

Assessment & optimisation of STCA performance: Using the Pareto-optimal receiver operating characteristic

William Reckhouse*, Richard Everson*, Jonathan Fieldsend*,
David Bush†, Trevor Arnold†, Richard Hayward†, Keith Slater†

*School of Engineering, Computing and Mathematics, University of Exeter

†Operational Analysis Department, NATS.

Abstract

Short Term Conflict Alert (STCA) systems are complex software programs, with many parameters that must be adjusted to achieve best performance. We describe a simple evolutionary algorithm for optimising the trade-off between wanted alerts and nuisance alerts. The procedure yields an estimate of the Pareto optimal Receiver Operating Characteristic for the STCA system and we discuss additional uses of this for characterising and comparing the performance of STCA systems and airspace.

1 Introduction

The recent European move towards standardisation of Short Term Conflict Alert (STCA)-type systems means that by the end of 2008 all European ANSPs (Air Navigation Service Providers) will be expected to implement a system which at least conforms to the minimum requirements laid out in the Eurocontrol specification [EUROCONTROL, b]; see [EUROCONTROL, a] for additional guidance material. STCA systems are essentially classifiers that classify aircraft pairs into ‘operationally relevant alerts’ or true positives (TP) and false positive (FP) or ‘nuisance alerts’ – alerts that unnecessarily attract the controller’s attention. Central to the safety case for STCA are the two criteria [EUROCONTROL, b]:

1. “the proportion of conflicts detected by the Controller in time for controller resolution will be enhanced by the use of STCA”; and
2. “any negative effects on safety shall be small compared with the safety benefit and reduced as far as reasonably practical”.

Thus the Eurocontrol specification stipulates that an STCA system shall provide alerts for operationally relevant conflicts, while keeping false and nuisance alerts to an effective minimum.

The challenge for system analysts and maintenance personnel is in making informed decisions about a classifier’s likely performance and selecting an appropriate trade-off between ‘operationally relevant’ alerts, and those which are deemed nuisance or false. STCA systems are generally complex and tuning them manually involves adjustment of a great many (typically over 1000) operational parameters. In this paper we briefly describe how automation of the optimisation process, using a multi-objective evolutionary algorithm, can produce a good approximation to the best possible set of trade-offs between true and false positive rates, and we explore further uses of this optimal set of trade-offs.

2 ROC analysis & Pareto optimality

Regarding the STCA system as a classifier $g(\mathbf{x}; \boldsymbol{\theta})$ which gives an estimate of the probability that a feature vector \mathbf{x} (for STCA the radar tracks of a pair of aircraft) belongs to one of two classes. We assume that the classifier depends upon a vector of adjustable parameters $\boldsymbol{\theta}$, and we denote by $T(\boldsymbol{\theta})$ the classifier's true positive classification rate, while the false positive rate is denoted by $F(\boldsymbol{\theta})$.

If the costs of an incorrect classification were known it would be straightforward to calculate the expected cost for any particular parameter and data set Duda and Hart [1973]. It would then be possible to adjust the parameters to minimise the expected cost. However, this procedure requires accurate specification of the misclassification costs which are seldom known accurately. However, the Receiver Operating Characteristic (ROC) curve displays the trade-off between true and false positive rates as the ratio of misclassification costs is varied for a particular set of parameters (see Fawcett [2006] for a recent review of ROC methods). As the misclassification cost ratio is varied a non-decreasing ROC curve in the (F, T) plane is obtained for any particular fixed set of parameters, and different, possibly intersecting, ROC curves are obtained for different parameters. With the ROC curves on hand the user can select the operating point with a full knowledge of the possible trade-offs involved.

Of course, all these measures based upon the ROC curve require knowledge of the ROC curve, which hitherto has been unavailable for the STCA system. In this section we show how multi-objective evolutionary algorithms may be used to derive the ROC curve for the STCA system optimised over all possible parameter values. That is, we seek to discover the set of parameters that simultaneously minimise $F(\boldsymbol{\theta})$ and maximise $T(\boldsymbol{\theta})$.

