Comparison of LSTM-based Prediction strategies for Grid Modal Parameters Forecast

Carlo Olivieri
Department of Industrial and Information Engineering and Economics
University of L'Aquila
L'Aquila, Italy
carlo.olivieri@univaq.it

Giorgio Giannuzzi
TERNA S.p.A.
Rome, Italy
giorgio.giannuzzi@terna.it

Francesco de Paulis
Department of Industrial and Information Engineering and Economics
University of L'Aquila
L'Aquila, Italy
francesco.depaulis@univaq.it

Abstract— **The high penetration of renewable energy sources poses great challenges for transmission system operators, especially concerning the detrimental phenomenon of electromechanical Inter-Area Oscillations. Although the actual monitoring techniques can offer a useful baseline in order to fight against such phenomena, predictive features are highly desirable in this context. This work presents a preliminary comparative study of two prediction strategies suitable to forecast the short-term values of the grid modal parameters. The considered strategies are based on the proper integration of the Dynamic Mode Decomposition technique with Machine Learning techniques such as Long-Short-Term Memory units and Ensemble methods. The development steps of both techniques are fully illustrated and the performance comparison is done by accounting for some key performance indicators. Two assessment scenarios are considered, based on the availability of some real measurement data.**

Keywords—Inter-Area Oscillations, Modal Analysis, PMU Monitoring, LSTM, Ensemble, Dynamic Mode Decomposition

I. INTRODUCTION

The inherently present difficulties in the management of modern power grids are, at present, continuously increasing with the natural development of the energy transition process. In particular, the reduction of the power grid inertia and the contextual increase of the power flow over long distances brings unavoidable stability problems. The occurrence of Inter-Area Low-Frequency Oscillations (LFO) [1] is a concrete and recurrent problem that has to be faced day-by-day by Transmission System Operators (TSOs), most of the time with the desirable aid of sophisticated monitoring techniques. In this context, the standard and more mature power grid stabilization techniques, even though useful, are not very well suited to face oscillatory phenomena which lie in the sub-hertz low-frequency range (we refer herefter to the typical range of 0.1 to 0.8 Hz [2]). As known, most of the time, the problem of LFO control (i.e. damping) is approached through the preliminary step of identifying the occurrence of such critical oscillations [3-4]. Plenty of techniques have been proposed for LFO identification [4-8], however, in a large part of these approaches, the proposed methods are fundamentally based on the proper estimation of the so-called "modal parameters" of the oscillatory phenomenon [6-8]. Basically, they are triplets of values identifying the oscillatory behavior of a certain electromechanical mode. Each triplet of modal parameters is composed of the values of Frequency (f_i), Damping ratio (α_i), and Amplitude (a_i), for the specific *i-th* electromechanical mode under consideration. Several approaches already present in the literature can provide the analysis of the LFO phenomena by continuously monitoring the values of such parameters over time [4],[8]. Despite what has been proposed in the past decades, at present, the ever-increasing development of sophisticated measurement devices and architectures disclosed a wide set of more advanced monitoring capabilities, this is the case for instance of Phasor Measurement Units (PMU) and Wide Area Meaurement Protection and Control (WAMPAC) solutions [8-10]. Even though the efforts spent to improve the existing modal parameters estimation techniques, based on these new capabilities, brought several successful results [8], the prediction of the time evolution of LFO is still a hard and challenging matter [9]. In this context, the parallel development of sophisticated forecasting algorithms [10-12], could pave the way for interesting predictive monitoring techniques. Machine Learning (ML) techniques seem to offer promising results for different estimation or forecasting applications [12-14], however, their application for the predictive monitoring of modal parameters seems to be still an unexplored field of study. In this paper, two different prediction strategies, based on the use of Long-Short-Term Memory (LSTM) units, have been comparatively assessed for the purpose of predicting the short-term behavior of grid modal parameters during two different operation scenarios. One strategy accounts for the use of an Encoder-Decoder LSTM structure whereas the other strategy exploits the use of Ensemble methods. The general approach and the specific building steps of both the two strategies are illustrated in Sections II and III, respectively. The two evaluated operating scenarios refer to the occurrence of LFO ringing events or to rated operating conditions of the grid, let's say "ambient" working conditions. The data related to the considered scenarios and coming from real measurements operated by the TSO is detailed in Section IV. The results obtained from the comparative analysis are concisely presented in Section V and some preliminary conclusions are outlined at the end.

