Kalman filtering based prognostics for automatic transmission clutches

Agusmian Partogi Ompusunggu, Jean-Michel Papy and Steve Vandenplas

Abstract—Demands of low-cost prognostics tool for automatic transmission clutches (i.e., based on measurement data from sensors typically available) by industry have increased since the last few years. In this paper, a prognostics tool is developed by fusing a newly developed degradation model with the pre-lockup feature from measurement (i.e., engagement duration) under the extended Kalman filtering framework. As this feature can be extracted from sensory data typically available in wet clutch applications, the developed prognostics tool hence does not require extra cost for any additional sensor. New history data of commercially available wet clutches obtained from accelerated life tests using a fully instrumented SAE#2 test setup have been acquired and processed. The experimental results show that the prognostics algorithm developed in this work outperforms the early developed prognostics algorithm, which is based on weighted mean slope method (i.e., data-driven approach). It is shown that the clutch remaining useful life estimations with the novel prognostics algorithm remain in the desired accuracy region of 20% with relatively small uncertainty interval in comparison with the early developed prognostics algorithm.

Index Terms—Automatic transmissions, Wet friction clutches, Degradation model, Extended Kalman filter, Condition monitoring, Prognostics and Health Management (PHM).

I. INTRODUCTION

Nowadays, automatic transmissions have become a popular choice in commercial vehicles and been widely used in off-road/heavy duty vehicles. As is obvious from its name, an automatic transmission is a transmission that shifts power or speed by itself. The key element that enables automatic power-shifting or speed-selection in automatic transmissions is a wet friction clutch. The power transmission from the engine to the wheels through wet friction clutch is based on the friction occurring in lubricated contacting surfaces.

A wet friction clutch (hereafter called wet clutch) is lubricated by an automatic transmission fluid (ATF) having a function as a cooling lubricant cleaning the contacting surfaces and giving smoother performance and longer life. For high power applications, the clutch is typically assembled with multiple friction and separator discs, see Fig. 1(a). The friction disc is made of a steel-core-disc with friction material bonded on both sides and the separator disc is made of plain steel. An electro-mechanical-hydraulic actuator is usually used for engaging/disengaging a wet clutch. This actuator consists of some main components, such as a piston, a return spring which is always under compression and a hydraulic group consisting of a control valve, an oil pump, etc. To engage the clutch, pressurized ATF is controlled by the valve to generate a force acting on the piston. To disengage the clutch, the pressurized ATF is released such that the return spring is allowed to push the piston back to its rest position.

The duty cycle of a wet clutch can be divided into four consecutive phases: (i) fully disengaged, (ii) filling, (iii) engagement and (iv) fully engaged phase, as illustrated in Fig. 1(b). In the fully disengaged phase \((t < t_f)\), i.e., prior to the clutch actuation, the return spring holds the piston at its...
rest (re-tracked) position so the two elements, i.e., friction and separator discs, are to rotate independently with the rotational speeds of $\omega_s$ and $\omega_d$, see the top leftmost scheme in the figure. In the filling phase which occurs between the time instants $t_f$ and $t_e$, the actuator is activated when the relative rotational speed between the input and output shafts is at a certain value $\omega_{rltr}$, pushing the piston to move as quick as possible before making contact with the neighbor disc. During the engagement phase ($t_e < t < t_l$), the ATF pressure is gradually increased such that gentle contacts between the friction and separator discs can be established, which eventually leads to gradual increase on the transmitted friction torque. As consequence, the $\omega_r$ gradually decreases until it reaches zero value ($\omega_r \rightarrow 0$), meaning that the discs are now to rotate with about the same rotational speeds of $\omega_l$ as illustrated by the top rightmost scheme in Fig. 1(b). Since sliding (rubbing) in the engagement phase constitutes an irreversible process, some portion of the transmitted energy is converted into heat which consequently results in an increase of the ATF temperature. The time instant when the sliding velocity reaches zero value for the first time is called the lockup time $t_l$. After this time instant, the clutch enters the fully engaged phase ($t > t_l$) wherein the relative rotational speed remains around zero value. From now on, the event before the lockup time instant $t_l$ is referred to as the pre-lockup phase while otherwise is referred to the post-lockup phase.

After production, it is desired that a wet clutch can transmit certain power under a smooth and fast engagement with minimal shudder. However due to unavoidable degradations, the clutch frictional characteristics change, thus altering its desired performance and eventually affecting the performance of the vehicles. As the degradation proceeds, failure can unexpectedly occur, which eventually leads to the total breakdown of the vehicles. An unexpected breakdown can put human safety at risk, possibly cause long term vehicle down times, and result in high maintenance costs. Hence, prognostics and health management (PHM) tool for a wet clutch has become important for vehicles equipped with automatic transmissions as it can significantly increase safety and availability/reliability as well as reduce the maintenance cost of the vehicles.

A. Related work

As the performance/condition of wet clutches is highly dependent on its frictional characteristics, a more natural way to monitor the condition of wet clutches is by tracking the mean coefficient of friction (COF) calculated at each duty cycle ([1], [2], [3], [4]). To extract the mean COF for clutch condition monitoring, torque and oil pressure sensors are required. A torque sensor is needed to measure the transmitted torque, while oil pressure sensor is used to estimate the axial force acting on the contacting separator-friction discs. However, this approach is not feasible in practice since torque sensor is typically not available in today’s transmissions. Therefore, prognostics algorithms based on the mean COF would not be workable solution for practical applications.

