

CAA consultation on NERL's flight efficiency performance regime: NERL response

This response document contains NERL's response to the CAA's consultation on a flight efficiency performance regime. NERL's original proposal for a CP3 flight efficiency metric is attached and remains our core proposal. This represents the culmination of extensive analytical work and consultation with customers through the CP3 review Customer Consultation and also through the Operational Partnership Agreement.

We would welcome further opportunity to discuss the core proposal and consultation response with the CAA, should this add value to the CAA's process.

Do you consider that NERL has used best endeavours to develop a flight efficiency regime?

We first presented our environmental programme and targets to airline customers in the CCWG process in June 2009. These included a commitment to deliver a target of a 10% reduction in CO₂ emissions on average per aircraft by 2020 against a baseline of 2006.

Since that time NERL has devoted very significant efforts to a structured process of agreeing metric criteria, testing existing flight efficiency metrics against those criteria and subsequently developing an entirely new flight efficiency metric.

Having developed that entirely new metric NERL has spent the last 12 months collecting data and analysing the metric – establishing a track record on which we have now based our proposals for an incentivisation scheme design. Through this process we have engaged extensively with our NERL operational managers, CAA/RPG, our airline customers, ICAO, EUROCONTROL, CANSO and SESAR to seek peer review and challenge on the metric structure and design.

The resources dedicated to this effort have been extensive. For example, at the start of the process it took us one month to process a 36 day sample. Although we were able to increase the efficiency of this process, processing the data for 2010 still took 3 months, incorporating data for 2.1 million flights and over 1 billion radar points.

For all these reasons, we consider that we have used our best endeavours to develop a flight efficiency regime.

Do you agree that there should be a flight efficiency performance regime from the start of 2012 and that it should be on the 3Di

NERL strongly supports the deployment of 3Di as a flight efficiency performance scheme from the start of 2012, on the basis of the core proposal presented in the paper appended to the CAA consultation letter dated 2nd August 2012.

We believe that the flight efficiency performance regime should be established on the 3Di basis, as this measures efficiency on both the horizontal and vertical dimensions. A one dimensional dimension (e.g. horizontal only) would not, in our opinion, provide the full capture and measure of flight efficiency required.

Do you agree with NERL's proposals for the par value and deadband? If not, on what basis should the par value and deadband be set?

NERL believes that the par value and deadband set out in NERL's core proposal are entirely appropriate given the developing maturity of the metric. NERL believes that its core proposal is as far as possible based on evidence, where that evidence is available.

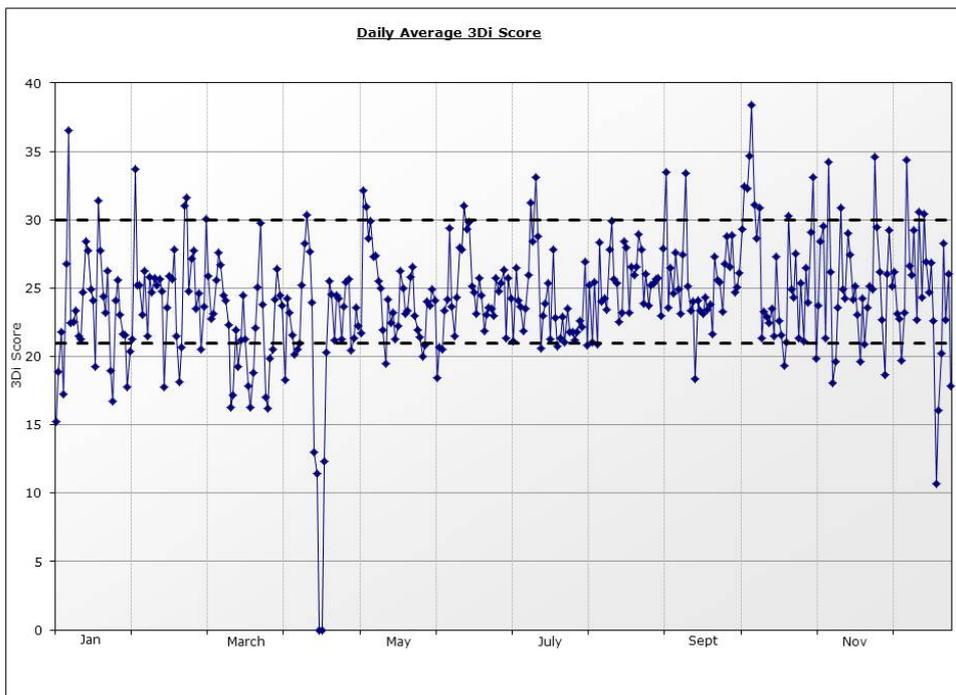
We have conducted a number of checks on the appropriateness of the deadband size:

- i. **Historical:** Historical daily 3Di score consistency; and
- ii. **Precedent:** Whether there is precedent for setting the size of the deadband; and
- iii. **Future:** Scenario testing for future 3Di score consistency.

Historical testing

We have traced performance in the 3Di score using 2010 data. The figure below shows a level of daily variation around a central level of performance.

Figure: Daily 3Di score variation from volume and unanticipated events in 2010



Note: Peaks and troughs in performance can in most instances be explained by uncontrollable circumstance, for example ash (April), snow (January) or runway maintenance (October).

Precedent testing

The current T1 delay performance regime features a deadband. We have calculated that the size of the T1 deadband relative to the metric is approximately the same as this proposed deadband for the 3Di score.

Future testing

We have tested this historical data against expected future performance. The 3Di score will be affected by the level of traffic, together with unexpected events¹. Specifically, given the base case traffic forecast combined with the historical levels of unexpected events, we could expect the historical average 3Di level of 25.5 units to increase to c.30 units (c.31 units if the high case traffic forecast is used), if we make no improvement in the network.

For clarity, this scenario is shown in the figure below.

