USE OF SURVIVAL MODELS IN A REFINERY

Authors: SÍLVIA MADEIRA

- Phd in Mathematics, ECT of University of Évora, Portugal silvia.madeira@galpenergia.com
- PAULO INFANTE
- CIMA/DMAT, ECT of University of Évora, Portugal pinfante@uevora.pt
- FILIPE DIDELET
- EST, IPS Setúbal, Portugal filipe.didelet@estsetubal.ips.pt

Abstract:

• Statistical methods are nowadays increasingly useful in industrial engineering. From plant design reliability to equipment analysis, there is much to cover with statistical models in order to improve the efficiency of systems. At Sines refinery we found it useful to apply a Cox model to a particular critical equipment trying to find process variables that cause its vibration as well as to apply well known distributions to baseline hazard rate.

Key-Words:

• reliability; maintenance; time between failures; Kaplan-Meier; Cox proportional hazards.

AMS Subject Classification:

• 62N01, 62P30, 90B25.

1. INTRODUCTION

As industries became more competitive, their methods to improve results are intimately related to efficiency. Although efficiency can be achieved with technology, it only can be reliable based on the fact that the system is prepared for failures, managing them the best way. Failures can be classified as follows: avoided, predictable or inevitable. In each case, knowing equipment behaviour helps companies to manage spare parts, man-hours and maintenance issues that will improve reliability and save money [12]. Reliability studies are then a priority and this work has been written to support decisions based on reliability models and according to standards [8].

Sines refinery has been concerned with the shut-downs made by the Turbo-Expander equipment in the FCC unit (Fluid Catalytic Cracking). As it was having problems due to a vibration failure mode causing FCC unit shut-down since year 2000, it was decided to investigate the origin of the vibration. Expander manufacturers and other entities that have this kind of equipment have investigated this problem as it is a big economical issue for the companies as we can see in [4] and [6]. They highly recommend investigating efficiency in order to detect scaling deposition in expander rotor blades. It is believed that the composition of some particles resulting from process reaction are the key for scaling, not just erosion. On 2011 turnaround, the procedure of changing the expander's rotor and shroud was not complete. Only the rotor was changed. This has caused a slight gap between the shroud and the rotor blades due to shroud erosion. This gap is believed to cause less resistance and then, less particles deposition. However, results of Cox Proportional Hazards [2] [3] demonstrate components in these particles to be linked with high values of expander vibration and with times to failure. In Section 3 we have a brief description of the equipment and contextualization of the subject. In Section 4 we will refer to the goal of this work and variables definition. Then we will present some parametric and non-parametric approaches using Kaplan–Meier estimator and Cox Proportional Hazards in Section 5, and a parametric adjustment to the null model with parametric models. Finally, Section 6 is dedicated to some general considerations of this work.

2. MODELS

Several approaches were tried and reviews for different models were studied, although not all of them could fit on the data and conditions of our study. Cox Proportional Hazards had a strong focus because of its flexibility, but some approaches using Competing Risks Theory as we can see in Fine and Gray [5] or Lunn and McNeil [11] were not possible due the complexity of the system. As we have both time-dependent and independent events it was very difficult to articulate a model and find covariates that could meet the assumptions needed. Additive interaction used by Li and Chambless [10] was proposed but due to the nature of the data and the way that some covariates are monitored make it impossible to use. We will then use the best possible models to fit our data that meet the required conditions for the assumptions needed.

3. FRAMING AND DEFINITIONS

3.1. Configuration of the FCC Power Recover Unit at Sines refinery

A PRU (Power Recover Unit) is composed of an expander, a main air blower, a turbine, a gear box and a motor/generator which recovers the flue gas from the process to generate energy and steam. At Sines refinery we can find a particular configuration of the PRU as we can see in Figure 1. This configuration



Figure 1: FCC and PRU train.

brings an additional problem to the system. The fact that the expander is coupled with the other equipment leads to unwanted shut-downs of all FCC units whenever the expander has a shut-down. As a shut-down implies high costs, the main goal is to avoid it. Pareto analysis was done on all PRU failure modes so the main failure mode can be easily identified. It was clear that the vibration failure mode was the principal reason for the problems in the expander. Actually, this equipment is supposed to have a reliability of 99%, and its only intrinsic failure is vibration.

