# Supplementary material
# Technical Report n. TR-1-2011
# Early Prediction of the Highest Workload in Incremental Cardiopulmonary Tests

Elena Baralis [1], Silvia Chiusano [1], Vincenzo D'Elia [1], Riccardo Molinari [2], Davide Susta [3]

## I. Introduction

This appendix reports additional experiments to support the evaluation of the proposed approach for the early prediction of the $W_{peak}$ value. These experiments have not been included in the paper due to the lack of space.

The appendix includes the following three sections. Section II describes the datasets used for the experiments. Section III compares the performance between our approach and a naive predictor. Section IV analyses the impact of parameters setting on the prediction error. Sections II and III are an extended version of the corresponding sections available in the paper. Section IV is instead a new section.

## II. Datasets

To gather representative examples, tests have been done with two protocols commonly used in endurance sport testing to elicit the highest workload in both elite and intermediate athletes. Protocols are 50 $W \times 2$ min and 25 $W \times 2$ min, and the corresponding tests have been collected in two datasets named $D_{50\times2}$ and $D_{25\times2}$, respectively. Table I reports the main characteristics of both datasets, and Figure 1 shows the distribution of Age, BMI, BSA, and $W_{peak}$ values. Datasets include male athletes, both amateur and elite athletes.

Protocol 50 $W \times 2$ min is more stressful for the athlete body, because of the higher increment of workload applied at each step. For this reason, athletes in $D_{50\times2}$ are typically younger and more trained than in $D_{25\times2}$, and they usually reach higher $W_{peak}$ values.

Protocol 25 W × 2 min is applicable to diverse athletes, due to the lower increment of workload applied. Athletes in $D_{25\times2}$ are more heterogeneous with respect to age, BMI, and BSA values describing them (in particular when considering the joint distribution of these values).

| Dataset | No. of Tests | No. of Athletes | Parameters: Name | Parameters: Mean ± SD |
|---|---|---|---|---|
| $D_{50\times2}$ | 231 | 202 | Age | 32.3 ± 12.2 years |
| | | | BMI | 22.8 ± 2.3 Kg/cm$^2$ |
| | | | BSA | 1.8 ± 0.1 m$^2$ |
| | | | $W_{peak}$ | 333 ± 68 W |
| $D_{25\times2}$ | 184 | 139 | Age | 38.7 ± 12.5 years |
| | | | BMI | 23.9 ± 2.6 Kg/cm$^2$ |
| | | | BSA | 1.8 ± 0.1 m$^2$ |
| | | | $W_{peak}$ | 262 ± 62 W |

TABLE I
Characteristics of the two datasets. For all parameters, the mean and the standard deviation (SD) values are reported.

## III. Comparison with a naive predictor

To our knowledge, an approach for the early prediction of the $W_{peak}$ value in incremental tests has not been proposed yet. Thus, we compared our approach with a naive predictor which works as follows. Consider a new test $Q$, where the workload currently assigned in the test is $W_Q$. The naive predictor selects all tests with $W_{peak} \geq W_Q$ from the dataset with the same protocol as test $Q$. The average $W_{peak}$ on the selected tests is the $W_{peak}$ value predicted for test $Q$.

We compared the MAE value achieved with our approach and the naive predictor. For both techniques, the leave-one-out cross-validation method [1] is used for prediction error evaluation.

Experimental results reported in Figure 2 show that our approach is always more accurate (except in dataset $D_{25\times2}$ at time 26 min, where the error by the naive predictor is slightly lower).

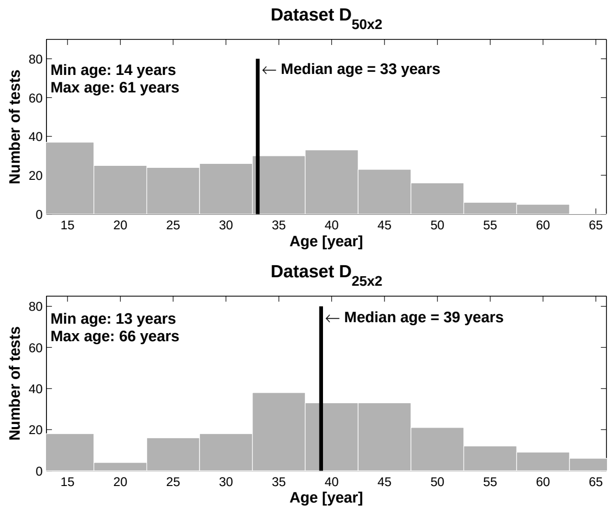
We also compared the MAE value obtained by analyzing separately tests reaching the same $W_{peak}$. Figures 3 and 4 plots the results for datasets $D_{50\times2}$ and $D_{25\times2}$, respectively. The naive predictor is characterized by the following behavior.