A general multi-objective optimisation problem seeks to simultaneously extremise D objectives:

$$y_i = f_i(\boldsymbol{\theta}), \quad i = 1, \dots, D \quad (1)$$

where each objective depends upon a vector $\boldsymbol{\theta}$ of P parameters. It is convenient to assume that all the objectives are to be minimised, so for the STCA system we minimise the pair of objectives $(-T(\boldsymbol{\theta}), F(\boldsymbol{\theta}))$. The multi-objective optimisation problem may be conveniently expressed as:

$$\text{minimise } \mathbf{y} = \mathbf{f}(\boldsymbol{\theta}) = (f_1(\boldsymbol{\theta}), \dots, f_D(\boldsymbol{\theta})) \quad (2)$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_P)$ and $\mathbf{y} = (y_1, \dots, y_D)$. Note that the parameters may also be subject to some constraints, such as being within a previously explored range.

When faced with only a single objective an optimal solution is one which minimises the objective given the constraints. However, when there is more than one objective to be minimised solutions may exist for which performance on one objective cannot be improved without sacrificing performance on at least one other. Such solutions are said to be *Pareto optimal* Coello Coello [1999], Veldhuizen and Lamont [2000] and the set of all Pareto optimal solutions is said to form the Pareto set.

The notion of *dominance* may be used to make Pareto optimality clearer. A parameter vector $\boldsymbol{\theta}$ is said to *strictly dominate* another $\boldsymbol{\phi}$ (denoted $\boldsymbol{\theta} \prec \boldsymbol{\phi}$) if and only if all the objectives corresponding to the parameter $\boldsymbol{\theta}$ are no worse than those obtained with $\boldsymbol{\phi}$ and at least one objective is better, that is:

$$\begin{aligned} f_i(\boldsymbol{\theta}) &\leq f_i(\boldsymbol{\phi}) \quad \forall i = 1, \dots, D \quad \text{and} \\ f_i(\boldsymbol{\theta}) &< f_i(\boldsymbol{\phi}) \quad \text{for some } i. \end{aligned} \quad (3)$$

Algorithm 1 Multi-objective optimisation of STCA.

```
1:   $E := \text{initialise}()$ 
2:  for  $n := 1 : N$                                 Loop for  $N$  generations
3:     $\theta := \text{select}(E)$                           Select parent to perturb
4:     $\theta' := \text{perturb}(\theta)$                     Perturb parameters
5:     $(T(\theta'), F(\theta')) := \text{STCA}(\theta')$       Evaluate classification rates
6:    if  $\theta' \not\leq \phi \forall \phi \in E$ 
7:       $E := \{\phi \in E \mid \phi \not\leq \theta'\}$         Remove dominated elements
8:       $E := E \cup \{\theta'\}$                        Insert  $\theta'$ 
9:    end
10: end
```

Less stringently, θ weakly dominates ϕ if the objectives corresponding to θ are all at least as good as those corresponding to ϕ :

$$f_i(\theta) \leq f_i(\phi) \quad \forall i = 1, \dots, D. \quad (4)$$

A set E of decision vectors is said to be a *non-dominated set* if no member of the set is dominated by any other member:

$$\theta \not\leq \phi \quad \text{for all } \theta, \phi \in E. \quad (5)$$

A solution to the minimisation problem (2) is thus *Pareto optimal* if it is not dominated by any other feasible solution, and the non-dominated set of all Pareto optimal solutions is the Pareto front. Recent years have seen the development of a number of evolutionary techniques based on dominance measures for locating the Pareto front; see Deb [2001], Zitzler et al. [2003] for recent reviews.

2.1 ROC optimisation

The basis of our algorithm for locating the Pareto front for STCA is a simple multi-objective evolutionary algorithm. In outline, the procedure operates by maintaining an archive, E , of mutually non-dominating solutions, θ , which is the current approximation to the Pareto front/ROC curve. The algorithm is iterative; at each step some solutions in E are copied and perturbed. Those perturbed solutions that are dominated by members of E are discarded, while the others are added to E and any dominated solutions in E are removed. In this way the estimated Pareto front E can only advance towards the true Pareto front.

Algorithm 1 describes in more detail the algorithm as applied to the optimisation of the STCA system; see Everson and Fieldsend [2006], Reckhouse et al. [2008] for additional details.

The archive or frontal set E is initialised by drawing parameters for the STCA system uniformly from their feasible ranges; in addition the current ‘best’ parameter set from manual tuning θ^* is added to E . Of course many of these randomly selected parameter vectors are dominated by other parameter vectors and these dominated parameters are deleted from E so that E is a non-dominated set (5).

