

# Incorporating Wind Modeling into Electric Grid Outage Risk Prediction and Mitigation Solution

Rashid Baembitov<sup>1</sup>, (Graduate Student Member, IEEE),  
Mladen Kezunovic<sup>1</sup>, (Life Fellow, IEEE), Keith A. Brewster<sup>2</sup>, and Zoran Obradovic<sup>3</sup>

<sup>1</sup>Department of Electrical and Computer Engineering Texas A&M University College Station, TX, USA

<sup>2</sup>Center for Analysis and Prediction of Storms University of Oklahoma Norman, OK, USA

<sup>3</sup>Computer and Information Sciences Department Temple University Philadelphia, PA, USA

Corresponding author: Rashid Baembitov (bae\_rashid@tamu.edu).

**ABSTRACT** Electric grids are vulnerable to the impacts of extreme weather. Utility companies face the necessity to reduce the number of power outages caused by weather. This paper expands the approach of predicting weather outages in the distribution grid by incorporating wind modeling. The models for the grid outage State of Risk (SoR) prediction are used by utilities to mitigate potential impacts and reduce outage durations. We study the performance of such models when they are enhanced by incorporating data from Wide Area Fine Grid Wind Modeling (WAFGWM). For a given period, WAFGWM produces wind fields that characterize the direction and speed of the wind over the area of interest. The process of extracting features for the Machine Learning (ML) algorithm is described. The new solution is tested utilizing actual grid performance data from a utility company. The results from nested cross-validation obtained on three years of data reveal that the proposed method improves model performance.

**INDEX TERMS** Machine learning, state of risk, outage prediction, wind modeling

## I. INTRODUCTION

Network outages have been reported to cause significant equipment damage and socio-economic losses [1]. Addressing the problem of reducing the number and duration of outages remains among the top priorities of utility companies. Strong winds may cause short circuits in distribution systems by tree branches being blown into feeders [2]. At the same time, technology advances raise consumer needs and expectations of reliable electrical supply [3]. Given the importance of enhancing grid resilience, the industry focus has shifted to predicting outage occurrences through enabling *a priori* mitigation measures for avoiding or reducing the impacts [4, 5]. This paper incorporates wind modeling into the process of assessing the State of Risk (SoR) of outages caused by impacts of weather and other environmental conditions.

Traditionally, utilities deal with outage impacts *a posteriori*, i.e., restoration actions are taken as a *reactive* measure after an outage occurs [6]. The *proactive* approach where the risk of an outage is predicted in advance has been made possible with the advancements in Geographic Information Systems (GIS) and Machine Learning (ML) algorithms, complemented by Big Data technologies [7-11].

Using a proactive approach, the utilities can reduce the risk with preventive actions such as tree trimming, work crew positioning, customer notifications, and back-up generator utilization. Such mitigation measures are increasing resilience, optimizing capital and operational expenditures while limiting socio-economic impact and increasing the overall satisfaction of utility customers.

Recent research [12-14] describes approaches to estimate the number of outages occurring due to thunderstorms using an optimized linear combination of ML models that include Random Forest (RF), Bayesian Additive Regression Tree, Ensemble regression, and Gradient Boosted Tree. Predicted SoR levels [15] are utilized for assessing the resilience of the distribution grid by adopting a regression model combined with the Naive Bayesian model [16]. The risk of distribution transformer failures due to weather is studied in [17]. Prediction of damages to the grid from extreme weather is discussed in [18]. These studies provide essential information on outage prediction in the grid, but they fall short of considering wind behavior comprehensively. Adding local topographic and vegetation effects to SoR models results in a more accurate wind representation.

We hypothesize that the addition of Wide Area Fine Grid Wind Modeling (WAFGWM) features to these methods can boost the performance of outage SoR prediction, allowing grid operators to make a more focused mitigation plan. Our novel refined wind modeling allows predicting possible wind impacts on the grid (e.g., broken wires, tree limbs contacting wires, and debris blown on feeders).

Our contribution is in improving the SoR prediction model with new features extracted from the WAFGWM process, which allows the selection and implementation of mitigation actions to alleviate the detrimental impact of the forced outages on the customers and utilities. The proposed methodology is evaluated on historical data from a utility company. We illustrate the advantages and shortcomings of our enhanced model by comparing its performance to the metrics of the baseline model.

Section II discusses the problem background. Section III describes the process to enhance the SoR prediction accuracy using WAFGWM. The software tools for wind modeling are discussed in Section IV. Sections V and VI deal with data preparation and feature extraction. Prediction model training and testing are presented in Section VII, and conclusions are drawn in Section VIII.

