A NOVEL FRAMEWORK TO LINK PROGNOSTICS AND HEALTH MANAGEMENT AND PRODUCT-SERVICE SYSTEMS USING ONLINE SIMULATION

ABSTRACT

Product Service Systems (PSS) and Prognostic Health Management (PHM) have so far been researched individually in different domains and as unrelated research theme. However, to guarantee the availability of the asset, which is a typical demand in some PSS contracts, it is fundamental for PSS providers to be able to properly manage the asset’s lifetime variability in order to avoid unscheduled downtimes and contract penalties. This paper describes part of a research project to investigate how PHM can support more effective fulfilment of some PSS contracts. In particular, this paper aims to present a novel framework to link PHM and PSS using online simulation. The paper also presents a prototype of the online simulation model and three experimental cases comparing the outcomes of the online simulation model against those obtained from the traditional simulation model.

Keywords: Product-Service System, Prognostics and Health Management, Online simulation, Dynamic behaviour
1. INTRODUCTION

Product-Service System (PSS) has emerged as a new manufacturing strategy whereby manufacturers shift their business model from selling ‘only product’ to selling ‘an integrated of product and services’ (Tan et al, 2009; Mont, 2002). In PSS, customers are typically no longer buying the product (asset) but instead buying the availability or capability of that asset (Cook et al, 2006). In PSS, the Original Equipment Manufacturers (OEMs), typically, do not offer product ownership but an integrated solution combining product and service provision. OEMs are therefore responsible for managing the asset’s lifetime, spare parts and maintenance services in order to maximise the asset’s performance and to minimise the operational cost (Spring and Araujo, 2009). Enabling technologies, such as Prognostics and Health Management (PHM), are often needed to manage the asset’s lifetime and to execute the appropriate maintenance activities to avoid catastrophic failures or unscheduled downtimes (Khalak and Tierno, 2006).

Fault diagnosis and fault prognosis functions (typically found in PHM technology) are often integrated in many high value products or assets to open up new service opportunities (Neely, 2007). Using those technologies, the companies can capture unforeseen performance variations during the asset’s lifetime. This is particularly useful in high-tech sectors, e.g. aerospace, military, etc., where operational costs have become a significant source of revenue generation for the PSS providers. Indeed, an effective PHM helps increase the asset’s reliability and safety by using a concept known as autonomic logistics (Banks et al, 2005). This concept considers a PHM system as an autonomic response or like a subconscious reflex of human being’s nervous systems to react automatically to an unforeseen potential problem.

The Rolls-Royce’s TotalCare® is one of the most widely reported cases of the integrated PSS and PHM. The TotalCare® is considered as “a flexible approach to achieving an engine support service that has the correct fit and scope of services to meet the operator’s specific needs”¹. In this case, not only does Rolls-Royce supply the aero engines, they also use the engine health monitoring to make an advance fault prediction in order to avoid the expensive cost of operational disruption and unnecessary engine repairs (Roll-Royce, 2010). PHM is also reported to be the backbone of the maintenance and logistics management for the Joint

Strike Fighters (JSF) (Malley, 2001). In this case, PHM components aid the logistic policy change of the US Department of Defense (DoD). Another example is the method to perform PSS for the machine tool proposed by Zhu et al (2011). In the Industrial Product-Service System for CNC machine tool (mt-iPSS), the capability of the machine tool and its attachment are sold as an intangible service. The business core is to provide machining capability rather than product ownership. In this case, the service provider, usually, retains the maintenance programme ensuring machine utilisation over a given period of time (Azarenko et al, 2009).

Jazouli and Sandborn (2010) proposed a new method to determine an unknown system attribute to fulfil a specific availability constraint. They demonstrated that the use of PHM technology to increase the availability of an asset can provide a value beyond failure avoidance and minimisation of cost. Greenough and Grubic (2011) investigated the use of Condition-Based Maintenance (CBM) in ‘servitization’ contracts. Using the tools they developed, i.e. machine tool health simulation (MATHS) and life-cycle risk evaluation of machine tool health (LIKEMATH), they demonstrated considerable benefits from PHM technology gained by the OEMs in terms of asset’s availability and utilisation.

Even though existing literature has revealed substantial benefits offered by PHM technology to support operational decisions, very few of them actually considered the effects of dynamic behaviour to the asset’s performance in their analysis. The dynamic behaviour concerns operational and environmental disturbances that can influence the expected asset’s lifetime and the scheduled performance goals. One characteristic of dynamic behaviour is that it cannot be forecast with some degree of certainty. Under the PSS contract where the ownership of the asset remains within the OEMs, the uncertainty is also transferred from the customers to the OEMs, and consequently there is an ever increasing need for the OEMs to better manage the uncertainty especially in the case of availability contracts (Erkoyuncu et al, 2009). In addition to that, there is a great deal of risk that the contractual targets defined at the PSS design phase cannot be achieved when the assets are severely affected by the dynamic behaviour.

