Small-Scale Wind Turbine Recurrence and Cost Modeling as a Function of Operational Covariates from Supervisory Control and Data Acquisition Systems

Michael S. Czahor[1], William Q. Meeker[2]

[1,2] *Iowa State University, Ames, Iowa, 50011*
czahor02@iastate.edu
wqmeeker@iastate.edu

ABSTRACT

Small-scale wind turbine (SWT) installations saw a dramatic increase between 2008 and 2012. Recently, the trend within industry has shifted towards installing larger wind turbines, leaving little attention for installed SWT reliability. Unfortunately, multiple downtime events raise concerns about the reliability and availability of the large number of installed SWTs. SWTs are repairable systems that can return to an operational state after a downtime or repair event. When a SWT experiences multiple events over time, these are known as recurrent events. The reliability of SWTs is examined in this paper using data from 21 individual 100 kW wind turbines. SWTs periodically record dynamic covariate data in the form of a vector time series using supervisory control and data acquisition (SCADA) systems. One type of event experienced by SWTs is known as a "service event," which is a time when a SWT is put into service mode for a repair or false alarm. Due to the proprietary nature of the data used in this paper, different kinds of service events are combined, even though different failure modes and event types exist. We explore recurring service events and the associated cost of each "service event" and propose methodologies to link dynamic covariate data to downtime costs to assist in quantifying the variation of downtime across wind turbines. Data used in this work was provided from a power systems company in the United States. We outline a nonhomogenous Poisson process (NHPP) model with a Bayesian hierarchical power law structure for the count process and an autoregressive time series use rate model with a Bayesian framework to describe posterior parameter distributions. Using the posterior results, we develop a conditional and unconditional method to predict downtime mean cumulative functions (MCFs) for wind turbines.

1. INTRODUCTION

1.1. Background

Enhancing the reliability of wind turbines has been a collaborative effort between industry and academia for the past two decades. Recently, the Department of Energy (DOE) has laid out their vision for achieving this goal with a five-part plan that includes a well-developed database on wind farm operations under normal operating conditions. Gould (2014) summarizes DOE's *Wind Vision* report and outlines the goals set forth by the DOE, which include developing a world class database, ensuring reliable operation in severe operating environments, and developing and documenting the best practices in the wind industry to improve reliability and increase service life. Bertling and Wennerhag (2012) provide an international perspective in contrast to the US-centric DOE reports, with a compilation of reports that survey the development and research needs for wind turbine operation and maintenance across Europe. Reports within this survey include component specific reliability reports, maintenance strategy reports based on reliability modeling results, and database development needs for future work.

Limited work in applying reliability-based statistical methodologies to wind farm reliability data has been seen in the literature and can potentially assist in the DOE's effort. The *Reliawind* study by Wilkinson, Hendricks, Harman, Spinato, and van Delft (2011) identifies critical failure modes, summarizes the potential of SCADA systems, and highlights the benefits of having access to service records and alarm logs. Arifujjaman (2013) conducts a component-specific reliability analysis on grid-connected permanent magnet generator-based wind turbines and establishes a method to relate wind speed and power losses to the reliability of power electronic converters. Fischer, Besnard, and Bertling (2012) present results on a reliability-centered maintenance (RCM) study that utilizes failure data and industry expert opinions to improve the reliability, availability, and profitability of wind turbines.

There is an abundant number of wind turbine component-specific papers that implement statistical methodologies that are not reliability-based, but can be useful for future reliability work as they convey the benefits of SCADA data. Sajid and Hossam (2013) focus on predicting gearbox health using a nonlinear autoregressive model with exogenous inputs. Al-Tubi, Long, Tavner, Shaw, and Zhang (2015) investigate the probabilistic risk of gear flank micropitting risk with the use of SCADA data. Matthews and Godwin (2013) develop classification methods to detect wind turbine pitch faults using SCADA data.

1.2. Big Data in the Wind Energy Industry

Wind turbines are commonly outfitted with many sensors to assist in tracking operational and environmental conditions. According to Kashyap (2014), a typical wind turbine has 125 to 200 sensors that generate data at a rate of approximately 2000 observations per minute. At this rate, a single wind turbine can generate upwards of one terabyte of data in one week. Table 1 provides an example of types of data that SCADA systems capture from wind turbines. These time series data are generally averaged values over 10-minute intervals with an attached chronological timestamp. Ciang, Lee, and Bang (2008) and Falukner, Cutter, and Owens (2012) describe the use of such sensor data for system health monitoring. Tautz-Weinert and Watson (2016) provide a review of using SCADA data for wind turbine condition monitoring. Industry and academia have made progress in the prognostic realm of wind energy with advances in condition monitoring techniques, wind turbine sensor placement, and communication capabilities (e.g., allowing wind turbines to use an IP address to send a live feed of data to a centralized location). Saxena et al. (2008) summarizes different prognostic techniques that are being used across industries and highlights the benefits of having historical covariate data in correspondence with life data. We use such data for SWT analysis in this paper.

SCADA data contains information on the state of each wind turbine. We focus on a state that indicates when a wind turbine is in service mode. Programmable logical controllers (PLCs) continuously log state data and when a component of the wind turbine exhibits unusual behavior (values exceed predefined tolerances), the wind turbine will change states to let an owner or operator know of the event via an alarm. Such alarms or state changes may serve as a precursor to failure events and are of interest to owners and operators to minimize financial burdens that are experienced because of unplanned maintenance.

