August 27-31, 2023
FLORENCE - ITALY
ICWE 16
16th International Conference on Wind Engineering

Machine learning identification of thunderstorm outflow events from anemometric records

Xiao Li [1], Maria Pia Repetto [2]

[1]*University of Genoa, Genoa, Italy, xiao.li@edu.unige.it*
[2]*University of Genoa, Genoa, Italy, repetto@dicca.unige.it*

SUMMARY: (10 pt)
This study proposes a series of machine learning (ML) algorithms for thunderstorm identification utilizing the anemometric records collected at 15 measurement sites from 2012 to 2020. The training dataset of these ML algorithms comprised 103 thunderstorm (TS) outflow events and an equivalent number of non-thunderstorm (NTS) events. In particular, 86 TS-like NTS events such as gust fronts, which can be easily misidentified as TS events by conventional gust factor-based approaches, were deliberately selected for the training data. The ML algorithms were developed using the support-vector machine technique and their performances were evaluated by k-fold cross validation. Results show that the ML algorithms had a significantly higher accuracy for TS identification than the conventional approaches, and this is attributed mainly to their superior abilities to distinguish the TS events from TS-like NTS events. This study aims to investigate the feasibility of using ML as an effective tool to identify the TS events from anemometric records, and so as to further the application of ML in wind engineering studies.

Keywords: Machine learning, Thunderstorm, Extreme wind events, Anemometric database

INTRODUCTION

Wind loads for structural designs are typically obtained by analyzing historical anemometric records, especially those of extreme wind events such as thunderstorm (TS) outflows. Since the TS outflows are featured with unique wind profiles and turbulence characteristics that are different from other extreme events at synoptic scales, it is required to conduct separate analyses on TS outflows and non-thunderstorm (NTS) high wind events (Gomes and Vickery, 1978). Efforts have been devoted to identify TS outflows from anemometric records over the past few decades (e.g., Twisdale and Vickery, 1992; Lombardo et al., 2009). Notably, literature pointed out that a NTS event may be misidentified as a TS event if it exhibits a similar time-varying trend of wind speed (i.e., a rapid increase and decrease in speed within a short period of time), resulting in a high false positive rate of the identification results. For instance, Kasperski (2002) noted that the gust fronts, as a type of these TS-like NTS events, can hardly be separated from TS outflows. To this end, several approaches based on the gust factor (GF), a conventional wind parameter that describes the intensity of wind speed variations, were proposed by previous studies (e.g., De Gaetano et al., 2014), whereas the TS identification performances of these approaches still needs to be further validated. In recent years, studies in the meteorology field (e.g., Zhang et al., 2016) developed a few machine learning (ML) algorithms to identify TS events based on a variety of meteorological data (satellite, radar, lightning, precipitation, etc.). However, the ML algorithms proposed based

on anemometric data, which meets the specific need for wind-resistant structural designs (e.g., Arul et al., 2022), are still lacking. Hence, this study developed a series of ML algorithms for TS identification based on anemometric records; a large number of TS-like NTS events were deliberately selected for the training of ML algorithms to enhance their abilities to accurately distinguish TS events from NTS events. The paper is structured as follows: Section 1 introduces the wind monitoring network that collected the anemometric records and the selected records of TS and NTS events; Section 2 presents the development and identification performances of the ML algorithms, followed by the conclusions.

1. ANEMOMETRIC RECORDS

1.1. Wind monitoring network

Established by the University of Genoa, the wind monitoring network started to monitor the wind events in the TS-prone areas of the High Tyrrhenian Sea since 2012 (Solari et al., 2012; Repetto et al., 2018). This study examined over 800,000 hours of wind records collected at 15 anemometer sites, as shown in Figure 1(a), from 2012 to 2020. Each site is instrumented with a bi- or tri-axial ultrasonic anemometer that can measure the wind at a speed of up to 45 m/s, which is adequate for the monitoring of natural extreme wind events such as thunderstorms. The sampling frequency is set as 10 Hz for the purpose of capturing the turbulence characteristics of wind. To measure the undisturbed winds that are not affected by adjacent obstacles, the anemometers are all mounted atop tall structures (e.g., towers and poles) at least 10 m above ground level.

