ARE THERE GOOD AIR TRAFFIC MANAGEMENT SAFETY INDICATORS

FOR VERY SAFE SYSTEMS?

Peter Brooker

Cranfield University

ABSTRACT

European Air Traffic Management is extremely safe. The drawback to this safety record is that it is very difficult to estimate what the 'underlying' accident rate for mid-air collisions is now, or to detect any changes over time. The aim is to see if it possible to construct simple ATM safety indicators that correlate with this underlying accident rate. A perfect indicator would be simple to comprehend and capable of being calculated by a checklist process. This problem has been examined by a combination of analogies with simple 'defensive' systems with Markov process properties. An important concept is that of 'system control': the ability to determine the outcome against reasonably foreseen changes and variations of system parameters The statistical distribution of future incidents has been analysed by focusing on an index – the CPI – of separation at the Closest Point of Approach. A promising indicator is 'Incident Not Resolved by ATC', INRA, incidents in which the ground ATC defences have been 'used up'. ATM Incidents can also be categorised in other ways: two examples are reviewed: the risk-bearing category for Airproxes and 'risk of collision/severity' scores. The second is more promising conceptually, but the existing scoring system has not been demonstrated to have the properties necessary to derive risk estimates. The key question is: if someone says they know how to make a good estimate of the underlying accident rate, then how could this claim be tested? If it correlates very well with INRA, then what would be the argument for saying that it is a better indicator?

1.   INTRODUCTION

Suppose you have just been appointed the Chief Executive of an air traffic control (ATC) provider. You must be a person of integrity and vision, and possess strategic problem-solving abilities. What is the question that should always be in your mind? The question is probably something like:

   "Is the system safe?"

But getting a good answer to this simple-formed question is not easy. The Chief Executive will be told that the problem is 'complex and difficult'; the data presented to him may be voluminous and indirect; the work of safety analysts may involve more Greek mathematical symbols than General Relativity papers. But the Chief Executive wants to see definitive conclusions. Is there an answer, or does the Chief Executive have to say: "We're doing the best anyone can: we're following 'best practice.'"

The real problem – a very good problem to have – is that modern ATC systems in developed countries have very few accidents. European statistics show that there were three air traffic management (ATM) related accidents involving commercial aircraft in the last twenty years, of which one was a mid-air collision. This means that attention has to focus on 'incidents' – safety occurrences somehow 'near' to accidents. Unfortunately, there are very many different kinds of incidents, and it is not always obvious which incidents are 'near' to accidents, for example see Spouge and Perrin (2005).

These problems are part of wider problems, eg (PRC, 2006 – quoting earlier material):

> "1. The quality, quantity and consistency of ATM incident reporting and safety information are wholly inadequate for the purposes of safety management and performance review across Europe;
>
> 2. There are no reliable key performance indicators for ATM safety;
>
> 3. There is no Europe-wide analysis of ATM-related incidents."

The aim here is to see if it possible to construct good ATM safety indicators. The problem is analysed in several different ways.

2.   WHAT DOES 'IS THE SYSTEM SAFE?' MEAN?

The Chief Executive's question 'Is the system safe?' could mean several different things, so it needs to be examined carefully. The words 'the system' 'safe' and 'is' all require some kind of definition.

What is 'the system'? There is considerable scope for debate about what precisely ATC and ATM should cover: Brooker (2006) presents one view and gives references to 'official' definitions. ATC is usually used to mean ground-based ATC and is supporting systems, and ATM is a wider term, which may include safety regulation and airborne collision avoidance. For present purposes, the system is restricted to en route ATC, ie not including accidents operating at airports, and the focus is on mid-air collisions. This excludes, for example, accidents occurring if controllers were to provide incorrect information about severe weather conditions.

There are instances where mid-air collision risks are entirely due to pilot actions. These can occur wholly independently of ATC, and could be near impossible for ATC to remedy. Mid-air collision and other airborne accidents could also arise from terrorism or hijacking, ie where there is some intentionality to cause an accident: it is vitally important to defend against such occurrences, but they are not failures of the ATM system *per se.*

What is 'safe'? There is an extensive literature about the word 'safe' and related words such as 'risk'. For present purposes, safe will be taken to relate to the rate of mid-air collision accidents occurring over a given period. Usually, this period covers tens of millions of aircraft flight hours while under the supervision of controllers. By supervision here is meant control in the normal sense, eg information and advisory

services would not be included, and the system should have 'current best practice' radar and ground/air conflict detection and alerting systems in place.

'Safe' is usually a statement that this rate is less than some declared safety target (eg see Brooker, 2004). For present purposes – this does not affect the main arguments – 'safe' will be measured by mid-air collisions during a particular year for the country or region in question. This is the 'underlying accident rate'.

The 'is' definition actually needs questions to be answered about the motivation of the Chief Executive? There are at least three different kinds of motive. The first is moral or ethical: nobody wants to have people's deaths on their conscience. The second is professional: the Chief Executive has objectives set by the ATC provider management board and government: to keep the system 'safe' means doing the job well. The third is consequentialism: a mid-air collision would be seen as the ultimate failure by an ATC provider's management, and would damage them and the organisation. [J. Bruce Ismay, the managing director of the firm which owned the Titanic, was unfortunate enough to travel on its maiden voyage, but far more unfortunate through surviving its sinking. His last years were spent largely in seclusion.]

Given these motives, it is reasonable to suppose that the Chief Executive will ask a second question: "What is the chance of a mid-air collision this year?" Leaders of ATC providers do ask this question. In the days before collision avoidance systems, an answer was once given: "Perhaps one in twenty-five years." These are not very good odds: a cynical view – but hardly wrong if it spurs people to safety-improving action.

So the Chief Executive's second question largely answers the question: "What is 'is'?" The concern is not with the 'is' of the present, or rather recent history, but with what information such incidents convey about the risks inherent in the system, so as to enable predictions to be made about the future. The Chief Executive wants this incident information – supplemented by what other knowledge? – somehow to be processed in a way that provides estimates of future system risk. Put bluntly: "What are the odds of a mid-air collision next year?"

Moreover, the Chief Executive cannot simply trust what he or she is told. It is important to be able to understand in broad terms how collision risk estimates are made. It is not sufficient to believe what one is told without questioning the nature of the estimates.

3. MODELLING SAFETY INDICATORS: ANALOGIES

How does one know actual safety? Is it possible to construct simple safety indicators for ATM? Indicators need to be able to tell a Chief Executive:

      If safety targets are being met this year – absolute indicator

      If safety is improving or not, year by year – relative indicator

The first is much more desirable – but also much more difficult.

What is a safety indicator? It is something that tells us about system performance. It enables us to predict the future frequency of critical system failures. A perfect indicator would be simple to comprehend and capable of being calculated by a checklist process. It should not require complex modelling calculations to be carried out in order to 'weight' the data appropriately. It should be 'obvious', in the sense that people would agree that it was a sensible thing to measure – what the psychologists call 'face validity'.

The quality of the prediction does not have to attain perfection, but it needs to be reliable in a statistical sense. A simple way of measuring this kind of reliability is the standard correlation coefficient, which ranges from zero, for a useless indicator, to unity, for a perfect indicator. Thus, if the underlying level of safety were to go up and down a hundred times over a long period, a perfect indicator would go up and down a hundred times with exactly the same pattern of ups and downs. A useless indicator would match the underlying pattern half the time, ie show no better than a chance correlation. A useful indicator might match the underlying patterns changes 80% of the time and fail to match 20% of the time.

The major problem with making such predictions for ATM is that the observed rate of mid-air collisions is extremely small. They generally arise from complex combinations of factors, so it is incredibly difficult to be confident that one understands the mechanisms that relate incidents to accidents. So how confident can one be that an indicator is giving the 'right' answers? If one wants to predict with confidence the likelihood of something happening, one needs strong evidence about the causation and measured frequency characteristics of that kind of event. This argues for few assumptions to be made and for calculations that do not require arbitrary or judgemental assumptions.

Suppose that a collision could only be caused by one particular set of circumstances. Suppose also that near-collisions were a well-defined category that described events that could only happen in a finite limited number (N – 1) of occasions per collision. This would correspond to the picture in Figure 1. In this analogy, a collision is the dark hexagon, a near-collision is a lighter shaded hexagon, and all the situations with no risk are empty hexagons. In reality, there are a very large number of empty hexagons, extending in all directions, because collisions and near collisions are rare.

