Variational Mode Decomposition as Trusted Data Augmentation in ML-based Power System Stability Assessment [★]

Tetiana Bogodorova [*] **Denis Osipov** [**] **Luigi Vanfretti** [*]

[*] *Electrical, Computer, and Systems Engineering Department, Rensselaer Polytechnic Institute, Troy, NY 12180 USA, (e-mail: bogodt2@rpi.edu; vanfrl@rpi.edu).*
[**] *System Planning & Analysis, New York Power Authority, Albany, NY 12207, USA (e-mail: Denis.Osipov@nypa.gov)*

Abstract: Balanced data is required for deep neural networks (DNNs) when learning to perform power system stability assessment. However, power system measurement data contains relatively few events from where power system dynamics can be learnt. To mitigate this imbalance, we propose a novel data augmentation strategy preserving the dynamic characteristics to be learnt. The augmentation is performed using Variational Mode Decomposition. The detrended and the augmented data are tested for distributions similarity using Kernel Maximum Mean Discrepancy test. In addition, the effectiveness of the augmentation methodology is validated via training an Encoder DNN utilizing original data, testing using the augmented data, and evaluating the Encoder's performance employing several metrics.

Keywords: Convolutional neural networks, data augmentation, deep learning, power system stability assessment, variational mode decomposition.

1. INTRODUCTION

Power grid measurements, even if regularly collected and stored, do not contain enough event recordings from where dynamics can be learnt. This poses a challenge to train a deep neural networks (DNNs) to perform power system security assessment tasks that require to recognize, predict or classify power system dynamics with high accuracy Narasimham Arava and Vanfretti (2018). Thus, the hybrid solution of using synthetic and real measurements mixture has recently been explored for this purpose Zhu and Hill (2022), Bogodorova et al. (2020); Dorado-Rojas et al. (2023). Such synthetic data is generated using extensive simulation physics-based power system models, which are challenging to maintain validated Podlaski et al. (2022). In this context, data augmentation arises as an attractive technique, to enlarge a pool of the relevant data for the DNN training, which has proven successful in computer vision Nie et al. (2021), Iglesias et al. (2023).

Ultra-fast performance of DNNs when deployed needs to be balanced against the training effort for the NN to perform well in real-world applications. Therefore, to speed up the process of learning and recycle the trained networks, transfer learning has been proposed. This concept suggests to recycle the trained DNNs when the data has a similar distribution, even though the NNs might be employed for slightly different tasks. This approach has been successfully applied in the image recognition and classification fields, for example, in classification of medical data, while in power system analysis it has been proposed to address event identification and dynamic security assessment tasks Li et al. (2022), Ren and Xu (2020).

For power system small signal stability assessment (SSSA) the oscillatory pattern within the data can be used to determine the damping ratios of different modes Osipov et al. (2023) and, therefore, the operational state of the system. To this end, data preprocessing methods Vanfretti et al. (2015) are applied by grid operators' tools to detect unwanted behavior.

In this work we propose to benefit from such processing step to enlarge the measurements of interest, use it to train the DNN and to direct the machine learning model to the features in the data that are important to learn. In addition we suggest to use the trained models by efficiently recycling them, i.e. use their parameters and their fixed structure for the repeated training process when new data arrive. Moreover, we evaluate if the preprocessing of the new collected data has a significant impact on the DNNs output. To this end, the variational mode decomposition (VMD) technique, that decouples the signal into meaningful modes, is exploited to augment data that is preprocessed via detrending. The verification methodology that is applied to check if the augmentation is successful is "Train on Synthetic, Test on Real Data" described in Esteban et al. (2017).

