2018 European Control Conference (ECC)
June 12-15, 2018. Limassol, Cyprus

Interval Predictor based on Supporting Hyperplanes

J.M. Bravo[1], E. Cojocaru[1], M.Vasallo[1], and T. Alamo[2]

Abstract— A new interval predictor for dynamical systems is presented in this work. The aim is to predict the future output of a dynamical system using a prediction model. This work focuses on predictors that return an interval bound. The interval prediction provides upper and lower bounds of the future system output. Given a set of input-output data of the dynamical system, the interval predictor is obtained using supporting hyperplanes of this set. An inner point of this interval can be used as point prediction. The main goodness of the proposed predictor is to provide a trade off between the width of the interval prediction and the prediction error of the point prediction. A design parameter can be used to balance both objectives. The work proposed a cross-validation methodology to tune this parameter. An example with real data is included to illustrate the proposed interval predictor.

I. INTRODUCTION

In the context of dynamical systems, a prediction method provides information about the future behavior of the system. Fault diagnosis and detection methods [1] and advanced control systems as model predictive control [2] use some kind of prediction method in order to improve their performance. A widely extended prediction method is to use a mathematical model of the system. This model can be obtained by first principles. However, this requires depth knowledge of the system and a parametric adjustment of the model. Another option is to use system identification methods [3]. System identification methods infer models from input-output data. One of the most widely used system identification method is the prediction error approach. Given a parameterized family of models and an input-output dataset, prediction error methods choose the parameter values that minimize a function of the prediction error, i.e., the quadratic prediction error. Once an identified model is available, the future system behavior can be predicted using it. It is important to remark that the dataset used in the identification step is not required in the prediction step.

Interval predictors are an interesting extension of classical prediction methods. Interval predictors provide an upper and lower bound of the prediction, in this sense, they bound the future system behavior taking into account disturbances, noise or non modeled dynamics. Furthermore, any inner value of the interval prediction can be used as point prediction. Interval predictors are important components of robust control systems [4] and fault detection systems under uncertainty [5], [6], [7].

A method to obtain a prediction interval is to consider an additive error term in the measurements. This term is assumed to be a stochastic variable with some a priory known properties [8]. This properties and a parametric model with certain structure are used to infer the confidence interval of the prediction from a probabilistic point of view. A different method to obtain interval predictors is to use system identification methods based on bounded-error [9]. These methods consider an unknown-but-bounded additive noise added to the system output. Assuming a parameterized family of functions, a set of measurements and the bounded noise, these methods obtain the set of parameters that is consistent with the measurements, the family of models and the bounded-error. In order to bound this set of parameters, different geometric figures has been considered. Ellipsoids, paralelotopes and zonotopes are well-known [10], [11], [12]. Bounded-error methods can be formulated from a nonparametric point of view. In [13] a set of functions with a bounded Lipschitz constant is considered. A family of interval predictors based on stored data and norm-1 is considered in [14].

This work proposes an alternative method to obtain an interval predictor. Given an input-output measurements dataset, the proposed predictor obtains a lower and upper supporting hyperplane to this set and uses both to provide an interval prediction. These hyperplanes are obtained solving a convex constrained optimization problem. An inner value of the interval prediction can be used as point prediction. In order to balance the predictor performance a parameter is included. An extreme value of the parameter provides the tightest interval prediction obtained by the proposed method. The opposite extreme value provides an interval prediction that minimizes the quadratic error of the point prediction. Intermediate values provide a trade off between these extreme options. This paper proposed a cross-validation methodology [15] to estimate a suitable parameter value.

The proposed method presents some advantages over other known interval methods. Unlike bounded-error methods, the proposed interval prediction does not assume a bound in the measurements errors. Neither is assumed known stochastic properties of the measure-

*This research has been supported by DPI2013-48243-C2-2-R, DPI2016-76493-C3-1-R and DPI2016-76493-C3-2-R of Ministerio de Economía y Competitividad (Spain).

[1]J.M. Bravo, E. Cojocaru and M. Vasallo are with Departamento de Ingeniería Electrónica, de Sistemas Informáticos y Automática, Universidad de Huelva, Escuela Técnica Superior de Ingeniería Campus de La Rábida, 21819 Palos de La Frontera, Spain. caro@diesia.uhu.es ,emiliancojocaru@gmail.com,manuel.vasallo@diesia.uhu.es

[2]T. Alamo is with the Departamento de Ingeniería de Sistemas y Automática. Universidad de Sevilla. Camino de los Descubrimientos s/n. 41092 Sevilla. Spain. talamo@us.es

ments error, assumption used by stochastic methods to obtain a confidence interval. On the other hand, the proposed method is related to quantile regression [16]. In quantile regression, an interval prediction is obtained solving an optimization problem that penalizes the positive or negative sign of the prediction error. The main advantage of the proposed method over the quantile regression is the possibility to trade off between the performance of the point prediction and the interval prediction width.

