11th World Congress on Computational Mechanics (WCCM XI)
5th European Conference on Computational Mechanics (ECCM V)
6th European Conference on Computational Fluid Dynamics (ECFD VI)
E. Oñate, J. Oliver and A. Huerta (Eds)

RARE EVENT ANTICIPATION AND DEGRADATION TRENDING FOR AIRCRAFT PREDICTIVE MAINTENANCE

S. ALESTRA[1], C. BORDRY[2], C. BRAND[1], E. BURNAEV[1,3,4], P. EROFEEV[1,3,4], A. PAPANOV[1,3] AND C. SILVEIRA-FREIXO[2]

[1]DATADVANCE, Pokrovsky blvd. 3, Moscow, 109028, Russia,
e-mail: stephane.alestra@datadvance.net, christophe.brand@datadvance.net,
evgeny.burnaev@datadvance.net, pavel.erofeev@datadvance.net,
artem.papanov@datadvance.net
web page: http://www.datadvance.net

[2]AIRBUS, St. Martin du Touch, 316 route de Bayonne, 31060 Toulouse Cedex 9, France,
e-mail: christophe.bordry@airbus.com, cassiano.freixo@airbus.com
web page: http://www.airbus.com

[3]IITP RAS, Bolshoy Karetny per. 19, Moscow, 127994, Russia,
web page: http://www.iitp.ru

[4]PreMoLab, MIPT, Institutsky per. 9, Dolgoprudny, 141700, Russia,
web page: http://www.premolab.ru

Key words: Predictive Maintenance, Data Mining, Rare Event Anticipation, Degradation Trending, Aircraft Health Management

Abstract. In this paper we examine problem of predictive maintenance in complex technical systems. We propose two approaches for anticipation of rare events (typically faults): 1) degradation detection and trending, and 2) failure discrimination based on classification techniques. The applicability of the approaches is illustrated on the real-world test cases from aircraft operations based on the data granted by AIRBUS.

1 INTRODUCTION

In recent years the concept of predictive maintenance in complex technical systems is gaining popularity. It is designed to help determine the condition of in-service equipment in order to predict when maintenance should be performed. This approach promises cost savings over routine or time-based preventive maintenance, because actions are performed only when warranted. To the date there exist several successful application of the concept in the different areas of technology including US navy cost reduction for maintenance [1], increase of safety and reliability of distributed power systems [2, 3] and automated search for faults in power systems [4]. In this study we examine problem of the effective aircraft maintenance in operation.

Aircraft is a complex systems with tens of interacting subsystems and thousands of underlying parts. Naturally, as an aircraft life progresses, these parts are exposed to different types of stress which, eventually, originate faults, i.e. abnormal conditions, or non-permitted deviations regarding some fundamental properties of the components.

Sometimes faults lead to failures, i.e. the permanent interruption of the aircraft's ability to perform some required functions under specified operating conditions, and therefore the maintenance appears as a fundamental activity to preserve aircraft operability.

From a maintenance standpoint, the first approach to preserve this aircraft operability is to perform scheduled activities, such as structural inspections or electronic tests. This is when unscheduled maintenance activities take place, aiming at restoring, as quickly as possible, some aircraft functions. Usually the cost of this unpredicted maintenance is problematic and heavy, and such events have to be avoided as best as possible.

To try to decrease these unscheduled costs, AIRBUS is interested by predictive maintenance concept as a methodology of failure anticipation and warning monitoring function to decide whether a operability-related failure is present in the aircraft before a fault actually occurs.

Predictive maintenance for aircrafts involves data collection, handling and processing. In this paper we describe two approaches for aircraft failure anticipation illustrating them applied to two 'prove-of-the-concept' examples from real aircraft operations. The goal of the ongoing project in Airbus is to develop a full support automated system for the early warnings for possible costly faults.

In order to build up such a system several problems should be considered. The first is a multidimensional data trending to be able to trace the known degradation processes. And the second is a problem of rare event prediction to be able to anticipate some specific families of faults.

Classical statistical approaches are ineffective for low frequency and high consequence events because of their rarity. In this paper we try to adapt existing approaches for rare event prediction.

