NAVAL
POSTGRADUATE
SCHOOL

MONTEREY, CALIFORNIA

THESIS

METRICS OF METOC FORECAST PERFORMANCE AND OPERATIONAL IMPACTS ON CARRIER STRIKE OPERATIONS

by

Jeremy Callahan

September 2006

Thesis Advisor:	Tom Murphree
Co-Advisor:	Rebecca Stone

Approved for public release; distribution is unlimited

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** September 2006	**3. REPORT TYPE AND DATES COVERED** Master's Thesis

4. TITLE AND SUBTITLE Metrics of METOC Forecast Performance and Operational Impacts on Carrier Strike Operations	**5. FUNDING NUMBERS**
6. AUTHOR(S) Callahan, Jeremy J.	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release, distribution is unlimited	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

We have developed metrics of the performance and operational impacts of METOC support to strike operations conducted on operational aircraft carriers (CVs). Our goal was to assess that support and make recommendations on improving it. We adapted an existing automated real time METOC metrics system, which was developed for land based training missions, for use on CVs by developing a new data collection form, new metrics, and new collection, analysis, and reporting architecture for the remote entering of sensitive mission data without compromise. The weather support element of a CV, the OA division, does not provide strike mission planning support, but does provide situational awareness to pilots. Our system allows that situational awareness to be measured and assessed using metrics that quantify the performance of the forecasts, the relationship of the forecasts to the mitigating actions taken by pilots due to adverse weather conditions, and the effects of individual weather phenomena on the execution of strike missions. A key element of the data collection, analysis, and reporting system developed in this study is the collection of METOC related data from pilots during their intelligence debriefings. This system is readily adaptable for the assessment of METOC support to other warfare areas.

14. SUBJECT TERMS Meteorology, METOC, metrics, quality, performance metrics, operational impacts, impacting phenomena, situational awareness, strike warfare, carrier operations, METOC support	**15. NUMBER OF PAGES** 116
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

METRICS OF METOC FORECAST PERFORMANCE AND OPERATIONAL IMPACTS ON CARRIER STRIKE OPERATIONS

Jeremy J. Callahan
Lieutenant, United States Navy
B.S., Texas A&M University, 1999

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY AND PHYSICAL OCEANOGRAPHY

from the

NAVAL POSTGRADUATE SCHOOL
September 2006

Author:	Jeremy Callahan
Approved by:	Tom Murphree Thesis Advisor
	Rebecca Stone Co-Advisor
	Professor Philip A. Durkee Chairman, Department of Meteorology

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

We have developed metrics of the performance and operational impacts of METOC support to strike operations conducted on operational U.S. Navy aircraft carriers (CVs). Our goal was to assess that support and make recommendations for improving it. We adapted an existing automated real time METOC metrics system, which was developed for land based training missions, for use on CVs by developing a new data collection form, new metrics, and new collection, analysis, and reporting architecture for the remote entering of sensitive mission data without compromise. The weather support element of a CV, the OA division, does not provide strike mission planning support, but does provide situational awareness to pilots. Our system allows that situational awareness to be measured and assessed using metrics that quantify the performance of the forecasts, the relationship of the forecasts to the mitigating actions taken by pilots due to adverse weather conditions, and the effects of individual weather phenomena on the execution of strike missions. A key element of the data collection, analysis, and reporting system developed in this study is the collection of METOC related data from pilots during their intelligence debriefings. This system is readily adaptable for the assessment of METOC support to other warfare areas.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AFW	Air Force weather
AFWSPV	Air Force Weather Strategic Plan and Vision
AMC	Air Mobility Command
ATO	Air tasking order
CAOC	Combined Air Operations Center
CNMOC	Commander, Naval Meteorology and Oceanography Command
CNO	Chief of Naval Operations
CO	Commanding officer
CV	Aircraft carrier
CVIC	Aircraft Carrier Intelligence Center
CWT	Combat weather team
DOD	Department of Defense
DoN	Department of the Navy
FNMOC	Fleet Numerical Meteorology and Oceanography Center
GWOT	Global war on terrorism
ISR	Intelligence, surveillance, and reconnaissance
IT	Information technology
LHA	Amphibious assault ship (*Tarawa* Class)
LHD	Amphibious assault ship (*Wasp* Class)
METOC	Meteorology and oceanography
MRM	METOC related mishap
NIPRNet	Non-secure Internet Protocol Router Network
NLMOC	Naval Atlantic Meteorology and Oceanography Center
NPMOD	Naval Pacific Meteorology and Oceanography Detachment, Fallon
NPS	Naval Postgraduate School
NSAWC	Naval Strike and Air Warfare Center
NWS	National Weather Service
OIF	Operation IRAQI FREEDOM

OMB	Office of Management and Budget
ONR	Office of Naval Research
PART	Program Assessment Rating Tool
SA	Situational awareness
SGOT	Strike Group Oceanography Team
SIPRNet	Secure Internet Protocol Router Network
SPAWAR	Space and Naval Warfare Systems Command
XO	Executive officer

GLOSSARY OF SPECIAL TERMS USED IN THIS STUDY

Bias: Bias is a forecast performance metric that compares forecasted severe negative weather conditions to observed severe negative weather conditions to determine if is the conditions are being over or under forecasted. Bias uses the following calculation based on the standard contingency table:

$$\mathrm{Bias} = 100 \times \frac{A + C}{A + B}$$

False alarm rate (FAR): FAR is a forecast performance metric that measures the rate at which incorrect forecasts are issued by comparing incorrect forecasts of severe negative weather conditions to the total number of forecasts of these conditions. FAR is determined using the following calculation based on the standard contingency table:

$$\mathrm{FAR} = 100 \times \frac{C}{A + C}$$

Forecast accuracy (FAC): FAC is a forecast performance metric that measures the rate at which correct forecasts are issued by comparing correct forecasts to the total number of forecasts. FAC is determined using the following calculation based on the standard contingency table:

$$\mathrm{FAC} = 100 \times \frac{A + D}{A + B + C + D}$$

Metrics: Quantitative measures of variables and their relationships with each other that are critical in assessing the performance of an operation or organization. Example: FAC is a metric of the performance of forecasts and of the METOC units that generate those forecasts.

Missions canceled: An operational impacts metric equal to the percentage of total missions that were cancelled due to severe negative weather impacts. This metric is a subset of missions requiring mitigation that specifically shows which missions were unable to complete their primary objective due to severe negative weather impacts.

Mission canceling phenomena: An individual negatively impacting phenomena metric equal to the percentage of missions that were cancelled due to an individual weather phenomenon. This metric is useful in determining which phenomena have large operational impacts and are therefore especially important to correctly forecast.

Mission impacting phenomena: An individual negatively impacting phenomena metric equal to the percentage of mission impacts that were due to an individual weather phenomenon. This metric is useful in determining which phenomena have large operational impacts and are therefore especially important to correctly forecast.

Missions placed at risk: An operational impacts metric equal to the percentage of total missions flown for which the pilots encountered a severe negative weather impact.

Missions requiring mitigation: An operational impacts metric equal to the percentage of missions that encountered a severe negative weather impact and for which the pilots took mitigating action. This metric helps distinguish the mission that encountered adverse weather conditions and were able to take mitigating action from those that encountered adverse weather conditions but could not take mitigating action.

Mitigation rate: An operational impacts metric equal to percentage of total missions requiring a mitigating action by the pilot. This metric gives an overall picture of the extent to which weather impacts operations.

Mitigated received negative: An operational impacts metric equal to the number of missions with a correctly forecasted severe negative weather impact and a mitigating action taken by the pilot divided by total number of missions forecasted to have severe negative weather impacts.

OA Division: Designation for the division on a ship, if so equipped, that is responsible for providing METOC support. The “O” designates the operations Department and the “A” designates the Aerography Division.

Probability of detection (POD): POD is a forecast performance metric equal to the number of times a given type of severe negative weather impact was correctly forecasted divided by the number of times that impact was predicted. POD is determined using the following calculation based on the standard contingency table:

$$\mathrm{POD} = 100 \times \frac{A}{A + B}$$

Received negative: An operational impacts metric equal to the number of missions forecasted to have a severe impact divided by the total number of missions.

Situational awareness (SA): SA is the understanding prior to mission execution of the actual and potential, and the controllable and uncontrollable, aspects of a mission. Pilots with SA are better able to prepare mitigation and contingency actions in the event that the mission cannot proceed flawlessly or as planned.

Targets changed: An operational impacts metric equal to the percentage of total missions for which the target was changed due to severe negative weather impacts. This metric is a subset of missions requiring mitigation that specifically shows which missions required a target change due to severe impacts.

Target changing phenomena: An individual negatively impacting phenomena metric equal to the percentage of missions for which an individual negatively impacting phenomenon led to a target change. This metric is useful in determining the phenomena that, from the perspective of the end user, are most important to accurately forecast.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

This study could not have been possible if it were not for my advisors, Professor Tom Murphree and CDR Rebecca Stone. Their patience with me and knowledge of the topic were instrumental in steering my efforts towards a useful conclusion. Our research and development effort was supported, in part, by Rapid Transition Program funding from ONR and SPAWAR.

I would also like to acknowledge the support of CDR Steve Woll and LCDR Gregory Schmeiser from SGOT Norfolk. Their insight, experience, and organizational influence were critical elements needed in determining what data to collect, how to collect it, and how to enlist the support of the OA divisions. Without CDR Woll and LCDR Schmeiser, we might still be asking the wrong questions of the wrong people. I would like to thank Bruce Ford of Clear Science Inc., His prior METOC experience, computer programming skill, and flexibility allowed for us to efficiently create a working collection and reporting system. His patience with me should also be noted, as I tried his math ability over and over. I would also like to acknowledge William (Billy) Little, an intern during the summer of 2006 at NPS. His assistance in creating a test database and hand checking calculations was a great help.

I would like to also thank the participating afloat units for their participation and feedback. They allowed us to conduct trials to learn what was right and wrong with our questions, and provided us with the data for our calculations. The participating units include the OA division from the USS Enterprise, under LCDR Keith Everett, the OA division from the USS Eisenhower, and the OA division on the USS Stennis under LCDR Matt Moore.

Finally, I'd like to thank my ever loving and understanding wife Christina. She had to put up with long nights, cranky evenings, limited access to the home computer, and my inexplicable desire to lug all my study materials to Lake Tahoe. Her support was needed and selfless.

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

1. Introduction

A metrics program involves objectively measuring and evaluating an organization's critical variables, and their relationships to each other, to help in assessing and improving the organization's products and services. Metrics can be applied to assessing the Navy meteorology and oceanography (METOC) community's support of strike warfare by measuring the accuracy of METOC products provided in support of strike warfare and the operational impacts of those products on strike warfare. A standard set of Navy METOC community metrics will help rate the forecasting skill of the community and of its individual units, it will track the value of the community to the war fighter, and it will assist war fighters in their mission planning.

We have conducted a research, development, and transition study to help lay the foundation for a METOC metrics program. This study is part of a long line of closely related METOC metrics studies conducted at the Naval Postgraduate School (NPS) during the last six years. This study had three main goals. The first goal was to adapt the metrics collection and analysis system developed by Butler (2005) for use in assessing operational strike warfare support and for eventual use across a broad range of warfare types. The second goal was to implement our adapted system on an operational CV. The third goal was to conduct an initial evaluation of the accuracy of METOC products provided in support of strike warfare and the operational impacts of those products on strike warfare.

2. Data and Methods

We focused our metrics development efforts on: (a) the relatively short term mission execution weather forecasts provided to pilots by the OA divisions on the USS Enterprise, USS Eisenhower, and USS Stennis; and (b) the actions taken by those pilots in response to severe negative weather impacts. Pilots are the main customer of the OA division on an aircraft carrier. They receive a forecast from the OA division that provides them with situational awareness of the weather expected during the mission. We were able to measure the accuracy of the flight path weather forecasts and to infer their value to the pilots' situational awareness during mission execution.

The data collected for this project had to be relevant to missions flown by strike aircraft and had to be observable by the pilots. In discussions with members of SGOT Norfolk we decided to collect forecast and observational data for general aviation and weapons impacts for the launch, transit, refueling, over target, and recovery phases of the missions. This is the data that is routinely provided by OA divisions to pilots for individual missions, so collecting this data did not impose much burden on the division.

One of our main challenges was that standard meteorological data for verifying mission forecasts is generally not available, except at the carrier. This meant we needed to rely on the pilots for their assessments of the conditions they experienced during their missions. In addition, we also had to rely on them to provide us with data on the actions they took to mitigate severe negative weather impacts. Prior studies in which METOC personnel have attempted to collect such METOC and operational data from pilots have shown that it can be very difficult get pilot cooperation. So we needed to develop a method for collecting enough data to accurately describe what happened in the environment but not so much that we would lose the cooperation of the pilots. We settled on an approach for collecting pilot data that was based on three main questions:

1. Were any inconsistencies between weather that impacted the mission and weather in the mission execution forecast?
2. What, if any, part of the mission was negatively impacted by weather?
3. What if any mitigating actions were taken due to adverse weather conditions?

OA division personnel obtained the answers to these basic questions mainly by sitting in on intelligence debriefings of the pilots, supplemented by follow-up questions if needed. The answers to these questions provided us with the data we needed without asking pilots to recount specific meteorological values, and while imposing little or no extra work on the pilots. .

The majority of our metrics calculations are based on a contingency table similar to the one in Figure 1. This contingency table allows for easy categorization and grouping of forecasts.

Forecasted **Observed**	**Severe Impact**	**No Severe Impact**	**Total**
Severe Impact	Hit **A**	Miss **B**	A+B
No Severe Impact	False Alarm **C**	Correct Rejection **D**	C+D
Total	A+C	B+D	A+B+C+D

Figure 1. Contingency table. Forecasts are organized into one of four categories in the table, A-D. Each category corresponds to one the four possible combinations of forecasted conditions and observed conditions.

The four categories of the contingency table we used were based on our interest in assessing forecasts of severe negative weather conditions, the forecasts that prior studies have shown are especially challenging for forecasters, and especially important to war fighters (e.g., Jarry 2005). The four categories are: A: severe impact forecast and observed; B: severe impact not forecasted but observed; C: severe impact forecasted but not observed; and D: severe impact not forecasted and not observed. We determined whether forecasts and observations represented severe impacts by comparing them to published weather related thresholds for different strike mission air frames, weapons, sensors, etc. A severe weather impact threshold represents the conditions beyond which successful mission completion is very doubtful. By comparing the number of forecasts in each category to each other and to the total number of forecasts, forecast performance metrics can be directly calculated, and other types of operational impact and weather phenomena metrics can be determined.

We calculated three types of metrics: forecast performance metrics, operational impacts metrics, and impacting phenomena metrics. The forecast performance metrics were calculations based only on the standard contingency table in Figure 1 and were used to describe the performance of the forecasters in predicting events that would have resulted in significant impacts to operations. Operational impacts metrics represent how the forecast for the mission affected operations. Finally, we used impacting phenomena metrics to describe how individual weather phenomena impacted missions. An example of the output from a performance metric calculation showing probability of detection (POD) can be seen in Figure 2.

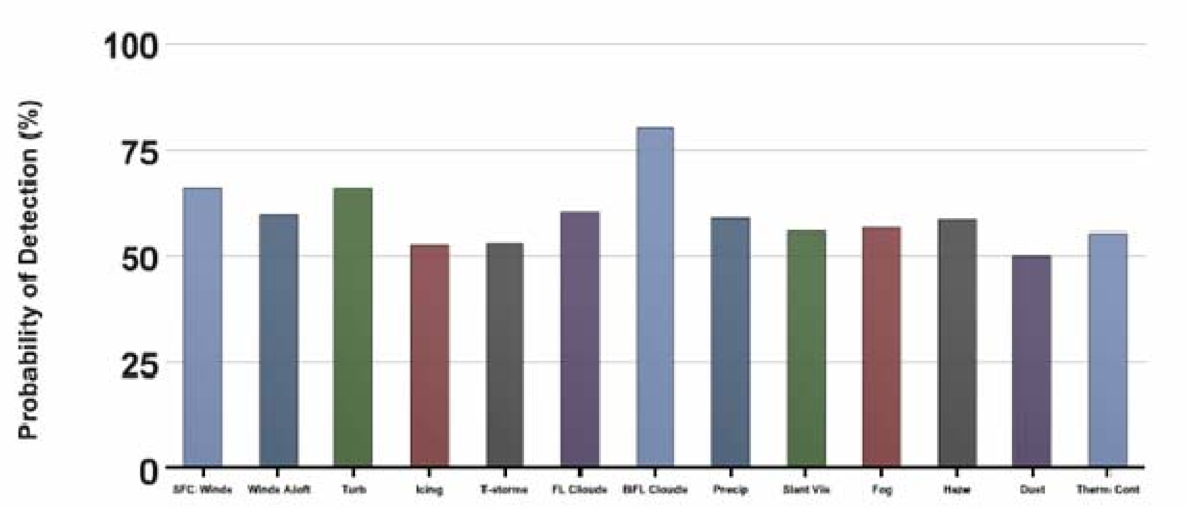
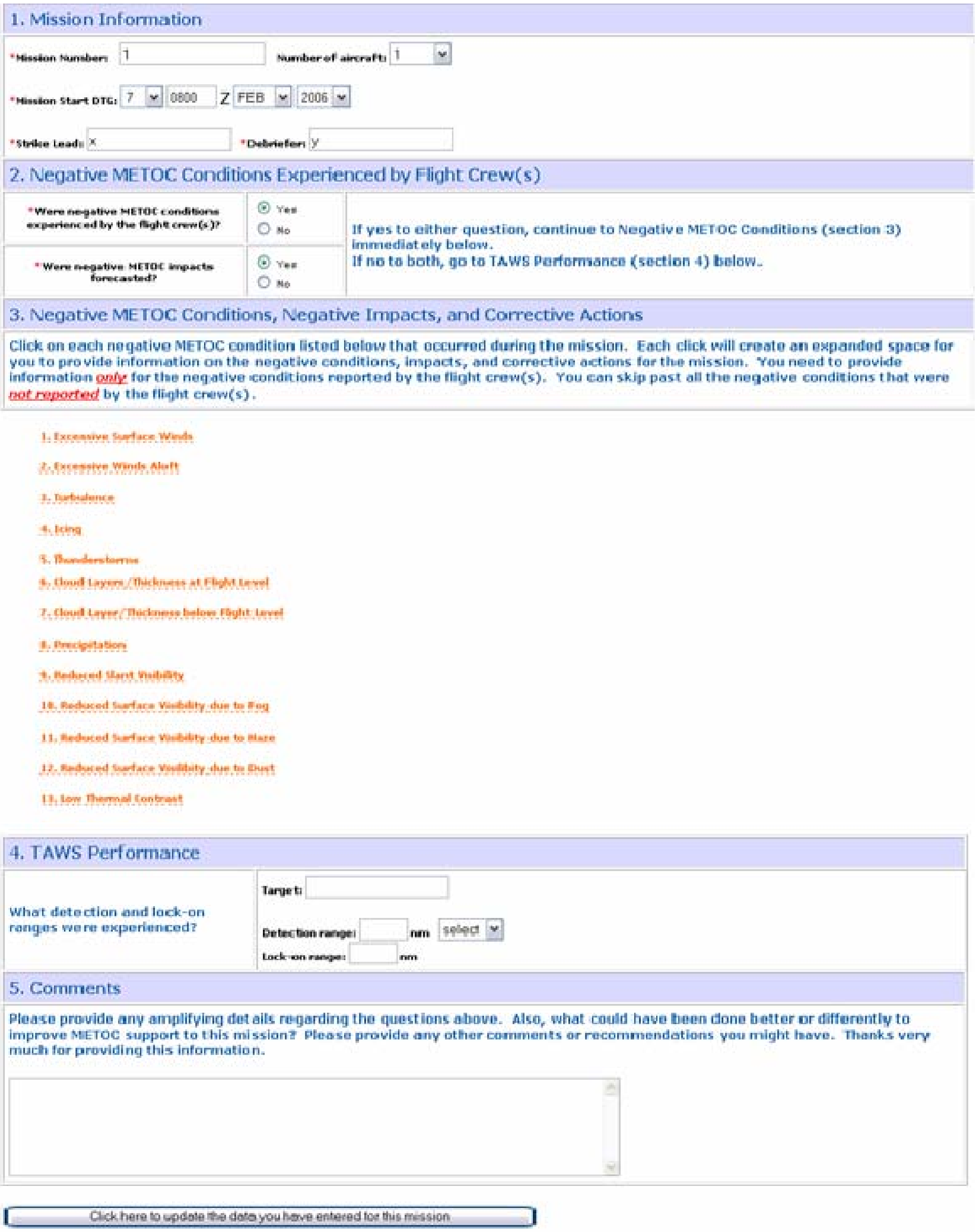


Figure 2. Example of POD metric output based on test data analyzed. Each bar indicates the POD for 13 different weather phenomena types and for all mission phases. Results based on classified real world data are similar.

We developed our data collection, analysis, and report system on the internet to allow for easy development of the system by civilian contractors. We then moved the system to the NPS SIPRNet server for further testing and operational use by the OA divisions. We implemented the system on the SIPRNet to ensure protection of sensitive mission data and to allow maximum accessibility to aircraft carriers at sea. We propose that our system eventually be transferred to an operational METOC center (e.g., FNMOC). Due to the Navy's IT requirements, current hosting of the system is limited to NPS and FNMOC because these locations have developmental systems that can host the open source PHP and PostgreSQL software used in our system.

3. Results

Designing and developing our system involved three main steps. The first step was to design a set of questions on a paper form that could be put through a trial with an OA division to experiment with content and wording. The second step was to develop the user interface seen in Figures 3 and 4 for entering the collected data into a database. The final step was to test the, collection, analysis, and reporting functions of the web based system using real world operational data.



1. Mission Information

*Mission Number: 1 Number of aircraft: 1

*Mission Start DTG: 7 0800 Z FEB 2006

*Strike Lead: X *Debriefer: Y

2. Negative METOC Conditions Experienced by Flight Crew(s)

*Were negative METOC conditions experienced by the flight crew(s)? ● Yes ○ No

*Were negative METOC impacts forecasted? ● Yes ○ No

If yes to either question, continue to Negative METOC Conditions (section 3) immediately below.
If no to both, go to TAWS Performance (section 4) below.

3. Negative METOC Conditions, Negative Impacts, and Corrective Actions

Click on each negative METOC condition listed below that occurred during the mission. Each click will create an expanded space for you to provide information on the negative conditions, impacts, and corrective actions for the mission. You need to provide information *only* for the negative conditions reported by the flight crew(s). You can skip past all the negative conditions that were *not reported* by the flight crew(s).

1. Excessive Surface Winds
2. Excessive Winds Aloft
3. Turbulence
4. Icing
5. Thunderstorms
6. Cloud Layers/Thickness at Flight Level
7. Cloud Layer/Thickness below Flight Level
8. Precipitation
9. Reduced Slant Visibility
10. Reduced Surface Visibility due to Fog
11. Reduced Surface Visibility due to Haze
12. Reduced Surface Visibility due to Dust
13. Low Thermal Contrast

4. TAWS Performance

What detection and lock-on ranges were experienced?

Target:

Detection range: nm select

Lock-on range: nm

5. Comments

Please provide any amplifying details regarding the questions above. Also, what could have been done better or differently to improve METOC support to this mission? Please provide any other comments or recommendations you might have. Thanks very much for providing this information.

Figure 3. The online data collection form, similar to the paper data collection form. The online form was designed to be easy to use by having the same questions as the paper form in the same physical location. The form also hides the individual weather impact fields until the user needs them, resulting in a less confusing form.

1. Excessive Surface Winds

Negative METOC Condition	Mission Phase (Check all that apply)	Was Negative METOC Condition Forecasted?	Impacts to Mission (Check all that apply)	Corrective Actions Taken (Check all that apply)	
Excessive Surface Winds	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ Ejection limit ☐ Couldn't use glide weapon ☐ Other. Provide details in section 5 comment box below.	☐ None ☐ Change target/location	☐ Cancel mission ☐ Other. Provide details in section 5 comment box below.

Figure 4. Expanded weather impact field from the online data collection form. The check boxes are in the same layout as the paper form to allow for easy transfer of information.

These steps were accomplished with the assistance of the USS Enterprise OA division over a period of 2 months and resulted in the following lessons learned:

- Detailed understanding of both the METOC support process and the war fighter process is necessary to properly design the data collection portion of the system. In this study, we found that the CAOC, not the carrier OA division, is the provider of METOC support during mission planning. To fully track the impact of weather forecasts on missions, data from the CAOC must also be collected and analyzed.
- The shipboard METOC personnel have limited options for mitigation recommendations. Pilots have limited options for mitigating action.
- SIPRNet connectivity is not as reliable as we had hoped.
- PostgreSQL was too slow on the metrics calculations; PHP is a better choice for the calculation task.
- Collection of data from pilots was not as difficult as we had anticipated.

After the initial trials, we collected data from USS Enterprise for a short at-sea period, and then introduced the system to personnel on USS Eisenhower and USS Stennis. Due to the short data collection period, we were only able to perform preliminary assessments of the performance and operational impacts of the OA divisions' forecasts. These results show FAC for all forecasts of 99%, POD for severe negative weather conditions of 40%, and mitigation rate of 14%. These results are very similar to those obtained in closely related studies by Hinz (2004), Jarry (2005), and Darnell (2006).

