Presented at ARTS Conference, Loughborough, April 2007

Reliability-Growth Analysis of Wind Turbines from Fleet Field Data

Fabio Spinato, Peter Tavner
New & Renewable Energy Group
School of Engineering, Durham University, Durham, DH1 3LE, UK
Gerard van Bussel
Wind Energy Group
Technical University, Delft, Netherlands

Abstract

The availability of failure data for complex and commercially relevant systems, like wind turbines, is often restricted to a few publicly accessible databases, which collect failures in grouped form for entire fleets of units.
Despite the deficiencies of this data, reliability-growth analysis methods do allow the extraction of reliability trends over an observed period. The analysis can also differentiate between

- subassemblies in a system subject to human-driven reliability improvement and mature technology
- subassemblies which are deteriorating, characterised by an increasing failure intensity.

For this purpose the Crow-AMSAA mathematical model, based on the Power Law Process (PLP), has been implemented on fleet failure data from wind turbines in Germany over a 10 year period.
Better design, maintenance or repair activities will spread technological innovation throughout a fleet of systems with a rapidity that is reflected by the value of the shape parameter of the failure intensity function. This is demonstrated in the analysis of the wind turbine fleet failure data.
In fleets or subassemblies with more static technology, analysis of the failure data should exhibit technology maturity. This is demonstrated in the analysis by the negative reliability growth encountered.
Two statistical tests were applied to the wind turbine data, the goodness of fit for the Power Law Process and the null hypothesis of constant intensity function for the same PLP, which is the Homogeneous Poisson Process (HPP).
The results from the intensity function analysis were used to approximate the expected failure intensity for different designs of wind turbine and were subsequently compared with failure from other surveys to improve confidence in the validity of the analysis.
The results from this type of wind turbine subassembly failure analysis can indicate the reliabilities which could be expected in the prospective off-shore installation of wind energy turbines, showing the wind turbine concepts which will offer a greater potential for reliability improvement in the future.

1. Introduction

The impending challenge of achieving cost-effective, reliable wind turbines (WT) in the offshore environment can be realised only if there is a substantial improvement over the reliability measured on today's onshore wind turbines.
The limited accessibility to offshore wind turbine sites will increase operational costs by 1000%, compared with the equivalent operations onshore, due to the necessary utilisation of costly crane vessels for maintenance and operation and maintenance (O&M) costs will add 30% to the final cost of energy from offshore installations [1, 2, 3, 4].
Improving of availability of offshore wind turbines requires the adoption of a number of design and operational measures, for example the choice of the most effective turbine architecture, the installation of effective condition monitoring and the application of appropriate O&M programs. However, the most important measure will be to use turbines of proven reliability.
This paper is intended to contribute by analysing the reliability of different designs of turbine at present installed onshore in Germany.
Despite the crucial importance that reliability plays, very little is known about real failure occurrences figures. Current analysis is usually based upon unscientific methods, relying heavily on expert knowledge, 1 or 2 year averages of restricted failure data and operational value judgements.
The reason for this is the non-availability of accurate failure data, because of its significance in multi-million pound wind turbine investments, where manufacturers and operators apply confidentiality to their commercially-sensitive failure data.
Researchers have therefore focused on developing economic or O&M mathematical models of wind turbines, without being able to investigate the "quality" of the reliability data on which their mathematical models are founded.
It is clear that, the variability of weather and other "environmental" conditions, such as grid disturbances, contribute to the failures which turbines experience. This variability heavily affects the reliability of the turbine subassemblies, making the process of averaging the failures over any period potentially inaccurate, especially if averages are made over short periods [5].
In addition to the uncertainty caused by environmental conditions, field failures are also affected by decreasing or increasing time trends, due respectively to designers' efforts to improve the turbine, or the natural deterioration of subassembly properties.
While the effect of environmental covariates can only be analysed with dedicated statistical tools, time trends can be studied by applying well-known methods of reliability growth analysis.
These are the analytical tools normally used to support reliability growth management programs. For the application of such methods, military handbooks and standards are, in this field, the most widely adopted protocols [6].
This paper presents the methods and results of the implementation of one of the method's cardinal concepts; the so called, demonstrated reliability, applied to German wind turbine fleet failure data.
Once the mathematical model has been implemented on a set of failure data, the reliability of the objects in analysis can be more easily compared. The intensity function $\lambda(t)$, the main result of the mathematical model, is a continuous function representing the instant intensity of failure that allows both a rapid visual comparison and a more systematic evaluation of particular features of the data, as:

- Validation of the reliability figures. In the case of early failure rate, the value of the instant failure intensity at the end of the period of analysis, $\lambda(T)$ can be assumed to be the expected failure intensity to the end of the development.
- Maturity of the technology. The observation of future failure rates assumed from the shape parameter of the failure intensity function, the feasibility of further reliability improvement can then be estimated.
- Random failures hypothesis. In case of observing significantly constant failure intensity function, the assumption of HPP as a valid reliability mathematical model.
- Deterioration trend. For increasing trends of the failure intensity function the expected initial failure rate can be estimated. The same trend is also quantified.

The Military Standard [6] includes statistics that are suitable for the particular kind of wind turbine fleet data that are to be analysed.

2. Data Set

The data set which has been analysed has been extracted from [7], a publicly available report issued in Germany and reporting failure data from wind turbines installed in the region of Schleswig-Holstein from 1994 to date.
Data are collectively reported by year (grouped data), for each turbine of a given model and segregated for each of the turbine subassemblies. The subassemblies constitute the lowest level of indenture available and no further information is available about the failure modes which caused failure. For obvious reasons the data concerns turbines in service under incompletely specified environmental conditions. Data that are characterised by the above features are often referred as fleet field data.
It is known that two major categories of phenomena affect the reliability of the wind turbines;

- Mechanical subassemblies are heavily affected by the dynamic loads due to wind turbulence and gustiness.
- Disturbances of the electrical grid influence the reliability of electrical related subassemblies.

The number of hours lost due to subassembly failures and outages are also reported, allowing the calculation of the actual fleet running time or total time on test, in the language of the Military Standard, which has been used as the age variable in this analysis.
For each turbine model the population changes as a result of new turbines included in and non-operational turbines excluded from the survey and the number of turbines involved is not constant. As specified later, this does not affect the mathematical model, but makes some conceptual assumptions necessary. It is assumed that, in the case of positive reliability growth, maintenance and design & operational experience gained from previous turbines, spreads the growth of reliability over the population of turbines, with an overall perceivable effect. The rapidity of reliability growth depends on the efforts of both operators and designers, but being a feature of that particular population, results cannot be generalised. The mathematical model aims to detect this collective, fleet trend and to evaluate the feasibility of further reliability improvement.

Seven wind turbines models have been chosen for the analysis, allowing the comparison of three categories of machines ratings; small (rated at about 250 kW), medium (rated at about 500 kW) and large (rated at about 1 MW). Table 1 is an example of the available data:

Subassembly	1998	1999	2000	2001	…
Blade	2	17	6	0	…
Rotor Brake	0	0	0	0	
Pitch control	7	10	5	5	
Main Brake	0	2	0	3	
Shaft	0	0	0	2	
Gearbox	6	15	5	7	
…	…				
Number of Turbines	61	52	45	47	
Hours Lost	2936	3317	3534	3571	

Table 1: LWK failure data extract, for a Vestas V39 (500 kW) turbine.

The presence of many time cells containing less than 5 failures is common, given the numbers of turbines involved, representing an obstacle to the implementation of the statistical model. Nevertheless, whenever this problem arises, the aggregation of contiguous time cells represents an effective and rapid method to overcome this deficiency.

On previous works the authors implemented the mathematical model on a similar data set [8, 9] with satisfactory results for the purposes of comparison between two national wind turbine populations.

The access in the LWK fleet to data segregated by wind turbine model constitutes a further and valuable improvement in understanding of wind turbine reliability data.

3. The Reliability Growth Model

The Crow-AMSAA mathematical model is based on the Power Law, or Weibull, Process (PLP), a particular case of Poisson Process. This statistical tool has been adopted for modelling reliability improvement in the early failures, constant failures and deterioration phases, and its flexibility in representing these three phases of the so-called "bathtub curve" is at the origin of its adoption.