(i) It is slightly more accurate than our approach for tests reaching $W_{peak}$ values close to the average $W_{peak}$ in the dataset. In dataset $D_{50\times2}$, with average $W_{peak}$ 333 W, it occurs for tests characterized by $W_{peak}$ equal to 300 W and 350 W (Figures 3(b) and 3(c)). In dataset $D_{25\times2}$, with

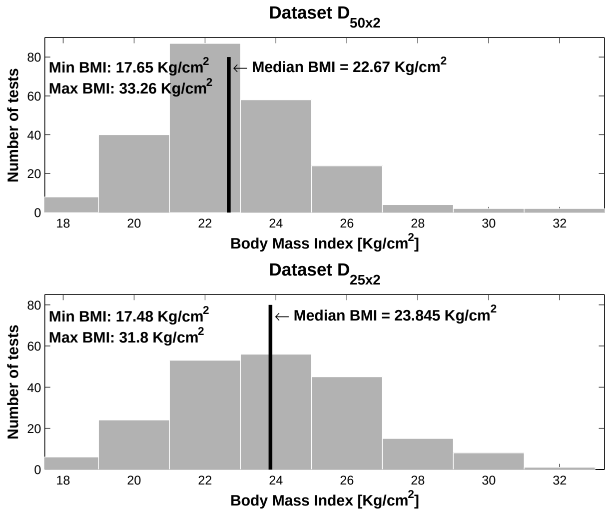
[1] Department of Control and Computer Engineering, Politecnico di Torino, Italy (email: elena.baralis@polito.it; silvia.chiusano@polito.it; vincenzo.delia@polito.it)

[2] CSA, Sport Training Center, Eupilio, Como, Italy (email:riccardo@csa4sport.it)

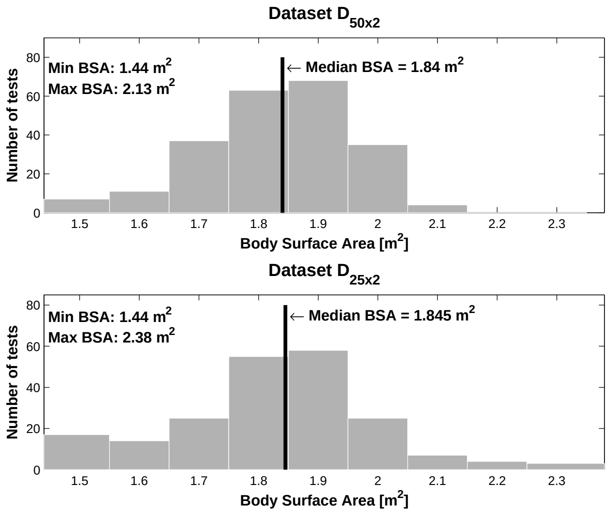
[3] School of Health and Human Performance, Faculty of Science and Health, Dublin City University, Ireland (email:davide.susta@dcu.ie)



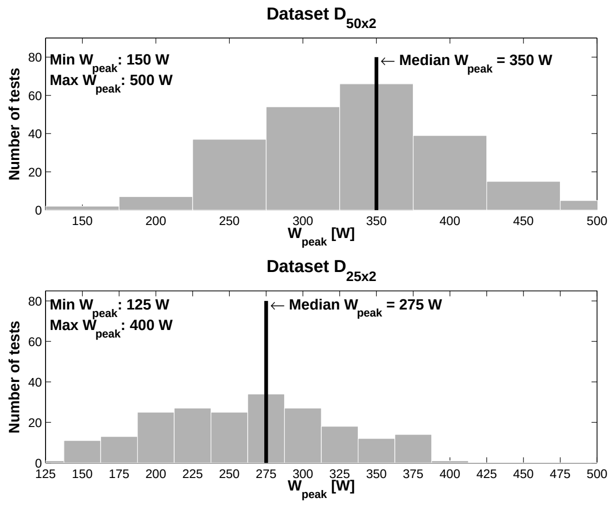

(a) Distribution of age




(b) Distribution of Body Mass Index (BMI)




(c) Distribution of Body Surface Area (BSA)




(d) Distribution of the Highest Workload ($W_{peak}$)

Fig. 1. Distribution of age, Body Mass Index (BMI), Body Surface Area (BSA), and Highest Workload ($W_{peak}$) in the datasets

average $W_{peak}$ 262 W, it occurs for tests reaching $W_{peak}$ between 250 W and 300 W (Figures 4(a), 4(b), and 4(c)). Otherwise our approach is always more accurate. For example, it reduces the MAE value up to 67% in dataset $D_{50\times2}$ for tests with $W_{peak}$=450 W (Figures 3(e)), and up to 60% in dataset $D_{25\times2}$ for tests with $W_{peak}$=375 W (Figures 4(f)).

