Chapter-4

Data driven Methodology-Mathematical Background and Associated Computation

4.1 Introduction

To survive in the competitive market, optimization not only on product performance and price, but on product life must be addressed. The present chapter addresses the two associated data-driven methodologies for motor health prognosis namely:

a) Stochastic Approach - Exponential Degradation Model
b) Adaptive filtering Techniques - Particle filter, Unscented Particle filter and Improved Unscented Particle filter

Review of data driven approaches as presented in chapter 1 evidenced that the generated data from the experiment in the previous chapter 3 is highly nonlinear and temporal. The ability to handle the nonlinearity associated with the real-time data obtained during continuous monitoring and thereby the RUL prediction necessitates the use of the two approaches. The Fig. 4.1 provides a schematic from data collection to model development.

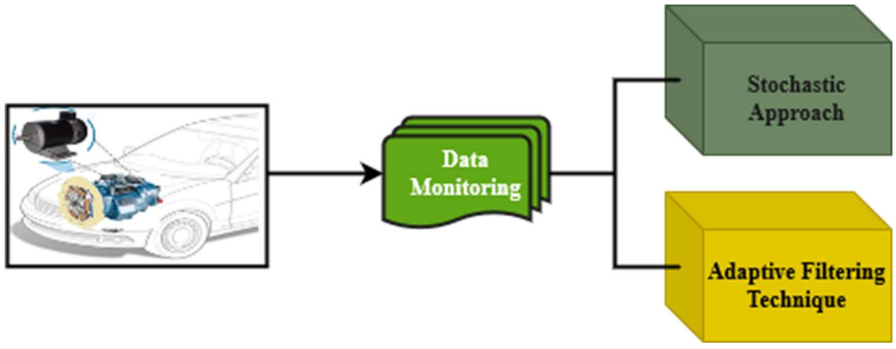

Figure 4.1 Schematic of data-driven RUL estimation methodologies

Most degradation phenomena are complicated (owing to nonlinearities, stochasticity, non-stationarity, etc.) and difficult to represent analytically in practice. The time series data features using signal processing for engineering applications have been earlier studied (Antonino-Daviu 2020; Mohanty 2014) in order to obtain a robust signal characteristic. A random signal with time varying statistical property is said to possess a strictly nonstationary characteristic. The presence of nonstationary stochastic property is mostly observed when there occurs a stochastic change in statistical property or integrating type non-stationarity (Tangirala 2015). The inherent trend could be visually observed, while the signal characteristic can be identified only after signal processing.

Data obtained from the experiments were analyzed after a longer duration and the non-stationary dataset for the dynamical system indicated that the data lacks ergodicity. The entropy offers a measure of uncertainty as well as the maximum amount of useful information one can get out of a dataset. Furthermore, if the entropy offers a monotonic measure, the dataset is said to possess finite learning ability that aids in devising intelligent machines (Looney et al. 2015). To get around this, researchers may further employ learning methods to create a conceptual framework that depicts the degradation in the form of latent health risks, including such PF models or exponential degradation models. Setting the assumptions and defining the technique's limits are necessary for each modelling and implementation process. To construct the conceptual framework in adaptive filtering, several elements must be defined, such as the number of states, the data set, the amount of training data, and so on. Furthermore, the measurements must be persistent but not limited in size. In practice, however, the data size has an impact on algorithms learning and inference capability.

In the realms of failure diagnostics and prognostics, the stochastic approach provides an efficient online fault prognosis by updating the most recent observations for dynamical systems automatically. Currently, the electromechanical actuator's inputs are routed through a 32bit controller with sufficient set-points, limiting the controller's ability to work beyond a safe limit. As a result, anticipating the system RUL with any further change in the system's input-output parameter values is a challenge. Adaptive filtering techniques have proven to be a good tool for incorporating the necessary adjustments made by the controller during operation, whilst it enables to mimic the actual product's degradation utilizing continuous measurements offered by the various sensors.

4.2 Overview of the Data Pre-Processing

The proposed methodology (see Fig. 4.2) involves the use of the data driven approach (Coble 2010; Jouin et al. 2016; Lei et al. 2018; Zhu et al. 2020) for meaningful interpretation of the changes occurring during the period of transient operation. Fig. 4.2 describes the prognostics protocol module from data import to RUL prediction on the basis of model accuracy. Chapter 3 elaborates the ALCT approach towards data collection for three different working conditions.

