

CHAPTER XX

Data-driven evaluation of a flight re-route air traffic management decision-support tool*

Lianna M. Hall, Ngaire Underhill, Yari Rodriguez, Richard DeLaura

Weather Sensing Group, MIT Lincoln Laboratory
Lexington, MA
lianna.hall@ll.mit.edu

ABSTRACT

Air traffic delays in the U.S. are problematic and often attributable to convective (thunderstorms) weather. Air traffic management is complex, dynamic, and influenced by many factors such as projected high volume of departures and uncertain forecast convective weather at airports and in the airspace. To support the complexities of making a re-route decision, which is one solution to mitigate airspace congestion, a display integrating convective weather information with departure demand predictions was prototyped jointly by MIT Lincoln Laboratory and the MITRE Corporation. The tool was deployed to twelve air traffic facilities involved in handling New York area flights for operational evaluation during the summer of 2011. Field observations, data mining and analyses were conducted under both fair and convective weather conditions. The system performance metrics chosen to evaluate the tool's effectiveness in supporting re-route decisions include predicted wheels-off error, predicted wheels-off forecast spread, and hourly departure fix demand forecast spread. The wheels-off prediction errors were near zero for half the flights across all days, but the highest 10% errors exceeded 30

*This work was sponsored by the Federal Aviation Administration under Air Force Contract No. FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

minutes on convective weather days. The wheels-off forecast spread exceeded 30 minutes for 25% of forecasts on convective weather days. The hourly departure demand forecast spread was 9 flights or less for 50% of departures across all days except one. Six out of the seven days having the highest hourly departure demand forecast spreads occurred in the presence of long-lived weather impacts.

INTRODUCTION

Air traffic delays in the National Airspace System (NAS) are problematic. The New York region's airspace is highly congested, and problems originating from the region contribute to nearly three-quarters of NAS delays (Partnership for New York City, 2009). The Federal Aviation Administration (FAA) reports that 70% of delays are attributed to weather, a large portion of which are due to convective activity that is often localized and difficult to forecast (Hughes, 2011). Air traffic delays have increased during the months of the year in which convective weather predominates (Evans, 2001). Convective weather is unpredictable, greatly impacts air traffic, and adds challenges to air traffic management.

When convective weather occurs near airport terminal areas, traffic managers have difficult decisions to make to balance demand with capacity, such as choosing to delay flights, run traffic through impacted airspace, or reroute flights from impacted routes to non-impacted ones should they be available. Multiple factors are considered in departure planning, which include airspace resources such as departure fixes and jet routes. A departure fix is the first airspace location a flight passes through upon departing to a destination airport along a jet route. One departure fix may serve multiple air routes in the congested NY-area airspace. Demand includes the number of departures predicted over each fix and along each route; forecast weather impacts are considered for these resources and airports (Song, Bateman, Masalonis, & Wanke, 2009).

Departure management consists of planning and implementation activities among many stakeholders. Demand volumes over fixes and routes are considered strategically by national and regional air traffic managers to implement plans to mitigate congestion. Such plans, which may include reroutes, are subsequently implemented by regional and airport air traffic controllers in coordination with airline dispatchers and pilots (Beatty, Smith, McCoy, & Billings, 2000; Smith, Spencer, & Billings, 2007). To implement reroutes, impacted flights are identified, alternative routes are sought, sequencing of flights on the airport surface is considered, and flight plans are updated (Song et al., 2009). Air traffic managers coordinate with additional stakeholders such as pilots and airline dispatchers, but do so without the benefit of integrated tools that provide situational awareness of air traffic demand and forecasted weather impacts (Beatty et al., 2000; Song et al., 2009).

Decision-support tool components

To support the complexities of making a re-route decision, a display integrating convective weather information with departure demand predictions was prototyped jointly by MIT Lincoln Laboratory and the MITRE Corporation. The display and underlying system components are collectively called the Integrated Departure Route Planning (IDRP) tool, which provides real-time, integrated departure information (DeArmon, Jackson, Bateman, Song, & Brown, 2010). The purpose of the IDRP tool is to reduce the workload required to identify, plan, coordinate, and implement re-routes during convective weather (Masalonis et al., 2008).