(ii) The MAE value is constant when the prediction is performed in the first 4 min of the test in dataset $D_{50\times2}$, and in the first 8 min in dataset $D_{25\times2}$. In $D_{50\times2}$, all tests reach 150 W (4 min). It follows that, when the instant of prediction is within the first 4 min, the entire dataset is used to predict $W_{peak}$, and the $W_{peak}$ value is computed as average $W_{peak}$ on all tests in the dataset. Analogously in dataset $D_{50\times2}$, where all tests reach 125 W (8 min).

(iii) The MAE value usually shows a sharp increase when the prediction is performed in the last step of the test. The workload $W_Q$ assigned in the test at this step is the highest workload ($W_{peak}$) actually reachable in the test. To perform prediction, the naive approach computes the average $W_{peak}$ on tests with $W_{peak} \geq W_Q$. In the last step this average value is likely to be greater than $W_Q$, and thus the naive predictor tends to overestimate the actual $W_{peak}$ of the test. Instead, our solution always achieves a prediction error that progressively tends to zero when approaching the test end.

## IV. Effect of parameters setting on the Mean Absolute Error

The framework configuration depends on the following four parameters:

- the number of singular values of matrix $T$, used to compute $\varepsilon_T$ with the SVD technique;
- the *warping_band* value, to compare two $CPE_{peaks}[t]$ sequences using the DTW technique;
- the weight $w$ of the contributions from factual and dynamic data to evaluate test similarity;
- the number $k$ of the nearest tests, which are selected to predict the $W_{peak}$ value.

The reference configuration of parameters is the following: the number of considered singular values is 2; *warping_band* is set to 2 min; the contributions from factual and dynamic data have the same weight ($w = 0.5$); the 5-nearest tests are selected for $W_{peak}$ prediction ($k = 5$). This reference configuration allows achieving a limited prediction error in both datasets. Results may be further improved by tuning a specific configuration for each dataset.

To analyze how the variation of parameters impacts on the prediction accuracy, we fixed in turn three parameters to their reference values, and varied the remaining one. Besides the choice of parameters, the prediction accuracy is affected by the time when the prediction is performed. Thus, experiments are repeated by considering different instants of prediction. The leave-one-out cross validation method is used for error evaluation [1].

### A. Analysis of the number of singular values of matrix $T$

Experiments have been run with 1, 2, 3, and 9 singular values. Increasing the number of singularities reduces the action of signals filtering performed by means of the SVD technique. In this paper, the maximum number of possible singular values is 9, being 9 the physiological signals used to compute $\varepsilon_T$. The use of all available singularities corresponds to a non-truncated SVD solution, where no filtering step is performed.

Figure 5 show that, in both datasets, our approach is slightly sensitive to the number of singular values. The MAE value is slightly higher when increasing this number (e.g., 3 or 9 in Figure 6), suggesting that a filtering stage is important to deal with possible noise components.

### B. Analysis of the warping_band parameter

Figure 7 plots the MAE value when varying the *warping_band* parameter. Experimental results show that the DTW technique allows reducing the prediction error, especially when selecting small values of warping_band.

In both datasets, the highest prediction error is achieved when no time shifting is allowed in evaluating the similarity between two $CPE_{peaks}[t]$ sequences (*warping_band*=0 min). In this case, the test similarity is computed simply as the Euclidean distance between two $CPE_{peaks}[t]$ sequences having exactly the same length.

When *warping_band* is greater than 0 min, the DTW technique is properly exploited for sequences comparison, supporting a non-linear alignment between two sequences. Similar performance are obtained when the maximum tolerated time shifting is between 2 min and 6 min (between 1 and 3 steps in both protocols). The minimum error is achieved when limiting *warping_band* to 2 min, i.e., the reference value in our configuration.

### C. Analysis of the $w$ parameter

Figure 8 plots the MAE value when varying parameter $w$. Experimental results showed that in both datasets, the predicted $W_{peak}$ value well approximates the actual one when the same weigth is assigned to the similarity between factual and dynamic data in the two tests; otherwise, the prediction is slightly less accurate. Dataset $D_{25\times2}$ is more sensitive to the variation of parameter $w$, because of the characteristics of protocol $25 \times 2$.

When increasing $w$, the k-nearest tests are mainly selected based on the characteristics of the athletes performing the tests. As a consequence, tests in the knowledge base having body responses close to the running test Q, but done by athletes (slightly) dissimilar from the athlete of Q, may be missed. On the other hand, when decreasing $w$, the test similarity is mainly evaluated by comparing the athlete body responses to exercise. Increasing the relevance of this comparison could be misleading in the first steps of the test, when sequence $CPE_{peaks}$[t] still contains few peaks. This comparison could be misleading in particular in dataset $D_{25\times2}$. Since protocol $25\times2$ is less stressful for the athlete body, tests reaching different $W_{peak}$ values may initially show similar $CPE_{peaks}$[t] sequences.