4. Recommendations and Future Work

The next step in our METOC metrics for strike warfare program will be to transfer our system to the Naval Oceanography Enterprise for continued hosting and maintenance. Hosting the interface and database at an operational command will ensure continued data collection and encourage accrual of a more complete dataset that will accurately provide robust assessments of the quality of the support that the Navy METOC community gives to operational units.

Our metrics system was developed using a bottom-up approach. Our development of the system was encouraged by CNMOC, but its design was dictated by those who will use the system and those who will analyze the results. With the conclusion of this study there is now an extensive set of METOC metrics concepts and tools methods suitable for assessing METOC support for a wide range of warfare areas. Thus, METOC metrics research and development have reached a point where many METOC Units can participate, especially in routine data collection. We recommend that this be mandated and financially supported by the leadership of the METOC enterprise.

Resource shortages are rapidly pushing the METOC community toward an automated and integrated forecast creation and verification system, including a weather briefing generator, a forecast verification module, and a metrics module that tracks forecasts and mitigating actions during both mission planning and mission execution. The implementation of our data collection, analysis, and reporting system is an important step towards this larger goal.

Future work should extend our metrics system to incorporate both the planning and execution portions of the mission, including implementation of the system in the CAOC. This future work must collect metrics data from earlier in the planning process to better understand how planning forecasts impact mission generation, ATO creation, and, ultimately, mission success.

Future work should also extend this system to incorporate automatic capture of forecast data from a brief generator tool. Likewise, pulling observation data from existing automated sources should be pursued to make the observed weather data collection invisible to the pilot and the OA division. As much as possible, we should strive to make data collection a constant background process, so that we can get on to the business of putting the metrics to work to help improve both METOC support and decision processes.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. OVERVIEW

Corporations have developed quantitative analysis tools to identify how they can make their processes more efficient. These tools help them improve production numbers, cost-ratios, and ultimately, the bottom line by evaluating specific parameters of their operations processes and making targeted improvements. These tools developed by civilian industry are applied to individual processes to quantify their value to the company. Processes that demonstrate poor efficiency can be retooled, improved, or eliminated to save costs. The quantitative assessments that result of measuring and evaluating specific parameters, and quantifying the parameters' impact on end processes, are referred to as performance metrics, or simply *metrics*. This is the basic definition we will use when referring to metrics, and in this study we will apply these business-oriented ideas to measure specific parts of the U.S. Navy's strike warfare process.

With shrinking budgets and ever mounting criticism of government spending, the concept of improving process efficiency has found its way into the federal budget. The Budget of the United States Government, Fiscal Year 2007 has outlined a metrics related initiative started in 2004 called the Program Assessment Rating Tool (PART) (Executive 2006). In defining PART, the FY2007 Budget document said:

> The PART helps assess the management and performance of individual programs. With the PART, agencies and [Office of Management and Budget (OMB)] evaluate a program's purpose, design, planning, management, results, and accountability to determine its overall effectiveness. Recommendations are then made to improve program results.

PART encompasses most programs in the executive branch of the Federal Government, including the Department of Defense. Because of this, the concept of applying metrics to evaluate programs extends down to the Navy and the Meteorology and Oceanography (METOC) community within the Navy. The

Assistant Secretary of the Navy in a memo to the Chief of Naval Operations (CNO) stated “The [Department of the Navy (DoN)] must establish performance metrics that correctly reflect management and funding priorities … Recommended metrics beyond those proposed in [the listing of proposed DoN metrics] are encouraged. For optimal effect, metrics proposed by the DoN should be those most meaningful to each of you in monitoring and assessing performance in each of your functional areas of responsibility.” (Assistant Secretary of the Navy (ASECNAV) 2002)

LCDR Jake Hinz, while a master’s candidate at the Naval Postgraduate School (NPS), noted that the Commander, Naval Meteorology and Oceanography Command (CNMOC) had been unsuccessfully trying to measure the METOC community’s value to the Navy for over 10 years. CNMOC’s failures at previous metric efforts were traced back to a poor definition of metrics and attempts to link METOC support to operational customers through links that are “ill defined, too complex or do not exist at all” (Hinz 2004). Hinz (2004) set out to develop a prototype metrics system for the METOC community based on those developed by the National Weather Service (NWS). The NWS metrics program was chosen as a model due to the NWS’s similarity to the Navy METOC program. Both groups are government run weather services, and consequently have similar measurable processes. The NWS metrics process is based on metrics of success recognized by the end users of NWS products (e.g., emergency managers who rely on NWS products, such as tornado warnings). For these metrics, NWS identifies a goal (e.g., improved tornado warning lead times) and then measures how NWS forecasters are doing in achieving that goal (Hinz 2004). Hinz went on to define metrics:

> Performance measurement is the term used by OMB to describe the key processes used by the NWS and other agencies to measure their results based on previous performance baselines, projected goals, and the current status of management efforts to reach those goals. The quantitative results of these processes are called performance metrics. These metrics are then applied in management level decision making.

The system Hinz (2004) created, referred to as the NPS Metric Method, was used to provide information to CNMOC on the performance of their forecasters during Operation IRAQI FREEDOM (OIF) and contributions the forecast made to the planning and execution of combat missions. His system was based on one developed by the National Weather Service (NWS) and was adapted for use in tracking Navy weather forecasts through the use and development of performance metrics. His use of performance metrics, metrics that measure the performance of a forecaster's ability, laid the ground work for calculating how the forecaster's performance affected operations conducted by the Navy.

The Federal Government PART initiative has also resulted in a similar effort towards metrics by the Air Force. The Air Force metrics program is governed by Air Force Instruction 15-114, Functional Resource and Weather Technical Performance Evaluation 2001. This instruction outlines a metrics program with the goal of understanding the overall effectiveness and value of the Air Force weather system at all levels of supported operations. The program has outlined that mission execution forecasts, terminal aerodrome forecasts, and warnings should all be verified and the results tracked. Captain Jeff Jarry in 2005 and Major Karen Darnell in 2006 both did metrics studies, reviewed later in this chapter, that concentrated on the verification and tracking of mission execution forecasts while they both were students at NPS (Jarry 2005, Darnell 2006).

1. METOC Community Metrics

Due to the requirements imposed throughout the Federal Government and the Department of Defense (DOD) by the PART system, the Navy has a need for a fleet wide metrics program. The ability of the fleet to deploy and fulfill its duty is now tracked with the Ship Depot Maintenance metric which measures quantifiable elements of the ship's maintenance programs (DoN 2005). The Navy's tobacco cessation program is even tracked with metrics (Long 2003). The METOC community has also found itself in need of quantifiable performance metrics after having recently retooled its support for the war fighter (OPNAV

2006). The original structure was aligned along geographic boundaries, which often led to poor use of manpower and assets, whereas the new structure is aligned with warfare communities.

Most of the production of forecasts originally came from large, regional centers that produced all environmental support for units that did not have a ship's company METOC support cell onboard. In this case, one group of forecasters became regional experts, but their energy was spread through many different disciplines of combat. One METOC community member could have been tasked to create a flight plan forecast, ocean sound speed propagation profile, and a special operations wave and surf forecast during one watch cycle. This meant that forecasters needed to be proficient at many types of atmospheric and oceanic forecasting for their region. When a unit that was being supported by one regional center transitioned to another region, the unit would be handed off to a new center, often with a dramatic change on how support was provided. Additionally, the organic METOC division (OA division) on an aircraft carrier (CV) or an amphibious assault ship (LHA/LHD) reported to the CO of the ship and not to the METOC community. This led to non-standard support between OA divisions due to a lack of unifying command.

To resolve the shortcomings of having the METOC community aligned geographically, the community has been realigned by warfare types, sometimes called business lines. Nine warfare directorates have been established to better serve the customer. The nine warfare directorates are: Anti-Submarine Warfare; Mine Warfare; Naval Special Warfare; Fleet Support; Maritime Weather; Aviation Safety; Intelligence, Surveillance, and Reconnaissance (ISR); Navigation; and Precise Time and Astronomy. By aligning with a single warfare type, METOC personnel in each directorate can tailor their skill sets for the customers they are supporting. This reduces the time needed to learn many different disciplines while working at one assignment and frees up more man-hours for production of support products. The alignment with warfare types was implemented at the same time that a force reduction was implemented, but it is believed that alignment by warfare type approach will ultimately give better support (OPNAV 2006).

The Maritime Weather directorate has taken over the role of most of the previous ship's-company OA divisions, and the OA divisions' personnel are now assigned to two centralized commands, one in San Diego and one in Norfolk, called Strike Group Oceanography Teams (SGOTs). The members of the divisions will report to a Commanding Officer (CO) in the METOC community and will be attached to the CVs and LHA/LHDs in a temporary duty status during a major deployment and associated training cycle. This allows for centralized meteorology and METOC support training, a fluid manpower pool, and eliminates the inefficient use of forecasters being assigned full-time to ships in dry-dock and in maintenance availabilities. The result of the SGOT, again, is fewer people giving better support.

The METOC community realignment raises two questions: (1) Can the realigned METOC community provide quality weather support to the war fighter? (2) If the METOC community was poorly aligned before and no one suffered, does that imply that the community does not provide a tangible benefit? The answer to both of those questions can be found through the same source: the metrics methods developed by Hinz (2004) that were originally intended as a response to the Federal Government's use of PART. By analyzing the forecasting process using metrics and linking it to something tangible, such as impacted missions, the value of the METOC community to the war fighter can be measured.

2. The Importance of Metrics Today

The use of metrics to analyze forecast value has importance beyond justifying budgets. Some of the important functions of METOC support to DoD operations are to enable and enhance combat effectiveness, to protect the life and property of US personnel and assets, and to ensure the safety and effectiveness of peacetime operations such as training and humanitarian assistance. All of these functions, or purposes, imply that excellence in METOC support will help protect lives and other resources. Metrics are needed to help determine how well METOC community fulfills those functions, and how the community can fulfill those functions better.

The way we fight wars has changed, and the way in which the METOC community supports wars has also changed. Sometime between the Korean War and the Global War on Terrorism (GWOT), wars have gone from being defined by large armies and theater-wide fronts to being associated with guerilla tactics, urban warfare, and small scale operations. The use of weapons has changed from carpet bombing to a combination of precision weapons and the extensive use of intelligence. Along with this shift in the war fighting paradigm, there has been a need to change the way METOC support of warfare is provided. Forecasts have gone from broad regional horizontal weather depictions and prognostic blends to specialized products, such as lunar illumination predictions and ISR cloud coverage forecasts that support very specialized missions.

Modern forecasts must provide improved depictions of mesoscale and microscale phenomenon. Broad overviews of the weather neglect small, but critical, effects that can cripple modern warfare operations. Average winds over a region do not accurately represent localized effects that can include higher winds and dust lifted from the surface that would wreak havoc on laser guided munitions. Using open ocean sea swell height for a coastal region could ruin a SEAL insertion onto the beach if breaking wave height was unanticipated. Specialized tactics and weapons are often more sensitive to environmental impacts than conventional weapons and forces. Due to the large scale and large variability associated with traditional operation area forecasts, tailored products are needed for small, sensitive operations. Given the narrow band of operability for many sensitive operations, to have poor specialized support can be as bad as or worse than having no support.

This leads to back to the other benefits of metrics beyond those associated with budgeting. Since it is no longer useful to categorize an entire theater of operations with one blanket forecast, more man power is needed to provide specialized, tailored forecasts. METOC personnel now have to spend more man hours understanding the needs of many specialized customers, whereas before they could provide a more generic support product. This extra support costs extra resources: not only must the cost versus value be justified, but the resources must be spent as efficiently as possible. The ability to measure and improve specific mission support, such as strike operations or humanitarian assistance requiring beach operations, will save lives and could serve to validate increased expenditures associated with tailored support, if it can be shown that the more expensive tailored support is needed.

B. BACKGROUND

In this section we will review several studies that were conducted as part of the NPS METOC Metrics Project, and that have examined ways of statistically linking forecasts, weather impacts, and costs. The logical progression of these studies begins with work on the costs of METOC related mishaps (MRMs). Later studies brought together lessons learned from the PART concept and the MRM work to find statistical links between forecasts and operational success by the war fighter. The most recent studies have explored the concept of collecting operational metrics in real-time and exploiting them as a scorecard for operational weather support.

1. METOC Related Mishaps

The first two studies of the NPS METOC Metrics Project laid the groundwork for understanding the linking weather forecasts to operational impacts. Specifically, LCDR Ruben Cantu identified the role of weather in Class A weather mishaps (Cantu 2001), and LCDR Brett Martin sought to identify the importance of METOC conditions in conducting operational risk management (ORM) (Martin 2002). The goal of both of these studies was to identify how weather, and what kind of weather, contributes to mishaps, and how much MRMs cost the military. The studies did not intend to collect metrics data, but instead they used data collected by other sources. Cantu (2001) sought to identify which Class A mishaps (mishaps that result in a loss of over $1,000,000, loss of aircraft, or death/permanent total disability) from 1990 to 1998 were MRMs. Martin (2002), on the other hand, studied Navy ORM practices and their ability to reduce costly mishaps through managing risk using all available information. Martin reviewed all Class A, B, and C afloat mishaps from March of 1997 to March of 2002 and found that METOC factors contributed to a number of afloat mishaps. The result of both studies was that a significant number of mishaps are related to METOC factors. The main recommendation from both studies was also very similar: improve forecasts and provide the customer with a better understanding of the operational significance of the forecast to reduce mishaps.

2. Operation Based Metrics

Later studies focused on other types of forecasts besides operational impacts. By doing statistical analyses of operations during OIF and applying lessons learned about the value of accurate and believed forecasts, LCDR Jake Hinz and Capt. Jeff Jarry were able to develop metrics describing the impacts of weather and weather forecasts to combat missions.

Hinz (2004) applied NWS metrics to the Navy forecasting problem. His goal, as mentioned earlier, was to apply lessons learned during the NWS's implementation of the PART initiative to quantify the value of METOC support to operational units. He applied a variation of the NWS metrics concept to create a series of performance metrics he called the NPS Metric Method. His performance metrics focused on evaluating forecaster performance independent of mission impacts. The core element of the NPS Metric Method was the adaptation of two key calculations from the NWS for use in Navy METOC applications. The first calculation was forecast accuracy (FAC), which is a comparison of the number of correct forecasts for a certain event compared to the total number of forecasts for that event (Hinz 2004). This provides an objective measurement of the forecasters' ability to predict a category of events. The FAC measurement has a down side in that a forecaster could miss a rare event consistently and still have a high FAC score. In this case, the high FAC score conveys no information about the forecaster's ability to predict the rare event (Hinz 2004). For this reason, Hinz (2004) also applied a probability of detection (POD) calculation to provide a check and balance. POD represents the forecasters' ability to predict an event before it happened. The POD calculation, when combined with a FAC calculation, would prevent forecasters' scores from over-representing their ability if they were to hedge a forecast by not forecasting a particular event or had trouble predicting a particular event. Hinz then applied his metrics method to historical data collected at the Combined Air Operations Center (CAOC) and other forecasting entities involved in the early stages of OIF (Figure 5-6). The main data set that Hinz used was incomplete and inconsistent for the purposes of his study. This is in part because the data collection was not designed to support specific metrics analyses.

Jarry (2005) focused his efforts on an analysis of data collected from the Air Mobility Command (AMC) and its Combat Weather Teams (CWT) during fiscal year 2004. His goal was to advance the application of metrics to include operational impacts metrics. Operational impacts metrics describe how forecasts affect the planning and execution of operations. Jarry developed the missions saved metric that measures the number of missions that were successful because the mission was changed in response to a forecast of adverse or negative weather conditions (Figure 7). This gave Jarry the ability to show when Air Force Weather (AFW) was able to provide good decision making inputs to the war planners. By applying missions saved and variations of missions saved, Jarry was also able to achieve a measure of the value of weather forecasts to the war fighter. However, Jarry ran into problems similar to those encountered by Hinz. Jarry also used a historical dataset, and his data was collected by the CWTs per the direction of AFW. His dataset was more complete, but he encountered problems with inconsistent collection procedures and different interpretation of the collection by different organizations. Again, these problems were due mainly to the data collection not being designed to support the metrics analyses that Jarry conducted.

Both of these studies developed and applied metrics to show the correlation between good forecasts and mission successes, bad forecasts and mission failures, and how operational impacts vary from one forecast to another depending on the forecast accuracy and the mission supported by the forecast. The recommendations from both studies were very similar: adopt a service-wide metrics program to measure the value of weather support to operational units. Both studies reached similar conclusion on procedures for putting together a metrics program. Both studies suffered from difficulties with their historical datasets, and both were unable to give their metrics results until several months after the missions were completed. Their difficulties with datasets and long delays in issuing results highlighted the need for a standardized collection program and the ability to compute results in near real-time.

3. Real-Time Metrics

The most recent stage of NPS metrics work has sought to apply the concepts developed by Hinz (2004) and Jarry (2005) to the real time collection and analysis of data on current operations. By collecting and analyzing mission data in near real-time, forecasters are able to grade their forecasts, and war fighters can better understand the value of the weather forecast as it applies to mission planning and execution. LCDR Mark Butler was the first to develop such a metrics system with his system to collect and analyze data collected at a training command. Butler (2005) made the first logical step to collect data on forecasts issued by the Naval Pacific Meteorology and Oceanography Detachment, Fallon, NV (NPMOD) for training missions flown at the Naval Strike and Air Warfare Center (NSAWC). His goal was to develop a system, based on the Hinz (2004) and Jarry (2005) concepts, which would collect forecast and mission data from NPMOD and NSAWC and provide results of the metrics calculations in near real-time. Near real-time metrics enable analyses that can be used at the end of the day, instead of months later, to help quantify value and improve METOC support. Maj. Karen Darnell (2006) implemented a version of the Butler (2005) system to analyze forecasts and impacts to operations at CWTs supporting Air Force fighter aircraft missions. Both studies used concepts from both Hinz (2004) and Jarry (2005) to develop a set of performance metrics and operation impact calculations to meet the needs of their customers. The Butler study was able to successfully provide real-time metrics analyses to the forecasters and the customers by having the forecasters collect forecast and impacts data for the calculation. NPMOD found the data extremely useful, and uses it to help NSAWC planners prepare future training missions. The Darnell (2006) study was a first attempt to provide real time metrics support for real world operations. A working system was successfully put in place for Air Force Weather (AFW). There were some problems with the data collection methods, but those problems provided some valuable lessons. The data collection process from the Darnell study relied on a long, four page questionnaire given to the pilots to assess the forecasts and the impacts (see Figure 5 of Darnell (2006)). The questionnaire proved to be too ambitious for the current life stage of implementing metrics into operational use. The recommendations for the projects were also different. Butler (2005) concluded that continued use of the collection system at NPMOD and NSAWC would continue to familiarize the forecasters that go through the training with the pilots, and prepare them to collect metrics in the fleet. Additionally, he recommended that the system be implemented outside of training commands. The Darnell (2006) study recognized the shortcomings of the information collected in that study and recommended that the data collection process be revised to better support the CWTs.

C. METOC METRICS NEEDS

We have reviewed the need for a Navy METOC community metrics program from several perspectives: (1) the federal government requirement outlined in PART that supports tracking the value of the Navy METOC community to its customers; (2) the need to quantify the benefits that more expensive tailored METOC support can bring to modern warfare; and (3) the need to use to improve the process of producing accurate forecasts and the process of using forecasts in the mission planning process. A standard set of Navy METOC community metrics will help rate the forecasting skill of the community and of the individual units, it will track the value of the community to the war fighter, and it will assist the customer in their mission planning and mission evaluation. For example, the customer can weigh the value of a forecast in their planning by knowing the accuracy of different types of forecasts. If the customer knows that the forecaster accuracy (FAC) for high winds over the target area tends to be near 100%, and high winds are forecasted during the mission time, the customer can weigh the forecast appropriately in planning. Conversely, customers can choose not to act on a forecast for a less predictable event if circumstances warrant. Ideally, any customer could find out the skill with which any individual phenomenon tends to forecasted and be able to immediately identify the value that the forecast information adds to the planning process. Implementing a metrics system is not without its hurdles. A lack of manpower, leadership, and interest has likely stymied many previous efforts. Current efforts to implement a metrics program also have obstacles, such as garnering participation, lack of understanding of the usefulness of metrics, and limited access to critical information needed to complete metrics analyses.

1. Standardization

It is important to standardize the METOC metrics process throughout the entire METOC community. The metrics programs we have developed in this study relate to forecasting for air-strike missions, but this concept can be applied throughout the forecasting spectrum. If senior leadership becomes familiar with a standard type of valuing system for Navy METOC forecasts, then decision makers can make better use of forecasts across different warfare types. This would be invaluable in large scale operations requiring multiple types of forecasts, such as emergency non-combatant evacuation operations, where air-strike is working in conjunction with Marines on the ground, landing craft in the surf-zone, small boat operations, and naval gun fire support.

2. Navy Benefit

Once a standardized system of measuring METOC forecasting performance and operational impacts is implemented in the Navy, the results can be used to monitor how the METOC community is doing and to identify how it can improve. First, the data from the metrics program can be used within the Navy METOC community to track and identify shortcomings in forecasting skill.

Second, metrics can also help quantify the value of the METOC community to the war fighter by directly linking mission performance to mission impacts that were predicted by the forecaster. The linking of mission performance and forecasted mission impacts allows the METOC community to identify ways in which the METOIC community can better support war fighters. Finally, decision makers can use metric data to better weight forecasts in their decision making processes based on the past performance statistics of similar forecasts.

The first identified benefit, tracking the forecasting skill of individual units, or of particular forecast parameters and situations, could be a great educational tool if used correctly. It would help the Navy METOC community identify training areas and forecast challenges that need improvement. Measuring how well different regions of the world are supported would give insights into the Navy METOC community's understanding of that region. Once a representative data set for forecasting skill is available from a Navy METOC metrics program, CNMOC could implement improved and focused training to resolve any forecasting shortfalls that have been identified through the metrics program. The temptation must be resisted to rate individuals against their peers. Not only do the chances for corrupted metrics data exists when the data is being collected from those who are being graded, but it must be remembered that it is the entire METOC support system that is producing the forecast, not just the mental processes and work habits of the individual forecasters. The presence or lack of observations, the capabilities or limitations of the numerical model, and even the timeliness or latency of communications has effects on the forecast *grade* that may outweigh the effects of having the best forecaster working the weather impacts problem.

The second benefit, the ability to track the value of the Navy METOC community through the eyes of the war fighter using a metrics program, is easier to grasp than any other benefit. By incorporating the missions saved metric from Jarry (2005) and the idea of potential saves from Butler (2005) and Darnell (2006) into a METOC community metrics program, a number of values that are understandable to non-meteorologists can be created. Missions saved can give pilots and mission planning cells an idea of the mitigation value provided by the METOC unit. The missions saved numbers can also be somewhat translated into dollar amounts that can have an impact on budget decisions (Jarry 2005). By counting both missions that have been changed due to a forecaster recommendation and forecasts that were not changed and unsuccessful, a rough dollar “value” can be assigned to the METOC community’s forecasting. The value of a mission saved would account for fuel costs, maintenance costs from added hours on equipment, lost man hours from work supporting the mission, and lost weapons from a mission attempt (Jarry 2005). This calculation of the value of a mission saved would be imperfect, given the large number of unknowns regarding mission costs, but the rough monetary value of the Navy METOC community could be at least estimated.

Improved mission support and forecasts are clear benefits to customers that could be achieved through a METOC metrics program. The less obvious benefit to the customer is in improved understanding of how to weight a forecast against other planning inputs. Mission planners, and other high level decision makers, operate by weighing all of the inputs that come to them through various channels. Each input the decision maker receives is information that affects the uncertainty of the decision to be made. But it can be very difficult to quantify the amount of uncertainty associated with each input. Decision makers need to know how much uncertainty exists in all the inputs, including the forecasts, in order to optimize their decisions. Forecasts should not be considered completely accurate; rather they should be considered as providing probability information. However, often times they are considered by decision makers as either *absolute truth* of future environmental conditions or they are considered worthless and ignored completely. A well executed metrics program could help both forecasters and decision makers understand the probabilistic nature of the forecasts, and shed light on which forecasts are of greater value and which ones are of lesser value. Using metrics such as POD or FAR (Hinz 2004, Jarry 2005), the chance of a certain phenomenon being detected can be estimated and a confidence value can be assigned to the forecast inputs. The mission planner will then have a better understanding of the uncertainty involved in the decisions to be made and therefore be better equipped to make the decisions (Environmental 2003).

3. Past Challenges

There were attempts prior to 2001 to create a METOC metrics program, but those efforts were poorly documented and had almost no lasting effects. Large scale efforts to create a METOC metrics system beyond those outlined earlier in this study do not seem to have taken place, and it is the large scale efforts that are needed to reach across the CNMOC claimancy as a whole. Individual commands have been required to maintain forecast verification programs, but it was to be of their own design and administration. These programs disappeared for the most part when the command inspection requirement was cancelled.

The large scale efforts needed to create a metrics system have likely not existed because of a series of suppressing factors. The first factor was a lack of man power. Somebody would actually have to do the work to create and develop the system. In that case the Navy would either have to use sailors that are already tasked to their limit, or would have to have procured funding for an outside agency to complete the metrics program development. The second likely hurdle was lack of expertise on how to develop and implement a metrics program. Most people in the METOC community have little or no education or training on metrics in general or METOC metrics in particular. The third likely hurdle to a large scale metrics system was a lack of centralized control. The lack of a higher entity directing a specific metrics effort resulted in ambiguous verbiage calling for subordinate commands to “improve output metrics to better define our requirements and resource needs” (Clark 2004). Without clear, unifying direction there was no chance that a standard METOC community wide metrics system would be developed. The fourth likely hurdle was a lack of broad-based interest for a METOC metrics program. We have identified a pattern in the METOC community that unless a customer specifically complains about service, the community assumes all is well. Unless a customer actually asked for performance metrics, the chances of metrics being created would have been very slim. The four hurdles outlined here have been partially overcome in a series of NPS studies funded by the Office of Naval Research (ONR) and Space and Naval Warfare Systems Command (SPAWAR). Funding in the form of grants, and manpower in the form of research students and contract employees, has been provided to do the work. Control of the project is being handled by a forum of Navy METOC leadership and NPS professors. Interest in METOC metrics has developed both in the Federal Government in the form of the PART initiative and by customers of Navy METOC (Woll 2006).

