	4.97±0.42
	24	2.26±0.51	2.92±0.86	2.73±0.79	3.89±1.13	4.23±0.89	3.69±0.94	3.78±1.14	3.49±0.61
	55	0.80±0.49	1.08±0.60	1.32±0.98	1.61±0.88	1.40±0.46	1.27±0.57	1.33±0.48	1.65±0.68
	130	—	0.41±0.15	0.46±0.15	0.41±0.18	0.61±0.14	0.51±0.11	0.61±0.14	0.60±0.08
	307	—	0.25±0.08	0.24±0.07	0.26±0.13	0.27±0.04	0.28±0.03	0.28±0.03	0.34±0.04
	722	—	0.14±0.03	0.14±0.05	0.15±0.05	0.17±0.02	0.18±0.02	0.17±0.01	0.20±0.02
	1700	—	0.09±0.03	0.09±0.01	0.09±0.03	0.10±0.01	0.12±0.01	0.12±0.01	0.14±0.01
4000	—	—	—	—	0.07±0.01	0.08±0.01	0.09±0.01	0.10±0.00	

Table 6: REG-PROB: MAPE of models learned with the different algorithms (DNN, SNN, ELM, SVR, KRLS, KNN, and GP) when varying n for the different problems $P1, \dots, P15$ and for the HLL NPS component.

	n	P2	P5	P8	P9	P11	P12	P14	P15
		GTC	GT, GTC	GTC, HLL	GTC, PRP	GT, GTC, HLL	GT, GTC, PRP	GTC, HLL, PRP	GT, GTC, HLL, PRP
KRLS	10	1.29±0.30	1.32±0.29	1.32±0.27	1.39±0.38	1.40±0.19	1.43±0.37	1.39±0.11	1.44±0.17
	24	1.00±0.45	1.10±0.25	1.23±0.31	1.29±0.41	1.35±0.20	1.34±0.24	1.35±0.12	1.38±0.10
	55	0.60±0.30	0.75±0.27	0.93±0.28	1.03±0.15	1.10±0.22	1.19±0.22	1.26±0.07	1.26±0.14
	130	—	0.40±0.22	0.41±0.20	0.47±0.21	0.59±0.08	0.64±0.11	0.70±0.09	0.79±0.06
	307	—	0.21±0.15	0.19±0.10	0.22±0.10	0.31±0.05	0.33±0.07	0.34±0.04	0.45±0.04
	722	—	0.10±0.07	0.08±0.06	0.11±0.05	0.15±0.01	0.16±0.02	0.15±0.03	0.24±0.02
	1700	—	0.04±0.03	0.04±0.04	0.05±0.03	0.07±0.01	0.08±0.01	0.07±0.01	0.12±0.02
4000	—	—	—	—	0.03±0.00	0.04±0.01	0.04±0.00	0.06±0.00	
ELM	10	1.36±0.41	1.37±0.64	1.44±0.25	1.43±0.31	1.39±0.09	1.52±0.36	1.38±0.20	1.42±0.17
	24	0.66±0.80	0.66±1.07	0.88±0.73	0.91±0.98	0.84±0.69	0.89±0.88	1.17±0.89	1.03±0.62
	55	0.07±0.09	0.18±0.38	0.10±0.12	0.14±0.10	0.15±0.11	0.23±0.22	0.25±0.39	0.19±0.08
	130	—	0.03±0.02	0.03±0.02	0.05±0.04	0.05±0.02	0.07±0.02	0.06±0.02	0.08±0.02
	307	—	0.02±0.01	0.02±0.01	0.03±0.01	0.03±0.01	0.04±0.01	0.04±0.01	0.06±0.02
	722	—	0.01±0.01	0.01±0.00	0.02±0.01	0.02±0.00	0.03±0.00	0.03±0.01	0.04±0.01
	1700	—	0.01±0.00	0.01±0.00	0.02±0.00	0.01±0.00	0.02±0.00	0.02±0.00	0.03±0.00
4000	—	—	—	—	0.01±0.00	0.01±0.00	0.02±0.00	0.02±0.00	
KNN	10	1.70±0.62	1.62±0.43	1.51±0.18	1.57±0.46	1.61±0.41	1.63±0.39	1.63±0.41	1.69±0.43
	24	1.53±0.44	1.40±0.32	1.42±0.41	1.53±0.34	1.60±0.30	1.51±0.26	1.52±0.27	1.52±0.26
	55	0.88±0.50	1.10±0.43	1.29±0.42	1.41±0.27	1.37±0.20	1.40±0.09	1.40±0.07	1.39±0.08
	130	—	0.66±0.35	0.70±0.38	0.81±0.29	0.94±0.13	1.05±0.28	1.23±0.21	1.36±0.12
	307	—	0.44±0.24	0.47±0.09	0.60±0.15	0.69±0.05	0.75±0.05	0.87±0.07	1.01±0.08
	722	—	0.28±0.17	0.32±0.08	0.48±0.16	0.49±0.02	0.56±0.04	0.63±0.04	0.80±0.06
	1700	—	0.16±0.11	0.28±0.07	0.38±0.12	0.36±0.03	0.43±0.03	0.47±0.02	0.63±0.03
4000	—	—	—	—	0.27±0.02	0.32±0.02	0.36±0.02	0.49±0.02	
RF	10	1.45±0.40	1.48±0.43	1.46±0.32	1.54±0.45	1.47±0.18	1.52±0.27	1.50±0.17	1.55±0.28
	24	1.35±0.49	1.36±0.43	1.38±0.32	1.46±0.39	1.41±0.15	1.47±0.26	1.43±0.11	1.46±0.17
	55	0.87±0.22	0.99±0.27	1.13±0.23	1.23±0.30	1.23±0.14	1.32±0.15	1.28±0.12	1.33±0.11
	130	—	0.58±0.17	0.71±0.28	0.94±0.30	0.90±0.08	1.09±0.16	1.02±0.09	1.13±0.11
	307	—	0.37±0.09	0.41±0.23	0.55±0.14	0.59±0.05	0.73±0.07	0.71±0.09	0.84±0.07
	722	—	0.23±0.07	0.21±0.12	0.31±0.06	0.38±0.03	0.49±0.04	0.41±0.05	0.57±0.05
	1700	—	0.12±0.04	0.11±0.04	0.17±0.06	0.24±0.02	0.32±0.02	0.22±0.03	0.38±0.02
4000	—	—	—	—	0.15±0.01	0.20±0.01	0.12±0.01	0.24±0.01	
GP	10	1.28±0.30	1.32±0.27	1.31±0.27	1.36±0.36	1.35±0.11	1.38±0.18	1.37±0.11	1.40±0.13
	24	1.08±0.37	1.15±0.18	1.23±0.25	1.29±0.40	1.34±0.16	1.32±0.19	1.34±0.13	1.36±0.12
	55	0.82±0.29	0.91±0.20	1.06±0.22	1.10±0.23	1.17±0.11	1.21±0.14	1.26±0.07	1.29±0.05
	130	—	0.66±0.23	0.77±0.26	0.82±0.27	0.91±0.08	0.97±0.08	1.05±0.06	1.11±0.08
	307	—	0.46±0.19	0.47±0.15	0.54±0.17	0.63±0.06	0.69±0.08	0.77±0.05	0.84±0.06
	722	—	0.30±0.13	0.26±0.10	0.33±0.09	0.41±0.03	0.45±0.05	0.49±0.05	0.59±0.05
	1700	—	0.19±0.10	0.16±0.07	0.20±0.07	0.25±0.01	0.28±0.02	0.29±0.03	0.38±0.03
4000	—	—	—	—	0.15±0.01	0.17±0.02	0.17±0.02	0.23±0.02	
SVR	10	1.24±0.28	1.24±0.28	1.31±0.25	1.35±0.36	1.34±0.17	1.42±0.35	1.36±0.11	1.44±0.17
	24	0.99±0.41	1.03±0.24	1.12±0.30	1.24±0.37	1.32±0.18	1.32±0.22	1.35±0.12	1.33±0.10
	55	0.55±0.30	0.69±0.27	0.87±0.25	0.96±0.14	1.06±0.20	1.10±0.21	1.20±0.06	1.16±0.13
	130	—	0.40±0.20	0.38±0.19	0.42±0.20	0.53±0.07	0.58±0.11	0.67±0.09	0.74±0.06
	307	—	0.19±0.15	0.19±0.10	0.21±0.10	0.29±0.04	0.31±0.06	0.32±0.04	0.45±0.04
	722	—	0.09±0.07	0.08±0.05	0.10±0.05	0.14±0.01	0.15±0.02	0.15±0.03	0.23±0.02
	1700	—	0.04±0.03	0.04±0.03	0.05±0.02	0.07±0.01	0.07±0.01	0.07±0.01	0.11±0.01
4000	—	—	—	—	0.03±0.00	0.04±0.01	0.03±0.00	0.05±0.00	
DNN	10	1.28±0.18	1.30±0.28	1.29±0.11	1.31±0.14	1.27±0.04	1.30±0.17	1.25±0.09	1.29±0.07
	24	0.60±0.36	0.62±0.49	0.79±0.32	0.80±0.46	0.72±0.30	0.78±0.39	0.99±0.42	0.93±0.29
	55	0.06±0.04	0.16±0.18	0.09±0.05	0.13±0.05	0.14±0.05	0.21±0.10	0.23±0.17	0.18±0.04
	130	—	0.03±0.01	0.03±0.01	0.05±0.02	0.04±0.01	0.06±0.01	0.05±0.01	0.08±0.01
	307	—	0.02±0.00	0.01±0.00	0.03±0.01	0.03±0.00	0.03±0.00	0.03±0.00	0.05±0.01
	722	—	0.01±0.00	0.01±0.00	0.02±0.00	0.02±0.00	0.02±0.00	0.02±0.00	0.04±0.00
	1700	—	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.02±0.00	0.02±0.00	0.03±0.00
4000	—	—	—	—	0.01±0.00	0.01±0.00	0.01±0.00	0.02±0.00	
RFR	10	1.18±0.17	1.20±0.18	1.24±0.14	1.35±0.19	1.19±0.08	1.25±0.11	1.32±0.07	1.36±0.12
	24	1.11±0.20	1.18±0.19	1.21±0.14	1.21±0.16	1.21±0.07	1.31±0.11	1.21±0.05	1.17±0.07
	55	0.78±0.10	0.82±0.12	0.91±0.09	1.03±0.13	1.03±0.06	1.16±0.06	1.04±0.05	1.17±0.05
	130	—	0.52±0.07	0.63±0.12	0.79±0.13	0.78±0.03	0.92±0.07	0.83±0.04	0.98±0.05
	307	—	0.31±0.03	0.35±0.10	0.45±0.06	0.51±0.02	0.66±0.03	0.62±0.04	0.72±0.03
	722	—	0.20±0.03	0.17±0.05	0.25±0.03	0.31±0.01	0.40±0.02	0.36±0.02	0.49±0.02
	1700	—	0.10±0.02	0.09±0.02	0.15±0.02	0.19±0.01	0.26±0.01	0.20±0.01	0.32±0.01
4000	—	—	—	—	0.13±0.00	0.18±0.01	0.11±0.00	0.21±0.01	
SNN	10	1.31±0.20	1.34±0.30	1.37±0.12	1.40±0.15	1.32±0.04	1.47±0.17	1.25±0.09	1.42±0.08
	24	0.66±0.40	0.60±0.50	0.80±0.36	0.86±0.48	0.82±0.33	0.82±0.44	1.13±0.42	0.94±0.31
	55	0.07±0.04	0.17±0.18	0.10±0.06	0.13±0.05	0.14±0.05	0.21±0.10	0.24±0.19	0.18±0.04
	130	—	0.03±0.01	0.03±0.01	0.05±0.02	0.05±0.01	0.06±0.01	0.06±0.01	0.08±0.01
	307	—	0.02±0.01	0.01±0.00	0.03±0.01	0.03±0.00	0.03±0.00	0.04±0.00	0.05±0.01
	722	—	0.01±0.00	0.01±0.00	0.02±0.00	0.02±0.00	0.02±0.00	0.03±0.00	0.04±0.00
	1700	—	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.02±0.00	0.02±0.00	0.03±0.00
4000	—	—	—	—	0.01±0.00	0.01±0.00	0.01±0.00	0.02±0.00	

Table 7: REG-PROB: MAPE of models learned with the different algorithms (DNN, SNN, ELM, SVR, KRLS, KNN, and GP) when varying n for the different problems $P1, \dots, P15$ and for the GTC NPS component.