A number of researchers have explored and developed model-based prognostics techniques for wet clutches. Yang et al. ([5], [6]) developed a physics-based prognostics model by considering that the degradation occurring in a wet clutch is due to thermal effect alone in the friction materials. To this end, a dedicated invasive and destructive test, i.e. Thermal Gravimetric Analysis (TGA), is required for identifying some parameters for the prognostics mode. Since the degradation mechanism occurring in the clutch friction material is not only due to thermal effect but also another major mechanism namely adhesive wear ([2], [7], [8]), the assumption made within the prognostics method in Refs. ([5], [6]) is therefore too oversimplified. Moreover, this approach would be difficult to implement by the end users when the complete design data of a wet clutch system is not available.

Prognostics algorithms for the ATF (i.e. lubricant) of wet clutches have been also developed and reported in the literature. Calcut et al. [9] and Sarkar et al. [10] developed an empirical degradation model for predicting the lifetime of ATF based on an SAE#2 modified plate test, in which the energy per shift and bulk lubricant temperature are used as input parameters. The degradation model applies only to a specific ATF under certain operating conditions. Furthermore, Hirthe et al.[11] developed a prognostics methodology using extra lubricant sensor which is immersed in ATF. The sensor provides an electrical signal indicating in real time the chemical condition of the lubricant to be monitored. Three parameters, namely (i) Total Acid Number (TAN), (ii) delta Oxidation (OX) and (iii) HPDSC induction time (MIN), can be derived from the sensor readings. An empirical model was developed to predict the RUL of ATF based on the three parameters.

Because of its robustness, hybrid prognostics approach under the Bayesian framework (e.g Kalman filtering) has been attracting a number of researchers nowadays and been successfully applied to various applications like bearings, batteries, material crack growth, electrolytic capacitors, etc. ([12], [13], [14], [15]). In this framework, empirical or physics-based degradation models are fused with measurement data (i.e. feature) in order to improve the RUL estimation. To the authors’ knowledge, no work has been posed in literature on applying the Bayesian based filtering (i.e. Kalman filtering) to wet clutch prognostics.

B. Problem statement

Recently, we developed a prognostics algorithm for wet clutches based on a data-driven technique [16]. This prognostics framework is the extension of the pre-lockup feature based condition monitoring method, where the pre-lockup feature is extracted from signals typically measured in automatic transmission applications with low computational complexity. The main advantage of this condition monitoring method as described in detail elsewhere ([17], [18]) is that it does not require extra sensors. Hence, prognostics framework based on the pre-lockup condition monitoring method offers a low-cost solution and is suitable for practical applications.

Notably, it was observed earlier that the pre-lockup feature, extracted from limited amount of lifetime data of some commercial wet clutches, exhibits quasi-linear trend as the degradation proceeds. On the basis of this observation, the weighted
mean slope (WMS) method [19] was applied in the early developed data-driven prognostics method [16]. However, as will be shown later on in this study, this early developed prognostics method is not accurate enough when applied to wet clutches where the pre-lockup feature does not highly exhibit quasi linear-trend. For the sake of reader’s convenience, the pre-lockup feature based condition monitoring method and the early developed prognostic algorithm are briefly reviewed in Section II and Section III, respectively.

C. Objective

This research work aims at developing a novel prognostics algorithm for wet clutches under the extended Kalman filtering framework and furthermore comparing the performance with the one that was developed in the previous work.

D. Paper organization

The remainder of the paper is organized as follows. In Section II and III, the pre-lockup feature based clutch condition monitoring method and the early prognostics algorithm are revisited, respectively. In Section IV, the novel prognostics algorithm development and performance metric used for quantitative evaluation of the developed algorithms are presented. In Section V, the experimental methodology including the test setup and procedure for data generation is discussed. Section VI demonstrates the effectiveness of the new prognostics method and compares its performance to the performance of the early developed prognostics method. Section VII finally concludes important findings obtained from this work and suggests a possible direction for future work.

II. PRE-LOCKUP FEATURE BASED CONDITION MONITORING - REVISITED

Fig. 2 schematically summarizes the pre-lockup feature based condition monitoring method. The pre-lockup feature corresponds to the engagement duration which is calculated as follows. As shown in the figure, both the rotational speed signals ($\omega_l$ and $\omega_r$) measured on the input and output shafts of a wet clutch are used to calculate the slip velocity $\omega_s$. This slip velocity is then used to estimate the lockup time instant $t_l$. At the same time, the pressure signal $p$ is used to estimate a reference time instant $t_f$. For detailed explanation, the interested readers are referred to our previous publication ([17], [18]). Once both $t_l$ and $t_f$ are calculated based on the procedure shown in the figure above, the engagement duration feature $\tau$ can then be calculated according to the following equation:

$$\tau = t_l - t_f.$$  

(1)

In dimensionless form, the feature can be normalized according to the following equation:

$$\tilde{\tau} = \frac{\tau - \tau_0}{\tau_0},$$  

(2)

with $\tilde{\tau}$ denoting the normalized feature and $\tau_0$ denoting feature measured in initial state.

