How to Use r and LaTeX to Write a Dynamic Report

An Enthusiastic of $\iiint_{Birth}^{Life} Learning\,dy\,dm\,dt$

August 5, 2015

Contents

1 TF34-100 Engine Sensor Locations

The locations of certain engine sensors are shown in Figure [1].

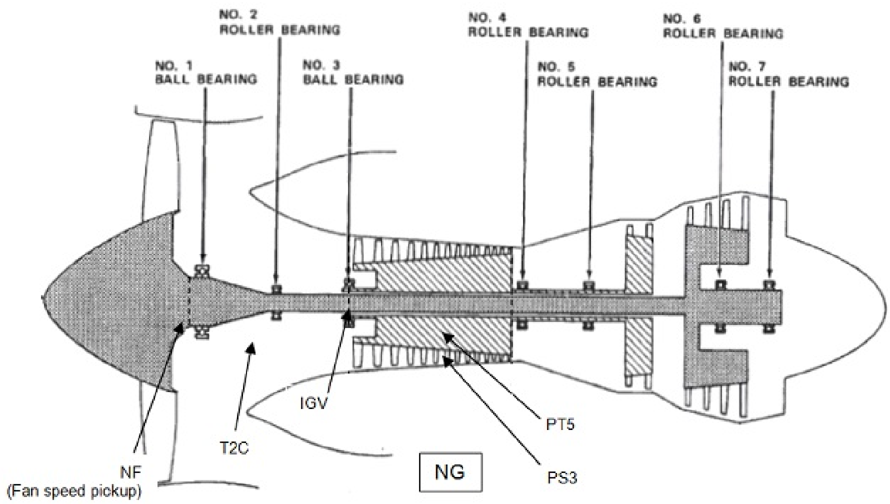

Figure 1: Locations illustrated for three interested sensors (T2C, NG and IGV)

2 Type of Engine Compressor Stall Fault

See Table [1].

Table 1: Type of compressor stalls

	WESA	AoA in Envelope
PLA	>= 12°	>= 12°
PS3(% drop over 500ms)	>25 %	>50%
NF	7%	
AoA	in envelope for mach number	in envelope for mach number

3 Data Definition

Reference [3], Reference [2] and Reference [1]

Following is the definition for the data used in this Dynamic Report.

```
> seekData <- function (  workDir = "C://simplifiedthesis/data",
+                         Num = 4,   ## number of flights combined
+                         K = 3      ## the Kth data ie. the Kth Aircraft Series Number
+ ) {
+ pattern <- paste("comb_", Num, sep = "")
+ filez <- list.files(path = workDir, pattern = pattern, full.names = TRUE)
+ faultStatus <- substr(filez[K], 44, 51)
+ Dat <- read.csv(filez[K])
+ Dat <- Dat[which(Dat[,2] < 9.999 & Dat[,2] > -9.999), ]
+ Dat <- ts(Dat[,2])
+ return(Dat)
+ }
```

Table 2: Defination code for data used for this report.

4 Plots of ts, diff, acf and pacf

Fllowing 2 figures (Figure 2 and Figure 3) demostrate how the R codes would work here.

```
[1] "For 7-Combined-Flights data; Aircraft 3th in ASN."
```