This work is organized as follows. Section II formulates the addressed problem. The proposed predictor is presented in section III. Some practical considerations can be consulted in section IV. Section V presents an example to illustrate the proposed method. Finally, some conclusions and future works are shown in sections VI and VII.

II. PROBLEM FORMULATION

Consider a discrete time dynamical system. At time instant k, $y_k \in \mathbb{R}$ denotes a measured system output and $x_k \in X \subseteq \mathbb{R}^{n_x}$, being X a compact set, a vector of system measurements. This vector can include, for example, past input-output system measurements. The mathematical formulation that relates y_k and x_k is unknown. Given a vector x_k, the aim of this work is to obtain an interval $IP_{\mathcal{D}}(x_k) = [\underline{y}_k, \overline{y}_k]$ such that $\underline{y}_k \leq y_k \leq \overline{y}_k$.

The proposed method uses a set $\mathcal{D}$ of input-output system measurements to obtain the interval predictor. Assumption 1 establishes this point.

Assumption 1: A set of input-output system measurements $\mathcal{D} = \{(x_i, y_i),\ with\ i = 1, ..., N\}$ is available.

III. PROPOSED PREDICTOR

The proposed interval predictor is based on the concept of supporting hyperplane [17]. A definition is given in this sense.

Definition 1 (Supporting hyperplane): Given a set $\mathcal{C}$ and a point $z_o \in \mathrm{bd}\mathcal{C}$ in the boundary of $\mathcal{C}$, if $\beta \neq 0$ satisfies $\beta^T z \leq \beta^T z_o$ for all $z \in \mathcal{C}$, then the hyperplane $\{z\ :\ \beta^T z = \beta^T z_0\}$ is a supporting hyperplane to $\mathcal{C}$ at point z_0.

As commented above, the proposed interval method is based on supporting hyperplanes to set $\mathcal{D}$. The key idea is to obtain a supporting hyperplane that minimizes a quadratic error function. Definition 2 provides the exact expression and Property 1 shows that the proposed expression provides a supporting hyperplane to set $\mathcal{D}$.

Definition 2: Given a set $\mathcal{D}$, the hyperplane $\mathcal{H}_{\mathcal{D}}$ is defined as

$$\mathcal{H}_{\mathcal{D}} = \{\left[x^T\ y\right]^T\ :\ [x^T\ 1]\theta^* - y = 0\}$$

where θ^* is obtained solving the optimization problem

$$\begin{aligned} \min_{\theta} \quad & \sum_{i=1}^{N} (y_i - [x_i^T\ 1]\theta)^2 \\ s.t. \quad & [x_i^T\ 1]\theta \leq y_i \quad i = 1, ..., N. \end{aligned} \tag{1}$$

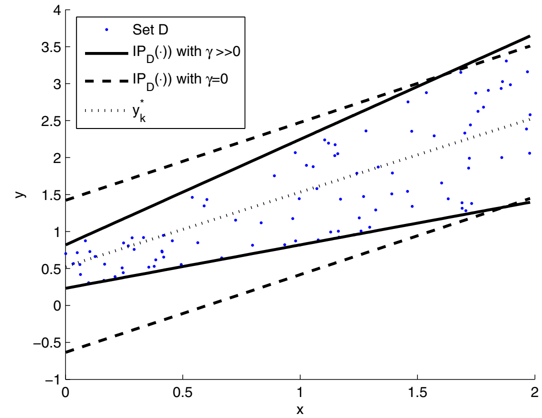

Fig. 1. Example of application of $IP_D(\cdot)$.

Property 1: The hyperplane $\mathcal{H}_{\mathcal{D}}$ is a supporting hyperplane to set $\mathcal{D}$.

Proof. The proof is omitted by lack of space.

Remark 1: By definition we have that $[x_i^T\ 1]\theta^* \leq y_i$ with $i = 1, ..., N$. So, the hyperplane $\mathcal{H}_{\mathcal{D}}$ is a lower bound of the convex hull of $\mathcal{D}$ in the component y. Furthermore, by Property 1, this bound is tangent to $\mathcal{D}$ because there exists an index s such that $[x_s^T\ 1]\theta^* = y_s$.

