3-5-2004

Verification of the Mountain Wave Forecast Model's Stratospheric Turbulence Forecasts Using Sounding Data and Pilot Reports

Scott M. Miller

VERIFICATION OF THE MOUNTAIN WAVE FORECAST MODEL'S STRATOSPHERIC TURBULENCE FORECASTS USING SOUNDING DATA AND PILOT REPORTS

THESIS

Scott M. Miller, Captain, USAF

AFIT/GM/ENP/04-10

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government.

AFIT/GM/ENP/04-10

VERIFICATION OF THE MOUNTAIN WAVE FORECAST MODEL'S STRATOSPHERIC TURBULENCE FORECASTS USING SOUNDING DATA AND PILOT REPORTS

THESIS

Presented to the Faculty

Department of Engineering Physics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Meteorology

Scott M. Miller, B.S.

Captain, USAF

March 2004

AFIT/GM/ENP/04-10

VERIFICATION OF THE MOUNTAIN WAVE FORECAST MODEL'S STRATOSPHERIC TURBULENCE FORECASTS USING SOUNDING DATA AND PILOT REPORTS

Scott M. Miller, B.S.
Captain, USAF

Approved:

Michael K. Walters (Chairman)	4 MAR 04 Date
Ronald P. Lowther (Member)	5 MAR 04 Date
Daniel E. Reynolds (Member)	5 Mar 04 Date

AFIT/GM/ENP/04-10

Abstract

Since stratospheric turbulence (Stratoturb) is becoming an increased concern to the Air Force, the threat of damage to aircraft must be addressed. Therefore, the Air Force Weather Agency (AFWA) requests an accurate Stratoturb forecast model.

In 2002, The Mountain Wave Forecast Model (MWFM) was modified in order to develop a Stratoturb forecast tool. Turbulence forecasts generated twice daily by the MWFM for locations over East Asia over a period of thirty days were compared to output from the Rawindsonde Observation (RAOB) program to determine if the model agreed with the program output. Although the results were promising, verification by aircraft crews flying through the stratosphere would improve the confidence of this forecast model, improving the forecaster's ability to warn pilots and alleviate the potential danger associated with flying through areas of Stratoturb.

This thesis continues that research. Three major changes were made. Pilot reports (PIREPs) were collected for verification of MWFM forecasts, the model's time resolution was increased for better comparison to PIREPs, and data were collected for nearly a year to determine season performance. Model performance at ten sounding locations was analyzed to determine if performance improved over a certain terrain type. Model performance at three atmospheric levels (100-70mb, 70-50mb, and 50-30mb) was also compared to determine if the model performed better at a certain altitude.

Results suggest that the MWFM is superior to previous methods of detecting Stratoturb. Therefore, the MWFM is recommended to AFWA for operational use.

Acknowledgments

I would like to thank all of the people who made this work possible. First, I would like to thank my thesis advisor, Lt Col Michael Walters, for his guidance and support throughout this effort. I would also like to express my gratitude to the other members of my committee, Lt Col Ronald Lowther and Mr. Daniel Reynolds, for the expertise they provided.

Two of my classmates were also extremely helpful and saved me days of work. Capt Brian Belson and Capt Lou Lussier patiently assisted with troubleshooting and programming. I would also like to thank those who made data available for comparison. Their help added a most needed new dimension to this work.

Finally, I would like to express thanks to my wife and children, who were patient and understanding during the long hours I spent on this research.

Scott M. Miller

Table of Contents

Page

List of Figures

List of Tables

VERIFICATION OF THE MOUNTAIN WAVE FORECAST MODEL'S STRATOSPHERIC TURBULENCE FORECASTS USING SOUNDING DATA AND PILOT REPORTS

I. Introduction

1.1 Background

Turbulence has long been known to be dangerous to aircraft. Tropospheric turbulence has been the primary focus of turbulence research, because most aircraft fly in the troposphere.

Turbulence forecasting is being accomplished by various agencies at tropospheric levels. Far less turbulence forecasting is being accomplished at stratospheric levels. Data from levels above the troposphere are often not included in model runs, since data from such high levels have been of no operational consequence until recently. Another reason for the limited amount of turbulence forecasting performed at stratospheric levels is the lack of aircraft flights at these levels. Since there has been an increase in aircraft flying in the lower stratosphere, it has become important to include stratospheric turbulence (Stratoturb) in research efforts.

In the fall of 2002, Capt Mark Allen performed research in order to produce a Stratoturb turbulence forecasting tool, producing a reasonably good Stratoturb forecast model. This research is a continuation of Allen's, performed with slight modifications. A brief summary of his research follows in the next section.

This thesis follows the journal guidelines set forth by the American Meteorological Society.

Stratoturb has been well-correlated with wind flow over mountainous terrain (Waco 1972). A good example of this correlation is over East Asia, where highlyanisotropic terrain features located in this region are known to be associated with increased levels of Stratoturb. This correlation is particularly true in the winter months, when the jet stream migrates, positioning itself orthogonal to ridge axes.

Stratoturb poses a serious threat to U-2 crews and unmanned aerial vehicles, which fly at stratospheric levels. Therefore, development of an accurate Stratoturb forecast model is important in order to avoid aircraft mishaps associated with Stratoturb.

1.2 Problem Statement

The Air Force Weather Agency (AFWA) has requested a tool to aid in the automated forecasting of Stratoturb. To fulfill this request, research began in the fall of 2002 by Capt Mark Allen.

The Naval Research Laboratory (NRL) has developed two versions of the Mountain Wave Forecast Model (MWFM), which are being used for tropospheric levels. Allen's research developed a Stratoturb forecast process by modifying this model for use at stratospheric levels over East Asia. In order to accomplish his research, the model was compiled and run using output data from the National Center for Environmental Prediction's (NCEP) Operational Global Forecast System (GFS) and the Fifth Generation Mesoscale Model (MM5). After gathering, ingesting and analyzing the data, the MWFM developed atmospheric profiles and determined locations of forecasted turbulence. Graphical and text output was produced for comparison to Environmental Research Services' Rawinsonde Observation (RAOB) program output. Allen's research used the

MWFM to forecast Stratoturb over locations in the Republic of Korea (ROK) and Japan over a 30-day period. MWFM model output forecasts over each of these locations were compared with output from RAOB (described in Chapter 3) which analyzes rawindsonde data to determine if the MWFM output agreed with RAOB output (Allen 2003). One problem inherent to this process is due to the fact that output from this program is subjective, that is, it has not been confirmed by comparison to in situ observations. While flying units are currently using this program to attempt to locate turbulence, it is only diagnostic, not prognostic. That is, the program analyzes sounding data, and does not have the ability to forecast future turbulence locations (unless forecasted sounding data were used as input).

To aid in fulfilling AFWA's request, Allen's research has been continued. Much of the structure of the research has been left unchanged, while significant changes will help refine the interpretation of research results.

Forecasts derived through research have been difficult to verify, due to the lack of in situ stratospheric turbulence observations. Objective verification of the MWFM, however, needs to be accomplished. Increased effort was made to rectify this situation, and is a focus of this research. The confidence of model output would be greatly increased if there is a high correlation between the model forecasts and the observed stratospheric turbulence reported via PIREPs from crews who had flown at the same time and location as that of the model. After great effort and many roadblocks, PIREPs were obtained from U2 crews flying over East Asia for comparison to MWFM output for a period of about 60 days. Other changes include collecting data and running the model for ten months, and increasing the time resolution of the MWFM to better coincide with PIREPs.

While Allen used both GFS and MM5 model data as input, this research only used GFS model output as input for the MWFM. This decision was made because, as a result of previous research, AFWA has begun using the MWFM, and has chosen to use GFS data as MWFM input.

1.3 Research Objectives

The ultimate goal of this research was to provide AFWA with a recommendation, which will aid them in determining the usefulness of the MWFM. Specific goals of this research are:

1) Run the MWFM using output from NCEP's GFS model at 00Z and 12Z over a period of nearly a full year,
2) Obtain PIREPs from U2 crews flying at stratospheric levels in the theatre being considered,
3) Collect rawinsonde sounding data over a period of nearly a full year to be analyzed by RAOB for comparison to model output,
4) Conduct objective analysis comparisons between MWFM output, PIREPs, and RAOB analyses,
5) Determine if one version of the MWFM is better than the other overall, or if out-performance of one version is correlated to topographical features of each location, and
6) Determine if weaknesses of the MWFM and RAOB can be identified, in order to aid forecasters in future operational use.

1.4 Research Approach

Model output from the GFS is required as input by the MWFM to develop atmospheric profiles of wind speed and direction, density and stability. The MWFM then produces a mountain wave activity forecast and determines locations of turbulence.

The best way to verify these determined locations of turbulence involves the use of real-time Stratoturb PIREPs. Unfortunately, PIREPs are only available for about half of the model runs. In order to provide an evaluation procedure that allows comparisons when PIREPs are unavailable, analyses based on rawinsonde balloon soundings are also used. These analyses are regularly used by flying units in East Asia. Data from these soundings are analyzed by the RAOB program, which is discussed in Chapter 3.

Several comparisons are then made. Since the MWFM has been released in two versions, their performances are compared. Since three different atmospheric levels are analyzed, MWFM performance is compared between these levels. The MWFM produced initial, 6-hr, 12-hr, 18-hr and 24-hr forecasts. MWFM performance is compared between these forecast hours. Since each of the ten locations being compared have widely varying upstream terrain, MWFM performance is compared between these locations. Since there are large seasonal variations in turbulence over East Asia, MWFM performance is compared between seasons. Comparison procedures are discussed in detail in Chapter 3. Results are discussed in Chapter 4. Recommendations are made in Chapter 5.

1.5 Research Challenges

Several challenges exist with this research. Primarily they concern whether or not the data used for evaluation and verification are a good representation of actual turbulence.

Setting up a method of collecting PIREPs was laborious. The first question that had to be answered was whether or not it is reasonable to expect there to be a consistent, reliable method of collecting PIREPs for use in this research and for continued verification after research is complete. The main reason Allen's research was unable to collect PIREPs was because of classified flight times and locations. The Assistant Director of Operations and the commander of the flying squadron located in the ROK were very willing to have their pilots report turbulence locally and pass this information in an unclassified way. Data were transmitted by breaking the theater into 6 large areas, and then transmitting PIREPs based on them. For example, "LGT-Turb at FL700 in sector B at 21/1200Z." Now that AFWA is running the MWFM, they have requested help in developing a method of verifying MWFM output. Establishing a PIREP reporting procedure for this research should lead to a reporting procedure from the flying unit directly to AFWA. A more detailed discussion about PIREPs and other interaction between AFWA and operational units is included in the recommendations section of this paper. Fortunately, headway made with this research will smoothly transition into a verification procedure for AFWA.

It must be determined whether or not turbulence was under-reported. There may be a tendency for pilots to only submit positive reports of turbulence. Further, it must be determined whether the turbulence reported in a PIREP represents the highest intensity of turbulence over an area. It is expected that each flight will not fly through the areas of the most severe turbulence over an area. The result would be turbulence reports which indicate less severe turbulence than what is the strongest in the area. Therefore, it was necessary to simplify the report of turbulence from specific levels of turbulence to a simple "yes" or "no", signifying that there either was, or was not turbulence present at a specific level and time. When PIREPs are not reported for each model run, as was the case in this research, and only reported positive reports of turbulence, analysis must proceed carefully. This is fully discussed in Chapter 3.

