**Monograph:**

Marriott, S. and Harrison, R.F. (1998) A Novel Algorithmic Approach to the Integration of Posterior Knowledge into Condition Monitoring Systems. Research Report. ACSE Research Report 721 . Department of Automatic Control and Systems Engineering

# A Novel Algorithmic Approach to the Integration of Posterior Knowledge into Condition Monitoring Systems

**S. Marriott and R. F. Harrison**

**Research Report Number 721**

**August 1998**

Department of Automatic Control and Systems Engineering
The University of Sheffield
Mappin Street
Sheffield, S1 3JD, U.K.

Contact author: Dr. S. Marriott
E-Mail s.marriott@sheffield.ac.uk

The financial support of Rolls-Royce plc and the EPSRC (GR/L16651) is gratefully acknowledged

# A Novel Algorithmic Approach to the Integration of Posterior Knowledge into Condition Monitoring Systems:

**S. Marriott and R. F. Harrison**

The University of Sheffield, U.K.



Contact author: Dr. S. Marriott
E-Mail s.marriott@sheffield.ac.uk

The financial support of Rolls-Royce plc and the EPSRC (GR/L16651) is gratefully acknowledged

Indexing Terms: Condition monitoring. fault diagnosis. posterior knowledge integration.

**Abstract**

This paper considers the problem of the integration of 'posterior knowledge' into condition monitoring systems from both the theoretical and practical points of view. The work is presented in the context of aircraft engine maintenance. A methodology for updating posterior probabilities is proposed for cases where fault conditions are rejected or retained on the basis of external knowledge supplied by an end-user—the posterior knowledge. A possible fault class ranking is generated following the specification of fault class posterior probability functions. Context-free simulations are used to show the effect of posterior knowledge as part of a maintenance strategy. The simulations are independent of any specific condition-monitoring situation. Preliminary results indicate that posterior knowledge reduces the number of sub-unit inspections required for isolation of all faults. This has the potential to result in real maintenance cost savings.

## 1. Introduction

A common approach to identifying faults or conditions in dynamical systems is to develop a statistical classifier based upon historical data. Much effort—and possibly expense—can go into the development of such classifiers which form the basis of diagnostic systems. A set of fault-condition (FC) posterior probabilities is generated upon which diagnoses are made for example using maximum *a posteriori* (MAP) or Bayesian risk weighted decision criteria (e.g. Melsa and Cohn, 1978). In short, this is a *statistical* viewpoint on condition monitoring. When given a ranked set of FC probabilities representing the most likely FCs to have occurred, if the most probable FC is known not to have occurred then what should further decisions be based upon? Does it make sense always to choose the next most likely FC or set of FCs?

The knowledge that FCs have (or have not) occurred is deterministic, not available to the statistical classifier and is *specific* to the current situation. It cannot be made part of the historical data until the *complete* set of FCs is known for that particular input vector. Furthermore, the situation-specific data may become 'swamped' by the rest of the historical set in which it will be included. The main issue then, is the problem of integrating deterministic situation-specific data with historical, probabilistic data in a condition monitoring context.

This paper addresses the issue of the post-processing of condition monitoring information when external evidence is available to inform the fault diagnosis process. The key objective is to devise a mechanism for the integration of such evidence into predictive systems to allow the update of FC probabilities that have been generated without reference to that knowledge. Posterior knowledge integration has potentially widespread applications in the field of condition monitoring (fault detection and isolation) as explored in this paper. Incorporation of deterministic situation-specific knowledge about a monitored plant, not available in developing the condition monitoring system, will facilitate a more informed choice of maintenance strategy. Such a post-processing system could augment available condition monitoring systems which generate probabilistic data following fault classification by pattern recognition.

There is a growing interest in automated condition monitoring systems as the number and complexity of monitored plants increases to keep pace with the demands of modern technology. This interest is reflected in the number of fault detection and isolation methods appearing in the literature (e.g. Iserman, 1997; Rodd, 1997; Ruokonen, 1994). Such methods usually entail the monitoring of key system features—with or without a reference model—for pre-defined anomalies or novel operating conditions. A discussion of specific condition monitoring methods is not relevant to this paper. The emphasis of this work is the post-processing of probabilistic fault data regardless of the fault detection and isolation methods employed.

In general, condition monitoring systems are confined to the actual tasks of detecting and isolating faults and alerting an end-user to their possible existence and location. These systems may or may not give probabilistic estimates of FC probabilities to allow the end-user to decide an appropriate course of action. It is clear that such a methodology is "open-loop" in that end-users are given a final analysis, upon which to base operational decisions, without having the opportunity to feed their *observations* or their knowledge back into the process.

What if the end user has external information (not available to the condition monitoring system) which would alter specific fault diagnoses? It is obviously desirable to maximise the use of available information. The feedback of external information to a condition monitoring process makes it a "closed-loop" process as shown in Figure 1.

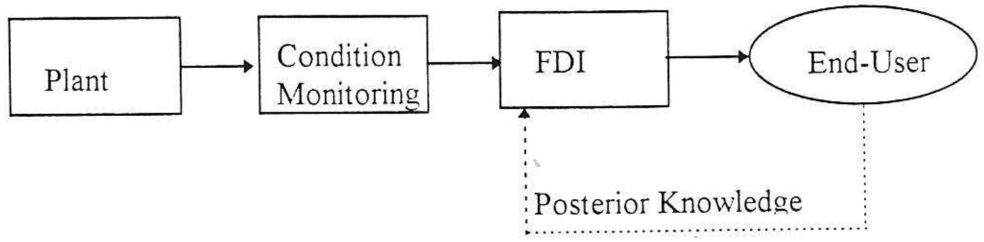



Figure 1. The condition monitoring feedback-loop. Posterior knowledge supplied by an end-user may be integrated into the condition monitoring process to improve FC isolation. The fault diagnosis and isolation (FDI) block is where the decisions are made.

In a condition monitoring situation, the end-user may say, "The condition monitoring system indicates the possibility of faults x,y and z. I have just checked y and can discount the possibility of a fault there. How does this affect the probability of faults x and z having occurred?".

The checking of y is not included in the monitored plant features and occurs after the condition monitoring system has made its predictions concerning possible fault scenarios. This external knowledge is given the name "posterior knowledge" to distinguish it from any other knowledge about the monitored plant. Posterior knowledge is knowledge about the outcome supplied by an operator, or some other source, and which is not available to the predictive system at the time of prediction. It is new evidence about the posterior probabilities which have been predicted for the current classification in the form of an updated output classification and differs from the new evidence about the state of the system which is typically encountered in sequential decision theory (e.g. Melsa and Cohn, 1978), i.e. updated evidence vectors. Posterior knowledge is deterministic; it is about *known* outcomes. Subjective probabilities could be used but are not considered here. The incorporation of this knowledge into the feedback loop of condition monitoring allows fault analysis to be adjusted towards a more accurate picture of the current plant status.

The main principle behind posterior knowledge integration is the fact that dynamical systems usually consist of a set of interconnected sub-systems which are causally related in some way. This means that information about a particular sub-system may have an effect on the prediction probabilities of other sub-system faults mediated by the causal connection, i.e. that multiple fault scenarios may be indicated simultaneously. This differs from the usual assumptions of exclusivity or of conditional independence made in conventional pattern recognition situations.

One particular application area for fault diagnosis and isolation methods is that of aircraft jet engines (e.g. Patton and Chen, 1997; Nairac *et al*, 1997). These are complex systems comprising distinct interacting sub-units which include electronic feedback control and monitoring devices. The posterior knowledge integration problem, as considered in this paper, is discussed in the context of aircraft jet engine monitoring. The integration of posterior knowledge into condition monitoring systems applied to jet engines is motivated by a need to reduce costly no fault found (NFF) conditions. NFF conditions occur when one or more faults are flagged and subsequent tests of sub-units fail to locate a problem. For example, a fault may be logged in-flight and when the plane lands at airport x, the supposedly faulty unit is replaced. The same fault is flagged during the onward flight and the unit is replaced at airport y. Subsequent analyses of both replaced units show no signs of malfunction because the alarm was, perhaps, owing to a faulty connection. However, the units have to be re-certified for future use which is a very expensive process. The generation of fault rankings—capable of being updated by posterior knowledge—may allow better-informed decisions about which sub-units and/or components to remove and test.

## 2. Posterior Knowledge Integration From an Engineer's Point of View

A condition monitoring system will typically provide an end-user with a set of predictions indicating one or more possible FCs. Merely choosing a single FC, on the basis of its associated probabilities, may be too simplistic. Furthermore, the end-user's knowledge may come to bear on the problem, as posterior knowledge, and be used to modify the original condition monitoring system diagnosis. A simple example will illustrate this (Marriott and Harrison, 1998a):

*A gas turbine vibration monitoring system has detected several features that correspond to one of three conditions: "Bearing wear in IP shaft" with probability 0.65, "Out-of-balance in LP compressor" with probability 0.20, and "Out-of-balance in HP compressor" with probability 0.15. However, the user knows from additional knowledge that a recent change of bearing rules out condition "Bearing wear in IP shaft". Is the most likely diagnosis now "Out of balance in LP compressor"?*[1]

If the above conclusions are based on *dependent* probability distributions then it may not be sufficient simply to redistribute the probabilities between FCs "*Out-of-balance in LP compressor*" and "*Out-of-balance in HP compressor*"; this issue will be discussed further in Section 4. Indeed it is possible that the suggestion *"Out-of-balance in LP compressor"* is based on vibration phenomena attributed to bearing wear that also produces the out-of-balance as a side-effect. Eliminating bearing wear as a possible diagnosis could remove the possibility of the LP out-of-balance. The engineer may, therefore, conclude that the correct diagnosis is "*Out-of-balance in the HP compressor*". This example illustrates some of the issues concerning the manner in which this posterior knowledge can be incorporated by the system for re-evaluation and future reference.

---

[1] Suggested by Dr. Steven King of Rolls-Royce plc. Applied Science Laboratory, Derby

## 3. Posterior Knowledge: Representation and Integration

It has been stated that posterior knowledge is deterministic knowledge external to the condition monitoring system. Two questions naturally arise from this: how can posterior knowledge be quantified and how is it to be integrated with the information contained within the condition monitoring system based upon the key plant features? This paper explores these two questions and then presents simulation evidence to show that posterior knowledge integration represents a technique with potential application in the condition monitoring field. As stated in the introduction, the key objective is to develop a method of automating the knowledge integration and updating process which follows logically from the fault detection and isolation tasks.

