

## The application of big data-driven prognostic and health management on complex equipment based on internet of things

Guo-Shun Chen<sup>†</sup> and Gang Niu

*Institute of Ordnance Technology, Shijiazhuang, P. R. China*

<sup>†</sup>*wa\_2003@126.com*

A novel big data-driven Prognostic and Health Management (PHM) method of Unmanned Air Vehicles (UAV) based on IoT has recently been proposed. This method is innovatively using big data of IoT analysis to health conditions, its goal was to build the reference indicator through probability-based techniques, and to generate a degradation model to health monitoring of UAV systems. This method involves trend analysis and regression techniques, using for estimation of the future conditions of IoT systems and forecast of the time-to-failure. The simulation results of the proposed method shows that it could provide the health index of UAV systems visually, and was proved to be conforming to reality and easy to implement in engineering.

**Keywords:** Big data-driven prognostic; Prognostic and Health Management(PHM); complex equipment based on Internet of Things(IoT); big data; equipment health management)

### 1. Introduction

For the past few years, big data-driven Prognostics and Health Monitoring (PHM) of system based on IoT has been focusing on monitoring the lifetime and condition of equipment and systems. By this new technology and application, Great savings in equipment maintenance have been provided, delays in equipment maintenance have been reducing, what's more, operational safety has been improved. The understanding of the current and the prediction of the future health condition of equipment components may provide guidelines to the maintenance activities and spare components logistics [1].

As increasing number of big data from IoT become available, and with the availability of related big data of failure reports and the maintenance actions history, the goal of monitoring the current health state and prediction of the future health condition of equipment components is gradually being achieved. The Using of big data-driven PHM is about to make equipment more powerful

through making them more maintainable, and will probably represent a remarkable competitive advantage before long [2].

To make the decision around the health status of equipment, we must have two types of information: indication of the current state of equipment and reference indicator. The indication of the current condition of equipment is obtained through the failure logs of parameters associated with failure modes, which are the results of the degradation of equipment. The reference indicator can be obtained from a model which deals with the physical limits or a known tolerance threshold of the equipment in use. Normally speaking, such reference value neither exists nor is available to estimate the length of equipment lifetime. Therefore, arbitrary definitions of degradation are produced [3].

- In the case of complicated PHM models are not warranted because of the lower level of criticality or low rates of failure occurrence to diagnose health condition, the only alternative could be providing a big data statistical reliability or developing a usage-based prognostic approach. It is simplistic, though, a statistical reliability-based prognostic distribution can be used to drive interval-based maintenance practices, and that distribution can then be updated on regular intervals [4]. Generally, in order to fit a probability distribution of time-to-failure, error and inspections data are compiled and applied. Weibull distributions are typically employed for this purpose [5].

Providing a big data-based methodology to produce the reference health value and generating a degradation model to be used for condition monitoring on UAV systems based on IoT is this paper's purpose.

## **2. Structure model of UAV**

The UAV is one part of a total system which composed of aerospace vehicle sub-system, control and navigation sub-system, integrated radio sub-system and mission facility sub-system, and this is the important distinction between a UAV and a conventional aircraft. All the sub-systems include some devices and parts, each with specific maintenance requirements[6]. In order to confirm the error propagation path of devices, a topological structure map of UAV system is necessary, and the structure model of UAV system is based on this map, as seen in Fig.1.

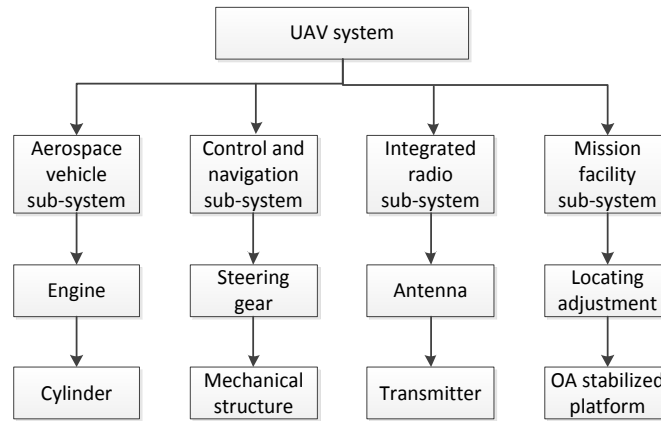


Figure 1. Structure model of UAV

The UAV system are connected each other through IOT to form UAV IoT systems, as seen in Fig.2.