Finally, it should be considered that one of the two versions of the MWFM may perform better than the other based on topographical differences of locations being studied. Although gravity waves may be initiated by a number of means, the trigger we are most concerned with in this research is topography-induced mountain waves. Mountain waves may be analyzed using hydrostatic and non-hydrostatic models, depending on the situation. Because of this difference, the MWFM has been released in two versions. Version 1.1 is a two-dimensional hydrostatic gravity wave model, while Version 2.1 takes into account three-dimensional, non-hydrostatic effects on gravity waves. Recent research recommends further comparison between the two versions (Allen 2003). It may be concluded that Version 1.1 is a better forecast tool for locations with large-scale features, while Version 2.1 is a better forecast tool for locations with individual mountain peaks and ridges. A more detailed comparison of the two versions of the MWFM is made in Chapter 2.

Considering each of these anticipated research challenges, a collective recommendation was reached. Using these PIREPs, MWFM output can be objectively verified. Results show that the MWFM is a superior product to what has been used in the past. Of course, ideally, 100% agreement between MWFM output, RAOB analysis, and PIREPs is desired. However, disagreement between the products can be expected. Analysis of the resulting data from research should be able to show weaknesses of both tools, in order for future forecasters to better use both tools concurrently. The forecaster's skill will then be used to determine which tool is best for any particular situation.

II. Literature Review

Allen's research done on this topic provides an excellent analysis of mountain waves and mountain wave forecasts (Allen 2003). The literature review contained here is intended to augment that done by Allen. While Allen's research focused mainly on the dynamic principles involved in mountain waves, this literature review is focused on magnitude of scale and physical effects of terrain temporally and spatially followed by a discussion of dynamic principles regarding linear versus nonlinear analysis.

2.1 Mountain Waves

Gravity waves are disturbances in the atmosphere propagated by the force of buoyancy (Wurtele et al. 1993). Mountain waves are simply gravity waves forced by terrain features. These waves are known to propagate into the lower stratosphere even when initiated by individual islands in the middle of the ocean (Balsley and Carter 1989). It has been shown that any mountain higher than about 1 km, no matter how gentle the slope, can, under typical atmospheric conditions, produce waves too large for linear theory (Smith 1977).

Recognition that turbulence in the middle and upper atmosphere was likely to have its source in atmospheric gravity waves came in the early 1960s (Hines 1963), the emphasis then being placed on the requirement for a dynamic instability and on the wind shears that might produce this condition (Hines 1988a). The induced turbulence causes danger to aircraft flying at tropospheric and lower stratospheric levels (Wurtele 1993, Weinstock 1987).

When the mean wind flow passes over mountainous terrain, air is displaced vertically, transporting momentum and energy with it. The behavior of this displaced air depends on several inter-related factors, most importantly buoyancy, whether or not the environment is hydrostatic, and the size and shape of the mountains. In a stably stratified environment, this displaced air will tend to sink, returning to its equilibrium level. While continuing to sink and move downstream, this air will tend to overshoot its equilibrium level. Its downward vertical velocity will lessen as it continues to sink once it has overshot its equilibrium level. This results in a tendency to eventually rise once again to its equilibrium level. As this oscillation repeats, it is damped and eventually becomes insignificant in comparison to the mean flow. This oversimplification of the oscillation process describes what is termed mountain waves.

2.2 Characteristics of Mountain Waves

There are many characteristics of mountain waves that must be considered when performing analysis, including whether or not the flow may be analyzed by assuming the environment is hydrostatic, and the shape, extent and orientation of the terrain in relation to mean wind flow. Once a wave breaks, it must be studied differently.

Complexities of mountain flow have contributed to the limited development of useful forecasting tools. This is due, in part, to the complicated nature of topographic forcing. Topography is anisotropic (i.e., there are directional differences in magnitude and scale of topographic variance), which is evident in the fact that much of the earth's topography is organized into long, narrow ridges. The orientation of these ridges is an important variable in determining the wave response of the atmosphere to topography (Bacmeister 1994).

Atmospheric winds are usually decelerated to some degree when passing over rough terrain. If the fluid is not thermally stably stratified the slowing down of the winds occurs by the effect of eddies developing in the region of the rough terrain, moving away from the surface, causing momentum to be moved vertically. In stably stratified fluids a more subtle but still effective process occurs through the agency of vertically and horizontally propagating internal gravity waves. Generated by the flow over the surface roughness elements, especially mountains, such waves transport momentum downward through the otherwise undisturbed fluid by means of pressure forces (Lilly and Kennedy 1973).

Gravity waves can exist only in the atmosphere under stably stratified conditions. Then, a fluid parcel displaced vertically will undergo buoyancy oscillations (Holton 1992). The stratosphere is a very stratified region of the atmosphere. In the lower stratosphere, temperature tends to be nearly isothermal, and wind speed generally decreases to minimum values at altitudes between 20 and 25 km. Buoyancy forces in the stratosphere act as a stiff spring, and are much stronger than in either the troposphere or the mesosphere (Ehernberger 1992). Mountain-wave-induced vertical velocity perturbations persist for many hours (Balsley and Carter 1989). The deceleration or drag effect on the atmosphere may actually appear at a distance of many kilometers above or beyond the mountain, often in excess of 30 km downstream (Balsley and Carter 1989, Lilly and Kennedy 1973).

Clear air turbulence (CAT) probabilities are significantly higher over mountains than flat terrain. Over mountains, the probability of CAT is greatly increased by large temperature gradients (Bender et al. 1976). The role of lower altitude wave activity has been empirically established for a significant portion of high altitude turbulence cases encountered by both subsonic and supersonic aircraft. It has been analyzed and demonstrated that CAT enhancement by mountain-wave-induced vertical displacement of shear layers causes Kelvin-Helmholtz wave amplification and instability. Available data for turbulence encountered by aircraft in the lower stratosphere often show an association with lower-altitude mountain-wave activity (Ehernberger 1992). Therefore, it is important to gain understanding of upward wave propagation processes in order to understand, study, and forecast turbulence.

2.3 Mountain Wave Propagation

Mountain wave propagation may occur horizontally and vertically. When propagating horizontally, mountain waves propagate into a parabolic-shaped region that spreads outward transverse to the mean flow as the disturbances move downstream (Smith 1980, Hines 1988b). This parabolic-shaped region will be influenced by the size and shape of the mountain ridge. In other words, flow over a wide, broad ridge will have a significantly different downstream region of influence when compared to flow of equal magnitude over a single peak or a narrow ridge.

Some gravity waves propagate vertically, while others are evanescent, or trapped. Stable stratification, wide ridges, and comparatively weak zonal flow provide favorable conditions for the formation of vertically propagating topographic waves. In this case, the line of maximum upward displacement tilts back (upstream) and amplitude is independent of height (Figure 1).

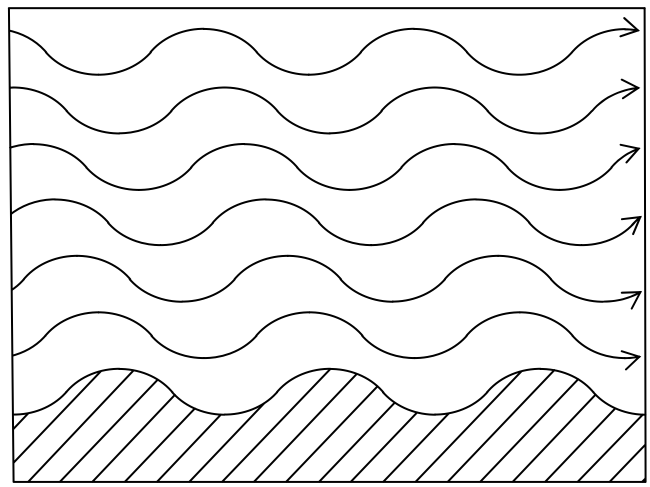

Figure 1. Vertically Propagating Waves. Vertical motion phase tilts with height. Wave amplitude is independent of height.

Thus, vertically propagating waves are not in phase at all heights. This is true of mountain ridges having a characteristic width of 50 to 200 km. For vertically propagating waves, the vertical wave number is real. Whether or not the flow may be analyzed by assuming the environment is hydrostatic is determined by the relationship between the mountain width parameter L, and U/N, where U is the mean horizontal flow and N is the Brunt-Väisälä frequency. If L is greater than U/N (but not so large that rotational effects are important), the flow may be analyzed by assuming the environment is hydrostatic, provided that the generated waves have

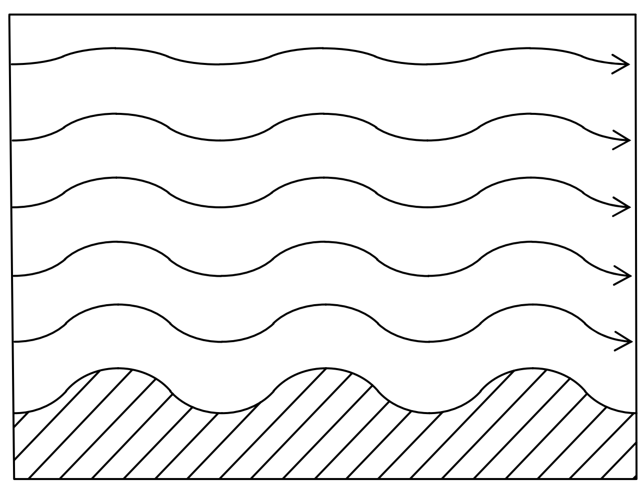

Figure 2. Evanescent Waves. Vertical motions are in phase at all heights. Wave amplitude decreases with height.

horizontal wavelengths much greater than their vertical wavelengths. In two-dimensional hydrostatic flow, mountain waves are found only above mountains, since the group velocity is vertical (Ehernberger 1992). Since the energy source for these waves is at the ground, they transport energy upward.

Narrow ridges and isolated peaks, on the other hand, provide favorable conditions for the formation of evanescent topographic waves. The maximum upward displacement occurs at the ridge tops and the amplitude of the disturbance decays with height (Figure 2). For evanescent waves, the vertical wave number is complex. The real part of the vertical wave number describes the sinusoidal variation in the vertical, and the imaginary part describes exponential growth or decay, depending on whether it is positive or negative. For evanescent waves, vertical motions are in phase at all heights, and the rate of decrease of intensity with height is inversely proportional to the wavelength (Gill 1982).

When waves transport momentum and energy upward, wave instability and breakdown occur. A substantial drag may be exerted on upper level circulation and associated CAT may be hazardous to aircraft. At certain locations in the lee of large mountain ranges, intense and damaging surface winds arise when these waves attain large amplitude (Klemp and Lilly 1978). Therefore, resulting turbulence depends on the amplitude of these waves and whether or not instability and breakdown can be expected.

2.4 Linear Versus Nonlinear Analysis

Linear analysis of wave propagation has provided ample insight into wave behavior. Even so, nonlinearity of these systems must be considered.

The concept of "wave breaking" implied by the onset of static instability has been widely employed using linear perturbation theory in middle-atmosphere studies, as by Lindzen (1981) in his modeling of momentum deposition by gravity waves. There are two serious limitations to this concept.