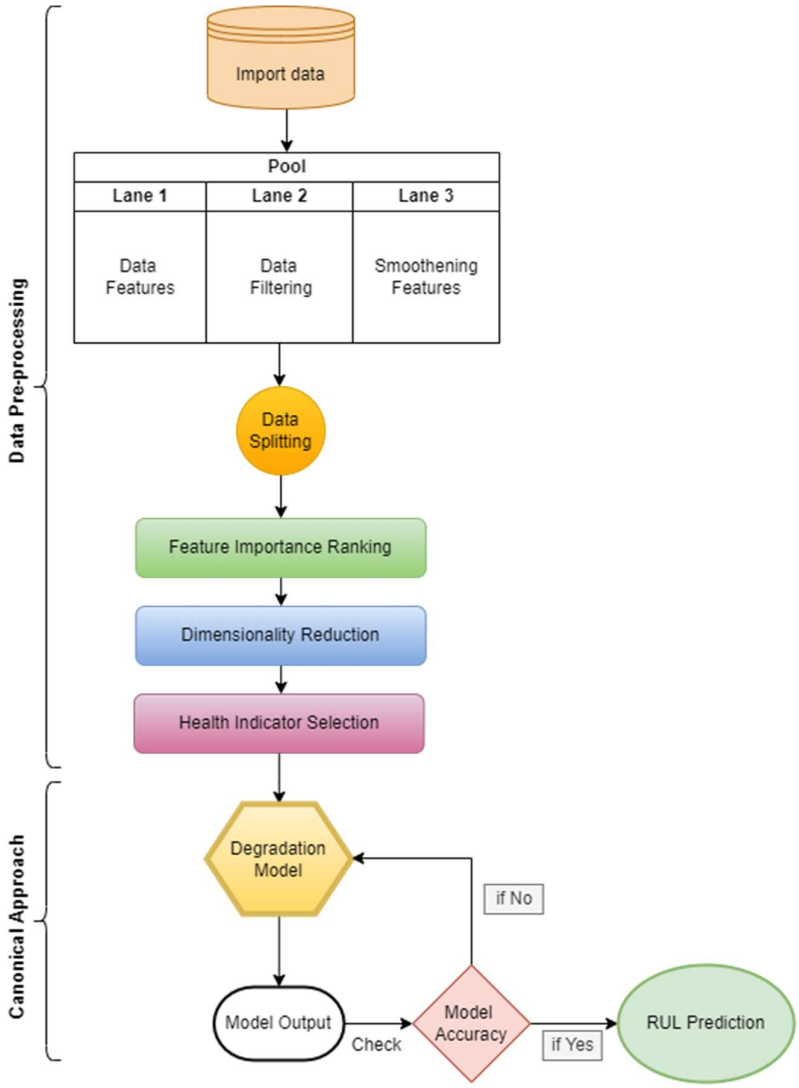

Figure 4.2 Prognostics Protocol Module

The initial step to the proposed approach commences with the feature extraction, which is then followed by the filtering, and smoothening. Assuming only 60% of the data has been used for training purposes the feature importance ranking (FIR) has been developed. The raw data was processed to extract the statistical features followed by smoothening to attenuate the effect of noise. Monotonocity and trendability metrics were applied on the part of the smoothened features (training data) for ranking statistical features (Qian and Niu 2016). Feature fusion (Aymaz and Kose 2018; Jardine et al. 2006) was carried out using parameters of the training dataset and PCA. Matrix diagonalization was carried out for estimation of the eigen values and vectors. The largest Eigen vector was selected for fusion technique which represented the ideal signal. The PCA (Yang et al. 2019) based fused features were applied for the selection of the appropriate CI. Thus, selected

features from FIR with a target score qualified for the PCA enabled feature fusion to produce a healthy CI. Degradation model fitting was carried out on the health indicator. The proposed approach used a known threshold value of the CI for determining the RUL (Chen et al. 2020). Step wise description for the same can be visualised in Fig. 4.3 below.

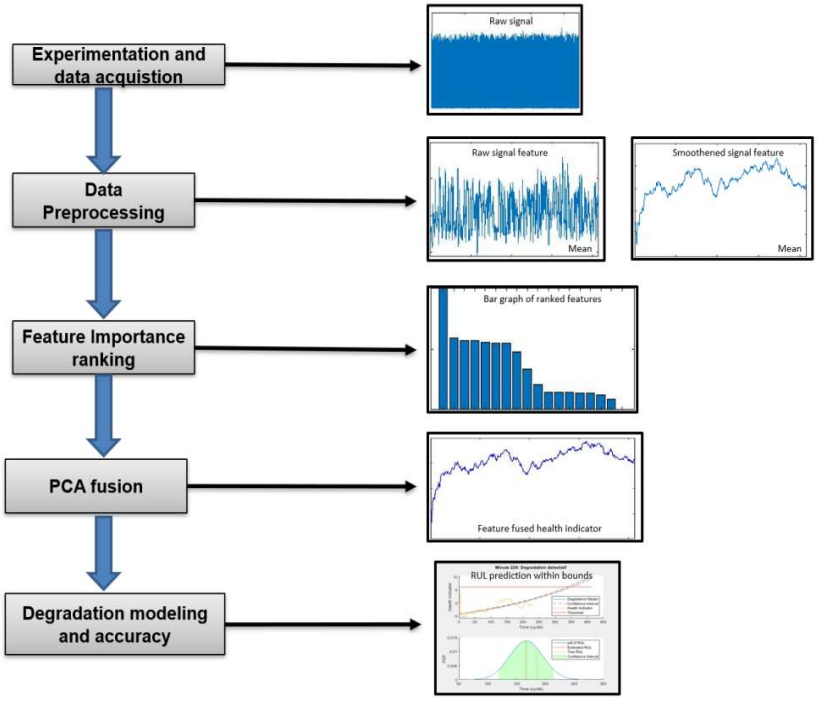

Figure 4.3 Flowchart of the proposed approach

Each step of data processing, from the generation of the CI through the RUL finding, is described in depth in the subsections below.

4.2.1 Data Processing

Raw data from the sensor loggers were pre-processed to be applied to the degradation model. Statistical features, as described in Table 4.1 were extracted from the noisy current signal dataset. The moving average filtering was implemented on the statistical features of the dataset. The filtering was intended to smoothen the prime characteristic of the statistical features. The smoothened datasets were normalized using statistical parameters (mean and standard deviation). Detailed explanation to feature extraction and smoothening process are described hereunder.

4.2.1.1 Feature Extraction

The statistical characteristics (see Table 4.1) of the current (I_a) signal were determined using the logged dataset in time, frequency and the time-frequency domains. The equations provided in Table 4.1 depicts the statistical features employed in this study.

