

Method for evaluating the compliance of a tracking system using a restricted number of measurements and associated devices

5 FIELD OF THE INVENTION

The present invention relates to a method for evaluating the compliance of a tracking system with a set of requirements. The present invention also relates to an evaluation module allowing to implement the method and a tracking system including such an evaluation module.

10

TECHNOLOGICAL BACKGROUND OF THE INVENTION

The proposed invention is in the field of air traffic control.

The control of airspace is ensured by tracking all aircraft. It involves supervising air traffic to prevent collisions between aircraft and to control traffic, as much when cruising in flight as near airports – takeoff and landing.

15

To implement such a control, air traffic controllers use an air traffic surveillance system based on aircraft tracking called a tracking system. The tracking system is able to, in real-time, estimate from data coming from sensors, the best possible estimate of the position, heading, and speed of each aircraft in a given flight information region.

20 The flight information region is often referred to by the acronym FIR, referring to the corresponding English term "Flight Information Region."

All estimated data form a track.

The quality of the track estimation (tracking) varies depending on the configuration of the tracking system, the quality of the data acquired by the sensors, or even environmental data such as terrain topology or weather.

25

As such, it is desirable for an air traffic controller to know the quality of the track estimates reconstructed by the tracking system.

For this, it is known to use the requirements of the ESASSP standard, which is detailed in the document entitled "*EUROCONTROL Specification for ATM Surveillance System Performance*" (Volume 1) the ISBN number of which is, 978-2-87497-022-1 and which was published in March 2012.

30

More specifically, a set of requirements defined by this ESASSP standard is evaluated by comparing the outputs of the tracking system with an estimate of the ideal trajectory of the aircraft.

However, calculating the requirements of the standard requires obtaining a large number of measurement points, typically between 50,000 and 100,000.

This poses a practical problem for obtaining an on-the-fly evaluation and notably detecting potential failures during operation.

5

SUMMARY OF THE INVENTION

There is therefore a need for a method allowing to evaluate the quality of tracking performed by a tracking system with a reduced number of measurements.

To this end, the description describes a method for evaluating the compliance of a tracking system with a set of requirements, at least one requirement being a mandatory requirement and at least one requirement being a recommended requirement, each requirement imposing that a physical quantity be greater than or equal to a threshold or that a physical quantity be less than or equal to a threshold, the method for evaluating being implemented by computer, the method for evaluating including at least the following steps:

15 - acquisition of a set of values, the set of values comprising the values accessible to the tracking system over a predefined time interval,

 - for each requirement, calculation of a robustness value, the robustness being calculated as the value p of a statistical test indicating whether the requirement considered is respected or not,

20 - for each requirement, comparison of each robustness value to a predefined threshold according to a respective comparison criterion, and

 - determination of the requirements to which the tracking system is compliant, the requirements determined as compliant being the requirements for which, at the comparison step, it was determined that the comparison criterion is respected.

25 According to particular embodiments, the method for evaluating presents one or more of the following features, taken alone or according to any technically possible combination:

 - the method further includes, for each robustness value not respecting the comparison criterion, a step of determining the cause of the non-compliance with the comparison criterion, the cause being either that the requirement is not respected or that the number of values is too low to guarantee a predefined reliability threshold for evaluating compliance with the requirement considered.

30 - when the determined cause is a number of values too low, the method includes a step of deducing a number of additional values to be acquired to obtain an evaluation reliability greater than the predefined reliability threshold.

35

- the method includes an iteration of the acquisition, calculation and comparison steps until the number of values allows to obtain an evaluation reliability greater than the predefined reliability threshold or reaches a maximum predefined number with an evaluation reliability lower than the predefined threshold.

5 - the method further includes a step of issuing an alert for requirements not respected or for requirements the number of values of which, from the acquisition, calculation and comparison steps has reached the maximum predefined number with an evaluation reliability lower than the predefined threshold.

- the maximum number is between 40,000 and 60,000.

10 - during the calculation step, an integral of a regularized incomplete beta function is calculated.

- during the calculation step, a p-value is equal to the result of the integral and another p-value is a linear function of the result of the integral.

- during the calculation step, a cumulative distribution function of the standard
15 normal distribution is calculated.

- the method further includes:

- a step of applying an evaluation function specific to each requirement to the p-values to obtain evaluation values, and

- a step of using the evaluation values to obtain a global score evaluating
20 the compliance of the tracking system with all the requirements.

- the number of requirements is greater than or equal to 20.

- the tracking system provides measurement estimates from previous data coming from multiple sensors, the set of values comprises the measurement estimates.

The description also describes a module for evaluating the compliance of a tracking
25 system with a set of requirements, at least one requirement being a mandatory requirement and at least one requirement being a recommended requirement, each requirement imposing that a physical quantity be greater than or equal to a threshold or that a physical quantity be less than or equal to a threshold, the evaluation module being able to:

- acquire a set of values, the set of values comprising the values accessible
30 to the tracking system over a predefined time interval,

- for each requirement, calculate a robustness value, the robustness being calculated as the p-value of a statistical test indicating whether the requirement considered is respected or not,

- for each requirement, compare each robustness value to a predefined
35 threshold according to a respective comparison criterion, and

- determine the requirements to which the tracking system is compliant, the requirements determined as compliant being the requirements for which, at the comparison step, it was determined that the comparison criterion is respected.

The description also proposes a tracking system comprising an evaluation module.

5 In the present description, the expression "able to" means interchangeably "adapted for," "adapted to," or "suitable for."

BRIEF DESCRIPTION OF THE DRAWINGS

10 Features and advantages of the invention will appear from reading the following description, given solely by way of non-limiting example, and made with reference to the appended drawings, in which:

- Figure 1 is a schematic representation of an example of a tracking system in interaction with a set of sensors,

15 - Figure 2 is a flowchart of an example of implementing a method for evaluating the compliance of the tracking system of Figure 1 with a set of requirements, and

- Figure 3 graphically illustrates the variation of the p-value as a function of the observed value for a given requirement as an example.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

GENERAL CASE

20 Figure 1 schematically illustrates a tracking system 10 in interaction with a set of sensors 12.

The tracking system 10 is able to observe an airspace to determine the trajectories of aircraft passing through an observed airspace.

25 This makes it possible to perform air traffic control in the considered airspace.

For this, the tracking system 10 is able to collect data from all the sensors and analyze the collected data.

The tracking system 10 emits as output the calculated data.

30 The calculated data are, for example, aircraft trajectories in the observed airspace, in other words, position, speed, or heading data for aircraft passing through the observed airspace.

The tracking system 10 forms with the set of sensors 12 an air traffic surveillance system.

Such a system is often referred to by the acronym SSTA.

35 The set of sensors 12 is configured to obtain data in the observed airspace.

According to the described example, the set of sensors 12 includes an ADS unit 14, a WAM unit 16, and a radar 18.

The ADS unit 14 is a cooperative surveillance system for air traffic control and other related applications. An aircraft equipped with an ADS unit 14 determines its position by a satellite positioning system (GPS) and periodically sends this position and other information to ground stations.

The abbreviation ADS refers to the corresponding English term "*Automatic Dependent Surveillance*."

Such an ADS unit 14 is sometimes also called a B-unit, the abbreviation ADS-B referring to the corresponding English term "*Automatic Dependent Surveillance-Broadcast*."

A WAM unit 16 uses data from multiple sensors to obtain the location of an aircraft.

The abbreviation WAM refers to the corresponding English term "*Wide Area Multilateration*" and designates an aircraft surveillance technology based on the principle of the time difference of arrival which is used at an airport.

For example, the WAM unit 16 collects data coming from multiple ground antennas to apply mathematical calculations allowing to obtain the position of the aircraft.

A radar 18, allows to detect the presence of aircraft in the sky and determine their position. The radar 18 emits electromagnetic pulses into the sky, and the detection and location of an aircraft are obtained by analyzing the wave reflected by the aircraft and retransmitted in the direction of the radar 18.

According to the example of Figure 1, the tracking system 10 includes an analysis module 20, an evaluation module 22.

