Future Concepts for Measuring Aviation Safety, Performance, & Effectiveness

Thomas Michael Lintner; Federal Aviation Administration; Washington, D.C., USA

Steven D. Smith; Federal Aviation Administration; Washington, D.C., USA

Keywords: Aerospace Performance Factor, APF

This paper presents the Aerospace Performance Factor, APF, a balanced approach to measurement using multiple indicators, weighted by expert judgment, and normalized against system operations.

The paper introduces the philosophy and some early concepts that led to the development of the beta version of the APF. Conceptually, the APF should allow any organization that deals with risk management to maximize their operation while maintaining an acceptable level of risk.

This project is resident within the US Federal Aviation Administration Air Traffic Operations Safety organization and also involves the US Naval Safety Center, easyJet Airlines (UK), the Imperial College of London, MITRE, and the US Department of Transportation's VOLPE Center. Although there are concurrent efforts underway to use the AFP methodology in naval flight operations and airline safety, this paper will focus on the perspective of the air navigation service provider (ANSP).

Historically, the aviation system has relied on basic metrics associated with traffic counts, delays, flight cancellations, incidents, accidents, etc. to gauge performance These measurements never really evolved past the 'sophistication' of ratios and have never been fully integrated into a system-wide performance measurement tool.

Unfortunately to this day we have not made much progress and could be criticized for placing more emphasis in trying to redefine events to lessen their significance rather than analyzing the data and taking system wide corrective actions.

It is clear that for any organization to operate at its most efficient, with the lowest level of risk, accurate data reporting, using multiple sources of information is necessary. That data must reflect quantified information, allow for repeatable and consistent assessment of the experience and judgment of subject matter experts, and it must be balanced in the way it reflects positive and negative events. By following that approach, where all data is placed "on the table" for all to see, we develop an 'honest indicator' from which we can identify long term trends and better manage a system.

That is the design goal of the APF. To present a graphical representation of a system's performance over a long period of time compared to a specific baseline. The intent is to include all data so we don't just focus on one aspect of a large scale operation. By showing all data, and the relative significance, or weighing, of one data set to another, we balance the measure.

The APF is designed to operate, not as a sole indicator of a systems performance, but as part of a larger system of SMS's, data mining and analysis, organizational goals measured as an integrated set of metrics, and cost.

An example of this concept can be seen in financial charts, most notably the Dow Jones Industrial Average. Using multiple sets of data, (daily closing prices of different stocks), the DJIA presents a graphic representation of performance over time. The reader is able to assess relative trends and maintain a broad, long term perspective of performance. The APF follows a similar concept in that it allows for the identification of gradual changes which, if measured simply by year-vs.-year, may seem small or go unnoticed. Ideally the APF will also establish an environment within an organization to allow accurate analysis of causes, and implementation of effective corrective actions, instead of a tactical, and generally incomplete, focus on a specific event without system wide follow-up.

Recognizing there would be resistance to developing new data measurement sources, for any number of valid reasons, we elected to use all the current data sets that the FAA has historically used to measure the safety of the system. Those databases are comprised of the following:

- Operational Errors (subdivided as Category A, B, C, and D (or PE)
- Operational Deviations
- Pilot Deviations
- Runway Incursions (subdivided as Category A, B, C, and D)
- Near Midair Collisions
- Vehicle and Pedestrian Deviations

A significant note is the acknowledgement that these metrics, by themselves, are poor indicators of safety. In fact it could be argued that they are indicators of failure, not safety. However, by using these metrics as a starting point, we eliminated arguments regarding the validation of some new type of data source. Instead we capitalized on the historical data that could be translated into a long term trend line. Like the DJIA, we can see the performance of the system over time and thus establish a view of the historical performance and develop an expectation of future direction.

Once we identified the historical safety data we were going to use, we began to develop the mathematical approach we would follow. The APF, at this point in its evolution is a very simple sum of sum methodology. We took the individual safety metrics, subdivided into whatever classification was established (Category A, B, C, etc.), and set them in the numerator.

The denominator was the total aircraft operations for the NAS, segregated by airport (terminal) operations, and enroute (air traffic control center) traffic activity.

We then addressed the most critical aspect of the formulation, the establishment of subjective weighting factors that could be turned into quantified multipliers within the equation. That step first required the development of weighting factor mind-maps. The APF mindmap is shown below and includes all data sets and subsets that comprise the FAA Air Traffic Operations Safety APF.

