Data-driven Feature Description of Heat Wave Effect on Distribution System

This pre-print refers to the paper presented in 2019 IEEE Milan PowerTech, June 23-27, 2019, DOI: 10.1109/PTC.2019.8810712

Data-driven Feature Description of Heat Wave Effect on Distribution System

Yang Zhang, Andrea Mazza, Ettore Bompard
Department of Energy, Politecnico di Torino, Torino, Italy
{yang.zhang, andrea.mazza, ettore.bompard}@polito.it

Emiliano Roggero, Giuliana Galofaro
Ireti SpA, Gruppo IREN, Torino, Italy
{emiliano.roggero,giuliana.galofaro}@ireti.it

***Abstract*—During the last years, the effects of the climate change have become more and more evident. In particular, urban regions, where is more common the use of underground cables, are experiencing the strong effect of extremely high temperature conditions and low humidity. This phenomenon, known in literature as "heat wave", should be properly evaluated for highlighting its effect on the system operation and planning, as well as for properly scheduling appropriate maintenance interventions. This paper presents a three-step procedure aiming to characterize the heat wave phenomenon in terms of "most significant features" and, on this basis, recognizing the days as "critical" and "non-critical". The weather conditions of the city of Turin (Italy) and the faults that have affected the local network in the last 10 years have been considered. This approach will be useful for system operators for integrating the weather information in distribution system operation and planning procedures.**

***Index Terms*--Data analytics, distribution system, heat wave, resilience, kernel density estimation, Gaussian mixture model.**

VI. INTRODUCTION

Currently, about 55% of the entire world's population live in cities and, in 2050, it is expected that this share will increase up to 68% [1]. In particular in Europe, the urban areas are characterized by an electrical infrastructure based on the use of underground cables: the widespread use of cables in the European distribution networks is evident from the available data, which indicated that the share of them in the voltage range falling between 1 kV and 100 kV reached 41% of the entire lines length, and even 55% for lines operated at voltage lower than 1 kV [2]. The data above show how important is the study of the causes that can affect the underground cables (in particular in urban areas), due to the impact that the faults can have on the overall quality of the life of the people living there. This kind of analysis involves new calculation approach, in such a way that the system resilience assessment may be properly done [3][4][5]. Many power grid components, such as the joints of different sections in a cable, are exposed to various weather conditions and at a certain risk [7][5]. In recent years, a significant increase of failures in distribution systems have occurred during summer in different Italian urban areas and was assumed that it could be an effect of the heat wave occurring in the region. This kind of effect was registered in UK on 132 kV system [8], and was evident how the moisture level of the sand, used in filling the digs, could affect the correct operation of the lines. In distribution system, this phenomenon is quite new and, up to now, there is no wider distribution system-monitoring infrastructure for studying its actual effect on the electrical system. A convenient and cost-effective way for approaching this problem is to use the existing weather monitoring system to check the relations between environment and the failures in distribution system [9][10]: in fact, by processing the historical weather records and the number of failures in the same period, it is possible to assess the criticality of weather conditions for the distribution system on a given territory.

Some works existing in literature focus on this issue. For example, an extreme weather stochastic model is combined to a realistic cascading failure simulator in [11]. The fuzzy clustering method is applied in [12] to evaluate the adverse weather effects in power system reliability. The real-time impact of severe weather on the failure probabilities is described by a fragility model of individual components in [13]..