The analysis module 20 is able to analyze data from the sensors to estimate new data.

The analysis module 20 thus regularly receives data relating to aircraft in the observed environment, notably their altitude, horizontal position, or speed.

In particular, the analysis module 20 is able to estimate the trajectory of an aircraft.

Such an analysis module 20 is often referred to by the English term "*tracker*."

The analysis module 20 is also able to estimate measurements from previous data from the set of sensors 12.

The evaluation module 22 is able to implement the steps of a method for evaluating the compliance of the tracking system 10 with a set of requirements.

According to the described example, the evaluation module 22 comprises a calculation sub-unit 24 and an alert sub-unit 26 the roles of which will appear in the following description.

An example of implementing the method for evaluating is now described with reference to Figure 2.

The method for evaluating aims to evaluate the compliance of the tracking system 10 with a set of requirements.

5 From among this set of requirements, at least one requirement is a mandatory requirement and at least one requirement is a recommended requirement.

By definition, a mandatory requirement is a requirement the non-compliance of which results in the disqualification of the tracking system 10. The term necessary requirement is sometimes used to designate this type of requirement.

10 By definition, a recommended requirement is a requirement the non-compliance of which does not result in the disqualification of the tracking system 10 but the compliance of which is desirable. The term desirable requirement is sometimes used to designate this type of requirement.

15 According to the proposed example, the number of requirements is greater than or equal to 20.

For the following, it is assumed that each requirement fixes a threshold to be respected. This means that each requirement can be formulated as a physical quantity greater than or equal to the threshold or a physical quantity less than or equal to the threshold.

20 According to the described example, the method for evaluating includes an acquisition step E100, a calculation step E102, a comparison step E104, a first determination step E106, a second determination step E108, a deduction step E110, an iteration step E112, an emission step E114, an application step E116, and a use step E118.

25 From among these steps, only the emission step E114 is implemented by the alert sub-module 26, the other steps being implemented by the calculation sub-module 24.

During the acquisition step E100, the evaluation module 22 acquires a set of values.

The set of values comprises the values accessible to the tracking system 10 over a predefined time interval.

30 According to the described example, the accessible values are the values coming from the set of sensors 12.

In addition, the values also include the data estimated by the tracking system 10, and notably the measurement estimates obtained from previous data by the analysis module.

35 The predefined time interval is chosen according to the imperatives related to the compliance evaluation.

For example, if regular evaluation is desired, the predefined time interval will correspond to the time interval chosen between two evaluations.

In particular, a predefined time interval between 10 minutes and 1 hour can be chosen. The predefined time interval chosen will depend in practice on the requirement considered.

As detailed later, the number of values will usually be chosen to allow near-real-time evaluation and is generally less, even much less, than the number of measurement points usually required by the evaluation standards for compliance with requirements.

During the calculation step E102, the evaluation module 22 calculates a robustness value for each requirement.

For this, the robustness is calculated as the p-value of a statistical test indicating whether the requirement considered is respected or not.

A statistical test is a mathematical tool allowing to verify whether the data allows a hypothesis to be rejected or not.

This hypothesis that we seek to reject is called the "null hypothesis."

The implementation of the statistical test gives notably a value p called "p-value," located between 0 and 1, which will serve to reject or not the hypothesis.

The "value p " is also referred to as "p-value" in reference to the corresponding English term "p-value."

In the following, the term "p-value" is used.

The p-value is defined as the probability, assuming that the null hypothesis is true, of observing the data, or even more "extreme" data.

As will be explained later, each p-value is associated with a metric depending on the requirement.

This metric generally includes moments of the assumed probability distributions on the acquired data.

In such a case, it may be favorable to calculate each moment first and then the p-values to avoid potential redundant calculations.

During the comparison step E104, the evaluation module 22 compares each robustness value to a predefined threshold according to a respective comparison criterion.

In other words, the evaluation module 22 compares the p-value to a predefined threshold.

Typically, the term "p-value" is compared to a threshold chosen at 5%.

This value of 5% is given as an example, the threshold depending on the needs of the user of the tracking system 10.

During the determination step, the evaluation module 22 determines the requirements to which the tracking system 10 is compliant.

For this, the evaluation module 22 considers that the requirements determined as compliant are those for which at the comparison step E104, the evaluation module 22
5 determined that the comparison criterion is respected.

If the p-value is less than the threshold, it is considered very unlikely to have observed such data, knowing that the null hypothesis is true. This leads to rejecting the hypothesis with great certainty.

To better explain this idea, the simple case of a coin toss is developed.

10 Let a coin, with one side heads, and one side tails. The following hypothesis is made: "the coin is balanced and can therefore land on heads or tails with a probability of $\frac{1}{2}$."

According to a first example, the coin is tossed 10 times, and 7 heads and 3 tails are obtained. These are the acquired data.

The question is then to know whether the null hypothesis can be rejected or not.

15 The probability of this result is then calculated, as well as more extreme results (namely, the probability of having had a number of heads between 7 and 10 inclusive). A probability of 0.17 is obtained, in other words, a p-value of 0.17.

This means that the calculation step E102 leads to a 17% chance of getting 7 heads or more if the coin was balanced.

20 This p-value is greater than 5%, so it is not possible to reject the hypothesis with sufficient certainty.

According to a second example, the coin is tossed 10 times and 9 heads and 1 tail are obtained. These are the acquired data.

25 The probability of getting 9 or 10 heads, knowing that the coin is balanced, is then calculated. This leads to a p-value of 0.01.

With the threshold of 0.05, this leads to determine that the hypothesis is to be rejected. In other words, it can be considered that, knowing the coin is balanced, it is very unlikely to observe so many heads.

30 A third example can also illustrate the fact that the p-value also evolves depending on the number of tosses.

Suppose 1000 tosses are made and 700 heads are obtained.

In such a case, the p-value is less than 10^{-5} , which corresponds to the fact that the probability of getting 7 heads or more on 10 tosses (first example) is much higher than that of getting 700 heads and more on 1000 tosses, knowing a balanced coin.

A calculation allows to show that obtaining 527 heads on 1000 tosses is enough to conclude with a threshold of 5% that the null hypothesis is to be rejected, in other words, that the coin is biased.

More elaborate examples but based on what has just been described are given in the section "Application to the specific case of the ESASSP standard." In this section, specific cases related to the evaluation criteria of the tracking system are developed.

In this section, it will be explained how to derive from a given requirement the expression for calculating the p-value.

More precisely, a given requirement generally corresponds to the calculation of a metric and the compliance of this metric with a threshold value.

By analyzing the evolution of the p-value as a function of the requirement value, it is possible to derive a suitable expression for this p-value.

Such a construction is done before the implementation of the method by a so-called a priori analysis.

In this analysis, as explained later, it is also possible to derive the number of values to be obtained to determine each requirement to ensure a high level of confidence in the requirements, while applying to reduced geographical areas and reduced time intervals.

Such a number of values corresponds to a predefined reliability threshold that will be used during the second determination step E108.

During the second determination step E108, the evaluation module 22 determines the cause of the non-respect of the comparison criterion for each robustness value not respecting the comparison criterion.

According to the described example, there are two possible causes.

The first possible cause corresponds to the fact that the requirement is not respected.

Indeed, it may be enough to know that the requirement will not be respected even if the number of measurement points would be higher.

The second cause corresponds to the fact that the number of acquired values is too low to guarantee a predefined reliability threshold for evaluating compliance with the requirement considered.

This may be due to lower than expected aircraft traffic in the considered geographical area (off peak hours or strike, for example).

It is assumed for the following that, for at least one requirement, the determined cause is the second cause.

During the deduction step E110, the evaluation module 22 calculates the number of additional values to be acquired to obtain an evaluation reliability greater than the predefined reliability threshold.

For this, the evaluation module 22 uses the number of points ensuring respect of the reliability threshold, this number of points having been calculated before the implementation of the method.

The evaluation module 22 then reads the value in a memory of the tracking system 10 and deduces, for example, by subtraction, the number of additional values to acquire.