Generally, effective design of a complex system is a challenging task, simply because the system designers usually do not have early visibility of the system performance. For this
reason, simulation modelling has traditionally been used to aid the validation of the systems (or sub-systems) design in order to improve the overall system performance (Ingemansson et al., 2002). Simulation modelling can be intuitive and visually appealing to use to support decision making (Benedettini and Tjahjono, 2009), and in this instance, simulation modelling can also be considered as a tool to support PSS design. As pointed out by Phumbua and Tjahjono (2011), however, simulation modelling tools are usually employed only during the design phase of PSS. Furthermore, in traditional discrete-event simulation models, an asset’s downtime uncertainty is only represented by random variability (Banks et al., 2009). During the execution of the PSS, which is typically in the form of contractual agreement, various unpredicted events, such as machine overload due to extreme weather conditions and improper use, can affect the asset’s expected lifetime. The impact from these unpredictable events, unfortunately, cannot be properly modelled with random variability and Mean Time Between Failure (MTBF) information alone. Any maintenance activities that do not consider modifications in the operational environment lead to many unexpected system and component failures (Markeset and Kumar, 2003).

As there is no direct coupling between the simulation model and the actual systems, the model usually runs from an initial or empty state. As the model reaches the steady state, experimentation can progress and various what-if scenarios can be performed. Nonetheless, it is very common that this state does not necessarily correspond to the state which the actual system is currently at, and it is challenging to bring the model to this initial state. One way to address those issues is by coupling the simulation model with the actual system in an online mode. Online simulation, or real-time simulation as it is also known, uses some kind of feedback mechanism to couple the model and the actual system. It adopts similar principles used in real-time control systems where the simulation model is applied as a feedback loop. One distinct feature of the online simulation is that the parameters obtained from the actual system become a set of current state parameters of that system which in turn will be used to initialise the simulation model. This feature consequently enables the simulation model to be used not only during the design phase of PSS but more importantly can be extended to the operational phase where the simulation model can now be used as a day-to-day operational tool.
The main aim of this paper is to propose a novel framework to link PHM and PSS technologies using online simulation. The paper is organised into six sections. Section 2 presents a brief literature review of the research domain. Section 3 presents the proposed framework and the detailed explanation on its modules. Section 4 presents the framework implementation. Section 5 describes the three case studies and their respective experimental results. Finally, Section 6 lists the concluding remarks and future work.

2. LITERATURE REVIEW

2.1. Prognostics and Health Management

Prognostics and Health Management (PHM) is considered as key to improved safety and reliability of components through an intelligent and autonomous detection, and isolation of fault to estimate the Remaining Useful Life (RUL) of such component with the aim to reduce the operational and support cost (Shen et al., 2010). This enabling technology shows the potential to address reliability problems that have been manifested due to the complexities in design, manufacturing, environmental and operational use conditions and maintenance (Pecht and Jaai, 2010). Manufacturers have been often driven to embrace a PHM programme in order to improve their product performance, improving availability, improving maintenance efficiency and effectiveness and differentiating from their competitor’s products (Grubic et al., 2011). In other words, the main goal of a PHM technology is to provide the most up-to-date health status of the assets in order to support proactive actions to improve the system’s performance.

The PHM programme, typically incorporates data acquisition, data processing, diagnostic and prognostic disciplines. During the data acquisition, sensor technology captures one or more performance conditions (for example, temperature, vibration, shock, pressure, etc.) of the critical component. Complex systems often require several parameters to be monitored in the whole asset’s lifetime to provide the information entailed by PHM programme (Cheng et al., 2010). A pre-processing stage is also needed to clean and to filter the data for further data analysis. In the next stage, the filtered data must be transformed and manipulated. Numerous models, algorithms and tools for data analysis (time domain analysis, frequency domain analysis, time-frequency domain analysis etc.) exist in the current literature. Afterwards, a
fault diagnosis technique determines the condition of the system or the critical component based on processed data. The fault diagnosis procedure is a method for detecting, isolating and identifying a failure condition of a system, while its critical components are operating even though they are in a degradation mode (Caesarendra et al, 2010). Statistical, Artificial Intelligence and Model-based approaches are the most common methods for fault diagnostic purposes (Jardine et al, 2006).

Based on fault diagnosis analysis, the fault prognosis technique estimates the RUL. The RUL information can be used to plan, in advance, the entailed maintenance activities (Banjevick, 2009). The accuracy of its estimation depends on what prognostic technique was employed. Prognostic techniques can be grouped into physics-based, trend-based evolutionary and experience-based (Tran et al, 2008; Muller et al, 2008). According to Vachtsevanos et al (2006), physics-based models remain the most accurate but expensive as they provide a straightforward method to calculate the damage of the critical component based on operating conditions and to assess the cumulative effects in terms of component life usage. In addition, model-based approach can be robust to sensor loss and still work under limited sensing environment with an accurate model (Daigle and Goebel, 2010).

When a direct relationship amongst fault evolution mechanism and the critical component operating conditions is hard to find, the trend-based evolutionary approach can be adopted. This modelling technique is often used when there is a known trend in terms of fault evolution and operating conditions. A data-driven model is considered as the most applicable for trend-based approach. It uses a learning mechanism procedure (e.g. supervised or unsupervised training) to empirically understand the relationship between inputs and outputs without considering an explicit mathematical relationship. Neural networks (Wang et al, 2004) and fuzzy logics (Al-Najjara and Alsyouf, 2003) are the most popular data-driven techniques to estimate the RUL in the literature.