1.3. Key Feature of Wind Energy Field Data

Large vectors of time series data from wind turbines are periodically recorded, letting owners, operators, and researchers study differences between turbines at the individual and fleet levels. These dynamic covariate values are commonly referred to as system operating/environmental (SOE) data. Such SOE data have the potential to increase the reliability and availability of wind turbines with a minimal cost (Meeker & Hong, 2014). We look at SOE data from 100 kW wind turbines and show the benefits of analyzing such data with recurrence methods and highlight the future of similar work in the wind energy industry.

Table 1. Examples of data logged by SCADA systems.

Subsytem	**Data Collected**
Rotor and Blades	Pitch angle and rotor speed.
Gearbox	Oil, bearing, and hydraulic temperatures. Vibration, force, and rotational speed.
Generator	Stator and rotor voltages and currents. Power factors, rotor and grid frequencies, cabinet temperature, and generator speed.
Nacelle	Position, frame temperature, yaw break pressure, etc.

1.4. O&M Costs and Availability

According to Morthorst and Awerbuch (2009), O&M costs typically account for 20 to 25% of a wind turbine's total levelized cost of energy (LCOE). O&M costs have gone down drastically in the last 30 years due to advances in engineering, condition monitoring approaches, and preventive maintenance strategies. The International Renewable Energy Agency reports that O&M costs are not uniform across wind farms, suggesting that factors other than location (i.e., wind turbine manufacturer, model, etc.) influence the number and cost of downtime events, which are the main contributor to O&M costs. One must consider that wind turbines' O&M costs also change over time. The probability of failure increases, making failures more likely to occur outside of warranty periods, which increases the cost to return wind turbines to operational status after a downtime event. Refer to Gielen (2012) for more information on O&M costs for wind turbines.

One key O&M metric is wind turbine availability, which helps compare turbine to turbine performance. The method used to determine how availability is determined is up to the owner and operator. A large majority of operators do not have the capability to process terabytes of SCADA data to determine the true availability and resort to a simple method based on time. Wind turbine availability based on time is computed by

$$A_{Time} = \frac{T_{Operation}}{T_{Period}} \tag{1}$$

where $T_{Operation}$ is the time that a wind turbine is operating (i.e., generating power) and T_{Period} is the total time that a wind turbine could have been operating. This is an easy method to compute and is used widely within industry.

Maintenance events to be considered in (1) include preventive maintenance, corrective maintenance, and scheduled shutdowns. These are examples of events that lower A_{Time}, since $T_{Operation}$ decreases when such maintenance events occur . Throughout the paper we will discuss a collection of these events and provide an illustrative example of the associated downtime for such events.

1.5. Overview

Adapting some of the methods from Ryan, Hamada, and Reese (2011), the rest of the paper is organized as follows. Section 2 introduces wind turbine datasets provided by a United States energy company as a motivating example. Section 3 introduces notation for a nonhomogeneous Poisson process (NHPP) with power law intensity function. In Section 3, we also introduce a model to analyze mean cumulative functions for a single wind turbine and multiple wind turbines, while identifying the posterior parameter distributions. In Section 4 we develop an autoregressive time series model under a Bayesian framework to obtain posterior parameter distributions of the time series parameters. In Section 5, we propose methods to predict downtime MCFs using results from previous sections and provide results from a simulation study. In Section 6, we conclude and discuss the compromise between unconditional and conditional prediction intervals for downtime MCFs.

2. DATA

2.1. Company P2 Wind Turbine Data

The illustrative application is based on 21 wind turbines randomly located throughout the United States. The data were collected over the course of a four-year period from 2012 through 2016. The 21 wind turbines all have the same model specifications, with a generating capacity of 100 kW. All 21 wind turbines have unique starting times (install dates), but have a common data freeze date (DFD) in October 2016. During the observation period, each wind turbine had a SCADA system automatically record operational and environmental dynamic covariate data (e.g. wind speed, ambient temperature, etc.) as 10-minute averages for each variable. From the time of installation until the DFD, there is an entire covariate history. Each wind turbine's state information was also periodically recorded over the observation period, where a state code corresponds to a wind turbine's operational status

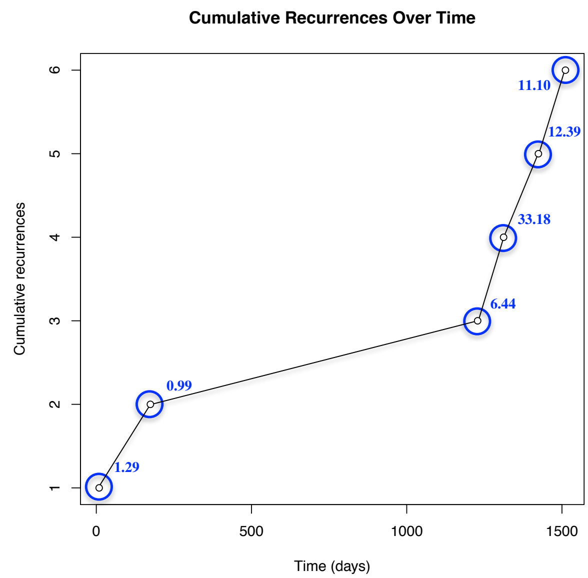

Figure 1. Cumulative recurrences vs. time (days) with associated downtimes for each service event.

during each 10-minute time frame. Because different manufacturers use different systems of state codes and because the exact coding method could be sensitive, we refer to states by name instead of code.

