



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

**COMMUNICATING OPTIMIZED DECISION INPUT
FROM STOCHASTIC TURBULENCE FORECASTS**

by

Jeanne R. Szczes

March 2008

Thesis Advisor:
Second Reader:

F. Anthony Eckel
Patrick A. Harr

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE March 2008	3. REPORT TYPE AND DATES COVERED Master's Thesis	
4. TITLE AND SUBTITLE Communicating Optimized Decision Input from Stochastic Turbulence Forecasts			5. FUNDING NUMBERS	
6. AUTHOR(S) Jeanne R. Szczes				
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	
13. ABSTRACT (maximum 200 words) The uncertainty of weather forecasts contributes to mission risk. Ensemble data can improve combat capability by incorporating forecast uncertainty into the warfighter decision process. The study transforms raw ensemble data into optimized decision inputs for upper level turbulence using ORM principles and decision science. It demonstrates the methodology and importance of incorporating ambiguity, the uncertainty in forecast uncertainty, into the decision making process using the Taijitu method to estimate ambiguity. Comparing ambiguity and risk tolerance uncertainty intervals produces a more appropriate decision input compared to currently existing methods. Significant differences between the current and research derived decision input products demonstrate potential value added to decision making by incorporating ambiguity information. An effective visualization is devised for varying levels of risk tolerance and mission thresholds that is educational and practical for users. Research procedures and results can serve as an example to further education and development of stochastic methods in the Air Force and Department of Defense.				
14. SUBJECT TERMS Stochastic, Ensemble, Uncertainty, Probability, Ellrod-Knapp Turbulence Index, Uniform Ranks Method, Risk Tolerance, Optimum Decision Threshold, Cost-Loss Ratio, Ambiguity, Taijitu Method, Decision Input			15. NUMBER OF PAGES 159	
			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 UU	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)

Prescribed by ANSI Std. Z39-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

**COMMUNICATING OPTIMIZED DECISION INPUT FROM STOCHASTIC
TURBULENCE FORECASTS**

Jeanne R. Szczes
Captain, United States Air Force
B.S., Florida State University, 2002

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY

from the

**NAVAL POSTGRADUATE SCHOOL
March 2008**

Author: Jeanne Szczes

Approved by: F. Anthony Eckel
Thesis Advisor

Patrick A. Harr
Second Reader

Philip A. Durkee
Chairman, Department of Meteorology

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The uncertainty of weather forecasts contributes to mission risk. Ensemble data can improve combat capability by incorporating forecast uncertainty into the warfighter decision process. The study transforms raw ensemble data into optimized decision inputs for upper level turbulence using ORM principles and decision science. It demonstrates the methodology and importance of incorporating ambiguity, the uncertainty in forecast uncertainty, into the decision making process using the Taijitu method to estimate ambiguity. Comparing ambiguity and risk tolerance uncertainty intervals produces a more appropriate decision input compared to currently existing methods. Significant differences between the current and research derived decision input products demonstrate potential value added to decision making by incorporating ambiguity information. An effective visualization is devised for varying levels of risk tolerance and mission thresholds that is educational and practical for users. Research procedures and results can serve as an example to further education and development of stochastic methods in the Air Force and Department of Defense.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
II.	BACKGROUND	3
A.	ATMOSPHERIC PREDICTION.....	3
1.	Atmospheric Uncertainty	3
2.	Methods for Prediction.....	6
B.	ENSEMBLE FORECASTING (EF)	8
1.	General Description	8
2.	Range of Possibilities	9
3.	Ambiguity	10
C.	WEATHER RISK MANAGEMENT.....	14
1.	Value and Cost-Loss Analysis.....	14
a.	<i>Measure of Value</i>	14
b.	<i>Risk Tolerance and Cost-Loss Ratio</i>	18
2.	Operational Risk Management.....	19
D.	CALCULATING AND COMMUNICATING FORECAST TURBULENCE AND ITS UNCERTAINTY	20
1.	Today's Deterministic Methods.....	20
2.	WRAP	22
3.	Conclusion	25
III.	METHODOLOGY	27
A.	PART ONE	28
1.	Model Description and Ingest.....	28
2.	Turbulence Index (TI) Calculation	29
3.	Interpolation.....	32
4.	Forecast Probability Calculations	36
5.	Validation.....	39
B.	PART TWO.....	41
C.	PART THREE.....	53
IV.	RESULTS	61
A.	EVOLUTION OF TAIJITU CURVES THROUGH TIME WITH INCREASING LEAD TIME	61
B.	ANALYSIS OF DECISION INPUT (DI) WITH ADDED UNCERTAINTY INFORMATION	65
1.	Simulated DI Product and Analysis Construction.....	65
2.	Simulated IWEDA versus Simulated WRAP.....	69
a.	<i>Fall Case: FH 06</i>	69
b.	<i>Fall Case: FH 84</i>	76
3.	Simulated WRAP versus RSCH Results.....	82
a.	<i>Fall Case: FH 06</i>	82
b.	<i>Fall Case: FH 84</i>	90

V.	CONCLUSIONS	97
A.	FINAL REMARKS.....	97
B.	FURTHER RESEARCH AND RECOMMENDATIONS	99
	APPENDIX A: RESEARCH DI WITH DIFFERENT CI OF AMBIGUITY.....	103
	APPENDIX B: WINTER CASE: SIMULATED IWEDA VERSUS SIMULATED	
	WRAP	107
A.	FORECAST HOUR: 06	107
B.	FORECAST HOUR: 84	114
	APPENDIX C: WINTER CASE: SIMULATED WRAP VERSUS RSCH	
	RESULTS	121
A.	FORECAST HOUR: 06	121
B.	FORECAST HOUR: 84	128
	LIST OF REFERENCES	135
	INITIAL DISTRIBUTION LIST	137

LIST OF FIGURES

Figure 1.	a) Dynamical System with instabilities b) Stable Dynamical System with stationary or periodic orbits [From Kalnay, 2003].	5
Figure 2.	Current deterministic format for USAFE Upper Level Turbulence Chart; Intensities of turbulence (TURBC): MDT=moderate, SVR=severe, EXTRM=extreme; FL=Flight Level; CAT II ACFT=Category II aircraft [From USAFE OWS, 2007].	7
Figure 3.	JEFS Non-Operational Turbulence Chart with varying probability to turbulence occurrence in southeast Asia [From JEFS, 2007].	9
Figure 4.	Sampling distributions of the (a) ensemble mean and (b) ensemble spread (i.e., standard deviation), labeled by the number of ensemble members [From Eckel, 2008].	13
Figure 5.	Variation of value V with decision threshold (p_t) for ECMWF EPS probability forecasts for users with $C/L = 0.2$ (solid line) and $C/L = 0.8$ (dashed line) [From Jolliffe, 2003].	16
Figure 6.	U.S. Air Force Risk Management Goals [From AFPAM 90-902].	20
Figure 7.	Examples of decision input visualizations from (a) medium risk tolerant user and (b) from low risk tolerant user. Adapted from the Weather Risk Analysis and Portrayal Phase II Prototype Risk Tolerance Chart. [From Next Century, 2007].	23
Figure 8.	PDF representation of the risk of exceeding marginal and unfavorable thresholds using a $RT = 40\%$. a) Green decision input. b) Red decision input. c) Amber decision input. [After Eckel, 2006].	24
Figure 9.	Three part process of generating products communicating flight level turbulence risk.	28
Figure 10.	Three pressure level computation method. Red line indicates location of pressure level computation. Circled points of data used in calculation of TI at square point. Note Δy is not shown, but perpendicular to the page.	31
Figure 11.	Skew-T Log-P diagram showing 50 mb increment pressure level data locations (red text and lines) and interpolated levels (blue text and lines). For reference, standard atmospheric heights are displayed (green text).	33
Figure 12.	Piecewise cubic spline (black) and piecewise linear segments (red) [From Weisstein, 2008]	33
Figure 13.	Interpolated TI data using “linear” and “spline” methods versus the original TI data, Latitude/Longitude: a) 35°N/245°W b) 36°N/243°W c) 33°N/239°W.	35
Figure 14.	Example of raw TI output on a global grid. FH: 024 = Forecast Hour 24 , FL350 (MSL) = Flight Level 35,000 feet (Mean Sea Level)	36
Figure 15.	Uniform Ranks Method: calculating turbulence FP > MDT (TI>5) using 10-member ensemble.	37
Figure 16.	Example of uncalibrated FP output for EUCOM grid.	39
Figure 17.	FH 24 of the January 30, 2008, 00Z run; valid January 31, 2008, 00Z.	40

Figure 18.	Example comparison of a true and an ensemble forecast PDF (upper plot) and CDF (lower plot). An error of -13.9% in forecast probability for the chance of temperature $\leq 0^{\circ}\text{C}$ is the difference in the PDF's shaded areas, or the difference in the two CDFs indicated by the double arrow [From Eckel, 2008].	42
Figure 19.	Sample standardized error distributions for FH 84.	43
Figure 20.	Error (%) in FP for a) 10 and b) 1000 random errors in the ensemble mean and spread for FH 84.	47
Figure 21.	Histogram of 1000 sample FP errors, with fitted beta distribution, resulting from the ensemble PDF error distributions in Figure 19 for FP of (a) 25.0%, (b) 75.0%	48
Figure 22.	Uncalibrated 5 th , 50 th , 95 th percentiles of the beta PDFs of possible FP error for FH 84.	49
Figure 23.	Ambiguity and Calibration Curve for FH 84, 90% CI: The FP calibration (thick line) is simply the opposite of the 50 th percentile of the ensemble error distribution (beta PDF). The upper and lower bounds of ambiguity (thin lines) are the 90% confidence interval of the true forecast probability, found by taking the difference of the 5 th and 95 th percentiles from the 50 th percentile. Displayed is calibration factor for FP=10%, upper and lower ambiguity bounds for calibrated FP (16%)[After Eckel, 2008].	49
Figure 24.	Ambiguity Curves for FH 84, CIs a) 70% b) 80% c) 90% d)95% e) 99%.	52
Figure 25.	Risk Tolerance Uncertainty Curve.	53
Figure 26.	(a-c) Calibrated FP and user RT (center of error bar) and associated confidence intervals showing different DIs	54
Figure 27.	Prototype GUI.	55
Figure 28.	Prototype GUI: Drop down menu of a) generic risk tolerances b) mission limiting turbulence intensities.	57
Figure 29.	Prototype GUI: a) Drop down menu of forecast hours b) Drop down menu of flight levels	58
Figure 30.	EUCOM Mission Success Chance (FP calibrated) due to MDT turbulence for FH 24 of the January 30, 2008, 00Z run; valid January 31, 2008, 00Z	59
Figure 31.	EUCOM Worst Case Potential FP Error of MDT turbulence for FH 24 of the January 30, 2008, 00Z run; valid January 31, 2008, 00Z	59
Figure 32.	EUCOM DI for RT: 0.3 of MDT turbulence for FH 24 of the January 30, 2008, 00Z run; valid January 31, 2008, 00Z.	60
Figure 33.	a-o) Calibration and Ambiguity Curves through all forecast lead times, CI: 90%	63
Figure 34.	Deterministic MDT turbulence forecast using control member for the fall run, FH: 84, FL: 350	66
Figure 35.	Fall run, FH 06, FL 350 MSL. a) Deterministic MDT turbulence forecast b) Stochastic MDT turbulence forecast using FP	70
Figure 36.	Fall run, FH 06, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.5.	71
Figure 37.	Fall run, FH 06, FL 350 MSL. Simulated IWEDA MDT turbulence DI	72

Figure 38.	Fall run, FH 06, FL 350 MSL. Simulated WRAP MDT turbulence DI with RT=0.5	73
Figure 39.	Fall run, FH 06, FL 350 MSL. Simulated WRAP MDT turbulence DI with RT=0.1	74
Figure 40.	Fall run, FH 06, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.1.....	74
Figure 41.	Fall run, FH 06, FL 350 MSL. Simulated WRAP MDT turbulence DI with RT=0.9	75
Figure 42.	Fall run, FH 06, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.9.....	75
Figure 43.	Fall run, FH 84, FL 350 MSL. a) Deterministic MDT turbulence forecast b) Stochastic MDT turbulence forecast using FP	76
Figure 44.	Fall run, FH 84, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.5.....	77
Figure 45.	Fall run, FH 84, FL 350 MSL. Simulated IWEDA MDT turbulence DI	78
Figure 46.	Fall run, FH 84, FL 350 MSL. Simulated WRAP MDT turbulence DI with RT=0.5	78
Figure 47.	Fall run, FH 84, FL 350 MSL. Simulated WRAP MDT turbulence DI with ..	79
Figure 48.	Fall run, FH 84, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.1.....	80
Figure 49.	Fall run, FH 84, FL 350 MSL. Simulated WRAP MDT turbulence DI with RT=0.9	81
Figure 50.	Fall run, FH 84, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.9.....	81
Figure 51.	Fall run, FH 06, FL 350 MSL. a) FP of MDT turbulence b) Maximum (Worst Case) Potential Error for FP of MDT turbulence	83
Figure 52.	Fall, FH 06, FL 350 MSL, RT=0.1. DI Bias: WRAP minus RSCH chart.....	84
Figure 53.	Fall run, FH 06, FL 350 MSL, RT=0.1 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI.....	85
Figure 54.	Fall, FH 06, FL 350 MSL, RT=0.5. DI Bias: WRAP minus RSCH chart.....	86
Figure 55.	Fall run, FH 06, FL 350 MSL, RT=0.5 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI.....	87
Figure 56.	Fall, FH 06, FL 350 MSL, RT=0.5. CI=0.95: a) RSCH DI b) DI Bias: WRAP minus RSCH chart; CI=0.99: c) RSCH DI d) DI Bias: WRAP minus RSCH chart.	88
Figure 57.	Fall, FH 06, FL 350 MSL, RT=0.9. DI Bias: WRAP minus RSCH chart.....	90
Figure 58.	Fall run, FH 84, FL 350 MSL. a) FP of MDT turbulence b) Maximum (Worst Case) Potential Error for FP of MDT turbulence	91
Figure 59.	Fall, FH 84, FL 350 MSL, RT=0.1. DI Bias: WRAP minus RSCH chart.....	92
Figure 60.	Fall run, FH 84, FL 350 MSL, RT=0.1 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI.....	92
Figure 61.	Fall, FH 84, FL 350 MSL, RT=0.5. DI Bias: WRAP minus RSCH chart.....	94
Figure 62.	Positive Definite TI PDF with high probability of high TI	94
Figure 63.	Positive Definite TI PDF with high probability of low TI.....	95
Figure 64.	Fall, FH 84, FL 350 MSL, RT=0.9. DI Bias: WRAP minus RSCH chart.....	96

Figure 65.	(a-e) RSCH DI for MDT turbulence using a range of CIs; fall run, FH 24, FL 350 (MSL), RT=0.5.....	103
Figure 66.	(a-e) RSCH DI for MDT turbulence using a range of CIs; fall run, FH 48, FL 350 (MSL), RT=0.5.....	104
Figure 67.	(a-e) RSCH DI for MDT turbulence using a range of CIs; fall run, FH 84, FL 350 (MSL), RT=0.5.....	105
Figure 68.	Winter run, FH 06, FL 350 MSL. a) Deterministic MDT turbulence forecast b) Stochastic MDT turbulence forecast using FP.....	107
Figure 69.	Winter run, FH 06, FL 350 MSL. a) Simulated IWEDA MDT turbulence DI b) Simulated WRAP MDT turbulence DI with RT=0.5.....	108
Figure 70.	Winter run, FH 06, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.5.....	109
Figure 71.	Winter run, FH 06, FL 350 MSL. a) Simulated IWEDA MDT turbulence DI b) Simulated WRAP MDT turbulence DI with RT=0.1.....	110
Figure 72.	Winter run, FH 06, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.1.....	111
Figure 73.	Winter run, FH 06, FL 350 MSL. a) Simulated IWEDA MDT turbulence DI b) Simulated WRAP MDT turbulence DI with RT=0.9.....	112
Figure 74.	Winter run, FH 06, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.9.....	113
Figure 75.	Winter run, FH 84, FL 350 MSL. a) Deterministic MDT turbulence forecast b) Stochastic MDT turbulence forecast using FP.....	114
Figure 76.	Winter run, FH 84, FL 350 MSL. a) Simulated IWEDA MDT turbulence DI b) Simulated WRAP MDT turbulence DI with RT=0.5.....	115
Figure 77.	Winter run, FH 84, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.5.....	116
Figure 78.	Winter run, FH 84, FL 350 MSL. a) Simulated IWEDA MDT turbulence DI b) Simulated WRAP MDT turbulence DI with RT=0.1.....	117
Figure 79.	Winter run, FH 84, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.1.....	118
Figure 80.	Winter run, FH 84, FL 350 MSL. a) Simulated IWEDA MDT turbulence DI b) Simulated WRAP MDT turbulence DI with RT=0.9.....	119
Figure 81.	Winter run, FH 84, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.9.....	120
Figure 82.	Winter run, FH 06, FL 350 MSL. a) FP of MDT turbulence b) Maximum (Worst Case) Potential Error for FP of MDT turbulence	121
Figure 83.	Winter run, FH 06, FL 350 MSL, RT=0.1 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI.....	122
Figure 84.	Winter run, FH 06, FL 350 MSL, RT=0.1. DI Bias: WRAP minus RSCH chart.....	123
Figure 85.	Winter run, FH 06, FL 350 MSL, RT=0.5 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI.....	124
Figure 86.	Winter run, FH 06, FL 350 MSL, RT=0.5. DI Bias: WRAP minus RSCH chart.....	125

Figure 87.	Winter run, FH 06, FL 350 MSL, RT=0.9 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI.....	126
Figure 88.	Winter run, FH 06, FL 350 MSL, RT=0.9. DI Bias: WRAP minus RSCH chart.....	127
Figure 89.	Winter run, FH 84, FL 350 MSL. a) FP of MDT turbulence b) Maximum (Worst Case) Potential Error for FP of MDT turbulence	128
Figure 90.	Winter run, FH 84, FL 350 MSL, RT=0.1 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI.....	129
Figure 91.	Winter run, FH 84, FL 350 MSL, RT=0.1. DI Bias: WRAP minus RSCH chart.....	130
Figure 92.	Winter run, FH 84, FL 350 MSL, RT=0.5 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI.....	131
Figure 93.	Winter run, FH 84, FL 350 MSL, RT=0.5. DI Bias: WRAP minus RSCH chart.....	132
Figure 94.	Winter run, FH 84, FL 350 MSL, RT=0.9 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI.....	133
Figure 95.	Winter run, FH 84, FL 350 MSL, RT=0.9. DI Bias: WRAP minus RSCH chart.....	134

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table 1.	Expense Matrix: Costs (C) and Losses (L) for different outcomes in C/L decision model [From Table 8.1, Jolliffe, 2003].	15
Table 2.	All Fractional Error Spread (FES) values for forecast hour (FH) 84.	45
Table 3.	Mean and standard deviation values for the EF spread fractional error distribution	62
Table 4.	IWEDA Simulation: TI Thresholds and corresponding color values, color assignments, and DI	67
Table 5.	WRAP Simulation: FP vs RT and corresponding color values, color assignments, and DI	67
Table 6.	DI Bias Colorbar Key. Note: WRAP and RSCH never differ by more than one DI category	68
Table 7.	Prototype Mission Decision Input Matrix	101

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AF	Air Force
AFWA	Air Force Weather Agency
AOR	Area Of Responsibility
C	Cost
C/L	Cost-Loss Ratio
CAT	Clear Air Turbulence (non-convective)
DI	Decision Input
DoD	Department of Defense
ECMWF	European Centre for Medium-range Weather Forecasts
EF	Ensemble Forecast(ing)
EPS	Ensemble Prediction System
EUCOM	European Command
F	False-alarm
FH	Forecast Hour
FL	Flight Level
FNMOCC	Fleet Numerical Meteorology and Oceanography Center
FP	Forecast Probability
GEFS	Global Ensemble Forecast System
H	Hit
INT	Intensity
IWEDA	Integrated Weather Effects Decision Aid
L	Loss
lat	latitude
LGT	Light (Intensity)
lon	longitude
MAT	Microsoft Access T....(MATLAB data file)
MDT	Moderate (Intensity)
MSL	Mean Sea Level
NCEP	National Centers for Environmental Prediction
NOGAPS	Navy's Operational Global Atmospheric Prediction System
NWP	Numerical Weather Prediction
OWS	Operational Weather Squadron
PDF	Probability Density Function
PE	Potential Error
RT	Risk Tolerance
SVR	Severe (Intensity)
TI	Turbulence Index
TURBC	Turbulence
UTC	Coordinated Universal Time
V	Value
VS	Value Score
VT	Valid Time

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I owe special acknowledgements to several people who help make my thesis success possible. First, I'm extremely grateful for my family and friends who have supported me with camaraderie, encouragement, and prayers throughout this daunting process. Despite the challenges, my advisor, Maj Tony Eckel, pushed me into learning new areas. His expertise and feedback laid the foundation for this thesis. I appreciate Professor Harr's and Professor Regnier's willingness to lend help and provide constructive feedback. I thank Bob Creasey for providing technical support and helpful advice for my voluminous data ingest and storage needs. Immense gratitude goes to Mary Jordan for her technical assistance and invaluable mentorship. Finally, I would be remiss to fail to thank Matt Sittel and Gordon Brooks from AFWA for providing vital data for the study's background and analysis.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Air Force Smart Operations 21 (AFSO21) initiated a movement to improve combat capability to include improve decision making. The uncertainty of weather forecasts contributes to mission risk. Combined with ORM principles, ensemble data can improve combat capability through better decision making (Eckel et al., 2007). The focus of this study is not to prove the value of ensemble data. Rather, the focus is to transform ensemble data into optimized decision input and communicate it an effective manner so it can be incorporated into the decision making process. ‘Optimized’ describes decision input that incorporates a complete picture of uncertainty by estimating the uncertainty of forecast probability (ambiguity) and risk tolerance information. Upper level turbulence data will serve as the example weather parameter.

The three main objectives are to: 1) Furnish a process to account for uncertainty in both the ensemble data and user risk tolerance for the decision input, 2) create an effective visualization to for varying levels of risk intolerance, mission thresholds for turbulence, flight levels, and forecast hours that is educational and practical for the user 3) demonstrate potential value added to the user when using the research derived decision input product through a comparison of current decision input products.

The thesis organized into three main chapters: Introduction (Chapter I), Background (Chapter II), Methodology (Chapter III), Results (Chapter IV), and Conclusions (Chapter V). Chapter III fulfills objective one through a description of a newly fashioned ambiguity calculation method, called “Taijitu” (Eckel, 2008). Chapter III also fulfills objective two by characterizing the process to create the visualizations and display it using a graphical using interface. Chapter IV fulfills objective three with two cases which compare the different decision input products using scenarios of different risk tolerances and forecast hours.

The main benefit to this study will be to demonstrate the capability of stochastic model data incorporation into the decision making process. It will serve as an example for further education and development of stochastic methods in the Air Force and Department of Defense to improve combat capability.

THIS PAGE INTENTIONALLY LEFT BLANK

II. BACKGROUND

A basic understanding of meteorology (with emphasis in numerical weather prediction), decision theory, statistics, and military weapon systems is important to understand the complex problem of applying ensemble forecast application products to DoD operations. In this chapter, background will be explicitly given on ensemble forecasting, weather risk management, and visualizations. It is assumed the reader will have some basic understanding of statistics and military weapon systems.

A. ATMOSPHERIC PREDICTION

1. Atmospheric Uncertainty

Prediction of the atmosphere is inherently uncertain. Prediction of a future state of a dynamical system, such as the atmosphere, is possible using the system's rules and current state (Eckel, 2007). From the atmospheric modeling perspective, the system's rules are the set of equations that govern atmospheric motion and the current state is the set of initial conditions that estimate the atmosphere's current state. Prediction is limited when knowledge of one or both is erred or incomplete, which results in exponentially growing forecast errors (Eckel, 2007). This uncertainty of prediction stems from the chaotic nature of the atmosphere. The mathematical definition of chaos is "a dynamical system that has a sensitive dependence on its initial conditions" (American Heritage Dictionary, 2004). Edward Lorenz (1963 a, b) described "dynamical chaos" as the time evolution of "a nonlinear, deterministic dynamical system, which is extremely sensitive to the initial conditions of the system" (Wilks, 2006). Error in this initial description of the atmosphere is the primary contributor to the lack of atmospheric predictability (Lorenz, 1963 a, b).

Initial conditions (ICs) are a set of values for the dependant variables that describe the state of atmosphere at certain point in time. There exists two sets of ICs. The first set of ICs describes the true state of the atmosphere. The second set of ICs is analyzed from observation of the atmosphere. The state variables measured include temperature, dew point, relative humidity, three dimensional wind speed components, and pressure. The second set of ICs are inherently erred because the atmosphere is never completely and accurately observed. The data network is incomplete with course temporal and spatial resolution and the observations are imprecise and/or erred. Hence, the observed/analyzed ICs will never be the same as the true ICs (Wilks, 2006).

Forecast error arises when the observed (incomplete/erred) ICs are fed into an imperfect numerical weather prediction (NWP) model. A NWP model is composed of a set of differential equations, with time as an independent variable, which govern the atmosphere's motion. It integrates these atmospheric equations using complex algorithms to simulate the atmospheric behavior over time, generating a forecast with varying degrees of accuracy (Eckel, 2007). The model is unable to simulate the true behavior of the atmosphere because of inexact numerical and computational methods, boundary conditions, and physics within the model. Errors in the ICs and model may be small (even undetectable), but over a typical forecast period, they lead to a cascade of error that makes the forecast diverge exponentially from the truth.

Predictions made by the deficient NWP model, which is initialized using the erred ICs, are reasonably accurate in the short-term, but the solutions eventually diverges from truth and the ability to forecast accurately is lost (Eckel, 2007). Kalnay (2003) states even though two ICs may be almost identical initially, an unstable dynamical system, such as the atmosphere, will cause the trajectories to drift apart (Figure 1a). Conversely, a stable dynamical system with “stationary or periodic orbits” will cause two ICs initially on different trajectories to converge together (Figure 1b). On the synoptic scale, the solutions may start at similar point and be close for two to three days, but further in time, the two solutions diverge (Kalnay, 2003).

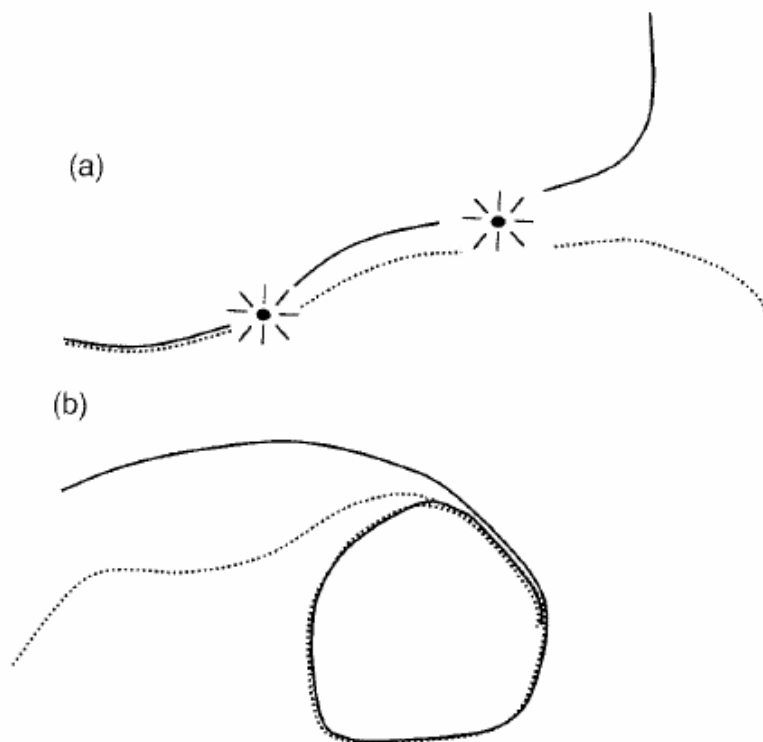


Figure 1. a) Dynamical System with instabilities b) Stable Dynamical System with stationary or periodic orbits [From Kalnay, 2003].

Lorenz proved this concept when he performed an “identical twin” experiment, where he compared two runs of the same model but with slightly different ICs. He found that the two runs’ solutions diverged dramatically and exhibited totally different trajectories (Kalnay, 2003). The initial round-off errors were amplified until they dominated the solution (Lorenz, 1993). Lorenz (1963 a, b) ascertained the atmosphere, similar to any dynamical system with instabilities, has a ‘finite limit of predictability’.

Atmospheric predictability is the degree to which a phenomenon can be foretold with an average accuracy greater than that of an unskilled reference forecast (Eckel, 2007). Generally, prediction decreases with increasing lead-time. When predictability is high, the theoretical limit for forecasting is two weeks (Kalnay, 2003). The long term average (climatology) can be viewed as a reference forecast, but like any other forecast, it’s predictability deteriorates with time (years to decades). In the near term, weather

forecasters focus on specific atmospheric phenomena that are generally anomalies from the average. It is then logical to use the climatology as a baseline for predictability. Errors become “saturated” (predictability is lost) when the error equals the average error of a climatology forecast (Eckel, 2007).

2. Methods for Prediction

The common approach to atmospheric prediction uses a deterministic, single solution. Deterministic forecasts have greatly improved with the development of NWP and since then, but the inherent uncertainty remains (Kalnay, 2003). A consistently perfect deterministic forecast is impossible to produce (Lorenz, 1963). Eric Eady was the first meteorologist to formally voice concern over the strict deterministic (one solution) method, in which the future is determined by the initial state without an account of error in the state (Lewis, 2005).

The graphics product in Figure 2 is a deterministic forecast of upper level turbulence (10,000 feet and above) for the region of interest applied in this research. The valid time (VT) of the graphics product is forecast hour 24 of case two. The term “Category II aircraft” refers to an instrument landing system capable of approaches to the runway ≥ 30 meters above ground and a Runway Visual Range (RVR) of ≥ 360 meters. The product implies complete confidence in the forecast. Although it offers a straightforward application, it neglects the importance and value of forecast uncertainty. Users may apply their own subjective measure of uncertainty to this product, which results in sub-optimal decisions (Wilks, 2006). Wilks (2006) asserts, “Deterministic forecasts of future atmospheric behavior will always be uncertain, and probabilistic methods will always be needed to adequately describe that behavior.”

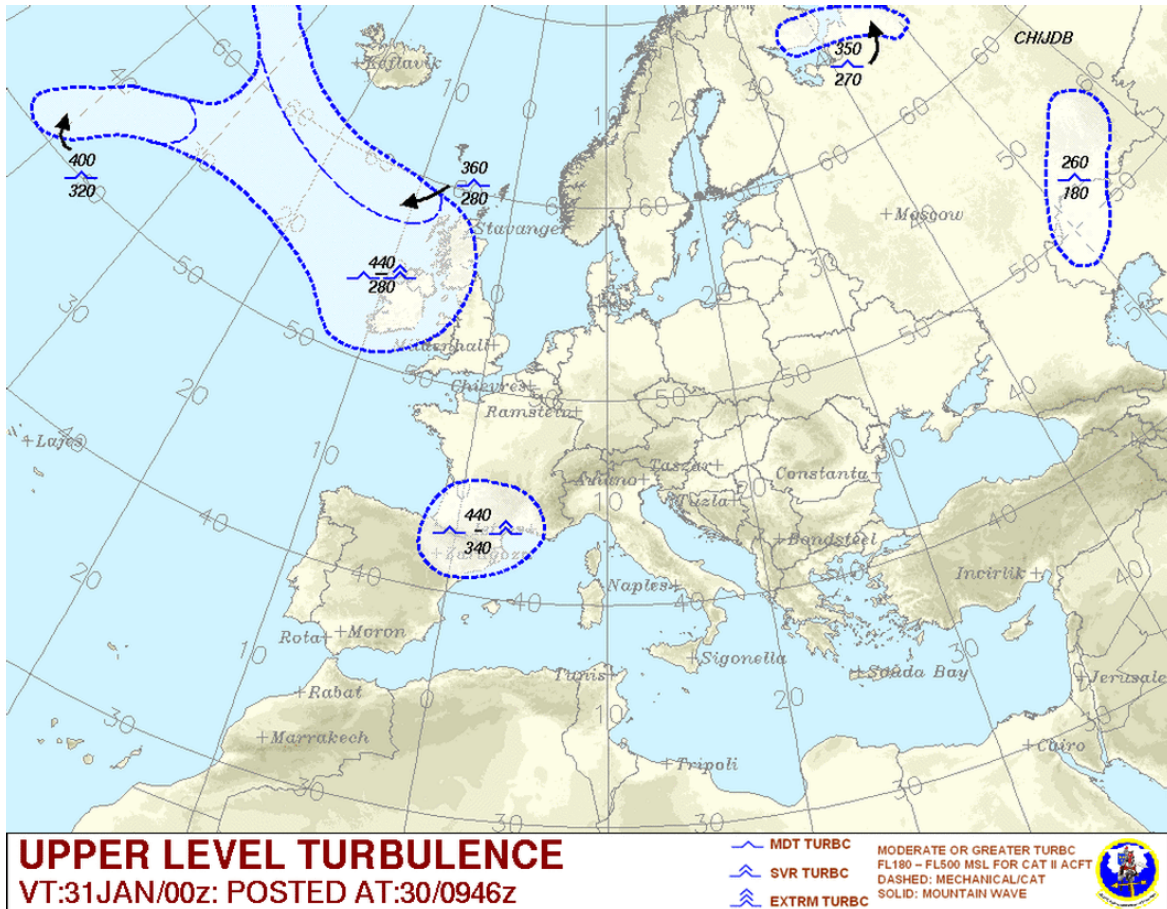


Figure 2. Current deterministic format for USAFE Upper Level Turbulence Chart; Intensities of turbulence (TURBC): MDT=moderate, SVR=severe, EXTRM=extreme; FL=Flight Level; CAT II ACFT=Category II aircraft [From USAFE OWS, 2007].

An alternative is the stochastic, multi-solution approach, which conveys the uncertainty of the forecast solution objectively. “No weather forecast is complete without a description of this uncertainty” (National Research Council, 2006). An objective description of the uncertainty (potential forecast error) is necessary to avoid unnecessary vulnerabilities, possible costly mistakes, and resource waste. Quantifying uncertainty can improve a weather forecast user’s ability to make more economically effective decisions based on their own utility function (Zhu et al. 2002). Stochastic forecasting does not aim to correct the forecast error but uses it to estimate error growth and describe the day-to-day, flow-dependant forecast uncertainty (Eckel, 2007).

B. ENSEMBLE FORECASTING (EF)

1. General Description

One method to produce a stochastic forecast is to run an ‘ensemble’ of model solutions. An ensemble forecast (EF) system uses “different but concurrent NWP model runs with perturbations to the ICs and the model to make a spectrum of possible solutions, or a distribution of likely future states of the atmosphere” (Eckel, 2007). A typical operational ensemble model runs 20 or more members (individual model runs), each time with a different modification (perturbations within the bounds of the suspect error) of ICs and/or the model itself to formulate different possible outcomes (Eckel, 2007). Ideally, each member has an equal chance of verifying. The greater the number of members in the ensemble gives a greater chance for the ensemble model to consistently capture the truth (Eckel, 2007). The aim of EF is to express the range of possibilities and define potential error in the deterministic forecast (Eckel, 2007).

The graphics product displayed in Figure 3, generated from the Joint Ensemble Forecast System (JEFS), is a typical output from an ensemble model. JEFS is a multi-year pilot project directed by the Air Force Weather Agency (AFWA) and the Fleet Numerical Meteorology and Oceanography Center (FNMOC) to prove the value and operational ability of EF for enhancing DoD operations and to investigate application to war-fighter decision making. This prototype graphics product displays the probability of turbulence occurrence and is one of the foundations of this research. The upper left hand image gives the ensemble consensus (mean of all the members) of geopotential heights and winds speeds. Using the ensemble data and a turbulence algorithm, JEFS produced probabilities for light, moderate, and severe turbulence thresholds in the following images. Higher probabilities are denoted by the warmer colors. Using decision science (discussed in section C) an educated user can objectively make decisions based upon the probabilities associated with each turbulence category.

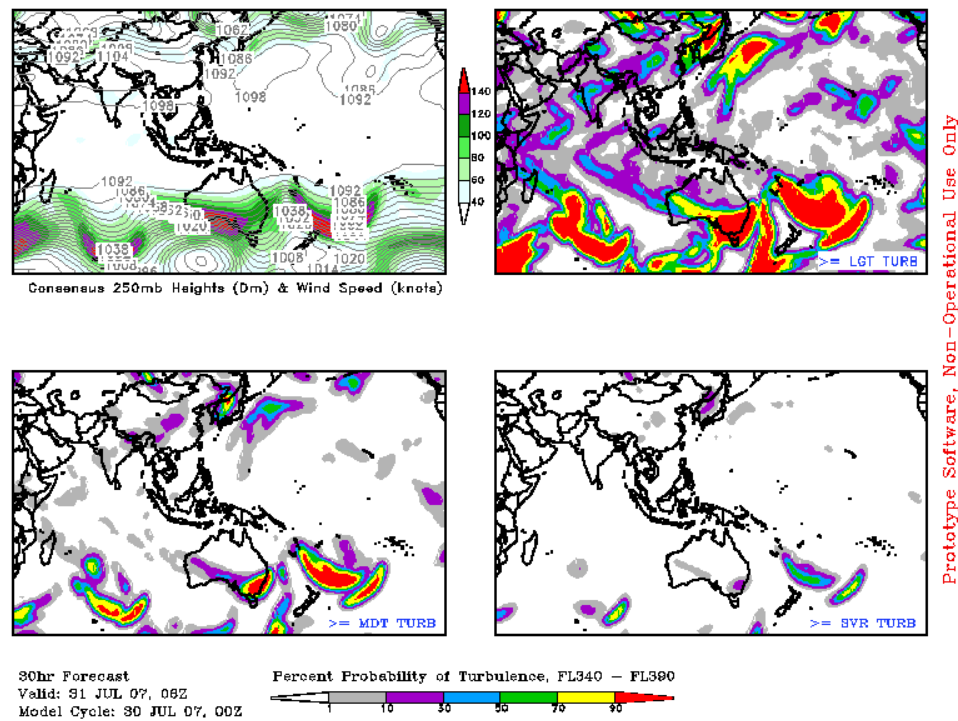


Figure 3. JEFS Non-Operational Turbulence Chart with varying probability to turbulence occurrence in southeast Asia [From JEFS, 2007].

2. Range of Possibilities

Eady emphasized consideration an ensemble of all possible atmospheric solutions, and these solutions must be expressed in terms of probabilities (Lewis, 2005). The atmosphere can be described as a range of possibilities in a probability density function (PDF), contrary to the singular deterministic solution (Eckel, 2007).

Like deterministic methods, stochastic methods employs analyses, forecasts, and climate data to ascertain the future state of the atmosphere, but do so using PDFs. The analysis PDF is a frequency distribution of potential analyses, all of which are random samples from the PDF trying to estimate the true PDF (Eckel, 2007). The size and shape of the analysis PDF is a description of the uncertainty in the analysis error. The forecast PDF works the same way, but it is a frequency distribution of potential forecast states, rather than potential analyses (Eckel, 2007). As a general principle, the more narrow the

PDF, the more certain the solutions and the more wide the PDF, the less certain the solutions (Eckel, 2007). Wider PDFs occur with the elapse of forecast lead time as the ensemble members spread to represent the growing error. As described in section 1, analysis and model errors contribute to the forecast error and inherent uncertainty.

Ensembles produce forecast probability for a particular event (i.e, turbulence $\geq$ MDT) that attempt to estimate the true probability from a hypothetical true PDF that is relative to the model system (Eckel, 2008). No one ‘true PDF’ exists for all models. A deterministic forecast is a guess at the unknown, true state of the atmosphere at a particular moment found by taking one sample from the forecast PDF. A perfect ensemble models all sources of uncertainty and takes infinite samples from the forecast PDF to produce the “true PDF” for that particular model. Real world ensemble models attempt to model many sources of uncertainty by taking limited samples from the forecast PDF. The goal of ensemble forecasting (EF) is to produce a forecast PDF of all possible future states of the atmosphere from which the true state is consistently a random sample (Eckel, 2007).

3. Ambiguity

Ambiguity is defined as the uncertainty of the uncertainty prediction, which can be conceptualized as errors bars about a probability forecast. Alternatively, ambiguity can be considered as the potential error in the ensemble’s forecast PDF (Eckel, 2008). Ambiguity arises when any contribution to forecast uncertainty is not sufficiently accounted for in the ensemble.

Several factors contribute to the ambiguity of ensemble data, the first of which is limited sampling. Due to computational constraints, only so many members can be processed at one time. Choosing only a few random samples from a forecast PDF fails to consistently reflect the breadth of uncertainty within the true PDF, therefore creating ambiguity (Eckel, 2007). Another contribution to ambiguity is a coarse resolution model. The smallest atmospheric motion wavelength that a model can resolve is seven to eight

grid points in length (Kalnay, 2003). The ensemble cannot quantify the uncertainty from atmospheric motions not represented in the model, hence creating ambiguity (Eckel, 2007).

Additionally, poorly simulated aspects of the IC and model uncertainty by the ensemble add to the ambiguity. A well-designed EF system attempts to models all possible sources of uncertainty (Eckel, 2007). The desire is to have at least different 20 members (perturbations) in order to ensure the ensemble disperses enough to encompass the truth and sufficiently account for analysis and model design uncertainty. The term ‘encompass the truth’ describes the ability of the ensemble to bound the verification value with the span of the members (Eckel, 2007). Three methods exist to create model perturbations: 1) IC/boundary condition perturbations 2) ensemble members that have different numerical schemes from the model 3) a combination of methods one and two (Cunningham, 2006). Using only one of first two methods contributes to ambiguity. For example, if an ensemble fails to account for the uncertainty within the cumulus scheme by only perturbing IC/boundary conditions, then the members will evolve cumulus clouds in the same manner (Eckel, 2007). The members remain too similar to one another as time elapses. The insufficient internal model perturbations lead to underdispersion and a false indication of certainty. The unsimulated variability of dispersion over time and space creates ambiguity.

Other possible contributors of ambiguity include post-processed diagnostics and decision threshold determination, all of which will be discussed in Chapter III. The challenge is to estimate this ambiguity and incorporate it into the decision making process so it does not hinder the utility of the ensemble data (Eckel, 2007).

As described previous section, the ensemble mean and spread (μ_e, σ_e) are estimates of the true PDF mean and spread (μ_T, σ_T) . In this research, the term “spread” is used interchangeably with “ensemble standard deviation”. Finite sampling hampers the ability of an ideal ensemble to represent the true PDF. The random nature of sampling causes the error to be variable and non-systematic, generating ambiguity. The variability

of μ_e and σ_e about the true values (potential error) is described using sampling distributions which are standardized with respect to a normal PDF where $\mu = 0$ and $\sigma = 1$ (Eckel, 2008).

$$\mu_e' = \frac{\mu_e - \mu_T}{\sigma_T} \quad (1)$$

$$\sigma_e' = \frac{\sigma_e}{\sigma_T} \quad (2)$$

Hence, μ_e' and σ_e' are standardized potential errors in the EF mean and variance respectively. Standardizing the error allows the error to be proportional across the subsets of data, making for easier analyzation and implementation. Figure 4 shows the sampling distributions from the ensemble mean (a) and spread (b).

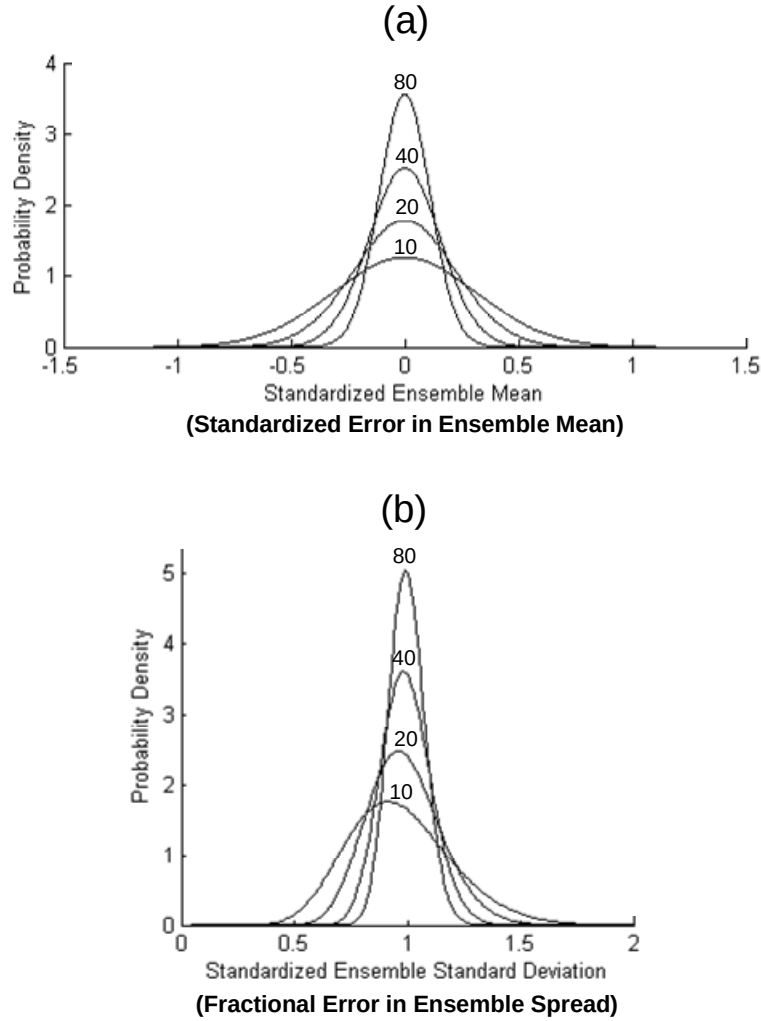


Figure 4. Sampling distributions of the (a) ensemble mean and (b) ensemble spread (i.e., standard deviation), labeled by the number of ensemble members [From Eckel, 2008].

