Dynamic Weighted PSVR-Based Ensembles for Prognostics of Nuclear Components

Jie Liu[1], Valeria Vitelli[2], Redouane Seraoui[3] and Enrico Zio[4]

[1]*Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy-Électricité de France, CentraleSupélec, Paris, France*
jie.liu@ecp.fr

[2]*Department of Biostatistics, University of Oslo, Oslo, Norway*
valeria.vitelli@medisin.uio.no

[3] *EDF R&D/STEP Simulation et Traitement de l'information pour l'exploitation des systemes de production, Chatou, France*
redouane.seraoui@edf.fr

[4]*Energy Departement, Politecnico di Milano, Milano, Italy. Chair on System Science and the Energetic Challenge, European Foundation for New Energy – Electricite de France, CentraleSupélec, Paris, France*
enrico.zio@ecp.fr

ABSTRACT

Combining different physical and / or statistical predictive algorithms for Nuclear Power Plant (NPP) components into an ensemble can improve the robustness and accuracy of the prediction. In this paper, an ensemble approach is proposed for prediction of time series data based on a modified Probabilistic Support Vector Regression (PSVR) algorithm. We propose a modified Radial Basis Function (RBF) as kernel function to tackle time series data and two strategies to build diverse sub-models of the ensemble. A simple but effective strategy is used to combine the results from sub-models built with PSVR, giving the ensemble prediction results. A real case study on a power production component is presented.

1. INTRODUCTION

Combining various data-driven approaches into an ensemble is a relatively recent direction of research, aimed at improving the robustness and accuracy of the final prediction. The models which compose the ensemble are called sub-models. Various strategies have been proposed for building sub-models, including error-correcting output coding, Bagging, Adaboost, and boosting (Kim, Pang, Je, Kim & Bang, 2003; Hu, Youn, Wang & Yoon, 2012). Similarly, several methods for aggregating the prediction results of the sub-models have been proposed, such as majority vote, weighted vote, Borda count, Bayes and probabilistic schemes, etc (Polikar, 2006).

Support Vector Machine (SVM) is a popular and promising data-driven method for prognostics. SVM-based ensemble models have been proposed for classification. Chen, Wang and Zuylen (2009) use ensemble of SVMs to detect traffic incidents. The sub-models use different kernel functions and parameters, and their outputs are combined to improve the classification performance. Acar and Rais-Rohami (2009) treat the general weighted-sum formulation of an ensemble as an optimization problem, and then minimize an error metric to select the best weights for the sub-models of SVM. Kurram and Kwon (2013) try to achieve an optimal sparse combination of the sub-model results by jointly optimizing the separating hyperplane obtained by each SVM classifier and the corresponding weights of sub-decisions. Valentini and Dietterich (2003) prove that an ensemble of SVMs employing bagging of low-bias algorithms improves the generalization power of the procedure with respect to single SVM. The ensemble of SVMs built with bagging and boosting can greatly outperform a single SVM in terms of classification accuracy (Kim *et al*., 2003).

SVM can also be treated as a Bayesian inference problem with Gaussian priors. The Maximum A Posteriori (MAP) solution to this problem can contextually give an estimate of the model parameters and also of the underlying function (Sollich, 1999). Within the Bayesian treatment of SVM, an error bar for the prediction, i.e. the variance of the predicted outcome, can also be obtained (Liu *et al.*, 2012). This

Bayesian interpretation of SVM for regression is called Probabilistic Support Vector Regression (PSVR).

In this paper, we focus on the combination of multiple PSVR sub-models (Liu, Seraoui, Vitelli & Zio, 2012). The case study addressed in this paper concerns the monitoring of a component in the Reactor Coolant Pump (RCP) of a Nuclear Power Plant (NPP), with real data collected from a sensor.

An ensemble model of PSVRs is proposed in this paper with a dynamic weighting strategy. The elements of novelty of the method here proposed are various. In the previously mentioned ensembles of SVMs, all the weights were calculated during the training part and fixed for testing. However, a sub-model may perform well only on a part of the data set. Hence, the weights need to be updated considering the different data sets involved in the case study, and even different input vectors. A dynamic weighting strategy, based on local fitness calculation (Baudat & Anouar, 2003) is proposed in this paper. A dynamic weighting method is also used in Muhlbaier, Topalis and Polikar (2009), Yang, Yuan and Liu (2009) and Razavi-Far, Baraldi and Zio (2012), to add a new classifier to the ensemble model, but weights are not adjusted to the different input vectors. Moreover, in order to build an ensemble of PSVRs on different failure scenarios, a modified Radial Basis Function (RBF) is also proposed and used in this paper. In addition, a simple but efficient aggregating method is proposed to combine the outputs of the sub-models, including predicted values and associated error bars. Finally, two different strategies are proposed to form the training data set of each sub-model on the basis of the characteristics of the data. All the novel strategies are tested in the case study concerning a component of the RCP in a NPP.

