Health Monitoring and Prognostics

Health Monitoring exists for a purpose – raising business revenue. It can be used to deliver Business Value by increasing availability (avoiding unscheduled maintenance/shop visits, reducing planned shop visit frequency/matched to use) or improve the user experience (which in turn drives brand allegiance), but also, in certain circumstances, can be used to modify the user's behaviour to help achieve the customer's goal.

Historic Context

Traditional mechanical control applications selected electronic hardware based around the number of outputs that were needed to drive the solenoids/functions on the machine. This was the primary restricting factor. Adding new features from the software functionality viewpoint was not restricted since there was, in most cases, sufficient memory capacity.

As feature content has risen, it has outpaced the capability of the silicon vendors to keep up in the most demanding applications of electrical motor/generator control and automotive powertrain control. These additional demands have driven the number of sensing inputs and computational needs to break the real-estate and dissipation limits of the devices.

This has influenced systems towards distributed architectures, as traditional components (like motors, pumps and valves) take on intelligence, they allow the system architect to optimize performance, ensuring that power is produced and/or supplied only when needed, which reduces fuel consumption and emissions. This approach boosts performance while decreasing the cost of a central controller, and reduces the cost of 'shoe-horning' code into limited central capability, which can cause development effort to be disproportionately expensive.

The increased use of sensors for all mechanical and electrical systems provides opportunities for a major expansion of automatically generated prognostics, with contributions from intelligence (computers) collected from the (potentially distributed) system.

The increased availability of communications capacity (notably terrestrial wireless telephone and data networks) opens up significant opportunities for equipment designers. The use of "telematics" to connect products to service providers and increasingly with access to the cloud provides an enhanced opportunity for prognostics. Onboard ECUs (electronic control units) are not traditionally designed to do extensive data crunching prognostics because of their limited RAM, ROM, and CPU capabilities, and ubiquitous connectivity offers opportunities to get data out.

As we push our systems ever closer to their design limits with complex, non-linear, computationally intensive digital controls, we must ensure the machine stresses are well-managed, that we have design margin consistent with the error bands of our design and its environment and yet an overall approach to prognostics that ensures accuracy of detection of impending failure with minimum of false positives to avoid poor value delivery, or customer experience.

Prognostics in Automotive

Even before the age of digital electronic control systems, automotive architects were attempting to solve their customers' increasing sensitivity to the 'cost of ownership' rather than initial purchase price of a vehicle. This initially led to two trends:

i) reduced serviceable component costs and a burgeoning (and profitable) aftermarket

ii) development of long-life, non-serviceable parts

The latter element required a response to improve the designs in response to the causes of failure, and a better understanding of the mechanisms of degradation.

As electronic digital controls became more prevalent (1980s) the automotive standards progressively addressed these issues in attempt to harmonise the multitude of brand-specific monitoring and diagnostic service solutions.

To a modern vehicle user there are obvious indications of these efforts. Many consumable items have 'service indicators' propagated to the user based on condition e.g. brake pads, belts, lubricants (quantity, contamination), batteries, tyres, bulbs, fluids (levels and contaminants) and even larger items such as emissions control equipment. Additionally most owners are aware that service centres are able to diagnose failures, including intermittent issues, through Electronic Control Unit (ECU) supported 'trouble code' reports (through various evolutions of on-board diagnostic (OBD) standards, the most prominent being OBD-II/SAE J1979).

As vehicle systems have become more complex, owners are discouraged from all but mundane user-servicing (consumable levels such as washer-bottle refilling). We now accept monitoring of 'difficult to reach' components, or 'inconvenient' user actions (monitoring dirty oil levels) and in turn the manufacturers have altered part service life for those expensive to service/maintain due to cost of access, rather than cost of periodic service of part.

e.g. 'Buried' spark plugs – long-life, no longer maintenance item, individual coil also helps remove/reduce disturbance of 'fragile' HT components.

