DOCTORAL THESIS

Multi-Model Ensemble Predictions of Atmospheric Turbulence

Author: Luke N. Storer

Supervisor: Prof. Paul D. Williams and Dr. Philip G. Gill

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy

in the

Department of Meteorology
School of Mathematical, Physical and Computational Sciences

April 6, 2019

Declaration of Authorship

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Luke N. Storer

Abstract

Atmospheric turbulence is a major aviation hazard, costing the aviation industry millions of dollars each year through aircraft damage and injuries to passengers and crew. In this thesis we compare reanalysis data to climate model simulations to understand how well climate models predict the location of Clear-Air Turbulence (CAT). We then model how climate change will impact CAT on a global scale, in all four seasons and at multiple flight levels. This provides ample motivation for the second half of the thesis which aims to improve aviation turbulence forecasting by testing a multi-model ensemble forecast by combing the Met Office Global and Regional Ensemble Prediction System (MOGREPS-G) and the European Centre for Medium Range Weather Forecasting (ECMWF) Ensemble. The main results found are that climate models are able to skilfully predict the location of CAT, with the main uncertainty of the location of CAT coming from which turbulence index is the best and not from the use of a climate model. We also found CAT will increase globally in the future with climate change, for multiple aviation-relevant turbulence strength categories, at multiple flight levels and in all seasons. For ensemble forecasting we started with a single-diagnostic equally weighted multi-model ensemble and found it is at least as skilful as the single-model ensembles. This lack of significant improvement in the forecast skill could be because when increasing the forecast spread, we capture more turbulence events but also more false alarms. The relative economic value of the forecast is improved for the multi-model ensemble, particularly at low cost/loss ratios. Through combining two ensembles we gain consistency, gain more operational resilience and create one authoritative forecast whilst maintaining skill and increasing value. Extending this work further, it is found that these results apply more generally for multiple turbulence diagnostics, as the multi-model ensemble was more skilful than either of the single-model ensembles. When combining the predictors, the multi-diagnostic multi-model ensemble was more skilful than the two single-model ensembles. It was also found that an optimised 12-member ECMWF and MOGREPS-G multi-diagnostic ensemble was more skilful than the 51-member multi-diagnostic ensemble. What this therefore indicates is that a smaller ensemble spread for the individual diagnostics within a multi-diagnostic ensemble is important for optimising operational forecasts in the future, which could reduce computational costs for turbulence forecasting.

Acknowledgements

I would like to thank the Natural Environment Research Council SCENARIO Doctoral Training Partnership which funded this PhD studentship and the Met Office for their support in providing forecast and observational data, as without it this project would not be possible. Also for their roles in producing, coordinating, and making available the CMIP5 model output, we acknowledge the Met Office Hadley Centre, the World Climate Research Programme's (WCRP) Working Group on Coupled Modelling (WGCM), and the Global Organization for Earth System Science Portals (GO-ESSP).

I would like to thank my supervisors Paul Williams and Philip Gill for their guidance, support and knowledge that has been invaluable throughout. I would also like to thank Jacob Cheung for his role in making available the ensemble data at the Met Office and his technical support in creating the probability forecasts.

Finally I would like to thank my family and friends for their support and my wife and parents who helped proof read my papers, as well as listening (even if pretending) to my project troubles. To all those who proof read my work, I can't thank you enough for volunteering your time.

Contents

List of Figures

List of Tables

List of Abbreviations

AMO	**A**tlantic **M**ulti-decadal **O**scillation
ATC	**A**ir **T**raffic **C**ontrol
AUC	**A**rea **U**nder (the) **C**urve
CAPE	**C**onvective **A**vailable **P**otential **E**nergy
CAT	**C**lear-**A**ir **T**urbulence
CFCs	**C**hloro**F**luoro**C**arbons
CIT	**C**onvectively **I**nduced **T**urbulence
CMIP3	Third **C**oupled **M**odel **I**ntercomparison **P**roject
CMIP5	Fifth **C**oupled **M**odel **I**ntercomparison **P**roject
CMIP6	Sixth **C**oupled **M**odel **I**ntercomparison **P**roject
CO_2	Carbon dioxide
DEVG	**D**erived **E**quivalent **V**ertical **G**ust
DJF	**D**ecember **J**anuary **F**ebruary
ECMWF	**E**uropean **C**entre (for) **M**edium **R**ange **W**eather **F**orecasting
EDR	**E**ddy **D**issipation **R**ate
Ellrod TI1	Ellrod & Knapp (1992) **T**urbulence **I**ndex **1**
ENDGame	**E**ven **N**ewer **D**ynamics (for) **G**eneral **a**tmospheric **m**odeling (of the) **e**nvironment
ENSO	**E**l **N**iño **S**outhern **O**scillation
EPS	**E**nesmble **P**rediction **S**ystem
FAA	**F**ederal **A**viation **A**dministration
GEFS	**G**lobal **E**nsemble **F**orecast **S**ystem
GTG	**G**raphical **T**urbulence **G**uidence system
hPa	**h**ecto**Pa**scals
IPCC	**I**ntergovernmental **P**anel (on) **C**limate **C**hange
JJA	**J**une **J**uly **A**ugust
LIDAR	**LI**ght **D**etection **A**nd **R**anging
MAM	**M**arch **A**pril **M**ay
MCS	**M**esoscale **C**onvective **S**ystem
MoG	**M**oderate-**o**r-**G**reater
MOGREPS-G	**M**et **O**ffice **G**lobal (and) **R**egional **E**nesmble **P**rediction **S**ystem
MWT	**M**ountain **W**ave **T**urbulence

NAO	**N**orth **A**tlantic **O**scillation
NCEP	**N**ational **C**enters (for) **E**nvironmental **P**rediction
NCSU1	**N**orth **C**arolina **S**tate **U**niversity index **1**
NOAA	**N**ational **O**ceanic (and) **A**tmospheric **A**dministration
NTSB	**N**ational **T**ransport **S**afety **B**oard
O_3	Atmospheric ozone
picontrol	**p**re-**i**ndustrial **control**
PIREPs	**PI**lot **REP**orts
RADAR	**RA**dio **D**etection **A**nd **R**anging
RCP8.5	**R**epresentative **C**oncentration **P**athway **8.5**
Ri	**Ri**chardson number
ROC	**R**elative **O**perating **C**haracteristic plot
SIGWX	**SIG**nificant **W**eather chart
SON	**S**eptember **O**ctober **N**ovember
TI2	Ellrod & Knapp (1992) **T**urbulence **I**ndex **2**
TIGGE	**T**he **I**nternational **G**rand **G**lobal **E**nsemble project
USA	**U**nited **S**tates (of) **A**merica
V	Relative economic **V**alue
WAFC	**W**orld **A**rea **F**orecast **C**entre
WMO	**W**orld **M**eteorological **O**rganisation

List of Symbols

A	Aircraft specific parameter	
c	Phase speed	s^{-1}
d_s	Near surface diagnostic	$\mathrm{m\,s}^{-1}$
f	Coriolis parameter	s^{-1}
g	Gravitational acceleration (9.81)	$\mathrm{m\,s}^{-2}$
H	Altitude ($\times 10^3$)	feet
k	Wave number	
m	Total mass of aircraft	metric tonnes
$\overline{m}$	Reference mass of aircraft	metric tonnes
n	Peak deviation of aircraft acceleration	g
N	Brunt–Väisälä frequency	s^{-1}
p	Pressure	Pa
R_E	Radius of Earth (6371000)	m
R_s	Specific gas constant for dry air (287.058)	$\mathrm{J\,kg}^{-1}\,\mathrm{K}^{-1}$
T	Temeprature	K
u	Horizontal wind velocity in the East-West direction	$\mathrm{m\,s}^{-1}$
U	Horizontal wind velocity	$\mathrm{m\,s}^{-1}$
$U(z)$	Vertical wind shear	s^{-1}
v	Horizontal wind velocity in the North-South direction	$\mathrm{m\,s}^{-1}$
V	Aircraft calibrated airspeed	knots
V_a	Aircraft true airspeed	$\mathrm{m\,s}^{-1}$
w	Vertical wind velocity	$\mathrm{m\,s}^{-1}$
x	Distance in the East-West direction	m
y	Distance in the North-South direction	m
z	Altitude	m
θ	Potential temperature	K
ϕ	Latitude	$^\circ$
Ω	Rotation rate of Earth	$\mathrm{rad\,s}^{-1}$
β	Coriolis frequency gradient	$\mathrm{m}^{-1}\,\mathrm{s}^{-1}$
σ_w	Running mean vertical wind shear	s^{-1}
ξ	Relative vorticity	s^{-1}

ω	Wave frequency	s^{-1}
ω_1	Cut-off frequency (0.1)	Hz
ω_2	Cut-off frequency (2)	Hz
Ri	Richardson number	
$DEVG$	Derived Equivalent Vertical Gust	$m\,s^{-1}$
EDR	Eddy Dissipation Rate	$m^{2/3}\,s^{-1}$
$TEMPG$	Horizontal temperature gradient	$K\,m^{-1}$

Chapter 1

Introduction

1.1 Turbulence and its Impact on Aviation

Atmospheric turbulence is a major aviation hazard, causing damage to aircrafts and injury to passengers and crew. Turbulence is part of the chaotic atmosphere, and the chaotic nature poses a great challenge in understanding and forecasting turbulence. Aviation turbulence is incompletely understood and difficult to forecast, making it a significant hazard. The USA National Transportation Safety Board (NTSB) records the average number of air-carrier related injuries as 58 per year (FAA, 2017b). However Sharman et al. (2006) suggest this number is an underestimate, as not all injuries are reported. They state that in the period 1980–2008 there were 234 turbulence related accidents, resulting in 298 serious injuries and three fatalities on United States operated air carriers. 184 of the serious injuries involved flight attendants and 114 involved passengers (FAA, 2017b). These turbulence injuries will come at a cost to the airlines in two ways: (i) through compensation being paid, which amounts to over $10 million per year, and (ii) through lost working days by injured cabin crew, which is equivalent to over 7,000 days per year (Sharman & Lane, 2016). Although most of the flight is spent at the cruise phase (around 33,000–39,000 feet), this is also the most vulnerable part of the flight as passengers and crew are unbuckled, making any encounter of turbulence more likely to result in injury. As a result, most incidents occur above 10,000 feet (Sharman et al., 2006).

Kauffmann (2002) analysed the impact turbulence has on the aviation industry in detail. He presents figures from the USA Federal Aviation Administration (FAA), who state there were 342 reports of turbulence affecting flights on major air carriers over the period 1981–1997, with three fatalities, 80 serious injuries and 769 minor injuries. These figures are similar to those of Sharman et al. (2006) and, again, may underestimate the real statistics of turbulence injuries. These values could also be higher in the future, as climate change

is likely to increase the frequency of Clear-Air Turbulence (CAT) around the world, particularly in the mid-latitudes (Williams & Joshi, 2013; Williams, 2017). The cost of turbulence to the aviation industry is significant and comes from many sources, one of which is preventing aircraft from flying on the optimum route. This cost is as much as $16 million a year, as it is estimated that 5% of flights are forced to fly non-optimal routes (Search Technology, 2000). More importantly, it is also suggested that 15% of the diversions could have been avoided with improved turbulence detection. One form of turbulence detection could be using LIght Detection And Ranging (LIDAR) technology on the front of an aircraft as an in-flight avoidance mechanism. The current RAdio Detection And Ranging (RADAR) technology used by aircraft is unable to detect turbulence unless hydrometeors are present, whereas for CAT they are absent. However, LIDAR technology is able to sense CAT using non-hydrometeor particles. This technology could warn pilots of turbulence ahead, enabling them to divert or put on the seatbelt sign so passengers and crew are aware of the approaching danger. However currently, Kauffmann (2002) calculates that LIDAR technology costs more to install than the saving would be, and therefore rolling it out is not worth the investment. Satellites can be used to avoid some types of turbulence, particularly turbulence associated with convection (Mecikalski et al., 2007; Francis & Batstone, 2013). There is also the potential for the new generation of satellites (e.g. GOES-16) with higher spatial and temporal resolutions to improve the avoidance of turbulence. This improvement could come from better identification of deep convection and the ability to resolve gravity waves that would otherwise be invisible to the older generation GOES satellites (Wimmers et al., 2018; Nunez, 2018).

Forecasting turbulence is another possibility to mitigate injuries and damage to the aircraft by allowing pilots and flight planners to avoid regions containing turbulence. Turbulent eddies in the atmosphere occur on scales ranging from the planetary scale down to millimeters, but only eddies of approximately 100 m in size impact aviation (Sharman et al., 2006). This is a scale that is not explicitly simulated in forecasts except for a few detailed case studies (e.g. Lane et al., 2012), because the numerical models have resolutions too coarse to resolve the individual eddies. Therefore, turbulence diagnostics are used. These diagnostics generally work on the principle that the energy associated with turbulence on aviation-affecting scales cascades down from the larger scales that can be explicitly resolved by numerical models.

The different kinds of turbulence that impact aviation are outlined in Figure 2.1. Vertical wind shear instabilities, mountain waves and convection are

the three main sources that we consider in this thesis. Turbulence in and near clouds can also cause injuries to passengers and crew, but can easily be detected visually by pilots using on-board RADAR. Boundary-layer turbulence is not considered in this thesis, because it influences only a small portion of the flight after take-off and before landing.

1.2 Aims of this Thesis

The main aim of this thesis is to improve aviation turbulence forecasting by testing a method currently used in other areas of meteorology. This method is multi-model ensemble forecasting and is currently used in hurricane forecasting. We know that aviation turbulence has a large impact on the aviation industry, as discussed in Section 1.1, however climate model simulations are predicting an increase in turbulence frequency over the North Atlantic in the future due to strengthening vertical wind shear in the jet stream (Williams & Joshi, 2013; Williams, 2017). This expected increase provides ample motivation to develop improved turbulence forecasting. However, there is currently no literature assessing how well climate models predict CAT globally and therefore no way of knowing how well climate models predict the location of CAT. To address this, we compare reanalysis data to climate model simulations to understand how well climate models predict the location of CAT and its response to changes in greenhouse gas emissions. Then using this understanding we can extend the work conducted by Williams & Joshi (2013); Williams (2017) to model how climate change will impact CAT on a global scale, in all four seasons and at multiple flight levels. Using these results as further motivation, in the second half of this thesis we combine the Met Office Global and Regional Ensemble Prediction System (MOGREPS) ensemble, with the European Centre for Medium Range Weather Forecasting (ECMWF) Ensemble Prediction System (EPS) to see if a multi-model ensemble can improve the forecast skill, value and reliability of the forecasts.

1.3 Thesis Structure

The thesis will be structured as follows: Chapter 2 will present the current literature on aviation turbulence including the dynamics of turbulence, climate change impacts and how it is currently forecast for aviation. Then Chapter 3 will study if climate models can successfully diagnose clear-air turbulence and its response to climate change. Chapter 4 will show how climate change will

impact CAT globally in all four seasons and at multiple flight levels. Chapter 5 will show if the multi-model ensemble method can provide more skill, value and reliability of the forecast than a single model ensemble approach for a single turbulence diagnostic. Chapter 6 will introduce multiple turbulence diagnostics forecasting shear turbulence, Mountain Wave Turbulence (MWT) and Convectively Induced Turbulence (CIT), to show the full capability of a multi-diagnostic multi-model ensemble forecast of aviation turbulence. Finally Chapter 7 will present the conclusions of the thesis and discuss areas where more research is needed.

Chapter 2

Literature Review

This Chapter is based on the paper 'Aviation Turbulence: Dynamics, Forecasting, and Response to Climate Change' published by Storer et al. (2018) in *Pure and Applied Geophysics*.

2.1 Turbulence sources

2.1.1 Clear-Air Turbulence

Clear-Air Turbulence (CAT) is defined as high-altitude aircraft bumpiness in regions devoid of significant cloudiness and away from thunderstorm activity (Chambers, 1955). Far from mountains, CAT is generally accepted to result from shear instabilities. Wind shear is therefore a major source of CAT and is one of the best understood sources. Figure 2.1 indicates this type of turbulence and its association with the jet stream. To understand why shear causes turbulence, we must define the Richardson number (Ri):

$$\mathrm{Ri} = \frac{N^2}{(\partial U/\partial z)^2} = \frac{(g/\theta)(\partial\theta/\partial z)}{(\partial U/\partial z)^2}, \tag{2.1}$$

where N^2 is the Brunt–Väisälä frequency squared, U is horizontal wind speed, z is altitude, g is gravitational acceleration and θ is potential temperature. The Richardson number is a nondimensional number with the numerator representing the stratification and the denominator representing the vertical wind shear. It follows from theoretical considerations that instability occurs when Ri$<$0.25, so instability is favoured by large vertical wind shear (denominator) and weak stratification (numerator). In computational calculations using gridded data, numerical models rarely reach Ri=0.25 due to the coarse resolutions and therefore thresholds of turbulence are model specific. To overcome this, Williams (2017) chose thresholds based on the distribution of turbulence in the atmosphere. For example, he assumes severe turbulence is found in 0.01% of

the atmosphere, and therefore he takes the top 0.01% (99.9-100%) of the probability distribution to be severe turbulence. Therefore each threshold is specific to each model and resolution.

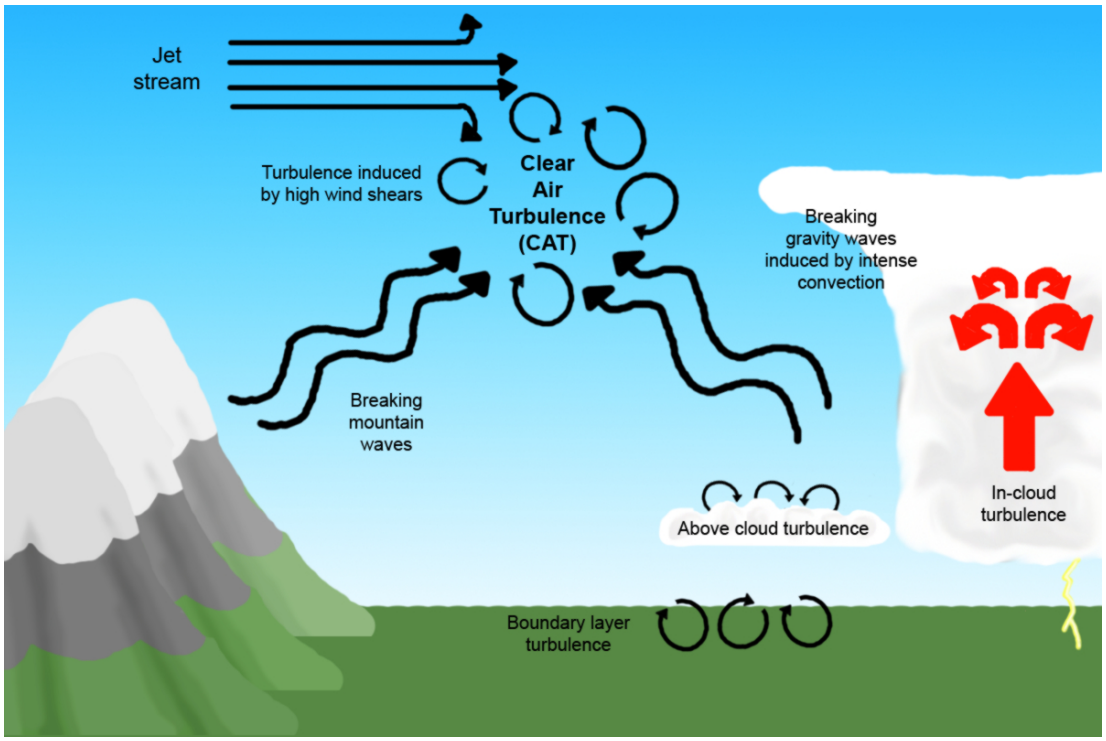

FIGURE 2.1: Plot of the main sources of turbulence that impact aviation: shear turbulence caused by high wind shear, breaking gravity waves induced by intense convection and breaking mountain waves. From Marlton (2016).

It is possible for turbulence to be produced when the environmental Richardson number is much larger than the theoretical critical value, if a local effect reduces the Richardson number locally. For example, gravity waves can cause CAT by reducing the Richardson number locally in an environment that would not normally produce turbulence, initiating the Kelvin–Helmholtz shear instability and leading ultimately to overturning and breaking billows. The various sources of gravity waves are discussed by Williams et al. (2003, 2005, 2008). In particular, gravity waves can be produced by convection (as discussed in Section 2.1.2) and spontaneous loss of geostrophic balance as the flow evolves, as described by the Lighthill–Ford theory (Lighthill, 1952; Ford, 1994; Knox et al., 2008; McCann et al., 2012). There is a direct connection in certain circumstances that links deformation to the generation of inertia–gravity waves (Knox et al., 2008). In some cases, this could explain the success of deformation-based CAT diagnostics such as Ellrod & Knapp (1992) TI1 and TI2 (Knox, 1997).

It is also possible for the environmental Richardson number to be reduced on a much larger scale, in regions of the atmosphere with particularly strong wind shear such as the jet streams. Strong vertical wind shears around the jet stream increase the denominator in Equation 2.1, which therefore decreases the Richardson number until it reaches a critical value and turbulence is produced. Therefore, understanding the behaviour of the jet stream will help researchers and forecasters understand how CAT may change. This is one of the reasons for the success in forecasting turbulence of CAT diagnostics containing vertical wind shear (e.g. Colson–Panofsky index (Colson & Panofsky, 1965), Ellrod TI1, TI2 (Ellrod & Knapp, 1992)).

Kim et al. (2016) studied the impact that the North Atlantic Oscillation (NAO) has on aviation turbulence. The NAO is a measure of the relative strength of the Icelandic low and the Azores high. The positive NAO phase implies a stronger than normal Icelandic low and Azores high, so there is a strong pressure difference between the two. In the negative NAO phase, the opposite is true and the pressure difference is weaker. In the positive NAO phase, we see a stronger jet stream and a more northerly path, whereas the negative NAO phase is associated with a weaker jet stream and a more southerly path. This change to the jet stream therefore has an impact on turbulence for trans-Atlantic flights. Kim et al. (2016) used wind-optimal routes to find the fastest possible flight path between London (LHR) and New York (JFK). The study found that eastbound flights fly more frequently through regions of CAT than westbound flights, and therefore experience more turbulence in both the positive NAO phase and the negative NAO phase. The reason for this is that eastbound flights utilise the jet stream more to benefit from strong tailwinds, so the wind-optimal routes would fly in stronger vertical shear regions more for eastbound flights than westbound flights. Westbound wind-optimal routes avoid the strong headwinds of the jet stream, and would also avoid the stronger vertical shear associated with the jet stream and therefore encounter less turbulence.

According to Kim et al. (2016), in the positive NAO phase, westbound flights experience more moderate-or-greater (MoG) CAT than in the negative NAO phase, because some of the westbound flights detour northward to be on the cyclonic shear side of the northerly shifted jet stream, which is more susceptible to MoG turbulence. In contrast, eastbound flights in the negative NOA phase fly through the cyclonic shear side of the southerly shifted jet. Therefore, the study suggests that westbound flights are more prone to MoG CAT in the positive NAO phase, whereas eastbound flights are more prone to

MoG CAT in the negative NAO phase. This information is important for flight planners, as for example in a positive NAO phase, an eastbound flight would fly on the anticyclonic shear side of the jet stream, still using the tailwind to reduce flight time. However, a westbound flight that would normally fly north to avoid the strong headwinds, could fly south and avoid the cyclonic shear side of the jet stream, reducing flight time but also the chances of strong turbulence. Information such as this can be used to avoid the strongest turbulence, while still attempting to reduce flight times and therefore fuel consumption.