The optimization problem (1) can be modified to obtain a hyperplane that is an upper bound of $\mathcal{D}$ in the component y. We only need to change the inequalities, that is, $[x_i^T\ 1]\theta \geq y_i$ with $i = 1, ..., N$. This idea enables to propose an interval predictor of the system output. In order to obtain the interval prediction, the proposed method computes a pair of hyperplanes that are a lower and upper bound of the convex hull of set $\mathcal{D}$ in the component y.

Definition 3: Given a vector x_k an interval prediction $IP_{\mathcal{D}}(x_k) = [\underline{y}_k, \overline{y}_k]$ is defined by

$$\underline{y}_k = [x_k^T\ 1]\underline{\theta}^* - \underline{\alpha}^*$$

$$\overline{y}_k = [x_k^T\ 1]\overline{\theta}^* + \overline{\alpha}^*$$

where $\underline{\theta}^*$,$\overline{\theta}^*$, $\underline{\alpha}^*$ and $\overline{\alpha}^*$ are the optimal solutions of the following optimization problems

$$\begin{aligned} \min_{\underline{\theta},\underline{\alpha}} \quad & \sum_{i=1}^{N} (y_i - [x_i^T\ 1]\underline{\theta})^2 + \gamma\underline{\alpha} \\ s.t. \quad & [x_i^T\ 1]\underline{\theta} \leq y_i + \underline{\alpha} \quad i = 1, ..., N \\ & \underline{\alpha} \geq 0 \end{aligned} \tag{2}$$

$$\begin{aligned} \min_{\overline{\theta},\overline{\alpha}} \quad & \sum_{i=1}^{N} (y_i - [x_i^T\ 1]\overline{\theta})^2 + \gamma\overline{\alpha} \\ s.t. \quad & [x_i^T\ 1]\overline{\theta} \geq y_i - \overline{\alpha} \quad i = 1, ..., N \\ & \overline{\alpha} \geq 0 \end{aligned} \tag{3}$$

where $\gamma > 0$ is a design parameter. Furthemore, a point prediction can be obtained by expression

$$y_k^* = \frac{1}{2}[x_k^T \ 1](\overline{\theta}^* + \underline{\theta}^*).$$

Figure 1 shows a numeric example of the proposed predictor. The set $\mathcal{D}$ is showed as points. The interval bounds obtained with parameters $\gamma = 0$ and $\gamma >> 0$ are showed with dashed and solid lines respectively. The dotted central line is the pointed prediction obtained with $\gamma = 0$.

Remark 2: As we show in what follows, one can get rid of the variables $\bar{\alpha}, \underline{\alpha}$ in (2) and (3). Applying equalities $\bar{\alpha} = \max\limits_{i=1,\ldots,N}\{0, y_i - [x_i^T \ 1]\bar{\theta}\}$ and $\underline{\alpha} = \max\limits_{i=1,\ldots,N}\{0, [x_i^T \ 1]\underline{\theta} - y_i\}$ is possible to reformulate the optimization problems (2) and (3) as the following convex and unconstrained optimization problems,

$$\min_{\underline{\theta}} \quad \sum_{i=1}^{N}(y_i - [x_i^T \ 1]\underline{\theta})^2 + \ \gamma \max_{i=1,\ldots,N}\{0, [x_i^T \ 1]\underline{\theta} - y_i\} \quad (4)$$

$$\min_{\bar{\theta}} \quad \sum_{i=1}^{N}(y_i - [x_i^T \ 1]\bar{\theta})^2 + \gamma \max_{i=1,\ldots,N}\{0, y_i - [x_i^T \ 1]\bar{\theta}\}. \quad (5)$$

It is interesting to remark that if $\gamma = 0$, the optimization problems (2) and (3) can be formulated as an only one optimization problem,

$$\min_{\theta} \quad \sum_{i=1}^{N}(y_i - [x_i^T \ 1]\theta)^2, \qquad (6)$$

that is, (2) and (3) reduce to a standard identification of a linear system by means of ordinary least squares. On the other hand, formulations (2) and (3) are very close to the well-known Quadratic Programming (QP) formulation, making clear a computer implementation.