3.2. Expander — What is it and how does it work?

The Turbo-Expander (Figure 2) is composed of the nose cone, the rotor blades, the stator ring, the shaft and the casings. Process flue gas reaches the expander rotor blades at a pressure of about 2.1 barg and a temperature of 700° Celsius degrees. The flow at this pressure and temperature is here transformed



Figure 2: Turbo-Expander.

into mechanical energy, making the rotor blades rotate as well as the shaft at approximately 5700 rpm, held by the steam turbine. This mechanical energy is thus transformed in electrical energy through the generator that is coupled with the expander in the same shaft. In Figure 3 we can see process flue gas come inside the expander casing (grey arrows) that reaches the rotor blades being cooled (white arrow), and then, exhausting through the exhaust casing (black arrows). During this process some particles can set down on the rotor blades. As deposition may not be uniformly distributed by the blades, it will cause imbalance at high rotation, and thus, cause vibration. The trip value is set to $v \mu m$ (microns) — where v is a predefined target — and when this value is reached, either the trip (shut-down due to go beyond the threshold values) can occur or operational staff can choose to try to make a controlled shut-down once it is unavoidable.



Figure 3: Turbo-Expander flue gas flow.

4. TIMES TO EXPANDER VIBRATION FAILURE MODE

In order to investigate the root cause of the problem, all shut-down events were recorded and variables that were suspected to be related with the expander vibration were recorded through the on-line monitoring system of all instrumented variables, since year 2000. There were also concerns with the fact that vibration due to imbalance in the rotor blades caused by deposition can be due to mechanical reasons, chemical reasons, or a combination of both. As flue gas is the product of combustion of coke, compounds that are produced in this reaction can be carried out with flue gas and reach the expander. Tiny particles as well as some chemical combinations that can produce a kind of glue effect can cause deposition. This is a theory supported by the industrial community [4] and the present work has demonstrated some compounds and their characteristics to be important reasons. However, in the past, some internal mechanical damage either in the regenerator and 3rd stage separator have been shown to be a good reason for particles to easily reach the turbo expander and cause vibration. To explain the reasons that lead to scaling is not the subject of this work, but the fact that it may exist. However, it is important to refer that we have several hypotheses for scaling and none must be rejected.

Reasons for scaling:

- a) Erosion combined with more tiny particles can cause deposition;
- **b**) Some chemical compounds combined with each other can act as glue and cause deposition;
- c) Vapour is not at the correct temperature/pressure and will cause scaling combined with flue gas particles;

Reasons for high concentrations of tiny particles to reach the expander:

a) Internal damage in the regenerator, or internal damage in the $3^{\rm rd}$ stage separator.

Internal damage is not easily detected which makes this a major challenge. Scaling can be due a combination of factors that we actually don't know if they are happening or not at the same time. Another problem is that the point of collection of particles for analysis is not immediately before the expander so, in practice, we can only infer what's reaching expander from particles that are being regenerated.

From all possible variables that can influence the system, we have:

variable	description	variable	description
vnd	Vanadium (ppm)	с	Carbon (%)
ni	Nickel (ppm)	re	RE_2O_3 – Rare-earth oxides (%)
fe	Iron $(\%)$	abd	Apparent Bulk Density (g/cc)
cu	Copper (ppm)	alo	Al_2O_3 Alumina Oxide (%)
pb	Lead $(\%)$	aps	Average Particle Size (microns)
na	Sodium (%)	pv	Particules' Pore Volume (cc/g)
po	P_2O_5 – Phosphorus Pentoxide (%)	mat	Micro activity Test
mgo	MgO – Magnesium Oxide (ppm)	sa	Particles' Surface Area (m ² /g)
cf	Coke Factor	vpb	Vapour pressure (barg)
gf	Gas factor	vib	Vibration Values (microns)
pin	Inlet Pressure (barg)	car	Main Air Blower Flow (%)
pout	Exhaust Pressure (barg)	cr	Reactor Feedstock (t/h)
tpin	Inlet Temperature (°C)	bp	Bypass Valve $(\%)$
tpout	Exhaust Temperature (°C)	pt	Particles Size $(\%)$

 Table 1:
 Variables description.

