

A Statistical Forecast Model of Weather-Related Damage to a Major Electric Utility

BRIAN J. CERRUTI* AND STEVEN G. DECKER

*Department of Environmental Sciences, Rutgers, The State University of New Jersey,
New Brunswick, New Jersey*

(Manuscript received 5 January 2011, in final form 13 October 2011)

ABSTRACT

A generalized linear model (GLM) has been developed to relate meteorological conditions to damages incurred by the outdoor electrical equipment of Public Service Electric and Gas, the largest public utility in New Jersey. Utilizing a perfect-prognosis approach, the model consists of equations derived from a backward-eliminated multiple-linear-regression analysis of observed electrical equipment damage as the predictand and corresponding surface observations from a variety of sources including local storm reports as the predictors. Weather modes, defined objectively by surface observations, provided stratification of the data and served to increase correlations between the predictand and predictors. The resulting regression equations produced coefficients of determination up to 0.855, with the lowest values for the heat and cold modes, and the highest values for the thunderstorm and mix modes. The appropriate GLM equations were applied to an independent dataset for model validation, and the GLM shows skill [i.e., Heidke skill score (HSS) values greater than 0] at predicting various thresholds of total accumulated equipment damage. The GLM shows higher HSS values relative to a climatological approach and a baseline regression model. Two case studies analyzed to critique model performance yielded insight into GLM shortcomings, with lightning information and wind duration being found to be important missing predictors under certain circumstances.

1. Introduction

The weather can have a significant impact on electric utility operations. Hurricanes can cause widespread equipment damage, resulting in power outages that often last for days (Liu et al. 2008; Han et al. 2009). Ice and snowstorms have been responsible for significant outages, which typically occur when ice- and snow-laden trees fall or are blown onto overhead electrical equipment or when ice accrues onto transmission lines (Changnon and Karl 2003; DeGaetano et al. 2008). Thunderstorms can also cause destruction of electrical equipment through damaging wind gusts, large hail, and lightning (Rakov and Rachidi 2009; Li and Treinish 2010; Treinish et al. 2010). The impact of a storm, in terms of total overhead

equipment damage, is often exacerbated in densely populated areas because the equipment coverage is more dense, allowing an isolated event to cause more damage than would a similar event that occurs in an area serviced by a rural utility company (Cerruti et al. 2009; Li and Treinish 2010; Treinish et al. 2010).

Public Service Electric and Gas (PSE&G) is the largest public utility in New Jersey, the most densely populated state in the United States (U.S. Census Bureau 2010), and provides service to several large urban areas. PSE&G also serves some of the largest waste management facilities, hospitals, and schools in the state. Electrical service to such venues is dependent upon prompt restoration time, which is typically a function of workforce preparedness. Adverse weather can have a dramatic impact on PSE&G workforce operations; a typical overtime crew can cost \$2,400 per hour. A large cleanup effort may cost over \$1,000,000, not counting the costs of materials, especially if the utility is unprepared to handle the event (Cerruti et al. 2009).

This study presents a statistical damage model, hereinafter called the *infrastructure damage model*, which can be used as a guidance tool to allow forecasters to

* Current affiliation: NOAA/NWS/Meteorological Development Laboratory, Silver Spring, Maryland.

Corresponding author address: Steven G. Decker, Dept. of Environmental Sciences, Rutgers, The State University of New Jersey, 14 College Farm Rd., New Brunswick, NJ 08901.
E-mail: decker@envsci.rutgers.edu

predict the damage to utility equipment in advance of adverse weather. In this case, the statistical relationships arise from multiple linear regression, with weather observations (or forecasts thereof) as the predictors and PSE&G infrastructure damage as the predictand. A well-constructed infrastructure damage model should allow utility companies to be more adequately prepared in terms of staffing and materials for impending adverse weather. Section 2 provides background material that places the current study in the context of related work. Section 3 describes the data and methods that were used in developing the infrastructure damage model. Section 4 validates the model and finds good overall performance, whereas section 5 finds shortcomings within the context of two case studies. Section 6 discusses issues related to implementing the model in practice, and section 7 concludes the work.

2. Background

a. Regression methods

Many attempts have been made to use statistical methods for weather prediction, most notably the method known as model output statistics (MOS). This method uses multiple linear regression to correct deterministic model forecast errors at various lead times (Glahn and Lowry 1972; Glahn 1985). The main advantage to MOS is that systematic errors in the model calculations, including model bias, are statistically corrected, allowing for a more accurate forecast (Glahn and Lowry 1972).

MOS is not without shortcomings, however. First, MOS needs a large dataset of model forecast variables as input for the regression. Second, the model itself should remain constant, which implies that either new equations be derived for every version of the model or improvements not be implemented to preserve the integrity of the statistical forecasts. An additional drawback is that, at longer lead times, MOS approaches the local climatological mean (Vannitsem and Nicolis 2007).

The perfect-prognosis (PP) approach avoids these drawbacks, and the following provides justification for selecting the PP method in this study. Perfect-prognosis forecasts can be similar to MOS in terms of using regression to make a statistical forecast, but the PP method uses observed fields, and not model-predicted fields, to create the statistical forecast equations, eliminating the need for a model dataset. Thus, unlike MOS, the PP method has an advantage because the numerical weather prediction model may be upgraded as needed without affecting the statistical model. In fact, whereas model upgrades may degrade the reliability and accuracy of MOS-based forecasts, the PP method will likely yield

a better forecast, assuming that the model upgrades improve the accuracy of the output (Wilks 2006). The exclusion of model input data during regression development also makes the PP approach more computationally efficient. The PP approach is not without drawbacks, however; the observed fields used to develop the PP statistical model must be replaced with predictions of those fields when put into practice. For example, if temperature is a predictor in the PP model, the performance of the PP model will be degraded if the temperature forecasts used to drive the model are of poor quality.

The PP approach has recently been applied to make short-term lightning-activity predictions in the United States (Bothwell 2002; Shafer and Fuelberg 2006). It has also been used to develop the regression equations in Hansen (2007) that predict ceiling and visibility on the basis of temperature, humidity, wind speed and direction, and precipitation. Obled et al. (2002) used the PP assumption to validate an analog quantitative precipitation forecast method with geopotential height fields across western Europe as the discriminator for identifying the proper analog, and they found little difference in terms of model performance between using short-term forecasts of geopotential height and using observations.

b. Statistical forecasting applications for utilities

In general, statistical models in the utility industry are for predicting peak electrical loads at various lead times for varying time windows (e.g., Hippert et al. 2001). Several studies have attempted to enhance utility response to significant events, however. Brown et al. (1997) studied the reliability of distribution systems during high-wind events using Monte Carlo simulations and found wind speed and wind duration to be essential for assessing system reliability. Han et al. (2009) used a general additive model to predict power-outage risk from measured characteristics of hurricanes and were able to outperform regression-based approaches. The potential for ice storms and hurricanes to produce power outages is addressed by Liu et al. (2008) with a generalized linear mixed-model approach, which found maximum wind gust and maximum ice thickness to be important predictors for hurricanes and ice storms, respectively. DeGaetano et al. (2008) used a precipitation-type algorithm adjustment to 6–12-h Weather Research and Forecasting (WRF) model forecasts to predict ice accretion to utility equipment in terms of outages. Although limited outage data were available for model verification, the model results generally matched the observations. Zhu et al. (2007) formulated a cumulative outage model as a baseline prediction that could be augmented by an observer to correct model errors. This method provided accurate

outage forecasts, but with a lead time of only 5 h because of the need for real-time outage data to be available for the augmentation. In general, these models are designed to aid distribution managers in deciding what additional staffing and equipment may be needed in advance of a weather event.