	n	P1	P5	P6	P7	P11	P12	P14	P15
		GT	GT, GTC	GT, HLL	GT, PRP	GT, GTC, HLL	GT, GTC, PRP	GTC, HLL, PRP	GT, GTC, HLL, PRP
KRLS	10	0.64±0.17	0.68±0.17	0.66±0.13	0.63±0.08	0.69±0.07	0.69±0.06	0.70±0.10	0.69±0.04
	24	0.49±0.22	0.61±0.13	0.63±0.12	0.62±0.13	0.67±0.03	0.66±0.03	0.68±0.10	0.67±0.02
	55	0.24±0.15	0.49±0.09	0.53±0.12	0.52±0.11	0.64±0.06	0.63±0.07	0.64±0.05	0.68±0.03
	130	—	0.29±0.13	0.25±0.09	0.30±0.13	0.45±0.03	0.44±0.04	0.42±0.05	0.55±0.13
	307	—	0.19±0.10	0.12±0.03	0.16±0.11	0.26±0.03	0.27±0.03	0.24±0.02	0.34±0.03
	722	—	0.10±0.05	0.06±0.03	0.08±0.04	0.14±0.01	0.15±0.02	0.12±0.02	0.20±0.02
	1700	—	0.05±0.03	0.02±0.02	0.03±0.02	0.07±0.01	0.08±0.01	0.06±0.01	0.11±0.01
4000	—	—	—	—	0.03±0.00	0.04±0.01	0.03±0.00	0.05±0.00	
ELM	10	0.69±0.18	0.73±0.18	0.74±0.30	0.73±0.22	0.70±0.09	0.71±0.18	0.68±0.05	0.77±0.20
	24	0.61±0.32	0.66±0.28	0.67±0.20	0.68±0.22	0.68±0.10	0.67±0.13	0.69±0.28	0.69±0.11
	55	0.13±0.26	0.24±0.25	0.10±0.10	0.20±0.19	0.19±0.18	0.19±0.10	0.24±0.12	0.34±0.28
	130	—	0.04±0.04	0.02±0.02	0.04±0.02	0.04±0.01	0.05±0.02	0.05±0.02	0.06±0.01
	307	—	0.01±0.01	0.01±0.01	0.02±0.01	0.02±0.01	0.02±0.00	0.02±0.00	0.03±0.00
	722	—	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.02±0.00
	1700	—	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00
4000	—	—	—	—	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	
KNN	10	0.82±0.32	0.76±0.27	0.73±0.08	0.79±0.31	0.78±0.23	0.83±0.23	0.82±0.27	0.79±0.25
	24	0.71±0.25	0.68±0.14	0.70±0.17	0.74±0.22	0.77±0.20	0.75±0.11	0.72±0.15	0.70±0.03
	55	0.37±0.20	0.68±0.14	0.68±0.14	0.65±0.11	0.69±0.02	0.71±0.09	0.69±0.06	0.69±0.04
	130	—	0.51±0.19	0.42±0.14	0.47±0.13	0.67±0.04	0.69±0.03	0.69±0.05	0.69±0.04
	307	—	0.36±0.13	0.28±0.07	0.36±0.13	0.51±0.04	0.54±0.11	0.49±0.03	0.69±0.03
	722	—	0.27±0.09	0.23±0.07	0.29±0.07	0.40±0.02	0.41±0.03	0.38±0.03	0.57±0.06
	1700	—	0.22±0.06	0.18±0.03	0.22±0.07	0.30±0.01	0.32±0.02	0.29±0.02	0.45±0.02
4000	—	—	—	—	0.24±0.01	0.25±0.02	0.23±0.01	0.36±0.02	
RF	10	0.76±0.14	0.79±0.13	0.73±0.22	0.71±0.17	0.77±0.11	0.73±0.07	0.74±0.12	0.76±0.13
	24	0.66±0.17	0.72±0.21	0.73±0.20	0.70±0.15	0.72±0.05	0.73±0.05	0.72±0.07	0.74±0.09
	55	0.46±0.15	0.59±0.13	0.62±0.12	0.62±0.17	0.67±0.10	0.69±0.04	0.65±0.05	0.72±0.05
	130	—	0.44±0.08	0.43±0.12	0.51±0.11	0.56±0.05	0.61±0.04	0.55±0.04	0.65±0.03
	307	—	0.29±0.04	0.27±0.09	0.34±0.10	0.42±0.04	0.49±0.04	0.40±0.03	0.53±0.02
	722	—	0.19±0.06	0.14±0.07	0.21±0.09	0.29±0.02	0.36±0.04	0.25±0.01	0.41±0.03
	1700	—	0.11±0.04	0.07±0.03	0.13±0.06	0.18±0.01	0.24±0.02	0.15±0.01	0.29±0.02
4000	—	—	—	—	0.11±0.00	0.16±0.02	0.08±0.01	0.19±0.01	
GP	10	0.63±0.15	0.67±0.13	0.65±0.13	0.62±0.08	0.68±0.05	0.68±0.04	0.68±0.07	0.68±0.03
	24	0.53±0.19	0.61±0.11	0.63±0.10	0.60±0.08	0.66±0.02	0.66±0.02	0.67±0.07	0.67±0.02
	55	0.40±0.16	0.54±0.10	0.56±0.11	0.55±0.08	0.64±0.03	0.63±0.03	0.64±0.04	0.67±0.03
	130	—	0.41±0.10	0.43±0.10	0.43±0.07	0.57±0.03	0.56±0.04	0.56±0.04	0.63±0.03
	307	—	0.30±0.09	0.27±0.04	0.30±0.09	0.45±0.03	0.45±0.02	0.44±0.03	0.53±0.03
	722	—	0.20±0.07	0.16±0.05	0.17±0.07	0.32±0.01	0.32±0.02	0.30±0.03	0.41±0.02
	1700	—	0.14±0.05	0.09±0.04	0.11±0.05	0.20±0.01	0.21±0.02	0.18±0.02	0.29±0.01
4000	—	—	—	—	0.11±0.01	0.13±0.01	0.11±0.01	0.18±0.01	
SVR	10	0.60±0.15	0.63±0.16	0.62±0.12	0.57±0.08	0.67±0.06	0.65±0.06	0.66±0.10	0.65±0.04
	24	0.47±0.20	0.58±0.13	0.62±0.11	0.62±0.13	0.62±0.03	0.60±0.03	0.64±0.10	0.66±0.02
	55	0.23±0.14	0.49±0.09	0.51±0.11	0.49±0.10	0.63±0.06	0.62±0.07	0.60±0.04	0.64±0.03
	130	—	0.26±0.12	0.24±0.08	0.27±0.12	0.41±0.03	0.41±0.04	0.40±0.04	0.54±0.12
	307	—	0.18±0.10	0.12±0.03	0.16±0.10	0.26±0.03	0.26±0.03	0.21±0.02	0.34±0.03
	722	—	0.09±0.04	0.06±0.03	0.08±0.04	0.14±0.01	0.14±0.02	0.11±0.02	0.19±0.02
	1700	—	0.05±0.03	0.02±0.02	0.03±0.02	0.06±0.01	0.07±0.01	0.06±0.01	0.10±0.01
4000	—	—	—	—	0.03±0.00	0.04±0.01	0.03±0.00	0.05±0.00	
DNN	10	0.64±0.08	0.67±0.08	0.67±0.14	0.63±0.10	0.65±0.04	0.64±0.08	0.61±0.02	0.69±0.09
	24	0.53±0.14	0.59±0.12	0.63±0.09	0.64±0.10	0.61±0.05	0.64±0.06	0.65±0.13	0.65±0.05
	55	0.12±0.12	0.21±0.12	0.09±0.04	0.18±0.08	0.17±0.08	0.18±0.04	0.22±0.06	0.31±0.13
	130	—	0.03±0.02	0.02±0.01	0.04±0.01	0.04±0.01	0.05±0.01	0.04±0.01	0.05±0.01
	307	—	0.01±0.00	0.01±0.00	0.02±0.00	0.02±0.00	0.02±0.00	0.02±0.00	0.02±0.00
	722	—	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00
	1700	—	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00
4000	—	—	—	—	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	
RFR	10	0.63±0.06	0.70±0.06	0.62±0.09	0.59±0.07	0.62±0.05	0.62±0.03	0.65±0.05	0.62±0.06
	24	0.58±0.07	0.58±0.08	0.64±0.08	0.61±0.06	0.60±0.02	0.64±0.02	0.63±0.03	0.63±0.04
	55	0.38±0.06	0.52±0.06	0.51±0.05	0.52±0.07	0.54±0.04	0.59±0.02	0.57±0.02	0.60±0.02
	130	—	0.37±0.03	0.37±0.05	0.46±0.05	0.45±0.02	0.53±0.02	0.46±0.02	0.54±0.01
	307	—	0.25±0.02	0.24±0.04	0.29±0.04	0.34±0.02	0.43±0.02	0.35±0.01	0.46±0.01
	722	—	0.17±0.02	0.12±0.03	0.18±0.04	0.25±0.01	0.30±0.02	0.21±0.01	0.34±0.01
	1700	—	0.10±0.02	0.06±0.01	0.10±0.03	0.15±0.00	0.22±0.01	0.13±0.00	0.25±0.01
4000	—	—	—	—	0.09±0.00	0.13±0.01	0.07±0.00	0.16±0.01	
SNN	10	0.68±0.09	0.71±0.09	0.73±0.13	0.71±0.11	0.63±0.04	0.69±0.08	0.63±0.02	0.70±0.10
	24	0.55±0.15	0.66±0.13	0.64±0.09	0.64±0.10	0.63±0.05	0.67±0.06	0.65±0.14	0.64±0.05
	55	0.12±0.13	0.22±0.12	0.10±0.05	0.18±0.09	0.18±0.09	0.19±0.05	0.21±0.06	0.34±0.13
	130	—	0.04±0.02	0.02±0.01	0.04±0.01	0.04±0.01	0.05±0.01	0.04±0.01	0.05±0.01
	307	—	0.01±0.00	0.01±0.00	0.02±0.00	0.02±0.00	0.02±0.00	0.02±0.00	0.03±0.00
	722	—	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.02±0.00
	1700	—	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00
4000	—	—	—	—	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	

Table 8: REG-PROB: MAPE of models learned with the different algorithms (DNN, SNN, ELM, SVR, KRLS, KNN, and GP) when varying n for the different problems $P1, \dots, P15$ and for the GT NPS component.

	Problem	P4	P7	P9	P10	P12	P13	P14	P15
	Decayed Components	PRP	GT, PRP	GTC, PRP	HLL, PRP	GT, GTC, PRP	GT, HLL, PRP	GTC, HLL, PRP	GT, GTC, HLL, PRP
KRLS	MAE	$1.26 \cdot 10^{-2}$	$1.60 \cdot 10^{-3}$	$2.66 \cdot 10^{-3}$	$1.53 \cdot 10^{-3}$	$2.03 \cdot 10^{-3}$	$1.30 \cdot 10^{-3}$	$1.40 \cdot 10^{-3}$	$2.27 \cdot 10^{-3}$
	MSE	$1.95 \cdot 10^{-2}$	$4.04 \cdot 10^{-3}$	$6.76 \cdot 10^{-3}$	$3.28 \cdot 10^{-3}$	$6.04 \cdot 10^{-3}$	$3.54 \cdot 10^{-3}$	$3.81 \cdot 10^{-3}$	$6.08 \cdot 10^{-3}$
	NMSE	$6.42 \cdot 10^{-1}$	$1.31 \cdot 10^{-1}$	$2.31 \cdot 10^{-1}$	$1.08 \cdot 10^{-1}$	$1.96 \cdot 10^{-1}$	$1.15 \cdot 10^{-1}$	$1.24 \cdot 10^{-1}$	$1.98 \cdot 10^{-1}$
	REP	2.05	$4.25 \cdot 10^{-1}$	$7.11 \cdot 10^{-1}$	$3.45 \cdot 10^{-1}$	$6.36 \cdot 10^{-1}$	$3.73 \cdot 10^{-1}$	$4.01 \cdot 10^{-1}$	$6.40 \cdot 10^{-1}$
	PPMCC	0.77	0.99	0.97	0.99	0.98	0.99	0.99	0.98
	MAPE	1.33	0.17	0.28	0.16	0.21	0.14	0.15	0.24
ELM	MAE	$4.46 \cdot 10^{-3}$	$9.02 \cdot 10^{-4}$	$8.53 \cdot 10^{-4}$	$5.88 \cdot 10^{-4}$	$7.91 \cdot 10^{-4}$	$7.96 \cdot 10^{-4}$	$8.21 \cdot 10^{-4}$	$1.06 \cdot 10^{-3}$
	MSE	$8.55 \cdot 10^{-3}$	$1.67 \cdot 10^{-3}$	$1.78 \cdot 10^{-3}$	$1.14 \cdot 10^{-3}$	$1.64 \cdot 10^{-3}$	$1.53 \cdot 10^{-3}$	$1.59 \cdot 10^{-3}$	$2.04 \cdot 10^{-3}$
	NMSE	$2.80 \cdot 10^{-1}$	$5.53 \cdot 10^{-2}$	$6.03 \cdot 10^{-2}$	$3.85 \cdot 10^{-2}$	$5.33 \cdot 10^{-2}$	$4.97 \cdot 10^{-2}$	$5.15 \cdot 10^{-2}$	$6.60 \cdot 10^{-2}$
	REP	$9.01 \cdot 10^{-1}$	$1.75 \cdot 10^{-1}$	$1.87 \cdot 10^{-1}$	$1.20 \cdot 10^{-1}$	$1.73 \cdot 10^{-1}$	$1.61 \cdot 10^{-1}$	$1.68 \cdot 10^{-1}$	$2.15 \cdot 10^{-1}$
	PPMCC	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.47	0.10	0.09	0.06	0.08	0.08	0.09	0.11
KNN	MAE	$1.78 \cdot 10^{-2}$	$3.31 \cdot 10^{-3}$	$5.55 \cdot 10^{-3}$	$4.55 \cdot 10^{-3}$	$5.21 \cdot 10^{-3}$	$4.49 \cdot 10^{-3}$	$5.13 \cdot 10^{-3}$	$8.63 \cdot 10^{-3}$
	MSE	$2.37 \cdot 10^{-2}$	$9.50 \cdot 10^{-3}$	$1.31 \cdot 10^{-2}$	$9.98 \cdot 10^{-3}$	$1.34 \cdot 10^{-2}$	$1.08 \cdot 10^{-2}$	$1.10 \cdot 10^{-2}$	$1.50 \cdot 10^{-2}$
	NMSE	$7.87 \cdot 10^{-1}$	$3.08 \cdot 10^{-1}$	$4.49 \cdot 10^{-1}$	$3.32 \cdot 10^{-1}$	$4.36 \cdot 10^{-1}$	$3.49 \cdot 10^{-1}$	$3.56 \cdot 10^{-1}$	$4.89 \cdot 10^{-1}$
	REP	2.50	$9.97 \cdot 10^{-1}$	1.38	1.05	1.41	1.13	1.16	1.58
	PPMCC	0.60	0.93	0.89	0.94	0.90	0.94	0.94	0.87
	MAPE	1.88	0.35	0.58	0.48	0.55	0.47	0.54	0.91
RF	MAE	$1.35 \cdot 10^{-2}$	$9.01 \cdot 10^{-4}$	$9.47 \cdot 10^{-4}$	$2.79 \cdot 10^{-3}$	$6.55 \cdot 10^{-4}$	$1.98 \cdot 10^{-3}$	$2.18 \cdot 10^{-3}$	$2.45 \cdot 10^{-3}$
	MSE	$1.79 \cdot 10^{-2}$	$2.36 \cdot 10^{-3}$	$2.18 \cdot 10^{-3}$	$4.15 \cdot 10^{-3}$	$1.79 \cdot 10^{-3}$	$3.41 \cdot 10^{-3}$	$3.55 \cdot 10^{-3}$	$3.81 \cdot 10^{-3}$
	NMSE	$5.75 \cdot 10^{-1}$	$7.87 \cdot 10^{-2}$	$7.75 \cdot 10^{-2}$	$1.41 \cdot 10^{-1}$	$5.80 \cdot 10^{-2}$	$1.11 \cdot 10^{-1}$	$1.15 \cdot 10^{-1}$	$1.23 \cdot 10^{-1}$
	REP	1.88	$2.49 \cdot 10^{-1}$	$2.29 \cdot 10^{-1}$	$4.35 \cdot 10^{-1}$	$1.88 \cdot 10^{-1}$	$3.59 \cdot 10^{-1}$	$3.73 \cdot 10^{-1}$	$4.01 \cdot 10^{-1}$
	PPMCC	0.85	1.00	1.00	0.99	1.00	0.99	0.99	0.99
	MAPE	1.43	0.10	0.10	0.29	0.07	0.21	0.23	0.26
GP	MAE	$1.73 \cdot 10^{-2}$	$4.65 \cdot 10^{-3}$	$6.00 \cdot 10^{-3}$	$5.08 \cdot 10^{-3}$	$5.06 \cdot 10^{-3}$	$4.55 \cdot 10^{-3}$	$4.63 \cdot 10^{-3}$	$5.86 \cdot 10^{-3}$
	MSE	$2.21 \cdot 10^{-2}$	$8.97 \cdot 10^{-3}$	$1.20 \cdot 10^{-2}$	$8.68 \cdot 10^{-3}$	$1.05 \cdot 10^{-2}$	$8.51 \cdot 10^{-3}$	$8.64 \cdot 10^{-3}$	$1.11 \cdot 10^{-2}$
	NMSE	$7.26 \cdot 10^{-1}$	$2.91 \cdot 10^{-1}$	$4.11 \cdot 10^{-1}$	$2.88 \cdot 10^{-1}$	$3.41 \cdot 10^{-1}$	$2.75 \cdot 10^{-1}$	$2.80 \cdot 10^{-1}$	$3.62 \cdot 10^{-1}$
	REP	2.32	$9.43 \cdot 10^{-1}$	1.27	$9.13 \cdot 10^{-1}$	1.11	$8.95 \cdot 10^{-1}$	$9.09 \cdot 10^{-1}$	1.17
	PPMCC	0.76	0.96	0.91	0.96	0.94	0.96	0.96	0.93
	MAPE	1.83	0.49	0.63	0.54	0.53	0.48	0.49	0.62
SVR	MAE	$1.23 \cdot 10^{-2}$	$1.58 \cdot 10^{-3}$	$2.60 \cdot 10^{-3}$	$1.47 \cdot 10^{-3}$	$1.95 \cdot 10^{-3}$	$1.28 \cdot 10^{-3}$	$1.29 \cdot 10^{-3}$	$2.23 \cdot 10^{-3}$
	MSE	$1.93 \cdot 10^{-2}$	$3.65 \cdot 10^{-3}$	$6.65 \cdot 10^{-3}$	$3.23 \cdot 10^{-3}$	$5.57 \cdot 10^{-3}$	$3.29 \cdot 10^{-3}$	$3.54 \cdot 10^{-3}$	$5.61 \cdot 10^{-3}$
	NMSE	$5.78 \cdot 10^{-1}$	$1.18 \cdot 10^{-1}$	$2.24 \cdot 10^{-1}$	$9.94 \cdot 10^{-2}$	$1.89 \cdot 10^{-1}$	$1.12 \cdot 10^{-1}$	$1.15 \cdot 10^{-1}$	$1.95 \cdot 10^{-1}$
	REP	1.91	$4.23 \cdot 10^{-1}$	$6.81 \cdot 10^{-1}$	$3.25 \cdot 10^{-1}$	$6.31 \cdot 10^{-1}$	$3.66 \cdot 10^{-1}$	$3.93 \cdot 10^{-1}$	$6.19 \cdot 10^{-1}$
	PPMCC	0.81	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	1.20	0.16	0.28	0.15	0.19	0.13	0.14	0.23
DNN	MAE	$3.85 \cdot 10^{-3}$	$8.24 \cdot 10^{-4}$	$7.68 \cdot 10^{-4}$	$5.34 \cdot 10^{-4}$	$7.22 \cdot 10^{-4}$	$7.41 \cdot 10^{-4}$	$7.66 \cdot 10^{-4}$	$9.96 \cdot 10^{-4}$
	MSE	$8.00 \cdot 10^{-3}$	$1.43 \cdot 10^{-3}$	$1.54 \cdot 10^{-3}$	$1.01 \cdot 10^{-3}$	$1.41 \cdot 10^{-3}$	$1.33 \cdot 10^{-3}$	$1.41 \cdot 10^{-3}$	$1.85 \cdot 10^{-3}$
	NMSE	$2.53 \cdot 10^{-1}$	$4.99 \cdot 10^{-2}$	$5.70 \cdot 10^{-2}$	$3.59 \cdot 10^{-2}$	$4.57 \cdot 10^{-2}$	$4.56 \cdot 10^{-2}$	$4.59 \cdot 10^{-2}$	$6.18 \cdot 10^{-2}$
	REP	$7.67 \cdot 10^{-1}$	$1.65 \cdot 10^{-1}$	$1.73 \cdot 10^{-1}$	$1.04 \cdot 10^{-1}$	$1.64 \cdot 10^{-1}$	$1.51 \cdot 10^{-1}$	$1.45 \cdot 10^{-1}$	$1.90 \cdot 10^{-1}$
	PPMCC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.43	0.09	0.08	0.06	0.08	0.08	0.08	0.10
RFR	MAE	$1.21 \cdot 10^{-2}$	$7.71 \cdot 10^{-4}$	$8.02 \cdot 10^{-4}$	$2.27 \cdot 10^{-3}$	$5.62 \cdot 10^{-4}$	$1.63 \cdot 10^{-3}$	$1.76 \cdot 10^{-3}$	$2.01 \cdot 10^{-3}$
	MSE	$1.47 \cdot 10^{-2}$	$1.98 \cdot 10^{-3}$	$1.91 \cdot 10^{-3}$	$3.64 \cdot 10^{-3}$	$1.54 \cdot 10^{-3}$	$2.99 \cdot 10^{-3}$	$3.00 \cdot 10^{-3}$	$3.10 \cdot 10^{-3}$
	NMSE	$5.02 \cdot 10^{-1}$	$6.96 \cdot 10^{-2}$	$6.43 \cdot 10^{-2}$	$1.14 \cdot 10^{-1}$	$5.09 \cdot 10^{-2}$	$9.84 \cdot 10^{-2}$	$1.00 \cdot 10^{-1}$	$9.90 \cdot 10^{-2}$
	REP	1.56	$2.11 \cdot 10^{-1}$	$1.87 \cdot 10^{-1}$	$3.81 \cdot 10^{-1}$	$1.59 \cdot 10^{-1}$	$3.19 \cdot 10^{-1}$	$3.09 \cdot 10^{-1}$	$3.25 \cdot 10^{-1}$
	PPMCC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	1.23	0.08	0.08	0.23	0.06	0.17	0.19	0.23
SNN	MAE	$4.38 \cdot 10^{-3}$	$8.79 \cdot 10^{-4}$	$7.81 \cdot 10^{-4}$	$5.72 \cdot 10^{-4}$	$7.70 \cdot 10^{-4}$	$7.21 \cdot 10^{-4}$	$7.55 \cdot 10^{-4}$	$9.89 \cdot 10^{-4}$
	MSE	$8.32 \cdot 10^{-3}$	$1.51 \cdot 10^{-3}$	$1.74 \cdot 10^{-3}$	$1.09 \cdot 10^{-3}$	$1.57 \cdot 10^{-3}$	$1.45 \cdot 10^{-3}$	$1.50 \cdot 10^{-3}$	$1.95 \cdot 10^{-3}$
	NMSE	$2.71 \cdot 10^{-1}$	$5.10 \cdot 10^{-2}$	$5.63 \cdot 10^{-2}$	$3.54 \cdot 10^{-2}$	$5.06 \cdot 10^{-2}$	$4.91 \cdot 10^{-2}$	$5.07 \cdot 10^{-2}$	$6.55 \cdot 10^{-2}$
	REP	$8.87 \cdot 10^{-1}$	$1.61 \cdot 10^{-1}$	$1.69 \cdot 10^{-1}$	$1.09 \cdot 10^{-1}$	$1.72 \cdot 10^{-1}$	$1.45 \cdot 10^{-1}$	$1.52 \cdot 10^{-1}$	$1.98 \cdot 10^{-1}$
	PPMCC	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.44	0.09	0.09	0.06	0.08	0.08	0.08	0.10