4. Current Challenges

With many of the past organizational barriers resolved, the challenges we expected to face in our project were mainly of an operational nature. We anticipated an initial struggle to garner participation, both by METOC units and customers. Without fully realizing the benefit that forecasters and customers stand to reap, participation in a metrics program would be viewed as a wasteful addition of work. Forcing participation could have resulted in bad data that would be hard to filter from good data, so the best solution was education about the aforementioned values of a METOC metrics program. We also anticipated problems regarding communications connectivity. Butler (2005) recognized the need for a metrics system to be centralized to allow for standardization, maintenance and upgrading. But this centralization requires that participating units have continuously reliable internet access to reach back to the metrics processing computer. When NPMOD Fallon was using the system, internet access was not a problem, but as we branched out to other units in the fleet, we found that unclassified internet access was not as reliable. Finally, we anticipated problems with training on use of the metrics system. We would need to ensure that only quality data was collected, but without extensive travel and time, it would not be possible for the main participants of this study to train all potential users. We concentrated on these three hurdles in this study while working towards our goals.

D. GOALS OF THE STUDY

There were three goals that this study intended to meet. The first goal was to adapt the metrics collection and analysis system developed by Butler (2005) to operational use for strike warfare support and for eventual use across a broad range of warfare types. A second goal was to implement this system on an operational CV. The final goal was to conduct an initial evaluation of the accuracy of METOC products provided in support of strike warfare and the operational impacts of those products on strike warfare.

The Butler (2005) study showed the feasibility of collecting forecast and impacts information and providing near real-time metrics calculations. His system was focused on collecting and analyzing data from an air wing training facility that operates at a limited capacity and on a more predictable schedule than do carrier based air wings, especially those in combat zones. But we determined that with some adaptation, the Butler study could be used for war fighter units conducting in real world (non-training) operations. The types of forecasts, missions, impacts, and mitigation procedures used in these real world operations are different than those used during training missions, but there are many similarities. The greatest challenge we anticipated was providing accessibility to operational units. The Butler (2005) study collected data from a training command with a fixed unclassified internet connection. In that study, the information being collected was unclassified by its nature, and there were no concerns that the central metrics system would not be accessible. In our study, however, the information was classified due to its operational association, and the internet was not always accessible while the ships were at sea. For these reasons, we had to adapt our system for use on the Secure Information Protocol Routing Network (SIPRNet) which is a government internet system that can pass classified information between users. The SIPRNet is certified to handle information up to a secret classification, and the network is generally available at all times while ships are at sea due to the high volume of operations critical traffic that is passed through it. There are always times that a network is down at sea, and for that reason a sub-goal was to produce a set of paper forms that collect the same information as our web form to be used in the event that the SIPRNet was not available.

The second adaptation we made to the Butler (2005) system adaptation was to ensure that it can be easily applied to a broad range of warfare types. Our initial application of the system was to be strike warfare. To facilitate standardization throughout the service, we wanted to design our system to look very similar when applied to amphibious warfare, undersea warfare, and other warfare types. As mentioned earlier, this standardization will lead to more efficient use of METOC metrics.

After these adaptations of the Butler (2005) system were achieved, our second goal in this study was to implement our system on an operational CV. The operational use by a deployed unit would give us feedback on our system and allow us to make corrections and fine tune it.

Our final goal was to conduct initial evaluations of the performance of METOC products provided in support of strike warfare and the operational impacts of those products on strike warfare. Implementing the collection system on a CV would also allow us to do initial calculations of forecast performance and operational impacts to determine the operational benefit of METOC support to strike operations. These calculations would allow us to conduct an early check on whether we were making the right calculations and looking at the right data. We hoped to provide through our study a tool set that can be used to: (1) measure the effectiveness of the Navy METOC system; (2) demonstrate the contributions Navy METOC provides to the war fighter; and (3) recommend ways that Navy METOC can be improved. .

The following chapters outline how we worked towards these goals. We start by outlining our data and methods of working with the data in Chapter II. We then look at the results of implementing our system and the results of the data calculations in Chapter III. Finally, we summarize our study and provide recommendations in Chapter IV.

II. DATA AND METHODS

A. DATA

Based on the recommendations for future work by Butler (2005) and his metrics work implementing an automated data collection and analysis system for NSAWC, the next goal in the overall NPS METOC metrics project was implementation of a metrics program on a U.S. Navy afloat unit. Our first step towards the goal of an afloat metrics program started with CDR Steve Woll, the CO of SGOT East in Norfolk, Virginia. CDR Woll held a similar vision of implementing a comprehensive metrics system onboard afloat units.

The ideal METOC metrics system for strike warfare is one that collects data on early mission planning forecasts given to Air Tasking Order (ATO) planners, collects mitigation recommendations given to planners based on poor environmental conditions, and tracks the mitigations taken by the planners before promulgating the final ATO, and any revisions before or during the ATO execution window. The ATO is a schedule of all flights and air combat missions in a specific theater or under a specific commander. The metrics system would track the forecasts given to the pilots conducting the missions, their observations during the mission, and mitigations taken by the pilots if there were adverse METOC conditions. The comprehensive nature of this ideal metrics system would eventually paint a very complete picture of how the Navy METOC community provides value to the whole of strike warfare (Woll 2006). We and CDR Woll hoped to quantify the value of Navy METOC to both planners and budget makers, information about observed METOC impacts to strike missions, the ability to forecast these impacts, and the mitigation of these impacts.

We started our study with the intention of tracking the planning forecasts given to the ATO planners, but we learned that the OA division’s role is currently to provide only mission briefs to the pilots conducting the missions on the ATO. The OA division does provide larger scale and longer lead-time weather briefs to other planning and intelligence personnel on the ship, but not to strike planners, and these longer lead-time briefings tend not to be mission specific. With an established pilot briefing and de-briefing system already in place, we re-focused our efforts on tracking the forecasts given to pilots and mitigations taken by pilots. Thus, the focus of our metrics development efforts was on the performance and operational impacts of the mission execution weather forecasts (lead times of 12 hours or less).

Future METOC metrics studies should focus on the performance and operational impacts of the forecasts provided to the CAOC (for more on this, see chapter IV). One of CDR Woll's goals is to increase the involvement of Navy METOC forecasters by incorporating them into the planning phase of missions. Involvement at the planning phase of a mission requires increased information flow among the OA division, the CAOC, and the CV Intelligence Center (CVIC) on the CV during the ATO planning phase. Changing the way ATO planning is done to accommodate increasing OA division involvement would require a large amount of diplomacy, planning, and time (Woll 2006, Schmeiser 2006).

The work CDR Woll has done with earlier stages of the current NPS METOC metrics project and his position within the METOC community allowed him to identify appropriate participants for our study. Based on ship schedules and the willingness of the ships' OA divisions to participate in our study, we decided to focus our initial development efforts on the aircraft carrier USS Enterprise, followed by a later data collection efforts on the USS Enterprise, USS Eisenhower, and USS Stennis.

We began by identifying the role of the OA division on a carrier and what products they provide to their customers. Once we knew the METOC support being given to the customers and how it was being used, we developed a data collection strategy that identified the data to be collected and any pitfalls that might be encountered during the collection process.

1. OA Division Role

Before we could collect data on how METOC forecasts affected strike operations, we needed to know what was provided by the OA division to the customers. We had to identify who the customer was, what products were produced, when they were provided to the customer, and how the customer used them.

Our original intended customer was a mission planning cell at the aircraft carrier level. We originally believed that the air wing command that had cognizance over all the aircraft embarked onboard the CV had a mitigation capability that could change missions in reaction to forecasted environmental conditions. By comparing the forecasted weather and mitigation recommendations given to the onboard air wing planning personnel against mitigated and non-mitigated missions, we could have a clear understanding of the value of the weather forecast and mitigation recommendations to the planner. That planning forecast and mitigation process is used at NSAWC and was analyzed by Butler (2005) who found it to be very useful in providing support to NSWAC planners. The Hinz (2004), Jarry (2005), and Darnell (2006) studies also analyzed similar types of planning forecast and mitigation recommendation customer support.

However, during the planning phase of our study we learned that the ATO is produced at the theater CAOC and then sent to the participating air wings operating in the theater, so that the embarked air wings are not heavily involved in the ATO planning. In most cases, the weather cell at the CAOC provides weather for the purposes of planning and writing the ATO (Schmeiser 2006) along with weather used for mitigating weather impacts on missions on the ATO. Once the ATO is received on the ship, only minor changes can be requested by the air wing. These changes are generally limited to changing the weapons carried by the aircraft, canceling missions, and in some cases altering the timing of the mission (Ford 2006, Schmeiser 2006).

Since the OA division was not forecasting for the planning of missions or for the mitigation of missions, we brought up the question as to what the OA division was forecasting for. Conversations with LCDR Gregory Schmeiser, the Executive Officer of SGOT Norfolk, revealed that “we give value added [mission information] to the pilots… Our value added is the situational awareness about the environment... We give safety of flight forecasts, tactical sensor predictions, etc.” (Schmeiser 2006). According to SGOT Norfolk, the OA division’s customer on the aircraft carrier is the pilot and the OA division provides a forecast of the weather that the pilot will be operating in.

With pilots currently as the main customer of the OA division, we were unable to develop metrics on the usefulness of planning forecast. Since there was a limited mitigation capability on the ship, there would be little opportunity to calculate some of the valuable metrics developed in prior studies such as missions saved (Jarry 2005, Butler 2005, Darnell 2006). Working with the pilots as the main customer required us to focus on how mission execution forecasts improve the ability of pilots to complete their mission, knowing that the forecasts come too late to have much impact on mission planning. As mentioned above, LCDR Schmeiser outlined that the pilot uses mission execution forecasts mainly for situational awareness (SA). This is consistent with the Air Force Weather Strategic Plan and Vision (AFWSPV) that states, “if you ‘boiled down’ the AFW business to its essence, you would find that we *integrate* into operations and intelligence, that our analyses provide *battlefield situational awareness*, and our predictions and tailored products enhance *decision superiority* for commanders at every level” (AFWSPV 2004). However, it is difficult to quantify the impacts of SA on mission execution. We presume that these impacts tend to be beneficial, but they tend to be somewhat intangible. The concept of intangible benefit from mission weather forecasts briefed by a forecaster was also outlined by Darnell (2006). In interviewing Air Force pilots, she learned that the pilots’ SA benefited greatly by having weather personnel available in-person to explain expected atmospheric conditions (Darnell 2006). SA is a hard entity to quantify, and we have made no attempt to do so. We can, however, intuitively understand the value of good SA to the pilot, and note that awareness of how weather might affect the route or area is acknowledged as an industry standard. Implicit in the regulations concerning receiving and filing of weather reports (FAA, 2006) is the assumption that situational awareness of weather by pilots is absolutely necessary. Although we cannot measure SA, we can measure whether forecasts provide accurate SA to pilots.

The pilots and strike leads are provided a flight plan weather brief by the OA division, generally 1.5 to 2.5 hours before the planned time of the mission. The weather brief, as shown in Figures 9 and 10, consists of information about several parameters, including flight level winds, icing, turbulence, ceilings, visibility, and atmospheric conditions that would affect sensors. The pilot uses the forecast to mentally prepare for a mission and to visualize what to expect during the mission. A good forecast reduces the range of conditions for which to prepare and allows the pilot to mentally rehearse options that can be used to alleviate a potential impact. As an example of how SA is perceived as valuable, consider the case of forecasted turbulence. If the pilot knew in advance that he/she was going to fly through an area of light to moderate turbulence, he/she probably would not change the intended mission track, nor would the forecaster recommend that the pilot change the track. But the pilot would mentally prepare for the possibility of encountering turbulence and anticipate how he/she would mitigate it if it was actually encountered. Thus, if moderate to severe turbulence was encountered, the pilot would be better prepared to take mitigating action, such as altering altitude, speed, or course. Knowing about the turbulence didn’t cause the pilot to avoid the impacting phenomenon, but it allowed a quick, preplanned action. Most importantly, the impact was not a surprise that resulted in several minutes of distracted thinking in a combat zone.

In a situation such as described above, there is an opportunity to collect mitigation data, although in this case the mitigating action occurs during mission execution rather than during mission planning (the type of mitigation examined by Jarry 2005, Butler 2005, and Darnell 2006). In our study, we were able to collect data on the major actions taken by pilots during mission execution to reduce the negative weather impacts on their missions. This data included information on what was predicted, what was encountered, and what mitigating actions had to be taken. The benefit of metrics based on this type of data is that they show when pilots were expecting negative impacts, mentally prepared for the impacts, and took mitigating action during the mission. Note that there is no implication in our data or analyses that the pilot should have changed the mission before hand.

So, despite the limited role of the OA division during the planning phase, we were able to devise data collection and calculations that measure benefit to strike operations. With the pilots identified as the primary customers of the OA division, we were able to measure the accuracy of the flight path weather forecasts and to infer their value to the pilots' SA in the combat zone. SA is difficult to quantify and collect data on, but we were able to measure inputs to SA and to quantify responses to weather impacts where accurate SA is believed to be useful.

2. Data Collection

After understanding the support provided by the OA division and how that support is used by the OA division's customers (the air wing pilots), we determined the data that we would need to collect to assess the performance and the operational impacts of that support. The data had to include information about: (1) the forecasted and observed weather phenomena that could lead to negative impacts; and (2) the mitigating actions taken by the pilots. The data needed to be collected in as consistent, objective, and accurate a manner as possible. Finally, the data needed to be put into a format that allowed basic metrics calculations to be performed.

The data needed for this project had to: (1) be relevant to missions flown by strike aircraft; (2) known in advance and consistent for all missions; (3) include information about the forecasted and observed weather conditions; and (4) include information about what the pilots planned to do and what they actually did. Additionally, we did not want to include any weather parameters beyond those already being forecasted by the OA divisions. In discussions with members of SGOT Norfolk, we decided to collect forecast and observational data on the following nine METOC parameters:

1. Surface winds
2. Winds aloft (up to 30,000 feet)
3. Turbulence
4. Icing
5. Cloud cells that are associated with thunderstorms
6. Cloud layers
7. Visibility
8. Precipitation
9. Thermal contrast

Data on the forecasts of these parameters were collected for five different mission phases: (1) aircraft launch; (2) enroute to and from mission area; (3) enroute refueling; (4) target area; and (5) aircraft recovery. As standard procedure, the OA division also recorded the observed values for the nine parameters during launch and recovery, except for icing and turbulence. The OA division was not already collecting observations from the three other mission phases. For observations of the conditions during these three phases, we relied on pilot observations, in particular, pilot observations of negative conditions associated with the nine parameters during phases 2-4.

Obtaining data proved to be one of our larger challenges. An ideal data set would give an analyst enough information to fully understand an environmental situation without having been in that environment. The problem with obtaining the ideal data set is that generally in real world operational settings, accurate observations of enroute and target conditions are hard to obtain. The pilots of course observe the conditions as they fly through them. Although they are not meteorological observers by trade, we decided to use their observations as the best available observations. This was based in part on informal analyses of by Darnell (2006) that indicated that pilots are relatively good observers of weather conditions, especially conditions that have a negative impact on their missions.

A major challenge to creating a data set for the purposes of our study was finding a balance of enough data to accurately describe what happened in the environment without overtaxing the pilots. Darnell (2006) observed that her “data collection form was too long and therefore aircrews were reluctant to complete it. The form was also not structured in a flow consistent with aircrew thought processes after they had just completed a mission.” The Darnell (2006) pilot mission debriefing forms, seen in Figure 8, asked many useful questions and provided data that could be used in a variety of analyses. The form lengthiness, however, was a liability. Hinz (2006) reported that the collection of pilot data by OA division personnel on the USS Nimitz was made much more effective and efficient by having an experienced AG attend the intelligence debriefings of the pilots. At these debriefings, the METOC data collector would mainly collect the needed METOC and operational impacts data by listening to the intelligence debrief and quickly making hand written notes that were later entered into an electronic database.

Based on these prior efforts, we decided to carefully limit the amount of data we would try obtain directly from interviewing pilots. Limiting the questions asked of the pilots would limit the chances that incorrect data would be given to the weather debriefer, and would increase the chances that pilots would be willing to cooperate in providing us with data. To keep the post-mission interview short, we decided to ask the pilots three basic questions after each mission. The first question asked was whether there were any inconsistencies between the forecasted weather phenomena and those that negatively impacted their missions. The second question was what, if any, part of their mission was negatively impacted by weather. The third question was whether they had to take any mitigating actions because of weather. Pilots generally will remember if they experienced weather that negatively impacted their missions along with any unplanned actions that they had to take because of the weather (Schmeiser, 2006). We learned that by asking pilots what they experienced that affected their mission they were more likely to participate than if we asked them to recount all the weather they experienced during the mission. To simplify the job of the OA division, we developed a series of questions for use in collecting data during pilot debriefings. It provided consistent wording and eliminated any variability that might have been caused if the pilots' observations were recounted in an informal, unstructured interview.

The final prerequisite was that the data be compatible with the analyses and metrics calculations that we wanted to conduct. Many forecast performance metrics are based on the use of a contingency table similar to the one in Figure 11. The contingency table allows for easy categorization and grouping of forecasts. The number of forecasts in each of the four categories (A-D) of the contingency table is used to calculate a wide range of forecast performance metrics (cf. Jarry 2005). Values in the A-D diagonal are forecasts that verified as accurate. Values in the B-C diagonal are forecasts that verified as inaccurate (cf. Jarry 2005). More in depth metrics can also be calculated using comparisons of the different values in the table, and these calculations are described in the methodology section of this chapter.

In order to be able to use contingency tables, the forecasts have to be recorded in the appropriate table category. Generally this assignment is based on whether a specific condition (e.g., rain) was or was not forecasted and was or was not observed. If the condition was forecasted and was observed, then the forecast is assigned to category A. If the condition was not forecasted but was observed, then the forecast is assigned to category B. If the condition was forecasted and was not observed, then the forecast is assigned to category C. If the condition was not forecasted and was observed, then the forecast is assigned to category D. To determine whether a condition was or was not forecasted, and was or was not observed, thresholds are used (e.g., minimum amount of precipitation, if rain is the condition).

For our study, we modified this general method due to the logistics of obtaining detailed meteorological observations from pilots in combat zones. We considered using Go/No Go recommendations provided by forecasters as the basis for our thresholds (cf. Hinz 2004, Jarry 2005). Go/No Go refers to forecaster recommendations that a mission be attempted or abandoned based on the predicted weather conditions. We had to set that idea aside when we leaned that the OA division does not make Go/No Go recommendations (Schmeiser 2006).

The OA division's role is to provide forecasted weather conditions to the pilots. The pilots then use the forecast to make their own execution decisions based on published METOC thresholds for the mission platform, sensors, and weapons. With inputs from LCDR Schmeiser at SGOT Norfolk, we collected forecasts from the OA division according to whether they indicated severe weather impacts to the missions. Severe impacts are those that prevent a mission from being completed without taking mitigating actions, and are generally categorized as a red condition on a red-yellow-green stoplight chart. They can be such events as visibility ranges that prevent use of laser guided weapons, winds that prevent the launch and recovery of aircraft, or turbulence that exceeds the limitations of the airframe. The severely impacted threshold represents the upper threshold of being able to complete a mission and is therefore a valuable measurement to the end users of metrics data. Additionally, a pilot is likely to be able to accurately report severe impacts to a mission debriefer. Thus, we chose to categorize the forecasts and the observations according to whether they indicated *severe impacts* or *no severe impacts*. For our study, the condition we analyzed is the condition of severe impact and the corresponding contingency table for analyzing this condition is shown in Fig. 11. This contingency table was applied to all nine METOC parameters (e.g., applied to analyze whether severe impacts from icing were or were not forecasted, and whether severe impacts from icing were or were not observed by the pilot).

3. Data Collection and Analysis IT Issues

One of the primary findings of the Butler (2005) study was that a real time metrics system should be hosted at a remote site. It would be easier to write a self contained program to run at the location of an operational unit, but the ability to review and compare data for various locations by anyone with authorization is far too valuable to sacrifice. Centralized control also allows for more efficient storage and updating, and a standardized reporting and calculation system. In addition, centralization means that local units do not have to become proficient in the analysis and computing issues associated with maintaining a metrics program. For these reasons we decided to develop a remote system similar to that of Butler (2005) that would collect data through a webpage, and then conduct all the data analyses (e.g., calculations of metrics) and analysis reporting via an NPS server hosting the system. The system architecture is such that the database program stores the data and computes the metrics, while separate software components run the web based online collection form and the interface between this user interface and the database. Three technical issues arose in remotely hosting our metrics calculation system. The first problem was accessibility of an internet-based system by an afloat operational user that would have limited internet access while at sea. The second problem was the protection of potentially sensitive mission data that was to be transmitted to the processing system. The third problem was ensuring that the remote site development was in compliance with the Navy’s information technology (IT) guidelines so that it would dovetail with other Navy IT development efforts and be transferable from the research and development oriented NPS to an operational METOC center.

The first and second problems were solved by putting the system on the SIPRNet. This allowed us to safely transmit sensitive mission data without the fear of compromise, and, since the SIPRNet is also used for quite a bit of other operational traffic by a CV, the SIPRNet is rarely inaccessible to OA divisions.

For the third problem we had to modify the software used in the Butler system. Most of the work done by Butler (2005) and Darnell (2006) used open source software, such as PHP and MySQL, to develop a database and interface for data collection, analysis, and reporting. The open source software they used is not cleared for use on operational computer systems in the military; but is allowed on developmental servers such as those at NPS. But hosting the metrics system at NPS indefinitely would pose a problem because the METOC community would not have direct control over the system and unlimited access could not be guaranteed. For this reason, we explored other options. One option was to host the system at the Naval Atlantic Meteorology and Oceanography Command (NLMOC) on their SIPRNet servers. This would provide the easiest access for members of SGOT Norfolk because it would be hosted in the same building as the SGOT command. Hosting at NLMOC was determined to be non-feasible because of the IT requirements of the Navy precluded the software we were using. NLMOC was bound by the Navy's requirements and we were unable to develop our system on approved software due to the relatively high cost of licensing requirements. Our second option was to target the system to be hosted at the Fleet Numerical Meteorology and Oceanography Command (FNMOC). FNMOC hosts most of the Navy's numerical weather models and computer-based research in the field of METOC modeling. Because of this, they are able to run developmental servers that can host open source software. They had already received approval for use of PHP in their development environment but had not received approval for MySQL, which had been used for the database in previous NPS METOC metrics projects. The approved open source software for databases was PostgreSQL. So we adapted our database to run with the approved software.

Our plan is to move our system, once fully developed, over to FNMOC. Hosting at FNMOC will allow us the METOC community to directly maintain the metrics system. To facilitate the development of our system, we conducted much of the development on a publicly hosted Internet site. We then migrated the system to an NPS SIPRNet site where small amounts of additional development were conducted. This NPS SIPRNet system is now being used to collect and analyze data from several OA divisions, and to report their metrics results to them. We expect the system to be moved to an FNMOC site sometime in 2007.

4. Data Collection Process

The actual process of collecting data would ideally be a fully automated system that pulled observations from existing pilot reports and debriefing databases, and pulled forecast data from databases of mission forecasts. The data collection would seamlessly integrate into the existing operational tempo of CV air operations. The results would be instantly available for review and no extra man hours would be spent extracting data, entering data, or doing quality control of the data. In practice, such databases are not widely used or integrated with each other. Additionally, making entries into such databases without an expert on meteorological conditions present might frequently result in the severity of the weather events being misinterpreted. The solution of how to get the data from the pilots to a database was to develop a paper form (see Figure 12) and set of instructions for use by OA division personnel to collect data before, during, and after pilot debriefings, and companion web based form into which METOC personnel would enter data from the paper form and make additional entries as needed. The paper form and instructions are designed to be used by a METOC specialist in obtaining the mission data from the pilots. The form allows for quick note taking and impact classification, and was much faster, more portable, and more flexible than data collection based only on a web based form would be. The paper form was taken to the intelligence debriefings of the pilots, and the debriefer, usually a forecaster or competent weather technician, would observe the debriefing conducted by the intelligence office. Most of the information the weather debriefer needed was obtained from simply listening to the intelligence debriefing. Any important parts of the weather picture that were not covered in the intelligence debriefing were quickly covered by a few questions from the weather debriefer to the pilot at the end of the intelligence debriefing. This method of collecting data from the pilot provided a relatively complete set of data on negative weather impacts during the mission while minimizing the time required of the pilot to provide that information. After the debriefing, the weather debriefer took the paper form to the OA division office where it was entered into the electronic form on the SIPRNet at the convenience of the division.