Next, some characteristics of the proposed predictor are commented. Definition 3 provides a family of interval predictors parameterized by γ. If $\gamma = 0$ then $\underline{\theta}^* = \overline{\theta}^* = \theta^{LS}$ where θ^{LS} is the optimal solution of (6), being $\theta^{LS} = (\Omega^T\Omega)^{-1}\Omega^T Y$ with $\Omega = [r_1 \ r_2 \ \ldots \ r_N]^T$, $r_i = [x_i^T \ 1]^T$, $i = 1, \ldots, N$ and $Y = [y_1 \ y_2 \ \ldots \ y_N]$. From this it is inferred that if $\gamma = 0$, the point prediction y_k^* is optimal because is the solution obtained by ordinary least squares. Furthermore, taking into account that $\underline{\alpha}^* = \max\limits_{i=1,\ldots,N}\{0, r_i^T\theta^{LS} - y_i\}$ and $\overline{\alpha}^* = \max\limits_{i=1,\ldots,N}\{0, y_i - r_i^T\theta^{LS}\}$, then $\overline{y}_k - \underline{y}_k \geq \underline{\alpha}^* + \overline{\alpha}^*$. This property is important to guarantee that for any $x_k \in X$ there is a $\gamma \geq 0$ such that $\underline{y}_k \leq y_k^* \leq \overline{y}_k$, that is, the consistence of the interval prediction is assured.

On the other hand, when $\gamma \to \infty$ we have that $\underline{\alpha}^*, \overline{\alpha}^* \to 0$. We can infer that the optimization problems (2) and (3) belongs to the family presented in Definition 2. Therefore, the optimal solutions $\underline{\theta}^*$ and $\overline{\theta}^*$ define a pair of supporting hyperplanes that provide an outer bound of the set $\mathcal{D}$. The obtained interval prediction is tight, however the optimality of the point prediction can be lost and some bias can appear.

To summarize the previous paragraphs, the proposed interval predictor includes a user defined parameter γ that enable to balance the size of the interval prediction and the prediction error of the point prediction. Next section provides an empirical method to tune this parameter using the set of data $\mathcal{D}$.

IV. PRACTICAL CONSIDERATIONS

The proposed method relies on parameter $\gamma \geq 0$. In this paper we propose a work scheme in which the value for γ is obtained through cross-validation. The checking scheme is based in the $\mathcal{D}^\mu$-consistency of the design parameter γ. Basically, fixed a value of γ, the idea is to check the number of outputs y_i of the set $\mathcal{D}$ that are predicted in a correct form, that is, $y_i \in IP_{\mathcal{D}_i}(x_i)$, using the proposed predictor and the set of points $\mathcal{D}_i = \mathcal{D} \setminus \{(x_i, y_i)\}$. Note that the pair (x_i, y_i) is not included in set $\mathcal{D}_i$ to obtain the leave-one-out validation. Definition 4 formalizes this idea.

Definition 4: Given an interval predictor $IP_{\mathcal{D}}(\cdot)$, the parameter γ is $\mathcal{D}^\mu$-consistent, where $\mu \in [0, 1]$, if

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{I}_{\{IP_{\mathcal{D}_i}(x_i)\}}(y_i) \geq \mu$$

where

$$\mathbb{I}_{\{A\}}(y) = \begin{cases} 1 & if \ y \in A \\ 0 & if \ y \notin A \end{cases}$$

is the indicator function and $\mathcal{D}_i = \mathcal{D} \setminus \{(x_i, y_i)\}$.

Note that other schemes based on alternative cross-validation methods, i.e. k-fold cross-validation, could be considered by users. The checking of the $\mathcal{D}^\mu$-consistency of a parameter γ can be computed by Algorithm 1.

```
Algorithm 1: D^μ-consistency(D, γ, μ)
    cont = 0;
    for i = 1:N
        Obtain D_i = D \ {(x_i, y_i)}
        if y_i ∈ IP_{D_i}(x_i)
            cont = cont+1
        endif
    endfor
    if cont/N ≥ μ
        return true
    else
        return false
    endif
End
```

Note that a value γ that is $\mathcal{D}^1$-consistent implies that $y_i \in IP_{\mathcal{D}_i}(x_i)$ for all y_i of set $\mathcal{D}$. However, a value $\mu = 1$ can be too strict for certain problems, i.e., a dataset including rare dynamics or values declared as outliers. In this case a predictor with poor performance can be obtained. This problem can be circumvented using values of μ lower that 1. On the other hand, if an upper bound of the noise included in the measurements y_i is known, this information can be added to the interval prediction to obtain robust supporting hyperplanes.

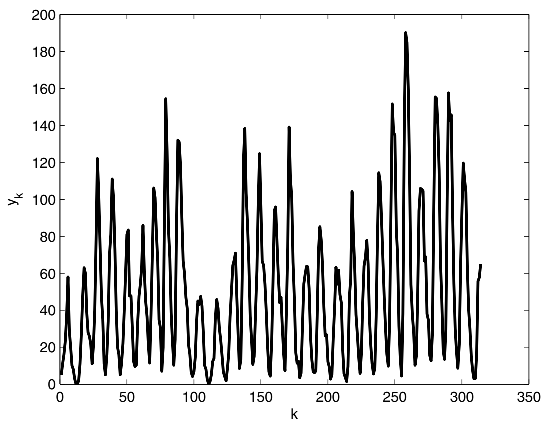

Fig. 2. Time Series Sunspot Numbers.