Suppose that the probability of an event is represented simply by the number of hexagons, and that all hexagons, empty, shaded or dark, are equally likely to happen. Could this occur in reality? Yes, for example when impaired performance was caused by a single (detectable) failure in a piece of doubly redundant equipment, and total failure by an independently occurring second failure of this equipment. This might occur from statistically predictable failure rates, eg show the regularity of mechanical failures that arise from metal fatigue.

Leaving statistical variations aside for the moment, and with 'near collisions' including any actual collisions, the picture says:

$$\text{Rate of collisions} = \frac{\text{Rate of collisions}}{\text{Rate of near collisions}} \times \text{Rate of near collisions}$$

Or:

$$\text{Rate of collisions} = \frac{1}{N} \times \text{Rate of near collisions}$$

This means that, if collisions are not observed during some time period, their rate can be estimated by scaling down the rate of observed near collisions by the factor N. The three key elements for this to work satisfactorily are: that near collisions can be identified; that they are of equal value in predicting the collision rate, ie there is a single factor for doing this scaling; and that the size of this factor is in principle knowable. Unfortunately, this nice mental model is a wholly inadequate description of ATM incidents and accidents.

Are there simple analogies for the operation of the much more complex kinds of defensive mechanisms and probabilistic effects that really occur in ATC systems? Table 1 shows a possible analogy. Here there are significant 'bad' Events, which might be predicted by Indicator 1 or Indicator 2. The Events and Indicator numbers are observed over 19 separate occasions. Indicator 1 and indicator 2 occurrences on each of these days are more frequent: the first by a factor of about eight and the second by a factor of about four. The ratio between the number of Events and Indicators 1 and 2 is shown, as a percentage, in the 'rate' columns:

There is a great deal of statistical variation between these variables. A simple way of quantifying this is by calculating the correlation coefficient between the number of Events and the Indicator 1 and 2 numbers – shown at the bottom of the Table. Both Indicators are correlated with the Events, but Indicator 2 is more highly correlated. Indicator 2 is not perfect, but it would not be a bad predictor of the Event numbers: the standard deviation of the difference between actual and predicted numbers is about one.

The data in Table 1 in fact derives from statistics on association football games. It is important to say immediately that winning or losing a soccer game has no comparison with the life-and-death issues of ATC safety. The relevance here is that this real-life data provides a statistical example of a system defended by people, one that has probabilistic elements. Hence, in some ways, it is a real-life simulation of key features of ATC safety. It is thus potentially a useful analogy, without the horrific connotations of aviation accidents.

The data in Table 1 are taken from University of Kentucky (2005), purely because USA sports statistics, even for non-professional teams, are widely available. They show nineteen games played. Events are goals scored by the University team; Indicator 1 is 'shots at goal'; ie attempting to score; Indicator 2 is 'shots on target', ie which either scored a goal or were saved by the opposition goalkeeper/defence. The message from Table 1 is that both these indicators predict the number of goals scored, but that 'shots on target' is a better predictor.

Thus, Table 1 shows statistical data on a people-defended system. To win a football match, it is necessary to concede fewer goals than your team scores. Goals occur when the goalkeeper/defence fails to deal effectively with a shot on goal, which has come about because the opposition have possession of the ball and are position to

make such shots. This has been modelled statically by several authors. Particularly relevant here is the Markov Process model constructed by Hirotsu and Wright (2003). It presents data on possession by the competing professional teams; 'possession' meaning that the team has control of the ball. The correlation coefficient between Goals and Possession is about 0.25, smaller than the two coefficients in Table 1.

The Hirotsu and Wright Markov Process model focuses on the changes in possession during a game. The opposition defends by trying to obtain possession, ie in effect to reduce the likelihood that the attacking team can get into position for a shot at goal; it also blocks balls if the attacking gets near enough to attempt a shot on goal, and the goalkeeper saves shots that would otherwise produce a goal. Thus the sequence is:

> Possession
>> Attack
>>> Shot at goal?
>>>> Shot on target?
>>>>> Not saved by goalkeeper?

The key point from the present task of trying to find good ATM indicators is that, the closer the indicator is to the actual event, the higher will generally be the correlation coefficient. Thus, for these examples, the highest correlation is with Indicator 2, representing shots on target, and the lowest is with minutes of possession of the ball.

One reason that Indicator 2 predicts the number of goals quite well is that there are few intervening factors between a shot on target and a goal. The observed rate over all the matches in Table 1 is 25%, ie the goalkeeper/defence prevents about three quarters of these shots. This is largely because the probability of scoring is mainly a geometric probability. Can the goalkeeper etc manoeuvre to reach the ball in the time interval from the shot the going in the goal? Thus, the chance of a goal will roughly be the ratio of the area that a goalkeeper etc can reach divided by the area of the goalmouth. This is why defences try to ensure that shots are made at a sharp angle to the goalmouth, because this reduces the effective area presented to the attacking team. Conversely, penalty shots usually have a much greater chance of producing a goal than shots from normal play because the whole area of the goalmouth is presented.

Another important statistical feature of Table 1 is how difficult it is to be confident that there are no changes over time in the frequency of goal scoring or the number of incidents, either type 1 or 2, per match. For the Table 1 data, these variables have small negative correlations with the day number. These correlations are not very significantly statistically – they could easily have arisen by chance. This shows the difficulty of identifying 'real' changes in even simply-modelled Markov processes (evidenced by Hirotsu and Wright's (2003) models and earlier published work, eg Dixon and Robinson (1998)), with Poisson-distributed variables having small statistical expectations).

It must also be stressed that the measured frequency of goals and indicator numbers are not necessarily fixed for all times, even for a particular kind of league. Higher or lower rates could be produced by a 'soft' system change: for example, if more points were awarded for wins, then players might change the balance of attack and defence, or by modifications to the rules of the game (eg see Everything 2 (2004)).

4.   COMPARISON OF SYSTEM DEFENCES

How useful is the football defence analogy in understanding the much more complex ATC safety system? What are the important differences and similarities in system terms?

ATC does not fail because of an actual opponent. All the controllers and pilots are on the same team, trying to steer balls (more than one of them, but not that many – eg see page 69 of PRC (2005)) towards the goal of a safe transit through the airspace and away from the other end of the pitch, where a goal would represent a mid-air collision. Incidents and accidents occur because people are fallible, so incidents and accidents are in essence 'own goals'. Everyone makes mistakes, fails to remember some things, sometimes mis-estimates what is likely to happen, etc. Controllers and pilots are carefully selected and trained for their jobs, but still make these kinds of errors – at a very low frequency.

Possession in football terms translates roughly into a shortfall of system control in ATC. People's errors, either solely or in combination, are the threats. The phrase 'system controlled' here covers all the means by which the system is defended against potential negative and serious consequences. As system control is rarely lost in practice, the proportion of time that the ATC system is under threat from even these 'root cause' precursor events (see Corcoran (2004) for a discussion) is very low.

ATC has highly structured processes and safety defences. Figure 2 is a simplified version of the control processes ensuring safety, in reality there is a very complex set of probabilistic feedbacks and interactions. Explanations of separation minimum, STCA (Short Term Conflict Alert) and ACAS (Airborne Collision Avoidance System) can found in Brooker (2005a) and its references; the symbols ① and ② are covered in a later section.] The existence of STCA – plus help from colleagues – means that the controller is warned about potential separation breaches, even if he or she does not notice them. Note that a separation breach can occur because the pilot deviates from the safe plan; or when the safe plan was not in fact safe, in terms of the required minimum separation between aircraft. The existence of ACAS means that the pilot is warned about possible collisions, and told what ascent or descent flightpath should remove the risk.

Hence, the ATC system has multiple error detection and correction mechanisms. Note that these defences allow for error corrections to be carried out before any separation is breached, ie the error would not be detected just from records of separation breaches alone. Remedial action in the ATC system is therefore diverse and in depth. Brooker (2005b) illustrates the complex combinations that have been observed in UK airspace, from an examination of Airproxes (UKAB, 2006).

The timescale for decisions and actions in football is very short. A fast shot in professional football can travel at around 90 feet per second (~60mph), while the length of a typical football pitch is ~360 feet. Thus, decisions have to be very quick – within a second or much less – and essentially require skilled and practiced physical reactions. In contrast, controllers generally have much longer periods to work with; hence, they can absorb quantitative information and make rational choices. They are selected and trained to ensure that they can process pictorial and database information efficiently and quickly. The need for a 'time buffer' so that these processing can be carried out is one the reasons that separation minima set to comparatively large values – 3Nm in terminal areas/5Nm en route, and 1000 feet vertical separation [NB: the higher en route figure corresponding to a higher aircraft speed compared with terminal areas]. Moreover, STCA and ACAS make huge differences to safety (Brooker, 2005b).