There are many possible ways of representing posterior knowledge. The representation problem is solved here by representing the posterior knowledge of possible system states and associated FCs as FC probabilities with discrete values of 1 or 0 depending upon whether a FC is known to occurred or not. Thus, although the posterior knowledge is deterministic, it is represented as a new set of FC probabilities, that is, a revised probability for each class which is influenced by external observations of the current situation only. The externally obtained information is then used to update the predicted FC ranking for the remaining probabilistic FCs. In other words, posterior knowledge about an FC is represented in the form of a probability indicating the occurrence or non-occurrence of that FC. In this way, deterministic data has been represented within a probabilistic framework. Note that this is not to be confused with the *posterior probability* of a fault occurring.

For example, a set of class posterior probabilities will be predicted for a single input datum (feature vector). If it is then possible to *exclude* one or more classes (i.e. the probability of those FCs occurring is zero) on the basis of knowledge or reasoning not available to the predictive system, then the current list of predicted FC probabilities may be revised. This will give a more accurate estimation of new FC *posterior* probabilities in the form of a revised probabilistic ranking. Similarly for the *inclusion* of known FCs.

Here, the integration of posterior knowledge is given in terms of classes which are known *not to have occurred* (excluded) or are known *to have occurred* (included) as indicated by the external knowledge. It is convenient to represent the updated posterior probabilities in terms of probabilities estimated from previous observations of system FCs, i.e. classes which *have* occurred; these probabilities are probabilities of *occurrence* and they can be estimated from empirical fault data. The limitation of probabilities involved in the representation of posterior knowledge to the class of null and certain events is a design choice made to facilitate implementation of the posterior knowledge integration process. This does not preclude the generalisation of the subsequent analysis to cases of subjective posterior knowledge between these extremes.

Three types of probabilities are involved in the posterior knowledge integration process:

1. the probability associated with the posterior knowledge that a FC has or has not occurred i.e. 1 or 0 representing the deterministic knowledge,
2. the exclusive (i.e. non-overlapping) singleton and joint probabilities generated by the condition monitoring system or FC frequencies of occurrence,
3. the (possibly) overlapping posterior probabilities of FCs reconstructed from the condition monitoring probabilities of (2).

These probabilities will be discussed further in the remainder of the paper.

In essence, the posterior knowledge integration process involves the modification of the FC posterior probabilities by the exclusion or inclusion of FC frequencies based upon the known occurrence or non-occurrence of FCs.

Both the inclusion and exclusion of FCs by posterior knowledge integration will be dealt with here. The revised posterior probabilities include the integration of external knowledge explicitly in the notation. The revised posterior FCs probabilities given posterior knowledge are represented by the expression $P\left(C_p \middle| \varepsilon \cap \mathbf{x}\right)$ where the posterior knowledge is given by $\varepsilon = \left(\bigcap_i C_{\delta_i}\right) \cap \left(\bigcap_e C^c_{\delta_e}\right)$ and $x$ is the feature vector. The subscripts $\delta_i$ and $\delta_e$ represent the indices of the sets of FC occurrences and non-occurrences respectively which comprise the posterior knowledge. The set of FCs occurrences and the set of non-occurrences are known as the *included* and *excluded* classes respectively. The expression $P\left(C_p \middle| \varepsilon \cap \mathbf{x}\right)$ denotes the predicted probability that FCs $p$ has occurred given the observation vector $x$ and posterior knowledge, $\varepsilon$.

## 4. The Probability Update Equation

The revised FC probabilities $P\left(C_p \middle| \varepsilon \cap \mathbf{x}\right)$ are computed using the *posterior probability update equation* which is given by

$$P\left(C_{\delta_r} \middle| \left(\bigcap_i C_{\delta_i}\right) \cap \left(\bigcap_e C^c_{\delta_e}\right) \cap \mathbf{x}\right) = \frac{P\left(C_{\delta_r} \cap \left(\bigcap_i C_{\delta_i}\right) \middle| \mathbf{x}\right) - P\left(\bigcup_e \left[C_{\delta_r} \cap \left(\bigcap_i C_{\delta_i}\right) \cap C_{\delta_e}\right] \middle| \mathbf{x}\right)}{P\left(\bigcap_i C_{\delta_i} \middle| \mathbf{x}\right) - P\left(\bigcup_e \left[\left(\bigcap_i C_{\delta_i}\right) \cap C_{\delta_e}\right] \middle| \mathbf{x}\right)} \qquad (1)$$

Equation (1) gives the new FC probability when posterior knowledge is available. It is based upon standard conditional probability theory. Thus, the new probabilities are computed given that some events are known to have occurred and others are known not to have occurred. The causal links between sub-systems are reflected in the joint probabilities present in the probabilities of set unions of the update equation which can be computed using the standard definition of set unions (e.g. Durret, 1994) with the appropriate terms substituted. This compact expression for the update of FC probabilities following the integration of posterior knowledge is derived formally in Appendix A. The union of the posterior probabilities of excluded classes is subtracted from the remaining (non-excluded) classes. Both the numerator and denominator of equation (1) represent what remains when any possibility of the excluded classes is removed. Note that all terms include the intersection with included classes (subscript $i$). This is because only those fault scenarios (involving multiple sub-units) including the faults known to have occurred are possible following posterior knowledge. Thus, the new probability of $P\left(C_{\delta_p}\middle|\mathbf{x}\right)$ given by Equation (1) is based upon the reduced sample space for the feature vector $\mathbf{x}$. Note that $\bigcup_k C_{\delta_k} = U$ (the universal set) for an exhaustive classification system where $k$ includes all class indices. The probabilities on the right hand side of Equation (1) are estimated from condition monitoring data. This is a non-trivial problem but is not considered further here. A brief description of the estimation problem is given in Section 7.

## 5. A Taxonomy of FCs: Some Rules.

Three specific exclusion or inclusion cases may be isolated from equation (1); these are, in increasing order of difficulty: Exclusive Class, Conditionally Independent Class, and Dependent Class exclusions or inclusions which reflect the division of FCs ( Marriott and Harrison, 1998 a, b). A simple renormalisation of posterior probabilities is only valid in restricted cases ( Marriott and Harrison, 1998 a,b). This accords with intuition in that exclusion or inclusion of FCs with dependent probability distributions may, in general, alter the position of classes in the ranking. Analysis of equation (1) indicates the conditions under which simple renormalisation is appropriate.

The following list of rules for the possible simplification of the probability update procedure is derived from the sixteen distinct cases arising from the non-existence, exclusivity, independence and dependence of excluded and included classes respectively. Table 1 shows graphically the outcomes for the sixteen cases. Derivations for the sixteen cases are found in Appendix B.

| | Included (i) N | E | I | D |
|---|---|---|---|---|
| Excluded (e) N | — | 0/1 | — | $P\left(C_{\delta_i} \middle| X_i \cap \mathbf{x}\right)$ |
| E | Renorm. | 0/1 | Renorm. | $P\left(C_{\delta_i} \middle| X_i \cap \mathbf{x}\right)$ |
| I | — | 0/1 | — | $P\left(C_{\delta_i} \middle| X_i \cap \mathbf{x}\right)$ |
| D | PUE (excluded) | 0/1 | PUE (excluded) | General |

Table 1. A table showing the sixteen possible combinations of included and excluded classes which can be non-existent, exclusive, independent or dependent respectively. These FCs are denoted by N, E, I, and D respectively. The '—' indicates that no changes are made to the remaining posterior probabilities. PUE (excluded) denotes the probability update equation under the given conditions of excluded FCs only. The intersection of included FCs is denoted by $X_i = \bigcap_i C_{\delta_i}$ . Renorm. indicates where renormalisation of probabilities is justified. General denotes the full equation, Equation (1).

### 5.1. *The Exclusive Included Class Rule* (EICR).

Where the *intersection* of the *included* classes is *exclusive*, the updated posterior probabilities will be one or zero depending upon whether the probabilities are in the included set (the intersection) or not.

In other words, if a set of exclusive events have occurred, then other, different, events cannot possibly occur, hence a probability of zero for those other events. See Figure 2.

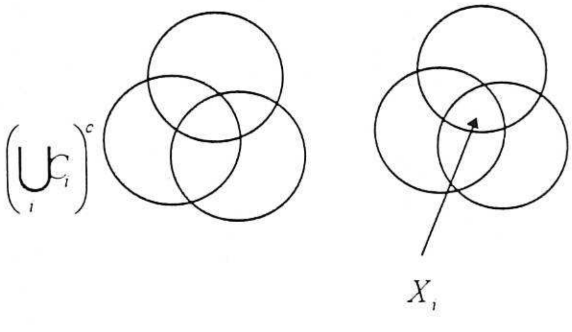
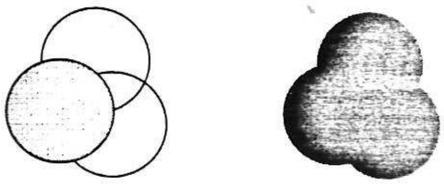



Figure 2. A schematic representation of the EICR. See text for details. The FC of interest is either included in $X_i = \bigcap_i C_{\delta_i}$ or not. If it is then the update posterior is 1, else it is 0 because of the exclusivity of the included classes

### 5.2. *The Exclusive Excluded Class Renormalisation Rule* (EECRR).

> Where the *included* classes are *non-existent* or *independent* and the union of the *excluded* classes is *exclusive,* the updated posterior probabilities of the remaining classes, following the exclusion of the set, will be given by a renormalisation of the remaining probabilities.

In other words, the excluded classes have no effect on the relative probabilities of the remaining FCs and can be removed giving a smaller probability space. See Figure 3.



$X_e^c$

$X_e$

Figure 3. A schematic representation of the EECRR. See text for details. The FC of interest is now conditional upon the remaining classes in $X_e^c$. Because there is no overlap, a simple renormalisation is justified.

### 5.3 *The Independence Rule* (IR).

> If either the *excluded* or *included* classes are *independent* of all other classes, then the remaining probabilities determined by other operations are unaffected.
>
> *Corollary*: If both the excluded and included classes are independent then there is no change to the posterior probabilities.