Some attempts at explicitly predicting infrastructure damage have recently been made. Takata et al. (2005) produced forecasts of damaged poles and transmission lines from typhoons on the basis of 29 cases. Using the typhoon forward speed, maximum wind speed, central pressure, and wind radii as input, they combined the use of neural networks and nonlinear regression (where *nonlinear* indicates that transformations upon the data were performed prior to regression). These two methods when combined outperformed each method individually, with an error decrease of up to 40%. Treinish et al. (2010) used weather observations from the New York City metropolitan area and power-outage information from a local utility company to create a model to predict the outage-related infrastructure damage caused by weather events. A statistical damage model was derived using meteorological data from Deep Thunder (DT; a version of WRF with three domains at 18-, 6-, and 2-km grid spacing) as the predictor and infrastructure damage as the predictand. The statistical damage forecast model was coupled with DT output to predict infrastructure damage in a MOS-style format operationally. Li and Treinish (2010) expanded on this work by creating a statistical model to predict the spatial extent, duration, and severity (as measured by number of customers without power) of power outages, again using DT forecasts as the predictor. The statistical outage model is stratified according to season and the magnitude of certain weather variables such as precipitation.

The DT-based models attempted to account for uncertainty from several areas, including DT initial-condition and model errors, errors in the observations used to calibrate a wind gust statistical model coupled to DT, and possible input variable errors in the training dataset used for the regression equations. These numerous and often substantial sources of uncertainty may be to blame for the model's modest performance and large uncertainty forecasts.

3. Data and methods used

This study uses multiple linear regression to obtain equations that relate surface weather observations to infrastructure damage observations. These equations are then used as a PP forecast tool that, given a *forecast* of the same weather observables used to create the regression equations, predicts the amount of infrastructure

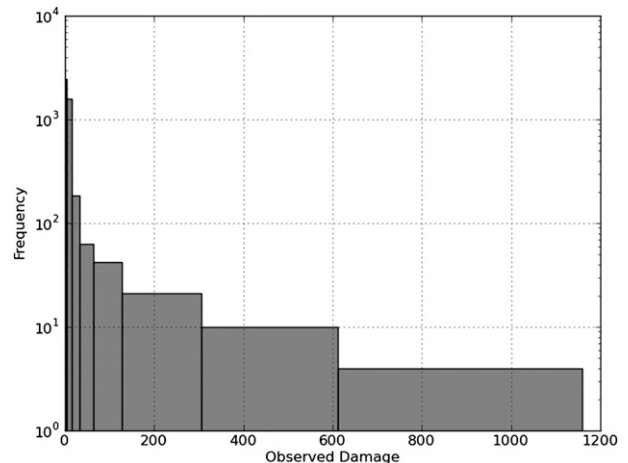


FIG. 1. Histogram of observed infrastructure damage counts over all days and regions.

damage that would be expected if the weather forecast were to verify. We choose multiple linear regression because it is often the basis for other statistical approaches (Wilks 2006).

a. Data sources

The training data for this study come from a unified damage database compiled by PSE&G that covers the period from 1 January 2003 to 31 October 2008. This database is composed of counts of damage to transformers, poles, trees, and service, secondary, and primary wires,¹ which are the six infrastructure elements for which forecasts of damage counts are desired. A frequency histogram of damage counts within this dataset is presented in Fig. 1, which shows the data to be positively skewed (i.e., Poisson distributed). This situation suggests that logarithmically transforming the predictand will produce the best regression results. The Poisson distribution is typical of count data and implies that a generalized linear model, as used in section 4, will outperform a least squares regression (Han et al. 2009).

PSE&G's electrical operations are divided into the four regions of Palisades, Metropolitan, Central, and Southern (Fig. 2). To simplify the process, one weather station per region supplies predictors, similar to original MOS methods of using the closest model grid point to represent a weather station (Glahn and Lowry 1972). The stations chosen to represent the four regions were Terborro Airport (TEB), Newark Liberty International

¹ Service wires connect poles to buildings at 120 V (service voltage). Secondary wires connect poles to poles at service voltage. Primary wires connect poles to poles at higher (distribution) voltages.

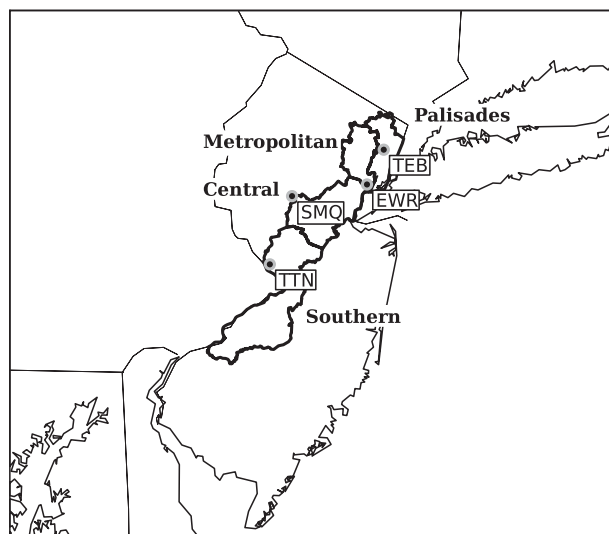


FIG. 2. Map showing the location of the PSE&G regions and the weather stations used to construct the infrastructure damage model.

Airport (EWR), Somerset Airport (SMQ), and Trenton Mercer County Airport (TTN), as shown in Fig. 2. These stations are not ideally located, but they represent the most reliable first-order stations within each region. In an ideal situation, multiple stations would have been utilized for equation development, but the lack of more-detailed damage-location information limits this study to assigning a single station per region; PSE&G records assign damage to its respective region and not to the closest weather station.

Daily surface data for the aforementioned stations were obtained from the National Climatic Data Center's daily climate summaries. These data included maximum and minimum temperature, average dewpoint, average temperature, liquid equivalent precipitation, maximum wind gust, and observed weather. To identify the occurrence of severe thunderstorms, local storm reports were obtained from the Storm Prediction Center (SPC) Storm Reports database (see online at <http://www.spc.noaa.gov/climo/online/#reports>).

b. Stratification

An initial multiple linear regression without stratification yielded poor coefficient-of-determination R^2 values of less than 0.16 for each infrastructure element. It was found that regression performance could be improved by using an objective assessment of the daily weather summary observations in each region to stratify each day into one of six weather modes. The values used as stratification thresholds, defined below, were taken from PSE&G weather-alert criteria (Cerruti et al. 2009) so that the infrastructure damage model would seamlessly

integrate with PSE&G operations. The method consists of assigning to each day² a particular weather mode for which a unique equation will be developed for each PSE&G region and infrastructure element; this approach results in 144 separate equations (six modes \times four regions \times six elements). The following criteria determine to which weather mode each day and region are uniquely assigned.

The "thunderstorm" mode was diagnosed when thunderstorms were observed at the station, when thunderstorms were observed near the station [code VCTS in an aviation routine weather report (METAR)], or if severe weather occurred within the region according to local storm reports. The inclusion of data from local storm reports partially accounts for events in which thunderstorms hit the region but not the station. As a result, thunderstorm-mode days can occur without reports of thunder at the designated station. Within the training dataset, this occurred on anywhere between 7 and 16 days, depending on the region.

The "warm" mode was diagnosed if the only form of precipitation observed was rain and accumulated precipitation was greater than 0.25 mm (0.01 in.). The "cold" mode was diagnosed if only wintry precipitation was observed (such as snow, freezing rain, or sleet) with precipitation of at least 0.25 mm (0.01 in.) liquid water equivalent. The "mix" mode was diagnosed if a combination of warm and cold modes occurred (rain *and* at least one of snow, sleet, or freezing rain) with liquid equivalent precipitation of at least 0.25 mm (0.01 in.). Precipitation of exactly 0.25 mm (0.01 in.) was included for cold and mix modes and excluded for the warm mode to account for the notable undercatch of precipitation gauges during times of freezing or frozen precipitation (Yang et al. 1998).