The rest of the paper is organized as follows. Section 2 gives details about the proposed ensemble approach and a modified RBF. Section 3 illustrates the case study, the available data and how the two proposed ensemble models are constructed. Section 4 presents the experimental results from the PSVR ensemble models and describes the comparison with a single PSVR model. Finally, conclusions with some considerations are drawn in Section 5.

2. DYNAMIC-WEIGHTED PSVR-BASED ENSEMBLE

The strategy underlying the use of ensemble-based methods in prediction problems is to benefit from the strength of different sub-models by combining their outputs to improve the global prediction performance if compared to the result of a single sub-model.

In this section, we give details about the proposed Dynamic-Weighted PSVR-based Ensemble (named DW-PSVR-Ensemble in short).

2.1. Probabilistic Support Vector Regression

Depending on the choice of the loss function, we can define different Gaussian versions of PSVR. The PSVR approach proposed in the previous work (Liu *et al.*, 2012) and used in the ongoing research makes use of the ε-insensitive Loss Function, which enables a sparse set of support vectors to be obtained.

2.1.1. PSVR with ε-Insensitive Loss Function

With limited length of the paper, we do not give mathematical details on the derivation of the PSVR approach that can be found in Gao, Gunn, Harriset and Brown (2002). But it is very important to recall that the output of PSVR is a Gaussian distribution of the predicted value.

2.1.2. Modified Radial Basis Function Kernel

The kernel function enables the mapping of an input vector in a higher-dimensional Reproducing Kernel Hilbert Space (RKHS). By calculating pairwise inner products between mapped samples, kernel functions return the similarity between different samples. In fact, only kernels that fulfill Mercer's Theorem (i.e. the kernel matrix must be positive semi-definite) are valid ones and, thus, can be used in SVM (Minh, Niyogi and Yang, 2006). The most common kernel functions include the linear kernel function, the polynomial kernel function and the Radial Basis Function (RBF).

In all these popular kernel functions, different inputs, i.e. different elements of $\boldsymbol{x}$(t), are treated equally in computing the inner product involved in RBF. For time series data, H historical values of the time series are normally chosen as inputs according to the partial autocorrelation analysis results. These values have, of course, different correlation structures with respect to the output. In order to reflect this difference, a modified RBF is proposed in this paper.

Supposing two input vectors $\boldsymbol{x}(i)$ and $\boldsymbol{x}(j)$, in order to calculate the inner product of these two input vectors in RKHS, the traditional RBF is $K(\boldsymbol{x}(i),\boldsymbol{x}(j)) = \exp(-\frac{\|\boldsymbol{x}(i)-\boldsymbol{x}(j)\|^2}{2\gamma^2})$, with γ the width of the kernel given by particular optimization algorithm, and the proposed modified RBF is $K(\boldsymbol{x}(i),\boldsymbol{x}(j)) = \exp\left(-\frac{\langle \boldsymbol{C_a}^2,(\boldsymbol{x}(i)-\boldsymbol{x}(j))^2\rangle}{2\gamma^2}\right)$. In general, $\boldsymbol{C}_a = (C_1,\dots,C_H)$ denotes the correlation between each input and the output, in our case between different temporal lags and the output of time series data. Suppose $\boldsymbol{A_i} = [x_i(t)], \boldsymbol{B} = [y(t)]$, with $x_i(t)$ the i-th input of $\boldsymbol{x}(t)$ and $t = 1,\dots,M$. Then, C_i is the correlation between $\boldsymbol{A_i}$ and $\boldsymbol{B}$, and so the correlation between $x_i(t)$ and $y(t)$. As $\boldsymbol{C}_a$ is constant for each sub-model, it is easy to prove that the modified RBF satisfies Mercer's Theorem. Thus, the modification of the RBF does not change the theoretical results on which the PSVR method is based.

By giving different weights to different inputs in the input vector, we can reduce the influence of the inputs less correlated with the output and make the more correlated ones more significant in the relation between the inputs and the output. Another advantage of the modified RBF is illustrated in Section 3, when dealing with multiple time series data.

2.2. Ensemble-Based Approach

An ensemble-based approach is obtained by training diverse sub-models and, then, combining their results with given strategies. It can be proven that this can lead to superior performance with respect to a single model approach (Bauer & Kohavi, 1999). A simple paradigm of a typical ensemble-based approach with N sub-models is shown in Figure 1. Ensemble models are built on three key components: a strategy to build diverse models; a strategy to construct accurate sub-models; a strategy to combine the outputs of the sub-models in a way such that the correct predictions are amplified, while the incorrect ones are counteracted. We focus here on the latter. Proper strategies to build diverse and accurate sub-models are described in relation to the case study.

In the DW-PSVR-Ensemble that we are proposing, the sub-models are built using the PSVR model presented in Liu *et al.* (2012). The reason for not using other data-driven approaches, including other SVMs, lies on the special output structure of PSVR. The output of each sub-model built with PSVR contains a predicted value and the associated variance, assuming that the predicted value follows a Gaussian distribution.