Another example of jet stream behaviour was studied by Trier et al. (2012), who discuss how moist convection influences the upper-level jet stream. This topic has been studied by Trier & Sharman (2009) for warm Mesoscale Convective Systems (MCSs). Latent heat release perturbs an anticyclone, and this mechanism accounts for nearly all of the magnitude of the upper-level jet stream. This is a similar mechanism found in a cold weather outbreak, so a midlatitude cyclone (like the MCS) enhances the downstream anticyclone. Trier et al. (2012) suggests that, although not as dominant as MCSs, midlatitude cyclones account for 30–50% of the strength of the southerly jet stream. Without the moist convection and the perturbation of the anticyclone, the wind shear and therefore resultant CAT would not be as strong. This is different to Convectively Induced Turbulence (CIT) because it is the indirect effect of convection on the jet stream and the increased shear that causes turbulence. CIT however is associated with strong convection in the mid-latitude cyclone with cloud tops below the flight level and they generate gravity waves. This mechanism is discussed further in Section 2.1.2.

2.1.2 Convective Turbulence

Understanding the relationship between buoyancy and shear is very important in understanding where and why turbulence forms. Lane et al. (2012) explore our current understanding of near-cloud turbulence or Convectively Induced Turbulence (CIT). They explain that the FAA regulations at the time were not sufficient for avoiding severe turbulence. For example, guideline 5 states: “Do avoid by at least 20 miles (laterally) any thunderstorm identified as severe or giving an intense radar echo, especially under the anvil of a large cumulonimbus”. Guideline 6 states: “Do clear the top of a known or suspected severe thunderstorm by at least 1000 ft altitude for each 10 knots of wind speed at the cloud top”.

A possible source of CIT is unstable upper tropospheric thunderstorm outflow, similar to that described in Trier & Sharman (2009). Those authors proposed a mechanism for turbulence formation after completing simulations using a convection-permitting model. Their proposed mechanism is the formation of strong vertical wind shear in the outflow regions of Mesoscale Convective Systems (MCS), which we know reduces Ri and leads to Kelvin–Helmholtz instability, which is a well known source of turbulence. In the simulations they also found that strong vertical wind shear created regions of strong static instability. This was caused by differentially advecting equivalent potential temperature gradients, which were influenced by the adiabatic cooling in the convective updrafts.

The static instabilities described above are different from the traditional CIT mechanisms that generally result from high-frequency gravity wave breaking (Lane et al., 2003; Lane & Sharman, 2008) or from reductions in Ri as a result of propagating gravity waves (Sharman et al., 2012). Propagating gravity waves induced by convection are indicated as a source of turbulence in Figure 2.1. Lane et al. (2003) studied the traditional mechanisms forming turbulence above deep convection, and found that gravity waves formed when the overshooting top returns back down. As these waves propagate up, the changes in vertical wind shear and buoyancy with height can change the wavelength of gravity waves as well as the tilt. The phase speed remains the same above the jet stream as within the jet stream based on linear theory, and the decreasing speed above the jet stream could reach the critical level (Lane et al., 2012). This is found when $U(z) - c = 0$, where $U(z)$ is vertical wind shear and $c = \omega/k$, with ω being the wave frequency and k being the horizontal wave number. So the gravity wave can break if wave frequency is low, or wind shear is high.

Since wind shear is an important part of the gravity wave breaking, the jet stream plays an important role. The smaller the jet width, the larger the horizontal wind shear, and therefore the closer to the cloud top the gravity wave breaking would be. Lane et al. (2012) also showed turbulence can be found in cirrus bands (banding caused by thermal instabilities), and also ducted gravity waves that propagate far from the updraft, reducing the Richardson number and causing turbulence much farther away than the 20 miles outlined in the guidelines. The new guidelines from FAA (2017a) no longer reference the acceptable height above the thunderstorm pilots can fly. Guideline 2 now states: "Don't attempt to fly under a thunderstorm even if you can see through to the other side. Turbulence and wind shear under the storm could be hazardous". Guideline 3 states "Don't attempt to fly under the anvil of a thunderstorm.

There is a potential for severe and extreme clear-air turbulence". Guideline 14 states "Do avoid by at least 20 miles any thunderstorm identified as severe or giving an intense radar echo. This is especially true under the anvil of a large cumulonimbus".

2.1.3 Mountain Wave Turbulence

Mountain Wave Turbulence (MWT) is similar to Convectively Induced Turbulence, as gravity waves produced by the terrain (instead of convection) propagate and break when a critical vertical wind shear value is reached (Nastrom & Fritts, 1992; Wurtele et al., 1996). This is also indicated in Figure 2.1 and shows its similarity to CIT. Clark & Peltier (1977) conducted a numerical simulation of a stably stratified fluid over an isolated obstacle. They found that there is a critical aspect ratio of the topography, above which an internal gravity wave is formed. They also discussed the amplification of the nonlinear mountain wave between the critical level and the ground by 'overreflection'. This results in an increase in wave drag and results in a strengthening of the downslope wind. Epifanio & Qian (2008) however, used a high-resolution ensemble to study turbulence produced by breaking mountain waves and found their results were consistent with previous work by Fritts et al. (1994, 1996), suggesting shear is an important factor in the breaking of gravity waves. Fritts et al. (1996) ran simulations to understand the turbulence mechanism in mountain waves, and found that the dominant source of instability, and therefore turbulence, was the wind shear. This came from the mean wind field, and the differential vertical advection of the mean shear by the wave field. Similar to convective turbulence, the vertical wind shear and its interaction with propagating gravity waves is an important mechanism in turbulence production. This understanding will help forecasters in the future.

Shutts & Gadian (1999) also produced numerical simulations to understand how Mountain waves respond to wind backing with height. In the study they describe a process where the flow over mountains can feed a background spectrum of quasi-inertia gravity waves, which can be thought of as 'unrealised mountain wave drag'. When the vertical wavelengths of the inertia-gravity waves are small enough, they will eventually break and form turbulence Shutts (1998).

Wolff & Sharman (2008) show that MoG MWT is typically found where topographic heights are above 1.5 km. They also analysed the low-level wind direction when mountain-wave turbulence is produced, and concluded that

the wind direction must be within 30° of the perpendicular to the mountain range orientation. For North America this makes the Rocky Mountains a prime location for MoG MWT, as the mountain range is oriented North–South and frequently experiences westerly low-level flow with mountain heights above 1.5 km. Greenland is also an area with high mountain ranges with the potential to cause turbulence for trans-Atlantic flights.

2.2 Climatology and Response to Climate Change

Jaeger & Sprenger (2007) used reanalysis data to understand upper tropospheric clear-air turbulence in the Northern Hemisphere, because of its role in stratosphere–troposphere exchange (Shapiro, 1976, 1978, 1980). The authors create a climatology of the Ellrod & Knapp (1992) Turbulence Index 1 (Ellrod TI1), Brunt–Väisälä frequency squared (N^2), Richardson number (Ri), and potential vorticity (PV). Hoskins et al. (1985) state that PV is able to deduce diagnostically all other dynamical fields. It is used as a good indicator for the dynamical processes behind synoptic and air-mass concepts such as cut-off cyclones and blocking anticyclones as well as processes such as Rossby wave propagation, baroclinic instability and many more. Negative potential vorticity is however a turbulence predictor as it is also related to isentropic inertial instability. Therefore it is Jaeger & Sprenger (2007) found that Ellrod TI1 is largest north of the jet stream, and Kelvin–Helmholtz instability indicated by Ri is near the jet streams, which is what we would expect with larger wind shear in that region. Symmetric instability is most frequent south of the jets, and is particularly associated with anticyclonic jets. Hydrostatic instability is only slightly dependent on the jet position, and is most common over land where convection and gravity wave activity are most prevalent (mountain wave and deep convection gravity waves). The study also showed that winter (December, January and February) has the highest turbulence frequency, which follows the understanding that the jet stream is stronger in winter. The study also found long-term trends in the frequency of diagnosed turbulence over the reanalysis period. Figure 2.2 shows that over the reanalysis period there is an increasing trend of turbulence in the Northern Hemisphere. They were also able to relate inter-annual variability to the North Atlantic Oscillation (NAO) and Pacific/North American flow, which we know from Section 2.1.1 influences CAT.

We know the climate system is changing due to anthropogenic (human) forcing, and these changes may have an impact on turbulence in the future.

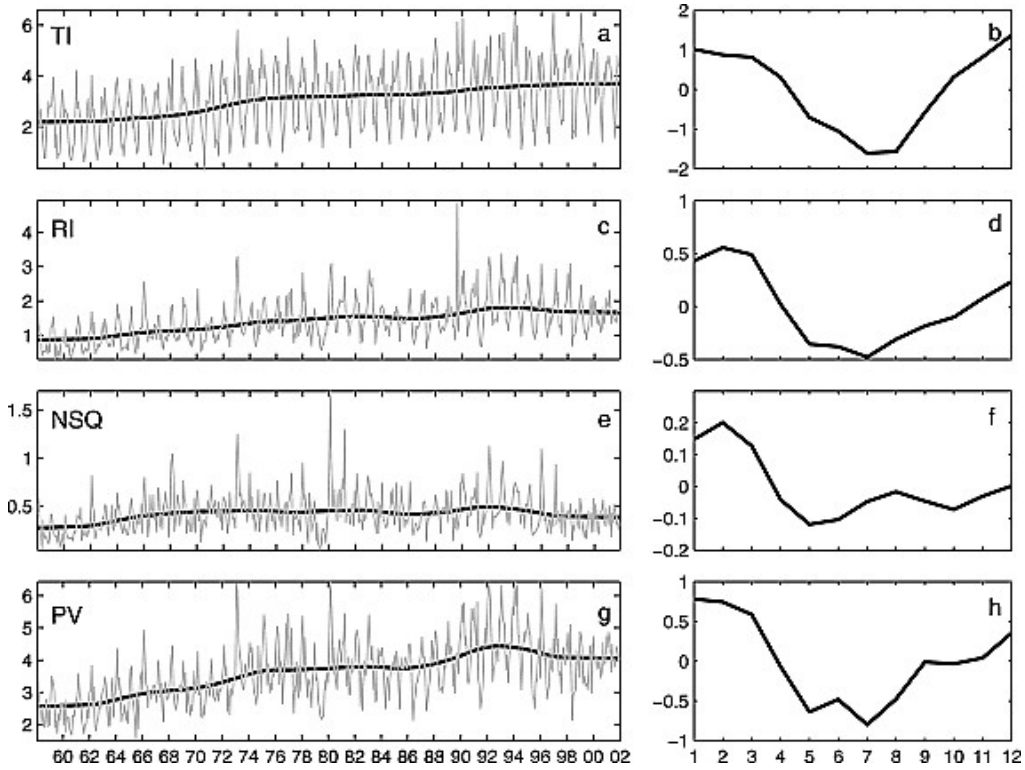

FIGURE 2.2: (left) Ellrod TI, Richardson number (Ri), Brunt–Väisälä frequency squared (N^2) and Potential Vorticity (PV) (from top to bottom) frequency time series (grey lines) and nonlinear trend estimates from seasonal trend decomposition procedure based on loess (STL) analysis (bold black line) for the north Atlantic sector from 90°W–10°E and 30–70°N in the tropopause region (%). (right) Mean seasonal cycle component of the turbulence indicators from STL decomposition (Δ%). All panels are for the time period 1958 to 2001. Note the different scales. From Jaeger & Sprenger (2007)

Collins et al. (2013) showed that the changes in temperature are not uniform around the world. Importantly for turbulence, the upper troposphere and lower stratosphere respond differently to anthropogenic forcing. The tropical upper troposphere is predicted to warm faster than the tropical surface, due to an increase in latent heat release. Latent heat is released during convection, and in a warmer climate the atmosphere can hold more moisture. As a result more convection and subsequent latent heat release will warm the troposphere. The lower stratosphere, however, will cool with the increasing greenhouse gases (Fels et al., 1980). This cooling is related to changes in infrared radiation.

The upper tropospheric and lower stratospheric changes discussed above

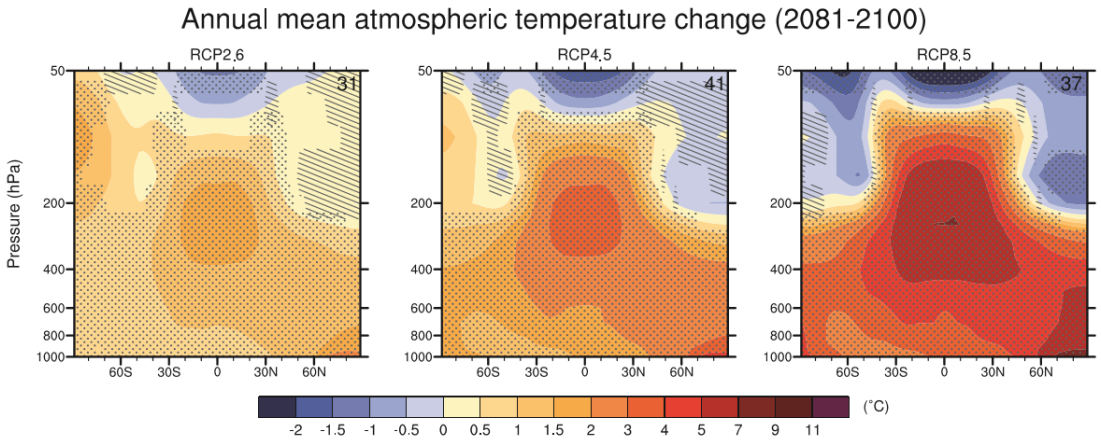

FIGURE 2.3: Modelled zonal-mean annual-mean ensemble-mean future temperature changes for climate change scenario RCP2.5 (left), RCP4.5 (middle), and RCP8.5 (right). Hatching indicates the regions where the ensemble-mean change is less than one standard deviation of internal variability. Stippling indicates regions where the ensemble-mean change is greater than two standard deviations of internal variability and where at least 90% of the models agree on the sign of the change. From Collins et al. (2013).

lead to an increase in the equator-to-pole temperature gradient at flight cruising levels, as shown in Figure 2.3. Particularly in the RCP8.5 simulation (right), where greenhouse gas emissions are highest, we see the warming in the tropics and the cooling at the poles, increasing the equator-to-pole temperature gradient. This increase in meridional temperature gradient will induce a thermal wind response, resulting in an increase in vertical wind shear and therefore turbulence in the mid-latitudes. These changes to the jet stream are shown in Figure 2.4, which is taken from Delcambre et al. (2013). The changes to the jet stream are predominant in the midlatitudes and at airline cruise altitudes, making their impact the largest in the busiest flight regions around the world. In addition to modifying turbulence, the increased wind speeds are also expected to modify flight times (Williams, 2016). Another impact of anthropogenic forcing is the twentieth-century release of chlorofluorocarbons (CFCs) which destroy atmospheric ozone (O_3). The loss of ozone reduces the lower stratospheric temperature, further increasing the equator-to-pole temperature

gradient. This effect enhances the predicted change arising from carbon dioxide.

Some studies have reported that climate change will act to increase CAT in the future, according to climate model simulations (Williams & Joshi, 2013; Williams, 2017). The first study to look at this (Williams & Joshi, 2013) focused on North Atlantic moderate-or-greater turbulence and showed that it would increase in frequency with climate change by an estimated 40–170%. Williams (2017) then furthered the study to see how climate change might influence turbulence in five strength categories from light to severe, finding that all would increase in frequency with climate change. Aircraft manufacturers and forecasters need to be prepared to prevent an increase in damage to aircraft and injuries to passengers and crew.

Currently there is a gap in the literature looking at the response of Convectively Induced Turbulence and Mountain Wave Turbulence to climate change. However, we know that deep convection can produce turbulence, and therefore it is plausible that CIT would increase if convection were increased. Price & Rind (1994) showed an increase of 5–6% of global lightning activity with every 1°C of global warming. Lightning activity has been used to characterise turbulence events (Gill & Stirling, 2013; Meneguz et al., 2016), and therefore if we see an increase in lightning, we would expect to see an increase in convection and possibly an increase in CIT as a result. Price & Rind (1994) show that lightning activity and convection will increase particularly around the tropics. This trend of increasing lightning with climate change is supported by Reeve & Toumi (1999) who show an increase in lightning activity of 40% for every 1 K of average land wet-bulb temperature. It could be possible to study the change in CIT to climate change. Convective precipitation accumulation was used by Gill & Stirling (2013) as a convective indicator for turbulence forecasts, therefore assessing the changes of this quantity in climate models could indicate how climate change might impact CIT.

2.3 Forecasting

Currently the World Area Forecast Centres (WAFCs) in London (Met Office) and Washington (NOAA) produce operational turbulence forecasts for aviation. This includes a forecaster produced T+24 hours significant weather (SIGWX) chart four times a day (0000, 0600, 1200, 1800 UTC) which displays multiple aviation hazards including icing, CAT as well as the location of convection. The WAFCs also produce a gridded turbulence forecast four times a

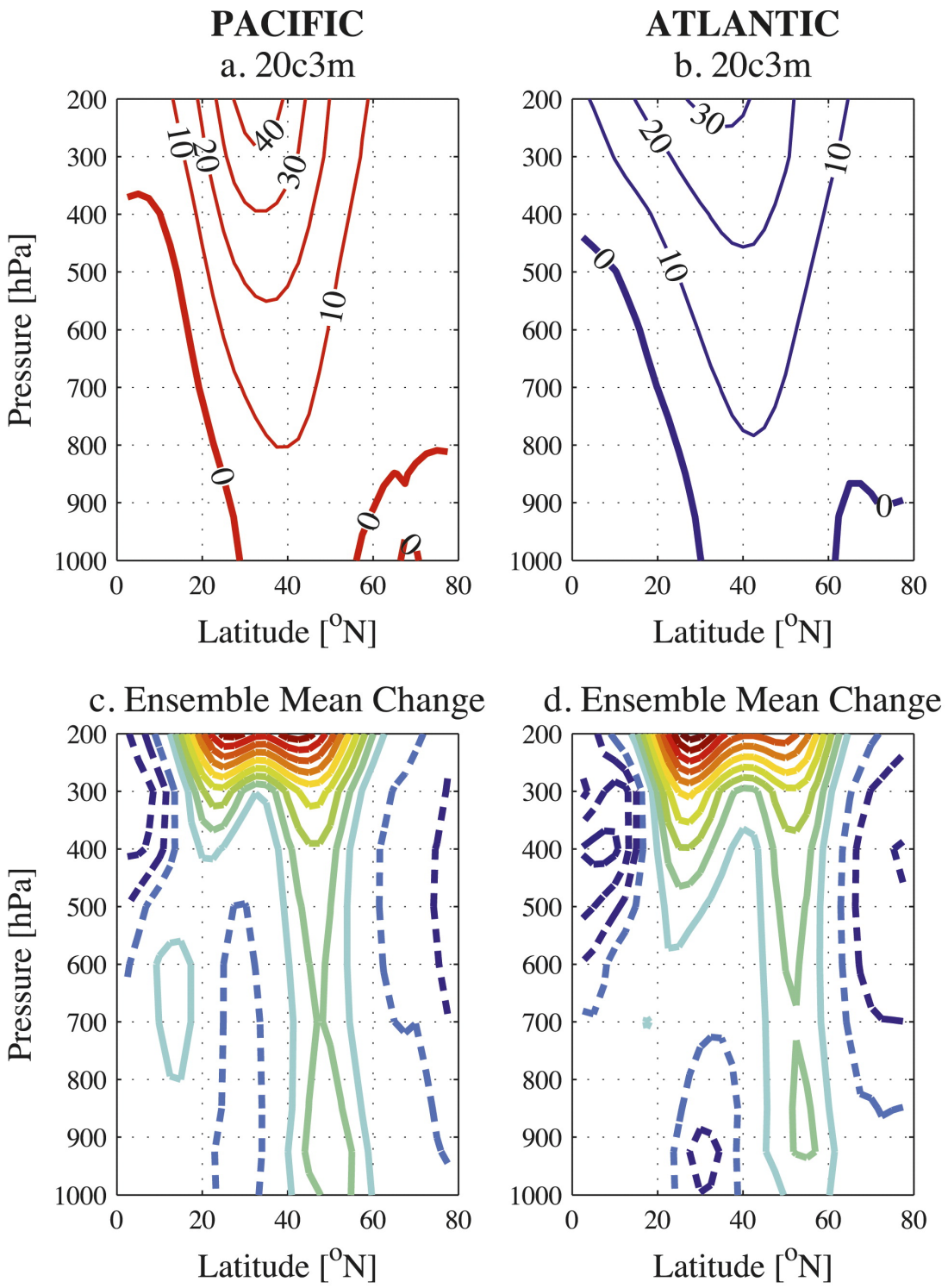

FIGURE 2.4: Modelled zonal-mean annual-mean ensemble-mean future zonal wind changes. The zonal-mean is taken over the Pacific (left) and Atlantic (right). The upper row shows a control period contoured every 10 m s^{-1} and the lower row shows the future change contoured every 0.25 m s^{-1}. From Delcambre et al. (2013).

day with a horizontal resolution of 1.25° at six pressure levels with a lead time T+6 to T+36 hours (ICAO, 2016).

2.3.1 Clear-Air Turbulence Forecasting

Using diagnostics to forecast clear-air turbulence allows pilots and flight planners to avoid turbulent regions. Fahey (1993) showed strategic planning can lead to not only a reduction in injuries, but also costs. The use of diagnostics is also the only way to operationally forecast turbulence, as the turbulent eddies that affect aviation are smaller than the resolution of global atmospheric models. Turbulence predictors that have a deformation term are particularly good at forecasting CAT. Knox et al. (2008) tried to understand the previous work that linked deformation to Kelvin–Helmholtz instabilities. This included frontogenesis that could initiate Kelvin–Helmholtz instabilities, or mesoscale waves that would then break down and form turbulence (Mancuso & Endlich, 1966). They found that neither could fully explain the relationship between deformation and turbulence, so instead found a link between deformation and inertia–gravity wave generation via the Lighthill–Ford theory. In certain atmospheric environments close to shear instability (low Richardson number), gravity waves emitted by imbalance destabilize the atmosphere, locally reducing Ri to below the critical value of 0.25 and leading to Kelvin–Helmholtz instabilities, generating waves which break down and form turbulence (Miles & Howard, 1964; Dutton & Panofsky, 1970). This explains the success of empirical diagnostics such as the Ellrod TI1:

$$\mathrm{TI1} = \left[\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)^2\right]^{1/2} \times \left[\left(\frac{\partial u}{\partial z}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2\right]^{1/2}, \quad (2.2)$$

where u is the horizontal wind velocity in the East-West direction, v is the horizontal wind velocity in the North-South direction, x is distance in the East-West direction, y is distance in the North-South direction and z is distance in the vertical. Ellrod TI1 found in equation (2.2) is one of the best shear turbulence forecast diagnostics, and has been found to forecast up to 75% of all CAT cases (Ellrod & Knapp, 1992). This is why a deterministic Ellrod TI1 forecast is routinely used by the WAFCs.

McCann et al. (2012) furthered the work by Knox et al. (2008) to try and improve the forecast based on spontaneous imbalance. They made an algorithm that could be used operationally by improving the forecast below FL200 (which was a problem with the current diagnostics). The apparent high bias

in turbulent kinetic energy dissipation above the tropopause is reduced by including the turbulent kinetic energy production from the environment and the locally produced turbulent kinetic energy. McCann et al. (2012) also converted the turbulent kinetic energy into an Eddy Dissipation Rate (EDR), which is now a standard measure of turbulence in the atmosphere as it does not depend on aircraft size (WMO, 2003).