In other words, independent classes have no effect on the outcomes of the remaining FCs leaving the posterior probabilities unchanged.

### 5.4 *The Dependent Inclusive Class Rule* (DICR).

> The remaining posterior probabilities are conditional upon the *included* class probability.

In other words, there may be overlap between the remaining posterior probabilities and the included classes. The area within the overlap compared to the area of the included classes gives the conditional probability of the posterior of the FC of interest occurring. See Figure 4.

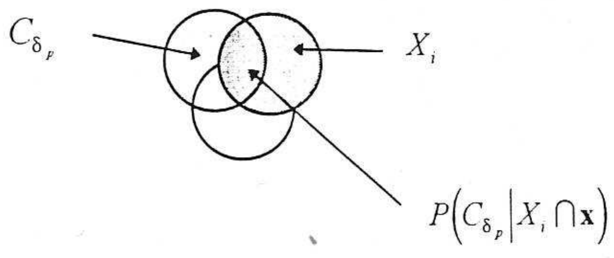



Figure 4. A schematic representation of the DICR. See text for details. The posterior probability of class $p$ occurring is now conditional upon the included FCs.

### 5.5. *The Dependent Excluded Class Rule* (DECR)

Where the included classes are non-existent or independent, the probability update equation is a special case for excluded classes only.

In other words it is as if there are no included sets involved in the posterior knowledge. See Figure 5.

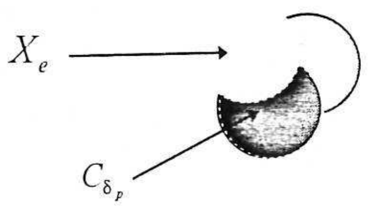



Figure 5. A schematic representation of the DECR. See text for details. Excluded classes are only of importance here. The probability of FC p occurring is proportional to ratio of the area of class p remaining to the whole remaining area. Note that it is not a simple renormalisation because some of the p th class is discounted owing to the joint probability "shared" with the excluded classes.

All the above rules are special cases of the general posterior probability update equation (1). It is to be noted that a simple renormalisation is not valid in *most* cases. This accords with intuition in that exclusion of FCs with dependent probability distributions will alter the position of FCs in the ranking. This idea will be illustrated in the next section.

### 6. Modified Fault Ranking Following Posterior Knowledge Integration

The original FC probabilities generated by the condition monitoring system are combined to give posterior probabilities of FC occurrence. The identified FC would be chosen on the basis of these posterior probabilities using whatever method such as MAP or weighted Bayes. Equation (1) is used to update the FC posterior probabilities.

We assume that a statistical model of the fault probability distributions is available via some estimation process (e.g. neural networks, mixture models etc.). The resulting model is fixed and does not give any information about how posterior knowledge is to be incorporated. This can lead to problems in complex situations where posterior knowledge may change the relative ranking of possible FCs. For example, the interrogation of a fixed classifier will provide a ranking of possible FCs based upon the computed posterior probabilities. If the indicated excluded faults are *exclusive* and the included faults are *non-existent* or *conditionally independent* of all other possible FCs, then a simple renormalisation of the probabilities of the remaining fault classes—following the exclusion of one or more faults on the basis of external information—is the obvious solution. Excluded *dependent* faults may, however, effect the fault ranking owing to interactions between faults; these interactions being mediated by causal connections between sub-units. This is illustrated in the following example (Marriott and Harrison, 1998 a,b):

A Gaussian three class problem was specified with the posterior probabilities as shown in Figure 6 (a). The classes in this synthetic problem might represent anomalous conditions such as *"Out-of-balance in LP compressor"*. Gaussian likelihoods are specified for the occurrences of FCs 1,2, and 3 alone, that is where a FC does not occur in conjunction with any other. Gaussian likelihoods are also specified for the joint events of classes 1 and 2 and classes 2 and 3. Priors are also specified for the classes. Using Bayes' theorem gives the posterior probabilities shown in Figure 6(b).

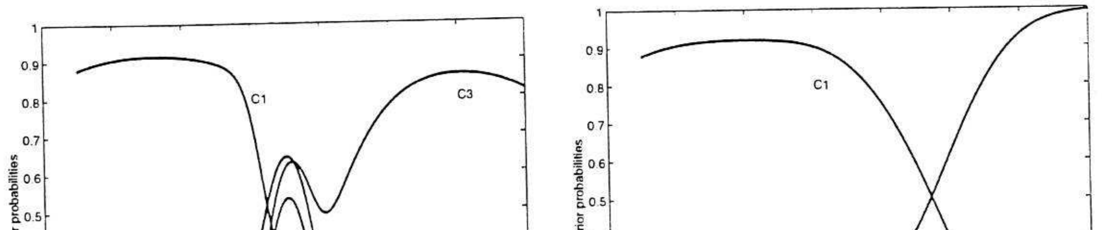



Figure 6. (a) The posterior probabilities for the three class example before exclusion of class 3. (b) the posterior probabilities following the exclusion of class 3 on the basis of posterior knowledge.

At the point $x = 5$ the posterior probabilities of fault occurrence prior to posterior knowledge are given by column 2 of Table 1 which shows the effect of external knowledge on the ranking of FCs. Given the *posterior* knowledge that class 3 is excluded, in this case, the updated posterior probabilities are given in column 3 of Table 2 and shown in Figure 6 (b). The exclusion of class 3 entails the removal of the likelihoods of class 3 alone and class 2 and 3. These revised probabilities have been calculated using Equation (1). Note that class 1 has risen to the top of the FC ranking following the integration of posterior knowledge into the probability adjustment process. A simple renormalisation would have placed class 2 at the top of the ranking which would have been incorrect.

The reason for the change in classification ranking following posterior knowledge is that faults C2 and C3 are very highly coupled at x=5 as shown by the posterior probability of 0.5343 for the two classes occurring together. At the point $x = 5$ the exclusion of C3 reduces the probability of C2 occurring by an amount significant enough to alter the class ranking. The joint probability distribution of C2&C3 accounts for a significant proportion of class C2 occurring at $x = 5$.

| FCs | Probs. prior to PK integration | Probs. following PK integration |
|---|---|---|
| $C_1$ | 0.3229 | 0.8518 |
| $C_2$ | 0.6444 | 0.2907 |
| $C_3$ | 0.6210 | — |
| $C_1 \cap C_2$ | 0.0540 | 0.2014 |
| $C_2 \cap C_3$ | 0.5343 | — |

Table 2. The fault class ranking before and after the integration of external evidence. The shaded box indicates the most probable FC.

## 7. The Estimation Problem

The unprocessed condition monitoring data will consist of monitored feature vectors with attached fault labels derived from a fault detection method. The integration of posterior information requires posterior probabilities to be estimated either directly, or indirectly from this data.

A common method of estimating posterior probabilities is to use an artificial neural network (e.g. Bishop, 1995; Richard and Lippmann, 1991). Where the FCs are exclusive, given N classes, there arises the 1 from N estimation problem, that is, for each input, one FC will be chosen on the basis of the posterior probabilities. Where the classes are non-exclusive, more than one FC can occur simultaneously giving rise to an *M* from *N* estimation problem. It has been shown (e.g. Bishop, 1995; Richard and Lippmann, 1991) that for both the mean squared error (MSE) and cross entropy (CE) measures, the neural networks will estimate the total Bayesian posterior probabilities of the form $P(C_i|\mathbf{x})$ only. Thus, although joint class information (*M* from *N*) is available in the training vectors, a conventional neural network classifier will not be able to estimate the joint probability function unless the output space is expanded to give an equivalent 1 from *many* problem. To capture class combination information in general, an augmented output vector consisting of $2^N$ outputs is required. The expansion is valid if the output space is treated as a collection of disjoint sets or *partition* (Halmos, 1974). The desired probabilities may then be reconstructed from the members of the partition.

Figure 7 shows the situation schematically. If the desired class training vector was $[1\ldots1\ldots1]'$ for example, joint class information would be available but would not be learned by the network. It is equivalent to having *N* decoupled networks with no correlation between the outputs, hence the *M* from *N* to a 1 from $2^N$ expansion at the worse possible case.

The whole process of posterior knowledge integration is shown schematically in Figure 8. Not all of the joint probabilities will be nonzero unless the worst case scenario occurs. The subset of relevant probabilities is chosen, forming a partition, and estimated using a neural network or other method. The fault scenario identification cycle is then entered.

The desired FC posteriors are reconstructed giving a ranking of sub-unit fault probabilities. This information is used to make a sub-unit inspection. If the fault scenario is identified—i.e. there are no other faults to be found—then the cycle ends. If the scenario is not yet identified, then the new inspection information is fed in as further posterior knowledge and the cycle continues. For the purpose of simulation, the number of fault in a scenario is known in advance to provide the stopping criteria.

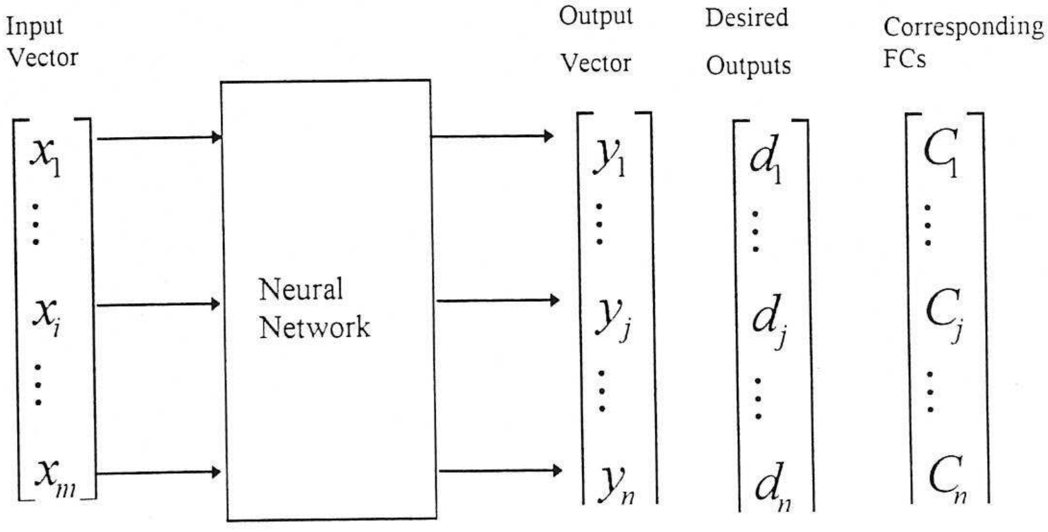



Figure 7. A schematic illustration of the process of using a neural network as a classifier. If a class is indicated, the relevant desired output is set to 1, otherwise it is left at 0.