Most users are probably unaware that those same standards support predictive monitoring systems (e.g. OBDII – Mode 6 (analysis from all fault counters , even if not yet reached 'trouble code' (report) thresholds)) and that ECUs routinely are required to capture black-box style "Freeze Frame" data in support of trouble code reports... including 3rd party data loggers that mine this (private) data (e.g. The US company 'Progressive'). Users often do not correlate the service reminder bulletin, or post, that uses this service information and knowledge of usage patterns to target them with timely reminders. These items have been de-emphasised publicly because of privacy concerns (use of vehicle), including their potential use in accident investigation to support assignment of liability.

The most significant, but understated, use of this diagnostic information, is to allow designers to modify control strategies in support of reducing system 'stress' to improve vehicle reliability, or to psychologically modify the users behaviour, by proffering reduced costs of ownership through condition-based maintenance indicators.

Examples include "stress" reduction control strategies:

1. torque "choreography" to reduce load stress/cycle fatigue - Air-con / Alternator and especially its interaction with stop/start engine
2. Diesel Cold start – 'Heater plugs' heat head/block, add load to alternator for faster warm-up, helps catalytic converter life – not just aid combustion/reduce fuelling
3. Spying - Modifying 'delivery' (usually power or torque output) if 'demand' is abusive/compromises warranty (Note that the OBD-II data logging standards report both 'demands' and 'delivery' to service/fleet operations) – BUT big privacy issues!

Other examples include 'seamless' torque delivery (removing driveline 'shunt' (backlash)) to reduce wear, improve (vehicle) driveability and (vehicle stability)control e.g.

1. Acceleration - dual clutch transmission e.g. 'Tiptronic' (Porsche/ZF/Bosch)
2. Retardation - Hybrid EV – regenerative braking, friction braking and abs/stability control/anti-skid, artificially elevated alternator load for increased retardation /fuel saving
3. Traction control – from 'bang-bang' to 'slip and control' for off-road (e.g. hill descent) decreases transmission stress and improves vehicle control

A number of other Controls enabled strategies exist as 'cross-checks' for predictive and condition maintenance support, as the typical ECU sensor inputs have high-bandwidth to support control, but need to be significantly filtered responses to support diagnostics and prognostics. E.g. For fuel injection:

i) Abnormal operating point (N.B. Control may still compensate and remain operational but may imply MIL (Malfunction Indicator Lamp) illumination.)

   a. Fuelling <> torque production
   b. Emissions <> fuelling
   c. Combustion irregular/ inconsistent across cylinders (fuel leak detection/ loss of cylinder pressure)

   ...and in the latest generation fuel injection systems...

   d. Combustion temperature/pressure/noise modelling <> fuelling/load

At a vehicle level those cross-checks would include

i) Differential speeds (rpm) from plant through gearbox to axle and wheel speeds (which have defined relationships in non-CVT applications) – clutch/transmission/differential issues

ii) Differentiated torques – generated versus sum of consumption (hydraulic, electric e.g. alternator, PAS, HVAC, supercharger, PTO etc).

iii) Peak stress and duration (enables control -limited “over-specification” excursions) (e.g. Mercedes-Benz and Ford Power output ratings)

iv) Deviant performance degradation (Degradation modelling of plant torque)

v) Electric/Hydraulic Steering rack protection

  a. slamming into stops

  b. Kinematic levelling (stop hunting – accelerating/decelerating during a move that should be constant)

As Vehicle architectures become more complex with the advent of Hybrid and Full Electric Vehicle (EV) (including Plug-in) a number of other cross-checks have been developed as the relationship between vehicle characteristics and plant is no longer directly related through shaft distributed power.

Additionally, for EVs the ‘plant’ and its life characteristic (which drives the major cost of ownership) has required more detailed monitoring. E.g. Batteries

i) electrolyte levels,

ii) temperature,

iii) voltage/current relationships (discharge & charge),

iv) cell-to-cell matching,

v) estimated cell/battery charge storage capacity,

vi) sulphation and pollutants/electrolyte chemistry

Automotive de-centralisation of diagnostic information (Distribution)

As electronic/electrical actuation has become more complex, so intelligence surrounding faults has migrated, as integrated sensing and fault reporting, closer to the point of actuation, or at least its power driver. Usually these mechanisms are supported by intelligent chipsets, who are able to apply different configurable characteristics (slew rate limiting, current limiting), protection (thermal shutdown, short-circuit current limit) or signal, via alerts or by interrogation down a signalling bus, a number of pre-defined conditions (E.g. short circuit load to ground or supply, open circuit load) that are cognisant of the applied configured characteristic behaviour.

