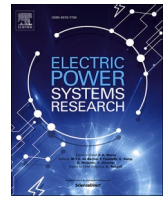# Unsupervised Deep Learning and Analysis of Harmonic Variation Patterns using Big Data from Multiple Locations

(article starts on next page)

Unsupervised deep learning and analysis of harmonic variation patterns using big data from multiple locations

Chenjie Ge [a], Roger A.D. Oliveira [b,*], Irene Y.H. Gu [a], Math H.J. Bollen [b]

[a] *Department of Electrical Eng, Chalmers University of Technology, 412 96 Göteborg, Sweden*
[b] *Electric Power Eng. group, Luleå University of Technology, 931 77 Skellefteå, Sweden*

ARTICLE INFO

ABSTRACT

This paper addresses the issue of automatically seeking and identifying daily, weekly and seasonal patterns in harmonic voltage from measurement data at multiple locations. We propose a novel framework that employs deep autoencoder (DAE) followed by k-mean clustering. The DAE is used for extracting principal features from time series of harmonic voltages. A new strategy is used for training the encoder in DAE from data at one selected location that is effective for subsequent feature extraction from data at multiple locations. To analyze the patterns, several empirical analysis approaches are applied on the clustered principal features, including the distribution of daily patterns over the week and the year, representative waveform sequences of individual classes, and feature maps for visualizing high-dimensional feature space through low-dimensional embedding. The proposed scheme has been tested on a dataset containing harmonic measurements at 10 low-voltage locations in Sweden for the whole year of 2017. Results show distinct principal patterns for most harmonics that can be related to the use of equipment causing harmonic distortion. This information can assist network operators in finding the origin of harmonic distortion and deciding about mitigation actions. The proposed scheme is the first to provide a useful analysis tool and insight for finding and analyzing underlying patterns from harmonic variation data at multiple locations.

1. Introduction

Power-quality monitoring can result in large amounts of data, especially where it concerns harmonics at multiple locations over a long period. For example, one year of monitoring of 39 harmonics and 40 inter-harmonics (3 voltages, 3 currents, 10 min values) results in about 31 million data points per location. In terms of the amount of data, [1] shows that a power quality measurement campaign in five different locations over two years recorded a total of 250 GB of data. Although manual analysis is possible, it is too time-consuming for multiple location measurements. Electric power systems often exhibit distinct harmonic patterns at different locations [2], which means that it is inaccurate to characterize a whole installation by the pattern in a particular location.

Patterns of daily variations in harmonic voltage are compared between locations in [3]. Visual inspection of the daily variations is used to study the propagation of harmonics through the network during resonance conditions. Reference [4] presents data on daily harmonic variations from a number of low-voltage locations. The paper performs manual analysis of the data, using mainly visual inspection of the harmonic currents versus time to identify patterns. No analysis of the voltage was presented. More commonly, statistical analysis is performed when it concerns multiple locations. A classic example of the analysis of harmonics at multiple locations is presented in [5]. The analysis results in statistics over the locations, mainly in the form of probability density functions of the voltage for each harmonic order. The same approach has been used in many later papers and reports, with the main development being the definition of single-side indices to quantify the harmonic distortion per site [6,7]. In [8], statistics on the prevailing phase angle of the harmonics are included, but the overall approach remains the same.

Methods to automatically identify common patterns in large amounts of harmonic voltage data from multiple locations are rare or non-existing in literatures. Machine learning, especially deep learning [9,10,11], provides automatic ways to extract information from a large amount of data. Machine learning tools have been applied to power quality data for the classification and recognition of distinct events [12–29]. Most algorithms are based on supervised learning that requires pre-labeled data, e.g., shallow neural networks [14–19]; and support

* Corresponding author.
E-mail address: roger.oliveira@ltu.se (R.A.D. Oliveira).

vector machines [20–23]. Recently, some deep learning algorithms have been developed for handling power quality data where extracting features is automatic rather than hand-crafted performed. Most deep learning approaches in the literature on power quality applications [24–29] are in the category of supervised learning. Within the literature on event classification, a substantial part of the approaches is based on synthetically generated and often non-realistic data, e.g. [27,28,29]. This limits the practical applicability of those methods, although they still contribute to the knowledge on supervised machine learning methods.

For unsupervised machine learning algorithms, a limited number of applications in power quality measurement analysis has been found, e. g., in [30] a clustering method is used for probabilistic evaluation of harmonic load flow; in [31] k-means clustering is used for identification of distributed generation contribution; in [12,13,14] expert systems are used for classifying power quality disturbances. None of these works has applied deep learning architecture for seeking underlying patterns from variation measurements, despite many works on unsupervised deep learning in other areas especially in computer vision areas, e.g. perceptual grouping using a ladder network for grouping of images [32], and generic adversarial networks (GANs) for image and text synthesis [33,34].

This paper proposes an approach using deep autoencoders (DAE) and feature clustering for seeking possible underlying patterns from long sequences measured from multiple locations in power system networks. The DAE is used for obtaining principal features from time series of harmonic voltages. For the approach being effective to data from multiple locations, we propose a new codebook training strategy where the codebook is trained by the dataset from the best selected location under a given criterion. To analyze the patterns, several empirical analysis approaches are employed after k-means clustering [35] of principal features. This results in a number of graphical presentations of patterns that can be used by the network operator to obtain information about patterns at different locations, without the need for time-consuming manual analysis. The main contributions of this paper include: (a) propose a framework that shows applicability for extracting underlying information and common patterns from harmonic variation measurements; (b) proposing a novel method employing deep autoencoder (DAE) and clustering for extracting principal features grouped by clusters; (c) introduce a criterion for training the DAE that is robust to apply subsequently on data from multiple locations, based on minimizing the distance to the mean cross-correlations; (d) present several empirical analysis approaches for revealing underlying common patterns from data in multiple locations, including feature distribution maps over weekdays and days of the year, and typical waveform sequences of individual patterns. To the best of our knowledge, this is the first successful attempt for automatically seeking harmonic variation patterns from big measurement data in multiple locations through unsupervised deep learning.

The remainder of this paper is structured as follows. Section 2 describes the proposed scheme in detail. Section 3 describes the large dataset and pre-processing used in our experiments. Section 4 shows the results and analysis. Finally, Section 5 concludes this paper.

2. The proposed scheme

This section describes a robust unsupervised deep learning and clustering scheme for seeking common underlying patterns in data sequences of harmonic voltages over a long period of time, from multiple locations. The basic idea behind the proposed scheme is to design a deep learning scheme that has a good generalization performance on extracting underlying intrinsic feature patterns from big data, related to the harmonic variations at multiple network locations in a power system. The scheme, when trained by data measured from one location, is able to find common underlying features using the data from multiple locations for which similarities in electromagnetic environment can be

expected, e.g. locations in the same low-voltage network or in different low-voltage network connected to the same medium-voltage network. To avoid time-consuming manual analysis and because every location has its own characteristics, the measurement data would be unannotated. Hence unsupervised learning and pattern prediction/analysis methods are necessary. The scheme used in this paper consists of two streams, as shown in Fig. 1, where the DAE is further detailed in Fig. 2.

