Sustainable Energy, Grids and Networks 41 (2025) 101604

Contents lists available at ScienceDirect

Sustainable Energy, Grids and Networks

journal homepage: www.elsevier.com/locate/segan

Enhancement of short-term prediction capabilities of inter-area grid oscillations with a multi-variate ensemble-based method

Carlo Olivieri [a,*], Francesco de Paulis [a], Lino Di Leonardo [a], Antonio Orlandi [a], Cosimo Pisani [b], Giorgio Giannuzzi [b]

[a] UAq EMC Laboratory, Department of Industrial and Information Engineering and Economics, University of L'Aquila, L'Aquila 67100, Italy
[b] TERNA S.p.A., v.le E. Galbani, 70, Rome 00156, Italy

A R T I C L E I N F O

Keywords:
Inter-area oscillations
Predictive monitoring
PMU
Modal analysis
Multi-variate prediction
Ensemble learning
Neural networks

A B S T R A C T

The actual and future even higher penetration of renewable energy sources into the power grid sets challenging issues for transmission system operators, especially concerning the hard-to-solve problem of inter-area electromechanical oscillations. Despite the useful existing monitoring systems, the possibility of having predictive monitoring features for such phenomena could be an appealing tool. The work presented in this paper aims to assess the possibility of enhancing the predictive monitoring features offered by machine learning techniques based on the combination of ensemble methods and Long-Short-Term Memory units using multi-variate methods. The development steps of a multi-variate prediction strategy are presented together with the assessment of its performance versus uni-variate solutions. The assessment takes into account different kinds of datasets, taken from real grid measurements, and strategy configurations. Either transient low frequency oscillation phenomena or normal grid operation are considered as representative cases of real-world scenarios. Finally, some preliminary considerations about improving prediction performance and the limitations are outlined.

1. Introduction

Due to the advent of the worldwide energy transition process, the management of modern power grids is becoming extremely challenging. In particular, the progressive reduction of the power grid inertia, due to the high penetration of renewable energy sources (RES), and the contemporary increase of power flows over long distances pose severe stability issues for the overall power system [1]. In this context, the occurrence of detrimental phenomena like the inter-area electromechanical Low-Frequency Oscillations (LFO), can heavily impair the process of RES integration. As shown from the literature [2–4], LFOs are difficult to control by using standard regulation techniques [5] because of some impactful aspects. First of all, the frequency range of such oscillations is very low, typically comprised between 0.1 and 0.8 Hz [6], and, in addition to this, the fact they occur on a geographical scale makes the application of already existing and mature control schemes not very effective [1–4]. Considering all of this, LFO phenomena constitute at today one of the major hot-topics for Transmission System Operators (TSO) since they either impair the integration of RES or prevent the operation of the grid at its maximum power transfer capacity. As well-known from the literature, the possibility of damping or even preventing dangerous LFO oscillatory (ringing) events is strictly related to the possibility of identifying such critical phenomena [3,4,7]. For what concerns LFO identification, several techniques have been proposed and successfully developed in the past to describe LFO ringing events in terms of a set of characteristic parameters [5,7–16]. All these techniques can be categorized mainly as model-based techniques [7–10] or data-driven (i.e. measurement-driven) techniques [3,11–16]. In a large part of these LFO identification methods, the core approach is mainly founded on the proper estimation of the so-called "modal parameters" of the oscillatory phenomenon [3,5,10,14–17]. Basically, they are triplets of values identifying the oscillatory behavior of a certain electromechanical mode of the LFO based on a model following an exponentially damped sinusoidal representation. The improvement of grid stability claims for twofold necessities. On one side, the damping of LFO ringing events requires enhanced control techniques, on the other side, the use of predictive monitoring features can support the anticipated identification of potentially critical situations. Several literature

* Corresponding author.
E-mail addresses: carlo.olivieri@univaq.it (C. Olivieri), francesco.depaulis@univaq.it (F. de Paulis), lino.dileonardo@univaq.it (L. Di Leonardo), antonio.orlandi@univaq.it (A. Orlandi), cosimo.pisani@terna.it (C. Pisani), giorgio.giannuzzi@terna.it (G. Giannuzzi).

https://doi.org/10.1016/j.segan.2024.101604
Received 15 July 2024; Received in revised form 14 December 2024; Accepted 14 December 2024
Available online 19 December 2024

approaches proposed the analysis of LFO phenomena through the tracking (also in real-time) of the associated modal parameters [3,4,14–16]. Even though the identification of detrimental LFO events can be of certain utility, the opportunity to have indications about the future behaviour constitutes a desirable next step for the building of sophisticated management tools suitable to support the TSO in keeping the grid working properly and at the edge of its capacity. In addition to this, the possibility of having accurate forecasts of the trend of modal parameters during sub-critical ringing events can help to trigger early-warning signals and then, in turn, more effective control actions. This results in a decreased vulnerability of the grid and a potentially increased resilience. Even though modal parameter estimation techniques are enough mature for LFO identification purposes, predictive monitoring techniques are still an open issue. On the other hand, the quick and ever-increasing development of sophisticated measurement systems could disclose a wide set of additional capabilities; this is the case for instance of Phasor Measurement Units (PMUs) and Wide Area Measurement Protection and Control (WAMPAC) solutions [3,8,14–17]. These systems can be used for the realization of well established measurement/control systems or for developing cutting-edge monitoring algorithms. PMU devices have been successfully used to either identify and track LFO modes in real-time [15–19] or to implement Machine Learning (ML) algorithms suitable to enable or support the damping of LFO phenomena [7,20]. ML techniques seem to be promising for both estimation and forecasting purposes [17–21], however, their application for the predictive monitoring of modal parameters associated with LFO phenomena is still an unexplored field of study, based on what is suggested by the actual literature. Most of the presented approaches are in fact not directly linked with the analysis of LFO ringing conditions or to the prediction of grid stability. Several contributions analyzed the use of ML-based methods for short-term load forecast (STLF) problems, either with standard techniques [22–24] or complex attention-based techniques [25], or for the prediction of different kind of faults [21,26] or weather related outages [27,28]. In other cases, instead, ML algorithms are applied to more specific applications like the evaluation of the frequency response of the grid [29] or its short-term dynamic behavior [30], the prediction of voltage stability [31], the prevention of line overloading [32], or system restoration. The ML methods applied in these contributions are mainly related to the use of Artificial Neural Networks (ANN) [32,33], Support Vector Machines (SVM) [22], Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) [25] or different kinds of Ensemble methods [28]. In particular, the use of Ensemble methods in some specific application domains has been extremely useful to manage situations characterized by the so-called "Concept Drift", that is the property of a system to change its dynamic behavior over time. The same Concept Drift (CD) concept is the one involved in problems where the ML-based techniques should learn under non-stationary conditions, that is the case of the actual domain where the grid is a kind of "live" object. Ensemble-based methods have been already used in the literature to face challenging non-stationary regression problems and on-line learning tasks as in [34–38]. There are recent contributions aimed to explore the suitability of ML-based methods for the prediction of the time evolution of LFO phenomena based on the short-term forecast of the associated modal parameters [39,40]. In one of them [40] a strategy based on the proper integration of Recurrent Neural Networks with an Ensemble method has been proposed to forecast the modal parameters' values few tens of seconds in advance in the future. In this case, the forecast has been done by setting up a 1-step time-series prediction problem exploiting a uni-variate structure. In further contributions [39] the possibility to have an extend prediction horizon has been evaluated by setting up a multi-step time-series prediction problem and the assessment of different prediction architectures has been also developed [39]. In all the approaches presented in [39,40] one strong limitation is the uni-variate prediction problem structure, lacking the possibility to exploit all the informative data embedded in the input datastream of the problem at-hand. The target of this paper is the evaluation of the possibility of enhancing the short-term prediction accuracy of the behaviour of the modal parameters by exploiting a multi-variate prediction problem structure. In particular, to have the possibility to compare the results coming from the multi-variate prediction problem addressed in this paper with the ones obtained with the uni-variate structure, the study has been restricted to single-step (1-step) predictions. Owing to this aim, a multi-variate prediction approach based on the joint use of RNN-based algorithms and Ensemble methods is described in this paper and the comparison of the prediction accuracy of the multi-variate approach with respect to the uni-variate one is also presented. Since the presented study is based on measurement data only, the possibility of assessing the generalization features of the predictive model against arbitrary CD issues is quite limited and out of the scope of this paper. As already done in previous works [39,40] the core business of the approach takes into account the use of a proper modal parameter estimation technique that is the Dynamic Mode Decomposition (DMD) [41–43] which is also used as a dimensionality reduction technique, in place of other techniques like the Principal Component Analysis (PCA) [44].

2. Application background and overall prediction approach

This section describes the background of the application scenario and recalls the main points about the detection and characterization of LFOs occurring in the power grid through the use of modal theory. This section also outlines the contour of the proposed LFO prediction approach and highlights the technical constraints characterizing the application at hand.

2.1. Low-frequency oscillations in electric power systems

As already mentioned in the introductory section, the tracking of the values assumed by the modal parameters has been retained an interesting tool [15] suitable to characterize the LFO phenomena. Basically, these parameters are constituted by triplets of values identifying the oscillatory behavior of a certain electromechanical mode, following an exponentially damped sinusoidal (EDS) model [5,41–44]. Each triplet (referred to a specific electromechanical mode) is composed by the following three well-known quantities: frequency (f_i), damping ratio (α_i), and amplitude (a_i), where the subscript i stands for the generic *i-th* electromechanical mode of interest. For the application under study, the desired target is a monitoring system capable of tracking the time evolution of the modal parameters and providing short-term predictions of their future values.

For what concerns the possibility to have a real-time or quasi-real-time tracking of the modal parameters, there are already mature and well-known estimation algorithm suitable to extract the modal parameters from measurements coming from PMU devices. Even though plenty of estimation algorithms exist [5,14,15,41] for this purpose, in the case of this paper the Dynamic Mode Decomposition method has been considered [41]. The reasons are mainly linked to the fact that it offers some additional advantages in the context of this specific application scenario. First of all, it has the possibility to operate a dimensionality reduction (as already mentioned in the introduction) based on the number of modes selected for the study (i.e. depending on the relevance of the specific electromechanical modes). In this paper the number of considered/selected modes is indicated by the parameter N_{mod}. Secondly, the DMD method has good numerical stability and limited computational burden, making it ideal for implementation and integration over the real grid monitoring system. According to the specific literature, the DMD method is founded on the collection of snapshots of the input data that has to be arranged in the form of a sequence having the structure recalled in (1)

$$X_1{}^N = \{x_1, x_2, \cdots, x_N\} \tag{1}$$

where x_i is the *i-th* snapshot of data and X_1^N is the resulting aggregate data matrix whose columns constitute the different snapshots of data. It has to be remarked that here each snapshot corresponds to the acquisition of the instantaneous frequency from a set of PMU devices placed along the power grid at different locations (the number of locations is arbitrary and indicated as p). Several literature works demonstrated that the DMD method is capable to successfully extract the modal parameters of electromechanical modes starting from instantaneous frequency measurements only [15,17]. The DMD method is based on the theory for which the modal parameters can be extracted by establishing a linear mapping between the *N-th* snapshot of data and the *N-1-th* one. This linear mapping can be stated as reported in (2)

$$X_2^N = AX_1^{N-1} + r \tag{2}$$

In this formulation, X_1^{N-1} and X_2^N are two subsets of the initial collection of data and the term r indicates the vector of residuals which accounts for the dynamic behaviours that cannot be captured by the linear mapping. From a computational point of view the DMD algorithm operates by executing three main steps: a Singular Value Decomposition (SVD), a linear projection, and, an Eigenvalue Decomposition (EVD).