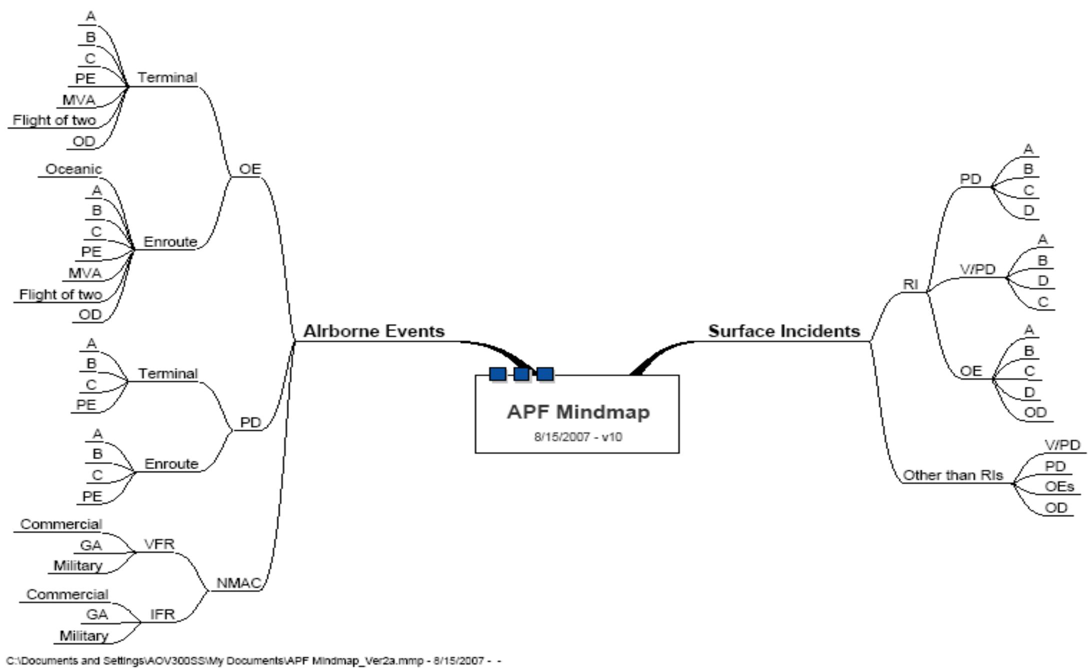

Figure 1 – Mindmap of APF

An immediate problem arose when we first looked at the data. There were operational errors in both the ground and airborne environments, some were more important (or weighted more heavily) but how could we determine how much more important they were? This situation also held for other data elements as well as the interactions between and among the data. We needed a way to "weight" the individual data that was agreeable to all. We decided to employ the Analytical Hierarchy Process and expert opinion.

The Analytical Hierarchy Process (AHP) enables a decision maker to portray the relationships between many facets of a complex problem, and incorporates both quantitative and qualitative information including experience and intuition. AHP does this by first arranging the problem into a hierarchy (*An abstraction of the structure of a system to study the functional interactions of its components and their impacts on the total system)* and a process called "pairwise comparisons", i.e., it takes the functions and compares them two at time (one against another) until all the interactions get a chance to be evaluated.

As we mentioned, the APF methodology can be applied to any organization that needs to assess risk, or performance, based on a complex set of metrics. The following discussion regarding the development of weighting factors is based on an example from the US Naval Safety Center's development efforts of their risk management tool.

We assisted in the development of the mindmap for the Navy and the following example provides a good indication of how non-ANSP data could be used.

