

Hill chart modelling using the Hermite polynomial chaos expansion for the performance prediction of pumps running as turbines

J. Delgado^{a,b,*}, L. Andolfatto^a, D.I.C. Covas^b, F. Avellan^a

^a Laboratory for Hydraulic Machines, École Polytechnique Fédérale de Lausanne, Switzerland

^b CERIS, Instituto Superior Técnico, Universidade de Lisboa, Portugal

ARTICLE INFO

Keywords:

Pump running as turbine
Characteristic curves prediction
Hill chart modelling
Hermite polynomial chaos expansion
Variable speed operation

ABSTRACT

Pumps running as turbines are suitable hydraulic machines for micro hydropower applications. The selection of the proper pumps to install in a given site still remains a major challenge, as pump manufacturers do not provide the characteristic curves data in the turbine mode. Also, the accurate prediction and modelling of the pumps running as turbines characteristic curves still remain a major difficulty as existing methodologies still lack accuracy, especially in the part load and full load operating regions. This paper proposes a new two-step methodology based on the Hermite polynomial chaos expansion for predicting the characteristic curves of pumps running as turbines and modelling their variable speed operation, aiming at improving the prediction accuracy. Firstly, bivariate continuous surrogate functions are established for predicting the turbine mode and the extended operation mode characteristic curves inside a closed interval of unit specific speed values. These surrogate functions are developed by calibrating empirical coefficients based on collected experimental data. Secondly, a hill chart model is determined for describing the variable speed operation of a given pump running as a turbine. This hill chart model allows identifying the discharge and the rotational speed set points for maximising efficiency for a given operating condition. The proposed prediction surrogate functions and the variable speed hill chart model are useful engineering tools for improving the design of pump as turbine hydropower plants and for optimising the pump running as turbine control settings to maximise the produced energy.

1. Introduction

Pumps running as turbines (PATs¹) are hydraulic machines used for micro hydropower applications such as the energy recovery in water trunk mains [1] or in water distribution networks [2], as well as for the rural electrification of remote areas [3]. The low capital investment, the immediate availability in several standardised sizes for a wide range of head and discharge values and the possibility of using the original induction motor as an asynchronous generator make PATs a cost-effective alternative to custom-made turbines for micro hydropower application [4].

The selection of the proper PAT to install in a given power plant is still a major challenge, as the turbine mode characteristic curves are not provided by pump manufacturers. For this purpose, several methodologies have been proposed to predict the turbine mode performance. However, the accurate prediction of the characteristic curves is still a

major difficulty, as existing methodologies lack accuracy, especially in the part load and full load operating regions. The best efficiency point (BEP) in the turbine mode is found for higher values of head and discharge than the BEP in pump mode, while the maximum efficiency in both modes has approximately the same value [5]. The shape of the PAT characteristic curves in the turbine mode is dependent on the unit specific speed parameter [6]. The knowledge of the turbine full operating range is of the utmost importance for the development of the control strategies [7] and for the assessment of project economic feasibility [8], as PATs efficiency is highly affected when the machine operates under variable discharge conditions [9].

Existing methodologies for predicting the turbine mode characteristic curves still lack accuracy and PATs are usually tested in laboratory conditions prior to on-site installation [10]. The most accurate predictions are achieved by empirical formulations, being its application limited, however, to the type of pumps and to the range of unit specific

* Corresponding author at: CERIS, Instituto Superior Técnico, Universidade de Lisboa, Lisboa 1049-001, Portugal.

E-mail addresses: joao.delgado@epfl.ch, joao.borga.delgado@tecnico.ulisboa.pt (J. Delgado).

¹ For notation simplicity: (i) PAT refer to the single version of the acronym – pump running as turbine; (ii) PATs refer to the plural version of the acronym – pumps running as turbines.

Nomenclature		Ψ	truncated basis (–)
<i>Latin symbols</i>		ρ	water density ($\text{kg}\cdot\text{m}^{-3}$)
D	impeller diameter at the high pressure section (m)	<i>Superscripts and subscripts</i>	
E	specific hydraulic energy ($\text{J}\cdot\text{kg}^{-1}$)	EO	extended operation in generating mode
g	gravity acceleration ($\text{m}\cdot\text{s}^{-2}$)	HC	hill chart
H	head (m)	P	pump mode
N	rotational speed (min^{-1})	R	rated conditions
n	runner rotating frequency (s^{-1})	T	turbine mode
n_q, n_s	unit specific speed (SI)	<i>Acronyms</i>	
N_s	measurement samples size (–)	3D	three dimensional
n_{ED}	IEC Speed factor (–)	AE	absolute error
P	shaft power (W)	AIC	Akaike information criterion
P_h	hydraulic power (W)	BEP	best efficiency point
P_{max}	maximum index p of the truncated basis (–)	IEC	International Electrotechnical Commission
P_{ED}	IEC Power factor (–)	PAT	pump running as turbine
Q	discharge ($\text{m}^3\cdot\text{s}^{-1}$)	PATs	pumps running as turbines
Q_{ED}	IEC Discharge factor (–)	PCE	polynomial chaos expansion
T	shaft torque ($\text{N}\cdot\text{m}$)	<i>Number sets</i>	
T_{ED}	IEC Shaft torque factor (–)	\mathbb{N}	set of natural numbers
<i>Greek symbols</i>		\mathbb{R}	set of real numbers
η	efficiency (–)		
λ	weighting coefficient (–)		

speed values for which the formulations were developed. The unit specific speed parameter n_q (or n_s depending on the definition) is defined by Eq. (1). A summary of the scope of application of the formulations found in the literature is provided in Table 1.

$$n_q = N \frac{Q^{\frac{1}{2}}}{H^{\frac{3}{4}}}$$

$$n_s = N \frac{P^{\frac{1}{2}}}{H^{\frac{5}{4}}} \tag{1}$$

where N is the rotational speed, Q is the discharge, H is the head and P is the shaft power. As the parameters n_q and n_s are unit dependent, the parameters N , Q , H and P should be described in (min^{-1}), ($\text{m}^3\cdot\text{s}^{-1}$), (m) and (kW), respectively.

Existing methodologies for predicting the turbine mode characteristic curves found in the literature are based on different methods. A methodology based on the specific speed-specific diameter charts and on a Hermite spline interpolation provided good results for the turbine mode BEP but errors are substantial for the part load region [11]. Quadratic and cubic polynomial interpolations with constant coefficients for the entire range of n_q values is proposed in [12]. The need for recomputing new coefficients for improving the prediction accuracy in the part load and in the full load region for new experimental data, for instance in [13], in [14] or in [15], has put into question the hypothesis of using constant coefficients for the entire range of n_q values. The computation of empirical coefficients as a function of the unit specific speed is suggested in [16] for improving prediction accuracy. However, the former methodology requires the previous knowledge of the turbine mode BEP, hindering its application based on information provided by pump manufacturers (e.g., rated parameters in the pump mode). Other methodologies based on the modified turbomachine affinity laws [17] or artificial neural networks [18] have been proposed, but results seem to lack accuracy for the operating points far from the BEP, both in part load and full load operation. Also, the lack of repeatability in the prediction accuracy of methodologies applied to new datasets inside the scope of application (see Table 1) are still not clear. Authors in [11] suggest different designs of the pumps and different manufacturers as the main factors.

Additionally, to the prediction of the turbine mode characteristic curves, to the authors' knowledge, there is no methodology for predicting the PAT characteristic curves in the extended operation of the generating mode. These extended operation characteristic curves, which are illustrated in Fig. 1, are fundamental for the analysis of the hydraulic transients in a PAT power plant caused for instance by the start-up and shut-down of the power plant or by the sudden load rejection of the generator [20]. In this research, the extended operation in generating mode includes the quadrants of operation III and IV of a pump (see Fig. 1), which, hereafter, are referred to as the *extended operation* for nomenclature simplicity. The modelling of the PAT performance during a hydraulic transient event is, usually, performed by assuming the extended operation characteristic curves of a pump with the closest value of unit specific speed, if the actual performance of the installed PAT is not available. However, this assumption should be used with caution, as it may introduce errors during the analysis of the hydraulic transients.

The characteristic curves presented in Fig. 1 are described by the variation of the International Electrotechnical Commission (IEC) factors of discharge and torque, Q_{ED} and T_{ED} , respectively, with respect to the

Table 1
Domain of application of empirical formulations for predicting the PAT characteristic curves in the turbine mode.

Reference	Type of pumps	Range of application
[11]	13 end-suction single-stage centrifugal pumps	$n_q = 14.6 - 79.1$
[12]	7 end-suction single-stage centrifugal pumps	$n_q = 14.6 - 55.6$
[13]	7 pumps of several types	$n_q = 10.2 - 19.6$
[14]	1 multi-stage and 1 single-stage centrifugal pump	$n_q = 26.6 - 37.8$
[15]	12 end-suction single-stage centrifugal pumps	$n_q = 9.1 - 64.1$
[16]	113 pumps comprising radial, semi-axial and radial pumps	Not specified
[17]	6 submersible semi-axial pumps	$n_s = 120 - 162$
[18]	32 pumps (see [19])	$N_s = 0.28 - 2.24^{(1)}$

⁽¹⁾ Definition by the authors in [18]: $N_s = \sqrt{Q}/\sqrt[4]{(g\cdot H)^3}$

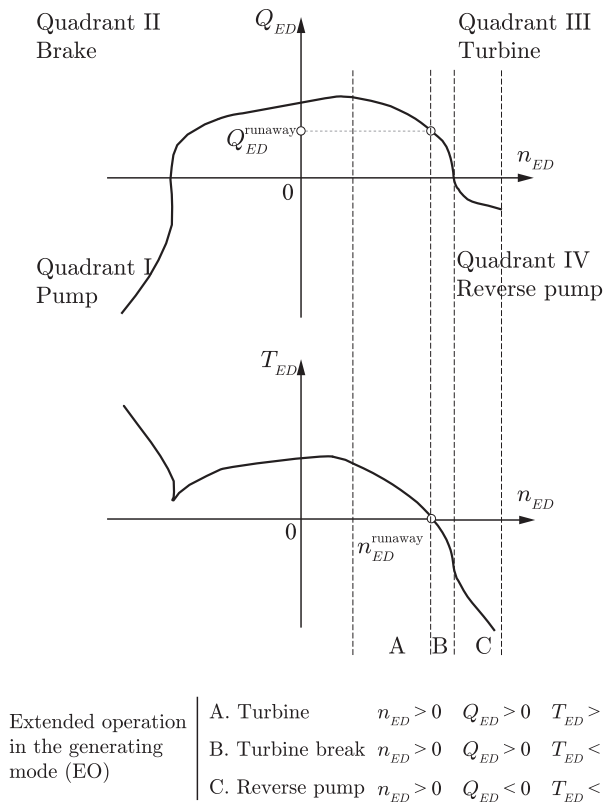


Fig. 1. Four-quadrants of operation of a pump. Identification of the operating zones of interest for the PAT performance modelling of the extended operation in the generating mode, hereafter referred to as *extended operation*. The zones of interest are: Turbine (A); Turbine break (B); and Reverse pump (C), which correspond to the quadrants of operation III and IV.

IEC speed factor n_{ED} , being these parameters defined in Eq. (2) [21]. The IEC power factor P_{ED} is also defined in this equation. These factors provide the condensed information regarding the runner rotating frequency n , the discharge Q , the specific hydraulic energy $E = gH$ (being g the gravity acceleration and H the head), the shaft torque T , the shaft power P , and the impeller diameter D .

$$\begin{aligned}
 n_{ED} &= \frac{nD}{\sqrt{E}} \\
 Q_{ED} &= \frac{Q}{D^2\sqrt{E}} \\
 T_{ED} &= \frac{T}{\rho D^3E} \\
 P_{ED} &= \frac{P}{\rho D^2\sqrt{E^3}}
 \end{aligned} \tag{2}$$

where ρ is the water density.

This paper proposes a new two-step methodology for predicting the characteristic curves of pumps running as turbines and modelling their variable speed operation. The aim of the paper is to provide an engineering methodology for practical application, rather than a scientific theoretical framework. In the first step, empirical continuous surrogate functions are developed based on the Hermite polynomial chaos expansion (PCE) for predicting the characteristic curves, both in the turbine mode and in the extended operation. In the second step, the same PCE method is used to model the variable speed hill chart (HC) of a given PAT, which provides a continuous surrogate function that describes the variation of the parameters such as the specific hydraulic energy E and the efficiency η^T , with respect to the discharge Q and the rotational speed N .