There are many turbulence diagnostics, each with its own strengths and weaknesses. All forecast CAT and are unable to forecast convective turbulence or mountain wave turbulence. Since each individual diagnostic is unable to forecast every CAT event, Sharman et al. (2006) generated the Graphical Turbulence Guidance (GTG) system for forecasting turbulence. They selected multiple turbulence diagnostics and then used PIlot REPortS (PIREPS) to understand the diagnostics' performances. A weighting system can then be added depending on the skill of each diagnostic to produce the best overall forecast possible. The study showed that this was an improvement over the single turbulence diagnostic, and is therefore worth doing. Over the years this forecast has been improved, and more recently mountain wave turbulence predictors have been added to the GTG system. At the moment CIT is not forecast by the GTG system, but it has been shown it would be a useful addition (Gill & Stirling, 2013).

2.3.2 Convective Turbulence Forecasting

Convection is one of the main sources of turbulence. Convective turbulence is not yet forecasted using numerical models operationally, but instead forecasters manually put the location of convection on the significant weather (SIGWX) charts used by pilots and flight planners. Gill & Stirling (2013) showed that using a convective diagnostic from numerical weather prediction output can forecast many convective turbulence events. Therefore, combining the shear turbulence predictors and the convective predictors (similar to GTG), offers the promise to greatly improve the forecast skill. The convective indicators used in the study included convective precipitation rate, convective precipitation accumulation and Convective Available Potential Energy (CAPE). Gill & Stirling (2013) concluded that it is possible to improve turbulence forecasts by not only combining different predictors for the same turbulence source (e.g. Ellrod TI1, Brown index, etc.) but also combining predictors for different turbulence sources as well (e.g. clear-air turbulence and convective turbulence). They suggest more work is needed in this area, but there is the potential to

greatly increase the forecast skill by adding in these extra turbulence diagnostics and forecasting a greater proportion of turbulence events.

2.3.3 Mountain Wave Turbulence Forecasting

An algorithm for forecasting mountain wave turbulence was described by Turner (1999). Mountain waves are a special case of gravity waves and they are smaller than the numerical resolution of the models, so therefore the wave stress has to be parameterised. To do this, the surface winds and stability are used to work out a wind stress, and then the wind stress is used to work out the vertical profile. The wind stress is computed for every grid point, taking into account the abruptness of orography (as that is the location of the mountains). The surface stress is passed upwards from the surface (unless a hydraulic jump, critical level or saturation stress is diagnosed), and therefore less stress is transfered to the next level. Above the maximum point that gravity wave stress can be sustained, the waves break and thus turbulence forms. The models can use the wave stress at each model level to calculate the location and height at which the gravity wave stress is too high, indicating the turbulence location in the model. For mountain wave formation, the higher the gravity wave propagates, the more likely it is to break. This is because the maximum gravity wave stress the atmosphere can hold decreases with increased wind shear and lower air density. Mountain wave turbulence is found to be associated most strongly with the mid-latitude westerlies and mountains perpendicular to the flow.

The new GTG3 system includes mountain wave turbulence diagnostics alongside the typical CAT diagnostics (Sharman & Pearson, 2017). These authors have also used a different approach to forecasting MWT and use a combination of low-level wind combined with CAT diagnostics. The authors produced 14 new MWT diagnostics, which showed an increased forecast skill in MWT prone areas, supporting their use in the new GTG3 system.

2.3.4 Ensemble Forecasting

One of the main problems with the turbulence forecasting methods outlined in Sections 2.3.1, 2.3.2, and 2.3.3 is that they do not convey uncertainty. To resolve this issue, Gill & Buchanan (2014) and Buchanan (2016) trialled the use of ensemble forecasting for aviation turbulence. An ensemble is a collection of forecast runs, each of which is considered to be equally likely. By using an ensemble, the uncertainty in the forecast can be found. In these studies,

the Met Office Global and Regional Ensemble Prediction System (MOGREPS) was used, which was made operational in 2008 (Bowler et al., 2008). Multiple turbulence predictors were considered, including the Dutton (1980) index, the Ellrod & Knapp (1992) indices, the Brown (1973) index, the Lunnon index (Roach & Dixon, 1970), the Richardson number and two convective predictors. Using the Derived Equivalent Vertical Gust (DEVG) as a truth in the verification, the studies found that by using the ensemble forecast they increased the forecast skill for most of the turbulence diagnostics.

The studies by Gill & Buchanan (2014) and Buchanan (2016) then combined the predictors in a manner similar to Sharman et al. (2006), by using an iterative scheme to maximise the forecast skill. Again, the studies showed that the ensemble forecast was more skilful than a single model deterministic forecast. A probabilistic forecast would also be more useful for pilots and flight planners, as understanding the uncertainty of the forecast can help them to make the best decision possible to maintain safety of passengers and crew, while still flying the optimum routes and reducing flight times and fuel consumption. Further work is needed in ensemble forecasting before it can be used operationally, such as studying whether multi-model ensembles can improve the forecast. Also more research is needed to include a mountain wave turbulence predictor alongside the convective and shear turbulence predictors. However, it is clear that ensembles can be beneficial and should eventually be used operationally.

2.3.5 Multi-model Ensemble Forecasting

Multi-model ensemble forecasting is a technique where two or more ensemble models are combined to create either a standard equally weighted multi-model super ensemble, or an optimised weighted multi-model ensemble. The weighted multi-model ensemble uses the performance of each individual ensemble and tries to optimise the forecast to maximise the skill. Park et al. (2008) showed that when combining multiple ensembles, some of the results didn't see a large improvement when one of the models already had a well tuned spread (which was the European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System (EPS) in this example). However, each model has its own strengths and weaknesses, and so it is possible to benefit from combining ensembles. For example, the ECMWF EPS does not have a large spread in the tropics, and therefore combining multiple ensembles could

improve the forecast skill here. Buizza & Palmer (1998) also showed that increasing model resolution and not just ensemble size can increase forecast skill. However it is also not just forecast skill that can be improved by increasing ensemble size. Richardson (2001) suggests the relative economic value of a forecast can be greatly improved by increasing the number of ensemble members even if the skill doesn't improve greatly. By creating a multi-model ensemble, it is a way of adding ensemble members without the computation for the forecast centre, and even if forecast skill is not greatly increased, the forecast value might increase.

Atger (1999) studied the skill of the EPS, and found a poor person's ensemble (ensemble of single deterministic output from multiple-models) was more skilful than either of the ECMWF EPS or the National Centers for Environmental Prediction (NCEP) EPS up to T+144 hrs. This is important to note because it suggests that combining models is more important than simply increasing the number of ensemble members in an EPS. They discuss the reasons for the improved skill and suggest it is likely because the resolution of the poor person's ensemble is higher than either the ECMWF or NCEP EPS, however the use of multiple centres and not just one EPS does show some benefit. Ziehmann (2000) also showed that a multi-model approach performs better in most aspects over a single model ensemble. Other areas of meteorology such as The International Grand Global Ensemble (TIGGE) project (Swinbank et al., 2016) have already looked at combining ensembles from different centres around the world. A particular research area using multi-model ensembles operationally is tropical cyclone forecasting (Krishnamurti et al., 2000; Vitart, 2006; Titley & Stretton, 2016).

2.3.6 Forecast Verification

Turbulence forecast verification is difficult because objective verification datasets are limited. Some previous studies resolved this issue by using PIlot REPorts (PIREPs) (Tebaldi et al., 2002; Kim & Chun, 2011) to identify regions of turbulence based on a semi-quantitative scale from light to extreme, but these can be unreliable (Schwartz, 1996; Kane et al., 1998; Sharman et al., 2014). PIREPs are subjective, in the sense that a more experienced pilot may catagorise an event as moderate, but an inexperienced pilot may record it as severe. Also PIREPs are aircraft dependent, so a smaller aircraft will experience stronger turbulence than a larger aircraft in the same volume of air. PIREPs also have poor spatial

reliability as they tend to be located in turbulence, so null turbulence events are rarely recorded (Kane et al., 1998).

To avoid the above problems, there are two main aircraft-independent measures, which are the Eddy Dissipation Rate (EDR) and the Derived Equivalent Vertical Gust (DEVG). These measures are calculated by using high-resolution automated aircraft data. The EDR is calculated following Haverdings & Chan (2010) using:

$$\text{EDR} = \frac{\sigma_w}{\sqrt{1.05 V_a^{\frac{2}{3}} (\omega_1^{-\frac{2}{3}} - \omega_2^{-\frac{2}{3}})}}, \tag{2.3}$$

where σ_w is the running mean of the vertical wind shear over a 10 second time window, V_a is the true airspeed, and ω_1 and ω_2 are cut-off frequencies set at 0.1 and 2 Hz respectively. The DEVG (Tenenbaum, 1991; Gill, 2014, 2016; Kim et al., 2017) is calculated following Truscott (2000) using:

$$\text{DEVG} = \frac{Am|\Delta n|}{V}, \tag{2.4}$$

where $|\Delta n|$ is the peak modulus value of the deviation of the aircraft acceleration from $1g$ in units of g, m is the total mass of the aircraft (metric tonnes), V is the calibrated airspeed at the time of the observation (knots), and A is an aircraft-specific parameter that varies with flight conditions and can be calculated using:

$$A = \overline{A} + c_4(\overline{A} - c_5)\left(\frac{m}{\overline{m}} - 1\right) \tag{2.5}$$

and:

$$\overline{A} = c_1 + \frac{c_2}{c_3 + H(kft)}, \tag{2.6}$$

where H is the altitude (thousands of feet), $\overline{m}$ is the reference mass of the aircraft (metric tonnes), and parameters c_1 to c_5 depend on the aircraft's flight profile as outlined by Truscott (2000).

Unlike PIREPs, DEVG is aircraft independent and therefore all observations from multiple aircraft can be combined to create a consistent observational database. Also unlike PIREPs, DEVG reports all results, including null reports. Because it is automated, the entire flight is recorded and therefore the only limit in coverage is the location the aircraft fly. So if every commercial aircraft had this capability, then all turbulence events could be recorded and forecast verification improved. There are limitations to using the DEVG data set, as aircraft manoeuvres and active control techniques can enhance or dampen vertical accelerations of aircraft, leading to over- or under-representation of

the vertical gusts (WMO, 2003). However, despite these limitations, using an aircraft-independent measure (such as EDR or DEVG) is preferable to pilot reports (PIREPs).

Chapter 3

Reanalysis

This chapter is based on our paper 'Can climate models successfully diagnose clear-air turbulence and its response to climate change?' submitted to the *Quarterly Journal of the Royal Meteorological Society*. Climate models have been used to predict how climate change will impact Clear Air Turbulence (CAT) over the North Atlantic (Section 2.2), however no study has assessed how well climate models can predict the location of CAT and its response to changes in the climate system. Therefore this chapter will compare a climate model (HadGEM2-ES) to reanalysis data (ERA-Interim), which is a record of the global atmospheric circulation (Dee et al., 2011), to understand how well the climate model performs on a global scale. It could in principle be the case that all climate models forecast CAT in a similar location, but that location is wrong due to biases in the models. This chapter will answer the important question of whether climate models can diagnose the location of CAT as found in a reanalysis data set. Can the climate models predict the location of the changes in CAT due to the response to climate change over the reanalysis period? And finally, does the climate model correctly predict the global percentage change in CAT with climate change over the reanalysis period? Understanding how the climate model compares to the true atmosphere will help us understand if the results found in previous studies are correct and if the warnings of an increase in CAT with climate change are justified.

3.1 Methodology

The data used in this study is part of the historical climate simulation using the Met Office Hadley Centre HadGEM2-ES model (Jones et al., 2011) which is part of the fifth Coupled Model Intercomparison Project (CMIP5) ensemble (Taylor et al., 2012). This data set is the only CMIP5 model for which six-hourly output fields have been archived on a suitable set of upper tropospheric and

lower stratospheric pressure levels. This data is then compared to the ERA-Interim reanalysis data from the European Centre for Medium Range Weather Forecasting (ECMWF) (Dee et al., 2011). We use 38 years of data for each of the two data sets in this study. From the HadGEM2-ES model we use the 'historical' data which is based on climate forcings from 1968-2005, and the reanalysis data set is from 1979-2016. The time periods are not the same as ERA-Interim does not start until 1979 and the historical data set finished in 2005. Therefore to have as large a data set as possible (which is 38 years from the reanalysis data) the time periods are slightly offset.

We study two flight levels, one at 200 hPa (12 km / 39,000 ft. / FL390) and 250 hPa (10 km / 34,000 ft / FL340) using data captured every 6 hours. The climate model has a horizontal grid of 1.25° in latitude and 1.875° in longitude, giving 192 × 144 grid boxes globally. The vertical resolution available to us is limited with only two upper tropospheric levels (200 and 250 hPa) however we need to calculate vertical derivatives. To do this we have to use 200 and 250 hPa which therefore means it is centred around 225 hPa rather than the flight level we are studying. This is a much coarser resolution compared to the reanalysis data, which has a horizontal grid 0.70° in latitude and 0.70° in longitude, giving 512 × 256 grid boxes globally. The vertical resolution is also much higher for the reanalysis data as there is data at 300, 250, 200, 150 hPa meaning vertical derivatives can be centred around each flight level. These differences could impact the location of the predicted CAT. To understand how horizontal and vertical resolution impacts the location of CAT, we have used Iris (Met Office, 2013) to regrid the data through linear analysis to have the same horizontal grid as the climate model before calculating the turbulence diagnostics (1.25° in latitude and 1.875° in longitude with only two vertical levels 200 and 250 hPa).

This study uses seven turbulence diagnostics used by Williams & Joshi (2013) that are also part of the Graphical Turbulence Guidance (GTG) system (Sharman & Pearson, 2017). This includes the negative Richardson number (Ri), Colson–Panofsky index, frontogenesis function, Ellrod & Knapp (1992) TI1, wind speed times directional shear, nonlinear balance equation and North Carolina State University index 1. This study also follows on from Williams (2017) by comparing five turbulence strength categories: light, light-to-moderate, moderate, moderate-to-severe and severe. The thresholds are based on the cube root of the eddy dissipation rate, which is proportional to the vertical acceleration of an aircraft experiencing turbulence (MacCready Jr, 1964). The thresholds used for each of the models will be different because the turbulence

diagnostics are calculated using horizontal and vertical gradients. Therefore the maximum value that is calculated using one model, which would be severe turbulence, might not be equivalent to light turbulence in another model. To compare consistently, we calculate the probability distribution of all seven turbulence diagnostics for the first 20 years of each data set individually. The top 0.1% (99.9–100%) of the probability distribution is taken to be severe turbulence. The next 0.1% (99.8–99.9%) is moderate-to-severe, the next 0.2% (99.6–99.8%) is moderate, the next 0.5% (99.1–99.6%) is light-to-moderate and the next 2.1% (97–99.1%) is light turbulence. We calculate the thresholds based on the first 20 years of data for each of the data sets. After this, when running the analysis again for the full 38 years, any time a value exceeds one of these thresholds it is classed as a turbulence event. It is then possible to compare each turbulence diagnostic, climate model and turbulence strength category, flight level and season to understand how well the climate model can represent CAT. It is also possible to separate the data set into two 19 year periods to measure the climate change impact of CAT. We can see how changes in the recent climate system have already impacted CAT by calculating the percentage change between the first and second half of the data set.

3.2 Results

3.2.1 Global CAT Distribution

Understanding how the climate model compares to the real atmosphere is vital in supporting previous studies (e.g. Williams & Joshi, 2013; Williams, 2017). To do this we have compared 7 turbulence diagnostics currently used in the GTG system (Sharman & Pearson, 2017) between the climate model HadGEM2-ES and reanalysis data from ERA-Interim. An example plot of the global spatial distribution in Northern Hemisphere winter (December January February, DJF) at 200 hPa is shown in Figure 3.1. The left column shows the historical data from the HadGEM2-ES climate model, the right column is the reanalysis data from ERA-Interim, and the middle column shows the reanalysis data but regridded to have the same spatial resolution as the historical data from the climate model before computing the CAT diagnostics. This allows us to not only compare how well the climate model correctly diagnoses the location of CAT, but also how the resolution of the data set impacts the location of CAT as well. Looking at Figure 3.1 we first observe that some of the diagnostics are

more confined to the tropics and others are distributed more around the mid-latitudes. For example, the negative Richardson number, which is a measure of buoyancy and vertical wind shear, is predominantly found in the tropics, where buoyancy is lower than in the mid-latitudes. This suggests some regions have light turbulence about 20% of the time. As stated in section 3.1, the thresholds are based on a percentile and therefore the global average of occurrence will be 2.1%. For Ri and Colson–Panofsky, most of the results occur in the tropics and therefore there are very few events in the mid-high latitudes. Compared to the Ellrod TI1, this is predominantly found in the mid-latitudes and circumnavigates the Earth near the jet stream. This is what we would expect since TI1 is deformation $\times$ vertical wind shear, which will be strongest around the high wind speeds of the jet stream.

When comparing the historical data set from the climate model to the two reanalysis data sets, the spatial distribution is more similar to the reanalysis data regridded onto the climate grid than it is to the reanalysis on its native grid. This is an important result because it shows that the resolution of the model used will have an impact on spatial distribution and therefore the location of turbulence forecast. The impact of resolution on the forecast would need to be investigated fully using a verification method to understand if a higher resolution gives a more skilful turbulence forecast. However, what we can see from Figure 3.1 is that the resolution has an impact on the spatial distribution and as a result, the same resolution is a better match to the climate model as would be expected.

This study has looked at the global spatial distribution of CAT throughout the year, and Figure 3.2 is a plot of the spatial distribution of the seven turbulence diagnostics across the three models in the Northern Hemisphere summer (June July August, JJA) at 200 hPa. The results are similar to Figure 3.1 but as would be expected for a different season, they are not the same. One particular difference to note is the increase in Ellrod & Knapp (1992) TI1 over the Southern Ocean. Although this is not a busy airspace it does see an increase in turbulence as the jet stream is typically stronger in the winter and JJA is the Southern Hemisphere winter. Some of the other diagnostics also see a shift, for example around the tropics the Colson–Panofsky index and frontogenesis function have shifted north. This shift is possibly due to the changes in the Hadley cell (which is a pair of thermally direct circulations, either side of the equator and move north or south depending on the season (Dima & Wallace, 2003)), as it moves north in the Northern Hemisphere summer. The results again show that the climate model is diagnosing the location of the turbulence

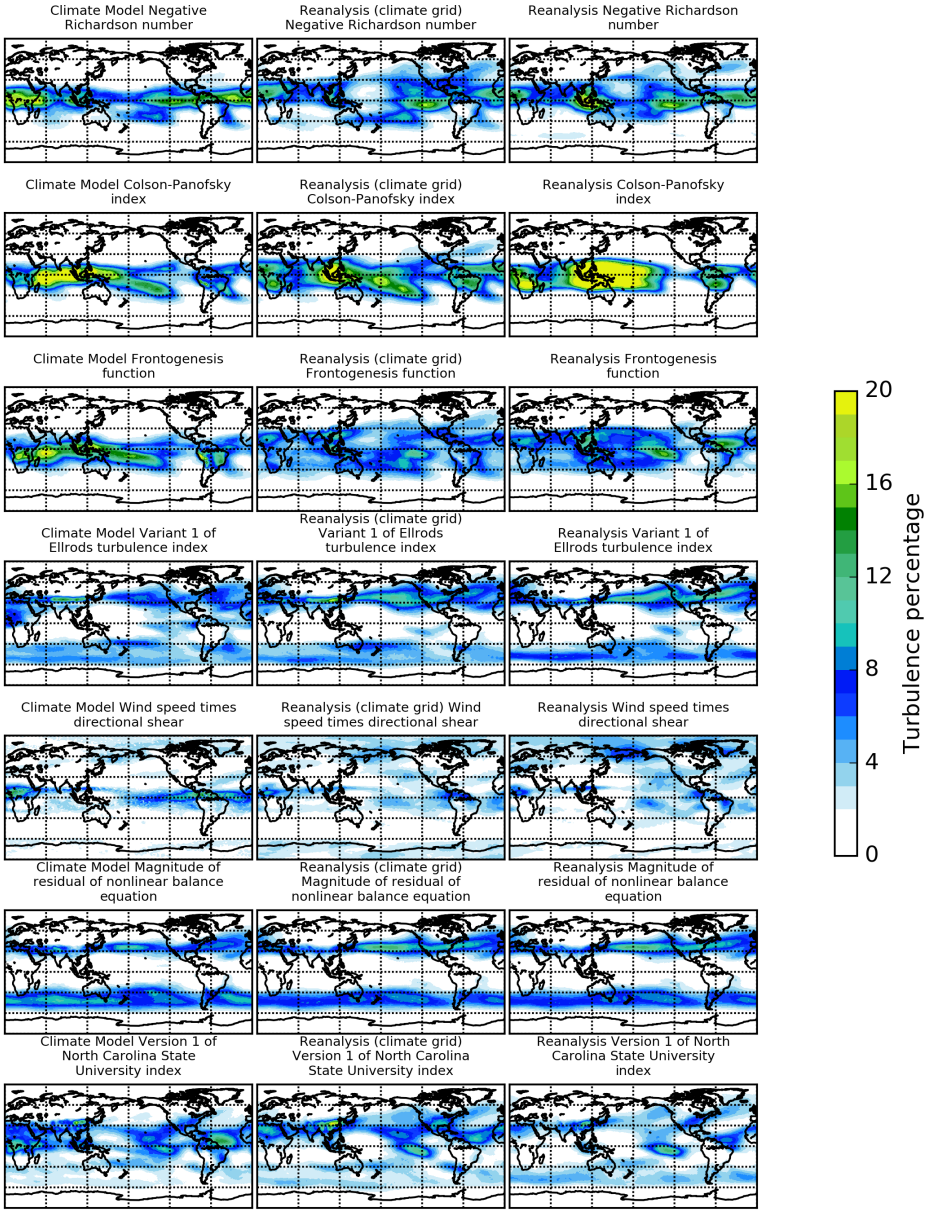

FIGURE 3.1: Plot comparing the spatial distribution of the probability of encountering clear-air turbulence for the seven turbulence diagnostics in December, January and February (DJF) at 200 hPa for light turbulence using the HadGEM2-ES historical data set reanalysis data with the resolution of the climate model and the reanalysis data with its original resolution.