Table 4.1. List of extracted features considered in the study

Features	Equations	Features	Equations
Mean	$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$	**RMS**	$\sqrt{\frac{1}{\mathrm{N}}\left(\sum_{i=1}^{N}\lvert X_i\rvert^2\right)}$
Standard Deviation	$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{n}(X_i-\mu)}$	**Crest Factor**	Peak2Peak/RMS
Skewness	$\frac{\mathrm{E}(X_i-\mu)^3}{\sigma^3}$	**Shape Factor**	$\frac{RMS}{\frac{1}{\mathrm{N}}\left(\sum_{i=1}^{N}\lvert X_i\rvert\right)}$
Kurtosis	$\frac{\mathrm{E}(X_i-\mu)^4}{\sigma^4}$	**Impulse Factor**	$\frac{Peak2Peak}{\frac{1}{\mathrm{N}}\left(\sum_{i=1}^{N}\lvert X_i\rvert\right)}$
Peak2Peak	$\frac{1}{2}(\max(X_i)-\min(X_i))$	**Margin Factor**	$\frac{Peak2Peak}{\frac{1}{\mathrm{N}}\left(\sum_{i=1}^{N}\sqrt{\lvert X_i\rvert}^2\right)}$

Where $\boldsymbol{X_i}$, μ, σ, and N respectively represent the raw sensor data, mean of the data, standard deviation, and the total number of cycles. The time domain features are considered for the less complexity in calculations. However, significant changes in the statistical properties of the features were observed for faulty conditions of the motor in contrast to the normal operating regime. These features need to be sensitive to the faults besides being significant in relation to the noise. Such dynamic feature of the system helps in prior diagnosis of the faults. The time domain features obtained by increased order of moments also help in categorizing the system's health state. Thus, the noisy masked signal in time domain may offer sufficient results for diagnosis of the system. Although much had been spoken about the benefits of frequency domain and its use in identifying the change in process variables (Pappachan et al. 2017), but the drawback of the

time to frequency domain transformation appears in terms of the increase in response time besides increased computational efforts. Thus, the transformed data in frequency domain poses a challenge in modelling the noise for a stochastic signal. The power of the spectrum for current signal was estimated using spectral entropy and the uncertainty was determined by information entropy of the stochastic signal. Further, the Shannon entropy principle states that the unlikely events will have higher entropy than the most likely ones. For a known time-frequency power spectrum *S(t, f)*, the probability distribution (*P(m)*) and spectral entropy (*H*) are:

$$P(m) = \frac{\sum_t S(t,m)}{\sum_f \sum_t S(t,f)} \tag{4.1}$$

$$H = -\sum_{m=1}^{N} P(m)\, log_2\, P\,(m) \tag{4.2}$$

The Shannon entropy therefore finds significance in the event of complex noisy data. Since the fourth moment (kurtosis) of the dataset reflects the peak of the signal, the complex signals produced by an electromechanical actuator in transient state will display high amplitude characteristic. This way, for the sampling of data at 10 milli-seconds interval, kurtosis (see Eq. 3) gave a better representation of the data characteristic. The kurtosis is given by;

$$Kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{X}_i - \mu)^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(\mathbf{X}_i - \mu)^2\right)^2} \tag{4.3}$$

4.2.1.2 Smoothing

The features extracted from the dataset are usually associated with noise. These noises may adversely affect the performance of degradation model predicting RUL. The "causal" moving average filter of window length *M* involves the average of every *N* consecutive samples of the waveform while ignoring the future values. They are used to compensate the unavailability of future sample points during the real-time smoothening of data. This kind of linear time invariant filtering technique is applied to low pass filter that restricts higher frequency signals. The expression for moving average filter is:

$$y(i) = \frac{1}{M+1}\sum_{j=0}^{M-1} \mathbf{X}(i-j) \tag{4.4}$$

Where, $\frac{1}{M+1}$ is a coefficient, known as window length and M denotes sample size. The order of the causal moving average is determined on the basis of global trend and local variations.

4.2.2 FIR and Health Indicator Determination

Monotonocity (Coble 2010; Liao 2014, Gomes et al. 2016) and Trendability (Coble 2010) metrics were used to rank the features. The non-self-healing property and underlying positive/negative trend were characterized using monotonocity. The variables p_m and p_n used in Eq. (5) respectively represent positive and negative differences between the features, whereas, n shows the number of observations. Monotonocity is low in the presence of noise, and it is found to be 1, if it increases or decreases. The linear correlation amongst the features and their absolute values are the measure of trendability. The monotonocity and trendability are expressed as:

$$Monotonocity(x) = \left| \frac{+\mathrm{p}_m}{n-1} - \frac{-\mathrm{p}_n}{n-1} \right| \tag{4.5}$$

$$Trendability = \min\left(\left|corrcoef\right|\right) \tag{4.6}$$

Whereas, FIR has been defined as: $FIR = Monotonocity + Trendability$ (4.7)

Post normalization, higher scores above 0.6 for the are qualified to be well behaved, however, the value of unity (1) is considered to be desirable for the fitness metric. Peng et al. (2005) addressed the importance of feature selection in pattern recognition problems by introducing maximal statistical dependency criterion based on mutual information. Unsupervised situation (Hamaide and Glineur 2021; Peng et al. 2005) often makes use of the criterion of minimal classification error metric thereby inducing statistical dependency. Statistical dependency is realized using relevance which is a measure of correlation. Trendability makes use of correlation coefficient in assisting the best possible characteristic feature from a nonstationary signal. For the selection of a good condition indicator, the metrics (i.e., monotonocity and trendability) need to be consistent, i.e., either increasing or constant but not decreasing.

The involvement of Eigen vector for feature fusion simplified the selection of health indicator and is discussed in the subsequent section.

4.2.3 Feature Fusion using PCA

The training dataset was finalized on the basis of fitness function (FIR). Statistical characteristics of the training dataset were estimated and the best features were selected in accordance to the FIR. The selected best features were normalized using parameters (mean and standard deviation) of the training dataset and PCA was implemented thereupon. The overview of the PCA (Yang et al. 2019) is described below:

The PCA is an unsupervised dimensionality reduction technique. Estimation of the mean, covariance, eigen values and vectors is carried out to implement the PCA.