Following initialisation, the loop on lines 2–10 of Algorithm 1 is repeated for N iterations. At each iteration a single parameter vector θ is selected from E ; selection may be uniformly

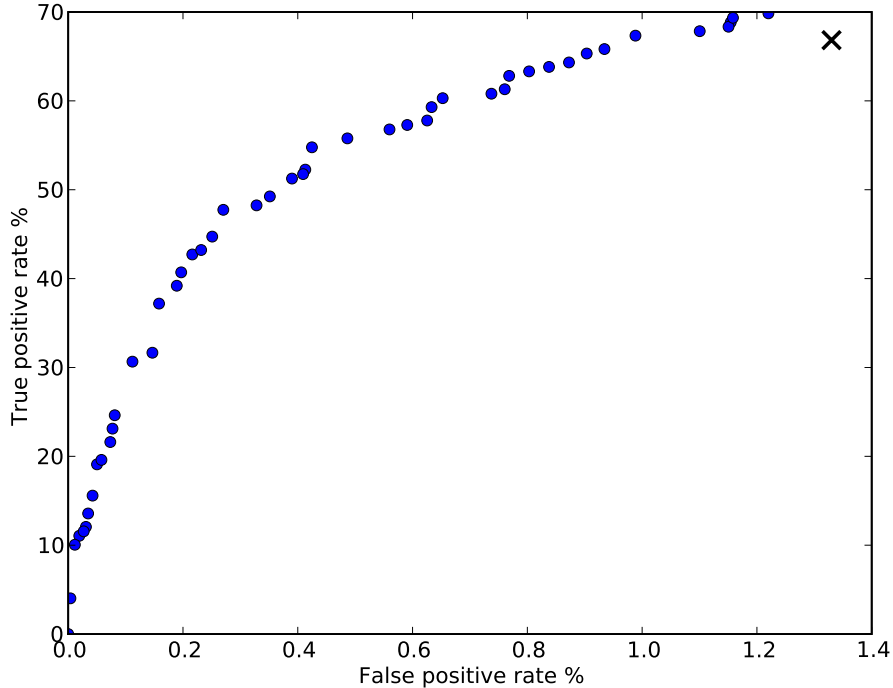


Figure 1: Conservative optimisation. Circles show estimates of the Pareto optimal ROC curve for STCA obtained after 12000 evaluations of the multi-objective optimiser. The cross indicates the manually tuned operating point θ^* .

random, but a scheme based on ‘uniselect’ method [Smith et al., 2008] was used here to promote exploration of the front. The selected parent vector is perturbed (by addition of a random number from a heavy-tailed distribution to aid exploration of parameter space and escape from local minima) to generate a single *child* (line 4). Each individual parameter in the parent vector is perturbed with equal probability.

The true $T(\theta')$ and false $F(\theta')$ positive rates for the perturbed vector are evaluated by running the STCA system with parameters θ' on a test database of track pairs. If the child θ' is not dominated by any of the parameter vectors in E , any parameter vectors in E that θ' dominates are deleted from the archive (line 7) and θ' is added to E (line 8). These two steps ensure that E is always a non-dominated set whose members dominate any other solution encountered thus far in the search.

3 Illustration

We illustrate the method by showing the results from a conservative optimisation of the STCA system for the Manchester Area Control Centre (MACC). It is conservative in that the ranges of parameters to be varied are limited to lie between the minimum and maximum that have previously been used by NATS so that the parameters are confined to regions of decision space with which personnel at NATS have considerable experience. Although we could adjust more parameters and adjust parameters over a greater range, the strategy adopted here provides an assurance that the optimised system is still operating within the