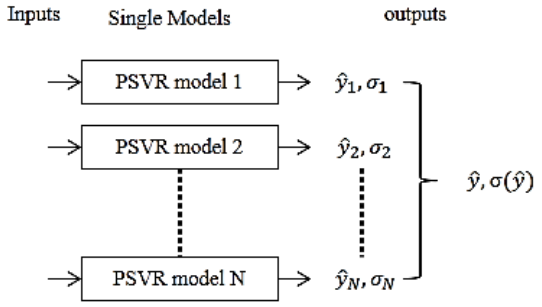

Figure 1. Paradigm of a typical ensemble method.

A dynamic weighted-sum strategy is proposed to combine the outputs of the sub-models. As mentioned in Section 1, different methods can be applied to calculate the weights for the sub-models. In the methods that can be found in the literature, the weights are normally fixed after the ensemble model is built. They are only updated when new sub-models are added to the ensemble or when some sub-models are changed. In some real applications with fast changing environmental and operational conditions, the performance of the ensemble model may degrade rapidly. This degradation is not always caused by the low robustness or capability to adapt of the ensemble model, but can be due to the fact that the best sub-models are not given proper weights. In this paper, a dynamic weighting strategy is thus proposed. The weights are no longer constant during the prediction, but dependent on the input vector. They are recalculated each time a new input vector arrives. Inspired by the work of Baudat and Anouar (2003) and considering the characteristics of PSVR, a local fitness calculation is implemented in this paper to calculate weights of different sub-models for each input vector.

2.2.1. Local Fitness Calculation

In Baudat and Anouar (2003), the authors define a global and local criterion to characterize the feature space in SVM. The proposed local fitness can describe the linearity between the mapping of a new input vector and the mapping of all the Feature Vectors (FVs) of the model: if a linear combination of the mapping of the FVs can better approach the mapping of the new input vector, the model gives better approximation of the output of the new data point; otherwise, the model performs worse for this data point. Thus local fitness can be implemented to derive the weight of each sub-model for each input vector.

Suppose $(\boldsymbol{x}_i, y_i)$, for $i = 1, 2, \dots, M$ are the training data points, and the mapping $\varphi(\boldsymbol{x})$ maps each input vector $\boldsymbol{x}_i$ into RKHS with the mapping $\boldsymbol{\varphi}_i$, for $i = 1, 2, \dots, M$. $k_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is the inner product between $\boldsymbol{\varphi}_i$ and $\boldsymbol{\varphi}_j$. The FVs of this model, selected with the method proposed in Baudat and Anouar (2003), are $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_L\}$, with the corresponding mapping $S = \{\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \dots, \boldsymbol{\varphi}_L\}$. $\boldsymbol{\varphi}_N$ is the mapping of the new input vector $\boldsymbol{x}_N$. According to Baudat and Anouar (2003), the calculation of the local fitness of this new input vector amounts to finding $\{a_{N,1}, a_{N,2}, \dots, a_{N,L}\}$, which gives the minimum of Eq. (1).

$$\delta_N = \frac{\left\| \boldsymbol{\varphi}_N - \sum_{i=1}^{L} a_{N,i} \boldsymbol{\varphi}_i \right\|}{\|\boldsymbol{\varphi}_N\|} \quad (1)$$

The minimum of δ_N can also be expressed with an inner product as shown in Eq. (2).

$$\min \delta_N = 1 - \frac{K_{S,N}^t K_{S,S}^{-1} K_{S,N}}{k_{N,N}} = J_S \quad (2)$$

where $K_{S,S} = \left(k_{i,j}\right), i, j = 1,2, \dots, L$ is the kernel matrix of S and $K_{S,N} = \left(k_{i,N}\right), i = 1,2, \dots, L$ is the vector of the inner product between $\boldsymbol{\varphi}_N$. J_S is the local fitness of $\boldsymbol{x}_N$ for this model.

With Eq. (2), for a new coming data point at time *t*, we can calculate the local fitness $J_i(t)$ for the *i*-th sub-model. And the weight of the *i*-th sub-model for this data point is calculated as $\omega_i(t) = \frac{1/J_i(t)}{\sum_{j=1}^{N} 1/J_j(t)}$.

2.2.2. Combining Sub-Models Outputs

Figure 2 shows the paradigm of DW-PSVR-Ensemble, where *N* is the number of sub-models, $\boldsymbol{x}(t)$ is a new input vector arriving at time t, $w_j(t)$ is the weight assigned to the *j*-th sub-model for the new input vector, $\hat{y}_j(t)$ and $\sigma_j{}^2(t)$ are the predicted value and corresponding variance for the *j*-th sub-model given by PSVR, and $\hat{y}(t)$ and $\sigma^2(t)$ are the final outputs of the ensemble model.