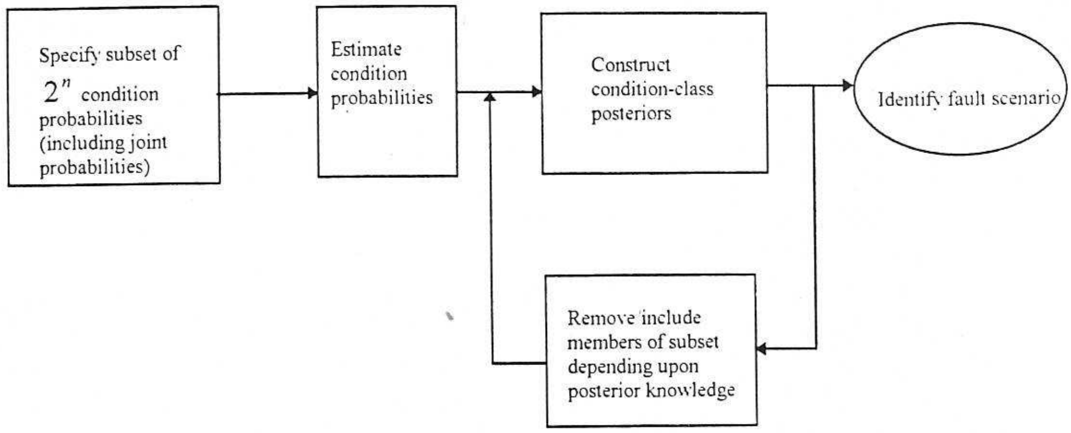



Figure 8. The posterior knowledge integration cycle. The posterior knowledge feedback occurs until all faults are isolated.

## 8. Assessing the Utility of Posterior Knowledge Integration

Of primary interest here is the development of more informed maintenance strategies which reduce the amount of maintenance required. By using posterior knowledge, revised fault probabilities will lead to a more efficient sub-unit checking order. Without posterior knowledge integration, the probability estimates—generated from the condition monitoring data—give a fixed fault *ranking* via the posterior probabilities of fault occurrences. The theory discussed so far indicates that revised posterior probabilities may alter the fault ranking and give a more accurate prediction of the current fault scenario. How can posterior knowledge integration enhance maintenance strategies in practice? Furthermore, how effective is the use of posterior knowledge and how can this be quantified? These questions and related issues will be explored in the remainder of this paper.

In general, in a condition monitoring situation, there will be a *search path* followed by maintenance engineers to detect and isolate all current faults. In terms of aircraft maintenance, this entails using all available fault indicators and maintenance experience/procedures to detect the faulty line-replaceable units (LRUs). The posterior knowledge integration technique has been developed to reduce search-path lengths during maintenance.

Posterior knowledge integration is an abstract technique designed, in theory, to be a post-processing stage with general applicability to a wide-range of condition monitoring techniques which produce probabilistic FC data. Consequently, the assessment of this technique should be, at least initially, *context-free*. In other words, its utility should be indicated without reference to a *specific* condition monitoring situation. A technique for context-free simulation has been developed for this purpose. Using context-free simulations means that the results are not limited to a specific set of FC relative frequencies. Using such a specific set may give a misleading impression of the possible utility of posterior knowledge integration.

Context-free simulations use a number of individual sets of relative frequencies to explore how the posterior knowledge technique functions across a range of conceivable condition monitoring situations. Each simulation is based upon a single set of relative frequencies generated at random; this set represents some possible set of condition monitoring data for a single plant such as an individual aircraft. By applying the integration technique to each of the relative frequency sets, an *ensemble* of results is obtained which can be summarised using appropriate ensemble measures. Performance measures will be discussed in Section 10. The ensemble results may represent many individual items of the same plant type (e.g. a fleet of aircraft) or, more generally as applied in this work, a heterogeneous set of plants. For the purposes of simulation, multiple instantiations of a plant may be characterised by using a narrow probability distribution for the relative frequency vectors. In other words, relative frequencies of individual plant items of the same basic type obviously do not vary to the extent of those of different plants.

## 9. A Single Simulation

For a single simulation of an ensemble, a set of FC frequencies are generated at random which represent a possible estimated set from the real-world. The real-world counterpart is shown schematically in Figure 9. Here, in the context-free simulations the input features do not exist because the FC frequencies may be associated with any possible input and represent some possible situation. The relative frequencies represent the probabilities of fault *scenarios* for the plant. An example scenario may be when sub-units (LRUs) 1 and 5 are faulty and all the rest are operating normally.

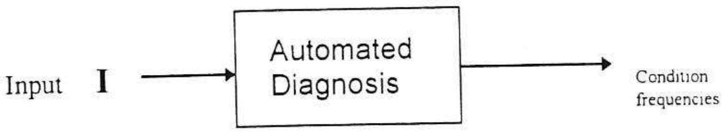



Figure 9. A schematic diagram of a "real-world" counterpart to a single context-free simulation. A set of observations is fed into a pattern recognition system which generates a set of FC frequencies corresponding to the predicted probabilities of fault scenarios. These probabilities will then be used to identify the actual fault scenario.

A fixed set of FC frequencies are generated for a given simulation from the ensemble. Each of the fault scenarios represented by a non-zero probability is taken in-turn as the actual fault scenario to be detected. The FC frequencies are used systematically—in conjunction with the probability update equation—to identify the actual fault scenario. Finally, after each scenario, relevant information is recorded which allows the single simulation measures ( e.g. for a single aircraft) or the ensemble measures (e.g. for a fleet) to be calculated. Each fault scenario of the individual simulation entails a fault search which results in a fault path representing the number of inspections required before all faults are identified.

Figure 10 illustrates what happens for a single fault scenario awaiting identification using posterior knowledge integration. The posterior probabilities of occurrence for each fault are reconstructed using the scenario frequencies. These posterior probabilities are ranked in descending order of magnitude and represent the probability of a given fault occurring. The sub-unit is chosen for inspection with the highest fault probability and inspected for that fault. The posterior knowledge following inspection is then used to specify the actual form of the probability update equation which depends upon the FCs to be included or excluded. The probability update equation is used to give the revised posteriors. If the fault scenario has been identified then the search is halted. Otherwise, the revised posteriors are used and the process is continued until the scenario is identified or the maximum number of sub-units is inspected.

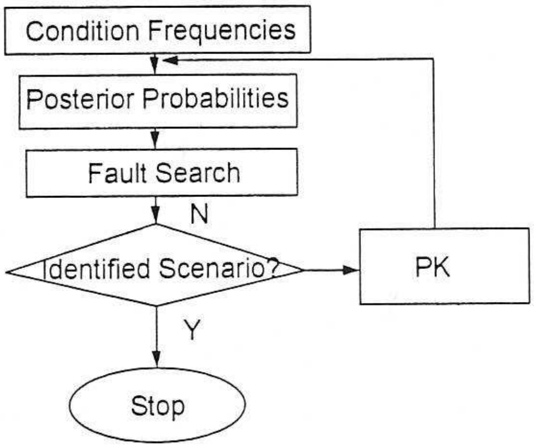


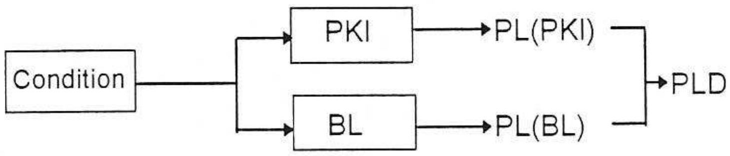

Figure 10. A flow diagram of the identification process for a single binary-coded fault scenario. A number of scenarios is identified for a given set of FC frequencies generated for a single simulation.

In order to show any possible utility of the posterior knowledge integration technique, a comparison has to be made with a baseline method which does not use this external evidence. A simple comparison method is to fix the FC posterior probabilities as they are obtained from the condition monitoring system. The fault search entails inspecting the sub-units in descending order which is equivalent to a simple renormalisation of the MAP decision. Here, the comparison technique is referred to as the *baseline method*.

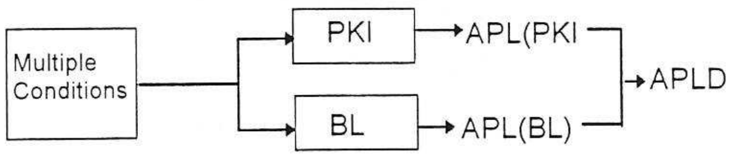
For comparison purposes, the posterior knowledge is included *systematically*, that is, the fault status of an inspected unit is fed into the probability update equation to give the new posteriors after each inspection. Both the posterior knowledge and baseline methods are carried out for each scenario until all faults are found, i.e. the scenario is identified.

## 10. Performance measures

A number of performance measures are possible which allow a comparison between methods. Of direct interest and applicability are measures involving the path length (PL) or *path length difference* (PLD) between the same scenario identified using the two techniques. The path length difference is the difference between the baseline and posterior knowledge integration techniques for a given scenario. A positive path length difference indicates that the baseline method took more steps (sub-unit inspections) to identify the scenario as compared to the PKI method and vice versa. Results are recorded for each scenario, each simulation and each ensemble for a given set of simulation FCs. Thus, in the context of aircraft maintenance, a set of measures is calculated for each situation (FC), each aircraft and each fleet. Figure 11 shows this schematically for a single FC and a single simulation.




(a)




(b)

Figure 11. (a) for a single FC, the posterior knowledge and baseline methods are applied to give path-lengths signifying the number of sub-unit (LRU) inspections required to identify the scenario. The path length difference is then calculated from the two path lengths. (b) For multiple FCs, the average path lengths and average path length differences may be calculated.