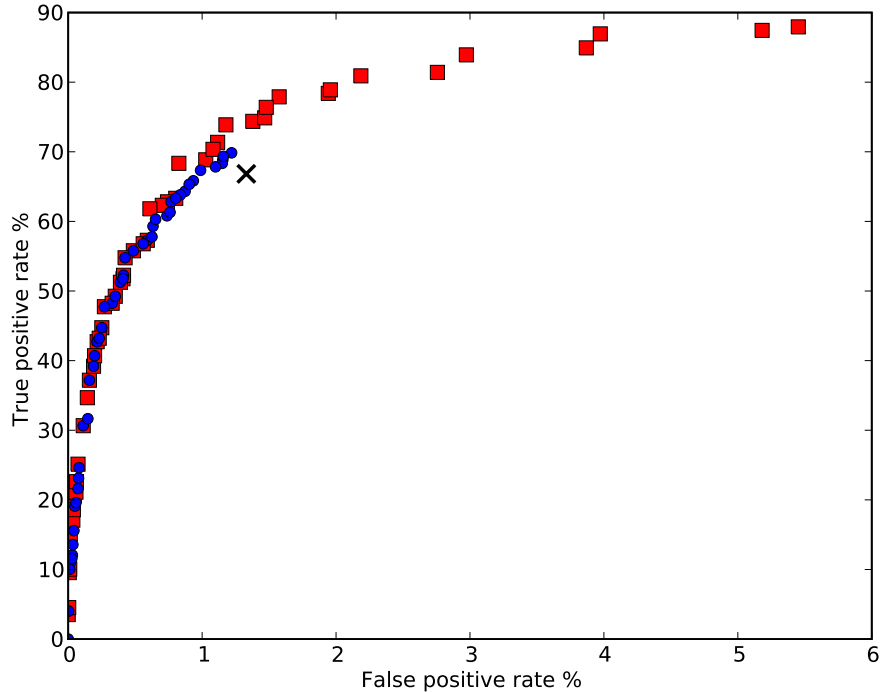


Figure 2: Aggressive optimisation for the data used in Figure 1. Red squares show the estimates of the Pareto front using aggressive optimisation, while the blue circles show the Pareto front from a conservative optimisation (Figure 1). The cross indicates the manually tuned operating point θ^* .

usual parameter ranges.

We optimised the true and false positive rates for a database comprised of manually and semi-automatically categorised encounters. The database included historical track pairs taken from the first two weeks of June 2007. This amounts to roughly 50,000 track pairs which, for computational reasons, is approximately a third the number of pairs generally used for manual tuning.

Even this conservative optimisation produces some striking results. Figure 1 shows the estimates of the Pareto optimal ROC curve obtained using the multi-objective optimiser after $N = 12000$ evaluations. The current NATS operating point is also plotted as a cross. The optimisation has located an ROC curve consisting of 48 points with true positive rates up to 70% and false positive rates up to 1.16%. In addition the manually tuned STCA operating point θ^* lies *behind* (is dominated by) operating points on the estimated ROC curve. Although the improvement over θ^* is relatively small in percentage terms, the quantity of track pairs processed by the STCA system means that a significant reduction in the *number* of false alerts could be achieved while maintaining the current genuine alert rate. We remark that this optimisation was initialised from 100 randomly chosen parameterisations, and including the current NATS operating point in the initialisation can result in both a more rapid convergence and an estimated Pareto front that extends to higher false positive rates than those found here; see, for example, additional results in [Everson and Fieldsend, 2006, Reckhouse et al., 2008].

More important, however, than the improvement in true and false positive rates is the production of the optimal ROC curve itself, because it reveals the true positive versus false positive trade-off, permitting the operating point to be chosen with a knowledge of the range of best possible trade-offs.

Figure 2 shows the result of an aggressive optimisation of the same data. (Initialised from 100 random parameterisations and the conservative front in Figure 1.) This optimisation is dubbed ‘aggressive’ because, in contrast to the conservative optimisation, we did not constrain the parameters to lie within ranges already used when manually tuning STCA. As Figure 2 shows, the additional freedom allowed to the parameters in the aggressive case permits an optimal ROC curve to be found that dominates the conservatively optimised curve in all places, although at low TP and FP rates the conservative and aggressive fronts virtually coincide. The aggressive front extends to a FP rate of 14.3% and 93.5% TP rate, although for clarity at low TP and FP these are not shown in the figure. Although these parameter settings must be scrutinised by NATS personnel before operational use, these results show how our evolutionary procedure can automatically locate operating points with performance well beyond those found manually. In fact it may be observed that the current operating point θ^* is close to the corner of the Pareto optimal curve. Choosing an operating point to the left of the corner would result in a rapidly diminishing genuine alert rate for little gain in the nuisance alert rate; whereas operating points to the right of the corner provide small increases in the true positive rate at the expense of relatively large increases in the false positive rate.

4 Applications

Perhaps the principal use of our evolutionary algorithm is to optimise STCA performance for a particular airspace and, with the optimal ROC curve on hand, personnel may make a reasoned choice as to the false alert rate that must be tolerated to achieve a particular true positive alert rate, or vice versa. However the optimal ROC curve also has a number of other potential applications which we discuss.

4.1 Visualisation

We emphasise again that the optimal ROC curve for a particular STCA system and particular locale (airport and operating procedures for the airspace) provides a readily interpretable graph of the best possible trade-offs between true and false positive rates. This permits STCA operating parameters to be chosen with a full knowledge of the alternatives, something that was not possible before the production of the optimal ROC curve.