Experimental results reported in Figure 8 show how the $w$ value affects the prediction error at different prediction instants. At the beginning of the test, a limited information is available about dynamic data. For this reason, better results are obtained when the test similarity is mainly evaluated by comparing athletes characteristics (i.e., the value of $w$ is high). Later, the running test is more precisely described by a longer $CPE_{peaks}$[t] sequence, and better results are obtained when the similarity is mainly evaluated based on the dynamic data (i.e., the value of $w$ is low). This behavior is particularly clear in dataset $D_{25\times2}$. However, in both datasets good results are obtained with respect to the entire test duration when $w = 0.5$.

The mean error reported in Figure 8 is computed by considering the prediction errors for all tests in the dataset, characterized by different durations. To better evaluate the impact of the $w$ parameter, we analyzed dataset $D_{25\times2}$ as a reference example and we considered separately a set of short tests ($W_{peak}$ 225 W) in Figure 9(a), and set of long tests ($W_{peak}$ 375 W) in Figure 9(b). In both the cases, the prediction is more accurate when $w = 0.5$.

### D. Analysis of the $k$ parameter

Figure 10 plots the MAE value when varying the $k$ parameter.

In both datasets, the highest error is obtained when a *single* nearest test is selected to predict $W_{peak}$ ($k = 1$).

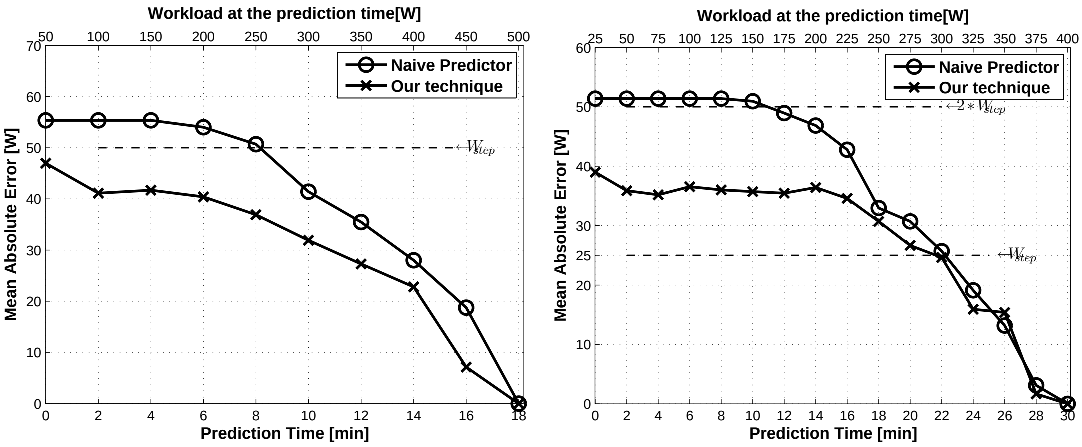
Values of $k$ greater than 1 (i.e., 5, 10, 15, 20) give similar performance in both datasets. Values $k = 5$ and $k = 10$

provide close results, even though sometimes $k = 5$ slightly improves $k = 10$ or vice versa, depending on both the dataset and the prediction time. Instead, values $k$=20 and $k$=15 provide a sudden increase in the MAE value when postponing the prediction time, i.e., for $k$=20 at time = 16 min in dataset $D_{50\times2}$ and for $k = 15$ at time = 24 min in dataset $D_{25\times2}$.

We observe that when parameter $k$ is close to the dataset size, our approach provides similar performance than the naive approach. In this case, all tests in the knowledge base contribute to predict $W_{peak}$, instead of using only the small subset of the most similar ones.

## References

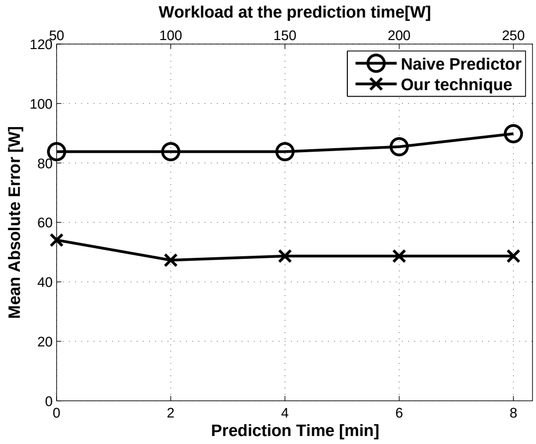
[1] P.N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Addison Wesley, 2005.