As stated by fundamental works on this topic [41–44], the dimensionality reduction feature of the DMD method is provided by the fact that the EVD procedure can be executed by considering a reduced order matrix with dimensions much smaller than the ones of the full dynamical matrix A in (2). This aspect translates into the possibility to have a very high spatial sampling of the measurements (high number of PMUs) with a low number of estimated electromechanical modes and high computational efficiency. This is an appealing point with respect to other dimensionality reduction techniques like PCA [44], Independent Component Analysis (ICA) [45] or other feature extraction techniques.

2.2. Predictive monitoring application scenario and constraints

Considering all the background information already presented up to here, the specific application scenario of interest in this paper can be summarized with a general scheme like the one reported in Fig. 1. As can be seen from here, the instantaneous frequency measurements coming from a subset of the available PMU devices (indicated as *PMU 1* up to *PMU p*) are sent to a proper prediction strategy after a preliminary signal preconditioning stage. The pre-processed input data stream enters the ML-based prediction strategy and, at this stage, for each electromechanical mode, the associated modal parameters are computed through the use of the DMD algorithm [41]. As depicted in Fig. 1, the ML-based prediction algorithm outputs a triplet of forecasted values $(\widehat{f}_i, \widehat{\alpha}_i, \widehat{a}_i)$ for each electromechanical mode of interest (i.e. from mode #1 up to N_{mod}) and for the required number of prediction steps accounted for the forecast problem at hand.

The details about the pre-processing stage of the input data and the core ML-based prediction strategy are given and fully described in the following Section 3.

In this subsection, we want to stress instead which are the constraints set by the depicted application scenario and which are the operational hypotheses on which the multi-variate prediction strategy illustrated in Section 3 has been built.

As already mentioned the first relevant constraint of the application scenario is that only instantaneous frequency measurements can be gathered from PMU devices in order to set up the multi-variate prediction problem. A second constraint is that the prediction strategy has to be deployed in a centralized hub where the overall raw input datastream from PMU devices is directed through a central concentration hub (as depicted in Fig. 1). Some simplification assumptions are set on the data collection technique, in particular, the time delays due to the communication of sampled instantaneous frequency from the different PMU locations can be neglected, and, in addition to this, the data collection process is supposed to be such that no time overhead is introduced to pack the final datastream (this is reasonably true due to the high efficiency of modern monitoring infrastructures). All these hypotheses can be resumed stating that the overall strategy can evolve in quasi-real-time. Under these assumptions, the timings required to output the predicted values are set by the specific data segmentation operated on the input datastream, which is better described in the following Section 3.

It is worthy noticing that critical ringing events in the grid occur quite rarely, so, the measurement data collected in the past and used for the development of this work represents singleton cases. This is an application constraint that does not enable the assessment of the generalization capabilities of the method for every possible ringing event. This assessment requires a separate study.

3. Proposed enhanced multi-variate prediction approach

In analogy with other previously proposed approaches [39,40], here the basic idea is to make use of the past values assumed by the modal parameters in order to formulate a multi-variate prediction problem with the aim of forecasting the values assumed by the modal parameters in the near future, meaning tens of seconds in this specific application scenario. The prediction problem is formulated as a multi-variate

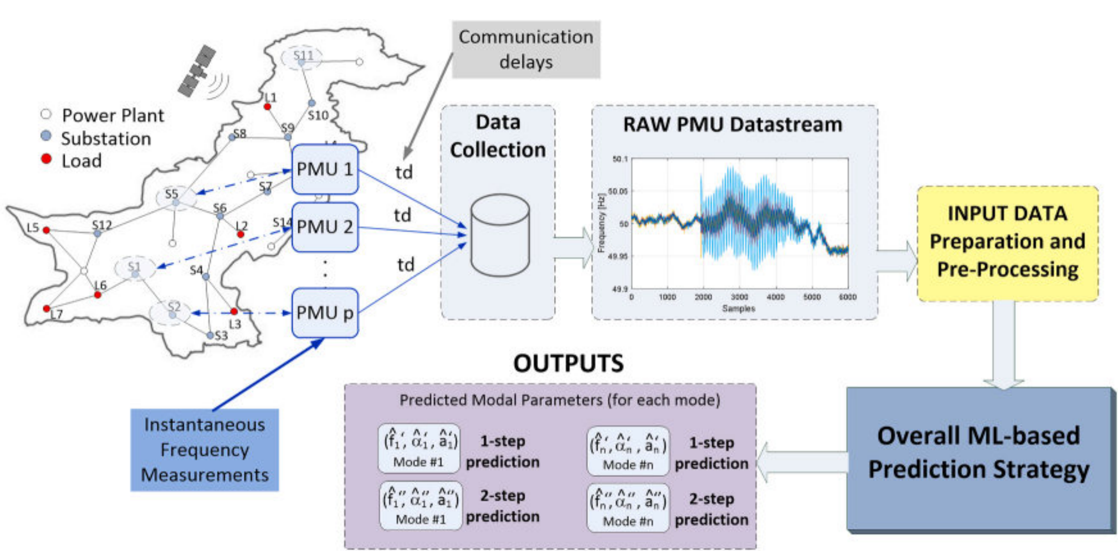

Fig. 1. Top- level description of the application domain.

time-series prediction problem where the quantities involved in the time-series are given by the three variables composing the triplet of modal parameters (f_i, α_i, a_i) for each *i-th* electromechanical mode. Differently from previous approaches [40], where the overall prediction problem is treated as the assembly of three separate univariate problems, meaning that each modal parameter is predicted by setting up an independent prediction problem, in this case, the prediction of each modal variable is made based on the past values of all the three modal variables. It has to be remarked that the proposed approach, fully described in the following of this section, is aimed only to provide a helpful tool to the control room operators in identifying potentially critical situations or LFO ringing events a short time in advance, from this, it follows the necessity to make only short-term modal parameters forecasts. The approach described in the following exploits the use of RNN structures and, in particular, Long-Short-Term Memory (LSTM) unit structures in order to show also the impact of using multi-variate predictions with respect to simpler uni-variate ones,already analyzed in other works [39,40]. The use of LSTM units as the base of the approach is motivated by the fact that this is one the simplest algorithms suitable to make prediction over time-series problems, according to what has been discussed also in previous works [39,40]. In the following of this section the overall LSTM-based strategy is described together with its integration with a proper Ensemble-based method.

3.1. Overall prediction approach architecture

It is quite straightforward that the starting point of the monitoring and predictive monitoring strategy needs to be the same as other previous approaches, that is, in our case, the collection of the frequency measurements coming from some PMU devices and their proper elaboration through suitable estimation or prediction algorithms.

A more detailed view of the overall approach is reported in the following Fig. 2. The initial pre-processing step consists of a data detrending operation and a band-pass filtering action, suitable to isolate the frequency content of interest, excluding all the spurious measurement signals. The filtering action is realized through an Hilbert filter which is properly configured in order to have transition bands centered on 0.1 and 0.5 Hz.

Preliminarily to the pre-processing using the mentioned detrending and filtering action, the incoming input stream gathered from the PMUs is unpacked by using a sliding window framing mechanism. In this context the basic element used in the proposed approach is called "data window" and represents the minimum collection of data on which the modal parameters can be estimated using the DMD algorithm. In this case the length of each data window is defined by the number of samples considered, here in the following indicated by the parameter *Lw*. Then, the pre-processed input data stream enters the ML-based prediction strategy and, at this stage, for each window of data the associated modal parameters are computed using the DMD algorithm.

For the sake of clarity, in order to avoid confusion, the triplet of modal parameters estimated by the DMD is indicated in the following as ($\tilde{f}_i$, $\tilde{\alpha}_i$, $\tilde{a}_i$). The number of electromechanical modes of interest can be selected by using the already mentioned parameter N_{mod}, to be passed to the DMD computation routine. It is worthy to point out that, in order to build the multi-variate time-series prediction problem, it is necessary to set up a proper buffering strategy. The role of the buffering strategy is twofold in the proposed approach. On one side, the buffering strategy is necessary to complete the time-series information required to build the prediction problem; on the other side, it is fundamental also to handle the continuous input datastream coming from the PMU devices. In effect, the application at hand presents the typical elements of an Incremental Learning (IL) problem since the input data feeds continuously the prediction method, and it should be treated formally as a nonstationary regression problem. For what concerns with the first scope of the buffering strategy it has to be underlined that it is useful to arrange the incoming datastream, that is a sequence of data windows, into bigger slices of data, called hereafter "data chunks". The length of this data segment is called hereafter "chunk size" and is expressed in terms of the number of considered data windows. The chunk size is indicated by the parameter L_{ch}. For what concerns with the second scope of the buffering strategy, it has to be clarified that the policy used to manage the buffering strategy is applied in conjunction with the implementation of the Ensemble method proposed by this approach in order to realize an approximation of an IL strategy, as will be better detailed in the

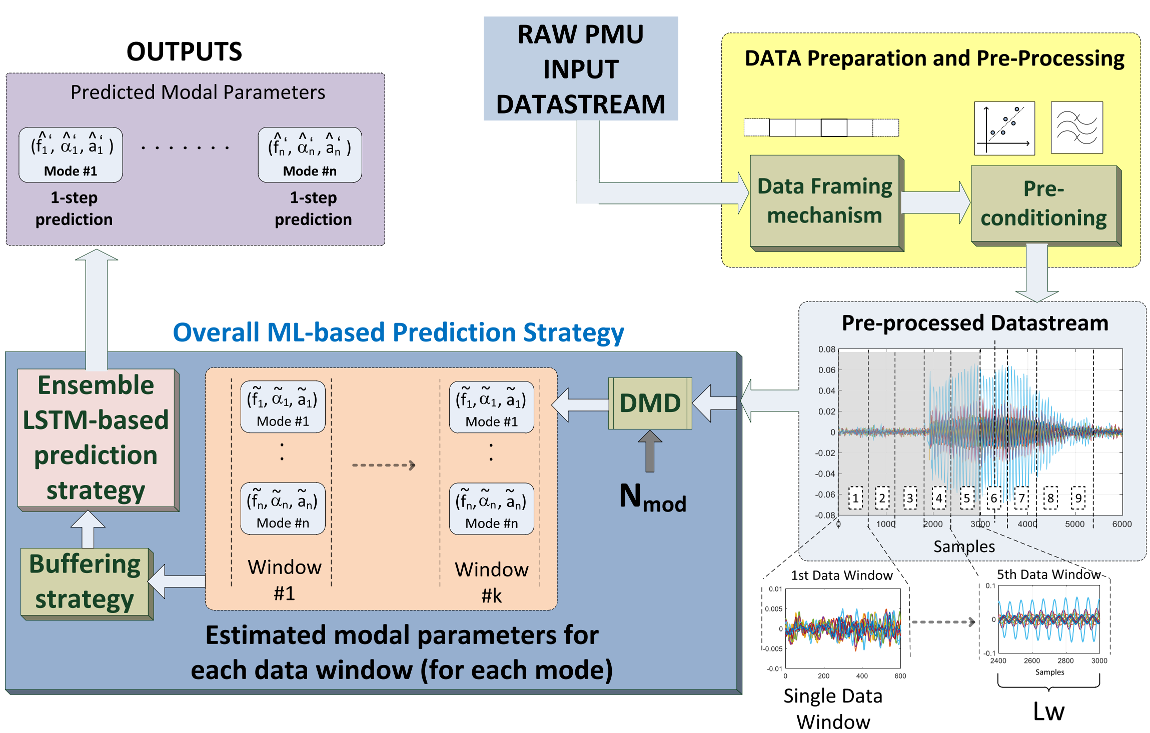

Fig. 2. Overall description of the proposed prediction approach.