The 2011 IDRP prototype is the second version of the system (DeLaura, Underhill, Hall, & Rodriguez, 2011); the first version was deployed in 2010 for a limited field study. Underlying the IDRP prototype are the Route Availability Planning Tool (RAPT) and the Corridor Integrated Weather System (CIWS), both of which are presented in Figure 1. RAPT provides 30-minute forecast weather impacts in 5-minute increments for each departure route (Robinson, DeLaura, & Underhill, 2008), and it has undergone a series of field evaluations to solidify and expand its operational concept (Davison Reynolds, DeLaura, & Robinson, 2010; Robinson et al., 2008; Robinson, DeLaura, Evans, & McGettigan, 2008). CIWS provides forecast weather on a geospatial display that allows detailed weather information to be provided (Robinson et al., 2008).

The 2011 IDRP prototype calculates departure demand forecasts using filed flight plans and current aircraft locations on the airport surface from the Airport Surveillance Detection Equipment, Model X (ASDE-X) where available. Course- and fine-grained departure forecasts across four visual components (DeLaura et al., 2011) are available; the first two are presented in Figure 1:

1. a fix list giving predicted departure demand and congestion alerts for each departure fix,
2. predictions of departure demand on each RAPT departure route,
3. a departure demand flight list that provides origin, destination, fix, flight plan, predicted departure time and RAPT status, and
4. a reroute alternative list with RAPT forecast and additional miles flown for each flight in the flight list.

The aggregate departure demand predictions, shown as totals in the right-most column of the fix list and alongside each route, provide course-grained information to support ATMs' strategic re-route planning. The aggregate departure demand predictions, shown in 15-minute bins in the fix list, along with the detailed flight and re-route lists (not shown), support fine-grained planning, coordination, and implementation of re-routes.

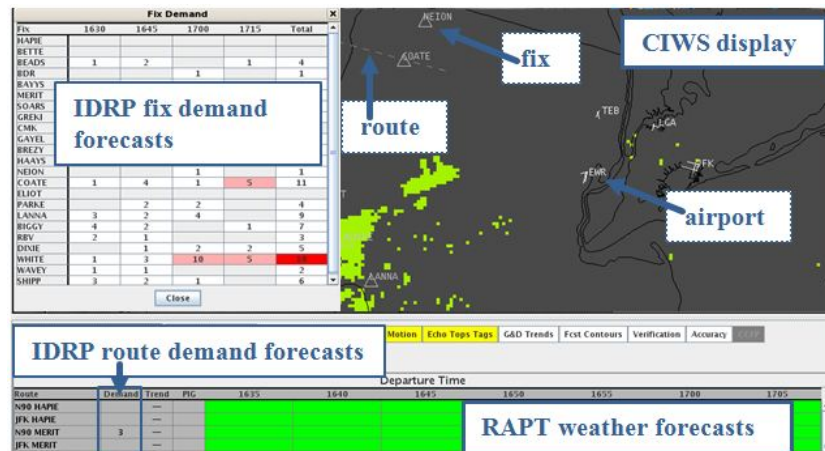


Figure 1. IDRP views for aggregate fix and route demand forecasts; NYC-area airspace.

Tool evaluation

The IDRP tool was deployed in the ZNY New York Air Route Traffic control Center (ARTCC), N90 New York Terminal Radar Approach Control (TRACON) facility, and ten other facilities for operational evaluation during the summer of 2011 (DeLaura et al., 2011). Field observations were conducted across the facilities, and data mining and analyses were performed on underlying system components, supplemented with reported flight departure (wheels-off) times and observed (true) weather impacts, to assess forecast accuracy and reliability. System stability, expressed using accuracy (correctness) and precision (consistency, reliability) measures, may impact the adoption or value of a decision-support tool. Disuse, an underutilization of automation, may result from an unstable system that can cause users to distrust the information (Parasuraman & Riley, 1997; Lee & See, 2004). Conversely, misuse, an overreliance of automation, may occur if users accept the information literally without regard for its reliability; in effect, causing an over trust of the information (Parasuraman et al., 1997; Lee et al., 2004; Smith, McCoy, & Layton, 1997).

METHODS

The tool's effectiveness in supporting re-route decisions was assessed using IDRP departure demand forecast issuances that were updated at a frequency of once every minute. The forecast performance metrics were evaluated separately for fair and convective weather days, as operations differ in the presence of convective weather. Departures from five major New York area airports were included in the analysis: Newark Liberty International, NJ (EWR); LaGuardia, NY (LGA); John F. Kennedy International, NY (JFK); Teterboro, NJ (TEB); White Plains, NY (HPN).