Figure: Depiction of future scenarios



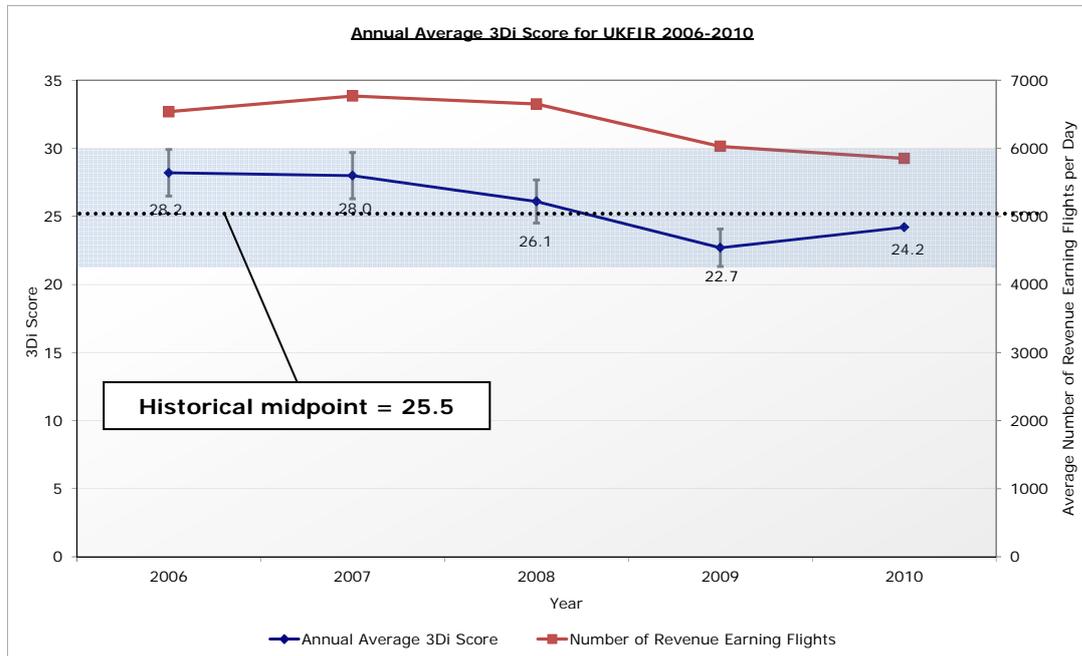
Therefore, the deadband is designed to prevent undue bonuses or penalties where non-controllable volume and unexpected events could influence the 3Di score. The size of the deadband is based on observed variances due to volume and unexpected events. We believe that the size of the deadband is appropriate because it encompasses the majority of daily average variances which can be attributable; it is approximately the same size as the T1 deadband; and because future scenario testing reveals that the deadband encompasses the effect of reasonable base case traffic and unexpected events, if we make improvement in the network.

We have proposed that a par value should be based on expectations of stretching levels of performance, given the limitations of a recovering market and historical experience. The figure below shows the historical annual average 3Di score between 2006 and 2010. Before the recession, the 3Di score averaged to c.28 units, while this fell back to c.23 units in 2009, before increasing in 2010 to c.24 units, with a recovering market (remembering that a lower score indicates improved performance).

Given the relationship between the market and the 3Di score, an increase in the 3Di score to c.30 units could be anticipated by 2014, as base case traffic forecasts recover to pre-recession levels, together with the effects of unexpected events (the high case forecast, would imply a c.31 units level by 2014). These levels exceed 2010 performance levels of c.24 units.

¹ Additional analysis of the historical traffic to 3Di relationship over the last 5 years shows a maximum variance of 7.5%.

Figure: Average annual 3Di Score 2006-2010 (showing inferred 99% confidence interval for sampled data 2006-2009)



Naturally, this assessment of future par values ignores the potential reduction in 3Di score that we expect to achieve through our efforts. Given the immaturity of the metric, it is difficult to assess the effect of our efforts on the 3Di score. However, our expectation is that our efforts to improve the network can, at best, reverse the expected effect of traffic recovery. Therefore, we propose a par value set at the historical midpoint level of 25.5 units.

NERL is strongly committed to the deployment of an incentivised flight efficiency metric that encourages the deployment of network improvements to drive fuel burn and emissions reductions, investments for the majority of this improvement is planned to start at the end of CP3. For this reason, NERL is recommending a par value and deadband range which reduce the risk of undeserved bonuses or penalties relating to factors outside of our control (such as unanticipated weather phenomena). If the 3Di metric is undermined by the award of undeserved bonuses or penalties NERL is not aware of an alternative metric that could be deployed to meet the fuel burn and emissions reduction outcomes required by airlines.

Do you agree that flight efficiency incentives should be set at £0.2 million per unit capped at 20% of available money at risk? If not, on what basis should payment rates be set?

NERL believes that the flight efficiency incentives of £0.2 million per 3Di unit capped at 20% of available money at risk represents an appropriate incentive. The bonus and penalty gradient is evidence based in that it is derived by dividing the maximum bonus of £2.4m by the 12 units difference between the maximum bonus score and the deadband. NERL believes that it is appropriate for a symmetrical gradient should be applied to balance the risk and opportunity incentives.

Do you agree with the adjustments proposed by NERL? Are there any other adjustments that should be made?

We continue to believe that the alternatives suggested increase the risk/opportunity of achieving a bonus/penalty, and/or for the value of that risk/opportunity to be increased. We do not believe that it is wise to increase this risk/opportunity given the lack of experience of using the metric.

We believe further that flight efficiency improvements are important for our industry. Such improvements may not be incentivised if a metric is discredited through unwarranted bonuses/penalties. NERL's core proposal uses historical evidence of factors beyond NERL's control, to calculate the proposed structure. Therefore, we continue to believe that the core proposal is appropriate and establishes a secure base on which to continue to develop the metric in RP2.

However, we would encourage the CAA to ensure that any adjustments it proposes are based on evidence, and act in a way that balances the incentives for flight efficiency improvements to ensure that the metric is not discredited and improvements occur.

Do you agree with the annual review process proposed and the threshold for the test?

As stated in its core proposal, we believe that the annual review process to test the continued appropriateness of the regression coefficients that underpin the 3Di score remains an appropriate and important protection for 3Di as a relatively immature metric. We also recognise that the annual review process could be a vehicle to consider the on-going appropriateness of other elements of the performance framework around the 3Di metric. However, NERL would note that annual (and possibly protracted) debate around the performance framework would represent a significant resource burden on CAA, airlines and NERL, at a time in which analysis and discussion about the RP2 successor metric should be focused upon.

Therefore, we encourage the CAA to set limited parameters around an annual review process, to ensure that gross errors in the metric do not result in undeserved bonuses/penalties, but the effort needed to conduct the review does not distract the CAA, NERL and customers from making improvements in performance or in the metric, ready for RP2.

NERL's original proposal for a flight efficiency metric in CP3, July 2011

Introduction

This paper provides NERL's proposal for a flight efficiency metric. At the last customer workshop in July 2010, NERL confirmed with customers the detail of the 3D inefficiency ("3Di") metric, while recognising that it was not in a position to propose a performance regime structure. This was because it did not have enough data to analyse which factors affected the 3Di score and whether these factors were under NERL's control or not, to ensure that unexpected events did not lead to undue NERL bonuses/penalties. The CAA confirmed the need for this additional analysis in its October 2010 CP3 formal proposals, with an aim to establish a metric on 1 January 2012.

We have spent the last 12 months collecting data and analysing the metric. We now have a greater understanding of the metric. We are now confident that based on its proposals a metric can and should be established for CP3, starting 1 January 2012.

In its October 2010 CP3 formal proposals, the CAA stated its intention to track NERL's progress in developing the metric, as well as convening workshops, as required and to consult on NERL's proposed approach. This consultation proposal sets out the work undertaken in developing the metric and also the proposal for its implementation with financial incentives. The core 3Di metric is unchanged from July 2010 and this description is reproduced in the appendix for ease of memory.

This paper discusses the background to developing the metric and performance regime in section 1; while section 2 sets out NERL's proposal for the performance regime, together with alternative scenarios for customers to consider. The appendix reproduces a description of the 3Di flight inefficiency score metric shared with customers in July 2010.