Control Sensor Suite or Dedicated Health Monitoring sensors?

Control Sensors

Using Controls System Sensors (Prognosis information synthesis) offers a cheap (Works Cost) solution to capturing information, but may be difficult to extract the appropriate detail due to large ‘noise’ relative to the wanted trend signal. Pre-conditioning/Filtering of the signal for control purposes may mean loss of information. An Automotive example:

- “knock” sensor – used in control strategy to detect start of combustion and synthesize injector opening point - traditionally signal ‘processed’ (4th to 8th order band-pass filter!) in knock sensor chip to give envelope of characteristic tuneable centre from 1 to 20KHz (typically ~7KHz) pre-ignition signal
- can also be used to ‘listen’ for cam wear, crank wear, mis-fuelling or poor/uneven/unbalanced combustion and even flywheel anomalies

Dedicated Sensors

The major advantage of dedicated sensing is that it can be close-coupled in its connection to the “Physics of Failure” (sensing matches expected mechanism of failure). The explicit understanding of how the failure is to be detected implicitly means that the onset of failure characteristic is understood and the signalling can be significantly decreased to surrounding just changes in that characteristic.

As with any sensing, dedicated sensing may include local conditioning which may be passive (e.g. for noise reduction), active (for improving signal strength or signal-to-noise ratio), or intelligent (to apply a complex filter, or extract a particular signature).

Engineering Issues

Detection

Engineers are frequently focussed on the physical components and their view of ‘what can be done’ probably based on existing sensor sets, rather than business reactive to ‘What needs to be monitored/detected’ to determine business value. This increases content (including data collected) but may not add value (unless you strike lucky).

The emphasis should always be on ‘what ails you’... and put in place a strategy for collecting data that allows you to diagnose the ailment. Look at your top ten causes for customer distress or downtime... have you got a detection mechanism for those problems?

Management

Most control systems manage a mechanical product. Inherently they have the ability to manage stress (where not in conflict with desired performance or fundamental operation). To do this successfully, they must be able to reliably detect the stressful situation (or have a well-defined set of boundary conditions from design or analysis). Product stress potentially reduces life, or at least will adapt your maintenance period in the wrong direction for value (unless you are in the business of selling routine maintenance)!

You can apply the same techniques to other stressful situations (e.g. step or shock load, oscillatory load, availability of torque, optimal operating point, compromise stability margin)

Diagnosis should always be aimed at Lowest Replaceable Unit (LRU) (aimed at Service initial needs i.e. returns, Logistics etc) . Return to base diagnostics should exercise the LRU sufficiently for full potential at diagnosing sub-LRU failures (e.g. motor winding short, seized actuator) and allow for any potential repair.

In our experience, most businesses are not yet apparently mature enough (or we lack confidence in the system's ability) to accurately and reliably identify the 'failed' or 'failing' LRU. Unfortunately the Control system, as the 'messenger' frequently gets the blame. See how many No Fault Found (NFF) Control System boards you have being returned!

Setting Thresholds

In most of our environments, the readings have an element of 'noise'. What works in a sterile lab environment for detection will often lead to over-exuberant false positives. Detuning the threshold on the other hand, may mean a faulty device is not flagged until late and the stress of that fault may create further stresses on other components.

Characterise your real Electrical Noise/Error Environment and use those results to design in some intelligent system of filtering that hides transient false positives, yet gives you sufficient early indication to not compromise the rest of the system.

Digital Control Systems in a Mechanical World

OK, so you recognise that prognostic indicators for digital systems are almost non-existent. The digital systems either work, or they are dead. If they are in any other state then you can't trust them to give a diagnosis anyway, so they are effectively dead. (So why do we still see so many control system ECUs as LRU returns?)