B. METHODOLOGY

Once we had started collecting data, we were ready to calculate metrics based on that data. We calculated metrics that allowed us to assess the quality of the forecasts and their contribution to the situational awareness of the pilots. Three main categories of metrics were used. The first type was forecast performance metrics that were designed to describe the quality of the forecast. The second type of metric was operational impacts metrics, used to describe how information in the forecast and the actual weather phenomena experienced during the mission affected or could have affected operations. The third type of metric used in this study was phenomena metrics; these relate individual weather phenomenon to forecast performance and impacted operations. For all of the metrics, our focus was on situations that were forecasted to, or actually did, have severe negative impacts on the missions (e.g., the performance of the forecasts in predicting phenomena with severe negative impacts).

1. Forecast Performance Metrics

The forecast performance metrics are based on the contingency table in Figure 11 and are used to describe the performance of the forecasts in predicting events with the potential to cause significant negative impacts to operations. Although these metrics are flexible and can be used for any chosen threshold, we have used the threshold between *moderate impact* and *severe impact* as defined by the operational personnel. We will refer to a forecast of conditions above the threshold as a forecast for “severe impact” and a forecast for conditions below the threshold as a forecast for “no severe impact.” Note that the forecasts with which we were dealing were not predictions of impacts but rather predictions of conditions that, based on established thresholds, would likely lead pilots to infer severe negative impacts from the forecasts. Also, sometimes we have shortened the term “no severe negative impact” to “no impact,”, but both terms mean the same thing: the impact of the forecasted weather was expected to be less than severe, i.e., *no impact* or *moderate impact.* We calculated four basic performance metrics that are outlined in the following sections.

a. Forecast Accuracy Metric

The forecast accuracy (FAC) metric compares the number of correct forecasts in a given category to the total number of forecasts in a given category (Hinz 2004) and gives the result as a percentage of correct forecasts. In relation to the standard contingency table, the calculation is given as:

$$\mathrm{FAC} = 100 \times \frac{A + D}{A + B + C + D}$$

A perfect FAC score would then be 100% if every forecast for severe impacts and for no severe impacts was correct. A serious drawback to this metric is evident when it is used to measure a forecaster’s accuracy on predicting rare events. If, as illustrated in Table 1, a particular event occurs in 3% of forecast periods, a forecaster could maintain a 97% FAC simply by never forecasting the event.

Forecasted **Observed**	**Severe Impact**	**No Severe Impact**	**Total**
Severe Impact	Hit **0**	Miss **3**	3
No Severe Impact	False Alarm **0**	Correct Rejection **97**	97
Total	0	100	100

Table 1. Contingency table completed to demonstrate how a forecaster could maintain a 97% FAC by never forecasting for a rare event.

b. Probability of Detection Metric

The probability of detection (POD) metric serves as a useful complement to FAC, because it accounts for a failure to forecast rare events. POD compares the number of times a type of event was correctly predicted with the number of times that type of event was observed (Hinz 2004) using the following equation:

$$\mathrm{POD} = 100 \times \frac{A}{A+B}$$

POD shows us how good forecasters are at predicting a particular type of event. A score of 100% would mean that the forecaster correctly predicted the event every time it was observed. Note that a high POD score tends to be hard to get for rare impacts. By including the POD for rare events as a forecast performance metric, forecasters are discouraged from ignoring rare events in an attempt to get a higher FAC score. However, a POD score can be a misleading skill indicator if the forecaster has a tendency to over-forecast a certain type of event. Take the example in Table 2 of 100 forecasts where 40 of them are a correct forecast for a severe impact (A in the contingency table), 20 are incorrect severe impact forecasts (C), zero are for incorrect no severe impact forecasts (B), and 40 are for correct no severe impact forecasts (D).

Forecasted **Observed**	**Severe Impact**	**No Severe Impact**	**Total**
Severe Impact	Hit **40**	Miss **0**	40
No Severe Impact	False Alarm **20**	Correct Rejection **40**	60
Total	60	40	100

Table 2. Contingency table completed to demonstrate how a forecaster could maintain a high POD by over forecasting for an event.

In this example the FAC would be 80%, and the POD would be an impressive 100%. The problem is that even though all the observed severe impacts were correctly forecasted, 33% of the severe impact forecasts were incorrect. Even though the POD and FAC numbers were rather good, a reasonable perception by the customer who received the many extra severe impacts forecasts that never came to fruition might be that the forecasters hedge their bets to stay on the safe side.

c. False Alarm Rate Metric

The false alarm rate (FAR) metric indicates the rate at which a forecaster predicts severe impacts that are not observed and serves as a final partner to FAC and POD. FAR compares the number of incorrect forecasts of an event with the total number of times the event was forecasted in the following calculation:

$$\mathrm{FAR} = 100 \times \frac{C}{A+C}$$

The FAR calculation gives the percentage of severe impact forecasts that are incorrect. A FAR score of 0% would indicate that no severe impact forecasts were incorrect, and a high FAR score could indicate that the forecaster was over forecasting a severe impact. When combined with POD, a more complete picture can be formed about the forecaster's skill (Hinz 2004). A high POD and low FAR would indicate that a forecaster was skilled at accurately forecasting a particular type of severe impact. A low POD and high FAR would indicate that a forecaster was having trouble forecasting for a specific impact because they were missing impacts that were observed and predicting impacts that were not observed. Finally, a low POD and low FAR would mean the forecaster was under forecasting the impact, and a high POD and high FAR would mean the forecaster was over forecasting the impact.

d. Bias Metric

A more concise way to determine under and over forecasting is through the use of a bias calculation. Bias compares all severe impact forecasts to all severe observations in the following calculation:

$$\text{Bias} = 100 \times \frac{A+C}{A+B}$$

The bias calculation gives the percentage of observed severe impacts that were forecasted. A score of 100% would be a neutral bias and would indicate that there were an equal number of forecasted impacts and observed impacts. It does not necessarily mean that the forecasted and observed impacts correlated with each other. A bias score of 125% would indicate that there were 25% more forecasted impacts than observed impacts, and that there is a tendency to over forecast the particular impact. One problem with this definition of bias is that a given difference between the number of forecasts (A+C) and the number of observed events (A+B) can lead to very large differences in the magnitude of the bias, depending on which is larger. Tables 3 and 4 provide an example of this by showing the bias results when the difference between the number of forecasts and the number of observations is 80. In the first case (Table 3), A+C = 10 and A+B = 90, leading to a bias of 11%. In the second case, A+C=90 and A+B = 10, leading to a bias of 900%. So although the difference between A+C and A+B is the same in both cases, the biases are very different.

Forecasted **Observed**	**Severe Impact**	**No Severe Impact**	**Total**
Severe Impact	Hit **0**	Miss **90**	90
No Severe Impact	False Alarm **10**	Correct Rejection **0**	10
Total	10	90	100

Table 3. Contingency table completed to demonstrate how an under-bias of 11% can be calculated.

Forecasted / **Observed**	**Severe Impact**	**No Severe Impact**	**Total**
Severe Impact	Hit **0**	Miss **10**	10
No Severe Impact	False Alarm **90**	Correct Rejection **0**	90
Total	90	10	100

Table 4. Contingency table completed to demonstrate how an over-bias of 900% can be calculated.

To address this shortcoming, we have developed a modified bias calculation that centers the bias score on zero and uses two different calculations depending on whether there is an over or under forecasting bias. For an under forecast bias, where (A+C) < (A+B) the following calculation would be used:

$$\text{Bias} = (-1) \times 100 - 100 \times \frac{A+C}{A+B}$$

For an over forecast bias, where (A+C) > (A+B) the following calculation would be used:

$$\text{Bias} = 100 - 100 \times \frac{A+B}{A+C}$$

This modified version of the bias metric allows an easier comparison of over and under forecast tendencies because values with an equal difference between the number of forecasts and the number of observations are represented by an equal, but opposite, tendency to over or under forecast an event. In this version, a 0% bias is neutral, positive numbers indicate an over forecasting bias and negative numbers indicate an under forecasting bias.

2. Operational Impacts Metrics

The second category of metrics calculations are operational impacts metrics that represent how the mission was affected by the forecast or by the weather events encountered during the mission. We used metrics similar to those used by Jarry (2005) and Butler (2005), in particular, metrics similar to the missions saved metric. These metrics describe how a forecast for adverse conditions that led to a mitigation recommendation affected the mission planning. Because the OA division does not typically make recommendations that affect mission planning we did not calculate the missions saved or weapons saved metrics. However, we were able to create some metrics based on similar concepts, as outlined in the rest of this section.

a. Received Negative Metric

The missions that received a negative impact forecast (received negative) metric is the number of missions forecasted to have a severe impact divided by the total number of missions. This metric shows the percentage of all missions for which pilots were warned about a severe impact.

b. Mitigated Received Negative Metric

Based on the missions potentially saved metric in Butler (2005), we developed a similar metric, the missions that received a negative impact forecast and took mitigating action in response to encountering negative phenomena metric, which we simplified to: the mitigated received negative metric. This metric is based on the same calculation as the missions potentially saved metric, with the exception that there was no mitigating action recommended by the forecaster. The need for mitigating action is implied by the forecast of negative impacts. The calculation is done by totaling the number of missions with a correctly forecasted severe impact and a mitigating action by the pilot due to that impact divided by total number of missions. This metric provides a measure of the accuracy of the SA given to the pilot by the forecasters.

The negative received metric indicates the percentage of times that forecasters warned pilots of severe negative impacts. The mitigated received negative metric indicates the percentage of times that the forecaster warnings were correct. Thus, the closer the two metrics are to each other, the better the forecasters are doing in providing accurate SA to the pilots about severe negative impacts. These forecasts used in the calculation of the mitigated received negative metric can be seen as those that would tend to increase forecaster credibility. The forecasts of severe negative impacts that do not go into this calculation (i.e., those that are incorrect) would tend to decrease forecaster credibility. The other type of forecast error that erodes credibility, that of forecasting no impact, or moderate impact, when severe impacts ultimately occur, was not calculated or considered in this study, but is obtainable from the data we collected.

c. Missions Placed at Risk Metric

The missions placed at risk metric is the percentage of total missions flown for which pilots encountered severe negative weather impacts during mission execution. The metric is independent of the forecasted impacts and gives the user of the metric an idea how weather impacted missions flown during the data collection period.

d. Mitigation Rate Metric

The mitigation rate metric is the percentage of the total number of missions in which pilots took mitigating action in response to severe negative weather impacts. This metric gives an overall picture of how weather led to mitigating actions.

e. Missions Requiring Mitigation Metric

The missions requiring mitigation metric is the percentage of missions that encountered severe negative weather impacts for which mitigating actions were taken due those impacts. This metric is similar to the mitigation rather but is based on a subset of all missions, rather than all missions. Pilots are not always able to take mitigating actions in response to the severe negative weather impacts they encounter. For example, air space constraints above the planned flight level, and enemy threats below that level, may prevent a pilot from avoiding negative impacts from turbulence at the planned flight level. Or, as another example, the requirements of combat may make a mission a must fly mission despite severe negative impacts (cf. Hinz 2004, Jarry 2005). Thus, encountering a negative impact does not necessarily mean that mitigating action was taken. So the missions requiring mitigation metric is useful in helping to distinguish the situations in which severe impacts were encountered and mitigating action was taken from the situations in which severe impacts were encountered but no mitigating action was taken.

f. Missions Canceled Metric

The missions canceled metric represents the percentage of total missions that were cancelled due to severe impacts. This metric is a subset of missions requiring mitigation metric that focuses on missions that were cancelled due to severe negative weather impacts.

g. Targets Changed Metric

The targets changed metric represents the percentage of total missions for which the target was changed due to severe impacts. This metric is a subset of the missions requiring mitigation metric that focuses on missions that required a target change due to severe negative weather impacts.

3. Impacting Phenomena Metrics

The final category of metrics in this study is negatively impacting phenomena metrics. These metrics are used to describe how specific weather phenomena impacted missions, independently of how well the impacts were forecasted. These metrics are useful in determining the phenomena on which forecasters and pilots may need to focus extra effort; for example, extra effort in improving forecasts, mission planning, and pilot education and training (cf. Jarry 2005, Butler 2005)

a. Mission Impacting Phenomena Metrics

The mission impacting phenomena metrics show the percentage of mission impacts due to individual weather phenomena.

b. Mission Canceling Phenomena Metrics

The mission canceling phenomena metrics show the percentage of mission cancellations due to individual weather phenomena.

c. Target Changing Phenomena Metrics

The target changing phenomena metrics show the percentage of target changes due to individual weather phenomena.

III. RESULTS

A. OVERVIEW

In this chapter, we review the results of the implementation into the fleet of our metrics data collection, analysis, and reporting system. We begin by evaluating the system we implemented, including the development of the data collection model, the data collection by the user, the value of the metrics calculations to the user, and the lessons learned in the development and implementation of our metrics system. We then review the results of our data analyses, especially the forecast performance and operational impacts metrics. We also provide an assessment of what the metrics reveal about how the METOC community contributes to the Navy, what data still needs to be collected to have a more complete picture of METOC community contributions, and additional steps needed in the on-going metrics program of which this study is a part.

B. SYSTEM RESULTS

One of our goals with this project was to create and implement a system that collected forecasted and observational data and provided near real time metrics reports in text and graphical formats to the user. To achieve this goal, we conducted a series of experiments to determine the best data to collect, the best way to conduct the data collection, and the best analyses and metrics to conduct, and the best reports to provide.

The experiments included a series of trials and redesigns of our system as the system evolved into its final form for this study. The first series of trials and changes occurred in the development of the data collection questions and forms. The second series of trials and changes occurred during the course of determining the best process for using the form to collect data on an operational CV. The third series of trials and changes occurred in the development of the analysis tools and metrics calculations, and the reporting components. Our final step was to assess our full development process to identify the major lessons we learned that should be considered in planning future METOC metrics work.

1. Implementation

Implementing the metrics program took three steps. The first step was to design a set of questions that could fit on a 1-2 page paper form and have the form assessed by an OA division for content, wording, and ease of use. The second step was to develop an electronic form that closely resembled the paper form to allow for computer computations of collected data. The final step was to run the electronic form through a test run with real operational data.

Our first step in implementing a metrics system on an operation CV was to do a series of trials with paper forms. Through a weekly conference call with CDR Woll and LCDR Schmeiser from SGOT Norfolk, Bruce Ford (a retired Navy METOC officer and contracted computer programmer), Dr. Tom Murphree (the head of the NPS METOC metrics program), and the author, we developed a series of paper forms for discussion and testing (Woll 2006, Schmeiser 2006, Ford 2006, Murphree 2006). The original paper forms were used to test ideas on what information needed to be collected, what information could be collected, and how much information should be collected. We applied many of the lessons learned by Butler (2005) on designing a data collection form and process, by Darnell (2006) on the length of the data collection form and the aircrew time required during data collection, and by Hinz (2006) on using the intelligence debriefings of pilots as important opportunities to collect data for METOC metrics analyses. Early versions of the collection forms were distributed amongst the members of the conference call party for review. The first version of the form came from the consensus of the conference call members, and the form reflected our original desire to collect data on mission planning forecasts along with mission execution forecasts. The original data collection form can be seen in Figure 13.

We did a trial of the first iteration of the paper data collection form with the OA division on the USS Enterprise for three weeks during a pre-deployment exercise in early 2006. From this first trial, we learned that our form did not align well with how operations were conducted onboard the CV. Many fields of the form were left blank due to the lack of a planning element on the CV. A large section of the form concentrated on collecting data from two forecasts, a planning forecast issued more than 12 hours prior to mission takeoff, and a mission execution forecast issued within 12 hours of takeoff. However, since there was no planning element on the CV that received an OA division planning forecast, there was no data to collect regarding a planning forecast or mitigations based on a planning forecast. Additionally, we included many questions on the initial form regarding air wing initiated mitigations during mission planning, but learned that the air wing planners had very limited mitigation capability. Most mitigating actions were made by the pilot during mission execution, and we asked very few questions regarding these pilot initiated mitigating actions.

We took the lessons learned from the trial of our first paper form and developed a second paper form that can be seen in Figure 12. The second iteration of the paper form, as described in Chapter II, focused only on the mission execution forecasts and impacts experienced by the pilot. This second iteration of the form collected data on the forecasted severe impacts, severe impacts observed by the pilots, and mitigating actions taken by the pilots. This second form went through several stages before it was tested. Originally it was set up to track in what phase of the mission each severe impact and mitigating action was experienced. This proved to be very cumbersome and led us to believe it would be a daunting form to complete by the mission debriefer and a code intensive form to put onto a SIPRNet site. A version of the second form was presented to the OA division of the USS Enterprise. Feedback from the second iteration was much more encouraging (Everett 2006). A few changes were suggested by the OA division and the participants in the weekly conference calls. This paper form ended up being the form we then used to develop the electronic data collection form.

The electronic form was developed to parallel the final paper form. The electronic collection form consisted of the same questions as the paper form in the same basic layout to allow for easy transfer of the paper form data onto the SIPRNet. The electronic data collection form can be seen in Figures 14 and 15 and on the SIPRNet at http://web.ntsstl.nps.navy.smil.mil/Metrics/

shipboard_metrics/strike_metrics/index.php as of 22 September 2006. We found it beneficial to develop the electronic form on the internet before placing it on the SIPRNet. Developing it on the internet allowed for easier interaction with civilian contractors and convenient access for all parties that were reviewing the system. Having the electronic form developed on the internet also allowed us to populate the database with a large amount of test data inputted by civilian staff from NPS. The test data was used to test the operational capabilities of the system, to develop our calculations and text and graphical reports using a dataset with prescribed properties. Developing the calculations in this way allowed us to hand check our calculation against the computer calculations without having to contend with a real-time, changing database. One of the problems we encountered during our initial test with the test data was that the PostgreSQL database software was very slow at completing the many calculations involving nested loops. We felt this would become a liability once large amounts of operational data were collected. During the development of the metrics calculations, we decided it would be more efficient to have the PostgreSQL database software only collect and store data, and to have the PHP interface software do the analysis calculations.

Once the data collection form and database were functioning on our development internet site, the system was replicated at the NPS METOC Metrics program SIPRNet site. While developing the analysis and reporting functions of our system, we started collecting data from two operational CVs. The USS Enterprise and the USS Eisenhower both participated in small scale collections in August of 2006 that helped us assess access to the system from operating CVs and the usability of the system to new OA divisions with little or no prior knowledge of our program or online system. Overall feedback indicated that the system was accessible and functional. Data could be entered in via the SIPRNet form and archived on the hosting server at NPS with no complications.

2. Interaction

The OA divisions that participated in the August 2006 trial of the electronic form reported favorable opinions of the online system. The divisions reported that the questions asked of the pilots and the information collected provided an adequate representation of what the pilots experienced (Everett, 2006). The divisions also reported that they had adequate manpower to do the collections without having to overly strain personnel resources. Additionally, the method of collecting data during the intelligence debriefing of the pilot proved successful and was generally considered the best practice for collecting mission data (Hinz,2006, Everett 2006). Some of the mission debriefers also reported success debriefing individual pilots at times outside of the intelligence debriefing per the pilots' individual preferences (Woll, 2006). The OA divisions also noted that during August 2006, the weather in the ships' operating areas was generally benign and that their feedback to us about using our system might change when weather conditions become more adverse and lead to an increase in the OA divisions' workloads.

The negative feedback we received from the August 2006 data collection effort all involved SIPRNet connectivity issues. On several occasions, OA division members had trouble entering data into the online collection form because either they could not access the form hosted at NPS or because poor connectivity interfered with data submission. We were unable to identify an immediate solution for the connectivity issues but have some ideas for future work (see chapter IV). The best short term solution we could provided was for the OA division to archive their paper forms until SIPRNet connectivity improved and they were able to submit their data online. Once we determined that the electronic collection form was well matched with the needs and abilities of the OA divisions, we focused our efforts on the data analysis and reporting components of the system.

3. Data Analyses

Butler (2005) and Darnell (2006) both developed automated data collection, analysis, and reporting systems. Their systems allowed reports to be issued to users within seconds of data being entered. This was much faster than the many months that prior studies, such as those from Jarry (2005) and Hinz (2004), needed to provide the same type of calculations. Studies based on automated real time analyses and on delayed analyses have provided useful metrics. The delayed results studies have yielded historical analyses on missions that took place months to years in the past, whereas the near real time studies have provided current analyses of how the METOC community was supporting missions. For our study, we adapted the systems developed by Butler (2005) and Darnell (2006) to help provide real time information on the performance of forecasts, the value of forecasts to pilots, and how missions are being affected by specific weather phenomena.

We developed our set of metrics (Chapter II, Section B) from those in prior near real time studies. We only made slight, customer specific changes to most of the prior metrics. The metrics in this set were chosen to give an accurate real time picture of the skill of forecasters, the value forecasts give to the mission, and what types of weather are affecting missions.

Figures 16-32 show examples of the output from our automated real time system using test data analyzed at the internet development site. We used test data and an unclassified internet system development site to error check our data analyses, and our text and graphical outputs, prior to moving the system to the SIPRNet site. These output examples using test data are representative of the classified output available on the SIPRNet. The metrics output is displayed in tabular, graphical, and text formats, and shows the individual values based on type of weather impact, phase of the mission, and averages of impacts and/or phases. The output formats were selected based on recommendations from a range of METOC personnel. Classified output from the SIPRNet version of the system using real world operational data is available in the classified Appendix of this report and at the NPS METOC metrics SIPRNet site (http://web.ntsstl.nps.navy.smil.mil/Metrics/shipboard_metrics/strike_metrics/index.php).

Figures 16–25 show output examples from the forecast performance metrics: FAC, POD, FAR and bias. Examples of operational impacts metrics based on test data are shown in Figures 26 and 27. As discussed in Chapter II, Section B.2, the received negative metric and the mitigated received negative metric are based on the concept of missions saved developed by Jarry (2005) and missions potentially saved developed by Butler (2005). Examples of output for the impacting phenomena metrics are shown in Figures 28-32. The primary challenges in developing the process for generating real time displays of the three different types of metrics (forecast performance, operational impact, and impacting phenomena) was refining just what the metrics should represent, and how they should be calculated from the data collected using the forms shown in Figures 12, 14-15. This was especially true for the operational impacts metrics. Our secondary challenge was to realistically create the graphical outputs in an automated manner.

4. System Lessons Learned

Several lessons were learned from the development of the metrics system. These lessons mainly became evident during the trials of the paper and electronic data collection processes. We feel these lessons are not unique to working with OA divisions on CVs, but have wide application in developing many types of METOC metrics systems. Specifically, we learned the following.

- Prior to designing a system, it is critical to do a thorough assessment of the processes used, and not used, by the METOC unit's customers, and by the METOC unit itself. In our case, we spent much more time on this task then we anticipated needing. We did not initially know that CV air wings have essentially no mission planning capability and very little pre-mission mitigation capability. Nor did we initially know that OA divisions do not usually provide planning weather forecasts. Instead, the CAOCs tend to provide almost all of the mission planning and the planning forecasts. We learned that the main benefit the OA division provides is weather situational awareness for pilots. This required us to make some significant changes to our plans for collecting and analyzing data, in particular because SA is difficult to quantify and assess.

- Getting the data we needed, especially getting data from pilots, was not as hard as we initially believed it would be. We found that we were able to glean most of the data we needed from the intelligence debriefing, and that the demand on the pilots' time was not as large as we had expected. We attribute our success in collecting data to a combination of applying lessons learned from Darnell (2006) regarding asking for too much information, from Hinz (2006) regarding how to use intelligence debriefings of pilots to collect data, and the fact that CV operations required fewer questions than the missions on which Butler (2005) and Darnell (2006) based their data collection.

- Development of the collection system on the Internet made the development process much more efficient. Internet use allowed for civilian contractors to be involved in the programming of the system and allowed all parties to access the development system from non-secure computers, thus facilitating feedback.
- The PostgreSQL database software struggled to compute the many calculations involving nested loops within a reasonable time, and was probably not the best software choice for this purpose. During the development of the metrics calculations, we decided it would be more efficient to have the PHP interface software complete the calculations and have the PostgreSQL database software only collect and store data.
- SIPRNet connectivity problems affected data transfer from the CVs to NPS. Users reported that on several occasions they had trouble submitting mission information due to an intermittent SIPRNet connection. This occurred even though we designed a system that required relatively little data to be transferred per mission.

Based on these lessons, we have developed suggestions for future METOC metrics efforts (see Chapter IV).

C. METRICS RESULTS

1. Results

We were able to conduct an initial assessment of the METOC support given by one of the aircraft carriers in this study. This CV provided a sufficient amount of data for doing a very preliminary assessment of the performance and operational impacts of its forecasts. Our initial assessment revealed gave results that were similar to those from Hinz (2004) and Jarry (2005). These results include: a FAC for all forecasts of 99.4%, a POD for all forecasts of severe negative weather impacts of 40.0%; and a 13.6% mitigation rate. The similarity between these metrics and those from prior studies suggest that our data collection, analysis, and reporting system is functioning as intended. A more complete assessment of the performance and operational impacts of the OA division forecast will be possible once a larger quantity of data has been collected. These assessments will be available at the NPS METOC metrics SIPRNet site in real time as the data is collected.

2. Limited Data

The metrics results we obtained in this study are based mainly on data from two CVs and from brief periods of time (August-September 2006 for most of the data). Thus, our metrics results are not yet extensive enough to make well founded assessments of how well the OA divisions are doing or how CV missions are being affected by weather forecasts and phenomena. But, as the metrics system we created is introduced to more operational units and used over a longer period of time, a larger dataset will be produced from which robust results can be generated.

The time allowed for our collections ended up being mainly August - September 2006. The number of missions flown during this time was not high, and the weather conditions were very rarely adverse. Thus, there were few missions that experienced severe negative weather impacts.