5.   ATM SAFETY INDICATORS

The lessons from the previous sections are that ATM is a very strongly defended safety system, but one that shares some structural and statistical characteristics with simpler systems. If these analogies hold, then mid-air collisions will be most highly correlated with indicators that are closest in nature to collisions, and much less correlated with indicators that are nearer to root causes or initiating precursors. How could this be proved? This conclusion cannot be rigorously proved in statistical terms, because there thankfully is not the data on accidents to analyse – although a Markov process simulation of ATM could be used to demonstrate the reasonableness of the assertion. This is part of a wider problem: that the rarity of accidents implies there is generally not enough data to correlate with causal processes leading to accidents and incidents, and hence one cannot easily use statistical modelling for the whole process (Brooker, 2006).

To reiterate, the problem is that of estimating the frequency of (almost completely) non-observed catastrophic occurrences from the frequency of related observed occurrences. Good predictions necessarily rely on regularity and consistent patterns, so that past observations can be extrapolated in appropriate ways. What is the best – what is 'best'? – categorisation of such related occurrences, and how are they to be scaled/weighted to estimate the frequency of catastrophic occurrences?

Return to the perspective of the Chief Executive introduced earlier. The kind of question he or she is asking is: "If things are the same next year, will there be an accident?" But they will not be the same next year. If one could be sure that things would be exactly the same, then there would obviously be no accidents next year. But empirically there are changes in the number and types of incidents observed from year to year, even if total traffic is constant. There will be a variation in the number and characteristics of initiating factors, and in the relative orientation and timings of the aircraft flightpaths concerned. If there were to be a near-repeat of an incident's characteristics, there would still be variations in how the controller and pilot involved, added by colleagues and the alerting systems, will handle such an occurrence.

A simple characterisation of incidents is into Circumstances and Performance:

'Circumstances' describes the differences in the initiating event and flightpaths that lead to a breach of separation. Circumstances covers variations in the physical parameters, such as take-off time and entry to an airspace sector.

'Performance' is what the system defences of Figure 2 do about those Circumstances. Performance covers variation of humans and equipment in generating, detecting, and resolving the problem.

Thus, the incident data provides a particular set of Circumstances followed by a particular set of Performance actions dependent on those particular Circumstances. Incident data – notably the Closest Point of Approach (CPA) between the aircraft – generally corresponds to some kind of intervention.

As regards predicting next year's incidents and the probability of an actual accident, the modelling difficulty is that both the Circumstances and the Performance are essentially samples from an unknown statistical distribution. This distribution corresponds to the coming year's traffic volume and general pattern repeated many, many times, with all the variations of real life, producing different incidents and accidents, both in terms of Circumstances and Performance. What are the general characteristics of the frequency distribution – and is knowing something about them actually helpful?

The first step is to simplify incident CPA descriptions by using a method previously introduced in studying Airproxes, STCA and ACAS (Brooker, 2005b). Define H (Nm) and V (feet) to be the miss distances at the CPA. From Airprox data, H and V appear to be statistically independent variables, ie high values of one are not associated with either high or low values of the other. This offers the opportunity of combining the H and V values – a single indicator of close proximity is much easier to deal with analytically than a two-dimensional array. What should be the relative weightings in such a combination? The two parameters need to be re-scaled, possibly relative to the corresponding velocity components (a climb/descent manoeuvre in an ACAS Resolution Advisory (RA) takes the aircraft to a vertical speed of 1500 feet/minute; and aircraft might be travelling at 240 knots in a terminal area and 480 knots in en route airspace).

The simplest thing is to assume that the weighting should be based on the proportional deviation from the separation minimum, and to use terminal airspace criteria (because the great majority of Airproxes occur in TMAs). Thus, as the horizontal minimum is 3 Nm and the vertical minimum is 1000 feet, the 1 Nm horizontal CPA can be taken to be equivalent to 333 feet CPA. The simplest combination of the weighted H and V is just to add them together, ie:

$$CPI = 333 \times H + V$$

Here CPI stands for Close Proximity Indicator. There are obviously variants on this: for example, the H value of en route incidents could be scaled down by the ratio 3 to 5, to reflect the difference in minima.

It cannot be emphasized enough that the CPA parameters and the CPI value are measured <u>after</u> intervention. The CPI does not provide information on how close the aircraft would have been had there been no intervention or a 'standard intervention'.

Figure 3 is a schematic picture – note the non-linear scale – of the unknown frequency distribution describing the proportion of incidents with a particular CPI value – the x-axis. The actual number of incidents in a particular 'constant traffic' year will be sampled from some other kind of distribution, probably more Gaussian in shape. The distribution in Figure 3 is a very long-term average, ie would be observed over a very long period of operation of the current year's traffic, repeating the year almost endlessly. Why should the distribution have this kind of shape?

Why is the distribution drawn as monotonic increasing? The empirical reason is that data such as Airproxes do show this form: smaller CPI values are less likely than larger ones. In abstract terms, potentially small CPI values can be detected by the controller, STCA, the pilot, and ACAS; so there will be strong safety defences in place to prevent them occurring. Remember also that the CPI can be viewed in terms of the deviation from the separation minima.

Why is the region from origin to x_c flat and marked as 'Random Distances'? This is because these small CPIs will represent incidents in which 'system control' has been largely lost. The relative distances between the aircraft when they pass will be a function of the chance orientations of their flightpaths, even if modified by ATC. For such an occurrence, is there any reason to believe that a CPI of 100 feet is more or less likely than one of 200 feet or 300 feet?

Formally, the phrase 'system control' means something like:

> <u>system control</u> – the ability to determine the outcome against reasonably foreseen changes and variations of system parameters, such as the abilities of the participant(s), the environment (in the largest sense), and the safety mechanisms in place.

So where exactly does the flat region start to turn upwards? This is a very difficult question to answer, because it is essentially asking about the degree of control that a controller or pilot has about the flightpaths of aircraft.

What is the 'Collision occurs' region? This represents the CPI values that are less than the aircraft dimensions. Thus, these CPI values are collisions not incidents. The key fact about the frequency distribution is that the flat shape near the origin, which includes collisions, cannot simply be extrapolated from the increasing section of the curve – note again the logarithmic scale. Simple extrapolations from the observed part of the CPI distribution are likely to be very cautious – eg see Brooker (2005b). Knowledge of the distribution of CPIs for regions where there is a substantial degree of system control cannot then be extrapolated with great confidence to estimate the shape – more important the height – of the curve f(x) in the 'Random Distance' region. To be useful in estimating f(x) for x values near zero, ie to provide evidence about the flat section of the curve, it would be necessary to eliminate incidents from consideration that had any degree of system control. These kinds of arguments return to the discussion in Section 3, ie that the best estimates of

the rate of critical events are likely to be made from indicators that are very near to those events and which differ from them largely through geometrical factors.

Does this Figure show the full structure of incidents? No, it does not: the full picture cannot be mapped onto a simple two-dimensional diagram. The degree of system control tends to be higher for the higher CPI values, but there is not a one-to-one relationship. Incidents with the same CPI value can represent widely differing degrees of system control: thus, a large CPI value might merely indicate that the aircraft were on widely separated flightpaths rather than being the consequence of a swift control action to keep them apart.

6.   RISK ESTIMATES FROM ATM INDICATORS

From the argument in the previous Section, it appears that the simplest indicator of collision risk would need to use data from incidents in the 'Random Distances' region. Leaving aside for the moment how one might identify the extent of this region, how could an estimate of collision be made from such data?

Figure 4 shows a real-life frequency distribution histogram of Airprox incidents recorded in UK airspace. [Note that the vertical scale here is linear not logarithmic.] This data is taken from Brooker (2005b), and covers five years from 1999 to 2003. In this data set there were 29 incidents with CPI < 1000, of which one could well have a CPI small enough to fall into the 'Random Distances' region. That incident was Airprox 2001/052, in which a pilot took a wrong instruction to descend, which was undetected by ATC because of simultaneous transmissions; its aircrew then ignored RA and used 'visual' avoidance action. The aircraft in this incident were surely at best on the borderline of system control.

What would be a possible cut-off value for the 'Random Distances' region of x_C? It might be derived from an analysis of many incidents, or it might be chosen by safety managers/regulators as a standard value. For illustration, assume it is 500 feet. The CPI value for the size of an aircraft might be of the order of 50 feet. As f(x) is a flat distribution for values up to x_C, this implies that the CPIs for incidents within the range zero to x_C are equally probable. Thus, the probability of a CPI value falling within the Collision Region would be approximately 50, the aircraft dimension, divided by 500, the size of the 'Random Distances' region, ie 1/10. If it were believed that the risk of a collision is constant, and there have been (say) ten years without any other incidents in the 'Random Distances' region, then the probability of a collision in the next year would be estimated about 1/100.