We focus on the "service mode" state. Being in this state implies that the wind turbine was being treated for preventive maintenance, suspect to a false alarm, or correcting a failure event. Service events are recurring events that each result in downtime, for which cost accrues over time. An example of service event data for a single wind turbine is shown in Figure 1. The numbers next to each event represent the associated amount of downtime (in days). For more information on state codes, please see Kusiak and Verma (2010).

3. SERVICE EVENT AND COST MODEL

3.1. A Nonparametric View of the Cost and Count Data

With access to the entire life history up until the DFD, we have a situation where n_j service events are observed for wind turbine j. A time truncated design is appropriate for such a scenario where the end of observation time is denoted by t_c, which is in real time the DFD. Refer to Dai and Wang (2017) for more information on truncated frameworks. The observed service events that occur for wind turbine j are $0 < t_1 < ... < t_{n_j} < t_c$. We plot nonparametric estimates for each wind turbine's mean cumulative function (MCF) with respect to the number of service events. Chapter 16 of Meeker and Escobar (1998) provides an algorithm to compute MCF point estimates and standard errors that allow one to compute pointwise approximate confidence intervals

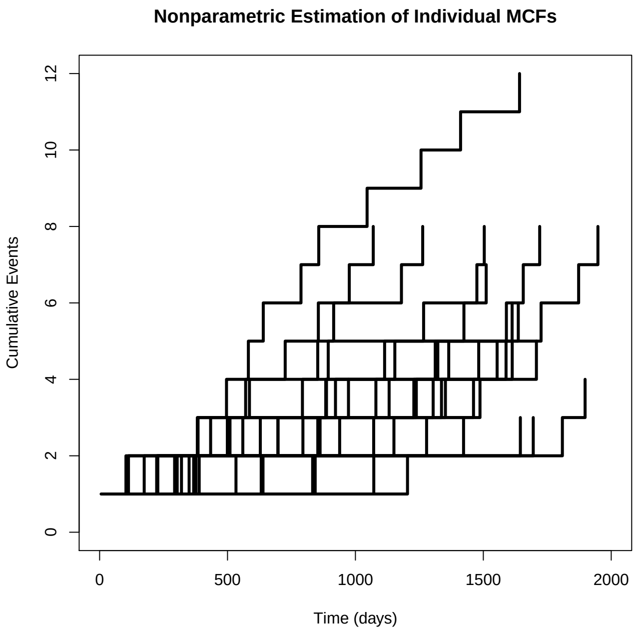

Figure 2. Cumulative events vs. time (days) for each wind turbine.

for each MCF. Figure 2 displays the individual MCFs for service events experienced by each wind turbine. See Nelson (2003) for more information on computing MCFs.

3.2. Nonhomogenous Poisson Process with a Power Law Intensity Function

Parametric statistical models for recurrent events have been studied by Rigdon (2000) and Bain and Engelhardt (1991). With minimal repair maintenance (as opposed to system renewal), the nonhomogenous Poisson process (NHPP) model is often appropriate. An NHPP model has a nonconstant recurrence rate $\nu(t)$. We consider an NHPP with a power law intensity rate function

$$\nu(t;\phi,\eta) = \frac{\phi}{\eta}\left(\frac{t}{\eta}\right)^{\phi-1}, \phi > 0, \eta > 0 \qquad (2)$$

with a mean function

$$\lambda(t) = E\big[N(t)\big] = \textstyle\int_0^t \nu(u)du = \left(\frac{t}{\eta}\right)^{\phi},$$

where η and ϕ can be estimated by using the method of maximum likelihood (ML). The likelihood function corresponding to (2) is

$$L(\phi,\eta) = \left(\frac{\phi}{\eta}\right)^{r} \times \textstyle\prod_{j=1}^{r} t_j^{\phi-1} \times \exp\left[-\mu\big(t_a : \phi, \eta\big)\right]$$

where $\mu\big(t : \phi; \eta\big) = \left(\frac{t}{\eta}\right)^{\phi}$ is the mean number of service events up to a certain time, t_a represents the time of the last service event, and r is the number of service events. The resulting ML estimates are

$$\hat{\phi} = \frac{r}{\sum_{h=1}^{r} \log\big(t_a/t_h\big)}$$

$$\hat{\eta} = \frac{t_a}{r^{1/\hat{\phi}}}.$$

For more information on NHPP estimation procedures see Chapter 16 of Meeker and Escobar (1998).

3.3. Model for a Single Wind Turbine

We replace the parameter η with $\lambda = \lambda_j = \lambda\big(t_{c_j}\big) = \left(\frac{\eta}{c}\right)^{-\phi}$. which is the mean number of service events up to time t_{c_j}. To move between two parameterizations we let

$$\eta = c\lambda^{-1/\phi}$$

The development of a single wind turbine model will assist in the multiple turbine model that is necessary to describe the $J = 21$ wind turbine datasets of interest. The derived likelihood function for the time truncated design can be seen in Ryan et al. (2011).