The reminder of this paper is organized as follows. Section 2 formulates a hypothesis. Section 3 presents the data augmentation process, the DNN to be trained and the DNN performance assessment metrics. Section 4 describes the statistical tests for the hypothesis testing. Section 5 presents the evaluation of the proposed hypothesis on the
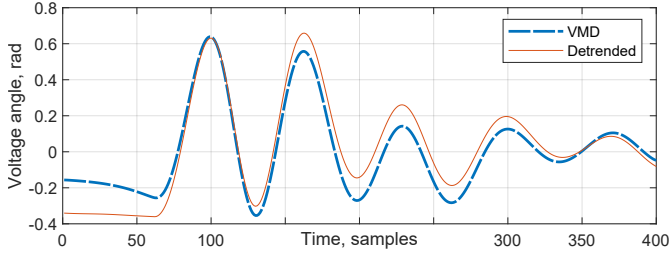
Fig. 1. Examples of input signals

SSSA task for power system data, such as voltage phasor angle. Finally, Section 6 concludes this work.

2. PROBLEM FORMULATION

Variational Mode Decomposition (VMD) is a method that extracts modes from a time-series data that has recently been explored in the analysis of power system modes Osipov et al. (2023). VMD allows to decompose the signal and extract a meaningful pattern in the form of Intrinsic Mode Functions (IMFs). The combination of these functions limits the bandwidth of the original signal, however, filters out the unnecessary information or noise. This ability of VMD may be utilized when training a DNN as a feature extraction step, thus shaping the pattern of the input training data to form another pull of data. Therefore, we wish to validate (or reject) the hypothesis if such data can be used as additional augmented data for training for the tasks where oscillations heavily influence the output of classification such as power system security assessment.

Hypothesis: Can the VMD-decomposed data (e.g. as illustrated in Fig. 1) be used as augmented data when the detrended data are used as input for a deep learning model, if the two distributions of the data are similar enough?

In this case if the hypothesis is validated, the augmented data can be considered as trusted augmentation since the method of data processing is known and unambiguous, which helps with data credibility.

3. DATA GENERATION AND DNN TRAINING

The data to be used to train DNN is processed using detrending or VMD decomposition. When applying VMD, all components are summed up except the last component that contains nonstationary behavior. Detrending is a typical preprocessing step Vanfretti et al. (2015), however, the VMD is used here as an alternative since it has been shown to be effective in oscillation detection applications Osipov et al. (2023).

3.1 Variational Mode Decomposition

In this method the signal is decomposed into band-limited IMFs that are defined as:

$$u_k(t) = A_k(t)cos(\phi_k(t)) \qquad (1)$$

where the phase $\phi_k(t)$ is a non-decreasing function $(\phi'_k(t) \geq 0)$, the envelope is non-negative $A_k(t) \geq 0$, and the instantaneous frequency $\omega_k(t)$ varies much slower than the phase $\phi_k(t)$ Dragomiretskiy and Zosso (2013).

The constrained variational problem is the squared L^2-norm of the gradient: $\partial_t(\cdot)$:

$$\min_{\{u_k\},\{\omega_k\}} \left(\sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right)$$

$$s.t. \sum_k u_k = f \qquad (2)$$

where $\{u_k\} := \{u_1, ..., u_K\}$ and $\{\omega_k\} := \{\omega, ..., \omega_K\}$ are the modes and their center frequencies, $\delta(t)$ is the Dirac distribution, $*$ denotes convolution.

The solution to (2) is obtained by utilizing the augmented Lagrangian multiplier with a series of iterative sub-optimizations known as the alternate direction method of multipliers (see Dragomiretskiy and Zosso (2013)).

3.2 Data Processing using Detrending

Voltage angle measurements naturally contain richer observability of modes that are excited after a contingency than other measurements Vanfretti and Chow (2010). Therefore, voltage angle measurements are used as input data for a deep learning model after pre-processing Vanfretti et al. (2015).

The following data preprocessing steps are performed for the voltage angle signal collected at the location of each electric power generator bus: a) subtraction of the center of angle that is defined as the inertia weighted average of all rotor angles Tavora and Smith (1972); b) unwrapping; c) subtracting the initial value to obtain a deviation signal; d) linear detrending. Let the (raw) input data be defined as $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T]$, where T is the number of time-series signals. For the voltage angle data, the deviations of the signals are given by:

$$\mathbf{x}_t = [\angle\theta_{1,t} - \angle\theta_{1,0}, ..., \angle\theta_{i,t} - \angle\theta_{i,0}, \angle\theta_{N,t} - \angle\theta_{N,0}] \quad (3)$$

where N is the total number of buses, $\angle\theta_{i,t}$ is the voltage angle at bus i of length t, and $\angle\theta_{i,0}$ is the initial value of voltage angle for each bus i.