Table 9: REG-PROB: different indexes of performances (MAE, MSE, NMSE, REP, PPMCC, and MAPE) of the models learned with the different algorithms (DNN, SNN, ELM, SVR, KRLS, KNN, and GP) when n is the largest possible for the different problems $P1, \dots, P15$ and for the PRP NPS component.

	Problem	P3	P6	P8	P10	P11	P13	P14	P15
	Decayed Components	HLL	GT, HLL	GTC, HLL	HLL, PRP	GT, GTC, HLL	GT, HLL, PRP	GTC, HLL, PRP	GT, GTC, HLL, PRP
KRLS	MAE	$1.80 \cdot 10^{-2}$	$1.92 \cdot 10^{-3}$	$2.60 \cdot 10^{-3}$	$1.69 \cdot 10^{-3}$	$1.68 \cdot 10^{-3}$	$1.57 \cdot 10^{-3}$	$1.81 \cdot 10^{-3}$	$2.31 \cdot 10^{-3}$
	MSE	$3.01 \cdot 10^{-2}$	$4.49 \cdot 10^{-3}$	$6.56 \cdot 10^{-3}$	$3.34 \cdot 10^{-3}$	$4.77 \cdot 10^{-3}$	$3.84 \cdot 10^{-3}$	$4.46 \cdot 10^{-3}$	$5.52 \cdot 10^{-3}$
	NMSE	$4.94 \cdot 10^{-1}$	$7.58 \cdot 10^{-2}$	$1.12 \cdot 10^{-1}$	$5.70 \cdot 10^{-2}$	$7.72 \cdot 10^{-2}$	$6.16 \cdot 10^{-2}$	$7.22 \cdot 10^{-2}$	$8.96 \cdot 10^{-2}$
	REP	2.74	$4.07 \cdot 10^{-1}$	$5.94 \cdot 10^{-1}$	$3.03 \cdot 10^{-1}$	$4.33 \cdot 10^{-1}$	$3.48 \cdot 10^{-1}$	$4.05 \cdot 10^{-1}$	$5.01 \cdot 10^{-1}$
	PPMCC	0.88	1.00	0.99	1.00	1.00	1.00	1.00	1.00
	MAPE	1.66	0.17	0.24	0.15	0.15	0.14	0.16	0.21
ELM	MAE	$9.05 \cdot 10^{-3}$	$1.10 \cdot 10^{-3}$	$1.05 \cdot 10^{-3}$	$1.05 \cdot 10^{-3}$	$7.91 \cdot 10^{-4}$	$9.69 \cdot 10^{-4}$	$9.77 \cdot 10^{-4}$	$1.10 \cdot 10^{-3}$
	MSE	$1.51 \cdot 10^{-2}$	$1.90 \cdot 10^{-3}$	$1.76 \cdot 10^{-3}$	$1.70 \cdot 10^{-3}$	$1.43 \cdot 10^{-3}$	$1.78 \cdot 10^{-3}$	$1.74 \cdot 10^{-3}$	$1.94 \cdot 10^{-3}$
	NMSE	$2.53 \cdot 10^{-1}$	$3.18 \cdot 10^{-2}$	$2.86 \cdot 10^{-2}$	$2.83 \cdot 10^{-2}$	$2.32 \cdot 10^{-2}$	$2.87 \cdot 10^{-2}$	$2.84 \cdot 10^{-2}$	$3.12 \cdot 10^{-2}$
	REP	1.37	$1.73 \cdot 10^{-1}$	$1.59 \cdot 10^{-1}$	$1.54 \cdot 10^{-1}$	$1.30 \cdot 10^{-1}$	$1.61 \cdot 10^{-1}$	$1.58 \cdot 10^{-1}$	$1.76 \cdot 10^{-1}$
	PPMCC	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.83	0.10	0.09	0.10	0.07	0.09	0.09	0.10
KNN	MAE	$2.66 \cdot 10^{-2}$	$6.76 \cdot 10^{-3}$	$6.62 \cdot 10^{-3}$	$1.28 \cdot 10^{-2}$	$7.64 \cdot 10^{-3}$	$9.37 \cdot 10^{-3}$	$1.13 \cdot 10^{-2}$	$1.37 \cdot 10^{-2}$
	MSE	$3.30 \cdot 10^{-2}$	$1.69 \cdot 10^{-2}$	$1.47 \cdot 10^{-2}$	$1.76 \cdot 10^{-2}$	$1.98 \cdot 10^{-2}$	$1.75 \cdot 10^{-2}$	$2.04 \cdot 10^{-2}$	$2.43 \cdot 10^{-2}$
	NMSE	$5.43 \cdot 10^{-1}$	$2.89 \cdot 10^{-1}$	$2.53 \cdot 10^{-1}$	$3.04 \cdot 10^{-1}$	$3.20 \cdot 10^{-1}$	$2.81 \cdot 10^{-1}$	$3.30 \cdot 10^{-1}$	$3.95 \cdot 10^{-1}$
	REP	3.00	1.53	1.33	1.59	1.80	1.59	1.85	2.21
	PPMCC	0.85	0.95	0.96	0.95	0.95	0.96	0.94	0.92
	MAPE	2.45	0.61	0.60	1.16	0.70	0.85	1.03	1.25
RF	MAE	$2.13 \cdot 10^{-2}$	$1.52 \cdot 10^{-3}$	$1.84 \cdot 10^{-3}$	$6.59 \cdot 10^{-3}$	$1.19 \cdot 10^{-3}$	$4.02 \cdot 10^{-3}$	$4.34 \cdot 10^{-3}$	$4.69 \cdot 10^{-3}$
	MSE	$2.94 \cdot 10^{-2}$	$4.04 \cdot 10^{-3}$	$4.37 \cdot 10^{-3}$	$1.01 \cdot 10^{-2}$	$3.17 \cdot 10^{-3}$	$6.65 \cdot 10^{-3}$	$7.10 \cdot 10^{-3}$	$7.77 \cdot 10^{-3}$
	NMSE	$4.99 \cdot 10^{-1}$	$6.28 \cdot 10^{-2}$	$7.22 \cdot 10^{-2}$	$1.64 \cdot 10^{-1}$	$5.16 \cdot 10^{-2}$	$1.08 \cdot 10^{-1}$	$1.15 \cdot 10^{-1}$	$1.26 \cdot 10^{-1}$
	REP	2.66	$3.65 \cdot 10^{-1}$	$3.94 \cdot 10^{-1}$	$9.13 \cdot 10^{-1}$	$2.87 \cdot 10^{-1}$	$6.04 \cdot 10^{-1}$	$6.44 \cdot 10^{-1}$	$7.05 \cdot 10^{-1}$
	PPMCC	0.88	1.00	1.00	0.99	1.00	0.99	0.99	0.99
	MAPE	1.93	0.14	0.17	0.60	0.11	0.37	0.40	0.43
GP	MAE	$2.89 \cdot 10^{-2}$	$7.58 \cdot 10^{-3}$	$8.70 \cdot 10^{-3}$	$8.38 \cdot 10^{-3}$	$6.65 \cdot 10^{-3}$	$6.64 \cdot 10^{-3}$	$7.35 \cdot 10^{-3}$	$8.14 \cdot 10^{-3}$
	MSE	$3.86 \cdot 10^{-2}$	$1.39 \cdot 10^{-2}$	$1.65 \cdot 10^{-2}$	$1.39 \cdot 10^{-2}$	$1.38 \cdot 10^{-2}$	$1.22 \cdot 10^{-2}$	$1.37 \cdot 10^{-2}$	$1.51 \cdot 10^{-2}$
	NMSE	$6.33 \cdot 10^{-1}$	$2.38 \cdot 10^{-1}$	$2.81 \cdot 10^{-1}$	$2.36 \cdot 10^{-1}$	$2.24 \cdot 10^{-1}$	$1.95 \cdot 10^{-1}$	$2.22 \cdot 10^{-1}$	$2.44 \cdot 10^{-1}$
	REP	3.51	1.26	1.50	1.26	1.26	1.10	1.25	1.37
	PPMCC	0.86	0.97	0.96	0.98	0.98	0.98	0.98	0.97
	MAPE	2.66	0.69	0.79	0.76	0.61	0.61	0.67	0.74
SVR	MAE	$1.75 \cdot 10^{-2}$	$1.83 \cdot 10^{-3}$	$2.57 \cdot 10^{-3}$	$1.53 \cdot 10^{-3}$	$1.65 \cdot 10^{-3}$	$1.45 \cdot 10^{-3}$	$1.80 \cdot 10^{-3}$	$2.25 \cdot 10^{-3}$
	MSE	$2.81 \cdot 10^{-2}$	$4.41 \cdot 10^{-3}$	$6.07 \cdot 10^{-3}$	$3.29 \cdot 10^{-3}$	$4.60 \cdot 10^{-3}$	$3.70 \cdot 10^{-3}$	$4.02 \cdot 10^{-3}$	$5.21 \cdot 10^{-3}$
	NMSE	$4.78 \cdot 10^{-1}$	$7.05 \cdot 10^{-2}$	$1.05 \cdot 10^{-1}$	$5.46 \cdot 10^{-2}$	$7.21 \cdot 10^{-2}$	$5.95 \cdot 10^{-2}$	$6.62 \cdot 10^{-2}$	$8.34 \cdot 10^{-2}$
	REP	2.55	$3.82 \cdot 10^{-1}$	$5.46 \cdot 10^{-1}$	$2.91 \cdot 10^{-1}$	$4.01 \cdot 10^{-1}$	$3.34 \cdot 10^{-1}$	$3.93 \cdot 10^{-1}$	$4.87 \cdot 10^{-1}$
	PPMCC	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	1.54	0.17	0.23	0.14	0.14	0.14	0.15	0.20
DNN	MAE	$8.29 \cdot 10^{-3}$	$1.02 \cdot 10^{-3}$	$9.87 \cdot 10^{-4}$	$9.83 \cdot 10^{-4}$	$6.92 \cdot 10^{-4}$	$8.69 \cdot 10^{-4}$	$9.10 \cdot 10^{-4}$	$9.52 \cdot 10^{-4}$
	MSE	$1.31 \cdot 10^{-2}$	$1.75 \cdot 10^{-3}$	$1.54 \cdot 10^{-3}$	$1.59 \cdot 10^{-3}$	$1.27 \cdot 10^{-3}$	$1.58 \cdot 10^{-3}$	$1.49 \cdot 10^{-3}$	$1.76 \cdot 10^{-3}$
	NMSE	$2.18 \cdot 10^{-1}$	$2.82 \cdot 10^{-2}$	$2.66 \cdot 10^{-2}$	$2.49 \cdot 10^{-2}$	$2.20 \cdot 10^{-2}$	$2.66 \cdot 10^{-2}$	$2.52 \cdot 10^{-2}$	$2.88 \cdot 10^{-2}$
	REP	1.25	$1.48 \cdot 10^{-1}$	$1.50 \cdot 10^{-1}$	$1.42 \cdot 10^{-1}$	$1.21 \cdot 10^{-1}$	$1.43 \cdot 10^{-1}$	$1.46 \cdot 10^{-1}$	$1.62 \cdot 10^{-1}$
	PPMCC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.76	0.09	0.09	0.08	0.06	0.08	0.08	0.09
RFR	MAE	$1.76 \cdot 10^{-2}$	$1.33 \cdot 10^{-3}$	$1.63 \cdot 10^{-3}$	$5.75 \cdot 10^{-3}$	$1.02 \cdot 10^{-3}$	$3.45 \cdot 10^{-3}$	$3.71 \cdot 10^{-3}$	$3.95 \cdot 10^{-3}$
	MSE	$2.51 \cdot 10^{-2}$	$3.57 \cdot 10^{-3}$	$3.90 \cdot 10^{-3}$	$8.20 \cdot 10^{-3}$	$2.56 \cdot 10^{-3}$	$5.51 \cdot 10^{-3}$	$6.31 \cdot 10^{-3}$	$6.41 \cdot 10^{-3}$
	NMSE	$4.01 \cdot 10^{-1}$	$5.53 \cdot 10^{-2}$	$6.50 \cdot 10^{-2}$	$1.44 \cdot 10^{-1}$	$4.24 \cdot 10^{-2}$	$8.87 \cdot 10^{-2}$	$9.24 \cdot 10^{-2}$	$1.05 \cdot 10^{-1}$
	REP	2.30	$3.28 \cdot 10^{-1}$	$3.33 \cdot 10^{-1}$	$7.65 \cdot 10^{-1}$	$2.35 \cdot 10^{-1}$	$5.03 \cdot 10^{-1}$	$5.24 \cdot 10^{-1}$	$6.18 \cdot 10^{-1}$
	PPMCC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	1.55	0.11	0.14	0.51	0.10	0.31	0.33	0.37
SNN	MAE	$8.39 \cdot 10^{-3}$	$1.03 \cdot 10^{-3}$	$9.97 \cdot 10^{-4}$	$1.03 \cdot 10^{-3}$	$7.86 \cdot 10^{-4}$	$8.89 \cdot 10^{-4}$	$9.35 \cdot 10^{-4}$	$1.07 \cdot 10^{-3}$
	MSE	$1.45 \cdot 10^{-2}$	$1.89 \cdot 10^{-3}$	$1.61 \cdot 10^{-3}$	$1.67 \cdot 10^{-3}$	$1.35 \cdot 10^{-3}$	$1.64 \cdot 10^{-3}$	$1.74 \cdot 10^{-3}$	$1.84 \cdot 10^{-3}$
	NMSE	$2.48 \cdot 10^{-1}$	$2.88 \cdot 10^{-2}$	$2.83 \cdot 10^{-2}$	$2.65 \cdot 10^{-2}$	$2.25 \cdot 10^{-2}$	$2.65 \cdot 10^{-2}$	$2.61 \cdot 10^{-2}$	$2.98 \cdot 10^{-2}$
	REP	1.36	$1.65 \cdot 10^{-1}$	$1.44 \cdot 10^{-1}$	$1.51 \cdot 10^{-1}$	$1.19 \cdot 10^{-1}$	$1.49 \cdot 10^{-1}$	$1.48 \cdot 10^{-1}$	$1.58 \cdot 10^{-1}$
	PPMCC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.80	0.09	0.09	0.09	0.07	0.08	0.09	0.10

Table 10: REG-PROB: different indexes of performances (MAE, MSE, NMSE, REP, PPMCC, and MAPE) of the models learned with the different algorithms (DNN, SNN, ELM, SVR, KRLS, KNN, and GP) when n is the largest possible for the different problems $P1, \dots, P15$ and for the HLL NPS component.