II. BACKGROUND AND GENERAL PREDICTION APPROACH

As already mentioned in the introductory section, the continuous tracking of the values assumed by the so-called modal parameters, that is the triplet (f_i, α_i, a_i), for each relevant *i-th* LFO mode of the power grid, can constitute an accurate and powerful monitoring technique suitable to discover critical LFO modes in real-time [6]. The additional required feature to realize a predictive monitoring system can be the possibility to make accurate forecasts of the time evolution of such parameters over a certain time horizon. It is quite straightforward that the starting point of monitoring and predictive monitoring strategies need to be the same, that is in our case, the collection of the frequency measurements coming from some PMU devices (done at certain points of the power grid) and their proper elaboration through suitable estimation or prediction algorithms. In the specific context of this paper, it can be made reference to the top-level block diagram reported in the following 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. This initial preprocessing consists of a data detrending operation and a band-pass filtering action, suitable to isolate the frequency content of interest and to exclude all the spurious measurement signals. The filtering action is properly configured in order to have transition bands set at 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, where the length of each data window is defined by the number of samples considered, here in the following indicated by the parameter L_w. 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 through the use of the Dynamic Mode Decomposition (DMD) algorithm [13]. Although plenty of numerical methods are available to estimate the modal parameters starting from a set of frequency measurements, here we chose to use this algorithm since it offers several interesting features. First of all, it has the possibility to operate a dimensionality reduction, based on the number of modes selected for the study, indicated by the parameter N_{mod}, secondly, it has good numerical stability and limited computational burden, these last features make it ideal for the implementation on a real monitoring system. According to what stated by the related literature, the DMD algorithm is based on the collection of snapshots of the input data arranged as a sequence of the following form:

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

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. In this application, each snapshot is constituted by the instantaneous acquisition of frequency over the *p* PMU locations. The DMD method is founded on the theory for which the modal parameters can be properly found by establishing a linear mapping between the *N-th* snapshot and the *N-1-th* snapshot of data, following the relationship (2), which is the core of the DMD technique.

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

Here $X_1^{N-1} = \{x_1, x_2, \cdots, x_{N-1}\}$ and $X_2^N = \{x_2, x_3, \cdots, x_N\}$ are two subsets of the initial collection and the term *r* stands for the vector residuals accounting for the dynamic behaviors that cannot be completely modeled by the linear mapping. From a computational point of view, the DMD algorithm extracts the values of the modal parameters by executing three main steps: a Singular Value Decomposition (SVD), a linear projection, and an Eigenvalue Decomposition (EVD). As depicted in Fig.1, the DMD algorithm ouputs a triplet of values (f_i, α_i, a_i) for each data window under consideration and for each mode of interest (from mode #1 up to N_{mod}).

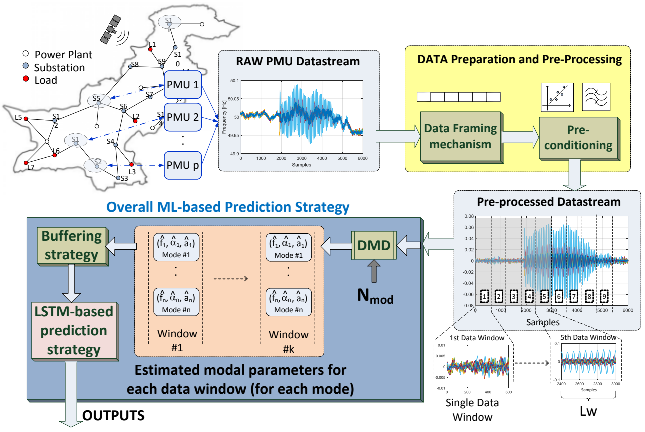

Fig. 1. Top-level description of the prediction approach.

The proposed prediction approach is such that, given the past values of the modal parameters (computed by the DMD) known at a certain time instant t*, the method can output the prediction of the values assumed for a given number of future time windows, i.e. the predicted values for a certain number of time steps in advance in time, indicated as N_w. For each time window there is a set of triplets of values, that constitutes a sample of the modal parameters over a time horizon of length L_w, in this case study we assume N_w=2. The key point in the proposed prediction approach is that the forecast of the next values of the triplets is obtained by executing a multi-step prediction over a set of univariate time series, as highlighted in the following Fig.2. That is, the three modal parameters of the triplet (f_i, α_i, a_i) of a given mode are treated in such a manner that they define three independent time series prediction problems over the time evolution of the input data windows. In order to manage the continuously incoming datastream from the PMU devices in the real application, it is necessary to introduce a proper buffering strategy, suitable to have the availability of a certain number of past values of the modal parameters for the given number of modes of interest. For the sake of simplicity but without loss of generality, here in this paper only the so-called "dominant" mode, which is the one with the highest energy, has been considered for prediction purposes, i.e. mode#1.The buffering strategy, as reported in Fig. 2, is merely constituted by a set of FIFO buffers, each one built for the specific modal parameter to be forecast (for each mode) and suitable to accomplish two main roles: on one side, it enables the arrangement of the past values of the modal parameter of interest into bigger slices of data, namely "data chunks", of length L_{ch} and, on the other side, it realizes a queue mechanism suitable to have a fixed number of past values for each iterative step of prediction. Since the overall strategy evolves over time, when new fresh data is available from the PMUs, it is added into the buffer and older data is discarded.