Furthermore, the ROC curve provides a means of illustrating to non-specialist personnel what effect a change in true positive rate has on false positive rate, in a way that is instantly comprehensible. This aids communication between personnel responsible for the maintenance of the safety net, users, such as air traffic controllers, and managers responsible for strategic decisions.

The optimisation of two objectives may also be straightforwardly extended to include a third objective, such as the warning time given before an alert is raised [Everson and Fieldsend, 2006]. This trade-off is conveniently displayed as a surface (rather than a curve), allowing an operating point trading-off three objectives to be selected.

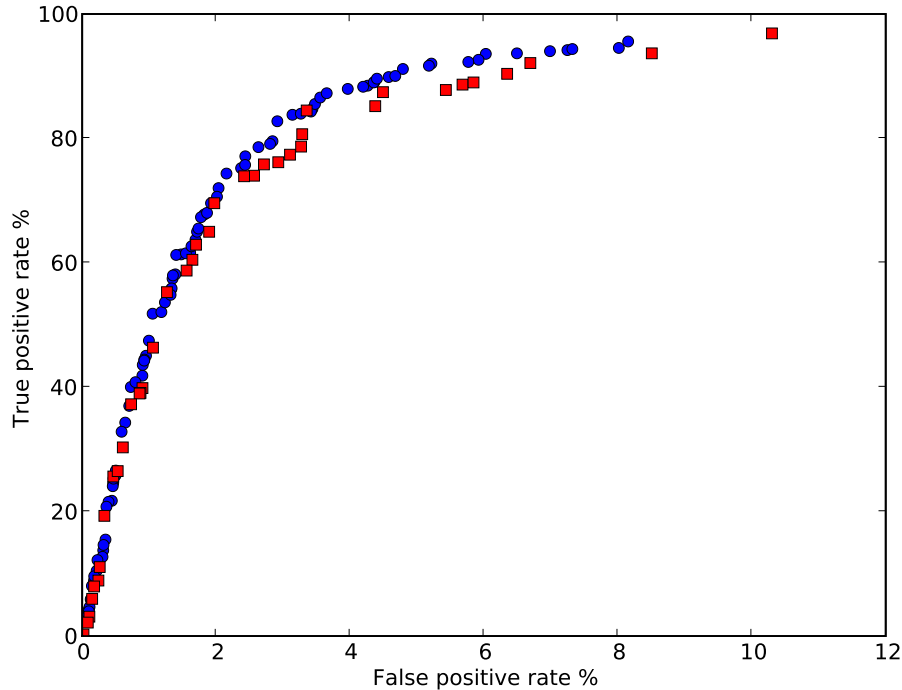


Figure 3: Comparison of Pareto optimal ROC curves for STCA (blue circles) and ESTCA (red squares).

4.2 Assessing STCA systems

It is important to be able to assess the changes, such as software upgrades, made to STCA systems themselves. The Pareto optimal ROC curve provides a tool for comparing STCA systems. By generating the optimal ROC curves for a pair of systems (using the same training track pairs) the performance of two systems may easily be assessed. Clearly, if one system's ROC curve dominates another's then the first is to be preferred.

As an illustration we display in Figure 3 the Pareto optimal ROC curves for STCA and Enhanced STCA (ESTCA). ESTCA contains a number of modifications and enhancements to the previous Issue 4 specification of NATS's STCA. Primarily ESTCA introduces a new fine filter to predict the lateral path of aircraft holding in stacks; and secondly it allows downlinked selected flight level information to be used to provide improved alerting performance. There are also smaller changes to existing STCA logic. As a result there are approximately 3800 parameters that may be optimised, in contrast to about 1400 for STCA. Here we used approximately 73,000 track pairs from the London Terminal Control Centre (LTCC) during July 2006. In order to increase the number of wanted alerts in the dataset, a fortnight's worth of track pairs for which an alert is wanted were combined with a weeks pairs for which an alert should not be raised.

The front for STCA (blue circles) is well converged, however the large number of parameters in ESTCA means that after $\approx 16,000$ iterations its front is still not yet converged and comparison of the two systems is not yet possible. It is to be expected that ESTCA, which is more flexible than STCA, should be capable of yielding a superior performance.

Although, on the basis of these data we are unable to quantitatively compare the systems' performance, these fronts do illustrate the importance of ensuring the systems are properly tuned before comparison.

We emphasise that the ROC curve gives a quantitative way of comparing STCA systems, but in a particular locale; operation under other circumstances may be different.