The top stream of Fig. 1 is designed for codebook training, where a deep autoencoder is employed for unsupervised training of codebook coefficients using measurement data from one location. The aim of choosing a DAE is twofold, one is for automatic unsupervised feature learning, and the other is for obtaining principal features. The bottom stream is designed for deep feature extraction and clustering, followed by pattern analysis. For feature extraction, the same structured DAE is utilized, whose encoder coefficients are provided by the DAE trained from the top stream. This provides automatic feature extraction from data in multiple locations. The principal features from DAE are then formed into several classes of patterns through an unsupervised k-mean feature clustering, followed by some harmonic pattern analysis. A list of items for empirical analysis of harmonic patterns is then selected in order to find a meaningful number of patterns as well as the distribution of the daily patterns over the week and the year. In this way, the classification method is able to identify daily, weekly and seasonal patterns.

In the following subsections, details of the proposed scheme will be described.

2.1. Training deep autoencoder using data at one selected location

To find the suitable architecture for unsupervised deep feature learning as well as the coefficients of the deep learning method, we employ a DAE with data at one selected location (described in Section 2.3) as the input. We choose deep autoencoders for obtaining principal features since the input data sequence in our case is considered as a basic or the smallest component in a year-long data sequence. Another advantage of DAE is that it has an exponentially reduced computation of representing some functions, this may also lead to a decreasing number of required training data to learn the functions [11]. More details on the basic theory and method of AEs can be found in [36]. The main aim of training the DAE is to obtain a good structure as well as the coefficients of trained DAE through end-to-end training of encoder and decoder.

The DAE used in the training consists of an encoder and a decoder. The encoder consists of three layers, each layer is fully connected to the previous one followed by ReLU (Rectified Linear Unit) activation function [11] except the last encoding and decoding layers (as the reconstructed data sequence is not constrained by non-negative values). The input to the encoder is the original data sequence (144×1), and the output of the encoder is the feature vector (16×1, determined empirically). The decoder has a reverse structure as the encoder (three layers), where the output dimension in each layer is increased when the layer number increases, and the last decoder layer output is of size (144×1). The detailed architecture is shown in Fig. 3. (where the number of layers and neurons were obtained after many empirical tests, without overfitting or underfitting).

The autoencoder is trained using backpropagation similar to that in conventional neural networks. The loss function L for DAE is defined as the mean squared error between the original data sequence in the encoder input and the reconstructed data sequence in the decoder output, where x_i represents each data sequence, N is the total number of data sequences in one location used for training DAE, h_e and h_d represent the encoder and the decoder.

$$L = \frac{1}{N} \sum_{i=1}^{N} \| h_d(h_e(x_i)) - x_i \|^2 \tag{1}$$

where $h_e(x_i)$ is the encoder output, which is the feature vector $\mathbf{f}_i$, $\mathbf{f}_i = [f_{i,1} \cdots f_{i,m}]^T$ consisting of m components.
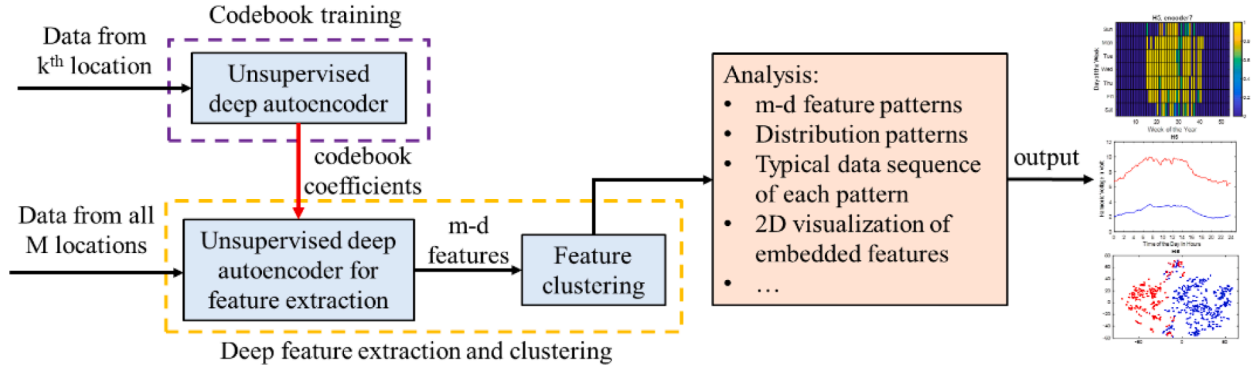
Fig. 1. Block diagram of the proposed scheme for seeking underlying patterns in power system measurements from multiple locations, m-d: multi-dimensional.
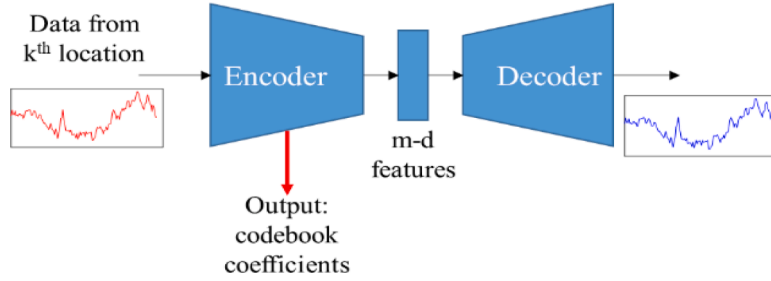

Fig. 2. Unsupervised feature learning by deep autoencoder (DAE).
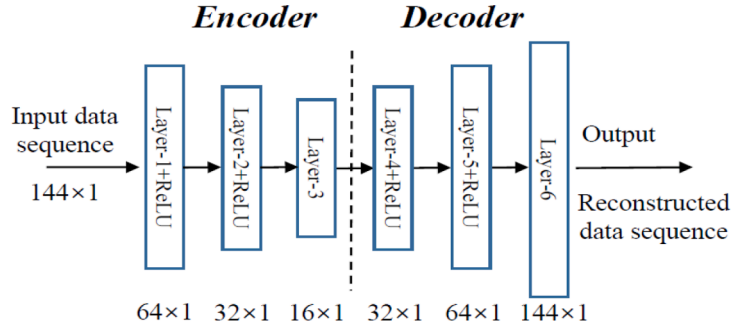

Fig. 3. The architecture of deep autoencoder (DAE) used in the proposed scheme.

2.2. Extracting and clustering features from data at multiple locations using trained DAE

Extracting features: To extract common features from the data measured in multiple locations, the DAE for feature extraction (see the bottom stream, Fig. 1) is set to have the same structure as that in the DAE encoder training (see the top stream, Fig. 1). The coefficients of DAE are fixed and provided from the encoder trained from data at one location (i. e., DAE in the top stream, Fig. 1). In such a way, features from measurement data at different locations are extracted based on the same codebook in the encoder (i.e. under the same m-dimensional space).

Clustering into patterns: Once the features are extracted from all M locations, the next step is to classify these features into several classes, each having a distinct pattern. Since the measurement data sequences are not annotated, unsupervised k-means clustering is employed on feature vectors of data for finding classes of patterns for all locations together. K-means clustering partitions $N \times M$ feature vectors into K clusters $C = \{C_1, C_2, ..., C_k\}$, that minimizes the Euclidean distance, $\arg\min_C \sum_{i=1}^K \sum_{\mathbf{f}_j \in C_i} \|\mathbf{f}_j - \boldsymbol{\mu}_i\|^2$, where $\boldsymbol{\mu}_i$ is the mean of all feature vectors $\mathbf{f}_j$ belonging to the cluster C_i. The k-means clustering consists of

assignment and update steps. Each feature vector $\mathbf{f}_j$ is assigned to a nearest cluster by examining the Euclidean distance between the feature vector and all cluster centroids. Once a new feature vector is added to a cluster, the cluster centroid is then updated by re-calculating its mean value. These steps repeat until the algorithm converges.