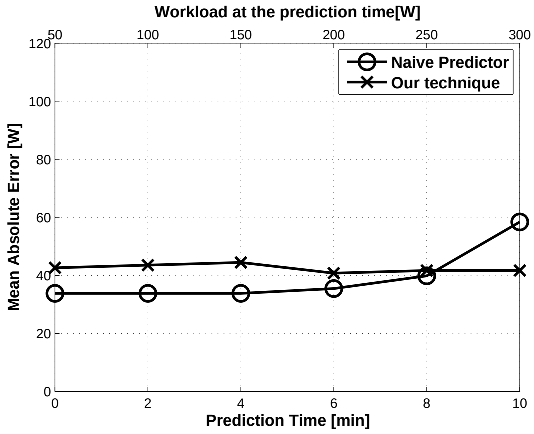

(a) Dataset $D_{50\times 2}$

(b) Dataset $D_{25\times 2}$

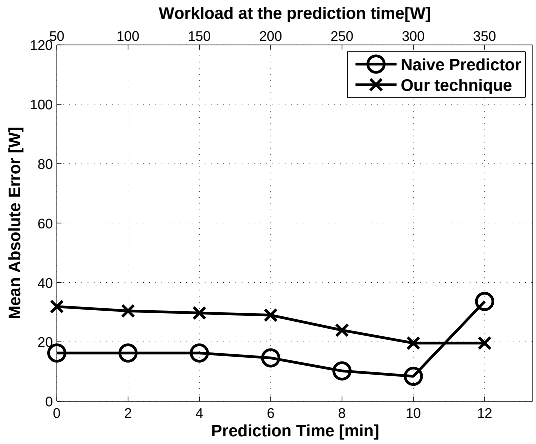
Fig. 2. Comparison with a Naive Predictor: trend of the Mean Absolute Error (MAE).