2 RARE EVENTS ANTICIPATION PROBLEM STATEMENT

From mathematical point of view the rare event anticipation problem can be formulated in the following way [5]. We observe in real time (with some frequency, possibly not uniform) starting from moment T multivariate time series of system performance parameters $X_t \in \mathbb{R}^d, t > T$. The time series before moment T are also known, i.e. we have also a *historical data set* $D = \{X_t\}_{t=1}^{T}$. The task is to predict the new events $Y_t \in [0, 1]$[1], $t > T$, where Y_t is an indicator of event occurrence at moment t.

[1]Here $Y_t = 1$ means that event happened at moment t and $Y_t = 0$ – event did not happen.

Based on information available two possible problem statements are considered:

1. **Unsupervised case**. Only the set of parameters X_t is considered at each moment of time t and we try to detect that system state has changed and correlate that possible changes to Y_t. In this case we build up some statistical model using historical data set $\{X_t\}_{t=1}^T$ and given new observations of $X_t, t > T$ we check if they expose the same statistical properties as those in historical period and obtain probabilities of 'abnormal' behavior p_t. Given the particular threshold on probability, $\nu \in [0, 1]$, one can calibrate the final prediction: $F_t = I\{p_t > \nu\}$[2] – 'abnormality' flag for moment $t > T$.

2. **Supervised case**. Both X_t and Y_t are considered and we try to predict directly Y_t based on previous history of X_t and $Y_t,\ t < T$. In this case we have the historical data set $\{X_t, Y_t\}_{t=1}^T$ and build the predictive model

$$f\colon X_{t-l-h}, X_{t-l+1-h} \ldots, X_{t-h} \to [0, 1],$$

where l is a lag in history and h *is a prediction horizon*, i.e. the number of time moments ahead we try to predict the event. Given new observations, $X_t, t > T$, we predict: a) probability of events with the model $p_t = f(X_{t-h}, X_{t-h-1}, \ldots, X_{t-h-l}), t > T$; and b) flag of early warning $F_t = I\{p_t > \nu\}$, given the threshold value $\nu \in (0, 1)$.

2.1 SOLUTION QUALITY METRICS

To measure the quality of the prediction we consider the following metric: accuracy of prediction of failure event on particular moment of time (flight, day, week, etc.), i.e. each moment of time we have a prediction of failure (or probability of failure) and we need to assess the accuracy. If $Y_t = F_t$ than the prediction is correct, otherwise not.

The possible outcomes of the prediction are summarized in the Table 1. And the requirement on predictive algorithm is to minimize the level of Type II errors (false alarms) while given the reasonable prediction of failure events.

2.1.1 Precision-recall curve

The well-known classification quality metrics, receiver operating characteristics (or simply ROC curve), is not applicable in our case due to rarity of one class – events. It gives quite optimistic and misleading representation of false alarms as the number of non-event states marked as corresponding to events (see e.g. [7]). Even if we have only 1% of such false alarms, in absolute values it could be comparable to the number of events itself and not feasible.

[2] Hereinafter notation $I\{bool\}$ is an indicator of boolean condition *bool*, i.e. $I\{bool\} = 1$ if *bool* holds and $I\{bool\} = 0$ otherwise.

		Condition		
		No Failure	Failure	
Prediction	Failure	Type I error False positive	Correct outcome True positive	**Precision** = $\sum$ True positive / $\sum$ Predicted Failure
Prediction	No Failure	Correct outcome True negative	Type II error False negative	**Negative predicted** = $\sum$ True negative / $\sum$ Predicted No Failure
		Specificity = $\sum$ True negative / $\sum$ No Failure	**Sensitivity** = $\sum$ True positive / $\sum$ Failure	**Accuracy**

Table 1: Classification error definitions.

A precision-recall curve is much more applicable in our case [7]. It is another graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold ν is varied. But compared to the ROC it is more useful in the case when one class has very limited number of examples (e.g. failure events are quite rare). It is created by plotting the fraction of true positives out of the total actual positives, precision of classifier, vs. the fraction of true positives out of all positive answers of classifier which is called recall.

Using the curve one can decide which algorithm is more suitable for the particular classification task and also choose the optimal threshold ν^* based on the cost-benefit analysis, i.e. fix the false alarm rate to get the corresponding precision or vice versa. That is how the user of the tool can finally control over false alarm and precision trade-off.