The prediction surrogate functions proposed in the first step aim at overcoming the major challenges found in literature, namely: (i) the requirement of the previous knowledge of the turbine mode BEP, which

is not provided by pump manufacturers; and (ii) considering the effect of the unit specific speed on the shape of the characteristic curves, aiming at improving the prediction accuracy in the part load and the full load operation. The first challenge is addressed by propagating the known IEC characteristic curves in the turbine mode and in the extended operation, obtained from experimental data, inside a range of unit specific speed computed using the rated values in the pump mode n_q^P . In this way, this empirical model has the advantage of only requiring as input data the pump rated parameters and the impeller diameter, which are usually provided by pump manufacturers. The second challenge is addressed by using the bivariate Hermite polynomials as a functional basis of the surrogate functions.

The major difficulty for determining the prediction surrogate functions is the identification of the best set of Hermite polynomials (i.e., the truncated basis cardinality) and the corresponding empirical weighting coefficients applied to each polynomial, such that the surrogate function accurately describes the experimental data. Several fitting metrics are proposed to gauge different candidate models, which are defined by the truncated basis cardinality, aiming at avoiding underfitting and overfitting issues between the collected data and model results. The weighting coefficients are computed by minimising the mean-squared error between the data collected and the model results.

The modelling of the PAT hill chart may be carried out by using the characteristic curves predicted by the previous surrogate functions, or by using data acquired in laboratory conditions. The representation used for this hill chart of the PAT performance allows identifying the discharge and the rotational speed set-points that maximise the efficiency, for a given operating condition (e.g., specific energy available).

The application of the methodology for the PAT characteristic curves prediction is shown for the collected experimental data for three end-suction single-stage centrifugal pumps in both the turbine mode and in the extended operation. The methodology is also validated using data published in the literature. Collected data and published data in the literature are presented in Section 2. The methodology for the PAT performance prediction and the Hermite polynomial chaos expansion are presented in Section 3. The application of the methodology for predicting the turbine mode and the extended operation performance, and for modelling the variable speed hill chart is shown in Section 4. Finally, the main conclusions and future perspectives are drawn in Section 5.

2. Experimental data

Experimental data are collected to measure the turbine mode characteristic curves and the extended operation characteristic curves of three end-suction single-stage closed-impeller centrifugal pumps. The three pumps feature unit specific speed n_q values of 23.1, 41.0 and 67.3, respectively. Additionally, the methodology is validated using the characteristic curve data reported in the literature [15] and the corresponding diameter values reported in [22]. This set of data set is selected, as the pumps are of the same type as the ones from the experimental tests, but from a different manufacturer. Authors provide regression coefficients, which allows the computation of the PATs IEC factors characteristic curves in the turbine mode. No data is available in the literature for the extended operation characteristic curves. The rated values for the pumps experimentally tested and for the pumps reported in the literature are presented in Table 2.

In this research, pumps are identified by the unit specific speed values computed using the rated values in the pumping mode n_q^P . Note that the unit specific speed referent to the turbine mode n_q^T can only be computed after the prediction of the turbine mode BEP has been performed, as this information is not provided by pump manufacturers.

The experimental investigation for collecting the data is performed in the closed-loop test-rig of Instituto Superior Técnico in the University of Lisbon, Portugal. The description of the experimental facility and the measurement equipment can be found in [23]. The characteristic curves

Table 2
Rated characteristics in the pumping mode of the PATs used in this research.

Pump ID	n_q^p (SI)	N_R^p (min^{-1})	Q_R^p ($\text{l}\cdot\text{s}^{-1}$)	H_R^p (m)	P_R^p (kW)	D (mm) ^a
NK40-160/158	23.1	2910	10.9	31.1	5.0	158
NK40-125/127	41.0	2900	13.0	16.2	2.8	127
NK65-125/127	67.3	2910	29.8	14.6	5.5	127
KSB40-335	9.1	1450	7.4	33.0	5.4	330
KSB40-250	12.8	1450	7.0	20.0	2.5	260
KSB50-160	28.7	1450	9.7	8.5	1.2	174
KSB80-220	30.3	1450	24.2	14.5	4.6	219
KSB80-200	34.1	1450	23.2	12.1	3.8	198
KSB100-160	64.1	1200	35.0	5.3	2.3	185

^a D refers to the impeller diameter at the high pressure side.

measured for the turbine mode and for the extended operation are presented in Figs. 2 and 3, respectively. Data presented in these figures is collected for values of rotational speed N between 2100 min^{-1} and 3000 min^{-1} with steps of 300 min^{-1} . All measurements are performed under stationary conditions. The turbine mode characteristic curves cover the part load, the BEP and the full load operating range of the three PATs. The extended operation characteristic curves describe the turbine mode, the turbine break and the reverse pump operating zones (see Fig. 1).

The characteristic curves shown in Figs. 2 and 3 make evidence that the IEC factors allow describing the variable speed performance of PATs using a single characteristic curve. Results provided in these figures also show that the unit specific speed value of the PAT plays a key role in the shape of the characteristic curves, both in the turbine and in the extended operation modes. These results put in evidence that the hypothesis adopted in the literature of considering constant coefficients for describing the performance of different PATs in a given range of unit specific speed values may provide large deviations between prediction model results and actual PAT performance.

3. Hermite polynomial chaos expansion modelling

The bivariate Hermite polynomials are used as a functional basis for the PCE. In this research, the Hermite PCE is used to define prediction surrogate functions that propagate the IEC characteristic curves in between of the unit specific speed values, as well as to model the variable speed hill chart performance of the PAT. Other examples of the PCE method applied to different hydraulic machines and different scopes of application can be found in [24] for the modelling of the analytical hill chart of a counter-rotating microturbine and in [25] for the indirect discharge measurement of a Francis turbine prototype.

The general modelling approach and the rationale of the parameters used for defining the prediction methodology are presented in Section 3.1. The definition of the bivariate Hermite polynomials is presented in Section 3.2, while the methodology for the identification of the surrogate function (i.e., model selection) is shown in Section 3.3.

3.1. Modelling approach

The prediction of the PAT performance aims at providing the functional means to estimate the hydraulic and mechanical characteristic curves for a given unit specific speed value of interest. The key parameters for the modelling of the prediction surrogate functions are thus: (i) the parameters that describe the characteristic curves of the PATs; (ii) the interpolating parameter; and (iii) the parameters that describe the surrogate functions. The rationale for the selection of these key parameters is explained in the following paragraphs.

The PAT characteristic curves are described by the IEC factors n_{ED} , Q_{ED} , T_{ED} and P_{ED} . These parameters are selected for three main reasons: (i) the parameters provide the condensed information in non-

dimensional parameters with respect to the rotational speed n , the discharge Q , the shaft torque T , the shaft power P , the specific hydraulic energy E and the impeller diameter D , (ii) these parameters are obtained by experimental testing the PATs in laboratory conditions. Thus, these parameters are fixed for every experimental dataset; and (iii) by using these parameters for describing the PAT performance, the model does not require the previous knowledge of the turbine mode BEP, as discussed in the Introduction Section.

The unit specific speed computed based on the pump mode rated parameters n_q^p is selected as the interpolating parameter, as results provided in Figs. 2 and 3 suggest a trend between the characteristic curves of the different PATs with respect to this parameter. Also, the unit specific speed computed based on the turbine mode rated values n_q^t can only be computed after the prediction of the turbine mode performance, which would require the previous knowledge of the turbine mode BEP. Nevertheless, results provided in literature seem to point to a linear relationship between n_q^p and n_q^t [26].

Therefore, the continuous surrogate functions for the turbine mode characteristic curves prediction are developed by propagating the $Q_{ED} - n_{ED}$ and the $P_{ED} - n_{ED}$ characteristic curves, in between of the tested unit specific speed values n_q^p . Accordingly, this approach consists of finding two bivariate surrogate functions, namely $f_{QED}^T(n_{ED}, n_q^p) = Q_{ED}$ and $f_{PED}^T(n_{ED}, n_q^p) = P_{ED}$, for the hydraulic and mechanical characteristic curves, respectively. The Q_{ED} and P_{ED} are selected to describe the PAT characteristic curves in the turbine mode as they allow easily computing, afterwards, the turbine mode efficiency $\eta^T = P_{ED}/Q_{ED}$. Note that the shaft torque can be easily computed by $T = P/(2\pi n)$.

Similarly, for the extended operation characteristic curves prediction, the methodology is based on finding the bivariate surrogate functions $f_{QED}^{EO}(n_{ED}, n_q^p) = Q_{ED}$ and $f_{TED}^{EO}(n_{ED}, n_q^p) = T_{ED}$, for the hydraulic and mechanical characteristic curves, respectively. The T_{ED} parameter is selected for the extended operation, as the PAT modelling of hydraulic transient events considers the $T_{ED} - n_{ED}$ characteristic curve (see Fig. 1).

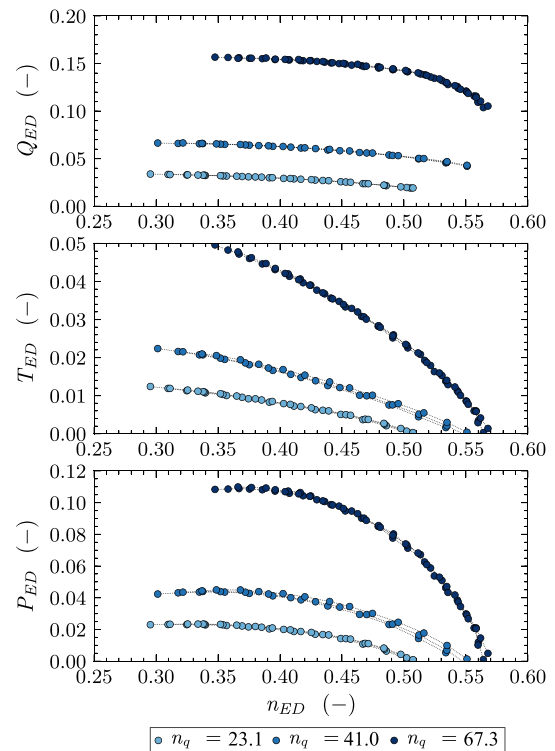


Fig. 2. Characteristic curves of the turbine mode stationary performance of the three PATs experimentally tested described by the variation of the IEC factors Q_{ED} , T_{ED} and P_{ED} with respect to n_{ED} .

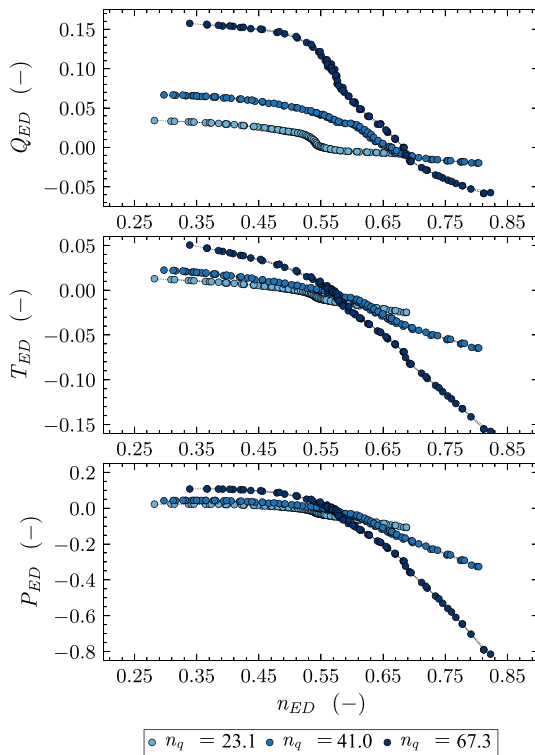


Fig. 3. Characteristic curves of the extended operation performance of the three PATs experimentally tested described by the variation of the IEC factor Q_{ED} , T_{ED} and P_{ED} with respect to n_{ED} .

The Stone-Weierstrass theorem states that there is a surrogate polynomial function f^* , which approximates as closely as possible a continuous real function f defined on a bounded closed interval of real numbers [27]. The continuous real function f is represented, in this application, by the discrete points of the experimental data of the PATs measured in test-rigs. The surrogate function f^* is computed using the polynomial chaos expansion, defined by Eq. (3) [28]. The bivariate Hermite polynomials He_k are chosen for the polynomial basis aiming at describing the effect of the unit specific speed on the characteristic curves shape.

$$f^* = \sum_k \omega \lambda_k He_k \tag{3}$$

where λ_k are the weighting coefficients and He_k are the bivariate Hermite polynomials.

