8-2008

An integrated decision– making approach for improving European air traffic management

Yael GRUKSHA-COCKAYNE

Bert DE REYCK
Singapore Management University, bdereyck@smu.edu.sg

Zeger DEGRAEVE

MANAGEMENT SCIENCE
Vol. 54, No. 8, August 2008, pp. 1395–1409
ISSN 0025-1909 | EISSN 1526-5501 | 08 | 5408 | 1395
DOI 10.1287/mnsc.1080.0878

An Integrated Decision-Making Approach for Improving European Air Traffic Management

Yael Grushka-Cockayne
Department of Management Science and Operations, London Business School,
London NW1 4SA, United Kingdom, ygrushka.phd2003@london.edu

Bert De Reyck
Department of Management Science and Innovation, University College London,
London WC1E 6BT, United Kingdom, and
Department of Management Science and Operations, London Business School,
London NW1 4SA, United Kingdom, bdereyck@london.edu

Zeger Degraeve
Department of Management Science and Operations, London Business School,
London NW1 4SA, United Kingdom, zdegraeve@london.edu

We develop a multistakeholder, multicriteria decision-making framework for Eurocontrol, the European air traffic management organization, for evaluating and selecting operational improvements to the air traffic management system. The selected set of improvements will form the master plan of the Single European Sky initiative for harmonizing air traffic, in an effort to cope with the forecasted increase in air traffic, while maintaining safety, protecting the environment, and improving predictability and efficiency. The challenge is to select the set of enhancements such that the required performance targets are met and all key stakeholders are committed to the decisions. In this paper, we develop and implement a model to identify a preferred set of improvements to the arrival and departure procedures to and from airports. We provide an integrated approach for valuing a large number of alternatives, while considering interactions among them. The model combines quantitative and qualitative expert assessments of the possible enhancements and identifies commonalities and differences in the stakeholders' perspectives, ultimately recommending a preferred course of action. The model is currently being adopted by Eurocontrol as the formal trade-off analysis methodology supporting all enhancements' decision-making discussions throughout the construction of the master plan.

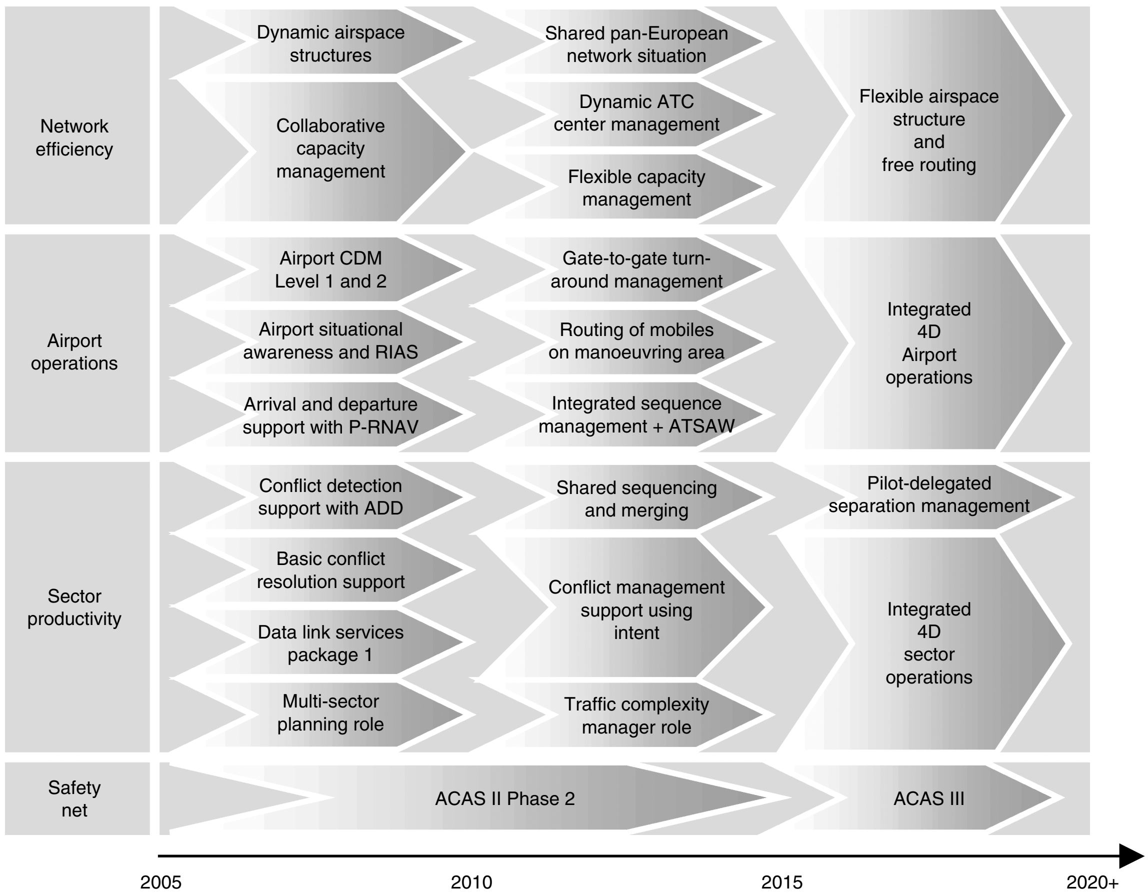
Key words: government; agencies; transportation; environment; research and development; project selection; decision analysis; multiple criteria
History: Accepted by Linda V. Green, public sector applications; received December 7, 2006. This paper was with the authors 7 months for 2 revisions. Published online in *Articles in Advance* June 20, 2008.

1. Introduction

Airport and air traffic network capacity is becoming an increasingly constraining factor in European air transportation, resulting in delays costing airlines between €1.3 and €1.9 billion per year (European Commission 2005). Forecasts predicting a doubling of air traffic within 15 years will result in an estimated 3.7 million unaccommodated flights per year. Therefore, Eurocontrol, the European air traffic management (ATM) organization, is developing a seamless pan-European ATM system, capable of coping with the forecasted growth in air traffic while maintaining a high level of safety, reducing costs, and preserving the environment. In spite of much effort to modernize and streamline European ATM systems, the current system still consists of heterogeneous elements that are based on national practices. Eurocontrol's Single European Sky (SES) initiative aims to lay the foundations of a unified system, eliminating borders in the sky. The SES master plan will cover the technological, economic, and regulatory aspects and will synchronize the implementation of new equipment.

Several harmonization and improvement activities have been proposed to enhance the performance of the European ATM system. These improvement initiatives are grouped into operational improvement (OI) clusters, which describe the operational, technical, and institutional improvements to be applied. The European ATM Strategic Roadmap of OI clusters, seen in Figure 1, describes these strategic changes, foreseen for 2005–2020. Each of the arrows in Figure 1 represents an OI cluster, and the clusters are vertically grouped by category (Network Efficiency, Airport Operations, etc.). The clusters are also grouped, horizontally, in chronological order. Beginning in §3 we will focus our analysis on the near-term cluster, *arrival and departure support with precision area navigation*.

Figure 1 European ATM Strategic Roadmap of OI Clusters



Each of the OI *clusters* consists of a set of candidate projects, hereafter referred to as *components*. Eurocontrol must decide which components to include in each OI cluster, taking into account performance targets that need to be met. Also, any proposed improvement projects, put forward as part of the SES master plan, have to conform to regulators' plans and be agreed upon by all aviation stakeholders that are likely to be affected, including airlines, air navigation service providers (ANSP), aircraft manufacturers, and airports.

The decisions concerning which components to select for each of the OI clusters are complicated by several factors. First, the selection of specific components results in trade-offs among the objectives, referred to as *key performance areas*. Second, interactions exist among the components as the inclusion of one component might affect the impact delivered by others. A third complicating factor is that each component can be implemented in one of several ways, each characterized by a different cost and expected performance impact. Fourth, the different stakeholders may associate a different priority to each key performance area and also disagree on the expected costs and performance benefits of each component. Fifth, several key performance areas require qualitative assessments, because they cannot easily be expressed in quantitative measures, or because the required information is not disclosed by the stakeholders. Finally, performance targets must be taken into consideration. The targets set by the Roadmap include (a) a threefold increase in capacity, (b) an improvement of safety performance by a factor of 10, (c) a 10% reduction in environmental effects, and (d) a reduction of ATM services costs by 50%.

In this study, we provide Eurocontrol with an integrated framework and model to support the strategic decision-making problem of selecting sustainable ATM system enhancements. By *integrated* we mean that we combine methodologies such as problem-structuring methods, qualitative and quantitative multicriteria decision analysis (MCDA) techniques, and large-scale optimization tools in a single decision-making framework. By *integrated* we also refer to the fact that within this framework we group multiple stakeholder views with multiple objectives, making decisions on a combined portfolio of possible technological enhancements, rather than evaluating technologies individually and separately for each stakeholder. We propose methods for reaching an overall consensus among all stakeholders and make recommendations on a compromise solution for a specific OI cluster. We applied the model between September 2005 and September 2006, analyzing the improvements suggested to the management of arrival and departure of aircraft to and from airports.