Sometimes, there is not enough information to apply the trend-based evolutionary or physics-based model, especially when the critical components are not monitored by sensors or when the failure is intermittent. In this case, the failure information can be obtained from the manufacturer’s component data sheet. In the experience-based approach, historical data is collected to fit the probability distribution and MTBF information. Although it is the simplest
approach, it can be used to drive preventive maintenance practices that can then be updated at regular intervals (Byington and Roemer, 2009). Despite the advancement of technology, the means in which the assets may fail and the operation principles of critical components are too specific and can vary according to particular component characteristics (e.g. material, physical conditions, statistical correlations, etc.). As a consequence, designing a generic framework to implement PHM solution is challenging and although prognostic techniques can provide the estimation of Remaining Useful Life (RUL), its implementation will always be application-specific (Roemer et al, 2005).

2.2. Reliability definition and concepts

Reliability, also known as survival function, is related to the ability of an item to perform a required function under a given environment and operating conditions and for a stated period of time (NIST/SEMATECH, 2011). The reliability accounts whether or not an item performs at or above a specific standard, how long it can perform at that standard and the conditions under which it operates (Hamada et al, 2008). The reliability calculation aims to verify if such a system, working under a set of operational conditions, can survive until a specified time. It becomes an important evaluation criterion because the estimation of RUL itself does not provide enough information to support the decision maker (Feldman et al, 2009). Reliability can therefore be mathematically defined as:

\[
R(t) = P(T > t) = \int_{t}^{\infty} f(\xi) \, d\xi
\]  

(1)

where:

- \( R(t) \) = Reliability function
- \( T \) = Time to failure
- \( t \) = Time until the system has survived
- \( f(\xi) \) = Probability density function

The integral in the Equation 1 is only defined when \( T \) is continuous and its density function and hazard function exist.

---

7
The abovementioned formula depicts a mathematical relationship between reliability and Time To Failure (TTF). TTF is a random variable influenced by the operational and environment conditions that can vary during the whole component’s lifetime. For this reason, in practice, conditional reliability is preferred rather than reliability. Indeed, the conditional reliability function (R) contains all information required for prediction and planning of future activities that are depending on RUL (Banjevic, 2009).

Considering that such a component has survived until the current time (s), mission reliability is the conditional probability that the component will survive until future time (T). Splitting the remaining time to failure \(X_r = T - t\) i.e. the Remaining Useful Life (RUL) of the component as small as possible \(X_r = n \cdot \Delta t\, |\Delta t \to 0; n \to \infty\), each time step can be seen as mission reliability and treated as series across the overall interval. Thus, the conditional reliability obtained from the entire interval can be expressed as (Lu et al, 2001a):

\[
R(X_r|T > t) = \prod_{i=1}^{n} \Pr [X_{ri}|T_i > t_i]
\]  

(2)

where:

- \(R(X_r)\) = Conditional reliability
- \(T\) = Future time
- \(t\) = Time until the system has survived
- \(X_r\) = Remaining useful life of such component

3. FRAMEWORK

The design of the online simulation framework was based on the closed-loop control system (Ogata, 2009), where the asset performance degradation data is captured, manipulated and propagated (as a feedback signal) for each framework module with the ultimate goal to support operational decisions during PSS contract execution. Figure 1 shows a simplified block diagram representation of the proposed framework.
The framework is made up by the following modules:

- Prognostics and Health Management Module
- Reliability Estimation Module
- Operational Simulation Module
- Analysis and Decision Support Module

3.1. Prognostics and Health Management Module

PHM technology provides the prediction of future behaviours (based on the current system state which supports timely maintenance actions) in order to maintain an acceptable level of asset’s health degradation (Vachtsevanos et al., 2006). Furthermore, this technology offers an intelligent way to detect, to isolate faults and to estimate the RUL so as to improve the asset’s safety and reliability (Shen et al., 2010). In order to provide asset’s health performance degradation data, the PHM module must implement the functionalities supported by PHM programme (e.g. conditional monitoring, fault diagnosis, fault prognosis, etc.). Conditional monitoring (CM) and fault diagnosis (FD) functions require continuous comparison between the current performance degradation data and operating boundaries. Besides, additional evaluation made by FD function matches current health degradation level into known health states. Using the CM and FD outputs, fault prognosis function estimates the RUL of the asset.

Even though the PHM module implements several functionalities which require particular methods and algorithms, a detailed explanation about them is out of scope of this paper. Further information can be found in Vachtsevanos et al. (2006) and Jardine et al. (2006). Indeed, from the framework viewpoint the PHM module supplies conditional monitoring data and event data related to the asset’s health performance degradation. In other words, it continuously monitors the asset’s performance conditions and converts them into performance degradation data or, possibly, into RUL estimation. Critical levels are also embedded into PHM data to guide reliability estimation and maintenance decision making. Moreover, a
dedicated communication interface is also needed to promote data exchange between PHM module and the reliability estimation module.

### 3.2. Reliability Estimation Module

The Reliability Estimation Module (REM) predicts the health of the assets in terms of their reliability using the latest PHM data. It determines the likelihood of failure based on a predefined operating time period. Current asset’s performance data are measured and converted into reliability estimation using thresholds defined beforehand (Xu et al., 2009). Based on the asset’s data and operating boundaries, the REM estimates what the *performance reliability* will be. The performance reliability, also known as mission reliability can be defined as the conditional probability from such component/system to execute its function effectively for the next time frame based on current degradation level (Lu et al., 2001a).