Then, the angle unwrapping is performed by computing:

$$\angle\theta_{j,i} = \angle\theta_{j,i} + (2\pi k) \text{ if } (\angle\theta_{j,i} - \angle\theta_{j,i-1}) \geq \pi \qquad (4)$$

where j is the sample number in the data set, i is the identifier of a measurement at a particular moment in time, and k is a coefficient that is updated after every large jump in the phase value Venkatasubramanian (2016). Measurements are assumed to be collected at key system locations where Phasor Measurement Units are installed or synthetic data are obtained by simulating a power system model (as in this paper) Vanfretti and Chow (2010).

In the last step of this process, i.e. labeling, VMD is applied and the IMF of the largest energy that contains the dominant modes is extracted. Then, Prony's method Sanchez-Gasca and Chow (1999) is employed to the extracted IMF in order to identify modes. Next, calculate the damping of the mode closest to the critical system mode (i.e. 0.8 Hz is the frequency of the inter-area mode in the examples herein).

In sum, the data generation and labeling approach using simulation models used herein is:

(1) Calculate the initial condition of the power system (i.e. obtain a power flow solution).

(2) Sample the contingency to be applied using realistic contingency generation Bogodorova et al. (2020) and simulate the behavior of the power system.

(3) The measured voltage angles in the buses of interest are pre-processed. The signal is detrended as shown in Fig. 1.

(4) Label the trajectories identifying the state of the system. An example of the VMD decomposition and the detrended signal is shown in Fig. 1

3.3 Offline Training

To validate the formulated hypothesis (see Section 2), an Encoder Serrà et al. (2018), a hybrid DNN architecture that is designed for time-series data classification tasks, is chosen. Encoder (Fig. 2) consists of fully connected layers with an attention layer. Each of three convolutional layers includes respectively 128, 256, 512 filters, with the length of a 1D convolution window of sizes 5, 11, and 21 correspondingly. The operations within the hidden layers are given by:

$$x = Conv_{k_3}(Conv_{k_2}(Conv_{k_1}(x_0)))$$
$$u = ATN(x[:,:,:256] * S(x[:,:,256:])) \qquad (5)$$
$$y = S(IN(\mathbf{W} \otimes \mathbf{u} + \mathbf{b}))$$

where IN is the Instance Normalization operation, k_1, k_2, k_3 is the number of output filters; ATN is the attention mechanism, S is the softmax function. IN normalizes and scales outputs of the previous layer. In contrast to batch normalization, this normalization does not operate on batches, but rather normalizes the activation of a single sample, making it suitable for recurrent neural networks Ulyanov et al. (2016). Meanwhile ATN directs the NN to pay more attention to the small, significant parts of the data. To implement ATN, the data is divided into two equal parts: $x[:,:,:256]$ and $x[:,:,256:]$ (see (5)). The softmax function is then applied to one of the parcels, and the two parcels are multiplied. This allows each element of the softmax-treated parcel to act as a weight for the other one. This mechanism enables the model to learn which parts of the time series are essential for the classification task.

The goal of the deep learning model is to minimize a loss function known as categorical cross entropy. The training set is composed of a collection of values $\{x_{(n)}, y_{(n)}\}_{n=1}^N$. The objective is to find the parameters of the model (e.g. $\mathbf{W}$, $\mathbf{b}$ in equations (5)) that minimize the categorical cross entropy error function L_{CE}, and is given by

$$L_{CE} = \min \sum_{n=1}^N \sum_{c=1}^C y_{c,(n)} log(\hat{y}_{c,(n)}) \qquad (6)$$

where $\hat{y}_{(n)}$ is the classification result of the input values $x_{(n)}$ for the trained model, C is the number of classes, and N is the number of training cases.

To solve (6), the Adam optimizer was chosen with a learning rate equal to 0.00001.