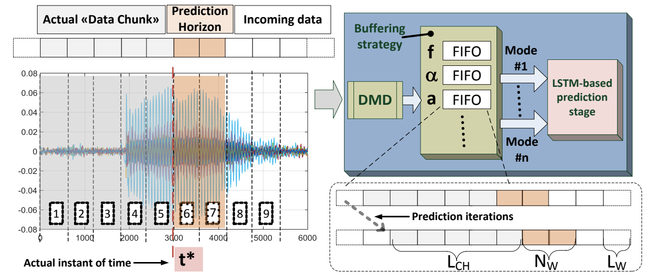

Fig. 2. Buffering strategy and input data segmentation using "chunks of data". The prediction horizon is equal to N_w =2.

III. EVALUATED PREDICTION STRATEGIES

The core business of the prediction approach is constituted by the LSTM-based prediction stage of Fig.1, having the inner structure detailed in Fig.3. As can be seen, it consists of an array of Prediction Units (PUs), which output the ultimate modal parameters forecasts based on the times series built for each modal parameter through the use of the buffering strategy. Each PU is implemented by using one of either the evaluated prediction strategies. In this paper two different prediction strategies are considered for the generation of the forecasts, they are described in the following of this section and qualitatively compared in Section V. The first one (namely strategy "S1") is an LSTM-based strategy using an Encoder-Decoder model whereas the second one ("S2") is a custom solution combining a set of weak LSTM predictors by using an ensemble method.

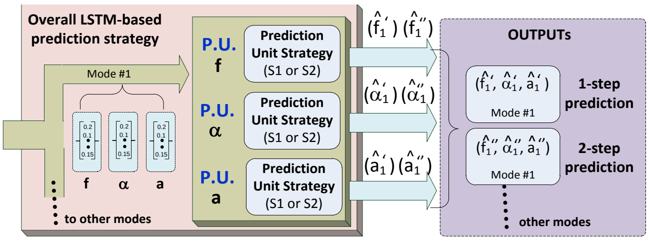

Fig. 3. Functional description of the LSTM-based prediction strategy block. The 1-step and 2-step predictions of each modal parameter are provided by a specific Prediction Unit (PU).

A. Encoder-Decoder LSTM-based P.U. (strategy S1)

The first considered strategy is the one where each PU is constituted by a single LSTM-based Encoder-Decoder model (with univariate input), having the structure detailed in the following Fig.4. The Encoder section is constituted by a stack of a given number of vanilla LSTM units (in this case parameterized by the variable N_k), they are used to process the input data and to provide a fixed lenght representation of it. The Encoder Vector is the final state produced from the encoder part; the hidden states are calculated with the rule:

$$h_t = f(W^{(hh)}h_{t-1} + W^{(hx)}x_t) \qquad (3)$$

The decoder section is constituted by a stack of LSTM units where each one predicts an output y_t at a given instant t, each LSTM unit accepts a hidden state coming from the unit and produces an output equal to its own hidden state. The final output y_t at time step t is computed by using the underlying well-known relationships (4).

$$h_t = f(W^{(hh)}h_{t-1}), \quad y_t = \mathrm{softmax}(W^S h_t) \qquad (4)$$

The output of the Encoder-Decoder prediction unit is composed of two quantities: the one-step prediction of the specific modal parameter of interest and the two-step one.

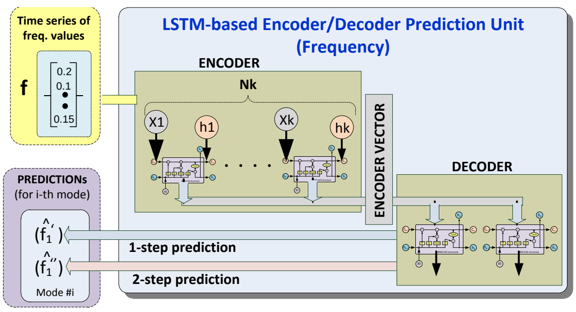

Fig. 4. Schematic diagram of the frequency prediction unit making use of the Encoder-Decoder LSTM-based prediction strategy (strategy S1).