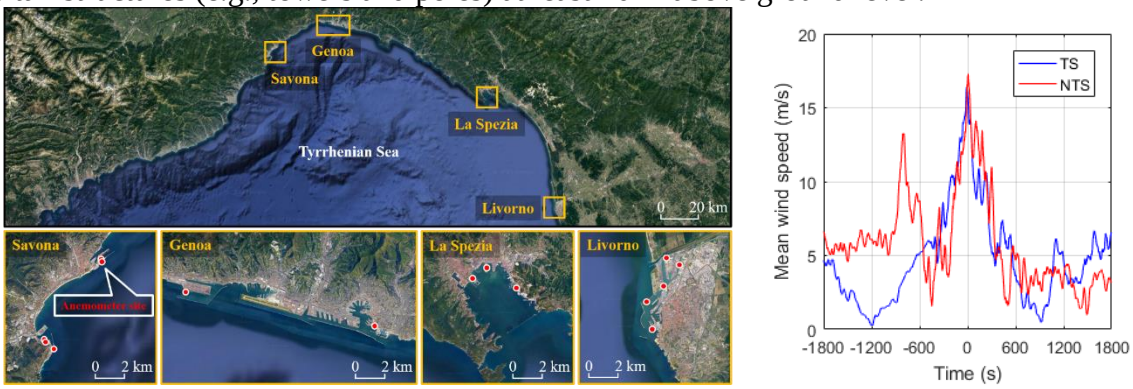

Figure 1. Anemometric database: (a) Locations of anemometer sites; (b) examples of TS and TS-like NTS events

1.2. Thunderstorm and non-thunderstorm records

From the anemometric database of the wind monitoring platform, this study extracted 103 wind speed and direction record segments of TS outflows and an equivalent number of record segments of NTS events. Each of these 206 record segments is one hour in length and centered at the time instance corresponding to the maximum instantaneous wind speed. To ensure the classification accuracy of the selected TS and NTS records, lightning data and infrared satellite images collected during the same periods were carefully examined to detect the existence of lightning activities and severe convections in the vicinity of measurement sites for all TS events (and the nonexistence of these phenomena for NTS events) (Burlando et al. 2018). Notably, with the aim to investigate the difference between TS events and TS-like NTS events such as gust fronts, this study deliberately selected 86 NTS events that were classified as TS events by a conventional GF-based approach (De Gaetano et al., 2014). Figure 1(b) plots the 30-s mean wind speed time histories of a typical

TS event and a TS-like NTS event, showing that the TS-like NTS event demonstrates a very similar time-varying trend of wind speed as the TS event.

2. MACHINE LEARNING IDENTIFICATION

2.1. Feature selection

To develop ML algorithms, the 206 selected anemometric record segments were utilized as the training dataset. It is a widely adopted approach to classify the wind events into TS and NTS categories based on specific wind parameters, and this approach is also followed by the development of ML classification algorithms, in which context the wind parameters and the event categories are referred to as features and labels, respectively. This study obtained a total number of 207 features of the training dataset, including the mean and extreme values of wind speed, turbulence intensity, gust factor, kurtosis and skewness of wind speed, turbulence integral length scale, the variation amplitude of wind direction, etc. To remove the features that are redundant or irrelevant to the labels (i.e., dimensionality reduction), this study conducted neighborhood component analysis (Goldberger, et al., 2004) and the learned feature weights, which quantify the relevancies of features (weights of irrelevant features are zero), are plotted in Figure 2(a). By setting a threshold of 1.2 for the calculated feature weights as presented in, four features with the highest weights were selected for the development of ML classification algorithms.

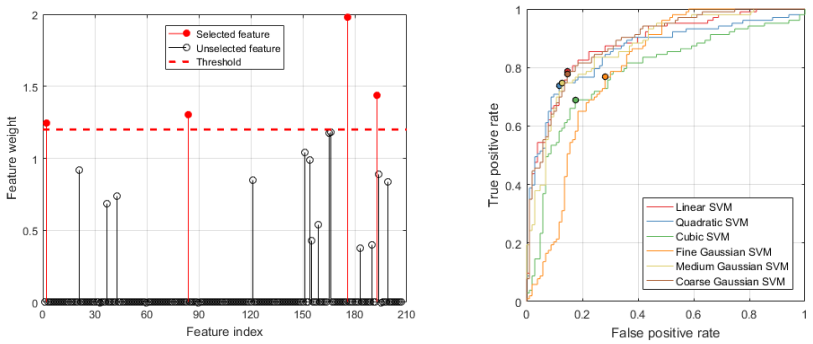

Figure 2. Development of ML algorithms: (a) Feature weights calculated by neighborhood component analysis; (b) Receiver operating characteristic curves and optimal operating points of ML algorithms

2.2. Identification performances

Based on the training dataset introduced above and the four selected features, this study developed six ML classification algorithms using the linear support-vector machine (SVM), first proposed by Vapnik (1963), and non-linear SVMs with polynomial and Gaussian kernels. Figure 2(b) illustrates the optimal operating points on the receiver operating characteristic curves of these SVM-based ML algorithms. Table 1 summarizes the identification performances of the trained ML algorithms, obtained by k-fold cross validation with k = 5, and a conventional GF-based TS identification approach (De Gaetano et al., 2014). Results show that the identification accuracies of the ML algorithms, ranging from 72.8% to 82.0%, were significantly higher than that of the GF-based approach, 55.3%. Such significant improvement in accuracy is mainly because the true negative rates of the ML algorithms reached 68.9% to 85.4%, which were much higher than that of the GF-based approach, 16.5%. These high true negative rates indicate that the ML algorithms, comparing with the GF-based approach, can distinguish the TS events from the TS-like NTS events at a much higher accuracy.