III. EARLY DEVELOPED PROGNOSTICS ALGORITHM - REVISITED

In this method, the slope at the current state (i.e. arbitrary time step to do prediction $t_N$) is recursively computed by summing up all the positive local slopes weighted by a certain function, where the weighting factor of the most recent data is the greatest. Under quasi-linear trajectory assumption, the future trend is predicted and finally the clutch remaining useful life (RUL) can be estimated. Let $\tau = \{\tau_1, \tau_2, \ldots, \tau_N\}$ and $t = \{t_1, t_2, \ldots, t_N\}$ be respectively the observed feature values and the corresponding time sequence at $t_N$. The WMS $b_w$ at $t_N$ is calculated according to the following equation:

$$b_w = \sum_{l=1}^{L} \psi_l b_l^+, \quad (3)$$

with $b_l^+ = \{\forall b_k | b_k \in \mathbb{R}, b_k > 0\} = \{b_1^+, b_2^+, \ldots, b_L^+\}$, $L < N$

$$b_k = \frac{\tau_k - \tau_{k-1}}{t_k - t_{k-1}} \quad (k = 2, 3, \ldots, N),$$  

(4)

and

$$\psi_l = \frac{l}{\sum_{l=1}^{L} l},$$  

(5)

where $b_k$ and $\psi_l$ respectively denote the local slope and the corresponding weighting factor. The standard deviation $\sigma_b$ of the WMS parameter $b_w$ at the time step $t_N$ is calculated according to the following equation:

$$\sigma_b = \sqrt{\sum_{l=1}^{L} \psi_l (b_l^+ - b_w)^2}. \quad (6)$$

For 95% confidence interval, the uncertainty bound of the WMS parameter, i.e. $\Delta b_w$, can be calculated as follows [20]:

$$\Delta b_w = 1.96 \frac{\sigma_b}{\sqrt{L - 1}},$$  

(7)

Suppose that the threshold of the end-of-useful-life $\tau_{lim}$ is given. The expected RUL $r^l$ at the time step $t_N$ can then be expressed as follows:

$$r^l = \frac{\tau_{lim} - \tau_N}{b_w},$$  

(8)
Based on the WMS uncertainty bound expressed in Eq. (7), the uncertainty interval of the RUL (i.e. $\Delta r^j$) can be approximated as follows:

$$\Delta r^j = \left| -\frac{r^j}{b_{uw}} \Delta b_{uw} \right|. \quad (9)$$

IV. DEVELOPMENT OF A NOVEL PROGNOSTICS METHOD

This section first discusses the development of a degradation model that is physically-inspired from typical tribological characteristics of wet clutches during the useful lifetime ([4], [17], [21]). Due to the non-linear nature of the degradation model, a novel prognostics method for wet clutches is then developed under the extended Kalman filtering framework.

A. Degradation Model: Physically-Inspired Derivation

Adhesive wear and thermal degradation are the major causes of wet clutch failure. As the degradation progresses, the friction material surface becomes smoother ([7], [3]) and the shear strength decreases [22]. As a result, the coefficient of friction decreases [22]. As a result, the coefficient

\[ = 0 \]

\[ \tau \]

Depending on the used friction material, ATF and operating condition, previous studies [17] show that the engagement duration (i.e. pre-lockup feature) increases in either quasi-linear or asymptotic way. To predict possible trajectories of the engagement duration feature at an arbitrary time step (i.e. duty cycle) $t_k$, for this purpose, a degradation model was developed as expressed:

$$\tau = \tau_0 + \alpha \left( 1 - e^{-\beta t_k} \right), \quad (10)$$

where $\tau_0$, $\alpha$, $\beta$ are model parameters that can be identified with a non-linear regression technique [23]. Fig. 3 illustrates the effect of the model parameters on the feature trajectory assuming $\tau_0 = 0$. As seen, when $\beta$ is small and $\alpha$ is large, the model can capture quasi-linear trends. On the other hand, the model predicts asymptotic trends when $\beta$ is large and $\alpha$ is small.

**Fig. 3.** Graphical illustration of the influence of the degradation model parameters on the feature trajectory.

B. Prognostics based on Kalman filtering

As the operating condition of a wet clutch may vary, the model parameters in Eq. (10) are not necessarily constant. To accommodate this issue, we have chosen here the flexibility of the Kalman filter having the nature to recursively update the model parameters as measurement data are progressively available. In case of the model parameters are not given, one can also use the Kalman filter as an estimator which is clearly an asset in the context of fast on-line estimation [24]. Unless stated otherwise, in this section we use the following convention: scalars and functions are written in small font, vectors are written in regularly weighted capital font and matrices are written in bold capital font.