The objective of this study is to build a proper evaluation system aiming to assess the criticality of extreme weather conditions caused by the heat waves which can affect the proper operation of the system. This specific activity was developed in partnership between Politecnico di Torino and IRETI SpA as part of the investigations required by the the Italian Electricity, Gas, and Waste Authority (ARERA) regarding the increase of the resilience of the network[14]. Clustering techniques, as a classic but effective data-driven method for unsupervised learning, has been already successfully applied in weather-related analysis [15][16]. This paper focuses on the three-step statistical analysis of weather conditions in the Turin urban area based on the 10-year records as shown in Figure 9. , which is then compared with the failures recorded in the same period in the urban distribution system of Turin. The first step represents the pre-processing, where the data contained in the two databases (i.e., the weather information and the failures affecting the network) are cleared and filtered. The second step aims to select the most effective features for

978-1-5386-4722-6/19/$31.00 ©2019 IEEE

properly describing the weather conditions, by evaluating the probability density function (PDF) curves derived from the Kernel Density Estimation (KDE). Then, in the third step, a Gaussian Mixture Model (GMM) is used for grouping the days of the different years in clusters based on the features found so far. The output of the procedure is the recognition of critical and non-critical conditions during the different years under analysis. The built model is then applied to the daily weather conditions for a better arrangement of the maintenance priority based on the criticality of the weather conditions. Moreover, the results can be useful for being used in "a posteriori" evaluation for assessing the convenience to make an investment reinforcing the network against the heat wave.

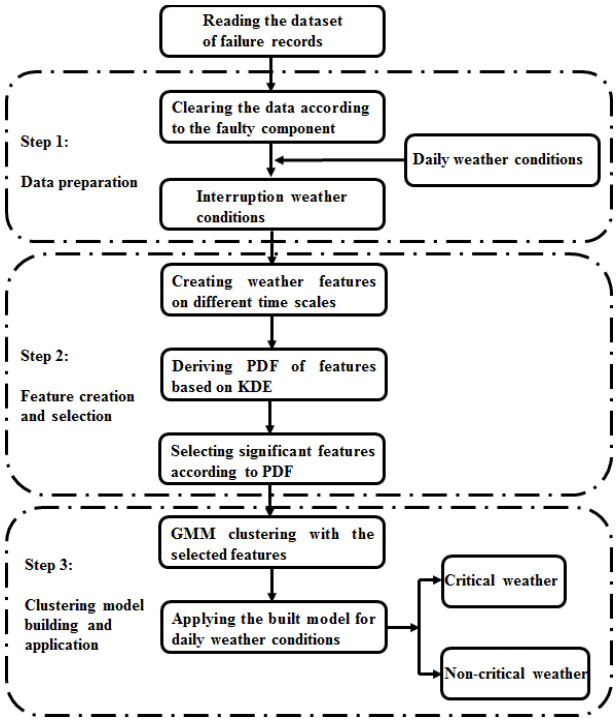


Figure 9. Three-step statistical analysis for weather-related failures in distribution system

VII. Weather Impact on the Failures in Urban Distribution System

In the urban distribution system, most of the feeders are composed of underground cables, whose joints and terminals are usually regarded as fragile components and sensitive to the environmental conditions. According to the local utility, during the last 10 years almost 500 interruptions were related to the failure of the above components in the urban distribution system. This number should be considered as a rough indication, because the total number of faults was much higher, and for many records the indications were generic (e.g., cable) and not indicating the specific part (e.g., conductor, terminals, joints). The annual number of records related to the failure of joints are plotted in the Figure 10. .

As can be seen above, the year of 2017 has a significant increase of failures with respect to the other years. At the same time, an extraordinary summer season characterized of very high temperatures and low rain has been observed in the region of Turin, which leads to think the presence of a strong correlation between the weather conditions and the failure of distribution network.

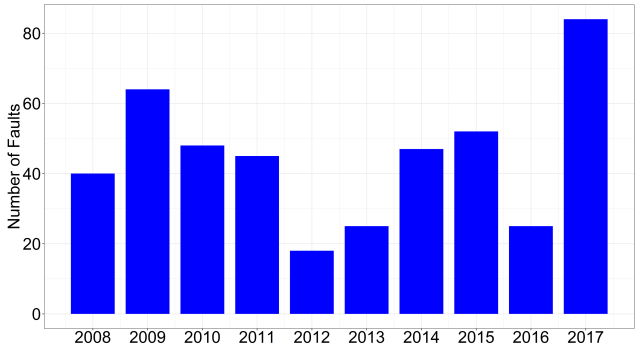


Figure 10. Annual number of failures related to joints or terminals in urban distribution system