3.4 Online Assessment

For online SSSA, the previously trained Encoder is employed to classify the test cases using voltage angle measurements. With a set of bus voltage angle phasor measurements, it is straightforward to calculate the corresponding

layer output values $h_1, h_2, ..., y_T$ given by (2) using the trained NN parameters. To transform the output of the Encoder, $y_T \in (0, 1)$, into a stability index, a class separation threshold $\delta = 0.5$ is defined so that the test cases with $y_T < \delta$ are considered stable and otherwise unstable.

3.5 Evaluation Metrics

The training of the DNN has been performed using the accuracy metric, while precision and recall were measured on the testing data to validate the resulting model's quality. The class in which unstable cases belong is assigned as a positive class. The choice of positive class greatly influences on which class the precision and recall metrics will be focused on.

Accuracy. This metric defines a general performance of the model across all classes.

$$accuracy = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{FP} + N_{TN} + N_{FN}} \qquad (7)$$

where N_{TP} is the total of unstable cases correctly classified as unstable; N_{TN} is the total of stable cases correctly classified; N_{FP} is the number of stable cases misclassified as unstable; N_{FN} is the number of unstable cases misclassified as stable.

Precision. This metric evaluates accuracy of the model in classifying the data as a positive sample, which is, by our choice, an unstable case. Precision is defined as the ratio between true unstable samples and the total number of samples that are classified by the model as unstable, including those that are false unstable.

Recall. This metric evaluates the number of correct unstable predictions on all relevant unstable predictions.

$$precision = \frac{N_{TP}}{N_{TP} + N_{FP}}, \ recall = \frac{N_{TP}}{N_{TP} + N_{FN}} \qquad (8)$$

In other words, recall evaluates missed correct predictions from the class that is labeled as unstable that is more important to be classified correctly.

Train on Synthetic, Test on Real Data (TSTR) & Train on Real, Test on Synthetic Data (TRTS) Esteban et al. (2017). These are the techniques used to assess whether the augmented data (referred to as "synthetic data" in the names of the methods) are suitable for use as additional training data for the selected NN. For the purposes in this work, the approach can be applied in two ways: when the VMD-decomposed data is used for training but the original detrended data is applied for testing, and vice versa. The resulting evaluation metrics, such as accuracy, recall, and precision, are compared with those received in the original data.

4. DISTRIBUTIONS SIMILARITY KERNEL MAXIMUM MEAN DISCREPANCY TEST

To verify whether the VMD-decomposed data are good to serve as augmented data for NN training, the input data distribution similarity test has to be applied. The Kernel Maximum Mean Discrepancy (KMMD) is one of the recent statistical tests developed to determine if two samples of input data are drawn from different distributions and has
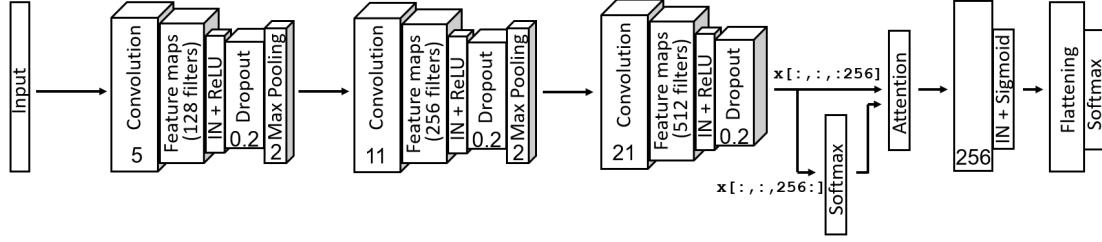
Fig. 2. Encoder neural network architecture

been effective in machine learning applications Gretton et al. (2006).