The polynomial basis can only be made of a finite set of polynomials, as there are only a limited number of experimental observations. The main challenge for the identification of each surrogate function consists of finding both the most appropriate truncated basis of polynomials Ψ and the corresponding weighting coefficients λ_k that better represent these experimental observations. Accordingly, the identification of Ψ and λ_k (i.e., the model selection) aims at minimising the error between observations and predictions (i.e., avoid underfitting) and at guaranteeing the realistic behaviour of the surrogate functions (i.e., avoid overfitting). Underfitting issues are characterised by large deviations between measured data y and the response of the surrogate function y^* . Overfitting issues are associated with a surrogate function which exhibits unrealistic oscillations in the function operating domain, despite providing reduced deviations between y and y^* .

3.2. Hermite polynomials basis

The bivariate Hermite polynomials [29] $He_{n,m}$, with indexes $n, m \in \mathbb{N}$ and $m \in \llbracket 2, n + 1 \rrbracket$, being \mathbb{N} the set of natural numbers, can be

recursively defined by Eq. (4) and by the boundary conditions given in Eq. (5).

$$\begin{cases} He_{n+1,1} = a_n \cdot X_1 \cdot He_{n,1} + c_n \cdot He_{n-1,1} \\ He_{n+1,m} = He_{m-1,m} \cdot He_{n-m+2,1} \\ He_{n+1,n+2} = a_n \cdot X_2 \cdot He_{n,n+1} + c_n \cdot He_{n-1,n} \end{cases} \tag{4}$$

$$\begin{cases} He_{0,1} = 1 \\ He_{1,1} = X_1 \\ He_{1,2} = X_2 \end{cases} \tag{5}$$

where a_n and c_n are coefficients computed for each polynomial, which depend on the index $n \in \mathbb{N}$ and are defined in Eq. (6). $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 \sim \mathcal{N}(0, 1)$ are two input independent Gaussian random variables. For notation simplicity, the two input variables are gathered inside the vector $\mathbf{X} = (X_1, X_2)$. The Hermite polynomials are hereafter referred to as He_p with $p \in \mathbb{N}$, using the change of index presented in Eq. (7).

$$\begin{cases} a_n = \sqrt{\frac{1}{n+1}} \\ c_n = -\sqrt{\frac{n}{n+1}} \end{cases} \tag{6}$$

$$p = \frac{n(n+1)}{2} + m - 1 \tag{7}$$

The Hermite polynomials family He_p form a complete basis of functions inside the Hilbert space that are orthogonal with respect to the scalar product [30] defined by:

$$\langle He_A, He_B \rangle = \int_{\mathbb{R}^2} He_A(\mathbf{X}) \cdot He_B(\mathbf{X}) \cdot p(\mathbf{X}) d\mathbf{X} \tag{8}$$

where \mathbb{R} is the set of real numbers, $He_A(\mathbf{X})$ and $He_B(\mathbf{X})$ are any polynomial of the He_p family and $p(\mathbf{X})$ is the standardised bivariate probability density function defined by Eq. (9). The graphical representation of the first 15 polynomials of the family He_p is depicted in Fig. 4. Both indexes representations are shown in the figure, though only the p notation is used hereafter.

$$p(\mathbf{X}) = \frac{1}{2\pi} \exp\left(-\frac{\mathbf{X}^2}{2}\right) \tag{9}$$

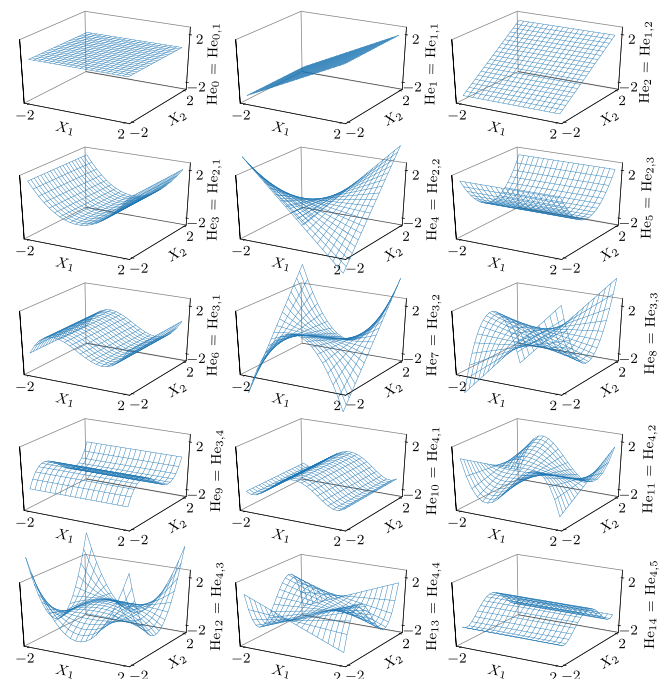


Fig. 4. First 15 bivariate polynomials of the Hermite polynomial family He_p . Adapted from [25].

The input data vector $\mathbf{x} = (x_1, x_2)$ collected during the experimental tests is usually not suitable for the Hermite PCE modelling. The polynomials He_p are only orthogonal with respect to the standardised bivariate normal probability density function. Thus, the standardisation function $g(\mathbf{x})$, defined in Eq. (10), is applied to the physical input data. The coefficients a, b and c and the vector $\boldsymbol{\mu}$ of $g(\mathbf{x})$ are computed by Eqs. (11)–(13). The first equation guarantees the average values $\bar{X}_1 = 0$ and $\bar{X}_2 = 0$. The second equation ensures the standard variation $\sigma_{x_1} = 1$. The solution of the system of equations of Eq. (13) satisfies the standard variation $\sigma_{x_2} = 1$ and the covariance $\text{cov}(X_1, X_2) = 0$.

$$g: \begin{cases} \mathbb{R}^2 & \mapsto & \mathbb{R}^2 \\ \mathbf{x} & \mapsto & \mathbf{X} = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \cdot (\mathbf{x} - \boldsymbol{\mu}) \end{cases} \quad (10)$$

$$\boldsymbol{\mu} = \frac{1}{N_s} \sum_{j=1}^{N_s} (j)\mathbf{x} \quad (11)$$

$$a = \sqrt{\frac{1}{\frac{1}{N_s-1} \sum_{j=1}^{N_s} ((j)x_1 - \mu_1)^2}} \quad (12)$$

$$\begin{cases} \frac{1}{N_s-1} \sum_{j=1}^{N_s} [b^{(j)}x_1 - \mu_1 + c^{(j)}x_2 - \mu_2]^2 = 1 \\ \sum_{j=1}^{N_s} (j)X_1 \cdot [b^{(j)}x_1 - \mu_1 + c^{(j)}x_2 - \mu_2] = 0 \end{cases} \quad (13)$$

where N_s is the total number of measurement samples (i.e., experimental observations) and the superscript $(k) \in \llbracket 1, N_s \rrbracket$ refer to the number of the measurement sample.

3.3. Identification of the surrogate function

For a given truncated basis of Hermite polynomials selected *a priori* $\Psi = \{\text{He}_0, \text{He}_1, \dots, \text{He}_{p_{\max}}\}$, the response of the surrogate function y^* is defined by Eq. (14). The weighting coefficients λ_k are computed based on minimising the absolute mean-squared error between experimental observations \mathbf{y} and the response of the surrogate function y^* , $1/N_s \sum_{i=1}^{N_s} (y - y^*)^2$. The solution to this problem is given by the least-square solution, given by Eq. (15) [31]. This solution is only valid for an overdetermined system of equations such that $N_s > p_{\max}$, where p_{\max} is the index of the maximum order polynomial of the family He_p of the truncated basis Ψ . This p_{\max} index defines the cardinality of the truncated basis.

$$y^* = f^*(\mathbf{x}) = \sum_{k \in \Psi} \lambda_k \cdot \text{He}_k(g(\mathbf{x})) \quad (14)$$

$$\boldsymbol{\lambda} = [\lambda_0 \ \lambda_1 \ \dots \ \lambda_{p_{\max}}]^T = (\mathbf{He}^T \mathbf{He})^{-1} \mathbf{He}^T \mathbf{y} \quad (15)$$

where $\boldsymbol{\lambda} \in \mathbb{R}^{p_{\max} \times 1}$ is the vector of the weighting coefficients, $\mathbf{He} \in \mathbb{R}^{N_s \times p_{\max}}$ is the matrix of the Hermite polynomials evaluated in the domain of \mathbf{X} defined in Eq. (16) and $\mathbf{y} \in \mathbb{R}^{N_s \times 1}$ is the vector of the experimental observations defined in Eq. (17).

$$\mathbf{He} = \begin{bmatrix} \text{He}_0^{(1)}(\mathbf{X}) & \text{He}_1^{(1)}(\mathbf{X}) & \dots & \text{He}_{p_{\max}}^{(1)}(\mathbf{X}) \\ \text{He}_0^{(2)}(\mathbf{X}) & \text{He}_1^{(2)}(\mathbf{X}) & \dots & \text{He}_{p_{\max}}^{(2)}(\mathbf{X}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{He}_0^{(N_s)}(\mathbf{X}) & \text{He}_1^{(N_s)}(\mathbf{X}) & \dots & \text{He}_{p_{\max}}^{(N_s)}(\mathbf{X}) \end{bmatrix} \quad (16)$$

$$\mathbf{y} = [{}^{(1)}y \ {}^{(2)}y \ \dots \ {}^{(N_s)}y]^T \quad (17)$$

In this research, the truncated basis of each candidate model contains all polynomials between the index $p = 0$ up to p_{\max} . Sparse solutions are not taken into account (i.e., values of $\lambda_k = 0$), as this is a relatively simple problem of model selection. The choice of the truncated basis is performed by comparing several fitting metrics computed for the different candidate models. Metrics used are the maximum absolute error (Max. AE), the mean absolute error (Mean AE), the

standard deviation of the error σ_e (see Eq. 18), for evaluating the underfitting, as well as the determination coefficient R^2 (see Eq. 19) and the Akaike Information Criterion (AIC) [32], for evaluating overfitting. The latter estimates the maximum likelihood of a given model obtained using the least squares estimation and is penalised by the number of model parameters aiming at finding overfitted models (see Eq. 20). The AIC metric is further corrected (AIC_c) according to Eq. (21) to account for the number of parameters with respect to the existing number of samples [33]. As the results given by the AIC metric are dependent on the measurement samples, the absolute values of this metric do not provide relevant information. Results of the candidate models are further re-scaled with respect to the model with the minimum AIC_c value, namely $\text{AIC}_{c,s} = \text{AIC}_c - \text{AIC}_{c,\min}$.

$$\sigma_e(y, y^*) = \sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} (y_i - y_i^*)^2} \quad (18)$$

$$R^2(y, y^*) = 1 - \frac{\sum_{i=1}^{N_s} (y_i - y_i^*)^2}{\sum_{i=1}^{N_s} (y_i - \bar{y})^2} \quad (19)$$

$$\text{AIC}(y, y^*) = N_s(\log \sigma_e^2) + 2p_{\max} \quad (20)$$

$$\text{AIC}_c(y, y^*) = \text{AIC} + \frac{2p_{\max}(p_{\max} + 1)}{N_s - p_{\max} - 1} \quad (21)$$

The choice of the truncated basis Ψ for describing the surrogate function is performed by plotting the previous metrics with respect to the maximum index of the Hermite polynomial basis p_{\max} . Best models should minimize both the maximum absolute error, the average absolute error, the standard deviation of the error and the AIC_{c,s} while maximizing the R^2 value. The final model is selected and plotted to verify if underfitting or overfitting have occurred. This process is further illustrated in the application of the methodology.

Overall, this set of fitting metrics has proved effective and of a simple application for identifying underfitted and overfitted models and gauging between different candidate models in this particular

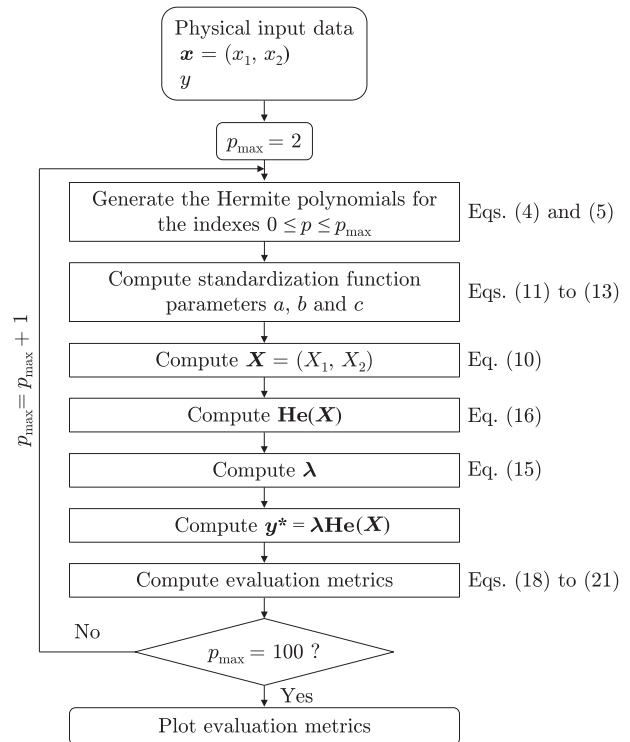


Fig. 5. Flow chart for the identification of the truncated basis for describing the surrogate function of a given variable of interest y with respect to the interpolation domain $\mathbf{x} = (x_1, x_2)$.