	Problem	P2	P5	P8	P9	P11	P12	P14	P15
	Decayed Components	GTC	GT, GTC	GTC, HLL	GTC, PRP	GT, GTC, HLL	GT, GTC, PRP	GTC, HLL, PRP	GT, GTC, HLL, PRP
KRLS	MAE	$5.90 \cdot 10^{-3}$	$4.35 \cdot 10^{-4}$	$3.64 \cdot 10^{-4}$	$4.94 \cdot 10^{-4}$	$3.24 \cdot 10^{-4}$	$3.90 \cdot 10^{-4}$	$3.41 \cdot 10^{-4}$	$5.45 \cdot 10^{-4}$
	MSE	$9.46 \cdot 10^{-3}$	$8.45 \cdot 10^{-4}$	$8.33 \cdot 10^{-4}$	$9.67 \cdot 10^{-4}$	$6.18 \cdot 10^{-4}$	$6.88 \cdot 10^{-4}$	$5.91 \cdot 10^{-4}$	$8.41 \cdot 10^{-4}$
	NMSE	$6.22 \cdot 10^{-1}$	$5.62 \cdot 10^{-2}$	$5.43 \cdot 10^{-2}$	$6.47 \cdot 10^{-2}$	$3.99 \cdot 10^{-2}$	$4.45 \cdot 10^{-2}$	$3.84 \cdot 10^{-2}$	$5.45 \cdot 10^{-2}$
	REP	$9.71 \cdot 10^{-1}$	$8.67 \cdot 10^{-2}$	$8.55 \cdot 10^{-2}$	$9.93 \cdot 10^{-2}$	$6.34 \cdot 10^{-2}$	$7.06 \cdot 10^{-2}$	$6.06 \cdot 10^{-2}$	$8.62 \cdot 10^{-2}$
	PPMCC	0.81	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.60	0.04	0.04	0.05	0.03	0.04	0.04	0.06
ELM	MAE	$7.17 \cdot 10^{-4}$	$9.48 \cdot 10^{-5}$	$8.03 \cdot 10^{-5}$	$1.47 \cdot 10^{-4}$	$9.20 \cdot 10^{-5}$	$1.28 \cdot 10^{-4}$	$1.52 \cdot 10^{-4}$	$2.38 \cdot 10^{-4}$
	MSE	$1.42 \cdot 10^{-3}$	$1.35 \cdot 10^{-4}$	$1.11 \cdot 10^{-4}$	$2.13 \cdot 10^{-4}$	$1.33 \cdot 10^{-4}$	$1.91 \cdot 10^{-4}$	$2.90 \cdot 10^{-4}$	$4.42 \cdot 10^{-4}$
	NMSE	$9.69 \cdot 10^{-2}$	$9.27 \cdot 10^{-3}$	$7.30 \cdot 10^{-3}$	$1.36 \cdot 10^{-2}$	$8.62 \cdot 10^{-3}$	$1.23 \cdot 10^{-2}$	$1.89 \cdot 10^{-2}$	$2.86 \cdot 10^{-2}$
	REP	$1.46 \cdot 10^{-1}$	$1.39 \cdot 10^{-2}$	$1.13 \cdot 10^{-2}$	$2.19 \cdot 10^{-2}$	$1.36 \cdot 10^{-2}$	$1.96 \cdot 10^{-2}$	$2.98 \cdot 10^{-2}$	$4.53 \cdot 10^{-2}$
	PPMCC	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.07	0.01	0.01	0.02	0.01	0.01	0.02	0.02
KNN	MAE	$8.61 \cdot 10^{-3}$	$1.56 \cdot 10^{-3}$	$2.73 \cdot 10^{-3}$	$3.74 \cdot 10^{-3}$	$2.64 \cdot 10^{-3}$	$3.14 \cdot 10^{-3}$	$3.54 \cdot 10^{-3}$	$4.80 \cdot 10^{-3}$
	MSE	$1.20 \cdot 10^{-2}$	$3.12 \cdot 10^{-3}$	$3.18 \cdot 10^{-3}$	$4.51 \cdot 10^{-3}$	$3.77 \cdot 10^{-3}$	$4.35 \cdot 10^{-3}$	$4.65 \cdot 10^{-3}$	$6.32 \cdot 10^{-3}$
	NMSE	$7.91 \cdot 10^{-1}$	$2.13 \cdot 10^{-1}$	$2.13 \cdot 10^{-1}$	$3.02 \cdot 10^{-1}$	$2.44 \cdot 10^{-1}$	$2.82 \cdot 10^{-1}$	$3.03 \cdot 10^{-1}$	$4.10 \cdot 10^{-1}$
	REP	1.23	$3.20 \cdot 10^{-1}$	$3.26 \cdot 10^{-1}$	$4.64 \cdot 10^{-1}$	$3.87 \cdot 10^{-1}$	$4.46 \cdot 10^{-1}$	$4.78 \cdot 10^{-1}$	$6.49 \cdot 10^{-1}$
	PPMCC	0.58	0.98	0.98	0.96	0.97	0.96	0.95	0.91
	MAPE	0.88	0.16	0.28	0.38	0.27	0.32	0.36	0.49
RF	MAE	$8.47 \cdot 10^{-3}$	$1.17 \cdot 10^{-3}$	$1.03 \cdot 10^{-3}$	$1.68 \cdot 10^{-3}$	$1.44 \cdot 10^{-3}$	$1.96 \cdot 10^{-3}$	$1.15 \cdot 10^{-3}$	$2.38 \cdot 10^{-3}$
	MSE	$1.15 \cdot 10^{-2}$	$1.71 \cdot 10^{-3}$	$1.49 \cdot 10^{-3}$	$2.44 \cdot 10^{-3}$	$2.09 \cdot 10^{-3}$	$2.68 \cdot 10^{-3}$	$1.80 \cdot 10^{-3}$	$3.19 \cdot 10^{-3}$
	NMSE	$7.86 \cdot 10^{-1}$	$1.15 \cdot 10^{-1}$	$1.00 \cdot 10^{-1}$	$1.62 \cdot 10^{-1}$	$1.37 \cdot 10^{-1}$	$1.73 \cdot 10^{-1}$	$1.17 \cdot 10^{-1}$	$2.08 \cdot 10^{-1}$
	REP	1.18	$1.75 \cdot 10^{-1}$	$1.53 \cdot 10^{-1}$	$2.51 \cdot 10^{-1}$	$2.15 \cdot 10^{-1}$	$2.75 \cdot 10^{-1}$	$1.85 \cdot 10^{-1}$	$3.27 \cdot 10^{-1}$
	PPMCC	0.65	0.99	1.00	0.99	0.99	0.99	0.99	0.98
	MAPE	0.87	0.12	0.11	0.17	0.15	0.20	0.12	0.24
GP	MAE	$7.98 \cdot 10^{-3}$	$1.84 \cdot 10^{-3}$	$1.53 \cdot 10^{-3}$	$1.99 \cdot 10^{-3}$	$1.50 \cdot 10^{-3}$	$1.66 \cdot 10^{-3}$	$1.66 \cdot 10^{-3}$	$2.24 \cdot 10^{-3}$
	MSE	$1.06 \cdot 10^{-2}$	$2.75 \cdot 10^{-3}$	$2.47 \cdot 10^{-3}$	$3.13 \cdot 10^{-3}$	$2.28 \cdot 10^{-3}$	$2.44 \cdot 10^{-3}$	$2.48 \cdot 10^{-3}$	$3.06 \cdot 10^{-3}$
	NMSE	$6.93 \cdot 10^{-1}$	$1.82 \cdot 10^{-1}$	$1.63 \cdot 10^{-1}$	$2.09 \cdot 10^{-1}$	$1.47 \cdot 10^{-1}$	$1.58 \cdot 10^{-1}$	$1.61 \cdot 10^{-1}$	$1.98 \cdot 10^{-1}$
	REP	1.08	$2.82 \cdot 10^{-1}$	$2.53 \cdot 10^{-1}$	$3.22 \cdot 10^{-1}$	$2.33 \cdot 10^{-1}$	$2.51 \cdot 10^{-1}$	$2.54 \cdot 10^{-1}$	$3.14 \cdot 10^{-1}$
	PPMCC	0.80	0.99	0.99	0.98	0.99	0.99	0.99	0.99
	MAPE	0.82	0.19	0.16	0.20	0.15	0.17	0.17	0.23
SVR	MAE	$5.44 \cdot 10^{-3}$	$4.35 \cdot 10^{-4}$	$3.40 \cdot 10^{-4}$	$4.55 \cdot 10^{-4}$	$2.99 \cdot 10^{-4}$	$3.62 \cdot 10^{-4}$	$3.11 \cdot 10^{-4}$	$4.96 \cdot 10^{-4}$
	MSE	$9.07 \cdot 10^{-3}$	$7.64 \cdot 10^{-4}$	$8.25 \cdot 10^{-4}$	$9.35 \cdot 10^{-4}$	$5.89 \cdot 10^{-4}$	$6.75 \cdot 10^{-4}$	$5.49 \cdot 10^{-4}$	$8.35 \cdot 10^{-4}$
	NMSE	$6.07 \cdot 10^{-1}$	$5.35 \cdot 10^{-2}$	$5.00 \cdot 10^{-2}$	$6.19 \cdot 10^{-2}$	$3.99 \cdot 10^{-2}$	$4.21 \cdot 10^{-2}$	$3.62 \cdot 10^{-2}$	$5.28 \cdot 10^{-2}$
	REP	$9.63 \cdot 10^{-1}$	$8.45 \cdot 10^{-2}$	$8.43 \cdot 10^{-2}$	$9.33 \cdot 10^{-2}$	$5.89 \cdot 10^{-2}$	$6.36 \cdot 10^{-2}$	$5.50 \cdot 10^{-2}$	$7.81 \cdot 10^{-2}$
	PPMCC	0.85	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.55	0.04	0.04	0.05	0.03	0.04	0.03	0.05
DNN	MAE	$6.65 \cdot 10^{-4}$	$8.19 \cdot 10^{-5}$	$7.03 \cdot 10^{-5}$	$1.34 \cdot 10^{-4}$	$7.94 \cdot 10^{-5}$	$1.10 \cdot 10^{-4}$	$1.33 \cdot 10^{-4}$	$2.17 \cdot 10^{-4}$
	MSE	$1.34 \cdot 10^{-3}$	$1.16 \cdot 10^{-4}$	$1.04 \cdot 10^{-4}$	$1.92 \cdot 10^{-4}$	$1.17 \cdot 10^{-4}$	$1.78 \cdot 10^{-4}$	$2.56 \cdot 10^{-4}$	$4.08 \cdot 10^{-4}$
	NMSE	$8.93 \cdot 10^{-2}$	$8.35 \cdot 10^{-3}$	$6.49 \cdot 10^{-3}$	$1.24 \cdot 10^{-2}$	$8.10 \cdot 10^{-3}$	$1.06 \cdot 10^{-2}$	$1.61 \cdot 10^{-2}$	$2.51 \cdot 10^{-2}$
	REP	$1.25 \cdot 10^{-1}$	$1.26 \cdot 10^{-2}$	$1.06 \cdot 10^{-2}$	$2.04 \cdot 10^{-2}$	$1.18 \cdot 10^{-2}$	$1.67 \cdot 10^{-2}$	$2.68 \cdot 10^{-2}$	$3.97 \cdot 10^{-2}$
	PPMCC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.06	0.01	0.01	0.01	0.01	0.01	0.01	0.02
RFR	MAE	$7.24 \cdot 10^{-3}$	$1.02 \cdot 10^{-3}$	$9.25 \cdot 10^{-4}$	$1.35 \cdot 10^{-3}$	$1.18 \cdot 10^{-3}$	$1.70 \cdot 10^{-3}$	$9.47 \cdot 10^{-4}$	$2.13 \cdot 10^{-3}$
	MSE	$9.94 \cdot 10^{-3}$	$1.41 \cdot 10^{-3}$	$1.27 \cdot 10^{-3}$	$1.96 \cdot 10^{-3}$	$1.77 \cdot 10^{-3}$	$2.34 \cdot 10^{-3}$	$1.46 \cdot 10^{-3}$	$2.65 \cdot 10^{-3}$
	NMSE	$6.93 \cdot 10^{-1}$	$9.60 \cdot 10^{-2}$	$8.53 \cdot 10^{-2}$	$1.38 \cdot 10^{-1}$	$1.12 \cdot 10^{-1}$	$1.42 \cdot 10^{-1}$	$9.79 \cdot 10^{-2}$	$1.78 \cdot 10^{-1}$
	REP	$9.78 \cdot 10^{-1}$	$1.43 \cdot 10^{-1}$	$1.33 \cdot 10^{-1}$	$2.20 \cdot 10^{-1}$	$1.72 \cdot 10^{-1}$	$2.39 \cdot 10^{-1}$	$1.50 \cdot 10^{-1}$	$2.74 \cdot 10^{-1}$
	PPMCC	0.74	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.78	0.10	0.09	0.15	0.13	0.18	0.11	0.21
SNN	MAE	$6.56 \cdot 10^{-4}$	$8.74 \cdot 10^{-5}$	$7.60 \cdot 10^{-5}$	$1.34 \cdot 10^{-4}$	$8.59 \cdot 10^{-5}$	$1.18 \cdot 10^{-4}$	$1.39 \cdot 10^{-4}$	$2.20 \cdot 10^{-4}$
	MSE	$1.29 \cdot 10^{-3}$	$1.32 \cdot 10^{-4}$	$1.06 \cdot 10^{-4}$	$1.99 \cdot 10^{-4}$	$1.22 \cdot 10^{-4}$	$1.82 \cdot 10^{-4}$	$2.62 \cdot 10^{-4}$	$4.17 \cdot 10^{-4}$
	NMSE	$9.00 \cdot 10^{-2}$	$8.89 \cdot 10^{-3}$	$6.64 \cdot 10^{-3}$	$1.26 \cdot 10^{-2}$	$7.76 \cdot 10^{-3}$	$1.14 \cdot 10^{-2}$	$1.79 \cdot 10^{-2}$	$2.66 \cdot 10^{-2}$
	REP	$1.37 \cdot 10^{-1}$	$1.35 \cdot 10^{-2}$	$1.13 \cdot 10^{-2}$	$2.01 \cdot 10^{-2}$	$1.24 \cdot 10^{-2}$	$1.91 \cdot 10^{-2}$	$2.70 \cdot 10^{-2}$	$4.22 \cdot 10^{-2}$
	PPMCC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.07	0.01	0.01	0.01	0.01	0.01	0.01	0.02

Table 11: REG-PROB: different indexes of performances (MAE, MSE, NMSE, REP, PPMCC, and MAPE) of the models learned with the different algorithms (DNN, SNN, ELM, SVR, KRLS, KNN, and GP) when n is the largest possible for the different problems $P1, \dots, P15$ and for the GTC NPS component.