The "heat" mode was diagnosed if maximum temperatures exceeded 32°C (90°F) and measured precipitation was no more than 0.25 mm (0.01 in.). A no-weather ("none") mode was diagnosed if none of the previous criteria were met. Upon further investigation into the none mode, it was found that stratification by wind gust magnitude greater than 12 m s⁻¹ (27 mi h⁻¹) increased R^2 values; thus, the "wind" mode was created. The none-mode days were omitted from the regression since the damage on these days is presumably not weather related. An eighth mode ("questionable") emerged as a result of our use of daily weather summaries. These data only record the sensible weather type if it occurs at the top of the hour. Therefore, precipitation falling between these

² In this study, *day* refers to the 24-h period extending from midnight to midnight LST.

TABLE 1. The relative frequency of each weather mode (percentage of days) for the training dataset (1 Jan 2003–31 Oct 2008) after removing days deemed in the questionable mode (defined in section 3b). Percentages for questionable are relative to all days. Regions are abbreviated so that CEN, MET, PAL, and SOU represent Central, Metropolitan, Palisades, and Southern, respectively.

Mode	CEN	MET	PAL	SOU	Overall
Thunderstorm	8	6	6	8	7
Warm	22	27	26	23	25
Mix	2	2	2	2	2
Cold	5	4	5	5	5
Heat	4	5	5	3	4
Wind	7	17	12	9	11
None	52	39	43	50	46
Questionable	9	2	3	8	5

observations is recorded by the rain gauge, but the exact precipitation type is unknown. Thus, the storm mode cannot be known exactly, and these days are disqualified from the training dataset. Table 1 presents a summary of the frequency of occurrence of each mode within the training dataset.

c. Postprocessing

The damage database is a collection of field reports that have been compiled on a daily basis. The data are collected as they are reported, which means damage reports from a significant event, such as widespread severe thunderstorms or a tropical cyclone, may not appear until the next day. For particularly extensive events, it may take several days for all of the damage from that event to be reported since restoration efforts can be impeded by flooded roads, snow drifts, stalled cars, and downed trees blocking roadways (Cerruti et al. 2009; Treinish et al. 2010). Thus, the infrastructure element damage data were postprocessed to account for any potential lag in reporting during extreme weather. Significant events considered for postprocessing include heavy snowstorms, ice storms, thunderstorms, and cyclones that brought flooding or damaging winds.

Days may be postprocessed if the surface observations indicate a substantial storm occurred but damage is higher on the following day(s) and these following day(s) are classified as being in the none mode. The number of days available for postprocessing is limited to all consecutive none-mode days of observed damage with greater than 20 total damaged elements until a weather mode other than none is diagnosed. The damage observations may not be adjusted for days *not* diagnosed as none mode. In cases in which adjacent days incur damage and the none mode is not diagnosed on the following day, each day is assigned to the proper weather mode as per section 3b. Damage-element counts were set

to zero for all nonevent days that were postprocessed, and these days were omitted from the regression.

For example, 10 June 2008 was a high-end severe-weather day that required postprocessing. On that day, the thunderstorm mode was diagnosed for the Central, Metropolitan, and Palisades regions followed by three consecutive none-mode days, but the observed damage was spread out over multiple days. In Palisades, for example, 287 elements were reported as damaged on 10 June, with 517, 237, and 78 damaged elements reported for 11, 12, and 13 June, respectively. Similar patterns existed in the Central and Metropolitan regions. During postprocessing, the damage from 11 to 13 June was set to zero, these days were changed from none mode to questionable, and the sum of the damage from 10 to 13 June replaced the original damage observed for 10 June. Therefore, for Palisades, the 10 June damage is now 1119 elements. A total of 106 days over the four regions were adjusted according to this method, representing 1.25% of the 7508 days in the dataset when considering each of the four regions separately. Thunderstorm-mode days represent 91.5% of these adjustments.

d. Predictors

The regression analysis performed here used multiple predictors as a starting point in each case. The traditional meteorological variables used are maximum wind gust V_{\max} , maximum temperature T_{\max} , and liquid-water-equivalent precipitation LWE_d . Additional parameters were added to this study in an attempt to improve upon the R^2 values: 10-day accumulated liquid-water-equivalent precipitation LWE_{10} , 3-day maximum temperature sum T_3 , the number of severe-weather reports in a given region, and various storm factors SF_x .

The LWE_{10} predictor is the sum of the precipitation amounts from 1 to 10 days prior to the forecast day and was intended to serve as a proxy for the amount of moisture in the top layer of the soil, which is thought to be a contributor to downed trees and poles (Liu et al. 2008; Han et al. 2009). Soil moisture would have been used explicitly, but such a dataset with the necessary spatial coverage within the PSE&G regions did not exist at the time of model development. Several different lag times were considered during preliminary investigation, and 10 days outperformed other lag times by yielding higher R^2 values. This predictor proved particularly useful for the cold and wind modes.

The T_3 predictor was considered as a proxy for the cumulative heat to which equipment has been exposed. It is defined as the sum of the maximum temperature for the previous 2 days and the current forecast day with no regard for when in the day the maximum temperature

occurred. This method outperformed other combinations of lagged temperature sums. Maximum temperature was selected as the appropriate variable to lag because in preliminary investigations it outperformed other potential predictors such as daily average temperature, minimum temperature, and daily average wet-bulb temperature.

An analysis of these variables revealed a stronger relationship if the product of certain variables was considered. These products, the SF_x , include the product of the wind gust and liquid equivalent precipitation (SF_1), the product of the wind gust and 10-day accumulated precipitation (SF_2), and the product of the wind gust and maximum temperature (SF_3). A transformation in this manner allows for the inclusion of the interaction between two predictors. A similar method is used by Takata et al. (2005) in their nonlinear regression.

Storm factor 1 is meant to provide a measure of overall weather adversity, given that several studies cite the interaction between wind and precipitation as causing damage to overhead electrical equipment (Rooney 1967; Takata et al. 2005; Liu et al. 2008; Han et al. 2009). Storm factor 2 is meant to provide a measure of the susceptibility of the tree and root system during storms with relatively low precipitation totals. It was discovered by analyzing scatterplots of wind gusts and damage that LWE_{10} modified the outliers to improve the fit of a linear model across several storm modes and regions. This variable combines an aspect of tree and root susceptibility (LWE_{10}) with a destructive property of storms (wind gust) to create a measure of the potential for tree failure.

Storm factor 3 was discovered in the same manner as SF_2 , except by initially analyzing the thunderstorm mode. Although high temperatures surely will not cause trees to fall or poles to snap, the maximum temperature represents a proxy for several conditions favorable for convection (all else being equal, higher temperatures indicate the possibility for higher equivalent potential temperature, which suggests the possibility of higher convective available potential energy) and strong convective wind gusts (higher maximum temperatures can signify a well-mixed boundary layer, potentially steep low-level lapse rates, and possibly reduced convective inhibition, particularly on thunderstorm-mode days). Similar transformations between the wind gust and daily averaged dewpoint, daily averaged wet-bulb temperature, and minimum temperature provided less-encouraging results when included in the thunderstorm-mode regression.

