ASSESSING TRAJECTORY PREDICTION PERFORMANCE – METRICS DEFINITION

Stéphane Mondoloni, CSSI Inc., Washington, DC
Sipke Swierstra, Eurocontrol HQ, Brussels, Belgium
Mike Paglione, FAA WJHTC, Atlantic City, NJ

Abstract

As one of the central components of Air Traffic Management (ATM) automation tools, trajectory predictors have a significant impact on the performance of ATM automation and hence the ATM system. Building on prior efforts, this paper applies a framework to assist in the development of performance metrics for trajectory predictors. Key performance areas are defined for trajectory prediction based upon existing trajectory predictor (TP) performance evaluations. Basic metrics are described within the accuracy key performance area for both input and output metrics. The basic metrics rely on precise definitions of events. Issues associated with the definition of events are discussed and approaches are provided for dealing with these. Specifiers are defined to further refine the definition of metrics in all performance areas. The application of basic metrics, events and specifiers is illustrated by drawing upon examples from the literature.

An example of the impact of trajectory prediction accuracy on conflict probe performance is provided to illustrate considerations that must be given to the impact on higher-level systems when developing metrics for TP performance.

Introduction

The future of Air Traffic Management (ATM) relies, in part, on the use of decision support tools (DST) in order to deliver improved services to the user community under increasing traffic demand [1-3]. These decision support tools include tools capable of achieving conflict detection and resolution, trial planning, controller advisories for metering and sequencing and traffic load forecasting and weather impact assessment. One common theme in all of these DST is the need for trajectory prediction. Even concepts relying on 4-D contracts and required times of arrival require trajectory prediction to: determine valid contracts, respond to perturbations, evaluate paths between contract points, deal with unequipped flights during transition and provide information necessary for backup of failures in airborne systems and communications. The result of this reliance on TP is that the performance of the DST, and hence of the ATM system, is affected by the performance of the TP.

The cascading of performance is illustrated in the ICAO performance hierarchy as shown in Figure 1. TP performance resides in the lower levels of the hierarchy and affects the performance of DSTs as previously discussed. Multiple DSTs may be affected by the performance of a single TP. When combined with other systems, the DST would contribute to the performance of the system in terms of the delivery of specific services and system functions. The effect of interoperability issues would need to be considered as well. Eventually, the performance of the system at the highest level would be impacted (e.g., safety, cost).

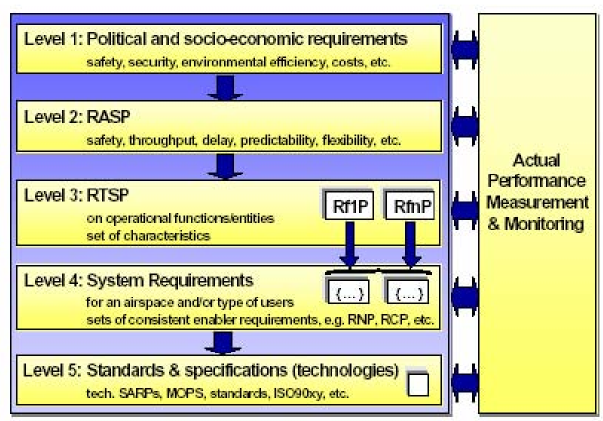

Figure 1. ICAO ATM Performance Hierarchy [4]

While dependency between the levels is established, the details of the dependency are omitted. The structure does not require metrics at one level to be expressible as simple relationships of the metrics at the lower level. However, it is clear that the purpose of measurement of performance at the lower levels is to understand and improve the behavior of the system at a higher level. This leads to the need for interaction between designers of the overall system concept and of its more detailed components. In this instance, it emphasizes the need to develop metrics for TP performance that consider the impact on performance of the client DST applications. Further, the performance aspects of the DST being considered ought to eventually impact the highest level of the hierarchy. This is equivalent to stating that TP performance metrics are only meaningful if they can be traced to a higher-level system objective.

The ICAO performance hierarchy has been developed as a result of the emphasis by Air Navigation Service Providers (ANSP) on performance-based ATM. This approach imposes a need for measurement and understanding of ATM performance. The dependency of system performance on TP performance ultimately requires the development of TP performance metrics to be able to quantify performance in a manner that can be commonly understood.

Prior metrics developed for TP performance have largely been accomplished in an ad hoc fashion, primarily by DST developers interested in TP-led improvements to their DSTs (e.g., [5-16]). As part of the joint FAA/Eurocontrol R&D Action Plan 16 investigating common trajectory prediction capabilities, we have sought to harmonize these metrics. In this paper, we apply the work of [17] and [18] to develop an approach for defining consistent metrics for trajectory predictors. By using this approach, TP metrics can be defined in a unique and consistent manner easily suitable for automation. In this paper, we also describe an example of the impact of trajectory prediction on the next higher-level system and the implications that this impact analysis has on TP metrics.

Metrics Taxonomy

The quality of service (QoS) framework, ISO-13236 [17], has been discussed in the context of air traffic management problems in [18]. Here, we apply this framework to the development of metrics for the trajectory prediction problem.