To make inferences on the parameter vector $\theta = \big(\lambda, \phi\big)$ we use the posterior distribution, $\pi\big(\theta|DATA\big)$, which is proportional to

$$\mathcal{L}(DATA|\theta)\pi(\theta), \qquad (3)$$

where θ is a parameter vector and $\pi(\theta)$ is a prior distribution of θ. When there is little or no prior information we use a diffuse prior distribution. See, for example, Gelman et al. (2013) for detailed information about Bayesian analysis. The posterior distribution in (3) provides an update of information on the parameters based on the observed data.

In our analysis we assume no knowledge of the parameters and use diffuse Jefferey's priors (Dodge & Whittaker, 1992) for the parameters where

$$\pi(\lambda,\phi) \propto \frac{1}{\lambda\phi}.$$

Following Ryan et al. (2011), direct use of Bayes' Theorem results in independent posterior distributions

$$\lambda|t_1, ...t_n, t_c \sim \text{Gamma}(n, 1)$$

$$\phi|t_1, ...t_n, t_c \sim \text{Gamma}\Big(n, \textstyle\sum_{i=1}^{n} \ln(t_c/t_i)\Big)$$

for λ and ϕ where an estimate of the expected number of wind turbine service events at the end-of-observation time t_{c_j} is equal to n. Ryan et al. (2011) discusses computing credible intervals for functions of λ and ϕ and provides a numerical example. We note that the form of the Gamma density used throughout the paper is Gamma$(a, b) = b^a x^{a-1} e^{-xb}/\Gamma(a)$.

3.4. Multiple Wind Turbine Hierarchical Model

We now consider the service event data from multiple wind turbines. Similar to Ryan et al. (2011), we use a fully hierarchical model based on a power law process for multiple repairable systems (i.e., wind turbines) and simple posterior inference methods for time-truncated designs by taking the cost of service events into consideration and adding dynamic covariate information to the analysis.

We start by considering the J wind turbines, which have unique install times and are observed through the time t_{c_j}. For wind turbine j, n_j service times are observed and denoted by $t_j = \left(t_{j1}, t_{j2}, ..., t_{jn_j}\right)$. We assume that the service event times from wind turbine j follow an intensity model (2) and parameters λ_j and ϕ_j. We consider a hierarchical model to allow data from all of the wind turbines to be pooled, and allow all wind turbines to have their own intensity parameters. The statistical notion of pooling data is commonly known as "borrowing strength" and is outlined, for example in Draper et al. (1992). This notion assists in describing relationships involving the observed data and unobserved parameters of interest.

To motivate a model similar to the one developed in Ryan et al. (2011), we look at the empirical distributions of fitted parameter estimates for the individual wind turbines and find that Gamma distributions are also appropriate in this application. These distributions can be seen in Figure 3.

Thus let the distributions for λ and ϕ be iid Gamma distributions denoted by

$$\lambda_j \sim \text{Gamma}\left(\alpha_\lambda, \beta_\lambda\right) \tag{4}$$

$$\phi_j \sim \text{Gamma}\left(\alpha_\phi, \beta_\phi\right) \tag{5}$$

for $j = 1, 2, ...J = 21$. With no information on the hyperparameters in (4) and (5), we propose

$$\alpha_\lambda \sim \text{Gamma}\left(a_1, b_1\right),$$

$$\beta_\lambda \sim \text{Gamma}\left(a_2, b_2\right),$$

$$\alpha_\phi \sim \text{Gamma}\left(a_3, b_3\right),$$

$$\beta_\phi \sim \text{Gamma}\left(a_4, b_4\right),$$

priors for α_λ, β_λ, α_ϕ, and β_ϕ to guarantee that the parameters are positive, where choices for $a_1, a_2, a_3, a_4, b_1, b_2, b_3$, and b_4 are selected to make the priors diffuse, allowing analysis on the $J = 21$ wind turbines to be data driven.

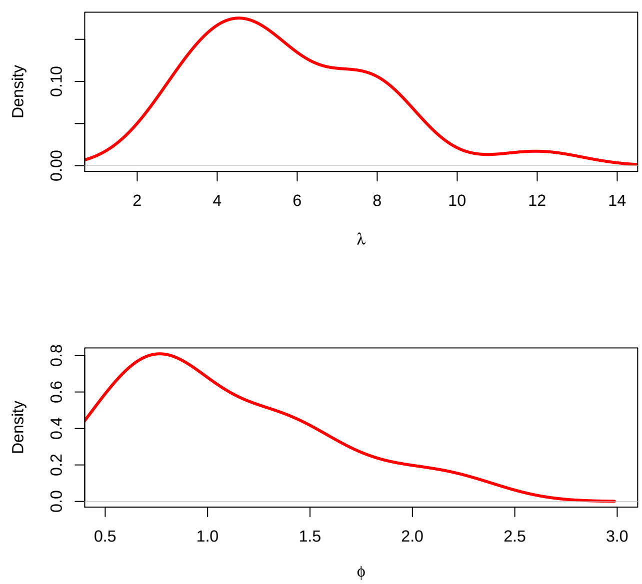

Figure 3. Empirical parameter distributions from individual NHPP model fits

A fully specified likelihood and prior for the multiple systems model can be seen in Ryan et al. (2011), followed by a multiple systems Metropolis-Hastings in Gibbs sampler. Here we use RJAGS for our analysis. For more information on Metropolis-Hastings and Gibbs samplers see Upton and Cook (2014).