So when reading the ensuing chapters, be clear that you are looking for signs of the mechanical system under control degrading, not the digital components of the control system itself. Obviously the control sensors and control actuators, as physical systems with corresponding mechanisms that are measurable, should be within the 'mechanical system' scrutiny.

Condition-based maintenance

Plagiarised from Wikipedia and other sources and much modified!

Condition-based maintenance (CBM), shortly described, is maintenance when need arises. This maintenance is performed after one or more indicators show that equipment is going to fail or that equipment performance is deteriorating (usually to the extent where service would restore a better economic value).

Condition-based maintenance techniques were introduced to try to maintain the correct equipment at the right time. CBM is based on using real-time data to prioritize and optimize maintenance resources. Observing the state of the system is known as condition monitoring. Such a system will determine the equipment's health, and act only when maintenance is actually necessary. Ideally condition-based maintenance will allow the maintenance personnel to effect corrective repairs and other that might be economically justified by the overhead of the maintenance activity, minimizing spare parts cost, system downtime and time spent on maintenance.

Value potential

Equipment failures may not have a defined relationship to length of time in-service. Some failures are unpredictable, but may exude a ‘signature’ preamble before becoming either un-usable or causing disruption. Early detection allows cost effective repairs to be planned and disruptions to be minimised, whilst avoiding potentially unnecessary, and costly periodic preventative maintenance. Various studies have shown that failure is related to age and the stresses of the way it is used, which may be differentiated by application, environment or user. Getting more insight into the actual use enables improved design margins, and increasingly better understanding of the system’s transient responses rather than steady-state behaviour (albeit with potentially significant privacy issues). In some circumstances, gathering historical data is the only way to validate design limits... the margins and sensitivity required to deliver true value cannot be determined from simulation, this is evidenced by suppliers implementing ‘private’ fault counters, to establish ‘production’ limits that support an ‘acceptable’ level of false alarms.

As systems get more costly, and instrumentation and information systems tend to become cheaper and more reliable, CBM becomes an important tool for running a plant in an optimal manner. Better operations will lead to lower production cost and lower use of resources. Lower use of resources may be one of the most important differentiators in a future where environmental issues become more important by the day.

Good CBM systems recognize the implications (cost/benefit or risk) to the availability of the larger system and monitor critical components that are not economically maintained by other mechanisms, to ensure appropriate availability.

The most common cause of failing to achieve business benefit through CBM systems generally is due to the lack of cultural implementation within organisations.

Advantages and disadvantages

If appropriately targeted, CBM has some advantages over planned maintenance:

- Improved system reliability
- Decreased maintenance costs

- Decreased number of maintenance operations causes decreasing of human error influence
- Improved Customer experience

However the disadvantages are:

- High installation costs, for minor equipment items, it may be more than value of equipment
- Unpredictable maintenance periods cause costs to be divided unequally
- Increased number of parts (CBM installation) that need maintenance/checking and may affect overall system reliability

Targeting Condition Based Maintenance

I liked a lot of the stuff that Mike Sondalini at Lifetime Reliability Solutions has to say. So a lot of this explanatory text is 'borrowed' from their pages – check them out here http://www.lifetime-reliability.com/free-articles/maintenance-management/condition-based-maintenance.html

Starting from new, a part properly built and installed into equipment without any errors, will operate at a particular level of performance, which ideally is at its design requirement. As its operating life progresses degradation occurs. Regardless of the reasons for degradation, the item can no longer meets its original service requirements and its level of performance falls. By detecting the loss-in-condition of the item you have advanced warning that degradation has started. If you can detect this change in performance level you have a means to forecast a coming failure.

Figure 2 below represents the 'typical' degradation process experienced by equipment. Following a period of normal operation, where the item has been running smoothly, a change occurs that affects its performance. The earliest time where you can detect the degradation is called the P-point, or potential failure point, since after this time the item can potentially fail at any time. This change gradually, or rapidly in some cases, worsens to the point that the equipment cannot reliably and safely deliver its service duty. At this stage the item has functionally failed, i.e. it is not delivering its required performance, and the location is called the functional failure point or F-point. If it continues in operation the part will fail and the equipment will stop working, or worse may cause other consequential damage or compromise a large system.