The first limitation concerns the fact that the middle-atmosphere spectrum of waves is only rarely represented by a single dominant member, so competing processes lead to nonlinear solutions. For example, if the amplitude of horizontal perturbation speed equals the horizontal phase trace speed, which can reasonably be expected, conditions are inadequate to produce nonlinearity in a single wave mode. However, it imposes severe nonlinearity when two modes having significantly different wave numbers are present simultaneously, even if their amplitudes are comparable (Hines 1960). The resulting nonlinear interaction may draw away wave energy and propagate it away in advance of the onset of instability (Hines 1988a). This situation is most often the case when considering mountain-induced waves.

Studies have shown that, even when confined to a single wave, there is a further limitation. When waves are analyzed as having only vertical gradients in the wave system are often also found to possess horizontal gradients which may be just as large in amplitude. Therefore, these waves should be considered to be relevant to the generation of turbulence, producing slantwise instability, whether static or dynamic (Hines 1971). Conclusions have been reached that the criterion for slantwise static instability is less demanding than that for vertical static instability. However, the time-scale of growth of slantwise static instability is inherently long, and is perhaps too long for the instability to develop into turbulence, if the dynamics of a wave system are continually changing (Hines 1960).

Simulations for small and fairly large amplitude mountains have been compared to linear and nonlinear analytic solutions for a one-layer atmosphere to test the validity of the numerical representations. Simulations of real data cases using a linear steady-state hydrostatic model demonstrated a strong positive correlation between model results and observations using the intensity of surface winds as the basis for comparison (Klemp and Lilly 1978).

Therefore, linear analysis is often adequate for study, as long as it is realized that nonlinearities may exist and have been considered to have negligible impact on the area of study.

2.5 The Mountain Wave Forecast Model (MWFM)

NRL's MWFM was released in two versions. Version 1.1 is described in the next section, followed by the major differences between Versions 1.1 and 2.1.

2.5.1 Version 1.1 Version 1.1 is a hydrostatic gravity wave model. Topographic data used in this version includes latitude, longitude, ridge orientation, altitude, and width of ridges. Topographic forcing in this version is based on a box-by-box analysis of topographic features with scales between 50 and 100 km, with only one ridge assumed to be within each grid box. No attempt was made to segregate features by width or to identify features smaller than 50 km. First, the profile of the wind component perpendicular to a ridge is calculated using the ridge orientation. Then a profile of the stratification frequency or buoyancy frequency above each ridge $N_k(z)$ is estimated from the local potential temperature profile $\Theta_k(z)$ according to

$$N_k(z) = \sqrt{\frac{g}{\Theta_k}\frac{\partial \Theta_k}{\partial z}} \,. \tag{1}$$

Waves launched by each ridge are assumed to be in steady state and purely two-dimensional with wave crests parallel to the generating ridge at all levels. The atmosphere is assumed to be hydrostatic so that the wave activity is generally localized over forcing topography. The average momentum flux profile over any ridge can be approximated in terms of the wave vertical displacement profile

$$\varphi_k(z) = \alpha\rho(z)N_k(z)U_{\perp;k}(z)\frac{\delta_k(z)^2}{L}\,, \tag{2}$$

where α is a dimensionless factor that depends on ridge shape, $\rho(z)$ is the background atmospheric density profile, which is assumed to be proportional to pressure, $U_{\perp;k}$ is the component of the horizontal wind which is perpendicular to the k^{th} ridge, $\delta_k(z)$ is the profile of the wave-induced vertical profile above the k^{th} ridge, and L is the horizontal length representing the extent of the wave disturbance. Wave momentum flux is assumed to remain constant with height until wave breaking occurs. This constant vertical wave momentum flux allows the environment to be analyzed using the hydrostatic assumption. The criterion for wave breaking is based on simulations of two-dimensional flow over topography, which suggests that wave amplitudes don't exceed the local saturation limit given by

$$\delta_{sat;k}(z) \equiv \frac{U_{\perp;k}(z)}{N_k(z)}. \qquad (3)$$

After the average momentum flux profile is approximated and the criterion for wave breaking is determined, an approximate wave displacement profile is constructed using (2) and (3). Wave-induced turbulence is forecasted whenever saturation is invoked to limit wave amplitudes. The intensity of turbulence is assumed to be proportional to the amount of momentum flux lost by the wave within that layer. The disadvantage of this version of the model is due to the hydrostatic assumption. Therefore, this version doesn't treat narrow ridgelines correctly, leading to over-forecasting the intensity of the mountain waves directly over narrow ridgelines, which should produce evanescent waves, and under-forecasting the intensity downstream, as waves are limited to the vertically propagating type (Bacmeister et al. 1994).

2.5.2 Version 2.1 Unlike Version 1.1, horizontal wavenumbers must be computed explicitly in Version 2.1 calculations. Each ridge is assigned two wavenumber harmonics. For each harmonic, rays are launched at six equispaced azimuths with respect to the ridge axis angle, yielding wave vectors spanning the 180 degree range. So, twelve rays are launched from each ridge feature, with the largest amplitude assigned to the ray directly orthogonal to the long axis orientation of the ridge. Rays at other angles are scaled down in amplitude according to the shape of the ridge (Marks and Eckerman 1995).

2.6 Statistical Terminology

A typical contingency table used to verify turbulence forecasting is presented in Table 1. In this table, a and d represent correct forecasts for turbulence and no turbulence, respectively, while b and c represent incorrect forecasts for turbulence and no turbulence, respectively. The total number of forecasts compared to observations is represented by N. Statistical analyses used in Chapters 3 and 4 are based on this convention. Terms describing these statistics are introduced in Table 2.

Table 1. Typical Turbulence Forecast Verification Contingency Table

		Observation		
		Yes	No	Total
Forecast	Yes	a	b	
	No	c	d	
	Total			N

2.7 Using PIREPs to Verify Turbulence Forecasts

Pilot reports (PIREPs) are often used to verify forecasts of turbulence. Even though they have many characteristics that make them difficult to use for verification, they are still the best observations currently available for evaluation of turbulence forecasts (Brown and Young 2000).

Table 2. Common Contingency Table Statistics (Wilks 1995)

Hit rate (HR)	Proportion of all forecasts which were forecasted correctly	$\frac{a+d}{N}$
False Alarm Rate (FAR)	Proportion of forecasted turbulence events which were forecasted incorrectly	$\frac{b}{a+b}$
Probability of Detection (POD)	Proportion of turbulence events which were forecasted correctly	$\frac{a}{a+c}$
Bias	Measures over- and under-forecasting	$\frac{a+b}{a+c}$
Critical Success Index (CSI)	Proportion of forecasted and/or observed turbulence events which were forecasted correctly	$\frac{a}{a+b+c}$
Heidke Skill Score (HSS)	Proportion of correct forecasts after eliminating those which would be correct due to chance	$\frac{2(ad-bc)}{(a+c)(c+d)+(a+b)(b+d)}$
Chi-Squared (χ^2) Test for homogeneity or statistical significance	Test whether the proportions for each cell in the contingency table are equal across both populations or independent	$\frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$

Among the characteristics which make PIREPs difficult to use, are their subjective nature and their spatial and temporal biases (Kane et al. 1998). Further, unlike METAR observations and RAOB soundings, it cannot be know in advance whether PIREPs will be reported over a particular location, at a particular elevation, or at a particular time. If PIREPs were to be compared to a forecast grid, this inconsistent reporting of PIREPs would not provide a representative sample of the forecast grid.

It has been shown that it is inappropriate to calculate the false alarm rate (FAR), bias and various other measures when using PIREP data for verification of turbulence forecasts (Brown 1996). Turbulence forecast verification techniques usually employ the use of a contingency table (Table 1). If the statistics in Table 1 are considered to be functions of the joint distribution of forecasts and observations, then this joint distribution can also be represented by various conditional and marginal distributions and probabilities. For example, the probability of detection of Yes observations (PODy) given by $\frac{a}{a+c}$ is an estimate of the conditional probability that the forecast is Yes, given that the observation is Yes. That is, PODy is conditioned on the observations. On the other hand, FAR is an estimate of the probability that the observation is No, given that the forecast is Yes. That is, FAR is conditioned on the forecasts. This conditioning leads to difficulties, since PIREPs do not adequately sample the forecast grid (Brown 1996).

Therefore, FAR is strongly related to the relative frequencies of Yes and No PIREPs. That is, when either the number of Yes or No PIREPs is changed, the FAR also changes. In contrast, some other statistics, such as POD, change very little when the number of PIREPs change. The underlying difficulty is that the distribution of Yes and No PIREPs at any given time is unlikely to appropriately represent the actual distribution of turbulence in the atmosphere. But FAR is not the only statistic that is strongly affected. Others include the bias, the Critical Success Index and the Heidke skill scores (Brown 2000). Therefore, when comparing MWFM output with PIREPs, conclusions made from this research are made only using the hit rate and probability of detection.

2.8 Summary

It is widely known that mountain waves are gravity waves induced by flow over rough terrain. Even terrain that is comprised of small mountains or islands in the ocean is known to propagate waves into the stratosphere. Aspects of these waves must be carefully considered when turbulence analysis is performed in order to determine whether a hydrostatic or nonhydrostatic model is used and whether to use linear or nonlinear techniques. It is reasonable to expect that model output can then be used to determine locations of Stratoturb.

III. Methodology

3.1 Overview

One of the goals of this research is to verify the reliability of the Stratoturb forecasts produced by the MWFM. The efforts of this research are a continuation and expansion of research performed by Capt Mark Allen in the fall of 2002. Throughout each of the following sections, Capt Allen's research is summarized, followed by a summary of this research, detailing the changes that were made. This research used model data beginning in early April 2003, and continued through mid-March 2004.

3.2 Data

During Allen's research, MWFM output was compared to RAOB analyses, providing insight to the consistency of the MWFM. During this research, MWFM output was also compared to PIREPs. Although PIREPs are not always available, and are subject to pilot acuity, they are the only observational data available for use. Since operational units use the High Altitude Clear Air Turbulence (HiCAT) output from the RAOB program as their primary tool in determining Stratoturb, it was also used during this research for comparison with MWFM output.

3.2.1 MWFM Input Data The MWFM requires data from a larger scale model, namely, it requires absolute temperature, geopotential height, and zonal and meridional components of the wind. These data are used to create atmospheric profiles of wind speed, wind direction, density, and stability in order to make the wave forecasts. Allen's research used output from both the GFS and the MM5. Output from the GFS model is currently used at AFWA as model input to the MWFM, and so was chosen as the only data source for this research. The GFS data used as input to the MWFM were downloaded from NCEP via ftp twice daily, at 00Z and 12Z. The MWFM extracted the required input data using WGRIB, a GRIB data file management program. While Allen's research used model forecasts made in twelve-hour intervals through the 48-hr forecast point, this research used model forecasts made in six-hour intervals through the 24-hr forecast point. Data covered the entire globe at a $1° \times 1°$ resolution, up to the 10mb level.

3.2.2 RAOB Input Data During both Allen's research and this research, sounding data were obtained for the same ten locations from the same source and analyzed the same way. Rawinsonde sounding data were obtained from the University of Wyoming and the Florida State University archives in text format. Ten stations in East Asia were chosen for comparison. Sounding data were collected twice daily, at 00Z and 12Z. These stations are listed in Table 3; their locations are shown in Figure 3.