	Problem	P1	P5	P6	P7	P11	P12	P14	P15
	Decayed Components	GT	GT, GTC	GT, HLL	GT, PRP	GT, GTC, HLL	GT, GTC, PRP	GTC, HLL, PRP	GT, GTC, HLL, PRP
KRLS	MAE	$2.41 \cdot 10^{-3}$	$4.89 \cdot 10^{-4}$	$2.38 \cdot 10^{-4}$	$3.40 \cdot 10^{-4}$	$3.19 \cdot 10^{-4}$	$3.93 \cdot 10^{-4}$	$3.17 \cdot 10^{-4}$	$5.28 \cdot 10^{-4}$
	MSE	$3.85 \cdot 10^{-3}$	$1.07 \cdot 10^{-3}$	$5.68 \cdot 10^{-4}$	$7.23 \cdot 10^{-4}$	$6.11 \cdot 10^{-4}$	$7.67 \cdot 10^{-4}$	$6.12 \cdot 10^{-4}$	$8.96 \cdot 10^{-4}$
	NMSE	$5.05 \cdot 10^{-1}$	$1.44 \cdot 10^{-1}$	$7.43 \cdot 10^{-2}$	$1.03 \cdot 10^{-1}$	$7.93 \cdot 10^{-2}$	$9.95 \cdot 10^{-2}$	$7.94 \cdot 10^{-2}$	$1.16 \cdot 10^{-1}$
	REP	$3.90 \cdot 10^{-1}$	$1.08 \cdot 10^{-1}$	$5.75 \cdot 10^{-2}$	$7.32 \cdot 10^{-2}$	$6.18 \cdot 10^{-2}$	$7.77 \cdot 10^{-2}$	$6.20 \cdot 10^{-2}$	$9.07 \cdot 10^{-2}$
	PPMCC	0.88	0.99	1.00	1.00	1.00	1.00	1.00	0.99
	MAPE	0.24	0.05	0.02	0.03	0.03	0.04	0.03	0.05
ELM	MAE	$1.24 \cdot 10^{-3}$	$7.34 \cdot 10^{-5}$	$5.59 \cdot 10^{-5}$	$8.89 \cdot 10^{-5}$	$6.39 \cdot 10^{-5}$	$7.67 \cdot 10^{-5}$	$6.47 \cdot 10^{-5}$	$8.54 \cdot 10^{-5}$
	MSE	$2.02 \cdot 10^{-3}$	$1.10 \cdot 10^{-4}$	$8.21 \cdot 10^{-5}$	$1.35 \cdot 10^{-4}$	$1.04 \cdot 10^{-4}$	$1.26 \cdot 10^{-4}$	$1.02 \cdot 10^{-4}$	$1.34 \cdot 10^{-4}$
	NMSE	$2.70 \cdot 10^{-1}$	$1.48 \cdot 10^{-2}$	$1.06 \cdot 10^{-2}$	$1.80 \cdot 10^{-2}$	$1.34 \cdot 10^{-2}$	$1.63 \cdot 10^{-2}$	$1.32 \cdot 10^{-2}$	$1.73 \cdot 10^{-2}$
	REP	$2.04 \cdot 10^{-1}$	$1.11 \cdot 10^{-2}$	$8.31 \cdot 10^{-3}$	$1.36 \cdot 10^{-2}$	$1.05 \cdot 10^{-2}$	$1.28 \cdot 10^{-2}$	$1.03 \cdot 10^{-2}$	$1.36 \cdot 10^{-2}$
	PPMCC	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.13	0.01	0.01	0.01	0.01	0.01	0.01	0.01
KNN	MAE	$3.64 \cdot 10^{-3}$	$2.17 \cdot 10^{-3}$	$1.78 \cdot 10^{-3}$	$2.19 \cdot 10^{-3}$	$2.38 \cdot 10^{-3}$	$2.52 \cdot 10^{-3}$	$2.28 \cdot 10^{-3}$	$3.57 \cdot 10^{-3}$
	MSE	$4.71 \cdot 10^{-3}$	$2.69 \cdot 10^{-3}$	$2.15 \cdot 10^{-3}$	$2.67 \cdot 10^{-3}$	$3.06 \cdot 10^{-3}$	$3.38 \cdot 10^{-3}$	$3.06 \cdot 10^{-3}$	$4.72 \cdot 10^{-3}$
	NMSE	$6.24 \cdot 10^{-1}$	$3.62 \cdot 10^{-1}$	$2.85 \cdot 10^{-1}$	$3.80 \cdot 10^{-1}$	$3.98 \cdot 10^{-1}$	$4.39 \cdot 10^{-1}$	$3.97 \cdot 10^{-1}$	$6.11 \cdot 10^{-1}$
	REP	$4.77 \cdot 10^{-1}$	$2.72 \cdot 10^{-1}$	$2.17 \cdot 10^{-1}$	$2.70 \cdot 10^{-1}$	$3.10 \cdot 10^{-1}$	$3.43 \cdot 10^{-1}$	$3.10 \cdot 10^{-1}$	$4.78 \cdot 10^{-1}$
	PPMCC	0.79	0.93	0.96	0.94	0.92	0.90	0.92	0.80
	MAPE	0.37	0.22	0.18	0.22	0.24	0.25	0.23	0.36
RF	MAE	$4.58 \cdot 10^{-3}$	$1.11 \cdot 10^{-3}$	$7.15 \cdot 10^{-4}$	$1.26 \cdot 10^{-3}$	$1.12 \cdot 10^{-3}$	$1.58 \cdot 10^{-3}$	$8.10 \cdot 10^{-4}$	$1.90 \cdot 10^{-3}$
	MSE	$5.89 \cdot 10^{-3}$	$1.90 \cdot 10^{-3}$	$9.93 \cdot 10^{-4}$	$1.81 \cdot 10^{-3}$	$1.59 \cdot 10^{-3}$	$2.20 \cdot 10^{-3}$	$1.29 \cdot 10^{-3}$	$2.58 \cdot 10^{-3}$
	NMSE	$7.61 \cdot 10^{-1}$	$2.49 \cdot 10^{-1}$	$1.28 \cdot 10^{-1}$	$2.38 \cdot 10^{-1}$	$2.07 \cdot 10^{-1}$	$2.84 \cdot 10^{-1}$	$1.68 \cdot 10^{-1}$	$3.31 \cdot 10^{-1}$
	REP	$5.96 \cdot 10^{-1}$	$1.92 \cdot 10^{-1}$	$1.01 \cdot 10^{-1}$	$1.83 \cdot 10^{-1}$	$1.61 \cdot 10^{-1}$	$2.23 \cdot 10^{-1}$	$1.31 \cdot 10^{-1}$	$2.61 \cdot 10^{-1}$
	PPMCC	0.67	0.97	0.99	0.98	0.98	0.97	0.99	0.96
	MAPE	0.46	0.11	0.07	0.13	0.11	0.16	0.08	0.19
GP	MAE	$3.95 \cdot 10^{-3}$	$1.34 \cdot 10^{-3}$	$9.05 \cdot 10^{-4}$	$1.09 \cdot 10^{-3}$	$1.12 \cdot 10^{-3}$	$1.27 \cdot 10^{-3}$	$1.07 \cdot 10^{-3}$	$1.80 \cdot 10^{-3}$
	MSE	$5.08 \cdot 10^{-3}$	$2.28 \cdot 10^{-3}$	$1.48 \cdot 10^{-3}$	$1.76 \cdot 10^{-3}$	$1.76 \cdot 10^{-3}$	$2.03 \cdot 10^{-3}$	$1.65 \cdot 10^{-3}$	$2.55 \cdot 10^{-3}$
	NMSE	$6.65 \cdot 10^{-1}$	$3.06 \cdot 10^{-1}$	$1.95 \cdot 10^{-1}$	$2.50 \cdot 10^{-1}$	$2.29 \cdot 10^{-1}$	$2.64 \cdot 10^{-1}$	$2.14 \cdot 10^{-1}$	$3.29 \cdot 10^{-1}$
	REP	$5.15 \cdot 10^{-1}$	$2.31 \cdot 10^{-1}$	$1.50 \cdot 10^{-1}$	$1.78 \cdot 10^{-1}$	$1.78 \cdot 10^{-1}$	$2.06 \cdot 10^{-1}$	$1.67 \cdot 10^{-1}$	$2.58 \cdot 10^{-1}$
	PPMCC	0.84	0.96	0.99	0.98	0.98	0.97	0.98	0.96
	MAPE	0.40	0.14	0.09	0.11	0.11	0.13	0.11	0.18
SVR	MAE	$2.37 \cdot 10^{-3}$	$4.60 \cdot 10^{-4}$	$2.34 \cdot 10^{-4}$	$3.12 \cdot 10^{-4}$	$3.18 \cdot 10^{-4}$	$3.91 \cdot 10^{-4}$	$3.05 \cdot 10^{-4}$	$4.86 \cdot 10^{-4}$
	MSE	$3.81 \cdot 10^{-3}$	$1.05 \cdot 10^{-3}$	$5.41 \cdot 10^{-4}$	$7.09 \cdot 10^{-4}$	$5.66 \cdot 10^{-4}$	$7.51 \cdot 10^{-4}$	$5.93 \cdot 10^{-4}$	$8.09 \cdot 10^{-4}$
	NMSE	$4.74 \cdot 10^{-1}$	$1.33 \cdot 10^{-1}$	$6.93 \cdot 10^{-2}$	$9.55 \cdot 10^{-2}$	$7.78 \cdot 10^{-2}$	$9.15 \cdot 10^{-2}$	$7.57 \cdot 10^{-2}$	$1.16 \cdot 10^{-1}$
	REP	$3.82 \cdot 10^{-1}$	$1.04 \cdot 10^{-1}$	$5.66 \cdot 10^{-2}$	$6.72 \cdot 10^{-2}$	$5.93 \cdot 10^{-2}$	$7.28 \cdot 10^{-2}$	$6.01 \cdot 10^{-2}$	$8.82 \cdot 10^{-2}$
	PPMCC	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.23	0.05	0.02	0.03	0.03	0.04	0.03	0.05
DNN	MAE	$1.10 \cdot 10^{-3}$	$6.55 \cdot 10^{-5}$	$5.29 \cdot 10^{-5}$	$7.85 \cdot 10^{-5}$	$5.89 \cdot 10^{-5}$	$6.54 \cdot 10^{-5}$	$5.71 \cdot 10^{-5}$	$7.97 \cdot 10^{-5}$
	MSE	$1.79 \cdot 10^{-3}$	$1.01 \cdot 10^{-4}$	$7.70 \cdot 10^{-5}$	$1.22 \cdot 10^{-4}$	$9.10 \cdot 10^{-5}$	$1.18 \cdot 10^{-4}$	$9.24 \cdot 10^{-5}$	$1.16 \cdot 10^{-4}$
	NMSE	$2.50 \cdot 10^{-1}$	$1.35 \cdot 10^{-2}$	$9.63 \cdot 10^{-3}$	$1.59 \cdot 10^{-2}$	$1.21 \cdot 10^{-2}$	$1.43 \cdot 10^{-2}$	$1.25 \cdot 10^{-2}$	$1.63 \cdot 10^{-2}$
	REP	$1.84 \cdot 10^{-1}$	$9.99 \cdot 10^{-3}$	$7.67 \cdot 10^{-3}$	$1.17 \cdot 10^{-2}$	$8.97 \cdot 10^{-3}$	$1.16 \cdot 10^{-2}$	$9.20 \cdot 10^{-3}$	$1.25 \cdot 10^{-2}$
	PPMCC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.12	0.01	0.01	0.01	0.01	0.01	0.01	0.01
RFR	MAE	$3.85 \cdot 10^{-3}$	$9.39 \cdot 10^{-4}$	$6.10 \cdot 10^{-4}$	$1.07 \cdot 10^{-3}$	$9.39 \cdot 10^{-4}$	$1.42 \cdot 10^{-3}$	$6.73 \cdot 10^{-4}$	$1.62 \cdot 10^{-3}$
	MSE	$4.84 \cdot 10^{-3}$	$1.60 \cdot 10^{-3}$	$8.30 \cdot 10^{-4}$	$1.63 \cdot 10^{-3}$	$1.30 \cdot 10^{-3}$	$1.84 \cdot 10^{-3}$	$1.04 \cdot 10^{-3}$	$2.26 \cdot 10^{-3}$
	NMSE	$6.30 \cdot 10^{-1}$	$2.13 \cdot 10^{-1}$	$1.10 \cdot 10^{-1}$	$2.13 \cdot 10^{-1}$	$1.72 \cdot 10^{-1}$	$2.44 \cdot 10^{-1}$	$1.44 \cdot 10^{-1}$	$2.72 \cdot 10^{-1}$
	REP	$4.89 \cdot 10^{-1}$	$1.65 \cdot 10^{-1}$	$8.27 \cdot 10^{-2}$	$1.59 \cdot 10^{-1}$	$1.33 \cdot 10^{-1}$	$1.96 \cdot 10^{-1}$	$1.05 \cdot 10^{-1}$	$2.13 \cdot 10^{-1}$
	PPMCC	0.79	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.38	0.10	0.06	0.10	0.09	0.13	0.07	0.16
SNN	MAE	$1.15 \cdot 10^{-3}$	$6.96 \cdot 10^{-5}$	$5.57 \cdot 10^{-5}$	$8.50 \cdot 10^{-5}$	$6.10 \cdot 10^{-5}$	$7.60 \cdot 10^{-5}$	$6.17 \cdot 10^{-5}$	$7.79 \cdot 10^{-5}$
	MSE	$1.98 \cdot 10^{-3}$	$1.02 \cdot 10^{-4}$	$7.54 \cdot 10^{-5}$	$1.23 \cdot 10^{-4}$	$9.90 \cdot 10^{-5}$	$1.17 \cdot 10^{-4}$	$9.60 \cdot 10^{-5}$	$1.26 \cdot 10^{-4}$
	NMSE	$2.45 \cdot 10^{-1}$	$1.43 \cdot 10^{-2}$	$9.84 \cdot 10^{-3}$	$1.77 \cdot 10^{-2}$	$1.24 \cdot 10^{-2}$	$1.47 \cdot 10^{-2}$	$1.30 \cdot 10^{-2}$	$1.65 \cdot 10^{-2}$
	REP	$2.04 \cdot 10^{-1}$	$1.04 \cdot 10^{-2}$	$7.74 \cdot 10^{-3}$	$1.34 \cdot 10^{-2}$	$1.01 \cdot 10^{-2}$	$1.18 \cdot 10^{-2}$	$9.78 \cdot 10^{-3}$	$1.35 \cdot 10^{-2}$
	PPMCC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MAPE	0.12	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Table 12: REG-PROB: different indexes of performances (MAE, MSE, NMSE, REP, PPMCC, and MAPE) of the models learned with the different algorithms (DNN, SNN, ELM, SVR, KRLS, KNN, and GP) when n is the largest possible for the different problems $P1, \dots, P15$ and for the GT NPS component.

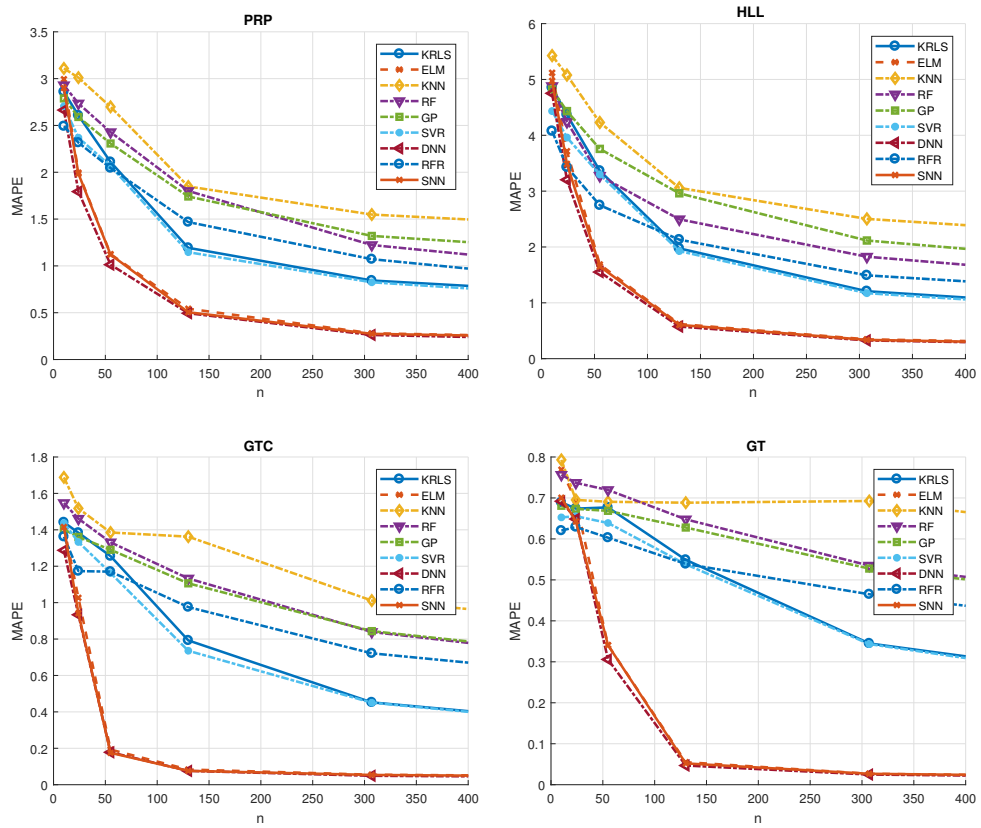


Figure 7: REG-PROB: MAPE of the models learned with the different algorithms (DNN, SNN, ELM, SVR, KRLS, KNN, and GP) when varying n for $P15$ and for the four main NPS components.