The number of severe thunderstorm reports in each region SVR is used as an additional predictor for the

thunderstorm mode. It partially augments surface meteorological observations to account for equipment damage incurred from a thunderstorm that was not observed at the surface station for a given region. The reports do contain errors with timing and location (Witt et al. 1998), but they prove to be a useful predictor, as will be shown shortly. The latitude and longitude of each storm report were mapped to assign each report to an individual region. All possible severe convective events were considered, including tornadoes, hail of at least 19 mm (0.75 in.), and wind gusts of greater than 26 m s^{-1} (50 kt). If a severe-weather report fell outside all regions, it was discarded. A severe-wind report will occasionally contain comments such as “wires down” that indicate the predictor (SVR) depends in some way on the predictand (wires). Fears of such circularity may be allayed by noting that the number of wires down is rarely indicated in such ancillary information, nor is the type of wire usually indicated, and it is this information that determines the predictands. Furthermore, many severe-wind reports do not contain such comments. Thus, we consider SVR to contain information that is not present in our observations of the predictands themselves.

e. A note on using the infrastructure damage model in practice

Although all of the predictors defined previously are available as observations and can be used as such during model training and validation, applying the infrastructure damage model in practice requires forecasts of those predictors. These forecasts traditionally come from models themselves (e.g., maximum temperature could come from any NWP model), but the storm-report predictor used here is not available from any standard model output. Similarly, determining whether to apply thunderstorm mode to a future day depends on whether thunderstorms are forecast. Therefore, additional techniques augmenting NWP output are required to provide the forecasts that drive the damage model. To the extent that these techniques produce low-quality forecasts, the damage model itself will provide forecasts of degraded quality. We return to these considerations in section 6.

4. Results

a. Regression equations

The general form of the regression equations is

$$\hat{y}_{rind} = b_{0rind} + b_{1rind}k_1 + b_{2rind}k_2 + \dots + b_{nrind}k_n. \quad (1)$$

Here, \hat{y} is the model-predicted infrastructure damage and r denotes the various regions, i denotes the infrastructure

element under consideration, m denotes the weather mode, and d denotes a particular day. The regression coefficients are denoted by b and the predictors are denoted as k_1, k_2, \dots, k_n . Multiple models were developed, including 1) a model with no backward elimination, 2) a model with backward elimination, 3) a model with no backward elimination but with log-transformed predictands, and 4) a model with both backward elimination and log-transformed predictands. The fourth model performed best (not shown), and because it includes a log transformation we refer to this regression method as the generalized linear model (GLM) hereinafter.

The backward regression method consisted of repeatedly removing the predictor with the highest p value from the regression until all of the p values that remained were significant at the 95% confidence interval (Draper and Smith 1998). Because many of the predictors will be mutually correlated, a filtering technique was developed to eliminate these variables as follows. 1) After the initial backward elimination regression, all predictors are tested for mutual correlation. 2) The pair displaying the highest correlation is then tested to investigate which single predictor has the highest correlation with the predictand. 3) Of the two predictors under investigation, the one with the least significant relationship to the predictand (the one with the highest p value) is dropped from the pool of available predictors. 4) This process is repeated until all predictors in the equation display a mutual correlation coefficient of less than 0.3. In some instances, no significant predictor remained in the equation—in these cases, only the constant was taken as the equation. This constant forecast represents the climatological value (commonly referred to as the “climatology”) of infrastructure damage. Climatology is an acceptable method for forecasting and often represents a reference forecast that other methods try to beat (Wilks 2006).

As suggested in section 3a, a log transformation of the predictands was found to produce better results in terms of higher R^2 values. This log transformation is accomplished by writing (1) with \hat{Y}_{rimd} instead of \hat{y}_{rimd} , where \hat{Y}_{rimd} represents a log-transformed predictand. It is related to the original infrastructure damage predictand via

$$\hat{Y}_{rimd} = \ln(\hat{y}_{rimd} + 1). \quad (2)$$

b. Model training

The commercial R software package was used to derive coefficients for the 144 equations constituting the infrastructure damage model by using the GLM approach. Coefficients for all 144 equations are available from the

authors upon request. Many equations account for reasonable amounts of the variance of their predictands, with a maximum R^2 of 0.855 for the equation describing damage to secondary wires in the Palisades region during thunderstorm mode. On the other hand, equations in the Metropolitan region, where much of the equipment is underground, tended to account for less variance. An extreme example of this was the equation for service wires in the Metropolitan region during cold mode, where R^2 was nearly zero.

To demonstrate that the GLM is a useful model, we trained two additional models to set various performance standards. The first model (“CLIMO”) was based on climatology. CLIMO was derived by stratifying the training dataset into the same weather modes used for GLM development, but an average of the damage counts for a given region, weather mode, and infrastructure element served as the forecast instead of regression equations. The second model (“BASE”) was developed with the same training dataset but used a multiple-linear-regression method with no backward elimination, no stratification, and no log transformation.

c. Model validation

Independent data for the period 1 November 2008–15 November 2009 (hereinafter referred to as the validation dataset) were provided by PSE&G so that the model could be validated. The validation dataset is necessary because the many screening procedures that were used to define predictors and to carry out backward elimination introduce artificial skill into the model if not accounted for. Cross validation is not enough to eliminate these biases in model performance evaluation (DeSole and Shukla 2009). The validation dataset is much smaller than the training dataset for two reasons. First, a larger training dataset allows for the most robust equations possible, because having more data means having less opportunity for overfitting. Second, PSE&G valued the development of robust equations and rapid project completion over taking the time necessary to collect the additional data needed to form a validation dataset of comparable size.

Daily weather observations and local storm reports were obtained for the appropriate weather stations and PSE&G regions, respectively, and the infrastructure damage data were postprocessed according to the methods described in section 3c. A summary of weather-mode frequency for the validation dataset can be found in Table 2. The training and validation datasets both show similar properties, with the none mode occurring most often, warm mode occurring second most often, and the cold and heat modes occurring less frequently. The GLM, CLIMO, and BASE models discussed

TABLE 2. As in Table 1, but for the validation dataset. These data represent the period 1 Nov 2008–15 Nov 2009. A total of 380 cases were analyzed in each region.

Mode	CEN	MET	PAL	SOU	Overall
Thunderstorm	7	6	3	7	6
Warm	29	31	35	31	32
Mix	2	2	4	2	3
Cold	6	5	5	5	6
Heat	1	2	1	1	1
Wind	9	16	11	11	12
None	45	38	39	42	41
Questionable	18	4	12	9	10

previously were applied to produce hindcasts of infrastructure damage, which were compared with the observed damage.

Figure 3 shows a box plot of the damage counts in each of the six modes in the training and validation datasets. Notice that the validation dataset has fewer high-end damage events than the training dataset does for most storm modes, suggesting that model performance as measured by the validation dataset may be hindered by sampling error in these modes. The cold and mix storm modes display the most obvious sampling error as their respective maximum damage counts differ by an order of magnitude. The thunderstorm and warm modes also suffer from sampling error, but to a slightly lesser degree.

The ability of the CLIMO, BASE, and GLM models to make accurate predictions for more than 5, 10, 20, 30, 40, 50, 75, or 100 total damaged infrastructure elements was assessed using the probability of detection (POD), false-alarm ratio (FAR), critical success index (CSI), and Heidke skill score (HSS). Specifically, these scores assess the models' ability to forecast a binary event: Given a forecast above some threshold, was the observed damage above some threshold? For more information on these scores, the reader is referred to Roebber (2009). To provide additional information on forecast quality, curves of relative operating characteristic (ROC) were calculated for each observed damage threshold, and the area under the ROC curve (AUC) was computed. A perfect forecast system would have an AUC score of unity, whereas $AUC = 0.5$ corresponds to skill equivalent to random chance. See Marzban (2004) for more information on ROC curves. To provide a summary of model performance, verification scores were determined by using the sum of the damage counts for the six infrastructure elements, and the POD, FAR, CSI, and HSS diagnostics were derived from the contingency table corresponding to the point on each ROC curve at which the HSS was maximized. This method was applied to all days of the independent dataset for each of the three models discussed.