The raw sounding data include the temperature, dew point temperature, pressure, wind speed and wind direction. Temperature and pressure measurements are used by RAOB. Equipment failure or premature popping of the balloon may prevent data collection for the entire atmospheric column. Since comparisons were made based on atmospheric layers of 100-70mb, 70-50mb, and 50-30mb, if data were not collected for

Table 3. East Asia Rawinsonde Stations

WMO Number	Station Name	Country	Latitude (N)	Longitude (E)
47122	Osan AB	ROK	37° 06'	127° 02'
47158	Kwangju AB	ROK	35° 07'	126° 49'
47138	Pohang	ROK	36° 02'	129° 23'
47580	Misawa AB	Japan	40° 41'	141° 23'
47681	Hamamatsu AB	Japan	34° 44'	137° 40'
47412	Sapporo	Japan	43° 03'	141° 20'
47600	Wajima	Japan	37° 23'	136° 54'
47646	Tateno	Japan	36° 03'	140° 08'
47778	Shionomisaki	Japan	33° 27'	135° 46'
47807	Fukuoka	Japan	33° 35'	130° 23'

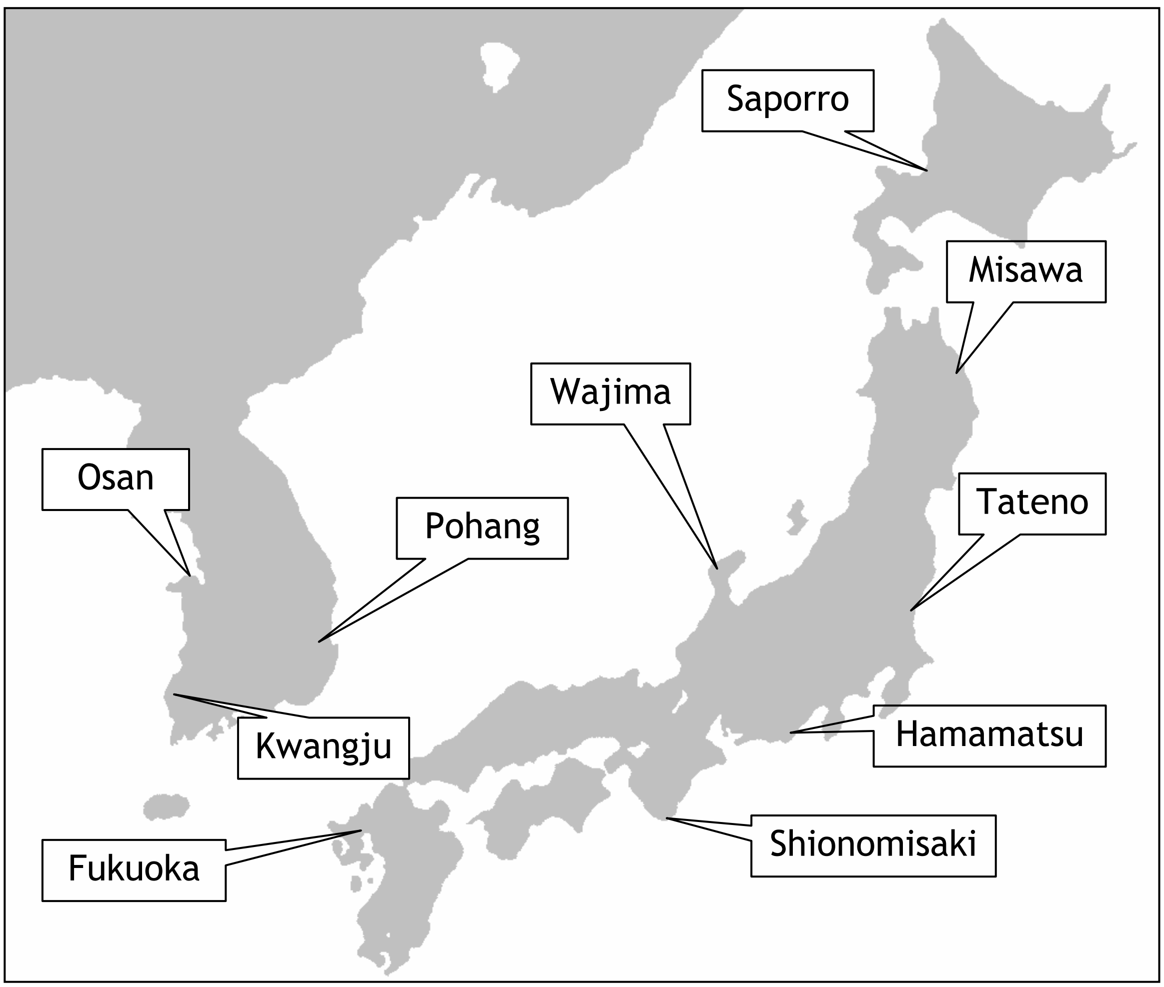

Figure 3. East Asia Rawinsonde Stations

an entire layer, the entire layer was considered missing, unless the partial layer's analysis indicated turbulence.

The RAOB program analyzes data to identify the existence of turbulence by looking for three distinct layers in the atmosphere, such that the upper and lower layers are inversions which have a mixing layer in between, forming an 'S' shape on the temperature trace sounding, as shown in Figure 4. Sinclair and Kuhn (1991) showed that the 'S' layer model was verified 93.8% of the time, and that all of the turbulence identified was within the mixing layer part of the 'S' layer. Further analysis has shown that turbulence intensity is directly related to the mixing layer temperature lapse rate and the vertical temperature difference, while intensity is inversely related to the depth of the 'S' layer. The intensity of the turbulence is depicted by the width of the rectangular area along the left vertical axis in Figure 5 (Sinclair and Kuhn 1991).

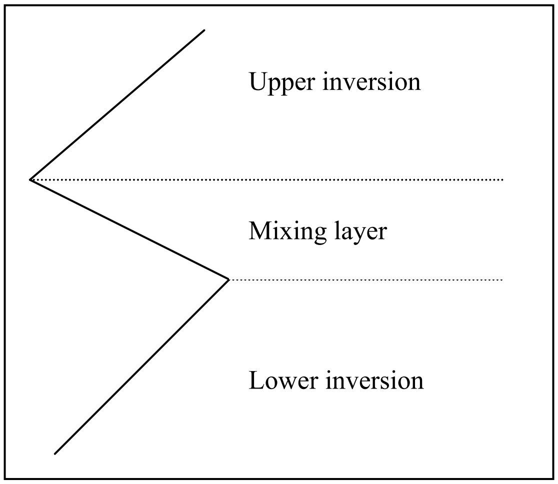

Figure 4. The 'S' Layer. Turbulence has been highly correlated with the mixing layer.

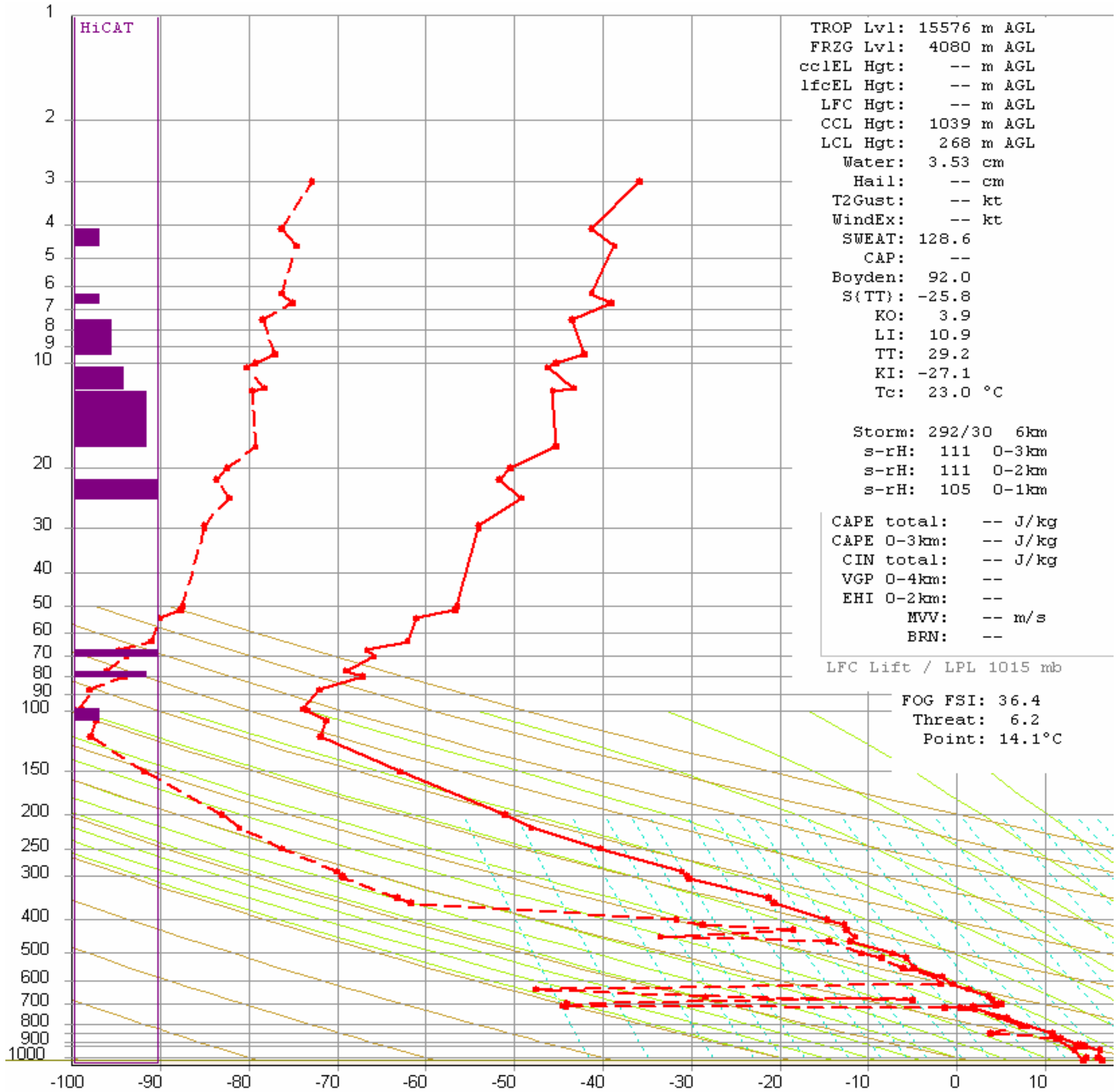

Figure 5. RAOB Graphical Turbulence Analysis. The dark colored blocks on the left side of the diagram show the HiCAT analysis. The extension of the bars towards the right indicates the intensity level of turbulence. From this image, the use of the 'S' layer model is evident, with turbulence located in the mixing layers.

3.2.3 PIREPs PIREPs were not collected during previous research due to data classification issues, which were resolved during the early stages of this research. Dates, times and specific locations of aircraft flights over the ROK are classified. In order to transmit reports of turbulence without compromising classified data, the general area of the ROK was divided into sectors. PIREPs from flights over these sectors were transmitted in an unclassified way. Sectors were designated by the flying unit in the ROK and were made as detailed as possible for comparison during this research without being so specific that classified data were compromised. Classification issues and the operational tempo of the unit prevented transmission of PIREPs during the summer and fall months. PIREPs were transmitted from December 2003 through the end of the research period.