The maximum mean discrepancy (MMD) is a measure of similarity between two distributions with given observations $X = \{x_1, ..., x_m\}$ and $Y = \{y_1, ..., y_n\}$. It is determined by using a function from the class $\mathcal{F}$. Let $\mathcal{F}$ be a unit ball in a universal reproducing kernel Hilbert space $\mathcal{H}$, defined on the compact metric space $\mathcal{X}$ with associated kernel $k(\cdot, \cdot)$, then $MMD[\mathcal{F}, X, Y]$ is defined as:

$$MMD[\mathcal{F}, X, Y] = \Big[\frac{1}{m^2} \sum_{i,j=1}^{m} k(x_i, x_j) - \frac{2}{mn} \sum_{i,j=1}^{m,n} k(x_i, y_j) + \frac{1}{n^2} \sum_{i,j=1}^{n} k(y_i, y_j)\Big]^{\frac{1}{2}}$$

(9)

where $k(x, x') = \langle \phi(x), \phi(x') \rangle$ - kernel in kernel Hilbert space. The most common kernel function that is applied in this test is the Radial Basis kernel function (Gaussian). The similarity metric $MMD[\mathcal{F}, X, Y] = 0$ if two distributions are equal $X = Y$.

The uniform convergence bound for the empirical MMD (see (9)) that defines a threshold of the hypothesis test is based on the Rademacher complexity Cortes et al. (2013) and presented below under the assumptions that $m = n$, $|k(x, y)| \leq K$, distributions are the same:

$$MMD_b[\mathcal{F}, X, Y] > m^{-\frac{1}{2}} \sqrt{2\mathbf{E}[k(x, x) - k(x, x')]} + \epsilon$$
$$> 2(K/m)^{1/2} + \epsilon$$

(10)

both with probability less than $exp(-\frac{\epsilon^2 m}{4K})$. Thus, a hypothesis test of level α for the null hypothesis $X = Y$ (i.e. $MMD[\mathcal{F}, X, Y] = 0$) has acceptance region $MMD_b[\mathcal{F}, X, Y] < \sqrt{2(K/m)}(1 + \sqrt{2\log\alpha^{-1}})$ according to Corollary 16 in Gretton et al. (2008). The boundary of this region is later named Rademacher. This boundary is relaxed compared to the asymptotic boundary of the unbiased estimate of MMD_u^2. According to Corollary 18 in Gretton et al. (2008), a hypothesis test of level α for the null hypothesis that two distributions are the same $X = Y$ has the acceptance region $MMD_u^2 < (4K/\sqrt{m})\sqrt{\log(\alpha^{-1})}$. The boundary of this region is later named the asymptotic boundary.

5. CASE STUDIES AND ANALYSIS

Several case studies have been developed to test the hypothesis of using VMD-decomposed data as augmented data for SSSA. The KMMD distribution similarity test is applied in two variants of the relaxed Rademacher boundary and the tight asymptotic boundary. The test is validated on the same distribution data that were split into two sets, on the original detrended data and randomly generated, and on the original and the augmented set to test the hypothesis. If the VMD-decomoded data come from a distribution similar to the original detrended voltage angle phasor data, additional DNN performance metrics are computed for the original and the augmented data. These are TSTR and TRTS metrics that are explained in Section 3.5.

5.1 Testing distributions similarity using Kernel Maximum Mean Discrepancy Test

Distribution similarity testing results comparing the detrended and decomposed voltage angle data are presented in Fig. 4 and Table 1. Each data set consists of 7878 samples with each sample length of 400 points with the recording rate 60 samples per second.

The KMMD similarity tests are performed as detailed in Section 4. The MMDs calculated with the bounds are presented in the same color showing the correspondence with the same test result. When the MMD value is smaller than the boundary value, the hypothesis rejection is considered true, otherwise false. The tests are evaluated for different confidence levels that correspond to the significance level α. The meaning of significance level is the probability to reject hypothesis when it is true. In Figure 4 the case of comparison of the distributions of the decomposed data and the randomly generated data is used as a baseline. In this case, the values of MMD or MMD^2 are significantly higher than the relaxed Rademacher and tight asymptotic boundaries, respectively. Thus, the hypothesis that the data originate from the same distribution is rejected.

Another test is performed to validate the proposed methodology on the detrended data that is divided into two parts. Both tests with the relaxed and tight bound have shown that the hypothesis cannot be rejected for all levels of significance. Thus, both case studies reflect the expected performance of the distribution similarity tests.