J. L. Devore determined the μ_e distribution is Gaussian, has no bias with a standard deviation equaling $\sigma/\sqrt{n}$, and is independent of the (sampled) true PDF (Eckel, 2008). The σ_e distribution is gamma with a spread dependent on shape of the sampled PDF (Eckel, 2008). The minimum error variance is linked to finite sampling (ensemble size), but the error variance is likely to increase due to model deficiencies (Eckel, 2008).

C. WEATHER RISK MANAGEMENT

1. Value and Cost-Loss Analysis

a. *Measure of Value*

Weather forecasts give value to the user's decision-making process when the forecasts give the user the ability to take some action that reduces the average long-term expense due to an adverse weather event. Stochastic forecasts are difficult to apply since people often do not think in terms of probability (National Research Council, 2006). Since most decisions to take action are binary decisions (either "go/no-go" or "protect/don't protect), the forecast probability must be converted to a binary decision. Such a decision can be arrived at by choosing a "probability decision threshold" or just decision threshold for short. For example, an aggressive decision maker may use a decision threshold of 80% so that he will only decide the mission is a no-go when the risk of hazardous weather is greater than 80%. A key concept within this research is that there is an optimal decision threshold, which is a particular to a user.

The value score (V) helps illustrates the concept of the optimum decision threshold by measuring the utility of the forecast in the context of the cost-loss decision model (Wilks, 2006). To explain, each user has a unique sensitivity to an event -- an adverse weather occurrence defined by a specific threshold (e.g. 50 knots of wind or greater). Cost (C) is the amount of resources it takes to take protective action for the event, while loss (L) is the amount of resources lost when the event occurs without protection. Cost and loss are not necessarily limited to monetary value but may include other resources such as people, morale, and mission priority. As discussed previously, most decisions are binary so there are four possible outcomes (Table 1). A 'yes' forecast that verifies is hit (a). A 'yes' forecast that does not verify is a false alarm (b). A 'no' forecast that does not verify is a miss (c). A 'no' forecast that does verify is a correct rejection (d). After many forecasts have been complete, two quantities can be calculated: the hit rate (H) and false alarm rate (F), (Jolliffe, 2003):

$$H = \frac{a}{a+c} \quad (3)$$

$$F = \frac{b}{b+d} \quad (4)$$

Applying a cost-loss decision model, a decision maker can either take protective action or do nothing at all with the weather forecast, incurring an associated expense. Hence, from these four possible outcomes come four possible expenses, which are summarized in Table 1.

ACTION TAKEN	EVENT OCCURS	
	Yes	No
Yes	C (a)	C (b)
No	L (c)	Zero Expense (d)

Table 1. Expense Matrix: Costs (C) and Losses (L) for different outcomes in C/L decision model [From Table 8.1, Jolliffe, 2003].

The value score (V) of the forecast is a function of the quality of the forecast system (H and F), the user's C/L (α), and the climatology rate (s). The climatology rate is simply the average rate of occurrence of the weather event derived from past records of the event, independent of forecaster input. Jolliffe (2003) gives the value score equation:

$$V = \frac{\min(\alpha, s) - F(1-s)\alpha + Hs(1-\alpha) - s}{\min(\alpha, s) - s\alpha} \quad (5)$$

A value of $V = 1$ signifies a perfect forecast, while $V < 0$ signifies that climatology forecasts give better decision input. The highest possible V for a particular user is achieved when the decision threshold is equal to the C/L (Figure 5), thus indicating the optimal decision threshold.

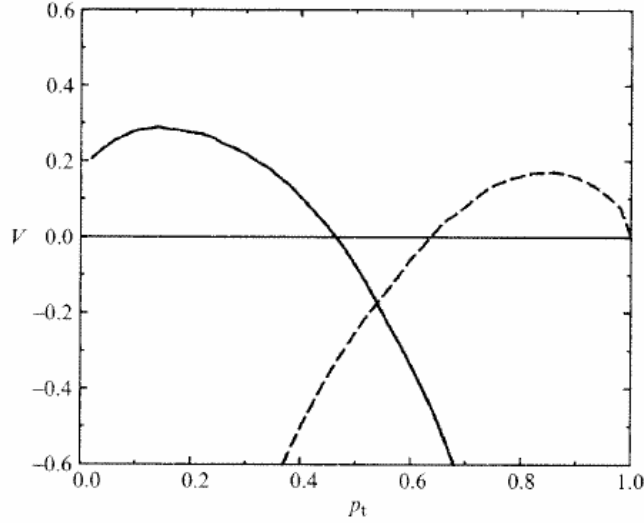


Figure 5. Variation of value V with decision threshold (p_t) for ECMWF EPS probability forecasts for users with $C/L = 0.2$ (solid line) and $C/L = 0.8$ (dashed line) [From Jolliffe, 2003].

To demonstrate this concept, consider the following example. Two different users have the same expenses for the event of surface winds greater than 50 knots: $C = \$150K$, $L = \$1,000K$; hence the user's $C/L=0.15$. User A is hesitant and takes action to protect anytime the FP is greater than 0%. User B is daring and will only take action when the FP is 100%. Both users are assumed to know nothing about decision science. The forecasting system is perfectly reliable, so the number of event occurrences on average (observed rate of the event) is equal to the product of the forecast probability (FP) and number of forecast cases. The following expenses (E) are calculated for three different sets of 100 forecast cases where ≥ 50 knots of wind was possible.

FP for ≥ 50 knots = .20

User A (always protect):

$$E = (100 \text{ cases}) * (\$150\text{K cost}) \\ = \$15,000\text{K}$$

→ WINNER

User B (never protect):

$$E = 0.20 * (100 \text{ cases}) * (\$1,000\text{K loss}) \\ = \$20,000\text{K}$$

FP for ≥ 50 knots = .10

User A (always protect):

$$E = (100 \text{ cases}) * (\$150\text{K cost}) \\ = \$15,000\text{K}$$

User B (never protect):

$$E = 0.10 * (100 \text{ cases}) * (\$1,000\text{K loss}) \\ = \$10,000\text{K}$$

→ WINNER

FP for ≥ 50 knots = .15

User A (always protect):

$$E = (100 \text{ cases}) * (\$150\text{K cost}) \\ = \$15,000\text{K}$$

User B (never protect):

$$E = 0.15 * (100 \text{ cases}) * (\$1,000\text{K loss}) \\ = \$15,000\text{K}$$

←TIE→

In this last scenario, the total expense is equal for both users when the $FP = C/L = 0.15$. User B expense is less than user A expense only when $FP < C/L$. Therefore, an astute user would act like User B (take no protective action) when $FP < C/L$ and would act like User A (take protective action) when $FP > C/L$. Over the long term, a user can minimize expense and yield the greatest value by noting that a decision threshold $= C/L$ is the optimal decision threshold.

Realistically, some users may find it difficult to define their C/L when their expenses are too complex or difficult to quantify. A user can still attain value from the forecast by using a decision threshold within the “ballpark” of their actual C/L ratio. As seen in Figure 5, a decision threshold approximately between 0.02 and 0.45 still yields value for the $C/L=0.2$ user. This concept is important to the integration of stochastic

forecasting into DoD operations. Users can still attain value from an ensemble forecast without having to invest resources for an in-depth analysis to determine their actual C/L ratio.

It is important to note that no single decision threshold is optimal for a range of users with different measures of risk. For example, if the $C/L=0.8$ user, who has a high tolerance to risk, acts on $C/L=0.2$ user's optimal decision threshold, he will attain negative value from the forecast and miss opportunities to act by protecting unnecessarily. The optimum decision threshold = C/L concept only applies to properly calibrated forecast probabilities; the user cannot simply optimize his/her decision threshold otherwise.

When the FP exceeds the decision threshold, protective action should be taken. In this way, the probability forecast is converted to a binary decision input (Jolliffe, 2003).

b. Risk Tolerance and Cost-Loss Ratio

A cost-loss analysis (C/L ratio) is a way to quantify the user's tolerance to Risk, or 'risk tolerance'. The risk tolerance is a measure of the amount the risk the user is willing to accept. For example, a user with a small C/L has relatively large potential losses and benefits by taking action at relatively low forecast probabilities. Users with small C/L are highly sensitive to adverse weather events; hence, their risk tolerance (RT) is lower. Conversely, a user with a high C/L needs a higher probability to justify taking protective action for an event since the cost of protection is not much less than the potential loss. Users with high C/L are more tolerant to adverse weather events; hence, their RT is higher. In this research, the term 'RT threshold' is interchangeable with the term 'decision threshold'. The term 'mission threshold' describes those weather event criteria that restrict mission accomplishment and may pose a threat to safety. The best approach is to combine knowledge of mission thresholds and RT with a skilled stochastic forecast to give an optimum decision input that minimizes loss and/or maximizes gains (Eckel, 2007).

2. Operational Risk Management

In addition to economic benefit, military decision makers are also concerned with risks such as loss of life and benefits such as mission success (Cunningham, 2006). The ability of stochastic forecasting to quantify uncertainty serves to strengthen the operational Risk Management (ORM) system. The foundation of ORM is to accept risk when the potential benefits are greater than the cost and avoid risk when potential losses are unaffordable in order to maximize gain/minimize loss in the long run (AFI 90-901; Zhu et al. 2002). Two types of ORM exist: offensive ORM acts to maximize gains (benefits) and defensive ORM acts to minimize loss (risk). From the cost-loss analysis perspective, the C and L are defined in terms of benefits and risks respectively. In order for the concept of stochastic forecasting to be accepted and implemented by the DoD, then it has to be part of the ORM decision-making process (Eckel et al, 2008).

The Air Force (AF) defined risk management guidelines in AF Instruction 90-901 and elaborated the guideline in AF Pamphlet 90-902. “The USAF aim is to increase mission success while reducing the risk to personnel and resources to the lowest practical level in both on- and off-duty environments” (AFPAM 90-902). ORM has become a mode of thinking in the AF with yearly training, safety days, and in everyday mission checklists (Cunningham, 2006). The entire set of AF risk management goals are listed in Figure 6. Along with these goals, are also 4 guidelines 1) accept no unnecessary risk, 2) make decisions at the appropriate level, 3) accept risk when benefits outweigh the cost, 4) integrate ORM into Air Force doctrine at all levels (AFI 90-901). Other DoD service components follow similar ORM principles.



Figure 6. U.S. Air Force Risk Management Goals [From AFPAM 90-902].

The ability of a stochastic forecast to quantify uncertainty is key to giving the DoD decision-maker a complete understanding of the weather situation. Deterministic forecasts omit uncertainty, while stochastic forecasting objectively measures uncertainty. This objective measure allows the user to apply their own set of risk tolerances to make an optimized, unbiased decision to minimize risk and/or maximize benefits. Ensemble-based probabilistic forecasts have a place in DoD operations by incorporating vital information into the decision-making process.

D. CALCULATING AND COMMUNICATING FORECAST TURBULENCE AND ITS UNCERTAINTY

1. Today's Deterministic Methods

High altitude clear-air turbulence (CAT) can be an extremely detrimental weather event for aviation. The number of fatal accidents involving CAT encounters are relatively low (McLean, 1986), but serious injury and structural damage have occurred with severe (SVR) to extreme CAT. Rerouting and loitering due to CAT and flying through CAT increases fuel expense and mission times (Ellrod and Knapp, 1992).

Significant CAT is prevalent in areas of vertical wind shear, horizontal shear, convergence, horizontal deformation, lapse rate discontinuities, and strong horizontal thermal gradients (Ellrod and Knapp, 1992). NWP models cannot practically predict turbulence because the resolution required for aircraft scale turbulence remains too high. Diagnostics and rules of thumb (ROT) have been developed to produce turbulence forecasts by correlating synoptic and mesoscale patterns to CAT occurrences (Cunningham, 2006). The current mainstream practice of forecasting turbulence is largely deterministic. Operational forecasters use a combination of tools like raw model data, model-derived turbulence diagnostics, and forecast ROT with their skill and experience to produce turbulence forecasts like the one illustrated in Figure 2, which is an example of the most basic decision input in which objective uncertainty is ignored.

The Integrated Weather Effects Decision Aid (IWEDA) is a baseline decision input tool that compares mission limiting weather criteria with deterministic forecasts to produce decision inputs. IWEDA has marginal and unfavorable thresholds established for different assets (resources such as people or equipment) depending on the assets' vulnerability to the weather parameter. IWEDA defines the mission threshold as "the point where the occurrence of a meteorological element causes a significant (moderate or severe) impact on a military operation, system, subsystem, or personnel" (Shirkey and Gouveia, 2002). IWEDA determined the mission thresholds by "doctrine, safety, or engineering factors (people, modeling, or testing)" (Shirkey and Gouveia, 2002).

Decision inputs are color coded. Red indicates conditions in which operations "are severely impacted: There is either a total or severe degradation or the operational limits or safety criteria have been exceeded" (Shirkey and Gouveia, 2002). Amber indicate conditions in which operations "are marginal and the operational capability is degraded, or there is a marginal degradation" (Shirkey and Gouveia, 2002). Personnel can still proceed with operations, but need to proceed with caution. Green indicates that no operational restrictions exist (Shirkey and Gouveia, 2002). To give an example of an unfavorable (red) mission threshold: personnel cannot safely perform a static line parachute jump in greater than 13 knots of surface wind. The marginal (amber) threshold is 9 knots.

IWEDA in a sense is giving a hint of uncertainty information with this marginal (amber) mission thresholds by inferring an average forecast error of 4 knots. IWEDA includes no objective uncertainty information. Other limitations of IWEDA include no consideration of the user's risk tolerance and complete ignores ambiguity.

2. WRAP

Methods to incorporate objective uncertainty into the decision making process are in development. One approach to produce and visualize stochastic weather-based decision input is the prototype Weather Risk Analysis and Portrayal (WRAP) tool. The Army Research Laboratory funded Next Century Corporation to develop the WRAP tool to address the need of military users to correctly and efficiently interpret ensemble forecasts (Next Century, 2007). WRAP processes ensemble data, users' RT, and IWEDA's mission thresholds to create a decision input which is superior to IWEDA. WRAP uses objective uncertainty information to formulate unique decision inputs that vary with different risk tolerances whereas IWEDA uses subjective uncertainty information to generate generic decision inputs. WRAP defines the user's RT as the level of risk the user is willing to accept (Next Century, 2007). WRAP uses the familiar stop-light (red, amber, green) format in their graphics product to easily convey the user-specific decision input. A user with a high risk tolerance is going to have different decision input that a user with a low risk tolerance (Figure 7).

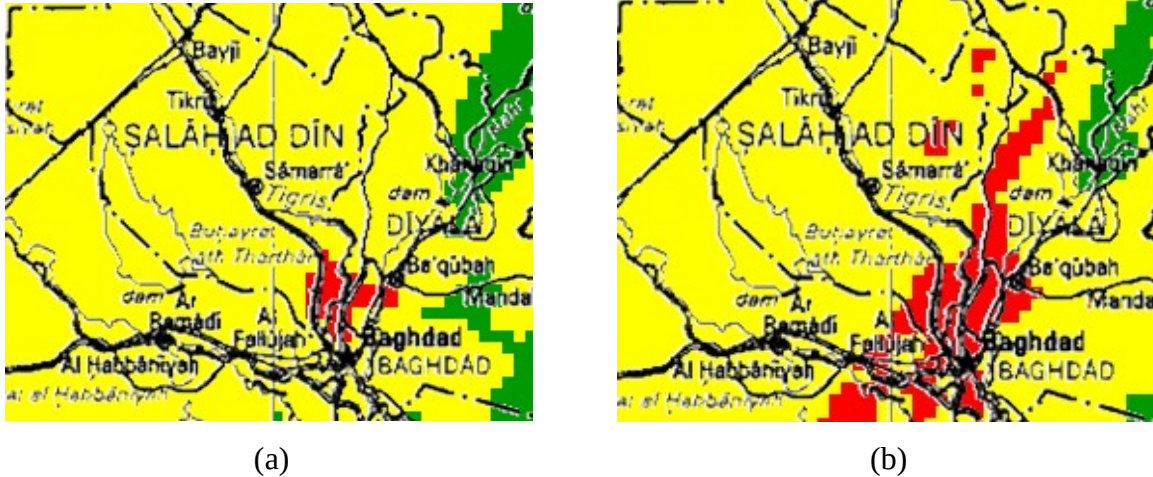
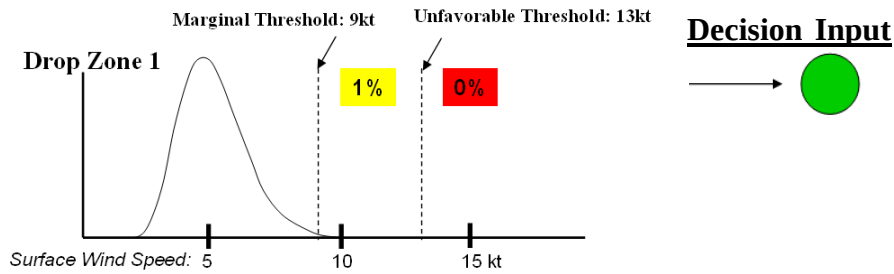


Figure 7. Examples of decision input visualizations from (a) medium risk tolerant user and (b) from low risk tolerant user. Adapted from the Weather Risk Analysis and Portrayal Phase II Prototype Risk Tolerance Chart. [From Next Century, 2007].

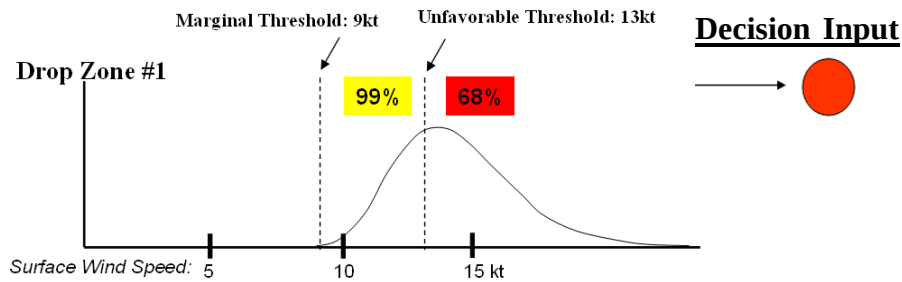
The example below illustrates WRAPs methodology for determining the red (no-go), amber (marginal), or green (go) colored decision inputs. From the previous IWEDA example, the weather parameter of concern is surface wind for a static line parachute jump into a drop zone. The user has a medium tolerance to risk with a RT of 40%. The PDFs displayed in Figure 8 represent the range of possible outcomes for surface wind speed as defined by an ensemble forecast. The risk of exceeding the marginal or unfavorable thresholds is the area under the PDF to the right of the threshold.

In Figure 8a, nearly all of the PDF lies below the marginal threshold. The risk of > marginal and > unfavorable thresholds is one and zero percent respectively. Since the RT exceeds the risk for either threshold, the user is given a “green” decision input. In Figure 8b, the majority of the PDF lies above the marginal threshold (99%). Since the risk of > unfavorable threshold (the worse condition) is 68% and exceeds the RT, the user is given a “red” decision input. In Figure 8c, the majority of the PDF lies above marginal threshold. The risk of > unfavorable conditions is 11 percent. The risk of > marginal threshold is 63 percent. Since this risk exceeds the RT, the user is given a “amber” decision input.

a)



b)



c)

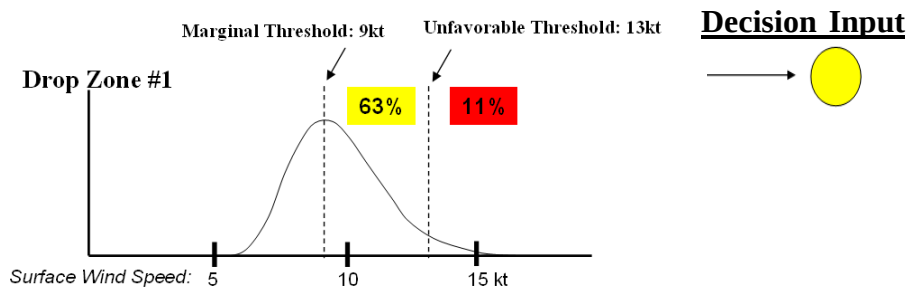


Figure 8. PDF representation of the risk of exceeding marginal and unfavorable thresholds using a RT = 40%. a) Green decision input. b) Red decision input. c) Amber decision input. [After Eckel, 2006].

WRAP is limited by its use of the deterministic IWEDA marginal thresholds in a stochastic context to derive ambiguity. Using the example from the static line parachute jump, WRAP is giving a hint of ambiguity information using two thresholds. The marginal threshold (> 9) infers there is an average forecast error of four knots for the unfavorable mission threshold (≥ 13 knots). WRAP makes no formal calculations to determine the ambiguity, rather it uses the marginal threshold as a proxy. The concept of

marginal threshold only makes sense in a deterministic context. The amber regions on stochastic weather-based decision input graphic should denote regions of “too close to call” not regions of “proceed with caution.” Only one threshold, the mission (unfavorable) threshold, should be applied to the stochastic context.

3. Conclusion

Today’s method to give decision input is deterministic with no uncertainty information or consideration of RT included. IWEDA is a more advanced deterministic tool with its application of marginal and mission thresholds to give decision inputs. WRAP, the prototype stochastic weather decision aid improves upon IWEDA, by incorporating both forecast probability uncertainty and RT but poorly estimates ambiguity and does not consider the uncertainty in RT. This research demonstrates the same capabilities of WRAP, but objectively estimates ambiguity and subjectively estimates RT uncertainty to incorporate the possibly vital information into the decision making process.

THIS PAGE INTENTIONALLY LEFT BLANK

III. METHODOLOGY

Using the previous background research as a foundation, this study demonstrates capabilities that will serve to further education, development, and utilization of stochastic methods in the Air Force and Department of Defense. Generating products communicating flight level turbulence risk to the user for optimal decision making requires a three part process, outlined in Figure 9. Specifically, part one generated non-calibrated forecast probabilities of varying levels of turbulence intensities using the uniform ranks method. Part two generated calibrated forecast probability, ambiguity (using the Taijitu method), and decision input data by comparing the risk and the risk tolerance. Part three introduced a user-friendly graphical user interface (GUI) to visualize the information from Part two.

The researcher used Matlab, version 7.0 or later, as the computer programming platform. The study's methods can be applied to other computer programming platforms as well, especially in the visualization realm. Matlab has limited graphics capability. A more sophisticated programming platform can produce a more advanced visualization to communicate flight level uncertainty.

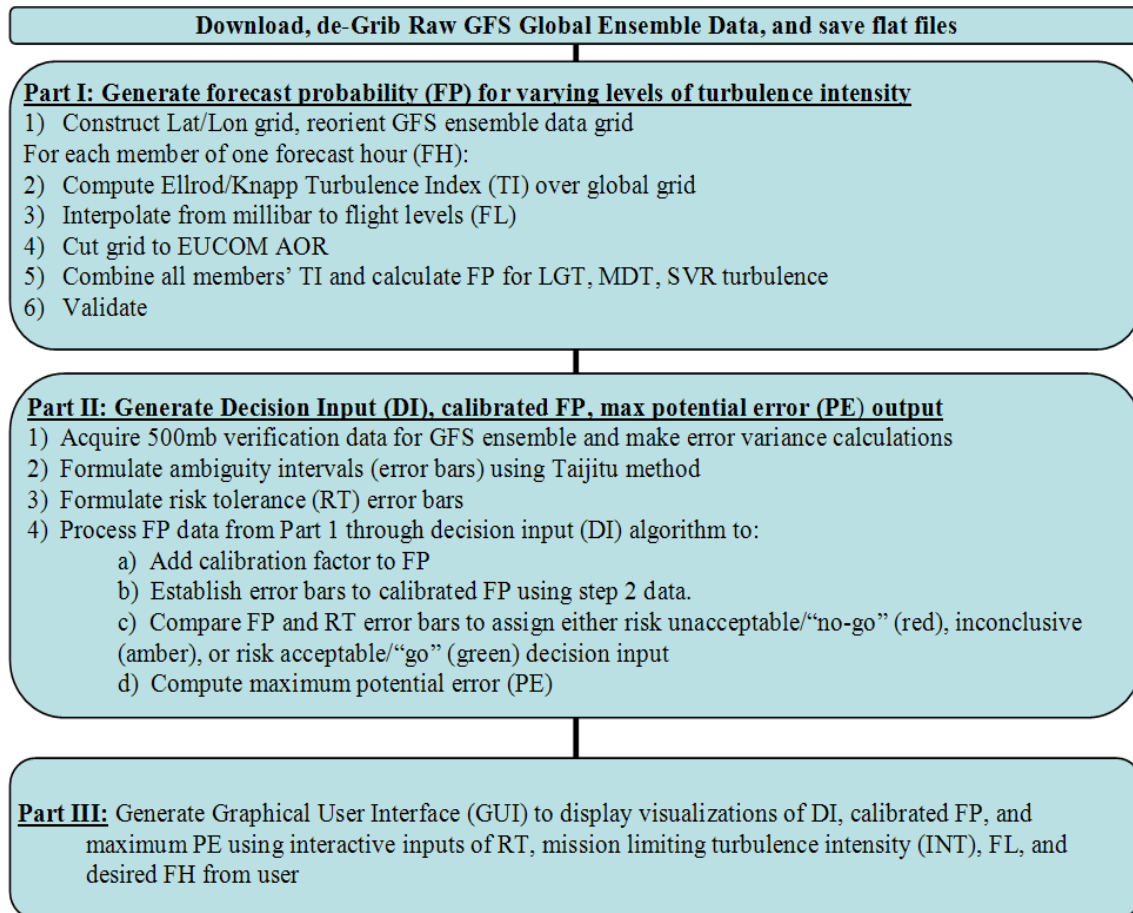


Figure 9. Three part process of generating products communicating flight level turbulence risk

A. PART ONE

1. Model Description and Ingest

The Global Ensemble Forecasting System (GEFS), used in this study, is a worldwide (360 x 181) model grid generated by the National Centers for Environmental Prediction (NCEP). The GEFS dataset has 21 members, one degree grid spacing, 50 millibar (mb) pressure increments, six hour forecast intervals, and a 384 hour forecast length. In this research, the maximum forecast length is 84 hours. The 50 mb pressure increment essential for calculating the turbulence index is only available up to forecast

hour (FH) 84. The GEFS uses an ensemble transform to create initial condition perturbations; no internal model perturbations are applied.

GEFS is available in gridded binary version two (GRIB2) format on the GEFS ftp website. For the purposes of the study, three variables (3 GRIB2 files) for each member were downloaded and “de-gribbed” into a usable format: u and v components of the wind (meters/second), and the geopotential heights (meters). The latitude, longitude, and variable spatial grids were reoriented so the coordinate 90°N and 0°E is positioned at the upper left hand corner of the grid. In order to explore risk analysis for different amounts of turbulent areas, the researcher downloaded two GEFS model runs: one taken from the fall season (02 OCT 07, 00Z) and one taken from the winter season (30 JAN 08, 00Z). A single sample forecast hour, flight level, forecast probability, and risk tolerance from the winter case will be used throughout this chapter to exhibit the methodology.

2. Turbulence Index (TI) Calculation

The Ellrod and Knapp (1992) turbulence index (TI) calculation is the diagnostic of choice in this study because of its skill and operational history with NCEP’s and AFWA’s model derived turbulence forecasts. The Ellrod TI was developed from concepts derived from 30 years of turbulence studies starting in the 1950s (Ellrod and Knapp, 1992). The physical foundation uses the concept of frontogenic intensity. Frontogenesis increases vertical wind shear (VWS) and therefore increases the threat of clear air turbulence (CAT) (Ellrod and Knapp, 1992). The product of VWS with the combination of horizontal deformation (DEF) and convergence (CVG) forms the basis the Ellrod TI equation (Ellrod and Knapp, 1992):

$$TI = VWS * (DEF + CVG) \quad (6)$$

The VWS term describes a rapid change in wind speed and/or direction with height (Ellrod and Knapp, 1992). The equation for VWS is:

$$VWS = \frac{(\Delta u^2 + \Delta v^2)^{1/2}}{\Delta z^2} \quad (7)$$

where Δu and Δv are the differences of wind speeds between grid points, and Δz is the difference in geopotential heights (layer thickness).

The DEF term is characterized by a “property of a fluid that transforms a circular-shaped area of fluid to an elliptical shape” (Ellrod and Knapp, 1992). The equation for DEF is:

$$DEF = (DST^2 + DSH^2)^{1/2} \quad (8)$$

where DST is the stretching deformation and DSH is the shearing deformation, given by the following equations:

$$DST = \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \quad (9)$$

$$DSH = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \quad (10)$$

Ellrod and Knapp (1992) describe CVG as “a compaction of a fluid caused by the confluence of streamlines and/or deceleration of air parcels.” CVG is the negative of divergence and is given by the following equation:

$$CVG = -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \quad (11)$$

AFWA uses the Ellrod index operationally in an algorithm to automatically compute turbulence forecasts from the Mesoscale Model (MM5) (G. Brooks, 2008, personal communication).

The researcher acquired and manipulated AFWA’s turbulence algorithm to construct the algorithm used in this research. The researcher used a three pressure level center differencing approach where Δz is computed at one pressure level (p2) using the data from the pressure level above (p3) and below (p1) and divided by the total depth (Figure 10). The TI for the top pressure level was computed using one side differencing.

The TI for the bottom pressure level (850 mb) was negated because it was of little consequence to the interpolation scheme, discussed in the next section.

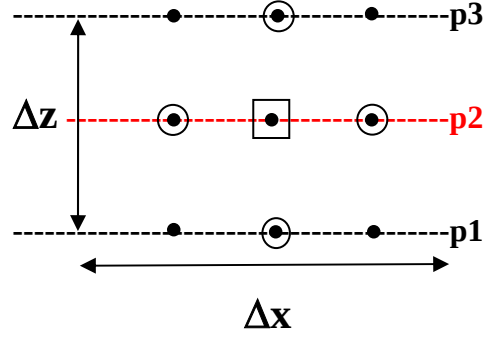


Figure 10. Three pressure level computation method. Red line indicates location of pressure level computation. Circled points of data used in calculation of TI at square point. Note Δy is not shown, but perpendicular to the page.

The horizontal distance between the grid points was computed for the one degree global grid using the equations:

$$\partial x = \left(\frac{2 * \Pi * R_e}{360} \right) * \cos(\text{degree of latitude}) \quad (12)$$

$$\partial y = \left(\frac{2 * \Pi * R_e}{360} \right) \quad (13)$$

where R_e is the average earth radius = 6,372,797 meters. The end TI result was multiplied by the scaling factor 1×10^7 (Ellrod and Knapp, 1992).

The Ellrod index thresholds for turbulence intensity are dependant on the model used. The study's TI thresholds were taken from the JEFs Joint Global Ensemble (JGE) upper level turbulence algorithm which is also applied to the GFS ensemble. The

intensity thresholds for light (LGT), moderate (MDT), and SVR (SVR) turbulence are 2, 5, 10 (unitless) respectively (G. Brooks, 2008, personal communication).

3. Interpolation

After the TI was calculated, the researcher interpolated from pressure level data to constant flight levels. Interpolation is a necessary step to convert the turbulence data to standard flight levels so the data can easily be applied to DoD operations. The interpolated heights were chosen with consideration of the logarithmic pressure spacing of the atmosphere. Since the GEFS data levels are incremented every 50 millibars, the researcher chose a logical height level scheme to minimize over interpolation and still yield useful flight level products. The flight levels above 20 thousand feet are incremented every five thousand feet and flight levels between 10 and 20 thousand feet are incremented every two thousand feet (MSL) (Figure 11).

Matlab has routines to automatically interpolate data from one height to another. The Matlab one-dimensional interpolation routine (`interp1`) has several options to include 'linear' and 'spline.' The linear method, as it states, interpolates data linearly from the original level to the desired level. The spline method interpolates data using a cubic spline (Figure 12), which is constructed using third order piecewise polynomials (Weisstein, 2008).

To compare the performance of the methods, the researcher took a group of cross sections lying within a strong TI gradient in WRN CONUS from the same model day, time, and forecast hour pictured in Figure 14. The two methods were interpolated to the research height levels (Figure 13). Interpolated TI results for both methods were similar below 25,000 feet. Larger differences occur above 25,000 feet because of larger distances exist between the original data points and the interpolate points due to logarithmic spacing. Since the upper level winds generally do not decrease and increase with height linearly, the spline method interpolated the TI in between the large gaps (separating the original data points) more realistically. For this reason, the researcher implemented the spline method.

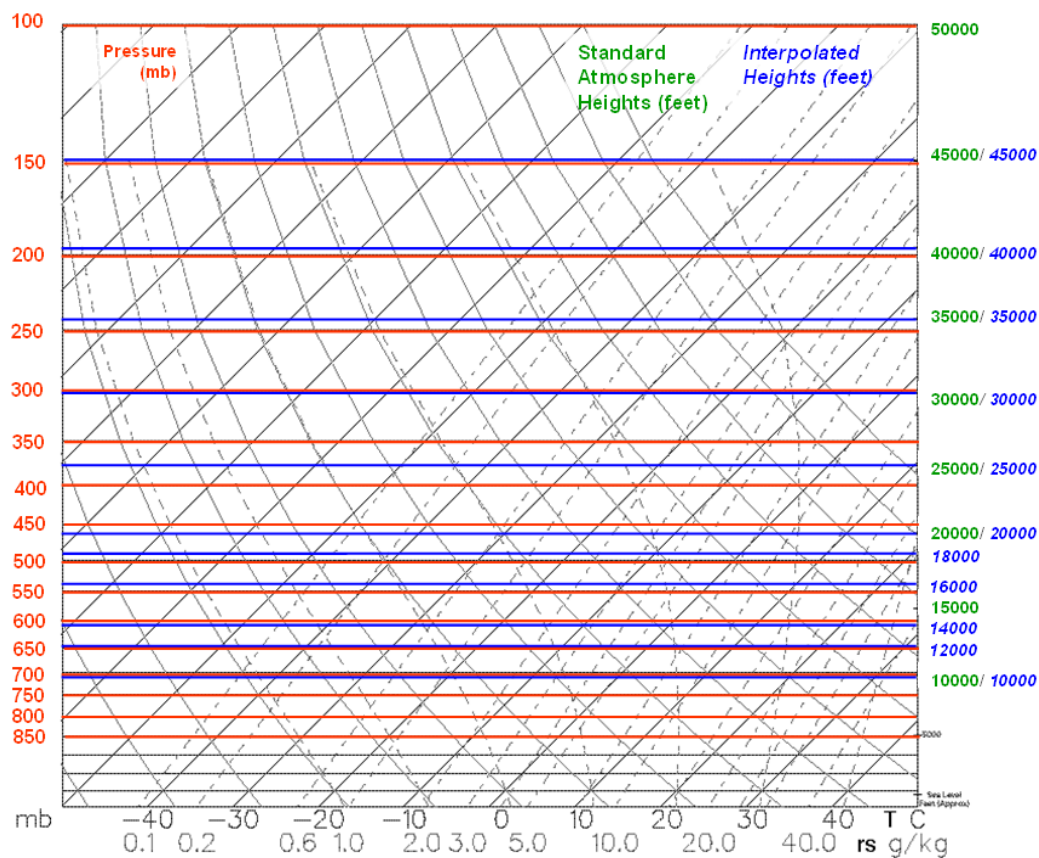


Figure 11. Skew-T Log-P diagram showing 50 mb increment pressure level data locations (red text and lines) and interpolated levels (blue text and lines). For reference, standard atmospheric heights are displayed (green text).

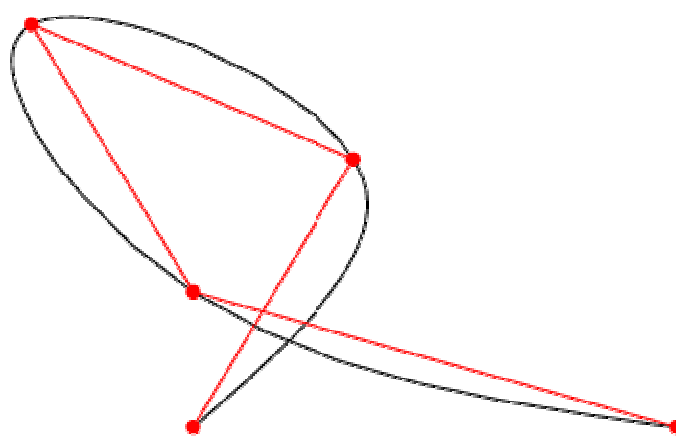


Figure 12. Piecewise cubic spline (black) and piecewise linear segments (red) [From Weisstein, 2008]

After the GEFS model ingest, TI calculation, and interpolation of height levels, the resultant output was TI (no units) data for all 21 members. Member number two from the January 31, 2008 00Z run is shown in Figure 14.

Because of the extremely large amount of data, the researcher cut the global grid ($181^{\circ} \times 360^{\circ}$) down to a specific area of responsibility (AOR), the European Command (EUCOM), which measures ($41^{\circ} \times 71^{\circ}$). The EUCOM AOR serves a primary location for refueling and transport missions supporting Middle East Operations, which are particular vulnerable to upper level turbulence.

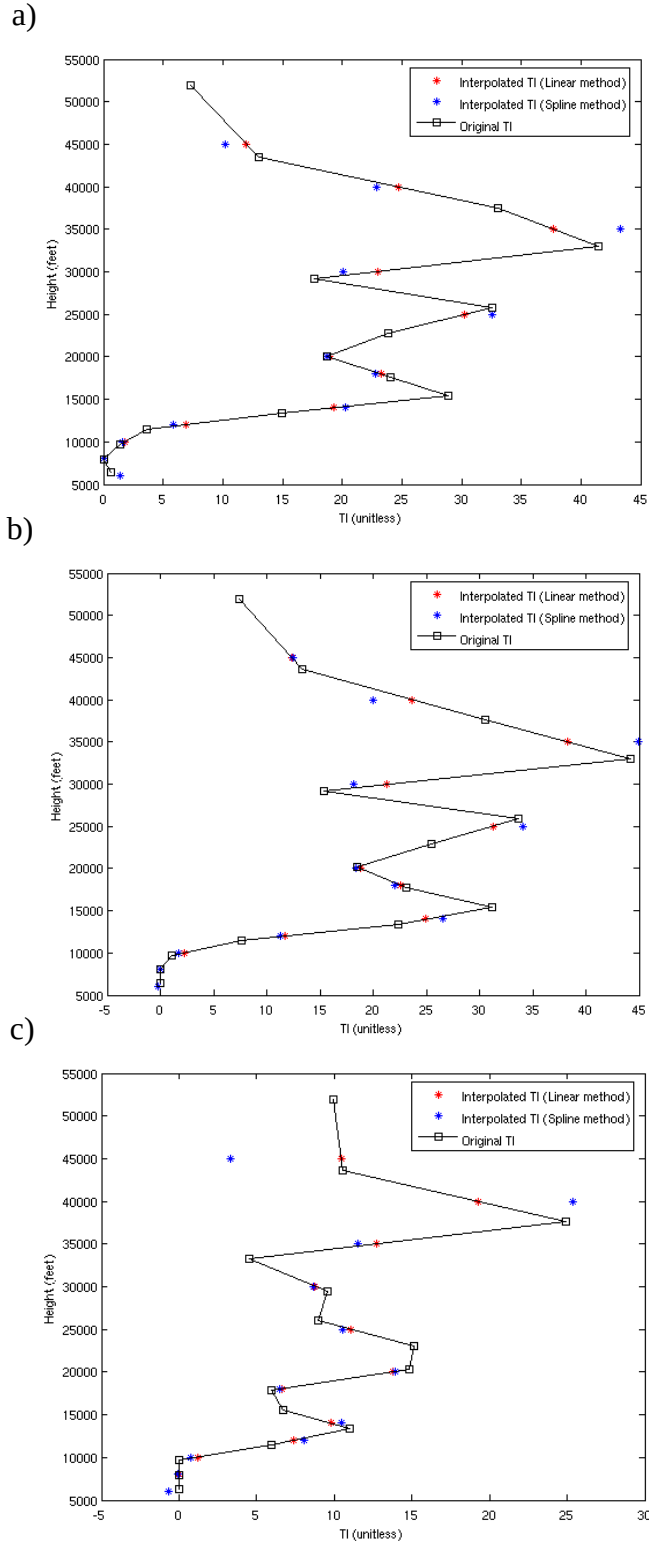


Figure 13. Interpolated TI data using “linear” and “spline” methods versus the original TI data, Latitude/Longitude: a) 35°N/245°W b) 36°N/243°W c) 33°N/239°W.

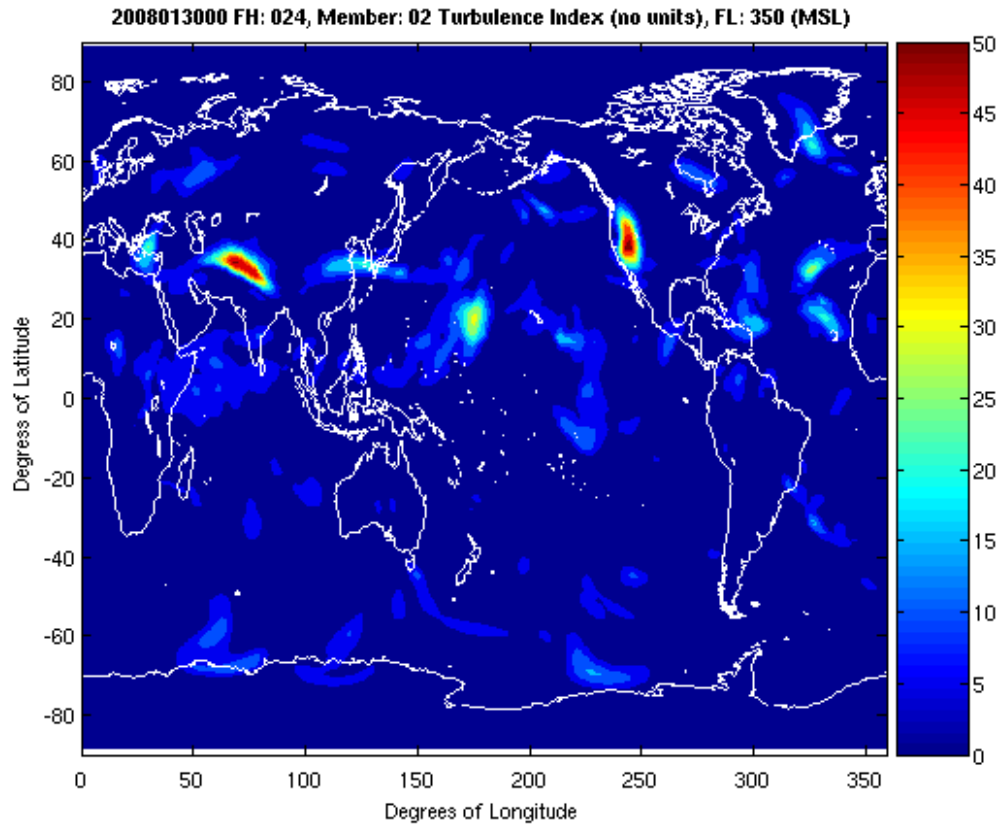


Figure 14. Example of raw TI output on a global grid. FH: 024 = Forecast Hour 24 , FL350 (MSL) = Flight Level 35,000 feet (Mean Sea Level)

4. Forecast Probability Calculations

To calculate forecast probability (FP), the most straightforward method is the counting method, which biases FP toward the extreme probabilities (Eckel, 2007). To convert the 21 member TI output to forecast probability, the researcher employed a more advance method called uniform ranks. The ensemble members' output of TI at one grid point is assumed to be equally likely and ranked ordered (sorted from least to greatest). Uniform rank's total probability is divided into $n+1$ bins (n =total number of members) which equal the number of possible ranked positions of the threshold when combined with the members (Hamill and Colucci, 1997). The rank-probability bins that exceed the threshold value are summed to produce the probability of the turbulence exceeding the threshold.

The uniform ranks method considers the probability existing partially in adjacent bins by linearly interpolating the distance between the threshold value and value in the adjacent bin. Eckel (1998; 2003) gives the probability of the partial bin:

$$P(T < V < x_i) = \left(\frac{x_{i+1} - T}{x_{i+1} - x_i} \right) RP_{i+1}. \quad (14)$$

In the preceding equation, T is the threshold value, V is the verification value, x_i is the value of the ensemble member with rank i , x_{i+1} is the value of the ensemble member of $i+1$, and RP_{i+1} is the amount of the probability of the verification rank $i+1$ (that is $RP_{i+1}=1/(n+1)$). Using the uniform ranks method, the following equation calculates forecast probability. Eckel (1998, 2003) adapted the uniform ranks method from Hamill and Colucci (1997).

$$FP = P(T < V < x_i) + (\#ofmembers > T)/(n + 1) \quad (15)$$

In the following example (Figure 15), a ten-member ensemble is used. The output parameter is a turbulence index (TI). The threshold is moderate (MDT) turbulence, measured by TI equal to or greater than five. The resultant forecast probability is 77%.