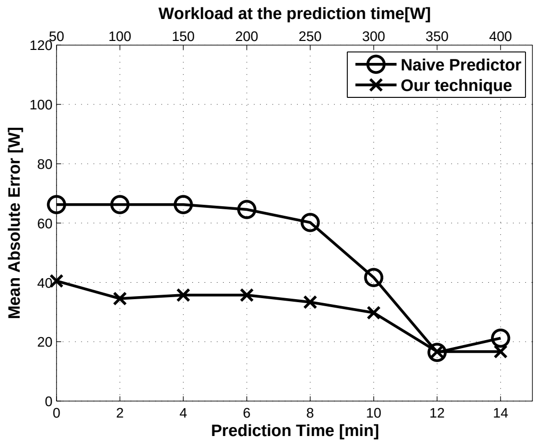

(a) $W_{peak}$ = 250 W



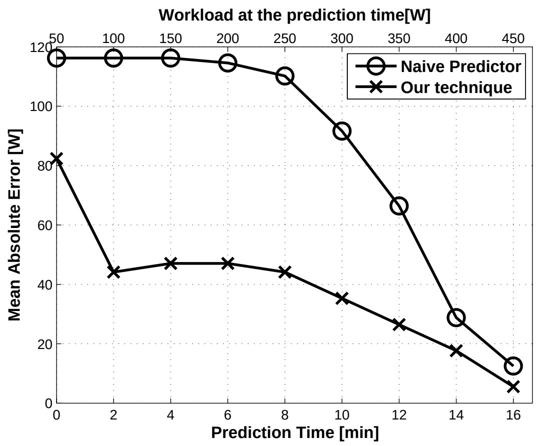

(b) $W_{peak}$ = 300 W




(c) $W_{peak}$ = 350 W

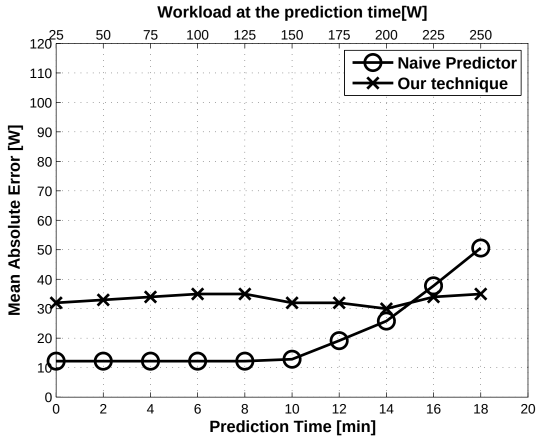



(d) $W_{peak}$ = 400 W

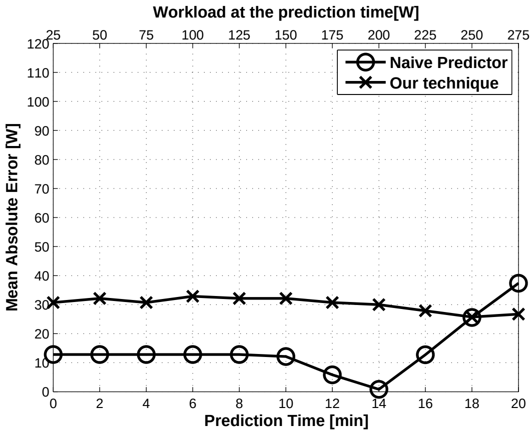



(e) $W_{peak}$ = 450 W

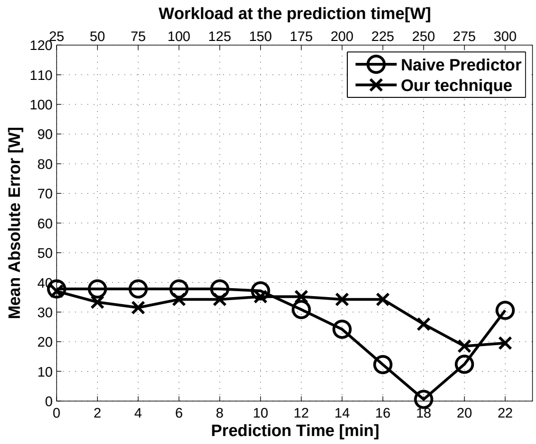
Fig. 3. Comparison with a Naive Predictor: trend of the Mean Absolute Error (MAE) for tests with different $W_{peak}$ values in dataset $D_{50\times2}$



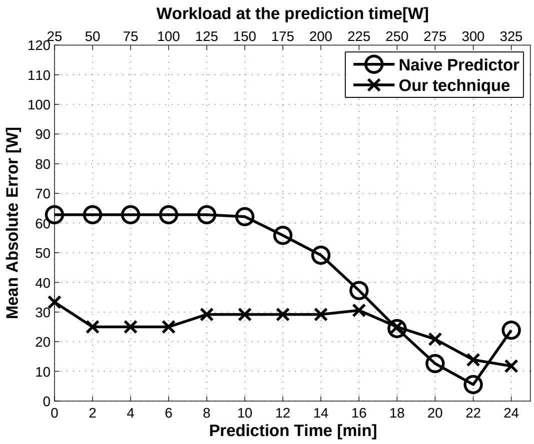

(a) $W_{peak}$ = 250 W



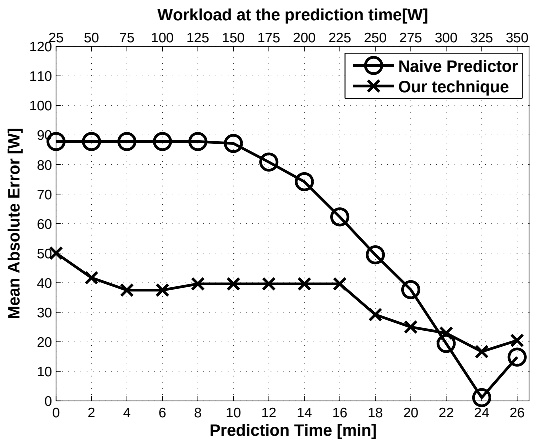

(b) $W_{peak}$ = 275 W



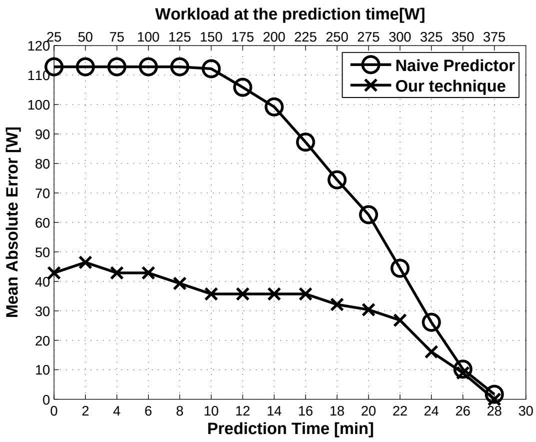

(c) $W_{peak}$ = 300 W




(d) $W_{peak}$ = 325 W

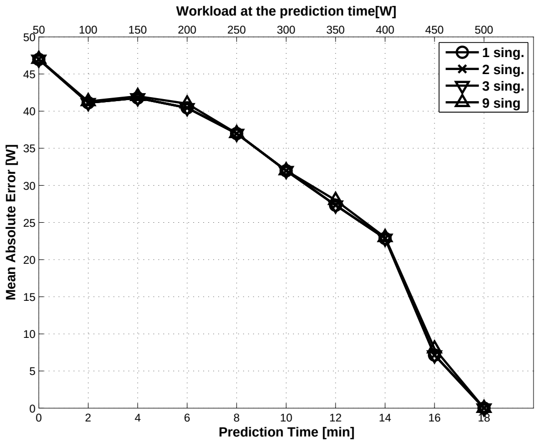



(e) $W_{peak}$ = 350 W

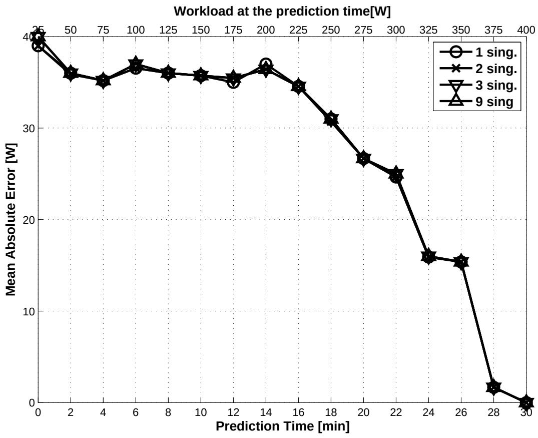



(f) $W_{peak}$ = 375 W

Fig. 4. Comparison with a Naive Predictor: trend of the Mean Absolute Error for tests with different $W_{peak}$ values in dataset $D_{25\times 2}$