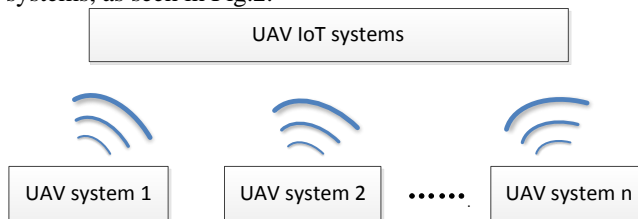


Figure 2. Structure model of UAV IoT systems

As seen above, the PHM of system level mainly acquires the monitoring result of health state of all sorts of equipment and its running tendency of the whole UAV IoT systems, and implements maintenance decision based on the health condition of UAV IoT systems, and then provides maintenance advice to the maintainers.

- A. The PHM of sub-system level conducts relativity analysis with all sorts of condition big data, solves the disparity of those data, and then confirms and isolates errors, and receives more reliable information about health status of sub-systems. The health management of device level plays the role of real time actuator, acquires the sensor information of crucial components of IoT, and implements fusion treatment to detect abnormality, improved diagnose error, assess health status, and predict the life-time of the sub-systems. By reading sensor big data real-time of crucial components and performing feature extraction for those data, the health management of crucial components level realizes security monitoring and error early-warning.

Without list all sub-system failure modes and health assessment methods, this paper focuses only on fuel sub-systems of engine, which is the critical

component of aerospace vehicle sub-system and has a certain reference, other sub-systems can refer to the analysis and the evaluation methods for analysis.

In light of function, the modern UAV fuel system was made of fuel supply sub-system, refueling sub-system, fuel delivering sub-system, the lower fuel delivering controller, lower fuel supply controller and the host computer controller sub-system, and so on. \*

In view of practical operation, common errors of fuel system of UAV based on IoT are generally manifested in two aspects: First, the error of the fuel components, in other words, the error occurred in UAV fuel delivery process, was caused by a mechanical failure due to the operational parts (mainly for pumps, valves, regulators, and sensors). Second, the sub-systems' dysfunction is the result of the sub-system failure (such as controllers), and the sub-systems multiple attachments failure.

### 3. Big data-based Method

To all means attempting to fit a model, regression is a common term, to observed big data, and the goal of the observation is to quantify the relationship between two groups of variables. Regression analysis deals with techniques for the modeling and analysis of numerical data in statistics, it is composed of values of one or more independent variables and of a dependent variable. And then, the model suited may be used either to describe the relationship between the variables, or to predict new values.

From the empirical, the relationship between degradation index  $y$  and the time  $t$  can be expressed as

$$y = \beta_0 + \beta_1 t + \varepsilon \quad (1)$$

in the equation,  $\varepsilon$  is the model residue,  $\beta_0$  is the linear coefficient, and  $\beta_1$  is the angular coefficient. The residue is unknown, but for each  $y$  observation, it is different. The coefficients  $\beta_0$  and  $\beta_1$  are also unknown, but they are constants. Estimates of  $\beta_0$  and  $\beta_1$  can be obtained by least-squares. Normally, as assumed in Eq. 1, the long-term evolution of degradation with time will not be linear. Although simplified, this model may be adequate for medium and short-time predictions, which may also be enough for CBM purposes.

To estimate, the following notation will be applied:

$$\begin{aligned} b_0 &= \widehat{\beta_0} \\ b_1 &= \widehat{\beta_1} \\ \hat{y} &= b_0 + b_1 t \end{aligned} \quad (2)$$

in the equations,  $\hat{y}$  refers to the predicted value of  $y$  for a given  $t$ . Thus, given a

failure threshold  $L$  for the degradation index, the prediction of the time of failure  $\hat{t}_f$  can be expressed as

$$\hat{t}_f = \frac{L - b_0}{b_1} \quad (3)$$

In the case of the coefficients  $b_0$  and  $b_1$  have been obtained, it is important to analyze whether the data's observed linear trend is statistically significant. More on this coming up.