Once the current asset’s performance degradation status has been updated into the REM internal buffer, it estimates the performance degradation trend and its error variance for $k$-steps-ahead. The number of steps-ahead is related to the completion horizon specified on the operational simulation model. Completion horizon is an explicit rule which defines when the simulation engine must interrupt the simulation execution (Wynn et al., 2008). It can be set on the simulation model as a function of jobs completed, absolute time duration or asset failure. It can also be compressed and relaxed accordingly. Using the estimation outcomes and operating boundaries, the conditional reliability of the asset can be calculated. Indeed, prediction of asset’s residual life and their variances based on operating and environmental conditions has inherently better accuracy than the constant MTBF information (Tu et al., 2007).

The selection of the estimation technique to implement the REM kernel entails striking a balance between technical constraints, computational power, desired estimation accuracy, etc. Moreover, due to inherent uncertainty around the asset’s measured data, the candidate to implement REM kernel must be selected from the stochastic estimators group (such as Kalman Filter (Lu et al., 2001b), Particle Filter (Xu et al., 2009), Holt-Winter smoothing (Xu et al., 2009; Lu et al., 2001a), etc.). For instance, the Kalman Filter is considered to be the most appropriate approach for assets with linear failure mode (and a known mathematical
representation) and a measurement error which is distributed normally. Nonetheless, for non-linear failure mode with non-normally distributed measurement error, Particle Filter is preferable. Holt-Winter’s exponential smoothing is typically suitable for stochastic dynamic failure mode with continuous degradation which can be represented by time series.

3.3. Operational Simulation Module

The Operational Simulation Module (OSM) is an online simulation model of the assets covered by the PSS contract. Each element of the model is an asset’s representation (i.e. machines, service managers, technicians, data acquisition, etc.) with attributes and behaviours typically found in manufacturing systems. Furthermore, asset’s model assisted by PHM programme has their time to failure obtained from the REM outputs rather than probability distributions and MTBF information. As a consequence, OSM outcomes will become more reliable and sensitive to the effects of dynamic behaviours, which enable the maintenance team to react in time due to the unforeseen circumstances. Figure 2 illustrates how performance degradation data can be used to update the simulation model.

[Insert Figure 2 here]

Figure 2(A) shows an expected and current asset’s health performance curves. The expected curve is often obtained from OEM’s historical data whereas the current curve is taken from PHM data. A deviation between the expected curve and the current curve indicates an unexpected asset’s usage time variation \( t_1 - \text{current failure time}, \ t_2 - \text{expected failure time} \).

If this variation has not been considered by the maintenance service team, it can affect the fulfilment of the agreed asset’s performance, leading the OEM to possible contract penalty. In fact, the only way to compute MTBF information to be equivalent to the asset’s service life is to wait until the entire sample population of assets has achieved their end-of-life (Torell and Avelar, 2004) which is, clearly, unfeasible for proactive reaction purpose. Indeed, conditional monitoring data can provide information on current working age and state of the system measured by some diagnostic variables and also the environment conditions that may affect its future life (Banjevic, 2009).
Forecasting strategy is the main function supported by OSM. Using this function, service providers can test a particular maintenance regime for each individual asset. In order to do that, the asset’s performance degradation curve can be mapped into four distinct zones (Hess, 2002). In the first zone (ZN1), the asset operates under a normal operating period and constant hazard rate. Moreover, no service action is needed during this asset’s operating time. Second zone (ZN2) is reached when the asset begins to wear out. The uncertainty around reliability estimation often prevents the asset’s repair at this stage. Nonetheless, some operating tasks (such as spare part acquisition, labour allocation, etc.) can be done in advance which considerably reduce the asset downtime. Once performance degradation reaches Zone 3 (ZN3), the reliability estimation is at its highest confidence level and the repair tasks may start. If the asset’s repair does not start until the Economical Removal Point (ERP), then the asset’s performance degradation will reach Zone 4 (ZN4) leaving it to an imminent risk of failure.

Figure 2(B) shows the relationship between the forecasting strategy and the simulation execution. The forecasting strategy function (FSF) also enables short completion horizon extension using adaptive time window and reliability bounds supervision. For each asset’s conditional reliability estimation, the FSF checks whether the obtained estimation is within the reliability bounds. Once this condition is checked out, the completion horizon of the OSM is extended which enables reliability estimations for more steps-ahead until the estimation achieves the lower bound limit. In order to avoid rapid uncertainty increase around reliability estimation for large completion horizon, the confidence function implemented into OSM often checks the error variance of the mission reliability estimated. If the estimation confidence is less than the expected threshold defined beforehand, the simulation engine will interrupt the simulation execution, and warns the user for the loss of confidence for further simulation outcomes.

The execution mode function provides a powerful control over the simulation execution. The simulation engine can run the OSM either in a continuous execution mode or a step execution mode. In the former case, the simulation engine automatically restarts the execution of the simulation model for each new available condition data. As a consequence, the user does not have to initialise the simulation model every time the OSM is interrupted. Nonetheless, in this execution mode the simulation model cannot be rolled back for further analysis. In the latter
case, the simulation engine runs the model until the stoppage condition. Once the simulation stops, the user can modify the simulation parameters before the next simulation execution. In order to maintain synchronisation between the simulation model and the current asset’s condition, the OSM prevents the rollback of the simulation model before the current time.