Background

In January 2008, NERL became the first ANSP to set itself an environmental target in terms of ATM CO₂ reduction. By March 2009 we launched our ATM CO₂ Plan and published its emissions baseline of circa 25Mt CO₂ for 2006. The plan identifies how CO₂ emissions are distributed across the different phases of flight, in order to understand where we can achieve reductions. It goes on to map broadly the sorts of initiatives required to meet the target.

During the CP3 Customer Consultation in 2009, NERL asked customers about their environmental requirements. We presented to airline customers our ATM CO₂ targets and its plan of action to reduce ATM CO₂ by an average of 10% per flight by 2020, against a 2006 baseline. We stated that we are strongly committed to delivering real environmental and fuel efficiency benefits to customers and, for this reason, we supported the concept of financial incentivisation based on CO₂ performance. However, because the development of environmental metrics was at a relatively immature stage at that time, we recognised that the implementation of the wrong metric could have potential to drive the wrong behaviours and lead to adverse outcomes for our customers.

Recognising the importance that airlines attached to fuel efficiency, NERL proposed a Roadmap Approach (shown in the table below), which outlined the steps that needed to be achieved in order to gain confidence on flight efficiency performance metrics for incentivisation during CP3.

Table: CP3 Customer Consultation roadmap for developing a flight efficiency metric

Step	Conditions	Indicative timescale
Step 0: Agree Roadmap Approach and timescales	NATS / Customer agreement	Confirmed through Customer Consultation
Step 1: Agree Principles for definition of suitable metrics	Principles documented and agreed	Confirmed through Customer Consultation
Step 2: Agree candidate metrics for tracking performance	Proposed candidate metrics reviewed; metrics clearly defined, documented and agreed	Confirmed through Customer Consultation (need to remain aware of alternative emerging candidate metrics)
Step 3: Track candidate metrics	Frequency, reporting mechanism, analysis methodology / data sources agreed through OPA including using airline FDR data for operational validation	From Jan '10
Step 4: Operational trials used to support tracking process	Customer support with trials programme coordinated through OPA	Until Jan '11
Step 5: Evaluate candidate metrics	Evaluation framework and criteria agreed	From Jan '11
Step 6: Agree Financially Incentivised Metric	Principles respected; financial framework, change control mechanism, reporting schedule agreed	Mid CP3 or as soon as possible before
Step 7: Implement with change control	Agreed triggers for changing incentivisation regime	Mid CP3 or as soon as possible before

NERL has held to this roadmap timetable.

During Summer 2009, NERL started to develop a metric concept. In April 2010, we introduced the 3Di score metric. This metric aimed to measure the environmental impact of NERL's work, which is flight efficiency, rather than an outcome dependent on fleet choice, such as CO₂. This was analysed over the Spring in 2010 and finalised at the consultation meeting in July 2010.

At this meeting, NERL recognised that it was not in a position to propose a performance regime structure because it did not have enough data to analyse which factors affected the 3Di score and whether these factors were under NERL's control or not. If factors beyond NERL's control were incorporated into the metric, then there is a risk that unwarranted bonuses/penalties are paid.

Therefore, NERL was granted additional time to collect and analyse data on the 3Di score. The CAA's CP3 October 2010 formal proposals requested NERL to continue to develop and test our proposed flight efficiency metric, ahead of potential implementation in 2012. Following meetings with the CAA earlier in 2011, we committed to proposing a metric structure and plan for annual CAA reviews, based on analysis of 2010 data, ahead of CAA consultation during Summer 2011.

2) Proposed performance regime

This section provides NERL's proposal for the:

- 1) Structure of the metric;
- 2) Par values and risk mitigation; and
- 3) CAA review.

The 3Di score is a new metric. It is based on a complex calculation. Having a bonus/penalty linked to a par value on a new metric opens the risk that bonuses/penalties are paid which are not deserved. NERL's proposal aims to mitigate this risk.

At the CAA's request, in this section we also lay out alternatives for the structure of the metric for consultation, which could be implemented, depending on customer's preference for risk.

Structure of the metric

Following detailed operational analysis, we have identified the factors that could affect the 3Di score. Some of these factors are within NERL's control and some are non-controllable. Such non-controllable factors pose a risk that the score could produce undue bonuses or penalties. This paper proposes mitigations for these risks, based on the precedent set in the CAA's delay metrics.

Controllable factors

NERL's operational personnel have identified actions that can improve the network's flight efficiency. These include short term innovations in areas such as tactical procedures and traffic management, as well as longer term improvements, such as better airspace design and enhanced controller tools.

Many of these improvements were discussed in detail with customers during Customer Consultation in 2009, and more regularly at OPA meetings. We believe that the metric could act as a financial incentive to fulfil these plans.

Non-controllable factors

Factors outside of NERL's control could affect the 3Di score; thereby incurring undue bonus or penalty. NERL's proposed structure mitigates this risk.

Generally, the 3Di metric is immature. Our understanding of the 3Di score has benefitted from additional analytical time during 2011. However, there remains a significantly higher degree of uncertainty around the performance of the 3Di score, compared with the well-established T1 delay metric.

These factors include:

- 1) **Traffic volume and unanticipated events:** NERL's network performance is affected by the actions of both airlines and airports. We can re-optimize our network, given additional operational complexity, but a change in daily volumes results in an effect on network efficiency, while this is achieved.

Similar effects arise from events which cannot be anticipated, such as weather effects and airport system changes. The impacts from such events are not consistent on our network.

- 2) **Traffic mix:** The ratio of types of NERL's traffic can vary, e.g. a proportional rebalancing from domestic to overflying flights. Different types of traffic require different levels and complexity of service from NERL. Therefore, 3Di scores vary by category of traffic. For example, the typical 3Di scores of aircraft overflying UK airspace are much smaller than those for aircraft arriving or departing from UK terminals. Thus, any change in the traffic mix will have a reweighting effect on the 3Di score and will affect bonuses/penalties artificially.
- 3) **Olympic and Paralympics:** The UK aviation industry is expecting an exceptional period during Summer 2012. As recognised in the CAA's delay metrics, the Olympics and Paralympics are expected to generate additional traffic, during the peak Summer season, with unusual disruption caused by security and temporary airspace changes. Owing to the immaturity of the metric, NERL is not able to fully anticipate the effect of this period on flight efficiency network performance.