5. SURVIVAL MODELS

Several approaches were tried to correlate some particles' compounds and attributes as well as with process variables such as temperature, pressure, among others, with times to vibration failures using survival models.

- First approach Use the times to expander failure and correlate them with particles' compounds and attributes and with process physical variables. For this, we have recorded all shut-down events since year 2000, censoring those that were not by vibration. Whenever there is a shut-down, a "as good as new condition" is reached for turbo-expander due to the thermal shock produced by equipment cooling.
- 2. Second approach Use times to high vibration and correlate them with particles' compounds and attributes and with process physical variables, as high vibration values can end in a shut-down or not. High vibration values are harmful for the equipment, and they may not cause a shut-down but they must be avoided, and are here treated as an event. For this, we have recorded all shut-down events and high vibration values (higher than $v/3 \mu m$) since year 2000, censoring shut-downs that were not by vibration. Whenever there is a shut-down or a high vibration value and the value drops again after some short time, a good as new condition is considered.

5.1. Kaplan–Meier estimators

First, we have made a non-parametric approach using the Kaplan–Meier estimator [9] (using software R) to obtain the survival curves for times to expander vibration failure mode. As we have right censored data and the intervals between events are typically non uniform, Kaplan–Meier is a good approach.

Let R(t) be the probability that a member from a given population will have a lifetime exceeding t. For a sample of size N from the list of observations, let the observed times until the shut-down of the N sample observations be $t_1 \leq t_2 \leq t_3 \leq \cdots \leq t_n$. Corresponding to each t_i is n_i , the number "at risk" just prior to time t_i , and d_i , the number of shut-downs at time t_i . The Kaplan–Meier estimator is the non-parametric maximum likelihood estimate of R(t). It is a product of the form:

(5.1)
$$\hat{R}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$$

5.1.1. First Approach — Only shut-downs are considered

In these Kaplan–Meier curves for times to expander vibration failure mode, we know that reliability keeps high for values under $v/2 \mu m$. Therefore, one of the goals here is to predict when vibration values become dangerous and a potential spark for the fast increase in its values. Although reliability never drops below 50% (Figure 4), it is important to find the reasons for the shut-downs as they are an important economical factor.



Figure 4: Kaplan–Meier for the null model — Approach 1.

5.1.2. Second Approach — High vibration values are considered

The curve in Figure 5 shows us that high vibration values are recurrent although shut-downs are not. The reliability values quickly decrease and when t_1 is reached, we have less than 50% of reliability for high vibration values.



Figure 5: Kaplan–Meier for the null model — Approach 2.

The question here, is that high vibration values can lead to one of two situations: — particles deposited in rotor blades can drop due to imbalance caused by high vibration, or — imbalance can increase as a result of high vibration and get out of control leading to trip vibration values and cause a shut-down. Here is just a matter of "mechanical free will". That's why it is important to check causes for high vibration values.

5.2. Cox Proportional Hazards Model

Survival models were studied, first with a non-parametric approach with Kaplan–Meier, and then, a semi-parametric approach was tried using Cox proportional hazards. Cox model allows analysis of life times in which the outcome variable is the time until the latest event, being censoring or failure, and is characterised by the coefficients (β 's) which measure the effects of covariates on the hazard rate:

(5.2)
$$h(t) = h_0(t) \exp\{\beta_1 x_1 + \dots + \beta_p x_p\},\$$

with:

- h_0 : baseline hazard rate function;
- β_1, \ldots, β_p : model parameters;
- x_1, \ldots, x_p : explanatory covariates.