The remainder of the paper is organized as follows. A review of related literature is provided in §2. Section 3 describes in detail a specific OI cluster selection problem and the components that are considered. In §4 we outline the problem formulation and methodology for evaluating and selecting the possible system enhancements. Subsequently, §5 reports the main results of the study. Section 6 describes the implementation of the model and concludes.

2. Related Work and Contributions

The decisions concerning which components to select for each OI cluster are multistakeholder, multicriteria decision problems. In these problems, trade-offs need to be made between conflicting objectives and stakeholders, who put forward different objective hierarchies, priorities, and component evaluations. MCDA methods have been introduced as a means for solving decision problems with multiple objectives. Belton and Stewart (2002) provide an extensive review of existing MCDA methods. We also refer the reader to Figueira et al. (2005) for a recent overview of state-of-the-art methods and applications.

Since the introduction of MCDA, there has been an increasing demand for developing integrated frameworks. Belton and Stewart (2002) highlight that the application of integrated solutions is essential to the growth and success of MCDA. They encourage the integration of MCDA methods within a broader framework of problem structuring and organizational intervention, and the application of such solutions in complex real-world problems. Our work supports this call, expanding the implementation of integrated methods within a real-world setting, specifically in the aerospace domain, extending the work of others in this area, e.g., Parnell et al. (1998).

Three streams of literature are related to our work. The first explores the use of MCDA methods when trying to reach a consensus among multiple stakeholders. Butterworth (1989) represents the different stakeholders as high-level attribute categories in a hierarchical value tree. Similarly, Merrick et al. (2005) use weights to trade-off among stakeholders. Stirling and Mayer (1999) present the multicriteria mapping (MCM) approach to identify robust alternatives that are not necessarily optimal for any group of stakeholders but are broadly acceptable to most, by avoiding a premature focus of the analysis around a single perspective. Beck and Lin (1983) suggest using the maximize-agreement heuristic (MAH) to form a consensus ordering that reflects a collective agreement of alternative ranking. Tavana (2003) applies this heuristic within an MCDA framework, allowing for different departments within NASA to reach an agreement as to the prioritization of projects. Bana e Costa et al. (2000) use the MACBETH (measuring attractiveness by a categorical based evaluation technique) approach, an interactive questioning procedure that compares two elements at a time, to resolve a long-lasting construction conflict in Lisbon. Reaching consensus and buy-in is a primary challenge for Eurocontrol. Our work contributes to this domain by integrating these techniques and offering a framework in which each stakeholder's perspective is fully represented, as described in §4, and generating an efficient subset of compromise solutions, as we demonstrate in §5.

In practice, the comparative evaluation of alternatives, which occurs in most of the MCDA methods, can only be applied to a relatively small number of discrete options. However, in certain situations, a large number of alternatives may exist. Gershon et al. (1982) demonstrate that, by considering all combinations of a few discrete options, a large set of alternatives might exist. Methods for tackling this issue constitute the second stream of literature related to ours. The multiobjective design problem (Belton and Stewart 2002) requires specific techniques for screening and reducing the dimensionality of the problem. When the number of combinations is large, Stewart and Scott (1995) demonstrate how interactive MCDA procedures can reduce the number of alternatives. The elicitation approach and innovative heuristics we present in §4.3 for assessing the performance of combinations of components that can be selected, adds to this literature and serves to further combine multiobjective programming with MCDA techniques (Ehrgott and Wiecek 2005).

Finally, the need to consider interactions among the components links our work to a third stream of literature, namely that of project portfolio optimization. Difficulties in understanding the nature of the interactions among projects within the same portfolio have often limited the practical use of theoretical portfolio optimization models (Shane and Ulrich 2004). Fox et al. (1984) distinguish among different types of interactions: (1) cost or resource utilization; (2) outcome, probability, or technical; and (3) benefit, payoff, or effect. Here, we are concerned with the third type. Fox et al. (1984) propose a mathematical programming modeling framework in which interactions among projects can be assessed indirectly, by modeling the projects' impacts on profit. Santhanam and Kyparisis (1996) consider interdependencies among more than two projects and develop a nonlinear polynomial programming model. However, in their model, all benefit interdependencies are represented by monetary values. Dickinson et al. (2001) use a dependency matrix to quantify the interdependencies between development projects at Boeing and propose a multicriteria nonlinear optimization model for selecting an optimal portfolio. More recently, Medaglia et al. (2007) suggest a multiobjective evolutionary approach, applicable to project selection problems with partially funded projects, multiple objectives, project interdependencies, and constrained resources. We continue this line of research by suggesting ways to model interactions among alternative projects, and by developing a practical heuristic approach, further promoting the use of theoretical models in practice. Chien (2002) highlights several difficulties in developing practical portfolio selection methods including (1) acknowledging a difference between portfolio objectives and project objectives, (2) inadequate treatment of the interrelationships among alternatives, and (3) the number of assessments required for exploring all possible portfolios. In our work, we address the second and third concern, adequately representing interrelationships among alternatives and reducing the number of assessments required.

Our contributions are fourfold. First, we demonstrate how decision analysis models can be applied for high-impact, strategic, global decisions, promoting the recognition of issues such as the environment and safety. Second, we extend the use of MCDA models in situations in which a single course of action needs to be agreed upon by multiple stakeholders, who have conflicting objectives and different perspectives. Third, our model combines large-scale optimization methods like integer programming using branch and price with MCDA methods such as pairwise comparisons. Thus, we advocate the use of an integrated MCDA model, including problem-structuring methods, combining qualitative and quantitative assessment methods, and allowing for interactions among the alternatives. In this way, we also contribute by suggesting an innovative approach for applying decision analysis techniques to project portfolio problems. Finally, our fourth contribution is providing an elicitation technique for solving multicriteria problems with many alternatives.

3. Airport Arrival and Departure

In this paper, we focus on Eurocontrol's CL-03-02 cluster: *arrival and departure support with precision area navigation*. This cluster proposes changes to the arrival and departure processes at airports to improve the safety, capacity, and efficiency of terminal area airspace operations. The main stakeholders involved are airlines, airports, and ANSP. Naturally, other stakeholders will also be affected by the proposed changes, including the communities around airports and aircraft and equipment manufacturers. The following five components are being considered as part of the CL-03-02 cluster initiative:

Precision area navigation (P-RNAV) offers the ability to define routes for the onboard flight information management system that best meet the needs of the airport, the air traffic controller, and the pilot. This will result in shorter, more-direct routes and improved route adherence and predictability. When environmental issues play a major role, the route can be designed to bypass densely populated areas. Although all stakeholders believe that there are clear environmental and safety benefits from implementing P-RNAV, the larger airports fear a negative impact on capacity because they anticipate a reduction in the number of holding aircrafts and thus a possible reduction of runway utilization.

Arrival and departure manager (AMAN and DMAN) assist the air traffic controller by recommending the optimal arrival and departure sequence. Airports hope that its implementation will result in increased capacity, whereas airlines anticipate a better prediction of departure and arrival times.

Wake vortex (WV). An aircraft encountering the wake vortex of a preceding aircraft may lose control or suffer structural damages. To avoid wake vortex encounters, air traffic control regulations prescribe longitudinal separation minima between aircraft that are considered conservative. According to recent research, considerable operational improvements can be made through a more accurate knowledge of wake vortex behavior, which could yield important capacity gains by allowing planes to fly closer to each other.

Time-based separation (TBS). If separation between aircraft is set according to time intervals instead of distance, this may allow, under strong headwind conditions, for an increase in airport capacity while maintaining safety levels. TBS can be implemented with fixed or variable time intervals.

Basic continuous descent approach (B-CDA), or vectored CDA, reduces noise and fuel consumption by keeping aircraft at higher levels longer than conventional techniques, thus eliminating level segments and associated thrust transients (Reynolds et al. 2005). B-CDA will produce environmental and cost benefits. Airports, however, are concerned that B-CDA might have a negative impact on capacity.

4. Problem Description and Formulation

4.1. The Problem

We now present key definitions and a formulation for the OI cluster component selection problem. In the CL-03-02 cluster presented in the previous section, the five components under consideration are P-RNAV,

Table 1 Components and Implementation Modes Considered in the CL-03-02 Cluster

Components	Implementation modes				
P-RNAV	**A0** Stay with current	**A1** Implement during departure and arrival			
AMAN and DMAN	**B0** Stay with current	**B1** AMAN	**B2** DMAN	**B3** B1 and B2	**B4** B3 and sequencing tool
WV	**C0** Stay with current	**C1** Reduced WV minima	**C2** Improved onboard WV visualization	**C3** C1 & C2	**C4** Alternative procedures
TBS	**D0** Stay with current	**D1** Fixed-minimum TBS	**D2** Varying TBS		
B-CDA	**E0** Stay with current	**E1** Implement			

AMAN and DMAN, WV, TBS, and B-CDA. Each of the components can be implemented in one of several mutually exclusive *implementation modes*, which describe the operational characteristics of the implementation of each component. Essentially, the implementation modes entail different technologies, each characterized by a cost and an effect on each of the key performance areas, including capacity, efficiency, environment, predictability, and safety. The cost of each technology may differ across the stakeholders. Note that the terms *cluster*, *component*, *implementation modes*, and *key performance areas* correspond to *portfolio*, *project*, *options*, and *objectives* in the standard MCDA literature, respectively. The implementation modes considered for the five components in the CL-03-02 cluster are detailed in Table 1.