4.3 Assessing airspace reconfiguration

The ROC curve may also be used to monitor the effect of airspace reconfiguration. If the optimal ROC curve is known before the change it may be compared with the curve after the change. This permits STCA maintenance personnel to assess changes in safety-net performance secure in knowledge that any change observed is due to the airspace change rather than a sub-optimal configuration of the STCA system. However, it should be noted that it will not be possible to use the same data for optimisation before and after the airspace reconfiguration due to changes in traffic patterns resulting from the reconfiguration; consequently data sets for comparison must be carefully selected. Everson and Fieldsend [2006] describe a straightforward bootstrap procedure for estimating the variability in the ROC curve due to the size of the training data, which may be used to assess the magnitude of errors due to different data sets.

In addition, the particular region of airspace that yields changes in performance may be identified, possibly leading to further reconfigurations. The evolutionary algorithm may also be adapted to optimise objectives in addition to true and false positive rates; for example, alerts in a particular sector might be targeted, together with the average warning time given.

4.4 Airspace comparison

The Pareto ROC curve may also be used to compare airspaces. An airspace whose Pareto ROC curve is entirely dominated by the curve from another airspace is an airspace in which there are necessarily more false alerts for any given wanted alert rate. Although it is probably unrealistic to specify *a priori* the ROC curve for an airspace, the configuration of airspaces with dominated ROC curves may be candidates for early review.

5 Conclusion

We have briefly described a straightforward evolutionary algorithm for the simultaneous optimisation of the trade-off between genuine and nuisance alerts for STCA systems. The methodology is applicable to any STCA system and airspace and is currently being put into use at NATS as part of the parameter review and optimisation process.

As Europe moves towards standardisation of STCA systems and their operation, an exciting possibility is that the optimal ROC curve may serve as a measurement tool for comparing STCA systems and airspaces. It is well known from statistical pattern recognition that the area under the ROC curve is a measure of a classifier's ability to distinguish between two classes (here wanted and nuisance alerts) [Fawcett, 2006]. Although the total area under the Pareto ROC curve is likely to be less relevant than the area dominated by the region close to operational parameterisations, the optimal ROC curves of two STCA systems optimised on the same data provide a direct comparison of the efficacy of the two systems. We recognise that curves are only comparable for systems optimised on the same airspace using common data. Thus the lack of common data is likely to be the greatest

barrier to assessing systems in this way, but we look forward to a common standard for evaluating STCA systems over a range of operational conditions.

Acknowledgement

This work was supported by a Knowledge Transfer Partnership awarded to NATS and the University of Exeter.

References

- C.A. Coello Coello. A Comprehensive Survey of Evolutionary-Based Multiobjective Optimization Techniques. *Knowledge and Information Systems. An International Journal*, 1(3):269–308, 1999.
- K. Deb. *Multi-Objective Optimization Using Evolutionary Algorithms*. Wiley, Chichester, 2001.
- R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis*. Wiley, New York, 1973.
- EUROCONTROL. Guidance material for short term conflict alert. Eurocontrol, 2007a. Available from http://www.eurocontrol.int/safety-nets/public/standard_page/stca_01.html.
- EUROCONTROL. Specification for short term conflict alert (updated: Edition 1.0), 22 November 2007b. Available as http://www.eurocontrol.int/safety-nets/gallery/content/public/es_STCA-10.pdf.
- R.M. Everson and J.E. Fieldsend. Multi-objective optimisation of safety related systems: An application to short term conflict alert. *IEEE Transactions on Evolutionary Computation*, 10(2):187–198, 2006.
- T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27:861–874, 2006.
- W.J. Reckhouse, R.M. Everson, J.E. Fieldsend, R. Hayward, and T. Arnold. Multi-objective optimisation of Short Term Conflict Alert Systems. *Engineering Optimisation*, 2008. (In preparation).
- K.I. Smith, R.M. Everson, and J.E. Fieldsend. Simulated annealing and greedy search for multi-objective optimisation. *IEEE Transactions on Evolutionary Computation*, 2008. (Accepted).
- D. Van Veldhuizen and G. Lamont. Multiobjective Evolutionary Algorithms: Analyzing the State-of-the-Art. *Evolutionary Computation*, 8(2):125–147, 2000.
- E. Zitzler, L. Thiele, M. Laumanns, C.M. Fonseca, and V. Grunert da Fonseca. Performance Assessment of Multiobjective Optimizers: An Analysis and Review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, 2003.