The number of participating ships also limited our dataset. The primary ship we collected from was the USS Enterprise while she was on her bi-annual deployment in the Middle East. We collected data from the USS Eisenhower as its air wing participated in pre-deployment exercises and while the air-wing was training at NSAWC. Finally, a small amount of data was collected from the air-wing of the USS Stennis as it was training at NSAWC. We did not attempt to extend our data collection beyond the aforementioned ships. This is mainly because of time constraints imposed by the deployment schedule for the CVs, and because our project was aimed at developing a well tested prototype system, not at collecting large amounts of data from many ships,

3. Limited Metrics

A side effect of the limited amount of data we were able to collect was the limited number of metrics we were able to compute. The ability to accurately predict weather varies not only by forecaster and type of unit supported. Forecasting accuracy is also affected by regions, seasons, and climate variations. Metrics that measured quality of forecasts in different geographical regions, seasons, and climate variations could be used to determine how these variables affect, for example, METOC contributions to its customers, and how to improve forecaster training. In order to calculate metrics that describe spatial and temporal forecasting differences, more data must be collected. The data must also be linked to the region in which it was collected, something we did not attempt to do in this study. Once a large dataset is created, most likely over several years, a good understanding of the spatial and temporal variations in forecast performance, operational impacts, and impacting phenomena can be developed. The effects of slowly varying factors (e.g., El Nino, Pentagon budget variations, CONOPS variations, etc.) will require decades worth of data to be fully understood.

4. Value

One of our goals was to conduct an initial evaluation of the performance of METOC products provided in support of strike warfare and the operational impacts of those products on strike warfare. Our dataset allowed for only a limited assessment of the quality of the products that were being provided to pilots in support of their missions, and we were only able to look at the METOC products provided to pilots, not those provided to mission planners. Initially, we had hoped to be able to asses both types of products.

To analyze planning products for CV missions, we would need to collect and analyze data from CAOCs (as was done by Hinz (2004)). However, we determined that collecting and analyzing data on mission execution forecasts and outcomes would be challenging in several respects, and that working with CAOC data would be beyond the scope of this study. To complete the depiction of METOC support for Naval strike operations, future extensions of this study will need to be conducted using CAOC forecast and planning data (see Chapter IV).

THIS PAGE INTENTIONALLY LEFT BLANK

IV. SUMMARY AND RECOMMENDATIONS

A. OVERVIEW

We started this study with three goals (see Chapter I). The first goal was to adapt the metrics collection and analysis system developed by Butler (2005) to operational use for strike warfare support. We met this goal through the development of a SIPRNet based data collection, analysis, and report system. The second goal was the implementation of the metrics system on three operational aircraft carriers, which we achieved. Our third goal was to conduct an initial evaluation of the accuracy and impact of METOC products provided in support of strike warfare. This goal was also met, primarily in the form of our classified data analyses and reports.

B. SUMMARY OF RESULTS

While striving to achieve our overall goals for the project we proceeded through six major phases. Each phase was a necessary step toward the implementation of our intended metrics. The phases were:

1. Determination of METOC Support Focus Area

The first step towards supporting aircraft carrier strike missions was to determine what aspect of strike operations we wanted to focus out efforts on. Developing a system that touched every aspect of carrier based strike operations would not have been feasible within our study. Through initial trials and discussions, we focused our metrics efforts on mission execution forecasts given to carrier based pilots, the situational awareness those forecasts provided, and the mitigating actions pilots took due to the corresponding observed conditions.

2. Dataset Development

Specific data had to be identified for collection. The data we collected had to be complete enough to make basic metrics calculations, and succinct enough that it could be reasonably extracted from pilot debriefings without imposing a burden on the pilots. A short form was developed to collect data from pilots on severe negative weather impacts they encountered. The key issues on which we sought data from the pilots and/or the OA personnel were: (1) the severe impacts encountered during the mission; (2) the severe impacts that were forecasted; (3) how the severe impacts affected the mission; and (4) the mitigating action pilots took in response to the severe impacts. We focused on severe negative weather impacts, even though they tend to be rare, because they are: (a) impacts pilots are likely to remember; (b) impacts that are most difficult to forecast; and (c) impacts that are critical to accurately forecast (cf. Hinz (2004), Jarry (2005)). Severe impact forecasts and observations were also well suited to binary 4x4 contingency table analyses, one of the standard tools in forecast verification..

3. Development of Metrics Analyses

The studies outlined in Chapter I, Section B extended considerable effort in developing metrics to use in measuring METOC support to combat operations. We used many of the concepts they studied, and made some alterations along with developing several new metrics of our own. Our performance metrics (FAC, POD, FAR, and bias) were all adapted for use in combat support in prior studies. We used the performance metrics unaltered, with the exception of bias which we retooled slightly for easier understanding by end users. We also adapted the previously developed concepts of missions saved and missions potentially saved (Jarry 2005; Butler2005) and adapted them for use in an environment where no mitigating actions were recommended. Finally, we used variations of many existing impacting phenomenon metrics and adapted them for tracking missions in which pilots took mitigating action during mission execution.

4. Preliminary Analyses

With the structure of the dataset and metrics calculations in place, we developed the online data collection, analysis, and reporting component of our system. The system was initially developed on the internet for testing and revising. While hosted on the internet we populated the developmental system with test data to check our metrics analysis and reporting processes, we solicited comments and ideas from parties involved in the development of the overall metrics project, and we operationally tested the functionality of the code and software.

5. Implementation of a Highly Adaptable SIPRNet System

Once the developmental system on the internet was approved, we placed the system on the SIPRNet for use by operational units. The first operational run of the system consisted of three CVs collecting data from their missions and entering that data into the NPS SIPRNet database. The data we collected dealt with four main issues: (1) the severe impacts encountered during the mission; (2) the severe impacts that were forecasted; (3) how the severe impacts affected the mission; and (4) the mitigating action pilots took in response to the severe impacts. These questions and answer options we developed for these four issues are easily changed to meet the needs of other warfare areas without having to redevelop the programming code or retraining system operators and users. Thus, the foundation component of our system, data collection, along with the supporting IT infrastructure, is highly adaptable.

6. Feedback

Once the operational collection system was in place, we were able to solicit feedback from METOC personnel that were using the system. Their feedback was useful in fine-tuning the process of collecting mission data from the pilots. The feedback from the users also showed us some of the shortcomings of the system, such as unreliable SIPRNet connectivity, that can be applied to future METOC metrics efforts.

C. LESSONS LEARNED

We learned lessons that should be applied to any future work in future METOC metrics efforts. Our lessons came from two distinct parts of our project: the conceptual development phase and the transitioning to operational use phase.

1. Conceptual Development

During conceptual development we originally wanted to collect METOC metrics data and perform metrics calculations on forecasts and mitigating actions during the planning portions of a strike missions. This would include weather support given to the ATO planners and changes made to the ATO up to forecasts given to pilots for mission execution and mitigating actions taken before and after aircraft launch. What we learned was that we did not have a good understanding of the process of moving a mission from the planning stages through execution. Our incomplete understanding cost us time and man-hours while we were trying to develop a system to collect data that did not exist within the framework of OA division METOC support. Our take-away from the process was that we should have invested more initial project time and effort into fully understanding the process we were trying to track, and a complete understanding could have led to a better end product.

The second lesson learned during the conceptual development was that the CAOC provides critical support to mission planners and pilots. The METOC support given to ATO planners ultimately supports air wing pilots flying missions on the ATO. To fully track the weather forecasts given to the pilot and all mitigating actions that affect a mission, a data collection, analysis, and reporting system must be created for and implemented in the CAOC.

2. System Implementation

The implementation of our data collection, analysis, and reporting system in CV operations also led to a series of lessons learned. These lessons were outlined in Chapter III, Section B, and are reviewed here:

- We found that we were able to glean most of the data we needed from the intelligence debriefing using a senior AG or a junior AG with a strong understanding of the strike mission process, and that the demand on pilots' time was acceptable to the pilots. Based on lessons learned from prior studies (Butler (2005), Darnell(2006), Hinz (2006)), we set out to minimize the impacts of our data collection on the pilots' routine by integrating our collection into an existing debriefing rather asking pilots to participate in a new METOC focused debriefing.
- Development of the collection system on the Internet made the process much more efficient. Internet use allowed civilian contractors to readily be involved in the programming of the system, and allowed all parties to access the system from non-secure computers, thus facilitating feedback.

- The PostgreSQL database software had difficulty quickly computing the many metrics, and may not have been the best choice of software for this purpose. We decided it would be more efficient to have the PHP interface software do the calculations and have the PostgreSQL database software only collect and store data.
- Despite our efforts to require the least amount of data to be transferred through a remote site, and our use of the most available connection on the CV, we found that poor SIPRNet connectivity still affected data transfer. Users reported that on several occasions they had trouble submitting mission information due to intermittent SIPRNet connectivity.

D. RECOMMENDATIONS

The first step to the continuation of the current metric program of collecting aircraft carrier strike mission data is the transfer of our system to an operational METOC center for continued hosting and maintenance. The server at NPS was always considered a developmental host and was not intended to be used indefinitely. Access to the server is limited and the manpower at NPS cannot support maintenance of the system. Personnel at an operational METOC center need to be trained on the operation of the system and how to maintain it for future use. Hosting it at such a center will ensure prolonged data collection and establishment of a long term dataset with which to represent the quality of support Navy METOC gives to its operational customers. One prime candidate center is FNMOC.

Our METOC metrics system, along with the prior systems that were developed by Butler (2005) and Darnell (2006), was developed with a bottom-up approach. The idea for it originated at an operational unit in the Navy and its design was dictated by those who would be using the system, especially those who would be analyzing the results. The concept of developing a system to collect METOC support data and provide metrics results is supported by CNMOC leadership, but is not yet directed by them. This study adds to an extensive set of studies on the reasons for, and means to, a comprehensive METOC metrics program. The net results of these studies clearly indicate the need for a governing body of the METOC community to develop a directed and standardized METOC metrics program.

Under direction from CNMOC, the results of our project should be adapted for use in other warfare areas. Most work in the field of METOC support metrics up to this point have been concentrated on fixed wing aviation missions. The concepts developed in this and past studies could be adapted to special warfare, amphibious assault, and even into the realm of oceanography support for mine warfare and antisubmarine warfare. The methods and infrastructure we have developed are highly adaptable to other warfare areas.

Beyond Navy METOC, other branches of the military can benefit from the results of our study. We have already seen some success with the closely Darnell (2006) system in the Air Force. The improvements we have made to the Darnell approach can readily yield additional benefits to AFW and the customers it supports. Beyond the Navy and the Air Force, the Army and the Marine Corps use weather support in the planning of operations. Larger scale operations might require a different approach than that used for individual aircraft missions. But the basic methods and benefits to the user are the same: (1) efficient collection and analysis of data on (a) forecasts and verifying observations and (b) customer plans and outcomes; and (2) an understanding of: (a) how weather and weather support products are used in mission planning and execution and (b) how weather impacts operations.

E. FUTURE WORK

One of the major projects we recommend for future research and development is the creation of a metrics system that incorporates both the planning and execution portions of the mission process based on data collected at the CAOC. We also recommend continuing the effort toward seamless integration of different METOC related data collection efforts. Pulling forecasted data from a briefing generator would reduce the work done by the OA division by making the forecast data collection invisible to the operator. Likewise, pulling observation data from existing intelligence databases would make the collection of observational data invisible to the pilot and the OA division, and make the collection of the data needed for a community wide METOC metrics program much more efficient, and would lead to much more extensive data sets than would otherwise be possible. Such an automated system would be of great benefit to forecasters, planners, and pilots alike.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Air Force Weather Strategic Plan and Vision (AFWSPV), cited 2004. [Available online at [http://www.e-publishing.af.mil/pubfiles/af/dd/afdd2-9.1/afdd2-9.1.pdf]. Accessed 25 July 2006.

The Assistant Secretary of the Navy (ASECNAV), 2002. *Incorporating Performance Metrics Into Fiscal Year 2004-2009 Programs and Budget Process.* (Memo to the Chief of Naval Operations), May 2002.

Butler, M., 2005. *Automated Metrics of METOC Forecast Performance and Operational Impacts.* M.S. Thesis, Department of Meteorology, Naval Postgraduate School.

Cantu, A., 2001. *The Role of Weather in Class A Naval Aviation Mishaps.* M.S. Thesis, Department of Meteorology, Naval Postgraduate School.

Darnell, K., 2006. *Analysis of Weather Forecasts Impacts on United States Air Force Combat Operations.* M.S. Thesis, Department of Meteorology, Naval Postgraduate School.

Department of the Navy, 2005. *Department of the Navy Fiscal Year (FY) 2006/FY 2007 Budget Estimates, Justification of Estimates.* February 2005.

Everett, K., 2006. Personal communication with LCDR Keith Everett, OA Division Officer, USS Enterprise, and the author, between May 2006 and August 2006.

Executive Office of the President of the United States (Office of Management and Budget), 2006, *Analytical Perspectives, Budget of the United States Government – Fiscal Year 2007.* Document available at http://www.whitehouse.gov/omb/budget/fy2007/pdf/spec.pdf as of July 21, 2006.

Ford, B., 2006. Personal communication with Bruce Ford, a retired METOC officer and computer programming contractor, and the author, between February 2006 and September 2006.

Hinz, J., 2004. *Developing and Applying METOC Metrics to Sea Strike: A Case Study of Operation Iraqi Freedom.* M.S. Thesis, Department of Meteorology, Naval Postgraduate School.

Hinz, J., 2006. Personal communication with LCDR Jake Hinz, OA Division Officer, USS Nimitz, and the author's thesis advisor, Dr. Tom Murphree, between May 2006 and September 2006.

Jarry, J., 2005. *Analysis of Air Mobility Command Weather Mission Execution Forecasts: Metrics of Forecast Performance and Impacts on War Fighting Operations.* M.S. Thesis, Department of Meteorology, Naval Postgraduate School.

Long, M., 2003. *Navy Tobacco Program Metrics – Summary Report.* (Summary of the MTF Tobacco Program data reported to the Navy Environmental Health Center), December 2003.

Martin, B., 2002. *METOC and Naval Afloat Operations: Risk Management, Safety, and Readiness.* M.S. Thesis, Department of Meteorology, Naval Postgraduate School.

Murphree, T., 2006. Personal communication with Tom Murphree, the NPS Metrics Project lead, and the author, between October 2005 and September 2006.

OPNAVNOTE 5450.0336, 2006. *Establishment of Strike Group Oceanography Team (STRKGRUOCEAN TEAM) Norfolk, VA and San Diego, CA.* Office of the Chief of Naval Operation serial number DNS-33/6U827251, May 2006.

Schmeiser, G., 2006. Personal communication with LCDR Gregory Schmeiser, Executive Officer, SGOT Norfolk, and the author, between May 2006 and September 2006.

U.S. Department of Transportation, Federal Aviation Administration (FAA), 2006. Order 7110.10S, Flight Services. February, 2006.

Woll, S., 2006. Personal communication with CDR Steve Woll, Commanding Officer, SGOT Norfolk, and the author, between January 2006 and September 2006.

APPENDIX — FIGURES

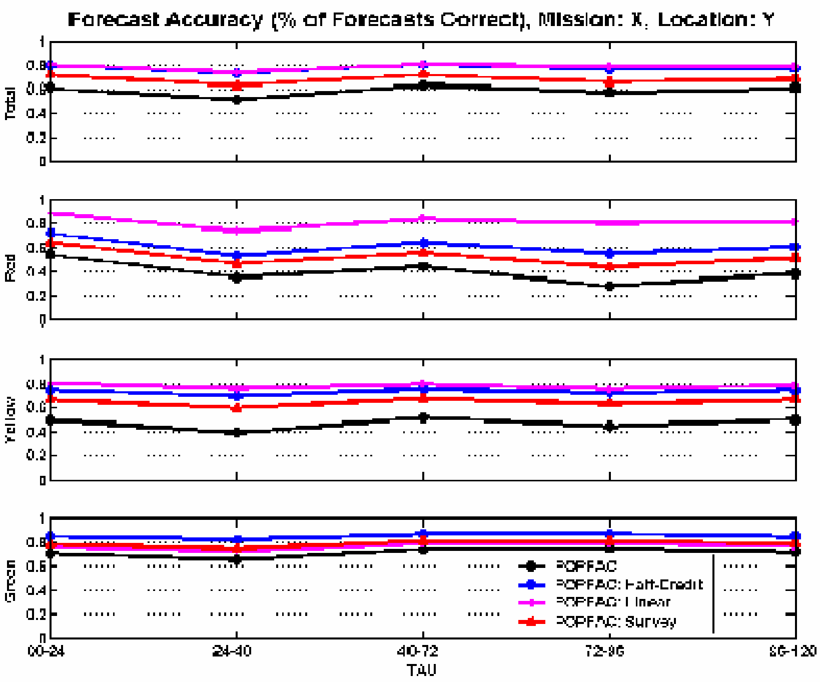


Figure 5. From Hinz (2004). FAC computed for a classified mission at a classified location. Vertical axis indicated percentage of perfect values divided by 100, and horizontal axis indicates forecast taus. Top panel is total accuracy, and the lower panels are for Red, Yellow, and Green forecasts accordingly. Hinz used a Red/Yellow/Green forecast structure instead of the Severe/Not Severe structure of this study.

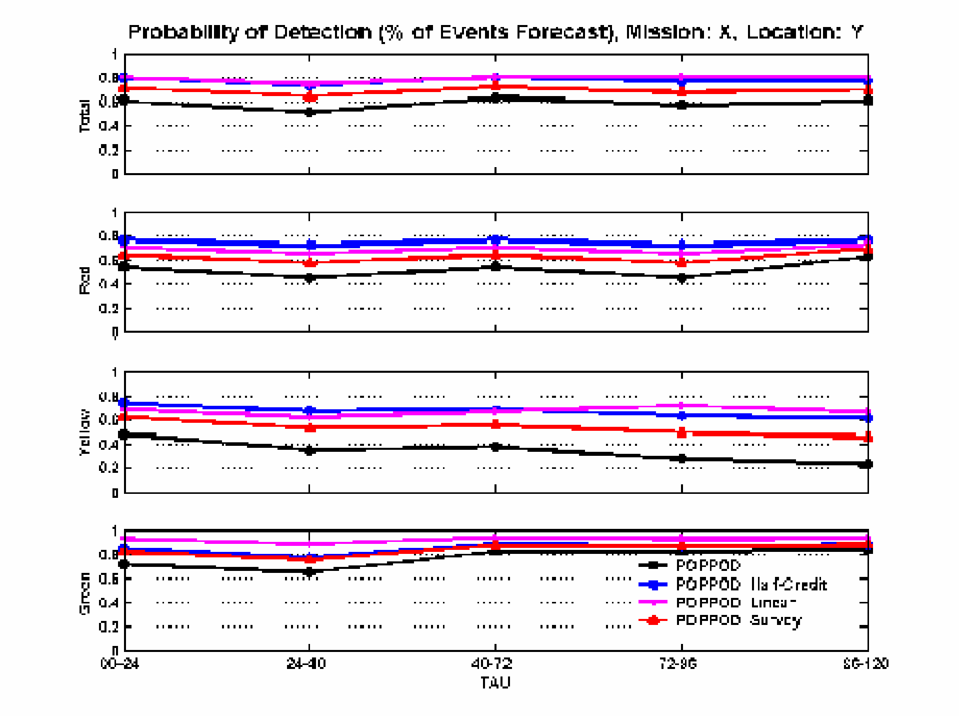


Figure 6. From Hinz (2004). Same as in Figure 1 but for POD.

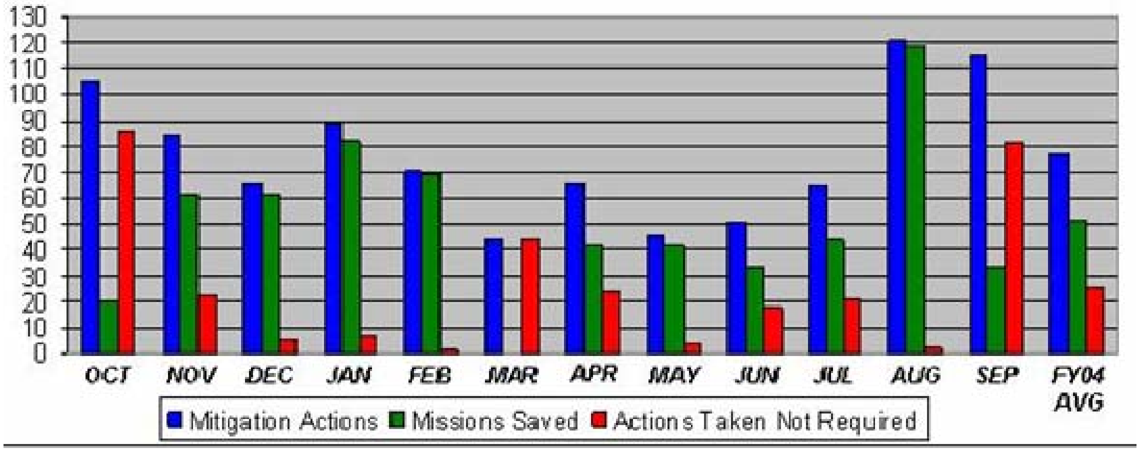
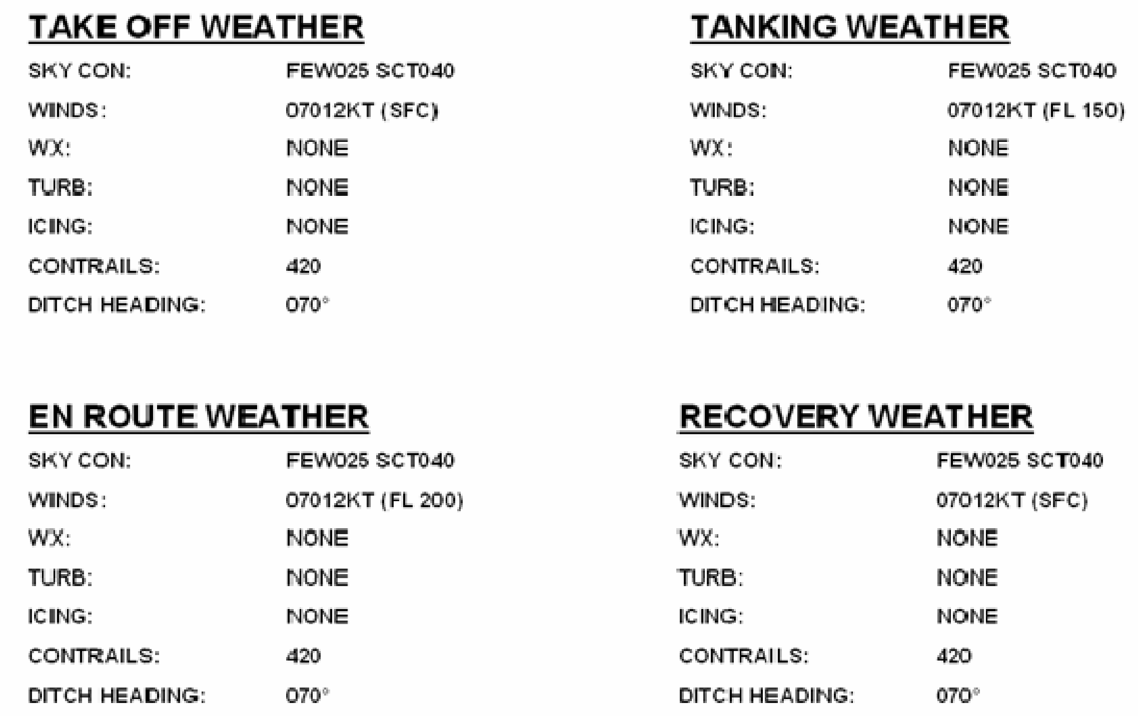


Figure 7. From Jarry (2005). 15th Operational Weather Squadron Global Mobility Weather Flight FY2004 monthly and annual mean number of missions in 2004 that took mitigating actions, estimated number of missions saved from weather delays, and estimated number of missions that took mitigating actions that were not required.