Three of these airports (EWR, LGA, and JFK) have ASDE-X systems that provide information about aircraft location on the airport surface and in the immediate airspace. Forecast issuances for twenty-three hours of each day, ranging from midnight to 11:00pm local time (04:00Z to 02:59Z the next day), were included in the analysis.

Three forecast performance metrics were defined for the evaluation: predicted wheels-off error, predicted wheels-off spread, and predicted fix demand spread. The IDRP flights included in the predicted wheels-off analyses were limited by two additional criteria. First, the flights must correlate to reported Aviation System Performance Metrics (ASPM) flights for actual wheels-off times. Second, the flights must not have been rerouted, because their departure times may have significant changes due to the additional coordination required to implement reroutes. The predicted wheels-off forecasts were evaluated with respect to a planning horizon, which represents a decision maker’s point of view for the time during which proactive reroute planning occurs to avoid convective weather and congestion. For the evaluation, a 30-minute planning horizon was used to align with the current RAPT status forecast limit. Figure 2 illustrates the forecasted wheels-off issuances, and identifies the relationship between the forecasted wheels-off times and the planning horizon.

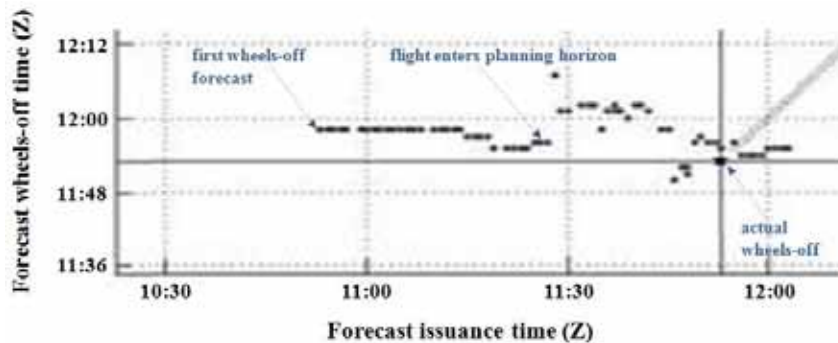


Figure 2. The decision-making planning horizon for a single flight’s forecasts. The flight enters the 30-minute planning horizon at 11:27:00Z when the forecast wheels-off time is 11:56:21Z. Actual wheels-off time is 11:54:00Z; forecasts range from 11:51:00Z (earliest) to 12:07:39Z (latest).

The forecast issuance when the flight entered the planning horizon was used to calculate the predicted wheels-off error and magnitude metrics. For each flight, the predicted wheels-off error was calculated as the difference between the flight’s actual wheels-off time and the predicted wheels-off time at the time the flight entered the 30-minute planning horizon. Given the predicted wheels-off issuances provided in the example flight illustrated in Figure 2, the wheels-off error is -00:02:41Z, which represents a late forecast. A highly volatile forecast may be difficult to use in planning, so the wheels-off forecast spread was assessed. The

wheels-off forecast spread for a particular flight was defined as the difference between the latest and earliest predicted wheels-off times for that flight over the interval of time from the flight's entrance into the planning horizon until its wheels-off time. Given the predicted wheels-off issuances provided in the example flight illustrated in Figure 2, the wheels-off error spread is 00:16:39Z.

A highly volatile wheels-off forecast may likely manifest itself in the forecasted fix demand counts, so the aggregate hourly fix demand forecast spread was assessed. The hourly fix demand forecasts were selected instead of the more fine-grained 15-minute forecasts to focus on this strategic decision-support aspect. The total predicted hourly demand across twenty-four NY-area departure fixes, equal to the sum of hourly demand forecast from all twenty-four fixes, was calculated for each forecast issuance. The hourly fix demand forecast spread was defined as the difference between the largest and smallest total hourly fix demand for forecast issuances within a 15-minute time period. .

RESULTS

The forecast performance metrics were evaluated for a total of twelve days spanning the summer of 2011, two fair weather and ten convective weather days. The scale (widespread or local) and duration (long-lived or short-lived) of convective weather impacts were reported in (DeLaura et al., 2011), which included: four days having short-lived, widespread impacts; one day having long-lived, local impacts; five days having periods of long-lived, widespread impacts.