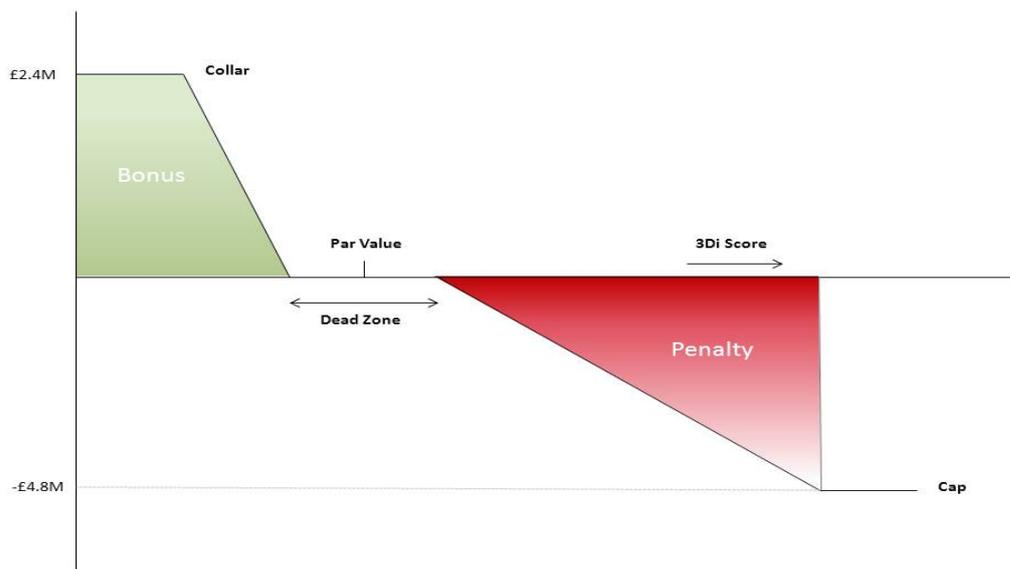
Risk mitigations

Given these risks, we have designed a metric structure that incorporates necessary mitigations. These tools are based on the precedent set by the CAA's delay metrics:

- A) **Traffic volume and unanticipated events = Deadband:** Given the level of uncertainty seen in the historical data caused by traffic volume and unanticipated events, we propose a deadband around the par value; performance within which would not result in a bonus or a penalty. This reflects the historical expectation that effects within this range are outside of NERL's control. This follows a traffic related precedent from the CAA's T1 delay term;
- B) **Traffic mix = Par value modulation:** Given potential trends in traffic mix, we propose an automatic function that would change the par value depending on the ratio of key traffic categories. This reflects the expectation that NERL should not benefit purely from a greater proportion of more favourable traffic, and vice versa. This follows precedent from the traffic related par value modulation for T1; and
- C) **Olympics and Paralympics = Exclusion period:** Given the uncertain effect of the games on NERL's network, we propose an exclusion period, during which the metric would not apply. This follows precedent from T1 and T2, for which a different treatment is enforced for the 2.5 month period.

The figure below shows a graphical representation of the metric, par value and risk mitigations.

Figure: Proposed structure of the 3D inefficiency metric



Note: The minimum and maximum bonus/penalties reflect the CAA's CP3 decision on financial exposure for the metric and are listed in 2006 prices.

The calculation of the par values and relevant figures for the risk mitigations is given below.

Par value and risk mitigation

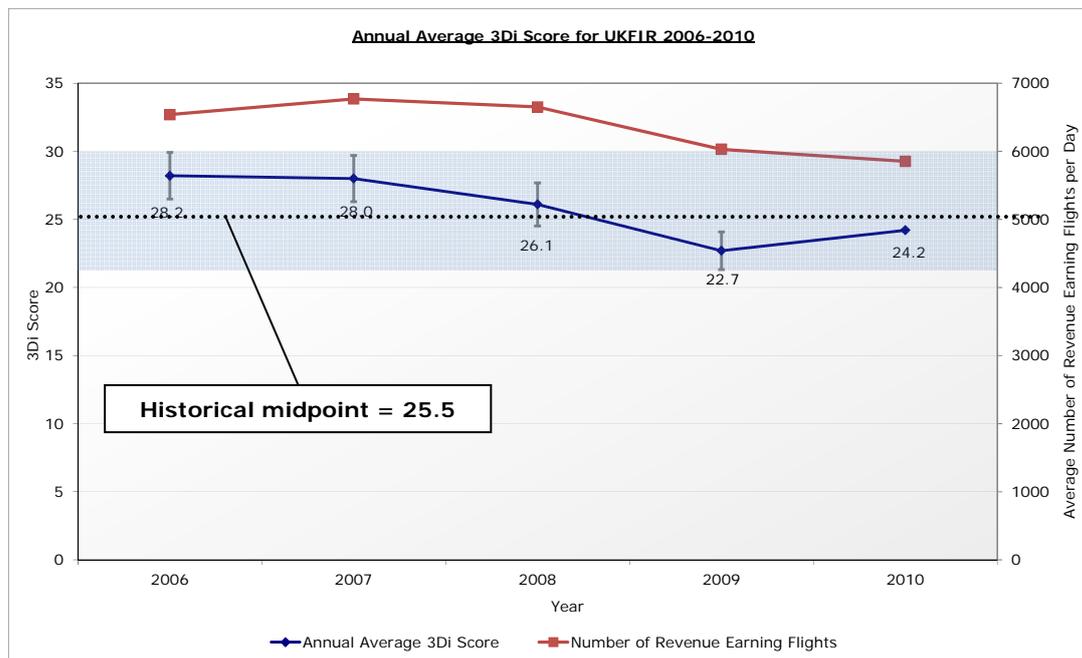
We have used the data collected and analysed over the last 3 years to create proposals for the par value and other values linked to the performance regime and risk mitigation tools.

Par value

We propose that a par value should be based on expectations of stretching levels of performance, given the limitations of a recovering market and historical experience. The figure below shows the historical annual average 3Di score between 2006 and 2010. Before the recession, the 3Di score averaged to c.28 units, while this fell back to c.23 units in 2009, before increasing in 2010 to c.24 units, with a recovering market (remembering that a lower score indicates improved performance).

Given the relationship between the market and the 3Di score, an increase in the 3Di score to c.30 units could be anticipated by 2014, as base case traffic forecasts recover to pre-recession levels, together with the effects of unexpected events (the high case forecast, would imply a c.31 units level by 2014). These levels exceed 2010 performance levels of c.24 units.

Figure: Average annual 3Di Score 2006-2010 (showing inferred 99% confidence interval for sampled data 2006-2009)



Naturally, this assessment of future par values ignores the potential reduction in 3Di score that we expect to achieve through our efforts. Given the immaturity of the metric, it is difficult to assess the effect of our efforts on the 3Di score. However, our expectation is that our efforts to improve the network can, at best, reverse the expected effect of traffic recovery. Therefore, we propose a par value set at the historical midpoint level of 25.5 units.

Maximum bonus/penalty

The CAA set the maximum financial exposure due to the metric in the CP3 decision. These maximum bonus/penalty values need to be linked to 3Di score values. The level of these values also dictates the rate at which 1 unit relates to a proportion of bonus or penalty. Our proposal is as follows:

- Maximum bonus:** Since we cannot achieve a zero 3Di score, we propose a maximum bonus set at a realistic, yet stretching level, to act as an incentive. The best *daily* average 3Di score between 2006 and 2010 was 9 units on Christmas Day 2009. Therefore, we propose to set this level to achieve the maximum bonus for the *annual* average 3Di score.

Achieving Christmas Day performance linked to low traffic level, on average throughout the year would be impossible without significant operational improvement. Therefore we believe that setting this score level for the maximum bonus is based on real precedent, yet is stretching and a good incentive.

This maximum bonus score of 9 units results in a bonus gradient of £200K per unit improvement².