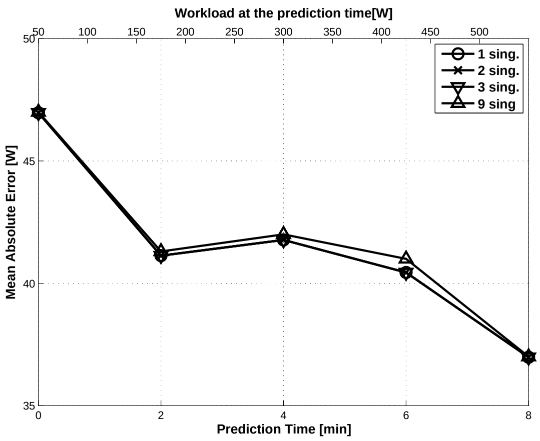



(a) Dataset $D_{50\times 2}$

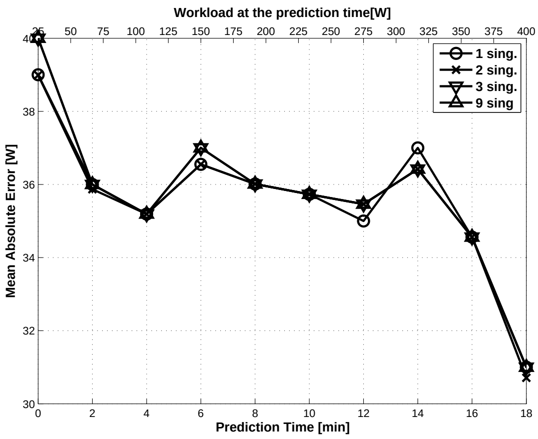



(b) Dataset $D_{25\times 2}$

Fig. 5. Impact of the number of considered singular values on the Mean Absolute Error.

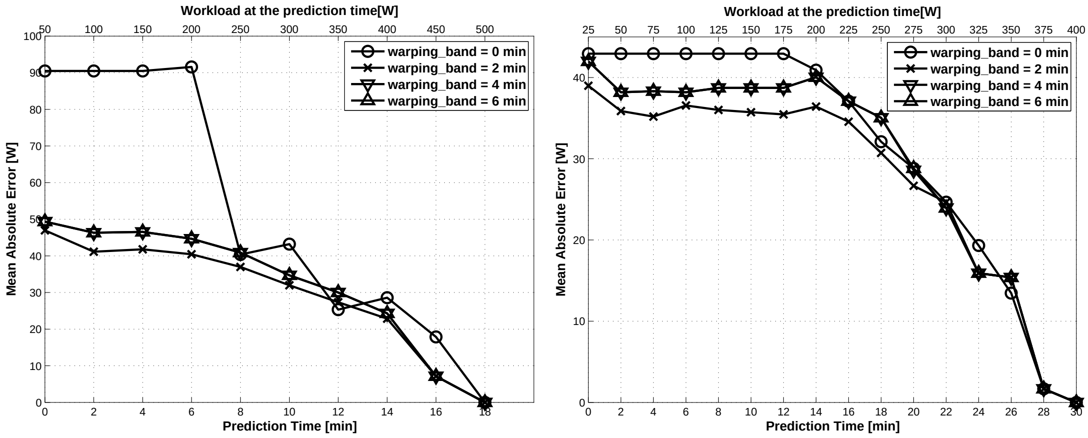



(a) Dataset $D_{50\times 2}$. Prediction time between 0 and 8 min.



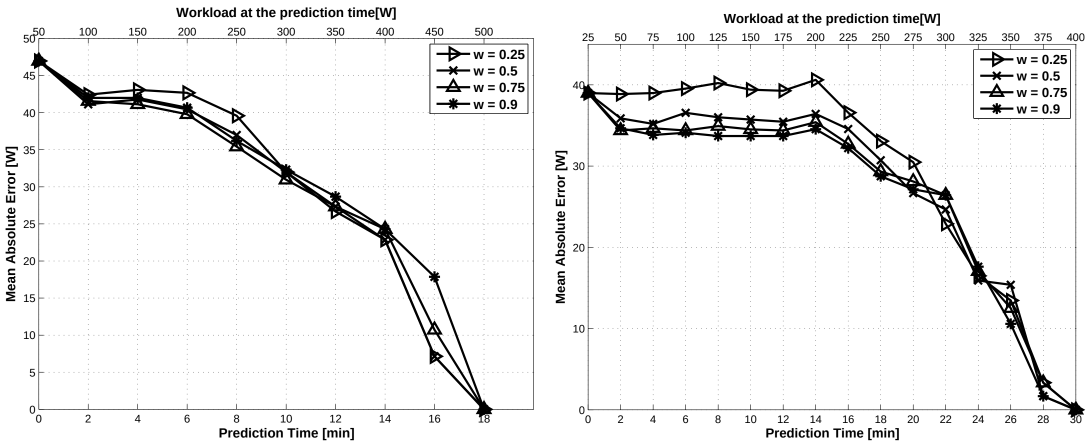

(b) Dataset $D_{25\times 2}$. Prediction time between 0 and 18 min.