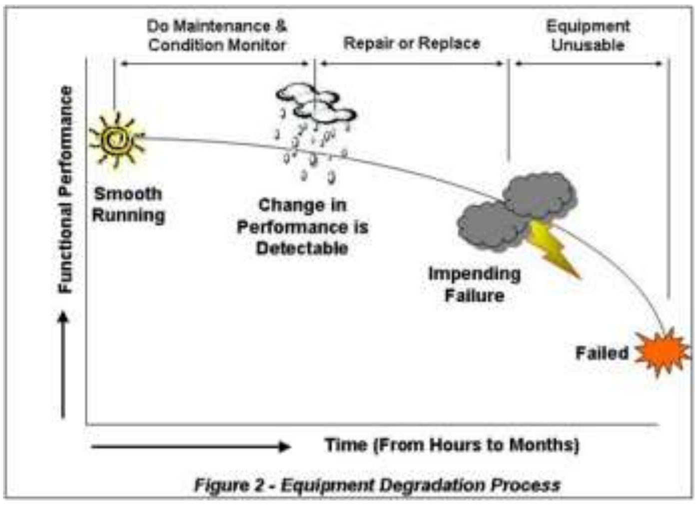

Figure 2 - Equipment Degradation Process

By using the 'tell-tale' evidence of changing equipment performance due to degradation, you can detect a failure and act to address it before an unplanned production disruption occurs. How soon you can detect the change in performance and spot the P-point depends on the condition monitoring technology that you use.

Inappropriate application of Condition-based and Predictive Maintenance solutions

Condition based maintenance can only be used on machines or materials that show deterioration in some way and over a time period that allows a maintenance response before they fail. A gearbox is a good example where we might expect vibration, overheating, noise as signs of deterioration before it grinds to a halt. A light bulb, although it has failure characteristics, fails over such a short timescale that it is an impractical subject for condition based maintenance (although a lighting system may be suitable).

The time between deterioration setting in and failure is known as the p-f interval. Clearly then, equipment failures whose p-f interval is abrupt when compared with the acceptable operation in a degraded state, are not amenable to Condition-based maintenance strategies. Similarly, equipment failures which have significant variation in the slope of the degradation during the p-f interval may also not be amenable to predictive maintenance strategies.

See **Don't Waste Your Time and Money with Condition Monitoring** (internet) or

The main thing condition monitoring does for you is to tell you that you have a problem in time to deal with it in a low cost way. It does not stop the problem! To reduce your maintenance costs, you must remove the failure mode. Having discovered a cause of failure through condition monitoring, you can use the feedback to improve the product strategy (or maintenance strategy) and improve the system life (and reduce maintenance costs still further).

Automotive OEM examples

BMW’s Condition Based Service concept for premium segment Service is scheduled based on the status of the vehicle (its condition). The vehicle checks continuously the status of parts, fluid levels and mileage data and makes this information available to drivers. The customer is prompted when service is needed. Vehicle data is saved in the key. ”TeleService” sends the data to the BMW service partner to plan service to be more effective.

Volvo Pilot project: Service Notification Objectives - To improve Customer Satisfaction - To improve planning and scheduling of service in the workshop - To reduce administration for service booking - To provide service for customers more rapidly. Service Notification Approx. 200 cars for 1 year Software added to phone module SMS every 1 000 km. Conclusion: Customers pleased with this way of booking services SMS from car to VCC back office. Service need assessment by server software. Customer gets email with web form. Web form filled in and sent to workshop. Workshop propose time for service. Customer confirms Service of car.

So next time your vehicle flags itself for maintenance, don’t be surprised when the service centre knows what it is about your driving style that has caused the failures, can tell you of the significant events and even the conditions prevailing at the time!

The military routinely don’t allow health monitoring data to leave their confines, as too much information may enable a deployment scenario to be established... well Big Automotive Brother is already watching you!