1) Kalman filter (KF): The Kalman filter is an optimal state estimator for linear systems perturbed by white Gaussian noise. These systems are called stochastic dynamical systems, where discrete linear form are written as follows:

$$X_k = AX_{k-1} + BU_{k-1} + GW_{k-1} \quad (11a)$$

$$Y_k = CX_k + DU_k + V_k \quad (11b)$$

where $k$, the time index, is such that $X_k = X(t_k) = X(t_0 + kT_s)$ in which $t_0$ denotes the origin of time (generally 0) and $T_s$ is the sampling time interval. Equation (11a) is called the state transition equation. This equation determines how each element of the state vector evolves over time. In the discrete linear case each element of the state vector at time index $k+1$ is expressed as a linear combination of all the elements of the state vector at time $k$ plus some linear combination of the elements of an input vector $U_k$. The matrix $A$ is called state transition matrix and the matrix $B$ is called input matrix. Because in general the states themselves are not directly measurable, Eq. (11b) indicates how the measurements are linked to the system states. More precisely this equation expresses each element of the measurement vector $Y$ as a function of the state vector elements plus some input. This is a static equation. The matrix $C$ is called measurement (or output) matrix and the matrix $D$ is called feed through (or feed forward) matrix. The additional stochastic input $W_k$ and $V_k$ are respectively called process noise and measurement noise. The process noise $W_k$ is a vector of zero-mean independent and identically distributed (iid) random variables (RVs), $G$ is the stochastic input matrix and $\mathbb{E}\{W_jW_k^T\}$ is the covariance matrix of the stochastic disturbance processes such that:

$$\mathbb{E}\{W_jW_k^T\} = \left\{ \begin{array}{ll} Q_W & \text{for } j = k \\ 0 & \text{for } j \neq k \end{array} \right. \quad (12)$$

where $\mathbb{E}\{\}$ is the expectation operator and $Q_W$ is a positive-semidefinite matrix. The elements in $V_k$ are Gaussian random variables expressing the fact that measurements are corrupted by noise. We note $\mathbf{R} = \mathbb{E}\{V_jV_k^T\}$ its covariance matrix. Note that the (auto) covariance of a multidimensional random variable $Z$ is expressed as $\text{Cov}(Z) = \mathbb{E}\{(Z - \mathbb{E}\{Z\})(Z - \mathbb{E}\{Z\})^T\}$ but in the case of zero-mean variable it reduces to $\text{Cov}(Z) = \mathbb{E}\{Z_jZ_k^T\}$. Hence, the Kalman filter aims at providing

1) $\hat{X}_k = \mathbb{E}\{X_k|X_1:k\}$, an estimate of the true state mean at any time $t_k > t_0$ given all the previous measurements $Y_i, i = 1, \ldots, k$;

2) $\text{Cov}(X_k - \hat{X}_k)$ an estimate of the state error covariance given all previous measurement.

This latter quantity, where $X_k - \hat{X}_k$ is the estimator error, can be seen as a measure of the accuracy of the estimate. We
will use it later for RUL estimation accuracy. If the random processes \(W_k\) and \(V_k\) have a Gaussian distribution then the conditional state distribution \(p(X_k|Y_{1:k})\) has also Gaussian distribution:

\[
p(X_k|Y_{1:k}) \sim \mathcal{N}\left(\mathbb{E}\{X_k|Y_{1:k}\}, \text{Cov}(X_k - \hat{X}_k)\right)
\]  

(13)

The algorithm for recursively estimating the state mean and error covariance for linear systems as expressed in Eqs. (11a)-(11b) is called a Kalman algorithm and can be outlined as follows:

- Initialize \(X_0^+\) and \(P_0^+\)
- For \(k = 1, 2, \ldots\) do

\[
\begin{align*}
\hat{X}_k^- &= A\hat{X}_{k-1}^+ + BU_{k-1}, \quad (14a) \\
\hat{P}_k^- &= AP_{k-1}^+A^T + Q, \quad (14b) \\
K_k &= P_k^- C^T \left( CP_k^- C^T + R \right)^{-1}, \quad (14c) \\
\hat{X}_k^+ &= \hat{X}_k^- + K_k \left( Y_k - C\hat{X}_k^- \right), \quad (14d) \\
P_k^+ &= (I - K_k C)P_k^-, \quad (14e)
\end{align*}
\]

where \(Q = GG^T\) is the process noise covariance matrix, \(Q_W\) being defined in Eqs. (12) with \(\mathbb{E}\{W_k\} = 0 \forall k > 0\), \(R = \mathbb{E}\{Y_k\}\) is the measurement noise covariance matrix with \(\mathbb{E}\{Y_k\} = 0\) and \(K\) is the Kalman gain. Note that in the Kalman filter algorithm, system matrices (\(A, B, C, D\)) as well as covariance matrices (\(Q, R\)) may vary over time as long as their values are known. Eqs. (14a) and (14b) are called the prediction equations. They provide the \textit{a priori} state estimate \(\hat{X}_k^- = \mathbb{E}\{X_k|Y_{1:k-1}\}\) as well as \(P_k^- = \text{Cov}(X_k - \hat{X}_k^-)\) the \textit{a priori} estimate covariance. Eqs. (14c), (14d) and (14e) are called the update equations. They provide the \textit{a posteriori} state estimate \(\hat{X}_k^+ = \mathbb{E}\{X_k|Y_{1:k}\}\) as well as \(P_k^+ = \text{Cov}(X_k - \hat{X}_k^+)\) the \textit{a posteriori} estimate covariance. Hence, the Kalman filter uses two sources of information: (1) a dynamic model expressing the way the system evolves over time and (2) sensory data (measurements). The information from these sources is fused in order to infer an optimal state estimate at every time step.