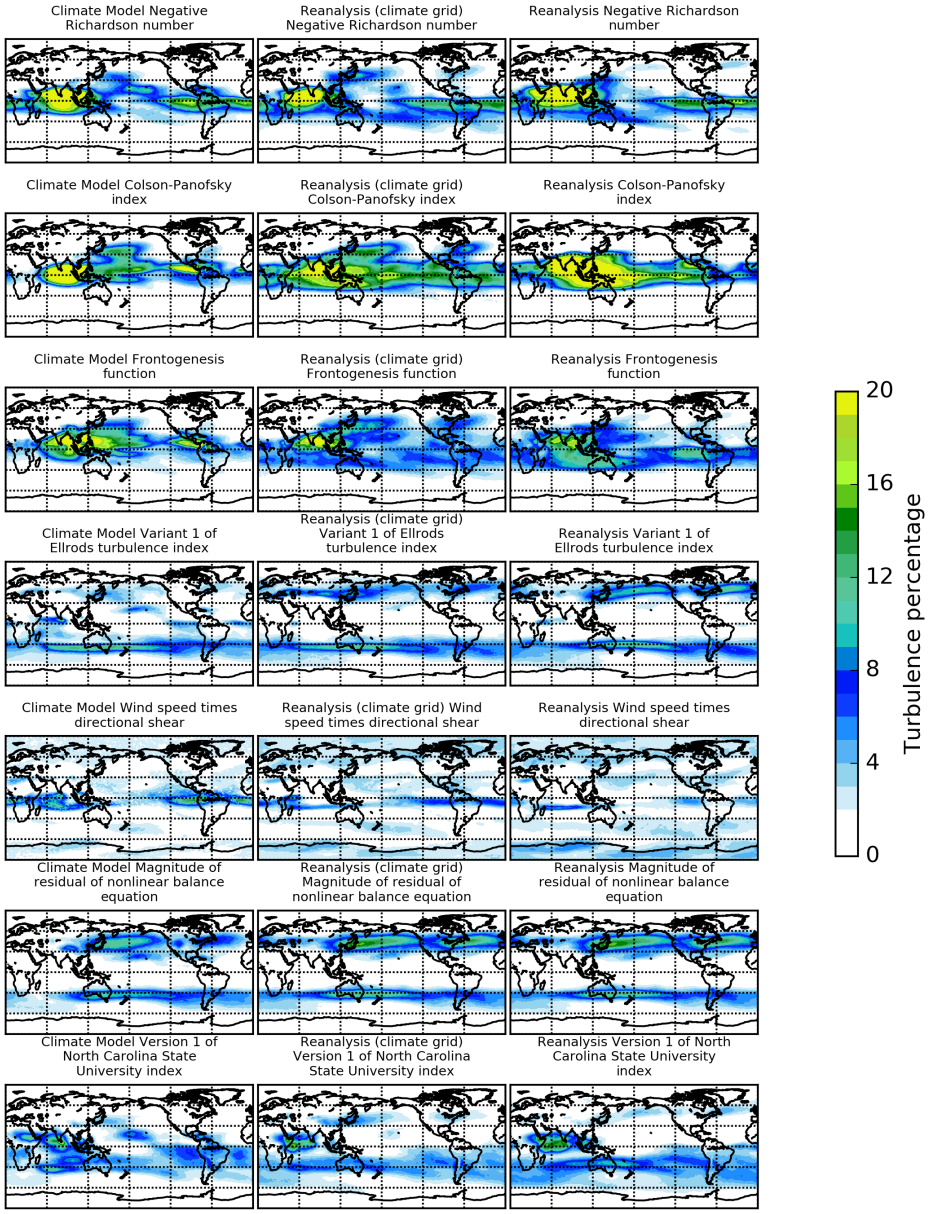

FIGURE 3.2: Plot comparing the spatial distribution of the probability of encountering clear-air turbulence for the seven turbulence diagnostics in June, July and August (JJA) at 200 hPa for light turbulence using the HadGEM2-ES historical data set, reanalysis data with the resolution of the climate model and the reanalysis data with its original resolution.

very well, supporting our use of it in other studies.

3.2.2 Global CAT Comparison

We want to quantify whether the main differences in Figure 3.1 and 3.2 are along the rows or down the columns, and therefore identify whether the main source of uncertainty is the use of data sets or the use of seven diagnostics. To quantify how similar the models are for each turbulence index and how similar each index is within the models we have calculated the standard deviation (which is a measure of how similar the data sets are, the lower the number, the more similar the models). Table 3.1 shows the standard deviation across each model for each index (left to right in Figure 3.1) and we have included the cosine latitude factor to down-weight high-latitude grid boxes with smaller areas. We have calculated this for each season and each turbulence strength category but Table 3.1 only shows the four seasons for 200 hPa at the 97% threshold (light turbulence). The lower three rows in Table 3.1 are the standard deviation across each turbulence diagnostic for each model (top to bottom in Figure 3.1). The results show that the standard deviation across the models is typically much lower than the standard deviation between indices in all seasons. This is what we expected to see looking at Figures 3.1 and 3.2, as the models agree quite well in the location of turbulence and each turbulence index forecasts turbulence in very different locations as discussed in section 3.2.1. It is clear from Table 3.1 that there is a lower standard deviation in the Northern Hemisphere spring and autumn and the nonlinear balance has the best agreement between models with Colson–Panofsky having the worst agreement between models.

To summarise the results in Table 3.1, we compare the standard deviation between each season and turbulence strength category. We do this by calculating the ratio between the standard deviation across models and across diagnostics within each model. To do this we average the standard deviation in the table for the upper seven rows of the table (across models for each diagnostic, left to right in Figure 3.1 and 3.2) and the lower three rows of the table (across diagnostics for each model, top to bottom in Figure 3.1 and 3.2). Then we calculate the ratio between these two averages and the results are shown in Table 3.2. The higher the ratio, the lower the standard deviation is across models for each diagnostic (left to right in Figure 3.1 and 3.2) than across diagnostics for each model (top to bottom in Figure 3.1 and 3.2). The top line for light turbulence is the ratio calculated from the results in Table 3.1 and shows

Turbulence index	DJF	MAM	JJA	SON
Richardson	0.68	0.61	0.69	0.50
Colson–Panofsky	1.00	0.96	1.06	0.87
Frontogenesis	0.75	0.62	0.78	0.63
Ellrod TI1	0.41	0.36	0.48	0.33
Wind*Directional shear	0.38	0.30	0.40	0.32
Nonlinear balance	0.22	0.20	0.27	0.24
NCSU1	0.53	0.48	0.50	0.44
Historical	1.46	1.23	1.44	1.22
Reanalysis climate grid	1.46	1.37	1.63	1.36
Reanalysis	1.75	1.62	1.84	1.56

TABLE 3.1: Table of the global standard deviation of the probabilities in Figures 3.1 and 3.2 in units of % at 200 hPa in December January February (DJF), March April May (MAM), June July August (JJA), September October November (SON) for light turbulence across three data sets (HadGEM2-ES historical data set, reanalysis data with the resolution of the climate model and reanalysis data with its original resolution) for each of the seven turbulence diagnostics (left to right in Figure 3.1 and 3.2) and across the seven turbulence diagnostics for each of the three data sets (top to bottom in Figure 3.1 and 3.2).

the standard deviation across diagnostics for each model is nearly three times larger than across the models for each diagnostic. This is an important result as it suggests that the main uncertainty in our results comes from deciding which turbulence diagnostic is best to use rather than the model we have selected. Table 3.2 shows the results for both 200 and 250 hPa, and shows that our findings are consistent across multiple flight levels, all turbulence strength categories and throughout all seasons. What we do see in Table 3.2 is the ratio is much higher for the lower turbulence strength categories. Even for the highest turbulence strength category it is nearly twice the standard deviation and therefore still agrees with our conclusion that the greatest uncertainty in turbulence location comes from selecting which turbulence diagnostic to choose and not the climate model. This helps support the use of a climate model in climate change studies that have tried to predict the changes in CAT with climate change (e.g. Williams & Joshi, 2013; Williams, 2017).

Strength Category	DJF		MAM		JJA		SON	
	200 hPa	250 hPa	200 hPa	250 hPa	200 hPa	250 hPa	200 hPa	250 hPa
Light	2.76	2.58	2.80	2.65	2.77	2.40	2.90	2.46
Light-to-moderate	2.36	2.32	2.35	2.27	2.51	2.29	2.56	2.26
Moderate	2.10	2.13	2.14	2.08	2.31	2.16	2.38	2.10
Moderate-to-severe	1.94	2.01	2.00	1.97	2.15	2.06	2.22	2.00
Severe	1.81	1.86	1.88	1.80	1.89	1.87	2.09	1.86

TABLE 3.2: Table of the ratio between the global-mean standard deviation across the three data sets for each turbulence diagnostic (left to right in Figure 3.1 and 3.2) and across each turbulence diagnostic for each data set (top to bottom in Figure 3.1 and 3.2) at 200 and 250 hPa for December January February (DJF), March April May (MAM), June July August (JJA), September October November (SON) and five turbulence strength categories.

3.2.3 Global CAT Change Distribution

To understand how CAT has changed with the changing climate, we divide the 38 years into two 19 year periods (e.g. 1979-1997 and 1998-2016 for the reanalysis data set) so we can measure the number of turbulence 'hits' in each time period and then find the percentage change between the two. Williams & Joshi (2013); Williams (2017) both showed there is an increase in CAT with climate change over the North Atlantic and here we will see if a climate model can predict the changes in CAT that is observed in reanalysis data. We would not expect to see a large increase here due to the small gap between study periods, and the climate system has had less greenhouse gas emissions and a shorter time to equilibrate than both Williams & Joshi (2013); Williams (2017). However we might expect to see some impact of climate change on CAT as we have seen the impact of climate change in other areas of meteorology e.g. temperature change and sea ice loss.

The results for the Northern Hemisphere winter (DJF) global percent change at 200 hPa for light turbulence is shown in Figure 3.3. The layout of the figure is the same as Figure 3.1 with the historical data from the climate model in the left column, the reanalysis data on its native grid in the right column, and the reanalysis data regridded onto the climate model grid in the middle column. It is clear in Figure 3.3 that the climate change response is less dependent on resolution than the spatial distribution (Figures 3.1 and 3.2) as the two reanalysis data sets are almost identical. Also the percentage change for the reanalysis data set is larger than the climate model (discussed in section 3.2.5). Differences along the rows compared to differences down the columns are evidently larger than in Figures 3.1 and 3.2. One of the main regions that show a surprising result is that in the North Atlantic the increases are not as large as in other

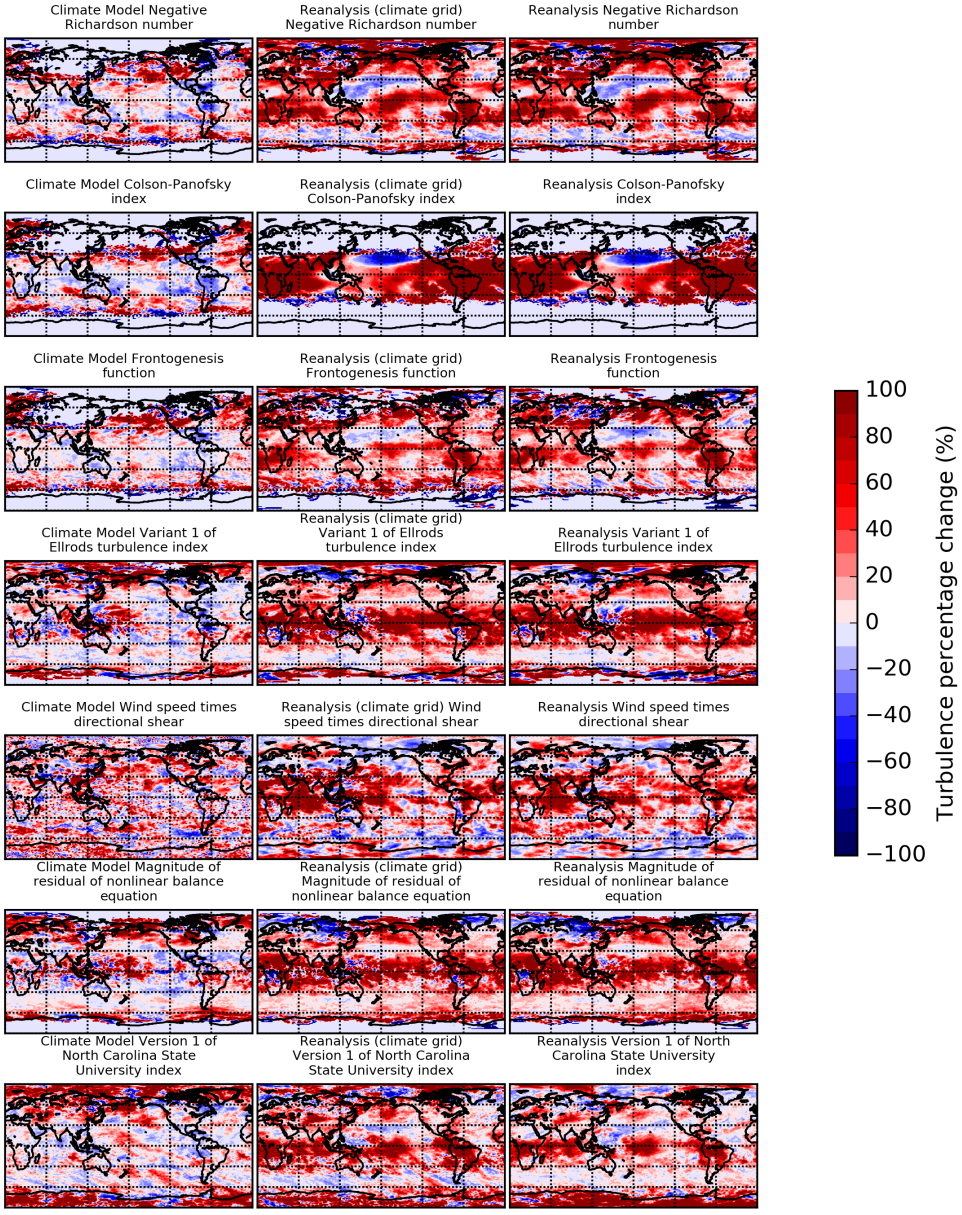

FIGURE 3.3: Plot comparing the percentage change in the probability of encountering CAT between the first and second half of the data set for the seven turbulence diagnostics in December, January and February (DJF) at 200 hPa for light turbulence using the HadGEM2-ES historical data set, reanalysis data with the resolution of the climate model and the reanalysis data with its original resolution.

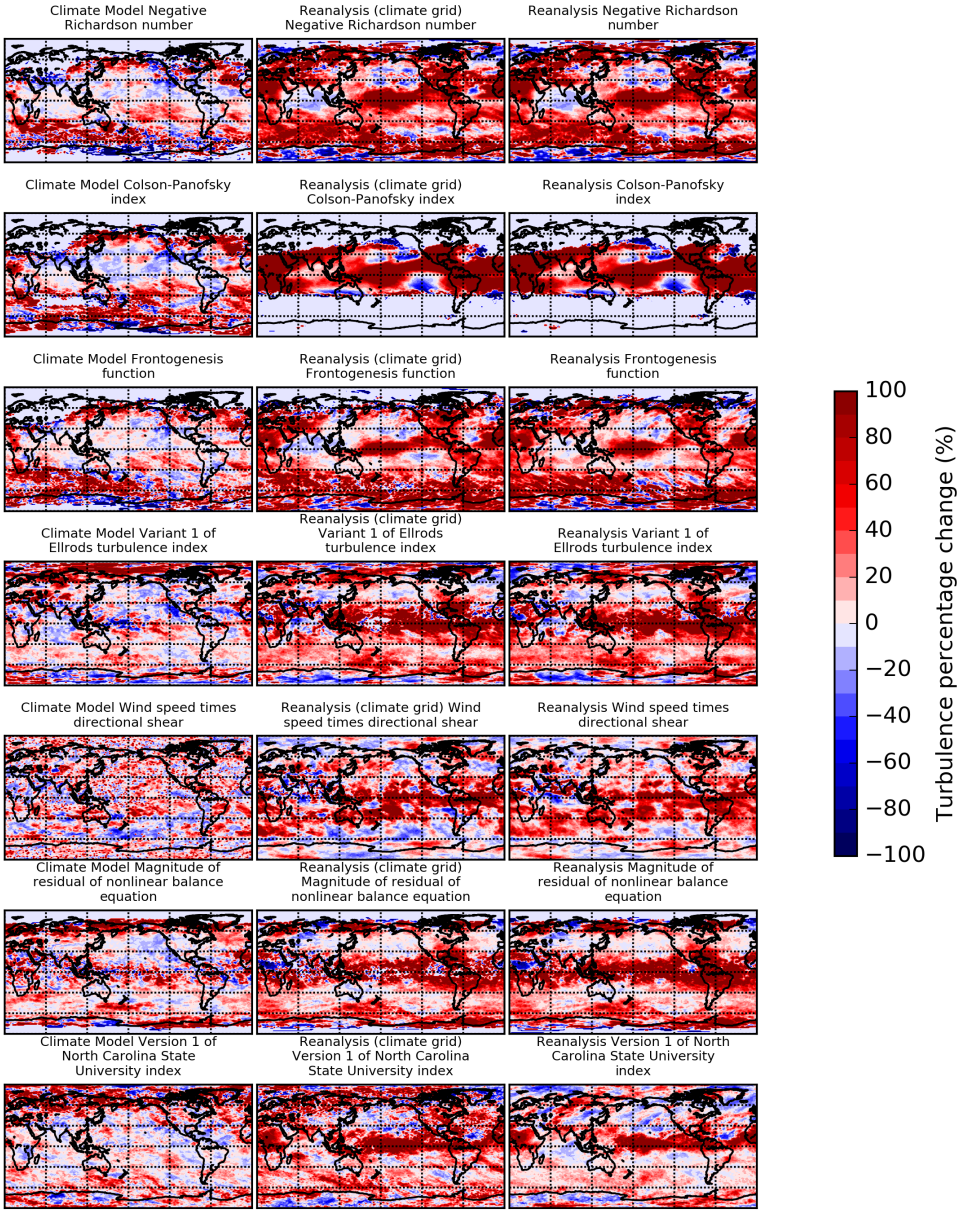

FIGURE 3.4: Plot comparing the percentage change in the probability of encountering CAT between the first and second half of the data set for the seven turbulence diagnostics in June, July and August (JJA) at 200 hPa for light turbulence using the HadGEM2-ES historical data set, reanalysis data with the resolution of the climate model and the reanalysis data with its original resolution.

areas, given that both Williams & Joshi (2013); Williams (2017) showed large increases in CAT over the North Atlantic.

Figure 3.4 is a plot of the Northern Hemisphere summer (JJA) at 200 hPa for light turbulence and it shows a similar distribution as DJF. There is however a slight increase in the Southern Hemisphere which, as in Figure 3.2, is what we would expect as typically the jet stream (which is a dominant cause of CAT) is strongest in the winter, which is JJA for the Southern Hemisphere. Therefore the predicted change in the upper tropospheric jet stream with climate change (Delcambre et al., 2013) would be strongest in the winter, and therefore a larger increase in the winter for both Hemispheres is expected. In both seasons the tropics see a large increase with some increases in the mid latitudes as well. This increase in the tropics is particularly interesting for the Ellrod & Knapp (1992) TI1 as there are few events in the tropics (as shown in Figure 3.1 and 3.2). Therefore this increase is probably an increase of only a few events, but because the initial values are so low, any increase will lead to a large percentage change. It is also clear in the figures that the climate model is not able to predict the changes in the tropics, where CAT is less relevant compared to CIT, however the mid-high latitudes has good agreement for most indices.

3.2.4 Global CAT Change Comparison

We can compare the spatial distribution of the response to climate change as we did for the spatial distribution in Section 3.2.2. The results for the standard deviation across the models for each diagnostic and across all diagnostics for each model with light turbulence (97%-99.1%) for all seasons at 200 hPa are found in Table 3.3. The results show more variability in which season has the highest standard deviation. For the spatial distribution it was found that DJF and JJA had a higher standard deviation for all models and indices. However for the climate change distribution, we do not find the same consistency, and for some indices it is March April May (MAM) and September October November (SON) that had the higher standard deviation while others still have it as DJF and JJA.

The differences along the rows is larger than in Figures 3.1 and 3.2. Table 3.4 shows the ratio between the average standard deviation across models for each diagnostic (left to right in Figures 3.3 and 3.4), and the average standard deviation across diagnostics for each model (top to bottom in Figures 3.3 and 3.4) at both 200 and 250 hPa. When comparing to Table 3.2, the ratio is a lot lower. Therefore the uncertainty in the model we used is slightly higher. This

Turbulence index	DJF	MAM	JJA	SON
Richardson	20.92	20.63	23.33	26.49
Colson–Panofsky	32.12	43.91	38.47	46.82
Frontogenesis	23.19	27.59	30.00	30.16
Ellrod TI1	25.10	30.16	29.75	32.46
Wind*Directional shear	27.29	21.88	27.34	22.65
Nonlinear balance	23.61	27.17	32.57	29.20
NCSU1	15.38	14.49	16.63	15.07
Historical	29.05	25.90	30.88	28.85
Reanalysis climate grid	37.10	43.90	44.39	44.52
Reanalysis	43.12	54.77	56.25	59.56

TABLE 3.3: Table of the global standard deviation of the climate change signal shown in Figures 3.3 and 3.4 in units of % at 200 hPa in December January February (DJF), March April May (MAM), June July August (JJA), September October November (SON) for light turbulence across three data sets (HadGEM2-ES historical data set, reanalysis data with the resolution of the climate model and reanalysis data with its original resolution) for each of the seven turbulence diagnostics (left to right in Figure 3.3 and 3.4) and across the seven turbulence diagnostics for each of the three data sets (top to bottom in Figure 3.3 and 3.4).

is likely caused by the poor performance of the climate model in the tropics. The mid-high latitudes performs quite well for most diagnostics, which is why the uncertainty in which model is used is still lower than which index is used. What we do find however is the greater the turbulence severity, the better the model performs. This was the opposite to the spatial distribution as we found the more severe the turbulence, the lower the ratio. We also see that the upper flight level (200 hPa) has the greatest ratio which suggests the model performs better at the higher altitudes than the lower altitudes.

3.2.5 Climate Change

We can see from Sections 3.2.3 and 3.2.4 that there has been an increase in CAT during the reanalysis period. The change is based on the percentage difference between the first 19 (e.g. 1979-1997 for the reanalysis data set) years and the second 19 years (e.g. 1998-2016 for the reanalysis data set) of the 38 year data set. Table 3.5 shows the annual global percentage change at all turbulence strength categories and at both 200 and 250 hPa for the historical data from the climate model (left), the reanalysis data on its native grid (right) and the

Strength Category	DJF		MAM		JJA		SON	
	200 hPa	250 hPa	200 hPa	250 hPa	200 hPa	250 hPa	200 hPa	250 hPa
Light	1.52	1.42	1.56	1.42	1.55	1.47	1.53	1.41
Light-to-moderate	1.53	1.49	1.59	1.49	1.55	1.51	1.56	1.49
Moderate	1.58	1.51	1.63	1.52	1.58	1.53	1.64	1.52
Moderate-to-severe	1.59	1.52	1.64	1.54	1.63	1.54	1.65	1.54
Severe	1.60	1.62	1.57	1.61	1.61	1.63	1.61	1.64

TABLE 3.4: Table of the ratio between the global-mean standard deviation of percentage change in turbulence across the three data sets for each turbulence diagnostic, and across each turbulence diagnostic for each data set at 200 and 250 hPa for December January February (DJF), March April May (MAM), June July August (JJA), September October November (SON) and five turbulence strength categories.

reanalysis data on the climate model grid (middle). Here we see the climate model underpredicts the percentage increase in CAT, with the reanalysis data showing more than five times larger percentage increase. This is an interesting result and suggests that the climate model is sluggish in its climate change response, but also as seen in Figures 3.3 and 3.4, the response to climate change that has already occurred is on a much more global scale than predicted by the climate model. The reanalysis data suggests the increase has occurred all around the world and particularly around the tropics. This could therefore indicate that any global estimates found in Chapter 4 for the CAT response to climate change could be underestimates and therefore the ability to forecast turbulence will become even more important than first thought.