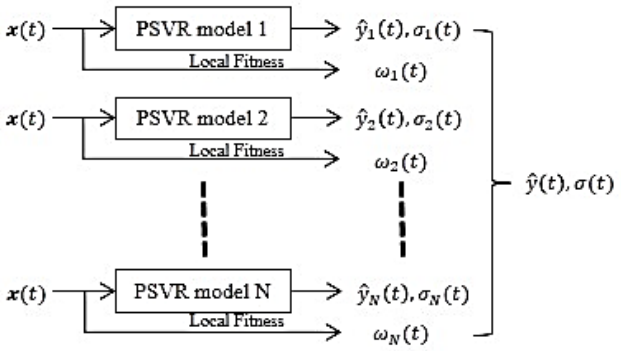

Figure 2. Paradigm of the proposed DW-PSVR-Ensemble.

The output of each PSVR-based sub-model is a Gaussian distribution predicted value. The proposed simple but efficient strategy for combining sub-models results is by taking a weighted-sum of Gaussian distributions, which means that $N(\hat{y}(t), \sigma^2(t)) = \sum_{j=1}^{N} \omega_j(t) N(\hat{y}_j(t), \sigma_j{}^2(t))$, with $N(\hat{y}(t), \sigma^2(t))$ denoting a Gaussian distribution with mean value $\hat{y}(t)$ and variance $\sigma^2(t)$. From this, we can derive the fact that $\hat{y}(t) = \sum_{j=1}^{N} \omega_j(t) \hat{y}_j(t)$ and $\sigma(t) = \sqrt{\sum_{j=1}^{N} \omega_j{}^2(t) \sigma_j{}^2(t)}$, if we assume sub-models results to be uncorrelated.

Note that all the sub-models weights and outputs are a function of t, which means that they are all dependent on the input vector of the ensemble model.

3. CASE STUDY DESCRIPTION

The real case study considered in this paper concerns the 1-day prediction of leak flow of the first seal of the RCP of a NPP.

In this section we describe the time series data and briefly recall the data pre-processing steps. We also detail the strategies to build accurate and diverse sub-models.

3.1. Data Description and Pre-Processing

In the data provided, there are 20 failure scenarios concerning the leak flow from 10 different NPPs. Each failure scenario contains a time series data of the leak flow. They are named Scenario 1, Scenario 2, …, Scenario 20 in the following sections of the paper. These data are monitored every four hours. As these data are time-dependent and recorded within different time windows, only failure scenarios coming from the same NPP have the same size. From the second column of Table 1, we can see that the size of the failure scenarios can vary from 389 to 3129 data points. In some of the scenarios, there are missing data points and outliers.

Table 1. Characteristics of raw and reconstructed scenarios

Scenario	Size of Raw Data	Best Number of Historical values H	Size of Reconstructed Data
1	2277	17	2265
2	385	3	373
3	385	3	373
4	2027	14	2015
5	2027	8	2015
6	2027	8	2015
7	1391	13	1379
8	1391	4	1379
9	1391	4	1379
10	1391	4	1379
11	3124	12	3112
12	562	7	550
13	562	9	550
14	562	9	550
15	964	2	952
16	2767	8	2755
17	2767	7	2755
18	1061	7	1049
19	1061	12	1049
20	861	9	849

Since the dataset we are going to analyze contains both missing data and outliers, we have to deal with both these issues. First of all, we will remove anomalous data, since their extreme values would affect the results of the analysis. Outliers can be easily detected by deciding some constraints, e.g. the limits $\bar{x} \pm 3 * \sigma_x$ where $\bar{x}$ is the mean of all the data points and σ_x is their standard deviation. These limits are needed to detect the outliers, selected as those data points bigger than $\bar{x} + 3 * \sigma_x$ or smaller than $\bar{x} - 3 * \sigma_x$, and subsequently removed. Note that we used such constraints, rather than the usual ones based on the median and the InterQuartile Range (IQR), to be more conservative in the outlier selection, due to the dependence among data.

Secondly, we want to reconstruct missing data. Note that, after the outlier selection and elimination procedure, the number of missing data has increased. A possible way to deal with the reconstruction of missing data is the local polynomial regression fitting. This local least squares regression technique estimates effectively the values when there are missing data points. Moreover, it can also be used to perform the smoothing of the available observations, in order to reduce noise. We will thus use this technique both

to reconstruct data where missing, and to obtain a smoother and less noisy time series in all remaining time instances. All details can be found in Liu *et al.* (2012).

All the time series data of all failure scenarios are, then, normalized from 0 to 1.

3.2. Strategies to Build Sub-Models

Since we have a time series data set and since there is no other information available related to the target except for a set of monitored data directly related to the condition of the component of interest, the input vector of the model can only be a set of historical values. Before building the sub-models of the ensemble, we, thus, need to decide the best number of historical values to be used as inputs.