² £200k is derived by dividing the maximum bonus £2.4m (expressed in 2006 prices in the CAA's CP3 decision) by the 12 units difference between the maximum bonus score and the deadband).

- **Maximum penalty:** We propose to set the maximum penalty based on a fair penalty gradient which is symmetrical to the bonus gradient. This would set the maximum penalty score at 54 units.

The figure at the end of this section depicts the maximum bonus and penalty scores.

Deadband

As described above, the deadband is designed to prevent undue bonuses or penalties where non-controllable volume and unexpected events could influence the 3Di score. We propose to set the deadband according to observed levels of these effects traced in the historical 3Di score:

- **Maximum value:** The maximum score attributable to volume and unexpected events was c.30 units in 2006 (annual average, including inferred 99% confidence intervals), at the peak of recent traffic levels; and
- **Minimum value:** The minimum score attributable to volume and unexpected events was c.21 units in 2009 (annual average, including inferred 99% confidence intervals).

Therefore, we propose a deadband between 21 and 30 units.

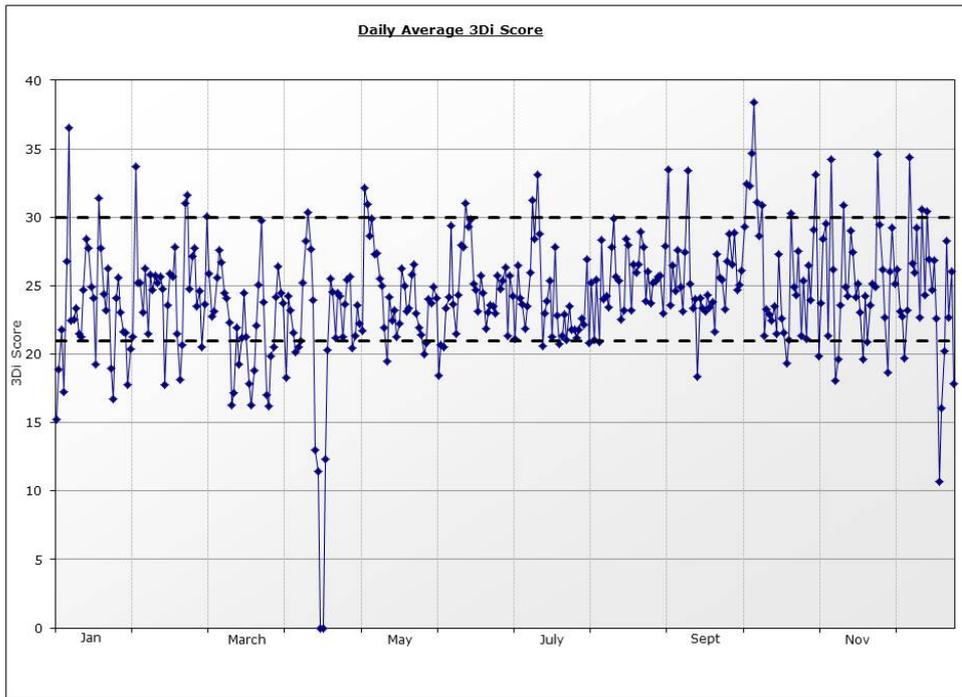
We have conducted a number of checks on the appropriateness of the deadband size:

- i) **Historical:** Historical daily 3Di score consistency; and
- ii) **Precedent:** Whether there is precedent for setting the size of the deadband; and
- iii) **Future:** Scenario testing for future 3Di score consistency.

Historical testing

We have traced performance in the 3Di score using 2010 data. The figure below shows a level of daily variation around a central level of performance.

Figure: Daily 3Di score variation from volume and unanticipated events in 2010



Note: Peaks and troughs in performance can in most instances be explained by uncontrollable circumstance, for example ash (April), snow (January) or runway maintenance (October).

Precedent testing

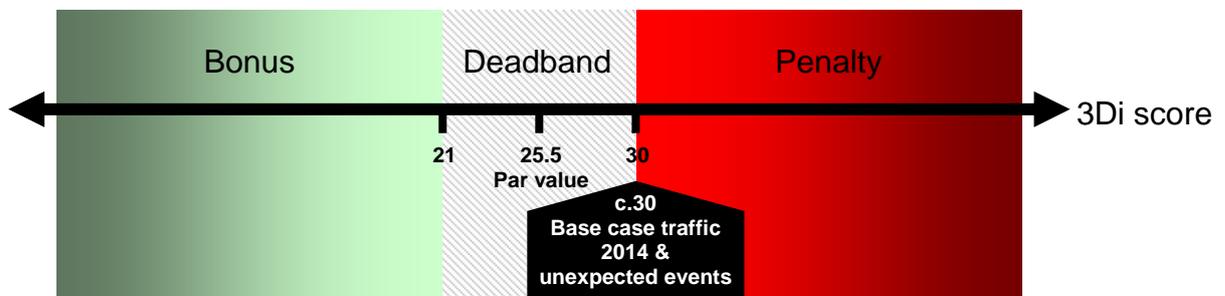
The current T1 delay performance regime features a deadband. We have calculated that the size of the T1 deadband relative to the metric is approximately the same as this proposed deadband for the 3Di score.

Future testing

We have tested this historical data against expected future performance. The 3Di score will be affected by the level of traffic, together with unexpected events³. Specifically, given the base case traffic forecast combined with the historical levels of unexpected events, we could expect the historical average 3Di level of 25.5 units to increase to c.30 units (c.31 units if the high case traffic forecast is used), if we make no improvement in the network.

For clarity, this scenario is shown in the figure below.

Figure: Depiction of future scenarios



³ Additional analysis of the historical traffic to 3Di relationship over the last 5 years shows a maximum variance of 7.5%.

Therefore, the deadband is designed to prevent undue bonuses or penalties where non-controllable volume and unexpected events could influence the 3Di score. The size of the deadband is based on observed variances due to volume and unexpected events. We believe that the size of the deadband is appropriate because it encompasses the majority of daily average variances which can be attributable; it is approximately the same size as the T1 deadband; and because future scenario testing reveals that the deadband encompasses the effect of reasonable base case traffic and unexpected events, if we make improvement in the network.

We have made a number of further observations:

- 1) **Par value:** Given traffic expectations, we will have to make significant improvements in the network to reach the par value level;
- 2) **Bonus:** A bonus would only reflect additional hard work. Given the effect of growing traffic we anticipate the 3Di score to increase to c.30 units, in order for NERL to achieve a bonus, we will have to improve the score by over 9 units. To do this, first, we would have to reverse the effect of additional volume on the network and then second, work even harder than planned. Therefore the bonus will incentivise additional future initiatives; and
- 3) **Penalty:** If we do not act to improve the network, then the risk of a penalty is greater than the opportunity of generating a bonus, given the expected base case traffic forecast position of c.30 units. Therefore, we will be strongly incentivised to continue to pursue flight efficiency initiatives and to avoid a penalty.

Par value modulation

As described above, par value modulation is designed to prevent undue bonuses or penalties where non-controllable changes in traffic mix could influence the 3Di score. Therefore, we propose that the par value would modulate automatically, where the proportion of overflight traffic alters by 5 percentage points or more (e.g. an increase from 10% to 15% of all traffic).