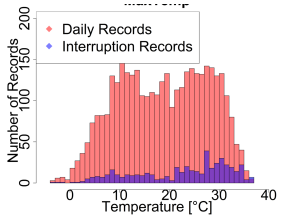
A. *Weather Conditions*

This section aims to describe, in statistical terms, the weather conditions in the last 10 years. In particular, the daily maximum temperature and minimum humidity from 2008 to 2017 are analyzed. The impact of temperature and humidity on terminals and joints may be cumulative, in the sense that a long period of critical conditions can increase the chance of faults. For this reason, the average weather conditions have been defined according to different time windows with respect to the day in which the fault happened, i.e., 5, 10, and 15 days. There are in total 3653 *daily weather condition*s in the records and, among them, those weather records with at least one selected failure are in the following called *interruption weather conditions.*

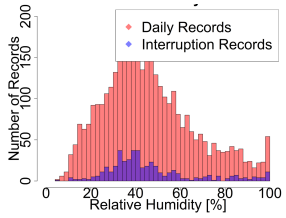
Since several interruptions could happen in the same day, there are some repeated weather records among the interruption weather. Furthermore, if some weather conditions have more impact for an interruption, the density of such cases is supposed to be higher than the other ones, which is a kind of evidence indicating the weather impact. An effective way to demonstrate this assumption is to create the histograms of both daily records and interruption records regarding the maximum temperature and the minimum humidity, as shown in Figure 11.

As can be seen from the range of weather features in Figure 11. , both the distributions of the two weather conditions become denser with more days taken into consideration. For example, the minimum relative humidity in Figure 11. (b) ranges from 5% to 100% while the average value for 15 days ranges only from 20% to 90% in Figure 11. (h). It is because the increase of days averaged the extreme values in a period. In most cases, an extreme weather condition may be not as harmful as a continuous impact in a period. The distribution of minimum humidity of the days when a failure happened is not significantly different from the daily weather records (Figure 11. (b)), while the average value calculated by considering the minimum humidity of the 10 days prior to an interruption is much more skewed compared with the corresponding daily weather records (Figure 11. (f)).

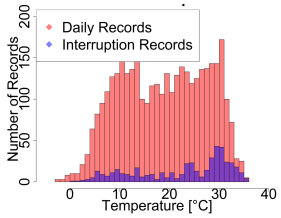
Based on the differences between daily weather features and interruption weather features, the weather records could be used as an indicator for the reliability and resilience of distribution network. As the second step in Figure 9. , the effective weather feature selection is to be introduced in the following part with the help of smooth probability density curves derived from the KDE. Then, on this basis, the weather conditions could be clustered into two groups depending on the selected features.



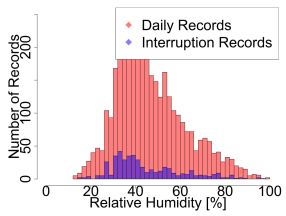
(a) Maximum temperature



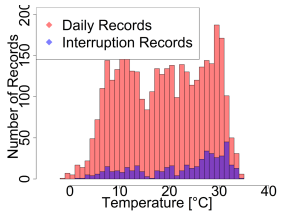
(b) Minimum humidity



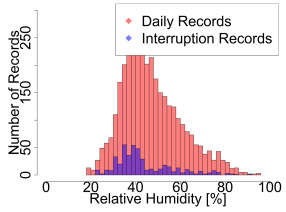
(c) Average value of maximum temperature in 5 days



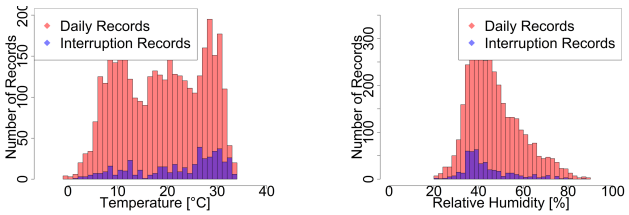
(d) Average value of minimum humidity in 5 days