3 PROPOSED APPROACHES

We distinguish two types of problems described below: degradation detection and fault prediction. The first step in both approaches is accurate selection of features describing the system behavior (see the next subsection for details).

3.1 PHASE SPACE NOTATION

In order to define correct mathematical terms for analysis one should define the so called *phase space* of the problem. This terms should represent the most important features of the system that expected to be the subjects of some change preceding the failure events. This step is done just after the selection of particular physical parameters to observe. There is no generic procedure for construction of phase space because it is problem specific and relies on the physical knowledge. But there exist several generic approaches, including

- Usage of raw parameter values. In case of no additional physical information or the parameters describe the system in the best way.
- Calculation of first differences. In some cases it is more important to look not at the absolute values of some parameters but at the empirical derivative of the process, i.e. the first differences. For a given parameter time series $x_t, t = 1, \ldots, T$, the series of first differences is defined as $\tilde{x}_0 = 0, \tilde{x}_t = x_t - x_{t-1}$ for $t = 2, \ldots, T$. Consider the case of degradation when the absolute level of the parameter is not much important but the crucial feature is a sequential decrease in parameter value corresponding to some kind of 'leakage'.
- Calculation of some statistical properties like second moment (variance), third moment (kurtosis), etc. These features could be representative in case of complex but not obvious changes in the system.
- Correlation analysis. In some cases the most important feature of the data is its temporal correlation. For two time series x_t and z_t for $t = 1, \ldots, T$, the temporal correlation is defined as

$$\mathrm{Corr}(x_t, z_t) = \frac{\sum_{t=1}^{T}(x_t - \overline{x})(z_t - \overline{z})}{\sqrt{\sum_{t=1}^{T}(x_t - \overline{x})^2 \sum_{t=1}^{T}(z_t - \overline{z})^2}},$$

  where $\overline{x}$ is an average value of corresponding parameter x. Consider, e.g., the case when one part of the system is expected to be broken and loose its connection to the other parts which are supposed to operate in normal way.
- Dimensionality reduction [8, 9]. Sometimes it is not obvious which parameters or combination of parameters are the most important for the system. In such cases one can use a generic dimensionality reduction techniques both to reduce the number of parameters to analyze and to detect the most important ones from the statistical point of view.

3.2 DEGRADATION DETECTION AND TRENDING

This approach covers simple cases when we know which parameters are involved in multidimensional degradation process (e.g. leakage, increasing pressure due to filter clogging, etc.) and we know the threshold of degradation before it becomes critical. In this case, we propose to build up a statistical model of 'normal' behavior and then apply it to the the new-coming data. In case degradation ('abnormal' behavior) is detected, the next step is to predict time before the fault. From mathematical point of view we are in unsupervised problem statement (see section 2).

3.2.1 Degradation detection

For detection of degradation we use One-class Support Vector Machine [10]. This method is extensively used for *novelty detection* and basically separates all the data points from the origin (in feature space) with a hyperplane and maximizes the distance from this hyperplane to the origin. This results in a binary function which captures regions in the input space where the probability density of the data lives. Thus the function returns 0 in a 'small' region (capturing the training data points) and 1 elsewhere (anomalies).

The quadratic programming minimization function in this case:

$$\min_{w,\xi_i,\rho} \frac{1}{2}\|w\|^2 + \frac{1}{\nu T}\sum_{i=1}^{T}\xi_i - \rho, \text{ such that}$$

$$\begin{aligned} &(w, \phi(x_i)) \geq \rho - \xi_i, && i = 1, \dots, T \\ &\xi_i \geq 0, && i = 1, \dots, T \end{aligned}$$

In this formula parameter ν characterizes the solution:

- it sets an upper bound on the fraction of outliers (training examples regarded out-of-class) and,
- it is a lower bound on the number of training examples used as Support Vector.

Due to the importance of this parameter, this approach is often referred to as ν-SVM.