following subsections. Without loss of generality, for the sake of simplicity, in the specific case of this work we restrict the prediction time horizon to a single data window in the future, as depicted in Fig. 2, so the output forecasts will be only the 1-step predictions of the modal parameters. The prediction horizon, in terms of data windows, is indicated as N_w (N_w=1). Basically, when fresh data related to a new data window is gathered from the PMUs, meaning after the application of the preliminary pre-processing stage of Fig. 2, it is added to the actual chunk of data and its related modal parameters are extracted through the DMD algorithm. The triplets ($\widetilde{\mathrm{f}}_\mathrm{i}$, $\widetilde{\alpha}_\mathrm{i}$, $\widetilde{\mathrm{a}}_\mathrm{i}$) related to the *i-th* mode and computed for each new incoming data window are placed into a proper buffering structure, observing the aforementioned buffering policy. The management of such buffer follows a typical first-in-first-out (FIFO) buffering policy; the length of the resulting queue is practically set equal to the length of the chunk in order to have a circular completion of the temporal sequence of data. Once the necessary information has been got, meaning that the buffer structure carrying the actual chunk of data is full, the triplets linked to a specific LFO mode and related to the past L_{ch} data windows are used to make the prediction of the triplet of values for the future data window, as shown also in the following Fig. 3. In particular, at a certain time, all the data contained in the actual data chunk is used to perform the training of the subsequent LSTM-based prediction stage, then, in a second step, this last one generates the set of triplets related to the forecast of the modal parameters for the next data window and for each mode. As depicted also in Fig. 3, the overall strategy is thought to output the forecast each time a new data window is available from PMUs and the related modal parameters triplets enter the data chunk. In the case of Fig. 3 the past five data windows (and their associated triplets) are used to predict the set of triplets associated to the future data window (window #6 in Fig. 3). It clearly follows from here that the computation has to be done iteratively and that the time required to output the predicted values has to be much smaller than the time duration of the prediction horizon (i.e. a single data window). Actually, a preliminary rough assessment of the computation time suggests that the elapsed time should be lower than half the duration of a time window. This is motivated by the fact that in the practical case (worst case) for which the data stream cannot be stored in parallel to the computation of the forecasts and if the completion of the data frame with additional incoming data is not feasible, the residual data of the frame should be sufficient to correctly estimate the modal parameters. In order to be able to catch the modal parameters of a sufficient number of modes of LFO oscillations, also the ones having typical frequencies near and below 0.1 Hz for instance, it is required to feed the DMD procedure with at least 10 s of data. This motivates, in this application case, the use of segments of data at least containing half a data window.Differently from previous approaches [39,40], in this case we have a logical buffer structure for each mode and the values of all the three modal quantities are used to make the prediction of the triplet for the next window.The details of how the values of the past triplets of a specific mode are used to make the forecast of the future window are detailed in the following subsection.

3.2. Multi-variate LSTM-based prediction stage

Given a specific mode, the past values of the triplets ($\widetilde{\mathrm{f}}_\mathrm{i}$, $\widetilde{\alpha}_\mathrm{i}$, $\widetilde{\mathrm{a}}_\mathrm{i}$) are provided to the aforementioned LSTM-based prediction strategy sketched in the following Fig. 4. As can be seen, from a top-level point of view, the strategy is founded on the use of three predictors, namely Prediction Units (PU), formulating three independent multi-variate prediction problems, one for each variable to forecast: frequency, damping, amplitude.

The core business of each PU is based on using a custom prediction solution combining a set of weak LSTM predictors by using an ensemble method, as better detailed in the following subsection. The choice of considering three independent PUs for outputting the forecast triplets is motivated by the fact that different settings of the strategies associated with each PU can be accounted for (higher flexibility of the method). LSTM structures have been selected due to their suitability to catch long-

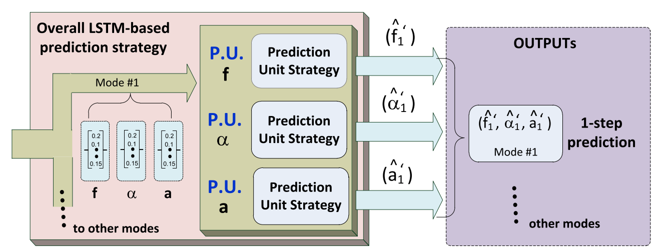

Fig. 4. General structure of the LSTM-based prediction strategy.

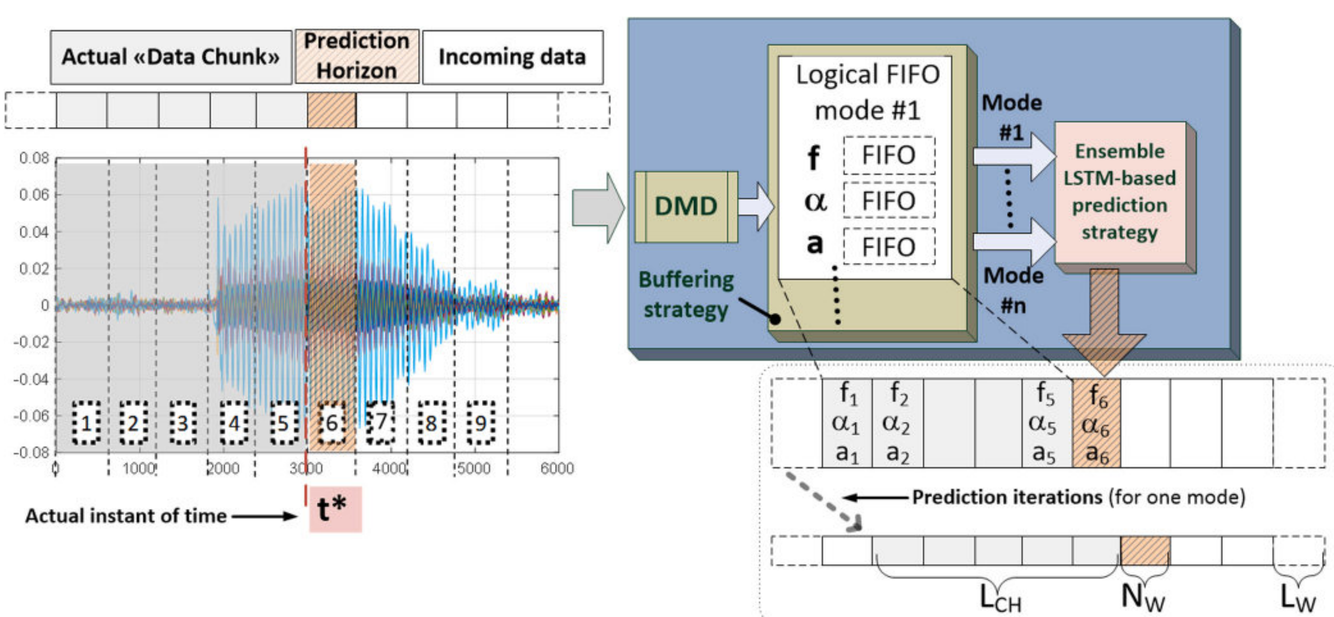

Fig. 3. High-level depiction of the multi-variate prediction problem with an example prediction iteration.

term dependencies. For each PU a data standardization procedure is considered, based on a proper regularizing transformation of the input data accounting for a rescaling of the input by its variance and mean value, as already done in [40].

3.3. Prediction unit inner structure and multi-variate prediction ensemble

The basic structure proposed here for a given PU involves a solution combining a set of weak predictors through the use of an Ensemble method having a custom average function as its aggregation function. Although ensemble methods have been used to face challenging non-stationary regression problems [34,37] and IL problems [28,36,46,47], in this paper a custom solution has been investigated to handle specific learning tasks. It has to be pointed out that the overall strategy is not intended here to handle complex Concept Drift problems [47] but only to provide some beneficial effects, in particular: reduce the over-fitting issues and improve the statistical properties of the prediction error. In addition to this objective, the LSTM-based ensemble structure of the prediction unit has been selected also to make some comparisons with previously shown results [40] in order to explore the improvements that can be obtained by using multi-variate predictions. In the proposed multi-variate ensemble approach, each weak predictor is constituted by an LSTM model with a predefined number of layers, indicated by the parameter N_L, whereas the number of members in the ensemble is parameterized by the variable N_E. The overall view of the Prediction Unit structure is depicted in Fig. 5.

The proposed approach exploits the Parameter Diversity paradigm of ensemble methods [28,46], since the involved LSTM-based weak learners are different in terms of the number of attributed hidden units, indicated by the parameter N_{HU}. The number of hidden units for each LSTM model composing the ensemble N_{HU} is set with the rule reported in (3).

$$\begin{cases} N_{HU}^{(i)} = \left(\left\lceil \dfrac{(N_{HU}^{M} - N_{HU}^{m})}{N_E} \right\rceil + 1 \right) \cdot (i-1) + N_{HU}^{m} & i = 1, \ldots, N_E - 1 \\ N_{HU}^{(i)} = N_{HU}^{M} & i = N_E \end{cases} \quad (3)$$

where i is the index of the ensemble member, ranging from 1 to N_E, while N_{HU}^{M} and N_{HU}^{m}indicate the maximum and minimum number of hidden units. This is done in order to uniformly distribute the number of hidden units among the ensemble members. In all the cases analyzed in Section 4 and Section 5, the minimum number of units is always set equal to 5, and the parameter N_{HU} is referenced in place of N_{HU}^{M}. In addition to this, the strategy considers a final predicted value that is computed by calculating a proper dynamic weighted average (DWA) of the forecasts provided by the ensemble members, which constitutes here the proposed aggregation function. The DWA calculation considers a weighted average of the predicted values by each ensemble member but includes a "premium" mechanism based on the prediction error at the preceding forecast iteration. Basically, the weights are initially set to the standard value $W_{init} = 1/N_E$, then, for each prediction iteration i, the weights are updated in such a manner that the weight of the model with the best fit (lower prediction error) at iteration i-1 is increased with a predefined amount, ΔW, while the other weights are decreased by a predefined amount equal to $\Delta W/(N_E\text{-}1)$. Other fine-grained details about the DWA algorithm are described in the work [40].