Let Y_i denote the observed time (either censoring time or event time) for observation i, and let C_i be the indicator that the time corresponds to an event (i.e. if $C_i = 1$ the event occurred and if $C_i = 0$ the time is a censoring time, which is, a general shut-down). We used the partial likelihood method for the parameters using Breslow's [1] estimate:

(5.3)
$$L(\beta) = \prod_{i:C_i=1} \frac{\theta_i}{\sum_{j:Y_j \le Y_i} \theta_j} ,$$

with

(5.4)
$$\theta_j = \exp\{\hat{\beta}_j\}$$

As it is a multivariate regression, correlation between covariates must be carefully studied. Once vibration can be consequence of the other covariates, it was not included for now in our models. Models were developed in order to avoid joining correlated covariates, which had a Pearson's correlation value above 0.75 or because experts show us that chemically speaking, when some covariate increases, another one will also increase (or vice-versa). All variables here presented are continuous, and whenever a covariate is as significant as when categorized as when continuous, it will be used in its categorized form, in order to simplify model interpretation. Models adjustments were made according Hosmer and Lemeshow [7] modelling stages.

5.2.1. First Approach — Only shut-downs are considered

Univariate analysis was made to the covariates and those with significance below 25% were considered in the multivariate model. Both forward and backward methods were tested using R software, and the best fit was achieved for each. Because this is a dynamic system, each time there is a shut-down, the event is uploaded to the database and the model is tested again. Some variables were added to the initial tested model because they were shown to be relevant, and have, individually, good explanatory values. However, sometimes when together, they have poor explanatory values. Variable vib corresponds to the vibration and shall not be accompanying other variables in the same model due to collinearity. From the possible 28 covariates, only 22, according the criterion of correlation, can be used. Only 10 from these variables are significant at 10%. Choosing only those variables with p-values below 25% of significance and after using backward and forward techniques, only 8 of them were used, and Model 1 was reached as shown in Table 2. Previously, in an initial approach we have used two different models, one for chemical and another one for physical variables, but with the introduction of new variables, this has shown not to be the best solution.

variable	β	$se(\beta)$	<i>p</i> -value
nix	-3.12	0.88	0.0004
fe	-11.18	5.68	0.0492
sa	0.11	0.03	0.0003
na	-24.01	6.63	0.0003
vpb	0.86	0.36	0.0179
mgo	43.53	20.26	0.0316
cfx	-328.5	97.29	0.0007
tpin	-0.24	0.06	0.0000
cfx:tpin	0.46	0.14	0.0007

Table 2:Cox model 1.

62.8% of the variation can be explained by this model, with a concordance value of 0.899 and a likelihood ratio test with a *p*-value effectively zero. Variable *sa* is the surface area of the particles and seems to be an important variable to take into consideration. This variable explains about 18.4% of data variation in its

univariate analysis. An increase in the surface area of the particles means that they are more able to break and produce fine particles, and the increase of one unit of sa increases the risk by about 5.5%. Coke factor (cf) and vanadium vndare important variables as well as they explain 15% and 11.5% of data variation respectively in the univariate analysis. In Model 1 sa has an associated risk of 11.6%. Except for sa, mgo and vpb covariates, all the other covariates that are not in an interaction (*ni*, *fe*, *na*), are indicating that if they are trending downward, the risk will largely increase. We know also that for the increase in one unit of vpb the associated risk increases almost 2.5 times keeping the other covariates constant. Magnesium can be a contaminant metal when combined with other components, and makes sense that the increase of one unit will exponentially increase the covariate effect. Variable cfx is the categorized variable (0 and 1) for coke factor and the cut-off point used for it was its mean because of its better interpretation, and if it is trending downward it may indicate a higher risk. Inlet temperature tpin can be a good monitoring variable for the same reason as cf. Coke factor is related with temperature in the regenerator and thus, also with the quantity of contaminant metals, so this interaction makes all sense. Statistically speaking, we can say that for the cfx value of 0, the risk will decrease regardless of the value of the inlet temperature (tpin), but if the cfx value is 1, the risk will increase, with a higher risk (287 times) for inlet temperature values under the 1st quartile. Also this means that for inlet temperatures above its mean, and when cfx goes from 0 to 1, the risk starts to increase as we can see in Figure 6.