## II. BACKGROUND

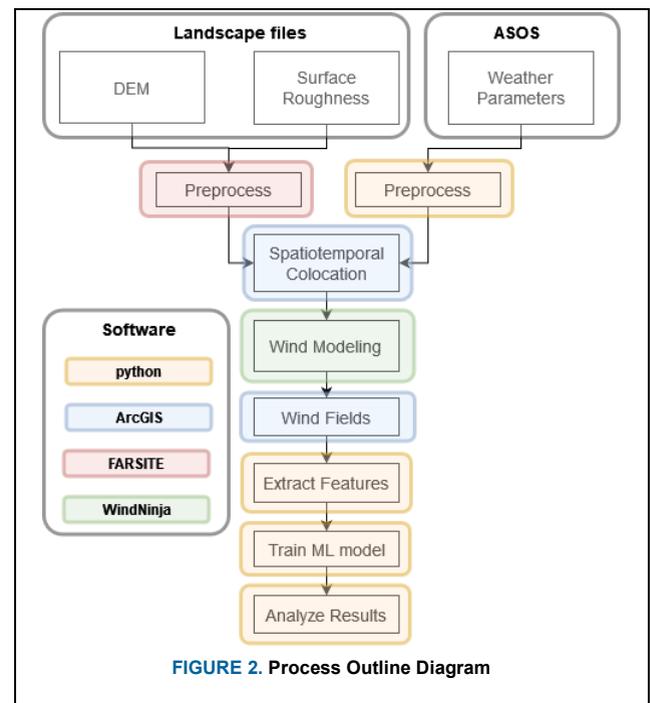
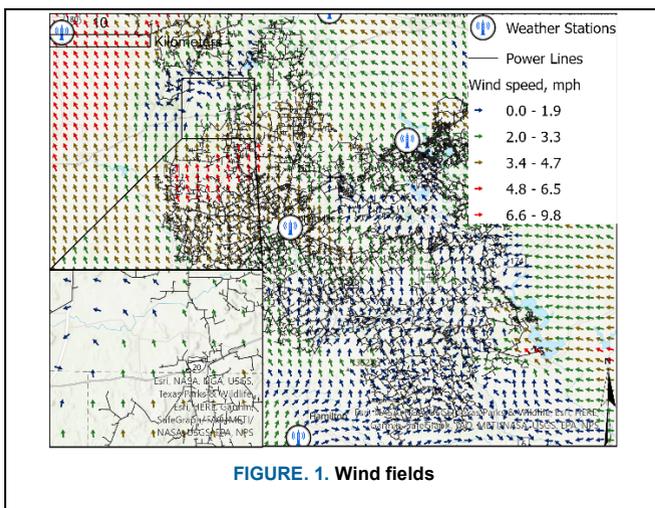
Atmospheric wind analysis and forecast are extensively used in several applications, such as assessing air transport efficiency and safety [19], evaluating air pollution dispersion [20], and understanding the behavior of wildfires and their possible progression scenarios [21]. In the utility industry, 3D wind forecasting is often used to estimate wind power plant efficiency and predict wind power production [22]. In our study, we use atmospheric wind analysis to enhance the performance of the SoR outage prediction model in power grids, which gives the probability of outage occurrence in a part of the network under given weather conditions.

A forecasted wind field example is given in Fig. 1. The wind field is a set of points on a regularized grid that characterizes the horizontal wind vectors at a given height.

Our study uses a height of 2 ft above treetops, near the typical distribution line height. There are many approaches to wind modeling. They range between two extremes: extrapolation and interpolation of available observations and comprehensive numerical solutions based on the Navier-Stokes equations for fluid motion [23]. One type of model is a mass-consistent model, which is relatively fast and attempts to balance the simplicity of the interpolation approach and intricacy of the complete set of equations by meeting the requirement of conservation of mass [24]. Another class comprises linear models, which are designed to keep both: the conservation of mass and conservation of momentum [25], [26]. We are using the first type of model in our study.

## III. WAFGWM PROCESS

We define WAFGWM as a process of calculating wind direction and speed at each small cell in the area of interest with consideration of elevation and underlying type of terrain. The process summarized in Fig. 2 consists of several steps. First, one needs to prepare the datasets necessary for modeling. The data are correlated in time and space, anomalous values are discarded, and the data are fed to the wind modeling software. In our case, we are using WindNinja [27]. It supports several types of wind modeling, each requiring different weather input data. In general, the inputs include Digital Elevation Model (DEM), surface roughness, and 10-m above-ground-level (AGL) wind observations from the Automated Surface Observing Systems (ASOS) [28], as detailed in the next section. The datasets need to be subsetted for the appropriate domain and fetched (downloaded) to a local machine (or cloud-based storage) for easier access. Big Data can become overwhelmingly space-consuming without proper subsetting,



quickly growing over several terabytes [7].

After the datasets are prepared, cleaned, transformed, and wrangled into a suitable format (detailed below), the software may start modeling wind fields over a specified area at specified points in time. That process may be executed manually or can be automated. When the number of modeled cases is low, manual operation through a Graphical User Interface (GUI) is acceptable. However, for the tasks of implementing and training ML models, a large number of cases are needed, so in our study this process is automated.