To allow for the key performance areas to be measurable, key performance *indicators* have been identified. For instance, capacity can be measured by the number of unaccommodated flights. In our study, the performance of the system is measured by 14 performance indicators, as detailed in Table 2. Some indicators use qualitative scales, e.g., the level of compliance with environmental rules and constraints, and others are quantitative, e.g., cost in Euros, or the number of people affected by noise pollution. In some cases, qualitative scales were explored because there were disagreements among the stakeholders as to appropriate quantitative measures. When defining the list of key performance areas and indicators, we were guided by Clemen and Reilly (2001), who highlight that the selected objectives and criteria should consist of different and decomposable objectives, such that the stakeholders are able to consider each separately, without having to consider the other objectives. This requirement is also referred to as *preferential independence* (Merrick et al. 2005). Note that this concept is less restrictive than statistical independence. Hence, two criteria can be statistically correlated, yet be preference independent.

Disagreement might exist concerning the anticipated performance impact of a certain implementation mode, thus resulting in different parameters for each stakeholder. To ensure that the performance impact is comparable across the various indicators, we use a value function (Keeney and Raiffa 1976), a mathematical representation of preferences. This function represents the stakeholders' preferences for each indicator, on a common scale, reflecting the relative importance of achieving different performance levels for each indicator.

At the heart of the problem are potential stakeholder disagreements regarding the importance of the different objectives. Therefore, we assign a weight to each indicator per stakeholder. Multiplied by the values assigned by the value function and summed over all performance indicators, the weights allow us to derive an overall performance score and rank for each combination of components, according to each stakeholder's perspective. Such linear additive models are widely employed to assess the overall value of alternatives when multiple, mutually preferentially independent criteria are to be taken into account (Figueira et al. 2005).

We assume that each stakeholder wants to maximize its perceived performance improvement to the ATM system, achieved by the selected component implementation mode combination. Due to the different expectations of cost and performance impact, a combination that is optimal for one stakeholder might not be optimal for another. In support of the SES master plan harmonization activities, Eurocontrol has as its prime objective to reach a joint commitment to action by all stakeholders. As such, Eurocontrol's overall objective is to maximize the anticipated performance score over all the stakeholders. This can be achieved by maximizing a weighted average of the anticipated performance scores, where the weights, determined by a steering committee headed by Eurocontrol, are used to represent trade-offs among stakeholders. When political concerns make it difficult to assign weights to

Table 2 List of Key Performance Areas and Performance Indicators

Index	Key performance areas and indicators	Preference	Measure
1	*Capacity*		
1.1	Exploit resources so as to maximize use of existing inherent capacity	Increasing	Qualitative
1.2	Increase capacity to meet projected traffic growth		
1.2.1	Unaccommodated demand	Decreasing	Flights per year
2	*Cost effectiveness*		
2.1	Cost-effective improvements of ATM services		
2.1.1	Cost of the ATM investment	Decreasing	Cost in million Euros
3	*Efficiency*		
3.1	On-time flight operations (reduction of departure and arrival delays)		
3.1.1	On-time gate arrivals	Increasing	% flights arriving within 15 minutes of schedule
3.1.2	On-time gate departures	Increasing	% flights departing within 15 minutes of schedule
3.2	User-preferred routes (reduction of excess flight distance)		
3.2.1	Average holding time	Decreasing	Minutes per flight
4	*Environment*		
4.1	Limit or reduce gaseous emissions on the ground and in the air		
4.1.1	Local air quality	Increasing	Qualitative
4.1.2	Level of compliance with environmental rules and constraints	Increasing	Qualitative
4.2	Limit or reduce the effect of noise during departures and arrivals		
4.2.1	Noise exposure	Decreasing	Number of people affected by a certain level of noise
4.2.2	Person-Event Index (PEI)	Decreasing	Noise exposure * (Number of events beyond a noise threshold)
5	*Predictability*		
5.1	Improve predictability of departure and arrival time		
5.1.1	Predictability of arrival time	Decreasing	Delay variance (minutes per flight)
6	*Safety*		
6.1	Achieve the lowest possible accident rate and constantly improve safety		
6.1.1	Fatal accident rates	Decreasing	Number of accidents per 100,000 departures
6.1.2	Runway incursions	Decreasing	Number of events per million movements
6.1.3	Safety events with ATM/CNS as the primary cause	Decreasing	Number of events per million flights

the stakeholders, an alternative approach, discussed in §5.3, can be used.

4.2. Notation and Formulation

To formally describe the Eurocontrol cluster selection problem, we introduce the following notation:

S: set of stakeholders, index k;
P: set of performance indicators, index j;
O: set of components, index q;
$I(q)$: set of implementation modes for component q, index $m(q)$.

The parameters of our model are as follows:

$c^k_{q,\,m(q)}$: cost of implementation mode $m(q)$ of component q for stakeholder k, $k = 1, \ldots, |S|$, $q = 1, \ldots, |O|$, $m(q) = 1, \ldots, |I(q)|$;
$e^k_{q,\,m(q),\,j}$: anticipated performance impact of implementation mode $m(q)$ of component q on performance indicator j by stakeholder k, $k = 1, \ldots, |S|$, $j = 1, \ldots, |P|$, $q = 1, \ldots, |O|$, $m(q) = 1, \ldots, |I(q)|$;
w^k_j: weight of performance indicator j, as seen by stakeholder k, $k = 1, \ldots, |S|$, $j = 1, \ldots, |P|$, where $\sum_{j=1}^{|P|} w^k_j = 1$, $k = 1, \ldots, |S|$;
b^k: budget of stakeholder k, $k = 1, \ldots, |S|$;
r^k_j: performance target for performance indicator j for stakeholder k, $k = 1, \ldots, |S|$, $j = 1, \ldots, |P|$.

The performance targets for each stakeholder are either determined exogenously by European regulators such as the European Commission, or internally by the stakeholders themselves for strategic reasons. For instance, in spite of the prospected growth in air traffic, the current level of accident risk was set as a performance target for the future level of safety. Targets are typically set as a percentage of change from the status quo.

We define the following decision variables:

$x_{q,\,m(q)}$: $= 1$ if implementation mode $m(q)$ will be selected, $= 0$ otherwise, $q = 1, \ldots, |O|$, $m(q) = 1, \ldots, |I(q)|$.

For each combination of a specific implementation mode for each component, $(m(1), \ldots, m(|O|))$, we denote:

$h^k_j(e^k_{1,\,m(1),\,j}, \ldots, e^k_{|O|,\,m(|O|),\,j})$: total anticipated performance impact of combination $(m(1), \ldots, m(|O|))$ on indicator j according to stakeholder k, $k = 1, \ldots, |S|$, $j = 1, \ldots, |P|$;

$v_j^k(h_j^k(e^k_{1,m(1),j},\ldots,e^k_{|O|,m(|O|),j}))$: preference value of the total anticipated performance impact of combination $(m(1),\ldots,m(|O|))$ on indicator j according to stakeholder k, $k=1,\ldots,|S|$, $j=1,\ldots,|P|$;

λ^k: weight associated with stakeholder k, $k=1,\ldots,|S|$, where $\sum_{k=1}^{|P|}\lambda^k=1$.

The problem is modeled as follows:

(i) Maximize the overall performance score of the selected set of implementation modes, weighted by the stakeholders' importance weights:

$$\max_{\mathbf{x}} \sum_{m(1)=1}^{|I(1)|}\cdots\sum_{m(|O|)=1}^{|I(|O|)|}\sum_{k=1}^{|S|}\lambda^k\sum_{j=1}^{|P|}w_j^k \cdot v_j^k(h_j^k(e^k_{1,m(1),j},\ldots,e^k_{|O|,m(|O|),j}))\prod_{q=1}^{|O|}x_{q,m(q)}. \quad (1)$$

Subject to

(ii) Ensure that the performance targets are achieved:

$$\sum_{m(1)=1}^{|I(1)|}\cdots\sum_{m(|O|)=1}^{|I(|O|)|}h_j^k(e^k_{1,m(1),j},\ldots,e^k_{|O|,m(|O|),j})\prod_{q=1}^{|O|}x_{q,m(q)}\geq r_j^k \quad k=1,\ldots,|S|,\ j=1,\ldots,|P|. \quad (2)$$

(iii) Each stakeholder has a limited available budget, which must not be exceeded:

$$\sum_{q=1}^{|O|}\sum_{m(q)=1}^{|I(q)|}c^k_{q,m(q)}x_{q,m(q)}\leq b^k \quad k=1,\ldots,|S|. \quad (3)$$

(iv) Enforce the selection of a single implementation mode for each component:

$$\sum_{m(q)=1}^{|I(q)|}x_{q,m(q)}=1 \quad q=1,\ldots,|O|. \quad (4)$$

(v) Integrality constraints:

$$x_{q,m(q)}\in\{0,1\} \quad q=1,\ldots,|O|,\ m(q)=1,\ldots,|I(q)|. \quad (5)$$

Note that in (3) the costs of the implementation modes are assumed to be additive. This assumption was put forward by Eurocontrol, although it is likely to be conservative, because the simultaneous implementation of several components could probably result in savings. If data supporting this would become available, the model can be modified using the approach described in §4.3.