The covariance $\mathbf{C_x}$ is determined for the matrix **X** that has mean centred and scaled sensor data:

$$C_x = \frac{1}{M-1}\left(XX^T\right) \tag{4.8}$$

The eigenvalues and eigenvectors are determined by using matrix diagonalization as:

$$\det(X - \lambda_i I) = 0 \tag{4.9}$$

$$C_x V_i = \lambda_i V_i \tag{4.10}$$

$\mathbf{V}_i$ is the eigen vectors of the sample covariance matrix $\mathbf{C_x}$. The λ_i represents eigen values. **I** is the identity matrix.

The eigen vector represent the direction in which the variables strongly correlate. The eigen vector with the largest eigen value represents the first principal component and it possesses the maximum variance. These principal components share a relation of orthogonality and a few principal components are usually sufficient to explain the behaviour of the dataset, which, in turn, assist in overcoming the challenge of overfitting due to redundant variables. The steps for fusion of the features using PCA are described below:

Step 1: 60% of smoothened features data was chosen to form a data cluster

Step 2: Determination of FIR metric for the data cluster chosen in step 1

Step 3: FIR score larger than 0.6 was selected to obtain the training dataset from the data cluster of step 1

Step 4: Calculation of the statistical parameters of the 'training dataset' for normalization

Step 5: Determination of the eigen vectors for the normalized training dataset

Step 6: The entire smoothened features data was now normalized using the parameters of training dataset

Step 7: The entire data obtained from step 6 is multiplied with the eigen vector to yield the principal components

Step 8: Plotting and visualization of the principal components

The first principal component held the maximum variance, therefore, used as a tool to obtain the promising CI for the degrading motor. Data fusion works (Ompusunggu et al. 2012) has been earlier attempted for wet friction clutches in which the three captured degradation signals are utilised to obtain an single health index. It is the CI, which depicts the actual trend in the degradation behaviour of the motor that will be used for estimating the RUL by means of the degradation modelling and subsequently the adaptive filtering approach.

4.3 Stochastic Approach

4.3.1 Exponential Degradation Modelling and RUL Prediction

Linear, polynomial, power and exponential models are available in literature to fit the data characteristics of the degrading systems. The continuously changing pattern of the data signal over the period of observation reflected the prevalence of cumulative (Gebraeel 2006) deteriorating behavior of the motor. Therefore, exponential degradation model has been adopted to model the behavior of the system. Assessment of the prior distribution (Gebraeel 2009) of the stochastic parameters is the initial step to progress towards modelling. The online estimation and updating (Yu et al. 2017; Zhou et al, 2011; Gebraeel 2006) of the parameters of this model is carried out using Bayesian framework (Gebraeel et al. 2005). Maximum Likelihood Estimation and expectation maximization (Saxena et al. 2008; Celaya et al. 2011) are commonly applied approaches to estimate the parameters of continuously changing signal. These parameters are the respective state variables that change with respect to time. Thus, it constitutes a model for non-linear dynamic system with online tracking algorithm.

The general form of the exponential degradation model is:

$$S(t) = \phi + \theta \exp\left(\beta t + \varepsilon(t) - \frac{\sigma^2 t}{2} \right) \tag{4.11}$$

Where, $\boldsymbol{\Phi}$ is a constant term representing the initial level of degradation, stochastic parameter θ is a lognormal random variable with mean μ_0 and variance σ_0^2, β is a normal random variable with mean μ_1 and variance σ_1^2, and $\varepsilon(t) = \sigma W(t)$ is an error term with mean 0 and variance $\sigma^2 t$. W(t) is a Brownian error. The random error term $\varepsilon(t)$ explains the noise and the randomness. The variables θ, β and ε(t) are mutually independent and modelled as a stochastic process. Stochasticity $\{\mathbf{S}(t), t>0\}$ is presumed to be present in the underlying model parameters that helps in accurately specifying the path followed by the model. The present study considers the error to be continuous

stochastic process to capture the complexity in the data. The following steps have been adopted in degradation modelling.

Step I: Estimation of the prior distribution of stochastic parameters of degradation signal, $p(\theta, \beta)$

Step II: Updation of the prior distributions using Bayesian method

Step III: Computation of the posterior, $p(\theta, \beta \mid S(t_i))$ for RUL estimation

4.3.2 RUL Computation

For the stochastic signal, $S(t_i)$ of the motor describing the path of the degradation signal: $S(t_i) = \gamma(t_i \mid \beta, \phi) + \varepsilon(t_i)$ i = 1, 2, 3…, . $\boldsymbol{\Phi}$ is the deterministic parameter, and stochastic coefficient, β follows a prior distribution with normally distributed error, $\mathbf{N}(0, \sigma^2)$. The *apriori* failure threshold '*x_threshold*' in advance at time '*t*', the expected RUL, '*RUL*' at any arbitrary time instant can be expressed as:

$$RUL_T = P\left(T \le t\right) = P\left[\gamma\left(t \mid \beta, \phi\right) + \varepsilon(t) \ge x_{threshold}\right] \tag{4.12}$$

The uncertainty associated can be computed using expressions likewise as:

$$\begin{aligned} \Delta RUL^{lower} &= RUL - RUL^{lower} \\ \Delta RUL^{upper} &= RUL^{upper} - RUL \end{aligned} \tag{4.13}$$

4.4 Adaptive Filtering Technique

The methodology of motor RUL estimation using adaptive filtering technique comprises of three variants of particle filters. The flowchart in Fig. 4.4 consists of three stage description of the filtering process. Importing the developed CI is the initial step. Next follows the PF, UPF and i-UPF stages that includes the respective particle generation, the weight updating followed by the resampling and thereafter the motor's state prediction. The third stage is the motor RUL calculation, which is done utilizing the predicted states and the failure threshold. The subsequent sections will provide a detailed description of the allied PF variants. Then the four resampling techniques are discussed. RUL estimation is discussed at last.