At the highest level, a set of performance characteristic groups is defined. These corresponding characteristics define the "fundamental aspect of QoS that is to be managed." Groups are referred to herein as TP key performance areas (TP-KPA) and are discussed below. Within each key performance area, one or more basic metrics are defined that capture different aspects or dimensions of performance within the TP-KPA.

Basic metrics are used as concepts to create actual metrics through the application of *specialization* and *derivation*. We can consider an analogy to object-oriented programming in which basic metrics define an abstract class. Just as we must extend the class to instantiate it, we must specialize the basic metrics to apply them. Figure 2 illustrates the hierarchy of metrics as described above.

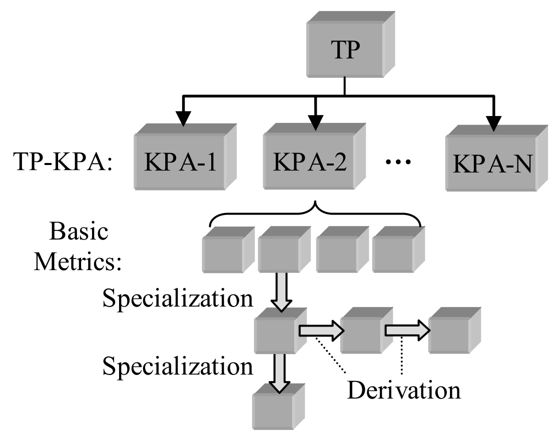

Figure 2. Hierarchy of Metrics

Specialization

Basic metrics can be applied to a wide variety of circumstances. For example, under the key performance area of accuracy, one basic metric would be altitude error. However this altitude error can be computed: at a 10-minute look-ahead from the prediction time, upon crossing a specified location, for a specified aircraft type, or for a variety of other circumstances. Specialization is the means through which we define these circumstances.

Specialization is accomplished through the sequential application of *specifiers*. These can be categorized into the following broad categories, consistent with the ISO categorization:

- *Object/entity* over which the characteristic is established
- *Location* at which the characteristic is described
- *Process* in which the characteristic is defined
- *Event* that defines the extent being considered.

The categorization of specifiers allows further understanding, but the classification of specifiers is subjective and can be altered in the future based upon community consensus.

Object/Entity Specifiers

These specifiers focus the application of a metric onto specific entities. For example, an entity specifier can focus an error on only the B757-200 models, only the turboprops, or only those flights equipped with FMS. Some example entity specifiers include: aircraft model, engine model, equipage level, operator, airport, fix, air traffic control center, sector, or flight level.

Location Specifiers

These specifiers are used to limit the geographical scope of a metric. For example, prediction errors can be measured within a specified geographical boundary, or within a certain distance of a specified point. Location specifiers include two dimensional points (latitude/longitude), three-dimensional points (2-D plus altitude), gates, and airspace volumes.

Many entity specifiers also have a location specifier associated with them. For example, a center and a sector have associated locations, but also represent entities in the system.

Process Specifiers

Processes have a start and end point that are described through events. Within a process, a consistent set of behavior can be established. Process specifiers can be used to describe aircraft that are controlled in a similar fashion. Examples of process specifiers include: climb, cruise, step climb, hold, initial descent, fly-by turn, LNAV, VNAV, constant CAS climb, etc.

The above discussion has focused on the aircraft. However, process specifiers can also be used to describe specific processes within TP evaluation (e.g., the preparation process). These are useful when evaluating such performance areas as TP calculation speed and prediction stability.

One can apply process specifiers during TP evaluation to consider flights that are operating in a similar manner.

Events

For TP performance measurement, events can be associated with both processes and locations. Events designate the start and the end of a process; for example, the cruise segment is delineated by the top-of-climb and the top-of-descent (TOD) events. Events can be associated with locations; for example arriving at a location, crossing into a volume, or crossing a gate.

Events can also occur during the production of a trajectory prediction. For example, when measuring calculation speed, the output of a prediction would be a relevant event.

Examples of events specifiers include: the initial condition, top-of-descent, crossing an intermediate altitude, crossing a metering fix reference, reaching the Mach/CAS transition point, reaching a specified look-ahead time, or receipt by the TP of the last data element required for prediction.

Derivation

Derivation typically describes the application of statistical methods to existing metrics in order to derive a new metric. These statistical measures are applied to a *set* of data. This set of data can be obtained through the application of various specifiers (e.g., all flights through a certain fix, all jets, all non-FMS equipped jets, etc.). Examples of derivation include maximum, mean, standard deviation, root-mean-square (rms), and N^{th} percentile.

Metric Signature

Once a metric has been fully defined through the application of specifiers, the metric can be uniquely identified through its *signature*. This term simply refers to a definition that includes:

- Name of the metric
- Definition explaining its purpose and area of application
- Statement indicating how the measure is quantified
- An indication of how it is derived
- An indication of specifiers
- A description of how the measurement is calculated

Consider the basic metric of altitude error under the key performance area of accuracy. To be useful, this metric requires specifiers. We must specify the prediction event, e.g. upon crossing the 10,000 ft mark on climb. We must also specify the measurement event, e.g., at a look-ahead time of 5 minutes.

