Proceedings of the ASME 2018 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2018
August 26-29, 2018, Quebec City, Quebec, Canada

DETC2018-85127

COMPARING MACHINE LEARNING REGRESSION TECHNIQUES FOR TRANSMISSION-RELATED STORM OUTAGES

Caitlyn E. Clark
PhD Candidate
Department of Mechanical Engineering
Oregon State University
Corvallis, Oregon 97330
Email: clarkc@oregonstate.edu

Bryony DuPont*
Assistant Professor
Department of Mechanical Engineering
Oregon State University
Corvallis, Oregon 97330
Email: bryony.dupont@oregonstate.edu

ABSTRACT

In this study, we characterize machine learning regression techniques for their ability to predict storm-related transmission outages based on local weather and transmission outage data. To test the machine learning regression techniques, we use data from the central Oregon Coast—which is particularly vulnerable to storm-related transmission outages—for a case study. We test multiple regression methods (linear and polynomial models with varying degrees) as well as support vector regression methods using linear, polynomial, and Radial-Basis-Function kernels. Results indicate relatively poor prediction capability by these methods, but this is attributed to the lack of outage data (characteristic of low-probability, high-risk events), and a cluster of data points representing momentary (<0 seconds) outages. More long-term outage data could lead to better characterization of the models, enabling others to quantify the frequency of storm-related transmission outages based on local weather data. Only by understanding the frequency of these occurrences can a cost-benefit analysis for potential transmission upgrades or generation sources be completed.

*Corresponding author

INTRODUCTION

Storm-related transmission outages can be costly, resulting in widespread effects and cascading damages in affected communities. Improving electrical resiliency in these systems through transmission or generation enhancements can help avoid outage consequences. To compare the costs of potential improvements with the cost of outages, stakeholders of electrical equipment management need to know the frequency and duration of local outages.

Previous work using extreme weather [1–4] or normal conditions [5, 6] as explanatory variables have mainly focused on predicting number of outages. Of those methods used for storm-related outages, Zhu et al. propose an exponential model [1], while Zhou et al. use Poisson regression and Bayesian network models [2]. Liu et al. propose two statistical models (a negative binomial regression model and a spatial generalized mixing model) to predict number of outages due to hurricanes and ice storms [3, 7]. Kankanala et al. focus on outage occurrence, testing linear, quadratic, and exponential regression models, multi-layered neural networks, a mixture of experts methods, and an ensemble-learning approach based on a boosting algorithm [6, 8–10]. While these methods inform our under-

Copyright © 2018 by ASME

standing of how to predict outage occurrence, methods for predicting outage duration are less explored.

Only more recently have algorithms been proposed for predicting storm-related outage duration, and none have focused on transmission outages. Chow et al. test the statistical significance of explanatory variables for distribution outage duration, but no forecasting algorithm is given [11]. Adibi et al. develop an estimation method for system outage duration, but it requires detailed information about the outage and its restoration [12]—information that is often unavailable. Rodriguez et al. provide a fuzzy logic technique that uses human input to determine the relative importance of the possible explanatory variables in distribution system outages [13]. Jaech et al. leverage natural language processing to improve the performance of neural networks to predict the duration of storm-related distribution outages [14]. In these studies, only Adibi et al. include transmission outages in their analysis. Our study will focus on transmission outages because of the spatial and temporal extent a failure in the transmission system causes, as well as because transmission systems are less reliable during storm events in the area of interest (Oregon) than distribution systems [15].

In this work, we characterize machine learning regression techniques for their ability to predict storm-related transmission outages using local weather and transmission outage data on the Oregon Coast. First, we will describe how outages in the U.S. affect coastal communities, and what is currently being done to avoid or reduce these consequences. Then, we will detail the case study location on the central Oregon coast, and why it is particularly susceptible to outages. We will follow this by reporting which machine learning regression techniques we used and their results. Finally, we will discuss the methods' abilities to predict storm-related transmission outages and areas for further research.

Electrical Outages in the U.S.

Coastal communities are especially vulnerable to electrical outages. In the United States (U.S.), over half the population lives within 50 miles of a coast [16], and 78% of energy usage occurs in these coastal areas [17]. Because of this concentration of population and energy use, outages in coastal areas have relatively higher impacts than less-populated inland areas. Furthermore, coastal population centers face specific low-probability, high-risk hazards which can effect electrical infrastructure. These specific hazards include coastal flooding, tsunamis, and hurricanes, and have impacts that are exacerbated by human action and environmental change [18–21].

Hazard-related outages are costly, especially in the U.S. where electrical infrastructure is often aged. Estimates of economic loss due to weather-related outages average between $18 and $33 billion annually, and are higher during years of major storm events, such as Hurricane Ike in 2008 ($40-75 billion) and Hurricane Sandy in 2012 ($27-52 billion) [22]. Increasing grid resiliency, or the ability of electrical assets and facilities to anticipate, resist, absorb, respond to, and recover from low-probability, high-risk disturbances [23], can help avoid costs and outage impacts in coastal communities.