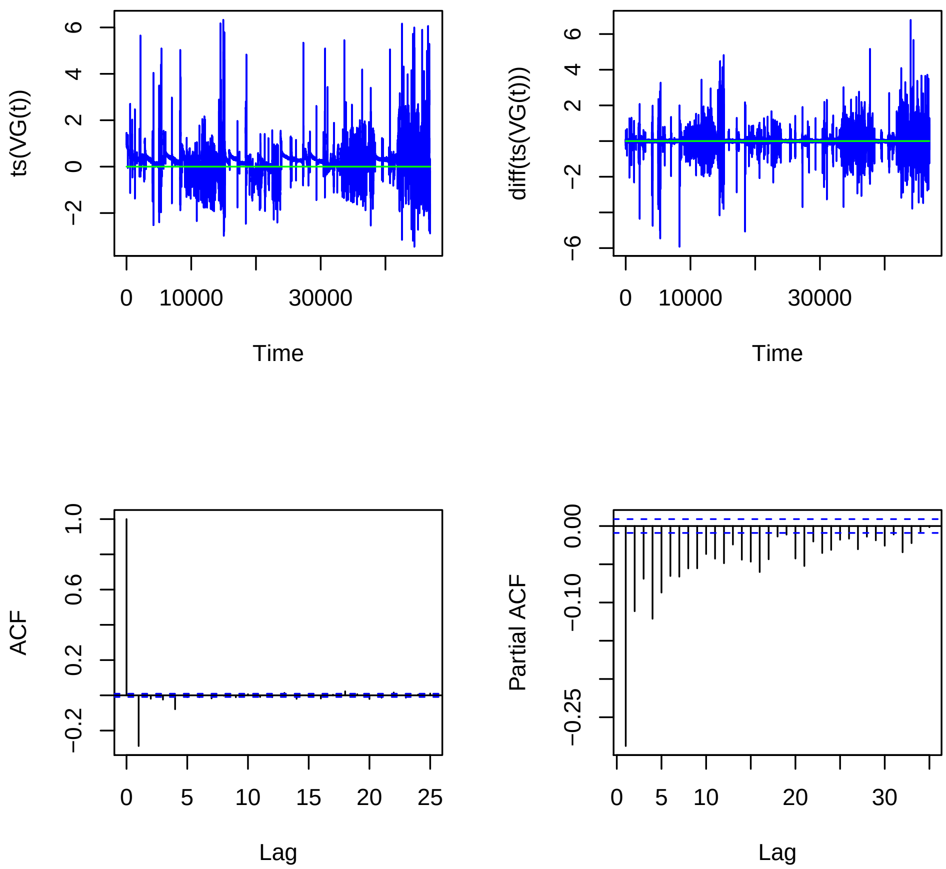

Figure 2: The ts(VG(t)), diff(ts), acf(diff(ts) and pacf(diff(ts))).

```
[1] "For 7-Combined-Flights data; Aircraft 18th in ASN."
```

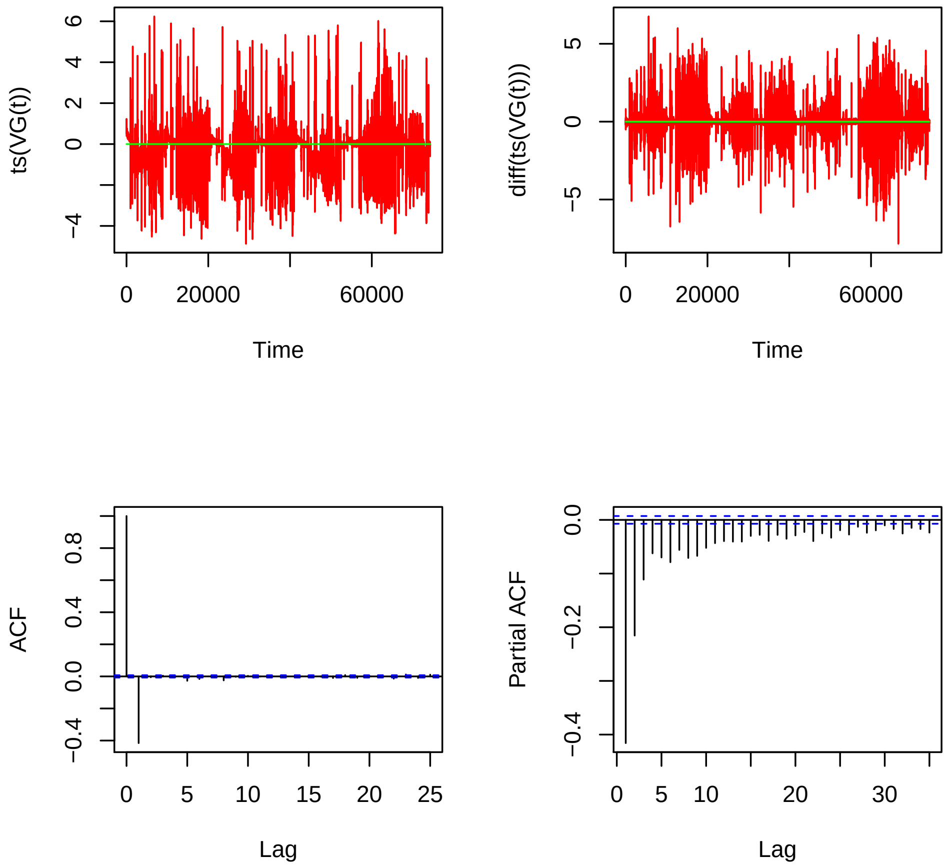

Figure 3: The ts(VG(t)), diff(ts), acf(diff(ts) and pacf(diff(ts))).