Derivation will often require a set of data. This set may be derived from many measurements on a single flight, e.g., peak co-temporal altitude error during the climb phase for a prediction initiated upon crossing the 10,000 ft point. The set may also be derived from measurements on multiple flights, such as the standard deviation of altitude error at a look-ahead time of 5 minutes over all jets. Note the application of specifiers to help define the set of data in a manner consistent with Figure 2.

Treatment of Events

Precise metrics require the precise definition of events, since metrics are often based upon measurement of or at events. It is useful to classify events as follows:

- Those that occur upon reaching a value of a specific parameter. Examples of this include reaching a specified speed, reaching a specific altitude or reaching a certain look-ahead time.
- Those that occur at a discontinuity. Examples of these events include top-of-climb, sequencing a waypoint, and changes in aircraft configuration.
- Those that occur at an extreme. One example of such an event is the closest point of approach to a point when sequencing a waypoint.
- Those that occur upon initiation or completion of a portion of the trajectory prediction process. This type of event would be useful in the definition of calculation speed-type metrics.

Figure 3 illustrates the time of occurrence of several of these types of events when they can be measured by a parameter.

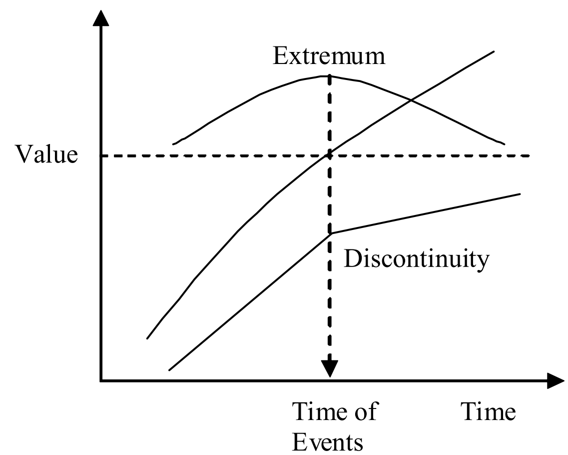

Figure 3. Illustration of Types of Events

When defining events of each type defined above, the precise location of events can be complicated by several factors:

1. Measurement noise – noise on the signal used to detect the event may make it difficult to precisely determine the location of the event.
2. Discrete samples – discrete measurements limit the precision with which either of the preceding events can be established.
3. Smoothing around discontinuities / Events of finite duration – smoothing removes discontinuities and spreads an event so that it no longer occurs at a discrete point. Many situations labeled as events are not instantaneous. The duration of these situations can lead to

errors on metrics based upon the precise location of events. For example, configuration changes can be considered an event in a model; however, in reality, there is a starting event and an ending event corresponding to the configuration change. Other such cases include: start and end of turn initiation, flap retraction, rotation, and power changes requiring spooling of engines.

4. Small offsets from a steady state – steady-state errors can make it difficult to identify events, depending on the method used to identify them. An example includes a course intercept with an offset error from the intercepted course.

Measurement noise can complicate the value-reaching type events by rendering the reaching of a value ambiguous (see Figure 4). Noisy measurement examples might include a Mach/CAS transition point where the speed is derived through radar data. Filtering and thresholds can be applied to improve the identification of the event. However, filtering schemes will often introduce lags, so care must be taken to ensure that events are not biased as a result of the filtering.

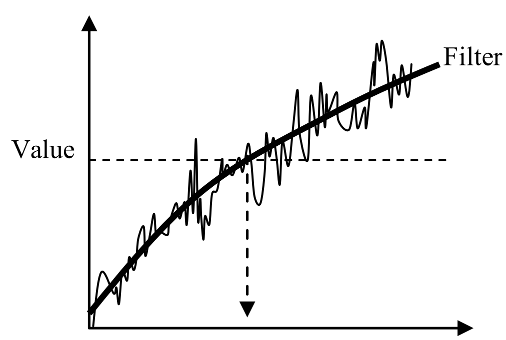

Figure 4. Impact of Noise on Value-Event Identification

Discrete signals can be converted into continuous signals through an appropriate choice of interpolation. The choice of interpolation should consider the physics of the situation and the event being measured. For example, linear interpolation during a turn with few points will not be appropriate if we are seeking the minimum distance to a fly-by waypoint. The interval of the discrete signal will impact the accuracy with which an event can be identified. Ultimately, this is but one component in measurement error. Once the discrete signal has been converted into a continuous one, the remaining problems are identical to those associated with continuous signals.