Such a calculation can be performed more elaborately by calculating the number of points necessary to pass the requirements, while dynamically taking into account the values observed so far. This corresponds to a real time recalculation.

This is now illustrated for examples developed in the section "application to the specific case of the ESASSP standard."

Thus, for the case of requirements of the type probability of update, it is assumed, that for the moment, observed values k^* from among acquired values n^* , it is possible to calculate a p-value p^* on the test that aims to verify if the minimum required value p_r is exceeded in a statistically significant way.

It is then possible to calculate the minimum number of additional values to acquire for the requirement to be passed with a given threshold. This corresponds to assuming that all the next values are good, and to give the minimum value m such that success $m + k^*$ from among $m + n^*$ reports allows to pass the requirement with a sufficiently low p-value.

This number corresponds to the minimum number of new points (new values) necessary to reject the null hypothesis.

Alternatively, or in addition, the evaluation module 22 modifies the spatial-temporal discretization of the zones on which the calculation of the p-value is performed.

This spatial-temporal discretization is generally provided with the calculation of the p-value associated with the requirement considered but can be slightly modified without altering the quality of the requirement evaluation.

During the iteration step E112, at least the acquisition, calculation, and comparison steps are reiterated until a condition is verified.

This is indicated by arrow 120 on Figure 2.

Other steps such as the deduction step E110 can also be iterated.

In the proposed example, the condition is to respect either a first condition or a second condition.

The first condition is to obtain an evaluation reliability greater than the predefined reliability threshold.

This is obtained by exceeding the minimum number of additional values to acquire if it has been determined only once or when the minimum number of additional values to acquire is determined as zero at an iteration of the deduction step E110.

The second condition is to reach a maximum predefined number with an evaluation reliability lower than the predefined threshold.

This corresponds to the case where the p-value is too high for the requirement to be satisfied with sufficient certainty, but where the values observed so far seem to corroborate the fact that the requirement is satisfied (formally in the update example described previously: $\frac{k^*}{n^*} > p_r$, but n^* is too low for us to have enough certainty).

It is then possible to calculate the number of points necessary to pass the requirement with reasonable certainty, assuming that the success ratio will remain the same. Formally, by noting $p^* = \frac{k^*}{n^*}$, the number of reports m (acquired values) are sought, such that observing successes p^*m from among reports m allows to obtain a p-value lower than the required threshold.

In practice, the maximum predefined number is often between 40,000 and 60,000.

During the emission step E114, the evaluation module 22 emits an alert for requirements not respected or for requirements the number of values of which from the acquisition, calculation and comparison steps have reached the maximum predefined number with an evaluation reliability lower than the predefined threshold.

These cases correspond to the fact that neither the first condition nor the second condition of the iteration step E112 are respected.

The alert is, for example, a sound alert or a visual alert issued to an air traffic controller.

During the application step E116, the system applies an evaluation function specific to each requirement to the p-values to obtain evaluation values.

Each evaluation function specific to each requirement is a utility function.

As explained above, the p-value noted ρ is interpreted as meaning that the requirement is satisfied if the hypothesis H_0 is rejected and therefore if the p-value ρ is lower than a threshold ε .

It is possible to convert the p-value into a utility function giving a value between 0 and 1 depending on the value of the p-value.

For example, the utility function is a piecewise linear function worth 1 as soon as $\rho \leq \varepsilon$ (case corresponding to compliance with the requirement) then decreases linearly to 0 for p-values ρ greater than the threshold.

The linear function is, for example, the following:

$$1 - \rho$$

Alternatively, it may be considered to normalize the decreasing utility function by including a supervised learning phase allowing to learn this function.

Advantageously, a decreasing normalization function with thresholds is used.

For example, the evaluation module 22 may impose that if the p-value is greater than a value to be defined by an expert, then the utility is zero.

The value defined by the expert is determined before the implementation of the method, so that it is accessible to the evaluation module 22 by reading its value in a memory of the tracking system 10.

Then, the evaluation module 22 uses the evaluation values to obtain a global score during the use step E118.

The global score evaluates the compliance of the tracking system 10 with all the requirements.

Such a score is often referred to as QoS in reference to the English term "Quality of Service."

For this, the evaluation module 22 can implement different techniques.

For example, with the previous utility function, the utility is worth 1 if the p-value is lower than the threshold and with very high confidence that the requirement is passed and otherwise has a value strictly less than 1.

In doing so, the evaluation module 22 can use an aggregation function.

According to one example, the aggregation function is a strictly increasing compensatory function.

By applying only this technique to the case of mandatory requirements, this allows to obtain a QoS score equal to 1 if all mandatory requirements are passed with high confidence and strictly less than 1 if at least one mandatory requirement is not completely respected.

According to another example, the use step E118 comprises supervised learning of the compliance evaluation functions for a requirement in the form of a piecewise linear function, the linear function including at least three pieces, preferably less than six pieces, to obtain the learned evaluation functions, the application of the learned evaluation functions to the tracking system to obtain evaluation values, and the use of the evaluation values to

obtain a global score evaluating the compliance of the tracking system 10 with all the requirements.

The evaluation module 22 then calculates the global score for each type of requirement by aggregating the evaluation functions of all the requirements of the considered type, the global score evaluating the compliance of the tracking system with all the requirements being obtained by aggregating the global scores by type of requirement.

Alternatively, instead of this calculation, the use step E118 comprises supervised learning of aggregation functions of the evaluation values, the aggregation function evaluating the compliance of the tracking system with a part of the requirements of the set of requirements corresponding to a particular quantity, and a use step comprising the use of the evaluation values and the learned aggregation functions to obtain a value for each part of the requirements of the set of requirements corresponding to a particular quantity and the use of the obtained values during the use as well as the aggregation functions to obtain the global score.

These techniques are also applicable to the case of recommended requirements.

According to another approach, the evaluation module 22 aggregates differently the values QoS^O (value for mandatory requirements) and QoS^R (value for recommended requirements) to obtain the global QoS. The normalized scores here are interpreted as the level of confidence according to which the requirements are passed.

Given the mandatory nature of the underlying requirements to QoS^O , it is necessary that the slightest failure of a mandatory requirement is reflected in the global score and cannot be compensated by the recommended requirements.

For this, in this approach, the evaluation module 22 uses the following formulas:

$$QoS = \begin{cases} \frac{QoS^O}{2} & \text{si } QoS^O < 1 \\ \frac{1 + QoS^R}{2} & \text{si } QoS^O = 1 \end{cases}$$

In the previous techniques, the aggregation functions are compensatory, which implies that poor respect of one requirement can be compensated by good respect of another requirement.

In some cases, it may be desired that this compensation is not possible, notably in what concerns mandatory requirements. As soon as at least one mandatory requirement is not respected, it would then no longer be possible to compensate for this non-respect even if all the other requirements are respected.

For this, another type of aggregation function called t-norms can be used.

These are binary operators $T(x, y) \in [0, 1]$ where x, y are also valued in $[0, 1]$.

A binary operator $T: [0, 1] \times [0, 1] \rightarrow [0, 1]$ is a t-norm if it satisfies the following properties:

- (neutral element) $T(1, x) = x$ for all $x \in [0, 1]$,
- 5 • (commutativity) $T(x, y) = T(y, x)$ for all $x, y \in [0, 1]$,
- (monotonicity) $T(x, y) \leq T(u, v)$ for $0 \leq x \leq u \leq 1$ and $0 \leq y \leq v \leq 1$, and
- (associativity) $T(x, T(y, z)) = T(T(x, y), z)$ for all $x, y, z \in [0, 1]$.

The function T satisfies certain properties. This function provides a pessimistic evaluation relative to a compensatory function because $T(x, y) \leq \min(x, y)$, in other words, it is never possible to have a better score than the worst score. Several bad scores worsen the overall score.

It can be shown that the binary operator $T: [0, 1] \times [0, 1] \rightarrow [0, 1]$ is a strict and Archimedean t-norm, if and only if, there exists a function $f: [0, 1] \rightarrow [0, \infty)$ continuous, strictly decreasing function with $f(1) = 0$ such that $T(x, y) = f^{-1}(\min(f(x) + f(y), f(0)))$.