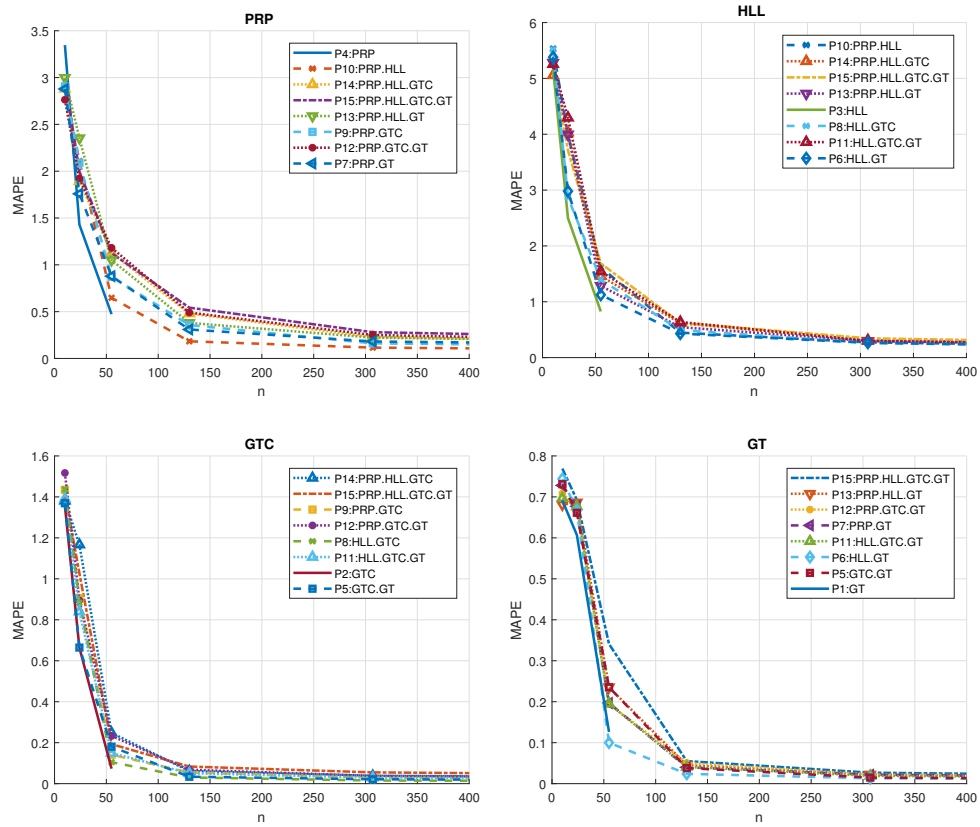


Figure 8: REG-PROB: MAPE of the models learned with DNN when varying n for the different problems P_1, \dots, P_{15} and for the four main NPS components.

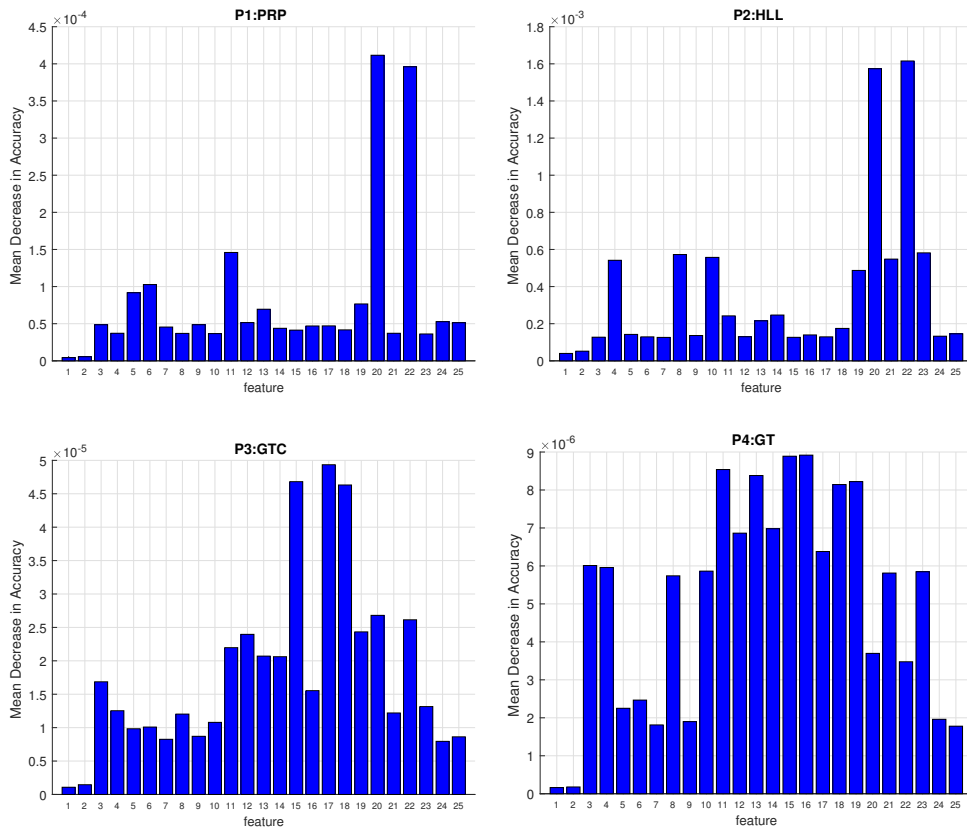


Figure 9: REG-PROB: FS performed with RF for the four main NPS components for problem $P1$, $P2$, $P3$, and $P4$

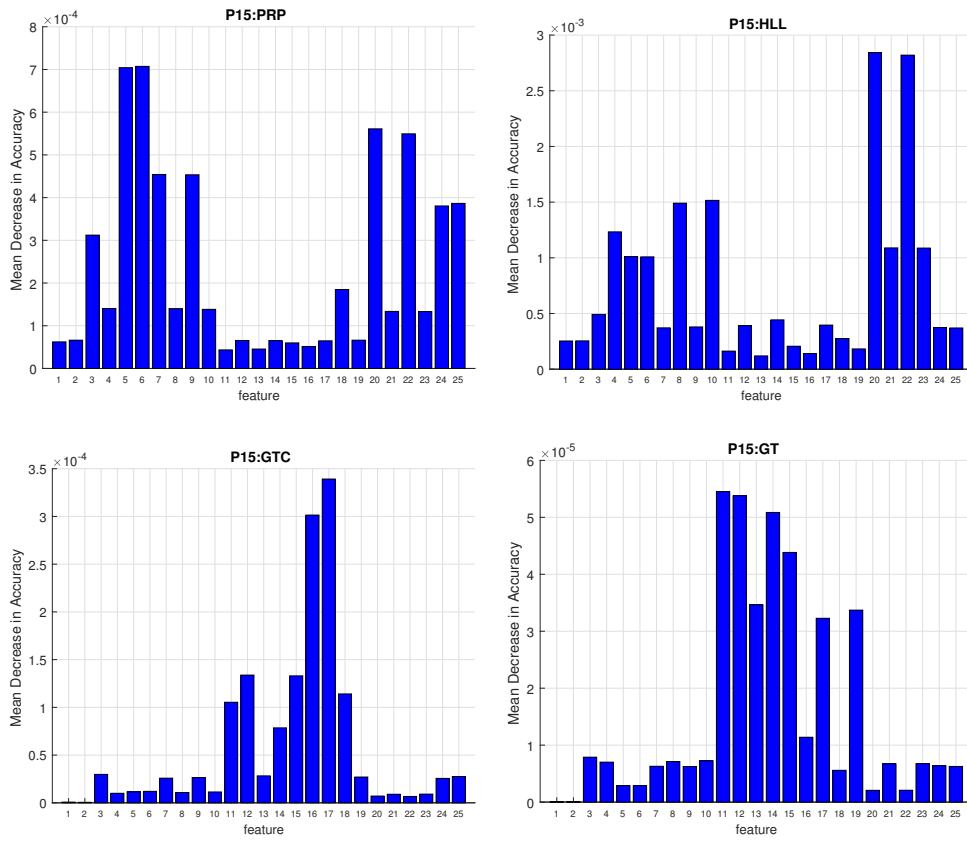


Figure 10: REG-PROB: FS performed with RF for the four main NPS components for problem *P15*.

- similarly to the REG-PROB, the larger is n the better performances are achieved by the learned models (see Table 13 and Figure 11) and the models learned with ELM, SNN, and especially DNN generally show the best performances (see Tables 14 and Figure 11);
- as expected, in order to achieve a reasonable AMR a smaller number of samples is needed with respect to the REG-PROB making this regression-based approach not feasible in practice.

Finally in Figure 13 the FS phase is reported, performed for $P15$. In particular, for each problem and each feature (see Table 3) the mean decrease in accuracy as described in Section 4.3 is reported.

From Figure 13 it is possible to observe that, oppositely from the REG-PROB, several features are always necessary to forecast the decay state of each component. This result indicates that, from a data driven perspective, the decay state of each component influences different phases of the NPS behavior. Figure 13 clearly shows that the interaction between the main components cannot be easily modeled with a physical approach, considering the large number of variables that affect the final behavior of each component. Instead, DDMs, by making use of these variables, can outperform physical models as they have the capability to take into account all the available sensors measurements to build effective and accurate predictors as reported in Table 14 and Figure 11.

5.3. ANOMALY-PROB

In this section the results on the ANOMALY-PROB are reported. In Table 15, respectively for PRP, HLL, GTC, and GT, the AMR of the models learned with the different algorithms (OCSVM and GKNN) is reported, when the number of unlabeled samples in the learning set is $l = 4000$ and when varying the number of labeled samples in the validation set $v \in \{10, 20, 30\}$ (half positively and half negatively labeled) for $P15$. In Table 16, respectively for PRP, HLL, GTC, and GT, the AMR of the models learned with the different algorithms is reported, when $v = 30$ and when $l \in \{1500, 2000, 3000, 4000\}$ for $P15$. In Table 17, respectively for PRP, HLL, GTC, and GT, the different indexes of performances (AMR, TP, TN, FP, and FN) of the models learned with the different algorithms are reported when $n = 4000$ and $v = 30$ for $P15$.

From the tables it is possible to observe that:

- both OCSVM and GKNN perform quite well on the problem and there is no clear winner;

	n	PRP	HLL	GTC	GT
KRLS	10	0.50±0.06	0.48±0.06	0.46±0.10	0.44±0.09
	24	0.43±0.10	0.43±0.10	0.43±0.07	0.43±0.07
	55	0.29±0.06	0.25±0.07	0.37±0.06	0.42±0.08
	130	0.17±0.05	0.17±0.06	0.25±0.12	0.32±0.09
	307	0.13±0.02	0.11±0.02	0.13±0.03	0.19±0.03
	722	0.10±0.02	0.07±0.02	0.08±0.03	0.12±0.03
	1700	0.06±0.01	0.04±0.02	0.05±0.02	0.07±0.02
	4000	0.03±0.01	0.02±0.01	0.04±0.02	0.05±0.01
ELM	10	0.40±0.22	0.44±0.13	0.46±0.16	0.45±0.09
	24	0.28±0.14	0.31±0.18	0.31±0.20	0.42±0.16
	55	0.19±0.08	0.23±0.09	0.16±0.07	0.22±0.07
	130	0.12±0.05	0.11±0.05	0.08±0.03	0.12±0.03
	307	0.09±0.04	0.06±0.03	0.06±0.02	0.07±0.01
	722	0.05±0.04	0.04±0.02	0.05±0.02	0.05±0.02
	1700	0.03±0.01	0.03±0.01	0.03±0.01	0.03±0.01
	4000	0.01±0.01	0.01±0.01	0.01±0.01	0.02±0.02
KNN	10	0.50±0.07	0.47±0.06	0.46±0.10	0.43±0.08
	24	0.46±0.08	0.43±0.15	0.43±0.08	0.43±0.07
	55	0.29±0.07	0.28±0.12	0.39±0.05	0.42±0.08
	130	0.21±0.05	0.21±0.06	0.30±0.03	0.40±0.07
	307	0.17±0.02	0.15±0.02	0.24±0.04	0.33±0.04
	722	0.15±0.03	0.12±0.02	0.18±0.03	0.26±0.02
	1700	0.13±0.02	0.10±0.02	0.13±0.02	0.20±0.02
	4000	0.09±0.01	0.07±0.01	0.10±0.02	0.15±0.02
RF	10	0.47±0.07	0.43±0.12	0.47±0.09	0.47±0.07
	24	0.40±0.12	0.36±0.15	0.45±0.07	0.45±0.04
	55	0.31±0.08	0.25±0.06	0.41±0.04	0.45±0.03
	130	0.19±0.04	0.19±0.05	0.31±0.06	0.38±0.04
	307	0.13±0.04	0.12±0.04	0.21±0.05	0.30±0.04
	722	0.07±0.02	0.07±0.02	0.13±0.01	0.21±0.03
	1700	0.04±0.02	0.04±0.01	0.09±0.02	0.13±0.02
	4000	0.02±0.01	0.02±0.01	0.06±0.02	0.08±0.02
GP	10	0.49±0.05	0.47±0.07	0.46±0.07	0.44±0.05
	24	0.42±0.11	0.40±0.11	0.42±0.07	0.45±0.09
	55	0.29±0.05	0.27±0.05	0.37±0.04	0.44±0.07
	130	0.20±0.05	0.20±0.03	0.30±0.03	0.39±0.08
	307	0.17±0.02	0.14±0.03	0.22±0.02	0.32±0.06
	722	0.13±0.03	0.11±0.03	0.15±0.03	0.24±0.04
	1700	0.11±0.02	0.07±0.02	0.10±0.03	0.17±0.03
	4000	0.08±0.01	0.04±0.01	0.07±0.02	0.11±0.02
SVM	10	0.46±0.06	0.46±0.07	0.44±0.11	0.43±0.10
	24	0.40±0.11	0.41±0.11	0.41±0.07	0.40±0.08
	55	0.27±0.06	0.24±0.07	0.35±0.07	0.41±0.09
	130	0.17±0.06	0.17±0.07	0.23±0.14	0.31±0.10
	307	0.13±0.03	0.11±0.02	0.12±0.03	0.18±0.04
	722	0.10±0.02	0.07±0.02	0.07±0.03	0.11±0.03
	1700	0.06±0.02	0.04±0.02	0.05±0.02	0.07±0.02
	4000	0.03±0.01	0.02±0.01	0.03±0.02	0.05±0.01
DNN	10	0.35±0.12	0.42±0.08	0.42±0.09	0.41±0.05
	24	0.25±0.08	0.28±0.10	0.28±0.11	0.37±0.09
	55	0.18±0.05	0.20±0.05	0.14±0.04	0.20±0.04
	130	0.11±0.03	0.09±0.02	0.08±0.02	0.11±0.02
	307	0.08±0.02	0.06±0.02	0.05±0.01	0.06±0.01
	722	0.04±0.02	0.04±0.01	0.04±0.01	0.04±0.01
	1700	0.02±0.01	0.02±0.01	0.02±0.01	0.03±0.01
	4000	0.01±0.00	0.01±0.01	0.01±0.00	0.02±0.01
RFR	10	0.40±0.04	0.36±0.07	0.39±0.05	0.39±0.04
	24	0.33±0.06	0.30±0.08	0.38±0.04	0.38±0.02
	55	0.26±0.04	0.22±0.03	0.35±0.02	0.40±0.02
	130	0.16±0.02	0.17±0.03	0.27±0.03	0.34±0.02
	307	0.11±0.02	0.11±0.02	0.19±0.03	0.26±0.02
	722	0.06±0.01	0.06±0.01	0.11±0.01	0.19±0.02
	1700	0.04±0.01	0.03±0.01	0.07±0.01	0.11±0.01
	4000	0.02±0.01	0.02±0.01	0.05±0.01	0.07±0.01
SNN	10	0.38±0.12	0.43±0.08	0.43±0.09	0.42±0.05
	24	0.25±0.08	0.30±0.10	0.29±0.10	0.40±0.09
	55	0.19±0.05	0.22±0.05	0.16±0.04	0.21±0.04
	130	0.12±0.03	0.10±0.02	0.08±0.02	0.11±0.02
	307	0.08±0.02	0.06±0.02	0.06±0.01	0.07±0.01
	722	0.04±0.02	0.04±0.01	0.04±0.01	0.04±0.01
	1700	0.02±0.01	0.02±0.01	0.03±0.01	0.03±0.01
	4000	0.01±0.00	0.01±0.01	0.01±0.00	0.02±0.01

Table 13: CLASS-PROB: AMR of the models learned with the different algorithms (DNN, SNN, ELM, SVM, KRLS, KNN, and GP) when varying n for $P15$ and for the four main NPS components