3.2.1. Sub-Model Identification

For time series data, the inputs are normally a number of historical target values. Suppose $a(t)$ represents an instance of the time series data of one failure scenario. For 1-day ahead prediction, the output $y(t)$ is $a(t+6)$, because the signals are monitored every four hours. In order to decide the best H for selecting the input vector $\boldsymbol{x}(t) = (a(t-H+1), \dots, a(t))$ most related to the output, a partial autocorrelation analysis is carried out on each failure scenario, i.e. the correlation between the output and different temporal lags is computed. Figure 3 shows the results of this analysis on Scenario 1, where the x and y axis represent, respectively, the temporal lag (a multiple of four hours) and the corresponding empirical partial autocorrelation. The bounds of a 95% confidence interval are also shown with dashed lines in the Figure. The correlation decreases with the lag (although not linearly), and after a lag of 17 time steps, for Scenario 1 it is no longer comparable with the values observed for lags smaller than 17, i.e. the best choice is $H_1 = 17$.

A best value H_i is, thus, found for Scenario i, for $i = 1, 2, \dots, 20$; but this value is not the same for all scenarios, as shown in the third column of Table 1. When building an ensemble model, however, a unified size of input vector would simplify the model, since a single value of H is applied for all scenarios to reconstruct the data. If we choose a small H, some useful information would be ignored for those scenarios with larger best H; in contrast, choosing a large H would bring some perturbations to scenarios with smaller best H. In order to solve this problem, we propose the modified RBF, where $\boldsymbol{C}_a$, calculated by partial autocorrelation analysis, controls the contribution of each variable of the input vector, when H is chosen as the largest of all the failure scenarios. For one scenario with smaller best H_i, the values for the last $H - H_i$ elements of the vector $\boldsymbol{C}_a$ are very small compared to the first H_i elements, because their correlations with the output are very weak. In this case study, we choose the biggest time step H of all the scenarios, i.e. $H = 17$.

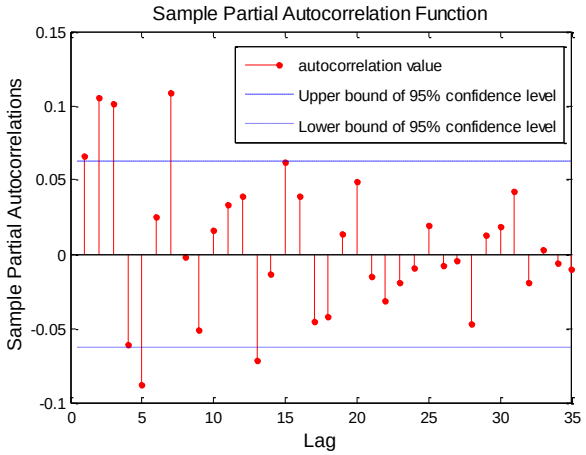

Figure 3. Partial autocorrelation function of Scenario 1 with respect to time lags (multiples of four hours). Dotted lines are bounds of a 95% confidence interval.

3.2.2. Two Strategies to Build Sub-Models

Bagging and boosting are two of the most popular strategies to build diverse sub-models of an ensemble. However these methods are more suitable with scarce data. In our case, there are enough data (20 failure scenarios), so that two simple but efficient and reasonable strategies can be proposed.

Thanks to the sub-model identification process described before, the data for each failure scenario has been reconstructed with same structure, where the input vector is $\boldsymbol{x}(t) = (a(t-16), \dots, a(t))$, and the corresponding output is $y(t) = a(t+6)$, and t takes every possible value in each scenario. The size of each failure scenario after reconstruction is listed in the fourth column of Table 1.

With multiple failure scenarios available, the simplest and most immediate strategy is to build a sub-model on each failure scenario, so that the number of sub-models equals the number of failure scenarios. Because of the frequently changing operational and environmental conditions in NPP, each scenario can represent a specific process, and thus sub-models built in such a way show enough diversity between each other. Another simple but effective strategy is to mix all the data points from all failure scenarios, and then divide them into different groups according to their target values $y(t)$. A sub-model is, then, trained on each group. This strategy is inspired by the intrinsic structure of SVM/PSVR. Performance of SVM depends highly, although not only, on the training data set (or support vectors). Sub-models built on training data set considering different ranges of output values can strengthen the specialty of each sub-model on particular characteristics of the input vectors. This strategy can make the sub-models perform well on different text

examples but worse on others. The proposed weighted-sum strategy to combine the outputs of sub-models will be shown to outperform the individual model. These two strategies are named Ensemble 1 and Ensemble 2, for convenience.

3.2.3. Comparison of DW-PSVR-Ensemble with Single PSVR

The ensemble model is expected to give better results than a single PSVR model. To verify this claim, a comparison between a single PSVR model and the proposed DW-PSVR-Ensemble is carried out on the considered case study.