The path length difference ranges between $-(n-1)$ through 0 to $(n-1)$ where $n$ is the number of sub-units. One ensemble measure is to count the individual PLD's to give quasi-histograms of PLD frequencies. The counts are weighted to reflect the relative frequencies of the scenarios which gave rise to those path length differences. For a single scenario, the path length difference is calculated and weighted by the scenario probability. The weighted PLD counts are then presented in a bar chart as a quasi-histogram. Positive PLD's indicate that the baseline method requires more sub-unit inspections to identify the scenario. Thus, a quasi-histogram skewed in the positive direction indicates that the posterior knowledge method is more effective at scenario identification and requires fewer sub-unit inspections. This may translate into maintenance savings. In the aircraft industry, this would mean fewer LRU inspections. LRU's are usually removed and replaced which can be a costly process in terms of LRU recertification. NFF situations mean that non-faulty LRU's have to be tested thoroughly prior to re-use.

## 11. Simulations

One thousand simulations were carried out where each simulation generated a set of FC frequencies and applied the posterior knowledge and baseline methods to the scenarios within each simulation. For the simulations described here, a total of eight sub-units was used to represent a hypothetical plant; this gave 256 outcomes where the fault scenarios were represented by 8-bit binary strings. Here, the 256 fault scenarios were equally likely. Future work will consider different distributions of fault scenarios. Figure 12 shows a quasi-histogram for a set of simulations in which *all* 255 scenarios are possible which have one or more faults. The 256th scenario has zero faults.

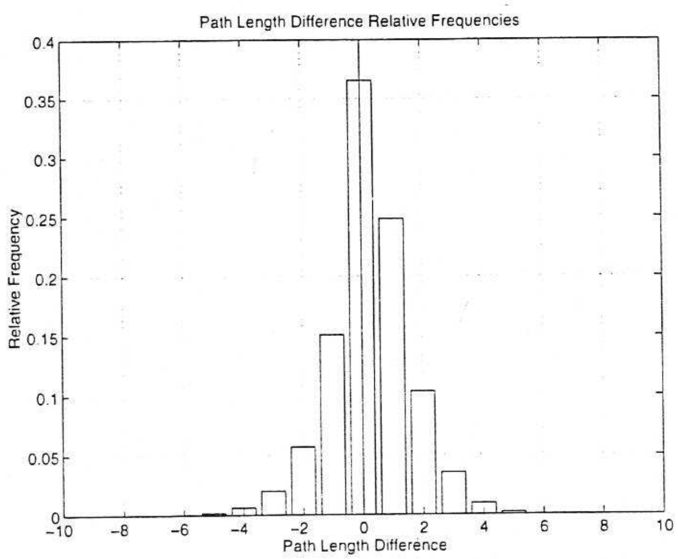



Figure 12 A quasi-histogram of path length differences over a thousand simulations. Each simulation provided a number of scenarios for identification using both the posterior knowledge and baseline methods. The histogram is skewed towards positive path length indicating that posterior knowledge integration results in shorter path lengths overall. The height of the bar indicates the relative frequency of occurrence for a particular PLD

Comparing related columns (PLD of same *magnitude*) such as +1 and -1 sub-unit inspections respectively reveals that the quasi-histogram is skewed towards positive path length differences, i.e. that positive PLDs are more likely.. This means that the sequential integration of posterior knowledge has reduced the number of required sub-unit inspections. As 255 fault scenarios were possible, the scenario predictions were maximally ambiguous, that is, when an actual scenario is to be identified, it can be any 1 of 255 possible scenarios.

Figure 13 shows the same protocol but with the number of scenarios reduced to any 64 out of 256. Comparison of figures 12 and 13 reveals that the quasi-histogram skewing is more pronounced—as expected—because the number of possible scenarios for a given diagnosis is reduced. In reality, the number of fault scenarios predicted for a given feature vector will invariably be lower than the maximum possible; the prior distribution of scenario frequencies will be dependent upon the dynamical system being monitored. For example it is conceivable that multiple fault scenarios will be much less prevalent than simple fault scenarios thereby reducing the number of fault scenarios associated with a given input. Furthermore, associations between fault scenarios and input vectors depend upon the key features monitored. If ambiguity is high, then it is likely that the choice of monitored features is not optimal for predictive disambiguation of fault scenarios.

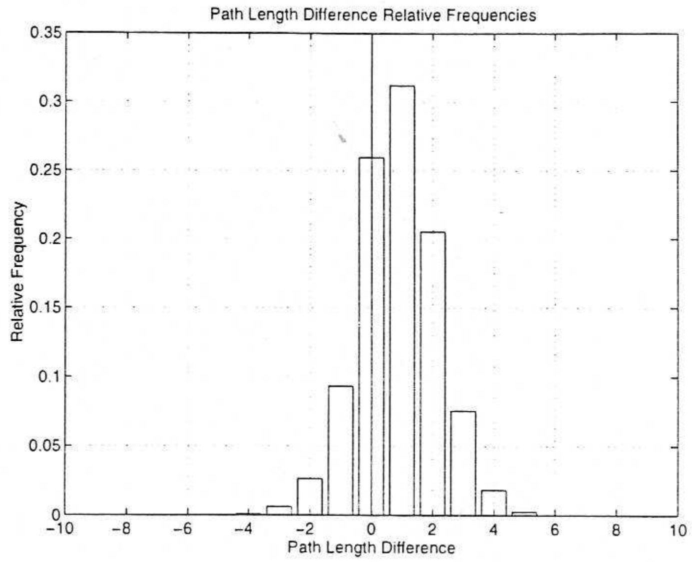



Figure 13. The results of the same simulations carried out using a reduced number of scenarios of 64 out of 256. The skewing is more pronounced. See text for details.

Figure 14 shows the same protocol again but this time with only 8 out of a possible 256 scenarios. Note that the skewing is even more pronounced than in the previous simulations. This is owing to a further substantial reduction in target scenario ambiguity.

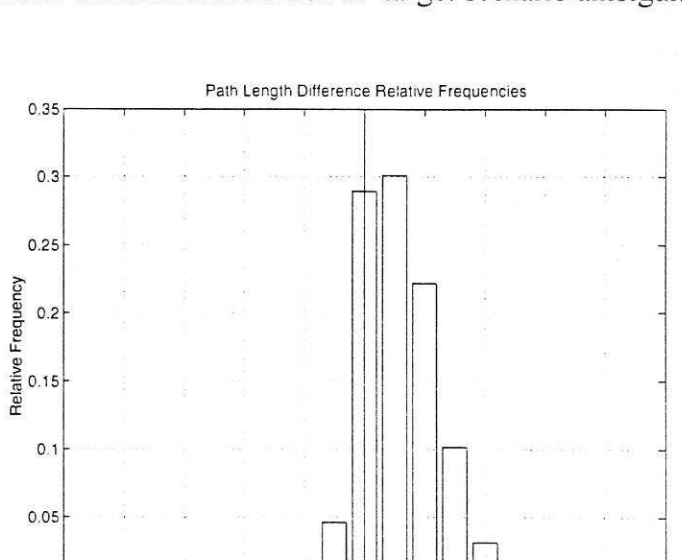



Figure 14. The results of the same simulations carried out but using a reduced number of scenarios of 8 out of 256. The skewing is more pronounced. See text for details.

Figure 15 shows the total path-length difference relative frequencies of the left and right sides of the quasi-histograms for an increasing number of scenarios. These sums represent the total relative frequencies of the left and right sides of the quasi-histograms. A larger positive PLD relative frequency indicates that the path-length for posterior knowledge integration is shorter than for the baseline method. Note that the positive PLD sum is consistently larger. As the number of scenarios allowed for each simulation increases, the difference between the positive and negative PLDrelative frequencies diminishes. The number of possible scenarios for a given input increases and represents an increase in predictive ambiguity. An expected consequence is that the effectiveness of the posterior knowledge technique diminishes. A high degree of predictive ambiguity is not expected in a "real world" situation because it would indicate a problem with fault resolution. A consistently larger *positive* PLD relative frequency indicates that the posterior knowledge integration technique always outperforms the baseline technique. The main point is that the posterior knowledge integration technique is superior even with high predictive ambiguity.

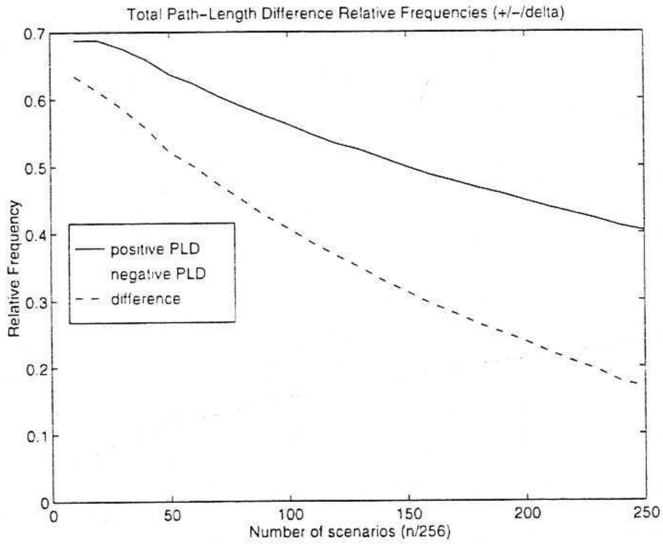



Figure 15. A graph of the PLD sum relative frequency versus the number of scenarios allowed for each simulation. The positive PLD sum relative frequency is a total relative frequency reflecting how many times the baseline method path length has exceeded that of the posterior knowledge integration technique (positive PLD) i.e. it the total relative frequency for the right-hand side of the histogram. This shows where the integration of posterior knowledge has shortened the path-length.

Where there are only singleton classes, i.e. no joint probabilities, there will be no gain using the PKI method. This is because no information is given about other FCs or sub-units.

## 12. Conclusions

In general, condition monitoring involves the detection of anomalous conditions that arise during the operation of some plant or process. Condition monitoring techniques usually end at the point of providing information about which sub-units of a given plant are suspected as being faulty. The indication of the most likely fault and its estimated probability by a fixed pattern recognition system is not necessarily the end-point. In reality, condition monitoring is a continuous, closed-loop process involving an end-user who ultimately decides how to use the information generated by the condition monitoring system. The end-user may, in turn, require a mechanism of incorporating his or her observations into the condition monitoring system for a more accurate diagnosis. The incorporation and utilisation of posterior knowledge presents a difficult problem. This paper has attempted both to articulate the problem and to provide a framework for its solution.