4.3. Discussion

A difficulty with the formulation above is the nonlinearity of the model caused by the multiplication of decision variables. This is necessary because the total anticipated performance impact of each combination of implementation modes, $h_j^k(e^k_{1,m(q),j},\ldots,e^k_{|O|,m(|O|),j})$, is not necessarily a linear function of the individual implementation modes' performance impact. Although independence among the components might be true for some indicators, interaction effects do exist among most components. Therefore, our model requires the assessment of the anticipated performance impact of each of the possible combinations of implementation modes on each of the performance indicators. The resulting data-gathering process is likely to be challenging and time consuming. In addition, some preferences are assessed by qualitative pairwise comparisons. Obtaining these assessments is an even more cumbersome task, as the number of comparisons required would be enormous. Although technically possible, it is unlikely that this can be done in a consistent manner (Dyer 1990).

In some circumstances all interactions can be explicitly assessed, e.g., when assessing the impact of combining several cancer chemotherapy treatments (Petrovski and McCall 2001). In our case, however, the dimensions of the problem would make this approach infeasible. Other approaches, such as financial models for portfolio optimization, have limited use in a project portfolio setting, because returns and risks of different technologies cannot be estimated using historical values (Solak et al. 2008). Therefore, the use of heuristics to estimate the interactions is warranted.

We estimate the performance of a combination of implementation modes based on the separate performance assessments of the individual implementation modes, obtained from stakeholder interviews, expert assessments, qualitative pairwise comparisons and existing documentation. In this way, we are able to reduce the required number of assessments dramatically. For some key performance areas, the components' impacts are independent, resulting in an additive total performance impact. Other indicators are antagonistic with a joint impact lower than the sum of the individual effects, as is the case for capacity enhancements. Alternatively, for several of the performance indicators, the joint impact is higher than the sum of the individual component effects. This is the case when synergies exist among the components, e.g. for improved predictability due to the simultaneous implementation of AMAN and P-RNAV. In addition, qualitative assessments must also be combined to represent the total performance impact of a set of components.

The following heuristics were developed in collaboration with Eurocontrol experts and validated in a joint workshop in which the estimates obtained were compared with direct assessments. We denote:

- P_a: set of additive quantitative performance indicators,
- P_-: set of antagonistic quantitative performance indicators,
- P_+: set of synergistic quantitative performance indicators,
- P_q: set of qualitative performance indicators,

where $P=P_a\cup P_+\cup P_-\cup P_q$.

Additive. If the impact of a combination of implementation modes on a key performance area is additive, the total performance impact for the cluster equals

$$h_j^k(e^k_{1,m(1),j},\ldots,e^k_{|O|,m(|O|),j})=\sum_{q=1}^{|O|}e^k_{q,m(q),j}\qquad j=1,\ldots,|P_a|,\ k=1,\ldots,|S|.$$

Antagonism. The underlying reason for an antagonistic effect is that the impact of a component is reduced due to the implementation and impact of other, already implemented components. This can be modeled by applying its effect only to the remaining value of a key performance area (Degraeve and Koopman 1998). We assume, without loss of generality, that the individual component effects are ordered as follows: $e^k_{1,m(1),j}\geq\cdots\geq e^k_{q,m(q),j}\geq\cdots\geq e^k_{|O|,m(|O|),j}$, with $e^k_{q,m(q),j}$ representing a percentage improvement achieved by component q in implementation mode $m(q)$, on performance indicator j according to stakeholder k. We use $h_j^{k,q}$ to represent $h_j^{k,q}(e^k_{1,m(1),j},\ldots,e^k_{q,m(q),j},e^k_{q+1,1,j},\ldots,e^k_{|O|,1,j})$, the total effect of implementing up to component q, where implementation mode 1 represents the status quo. We compute antagonistic effects as follows:

$h_j^{k,q}=h_j^{k,q-1}+((1-h_j^{k,q-1})^2e^k_{q,m(q),j})$: the total effect of implementing up to component q, $j=1,\ldots,|P_-|$, $k=1,\ldots,|S|$, $q=2,\ldots,|O|$;

$h_j^{k,1}=e^k_{1,m(1),j}$: the total effect of implementing only the component with the largest effect, $j=1,\ldots,|P_-|$, $k=1,\ldots,|S|$;

$h_j^k(e^k_{1,m(1),j},\ldots,e^k_{|O|,m(|O|),j})=h_j^{k,|O|}$: the total effect of implementing $|O|$ components $j=1,\ldots,|P_-|$, $k=1,\ldots,|S|$.

For instance, from the airports' perspective ($k=2$), a 25% improvement in capacity (indicator 1.2.1, index numbering as in Table 2, $j=1$) due to implementing DMAN ($q=2$, $m(2)=3$), a 15% improvement as a result of implementing Reduced WV minima ($q=3$, $m(3)=2$), and a 10% improvement as a result of implementing fixed TBS ($q=4$, $m(4)=2$) result in an estimated total effect of $h_1^2(e^2_{2,3,1},e^2_{3,2,1},e^2_{4,2,1})=38\%$ as follows:

$$h_1^{2,2}=25\%,\quad h_1^{2,3}=0.25+((1-0.25)^2\times0.15)=33\%,$$
$$h_1^{2,4}=0.33+((1-0.33)^2\times0.10)=38\%.$$

Synergism. We model synergy effects as follows:

$$h_j^k(e^k_{1,m(1),j},\ldots,e^k_{|O|,m(|O|),j})=\prod_{q=1}^{|O|}(1+e^k_{q,m(q),j})-1\qquad j=1,\ldots,|P_+|,\ k=1,\ldots,|S|.$$

For instance, consider the anticipated effect of the system enhancements on on-time arrivals (indicator 3.1.1, $j=1$), again from the airports' perspective ($k=2$). The combined effect of implementing P-RNAV ($q=1$, $m(1)=2$), AMAN ($q=2$, $m(2)=2$), and varying TBS ($q=4$, $m(4)=3$), is expected to be synergistic. Therefore, if P-RNAV is expected to deliver a 25% improvement, AMAN a 15% improvement, and the varying TBS a 10% improvement, the net effect is estimated to be $h(e^2_{1,2,1},e^2_{2,2,1},e^2_{4,3,1})=(1.25\times1.15\times1.1)-1=58.125\%$.

Qualitative Multiplication (QM). The need for an overwhelmingly high number of pairwise comparisons when dealing with multiple alternatives has been the subject of much research (Dyer 1990). Saaty (1990) suggests *clustering*, i.e., grouping alternatives with respect to a common property and assessing the relative performance of each group. Millet and Harker (1990), building on Harker's incomplete pairwise comparison technique (Harker 1987), suggest reducing the effort through a more effective elicitation process with stopping rules when no additional data is necessary. Hotman (2005) suggests the base reference analytical hierarchy process (BR-AHP), which enhances the AHP by using comparisons to a single base-case alternative. However, applying these techniques in our model would still result in an excessive amount of pairwise comparisons. Instead, we estimate the total effect of a combination of components using the mathematical product of the individual, qualitative, implementation mode ratings on a ratio scale:

$$h_j^k(e^k_{1,m(1),j},\ldots,e^k_{|O|,m(|O|),j})=\prod_{q=1}^{|O|}e^k_{q,m(q),j}\qquad j=1,\ldots,|P_q|,\ k=1,\ldots,|S|.$$

For instance, consider the anticipated effect of the system enhancements on *Exploit resources so as to maximize use of existing inherent capacity* (indicator 1.1, $j=1$), again from the airports' perspective ($k=2$). The combined effect of implementing fixed TBS ($q=4$, $m(4)=2$), which when compared with other TBS implementation modes received a rating of 0.27, and implementing B-CDA, which when compared with other B-CDA implementation modes received a rating of 0.36, is estimated to be $h(e^2_{4,2,1},e^2_{5,2,1})=(0.27\times0.36)=0.0972$.

We have validated this heuristic by comparing the estimated effects with those assessed by direct pairwise comparisons solicited from Eurocontrol experts. The resulting expert rankings, once the pairwise comparisons were corrected for consistency, were almost identical to those generated by the qualitative multiplication heuristic.

4.4. Amended Formulation: Linear Model

Applying the heuristics described above, a linear model can be constructed by defining a different set of decision variables. Instead of selecting an implementation mode for each component independently, we now look at bundles of implementation modes, hereby referred to as *cluster versions*, whereby each cluster version defines a specific combination of implementation modes, one for each component. For instance, two possible cluster versions in our example would be: (i) do not invest in any component (A0, B0, C0, D0, E0) or (ii) invest in P-RNAV, AMAN, reduced WV minima, fixed TBS, and B-CDA (A1, B1, C1, D1, E1). Additional notation is necessary:

N: set of all cluster versions, index i;
tc_i^k: total cost of cluster version i to stakeholder k, $i=1,\dots,|N|$, $k=1,\dots,|S|$;
$te_{i,j}^k$: total anticipated performance impact of cluster version i on indicator j by stakeholder k, $i=1,\dots,|N|$, $k=1,\dots,|S|$, $j=1,\dots,|P|$. We denote the corresponding preference value by $v_j^k(te_{i,j}^k)$;
z_i: $=1$ if cluster version i is selected, $=0$ otherwise, $i=1,\dots,|N|$.