The IDRPs prototype was predominantly used in one facility, the New York TRACON (N90), out of the twelve facilities deployed. The supervisors and traffic management coordinators at N90, responsible for the airspace surrounding the NYC airports, were observed using the tool to monitor trends in fix demands, detect capacity overloads, and identify possible reroutes to avoid congested fixes. The field observer noticed flickering of the fix demand forecasts, but subjects did not explicitly comment on this tool behavior.

Forecast Wheels-Off Analyses

Over 15,000 departure flights were included in the two predicted wheels-off analyses. Wheels-off prediction times were generally constant (or infrequently changing) until aircraft entered into ASDE-X coverage, as shown in the example flight in Figure 2. From this point, wheels-off prediction times changed every minute. Several different prediction behaviors were observed. In some instances, predictions steadily converged toward the actual wheels-off time, and errors decreased as the actual wheels-off time approached. However, forecasts often showed considerable volatility, as the wheels-off time forecasts moved later and earlier, sometimes not approaching the actual wheels-off time until just a few minutes before takeoff.

Predicted wheels-off error measurements were made separately for convective

and fair weather days, and are presented as a histogram and a line overlay, respectively, in Figure 3. Median errors are near zero minutes for both datasets. A negative wheels-off forecast error indicates that a flight departed before the predicted wheels-off time (a ‘late’ forecast). The error distribution falls off more slowly for convective days than fair weather days. Half of the wheels-off prediction errors on convective days fell within the error bound envelope of -10 and plus 12 minutes (except for August 25, when the error envelope reached 20 minutes). The extreme error bound – the ceiling for the highest 10% errors – ranged from 30 to 50 minutes (with the exception of August 25 and June 22, when the extreme error bounds were 70 minutes and 23 minutes, respectively).

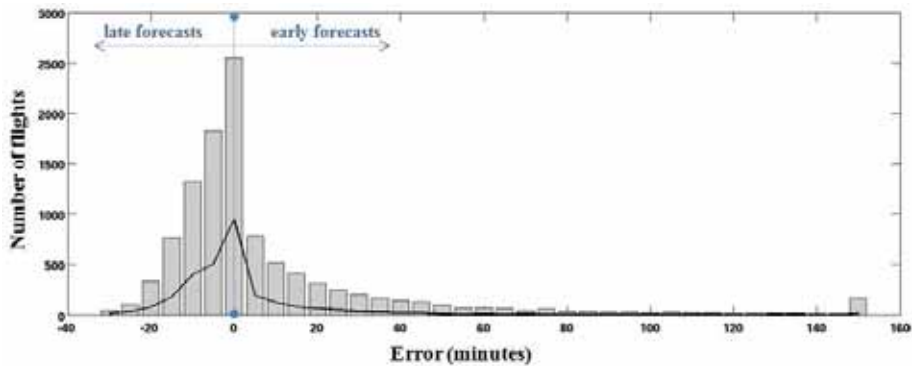


Figure 3. Histogram of wheels-off forecast errors for convective (bars) and fair (line) weather days

Wheels-off forecast spread measurements were made separately for convective and fair weather days. On convective days, the spread was typically 20 minutes or less for many flights, but there was a very long tail to the distribution. Wheels-off forecast spreads on fair weather days was 20 minutes or less for the majority of forecasts. The spread of wheels-off forecasts on convective days was generally around 30 minutes or less for 75% of departures (with the exception of August 25, when the spread was approximately 45 minutes). The extreme spread ranged from 50 to 70 minutes on convective days (with the exception of August 25, when the extreme spread was approximately 90 minutes).

Hourly Fix Demand Forecast Analysis

The hourly fix demand forecast spread statistics (10th, 25th, 50th, 75th, and 90th percentiles) were calculated for each individual day. Predicted hourly fix demand spread was 9 flights or less for 50% of departures (except for September 7, when the spread reached 14 flights). The hourly fix demand forecast spread on convective days was 19 flights or less for 75% of departures (except for two convective weather days, July 29 and September 7, whose spreads were 28 and 34 flights,

respectively). The extreme forecast spread ranged from 17 to 55 flights on convective days, and ranged from 8 to 19 flights on fair weather days. Six out of the seven days having the largest spreads of the two most extreme spread bounds (75th and 90th percentiles) occurred in the presence of long-lived weather impacts. Three out of the four days having the smallest spreads of the two most extreme spread bounds incurred widespread weather impacts of short duration. An example of large forecast spreads in the presence of impacted weather is illustrated in Figure 4, which shows increased forecast spread starting 19:45Z after locally impacted weather starts around 18:00Z and continues for 8 hours.