Table 3 shows that CLIMO proved to produce positive skill ($HSS > 0$ and $AUC > 0.5$ for all thresholds). Therefore, other models scoring higher can be said to be skillful relative to climatology, a common benchmark (Wilks 2006). The BASE model shows skill and accuracy for all metrics as well; therefore, it represents a regression-based standard for the GLM to beat. According to Table 3, the HSS skill metric for the BASE model is higher than the values presented for the GLM for thresholds of 20, 30, 40, and 50 total damaged elements. However, the AUC values are always higher for the GLM relative to the BASE and CLIMO models. Figure 4 shows the actual ROC curves for each model using a threshold of 30 total damaged elements from which the skill scores were calculated. Although BASE is superior for intermediate forecast thresholds (the BASE curve is slightly farther to the left than GLM at center left), the GLM more closely follows the left edge of the plot for POD values of less than 0.4 and greater than 0.7. This indicates that, given at least 30 observed damage elements, the GLM has a superior ability to limit false alarms when forecast thresholds are low. In addition, the GLM outperforms BASE in all metrics for thresholds of 75 and 100, which demonstrates that the GLM is superior to the BASE model for higher damage thresholds. Therefore, on the basis of analysis of Table 3 and Fig. 4, the GLM has a superior ability to limit false alarms for low- and high-impact events and to detect high-impact events more skillfully than the CLIMO and BASE models.

Now that the superiority of the GLM is established in terms of more-skillful forecasts, with an emphasis on higher-damage events and minimizing false alarms, we turn our attention to a measure of reliability similar to that employed by Glahn et al. (2009). Figure 5 shows a histogram of the bias for both the GLM and BASE models, where the bias has been conditioned on particular forecast intervals. The bias is expressed in terms of the relative median count, that is, the median observation corresponding to a particular forecast bin divided by the midpoint of that bin. Values of the relative median count that are close to 1 indicate low conditional bias and, hence, high forecast reliability. Figure 5 reveals a clear bias in the BASE model, which overforecasts damage counts in each bin. On the other hand, the GLM shows relatively little bias.

Overall reliability can be assessed through the squared bias in relative frequency (SBRF) score, given by

$$\sum_i (1 - \text{bias}_i)^2,$$

where i represents a particular forecast bin. In this case, scores near 0 indicate reliable forecasts. As suggested by

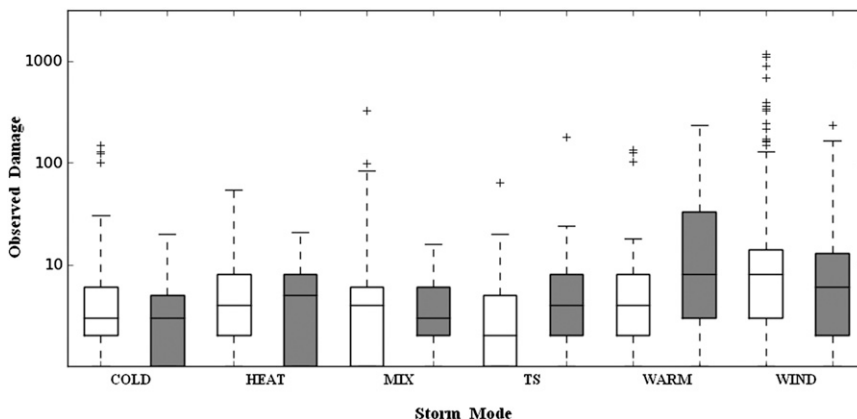


FIG. 3. Box plot showing the total observed damage for all regions in each mode for the training (white) and validation (gray) datasets. The top and bottom of each box represent the first and third quartiles, respectively, and the middle line represents the median. The bottom and top whiskers extend to points no further than 1.5 times the interquartile range away from the first and third quartiles, respectively. The points denoted by plus signs are outliers.

Fig. 5, the GLM is the more reliable model overall, with an SBRF score of 0.214 as compared with 1.283 for the BASE model.

It is important that the GLM’s skill is higher than the BASE model for the high-end events; this situation demonstrates the goal of a utility to discriminate properly between high-end events and non-high-end events is attainable. Because we have demonstrated the GLM’s skill in reliably detecting and discriminating high-end events from non-high-end events, we conclude that our model would provide useful information upon application. We now focus on applications and shortcomings of the model by analyzing case studies from the validation period.

5. Case studies

The following is a pair of case studies intended to show the functionality of the infrastructure damage model on a per-region basis for two selected storm modes. The authors chose these cases to highlight shortcomings specifically. Therefore, the results from these isolated cases do not reflect the overall successful performance of the GLM, as discussed in section 4. For each case study, the model results are compared with observed infrastructure damage per region, and a discussion follows.

a. Thunderstorm mode: 9 June 2009

Thunderstorms affected the PSE&G regions in the early morning hours of 9 June 2009 with frequent lightning, wind gusts of 7–14 m s⁻¹ (16–31 mi h⁻¹), rainfall of 5–27 mm (0.2–1.05 in.), and a report of severe hail [19 mm (0.75 in.)] in the Southern region. The

appropriate regression equations were applied for each region to the surface observations from each weather station using the thunderstorm mode. The results are summarized in Table 4, with an example of the usage of a particular equation given in Table 5.

The model performed well in the Central, Palisades, and Metropolitan regions by correctly predicting a low damage total. The model badly underestimated the damage for the Southern region, however. Other local stations, including Philadelphia International Airport (PHL), South Jersey Regional Airport (VAY), Northeast Philadelphia Airport (PNE), Millville Municipal Airport (MIV), and McGuire Air Force Base (WRI) were interrogated to determine the cause of model error. Maximum wind gusts were only 8–11 m s⁻¹ (18–25 mi h⁻¹) at these additional stations, and daily precipitation was generally below 7.6 mm (0.30 in.), except at PHL where 25 mm (1.00 in.) fell.

Because these values do not differ substantially from the values observed at TTN, an analysis of nearby local severe-storm reports was carried out to determine whether severe weather was reported close to the Southern region. This analysis yielded two wind-related storm reports in northern Delaware. The timing of the Delaware severe thunderstorms differs by approximately 5 h from the timing for the New Jersey thunderstorm observations, which indicates two separate thunderstorm areas occurred. Radar observations confirm this. Further analysis of New Castle Airport’s (ILG) observations indicated that the main lightning activity from this storm remained south of the station according to METARs. An analysis of lightning activity recorded by the aforementioned stations yielded multiple hours of lightning activity in and

TABLE 3. Summary of forecast verification scores for thresholds of 5, 10, 20, 30, 40, 50, 75, or 100 observed infrastructure damage elements in a day for each PSE&G region. The values shown are derived from ROC curves and represent the information corresponding to the point on the ROC curve with the maximum HSS values. The model types are described in section 4a.