5 Mathmatical Equations

$$X_t = \delta + AR_1 X_{t-1} + AR_2 X_{t-2} + \cdots + AR_p X_{t-p} + A_t - MA_1 A_{t-1} - MA_2 A_{t-2} - \cdots - MA_q A_{t-q} \quad (1)$$

Where $X_t = X(t) = diff(ts(VG(t)))$ is the first order differenced value of the ts transformed VG data at the time , $AR_i$is the AutoRegression (AR) coefficient, and MA_j is the Moving Average (MA) coefficient. This model is denoted as $arima(p, 0, q)$.

$$p(CompressorStall|N_{CombinedFlights}) = \beta_0 + \sum_{i=1}^{p} \beta_i * AR_i + \sum_{j=1}^{q} \beta_{j+p} * MA_j + \epsilon) \tag{2}$$

Where Nindicates number of flights data used, currently, $N = 1, 2, ..., 7$. AR_i is the AR coefficients ($i = 1, 2, ..., p$) and MA_j is the MA coefficients ($j = 1, 2, ..., q$). This Probability Model of Compressor Stall Fault Event is based on the data as described below:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \\ y_{m+1} \\ \vdots \\ y_n \end{pmatrix} \sim \begin{pmatrix} AR_{1_1} & AR_{2_1} & \dots & AR_{p_1} & MA_{1_1} & MA_{2_1} & \dots & MA_{q_1} \\ AR_{1_2} & AR_{2_2} & \dots & AR_{p_2} & MA_{1_2} & MA_{2_2} & \dots & MA_{q_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ AR_{1_n} & AR_{2_n} & \dots & AR_{p_n} & MA_{1_n} & MA_{2_n} & \dots & MA_{q_n} \end{pmatrix} \tag{3}$$

Where the response variable (independent variable) $(y_1, y_2, ..., y_m)$ indicates compressor stall fault event DID NOT happen for engine 1 to engine m, and $(y_{m+1}, ..., y_n)$ indicates compressor stall fault event DID happen for engine (m+1) to engine n; $AR_{i,k}$ $(i = 1, 2, ..., p; k = 1, 2, ..., n)$ and $MA_{j,k}$ $(j = 1, 2, ..., q; k = 1, 2, ..., n)$ are the regressor variables (explanatory variables) which are AutoRegression (AR) coefficients and Moving Average (MA) coefficients of the fitted $arima(p, 0, q)$ model from each set of first-order differenced VG data. Currently m = 22 and n = 28. This reflects that the data consists of 14 aircraft (equipped with a total of 28 engines), and that 6 of the engines have compressor stall fault events.

VG time series data is generated by combining flights 1 (one) through 7 (seven). Following equations have been used to explain how the N-Combined-Flight Data constructured.

For example, for 1-Flight Data:

$$VG(t)_{aircraft_i,1Flight} = f(T2C(t)_{aircraft_i,flight_j} + NG(t)_{aircraft_i,flight_j} + IGV(t)_{aircraft_i,flight_j}) \tag{4}$$

Where this is aircraft i, the j is the highest flight number in our selection of this aircraft.

For 2-Combined-Flights Data:

$$VG(t)_{aircraft_i,2CombinedFlights} = VG(t)_{aircraft_i,1Flight_j} + VG(t)_{aircraft_i,1Flight_{(j-1)}} \tag{5}$$

Where this is aircraft i, the $(j-1)$ is the second highest flight number in our selection of this aircraft.

For N-Combined-Flight Data: $(N = 2, 3, ..., 7)$

$$\begin{aligned} VG(t)_{aircraft_i, NCombinedFlights} =& VG(t)_{aircraft_i, 1Flight_j} \\ &+ VG(t)_{aircraft_i, 1Flight_{(j-1)}} + ... + VG(t)_{aircraft_i, 1Flight_{(j-N+1)}} \end{aligned} \tag{6}$$

Where this is aircraft i, the $(j-1)$ is the second highest flight number in our selection of data for this aircraft, the $(j-N+1)$ is the $(N-1)^{th}$ lower that the highest flight number in our selection of data for this aircraft. The first order differenced VG time series data was constructed as following equations shown.

$$\begin{aligned} diff\{VG(t)_{aircraft_i, NCombinedFlights}\} =& VG(t)_{aircraft_i, NCombinedFlights} \\ &- VG(t-1)_{aircraft_i, NCombinedFlights} \end{aligned} \tag{7}$$