B. Integrated LSTM-based Ensemble P.U. (strategy S2)

The second evaluated prediction strategy is an integrated solution combining a set of weak predictors through the use of an Ensemble method having a mere average function as its aggregation function. Each weak predictor is constituted by a vanilla LSTM model with a predefined number of layers, indicated by the parameter N_L. The number of members in the ensemble is parameterized by the variable N_{EM}. The general view of the strategy is reported in Fig. 5.

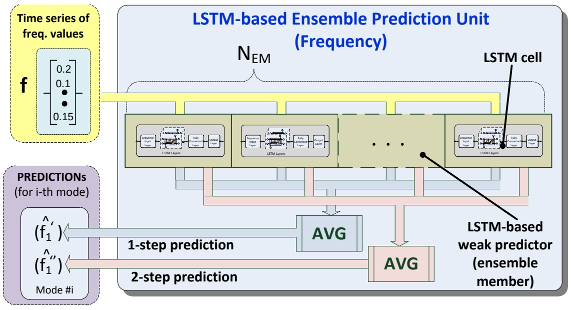

Fig. 5. Schematic diagram of the frequency prediction unit making use of the proposed LSTM-based integrated ensemble strategy (strategy S2).

The main purpose of this second kind of strategy is to check if there is the possibility to get an improved prediction behavior by following the paradigm of Parameter Diversity [14] since the overall prediction approach has to face a continuous input data stream in its final application domain. Even though this should be formally treated as an Incremental Learning problem, for the actual specific case study also a bare performance improvement can be retained of interest. Starting from the related literature [14-16], ensemble methods have been demonstrated to be effective in obtaining better performance in case of Concept Drift [14]. Following the paradigm of Parameter Diversity, the number of hidden units for each LSTM model composing the ensemble (N_{HU}) has been set with the following rule.

$$\begin{cases} N_{HU}^{(i)} = \left(\left[\dfrac{\left(N_{HU}^{M} - N_{HU}^{m}\right)}{N_{EM}} \right] + 1 \right) \cdot (i-1) + N_{HU}^{m} & i = 1, \dots, N_{EM} - 1 \\ N_{HU}^{(i)} = N_{HU}^{M} & i = N_{EM} \end{cases} \qquad (5)$$

where i is the index of the ensemble member, ranging from 1 to N_{EM}, while N_{HU}^{M} and N_{HU}^{m} indicate the maximum and minimum number of hidden units. In this case study we can consider for simplicity the minimum number equal to 5 and the maximum number equal to the parameter N_{HU}, with equal number of hidden units among all the used layers. The rule in (5) is used to uniformly distribute the number of hidden units among the ensemble members.

The final 1-step and 2-step predicted values are obtained from the computation of the average value from all the ensemble members, as clearly depicted in Fig. 5.

IV. CONSIDERED APPLICATION DATA AND TEST SCENARIOS

As already mentioned, the two considered strategies have been assessed by checking their prediction performance in two different application scenarios. They are aimed to compare the performance of the two strategies under two real working conditions, that are: during short ringing events and over long periods of rated working conditions of the grid. The measurement data related to these two scenarios have been provided by the national TSO and are detailed in the following two subsections, they refer to real grid events [17].

A. Ringing Events

Two real ringing events have been recorded by the TSO and are used in the following for the comparison of the prediction performances over short time durations. The first ringing event (namely, dataset "R1") is described by the frequency measurements plotted in the following Fig. 6 a). It considers 30 PMU locations and a time duration of almost 20 minutes, with a sampling time of 100 ms. By considering also the length of the data windows involved in the sliding-window framing mechanism equal to L_w=300 samples, this translates to a total of 39 windows of data for the assessment, since the first window has been cut in order to eliminate transient effects due to the filtering stage depicted in Fig.1. The second ringing event (namely "R2") is characterized by a shorter time duration (about 10 mins), as depicted in Fig. 6 b), and hence a dataset including only 19 windows.

a)

b)

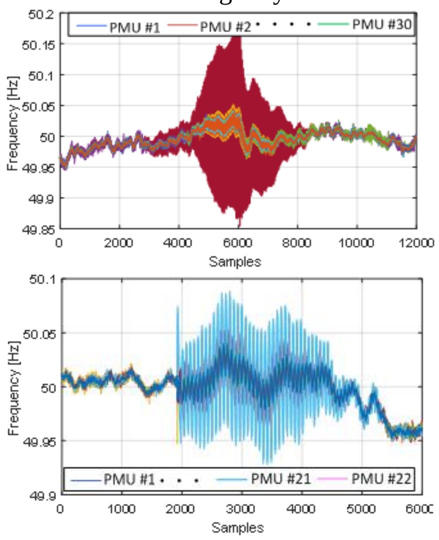

Fig. 6. Measurement data related to: a), the first considered ringing event "R1", and, b), to the second rining event "R2".