Strength Category	Climate model		Reanalysis climate grid		Reanalysis	
	200 hPa	250 hPa	200 hPa	250 hPa	200 hPa	250 hPa
Light	8.4%	7.4%	29.3%	22.9%	27.7%	21.8%
Light-to-moderate	8.6%	7.7%	36.2%	29.1%	37.1%	24.4%
Moderate	7.5%	6.3%	41.2%	30.6%	41.2%	25.6%
Moderate-to-severe	6.5%	5.6%	38.9%	27.9%	40.2%	25.8%
Severe	7.2%	6.3%	42.5%	27.4%	45.8%	29.6%

TABLE 3.5: Table of the annual global percentage change in turbulence for 200 & 250 hPa and five turbulence strength categories using the HadGEM2-ES historical data set, reanalysis data with the resolution of the climate model and the reanalysis data with its original resolution.

The results in Table 3.5 suggest CAT at 200 hPa will increase more than at 250 hPa which could become a problem if the next generation of aircraft prefer to fly at higher altitudes due to the improved fuel efficiency. The results also show that severe turbulence has increased more than light turbulence

over the reanalysis period. This study does not include the changes to Convectively Induced Turbulence (CIT) which may also be set to increase, as climate change increases deep convection, particularly around the tropics (Price & Rind, 1994; Reeve & Toumi, 1999). This again suggests that the previous studies by Williams & Joshi (2013); Williams (2017) are underestimates and CAT could become much more of an aviation hazard than it already is.

3.3 Summary and Discussion

Clear-Air Turbulence (CAT) is a major aviation hazard and recent research has suggested the frequency of CAT will increase in the future as the climate system changes (Williams & Joshi, 2013; Williams, 2017). One major question posed after these studies is: 'Can climate models accurately predict the location of CAT?' This is a very important question as if the climate models are not able to predict the location of CAT, then the results would be in doubt. This study aimed at answering this question by comparing the climate model HadGEM2-ES and its historical emissions scenario, to reanalysis data from ERA-Interim on its full grid and with the same data but on the same grid as the climate model.

The initial results were promising with the climate model showing very similar spatial distribution of turbulence, therefore the answer to the question: 'Can the climate model predict the location of CAT as found in a reanalysis data set (ERA Interim)?' is yes. Climate models can predict the location of CAT with the main uncertainty in the location of CAT coming from which turbulence diagnostic is the best, and not from which climate model was used. The climate model was a better fit for the reanalysis data on the climate grid than the full resolution, which suggests resolution does have an impact on forecasted CAT and therefore further study is needed to understand how resolution of the forecast models might impact the forecasts of CAT.

This study also compared how the models predict the impact of climate change. What was interesting is the climate model is more sluggish in its response to climate change and the results from the reanalysis data suggested that the response of CAT to climate change is on a much more global scale, than predicted in the climate model. Therefore the answer to the question: 'Can the climate models predict the location of the changes in CAT due to the response to climate change over the reanalysis period?' is yes for the most part, but there are some large underestimates in the tropics. These results show that the global response of CAT to climate change in Chapter 4 could be an underestimate.

So the answer to a slightly different question: 'Does the climate model correctly predict the global percentage change in CAT with climate change over the reanalysis period?' is no in this study, as there is a large underestimate of the global response. However, we have only considered two 19 year periods and not taken into account inter-annual and decadal variability. For example the Atlantic Multi-decadal Oscillation (AMO), El Niño Southern Oscillation (ENSO) and North Atlantic Oscillation (NAO) that could all impact the occurrence of aviation turbulence. This is an important area of study in the future because our results suggest that there might be more urgency to prepare in the future for more turbulence around the world and make sure that systems are in place to maintain a high level of safety in the aviation industry. Therefore it is important to understand how these annual and inter-annual variabilities impact our results and therefore how climate change is impacting CAT.

This study has answered some important questions, however a few still remain. We must understand how the model resolution impacts the location of CAT and therefore how forecasts might change in the future with increasing model resolution. Also with an increase in turbulence due to climate change, better turbulence forecast techniques must be implemented as well as the airframe manufacturers making sure the aircraft are capable of withstanding more frequent turbulence events. Other studies of how aviation turbulence as a whole, including changes in convection, could be impacted by climate change is vital to create a better picture for stakeholders on what needs to happen now to prepare for the future.

Chapter 4

Global Response to Climate Change

This chapter is based on our paper 'Global Response to Clear-Air Turbulence to Climate Change' published by Storer et al. (2017) in *Geophysical Research Letters*. We have seen in Section 2.2 that climate change will have an impact on Clear-Air Turbulence (CAT) in the future over the North Atlantic. We have also seen in Chapter 3 that over the last 38 years there has been a global increase in CAT as shown in the reanalysis data set ERA-Interim. However what has not been studied is the future global increase in diagnosed CAT with climate change at multiple flight levels, in all seasons and all turbulence strength categories.

4.1 Methodology

We use climate simulations that were performed with the Met Office Hadley Centre HadGEM2-ES model (Jones et al., 2011), which forms part of the fifth Coupled Model Intercomparison Project (CMIP5) ensemble (Taylor et al., 2012) and is the same model used in Chapter 3. This is again because it is the only CMIP5 model for which six-hourly output fields have been archived on a suitable set of upper tropospheric and lower stratospheric pressure levels. The six-hourly snapshots resolve the diurnal cycle and are therefore expected to provide a better representation of wind shear than the daily-mean CMIP3 fields used by Williams & Joshi (2013) and Williams (2017). The multiple pressure levels make it possible to calculate the vertical wind shear using second-order centred finite differences at both 200 hPa and 250 hPa, which correspond to typical cruising altitudes of approximately 12 km (39,000 ft or FL390) and 10 km (34,000 ft or FL340) respectively. The atmospheric model has a horizontal grid spacing of 1.25° in latitude and 1.875° in longitude, giving 192×144 grid boxes globally. The model used by the previous studies Williams & Joshi (2013) and Williams (2017) was the GFDL-CM2.1 model which formed part of the CMIP3 ensemble. The model had a coarser resolution of 2.0° by 2.5° with 24 vertical levels and daily average data, rather than the 6 hourly snapshot

found in the HadGEM2-ES model used in this study (Delworth et al., 2006; Gnanadesikan et al., 2006).

Two HadGEM2-ES simulations are analysed to calculate how climate change could impact CAT in the upper troposphere and lower stratosphere in future. Specifically, a pre-industrial control simulation (picontrol) is compared with a climate change simulation using the Intergovernmental Panel on Climate Change (IPCC) Representative Concentration Pathway 8.5 (RCP8.5) (Flato et al., 2013). The picontrol run is a base state that uses constant pre-industrial greenhouse gas concentrations to simulate the global climate before the industrial revolution. The RCP8.5 run assumes a net radiate forcing increase of 8.5 W m^{-2} by 2100 (Van Vuuren et al., 2011), which implies greenhouse-gas concentrations equivalent to around 1,370 ppmv of CO_2. We analyse 30 years of data for the future period 2050–2080 from RCP8.5 compared to 30 years of historic data from picontrol.

The present study focuses on CAT generated by wind shear and loss of balance, disregarding mountain waves and remote convection. For consistency, we calculate the same basket of CAT diagnostics as Williams & Joshi (2013) and Williams (2017), except that we exclude the potential vorticity diagnostic because it was found to give unrealistic results. They were unrealistic because the turbulence locations were confined to the summer hemisphere pole region which is not consistent with our understanding of the location of CAT. We define a threshold for each turbulence strength category and each CAT diagnostic in HadGEM2-ES, following Williams (2017) and Chapter 3. The thresholds are appropriate for a large, commercial airliner. The calibration is based on the cube root of the eddy dissipation rate, which is proportional to the vertical acceleration of an aircraft experiencing turbulence (MacCready Jr, 1964), and it is implemented as follows. First, the probability distribution is calculated for each of the 20 CAT diagnostics, using six-hourly global fields from the 30-year picontrol run of HadGEM2-ES in all seasons on a given pressure level. Then, the top 0.1% (99.9–100%) of the probability distribution for each diagnostic is taken to represent severe turbulence, the next 0.1% (99.8–99.9%) is moderate-to-severe turbulence, the next 0.2% (99.6–99.8%) is moderate turbulence, the next 0.5% (99.1–99.6%) is light-to-moderate turbulence, and the next 2.1% (97–91.1%) is light turbulence. The percentiles are derived from an assumed log-normal probability distribution that is fitted to observations. It follows from these percentiles that 0.1% of the global atmosphere at aircraft cruising altitudes contains severe-or-greater turbulence, 0.4% contains

Moderate-or-Greater (MoG) turbulence and 3.0% contains light-or-greater turbulence. Extreme turbulence is neglected because of its rarity. To calculate regions of CAT, every time an index exceeds one of the above threshold values, it is classed as a turbulence event.

4.2 Results

4.2.1 Climate Model Comparison

As an initial check on the similarity between turbulence increases in different climate models, Figure 4.1 shows a scatterplot of the percentage change in the prevalence of MoG CAT in the North Atlantic in winter, as calculated from the CMIP3 GFDL-CM2.1 model used by Williams & Joshi (2013) and the CMIP5 HadGEM2-ES model used in the present study. To ensure a fair comparison, the six-hourly wind and temperature fields from HadGEM2-ES are first time-averaged to match the daily mean fields from GFDL-CM2.1, before calculating the CAT diagnostics. There is a clear correlation between changes in MoG CAT in the two models, as indicated by the correlation coefficient of 0.72. Most of the indices (17 out of 20) appear in the upper-right quadrant of the Cartesian plane, indicating increases in both models, although there is some scatter around the line of best fit.

Figure 4.1 shows that the CAT increases in HadGEM2-ES are on average 30% smaller than in GFDL-CM2.1, possibly because of the different anthropogenic forcing used in the climate change simulations. Specifically, the GFDL-CM2.1 climate change simulation was allowed to equilibrate after the CO_2 loading had been instantaneously doubled. In contrast, the HadGEM2-ES climate change simulation was a transient RCP8.5 run in which the radiative forcing was gradually increased, and so the atmospheric circulation is not expected to be in equilibrium with the contemporary radiative forcing. Therefore, the comparison is not strictly like-for-like. Nevertheless, the comparison shows for the first time that the projected increase in trans-Atlantic turbulence is robust, because it occurs across multiple climate models and to first order it does not depend on the parameterized physics, model resolution, or greenhouse-gas scenario. When the comparison is repeated after first down-scaling the HadGEM2-ES wind and temperature fields to GFDL-CM2.1 resolution, before re-computing the MoG thresholds and re-calculating the turbulence increases, the scatterplot is essentially unchanged.

We also showed in Chapter 3 that the climate model we used in this study (HadGEM2-ES) is a good representation of the location of turbulence. We also showed that in the historical period, HadGEM2-ES underestimated the change in turbulence with the changing climate and therefore this model might underestimate the projected change in turbulence with climate change.

4.2.2 Global Response to Climate Change

Global geographic maps of the percentage change in the prevalence of moderate turbulence in the HadGEM2-ES simulations at 200 hPa in December, January, and February (DJF) are shown in Figure 4.2 for each of the 20 CAT indices. The percentage change refers to the period 2050–2080 compared to pre-industrial times. The indices are ranked in descending order according to the global-mean percentage change. (All geographic averages in this study include the cosine (latitude) scaling factor, to down-weight the smaller high-latitude grid boxes compared to the larger low-latitude ones.) Previous findings about CAT increasing in the North Atlantic evidently apply to other parts of the planet too. In the tropical regions (30°S–30°N), the percentage changes are generally smaller and there is less agreement between the diagnostics. Outside the tropics, in the middle- and high-latitude regions, the percentage changes are generally larger and there is more agreement between the diagnostics.

To assess which features are robust amongst the different diagnostics, the 20 estimates of the percentage changes in CAT shown in Figure 4.2 for DJF are averaged and shown in the upper panel of Figure 4.3. The remaining three panels in Figure 4.3 show the corresponding averages for March, April, and May (MAM), June, July, and August (JJA), and September, October, and November (SON). The averages being taken here are equally weighted, under the assumption that each of the 20 estimates is equally plausible. The percentage changes generally display relatively little seasonality, with the bulk spatial patterns occurring in all four seasons, although there does appear to be a moderate seasonal amplitude modulation locally in some regions. These bulk changes include large increases of several hundred per cent in the midlatitudes in both hemispheres. In the Southern Hemisphere, these increases peak at around 45–75°S and are fairly zonally symmetric. In the Northern Hemisphere, the increases peak at around 45–75°N but they display more zonal variability, which appears to be associated with the presence of land masses. The bulk features also include small and statistically insignificant decreases

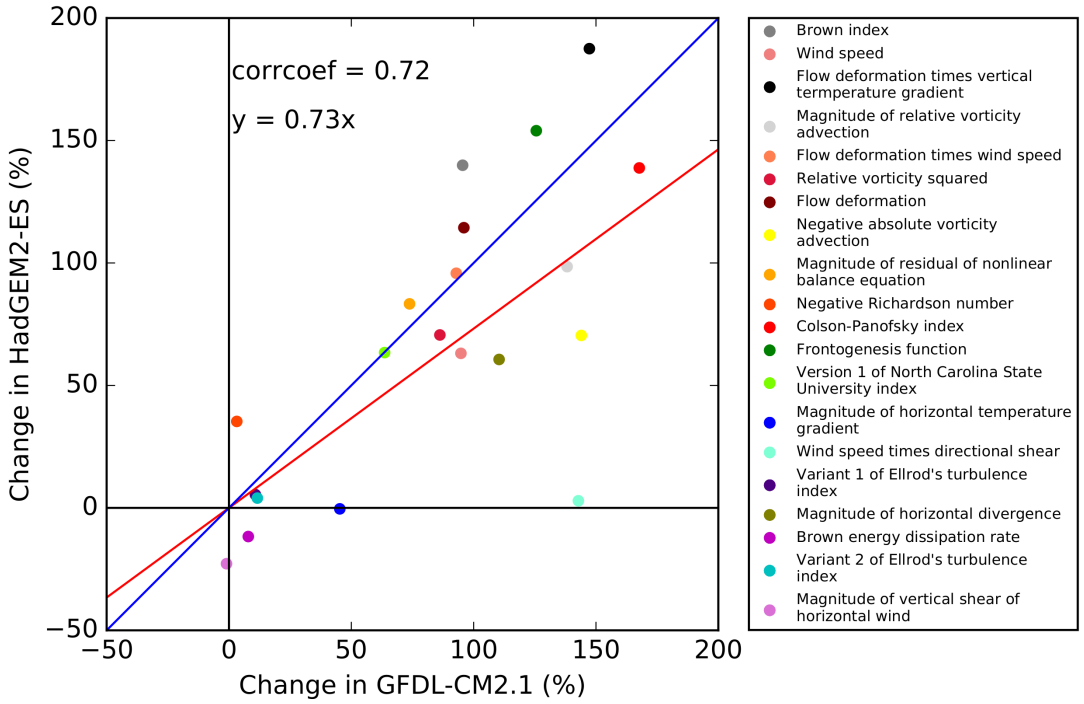

FIGURE 4.1: Scatterplot comparing the HadGEM2-ES simulations from the present study with the GFDL-CM2.1 simulations from Williams & Joshi (2013). The plot shows the percentage change in the prevalence of MoG turbulence for 20 CAT diagnostics calculated at 200 hPa over the north Atlantic (50–75°N and 10–60°W) in winter (December, January, and February; DJF). In this figure only, to ensure the fairest possible comparison, the HadGEM2-ES MoG thresholds are calculated from the control run in exactly the same way that the GFDL-CM2.1 thresholds were calculated by Williams & Joshi (2013), i.e. using the 99th percentiles of CAT diagnosed from daily mean fields in the above geographic box and on the above pressure level in winter. The blue line ($y = x$) indicates parity and the red line ($y = 0.73x$) is a least-squares fit constrained to pass through the origin.

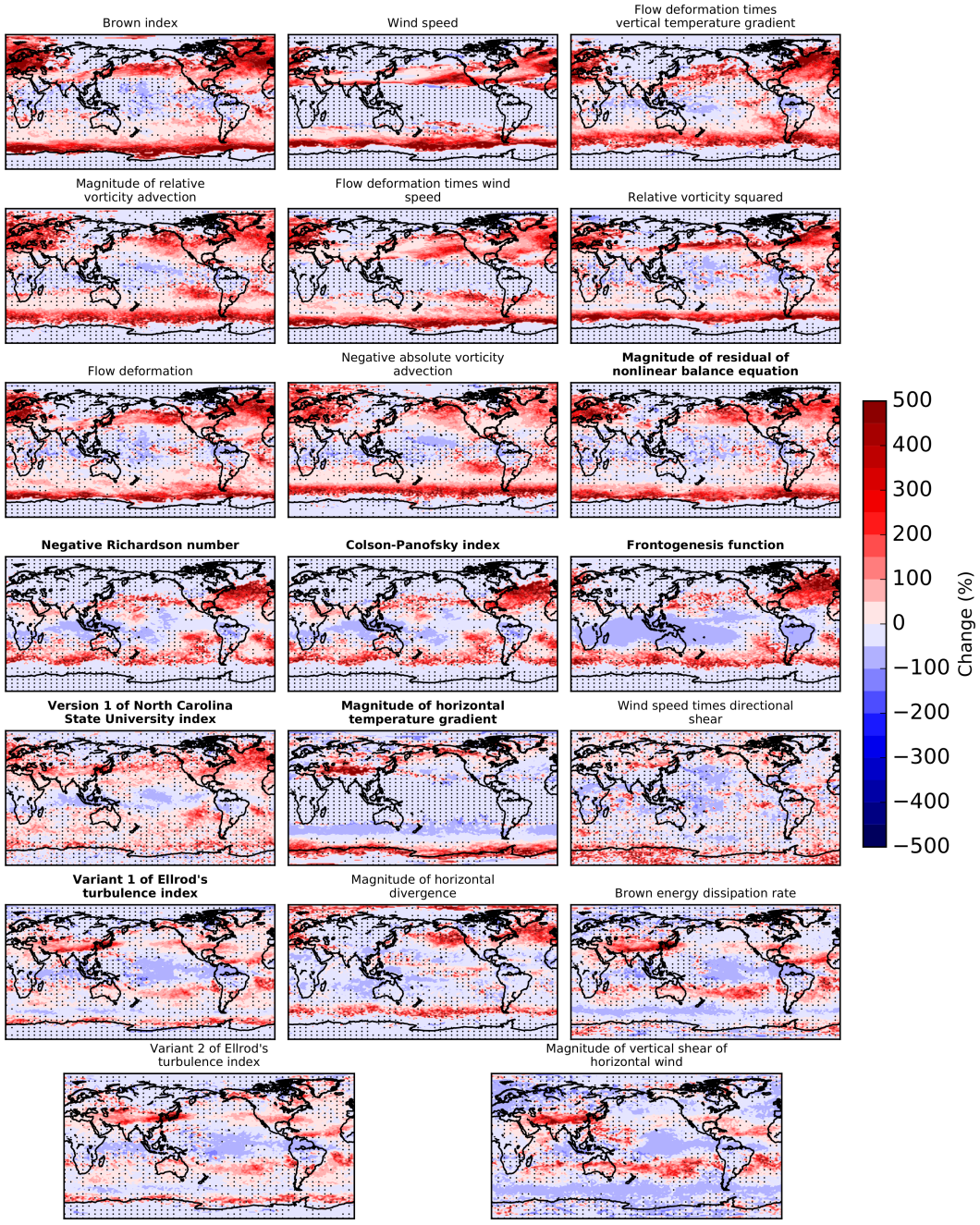

FIGURE 4.2: Maps of the percentage change in the amount of moderate CAT from pre-industrial times (picontrol) to the period 2050–2080 (RCP8.5). The maps are calculated for all 20 CAT diagnostics at 200 hPa in December, January, and February (DJF) using the HadGEM2-ES climate model. The maps are ordered (from left to right and top to bottom) from the largest to smallest global-mean percentage change. Bold titles indicate the seven GTG2 upper-level diagnostics that are used operationally (Sharman et al., 2006). Stippling indicates regions where the percentage change is not statistically significant at the 90% level according to the two-tailed binomial test.

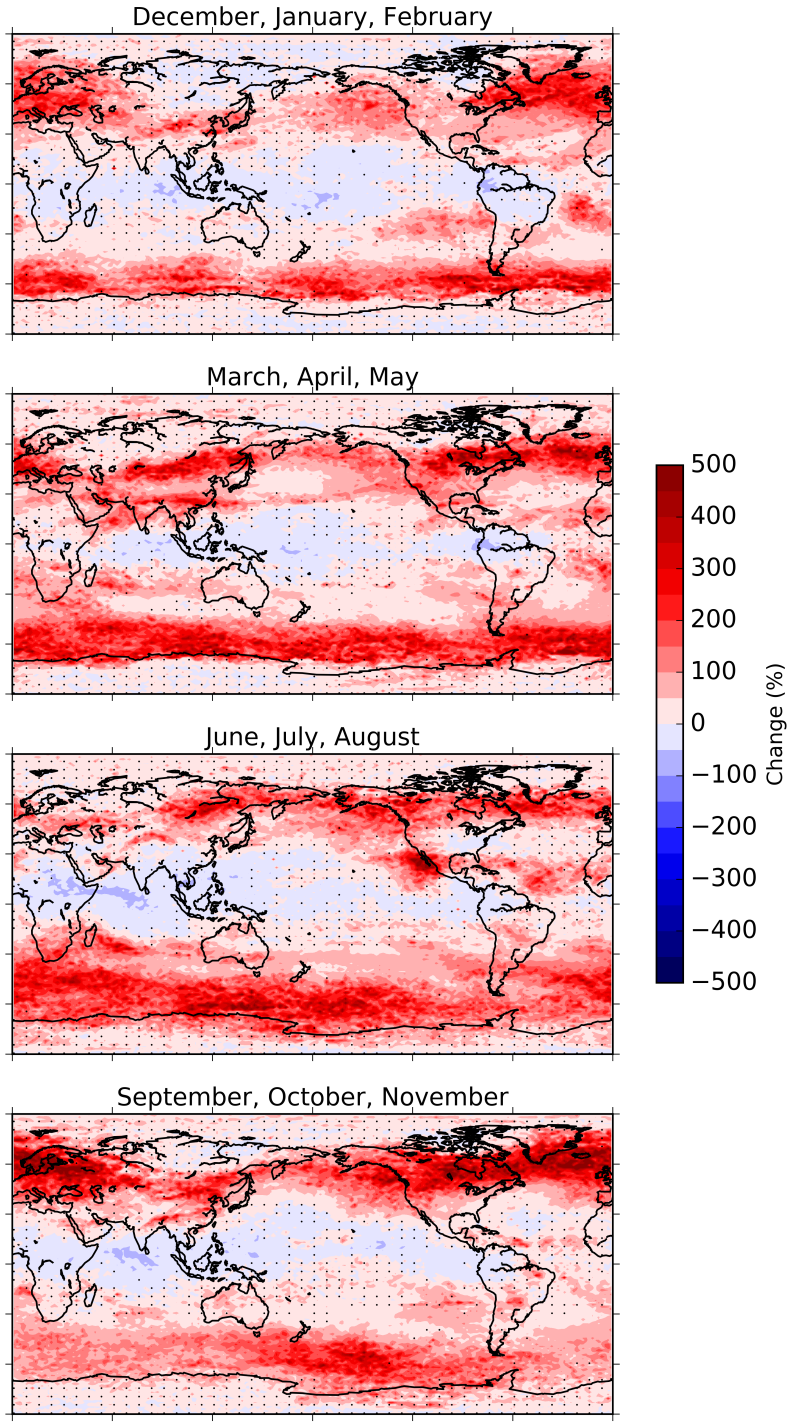

FIGURE 4.3: Maps of the average percentage change in the amount of moderate CAT from pre-industrial times (picontrol) to the period 2050–2080 (RCP8.5) at 200 hPa in each season. The average is taken over all 20 CAT diagnostics, which are equally weighted. The upper panel for December, January, and February is the average of the 20 panels in Figure 4.2. Stippling indicates regions where the average percentage change is not significantly different from zero at the 90% level according to the one-sample, two-tailed t-test.