2) Extended Kalman filter (EKF): In our use case we aim to estimate the model parameters which are the Kalman filter states. However, we shall notice that the lockup duration which is measured depends nonlinearly on these parameters. In other words we have a nonlinear measurement equation, which makes the linear Kalman filter algorithm defined above non applicable since it assumes a linear relationship between states and measurements. In this case a simple procedure consists in linearizing the nonlinear equations around the current operating point i.e. the current (last known) state value. Hence the states of a system of the type:

\[
\begin{align*}
X_k &= f(X_{k-1}, U_{k-1}) + GW_{k-1}, \quad (15a) \\
Y_k &= h(X_k, U_k) + V_k, \quad (15b)
\end{align*}
\]

could be estimated using the following algorithm:

- Initialize \(X_0^+\) and \(P_0^+\)
- For \(k = 1, 2, \ldots\) do

\[
\begin{align*}
\hat{X}_k^- &= f(\hat{X}_{k-1}^-, U_{k-1}) \\
\hat{P}_k^- &= A_{k-1} P_{k-1}^+ A_{k-1}^T + Q \\
K_k &= P_k^- C_k^T \left( C_k P_k^- C_k^T + R \right)^{-1} \\
\hat{X}_k^+ &= \hat{X}_k^- + K_k \left( Y_k - h(\hat{X}_k^-, U_{k-1}) \right) \\
P_k^+ &= (I - K_k C_k)P_k^-
\end{align*}
\]  

(16a-16e)

\[
\hat{X}_k^- = f(\hat{X}_{k-1}^-, U_{k-1})
\]

\[
P_k^- = A_{k-1} P_{k-1}^+ A_{k-1}^T + Q
\]

\[
K_k = P_k^- C_k^T \left( C_k P_k^- C_k^T + R \right)^{-1}
\]

\[
\hat{X}_k^+ = \hat{X}_k^- + K_k \left( Y_k - h(\hat{X}_k^-, U_{k-1}) \right)
\]

\[
P_k^+ = (I - K_k C_k)P_k^-
\]

(16a-16e)

with

\[
A_{k-1} = \frac{\partial f}{\partial X} \bigg| _{\hat{X}_{k-1}^-, U_{k-1}},
\]

denoting the Jacobian matrix of \(f(X, U)\) with respect to the states \(X\) evaluated at the current estimated state value \(\hat{X}_{k-1}^+\) and current input value \(U_{k-1}\), and with

\[
C_k = \frac{\partial h}{\partial X} \bigg| _{\hat{X}_{k-1}^+, U_{k-1}},
\]

denoting the Jacobian matrix of \(h(X, U)\) with respect to the states \(X\) evaluated for the current state value \(\hat{X}_{k-1}^+\) and for the current input value \(U_{k-1}\). Note that (i) the algorithmic structure of the Extended Kalman Filter is very similar to the one of the linear Kalman Filter and as for the linear case Eqs. (16a) and (16b) are called the prediction equations and Eqs. (16c), (16d) and (16e) are called the update equations, (ii) the KF is strictly equivalent to the EKF for linear functions if for instance \(f(X_k) = AX_k\) or \(h(X_k) = CX_k\) we have respectively \(A_k = \frac{\partial f}{\partial X} = A\) or \(C_k = \frac{\partial h}{\partial X} = C\). Hence we can very naturally consider systems that are partially nonlinear. For instance the states of a system consisting of a linear stochastic state transition Eq. (11a) and a nonlinear measurement Eq. (15b) would be estimated by the following sequence: [i] Eq. (14a), [ii] Eq. (14b), [iii] Eq. (16c), [iv] Eq. (16d), [v] Eq. (16e). The algorithm handling the inverse situation, i.e. nonlinear state equation and a linear measurement equation are trivial to derive.

3) Prognostics algorithm: For estimation of the degradation model parameters of Eq. (10) in the Kalman filter framework we are going to consider them as states being constants or evolving slowly over time. This is expressed by the following state transition equation:

\[
X_k = X_{k-1} + W_{k-1}
\]  

(18)

with

\[
X_k = \begin{bmatrix} \tau_0, k \\ \alpha_k \\ \beta_k \end{bmatrix}
\]

(19)

and where

\[
\mathbb{E}\{W_k W_k^T\} = Q_W = \begin{bmatrix} q_1 & 0 & 0 \\ 0 & q_2 & 0 \\ 0 & 0 & q_3 \end{bmatrix}
\]

(20)

The structure of the process noise covariance matrix \(Q\) indicates that we assume no correlation between states. This is a linear state equation therefore we can use Eqs. (14a)-(14b) to propagate the states and the covariance. The meaning of Eqs. (18)-(19) is that all parameter estimated at time \(t_k\) is equal to parameter estimated at time \(t_{k-1}\) up to some perturbation. In
reality this perturbation is not a random variable drawn from a Gaussian distribution but rather a step aiming at allowing a variation in the estimation. This variation is necessary to follow underlying variations of the true parameter and more generally to allow a good convergence of the Kalman filter. Note that a change in the true parameter requires convergence of the estimate to the new value. The measurement equation is directly given by Eq. (10) with some additive measurement noise. Therefore we have:

\[ Y_k = h(X_k, t_k) + V_k, \]

with \( Y_k = \tau_k \) being the measured engagement duration feature and

\[ h(X_k, U_k) = X(1)_k + X(2)_k \left( 1 - e^{-X(3)_k U_k} \right), \]

where \( X(1)_k = \tau_{0,k}, X(2)_k = \alpha_k, X(3)_k = \beta_k, U_k = t_k \) and \( V_k \) is the measurement noise whose variance denoted by \( R \). In order to update the states with respect to this nonlinear measurement equation we need Eqs. (16c)-(16b)-(16a).