The failure intensity function $\lambda(t)$ of the Crow-AMSAA mathematical model is represented by the following exponential expression:

$$\lambda(t) = \rho \, \beta \, t^{\beta-1} \tag{1}$$

The shape parameter β , determines the trend of the curve, while ρ is the scaling parameter, so for $\beta<1$, $\beta=1$ and $\beta>1$ the mathematical model represents respectively improvement, constant failures and deterioration. Figure 1 gives an impression of the shape of the intensity function in the three phases.

Figure 1: the three phases of the bathtub curve. From left, Early Failures, Constant Failures and Deterioration. The first is asymptotic to the ordinate axis.

In the case of grouped data, that is failure collectively grouped in periods, for which times to failure are interval censored, the Maximum Likelihood Estimators (MLE) of the shape parameter and the scale parameter, β and ρ, are given by:

$$\sum_{i=1}^{I} n_i \left(\frac{t_i^{\hat{\beta}} \ln t_i - t_{i-1}^{\hat{\beta}} \ln t_{i-1}}{t_i^{\hat{\beta}} - t_{i-1}^{\hat{\beta}}} - \ln t_I \right) = 0 \tag{2}$$

$$t_0^{\hat{\beta}} = \ln t_0 = 1 \tag{3}$$

$$\hat{\rho} = \frac{\sum_{i=1}^{I} n_i}{t_I^{\hat{\beta}}} \tag{4}$$

Where the cap, like for $\hat{\rho}$ distinguishes the MLE from the relative random variable, I is the total number of intervals and n_i is the number of failures in the i^{th} interval, which is defined by the set [t_{i-1}, t_i].
For fleet grouped data, it is convenient to use the total time on test (TTT) as an age variable, calculated as the cumulative integral of the running hours of the considered population.
The non linear equation (2) has been solved in β, with an iterative procedure then, using the equation (4), the shape parameter's MLE is obtained.

4. Tests

A goodness of fit (GoF) test for the mathematical model can be built from the expected number of failures in the *ith* interval, e_i,, that is:-

$$e_i = \hat{\rho}(t_i^{\beta} - t_{i-1}^{\beta}) \tag{5}$$

The statistic:-

$$\chi^2 = \sum_{i=1}^{I} \frac{(n_i - e_i)^2}{e_i} \tag{6}$$

is distributed as a chi-square random variable with $I-2$ degrees of freedom, for which the critical values can be found in common tables. The hypothesis is accepted if the statistic assumes values smaller than the critical value, which is calculated for the level of confidence, α, chosen.
The hypothesis of no growth, that is equivalent of considering the times between failures exponentially distributed, can be tested with the following statistics

$$y^2 = \sum_{i=1}^{I} \frac{(n_i - NP_i)^2}{NP_i} \tag{7}$$

$$P_i = \frac{(t_i - t_{i-1})}{T} = \frac{T_i}{T} \tag{8}$$

T being the total test time (TTT). The random variable *y* is distributed as a chi-square random variable with *I-1* degrees of freedom.

The final result, early failure, constant intensity or deterioration phase, is deduced from the evaluation of the combined results for beta and the trend hypothesis test according to the following table:

Shape parameter	Ho=HPP	Result
$\beta < 1$	Rejected	early failures
$\beta > 1$	Rejected	deterioration
$0.88 < \beta < 1.2$	Accepted	constant failures
$\beta < 0.88$ or $\beta > 1.2$	Accepted	unknown

Table 2: combinations of results

Obviously, the rejection of the goodness of fit tests simply means the non-applicability of the PLP for the set of data under examination.

5. Assumptions

It is necessary to distinguish between the asymmetry of the early failure and deterioration phases.

In the early failures phase the improvement is substantially “human driven”, the results of actual changes to the subassemblies of the turbine systems, as operational experience is gained. The improvement depends on the effort put into this task and is strictly dependent upon the fleet being observed. The “washing out” effect of subassemblies prone to early failure, is typical of non-repairable, or “one shot” systems, but is also present in costly and highly repairable machinery, such as wind turbines. Furthermore the final failure intensity value $\lambda(T)$ is achieved after a period of development, involving the repair and replacement of less reliable subassemblies, and can be adopted as the initial expected intensity of failures for similar new systems.

