Hybrid Models and Digital Twins for Condition Monitoring: HVAC System for Railway

Antonio Gálvez[1,2*], Jokin Rubio[1], Dammika Seneviratne[1], Asier Gonzalez[1],
Alberto Jimenez[1], Unai Martinez-de-Estarrona[1], Diego Galar[1,2], Esko Juuso[3]

[1]TECNALIA, Basque Research and Technology Alliance (BRTA), 48170 Derio, Spain; *antonio.galvez@tecnalia.com,
 Dammika.seneviratne@tecnalia.com, diego.galar@tecnalia.com

[2]Division of Operation and Maintenance Engineering, Department of Civil, Environmental and Natural Resources
 Engineering, Luleå University of Technology, 971 87 Luleå, Sweden; *antonio.galvez@ltu.se, diego.galar@ltu.se

[3]Control Engineering Group, Faculty of Technology, University of Oulu, 90014 Oulu, Finland; esko.juuso@oulu.fi

Abstract. Safety passenger transportation is more important than efficiency or reliability. Therefore, it is vital to maintain the proper condition of the equipment related to the passengers' comfort and safety. This manuscript presents the methodology of complete development and implementation of both hybrid model and digital twin 3.0 for an HVAC in railways. The objective of this is to monitor the condition of the HVAC where it matters to the comfort and safety of the passengers in the trains. The level 3.0 of digital twin will be developed for the diagnosis and prognosis of HVAC by using hybrid modeling. The description illustrated in this paper is focused on the methodology used to implement a hybrid model-based approach, and both the need and advantages of using hybrid model approaches instead of data-based approaches. The development considers the importance of safety and environmental risks, which are included in the risk quantification of failure modes. Railway's maintainers replace critical components in early stages of degradation; thus, the use of a data-driven model loses essential information related to advanced stages of degradation which might decrease the accuracy of the maintenance instructions provided. Physics-based model can be used to generate synthetic data to overcome the lack of data in advanced stages of degradation, and then, the synthetic data can be combined with the real data, which is collected by sensor located in the real system, to build the data-driven model. The combination leads to form hybrid-model based approach with a large number of failure modes that were unpredictable. Finally, the outcome is beneficial for the proper functioning of systems; hence, safety of the passengers.

Introduction

The diagnostics and prognostics are the main processes in condition-based maintenance (CBM). These processes are broadly used in industrial assets, mainly, at levels of part, components, sub-systems, and systems. Fault detection and diagnostics (FDD) is the identification of a faulty component through the detection and isolation of a fault. Once a fault appears in a system, diagnostics process might detect that fault and identify the faulty part. Prognostics process is performed for estimating the remaining useful life (RUL) of a system. It is estimated using the behavior data while the system works, allowing maintainers to avoid the corrective maintenance [1]. Thus, appropriate implementation of CBM leads to reduce costs and increase the reliability and safety of systems.

As it is shown in Figure 1, there are highlighted four techniques for estimating the RUL and, on a consequence, for implementing FDD and prognostics processes [1]: experience-based approaches, model-based approaches, data-driven approaches, and hybrid model-based approaches (HyMAs).
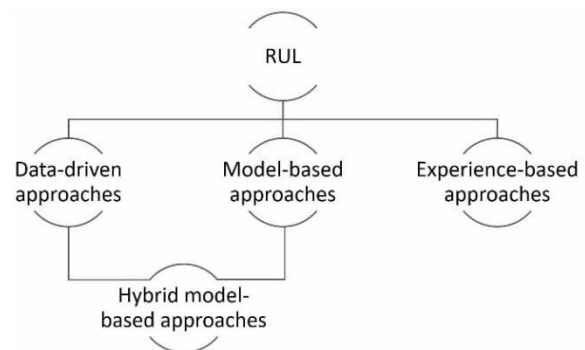

Figure 1: Remaining useful life (RUL) estimation models.