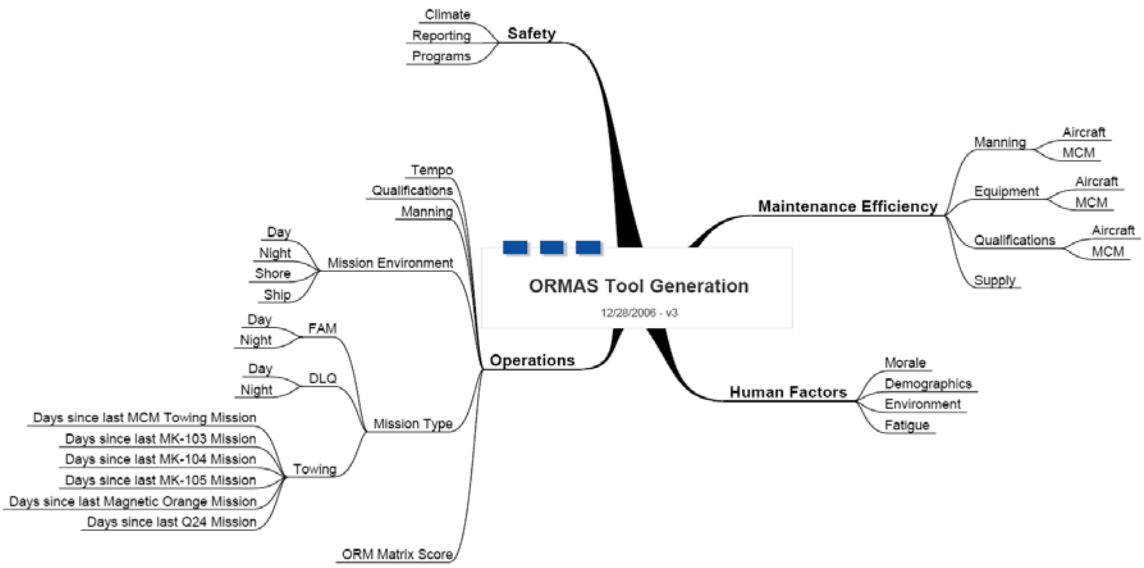

Figure 2 – Mindmap of Navy ORMAS Tool

An Example: **Flight Risk for Naval Aviators**

For this example, look at the data contained in "Mission Environment," a subset of "Operations," in Figure 2. In order to determine the relationship of risk between, the functions of Ship, Shore, Day and Night flying is considered by the expert. Do all four functions carry equal weight or… does one function have more impact on which category of flight exposes the pilot and/or craft to the most risk? In order to determine how much weight each function has, we must compare: 1. Day to Night, 2.Day to Shore, 3. Day to Ship, 4. Night to Shore, 5. Night to Ship and finally 6.Shore to Ship. We do this with the aid of a matrix (Table 1) and a value (or scale) table (Table 2). We use the values (or scales) table to determine the value to place in each cell.

Table 1 – Decision Matrix

	Day	Night	Shore	Ship
Day				
Night				
Shore				
Ship				

Table 2 – Value (or Scales) Table

Intensity	Definition	Explanation
1	Equal importance	Two activities contribute equally
3	Weak importance of one over another	Experience and judgment slightly favor one over the other
5	Essential or strong importance	… strongly favor one over another
7	Very Strong and Demonstrated	…strongly favored and its dominance is demonstrated in practice
9	Absolute importance	Evidence favoring one over another is of the highest possible order
2,4,6,8	Intermediate values between scales	When compromise is needed

Table 3 – Decision Matrix

	Day	Night	Shore	Ship
Day		1/5	1	1/7
Night			5	1/2
Shore				1/5
Ship				

In the above example (Table 3), we have made a judgment that a Night operation is strongly more risky than a Day operation, that a Day and Shore operation are equally important, and that a Ship operation is very strongly more risky than a Day operation. Additionally we felt that a Night operation was strongly more risky than a Shore operation and a Ship operation was weakly more risky than a Night operation. Finally, we felt that a Ship operation was strongly more risky than a Shore operation. You can observe from the matrix that if the row is more important than the column, we use an integer; if the column is more important than the row, we use a fraction. This is for convenience in order to differentiate between the two conditions.

When the matrix is solved, the following risk weights are exposed:

Day Operations – 0.072
Night operations – 0.333
Shore Operations – 0.079
Ship Operations – 0.515

In the case of the Navy's first beta version these values were from only one expert. But in a true AHP expert choice panel, many experts should be polled and their input evaluated as a group.

Returning to the example of the FAA APF, we used a panel of experts and conducted the pair-wise comparison for all the data sets shown on the APF mindmap. Once the weighting factors were determined, the raw numbers from the various databases were populated into the application, multiplied by the weights and then summed. This established the numerator.

The denominator, designed to reflect the 'positive' outcome of the service provider, includes VFR and IFR, and all types of operations (air carrier, general aviation, and military). The traffic is sorted by: terminal surface traffic ("airport operations"), terminal airborne traffic ("airport operations" + "Instrument operations"), and Enroute ("Center Activity").

Once the numerator and denominator were married, the formula produced the APF lines below.