(e) Average value of maximum temperature in 10 days



(f) Average value of minimum humidity in 10 days



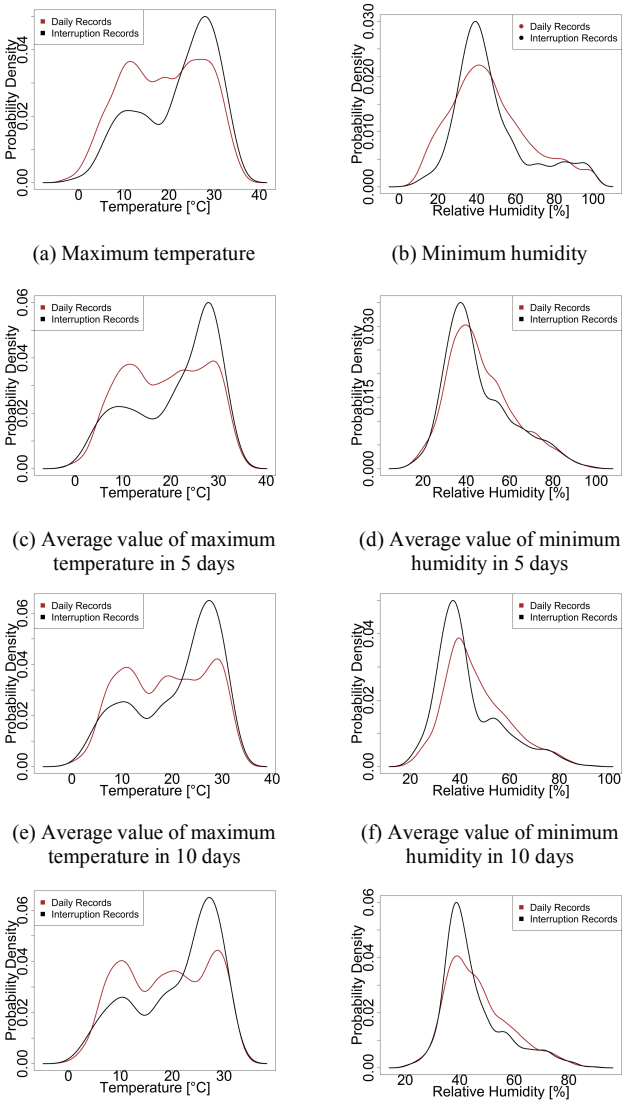
(g) Average value of maximum temperature in 15 days

(h) Average value of minimum humidity in 15 days

Figure 11. Maximum temperature and minimum humidity of both daily and interruption weathers

B. *Kernel Density Estimation*

The KDE is a widely used method in the probability theory as a non-parametric method without priori assumptions about the sample data distribution [17]. Especially in our case, the histograms in Figure 11. are far away from the known classic distributions like Gaussian or binomial distribution. To better display the differences between the distributions of daily weather and interruption weather conditions, the probability density curves derived from KDE are to be used for all the weather features. The KDE method can eliminate the errors in analysis compared to parametric methods and enhance the robustness. In addition, the normalized PDF curves with an enclosed area as 1 are more convenient for comparison. The estimated probability densities are shown in Fig. 4.