The next steps are to analyze the resulting wind fields (ensure the quality of output, estimate the total computation time) and transform them into table-like datasets. That allows them to be joined with other features (dimensions) that are already present in the training dataset for the SoR prediction model. Since the outputs of WAFGWM are wind fields valid at several different times, spatiotemporal collocation is performed to match features to the events in the distribution grid. GIS ArcGIS Pro and its python library arcpy software are utilized in our case [29].

The final steps are to merge the new features with the training and test datasets, run model training and testing, and calculate performance metrics. Based on these metrics, one can assess the impact and importance of new features on the model performance.

#### IV. WIND MODELING TOOL SELECTION AND USE

A variety of software packages are available for modeling the wind at distribution line height on a grid from scattered 10-m AGL observations. We have chosen WindNinja (WN), which is developed by the USDA Forest Service, Rocky Mountain Research Station [30], for use by wildfire management crews. WN accounts for our need to include the effect of terrain on the wind flow. Two solvers are available in WN for the calculation of wind fields: conservation of mass solver and conservation of mass and momentum solver. The second solver is generally more accurate, but the computational time required is significantly longer. It was shown that WN has difficulties modeling lee-side flow recirculation during externally forced high wind events [31], which is linked to the absence of a momentum equation in the first mode of modeling [30]. Additionally, point initialization is not available for the mass and momentum solver. In this study, we are using the first solver, mainly due to a large number of cases, since it takes less than one minute for the first solver to calculate one case compared to anywhere between 10-30 min for the second [32].

Required datasets include a DEM file for the modeled area, surface (10-m AGL) wind speed and direction, and specifications for the vegetation type in the area. Several output formats are supported in WN; here, shapefiles (\*.shp) are used.

Another indispensable feature of WN is the command line interface (CLI) that allows one to automate WN runs for several timestamps or different wind conditions. We are

using WindNinja's CLI through a python interface, which allows us to keep the entire process of wind modeling and subsequent ML model training and testing in the same environment.

#### V. DATA PREPARATION

This section discusses the data preparation for each of the input datasets: DEM, Vegetation Type, and Wind Fields.

##### A. DIGITAL ELEVATION MODEL

In our application, we decided to use the FARSITE comprehensive landscape file (\*.lcp) format because it also contains information about vegetation in the area as one of the data layers. Such landscape files are primarily used to model fire behavior. These types of files include information on elevation, slope, aspect, fuel model, canopy cover, and optionally can have stand height, height to live crown base, crown bulk density, duff loading, and coarse woody profiles [33]. We are using LANDFIRE (LF) 2016 Remap [34] products, specifically LCP 40 Fire Behavior Fuel Models-Scott/Burgan. The dataset reflects ground conditions for a time period close to 2016 for which we had records of outages. The spatial resolution of the acquired file is 30 by 30 meters.

Generally, the output cell size for the wind field is selected with consideration of the task at hand. For SoR predictions, we used 900 by 900 meters output resolution, which allowed us to balance the computational load and modeling accuracy. Smaller cell size substantially increases computation time, while bigger cell size yields coarser, less accurate results.

Another way of decreasing computation time is dividing the area into several smaller parts and merging the resulting wind fields afterward. However, that approach may yield inconsistencies in wind field values along the edges of the modeled region parts. In this paper, the problem of SoR prediction is solved with reference to separate feeders in the grid, which are not uniform in length, location, or shape. Thus, it is natural to model the entire area in a single run.

##### B. VEGETATION TYPE DATASET

Vegetation type in the area of modeling determines the surface drag due to vegetation and whether the diurnal flow or non-neutral stability is considered during simulation. It is also used for heat flux parameters by WN. By default, three vegetation types are accessible to the users: grass, brush, and trees. One downside of the WN is that the standard way of specifying the vegetation type is by selecting the dominant vegetation type in the entire area, which leads to inaccurate modeling of wind fields when large areas containing disparate vegetation types are considered.

To overcome that challenge, we used previously mentioned landscape files to specify the underlying terrain. When that approach is utilized, WN is forced to calculate surface drag based on the canopy and fuel information enclosed in the landscape file.

Other software for wind modeling may offer an easier way to define drag over the area. For example, Continuum [35] allows users to use well-known National Land Cover Database (NLCD) land cover GeoTiff files [36] as a source of information for the surface drag.

### C. WIND SPEED AND DIRECTION

Initialization values for wind speed and direction in the area are used as reference points to “fit” the resulting wind field. WN supports three ways of initializing winds in the area: domain average wind, point initialization, and weather model. Point initialization allows one to define wind speeds and direction at several arbitrary points in the area. The software then calculates a balanced wind field that matches the point observations. The calculation process is iterative and is stopped when the simulated values are within 0.1 m/s of the measured values at every observation point. The method is very convenient since one can acquire wind parameters from stationary weather stations that report weather with high time resolution.