Electrical infrastructure has conventionally been designed for reliability, or high-probability, low-risk events [24]. Consequently, designing for electrical resilience is a new focus for researchers and industry. Thus far, efforts to improve electrical resilience include prevention of damage, hastened recovery, and improvements in survivability (maintaining functionality despite damage) [25]. Only in some instances are outdated transmission systems replaced with modern, more robust systems.

To understand if the cost of improving electrical resilience outweighs the cost of equipment repair and collateral damages associated with an outage, the frequency of such outage events must first be quantified. This frequency is dependent on the conditions of the transmission infrastructure location. The next section will describe the specific conditions of coastal Oregon communities and their transmission system.

Electrical Outages in Oregon

From December 1-3, 2007, the Pacific Northwest experienced a series of 50-year storm events. These storms killed 38 people, caused over $488 million dollars in damage (2017 USD), and destroyed over 50,000 dwellings. In downtown Portland, wind gusts reached 116 mph, and other cities lost power for two to three weeks [26]. Although 550,000 electric customers from northern California to British Columbia lost power, northwest Oregon and southwest Washington were most affected, especially in coastal areas. Coastal areas and mountain ranges expe-

Copyright © 2018 by ASME

rienced greater damages from high winds (from 70-129 mph), landslides, mudslides, debris flows, storm surges, heavy rainfall, heavy snowfall, high turbulent seas, high surf, prolonged storm duration, and coastal and inland flooding of rivers, streams, and drainage basins [15].

Furthermore, the storm and subsequent environmental events caused infrastructure damage that left coastal Oregon communities without power and communications, disabling most critical services. During the three-day storm period, transmission and distribution electrical systems were flooded in low-lying coastal areas, undermined due to erosion and landslides, downed from falling trees and debris, or simply blown over or broken by the wind. Power outages caused collateral damage to wireless communications, weather stations, gas stations, emergency response services, pump stations, hospitals, medical clinics, and emergency number public safety answering points [15].

While part of the destruction to the electrical system was due to the storm intensity, the vulnerability of the coastal electrical system is also partly responsible for the resulting damages from power outages. Bonneville Power Administration and PacifiCorp service the lines that transmit electricity from inland areas over the coastal mountain range to coastal areas in Oregon and Washington. Oregon has six lines (rated at either 230 or 115 kV) that supply coastal Oregon areas with 100% of the area's power supply needs, sourcing all power from east of the mountain range. The coast has no generation capabilities.

Few lines connect the coast to the Willamette Valley due to the coast's relative isolation. Separated from the Willamette Valley by a heavily forested coastal range, building and maintaining transmission infrastructure is expensive and labor-intensive. During storms, this infrastructure is particularly prone to damage from wind, downed trees, rain-related flooding, and slope instability. If one of these transmission lines is damaged in a storm, a large area is impacted.

Important to note is the reliability of transmission versus distribution infrastructure during such storms. For instance, in the storms that occurred early in December 2007, distribution systems were more reliable than transmission systems, generally due to distribution systems being located in cleared public road right-of-ways, rather than on forested mountain sides prone to erosion [15]. This indicates that, if there existed local generation sources for Oregon coastal regions west of the coastal range, electric systems would perform better during large outages.

PROBLEM FORMULATION AND APPROACH

To quantify the frequency and duration of storm-related transmission outages, we develop a predictive model for outage duration based on weather data. The frequency and duration of these outages can be used to compare the value of loss load during an outage to the cost of potential solutions to improve electrical resiliency, such as emergency energy generation solutions or transmission system upgrades.

Machine Learning Techniques

We used two machine learning regression methods to correlate storm characteristics and power outages in Lincoln County, Oregon. We compared multiple regression methods (linear and polynomial models with varying degrees), as well as support vector regression methods using linear, polynomial, and Radial-Basis-Function kernels. Supervised regression techniques were chosen for the type of data being used (we have data with outcomes, and can leverage those outcomes to train our model), and for previous research that has shown regression techniques to appropriately model weather effects on outages.

In supervised regression techniques such as those used in this study, we assume independent and dependent variables x and y have an underlying function (called the target function) that relates them. To learn the target function, the regression method trains with a set of training data (x), in which the outcome or response (y) is known, and attempts to minimize the error between the calculated model outcome and the actual outcome.

The relationship between weather (x) and outage duration (y) could be linear or polynomial, thus we tested each of these hypothesis spaces. Kankanala et al. report that the effects of wind on outages can be modeled using second order regression models [8]. Islam describes regression relationships between weather conditions (wind, ice storms, heat storms, rain, lightning, temperature) and outage occurrence [27]. They describe a piecewise relationship between rain and the number of regional interruptions, and a third order relationship between wind speed and interruptions.