Ensemble Member TI Output	3.7	4.9	5.1	5.8	6.2	6.6	6.9	6.3	7.5	7.8	
Rank Probability Bins											
	1	2	3	4	5	6	7	8	9	10	11

Figure 15. Uniform Ranks Method: calculating turbulence $FP \geq MDT$ ($TI \geq 5$) using 10-member ensemble.

When T is ranked $n+1$, the fraction of probability is computed by assuming the total probability falls into the upper (right) tail of a Gumbel distribution (Eckel 1998,

2003). The Gumbel cumulative density function (CDF) gives the ability to characterize extreme probability events (Wilks 2006), which is given in the following equation:

$$F(x) = \exp\left\{-\exp\left\{-\frac{(x-\xi)}{\beta}\right\}\right\} \quad (16)$$

The Gumbel distribution parameters are estimated using the statistical method of moments, which employs the sample mean and standard deviation (Wilks, 2006):

$$\hat{\beta} = \frac{s\sqrt{6}}{\pi} \quad (17)$$

$$\hat{\xi} = \bar{x} - \gamma\hat{\beta} \quad (18)$$

$$\gamma = 0.57721. \text{ (Euler's Constant)} \quad (19)$$

With extreme low probability situations (where T falls in the highest bin), the forecast probability is calculated using the following equation adapted from Eckel (1998, 2003):

$$FP = P(T < V) = \left(\frac{1-F(T)}{1-F(x_n)}\right)RP_{n+1} \quad (20)$$

The opposite case of extremely high probability events, where T falls in the first bin is handled by a simpler function to ensure FP drops to zero by TI=0:

$$FP = P(T < V < x_1) + (\# of members > T) / n + 1 \quad (21)$$

where

$$P(T < V < x_1) = \left[1 - \left(\frac{T}{x_1} \right)^3 \right] \frac{1}{n+1} \quad (22)$$

After combining all the members and using the uniform ranks method to calculate the FP, the resultant output was an uncalibrated turbulence FP product for the EUCOM AOR.

Figure 16 shows the uncalibrated FP for FH 24 of the January 31, 2008 00Z run.

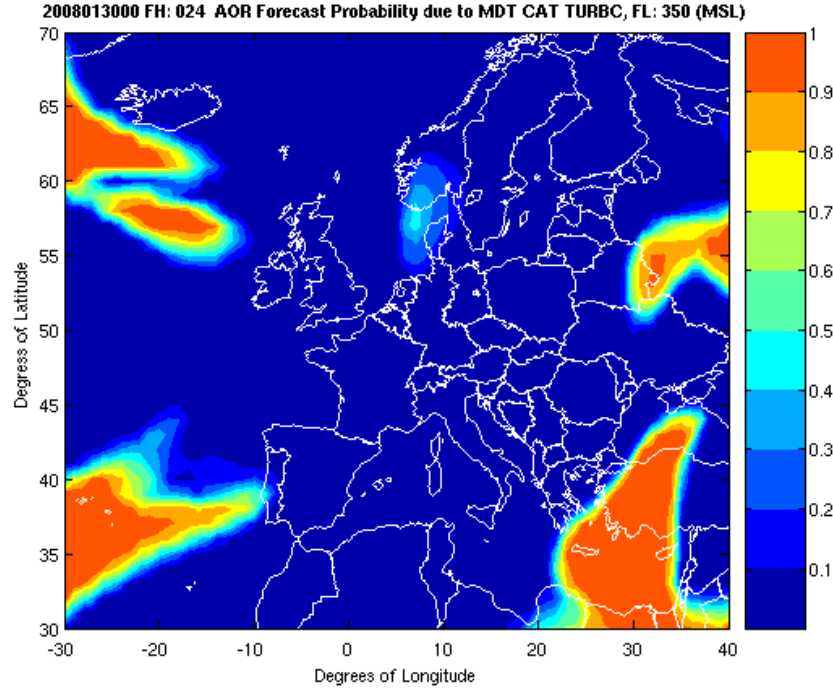


Figure 16. Example of uncalibrated FP output for EUCOM grid.

5. Validation

To validate of the study's TI and FP algorithm, the output of FP of LGT, MDT, SVR turbulence was visually compared against the JGE's upper level turbulence forecast

for the same forecast hour and for similar flight levels. The following comparisons made from Figure 17 are typical results found when the researcher compared other forecast hours and turbulence intensities.

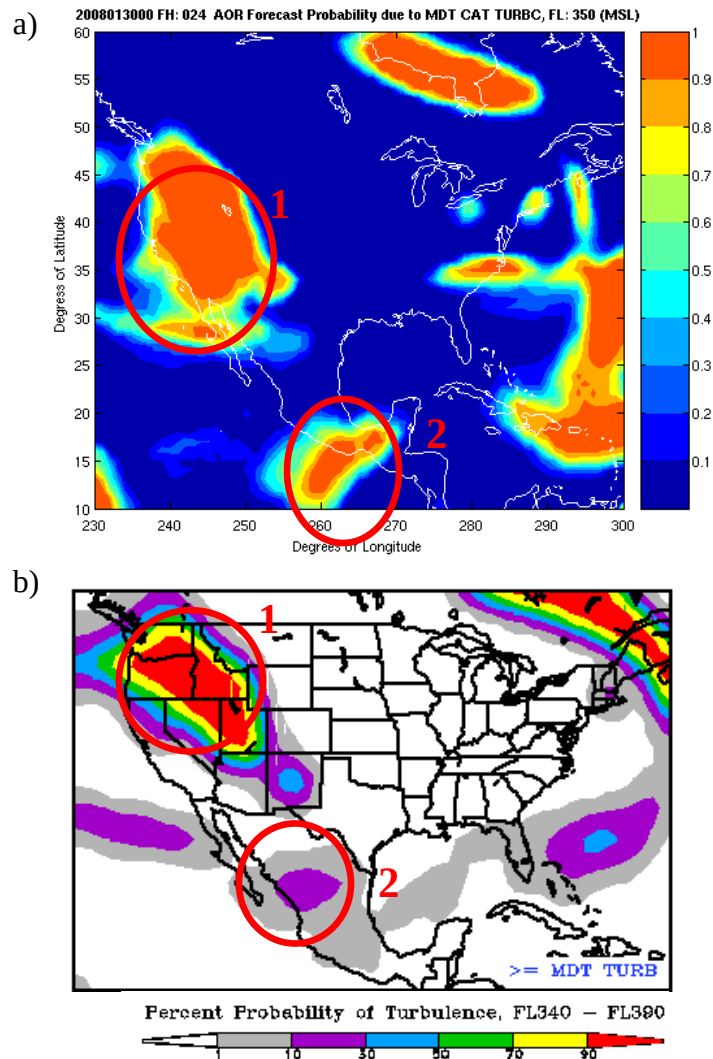


Figure 17. FH 24 of the January 30, 2008, 00Z run; valid January 31, 2008, 00Z.

a) CONUS FP (uncalibrated) of MDT turbulence b) JEFS CONUS FP of MDT turbulence

In area 1, the study's derived FPs are larger in coverage compared to JGE FP. In area 2, the derived FP is higher than JGE FP by 60 to 80%. This difference of FP in spatial coverage and intensity can be attributed to several factors. JEFS combines two

cycles per day to derive the forecast, whereas the research only employs one cycle. Additionally, the study used three levels of data to calculate the TI whereas JEFS uses only two. Furthermore, the researcher interpolated the data whereas JGE estimated the flight levels using the standard atmospheric heights. Although, the differences in FP spatial coverage and intensity noted are not of high concern, because the overall similarities generally validate the processing in the research. Regardless, the actual TI and FP output is not of critical importance. The processes behind FP and decision input calculations are the main focus.

B. PART TWO

This part of the process involves generating calibrated forecast probabilities, ambiguity, and decision input (DI) data. The researcher used a newly fashioned ambiguity and calibration calculation method called the “Taijitu” method (pronounced “ti-chee-tu”). Founded by F.A. Eckel (2008), it estimates the calibration and ambiguity (i.e., systematic and random error) in the ensemble FP caused by limited sampling and ensemble deficiencies. The method concentrates on the errors found in the first and second moments (mean and variance) of the ensemble PDF since they have the greatest affect on FP error (Eckel, 2008).

The GFS ensemble, used in this research, is non-ideal (no internal model perturbations) with finite sampling (only 21 members). GEFS forecasts are likely underdispersive and highly ambiguous. Dispersion is the ensembles representation of error growth (Eckel, 2007). The overall (average) underdispersion is the systematic (predictable) FP error that can be corrected for through calibration (Eckel, 2008). The unknown aspect of FP error (ambiguity) arises from the random variability of the quality of ensemble dispersion (Eckel, 2008).

Transforming from a PDF error to a FP error is a matter of comparing the probability of the ensemble and true PDF for the chosen threshold. For example, an ensemble forecast for 2-m temperature at one location is fit by a Gaussian PDF of $N(\mu_e = 1.4^{\circ}\text{C}, \sigma_e = 0.9^{\circ}\text{C})$. If the true PDF is $N(1.1^{\circ}\text{C}, 1.3^{\circ}\text{C})$, the unknown PDF errors are

0.3°C in μ_e and -0.4°C in σ_e . Example is taken from Eckel (2008). Figure 18 shows how to find the error in FP for the event of 2-m temperature $\leq 0^\circ\text{C}$.

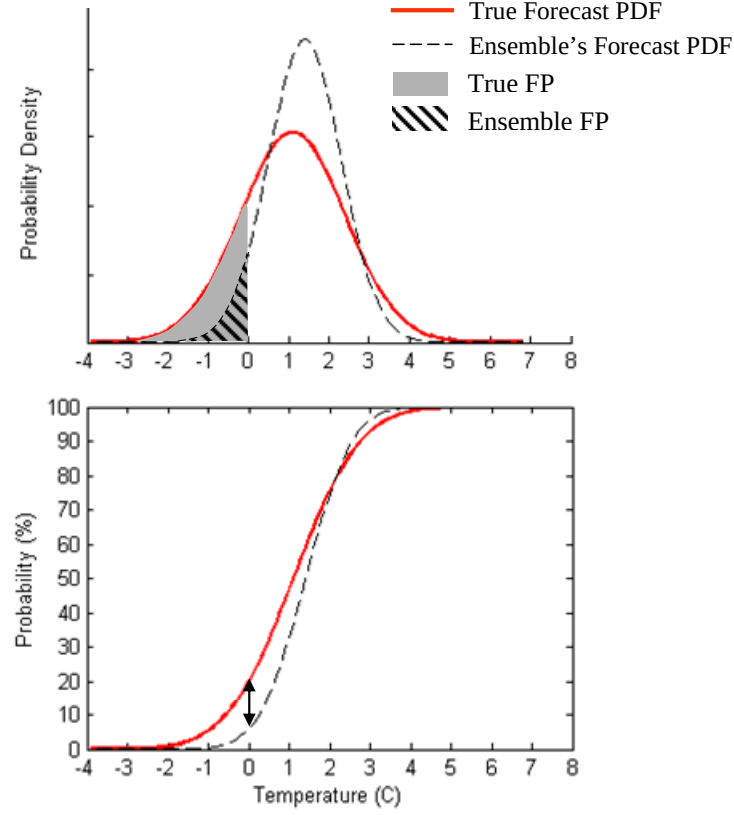


Figure 18. Example comparison of a true and an ensemble forecast PDF (upper plot) and CDF (lower plot). An error of -13.9% in forecast probability for the chance of temperature $\leq 0^\circ\text{C}$ is the difference in the PDF's shaded areas, or the difference in the two CDFs indicated by the double arrow [From Eckel, 2008].

The Taijitu method calculates ambiguity and calibration data using error in the first and second moments of the ensemble PDF:

- 1) the mean of the error in the ensemble mean (μ_{em}),
- 2) the standard deviation of the error in the ensemble mean (σ_{em}),
- 3) the mean of the error in ensemble spread (μ_{es}), and
- 4) the standard deviation of the error in the ensemble spread (σ_{es}).

Each ensemble PDF moment has its own error distribution determined by verifying the ensemble (Figure 19), which is standardized with respect to the average spread of the true PDF (Eckel, 2008). The standardized error distribution of the ensemble mean is assumed to be Gaussian, positioned on the long-term bias of the ensemble mean, μ_{em} , with a standard deviation, σ_{em} (Eckel, 2008). The fractional error distribution for ensemble spread is assumed gamma.

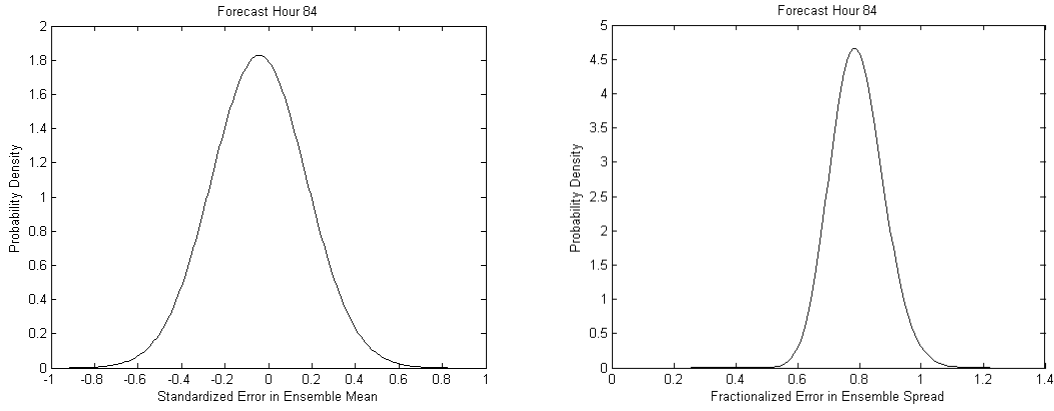


Figure 19. Sample standardized error distributions for FH 84.

In this research, ensemble mean bias (μ_{em}) was set to zero following the assumption that the TI thresholds were calibrated to account for any TI bias. The standardized σ_{em} was determined using a simplified process. Rather than computing the variance of error in ensemble mean from all the data subsets, the researcher used the σ_e' distribution whose variance is dependant on ensemble size (Eckel, 2008). Standardized to a normal PDF where $N(0,1)$, $\sigma_{em} = 1/\sqrt{n}$ (where $n=21$ members) ≈ 0.218 (Eckel, 2008).

To find the distribution of fractional error in ensemble spread, the researcher acquired a long-term verification dataset of GEFS for turbulence related parameters from AFWA. Ideally, the error distribution would be for TI itself, but this is not practical since TI observations are not available. The researcher used the GEFS geopotential height

output at the 500 hPa height level (H5) to estimate the ensemble TI error since 1) winds are generally geostrophic at this level in the middle to high latitudes and 2) the spacing of the geopotential heights indicate the speed of the wind with their inversely proportional relationship (Wallace and Hobbs, 1977). Since the TI calculation is based on vertical and horizontal wind shear, one can assume the tighter (looser) the geopotential height gradient (vertically and horizontally) is, the greater (lesser) the wind shear is. Hence, the ensemble error in the geopotential heights will reflect in the ensemble error in TI.

The long-term H5 spread variance data contains total variance and total mean squared error of the ensemble mean (MSE) information for the 30-60°N global latitude band (10,800 total grid points). The data spans from August 25 to September 29 (37 days total) for the 00Z model run. The researcher trimmed the data set to 20 days (by taking every other two days) to limit correlation among results. Twenty days worth of data was deemed sufficient to estimate error variability. From this 20 day set, approximately five percent of the data is missing since many days have one or more forecast hours (tau) missing. The researcher sorted the 20-day dataset by tau and calculated the fractional error in the spread (FES) for each day:

$$FES = \frac{\overline{\sigma_e}}{RMSE_{\bar{e}}} \quad (23)$$

where $\overline{\sigma_e}$ is the average ensemble spread and $RMSE_{\bar{e}}$ is the root mean square error of the ensemble mean. The following table shows the daily FES values for forecast hour 84.

DATE	FES (FH 84)
2007082500	0.719
2007082800	0.866
2007082900	0.792
2007090200	0.821
2007090300	0.799
2007090500	0.716
2007090600	0.799
2007090900	0.841
2007091000	0.877
2007091200	0.793
2007091300	0.774
2007091400	0.632
2007091600	0.899
2007091800	0.673
2007092000	0.784
2007092300	0.737
2007092500	0.803
2007092600	0.709
2007092800	0.812
2007092900	1.02
μ_{es}	0.793
σ_{es}	0.0859

Table 2. All Fractional Error Spread (FES) values for forecast hour (FH) 84.

To find the mean of the error in ensemble spread (μ_{es}), all the daily FES computations were averaged for one tau. Averaging all the FES for FH 84 yields $\mu_{es} = 0.793$. The μ_{es} is the systematic aspect of the spread error which can be corrected for through calibration (Eckel, 2008). To calculate the random aspect of spread error (σ_{es}), the standard deviation of the 20 FES values was computed. The σ_{es} represents the random aspect of the ensemble spread error (Eckel, 2008), which creates ambiguity. The standard deviation for FH 84 FES data yielded $\sigma_{es} = 0.0859$.

Figure 20a illustrates the effect on FP error for 10 random draws from the ensemble mean and spread error distributions. The standardized error distribution of the

ensemble mean represents the ensemble PDF location error, while the fractional error distribution for ensemble spread represents the ensemble PDF width error (with respect to the true PDF). Sample 1 (Figure 20a) is driven by primarily location error. Sample 2 is driven by primarily width error. Variation of samples 1 and 2 are driven by a combination of location and width errors.

Taking many random samples (1000) yields a distribution of potential error in FP or ambiguity (Figure 20b). The results for 1000 trials produced a skewed distribution which can be fitted by a beta PDF. Figure 21 displays example histograms of FP errors resulting from the FH 84 ensemble PDF error distributions (shown in Figure 19). The distribution of possible FP error skew away from the worst error; the direction of the skew is dependant on the FP. Using Figure 21 as an example, the beta distribution for FP=25% skewed to left (Figure 21a) and FP=75% skewed to right (Figure 21b).

The 5th, 50th, 95th percentiles of the beta PDFs for all FP values are plotted in Figure 22. The 50th percentile the beta PDF (Figure 21) equates to the mean of the ensemble error. Because the mean of the ensemble error represents a systematic error, calibration can be accomplished by adding the 50th percentile values to the raw FP. Thus ‘calibration’ is the reverse of the 50th percentile curve. In a 90 percent confidence interval (CI), the 5th and 95th percentile of the beta PDF equate to the spread of the ensemble error (Eckel, 2008) and represent the variability of the FP error (random error). The non-calibrated 50th percentile is subtracted from 95th percentile and the 5th percentile is subtracted from the non-calibrated 50th percentile to achieve the upper and lower ambiguity bounds respectively. The ambiguity bounds are added to the calibrated 50th percentile curve to compose the “Taijitu” curve pictured in Figure 23. The method takes its name from the resemblance it has to the Chinese taijitu symbol (Eckel, 2008).

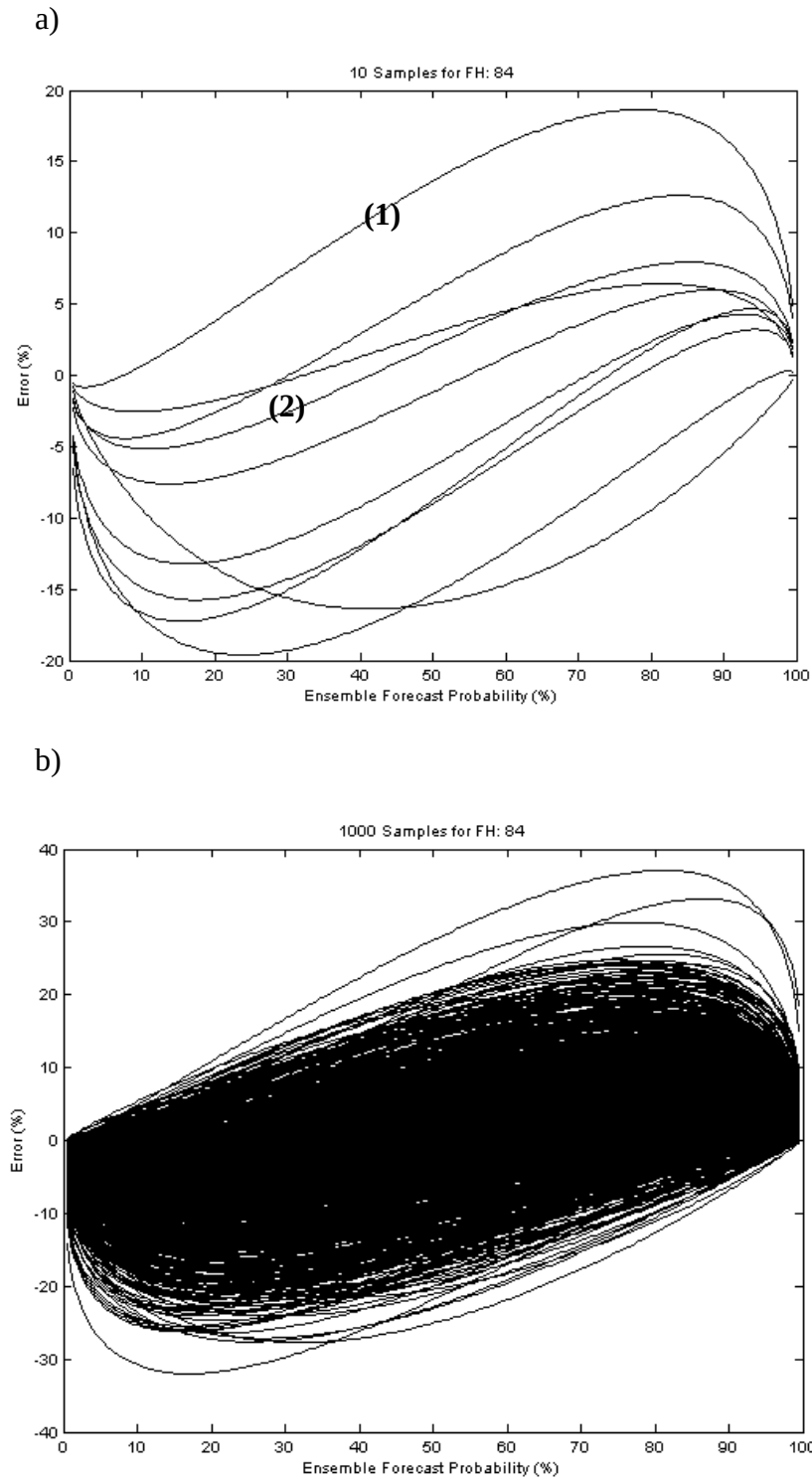
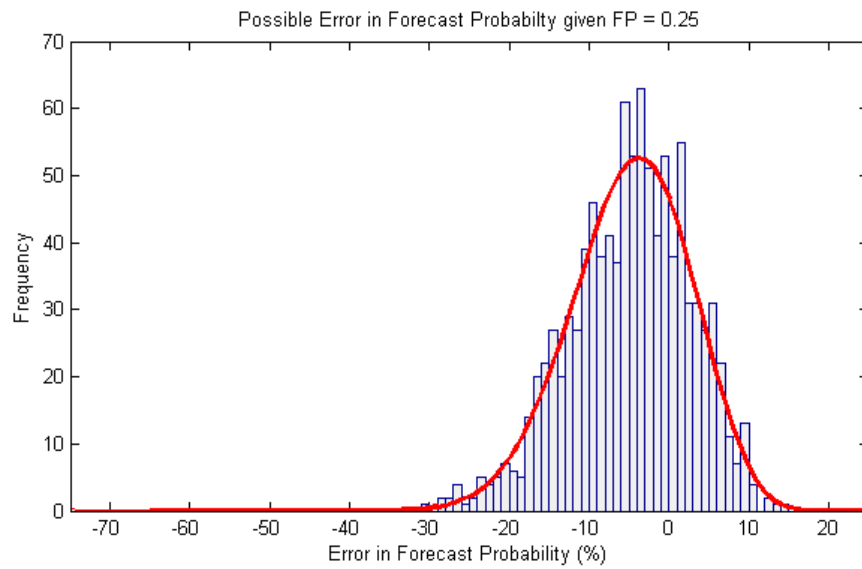


Figure 20. Error (%) in FP for a) 10 and b) 1000 random errors in the ensemble mean and spread for FH 84.

a)



b)

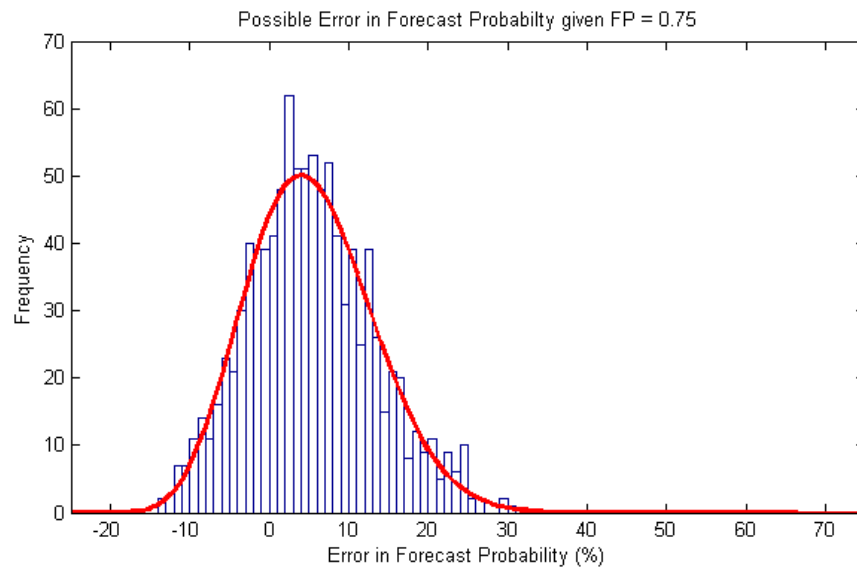


Figure 21. Histogram of 1000 sample FP errors, with fitted beta distribution, resulting from the ensemble PDF error distributions in Figure 19 for FP of (a) 25.0%, (b) 75.0%

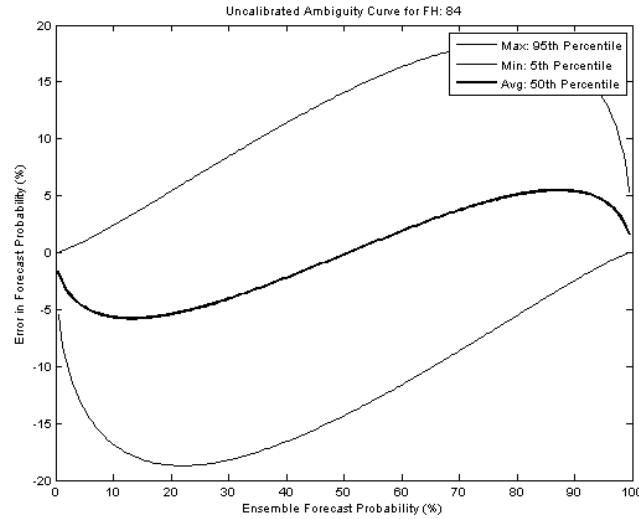


Figure 22. Uncalibrated 5th, 50th, 95th percentiles of the beta PDFs of possible FP error for FH 84.

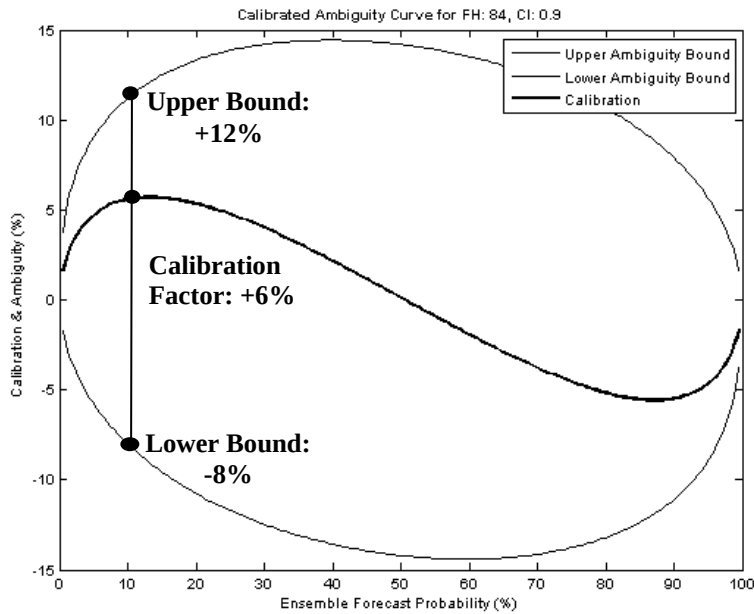


Figure 23. Ambiguity and Calibration Curve for FH 84, 90% CI: The FP calibration (thick line) is simply the opposite of the 50th percentile of the ensemble error distribution (beta PDF). The upper and lower bounds of ambiguity (thin lines) are the 90% confidence interval of the true forecast probability, found by taking the difference of the 5th and 95th percentiles from the 50th percentile. Displayed is calibration factor for FP=10%, upper and lower ambiguity bounds for calibrated FP (16%)[After Eckel, 2008].

To formulate the error bars, the calibration factor is added to the uncalibrated FP. The ambiguity intervals are computed from the uncalibrated FP. For example in Figure 23, the upper bound of the ambiguity interval is +12% and the lower bound is -8% for FP=10%. The calibration factor for FP=10% is +6%, which makes the calibrated FP=16%. Hence, ambiguity interval for FP=16% is 8-28%. Note, because of the “s” shaped curved, most upper and lower bounds of the ambiguity intervals are not equal.

In order to communicate optimized decision input, the proper confidence interval (for the beta PDF described above) for the ambiguity intervals must be chosen. The amber regions in the research decision input products are influenced by the extent of the ambiguity intervals. If the ambiguity intervals are too small, the amber regions will be smaller, giving the user a false sense of high certainty in the decision input. Conversely, if the ambiguity intervals are too large, the amber regions might be too large, which removes the ability of the product to give decisive recommendations.

The researcher tested CIs (70%, 80%, 90%, 95%, and 99%) to determine the affect on the DIs. The higher (lower) CI conveyed more (less) of the ambiguity. Figure 24 illustrates the maximum ambiguity bounds for each CI at FH 84. The CI with average maximum bound (approximately $\pm 15\%$) is the 90% CI. This CI is a balance between the 70% and 99% CI. It reveals a large degree of the calculated ambiguity, but still yields a reasonable decision aid. For these reasons, the 90% CI is the chosen CI for this study. See Appendix A for three forecast hour comparisons of varying CI.

To subjectively approximate uncertainty (error bars) for RT, the researcher assumed a $\pm 5\%$ confidence interval to represent the user’s inability to specify a precise RT. Using this confidence interval for all RT is not realistic. A user with a RT <0.05 would have a negative RT interval bound. A user with a RT >0.95 will have a RT interval bound >100%. The researcher used a second order polynomial curve to avoid this problem achieving the RT uncertainty intervals (Figure 25). As with FP, users with extreme high and low RT (i.e. RT=0.00 or 1.0) have the most confidence of their tolerance to risk (i.e. absolute tolerance to no or all risk). The RT=0.5 user may be unsure of his/her tolerance to risk. Hence, RT=0.5 has the most uncertainty with interval bounds of ± 0.05 .

To incorporate the ambiguity and calibration data into the graphical products, the data cycled through a DI algorithm to compare FP and RT, and produce stop-light DI data. Using look-up tables, the researcher added the calibrated factor to each of the grid's FP. The ambiguity bounds were added to the calibrated FP to formulate the ambiguity interval about FP. To formulate the stoplight DI product, the algorithm compared the confidence intervals of FP and RT. If the lower bound of the FP error bar is above the upper bound of the RT error bar, then the risk is deemed unacceptable (no-go recommendation) and the grid point is colored red (Figure 26a). If lower bound of the RT error bar is above the upper bound of the FP error bar, then risk is deemed acceptable (go recommendation) and the grid point is colored green (Figure 26b). Otherwise, the FP and RT error bars overlap, so there is no clear decision (inconclusive) and the grid point is colored amber (Figure 26c). The minimum confidence interval in the green and red DI (the FP and RT probability are independent of one another) is $(CI + (100-CI)/2)^2$. The minimum CI occurs when the FP and RT error bars are stacked atop one another (the end bounds just meet one another). With the research CI=90%, the minimum confidence is $95^2\% = 90.25\%$.

As stated previous, using CIs with different significance alters the length of the FP error bars (ambiguity intervals), therefore increasing or decreasing the amount of ambiguity conveyed. Inconclusive DI is risk that is too close to call; either a go or no-go DI may be acceptable. In addition to the calibrated FP and stoplight DI graphical product, an additional product displays the maximum or “worst-case” potential error (ambiguity) in the FP. In order to generate this product, lower and upper bounds of the FP ambiguity intervals are compared with each other at each grid point. Which ever bound is greater is deemed as the worst-case potential error (PE).

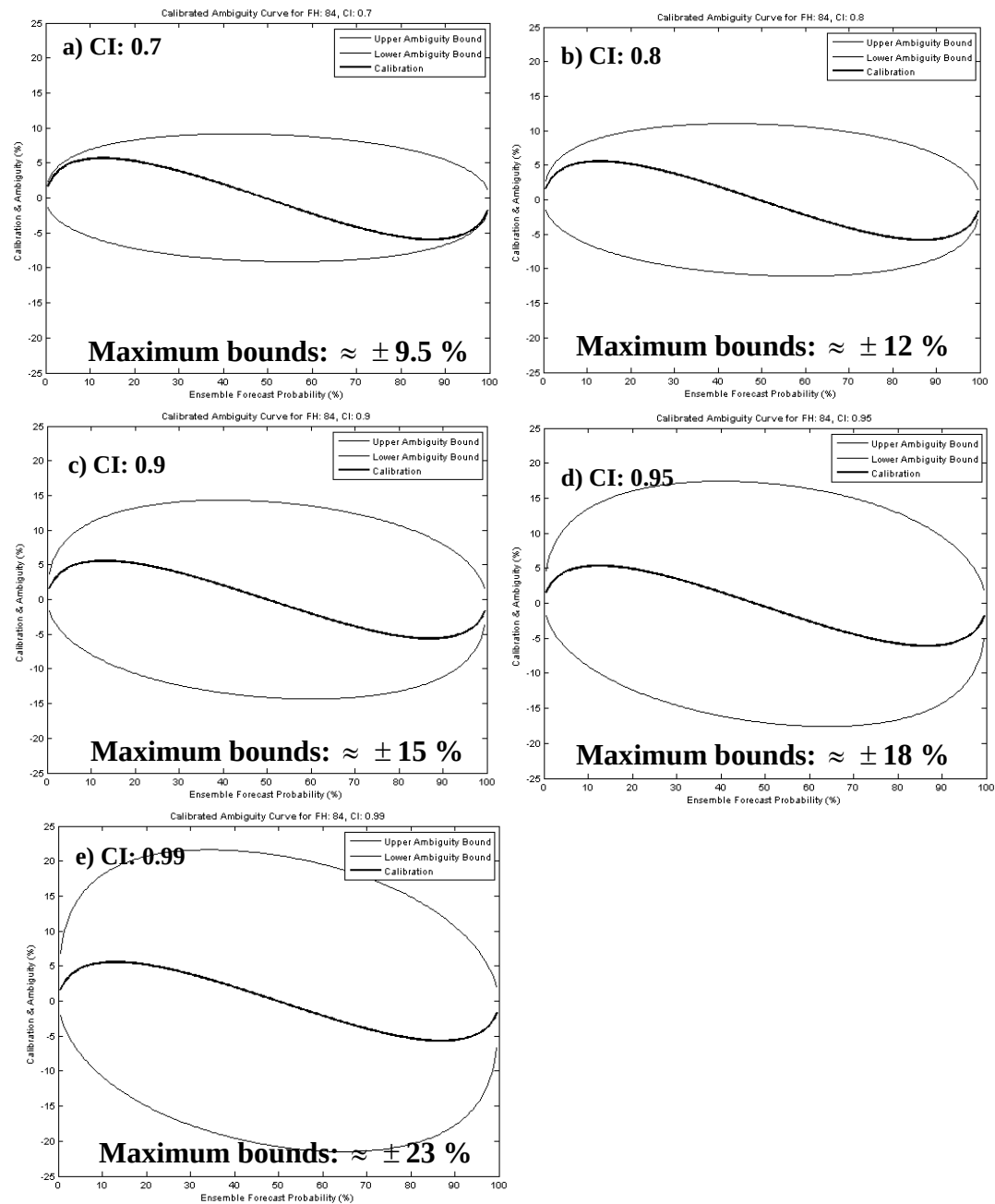


Figure 24. Ambiguity Curves for FH 84, CIs a) 70% b) 80% c) 90% d)95% e) 99%

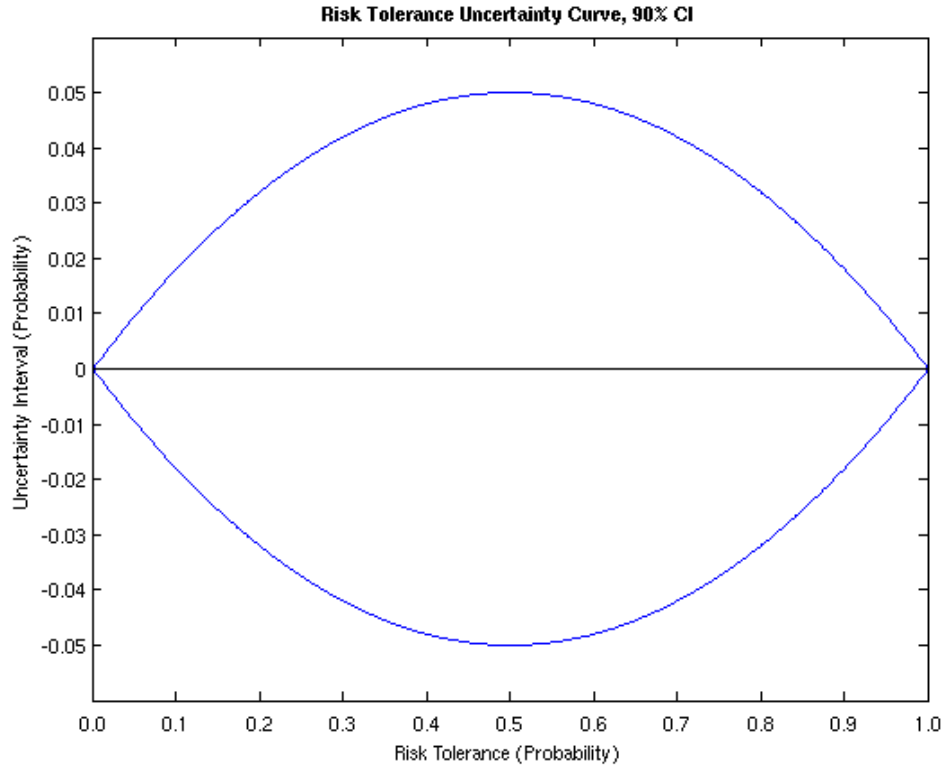
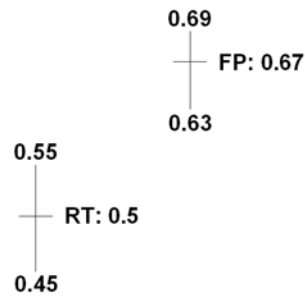


Figure 25. Risk Tolerance Uncertainty Curve.

C. PART THREE

In order to visualize DIs from Part two, the researcher devised a graphical user interface (GUI). If all the possible graphical products were generated for one risk tolerance (13 flight levels, 14 forecast hours, and 3 turbulence intensities), the total number of products would be 546. Using a simple GUI, a user can generate one set of graphical products for his/her specific risk tolerance, mission threshold turbulence intensity, forecast hour, and flight level (Figure 27) and eliminate unnecessary product generation.

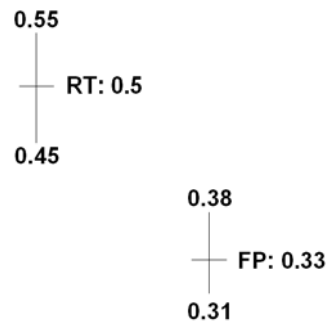
a)



Decision Input: Risk Unacceptable



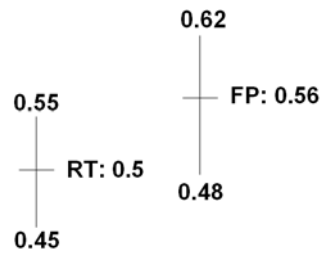
b)



Decision Input: Risk Acceptable



c)



Decision Input: Inconclusive



Figure 26. (a-c) Calibrated FP and user RT (center of error bar) and associated confidence intervals showing different DIs

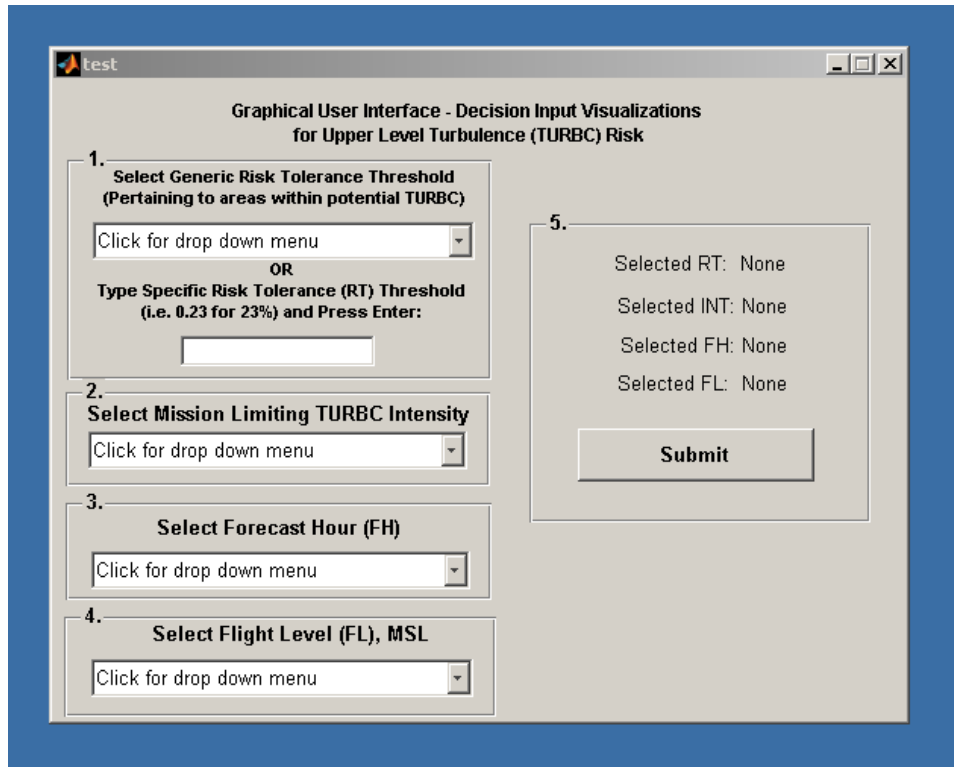


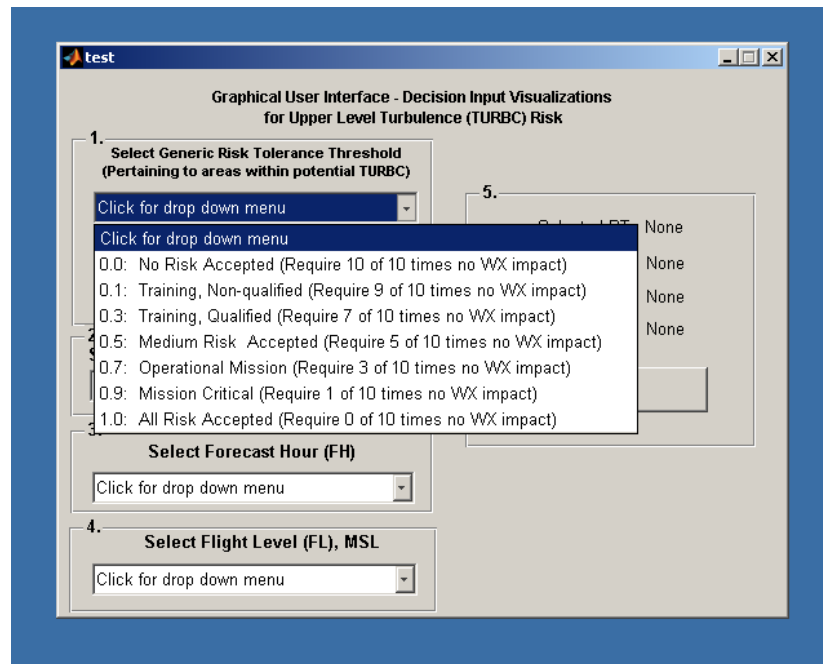
Figure 27. Prototype GUI.