Strength Category	DJF		MAM		JJA		SON	
	200 hPa	250 hPa	200 hPa	250 hPa	200 hPa	250 hPa	200 hPa	250 hPa
Light	+40.8	+23.5	+54.1	+32.5	+53.2	+32.4	+47.5	+27.0
Light-to-moderate	+36.8	+23.3	+54.8	+32.2	+52.7	+31.6	+47.4	+25.2
Moderate	+31.7	+20.5	+47.7	+30.8	+43.9	+29.3	+40.9	+24.5
Moderate-to-severe	+28.9	+18.2	+43.5	+28.9	+36.7	+26.4	+35.4	+22.5
Severe	+36.0	+21.3	+53.0	+35.1	+44.8	+31.1	+43.9	+26.9

TABLE 4.1: Global-mean percentage changes in the amount of CAT from pre-industrial times (picontrol) to the period 2050–2080 (RCP8.5). The changes are calculated for five turbulence strength categories, at two pressure altitudes, and in four seasons. The changes are averaged over 20 CAT diagnostics. DJF is December, January, and February; MAM is March, April, and May; JJA is June, July, and August; and SON is September, October, and November.

of several tens of per cent in parts of the tropics (where convection is a more important source of turbulence and CAT is less relevant). The global-mean percentage changes in moderate CAT at 200 hPa are dominated by the tropics and are +31.7% (DJF), +47.7% (MAM), +43.9% (JJA), and +40.9% (SON).

The global-mean percentage changes for all five turbulence strength categories (light, light-to-moderate, moderate, moderate-to-severe, and severe) and both pressure levels (200 hPa and 250 hPa) in all four seasons (DJF, MAM, JJA, and SON) are tabulated in Table 4.1. In all 40 cases the change is positive, indicating that CAT is intensifying across a range of strengths and altitudes and that it is intensifying throughout the year. The global-mean percentage changes are generally larger at 200 hPa than 250 hPa, largest for turbulence in the light strength category and largest in MAM.

4.2.3 Regional Response to Climate Change

The global averages discussed in Section 4.2.2 mask large regional variations, Table 4.2 tabulates the annual-mean percentage changes averaged within eight geographic regions, for all five turbulence strength categories and both pressure levels. The results indicate that the busiest international airspace around the middle and high latitudes (North Atlantic, North America, North Pacific, Europe and Asia) experience larger increases in CAT than the global average, with the volume of severe CAT approximately doubling at 200 hPa over North America (+115.5%), the North Pacific (+99.5%) and Europe (+166.7%). The less congested skies around the tropics (Africa, South America and Australia) generally experience smaller increases. Whereas globally it is light turbulence

that experiences the largest relative increase, locally it can be severe turbulence (e.g. Europe). For each strength category and geographic region, the percentage change is larger at 200 hPa than 250 hPa. To provide some context to aid with the interpretation of the magnitudes of these changes, in the North Atlantic (50–75°N, 10–60°W) at 200 hPa we find that (i) in winter, severe CAT by 2050–2080 will be as common as moderate CAT in the control period and (ii) for a range of turbulence strengths from light to moderate-to-severe, summertime CAT by 2050–2080 will be as common as wintertime CAT in the control period.

Strength Category	North Atlantic		North America		North Pacific		Europe	
	200 hPa	250 hPa	200 hPa	250 hPa	200 hPa	250 hPa	200 hPa	250 hPa
Light	+75.0	+47.8	+109.5	+71.0	+122.4	+82.8	+91.4	+60.3
Light-to-moderate	+124.1	+81.1	+115.4	+57.9	+111.1	+55.3	+132.6	+76.4
Moderate	+144.7	+75.2	+103.0	+51.0	+97.5	+43.8	+130.3	+61.9
Moderate-to-severe	+151.9	+72.2	+96.4	+48.2	+80.0	+38.7	+146.8	+67.8
Severe	+185.8	+90.1	+115.5	+60.6	+99.5	+44.2	+166.7	+94.0
Strength Category	South America		Africa		Asia		Australia	
	200 hPa	250 hPa	200 hPa	250 hPa	200 hPa	250 hPa	200 hPa	250 hPa
Light	+17.6	+13.2	+23.4	+20.0	+102.4	+64.9	+18.3	+10.1
Light-to-moderate	+25.4	+17.4	+25.8	+24.1	+93.5	+49.2	+23.9	+14.0
Moderate	+34.2	+22.8	+33.7	+26.6	+80.5	+49.1	+30.9	+19.8
Moderate-to-severe	+43.7	+24.4	+36.8	+27.7	+61.2	+49.5	+38.3	+26.2
Severe	+62.3	+32.2	+51.0	+40.6	+66.3	+57.0	+54.9	+37.6

TABLE 4.2: Annual-mean percentage changes in the amount of CAT from pre-industrial times (picontrol) to the period 2050–2080 (RCP8.5). The changes are calculated for five turbulence strength categories, at two pressure altitudes, and within eight geographic regions. The changes are averaged over 20 CAT diagnostics. The geographic regions are: North Atlantic (50–75°N, 10–60°W), North America (25–75°N, 123–63°W), North Pacific (50–75°N, 145°E–123°W), Europe (35–75°N, 10°W–30°E), South America (55°S–10°N, 80–35°W), Africa (35°S–35°N, 15°W–50°E), Asia (10–75°N, 45–140°E), and Australia (46–12°S, 113–177°E).

4.2.4 Summary and Conclusions

Using climate model simulations, this study has found large relative increases in the atmospheric volume containing significant CAT by the period 2050–2080 under the RCP8.5 greenhouse-gas forcing scenario. The increases occur throughout the global atmosphere but are most pronounced in the midlatitudes in both hemispheres. The increases occur in multiple aviation-relevant turbulence strength categories, at multiple flight levels and in all seasons. We

conclude that the intensification of CAT that has been calculated by previous studies, which considered only MoG CAT on trans-Atlantic flights in winter at altitudes of around 39,000 feet, apply more generally. We also conclude that the changes already experienced in the last 38 years, which are discussed in Chapter 3, will continue into the future and our results might be underestimated if the response of CAT in the HadGEM2-ES simulations is slower than the real atmosphere.

Our findings may have implications for aviation operations in the coming decades. Many of the aircraft that will be flying in the second half of the present century are currently in the design phase. It would therefore seem sensible for the airframe manufacturers to prepare for a more turbulent atmosphere, even at this early stage. Future aeronautical advances, such as remote sensing of clear-air turbulence using onboard LIght Detection And Ranging (LIDAR) technology, might be able to mitigate the operational effects of the worsening atmospheric turbulence (Vrancken et al., 2016). Our results also reinforce the increasingly urgent need to improve the skill of operational CAT forecasts. Despite containing useful information and demonstrably improving the safety and comfort of air travel, these forecasts continue to include a substantial fraction of false positives and missed events.

Future research should extend our results by quantifying the remaining uncertainties. Although the present study has captured uncertainties arising from gaps in our knowledge of turbulence generation, by computing 20 different CAT diagnostics, two key sources of uncertainty remain unquantified. First, future emissions of greenhouse gases depend on socioeconomic and political factors. The corresponding uncertainty in CAT should be quantified by using other forcing scenarios in addition to the RCP8.5 scenario used herein. Second, the jet streams in the upper troposphere and lower stratosphere in different climate models may respond differently to a given radiative forcing anomaly. The corresponding uncertainty in CAT should be quantified by using other climate models in addition to the CMIP5 model used herein, such as the next generation of CMIP6 models that will have substantially higher spatial resolutions.

Future studies could also use a more recent historical period as the baseline, instead of the pre-industrial control period similar to Section 3. Turbulence reports from commercial aircraft could be used for climate model verification purposes. The grid resolution of numerical weather prediction models could

be systematically degraded to match climate models to assess how the turbulence diagnostics depend on resolution. Whereas the present study has investigated the climate response of clear-air turbulence, which is prevalent in the mid-latitudes, future studies should investigate the climate response of convective turbulence which is more prevalent in the tropics. Finally, the response of clear-air turbulence to natural climate variability, such as the North Atlantic Oscillation (Kim et al., 2016), also deserves further study.

Chapter 5

Multi-Model Ensemble

This chapter is based on our paper 'Multi-Model Ensemble Predictions of Aviation Turbulence' which is accepted for publication in *Meteorological Applications*. As discussed in Chapters 3 and 4, climate change is going to increase the occurrence of Clear-Air Turbulence (CAT) in the future. Also the percentage changes shown in Chapter 4 could be underestimates as the climate model is more sluggish in its climate change response than reanalysis data (Chapter 3). Therefore improving turbulence forecasting is vital to help maintain high aviation safety standards. This chapter will introduce a multi-model ensemble for atmospheric turbulence, creating a probabilistic forecast.

Having a probabilistic forecast can help the pilots, flight planners and Air Traffic Controllers (ATC) manage their responses accordingly. An example of this would be if one out of ten ensemble members predict turbulence (i.e. there is a 10% probability of turbulence), in which case pilots might continue their route because the chances are still small. However if all 10 members predict that the threshold will be exceeded, then a pilot may divert around that region (expensive), change flight level (less expensive) or put the seat belt sign on (free) to avoid injury to passengers and crew. Choosing the appropriate action could reduce injuries and save costs. For example, if the turbulence is expected to be light or moderate or there is a low probability of the turbulence, then putting the seat belt sign on is a cost-free response but can impact passenger comfort. If the turbulence predicted is severe or there is a high probability, then the pilot might choose to change the flight level, which might cost money in terms of fuel usage but this cost would be less than a full diversion. If the turbulence predicted is on multiple flight levels, then a full diversion might be appropriate, which would be more expensive by increasing flight time and fuel usage, but this would be less expensive than damaging the aircraft or injuring passengers and crew.

The more ensemble members that are used, the larger the ensemble spread

and the understanding of the certainty of the forecast is improved. This approach therefore provides more information to pilots and flight planners about where turbulence is likely to be and therefore which regions they should avoid. However, increasing the forecast spread could capture more turbulence events, but could also increase the number of false alarms, and this trade-off is one that needs to be managed to maximise forecast skill. This study further expands on the use of an ensemble forecast, and follows other areas of meteorology such as The International Grand Global Ensemble (TIGGE) project (Swinbank et al., 2016) which looks at combining ensembles from different centres around the world. A particular research area using multi-model ensembles is tropical cyclone forecasting (Krishnamurti et al., 2000; Vitart, 2006; Titley & Stretton, 2016). All of these studies show that using multi-model ensembles improves the overall skill of the forecasts and therefore show a useful application that we will investigate for turbulence in this study.

By using at least two ensembles, not only is the spread broadened by increasing the number of forecasts but also a different numerical model is used, assimilating a differing set of observations that will have different strengths and weaknesses. These strengths and weaknesses can come from how the ensembles are perturbed. An example is the European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System (EPS), which starts each model run with the same initial conditions but adds dynamically defined perturbations to create the model spread (Molteni et al., 1996). In contrast, the Met Office Global and Regional Ensemble Prediction System (MOGREPS-G) uses different initial conditions and model perturbations to provide the ensemble spread (Bowler et al., 2008). The initial conditions are perturbed using the ensemble transform Kalman filter (Bishop et al., 2001) and the model perturbations are driven by two stochastic physics schemes; the random parameter scheme and the stochastic convective vorticity scheme. Bowler et al. (2008) showed as an example that the screen temperature Brier Skill Score (BSS) was higher for ECMWF-EPS compared to MOGREPS-G, but that for wind speed the MOGREPS-G ensemble was more skilful than the ECMWF-EPS. This shows each ensemble has its own strengths and weaknesses that we hope will increase the forecast skill. This study is the first time multi-model ensemble forecasting has been applied to turbulence. Both WAFCs plan to use a multi-model ensemble in the near future and therefore this study lays the foundations to make this possible.

5.1 Observations

In order to verify the forecasts we need to find a 'truth' data set and is the same method described in Section 2.3.6. Previous work used PIlot REPortS (PIREPS) (Tebaldi et al., 2002; Kim & Chun, 2011), but these can be unreliable (Schwartz, 1996; Kane et al., 1998; Sharman et al., 2014). PIREPS are subjective and are also aircraft dependent, so a smaller aircraft will experience more severe turbulence than a larger aircraft in the same volume of turbulent air. PIREPS also have poor spatial reliability as they tend to be located in turbulence, so null turbulence events are rarely recorded as there is no specified frequency (Kane et al., 1998). The location and time of PIREPS may also not be correct as they are manually reported after the event and for more severe events where action is required this may be some time later. To avoid these problems, this study will use aircraft data recorded on a fleet of Boeing 747 and 777 aircraft. This data has been used in other meteorological studies (Tenenbaum, 1991; Gill, 2014). High-resolution automated aircraft data, available at 4 second intervals, giving over 76, 000, 000 data points to calculate an aircraft-independent turbulence measure known as the Derived Equivalent Vertical Gust (DEVG) which is defined as:

$$\text{DEVG} = \frac{Am|\Delta n|}{V}, \tag{5.1}$$

where $|\Delta n|$ is the peak modulus value of the deviation of the aircraft acceleration from $1g$ in units of g, m is the total mass of the aircraft (metric tonnes), V is the calibrated airspeed at the time of the observation (knots), and A is an aircraft-specific parameter that varies with flight conditions and can be calculated using:

$$A = \overline{A} + c_4(\overline{A} - c_5)\left(\frac{m}{\overline{m}} - 1\right) \tag{5.2}$$

and:

$$\overline{A} = c_1 + \frac{c_2}{c_3 + H(kft)}, \tag{5.3}$$

where H is the altitude (thousands of feet), $\overline{m}$ is the reference mass of the aircraft (metric tonnes), and parameters c_1 to c_5 depend on the aircraft's flight profile as outlined by Truscott (2000).

DEVG is one of the World Meteorological Organization (WMO) recommended turbulence indicators and has a typical uncertainty of around 3-4% (WMO, 2003). DEVG is aircraft independent so values from all aircraft can be combined to create an observational database. Table 5.1 compares DEVG to Eddy Dissipation Rate (EDR) which is another aircraft independent measure

Turbulence Severity	DEVG (m s^{-1})	EDR ($m^{2/3}$ s^{-1})
None	DEVG ≤ 2	EDR ≤ 0.07
Light	$2 \leq$ DEVG ≤ 4.5	$0.07 \leq$ EDR ≤ 0.27
Moderate	$4.5 \leq$ DEVG ≤ 9	$0.27 \leq$ EDR ≤ 0.61
Severe	$9 \leq$ DEVG	$0.61 \leq$ EDR

TABLE 5.1: Turbulence severity for values of Derived Equivalent Vertical Gust (DEVG) (Truscott, 2000) and Eddy Dissipation Rate (EDR) (Sharman et al., 2014). For severe turbulence to be observed the DEVG value must be greater than or equal to 9 m s^{-1} and therefore $9 \leq$ DEVG.

that can both be used in turbulence verification as discussed in Section 2.3.6. There are limitations to using this data set, because aircraft manoeuvers and active control techniques can enhance or dampen vertical accelerations of aircraft leading to over- or under-representation of the vertical gusts (WMO, 2003). One of the other main issues with this data set is the typical spatial coverage. Figure 5.1 is a plot of aircraft data for May 2016 and shows the spatial coverage of our observations. It has very good coverage over the North Atlantic and Europe, but poorer coverage over Asia and the Pacific. Despite the uneven spatial coverage, this data set is still the best available source of truth data for verification which is why we have chosen to use it here. The Ellrod & Knapp (1992) Turbulence index 1 (Ellrod TI1) turbulence predictor used in this study only forecasts shear induced turbulence and is not able to predict convective turbulence. This study will therefore use a satellite-based convective product to filter out convective turbulence events (Francis & Batstone, 2013). By only looking at the non-convective events we should have a better representation of the forecast skill by removing events that we know will be missed by the turbulence diagnostic.

5.2 Forecast Data

This project uses an entire year of global ensemble data between May 2016 and April 2017 from two forecast centres: MOGREPS-G (Bowler et al., 2008) and the ECMWF EPS (Molteni et al., 1996). The forecast data is available with 3-hourly intervals and at the time of this study the MOGREPS-G ensemble had 12 members with forecasts every 6 hours, with 33 km resolution and 70 vertical levels, 10 of which are between 150 and 350 hPa. The ECMWF forecast had 51

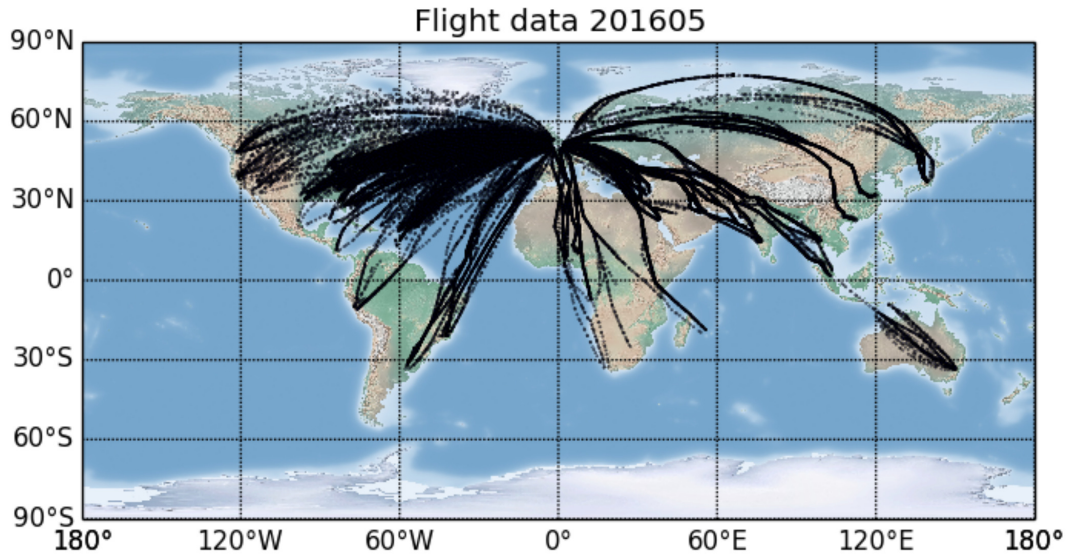

FIGURE 5.1: Plot of the spatial coverage of flight data from the fleet of Boeing 747 and 777 aircraft in May 2016.

ensemble members with 18 km resolution and 91 vertical levels, 14 of which are between 150 and 350 hPa. To use the ECMWF EPS system operationally we would have to extend the forecast range by approximately 12 hours due to a delay in accessing the forecast data. This is important to note because it means the results of the ECMWF EPS will be theoretical and in practice the skill would be reduced as we have to use longer lead times. This is shown by the WAFC verification website demonstrating how the forecast skill is reduced with forecast lead time (Met Office, 2018). We have not used the longer lead times in this study but understanding the impact this will have on the forecast skill would be an interesting area of further study. For both ensembles, this study only had access to the 0000 UTC model run between 01 May 2016 and 07 August 2016, which means for that period we forecast only half of the day. Between 07 August 2016 and 30 April 2017 we had both the 0000 UTC and 1200 UTC model run. The forecast lead time used throughout is T+24, T+27, T+30, T+33 hours.

This chapter focuses on non-convective turbulence, and uses the Ellrod TI1 predictor which is defined as:

$$\text{TI1} = \text{DEF} \times \text{VWS} = \left[\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)^2\right]^{1/2} \times \left[\left(\frac{\partial u}{\partial z}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2\right]^{1/2} \quad (5.4)$$

where u is the horizontal wind velocity in the East-West direction, v is the horizontal wind velocity in the North-South direction, x is distance in the East-West direction, y is distance in the North-South direction and z is distance in the vertical. This is the same index used in previous studies including Chapters 3 and 4 and is the turbulence diagnostic currently used by the WAFCs. The Ellrod TI1 combines deformation and vertical wind-shear as shown in Equation 5.4 and is a well-established shear turbulence diagnostic. Previous research has shown it predicts up to 75% of CAT events (Ellrod & Knapp, 1992), although it is typically only useful in the mid-to-high latitudes. Also the Ellrod TI1 was not developed to predict Convectively Induced Turbulence (CIT) or Mountain Wave Turbulence (MWT), which are both prominent turbulence sources.

To create a probability forecast, we use the ensemble data from both WAFCs and set thresholds based on literature. Exact numbers can't be used however as the data sets are not identical (e.g. different resolutions) and some are calibrated (e.g. ICAO, 2012) giving different turbulence thresholds. The turbulence threshold used in this project for the Ellrod TI1 is 3×10^{-7} s^{-2} which is equivalent to Moderate-or-Greater (MoG) turbulence (Gill, 2014). However we also calculate additional thresholds that range from light-or-greater to severe-or-greater turbulence. Using multiple thresholds to predict different turbulence strength categories is similar to the approach used by Williams (2017) and Chapters 3 and 4. Ellrod & Knapp (1992) also discuss the use of higher thresholds for moderate and severe turbulence with the actual values used being model specific. We use these additional thresholds to optimise the turbulence forecast as a lower threshold will capture more MoG turbulence events, but also more false alarms, and a higher threshold will forecast fewer MoG events but also fewer false alarms. The additional thresholds we used are 8×10^{-8} s^{-2}, 1×10^{-7} s^{-2}, 5×10^{-7} s^{-2}, 8×10^{-7} s^{-2}, 1.1×10^{-6} s^{-2}, and 2×10^{-6} s^{-2}. Above these thresholds, it is classed as an area of the atmosphere containing turbulence. The different thresholds could be used for different corresponding turbulence severities. By combining the forecasts for all the ensemble members we can calculate the probability of a grid point containing MoG turbulence. The more ensemble members that predict the occurrence of turbulence, the higher the probability forecast will be. It is then possible to combine both the ensemble forecasts to create a multi-model ensemble in two ways. The first is a standard equally weighted multi-model super ensemble, and the other is a weighted multi-model ensemble. This study uses the

simple equally weighted multi-model super ensemble. To create the multi-model ensemble we first create a probability field of turbulence for both the single-model ensembles based on exceeding a threshold of 3×10^{-7} s^{-2}. Then we average the two probability fields together, creating an equally weighter multi-model ensemble. This therefore means that although the ECMWF-EPS ensemble has more ensemble members, it does not have any more weight in the multi-model ensemble.

5.3 Verification Method

The verification method used in this study is outlined by Gill (2014, 2016). This study processes aircraft observations into 10-minute segments, which equates to approximately 100 km of flight. By analysing the DEVG values in each segment, when the maximum value exceeds a given threshold it is classed as a turbulence event. The aircraft data is constrained to $\pm$1.5 hours of the forecast time to ensure the aircraft observations reflect the forecast. Therefore only the 10 minute segment (observation) within the 3 hour time window is used for verification as beyond this the observations are not valid for the turbulence forecast. We then compare the turbulence observations to the forecast and a 2$\times$2 contingency table can be set up as shown in Table 5.2. One of the best ways to visualise these results is to use a Relative Operating Characteristic (ROC) plot (Jolliffe & Stephenson, 2012; Gill, 2016), which plots the hit rate against the false alarm rate which are defined as:

$$\text{Hit Rate} = \frac{A}{A+C} \tag{5.5}$$

$$\text{False alarm rate} = \frac{B}{B+D} \tag{5.6}$$

where A is a hit, B is a false alarm, C is a miss and D is a correct rejection. To create the ROC curve, thresholds are applied to the probabilities which then create binary yes/no forecasts with corresponding 2$\times$2 contingency tables yielding the hit rate and false alarm rate which are then plotted together. This produces a curve where the larger the Area Under the Curve (AUC), the more skilful the forecast is at discriminating between events and non-events.