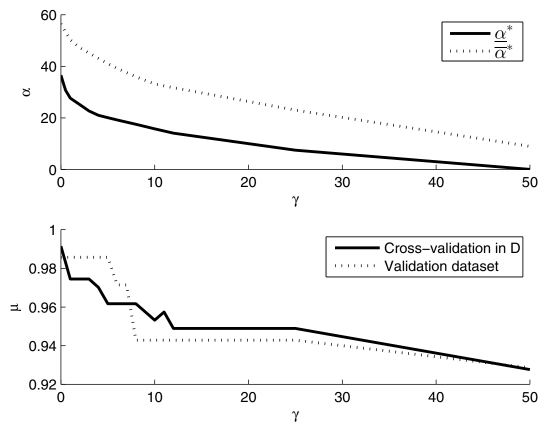

Fig. 3. Variable α^* and parameter μ.

At last, the user should select a value of γ. A slight modification of Algorithm 1 can be used to make the choice. Fixed a value of μ, we can take the value of γ that is $\mathcal{D}^{\mu}$-consistent and provides the best performance in some sense. For example, we can take a $\mathcal{D}^{\mu}$-consistent value that minimizes the mean of the width interval prediction. This paper proposes to use a value of γ with a trade off between the size of the interval prediction an the point prediction error. Next section provides an example of this point.

V. EXAMPLE

In order to illustrate the proposed interval predictor a time series named Sunspot Numbers has been considered (see Figure 2). This time series is interesting because is a real world example, the data are public and its prediction is relevant in many application fields. The time series includes 314 values, starting in year 1700. The first 244 data has been included in set $\mathcal{D}$. The last 70 data have been used to build a validation dataset. The aim of this example is twofold. Firstly, we want to illustrate that the proposed method provides a balanced predictor and second, we show that the scheme proposed in section IV provides suitable results. In this sense, section IV proposes a method to estimate a suitable value for parameter γ. A cross-validation approach in set $\mathcal{D}$ is used to estimate the best γ. The validation dataset is used to test the performance of different values of γ. We show in this example that the value for parameter γ obtained by cross-validation is close to the one that optimizes the performance index in the validation set.

A one-ahead predictor with $x_k = [y_{k-1}\ y_{k-2}\ \dots\ y_{k-9}]^T$ has been considered. The Root Mean Square Error ($RMSE$) is used as metric to evaluate the point prediction.

$$RMSE_\gamma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i^* - y_i)^2}. \tag{7}$$

On the other hand a second metric denoted INT and defined as

$$INT_\gamma = \frac{1}{N}\sum_{i=1}^{N}(\overline{y}_i - \underline{y}_i),$$

is used to evaluate the interval prediction.

A metric BM_γ that evaluates the balanced performance of metrics INT_γ and $RMSE_\gamma$ is used. This metric is defined by expression:

$$BM_\gamma = \left(\frac{INT_\gamma - \underline{INT}}{\overline{INT} - \underline{INT}} + \frac{RMSE_\gamma - \underline{RMSE}}{\overline{RMSE} - \underline{RMSE}}\right)$$

where $\overline{INT}$, $\underline{INT}$, $\overline{RMSE}$ and $\underline{RMSE}$ are the maximum and minimum values obtained by metrics INT_γ and $RMSE_\gamma$.

Figure 3 (upper case) shows the optimal coefficient $\underline{\alpha}^*$ and $\overline{\alpha}^*$ evolution with respect to the values of γ. If $\gamma = 0$, $\underline{\alpha}^*$ and $\overline{\alpha}^*$ obtain their maximum values because in this case the point prediction is given by $\underline{\theta}^* = \overline{\theta}^* = \theta^{LS}$, and the interval prediction width is defined by $\underline{\alpha}^*$ and $\overline{\alpha}^*$. In the opposite case, $\underline{\alpha}^*$ and $\overline{\alpha}^*$ tends to be zero and the optimal parameters $\underline{\theta}^*$ and $\overline{\theta}^*$ define the interval prediction. Figure 3 (lower case) shows the maximum value reached by μ with respect to the values of γ. As can be seen, the results obtained by cross-validation in $\mathcal{D}$ are very close to the results obtained in the validation dataset. On the other hand, as γ increases, the interval prediction width decreases and therefore, the obtained $\mathcal{D}^{\mu}$-consistency is reduced. As can be observed by the solid line, values of γ slower to 11 are $\mathcal{D}^{0.95}$-consistent.