B. Rated grid operation ("Ambient" data)

For what concerns with the second application scenario, the dataset used to compare the two strategies refer to the case of 24 hours of PMU recordings with the grid operating in rated conditions (meaning without noticeable oscillatory events), let's say with "ambient" conditions and over "long-runs". This dataset accounts for 4320 windows of length L_w=200, 100ms sampling and it is referred hereafter as "A1".

C. Summary of test scenarios

In this subsection there are summarized the proposed test cases, mapping the mentioned datasets with the two strategies and with a set of possible internal parameters configurations. First of all, the configurations used for the two strategies are concisely reported, in Table I for strategy S1 and S2. Table II describes instead the mapping of the proposed test scenarios. The first test scenario (TS#1) is aimed to select the most relevant configuration of strategy S1, whereas TS#2 is aimed to select the most relevant configuration of strategy S2. The third test scenario (TS#3) compares the two selected configurations of strategy S1 and S2 over the two ringing events R1 and R2. Finally, the test scenario TS#4 compares the two solutions over dataset A1.

V. Results

The results related to the prediction peformance of the two strategies are concisely summarized in the following subsections. The two strategies have been implemented by using the Python-based Anaconda environment and some of its related libraries: Keras, ScikitLearn[18]. The used training algorithm is the ADAM one. The key performance indicators used for the assessment are: the total error statistics (mean value and standard deviation of the sum of the 1-step and 2 step prediction errors), plus the Mean Absolute value, the RMSE and the CPU time for each prediction iteration.

TABLE I. PARAMETER CONFIGURATIONS FOR STRATEGY S1AND S2

Config. ID (S1)	Architectural Parameters	
	N_K	*Max Epochs*
C1	200	20
C2	200	50
C3	200	100
C4	100	100
C5	50	100
C6	150	100

Config. ID (S2)	Architectural Parameters			
	N_K	N_{HU}	N_{HL}	*Max Epochs*
C1	3	20	2	100
C2	5	20	2	100
C3	7	20	2	100
C4	15	20	2	100
C5	5	30	2	100
C6	5	10	2	100

TABLE II. SUMMARY OF PROPOSED TEST SCENARIOS

Test Scenario	Parameter	Dataset	Considered Configurations	*Lch*
TS#1	Frequency	R1	S1-C1, S1-C2, S1-C3, S1-C4, S1-C5, S1-C6	5
TS#2	Frequency	R1	S2-C1, S2-C2, S2-C3, S2-C4, S2-C5, S2-C6	5
TS#3	Frequency	R1	S1-C2, S2-C2	5
TS#3	Damping	R1	S1-C2, S2-C2	5
TS#3	Amplitude	R1	S1-C2, S2-C2	5
TS#3	Frequency	R2	S1-C2, S2-C2	5
TS#3	Damping	R2	S1-C2, S2-C2	5
TS#3	Amplitude	R2	S1-C2, S2-C2	5
TS#4	Frequency	A1	S1-C2, S2-C2	50
TS#4	Damping	A1	S1-C2, S2-C2	50
TS#4	Amplitude	A1	S1-C2, S2-C2	50

A. Test Scenario TS#1

In this test scenario the prediction performance of strategy S1 has been assessed with respect to the different configurations of Table I and in order to select the best performing set, the frequency modal parameter has been considered as reference in this case. The computation of the aforementioned KPIs for the six considered configurations of S1 brings to the results reported in the following Table III. In particular, as shown, depending on the selected strategy parameters, the RMSE value can be lowered by about 15% by increasing the computational burden, whereas, the error mean value is almost stable. Since this last one is quite low it follows that the RMSE value and the error standard deviation are similar. However, as shown, a possible "best trade-off" strategy configuration, accounting for both RMSE and CPU time, can be configuration S1-C2.

