# Building Resilience Into Urban Rail Transport Systems

Pierre Dersin, Ph.D., Alstom
Alban Péronne, Alstom

## *SUMMARY*

*Today's urban rail transport networks are an essential instrument for large metropolitan areas in coping with the growing demand for punctual, reliable and environment-friendly transport services. Alstom's solution to address that need is its Urbalis® communication-based train control system.*

*Resilience is the capacity to recover quickly from difficulties; in this context, it is the ability for the system to continue to perform its transport function, or to return quickly to nominal operation, after disturbances caused by unforeseen situations. The situations Urbalis® is designed to deal with are of three types: 1) hardware and software failures or degradations; 2) disruptions resulting from external perturbations, including passenger usage; 3) malevolent attacks.*

*To deal with hardware and software failures, Urbalis® relies on highly redundant architectures, in particular that of the data communication system which is its backbone, but also on the use of innovative maintenance and asset management strategies; in particular, predictive maintenance, supported by the HealthHub™ platform, aims at detecting degradations before they result in service-affecting failures.*

## 1 INTRODUCTION

### 1.1 Architecture

Communication system redundancy involves duplicated and fully separated wired and radio networks for end-to-end communication between trackside and trainborne equipment. Vital messages, essential for signalling, are sent through two different channels, and non-vital multi-media (Public Address, Passenger Information, Advertising, …) messages are carried by a third network. Wi-Fi communications between trackside and trainborne equipment are managing frequency reconfiguration in case of a radio perturbation or cyber-attack. On-line built-in test equipment with high detection rates enables the detection of partial failures before a function loss happens. Redundancy management mechanisms are made as simple as possible and attention has been paid to critical interfaces such as with power supply sources, to avoid single points of failure. Redundancy integrity and data flow are continuously monitored by the centralized maintenance system. Design protection mechanisms against common cause failures have been applied. The centralized equipment that supports key functions such as automatic train control, interlocking and train supervision can be duplicated to provide a standby redundancy backup that guarantees a short recovery time in case of catastrophic events such as a flood, fire, or malicious attack.

### 1.2 Maintenance & Operations

In addition, predictive maintenance is applied to wayside assets such as point machines, whose failures significantly impact service; it consists of performing maintenance operations based on the condition of the assets rather than scheduled maintenance (time-based or distance-based). Just as for the communication network, the goal is to avoid service-affecting failures as much as possible, this time by detecting evolving degradations and intervening before they result in a failure. Machine learning and domain knowledge are combined to construct health indicators which measure the distance to 'perfect health' and are used for degradation detection, diagnostics and prognostics. Beyond that, HealthHub™ paves the way for dynamic maintenance management based on evolving asset conditions.

Finally, adaptive traffic management algorithms ensure quick restoration of full nominal operation after a disruption (resulting from passenger flow fluctuations, passenger behaviour such as door obstruction, or external delays).

# 2 BUILDING RESILIENCE INTO URBAN RAIL TRANSPORT SYSTEMS

## 2.1 Abbreviations

- ATC Automatic Train Control
- ATS Automatic Train Supervision
- AXC Axle Counter
- CATS Central Automatic Train Supervision
- CBTC Communication-Based Train Control
- CC Carborne Controller
- CCF Common Cause Failures
- CER Central Equipment Room
- CIXL Central Interlocking
- CSP Cybersecurity Platform
- DCS Data Communication System
- FD Fault Detection
- LC Line Controller
- MSS Maintenance Support System
- NMS Network Management System
- OCC Operations Control Center
- OEM Original Equipment Manufacturer
- RBD Reliability Block Diagram
- SER Station Equipment Room
- SMIO Smart Input Output Module
- SPF Single Point Failure
- SPOF Single Point of Failures
- TRE Trackside Radio Equipment
- TWU Train Wake-up Unit
- UFT Undetected Fault Time
- ZC Zone Controller

## 2.2 Architecture

The key idea underlying communications-based train control (CBTC) is to enable a high capacity with as little wayside equipment as possible. Primary train detection is no longer achieved with trackside detection devices such as axle counters or track-circuits, but instead with bi-directional train-to-ground communication. CBTC makes it possible to know the positions of the trains more accurately than with the traditional signalling systems. This results in a more efficient and safe way to manage the railway traffic.