Model	Metric	Damage threshold							
		5	10	20	30	40	50	75	100
GLM	POD	0.64	0.44	0.60	0.62	0.58	0.47	0.50	0.56
	FAR	0.47	0.44	0.48	0.55	0.62	0.64	0.77	0.77
	CSI	0.41	0.33	0.38	0.35	0.30	0.26	0.19	0.19
	HSS	0.45	0.46	0.54	0.51	0.44	0.40	0.30	0.31
	AUC	0.82	0.88	0.93	0.96	0.95	0.95	0.95	0.94
BASE	POD	0.68	0.44	0.64	0.67	0.63	0.59	0.30	0.33
	FAR	0.63	0.65	0.45	0.52	0.59	0.66	0.82	0.82
	CSI	0.31	0.24	0.42	0.39	0.33	0.28	0.13	0.13
	HSS	0.26	0.35	0.58	0.55	0.49	0.42	0.21	0.22
	AUC	0.72	0.80	0.92	0.93	0.92	0.91	0.90	0.90
CLIMO	POD	0.92	0.64	0.37	0.43	0.31	0.33	0.16	0.21
	FAR	0.58	0.76	0.86	0.89	0.92	0.93	0.92	0.92
	CSI	0.41	0.21	0.11	0.09	0.07	0.06	0.05	0.06
	HSS	0.34	0.22	0.14	0.12	0.09	0.09	0.08	0.10
	AUC	0.73	0.72	0.69	0.69	0.68	0.68	0.67	0.72

around the Southern region that coincide with the timing of the observed thunderstorms. Therefore, the likely error source is the omission of lightning data in the damage model and not unreported severe thunderstorms. Indeed, ancillary data show that utility repair workers attributed much of the damage in this region to lightning.

b. Warm mode: 11 September 2009

A weak surface cyclone formed late on 10 September 2009 and tracked across New Jersey from south to north through the day of 11 September 2009 while dissipating. This system was responsible for wind gusts of $14\text{--}17\text{ m s}^{-1}$ ($31\text{--}37\text{ mi h}^{-1}$), rainfall of $13\text{--}38\text{ mm}$ ($0.5\text{--}1.5\text{ in.}$), and maximum temperatures of $18^{\circ}\text{--}20^{\circ}\text{C}$ ($64^{\circ}\text{--}68^{\circ}\text{F}$) across the area. The damage model's warm mode was applied for each region (Table 6).

The model underestimated the damage to all regions for this case, with the most-severe errors occurring for the Southern and Central regions. In all regions, the wind direction was northeasterly or easterly for the duration of the storm, whereas climatologically the winds are westerly in New Jersey. The winds during this event also were of a longer duration as all stations measured sustained winds of at least 9 m s^{-1} (20 mi h^{-1}) for longer than 11 h. The relatively high magnitude of the maximum wind gust combined with the anomalous direction and longer duration of the wind likely combined to cause damage that was higher than the model prediction since neither wind direction nor duration are predictors in the model.

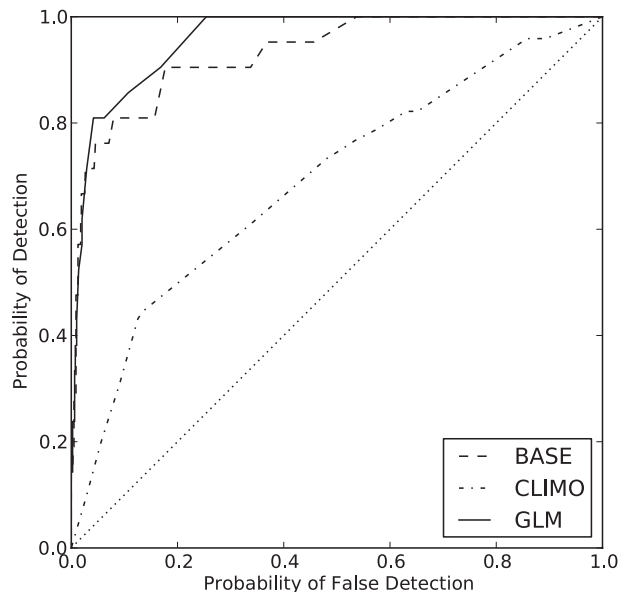


FIG. 4. ROC curve for the GLM, BASE, and CLIMO models for an observed damage threshold of 30 total damaged elements.

6. Discussion

Although the infrastructure damage model performs well during validation, applying the model to a real-world operational scenario necessitates forecasting the predictors. To the extent those forecasts are imperfect, the skill scores shown previously will not be attained. Although it is beyond the scope of this study to determine what source for those forecasts is best (i.e., which NWP or MOS product gives the best temperature forecasts?), the number of storm reports is not something that is routinely forecast by any NWP model, necessitating further discussion.

In lieu of model forecasts of storm reports, a reasonable method for forecasting the number of storm reports can be made by analyzing SPC convective outlooks, which are based in part on interpreting model output [for more information on SPC outlooks, see Ostby (1992) and Edwards et al. (2002)]. Specifically, the probabilistic forecasts for the occurrence of tornadoes, severe wind, and severe hail can be summed and divided by 20, ignoring the percent sign and rounding down. For example, on a low-end “slight risk” day, the forecast probabilities for tornadoes, severe wind, and severe hail may be 0%, 15%, and 15%, respectively. Summing and dividing leads to 1.5; rounding down results in a forecast of one severe report. On a high-end “moderate risk” day, the forecast probabilities may be 15%, 45%, and 60%, respectively, leading to a forecast of six severe reports. This method is based on a subjective analysis of past correspondences between SPC

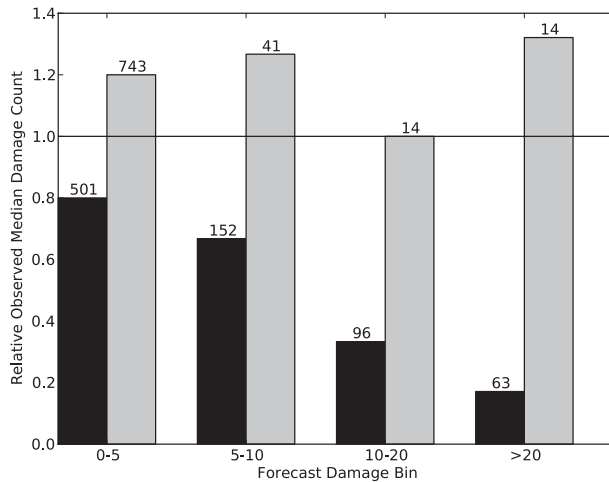


FIG. 5. Relative median count histograms for the BASE model (black) and GLM (gray). The numbers above each column indicate the case count.

convective outlooks and the resulting number of storm reports in each region.

Although current predictions of storm reports may be of questionable quality, advances in NWP suggest that higher-quality forecasts may soon be available. For instance, the High Resolution Rapid Refresh (HRRR) model generates hourly output at a storm-resolving 3-km grid spacing (Smith et al. 2008). The output includes products such as simulated radar reflectivity and updraft helicity that can be used to forecast thunderstorm location and severity (Stensrud et al. 2009). As such, a method predicting storm reports from HRRR output may be feasible in the near term. Development of this method is left for future work. Because of the PP approach used in developing the infrastructure damage model, no change to the underlying statistical model will be necessary to incorporate such improvements.

7. Conclusions

We have developed and presented verification of an infrastructure damage model for use by PSE&G in preparation for adverse-weather operations. The predictors are derived from daily weather summaries for four selected stations within the PSE&G service territory, one for each region. The predictors include traditional meteorological variables such as maximum temperature and maximum wind gust and transformations to account for interactions between the main variables. The three storm factors (one-third of the possible predictors) accounted for predictors in 95 (66%) of the GLM’s 144 equations, indicating their importance. The number of severe-thunderstorm reports was used in every equation

TABLE 4. Complete damage count data for 9 Jun 2009. Included are the observed damage and the corresponding model prediction for each damage element using the thunderstorm mode. The infrastructure damage elements are labeled on the left side as follows: TRANS is the number of transformers damaged, POLE is the number of poles damaged, SVC is the number of service wires replaced, SEC is the number of secondary wires replaced, PRI is the number of primary wires replaced, and TREE is the number of tree-related damage incidents. Boldface values indicate where the model error is two total infrastructure damage elements or less.