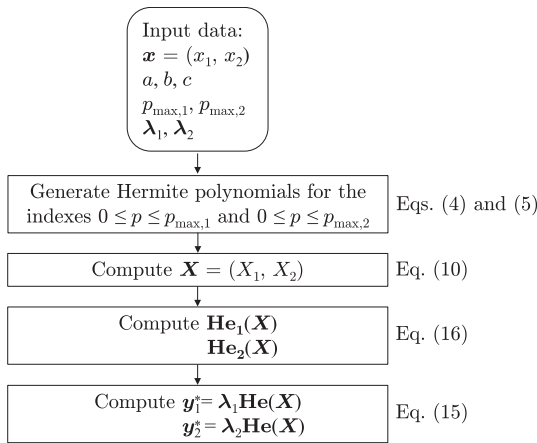


Fig. 6. Flow chart for computing the surrogate functions for the identified hydraulic and mechanical surrogate functions.

application. The polynomials in this application only have two dimensions, which makes model selection relatively simple. Other model identification methods more complex to implement and usually applied to higher dimensionality problems, such as the Least Angle Regression [34] or the Least Absolute Shrinkage and Selection Operator [35] could be used. However, the discussion of the most suitable method for model identification falls outside the scope of this research.

The flow chart for identifying the surrogate function truncated basis and the corresponding weighting coefficients for a given measured quantity y is presented in Fig. 5. The initial value of $p_{max} = 2$ is to take into account all polynomials from the boundary condition (see Eq. 5) in the first iteration. The maximum value of $p_{max} = 100$ may be increased, though this value is found suitable for this research, as the p_{max} values for the different surrogate functions analysed in the next Sections vary between 10 and 35.

After the identification of the truncated basis, the empirical surrogate functions for the propagation of the hydraulic and mechanical characteristic curves are computed based on the flow chart presented in Fig. 6. The input data are: (i) the variables used for defining the interpolating domain $\mathbf{x} = (x_1, x_2)$, which for the prediction methodology are $\mathbf{x}(n_{ED}, n_q^p)$; (ii) the values of the coefficients a, b and c and μ of the standardisation function $g(\mathbf{x})$; (iii) the truncated basis defined by the $p_{max,1}$ and the $p_{max,2}$ values and the corresponding vectors of the weighting coefficients λ_1 and λ_2 , for the identified hydraulic and

mechanical characteristic curves, respectively.

The identified surrogate functions allow estimating the hydraulic and mechanical characteristic curves, described by the IEC factors, by using as input data the value of the pump mode unit specific speed n_q^p for a given PAT of interest for the prediction, and the range of n_{ED} for which the characteristic curves are to be computed.

In the case of the turbine mode, the parameters of E, Q and η^T can be computed from the $n_{ED} - Q_{ED}$ and the $n_{ED} - P_{ED}$ characteristic curves by using the Eq. (22) for a given value of the runner rotating frequency n .

$$\begin{aligned}
 E &= \frac{n^2 D^2}{n_{ED}^2} \\
 Q &= n D^3 \frac{Q_{ED}}{n_{ED}} \\
 \eta^T &= \frac{P_{ED}}{Q_{ED}}
 \end{aligned}
 \tag{22}$$

The variable speed hill chart performance of the PAT defined by the variation of E and η^T with respect to N and Q can now be modelled using the procedures described in Fig. 5 for the model identification and Fig. 6 for computing each surrogate function, using as input data $\mathbf{x} = (N, Q)$ for the variables of interest $y = E$ and $y = \eta^T$. This variable speed hill chart of the PAT can be computed using the turbine mode prediction results provided by the surrogate functions or by using precise performance data of a PAT, given by a pump manufacturer or acquired in laboratory conditions.

4. Results and discussion

This Section shows the results and discussion of the application of the new two-step methodology for the PATs characteristic curves prediction and for the modelling of the variable speed hill chart. The identification of the truncated basis for a given surrogate function is illustrated in Section 4.1. The identification of the surrogate functions for the prediction of the turbine mode and for the extended operation are presented in Sections 4.2 and 4.3, respectively. Finally, the proposed hill chart model to describe the variable speed performance of a given PAT is presented in Section 4.4. The summary of the surrogate functions for the prediction of the turbine mode performance, the prediction of the extended operation and the modelling of the performance hill chart is given in Table 3.

Table 3
Surrogate functions for the prediction and the modelling of the PAT performance.

Operating mode	Surrogate function	Dataset	Operating domain	Variable of interest	Section
Turbine mode (T)	f_{QED}^T	Exp. data	$\mathbf{x} = (n_{ED}, n_q)$	$y = Q_{ED}$	4.2
	f_{PED}^T	Exp. data	$\mathbf{x} = (n_{ED}, n_q)$	$y = P_{ED}$	4.2
	$f_{QED,B}^T$	Lit. data	$\mathbf{x} = (n_{ED}, n_q)$	$y = Q_{ED}$	4.2
	$f_{PED,B}^T$	Lit. data	$\mathbf{x} = (n_{ED}, n_q)$	$y = P_{ED}$	4.2
Extended operation (EO)	f_{QED}^{EO}	Exp. data	$\mathbf{x} = (n_{ED}, n_q)$	$y = Q_{ED}$	4.3
	f_{TED}^{EO}	Exp. data	$\mathbf{x} = (n_{ED}, n_q)$	$y = T_{ED}$	4.3
Hill chart performance (HC)	f_E^{HC}	$n_q = 23.1$	$\mathbf{x} = (N, Q)$	$y = E$	4.4
	f_η^{HC}	$n_q = 23.1$	$\mathbf{x} = (N, Q)$	$y = \eta^T$	4.4
	f_E^{HC}	$n_q = 41.0$	$\mathbf{x} = (N, Q)$	$y = E$	4.4
	f_η^{HC}	$n_q = 41.0$	$\mathbf{x} = (N, Q)$	$y = \eta^T$	4.4
	f_E^{HC}	$n_q = 67.3$	$\mathbf{x} = (N, Q)$	$y = E$	4.4
	f_η^{HC}	$n_q = 67.3$	$\mathbf{x} = (N, Q)$	$y = \eta^T$	4.4

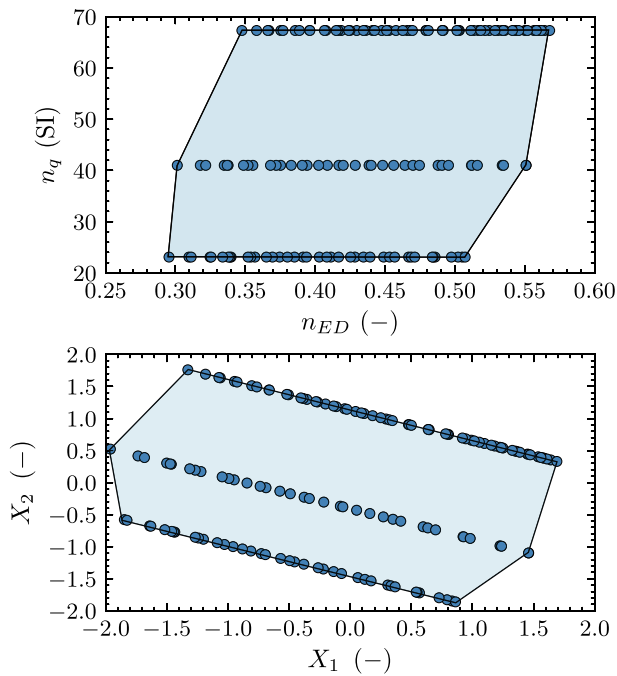


Fig. 7. Application of the standardisation function $g(\mathbf{x})$ to a physical input data set: (a) measured data $\mathbf{x} = (n_{ED}, n_q^P)$; (b) standardised data $\mathbf{X} = (X_1, X_2)$. The convex hull is indicated by the blue area.

4.1. Identification of the truncated basis

The application of the methodology for the model selection described in the flow chart of Fig. 5 is illustrated for the hydraulic surrogate function using the collected experimental data f_{QED}^T . The data transformation given by the standardisation function $g(\mathbf{x})$ is illustrated in Fig. 7. The evaluation of the surrogate function is only performed inside the convex hull of the input data, as shown in Fig. 7, for avoiding issues related with extrapolated results.

The comparison of the values computed for the fitting metrics of different candidate models to describe the surrogate function $f_{QED}^T(n_{ED}, n_q^P)$ is illustrated in Fig. 8. The use of truncated basis with $p_{max} > 24$ does not provide a decrease in the errors between the measured data Q_{ED} and the evaluation of the surrogate function Q_{ED}^* . Also, the $AIC_{c,s}$ criteria starts to increase for truncated basis with $p_{max} > 24$, which indicates possible overfitting.

The comparison between an underfitted model, the chosen model and an overfitted model is illustrated in Fig. 9. These candidate models correspond to p_{max} values of 5, 24 and 40, respectively. The underfitted model exhibits a large deviation between the surrogate function results and the experimental data. The overfitted model shows that despite the

reduced values of the maximum, average and the standard deviation of the error (see Fig. 8), the model provides significant oscillations in-between of the tested unit specific speed values and does not provide a smooth propagation between of the know characteristic curves. If this surrogate function would be used estimate the $Q_{ED} - n_{ED}$ characteristic curve, for instance for a pump with $n_q^P = 50.0$, model results would not feature a realistic behaviour of increasing Q_{ED} values with decreasing n_{ED} values (see Fig. 2). The chosen model, in this example, corresponds to the minimum $AIC_{c,s}$ value. This is not a mandatory criterion as the model corresponding to the $AIC_{c,min}$ value may feature overfitting. If overfitting occurs, a truncated basis with lower cardinality should be selected, considering the trade-off between the model accuracy and the expected likelihood of the surrogate function.

4.2. Turbine mode prediction surrogate functions

The identification of the turbine mode prediction surrogate functions is performed separately for the collected experimental data and the literature data [15]. Data preliminary analysis indicates no clear trend for the propagation of the characteristic curves between the PATs of the two datasets (i.e., different manufacturers). This trend is, however, easily observed if different surrogate functions are applied separately to each group of PATs of different manufacturers. The truncated basis of the identified surrogate functions and the corresponding values computed for the fitting metrics of the identified surrogate functions are indicated in Table 4. These surrogate functions are illustrated in Fig. 10 using a three dimensional (3D) view and, in Fig. 11 using a contour view.

Results provided in Figs. 10 and 11 make evidence that the Hermite PCE method proves effective to model the effect of the unit specific speed on the shape of the characteristic curves. The surrogate functions fit the different shapes of the PAT characteristic curves data (e.g., sharper Q_{ED} decrease with increasing n_{ED} for PATs with higher n_q^P values), while providing a smooth propagation between the unit specific speed values. In fact, the recursive definition of the Hermite polynomials (see Eq. 4) implies that the empirical coefficients are modified, depending on the constant n_q^P and consequently the X_2 values.

Also, results presented in Figs. 10 and 11 clearly show that the propagation of the characteristic curves between the two different datasets is not trivial. Firstly, the n_{ED} range of operation of PATs with similar n_q^P values is slightly different (e.g., n_{ED} values between 0.316 and 0.568 for the $n_q^P = 67.3$ PAT and n_{ED} values between 0.374 and 0.601 for the $n_q^P = 64.1$ PAT). Secondly, for the same n_q^P values, the surrogate functions found for the literature data provide higher values for both Q_{ED} and P_{ED} parameters. Nevertheless, the surrogate functions identified for the different manufacturers have approximately the same trend, which is clearly observed by the comparison of the isolines depicted in Fig. 11. This fact indicates that the Hermite PCE methodology proves effective for the propagation of the IEC characteristic curves if the surrogate functions are built independently for each pump manufacturer.