The availability of the CBTC system thus becomes a key factor of overall system availability and, to reach the very high system availability targets required by the market, it is necessary to design redundancy into that system.

In Urbalis®, the most common choice is active redundancy.

In addition however, the centralized equipment that supports key functions for a complete metro line, such as automatic train control (ATC), central interlocking (CIXL) and automatic train supervision (ATS) which are located in the Central Equipment Room (CER) and Operations Control Center (OCC), can be duplicated in a different geographical location (i.e. Backup CER and OCC) to provide a standby redundancy backup that guarantees a short recovery time in case of catastrophic events such as a flood, fire, power shutdown, or malicious attack, as illustrated in Figure 1.

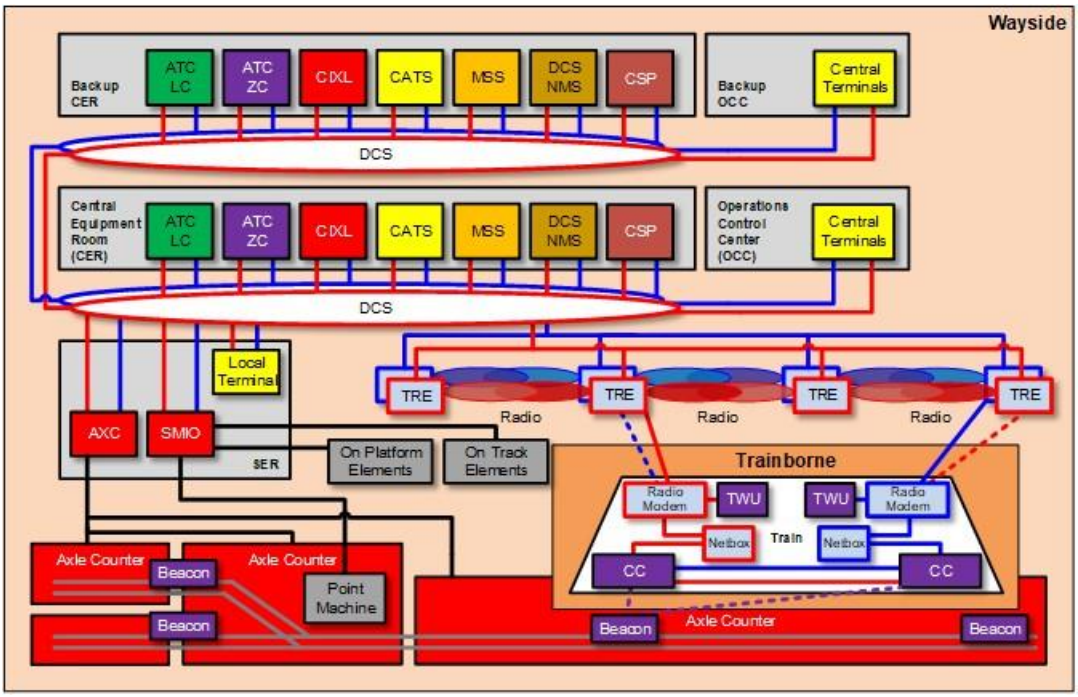


*Figure 1: Overview of the Urbalis® 400 system architecture*

Data communication system (DCS) redundancy involves duplicated and fully separated wired and radio networks, making a CBTC functional failure very unlikely. Vital messages, essential for signalling and train operations, are sent through two different channels (i.e. red and blue links), and non-vital messages such as non-critical media services are carried by a third network (green links, not represented on the figure).

The front on-board modem and the rear on-board modem always communicate with two different trackside radio equipment (TRE) access points, which considerably reduces the risk of interference.