4. Application of the proposed strategy to case study scenarios

To have a preliminary evaluation of the prediction performance obtained with the proposed approach, regardless of the assessment of the generalization capabilities of the method, some reference validation test cases have been analyzed in this section. In particular, the method has been tested by using two different kinds of input datasets. A first kind considers two short LFO ringing events, and a second kind considers a sample dataset reproducing 24 hours of normal grid operation, also indicated in the following as "ambient" conditions. The data related to both the two aforementioned kinds of datasets, namely, the ringing datasets and the "ambient" datasets, comes from real monitoring of the grid operated by the TSO in the past through PMU devices. The complete description of the test scenarios considered and the input data is given in the following subsection. For what concerns instead with the considered

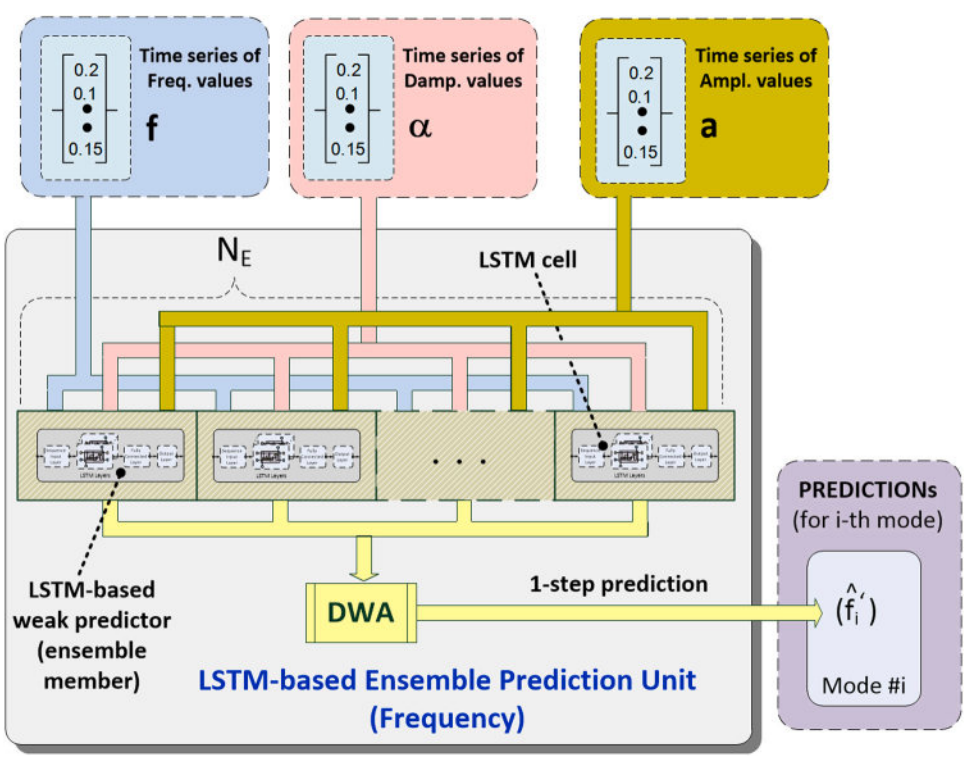

Fig. 5. Details of the inner structure of a prediction unit accounting for a multi-variate strategy.

evaluation metrics, all the considered scenarios have been assesses through the use of common key performance indicators (KPI), which are the error statistics (mean and standard deviation) plus the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE) and the CPU time for each prediction. For the sake of more clarity MAE and RMSE definitions are reported in (4),

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|\hat{y}_i - y_i| RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_i - y_i)^2} \quad (4)$$

where y_i barely stands for the *i-th* true (observed) value of the quantity under consideration and $\hat{y}_i$ indicates instead the predicted value for the same *i-th* sample, *n* represents the number of samples. For the analysis reported in the following, for both this section and the next section of the paper, the proposed prediction strategy has been implemented, as previously described in the text body of Section 3, by using the Python-based Anaconda/SPyder environment, exploiting the use of Keras/Tensorflow, NumPy, SciKitLearn libraries [48]. All the accounted test scenarios described in the following have been simulated on a middle-sized computing platform having 24 GB of RAM and 24 working threads with a CPU operating at 3 GHz (max clock frequency). The computing hardware accounts also for a GPU card but it was intentionally neglected in order to exclude the use of possible hardware acceleration (at least for the current study). The considered training algorithm is a standard Stochastic Gradient Descent-based (SGD) method, in particular the ADAM (Adaptive Moment estimation) one.

4.1. Test scenarios and input data

The two aforementioned kinds of datasets used for this study are briefly described hereafter and depicted in the following Fig. 6. Input datasets R1 and R2 describes two short ringing events with time duration of 20 and 10 minutes respectively. Dataset R1 accounts for the instantaneous frequency measurements coming from 30 PMU locations whereas dataset R2 accounts for 22 PMU measurement signals. In the case of dataset R1 the dominant electromechanical mode is mode #1, characterized by a frequency of nearly 0.25 Hz. The "Ambient" dataset A1 is characterized by about 48 hours of PMU measurement recordings with the power grid operating in its rated working conditions, that is without relevant LFO phenomena. The two datasets R1 and R2 have been measured through PMU devices located both in Italy and in other EU countries. All the datasets accounts for the same sampling time, in this case equal to 100 ms, but the duration of the data windows involved in the application of the proposed method are different for the two kinds of datasets. In particular for datasets R1 and R2 the length of the data windows is equal to L_w= 300 samples, corresponding to 30 seconds, whereas for dataset A1, the data windows have a length of 20 seconds, hence, L_w= 200.

It is worthy to note that the considered datasets do not take care of the existence of possible gappy data from the PMUs, that is a condition potentially occurring in real-world scenarios. The reason is that for the presented strategy the loss of some instantaneous frequency measurement samples from a PMU do not affect the entire prediction process, furthermore, in the case of highly gappy data, there are techniques alternative to DMD or extending the standard DMD algorithm and suitable to obtain the modal parameters estimates. By considering what already stated, dataset R1 is composed by 39 data windows and dataset R2 by 19 windows (considering also the discard of the first data window ascribed to the elimination of the transient introduced by the Hilbert filter used for the input data pre-processing, as described in Section 3). Dataset A1 is composed by 8640 data windows, since it considers directly the 8640 samples of the modal parameters over 48 hours. Without loss of generality dataset A1 can be restricted to 4320 windows.

The analysis described in the following takes into account different configurations of the prediction strategy, summarized in Table I, and several test scenarios, properly summarized in Table II. The accounted test scenarios are mainly aimed to either provide a preliminary overview of the prediction performance or to assess the impact of the various internal parameters of the strategy. There are basically five test scenarios, indicated as TS#1 up to TS#5, each one of them considers three subcases. The first three test cases TS#1-TS#3 show the prediction performance of the multi-variate prediction strategy over the dataset R1, for the three modal quantities respectively, exploring the impact of the number of hidden units N_{HU}, subcase *a*), the impact of the maximum number of training epochs (M_{ep}), subcase *b*), and the number of ensemble members (N_E), subcase *c*). The test scenario TS#4 evaluates the prediction performance over the dataset R2 by using two reference configurations, whereas TS#5 evaluates the performance for the same configurations over the ambient dataset A1. In the cases of test scenarios TS#4 and TS#5 the subcases a), b), c) are used to differentiate the analyzed modal quantity. The results of the assessment of all the test cases are resumed in the following subSection 4.2.

4.2. Results

The results coming from the analysis of the first test scenario TS#1, for all its subcases, are summarized in the following Fig. 7. The analysis

Table I
Considered architectural configurations and parameters of the prediction strategy.

Config. ID	Architectural Parameters			Training parameters	
	N_E	N_{HU}	N_{HL}	Learn Rate (LR)	Chunk Length (Lch)
C1	1	10	2	0.2	2
C2	1	25	2	0.2	2
C3	1	50	2	0.2	2
C4	1	100	2	0.2	2
C5	5	25	2	0.2	2
C6	10	25	2	0.2	2
C7	5	25	2	0.2	50

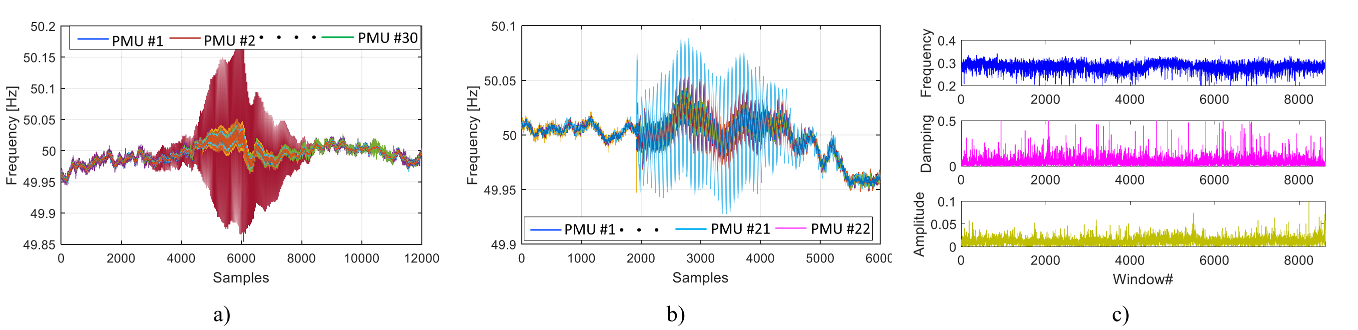

Fig. 6. Considered raw input data of the available datasets: instantaneous frequency measurements related to ringing events "R1", a), and "R2", b), and reference modal parameters of dataset "A1", c).

Table II
Summary of evaluated test scenarios.

Test Scenario	Parameter	Dataset	Considered Configurations	Max Epochs (Mep)
TS#1-a	Frequency	R1	C1, C2, C3, C4	100
TS#1-b	Frequency	R1	C4	50, 100, 200
TS#1-c	Frequency	R1	C2, C5, C6	100
TS#2-a	Damping	R1	C2, C3, C4	100
TS#2-b	Damping	R1	C4	50, 100, 200
TS#2-c	Damping	R1	C2, C5, C6	100
TS#3-a	Amplitude	R1	C1, C2, C3, C4	100
TS#3-b	Amplitude	R1	C4	50, 100, 200
TS#3-c	Amplitude	R1	C2, C5, C6	100
TS#4-a	Frequency	R2	C5	100
TS#4-b	Damping	R2	C5	100
TS#4-c	Amplitude	R2	C5	100
TS#5-a	Frequency	A1	C7	100
TS#5-b	Damping	A1	C7	100
TS#5-c	Amplitude	A1	C7	100

of the impact of the number of hidden units is described by the results of Fig. 7a) and b), for the case when the frequency modal variable is taken into account. As can be seen from here, there is a certain value of the parameter N_{HU}, in this case around 10, for which prediction performance is not heavily altered by the chosen number of hidden units. In effect, the analysis of the selected KPIs for the prediction error shows that the RMSE changes mainly between 10 and 25 hidden units, but then it is likely stable also after a large increase of N_{HU}. From the results reported in Fig. 7c) and d), it can be seen that the increase of the number of maximum epochs used for the training phase affects the level of obtained prediction error, in terms of RMSE. The use of a too much high value of N_{HU} can create overfitting issues, so, the choiche N_{HU}= 25 can be reasonable to avoid an unbearable increase of the computational burden of the overall strategy.

From the analysis of the results reported in Fig. 7e) and f), it follows that the use of a prediction method considering a degenerated strategy with only one ensemble member (N_E=1) has the worst prediction performance, in terms of both RMSE and MAE KPIs.