3.2.2 Degradation trending

For the modeling of multidimensional degradation behavior in time we employ Autoregressive Integrated Moving Average (ARIMA) model [11], that can be viewed as a 'cascade' of two models. The first is non-stationary drift:

$$Y_t = (1 - L)^d X_t,$$

where L is a lag operator, i.e. $L^k X_t = X_{t-k}$; while the second is wide-sense stationary:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) Y_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t,$$

where ϕ_i are the parameters of the autoregressive part of the model, θ_i are the parameters of the moving average part and ε_t are error terms. The error terms ε_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean.

3.3 DISCRIMINATION OF FAILURES

We propose to use two-class classification models in order to perform discrimination of parameter values corresponding to the event states (faults). From mathematical point of view we are in supervised problem statement (see section 2).

Logistic Regression is a popular and robust linear classification method. It's predictor function consists of a transformed linear combination of explanatory variables. In this model a probability of output Y to take the value $k \in \{0, 1\}$ has the form:

$$p(Y = k|X, \theta) = \frac{\exp(\theta_k^T X)}{\sum_j \exp(\theta_j^T X)}.$$

To fit this model we need to optimize the conditional log-likelihood:

$$\ell(\theta, D) = \sum_t \log p(y = y_t|X_t, \theta).$$

Typically some method like conjugate gradients can be used then to maximize log-likelihood.

In case of rare events one class in the history set is represented by very small number of examples and classifier that always predicts non-event will have a good average error equal to the percent of non-events. So in this case we need to perform a kind of set balancing [12].

3.3.1 Parameter Selection

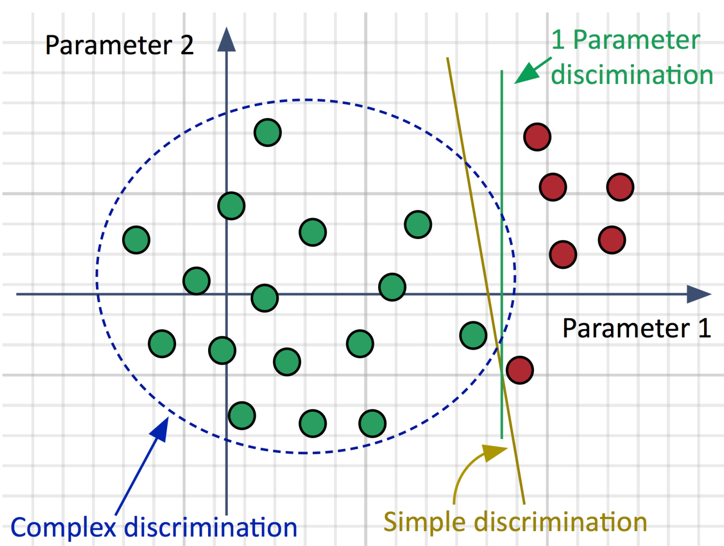


Figure 1: Parameter selection illustration.

Consider an example of events in a two-parameter system described on the figure 1. The red circles represent parameter value combinations corresponding to events while green – to ordinal behavior of the system. In order to discriminate events and ordinal behavior one can construct a complex discrimination curve (or surface in general), like the blue one on the figure, or rather simple linear rule, in yellow color on the figure) which does the same job. One can next notice that impact of the second parameter in the linear rule is small compared to those of the first parameter and actually events could be discriminated by the first parameter only (green line).

This idea lies in the core of parameter selection procedure which is based on regularized Logistic Regression model where likelihood is penalized by additional term:

$$\ell^*(\theta) = \ell(\theta) + \lambda J(\theta),$$

where $J(\theta) = \sum_k \alpha_k \psi(\theta_k), \alpha_k \geq 0,\ \psi(\theta_k) \geq 0$.

Usually the penalty function ψ is chosen to be symmetric and increasing on $[0, +\infty]$. Over-fitting often tends to occur when the fitted model has many feature variables with relatively large weights in magnitude. To prevent this situation we can use weight decay method (or ridge regression). Let $J(\theta) = \sum_k \theta_k^2$ (or ℓ_2). The result of using such a function $J(\theta)$ is classifier with smaller values of weights and often better generalization ability. We can also prune this classifier: weights with magnitudes smaller than some certain threshold can be considered redundant and removed.