Figure 4 compares the metrics INT_γ and $RMSE_\gamma$ obtained by $IP_{\mathcal{D}}(\cdot)$ in set $\mathcal{D}$ by cross-validation and the obtained in the validation dataset. As can be seen, the curves are similar. This is important because we can use the cross-validation results to infer suitable information. On the other hand, we can see that, as the design parameter γ increases, the metric INT_γ improves and the metric $RMSE_\gamma$ gets worse. In order to balance both metrics this work propose to use the alternative metric BM_γ. Note that $\gamma = 0$ is equivalent to consider an autoregressive model AR(9) adjusted by least squares.

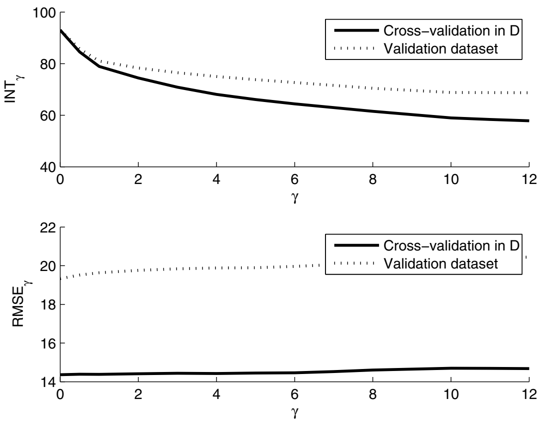

Fig. 4. Metrics INT_γ and $RMSE_\gamma$.

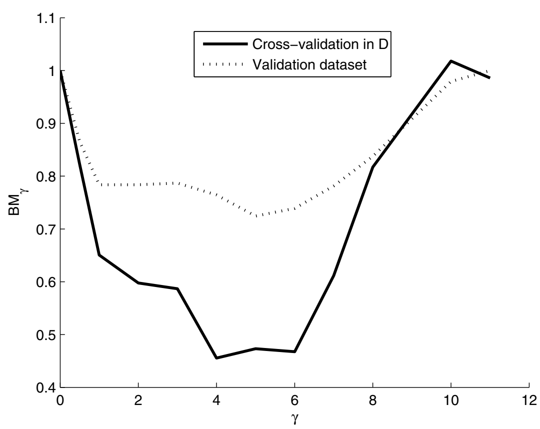

Fig. 5. Metric BM_γ.

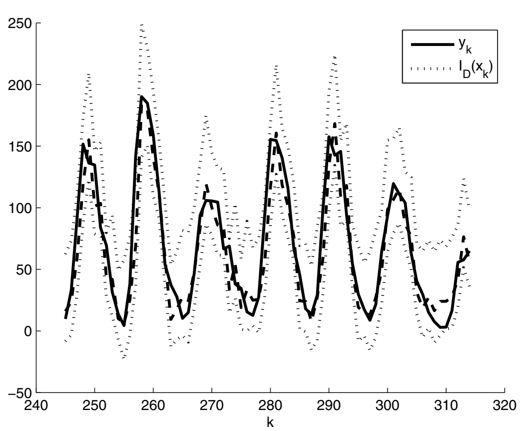

Fig. 6. Interval Prediction with $\gamma = 4$.

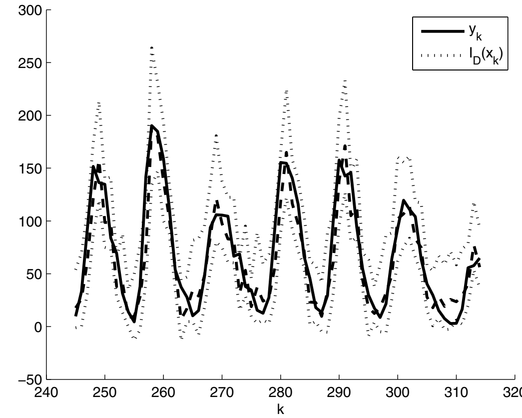

Fig. 7. Interval predictor with $\gamma = 11$.

Figure 5 compares the metric BM_γ obtained by cross-validation in set $\mathcal{D}$ and the obtained in the validation dataset by the proposed method using $\mathcal{D}^{0.95}$-consistent values of γ. As can be seen, the minimum value of BM_γ is obtained with $\gamma = 4$ in cross-validation. This value is very close to the minimum obtained in the validation dataset. Two conclusions can be obtained. The best value of γ to obtain a balanced interval and point prediction is a intermediate value (in this case γ close to 4) and therefore the proposed method is a suitable option. And second, the cross-validation scheme proposed in this paper can be used to infer a suitable value for γ in this real world example.