In order to update the state and pass through the nonlinear function \( \hat{X}_k \) is obtained by a nonlinear combination of state estimations \( \hat{X}_k \), \( \hat{X}_k \) and \( \hat{X}_k \) (i.e. when the feature value crosses the threshold, \( \hat{X}_k \) is determined based on the normalized feature, and to recursively update the model parameters using the last data sample and finally use the updated model parameters (i.e. posterior estimates) to predict the future trajectory of the normalized feature for RUL estimation. Notice that if this prior information is not available, one may also choose arbitrary values for the degradation model parameters because the proposed prognostics framework can also estimate the parameters, as formulated in Eq. (18). However, depending on the chosen initial values, the algorithm may take some time to converge to the expected degradation model parameters.

C. Off-line evaluation metric: \( \alpha - \lambda \) accuracy

To quantitatively justify and evaluate the performance of prognostics algorithms, a performance metric is required. Some metrics tailored for prognostics were proposed in the literature that can be used to effectively evaluate various algorithms [26]. One of the prognostics metrics that has been widely applied is the \( \alpha - \lambda \) accuracy metric, e.g. see Refs. ([27], [28]). In this present work, the \( \alpha - \lambda \) metric is considered for evaluating and comparing the newly developed clutch prognostics algorithm with the early developed one as will be presented in Section VI.

The \( \alpha - \lambda \) accuracy metric compares the ground-truth RUL \( r^{l}_{\lambda} \) to the estimated RUL \( r^l \) with converging \( \alpha \) bounds that provides an accuracy region at specific time instant \( t_{\lambda} \). Similar to Eq. (25), the ground-truth RUL \( r^{l}_{\lambda} \) at the time instant \( t_{\lambda} \) is calculated according to the following equation:

\[ r^{l}_{\lambda} = t_{col} - t_{\lambda}, \]

with \( t_{col} \) denoting the ground-truth end-of-useful lifetime. Depending on the applications, the \( t_{col} \) can be either known \textit{a priori} by human experts or be determined from available historical data. The \( \alpha \) bounds are application specific and expressed as a percentage of ground-truth RUL. The \( \alpha - \lambda \) accuracy, i.e. \( A_{\alpha - \lambda} \), is defined as a binary metric that evaluates whether the prediction accuracy at the time instant \( t_{\lambda} \) falls within specified \( \alpha \) bounds. This can be mathematically expressed as:

\[ A_{\alpha - \lambda} = \begin{cases} 1 & \text{if } r^{l}_{\lambda} [1 - \alpha] < r^l (t_{\lambda}) < r^{l}_{\lambda} [1 + \alpha] \\ 0 & \text{otherwise} \end{cases} \]

With this metric, the performance of prognostics algorithms can be visually evaluated through a graphical representation displaying three plots, namely (i) the ground-truth RUL \( r^{l}_{\lambda} \), (ii) the \( \alpha - \lambda \) accuracy region which is a shaded cone and (iii) the estimated RUL \( r^l \) with the uncertainty interval, in function of time. The closer the RUL estimations and the uncertainty intervals to the \( \alpha - \lambda \) accuracy regions at given time instant \( t_{\lambda} \), the better the prognostics algorithm will be.
V. EXPERIMENTAL METHODOLOGY

To demonstrate the effectiveness of the novel prognostics algorithm, historical data were acquired from two clutch packs subject to accelerated life tests (ALTs) using a fully instrumented SAE#2 test setup, as discussed below.

A. SAE#2 test setup

The SAE#2 test setup used in the present work, shown in Fig. 4, is different from the one used in our previous work, cfr. [17], [16], [29]. The current SAE#2 test setup is more compact and allows us to apply higher energy to a clutch pack.

The setup consists of three main sub-systems, namely driveline, control and measurement system. The driveline is constructed of several components: an AC motor for driving the input shaft (1), input velocity sensor (2), input flywheel (3), clutch pack (4), torque sensor (5), output flywheel (6), output velocity sensor (7), an AC motor for driving the output shaft (8), a hydraulic system (11-20) and a heat exchanger (21) for cooling the outlet ATF. An integrated control and measurement system (22) is used for controlling the ATF pressure (both for lubrication and actuation) to the clutch and for the initial velocity of both input and output flywheels as well as for measuring all relevant dynamic signals. The algorithm used to control the ATF pressure in the tests is the same as the one used in practice.

B. Test specification and procedure

Two sets of experiments with different clutch packs (i.e. hereafter called clutch pack A and B) were conducted. The two clutch packs are identical, but the friction material is different for each clutch pack. The lining materials of the friction discs are paper-based type, while the separator discs are made of identical steel material.