The model-based approach are mathematical models of the physical system [2]. These models incorporate such characteristics as material properties, thermodynamic, and mechanical responses. Furthermore, different operational conditions, components defect, and degradation must be modelled for fault detection. Thus, these models accurately describe the behaviour of systems [3] [4].

Data-driven approaches use operational data collected by sensors embedded in the system for obtaining information related to the degradation state of components. Thus, these approaches can trace the relationship between the data acquired from the system with its degradation [5] [6].

As Figure 1 shows, HyMAs is the combination of data-driven and model-based approaches. The main advantage of this combination is the reduction in both the historical information required to train a data-driven model and the information needed for a robust physics-based model. This data fusion aims to improve FDD and prognostics processes. The model-based approaches generates synthetic data to overcome the lack of historical data, thus improving the ability of a data-driven apprach to detect failure modes (FMs) and reducing the appearance of hidden FMs, metaphorically known as "black swan losses" [7].

1 Problem Discussion

The case study used in this research work is a heating, ventilation and air conditioning (HVAC) system which is installed in a high-speed passenger train.

Currently, many types of research are based on data-driven approaches but missing important information and failures. This occurs because maintainers replace critical elements in early stages of degradation for safety, reliability, economic, and environmental reasons. Therefore, it is challenging to acquire data in faulty stages of the system and advanced stages of degradation [3].

The solution proposed in this manuscript is to transfer the methodology to railway subsystems to scale, develop, and validate an HyMAs for railway equipment through the combination of modeling based on physical laws and modeling based on data. As a primary purpose, it will be developed to improve the diagnostics and prognostics processes of components of railway systems.

The outcome will be to provide railway companies with the necessary tools to reach a level 3 of digital twin based on HyMA. The Digital twin levels are listed in Table 1 and shown in Figure 2, Figure 3, and Figure 4.

Digital Twin	Data	Layer	Level
1.0	SCADA Data Online Data	OT (SCADA)	**Operational** Alarms based on anomalies
2.0	Adding: Offline data, GMAO, ERP, taxonomy, ontology management data	OT (SCADA) and IT (ERP, management data...)	**Tactical** Planning of corrective maintenance instructions.
2.1	Context data	Operative context	
3.0	Physics model Hybrid digital twin	OT + IT + Engineering technology	**Strategic** Renew of assets, performance improvements

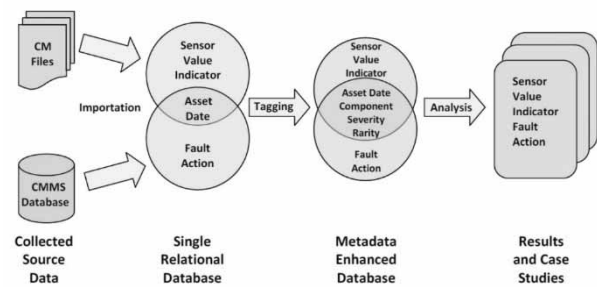**Table 1:** Digital twins levels description.

Figure 2: Digital twin 2.0, extracted from [8].

Figure 3 shows the Digital Twin 2.1, who combines the Digital Twin 2.0 (see Figure 2) and context data. Context-awareness is considered as an application's ability to adapt to changing circumstances and respond according to the context of use.
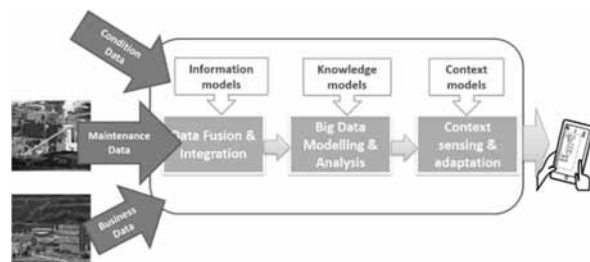

Figure 3: Digital twin 2.1, extracted from [8]).

Figure 4 shows the Digital Twin 3.0, who combines the Digital Twin 2.0 with a HyMA. It could be used from component to system level.