Whilst well founded in statistical and modelling terms, this Random Distances Estimate (RDE) is a disappointing calculation. The collision probability cannot be calculated without several years' data, and this data will be subject to the typical large fluctuations to be expected from Poisson-distributed events. What the Chief Executive really wants is something like a thermometer: absolute values are measured and the temperature goes up or down depending on the environment of interest. The Chief Executive wants to know quantitatively if the system is 'safe' now (ie for the next year), and if it is getting safer or less safe.

In an ideal world for the Chief Executive, the form of f(x) would be known on a yearly basis. Consider Figure 5 (based on Figure 3, but with a magnified vertical scale). The Figure shows two possible curves A and B, one of which might be the correct form for f(x) in a particular year. Estimating collision risk essentially means estimating the intercept on the vertical axis. Could this be done by measuring the size of the curve at some x-axis distance markedly above the value x_C, eg at x_I – where system control is weak but not wholly absent, thus benefiting from the accumulation of considerably more incident data? While it might not be possible to prove that the absolute estimate of the intercept was perfectly accurate, it might potentially be a very helpful way of monitoring relative collision risk. In other words, would a count of the annual frequency of small and medium-sized CPI values lead to a good indicator of collision risk in each year?

To explore this, define an incident Type as a specific set of necessary conditions sufficient to produce a particular incident (see Brooker (2006) for discussion and references on necessary and sufficient conditions). Each one of this set of causal factors, using the phrase loosely, can have a set of parameters associated with it. Thus, there could be a set of similar incidents that involved aircraft on converging routes, some of which might have two commercial aircraft and others with two executive jets, so that the speeds and climb performances would be different.

Figure 5 shows two incident Types that contribute to curve A, marked I and II. The curve will be made up of a large number of different incident Types. Every observed incident is a sample from all the 'similar' items, ie from that incident Type's probability distribution. Type I incidents are those in which, for some rare combinations of parameters, correspond to loss of system control; which implies that very small CPI values – less than x_C – can occur, and hence there is a risk of collision. Hence, Type Is have collision potential. In contrast, incident Type II can never lead to a collision: there is always some system control and CPI values never get as low as x_C. Hence, Type IIs have no collision potential. An examination of Airproxes shows that the bulk of them probably fall into this second category: separation has been lost for some reason, but is then restored through normal processes; that being one of the characteristics of that incident Type. Thus, no reasonable combination of parameters could produce a collision for this specific variety of incident.

Of course, an additional contributory factor could be seen as turning a Type II incident into one with collision risk potential; but this extra factor would then mean that the incident should be categorised in another (Type I) family. Data on Type II incidents does have safety (ie collision risk) value if the probability of such additional contributory factors can be estimated in some way. But this is leading to complex risk modelling rather than the creation of simple ATM safety indicators.

On what basis would the curve A and B values, measured at x_I, ie the points A and B, be a good estimate of the relative y-axis intercepts of the two curves? Do the curves have the same shape? A thought experiment shows this is a hard question. Do curves such as A and B retain their shape for increased traffic volumes, given exactly the same ATC system, controllers, pilots etc? Suppose traffic numbers increase by a factor k, but that all strategic traffic patterns over time and space remain the same. A given aircraft will encounter k times as many aircraft and hence

potentially k times as many incidents. Thus, the total number of incidents will be expected to increase by a factor k^2. So will the collision risk and the number count for CPI = x_I both increase by a factor of k^2? They might do – but they might not. A simple confounding element could be that, with an increased traffic volume, more ATC sectors would be 'above capacity' for short periods. This could increase the likelihood that controllers might be distracted and hence nor remember that particular aircraft flightpaths had to be revised on their progress through the sector (eg see Loft et al (2004)). This could increase the number of Type I incidents in comparison to Type II incidents, ie increase the likelihood of incidents with collision potential. This kind of change would mean that curves such as A and B would have different shapes for increased traffic levels, rather than being simply scaled-up. The values A and B would not therefore be in proportion to the y-axis intercepts.

The fact that something might happen does not mean that it occurs in practice. It might be that the value at x_I is generally a good estimate of relative collision risk. Two factors that might make it so are:

- That Type II incidents are systematically excluded in this calculation – because they are in themselves irrelevant to collision risk and cloud the Type I incident picture.

- That x_I is not too far from x_C – because the further the extrapolation from x_C the less reliable will be any quantitative predictions.

7. INCIDENT REPORTING AND SAFETY DEFENCE INDICATORS

If absolute ATM safety indicators for collision risk are difficult to construct, are there good indicators of the performance of ATC safety defences? If so, do they have merit as ATM safety indicators? Incident reporting is a very large subject, so the following just focuses on some key points.

There are many possible ways of classifying incident reports. For present purposes, a division into three schemes is useful: Individual reporting, event-related, and post-processed.

Individual reporting is the traditional kind of reporting. An operational person detects something that is unsatisfactory in safety terms and reports this to a central monitoring body. The body collects data about the incident and produces an explanation for what happened. The 'gold standard' for this in the UK is the UKAB (2006), whose work stretches back several decades. Originally, only pilots could report such incidents, but this was expanded to include controllers. The likelihood of someone reporting an incident very much reflects the ATC provider's organisational safety culture (eg Madsen, 2002).

Event-related reporting is triggered by system warnings or alerts. The main examples are STCA and ACAS, with other systems being used in different phases of flight, eg GPWS. Data on these incidents is collected and processed by States. Eurocontrol (eg Dean and Baldwin, 2005) has started to analyse ACAS events on a European basis. Again, there can be problems with reporting rates – some States

are either very, very safe or – more likely – just report a fraction of the total incidents (eg see Table 5 of Dean and Baldwin).

In *Post-processed reporting*, radar and related data is examined some time after actual operations, to determine if (eg) separation minima have been significantly breached. An early version of this was the USA FAA's 'snitch patch', which had a marked impact on reporting by individuals (Tamuz, 1987). The UK introduced its own version of this system – the Separation Monitoring Function (SMF) – about fifteen years ago. Eurocontrol has subsequently carried out work on this tool (eg Joyce and Fassert, 2002).

The combination of all the data from these three varieties of incident reporting system, enhanced by the pilot/controller recollections and other data (eg communication recordings), mean that a good picture can usually be obtained of the nature of any incident in controlled airspace. The exceptions are where there are equipment failures – this is very rare for UK Airproxes.

What are reasonable indicators that could be formed from these valuable incident databases? It must be stressed that these are not necessarily indicators of ATM safety *per se* but rather of the performance of the ATM system's safety defences. Starting from basics: for a collision to occur:

> Must have had separation breach – the aircraft was not flying to a safe plan or the plan was not in fact safe.

> Must have had failed or non-existent intervention(s) to remedy, even with assistance from colleagues and warning systems.

> Must have had the 'right' (post any intervention) flightpaths – traffic density, route/airspace construction are factors.

There are two obvious places for safety defence indicators apparent in Figure 2, indicated by at the symbols ① and ②. The first counts initiating events that produce a separation breach, and the second covers situations where the ground-based part of the system, ie ATC, has not resolved an incident. The first indicator, at about ① in Figure 2, is a count of what be termed here an 'Actual Separation Breach' – ASB. The second indicator, at about ② in Figure 2, will be termed here an 'Incident Not Resolved by ATC' (INRA): this focuses on incidents in which the ground ATC defences have been 'used up'.

There are many other potential indicators that might tell a Chief Executive something about the way the system's safety defences are performing, but these two represent decisive points in the safety defences, are very simple for anyone to understand, and can be measured reasonably consistently. This is because they correspond to measurable events or system states, rather than complex judgemental assessments of what might have taken place. In system terms, the first shows how frequently system control is impaired and the second when the final defensive line of system control necessarily has to be employed. To find another simple indicator in the flow chart between ① and ② is extremely difficult because of the variability and complexity of what can happen when ATC's defensive mechanisms restore full system control (eg see Spouge and Perrin (2005), SRC (2005a)).

It is essential to have clear rules for deciding what are ASB and INRA incidents and what are not, so their definitions need to cover both typical and 'pathological' cases'. The following sketches out some rules.