Figure 6: Interaction *cfx:tpin*, *tpin* fixed.

In Figure 7a we have the survival estimates for the quartiles and in Figure 7b we have the Kaplan–Meier estimate of model 1. Model 1 is quite satisfactory, with good results in its residuals analysis, and linear correlation tests were made and they suggest that none of the covariates violate the proportional hazards assumptions (see Appendix A). Realistic possible scenarios can be used and the survival estimates are given for two examples in Figure 8a and Figure 8b. As we can see, when the magnesium variable increases its concentration, reliability decreases.



Figure 7: Estimate analysis for model 1.



Figure 8: Reliability for different scenarios.

Model 2 uses the second approach and supports the idea that our event is "high vibration". Times to high vibration values and shut-downs are here analysed instead of only times to shut-down failures. Thus, using the same process variables, we achieve the Cox model 2 in Table 3. In this model, more than finding the reasons for shut-downs, is to find the reasons for scaling. High vibrations can be a spark for a shut-down and the line that bounds the two situations (shut-down or not) is just a question of the way the scaling is being heterogeneously distributed on the rotor blades, and how much imbalance it will cause. Comparing these results with the previous model (model 1), we found some variables in common.

^{5.2.2.} Second Approach — High vibration values are considered

variable	β	$se(\beta)$	<i>p</i> -value
fe	-6.49	2.01	0.0013
tpin	-0.05	0.01	0.0002
tpout	-0.01	0.01	0.0278
re	-0.95	0.28	0.0008
mgo	-66.48	21.33	0.0018
na	-202.8	58.24	0.0005
cf	-3.22	1.77	0.0691
gf	-2.81	0.69	0.0000
mgo:na	240.2	68.87	0.0005
cf: gf	1.65	0.49	0.0008

Table 3:Cox model 2.

Nevertheless, some covariates that are not in common seem to have individually, a high significance in both models, such as sa and vpb. However, when together in the models they lose their significance. Approximately 35% of data variation are explained by this model, with a concordance value of 0.74 and a likelihood ratio test with a *p*-value effectively zero. We can see that the common covariates with model 1 (*fe*, *tpin*, *mgo*, *na* and *cf*), are monitoring variables as well as exhaust temperature (*tpout*). Magnesium oxide (*mgo*) is a very problematic component when combined with sodium (*na*, which is a poison for the process and combined with some metals can act as "glue"). For *mgo* and *na* interaction (Figure 9) we



(b) Interaction mgo:na, na fixed.

Figure 9: Interaction *mgo:na* for model 2.

have that for magnesium values fixed above the 3^{rd} quartile, the risk can increase from 22% to 4 times as much depending on sodium increase, but we have a clear interaction when we set na at its maximum value and we see that the risk increases 18 times for values below mgo's 1^{st} quartile, although it always increases for all mgo quartiles' variations. Coke factor (cf) indicates the amount of produced coke in the process and it can be a monitoring variable as it isn't a protective factor. Gas factor gf is concerned with the amount of produced gases and it can be read as a monitoring variable as well. In Figure 10 we can see that for coke factor and gas factor interaction, if we set cf value for its maximum, we will have a large risk increase when we increase gf. At the same time, when we set the values for gf and make cf increase, the hazard rates will increase for high values of gf.



(b) Interaction *cf:gf*, *cf* fixed.

Figure 10: Interaction *cf:gf* for model 2.

We can see in Figure 11a, that we have a better survival estimate than in model 1, with a good fit to the quartiles. Linear correlation suggest that none of the covariates violate the proportional hazards assumption as for model 1, and the residual analysis didn't show problematic points that may be interfering in the model (see Appendix B). As with model 1, two different scenarios were considered for model 2. We can see in Figure 12a the reliability curve for model 2 considering the mean for all variables. As we increase cf (Figure 12b) keeping all other variables constant, we see that the reliability curve have a higher slope and decreases faster.



Figure 11: Estimate analysis for model 2.