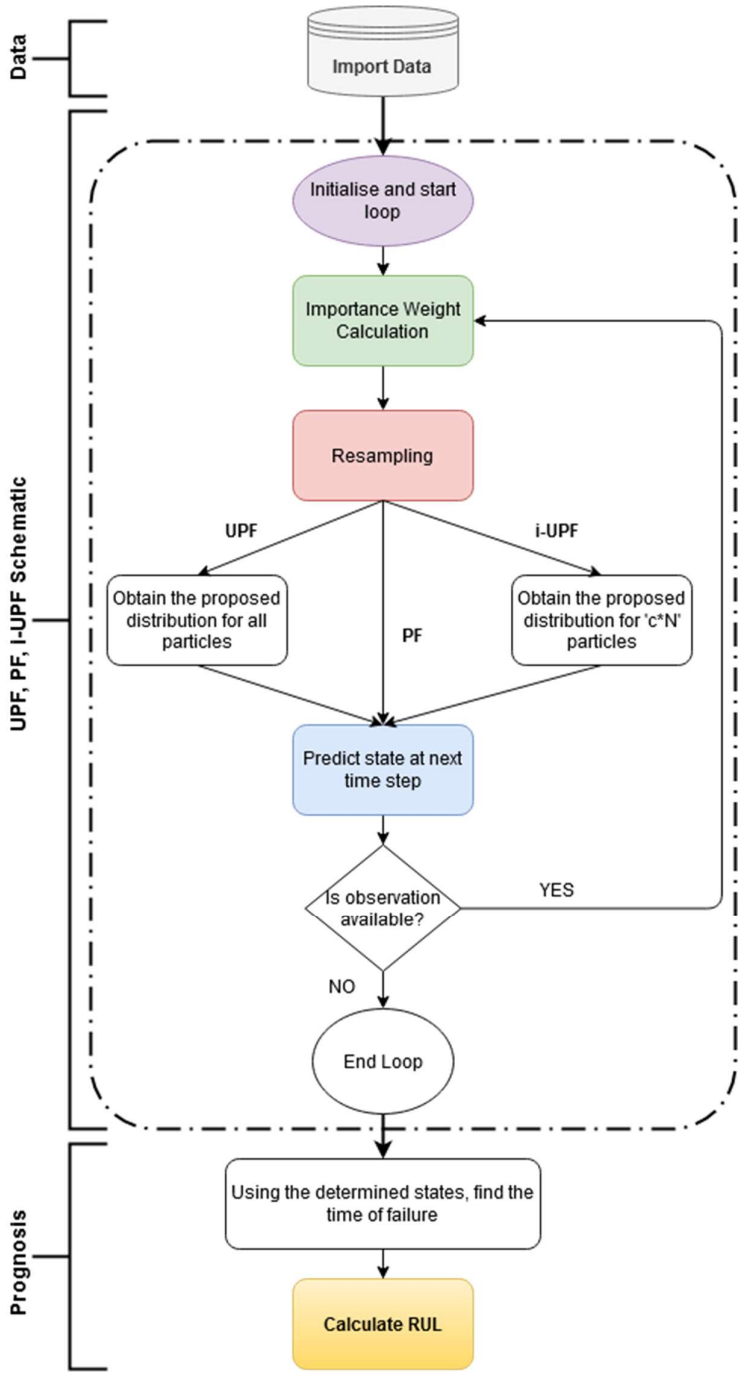

Figure 4.4 Illustration of particle filter variants in motor health prognosis

4.4.1 Particle Filter

PF approach to filtering involves linearizing the fitted models using Taylors expansion series to achieve the nearest approximations. The approximated nonlinear functions yielded less accurate estimations, thereby giving rise to approximating the distribution by introducing the state-space

model. The state variables follow a Markov first-order process. For Bayesian recursive filtering, the PF uses the posterior distribution (An et al. 2013a)of the state as per Bayes rule:

$$p(x_t \mid y_t) = L(y_t \mid x_t^j)\, p(x_t) \tag{4.14}$$

where X is a vector of unknown parameters (*a, b, c, d*) and y_t^j is a vector of observed data with the likelihood expressed as $\mathrm{L}(\mathrm{y}_t \mid \mathrm{x}_t^j)$ and $\mathrm{p}(\mathrm{x}_t)$ being the prior of the pdf.

A nonlinear state transition function (see Eq.15) is obtained using previous state vector, x_{t-1} and additive process noise, $u_t^j \sim \mathrm{N}(0, \sigma_u)$ to state, and can be expressed as:

$$x_t^j = h(x_{t-1}^j) + u_t^j \tag{4.15}$$

Predicting the measurement data for each particle *j* with nonlinear state vector x_t^j and adding a measurement noise, $v_t^j \sim \mathrm{N}(0, \sigma_v)$ to it

$$y_t^j = f(x_t^j) + v_t^j \tag{4.16}$$

The measurement equation (see Eq. 16) describes a relation between the observed (y_t^j) and unobserved state ($x_t^j \mid j = 1 \ldots \mathrm{M}$) variables. For prognostics, the state equation, on the other hand, reflects the dynamics of the state variables. The standard PF consists of three steps: particle generation, updating the weights, and resampling, which can be seen as expressed below. Sequential importance sampling (SIS) is used for particle generation and weight computation (Guo et al. 2015a). At each time step *t*,

1. Initialize particle generation - Initially, when $t = 0$, the weights are the same for all the (uniformly distributed) particles (j=1….M) and are generated according to the prior probability distribution, $p(x_0)$. It is expressed as: $w_0 = \frac{1}{M}$. The SIS finds the particles for t^{th} moment and adds them to t-1^{th} moment i.e., $x_{t-1} = \left\{ x_{0:t-1}^j, \mathrm{w}_{t-1}^j \right\}$. The proposal distribution (importance density) can be obtained using the prior, which is expressed as:

$$p(x_t^j \mid x_{t-1}) = p\left(x_t \mid x_{0:t-1}^j, \mathrm{y}_{1:t} \right) \tag{4.17}$$

The estimation and update are the two steps that follow the Bayesian conditional probability in a PF approach (Arulampalam et al. 2002)(Zhang et al. 2018).