Table I shows the obtained maximum and minimum values of INT_γ and $RMSE_\gamma$ in cross-validation and in the validation dataset using $\mathcal{D}^{0.95}$-consistent values of γ.

TABLE I

Maximum and minimum values

	$\overline{INT}$	$\underline{INT}$	$\overline{RMSE}$	$\underline{RMSE}$
Cross-validation	92.8	58.35	14.69	14.36
Validation	93.0	68.7	20.42	19.33

Figures 6 and 7 shows the interval predictions (dotted lines) obtained in the validation dataset with $\gamma = 4$ and $\gamma = 11$. Note that $\gamma = 4$ is the obtained optimal value and $\gamma = 11$ is near to the limit of the $\mathcal{D}^{0.95}$-consistency. As expected, the prediction obtained with $\gamma = 11$ provides a tighter interval but increases the error of the point prediction (dashed lines). The time series data are showed with solid lines.

Finally, a comparison with other interval predictors is included. In this sense, we define a point prediction $\hat{y}_k = [x_k^T \;\; 1]\theta^{LS}$, a prediction error $e_k = y_k - \hat{y}_k$, a sample mean $\bar{e} = \frac{1}{N}\sum_{i=1}^{N} e_i$ and a sample variance $\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(e_i - \bar{e})^2$. An interval prediction based on

TABLE II

Comparison with other methods

	$IP_{\mathcal{D}}(\cdot)$		$IP_{\mathcal{D}}^{CH}(\cdot)$	$IP_{\mathcal{D}}^{G}(\cdot)$		
γ	4	11	4.48	2	2.4	2.77
$\mu_{\mathcal{D}}$	0.97	0.95	1	0.95	0.97	0.98
μ_{vs}	0.98	0.94	0.98	0.87	0.94	0.98
$INT_{\mathcal{D}}$	68.0	58.3	131.1	58.5	70.2	81.1
INT_{vs}	75.0	68.7	130.8	58.4	70.0	80.9
$RMSE_{\mathcal{D}}$	14.4	14.6	13.6	13.6	13.6	13.6
$RMSE_{vs}$	19.8	20.4	19.3	19.3	19.3	19.3

the Chebyshev's inequality is defined as $IP_{\mathcal{D}}^{CH}(x_k) = [\hat{y}_k - \gamma\hat{\sigma}, \hat{y}_k + \gamma\hat{\sigma}]$ where γ is a constant. An interval prediction based on Gaussian assumptions is defined as $IP_{\mathcal{D}}^{G}(x_k) = [\hat{y}_k - \gamma\sigma^G, \hat{y}_k + \gamma\sigma^G]$ where $\sigma^G = \sqrt{\hat{\sigma}^2 + \hat{\sigma}^2[x_k^T\ 1](\Omega^T\Omega)^{-1}[x_k^T\ 1]^T}$ and γ is a constant.

Table II shows the obtained results. The results obtained by $IP_{\mathcal{D}}(\cdot)$ with $\gamma = 4$ and $\gamma = 11$ are shown in the first and second columns. We can see the indexes μ, INT and $RMSE$ obtained in the set $\mathcal{D}$ and the validation set. The third column shows the results obtained by $IP_{\mathcal{D}}^{CH}(\cdot)$ with $\gamma = 4.48$. By Chebyshev's inequality $\gamma = 4.48$ provides a $\mu = 0.95$. However the values $\mu_{\mathcal{D}}$ and μ_{vs} obtained are near to 1. An oversized interval prediction is obtained in this case (see the metric INT). The fourth column shows the results obtained by $IP_{\mathcal{D}}^{G}(\cdot)$ assuming Gaussian noise. In this case, by theory, $\gamma = 2$ proves a value of μ near to 0.95. In fact $\mu_{\mathcal{D}} = 0.95$ but in the validation set $\mu_{vs} = 0.87$. That is, a Gaussian assumption fails to provide a consistent interval prediction. Finally, columns five and six are included to compare the proposed interval predictor $IP_{\mathcal{D}}(\cdot)$ with $IP_{\mathcal{D}}^{G}(\cdot)$ under the same μ_{vs}. In both cases $IP_{\mathcal{D}}(\cdot)$ obtains tighter interval prediction than $IP_{\mathcal{D}}^{G}(\cdot)$. On the other hand, the metric $RMSE$ is constant in the $IP_{\mathcal{D}}^{CH}(\cdot)$ and $IP_{\mathcal{D}}^{G}(\cdot)$ predictors.