Paper-based materials are porous composites, basically containing some ingredients such as fibers, solid lubricant and friction modifiers, which are all saturated and cured at certain pressure and temperature, with a thermosetting resin acting as a binder [30]. These materials are favorable because of their good frictional characteristics and low production cost. The used separators are made of plain steel with the roughness $\sigma_{Ra}$ of less than 16 µin (= 0.41 µm) in order to prevent excessive abrasive wear on the friction material, as recommended by Fish and Lloyd [31]. Notice that all the friction discs, separator discs and ATF used in our experiments are commercially available on the market. The test specification is summarized in TABLE I.

| Number of friction discs in the clutch assembly [-] | 10 |
| Inner diameter of friction disc ($d_i$) [mm] | 99 |
| Outer diameter of friction disc ($d_o$) [mm] | 133 |
| ATF | Texamatic 7045 |
| Lubrication flow [liter/minute] | 30 |
| Inlet temperature of ATF [$^\circ$C] | 85 |
| Inertia of input flywheel [kgm$^2$] | 5.4 |
| Inertia of output flywheel [kgm$^2$] | 2.7 |
| Sampling frequency [kHz] | 12.8 |

In the ALT, the amount of energy applied to each clutch pack is relatively high compared to normal operating condition. Here, the energy level is adjusted by means of changing the inertia and/or initial rotational speed of the two flywheels. Fig. 5 shows representative measured signals a clutch duty cycle. Initially, both input flywheel (drum-side) and output flywheel (hub-side) are rotating at the same speed of 2,300 rpm, but in opposite direction. So, the relative rotational speed is 4,600 rpm. The two electric motors are powered-off and the pressurized ATF is simultaneously applied to a clutch pack at the time instant $t_f$. The oil thus actuates the clutch piston, pushing the friction and separator discs towards each other in the filling phase between the time instants $t_f$ and $t_e$. While the applied pressure increases, contact is gradually established between the separator and friction discs which builds up the transmitted torque and a simultaneous decrease of the relative rotational speed. Finally, the clutch is completely engaged when the relative rotational speed reaches zero at the lockup...
time instant $t_l$. As the inertia of the output flywheel is higher than that of the input flywheel, after $t_l$, eventually both flywheels rotate together in the same direction of the output flywheel. In order to prepare for the forthcoming duty cycle, both driving motors are braked at the time instant $t_b$, such that the driveline can stand still for a while.

It should be mentioned that before a clutch pack is subject to an ALT, a run-in test with lower energy level is carried out for 100 engagement cycles in order to stabilize the contact surface. The run-in test procedure is similar to the ALT procedure, but the initial relative rotational speed in the run-in test is lower than that of ALT. The ALT procedure discussed above is continuously repeated until the clutch has reached the end-of-useful-life, see Fig. 6. Notice that the used ATF is continuously cooled and filtered during all the tests, so it is reasonable to assume that the ATF remained fresh.

**VI. RESULTS AND DISCUSSIONS**

Measurement data obtained from the ALTs on clutch pack A and B were processed to compute the normalized engagement duration feature $\tilde{\tau}$ according to the procedure described in Section II. Figure 7 shows the evolutions of the feature obtained from the two ALTs, where the gray dashed-lines point to the ground-truth end-of-useful-lifetime $t_{eol}$. According to an experienced human operator, the ground-truth end-of-useful-lifetimes of clutch pack A and B are of 9,200 and 9,400 engagement cycles, respectively. Based on these ground-truth end-of-useful-lifetimes, the corresponding end-of-useful-life thresholds $r_{\ellag}$ for clutch pack A and B can be determined and the ground-truth RULs $r_{\ellag}$ for the two clutch packs can then be calculated at arbitrary duty cycles according to Eq. (27). Note that the ground-truth RULs $r_{\ellag}$ will be used later on as benchmarks for quantitatively evaluating the performance of the WMS and EKF based prognostics algorithms.

The variations of the averaged ATF outlet temperature $T_{outlet}$ and averaged pressure $\bar{p}$ at various duty cycles are shown in Fig. 8. It can be seen in the figure that the averaged outlet temperature $T_{outlet}$ obtained from the ALT of clutch pack A tends to increase ($\Delta T_{outlet} \approx 2$ °C), while it tends to decrease during the ALT of clutch pack B ($\Delta T_{outlet} \approx -4$ °C). The averaged pressure $\bar{p}$ obtained from the ALTs of clutch pack A and B have similarly increasing trends, with the respective average pressure increase of $\Delta \bar{p} \approx 0.4$ bar and $\Delta \bar{p} \approx 0.6$ bar.

To demonstrate the capability of the WMS and EKF based prognostics algorithms, the past trajectory estimations and the
Fig. 9. Estimation of the past trajectory and prediction of the future trajectory of the normalized engagement duration feature $\tilde{\tau}$ at 1,000th and 6,000th cycle. (a) and (b) with the WMS algorithm and the EKF algorithm on the clutch pack A respectively; (c) and (d) with the WMS algorithm and the EKF algorithm on the clutch pack B respectively.

Future trajectory predictions of the normalized feature $\tilde{\tau}$ at the 1,000th and 6,000th cycle of clutch pack A and B are shown in Fig. 9. Unlike the WMS algorithm, the EKF algorithm smooths the past trajectory and predicts more accurately the future trajectory as clearly shown in the figure.

Fig. 10 shows how the parameters of both the WMS and EKF based prognostics algorithms vary over time. The main causes of the ‘local variations’ on the four parameters ($b_w$, $\tau_0$, $\alpha$, and $\beta$) are due to the operating condition changes during the ALTs, namely temperature and pressure variations. As a result of the operating condition variations, some uncertainty on the degradation mechanism of the clutch packs, e.g. the degradation rate, may be introduced. In addition, the uncertainty in the mechanical properties and the surface profiles of the used paper-based friction materials, inherited from their nature as composite materials as discussed in Sections I and V, can also introduce some variations to the tribological characteristics of the two clutch packs that in turn affect the model parameters ($b_w$, $\tau_0$, $\alpha$, and $\beta$).