Finally, the test result on the detrended and VMD-decomposed data using the Rademacher boundary is that the hypothesis cannot be rejected. In other words, the distributions are locally similar enough to consider the data to be of close origin. The VMD-decomposed data can be used as augmented data. However, the test with the asymptotic boundary resulted in the hypothesis being rejected. Therefore, considering the different test results (see Table 1), deep learning model performance validation is performed using the TSTR and TRTS metrics.

Table 1. The hypothesis testing results

Data Set 1	Data Set 2	MMD_b with Rademacker bound	MMD_u with Asymptotic bound
Detrended	Detrended	Not Rejected	Not Rejected
Decomposed	Randomly generated	Rejected	Rejected
Detrended	Decomposed	Not Rejected	Rejected

5.2 Training of the deep learning model Encoder on detrended and decomposed data of the 769-bus power system model

To address the main hypothesis in this work (see Section 2), the metrics TSTR and TRTS (see Section 3.5) are employed. In addition, accuracy, precision and recall are employed to evaluate the Encoder's performance on different training and testing data. The idea is that if the data sets both contain the main features that distinguish the state of the system, and the difference in the distributions is not significant, the resulting performance do not change or change in an acceptable range. The case study has been carried out on 5252 training samples and tested on 2626 samples. The results are summarized in Table 2. The Encoder's training results for the solely detrended or decomposed data are very similar in terms of accuracy and recall, holding the largest difference of 1 % in precision value. The intuition behind this difference is that the ratio between the number of false unstable cases and true unstable cases is larger when the Encoder is trained on the decomposed data. However, for the other two cases that use different data for training and testing, the change in performance of the Encoder is more prominent, especially in precision and recall metrics values. Although the difference between cases with the same data and different data for training and testing in accuracy is around 1 %, the precision differs up to 5 % and the recall is up to 7 %. The biggest drop in performance is for the case where the training is performed on the VMD decomposed data, meaning less rich data than the original detrended data set. This result is logical if the Encoder learns fewer patterns during the training than it is present in the test data.
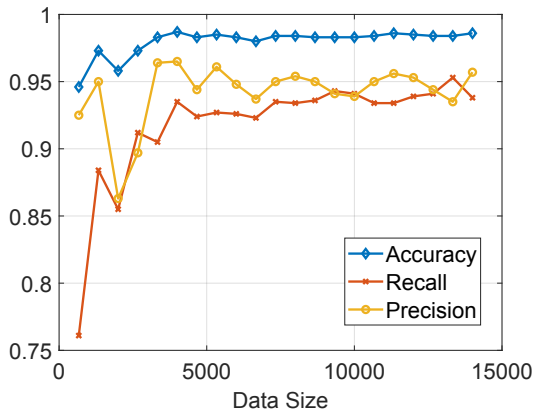

Fig. 3. Performance of the Encoder on data size

Finally, a study on the merged original and augmented data is carried out to observe the performance of the Encoder depending on the size of the data set. In Figure 3 the performance on the joint data (the original detrended and the augmented VMD-decomposed data) has shown a

Table 2. Performance of the Encoder using TSTR and TRTS metrics

	Train: Detrended, Test: Detrended	Train: Decomposed, Test: Decomposed	Train: Detrended, Test: Decomposed	Train: Decomposed, Test: Detrended
Accuracy, %	98.71	98.51	97.33	97.22
Precision, %	97.07	96.03	91.53	92.80
Recall, %	93.62	93.30	89.25	86.70

stable high performance when the data size is larger than 5500 samples. Therefore, even though the data sets are not completely from identical distributions according to the statistical test results in Table 1, the VMD-decomposed data merged with the original data set of the voltage angle data gives as good performance as the base case with the detrended data.

6. CONCLUSION

In this paper we propose to use variational mode decomposition to produce the augmented data for training the deep neural network to perform a small signal assessment for a large power system. By analogy to rotation, cropping, and flipping of an image, the decomposed components of the signal may serve as a means to augment the limited data set with valuable dynamics.