We then analyzed simple regression techniques against support vector regression methods to compare their efficiencies at solving for the target function. Linear regression

Copyright © 2018 by ASME

techniques assume a linear relationship, and then attempt to minimize the error between the guessed and actual outcomes though gradient-based or other exact optimization methods. In linear regression, every data point (the blue points in fig. 1) is used to calculate the error incurred by the guessed target function (the green function in fig. 1). Support vector regression methods use a margin to either side of the guessed target function to determine whether a data point contributes towards a cost function (those points that lie outside of the margin (ε) in fig. 1). Over time, the cost is minimized, as well as the width of the margin. Support vector regression techniques also differ from other regression techniques in that they learn from and assign a weight to the ith training example, penalizing the algorithm (through a penalty parameter, C) for those points which lay outside the margin, rather than learning some fixed set of parameters corresponding to the features of the inputs.

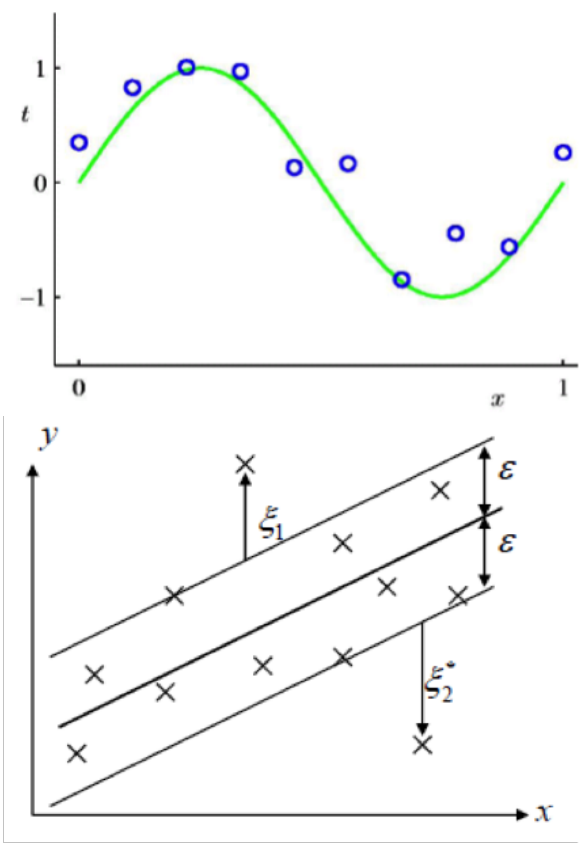

FIGURE 1: LINEAR (TOP) AND SUPPORT VECTOR (BOTTOM) REGRESSION METHODS [28]

In this study, we also test the efficacy of various kernels. Kernels transform the data into a higher dimensional feature space to make separation by the margin possible. Linear, polynomial, and Radial-Basis-Function kernels are used to transform the data into linear, polynomial, and variable dimensions.

Data Sources

We sourced outage data from Bonneville Power Administration [29], with outage duration measured in hours. We then filtered the outage data to only include transmission line outages that affect local transmission infrastructure in Lincoln County, Oregon. Therefore, only outages between Wendson, Toledo, and Albany were considered. This transmission section is shown in Fig. 2. This filtering of data resulted in a total of 86 outage examples. Outages ranged from momentary (which resulted in a 0 : 00 hr:min count) to sustained outages of over 250 hours.

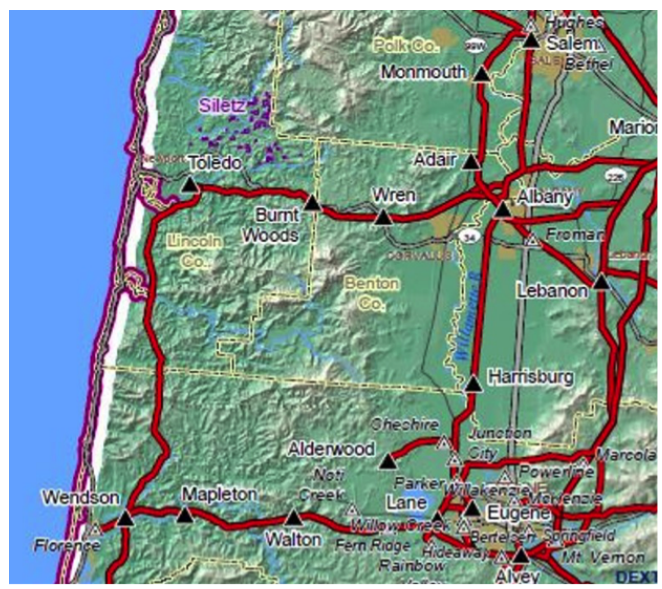

FIGURE 2: TRANSMISSION SYSTEM SECTION USED IN THIS STUDY [29]

We sourced weather data from Hatfield Marine Science Center in Newport, Oregon [30]. Using three features, we related the major causes of transmission infrastructure damage (wind, downed trees, flooding, and slope instability) to daily average wind speed (mph), maximum

Copyright © 2018 by ASME

gust speed (mph), and cumulative rainfall (in). Source data was measured every 5 minutes, but for the purpose of this study, this data was averaged over a 24-hour period. Fig. 3 shows this feature data to show the overall correlation between the three variables.