In addition, the validity time of messages (for location or end-of-authority) are of the order of several seconds; as a result, detailed modelling [7] has demonstrated that the probability of a spurious emergency brake due to interference is close to zero.

In case of TRE failure or radio perturbation (i.e. intermittent loss of radio coverage) the nominal communication path (e.g. between blue radio modem and blue TRE) is dynamically rerouted (i.e. between blue radio modem and red TRE) to finally be nominally rerouted at central equipment location.

In the very unlikely case of a radio communications failure, the CBTC function will be unavailable on a track section, triggering the emergency brakes, stopping trains, and causing delays due to the loss of driverless capability. To mitigate the effect of such major perturbations, the system will automatically switch to a degraded operation mode until CBTC function restoration. In this degraded mode, train detection is achieved by trackside secondary detection devices (i.e. axle counter or track circuit equipment).

The remote smart input/output modules (SMIO) are distributed in station equipment rooms (SER) to control and monitor wayside objects (e.g. point machines, platform screen doors, emergency plungers, etc.). Those modules can be internally redundant or not depending on the criticality of failures of the wayside objects they manage (for instance, a point machine located at a turnback will need a redundant SMIO, but a point machine used for cross-

over generally will not). This flexibility enables a high availability performance to be achieved whilst keeping corrective maintenance costs under control.

Achievement of high availability performance is mainly depending on redundancies, including hardware and software. Commonly, the K-out-of-N term defines the type of redundancy (e.g. 1 out of 2, 1 out of 3, etc.). Some frequently encountered examples are:

- 2- out of -3 redundancies with software diversity;
- 1- out of -2 redundancies of 2 out of 2 structures;
- 1- out of- 2 redundancies of coded mono-processors.

It is essential to correctly specify, design, manufacture, and test redundant architectures. Redundant structures are useful only if redundancy is correctly managed. This means that, when a channel fails, it must be restored before a function loss has occurred.

Figure 2 depicts a Markov model: circles represent system states, and arrows, failure or restoration transitions, respectively denoted λ and μ. This Markov model represents an ideal 1- out of- 2 active redundancy. State 1 is the system "complete up- state" (fully operational) and is the initial state. State 2 is a degraded state since one of the two channels has failed, but the main function is still performed. State 3 is a system down state when both channels have failed and therefore the main function is lost.

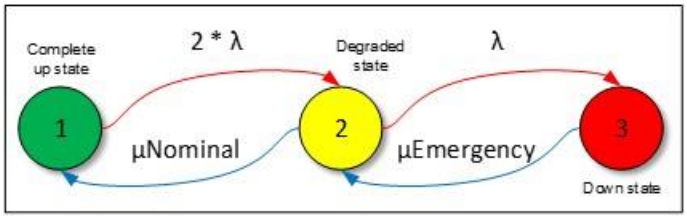


*Figure 2: Markov model of an ideal 1 out of 2 redundancy*

Testability is a key design characteristic for redundant structures since fault detection will trigger the redundancy management sequence and it is the first step of the fault restoration process. On-line built-in test equipment with high detection capabilities enables the detection of partial failures before a function loss happens. Such a detection can be made by means of built-in test equipment or periodic inspection, or a combination of both [1].

If testability is imperfect, the first failure of a channel might not be detected: this is a latent failure, therefore it will not be restored before a function loss has occurred. Figure 3 illustrates the testability sub-process in the context of the fault restoration process.