By definition, a t-norm is Archimedean if the following two conditions are satisfied:

- T is continuous, and
- $T(x, x) < x$ for all $x \in (0, 1)$.

This binary operator T is easily transformed into an n-ary operator:

$$T(x_1, \dots, x_m) = f^{-1} \left(\min \left(\sum_{i=1}^m f(x_i), f(0) \right) \right)$$

Where:

- $x_1, \dots, x_m$ are the normalized scores of the p-values of the mandatory requirements,
- $T(x_1, \dots, x_m)$ is the score QoS^O , and
- 25 • the function f is to be constructed by supervised learning.

Alternatively, the sum $\sum_{i=1}^m f(x_i)$ is replaced by an ordered weighted average, which is mathematically written as:

$$T(x_1, \dots, x_m) = f^{-1} \left(\min \left(\sum_{i=1}^m w_i f(x_{\pi(i)}), f(0) \right) \right)$$

Where π is a permutation on $\{1, \dots, m\}$ ordering the scores as follows:

$$x_{\pi(1)} \geq x_{\pi(2)} \geq \dots \geq x_{\pi(m)}$$

The method just described therefore allows to obtain a robust evaluation of the compliance of the tracking system 10 with a set of requirements.

In fact, rather than having a simple binary value (criterion satisfied or not), the method provides access to an estimate of the margin with which the criterion is satisfied, in other words, an estimate of the confidence that can be had in the estimate.

Furthermore, in the presence of too few values to have a reliable estimate, the criterion will always be considered as not satisfied, which is desirable in a critical context such as air traffic, where it is better to be alarmist wrongly than too confident.

Furthermore, the method gives the certainty that the result cannot instantly vary from "not satisfied " to " satisfied," the p-value evolving smoothly with new data. However, there can be a crossing above and below the threshold, in which case the alarm is triggered, but there is always a certain margin to cover before actually reaching the criterion violation.

Furthermore, the method uses a smaller number of points than those recommended in the standard.

This allows the amount of occupied memory to be reduced, as well as the stress on the hardware components of the computer system 10.

The method also presents the advantage that the evaluation is adaptive.

In particular, the method greatly reduces the data acquisition duration for a reliable score calculation, and therefore the compliance or non-compliance of the criteria. Thus, the data used are less old, and the quality of service estimate is therefore more responsive.

This is beneficial when a problem arises since in case of rapid degradation, the metric degrades faster, and thus the operator is informed more quickly.

The reverse is also true, when a problem is solved, the return to normal is faster.

This means that the method allows to take into account extreme drifts.

When the drift is minor, with the number of data fixed a priori, the method is not able to validate the requirement with high confidence. The number of points can then be temporarily increased to absorb this minor transient drift. This avoids raising an alert that is not necessary. To increase the number of measurement points, either the data collection duration is increased (if allowed), or geographical areas are grouped. These two measures are temporary.

When the drift is major, from the beginning of data collection over a time interval, the method allows to take into account that there is no chance of respecting the requirement given the excessively high number of bad data. Instead of waiting for the end of the information collection interval, the method raises an alert as soon as possible to the operator.

The method also defines a mechanism for controlling the number of measurement points, this control mechanism can be static (determined a priori) or/and dynamic (adapted from the measured values).

Thus, it is possible to implement the method in real-time over a time interval that can be of a fixed length configured in advance or dynamic depending on the value acquisition rate. This allows to detect the presence of an anomaly in the tracking system in real-time.

In each of the described examples, each module or sub-module is realized each in the form of software, or a software brick.

All the modules or sub-modules are then realized, in other words, in the form of a computer program, also called a computer program product, it is also able to be recorded on a medium, not represented, readable by a computer. The computer-readable medium is, for example, a medium able to store electronic instructions and to be coupled to a bus of a computer system. For example, the readable medium is an optical disk, a magneto-optical disk, a ROM memory, a RAM memory, any type of non-volatile memory (for example FLASH or NVRAM), or a magnetic card. On the readable medium is then stored a computer program comprising software instructions, stored in a memory executable by a processor.

In one alternative, not represented, each module or sub-module is realized in the form of a programmable logic component, such as an FPGA (*Field Programmable Gate Array*), or an integrated circuit, such as an ASIC (*Application Specific Integrated Circuit*).

The invention relates to any technically possible combination of the previously described embodiments.

APPLICATION TO THE SPECIFIC CASE OF THE ESASSP STANDARD

Presentation of the criteria

The method just described can be applied to any type of standard but will be illustrated in the following through the example of the ESASSP standard, which is detailed in the document entitled "EUROCONTROL Specification for ATM Surveillance System Performance" (Volume 1) with ISBN 978-2-87497-022-1 and published in March 2012.

The objective of this standard is to ensure the required quality to avoid collisions. This quality aims at better separation between aircraft in (civil) air traffic. In order to be able to increase air capacity, it is essential to have tracking tools that allow to ensure increasingly fine separation between aircraft. Two standards have been defined to ensure two distance separations: 5 NM and 3 NM.

The ESASSP standard defines 22 constraints named R1 to R22, in other words, 22 specific metrics. The 22 constraints are briefly explained in the following.

The first constraint R1 concerns the measurement interval for the probability of update (referred to in English as "Measurement Interval for Probability of Update"). The first
5 constraint R1 is divided into two requirements: a mandatory requirement and a recommended requirement.

The second constraint R2 concerns the probability of update of the horizontal position (referred to in English as "Probability of update of horizontal position"). The second
10 constraint R2 is divided into two requirements: a mandatory requirement and a recommended requirement.

The third constraint R3 concerns the ratio of time when the aircraft was not tracked (referred to in English as "ratio of missed 3D position involved in long gaps"). The third constraint R3 corresponds to a mandatory requirement.

The fourth constraint R4 concerns a Mean Squared error in the horizontal position, RMS error (referred to in English as "Horizontal position RMS error"). The RMS error is
15 often considered as a main error representative of measurement accuracy. The fourth constraint R4 is divided into two requirements: a mandatory requirement and a recommended requirement.

The fifth constraint R5 concerns a ratio of consecutive correlated errors (referred to
20 in English as "Consecutive correlated error ratio"). The fifth constraint R5 corresponds to a recommended requirement.

The sixth constraint R6 concerns the maximum time interval with close proximity (referred to in English as "Max delta time in close proximity"). Such a constraint corresponds to the fact that, on a radar, very close tracks that are not refreshed at the same time are
25 present. In the presentation to the air traffic controller, there should not be two tracks side by side with distant update dates. This could lead to collision risks, in other words, serious safety issues. The sixth constraint R6 corresponds to a recommended requirement.

The seventh constraint R7 concerns the probability of update of pressure altitude (referred to in English as "Probability of update of pressure altitude"). The seventh constraint
30 R7 corresponds to a mandatory requirement.

The eighth constraint R8 relates to the average age of forwarded pressure altitude data (referred to in English as "Average data age of forwarded pressure altitude"). The eighth constraint R8 corresponds to a mandatory requirement.

The ninth constraint R9 relates to the maximum age of forwarded pressure altitude data (referred to in English as "Max data age of forwarded pressure altitude"). The ninth constraint R9 corresponds to a mandatory requirement.

5 The tenth constraint R10 relates to the ratio of incorrect forwarded pressure altitude data (referred to in English as "Ratio of incorrect forwarded pressure altitude"). It should be noted that the tenth constraint R10 is an estimate of the error corresponding to a parasitic error. The tenth constraint R10 corresponds to a mandatory requirement.

10 The eleventh constraint R11 concerns unsigned pressure altitude errors (referred to in English as "Pressure altitude unsigned error"). It should be noted that the eleventh constraint R11 is an estimate of the main type error. The eleventh constraint R11 corresponds to a mandatory requirement.

The twelfth constraint R12 concerns the delay of appearance of emergency indicators (referred to in English as "Delay of apparition of the emergency indicator / SPI report"). The twelfth constraint R12 corresponds to a mandatory requirement.