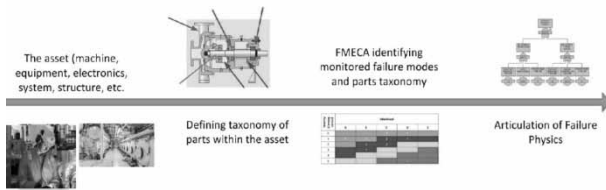**Figure 4:** Digital twin 3.0, extracted from [8].

Figure 5: Context-Driven CBM Digital twin 3.0, modified from [8].

2 Objectives

The aim of this research work is to provide maintainer of the HVAC system with the necessary tools to reach a level 3 of Digital Twin based on HyMA – see Table 1. The objectives of this development are listed below.

1. Study the different approaches to achieve a robust hybrid model of the HVAC system. Model must consider the number of variables monitored in the real system, the information contained in it, and the information available in other sources, such as expert knowledge, other papers, the history of failures and alarms, and work orders. The selection of information is crucial for developing a robust HyMA because it helps to identify the balance between the resources needed for modeling versus the accuracy of FDD and prognostics processes. Then, this leads to propose a methodology for combining data-driven and model-based approaches; thus, for implementing the HyMA. This approach will be firstly applied to HVAC use case.

2. Study the available models of the under-study asset, and then, develop models based on the physics of the monitored system that represent the response of the HVAC system. This includes electrical modeling of motors such as electric circuits, modeling of moving parts such as rotation of rigid solids, or modeling of heat exchangers through the energy balance, among others.

3. Definition the methodology for synchronizing the model with the real system. This step fixes the values of key unknown parameters of the physics-based models. The synchronization of the model is carried out using the data collected by the available sensors. The data used will be considered under normal operating behavior.

4. Validation of the HyMA will be carried out by comparing the response of the the physics-based model with the response of the real system. The validation will focus on fault detection, fault location, and fault identification.
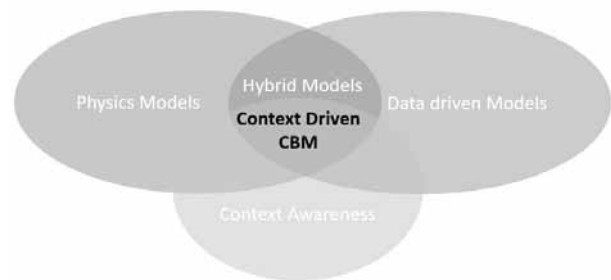
3 Advantages of the Hybrid Approach

Industries where safety is more important than efficiency or reliability, as it is the case of the railway sector, components are replaced in the initial stages of degradation, sometimes far from their maintenance limit, due to regulations and regulations, such as the standard EN 13848-5:2017 which stablishes strong replacement criteria for the railway sector.

This early substitution of components implies a lack of data of the complete degradation of replaced elements. This lack of data limits the visibility of FMs and the appearance of the aforementioned black swan losses, which are events associated with failures not foreseen or contemplated [8]. This can be avoided by using synthetic data generated by physics-based models, which is built for reproducing the behavior of the asset in degraded or unusual conditions and complete database on failures and degradation of systems. It should be noted that the assets are usually complex systems composed of several subsystems and their components, which makes it difficult to perform an accurate maintenance model. Consequently, the interaction between components and subsystems makes it challenging to obtain a global failure scenario due to sub-systems work together and they can fail individually or together. Thus, complex systems have many possible combinations of faults.