In addition to this also the mean value of the prediction error is lowered by increasing the number of ensemble members. Despite of this, the parameter N_e cannot be increased arbitrarily, since it affects directly the computational burden of the method and values greater than N_E= *10* seem to be not compliant with the application constraints. In the case of N_E= *10* the prediction time goes far beyond the maximum bound of half a data window.

With regard to the test scenarios TS#2 and TS#3, the prediction performance related to the other two modal quantities has been assessed in the following Fig. 8. It is reported in the plots of Fig. 8 the behavior of damping prediction over time (Fig. 8a) and its related prediction error (Fig. 8b), with particular reference to the analysis for a different number of hidden units, that is sub-case a. A similar analysis but referred to the amplitude modal variable is reported in Fig. 8c and d. It is worthy to note that in the case of damping prediction, Fig. 8a, the possibility of determining the approaching toward extremely low values of the damping parameter one time window in advance can be useful to trigger early-warning signals. Of course, it is expected that the longer the prediction horizon can be the more useful the strategy will be for enhancing the resilience of the grid against critical events. The overall results related to the test scenarios TS#2 and TS#3, for the considered KPIs, are detailed in the general plots reported in Fig. 8e and f respectively for the damping variable and for the amplitude variable. As can be seen from the plots of Fig. 8e and f, the number of hidden units has not a relevant impact on the prediction performance after a certain threshold (N_{HU}=25). Despite of this, the parameter M_{ep} seems to have a more weighty impact in the range 50–100.

According to what observed also for the frequency modal parameter, the results of Fig. 8e-f shows also that the number of ensemble members affects the accuracy of the prediction in terms of RMS, MAE, and also mean value of the error, in this case the effect is more evident on the damping parameter. For the test scenario TS#4 the results obtained from the application of the multi-variate prediction strategy to the second ringing dataset R2 are resumed in Fig. 9. In the specific case of test scenario TS#4 the analysis details directly the prediction performance over the various data windows composing the dataset (19 windows), furthermore, in order to stress the generality of the approach, in this case the method has been applied on the electromechanical mode #2, by properly selecting the mode of interest as reported in Fig. 2. As can be seen from the plots of Fig. 9a), b), c), the prediction performance of the frequency, damping and amplitude variables, respectively, of the second mode are fairly satisfactory even though the dataset is quite limited in terms of available data windows for the training phase. The results have been obtained by using a representative strategy configuration (i.e. C5) with good forecasting performances. The last evaluated test scenario is the one aimed to study the prediction performance over long-run datasets, that is dataset A1. The results related to test scenario TS#5 are summarized in the following Fig. 10. In this case only the behavior of frequency prediction has been reported in Fig. 10 with respect to the data window evolution.

Either the direct forecasts provided as output by the method or the computed prediction error and its related KPIs values are reported in Fig. 10a and b respectively. The statistics related to all the three sub-cases of TS#5 have been summarized in the plot of Fig. 10c. Since the dataset A1 is quite rich in terms of data points a proper variation of the presented strategies has been used to assess the prediction performance. In this case a specific configuration has been used to get the prediction results, indicated as C7, for which the length of the data chunk has been enlarged to 50 windows. As can be seen from the plots proposed in Fig. 10, the frequency prediction performance is such that the prediction error is limited and with a very small mean value, in this case the RMSE value and the error standard deviation are similar.

By looking at the overall results summarized in Fig. 10c, for all the three test sub-cases, also in the case of damping and amplitude variables the error mean values are very small and the standard deviation and RMSE values are enough low. The damping variable has a bigger value of MAE with respect to the other two variables. In terms of computational burden, some global results are outlined in Fig. 11, where the times required to generate a single forecast are reported for the different evaluated scenarios. The required times are always lower than the duration of half a data window.

As can be seen from Fig. 11a, the addition of hidden units does not have a great impact, whereas, the increase of maximum number of epochs scales almost linearly (Fig. 11b). The most impactful parameters are N_E and L_{ch} as clearly shown in Fig. 11 c-d.

5. Comparison of proposed technique against uni-variate technique

In this section the prediction performance of the proposed multi-variate technique has been compared with respect to a prediction technique similar in structure but exploiting only uni-variate time-series prediction capabilities. The comparison is aimed to show the enhancements obtained by using a multi-variate prediction technique, with a reasonable accuracy, in terms of the improvement of prediction error statistics.

5.1. Evaluated multi-variate and uni-variate strategies and parameters

In order to setup a comparison that can be retained of interest, a proper configuration of the presented multi-variate strategy has been defined, based on the architectural and training parameters of an already existing uni-variate technique. In particular, taking the solution

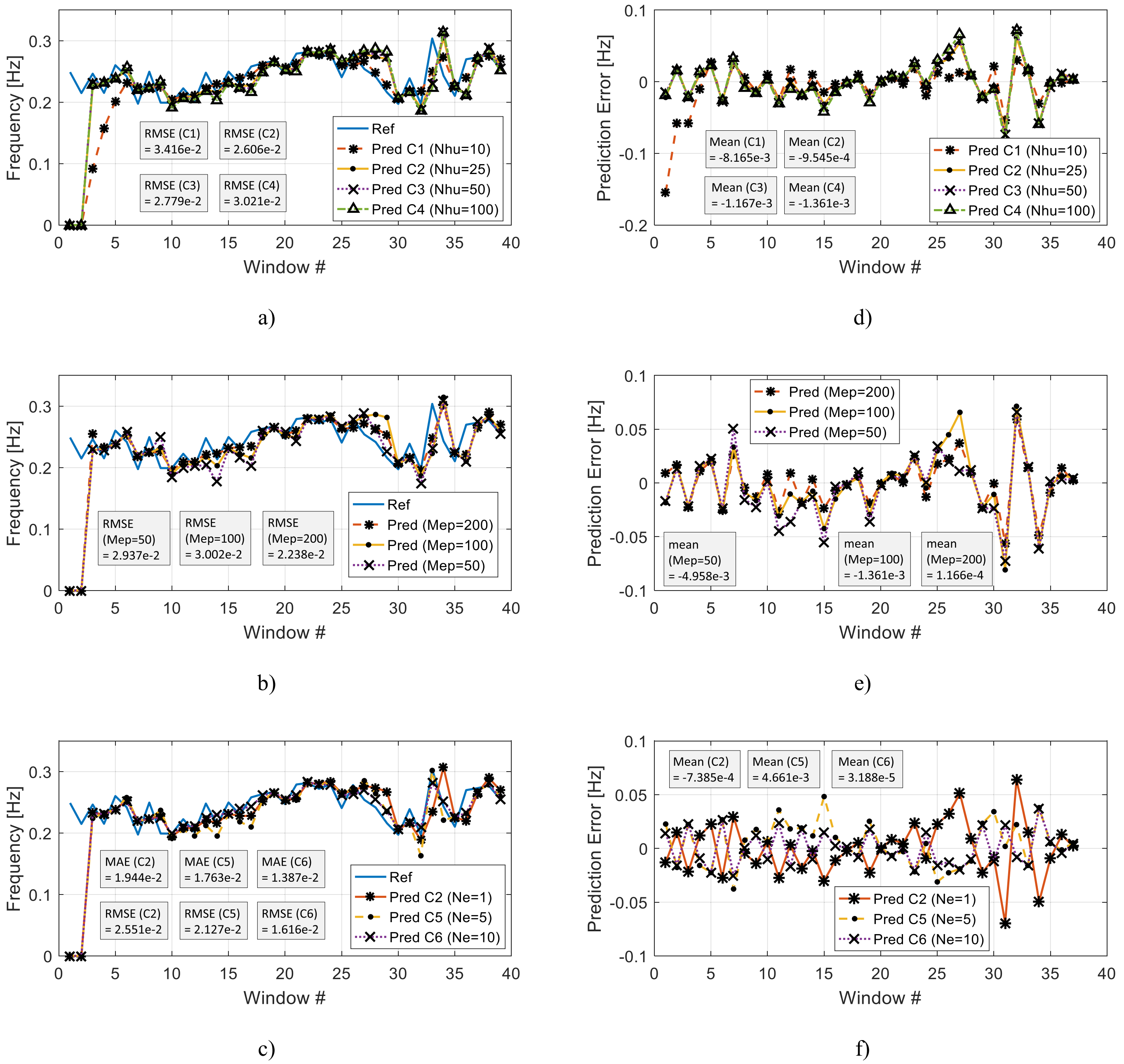

Fig. 7. Results related to test scenario TS#1: a), b), c), frequency prediction performance for the evaluated test cases, respectively for test cases TS#1-a, TS#1-b, TS#1-c. Prediction errors related to test cases TS#1-a, TS#1-b, TS#1-c, respectively: d), e), f).

called "T1" and already presented in [40] as a reference for comparison purposes, the configuration option indicated as "C8" has been set up. The parameters used for configuration C8 are functionally similar to the ones of configuration T1 in [40], they only differ in the number of hidden units, since it has been shown in Section 4 that the parameter N_{HL} is not very impactful in the range 20–30. For the sake of clarity the overall parameters of both configuration C8 and T1 are recalled in the following Table III. The two prediction strategies have been assessed on the two datasets R1 and A1 only, for the sake of conciseness, in the following subsection 5.2.

5.2. Comparison of results

The first comparison has been done on the dataset R1. The three modal parameters of the first electromechanical mode have been predicted 1-step forward with both the two strategies, uni-variate (T1) and multivariate (C8) and the obtained results are resumed, similarly to what done in Section 4 in the following Fig. 12, showing either the time evolution of predictions or the related statistical information about the prediction error.

From what reported in Fig. 12 a), b), c), it can be seen how the prediction performances of the two techniques are quite similar, especially for what concerns frequency and damping prediction. The overall prediction performance offered by the multi-variate technique seems to be slightly better than the one of uni-variate technique. In the case of amplitude prediction the multi-variate technique appears to be globally slightly better than the uni-variate one, in terms of the overall RMSE value, even though from the bare analysis of the prediction temporal behaviour the multi-variate one tends to underestimate the amplitude of the electromechanical mode under consideration.