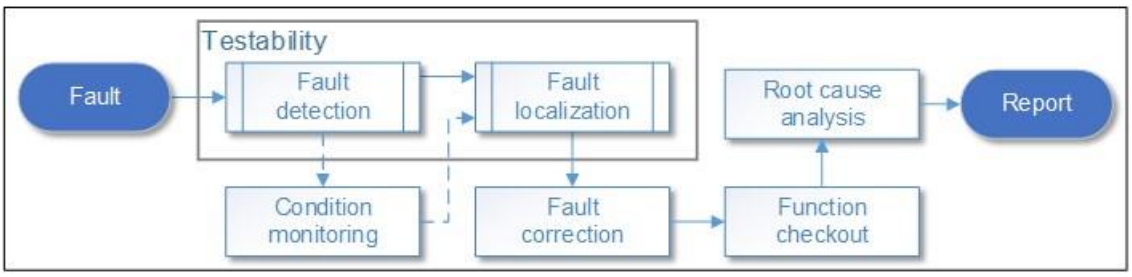


*Figure 3: Testability I the fault restoration process*

When modelling redundancies to estimate operational availability, various factors must be considered, as presented in [1] and [2]; such factors are:

- The redundancy management types (i.e. active, hot stand-by or cold stand-by);
- The testability design characteristic, (i.e. the fault detection rate);
- The switching probability and the warm-up time for stand-by redundancy;

- Maintenance resources available for restoration and the maintenance strategy (e.g. restoration priorities, possible deferred maintenance).

In Urbalis®, all those factors are modelled by means of Markov models or stochastic Petri nets, as shown in Figure 4 below, representing a Markov model of a 1- out- of- 2 redundancy considering imperfect detection. In this example, the Fault detection (FD) probability drives the system either to a detected degraded state (State 2) or a latent degraded state (State 3). If nothing is done, the system will inevitably fall into a down state (State 4), resulting in a loss of the function. One option to correct this weakness is to export a constraint for performing periodic scheduled maintenance inspections to limit the undetected fault time (UFT) and detect the failure to begin the restoration of the faulty channel (State 5). There is a trade-off between the necessary frequency of the maintenance inspections and the fault detection probability, as previously studied by the authors [2].

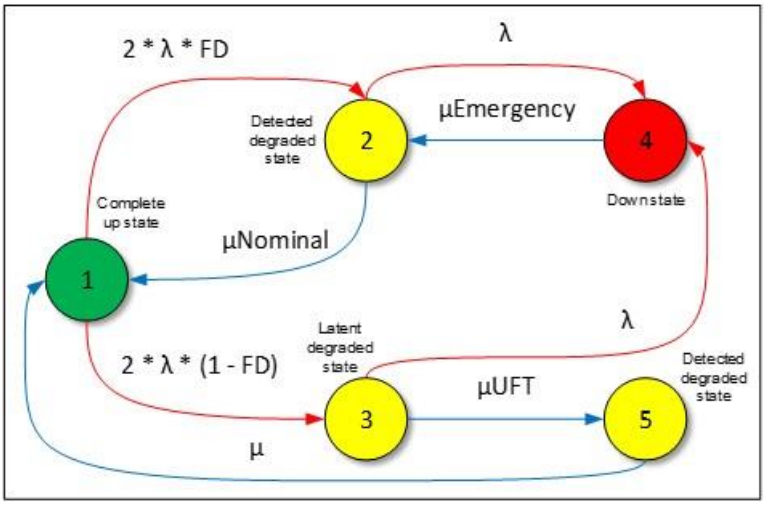


*Figure 4: Markov model of a 1 out of 2 redundancy considering imperfect detection*

For stand-by redundancy, the switching probability (i.e. from nominal to reserve channel) and the warm-up time must be considered. Figure 5 hereafter provides reliability block diagram (RBD) representation of the various redundancy management types.

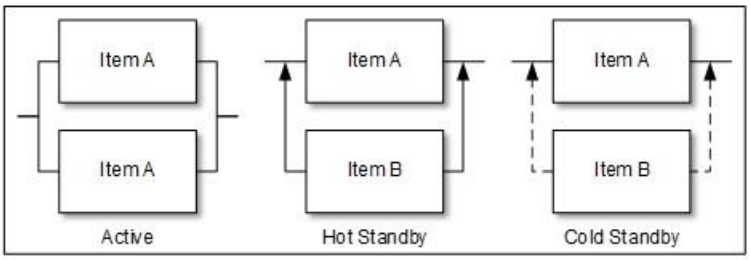


*Figure 5: RBD representation of the various redundancy management types*

The same equipment may be considered with hot standby redundancy or cold standby redundancy, depending on the application constraints.