For the deterioration phase, normally the consequence of material degradation, only the shape parameter β has a clear, directly applicable meaning and can be used to predict the behaviour of future similar systems. The final failure rate value is strictly dependent upon the fleet being observed, since the intensity curve grows indefinitely with the time and its absolute value depends on the length of the observation period.

The question is now what would be the initial expected failure rate value for systems that deteriorate? The intensity curve of a system that wears out can vary widely, so it is not possible to define univocally a single value. In this case the value that has been considered is the failure intensity at 5 turbine subassemblies/year, which is a completely subjective assumption that is, however, useful for the comparison of systems.

6. The Aggregation Process

In some of the cases a further aggregation of data has proved necessary. The aggregation is the process of joining two or more time cells, so that the number of failures of the new aggregated cell is the sum of the original ones. The aggregation is necessary in two cases

- To achieve the minimum requirement of 5 events in each time cell, as specifically requested in [5]
- In the attempt of turning the result of the goodness of fit test from rejection to acceptance, with a "rationalization" of the data

The manipulation of the data, as in the second case above, is acceptable whenever the estimate of the curve parameter does not change radically, compared to the original values, that is the shape parameter remain reasonably similar. This practice has been proven successful to overcome the volatility of the data, due to the effects of environmental variability, which has a negative effect on the GoF test. Sometimes, the aggregation of few time cells has proven sufficient to achieve an acceptance.

7. Results

The following figures 2, 3 and 4 illustrate three noteworthy subassemblies from 3 different wind turbine models, one for each combination of result shown in table 2, covering the three turbine rating groups chosen.