In the first block, the user has the option of either using the drop down menu or inputting a specific risk tolerance. The first option is geared toward those users that have not conducted a thorough study of their risk tolerance and/or are not educated in stochastic methods and decision science (a beginner user). The drop down menu includes generic risk tolerances and a description/example for the specific risk tolerance. In parenthesis, the relative frequency is displayed for the number of times out of ten the user requires no weather (WX) impact (Figure 28a). Research has shown using natural frequencies helps users make better statistical inferences and interpret the data more correctly than probabilities (Hoffrage et al, 2000). It is important to note, the users requirement to have no weather impact is only valid within potential areas of turbulence, not outside. In the second through fourth blocks, the user chooses his/her mission limiting turbulence intensity level (Figure 28), forecast hour (Figure 29a), and flight level (MSL) (Figure 29b).

After submitting all the inputs, the end result is the generation of three types of graphical products: calibrated FP, maximum PE in FP and, a stop-light DI product. Ideally, the FP and the PE data should be overlaid, but Matlab is too limited. Showing all these products together furthers education and gives validity to the DI product. It helps the uneducated user understand the process behind the final DI product. The examples in Figure 30 thru Figure 32 uses mission threshold turbulence intensity of MDT, forecast hour of 24, flight level of 35,000 feet (MSL), and a $RT=0.3$ or 30%. Instead of displaying FP for the chance of turbulence, the scale was reversed to show mission success chance due to turbulence for more operational relevance (i.e. probability of no impact from MDT turbulence), shown in Figure 30.

To illustrate, points A, B, C in Figure 30-Figure 32 represent three different decision inputs. Point A represents 10% chance of mission success (90% FP) in Figure 30. The worst case PE at this point is approximately 8% (Figure 31), so the maximum ambiguity interval is from 82-98%. The uncertainty interval for $RT=30\%$ (according to Figure 25) is $\pm 4\%$. With RT interval of 26-34% < FP=82-98%, the DI for point A is unacceptable risk (red) shown in Figure 32. Point B represents an approximately 60% chance of mission success (40% FP). The worst case PE at this point is approximately 12%, so the maximum ambiguity interval is from 28-52%. Since this interval overlaps the RT interval of 26-34% by 6%, point B is deemed inconclusive (amber). Point C represents a 90% chance of mission success (10% FP). The worst case PE at this point is approximately 7%, so the maximum ambiguity interval is 3-17%. With RT interval of 26-34% > FP=3-17%, the DI for point C is acceptable risk (green) shown in Figure 32.

a)



b)

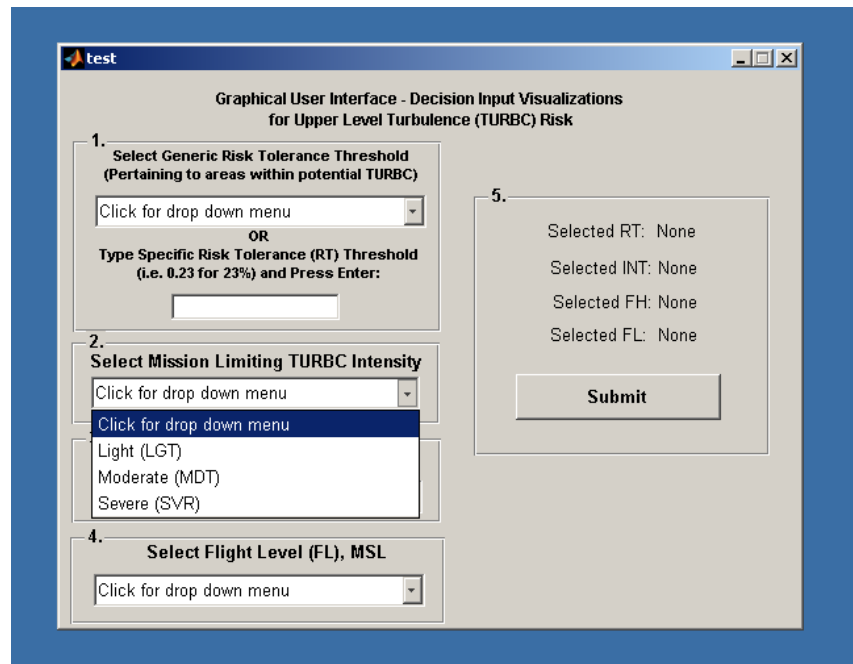


Figure 28. Prototype GUI: Drop down menu of a) generic risk tolerances b) mission limiting turbulence intensities

a)

The screenshot shows a window titled "test" with the subtitle "Graphical User Interface - Decision Input Visualizations for Upper Level Turbulence (TURBC) Risk". The interface contains five numbered sections:

- Select Generic Risk Tolerance Threshold (Pertaining to areas within potential TURBC)**: Includes a drop-down menu and a text input field for a "Type Specific Risk Tolerance (RT) Threshold (i.e. 0.23 for 23%) and Press Enter:". The drop-down menu is currently open, showing a list of values from 00 to 84 in increments of 6.
- Select Mission Limiting TURBC Intensity**: Includes a drop-down menu.
- Select Forecast Hour (FH)**: Includes a drop-down menu.
- Select Flight Level (FL), MSL**: Includes a drop-down menu.
- Summary Section**: Displays "Selected RT: None", "Selected INT: None", "Selected FH: None", and "Selected FL: None". A "Submit" button is located below these fields.

b)

The screenshot shows the same window as in (a), but with the "Select Flight Level (FL), MSL" drop-down menu open. The list of values displayed is: 450, 400, 350, 300, 250, 200, 180, 160, 140, 120, and 100. The other sections of the interface remain the same as in (a).

Figure 29. Prototype GUI: a) Drop down menu of forecast hours b) Drop down menu of flight levels

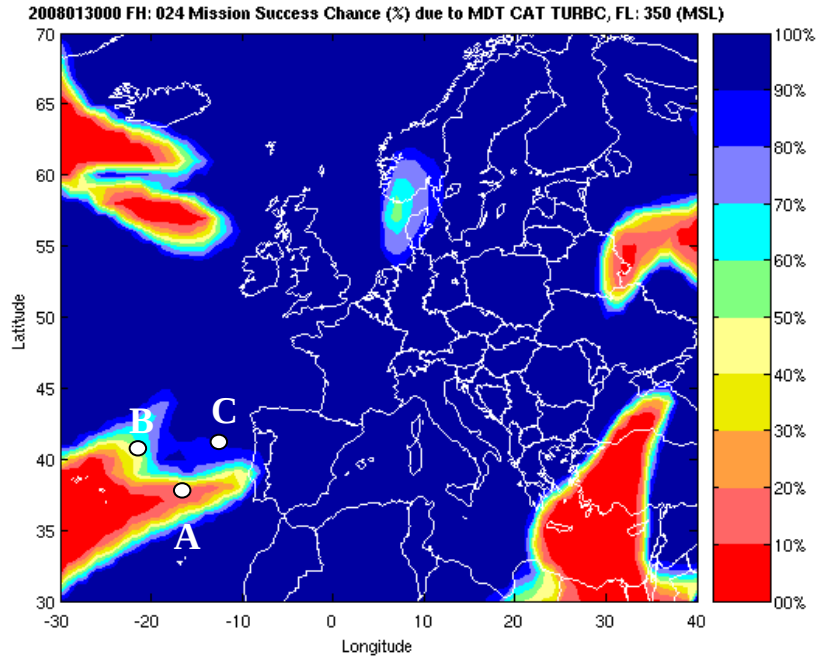


Figure 30. EUCOM Mission Success Chance (FP calibrated) due to MDT turbulence for FH 24 of the January 30, 2008, 00Z run; valid January 31, 2008, 00Z

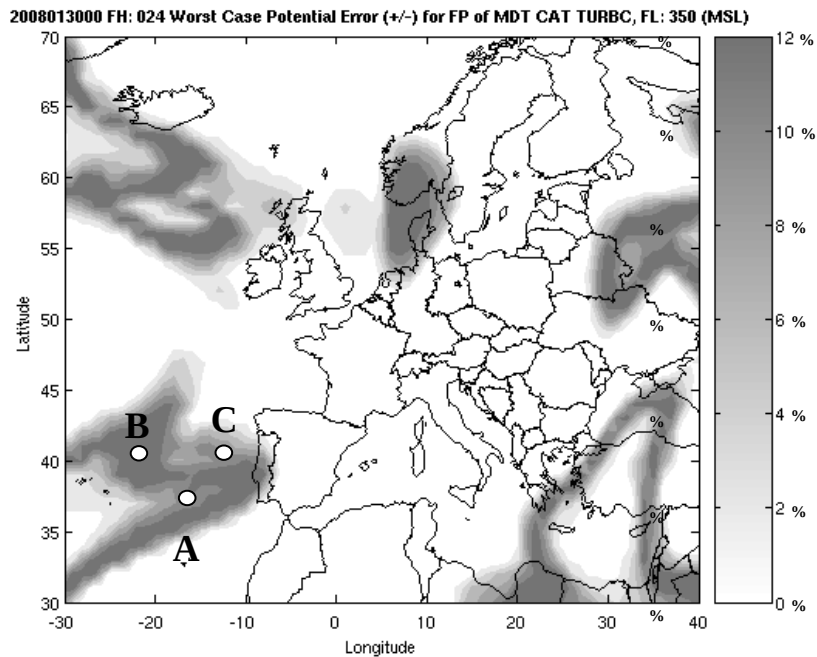


Figure 31. EUCOM Worst Case Potential FP Error of MDT turbulence for FH 24 of the January 30, 2008, 00Z run; valid January 31, 2008, 00Z

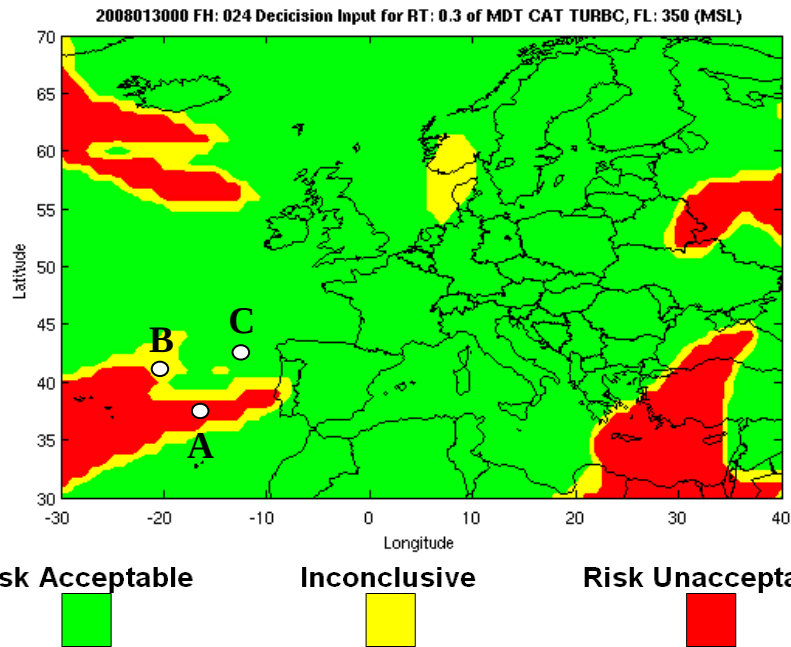


Figure 32. EUCOM DI for RT: 0.3 of MDT turbulence for FH 24 of the January 30, 2008, 00Z run; valid January 31, 2008, 00Z

IV. RESULTS

A. EVOLUTION OF TAIJITU CURVES THROUGH TIME WITH INCREASING LEAD TIME

Table 3 shows the mean and standard deviation results for all forecast hours for the distribution of the EF spread fractional error. The general reduction of mean in the EF spread is due to dispersion deficiencies in the GEFS. Early in the forecast period, FH 00 – 18, the magnitude of μ_{es} was greater than one. At FH 24, the magnitude was approximately equal to one. Beyond FH 24, the magnitude of μ_{es} dropped below one. This trend is due to the ensemble compensating for its underdispersion. As stated in the previous chapter, GEFS is underdispersive because of limited sampling and lack internal model perturbations. To somewhat compensate for the underdispersion, GEFS over perturbs its initial conditions. The actual error growth exceeded the ensemble dispersion, so by FH 24, the magnitude of $\mu_{es} < 1$. When compared to the true PDF, the ensemble PDF was too wide early in the forecast period, turning too narrow later in the forecast period. This reverse in of the calibration curve through the forecast period (Figure 33) was in response to the change in average ensemble error from >1 to <1 (driven by the downward trend of the ensemble spread errors). The calibration curve corrected the average error as explained in Chapter II.

Note the additional downward trend of the values for the standard deviation of error in EF spread, σ_{es} (Table 3). As forecast lead-time increased, the ensemble members diverged toward the climatological or climatic variance, the maximum limit of the ensemble spread. The climatic variance is the variance of all naturally occurring states (Eckel, 2007). As the ensemble spread reached this maximum spread, all the members asymptoted to this limit. Hence, the standard deviation of the error in EF spread decreased.

Even though the standard deviation of the ensemble spread error σ_{es} generally decreased, the ambiguity intervals remained almost constant. Recall that σ_{em} is dependant only on ensemble size and is fixed at $1/\sqrt{n}$ (where $n=21$ members) ≈ 0.218 (Eckel, 2008). The nearly constant ambiguity bounds for each CI reveal that σ_{em} dominated the random FP error in comparison to the contribution from decreasing σ_{es} .

Forecast Hour	Mean of Error in EF spread, μ_{es}	Standard Deviation of Error in EF Spread, σ_{es}
00	1.19	0.154
06	1.28	0.179
12	1.14	0.147
18	1.12	0.131
24	1.0077	0.125
30	0.985	0.115
36	0.921	0.112
42	0.9102	0.0998
48	0.862	0.1046
54	0.857	0.098
60	0.845	0.0755
66	0.838	0.0788
72	0.821	0.0774
78	0.816	0.0818
84	0.793	0.0859

Table 3. Mean and standard deviation values for the EF spread fractional error distribution

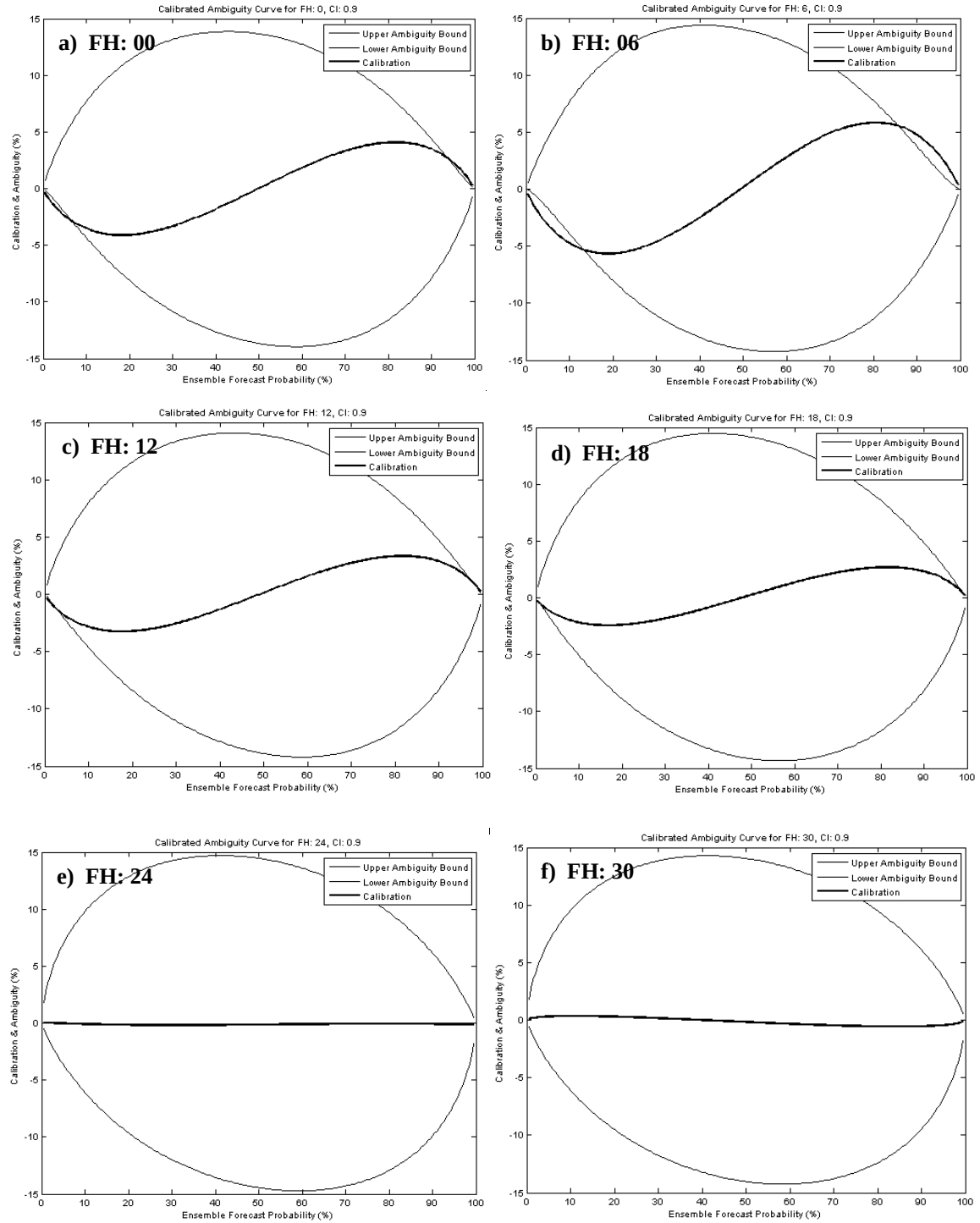


Figure 33. a-o) Calibration and Ambiguity Curves through all forecast lead times, CI: 90%

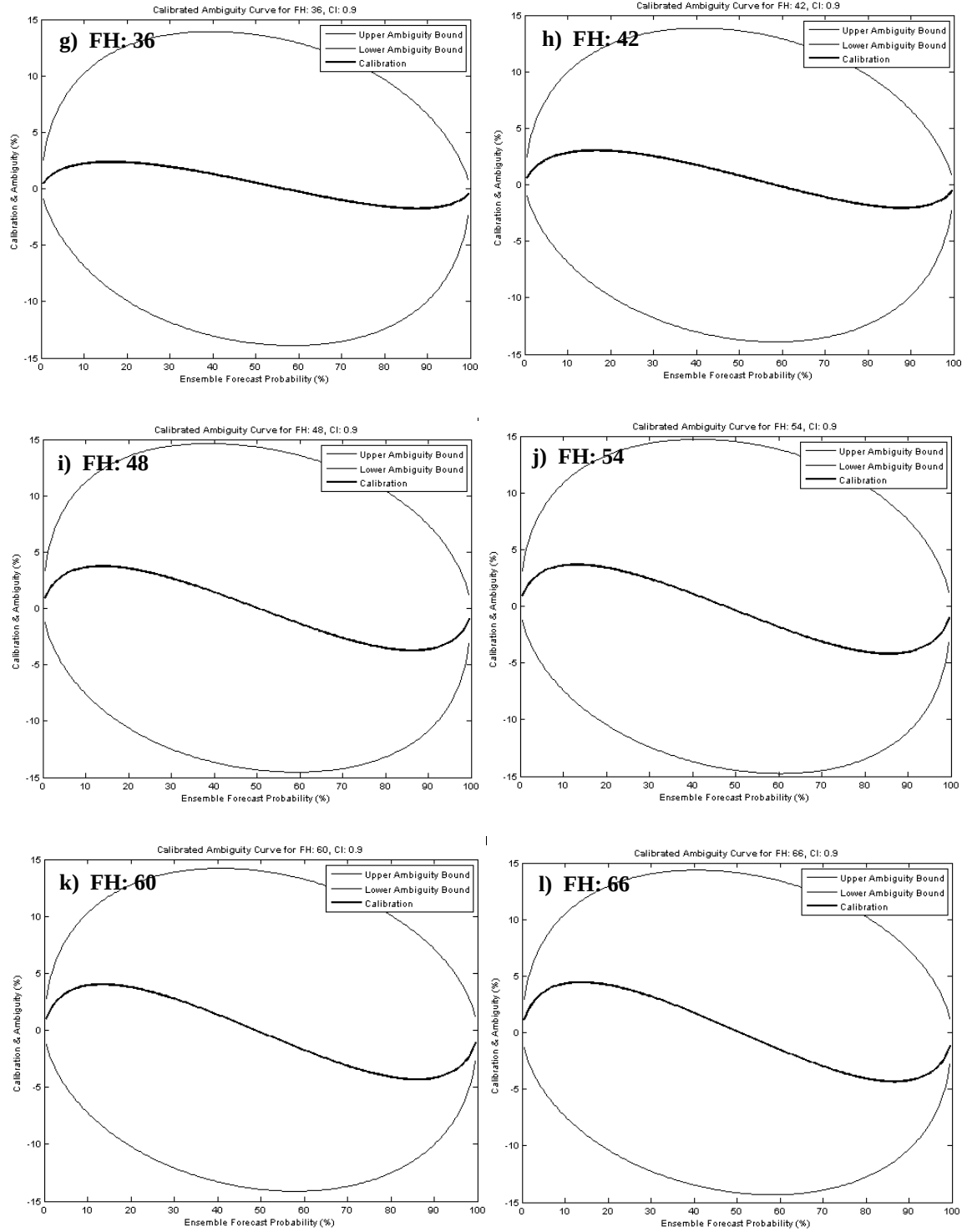


Figure 33 continued.

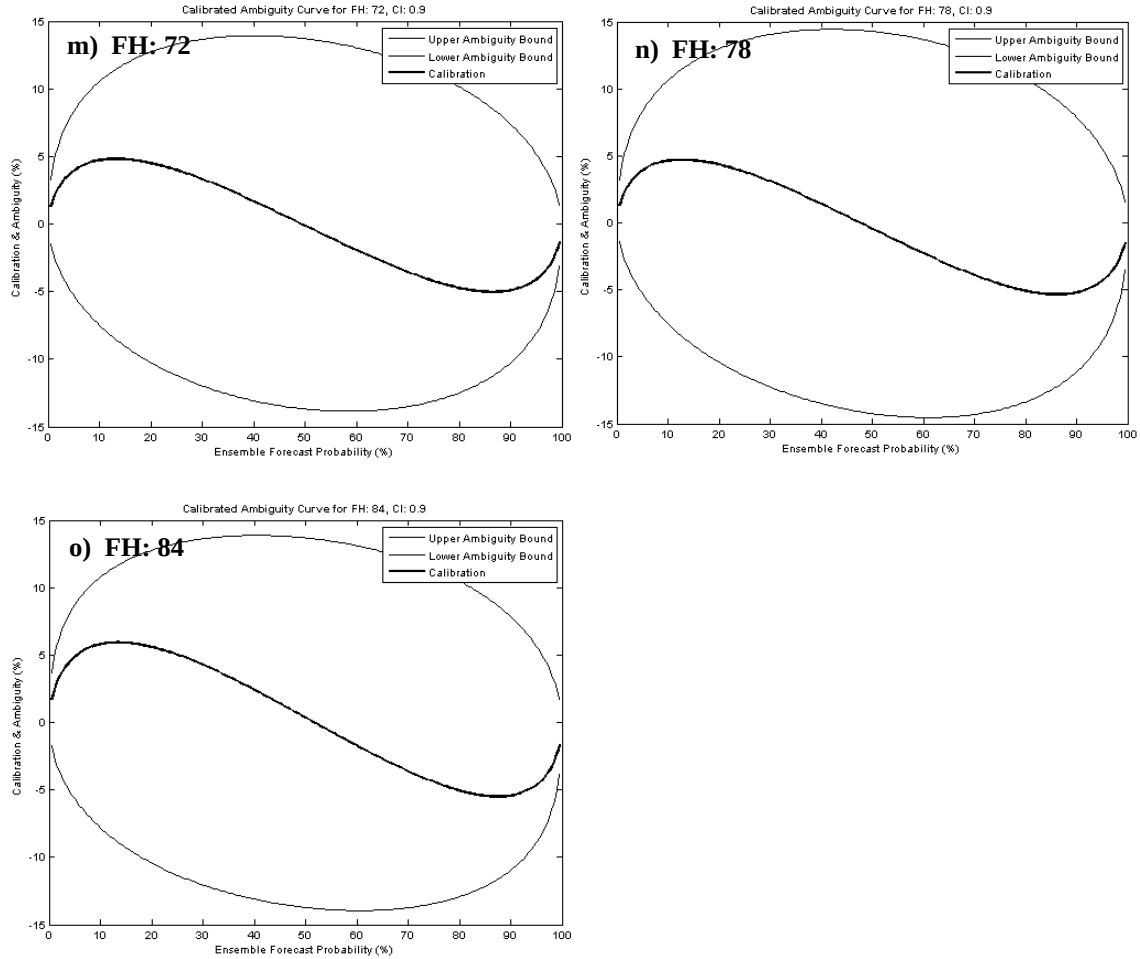


Figure 33 continued.

B. ANALYSIS OF DECISION INPUT (DI) WITH ADDED UNCERTAINTY INFORMATION

1. Simulated DI Product and Analysis Construction

In order to show potential value added and highlight differences between current decision input (DI) products and research results (RSCH), the researcher constructed products focused only on the risk of MDT turbulence to simulate IWEDA and WRAP to compare with RSCH. The unfavorable (mission) threshold of MDT turbulence used for

WRAP and IWEDA was $TI \geq 5$. The marginal threshold for IWEDA and WRAP was assumed to be half the magnitude between the MDT TI threshold and LGT TI threshold ($=2$), which equates to $TI \geq 3.5$.

To create a baseline deterministic product, all TI values in the control member ≥ 5 equated to a forecast of MDT turbulence. The product implies either zero or one hundred percent probability of turbulence (Figure 34).

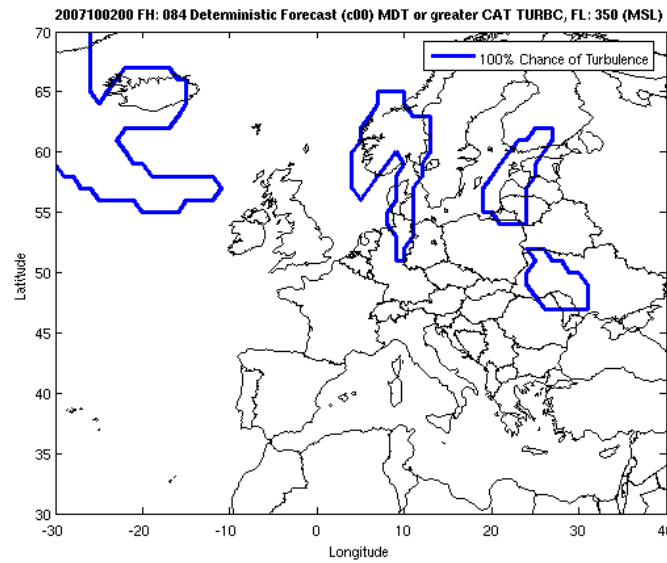


Figure 34. Deterministic MDT turbulence forecast using control member for the fall run, FH: 84, FL: 350

To simulate IWEDA, the researcher applied the marginal and mission MDT turbulence thresholds to the control member's TI. The researcher designated these specific values to each result in order to link them to the stop-light colors on the Matlab colorbar. If the TI for the grid point exceeded the marginal or mission threshold, then the grid point was set to the value of 1.5 (colored amber) or 3.0 (colored red) respectively. If neither of the thresholds are exceeded, the grid point was set to a value of zero (colored green). The green, amber, and red colors represent the decision inputs of acceptable risk, inconclusive, and unacceptable risk respectively (Table 4).

TI threshold	Color Value	Color Assignment	Decision Input
< 3.5	0.0	Green	Acceptable Risk
≥ 3.5 (marginal)	1.5	Amber	Inconclusive
≥ 5.0 (mission)	3.0	Red	Unacceptable Risk

Table 4. IWEDA Simulation: TI Thresholds and corresponding color values, color assignments, and DI

To simulate WRAP, the researcher compared the FP of exceeding the IWEDA marginal and mission thresholds to a fixed RT to assign the color values. If the FP for exceeding the mission threshold is greater than the RT, then the grid point was set to 3.0 (red). Otherwise, if the FP for exceeding the marginal threshold is greater than the RT, then the grid point was set to 1.5 (amber). Otherwise, the grid point was set to 0.0 (green) since neither FP exceeds the RT. See Table 5.

FP vs RT	Color Value	Color Assignment	Decision Input
FP (of either threshold) < RT	0.0	Green	Acceptable Risk
Marginal FP > RT	1.5	Amber	Inconclusive
Mission FP > RT	3.0	Red	Unacceptable Risk

Table 5. WRAP Simulation: FP vs RT and corresponding color values, color assignments, and DI.

To show differences between the IWEDA-like, WRAP-like, and RSCH derived DI products, these color values were subtracted in the following manner: IWEDA minus WRAP; WRAP minus RSCH. To illustrate, when the subtraction of WRAP from IWEDA yielded positive three, IWEDA forecasted red conditions and WRAP forecasted green conditions, translating to a high (pessimistic) decision input bias of IWEDA relative to WRAP. Conversely, when the subtraction yielded negative -1.5, this means

IWEDA forecasted green or amber conditions, and WRAP forecasted either yellow or red conditions respectively, which translated to IWEDA having a low (optimistic) decision input bias. Grey shading means equal decision inputs. Table 6 displays the DI bias colorbar key for IWEDA minus WRAP DI and WRAP minus RSCH DI products.

DECISION INPUT (DI) DIFFERENCE (BIAS) COLORBAR KEY:				
OPTIMISTIC BIAS		ZERO BIAS	PESSIMISTIC BIAS	
Dark Green (color value difference: -3)	Green (color value difference: -1.5)	Gray (color value difference: 0)	Red (color value difference: 1.5)	Dark Red (color value difference: 3)
IWEDA-WRAP OUTPUT:	IWEDA-WRAP or WRAP-RSCH OUTPUT:	IWEDA-WRAP or WRAP-RSCH OUTPUT:	IWEDA-WRAP or WRAP-RSCH OUTPUT:	IWEDA-WRAP OUTPUT:
 - 	 -   - 	 -   -   - 	 -   - 	 - 

Table 6. DI Bias Colorbar Key. Note: WRAP and RSCH never differ by more than one DI category.

Using the two cases of 0000 UTC, 2 October 2007 (fall) run and the 0000 UTC 30 January 2008 (winter) run, the researcher examined MDT turbulence decision input differences using FH: 06 and 84 and RT: 0.1, 0.5, 0.9. The fall case results are discussed in this chapter; the winter case resultant plots are in Appendix B and C.

2. Simulated IWEDA versus Simulated WRAP

IWEDA and WRAP use the same marginal and mission thresholds, but employ the thresholds through different methods to assume a DI. IWEDA depends on a single forecast whereas WRAP incorporates 21 forecast and user's risk tolerance information. The decision inputs are sometimes vastly different.

a. Fall Case: FH 06

To give a general comparison, the FP (stochastic) product conveys uncertainty information, while the deterministic product gives no uncertainty information. The deterministic chart implies one hundred percent chance of MDT turbulence inside the forecasted areas and zero percent chance outside the forecasted areas (Figure 35a). The FP product shows the spread amongst the ensemble's members with respect to the event (Figure 35b).

Area 1 (Figure 35) on the FP chart exhibited 40-60% probability of MDT turbulence, since some the members forecast $TI \geq 5$. The same region of area 1 in the deterministic forecast gave no hint to turbulence since the TI was below five in the control member. A user with a low risk tolerance needed the probability information to incorporate this into their mission planning. The potential loss for encountering moderate turbulence was much too high for a low risk tolerant user, so it is more beneficial if the user did not proceed with the mission.

Conversely, when the deterministic product gives a 100% chance of moderate turbulence (as in area 2 of Figure 35), no user with a mission threshold of MDT turbulence can operate. Using the FP product, the user can see the FP was not 100% but only 50-80%. Users with high risk tolerances are able to operate because they can afford the potential loss.

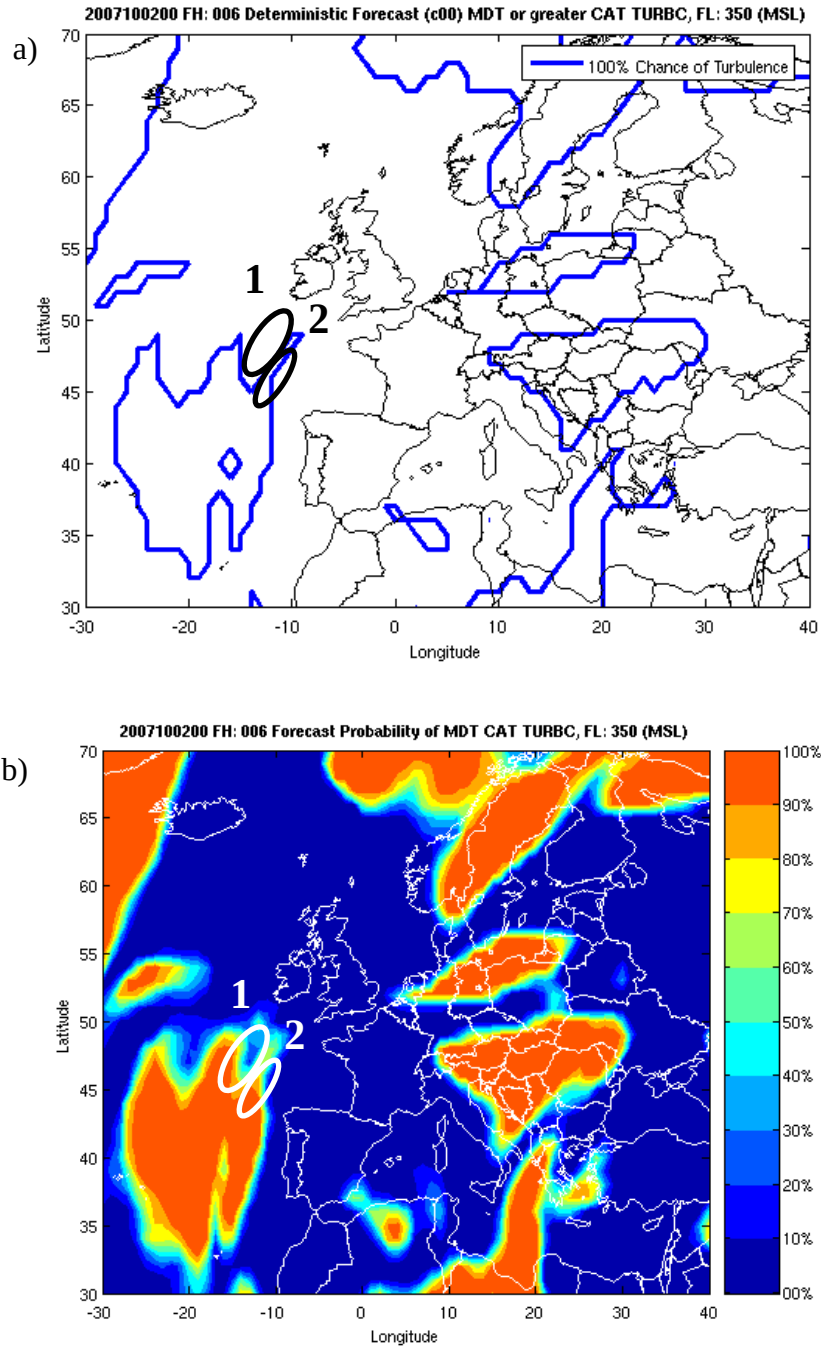


Figure 35. Fall run, FH 06, FL 350 MSL. a) Deterministic MDT turbulence forecast
b) Stochastic MDT turbulence forecast using FP

The differences between the IWEDA and WRAP, with $RT=0.5$, are minimal (Figure 36). The forecast evolution is only six hours. The members of the

ensemble have not begun to spread widely at this point, so for the majority of the region, high certainty exist (many areas with zero or 100 percent probability). So, in a way, the stochastic forecast is acting like the deterministic forecast. Note, the same areas marking 100% probability of turbulence in the deterministic product (Figure 35a) are the same areas deemed as unacceptable risk (red) in the IWEDA product (Figure 37). Similarly, the areas marking >50% probability of turbulence in stochastic product (Figure 35b) are the same areas deemed as red in the WRAP product (Figure 38).

The IWEDA-WRAP DI difference plot is a mixture of optimistic and pessimistic biases (Figure 36). The mixture is due to WRAP's $RT=0.5$, designed to emulate a deterministic decision maker. The $RT=0.5$ forces the stochastic DI to act like a deterministic forecast where the recommendation is binary (go or no-go). As stated in the previous section, the risk is inconclusive (amber) for any FP for exceeding the marginal threshold $> RT$ and unacceptable (red) for any FP exceeding the mission threshold $> RT$.

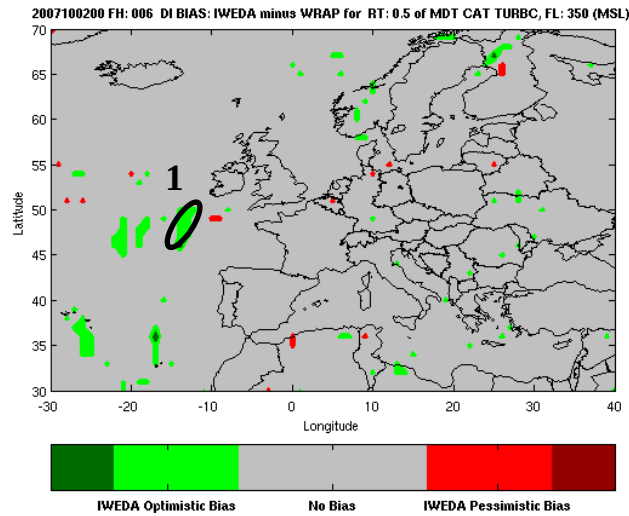


Figure 36. Fall run, FH 06, FL 350 MSL. DI Bias: IWEDA minus WRAP chart.
WRAP $RT=0.5$

In regions of area 1 (Figure 37), IWEDA forecasted green because the single TI solution < 3.5 . WRAP forecasted amber for the majority of area 1 because the FP for exceeding the marginal threshold > 0.5 (Figure 38).

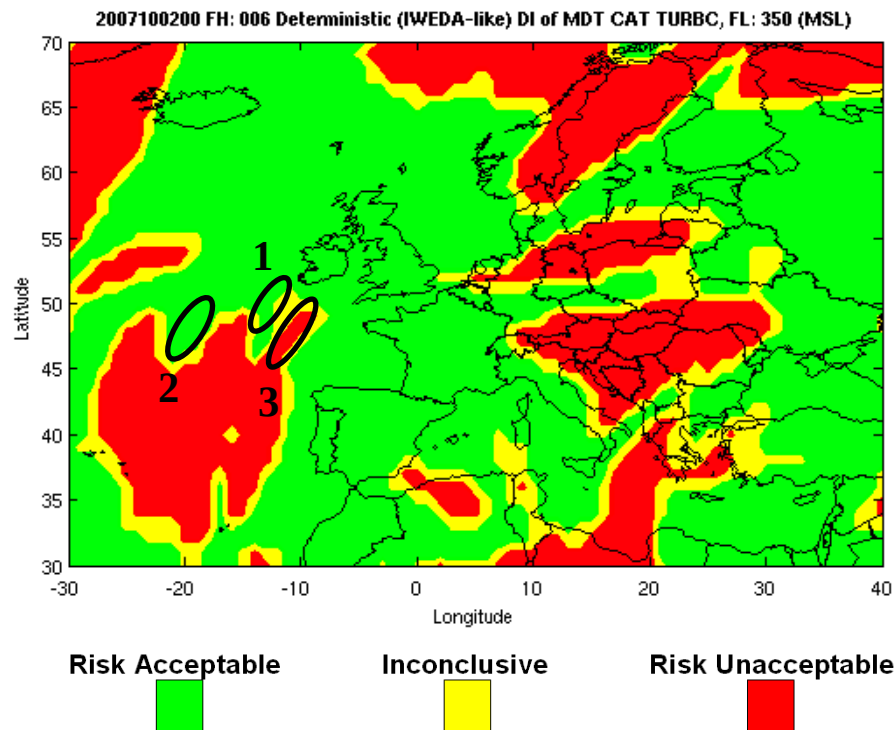


Figure 37. Fall run, FH 06, FL 350 MSL. Simulated IWEDA MDT turbulence DI

Lowering the RT to 0.1 for WRAP changed the DI dramatically (Figure 39). Because the user can not withstand much risk, $(FP) > 0.1$, many of the areas previously deemed acceptable or inconclusive were deemed unacceptable. Users with low risk tolerances can not afford to take large risks because potential losses are relatively too large, and the cost to protect is proportionally smaller. IWEDA DI remained the same. With predominantly more red and amber regions in WRAP DI, the IWEDA DI bias was all optimistic (Figure 40). Using IWEDA DI, a low risk tolerant user flying into area 1 was given a false sense of confidence that the mission will likely not be overcome by moderate turbulence in area 2 (Figure 37). WRAP revealed the mission can not afford to take the risk operating in area 2 in Figure 39.

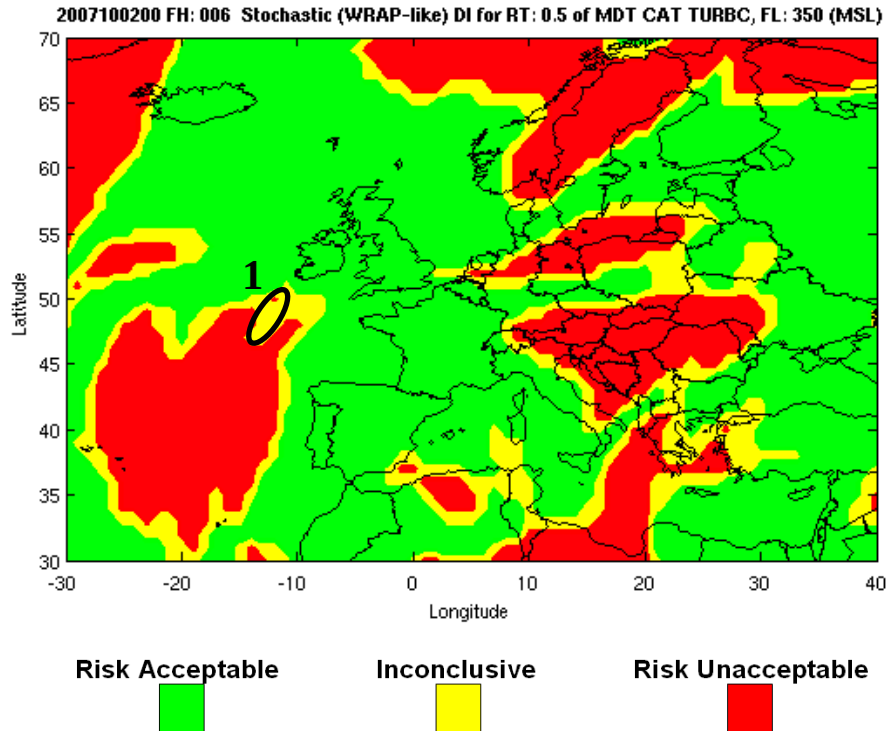


Figure 38. Fall run, FH 06, FL 350 MSL. Simulated WRAP MDT turbulence DI with RT=0.5

Increasing the RT to 0.9 for WRAP again changed the DI dramatically (Figure 41). A user with a risk tolerance of 0.9 can withstand the large risk; their potential losses are relatively small, almost equal to their cost to protect. Many of the areas previously deemed risk unacceptable/inconclusive were deemed risk acceptable. With the predominantly more green regions forecasted for WRAP, the IWEDA DI bias was all pessimistic (Figure 42). Using the IWEDA DI, a high risk tolerant user expected major impacts in parts of the area 3 (Figure 37). Using the WRAP DI, most of area 3 (Figure 41) is instead deemed inconclusive (amber) because of WRAP's inclusion of FP and RT.