It is, therefore, necessary to develop a strategy to overcome the lack of data which might provide information on the evolution of system degradation. This strategy must be hybrid, considering both model-based approaches, that is, models based on physical laws; and models based on data, in such a way that their combination can supply the shortcomings of each method. The primary objective of the project is the transfer of a methodology to build a HyMA for obtaining more information of faults and how these faults change over time.
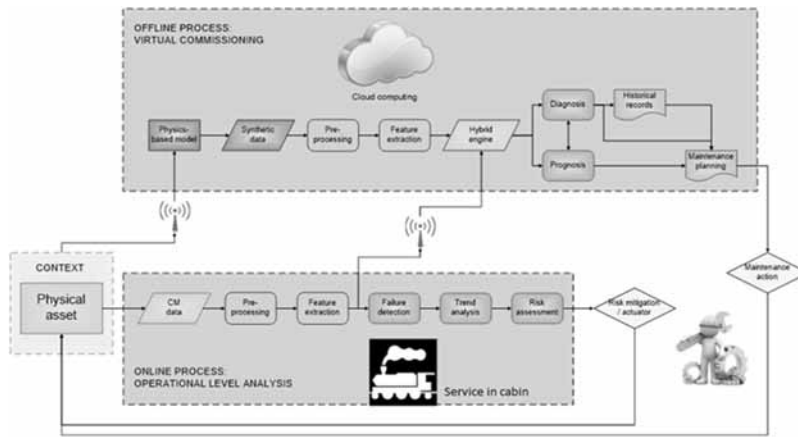
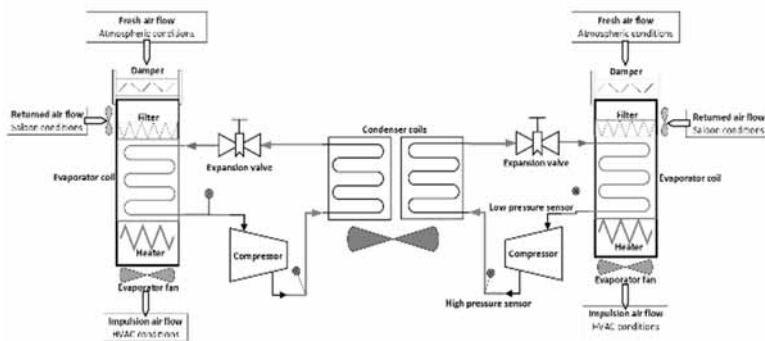**Figure 6**: Proposed architecture scheme (modif. from [9]).

Figure 7: HAC saloon scheme.

As it is shown in Figure 6, a dataset with operational data is generated; then, this data is used for performing diagnostics and prognostics processes. These processes generate the necessary information that constitutes the knowledge base for maintenance planning will be generated.

4 Case study on HVAC

The HVAC system is installed in a passenger train carriage. These systems are composed by many vital components. The main parts can be identified as compressor, condenser coil, evaporator coil, bypass valve, dehydrator filter, and airflow units (Figure 7).

Compressor. Move the refrigerant liquid from the low-pressure chamber to a high-pressure chamber by scroll mechanism without excessive noise and vibrations. The monitoring of compressor status can be performed indirectly through the parameters/variables, for example, inlet pressure, outlet pressure, engine current, vibration, the temperature of the compressor parts, etc.

Condenser coil and Evaporator coil. Heat exchanger between airflow and refrigerant liquid. These components do not usually have movable elements; their performance is mostly controlled by inlet and outlet temperatures.

Bypass valve. Regulates the refrigerant flow.

Dehydrator filter. Retains the moisture and solid particles existing in the refrigerant circuit.

Air diffusion unit. Composed mainly of fans connected to electric motors. This component can be monitored through the vibrations of the fans, the airflow and electrical consumption.

5 Methodology

The methodology shown in this section is a general approach used in many research studies where CBM is deployed in an industrial environment.
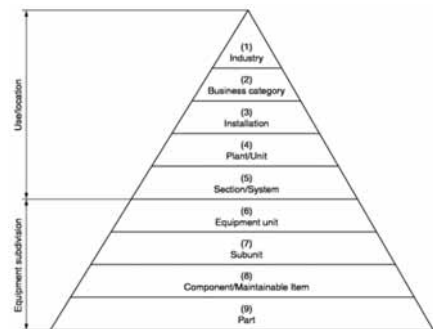

Figure 8: Taxonomy classification with taxonomic levels (from standard ISO 14224-2016).