Even if the system is made up of redundant structures it is important to consider critical interfaces with external systems, such as with power supply sources, to avoid single point of failures (SPOF).

Another aspect which must be systematically addressed and included in the redundant structure models, is the presence of common cause failures (CCF). CCF are failures of multiple items, which would otherwise be considered independent of one another, resulting from a single cause [1]. CCF belongs to a sub-category of multiples failures as illustrated by Figure 6.

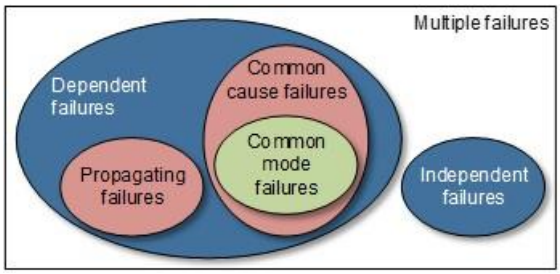


Figure 6: Types of multiples failures

CCF may result from a systematic fault (e.g. a design or specification mistake) or an external stress leading to an early residual hardware failure (e.g. an excessive temperature resulting from the residual hardware failure of a common cooling fan, which accelerates the life of the components or takes them outside their specified operating environment) or, possibly, a combination of both. Experience has shown that CCF occur even on simple systems, and sometimes years after equipment first deployment.

CCF and SPF drastically reduce the effectiveness of redundancy since they correspond to a block in series (i.e. first order failure) in the associated RBD, as shown in Figure 7.

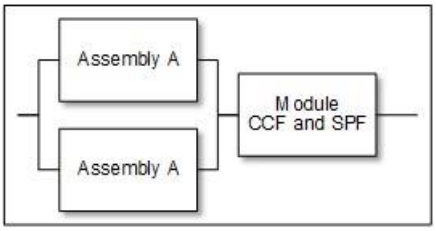


*Figure 7: Reliability block diagram of a 1 out of 2 redundancy considering CCF and SPF*

There are several methods to assess failure rate due to CCF, a famous one is the β-factor method because it is easy to use, and it is supported by the IEC 61508-6 [5] international standard. In addition, this standard provides interesting considerations on defence against CCF, after reducing the overall number of residual and systematic failures. The integration of such defence mechanisms at early development stage of redundant structures is essential to achieve efficient redundancy. It also makes the system more robust to causes which are beyond the scope of supply, such as human errors or external events. The defence mechanisms are grouped by categories:

- Separation/segregation (e.g. geographical separation);
- Diversity/redundancy (e.g. different technologies, test methods, or maintainers);
- Complexity/design/application/maturity/experience (e.g. simple, proven design, interfaces protection);
- Assessment/analysis and feedback of data (e.g. root cause analysis);
- Procedures/human interface (e.g. testability, human aspects);
- Competence/training/safety culture (e.g. training on common cause failures);
- Environmental control (e.g. personnel access limited, control of various stresses);
- Environmental testing (e.g. immunity to all relevant environmental influences).

Urbalis® 400 development applies those defence mechanisms as much as possible, naturally based on feasibility and considering cost constraints. Using backup for centralized equipment enters the separation/segregation category. Routing DCS red and blue cables at opposite sides of the tunnel section enters the same category, such consideration reduces the probability of having rodents cutting the two redundant communication cables. Automatic train control (ATC) redundancy between primary train detection and trackside secondary detection devices falls into the diversity/redundancy category.

Even though CCF must be considered since it is observed on the field and it is recommended by standards, the β-factor method seems to give pessimistic failure rate prediction results compared to the estimations based on test

and field data. Unlike Alstom, not all OEMs consider the CCF in their availability assessment, which can be a problem regarding competitiveness in tenders.