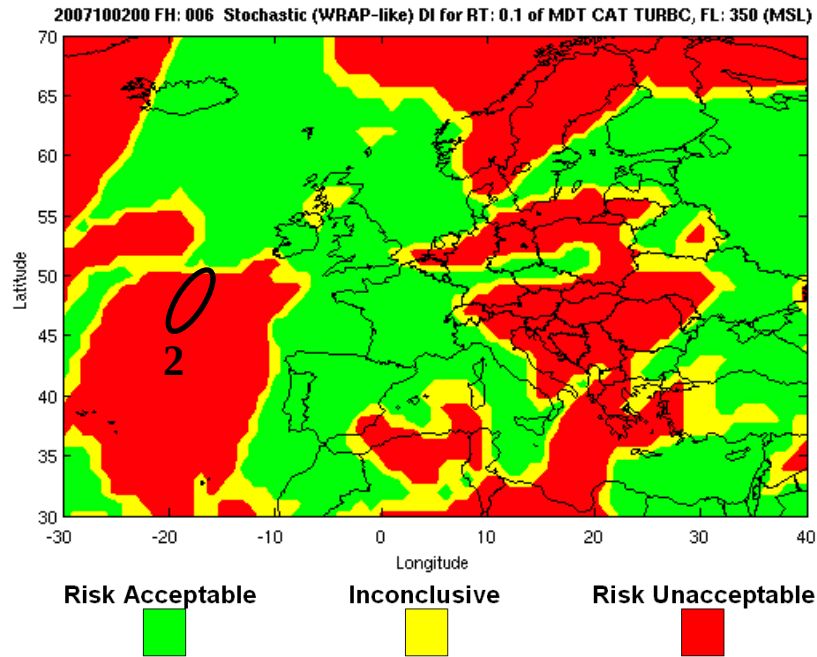


Figure 39. Fall run, FH 06, FL 350 MSL. Simulated WRAP MDT turbulence DI with RT=0.1

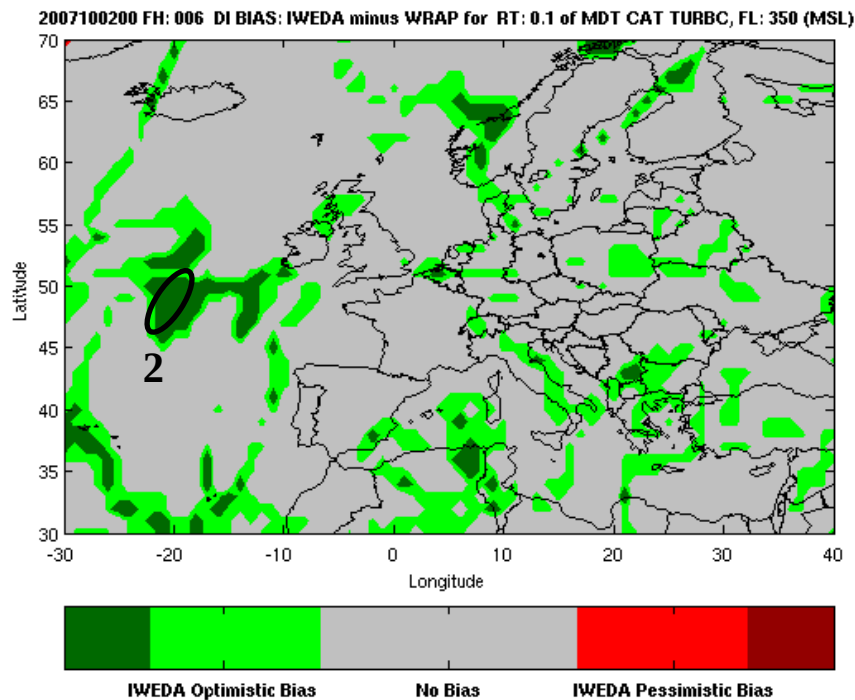


Figure 40. Fall run, FH 06, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.1

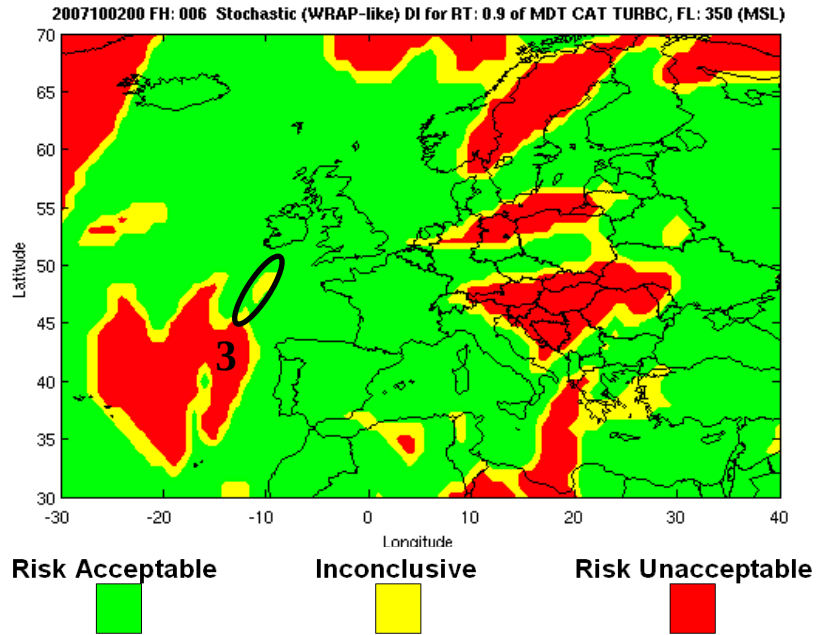


Figure 41. Fall run, FH 06, FL 350 MSL. Simulated WRAP MDT turbulence DI with RT=0.9

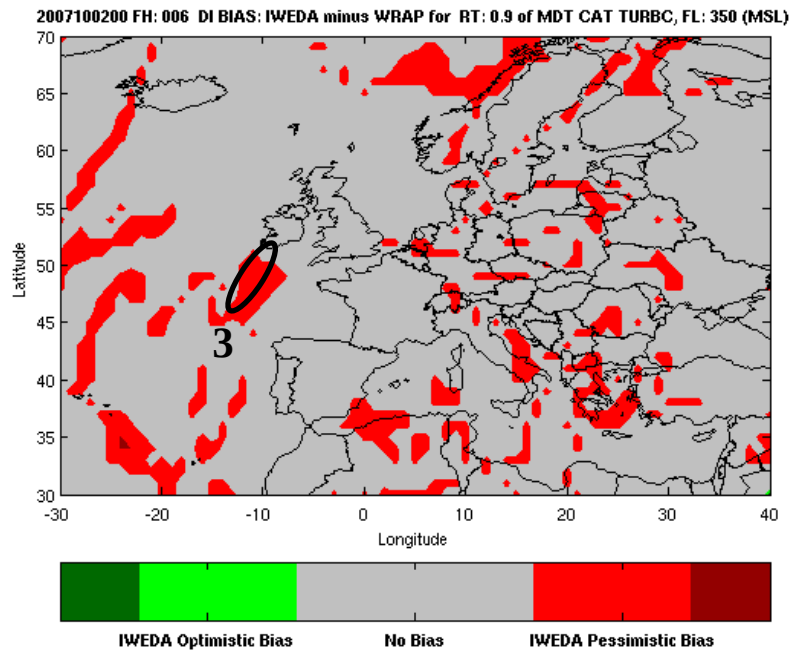


Figure 42. Fall run, FH 06, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.9

b. Fall Case: FH 84

By 84 hours into the forecast period, the ensemble members spread out considerably, revealing considerably more uncertainty. The forecast conveyed by the stochastic product (Figure 43b) is vastly different from the deterministic (Figure 43a).

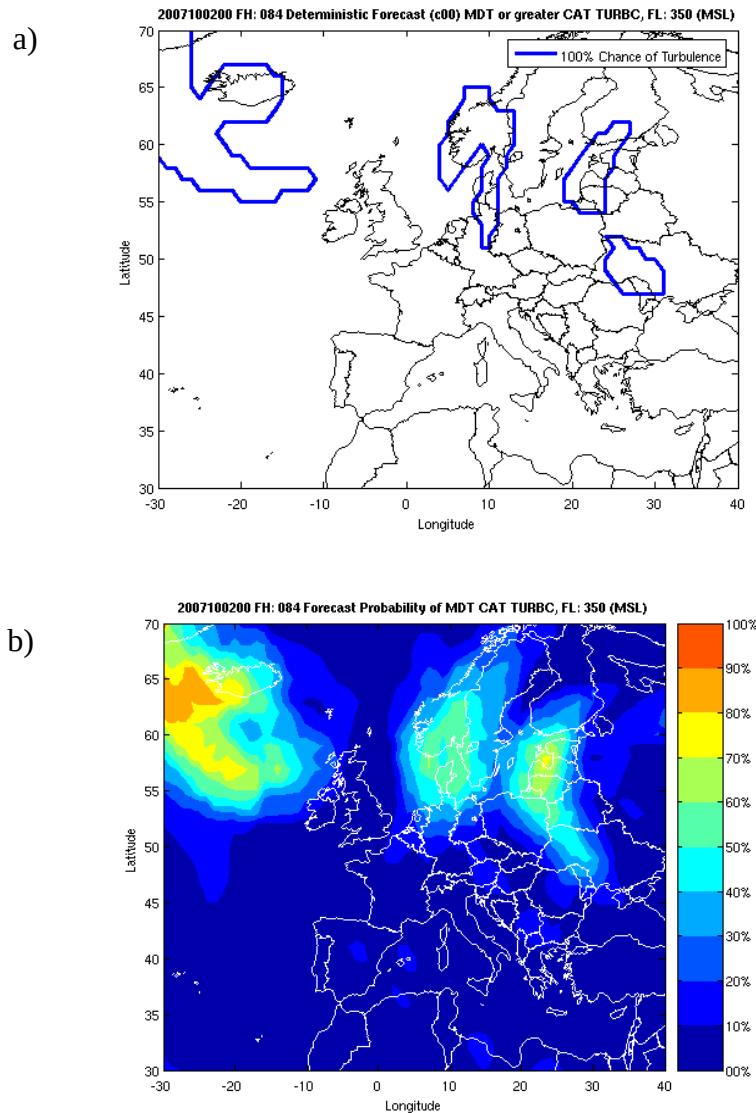


Figure 43. Fall run, FH 84, FL 350 MSL. a) Deterministic MDT turbulence forecast
b) Stochastic MDT turbulence forecast using FP

As time evolved in the forecast period, the differences between IWEDA and WRAP became larger (Figure 44). A two decision input bias exists in area 1 (Figure 44). IWEDA DI deemed unacceptable risk (Figure 45) whereas WRAP DI deemed acceptable risk (Figure 46). Hence, a medium risk tolerant user was given a no-go recommendation from IWEDA, but using the stochastic tool WRAP, the user could still operate in area 1.

In area 2, IWEDA forecasted green, but WRAP forecasted amber, a one decision input bias, shown by the optimistic bias (Figure 44). Using IWEDA, the medium risk tolerant user would be given a false sense of confidence that he/she can afford any possible impacts from MDT turbulence as opposed to using WRAP, which would communicate the possibility of intolerable impacts.

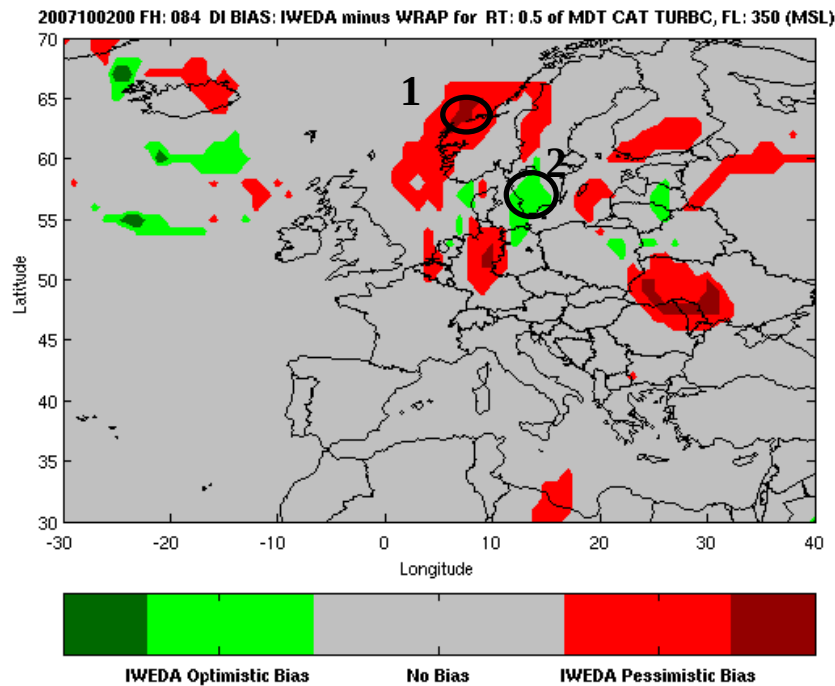


Figure 44. Fall run, FH 84, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.5

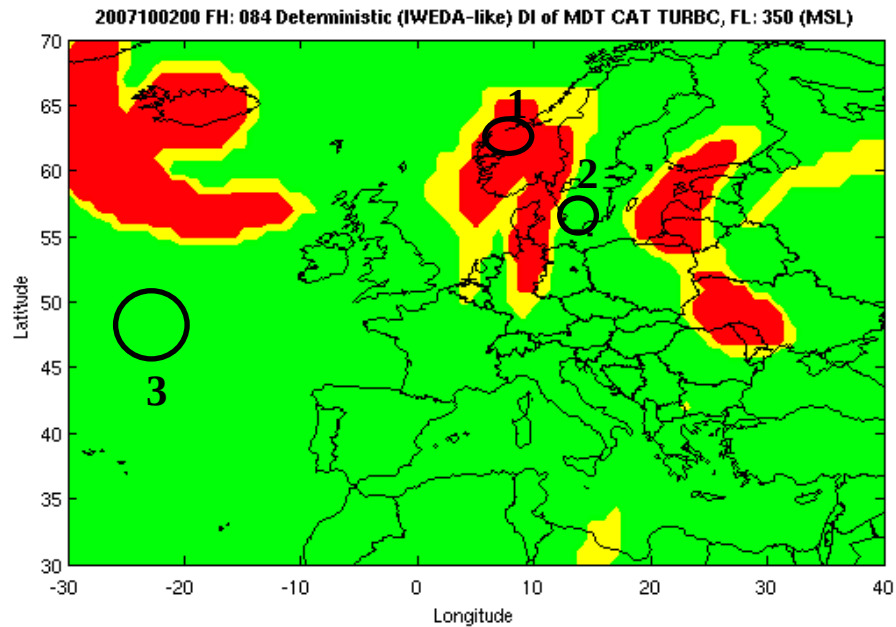


Figure 45. Fall run, FH 84, FL 350 MSL. Simulated IWEDA MDT turbulence DI

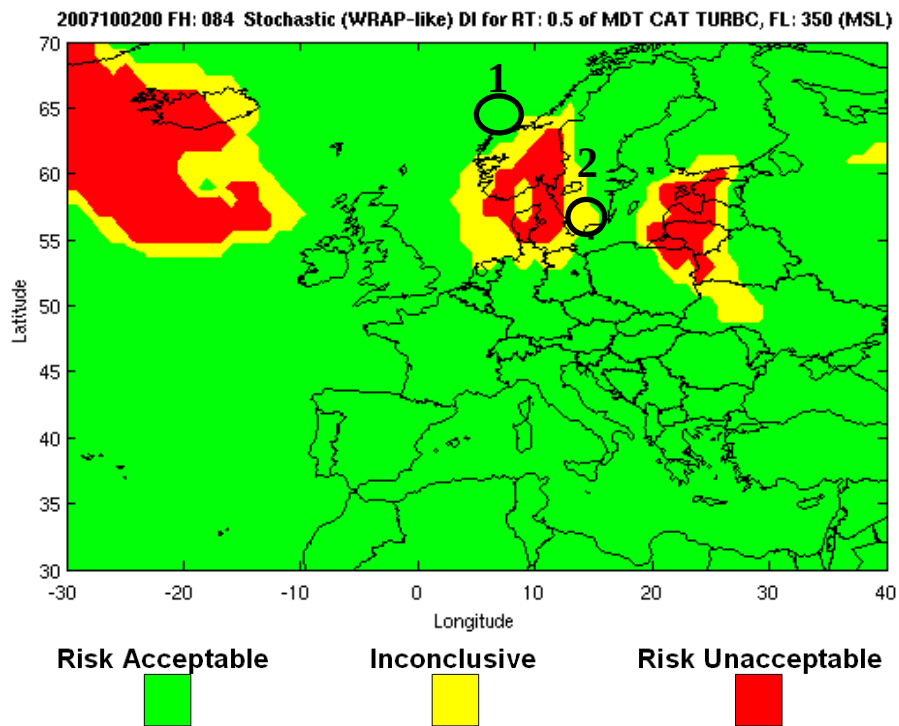


Figure 46. Fall run, FH 84, FL 350 MSL. Simulated WRAP MDT turbulence DI with RT=0.5

Similarly to FH 06, the IWEDA DI bias was all optimistic with WRAP's $RT=0.1$, but the differences are more stark (Figure 48). IWEDA was green in area 3 (Figure 45), while WRAP communicated unacceptable risk (red) according to the user's RT for the same area (Figure 47).

Using a stochastic stop-light DI tool in combination with the FP product can help the education process. To an uneducated user, the extent of unacceptable risk in

Figure 47) infers an abundance of MDT or greater turbulence. This user can look at the FP and discover many areas of the FP are highly uncertain. Using the simple concept $FP > RT$, the user can understand the unacceptable risk areas.

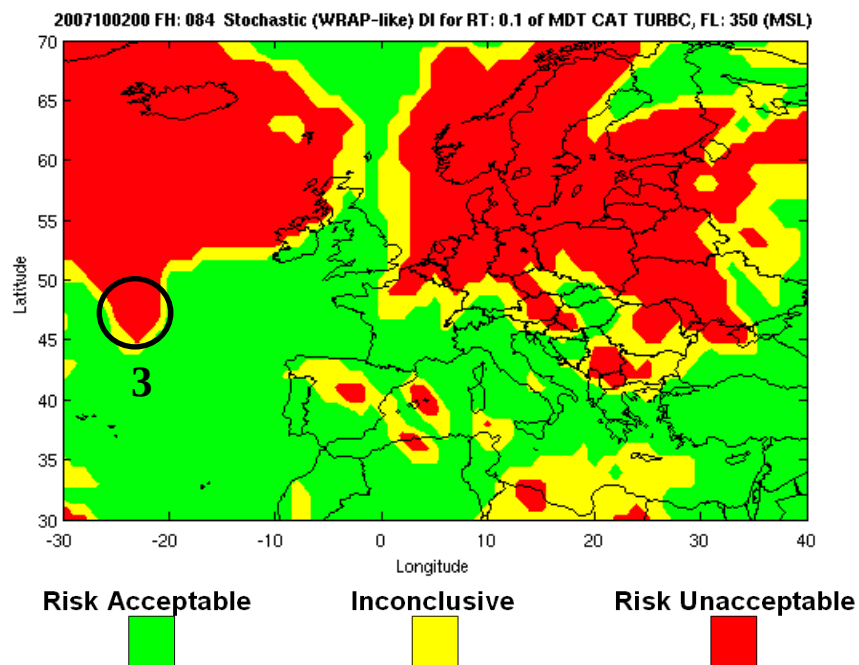


Figure 47. Fall run, FH 84, FL 350 MSL. Simulated WRAP MDT turbulence DI with $RT=0.1$

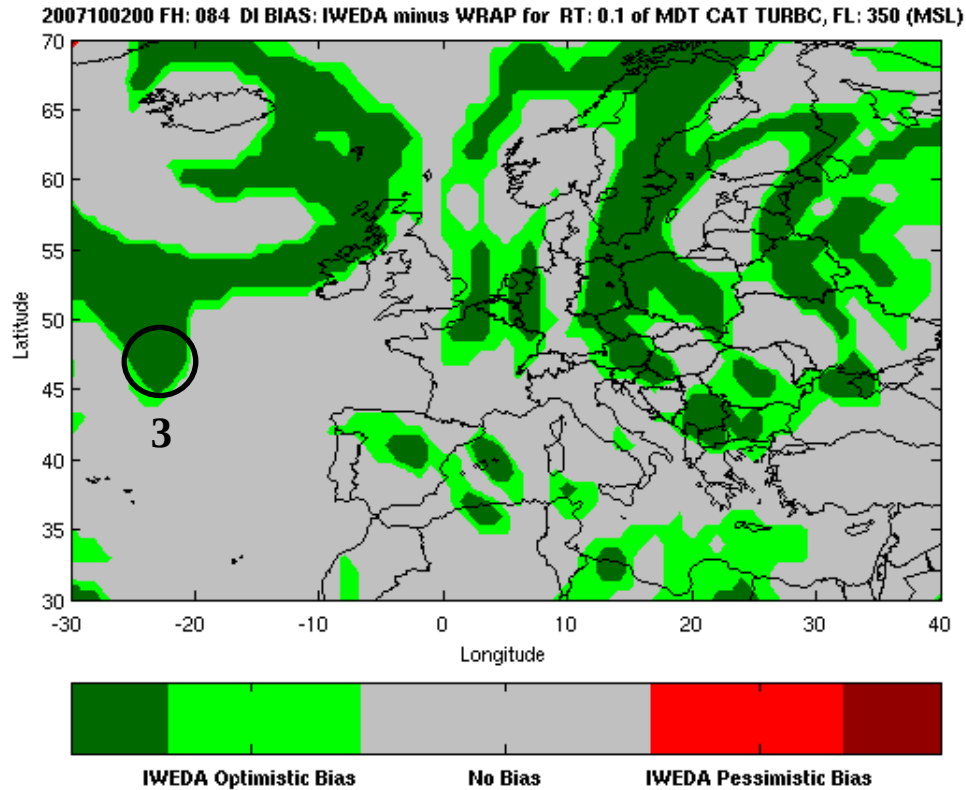


Figure 48. Fall run, FH 84, FL 350 MSL. DI Bias: IWEDA minus WRAP chart.
WRAP RT=0.1.

Again, the DI for WRAP changes vastly when the RT=0.9. According to the WRAP product, all users can operate across the entire AOR (Figure 49) while IWEDA communicates unacceptable risk across the northern portions of the AOR (Figure 45). IWEDA pessimistic bias is evident in these regions (Figure 50). Similar arguments from the preceding IWEDA-WRAP comparison apply to the winter case (Appendix B).

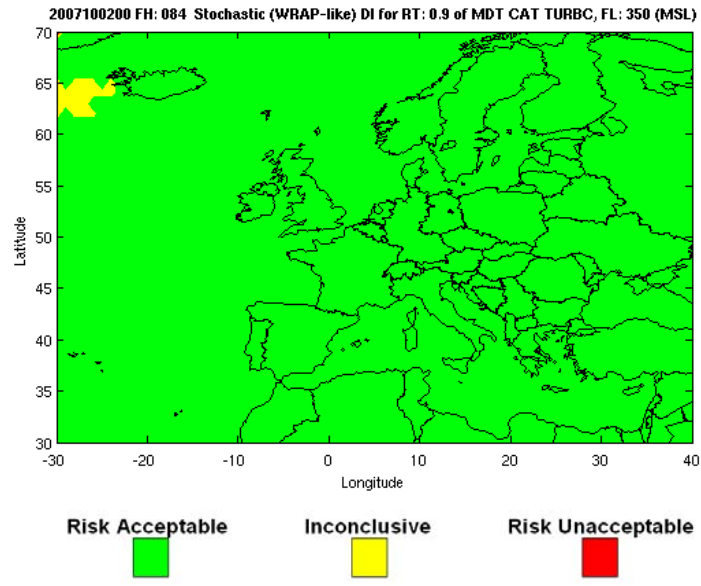


Figure 49. Fall run, FH 84, FL 350 MSL. Simulated WRAP MDT turbulence DI with RT=0.9

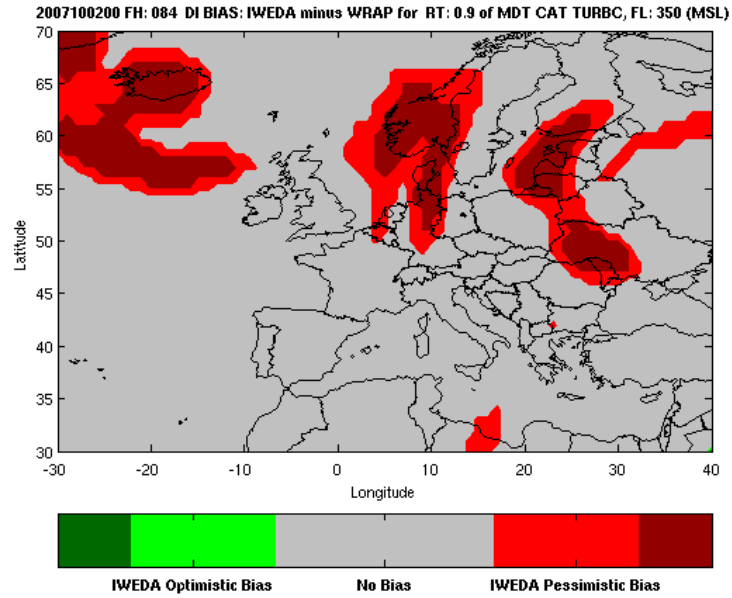


Figure 50. Fall run, FH 84, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.9.

3. Simulated WRAP versus RSCH Results

Since WRAP and the RSCH use the same RT and FP values, the maximum difference is only one decision input level. It is not possible for WRAP to forecast green and RSCH to forecast red and visa-versa. Differences in DI stem from the handling of inconclusive (amber) regions, as described in the methodology. WRAP compares just the FP for exceeding the marginal threshold with the RT while the RSCH method considers the uncertainty of FP (ambiguity) and RT by comparing FP and RT confidence intervals to determine the decision input. To aid in comparing WRAP and RSCH results, the worst case potential error in FP is plotted, which is a way to communicate ambiguity. The default CI is 90%.

a. Fall Case: FH 06

The forecast evolution is only six hours; low uncertainty still exists. When compared to the FP product (Figure 51a), worst case potential error (PE) areas (Figure 51b) are restricted to the sharp gradients separating the zero and 100% probability. Recall that PE decreases toward zero and as FP nears extremes (zero and 100 percent), so there is little to no ambiguity over much of the AOR. The most PE (%) is found in the 40-60 where uncertainty is at its maximum. For example, point A has FP of 50% where the PE is 14%. Hence, the CI for FP at this point can range 36-64%.

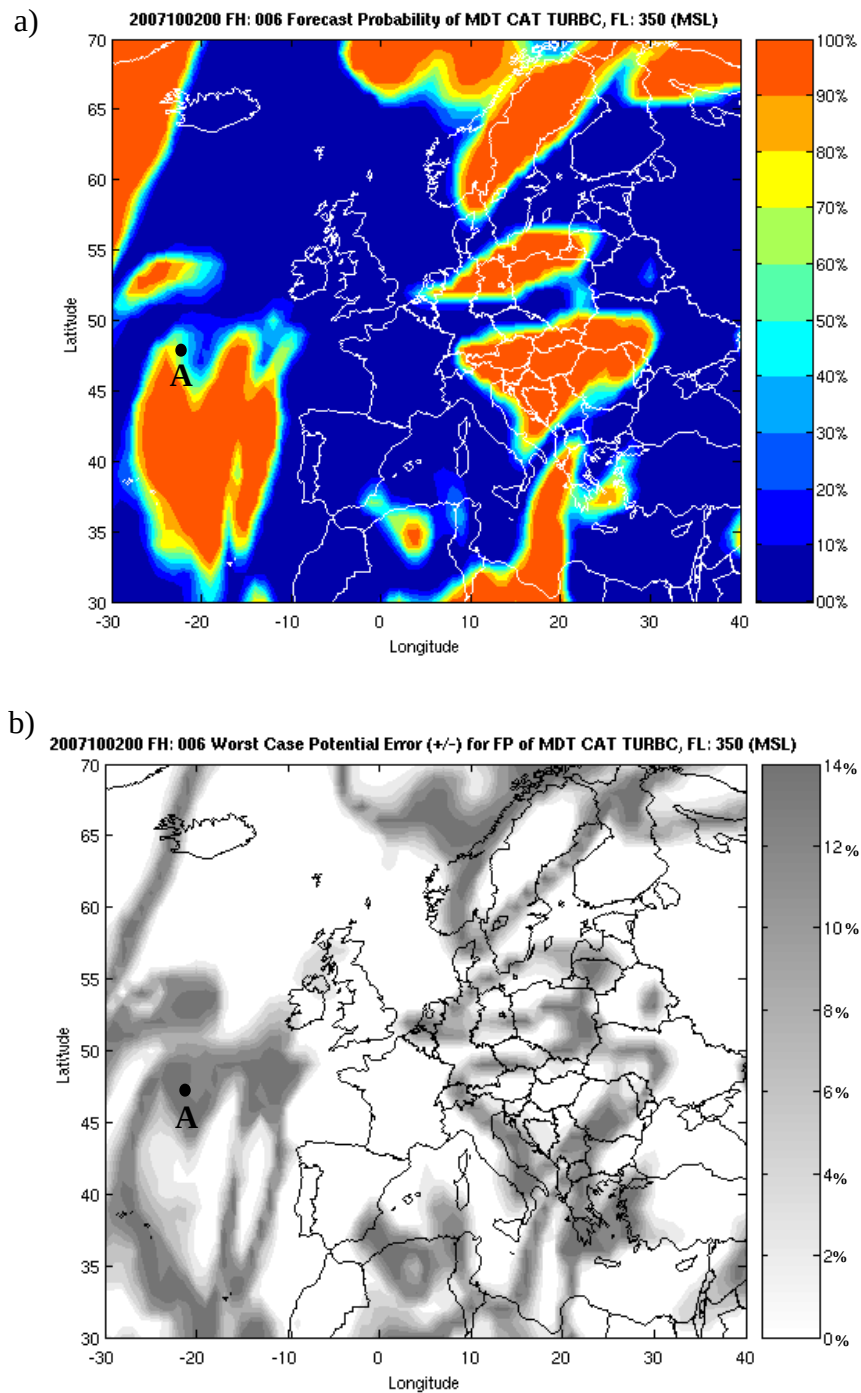


Figure 51. Fall run, FH 06, FL 350 MSL. a) FP of MDT turbulence b) Maximum (Worst Case) Potential Error for FP of MDT turbulence

The case in Figure 52 displays how the ambiguity and RT intervals used in RSCH can alter the decision input to the user. In area 1, WRAP forecasted green (Figure 53a), but RSCH forecasted amber (Figure 53b). Using WRAP DI, a user would appear to have affordable impacts from MDT turbulence in area 2, but using the RSCH, the DI is inconclusive.

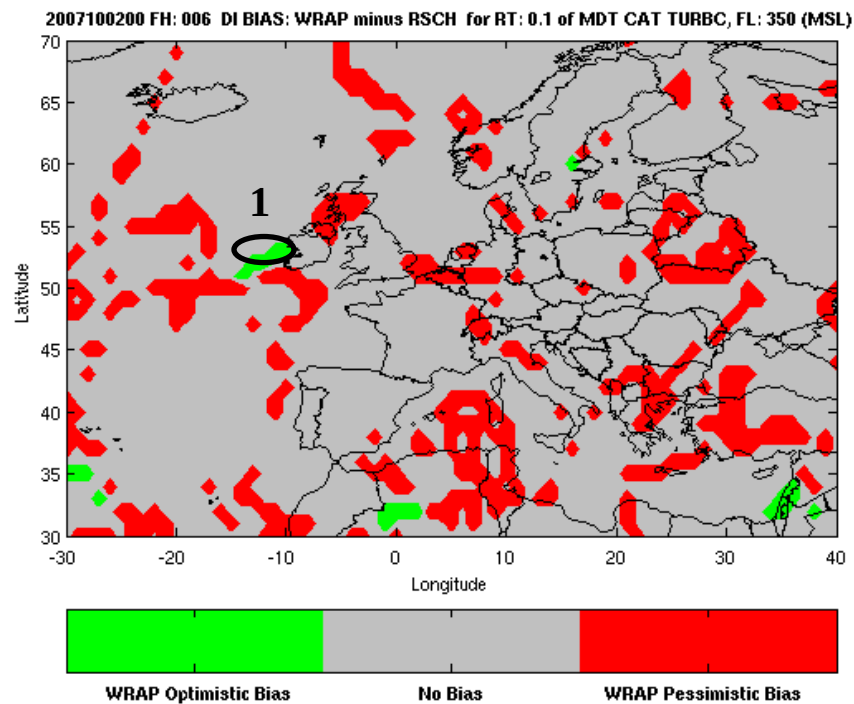


Figure 52. Fall, FH 06, FL 350 MSL, RT=0.1. DI Bias: WRAP minus RSCH chart.

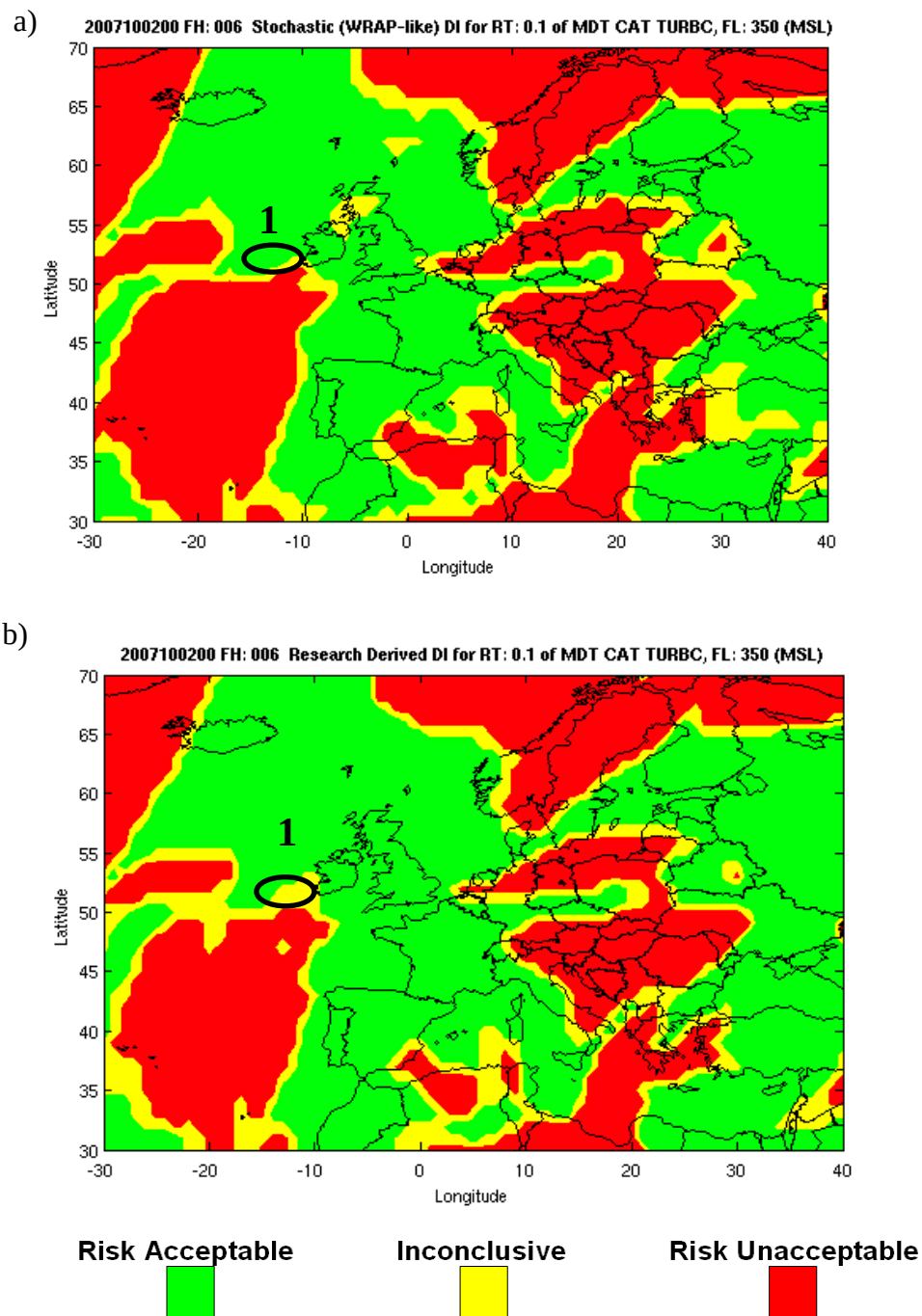


Figure 53. Fall run, FH 06, FL 350 MSL, RT=0.1 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI

In the case displayed in **Figure 54**, many of the differences (i.e. areas 1 and 2) occur because WRAP deemed areas inconclusive (**Figure 55a**) while RSCH deemed the same areas acceptable risk (**Figure 55b**). The RT=0.5, so areas greater than 50% risk of exceeding the marginal thresholds for the WRAP product are deemed inconclusive.

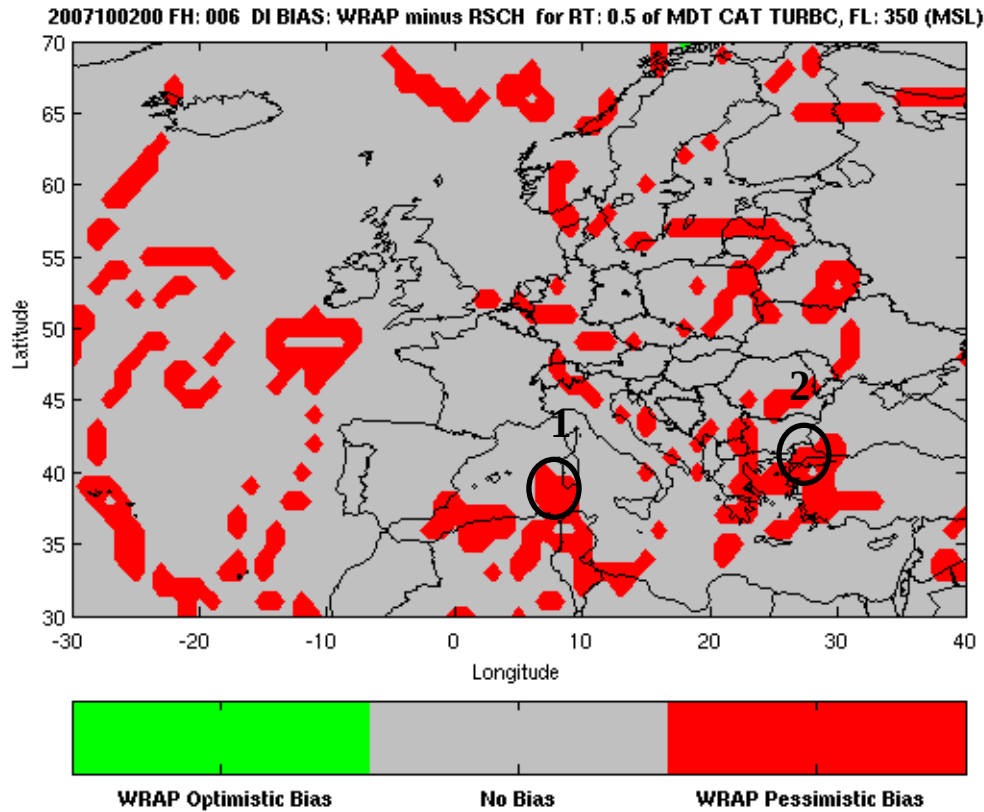


Figure 54. Fall, FH 06, FL 350 MSL, RT=0.5. DI Bias: WRAP minus RSCH chart.

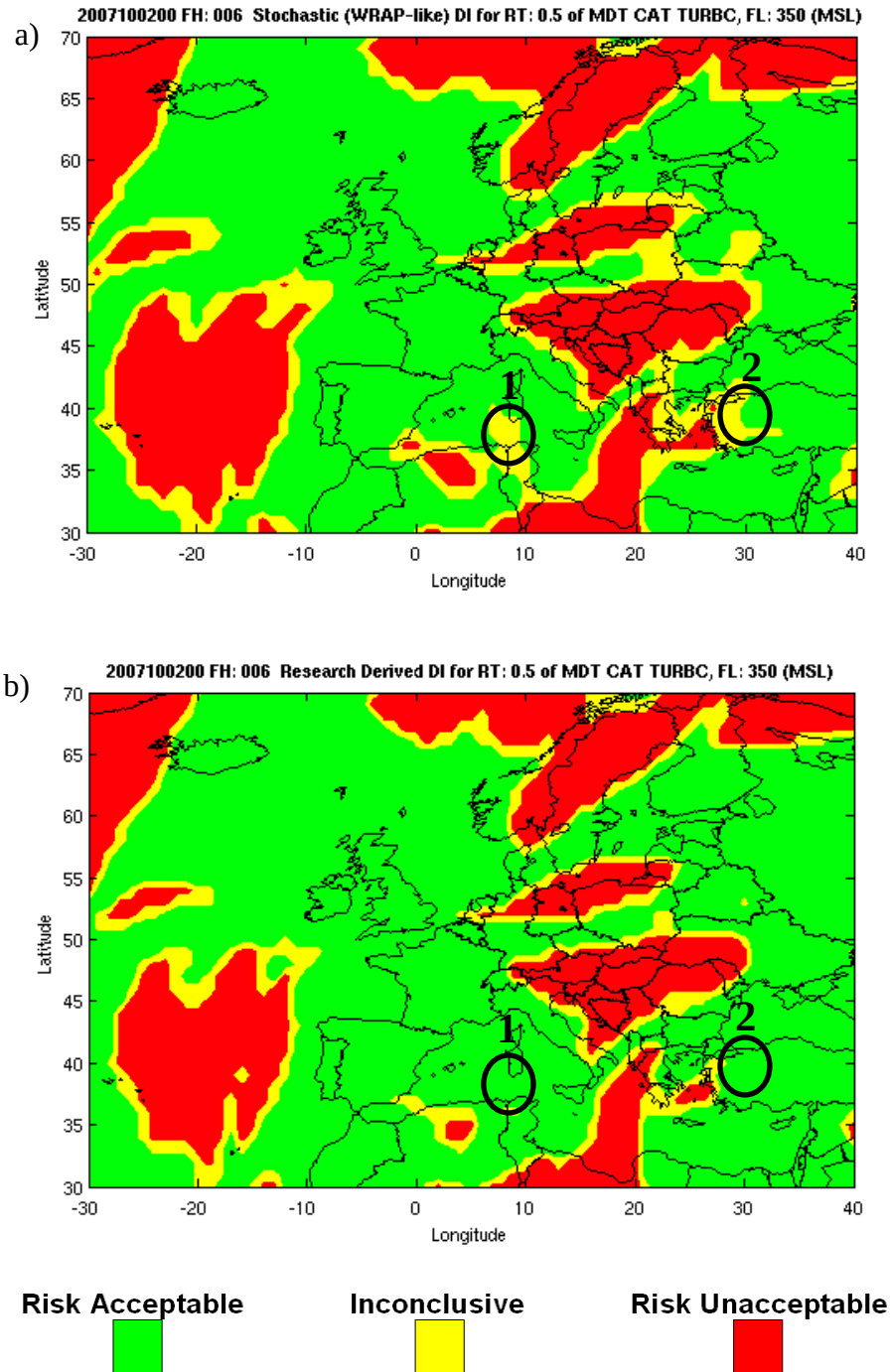


Figure 55. Fall run, FH 06, FL 350 MSL, RT=0.5 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI

If RSCH RT uncertainty interval and/or the ambiguity CI was higher, then areas 1-3 may be deemed inconclusive as well. The differences described may be a direct consequence of the assumed RT uncertainty interval and ambiguity CI. Further research is needed in the realm of decision science to objectively determined and test the RT interval. To explore the extent of the affect of higher ambiguity CIs on the WRAP bias, the researcher plotted the RSCH DI and DI Bias plots using CI=0.95 and 0.99 (Figure 56). The extent of amber regions in areas 1-3 increased slightly by incorporating the higher CI (Figure 56a,c), but the majority of pessimistic bias still exists (Figure 56b,d).

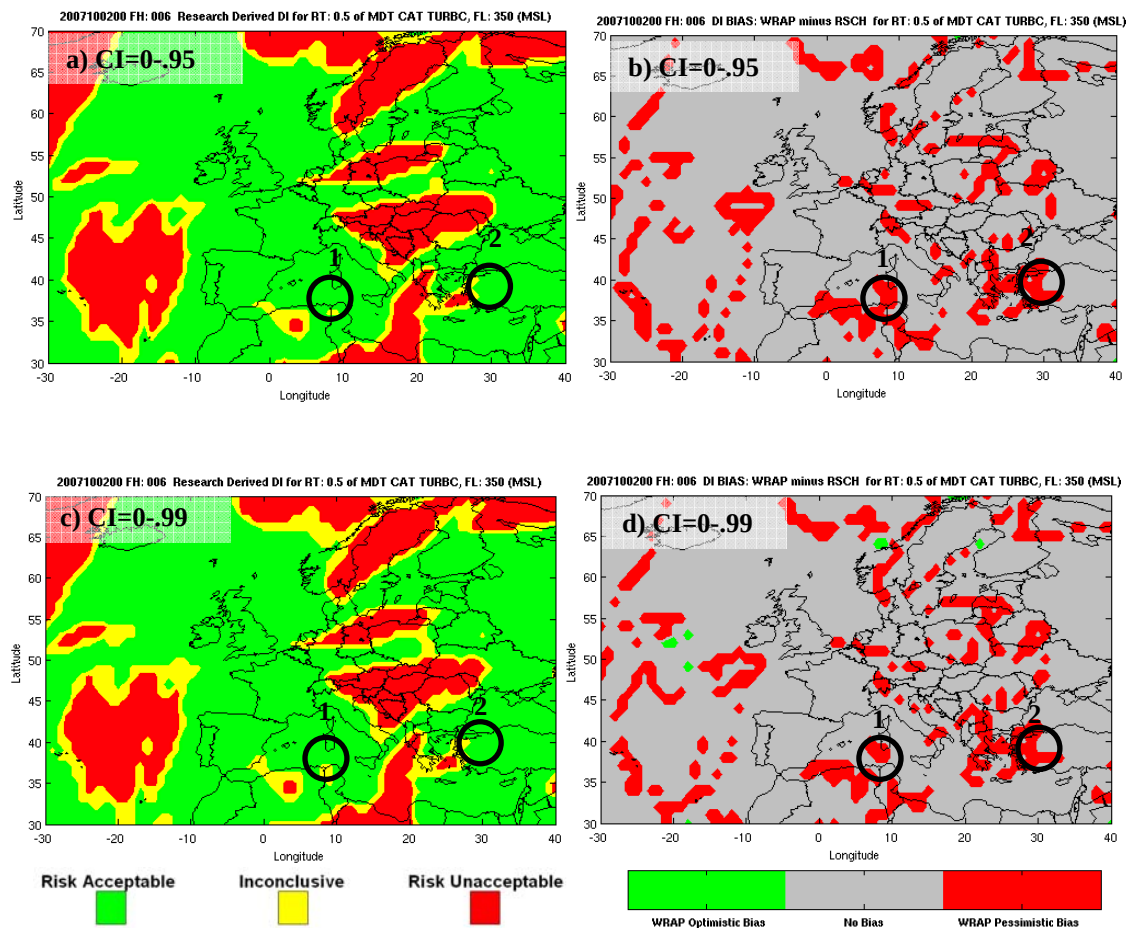


Figure 56. Fall, FH 06, FL 350 MSL, RT=0.5. CI=0.95: a) RSCH DI b) DI Bias: WRAP minus RSCH chart; CI=0.99: c) RSCH DI d) DI Bias: WRAP minus RSCH chart.