(a) Maximum temperature

(b) Minimum humidity

(c) Average value of maximum temperature in 5 days

(d) Average value of minimum humidity in 5 days

(e) Average value of maximum temperature in 10 days

(f) Average value of minimum humidity in 10 days

(g) Average value of maximum temperature in 15 days

(h) Average value of minimum humidity in 15 days

Figure 12. Probability density of maximum temperature and minimum humidity for both daily and interruption weathers

Given N independent variables $x_1, \ldots, x_N$ drawn from the identical distribution, the probability density of this data set estimated by the KDE has the follow expression:

$$\hat{f}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K_0\left(\frac{x - x_i}{h}\right) \tag{1}$$

where $K_0(\cdot)$ is the kernel function and h is the bandwidth. In this paper, the Gaussian kernel function in (2) is applied for the probability density estimation.

$$K_0(t) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right) \tag{2}$$

As can be seen from Figure 12. (b), the distribution of minimum humidity of the day when a failure happened is not far from the daily weather records. In particular, the two peaks correspond almost the same value of relative humidity. However, with a greater number of days for the average of the daily minimum humidity taken into consideration, the curve referring to the failures in distribution network becomes different from the daily weather curve as shown in Figure 12. (f). This indicates that the low relative humidity happened in a limited number of days corresponds to a greater number of interruptions. The more different these two curves are, the more effective this feature is supposed to be. By calculating the overlap areas under the two probability density curves in the figures above, the distortion of two curves could be quantified in an objective way. Because the less value of the overlap area indicates a larger distortion between the two curves. The maximum temperature in Figure 12. (a) and the minimum humidity in Figure 12. (f) have the largest distortion between the two estimated probability density curves, implying that these two features are more crucial to the failure of joints or terminals in cables of the urban distribution system. The following analysis will be implemented based on the selected two features to demonstrate the impact of weather conditions on the failures in distribution network.

VIII. Clustering of Weather Conditions based on Gaussian Mixture Model

As discussed above, the two most effective features are selected according to the "distortions" between the interruption weather and daily weather records. In our study, the weather conditions referring to failure records are used to define the different groups of weather status through clustering techniques. If a kind of weather condition appears more frequently in the failure records, it could be defined as a critical weather condition.

A. Gaussian Mixture Model (GMM)

The GMM is a classic statistic method for multimodal distribution problems and it is widely applied in the engineering problems for clustering, classification and density estimation [18]-[21]. In model-based clustering, the samples are assumed to be generated from a mixture of underlying probability distributions.

Assume a dataset containing n data points: $x_1, ..., x_n$ distributed into G clusters with the label for each sample as: $\gamma_1, \ldots, \gamma_n, \gamma_i \in [1, \ldots, G]$. For the multivariate Gaussian distribution, given the parameters of the cluster k (mean values vector μ_k and covariance matrix Σ_k), the probability density for a sample x_i belonging to a Gaussian distribution cluster k can be written as

$$f_k(x_i \mid \mu_k, \Sigma_k) = \frac{\exp\left\{-\frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1}(x_i - \mu_k)\right\}}{(2\pi)^{\frac{p}{2}}|\Sigma_k|^{\frac{1}{2}}} \tag{3}$$

where p is the number of dimensions of the dataset. With the probability of an observation belonging to the cluster k as τ_k, the likelihood of the mixture model is as follows:

$$L(\mu, \Sigma, \tau \mid x) = \prod_{i=1}^{n}\sum_{k=1}^{G}\tau_k f_k(x_i \mid \mu_k, \Sigma_k) \tag{4}$$

where the probability τ_k satisfies (5):

$$\sum_{k=1}^{G}\tau_k = 1, \quad \tau_i \geq 0 \tag{5}$$

The parameters of the GMM could be estimated by the expectation and maximization (EM) algorithm [22], which is also regarded as the information of clusters.

B. GMM Clustering

In our case, it is a Gaussian distribution modeling with two features corresponding to the maximum temperature and minimum humidity as *MaxTemp* and *MinHumidity10*, in Figure 11. (a) and (f). The PDF curves of these two features indicate that it is appropriate to group the interruption weather conditions into two clusters. Therefore, the number of clusters G=2, with the number of dimensions p=2.

The size and mean value of the two clusters of the interruption weather after the GMM clustering are listed in TABLE I.