2. Computing the particle weights and updating– The updating of particle weights helps to construct the posterior pdf that can be expressed below as:

$$w_t^j \propto w_{t-1}^j \frac{p\left(y_t \mid x_t^j\right) p\left(x_t^j \mid x_{t-1}^j\right)}{p\left(x_t^j\right)} \tag{4.18}$$

The normalized weights can be rewritten in the form: $w_t^j = \frac{w_t^j}{\sum_{j=1}^{M} w_t^j}$ (4.19)

The third step is the resampling in which the set of assigned particles $\left(x_t^j, \mathrm{w}_t^j\right)$ is replaced with another new set $\left\{x_{t-1}^j, \mathrm{w}_{t-1}^j\right\}$ (Li et al. 2015c) which is further used for obtaining the posterior pdf.

3. Resample the particles – The resampling is purely performed based on the weight of the particles. More is the weight of the particles, the likelihood of the particles within the range is higher. Therefore, evaluating the likelihood and CDF value for each particle is processed.

$$\mathrm{L}(\mathrm{y}_t \mid \mathrm{x}_t^j) = \sum_{t=1}^{T} \frac{1}{\sqrt{2\pi}\nu_t} \exp\left(-\frac{\left(\mathrm{y}_t - \mathrm{x}_t^j\right)^2}{2\nu_t^2} \right) \tag{4.20}$$

It is evident when the effective sample size, *N-eff* is below the given threshold, *N-th* the low weight particles are excluded, and resampling occurs. Thus, to proceed further:

$$N - eff = \frac{1}{\sum_{j=1}^{M} (w_t^j)^2} \tag{4.21}$$

The particles and unknown parameters are used to process the updating step in a sequential way. The paramount importance is in estimating the hidden states as the online data becomes available and modeling the measured data along with noise for the observed states. For a normally distributed measurement noise, the normally distributed likelihood function is used to update the measurement data.

4.4.2 Unscented Particle Filter and Improved Unscented Particle Filter

The requirement of predictability of the state dynamics under optimal bias is achieved via unscented transform (UT) and evaluating the sigma points. Recursive estimation of parameters (the mean, $\overline{x}$ and covariance, P_x) in a dynamic transient process with random variable(x) uses a

nonlinear function given as $\Phi = g(\mathrm{x})$. UT uses $2\,\mathrm{n}_x + 1$ weighted samples and can be described under the following steps:

i. Evaluate sigma points (χ)

$$\chi_{t-1}^{i} = \left[\bar{x}_{t-1}^{i},\ \bar{x}_{t-1}^{i} + \sqrt{(n_x + \lambda) P_{t-1}^{i}},\ \bar{x}_{t-1}^{i} - \sqrt{(n_x + \lambda) P_{t-1}^{i}} \right] \tag{4.22}$$

$$\lambda = \mathrm{n}_x(\alpha^2 - 1) \tag{4.23}$$

Where α and λ are constants. n_x denotes the number of sigma points that are found using the mean and covariance of the x.

ii. Propagating the sigma points using nonlinear transform

$$\gamma_i = g(\chi_i) \text{ for, } i = 1 \ldots 2n_x \tag{4.24}$$

iii. Evaluating the parameters of the transformation function

$$\begin{cases} \bar{\Phi} = \sum_{i=0}^{2n_x} w_i^{(m)} \gamma_i \\ P_{\Phi} = \sum_{0}^{2n_x} w_i^{(c)} (\gamma_i - \bar{\Phi})(\gamma_i - \bar{\Phi})^T \end{cases} \tag{4.25}$$

Weights of the sample mean, $w^{(m)}$ and sample covariance, $w^{(c)}$ are calculated using the following equations:

$$\begin{cases} w_0^m = \dfrac{\lambda}{n_x + \lambda} \\ w_0^c = \dfrac{\lambda}{n_x + \lambda} + (1 - \alpha^2 + \beta) \\ w_i^m = w_i^c = \dfrac{1}{2(n_x + \lambda)} \end{cases} \tag{4.26}$$

The constant $\alpha \geq 0$, and β denotes the prior distribution of x thus, incorporating higher order moments. For a Gaussian distribution $\beta = 2$. Smaller values are preferred in case of stronger nonlinearity.

4.4.2.1 Unscented Particle Filter (UPF)

The UPF attempts to produce an approximated posterior probability density function using a UT and an EKF approach. The information gain is achieved when the sigma points are propagated via UT.

i. Initial steps for implementing UPF are initialization (Eq. 4.27) and computing sigma points (see Eq. 4.28)

$$\begin{cases} \overline{x}_0 = E\left(x_0\right) \\ P_0 = E[\left(x_0 - \overline{x}_0\right)\left(x_0 - \overline{x}_0\right)]^T \end{cases} \tag{4.27}$$

$$\chi_{t-1} = \left[\, \overline{x}_{t-1},\ \overline{x}_{t-1} + \sqrt{(n_x + \lambda)P_{t-1}},\ \overline{x}_{t-1} - \sqrt{(n_x + \lambda)P_{t-1}} \,\right] \tag{4.28}$$

ii. The second step is the time update (to propagate particles in future steps) and are computed as:

$$\begin{cases} \chi'_{t|t-1} = h(\chi_{t-1}) \\ \overline{x}_{t|t-1} = \sum_{i=0}^{2n_x} w_i^{(m)} \chi'_{i,t|t-1} \\ P_{k|k-1} = \sum_{i=0}^{2n_x} w_i^{(c)} (\chi'_{i,t|t-1} - \overline{x}_{t|t-1})(\chi'_{i,t|t-1} - \overline{x}_{t|t-1})^T + Q^u \\ \chi_{t|t-1} = \left[\, \overline{x}_{t|t-1}, \overline{x}_{t|t-1} + \sqrt{(n_x + \lambda)P_{t|t-1}}, \overline{x}_{t|t-1} - \sqrt{(n_x + \lambda)P_{t|t-1}} \,\right] \\ \gamma_i = f(\chi_{t|t-1}) \\ \overline{\mathrm{y}}_{t|t-1} = \sum_{i=0}^{2n_x} w_i^{(m)} \gamma_{i,t|t-1} \end{cases} \tag{4.29}$$

iii. Finally, in the measurement update step (that adds the recent observations to tune the parameters automatically) the equations are calculated as:

$$\begin{cases} P_{\overline{y}_k \overline{y}_k} = \sum_{i=0}^{2n_y} w_i^{(c)} (\gamma_{i,t|t-1} - \overline{y}_{t|t-1})(\gamma_{i,t|t-1} - \overline{y}_{t|t-1})^T + Q^v \\ P_{\overline{x}_k \overline{y}_k} = \sum_{i=0}^{2n_y} w_i^{(c)} (\chi_{i,t|t-1} - \overline{x}_{t|t-1})(\gamma_{i,t|t-1} - \overline{y}_{t|t-1})^T \\ G_t = P_{\overline{x}_k \overline{y}_k} P^{-1}_{\overline{y}_k \overline{y}_k} \\ \overline{x}_t = \overline{x}_{t|t-1} + G_t(\mathrm{y}_t - \overline{y}_{t|t-1}) \\ P_t = P_{t|t-1} - G_t P_{\overline{y}_k \overline{y}_k} G_t^T \end{cases} \tag{4.30}$$

The Q^{u} and Q^{v} are the prior known noise covariances for process and measurement set. G_{t} is Kalman gain

4.4.2.2 Improved Unscented Particle Filter (*i*-UPF)

The *i*-UPF algorithm considers the latest set of observations to obtain the posterior for which it uses UKF to generate the proposal distribution. The improvement is in the reduced execution time. The idea of using particles in two-step frameworks (i.e., one generated from transition prior, M-*c*M* and another from UKF, *c*M*) is the strategy in implementing *i*-UPF. The knowledge for the selection of the number of particles in each step is quite cumbersome. However, our study introduces this aspect as one variable, '*c*'. The proper selection of '*c*' (Fasheng and Yuejin 2009) is by computing the mean and variances across multiple runs. It can be well designated as:

i. The initialization step is to draw the particles from a known prior and computing the system state vector's mean and covariance incorporating the noise.
ii. The second step is to follow the steps of UPF as described under section 4.3.2.1 to finally obtain the mean $\overline{x}_t$ and covariance P_t at time t for one set of particles from UKF. The subsequent step is to draw samples from the remaining set, i.e., transition prior.
iii. The third step is to draw another set of remaining particles from the transition prior and computing the $\overline{x}_t$ and P_t.
iv. Finally, with the involvement of resampling, the final particles and weights are calculated.

The state and the covariance are then expressed as:

$$\overline{x}_t^j = \sum_{j=1}^{M} w_t^j x_t^j \tag{4.31}$$

$$P_t^j = \sum_{j=1}^{M} w_t^j \left(x_t^j - \overline{x}_t^j\right)\left(x_t^j - \overline{x}_t^j\right)^T \tag{4.32}$$

4.4.3 Resampling Methodology

The motivation in using resampling methods in the present study is three-fold: the accuracy, robustness, and the ease of computational complexity (Kootstra and de Boer 2009) through a practical application. Accuracy is attained with the declining particle degeneracy (Jouin et al. 2016; Tulsyan et al. 2016). The increased ability to handle the heavy-tailed particles thereby contributes towards eradicating particle impoverishment (Li et al. 2015c), which adds to the results' robustness. An acceptable tradeoff between robustness and accuracy represents the novelty in the present work by overcoming the major obstacle in PF implementation. Both the deterministic and dynamic resampling schedule is chosen, and resampling is said to perform

whenever the variance of the importance weights exceeds the threshold. Comparative study with the resampling(Guo et al. 2015; Pugalenthi and Raghavan 2018) methods in PF shall provide a much better understanding in the purview of the above mentioned three aspects along with the increased efficacy of the algorithms. The notion of effective sample size (ESS) of a sequence of weighted observations is thereby required for this adaptive integration of the resampling technique. The present study uses the four traditional multinomial resampling, stratified resampling, residual resampling, and systematic resampling for the single distribution sampling (Li et al. 2015b) and considering only the weight (Li et al. 2012b) of the particles.

4.4.3.1 Multinomial resampling

One of the efficient methods in eradicating particle degeneracy was proposed in the formulation of multinomial resampling. Ideally, the weights should be equal during sampling. Below is the detailed operation of the steps:

i. The w_t^j represents the resampled *j*-th particle for the expected N_t^j number of times i.e.,

$$E(N_t^j \mid w_t^j) = Nw_t^j \tag{4.33}$$

ii. Generating *N* random numbers u_i' from uniformly distributed (0 1]set.

iii. Replicating particles from $\left(x_t^j, \mathrm{w}_t^j\right)$ using u_i' for *i*-th random sample, to obtain the ordered set,

$$u_i = \left(\sum_{t=1}^{j-1} w_t, \sum_{t=1}^{j} w_t \right] \tag{4.34}$$

The computational complexity is of the order O(*N*). The number of times a particle gets replicated (0 N], thereby increasing variance in the resampled particles, which can be reduced in the subsequent resampling methods.