In the past three cases, WRAP DI has consistently shown a pessimistic bias trend toward the RSCH DI. To explain, only three combinations of DI can constitute WRAP-RSCH biases (since WRAP and RSCH can only differ by one DI). The first two combinations comprise the “type-one differences,” which arise from the inclusion of ambiguity and RT uncertainty intervals in the RSCH DI. WRAP forecasted red (or green), but RSCH forecasted amber. The last combination of DI comprises the “type-two difference,” which arise from the use of WRAP marginal threshold. WRAP forecasted amber, but RSCH forecasted green. Biases such as area 1 in Figure 57 and areas 1 and 2 highlighted in Figure 54 are examples of the type-two difference.

The DI combination, WRAP =amber and RSCH = red, for the type-two difference is not possible. For example, if the $RT=0.3$, the FP of $TI=3.5$ (marginal threshold) must exceed 30% in order for WRAP to forecast amber. If the FP for exceeding the marginal threshold is greater than 30%, then the FP for exceeding the mission threshold must be less than 30%. Since, the RSCH DI uses the same mission threshold, the worst possible decision input is amber, not red.

The type-one difference has both an optimistic and pessimistic bias combination. Since it is not possible for WRAP=amber and RSCH=red, the type-two difference only has a pessimistic bias combination. Hence, the WRAP DI has an overall pessimistic bias towards the RSCH DI. If the marginal threshold was slighter higher, say 4.0 instead of 3.5, the probability of exceeding the marginal threshold would decrease and pessimistic bias would not be so prevalent. Consequently, if WRAP used no marginal threshold, the amount of optimistic and pessimistic biases should be almost equal, because the differences would be driven purely by the type-one differences. Therefore, by not considering the uncertainty of the FP and RT and using a marginal threshold, WRAP may have shifted the DI to a degree worse than it might actually be. Users may be unnecessarily protecting and/or hesitating when only portion of uncertainty information is included into the DI.

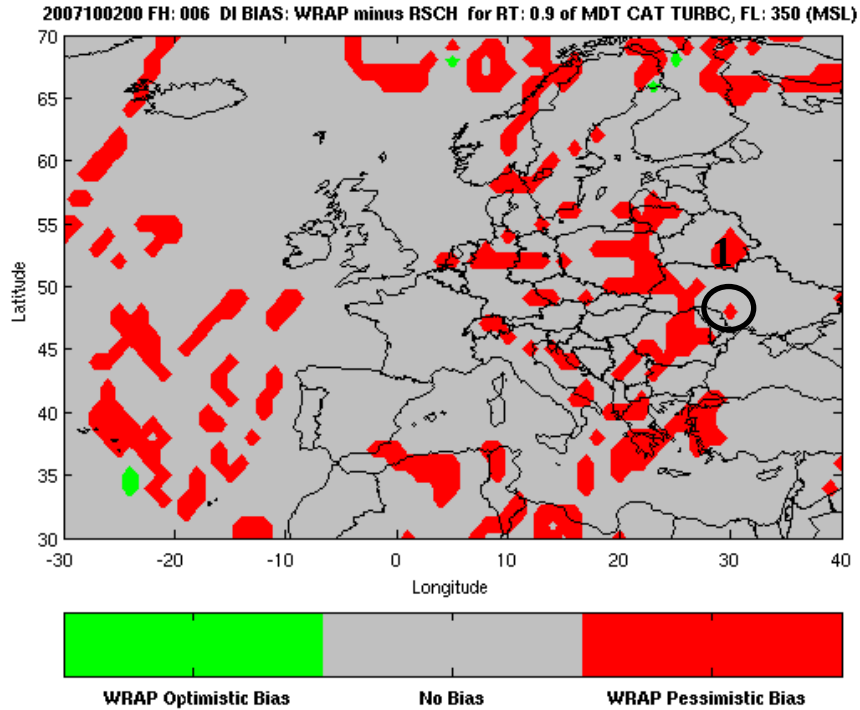


Figure 57. Fall, FH 06, FL 350 MSL, RT=0.9. DI Bias: WRAP minus RSCH chart.

b. Fall Case: FH 84

Much uncertainty evolved as time progressed to FH 84, reflected by the larger areas of FP 40-60% (Figure 58a) and the dense areas of high PE (Figure 58b). In response, the RSCH DI calculated more areas of amber and larger areas of difference between WRAP and RSCH DI exist.

The RT=0.1 case in Figure 59 is an example to WRAP's type-one difference. In area 1, the FP is 10-20%. According to WRAP, the DI is unacceptable risk (Figure 60a). The maximum PE in area 1 is 8% which makes the ambiguity interval 2-18%. The RT uncertainty interval is 8.3-11.7%. Since these two intervals overlap, the RSCH DI deemed the area inconclusive (Figure 60b).

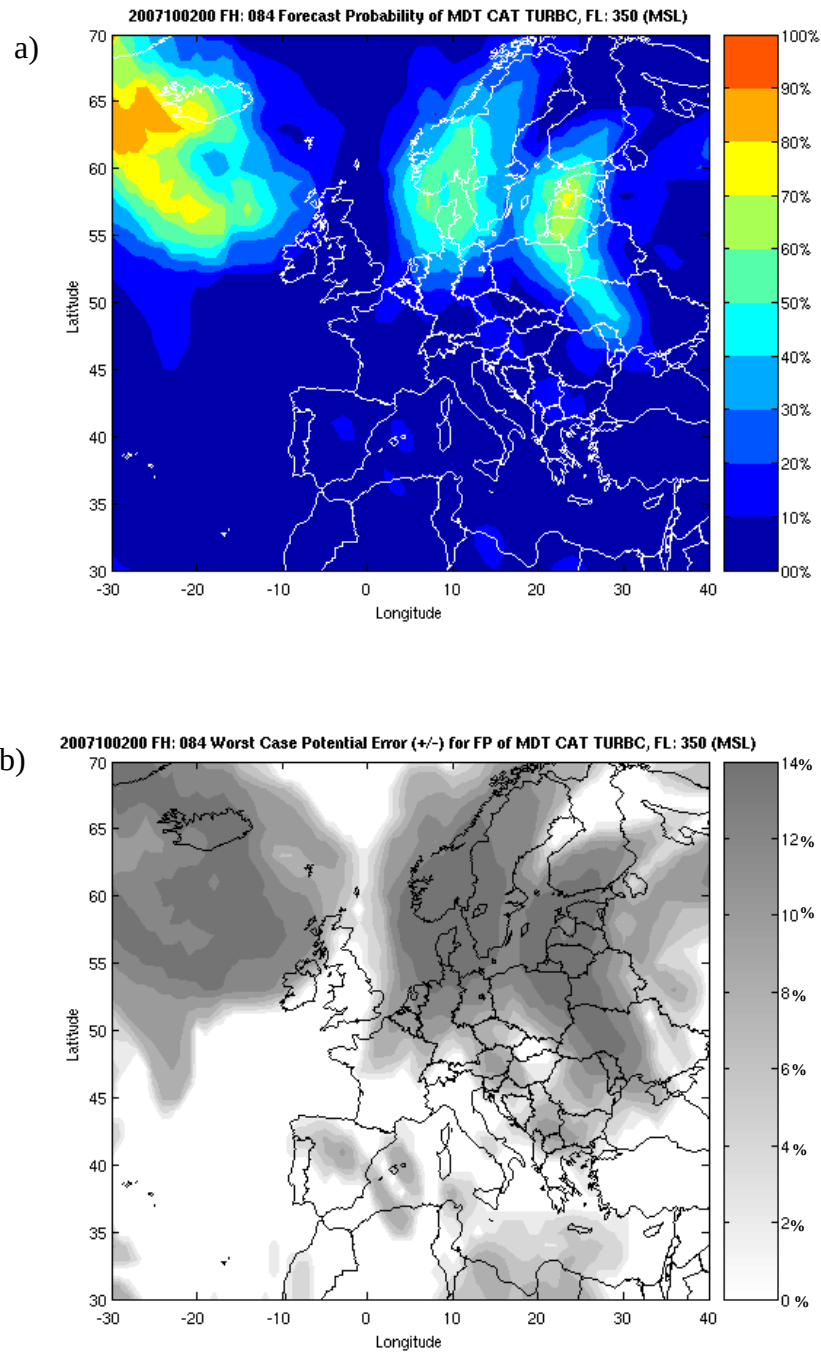


Figure 58. Fall run, FH 84, FL 350 MSL. a) FP of MDT turbulence b) Maximum (Worst Case) Potential Error for FP of MDT turbulence

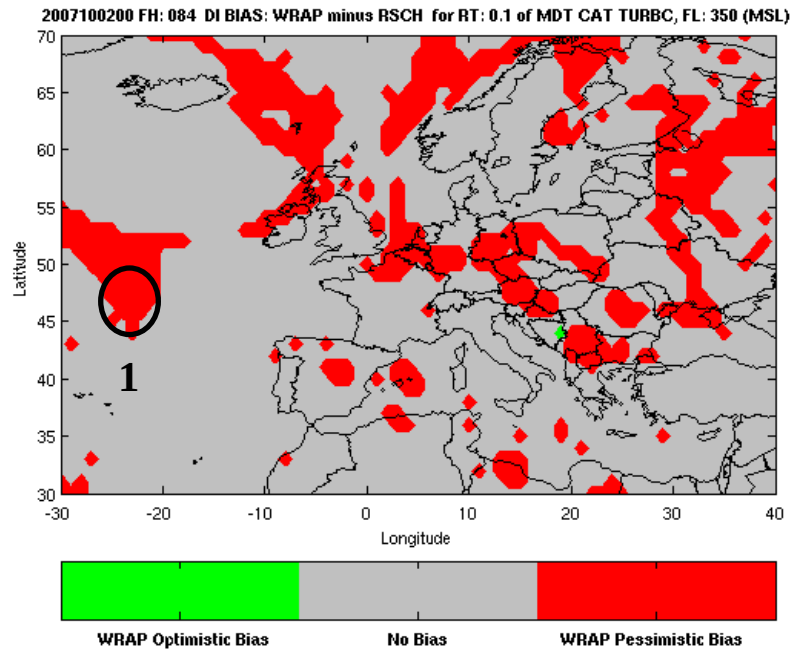


Figure 59. Fall, FH 84, FL 350 MSL, RT=0.1. DI Bias: WRAP minus RSCH chart.

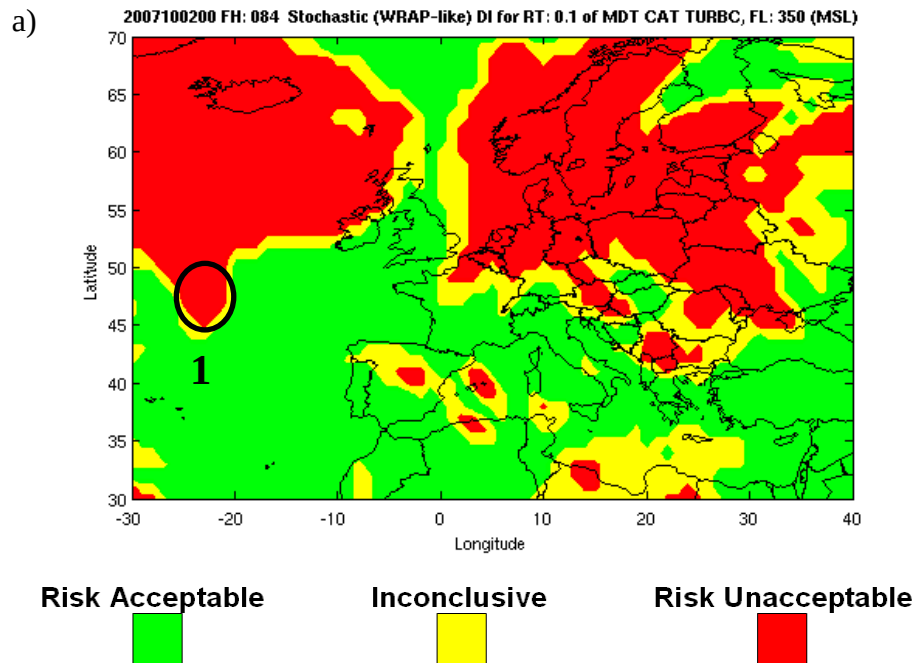


Figure 60. Fall run, FH 84, FL 350 MSL, RT=0.1 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI

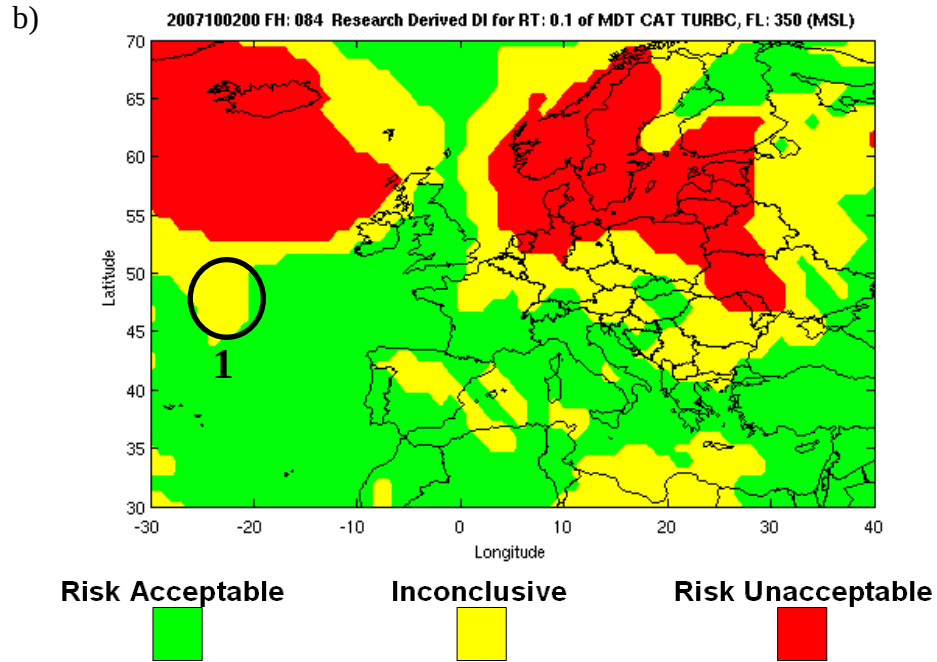


Figure 60 continued.

The case in Figure 61 is a rare balance of optimistic and pessimistic biases because the majority of the differences are type-one differences. In area 1, the maximum FP for MDT turbulence in area 1 is 60%, which WRAP deemed as unacceptable risk. The overlap of the ambiguity intervals (48-72%) the RT uncertainty intervals (45-55%) make the RSCH DI inconclusive. In this instance, WRAP may have shut down operations unnecessarily. In area 2, the maximum FP for MDT is just below 50%, hence WRAP deemed the area acceptable risk. The overlap of the ambiguity intervals (35-63) with the RT uncertainty intervals make the RSCH DI inconclusive once again. In this instance, WRAP may have left the user vulnerable to a possible threat because the DI is too close to call.

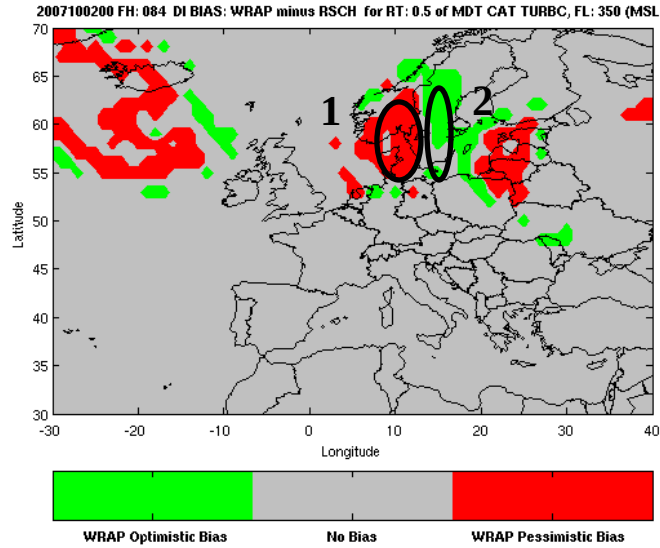


Figure 61. Fall, FH 84, FL 350 MSL, RT=0.5. DI Bias: WRAP minus RSCH chart.

Minimal differences exists between the WRAP and RSCH DI products in the case displayed in Figure 64. The shape of positive definite PDF drives this difference. In this highly uncertain situation, the positive definite TI PDF is spread out (for example, Figure 62). From Figure 62, with RT=0.9 and FP for exceeding the marginal threshold=90.5%, WRAP forecasts amber. RSCH, with its overlapping ambiguity and RT uncertainty intervals, also forecasts amber.

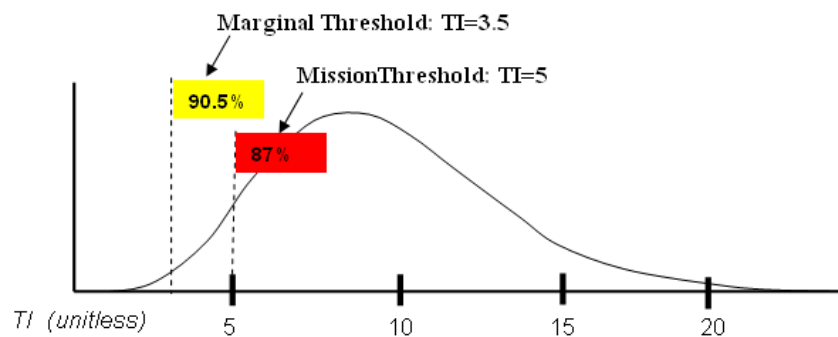


Figure 62. Positive Definite TI PDF with high probability of high TI

So the majority of the DI for both WRAP and RSCH are equal. Little value is added by incorporating ambiguity and RT uncertainty information when the forecast has high uncertainty and the RT is also high. In the case associated with Figure 59, the majority of the biases are in areas of FP ranging from 10-30%. The FP for mission threshold > RT, so WRAP forecasted red. The overlapping intervals for RSCH cause the DI=amber (Figure 63). The differences are much more prevalent at the low RT. Hence, in highly uncertainty weather situations, users with low RT such as those accomplishing training would benefit more using the RSCH DI than users with a high RT, such as those accomplishing mission critical operations. Similar arguments from the preceding WRAP-RSCH comparison apply to the winter case (Appendix C).

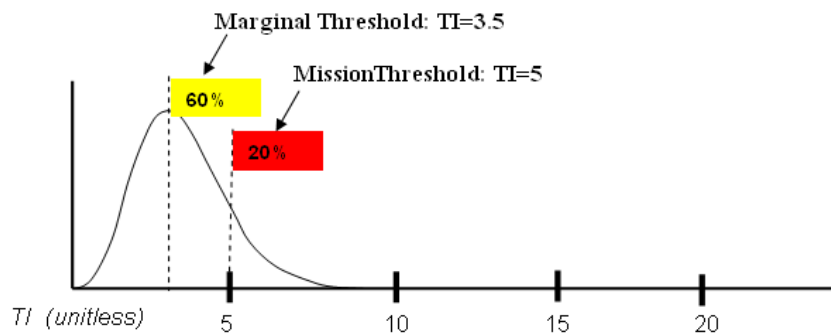


Figure 63. Positive Definite TI PDF with high probability of low TI

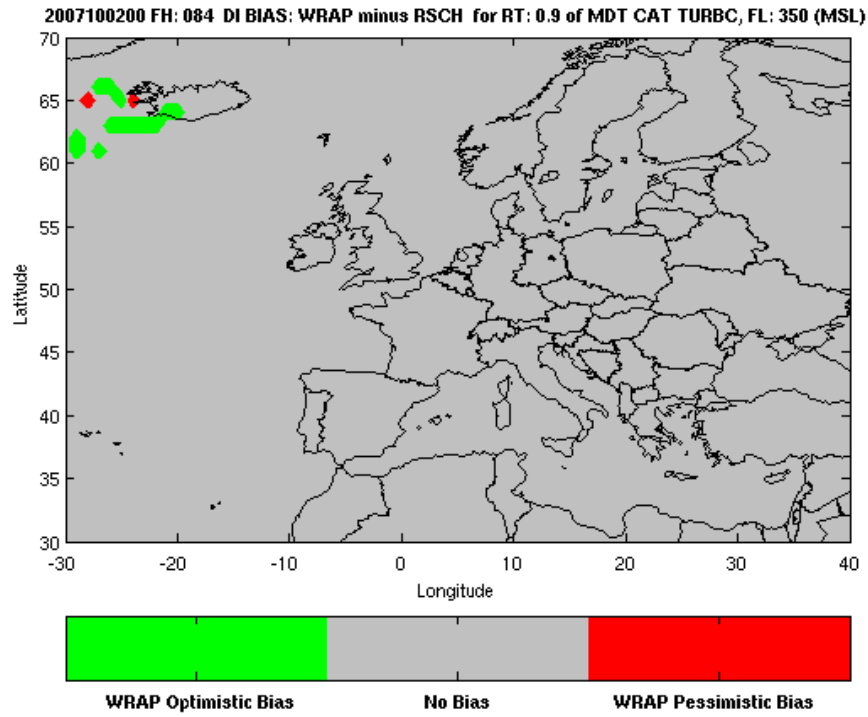


Figure 64. Fall, FH 84, FL 350 MSL, RT=0.9. DI Bias: WRAP minus RSCH chart.

V. CONCLUSIONS

A. FINAL REMARKS

The three main objectives of the thesis were to: 1) furnish a process to account for uncertainty in both the ensemble data and user risk tolerance for the decision input, 2) create an effective visualization to for varying levels of risk intolerance, mission thresholds for turbulence, flight levels, and forecast hours that is educational and practical for the user 3) demonstrate potential value added to the user when using the research derived decision input product through a comparison of current decision input products.

Chapter II described the background to atmospheric prediction, ensemble forecasting (including the ambiguity), weather risk management, and the calculation/communication of forecast uncertainty (from today's decision input tools). Using the background as a foundation, Chapter III laid out a three part process to calculate and produce visualization products for calibrated forecast probability, worst case potential error, and the optimized decision input. Through the Taijitu method, the forecast probability was calibrated and the worst case potential error was calculated. The decision input used objective estimates of the ambiguity intervals and subjective estimates of the risk tolerance uncertainty intervals. A prototype GUI was constructed so a non-educated user can easily access these products according to his/her unique RT, mission threshold of turbulence, flight level, and forecast hour. Chapter IV fulfilled the last objective with two cases which compared current decision input products with the research results using scenarios of different risk tolerances and forecast hours.

Using the ninety percent confidence interval for ambiguity communicated a large amount of inconclusiveness but still yielded definitive decision inputs. The maximum bounds for the ambiguity increased with increased CIs, but remained nearly constant through the forecast period, since the constant error variance of the ensemble mean appears to have dominated the ambiguity. As forecast period evolved, the average error in the ensemble spread varied due to weak ensemble dispersion and excessive initial

conditions perturbations. In response, the calibration curve reversed due to the change in average ensemble spread error from >1 to <1 . Incorporating uncertainty of FP and RT into the DI process can often alter the decision input. The differences between of the different decision tools show potential value added to the user when FP and RT uncertainty information is incorporated.

The differences shown in the comparisons of the current decision tools showed potential value added to the decision making process when increasing degrees of uncertainty information is included into the DI. IWEDA is a generic decision aid made to give DI to all users with one product but is limited by using only one forecast solution and does not consider RT. WRAP's inclusion of stochastic weather information and RT generates a unique DI which can benefit user in the long term by minimizing their expenses. The results confirmed the IWEDA DI bias is highly dependant on RT and degree of uncertainty. WRAP DI changes with different RT inputs. Increasing the RT gave more acceptable risk DIs because of the users ability to afford potential loss (high C/L) and decreasing the RT gave more unacceptable risk because of the user inability to afford the potential loss (low C/L). Compared with WRAP using a low (high) RT, IWEDA (the deterministic DI), had an optimistic (pessimistic) bias. As the forecast evolution progresses, the uncertainty increased and the areas of bias grew.

By comparing the error bars of FP and RT to create the DI, the RSCH product sometimes shifted the DI from WRAP. Overall, WRAP had a pessimistic bias relative to the RSCH DI. Only three combinations of DI can constitute WRAP-RSCH differences (since WRAP and RSCH can only differ by one DI). The first two combinations comprise the "type-one differences," which arise from the inclusion of ambiguity and RT uncertainty intervals in the RSCH DI. Type-one differences consist of both pessimistic and optimistic differences. The last combination of DI comprises the "type-two difference," which arise from the use of WRAP marginal threshold. Type-two difference consists of only a pessimistic difference. The DI combination, WRAP = amber and RSCH = red, for the type-two difference is not possible due to WRAP's marginal threshold. Therefore, an imbalance of pessimistic bias occurred. Increasing the CI for the ambiguity intervals only decreased WRAPs pessimistic bias slightly. Therefore, by not

considering the uncertainty of the FP and RT and using a marginal threshold, WRAP may have shifted the DI to a degree worse than it might actually be, preventing users from taking action. If the differences between WRAP and RSCH were only due to type-one differences, then there would be no overall bias.

WRAP-RSCH differences were few early in the forecast period because of the low uncertainty. As uncertainty increased, so did the ambiguity. The DI differences grew in spatial coverage because of the differences in calculating the inconclusive (amber) regions. In high uncertainty situations, the RSCH methods are potentially valuable in the low to middle RT range (i.e. training and some operational mission) but not high RT range (i.e. mission critical operations) due to the shape of the TI PDF.

The study explored an evolution of forecast honesty: from a deterministic DI that communicates no forecast uncertainty to a stochastic DI that considers forecast uncertainty (FP) and RT to another stochastic DI that considers uncertainty of both FP and RT. The differences in DI can affect the users' actions. The optimized decision input derived from this research incorporates a complete portrait of uncertainty so that the war-fighter can make the most cost effective decision. The inconclusive (amber) DI gives the decision maker the option to go/not-go based on other decision factors that may not be quantifiable like morale, importance of mission, and personnel safety. The research has demonstrated effective communication and integration of optimized decision input into the decision making process. Procedures and results of research can serve as an example to further education and development of stochastic methods in the Air Force and Department of Defense to help improve combat capability.

B. FURTHER RESEARCH AND RECOMMENDATIONS

1. Further Research

Although the Taijitu method introduced in this research uses sound mathematical and statistical methods, a study to verify and validate the method is needed. Additionally, decision science research needs to be accomplished to determine an objective RT uncertainty interval for military operations. Verification of the ensemble data was done for one season, but needs to be done for all four seasons in order to

accurately reflect the breadth of error in the ensemble mean and. To ascertain how the level of understanding of DI product attributes enhance or diminish the usability of the visualization, an objective process should be devised to solicit feedback from various operational users such as pilots. To quantify combat capability improvement, a robust case study using an operational units' data should be accomplished.

2. Recommendations

Results show potential value added if current DoD decision aid products include ambiguity and RT uncertainty. In order to help the transition to the stochastic mindset, user-friendly products need to be made available so that user can easily integrate the information into the decision making process. A possible method is an online GUI capability made available at the strategic level, so regional weather centers (Operational Weather Squadrons) and individual weather units (Combat Weather Teams) can access the stochastic weather information and calculate DI for their user. The resultant GUI product could be a three panel plot showing the calibrated FP, worst-case potential error, and stop-light decision input for one weather parameter. To show multiple parameters, a decision input matrix could be constructed according to the unit's requirements. The matrix would include the mission thresholds, RT for each threshold, FP of exceeding the threshold, and the DI. The highest (worst-case) recommendation will be the single input, overall recommendation for the mission (Table 7). The mission DI matrix could easily be formulated automatically from machine-to-machine inputs.

Mission Decision Input (DI) Matrix				
Mission Threshold	SVR Turbulence	MDT Rime Icing	15 knots Crosswinds	1 Statue Mile Visibility
Risk Tolerance/ Uncertainty Interval	10% ± 1.7%	30% ± 4%	20% ± 3%	40% ± 4.5%
Risk (FP)/ Ambiguity Interval	0% ± 0%	5% ± 0.5%	28% ± 3%	35% ± 4%
DI	Acceptable Risk	Acceptable Risk	Unacceptable Risk	Inconclusive
Overall Mission DI/ Recommendation				Unacceptable Risk / No-go

Table 7. Prototype Mission Decision Input Matrix

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A: RESEARCH DI WITH DIFFERENT CI OF AMBIGUITY

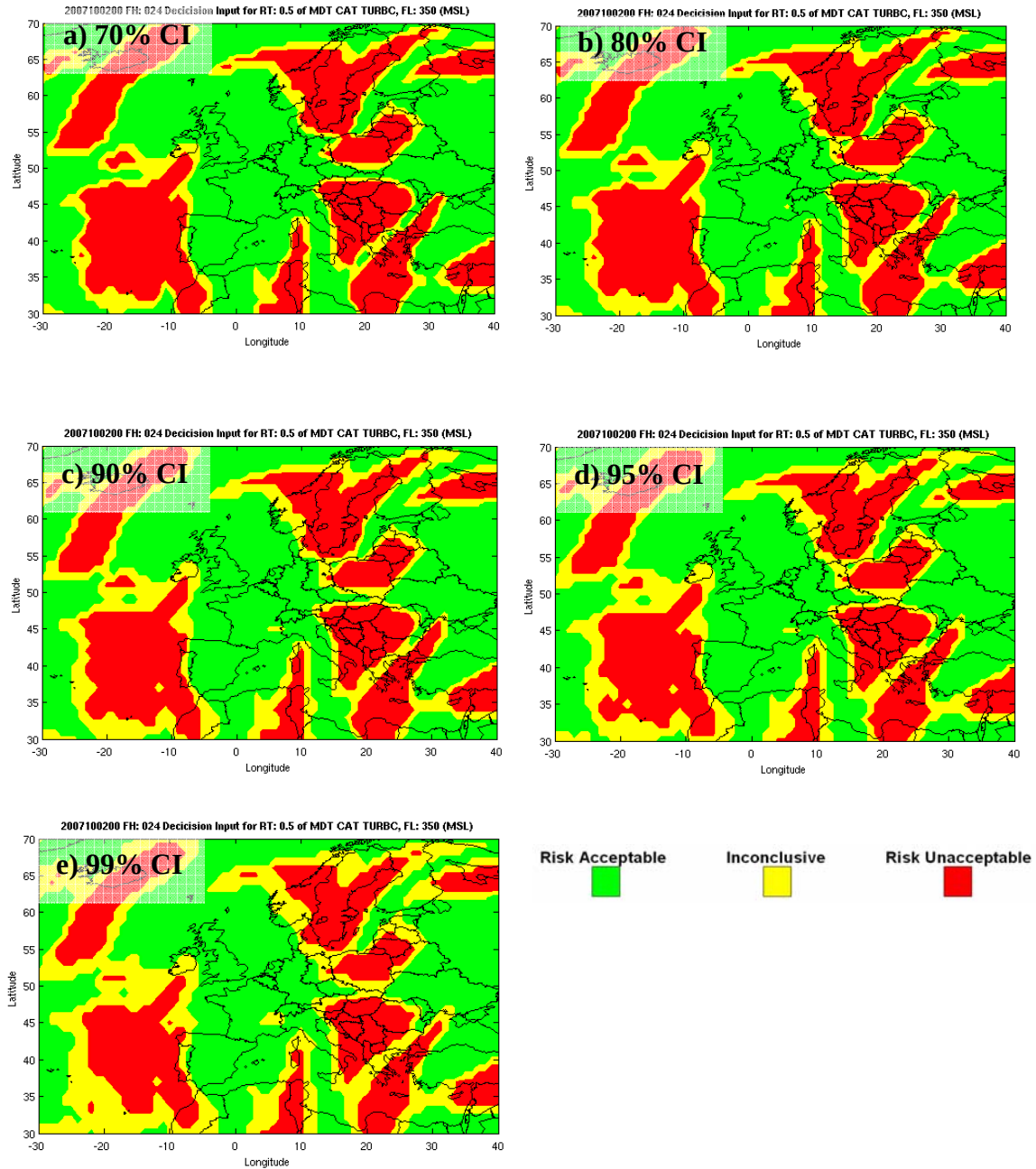


Figure 65. (a-e) RSCH DI for MDT turbulence using a range of CIs; fall run, FH 24, FL 350 (MSL), RT=0.5

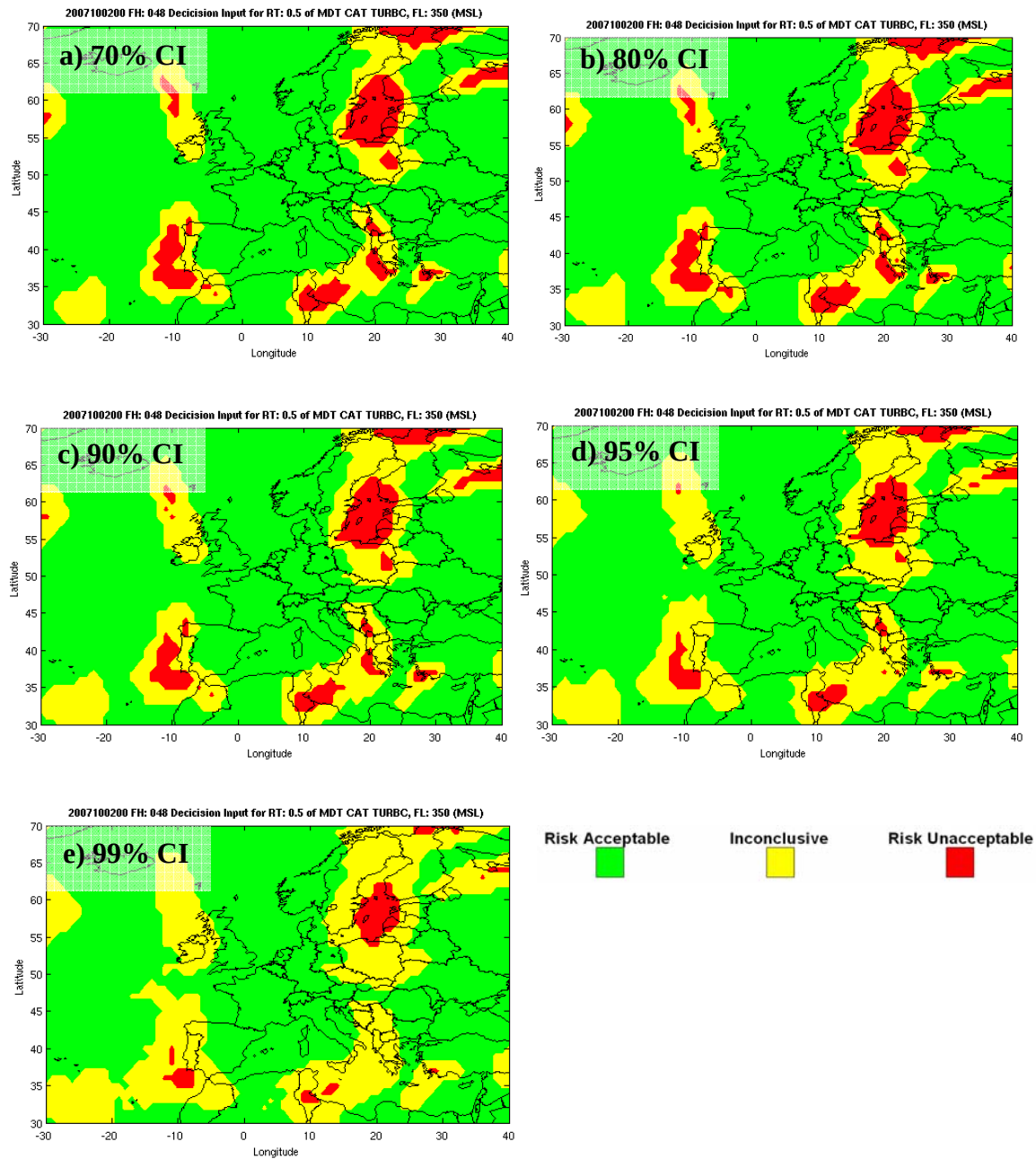


Figure 66. (a-e) RSCH DI for MDT turbulence using a range of CIs; fall run, FH 48, FL 350 (MSL), RT=0.5

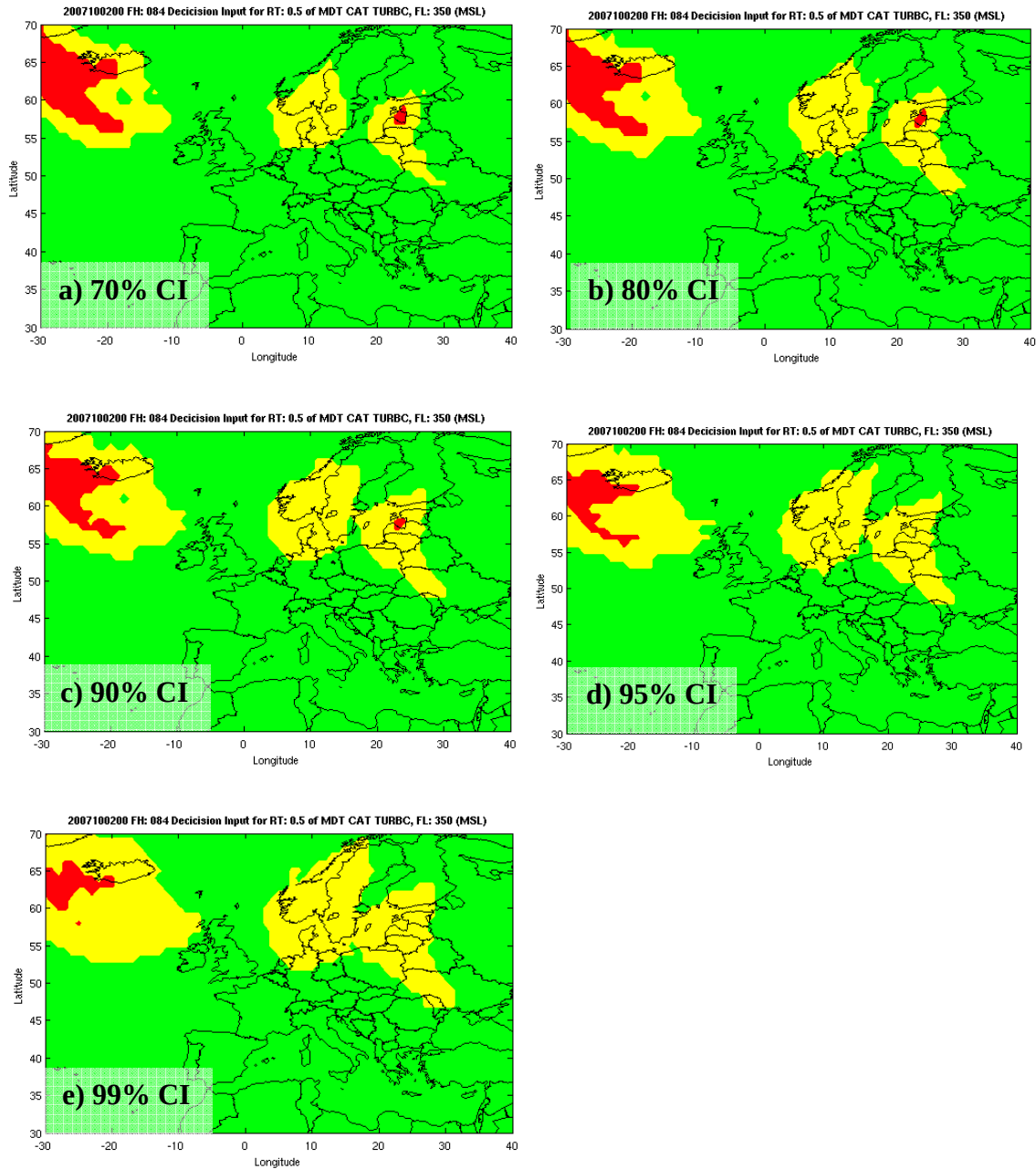


Figure 67. (a-e) RSCH DI for MDT turbulence using a range of CIs; fall run, FH 84, FL 350 (MSL), RT=0.5

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B: WINTER CASE: SIMULATED IWEDA VERSUS SIMULATED WRAP

The following are figures from the simulated IWEDA and WRAP comparisons of January 30, 2008, 00Z (winter) run. For reasoning, see similar cases in Chapter III, section B.2.

A. FORECAST HOUR: 06

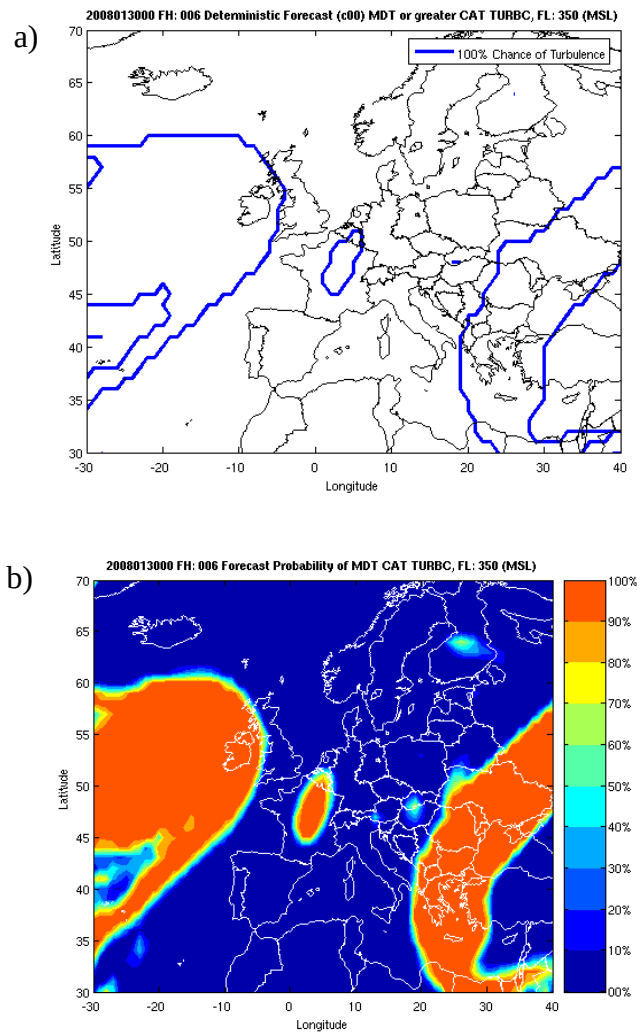


Figure 68. Winter run, FH 06, FL 350 MSL. a) Deterministic MDT turbulence forecast b) Stochastic MDT turbulence forecast using FP

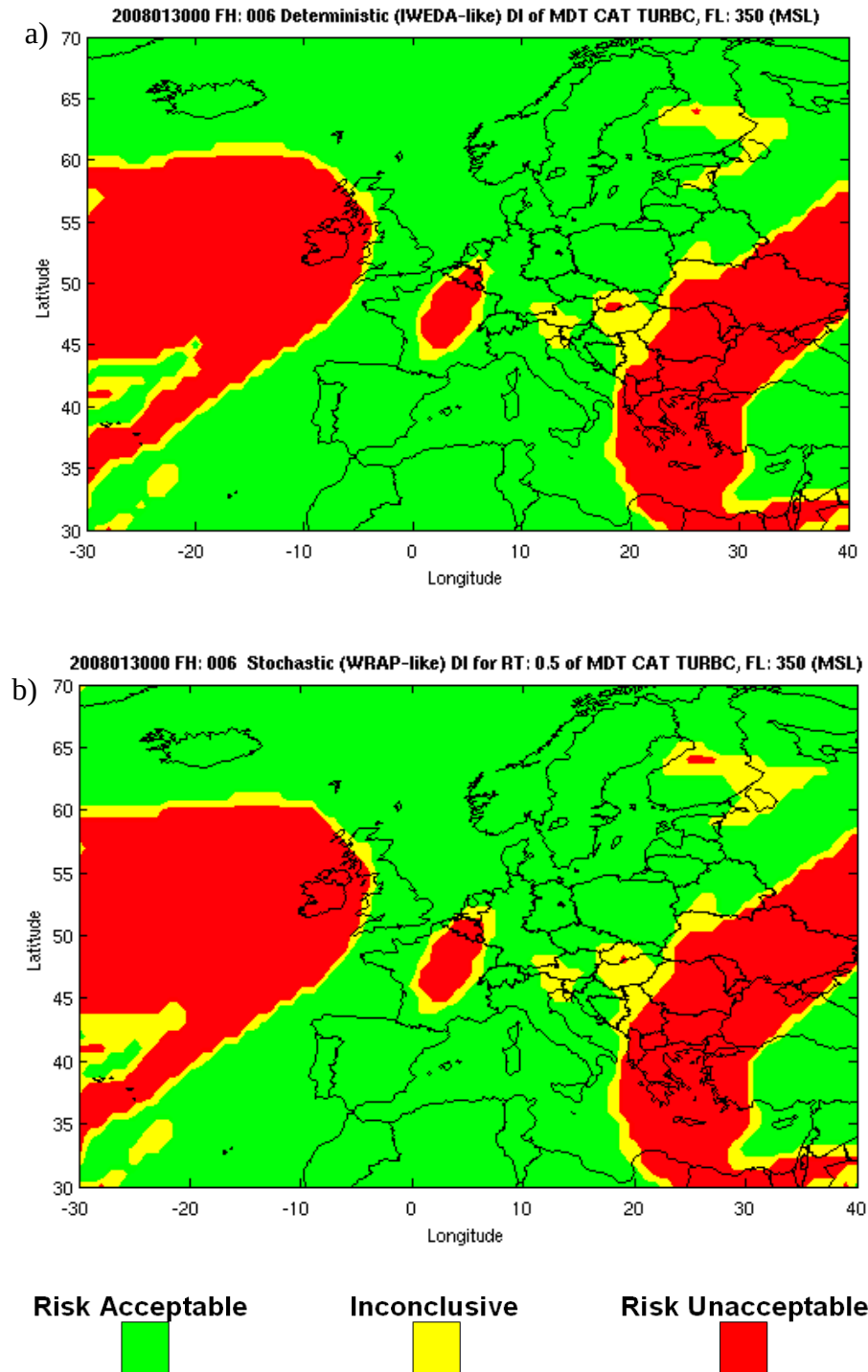


Figure 69. Winter run, FH 06, FL 350 MSL. a) Simulated IWEDA MDT turbulence DI b) Simulated WRAP MDT turbulence DI with RT=0.5

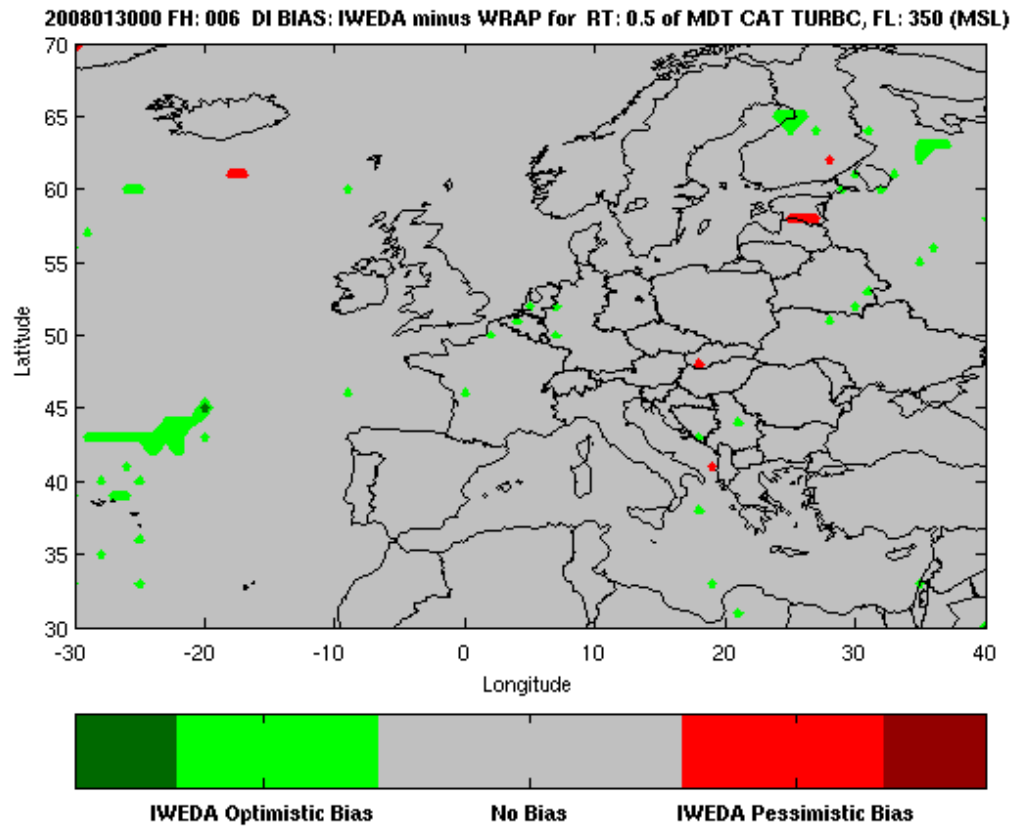


Figure 70. Winter run, FH 06, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.5.