Based on the discussions with Eurocontrol experts, it was determined that the stakeholders' preferences can be approximated using linear value functions (Dyer et al. 1998):

$$v_j^k(te_{i,j}^k)=\frac{te_{i,j}^k-(te_{i,j}^k)_{\min}}{(te_{i,j}^k)_{\max}-(te_{i,j}^k)_{\min}},$$

although the model can also support nonlinear value functions. The cluster selection problem can now be restated as follows:

(i) Maximize the overall performance score of the selected cluster version:

$$\max_{\mathbf{z}}\sum_{i=1}^{|N|}\sum_{k=1}^{|S|}\lambda^k\sum_{j=1}^{|P|}w_j^k v_j^k(te_{i,j}^k)z_i. \tag{6}$$

Subject to

(ii) Ensure that the performance targets are achieved:

$$\sum_{i=1}^{|N|}te_{i,j}^k z_i\ge r_j^k \quad k=1,\dots,|S|,\ j=1,\dots,|P|. \tag{7}$$

(iii) Each stakeholder's limited available budget cannot be exceeded:

$$\sum_{i=1}^{|N|}tc_i^k z_i\le b^k \quad k=1,\dots,|S|. \tag{8}$$

(iv) The selection of a single cluster version:

$$\sum_{i=1}^{|N|}z_i=1. \tag{9}$$

(v) Integrality constraints:

$$z_i\in\{0,1\} \quad i=1,\dots,|N|. \tag{10}$$

4.5. Solution Method

The downside of the linear integer program (6)–(10) is that it contains an exponential number of decision variables. The more components and implementation modes are considered, the larger the set of cluster versions to choose from, resulting in problems of intractable size. In such cases a column generation approach can be used. A small subset of the cluster versions, serving as the columns, are collected in a master program whose dual prices are used by subproblems to evaluate new cluster versions. This process is continued until a solution to the LP relaxation of the master is found, upon which a branch-and-price procedure is used to find the optimal integer solution. Barnhart et al. (1998) provide a comprehensive description of the branch-and-price methodology and the required conditions for optimality (see Lübbecke and Desrosiers 2005 for a recent review). Vanderbeck (2000) offers an alternative approach to the column generation, based on the discretization of the integer polyhedron associated with a subsystem of constraints (as opposed to its convexification), and describes the corresponding branching framework. Degraeve (1992) suggests that adding a fixed number of potentially "good" columns initially will result in excellent integer solutions. Specifically for our case, the initial columns could consist of a fixed number of cluster versions for each stakeholder that are lowest in cost and achieve the desired performance targets.

We denote:

L^k: subset of cluster versions generated for stakeholder k, $k=1,\dots,|S|$;
$L_{q,m(q)}^k$: subset of cluster versions generated for stakeholder k, containing implementation mode m_q for component q, $k=1,\dots,|S|$, $q=1,\dots,|O|$, $m(q)=1,\dots,|I(q)|$;
y_i^k: $=1$ if cluster version i is selected by stakeholder k, $=0$ otherwise, $i=1,\dots,|N|$, $k=1,\dots,|S|$.

The new decision variables, y_i^k, represent the selection of a specific cluster version by each of the stakeholders and allow us to decompose the problem and construct k subproblems, one per stakeholder. The overall selection of a single cluster will be enforced in the master problem by means of coordination constraints. These coordination constraints are analogous to nonanticipativity constraints, commonly found in multistage stochastic problems. Additionally, the selected cluster versions in each of the subproblems will provide insight into potential disagreements between the stakeholders concerning the preferred course of action, as will be discussed in §5.

The master program can be formulated as:

(i) Maximize the overall performance score of the selected cluster version, across the stakeholders:

$$\max_{\mathbf{y}} \sum_{k=1}^{|S|} \lambda^k \sum_{i=1}^{|L^k|} \sum_{j=1}^{|P|} w_j^k v_j^k (te_{i,j}^k) y_i^k. \tag{11}$$

Subject to

(ii) Ensure that global performance targets are achieved by all stakeholders (dual price φ_j^k)

$$\sum_{i=1}^{|L^k|} te_{i,j}^k y_i^k \geq r_j \quad k=1,\ldots,|S|,\ j=1,\ldots,|P|. \tag{12}$$

(iii) Coordination (dual price $\mu_{q,m(q)}^{k,k+1}$)

$$\sum_{i=1}^{|L_{q,m(q)}^k|} y_i^k = \sum_{i=1}^{|L_{q,m(q)}^{k+1}|} y_i^{k+1} \quad k=1,\ldots,(|S|-1),$$

$$q=1,\ldots,|O|,\ m(q)=1,\ldots,|I(q)|. \tag{13}$$

(iv) The selection of a single cluster version per stakeholder (dual price π^k)

$$\sum_{i=1}^{|L^k|} y_i^k = 1 \quad k=1,\ldots,|S|. \tag{14}$$

(v) Nonnegativity:

$$y_i^k \geq 0 \quad i=1,\ldots,|N|,\ k=1,\ldots,|S|. \tag{15}$$

Meeting the stakeholders' performance targets and ensuring that the stakeholders' individual budgets are not exceeded are considered in the subproblems, with subproblem k being:

(i) Minimize the reduced cost:

$$\min_{\mathbf{y}} \sum_{i=1}^{|N|} \Bigg(-\sum_{j=1}^{|P|} w_j^k v_j^k (te_{i,j}^k) - \sum_{j=1}^{|P|} \varphi_j^k + \sum_{q=1}^{|O|} \sum_{m_q=1}^{|I(q)|} \mu_{q,m(q)}^{k-1,k} - \sum_{q=1}^{|O|} \sum_{m(q)=1}^{|I(q)|} \mu_{q,m(q)}^{k,k+1} - \pi^k \Bigg) y_i^k. \tag{16}$$

Subject to

(ii) Ensure that stakeholders' performance targets are achieved:

$$\sum_{i=1}^{|N|} te_{i,j}^k y_i^k \geq r_j^k \quad j=1,\ldots,|P|. \tag{17}$$

(iii) Each stakeholder's limited available budget cannot be exceeded:

$$\sum_{i=1}^{|N|} tc_i^k y_i^k \leq b^k. \tag{18}$$

(iv) The selection of a single cluster version:

$$\sum_{i=1}^{|N|} y_i^k = 1. \tag{19}$$

(v) Integrality constraints:

$$y_i^k \in \{0,1\} \quad i=1,\ldots,|N|. \tag{20}$$

5. Results

The model is implemented in Lingo 6.0 (Schrage 2000), with a user interface in Microsoft Excel using VBA. In the case of the CL-03-02 cluster, we find that there is a total of 300 possible cluster versions, making a direct solution approach feasible.

5.1. Objectives Weights

Figure 2 presents the weights for the key performance indicators obtained from experts' responses to a combination of swing (Edwards and Barron 1994) and pairwise comparisons questions. The weights were solicited after the ranges of consequences were determined (Keeney 2002). The importance of safety to all three stakeholders is evident (indicators 6.1–6.3). We can see that airports view capacity (indicators 1.1, 1.2.1) as highly important. However, predictability seems to be less important to them (indicator 5.1.1). Airlines, on the other hand, do care about efficiency and predictability (indicators 3.2.1, 5.1.1). ANSP place a higher weight on reducing the number of accidents in which ATM serves as the primary cause (indicators 6.1.3) than the other stakeholders do. It is apparent that the range of values assigned to the environment indicators is narrow, compared to the range of weights assigned to the capacity indicators. This information allows Eurocontrol to identify the main sources of disagreement as well as to identify commonalities among the stakeholders, information that has proven valuable during the joint discussions with all stakeholders.