## 2.3 Predictive Maintenance

Whilst some failures are sudden (as mostly in the electronics control systems and communication networks), other are preceded by progressive degradations. For the latter, predictive maintenance attempts at spotting those degradations before they lead to function loss, i.e. failure, so as to schedule preventive maintenance operations to avoid such failures. Amongst trackside equipment, points are the most critical mainly due to delays and perturbations (and even sometimes accidents) which their failures cause, and the high maintenance costs they entail.

Alstom's predictive maintenance approach, HealthHub™, is based on a combination of physics of failures and machine learning. Health indicators are built with features extracted from raw signals (typically currents and voltage of the point machine) acquired from the monitored assets. Alerts are generated when the monitored asset begins to operate abnormally. The advantages of this method over more-traditional ones that rely on thresholds placed directly on raw signals are that the rate of missed and false alarms significantly drops. [3].

All assets are monitored individually, i.e. the normal operating behaviour is learnt individually for each asset, and operating context is considered. This approach enables the operator to obtain only the relevant information, and the lack of arbitrary thresholds on the acquired signal allows the method to be much more robust.

The process follows the phases outlined in the ISO 13374 [6], standard for Condition Monitoring and Diagnostics:

- Data Acquisition (DA);
- Data Processing (DP);
- State Detection (SD);
- Health Assessment (HA);
- Prognostic Assessment (PA);
- Advisory Generation (AG).

The process is now described briefly, together with the advantages derived from implementing the HealthHub™ approach to the point machines.

In the Data Acquisition phase, data is acquired from the on-field points, through a set-up of sensors and connections. The data, which is acquired for each manoeuvre of the machine, consists of specific signals recorded during the movement, as well as information on the system, such as temperature, total power consumed, direction of manoeuvre, etc. An example of a typical signal acquired from a point machine is shown in Figure 8 below.

Features are extracted from the raw data acquired. The relevant features for the monitoring and the health assessment need to be extracted from the signals and data acquired, as the raw data contains the information in a hidden and not evident way. This can be seen in Figure 8, bottom, where a zoom on a signal shows how only a small part of the signal holds relevant information. During the feature extraction procedure, the information which is not relevant is kept aside, to reduce the amount of data to be analysed to a specific set of information. The feature extraction process is illustrated in Figure 8, top.