2.3. Criterion for selecting a specific location

For training the encoder, a strategy for choosing the best suitable location is adopted. This is done by applying a criterion that is based on the mean cross-correlation. Let $\mathbf{f}_k^{(i,j)}$ be defined as the k-th extracted features from the data in location j ($j = 1, ..., M$, $M = 10$ in our case) where the codebook is trained from data in a selected location i, and $k = 1, \cdots, N$, and N is the total number of feature vectors or training samples in each location. We define the normalized cross-correlation as the feature vectors between all possible locations j and a fixed location i, averaged over all pairs of (j, i), $(j = 1...M, j \neq i)$ and all k, $(k = 1...N)$, as follows:

$$\rho\ (i) = \frac{1}{N(M-1)}\ \sum_{j=1}^{M}\ \sum_{k=1}^{N}\ \frac{\left\langle \mathbf{f}_k^{(i,j)},\ \mathbf{f}_k^{(j,j)} \right\rangle}{\sqrt{\parallel \mathbf{f}_k^{(i,j)} \parallel^2 \parallel \mathbf{f}_k^{(j,j)} \parallel^2}}, \tag{2}$$

$$(j \neq i)$$

where $\langle \cdot,\ \cdot \rangle$ is the inner product of two vectors, and $\mathbf{f}_k^{(i,j)}$ is the feature vector corresponding to data measured in j-th location, however extracted from DAE trained by the data in i-th location. The best location number i^* for training the encoder is chosen as the location that is nearest to the mean of $|\rho\ (j)|$:

$$i^* = \underset{i}{arg}\min\left(|\rho(i)| - \frac{1}{M}\ \sum_{j=1}^{M} |\rho(j)| \right) \tag{3}$$

The rational of choosing the location nearest to the mean cross-correlation value is that the encoder not only takes into account commonly shared patterns (i.e., strong patterns) but also includes some not so strong patterns that only show in some of the locations. In essence, (2) calculates normalized cross-correlation of feature vectors between all locations j ($j \neq i$) to i, where all features of j-th location data are extracted by a deep autoencoder trained by the dataset from a fixed i-th location. Then among all location i, the best autoencoder trained by dataset in location i^* (that can best represent data in all locations), is the one that is closest to the mean of all cross-correlation values.

2.4. Empirical analysis of harmonic patterns

Pattern distribution maps: From the class labels obtained from feature clustering, the distribution of the patterns of feature classes over time can be studied. In this study, the weekly and seasonal changes have been studied, from the daily patterns. To give an overview of the data/features of a given harmonic during a year, a pattern distribution map is utilized by accumulating the number of patterns in different classes for individual weekdays and weeks over the entire year. Such a map can be drawn for each individual harmonic.

Typical data sequences in each cluster/class: To further study the clustered patterns in K classes, one may examine and compare the typical or representative data sequences in several clusters. Since each feature is an m-dimensional vector ($m = 16$ in our case), comparing the typical feature vectors from different clusters in high dimension ($m = 16$) is difficult. Therefore, the feature vector in the centroid of each individual cluster is fed into the decoder part of DAE, which reconstructs a typical data sequence for the corresponding cluster. By comparing and examining typical data sequences in individual clusters, one can find out what are the main differences between different class patterns.

Visualization features by nonlinear embedding of high-dimensional features into 2D space: For visualizing high dimensional features, a commonly used method in machine learning is to nonlinearly embed the features into a low dimensional space by the t-distributed stochastic neighbor embedding (t-SNE) technique [37]. It models two neighboring feature vectors in the high-dimensional space by a Gaussian distribution, and their corresponding ones in a low-dimensional space by a student t-distribution, such that the Kullback–Leibler (KL) divergence measure is minimized [38]. More details for the basics of t-SNE method can be found in [37,38]. There exists a Matlab function *tsne* for realizing t-SNE. To mitigate the sensitivity of *tsne* to initial value/random seed settings, *tsne* is run several times with random initial seeds, where the 2D space associated with the smallest loss is then chosen. It is worth emphasizing that *tsne* is only for 2D visualization, and it does not affect the output from the proposed DAE and high dimensional clustering of patterns.

Pseudo-codes of the Scheme: Table I summarizes the pseudocodes for the proposed scheme.

Table I
Pseudo codes of the algorithm for seeking harmonic variation patterns.

Input: harmonic variation data from all locations (for one selected harmonic number, e.g., H5).
1. *Codebook Training and Deep Feature Extraction*
1.1 For i=1 to all locations {
    1.2 Train a DAE(i) on the measurement data from i-th location;
    1.3 For j=1 to all locations {
    1.4 Extract features $\mathbf{f}_k^{(i,j)}$ from data sequences in j-th location using DAE(i);}}
    1.5 Select the best DAE(i^*) associated with the best location i^* that minimizes (3);
    1.6 Output: Coefficients of encoder and decoder trained from the data in i^* location, feature vectors $\mathbf{f}_k^{(j)} := \mathbf{f}_k^{(i^*,j)}$, $j = 1, \cdots, M,\ k = 1, \cdots, N$;
2. *Clustering*
    2.1 K-means clustering of features vectors (m=16) in all locations;
3. *Empirical Analysis of Harmonic Variation Patterns*
    3.1 Calculate "weekdays vs. weeks of the year" distribution map from clustered m-dimensional features $\mathbf{f}_k^{(j)}$ in all locations;
    3.2 Reconstruct representative data sequences for all pattern classes;
    3.3. Visualize clustered feature points in 2D embedded space
(run *tsne* on features $\mathbf{f}_k^{(j)}$ (m=16) 50 times with random seeds $\in$ [1, 50], select a 2D space with minimum *tsne* error).

3. Application to a large dataset

The pattern-identification scheme, as presented in Section 2, has been applied to the datasets described in Section 3.1. Settings and other details of the application are given in Section 3.2. The results for a number of harmonics are presented in Section 4.

3.1. Dataset and pre-processing

For a given location j, the dataset $D^{(j)}$ consists of 49 sub-datasets corresponding to harmonics 2 to 50, $D^{(j)}=\{D_i^{(j)},\ i = 2,\ \ldots 50\}$, each being a one-year measurement data. There are M=10 locations for each harmonic voltage component (simplified as "harmonic" in the subsequent text). A simple extrapolation was used to fill the missing data sample values (rarely happened) by their nearest previous data value, when the missing samples values were shown as "NAN". Data samples were captured each 10 min for each harmonic, as defined by IEC 61000-4-30 Class A, over the year 2017 at the low-voltage side (immediately after the transformer) of 10 different MV/LV distribution transformers connected to the same Swedish MV network. The MV network consists of a number of feeders that originate in an urban area and continue into rural areas. We then partitioned the one-year long measurement sequence in each harmonic into many 24 h short sequences to form the dataset, where each sequence contains 144 samples, and the time interval Δ_t between 2 consecutive samples is 10 min. In such a way, 365 data sequences were obtained for each frequency component (harmonic) and each location. In such a way, we obtained 490 datasets (49 harmonics x 10 locations), each dataset containing 365 sequences (one sequence per 24 h x 365 days).