We analyzed outages that occurred between 0:00:00 January 1, 2011 to 23:59:59 December 31, 2015, which is concurrent with the time frame used for the weather data.

EXPERIMENTS

To find an accurate predictive model, we compared several regression methods. These models included regression with a linear model and polynomial model with degrees of 2, 3, and 4, as well as a support vector regression technique in which linear, polynomial (degree of 2), and RBF kernels were used. All models were used from Python's Scikit-Learn package [31]. Data was split into 90% training data and 10% testing data throughout all methods.

Linear Regression

The linear regression model produced with the Scikit-learn Linear_model module had coefficients of $[-5.74211874, 3.96718265, -25.4516585]$. Fig. 4 shows the linear model predictions in blue, while the actual outage data is shown in red.

The mean squared error and explained variance score of this model are 3038.5 and -901.16, respectively. While the mean squared error value is reasonable for data that has variation like the data in this study, the variance score indicates a poor model fit.

Polynomial Regression

After representing a linear regression with a degree of 1, further polynomials were represented, with degrees of 2, 3, and 4. Each model was determined with Scikit-Learn's Linear Regression and Ridge Regression modules. These modules are depicted in Fig. 5.

The mean squared error values of each degree polynomial model and the respective standard deviation was calculated using cross-validation calculations in Scikit-Learn's cros_val module. These values are summarized in Table 1 and Table 2.

Again, the mean squared error values are reasonable for the variation in the data, which is also reflected by the standard deviation in the polynomial models. Of the four models tried, the polynomial model with an order of 2 was the best fit, with the lowest average mean squared error.

TABLE 1: MEAN SQUARED ERROR (HRS) FOR POLYNOMIAL MODELS OF VARYING DEGREES IN THE APPLIED LINEAR REGRESSION TECHNIQUE

Linear	Degree = 2	Degree = 3	Degree = 4
3.91E+03	3.89E+03	4.59E+03	8.95E+03
3.66E+03	7.14E+03	1.28E+04	3.27E+03
3.89E+03	4.58E+03	4.92E+03	6.11E+03

TABLE 2: STANDARD DEVIATION (HRS) FOR POLYNOMIAL MODELS OF VARYING DEGREES IN THE APPLIED LINEAR REGRESSION TECHNIQUE

Linear	Degree = 2	Degree = 3	Degree = 4
3.45E+03	2.93E+03	3.72E+03	1.32E+03
3.77E+03	3.86E+03	4.89E+04	5.54E+03
3.96E+03	1.17E+03	2.75E+03	3.17E+03

Support Vector Regression

After testing linear and polynomial regression techniques, we tested the performance of a Support Vector Regression method. Three kernels were tested, including a linear, polynomial (of degree 2), and a Radial-Basis-Function (RBF) kernel. For all kernels, the penalty parameter C of the error term was set to 1.0E+03, while the kernel coefficient, γ, was set to 0.1 for polynomial and RBF. The comparison of these models is depicted in Fig. 6.