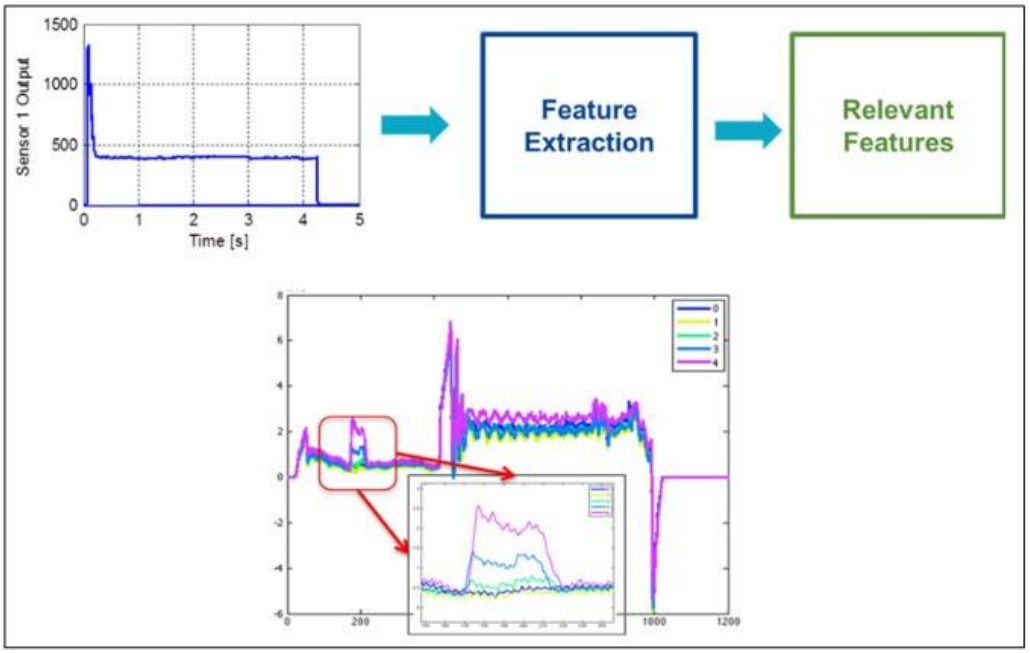


*Figure 8: Feature extraction process (above) and example of extracted feature (below)*

The State Detection phase aims at determining whether an acquired signal reflects a normal or abnormal behaviour. This is to promptly detect whether the asset being monitored may require more specific attention, or whether it is functioning normally. This process has an advantage with respect to the classical approach of arbitrary thresholds on signals, as it focuses on the relevant information acquired, i.e. the features, rather than all the information. Through this process, one can distinguish normal and abnormal conditions in a more precise manner, addressing the features which characterize the behaviour of the asset. This phase is illustrated in Figure 9.

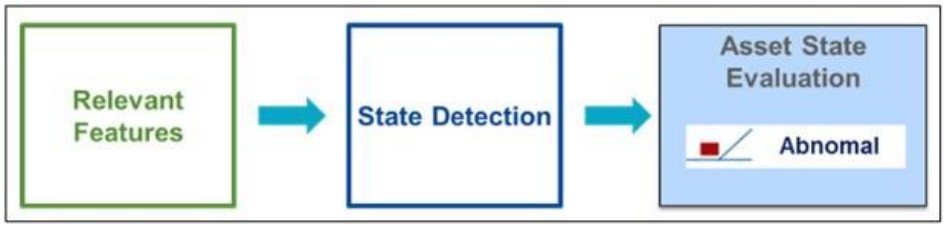


*Figure 9: Feature extraction process (above) and example of extracted feature (below)*

Health Assessment begins once the state of the asset is defined, to further understand and contextualize the detected state. This phase is done regardless of the asset state: if the state is detected to be normal, proceeding to the assessment of the health is beneficial as it allows for a historical overview which is of fundamental help for the end-user. The historical overview can help identify the first instances of deterioration easily, and aid in evaluating if an external or a common cause is possible. In this phase the health index is produced, which helps to quantify the behaviour of the asset and to diagnose what the degradation mechanism which is affecting the behaviour is. The diagnosis of the asset condition is an advantage, as it can help the end-user identify the target actions necessary to improve the functioning of the asset, and it can help identify common failure trends. This process is illustrated in Figure 10.


*Figure 10: Health Assessment Process*

In the Prognostic Assessment phase, the Remaining Useful Life (RUL) of the asset is evaluated. The remaining useful life allows for an estimation of the remaining time within which the end-user must perform a specific action to restore the asset health. This information is presented in the most convenient unit of measurement to the customer (e.g. working days, hours, operations, manoeuvres, etc.) and it is supported by a confidence interval, indicating the range of uncertainty in the prediction. This process is advantageous, as it helps to schedule the maintenance action with enough time delay before the failure begins to threaten the operation of the asset

Finally, in the Advisory Generation phase, the information processed and collected from the previous phases is combined to generate a work order. This work order is a specific action for a target asset with a specific time frame which can be used as input to include it in the maintenance schedule. The output of the work order can be modified to obtain possible suggestions as to the action required, it can be linked to the end-user maintenance manual in order to contain specific instructions as to how to proceed, and it can be connected to push-notifications (e.g. email, SMS, alerts, etc.) to have a real time monitoring of the machines without having to check the status continuously. This process combines the relevant information from the procedure and produces a simple output which is easily understandable. Another advantage is that it simplifies significantly the interpretation of the monitoring.