Fig. 6. Impact of the number of considered singular values on the Mean Absolute Error.



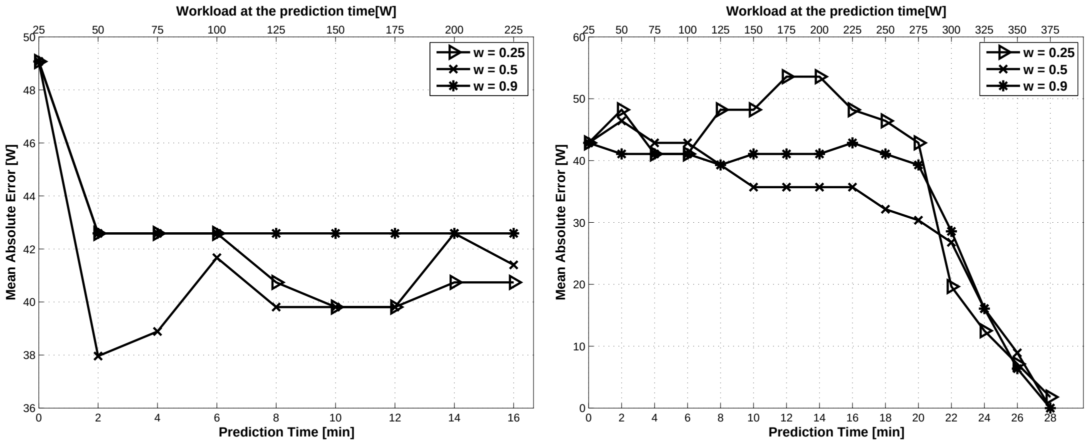

(a) Dataset $D_{50\times 2}$ (b) Dataset $D_{25\times 2}$

Fig. 7. Impact of the *warping_band* parameter on the Mean Absolute Error.




(a) Dataset $D_{50\times 2}$ (b) Dataset $D_{25\times 2}$

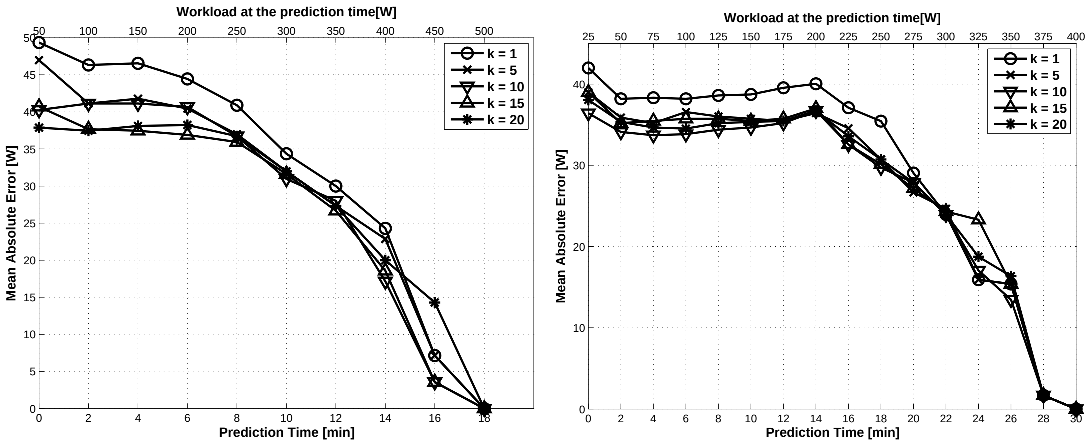
Fig. 8. Impact of parameter $w$ on the Mean Absolute Error.




(a) Impact of parameter $w$ on the MAE for tests with $W_{peak}$ = 225 W in dataset $D_{25\times2}$

(b) Impact of parameter $w$ on the MAE for tests with $W_{peak}$ = 375 W in dataset $D_{25\times2}$

Fig. 9. Impact of parameter $w$ on the Mean Absolute Error, for tests having different $W_{peak}$ in dataset $D_{25\times2}$.




(a) Dataset $D_{50\times2}$

(b) Dataset $D_{25\times2}$

Fig. 10. Impact of parameter $k$ on the Mean Absolute Error.