However, ℓ_2 usually leaves most of the weights θ_k non-zero that can be a problem when we have a lot of features and want to get sparse θ-vector. It can be also a native characteristic of our data, when we have a large amount of features, but only relatively small number of them is sufficient to learn the target concept. So, let $J(\theta) = \sum_k |\theta_k|$ (or ℓ_1). In this case a lot of weights can become zero thus giving a model with relatively sparse vector θ.

4 REAL WORLD CASES

To prove the proposed methodology we examined two anonymized real-world cases from aircraft operational performance area granted by AIRBUS. The first case is an application of multidimensional degradation trending for the detection and prediction of leakage-like behavior in an aircraft cooling system. Second case is an anticipation of some group of typical failures in an aircraft subsystem that happen with frequency 0.2-1%.

4.1 LEAKAGE DETECTION AND TRENDING

The leakage-like degradation process could be considered as a pathway to the failure event so that maintenance (change of equipment or replenishment of some liquid) should be planned when degradation comes to some threshold level. The main task here is to detect the abnormal behavior of the system (leakage) and then to predict time left before the predefined level is achieved, i.e. plan the maintenance actions.

Based on several examples of normal-behaved aircrafts we trained a one-class SVM model and applied it to the out-of-sample aircraft with (see figure 2a) and without (see figure 2b) degradation. The detection of anomaly (degradation process) was performed on the first differences of the corresponding time series, assuming that in case of degradation the subsequent differences expose unusual behavior.

For the prediction of degradation behavior in time we used ARIMA modeling (see figure 2c) separately for each aircraft. That approach allows to accurately predict number of flights left before the maintenance event is actually needed, i.e. plan the maintenance in advance.

4.2 PREDICTION OF RARE EVENTS BASED ON CLASSIFICATION

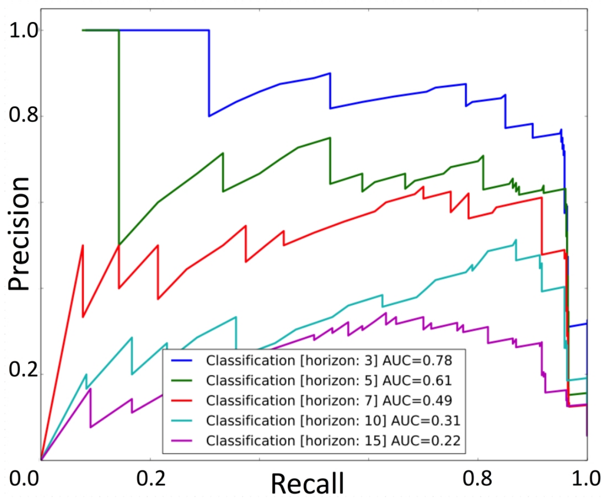


Figure 4: Precision-recall curves for proposed approach and simple parameter thresholding for different horizons of prediction

For the case of prediction of specific failures we used the parameter selection approach based on regularized Logistic Regression model (see section 3.3.1). Subsequently eliminating parameters, we ended up with one parameter only.

The selection of best history lag was done by performance assessment on separate validation set (based on area under corresponding precision-recall curves). The final prediction is done by constructing statistical model based on the selected parameters (see section 3.3). Example of failure prediction along with the selected parameter is shown on figure 3.

To evaluate the predictive ability and performance of the approach we compared it to simple thresholding (the warning of event is raised when the level of curve is below some predefined value) and obtained the precision-recall curves depicted on figure 4 for different horizons of prediction. Our approach significantly outperforms simple thresholding.

5 CONCLUSIONS

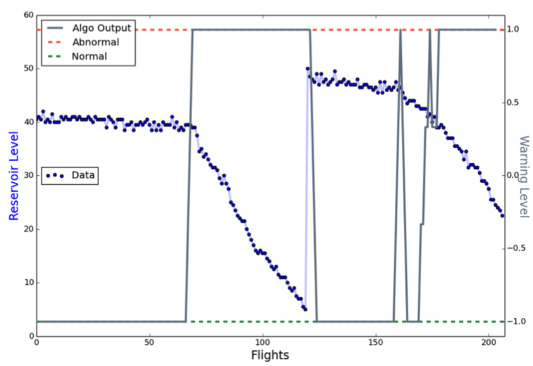
- We have proposed two purely data-driven approaches of fault prediction: 1) degradation detection and trending, and 2) rare event anticipation.