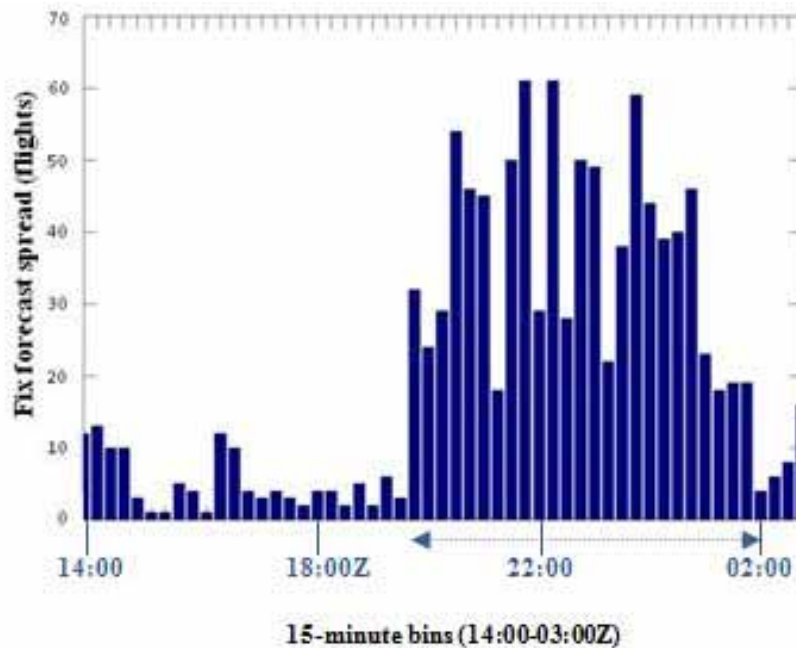


Figure 4. Hourly fix demand forecast spreads, in 15-minute bins, from July 19 at 14:00Z to July 20 at 03:00Z, in 15-minute bins; locally impacted weather from 18:00Z to 02:00Z, increased forecast spread from 19:45Z to 02:00Z.

DISCUSSION AND CONCLUSIONS

Wheels-off forecast accuracy and reliability are important because weather impacts on operations can vary greatly at different departure times and result in different traffic management decisions. Accuracy of the wheels-off prediction can influence the quality of traffic management decisions. Reliability of the wheels-off prediction, for individual flights and their contribution to aggregate hourly fix demand, may also impact the quality of a traffic management decision and may cause users to distrust the decision support.

The wheels-off forecast error metric revealed that, although half the flights had a near zero error across all days, over a quarter of flights had late predictions. The presence of late predictions can give a user the impression that a longer reroute decision time period is available than is actually the case. On convective weather days, errors for 10% of flights were beyond 30 minutes (i.e., 'early' predictions, where actual departure times were more than 30 minutes later than predicted) for all days except one; this exceeds the planning horizon available for users to proactively implement a reroute. The predicted wheels-off error was overall lower (the forecasts were more accurate) on fair weather days. On convective weather days, a quarter of the flights had a wheels-off forecast spread of 30 minutes or more, which increases the uncertainty of departure demand as it may give an inaccurate picture of congestion. The hourly fix demand forecast spread was generally lowest on the two fair weather days and on convective days having short-lived, widespread weather impacts. The forecast spread was highest on convective days characterized by long-lived weather impacts, where most days also had widespread weather impacts.

Overall, the departure demand forecasts were less accurate and reliable on severe convective weather days. Widespread weather impact conditions necessitate the use of impacted airspace to move departures, which itself has a high degree of uncertainty. The uncertainty of airspace capacity in turn makes departure capacity uncertain, which can make predictions about wheels-off times difficult to make. Although the perception of system performance was not explicitly measured, the system instability revealed in this study and noted by a field observer may cause a series of unanticipated consequences in the tool's use. What is not clear is how the system instability affects decision making and whether it causes over-control, paralysis, or poor decisions.