A more global analysis of the prediction errors statistics, reported in the following Fig. 13 shows that the two techniques are practically

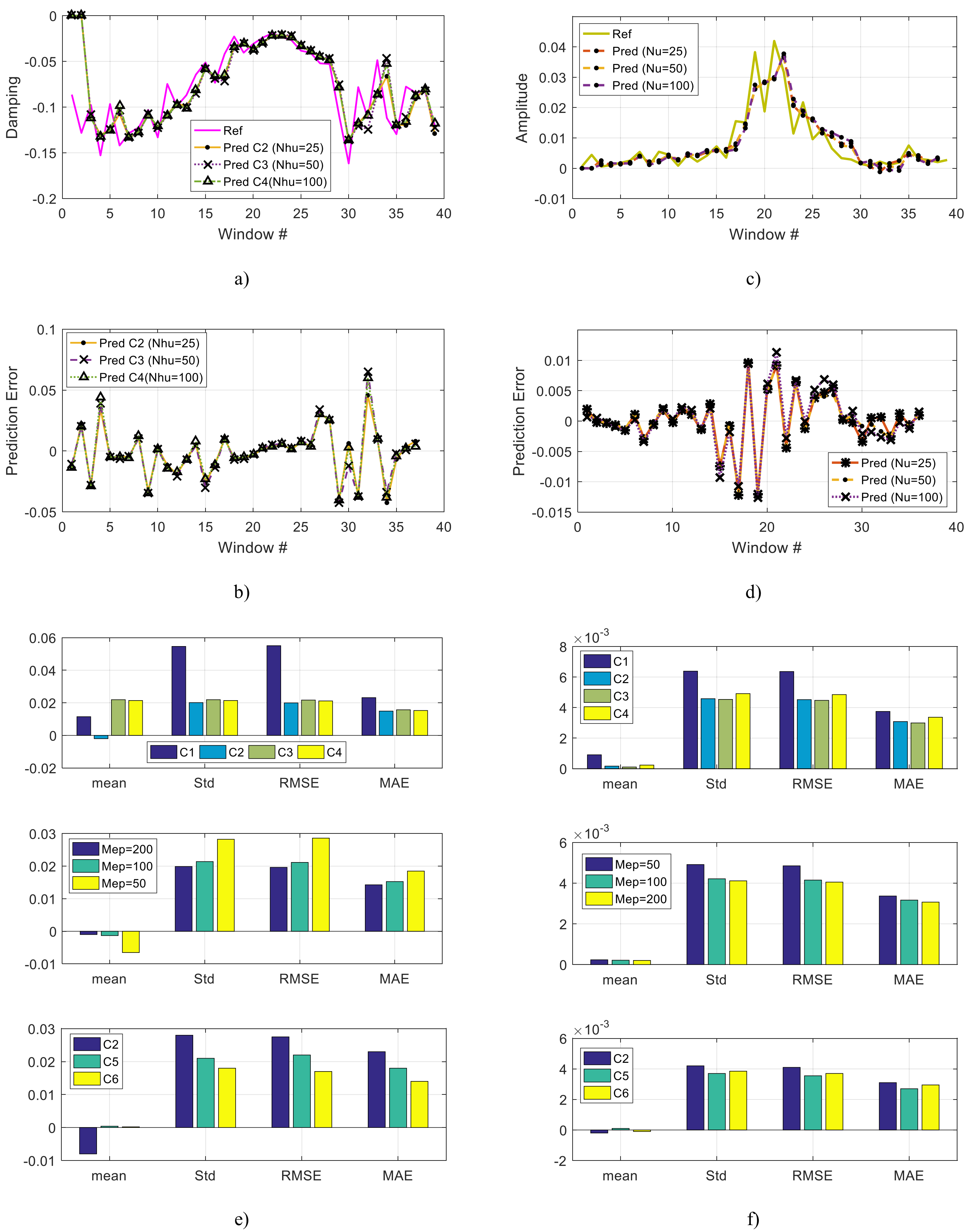

Fig. 8. Results related to the test scenarios TS#2 and TS#3 for all the test subcases. Damping prediction performance for TS#2-a, a), and related prediction error, b). Amplitude prediction performance c), and related error, d). Summary analysis of the KPIs for TS#2, e), and for TS#3, f).

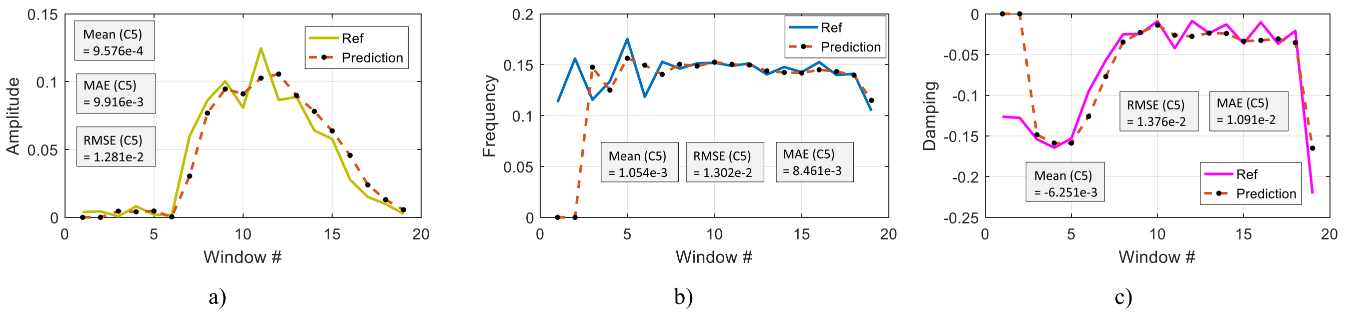

Fig. 9. Results for test scenario TS#4 and related subcases. Prediction performance of: a) Amplitude, b) Frequency, c) Damping.

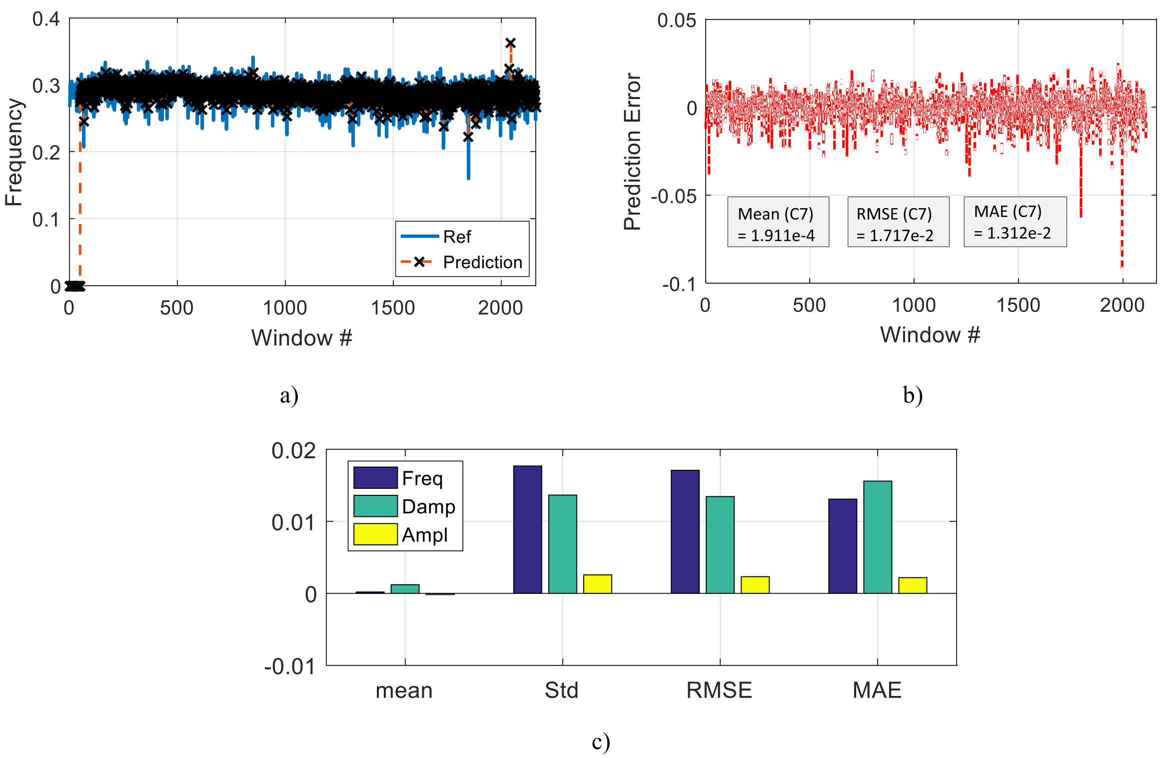

Fig. 10. Results for test scenario TS#5 and related subcases. Prediction performance for frequency, a), and related prediction error, b), plots limited to the first half of the considered data windows. Overall KPIs analysis for the test scenario TS#5, c).

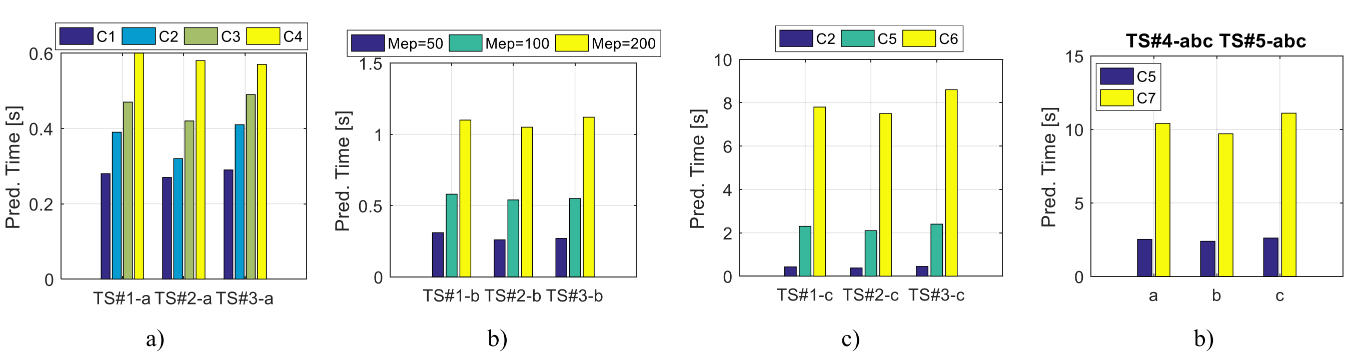

Fig. 11. Analysis of the elapsed times required to compute a single forecast for the different evaluated test scenarios and parameters. Evaluation of the impact of N_{HU}, a), and M_{ep} parameters, b). and of N_{E}, c), and of the chunk length for TS#4 and TS#5, d).

similar, although, by looking at the RMSE values only, the general indication is that there is a slightly better performance by using the multi-variate solution C8. The analysis of the computational burden for the two techniques, reported in the same Fig. 12, shows also that the multi-variate one is slightly more computationally intensive than the uni-variate one.

The application of the two techniques to the ambient dataset A1 brings to have the prediction performance reported in the following

Table III
Configurations and parameters considered for comparison.

Configuration ID	Parameter	Architectural Parameters			Training Parameters		
		N_E	N_{HU}	N_L	*LR*	*Mep*	*Lch*
C8	Frequency	5	25	2	0.2	300	2
	Damping	5	25	2	0.2	200	2
	Amplitude	5	25	2	0.2	150	2
T1 (*)	Frequency	5	24	2	0.2	300	2
	Damping	5	30	2	0.2	200	2
	Amplitude	5	20	2	0.2	150	2

(*): same configuration parameters of reference work [40].

Fig. 14. This last figure outlines the prediction behaviour and the related error for the three modal quantities by using histogram plots of the error KPIs, since the dataset is rich of data points and the specific punctual time behavior of these quantities is not of great interest in this case. The only prediction behavior reported in Fig. 14a is the one related to the damping. It is reported only to provide an example case suitable to show that the prediction error is limited also in the case of dataset A1 but its punctual value can be quite high.