We are using the point initialization method in our present study to be able to introduce forecast data in our SoR prediction model. The weather data for the area of interest has been obtained from the ASOS sites by using an API developed by Iowa Environmental Mesonet [37]. A pandas library is used to reformat the weather data for WN from the original ASOS report format. The set of files for one simulation consists of a .csv file for each weather station (WS) in the area, where each line represents one observation at the specified timestamp. Each .csv file needs to incorporate the following information:

- Unique weather station name
- Coordinate System
- Datum
- Latitude of WS
- Longitude of WS
- Height of measured wind above the surrounding vegetation
- Speed of wind
- Direction of wind
- Radius of influence (limits the spatial extent to which each WS can affect calculations)
- Units for each parameter
- Timestamp
- Optional parameters used for diurnal non-neutral stability sub-models are also included:
  - Temperature
  - Cloud Cover in Percent

In addition to the individual .csv file for each location, the summary file, which contains the list of all individual WS files, is created.

### D. CLI CONFIGURATION FILE

To make the SoR prediction process scalable to numerous modeling cases, we have automated WN simulation runs

through the CLI. The configuration file needed to specify the options for the modeling needs several important parameters are:

- *--num\_threads*: defines the number of virtual CPU cores used for simulation.
- *--elevation\_file*: specifies the path to the DEM file
- *--initialization\_method*: select from domain average wind, point initialization, and weather model.
- *--wx\_station\_filename*: path to the summary file with the list of WS files.
- *--start\_year, --start\_month, --start\_day, --start\_hour, --start\_minute, --stop\_year, --stop\_month, --stop\_day, --stop\_hour, --stop\_minute*: time period information. The period must be within the timestamps of the weather observations.

We used a python script to form the list of parameters that change for each run (mainly, the timestamp for each simulation) and then passed them as a command to the CLI of WN.

Once the simulation runs were configured, the simulation process was initiated. Initially, there was a total of 11230 timestamps, among which 5615 had one or more faults in the power distribution system. The time difference between subsequent faults resembles heavy-tailed exponential distribution with a mean of 306.4 minutes, a median of 22 minutes, a mode of 1 minute, and a standard deviation of 995.8 minutes. However, wind fields were not properly produced for every case. Some cases did not converge, and some timestamps did not have all the required weather variables for modeling, so no resulting wind field was obtained for them. The number of successfully modeled cases is 10564, which yields an overall 94% rate of success in modeling.

## VI. FEATURE EXTRACTION FROM WIND FIELDS

This section details the processing steps needed to extract features from wind fields and incorporate them with the rest of the datasets. The resulting wind field discussed in the previous section is represented as a shapefile, which is suitable for processing with ArcGIS Pro GIS software used for spatial analysis.

First, the DefineProjection function is used to set the projection for a shapefile with wind fields. The datum and projected coordinate system (PCS) (WN only works in PCS) is defined by the elevation file that was used for the modeling. In our case – the USA Contiguous Albers Equal Area Conic USGS version [38]. The PCS specifies the geographic location of the wind field with respect to the power distribution grid.

Next, to spatially correlate the feeders in the electric grid with the calculated wind field, the spatial join operation is performed (SpatialJoin). The goal is to match individual points in a wind field (each characterized by speed and direction) to the closest feeder in the network. As a result, each of the wind field points belongs to the closest feeder,

and the set of such points characterizes wind behavior for each feeder. To guarantee that each feeder would have at least one wind point assigned to it, we used the following parameters for spatial join [39]: `join_one_to_one`, `keep_common`, '1273 Meters'. The first specifies that each point must be matched only with one feeder, and the second instructs the program only to consider wind points within a specified distance, which is the last parameter. The spatial extent of 1273 meters is chosen out of consideration for the spatial resolution of the wind fields, which is a rounded length of the hypotenuse of a right-angled triangle with both sides equal to 900 meters.

To summarize the wind behavior for a feeder, we calculated several statistics from wind field values: Mean speed, Max speed, Min speed, Mean direction, Max direction, and Min direction. We exported the resulting attribute table as a .csv file, ingested it by pandas, and calculated the statistical values. In such a way, for each timestamp, each feeder gets six additional wind field features that aggregate wind information over the feeder and adjacent territory.

The described process took 11 hours and 48 minutes to run when utilizing 14 cores of Intel® Core™ i9-9900 CPU with 3.1 GHz and 64 GB of RAM. Statistics could also be calculated by ArcGIS by appropriate functions; however, we found that processing times are longer in that case.

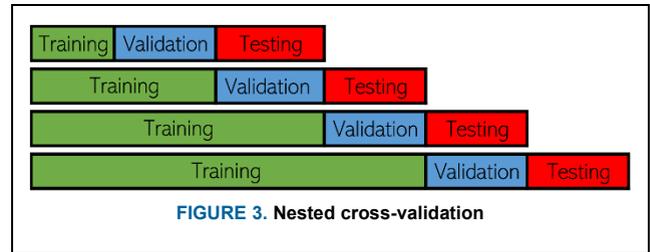
The last step is to augment the existing features with the new wind features. That is done by a simple join operation using the feeder's name and timestamp as join keys. The result is a training dataset enhanced with six additional wind field features. The forming of the former is discussed in the next section.