Developing a decision-support tool to enable air traffic managers to effectively manage highly impacted airspace is challenging given uncertainty in weather, pilot behavior, arrival and departure demand, and performance of individual air traffic managers and controllers. This study defined novel performance metrics to evaluate the IDRP tool from a user's point-of-view, exposed areas of system instability, and established specific areas of interest to investigate further. Models that relate errors and reliability in wheels-off, fix demand, and weather impact forecasts to departure throughput should be developed to assess the costs of forecast uncertainty and to determine meaningful forecast requirements. Algorithm improvements that trade dampened forecast response for improved stability should be explored. Finally, a detailed analysis of variations in the underlying flight and route lists may shed light on the usefulness of this tool component in support of reroute implementations.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Richard Ferris, Ngaire Underhill, Darin Meyer, Diana Klingle-Wilson, Hayley Reynolds, and Brad Crowe for their field

work in the facilities.

REFERENCES

- Beatty, R., Smith, P.J., McCoy, C.E., Billings, C.E. (2000). Design Recommendations for an Integrated Approach to the Development, Dissemination and Use of Reroute Advisories. *Institute for Ergonomics Technical Report*.
- Davison Reynolds, H.J., DeLaura, R.A., Robinson, M. (2010). Field & (Data) Stream: A Method for Functional Evolution of the Air Traffic Management Route Availability Planning Tool (RAPT). *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting, 27 September – 1 October, San Francisco, CA, 104-108*.
- DeArmon, J., Jackson, C., Bateman, H., Song, L., Brown, P. (2010). Benefits Analysis of a Routing Aid for New York Area Departures. *IEEE/AIAA 29th Digital Avionics Systems Conference (DASC), 3-7 October, Salt Lake City, UT, 1.D.3-1-1.D.3-8*.
- DeLaura, R.A., Underhill, N., Hall, L.M., & Rodriguez, Y. (2011). Evaluation of the Integrated Departure Route Planning (IDRP) Tool 2011 Prototype. *Project Report (in press), MIT Lincoln Laboratory, Lexington, MA*.
- Evans, J.E. (2001). Tactical Weather Decision Support To Compliment “Strategic” Traffic Flow Management for Convective Weather. *4th USA/Europe Air Traffic Management R&D Seminar, 3-7 December, Santa Fe, New Mexico, 1-11*.
- Hughes, D. (2011). Application-Oriented NextGen Weather Research. *Journal of Air Traffic Control, 54(1), 30-33*.
- Lee, J.D., See, K.A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors, 46(1), 50-80*.
- Masalonis, A., Bateman, H., DeLaura, R., Song, L., Taber, N., & Wanke, C. (2008). Integrated Departure Route Planning. *27th Digital Avionics Systems Conference, 26-30 October, St. Paul, MN*.
- Parasuraman, R., Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors, 39(2), 230-253*.
- Partnership for New York City (2009). Grounded: The high cost of air traffic congestion. Retrieved February 9, 2012, from http://www.pfnyc.org/reports/2009_0225_airport_congestion.pdf.
- Robinson, M., DeLaura, R.A., Evans, J.E., McGettigan, S. (2008). Operational Usage of the Route Availability Planning Tool (RAPT) During the 2007 Convective Weather Season. *Project Report ATC-330, MIT Lincoln Laboratory, Lexington, MA*.
- Robinson, M., DeLaura, R., Underhill, N. (2009). The Route Availability Planning Tool (RAPT): Evaluation of Departure Management Decision Support in New York during the 2008 Convective Weather Season. *Eighth USA/Europe Air Traffic Management Research and Development Seminar (ATM2009), Napa, CA*.
- Smith, P.J., McCoy, C.E., Layton, C. (1997). Brittleness in the Design of Cooperative Problem-Solving Systems: The Effects on User Performance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 27(3), 360-371*.
- Smith, P.J., Spencer, A.L., Billings, C.E. (2007). Strategies for designing distributed systems: case studies in the design of an air traffic management system. *Cogn Tech Work, 9, 39-49*.
- Song, L., Bateman, A., Masalonis, A., Wanke, C. (2009). Integrated Collaborative Departure Traffic Management. *9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO), 21-23 September, Hilton Head, SC, AIAA, Reston, VA*.