Because the NHPP parameters vary from turbine-to-turbine, estimates based on fully-pooled data are subject to large bias. If there is no pooling, we expect less bias, but an increased variance in our parameters. The hierarchical model allows for a tradeoff between a completely pooled analysis and an individual turbine analysis (Draper et al., 1992).

Results from RJAGS can be seen in Figure 4, Figure 5, and Table 2.

4. COST AND USE RATE MODEL

4.1. SCADA Data and Cost Relationship

After developing models for service event counts and corresponding downtimes, we are now interested in relating operating conditions to the amount of downtime that results from each event. The use rate, which is a two week average immediately before a service event, is measured in rotations per minute (rpms), has a linear relationship with Log$(cost)$. We plot the costs (measured in days) of the $N = 121$ versus

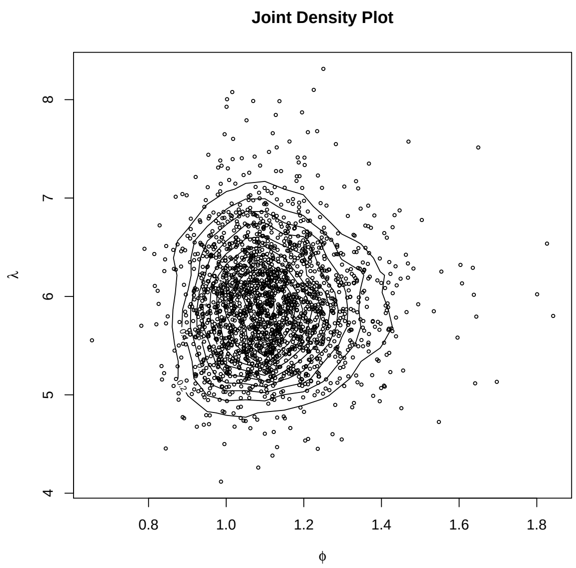

Figure 4. NHPP joint posterior distribution

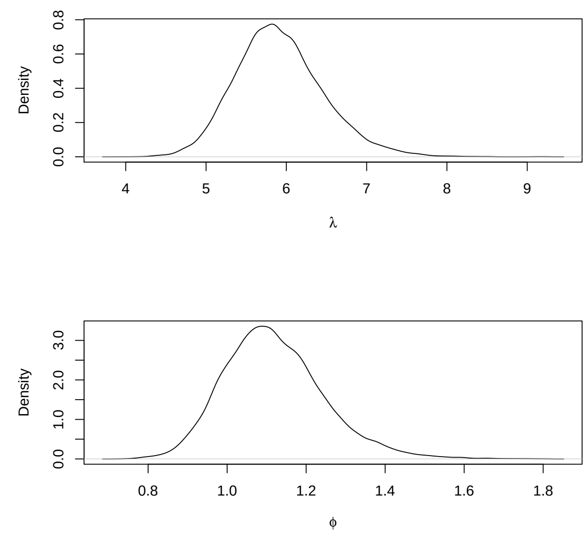

Figure 5. NHPP posterior distributions

the corresponding use rates, which are the two week turbine-specific rpm averages immediately before a service event. The relationship between cost and use rates can be denoted by

$$Z_i = \beta_0 + \beta_1 \times U_i + \epsilon_i$$

where Z_i is the log of the cost, U_i is the corresponding use rate, and β_0 and β_1 are population parameters that are estimated via a simple linear regression model. We plot the data in Figure 6.

Using standard graphical regression diagnostic checking of the usual linear model assumptions (i.e., constant variance of residuals, independence of residuals, normally distributed residuals, and a linear relationship between the explanatory and response variables) we found no evidence of serious departures. Also we note that there is a correlation of 0.51 between Z_i and U_i. Using the data, a simple linear regression model is

$$z_i = -2.31 + 0.11 \times u_i + \epsilon_i \tag{6}$$

where z_i is the observed log cost, u_i is an observed use rate, and ϵ_i is the random error associated with the prediction.

If we knew the use rate of future service events, we would be able to predict the corresponding downtimes for each event. We now develop a time series model for use rates.

Table 2. NHPP posterior parameter output.

Parameter	Median	95% Credible interval
η_j	401.2	(282.2 , 607.2)
λ_j	5.868	(4.945 , 7.084)
ϕ_j	1.107	(0.895 , 1.408)

4.2. Autoregressive Use Rate Model

We explore the time series structure of use rates for the $J = 21$ wind turbines using JMP Statistical Software. Using the "AR Coefficients" tool, we notice that an AR(2) model provides an adequate description of the use rate data.