The reliability of the forecasts can be assessed visually by using a reliability diagram (Jolliffe & Stephenson, 2012; Gill, 2016), where each probability is binned and the frequency of the event is calculated. The forecast probability should equal the observed frequency. For example, if the probability is 30%

	Turbulence observed	Turbulence not observed
Turbulence forecast	A (Hit)	B (False Alarm)
Turbulence not forecast	C (Miss)	D (Correct Rejection)

TABLE 5.2: A 2 × 2 contingency table showing the possible results of a turbulence forecast or event. The four possible outcomes include a Hit, Miss, False Alarm and Correct Rejection.

then turbulence in that region should be observed 30% of the time. A perfect forecast would result in a straight line, however in practice this is not the case and forecast probabilities tend to over-forecast the turbulence (below the line), or under-forecast turbulence (above the line). Understanding these biases allow us to implement a linear calibration which should bring the forecast probability more in-line with the observed frequency. Calibrating the forecast will not compromise the forecast skill, since ROC area discriminatory skill and reliability are independent (Gill, 2016).

A more practical analysis of the results for stakeholders would be to assess the relative economic value (V) of the forecast which is defined as:

$$V = \frac{min(\alpha, \bar{o}) - F\alpha(1-\bar{o}) + H\bar{o}(1-\alpha) - \bar{o}}{min(\alpha, \bar{o}) - \bar{o}\alpha} \tag{5.7}$$

where α = Cost/Loss, $\bar{o}$ is the fraction of occasions where the event occurred, F is the false alarm rate, H is the hit rate (Richardson, 2000; Jolliffe & Stephenson, 2012). This assigns a cost and loss for the elements in a contingency table (Table 5.3) where different outcomes depend on whether action was taken and if the event occurred or not. For a given model, the hit rate (H), false alarm rate (F) and fraction of occasions the event occurred $\bar{o}$ can be calculated using a 2×2 contingency table, and therefore varying the cost/ loss ratio gives a different value which can be plotted. The more skilful the model, the higher the maximum value will be (but the actual value will depend on the cost/ loss ratio of the user). If the value is higher for all cost/ loss ratios then that model will be the most useful for any consumer (as the cost/ loss ratio may vary depending on the consumer) and this is known as sufficiency (Ehrendorfer & Murphy, 1988). Gill & Buchanan (2014); Buchanan (2016) showed that probabilistic turbulence forecasts have greater value than deterministic forecasts, so this project will aim to show that by combining ensembles that we can further increase the value.

	Turbulence observed	Turbulence not observed
Turbulence forecast Action taken	Hit Cost	False Alarm Cost
Turbulence not forecast No action taken	Miss Loss	Correct Rejection

TABLE 5.3: A 2 × 2 contingency table assigning a cost to the possible results of a turbulence forecast or event. The four possible outcomes include a Hit (with a subsequent cost), Miss (with a subsequent cost), False Alarm (with a subsequent cost) and Correct Rejection (with no cost as no action was taken).

5.4 Results

5.4.1 Case Studies

Throughout this analysis we focus on shear turbulence, however MWT and CIT will be present in the observational truth data. To identify the source of turbulence, this study plots the aircraft data over a plot of orography, convection and shear turbulence. The orography plot uses a surface map that indicates terrain height and therefore mountain ranges. The orography map shows the height of the terrain and any event that occurs near high terrain could be caused by MWT, whereas the CIT plot uses a satellite product that indicates areas of deep convection (Francis & Batstone, 2013). The satellite product identifies regions of overshooting tops which indicate the regions of the strongest updraft, above the smooth anvil of a typical thunderstorm. To identify these regions they use two methods: the first method is the water vapour-infrared window brightness temperature difference method (Schmetz et al., 1997). The second is the infrared window texture method (Bedka et al., 2010). By using the infrared channel, it can be used in both the day and night which is important for aviation. We did not have full global coverage for the satellite product, and therefore only CIT events within that spatial coverage could be removed. For the shear turbulence we plot both the MOGREPS-G and ECMWF ensemble probability fields showing if shear turbulence is a likely cause and which ensemble products predicted turbulence. The plots have aircraft data ± 1.5 hours which will help identify the likely source of turbulence.

Figure 5.2 is a plot of a shear turbulence case study that was forecast by both the MOGREPS-G and ECMWF-EPS ensembles. The plot clearly shows the MoG turbulence event over the north Atlantic so MWT is not a factor, and the satellite product shows there is no deep convection in the area, although there is some convection much further south. This shows the turbulence event

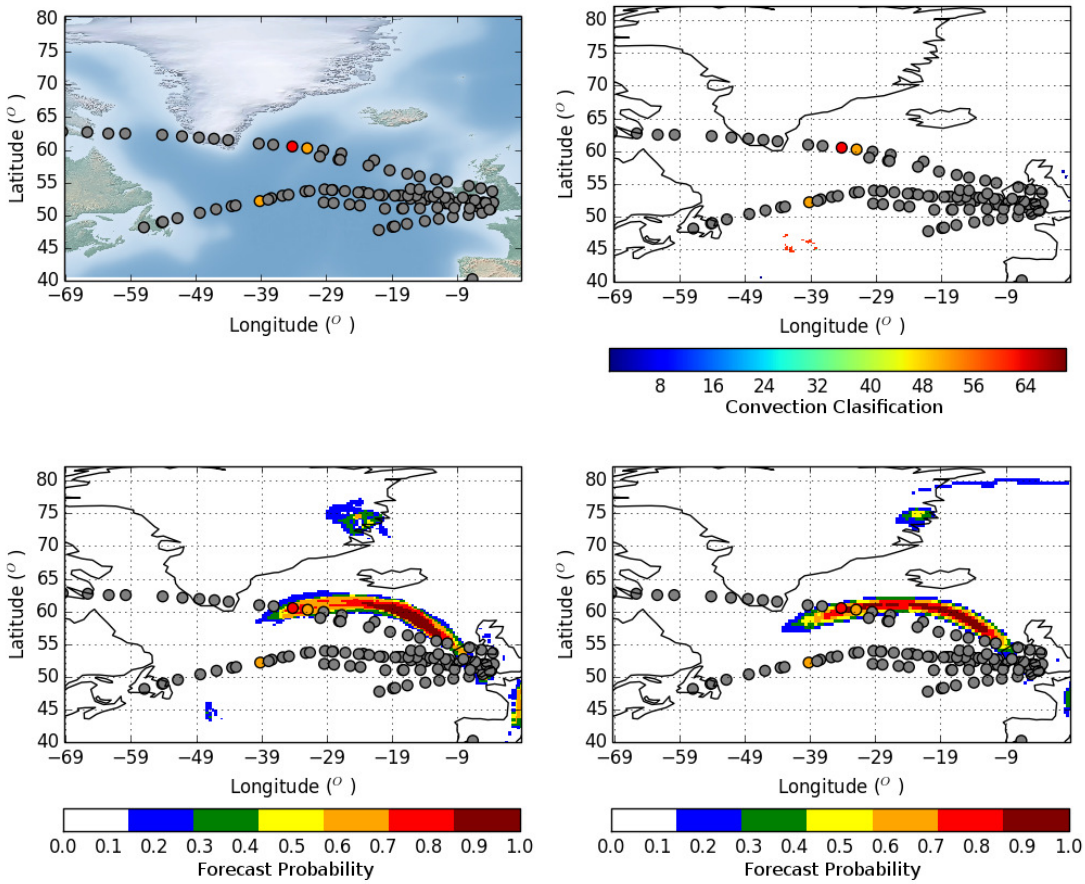

FIGURE 5.2: Plot of a moderate-or-greater turbulence event over the possible sources of turbulence: top left: orography, shear turbulence (bottom left: MOGREPS-G and bottom right: ECMWF EPS probability forecast), and top right: convection from satellite data (colour shading indicates deep convection). Both the MOGREPS-G and ECMWF-EPS ensembles forecast the shear turbulence event. The circles indicate turbulence observations with grey indicating no turbulence, orange indicating light turbulence and red indicating moderate or greater turbulence. The convective classification can be found in Francis & Batstone (2013).

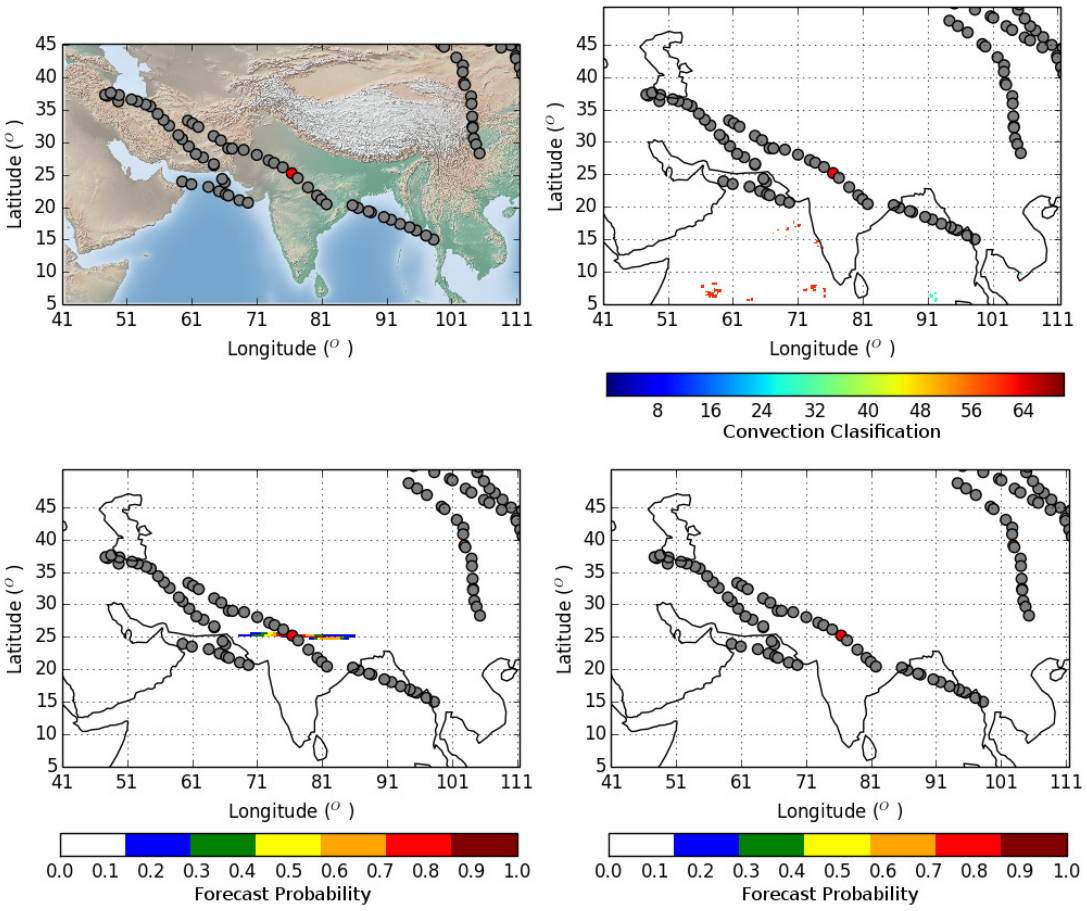

FIGURE 5.3: Plot of a moderate-or-greater turbulence event over the possible sources of turbulence: top left: orography, shear turbulence (bottom left: MOGREPS-G and bottom right: ECMWF EPS probability forecast), and top right: convection from satellite data (colour shading indicates deep convection). Only the MOGREPS-G ensemble forecast the shear turbulence event. The circles indicate turbulence observations with grey indicating no turbulence, orange indicating light turbulence and red indicating moderate or greater turbulence. The convective classification can be found in Francis & Batstone (2013).

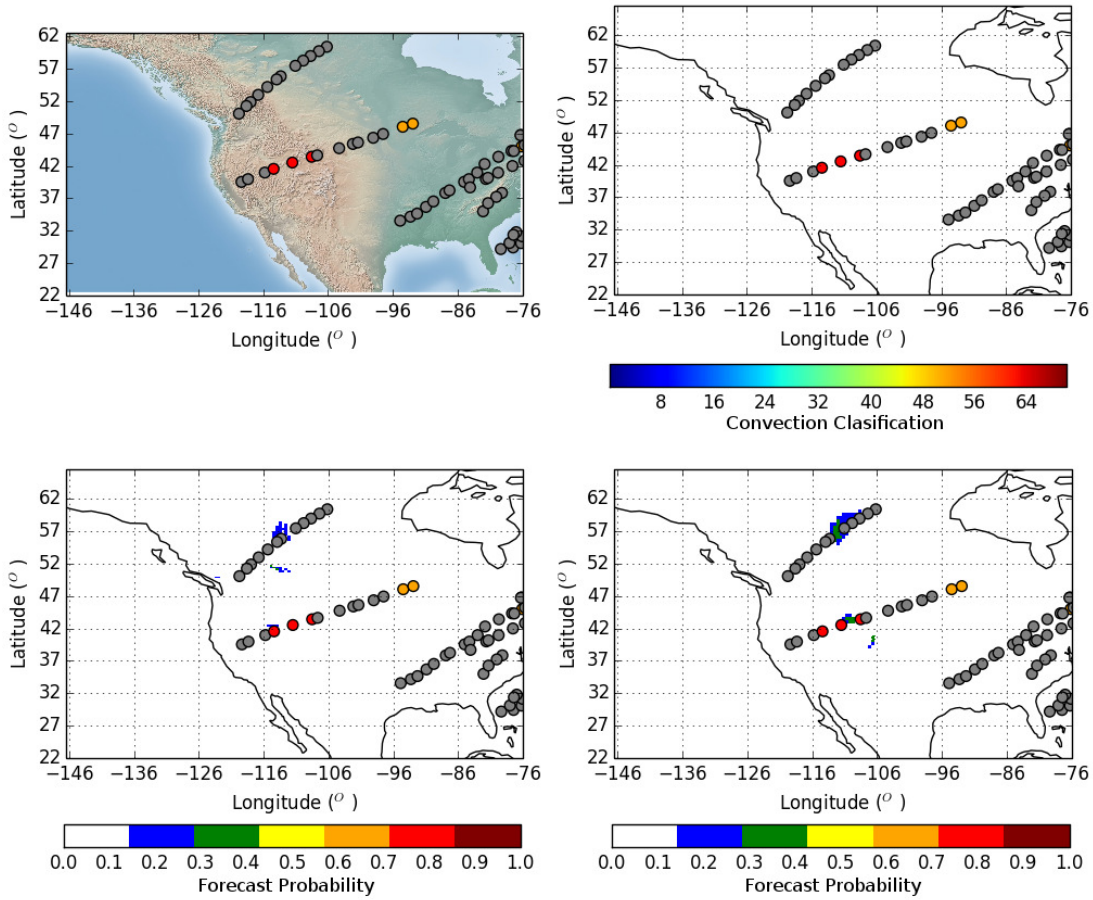

FIGURE 5.4: Plot of a moderate-or-greater turbulence event over the possible sources of turbulence: top left: orography, shear turbulence (bottom left: MOGREPS-G and bottom right: ECMWF EPS probability forecast), and top right: convection from satellite data (colour shading indicates deep convection). Only the ECMWF-EPS ensemble forecasts the shear turbulence event. The circles indicate turbulence observations with grey indicating no turbulence, orange indicating light turbulence and red indicating moderate or greater turbulence. The convective classification can be found in Francis & Batstone (2013).

was well forecasted by both the ensemble products. There is also a light turbulence event further south, which is not forecasted by either ensemble. This is to be expected as the threshold used for this figure is typical for MoG turbulence, and therefore not expected for this event. There is some convection just to the south, however it is too far away to cause this light turbulence event.

Looking at Figure 5.3 there is no CIT in the area, however this could be MWT as this event is over some smaller mountains, or shear turbulence as it is forecast by the MOGREPS-G ensemble. This is a case that shows the benefit of using the multi-model ensemble approach. If we had only the ECMWF-EPS ensemble then we would not have been able to forecast this event and as a result people could be injured. But since we have both the ensembles, we managed to forecast the event and therefore preventative action could have taken place, increasing passenger and crew comfort and safety.

Figure 5.4 shows another example where the multi-model ensemble approach is better, as the MOGREPS-G ensemble does not forecast the turbulence event but the ECMWF-EPS ensemble does. This case study is interesting however because the turbulence event is over the Rocky Mountains, so this could be MWT. It is probably a combination of shear and MWT and again shows the multi-model ensemble approach was a benefit, as if we only had the MOGREPS-G ensemble we would not been able to forecast this event. This case also reinforces the need to have a MWT diagnostic, as the severe turbulence observations spread further than the forecast indicated. Figure 5.4 also shows some of the problems with turbulence forecasting, as we see what could be a false alarm event over Canada. Both the ECMWF and MOGREPS-G ensemble predict turbulence, however there is no turbulence observed. This shows the benefit of using a probabilistic forecast because different end users can select the probability threshold of when they would take action. For example if the probability forecast for turbulence is 20%, then there is a 1 in 5 chance of turbulence being observed. However if the end user sets their threshold for taking preventative action at 10%, then this event would be classed as a false alarm because it was forecasted (as it was above the 10% threshold) and the event did not occur. If the threshold they used for taking preventative action was 30% however, this would be a correct rejection as turbulence would not be forecast as the threshold was not exceeded and turbulence did not occur. This helps to illustrate the trade-off between hits and false alarms and the ability the probability framework gives to its users to fine tune their response to optimise the forecast. So for this example, a higher threshold would result in a correct rejection but might also miss the MoG turbulence event over the Rocky

Mountains.

After plotting all of the 424 MoG turbulence events, we identified 98 cases that are likely to be CIT which Ellrod TI1 cannot forecast. To address this issue the CIT events from the rest of our study are removed in order to give the fairest possible test for the multi-model ensemble forecasts. We decided to keep all other MoG turbulence events in the study because we have no strong evidence that they are not shear related. An example is MWT, although we can identify these events occur over mountains, we are not able to prove wind shear is not a contributing factor. Using the satellite convection product we are more confident of the CIT events and therefore we have more confidence in removing them. If we were unsure in any way that it is not a CIT event, they were kept in the study. What this does highlight though is this study must be extended in the future to include a convective diagnostic and a MWT diagnostic because combined they count for a third of the events in this study.

5.4.2 Multi-Model Ensemble Trial

After removing the events that we have categorised as CIT only, we analysed the performance of the multi-model ensemble and the single-model ensembles. Figure 5.5 is a ROC plot showing the skill score for both the single-model ensembles and the combined multi-model ensemble. Typically the area under the curve is a good measure of discriminatory skill. However, in this study the MOGREPS-G ensemble has a shorter line than the both the ECMWF and multi-model ensemble. This is because a 12 member ensemble can't forecast the same lower probabilities as a larger ensemble. The 12 member ensemble can only predict probabilities as small as 1/12, whereas the ECMWF 51 member ensemble can forecast probabilities as low as 1/51. Therefore a simple AUC number could be biased towards the ECMWF forecast and multi-model ensemble forecasts as the longer line could (and does in this example) give them a larger AUC. Therefore it is better to focus on how steep the line is and therefore on low false alarm rates that are more useful for the aviation industry, although the best method of measuring statistical significance is to use the AUC.

Low false alarm rates are more important for this study because airline companies may have a limit on acceptable hit rates and false alarm rates, and therefore the lower false alarm rates are the ones they would focus on. Figure 5.5 shows that the ECMWF, MOGREPS-G and simple combined ensemble have almost the same skill. This is surprising because by combining the two

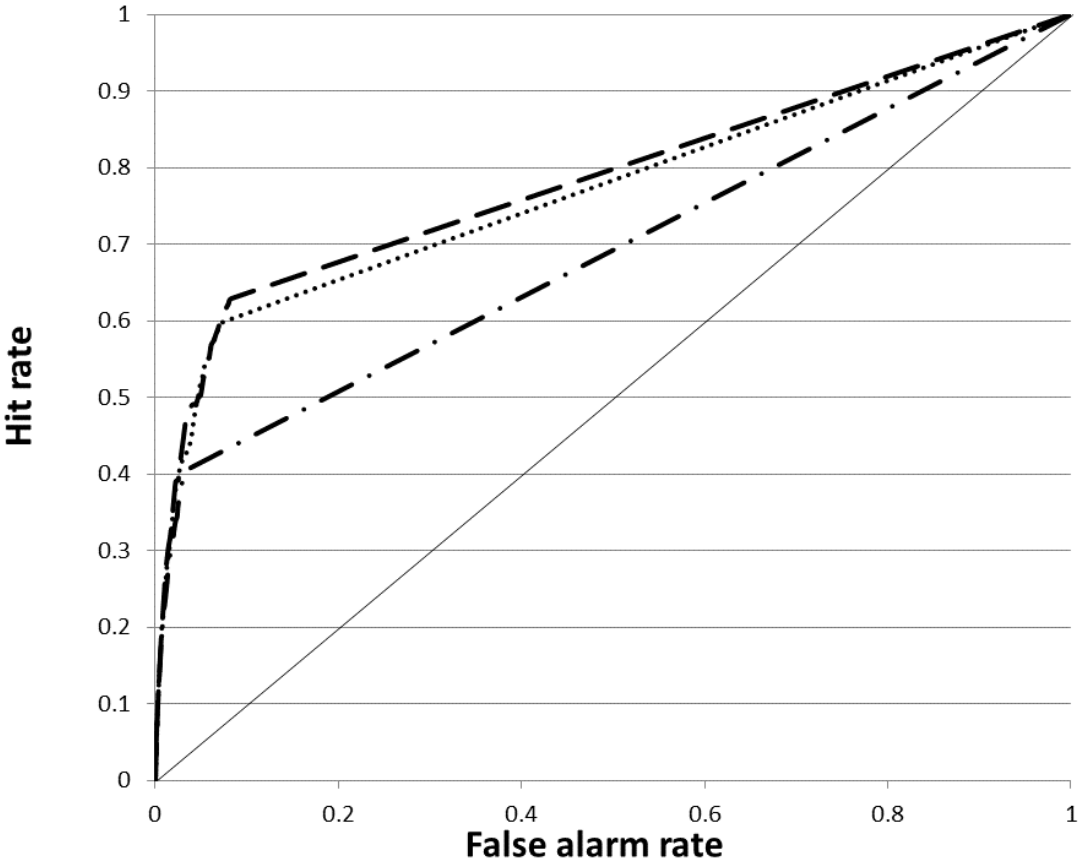

FIGURE 5.5: ROC plot of the global turbulence with the 98 convective turbulence cases removed showing the forecast skill of the MOGREPS-G (dot-dash) AUC=0.6881, ECMWF (dot) AUC=0.772 and combined multi-model ensemble (dash) AUC=0.7842. The data used has a forecast lead time between +24 hours and +33 hours between May 2016 and April 2017.

ensembles, the forecast spread has increased and therefore we capture more turbulence events, but consequently more false alarms. Because of this trade off we do not see a significant increase in skill. The AUC for the two single model ensembles are: ECMWF – 0.7712 and MOGREPS-G – 0.6881. The multi-model ensemble has an AUC of 0.7842 with the 95% confidence interval lower bound being 0.7538 and the 95% confidence interval upper bound being 0.8102. This therefore shows at the 95% confidence interval the multi-model ensemble is only significantly better than the MOGREPS-G ensemble, but not the ECMWF ensemble, and this is because the MOGREPS-G line is shorter. To understand the benefit of using a multi-model ensemble, Table 5.4 shows the number of MoG turbulence events where both models agree, and if they disagree, which model forecasted the turbulence event. Out of the 326 MoG turbulence events, 243 of them the models are in agreement so they either both

forecast the event or both do not forecast the event. That leaves 83 out of 326 MoG turbulence when the models do not agree. What we find however is this number is not split evenly and ECMWF forecasts 73 times when MOGREPS-G does not, and there are only 10 occasions where MOGREPS-G forecasts turbulence and ECMWF does not. This means there are 10 occasions where having both ensembles gives the ability to produce a non-zero forecast which a single ECMWF ensemble cannot. This could be part of the reason why we do not see the large improvement of forecast skill when combining ensembles. Most of the events are already forecast by one of the models, and therefore we only have a limited benefit to adding the second ensemble. This is similar to the conclusion Park et al. (2008) drew, as there is only a limited benefit to a multi-model ensemble when one ensemble is already well tuned. However, having a multi-model ensemble can improve the estimate of confidence in the forecast when both models forecast turbulence as well as capturing more turbulence events than a single model ensemble.