## VII. ML MODEL TRAINING AND TESTING

### A. BASELINE MODEL

The base model results for predicting SoR levels of outages in the system are considered to be a benchmark that the new model with additional features needs to be compared against. The foundation for the base model parameters comprises the weather variables obtained from ASOS observations. The training dataset uses the following features (dimensions): Air Temperature, Dew Point Temperature, Wind Direction in degrees from true north, Wind Speed, Wind Gust, and One-hour precipitation for the period from the observation time to the time of the previous hourly precipitation reset, Relative Humidity, Present Weather Codes.

To ensure the robustness of the resulting metrics, we used nested cross-validation (CV) for temporal data with ten folds. The process is illustrated in Fig. 3. Size of validation and testing folds are set to 500. A nested CV is known to provide a virtually unbiased measure of performance [40]. We train the model on the training part of the split, optimize hyperparameters on the validation part and, finally, calculate



metrics on the testing part. The data is then augmented with the next fold, and the process is repeated. To calculate the final metrics, we average the results of each fold. We note that the data is temporarily sorted, so there is no “leakage” of data. The loss function is set to LogLoss.

The following metrics are utilized for performance assessment: Precision, Recall, F1 Score, Area Under the Precision-Recall Curve (PRC AUC), and Area Under the Receiver Operating Characteristic (ROC AUC) [41]. We use Catboost [42] as our ML algorithm. It has demonstrated better performance when compared to alternative methods [43], is easy to implement, and has valuable features [44].

### B. MODEL TRAINING WITH WIND FIELD FEATURES

The next step is to evaluate the performance of the ML algorithm trained using enhanced data with incorporated wind field features. The new features are joined based on the feeder ID and timestamps, which ensures spatiotemporal correlation. Having the exact same rows (examples) for both models ensures that the results are comparable. Special care is exercised when preprocessing the datasets because often the preprocessing procedures include dropping rows with missing data. In such cases, the same rows must be dropped from both training datasets.

The newly created features enrich the dataset with information that was not available to the algorithm in the baseline case. The outage SoR levels are dependent on the speed and direction of the wind along the feeder.

### C. RESULTS DISCUSSION

The performance results for both models are presented in Table I. As can be observed, the Catboost model with wind field features has outperformed the baseline model. The difference in ROC AUC is 1.0 %, Precision did not change, Recall has improved by 2.7%, F1 Score increased by 1.7%, and PRC AUC had a change of 1.1%.

TABLE I.  
MODEL METRICS COMPARISON

Metric	Baseline	WAFGWM	Difference
ROC AUC	89.2	90.2	1.0
Precision	83.2	83.2	0.0
Recall	70.3	73	2.7
F1 Score	75.7	77.4	1.7
PRC AUC	88.1	89.2	1.1

The feature's importance is analyzed next to better understand the value of the new proposed features. This is achieved by measuring the difference between the Catboost loss function with and without individual features. Table II shows the feature importance where the new wind features are in bold text, while old features are in italic. One can conclude that some of the new features are more important than the original weather parameters: max speed and mean speed have a more considerable weight when compared to the original wind speed. That is most likely due to the fact that the high wind speeds cause outages in the system, and the WAFGWM process is capable of capturing more granular phenomena in the area. A notable fact is that all of the direction features are among the least important features. This shows that for outage prediction, wind direction may be less important than wind speed when considered in absolute terms.

### VIII. CONCLUSION

We concluded that the wind fields have a positive effect on the SoR model performance since the wind field characterizes the components that are strongly influenced by the topography and roughness length of the underlying surface that matches the topology of the feeders in the grid.

Specifically, we demonstrated that:

- The WAFGWM process improves the wind impact modeling on feeders since features from wind fields incorporate spatiotemporal dependencies.
- The performance of the ML prediction model with the new wind field features outperforms the baseline model with no wind features.
- The incorporation of the wind field features as one of the dimensions for the ML model boosts the SoR

TABLE II.  
FEATURE IMPORTANCE

Feature	Importance, %
Wind Gust	36.8
Precipitation	8.3
<b>Max speed</b>	<b>8.1</b>
<b>Mean speed</b>	<b>7.9</b>
<i>Orig. speed</i>	7.4
Dew Point Temp	7.3
<b>Min speed</b>	<b>6.9</b>
Air Temp.	3.2
Rel. humidity	3.1
Weather Codes	2.9
<b>Min direction</b>	<b>2.5</b>
<b>Max direction</b>	<b>2.3</b>
<b>Mean direction</b>	<b>1.9</b>
<i>Orig. direction</i>	1.4

model performance.

Given the findings, a promising approach is to account for wind direction in relation to feeder orientation. That is, a wind that is perpendicular to the feeder may have a higher impact than a wind directed along the feeder. Therefore, a closer analysis of wind directions is needed. For example, a more granular timescale needs to be used. These topics are left for future research.

### ACKNOWLEDGMENT

Data used in this project is provided by United Cooperative Services of North Texas, and funding is made available through the Texas A&M Engineering Experiment Station (TEES) Smart Grid Center Membership Agreement. Special thanks go to United staff, Mr. Jared Wennemark for the management guidance, and Mr. Cory Menzel for coordination of the technical support.