Copyright © 2018 by ASME

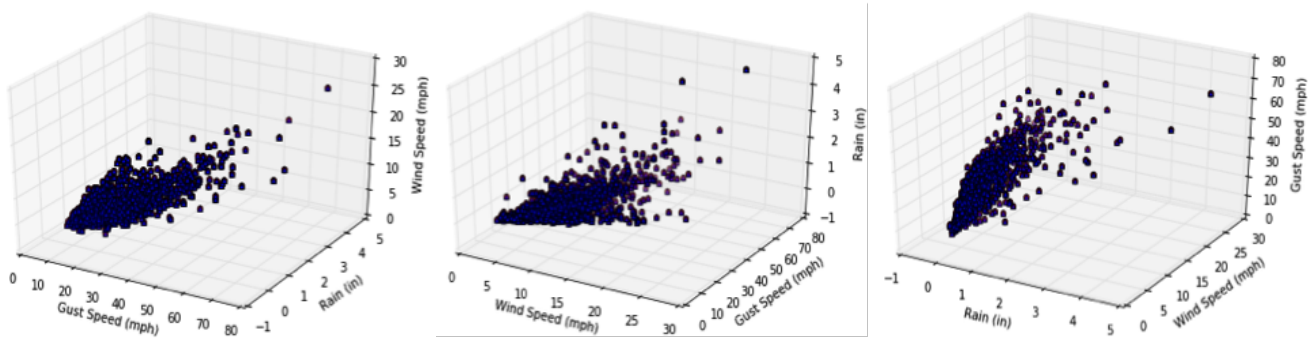

FIGURE 3: GUST WIND SPEED, AVERAGE WIND SPEED, AND RAINFALL DATA FOR ALL OUTAGES, IN THREE VIEWS

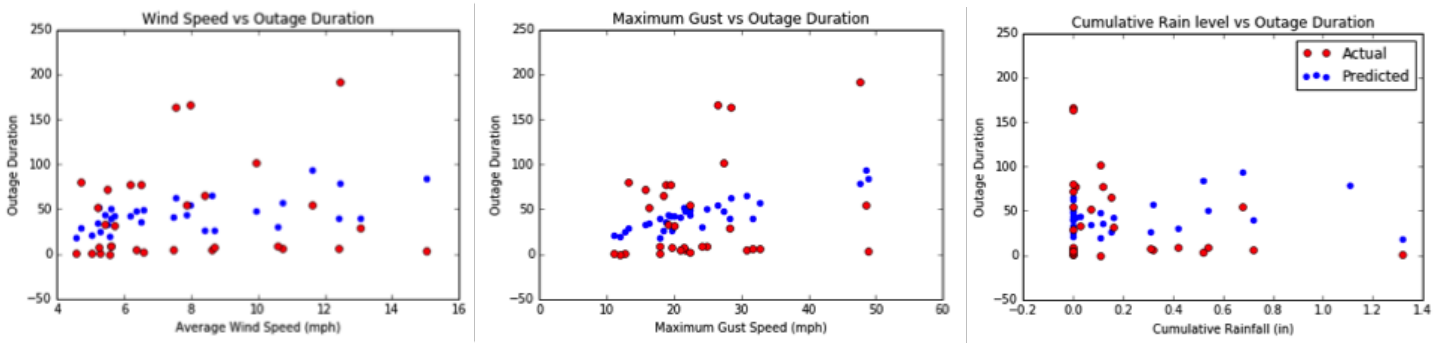

FIGURE 4: LINEAR REGRESSION MODEL WITH LINEAR FIT IN BLUE, WITH ACTUAL DATA IN RED

Copyright © 2018 by ASME

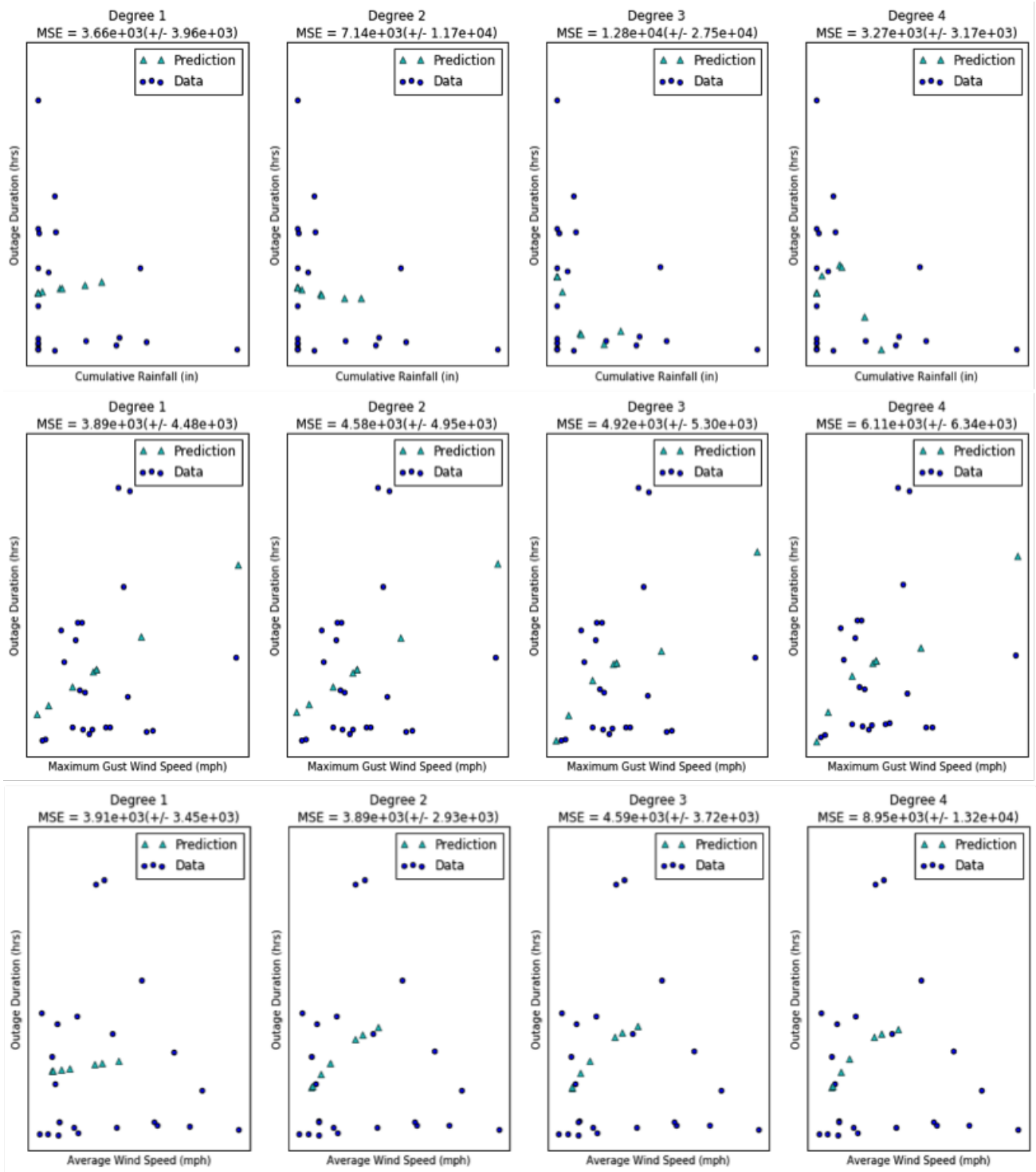

FIGURE 5: POLYNOMIAL REGRESSION WITH VARYING DEGREES

Copyright © 2018 by ASME

Mean squared error and explained variance values for these three techniques across each feature is summarized below in Table 3 and Table 4. Both calculations were performed via SciKit-Learn's metric module. The tables show mean squared error values consistent with the variation in the data, while the explained variance shows better fits than in the linear model. The best average mean squared error and explained variance is from the model with RBF kernel, especially when representing the outages and maximum gust data.

TABLE 3: MEAN SQUARED ERROR (HRS) COMPARISON OF DIFFERENT KERNEL METHODS FOR SVR

Linear Kernel	Polynomial Kernel	RBF Kernel
3468.6	3858.799	2713.043
3091.625	15000.17	936.5555
3157.746	3854.749	3232.379

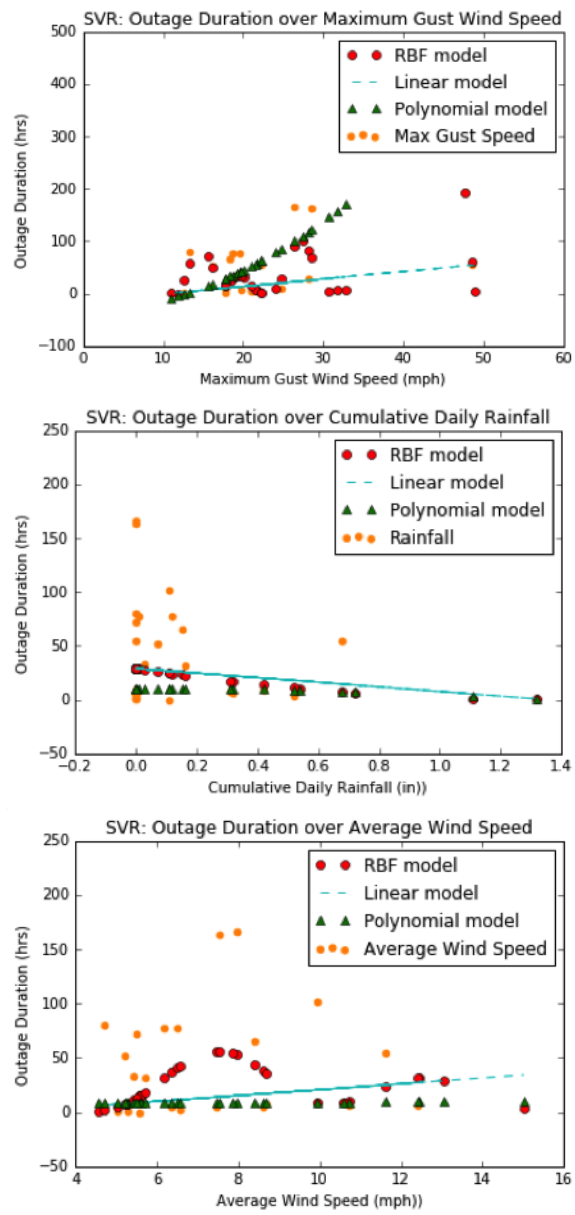

FIGURE 6: COMPARISON OF LINEAR, POLYNOMIAL, AND RBF KERNELS IN LINEAR REGRESSION

DISCUSSION

In this study, we compared machine learning regression techniques for their ability to predict storm-related transmission events from local weather and outage data. When comparing the models, the Support Vector Regression model with the Radial-Basis Function kernel best represented the data, especially the outages and maximum gust wind data. Fig. 4 through fig. 6 qualitatively show that even this fit by the SVR model with an RBF kernel does not adequately capture the data variation when making predictions, resulting in clustered outage predictions. The linear model did the best at capturing the variation of the data, with outage predictions clustering less (except for in the maximum gust speed data). In summary, none of the techniques ac-

Copyright © 2018 by ASME

TABLE 4: EXPLAINED VARIANCE (HRS) COMPARISON OF DIFFERENT KERNEL METHODS FOR SVR

Linear Kernel	Polynomial Kernel	RBF Kernel
0.018722	0.000967	0.130716
0.074963	-3.52251	0.676594
-0.03124	-0.00998	-0.02939

curately predicted the duration of an outage based on the weather conditions. This accuracy, however, cannot be attributed to the models alone. The volume and distribution of data is also a factor to consider when evaluating regression model performance.