Dealing with events at discontinuities that are actually finite-duration transients can require the definition of both a start and end event for precision. For these types of events, recasting the problem in terms of a variable with a discontinuity is useful for identification of the event. Examples of these include:

- Start / end of turn initiation. A discontinuity in heading is sought for a turn. Since the turn is finite-duration, a discontinuity will be replaced with a more gradual change (See Figure 5).
- Start / end of flap retraction to position X. Both are valid events. When comparing to an abstraction with an instantaneous change, judgment is required to determine if the mean is useful for comparison.
- Top of descent. A discontinuity in descent rate is sought for top-of-descent. At TOD, an instantaneous change in descent rate does not occur, thus some ambiguity exists regarding the top-of-descent location.

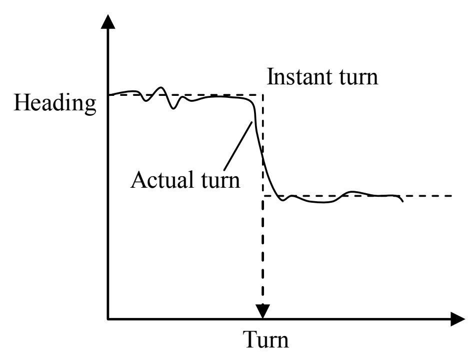

Figure 5. Ambiguity in Event Location

The introduction of uncertainty regarding the location of an event introduces an additional measurement error into the problem. The impact of the event location uncertainty can be quantified by considering the effect of shifts in the event location on the metric being considered. For example, errors in locating the top-of-descent will impact the peak altitude error in descent. Estimating the magnitude of the event-location error can be used to predict the measurement uncertainty due to such an error. If more precise measurements are needed, a more precise event definition may be required.

TP Key Performance Areas

When referring to TP performance, there are several aspects that may be considered; we have referred to these aspects as TP key performance areas (TP-KPA). These areas attempt to capture the fundamental areas of performance that can be evaluated. In order to identify the TP-KPA, we conducted a literature search of existing metrics typically used to evaluate TPs supporting DSTs (e.g., [5-16]). We also included some additional areas to incorporate issues raised in [19] regarding interoperability.

Below we describe the TP-KPAs. Within each TP-KPA, there can be multiple dimensions of performance.

Accuracy

When we speak of a TP being accurate, we are discussing its ability to precisely predict the future trajectory of a flight. Accuracy can be measured in multiple dimensions including lateral, vertical, longitudinal and temporal.

Confidence

Having high confidence in a TP indicates that a specified level of accuracy will be achieved with a high probability. Different values of confidence can exist for each accuracy measure.

Speed

The speed of the TP refers to the time required to deliver a prediction subsequent to a request. Speed measures may be conditional on the type of requests.

Stability

We speak of a stable TP if the current prediction will not undergo significant changes between predictions as a result of the prediction process.

Reliability

The reliability of the system represents its "ability to perform its required functions under stated conditions for a specified amount of time [20]".

Uniformity

This KPA only applies when investigating interoperability. Two trajectory predictors are said to be uniform if their predictions agree. This characteristic is analogous to accuracy with a comparison being made between predictors instead of between a predictor and "truth" data.

Synchronization

This characteristic only applies when investigating interoperability. Two predictors are synchronized when predictions occur simultaneously.

Input, Output, Outcome, Impact

The IOOI (input, output, outcome, impact) framework has been discussed in a variety of performance measurement settings [21, 22]. This framework is useful for structuring metrics associated with a system. In this case, our system, the TP, responds to certain inputs to produce a prediction as output. It is clear that measurement of TP performance cannot be conducted through metrics on output alone. Since the quality and conditions of the input will affect the output, measurements of both input and output are required to quantify TP performance. Outcome metrics refer to the combination of both input and output metrics that describe the performance of the TP. This does not imply an arithmetic combination, but may require applying input metrics as qualifiers on which output samples are valid.

Ultimately, the performance of the TP is being evaluated for the purposes of inclusion into a larger system, potentially subject to a Required Total System Performance (RTSP). It is important to ensure that lower-level TP measures are useful for determining the impact of TP performance in such a way as to achieve the RTSP.

Input Metrics

Input metrics attempt to describe the quality and conditions of TP input in order to make an assessment of the TP performance subject to that input. Many types of information are required to obtain a trajectory prediction. As the TP structure (see [23]) illustrates, input data includes:

- Aircraft performance
- Meteorological conditions
- Lateral intent
- Constraints
- Vertical / speed intent
- Initial aircraft state

These are broad categories that can encompass many sources of data from ground data to aircraft-derived information. Knowledge of the quality of input can be used in several ways:

- To provide values of input data quality when setting performance requirements on TP output,
- To provide requirements on input data types and performance in order to obtain a certain required performance level from a TP, or
- To identify areas of improvement in input data quality through methods such as airborne data exchange.

Basic Metrics

We provide examples of basic metrics in the performance area of accuracy. Input data is considered accurate when the data that is provided to the trajectory predictor replicates the value that the input data represents.

<u>Vector wind error</u> – The magnitude of the vector difference between the actual wind at a point and time and the predicted wind at the same point and time. This input error can significantly affect along-track errors and couple with altitude errors during climb and descent.

<u>Speed intent error</u> – The difference in the intended speed input to the TP versus the actual speed target applied by the aircraft. Note that this may apply to multiple speed intents during climb, cruise or descent. Speed intent errors will directly affect along-track errors and will couple with altitude errors.