Thereafter, the constructed first order differenced VG time series data was used to develop an $arima(p, 0, q)$ model in R. This process can be mathmatically expressed as

$$X(t) = diff\{VG(t)_{aircraft_i, NCombinedFlights}\} \tag{8}$$

$$x.fit < -arima(X(t), order = c(p, 0, q), optim.method = "Nelder - Mead") \tag{9}$$

$$AR_i = x.fit\$coef\,[i]\ where i = 1, 2, ..., p \tag{10}$$

$$MA_j = x.fit\$coef\,[j + p]\ where j = 1, 2, ..., q \tag{11}$$

6 Four ARIMA Performance Parameters

6.1 Sigma2

A first order difference was formed from the VG data files for the sampled 28 engines, including the 6 engines with compressor stall faults using the flight data immediately prior to the fault flight for that particular aircraft. The aircraft serial number (ASN) and engine ID were checked for consistency to ensure that this was the case. Then 1 to 7 combined-flights-data were formed for the specified ASN aircraft. The reason to form a multiple flight-combined- data is to attempt to achieve an "early alert".

An R script was developed to check a specified ARIMA(p,0,q) model's performance based on 4 (four) parameters.

First, check the sigma2 parameter by using the R code:

$$sigma2 = arima(x, order = c(p, 0, q), optim.method = "Nelder - Mead")\$sigma2 \tag{12}$$

Where "sigma2" stands for the maximum likelihood estimate (MLE) of the innovations variance. The difference between the expected mean at time t, given the time series prior to t, and the actual value is called the innovation. Measuring the variance of the innovation will give you a better idea of how "noisy" the process is.

6.2 Log-Likelihood

Second, check the Log-Likelihood parameter by using R code:

$$Log - Likelihood = arima(x, order = c(p, 0, q), optim.method = "Nelder - Mead")\$loglik \tag{13}$$

Where "Log-Likelihood" stands for a logarithm of likelihood function. In statistics, a Likelihoodfunction (often simply the Likelihood) is a function of the parameters of a statistical model. The likelihood of a set of parameter values, θ, given outcomes x, is equal to the probability of those observed outcomes given those parameter values, that is $\mathcal{L}(\theta|x) = \mathcal{P}(x|\theta)$.

6.3 AIC

Third, check the AIC parameter by using R code:

$$AIC = arima(x, order = c(p, 0, q), optim.method = "Nelder - Mead")\$aic \tag{14}$$

Where "AIC" stands for Akaike Information Criterion, which is a measure of the relative quality of a statistical model for a given set of data.

6.4 Percentage of Significant Coefficients (PSC)

Fourth, check the Percentage of Significant Coefficients (PSC) parameter. PSC represents the percentage of significant coefficients among the AR coefficients and MA coefficients of the fitted arima(p,0,q) model. Let x be the first order differenced ts data of VG, $x.fit$ be the fitted arima(p,0,q) model. That is $x.fit < -arima(X(t), order = c(p, 0, q), optim.method = "Nelder - Mead")$.

If one of the coefficients $x.fit\$coef\,[k]\,(k = 1, 2, ..., (p + q))$ of the fitted $arima(p, 0, q)$ model meets following condition, it is considered as significant.

$$\left|\frac{x.fit\$coef\,[k]}{x.fit\$var.coef\,[k]}\right| > 1.96 \tag{15}$$

In R programming environment, $x.fit\$coef\,[k]$ is the coefficients of the resulting $arima(p, 0, q)$ model, where $x.fit\$coef\,[k]$ be AR coefficients while $k = 1, 2, ..., p$, and $x.fit\$coef\,[k]$ be MA coefficients while $k = (p + 1), (p + 2), ..., (p + q)$. Correspondingly,$x.fit\$var.coef\,[k]$ be the estimated variances of coefficient . Therefore, the PSC is calculated as:

$$PSC = \frac{Sum\,of\,the\,Number\,of\,Significant\,x.fit\$coef\,[k]}{(p + q)} \tag{16}$$

The following is a demonstration of the parameter's performance for the various ARIMA models and the different sampling data:

For example, process the sample data as of Figure 4, is to fit to an $arima(12, 0, 4)$ model for the first order differenced values of ts transformed VG data. By using of an R script, the coefficients of $arima(12, 0, 4)$ is calculated as shown as Table 3.