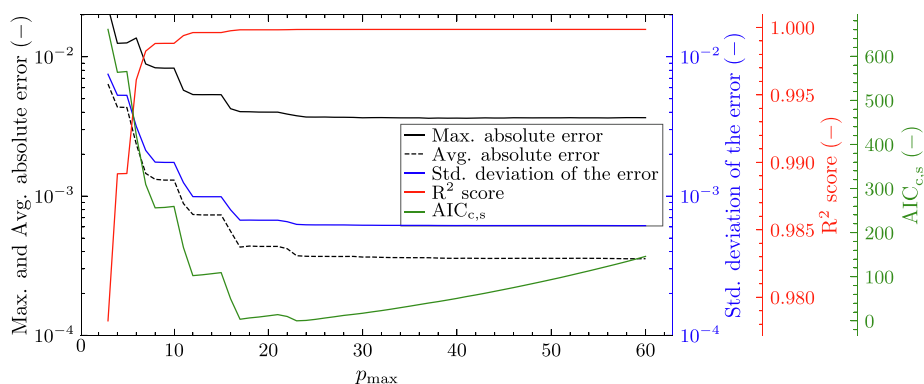


Fig. 8. Metrics evaluation for determining the surrogate function $f_{QED}^T(n_{ED}, n_q)$ applied to the experimental data.

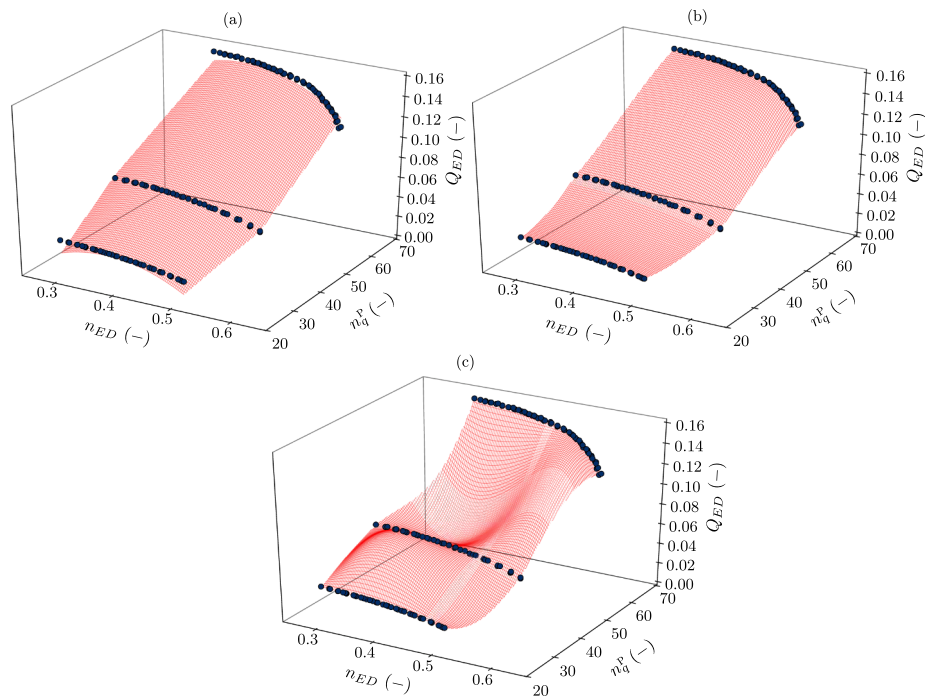


Fig. 9. Comparison of the truncated basis cardinality for computing the surrogate function f_{QED}^T : (a) underfitted model ($p_{max} = 5$); (b) chosen model ($p_{max} = 24$); (c) overfitted model ($p_{max} = 40$). Each candidate model is represented by the maximum index of the truncated basis p_{max} . Experimental data is indicated by the scatter points. Surrogate function results are represented by the red mesh.

Table 4
Truncated basis identified and fitting metrics values of the hydraulic and mechanical surrogate functions for the turbine mode prediction. Note that Q_{ED} and P_{ED} are dimensionless.

Parameter	f_{QED}^T	f_{PED}^T	$f_{QED,B}^T$	$f_{PED,B}^T$
p_{max}	23	17	18	18
Max. AE $\times 10^{-3}$ (-)	3.64	7.47	6.52	3.86
Mean AE $\times 10^{-3}$ (-)	0.04	0.87	0.58	0.67
$\sigma_e \times 10^{-3}$ (-)	0.64	1.40	0.96	1.02
R^2 (-)	0.999	0.998	0.999	0.998

The hypothesis of pumps of the same type from different manufacturers and with the same rated parameters featuring the same characteristic curves is likely to be far-fetched. Different manufacturers may use distinct component designs (e.g., impeller, volute, bearings) and/or materials,

which affect the turbine mode performance of each PAT. Therefore, the application of this methodology should be restricted to pumps of the same manufacturer, if possible, aiming at minimising the deviations between predictions and actual PAT performance. This restriction is a drawback compared to more general methodologies (e.g., [16]), as the application of the surrogate functions is restricted to pumps of the same type and the same manufacturer. One proposal for overcoming such drawback is the development of a surrogate functions database for different pump manufacturers and pump types. Nevertheless, the application of this prediction methodology only requires input data provided by pump manufacturers, as opposed to, for instance, the methodology provided in [16], which require the previous estimation of the turbine mode BEP.

Finally, the comparison of the evaluation of the surrogate functions with the original data from both datasets for the hydraulic and mechanical characteristic curves is presented in Figs. 12 and 13, respectively. Surrogate functions results are computed using the procedure presented in Fig. 6. Results provided in these figures make evidence

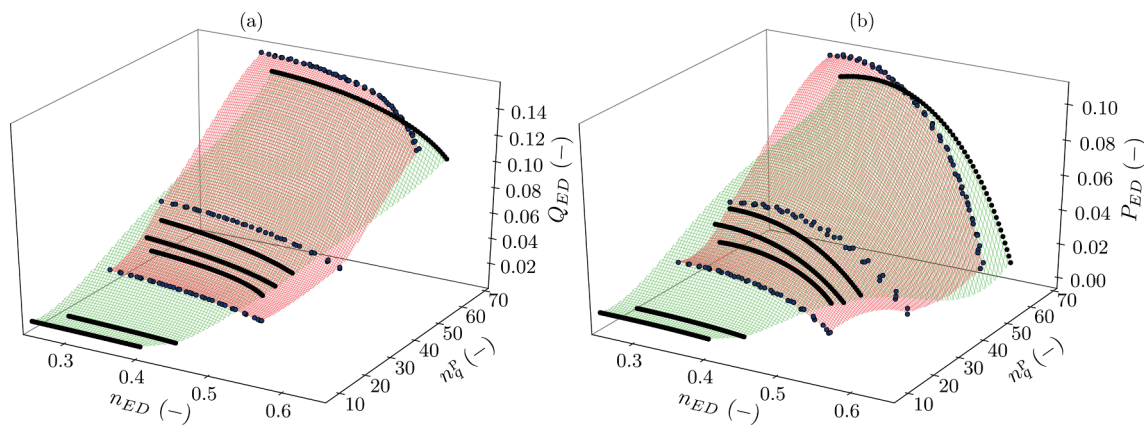


Fig. 10. 3D view of the identified surrogate functions for the propagation of the: (a) hydraulic and the (b) mechanical characteristic curves. The red mesh refers to the surrogate functions found for the collected experimental data f_{QED}^T and f_{PED}^T . The green mesh refers to the surrogate functions found for the literature data $f_{QED,B}^T$ and $f_{PED,B}^T$.

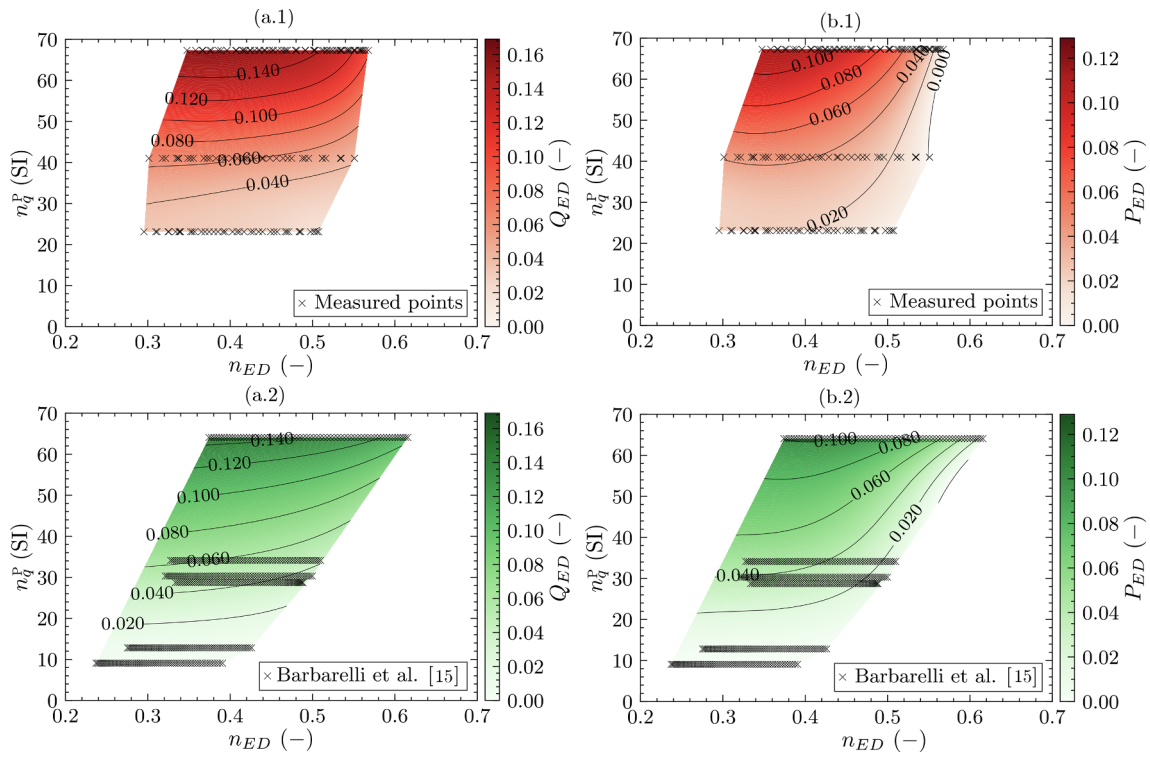


Fig. 11. Contour view of the identified surrogate functions for the propagation of the hydraulic and mechanical characteristic curves referent to the turbine mode. The red contours indicate the surrogate functions applied to the collected experimental data. The green contours indicate the evaluation of the surrogate functions applied to the literature data. (a.1) $f_{Q_{ED}}^T$; (b.1) $f_{P_{ED}}^T$; (a.2) $f_{Q_{ED,B}}^T$; (b.2) $f_{P_{ED,B}}^T$.

that the surrogate function results accurately describe the data of the PATs characteristic curves. The major difference is observed for the $n_q = 9.1$ PAT in the part load region, which could be explained by two reasons.

Firstly, the computation of the weighting coefficients λ_k by minimising the absolute mean-squared error $1/N_s \sum_{i=1}^{N_s} (y - y^*)^2$ (see Eq. 15), favours the surrogate functions fitting to higher values of the Q_{ED} and P_{ED} parameters and, consequently, to PATs with higher n_q^p values. For instance, the maximum absolute errors $\Delta Q_{ED}^{max} = \max(|Q_{ED}^{model} - Q_{ED}^{exp}|)$ of 0.0003 and 0.0004, respectively for the $n_q = 9.1$ and the $n_q = 34.1$

PATs, result in a maximum relative error $\delta Q_{ED}^{max} = \max(|Q_{ED}^{model} - Q_{ED}^{exp}|/Q_{ED}^{exp})$ of 9.6% and 0.78%, respectively for the two PATs. Secondly, the scope of radial impeller designs for pumps with n_q^p values between 9.1 and 64.1 is very wide [36]. This wide range of centrifugal impeller designs feature different shapes of characteristic curves, which may affect the accurate propagation of the characteristic curves in this unit specific speed range. This higher deviation between model predictions and experimental data for lower n_q^p values is a limitation of the methodology and results should be used carefully. Nevertheless, good fittings are observed for the remaining PATs.