<u>Lateral route error</u> – Measures the difference in the input route lateral path and the route actually flown. Lateral route error is the closest distance from the actual location of the flight at a specified event to the input route of flight. Lateral route errors will directly affect along-track errors and errors in the vertical profile.

<u>Input weight error</u> – The difference in the weight of the aircraft used for prediction and the weight of the aircraft at the initial condition. Weight errors affect vertical profiles during transition. Through the effect of error on fuel consumption, the impact is aggravated with look-ahead time.

<u>Initial condition positional error</u> – Measures the difference in location of the actual aircraft at the initial condition versus the input initial position. This difference may be the result of synchronization errors, or positional uncertainty in ground-based sensors. Initial condition uncertainty propagates forward.

Application of Specifiers

Input metrics are often not labeled as such in the trajectory prediction literature. However, prior efforts have often qualified trajectory prediction accuracy through the application of input metrics. Examples, from prior efforts, are described below including the specifiers that are applied.

- Standard deviation in thrust error at a specified altitude across all descent runs [24]. A basic metric (thrust error) has specifiers in location (altitude), and process (descent) applied, together with the standard deviation as derivation.
- Mean lateral route intent error for arrivals [25]. The basic metric of lateral route intent error is specified by process

(arrival) and the mean (derivation) is taken.

- Standard deviation of along track wind error at specified pressure altitude [12]. The pressure altitude specifier is applied to the along-track wind error basic metric together with the standard deviation as derivation.

Output Metrics

These metrics are obtained by taking measurements on the output of the TP. Here we describe the output accuracy of the TP as an example output metric. One measure of accuracy for a trajectory predictor is the difference between a predicted quantity versus a measured quantity at a specified event. These "quantities" can refer to the time at an event, or the measure of any of the aircraft state variables at the event. Prior to obtaining output accuracy metrics, we must establish reference predictions for comparison.

Basic Metrics

Example basic output metrics for trajectory prediction accuracy are described below.

Time error – Measures the difference in the time of occurrence of an event between a predicted trajectory and the corresponding actual, or "truth" trajectory. Tools using a prediction for synchronizing traffic (e.g., sequencing and merging) rely on time predictions at a point.

Altitude error – Expresses the difference in the predicted altitude of a flight and the actual altitude of the flight at a specified event. Tools using a prediction of aircraft location for separation, load-balancing, or altitude clearance (e.g., CD&R, traffic load forecasting) rely on altitude information.

Horizontal error – Measures the distance between the predicted location of the flight and the actual location of the flight in the horizontal plane. This error is decomposed into cross-track and along-track components.

Cross-track error – Measures the difference in the position of the predicted location of the flight and the actual location of the flight, projected onto a vector perpendicular to the actual course at a specified event (see Figure 6). Tools using a prediction of aircraft location for separation, or load-balancing rely on accurate positional information.

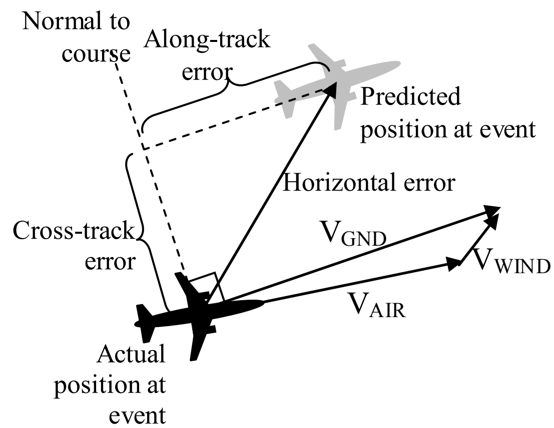

Figure 6. Cross-Track and Along-Track Errors

Along-track error – Measures the difference in the position of the predicted location of the flight and the actual location of the flight, projected onto the actual course at the time of a specified event. Tools using a prediction of aircraft location for separation, or load-balancing rely on accurate position information.

Application of Specifiers

Output metrics, as obtained from the literature illustrate how the application of specifiers and derivation to basic metrics can be used to express these metrics.

- Standard deviation of along-track error measured at a specified fix for FMS-equipped aircraft [12]. The along-track error basic metric has both a location (fix) and entity (FMS-equipped) specifier applied. Measurement is conducted at the event of crossing the fix. The standard deviation is an example of derivation.
- Probability of falling within a range of altitude error during vertical transition at a specified look-ahead time [26]. The altitude error basic metric has a process (vertical transition) specifier applied. The error is measured at a look-ahead time event. Derivation is applied by

investigating the probability of the error falling within a range.

- Count of segments within a range of along-track errors at a look-ahead time of 5 minutes for all MD80 flights [27]. The along-track error basic metric is measured at the 5-minute look-ahead event for only MD80 aircraft (object specifier). Derivation is further applied to obtain a count of segments meeting a condition.