(-0.1815, 0.0002387, 0.01500, 0.03247, 0.02954, -0.002241, 0.05759, 0.04050, 0.006447, 0.01128, 0.01244, 0.02152, -0.05026, 0.04650, 0.02096, -0.01320)

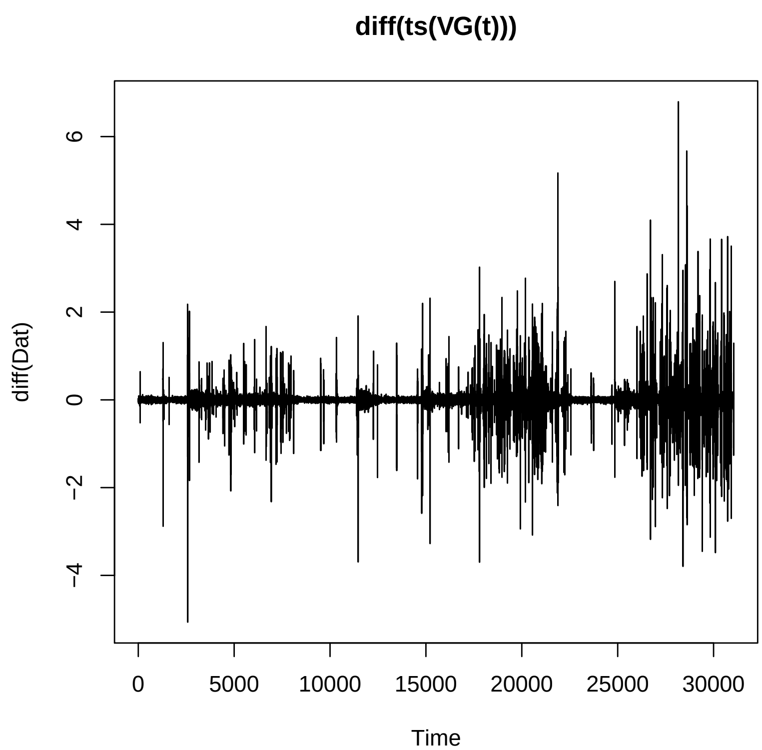

Figure 4: First order differenced values of ts transformed VG data for a 4-Combined-Flights.

```
[1] "For 4-Combined-Flights data; Aircraft 3th in ASN. p=12 & q=4 in arima(p,0,q) model."
```

ar1	ar2	ar3	ar4	ar5	ar6
-0.181470271	0.000238739	0.015003603	0.032469898	0.029538878	-0.002240709
ar7	ar8	ar9	ar10	ar11	ar12
0.057590585	0.040496339	0.006446809	0.011284292	0.012444539	0.021522452
ma1	ma2	ma3	ma4		
-0.050258616	0.046497112	0.020957831	-0.013201263		

Table 3: Coefficients of $arima(12, 0, 4)$ for a 4-Combined-Flights data.

Where the first 12 coefficients are called AR coefficients, and the last 4 coefficients are called MA coefficients, because $p = 12$ and $q = 4$ in an $arima(p, 0, q)$ model here.

And the variance of the coefficients of the $arima(12, 0, 4)$ is calculated as Table 4

```
[1] "For 4-Combined-Flights data; Aircraft 3th in ASN. p=12 & q=4 in arima(p,0,q) model."

 [1]  3.71987e-04 -2.02777e-04 -2.36069e-05 -2.78047e-04  1.43877e-05
 [6]  3.66777e-05  5.50166e-05  5.02489e-05  5.39051e-05  6.57346e-05
[11]  3.56272e-05  5.36436e-05 -3.38005e-04  3.33822e-04 -1.38888e-05
[16]  3.17574e-04
```

Table 4: Variances of coefficient of $arima(12, 0, 4)$ for a 4-Combined-Flights data.

(0.0003720, -0.0002028, -0.00002361, -0.0002780, 0.00001439, 0.00003668, 0.00005502, 0.00005025, 0.00005391, 0.00006573, 0.00003563, 0.00005364, -0.0003380, 0.0003338, -0.00001389, 0.0003176)

Thus, the ratio of coefficient to variance of coefficient $\frac{x.fit\$coef[k]}{x.fit\$var.coef[k]}$ can be calculated as Table 5.

```
[1] "For 4-Combined-Flights data; Aircraft 3th in ASN. p=12 & q=4 in arima(p,0,q) model."

       ar1         ar2         ar3         ar4         ar5         ar6         ar7
 487.84053     1.17735   635.56088   116.77849  2053.07125    61.09184  1046.78616
       ar8         ar9        ar10        ar11        ar12         ma1         ma2
 805.91444   119.59551   171.66453   349.29842   401.21157   148.69178   139.28730
       ma3         ma4
1508.97839    41.56904
```

Table 5: The absolute ratio of \$coef to \$var.coef of $arima(12, 0, 4)$ for a 4-Combined-Flights data.