TAKE OFF WEATHER

SKY CON:	FEW025 SCT040
WINDS:	07012KT (SFC)
WX:	NONE
TURB:	NONE
ICING:	NONE
CONTRAILS:	420
DITCH HEADING:	070°

TANKING WEATHER

SKY CON:	FEW025 SCT040
WINDS:	07012KT (FL 150)
WX:	NONE
TURB:	NONE
ICING:	NONE
CONTRAILS:	420
DITCH HEADING:	070°

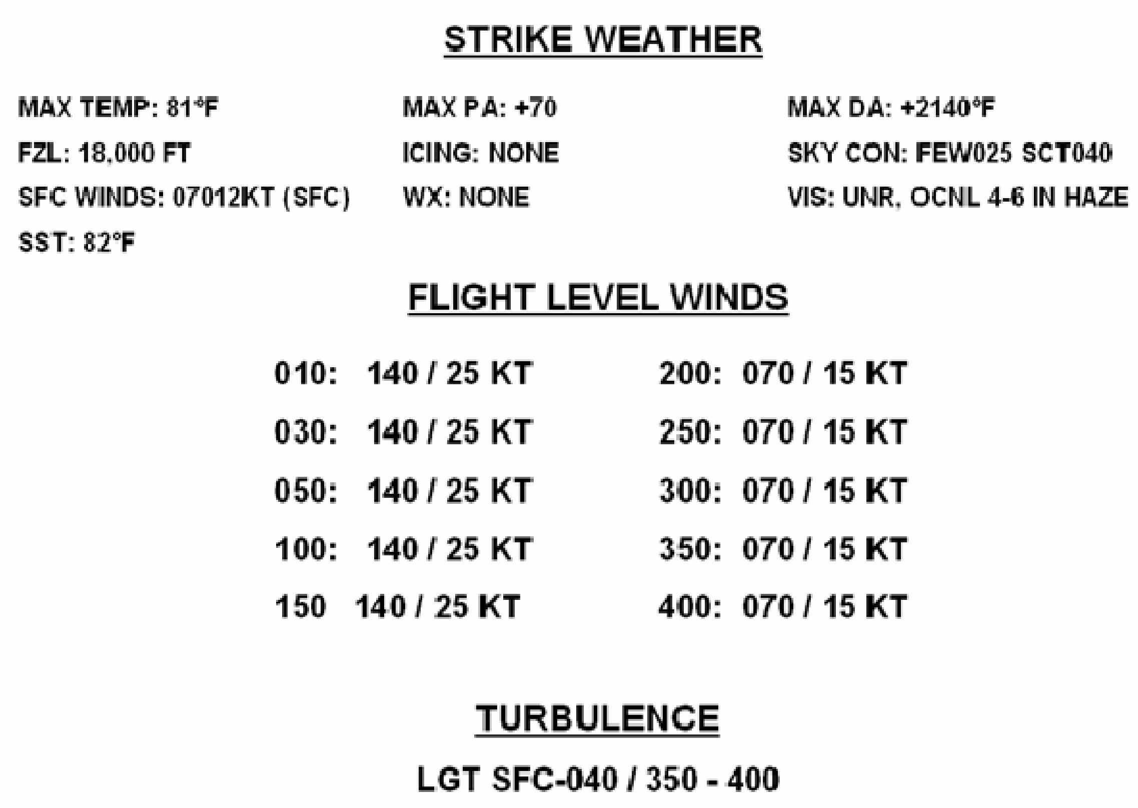
EN ROUTE WEATHER

SKY CON:	FEW025 SCT040
WINDS:	07012KT (FL 200)
WX:	NONE
TURB:	NONE
ICING:	NONE
CONTRAILS:	420
DITCH HEADING:	070°

RECOVERY WEATHER

SKY CON:	FEW025 SCT040
WINDS:	07012KT (SFC)
WX:	NONE
TURB:	NONE
ICING:	NONE
CONTRAILS:	420
DITCH HEADING:	070°

Figure 8. Example of the first page of a standard strike mission weather briefing given to pilots in support of individual missions flown from a CV showing forecasted weather conditions during four phases of the mission.



STRIKE WEATHER

MAX TEMP: 81°F
FZL: 18,000 FT
SFC WINDS: 07012KT (SFC)
SST: 82°F

MAX PA: +70
ICING: NONE
WX: NONE

MAX DA: +2140°F
SKY CON: FEW025 SCT040
VIS: UNR, OCNL 4-6 IN HAZE

FLIGHT LEVEL WINDS

010: 140 / 25 KT
030: 140 / 25 KT
050: 140 / 25 KT
100: 140 / 25 KT
150 140 / 25 KT

200: 070 / 15 KT
250: 070 / 15 KT
300: 070 / 15 KT
350: 070 / 15 KT
400: 070 / 15 KT

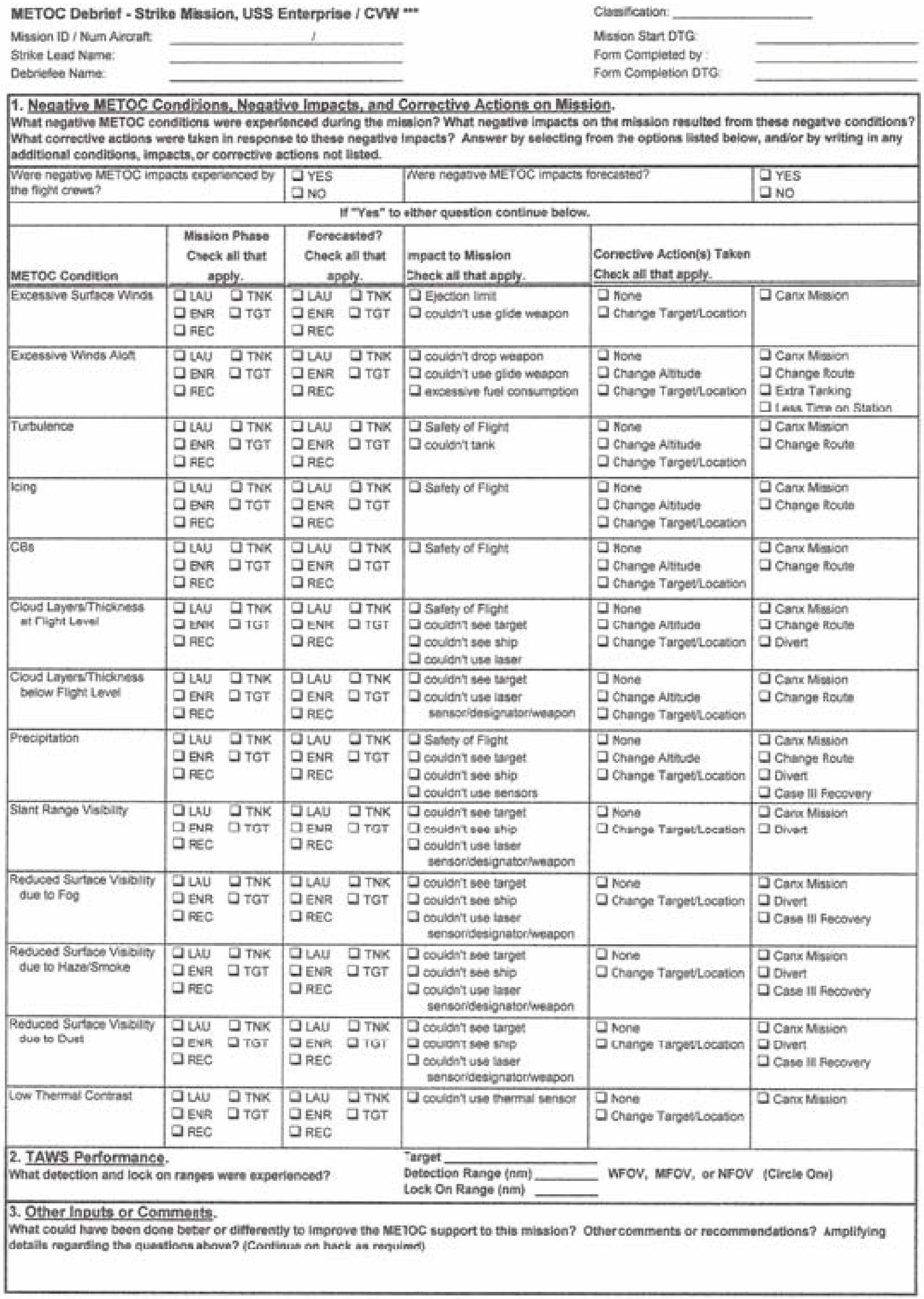
TURBULENCE

LGT SFC-040 / 350 - 400

Figure 9. Example of the second page of a standard strike weather briefing given to pilots in support of individual missions flown from a CV showing forecasted weather over the target area.

Forecasted **Observed**	**Severe Impact**	**No Severe Impact**	**Total**
Severe Impact	Hit **A**	Miss **B**	A+B
No Severe Impact	False Alarm **C**	Correct Rejection **D**	C+D
Total	A+C	B+D	A+B+C+D

Figure 10. Contingency table. Forecasts are assigned to one of four categories in this table (A-D) according to the operational impact indicated by the forecasted conditions and verifying observed conditions. The number of entries in each category, plus the sum of the entries in different groupings of the four categories, is used to calculate forecast performance metrics (FAC, POD, FAR, bias). For details, see Chapter II, Section B.1.



METOC Debrief - Strike Mission, USS Enterprise / CVW ***

Classification: ____________

Mission ID / Num Aircraft: ____________ / ____________

Strike Lead Name: ____________

Debriefee Name: ____________

Mission Start DTG: ____________

Form Completed by : ____________

Form Completion DTG: ____________

1. Negative METOC Conditions, Negative Impacts, and Corrective Actions on Mission.

What negative METOC conditions were experienced during the mission? What negative impacts on the mission resulted from these negatve conditions? What corrective actions were taken in response to these negative impacts? Answer by selecting from the options listed below, and/or by writing in any additional conditions, impacts, or corrective actions not listed.

Were negative METOC impacts experienced by the flight crews?	☐ YES ☐ NO	Were negative METOC impacts forecasted?	☐ YES ☐ NO

If "Yes" to either question continue below.

METOC Condition	Mission Phase Check all that apply.	Forecasted? Check all that apply.	Impact to Mission Check all that apply.	Corrective Action(s) Taken Check all that apply.	
Excessive Surface Winds	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ Ejection limit ☐ couldn't use glide weapon	☐ None ☐ Change Target/Location	☐ Canx Mission
Excessive Winds Aloft	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ couldn't drop weapon ☐ couldn't use glide weapon ☐ excessive fuel consumption	☐ None ☐ Change Altitude ☐ Change Target/Location	☐ Canx Mission ☐ Change Route ☐ Extra Tanking ☐ Less Time on Station
Turbulence	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ Safety of Flight ☐ couldn't tank	☐ None ☐ Change Altitude ☐ Change Target/Location	☐ Canx Mission ☐ Change Route
Icing	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ Safety of Flight	☐ None ☐ Change Altitude ☐ Change Target/Location	☐ Canx Mission ☐ Change Route
CBs	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ Safety of Flight	☐ None ☐ Change Altitude ☐ Change Target/Location	☐ Canx Mission ☐ Change Route
Cloud Layers/Thickness at Flight Level	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ Safety of Flight ☐ couldn't see target ☐ couldn't see ship ☐ couldn't use laser	☐ None ☐ Change Altitude ☐ Change Target/Location	☐ Canx Mission ☐ Change Route ☐ Divert
Cloud Layers/Thickness below Flight Level	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ couldn't see target ☐ couldn't use laser sensor/designator/weapon	☐ None ☐ Change Altitude ☐ Change Target/Location	☐ Canx Mission ☐ Change Route
Precipitation	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ Safety of Flight ☐ couldn't see target ☐ couldn't see ship ☐ couldn't use sensors	☐ None ☐ Change Altitude ☐ Change Target/Location	☐ Canx Mission ☐ Change Route ☐ Divert ☐ Case III Recovery
Slant Range Visibility	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ couldn't see target ☐ couldn't see ship ☐ couldn't use laser sensor/designator/weapon	☐ None ☐ Change Target/Location	☐ Canx Mission ☐ Divert
Reduced Surface Visibility due to Fog	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ couldn't see target ☐ couldn't see ship ☐ couldn't use laser sensor/designator/weapon	☐ None ☐ Change Target/Location	☐ Canx Mission ☐ Divert ☐ Case III Recovery
Reduced Surface Visibility due to Haze/Smoke	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ couldn't see target ☐ couldn't see ship ☐ couldn't use laser sensor/designator/weapon	☐ None ☐ Change Target/Location	☐ Canx Mission ☐ Divert ☐ Case III Recovery
Reduced Surface Visibility due to Dust	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ couldn't see target ☐ couldn't see ship ☐ couldn't use laser sensor/designator/weapon	☐ None ☐ Change Target/Location	☐ Canx Mission ☐ Divert ☐ Case III Recovery
Low Thermal Contrast	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ couldn't use thermal sensor	☐ None ☐ Change Target/Location	☐ Canx Mission

2. TAWS Performance.

What detection and lock on ranges were experienced?

Target ____________

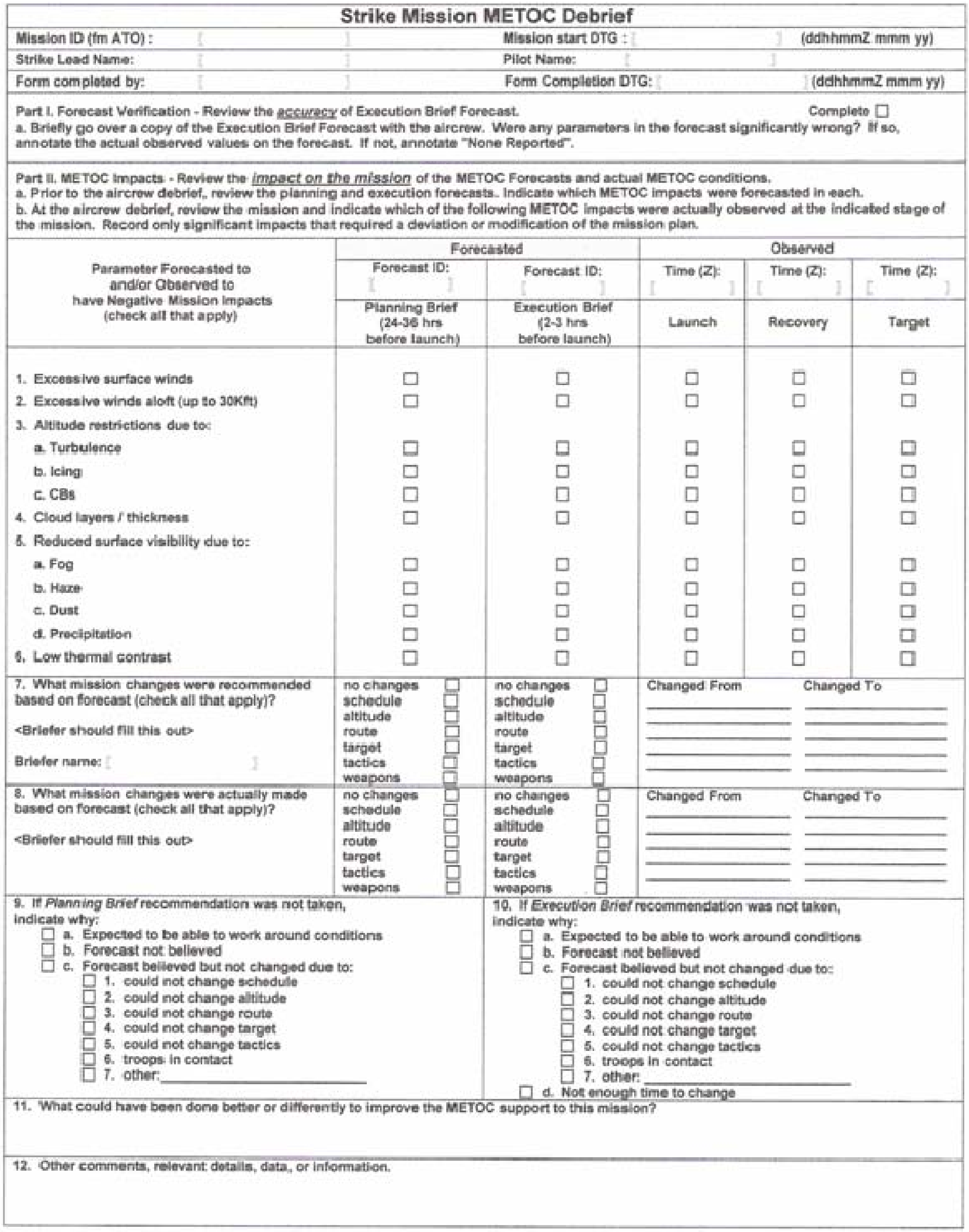
Detection Range (nm) ________ WFOV, MFOV, or NFOV (Circle One)

Lock On Range (nm) ________

3. Other Inputs or Comments.

What could have been done better or differently to improve the METOC support to this mission? Other comments or recommendations? Amplifying details regarding the questions above? (Continue on back as required)

Figure 11. Example of the paper data collection form that was used to collect post-mission data from the pilots for transfer to the online form at the NPS METOC metrics SIPRNet site. The form was designed such that most of the data only needed to be entered when there was a forecasted and/or observed severe negative weather impact. This allowed the form to be completed relative quickly for most missions.



Strike Mission METOC Debrief

Mission ID (fm ATO) : [ ] Mission start DTG : [ ] (ddhhmmZ mmm yy)

Strike Lead Name: [ ] Pilot Name: [ ]

Form completed by: [ ] Form Completion DTG: [ ] (ddhhmmZ mmm yy)

Part I. Forecast Verification - Review the *accuracy* of Execution Brief Forecast. Complete ☐

a. Briefly go over a copy of the Execution Brief Forecast with the aircrew. Were any parameters in the forecast significantly wrong? If so, annotate the actual observed values on the forecast. If not, annotate "None Reported".

Part II. METOC Impacts - Review the *impact on the mission* of the METOC Forecasts and actual METOC conditions.

a. Prior to the aircrew debrief, review the planning and execution forecasts. Indicate which METOC impacts were forecasted in each.

b. At the aircrew debrief, review the mission and indicate which of the following METOC impacts were actually observed at the indicated stage of the mission. Record only significant impacts that required a deviation or modification of the mission plan.

Parameter Forecasted to and/or Observed to have Negative Mission Impacts (check all that apply)	Forecasted		Observed		
	Forecast ID: [ ]	Forecast ID: [ ]	Time (Z): [ ]	Time (Z): [ ]	Time (Z): [ ]
	Planning Brief (24-36 hrs before launch)	Execution Brief (2-3 hrs before launch)	Launch	Recovery	Target
1. Excessive surface winds	☐	☐	☐	☐	☐
2. Excessive winds aloft (up to 30Kft)	☐	☐	☐	☐	☐
3. Altitude restrictions due to:					
a. Turbulence	☐	☐	☐	☐	☐
b. Icing	☐	☐	☐	☐	☐
c. CBs	☐	☐	☐	☐	☐
4. Cloud layers / thickness	☐	☐	☐	☐	☐
5. Reduced surface visibility due to:					
a. Fog	☐	☐	☐	☐	☐
b. Haze	☐	☐	☐	☐	☐
c. Dust	☐	☐	☐	☐	☐
d. Precipitation	☐	☐	☐	☐	☐
6. Low thermal contrast	☐	☐	☐	☐	☐

7. What mission changes were recommended based on forecast (check all that apply)? <Briefer should fill this out> Briefer name: [ ]	no changes ☐ schedule ☐ altitude ☐ route ☐ target ☐ tactics ☐ weapons ☐	no changes ☐ schedule ☐ altitude ☐ route ☐ target ☐ tactics ☐ weapons ☐	Changed From ________ ________ ________ ________ ________	Changed To ________ ________ ________ ________ ________
8. What mission changes were actually made based on forecast (check all that apply)? <Briefer should fill this out>	no changes ☐ schedule ☐ altitude ☐ route ☐ target ☐ tactics ☐ weapons ☐	no changes ☐ schedule ☐ altitude ☐ route ☐ target ☐ tactics ☐ weapons ☐	Changed From ________ ________ ________ ________ ________	Changed To ________ ________ ________ ________ ________

9. If *Planning Brief* recommendation was not taken, indicate why:

- ☐ a. Expected to be able to work around conditions
- ☐ b. Forecast not believed
- ☐ c. Forecast believed but not changed due to:
  - ☐ 1. could not change schedule
  - ☐ 2. could not change altitude
  - ☐ 3. could not change route
  - ☐ 4. could not change target
  - ☐ 5. could not change tactics
  - ☐ 6. troops in contact
  - ☐ 7. other:________________

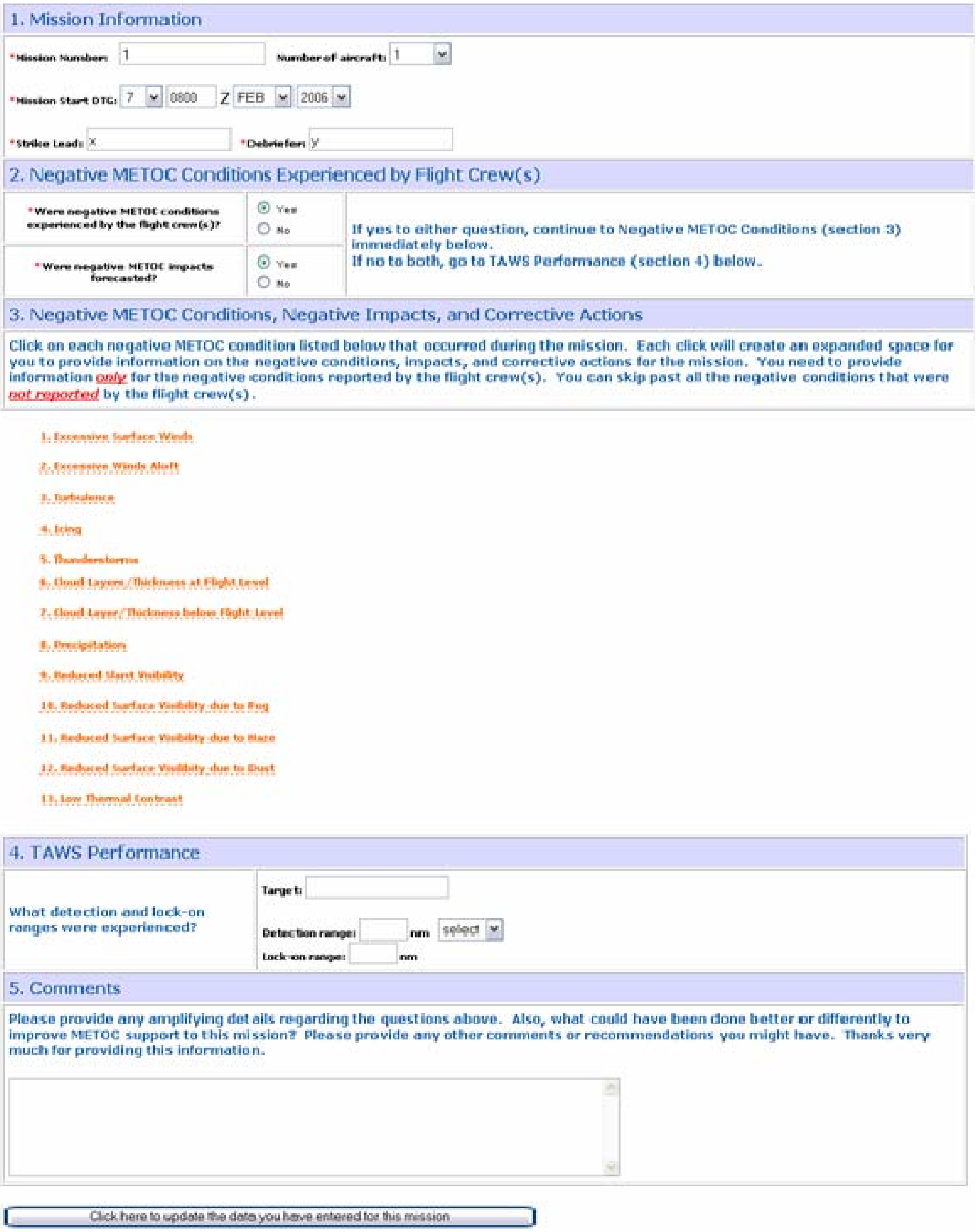
10. If *Execution Brief* recommendation was not taken, indicate why:

- ☐ a. Expected to be able to work around conditions
- ☐ b. Forecast not believed
- ☐ c. Forecast believed but not changed due to:
  - ☐ 1. could not change schedule
  - ☐ 2. could not change altitude
  - ☐ 3. could not change route
  - ☐ 4. could not change target
  - ☐ 5. could not change tactics
  - ☐ 6. troops in contact
  - ☐ 7. other: ________________
- ☐ d. Not enough time to change

11. What could have been done better or differently to improve the METOC support to this mission?

12. Other comments, relevant details, data, or information.

Figure 12. Example of the original proposed paper data collection form. The form proved to be inadequate due to the lack of mitigation recommendations from the OA division during mission planning.



1. Mission Information

*Mission Number: 1 Number of aircraft: 1

*Mission Start DTG: 7 0800 Z FEB 2006

*Strike Lead: X *Debriefer: Y

2. Negative METOC Conditions Experienced by Flight Crew(s)

*Were negative METOC conditions experienced by the flight crew(s)?	● Yes ○ No	If yes to either question, continue to Negative METOC Conditions (section 3) immediately below. If no to both, go to TAWS Performance (section 4) below.
*Were negative METOC impacts forecasted?	● Yes ○ No	

3. Negative METOC Conditions, Negative Impacts, and Corrective Actions

Click on each negative METOC condition listed below that occurred during the mission. Each click will create an expanded space for you to provide information on the negative conditions, impacts, and corrective actions for the mission. You need to provide information *only* for the negative conditions reported by the flight crew(s). You can skip past all the negative conditions that were *not reported* by the flight crew(s).

1. Excessive Surface Winds
2. Excessive Winds Aloft
3. Turbulence
4. Icing
5. Thunderstorms
6. Cloud Layers/Thickness at Flight Level
7. Cloud Layer/Thickness below Flight Level
8. Precipitation
9. Reduced Slant Visibility
10. Reduced Surface Visibility due to Fog
11. Reduced Surface Visibility due to Haze
12. Reduced Surface Visibility due to Dust
13. Low Thermal Contrast

4. TAWS Performance

What detection and lock-on ranges were experienced?