3.3 Comparison Procedure

During Allen's research, MWFM forecasts were accomplished twice daily extending through 48 hours at 12-hour intervals. During this research, MWFM forecasts were accomplished twice daily extending through 24 hours at 6-hour intervals. MWFM forecast data were collected for nearly an entire year. Rawinsonde data for the ten selected locations (Table 1) were collected at 00Z and 12Z for each day the model was run, allowing an extra day at the end to allow for comparison of the final forecast day. Although both graphical and text output were produced by the MWFM, text data from both versions at each forecast time period were used for comparison to provide objective analysis.

The RAOB program was used to analyze each of the rawinsonde soundings for the presence of HiCAT. See Figure 5 for an example of graphical output of HiCAT layers analyzed by RAOB. For this research, turbulence was considered to be present in a layer if a graphical indication was located anywhere within the layer.

PIREPs were collected on a nearly daily basis during the last three months of the research period. For this research, turbulence was considered to be present in a layer if a PIREP reported any level of turbulence.

For each station's RAOB analysis, the presence of turbulence within each layer was recorded as either 'yes' or 'no'. For each PIREP, the presence of turbulence within each layer was also recorded as either 'yes' or 'no'. Similarly, a 'yes' or 'no' was assigned to each forecast time, MWFM version, layer and station, with a 'yes' based on the presence of momentum flux deposition within a 1.5° × 1.5° box over each station. These boxes were positioned so that 90% of the area of the box was located downwind of the station, in order to capture turbulence forecasts from the most likely environment of the actual rawinsonde flight.

3.4 Statistical Methodology

Comparison techniques used during Allen's research were very similar to those employed during this research. Since PIREPs were not available for comparison during Allen's research, a few changes needed to be made. Comparisons techniques used to compare MWFM forecasts and RAOB analyses were also applied to comparisons between MWFM forecasts and PIREPs. Careful interpretation must be used when analyzing these statistics. Ideally, MWFM forecasts, RAOB analyses and PIREPs will indicate turbulence to a high degree of consistency. If the PIREPs are not reported in a consistent manner, and there is an overwhelmingly high percentage of PIREPs that indicate turbulence, care must be taken to determine whether or not pilots are simply submitting PIREPs only when they encounter turbulence. Since neither was the case during the research period, it is assumed that the PIREPs were reported objectively. Care must also be taken when considering the level of turbulence reported. For example, since pilots avoid flying through areas of known strong turbulence, it is very reasonable to conclude that a report of no turbulence or light turbulence may be reported when there is actually stronger turbulence in the sector. Considering this uncertainty, it may be reasonable only to consider comparison of data where less than severe or extreme turbulence is forecasted or analyzed. Another approach may be to simplify categories of turbulence. For example, it may be reasonable to consider it a "success" when a PIREP reports any turbulence and the MWFM forecasts any turbulence, regardless of the severity as well as when turbulence is not reported via PIREP and is not forecasted by the MWFM. This approach was used during this research.

When comparing MWFM output with RAOB analysis, 00Z, 06Z, 12Z, 18Z, and 24Z data were used. When comparing PIREPs with MWFM forecasts, the model forecast hour that was closest to the time of the PIREP was used. Results of comparisons were arranged in contingency tables (Figure 6) for analysis and testing.

These contingency tables were tested using the chi-squared (χ^2) test in order to determine significance of the results they present. Contingency tables determined to be statistically insignificant show no dependence between the factors, but significance implies that the numbers were not generated by chance, and that the values have some meaningful interpretation. If the contingency tables are found to be statistically significant, the cells of the table may be used to compute several measures of accuracy and skill. During this research, hit rate (HR), critical success index (CSI), false alarm rate (FAR), probability of detection (POD), Heidke Skill Score (HSS) and bias were all computed. The HR gives the percentage of the total number of forecasts resulting in a correct forecast, whether forecasting turbulence or not forecasting turbulence. The CSI is similar to the HR, however the incorrect forecasts for no turbulence are included. The FAR is the proportion of forecasts of turbulence, when turbulence did not occur. The POD is the percentage of events in which turbulence was both forecasted and observed. The HSS compares the results in the contingency table to a random forecast. The range of possible HSS is from -1 to +1, with a negative HSS representing a forecast that has less skill than a randomly-based forecast. Together, these indices help determine the amount of agreement between the MWFM forecasts and RAOB analyses and between MWFM forecasts and PIREPs.

		RAOB analysis	
		Yes	No
MWFM forecast	Yes	a	b
	No	c	d

		PIREPs	
		Yes	No
MWFM forecast	Yes	a	b
	No	c	d

		MWFM Ver 2.1	
		Yes	No
MWFM Ver 1.1	Yes	a	b
	No	c	d

Figure 6. Contingency Tables. These tables were used to compare MWFM forecasts with RAOB program analyses (top left), MWFM forecasts with PIREPs (top right), and MWFM Version 1.1 forecasts with MWFM Version 2.1 forecasts (bottom). The top two tables were used twice, once for MWFM Version 1.1 and once for MWFM Version 2.1. "Yes" represents a positive turbulence forecast, analysis or PIREP; "No" represents a negative turbulence forecast, analysis or PIREP.

IV. Results

4.1 Introduction

This chapter presents a summary of the contingency table statistical analyses conducted during this research. The analyses show differences between forecasts created using the two MWFM versions, and how they compare to RAOB analysis and PIREPs.

The MWFM forecasts were divided several ways, between MWFM version used, atmospheric layer, forecast hour, sounding location, and season of the year. Therefore, several comparisons needed to be analyzed. MWFM Version 1.1 forecasts were compared to Version 2.1 forecasts in order to determine if the forecasts were different, and if one version compared better to RAOB analysis and PIREPs than the other version. While comparing Version 1.1 to Version 2.1, it was also important to determine if the two versions' forecasts differed from one atmospheric layer to another, from one forecast hour to another, from one sounding location to another, and from one season of the year to another.

Before making comparisons between atmospheric layers, forecast times, sounding locations and seasons, the χ^2-test for significance was employed to determine statistical significance of the data in the contingency table. In each of the following sections, results of the χ^2-test for significance are summarized, followed by a brief discussion of the resulting statistics. Whenever the χ^2-test for significance showed that the data in a particular contingency table are insignificant, the Fisher Exact Test was also employed. In every instance, the Fisher Exact Test gave the same results as the χ^2-test for significance, unless otherwise noted. For both tests, a *p*-value of 0.05 was used in all cases, unless otherwise noted.

4.2 Comparisons Between MWFM Forecasts and RAOB Analyses

4.2.1 Comparison of MWFM Versions The χ^2-test for homogeneity tests whether or not the proportions for each class are equal across two populations and whether or not this is true for each class. This test was performed on contingency tables which represented forecasts performed by MWFM Versions 1.1 and 2.1, which were compared to RAOB analysis.

When comparing all forecasts based solely on model version, all forecasts were compiled into a contingency table, which was analyzed for homogeneity using the χ^2-squared test, which indicated that there was a statistically significant difference between the two versions.

Therefore, with high confidence, it can be stated that the two versions of the MWFM produce significantly different turbulence forecasts.

4.2.2 Atmospheric Layer Comparison When separating the dataset based on atmospheric layer, all forecasts were divided into six contingency tables; first by model version, then by atmospheric layer. The χ^2-test showed that all six tables were statistically significant. Table 4 shows the various accuracy measurements for the three atmospheric layers for Versions 1.1 and 2.1.

It is interesting to note that, for the most part, performance increases with height. This is true for both model versions, with the exception being the FAR of Version 2.1.

Table 4. Accuracy and Bias by Version and Level

		HR	CSI	FAR	POD	HSS	Bias
	50-30	65.45	41.29	31.58	51	0.3	0.75
Ver 1.1	70-50	56.44	32.45	39.51	41.18	0.13	0.68
	100-70	42.53	19.56	44.46	23.19	-0.04	0.42
	50-30	67.68	54.78	38.04	82.55	0.36	1.33
Ver 2.1	70-50	54.06	44	46.37	71.03	0.08	1.32
	100-70	47.34	37.46	43.16	52.35	-0.08	0.92

4.2.3 Forecast Hour Comparison When separating the dataset based on forecast hour, all forecasts were divided into ten contingency tables; first by model version, then by forecast hour. The χ^2-test showed that nine of the ten tables were statistically significant. The 18-hour forecast hour for Version 2.1 was shown not to be statistically significant. Table 5 shows the various accuracy measurements for the five forecast hours for Versions 1.1 and 2.1.

Table 5. Accuracy and Bias by Version and Forecast Hour

		HR	CSI	FAR	POD	HSS	Bias
	00 Hour	53.77	28.35	39.55	34.81	0.09	0.58
	06 Hour	52.78	33.47	29.94	39.06	0.12	0.56
Ver 1.1	12 Hour	55.03	29.96	39.02	37.07	0.11	0.61
	18 Hour	52.38	33.83	30.22	39.64	0.11	0.57
	24 Hour	55.58	30.81	38.52	38.18	0.12	0.62
	00 Hour	56.98	45.68	42.41	68.83	0.13	1.20
	06 Hour	53.88	42.81	36.48	56.77	0.06	0.89
Ver 2.1	12 Hour	56.67	45.08	43.16	68.53	0.13	1.21
	18 Hour	51.94	41.03	37.51	54.43	0.02	0.87
	24 Hour	55.81	44.43	44.13	68.44	0.11	1.22

It is interesting to note that forecast accuracy does not decrease with time, as one might expect. Instead, the statistics stay relatively fixed in time. It is also interesting to note that the statistics representing forecasts for the 06 Hour and the 18 Hour forecasts vary inconsistently. That is, for these forecast hours, HR, FAR, and Bias are consistently lower for both model versions, while CSI and POD are higher for Version 1.1, but lower for Version 2.1. The reason for this variance is the fact that out of the ten sounding locations, only two locations produce soundings at 06Z and 18Z. These are Kwangju, Japan and Osan, ROK. If all ten sounding locations produced 06Z and 18Z soundings, these statistics would most likely be more consistent.

4.2.4 Sounding Location Comparison When separating the dataset based on sounding location, all forecasts were divided into twenty contingency tables; first by model version, then by sounding location. The χ^2-test showed that eighteen of the twenty tables were statistically significant. The data from Pohang, ROK and Shionomisaki, Japan, Version 1.1 were shown not to be statistically significant. Table 6 shows the various accuracy measurements for the ten sounding locations for Versions 1.1 and 2.1.