- With proposed approaches we have covered some set of possible aircraft equipment failures and illustrated them by the two real data cases.

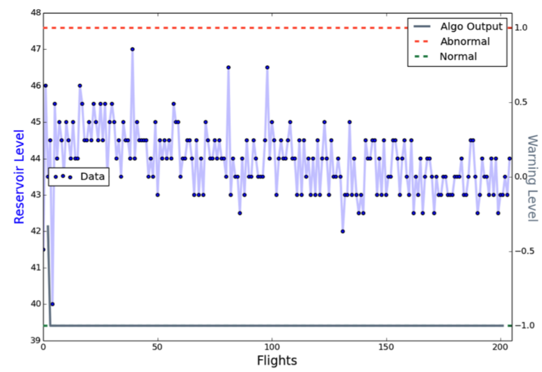
- The project is going on and in the next steps we are going to cover more possible cases of failure events and expand the overall methodology.

REFERENCES

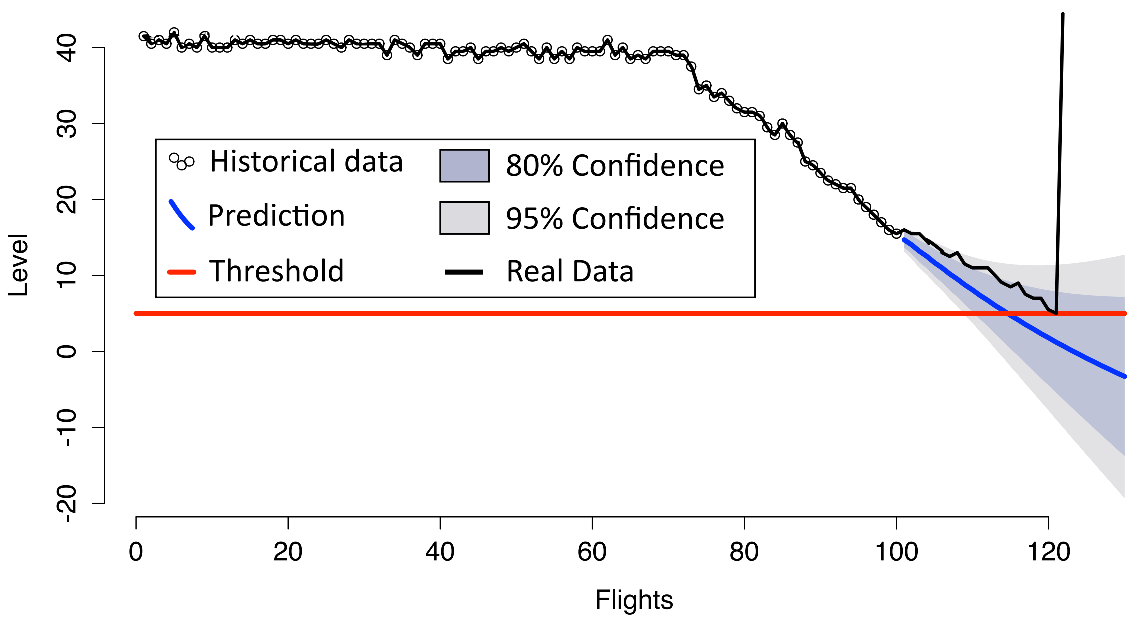
[1] Stascavage J. *Bringing the power of the US Navy to condition-based maintenance.* Marine Maintenance Technology International, (2012).