Target:

Detection range: nm select

Lock-on range: nm

5. Comments

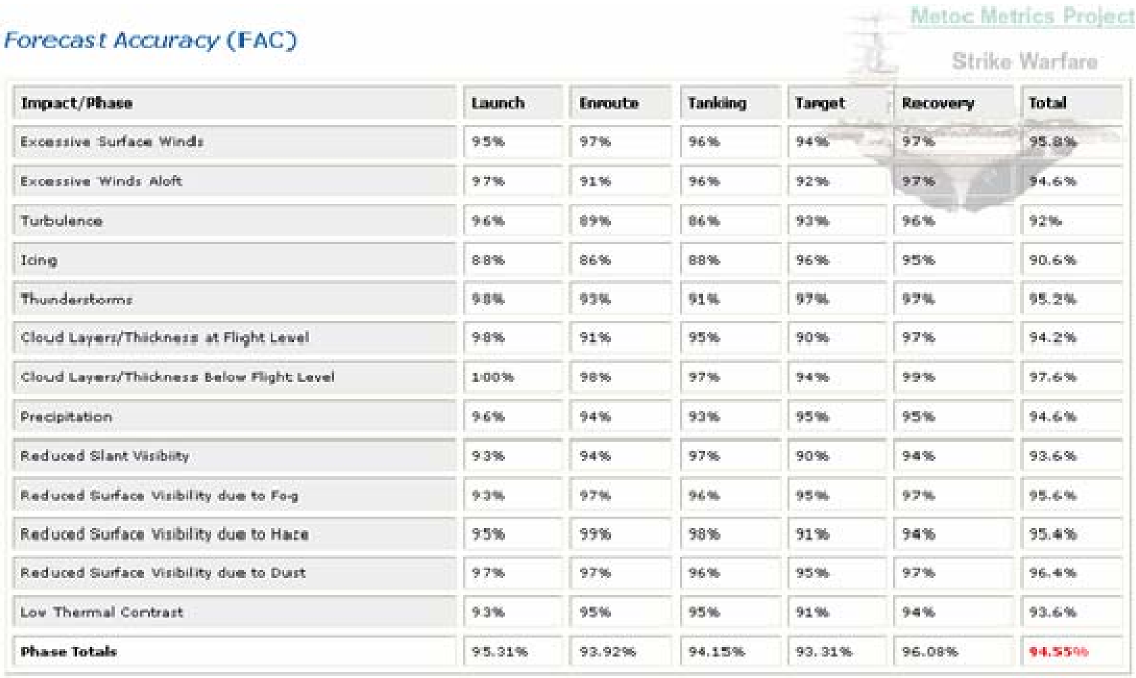
Please provide any amplifying details regarding the questions above. Also, what could have been done better or differently to improve METOC support to this mission? Please provide any other comments or recommendations you might have. Thanks very much for providing this information.

Figure 13. Snapshot of the online data collection form corresponding to the paper data collection form shown in Figure 12. The online form was designed to be as similar to the paper form as possible. The form also hides the individual weather impact fields until the user needs them resulting in a less confusing form.

1. Excessive Surface Winds

Negative METOC Condition	Mission Phase (Check all that apply)	Was Negative METOC Condition Forecasted?	Impacts to Mission (Check all that apply)	Corrective Actions Taken (Check all that apply)	
Excessive Surface Winds	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ LAU ☐ TNK ☐ ENR ☐ TGT ☐ REC	☐ Ejection limit ☐ Couldn't use glide weapon ☐ Other. Provide details in section 5 comment box below.	☐ None ☐ Change target/location	☐ Cancel mission ☐ Other. Provide details in section 5 comment box below.

Figure 14. Example of an expanded weather impact field from the online collection form. The check boxes are in the same layout as the paper form to allow for easy transfer of information.



Forecast Accuracy (FAC)

Metoc Metrics Project

Strike Warfare

Impact/Phase	Launch	Enroute	Tanking	Target	Recovery	Total
Excessive Surface Winds	95%	97%	96%	94%	97%	95.8%
Excessive Winds Aloft	97%	91%	96%	92%	97%	94.6%
Turbulence	96%	89%	86%	93%	96%	92%
Icing	88%	86%	88%	96%	95%	90.6%
Thunderstorms	98%	93%	91%	97%	97%	95.2%
Cloud Layers/Thickness at Flight Level	98%	91%	95%	90%	97%	94.2%
Cloud Layers/Thickness Below Flight Level	100%	98%	97%	94%	99%	97.6%
Precipitation	96%	94%	93%	95%	95%	94.6%
Reduced Slant Visibiity	93%	94%	97%	90%	94%	93.6%
Reduced Surface Visibility due to Fog	93%	97%	96%	95%	97%	95.6%
Reduced Surface Visibility due to Haze	95%	99%	98%	91%	94%	95.4%
Reduced Surface Visibility due to Dust	97%	97%	96%	95%	97%	96.4%
Low Thermal Contrast	93%	95%	95%	91%	94%	93.6%
Phase Totals	95.31%	93.92%	94.15%	93.31%	96.08%	94.55%

Figure 15. Example of tabular output from the development site for FAC using test data, not real world data. The actual results look similar but with real world operational data. FAC is the percentage of the total number of forecasts that were correct. 100% is a perfect score. The columns indicate the FAC at each phase for each impact, with the last column the total FAC for each impact. The bottom row indicates the total FAC per mission phase, and the red number is the total FAC for all phases and impacts. For details, see Chapter II, Section B.1.

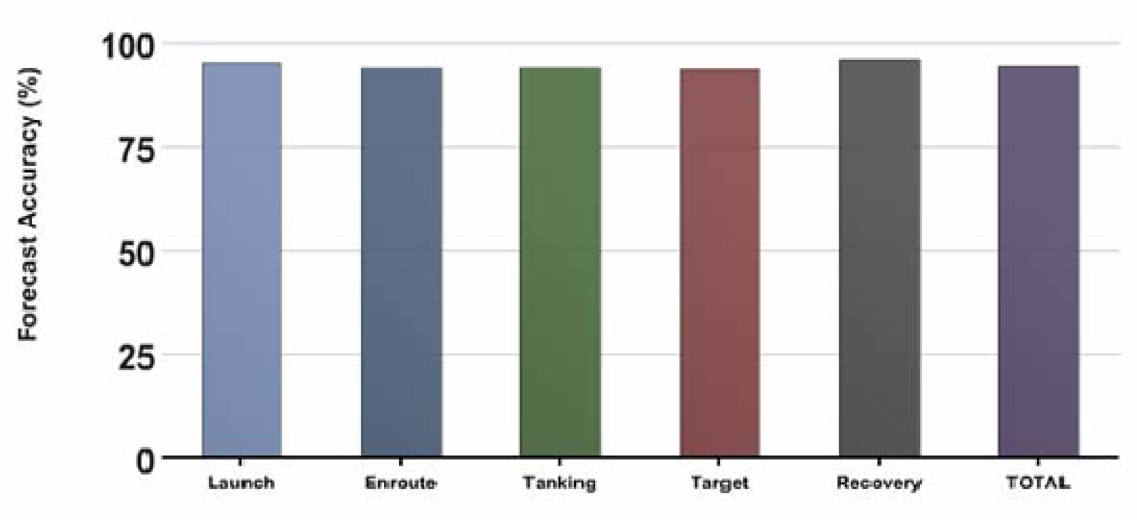


Figure 16. Example of graphical output from the development site for FAC using test data, not real world data. The actual results look similar but with real world operational data. The bars indicate the FAC for each mission phase and for all phases combined, based on all types of impacts. For details, see Chapter II, Section B.1.

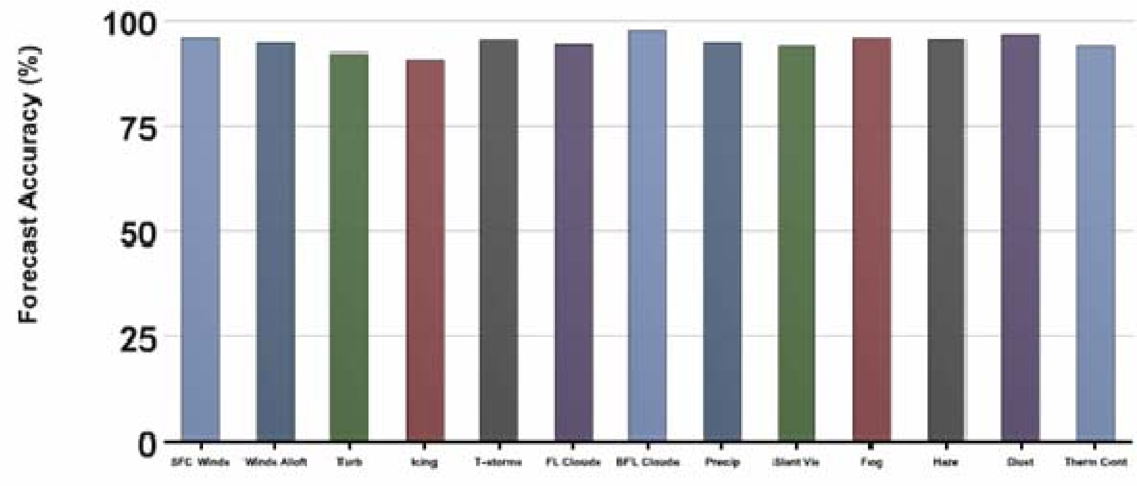
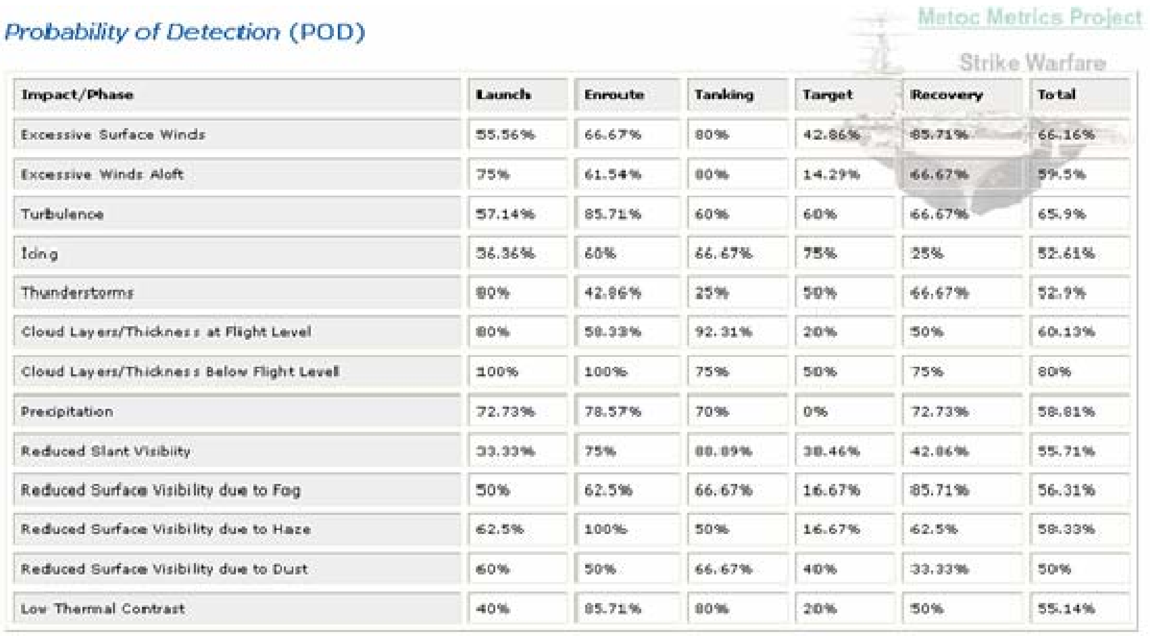


Figure 17. Example of graphical output from the development site for FAC using test data, not real world data. The actual results look similar but with real world operational data. The bars indicate the FAC for each type of impact, based on all mission phases. For details, see Chapter II, Section B.1.



Probability of Detection (POD)

Impact/Phase	Launch	Enroute	Tanking	Target	Recovery	Total
Excessive Surface Winds	55.56%	66.67%	80%	42.86%	85.71%	66.16%
Excessive Winds Aloft	75%	61.54%	80%	14.29%	66.67%	59.5%
Turbulence	57.14%	85.71%	60%	60%	66.67%	65.9%
Icing	36.36%	60%	66.67%	75%	25%	52.61%
Thunderstorms	80%	42.86%	25%	50%	66.67%	52.9%
Cloud Layers/Thickness at Flight Level	80%	58.33%	92.31%	20%	50%	60.13%
Cloud Layers/Thickness Below Flight Level	100%	100%	75%	50%	75%	80%
Precipitation	72.73%	78.57%	70%	0%	72.73%	58.81%
Reduced Slant Visibity	33.33%	75%	88.89%	38.46%	42.86%	55.71%
Reduced Surface Visibility due to Fog	50%	62.5%	66.67%	16.67%	85.71%	56.31%
Reduced Surface Visibility due to Haze	62.5%	100%	50%	16.67%	62.5%	58.33%
Reduced Surface Visibility due to Dust	60%	50%	66.67%	40%	33.33%	50%
Low Thermal Contrast	40%	85.71%	80%	20%	50%	55.14%

Figure 18. Example of tabular output from the development site for POD using test data, not real world data. The actual results look similar but with real world operational data. POD indicates the percentage of times an event happened that was forecasted to happen. 100% is a perfect score. The columns indicate the POD at each phase for each impact, with the last column showing the total POD for each impact. For details, see Chapter II, Section B.1.

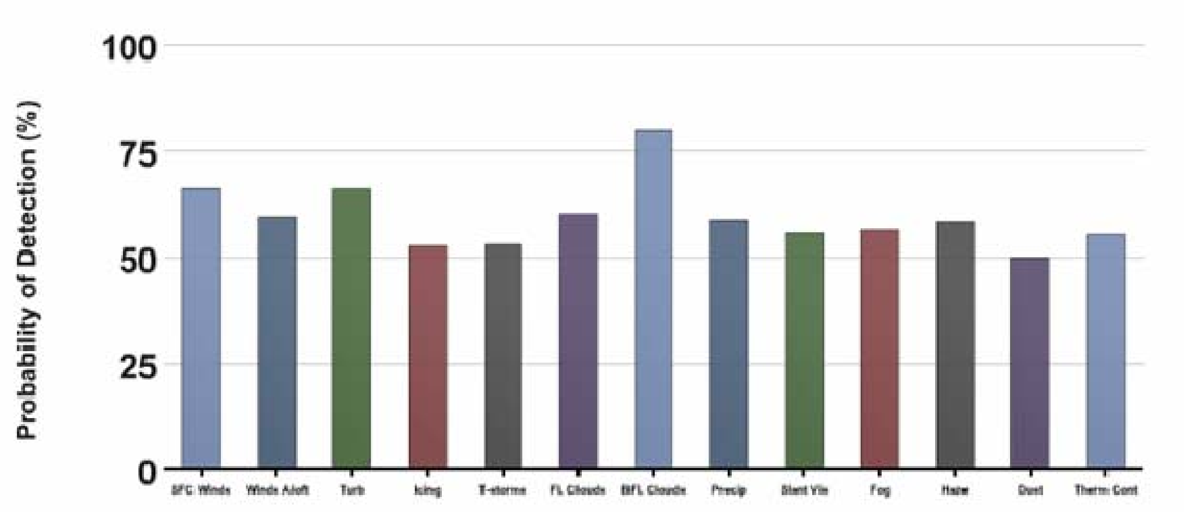
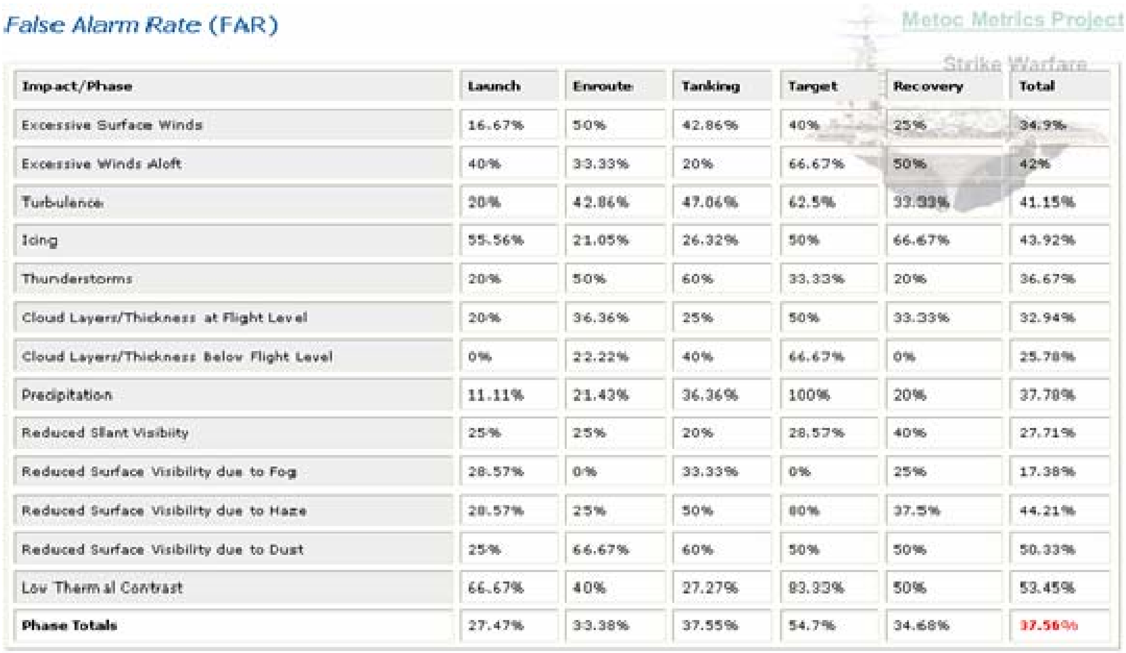


Figure 19. Example of graphical output from the development site for POD using test data, not real world data. The actual results look similar but with real world operational data. The bars indicate the POD for each mission phase and for all phases combined, based on all types of impacts. For details, see Chapter II, Section B.1.



Impact/Phase	Launch	Enroute	Tanking	Target	Recovery	Total
Excessive Surface Winds	16.67%	50%	42.86%	40%	25%	34.9%
Excessive Winds Aloft	40%	33.33%	20%	66.67%	50%	42%
Turbulence	20%	42.86%	47.06%	62.5%	33.33%	41.15%
Icing	55.56%	21.05%	26.32%	50%	66.67%	43.92%
Thunderstorms	20%	50%	60%	33.33%	20%	36.67%
Cloud Layers/Thickness at Flight Level	20%	36.36%	25%	50%	33.33%	32.94%
Cloud Layers/Thickness Below Flight Level	0%	22.22%	40%	66.67%	0%	25.78%
Precipitation	11.11%	21.43%	36.36%	100%	20%	37.78%
Reduced Slant Visibiity	25%	25%	20%	28.57%	40%	27.71%
Reduced Surface Visibility due to Fog	28.57%	0%	33.33%	0%	25%	17.38%
Reduced Surface Visibility due to Haze	28.57%	25%	50%	80%	37.5%	44.21%
Reduced Surface Visibility due to Dust	25%	66.67%	60%	50%	50%	50.33%
Low Thermal Contrast	66.67%	40%	27.27%	83.33%	50%	53.45%
Phase Totals	27.47%	33.38%	37.55%	54.7%	34.68%	37.56%

Figure 20. Example of tabular output from the development site for FAR using test data, not real world data. The actual results look similar but with real world operational data. FAR is an indication of the percentage of times an event was incorrectly forecasted to happen. Zero is a perfect score. The columns indicate the FAR at each phase for each impact, with the last column total FAR for each impact. The bottom row is the total FAR per phase and the red number is the total FAR for all phases and impacts.

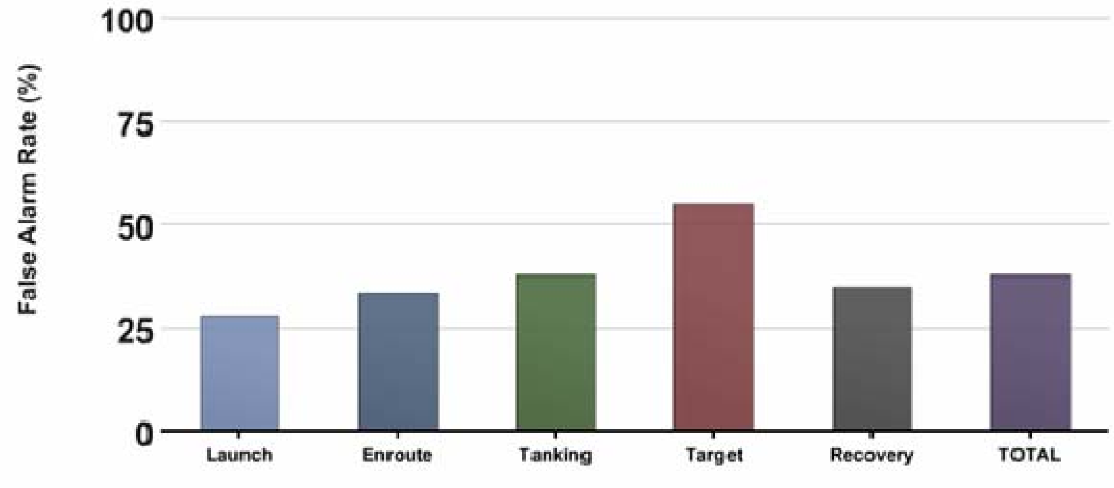


Figure 21. Example of graphical output from the development site for FAR using test data, not real world data. The actual results look similar but with real world operational data. The bars indicate the FAR for each mission phase and for all phases combined, based on all types of impacts. For details, see Chapter II, Section B.1.

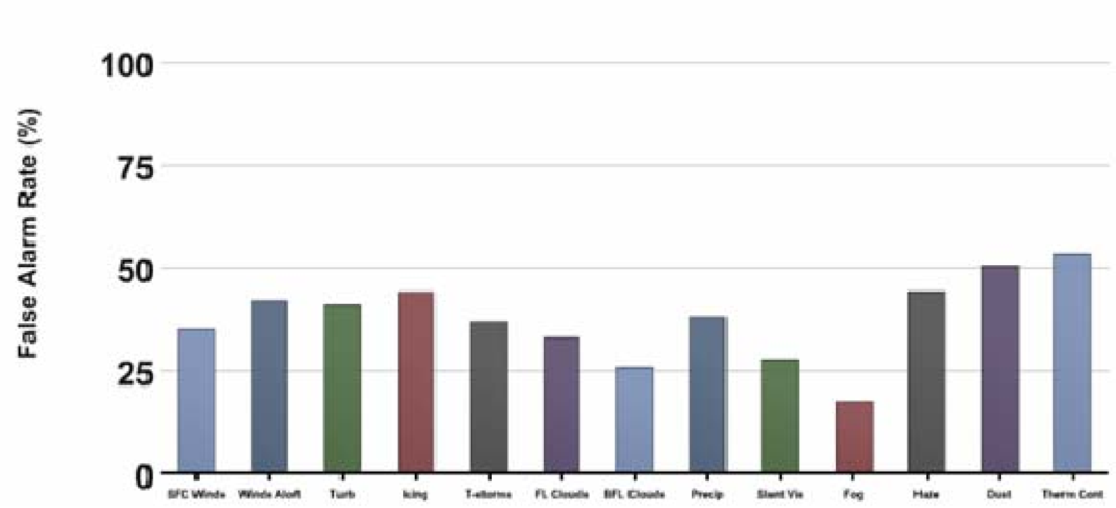
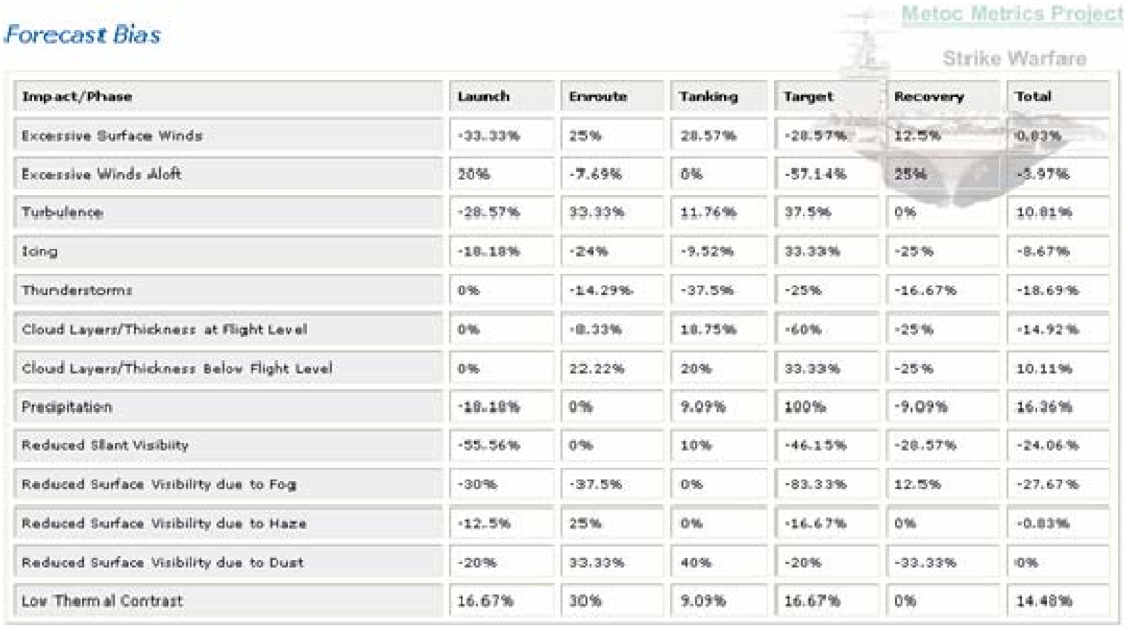


Figure 22. Example of graphical output from the development site for FAR using test data, not real world data. The actual results look similar but with real world operational data. The bars indicate the FAR for each type of impact, based on all mission phases. For details, see Chapter II, Section B.1.