Outcome Metrics

Outcome metrics combine information from both input and output metrics to provide trajectory predictor performance indicators. Since the quality of the input can have significant impact on TP output, indicators of TP performance must qualify the quality of the input in order to be meaningful. As an example, consider a single trajectory predictor operating on flight plan data with erroneous aircraft models, large errors in speed intent, and poor wind data. Under these circumstances, the TP will likely provide poor output accuracy. However, the same TP with improved input data quality could yield high output accuracy. In addition to data quality, other input factors such as the conditions under which the TP has been evaluated are critical determinants of the output metrics.

One method of assuring that we are evaluating TP performance, versus the effect of input data, is to evaluate the performance using reference data which has eliminated input errors. This may be achievable in some circumstances (e.g., intent information), for the purposes of evaluating the TP performance under ideal conditions. It is also necessary, for the purpose of eventually evaluating the impact at higher levels, to measure the performance of the TP under conditions representative of operational conditions.

We define outcome metrics as simply output metrics that are obtained under a controlled set of input conditions. Figure 7 illustrates the concept. A TP is evaluated under a sample of input data to produce a sample of output data. Input metrics can be measured for each input sample element and output metrics for each output sample element. Derivation can be applied to these metrics to obtain statistics on the input and output metrics. Outcome metrics are the derived output metrics (output statistics) obtained given an input sample with specified statistics on the input.

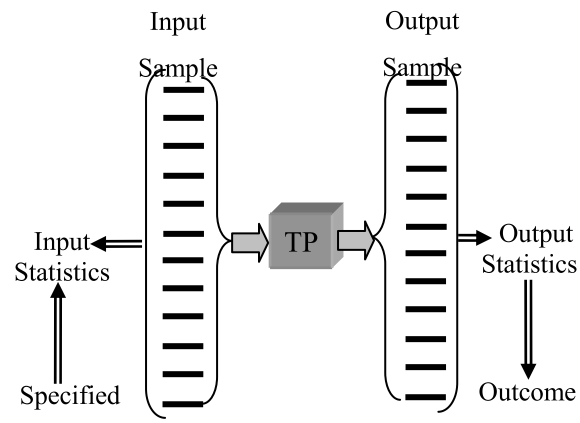

Figure 7. Output under Specified Conditions

As an example, we can consider output metrics (rms along-track error with a 10-minute look ahead) given a set of input data. The input data can be specified by describing input statistics such as:

- Traffic mix, including the ratio of specific aircraft models
- The frequency and mix of instructions provided by ATC
- The number and magnitude of restrictions in the airspace under consideration
- The mix of air carrier operators when significant differences in procedures exist
- Flight duration and the phase of flight under consideration
- Aircraft equipage if relevant to trajectory prediction, such as FMS-equipped aircraft
- The speed intent distribution (e.g., a typical mix of fast and slow aircraft)
- The atmospheric conditions (e.g., typical temperature and wind day)
- The accuracy of the input data (e.g., speed, flight path, wind, weight)

In practice, specification of the input data will be determined by the conditions under which the TP will be operating. Since one is ultimately interested in evaluating the impact on higher-level systems (e.g. decision support tools using TP functionality), the operational conditions must be considered within the context of the application. Different applications can operate under very different environments, and this must be considered when specifying input metrics. When evaluating nominal behavior of an en route system, input conditions corresponding to "typical cruise conditions" would be considered. If one is interested in operational acceptability, one may be concerned with TP performance under input errors at the 95th percentile.

Evaluation of Impact

The consideration of impact is critical in the determination of relevant metrics for any application. Without impact on higher-level systems, performance of a TP is strictly an intellectual exercise. Some of the reasons for considering the impact of TP are as follows:

- Improvements to trajectory prediction are possible and one wishes to determine the system-wide effect of these improvements.
- Applications relying on trajectory predictors require TP performance data to determine how these applications will perform.
- Development of performance requirements for TPs. These requirements help achieve performance in client applications that improve the ATM system.

Figure 8 illustrates the impact of TP performance on client application performance. Under specified input conditions, the TP will produce a certain performance, as expressed through TP metrics. The client performance can then be determined, given the TP performance and input conditions. This approach is ideal for imposing requirements on TP when the client is subject to minimum performance requirements under defined conditions. In practice, the complex interaction between the client application and the TP makes it difficult to obtain simple relationships between TP and client performance.