Each time one out of 20 failure scenarios is chosen as the test data set (named Observed Scenario), the other 19 failure scenarios (named Reference Scenarios) are used to construct the ensemble model with the two previously proposed strategies. A PSVR model is also trained on the Observed Scenario for comparison (it is named Single PSVR to be distinguished from the two ensemble models). The size of the training data set for all PSVR models is fixed at 200 for the fairness of comparison. The choice of the size is decided by trial and error in order not to increase too much the computational complexity in time and storage, which increases exponentially with the size of the training data set, and in order to guarantee the accuracy of the model.

The steps of comparison are the following:

1. Choose the training data set for Ensemble 1: 200 data points equidistantly distributed for each Reference Scenario are selected. Totally, 19 sub-models can be trained with PSVR, each trained on 200 data points from each scenario.
2. Choose the training data set for Ensemble 2: the normalized data of 19 Reference Scenarios are sorted according to the output value of each data point and then divided into 10 groups, with the output value in the intervals of [0, 0.1], [0.1, 0.2], …, [0.9, 1]. For each group, if the size is bigger than 200, 200 data points equidistantly distributed in the group are chosen, if not, all the points in the group are used in the training data set. For the first eight groups, the size of training data set is 200, while for the last 2, the training data sets contain only 90 and 33 data points. Ten sub-models are built with PSVR on these training data sets.
3. Choose the training data set for the single PSVR: the first 200 data points of the Observed Scenario are chosen to train one single PSVR model for regression on it.
4. Calculation of Mean Absolute Error (MAE), Mean Relative Error (MRE), width of Prediction Intervals (PIs) with 95% confidence level (PI_Width), and coverage percentage of PIs with 95% confidence level (PI_Coverage) of the outputs of Ensemble 1, Ensemble 2 and Single PSVR.
5. Comparison of Ensemble 1, Ensemble 2 and Single PSVR considering prediction accuracy, uncertainty of estimation and robustness.

The results and comparisons between these three models are presented in the next section.

4. RESULTS

In this section, the results from Ensemble 1, Ensemble 2 and Single PSVR are compared with respect to different aspects.

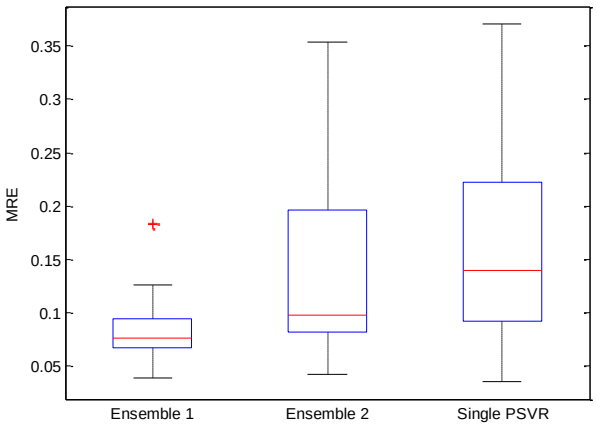

Figure 4. MAE of prediction results of Ensemble 1, Ensemble 2 and Single PSVR, for all 20 failure scenarios.

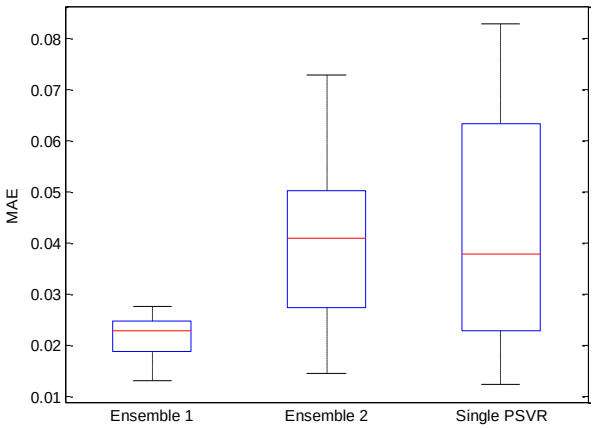

Figure 5. MRE of prediction results of Ensemble 1, Ensemble 2 and Single PSVR, for all 20 failure scenarios.

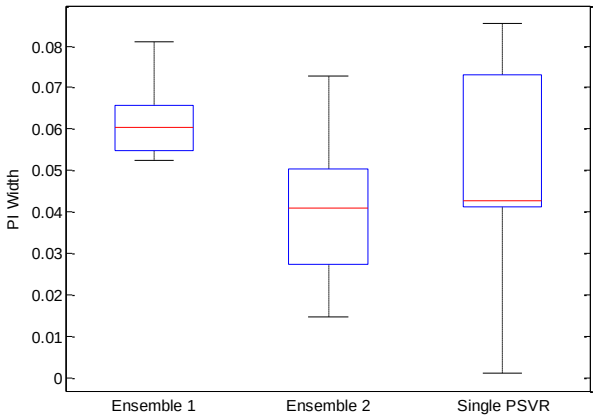

Figure 6. Width of PIs with 95% confidence level of prediction results of Ensemble 1, Ensemble 2 and Single PSVR, for all 20 failure scenarios.