The methodology followed is composed of the items below:

1. Analysis of the information related to the HVAC system, such as design, operational conditions, maintenance tasks, etc., and analysis of the data acquired by the sensors embedded in the system.

2. Evaluation of the current capabilities; the maintainers of the HVAC system understudy use a data-driven model to perform diagnostics and prognostics. Therefore, it is necessary to evaluate the limitations of the current data-driven model and the maturity of the deployed diagnostics and prognostics processes. Level 2.1 of the Digital Twin confirmed but improvable.

3. Taxonomy of the system considering the standard ISO 14224-2016. Before developing the taxonomy of the system, the taxonomic level of interest must be identified – see Figure 8 [7]. Identification of the subsystems, equipment, subunits, and components for building the hybrid model. This step reaches until level 8 of the taxonomy classification shown in Figure 8.

4. Analysis of the current list of FMs, their effects and causes. Moreover, study the possibility to add new FMs.

5. FMs classification by their criticality. Then, quantification of safety and environmental risks.

6. Selection of variables and parameters to perform the physics-based model, definition of appropriate models for each component based on the sensors installed in the system.

7. Identification of critical items and significant variables.

8. Analysis of models developed by the railway company. Adaptation of models, if possible, or development of new physics-based models and data-driven models; these models will be developed separately. Digital Twin 2.1 enhanced.

9. Integration of physics-based models and data-driven models to build the hybrid model considering subsystems. Level 3.0 of Digital Twin reached.

10. Validation of the final HyMA. Level 3.0 of Digital Twin reached and tested.

6 Planning

The research work will be addressed in two phases: the first phase will provide an early evaluation of the scope of the HyMA; and the second phase leads to develop the diagnostics and prognostics processes from model defined in the first phase. The scope of this research work includes both phases for the HVAC system.

6.1 Phase 1: Viable minimum product for evaluation of the scope of the hybrid approach on HVAC

The first phase involves the firsts contact between the technical team and the customer. This allows the technical team to study all the documentation, data, information related to the components and system understudy.

1. Visit and study the real system, analysis available data, and mapping of FMs and variables.

2. Estimation of diagnostics and prognostics performance.

3. Analysis of the models developed by the railway company; then, proposal of models for deploying both the diagnostics and prognostics performance.

6.2 Phase 2: Development of diagnostics and prognostics processes using a hybrid approach to HVAC (HyMA-HVAC)

After the study and proposal submitted in the first phase. The project should be narrowed down and ready to start this phase.

1. Modeling of the system and subsystems chosen. Determination of the most appropriate modeling in each case: compressors, heat exchangers, air diffusion, blowers....

2. Parametrization and fine-tune of models with available operational information.

3. Selection of FMs based on the performed analyses and modeling capabilities.

4. Giving an understanding of the physics of the selected FMs for integrating them in the model.

5. Generation of synthetic data under unique fault conditions, if possible, varying severities.

6. Study of the most suitable hybridization strategy in each case for the export of results to intermediate files; so that, the railway company provides data of failure to its system: limit values, rules, data export for data-based modeling... expanding the database of knowledge of FMs and operation parameters.

7. Implementation of diagnostics and prognostics processes based on the hybridization of data sources.

8. Transversely, transfer and integration of developed models, services, and work methodology; thus, the railway company may integrate the new strategy and tool in their current system.

7 Discussion

Once the digital twin 3.0 is tested and ready for the use, it can analyze the behavior of the system. The data will be taken from the sensors every thirty minutes; the system will record the operational data, parameters, and variables in a dataset as a fingerprint. It will be compared with the fingerprints previously saved. There are two possible situations: first, an alarm is not detected. Second, an alarm is detected. In the second case, the algorithm chooses the data with the biggest changes in the value in comparison with the previous value.

If it is a false alarm is detected, the fingerprint and the operational data will be deleted; if it is not a false alarm, the algorithm will estimate the RUL considering the operational data. The RUL will also be estimated when the algorithm does not detect any alarm.