(487.84, 1.18, 635.56, 116.78, 2053.07, 61.09, 1046.79, 805.91, 119.60, 171.66, 349.30, 401.21, 148.69, 139.29, 1508.98, 41.57)

The second AR coefficient (AR2) is not significant because its ratio of coefficient to variance of coefficient valued as of 1.18. The rest of the coefficients of arima(12,0,4) model are significant. Therefore the PSC can be calculated as

$$PSC = \frac{Sum\,of\,the\,Number\,of\,Significant\,x.fit\$coef\,[k]}{(p+q)} = \frac{15}{16} = 0.9375 = 93.75\%$$

Lower than 100% indicating some coefficients are not significant. It is estimated that compounding factors and non-flight-state-separated data play roles to make some of the arima model coefficients non-significant.

7 Linear Regression Model Adequacy

$fit.lm <- lm(Fault \ ., data = DATA, na.action = NULL)$

$R^2 = summary(fit.lm)\$r.squared$

$R^2_{adj} = summary(fit.lm)\$adj.r.squared$

$StandardError = summary(fit.lm)\$sigma$

$Model_{p-value} = pf(x[1], x[2], x[3], lower.tail = FALSE)$

Where $x < -summary(fit.lm)\$fstatistic$, $pfstats$ is the built-in F probability distribution function in R programming.

The examination of Linear Regression Model (LRM) adequacy based on various arima(p,0,q) are shown in Figure [5], Figure [6] and Figure [7].

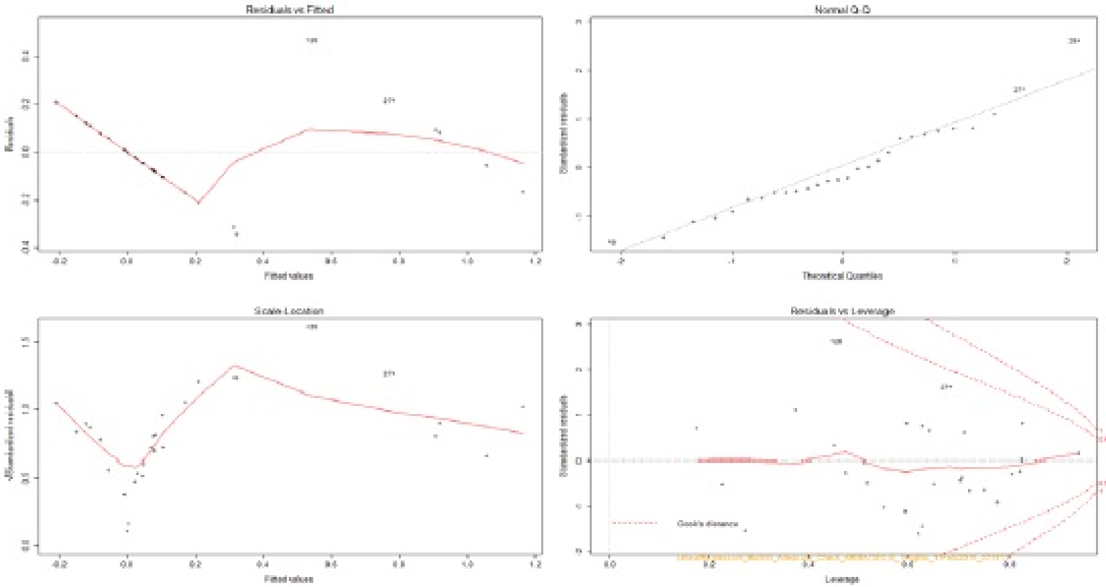

Figure 38: Check Linear Regression Model (LRM) adequacy with R built-in function $plot(lm\{stats\})$. This is an example for one-flight-data with arima(12,0,4).

arima_order	r_squared	adj_r_squared	standard_error	model_p-value
arima(12,2)	0.3528	-0.3443	0.4845	0.8902
arima(12,4)	**0.8656**	**0.67**	**0.24**	**0.008221**
arima(12,6)	0.6282	-0.1154	0.4413	0.6383
arima(14,2)	0.7846	0.4712	0.3039	0.06376
arima(14,4)	0.6257	-0.1229	0.4428	0.6452
arima(14,6)	0.5488	-0.7405	0.5513	0.9366
arima(16,2)	0.8493	0.548	0.2809	0.05775
arima(16,4)	0.7783	0.145	0.3864	0.4134
arima(16,6)	0.6217	-1.043	0.5972	0.951

Figure 39: The examination of Linear Regression Model (LRM) adequacy based on various arima(p,0,q) are shown here. For the one-flight-data, the "best fit" is arima(12,0,4).