Analysis of a range of changes to the proportion of flights in each category has identified that these threshold values typically alter the 3Di score by approximately 1 unit (4%), based on 2010 performance.

Therefore, we propose that there would be a stepped change in the par value if the composition of traffic changes to meet these thresholds, as shown in the table below. We believe that a stepped modulation is important to ensure that the par value only changes for significant traffic mix movements, rather than for small traffic mix movements, which could create too much uncertainty about the level of the par value.

However, if traffic mix continues to change once these thresholds have been breached, we propose a smooth modulation of the par value, using the same multiplier, as shown in the table below.

Table: Par value modulation

Traffic Proportion Scenario	Threshold Change in Proportion of Flight Category	Change in Par Value	Indicative Par Value (based on 25.5 proposal)	Additional Change In Par Value Per 1% Over Threshold
Overflights Increase	+5%	-4%	24.5	-0.8%
Overflights Decrease	-5%	4%	26.5	+0.8%

Please note: Traffic mix has remained largely unchanged over the last five years. However, since a change in traffic mix could have a significant effect on the 3Di score, we believe that the par value modulation is an important precaution.

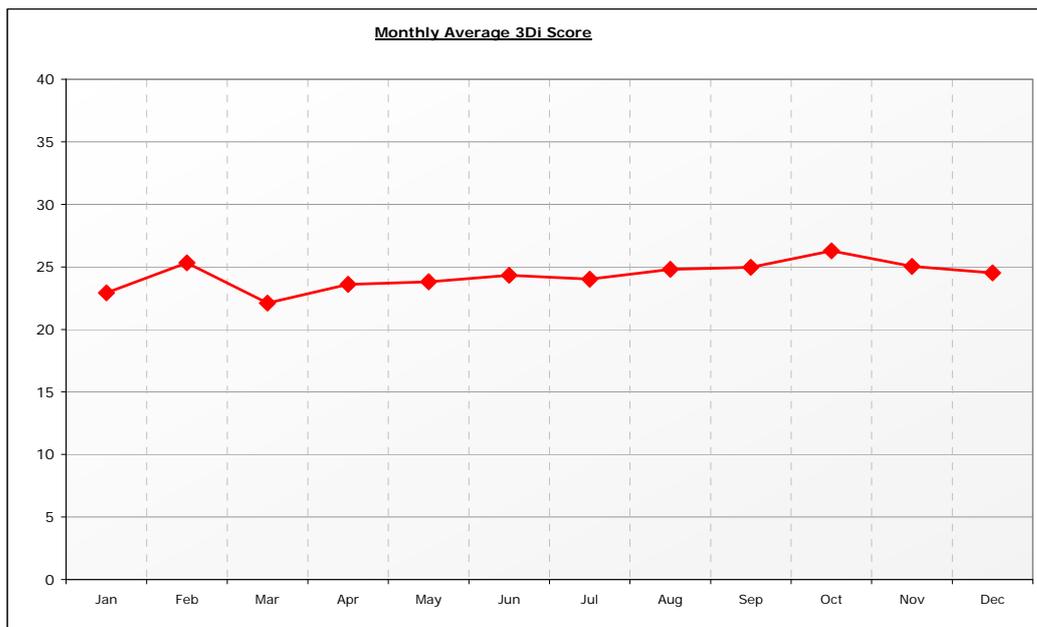
Exclusion period

As described above, the Olympic/Paralympic exclusion period is designed to prevent undue bonuses or penalties which could be generated through the non-controllable effect of the games in Summer 2012. We propose an exclusion period, during which the metric would not apply, covering the same 2.5 months as the T1 Olympic/Paralympic period.

Importantly, we note that there was no relationship between Summer seasonality and the 3Di score, in contrast with delays (as seen in the figure below). This implies that the average 3Di score in the remainder of 2012, should be no different from the average level seen in any other year which includes the Summer period, all other things held constant.

While this may appear counter-intuitive, we hypothesise that weather events are of course a key determinant of flight routings, and there are fewer weather constraints in the summer compared with the rest of the year, counteracting the effect of greater volume.

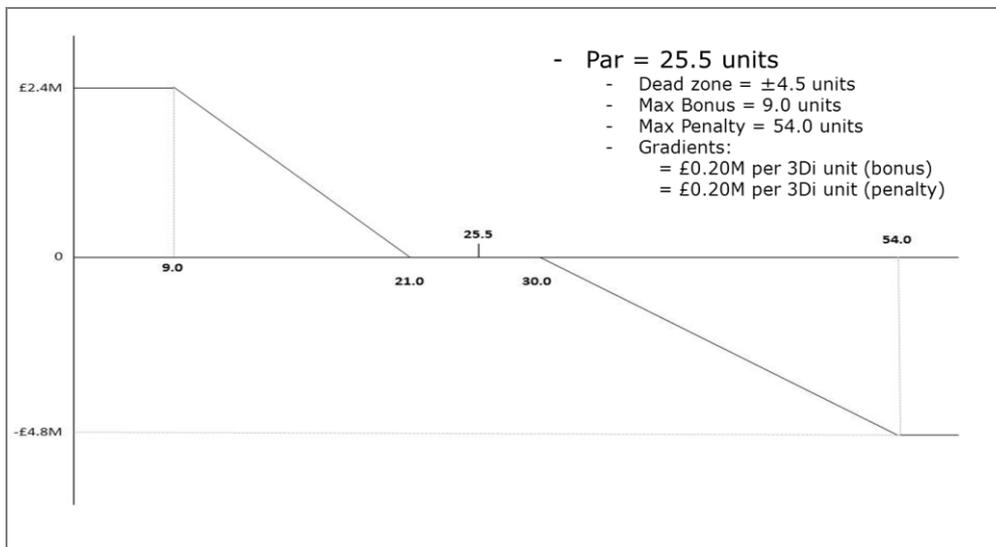
Figure: Monthly average 3Di score during 2010



Performance regime structure

Given these proposals, the 3Di score performance regime structure is represented in the following schematic below.

Figure: Proposed structure for the 3Di performance regime



CAA review

Following meetings with the CAA, NERL proposed the creation of an annual review, to allow for the removal of a bonus/penalty, should the metric basis prove to be unstable during CP3. We believe that the CAA review should be:

- 1) **Transparent:** The pre-specified test should be based on a transparent calculation, which CAA analysts have already performed, and thus, this would be a calculation that they could repeat; and
- 2) **Mechanically implemented:** The test should be pre-specified in the licence. Test failure (as described below) should result in removal of the bonus/penalty for the relevant year. This should happen without the need to re-open CP3, to change the licence or for specific additional CAA consultation.

Description of the CAA review test

The CAA review would test the continued appropriateness of the regression coefficients that underpin the 3Di score. While we continue to believe that these coefficients will remain stable during CP3, we acknowledge the small risk that undue bonuses/penalties could ensue from unstable coefficients, which could result from changes to the industry.