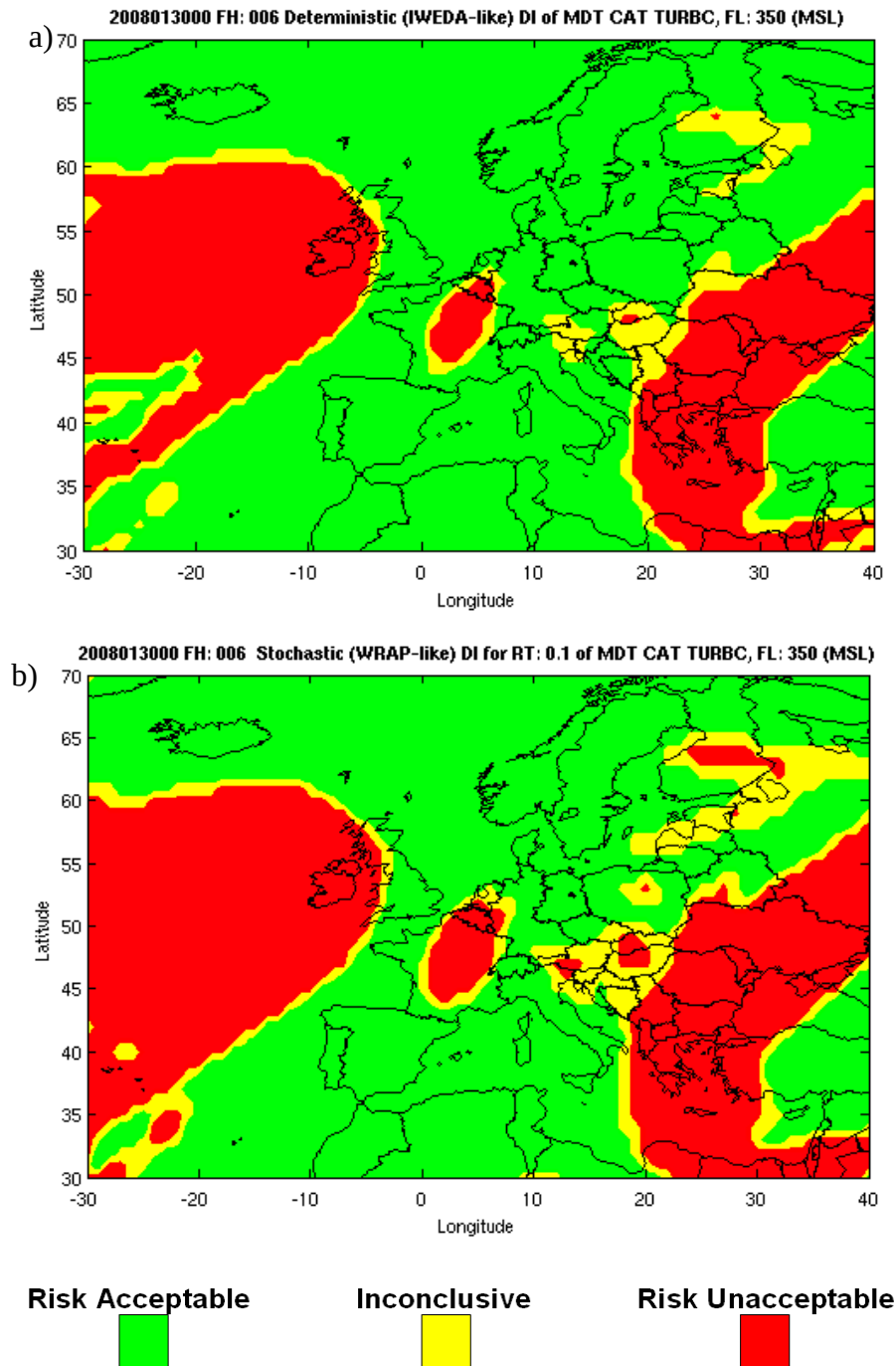


Figure 71. Winter run, FH 06, FL 350 MSL. a) Simulated IWEDA MDT turbulence DI b) Simulated WRAP MDT turbulence DI with RT=0.1

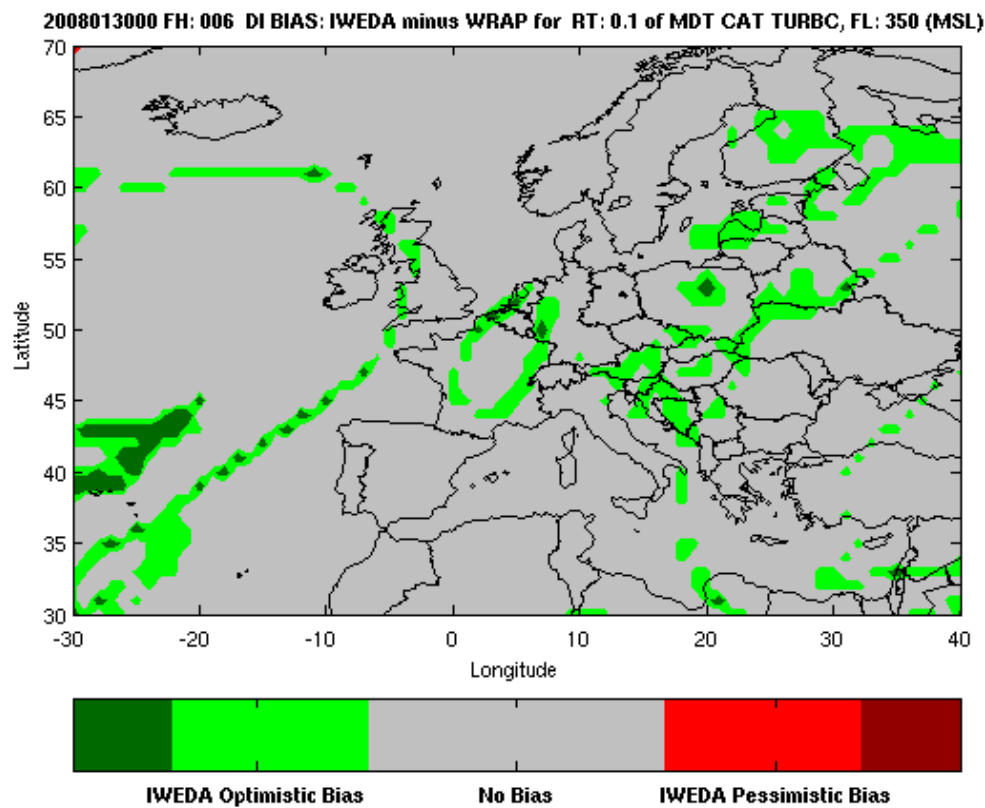
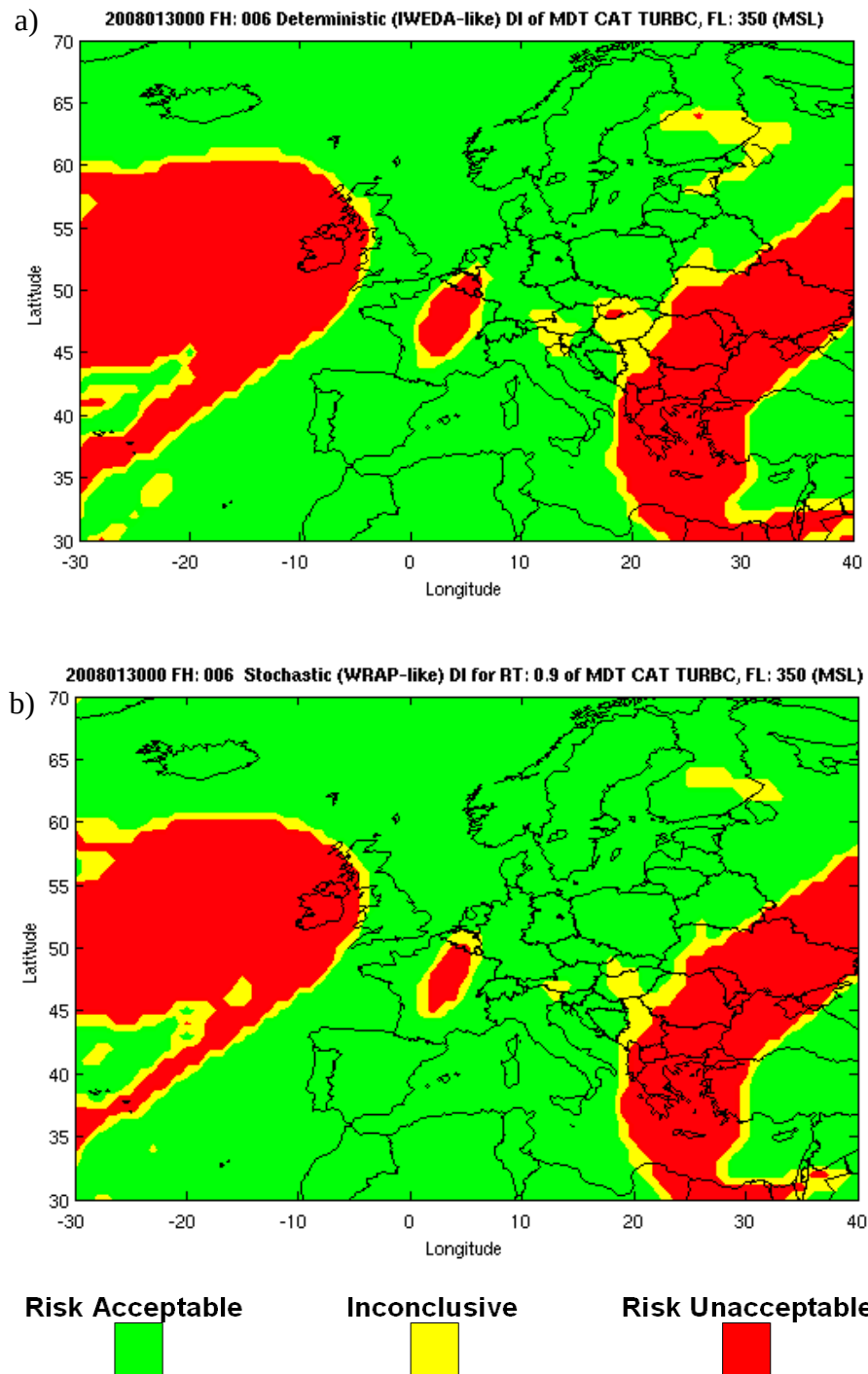


Figure 72. Winter run, FH 06, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.1.



\Figure 73. Winter run, FH 06, FL 350 MSL. a) Simulated IWEDA MDT turbulence DI b) Simulated WRAP MDT turbulence DI with RT=0.9

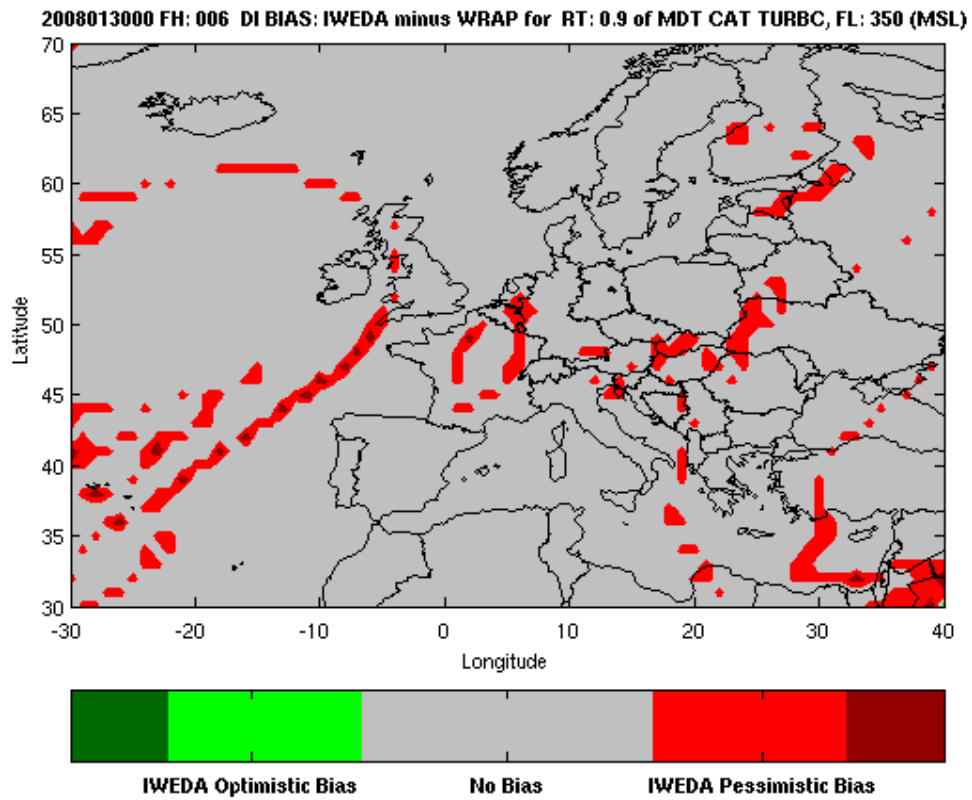


Figure 74. Winter run, FH 06, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.9.

B. FORECAST HOUR: 84

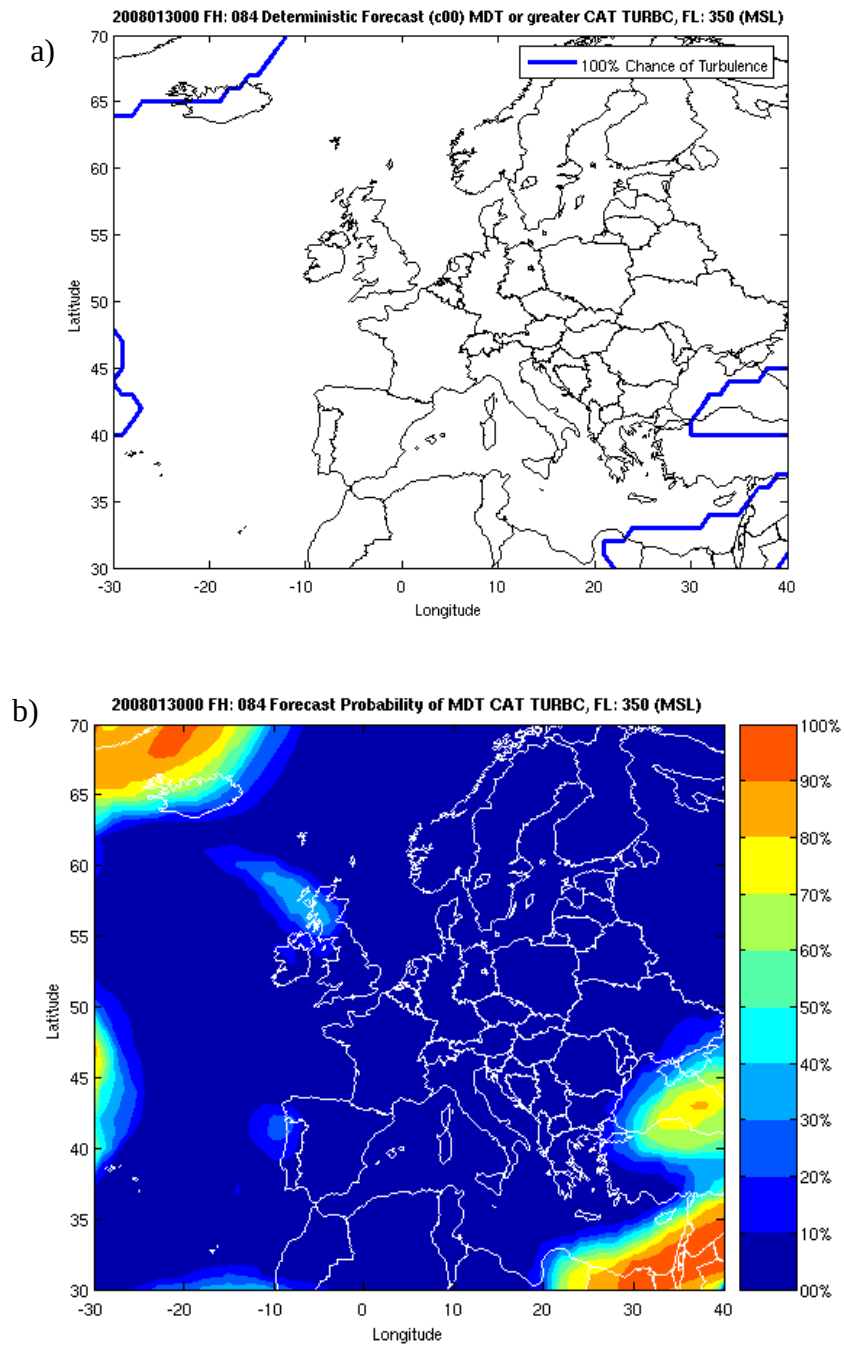


Figure 75. Winter run, FH 84, FL 350 MSL. a) Deterministic MDT turbulence forecast b) Stochastic MDT turbulence forecast using FP

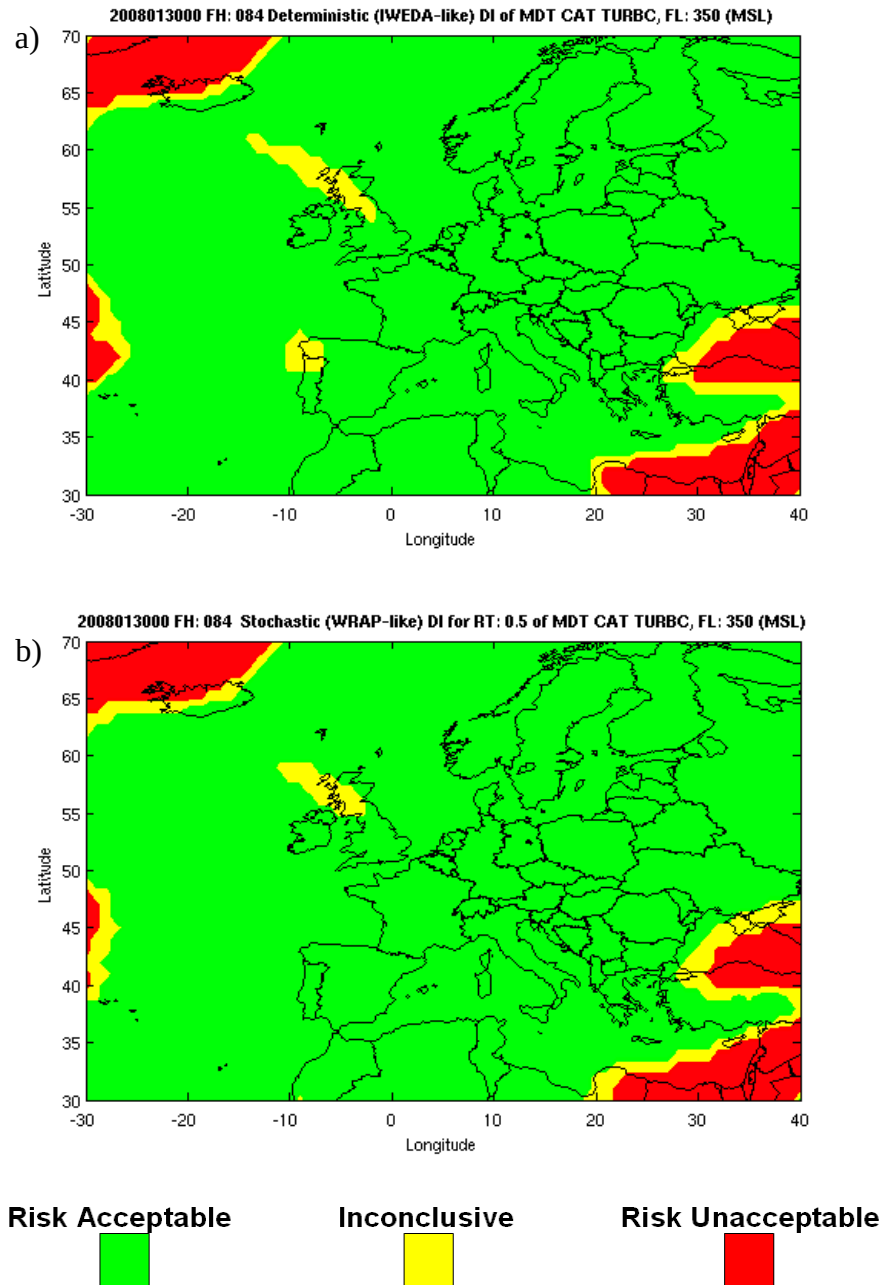


Figure 76. Winter run, FH 84, FL 350 MSL. a) Simulated IWEDA MDT turbulence DI b) Simulated WRAP MDT turbulence DI with RT=0.5

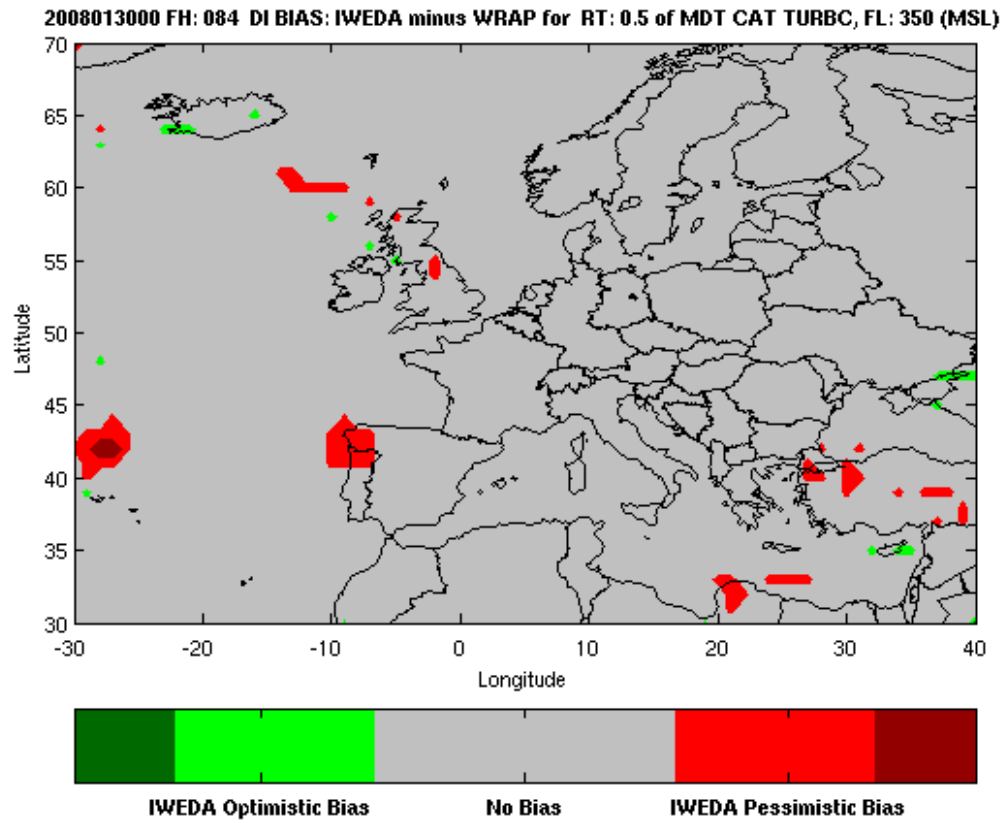


Figure 77. Winter run, FH 84, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.5.

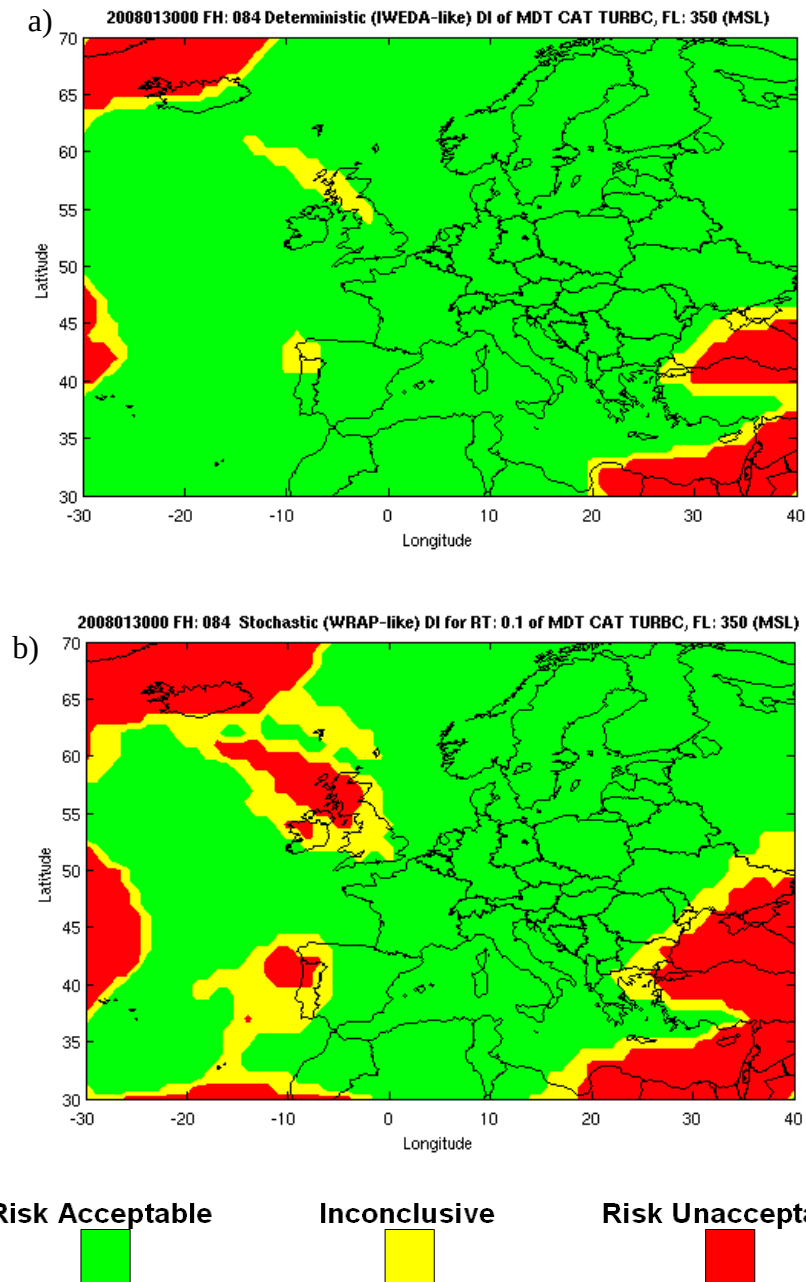


Figure 78. Winter run, FH 84, FL 350 MSL. a) Simulated IWEDA MDT turbulence DI b) Simulated WRAP MDT turbulence DI with RT=0.1

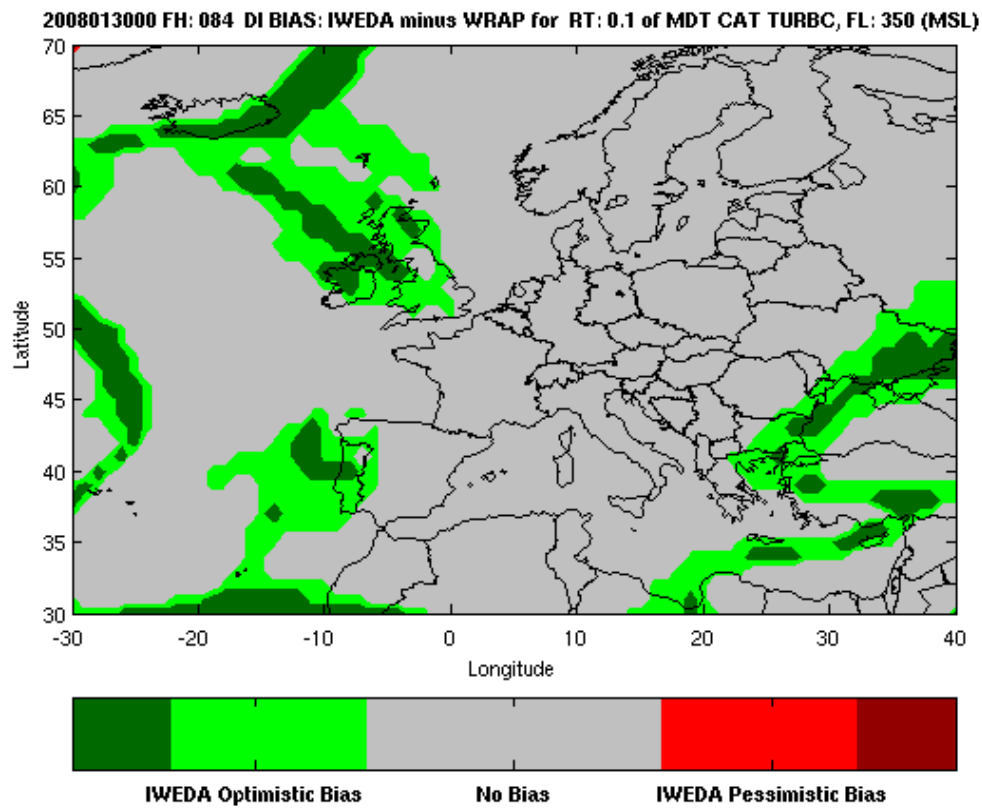


Figure 79. Winter run, FH 84, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.1.

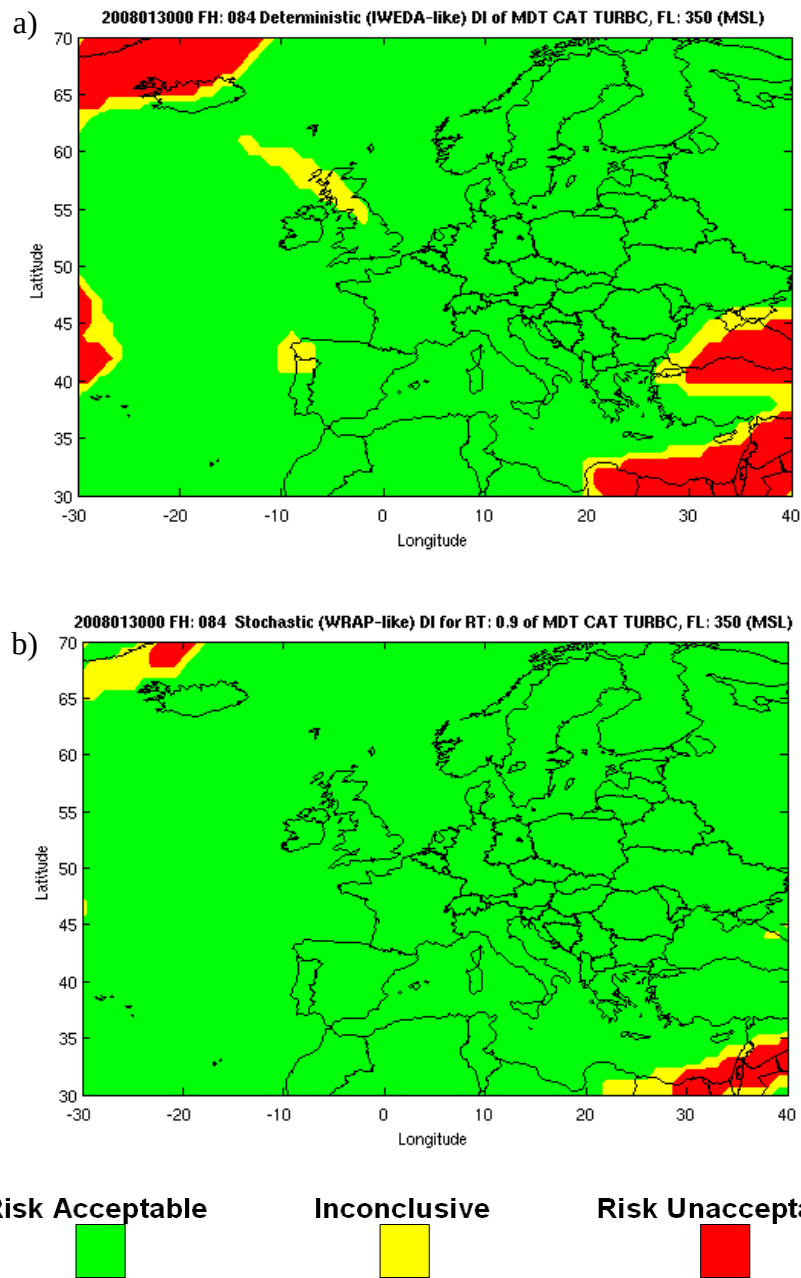


Figure 80. Winter run, FH 84, FL 350 MSL. a) Simulated IWEDA MDT turbulence DI b) Simulated WRAP MDT turbulence DI with RT=0.9

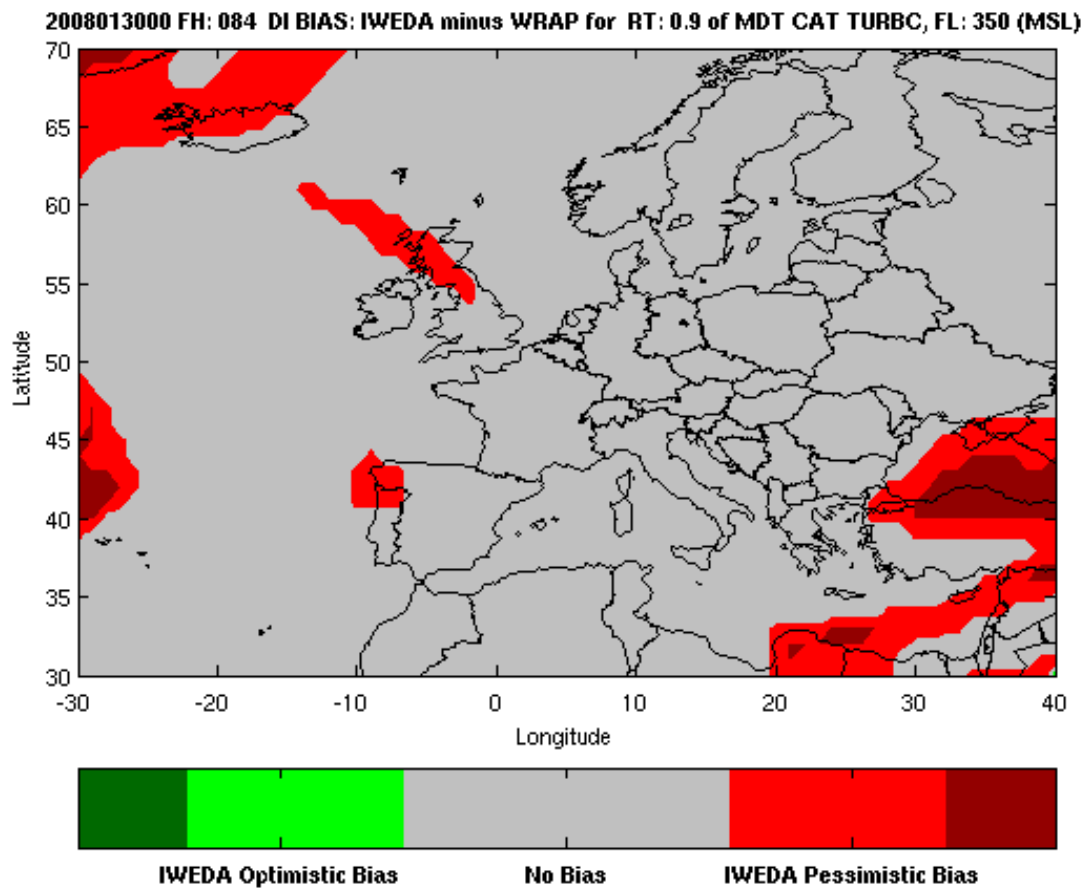


Figure 81. Winter run, FH 84, FL 350 MSL. DI Bias: IWEDA minus WRAP chart. WRAP RT=0.9.

APPENDIX C: WINTER CASE: SIMULATED WRAP VERSUS RSCH RESULTS

The following are figures from the simulated WRAP and research results (RSCH) comparisons of January 30, 2008, 00Z (winter) run. For reasoning, see similar cases in Chapter III, section B.3.

A. FORECAST HOUR: 06

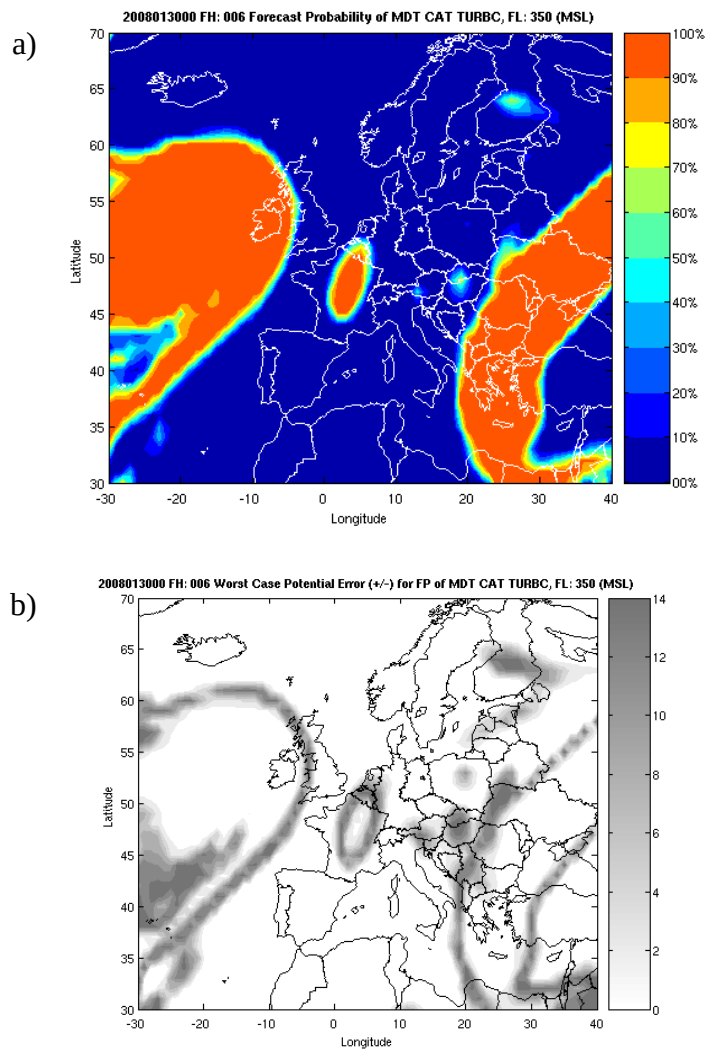


Figure 82. Winter run, FH 06, FL 350 MSL. a) FP of MDT turbulence b) Maximum (Worst Case) Potential Error for FP of MDT turbulence

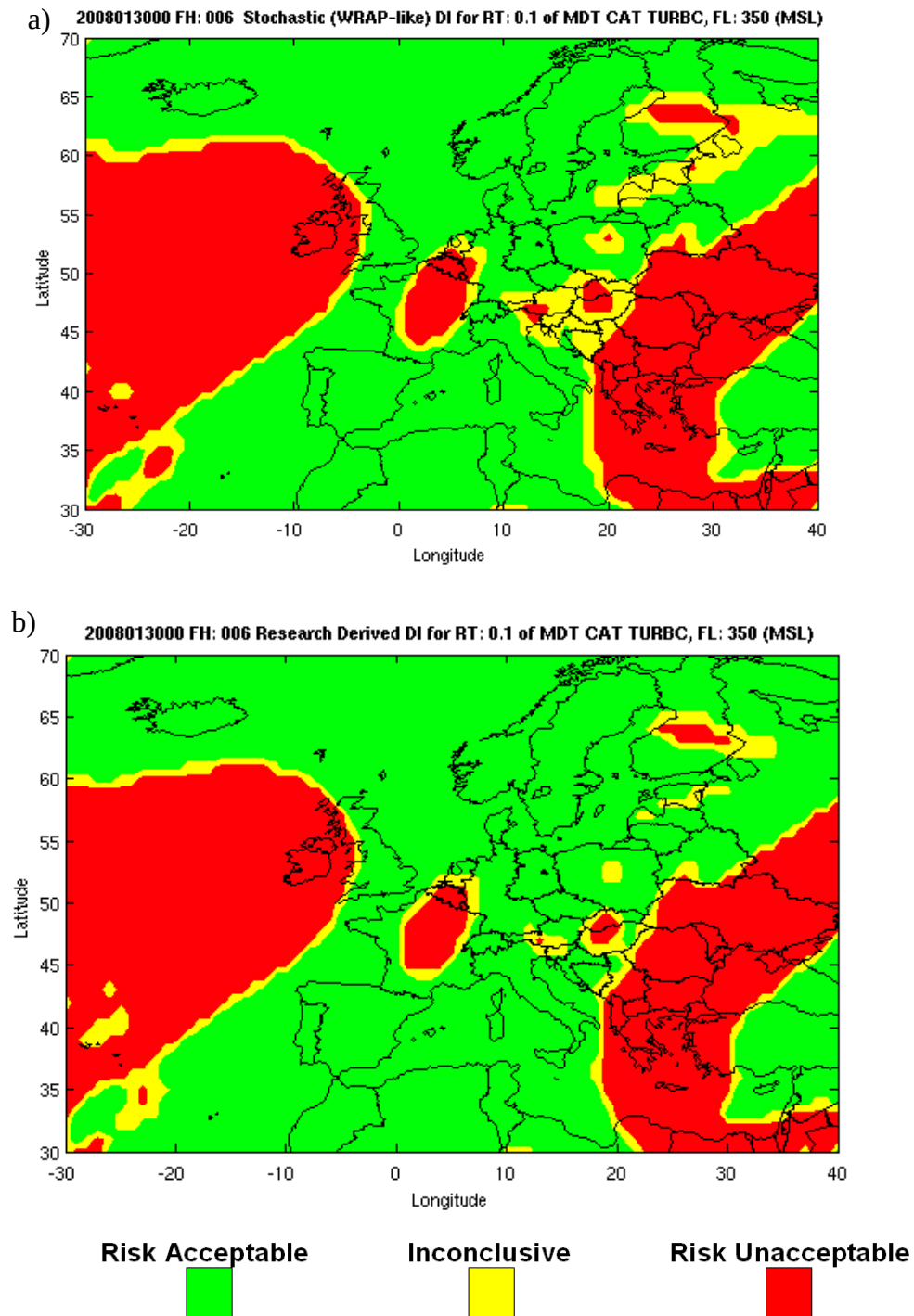


Figure 83. Winter run, FH 06, FL 350 MSL, RT=0.1 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI

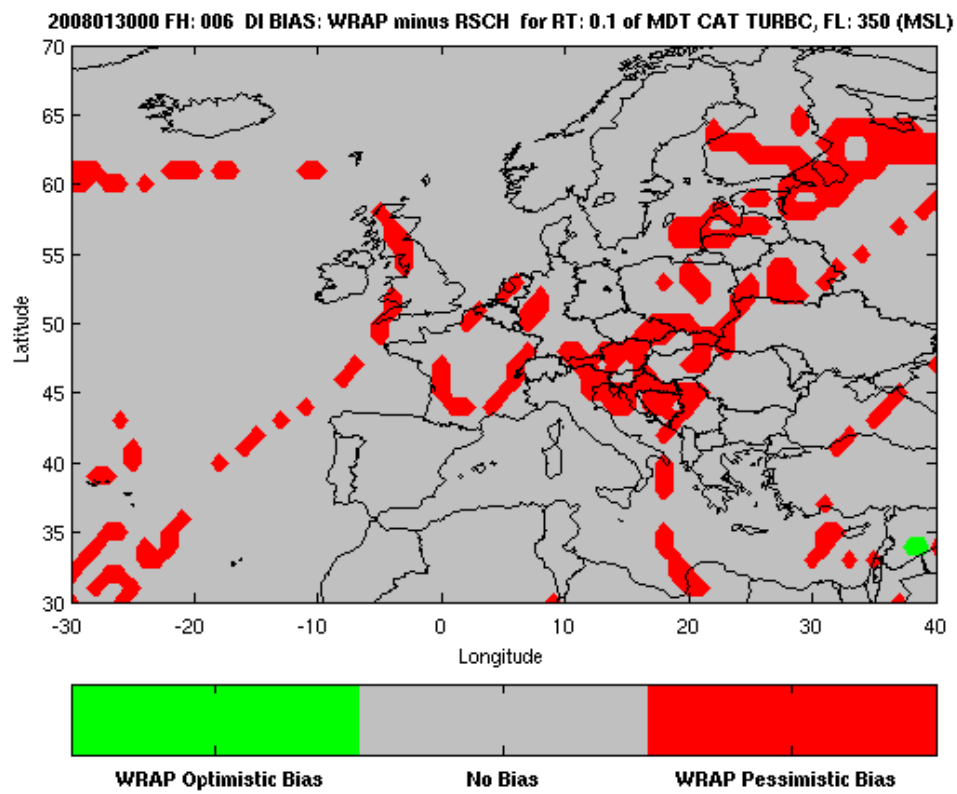


Figure 84. Winter run, FH 06, FL 350 MSL, RT=0.1. DI Bias: WRAP minus RSCH chart.