Figure 12: Reliability for different scenarios.

5.3. Parametric Models

In order to have a first parametric model, we have proceeded to parametric approaches for the baseline hazard rate determination. As we can see, none of the tested distributions are presenting a good fit, however, the log-normal has the lowest AIC (Figure 13). In order to find a better parametric fit, only times before t_1 were considered. We can see a better approach is achieved with the log-normal distribution in Figure 14, which has the lowest AIC. According to the AIC value we also have the log-normal distribution as a good fit for the baseline hazard as shown in Figure 15.



Figure 13: Parametric models for the null model (model 1).



Figure 14: Parametric models for the null model (model 1) until time t_1 .



Figure 15: Parametric models for the model 2.

6. CONCLUDING REMARKS

For the proposed problem in Section 4, the consistency of some covariates in tested models make them a subject of investigation. Some covariates such as surface area should be definitely monitored, as well as the inlet and exhaust temperatures trend. It seems that sodium, magnesium and iron are influential variables for the increase of the risk of high vibration values, and should be monitored, although all these variables are very difficult to control as they depend on the reactor feedstock and reactor temperatures. Investigation on this subject will be ongoing as it can help to find the reasons for the so non-welcome shutdowns. Future work will be done to mechanical equipment in order to optimize, if possible, predictive maintenance scenarios.

APPENDIX A

Here we have the residual analysis for model 1. We can see that proportionality assumption is verified when we analyse Figure 16. Deviance residuals and martingale residuals where analysed as well as score residuals and for all cases possible outliers or influential observations where removed to check the availability of the model. In all cases the coefficients never changed their values above 25%.



Figure 16: Schoenfeld residuals for model 1.

APPENDIX B

As for model 1 the proportionality assumption was checked via Schoenfeld residuals. We can see that proportionality assumption is verified when we analyse Figure 17. Just like model 1, model 2 has been checked for influential observations with no significant change in its coefficients.



Figure 17: Schoenfeld residuals for model 2.

ACKNOWLEDGMENTS

This study was supported on behalf of FCT (Portuguese Foundation for Science and Technology) and Galp Energia, the insertion company for the PhD. The second author is member of CIMA-UE, research center supported financially by the FCT.

REFERENCES

- [1] BRESLOW, N.E. (1972). Regression models and life tables (with discussion) by D.R. Cox, *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.
- [2] Cox, D.R. (1972). Regression models and life tables (with discussion), *Journal* of the Royal Statistical Society, Series B, **34**, 187–220.
- [3] COX, D.R. and OAKES, D. (1984). Analysis of Survival Data, Chapman and Hall, London.
- [4] DIANGU, H. (2011). Analysis on causes of scaling in flue gas turbine of FCCU and countermeasures, *China Petroleum Processing and Petrochemical Technology*, 13(1), 66–74.
- [5] FINE, J.P. and GRAY, R.J. (1999). A proportional hazards model for the subdistribution of a competing risk, *Journal of the American Statistical Association*, 94, 496–509.
- [6] GLADYS, N. and LAURA, V. (1999). Analysis of rotor-blade failure due to hightemperature corrosion/erosion, Surface and Coatings Technology, 120–121, 145– 150.
- [7] HOSMER, D.W. and LEMESHOW, S. (2000). *Applied Logistic Regression*, Wiley, New York.
- [8] ISO 14224:2006 (2006). Petroleum and Natural Gas Industries Collection and Exchange of Reliability and Maintenance Data for Equipment, International Standard Organization, Second Edition.
- [9] KAPLAN, E.L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations, J. Amer. Statist. Assn., 53, 457–481.
- [10] LI, R. and CHAMBLESS, L. (2007). Test for additive interaction in proportional hazards models, Annals of Epidemiology, 17(3), 237–236.
- [11] LUNN, M. and MCNEIL, D. (1995). Applying Cox regression to competing risks, *Biometrics*, 51(2), 524–532.
- [12] RAUSAND, M. and HOYLAND, A. (2004). System Reliability Theory Models, Statistical Methods and Applications, Second Edition, Wiley.