3.2. Setup

All experiments were performed on a workstation with Intel-i7 3.40 GHz CPU, 48G RAM and an NVIDIA Titan Xp 12GB GPU. DAE was implemented using KERAS library [39] with TensorFlow backend. In the proposed scheme, the DAE was trained with the following hyper-parameters: optimizer = mini-batch Adam [40] with batch size = 16, the number of epochs = 400, learning rate = 0.001. These hyperparameters were tuned through numerous empirical tests. Input data sequences were normalized to have the maximum value 1.0 for each harmonic from the whole year of data at $M = 10$ locations, before inputting to DAE. Each input feature vector consisted of $m = 16$ components, where m was determined empirically. Similar to other deep learning methods, features extracted from DAE do not associate with clear physical

meaning. Feature normalization was also applied so that each feature component was set to zero-mean and unit-variance on each feature component where features were extracted from data at all 10 locations. For visualizing the feature points, a nonlinear embedding method, t-SNE (Matlab function *tsne*) was used to map $m = 16$ dimensional features into 2 dimensional (2D) ones. The parameters in *tsne* were set as the default values (i.e., use Barnes-Hut algorithm [41], Euclidean distance, perplexity=30, random seed values $\in$ [1,50]). The best 2D embedded space was selected among the results from 50 runs that showed the minimum loss from the *tsne* function, so that the result is less sensitive to the initial seed settings.

A deep autoencoder (DAE) was first applied for training the DAE coefficients using the measurement data from one location. For a given harmonic dataset, the best location for encoder training was decided by using (3). Table II shows the location where the encoder was trained for the given harmonic dataset.

The feature extraction was then applied to the data for 10 locations, using the DAE whose encoder coefficients were obtained from training using data at one location. The output feature vectors ($m = 16$) from all 10 locations were then normalized as zero-mean unit-variance for each component before being fed into the k-means clustering. After many empirical tests, the number of clusters was chosen $K = 2$, as it generated some principal patterns associated with good physical explanations.

Program Time: Table III shows the time required for running the proposed method for training DAE, clustering and seeking principal patterns using the datasets from 10 locations. One can see from Table III that without using 2D embedded map, the algorithm was very fast and only needed about 71 s.

Furthermore, using the existing trained DAEs and two principal patterns, one can easily analyze the patterns of a new test dataset. Giving a new test dataset (consisting of 365 data sequences from one year measurements in a new location), the time required for analyzing the dataset and finding its new pattern distribution map was only about 0.142 s. Since for the new test dataset, the step "clustering" is converted to assigning features to a nearest cluster; the steps for "training DAE on data", "reconstruct data sequences for cluster centroids" and "compute 2D embedded feature map" are no longer needed.

4. Results and analysis

As previously described, one of the main tasks of the proposed method is to seek common underlying patterns from large measurement data, another one is that the method should be robust when applying to data measured from different locations. This section shows some results demonstrating that all these requirements are indeed satisfied. We show that 2 underlying principal patterns from the large harmonic variation data were found, also the method is robust in applying data measured from 10 locations, using the encoder trained by only one location.

4.1. Test results

To illustrate the method we have selected the dominant harmonics, 3, 5, 7 and 9, as well as harmonic 19 as it shows an interesting pattern. The method has been applied to other harmonics as well, but the results are not shown in this paper. Fig. 4 shows the results for harmonics 3, 5, 7, 9 and 19. For each of the harmonics, the figure shows the weekly and seasonal distribution of the patterns (Fig. 4(a)), the reconstructed data sequences for the cluster centers (Fig. 4(b)), and the visualization of 2D

Table II
Location of measurement data used for training the encoder according to (3), for different harmonics.

Harmonic dataset number	H3	H5	H7	H9	H19
Location for training encoder	8	7	4	4	7
$\mid \rho\ (i^*)\mid$ in (3)	0.0338	0.0706	0.0555	0.0413	0.0786

Table III
Time required for training DAE, clustering and seeking principal patterns using the datasets from 10 locations for each harmonic in the proposed scheme (split according to individual parts).

Name	Time (Seconds)
Training DAE on data in 1 location	66.09
Feature extraction for data in 10 locations	0.29
Compute the normalized cross-correlation	0.69
Clustering	0.50
Reconstruct data sequences for cluster centroids	1.96
Compute pattern distribution map	0.96
Compute 2D embedded feature map (50 runs of *tsne*)	1110.02
Total	1180.51

embedded feature space (Fig. 4(c)). The reconstructed data sequences in Fig. 4(b) are typical daily patterns in different clusters, hence are not associated with any particular days.

On weekly and seasonal pattern distribution

Fig. 4 (a) shows the pattern distribution maps indicating the fraction of locations with their daily pattern classed as cluster-1. For all harmonics, cluster-1 corresponds to low distortion (blue curve in Fig. 4(b) and blue dots in Fig. 4(c)) and cluster-2 (red) to high distortion. Thus, in the distribution maps, yellow (or, blue) corresponds to most locations having a high (or, low) distortion day. Fig. 4 (b) shows that there exist significant differences between two representative daily variations in these clusters.

For H3, one can see from Fig. 4(a) that H3 has no obvious patterns over the most time in the year, apart from some pattern for cluster-1 around week 10 and week 30. The number of locations falling into cluster-2 is small. It is reflected by Fig. 4(c) that the two clusters for H3 do not show clear gaps, which further confirms the result of distribution map in H3. A possible reason is the distribution transformers in this network being Dy connected, so H3 voltage remains mainly within the local LV network. Despite this electrical separation, there appears to be similar patterns for different locations.

For H5, the 10 locations have similar behavior. It happens often that the locations are all in cluster-1 or all in cluster-2. That points to a common cause affecting all locations in the same way. Whatever causes the high distortion at all locations, it is present during weekdays from week 15 to 28 and to a somewhat lesser extent week 32 to 40.

For H7, the results show a similar behavior as H5, but there are some differences. During the first period with high distortion, all locations show high distortion at the same days. During the second period, that is the case for only about half of the locations. Contrary to H5, there is another period, later in the year, with high-distortion days affecting all locations again. A possible conclusion is that there are different causes of the high distortion during these three periods. One of them affects H5 and H7 at all locations; one of them affects H5 at all locations and H7 at some locations; one of then affecting only H7 at all locations.

For H9, results show that the low-distortion pattern dominates; there are days when all locations have a low-distortion day, but no days with the high-distortion pattern for all locations. During summer months, the low-distortion pattern is even more likely than during winter, but the difference is small.

For H19, similar observations can be made as for H9, but with somewhat more low-distortion days during weekends and during summer.

The clustered patterns associated with low/high harmonic distortion and the occurrence of the patterns (time of day versus week of year) will assist network operators in finding the major contribution to the harmonic distortion, with the help of additional network information.