3.4. Analysis and Decision Support Module

The Analysis and Decision Support Module (ADSM) determines the proactive actions that must be executed based on the analysis of the simulation outcomes. In particular, the decision support function must take into account a systematic generation of alternatives to support more effective and efficient decisions even in complex situations (Yam et al., 2001). Most often, the decision support mechanism evaluates the possible alternatives using one-objective or multi-objective criteria to come up with the optimal decision. When multi-objective functions are specified (e.g. production cost minimisation, cycle time reduction, maximum reliability and quality, etc.), the search for an optimal solution becomes hard and can only be solved using optimisation search techniques, such as Newtonians methods, genetic algorithms, simulated annealing, etc. In fact, when the objectives are potentially conflicting, the selection of one alternative that best satisfies the decision maker involves the elicitation of various kinds of preferential information from the decision makers and in their applications to the decision making process (Iyer et al., 2006).

Following the analysis of the simulation results, the ADSM will provide the desired information to support the service supplier decision making. The decision support will result in modifications of maintenance strategies, e.g. modification on capacity (work force, staffs, etc.), facilities (tools, equipment, spares, workforce and location of workforce), etc. Within the business domain, it leads to changes of practice (such as additional training to operate the new equipment, review of the business requirements, etc.), additional services and better value provision for both the end-users and the manufacturers. Some of these actions will directly or indirectly affect the asset’s lifetime.

4. FRAMEWORK IMPLEMENTATION
At this stage, the focus is on the PHM technology and Online Simulation. Figure 3 shows the overall framework implementation scheme.

[Insert figure 3 here]

4.1. PHM technology

PHM technology comprises all the PHM modules for the purpose of asset health monitoring. In addition, each PHM module monitors a particular critical component and also implements a fault diagnosis or fault prognosis technique according to the component’s failure mode. Even though the PHM module implements distinct fault diagnosis or fault prognosis methods, those modules must always deliver normalised PHM data (which can be asset’s health degradation data or asset’s remaining useful life), where the asset’s data range varies from 100% to 0%. Those data are captured and used further to update the asset’s model assisted by PHM technology.

In the framework-based simulation model, those assets assisted by PHM technology are periodically updated to the real asset’s data. The simulation model will be updated with the field asset’s data thus improving the model representation and increasing the reliability of simulation outcomes. Whenever an asset’s health critical degradation limit is achieved, the simulation model notifies the ADSM module and triggers the operational decisions before the real asset’s fails.

4.2. Online simulation

The online simulation is a computer-based simulation environment where the model can be implemented and executed accordingly. In this computer-based simulation environment, PHM output data becomes input data to the simulation model which provides adaptive analysis over the current asset’s operating environment. In contrast to the traditional simulation methods where the focus is on long-term, steady-state behaviour, the online simulation focuses on the short-term operational decisions to control a deployed business process in response to contextual change or unforeseen circumstance (Wynn et al, 2008). Indeed, a real-time simulation environment provides dynamic input data to a simulation
model promoting an adaptive reaction due to the changes in operating environment and, it potentially improves the accuracy of the simulation outcomes (Song et al, 2008).

The online simulation environment was developed using Anylogic®, a java-based simulation tool. Anylogic® is a multi-modelling and simulation software tool based on the UML-RT paradigm. This multi-modelling simulation software supports highly sophisticated model based on distinct modelling paradigms, which include discrete-event simulation, system dynamics and agent-based simulation. Consequently, simulation elements might have custom functions running out of the simulation engine control which is needed to develop the Online Simulation functionalities.

4.2.1. **Standard asset model**

The adoption of fault diagnosis and fault prognosis functions requires addressing a range of challenges and developing a set of capabilities that relate to a business and cultural domain rather than to advances in technology (Grubic et al, 2011). Those technologies are particularly crucial when unplanned downtime on high value assets can lead to total system failure that is obviously costly and hazardous to the environment or even life-threatening (Tu et al, 2007). As a consequence, in a typical manufacturing environment, some assets that are not aided by a PHM programme will unavoidably interact with those with PHM-associated technology. For this reason, an asset’s model where breakdowns are driven by probability distribution and OEM’s information is needed even in an online simulation model.

The standard asset model (SAM) does not receive an asset’s field data to update the asset’s model. Furthermore, it does not have a PHM associated module or even a reliability estimation module. In SAM, downtime and repair time are fully based on deterministic and random time. Fixed time to failure or time to repair might be appropriate when the downtime is due to preventive maintenance (i.e. on a fixed schedule), otherwise, it should use random time (Banks et al, 2009). Random time is, typically, based on probability distributions and OEM’s information. Exponential and Weibull probability distributions are often used to determine an asset’s breakdowns.
4.2.2. Customised Asset Model

The Customized Asset Model (CAM) is an abstract representation where the asset model is updated using PHM data. In this model, asset’s breakdowns are based on current system data and mission reliability estimation rather than probability distribution and MTBF information. Figure 4 shows a screenshot of the online simulation model with customized assets and their class diagram.

The customized asset’s model embeds a PHM interface and a REM module. Each set of PHM interface and REM modules is associated with a monitored critical component. Indeed, a critical component is an electric or mechanic discrete unit with high risk of failure, such as an integrated circuit (Brown et al., 2007), or bearing (Zhang et al., 2011), etc., where the damage can substantially affect the asset’s functionalities. The PHM interface is a dedicated buffer embedded into CAM which stores the most recent PHM data received for the purpose of further reliability estimation. Once the PHM interface is set up with the latest asset’s data, REM estimates the conditional reliability of the critical component for the current completion horizon. This module also requires other input parameters values (such as critical levels, reliability bounds, etc.) that must be defined during the design of the simulation model. In CAM, assets will break down whenever current health’s performance of a critical component reaches its lower reliability bound. Based on this approach, the service provider can implement a distinct maintenance policy in order to replace only the damaged or faulty components, for instance.