In summary, the proposed method can be a suitable option to provide a balanced interval predictor in this example. The method provides consistency in the indexes $\mu_{\mathcal{D}}$ and μ_{sv} without assuming an additive error with some known probabilistic assumption. Furthermore, the proposed method obtains tighter interval predictions than $IP_{\mathcal{D}}^{CH}(\cdot)$ and $IP_{\mathcal{D}}^{G}(\cdot)$ in the same level of consistence μ. The price to pay is a bias in the point prediction that increases with the parameter γ.

VI. CONCLUSIONS

This work has presented a new methodology to design interval predictors that allows us to obtain an outer bound of the future output of a dynamical system. The method is based on the concept of supporting hyperplane. The interval prediction is obtained applying a prediction model that has been identified solving two convex optimization problems. The proposed method includes a parameter to balance the size of the interval prediction and the point prediction error. A suitable value of this parameter can be estimated using cross-validation in a dataset obtained from the system. Unlike other prediction methods, it is not necessary to assume a known outer bound of the noise associated to a measurement or a known stochastic distribution of this noise. Finally, it is important to remark that the proposed method encompasses the identification of linear models based as ordinary least squares as particular case.

VII. FUTURE WORK

A future line is to include weights in the cost function of the optimization problems (2) and (3) in order to take into account local information. The aim of this modification is to obtain tight supporting hyperplanes in a neighborhood of a specific vector of interest x_k. Other extension to bound non convex datasets, is to combine local information with variables $\underline{\alpha}$ and $\overline{\alpha}$.

References

[1] R. Isermann, Fault-Diagnosis systems. An Introduction from Fault Detection to Fault Tolerance. Springer, 2006.

[2] E. F. Camacho and C. Bordons, Model Predictive Control, 2nd ed. Springer Verlag, 1999.

[3] L. Ljung, System Identification. Prentice Hall, Upper Saddle River, NJ, 1999.

[4] J. M. Bravo, T. Alamo, and E. F. Camacho, “Robust MPC of constrained discrete-time nonlinear systems based on approximated reachable sets,” Automatica, vol. 42, no. 10, pp. 1745 – 175, 2006.

[5] A. Ingimundarson, J. M. Bravo, V. Puig, T. Alamo, and P. Guerra, “Robust fault detection using zonotope-based set-membership consistency test,” Int. J. Adapt. Control Signal Process, vol. 23, pp. 311–330, 2009.

[6] C. Combastel, “An extended zonotopic and gaussian kalman filter (ezgkf) merging set-membership and stochastic paradigms: Toward non-linear filtering and fault detection,” Annual Reviews in Control, vol. 42, pp. 232 – 243, 2016.

[7] J. Meseguer, V. Puig, and T. Escobet, “Approximating fault detection linear interval observers using λ-order interval predictors,” International Journal of Adaptive Control and Signal Processing, vol. 31, no. 7, pp. 1040–1060, 2017.

[8] D. P. Bertsekas and J. N. Tsitsiklis, Introduction to Probability. Athena Scientific, 2008.

[9] M. Milanese, J. Norton, H. Piet-Lahanier, and E. Walter, Bounding Approaches to System Identification. Plenum Press, New York, 1996.

[10] E. Fogel and Y. F. Huang, “On the value of information in system identification - bounded noise case,” Automatica, vol. 18, no. 2, pp. 229–238, 1982.

[11] A. Vicino and G. Zappa, “Sequential aproximation of feasible parameter sets for identification with set membership uncertainty,” IEEE Transactions on Automatic Control, vol. 41, no. 6, pp. 774–785, 1996.

[12] T. Alamo, J. M. Bravo, and E. F. Camacho, “Guaranteed state estimation by zonotopes,” Automatica, vol. 41, no. 6, pp. 1035–1043, 2005.

[13] M. Milanese and C. Novara, “Set membership prediction of nonlinear time series,” IEEE Transactions on Automatic Control, vol. 50, no. 11, pp. 1655–1669, 2005.

[14] J. M. Bravo, T. Alamo, M. Vasallo, and M. E. Gegundez, “A general framework for predictors based on bounding techniques and local approximation,” IEEE Transactions on Automatic Control, vol. 62, no. 7, pp. 3430–3435, 2017.

[15] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[16] R. Koenker and J. G. Bassett, “Regression quantiles,” Econometrica: journal of the Econometric Society, pp. 33–50, 1978.

[17] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, 2004.