The significance of regression can be evaluated by an analysis of variance (ANOVA). Such analysis is in light of the following identity:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

where:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (5)$$

And  $n$  stands for the number of available observations. The terms in Eq. 4 can be dealt with in standard ANOVA format, as seen in Table 1.

TABLE I. TERMS INVOLVED IN ANALYSIS OF VARIANCE FOR REGRESSION

Source of variation	Degrees of freedom	Sum of Squares (SS)	Mean Square (MS)
Due to regression	1	$SS_{Reg} = \sum_{i=1}^n (y_i - \bar{y})^2$	$MS_{Reg}$
About regression (residual)	$n-2$	$RSS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$S^2$
Total	$n-1$	$SYY = \sum_{i=1}^n (y_i - \bar{y})^2$	$Sy^2$

By comparing  $MS_{Reg}$  with  $S^2$ , the significance of the regression can be assessed. In line with the general assumptions for linear regression, it can be expressed that the ratio

$$F = \frac{MS_{Reg}}{s^2} \quad (6)$$

Follow with an F-distribution with 1 and  $(n-2)$  degrees of freedom if  $\beta_1=0$ . So, for the null hypothesis  $H_0: \beta_1 = 0$  (no linear trend), versus  $H_1: \beta_1 \neq 0$ , this fact stated above can be employed as a test Comparing the ratio (6) with the  $100(1-\alpha)\%$  point of the tabulated  $F(1, n-2)$  distribution, we can determine whether  $\beta_1$  can be thought as nonzero. Moreover, a p-value can be gotten as the area beneath the tail of the distribution  $F(1, n-2)$  for  $F > F_{calculated}$ . The p-value can be considered as the chance that the observed trend is fortuitous.

According to the general opinion for linear regression, the estimation of  $b_0$  follows a normal distribution with mean 0 and variance given by

$$\sigma^2 \left[ \frac{1}{n} + \frac{\bar{t}^2}{S_{TT}} \right] \quad (7)$$

In this equation,  $\sigma^2$  is the variance of the residue  $\varepsilon$ ,  $\bar{t}$  is the mean value of  $t$ . The mean and variance of estimated  $b_0$  and  $b_1$  are given by:

$$b_0 \sim N(b_0, \sigma^2 \left[ \frac{1}{n} + \frac{\bar{t}^2}{S_{TT}} \right]) \quad (8)$$

$$b_1 \sim N(b_1, \frac{\sigma^2}{S_{TT}}) \quad (9)$$

$S_{TT}$  is:

$$S_{TT} = \sum_{i=1}^n (t_i - \bar{t})^2 \quad (10)$$

Eq. 7 tells us, through the reduction of the noise level, the increase of the number of the observing values and the increase of the timespan of the historical record, we can substantially improve the accuracy of the estimations. We can also calculate the covariance value for  $b_0$  and  $b_1$ . Because the value of  $\sigma^2$  is not known in real operation, it can be replaced with the standard error (s) expressed as

$$s = \sqrt{\frac{RSS}{n-2}} \quad (11)$$

In the equation, we have to use a Student-t distribution for  $b_0$  and  $b_1$ . However, for large values of  $n$ , the Student-t distribution may be roughly calculated through a normal distribution.

To get a probability distribution pointily referring to the time of failure  $t_f$ , we can use a Monte Carlo simulation method, which is a sort of algorithms based on random sampling over and over again. As stated above, we may use a multivariate random number generator to acquire the distributed random values of  $b_0$  and  $b_1$ . Also, the random number generator can be used to accomplish different realizations of the residue  $\varepsilon$  over a series of instants of future time. For each realization of the residue and each pair  $(b_0, b_1)$ , we can get a succession of future degradation values. In this way, the failure time can be defined as the earliest instant, because the failure threshold  $L$  is smaller than the degradation.