On reconstructed data sequences from cluster centers

Fig. 4(b) shows the reconstructed data sequences from the feature vectors (dimension $m = 16$) from the two clustered centroids, by applying the decoder of DAE for the reconstruction. Each of the harmonics shows two distinctive patterns: a "low-distortion pattern"
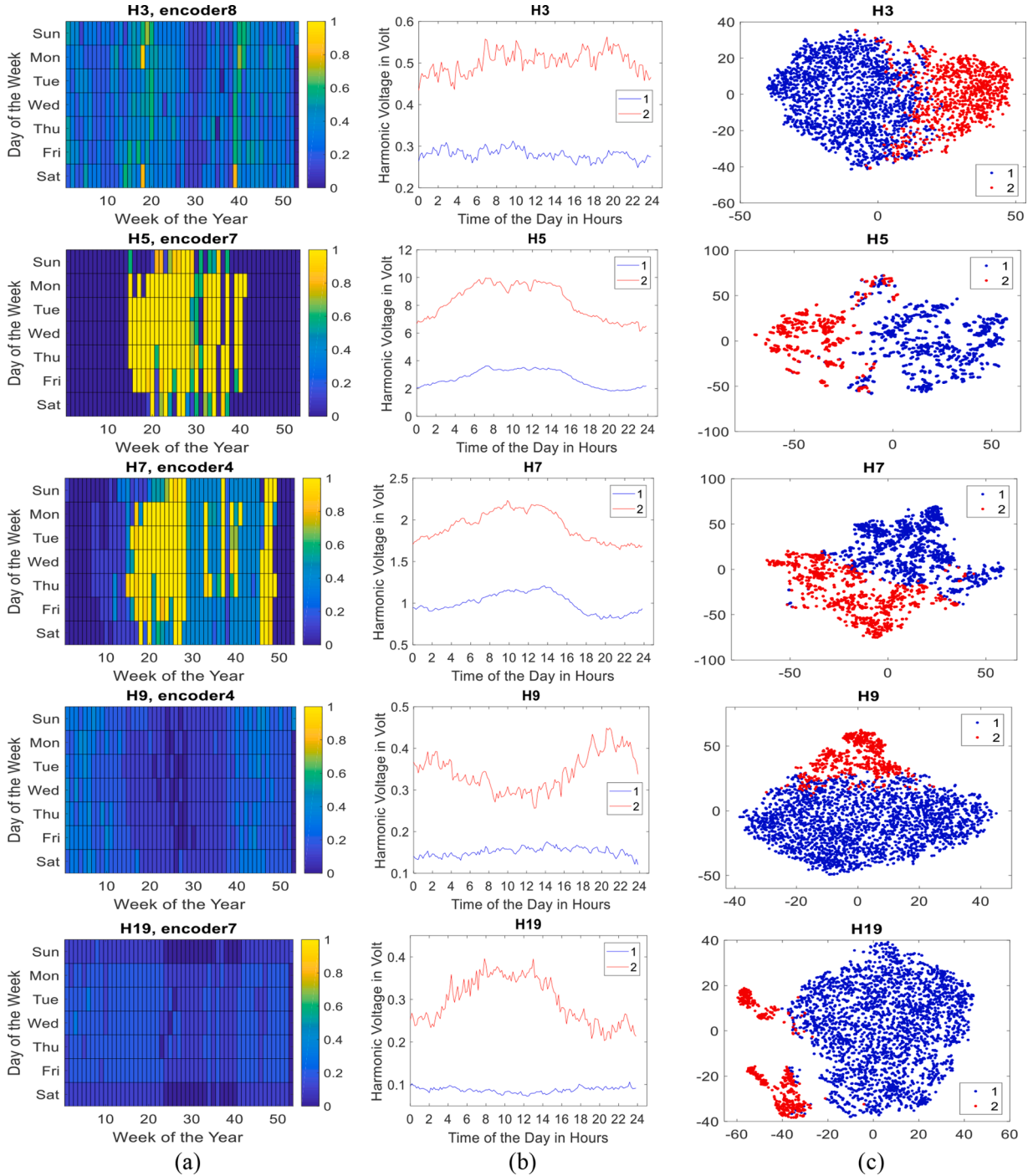
Fig. 4. Analysis of harmonic patterns on measurement data from 10 locations, where the encoder was trained on the data from the location described in Table II. From rows 1-5: harmonics H3, H5, H7, H9, H19. (a): pattern distribution maps, which shows the number of feature vectors in 10 locations falling into individual clusters at a specified time (day of the week vs. weeks of the year), and is described by a color that is defined by the pseudo color bar. In the color bar, the top yellow implies all 10 features are in cluster-2, while the bottom dark blue implies all 10 features are in cluster-1. (b): Reconstructed data sequences from 2 cluster centroids. (c): 2D embedding of high-dimensional features, where maps from MATLAB function *tsne* are associated with (smallest loss, initial random seed) as (2.0379,18), (0.5707, 12), (0.9550, 41), (1.9146, 49). (1.7934, 14), respectively, in 50 runs. Axes in (c) denote two embedded feature dimensions $\tilde{f}_1$ and $\tilde{f}_2$.

(pattern 1, blue) and a "high-distortion pattern" (pattern 2, red). For H3 both patterns are rather constant through the day, with the high-distortion one having a minor increase around 8 am and 8 pm, indicating domestic activity as a contributing factor. For H5 both patterns show a broad maximum around the middle of the day, pointing to office and commercial load as a major contribution. A similar observation can be made for H7. The high-distortion pattern for H9 shows a clear maximum in the evening, indicating domestic activity; for H19 around

the middle of the day, again indicating office or commercial activity. For both H9 and H19, the low-distortion patterns do not show a clear maximum.

On visualization of 2D embedded feature spaces

Observing Fig. 4(c), one can see that low dimensional embedded features of two clusters are reasonably well separated for H5, H7, H9, and H19 in the t-SNE map, while for H3 there is no clear gap between the two clusters in the t-SNE map that is consistent to the result in the

distribution map of Fig. 4(a).

## 4.2. Analysis of feature correlation between different locations

To compare the different spread of feature points from individual locations in a 10 locational feature map, Fig. 5 shows 2D feature maps from 10 locations for H7, and the contributions of feature points from each individual location to the 2D map in Fig. 5(a). This was done by selectively displaying embedded 2D feature map from one location only, while keeping the entire clustering and 2D feature embedding function fixed from that of 10 locations. In other words, by adding 10 embedded feature maps from Fig. 5(b)-(k), one obtains an exact 2D feature map as in Fig. 5(a)). In such a way one may gain more insight when learning the clusters and their centers, by examining the contribution of data from each individual location, e.g., whether or not it significantly deviates from the principal patterns learned from multiple locations. Observing Fig. 5, one can see more details on in the differences in distortion patterns between different locations. For example, high-distortion days (red

dots) are more dominant from locations 1, 2, 3, and 4; the low distortion days (blue dots) are more spread especially for locations 6 -10. Using feature maps associated with individual locations in Fig. 5 provides possible tools to further examine the origin of harmonic distortion.

## 4.3. Using multiple location data, versus using single location data

In this subsection, we compare the results from the proposed method applied on data from multiple locations (M=10), and on data obtained from one single location. For the single location case, the block diagram in Fig. 1 is simplified by merging the top and bottom flows as selection of the best training dataset for the encoder is no longer needed. The results are shown in Fig. 6 and Fig. 7 for H5 and H7, where the first and second row shows the results from data at 10 locations and one location, respectively.