A close look shows that the statistics representing forecasts for Wajima and Sapporo tend to be slightly better than those from the other stations, those representing forecasts for Misawa tend to be in the middle of the pack, while the statistics representing Version 2.1 forecasts for Shionomisaki tend to be worse than those from the other stations. Statistics representing Version 2.1 forecasts for Osan and Pohang were noticeably better than Version 1.1 forecasts for these locations. This is most likely due to the highly mountainous terrain surrounding these two locations. Since Version 2.1

Table 6. Accuracy and Bias by Version and Sounding Location

		HR	CSI	FAR	POD	HSS	Bias
Ver 1.1	Fukuoka	56.86	27.26	51.83	38.57	0.09	0.80
	Hamamatsu	58.98	34.47	49.76	52.33	0.16	1.04
	Kwangju	55.51	40.58	28.31	48.33	0.14	0.67
	Misawa	55.35	33.78	39.41	43.30	0.12	0.71
	Osan	50.61	27.93	27.22	31.19	0.11	0.43
	Pohang	34.24	0.51	29.41	0.51	0.00	0.01
	Sapporo	54.86	33.05	36.21	40.68	0.12	0.64
	Shionomisaki	60.64	0.00	—	0.00	0.00	0.00
	Tateno	55.65	24.70	48.53	32.19	0.07	0.63
	Wajima	62.84	50.38	35.96	70.26	0.25	1.10
Ver 2.1	Fukuoka	52.12	34.39	55.31	59.86	0.06	1.34
	Hamamatsu	52.07	38.91	54.94	74.05	0.10	1.64
	Kwangju	53.91	42.73	33.86	54.69	0.07	0.83
	Misawa	57.22	46.42	42.35	70.44	0.13	1.22
	Osan	56.78	46.08	34.11	60.51	0.11	0.92
	Pohang	58.39	50.78	30.18	65.06	0.10	0.93
	Sapporo	59.04	50.11	39.92	75.11	0.15	1.25
	Shionomisaki	51.70	31.52	58.37	56.49	0.05	1.36
	Tateno	56.17	45.59	49.07	81.31	0.16	1.60
	Wajima	62.30	54.09	39.02	82.73	0.22	1.36

allows propagation of turbulence downstream, it is more likely to model the atmosphere more accurately at these locations when compared to Version 1.1.

There is an interesting fact about the Version 1.1 forecasts for Shionomisaki. Version 1.1 never forecasted Stratoturb over Shionomisaki, which is the reason for the inability to calculate the FAR. When taking the terrain into consideration, this is reasonable, since Version 1.1 is a hydrostatic model and does not take into account any turbulence propagating downstream as Version 2.1 does.

4.2.5 Seasonal Comparison When separating the dataset based on season, all forecasts were divided into twenty contingency tables; first by model version, then by month. The

χ^2-test showed that twelve of the twenty tables were statistically significant. The Fisher Exact Test showed that fifteen of the twenty tables were statistically significant. Table 7 shows the various accuracy measurements for the months of April 2003 through Jan 2004 locations for Versions 1.1 and 2.1.

Table 7. Accuracy and Bias by Version and Month

		HR	CSI	FAR	POD	HSS	Bias
Ver 1.1	Apr*	45.16	32.51	33.30	38.81	-0.02	0.58
	May**	50.60	15.73	59.62	20.49	-0.04	0.51
	Jun	65.84	17.35	68.31	27.71	0.07	0.87
	Jul	75.13	11.33	80.65	21.48	0.06	1.11
	Aug	71.85	16.80	70.34	27.93	0.11	0.94
	Sep*	56.28	21.22	61.90	32.39	0.02	0.85
	Oct	47.60	30.55	31.13	35.45	0.05	0.51
	Nov*	41.72	24.00	29.80	26.72	0.01	0.38
	Dec*	43.45	30.26	25.96	33.85	0.02	0.46
	Jan	47.08	40.59	13.15	43.25	0.05	0.50
Ver 2.1	Apr	57.66	54.50	33.00	74.50	-0.04	1.11
	May*	47.78	34.17	55.86	60.19	-0.02	1.36
	Jun	54.03	27.93	68.03	68.84	0.13	2.15
	Jul	56.09	18.42	79.74	67.01	0.11	3.31
	Aug	61.42	24.52	71.07	61.64	0.16	2.13
	Sep	43.95	29.03	65.01	63.02	-0.03	1.80
	Oct**	54.63	47.33	34.12	62.70	0.02	0.95
	Nov	50.92	44.10	32.83	56.21	-0.04	0.84
	Dec**	55.39	49.25	26.27	59.73	-0.04	0.81
	Jan	65.21	62.75	14.44	70.19	0.07	0.82

*Shown to be statistically unsignificant by χ^2-test and Fisher Exact Test.
**Shown to be statistically unsignificant only by χ^2-test.

When making monthly comparisons, statistics representing forecasts made for the months of October through January were generally better than the other months. This is particularly true when considering Version 2.1 statistics. It is interesting to note that the statistics representing Version 1.1 forecasts made for July had the highest HR, but the worst CSI, FAR and POD.

This monthly data was further recombined into four groups of three months. This was done each of the three possible ways. The first possible way is Jan-Mar, Apr-Jun, Jul-Sep, and Oct-Dec; the second possible way is Feb-Apr, May-Jul, Aug-Oct, and Nov-Jan; the third possible way is Mar-May, Jun-Aug, Sep-Nov, and Dec-Feb. This monthly data was also recombined into three groups of four months. This was done each of the four possible ways. All of the statistics for each of the groups of months were compared to each of the other groups of months. These statistics are provided in Appendix A for review. When analyzing groups of months using Version 1.1 statistics, there was not any grouping that revealed insight to seasonal performance. However, this is not true for Version 2.1 statistics. The statistics representing Version 2.1 monthly grouping forecasts of Feb-May, Jun-Sep, and Oct-Jan were consistently better than the other groupings. Table 8 shows the various accuracy measurements for the monthly grouping of Feb-May, Jun-Sep, and Oct-Jan for Versions 1.1 and 2.1.

A close look at these statistics indicates that the model compared better to RAOB data during the months of December, January, February, and March using Version 2.1 than to the other monthly groupings using either version. The significance of this may be that Version 2.1 may be the model to use during the winter months, when turbulence is most persistent. Of course, this assumes that RAOB analysis is close enough to actual atmospheric conditions to be used as "truth".

Table 8. Accuracy and Bias by Version and Monthly Grouping

		HR	CSI	FAR	POD	HSS	Bias
Ver 1.1	Feb-May	46.83	28.44	38.66	34.65	0.00	0.56
	Jun-Sep	67.00	17.54	68.75	28.56	0.08	0.91
	Oct-Jan	45.74	33.07	23.04	36.70	0.05	0.48
Ver 2.1	Feb-May	54.63	48.92	39.05	71.25	0.00	1.17
	Jun-Sep	53.40	25.56	70.38	65.11	0.10	2.20
	Oct-Jan	57.58	52.40	25.69	63.98	0.04	0.86

4.3 Verification of MWFM Forecasts Using PIREPs

Extreme care must be taken when interpreting statistics taken from comparisons using PIREPs. As stated in Chapter 2, many statistics are inappropriate to calculate. Consider, for example, a forecast for turbulence and a PIREP which reports no turbulence. Since the MWFM output is designed to collect the maximum momentum flux in a 1.5° × 1.5° box, it is certainly conceivable that turbulence existed within that box, while the pilot was flying somewhere else in the box, where turbulence did not exist. Since pilots avoid turbulence, they would probably choose to avoid the leeward side of a mountain, where the MWFM would calculate turbulence to exist. Therefore, the MWFM forecast may be correct, and the PIREP may be accurate, and the comparison must be discarded. Then we are limited to comparisons between a negative turbulence forecast and a negative PIREP, a positive turbulence forecast and a positive PIREP, and a negative turbulence forecast and a positive PIREP. Since CSI, FAR, HSS, and Bias all include calculations using incidents of a positive turbulence forecast and a negative PIREP, they cannot be meaningfully calculated. This leaves HR and POD as the only meaningful statistics. In other words, the only meaningful statistics available are those which tell the proportion of correct forecasts to total forecasts, and how often a positive turbulence reported was correctly forecasted.

4.3.1 Comparison of MWFM Versions The χ^2-test for homogeneity was performed on contingency tables which represented forecasts performed by MWFM Versions 1.1 and 2.1, which were compared to PIREPs. Since approximately half of the forecasts for each model forecasted turbulence, the test indicated that there was not a statistically significant difference between the two versions. However, a larger sample size may produce a more accurate result, since this conclusion disagrees with the conclusion made earlier, when ten months' worth of forecasts were compared.

When all comparison between MWFM forecasts and PIREPs are divided into two contingency tables based on model version, the χ^2-test showed that statistics from neither version were statistically significant, but that the Version 2.1 *p*-value was much lower than that of Version 1.1. At first glance, it seems clear that statistics representing Version 1.1 forecasts are quite different from statistics representing Version 2.1 forecasts (see Table 9). However calculations show a higher HR for Version 1.1, and a significantly higher POD for Version 2.1.

4.3.2 Atmospheric Layer Comparison When the two data sets were further divided by atmospheric layer, six contingency tables were produced. The χ^2-test showed that only two tables were statistically significant.

Once again, we see that there is not a direct correlation between HR and POD (see Table 10). It is interesting to note that, in general, Version 1.1 had higher HRs, while Version 2.1 had higher PODs.

Table 9. MWFM vs PIREP Total Comparisons

Version 1.1				Version 2.1			
		Observed				Observed	
		Yes	No			Yes	No
Forecasted	Yes	51	52	Forecasted	Yes	86	109
	No	133	148		No	98	91
HR		51.82		HR		46.09	
POD		27.72		POD		46.74	
p-value		0.704		*p*-value		0.129	

Table 10. Verification by Atmospheric Layer

		Y Fcst Y PIREP	Y Fcst N PIREP	N Fcst N PIREP	N Fcst Y PIREP	HR	POD
Ver 1.1	50-30*	23	40	41	24	50.00	48.94
	70-50	23	11	49	45	56.25	33.82
	100-70*	5	1	58	64	49.22	7.25
Ver 2.1	50-30	30	65	16	17	35.94	63.83
	70-50*	34	31	29	34	49.22	50.00
	100-70*	22	13	46	47	53.13	31.88

*p-value greater than 0.05

4.3.3 Forecast Hour Comparison When the two data sets were divided by forecast hour, ten contingency tables were produced. The χ^2-test showed that only one of the tables was statistically significant. Of the rest of the *p*-values, only the Version 1.1 00-hour forecast was relatively low.

Once again, with the correct negative forecasts neglected by the POD, there is no correlation between HR and POD. It is also interesting to notice that forecast accuracy does not decrease with time, as is usually expected with forecasts.

Table 11. Verification by Forecast Hour

		Y Fcst Y PIREP	Y Fcst N PIREP	N Fcst N PIREP	N Fcst Y PIREP	HR	POD
Ver 1.1	00 HR**	10	6	32	21	60.87	32.26
	06 HR	9	14	30	31	46.43	22.50
	12 HR	13	8	30	30	53.09	30.23
	18 HR	9	14	29	32	45.24	21.95
	24 HR	10	10	27	19	56.06	34.48
Ver 2.1	00 HR	16	19	19	15	50.72	51.61
	06 HR	19	25	19	21	45.24	47.50
	12 HR	21	22	16	22	45.68	48.84
	18 HR*	15	25	18	26	39.29	36.59
	24 HR	15	18	19	14	51.52	51.72

*p-value less than 0.05, ** p-value = 0.107, all others much higher

4.3.4 Other Comparisons Comparisons between locations were not possible to maintain classification of military operations. Seasonal comparisons were not reasonable, since less than a month's worth of PIREPs were collected.

When making the comparisons between MWFM forecasts and RAOB analyses, an attempt was made to make a correlation with momentum flux deposition forecasts by the MWFM and the HiCAT calculated by RAOB. This was done by assigning a '0', '1', '2', or '3' to each level on each sounding. If there was no HiCAT analyzed, a '0' was assigned. If HiCAT was analyzed, but less than a third of the maximum amount of the column was shaded, a '1' was assigned. Similarly, if between a third and two-thirds were shaded, a '2' was assigned, and if more than two-thirds was shaded, a '3' was assigned. After averaging the momentum flux calculated by the MWFM over all of the times a '0', '1', '2', or '3' was assigned, there was no correlation between the momentum flux calculated by the MWFM and the HiCAT calculated by RAOB. Therefore, there was not an increase of momentum flux calculated by the MWFM with an increase in HiCAT calculated by RAOB.