The parameter trajectories of the WMS based prognostics algorithm $b_w$ for clutch pack A and B are quite different in the beginning of the clutch lifetimes (i.e. before around 2,000 duty cycles). Afterward, this parameter is prone to decrease as the number of duty cycles increases for the two clutch packs as shown in Fig. 10. The reduction of the WMS parameter $b_w$ indicates the asymptotic evolution of the feature value.

In the context of the EKF based prognostics algorithm, the trajectories of the $\alpha$ parameter are quite stable for clutch pack A and B. On the other hand, the trajectories of the $\tau_0$ and $\beta$ parameters are different for the two clutch packs. Besides the operating condition variations, we believe that the differences in the parameter trajectories are probably originating from different compositions of the friction materials used in clutch pack A and B.

Fig. 11 shows the RUL histograms of the two clutch packs obtained with the EKF based prognostics algorithm that are calculated at the 1,000th and 6,000th cycle as representative results. Due to the nonlinearity of the degradation model (see Eq. (10)), the resulting distribution is slightly skewed yielding to asymmetric uncertainty around the estimate. In the following paragraph, the uncertainty intervals of the RUL estimation at each duty cycle for the two clutch packs are presented.

The $\alpha - \lambda$ performance metric discussed in Section IV-C is calculated for quantitative evaluation and comparison of the performances of the WMS and EKF based prognostics algorithms. The accuracy bound of 20% is used in this paper (i.e. $\alpha = 0.2$). Fig. 12 shows the ground-truth RULs, the estimated RULs together with the uncertainty intervals using the EKF and WMS based prognostics algorithms, and the $\alpha - \lambda$ accuracy regions, for clutch pack A and B. Because of its nature as data-driven technique, the WMS based algorithm is only able to start doing prediction after some measurement data samples are available (in this case after 600 cycles). In
contrast, the EKF based prognostics algorithm can already start doing prediction since the beginning.

It is seen in Fig. 12 that the EKF based RUL estimations stay within the $\alpha - \lambda$ accuracy region, which is in contrast to the WMS based algorithm. Hence, it is reasonably to state that the EKF based prognostics algorithm outperforms the WMS based prognostics algorithm. Moreover, the figure also shows that the WMS based algorithm under estimates the ground-truth RULs. In practice, this underestimated RUL is not desired as it can lead to non-optimal maintenance actions.

VII. CONCLUSION AND OUTLOOK

In this paper, we have addressed the problem of remaining useful life (RUL) estimation of wet friction clutches used in vehicles equipped with automatic transmissions. The prognostics algorithm developed in this work extends the capability of the pre-lockup feature based condition monitoring method developed in the previous work. Here, the feature being defined as the engagement duration can be extracted from sensory data typically available in wet clutch applications (e.g. input and output shaft rotational speeds and applied pressure). The present prognostics algorithm is based on the extended Kalman filtering technique where a degradation model proposed in this paper is fused with measurement data under the Bayesian inference framework.

New history data of different commercially available wet clutches subject to accelerated lifetime tests (ALTs) using a fully instrumented SAE#2 test setup were acquired and processed both with the newly and early developed prognostics algorithms. The early developed prognostics algorithm [16] is based on weighted mean slope method (WMS), where the future trajectory of the feature is modeled under assumption of quasi-linear trend. The experimental results demonstrate that the EKF based prognostics algorithm outperforms the WMS based prognostics algorithm.

So far, the developed prognostics algorithm has been experimentally validated based on historical data obtained from the ALTs using the SAE#2 test setup under controlled operating conditions, where the variations of some operational parameters namely (i) the oil temperature, (ii) actuation pressure, (iii) initial relative rotational velocity and (iv) input/output torques, are relatively small. As theoretically shown in our previous study [18], the feature response is influenced by these operational parameters. In order to achieve an accurate RUL predictions for wet clutches, the variations of the operational parameters should be kept as small as possible. For practical implementation, this can be realized by embedding some control and acquisition related-routines into the Electronic Control Unit (ECU) that can satisfy the aforementioned requirements. As a concrete proposal, all the relevant signals for the feature calculation are only acquired by the ECU when the averaged values of the measured operational parameters are in predetermined ranges.

In order to extend the developed prognostics method into more generic situations, where all relevant signals can be measured at any conditions, the effects of the operational
parameters on the pre-lockup feature need to be further investigated in the future work. Profound understanding of the effects of the operational parameters on this feature may allow us to extend the degradation model proposed in this paper such that the variations of the operational parameters can be taken into account. This way, we believe that a more accurate prognostics and health management system for wet clutches can be achieved.

ACKNOWLEDGMENT

The authors are grateful for the experimental data provided by Dr. Mark Versteyhe of Dana Spicer Off Highway Belgium. Valuable suggestions and insightful comments of Dr. Tinne De Laet on the state estimation part using the extended Kalman filtering are appreciated. This work was performed within the frame of the Top Competence research project of the Smart diagnostics systems research program of Flanders’ Make.

REFERENCES