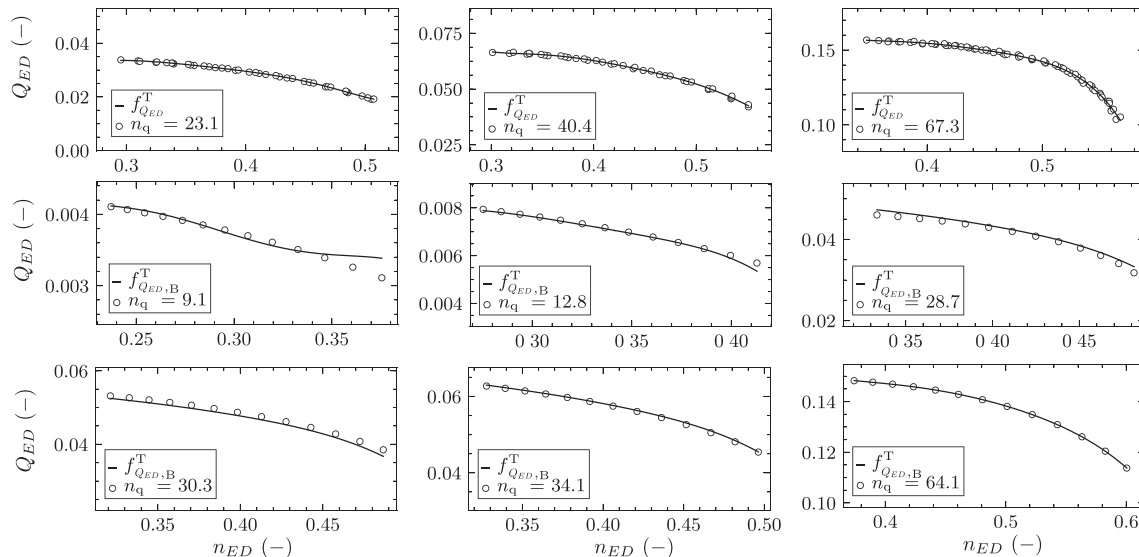


Fig. 12. Comparison between the experimental data, literature data and the evaluation of the hydraulic surrogate functions $f_{Q_{ED}}^T$ and $f_{Q_{ED,B}}^T$ (see Table 3 for the nomenclature of the surrogate functions). PATs are identified by the n_q^p values (see Table 2). Scatter points indicated experimental and literature data. Lines indicate the results of the surrogate functions.

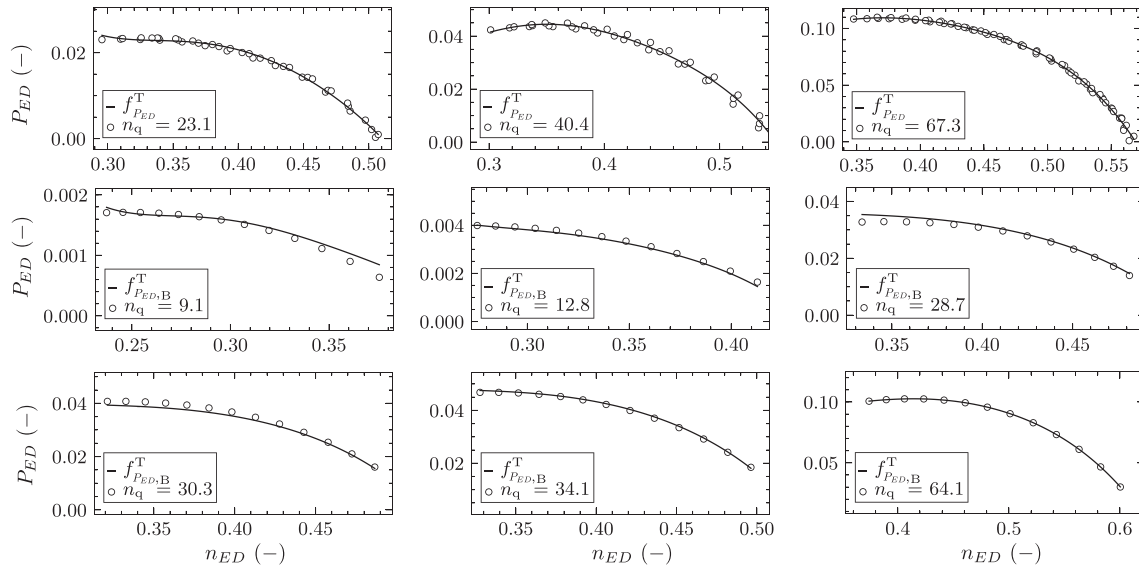


Fig. 13. Comparison between the experimental data, literature data and the evaluation of the mechanical surrogate functions f_{PED}^T and $f_{PED,B}^T$ (see Table 3 for the nomenclature of the surrogate functions). PATs are identified by the n_q^P values (see Table 2). Scatter points indicated experimental and literature data. Lines indicate the results of the surrogate functions.

Table 5

Truncated basis identified and fitting metrics values of the hydraulic and mechanical surrogate functions for the extended operation prediction. Note that Q_{ED} and T_{ED} are dimensionless.

Parameter	f_{QED}^{EO}	f_{TED}^{EO}
P_{max}	32	19
Max. AE $\times 10^{-3}$ (-)	9.20	21.18
Mean. AE $\times 10^{-3}$ (-)	2.58	4.64
$\sigma_e \times 10^{-3}$ (-)	3.26	5.96
R^2 (-)	0.994	0.998

The evaluation of the turbine mode surrogate functions, for a given value of unit specific speed of interest, as illustrated in the Figs. 12 and 13 provide the $n_{ED} - Q_{ED}$ and $n_{ED} - P_{ED}$ characteristic curves. Afterwards, Eq. (22) should now be applied to compute the parameters E , Q and η^T for a given constant value of N for obtaining the stationary performance of the PAT in the turbine mode.

4.3. Extended operation prediction surrogate functions

The same procedure is applied for determining the surrogate functions for the characteristic curves prediction in the extended operation with respect to the collected experimental data. The truncated basis and the values of the metrics for the surrogate functions for the propagation of the hydraulic and mechanical characteristic curves in the extended operation are presented in Table 5. These surrogate functions are illustrated in Fig. 14. Results in this figure make evidence that the Hermite PCE also proves effective for the smooth propagation of the characteristic curves inside the interval of tested unit specific speed values.

The comparison of the metrics values obtained for the turbine mode (Table 4) and for the extended operation (Table 5) show that the fitting errors regarding the latter are higher, while the R^2 coefficient values are lower. This is an expected result, as the interpolation of extended operation is more complex compared to the turbine mode, namely because of the higher curvature associated with the turbine break and reverse pump operating areas (i.e., the s-curve – see Fig. 1).

The comparison of the hydraulic and mechanical surrogate function results with the collected experimental data are shown in Figs. 15. Results presented in this figure show that the surrogate functions accurately describe extended operation characteristic curves, though a

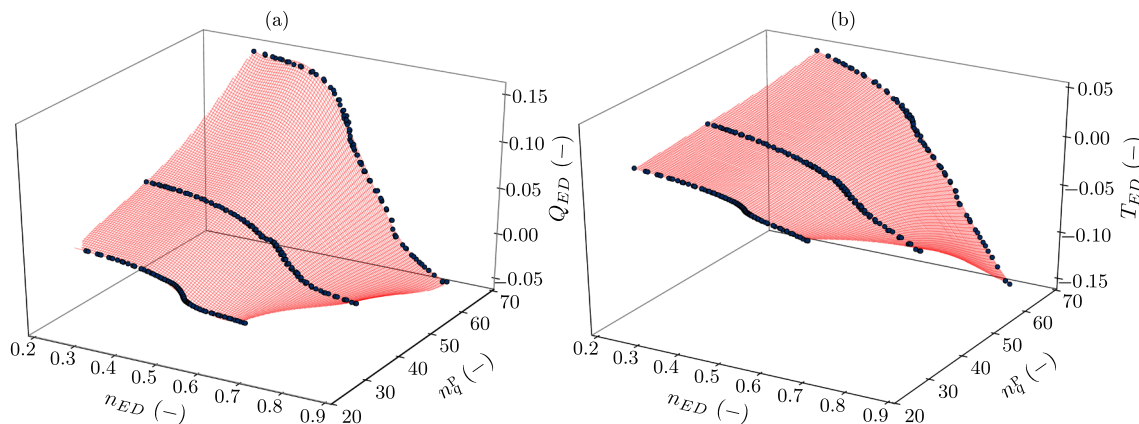


Fig. 14. 3D view of the identified surrogate functions for the propagation of the hydraulic (a) and the (b) mechanical characteristic curves referent to the extended operation. The red mesh refers to the surrogate functions found for the experimental data f_{QED}^{EO} and f_{TED}^{EO} .

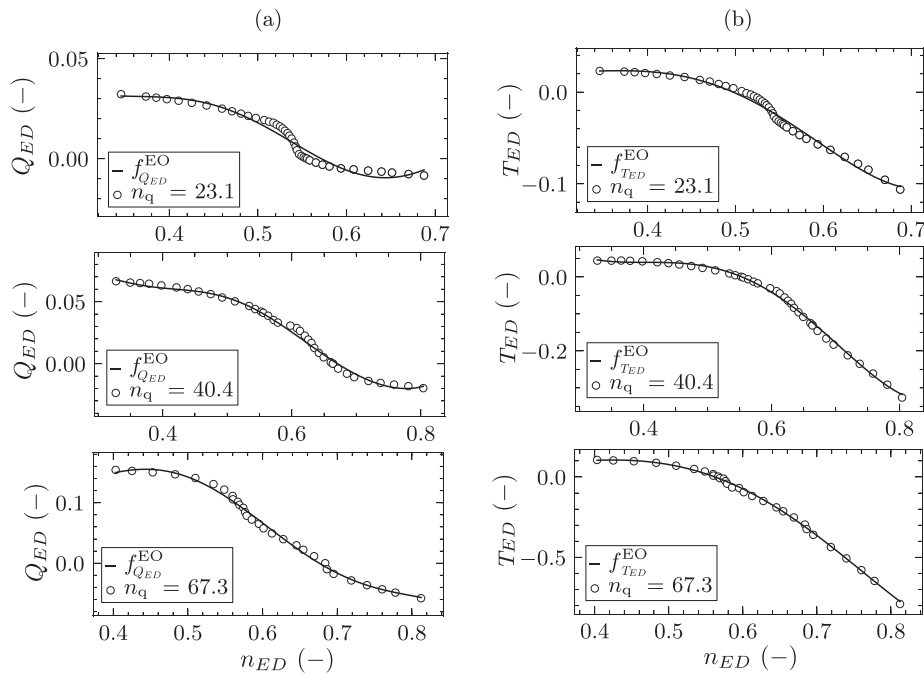


Fig. 15. Comparison between the experimental data and the evaluation of the: (a) hydraulic surrogate function f_{QED}^{EO} and; (b) mechanical surrogate function f_{TED}^{EO} (see Table 3 for the nomenclature of the surrogate functions). PATs are identified by the n_q^p values (see Table 2). Scatter points indicated experimental and literature data. Lines indicate the results of the surrogate functions.

large deviation is observed for the minimum unit specific speed $n_q^p = 23.1$ pump for n_{ED} values between 0.500 and 0.580. This operating region, near the runaway conditions (i.e., $T_{ED} = 0$) is specifically relevant for the hydraulic transients modelling caused by the sudden load rejection of the generator. As discussed in Section 4.2, this larger deviation could be partially explained by the computation of the λ_k coefficients favouring the adjustment of the surrogate function to higher n_q^p values. Further research is suggested for improving the accuracy of the surrogate functions in this operating mode.

4.4. Variable speed hill chart modelling

The modelling of the variable speed hill chart of the PAT aims at providing a continuous surrogate function that describes the variation of the variables of interest, such as the specific hydraulic energy E or the efficiency η^T with respect to the discharge Q and the rotational speed N . Other variables of interest such as the power or the shaft torque can be modelled. This modelling can be performed by using as input data the characteristic curves estimated using the surrogate functions for the turbine mode, or by using precise data obtained in laboratory conditions or given by the pump manufacturers (if available). Despite the Hermite PCE method has proved effective for describing the surrogate functions, the prediction results should only be used during preliminary engineering designs. The final design of the micro hydropower plant should be performed using precise data for the PAT characteristic curves.

The procedures presented in the flow charts of Figs. 5 and 6 are used to model the variable speed hill chart performance of the PATs experimentally tested, using as operating domain $x = (N, Q)$ and using the variables of interest E and η^T . The experimental data used consists of the measurements performed for values of rotational speed between 1200 min^{-1} and 3000 min^{-1} . The identified truncated basis and the respective values computed for the fitting metrics are presented in Table 6. The 3D views of the surrogate functions for describing the specific hydraulic energy E and the PAT efficiency η^T with respect to the discharge Q and rotational speed N are illustrated in Fig. 16 for the $n_q^p = 23.1$ pump. The contour view of the proposed hill chart model for

describing the PATs variable speed operation are presented in Figs. 17–19, respectively for the $n_q^p = 23.1$, $n_q^p = 41.0$ and the $n_q^p = 67.3$ pumps.