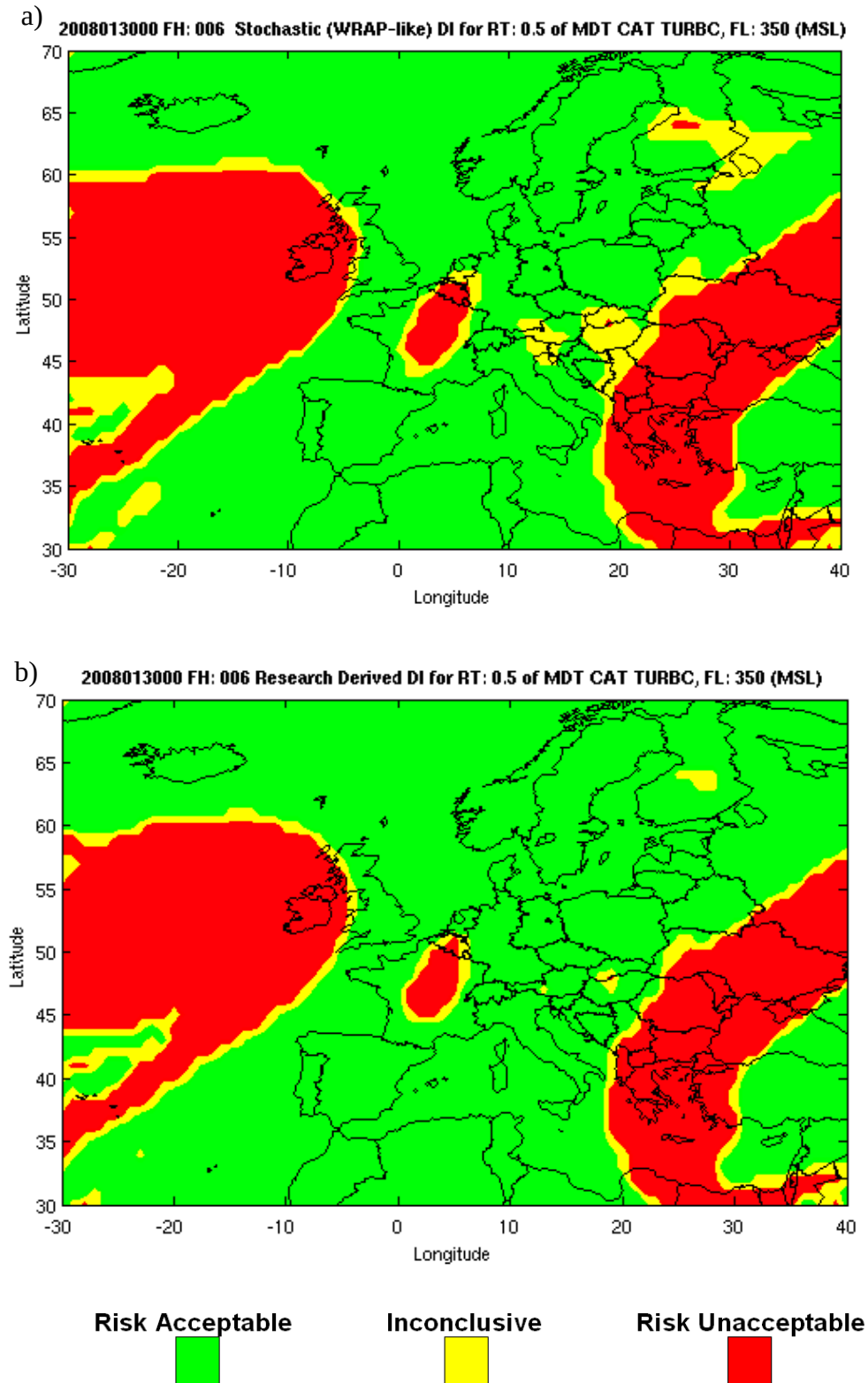


Figure 85. Winter run, FH 06, FL 350 MSL, RT=0.5 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI

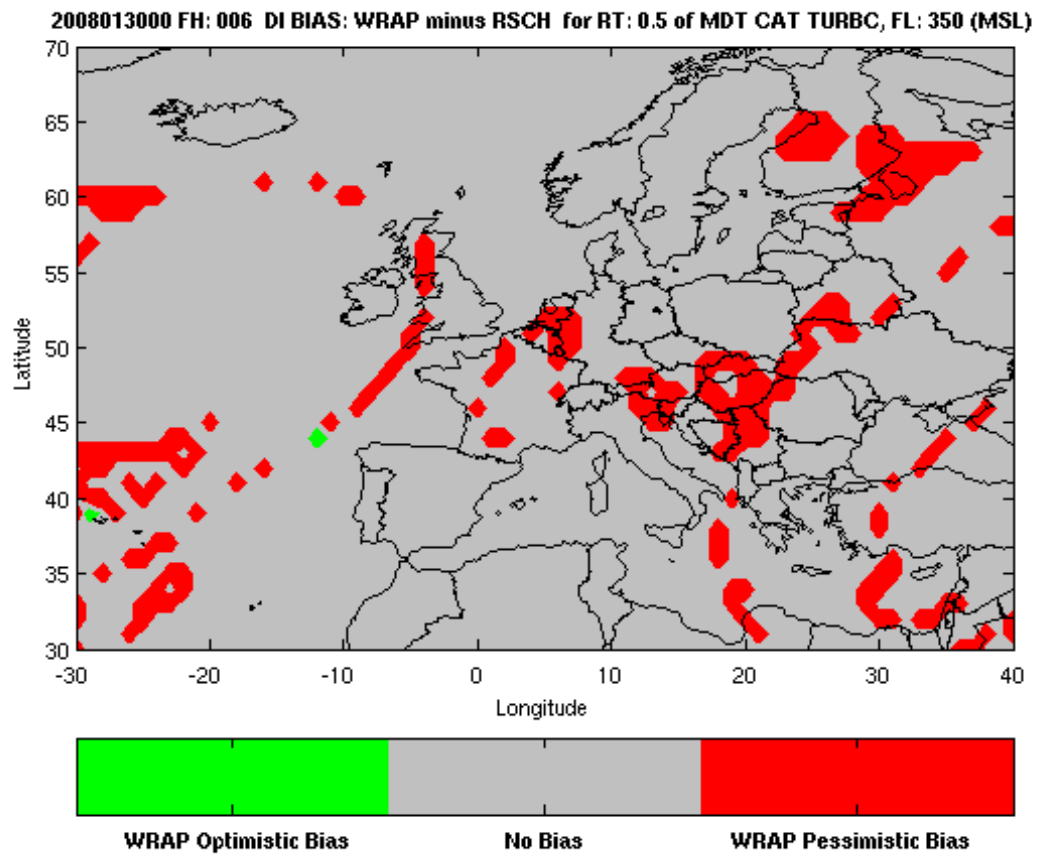


Figure 86. Winter run, FH 06, FL 350 MSL, RT=0.5. DI Bias: WRAP minus RSCH chart.

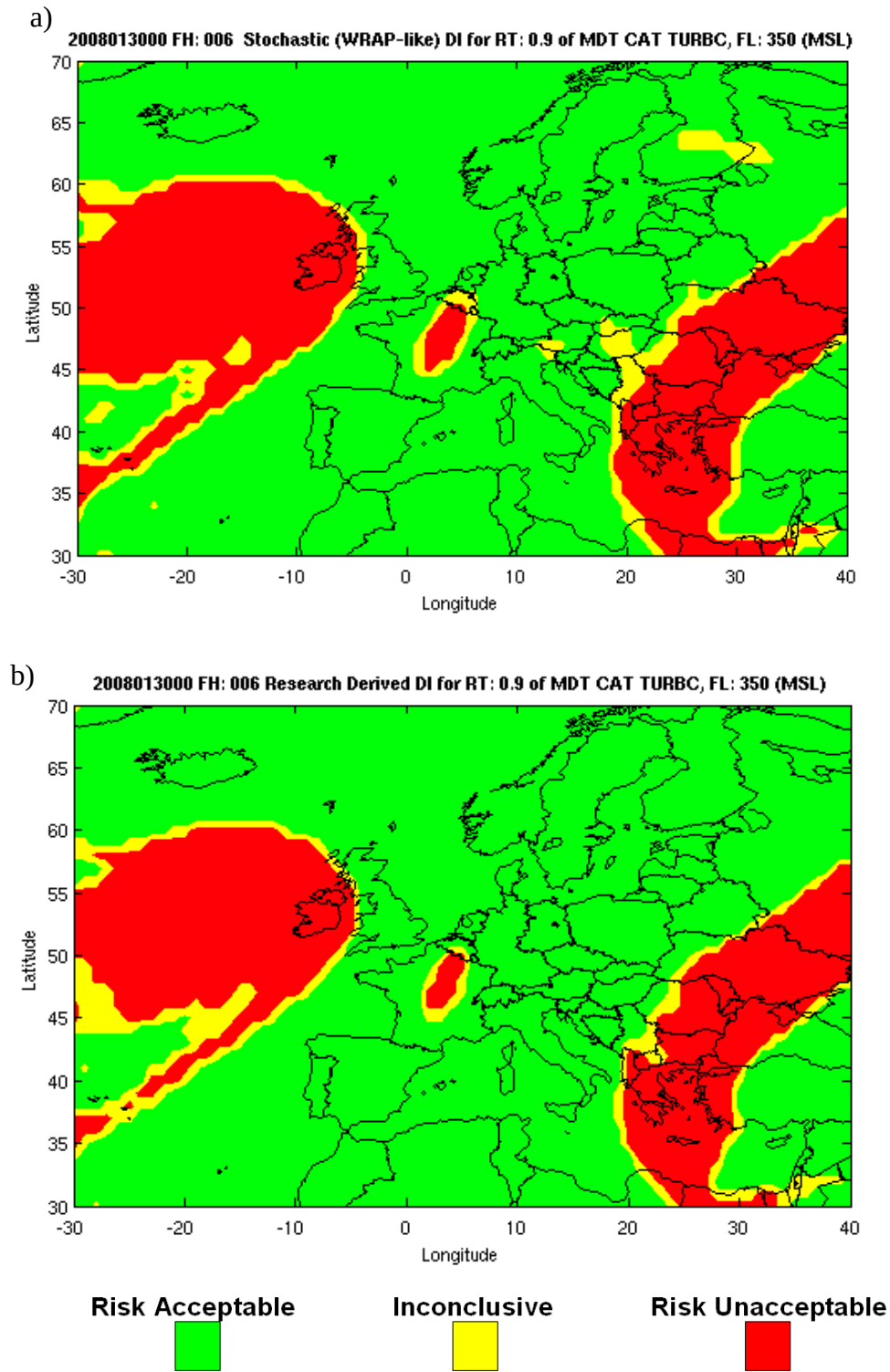


Figure 87. Winter run, FH 06, FL 350 MSL, RT=0.9 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI

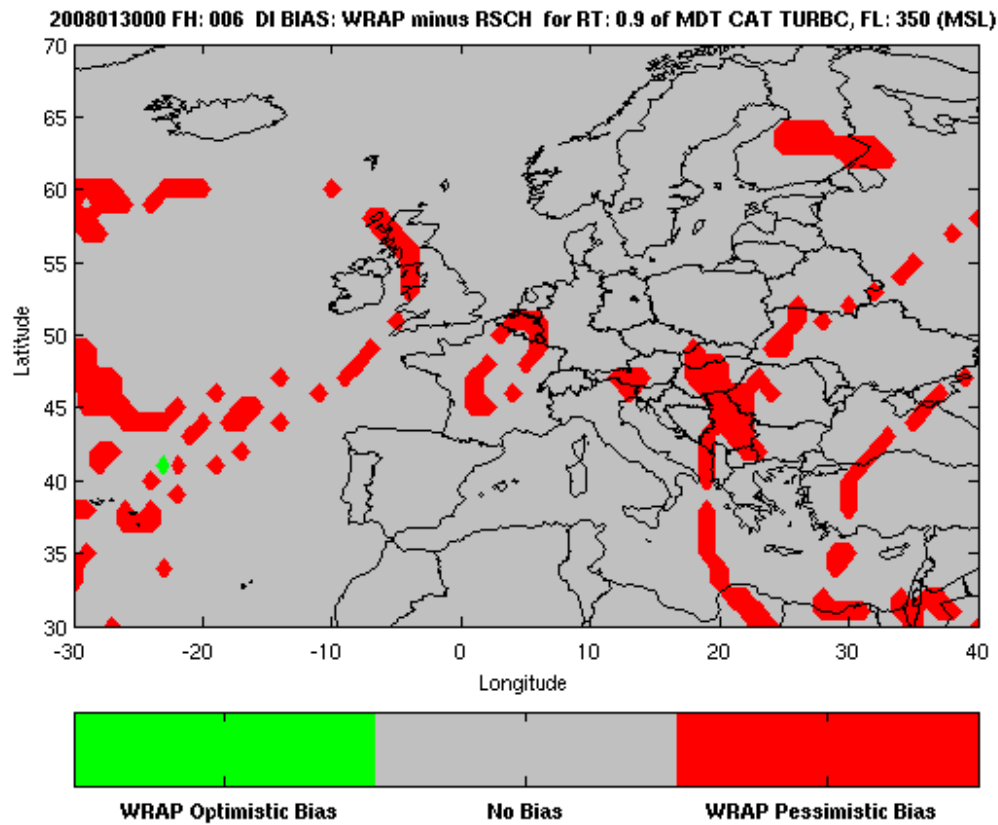


Figure 88. Winter run, FH 06, FL 350 MSL, RT=0.9. DI Bias: WRAP minus RSCH chart.

B. FORECAST HOUR: 84

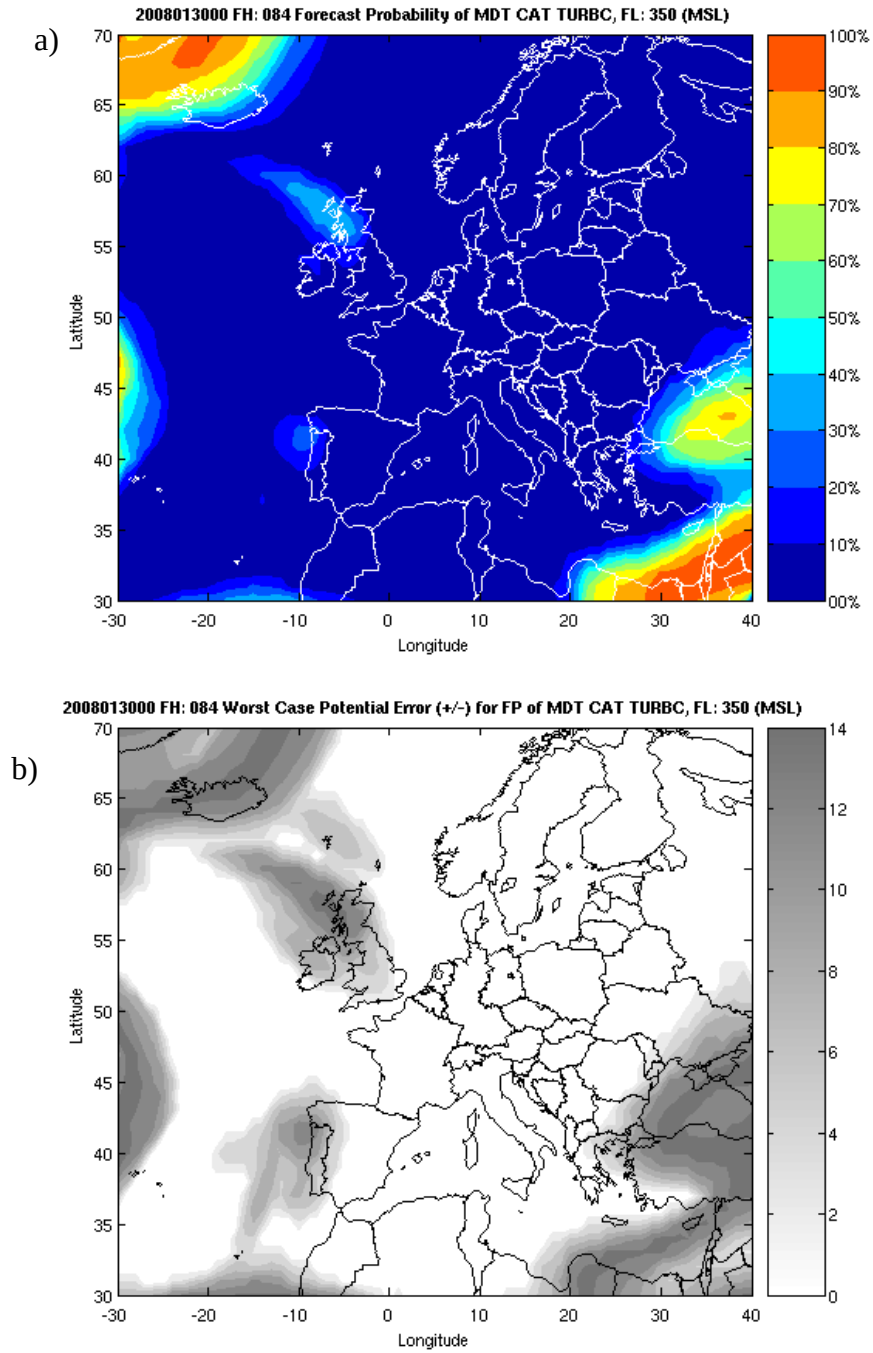


Figure 89. Winter run, FH 84, FL 350 MSL. a) FP of MDT turbulence b) Maximum (Worst Case) Potential Error for FP of MDT turbulence

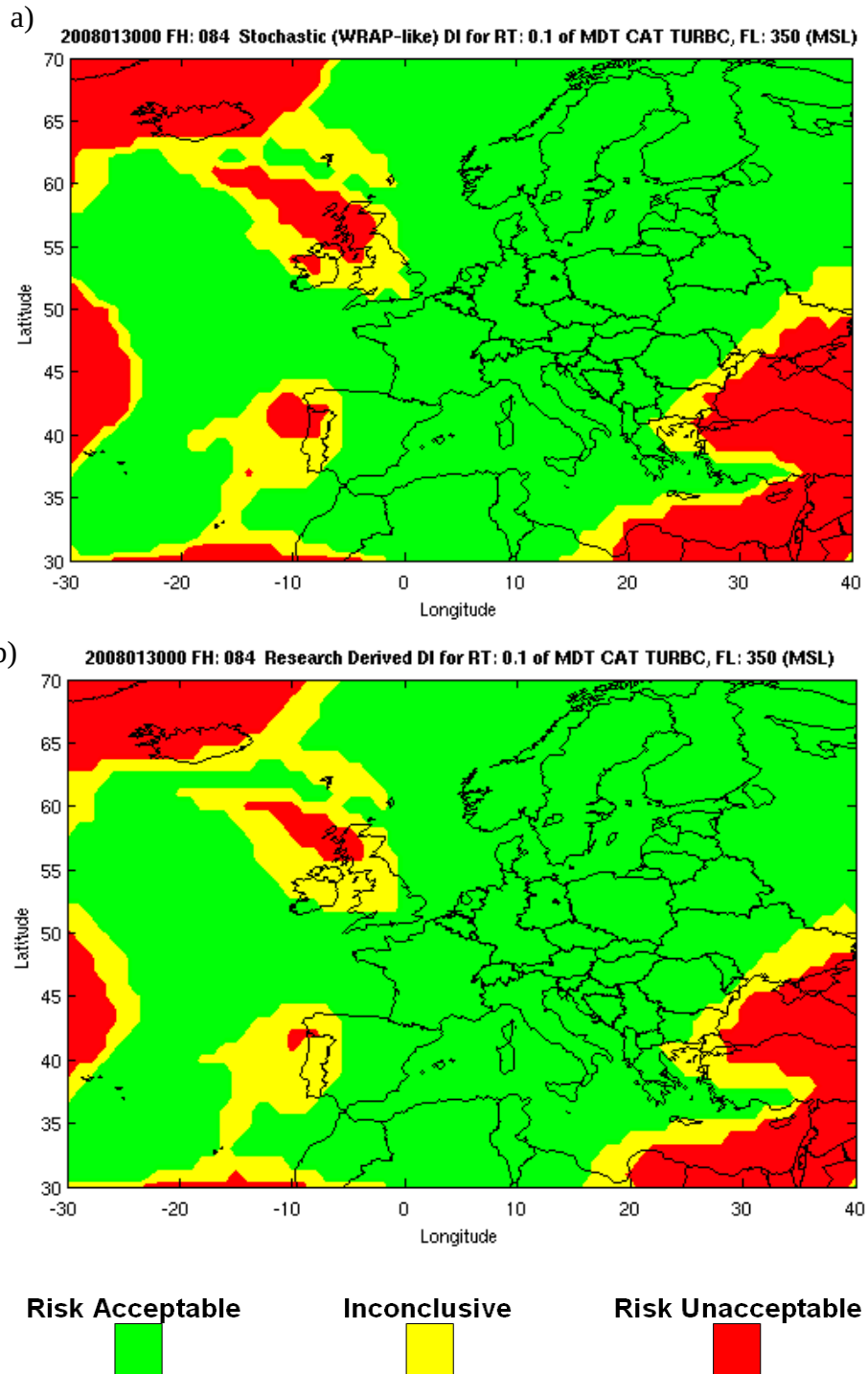


Figure 90. Winter run, FH 84, FL 350 MSL, RT=0.1 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI

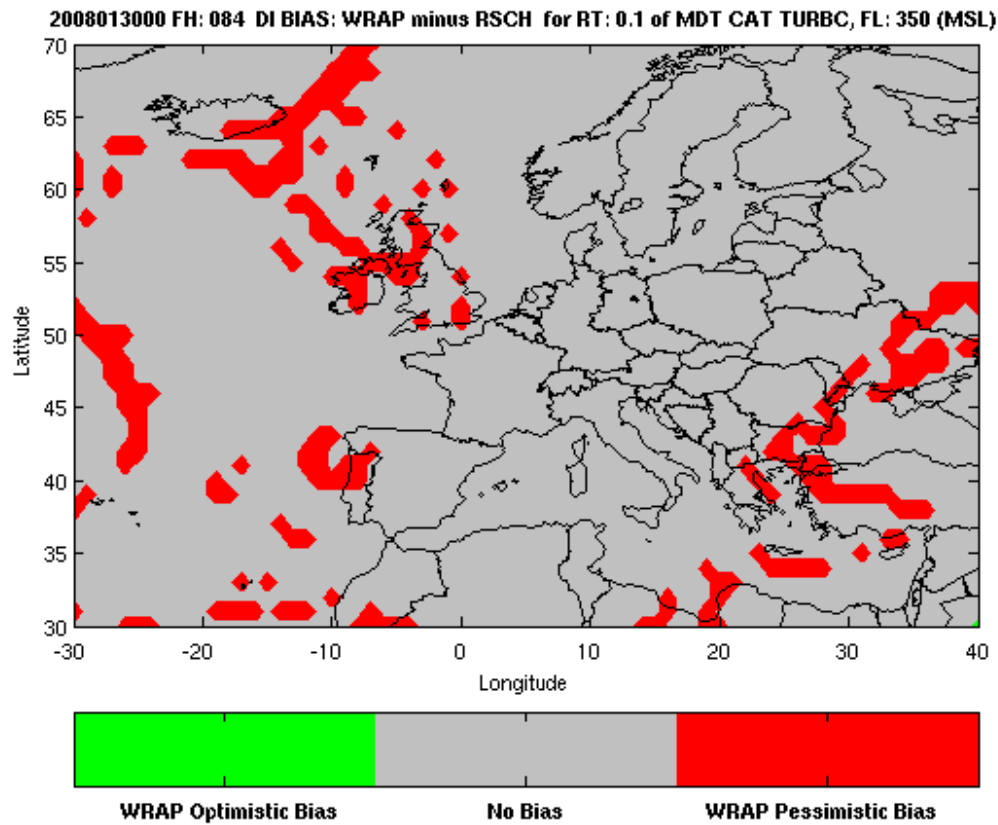


Figure 91. Winter run, FH 84, FL 350 MSL, RT=0.1. DI Bias: WRAP minus RSCH chart.

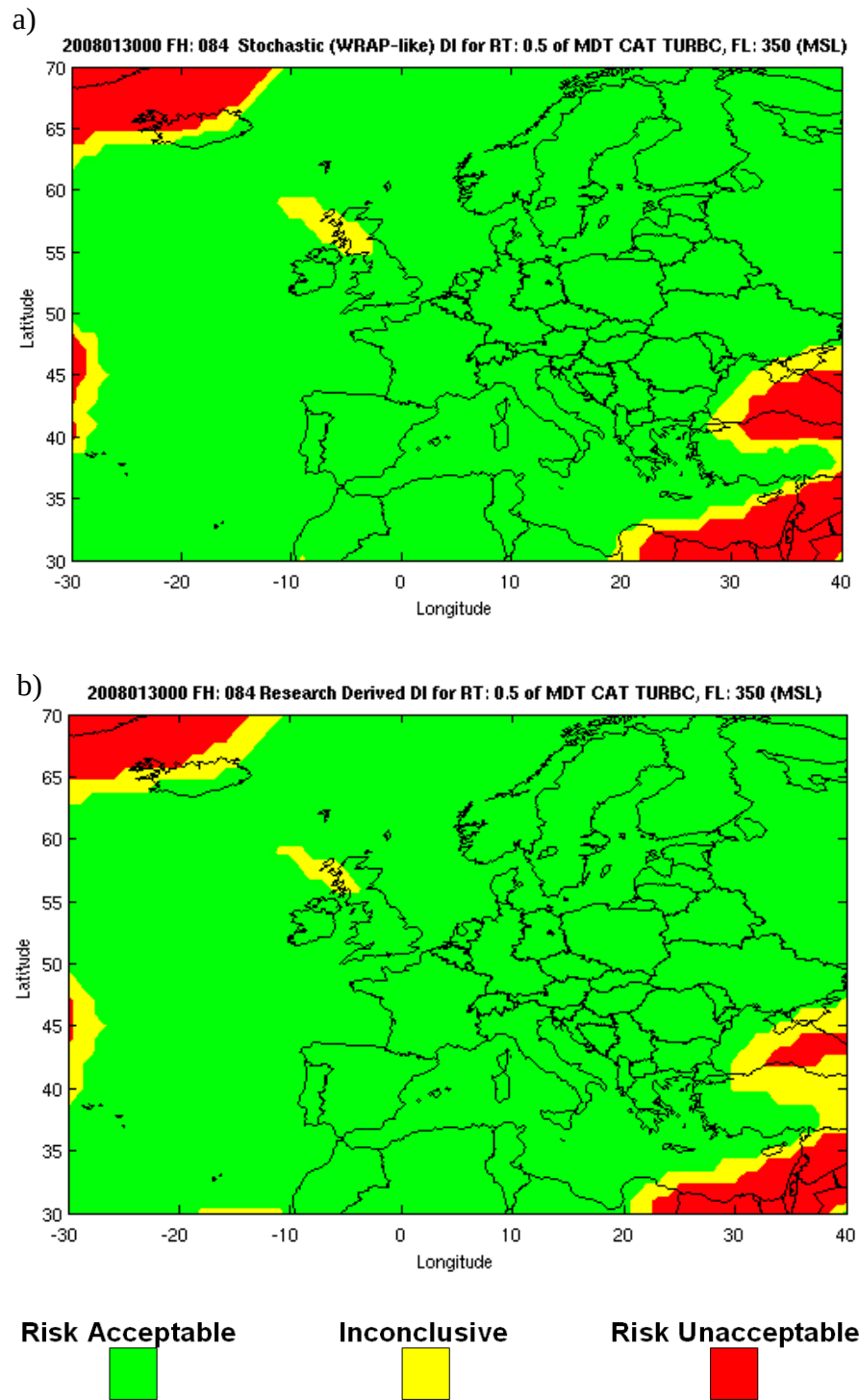


Figure 92. Winter run, FH 84, FL 350 MSL, RT=0.5 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI

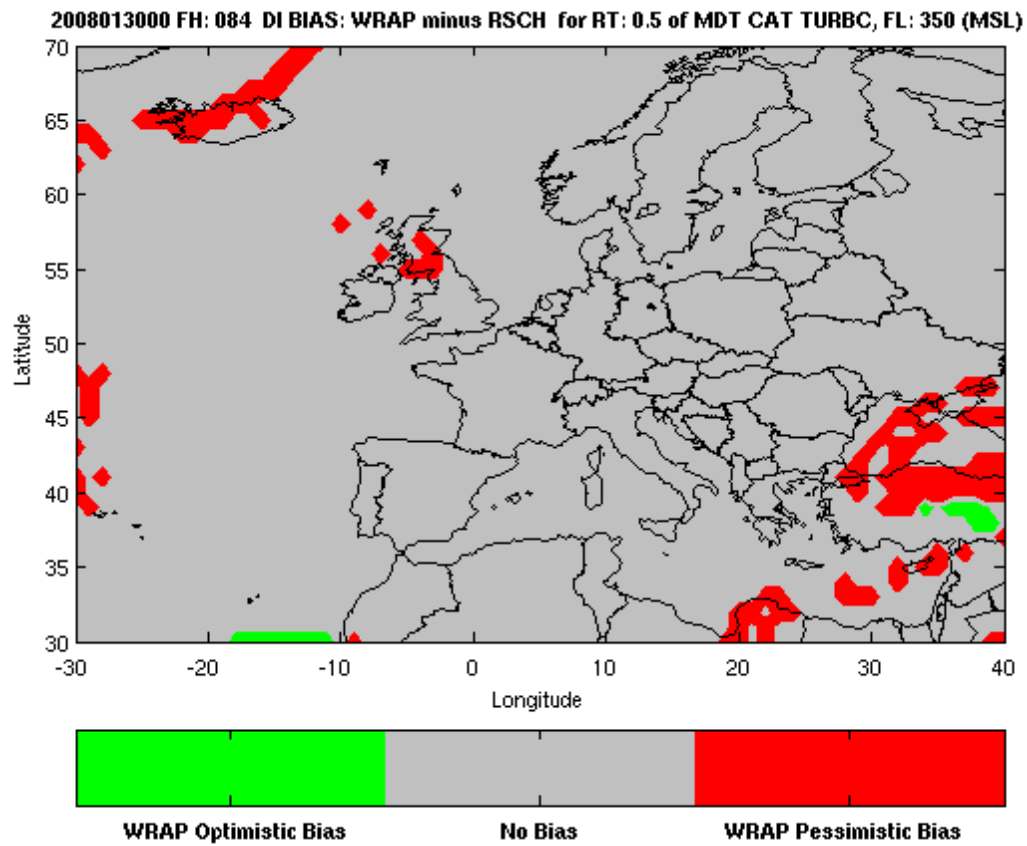


Figure 93. Winter run, FH 84, FL 350 MSL, RT=0.5. DI Bias: WRAP minus RSCH chart.

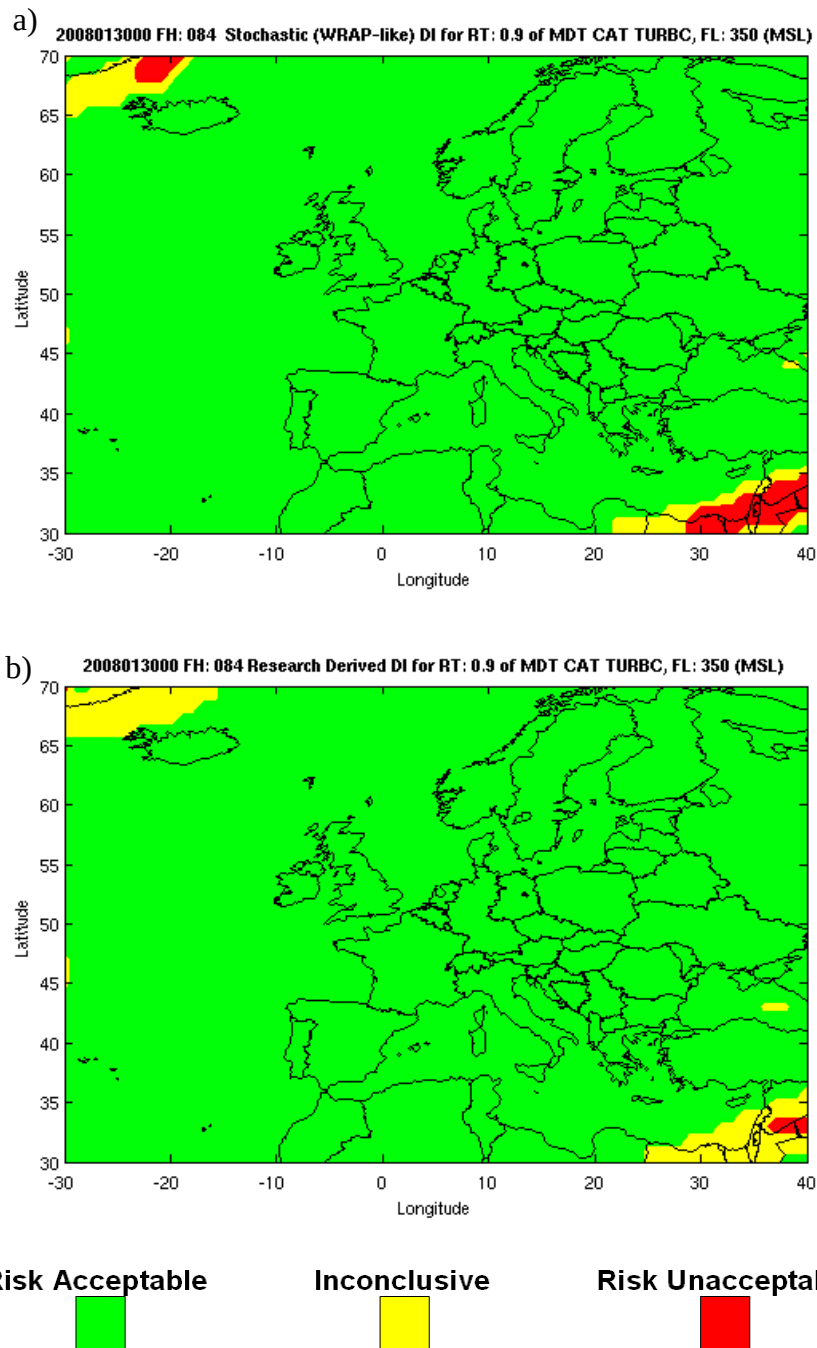


Figure 94. Winter run, FH 84, FL 350 MSL, RT=0.9 a) Simulated WRAP MDT turbulence DI b) RSCH derived MDT turbulence DI

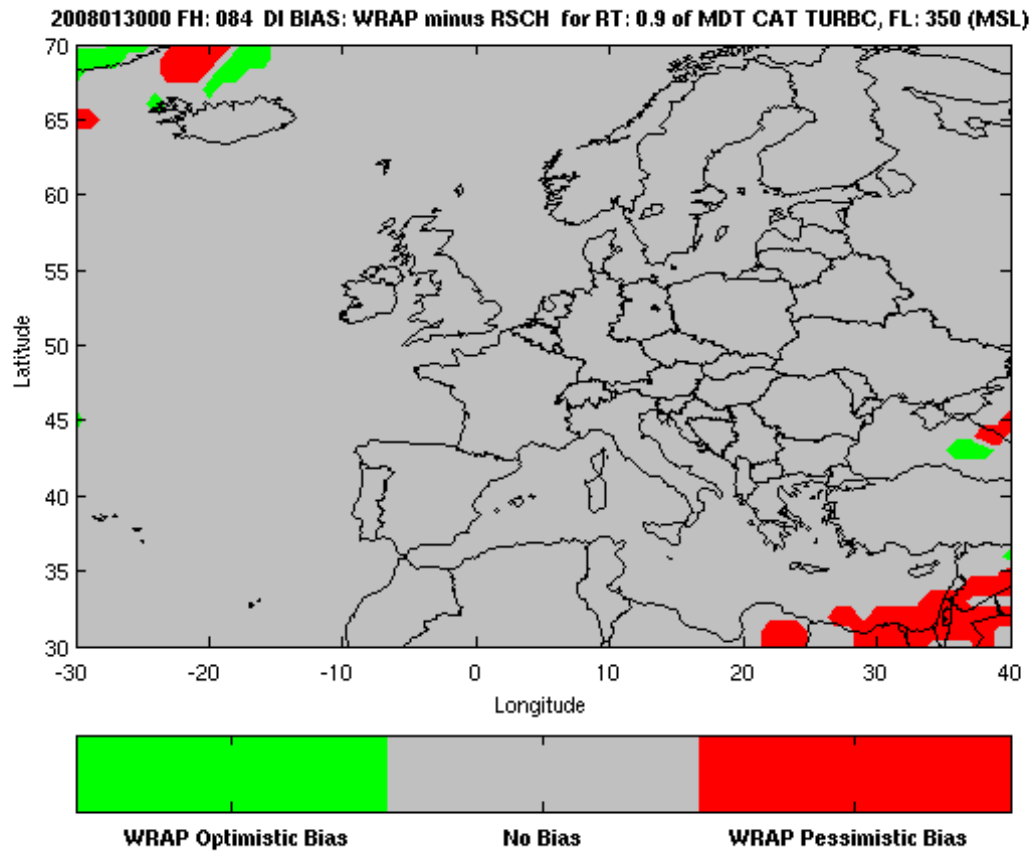


Figure 95. Winter run, FH 84, FL 350 MSL, RT=0.9. DI Bias: WRAP minus RSCH chart.

LIST OF REFERENCES

- Air Force Instruction 90-901 (AFI 90-901), cited 2000: Operational Risk Management (ORM). [Available online at <http://www.epublishing.af.mil/shared/media/epubs/AFI90-901.pdf>], (Accessed Mar 8, 2008).
- Air Force Pamphlet 90-902 (AFPAM 90-902), cited 2000: Operational Risk Management (ORM) – Guidelines and Tools. [Available online at [http://www.epublishing.af.mil/ shared/media/epubs/AFPAM90-902.pdf](http://www.epublishing.af.mil/shared/media/epubs/AFPAM90-902.pdf)], (Accessed Mar 8, 2008).
- . "chaos." *The American Heritage® Dictionary of the English Language*, Fourth Edition. Houghton Mifflin Company, 2004. Dictionary.com website. [Available online at: <http://dictionary.reference.com/browse/chaos>] (Accessed Jan. 29, 2008).
- Cunningham, J.G., 2006: *Applying Ensemble Prediction Systems to Department of Defense of Operations*. M.S. Thesis, Graduate School of Engineering and Applied Sciences, Naval Postgraduate School. 149 pp. [Available from the Defense Technical Information Center]
- Eady, E., 1949: Long waves and cyclone waves. *Tellus*, **1**, 33–52.
- . 1951: The quantitative theory of cyclone development. *Compendium of Meteorology*, T. Malone, Ed., Amer. Meteor.Soc., 464–469.
- Eckel, F.A., 2006: "FYI: Ensemble Forecasting". Air Force Weather Agency. Number 65.
- , 2007: *Introduction to Ensemble Forecasting*. Draft October 22, 2007. Naval Postgraduate School. 102 pp.
- . 2008: *Estimating Ambiguity in Ensemble Forecasts*. Draft January 25, 2008. Naval Postgraduate School. 32 pp.
- Ellrod, G. P. and D. I. Knapp, 1992: An objective clear-air turbulence forecasting technique: verification and operational use. *Weather and Forecasting*, **7**, 150–165.
- Hammond, K. R. 1996. *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. New York: Oxford University Press.

- Jolliffe, Ian T., Stephenson, David B., 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons Publishing.
- Joint Ensemble Forecasting System (JEFS), cited 2007: JEFS Non-Operational Turbulence Chart [Available online: <https://weather.afwa.af.mil/HOST/HOME/DNXM/JEFS/jefs.html>] (Accessed July 31, 2007).
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.
- Lewis, J. M., 2005: Roots of Ensemble Forecasting. *Monthly Weather Review*, **133**, 1865-1885.
- Lorenz, E. N., 1963: Deterministic Nonperiodic Flow. *Journal of Atmospheric Sciences*, **20**, 130-141.
- National Research Council, 2006: Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts Committee on Estimating and Communicating". National Academies. 120 pp.
- Next Century Corporation, 2007: *WRAP: Phase II: Final Report*. 53 pp.
- Sembach Operational Weather Squadron (OWS), cited 2007: Current deterministic format for USAFE Upper Level Turbulence Chart. [Available online at: <http://ows.public.sembach.af.mil>], (Accessed July 31, 2007).
- Stewart, T.R., and Lusk, C.M. 1994. Seven components of judgmental forecastingskill: Implications for research and the improvement of forecasts. *Journal of Forecasting*, **13**, 579-599.
- . 2007. Presentation: "The human forecaster: Components of Judgmental Skill" Workshop: Doing Something About the Weather. January 5-7, 2007. Naval Postgraduate School, Monterey, CA.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2d ed. Academic Press, 621 pp.
- Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Myline, 2002: The Economic Value of Ensemble-Based Weather Forecasts. *Bulletin of the American Meteorological Society*. Vol. 83, No. 1, pp. 73–83.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia
2. Dudley Knox Library
Naval Postgraduate School
Monterey, California
3. Major Tony Eckel, USAF
Naval Postgraduate School
Monterey, California
4. Dr. Patrick Harr
Naval Postgraduate School
Monterey, California
5. Dr. Philip Durkee
Naval Postgraduate School
Monterey, California
6. Dr. Eva Regnier
Defense Resources Management Institute
Monterey, California
7. Barbara Sauter
Army Research Laboratory
White Sands Missile Range, New Mexico
8. Major Timothy Nobis, USAF
Air Force Weather Agency
Offutt AFB, Nebraska