The dataset, which was gathered between January 1, 2011 and December 31, 2015, did not have enough data points to train the predictive models to the accuracy desired. Further, there was a group of data points that had outage duration of 0.00 hours (momentary outage), while sustained outages that lasted for several days were sparse. The effect on model prediction is exemplified in Fig. 4, in which model prediction of outage duration given average wind speed and maximum gust speed is influenced by momentary outages, despite a visible relationship between wind speeds and outage duration. The impact of momentary outages can be seen throughout the models; although higher-order models were penalized by the machine learning algorithms for not capturing momentary outages, values of error decreased slightly for those models which captured the higher-order relationship between weather and outage duration. If there had been more data points for these sustained outage events, and more even data distribution, the models would likely be more accurate.

Another reason for model inaccuracy could be that the outage data that was used to train these models did not account for the reason for the outage. For instance, a five-hour outage could be due to a storm event damaging equipment, or it could be due to a planned maintenance event. More thorough data collection would allow for filtering of outages to only include weather-induced events without reducing the models' accuracy.

In future work, more data needs to be collected, either by finding similar weather data previous to 2011 that can compliment the data currently being used, or by spatially extending the study to collate weather and outage data from neighboring coastal communities. With more data, there will be a higher probability of having more sustained outage events caused by weather to better train the models. Gathering more filtered data could also make multivariate analysis for the analysis of all three features concurrently feasible. Given that the models for maximum gust speed were less similar to the other features' models, the same comparison study with more filtered data should be repeated before completing multivariate analysis.

By improving the accuracy of these models to predict frequency and duration of storm-related transmission outages, electrical infrastructure stakeholders can make informed decisions about how to improve the system through transmission or generation solutions.

CONCLUSIONS

In this study, we compare machine learning regression models and their ability to predict storm-related transmission outages based on local weather data. The feature variables compared were daily cumulative rainfall, maximum daily wind gusts, and daily average wind speed. From this comparison study, we found that none of the above models adequately predicted outage duration. This is mostly attributed to the variation in data, which is influenced by the relatively high number of momentary outages that lasted less than five minutes. Another reason for model prediction inaccuracy is attributed to lack of data filtering for storm-related outages. Model prediction accuracy is hypothesized to improve with a larger number of outage data points, filtered by the reason for the outage and the outage duration (greater than five minutes). With more data and accurate models, stakeholders can predict outage duration based on local weather and transmission data, leading to more informed decision making about transmission infrastructure management and failure response.

ACKNOWLEDGMENT

Thanks go to Junki Hong for his helpful input for this study and the Avangrid Foundation for their financial support for this work.

Copyright © 2018 by ASME

REFERENCES

[1] Zhu, D., Cheng, D., Broadwater, R. P., and Scirbona, C., 2007. "Storm modeling for prediction of power distribution system outages". *Electr. Power Syst. Res.,* ***77***(8), pp. 973–979.

[2] Zhou, Y., Pahwa, A., and Yang, S. S., 2006. "Modeling weather-related failures of overhead distribution lines". *IEEE Trans. Power Syst.,* ***21***(4), pp. 1683–1690.

[3] Liu, H., Davidson, R. A., and Apanasovich, T. V., 2008. "Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms". *Reliab. Eng. Syst. Saf.,* ***93***(6), pp. 897–912.

[4] Alvehag, K., and Söder, L., 2011. "A reliability model for distribution systems incorporating seasonal variations in severe weather". *IEEE Trans. Power Deliv.,* ***26***(2), pp. 910–919.

[5] Domijan Jr., A., Matavalam, R., Montenegro, A., Wilcox, W., Joo, Y., Delforn, L., Diaz, J., Davis, L., and D'Agostini, J., 2005. "Effects of Normal Weather Conditions on Interruptions in Distribution Systems". *Int. J. Power Energy Syst.,* ***25***, pp. 54–61.

[6] Kankanala, P., Pahwa, A., and Das, S., 2012. "Estimation of Overhead Distribution System Outages Caused by Wind and Lightning Using an Artificial Neural Network". No. September.

[7] Liu, H., Davidson, R. A., Rosowsky, D. V., and Stedinger, J. R., 2005. "Negative binomial regression of electric power outages in hurricanes". *J. Infrastruct. Syst.,* ***11***(4), pp. 258–267.