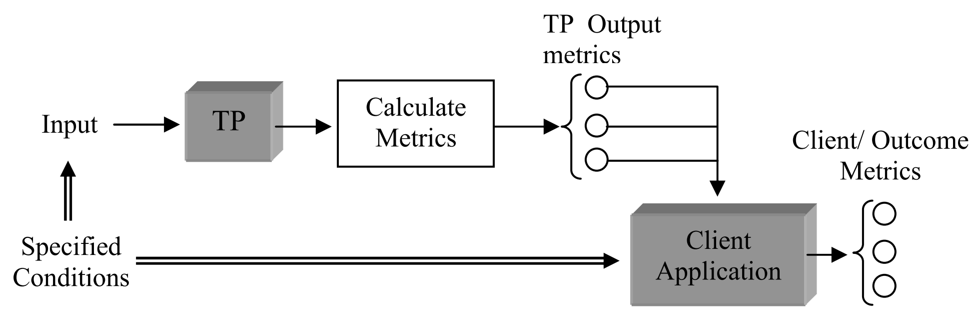

Figure 8. Illustration of Application of TP Metrics to Higher-Level Applications

As an example of the application of impact metrics, we consider a deterministic conflict probe decision support application. Trajectory prediction is used to forecast the position of all aircraft in a neighborhood within some look-ahead horizon. Predicted aircraft positions are compared to determine whether aircraft will violate separation at some future time. Since the prediction is subject to uncertainty, some conflicts will be predicted that will not occur (false alert), and some conflicts will not be predicted (missed alert). Through normalization as described in [26], one can define a false alert rate (FAR) and a missed alert rate (MAR) as performance measures for the conflict probe application. However, designers of conflict probe applications can often conduct a trade between missed alerts and false alerts through the application of buffers on detection. This trade space is expressed through "System Operating Characteristic" (SOC) curves as shown in Figure 9 (e.g., [28]).

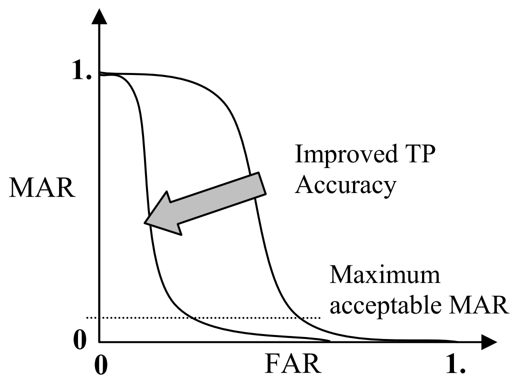

Figure 9. Conflict Probe SOC Curve

The desired operating point is to have a zero MAR and FAR; however, trajectory prediction error makes this impossible. Designers must determine the desired operating point. For example, if the missed alert rate must not exceed a level for operational acceptability, the minimum FAR that meets this goal would likely be selected as an operating point. The *impact* of improved accuracy of the trajectory predictor is the shifting of the curve to the left (closer to the desired operating point).

An example of a simple metric that has been used in the past (e.g., [5]) is the linear growth of along-track error. This scheme explicitly relies on the notion that the along-track error grows linearly. For an application in cruise, with well-defined input conditions, this approach can yield a straightforward numerical relationship between TP accuracy and the conflict probe SOC curves. However, additional dimensions are sometimes relevant (climb/descent and lateral errors), and along-track error will not grow linearly for all error sources. For these cases, metrics expressing the various dimensions of TP accuracy, and their correlation, can be used by the DST evaluators to estimate DST performance under the relevant operational conditions. This approach helps address the two issues of: evaluating DST performance improvements when TPs improve, and also investigating new DST performance given the level of TP performance. Regarding TP requirements, if a DST is subject to performance requirements, trade studies can be conducted to determine appropriate TP metrics that meet the DST performance requirements.

The example of the conflict probe illustrates the need to consider the higher-level application when developing metrics for TP, as also illustrated in [29]. In this case, co-temporal accuracy metrics are likely to be more directly relevant than co-spatial errors. Examples of other considerations would be the events being considered such as the look-ahead time. Certain applications requiring longer times (e.g., TFM) would require metrics tailored to these look-ahead times. Metering and sequencing applications are typically concerned with events at metering fixes.

Summary & Conclusion

This paper has defined an approach for defining TP metrics extending the work of [17, 18]. We have provided examples of specifiers and events to be applied to basic TP metrics. Issues associated with event definition have been presented together with suggested methods for resolving. Seven TP key performance areas have been identified and defined: accuracy, confidence, speed, stability, reliability, uniformity and synchronization. The last two apply to interoperability between TPs.

Basic metrics were presented in the key performance area of accuracy for both input and output. We propose that evaluation of TP performance requires consideration of TP output metrics under specified input conditions. The approach requires the user to consider the impact of TP performance on higher-level systems. This allows the determination of the relevant basic metrics, events and specifiers.

Despite describing the foundation for development of TP metrics, much work remains to be accomplished. Basic metrics for other TP key performance areas need to be developed. Within each performance area, certain key performance indicators should be defined that can apply to a broad class of applications. While these measures may not allow direct numerical calculation of impact, they should provide a relative ranking of TP performance when comparing different TPs for similar applications.