15 The thirteenth constraint R13 relates to the delay of change in aircraft identification (referred to in English as "Delay of change in Aircraft Id"). The thirteenth constraint R13 corresponds to a mandatory requirement.

The fourteenth constraint R14 concerns the probability of update with correct identification (referred to in English as "Probability of update of aircraft identity with correct value"). The fourteenth constraint R14 is divided into two requirements: a mandatory requirement and a recommended requirement.

25 The fifteenth constraint R15 relates to the ratio of incorrect aircraft identification (referred to in English as "Ratio of Incorrect Aircraft identity"). It should be noted that the fifteenth constraint R15 is an estimate of the parasitic type error. The fifteenth constraint R15 corresponds to a mandatory requirement.

The sixteenth constraint R16 concerns the RMS error of the rate of climb or descent (referred to in English as "Rate of climb/descent RMS error"). In this context, the estimated error is of the main type. The sixteenth constraint R16 corresponds to a recommended requirement.

30 The seventeenth constraint R17 concerns the RMS error of track velocity (referred to in English as "Track velocity RMS error"). In this context, the estimated error is of the main type. The seventeenth constraint R17 corresponds to a recommended requirement.

The eighteenth constraint R18 concerns the RMS error of the track velocity angle (referred to in English as "Track velocity angle RMS error"). In this context, the estimated

error is of the main type. The eighteenth constraint R18 corresponds to a recommended requirement.

The nineteenth constraint R19 concerns the density of uncorrelated false target reports (referred to in English as "Density of uncorrelated false target reports"). In this context, the estimated error is of the parasitic type. The nineteenth constraint R19 corresponds to a recommended requirement.

The twentieth constraint R20 concerns the hourly rate of false tracks close to true tracks (referred to in English as "Hourly rate of false tracks close to true tracks"). In this context, the estimated error is of the correlated type. The twentieth constraint R20 corresponds to a recommended requirement.

The twenty-first constraint R21 concerns continuity (referred to in English as "Continuity"). The twenty-first constraint R21 corresponds to a recommended requirement.

The twenty-second constraint R22 concerns manual investigations (manual analyses of results) that must be conducted when constraints R2, R4, R12, or R13 are not respected. The twenty-second constraint R22 is divided according to two requirements: a mandatory requirement and a recommended requirement.

To summarize, the following Table 1 can be established, with the presence of a cross indicating the presence of an associated requirement:

Constraints	English name used in the standard	Unit	Mandatory requirement	Recommended requirement
R1	Measurement Interval for Probability of Update	s	X	X
R2	Probability of update of horizontal position	%	X	X
R3	Ratio of missed 3D position involved in long gaps	%	X	
R4	Horizontal position RMS error	m	X	X
R5	Consecutive correlated error ratio	%		X

R6	Max delta time in close proximity	s		X
R7	Probability of update of pressure altitude with correct value	%	X	
R8	Average data age of forwarded pressure altitude	s	X	
R9	Max data age of forwarded pressure altitude	s	X	
R10	Ratio of incorrect forwarded pressure altitude	%	X	
R11	Pressure altitude unsigned error	%	X	
R12	Delay of apparition of the emergency indicator / SPI report	s	X	
R13	Delay of change in Aircraft Id	s	X	
R14	Probability of update of aircraft identity with correct value	%	X	X
R15	Ratio of Incorrect Aircraft Identity	%	X	
R16	Rate of climb/descent RMS error	m/s		X

R17	Track velocity RMS error	m/s		X
R18	Track velocity angle RMS error	degrees		X
R19	Density of uncorrelated false target reports	number		X
R20	Number per hour of false tracks close to true tracks	number		X
R21	Continuity	%		X
R22	Investigations		X	

Table 1: Association for each constraint of its English name, the associated unit, as well as the presence of a mandatory requirement and a recommended requirement

- 5 In summary, the ESASSP standard is a set of 14 mandatory requirements and 12 recommended requirements, which represent 26 requirements in total, these 26 requirements being estimated according to 22 distinct metrics.

Calculation of p-values associated with the criteria

- 10 For the calculation of p-values, the criteria defined in the ESASSP can be grouped into several families. These families are five in number and are exposed in the following.

The first family, groups criteria related to update probabilities, these criteria being criteria R2, R7, and R14.

- 15 These criteria R2, R7, and R14 have in common to focus on a proportion between the number of received reports containing a certain element (horizontal position, pressure altitude, and aircraft identity respectively) from among all received reports (which may or may not contain the element in question). These criteria impose that the value of the proportion be greater than a threshold value.

The second family, groups criteria related to the root mean square error, these criteria being criteria R4, R6, R16, R17, and R18.

Indeed, the criteria of the second family focus on a calculation of the average Euclidean distance between a value and a reference measurement, often calculated a posteriori by the SSTA. These criteria impose that the value of the proportion be less than a threshold value.

The third family, groups criteria related to ratios, these criteria being criteria R3, R5, R10, R11, and R15 and require that a ratio between two quantities presents a value less than a predefined value.

The fourth family, groups criteria related to averages, these criteria being criteria R8, R12, and R13.

The fifth family, groups criteria not falling under the four families already presented. These are criteria R19, R20, and R9.

It can be noted that these families do not group all criteria. In fact, the criterion R1 is a parameter of the tracking system 10 which is therefore fixed a priori, the criterion R9 is not a criterion, but a filter on the validity of the data, the criterion implies not having an incident over periods of several years (which cannot be evaluated over a short period) and criterion R22 is not a quantitative criterion.

It is now described how to calculate the probabilities of obtaining a new value not respecting the requirement given the acquired values, in other words, the p-values for each of the aforementioned families.

First family

For the first family, it is first explained how to determine the p-values for the criterion R2 involving a simple update probability.

The criterion R2 involves a metric calculated as the ratio between a number of observed trajectory portions (noted n) and the number of these trajectories containing a horizontal position (noted k). The probability is therefore:

$$P = k/n$$

Where:

- n is the number of measurement points (for example the number of segments of identical size in all trajectories, for the criterion R2) and
- k is the number of measurement points having a certain property (for example the number of trajectory segments in which the tracking system 10 has

received at least one update of the horizontal position of the aircraft, for the criterion R2).

It is assumed that the value of k is a realization of a random variable N_R following a binomial probability distribution p , with n trials. This assumes the independence of successive observations. The rule requires that this probability be above a certain threshold (noted p_r).

According to a simple embodiment, the null hypothesis can be chosen as:

$$H_0 : p \leq p_r$$

The calculation of the p-value noted ρ can then be implemented using the formula:

$$\rho = P(N_R \geq k | H_0)$$

Where N_R is a random variable according to a binomial probability p , with n trials.

This corresponds to the following formulation: "knowing that the actual probability does not satisfy the criterion, what is the probability that the tracking system observes an equal or greater number of reports containing a horizontal position?" If this value is sufficiently low, it is considered that the observation of k is probably not due to chance, and that the threshold is therefore passed with high probability.

According to a more elaborate embodiment and notably in the context of a critical system, it is preferable to impose larger margins.

For this, the null hypothesis becomes $H_0 : p = p_r$ and the calculation of the p-value as for the simple embodiment.

This corresponds to the following formulation: "knowing that the actual probability just satisfies the criterion, what is the probability that an equal or greater number of reports containing a horizontal position is observed?"

Thus, it is more difficult for the system to satisfy the criterion, and therefore greater robustness of the system is ensured.

For this criterion, and under these assumptions, the p-value is calculated:

$$\rho = 1 - I_{1-p_r}(n - k + 1, k)$$

where I is the regularized incomplete beta function.

The curve of the p-value thus obtained is represented in Figure 3 for the requirement R2. The x-axis represents the ratio $\hat{p} = \frac{k}{n}$, the y-axis represents the p-value. As a reminder, the requirement R2 requires that this ratio be greater than or equal to $p_r = 0.97$, materialized by the vertical dotted line. The horizontal dotted line corresponds to the commonly used 0.05 threshold on p-values. In other words, if the p-value is below this bar, then there is very high confidence that the requirement R2 is passed.