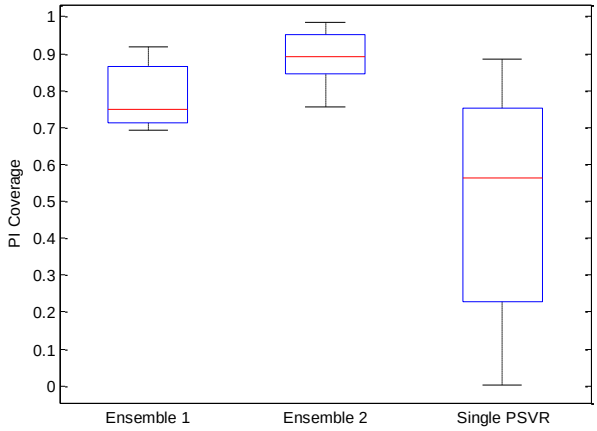

Figure 7. Coverage of PIs with 95% confidence level of prediction results of Ensemble 1, Ensemble 2 and Single PSVR, for all 20 failure scenarios.

4.1. Prediction Accuracy and Uncertainty Estimation

Figures 4, 5, 6 and 7 are the prediction results (including MAE, MRE, width and coverage of PI with 95% confidence level, i.e. $[\hat{y} - 1.97\sigma, \hat{y} + 1.97\sigma]$, where σ is the variance of the assumed Gaussian distribution of the predicted value), respectively from Ensemble 1, Ensemble 2 and Single PSVR. It is clear that Single PSVR gives worst results in this case study, i.e. on average, the MSE and MAE are bigger than the ensembles and PI_Coverage is lower compared to the ensembles. There is no such problem with Ensemble 1 and Ensemble 2, because the training data set contains more information than that of Single PSVR. Moreover, Ensemble 1 gives better results than Ensemble 2 considering the prediction accuracy, with more stable PIs. This is caused by the scarceness of the training data set for the last two sub-models of Ensemble 2, which are supposed to be experts on the prediction of the data points with output values in the intervals of [0.8, 0.9] and [0.9, 1.0].

We also notice that Single PSVR can give comparable prediction accuracy to the ensemble models for some failure scenarios, but not for all of them. The bad results of Single PSVR are caused by the fact that the prediction results are highly dependent on the training data set. Moreover, the hyperparameters optimization is also critical to the performance of PSVR. Well-chosen hyperparameters values can improve the performance of PSVR. However, the optimization method can easily converge to a local extreme, which results into a good performance at the beginning but very bad at the end of the scenario.

These unstable results from the Single PSVR prove the necessity of the ensemble approach for avoiding the limits of Single PSVR in attaining the desired accuracy and robustness of the model. The prediction results from Ensemble 1 and Ensemble 2 confirm the practicability and efficiency of the DW-PSVR-Ensemble approach.

4.2. Robustness

From Figures 4, 5, 6 and 7, it is seen that the ensemble models give more stable prediction results compared to the Single PSVR model. Single PSVR model cannot properly handle the noise in the data and it is difficult to find the global optimal values of the hyperparmeters, even with the modified RBF proposed in this paper. The weighted-sum ensemble models can decrease the influence of the noise by combining the prediction outputs of the sub-models; this is one reason for which ensemble models can give stable results, i.e. the ensemble models are more robust compared to the Single PSVR.

5. CONCLUSION

In this paper, we have proposed an innovative dynamic-weighted PSVR-based ensemble approach for short-term prediction (1-day ahead prediction) with multiple time series data. Local fitness calculation is integrated to calculate the specific weights of the sub-models of the ensemble for each new input vector without bringing too much computational burden. A modified RBF kernel is used to discriminate the different correlation of the different inputs with the output.

According to the characteristics of the available time series data in the case study, two strategies are proposed to form an ensemble model: one considering different scenarios and the other selecting different ranges of output values. In both cases, the proposed ensemble approach performs well in the real case study of signals recorded on a NPP component. Compared to the single model PSVR, the proposed ensemble models outperform on prediction accuracy, robustness and adaptability. This ensemble approach demands enough data on different pattern drifts.

Further research needs to be carried out, for optimizing the numbers of sub-models and for obtaining a more careful tuning of the hyperparameters.

REFERENCES

Acar, E., & Rais-Rohani, M. (2009). Ensemble of metamodels with optimized weight factors. *Structural and Multidisciplinary Optimization*, *37*(3), 279-294.

Bauer, E., & Kohavi, R. (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine learning*, *36*(1-2), 105-139.

Baudat, G., & Anouar, F. (2003). Feature vector selection and projection using kernels. *Neurocomputing*, *55*(1-2), 21-38.

Chen, S., Wang, W., & Van Zuylen, H. (2009). Construct support vector machine ensemble to detect traffic incident. *Expert systems with applications*, *36*(8), 10976-10986.