The RUL is estimated using the fingerprint, operational data, and context data [10]. It is important to highlight the importance of FMs related to critical components in train transportation. One of the peculiarities of this implementation is the effort made to evaluate the safety and environmental risk, this means that in this field, the implementation cannot consider just the operational and reliability risks. This is because a failure in this field can affect people or the environments; thus, safety and environmental risks become more important than reliability of systems.

The development allows us to know the behavior of the system and its progression. Thus, it shows and analyses how the degradation of the system varies in real-time, which reduces the time between maintenance instructions, and gives supports to maintenance area for managing the maintenance tasks.

8 Conclusions and Future Work

This research work starts with a rigorous study and evaluation of the maturity level of the current diagnostics and prognostics methodology developed for HVAC system. It is crucial for both knowing the utility of the available information and for having an overview of the personnel abilities related to CBM, which is important for the posterior staff training.

The data-driven approaches have a good result, but in train transportation, maintainers replace the critical components in early stages of degradation; thus, information is missing, and most of them are part of the advanced stages of degradation, this means that the approaches cannot identify several FMs in these stages. The HyMA proposed will use physics-based model, which uses real data, to generate synthetic data to overcome the aforementioned lack of data. Thus, many FMs in advanced stages of degradation will be identified by the proposed model.

The knowledge of more FMs and the possibility to predict accurately the RUL of a system reduce maintenance cost; a robust CBM supports decision making process; therefore, this allows maintainers to give components longer service life under required operational conditions.

Longer times between maintenance tasks implies the reduction of the spare parts inventory, and then, the reduction of maintenance cost.

The data-driven model is trained using data acquired from the real system and synthetic data, which is generated by the physics-based model. The synthetic data contains FMs that are already defined in the literature, experience, or other sources; thus providing a HyMA fully useful. Nevertheless, giving critical components a longer service life could results with the appearance of new FMs which have never been detected before, the aforementioned "black swan losses".

The future of this work is focusing on detecting and identify black swan losses for researching methods to identify the behaviour of every unknown failure.

References

[1] Galar D, Kumar U. eMaintenance: Essential Electronic Tools for Efficiency. Academic Press, 2017.

[2] An D, Kim N H, Choi J. *Practical options for selecting data-driven or physics-based prognostics algorithms with reviews.* Reliability Engineering and System Safety, vol. 133, pp. 223-236, 2015.

[3] Leturiondo U, Mishra M, Galar D, Salgado O. *Synthetic data generation in hybrid modelling of rolling element bearings.* Insight - Non-Destructive Testing and Condition Monitoring, vol. 57, nº 7, pp. 395-400, 2015.

[4] Heng A, Zhang S, Mathew J. *Rotating machinery prognostics: State of the art, challenges and opportunities.* Mechanical Systems and Signal Processing, vol. 23, nº 3, pp. 724-739, 2009.

[5] Turner W, Staino A, Basu B. *Residential HVAC fault detection using a system identification approach.* Energy and Buildings, vol. 151, pp. 1-17, 2017.

[6] Iyengar S, Lee S, Irwin D, Shenoy P, Weil B. *WattHome: A Data-driven Approach for Energy Efficiency Analytics at City-scale.* 2018.

[7] Aven T. *On the meaning of a black swan in a risk context.* Safety Science, vol. 57, pp. 44-51, 2013.

[8] Galar D. *Hybrid Models & Digital twins as prognosis enablers.* In 3rd Maintenance Congress, Bilbao, 2019.

[9] Leturiondo U, Salgado O, Ciani L, Galar D, Catelani M. *Architecture for hybrid modelling and its application to diagnosis and prognosis with missing data.* Measurement, vol. 108, pp. 152-162, 2017.

[10] Galar D, Thaduri A, Catelani M, Ciani L. *Context awareness for maintenance decision making: A diagnosis and prognosis approach.* Measurement, vol. 67, pp. 137-150, 2015.