Each curve corresponds to a number of points. Each curve takes the form of a decreasing sigmoid the inflection point of which is reached at $p_r = 0.97$. It is noted that the curve is all the steeper around this inflection point as the number of points is high.

This means that with many points, there is very high confidence that the requirement R2 is passed as soon as one is barely above the threshold $p_r = 0.97$. However, for a small number of points, it must be significantly above the threshold to be confident that the requirement is passed.

Such a curve can be realized for each of the requirements and leads to the results that will be detailed in the following. For simplicity, these curves are not introduced in the following to simplify the discussion, but the results obtained use these curves that have been realized by the Applicant.

In some cases, it is required that the calculated value for a criterion be reached for all tracked trajectories. This is notably the case for the "recommended" criterion of the criterion R2.

In this case, the values are used, which are calculated trajectory by trajectory. This means that, for a given trajectory, the values are calculated only on the points belonging to this trajectory.

Let ρ_t be the p-value associated with a given trajectory t and T the set of trajectories.

In the case of the criterion R2, it is a matter of verifying that, for all $t \in T$, the underlying update probability p_t verifies $p_t > 0.97$.

It is thus calculated for a trajectory t the p-value noted ρ_t of the test defined in the previous paragraph, for which only the information of the reports concerning the trajectory t . is used.

It thus comes:

$$\rho = P(\exists t \in T, N_R^t \geq \hat{k}^t | H_0)$$

This can be rewritten as:

$$\rho = 1 - P(\forall t \in T, N_R^t < \hat{k}^t | H_0) = 1 - \prod_{t \in T} P(N_R^t < \hat{k}^t | H_0)$$

Assuming that the probabilities are independent according to the different trajectories, it becomes:

$$\rho = 1 - \prod_{t \in T} (1 - P(N_R^t \geq \hat{k}^t | H_0))$$

Thus, the p-value of the test aiming to verify that all trajectories $p_t > 0.97$ verify is calculated as:

$$\rho = 1 - \prod_{t \in T} (1 - \rho_t)$$

It should be noted that this approach is applicable to all subsequent p-values for which it is desired to verify that all trajectories satisfy a given criterion.

It is then explained how to determine the p-values for the criteria R7 and R14 involving a double update probability.

The metrics involving a double update probability are calculated in the same way as for the metrics involving a simple update probability with one difference. More precisely, the probability is written:

$$P = z/k$$

Where:

- k is the number of reference realizations (for example, for the criterion R7, the number of time intervals where an altitude update is received), and
- z is the number of realizations having a certain property (for example, for the criterion R7, the number of time intervals where a non-erroneous altitude update is received)

In this case, the numbers k and z are considered as coming from binomial distributions. This assumes the independence of successive observations, and the fact that these observations are identically distributed.

According to one particular embodiment, the number k always corresponds to the number of reports containing a pressure altitude. As for the number z , it is, from among these k reports, those the pressure altitude value of which is correct (in the case of the criterion R7) or an aircraft identifier (the criterion R14).

Thus, the number k follows the binomial distribution N_R of the previous paragraph, with an unknown probability p_R , and the number z follows a binomial distribution of the parameters k and $\hat{p}$.

In this context, the value $\hat{p}$ is unknown and it is a matter of verifying that its value is greater than a given threshold p_r .

Always taking the null hypothesis $H_0 : \hat{p} = p_r$, the p-value is then calculated as follows:

$$\rho = 1 - \int_{l=0}^1 I_{1-lp_r}(n-z+1, z) \frac{\binom{n}{k} l^k (1-l)^{n-k}}{\int_0^1 \binom{n}{k} t^k (1-t)^{n-k} dt} dl$$

In practice, the integral can be calculated by numerical methods such as a Monte Carlo draw or a rectangular approximation method.

Second family

For the second family, it is a matter of calculating the p-value for root mean square errors.

A criterion of the second family is therefore a metric that is a root mean square error between an observed value and a reference value. The criterion is satisfied when this metric is less than a given value.

Let EQMR (RMSE) be the root mean square error. This quadratic error is written:

$$EQMR = \sqrt{\frac{1}{n} \sum_{t=1}^n E_t}$$

Where:

- E_t is the root mean square error on a given quantity for the t -th measurement. For example, for the criterion R4, the quantity is the error of the horizontal position of the aircraft.

Assuming that all root mean square errors E_t are identically distributed and independent, and that the corresponding distribution has a mean μ^2 and variance σ^2 (unknown), the criterion requires that the mean μ^2 be less than a certain value μ_r^2 , which is the requirement threshold.

Let $\hat{\mu}^2$ be the observed mean of the root mean square errors .

Making the null hypothesis: $H_0 : \mu^2 = \mu_r^2$ the central limit theorem allows to obtain the following relation:

$$\rho = P(EQMR < \mu_r \mid H_0) = \Phi \left(\frac{\hat{\mu}^2 - \mu_r^2}{\frac{\hat{\sigma}}{\sqrt{n}}} \right)$$

Where:

- Φ is the cumulative distribution function of the standard normal distribution.

For some criteria, the metrics are calculated as the minimum over different trajectories of the root mean square error calculated on each trajectory. This is for example the case of the criterion R4.

The approach is then the same as previously.

Third family

For the third family, as for the first family, there are two types of criteria in this family, a first type called simple ratio gathering criteria based on a metric involving a variable

divided by a constant and a second type called double ratio gathering criteria based on a metric involving a variable divided by another variable. For both types, the metric must be less than a predefined threshold.

It is first explained how to calculate the p-value for criteria based on a simple ratio, as is the case for the criteria R3, R5, and R11.

This calculation is quite similar to the case of the simple update probability since the ratios are of the form:

$$R = \frac{k}{n}$$

With:

- n is a given number (for example, the number of track updates received for the criterion R5), and
- k is a random variable following a binomial distribution of the parameters n and ρ (for example, concerning the criterion R, the number of track updates for which at least 3 consecutive updates of horizontal positions present an error on the same side of the trajectory).

It is a matter of testing if p is significantly less than p_r , the required value, given the observations.

Thus, taking as the null hypothesis, $H_0 : p = p_r$ and, for this metric, and under these assumptions, the p-value is calculated according to the following formula:

$$\rho = I_{1-p_r}(n - k + 1, k)$$

It can be noted here the symmetry with the simple update probability. This is explained by the fact that, in one case, it is a matter of maximizing the simple update probability while here, it is a matter of minimizing the simple ratios.

Concerning double ratios, the following approach can be followed. This approach is notably adapted for the criteria R10 and R15.

In the case of double ratios, the metric is written in the following form:

$$R = \frac{N_I}{N_V}$$

Where:

- N_I is the number of reference realizations (for example, the number of track updates containing a barometric altitude, for the criterion R10), and
- N_V is the number of realizations having a certain property (for example, the number of track updates containing an incorrect barometric altitude, for the criterion R10).

For these metrics, it is a matter of verifying that the ratio of two binomial variables is less than a threshold. By analogy to simple ratios, the p-value is calculated as the symmetric of those calculated for double update probabilities:

$$\rho = \int_0^1 I_{1-p_r l}(N_T - N_I, 1 + N_I) \frac{\binom{n}{k} l^{N_V} (1-l)^{N_T-N_V}}{\int_0^1 \binom{n}{k} t^{N_V} (1-t)^{N_T-N_V} dt} dl$$

5 Where:

- l plays the role of the observed probability of the intermediate value (namely the underlying probability of the law that determines the value N_V), as p_R was in the case of double update probabilities.

10 The previous expression of the p-value corresponds to integrating over all possible values of l , weighted by the likelihood of l given the value of N_V .

Fourth family

For the fourth family, the criteria involve metrics of the form:

$$M = \frac{1}{n} \sum_{i=1}^n X_i$$

15 Where:

- X is a variable of interest (for example, the age of the barometric pressure information when it is sent, for the criterion R8), given by the ratio i , and
- n is the number of considered reports.