TABLE I. SIZE AND MEAN OF CLUSTERS

		Cluster 1	**Cluster 2**
Mean value	***MaxTemp***	14.4°C	29.6°C
	MinHumidity10	51.6%	37.0%
Size	***Number of failure records***	195	253

As shown in TABLE I. , the two clusters are largely different: cluster 1 contains the cool and wet weather conditions, while cluster 2 contains the hot and dry days, as highlighted through the mean value of the two chosen features. Furthermore, the variances of the two Gaussian distributed clusters are largely different as shown in TABLE II.

TABLE II. COVARIANCES OF CLUSTERS

		MaxTemp	***MinHumidity10***
Cluster 1	***MaxTemp***	53.64	-40.55
	MinHumidity10	-40.55	221.42
Cluster 2	***MaxTemp***	13.83	-1.58
	MinHumidity10	-1.58	20.35

According to the results, cluster 2 has a denser distribution compared to cluster 1, showing that the 253 interruption weather records in the second cluster are closer to each other. This is an indicator for a good performance of the clustering results. For cluster 1, the large covariance values imply that the weather conditions are more various than in cluster 2, which means that the failures on joints and terminals of the cables are not as correlated to the weather conditions as in cluster 2. The probability density distribution of two clusters are also shown in Figure 13. . Although cluster 1 covers more area defined by the axes of maximum temperature and minimum humidity, the peak is not as obvious as the one of cluster 2. Taking the failure records into consideration, it is possible to find that almost 30% more interruptions happened under the weather conditions in cluster 2 compared to cluster 1.

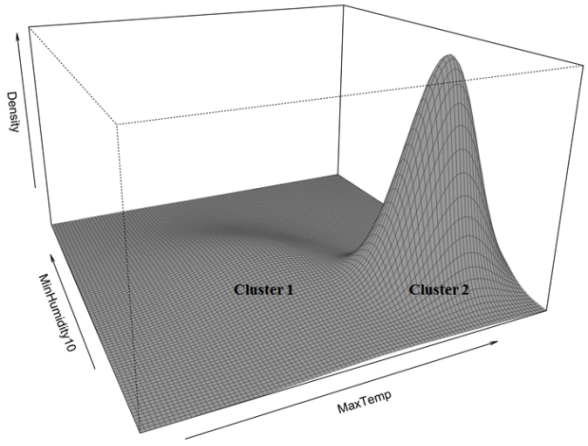


Figure 13. Probability density of two bivariate Gaussian distributed clusters

C. Definition of the critical and non-critical weather conditions

By applying the built clustering model to the 3653 daily weather records, the weather could be clustered into two groups as shown in TABLE III.

TABLE III. CLUSTERS OF DAILY WEATHER RECORDS

	Cluster 1	**Cluster 2**

Number of days	2385	1268
Failure rate	0.082	0.199

As shown in TABLE III. , the size of cluster 1 derived from the daily weather records is much larger than cluster 2: this fact differs from the clustering results of the interruption weather conditions. Even though the critical weather defined in cluster 2 is much less compared to the non-critical weather in real life, more than half of the failures in distribution network happened under this kind of conditions. Specifically, only 34.7% of the days among the total 10 years are in cluster 2, but it contains more than 56% of the failures. This highlights that the weather conditions defined in cluster 2 are crucial for the operation of the power grid. Another information reported in Table III is the failure rate: it has been calculated by dividing the number of interruptions over the total number of days belonging to each cluster. A lower failure rate of 0.082 in cluster 1 compared with the rate of 0.199 in cluster 2 shows the operation of the distribution system is somehow more easily to be affected by the weather condition defined in the cluster 2. In this study, the weather conditions belonging to the cluster 2 have been named as the critical weather and the rest is defined as non-critical weather. Since the critical weather is basically characterized through hot and dry conditions, this could be representative of the phenomenon of the heat wave. To verify this definition, the failure records of three years (2008, 2009 and 2017) are taken for analysis. The year of 2009 and 2017 are the two worst years for the large number of failures while the year of 2008 is a pretty normal year. The number of days belonging to the critical and non-critical weather clusters of the three years can be classified with the above method and the results are calculated and listed in TABLE IV.