We notice no significant differences across use rate distributions from turbine-to-turbine, but do notice some missing data. To simulate use rate data for additional wind turbines with the same model specifications, we used the arima.sim function in R to help produce posterior distributions to sample from. We used an AR(2) process with parameters $\left(\gamma_1, \gamma_2, \tau^2\right)$ where

$$U_t = \gamma_1 U_{t-1} + \gamma_2 U_{t-2} + \epsilon_t, \epsilon_t \sim N\left(0, \tau^2\right) \tag{7}$$

and the likelihood function for (7) is

$$f\left(U_1, U_2, ..., U_t\right) = f\left(U_1\right) \prod_{k=2}^{t} f\left(U_k | U_1, ...U_{k-1}\right).$$

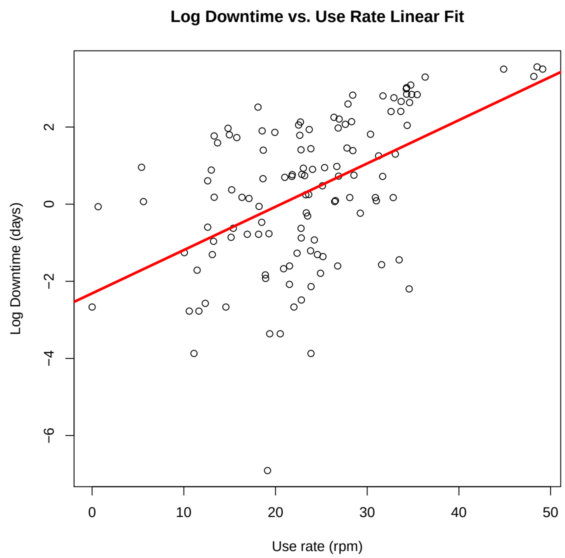

Figure 6. Log cost vs. use rate for $N = 121$ service events.

4.3. AR Model in JAGS via Bayesian Analysis

Similar to Section 3, we follow a standard Bayesian approach by using a likelihood and prior to obtain posterior distributions for our parameters of interest.

$$\pi\left(\tau^2, \gamma_1, \gamma_2 | U_1, U_2, ..., U_t\right)$$

$$= f\left(U_1, U_2, ..., U_t | \tau^2, \gamma_1, \gamma_2\right)\pi\left(\tau^2\right)\pi\left(\gamma_1, \gamma_2\right)$$

When trying to derive the full conditional distributions for τ^2 and $\left(\gamma_1, \gamma_2\right)$ using conjugate priors, the full conditional distribution for $\left(\gamma_1, \gamma_2\right)$ is complicated, making a Metropolis-Hastings algorithm appropriate. Depaoli, Clifton, and Cobb (2016) describe a software known as Just Another Gibbs Sampler (JAGS). We use the JAGS software to generate a candidate density and tune the distribution adaptively to ensure the chain samples correctly. We use noninformative uniform prior distributions for γ_1 and γ_2. Because $\tau > 0$, we use a gamma prior with shape and scale parameters of 0.001 and 0.001 respectively. This prior distribution places most of the mass close to 0 and lets the data from empirical observations dictate the shape of the posterior distribution. Table 3 and Figure 7 show posterior output from JAGS. For more information on prior distributions for variance parameters, see Gelman (2006).

5. PREDICTING BEHAVIORS OF A NEW WIND TURBINE

5.1. Assumptions

In this section we use the model fit to the 21 available wind turbines to predict the future cost of an addition turbine to be installed in the future. We call this Turbine 22.

Table 3. Use rate AR(2) Parameter estimates

Parameter	Median	95% Credible interval
γ_1	1.249	$(-1.122$, $1.384)$
γ_2	-0.262	$(-0.397$, $-0.135)$
τ	5.058	$(4.591$, $5.614)$

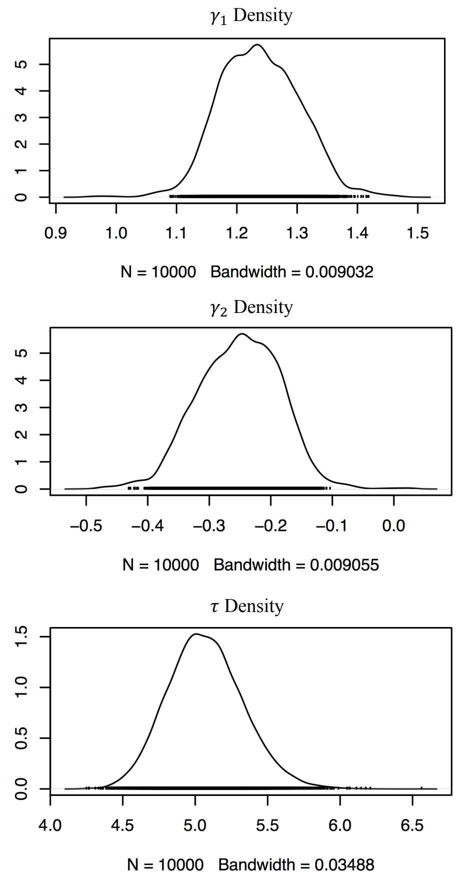

Figure 7. Posterior output from use rate AR(2) model.

The assumptions needed for the cumulative cost prediction are

1. The relationship between use rates and costs in (6) holds for Turbine 22.
2. Recurrence rates are independent of cost parameters
3. Turbine 22 comes from the same population of the originally observed wind turbines.