> ***Actual Separation Breach – ASB*** This is a count of post-processed incidents that breached the appropriate separation minimum, ie in which full system control was at least temporarily absent.
>
> It excludes those incidents in which the breach was 'small', eg a 2.5 Nm horizontal CPA when the minimum is 3 Nm – this is consistent with the triggers for SMF in the UK.
>
> It also excludes cases in which ATC management declare acceptable safe an operating procedure that breaches a minimum slightly, eg to cope with tight airspace constraints in the terminal area. However, these exceptions must have been documented in the ATC unit safety case or other formal safety document. Mere 'custom and practice' arrangements would not be a valid reason for excluding such incidents.
>
> ***Incident Not Resolved by ATC – INRA*** This is a count of Airprox and/or Event-related reports in which the ground based part of the system, ie ATC, has not resolved an incident. Thus, ground ATC cannot reasonably be said to have full system control. Was an ACAS RA then necessary to resolve safely?
>
> The simplest incidents to count in this category are those in which an ACAS RA is deemed by ATC to be 'justified'. Dean and Baldwin (2005) quote a figure of 27% of corrective RAs followed by the pilot being rated as 'justified' by the controller handling the aircraft. But 'only 26% of controllers expressed an opinion in the report', which indicates some safety culture issues. If the incident were so very short-term that an RA was not generated, eg a rapid descent to a small CPI, then that obviously would have to be included under this heading.
>
> Brooker's (2005b) analysis of Airproxes involving STCA and ACAS alerts comments that, of incidents producing RAs, that controller(s) are already aware of a potential conflict in about 90% of occasions.

The second indicator does require an assessment to be made by operational controllers, that the pilot action following the RA is justified. But it is an assessment that is restricted to the kinds of things that controllers actually experience. He or she is not being asked to extrapolate beyond this experience. The current low reporting rate of this judgement is of serious concern, and so will in any case need to be tackled as a part of continuing efforts to improve safety culture. In the absence of some fraction of controller opinions, the view of a UKAB-like body would be needed to ensure complete coverage.

The first indicator, ASB, provides an indication if the rate of 'initiating events' is changing. This is very important if there is continuing traffic growth. If flights increase by a factor of k, then, all other things being equal, the number of times that aircraft pairs need to be separated increases by a factor of k^2, and the incident/accident rate (as a proportion of total flying hours) increases by a factor of k. A five percent annual rate of growth would generate a doubling in the accident rate in

less than fifteen years. An annual count of ASBs would quickly show – even allowing for statistical fluctuations in the numbers from year to year – if this kind of trend was developing. ASBs are used by the Eurocontrol SRC (2005b) (although it is not obvious from the text if some 'acceptable' varieties of separation infringement are filtered from the counting).

The second indicator, INRA, has an important relationship to the first, ASB. The ratio of the counts INRA to ASB is a measure of the effectiveness of the ground ATC system in resolving initiating events. An improvement in this ratio would therefore demonstrate an improvement in ground-based ATC.

There are also good grounds for believing that INRA would be a good indicator of the underlying accident rate. This is because these events are, by definition, deemed to be 'justified' ACAS involvement by ATC, and so should generally be Type I incidents, have low CPI values (ie that the effective x_I is not too far from x_C), and reflect reduced levels of system control. These match the criteria for an extrapolated indicator put forward at the end of Section 6.

8.   AIRPROXES? INCIDENT WEIGHTINGS?

Can Airprox numbers be used as an indicator of ATM system safety? Airprox statistics are certainly viewed as being useful in painting the ATM safety picture, eg NLR's van der Geest et al (2003) and PRC (2006). The answer is that Airprox information is invaluable, but the use of Airprox statistics does require a 'health warning'. To understand this comment, it is necessary to go back to the basic definitions of terms used in assessing Airproxes. The following definitions and explanations are derived from UKAB (2006):

> *Definition of an Airprox* An Airprox is a situation in which, in the opinion of a pilot or a controller, the distance between aircraft as well as their relative positions and speed have been such that the safety of the aircraft involved was or may have been compromised.

> *Risk Ratings* Risk level assessments are made on the basis of what actually took place and not on what may or may not have happened. There are four categories, agreed at international level, as follows:

>> A   Risk of Collision: An actual risk of collision existed.

>> B   Safety not assured: The safety of the aircraft was compromised.

>> C   No risk of collision: No risk of collision existed.

>> D   Risk not determined: Insufficient information was available to determine the risk involved, or inconclusive or conflicting evidence precluded such determination.

Various subsets of Airprox numbers have been used as safety indicators, eg Category A, Category A + B (ie 'risk-bearing' Airproxes).

The second criterion implies that incidents are not being examined as representatives of the kinds of Types of incidents discussed above, ie no variation in the Circumstances of the incident is considered.

The meaning of a phrase such as 'risk of collision', and similar phrases about 'hazardous' events, are examined in Brooker (2005a). Saying that there was a risk of collision is essentially a judgement about a system state that existed in the past. Hazardous means 'near accident' in the sense that, if Circumstances/Performance had been slightly different, there could/would have been an accident. For a Category A event, the professionally competent members of the UKAB are effectively replaying the event in their minds eye; and detecting at some point in the replay that one could not being confident that a collision would be prevented.

On this view, the UKAB's attention is solely on the Performance of the ATM safety defences for these specific Circumstances. Thus, did the incident require the controller and/or pilot to perform exceptionally well? Was the intervention at the limits of human or warning system performance? Was protection operating by chance affects, eg a colleague monitoring: if this had not been a factor, then would the rest of the other safety defences have been likely to be triggered?

There is nothing inherently wrong with categorising Airproxes in this way, but it does have restricted usefulness in estimating the most likely future mid-air collision. The problem is that a 'low severity' Airprox might have been judged as such because the Circumstances just happened to be favourable for the Performance element not to be a major concern. But if the Circumstances, eg the relative flightpaths/sector timings of the aircraft, had been somewhat different, then the Performance required to prevent an accident might have been much more demanding. The UKAB is composed of experienced professionals, and recognises this kind of issue in its analysis rather than its categorisation. For example, Airprox 2000/029 is rated in the B category, but the UKAB noted in its report:

> "Board agreed that any separation was to a large degree fortuitous…with different geometry …could have been considerably more serious."

The UK predecessor of the present Airproxes, called Airmisses, in fact used a category 'B/Potential A' for some years. The reason for the present categorisation scheme is presumably the difficulty of specifying clearly the kinds and size of Cirumstances parameters that the UKAB would reasonably have to consider.

Could there be a better way of categorising or 'weighting' Airprox incidents? The problem is how to ensure that such weightings match the relative contribution to the future accident rate. In particular, how could account be taken of the fact that a mid-air collision might be the consequence of a parameter variation in a known incident Type's Circumstances/ Performance, or it could be an additional factor being applied to a hitherto 'risk-free incident' Type?. It would be very important to use explicit criteria rather than judgement, in order to get consistent results. How would essentially arbitrary weightings by 'experts' be validated?

What would an ideal scoring system – call it the 'Score', with a capital S – look like? The scheme would attach a numerical score to each Airprox, and the total of the scores in (say) a year – the Total Score (TS) – would say something about system safety – ideally about the underlying accident risk. The earlier Random Distances Estimate (RDE) meets this objective, because it directly relates to collision risk. However, it does this by discarding the vast bulk of incident data and applying the

same score to the few incidents that might be left. Is there a 'better' way of scoring that assigns some weight to a higher proportion of incidents?

What properties does the Score need to have, in terms of the characteristics of measurement scales (Stevens, 1946)? The numerical value assigned to each Airprox must at least enable 'interval' comparisons to be made, eg the increase from a Score of 7 to 10 must contribute the same amount to 'ATM risk' as the increase from a Score of 3 to 6. Put another way, an incident with a score of 10 must be equal in ATM risk terms as 10 incidents with a Score of 1.

Are there candidates – which might need to be adapted or further developed – for such a scoring system? One possibility is the Eurocontrol SRC's scheme for 'Harmonisation of Safety Occurrence Severity and Risk Assessment' (EAM2/GUI5 document, SRC (2005a)). EAM2/GUI5 is an extremely interesting and complex document, well worth reading by anyone interested in ATM safety, but sometimes very difficult to understand, largely because it does not offer fully-worked examples. It attempts to do very sophisticated analyses of incidents, eg using carefully structured decision trees, but this is based on traditional, ie often unclear, definitions of terms.

This cognitive dissonance is most easily explained by an example. EAM2/GUI5 uses a traditional definition:

> Risk of collision (from ICAO Doc 4444): Airprox - Risk of Collision: "The risk classification of an aircraft proximity in which risk of collision has existed".