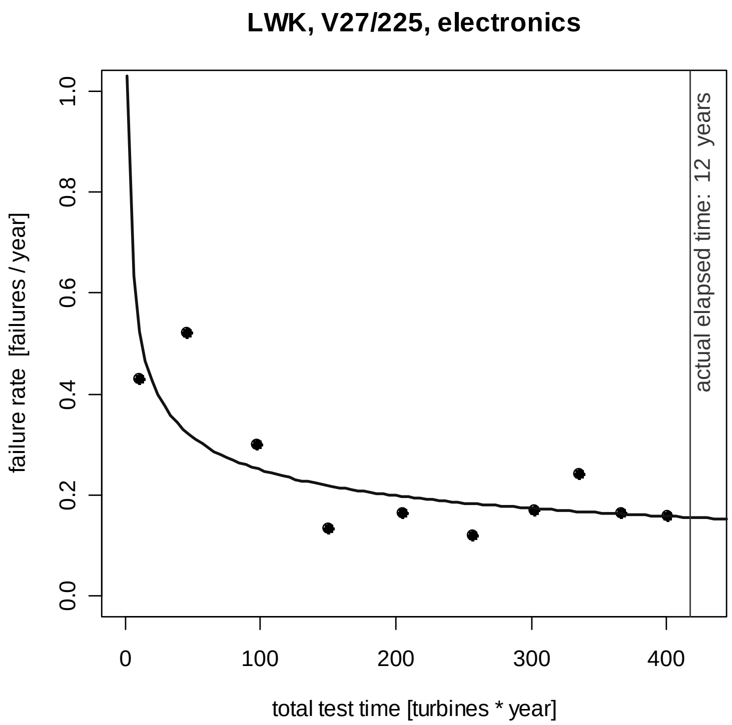

Figure 2: example of early failures, electronics of the Vestas V27, 225 kW

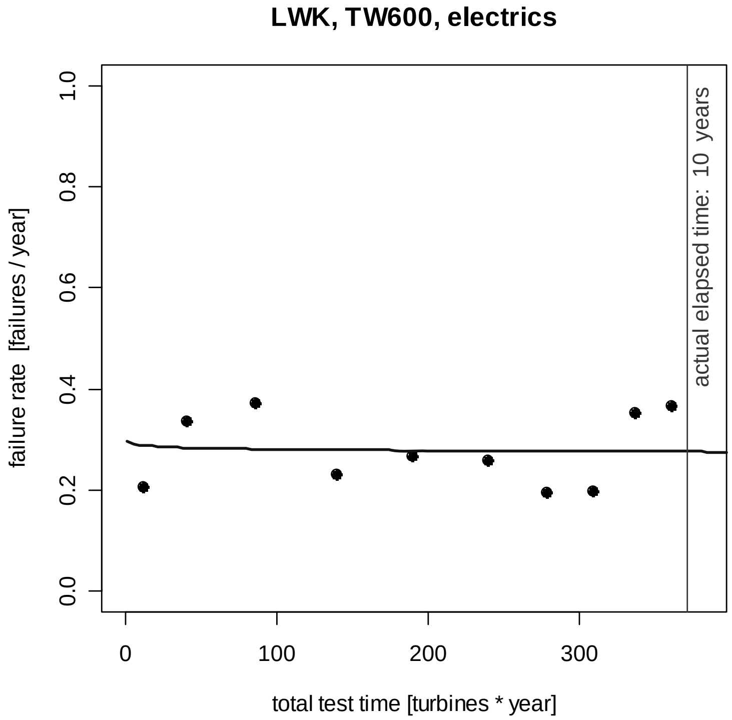

Figure 3: example of constant failures, electrics of the Tacke TW600 , 600 kW

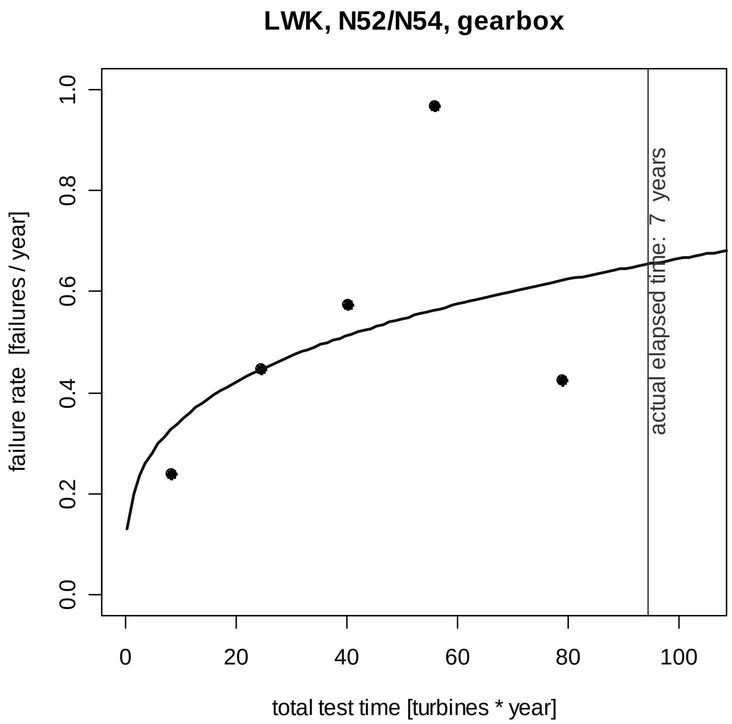

Figure 4: example of deterioration, gearbox of the Nordex N52-N54 , 0.8-1 MW

For all the figures the continuous line represents the resultant modelled failure intensity and the dots the average failure rate on each time cell, after the aggregation process. The quantities are plot versus the TTT, while a vertical cursor indicates the actual chronological time that has been elapsed over the test period. The average

number of turbines of each population is roughly the TTT divided by the elapsed time.
Strikingly, the results are consistent with other engineering evaluations made about the subassemblies and that is one of the objects of the analysis.
Figure 2 refers to the electronic equipment, and the downward trend typical of most electronic equipment is noted. Given the complexity of electronic systems with a large number of components, and the consequential difficulty of evaluating all problems during the design stage, the initial failure rate is understandably high. Reliability is improved after field data uncovers design deficiencies or weaknesses and new components or updated designs are implemented, and as shown the failure intensity then decreases. As explained in a previous section, the expected failure intensity is, in cases like this, the value determined by the intersection of the failure intensity curve with the vertical time cursor, at the end of the development phase.
Figure 3 refers to the electric equipment, which includes many different components and intrinsically a great deal of different failure modes. It is expected in this case that failures occur on a more random basis. For similar reasons, a constant failure rate is often found applying the mathematical model on failure data concerning the entire wind turbine.
Figure 4 refers to the gearbox of a large wind turbine. Gearboxes, due to their enormous impact on availability and the economics of the entire wind turbine, have attracted the interest of designers in the past whose effort have translated into improved reliability, so the technology of the gearboxes can be assumed to be "mature". At the same time it is likely, once a particular design is chosen for a certain turbine, that gearboxes are subject to a less developmental phase. Furthermore the massive mechanical torques, to which gearboxes are subjected, stresses internal components with a resultant perceivable degradation of materials. The expected result is that, for the reason, like the one shown in the picture where an increasing trend dominates the intensity of failure.