5.2. Simulating Draws from Posterior Predictive Distributions

In this section we present an approach to simulate from the posterior distributions in Section 3 and 4 to generate predictions for the cost of a new wind turbine. Consider different

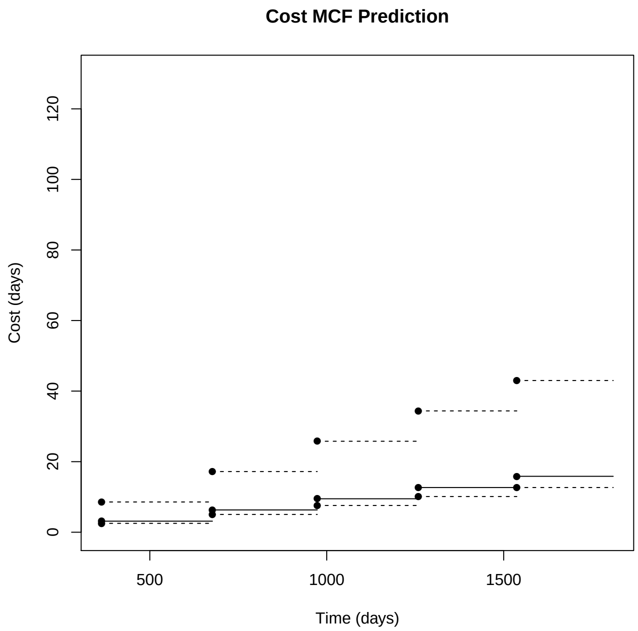

Figure 8. Conditional predicted cost MCF with 95% prediction intervals.

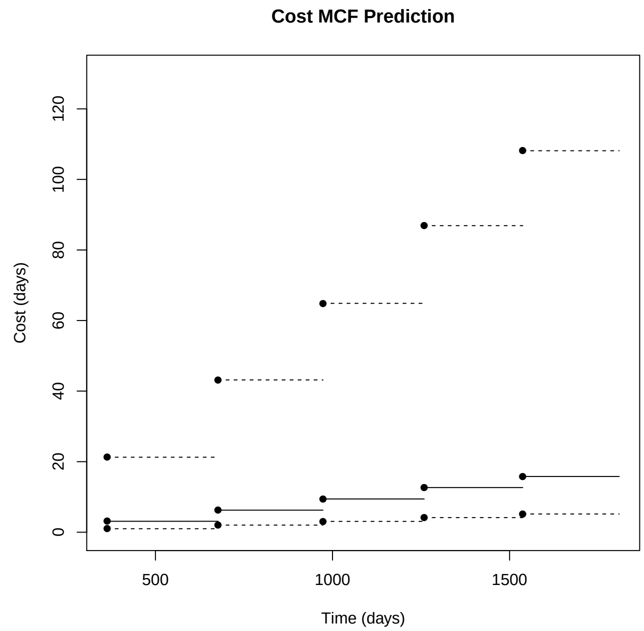

Figure 9. Unconditional predicted cost MCF with 95% prediction intervals.

methods: a conditional approach and an unconditional approach. For the conditional approach we specify $t_{c_{22}}$ and

(a) Draw λ_{22} and ϕ_{22} from the joint posterior distribution.
(b) Draw a realization from an AR(2) process
(c) Simulate NHPP events until $t_{c_{22}}$ resulting in n_{22} events
(d) For each event generate downtimes $d_1, ..., d_{n_{22}}$ using (6).
(e) Compute the MCF and accumulate
(f) Repeat steps b) - e) B_2 times and save the results
(g) Obtain the 0.025, 0.5, and 0.975 quantiles of the predictive distribution, giving a point prediction and 95% prediction intervals for each point in time.

The unconditional approach is similar, but we would generate a new λ and ϕ from the posterior distribution each time.

5.3. MCF Cost Function Results

After following the steps in Section 5.2, with $B_2 = 10000$, we obtain a cost MCF prediction with prediction intervals that are obtained by finding the 0.025 and 0.975 quantiles in part (g) of the conditional algorithm. Figure 8 shows the MCF prediction results, where the solid line is the median obtained in part (g). We notice that distance between the upper bound and MCF prediction increases quickly, due to the right skew in the distribution of costs. In Figure 8 the observation period is $0 < t < t_{c_{22}}$, where $t_{c_{22}} = 2000$.

The conditional distribution assumes we know MCF parameters of Turbine 22. Unconditionally, we expect to see wider prediction intervals since the NHPP parameters vary in an unconditional algorithm. Figure 9 shows the unconditional MCF prediction results. The prediction intervals generated in the unconditional approach are noticeably wider than the prediction intervals generated using the conditional approach.

6. DISCUSSION

6.1. A Compromise Between Conditional and Unconditional Approaches

Before we know anything about Turbine 22, the MCF cost prediction can be dealt with unconditionally. Once Turbine 22 begins to operate, we can use available data to update the prior distributions to get higher precision in our MCF predictions. One could perform a holdout method, where data is obtained over a period of time and MCF prediction intervals are generated via the unconditional approach and then use observed data to update the prior distributions up until the prediction time, t_{pred}. With updated information on Turbine 22's life characteristics, the prediction process becomes conditional on the prior information.