TABLE III. RESULTS FOR FREQUENCY PREDICTION WITH STRATEGY S1 AND BY USING DIFFERENT CONFIGURATIONS

Config. ID	*Mean*	*Std. Dev.*	*MAE*	*RMSE*	*Time [s]*
S1-C1	-0,01293	0,059758	0,066529	0,059649	2,8762
S1-C2	-0,01385	0,053281	0,065627	0,053286	6,7554
S1-C3	-0,0126	0,050287	0,059354	0,050253	14,0232
S1-C4	-0,01624	0,052492	0,062986	0,052408	7,3321
S1-C5	-0,01413	0,055375	0,0617	0,055298	6,1842
S1-C6	-0,01017	0,053927	0,062596	0,053868	9,1866

B. Test Scenario TS#2

The prediction performance of strategy S2 has been assessed for the various configurations of Table I and the results have been resumed in Table IV, similarly to TS#1. Also in this test scenario, depending on the selected strategy parameters, the frequency prediction RMSE value can be lowered by choosing a different configuration setup; the possible RMSE reduction can be nearly 18%, however, there is a considerable increase of the computational effort.

Hence, taking into account that 30 s is the duration of a time window and that there are also other extra timings that should be taken into account in the real application (which reduce the available time slot to make the prediction), configuration S2-C2 seems to be the best trade-off between RMSE and computational burden. It is worth noting that the mean value of the prediction error obtained with strategy S2 is dependent on the number of ensemble members used. Furthermore, the mean and the RMSE values are generally slightly lower than what can be obtained with strategy S1.

The results reported in Table IV show that strategy S2 has the potential to outperform strategy S1, according to what can be expected from other previous theoretical results [14-16]. However, the results have been obtained accounting only for the frequency parameter, so, additional comparisons are needed, this is the aim of test scenarios TS#3 and TS#4.

TABLE IV. RESULTS FOR FREQUENCY PREDICTION WITH STRATEGY S2 AND BY USING DIFFERENT CONFIGURATIONS

Config. ID	*Mean*	*Std. Dev.*	*MAE*	*RMSE*	*Time [s]*
S2-C1	-0.018018	0.0611586	0.0648157	0.0611564	2,029968
S2-C2	-0.0121022	0.0507147	0.0632632	0.0507106	4,148987
S2-C3	-0.0079017	0.0510788	0.0616197	0.0510759	8,057868
S2-C4	-0.009658	0.0501687	0.0614804	0.0501616	17,97441
S2-C5	-0.0099243	0.0519425	0.0634825	0.0519384	5,237083
S2-C6	-0.0122348	0.0561142	0.0633063	0.0561103	3,212854

C. Test Scenario TS#3

In this test scenario the prediction performance of strategy S1 versus strategy S2 have been compared by using the best performing configurations, that are: S1-C2 and S2-C2. The comparison related to the case of dataset R1 is detailed by the plots reported in Fig. 7., whereas for the dataset R2 the plots reported in Fig. 8 are used to synthesize the analysis of the main KPIs.

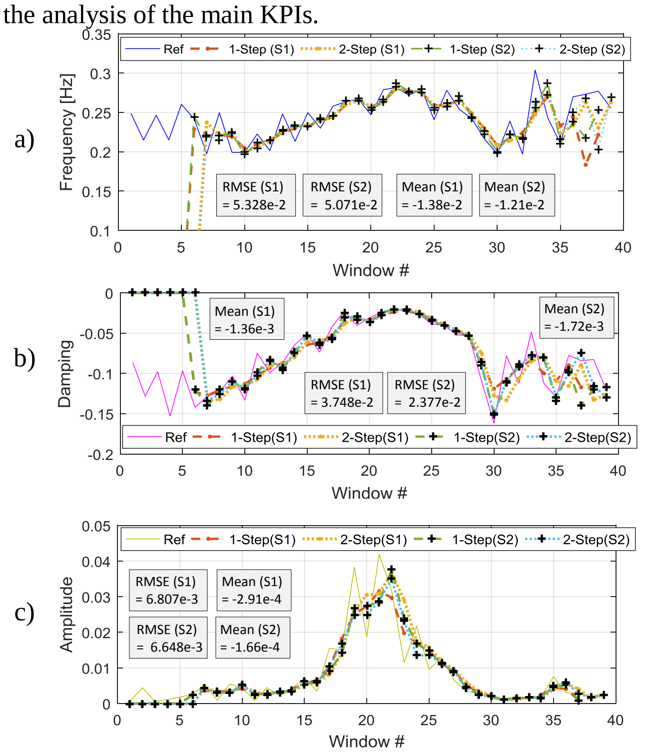

Fig. 7. Comparison of prediction performance of strategy S1 versus S2 for the different modal parameters: a) Frequency, b), Damping, c) Amplitude.