It has been demonstrated that posterior knowledge integration, as a post-processing technique, improves fault scenario identification. It is general in that it is applicable to condition monitoring systems which provide probabilistic fault scenario data. The end-user is able to feed back information into the condition monitoring process effectively, thus closing the loop. Context-free simulations provide a clear indication that, overall, posterior knowledge integration reduces path lengths in faulty sub-unit identification. This has potential payoffs in terms of maintenance costs, both direct and indirect. The skewing effect on the quasi-histograms is dependent upon the number of non-zero scenario probabilities. Here, the posterior knowledge integration was sequential, that is, was included after individual sub-units were inspected.

The above results are preliminary but they show that posterior knowledge integration has potential use in condition monitoring. Furthermore, the "closed-loop" method is independent of any predictive condition monitoring system. This stage follows on from the prediction of faults given a set of monitored features. The method requires a set of fault scenarios and their corresponding relative frequencies regardless of how they are estimated.

Now that the possible utility of the posterior knowledge integration technique has been demonstrated, a number of issues remain which have to be addressed.

The probability update equation has been applied to sets of FC frequencies as specified in the simulations. These FC frequencies determine both the initial fault scenario ranking and subsequent changes. In reality, the probabilities will be estimated from condition monitoring data and, as such, will be subject to estimation errors. The combined effects of these estimation errors may alter the scenario ranking and, consequently, change the maintenance strategy. The effect of estimation errors on the update equation must therefore be investigated.

At present, the path lengths are weighted only with respect to the scenario relative frequencies and are not weighted with respect to maintenance cost. In reality, the costs may rise significantly as time goes on; in the case of aircraft, foe example, long down-times can incur extra costs. The effect of cost weightings will be taken into account. Further weightings will also apply, e.g. the financial cost of replacing one LRU may be very much higher than replacing another. The simulations presented within this paper have equally weighted scenarios. This means that the prior probabilities of fault scenarios are the same and that scenarios with many faults are as equally likely as those with fewer faults. In practice, scenarios with multiple faults are less likely. This will be represented by using various prior distributions for the scenario frequencies.

The simulations presented here assume that the number of faults occurring in each simulation is known *a priori*. This is to ensure that performance comparisons between posterior knowledge integration and the baseline methods can be made. In the real-world, the number of faults will be unknown. Another possible benefit of posterior knowledge integration is that the modified probabilities may indicate whether or not it is sensible to search for other possible fault FCs. The baseline method will not supply any further information as to whether or not more faults remain. With posterior knowledge integration, a probability threshold may be used, below which any further search is terminated.

Posterior knowledge is currently included sequentially following each sub-unit inspection. In practice, information about one or more sub-units may be available prior to the fault search. A facility for "en masse" posterior knowledge integration will be included. There is also the possibility of bringing joint probability information into the system derived from engineering knowledge and practice, i.e. subjective probabilities other than simple 0/1.

The properties of exclusivity, independence may be used to pre-process the data before using the probability update equation. Furthermore, by identifying and simplifying dependencies, the probability estimation problem may be reduced.

## References


Durrett, R. (1994). *The Essentials of Probability* The Duxbury Press. Belmont California.

Halmos, P. R.(1974) *Naive Set Theory* Springer-Verlag

Isermann, R. (1997). Supervision, Fault-Detection and Fault-Diagnosis Methods-an Introduction, *Control Engineering Practice*: **5** (5) 639-651

Marriott, S, and Harrison, R. F. (1998a). The integration of posterior knowledge into statistical pattern recognition systems with particular application to fault diagnosis, *Proceedings of EIS'98* Proceedings Volume No. 3 (Artificial Intelligence) Pages 397 - 403

Marriott, S, and Harrison, R. F. (1998b) A Novel Approach to the Integration of Posterior Knowledge into Condition Monitoring Systems: Theory and Practice? *Proceedings of Control '98* Swansea, UK, In Press

Melsa, J. L. and Cohn, D. L. (1978) *Decision and Estimation Theory*. McGraw-Hill

Nairac, A, Corbett-Clark, T. A., Ripley, R., Townsend, N. W., and Tarassenko, L (1997), Choosing an appropriate model for novelty detection. *Proceedings of the 1997 5th International Conference on Neural Networks*, Cambridge UK

Patton, R. J. and Chen, J. (1997) Observer-Based Fault Detection and Isolation: Robustness and Applications, *Control Engineering Practice*, **5** (5) 671-682

Patton, R., J., Frank, P. M. and Clark, R. N. (eds.) (1989) *Fault Diagnosis in Dynamic Systems, Theory and Application* Control Engineering Series. Prentice hall, London

Rodd, M. G. (ed. in chief) (1997) *Control Engineering Practice*: Special section on fault detection and diagnosis. **5** (5)

Ruokonen, T (ed.) *IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes - SAFEPROCESS '94.*

## Appendix A. The Posterior Probability Update Equation

The general form of the posterior probability update expression can be written as $P\left(C_{\delta_p} \middle| \mathbf{x} \bigcap \varepsilon\left(C_{\delta_1} \ldots C_{\delta_k}\right)\right)$ where $\varepsilon\left(C_{\delta_1} \ldots C_{\delta_k}\right)$ is a function of the classes $C_{\delta_1} \ldots C_{\delta_k}$ and the three set operators of complement,union and intersection.

The probability update equation, Equation (1) can be proved formally as follows:

The general form can be expanded using the definition of conditional probability:

$$
\begin{aligned}
P\left(C_{\delta_p} \middle| \mathbf{x} \bigcap \varepsilon\left(C_{\delta_1} \ldots C_{\delta_k}\right)\right) &= \frac{P\left(C_{\delta_p} \bigcap \mathbf{x} \bigcap \varepsilon\left(C_{\delta_1} \ldots C_{\delta_k}\right)\right)}{P\left(\mathbf{x} \bigcap \varepsilon\left(C_{\delta_1} \ldots C_{\delta_k}\right)\right)} \\
&= \frac{P\left(C_{\delta_p} \bigcap \varepsilon\left(C_{\delta_1} \ldots C_{\delta_k}\right) \middle| \mathbf{x}\right)}{P\left(\varepsilon\left(C_{\delta_1} \ldots C_{\delta_k}\right) \middle| \mathbf{x}\right)} \qquad i \neq k
\end{aligned}
$$

The final expression can be rewritten in terms of positive probabilities. A linguistic analysis of the engineer's posterior knowledge regarding the occurrence or non-occurrence of faults can be coded to give an expression for $\varepsilon\left(C_{\delta_1} \ldots C_{\delta_k}\right)$ which, in turn can be expressed in terms of known (estimated) probabilities It is sensible to impose the restriction of $\varepsilon\left(C_{\delta_1} \ldots C_{\delta_k}\right)$ to expressions only involving complement and intersection. $\varepsilon\left(C_{\delta_1} \ldots C_{\delta_k}\right)$ will be of the form $\left[\left(\bigcap_a C_{\delta_a}\right) \bigcap \left(\bigcap_b C_{\delta_b}^c\right)\right]$.

Now,

$$
P\left(C_{\delta_p} \middle| \varepsilon \cap \mathbf{x}\right) = \frac{P\left(C_{\delta_p} \cap \varepsilon \cap \mathbf{x}\right)}{P(\varepsilon \cap \mathbf{x})} \qquad \text{(A1)}
$$

by the definition of conditional probability.

The external evidence is represented by the general form

$$
\varepsilon = X_i \cap X_e^c \qquad \text{(A2).}
$$

Substituting Equation (A1) into equation (A2) gives,

$$P\left(C_{\delta_p} \middle| X_i \cap X_e^c \cap \mathbf{x}\right) = \frac{P\left(C_{\delta_p} \cap X_i \cap X_e^c \cap \mathbf{x}\right)}{P\left(X_i \cap X_e^c \cap \mathbf{x}\right)}$$

$$= \frac{P\left(C_{\delta_p} \cap X_i \cap X_e^c \middle| \mathbf{x}\right) p(\mathbf{x})}{P\left(X_i \cap X_e^c \middle| \mathbf{x}\right) p(\mathbf{x})} \text{ by } P(A \cap B) = P(A|B)P(B)$$

$$= \frac{P\left(C_{\delta_p} \cap X_i \cap X_e^c \middle| \mathbf{x}\right)}{P\left(X_i \cap X_e^c \middle| \mathbf{x}\right)} \text{ by cancellation}$$

giving,

$$P\left(C_{\delta_p} \middle| X_i \cap X_e^c \cap \mathbf{x}\right) = \frac{P\left(C_{\delta_p} \cap X_i \middle| \mathbf{x}\right) - P\left(C_{\delta_p} \cap X_i \cap X_e \middle| \mathbf{x}\right)}{P\left(X_i \middle| \mathbf{x}\right) - P\left(X_i \cap X_e \middle| \mathbf{x}\right)} \qquad \text{(A3)}$$

by the identity $P\left(A \cap B^c \middle| \mathbf{x}\right) = P(A|\mathbf{x}) - P(A \cap B|\mathbf{x})$

The form of Equation (1) is obtained by substituting $X_e = \bigcup_e C_{\delta_e}$, $X_i = \bigcap_i C_{\delta_i}$ into Equation (A3) and applying the distributivity and associativity laws of set theory. The form of Equation (A3) is the most general. Appendix B gives details of the 16 special cases derivable from Equation (A3) by making the appropriate substitutions.