TABLE IV. CLUSTERS OF DAILY WEATHER RECORDS

Year	**Cluster 1**	**Cluster 2**	**Number of failures**	λ_{cd}	λ_{ncd}
2008	249	107	40	0.131	0.097
2009	226	139	64	0.245	0.133
2017	209	156	84	0.429	0.086

As shown in TABLE IV. , the number of failures in the three years changes in the similar manner with respect to the number of days in cluster 2, which is defined as days characterized by heat wave. It could be concluded that a longer period of heat wave may cause a larger number of failures in the distribution network, which is in accordance to our previous analysis. The failure rate λ_{cd} and λ_{ncd} are defined as the average number of failures per day under critical and non-critical weather conditions, respectively. As can be seen from TABLE IV. , although the performance of distribution network under non-critical weather conditions in 2017 has not changed much compared to the value in 2008, more failures happen under the critical weather with a significant increase of failure rate λ_{cd}. This increase could be also partially due to the different aging status of distribution system in the two years, as well as to the rise of electrical load in summer due to the high temperatures, which makes the system more and more fragile.

IX. CONCLUSIONS

This paper proposed a data-driven method for evaluating the weather impact on the failures of joints and terminals installed in underground urban distribution system. This strategy contains three steps, including weather and failure records data preparation, feature selection and clustering on the interruption weather conditions. As can be seen from the results, the failures are prone to happen under the critical weather conditions defined by cluster 2, which is also known as “heat wave”.

By combining the weather conditions to the failure records, the histograms of interruption weather and the daily weather records is compared. Instead of directly using the temperature and humidity, more features are created according to different time scales. As a normalized curve with the enclosed area equal to 1, the PDF curves are more easily to be compared than the histograms. In our study, without any a priori assumption of these distributions, the PDF curves of the weather features are derived from the KDE method. Finally, two features *MaxTemp* and *MinHumidity10* are selected as the most representative ones.

Based on the selected two features, a bivariate Gaussian mixture model are then built for clustering. The interruption weather features are divided into two different Gaussian distributions, whose parameters can be determined with the EM algorithm. Each Gaussian distribution represents one cluster. Since there are no predefined labels for the weather severity of every day, all the daily weather conditions are supposed to be evaluated with the Gaussian distributions derived before and get their indications.

Finally, the failure rate between the critical and non-critical years have been studied. There is a quite large difference between them, and this can be used as input datum for evaluating the resilience of the network under critical weather conditions, by integrating it into a cost-benefit analysis for investment evaluation. However, weather conditions are only part of the causes of the interruptions. For this reason, a more comprehensive research needs to consider the aging conditions of components in the distribution network and the loading status of the feeders, which will be the aim of the future work.

REFERENCES

[1] Department of Economic and Social Affairs, United Nations, “2018 Revision of World Urbanization Prospects”. [online]. Available: https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html

[2] EURELECTRIC, “Power Distribution in Europe Facts and Figures”. [online]. Available: https://www.eurelectric.org/news/eurelectric-publishes-facts-and-figures-on-power-distribution-in-europe