The criterion is satisfied when these averages are less than a threshold, noted μ_r .

20 Assuming the variables X are identically distributed and independent over the reports, and a sufficient number of reports, the central limit theorem can be applied.

Let μ be the mean and σ the standard deviation (unknown) of the variable X , the observed mean $\hat{\mu}$ and the standard deviation $\hat{\sigma}$. As indicated previously, it is a matter of μ is smaller than μ_r . Making the null hypothesis $H_0 : \mu = \mu_r$, by the central limit theorem, it becomes:

$$\rho = P(M \leq \mu_r | H_0) = \Phi\left(\frac{\hat{\mu} - \mu_r}{\hat{\sigma}/\sqrt{n}}\right)$$

Where:

- Φ is the cumulative distribution function of the standard normal distribution.

30

Fifth family

For the fifth family, each of the remaining criteria is examined.

Concerning the criterion R19, a metric corresponding to the number of false reports obtained for each one-hour interval and for each 900 NM² area is considered.

This gives an empirical distribution (representable in the form of a histogram), for each time interval.

5 Let D be this distribution, M its mean, and S its standard deviation.

This metric is calculated over one hour, giving at least 450 time intervals due to the fact that the length of the time interval is at most 8 seconds.

Furthermore, the metric is cumulative.

This implies that:

$$10 \quad K = \sum_{i \in I} d_i$$

Where:

- d_i denotes a realization of D over an interval i ,
- I denotes the set of intervals i that make up the observation hour, and
- K denotes the number of false tracks observed during the observation

15 hour, over a predefined area of 900 NM².

Considering a single area of 900 NM², for N intervals in one hour, the criterion corresponds to a maximum of false tracks n_r .

An empirical renormalized distribution D^* can then be defined. By definition, D^* is equivalent to D , but the values are multiplied by $\frac{n_r}{M \times N}$.

20 It thus becomes:

$$K^* = \sum_{i \in I} d_i^*$$

Where:

- d_i^* denotes a realization of D^* over an interval i , and
- K^* is a random variable of the variable of interest (number of false

25 tracks) knowing that the distribution of the number of false tracks per interval is distributed according to the distribution D^* .

The expectation of K^* is n_r , namely, the required threshold.

The null hypothesis is that the number of false tracks observed in the zone during an interval follows the distribution D^* .

30 The p-value is thus written:

$$\rho = P\left(K^* \leq \sum_{i \in I} n_i \mid d_i^* \sim D^*\right)$$

Where:

- ρ denotes the p-value,
- P denotes the probability,
- n_i the number of false tracks observed in the zone during the interval

i , and

- $\sim$ means "follows the distribution," and therefore here that d_i^* follows the distribution D^* .

As the number of intervals is several hundred intervals i , the central limit theorem can be applied, giving:

$$\rho = \Phi \left(\frac{\sum_{i \in I} n_i - n_r}{n_r S} M \sqrt{N} \right)$$

Where:

- Φ is the cumulative distribution function of the standard normal distribution.

Thus, for each of the 900NM² areas z , it is possible to calculate the p-value ρ_z on this area

As previously, it can be verified that it is indeed the maximum of these values over each of the areas that is less than the threshold.

Thus, the global p-value (over all areas $z \in Z$ where Z is the set of areas) is calculated as follows:

$$\rho_Z = 1 - \prod_{z \in Z} (1 - \rho_z)$$

Concerning the criterion R20, a metric corresponding to the number of false tracks close to true tracks for each one-hour interval is considered.

The same reasoning as for the criterion R19 leads to the following expression for the p-value:

the p-value:

$$\rho = P \left(K^* \leq \sum_{i \in I} n_i \mid d_i^* \sim D^* \right)$$

Where D^* is the empirical renormalized distribution of the number of false tracks close to true tracks on each of the intervals.

Concerning the criterion R9, a similar strategy can be applied.

Let:

- D be the observed distribution of the ages of reports containing a pressure altitude,

- D^* be the distribution rescaled to have a mean of 16 seconds (the required threshold), and
- C^* be the cumulative function of D^* .

The p-value of the hypothesis test, the null hypothesis of which, is that the variable of the age of reports containing a pressure altitude follows the distribution D^* can then be calculated.

The following expression for the p-value thus becomes:

$$\rho = 1 - \prod_{a \in R} (1 - C^*(t_a))$$

Where:

- R denotes a set of reports, each report including a pressure altitude,
- a denotes a report, and
- t_a denotes the age of the report a and is assumed to be independent of other ages.

Calculation of the optimal value of the number of points for the determination of a criterion

The parameters influencing the optimal number of points for each criterion as a function of the previous families are given in the following table.

<i>Requirement</i>	<i>Family</i>	<i>Parameters influencing the optimal number of points</i>
R2	1	- number of errors made - number of trajectories - number of errors made per trajectory
R3	3	- number of errors made
R4	2	- mean - standard deviation - number of trajectories - mean per trajectory - standard deviation per trajectory
R5	3	- number of errors made
R6	2	- mean - standard deviation
R7	1	- number of errors made - number of cases where the reference situation does not occur
R8	4	- mean - standard deviation
R9	5	- moments of empirical laws
R10	3	- number of errors made - number of cases where the reference situation does not occur
R11	3	- number of errors made
R12	4	- mean - standard deviation
R13	4	- mean - standard deviation

R14	1	- number of errors made - number of cases where the reference situation does not occur - number of trajectories - number of errors made per trajectory - number of cases where the reference situation does not occur per trajectory
R15	3	- number of errors made - number of cases where the reference situation does not occur
R16	2	- mean - standard deviation
R17	2	- mean - standard deviation
R18	2	- mean - standard deviation
R19	5	- empirical mean of past data - empirical standard deviation of past data - number of errors made
R20	5	- empirical mean of past data - empirical standard deviation of past data - number of errors made

Thus, to determine the number of points for each requirement as well as the spatial-temporal discretization, it is necessary to provide the values of the influence parameters.

5 For this, it is assumed that a sample D of normal acquired values (having to pass all the requirements) representative of the situation being studied (data from the same airport or the same En-Route zone) is available.

It is then calculated a priori, for each requirement, the average number of points necessary for the p-value to reach the desired threshold (if the data corroborate it). Indeed,
10 if the number of points is too low, the p-value will be too high to reject the null hypothesis

with sufficient certainty, even if the data are "perfect" (in other words, they tend to validate the requirement).

Consequently, assuming perfect data, it is possible to identify the minimum number of points necessary for the p-value to fall below the required threshold.

5 This then allows to decide to perform the calculation in real-time, based on a greater number of data. For this, with the help of an expert, a compromise is determined between a small number of data, which allows a much more reactive system since it contains fewer observed time intervals, and a large number of data, which reduces the system dynamics, but offers more precision (and therefore a better view of the situation).

10 Thus, the identification of the minimum number of points can be implemented by carrying out a series of operations.

During a first operation, the maximum desired surveillance duration is obtained in order to ensure sufficient reactivity. This maximum duration is provided by the user of the tracking system 10.

15 From this duration, a maximum number of points M_{max} that can be obtained over this duration is deduced.

During a second operation, a sample D of normal values is obtained, in other words, a sample respecting all the requirements. These normal values are called normal data in the following.

20 During a third operation, for each normal data $d \in D$, the values of the parameters influencing the minimum number of points for each requirement are calculated. The minimum number m_i^d of points needed to pass the requirement i on this data d is deduced.

In practice, three cases will be presented.

25 In a first case, if $m_i^d \leq M_{max}$ for a requirement i and all the data d , then it is the ideal case where all the data of which is available, passes the requirements with very high confidence, using M_{max} data.

It is also possible to use a number of data for the requirement i depending on $\max_{d \in D} m_i^d$ (to which a margin is added).

30 According to a second case, $m_i^d \leq M_{max}$ for the vast majority of requirements i and the majority of data d , but $m_i^d > M_{max}$ in a few isolated cases.