This is a circular definition, which adds no information (see Brooker (2005a) for a discussion on how best to interpret this inherently conditional concept). In contrast, the authors' thinking about incidents leads them explore much more fruitful concepts of 'controllability' (compare 'system control' here) and repeatability' (compare distribution of CPIs here).

The scoring methods proposed in EAM2/GUI5 are extremely sophisticated, and impossible to summarise in a few sentences – which would be a major concern if the aim to generate simple ATM safety indicators. An important point, however, is that it is not obvious that their final scorings have interval-like properties, ie that they weight things appropriately in 'potential collision risk' terms. An example is shown in Figure 6, reproduced from EAM2/GUI5. It shows the scoring component arising from infringements of the separation minimum. So does this scoring make sense in risk terms? Does having a separation of 0.5 Nm, ie in the bottom band (as 0.5 Nm = 10% of 5 Nm) merit a score of only 10, while a separation of 2.5 Nm (50% of 5 Nm) justifies one of 3?   Why does going from the middle band to the next one down increase the score by 4, but then going to the bottom one just increases by 3?

It is possible that a modified EAM2/GUI5 scoring system, removing these kinds of inconsistencies, could produce something near to the perfect Score. The questions would then be:

> How could it be shown that this does generate an interval-like Score that matches the underlying accident risk?

What correlation would it have with INRA values: if this is very high, then is the extra work required of benefit?

These are tough questions.

9. CONCLUSIONS

ATM in developed European countries is extremely safe. The drawback to this safety record is that it is very difficult to estimate what the 'underlying' accident rate for mid-air collisions is now, or to detect changes over time. The rate cannot be observed directly. The aim was to see if it possible to construct simple ATM safety indicators that correlate with this underlying accident rate. These indicators need to be able to tell (say) a Chief Executive if safety targets are being met and/or if safety is improving. A perfect indicator would be simple to comprehend and capable of being calculated by a checklist process. It should not require complex modelling calculations to be carried out to 'weight' the data appropriately. It should be 'obvious', in the sense that people would quickly agree that it was a sensible thing to measure.

This problem has been examined by a combination of analogies with simple 'defensive' systems with Markov process properties. An important concept is that of 'system control': the ability to determine the outcome against reasonably foreseen changes and variations of system parameters The statistical distribution of future incidents has been examined by focusing on an index – the CPI – of separation at the Closest Point of Approach. The indicator of collision risk that uses fewest modelling assumptions/extrapolations counts the number of incidents in the 'Random Distances' region of CPIs, where system control has been lost. However, the collision probability cannot be calculated without several years data, which is subject to the typical large fluctuations to be expected from a Poisson process. It does not tell us quantitatively if the system now is 'safe', or if it is getting safer or less safe.

Can ATM safety indicators be constructed from the performance of ATC safety defences? Two are examined. The first counts initiating events that produce a separation breach and the second covers situations where the ground based part of the system, ie ATC, has not resolved an incident. The first indicator is a count of what is termed here an 'Actual Separation Breach' – ASB. The second indicator counts 'Incident Not Resolved by ATC' – INRA: this focuses on incidents in which the ground ATC defences have been 'used up'.

ASBs are a useful measure for the frequency of initiating events for incidents, ie where full system control needs to be re-asserted. INRAs measure the number of times when the system is reliant on its final safety defensive layer, ACAS. There are good grounds for believing that INRA would be a good indicator of the underlying accident rate. This is because, by definition, these events are deemed 'justified' by controllers (and so should generally be incidents with collision potential), have low CPI values, and reduced levels of system control. However, there are issues about safety culture in reporting incidents that would need to be addressed, and/or a specialist UKAB-like body might be need to ensure complete coverage of such incidents.

ATM Incidents can be categorised in other ways. Two examples are the risk-bearing categories for Airproxes and the 'risk of collision/severity' scores proposed by the SRC. The first has gaps, because it (intentionally) does not consider variations in circumstances compared the observed incident. The second is more promising conceptually, but the existing scoring system has not been demonstrated to have the properties necessary to match the underlying risk level. The key methodological question is: if someone says they know how to make a good estimate of the underlying accident rate, then how could this claim be tested? In particular, if it correlates very well with INRA counts, then what would be the nature of any quantitative argument for saying that it is a 'better' indicator?

Indicators are very useful things to have – and it is very important to have systems that collect comprehensive data on incidents – but they are not solutions to safety problems. Accident rates will decease because organisations and the people in them understand the causes of the full range of potential accidents, and can think of ways of reducing the frequency of or eliminating these causes. Returning to the thermometer metaphor, very accurate temperature measurements tell us about the onset and progress of an illness, but they do not in themselves cure the patient.

Voltaire said "Le mieux est l'ennemi du bien" – "The best is the enemy of the good". The problem is a familiar one in general management – a good or very good plan is actually better than a perfect plan. To quote Beckwith (1997): "Rather than spending all your energy trying to get all the answers, running all the projections, getting all the data, and trying to achieve total consensus, you must reserve the majority of your energy for executing the plan". So: identify good, simple indicators, based on significant kinds of events or states of system control, and try to ensure that reporting rates are high.

ACKNOWLEDGEMENTS

REFERENCES

Beckwith, H. (1997). Selling the Invisible: A Field Guide to Modern Marketing. Warner Books.

Brooker, P. (2004). Consistent and Up-To-Date Aviation Safety Targets. Aeronautical Journal. July, 345-356.

Brooker, P. (2005a). Reducing Mid-Air Collision Risk in Controlled Airspace: Lessons from Hazardous Incidents. Safety Science. 43(9), 715-738.

Brooker, P. (2005b). STCA, TCAS, Airproxes and Collision Risk. Journal of the Institute of Navigation. 58(3), 389-404.

Brooker, P. (2006). Air Traffic Management Accident Risk Part 1: The Limits of Realistic Modelling. Safety Science *In press.*

Corcoran, W. R. (2004). Defining and Analyzing Precursors. Accident Precursor Analysis and Management: Reducing Technological Risk through Diligence. (Eds. Phimister, J. R., Bier, V. M., Kunreuther, H. C. The National Academies Press, Washington, D.C. http://www.nap.edu/books/0309092167/html/R1.html.

Dean, G. and Baldwin, T. (2005). European ACAS Operational Monitoring 2003 Report. EEC report 401.
http://www.eurocontrol.int/eec/gallery/content/public/documents/EEC_reports/2005/EEC_report_401.pdf

Dixon, M. and Robinson, M. (1998). A birth process model for association football matches. Journal of the Royal Statistical Society: Series D (The Statistician). 47(3), 523-538.

Everything 2 (2004). The Rules, and History of the Rules, of Association Football.
http://everything2.com/index.pl?node=The%20rules%20of%20soccer

Geest, P. J. van der et al (2003). Aviation safety management in Switzerland: Recovering from the myth of perfection. NLR-CR-2003-316. Netherlands National Aerospace Laboratory.
http://www.skyguide.ch/de/DownloadablesPDF/NLR_Report.pdf

Hirotsu, N. and Wright, M. (2003). An evaluation of characteristics of teams in association football by using a Markov process model. Journal of the Royal Statistical Society: Series D (The Statistician). 52(4), 591-602.

Joyce, A. and Fassert, C. (2002). Automatic Safety Monitoring' in Air Traffic Control - Achievements and Perspectives. Workshop on the Investigation and Reporting of Incidents and Accidents (Ed. Johnson, C. W.GIST Technical Report G2002-2.
http://www.dcs.gla.ac.uk/~johnson/iria2002/IRIA_2002.pdf

Loft, S., Humphreys, M. and Neal, A. (2004). Prospective memory in air traffic control. In, Edkins, G. and Pfister, P. (Eds.), Innovation and consolidation in aviation. Aldershot, UK Ashgate.

Madsen, M. D. (2002). A Study of Incident Reporting in Air Traffic Control - Moral Dilemmas and the Prospects of a Reporting Culture Based on Professional Ethics. [In IRIA 2002 – see Joyce and Fassert (2002).]