### REFERENCES

- [1] A. Vinogradov, A. Vinogradova, and V. Bolshev, "Analysis of the quantity and causes of outages in LV/MV electric grids," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 3, pp. 537-542, 2020, doi: 10.17775/CSEEJPES.2019.01920.
- [2] P. Kankanala, A. Pahwa, and S. Das, "Regression models for outages due to wind and lightning on overhead distribution feeders," in *2011 IEEE Power and Energy Society General Meeting: IEEE*, 2011, pp. 1-4.
- [3] S. Elphick, P. Ciuffo, V. Smith, and S. Perera, "Summary of the economic impacts of power quality on consumers," in *2015 Australasian Universities Power Engineering Conference (AUPEC): IEEE*, 2015, pp. 1-6.
- [4] M. Khoshjahan, R. Baembitov, and M. Kezunovic, "Impacts of weather-related outages on DER participation in the wholesale market energy and ancillary services," in *CIGRE Grid of the Future Symposium*, Providence, RI, USA, Oct. 2021, pp. 1-8.
- [5] M. Kezunovic, R. Baembitov, and M. Khoshjahan, "Data-driven State of Risk Prediction and Mitigation in Support of the Net-zero Carbon Electric Grid," in *11th Bulk Power Systems Dynamics and Control Symposium - IREP'2022*, Banff, Canada, 2022.
- [6] A. Arif, Z. Wang, J. Wang, and C. Chen, "Power Distribution System Outage Management With Co-Optimization of Repairs, Reconfiguration, and DG Dispatch," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4109-4118, 2018, doi: 10.1109/TSG.2017.2650917.
- [7] M. Kezunovic, P. Pinson, Z. Obradovic, S. Grijalva, T. Hong, and R. Bessa, "Big data analytics for future electricity grids," *Electric Power Systems Research*, vol. 189, p. 106788, 2020, doi: 10.1016/j.epr.2020.106788.
- [8] M. Kezunovic, Z. Obradovic, T. Djokic, and S. Roychoudhury, "Systematic Framework for Integration of Weather Data into Prediction Models for the Electric Grid Outage and Asset Management Applications," in *The Hawaii International Conf. on System Sciences - HICSS*, Waikoloa Village, Hawaii, USA, January 2018, pp. 2737-2746, doi: 10.24251/HICSS.2018.346.
- [9] M. Kezunovic and T. Dokic, "Big Data Framework for Predictive Risk Assessment of Weather Impacts on Electric Power Systems," in *Grid of the Future, CIGRE US Nat. Committee*, Atlanta, Nov., 2019.
- [10] R. Nematirad and A. Pahwa, "Solar Radiation Forecasting Using Artificial Neural Networks Considering Feature Selection," in *2022 IEEE Kansas Power and Energy Conference (KPEC)*, 25-26 April 2022, pp. 1-4, doi: 10.1109/KPEC54747.2022.9814765.
- [11] A. Ghasemi, A. Shojaeighadikolaei, K. Jones, M. Hashemi, A. G. Bardas, and R. Ahmadi, "A Multi-Agent Deep Reinforcement Learning Approach for a Distributed Energy Marketplace in Smart Grids," in *2020 IEEE International Conf. on Communications*,