	CEN		MET		PAL		SOU	
	Obs	Model	Obs	Model	Obs	Model	Obs	Model
TRAN	9	2	0	0	2	3	18	3
POLE	0	1	1	0	2	1	1	2
SVC	0	1	0	0	0	3	6	3
SEC	0	0	0	0	0	1	12	1
PRI	2	1	0	0	0	2	27	2
TREE	2	1	0	0	0	1	25	2

for the thunderstorm mode for the Central and Southern regions. Of interest is that it was found to be mutually correlated with T_{max} for the Palisades and Metropolitan areas during mutual correlation screening and was subsequently dropped from those equations, which provides further evidence for the usefulness of SF_3 since T_{max} is one of the multiplicands. Therefore, while we employ a subjective method for forecasting storm reports, this technique is only necessary for the thunderstorm mode in the Central and Southern regions, representing 12 out of the 144 total equations. The maximum observed daily wind gust V_{max} is included in 52 (36%) of the 144 equations, which demonstrates the importance of wind gusts for predicting infrastructure damage to overhead electrical distribution equipment.

Of the three models considered, the GLM was found to be the most reliable, accurate, and skillful. The relatively higher AUC values, HSS values of greater than 0 for all thresholds considered, superior HSS values for the GLM relative to the other two approaches, and a low SBRF score support these claims. Therefore, it is clear that, assuming accurate forecasts of the predictors can be made consistently, an electrical utility receiving the output from the GLM would be able to predict high-end damaging events with skill and have an ability to discriminate between high-end and low-end events, leading to increased preparedness and the most efficient use of resources and time.

a. Shortcomings

The infrastructure damage model documented in this study shows encouraging results, but a number of shortcomings have been revealed. For instance, unlike some previous work, the data were not stratified by season. This may affect the LWE_{10} predictor in particular because it is

TABLE 5. An example of the calculations required to use the infrastructure damage model, as applied to the number of damaged transformers expected in the Central region for thunderstorm mode on 9 Jun 2009. On this day, SMQ observed 0.2 in. of precipitation and a 16 mi h⁻¹ wind gust. Units of the coefficients are such that all units cancel; the final result is a damage count.

Variable	Value	Coef	Product
Constant	1	0.062	0.062
SF ₁	3.2 in. mi h ⁻¹	0.007	0.022
T ₃	238°F	0.004	1.024
SVR	0	0.329	0.000
		Total (\hat{Y})	1.11
		Forecast (\hat{y})	2.03

not corrected for precipitation type. This would likely yield an overestimate of soil moisture after snow events because the precipitation is in solid form and would not seep into the soil until the snowpack melts. This may cause some overestimations (underestimations) in damage forecasts following heavy snow events for equations in which the coefficients for LW_{E10} or SF₂ are positive (negative). The presence of snowfall or ice on overhead wires and tree limbs is also not accounted for, nor is the ability to determine whether trees have foliage. A seasonal stratification would help to account for these factors as well.

Two case studies assessed potential shortcomings in the model. Lightning can have a destructive impact on utility equipment, as shown by Bothwell (2002), Shafer and Fuelberg (2006), and Zhu et al. (2007). The lack of lightning activity in the model may be a substantial shortcoming, as was shown by the 9 June 2009 case study's underestimation of damage associated with several hours of lightning activity.

The 11 September 2009 case study revealed wind duration and direction to be possible causes of model error—a result that is also supported by the findings of Brown et al. (1997). When testing wind duration as a predictor, however, the number of consecutive hours with wind speeds of greater than 9 m s⁻¹ (20 mi h⁻¹) yielded only marginally better results in terms of R^2 values, and this duration predictor was often rejected via the backward elimination technique. We speculate that Brown et al. (1997) found a greater dependence on wind duration and direction because they studied an area in Washington where windstorms are by far the most damaging weather phenomena.

Because the predictors used in this study are to be a daily maximum value or representative of an entire day, and hourly wind direction changes can provide for difficult average wind direction forecasting, we leave accounting for wind direction for future work. Further stratification of the dataset into subsets by direction

TABLE 6. As in Table 4, but for 11 Sep 2009 and using the warm mode.

	CEN		MET		PAL		SOU	
	Obs	Model	Obs	Model	Obs	Model	Obs	Model
TRAN	3	3	0	0	4	1	4	2
POLE	10	2	0	0	3	1	8	2
SVC	13	2	0	0	7	1	14	2
SEC	1	1	0	0	0	0	4	1
PRI	6	2	0	0	1	1	14	2
TREE	13	4	3	1	0	0	9	1

resulted in case counts that were too low to obtain a reliable regression equation.

b. Future work

Future plans include implementing the infrastructure damage model through an Internet-based interface wherein a forecaster may enter a forecast of the necessary variables and obtain an infrastructure damage forecast for each region. The idea is to utilize the damage-model output as a forecast-guidance tool, similar to MOS. Human interaction with the damage model should produce more useful forecasts in certain situations, similar to how human forecasters outperform MOS forecasts (Glahn 1985). An example of possible improvements a human forecaster may provide is the use of a forecast of wind gusts and precipitation that represents the entire region as input for the model. In contrast to using a point-forecast approach, a forecast entered into the damage model that is representative of the impending weather may produce a better damage forecast.

The case studies presented herein pointed to the omission of wind duration and direction as being possible shortcomings. Duration may be accommodated in future work by combining days into events. This might improve results, because this method would allow for unbounded duration values. An event-based, hourly analysis method would likely provide the most improvement for powerful midlatitude-cyclone events, which often lasted multiple days according to the data analyzed for this study. For direction, stratification by direction of the maximum wind gust may yield better results, provided that sufficient data are available for such an analysis. Directional stratification may help to better detect wind events where topographic amplification occurs, such as the downslope windstorms discussed by Decker and Robinson (2011).

Infrastructure damage seems to be substantially more severe when it has been some time since the previous damaging storm. Incorporating as a predictor a lag factor to account for the amount of time between long-duration, high-magnitude wind events may also help to account for

the aging of vulnerable equipment and tree growth. The authors believe omitting such lag information from the current work may help to account for occasional large errors when the GLM predicts less than five total elements to incur damage (not shown). We leave for future work the process of accounting for the time between long-duration, high-magnitude events.

Including lightning observations from the National Lightning Detection Network could account for cases in which weak thunderstorms cause anomalously large counts of infrastructure damage, as seems likely in the 9 June 2009 case study, and would account for the substantial impact lightning has on electrical distribution operations (Cummins et al. 1998; Balijepalli et al. 2005; Zhu et al. 2007). In practice, however, one would need to predict lightning density or strike counts ahead of time, which is just as challenging as predicting storm reports, if not more so. Bothwell (2002) and Shafer and Fuelberg (2006) have generated lightning prediction models for the western United States and Florida, respectively; the SPC Short-Range Ensemble Forecast (SREF) system provides lightning guidance (Bright et al. 2005); and the lightning parameterization of Molinié et al. (2009) shows promise, but incorporating such information is left to future work.

Acknowledgments. This paper represents a portion of the first author's master's thesis. The research described here was funded in part by Public Service Enterprise Group and the New Jersey Agricultural Experiment Station. Frank Schwartz provided the damage data and user support for data applications. The first author thanks Bruce Veenhuis profusely for his help with the R programming language. We especially thank Tony Broccoli and Wayne Wittman as well as three anonymous reviewers whose comments and critiques have vastly improved upon previous versions of this manuscript.