PRC [Eurocontrol Performance Review Commission] (2005). ATM Cost-Effectiveness (ACE) 2003 Benchmarking Report," ACE 2003 Final Report.
http://www.eurocontrol.int/prc/gallery/content/public/Docs/ace2003/ACE%202003%20-%20FINAL%20+%20Cover%20page.pdf

PRC (2006). An Assessment of Air Traffic Management in Europe during the Calendar Year 2005. Performance Review Report, PRR 2005.
http://www.eurocontrol.int/prc/gallery/content/public/Docs/prr2005.pdf

Spouge, J. and Perrin, E. (2005). 2004 Baseline Integrated Risk Picture for Air Traffic Management in Europe.
http://www.eurocontrol.int/eec/gallery/content/public/documents/EEC_notes/2005/EEC_note_2005_15.pdf

SRC [Eurocontrol Safety Regulation Commission] (2005a). EAM 2 / GUI 5 Harmonisation of Safety Occurrence Severity and Risk Assessment.
http://www.eurocontrol.int/src/gallery/content/public/documents/deliverables/esarr2_awareness_package/eam2gui5_e10_ri_web.pdf

SRC (2005b). SRC Annual Safety Report 2005.
http://www.eurocontrol.int/src/gallery/content/public/documents/report/srcdoc37_e10_ri.pdf

Stevens, S. S. (1946). On the theory of scales of measurement. Science, 103, 677-680. [See http://en.wikipedia.org/wiki/Interval_measurement]

Tamuz, M. (1987). The impact of computer surveillance on air safety reporting. Columbia Journal of World Business. 22(1), 69-77.

UKAB [UK Airprox Board] (2006). Home Page.
http://www.caa.co.uk/default.aspx?categoryid=423

University of Kentucky (2005). Mens Soccer Media Guide.
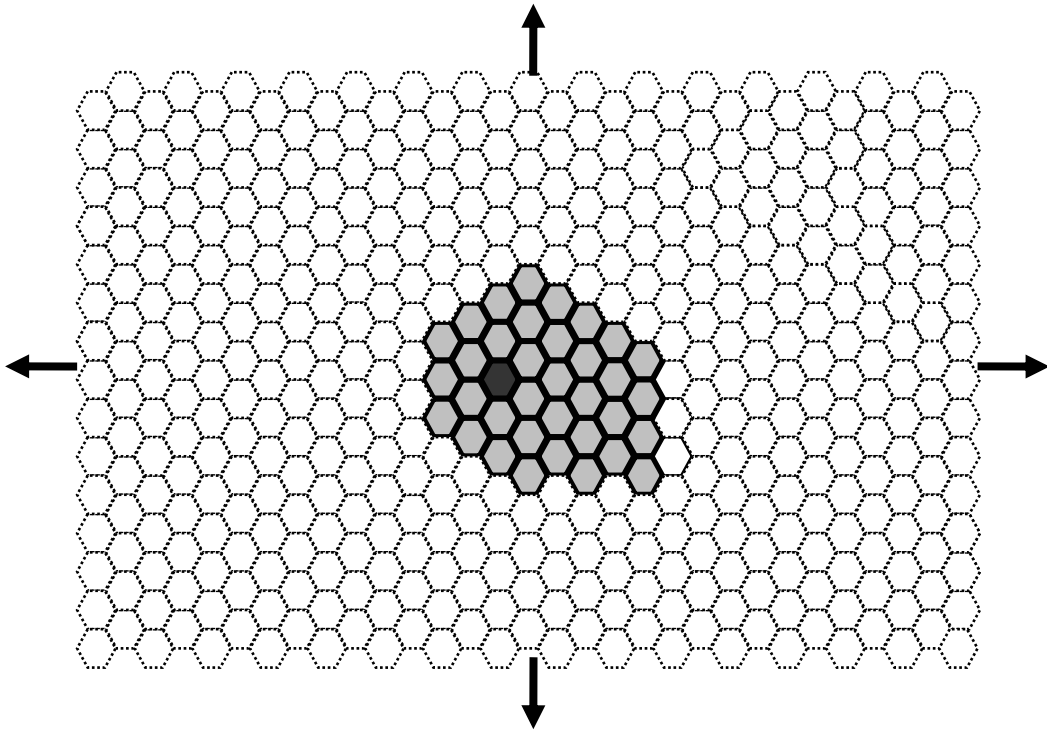http://www.ukathletics.com/doc_lib/socm_2005mg_complete.pdf

Figure 1. Idealised Accident/Incident Model – see text for explanation

Day	Events	Indicator 1	Indicator 1 – Rate %	Indicator 2	Indicator 2 – Rate %
1	1	9	11	8	13
2	1	9	11	8	13
3	2	7	29	5	40
4	3	15	20	9	33
5	3	15	20	10	30
6	1	19	5	2	50
7	2	14	14	5	40
8	2	21	10	6	33
9	2	21	10	6	33
10	0	8	0	1	0
11	3	12	25	10	30
12	0	8	0	6	0
13	0	12	0	2	0
14	5	19	26	9	56
15	1	23	4	12	8
16	1	14	7	5	20
17	2	9	22	9	22
18	2	15	13	9	22
19	0	9	0	1	0
Totals	**31**	**259**	**12**	**123**	**25**
Correlation coefficient with Events		**0.39**		**0.60**	

Table 1. Example of an Events indicator – see text for explanation
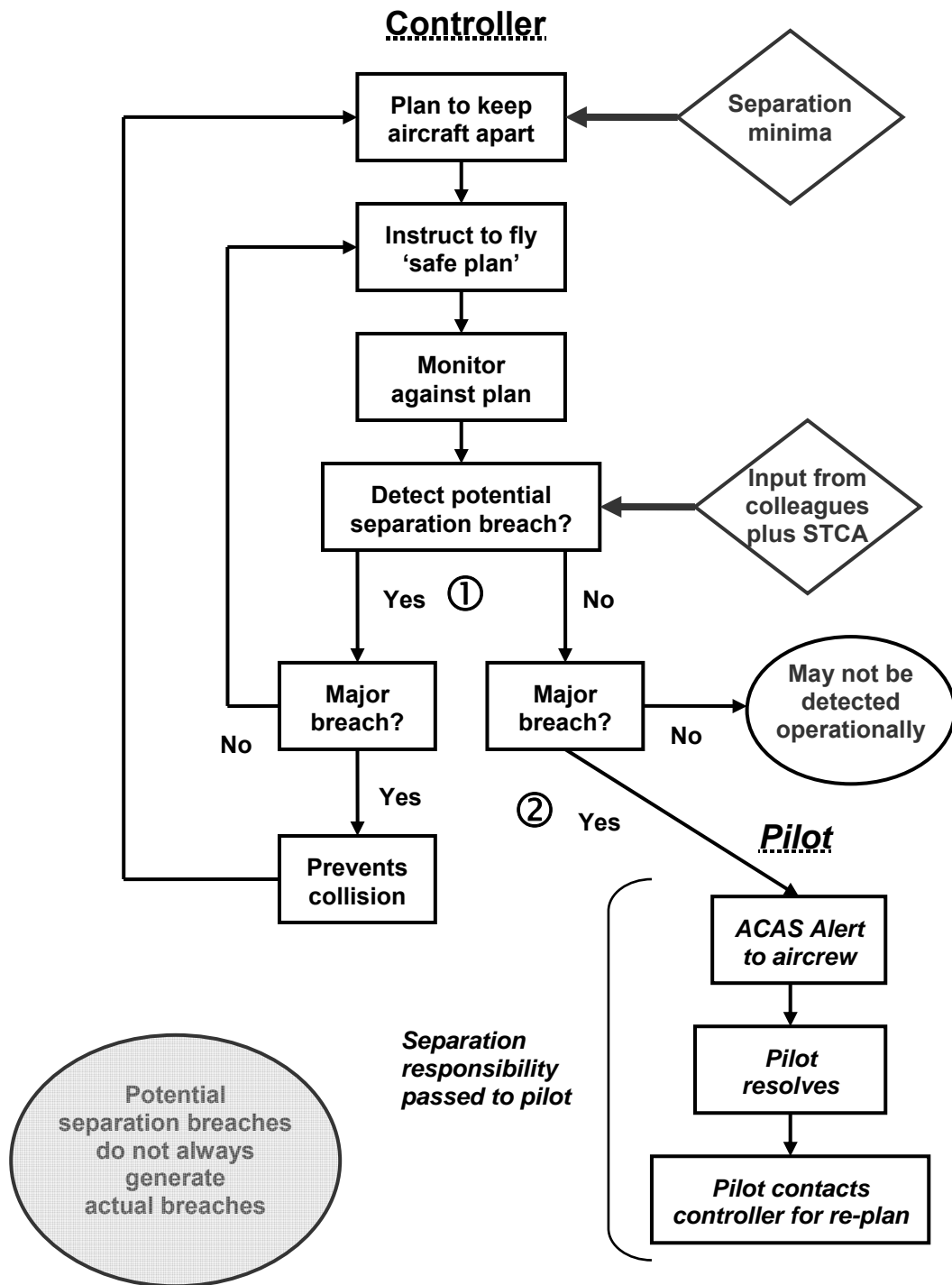
Figure 2. Simplified controller and pilot processes to prevent mid-air collisions