Observing Fig. 6(a), one can see that the pattern distribution map for H5 from 10 locations is very similar to that from one location, though small differences exist. Further, the reconstructed sequences from

(a)　　　　　　(b)　　　　　　(c)

(d)　　　　　　(e)　　　　　　(f)

(g)　　　　　　(h)　　　　　　(i)

(j)　　　　　　(k)
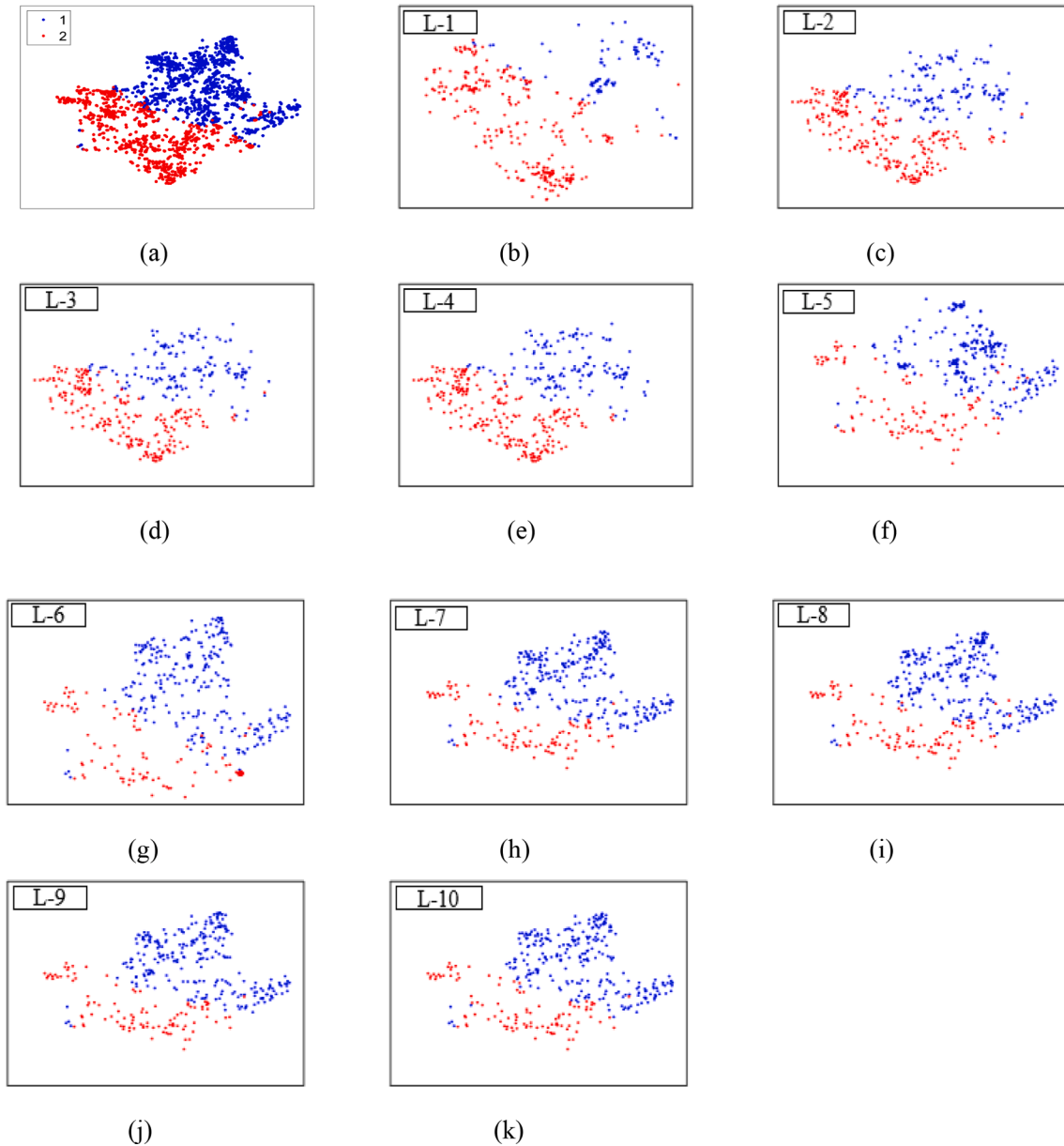
Fig. 5. Comparison of 2D maps of feature points on H7, by using MATLAB function *tsne*. (a) 2D embedded feature points from data in all 10 locations. (b)-(k) contribution of feature points from data in each single location, indicated by L-1 (location-1) to L-10 (location-10) on each sub-plot.
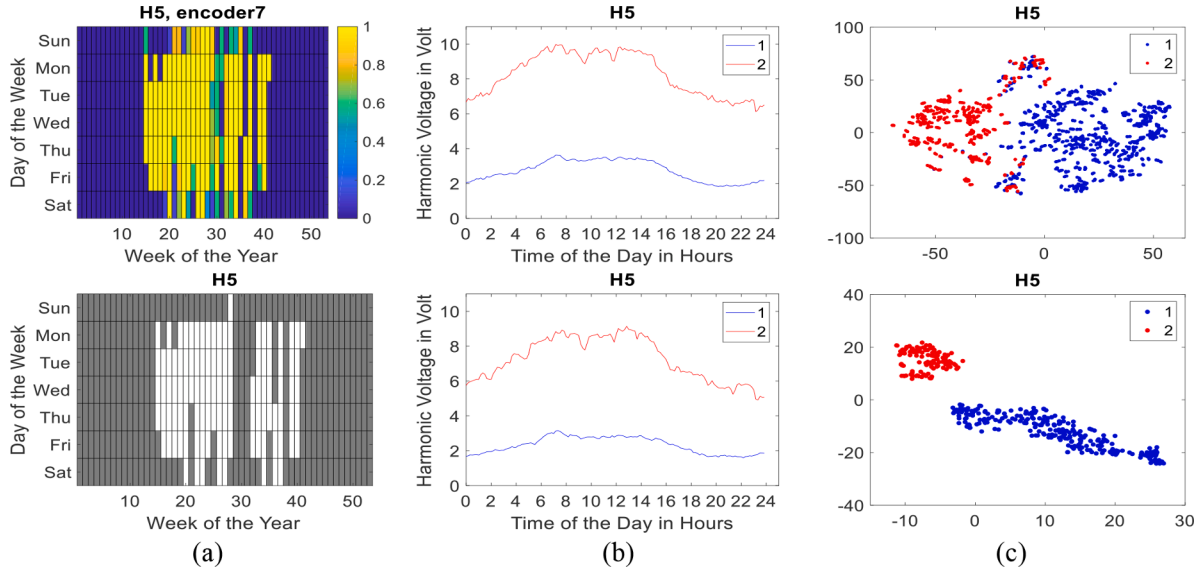
Fig. 6. Comparison of 2 class patterns of H5 from 10 locations (in first row) and one location (location-1) (in 2nd row). From column 1-3: (a) pattern distribution maps of weekdays vs weeks of a year; (b) reconstructed data sequences from 2 cluster centroids; (c) 2D embedded feature maps by using Matlab function *tsne*.
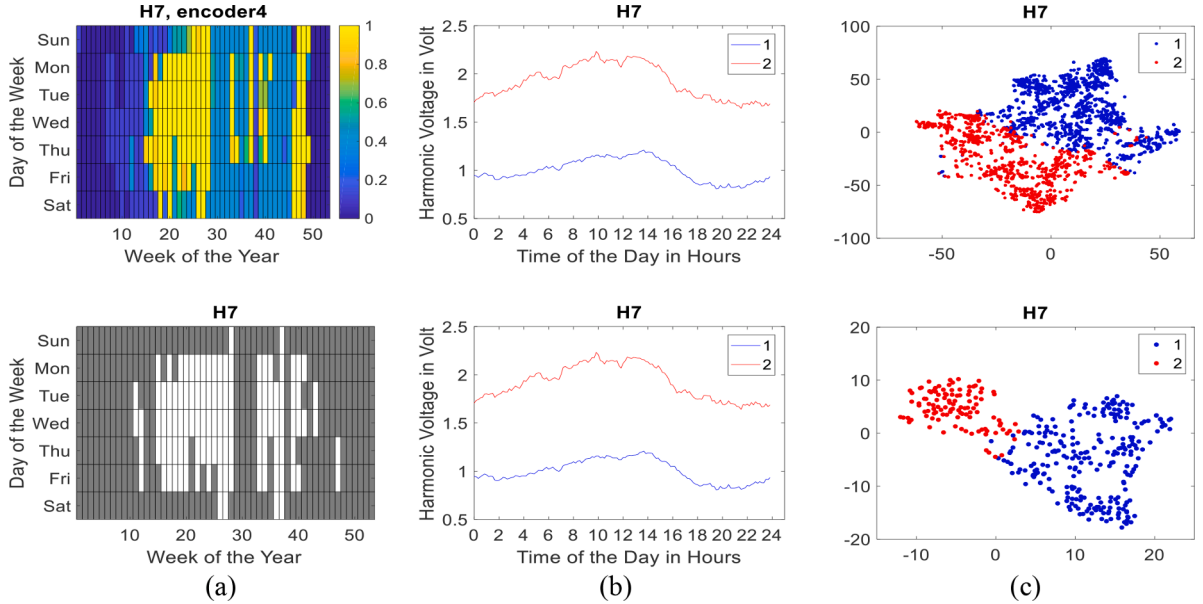
Fig. 7. Comparison of 2 class patterns of H7 from 10 locations (in first row) and one location (location-1) (in 2nd row). From column 1-3: (a) pattern distribution maps of weekdays vs weeks of a year; (b) reconstructed data sequences from 2 cluster centroids; (c) 2D embedded feature maps by using Matlab function *tsne*.