Fig.7 details the behaviour of either 1-step or 2-step predictions for the same dataset R1 and for both strategies, including also the values of the main KPIs. Fig.7 depicts the reference behavior of the specific modal parameter of interest, calculated though the application of the DMD algorithm over the full time horizon, and the quantitites predicted by the two strategies, for each of the three plots. The values assumed by the error standard deviation are not reported since they are fairly similar to the RMSE values, as already mentioned for test scenario TS#1, MAE values are neither reported. As can be seen, the strategy S2 can generally offer better performance in terms of both RMSE and mean value of the total prediction error, except in the case of mean value of damping modal parameter. Fig.8 details a comparison of the considered KPIs for all the three modal parameters and for both the strategies S1 and S2 in the case of dataset R2. From the obtained results it can be noted that strategy S2 offers slightly better performance for the cases of frequency and amplitude prediction but not for damping, where the encoder-decoder based strategy S1 seems to be offering a lower total error mean value and a lower RMSE.

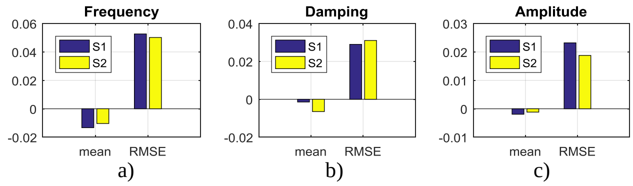

Fig. 8. Comparison of the main KPIs for the case of ringing event R2.

D. Test Scenario TS#4

In this last case, the two strategies have been compared by looking at the prediction performance over long runs. The plots reported in Fig. 9 provide the details of the frequency prediction performance (Fig.9a), with both strategy S1 and S2, and a summary analysis of the main KPIs, mean value of error and RMSE, for all the three modal parameters (Fig.9b).

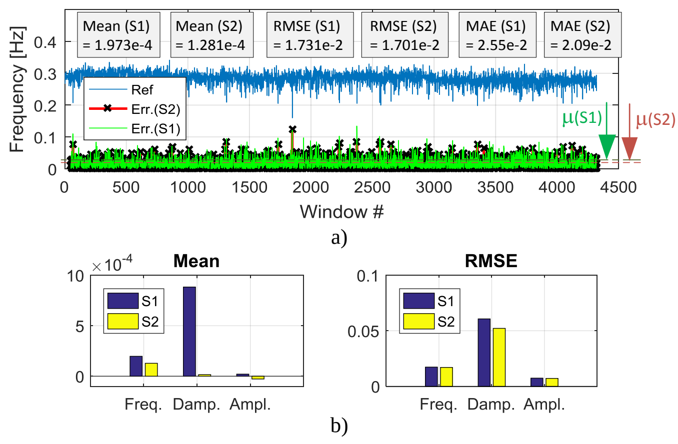

Fig. 9. Prediction performance for dataset A1: a) total absolute frequency prediction error for both strategies versus the reference frequency values, b) summary analysis of the KPIs for the three modal parameters.

As highlighted by the results of Fig.9 and in particular from the behavior of the total frequency prediction error for both strategies, it follows that, in the case of "ambient data", even though the prediction error is limited, the predictions can be retained acceptable only in a mean value sense, since the punctual prediction error can be quite relevant. Also in the cases of damping and amplitude modal parameters the total prediction errors are characterized by a behavior similar to that encountered for frequency prediction. This means that, due to the nature of this dataset, both the strategies are hardly put to the test in predicting the steep variations of ambient data. However, also with ambient data, the strategy S2 can offer a slightly lower error mean value and RMSE, especially in the case of the damping modal parameter. This better performance of strategy S2 can be ascribable to the beneficial effects provided by averaging the predictions put out by a committee of weak learners, following the theoretical outcomes of ensemble methods and in particular the bias-variance-covariance decomposition [15]. Despite of this, also by using strategy S2, the accuracy in the prediction of the modal parameters with ambient data claims for further improvements, since in the real application the prediction error has to be strictly related with the operational thresholds to be used in recognizing critical LFO phenomena. This is particularly true in the prediction of the damping parameter.