		PRP	HLL	GTC	GT
KRLS	AMR	0.03±0.01	0.02±0.01	0.04±0.01	0.05±0.00
	TP	45.11±1.10	45.68±1.79	58.06±1.67	57.29±1.37
	TN	51.72±1.02	52.33±1.97	38.31±1.45	37.77±1.50
	FN	1.64±0.39	0.84±0.28	1.96±0.74	2.81±0.55
	FP	1.53±0.47	1.15±0.35	1.67±0.49	2.13±0.37
ELM	AMR	0.01±0.00	0.01±0.01	0.01±0.00	0.02±0.01
	TP	45.05±1.87	46.32±0.85	60.18±1.14	57.51±1.80
	TN	53.60±1.67	52.22±1.08	38.33±1.23	40.20±1.53
	FN	0.80±0.31	0.64±0.30	0.73±0.24	1.94±0.75
	FP	0.55±0.22	0.82±0.41	0.76±0.31	0.35±0.27
KNN	AMR	0.09±0.01	0.07±0.01	0.10±0.01	0.15±0.01
	TP	42.06±0.63	43.10±1.75	55.23±1.68	52.65±1.70
	TN	48.46±1.03	49.56±1.93	34.92±1.53	31.96±1.81
	FN	4.69±0.51	3.42±0.53	4.79±0.71	7.45±0.68
	FP	4.79±0.59	3.92±0.40	5.06±0.79	7.94±0.82
RF	AMR	0.02±0.01	0.02±0.01	0.06±0.01	0.08±0.01
	TP	45.81±1.90	45.70±1.47	57.17±1.01	56.31±1.53
	TN	52.03±1.68	51.89±1.26	37.31±1.01	35.78±1.61
	FN	1.23±0.42	1.29±0.37	2.52±0.39	3.50±0.77
	FP	0.93±0.36	1.12±0.48	3.00±0.73	4.41±0.73
GP	AMR	0.08±0.01	0.04±0.01	0.07±0.01	0.11±0.01
	TP	42.33±0.84	44.17±1.94	57.43±1.55	57.07±1.51
	TN	49.99±0.91	51.55±1.90	35.71±1.32	32.39±1.45
	FN	4.42±0.67	2.35±0.85	2.59±0.74	3.03±0.63
	FP	3.26±0.86	1.93±0.51	4.27±0.73	7.51±1.03
SVM	AMR	0.03±0.01	0.02±0.01	0.03±0.01	0.05±0.00
	TP	45.15±1.21	45.70±1.95	58.16±1.90	57.36±1.48
	TN	51.79±1.13	52.43±2.15	38.37±1.57	37.95±1.65
	FN	1.60±0.42	0.82±0.30	1.86±0.83	2.74±0.59
	FP	1.46±0.53	1.05±0.39	1.61±0.53	1.95±0.39
DNN	AMR	0.01±0.00	0.01±0.01	0.01±0.00	0.02±0.01
	TP	45.10±2.08	46.40±0.96	60.25±1.20	57.65±2.02
	TN	53.63±1.83	52.30±1.23	38.40±1.38	40.24±1.75
	FN	0.75±0.35	0.56±0.34	0.66±0.26	1.80±0.85
	FP	0.52±0.25	0.74±0.45	0.69±0.34	0.31±0.31
RFR	AMR	0.02±0.01	0.02±0.01	0.05±0.01	0.07±0.01
	TP	46.05±2.12	45.96±1.63	57.56±1.08	56.68±1.64
	TN	52.21±1.88	52.05±1.43	37.70±1.07	36.45±1.72
	FN	0.99±0.48	1.03±0.39	2.13±0.45	3.13±0.88
	FP	0.75±0.40	0.96±0.53	2.61±0.83	3.74±0.78
SNN	AMR	0.01±0.00	0.01±0.01	0.01±0.00	0.02±0.01
	TP	45.10±2.00	46.33±0.93	60.19±1.25	57.56±1.89
	TN	53.64±1.90	52.25±1.17	38.36±1.41	40.20±1.63
	FN	0.75±0.34	0.63±0.32	0.72±0.28	1.89±0.79
	FP	0.51±0.25	0.79±0.47	0.73±0.34	0.35±0.29

Table 14: CLASS-PROB: the different indexes of performances (AMR, TP, TN, FP, and FN) of the models learned with the different algorithms (DNN, SNN, ELM, SVM, KRLS, KNN, and GP) when n is the largest possible for $P15$ and for the four main NPS components.

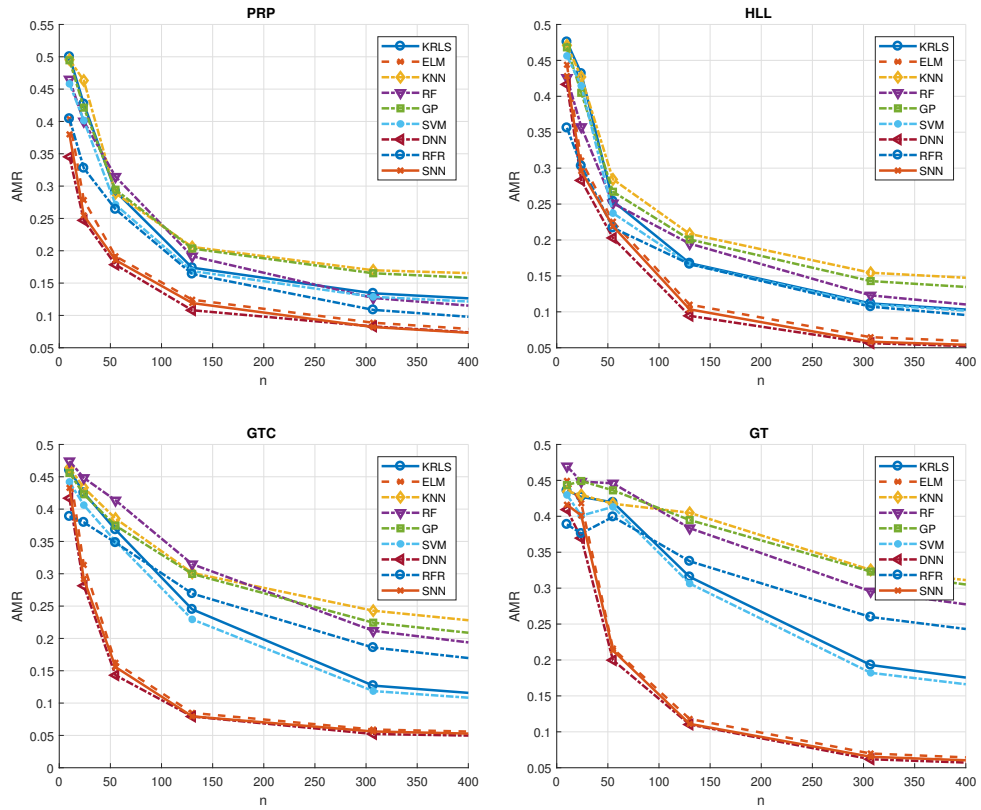


Figure 11: CLASS-PROB: AMR of the models learned with the different algorithms when varying n for $P15$ and for the four main NPS components.

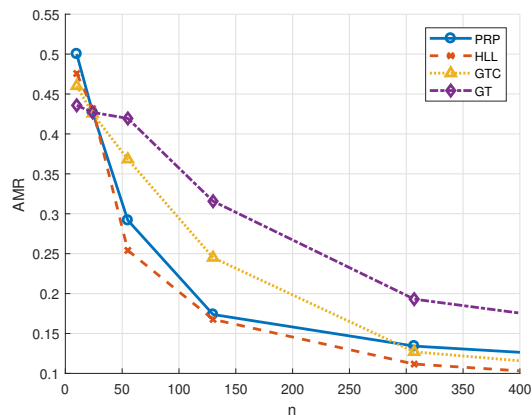


Figure 12: CLASS-PROB: AMR of the models learned with DNN when varying n for $P15$ and for the four main NPS components.

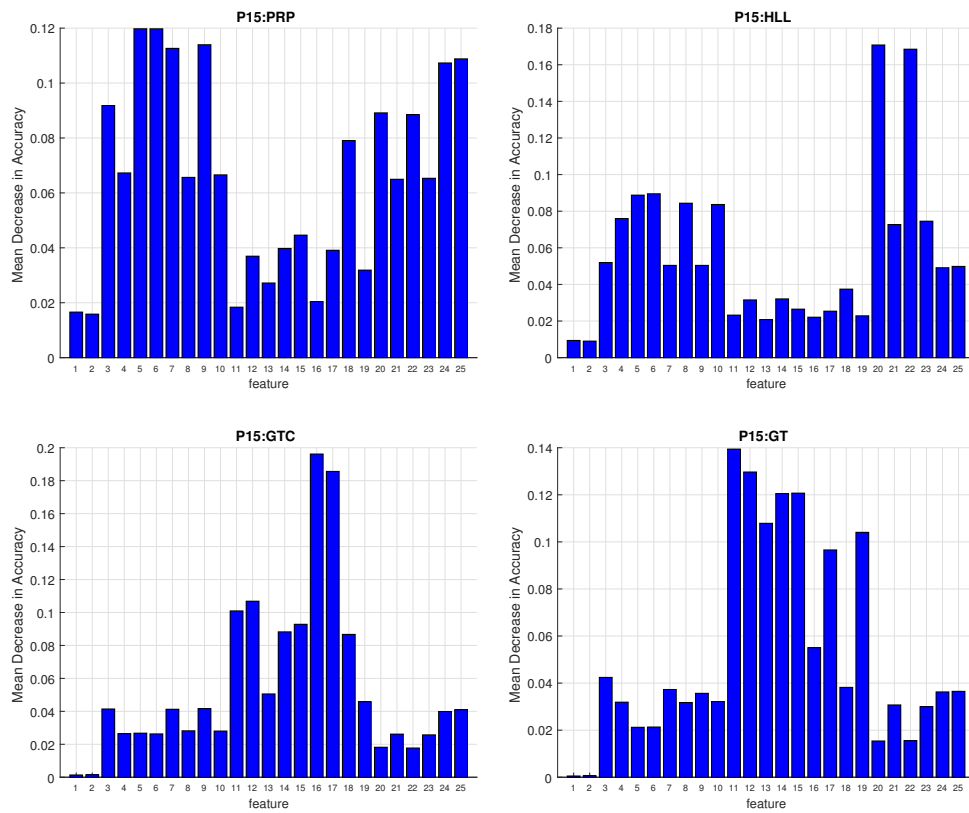


Figure 13: CLASS-PROB: FS performed with RF for the four main NPS components for problem $P15$.

	v	PRP	HLL	GTC	GT
OCSVM	10	0.08±0.08	0.07±0.09	0.05±0.07	0.11±0.06
	20	0.08±0.10	0.08±0.07	0.10±0.07	0.09±0.06
	30	0.08±0.07	0.12±0.08	0.10±0.07	0.09±0.03
GKNN	10	0.07±0.07	0.07±0.09	0.04±0.07	0.10±0.06
	20	0.08±0.10	0.08±0.07	0.08±0.07	0.07±0.06
	30	0.08±0.07	0.12±0.08	0.08±0.08	0.08±0.03

Table 15: ANOMALY-PROB: AMR of the models learned with the different algorithms (OCSVM and GKNN) when $l = 4000$ and $v \in \{10, 20, 30\}$ for $P15$ and for the four main NPS components.

	n	PRP	HLL	GTC	GT
OCSVM	1500	0.09±0.04	0.14±0.09	0.11±0.09	0.08±0.03
	2000	0.12±0.10	0.11±0.04	0.10±0.04	0.11±0.13
	3000	0.08±0.06	0.11±0.04	0.13±0.17	0.08±0.05
	4000	0.08±0.07	0.12±0.08	0.10±0.07	0.09±0.03
GKNN	1500	0.09±0.04	0.12±0.09	0.10±0.09	0.06±0.03
	2000	0.11±0.11	0.12±0.04	0.08±0.04	0.08±0.13
	3000	0.07±0.06	0.10±0.04	0.12±0.16	0.09±0.04
	4000	0.08±0.07	0.12±0.08	0.08±0.08	0.08±0.03

Table 16: ANOMALY-PROB: AMR of the models learned with the different algorithms (OCSVM and GKNN) when $l \in \{1500, 2000, 3000, 4000\}$ and $v = 30$ for $P15$ and for the four main NPS components.

- changing l or v does not remarkably affect the performance of the models;
- with just few labeled samples, around 10, it is possible to obtain satisfying accuracies and this is quite a remarkable result, since 10 samples can be easily manually labeled by an expert operator;
- FP and FN rate are quite balanced and this is a further indication of the quality of the result.

6. Conclusions

The behavior and interaction of the main components of Ship Propulsion Systems cannot be easily modeled with a priori physical knowledge, considering the large amount of variables influencing them. In fact, in this work, authors showed that by exploiting the most recent statistical techniques and the large amount of historical data collected by the current on-board automation systems it is possible to build effective Data-Driven Models which do not require any a priori knowledge. In particular, the developed models are able to continuously monitor the propulsion equipment to avoid Preventive or Corrective Maintenance and take decisions based on the actual condi-

		PRP	HLL	GTC	GT
OCSVM	AMR	0.08±0.07	0.12±0.08	0.10±0.07	0.09±0.03
	TP	46.68±8.88	45.29±9.25	47.89±3.17	46.59±5.49
	TN	45.47±3.88	42.56±7.54	42.42±9.36	44.56±4.50
	FN	3.32±8.88	4.71±9.25	2.11±3.17	3.41±5.49
	FP	4.53±3.88	7.44±7.54	7.58±9.36	5.44±4.50
GKNN	AMR	0.08±0.07	0.12±0.08	0.08±0.08	0.08±0.03
	TP	46.68±8.66	45.89±9.04	48.04±3.16	47.42±5.41
	TN	45.24±3.86	42.30±7.61	44.14±9.55	44.49±4.47
	FN	3.32±8.84	4.11±9.21	1.96±3.13	2.58±5.54
	FP	4.76±3.87	7.70±7.54	5.86±9.29	5.51±4.56

Table 17: ANOMALY-PROB: the different indexes of performances (AMR, TP, TN, FP, and FN) of the models learned with the different algorithms (OCSVM and GKNN) when $n = 4000$ and $v = 3$ for $P15$ and for the four main NPS components.

tion of the propulsion plant. Furthermore, since labelled sensor data can be quite expensive and in some cases unfeasible in this paper the authors investigated the problem of performing Condition-Based Maintenance both in a supervised and unsupervised fashion in order to minimize the feedback of the operators in labeling the sensor data. A naval ship, characterized by a COm-bined Diesel ELeCtric And Gas propulsion plant, has been exploited to show the effectiveness of the proposed approaches. Confidentiality constraints with the Navy forced the authors to use a real-data validated simulator and the dataset has been published for free use.

In conclusion, this study proved that it is possible to treat a Condition-Based Maintenance problem in an unsupervised fashion, with results close to the ones obtained with supervised techniques present in literature. These models can be adopted for real-time applications directly on-board, to easily and quickly identify maintenance necessities.

References

- [1] D. Anguita, A. Ghio, N. Greco, L. Oneto, S. Ridella, Model selection for support vector machines: Advantages and disadvantages of the machine learning theory, in: International Joint Conference on Neural Networks, 2010.
- [2] S. Boucheron, O. Bousquet, G. Lugosi, Theory of classification: A survey of some recent advances, ESAIM: probability and statistics 9 (2005) 323–375.
- [3] Y. Peng, M. Dong, M. J. Zuo, Current status of machine prognostics

in condition-based maintenance: a review, *The International Journal of Advanced Manufacturing Technology* 50 (1-4) (2010) 297–313.

- [4] A. Coraddu, L. Oneto, F. Baldi, D. Anguita, Vessels fuel consumption forecast and trim optimisation: A data analytics perspective, *Ocean Engineering* 130 (2017) 351–370.
- [5] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann, 2016.
- [6] E. V. Lewis, *Principles of naval architecture*, Vol. 2, The Society of Naval Architects and Marine Engineers, 1988.
- [7] V. N. Vapnik, *Statistical learning theory*, Wiley New York, 1998.
- [8] L. Györfi, M. Kohler, A. Krzyzak, H. Walk, *A distribution-free theory of nonparametric regression*, Springer Science & Business Media, 2006.
- [9] L. R. Newton, Data-logging in practical science: research and reality, *International Journal of Science Education* 22 (12) (2000) 1247–1259.
- [10] R. C. M. Yam, P. W. Tse, L. Li, P. Tu, Intelligent predictive decision support system for condition-based maintenance, *The International Journal of Advanced Manufacturing Technology* 17 (5) (2001) 383–391.
- [11] R. Langone, C. Alzate, B. De Ketelaere, J. Vlasselaer, W. Meert, J. A. K. Suykens, Ls-svm based spectral clustering and regression for predicting maintenance of industrial machines, *Engineering Applications of Artificial Intelligence* 37 (2015) 268–278.
- [12] R. Sommer, V. Paxson, Outside the closed world: On using machine learning for network intrusion detection, in: *IEEE Symposium on Security and Privacy*, 2010.
- [13] T. T. T. Nguyen, G. Armitage, A survey of techniques for internet traffic classification using machine learning, *IEEE Communications Surveys & Tutorials* 10 (4) (2008) 56–76.
- [14] L. Gómez-Chova, D. Tuia, G. Moser, G. Camps-Valls, Multimodal classification of remote sensing images: A review and future directions, *Proceedings of the IEEE* 103 (9) (2015) 1560–1584.