To validate the hypothesis that VMD-decomposed data can serve as augmented data for the neural network, the KMMD statistical test of distribution similarity is performed. The results of the testing have shown that the VMD-decomposed data can be considered as augmented data under relaxed conditions posed by the Rademacher boundary. The additional validation of the proposed augmentation has been performed using TSTR and TRTS metrics.

The outcome has shown good performance of the Encoder neural network used on the newly formed data. The Encoder has demonstrated a consistently positive performance with the increasing data size, which includes the additional "augmented" data.

REFERENCES

Bogodorova, T., Osipov, D., and Vanfretti, L. (2020). Automated design of realistic contingencies for big data generation. *IEEE Transactions on Power Systems*, 35(6), 4968–4971.

Cortes, C., Kloft, M., and Mohri, M. (2013). Learning kernels using local rademacher complexity. *Advances in neural information processing systems*, 26.

Dorado-Rojas, S.A., Fachini, F., Bogodorova, T., Laera, G., de Castro Fernandes, M., and Vanfretti, L. (2023). ModelicaGridData: Massive power system simulation data generation and labeling tool using Modelica and Python. *SoftwareX*, 21, 101258.

Dragomiretskiy, K. and Zosso, D. (2013). Variational mode decomposition. *IEEE transactions on signal processing*, 62(3), 531–544.

Esteban, C., Hyland, S.L., and Rätsch, G. (2017). Real-valued (medical) time series generation with recurrent conditional GANs. *arXiv preprint arXiv:1706.02633*.

Gretton, A., Borgwardt, K., Rasch, M., Schölkopf, B., and Smola, A. (2006). A kernel method for the two-sample-
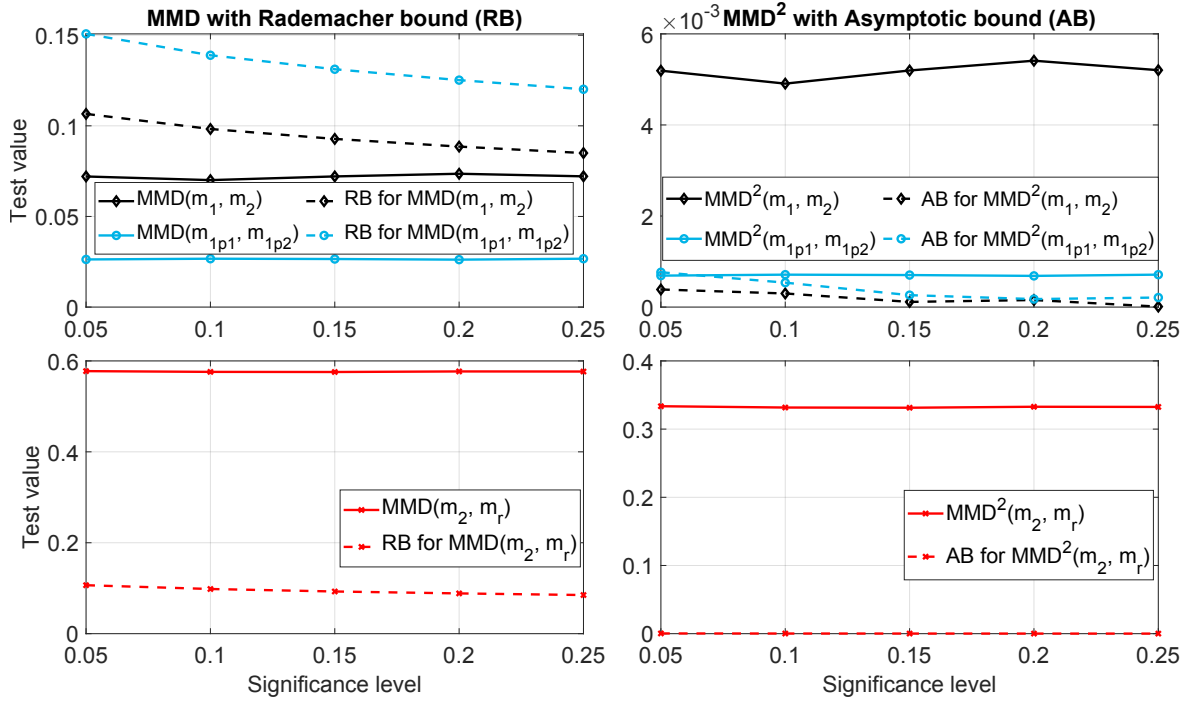
Fig. 4. Testing distributions similarity using Kernel Maximum Mean Discrepancy Test with the Rademacher and Asymptotic bound. m_1 - the decomposed data set, m_2 - the detrended voltage angle phasor data set, m_r - randomly generated data set, m_{1p1}, m_{1p2} - two parts of the same data set m_1