To assess the nature of the interaction among the components in our project, we used a series of questions of the following style: *Consider the impact of implementing* C1 *and* D1 *in terms of capacity* (*indicator* 1.2.1): *increase capacity to meet projected traffic growth, measured by unaccommodated demand. Assume the following*:

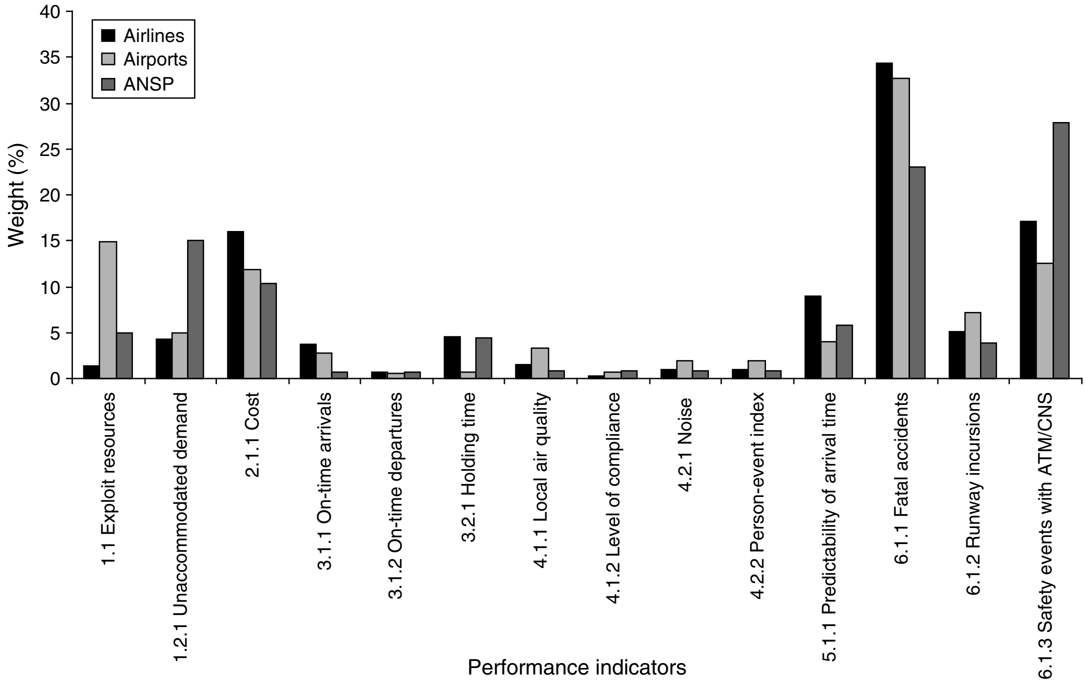
- *Status quo is* 1,000 *unaccommodated flights per year*.
- *Expected impact from* C1: 700 *unaccommodated flights per year, i.e., a* 30% *improvement*.
- *Expected impact from* D1: 800 *unaccommodated flights per year, i.e., a* 20% *improvement*.
- *What would be the overall impact of implementing them both?*

Several examples were considered, with varying parameters, to develop a heuristic to match the intuition of the experts. The heuristics were then validated in a workshop with all stakeholders. This approach led to a clear understanding and ownership of the heuristics by the members involved.

5.2. Recommended Cluster Version

Table 3 lists the top-ranking cluster versions according to each of the three stakeholders, when the model is solved using decision variables representing the selection of a specific cluster version by each of the

Figure 2 The Weights Assigned to the Performance Indicators, Representing the Importance of a Specific Performance Indicator to Each Stakeholder



stakeholders, y_i^k, without enforcing the coordination constraints (13). Note that the three top-scoring clusters include the implementation of P-RNAV (A1) and AMAN, DMAN, and a sequencing tool (B4), but the stakeholders seem to have different preferences concerning WV, TBS, and B-CDA. Also note that the implementation of P-RNAV (A1) is included in all clusters listed in Table 3.

The scatter plots in Figure 3 combine the perspectives of the different stakeholders, two by two. The efficient frontiers, appearing as dashed lines in each plot, highlight the cluster versions that are not dominated by any other. The search for the preferred cluster version can now focus on these options.

When solving our model using equal weights for the different stakeholders, cluster version 140 emerges as the recommended cluster version. It includes the implementation of P-RNAV (A1), AMAN, DMAN, and a sequencing tool (B4), reduced WV minima with the improved onboard WV visualization (C3), varying TBS (D2), but not B-CDA (E0). Recall that the main benefit and motivation in introducing the implementation of B-CDA is the environmental impact. Therefore, Eurocontrol and other regulating bodies may wish to focus on compensation related to B-CDA, if they wish to implement this component.

Table 3 Top-Ranking Cluster Versions According to the Individual Stakeholders

	Airlines		Airports		ANSP	
Rank	Cluster version	Components	Cluster version	Components	Cluster version	Components
1	290	A1 B4 C3 D2 E1	180	A1 B4 C2 D0 E1	140	A1 B4 C3 D2 E0
2	140	A1 B4 C3 D2 E0	230	A1 B4 C2 D1 E1	90	A1 B4 C3 D1 E0
3	240	A1 B4 C3 D1 E1	280	A1 B4 C2 D2 E1	290	A1 B4 C3 D2 E1
4	190	A1 B4 C3 D0 E1	178	A1 B3 C2 D0 E1	40	A1 B4 C3 D0 E0
5	90	A1 B4 C3 D1 E0	174	A1 B1 C2 D0 E1	138	A1 B3 C3 D2 E0
6	40	A1 B4 C3 D0 E0	130	A1 B4 C2 D2 E0	240	A1 B4 C3 D1 E1
7	288	A1 B3 C3 D2 E1	140	A1 B4 C3 D2 E0	134	A1 B1 C3 D2 E0

During the discussions of the results, Eurocontrol wished to ensure that no one stakeholder feels that their situation could be considerably improved if a different set of implementation modes was selected. While compensating stakeholders by transfer payments was experimented with by Eurocontrol, it was considered unsatisfactory, because stakeholders felt uncomfortable trading off nonmonetary objectives with cost transfers. The introduction of minimizing the maximum regret (French 1986) allows for including nonmonetary objectives and could thus generate solutions acceptable to all stakeholders. This can be modeled by

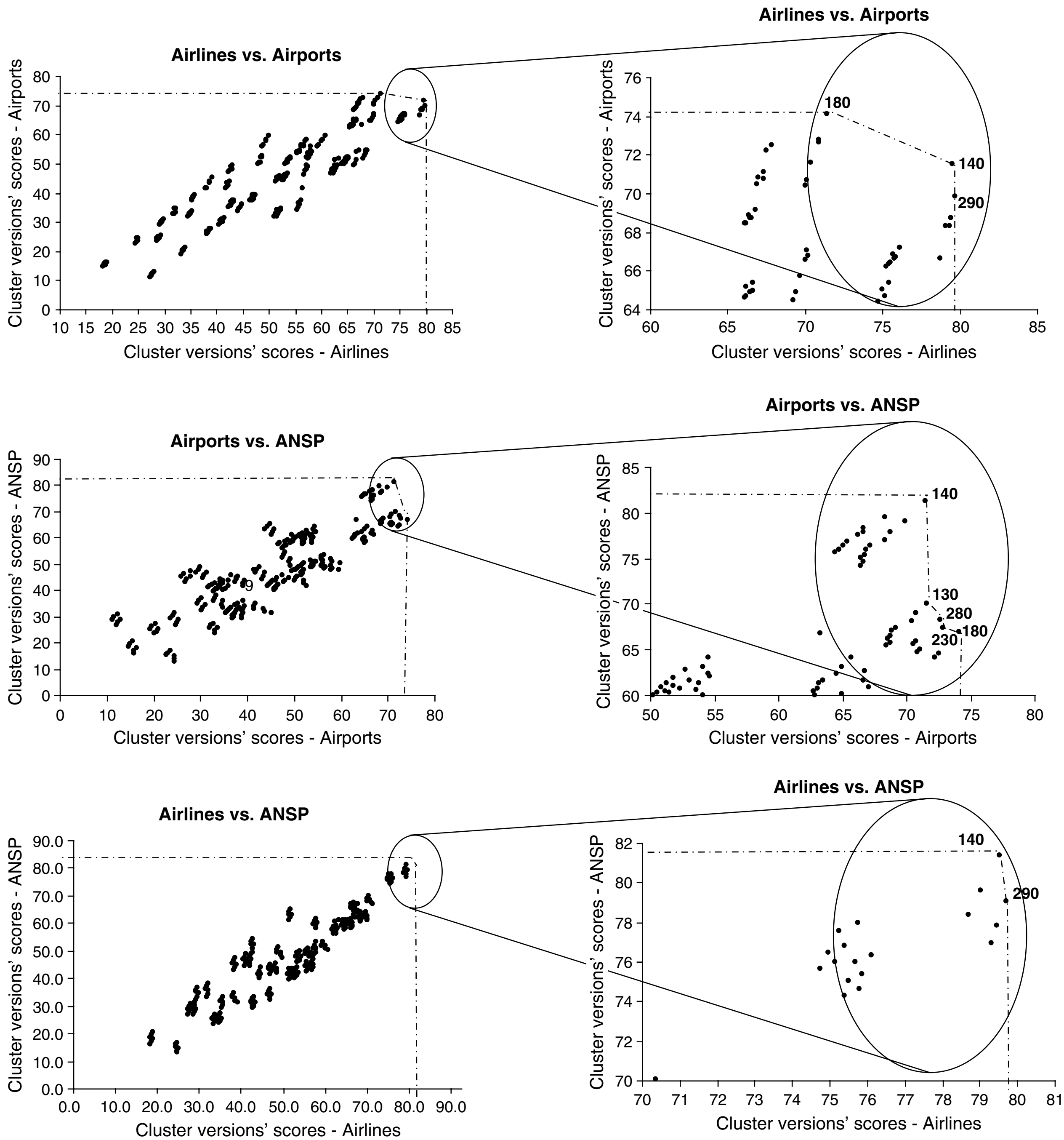
$r_i = \max_k(\max_i \sum_{j=1}^{|P|} w_j^k v_j^k(te_{i,j}^k) - \sum_{j=1}^{|P|} w_j^k v_j^k(te_{i,j}^k))$, the maximum regret across stakeholders if cluster i were selected. $i = 1, \ldots, |N|$.