- [15] I. Bose, R. K. Mahapatra, Business data mining-a machine learning perspective, *Information & management* 39 (3) (2001) 211–225.
- [16] K. Shin, T. S. Lee, H. Kim, An application of support vector machines in bankruptcy prediction model, *Expert Systems with Applications* 28 (1) (2005) 127–135.
- [17] A. Cielen, L. Peeters, K. Vanhoof, Bankruptcy prediction using a data envelopment analysis, *European Journal of Operational Research* 154 (2) (2004) 526–532.
- [18] G. D. Magoulas, A. Prentza, Machine learning in medical applications, in: *Machine Learning and its applications*, 2001.
- [19] A. Mannini, A. M. Sabatini, Machine learning methods for classifying human physical activity from on-body accelerometers, *Sensors* 10 (2) (2010) 1154–1175.
- [20] P. Lucas, Bayesian analysis, pattern analysis, and data mining in health care, *Current opinion in critical care* 10 (5) (2004) 399–403.
- [21] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: sentiment classification using machine learning techniques, in: *ACL-02 conference on Empirical methods in natural language processing-Volume 10*, 2002.
- [22] S. Das, M. Chen, Yahoo! for amazon: Extracting market sentiment from stock message boards, in: *Asia Pacific finance association annual conference (APFA)*, 2001.
- [23] J. P. Petersen, O. Winther, D. J. Jacobsen, A machine-learning approach to predict main energy consumption under realistic operational conditions, *Ship Technology Research* 59 (1) (2012) 64–72.
- [24] G. Budai, D. Huisman, R. Dekker, Scheduling preventive railway maintenance activities, *The International Journal of Advanced Manufacturing Technology* 57 (9) (2006) 1035–1044.
- [25] T. Smith, E. O’Keeffe, L. Aldous, P. Agnolucci, Assessment of shipping’s efficiency using satellite ais data, in: *International Council on Clean Transportation*, 2013.

- [26] G. Waeyenbergh, L. Pintelon, A framework for maintenance concept development, *International journal of production economics* 77 (3) (2002) 299–313.
- [27] G. S. Linoff, M. J. A. Berry, *Data mining techniques: for marketing, sales, and customer relationship management*, John Wiley & Sons, 2011.
- [28] L. Oneto, E. Fumeo, G. Clerico, R. Canepa, F. Papa, C. Dambra, N. Mazzino, D. Anguita, Advanced analytics for train delay prediction systems by including exogenous weather data, in: *International Conference on Data Science and Advanced Analytics*, 2016.
- [29] E. Cambria, M. Grassi, A. Hussain, C. Havasi, Sentic computing for social media marketing, *Multimedia tools and applications* 59 (2) (2012) 557–577.
- [30] J. K. Harter, F. L. Schmidt, T. L. Hayes, Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: a meta-analysis., *Journal of Applied Psychology* 87 (2) (2002) 268–279.
- [31] H. Chen, R. H. L. Chiang, V. C. Storey, Business intelligence and analytics: From big data to big impact., *MIS quarterly* 36 (4) (2012) 1165–1188.
- [32] S. Mills, S. Lucas, L. Irakliotis, M. Rappa, T. Carlson, B. Perlowitz, Demystifying big data: a practical guide to transforming the business of government, in: *TechAmerica Foundation*, Washington, 2012.
- [33] J. Carlton, *Marine propellers and propulsion*, Butterworth-Heinemann, 2011.
- [34] P. Andersen, W. Rządki, Hybrid ship propulsion system, *Google Patents*, uS Patent 7,241,192 (2007).
- [35] H. Wang, O. L. Osen, G. Li, W. Li, H. Dai, W. Zeng, Big data and industrial internet of things for the maritime industry in northwestern norway, in: *IEEE Region 10 Conference TENCON*, 2015.

- [36] L. Oneto, D. Anguita, A. Coraddu, T. Cleophas, K. Xepapa, Vessel monitoring and design in industry 4.0: A data driven perspective, in: IEEE International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow, 2016.
- [37] A. Coraddu, L. Oneto, F. Baldi, D. Anguita, Ship efficiency forecast based on sensors data collection: Improving numerical models through data analytics, in: OCEANS 2015-Genova, 2015.
- [38] A. Coraddu, L. Oneto, A. Ghio, S. Savio, D. Anguita, M. Figari, Machine learning approaches for improving condition-based maintenance of naval propulsion plants, Proceedings of the Institution of Mechanical Engineers Part M: Journal of Engineering for the Maritime Environment 230 (1) (2016) 136–153.
- [39] A. Widodo, B. Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mechanical systems and signal processing 21 (6) (2007) 2560–2574.
- [40] R. K. Mobley, An introduction to predictive maintenance, Butterworth-Heinemann, 2002.
- [41] G. B. Q. Management, S. S. Committee, B. S. Institution, British Standard Glossary of Maintenance Management Terms in Terotechnology, British Standards Institution, 1984.
- [42] G. Budai-Balke, Operations research models for scheduling railway infrastructure maintenance, Rozenberg Publishers, 2009.
- [43] R. Kothamasu, S. H. Huang, Adaptive mamdani fuzzy model for condition-based maintenance, Fuzzy Sets and Systems 158 (24) (2007) 2715–2733.
- [44] A. K. S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical systems and signal processing 20 (7) (2006) 1483–1510.
- [45] S. Poyhonen, P. Jover, H. Hyotyniemi, Signal processing of vibrations for condition monitoring of an induction motor, in: Symposium on Control, Communications and Signal Processing, 2004.

- [46] C. Bunks, D. McCarthy, T. Al-Ani, Condition-based maintenance of machines using hidden markov models, *Mechanical Systems and Signal Processing* 14 (4) (2000) 597–612.
- [47] S. Simani, C. Fantuzzi, R. Patton, *Model-based Fault Diagnosis in Dynamic Systems Using Identification Techniques*, Springer-Verlag London, 2003.
- [48] T. Palmé, P. Breuhaus, M. Assadi, A. Klein, M. Kim, New alstom monitoring tools leveraging artificial neural network technologies, in: *Turbo Expo: Turbine Technical Conference and Exposition*, 2011.
- [49] R. M. Singer, K. C. Gross, R. W. King, A pattern-recognition-based, fault-tolerant monitoring and diagnostic technique, in: *Argonne National Lab., Idaho Falls, ID (United States)*, 1995.
- [50] T. S. Akinfiyev, M. A. Armada, R. Fernandez, Nondestructive testing of the state of a ship’s hull with an underwater robot, *Russian Journal of Nondestructive Testing* 44 (9) (2008) 626–633.
- [51] S. Bagavathiappan, B. B. Lahiri, T. Saravanan, J. Philip, T. Jayakumar, Infrared thermography for condition monitoring-a review, *Infrared Physics & Technology* 60 (2013) 35–55.
- [52] O. C. Basurko, Z. Uriondo, Condition-based maintenance for medium speed diesel engines used in vessels in operation, *Applied Thermal Engineering* 80 (2015) 404–412.
- [53] I. Lazakis, K. Dikis, A. L. Michala, G. Theotokatos, Advanced ship systems condition monitoring for enhanced inspection, maintenance and decision making in ship operations, *Transportation Research Procedia* 14 (2016) 1679–1688.
- [54] J. Shawe-Taylor, N. Cristianini, *Kernel methods for pattern analysis*, Cambridge university press, 2004.
- [55] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain., *Psychological review* 65 (6) (1958) 386.

- [56] C. E. Rasmussen, Gaussian processes for machine learning, in: *Gaussian Processes for Machine Learning*, 2006.
- [57] W. Duch, Similarity-based methods: a general framework for classification, approximation and association, *Control and Cybernetics* 29 (4).
- [58] L. Breiman, Random forests, *Machine learning* 45 (1) (2001) 5–32.
- [59] M. Markou, S. Singh, Novelty detection: a review, *Signal processing* 83 (12) (2003) 2481–2497.
- [60] L. Swersky, H. O. Marques, J. Sander, R. J. Campello, A. Zimek, On the evaluation of outlier detection and one-class classification methods, in: *IEEE International Conference on Data Science and Advanced Analytics*, 2016.
- [61] M. Figari, M. Altosole, Dynamic behaviour and stability of marine propulsion systems, *Journal of Engineering for the Maritime Environment* 221 (4) (2007) 187–205.
- [62] G. Benvenuto, U. Campora, Performance prediction of a faulty marine diesel engine under different governor settings, in: *International Conference on Marine Research And Transportation*, 2007.
- [63] M. Altosole, G. Benvenuto, U. Campora, Numerical modelling of the engines governors of a codlag propulsion plant, in: *International Conference on Marine Sciences and Technologies*, 2010.
- [64] M. Altosole, G. Benvenuto, M. Figari, U. Campora, Real-time simulation of a cogag naval ship propulsion system, *Journal of Engineering for the Maritime Environment* 223 (1) (2009) 47–62.
- [65] G. Benvenuto, U. Campora, A gas turbine modular model for ship propulsion studies, in: *Symposium on High Speed Marine Vehicles*, 2005.
- [66] H. Saravanamuttoo, G. F. Rogers, H. Cohen, *Gas turbine theory*, Pearson Education, 2001.
- [67] U. Campora, M. Figari, Numerical simulation of ship propulsion transients and full-scale validation, *Journal of Engineering for the Maritime Environment* 217 (1) (2003) 41–52.

- [68] M. Altosole, U. Campora, M. Martelli, M. Figari, Performance decay analysis of a marine gas turbine propulsion system, *Journal of Ship Research* 58 (3) (2014) 117–129.
- [69] A. P. Tarabrin, V. A. Schurovsky, A. I. Bodrov, J. P. Stalder, An analysis of axial compressor fouling and a blade cleaning method, *Journal of Turbomachinery* 120 (2) (1998) 256–261.
- [70] C. B. Meher-Homji, A. B. Focke, M. B. Wooldridge, Fouling of axial flow compressors - causes, effects, detection, and control, in: *Eighteenth Turbomachinery Symposium*, 1989.
- [71] G. F. Aker, H. I. H. Saravanamuttoo, Predicting gas turbine performance degradation due to compressor fouling using computer simulation techniques, *Journal of Engineering for Gas Turbines and Power* 111 (2) (2006) 343–350.
- [72] Y. S. Khor, Q. Xiao, Cfd simulations of the effects of fouling and antifouling, *Ocean Engineering* 38 (10) (2011) 1065–1079.
- [73] M. Atlar, E. J. Glover, M. Candries, R. J. Mutton, C. D. Anderson, The effect of a foul release coating on propeller performance, in: *Marine Science and Technology for Environmental Sustainability*, 2002.
- [74] B. Wan, E. Nishikawa, M. Uchida, The experiment and numerical calculation of propeller performance with surface roughness effects, *Journal of the Kansai Society of Naval Architects, Japan* 2002 (238) (2002) 49–54.
- [75] A. Lindholdt, K. Dam-Johansen, S. M. Olsen, D. M. Yebra, S. Kiil, Effects of biofouling development on drag forces of hull coatings for ocean-going ships: a review, *Journal of Coatings Technology and Research* 12 (3) (2015) 315–444.
- [76] J. B. Hadler, C. J. Wilson, A. L. Beal, Ship standardisation trial performance and correlation with model predictions, in: *The Society of Naval Architects and Marine Engineers*, 1962.
- [77] K. Bache, M. Lichman, UCI machine learning repository (2013). URL <http://archive.ics.uci.edu/ml>

- [78] D. M. Hawkins, Identification of outliers, Springer, 1980.
- [79] T. Hastie, R. Tibshirani, J. Friedman, Unsupervised learning, in: The elements of statistical learning, 2009.
- [80] L. Rosasco, E. De Vito, A. Caponnetto, M. Piana, A. Verri, Are loss functions all the same?, *Neural Computation* 16 (5) (2004) 1063–1076.
- [81] L. Ghelardoni, A. Ghio, D. Anguita, Energy load forecasting using empirical mode decomposition and support vector regression, *IEEE Transactions on Smart Grid* 4 (1) (2013) 549–556.
- [82] D. M. Powers, Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation, *Journal of Machine Learning Technologies* 2 (1) (2011) 37–63.
- [83] R. Kohavi, et al., A study of cross-validation and bootstrap for accuracy estimation and model selection, in: International Joint Conference on Artificial Intelligence, 1995.
- [84] P. L. Bartlett, S. Boucheron, G. Lugosi, Model selection and error estimation, *Machine Learning* 48 (1-3) (2002) 85–113.
- [85] D. Anguita, A. Ghio, L. Oneto, S. Ridella, In-sample and out-of-sample model selection and error estimation for support vector machines, *IEEE Transactions on Neural Networks and Learning Systems* 23 (9) (2012) 1390–1406.
- [86] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, *The Journal of Machine Learning Research* 3 (2003) 1157–1182.
- [87] J. Friedman, T. Hastie, R. Tibshirani, The elements of statistical learning, Springer series in statistics Springer, Berlin, 2001.
- [88] Y. W. Chang, C. J. Lin, Feature ranking using linear svm., in: WCCI Causation and Prediction Challenge, 2008.
- [89] H. Yoon, K. Yang, C. Shahabi, Feature subset selection and feature ranking for multivariate time series, *IEEE transactions on knowledge and data engineering* 17 (9) (2005) 1186–1198.

- [90] S. J. Hong, Use of contextual information for feature ranking and discretization, *IEEE Transactions on Knowledge and Data Engineering* 9 (5) (1997) 718–730.
- [91] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors, *Cognitive modeling* 5 (3) (1988) 1.
- [92] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Mathematics of Control, Signals, and Systems (MCSS)* 2 (4) (1989) 303–314.
- [93] C. M. Bishop, *Neural networks for pattern recognition*, Oxford university press, 1995.
- [94] G. E. Hinton, S. Osindero, Y. W. Teh, A fast learning algorithm for deep belief nets, *Neural computation* 18 (7) (2006) 1527–1554.
- [95] Y. Bengio, Learning deep architectures for ai, *Foundations and trends[®] in Machine Learning* 2 (1) (2009) 1–127.
- [96] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, *IEEE transactions on pattern analysis and machine intelligence* 35 (8) (2013) 1798–1828.
- [97] G. Huang, G.-B. Huang, S. Song, K. You, Trends in extreme learning machines: A review, *Neural Networks* 61 (2015) 32–48.
- [98] G. B. Huang, Q. Y. Zhu, C. K. Siew, Extreme learning machine: theory and applications, *Neurocomputing* 70 (1) (2006) 489–501.
- [99] E. Cambria, G.-B. Huang, Extreme learning machines, *IEEE Intelligent Systems* 28 (6) (2013) 30–59.
- [100] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, in: *IEEE International Joint Conference on Neural Networks*, 2004.
- [101] N. Cristianini, J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*, Cambridge university press, 2000.

- [102] B. Scholkopf, The kernel trick for distances, in: *Advances in neural information processing systems*, 2001.
- [103] R. Rifkin, G. Yeo, T. Poggio, Regularized least-squares classification, *Nato Science Series Sub Series III Computer and Systems Sciences* 190 (2003) 131–154.
- [104] W. S. Lee, P. L. Bartlett, R. C. Williamson, The importance of convexity in learning with squared loss, *IEEE Transactions on Information Theory* 44 (5) (1998) 1974–1980.
- [105] S. S. Keerthi, C. J. Lin, Asymptotic behaviors of support vector machines with gaussian kernel, *Neural computation* 15 (7) (2003) 1667–1689.
- [106] L. Oneto, A. Ghio, S. Ridella, D. Anguita, Support vector machines and strictly positive definite kernel: The regularization hyperparameter is more important than the kernel hyperparameters, in: *IEEE International Joint Conference on Neural Networks*, 2015.
- [107] P. Germain, A. Lacasse, F. Laviolette, M. Marchand, J. F. Roy, Risk bounds for the majority vote: From a pac-bayesian analysis to a learning algorithm, *The Journal of Machine Learning Research* 16 (1) (2015) 787–860.
- [108] R. Blaser, P. Fryzlewicz, Random rotation ensembles, *The Journal of Machine Learning Research* 2 (2015) 1–15.
- [109] M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim, Do we need hundreds of classifiers to solve real world classification problems, *The Journal of Machine Learning Research* 15 (1) (2014) 3133–3181.
- [110] M. Wainberg, B. Alipanahi, B. J. Frey, Are random forests truly the best classifiers?, *The Journal of Machine Learning Research* 17 (110) (2016) 1–5.
- [111] I. Orlandi, L. Oneto, D. Anguita, Random forests model selection, in: *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2016.

- [112] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, Bayesian data analysis, Vol. 2, Chapman and Hall/CRC, 2014.
- [113] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE transactions on information theory 13 (1) (1967) 21–27.
- [114] S. Ramaswamy, R. Rastogi, K. Shim, Efficient algorithms for mining outliers from large data sets, in: ACM Sigmod Record, 2000.
- [115] B. Efron, R. J. Tibshirani, An introduction to the bootstrap, CRC press, 1994.
- [116] D. Anguita, A. Boni, S. Ridella, Evaluating the generalization ability of support vector machines through the bootstrap, Neural Processing Letters 11 (1) (2000) 51–58.
- [117] S. Arlot, A. Celisse, A survey of cross-validation procedures for model selection, Statistics surveys 4 (2010) 40–79.
- [118] P. Good, Permutation tests: a practical guide to resampling methods for testing hypotheses, Springer Science & Business Media, 2013.
- [119] F. Chollet, Keras, <https://keras.io> (2015).
- [120] A. Liaw, M. Wiener, Classification and regression by randomforest, R News 2 (3) (2002) 18–22.
- [121] A. Zeileis, K. Hornik, A. Smola, A. Karatzoglou, kernlab-an s4 package for kernel methods in r, Journal of statistical software 11 (9) (2004) 1–20.
- [122] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch, e1071: Misc functions of the department of statistics, probability theory group (formerly: E1071), tu wien, R package, r package version 1.6-7 (2015).