Figure 4 also illustrates the class diagram of the CAM model. A class diagram is, typically, an outcome from an early design phase and the foundation for later implementation of the required model (Genero et al., 2007). In the class diagram, a machine class has an aggregation relationship to the PHM and REM class with ‘1 to *’ cardinality, which means that each instance of CAM might have one or more of those modules linked to its critical components. A relationship of association between PHM and REM modules denotes a data transfer from the PHM to the REM module. Although a CAM representation has no limit for the number of
monitored critical components, it must be borne in mind that the simulation models run under Anylogic® platform and will be subject to the computational and memory constraints.

The Holt-Winter smoothing technique (also known as exponential smoothing) was the estimation technique adopted to implement the REM kernel, simply because it does not need a parametric model fitting (Gelper et al., 2010) and also it has similar estimation accuracy compared to other techniques, e.g. Kalman Filter and Extended Kalman Filter (LaViola, 2003). As the asset’s health degradation data for the framework is a continuous degradation process with stochastic dynamic failure mode, the seasonal index of the Holt-Winter smoothing is not considered. Therefore, the second order exponential smoothing (trend-exponential smoothing) can be mathematically described as (Yar and Chatfield, 1990):

\begin{align*}
L_t &= L_{t-1} + \alpha \cdot e_t \\
T_t &= T_{t-1} + \beta \cdot e_t \\
e_t &= y_t - \hat{y}_t
\end{align*}

where:
\begin{itemize}
  \item \( \alpha \) = Local mean smoothing constants;
  \item \( \beta \) = Local trend smoothing constants;
  \item \( L_t \) = Level-component of performance prediction at \( t \)
  \item \( T_t \) = Trend-component of performance prediction at \( t \)
  \item \( e_t \) = One-step ahead forecast error at time \( t \)
\end{itemize}

The mean and the trend constant can be selected within the interval (0;1] (NIST/SEMATECH, 2011). Once the smoothing constant values were chosen, the \( k \)-step ahead estimation can be obtained from the following formula:

\[ \hat{y}_t(h) = L_t + h \cdot T_t \]
where:
\[ \hat{Y}_t^k = k\text{-step ahead performance estimation at time } t. \]

To begin the calculation, the initial values can be estimated as:
\[ L_2 = y_{21}; T_2 = y_2 - y_1 \tag{7} \]

The \( k\)-step-ahead error variance can be obtained, recursively, from the one-step-ahead error variance as (Yar and Chatfield, 1990) (Lu et al, 2001a):
\[ s^2(k) = \left\{ 1 + \frac{2}{I} (k - 1) \alpha^2 \left[ 1 + k \beta^2 + k(2k - 1) \beta^2 \right] \right\} \cdot s^2(1) \tag{8} \]

and the \( s^2(1) \) is estimated as the average of squared 1-step forecast error:
\[ s^2(1) = \frac{\sum_{i=m+1}^{n-1} e_i^2(1)}{n-1-m} \tag{9} \]
\[ e_i(1) = y_{i+1} - \hat{y}_i(1) \tag{10} \]

where:
- \( s^2(k) \) = Estimated error variance for \( k \)-step-ahead prediction;
- \( s^2(1) \) = Estimated error variance for 1-step-ahead prediction;
- \( e_i^2(1) \) = Average of squared \( l \)-step forecast error;
- \( n - 1 \) = End time for the estimation;
- \( m + 1 \) = Start time for the estimation.

4.2.3. Data acquisition model

Data acquisition model (DAM) is the simulation element which provides synchronized data to the simulation model. It models a real time acquisition system with data acquisition channels. Each channel is connected to PHM module in order to capture the current asset’s health degradation data. Once a new asset’s data is available, it is stored in the DAM internal buffer.
The internal buffer is an abstract representation of the parent simulation model (Hanisch et al., 2005) where the model state corresponds to the current asset’s state. Even though the data acquisition channel might use a different sampling rate, the simulation model will only be updated after the elapsed time of the current completion horizon.

Numbers of channels, sampling rate and asset address are the input parameters from the DA module. Number of channels specifies how many channels that can be used to collect asset data. It must be set according to the number of available PHM modules. Sampling rate defines the frequency each channel must sample asset data. Asset address is a virtual address for each acquisition channel. Once a CAM is used to represent a machine, it must be set to the same address in order to receive the message from the DA module. An on/off button provides a manual control over execution of the overall simulation model (including functionalities running in and out of control of simulation engine).

### 4.3. Framework-based simulation execution

Figure 5 shows a flowchart describing the execution of the simulation model.

![Insert Figure 5 here]

Once the simulation experiment is started, the setting of input parameters is verified so as to avoid inconsistencies in the simulation model initialization. An invalid parameter setting (e.g. negative number of channels, invalid asset address, etc.) can lead to wrong simulation results. For this reason, any invalid configuration dispatches an illegal argument exception in order to stop the simulation experiment. With the appropriate parameter setting, the simulation experiment starts the real-time engine and the simulation engine. The former captures PHM data (according to sampling rate) and updates the DAM internal buffer whereas the latter monitors the real-time completion horizon. Indeed, the real-time engine is not connected to the simulation engine which means that it does not stop its activities whilst the simulation engine resumes.