The following objective function can then replace (6):
(i) Minimize maximum regret

$$\min_{\mathbf{z}} \sum_{i=1}^{|N|} r_i z_i.$$

Cluster version 140 is also the cluster version that minimizes the maximum regret.

Figure 3 Scatter Plots of Cluster Versions' Scores with Efficient Frontiers, According to Each Set of Two Stakeholders



Note. All axes represent cluster versions overall scores.

5.3. Sensitivity Analysis

To determine how robust the recommendations are, we carried out two types of sensitivity analyses. First, *within stakeholder* groups, we examined the sensitivity of the rankings of the cluster versions to the weights assigned by each stakeholder to each of the performance indicators, as well as to the performance impact assessments made by the stakeholders. Second, we explored the sensitivity *across stakeholders*, i.e., the robustness of the overall recommendations when the weights assigned to the stakeholders are varied.

Within Stakeholders. Many of the established MCDA methods focus on a series of one-dimensional sensitivity analyses to establish how robust model recommendations are. As highlighted by Butler et al. (1997), although these analyses do provide insights, they may be misleading because they ignore the potential interaction that can result from simultaneously manipulating multiple weights. Therefore, we use the random weights simulation technique suggested by Butler et al. (1997), in which weights assigned by the stakeholders to the performance indicators are randomly generated, according to a uniform distribution. From such an analysis we obtain a range of different rankings for each of the cluster versions. It is recommended that cluster versions that under certain combinations of weights rank first, should not be disregarded from the discussions because they might, under certain conditions, be preferred by one of the stakeholders. Twenty-six cluster versions qualified using this criterion, thus allowing for the more-detailed discussions to focus on this subset of cluster versions.

A simulation analysis allows us to provide recommendations in the absence of complete information regarding the expected performance impact of the implementation modes. For this purpose we use ranges to capture the anticipated impact of the components on the indicators, rather than single values. Also, we found that significant uncertainty exists due to lack of availability and disclosure of data by the stakeholders. Past experience of Eurocontrol has shown that when assessing the performance of system enhancements, some stakeholders prefer to provide ranges of expected costs and benefits in order to avoid the disclosure of strategic information. Running a simulation while allowing the performance impact ratings to randomly be drawn from the parameter ranges provided, according to a triangular distribution, we found that Cluster 140 remains the only common top-ranking cluster across the three stakeholder groups.

We also carried out a sensitivity analysis on the cost additivity assumption, exploring synergies of 10%, 20% and 50%, which revealed that there is no change in the top-ranking clusters.

Across Stakeholders. Due to political concerns, it is often difficult to assign weights to the stakeholders, as required for the calculation of the objective function value in our model. Using robust portfolio modeling methodology (RPM; Liesio et al. 2007), incomplete information concerning these weights can be used to exclude certain options that will never be preferred should more information become available. Such partial information could, for instance, specify that one stakeholder is more important than another, without the need for specifying exact weights.

We modify the RPM methodology for our problem setting as follows: we define the weights $\lambda = (\lambda^1, \ldots, \lambda^{|S|})$, such that

$$\boldsymbol{\lambda} \in S_\lambda^0 = \left\{ \boldsymbol{\lambda} \in \mathbb{R}^{|S|} \,\middle|\, \lambda^k \geq 0, \sum_{k=1}^{|S|} \lambda^k = 1 \right\}.$$

Incomplete information about the stakeholder weights is modeled by a set of feasible weights, denoted by $S_\lambda \subseteq S_\lambda^0$, a convex set of weight vectors constrained by a set of linear inequalities that correspond to Eurocontrol's statements regarding the stakeholders' relative importance. At the extremes, $S_\lambda = S_\lambda^0$ is the largest possible weight set, which corresponds to lack of any weight information, and when S_λ contains a single element, this corresponds to complete information.

Similarly, we define the set of feasible scores. Because incomplete information exists regarding the expected performance impact of the implementation modes, we obtain a range of scores for each cluster version, according to each stakeholder. Denoting the overall score of a cluster version by $\pi_i^k = \sum_{j=1}^{|P|} w_j^k v_j^k(te_{i,j}^k)$, $i = 1, \ldots, |N|$, and $k = 1, \ldots, |S|$, we denote the lower and upper bound of the score range $\underline{\pi}_i^k$ and $\overline{\pi}_i^k$ whereby $\underline{\pi}_i^k \leq \pi_i^k \leq \overline{\pi}_i^k$. The set of feasible scores is $S_\pi = \{\boldsymbol{\pi} \in \mathbb{R}^{n \times k} \mid \pi_i^k \in [\underline{\pi}_i^k, \overline{\pi}_i^k]\}$.

For a given cluster version, i, the selection of different feasible scores and stakeholder weights result in an interval for the overall cluster score such that for any $\boldsymbol{\lambda} \in S_\lambda$ and $\boldsymbol{\pi} \in S_\pi$,

$$\Pi(i, \boldsymbol{\lambda}, \boldsymbol{\pi}) \in \left[\min_{\boldsymbol{\lambda} \in S_\lambda} \sum_{k=1}^{|S|} \lambda^k \underline{\pi}_i^k, \max_{\boldsymbol{\lambda} \in S_\lambda} \sum_{k=1}^{|S|} \lambda^k \overline{\pi}_i^k \right].$$

Cluster version i dominates i' with regard to the information set $S = S_\lambda \times S_\pi$, denoted by $i \succ_s i'$, iff

$$\Pi(i, \boldsymbol{\lambda}, \boldsymbol{\pi}) \geq \Pi(i', \boldsymbol{\lambda}, \boldsymbol{\pi}) \quad \text{for all } (\boldsymbol{\lambda}, \boldsymbol{\pi}) \in S \quad \text{and}$$
$$\Pi(i, \boldsymbol{\lambda}, \boldsymbol{\pi}) > \Pi(i', \boldsymbol{\lambda}, \boldsymbol{\pi}) \quad \text{for at least one } (\boldsymbol{\lambda}, \boldsymbol{\pi}) \in S.$$

Without any assumptions on stakeholder importance, we identify 30 nondominated cluster versions with regard to the information set S, denoted $ND(S)$.

We define the *Core Index* (CI) of an implementation mode $m(q)$ as the proportion of cluster versions in the nondominated set that contain that implementation mode:

$$\mathrm{CI} = \frac{|\{i \in ND(S) | m(q) \in i\}|}{|ND(S)|}.$$

Implementation modes with $\mathrm{CI} = 1$ are robust choices in the sense that if additional information were to become available, they would definitely be recommended. In our case, A1 is such an implementation mode. *Exterior* implementation modes, where $\mathrm{CI} = 0$, can safely be rejected because they will never be included in the set of nondominated cluster versions when additional information becomes available. A0, B0, B2, C0, C1, and C4 are examples of exterior implementation modes. *Borderline* implementation modes with $0 < \mathrm{CI} < 1$, will require further analysis. B1, B3, B4, C2, C3, D0, D1, D2, E0, and E1 are borderline implementation modes.

6. Implementation and Impact

The model presented in this paper was implemented during September 2005 through September 2006.

Throughout the process we worked with a project team of 4 members and 11 additional experts from a diverse range of expertise within Eurocontrol. The framework we used was iterative, with looping back to previous stages and revising the model when necessary, until a requisite (Phillips 1984) representation of the problem situation was attained. A complicating factor in obtaining the requisite model in this project was the large number of components, as well as the large number of performance objectives and stakeholders. In order to gain the support of the stakeholders, the requisite framing of the problem was determined through joint workshops with representatives from all considered stakeholder groups.

The final decision concerning which cluster version to put forward for CL-03-02 is, at the time this paper is being written, being discussed within Eurocontrol. However, the focus is now on a subset of preferred cluster versions, especially 140 and 290, that scored highly in the model and performed well in all key sensitivity analyses. In addition, a compromise alternative, not originally identified, also emerged in the form of a possible partial implementation of B-CDA, balancing capacity and environment. As Goodwin and Weight (2004) emphasize, if at the end of the process no single best course of action has been identified, this does not suggest that the analysis was worthless. A good decision-making process is not necessarily characterized by reaching a final solution, but by the extent to which it has enhanced communication, developed a shared understanding of the problem, and achieved a joint commitment to action. Thus, a benefit of using the model was that, according to Peter Eriksen, Airport Research Area Manager at the Eurocontrol Experimental Centre, "this methodology has revealed combinations that we did not even think about." The analysis also exposed that although additivity among the components cannot be assumed, the interaction effects that are estimated to take place can be limited to four types, and these can be modeled. Finally, by exploring the source of nonlinearities, the stakeholders learned that, contrary to original estimates, the joint implementation of PRNAV and AMAN and DMAN is likely to be even more beneficial than originally thought in terms of increased efficiency, because the two technologies are highly synergistic.