Forecast Bias

Impact/Phase	Launch	Enroute	Tanking	Target	Recovery	Total
Excessive Surface Winds	-33.33%	25%	28.57%	-28.57%	12.5%	0.83%
Excessive Winds Aloft	20%	-7.69%	0%	-57.14%	25%	-3.97%
Turbulence	-28.57%	33.33%	11.76%	37.5%	0%	10.81%
Icing	-18.18%	-24%	-9.52%	33.33%	-25%	-8.67%
Thunderstorms	0%	-14.29%	-37.5%	-25%	-16.67%	-18.69%
Cloud Layers/Thickness at Flight Level	0%	-8.33%	18.75%	-60%	-25%	-14.92%
Cloud Layers/Thickness Below Flight Level	0%	22.22%	20%	33.33%	-25%	10.11%
Precipitation	-18.18%	0%	9.09%	100%	-9.09%	16.36%
Reduced Slant Visibiity	-55.56%	0%	10%	-46.15%	-28.57%	-24.06%
Reduced Surface Visibility due to Fog	-30%	-37.5%	0%	-83.33%	12.5%	-27.67%
Reduced Surface Visibility due to Haze	-12.5%	25%	0%	-16.67%	0%	-0.83%
Reduced Surface Visibility due to Dust	-20%	33.33%	40%	-20%	-33.33%	0%
Low Thermal Contrast	16.67%	30%	9.09%	16.67%	0%	14.48%

Figure 23. Example of tabular output from the development site for bias using test data, not real world data. The actual results look similar but with real world operational data. Bias is an indication of a tendency to over-forecast or under-forecast a type of event. Zero is a perfect score. Positive scores indicate over-forecasting and negative scores indicate under-forecasting. The columns indicate the bias at each phase for each impact, with the last column the total bias for each impact. For details, see Chapter II, Section B.1.

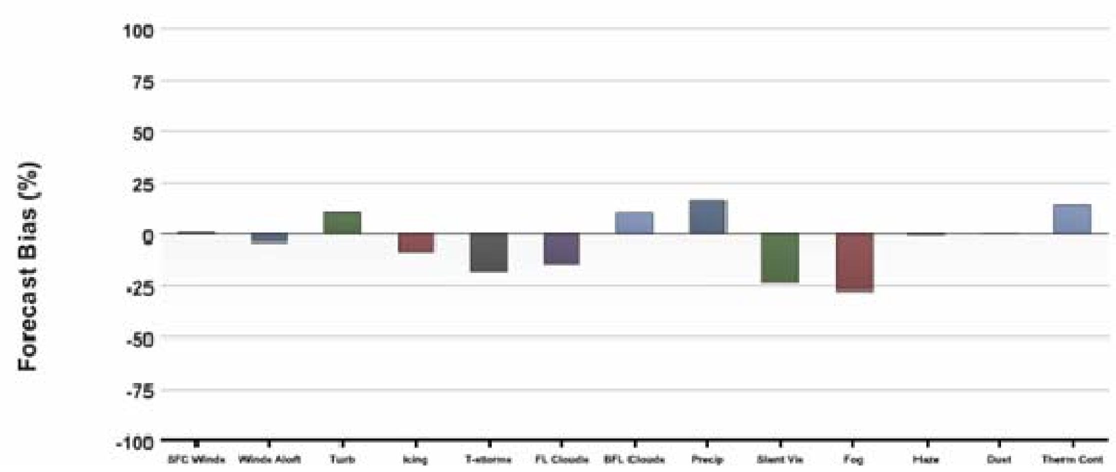


Figure 24. Example of graphical output from the development site for bias using test data, not real world data. The actual results look similar but with real world operational data. The bars indicate the bias for each type of impact, based on all mission phases. For details, see Chapter II, Section B.1.

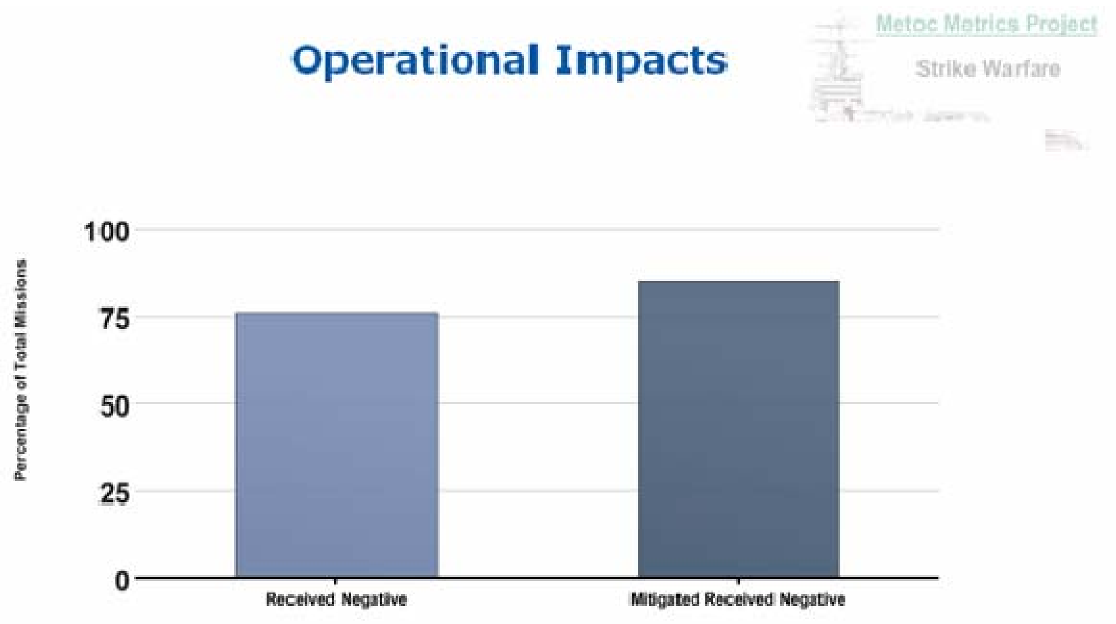


Figure 25. Example of graphical operational impacts output from the developmental site using test data, not real world data. The actual results look similar but with real world data. The bars indicate the received negative metric (indicating the percentage of mission in which the pilots received a negative impact) and the mitigated received negative metric (indicating the percentage of missions that received a negative impact forecast and the pilot took mitigating action in response to encountering negative phenomena). For details, see Chapter II, Section B.2.

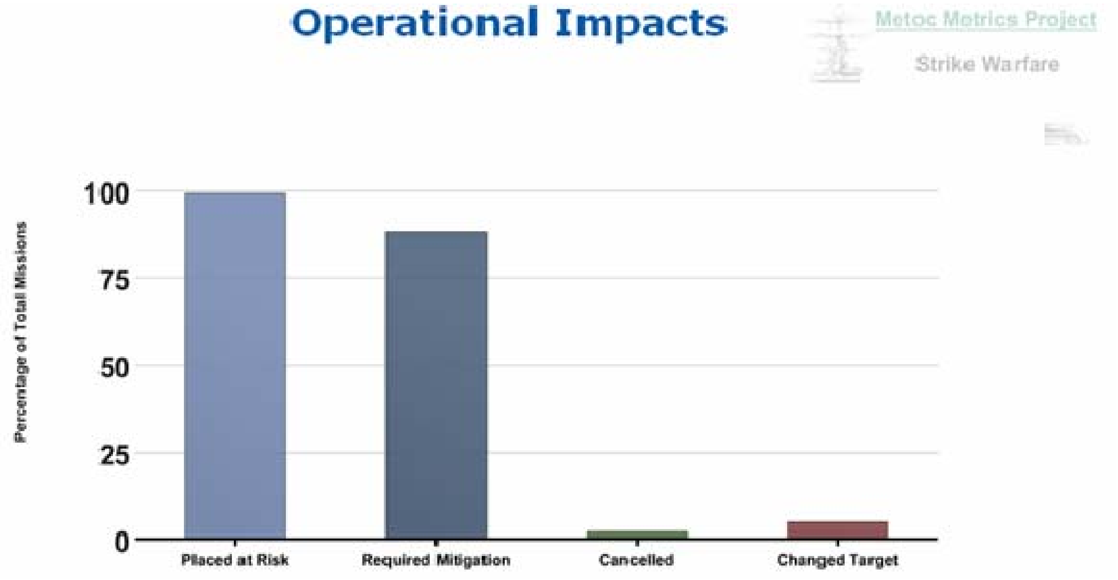


Figure 26. Example of the operational impacts graphic output from the developmental site. The actual results look similar but with operational data. The graph indicates the percentage of total missions that were affected by weather, had a mitigating action due to weather, that were canceled due to weather, and that had a target change due to weather.

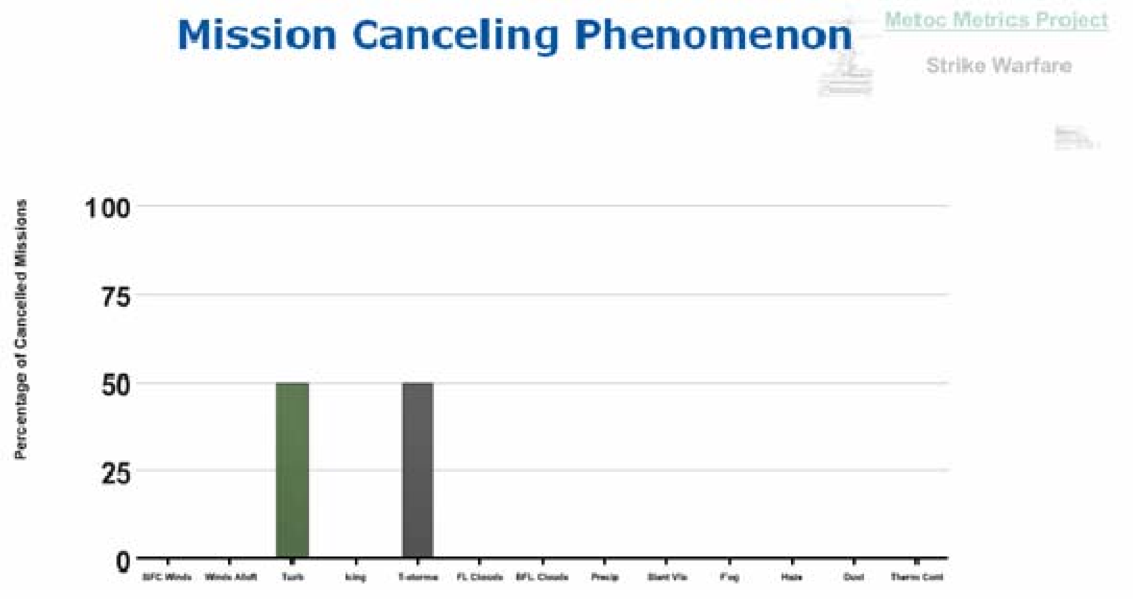


Figure 27. Example of the mission canceling phenomenon output from the developmental site. The actual results look similar but with operational data. The graph shows the percentage of cancelled missions that were caused by each type of weather impact.

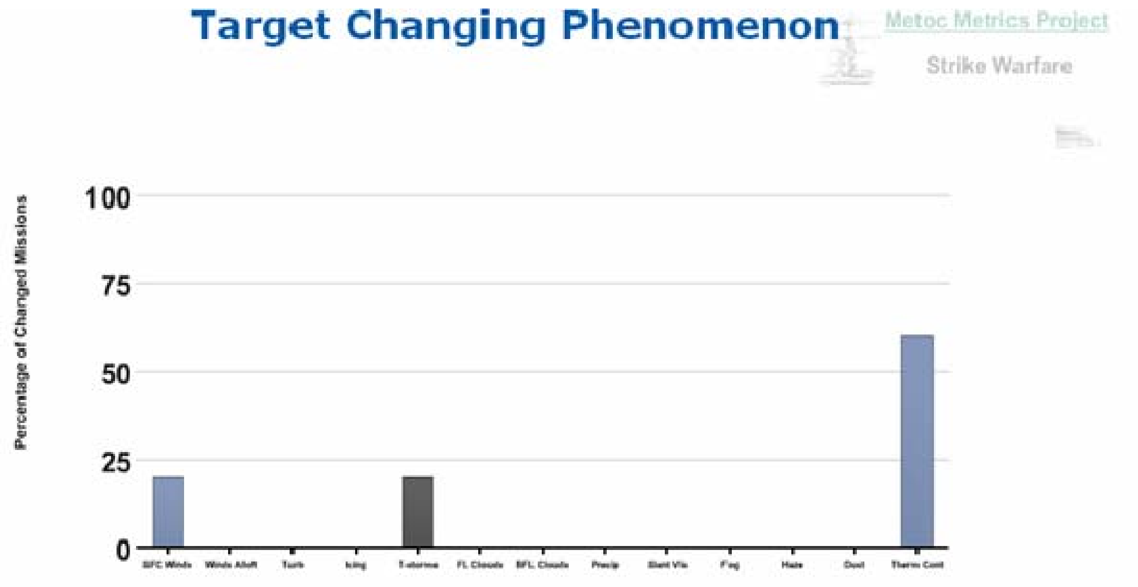


Figure 28. Example of the target changing phenomenon output from the developmental site. The actual results look similar but with operational data. The graph shows the percentage of missions with target changes that were caused by each type of weather impact.

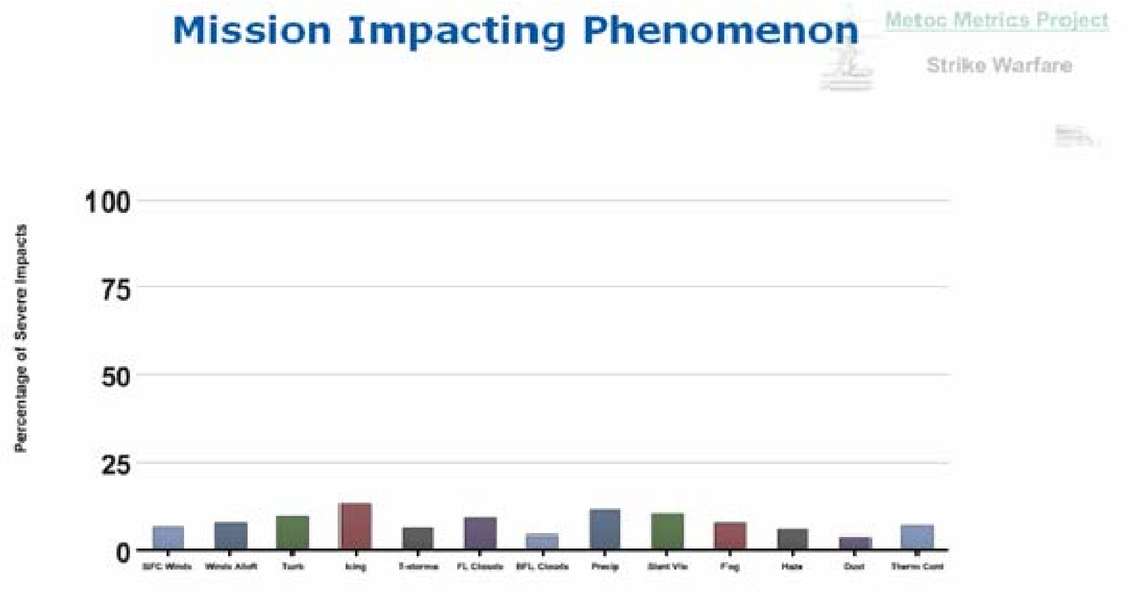
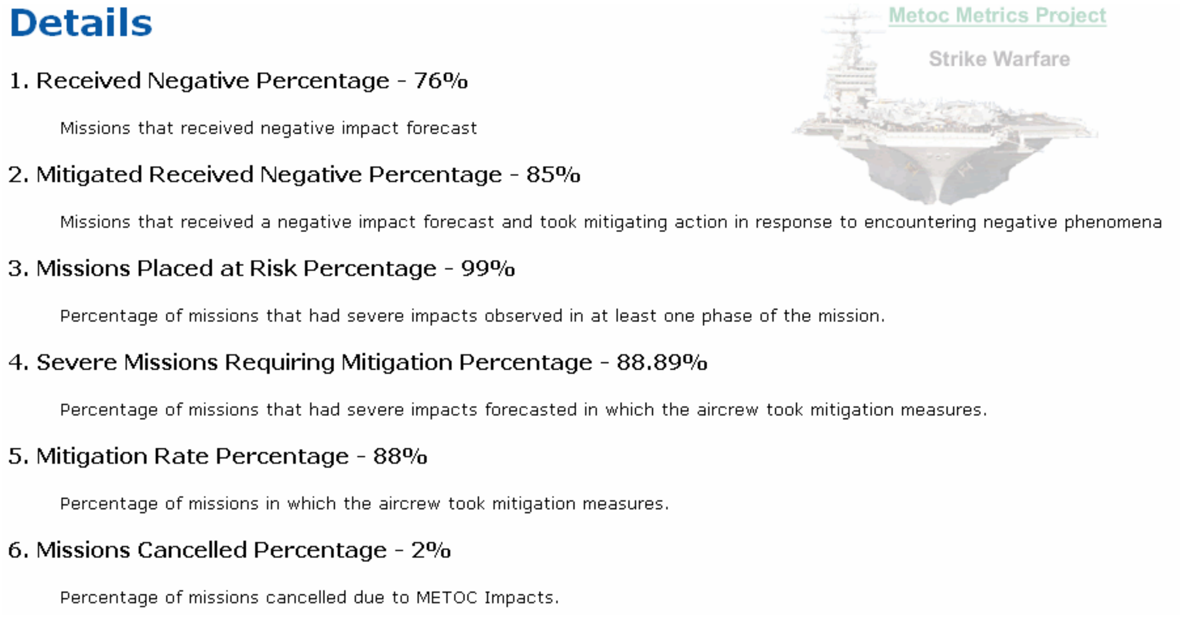


Figure 29. Example of the mission impacting phenomenon output from the developmental site. The actual results look similar but with operational data. The graph shows the percentage of total observed severe impacts that were caused by each type of weather impact.



Details

1. Received Negative Percentage - 76%

   Missions that received negative impact forecast

2. Mitigated Received Negative Percentage - 85%

   Missions that received a negative impact forecast and took mitigating action in response to encountering negative phenomena

3. Missions Placed at Risk Percentage - 99%

   Percentage of missions that had severe impacts observed in at least one phase of the mission.

4. Severe Missions Requiring Mitigation Percentage - 88.89%

   Percentage of missions that had severe impacts forecasted in which the aircrew took mitigation measures.

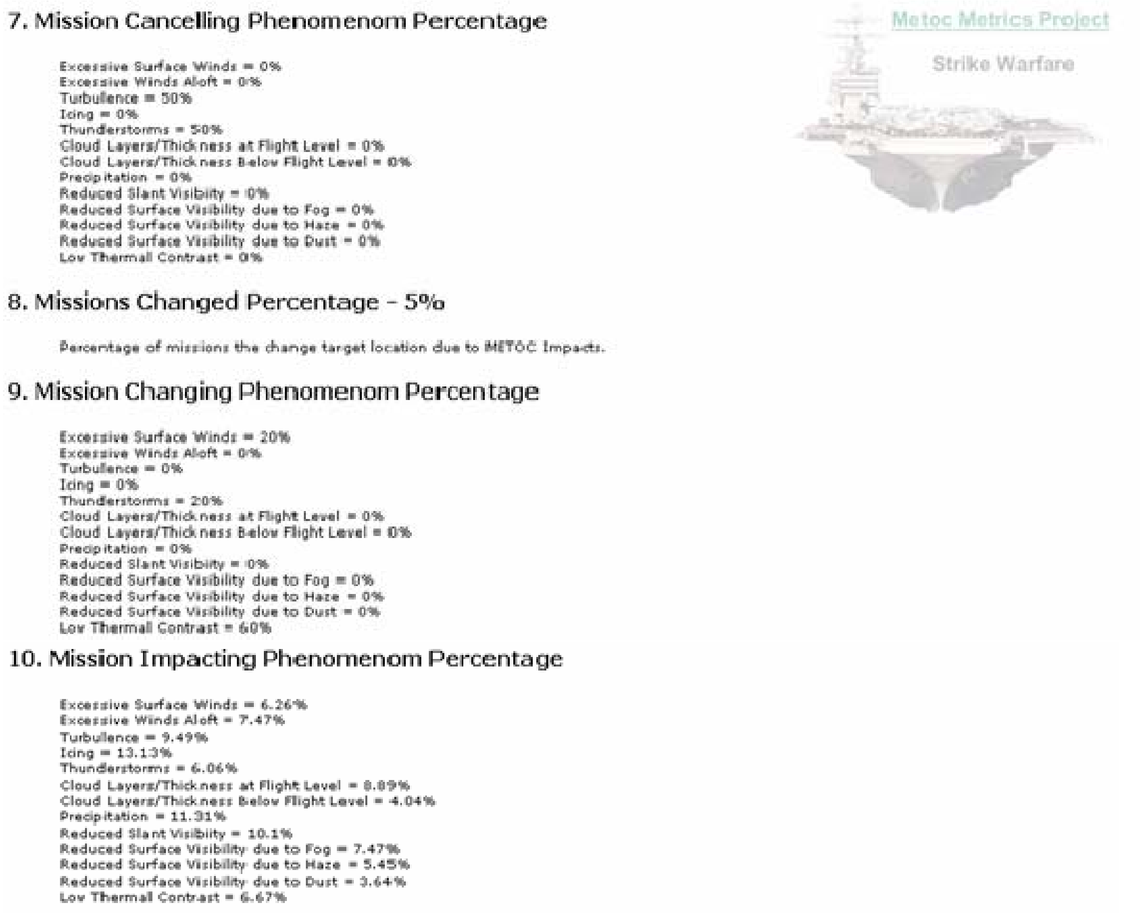
5. Mitigation Rate Percentage - 88%

   Percentage of missions in which the aircrew took mitigation measures.

6. Missions Cancelled Percentage - 2%

   Percentage of missions cancelled due to METOC Impacts.

Figure 30. Example of textual output from the development site using test data, not real world data. The actual results look similar but with real world data. The text outputs show the same results as in the operational impacts figures but in text format.



Metoc Metrics Project

Strike Warfare

7. Mission Cancelling Phenomenom Percentage

Excessive Surface Winds = 0%
Excessive Winds Aloft = 0%
Turbulence = 50%
Icing = 0%
Thunderstorms = 50%
Cloud Layers/Thickness at Flight Level = 0%
Cloud Layers/Thickness Below Flight Level = 0%
Precipitation = 0%
Reduced Slant Visibiity = 0%
Reduced Surface Visibility due to Fog = 0%
Reduced Surface Visibility due to Haze = 0%
Reduced Surface Visibility due to Dust = 0%
Low Thermall Contrast = 0%

8. Missions Changed Percentage - 5%

Percentage of missions the change target location due to METOC Impacts.

9. Mission Changing Phenomenom Percentage

Excessive Surface Winds = 20%
Excessive Winds Aloft = 0%
Turbulence = 0%
Icing = 0%
Thunderstorms = 20%
Cloud Layers/Thickness at Flight Level = 0%
Cloud Layers/Thickness Below Flight Level = 0%
Precipitation = 0%
Reduced Slant Visibiity = 0%
Reduced Surface Visibility due to Fog = 0%
Reduced Surface Visibility due to Haze = 0%
Reduced Surface Visibility due to Dust = 0%
Low Thermall Contrast = 60%

10. Mission Impacting Phenomenom Percentage

Excessive Surface Winds = 6.26%
Excessive Winds Aloft = 7.47%
Turbulence = 9.49%
Icing = 13.13%
Thunderstorms = 6.06%
Cloud Layers/Thickness at Flight Level = 8.89%
Cloud Layers/Thickness Below Flight Level = 4.04%
Precipitation = 11.31%
Reduced Slant Visibiity = 10.1%
Reduced Surface Visibility due to Fog = 7.47%
Reduced Surface Visibility due to Haze = 5.45%
Reduced Surface Visibility due to Dust = 3.64%
Low Thermall Contrast = 6.67%

Figure 31. Continuation of Figure 30.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia

2. Dudley Knox Library
Naval Postgraduate School
Monterey, California

3. Professor Tom Murphree
Naval Postgraduate School
Monterey, California

4. Commander Rebecca Stone
Naval Postgraduate School
Monterey, California

5. Colonel David Smarsh
Naval Postgraduate School
Monterey, California

6. Commander Steve Woll
Naval Atlantic Meteorology and Oceanography Center
Norfolk, Virginia

7. Captain Victor Addison
Fleet Numerical Meteorology and Oceanography Center
Monterey, California

8. Commander Grant Cooper
Commander, Navy Meteorology and Oceanography Command
Bay St. Louis, Mississippi

9. Commander Tim Gallaudet
Naval Special Warfare Command
San Diego, California

10. Lieutenant Commander Jake Hinz
United States Naval Observatory
Washington, District of Columbia

11. Major Don Harper
Air Force Weather Agency
Offutt Air Force Base, Nebraska

12. Major Jeff Jarry
Air Force Weather Agency
Offutt Air Force Base, Nebraska

13. Lieutenant Commander Brett Martin
NAVMETOCPRODEVCEN PAC DET
San Diego, California

14. Lieutenant Commander Mark Butler
USS Enterprise
Norfolk, Virginia

15. Lieutenant Commander Keith Everett
USS Enterprise
Norfolk, Virginia

16. Bruce Ford
Clear Science, Inc.
Jacksonville, Florida