When comparing the results of this research with that of Allen (2003), many similarities are noted. Specifically, during both periods of research, Versions 1.1 and 2.1 produce significantly different forecasts; Version 2.1 forecasted turbulence more often than Version 1.1. Further, during both periods of research, model correlation to RAOB analysis increases with height and does not decrease with time.

V. Conclusions and Recommendations

5.1 Conclusions

The primary purpose of this research was to provide AFWA with results which would help determine the usefulness of the MWFM. This goal has been reached through the evaluation of MWFM forecasts using RAOB analyses and PIREPs.

5.1.1 Evaluation of MWFM Forecasts Using RAOB Analyses Soundings were collected twice daily for nearly a full year, beginning in early-April 2003 and continuing through mid-March 2004. These soundings were analyzed using Environmental Research Services' RAOB program. This product was chosen because it was used by Allen (2003) and is currently used by operational weather units in the geographical region from which these sounding were taken.

Since analyses by this program are not direct measurements of turbulence, the comparison of MWFM forecasts to these analyses is simply a comparison of two turbulence products. Therefore, care must be taken when analyzing the statistical results of this research. The accuracy scores calculated from the contingency tables do not objectively describe the MWFM's ability to forecast Stratoturb. However, the objective comparison between the two products does provide the ability for the Air Force to determine the value of the MWFM.

The RAOB program has been criticized for overanalyzing the presence and intensity of turbulence (Allen 2003). Therefore, it is reasonable to conclude that when the MWFM's bias score is less than one, it is reasonable that the MWFM forecasts provide a more accurate assessment of the presence of Stratoturb. It is interesting to note that both versions of the MWFM have bias scores less than one during the winter months, when it is most critical to pilots to have accurate turbulence forecasts.

The MWFM is a forecast tool which allows both numerical and graphical output. The graphical output may be displayed over any geographical location and for any atmospheric layer available in the model input data. Further, the MWFM has the ability to create forecasts, while the RAOB program simply analyzes data which is already hours old by the time the forecaster receives it (even though it may be used to make a forecast using a forecasted sounding).

5.1.2 Verification of MWFM Forecasts Using PIREPs Because of the nature of military operations in the geographic region of research, PIREPs were classified, and could not be transmitted without declassification. A significant delay occurred when setting up a suitable format for relaying PIREP data for this research. PIREPs were not acquired for comparison until early-December. This allowed comparison between MWFM forecasts and PIREPs to cover a period of only a few months. Since specific flight paths could not be included in these PIREPs, regions were used instead. These regions had to be quite large. This introduced a serious problem for verification. One area of a particular region may have had turbulence at a particular time, when there may have been no turbulence in another part of the same region at the same time. Since the MWFM output used the maximum momentum flux in the region to determine whether turbulence existed in that region, and pilots try to fly in areas of minimum turbulence whenever possible, it is easy to see that verification could be performed much more accurately if the region was exactly the same as the flight path. This was not possible due to the classified nature of the military operations in the geographical location being analyzed. Further, due to the fact that pilots try to avoid turbulence whenever possible, verification of the MWFM forecasts by PIREPs is limited primarily to HR and POD.

5.1.3 Statistical Conclusions Statistical analysis between the different MWFM versions shows that the two versions are producing different forecasts. While statistics compiled by this research cannot definitively determine which is the more accurate of the two versions, it is reasonable to conclude that Version 2.1 is more effective than Version 1.1 since it has the ability to forecast both vertically propagating and evanescent waves. A more objective study of the two versions is needed to determine which of the two models is more accurate.

Further comparison between the two versions shows that statistics representing Version 2.1 forecasts were generally better than those representing Version 1.1 forecasts. This is true regardless of how the data set was divided, whether by atmospheric layer, forecast hour, sounding location, or season. Since turbulence occurrence is so much higher during the winter months, it is much more of a concern to pilots to have a forecast tool that performs well during the winter. Based on the statistical analyses, both version of the MWFM performed slightly better during the winter months. Further, statistics representing Version 2.1 forecasts were, in general, better than statistics representing Version 1.1 forecasts during these months.

When considering the fact that Version 2.1 has the ability to forecast both types of wave propagation and the statistical analyses, MWFM Version 2.1 is the best choice for use by the Air Force as a Stratoturb forecasting tool.

5.1.4 Quantification of Turbulence Intensity An extensive attempt was made to determine if there is a relation between the turbulence intensity forecasted by the MWFM and turbulence intensity analyzed by the RAOB program. The width of the box along the left vertical axis of the sounding analysis (see Figure 5) was used to assign a turbulence intensity value of zero through three. If there was no turbulence analyzed in a layer, a zero was assigned to the layer. If turbulence was analyzed, and less than a third of the column was colored, then a one was assigned to the layer. Similarly, if between a third and two-thirds was colored, then a two was assigned to the layer, and if more than two-thirds was colored, then a three was assigned to the layer. Then all MWFM forecasts were divided into four categories, each corresponding to the turbulence intensity value as described above. For each category, the momentum flux forecasts made by the MWFM were averaged. This analysis included dividing up the forecasts by atmospheric layer, forecast hour, location, and season. There was no correlation found between the turbulence intensity forecasted by the MWFM and turbulence intensity analyzed by the RAOB program, regardless of how the data were divided.

5.2 Recommendations

Since the MWFM is already in operational use at AFWA, and more objective analysis needs to be done in order to determine which version more accurately represents the atmospheric conditions being analyzed, it is recommended that PIREPs be made available to AFWA via secure mode (e.g., SIPRNet) for further comparison. This would allow more accurate comparisons between MWFM output and PIREPs, and would alleviate the problem of having regions which are too large.

Another option is to make graphical MWFM output available to the operational units providing PIREPs. Making this available would naturally lead to a relatively simple feedback process, allowing pilots to report back on the accuracy of the product and the

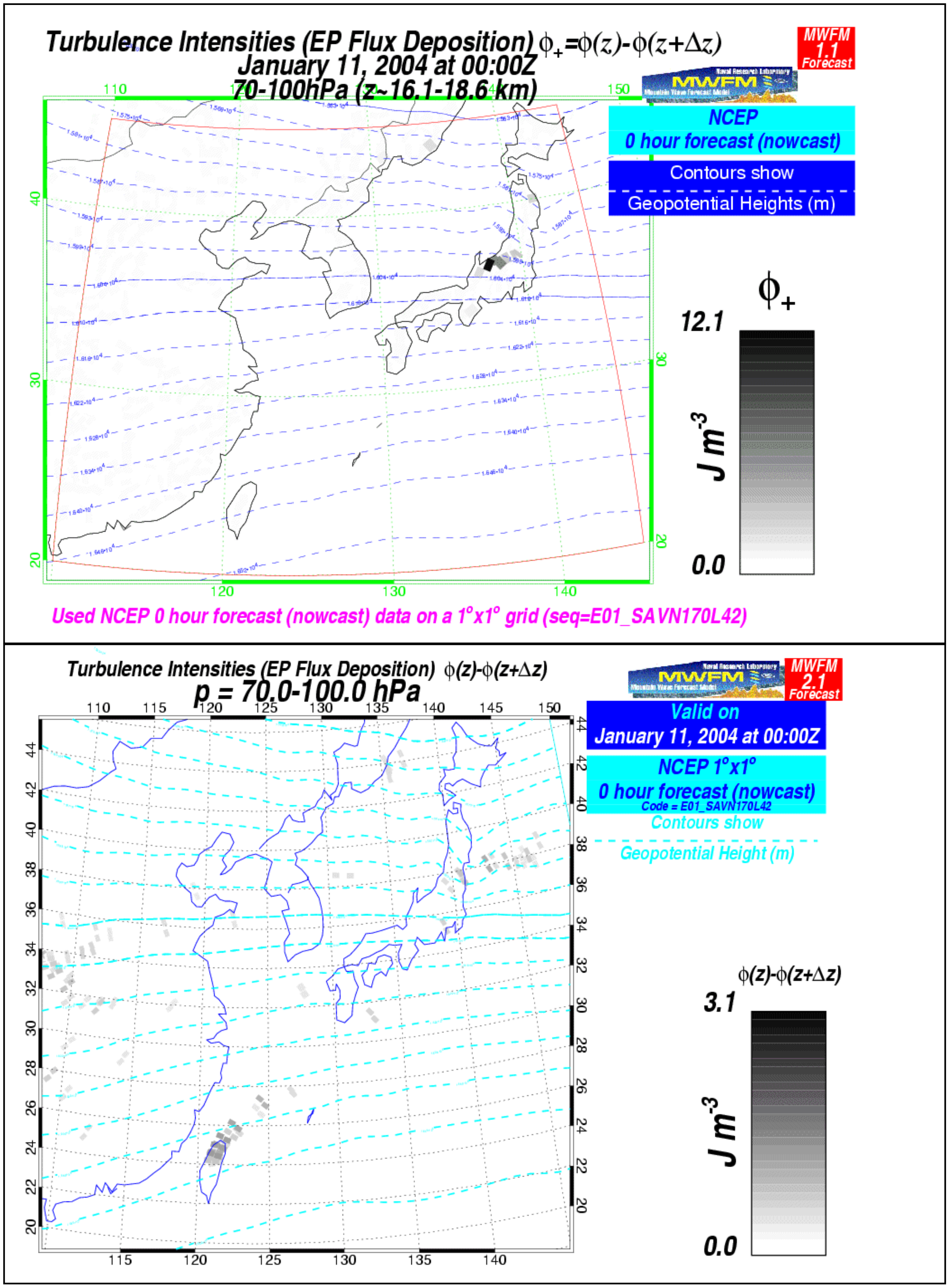

Figure 7. MWFM Graphical Output Comparison. Version 1.1 (top) with maximum momentum flux deposition of 12.1 J/m^3. Version 2.1 (bottom) with maximum momentum flux deposition of 3.1 J/m^3. Also evident is the fact that Version 1.1 does not allow for waves to propagate downstream like Version 2.1 does.

usefulness of having the product available during flight planning. There are two drawbacks to this option. First, the range of momentum flux deposition on the graphical output varies from zero to the maximum displayed, which is not the same on every product (see Figure 7). The fact that the ranges are different may lead to confusion, since the maximum displayed on one version's graphical output may represent light turbulence when compared to a model run which forecasts severe turbulence. Second, numerical analysis would be very difficult, since it could not be automated. Even with these drawbacks in mind, it would be a valuable evaluation tool to make these graphical products available to the operational units providing PIREPs.

Another recommendation is to research the relationship between the numerical output, which represents the momentum flux at a given location, and turbulence intensity. This would also require the submission of PIREPs from flying units to AFWA. Attempts were unsuccessful to make a correlation during this research. Even if a correlation were found, the correlation found would have either been between MWFM forecasts and RAOB analysis, which does not represent the actual atmospheric conditions, or between MWFM forecasts and PIREPs, which covered too large of an area to be accurate during this research. A study which uses MWFM output which is confined to the actual flight path would garner much more accurate, reliable and useful results.