Furthermore, in Fig. 14b, it is depicted a direct comparison between the uni-variate and the multi-variate prediction error. As can be seen from Fig. 14 b, the multi-variate prediction error is punctually greater than the uni-variate one, however, its statistical properties are very close to the uni-variate case, as reported in Fig. 14c, and in most of the cases its RMSE and MAE values are slightly lower than the one of the uni-

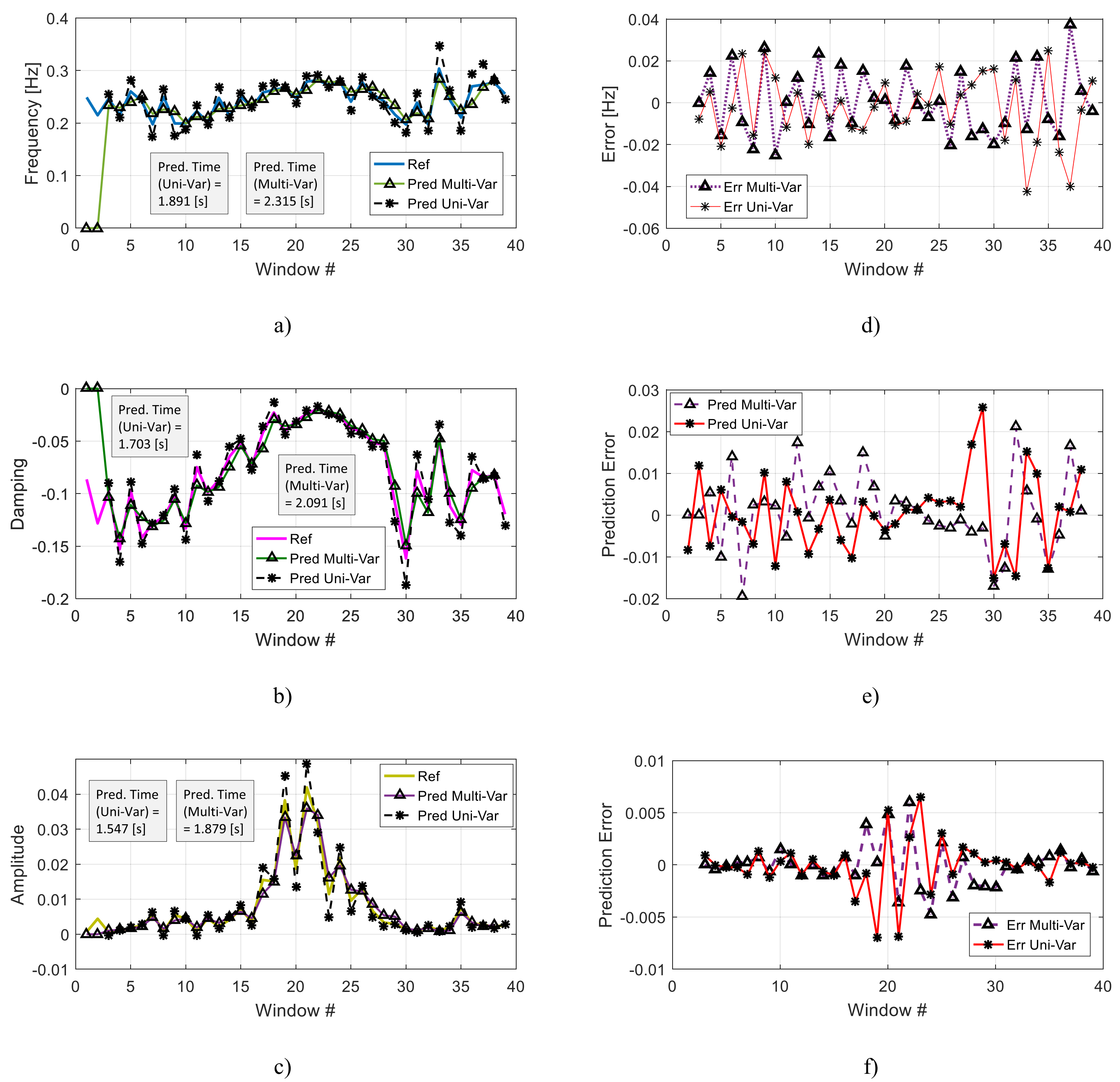

Fig. 12. Comparison of the prediction performance for uni-variate and multi-variate strategies in the case of dataset R1. Frequency, a), damping, b), and amplitude, c), prediction behavior. Comparison of computed prediction errors with both the uni-variate and multi-variate strategies for frequency, d), damping, e), and amplitude, f), modal parameters.

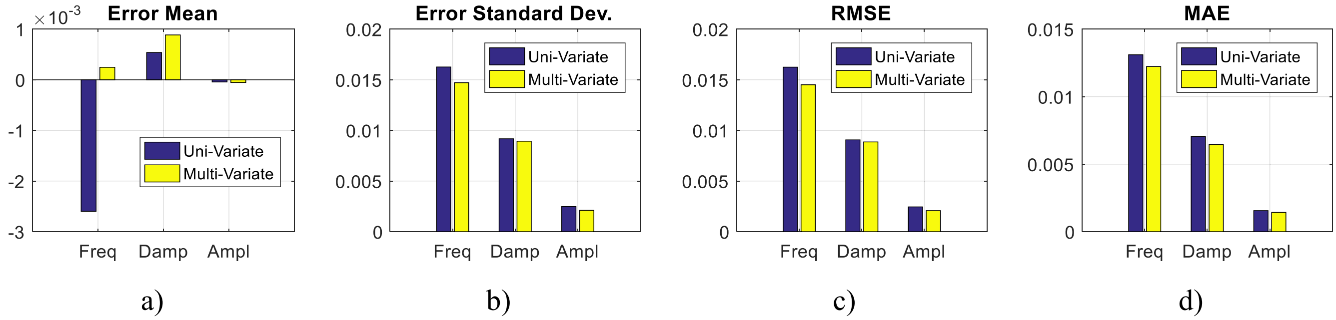

Fig. 13. Summary results of KPIs computation for the uni-variate and multi-variate technique prediction results for dataset R1.

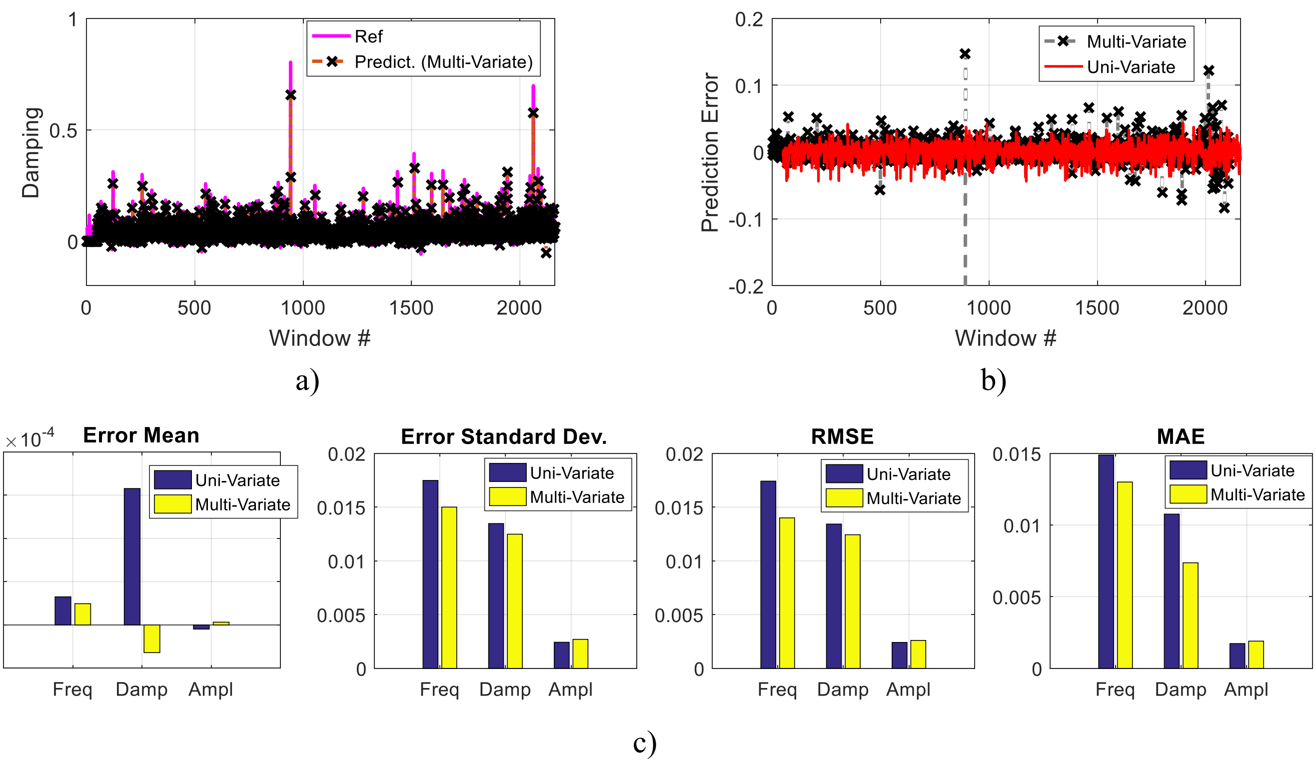

Fig. 14. Results of the comparison of the uni-variate and multi-variate prediction strategies over ambient dataset A1. Prediction performance for damping, a), with multi-variate technique and comparison of prediction errors for both the strategies, b), plots limited to the first half of the considered data windows. Summary comparative analysis of KPIs for all the modal parameters, c).

variate prediction strategy. So, basically, looking at Fig. 14, it can be said that for the dataset A1 the prediction performance is similar to the one of the uni-variate strategy, even though the RMSE values are slightly improved in the case of damping and frequency and worse in the case of amplitude, the punctual value of the prediction error can be higher.

6. Conclusions

In this work an LSTM-based ensemble strategy has been proposed and fully assessed to the aim of implementing a predictive monitoring system capable to identify the behavior of LFO phenomena some tens of seconds in advance. The strategy is founded on the use of modal parameters, extracted through the DMD algorithm, in order to build a proper 1-step multi-variate time-series prediction problem. This kind of approach is relatively new since there are few contributions in the literature exploiting the use of such parameters for predictive purposes. The development of the proposed multi-variate LSTM-based ensemble method has been described and its prediction performance has been assessed on two kinds of working scenarios: short LFO ringing events and long-run "ambient" conditions. Furthermore, the proposed multi-variate strategy has been comparatively assessed with respect to a uni-variate strategy. The obtained results show that the multi-variate technique is characterized by a prediction performance that is similar to the uni-variate strategy with no relevant increase of computational burden. In particular, for short ringing events, the multi-variate technique offers slightly better prediction performance, however, in the case of "ambient" datasets, even though it has similar statistical prediction performance than the uni-variate one, there is an increase of the number of points where the punctual value of the prediction error can be quite high. This result is ascribable to the fact that in the case of datasets with a behavior more similar to a random process, the use of multiple variables can be not effective. The use of a multi-variate technique can increase the number of cases where a steep variation on one variable, i.e. a modal quantity, can alter the prediction performance of another modal quantity. This last outcome shows how it is important, for the sake of the prediction problem, the use of additional electrical measurements (exogenous variables) in order to properly build the prediction problem. The initial assumption of using only frequency measurements from

PMUs seems to be a hardly limiting working condition for the sake of building effective predictive LFO monitoring systems. For future developments of the proposed predictive LFO monitoring system additional PMU measurement data will be considered in order to enrich enough the input dataset to be provided to the multi-variate prediction strategy, furthermore, also other more sophisticated prediction strategies will be considered.