Results provided in Fig. 16 for the $n_q^p = 23.1$ pump shows that the continuous surrogate functions describe the discrete experimental data without significant oscillations. Also, the standard deviation of the error of the surrogate functions for the hill chart modelling (see Table 6) is in the range of $0.56 \text{ J}\cdot\text{kg}^{-1}$ and $0.77 \text{ J}\cdot\text{kg}^{-1}$, for the specific hydraulic energy; and in the range of 0.40% to 0.47% for the PAT efficiency (see Table 6). The order of magnitude of data collected for the three PATs experimentally tested range from $50 \text{ J}\cdot\text{kg}^{-1}$ to $500 \text{ J}\cdot\text{kg}^{-1}$ for the specific hydraulic energy; and from 0% to 75% for the efficiency in turbine mode (see Figs. 17–19). Thus, the reduced standard deviation of the error, together with the non-existence of significant oscillations in the continuous surrogate functions obtained, prove that the Hermite PCE method is effective for modelling the hill chart for describing the variable speed operation of a PAT.

This representation of the variable speed operation using this hill chart model has two main advantages. Firstly, it allows identifying the best efficiency ridgeline, as illustrated in Figs. 17–19. This best efficiency ridgeline is of particular interest, as it provides the discharge and rotational speed set-points, for a given available specific hydraulic energy, which maximises the PAT efficiency. This set-points identification

Table 6

Truncated basis identified and fitting metrics values of the surrogate functions for modelling the variable speed hill chart of the PATs experimentally tested. Results of the fitting metrics of the surrogate functions referring to E are in $\text{J}\cdot\text{kg}^{-1}$, while the ones referring to η^T are in %.

Pump ID Parameter	$n_q = 23.1$		$n_q = 41.0$		$n_q = 67.3$	
	f_E^{HC}	f_η^{HC}	f_E^{HC}	f_η^{HC}	f_E^{HC}	f_η^{HC}
P_{max}	14	25	10	20	21	28
Max. AE (units)	2.7	1.30	1.86	1.57	2.9	1.82
Mean. AE (units)	0.56	0.30	0.40	0.29	0.9	0.34
σ_e (units)	0.77	0.40	0.56	0.43	0.59	0.47
R^2 score (-)	0.999	0.999	0.999	0.999	0.999	0.999

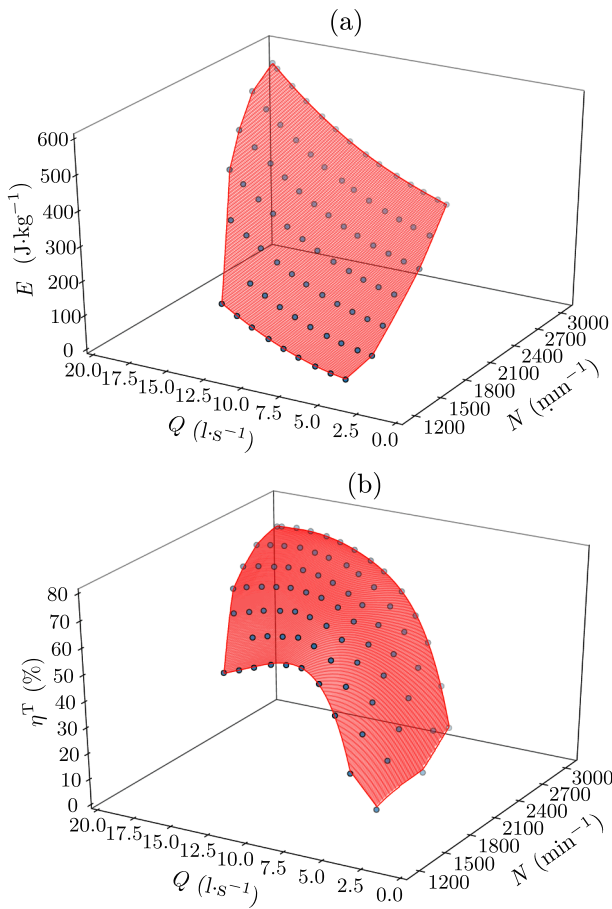


Fig. 16. 3D view of the surrogate functions for modelling the variable speed hill chart performance of the $n_q = 23.1$ PAT: (a) f_E^{HC} ; (b) f_{η}^{HC} .

is particularly convenient if the PAT is operated using a hydraulic and electric control scheme, as proposed in [7]. This case is illustrated for the $n_q^p = 67.3$ pump in Fig. 19 for two operating points corresponding to $E = 100 \text{ J}\cdot\text{kg}^{-1}$ and $E = 150 \text{ J}\cdot\text{kg}^{-1}$, respectively the operating points A and B. Secondly, as the surrogate function that describes the hill chart model is composed of the sum of polynomial terms, the computation of the derivatives with respect to the discharge or to the rotational speed is simple and expedite. Such fact is a major advantage for the

development of optimisation algorithms for controlling the PAT power plant, for instance, the Maximum Power Point Tracking [37], aiming at maximising the generated power based on the available specific energy and discharge at the inlet of single PAT power plant. Other optimisation methodologies can be found for multiple power plants installed in series (e.g., cascade hydropower as in [38]) or for optimising the dispatching of multiple hydropower sources as in [39]. Further investigation is suggested on this optimisation topic.

5. Conclusions and future perspectives

This research paper presents a new two-step methodology based on the bivariate Hermite PCE for predicting the PAT characteristic curves and modelling the variable speed hill chart performance. The PCE method allows defining empirical continuous surrogate functions in a closed-bounded interval, which are used in the scope of this research in two steps, namely for: (i) the identification of surrogate functions for predicting the PAT performance in the turbine mode and in the extended operation mode; and (ii) the modelling the variable speed hill chart performance of a given PAT.

In the first step, surrogate functions are developed aiming at propagating the hydraulic and the mechanical characteristic curves in-between of the unit specific speed values of the tested PATs. Regarding the extended operation, to the authors' knowledge, no methodology is available for predicting the PAT characteristic curves. In this research, the propagated characteristic curves are described by the IEC factors Q_{ED} , T_{ED} , P_{ED} and n_{ED} , as opposed to relative characteristic curves with respect to the turbine BEP, for instance as in [12] or in [16]. This formulation has the advantages of only requiring as input data information provided by pump manufacturers, namely the rated parameters in the pump mode and the impeller diameter, and does not require the previous knowledge of the turbine mode BEP. Overall, the developed methodology has proved effective for modelling the surrogate functions for the PAT characteristic curves prediction, both in the turbine mode and in the extended operation mode. The surrogate functions results featured reduced standard deviations of the error, with respect to the experimental data, while providing a smooth propagation of the characteristic curves in between of the unit specific speed values.

However, certain limitations of the Hermite PCE prediction methodology need to be taken into account. Firstly, the model performed inferiorly for PATs with lower values of unit specific speed. The computation of the weighting coefficients of the PCE based on minimising the relative mean-squared error instead of the absolute mean-squared error (see Eq. 15) would, possibly, improve the fitting of the surrogate

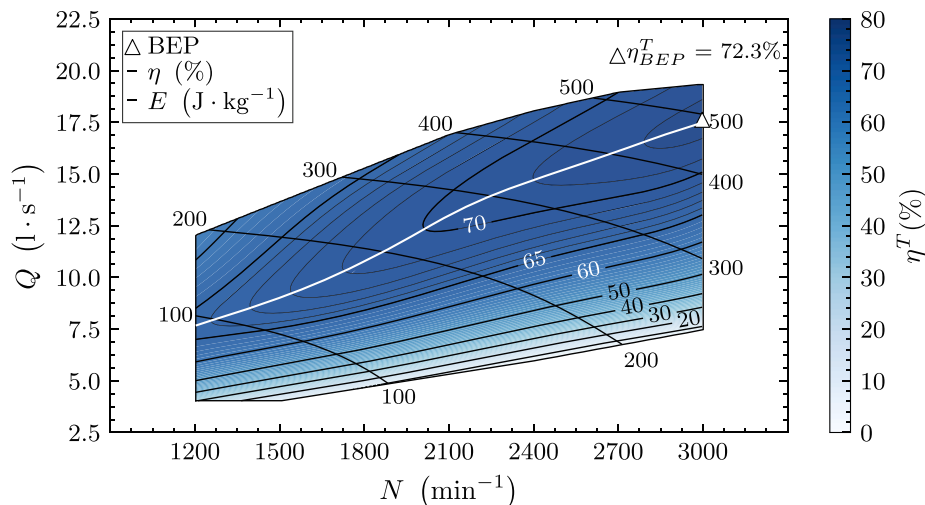


Fig. 17. Variable speed hill chart of the turbine mode performance of the PAT with $n_q^p = 23.1$. Variation of the specific hydraulic energy E and the efficiency η^T with respect to the discharge Q and rotational speed N . The best efficiency ridgeline is indicated in white.

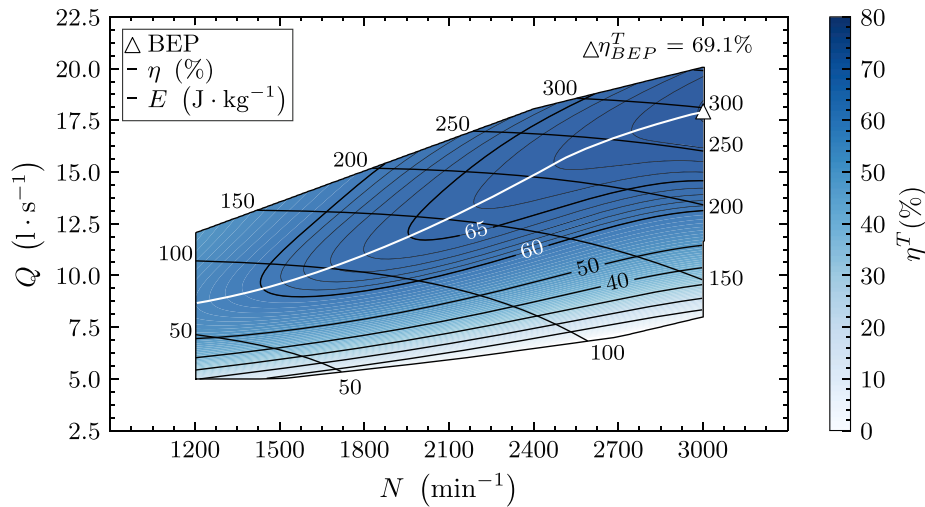


Fig. 18. Variable speed hill chart of the turbine mode performance of the PAT with $n_q^p = 41.0$. Variation of the specific hydraulic energy E and the efficiency η^T with respect to the discharge Q and rotational speed N . The best efficiency ridgeline is indicated in white.