(a) System with degradation



(b) System without degradation



(c) Trend prediction

Figure 2: Detection of degradation and trend prediction

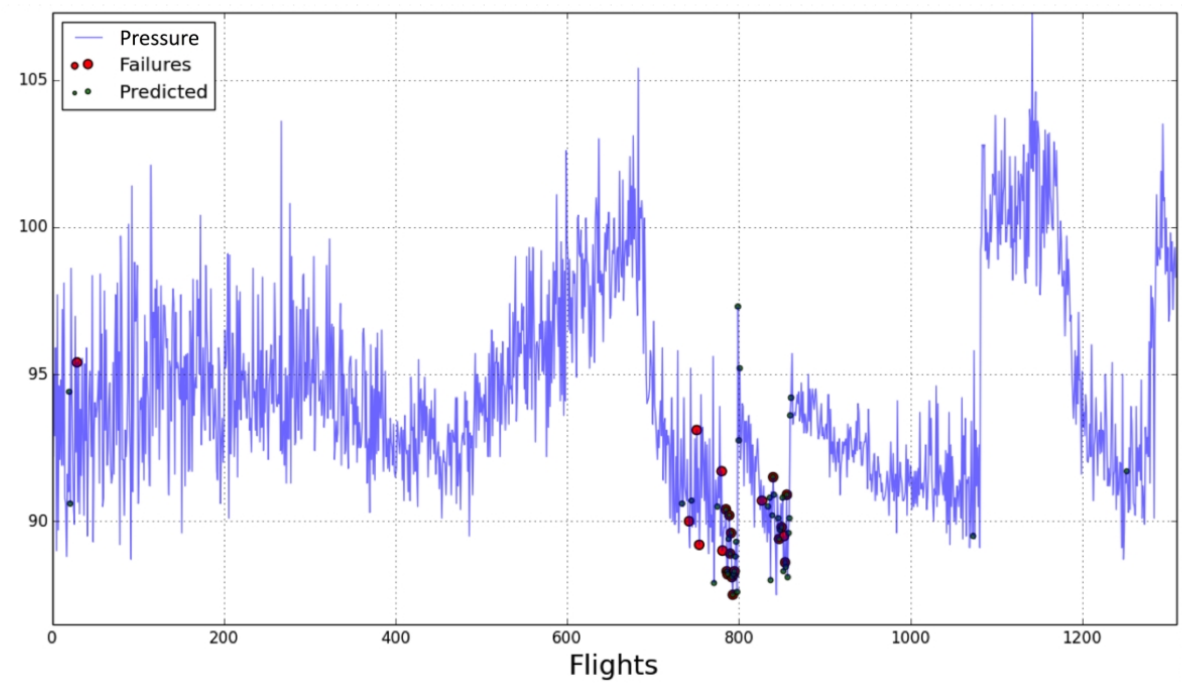


Figure 3: Parameter behavior in time with real and predicted failures

[2] Tremblay M., Pater R., Zavoda F., Valiquette D. and Simard G. *Accurate Fault-Location Technique based on Distributed Power-Quality Measurements*. 19th International Conference on Electricity Distribution, (2007).

[3] Silva P., Negrao M., Junior P.V. and Sanz-Bobi M.A. *A new methodology of fault location for predictive maintenance*. Electrical Power and Energy Systems, V. 42, pp. 568 574, (2012)

[4] Bashir N. and Ahmad H. *A Neural Network Based Method for the Diagnosis of Aging Insulators*. IEEE Symposium on Industrial Electronics and Applications (ISIEA), (2009).

[5] Basseville M. and Nikiforov I.V. *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1993).

[6] Weinberg J., Brown L.D. and Stroud J.R. *Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data*. Journal of the American Statistical Association, (2007).

[7] He H. and Garcia E.A. *Learning from Imbalanced Data*. IEEE Transactions on Knowledge and Data Engineering, vol. 21 no. 9 (2009).

[8] Kuleshov A., Bernstein A. and Yanovich Yu. *Asymptotically optimal method for Manifold estimation problem*. XXIX-th European Meeting of Statisticians, (2013).

[9] Bernstein A., Burnaev E. and Erofeev P. *Manifold Reconstruction in Dimension Reduction Problem*. Proceedings of the 9th International Conference 'Intelligent Information Processing', (2012).

[10] Schoelkopf B., Williamson R.C., Smola A. J., Shawe-Taylor J. and Platt J.C. *Support Vector Method for Novelty Detection*. NIPS, Vol. 12 (1999).

[11] Asteriou D, and Hall S.G. *ARIMA Models and the BoxJenkins Methodology*. Applied Econometrics (Second ed.). Palgrave MacMillan. pp. 265286 (2011).

[12] King G. and Langche Z. *Logistic Regression in Rare Events Data*. Political Analysis 9, pp. 137163 (2001).