Table 1. Identification performances of SVM ML algorithms and GF-BASED approach.

Identification approach	ACC (%)	TPR (%)	TNR (%)	PPV (%)	NPV (%)
Linear SVM	82.0	81.6	82.5	82.4	81.7
Quadratic SVM	80.1	85.4	74.8	77.2	83.7
Cubic SVM	72.8	76.7	68.9	71.2	74.7
Fine Gaussian SVM	74.3	63.1	85.4	81.3	69.8
Medium Gaussian SVM	78.2	78.6	77.7	77.9	78.4
Coarse Gaussian SVM	81.6	81.6	81.6	81.6	81.6
GF-based	55.3	94.2	16.5	53.0	73.9

Note: ACC – Accuracy; TPR – True positive rate; TNR – True negative rate; PPV – Positive predictive value; NPV – Negative predictive value.

CONCLUSIONS

Based on the anemometric dataset that comprises 103 TS outflow events and an equivalent number of selected NTS events (including 86 TS-like events), this study developed a series of ML algorithms using the SVM technique for TS identification and their performances were evaluated using k-fold cross validation. Results showed that the ML algorithm with the best performance identified the TS events at a satisfying accuracy of 82.0%. In particular, comparing with the conventional gust factor-based approaches, the ML algorithms were able to distinguish the TS events from the TS-like NTS events at a significantly higher accuracy.

ACKNOWLEDGEMENTS

This study is funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (No. 741273) for the project "THUNDERR - Detection, simulation, modelling and loading of thunderstorm outflows to design wind-safer and cost-efficient structures", supported by an Advanced Grant 2016.

REFERENCES

Arul, M., Kareem, A., Burlando, M., and Solari, G., 2022. Machine learning based automated identification of thunderstorms from anemometric records using shapelet transform. Journal of Wind Engineering and Industrial Aerodynamics 220, 104856.

Burlando M., Zhang, S. and Solari G., 2018. Monitoring, cataloguing, and weather scenarios of thunderstorm outflows in the northern Mediterranean. Natural Hazards and Earth System Sciences. 18, 2309–2330.

De Gaetano, P., Repetto, M.P., Repetto, T., Solari, G., 2014. Separation and classification of extreme wind events from anemometric records. Journal of Wind Engineering and Industrial Aerodynamics 126, 132–143.

Goldberger, J., Hinton, G. E., Roweis, S., and Salakhutdinov, R. R., 2004. Neighbourhood components analysis. Advances in neural information processing systems 17.

Gomes, L. and Vickery, B. J., 1978. Extreme wind speeds in mixed wind climates. Journal of Wind Engineering and Industrial Aerodynamics 2(4), 331-344.

Kasperski, M., 2002. A new wind zone map of Germany. Journal of Wind Engineering and Industrial Aerodynamics 90(11), 1271-1287.

Lombardo, F. T., Main, J. A., and Simiu, E., 2009. Automated extraction and classification of thunderstorm and non-thunderstorm wind data for extreme-value analysis. Journal of Wind Engineering and Industrial Aerodynamics 97(3-4), 120-131.

Repetto, M. P., Burlando, M., Solari, G., De Gaetano, P., Pizzo, M., and Tizzi, M., 2018. A web-based GIS platform for the safe management and risk assessment of complex structural and infrastructural systems exposed to wind. Advances in Engineering Software 117, 29-45.

Solari, G., Repetto, M. P., Burlando, M., De Gaetano, P., Pizzo, M., Tizzi, M., and Parodi, M., 2012. The wind forecast for safety management of port areas. Journal of Wind Engineering and Industrial Aerodynamics 104, 266-277.

Twisdale, L. A. and Vickery, P. J., 1992. Research on thunderstorm wind design parameters. Journal of Wind Engineering and Industrial Aerodynamics 41(1-3), 545-556.

Vapnik, V., 1963. Pattern recognition using generalized portrait method. Automation and Remote Control 24, 774–780.

Zhang, Y., Wistar, S., Li, J., Steinberg, M. A., and Wang, J. Z., 2016. Severe thunderstorm detection by visual learning using satellite images. IEEE Transactions on Geoscience and Remote Sensing, 55(2), 1039-1052.