6.2. Benefits of Linking Covariate Data to Event Data

Technological advancements, including SCADA systems, have the capability to minimize the uncertainty in predictions of reliability characteristics for wind turbines. It is desirable to have accurate reliability predictions in variable environments, such as wind farms, which are suspect to various environmental and operational conditions. Having access to individual wind turbine SCADA data, in addition to lifetime data, can be extremely useful for maintenance optimization and economic planning purposes.

6.3. Acknowledgements

We thank the power systems company who provided us with data on the $J = 21$ wind turbines used for this research.

REFERENCES

Al-Tubi, I., Long, H., Tavner, P., Shaw, B., & Zhang, J. (2015). Probabilistic analysis of gear flank micropitting risk in wind turbine gearbox using supervisory control and data acquisition data. *IET Renewable Power Generation*, *9*(6), 610-617.

Arifujjaman, M. (2013). Reliability comparison of power electronic converters for grid-connected 1.5kw wind energy conversion system. *Renewable Energy*, *57*, 348 - 357.

Bain, L., & Engelhardt, M. (1991). *Statistical analysis of reliability and life-testing models, theory and methods*. Marcel Dekker.

Bertling, L., & Wennerhag, P. (2012, October). *Wind turbine operation and maintenance* (Tech. Rep. No. 12:41). Elforsk.

Ciang, C. C., Lee, J., & Bang, H. (2008). Structural health monitoring for a wind turbine system: a review of damage detection methods. *Measurement Science and Technology*, *19*(12), 122001.

Dai, H., & Wang, H. (2017). *Analysis for time-to-event data under cenosring and truncation*. Elsevier.

Depaoli, S., Clifton, J. P., & Cobb, P. R. (2016). Just another gibbs sampler (jags). *Journal of Educational and Behavioral Statistics*, *41*(6), 628-649. doi: 10.3102/1076998616664876

Dodge, Y., & Whittaker, J. (1992). *Computational statistics* (Vol. 1). Physica-Verlag Heidelberg.

Draper, D., Gaver, D. P., Greenhouse, J. B., Hedges, L. V., Morris, C. N., Tucker, J. R., & Waternaux, C. M. (1992). *Combining information: Statistical issues and opportunities for research* (Tech. Rep.). Committee on Applied and Theoretical Statistics.

Falukner, P., Cutter, P., & Owens, A. (2012). Structural health monitoring systems in difficult environments. In *European workshop on stuctural health monitoring.* Dresden, Germany.

Fischer, K., Besnard, F., & Bertling, L. (2012, March). Reliability-centered maintenance for wind turbines based on statistical analysis and practical experience. *IEEE Transactions on Energy Conversion*, *27*(1), 184-195.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*(3), 515-534. doi: 10.1214/06-BA117A

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian Data Analysis* (3rd ed.). CRC Press.

Gielen, D. (2012). *Renewable energy technologies: cost analysis series* (Tech. Rep.). International Renewable Energy Agency.

Gould, E. B. (2014). *Wind vision: updating the DOE 20 % wind energy by 2030* (Tech. Rep. No. DOE/GO-102015-4557). NREL.

Kashyap, S. (2014). *White paper: Data analytics for wind farm performance improvement* (Tech. Rep.). http://algoengines.com/wind/: Algo Engines.

Kusiak, A., & Verma, A. (2010). *The future of wind turbine diagnostics.*

Matthews, P., & Godwin, J. (2013, January). Classification and detection of wind turbine pitch faults through scada data analysis. *International Journal of Prognostics and Health Management*, *4*(Sp2), 90-100.

Meeker, W. Q., & Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. Wiley.

Meeker, W. Q., & Hong, Y. (2014). Reliability meets big data: opportunities and challenges. *Quality Engineering*, *26*, 102-116.

Morthorst, P., & Awerbuch, S. (2009). *The economics of wind energy* (Tech. Rep.). EWEA.

Nelson, W. B. (2003). *Recurrent events data analysis for product repairs, disease recurrences, and other applications*. ASA-SIAM.

Rigdon, S. E. (2000). *Statistical methods for the reliability of repairable systems*. John Wiley.

Ryan, K., Hamada, M., & Reese, C. (2011). A bayesian hierarchical power law process model for multiple repairable systems with an application to supercomputer reliability. *Journal of Quality Technology*, *43*(3), 209-223.

Sajid, H., & Hossam, A. G. (2013). Fault diagnosis in gearbox using adaptive wavelet filtering and shock response spectrum features extraction. *Structural Health Monitoring*, *12*(2), 169-180.

Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., & Schwabacher, M. (2008, Oct). Metrics for evaluating performance of prognostic techniques. In *2008 international conference on prognostics and health management* (p. 1-17). doi: 10.1109/PHM.2008.4711436

Tautz-Weinert, J., & Watson, S. J. (2016). Using SCADA data for wind turbine condition monitoring - a review. *IET Renewable Power Generation*.

Upton, G., & Cook, I. (2014). *Oxford dictionary of statistics*. Oxford University Press.

Wilkinson, M., Hendricks, B., Harman, K., Spinato, F., & van Delft, T. (2011). *Measuring wind turbine reliability-results of the reliawind project.*