The simulation engine executes, in sequence, the following tasks: 1) update the asset’s model data; 2) estimate the asset’s reliability; 3) resume the simulation execution; and 4) check the
simulation completion horizon. In the first task, the asset’s data stored in the DA internal buffer is cloned into CAM. In other words, DAM broadcasts the asset’s data embedded into information message object (IMO) for all CAM connected on the PHM data bus. An IMO encrypts the current asset’s health degradation data and the asset’s address. A filter mechanism on each CAM accepts or rejects the IMO according to the asset’s address. Once IMO is accepted, CAM is updated and the asset’s reliability is estimated and, finally, the simulation execution is resumed. When the simulation completion horizon expires, current asset’s reliability is compared to the upper reliability bound and, if achieved, the simulation execution continues estimating the asset’s reliability until asset’s failure or a large reliability variance.

5. EXPERIMENTATION

Three experiments were carried out in order to compare performance outcomes obtained from the traditional simulation model to those obtained from the framework-based simulation model with asset’s lifetime variation in two experiments.

5.1. Statements for comparison purposes

One of the challenges in validating the framework was to understand how the outcomes of the proposed simulation approach could be compared to those obtained from a traditional simulation model. In the traditional simulation approach, there is no available information about the age from such asset, and the machine breakdown data are derived solely from probability distributions and MTBF information. Mostly, the MTBF value from the entire component’s population is taken whilst the majority of them are still operating in their service life. In those cases, the MTBF estimation does not match with the service life of a particular component. The only method to represent service life with MTBF value is to calculate it when the whole component’s population reach their end-of-life which is, clearly, unviable for proactive reaction purpose (Torrel and Avelar, 2004).

The first statement matches the service life of each critical component to their respective MTBF values. This hypothesis is in line with Moubray’s assumption (Moubray, 1997)
whenever the critical component’s service life does not encompass their wear-out phase. Moreover, the same hypothesis has already been put forward by Greenough and Grubic (2011) for similar comparative purposes. The second statement with regard to both simulation models must operate with equal setting values (e.g. number of components, execution time, input parameters, etc.). This is fundamental to avoid the influence of other input variations upon the simulation outcomes.

5.2. Experiments

In the first experiment, the asset is running under ideal operating and environment conditions where the asset’s lifetime varies only due to random influences. Likewise, the service contract is entirely executed without requirements modification. Although those hypotheses are rarely verified in practice, they are valid assumptions for comparative purposes. In the second experiment, dynamic behaviour affects the asset’s lifetime (i.e. the expected lifetime is reduced by 20 hours). It is a typical situation, where environment (e.g. temperature, humidity) and/or operating condition (e.g. unpredictable demands), affects the asset’s lifetime. In the third experiment, unforeseen circumstances also affect the expected asset’s lifetime, but this time, due to a better maintenance regime, the asset’s lifetime increases by 20 hours. This information is valuable for the service provider because it can negotiate a contract extension that leads to an additional source of revenue. Table 1 lists the input parameter values for both computer-based simulation models.

[Insert table 1 here]

From Table 1, it can be seen that both simulation models have the same input parameters. Additional input parameters (inserted into the framework-based simulation model) are needed to set the internal REM parameters and do not affect the simulation outcomes. For instance, local mean ($\mu$) and local trend smoothing constant ($\beta$) were set with the values 0.6 and 0.3 respectively. The upper and lower reliability bounds were set at 0.965 and 0.150 respectively. Three critical limits at 95%, 85% and 60% of the asset’s health degradation curve was set but, indeed, only the first critical limit was used because it was assumed that the experiments do not consider the wear-out period of an asset’s degradation curve. Table 2 shows the simulation outcomes for the three experiments.
In Experiment 1 (where the machines are running under ideal operating and environment conditions), the outcomes of both simulation models are similar. Indeed, the framework-based simulation model results indicate a reduction in the number of breakdowns which leads to an improvement in machine availability. Even though the difference, in terms of the number of breakdowns, is relatively low (18 breakdowns), this might become significant if the number of assets covered by PSS contract increases.

Results from Experiment 2 indicate different outcomes for both simulation models. As the traditional simulation model is not fed back with current asset’s data, it cannot capture systematic asset’s lifetime variation but only those from historical data and OEM’s information. Nonetheless, the framework-based simulation model, which is continuously updated with the current asset’s data, has shown a reduction of 23.62% in MTBF, i.e., from 80 hours expected to 61.097 hours actually measured. Likewise, the number of breakdowns increases (50 breakdowns) which leads to a possible contract penalty due to an unexpected maintenance regime. Indeed, this information can be used by the service team in order to find the potential source of the problem or possible future contract modification.

Results from Experiment 3 also point out distinct outcomes obtained from both simulation models. Once more, the results obtained from traditional simulation do not represent the true asset’s performance. The traditional simulation model output shows a reduction of 6% in MTBF, i.e., from 80 hours expected to 75.125 hours, which is clearly an unreliable simulation outcome. On the other hand, the framework-based simulation model has shown an increase of 21% in MTBF, i.e., from 80 hours expected to 96.932 hours actually measured. Furthermore, the number of breakdowns has reduced (32 breakdowns). Therefore, using the results obtained from the framework-based simulation model, the service provider may decrease the number of interventions leading to a more precise and timely maintenance regime, for instance.