In such a case, the number or percentage of data D that do not pass the requirements with high confidence is quantified. The user sets a tolerance threshold, below which it is tolerated that correct data raise alerts. In this case, M_{max} data are used, or a number of data for the KPI i depending on $\max_{d \in D} m_i^d$.

In the third case, a significant number of data D do not pass the requirements.

An example of a significant number in this context is, for example, half of the total data. Indeed, if one out of two measurements is bad, it can be considered that there is a malfunction of the tracking system 10 or at least that it is necessary to check its operation.

5 In such a case, it means that these requirements are not compatible with real-time evaluation. This is for example the case of rare event detection.

 It can be noted here that another possible gain in terms of the number of data points without extending the acquisition duration is to widen the geographical observation area. For this, two adjacent areas through which little traffic passes can be grouped, to
10 have, on this larger area, enough data to accurately estimate the satisfaction or not of the requirements.

 The initial areas (before grouping) are predetermined, notably chosen by the user of the tracking system 10.

 In the following, the number of points necessary to obtain a p-value less than or
15 equal to 0.05 depending on data degrading the compliance with the criterion is studied for certain criteria.

Criterion	Threshold to achieve	Number of errors			
		0	1	2	3
R2 more than 100 points	> 0.97	99	157	208	257
	> 0.99	299	473	628	773
R2 less than 100 points	> 0.97	99	>100	>100	>100
R3	< 0.005	598	947	1258	1549
R5	< 0.0003	9985	15812	20984	25844
R11 (case of a flight at constant altitude)	< 0.001	2995	4742	6294	7752
R11 (case of landing or takeoff)	< 0.015	199	315	418	515

Table 3: Number of points necessary to achieve a p-value of 0.05 or less for the criteria R2, R5, and R11

5 In Table 3, an error is a problematic report tending to degrade the criterion.

e_R	e_A							
	0	1	2	3	4	5	10	20
0	57	100	138	173	208	241	401	702
1	81	119	154	188	222	254	411	710
2	100	135	169	202	235	267	422	718
3	116	150	183	215	247	278	431	726
4	130	163	195	227	258	289	440	733
5	143	175	206	238	268	299	449	741
10	195	225	255	285	315	344	490	776
20	273	302	331	360	388	417	558	838

Table 4: Number of points necessary to achieve a p-value of 0.05 or less, for the criterion R7

In Table 4, e_R is the number of reports not containing a horizontal position and e_A is the number of reports containing a horizontal position but not containing pressure altitude.

e_R	e_A							
	0	1	2	3	4	5	10	20
0	115	200	276	348	418	485	806	1410
1	163	238	300	375	444	510	825	1422
2	200	271	339	405	470	534	844	1438
3	232	299	365	430	493	556	863	1453
4	260	325	389	453	515	577	881	1468
5	285	349	412	474	536	597	898	1462
10	388	448	508	567	626	685	977	1549
20	539	596	654	711	768	825	1107	1666

5 **Table 5:** Number of points necessary to achieve a p -value of 0.05 or less,
for the criterion R14

In Table 5, e_R is the number of reports not containing a horizontal position and
10 e_A is the number of reports.

p_R	ϵ_A								
	0	1	2	3	4	5	10	20	100
1.0	3000	4800	6300	7800	9200	10600	17000	29100	ND
0.98	3100	4900	6500	8000	9400	10800	17400	29700	ND
0.96	3200	5000	6600	8100	9600	11000	17700	30300	ND
0.9	3400	5300	7000	8700	10200	11700	18900	32300	ND
0.75	4000	6400	8400	10400	12300	14100	22700	38800	ND
0.5	6000	9500	12600	15600	18400	21100	34000	ND	ND
0.25	12000	19000	25200	31100	36700	42100	ND	ND	ND

Table 6: Number of points needed to achieve a p -value of 0.05 or less,
for the criterion R10

15 In Table 6, the number of points needed is rounded to the nearest hundred.
Additionally, p_R denotes the proportion of reports containing a pressure altitude (resp.

aircraft identifier), e_A is the number of reports containing invalid pressure altitude, and ND means this number is greater than 50,000.

The analysis of Tables 3 to 6 shows that, in most cases, the optimal number of points is less than 50,000.

- 5 The only cases where the optimal number exceeds 50,000 are cases where the number of errors is very high.

A large number of errors being a relatively rare event, these tables show that the method allows, in normal operations, to obtain compliance evaluations of the criteria with a reduced number of measurements.

CLAIMS

1. A method for evaluating the compliance of a tracking system (10) with a set of requirements, at least one requirement being a mandatory requirement and at least one requirement being a recommended requirement, each requirement imposing that a physical quantity be greater than or equal to a threshold or that a physical quantity be less than or equal to a threshold, the method for evaluating being implemented by computer, the method for evaluating including at least the following steps:

- acquisition of a set of values, the set of values comprising the values accessible to the tracking system (10) over a predefined time interval,
- for each requirement, calculation of a robustness value, the robustness being calculated as the value p of a statistical test indicating whether the requirement considered is respected or not,
- for each requirement, comparison of each robustness value to a predefined threshold according to a respective comparison criterion, and
- determination of the requirements to which the tracking system (10) is compliant, the requirements determined as compliant being the requirements for which, at the comparison step, it was determined that the comparison criterion is respected.

2. The method according to claim 1, wherein the method further includes, for each robustness value not respecting the comparison criterion, a step of determining the cause of the non-respect with the comparison criterion, the cause being either that the requirement is not respected or that the number of values is too low to guarantee a predefined reliability threshold for evaluating compliance with the requirement considered.

3. The method according to claim 2, wherein, when the determined cause is a number of values too low, the method includes a step of deducing a number of additional values to be acquired in order to obtain an evaluation reliability greater than the predefined reliability threshold.

4. The method according to claim 3, wherein the method includes an iteration of the acquisition, calculation and comparison steps until the number of values allowing to obtain an evaluation reliability greater than the predefined reliability threshold or reaches a maximum predefined number with an evaluation reliability lower than the predefined threshold.

5. The method according to claim 4, wherein the method further includes a step of issuing an alert for not respected requirements or for requirements whose number of values of the acquisition, calculation and comparison steps has reached the maximum predefined number with an evaluation reliability lower than the predefined threshold.
6. The method according to claim 4 or 5, wherein the maximum number is comprised between 40,000 and 60,000.
7. The method according to any one of claims 1 to 6, wherein, during the calculation step, an integral of a regularized incomplete beta function is calculated.
8. The method according to claim 7, wherein, during the calculation step, a p-value is equal to the result of the integral and another p-value is a linear function of the result of the integral.
9. The method according to any one of claims 1 to 8, wherein during the calculation step, a cumulative distribution function of the standard normal distribution is calculated.
10. The method according to any one of claims 1 to 9, wherein the method further includes:
- a step of applying an evaluation function specific to each requirement to the p-values to obtain evaluation values, and
 - a step of using the evaluation values to obtain a global score evaluating the compliance of the tracking system (10) with all the requirements.
11. The method according to any one of claims 1 to 10, wherein the number of requirements is greater than or equal to 20.
12. The method according to any one of claims 1 to 11, wherein the tracking system (10) provides measurement estimations from previous data from multiple sensors (12), the set of values comprising the measurement estimations.
13. An evaluation module (22) for the compliance of a tracking system (10) with a set of requirements, at least one requirement being a mandatory requirement and at least one

requirement being a recommended requirement, each requirement imposing that a physical quantity be greater than or equal to a threshold or that a physical quantity be less than or equal to a threshold, the evaluation module (22) being proper to:

- acquire a set of values, the set of values comprising the values accessible to the tracking system (10) over a predefined time interval,
- for each requirement, calculate a robustness value, the robustness being calculated as the p-value of a statistical test indicating whether the requirement considered is respected or not,
- for each requirement, compare each robustness value to a predefined threshold according to a respective comparison criterion, and
- determine the requirements to which the tracking system (10) is compliant, the requirements determined as compliant being the requirements for which, at the comparison step, it was determined that the comparison criterion is respected.

14. A tracking system (10) comprising an evaluation module (22) according to claim 13.