problem. *Advances in neural information processing systems*, 19.

Gretton, A., Borgwardt, K., Rasch, M.J., Scholkopf, B., and Smola, A.J. (2008). A kernel method for the two-sample problem. *arXiv preprint arXiv:0805.2368*.

Iglesias, G., Talavera, E., González-Prieto, Á., Mozo, A., and Gómez-Canaval, S. (2023). Data augmentation techniques in time series domain: a survey and taxonomy. *Neural Computing and Applications*, 35(14), 10123–10145.

Li, H., Ma, Z., and Weng, Y. (2022). A transfer learning framework for power system event identification. *IEEE Transactions on Power Systems*, 37(6), 4424–4435.

Narasimham Arava, V.S. and Vanfretti, L. (2018). Analyzing the Static Security Functions of a Power System Dynamic Security Assessment Toolbox. *International Journal of Electrical Power & Energy Systems*, 101.

Nie, Y., Zamzam, A.S., and Brandt, A. (2021). Resampling and data augmentation for short-term pv output prediction based on an imbalanced sky images dataset using convolutional neural networks. *Solar Energy*, 224, 341–354.

Osipov, D., Konstantinopoulos, S., and Chow, J.H. (2023). A cross-power spectral density method for locating oscillation sources using synchrophasor measurements. *IEEE Transactions on Power Systems*, 38(6), 5526–5534.

Podlaski, M., Bombois, X., and Vanfretti, L. (2022). Validation of power plant models using field data with application to the Mostar hydroelectric plant. *International Journal of Electrical Power & Energy Systems*, 142, 108364.

Ren, C. and Xu, Y. (2020). Transfer learning-based power system online dynamic security assessment: Using one model to assess many unlearned faults. *IEEE Transactions on Power Systems*, 35(1), 821–824.

Sanchez-Gasca, J. and Chow, J. (1999). Performance comparison of three identification methods for the analysis of electromechanical oscillations. *IEEE Transactions on Power Systems*, 14(3), 995–1002.

Serrà, J., Pascual, S., and Karatzoglou, A. (2018). Towards a universal neural network encoder for time series. In *CCIA*, 120–129.

Tavora, C.J. and Smith, O.J. (1972). Characterization of equilibrium and stability in power systems. *IEEE Transactions on Power Apparatus and Systems*, (3), 1127–1130.

Ulyanov, D., Vedaldi, A., and Lempitsky, V. (2016). Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*.

Vanfretti, L., Bengtsson, S., and Gjerde, J.O. (2015). Preprocessing synchronized phasor measurement data for spectral analysis of electromechanical oscillations in the nordic grid. *International Transactions on Electrical Energy Systems*, 25(2), 348–358.

Vanfretti, L. and Chow, J.H. (2010). Analysis of power system oscillations for developing synchrophasor data applications. In *2010 IREP Symposium Bulk Power System Dynamics and Control - VIII*, 1–17. doi: 10.1109/IREP.2010.5563289.

Venkatasubramanian, V. (2016). Real-time strategies for unwrapping of synchrophasor phase angles. *IEEE Transactions on Power Systems*, 31(6), 5033–5041.

Zhu, L. and Hill, D.J. (2022). Data/model jointly driven high-quality case generation for power system dynamic stability assessment. *IEEE Transactions on Industrial Informatics*, 18(8), 5055–5066.