Supported by Eurocontrol's Director General, this approach to decision making is currently being adopted by Eurocontrol as the formal trade-off methodology, supporting the highly visible European enhancement discussions throughout the construction of the Single European Sky master plan for the next-generation pan-European ATM system.

According to Robert Graham, Mid-Term Concept Validation Program Manager at the Eurocontrol Experimental Centre, the model has "improved Eurocontrol's understanding of how to bring together qualitative and quantitative components into a multicriteria decision frame while ensuring balance, clarity and equitable discussion between stakeholders, leading to implementation decisions involving significant European investment (several billions of Euros) over the next six years." Thus, the work described in this paper has made a positive contribution to Eurocontrol and to the European aviation community by successfully highlighting the benefits of using decision analysis techniques as part of the multistakeholder discussions. The methodology allows Eurocontrol to resume their visionary role as facilitators in the European aviation decision-making process and will assist Eurocontrol in structuring and formalizing the process and supporting the stakeholders in the assessment process.

The model presented, although developed specifically for the ATM discussions, is generalizable and applicable beyond the ATM domain. When multiple parties must reach a consensus on selecting among interdependent alternatives and decide on an acceptable set of actions, rules, policies or standards, our technique could be useful. The method is generalizable and flexible to support a multiparty multiobjective policy negotiation with many combinations of possible decisions.

Acknowledgments
The authors gratefully acknowledge Patricia Cauwenbergh, Peter Eriksen, Robert Graham, Henk Hof, and Paula Leal De Matos from Eurocontrol for their cooperation and assistance throughout the project. They also thank the associate editor and three anonymous referees for their constructive comments and suggestions for improving the paper.

References

Bana e Costa, C., F. Nunes da Silva, J. C. Vansnick. 2000. Conflict dissolution in the public sector: A case study. *Eur. J. Oper. Res.* **13** 388–401.

Barnhart, C., E. L. Johnson, G. L. Nemhauser, M. W. P. Savelsbergh, P. H. Vance. 1998. Branch-and-price: Column generation for solving huge integer programs. *Oper. Res.* **46**(3) 316–329.

Beck, M. P., B. Lin. 1983. Some heuristics for consensus ranking problem. *Comput. Oper. Res.* **10**(1) 1–7.

Belton, V., T. J. Stewart. 2002. *Multiple Criteria Decision Analysis: An Integrated Approach*. Kluwer Academic Publishers, Boston.

Butler, J., J. Jia, J. Dyer. 1997. Simulation techniques for sensitivity analysis of multi-criteria decision models. *Eur. J. Oper. Res.* **103** 531–546.

Butterworth, N. J. 1989. Giving up "The smoke": A major institution investigates alternatives to being sited in the city. *J. Oper. Res. Soc.* **40**(8) 711–717.

Chien, C. F. 2002. A portfolio-evaluation framework for selecting R&D projects. *R&D Management* **32**(4) 359–368.

Clemen, R. T., T. Reilly. 2001. *Making Hard Decisions*. Duxbury, Pacific Grove, CA.

Degraeve, Z. 1992. Scheduling joint product operations with proposal generation methods. Ph.D. dissertation, Graduate School of Business, The University of Chicago, Chicago.

Degraeve, Z., G. J. Koopman. 1998. Achieving air quality standards in the European union. *Oper. Res.* **46**(5) 643–654.

Dickinson, M. W., A. C. Thornton, S. Graves. 2001. Technology portfolio management: Optimizing interdependent projects over multiple time periods. *IEEE Trans. Engrg. Management* **48**(4) 518–527.

Dyer, J. 1990. Remarks on the analytic hierarchy process. *Management Sci.* **36**(3) 249–258.

Dyer, J. S., T. A. Edmunds, J. C. Butler, J. Jia. 1998. A multiattribute utility analysis of alternatives for the disposition of surplus weapons-grade plutonium. *Oper. Res.* **46**(6) 749–762.

Edwards, W., F. H. Barron. 1994. SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement. *Organ. Behav. Human Decision Processes* **60** 306–325.

Ehrgott, M., M. M. Wiecek. 2005. Multiobjective programming. J. Figueira, S. Greco, M. Ehrgott, eds. *Multiple Criteria Decision Analysis, State of the Art Surveys*, Chap. 17. Springer's International Series, New York.

European Commission. 2005. The single European sky. Retrieved February 1, 2008, http://ec.europa.eu/transport/air_portal/.

Figueira, J., S. Greco, M. Ehrgott, eds. 2005. *Multiple Criteria Decision Analysis, State of the Art Surveys*. Springer's International Series, New York.

French, S. 1986. *Decision Theory: An Introduction to the Mathematics of Rationality*. Ellis Horwood Limited, England and John Wiley & Sons, New York.

Fox, G., N. Baker, J. Bryant. 1984. Economic models for R&D project selection in the presence of project interactions. *Management Sci.* **30**(7) 890–902.

Gershon, M., L. Duckstein, R. McAniff. 1982. Multiobjective river basin planning with qualitative criteria. *Water Resources Res.* **18**(2) 193–202.

Goodwin, P., G. Weight. 2004. *Decision Analysis for Management Judgment*. Wiley & Sons, Chichester, England.

Harker, P. T. 1987. Incomplete pairwise comparisons in the analytic hierarchy process. *Math. Model.* **9** 837–848.

Hotman, E. 2005. Base reference analytical hierarchy process for engineering process selection. R. Khosla, R. J. Howlett, L. C. Jain, eds. *Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Computer Science*. Springer, New York, 184–190.

Keeney, R. L. 2002. Common mistakes in making value trade-offs. *Oper. Res.* **50**(6) 935–945.

Keeney, R. L., H. Raiffa. 1976. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. John Wiley, New York.

Liesio, J., P. Mild, A. Salo. 2007. Preference programming for robust portfolio modeling and project selection. *Eur. J. Oper. Res.* **181**(3) 1488–1505.

Lübbecke, M. E., J. Desrosiers. 2005. Selected topics in column generation. *Oper. Res.* **53**(6) 1007–1023.

Medaglia, A. L., S. B. Graves, J. L. Ringuest. 2007. A multiobjective evolutionary approach for linearly constrained project selection under uncertainty. *Eur. J. Oper. Res.* **179** 869–894.

Merrick, J. R. W., G. S. Parnell, J. Barnett, M. Garcia. 2005. A multiple-objective decision analysis of stakeholder values to identify watershed improvement needs. *Decision Anal.* **2**(1) 44–57.

Millet, I., P. T. Harker. 1990. Globally effective questioning in the analytic hierarchy process. *Eur. J. Oper. Res.* **48**(1) 88–97.

Parnell, G. S., H. W. Conley, J. A. Jackson, L. J. Lehmkuhl, J. M. Andrew. 1998. Foundations 2025: A value model for evaluating future air and space forces. *Management Sci.* **44**(10) 1336–1350.

Petrovski, A., J. McCall. 2001. Multi-objective optimisation of cancer chemotherapy using evolutionary algorithms. E. Zitzler, K. Deb, L. Thiele, C. A. Coello Coello, D. Corne, eds. *Evolutionary Multi-Criterion Optimization, Lecture Notes in Computer Science*, Vol. 1993. Springer-Verlag, London, 531–545.

Phillips, L. D. 1984. A theory of requisite decision models. *Acta Psychologica* **56** 29–48.

Reynolds, T. G., L. Ren, J.-P. B. Clarke, A. S. Burke, M. Green. 2005. *History, Development and Analysis of Noise Abatement Arrival Procedures for UK Airports*. AIAA 5th Aviation Technology, Integration and Operations Forum, Arlington, Virginia.

Saaty, T. L. 1990. How to make a decision: The analytic hierarchy process. *Eur. J. Oper. Res.* **48**(1) 9–26.

Santhanam, R., G. J. Kyparisis. 1996. Theory and methodology a decision model for interdependent information system project selection. *Eur. J. Oper. Res.* **89** 380–399.

Schrage, L. 2000. *Optimization Modeling with Lingo*. Lindo Systems, Chicago.

Shane, S. A., K. T. Ulrich. 2004. Technological innovation, product development, and entrepreneurship in management science. *Management Sci.* **50**(2) 133–144.

Solak, S., J.-P. B. Clarke, E. L. Johnson, E. Barnes, P. Collopy. 2008. *Optimization of Technology Development Portfolios*. Forthcoming.

Stewart, T. J., L. Scott. 1995. A scenario-based framework for multicriteria decision analysis in water resources planning. *Water Resources Res.* **31**(11) 2835–2843.

Stirling, A., S. Mayer. 1999. *Rethinking Risk*: *A Pilot Multi-Criteria Mapping of a Genetically Modified Crop in Agricultural Systems in the UK* (*SPRU Reports*). Science and Technology Policy Research, Brighton, UK.

Tavana, M. 2003. CROSS: A multicriteria group-decision-making model for evaluating and prioritizing advanced-technology projects at NASA. *Interfaces* **33**(3) 40–56.

Vanderbeck, F. 2000. On Dantzig-Wolfe decomposition in integer programming and ways to perform branching in a branch-and-price algorithm. *Oper. Res.* **48**(1) 111–128.