cluster centroids in two cases are largely similar. Overall, the conclusion is that the results obtained from using locational data is largely similar (time of day, time of week, time of year) to that from single locational data for H5. Whatever causes the patterns in H5, they have impacted all locations in a similar way. The origin should be sought somewhere at a higher voltage level.

Observing Fig. 7(a), the pattern distribution maps for H7 from 10 locations are largely similar to that from one location, though some differences exist. For example, pattern differences exist in weeks: V15-28: high distortion everywhere; V33-40 high distortion at location-1, but not in all other locations. V46-48; location-1 does not show high distortion, however, combining 10 locations results in high distortion at most locations.

Observing Fig. 6(c) and Fig. 7(c), one can see that 2D embedded feature maps for 10 locations and one location are somewhat different, despite the overall agreement to the distribution maps as well as the

reconstructive sequences for the 2 clusters in Fig. 6(b) and Fig. 7(b). This can be easily explained: Since t-SNE is a feature map through low-dimensional embedding of a high-dimensional space, it will naturally find different low-dimensional embedding spaces when the input domains are different (by using data from 10 locations vs. one location). In this sense, 2D feature maps by t-SNE, which were originally adopted purely for visual purpose, are not suitable for comparison results from training data with significant size difference. The original high-dimensional features should be considered.

5. Application of the method and results

The aim of the proposed method along with the empirical analysis is to seek whether the common underlying patterns in harmonic variations exist for multiple locations, and how they look like if the patterns exist. The deep-learning based pattern identification method results in

information that is presented in a number of graphical ways. The graphical presentations are chosen such that they can use by persons familiar with power-quality, without requiring deep insight in the machine-learning methods used. The patterns for the cluster centres and the clustering results versus time of week and time of year, can be easily interpreted. The t-SNE plot requires some experience, but even a non-experienced user can get an impression of the separability of different clusters. The graphically-presented information obtained from the proposed scheme can be used as part of an investigation after major contributions to harmonic distortion and after possible mitigation measures. The results from the proposed method will rarely immediately provide the required knowledge, but it will be a first step in the investigation. This first step is done automatically, to save time-consuming manual analysis. Additional steps will likely be rather straightforward methods like plotting harmonic versus time for specific parts of the year, correlations between different locations and between different harmonic orders, again for specific parts of the year. The observed daily, weekly and seasonal patterns should further be combined with knowledge on such patterns in sources of harmonics (e.g. operational hours of industrial installations) and network operational states (e.g. switching of capacitor banks). This requires detailed knowledge on the local condition, which is typically available to the network operator. However, this issue is beyond the scope of this paper.

It is also worth mentioning that the method illustrated here is applied at a distribution grid, but can equally be applied at transmission level. Although the example studied in this paper is on seeking underlying harmonic variation patterns, the proposed method can also be used for seeking patterns from other sets of large power-system data, e.g. power consumption from individual domestic customers, power consumptions at distribution transformer level, or production from renewable sources. The method can also be applied to other power-quality phenomena, like rms voltage, frequency, unbalance, rapid voltage changes, and harmonic currents.

6. Conclusion

A method using deep autoencoder for feature extraction from multiple locations followed by feature clustering for daily harmonic pattern analysis has been proposed and applied to a power system data measured from 10 locations at the low-voltage side (immediately after the transformer) of different MV/LV distribution transformers connected to the same Swedish MV network throughout the year 2017. Results from the experiments have shown that common underlying patterns and information can be obtained by using an encoder trained on data from one location and subsequently applying to seek patterns from data in multiple locations. The unsupervised clustering method has shown to be effective for pattern discovery. Empirical analysis, including the distribution patterns of the original data sequence, the typical data sequence of each pattern, and visualization of features in low-dimensional embedded space by t-SNE, shows the usefulness of the proposed scheme in finding underlying patterns from large measurement data.

The information obtained from the proposed scheme can be used by the network operator as part of an investigation after causes of high levels of harmonic distortion and after possible mitigation measures Although the example studied in this paper is on seeking underlying harmonic variation patterns, the proposed method can also be used for seeking patterns from other sets of large power-system data, e.g. power consumption and production from renewable sources. The method can also be applied to other power-quality phenomena, like rms voltage, frequency and harmonic currents. The results and comparison have demonstrated that the proposed scheme has indeed captured the common typical pattern distribution maps and data sequence patterns from different harmonic variation clusters. For the 2D feature maps, our results show that they can provide useful side information to visualize the separability of features between different clusters. Due to the nature of

low-dimensional embeddings of the high-dimensional feature space, such information should only be used as a reference. Future work will be on deepening the understanding of variation data, as well as towards their modeling.

Authorship contributions

Conception and design of study: All authors

Acquisition of data: Math Bollen

Analysis and/or interpretation of data: All authors

Drafting the first version of the manuscript: Chengjie Ge and Irene Y. H. Gu

Revising the manuscript: All authors

Approval of the version of the manuscript to be published: All authors

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.