4.4.3.2 Stratified resampling

i. The separately distributed subintervals are divided from within [0 1] range.

ii. The number of random numbers is equal to the subintervals, *N*. The independently drawn random numbers from each subinterval are $u_i = \left(\frac{i-1}{N}, \frac{i}{N} \right]$

iii. Particles $\{\mathrm{x}_t^j, \mathrm{w}_t^j\}$ are replicated based on the cumulative sum of the normalized weights as seen in Eq. 6

The deterministic relationship between the random numbers reduces the variance within the resampled particles. The computational complexity is again found to be the same as multinomial type, i.e., O(*N*).

4.4.3.3 Systematic resampling

The procedure to generate random numbers is the same as that of the stratified resampling. The significant difference is in using a single random number throughout the resampling and then adding the same to the initial point.

i. The subintervals are equidistantly distributed in [0 1] range.

ii. Sample data u is obtained in the range $u = \left[0, \frac{1}{N}\right]$

iii. Calculate $u_i = \frac{i-1}{N} + u, \quad i = 1, 2....N$ (4.35)

The variance of the resampled particles is smaller than that of the stratified resampling. The much stronger deterministic relationship and the generation of a smaller number of random numbers prefer its use over stratified, thereby enhancing the algorithm's performance and reducing the computational complexity.

4.4.3.4 Residual resampling

The attainment of reduced computational complexity with the use of two-stage resampling is aimed. The j^{th} particle is resampled $N_t^j + R_t^j$ times.

i. Deterministic replication is carried out for particles with a weight greater than *1/N,* thereby reducing the consequences by a multiple of *1/N,* i.e., $N_t^j = Nw_t^j$

ii. Random sampling with the remaining weights (a.k.a., residuals), i.e., $R_t^j = N - N_t^j$.

The importance weight can be expressed as: $w_t'^j = w_t^j - \frac{N_t^j}{N}$ (4.36)

The order for computation time is $O\left(N_t^j\right) + O\left(R_t^j\right)$. The first stage is multinomial type whereas, in the second stage, the variation in the number of times a particle is resampled is attained, reflecting the reduced time in computing.

The criterion for resampling is in proper threshold selection, which can be defined as 0.5 times the size of the particle-set. The performance of the three PF algorithms with the use of four resampling

techniques is presented while estimating the RUL of the motor. RUL_error as a metric is considered for evaluating the prognostic performance which is defined as a ratio of the difference in their absolute value of true RUL and predicted RUL to the true value of the RUL.

$$\text{RUL\_error} = \frac{|tRUL - eRUL|}{tRUL} \tag{4.37}$$

4.4.4 RUL Estimation Strategy

Prognosis deals with the estimated parameters and the observed states for predicting the remaining time, given the threshold value. The predicted states are used to calculate the failure and they are subtracted from the current time step. Extrapolation of the measurement equation is the basis for estimating the RUL.

The approximated Bayesian posterior probability for (t-1)th time instant can be expressed as:

$$p(x_{t-1} \mid \mathrm{y}_{1:t-1}) = \sum_{j=1}^{M} w_{t-1}^{(\mathrm{j})} \delta\left(x_{t-1} - x_{t-1}^{(\mathrm{j})}\right) \tag{4.38}$$

Where Dirac delta $\delta(\cdot)$ represents the summation of particle weights adding to unity. The state can be predicted thereafter as: $x'_{t-1} = \sum_{j=1}^{M} w_{t-1}^{j} x_{t-1}^{j}$ (4.39)

For a known failure, once the predicted value reaches the specified threshold, $\mathrm{x}_{threshold}$ the pdf of the RUL can be estimated as:

$$p(\mathrm{RUL}_t \mid \mathrm{x}_{1:n}, \mathrm{x}_{threshold}) = \sum_{j=1}^{M} w_n^{(\mathrm{j})} \left(RUL_t - RUL_t^{(\mathrm{j})}\right) \tag{4.40}$$

The state transition equation produces possible future states following the time-dependent degradation to make further predictions that can be expressed below as:

$$x_t = [p1_t; p2_t; p3_t; p4_t] \tag{4.41}$$

$$\begin{cases} p1_t = p1_{t-1} + w_{p1} & w_{p1} \sim N(0, \sigma_{p1}) \\ p2_t = p2_{t-1} + w_{p2} & w_{p2} \sim N(0, \sigma_{p2}) \\ p3_t = p3_{t-1} + w_{p3} & w_{p3} \sim N(0, \sigma_{p3}) \\ p4_t = p4_{t-1} + w_{p4} & w_{p4} \sim N(0, \sigma_{p4}) \end{cases} \tag{4.42}$$

$$\mathrm{f(t)}_t = p1_t * \mathrm{t}^3 + p2_t * \mathrm{t}^2 + p3_t * \mathrm{t} + p4_t + u_t \quad u_t \sim N(0, \sigma_u) \tag{4.43}$$

The predicted posterior density function is approximated using the uniformly distributed particle set and can be expressed as:

$$\mathrm{f}(t)_{t+\Delta t} = \sum_{j=1}^{M} p1_t^j * (\mathrm{t} + \Delta \mathrm{t})^3 + p2_t^j * (\mathrm{t} + \Delta \mathrm{t})^2 + p3_t^j * (\mathrm{t} + \Delta \mathrm{t}) + p4_t^j \tag{4.44}$$

The estimated posterior gives rise to a predicted pdf that can be evaluated using the expression

$$p(\mathrm{f(t)}_{t+\Delta t} \mid \mathrm{f(t)}_{0:t}) = \sum_{j=1}^{M} w_t^{(\mathrm{j})} \delta\left(\mathrm{f(t)}_{t+\Delta t} - \mathrm{f(t)}_{t+\Delta t}^{(\mathrm{j})}\right) \tag{4.45}$$

For a degradation dataset, the RUL_t pdf at cycle t is expressed as

$$p(\mathrm{RUL}_t \mid \mathrm{RUL}_{0:t}) = \sum_{j=1}^{M} w_t^{(\mathrm{j})} \delta\left(RUL_t - RUL_t^{(\mathrm{j})}\right) \tag{4.46}$$