[8] Kankanala, P., Pahwa, A., and Das, S., 2011. "Regression models for outages due to wind and lightning on overhead distribution feeders". *IEEE Power Energy Soc. Gen. Meet.*, pp. 1–4.

[9] Kankanala, P., Pahwa, A., and Das, S., 2011. "Exponential regression models for wind and lightning caused outages on overhead distribution feeders". *NAPS 2011 - 43rd North Am. Power Symp.*(September 2015).

[10] Kankanala, P., Das, S., and Pahwa, A., 2014. "AdaBoost+: An Ensemble Learning Approach for Estimating Weather-Related Outages in Distribution Systems". *IEEE Trans. Power Syst.,* ***29***(1), pp. 359–367.

[11] Chow, M. Y., Taylor, L. S., and Chow, M. S., 1996. "Time of outage restoration analysis in distribution systems". *IEEE Trans. Power Deliv.,* ***11***(3), pp. 1652–1658.

[12] Adibi, M. M., and Milanicz, D. P., 2000. "Estimating restoration duration". *Power Syst. Restor. Methodol. Implement. Strateg.,* ***14***(4), pp. 245–250.

[13] Rodríguez, J. R. A., and Vargas, A., 2005. "Fuzzy-heuristic methodology to estimate the load restoration time in MV networks". *IEEE Trans. Power Syst.,* ***20***(2), pp. 1095–1602.

[14] Jaech, A., Zhang, B., Ostendorf, M., and Kirschen, D. S., 2018. "Real-Time Prediction of the Duration of Distribution System Outages". pp. 1–8.

[15] Elliott, T., and Tang, A. K., 2012. *Pacific Northwest Storms of December 1-4, 2007: Lifeline Performance*. American Society of Civil Engineers.

[16] Crowell, M., Coulton, K., Johnson, C., Westcott, J., Bellomo, D., Edelman, S., and Hirsch, E., 2005. "An Estimate of the U.S. Population Living in 100-Year Coastal Flood Hazard Areas". *J. Coast. Res.,* ***21***(5), pp. 867–1084.

[17] Thresher, R., Robinson, M., and Veers, P., 2008. The Future of Wind Energy Technology in the United States. Tech. Rep. July, Glasgow, UK.

[18] St. Cyr, J. F., 2005. "At Risk: Natural Hazards, People's Vulnerability, and Disasters". *J. Homel. Secur. Emerg. Manag.,* ***2***(2).

[19] Turner 2nd, B. L., Kasperson, R. E., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., Eckley, N., Kasperson, J. X., Luers, A., Martello, M. L., Polsky, C., Pulsipher, A., and Schiller, A., 2003. "A framework for vulnerability analysis in sustainability science". *Proc Natl Acad Sci U S A,* ***100***(14), pp. 8074–8079.

[20] O'Keefe, P., Westgate, K., and Wisner, B., 1976. "Taking the naturalness out of natural disasters". *Nature,* ***260***(5552), pp. 566–567.

[21] Adger, W. N., 2005. "Social-Ecological Resilience to Coastal Disasters". *Science (80-.).,* ***309***(5737), pp. 1036–1039.

[22] Executive Office of the President, and Executive Office of the President of the U.S., 2013. Economic Benefits of Increasing Electric Grid Resilience To Weather Outages. Tech. Rep. August, White House Office of Science and Technology, Washington D.C.

[23] Carlson, L., Bassett, G., Buehring, M., Collins, M., Folga, S., Haffenden, B., Petit, F., Phillips, J., Verner, D., and Whitfield, R., 2012. "Resilience: Theory and

Copyright © 2018 by ASME

Applications". *Anl/Dis-12-1*(January), pp. 1–42.

[24] Panteli, M., and Mancarella, P., 2015. "Modeling and evaluating the resilience of critical electrical power infrastructure to extreme weather events". *IEEE Syst. J.,* ***PP***(99), pp. 1–10.

[25] Electric Power Research Institute, 2016. Electric Power System Resiliency: Challenges and Opportunities. Tech. rep., Electric Power Research Institute, Palo Alto, CA.

[26] Read, W., 2007. The Great Coastal Gale of December 1-3, 2007, dec.

[27] Islam, A., 2009. "Smart Grid Reliability Assessment Under Variable Weather Conditions by Arif Islam A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy Department of Electrical Engineering College of Engineering Univer". *UMI Diss. Publ.*, p. 172.

[28] Fern, X., 2017. CS 543 Machine Learning Lecture Slides.

[29] Bonneville Power Administration. Outage and Reliability Reports.

[30] Hatfield Marine Science Station. Hatfield Marine Science Center Weather Station.

[31] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., , and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A., and Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E., 2011. "Scikit-Learn: Machine Learning in Python". *J. Mach. Learn. Res.,* ***12***, pp. 2825–2830.

Copyright © 2018 by ASME