## 2.4 Adaptive Traffic Management

Urbalis® systems can be used with various types of regulation strategies: headway regulation (the goal being to keep a constant headway), or schedule regulation (the goal being to focus on departure and arrival times), or mixed schemes.

Perturbations of various kinds, caused not only by failures but also by passenger action, can potentially destabilise the system. For instance, at peak hour, passengers often block doors, causing the train to depart late from stations; such perturbations can propagate and lead to large delays in the absence of effective traffic management strategy.

Beyond rule-based regulation algorithms, which often have a purely local view, Alstom has developed goal-based algorithms, which aim at minimizing total system delay or other objective functions, subject to the various operational constraints. The control variables are station dwell time and speed between stations.

A simulation tool, based on stochastic Petri net modelling of the network and Monte Carlo simulations, has been developed [4] to compare different regulation strategies by means on their impact on key performance indicators (KPIs).

The most frequent KPIs are those of UITP:

- The Punctuality KPI defined by Equation 1; below
- The Regularity KPI defined by Equation 2; below
- The recovery time from an incident.

$$\text{Punctuality KPI} = \frac{\text{Number of train trips delayed by less than '}x\text{' minutes}}{\text{Actual number of train trips}} \text{ with } x = 60\text{s} \qquad \textit{Equation 1}$$

$$\text{Regularity KPI} = \frac{\text{Number of train departures at specified stations complying with planned headways within '}x\text{' minutes}}{\text{Actual train departures from specified stations}} \text{ with } x = 30\text{s} \qquad \textit{Equation 2}$$

The tool calculates the average value of those indicators corresponding to various regulation strategies and thus allows for selecting the best regulation algorithm.

# 3 CONCLUSION

Resilience aims at ensuring quick and efficient system recovery in response to a variety of aggressions, be they intrinsic or extrinsic.

Therefore, achieving resilience necessarily requires a variety of different risk mitigation measures.

It is an inherently multidisciplinary, system-wise endeavour which, in the case of Urbalis®, relies on both hardware and software and addresses the entire scope of design, operations and maintenance.

Perhaps a future direction for even further enhancement resides in the notion of 'learning system', with context-adaptive maintenance and operations policies and reconfiguration algorithms that would learn from past experience.

# 4 REFERENCES


1. Dersin P. & Valenzuela R., *Designing for Availability in Systems, and Systems of Systems*, Orlando, FL, USA, Annual Reliability & Maintainability Symposium (RAMS), 2019.

2. Dersin P. & Péronne A., *Probabilistic Characterization of Undetected Fault Time in a Redundant System with Imperfect Detection*, La Rochelle, France, Lambda-Mu18 Symposium, 2012.

3. Letot C., Dersin P., Pugnaloni M., Dehombreux P., Fleurquin G., Douziech C. & La-Cascia P., *A data driven degradation-based model for the maintenance of turnouts: a case study*, IFAC-PapersOnLine, vol. 48, pages 958-963, 2015.

4. Adeline B., Fabre E., Hélouët L., Kecir K. & Dersin P., *An efficient Evaluation Scheme for KPIs in Regulated Urban Train Systems*, International Conference on Reliability, Safety and Security of Railway Systems, Pistoia, Italy, RSSRAIL, 2017.

5. International Electrotechnical Commission (IEC), *Functional safety of electrical/electronic/programmable electronic safety-related systems – Part 6: Guidelines on the application of IEC 61508-2 and IEC 615083,* IEC 61508-6, Annex D, 2010

*6.* International Organization for Standardization (ISO), *Condition monitoring and diagnostics of machines Data processing, communication and presentation – Part 1: General guidelines,* ISO 13374, 2003

**7.** G. Neglia, S. Alouf, A. Dandoush, S. Simoens, P. Dersin, A. Tuholukova, J. Billion, & P. Derouet, *Performance Evaluation of Train moving Block Control*, Inria Sophia Antipolis, RR-8917, 2016