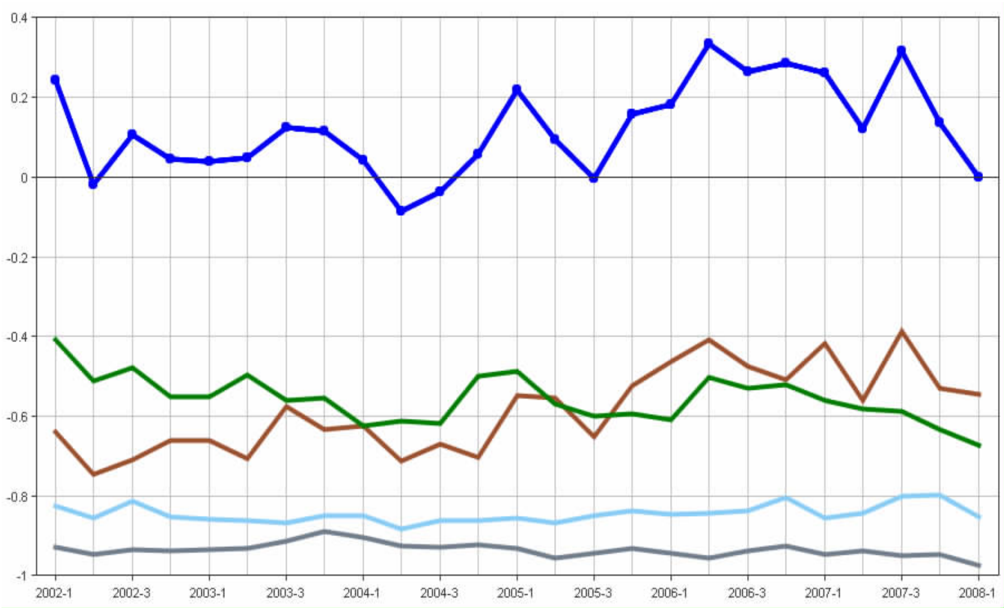

Figure 3 – ATO APF (Beta)

The data can be interpretated based on the following color keys (going from ***bottom line to top line***):

GREY:	Near Mid Air collisions / all traffic
LIGHT BLUE:	Airport surface incidents / airport traffic count.
BROWN:	Terminal operational errors + pilot deviations / airport traffic + terminal instrument operations
GREEN:	Enroute operational errors + pilot deviations / by traffic activity.
DARK BLUE:	Total APF (sum of four subcomponents)

Each of the lower lines represents an APF segment that is then summed into the total APF line shown in dark blue. For example, reading from the bottom, the chart presents the NMAC APF, the surface incident APF, the terminal APF, and the enroute APF. All culminating in the total system wide APF shown by the top line.

Initially, we had envisioned a much more complex equation whereby we would be able to include factors associated with various conditions, i.e., day/night, IFR/VFR, IMC/VMC, and traffic associated by Part 121, 135, 91, military.

However, we discovered that the current databases were not supportive of this level of detail and this aspect of APF development will require a future reassessment of what information is collected and how it's used. So for this phase of the APF development we remain focused on the macro level of the databases.

Key to the interpretation of any data, and the APF is no exception, is the need to establish two things – a baseline and an understanding of where the organization wants to go. This requirement represents one of the major assets of the APF's methodology of presenting data.

Too often we measure changes based on "current year vs. last year," but neglect the subtle year-over-year changes. For example, a 1% rise in air traffic operational errors may not seem to be a problem. But ten years of the same level of performance period equates to a 10% increase, which would probably be unacceptable.

So how did we select the baseline for the APF? It turned out to be more political than mathematical. In May 2005 the Administrator of the FAA established a new organization within the FAA, an internal regulator if you will, to provide oversight to the air traffic organization. At that time the Administrator stated, "We accept the system as it stands today…," and, by so doing, established a type of baseline.

Our logic was that we needed the ability to monitor the overall system performance as it was in May 2005 and track long term changes to make sure the system did not go below that level. Thus, the level of performance of the system as it was on that date became our beta baseline. Ideally, the goal would be to remain at or below the baseline (below being better performance).

If you look at the ATO APF (beta) above, you will see the 0 on the Y-axis set at 2005-3 (third quarter of FY2005). You will note that the ATO has remained above the baseline until the 1st quarter of FY08.