Figure 5: Linear Regression Model Adequacy Check 1)

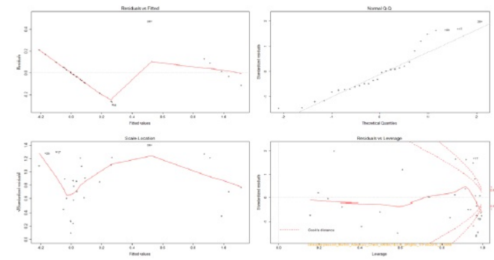

Figure 44: Check Linear Regression Model (LRM) adequacy with R built-in function *plot(lm{stats})*. This is an example for four-combined-flight-data with arima(14,0,6).

arima_order	r_squared	adj_r_squared	standard_error	model_p-value
arima(12,2)	0.6894	0.3549	0.3356	0.1007
arima(12,4)	0.3939	-0.4876	0.5096	0.9303
arima(12,6)	0.4542	-0.6375	0.5347	0.9459
arima(14,2)	0.4902	-0.2513	0.4674	0.7807
arima(14,4)	0.6323	-0.1031	0.4389	0.6268
arima(14,6)	**0.8826**	**0.5473**	**0.2812**	**0.09674**
arima(16,2)	0.3989	-0.8034	0.5611	0.9778
arima(16,4)	0.5448	-0.7556	0.5537	0.9399
arima(16,6)	0.69	-0.6738	0.5406	0.8778

Figure 45: The examination of Linear Regression Model (LRM) adequacy based on various arima(p,0,q) are shown here. For the four-combined-flight-data, the "best fit" is arima(14,0,6).

Figure 6: Linear Regression Model Adequacy Check 4)

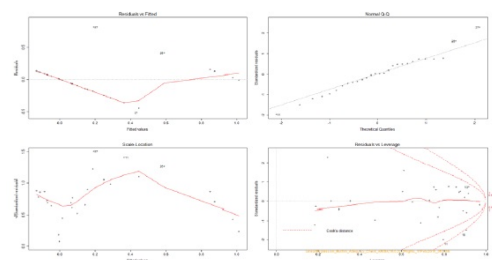

Figure 50: Check Linear Regression Model (LRM) adequacy with R built-in function *plot(lm{stats})*. This is an example for seven-combined-flight-data with arima(16,0,2).

arima_order	r_squared	adj_r_squared	standard_error	model_p-value
arima(12,0,2)	0.4666	-0.1077	0.4398	0.6485
arima(12,0,4)	0.4723	-0.2952	0.4755	0.8169
arima(12,0,6)	0.6658	-0.002485	0.4184	0.5284
arima(14,0,2)	0.2671	-0.799	0.5605	0.9937
arima(14,0,4)	0.5095	-0.4715	0.5069	0.8868
arima(14,0,6)	0.6191	-0.4694	0.5065	0.8479
arima(16,0,2)	0.6808	0.04247	0.4089	0.4824
arima(16,0,4)	0.4908	-0.9641	0.5856	0.9735
arima(16,0,6)	0.676	-0.7495	0.5527	0.8974

Figure 51: The examination of Linear Regression Model (LRM) adequacy based on various arima(p,0,q) are shown here. For the seven-combined-flight-data, the "best fit" is arima(16,0,2).