Acknowledgement

References

[1] A Mariscotti, Direct measurement of power quality over railway networks with results of a 16.7-Hz network, IEEE Trans Instr. Measurement 60 (5) (2011) 1604–1612.

[2] S Vlahinić, D Brnobić, D Vučetić, Measurement and analysis of harmonic distortion in power distributionsystems, Electric Power Syst. Res. 79 (7) (2009) 1121–1126.

[3] PS Wright, AE Christensen, et al., Multiple-site amplitude and phase measurements of harmonics for analysis of harmonic propagation on Bornholm Island, IEEE Trans. Instr. Measurement 66 (6) (2017) 1176–1183.

[4] W Yang, J Yong, Y Sun, W Xu, D Wong, Characteristics of harmonic distortions in residential distribution systems, IEEE Trans Power Delivery 32 (3) (2016) 1495–1504.

[5] AE Emanuel, JA Orr, D Cyganski, EM Gulachenski, A survey of harmonic voltages and currents at the customer's bus, IEEE Trans. Power Delivery 8 (1) (1993) 411–421.

[6] DD Sabin, DL Brooks, A Sundaram, Indices for assessing harmonic distortion from power quality measurements: definitions and benchmark data, IEEE Trans. Power Delivery 14 (2) (1999) 489–496.

[7] CIGRE WG C4.07, Power Quality Indices and Objectives, 261, CIGRE Technical Brochure, 2004.

[8] J Meyer, AB Castaneda, M Domagk, P Schegner, Assessment of prevailing harmonic current emission in public low voltage networks, IEEE Trans. Power Delivery 32 (2) (2016) 962–970.

[9] CM Bishop, Pattern Recognition and Machine Learning, Springer, Berlin, 2006.

[10] M Mohri, A Rostamizadeh, A Talwalkar, Foundations of Machine Learning, MIT Press, Cambridge, 2012.

[11] I Goodfellow, Y Bengio, A Courville, Deep Learning, MIT Press, Cambridge, 2016.

[12] M B I Reaz, F Choong, M S Sulaiman, F Mohd-Yasin, M Kamada, Expert system for power quality disturbance classifier, IEEE Trans. Power Delivery 22 (3) (2007) 1979–1988.

[13] P K Dash, S Mishra, MA Salama, AC Liew, Classification of power system disturbances using a fuzzy expert system and a Fourier linear combiner, IEEE Trans. Power Delivery 15 (2) (2000) 472–477.

[14] E Styvaktakis, MHJ Bollen, IYH Gu, Expert system for classification and analysis of power system events, IEEE Trans. Power Delivery 17 (2) (2002) 423–428.

[15] S Santoso, EJ Powers, et al., Power quality disturbance waveform recognition using wavelet-based neural classifier. I. Theoretical foundation, IEEE Trans. Power Delivery 15 (1) (2000) 222–228.

[16] J Huang, M Negnevitsky, DT Nguyen, A neural-fuzzy classifier for recognition of power quality disturbances, IEEE Trans. Power Delivery 17 (2) (2002) 609–616.

[17] L Gaing, Wavelet-based neural network for power disturbance recognition and classification, IEEE Trans. Power Delivery 19 (4) (2004) 1560–1568.

[18] S Mishra, CN Bhende, BK Panigrahi, Detection and classification of power quality disturbances using S-transform and probabilistic neural network, IEEE Trans. Power Delivery 23 (1) (2007) 280–287.

[19] M Valtierra-Rodriguez, RdJ Romero-Troncoso, et al., Detection and classification of single and combined power quality disturbances using neural networks, IEEE Trans. Indus. Electron. 61 (5) (2013) 2473–2482.

[20] H Erişti, A Uçar, Y Demir, Wavelet-based feature extraction and selection for classification of power system disturbances using support vector machines, Electric Power Syst. Res. 80 (7) (2010) 743–752.

[21] I Mitiche, G Morison, et al., Classification of EMI discharge sources using time–frequency features and multi-class support vector machine, Electric Power Syst. Res. 163 (2018) 261–269. Part A.

[22] PGV Axelberg, IYH Gu, MHJ Bollen, Support vector machine for classification of voltage disturbances, IEEE Trans. Power Delivery 22 (3) (2007) 1297–1303.

[23] D De Yong, S Bhowmik, F Magnago, An effective power quality classifier using wavelet transform and support vector machines, Expert Syst. Appl. 42 (15-16) (2015) 6075–6081.

[24] H Liu, F Hussain, Y Shen, S Arif, A Nazir, M Abubakar, Complex power quality disturbances classification via curvelet transform and deep learning, Electric Power Syst. Res. 163 (2018) 1–9. Part A.

[25] W Qiu, Q Tang, J Liu, Z Teng, W Yao, Power quality disturbances recognition using modified S transform and parallel stack sparse auto-encoder, Electric Power Syst. Res. 174 (105876) (2019).

[26] I Niazazari, H Livani, A PMU-data-driven disruptive event classification in distribution systems, Electric Power Syst. Res. 157 (2018) 251–260.

[27] H Liu, F Hussain, Y Shen, S Arif, A Nazir, M Abubakar, Complex power quality disturbances classification via curvelet transform and deep learning, Electric Power Syst. Res. 163 (2018) 1–9. Part A.

[28] K Cai, W Cao, L Aarniovuori, et al., Classification of power quality disturbances using wigner-ville distribution and deep convolutional neural networks, IEEE Access 7 (2019) 119099–119109.

[29] A Bagheri, IYH Gu, MHJ Bollen, E Balouji, A robust transform-domain deep convolutional network for voltage dip classification, IEEE Trans. Power Delivery 33 (6) (2018) 2794–2802.

[30] S Galvani, SR Marjani, J Morsali, MA Jirdehi, A new approach for probabilistic harmonic load flow in distribution systems based on data clustering, Electric Power Syst. Res. 176 (2019).

[31] M Jasiński, T Sikoeaki, SK Borkowski, Clustering as a tool to support the assessment of power quality in electrical power networks with distributed generation in the mining industry, Electric Power Syst. Res. 166 (2019).

[32] K Greff, A Rasmus, et al., Tagger: deep unsupervised perceptual grouping, Proc. NIPS (2016) 4484–4492.

[33] I Goodfellow, J Pouget-Abadie, et. al, Generative adversarial nets, Proc. NIPS conf. (2014) 2672–2680.

[34] Y LeCun, Y Bengio, G. Hinton, Deep learning, Nature (2015) 436–444, 521.7553.

[35] D MacKay, Chapter 20. An example inference task: clustering. Information Theory, Inference and Learning Algorithms, Cambridge University Press, 2003, pp. 284–292.

[36] GE Hinton, RR Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.

[37] L Maaten, G Hinton, Visualizing data using t-SNE, J. Machine Learn. Res. 9 (2008) 2579–2605.

[38] S Kullback, Information theory and statistics, Courier Corporation, Massachusetts (1997).

[39] F Chollet, et al. "Keras", [Online] Available: https://keras.io/.

[40] SJ Reddi, S Kale, S Kumar, On the convergence of Adam and beyond, 2019 in preprint arXiv:1904.09237.

[41] L. Maaten, Accelerating t-SNE using tree-based algorithms, J. Machine Learn. Res. 15 (1) (2014) 3221–3245.