Over the past several years of development, we have received numerous questions regarding the APF which has helped support development activity. In an effort to highlight key aspects of the APF as well as future plans, we would like to offer the information in a ***"Frequently Asked Questions"*** format.

What is the APF going to be used for?
The APF will be used for macro management of the system using a graphical representation of performance and safety measured off of an agreed baseline. Instead of having organizational goals targeted to only one aspect of a systems performance, future goals will be based on a holistic representation of safety and performance.

Can you determine why the APF line moved simply by looking at the chart?
You cannot. It remains the responsibility of the organization, after seeing a change in the APF line, to drill down into the data to determine why the data line moved. To support this task, a data mining tool is being developed that will allow the user to push down into the application, by segment of the operation, geographical region and facility, and type of incident, to pin point the cause of the change.

Can the APF tell if the system is safe?
The APF plots out the system's safety and performance over a long time frame and thus supports the establishment of a reference for a tolerable level of safety, or an acceptable level of risk. If we see the users of the system continue to fly, buy tickets, and operate aircraft, we could assume they, the ultimate judges of safety, are comfortable with the level of performance. If you accept that perspective, then by maintaining (or improving) the level of performance, we are in reality, operating at the user's acceptable level of safety and risk.

Are the changes statistically significant?
The APF is not a forecasting tool. It accepts the data for what it is and graphically presents it over time. In a similar fashion to the Dow Jones average, it shows historical performance and the user is left to judge whether past performance is an indication to future performance or risk. As such, it is up to the user to judge for him or herself whether the change is significant or not.

What else can the APF be used for?
The APF can be used as a model for future investment based decisions. For example, if you have €10M to invest in safety you could use the APF for comparative values. Assume you can use the €10M to reduce a safety event to zero. Now, go into the APF and set operational errors to zero and determine the total change of the APF. Then set runway incursions to zero and find the change of the total APF. Continuing the example, if the APF drops 3% with a total reduction of operational errors but drops 10% if

we eliminate runway incursions, we should focus on runway incursions for the better return on the investment of our safety Euros.

What are the next steps with the APF?

We will be applying the same concepts to the establishment of an APF for performance. By examining the impact of delays, weather, airport acceptance rates, holding, etc., we will be able to develop a balanced approach to operational performance. By incorporating cost data, and combining the safety APF, we will be able to develop a tool to provide the maximum level of operational performance, maintain it at an acceptable level of safety and risk, and within targeted cost levels. In time, we will be able to develop tactical metrics, including such measurements as dynamic sector density as a function of risk, and real-time changes in convective activity, to provide ANSPs a tactical tool, similar to a flight director, to maintain optimum efficiency and safety within the operation.

Will the APF be a stand alone tool for managing safety?

Effective safety management needs to be based on obtaining accurate data, an effective SMS process, in-depth operational analysis, and executing system wide steps to address identified risks. The APF is part of that system and will allow for trend identification and analysis, system wide data mining in future versions, and the ability for senior management to assess long term events vs. tactical event assessment.

Biography

Mr. Thomas Michael Lintner, Program Manager, Safety Services, Federal Aviation Administration, Air Traffic Organization, 490 L'Enfant Plaza Suite 7225, Washington, DC 20024, USA, telephone - +1.202.385.4536, email- Thomas.Lintner@faa.gov, T.Michael@neworkcityirish.us. Tom is a graduate from Dowling College, in Oakdale New York, where he majored in business management with a minor in aviation science and physics. He was an air traffic controller at New York City's LaGuardia Airport and the New York Terminal Radar Approach Control Facility prior to becoming the Manager of the Air Traffic Investigations Division in Washington, DC. An active pilot, Tom holds an Airline Transport Pilot license, is a Certified Flight Instructor, and an Airline Dispatcher.

Mr. Steven D. Smith, Program Manager, Runway Safety Program, Federal Aviation Administration, 490 L'Enfant Plaza, Suite 7225, Washington, D.C. 20024 USA Telephone - +1.202.510.7906, email – steven.d.smith@faa.gov. Steve is a graduate of St. Martins College, in Lacey Washington and attended graduate system safety program studies at the University of Southern California, Los Angeles. Steve is a member of the System Safety Society, Capital Chapter and was the General Conference Chair for the 1997 International System Safety Conference held in Washington, D.C. He has served as the Director of Government and Inter-Society Affairs for the System Safety Society.