Therefore, NERL proposes to conduct a test in February/March 2013 using calendar year data from 2012, and annually thereafter. The test uses a randomly chosen sample of 2012 data, with a target of 50,000 flights, and applies the methodology used to derive the 3Di metric coefficients. The test model is applied to the full calendar year data from 2012 and the mean score is compared to the actual 2012 mean 3Di score. If the difference between the mean scores rises above a pre-specified threshold (defined below), then the test fails and the 3Di bonus/penalty for 2012 would be cancelled. If the difference between the mean

scores falls within the pre-specified threshold, then the test is passed and the bonus/penalty is applied.

Frequency of the CAA review

The initial test would be performed in 2013 for the 2012 period, to provide a better understanding of the performance of the metric. If the review test is failed in a given year then the test would be repeated in the following year. This means that if the 2012 data fails the test, the metric is potentially re-established for 2013 and 2014 bonuses/penalties, dependent on the test being met for those years.

We believe that it is important to re-establish the metric following a test failure, to ensure that the metric is not brought down by an anomalous year. However, if the metric fails due to a long term industry change that permanently invalidates the metric, then annual tests should also fail and the risk of undue bonuses/penalties is mitigated.

NERL would conduct the test in the first instance and present the findings to the CAA for verification by 15 March 2013. The CAA could then confirm the result, or ask for further analytical work, in order to take a view on the validity of the bonus/penalty. The test result would be published by the middle of April 2013, to allow the financial statements to reflect the outcome.

CAA review test threshold

NERL proposes a test threshold which is statistically appropriate to ensure that bonuses/penalty is valid; the difference between the mean 3Di score⁴ and the mean test model score as a percentage of the mean 3Di score should be less than or equal to 6%. This threshold represents a maximum error of the metric due to the potential instability of the 3Di coefficients of approximately ± 1.5 3Di units (based on 2010 performance).

Alternatives

NERL's core proposal is based on the fact that the metric is new and still poorly understood. This means that we are proposing measures that mitigate significant risks that bonuses/penalties are paid for factors affecting the metric that cannot be attributed to NERL actions. However, we recognise that customers may have a different risk appetite, and this should be explored in this consultation.

Therefore, customers could choose the following alternatives to NERL's core proposal:

Alternative 1) Reduced deadband size (deadband between 22.5 and 28.5 units): Currently, the deadband (between 21 and 30 units) is based on a historical assessment of factors beyond NERL's control. Reducing the deadband would increase the likelihood of NERL paying a penalty, but also increase the likelihood of customers paying a bonus to NERL for factors beyond its control; AND/OR

Alternative 2) Reduced maximum bonus/penalty score: This would increase the bonus/penalty value of one unit from £200k to £250k⁵ (i.e. the maximum bonus/penalty would be paid at 11.4/49.2 units respectively). This means that the value of a penalty paid by NERL would increase, but also the value of a bonus paid by customers to NERL would also increase; AND/OR

⁴ This is the original regression model on which the metric is based.

⁵ 2006 prices, as per CAA's CP3 decision.

Alternative 3) Increased CAA review test threshold (equivalent to ± 6 units): Currently, the threshold level (equivalent to ± 1.5 units, based on 2010 figures) has a statistical basis in NERL's proposal. An increased threshold would prevent the cancellation of bonuses/penalties for less extreme instability in the underlying metric model. However, more generally, increasing the threshold would grow the likelihood of paying a bonus/penalty if changes to the industry result in some instability in the underlying metric model.

These alternatives increase the risk/opportunity of achieving a bonus/penalty, and/or for the value of that risk/opportunity to be increased. However, we do not believe that it is wise to increase this risk/opportunity given the lack of experience of using the metric.

We believe that flight efficiency improvements are important for our industry. Such improvements may not be incentivised if a metric is discredited through unwarranted bonuses/penalties. NERL's core proposal uses historical evidence of factors beyond NERL's control, to calculate the proposed structure. Therefore, we continue to believe that the core proposal is appropriate and establishes a secure base on which to continue to develop the metric in RP2/CP4.

NERL

29 July 2011

Appendix: 3Di flight inefficiency score metric

During the CP3 Customer Consultation meetings, NERL and the customers agreed a number of criteria that an incentivised metric should conform to. The principle criteria agreed were:

- It drives the right NERL behaviours – consistent with customer objectives for increased fuel efficiency
- It is fair and equitable across NERL's customer base
- It accurately reflects fuel/CO₂ performance outcomes from NERL initiatives and not be unduly affected by factors beyond NERL influence
- It does not lead to unintended consequences by incentivising actions that increase fuel use
- It is transparent, measurable and auditable

At the CAA and customer workshop in April 2010, NERL reported its assessment of the fit of a group of proposed metrics with the criteria (shown in the table below).

Table: Candidate Metrics vs. Criteria

	Drives NERL behaviours to reduce fuel use	Fair & equitable	Reflects Fuel/CO ₂ outcomes & not affected by factors beyond NERL's control	Free from unintended consequences	Transparent, measurable & auditable
Total NATS Operational CO ₂	Yes	No	No	No	No
CO ₂ per Flight	Yes	No	No	No	No
ATM Efficiency Index	Yes	No	No	No	No
Project Benefit Delivery	Yes	No	No	No	Yes
Eurocontrol PRU Route Extension	No	Yes	No	No	Yes

Also at the April 2010 workshop, NERL reported that it felt that its initial analysis showed that the newly developed 3D Inefficiency Score seemed to conform to these criteria. However it was also noted that NERL needed to develop an improved version of metric to inform an incentivisation scheme from the beginning of CP3. NERL also pointed out that the metric had not been tested in terms of the way it behaves to changes to the system outside of NERL's control and those as a result of NERL's actions.

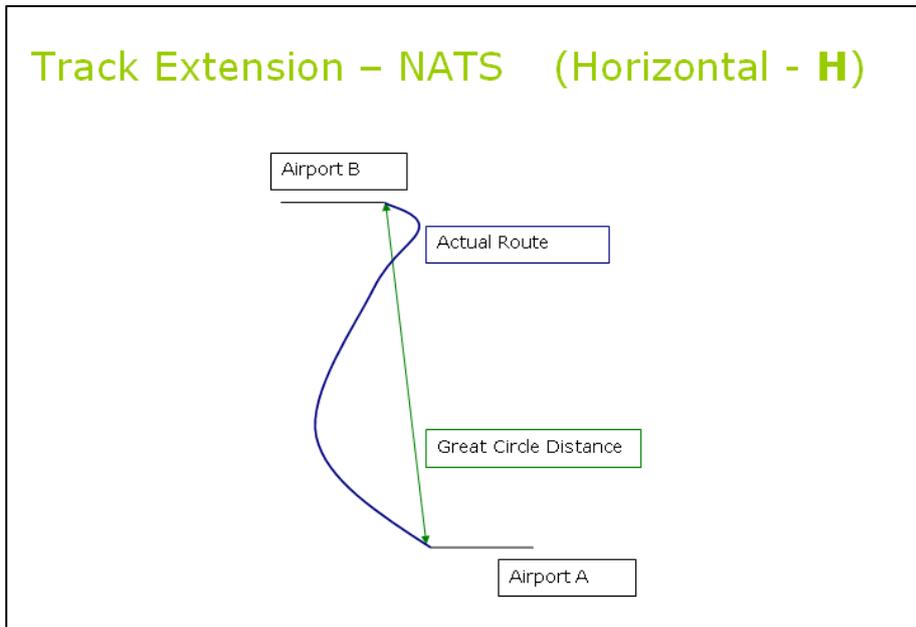
In addition, greater understanding of the metrics behaviour under day-to-day or year-to-year changes to the system would be needed to inform par value discussions. NERL undertook to develop the metric based on the key requirements to improve its predictive accuracy and our understanding of how the metric tracked by the end of June 2010. These regression analysis and metrics tracking tasks are described in the next sections.