The probabilities of set unions of Equation (1) can be computed using the standard definition (e.g. Durret, 1994) with the appropriate terms substituted; the numerator term is given by the expansion

$$
\begin{aligned}
&P\left(\bigcup_e \left[ C_{\delta_p} \cap \left( \bigcap_i C_{\delta_i} \right) \cap C_{\delta_e} \right] \middle| \mathbf{x} \right) = \\
&\sum_e P\left( C_{\delta_p} \cap \left( \bigcap_i C_{\delta_i} \right) \cap C_{\delta_e} \middle| \mathbf{x} \right) - \\
&\sum_{e<e'} P\left( C_{\delta_p} \cap \left( \bigcap_i C_{\delta_i} \right) \cap C_{\delta_e} \cap C_{\delta_{e'}} \middle| \mathbf{x} \right) + \qquad \text{(A4)} \\
&\sum_{e<e'<e''} P\left( C_{\delta_p} \cap \left( \bigcap_i C_{\delta_i} \right) \cap C_{\delta_e} \cap C_{\delta_{e'}} \cap C_{\delta_{e''}} \middle| \mathbf{x} \right) - \\
&\vdots \\
&+(-)^{N_e+1} P\left( \bigcap C_{\delta_1} \cap C_{\delta_2} \cap \cdots \cap C_{\delta_{N_e}} \right)
\end{aligned}
$$

to include the conditional probabilities of Equation (1). Equation (A4) can be proved easily by using the distributivity of set relations and substituting $C_{\delta_p} \cap \mathbf{x}$ for $C_{\delta_p}$ in the general form of $P\left(\bigcup_{s=1}^{K} C_{\delta_s}\right)$ (e.g. Durrett, 1994,Grimmet and Stirzaker, 1992) where K is the number of sets involved in the union:

$$
\begin{aligned}
P\left(\bigcup_{s=1}^{K} C_{\delta_s}\right) = & \sum_{i=1}^{K} P\left(C_{\delta_i}\right) \\
& - \sum_{i<j}^{K} P\left(C_{\delta_i} \cap C_{\delta_j}\right) \\
& + \sum_{i<j<k}^{K} P\left(C_{\delta_i} \cap C_{\delta_j} \cap C_{\delta_k}\right) \\
& \vdots \\
& + (-1)^{K+1} P\left(C_{\delta_1} \cap C_{\delta_2} \cap \ldots \cap C_{\delta_K}\right)
\end{aligned}
$$

in terms of probabilities of occurrence. Note that conditional probability has been used.

## Appendix B.

The *Axiom of Exhaustivity*,

The universal set of conditions is exhaustively covered by a finite number of condition classes, i.e. $U = \bigcap_{k} C_{\delta_k}$ , $1 \le k \le n$ where n is the number of *distinct* condition classes. In other words, there are no other condition classes.

Where there are excluded and included classes which can be non-existent, exclusive, independent or dependent there are 16 possible cases as shown in Table B1:

| Case Number | Excluded | Included |
|---|---|---|
| 1 | N | N |
| 2 | N | E |
| 3 | N | I |
| 4 | N | D |
| 5 | E | N |
| 6 | E | E |
| 7 | E | I |
| 8 | E | D |
| 9 | I | N |
| 10 | I | E |
| 11 | I | I |
| 12 | I | D |
| 13 | D | N |
| 14 | D | E |
| 15 | D | I |
| 16 | D | D |

Table B1. The 16 possible cases where excluded and included classes are possibly non-existent, exclusive, independent or dependent. These states are denoted by N, E, I and D respectively Note that it assumes that *all* excluded or included classes have the same state.

The cases of Table B1 will be dealt with in order.

**Case 1: (N/N)** there are *no* excluded or included classes

$$P_{\varepsilon}\left(C_{\delta_p}\middle|\mathbf{x}\right)=P\left(C_{\delta_p}\middle|\mathbf{x}\right)$$

where

$$P_{\varepsilon}\left(C_{\delta_p}\middle|\mathbf{x}\right)\equiv P\left(C_{\delta_p}\middle|\varepsilon\cap\mathbf{x}\right)$$

for notational convenience throughout the derivations.

In this case, no evidence available implies no change as expected.

**Case 2 (N/E)** there are *no* excluded classes but the included classes are *exclusive*.

Substituting $X_e = \phi$ into the PUE gives

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = \frac{P\left(C_{\delta_p} \cap X_i \middle| \mathbf{x}\right) - P\left(C_{\delta_p} \cap X_i \cap X_e \middle| \mathbf{x}\right)}{P\left(X_i \middle| \mathbf{x}\right) - P\left(X_i \cap X_e \middle| \mathbf{x}\right)}$$

$$= \frac{P\left(C_{\delta_p} \cap X_i \middle| \mathbf{x}\right)}{P\left(X_i \middle| \mathbf{x}\right)}$$

Now, by the axiom of exhaustivity, and exclusivity of the excluded classes:
$C_{\delta_p} = C_{\delta_i}$ for some $i \in \Delta I$ or $C_{\delta_p} = C_{\delta_k}$ for some $k \in \{1,\ldots,n\}$, $k \notin \Delta I$.

Where $C_{\delta_p}$ is not an included class:

$C_{\delta_p} = C_{\delta_k}, k \in \{1,\ldots,n\}$, $k \notin \Delta I \Rightarrow C_{\delta_p} \cap X_i = \phi$ giving $P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = 0$

which is the trivial case. In other words, $C_{\delta_p}$ will not happen.

For the former case, $C_{\delta_p}$ is an included FC so it has happened.

$C_{\delta_p} = C_{\delta_i}$ $i \in \Delta I \Rightarrow C_{\delta_p} \cap X_i = X_i$ giving $P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = 1$

Stated simply, case 2 entails that the included classes are exclusive which means that if exclusive classes are known to have occurred, the on the classes are ruled out. Hence a probability of one or zero depending upon whether the desired class is amongst those known to have occurred or not.

**Case 3 (N/I)** there are *no* excluded classes but the included classes are *independent*.

With no excluded classes, substituting $X_e = \phi$ into the PUE gives

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = \frac{P\left(C_{\delta_p} \cap X_i \middle| \mathbf{x}\right) - P\left(C_{\delta_p} \cap X_i \cap X_e \middle| \mathbf{x}\right)}{P\left(X_i \middle| \mathbf{x}\right) - P\left(X_i \cap X_e \middle| \mathbf{x}\right)}$$

$$= \frac{P\left(C_{\delta_p} \cap X_i \middle| \mathbf{x}\right)}{P(X_i|\mathbf{x})}.$$

Now, by conditional independence.

$$P\left(C_{\delta_p} \cap X_i \middle| \mathbf{x}\right) = P\left(C_{\delta_p} \middle| \mathbf{x}\right)P(X_i|\mathbf{x})$$

so,

$$P_\varepsilon\left(C_{\delta_p} \middle| \mathbf{x}\right) = \frac{P\left(C_{\delta_p} \middle| \mathbf{x}\right)P(X_i|\mathbf{x})}{P(X_i|\mathbf{x})}$$
$$= P\left(C_{\delta_p} \middle| \mathbf{x}\right)$$

When the included classes are independent, it follows that the occurrence of these classes will have no effect on the occurrence of the remaining classes.

**Case 4 (N/D)** there are *no* excluded classes but the included classes are *dependent*.

Again, with no excluded classes,

$$P_\varepsilon\left(C_{\delta_p} \middle| \mathbf{x}\right) = \frac{P\left(C_{\delta_p} \cap X_i \middle| \mathbf{x}\right)}{P(X_i|\mathbf{x})}$$

which cannot be simplified in any way unless $C_{\delta_p} = C_{\delta_i} \quad i \in \Delta I \Rightarrow C_{\delta_p} \cap X_i = X_i$ giving $P_\varepsilon\left(C_{\delta_p} \middle| \mathbf{x}\right) = 1$ which is trivial. For the non-trivial case, the included classes are dependent and restrict the probability space of $C_{\delta_p}$ occurring.

Events are restricted to those in the included class intersection, $X_i = \bigcap_i C_{\delta_i}$.

Now,

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right)=\frac{P\left(C_{\delta_p}\cap X_i\middle|\mathbf{x}\right)}{P\left(X_i\middle|\mathbf{x}\right)}$$

$$=\frac{P\left(C_{\delta_p}\cap X_i\cap\mathbf{x}\right)/P(\mathbf{x})}{P\left(X_i\cap\mathbf{x}\right)/P(\mathbf{x})}$$

$$=\frac{P\left(C_{\delta_p}\cap X_i\cap\mathbf{x}\right)}{P\left(X_i\cap\mathbf{x}\right)}$$

$$=P\left(C_{\delta_p}\middle|X_i\cap\mathbf{x}\right)$$

Thus the posterior probabilities are now conditioned on the included class space.

**Case 5 (E/N)** the excluded classes are *exclusive* but there are *no* included classes. Where there are no included classes, $X_i = U$. Substituting into the PUE gives,

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right)=\frac{P\left(C_{\delta_p}\middle|\mathbf{x}\right)-P\left(C_{\delta_p}\cap X_e\middle|\mathbf{x}\right)}{P\left(U\middle|\mathbf{x}\right)-P\left(X_e\middle|\mathbf{x}\right)}$$

By applying the axiom of exhaustivity, and the exclusivity of the excluded condition classes,

$C_{\delta_p}\cap X_e = C_{\delta_p}$ or $C_{\delta_p}\cap X_e = \phi$. The former case is trivial giving $P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right)=0$

Where, $C_{\delta_p}\cap X_e = \phi$,

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right)=\frac{P\left(C_{\delta_p}\middle|\mathbf{x}\right)}{P\left(U\middle|\mathbf{x}\right)-P\left(X_e\middle|\mathbf{x}\right)}.$$

Because the excluded classes are exclusive, the probability space is reduced without information affecting the ratios of the remaining classes. For example, if the class 1 and class 2 probabilities are in the ratio 2:1 then the ratio remains at 2:1 following the exclusion of the exclusive classes. Exclusive classes do not carry any joint information about the remaining classes.

**Case 6 (E/E)** both the excluded and included classes are *exclusive*.

The exclusiveness of the excluded classes gives

$X_i \cap X_e = \phi$, giving

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = \frac{P\left(C_{\delta_p} \cap X_i \middle| \mathbf{x}\right)}{P\left(X_i \middle| \mathbf{x}\right)}.$$

Now, by the axiom of exhaustivity and the exclusivity of the included classes,

$C_{\delta_p} = C_{\delta_i}$ for some $i \in \Delta I$ or $C_{\delta_p} = C_{\delta_k}$ for some $k \in \{1,\ldots,n\}$, $k \notin \Delta I$.