Through this kind of simulation over and over again, we can get a probability distribution of the forecasted failure time, and confirm that this probability distribution is Normal or Weibull, for instance, and the confidence interval can also be calculated.

#### 4. Simulation Analysis

To illustrate the technique described above, a case study including actual big data from a UAV fuel system based on IoT will be performed. From the features related to measurements of fuel content in equipment, the degradation index can be obtained. Under the circumstances, the degradation means the fuel consumption. Such data will be obtained through applicable forecast, and would be in form of a value, using to schedule a maintenance intervention conveniently (replenishment of the reservoir). The data is obtained over time, and is composed of a historical record of 250,000,000 values for degradation index.

To provide the results, an algorithm was used. Simply speaking, it calculates firstly values for  $b_0$  and  $b_1$ , using least square linear regression method, as seen in Eq. 2. Next, values for  $MS_{\text{Reg}}$  and  $s^2$  are calculated, see Table 1. An F-test was initially implemented to evaluate the significance of the linear regression. The essential value of the F distribution, for  $\alpha = 0.05$ , was calculated as  $F_{\text{crit}} = 3.9$ . The F-value calculated by using Eq. 5 was 1727, which is greatly larger than  $F_{\text{crit}}$ . So, we can use  $100(1-\alpha)\% = 95\%$  of confidence to reject the null hypothesis (there is no linear tendency). The related p-value was very close to zero, which means that one can be quite sure that the trend observed in the data record is not fortuitous.

Finally, the algorithm uses Eq. 8, 9 and 10, with the purpose to obtain the probability distribution of the estimated coefficients. Moreover, the dysfunction time predictions are simulated by Monte Carlo technique. A failure time is acquired from a linear extrapolation with noise for each interaction. We also provide a vector with the simulated failure times.

We can use the proposed technique to calculate the probability densities, for instance, and to calculate the probability of dysfunction within a period of time in the future. And we can utilize such result to promote a cost-benefit analysis, for the purpose of determining maintenance recommendations.

#### 5. Conclusion

Brief discussion of big data-driven PHM and the UAV structures based on IoT was the content of this paper, and it was described more fully theoretical and notional foundations, those foundations are correlated with regression-based forecasts. More to the point, a regression-based technique was presented and illustrated by a simulation study which simulated with real data obtained from a UAV IoT system. By using a Monte Carlo technique, the failure instant was calculated and the confidence related with this estimation was assessed.

Enhancing the armament operational readiness and maintenance support efficiency was the ultimate goal, and the paper's focus was on the decreasing of

the maintenance cost, and on realizing of the information-based autonomous logistics support of UAV IoT systems.

## References

- [1] Yin, S., Ding, S. X., Haghani, A., Hao, H., & Zhang, P. (2012). A comparison study of basic data-driven error diagnosis and process monitoring methods on the benchmark tennessee eastman process. *Journal of Process Control*, 22(9), 1567–1581.
- [2] Lee, S., Park, W., & Jung, S. (2014). Error detection of aircraft system with random forest algorithm and similarity measure. *The Scientific World Journal*, 2014.
- [3] JR Celaya, CS Kulkarni, G Biswas. 2011. “A Model-based Prognostics Methodology for Electrolytic CapacitorsBased on Electrical Overstress Accelerated Aging”, *Conference of the Prognostics & Health Management Society*, v. 27, 193-198.
- [4] Roemer, M. J., Byington, C. S., Kacprzynski, G. J., Vachtsevanos, G. “An Overview of Selected Prognostic Technologies with Reference to an Integrated PHM Architecture”. In *Proc. of the First Intl. Forum on Integrated System Health Engineering and Management in Aerospace*.
- [5] Groer, P.G. 2000. “Analyses of time-to-failure with a weibull model”. In: *MAINTENANCE AND RELIABILITY CONFERENCE (MARCON)*, 2000, Knoxville. *Proceedings* [S. l: s. n.].
- [6] JB Coble, P Ramuhalli, LJ Bond. *Prognostics and Health Management in Nuclear Power Plants: A Review of Technologies and Applications*. Office of Scientific & Technical Information Technical Reports, 2012, 81-88.