[3] A. Mazza, G. Chicco and M. Rubino, “Multi-objective distribution system optimization assisted by analytic hierarchy process”, International Energy Conference and Exhibition, ENERGYCON 2012
[4] R. Rocchetta, E. Zio and E. Patelli, “A Power-flow Emulator Approach for Resilience Assessment of Repairable Power Grids subject to Weather-induced Failures and Data Deficiency”, Applied Energy, vol. 210, pp. 339-350, 2018.
[5] W. Z. Wu, Y. Zhang, J. Li, “Compound sequence network model for steady state analysis of three-phase induction generator supplying single-phase load”, Transactions of China Electrotechnical Society, vol. 31, no. 9, pp. 156-161, 2016.
[6] S. Rubino, A. Mazza, G. Chicco, et al, “Advanced control of inverter-interfaced generation behaving as a virtual synchronous generator”, IEEE Conference on Powertech, Eindhoven, 2015.
[7] B. Ruszczak and M. Tomaszewski, “Extreme Value Analysis of Wet Snow Loads on Power Lines”, IEEE Trans. Power System, vol. 30, pp. 457-462, 2015.
[8] A. Arman, D. Cherry and L. Gosland, et al. “Influence of soil-moisture migration on power rating of cables in h.v. transmission systems”, Proceedings of the Institution of Electrical Engineers, Vol. 111, pp. 1000-1016, 1964.
[9] M. Reder, N. Y. Yurusen, J. J. Melero, “Data-Driven Learning Framework for Assessing Weather Conditions and Wind Turbine Failures”, Reliability Engineering and System Safety, vol. 169, pp. 554-569, 2018.
[10] M. Ervik, S. Fikke, "Development of a mathematical model to estimate ice loading on transmission lines by use of general climatological data", IEEE Trans. Power App. Syst., vol. PAS-101, no. 6, pp. 1497-1503, 1982.
[11] P. M. Quevedo, J. Contreras, A, Mazza, et al, “Reliability assessment of microgrids with local and mobile generation, time-dependent profiles and intraday reconfiguration”, *IEEE Trans. Indus. Appl*., vol. 54, no. 1, pp. 61-72, 2018.
[12] Y. Liu and C. Singh, “Nodal Reliability Evaluation of Impact of Hurricanes on Transmission and Distribution Systems”, Joint International Conference on Power Electronics, Drives and Energy Systems & 2010 Power India, New Delhi, India, Dec. 2010.
[13] M. Panteli, C. Pickering and S. Wilkinson, et al., “Power System Resilience to Extreme Weather: Fragility Model, Probability Impact Assessment and Adaptation Measures”, IEEE Trans. Power System, vol. 32, pp. 3747-3757, 2017.
[14] ARERA, “Incremento della resilienza delle reti di trasmissione e distribuzione dell’energia elettrica: Attività svolte e ulteriori orientamenti”, Discussion paper 645/2017/R/EEL, 2017.
[15] H. T. Ling and K. P. Zhu, "Predicting Precipitation Events using Gaussian Mixture Model", Journal of Data Analysis and Information Processing, vol. 5, pp. 131-139, 2017.
[16] A. Arroyo, A. Herrero and V. Tricio, et al., “Analysis of Meteorological Conditions in Spain by Means of Clustering Techniques”, Journal of Applied Logic, vol. 24, pp. 76-89, 2017.
[17] X. Yang, X. Ma, N. Kang, et al. “Probability Interval Prediction of Wind Power Based on KDE Method with Rough Sets and Weighted Markov Chain”, IEEE Access, 2018.
[18] M. Cui, C. Feng, Z. Wang, et al. “Statistical Representation of Wind Power Ramps Using A Generalized Gaussian Mixture Model”, IEEE Trans. Sustainable Energy, vol. 9, no. 1, pp. 261-272, 2018.
[19] H. Hino, H. Shen, N. Murata, et al. “A Versatile Clustering Method for Electricity Consumption Pattern Analysis in Households”, IEEE Trans. Smart Grid, vol. 4, no. 2, pp. 1048-1057, 2013.
[20] F. Ge, Y. Ju, Z. Qi, et al. “Parameter Estimation of a Gaussian Mixture Model for Wind Power Forecast Error by Riemann L-BFGS Optimization”, IEEE Access, vol. 6, pp. 38892-38899.
[21] L. Scrucca, M. Fop, T. B. Murphy, et al. “Mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models”, The R Journal, vol. 8, pp. 289-317, 2016.
[22] C. Fraley and A. E. Raftery, “How Many Clusters? Which Clustering Method? Answers via Model-based Cluster Analysis”, The Computer Journal, vol. 41, no. 8, pp. 578-588, 1998.