Appendix A: Monthly Grouping Statistics

		HR	CSI	FAR	POD	HSS	Bias
Ver 1.1	Jan-Mar	47.08	40.59	13.15	43.25	0.05	0.50
	Apr-Jun	53.94	25.67	47.02	33.25	0.06	0.63
	Jul-Sep	67.44	17.62	68.91	28.90	0.09	0.93
	Oct-Dec	45.07	28.71	29.60	32.66	0.03	0.46
Ver 2.1	Jan-Mar	65.21	62.75	14.44	70.19	0.07	0.82
	Apr-Jun	54.40	42.62	48.28	70.77	0.10	1.37
	Jul-Sep	53.17	24.65	71.30	63.60	0.10	2.22
	Oct-Dec	53.79	46.88	32.05	60.19	0.01	0.89

		HR	CSI	FAR	POD	HSS	Bias
Ver 1.1	Feb-Apr	45.16	32.51	33.30	38.81	-0.02	0.58
	May-Jul	66.34	15.19	70.21	23.66	0.05	0.79
	Aug-Oct	55.96	26.04	47.13	33.91	0.09	0.64
	Nov-Jan	44.80	34.21	19.20	37.24	0.05	0.46
Ver 2.1	Feb-Apr	57.66	54.50	33.00	74.50	-0.04	1.11
	May-Jul	53.55	26.37	69.33	65.27	0.11	2.13
	Aug-Oct	52.92	37.84	51.15	62.67	0.07	1.28
	Nov-Jan	59.07	54.83	21.51	64.53	0.04	0.82

		HR	CSI	FAR	POD	HSS	Bias
Ver 1.1	Mar-May	46.83	28.44	38.66	34.65	0.00	0.56
	Jun-Aug	70.81	15.44	72.61	26.14	0.09	0.95
	Sep-Nov	48.81	26.53	40.23	32.30	0.03	0.54
	Dec-Feb	45.95	37.58	16.54	40.61	0.05	0.49
Ver 2.1	Mar-May	54.63	48.92	39.05	71.25	0.00	1.17
	Jun-Aug	56.77	23.85	72.88	66.43	0.13	2.45
	Sep-Nov	50.47	41.29	43.78	60.86	-0.03	1.08
	Dec-Feb	62.13	58.70	17.77	67.22	0.07	0.82

		HR	CSI	FAR	POD	HSS	Bias
Ver 1.1	Jan-Apr	46.22	37.14	21.96	41.48	0.02	0.53
	May-Aug	67.59	15.51	70.24	24.47	0.06	0.82
	Sep-Dec	47.91	27.24	37.68	32.62	0.03	0.52
Ver 2.1	Jan-Apr	61.79	59.03	23.30	71.92	0.01	0.94
	May-Aug	55.33	26.01	69.66	64.58	0.12	2.13
	Sep-Dec	51.30	42.68	40.96	60.63	-0.02	1.03

		HR	CSI	FAR	POD	HSS	Bias
Ver 1.1	Feb-May	46.83	28.44	38.66	34.65	0.00	0.56
	Jun-Sep	67.00	17.54	68.75	28.56	0.08	0.91
	Oct-Jan	45.74	33.07	23.04	36.70	0.05	0.48
Ver 2.1	Feb-May	54.63	48.92	39.05	71.25	0.00	1.17
	Jun-Sep	53.40	25.56	70.38	65.11	0.10	2.20
	Oct-Jan	57.58	52.40	25.69	63.98	0.04	0.86

		HR	CSI	FAR	POD	HSS	Bias
Ver 1.1	Mar-Jun	53.94	25.67	47.02	33.25	0.06	0.63
	Jul-Oct	60.53	24.15	52.14	32.77	0.11	0.68
	Nov-Feb	44.80	34.21	19.20	37.24	0.05	0.46
Ver 2.1	Mar-Jun	54.40	42.62	48.28	70.77	0.10	1.37
	Jul-Oct	53.68	34.30	57.08	63.07	0.10	1.47
	Nov-Feb	59.07	54.83	21.51	64.53	0.04	0.82

		HR	CSI	FAR	POD	HSS	Bias
Ver 1.1	Apr-Jul	59.61	23.64	52.62	32.06	0.10	0.68
	Aug-Nov	53.14	25.55	44.00	31.97	0.07	0.57
	Dec-Mar	45.95	37.58	16.54	40.61	0.05	0.49
Ver 2.1	Apr-Jul	54.86	37.82	55.02	70.39	0.14	1.56
	Aug-Nov	52.52	39.23	47.58	60.92	0.05	1.16
	Dec-Mar	62.13	58.70	17.77	67.22	0.07	0.82

Bibliography

Allen, Mark S., 2003: Evaluation of the Mountain Wave Forecast Model's Stratospheric Turbulence Simulations. AFIT, 70 pp.

Bacmeister J.T., P.A. Newman, B.L. Gary, K.R. Chan, 1994: An Algorithm for Forecasting Mountain Wave-Related Turbulence in the Stratosphere. *Wea. Forecasting*, **9**, 241-253.

Balsley, B.B., and D.A. Carter, 1989: Mountain Waves in the Tropical Pacific Atmosphere: A Comparison of Vertical Wind Fluctuations over Pohnpei and Christmas Island Using VHF Wind Profilers. *J. Atmos. Sci.*, **46**, 2698-2715.

Bender, M.A., H.A. Panofsky, and C.A. Pelsen, 1976: Temperature Gradients and Clear-Air Turbulence Probabilities. *J. Appl. Meteor.*, **15**, 1193-1199.

Brown, B.G., 1996: Verification of in-flight icing forecasts: Methods and issues. *FAA International Conference on Aircraft In-flight Icing*, Report No. DOT/FFA/AR-96/81, II, 319-330.

Brown, B.G., and G.S. Young, 2000: Verification of Icing and Turbulence Forecasts: Why Some Verification Statistics Can't Be Computed Using PIREPs, Conference on Aviation, Range, and Aerospace Meteorology, 9th, 11-15 September 2000. Orlando, FL.

Ehernberger, L.J., 1992: *Stratospheric Turbulence Measurements and Models for Aerospace Plane Design*. NASA Technical Memorandum 104262, 10 pp.

Gill, A.E., 1982: *Atmosphere-Ocean Dynamics*. Academic Press, 662 pp.

Hines, C.O., 1960: Internal Atmospheric Gravity Waves at Ionospheric Heights, *Canadian Journal of Physics*., **38**, 1441-1481.

Hines, C.O., 1963: The Upper Atmosphere in Motion, *Quarterly Journal of the Royal Meteorological Society*., **89**, 1-42.

Hines, C.O., 1971: Generalizations of the Richardson Criterion for the Onset of Atmospheric Turbulence. *Quart. J. Roy. Meteor. Soc.*, **97**, 429-439.

Hines, C.O., 1988a: Generation of Turbulence by Atmospheric Gravity Waves, *J. Atmos. Sci.*, **45**, 1269-1278.

Hines, C.O., 1988b: On Ray Paths in Mountain Waves, *J. Atmos. Sci.*, **45**, 323-326.

Holton, J.R., 1992: An Introduction to Dynamic Meteorology. Academic Press, 511 pp.

Kane, T.L., B.G. Brown, and R.T. Bruintjes, 1998: Characteristics of pilot reports of icing. *Preprints, 14th Conference on Probability and Statistics in the AStmospheric Sciences*, 11-16 January 1998, Phoenix, AZ.

Klemp, J.B., and D.K. Lilly, 1978: Numerical Simulation of Hydrostatic Mountain Waves. *J. Atmos. Sci.*, **35**, 78-83.

Lilly, D.K., and P.J. Kennedy, 1973: Observations of a Stationary Mountain Wave and its Associated Momentum Flux and Energy Dissipation, *J. Atmos. Sci.*, **30**, 1135-1152.

Lindzen, R.S., 1981: Turbulence and Stress Owing to Gravity Wave and Tidal Breakdown. *J. Geophys. Res.*, **86**, 9707-9714.

Marks, C.J., and S.D. Eckerman, 1995: A Three-Dimensional Nonhydrostatic Ray-Tracing Model for Gravity Waves: Formulation and Preliminary Results for the Middle Atmosphere. *J. Atmos. Sci.*, **52**, 1959-1985.

Sinclair, P.C., and P.M. Kuhn, 1991: *Infrared Detection of High Altitude Clear Air Turbulence*, NOARL Technical Note 205, pp. 49.

Smith, R.B., 1977: The Steepening of Hydrostatic Mountain Waves. *J. Atmos. Sci.*, **34**, 1634-1654.

Smith, R.B., 1980: Linear Theory of Stratified Hydrostatic Flow past an Isolated Mountain. *Tellus*, **32**, 348-364.

Waco, D.E., 1972: Temperature Gradients in Stratospheric Turbulence, *J. Appl. Meteor.*,**11**, 99-107.

Weinstock, J., 1987: The Turbulence Field Generated by a Linear Gravity Wave. *J. Atmos. Sci.*, **44**, 410-420.

Wilks, D.S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press, 467 pp.

Wurtele, M.G., A. Datta, and R.D. Sharman, 1993: *Lee Waves: Benign and Malignant*. NASA Contractor Report 186024, 23 pp.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 074-0188
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, 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 the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**			
1. REPORT DATE *(DD-MM-YYYY)* March 2004	**2. REPORT TYPE** **Master's Thesis**		**3. DATES COVERED** *(From – To)* Jun 2003 – Mar 2004
4. TITLE AND SUBTITLE **VERIFICATION OF THE MOUNTAIN WAVE FORECAST MODEL'S STRATOSPHERIC TURBULENCE FORECASTS USING SOUNDING DATA AND PILOT REPORTS**			**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Miller, Scott M., Captain, USAF			**5d. PROJECT NUMBER** **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 641 WPAFB OH 45433-7765			**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT/GM/ENP/04-10
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFWA/DNXT Bruce Telfeyan 106 Peacekeeper Drive Offutt AFB, NE 68113-4039			**10. SPONSOR/MONITOR'S ACRONYM(S)** **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
12. DISTRIBUTION/AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.			
13. SUPPLEMENTARY NOTES			

14. ABSTRACT

Since stratospheric turbulence (Stratoturb) is becoming an increased concern to the Air Force, the threat of damage to aircraft must be addressed. Therefore, the Air Force Weather Agency (AFWA) requests an accurate Stratoturb forecast model.

In 2002, The Mountain Wave Forecast Model (MWFM) was modified in order to develop a Stratoturb forecast tool. Turbulence forecasts generated for locations over East Asia were compared to output from the Rawindsonde Observation (RAOB) program. Although the results were promising, verification by aircraft crews flying through the stratosphere would improve the confidence of this forecast model, improving the forecaster's ability to warn pilots and alleviate the potential danger of Stratoturb.

This thesis continues that research. Three major changes were made. Pilot reports (PIREPs) were collected for verification of MWFM forecasts, the model's time resolution was increased for better comparison to PIREPs, and data were collected for nearly a year to determine season performance. Model performance was analyzed to determine if performance improved over a certain terrain type or at a certain altitude.

Results suggest that the MWFM is superior to previous methods of detecting Stratoturb. Therefore, the MWFM is recommended to AFWA for operational use.

15. SUBJECT TERMS

Mountain Wave Forecast Model, Turbulence, Stratosphere, Stratospheric Turbulence, PIREP, Pilot Report, Verification, RAOB, Rawinsonde Observation

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Michael K. Walters, Lt Col, USAF (ENP)
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	UU	63	(937) 255-7419, ext 4681; e-mail: michael.walters@afit.edu

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18