REFERENCES

- Balijepalli, N., S. S. Venkata, C. W. Richter Jr., R. D. Christie, and V. J. Longo, 2005: Distribution system reliability assessment due to lightning storms. *IEEE Trans. Power Delivery*, **20**, 2153–2159.
- Bothwell, P. D., 2002: Prediction of cloud-to-ground lightning in the western United States. Ph.D. dissertation, University of Oklahoma, 178 pp.
- Bright, D. R., M. S. Wandishin, R. E. Jewell, and S. J. Weiss, 2005: A physically-based parameter for lightning prediction and its calibration in ensemble forecasts. Preprints, *Conf. on Meteorological Applications of Lightning Data*, San Diego, CA, Amer. Meteor. Soc., 4.3. [Available online at <http://ams.confex.com/ams/pdfpapers/84173.pdf>.]
- Brown, R. E., S. Gupta, R. D. Christie, S. S. Venkata, and R. Fletcher, 1997: Distribution system reliability assessment: Momentary interruptions and storms. *IEEE Trans. Power Delivery*, **12**, 1569–1575.
- Cerruti, B. J., S. G. Decker, L. A. Bowers, W. K. Wittman, and J. Carlson, 2009: Undergraduate forecasting and nowcasting for a major urban public utility. Preprints, *Eighth Symp. on the Urban Environment*, Phoenix, AZ, Amer. Meteor. Soc., J22.6. [Available online at <http://ams.confex.com/ams/pdfpapers/149046.pdf>.]
- Changnon, S. A., and T. R. Karl, 2003: Temporal and spatial variations of freezing rain in the contiguous United States: 1948–2000. *J. Appl. Meteor.*, **42**, 1302–1315.
- Cummins, K. L., E. P. Krider, and M. D. Malone, 1998: The U.S. National Lightning Detection Network and applications of cloud-to-ground lightning data by electric power utilities. *IEEE Trans. Electromagn. Compat.*, **40**, 465–480.
- Decker, S. G., and D. A. Robinson, 2011: Unexpected high winds in northern New Jersey: A downslope windstorm in modest topography. *Wea. Forecasting*, **26**, 902–921.
- DeGaetano, A. T., B. N. Belcher, and P. L. Spier, 2008: Short-term ice accretion forecasts for electric utilities using the Weather Research and Forecasting model and a modified precipitation-type algorithm. *Wea. Forecasting*, **23**, 838–853.
- DelSole, T., and J. Shukla, 2009: Artificial skill due to predictor screening. *J. Climate*, **22**, 331–345.
- Draper, N. R., and H. Smith, 1998: *Applied Regression Analysis*. Wiley-Interscience, 736 pp.
- Edwards, R., S. F. Corfidi, R. L. Thompson, J. S. Evans, J. P. Craven, J. P. Racy, D. W. McCarthy, and M. D. Vescio, 2002: Storm Prediction Center forecasting issues related to the 3 May 1999 tornado outbreak. *Wea. Forecasting*, **17**, 544–558.
- Glahn, B., M. Peroutka, J. Wiedenfeld, J. Wagner, G. Zylstra, B. Schuknecht, and B. Jackson, 2009: MOS uncertainty estimates in an ensemble framework. *Mon. Wea. Rev.*, **137**, 246–268.
- Glahn, H. R., 1985: Statistical weather forecasting. *Probability, Statistics and Decision Making in the Atmospheric Sciences*, A. H. Murphy and R. W. Katz, Eds., Westview Press, 289–335.
- , and D. A. Lowry, 1972: The use of model output statistics in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211.
- Han, S., S. D. Guikema, S. M. Quiring, K. Lee, D. Rosowsky, and R. A. Davidson, 2009: Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region. *Reliab. Eng. Syst. Saf.*, **94**, 199–210.
- Hansen, B., 2007: A fuzzy logic-based analog forecasting system for ceiling and visibility. *Wea. Forecasting*, **22**, 1319–1330.
- Hippert, H. S., C. E. Pedreira, and R. C. Souza, 2001: Neural networks for short-term load forecasting: A review and evaluation. *IEEE Trans. Power Syst.*, **16**, 44–55.
- Li, H., and L. A. Treinish, 2010: Statistical solution to forecasting electric power outages caused by severe storms. Preprints, *First Conf. on Weather, Climate, and the New Energy Economy/Eighth Users Forum on Weather and Climate Impacts*, Atlanta, GA, Amer. Meteor. Soc., J11.3. [Available online at <http://ams.confex.com/ams/pdfpapers/163237.pdf>.]
- Liu, H., R. A. Robinson, and T. V. Apanasovich, 2008: Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms. *Reliab. Eng. Syst. Saf.*, **93**, 897–912.
- Marzban, C., 2004: The ROC curve and the area under it as performance measures. *Wea. Forecasting*, **19**, 1106–1114.
- Molinié, G., J. Escobar, and D. Gazen, 2009: A stochastic lightning-flash scheme for 3D explicitly resolving cloud models. *Quart. J. Roy. Meteor. Soc.*, **135**, 113–124.

- Obled, C., G. Bontron, and R. Garcon, 2002: Quantitative precipitation forecasts: A statistical adaptation of model outputs through an analogues sorting approach. *Atmos. Res.*, **63**, 303–324.
- Ostby, F. P., 1992: Operations of the National Severe Storms Forecast Center. *Wea. Forecasting*, **7**, 546–563.
- Rakov, V. A., and F. Rachidi, 2009: Overview of recent progress of lightning research and lightning protection. *IEEE Trans. Electromagn. Compat.*, **51**, 428–442.
- Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608.
- Rooney, J. F., Jr., 1967: The urban snow hazard in the United States: An appraisal of disruption. *Geogr. Rev.*, **57**, 538–559.
- Shafer, P. E., and H. E. Fuelberg, 2006: A statistical procedure to forecast warm season lightning over portions of the Florida Peninsula. *Wea. Forecasting*, **21**, 851–868.
- Smith, T. L., S. G. Benjamin, J. M. Brown, S. S. Weygandt, T. Smirnova, and B. E. Schwartz, 2008: Convection forecasts from the hourly updated, 3-km High Resolution Rapid Refresh Model. Preprints, *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 11.1. [Available online at <http://ams.confex.com/ams/pdfpapers/142055.pdf>.]
- Stensrud, D., and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499.
- Takata, H., T. Hachino, and K. Komatsu, 2005: Prediction of electric power damage by typhoons in Japan via a two-stages predictor. *Dyn. Contin. Discrete Impulsive Syst.*, **1B**, 84–89.
- Treinish, L. A., A. P. Praino, H. Li, E. Novakovskaia, J. M. Drexel, R. Derech, and B. Hertell, 2010: Operational evaluation of a meso-scale weather and outage prediction service for electric utility operations. Poster, *First Conf. on Weather, Climate, and the New Energy Economy/Eighth Users Forum on Weather and Climate Impacts*, Atlanta, GA, Amer. Meteor. Soc., 506. [Available online at <http://ams.confex.com/ams/pdfpapers/161294.pdf>.]
- U.S. Census Bureau, cited 2010: GCT-PH1: Population, housing units, area, and density: 2000. [Available online at <http://www.census.gov/population/www/censusdata/density.html>.]
- Vannitsem, S., and C. Nicolis, 2007: Dynamical properties of model output statistics forecasts. *Mon. Wea. Rev.*, **136**, 405–419.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 648 pp.
- Witt, A., M. D. Eilts, G. J. Stumpf, E. D. Mitchell, J. T. Johnson, and K. W. Thomas, 1998: Evaluating the performance of WSR-88D severe storm detection algorithms. *Wea. Forecasting*, **13**, 513–518.
- Yang, D., B. E. Goodison, J. R. Metcalfe, V. S. Golubev, R. Bates, T. Pangburn, and C. L. Hanson, 1998: Accuracy of NWS 8" standard nonrecording precipitation gauge: Results of WMO intercomparison. *J. Atmos. Oceanic Technol.*, **15**, 54–68.
- Zhu, D., D. Cheng, R. P. Broadwater, and C. Scirbona, 2007: Storm modeling for prediction of power distribution system outages. *Electr. Power Syst. Res.*, **77**, 973–979.