Frequency distribution (logarithmic scale)

Collision Region

'Random Distances' region

Generally reducing system control of incidents

Curve f(x)

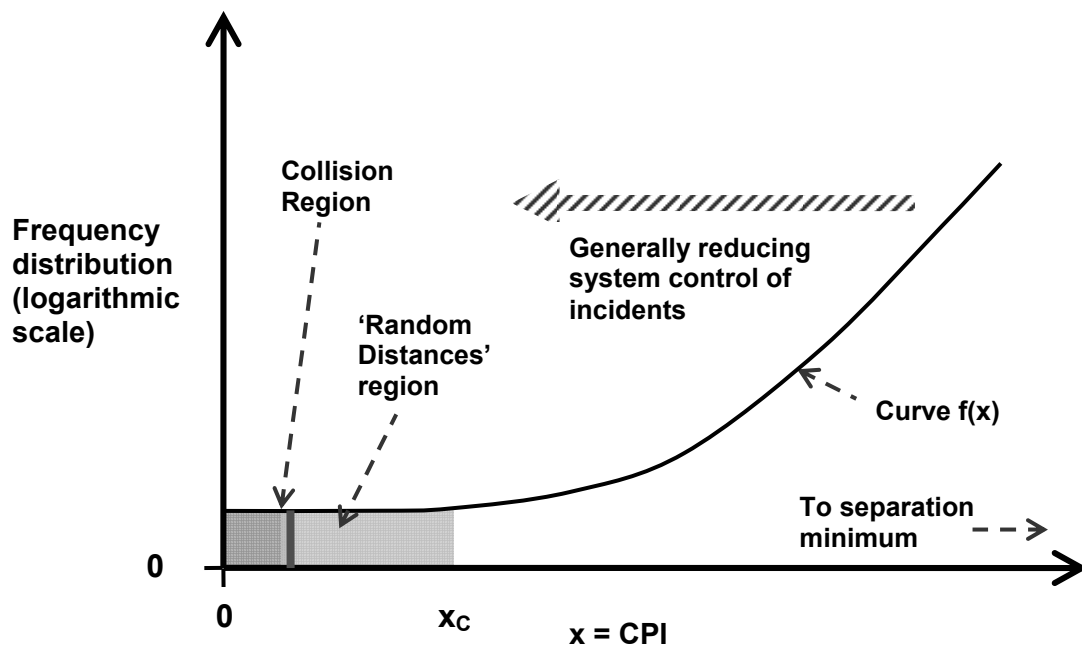**To separation minimum**

0

0

x_C

x = CPI

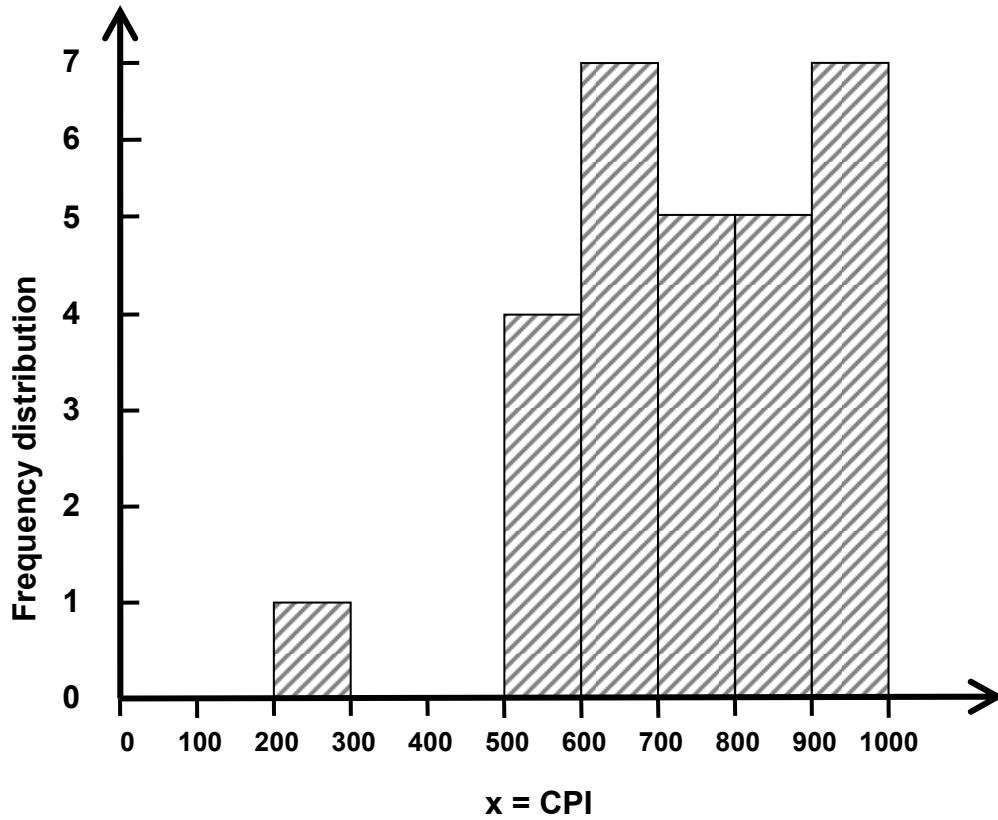Figure 3. Schematic frequency distribution for CPIs

Figure 4. Five Year's actual frequency distribution for CPIs
(From Brooker (2005b): UK Airprox involving CAT aircraft, with STCA and ACAS
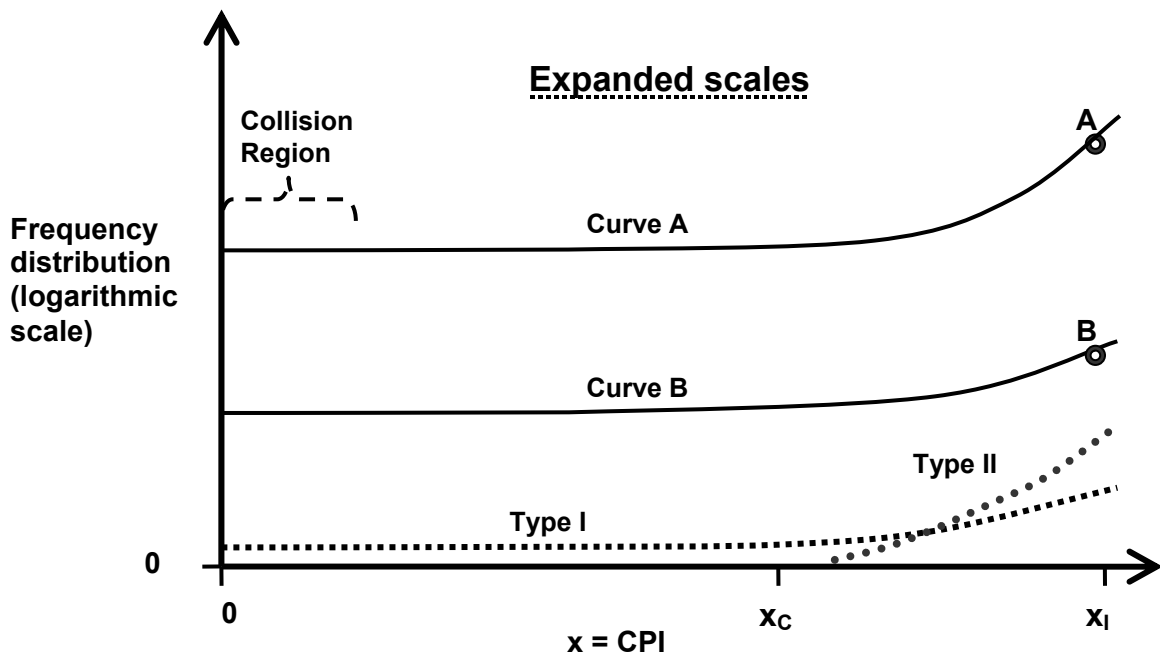indications. 1999 - 2003 inclusive)

Figure 5. Possible frequency distribution for CPIs plus Accident/incident components

Proximity scoring (1)		
Minimum separation		Score
Minimum separation achieved		**0**
+75% Minimum separation		**1**
+50%, =<75% Minimum separation		**3**
>25%, =<50% Minimum separation		**7**
=<25% Minimum separation		**10**

(Figure 14: Proximity scoring 1st criterion "minimum separation")

Figure 6. "Proximity scoring 1st criterion "minimum separation" From SRC (2005a)