6. CONCLUDING REMARKS AND FUTURE WORK
This paper proposes a novel framework to link PHM and PSS using online simulation. In particular, when the responsibility of the OEM is extended beyond manufacturing of the products to include maintenance support (e.g. through contractual agreements), the online simulation model has shown considerable benefits in supporting operational decisions during the contract execution. Using a framework-based simulation model, the service provider can evaluate, in advance, the impact of short-term operational decisions, and subsequently, to react to dynamic behaviours or unforeseen circumstances. This is particularly beneficial for the management of high value assets where untimely maintenance may lead to environmental risks or life-threatening hazard.

Unlike the traditional simulation methods where long-term business requirements are evaluated, the framework-based simulation tool can be used to support short-term operational decisions which typically occur in service contracts. In addition, the framework-based simulation tool can support service provider in making decisions considering the current asset’s health degradation state. If needed, some operational adjustment can be made, e.g. anticipating or delaying maintenance inspection, acquisition of spare parts, etc.

Three experiments were carried out to compare simulation outputs obtained from the traditional simulation model to the framework-based simulation model. The same input parameters spreadsheet was needed in order to evaluate the simulation outcomes solely from the viewpoint of asset’s lifetime variation. In the first experiment, where the asset’s lifetime did not vary, similar simulation outcomes were accomplished. Indeed, a small improvement was noticed on the framework-based simulation outputs and might become considerable when the number of assets increased. Nonetheless, in the second and in experiments, with the asset’s lifetime variation, the framework-based simulation results are closer to the expected results than to those obtained from the traditional simulation model. Those experiment results might be used to guide short-term operational modifications needed to achieve the defined business requirements.

Furthermore, the proposed framework enables the so-called ‘informating products’ (Zuboff, 1988), because company’s knowledge asset and key performance indicators are improved by current asset’s field data. In particular, this can lead to the opportunity to positively change the service delivery strategy for new PSS contracts or extensions. Industrial cases and more
numerical analyses will continue to allow profit analysis from the reliable contract execution. Further investigations will also be conducted in order to evaluate different maintenance strategies. Additional investigation into operational service strategies and more sophisticated repair models will also be needed in order to provide better operational availability estimation.
7. REFERENCES


26


Figure 1: Proposed framework to link PHM and PSS.

Figure 2: Interaction between REM and OSM module.
Figure 3: Framework implementation

Figure 4: Operational simulation model and customized asset representation
Table 1: Simulation input parameter.

<table>
<thead>
<tr>
<th>Simulations and component input data</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Traditional simulation</strong></td>
<td></td>
</tr>
<tr>
<td>MTBF = exponential(x)</td>
<td>80h</td>
</tr>
<tr>
<td>MTTR = exponential(x)</td>
<td>12h</td>
</tr>
<tr>
<td>Service contract duration</td>
<td>15,000h</td>
</tr>
<tr>
<td><strong>Online simulation</strong></td>
<td></td>
</tr>
<tr>
<td>Asset’s lifetime</td>
<td>80h</td>
</tr>
<tr>
<td>MTTR = exponential(x)</td>
<td>12h</td>
</tr>
<tr>
<td>Service contract duration</td>
<td>15,000h</td>
</tr>
<tr>
<td><strong>Reliability estimation module</strong></td>
<td></td>
</tr>
<tr>
<td>Constant smoothing</td>
<td>$\alpha = 0.6; \beta = 0.3$</td>
</tr>
<tr>
<td>Critical limits</td>
<td>CL1 = 95; CL2 = 85; CL3 = 60</td>
</tr>
<tr>
<td>Reliability bounds</td>
<td>upper = 0.965; lower=0.150</td>
</tr>
<tr>
<td><strong>Data acquisition model parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Completion horizon</td>
<td>5h</td>
</tr>
<tr>
<td>--------------------</td>
<td>----</td>
</tr>
<tr>
<td>Number of channels</td>
<td>1</td>
</tr>
<tr>
<td>Sample rate</td>
<td>1h</td>
</tr>
</tbody>
</table>

Table 2: Simulation results obtained from experiments

<table>
<thead>
<tr>
<th>Simulations outcomes and performance measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Traditional simulation</strong></td>
</tr>
<tr>
<td><strong>Availability (%)</strong></td>
</tr>
<tr>
<td>87.6</td>
</tr>
<tr>
<td><strong>Jobs complete</strong></td>
</tr>
<tr>
<td><strong>Utilization (%)</strong></td>
</tr>
<tr>
<td><strong>Breakdowns</strong></td>
</tr>
<tr>
<td><strong>MTBF (h)</strong></td>
</tr>
<tr>
<td><strong>MTTR (h)</strong></td>
</tr>
<tr>
<td><strong>Online simulation</strong></td>
</tr>
<tr>
<td><strong>Availability (%)</strong></td>
</tr>
<tr>
<td><strong>Jobs complete</strong></td>
</tr>
<tr>
<td><strong>Utilization (%)</strong></td>
</tr>
<tr>
<td><strong>Breakdowns</strong></td>
</tr>
<tr>
<td><strong>MTBF (h)</strong></td>
</tr>
<tr>
<td><strong>MTTR (h)</strong></td>
</tr>
</tbody>
</table>