CONCLUSION

This paper has developed a comparative analysis of two LSTM-based prediction strategies aimed to output the short-term forecasts of the grid modal parameters under different application scenarios. From the obtained results it seems that the ensemble-based strategy has generally a slightly better performance with respect to the encoder-decoder one and a comparable computational burden. However, it has to be clarified in advance that further research efforts are required in order to make the overall approach suitable for a real application in the field. As the first point, the prediction performance should be improved with respect to ringing scenarios and "ambient" conditions in particular. Actually, both the strategies have shown poorer performance with dataset A1, since LSTM structures can catch well long-term dependencies but within the ambient dataset there are no clear long-term dependencies. A possibility can be the use of different base architectures, other than LSTM networks, or, as an alternative, to try implementing effective incremental learning features, which could provide more accurate results. So, possible viable options suitable to improve the main presented approach should consider more sophisticated base prediction models, or, at least, multi-variate prediction strategies, in order to overcome the limited accuracy linked with forecasts executed in a separate way for each modal parameter. Furthermore, as a second point, the use of additional input data is advisable in order to increase the accuracy of the predictions by a proper extension of the method, for instance, the use of ROCOF or other grid measurements can be considered as useful auxiliary data.

REFERENCES

[1] L.L. Grigsby, “The Electric Power Engineering Handbook – Power System Stability and Control”, Taylor & Francis, NY, USA, 2007.

[2] S.M. Ustinov, J.V. Milanovic, V.A: Maslennikov, “Inherent Dynamic Properties of Interconnected Power Systems”, Electric Power and Energy Systems, vol. 24, pp. 371-378, 2002

[3] C. Kumar et al., "Detection of LFO and Evaluation of Damping Improvement Using Synchrophasor Measurement," IEEE PES Asia-Pacific Power & Energy Engineering Conference, 2018, pp. 701-706

[4] K. El-Naggar, "On-Line measurement of low-frequency oscillations in power systems", Measurement, 42, pp. 716-721, 2009

[5] Korba P, Larsson M, Rehtanz C. "Detection of oscillations in power systems using Kalman filtering techniques", Proc conf. on control applications, vol.1, p. 183–8 , 2003

[6] G. Giannuzzi, D. Lauria, C. Pisani, D. Villacci, "Real-time tracking of electromechanical oscillations in ENTSO-e Continental European Synchronous Area", Int. J. of Electrical Power & Energy Systems, Vol. 64, pp. 1147-1158, 2015

[7] A. Prakash, K. Kumar and S. K. Parida, "A Modal Transformation Approach to Design Reduced Order Functional Observer-Based WADC for Low-Frequency Oscillations," in IEEE Transactions on Power Systems, 2022, doi: 10.1109/TPWRS.2022.3196787.

[8] K. Sebaa, S. Moulahoum, H. Houassine, N. Kabache, "Model predictive control to improve the power system stability", 13th International Conference on Optimization of Electrical and Electronic Equipment (OPTIM), 2012, pp. 208-212,

[9] Tao Liu, Yue Song, Lipeng Zhu, David J. Hill, "Stability and Control of Power Grids" Annual Review of Control, Robotics, and Autonomous Systems, 2022 5:1, 689-716

[10] A. Lenzi, J. Bessac, and M. Anitescu, "Power grid frequency prediction using spatiotemporal modeling", Stat. Anal. Data Min.: ASA Data Sci. J. 14 (2021), 662–675

[11] V. Malbasa, C. Zheng, P. Chen, T. Popovic, M. Kezunovic, "Voltage Stability Prediction Using Active Machine Learning", IEEE Trans. on Smart Grid, vol.8, no.6, pp. 3117-3124, 2017.

[12] Olivieri C, de Paulis F, Orlandi A, Pisani C, Giannuzzi G, Salvati R, Zaottini R., "Estimation of Modal Parameters for Inter-Area Oscillations Analysis by a Machine Learning Approach with Offline Training", ENERGIES, 2020; 13(23):6410.

[13] Peter Schmid. Dynamic mode decomposition of numerical and experimental data. Journal of Fluid Mechanics, Cambridge University Press (CUP), 2010, 656 (August), pp.5-28.

[14] M. Lima, M. Neto, T. S. Filho, R. A. de A. Fagundes, "Learning Under Concept Drift for Regression—A Systematic Literature Review", IEEE Access, vol. 10, pp. 45410-45429, 2022

[15] Y. Ren, L. Zhang and P. N. Suganthan, "Ensemble Classification and Regression-Recent Developments, Applications and Future Directions", in IEEE Computational Intelligence Magazine, vol. 11, no. 1, pp. 41-53, Feb. 2016

[16] S. Kumar, P. Kaur and A. Gosain, "A Comprehensive Survey on Ensemble Methods", IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 2022, pp. 1-7,

[17] https://eepublicdownloads.entsoe.eu/clean-documents/SOC%20documents/Regional_Groups_Continental_Europe/OSCILLATION_REPORT_SPD.pdf

[18] https://scikit-learn.org/stable/./.