Note that for the deterioration phase a final failure rate value has no meaning, since the growth is indefinite while an initial failure intensity cannot be determined unequivocally, given the steep initial slope of the curve. The mathematical model parameters for the three cases are reported in table 5.

Case	β	λ	Ho = HPP	Result
V27 electronics	0.670	1.704	rejected	Early Failures
TW600 electrics	0.987	0.301	accepted	Constant Failures
N54/54 gearbox	1.280	0.140	rejected	Deterioration

Table 5: the results of the three analysed cases

In Table 6 the resulting failure intensity are compared with the averaged value over the, correspondent, entire period.

Case	Average	Model result	Comment
V27 electronics	0.156	0.233	Model Final Value
TW600 electrics	0.276	0.276	Average
N54/54 gearbox	0.509	0.140	Model Initial Value

Table 6: model results in comparison to the average

8. Conclusions

The Crow-AMSAA mathematical model has been applied successfully to German wind turbine fleet field data for the determination of subassembly failure intensity trends.
The method consists of the evaluation of the two parameters of the intensity function, with further support from two statistical tests; the first for the goodness of fit and the second for testing the no growth hypothesis.
Data volatility, which is due to environmental factors like wind gusts and grid disturbances and often causes the rejection of the hypothesis, can be overcome with a further aggregation of the data, whenever this does not drastically change the shape factor of the intensity function.
The data interpretation and the differentiation of the various cases, constitutes a fundamental part of the successful application of the mathematical model, in particular for the expected failure intensity:

- Early failures: the expected failure intensity is given by the mathematical model instant failure intensity at the end of the development phase, that is the observed period.
- Constant failures: the expected failure intensity is the average failure intensity, given by the ratio of the total number of failures and the TTT
- Deterioration: no unambiguous values can be extracted, conventionally the value of 5 turbine subassemblies/year has been chosen.

The assumption of constant failures and the adoption of average failure intensity is only valid in the case of no reliability growth.
When positive reliability growth occurs the final failure intensity must be chosen as the expected value. On the other hand when negative growth occurs the initial failure intensity should be used.
The distinction between these mathematical model results and average failure intensities, as calculated industrially, is important, because reliability trends cannot be neglected in the analysis of wind turbine reliability figures from fleet field data.

9. References

1. G.J.W. van Bussel: *Offshore Wind Energy, the Reliability Dilemma. P*roceedings of the World Wind Energy Conference, Berlin, Germany, 2-6 July 2002.
2. G.J.W. van Bussel*, State of the Art and Technology Trends for Offshore Wind Energy: Operation and Maintenance Issues,* Proceedings of the European Wind Energy Conference, Brussels, Belgium 10 – 12 December, 2001.

3. G.J.W. van Bussel, Chr. Schöntag, *Operation and Maintenance Aspects of Large Offshore Windfarms*, Proceedings of the European Wind Energy Conference, Dublin, Ireland, 1997
4. H.-J. Krokoszinski, *Efficiency and effectiveness of wind farms—keys to cost optimized operation and maintenance*, Renewable Energy, Volume 28, Issue 14, November 2003, Pages 2165-2178
5. P Tavner, C Edwards, A Brinkman, F Spinato, *Influence of Wind Speed on Wind Turbine Reliability* Wind Engineering, Volume 30, No 1, 2006, pp 55–72
6. MIL-HDBK-186, *Reliability Growth Management,* Department of Defence, Washington, 1981
7. W. Eggersglüß; *Wind Energie IX-XIV,* Praxis-Ergebnisse 1995-2004, Landwirtschaftskammer Schleswig-Holstein, Osterrönfeld, *Germany.*
8. *Windstats*, www.windstats.com
9. P Tavner, J. Xiang, F Spinato, *Reliability Analysis for Wind Turbines*, Wind Energy, Volume 10, Issue 1, January/February 2007.