Where, $C_{\delta_p} = C_{\delta_k}$, $k \in \{1,\ldots,n\}$, $k \notin \Delta I \Rightarrow C_{\delta_p} \cap X_i = \phi$ giving $P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = 0$, which is the trivial case. In other words, $C_{\delta_p}$ has not been included.

For the former case, $C_{\delta_p} = C_{\delta_i}$ $i \in \Delta I \Rightarrow C_{\delta p} \cap X_i = X_i$ giving $P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = 1$

Case 6 is similar to case 2 in that the exclusivity of the included classes precludes any other events. Similarly with cases 10 and 14.

**Case 7 (E/I)** the excluded classes are *exclusive* but included classes are *independent*.

The exclusiveness of the excluded classes gives

$X_i \cap X_e = \phi$, giving

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = \frac{P\left(C_{\delta_p} \cap X_i \middle| \mathbf{x}\right)}{P\left(X_i \middle| \mathbf{x}\right)}.$$

For the trivial case, $C_{\delta_p} = C_{\delta_k}, k \in \{1,\dots,n\}, \quad k \notin \Delta I \Rightarrow C_{\delta_p} \cap X_i = \phi$ giving $P_\varepsilon\left(C_{\delta_p} \middle| \mathbf{x}\right) = 0$.

Now, by conditional independence

$$P\left(C_{\delta_p} \cap X_i \middle| \mathbf{x}\right) = P\left(C_{\delta_p} \middle| \mathbf{x}\right) P\left(X_i \middle| \mathbf{x}\right)$$

so,

$$\begin{aligned} P_\varepsilon\left(C_{\delta_p} \middle| \mathbf{x}\right) &= \frac{P\left(C_{\delta_p} \middle| \mathbf{x}\right) P\left(X_i \middle| \mathbf{x}\right)}{P\left(X_i \middle| \mathbf{x}\right)} \\ &= P\left(C_{\delta_p} \middle| \mathbf{x}\right) \end{aligned}$$

**Case 8 (E/D)** the excluded classes are *exclusive* but included classes are *dependent*.

The exclusiveness of the excluded classes gives

$X_i \cap X_e = \phi$, giving

$$P_\varepsilon\left(C_{\delta_p} \middle| \mathbf{x}\right) = \frac{P\left(C_{\delta_p} \cap X_i \middle| \mathbf{x}\right)}{P\left(X_i \middle| \mathbf{x}\right)}.$$

which cannot be simplified in any way except for the trivial case, $C_{\delta_p} = C_{\delta_i}$ $i \in \Delta I \Rightarrow C_{\delta_p} \cap X_i = X_i$ giving $P_\varepsilon\left(C_{\delta_p} \middle| \mathbf{x}\right) = 1$ which is trivial. For the non-trivial case, the included classes are dependent and restrict the probability space of $C_{\delta_p}$ occurring.

**Case 9 (I/N)** the excluded classes are *independent* but there are *no* included classes.

Where there are no included classes,

$X_i = U$.

Substituting into the PUE gives,

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = \frac{P\left(C_{\delta_p}\middle|\mathbf{x}\right) - P\left(C_{\delta_p} \cap X_e\middle|\mathbf{x}\right)}{P(U|\mathbf{x}) - P(X_e|\mathbf{x})}$$

Now, by conditional independence,

$$P\left(C_{\delta_p} \cap X_e\middle|\mathbf{x}\right) = P\left(C_{\delta_p}\middle|\mathbf{x}\right)P(X_e|\mathbf{x})$$

giving,

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = \frac{P\left(C_{\delta_p}\middle|\mathbf{x}\right) - P\left(C_{\delta_p}\middle|\mathbf{x}\right)P(X_e|\mathbf{x})}{P(U|\mathbf{x}) - P(X_e|\mathbf{x})}.$$

$$P(X_e|\mathbf{x}) = 0$$

giving,

$$\begin{aligned} P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) &= \frac{P\left(C_{\delta_p}\middle|\mathbf{x}\right)}{P(U|\mathbf{x})} \\ &= P\left(C_{\delta_p}\middle|\mathbf{x}\right) \end{aligned}$$

**Case 10 (I/E)** the excluded classes are *independent* but included classes are *exclusive*.

Where the exclusive classes are independent,

$$P\left(C_{\delta_p} \cap X_e\middle|\mathbf{x}\right) = P\left(C_{\delta_p}\middle|\mathbf{x}\right)P(X_e|\mathbf{x})$$

giving,

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right)=\frac{P\left(C_{\delta_p}\cap X_i\middle|\mathbf{x}\right)-P\left(C_{\delta_p}\cap X_i\middle|\mathbf{x}\right)P\left(X_e\middle|\mathbf{x}\right)}{P\left(X_i\middle|\mathbf{x}\right)-P\left(X_i\middle|\mathbf{x}\right)P\left(X_e\middle|\mathbf{x}\right)}$$

$$=\frac{P\left(C_{\delta_p}\cap X_i\middle|\mathbf{x}\right)}{P\left(X_i\middle|\mathbf{x}\right)}$$

Now, by the axiom of exhaustivity and exclusivity of the included condition classes, $C_{\delta p} = C_{\delta_i}$ for some $i \in \Delta I$ or $C_{\delta_p} = C_{\delta_k}$ for some $k \in \{1,\ldots,n\}$, $k \notin \Delta I$.

$C_{\delta_p} = C_{\delta_k}, k \in \{1,\ldots,n\}$, $k \notin \Delta I \Rightarrow C_{\delta_p} \cap X_i = \phi$ giving $P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = 0$ which is the trivial case.

For the former case,

$C_{\delta p} = C_{\delta_i}$ $i \in \Delta I \Rightarrow C_{\delta p} \cap X_i = X_i$ giving $P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = 1$

**Case 11 (I/I)** both the excluded and included classes are *independent*.

Where the exclusive classes are independent,

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right)=\frac{P\left(C_{\delta_p}\cap X_i\middle|\mathbf{x}\right)}{P\left(X_i\middle|\mathbf{x}\right)}.$$

For the included classes, by conditional independence

$$P\left(C_{\delta_p}\cap X_i\middle|\mathbf{x}\right)=P\left(C_{\delta_p}\middle|\mathbf{x}\right)P\left(X_i\middle|\mathbf{x}\right)$$

so,

$$P_{\varepsilon}\left(C_{\delta_p}\middle|\mathbf{x}\right)=\frac{P\left(C_{\delta_p}\middle|\mathbf{x}\right)P\left(X_i\middle|\mathbf{x}\right)}{P\left(X_i\middle|\mathbf{x}\right)}$$
$$=P\left(C_{\delta_p}\middle|\mathbf{x}\right)$$

**Case 12 (I/D)** the excluded classes are *independent* but included classes are *dependent*.

Where the exclusive classes are independent,

$$P_{\varepsilon}\left(C_{\delta_p}\middle|\mathbf{x}\right)=\frac{P\left(C_{\delta_p}\cap X_i\middle|\mathbf{x}\right)}{P\left(X_i\middle|\mathbf{x}\right)}.$$

which cannot be simplified in any way except for the trivial case, $C_{\delta_p}=C_{\delta_i}$
$i\in\Delta I\Rightarrow C_{\delta_p}\cap X_i=X_i$ giving $P_{\varepsilon}\left(C_{\delta_p}\middle|\mathbf{x}\right)=1$ which is trivial. For the non-trivial case, the included classes are dependent and restrict the probability space of $C_{\delta_p}$ occurring.

**Case 13 (D/N)** the excluded classes are *dependent* but there are *no* included classes.

Where there are no included classes,

$X_i=U$.

Substituting into the PUE gives,

$$P_{\varepsilon}\left(C_{\delta_p}\middle|\mathbf{x}\right)=\frac{P\left(C_{\delta_p}\middle|\mathbf{x}\right)-P\left(C_{\delta_p}\cap X_e\middle|\mathbf{x}\right)}{P\left(U\middle|\mathbf{x}\right)-P\left(X_e\middle|\mathbf{x}\right)}$$

This is the probability update equation for excluded classes only.

**Case 14 (D/E)** both the excluded and included classes are *exclusive*.

The exclusiveness of the excluded classes gives

$X_i \cap X_e = \phi$, giving

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = \frac{P\left(C_{\delta_p} \cap X_i \middle| \mathbf{x}\right)}{P\left(X_i \middle| \mathbf{x}\right)}.$$

Now, by the axiom of exhaustivity and the exclusivity of the independent classes, $C_{\delta_p} = C_{\delta_i}$ for some $i \in \Delta I$ or $C_{\delta_p} = C_{\delta_k}$ for some $k \in \{1,\ldots,n\}$, $k \notin \Delta I$.

$C_{\delta_p} = C_{\delta_k}, k \in \{1,\ldots,n\}, \quad k \notin \Delta I \Rightarrow C_{\delta_p} \cap X_i = \phi$ giving

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = 0$$

which is the trivial case.

For the former case,

$C_{\delta_p} = C_{\delta_i} \quad i \in \Delta I \Rightarrow C_{\delta p} \cap X_i = X_i$ giving

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = 1$$

**Case 15 (D/I)** the excluded classes are *dependent* but included classes are *independent*.

Where the included classes are independent,

$$P_\varepsilon\left(C_{\delta_p}\middle|\mathbf{x}\right) = \frac{P\left(C_{\delta_p}\middle|\mathbf{x}\right)P\left(X_i\middle|\mathbf{x}\right) - P\left(C_{\delta_p} \cap X_e\middle|\mathbf{x}\right)P\left(X_i\middle|\mathbf{x}\right)}{P\left(U\middle|\mathbf{x}\right)P\left(X_i\middle|\mathbf{x}\right) - P\left(X_e\middle|\mathbf{x}\right)P\left(X_i\middle|\mathbf{x}\right)}$$

$$= \frac{P(C_{\delta_p}|\mathbf{x}) - P(C_{\delta_p} \cap X_e|\mathbf{x})}{P(U|\mathbf{x}) - P(X_e|\mathbf{x})}$$

Here, the included classes have no effect on the result as expected.

**Case 16 (D/D)** both the excluded and included classes are *dependent*.

Where both classes are dependent, Equation (A3) is used in its derived form where

$X_e = \bigcup_e C_{\delta_e}$, $X_i = \bigcap_i C_{\delta_i}$

From these 16 cases, a number of general rules can be noted as mentioned in the paper.