Gao, J. B., Gunn, S. R., Harris, C. J., & Brown, M. (2002). A probabilistic framework for SVM regression and error bar estimation. *Machine Learning*,*46*(1-3), 71-89.

Gurram, P., & Kwon, H. (2013). Sparse kernel-based ensemble learning with fully optimized kernel parameters for hyperspectral classification problems.*Geoscience and Remote Sensing, IEEE Transactions on*, *51*(2), 787-802.

Hu, C., Youn, B. D., Wang, P., & Taek Yoon, J. (2012). Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. *Reliability Engineering & System Safety*, *103*, 120-135.

Kim, H. C., Pang, S., Je, H. M., Kim, D., & Yang Bang, S. (2003). Constructing support vector machine ensemble. *Pattern recognition*, *36*(12), 2757-2767.

Liu, J., Seraoui, R., Vitelli, V., & Zio, E. (2013). Nuclear power plant components condition monitoring by probabilistic support vector machine. *Annals of Nuclear Energy*, *56*, 23-33.

Minh, H. Q., Niyogi, P., & Yao, Y. (2006). Mercer's theorem, feature maps, and smoothing. In *Learning theory* (pp. 154-168). Springer Berlin Heidelberg.

Muhlbaier, M. D., Topalis, A., & Polikar, R. (2009). Learn. NC: Combining Ensemble of Classifiers With Dynamically Weighted Consult-and-Vote for Efficient Incremental Learning of New Classes. *Neural Networks, IEEE Transactions on*, *20*(1), 152-168.

Polikar, R. (2006). Ensemble based systems in decision making. *Circuits and Systems Magazine, IEEE*, *6*(3), 21-45.

Razavi-Far, R., Baraldi, P., & Zio, E. (2012). Dynamic Weighting Ensembles for Incremental Learning and Diagnosing New Concept Class Faults in Nuclear Power Systems. *Nuclear Science, IEEE Transactions on*, *59*(5), 2520-2530.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and computing, 14(3), 199-222.

Sollich, P. (1999). Probabilistic interpretations and Bayesian methods for support vector machines. In *Artificial Neural Networks, 1999. ICANN 99. Ninth International Conference on (Conf. Publ. No. 470)* (Vol. 1, pp. 91-96). IET.

Valentini, G., & Dietterich, T. G. (2003, August). Low bias bagged support vector machines. In *ICML* (pp. 752-759).

Yang, X., Yuan, B., & Liu, W. (2009, November). Dynamic Weighting Ensembles for incremental learning. In *Proc. of IEEE conference in pattern recognition* (pp. 1-5).

BIOGRAPHIES

Jie Liu (B.Sc. in Mechanical Engineering, Beihang University, 2009; M.Sc. in Physics (at the same time, Engineer in Nuclear Energy), Ecole Centrale Pekin, Beihang University, 2012). He is now a PhD candidate of the Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy - EDF, CentraleSupélec, France. His current research interests concern kernel-based pattern recognition methods for prognostics of nuclear component's Estimated Time To Failure (ETTF).

Valeria Vitelli received the Ph.D. degree in Mathematical Models and Methods for Engineering, with a focus on statistical models for classification of high-dimensional data, in May 2012. She worked as a postdoc researcher within the Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy - EDF, CentraleSupélec, France from February 2012 to May 2013. She is currently a postdoc researcher in the Department of Biostatistics, University of Oslo, Norway. Her current research interests concern prediction and uncertainty quantification methods for complex data."

Enrico Zio (B.Sc. in nuclear engineering, Politecnico di Milano, 1991; M.Sc. in mechanical engineering, UCLA, 1995; Ph.D., in nuclear engineering, Politecnico di Milano, 1995; and a Ph.D., in nuclear engineering, MIT, 1998). He is the Director of the Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy - EDF, CentraleSupélec, France, and the full Professor of Computational Methods for Safety and Risk Analysis in Politecnico di Milano, Italy. He is President of Advanced Reliability, Availability and Maintainability of Industries and Services (ARAMIS) ltd., and Chairman of the European Safety and Reliability Association, ESRA. His research topics are: analysis of the reliability, safety and security of complex systems under stationary and dynamic conditions, particularly by Monte Carlo simulation methods; development of soft computing techniques (neural networks, support vector machines, fuzzy and neuro-fuzzy logic systems, genetic algorithms, differential evolution) for safety, reliability and maintenance applications, system monitoring, fault diagnosis and prognosis, and optimal design. He is co-author of five international books and more than 200 papers on international journals, and serves as referee of more than 20 international journals.

Redouane Seraoui works for EDF in the Research & Development division. Following the graduation from the French engineering school "Ecole Superieure d'Electricite," he works now on a series of large research projects involving collaboration between many leading academic and industry partners. These projects include the development of tools and methods for the monitoring of power plants' critical components, applied to the nuclear industry but also to the fossil and renewable sides.