CRediT authorship contribution statement

Carlo Olivieri: Writing – original draft, Software, Methodology, Conceptualization. **Francesco de Paulis:** Validation, Methodology, Data curation. **Cosimo Pisani:** Software, Resources, Methodology, Investigation. **Giorgio Giannuzzi:** Supervision, Resources, Project administration, Data curation. **Lino Di Leonardo:** Visualization, Investigation. **Antonio Orlandi:** Writing – review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

[1] L.L. Grigsby. The Electric Power Engineering Handbook – Power System Stability and Control, second ed., CRC Press, Taylor & Francis Group, New York, USA, 2007.
[2] S.M. Ustinov, J.V. Milanovic, V.A. Maslennikov, Inherent dynamic properties of interconnected power systems, Electr. Power Energy Syst. 24 (2002) 371–378.
[3] C. Kumar, et al., Detection of LFO and evaluation of damping improvement using synchrophasor measurement, IEEE PES Asia-Pac. Power Energy Eng. Conf. (2018) 701–706.
[4] K. El-Naggar, On-Line measurement of low-frequency oscillations in power systems, Measurement 42 (2009) 716–721.
[5] J.F. Hauer, Application of Prony analysis to the determination of modal Content and equivalent models for measured power system response, IEEE Trans. Power Syst. 6 (3) (1991) 1062–1068.
[6] Tao Liu, Yue Song, Lipeng Zhu, David J. Hill, Stability and control of power grids, Annu. Rev. Control, Robot., Auton. Syst. 5 (1) (2022) 689–716.
[7] K. Sebaa, S. Moulahoum, H. Houassine, N. Kabache, Model predictive control to improve the power system stability, in: 13th International Conference on Optimization of Electrical and Electronic Equipment (OPTIM), 2012, pp. 208-212.
[8] J. Ma, P. Zhang, H. Fu, Application of phasor measurement unit on locating disturbance source for low-frequency oscillation, IEEE Trans. Smart Grid 1 (3) (2010) 340–346.
[9] P. Korba, M. Larsson, C. Rehtanz, Detection of oscillations in power systems using Kalman filtering techniques, Proc. Conf. Control Appl. 1 (2003) 183–188.
[10] A. Prakash, K. Kumar, S.K. Parida, A modal transformation approach to design reduced order functional observer-based WADC for low-frequency oscillations, IEEE Trans. Power Syst. (2022), https://doi.org/10.1109/TPWRS.2022.3196787.
[11] T.A. Papadopoulos, A.I. Chrysochos, E.O. Kontis, Measurement-based hybrid approach for ringdown analysis of power systems, IEEE Trans. Power Syst. 31 (6) (2016) 4435–4446.
[12] K.P. Poon, K.C. Lee, Analysis of transient stability swings in large interconnected power systems by Fourier transformation, IEEE Trans. Power Syst. 3 (4) (1988) 1573–1581.
[13] J.L. Rueda, C.A. Juarez, I. Erlich, Wavelet-based analysis of power system low-frequency electromechanical oscillations, IEEE Trans. Power Syst. 26 (3) (2011) 1733–1743.
[14] D. Lauria, C. Pisani, Improved non-linear least squares method for estimating the damping levels of electromechanical oscillations, 8 1, IET Gener. Transm. Distrib. 9 (1) (2015) 1–11, https://doi.org/10.1049/iet-gtd.2014.0055.
[15] G. Giannuzzi, D. Lauria, C. Pisani, D. Villacci, Real-time tracking of electromechanical oscillations in ENTSO-e Continental European Synchronous Area, Int. J. Electr. Power Energy Syst. 64 (2015) 1147–1158.
[16] A.K. Gupta, K. Verma and K.R. Niazi, Wide-area PMU-ANN based monitoring of low frequency oscillations in a wind integrated power system, in: 8th IEEE India International Conference on Power Electronics (IICPE), jaipur, India, 2018, pp. 1-6.
[17] C. Olivieri, F. de Paulis, A. Orlandi, C. Pisani, G. Giannuzzi, R. Salvati, R. Zaottini, Estimation of modal parameters for inter-area oscillations analysis by a machine learning approach with offline training, ENERGIES 13 (23) (2020) 6410, https://doi.org/10.3390/en13236410.
[18] F. Sulla, E. Måsbäck, O. Samuelsson, Linking Damping of Electromechanical Oscillations to System Operating Conditions Using Neural Networks, IEEE Innovative Smart Grid Tech, 2014, pp. 1–6.
[19] V. Malbasa, C. Zheng, P. Chen, T. Popovic, M. Kezunovic, Voltage stability prediction using active machine learning, IEEE Trans. Smart Grid 8 (6) (2017) 3117–3124.
[20] A. Ioaneş and R. Tîrnovan, Power grid health assessment using machine learning algorithms, in: 2019 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE), Bucharest, Romania, 2019, pp. 1-4. doi: 10.1109/ATEE.2019.8724920.
[21] L. Maharjan, M. Ditsworth, M. Niraula, C.C. Narvaez, B. Fahimi, Machine learning based energy management system for grid disaster mitigation, IET Smart Grid 2 (2) (2019) 172–182.
[22] E. Ceperic, V. Ceperic, A. Baric, A strategy for short-term load forecasting by support vector regression machines, IEEE Trans. Power Syst. 28 (4) (Nov. 2013) 4356–4364.
[23] S.S. Reddy, C.-M. Jung, Short-term load forecasting using artificial neural networks and wavelet transform, Int J. Appl. Eng. Res 11 (19) (2016) 9831–9836.
[24] W. Kong, Z.Y. Dong, Y. Jia, D.J. Hill, Y. Xu, Y. Zhang, Short-term residential load forecasting based on LSTM recurrent neural network, IEEE Trans. Smart Grid 10 (1) (2019) 841–851.
[25] Pengfei Zhao, Weihao Hu, Di Cao, Zhenyuan Zhang, Wenlong Liao, Zhe Chen, Qi Huang, Enhancing multivariate, multi-step residential load forecasting with spatiotemporal graph attention-enabled transformer, Int. J. Electr. Power Energy Syst. 160 (2024).
[26] F.L. Grando, A.E. Lazzaretti, M. Moreto, and H.S. Lopes, Fault classification in power distribution systems using PMU data and machine learning, in: 2019 20th International Conference on Intelligent System Application to Power Systems, ISAP 2019, 2019.
[27] R. Eskandarpour, A. Khodaei, Machine learning based power grid outage prediction in response to extreme events, IEEE Trans. Power Syst. 32 (4) (2017) 3315–3316.
[28] P. Kankanala, et al., Adaboost: an ensemble learning approach for estimating weather-related outages in distribution systems, IEEE Trans. Power Syst. 29 (1) (2013) 359–367.
[29] A. Lenzi, J. Bessac, M. Anitescu, Power grid frequency prediction using spatiotemporal modeling, Stat. Anal. Data Min. ASA Data Sci. J. 14 (2021) 662–675.
[30] Y. Bai, Y. Liu, K. Liu, N. Sun, A. Bhatt, E. Del Rosso, Farantatos, X. Wang, Measurement-based correlation approach for power system dynamic response estimation, IET Gener. Transm. Distrib. 9 (2015) 1474–1484.
[31] R. Diao, K. Sun, V. Vittal, R.J. O'Keefe, M.R. Richardson, N. Bhatt, D. Stradford, S. K. Sarawgi, Decision tree-based online voltage security assessment using PMU measurements, IEEE Trans. Power Syst. 24 (832–8392014) (2009) 1–5.
[32] R. Fainti, M. Alamaniotis, L.H. Tsoukalas, Three-phase line overloading predictive monitoring utilizing artificial neural networks," in: 2017 19th International Conference on Intelligent System Application to Power Systems, ISAP 2017, 2017.
[33] C. Haseltine and E.E. Eman, Prediction of power grid failure using neural network learning, in: Proc. 16th IEEE Int. Conf. Machine Learning and Applications (ICMLA), Dec. 2017, pp. 505–510.
[34] S.G. Soares, R. Araújo, An on-line weighted ensemble of regressor models to handle concept drifts, Eng. Appl. Artif. Intell. 37 (Jan. 2015) 392–406.
[35] Y. Xu, R. Xu, W. Yan, Power plant performance modeling with concept drift" in: Proc. Int. Joint Conf. Neural Netw. (IJCNN), May 2017, pp. 2096–2103.
[36] S.G. Soares, R. Araújo, A dynamic and on-line ensemble regression for changing environments, Expert Syst. Appl. 42 (6) (Apr. 2015) 2935–2948.
[37] R. de Almeida, Y.M. Goh, R. Monfared, M.T.A. Steiner, A. West, An ensemble based on neural networks with random weights for online data stream regression, Soft Comput. 24 (13) (Jul. 2020) 9835–9855.
[38] D. Wang, J. Xiong, Z. Xiao, and X. Li, Short-term traffic flow prediction based on ensemble real-time sequential extreme learning machine under non-stationary condition, in: Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring), May 2016, pp. 1–5.
[39] C. Olivieri, G. Giannuzzi and F. de Paulis, Comparison of LSTM-based prediction strategies for grid modal parameters forecast, in: 2023 11th International Conference on Smart Grid (icSmartGrid), Paris, France, 2023, pp. 1-6.
[40] C. Olivieri, F. de Paulis, A. Orlandi, C. Pisani, S. Tessitore, G. Giannuzzi, Application of an integrated RNN-ensemble method for the short-term forecast of inter-area oscillations modal parameters, Electr. Power Syst. Res. 225 (2023).
[41] Peter Schmid. Dynamic mode decomposition of numerical and experimental data. Journal of Fluid Mechanics, Cambridge University Press (CUP), 2010, 656 (August), pp.5-28.
[42] J.N. Kutz, Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data, Oxford Univ. Press, New York, NY, USA, 2013.
[43] K.K. Chen, J.H. Tu, C.W. Rowley, Variants of dynamic mode decomposition: boundary condition Koopman and Fourier analyses, J. Nonlinear Sci. 22 (6) (Mar. 2012) 887–915.
[44] K. Chen, Indirect PCA dimensionality reduction based machine learning algorithms for power system transient stability assessment, in: 2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), Chengdu, China, 2019, pp. 4175-4179. doi: 10.1109/ISGT-Asia.2019.8881370.
[45] D. Niebur, E. Gursoy, H. Liao, Applied mathematics for restructured electric power systems. power electronics and power systems, in: Chow, H. J., F.F. Wu, J. Momoh (Eds.), Independent Component Analysis Techniques for Power System Load Estimation, Springer, Boston, MA, 2005.

[46] Y. Ren, L. Zhang, P.N. Suganthan, Ensemble classification and regression-recent developments, applications and future directions, IEEE Comput. Intell. Mag. 11 (1) (Feb. 2016) 41–53, https://doi.org/10.1109/MCI.2015.2471235.

[47] M. Lima, M. Neto, T.S. Filho, R.A. de, A. Fagundes, Learning under concept drift for regression—a systematic literature review, IEEE Access 10 (2022) 45410–45429, https://doi.org/10.1109/ACCESS.2022.3169785.

[48] ⟨https://scikit-learn.org/stable/⟩.