- Control, and Computing Technologies for Smart Grids (SmartGridComm), 2020, pp. 1-6, doi: 10.1109/SmartGridComm47815.2020.9302981.
- [12] F. Yang, D. Cerrai, and E. N. Anagnostou, "The Effect of Lead-Time Weather Forecast Uncertainty on Outage Prediction Modeling," *Forecasting*, vol. 3, no. 3, pp. 501-516, 2021, doi: 10.3390/forecast3030031.
- [13] D. Cerrai *et al.*, "Predicting Storm Outages Through New Representations of Weather and Vegetation," *IEEE Access*, vol. 7, pp. 29639-29654, 2019, doi: 10.1109/ACCESS.2019.2902558.
- [14] D. Cerrai, M. Koukoulou, P. Watson, and E. N. Anagnostou, "Outage prediction models for snow and ice storms," *Sustainable Energy, Grids and Networks*, vol. 21, p. 100294, 2020, doi: 10.1016/j.segan.2019.100294.
- [15] J. B. Leite and M. Kezunovic, "Risk Mitigation Approaches for Improved Resilience in Distribution Networks," in *2020 IEEE PES Transmission & Distribution Conf. and Exhibition - Latin America (T&D LA)*, 2020: IEEE, pp. 1-6, doi: 10.1109/TDLA47668.2020.9326210.
- [16] J. B. Leite, J. R. S. Mantovani, T. Dokic, Q. Yan, P.-C. Chen, and M. Kezunovic, "Resiliency Assessment in Distribution Networks Using GIS-Based Predictive Risk Analytics," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4249-4257, 2019, doi: 10.1109/TPWRS.2019.2913090.
- [17] E. H. Ko, T. Dokic, and M. Kezunovic, "Prediction Model for the Distribution Transformer Failure Using Correlation of Weather Data," in *5th International Colloquium on Transformer Research and Asset Management*, Trkulja B., Štih Ž., and J. Ž. Eds. Singapore: Springer, 2020.
- [18] R. Tervo, J. Karjalainen, and A. Jung, "Short-Term Prediction of Electricity Outages Caused by Convective Storms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 8618-8626, 2019, doi: 10.1109/TGRS.2019.2921809.
- [19] W. Frost and R. L. Bowles, "Wind shear terms in the equations of aircraft motion," *Journal of Aircraft*, vol. 21, pp. 866-872, 1984.
- [20] D. G. Ross, I. N. Smith, P. C. Manins, and D. G. Fox, "Diagnostic Wind Field Modeling for Complex Terrain: Model Development and Testing," *Journal of Applied Meteorology (1988-2005)*, vol. 27, no. 7, pp. 785-796, 1988. [Online]. Available: <http://www.jstor.org/stable/26183717>.
- [21] R. Jahdi, A. A. Darvishsefat, V. Etemad, and M. A. Mostafavi, "Wind Effect on Wildfire and Simulation of its Spread (Case Study: Siahkal Forest in Northern Iran)," *Journal of Agricultural Science and Technology*, vol. 16, no. 5, pp. 1109-1121, 2014.
- [22] S. Finardi, G. Tinarelli, P. Faggiani, and G. Brusasca, "Evaluation of different wind field modeling techniques for wind energy applications over complex topography," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 74-76, pp. 283-294, 1998, doi: 10.1016/S0167-6105(98)00025-7.
- [23] G. F. Homicz, Research Org.: Sandia National Lab. (SNL-NM), Albuquerque, NM (United States); Sandia National Lab. (SNL-CA), Livermore, CA (United States), "Three-Dimensional Wind Field Modeling: A Review," United States, 2002. [Online]. Available: <https://www.osti.gov/biblio/801406>
- [24] Y. Wang, C. Williamson, D. Garvey, S. Chang, and J. Cogan, "Application of a Multigrid Method to a Mass-Consistent Diagnostic Wind Model," *Journal of Applied Meteorology*, vol. 44, no. 7, pp. 1078-1089, 01 Jul. 2005 2005, doi: 10.1175/jam2262.1.
- [25] P. S. Jackson and J. C. R. Hunt, "Turbulent wind flow over a low hill," *Quarterly Journal of the Royal Meteorological Society*, vol. 101, no. 430, pp. 929-955, 1975, doi: 10.1002/qj.49710143015.
- [26] P. J. Mason and R. I. Sykes, "Flow over an isolated hill of moderate slope," *Quarterly Journal of the Royal Meteorological Society*, vol. 105, no. 444, pp. 383-395, 1979, doi: 10.1002/qj.49710544405.
- [27] *WindNinja*. (2020). US Forest Service. [Online]. Available: <https://www.firelab.org/project/windninja>
- [28] "Automated Surface Observing Systems." NOAA's National Weather Service. <https://www.weather.gov/asos/asostech> (accessed 2021).
- [29] *ArcGIS Pro*. (2021). Esri Inc.; [Online]. Available: [arcgis.com](https://www.esri.com)
- [30] J. M. Forthofer, B. W. Butler, and N. S. Wagenbrenner, "A comparison of three approaches for simulating fine-scale surface winds in support of wildland fire management: Part I. Model formulation and comparison against measurements," *International Journal of Wildland Fire*, 2014, doi: 10.1071/WF12089.
- [31] N. S. Wagenbrenner, J. M. Forthofer, B. K. Lamb, K. S. Shannon, and B. W. Butler, "Downscaling surface wind predictions from numerical weather prediction models in complex terrain with WindNinja," *Atmos. Chem. Phys.*, vol. 16, no. 8, pp. 5229-5241, 2016, doi: 10.5194/acp-16-5229-2016.
- [32] R. M. F. S. Laboratory, "Tutorial 1: The Basics," *WindNinja Tutorials*. [Online]. Available: <https://weather.firelab.org/windninja/tutorials>
- [33] "Landscape File (.LCP)." Systems for Environmental Management. [https://www.fire.org/downloads/farsite/WebHelp/referenceguide/pops/ups/landscape\\_file.htm](https://www.fire.org/downloads/farsite/WebHelp/referenceguide/pops/ups/landscape_file.htm) (accessed 2021).
- [34] "LANDFIRE Program: Data Products - Overview - LF 2016 Remap (LF 2.0.0)." U.S. Department of Agriculture Forest Service. [https://landfire.gov/lf\\_remap.php](https://landfire.gov/lf_remap.php) (accessed 8. Sep. 2021).
- [35] E. Walls, "Continuum Wind Flow Model on JSTOR," *Wind Engineering*, vol. 39, no. 3, pp. 271-298, 2015. [Online]. Available: <https://www.jstor.org/stable/90007070>.
- [36] C. G. Homer, J. A. Fry, and C. A. Barnes, "The National Land Cover Database," *Fact Sheet*, 2012, doi: 10.3133/fs20123020.
- [37] Iowa Environmental Mesonet: ASOS One Minute Data Download [Online]. Available: <https://mesonet.agron.iastate.edu/request/asos/1min.phtml>
- [38] ESRI, "ArcGIS Pro 2.8 Projected Coordinate System Tables." [Online]. Available: [https://pro.arcgis.com/en/pro-app/latest/arcpy/classes/pdf/projected\\_coordinate\\_systems.pdf](https://pro.arcgis.com/en/pro-app/latest/arcpy/classes/pdf/projected_coordinate_systems.pdf)
- [39] ESRI, "Spatial Join (Analysis)—ArcGIS Pro | Documentation." [Online]. Available: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/analysis/spatial-join.htm>
- [40] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformatics*, vol. 7, no. 1, p. 91, 2006/02/23 2006, doi: 10.1186/1471-2105-7-91.
- [41] D. M. W. Powers, "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation," *Mach. Learn. Technol.*, vol. 2, no. 1, 2008.
- [42] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: gradient boosting with categorical features support," *ArXiv e-prints*, 2018. [Online]. Available: <https://arxiv.org/abs/1810.11363v1>.
- [43] R. Baembitov, T. Dokic, M. Kezunovic, Y. Hu, and Z. Obradovic, "Fast Extraction and Characterization of Fundamental Frequency Events from a Large PMU Dataset using Big Data Analytics," in *54th Hawaii International Conference on System Sciences*, 2021, pp. 3195-3204, doi: 10.24251/HICSS.2021.389.
- [44] R. Baembitov, M. Kezunovic, and Z. Obradovic, "Graph Embeddings for Outage Prediction," in *The 53rd North American Power Symposium (NAPS 2021)*, College Station, TX, USA, 2021.