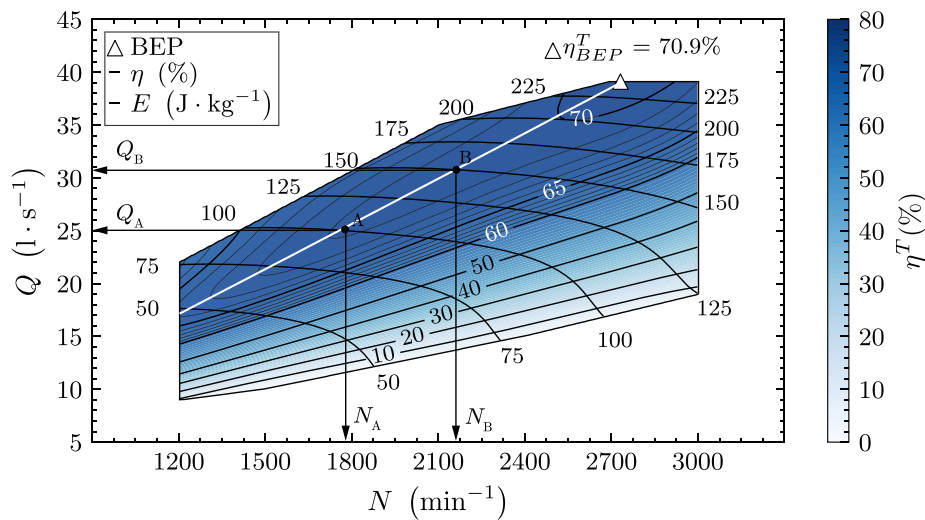


Fig. 19. Variable speed hill chart of the turbine mode performance of the PAT with $n_q^p = 67.3$. Variation of the specific hydraulic energy E and the efficiency η^T with respect to the discharge Q and rotational speed N . The best efficiency ridgeline is indicated in white. The operating points A and B indicate the discharge and rotational set-points that maximise the PAT efficiency for an available specific speed energy of $E = 100 \text{ J}\cdot\text{kg}^{-1}$ and $E = 150 \text{ J}\cdot\text{kg}^{-1}$. Operating point A: $Q_A = 25.0 \text{ l}\cdot\text{s}^{-1}$ and $N_A = 1770 \text{ min}^{-1}$; operating point B: $Q_B = 30.7 \text{ l}\cdot\text{s}^{-1}$ and $N_B = 2180 \text{ min}^{-1}$.

function in the domain of small n_q^p values. Further research is suggested on this topic. Secondly, the use of the IEC factors characteristic curves instead of the relative characteristic curves with respect to the turbine mode BEP (e.g., [12]) narrows the range of application of this methodology for which the surrogate functions were calibrated. The creation of a database of surrogate functions for several pump types, pump manufacturers or a range of n_q^p values is suggested to overcome this limitation. However, additional available data is required for the accomplishment of this step. The use of surrogate functions outside its range of application is not advised. Thirdly, by using the IEC factors characteristic curves, the surrogate functions are not taking into account the effect of increasing maximum efficiency with increasing impeller diameter [36] and, thus, the surrogate functions should only be applied in the range the impeller diameters considered. Including the effect of the impeller diameter on the characteristic curves prediction methodology would be a major improvement for this prediction methodology. Finally, the results accuracy of this methodology were not gauged on a blind test against other methodologies. This comparison would be interesting to assess the reliability of the methodology proposed in this research.

As a final note, methodologies for predicting the PAT performance (this methodology or the ones available in literature) should be only used during preliminary stages of the design of PAT power plants. The final design of the PAT power plants, this is, the design of the power plant layout and control settings, should be performed with reliable data, either acquired from model testing in test-rigs or by data provided by PAT manufacturers. This prediction methodology provides a surrogate function that provided characteristic curves prediction in a continuously closed domain of n_q^p values. Thus, an interesting application for this surrogate functions is the development of a numerical decision support tool that optimises the design of PAT power plants, in what concerns, for instance, their layout, number of units, unit specific speed of the installed units, among others.

The second step of this methodology consisted of modelling the hill chart that describes the variable speed operation of a given PAT. The obtained hill chart allows identifying the best efficiency ridgeline, this is, the discharge and rotational speed set-points that maximise the PAT efficiency for a given available specific hydraulic energy value. This feature is a major advantage if the PAT is operated using a simultaneous hydraulic control by using a by-pass and electric control by varying the

rotational speed [7]. Also, as the hill chart is described by a continuous function composed of the sum of polynomial terms, the computation of the derivatives of the hill chart is straightforward. This straightforward computation is a major advantage as it allows using the hill chart model for optimising control settings that maximise the energy produced by the PAT power plant. The use of the Hermite PCE hill chart model for describing the variable speed operation of a PAT is, however, more complex than simple turbomachine affinity laws. An interesting application to this PCE variable speed hill chart would be the development of a numerical model for using PATs to dynamically control the pressure in water supply system for its optimal management [40], while operating near the best efficiency ridgeline.

Acknowledgements

The authors acknowledge Portuguese national funding agency for science, research and technology (FCT) and the Laboratory for Hydraulic Machines for funding João Delgado PhD grant, under the joint doctoral programme H2Doc (PD/00206/2012) between Instituto Superior Técnico (IST) and École Polytechnique Fédérale de Lausanne (EPFL). Authors acknowledge Grundfos Portugal for kindly providing the pumps tested in the laboratory. Authors acknowledge EPFL Mechanical Engineer MSc João Gomes Pereira Júnior for the helpful discussion during the development of this research. Authors gratefully acknowledge the reviewers' comments and suggestions that enhanced the quality of this research paper.

References

- [1] Vilanova MRN, Balestieri JAP. Hydropower recovery in water supply systems: models and case study. *Energy Convers Manage* 2014;84:414–26. <https://doi.org/10.1016/j.enconman.2014.04.057>.
- [2] Samora I, Franca MJ, Schleiss AJ, Ramos HM. Simulated annealing in optimization of energy production in a water supply network. *Water Resour Manage* 2016;30(4):1533–47. <https://doi.org/10.1007/s11269-016-1238-5>.
- [3] Elbatran AH, Yaakob OB, Ahmed YM, Shabara HM. Operation, performance and economic analysis of low head micro-hydropower turbines for rural and remote areas: a review. *Renew Sustain Energy Rev* 2015;43:40–50. <https://doi.org/10.1016/j.rser.2014.11.045>.
- [4] Williams AA, Smith NPA, Bird C, Howard M. Pumps as turbines and induction motors as generators for energy recovery in water supply systems. *Water Environ J* 1998;12(3). <https://doi.org/10.1111/j.1747-6593.1998.tb00169.x>.
- [5] Jain SV, Patel RN. Investigations on pump running in turbine mode: a review of the state-of-the-art. *Renew Sustain Energy Rev* 2014;30:841–68. <https://doi.org/10.1016/j.rser.2013.11.030>.
- [6] Singh P, Nestmann F. A consolidated model for the turbine operation of centrifugal pumps. *J Eng Gas Turbines Power* 2011;133(6):063002. <https://doi.org/10.1115/1.4002270>.
- [7] Fecarotta O, Ramos HM, Derakhshan S, Giudice GD, Carravetta A. Fine tuning a PAT hydropower plant in a water supply network to improve system effectiveness. *J Water Resour Plan Manage* 2018;144(8):04018038. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000961](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000961).
- [8] Carravetta A, Fecarotta O, Sinagra M, Tucciarelli T. Cost-benefit analysis for hydropower production in water distribution networks by a pump as turbine. *J Water Resour Plan Manage* 2014;140(6):04014002. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000384](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000384).
- [9] Carravetta A, Derakhshan S, Ramos HM. *Pumps as turbines: fundamentals and applications*. Springer; 2017.
- [10] Muhammetoglu A, Karadirek IE, Ozen O, Muhammetoglu H. Full-scale PAT application for energy production and pressure reduction in a water distribution network. *J Water Resour Plan Manage* 2017;143(8):04017040. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000795](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000795).
- [11] Singh P, Nestmann F. An optimization routine on a prediction and selection model for the turbine operation of centrifugal pumps. *Exp Therm Fluid Sci* 2010;34(2):152–64. <https://doi.org/10.1016/j.expthermflusci.2009.10.004>.
- [12] Derakhshan S, Nourbakhsh A. Experimental study of characteristic curves of centrifugal pumps working as turbines in different specific speeds. *Exp Therm Fluid Sci* 2008;32(3):800–7. <https://doi.org/10.1016/j.expthermflusci.2007.10.004>.
- [13] Naldi G, Artina A, Bragalli C, Liserra T, Marchi A. Experimental investigation of characteristics curves of centrifugal pumps working as turbines 2010.
- [14] Pugliese F, Paola FD, Fontana N, Giugni M, Marini G. Experimental characterization of two pumps as turbines for hydropower generation. *Renew Energy* 2016;99:180–7. <https://doi.org/10.1016/j.renene.2016.06.051>.
- [15] Barbarelli S, Amelio M, Florio G. Experimental activity at test rig validating correlations to select pumps running as turbines in microhydro plants. *Energy Convers Manage* 2017;149:781–97. <https://doi.org/10.1016/j.enconman.2017.03.013>.
- [16] Novara D, McNabola A. A model for the extrapolation of the characteristic curves of pumps as turbines from a datum best efficiency point. *Energy Convers Manage* 2018;174:1–7. <https://doi.org/10.1016/j.enconman.2018.07.091>.
- [17] Fecarotta O, Carravetta A, Ramos HM, Martino R. An improved affinity model to enhance variable operating strategy for pumps used as turbines. *J Hydraul Res* 2016;54(3):332–41. <https://doi.org/10.1080/00221686.2016.1141804>.
- [18] Rossi M, Renzi M. A general methodology for performance prediction of pumps-as-turbines using artificial neural networks. *Renew Energy* 2018. <https://doi.org/10.1016/j.renene.2018.05.060>.
- [19] Rossi M, Renzi M. Analytical prediction models for evaluating pumps-as-turbines (PaTs) performance. *Energy Procedia* 2017;118:238–42. <https://doi.org/10.1016/j.egypro.2017.07.011>.
- [20] Nicolet C. Hydroacoustic modelling and numerical simulation of unsteady operation of hydroelectric systems [Ph.D. thesis]. Lausanne, Switzerland: École Polytechnique Fédérale de Lausanne; 2007. <https://doi.org/10.5075/epfl-thesis-3751>.
- [21] Hydraulic turbines, storage pumps and pump-turbines - model acceptance tests, Standard, International Electrotechnical Commission (IEC), Geneva, Switzerland; 1999.
- [22] Barbarelli S, Amelio M, Florio G. Predictive model estimating the performances of centrifugal pumps used as turbines. *Energy* 2016;107:103–21. <https://doi.org/10.1016/j.energy.2016.03.122>.
- [23] Delgado J, Ferreira JP, Covas DI, Avellan F. New test-rig for micro hydropower turbomachines. In: CCWI 2017-computing and control for the water industry. The University of Sheffield; 2017.
- [24] Andolfatto L, Delgado J, Vagnoni E, Münch-Alligné C, Avellan F. Analytical hill chart towards the maximisation of energy recovery on water utility networks with counter rotating micro-turbine. In: 36th IAHR world congress; 2015.
- [25] Gomes JP, Andolfatto L, Avellan F. Monitoring a francis turbine operating conditions. *Flow Measur Instrument* 2018. <https://doi.org/10.1016/j.flowmeasinst.2018.07.007>.
- [26] Yang S-S, Derakhshan S, Yang F-YK. Theoretical, numerical and experimental prediction of pump as turbine performance. *Renew Energy* 2012;48:507–13. <https://doi.org/10.1016/j.renene.2012.06.002>.
- [27] Stone MH. The generalized Weierstrass approximation theorem. *Math Mag* 1948;21(5):237–54. <https://doi.org/10.2307/3029750>.
- [28] Wiener N. The homogeneous chaos. *Am J Math* 1938;60(4):897–936. <https://doi.org/10.2307/2371268>.
- [29] Hermite C. Sur un nouveau développement en série des fonctions, Mallet-Bachelier; 1864.
- [30] Ogura H. Orthogonal functionals of the poisson process. *IEEE Trans Inf Theory* 1972;18(4):473–81. <https://doi.org/10.1109/TVT.1972.1054856>.
- [31] Lawson CL, Hanson RJ. Solving least squares problems. Society for Industrial and Applied Mathematics, vol. 15; 1995. <https://doi.org/10.1137/1.9781611971217>.
- [32] Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control* 1974;19(6):716–23. <https://doi.org/10.1109/TAC.1974.1100705>.
- [33] Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. *Sociol Methods Res* 2004;33(2):261–304. <https://doi.org/10.1177/0049124104268644>.
- [34] Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. *Ann Statist* 2004;32(2):407–99. <https://doi.org/10.1214/009053604000000067>.
- [35] Tibshirani R. Regression shrinkage and selection via the lasso. *J Roy Stat Soc Ser B (Methodol)* 1996;58(1):267–88.
- [36] Gülich JF. *Centrifugal pumps*. Springer; 2014.
- [37] Andolfatto L, Vagnoni E, Hasmatuchi V, Münch-Alligné C, Avellan F. Simulation of energy recovery on water utility networks by a micro-turbine with counter-rotating runners. *IOP Conf Ser Earth Environ Sci* 2016;49(10). <https://doi.org/10.1088/1755-1315/49/10/20212>.
- [38] Liu B, Cheng C, Wang S, Liao S, Chau K-W, Wu X, et al. Parallel chance-constrained dynamic programming for cascade hydropower system operation. *Energy* 2018;165:752–67. <https://doi.org/10.1016/j.energy.2018.09.140>.
- [39] Cheng C, Liu B, Chau K-W, Li G, Liao S. China small hydropower and its dispatching management. *Renew Sustain Energy Rev* 2015;42:43–55. <https://doi.org/10.1016/j.rser.2014.09.044>.
- [40] Wright R, Abraham E, Parpas P, Stoianov I. Control of water distribution networks with dynamic DMA topology using strictly feasible sequential convex programming. *Water Resour Res* 2015;51(12):9925–41. <https://doi.org/10.1002/2015WR017466>.