Regression Analysis

Having identified through the Roadmap approach (as described in page 2) that existing metrics were unsuitable for use as an incentivised service quality term NERL undertook to develop a predictive model for fuel inefficiency using regression analysis – a technique for modelling one variable from a combination of others. Using this approach enables the relative contributions from lateral and vertical components of flight inefficiency to be combined into a single measure of inefficiency, called the 3D Inefficiency Score.

The definition of the horizontal element of inefficiency, to add into the regression analysis, was relatively easy to derive – simply the difference between the lateral radar track and the great circle distance within UK airspace as illustrated in the figure below for the example of a domestic flight.

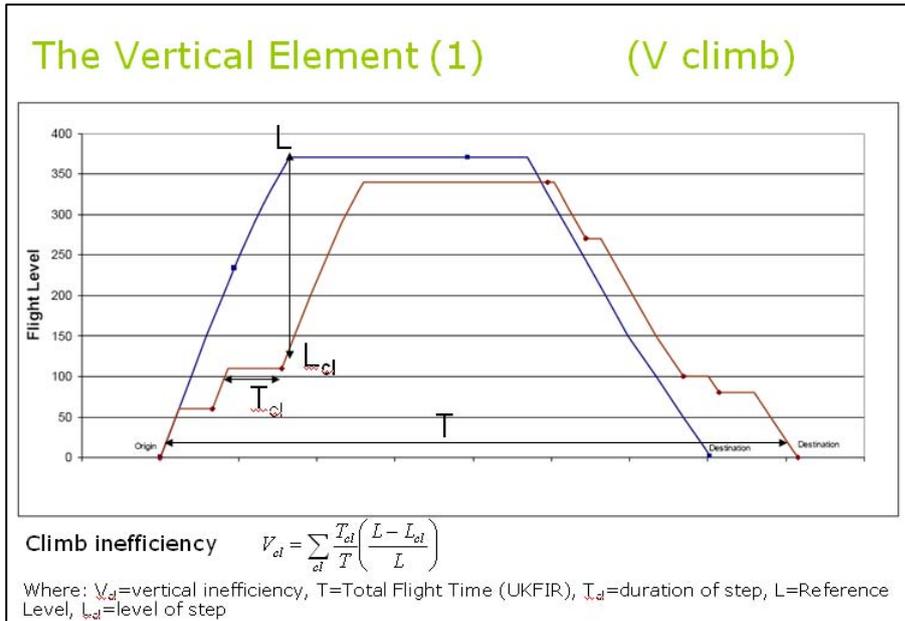
Figure: Horizontal Inefficiency



The NERL definition of horizontal inefficiency does not impose a 30NM exclusion zone around departure and destination airports (as per other track extension metrics), therefore taking account of horizontal inefficiency in the area close to the airport. When assessing the horizontal element of inefficiency for a flight the metric compares the actual entry and exit point into and out of UKFIR with the great circle distance for that flight to the same transit point. As such the impact of external factors that change the entry/exit point of aircraft into/out of the UKFIR do not adversely penalise the 3D Inefficiency Score.

Characterising vertical inefficiency was somewhat more complex and NERL has had to dedicate significant effort investigating methods to estimate this. The result has been the development of a vertical inefficiency term that is conceptually similar to the methodology used for horizontal track extension, i.e. a proxy measure of vertical inefficiency can be defined as the vertical deviation of a flight from the airline requested cruise flight level, taking consideration of the different fuel flow performance of aircraft across different phases of flight. Climb and descent inefficiency is shown diagrammatically in the figures below.

Figure: Vertical Inefficiency



Previous NERL studies (reported to the CAA and customer workshop on the 22nd April 2010) indicated that predictive accuracy could be improved by:

- greater consideration of phase of flight to account for differences in fuel flow rates
- accounting for the interaction terms, where vertical and horizontal inefficiency occur simultaneously e.g. during airborne holding where the aircraft is subject to both track extension and vertical inefficiency
- use a larger data sample when undertaking regression analyses

By July 2010, NERL regression studies to support the development of the 3D Inefficiency Score metric used a sample of data from over 174,000 individual flights within UK domestic airspace during 2009 and divided vertical inefficiency into three unique groupings by phase of flight (climb, cruise, and descent). Using different combinations of these explanatory variables, five regression models have been considered. The variables considered in each model are summarised in the table below:

Table: Regression Models

Model	Explanatory Variables Considered					
	Relative Track Extension (τ)	Combined Vertical Inefficiency (ν)	Vertical Inefficiency in Climb (ν_{CL})	Vertical Inefficiency in Cruise (ν_{CR})	Vertical Inefficiency in Descent (ν_D)	Interaction Terms
A	X					
B	X	X				X
C	X		X	X	X	
D	X		X	X	X	X*
E	X		X	X	X	X

* Model D contains only the interaction terms formed between track extension and individual vertical inefficiency terms, whilst model E contains all possible interaction terms.

Statistical tests show that all of the available models can be used to provide a predictive estimate of fuel inefficiency. In general, the addition of more explanatory

variables has been shown to increase how well the models predict fuel burn and CO₂ inefficiency.

Statistical tests have shown that Model D provides the most accurate and consistent predictions of fuel inefficiencies for individual flights. The more complex Model E was found to introduce significantly more complexity without a corresponding improvement in predictive accuracy. NERL is confident that Model D can be used to accurately predict fuel inefficiency within UK airspace. Model D takes the following form:

$$\varphi = A\tau + Bv_{CL} + Cv_{CR} + Dv_D + Ev_{CL}\tau + Fv_{CR}\tau + Gv_D\tau$$

Where

- $\varphi =$ 3D Inefficiency Score (i.e. predicted fuel inefficiency)
- $\tau =$ The proportion of track extension of a flight
- $v_{CL} =$ The vertical fuel efficiency of the climb phase of flight
- $v_{CR} =$ The vertical fuel efficiency of the cruise phase of flight
- $v_D =$ The vertical fuel efficiency of the descent phase of flight

A, B, C, D, E, F, and G are real constants.

The detail of the formation of the regression model supporting 3D Inefficiency Score has been reviewed by RPG.

Tracking the Metric

Since July 2010 NERL has spent a lot of analytical time calculating the effect of changes in internal and external factors on the metric, through the regression relationship. This work aims to understand the inherent risks involved in establishing a new metric for financial incentivisation, as per customer requests. The output of this work is