**RASHID BAEMBITOV** (S'19) received B.Sc. and M.Sc. degrees in Electrical Power Engineering and Economics and Business Administration from the National Research University "Moscow Power Engineering Institute", Moscow, Russia, in 2014 and 2016, respectively. Currently he is a PhD student and Graduate Research Assistant at the Texas A&M University. His main research interests are big data, artificial intelligence and machine learning for power system applications, power system risk assessment, grid integration technologies.



**MLADEN KEZUNOVIC** (S'77–M'80–SM'85–F'99–LF'17) received the Dipl. Ing. from University of Sarajevo, Sarajevo, Bosnia and Herzegovina, and M.Sc. and Ph.D. degrees in electrical engineering from University of Kansas, Lawrence, KS, in 1974, 1977, and 1980, respectively. He has been with Texas A&M University, College Station, TX, USA since 1987, where he is currently Regents Professor, Eugene E. Webb endowed Professor, and the Site Director of the “Power Engineering Research

Center” consortium. He is the Principal Consultant of XpertPower Associates, a consulting firm specializing in power systems data analytics. His expertise is in protective relaying, automated power system disturbance analysis, computational intelligence, data analytics, and smart grids. He has authored over 600 papers, given over 120 seminars, invited lectures, and short courses, and consulted for over 50 companies worldwide. Dr. Kezunovic is a CIGRE Fellow, Honorary and Distinguished member. He is Registered Professional Engineer in Texas. Dr. Kezunovic is a member of the US National Academy of Engineering.



**KEITH A. BREWSTER** ((Member, IEEE, '06) earned a BS in meteorology from the University of Utah ('81) and M.S. ('84) and PhD ('99) from the School of Meteorology at the University of Oklahoma. He is a Senior Research Scientist and Director of Operations at the Center for Analysis and Prediction of Storms at the University of Oklahoma, where he has worked since 1993, advancing research in weather radar data

analysis, data assimilation, high resolution numerical prediction and ensemble prediction for improving forecasts of thunderstorms, flash flooding, renewable energy production and winter storms and has published more than 150 journal and conference articles. He is a member of the American Meteorological Society and American Geophysical Union, serves as Co-Chair of the NOAA Unified Forecasting System Post-Processing Committee and is on the AMS Nationwide Network of Networks committee.



**ZORAN OBRADOVIC** (Senior Member, IEEE) is a Distinguished Professor, the Center Director at Temple University, an Academician with the Academia Europaea (the Academy of Europe), and a Foreign Academician with the Serbian Academy of Sciences and Arts. He mentored 46 Postdoctoral Fellows and Ph.D. students, many of whom have independent research careers at academic institutions and industrial research laboratories. His research results were published at about 400 data science

and complex networks articles addressing challenges related to big, heterogeneous, spatial–temporal data analytics motivated by applications in healthcare management, power systems, earth, and social sciences. He is also an Editorial Board Member at 13 journals and was the general chair, the program chair, or the track chair of 11 international conferences. He is the Steering Committee Chair of the SIAM Data Mining Conference. He is the Editor-in-Chief of the Journal of Big Data.