Figure 7: Linear Regression Model Adequacy Check 7)

8 Fit Goodness of ARIMA-LRM

The visualization for the goodness of fit for ARIMA-LRM Method is shown in Figure [8], Figure [9] and Figure [10]

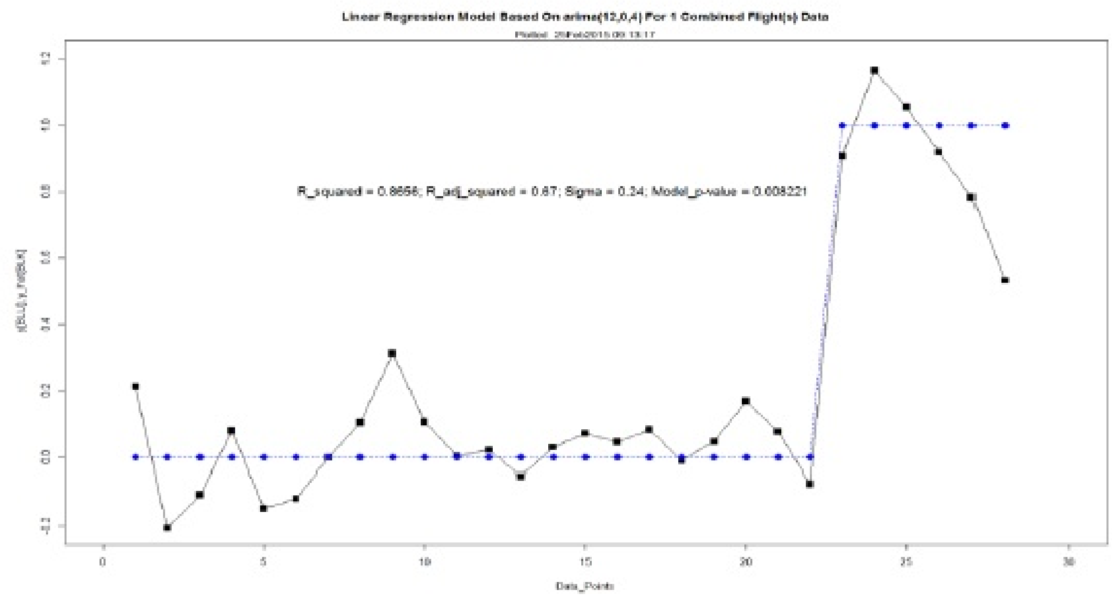

Figure 63: Check the goodness of Fit of proposed Probability Model for Compressor Stall Fault Event. This plot is a Linear Regression Model (LRM) based on arima(12,0,4) for one-Flight-Data.

Figure 8: The visualization for the goodness of fit for ARIMA-LRM Method 1)

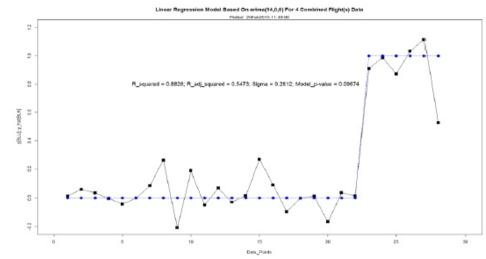

Figure 64: Check the goodness of Fit of proposed Probability Model for Compressor Stall Fault Event. This plot is a Linear Regression Model (LRM) based on arima(14,0,6) for four-Combined-Flight-Data.

Figure 9: The visualization for the goodness of fit for ARIMA-LRM Method 4)

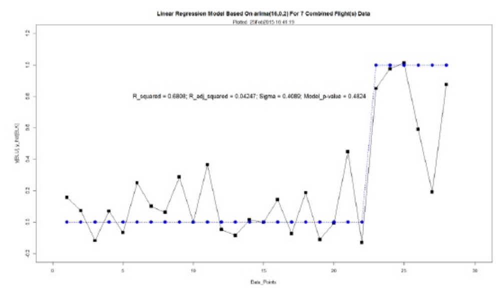

Figure 65: Check the goodness of Fit of proposed Probability Model for Compressor Stall Fault Event. This plot is a Linear Regression Model (LRM) based on arima(16,0,2) for seven-Combined-Flight-Data.

Figure 10: The visualization for the goodness of fit for ARIMA-LRM Method 7)

List of Figures

List of Tables

References

[1] S. Ganguli, S. Deo, and D. Gorinevsky. "Parametric Fault Modeling and Diagnostics of a Turbofan Engine". In: *IEEE CCA/ISIC/CACSD*. Taipei, Taiwan, 2004.

[2] Luca Marinai, Rita Singh , Barry Curnock, and Douglas Probert. "Detection and Prediction of the Performance Deterioration of a Turbofan Engine". In: *Proceedings of the International Gas Turbine Congress*. Tokyo, Japan, 2003.

[3] Field Engineering Unit/H564 Northrop Electronics Division. *Field Engineering A-10 TEMS Software Handbook, A-10 Turbine Engine Monitor System, for Version 54.1 and 56.1 software, section 3.1.4 VG Schedule*. 2010.