References

[1] Federal Aviation Administration, 2004, Operational Evolution Plan, Washington, DC

[2] RTCA, 2000, National Airspace System Concept of Operations, Washington, DC

[3] Eurocontrol, 2003, Eurocontrol ATM Strategy for the Years 2000+, Brussels, Belgium

[4] ICAO, 2003, ATM Operational Concept Document, AN-Conf/11-WP/4

[5] Erzberger, Heinz, R.A.Paeilli, D.R. Isaacson, M.M. Eshowl, 1997, Conflict Detection in the Presence of Prediction Error, FAA/Eurocontrol ATM R&D Seminar, Saclay, FR

[6] Brudnicki, D.J., A.L. McFarland, 1997, User Request Evaluation Tool (URET) Conflict Probe Performance and Benefits Assessment, FAA/Eurocontrol ATM R&D Seminar, Saclay, FR

[7] McNally, B.D., R.E.Bach, W. Chan, 1998, Field Test Evaluation of the CTAS Conflict Prediction and Trial Planning Capability, AIAA-1998-4480, AIAA GN&C Conference, Boston, MA

[8] Swensen, H.N., et al, 1997, Design and Operational Evaluation of the Traffic Management Advisor at the Fort Worth Air Route Traffic Control Center, FAA/Eurocontrol ATM R&D Seminar, Saclay, FR

[9] Coppenbarger, R., R. Lanier, D. Sweet, S. Dorsky, 2004, Design and Development of the En Route Descent Advisor (EDA) for Conflict-Free Arrival Metering, AIAA-2004-4875, AIAA GN&C Conference, Providence, RI

[10] Wanke, C., S. Mulgund, D. Greenbaum, L. Song, 2004, Modeling Traffic Prediction Uncertainty for Traffic Management Decision Support, AIAA-2004-5230, AIAA GN&C Conference, Providence, RI

[11] Kirk, D. B., et al, 2004, Problem Analysis, Resolution and Ranking (PARR) Development and Assessment, 4th FAA/Eurocontrol ATM R&D Seminar, Santa Fe, NM

[12] Green, S.M., M.P. Grace, D.H. Williams, 2000, Flight Test Results: CTAS and FMS Cruise/Descent Trajectory Prediction Accuracy, 3rd FAA Eurocontrol ATM R&D Seminar, Naples, IT

[13] Wanke, C., 1997, Using Air-Ground Data Link to Improve Air Traffic Management Decision Support System Performance, FAA/Eurocontrol ATM R&D Seminar, Saclay, FR

[14] Swierstra,S., S. Green, 2003, Common Trajectory Prediction Capability for Decision Support Tools, 5th USA/Eurocontrol ATM R&D Seminar, Budapest, Hungary

[15] Mondoloni, S., M. Paglione, S. Green, 2002, Trajectory Modeling Accuracy for Air Traffic Management Decision Support Tools, ICAS 2002 Congress, Toronto, ON

[16] Ryan, H.F., M. Paglione, S.M. Green, 2004, Review of Trajectory Accuracy Methodology and Comparison of Error Measurement Metrics, AIAA-2004-4787, AIAA Conference, Providence, RI

[17] International Standardization Organization, 1998, Information technology – Quality of Service: Framework, ISO/IEC 13236:1998(E)

[18] Williams, Almira, Steve Bradford, Diana Liang, Wim Post, Ulrich Borkenhagen, Jean-Marc Pomeret, Common Metrics Framework for ATM Performance Assessment and Monitoring, 23rd DASC, Salt Lake City, October 2004

[19] Many Participants, 2004, FAA/Eurocontrol Action Plan 16 Technical Interchange Meeting on TP Requirements and Validation, Toulouse, FR

[20] IEEE, 1990, IEEE Standard Computer Dictionary

[21] US GAO, 1999, Managing for Results, Strengthening Regulatory Agencies' Performance Management Practices, GAO/GGD-00-10 Washington, DC

[22]http://www.worldbank.org/html/opr/pmi/envindi4.html

[23] Mondoloni, S., S. Swierstra, C.G. Avello, S.M. Green, M. Paglione, A Structured Approach for Validation and Verification of Aircraft Trajectory Predictors, 23rd DASC, Salt Lake City, October 2004

[24] Williams, D.H., S.M. Green, 1998, Flight Evaluation of a Center-TRACON Automation System Trajectory Prediction Process, NASA-TP-1998-208439

[25] Jones, E., D.R. Schleicher, 2001, NASA/FAA En Route Data Exchange (EDX) Phase 2 Field Evaluation Final Report, prepared for FAA ADL Product Team

[26] Arthur, W., M.P. McLaughlin, 1998, User Request Evaluation Tool (URET) Interfacility Conflict Probe Performance Assessment, 2nd USA/Europe Air Traffic Management R&D Seminar, Orlando, FL

[27] Gong, C., D. McNally, 2004, A Methodology for Automated Trajectory Prediction Analysis, AIAA-2004-4788, GN&C Conference Providence, RI

[28] Paglione, M., M. L. Cale, H. F. Ryan, 1999, Generic Metrics for the Estimation of the Prediction Accuracy of Aircraft to Aircraft Conflicts by a Strategic Conflict Probe Tool, Air Traffic Control Quarterly, Vol. 7 No. 3, pp. 147-165

[29] Shakarian, A., A. Haraldsdottir, 2001, Required Total System Performance and Results of a Short Term Conflict Alert Simulation Study, 4th USA/Europe ATM R&D Seminar, Santa Fe, NM

24th Digital Avionics Systems Conference
October 30, 2005