Models in agreement	Models not in agreement	
	ECMWF forecasts turbulence	MOGREPS forecasts turbulence
243	73	10

TABLE 5.4: Categorising moderate or greater turbulence events between cases where both ECMWF and MOGREPS models are in agreement (both do/do not forecast turbulence), and where the models are not in agreement (one model does forecast turbulence and the other does not). When the models are not in agreement, the results are put into a sub category stating which ensemble did forecast the turbulence event.

The relative economic value of the forecast is shown in Figure 5.6 and the multi-model ensemble has greater value than the single model ensembles but only for low cost/loss ratios. This is important because depending on the relative importance of minimising misses and maximising hits for an airline company, defines the cost/loss ratio we focus on. We also know the cost of action is likely to be a great deal less than the cost of loss due to injuries or aircraft damage. Therefore the lower cost/loss ratios are likely to be more important for the airline companies, and therefore this study focuses on those here. So the multi-model ensemble is as skilful as the single model ensembles, but would be more useful for decision-making in an operational environment. This figure also shows the maximum value for all the thresholds

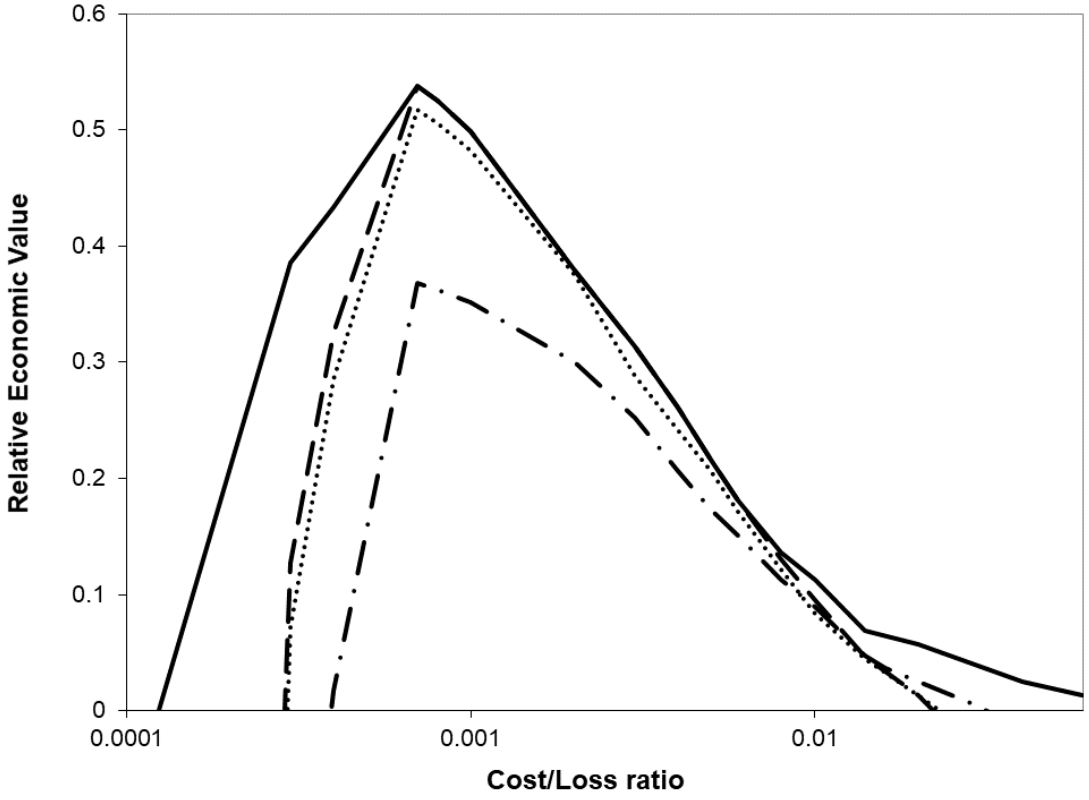

FIGURE 5.6: Value plot with a log scale x-axis of the global turbulence with the 98 convective turbulence cases removed showing the forecast skill of the MOGREPS-G (dot-dash), ECMWF (dot), combined multi-model ensemble (dash) and the maximum value using every threshold of the combined multi-model ensemble (solid). The data used has a forecast lead time between +24 hours and +33 hours between May 2016 and April 2017.

of the combined multi-model ensemble. As identified previously, the probability fields for many thresholds have been calculated, so this curve takes the highest value threshold for each cost/loss ratio. Figure 5.7 shows how this is done as it plots the relative economic value for each of the five thresholds for the MOGREPS-G ensemble, and then plots the maximum value (bold line) at each cost/ loss ratio. This indicates that different users might need a different threshold depending on what cost/loss ratio they operate at. For Figure 5.6 the bold line showing the maximum value for each cost/loss ratio is above the others, and as in Figure 5.7 there is more value in some of the other thresholds, therefore an optimised multi-model ensemble would provide more value and is worth pursuing in future studies. It is important to point out again that the ECMWF EPS value is theoretical and operationally would be lower since the availability of the data forces the use of a longer lead time. When comparing the single model ensembles to the previous study by Gill & Buchanan (2014)

we see the improvement in the relative economic value, showing a significant model improvement over the last few years. This improvement could be because the Met Office introduced the ENDGame (Even Newer Dynamics for General atmospheric modelling of the environment) dynamical core (Walters et al., 2014). This has provided a better forecast and in particular has resulted in a 'reduction of the slow bias in tropospheric windspeeds (Walters et al., 2014).' It is important to note that a direct comparison is not possible because each study looks at two different years, but this study shows a large improvement in MOGREPS-G value compared to the Gill & Buchanan (2014).

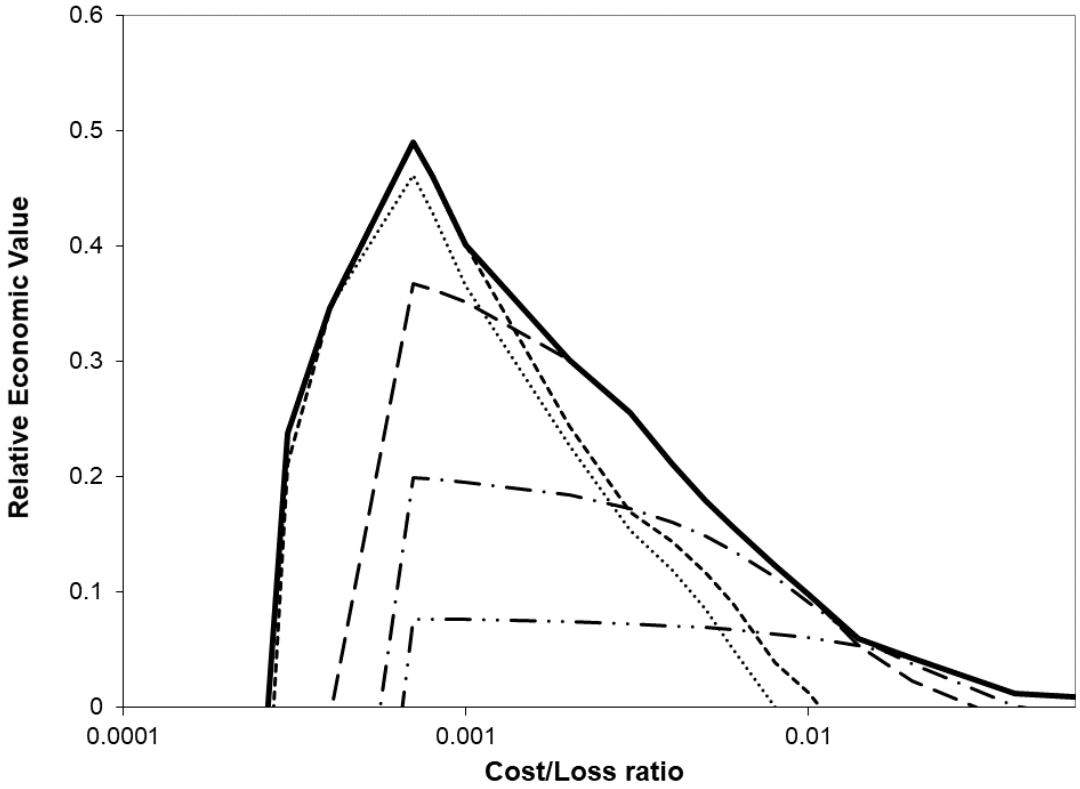

FIGURE 5.7: Value plot with a log scale x-axis showing the MOGREPS-G relative economic value for threshold 1 (dot), threshold 2 (short dash), threshold 3 (long dash), threshold 4 (dot dash), threshold 5 (dot dot dash) and the maximum value at each cost loss ratio (solid). The data used has a forecast lead time between +24 hours and +33 hours between May 2016 and April 2017.

Also plotted is a reliability plot shown in Figure 5.8. This figure shows that the MOGREPS-G, ECMWF-EPS and combined multi-model ensembles under-forecast the lower probabilities, but over-forecast the higher probabilities. This is shown by each ensemble being above the line of a perfect forecast

for the lower probabilities, but below the line for higher forecast probabilities. It is important to note that these plots have been calibrated because the forecast percentage from the ensembles is much higher than the observed frequency. This linear calibration is the forecast probability multiplied by a constant, which for this study is 1/17, bringing it more in-line with the observed frequency. Although a direct comparison can't be made, the forecast percentages and observed frequency in this example has increased and the reliability has improved over the last few years compared to Gill & Buchanan (2014). This indicates the turbulence forecast models have improved over the last few years and the multi-model ensemble is at least as reliable as the individual ensembles.

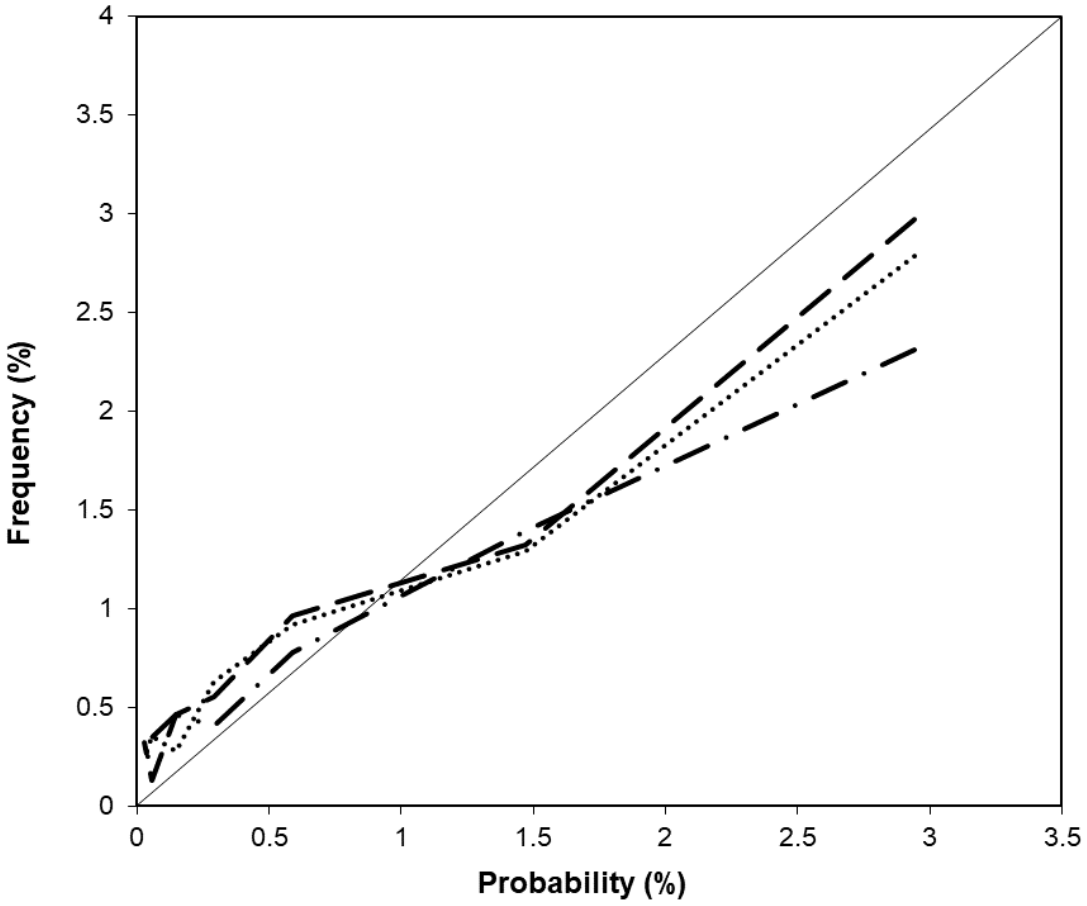

FIGURE 5.8: Reliability diagram of the MOGREPS-G (dot-dash), ECMWF (dot) and combined multi-model ensemble (dash). The data used has a forecast lead time between +24 hours and +33 hours between May 2016 and April 2017.

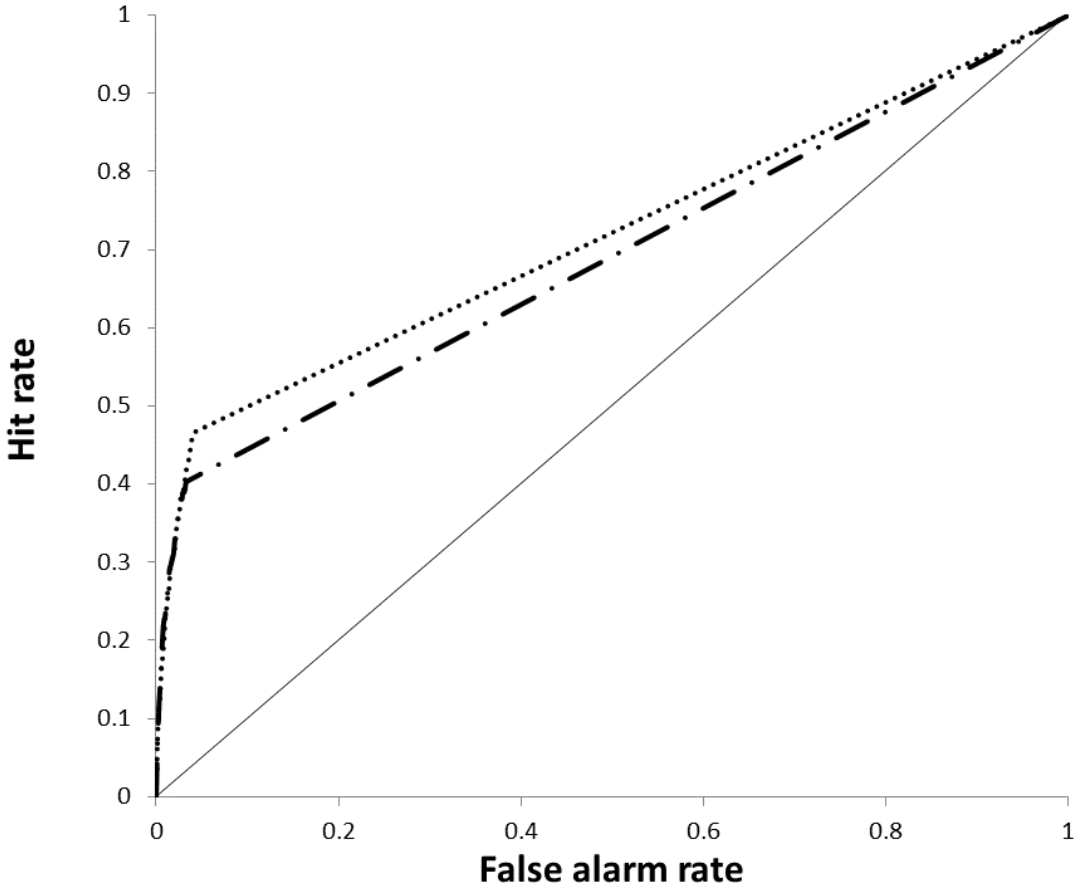

FIGURE 5.9: ROC plot of the global turbulence showing the forecast skill of the MOGREPS-G (dot-dash) and ECMWF 12 member ensemble (dot). The data used has a forecast lead time between +24 hours and +33 hours between May 2016 and April 2017.

5.4.3 Reduced Size Multi-Model Ensemble

So far it has been shown that combining two ensembles improves the forecast, but it is also important to understand how the individual ensembles compare to each other. To do this we must first reduce the ensemble size of the ECMWF forecast to make it a fair comparison. This is because a larger ensemble should give a larger forecast spread, and therefore improve the forecast result, and the ECMWF EPS has 51 members compared to the MOGREPS-G 12 members. To reduce the ensemble size we choose the first 12 members of the ECMWF ensemble. Each of the perturbed members is constructed to be equally likely, and each consecutive member has a 'pair-wise anti-symmetric perturbation' (Owens & Hewson, 2018). Therefore choosing consecutive members is a bias-free method for creating a sub sample. This is also how Buizza & Palmer (1998)

studied the impact of ensemble size on ensemble skill. They took pairs of perturbed members, so that each ensemble has pairs of members with the same positive and negative perturbation. Figure 5.9 is the ROC plot with the same ensemble size and can see that both models have almost the same forecast skill. When looking at Figure 5.10 we see that the ECMWF-EPS ensemble is more valuable. This would be useful when trying to combine the two ensembles using a weighted scheme to get the best forecast skill. As the ECMWF-EPS forecast is more valuable, a larger weight could be applied when creating the multi-model ensemble. But again in an operational system the ECMWF-EPS skill would be reduced as the longer lead time needed due to the time delay of the forecast would reduce the skill. So before an optimised weighted multi-model ensemble can be created, the ECMWF ensembles performance with the time delay would have to be analysed. This would then need to be extended to include all turbulence predictors to find the best multi-model ensemble forecast.

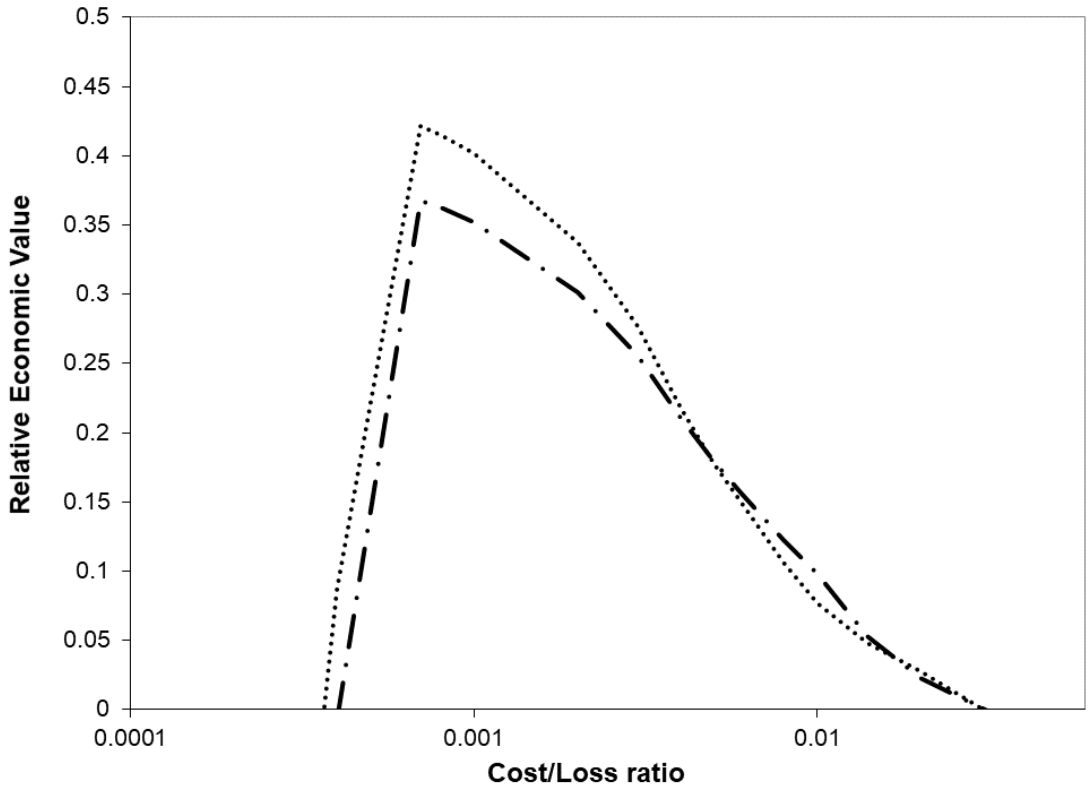

FIGURE 5.10: Value plot with a log scale x-axis of the global turbulence showing the forecast value of the MOGREPS-G (dot-dash) and ECMWF 12 member ensemble (dot). The data used has a forecast lead time between +24 hours and +33 hours between May 2016 and April 2017.

5.5 Conclusions and Further Work

This chapter has investigated the use of ensemble forecasts for aviation turbulence. By combining two ensembles to create a simple multi-model ensemble, we aimed to show improved forecast value and skill which could then be implemented operationally. To verify the forecasts, aircraft observations from a fleet of Boeing 747 and 777 aircraft were used and created a contingency table of results. From this the results were analysed to show the multi-model ensemble system is as skilful as a single model ensemble, which follows on from the work of Gill & Buchanan (2014); Buchanan (2016).

The results found indicate the forecast skill for the simple equally weighted multi-model ensemble is at least as skilful as the single model ensembles. This lack of significant improvement in the forecast skill was not expected, but this could be because when increasing the forecast spread we capture more turbulence events but also more false alarms. We would have to optimise this trade off to maximise the forecast skill, but in this study we are unable to show a significant improvement. However the value of the forecast is improved for the multi-model ensemble particularly at low cost/loss ratios, which are more important for operational use. Therefore to see an improvement in value at low cost/loss ratios shows it is worth implementing this multi-model approach as it would be more valuable in an operational setting. Our results also showed that the multi-model ensemble is as reliable as the single model ensembles, and therefore overall the multi-model ensemble is an improvement to the single model ensembles. Through combining two ensembles we gain consistency, gives more operational resilience and create one authoritative forecast whilst maintaining skill and increasing value, which would be particularly important in operational use in the future by the WAFCs.

Throughout the analysis, it is also found that the Ellrod & Knapp (1992) TI1 predictor is good at forecasting shear turbulence particularly. However not all the shear-induced turbulence events are forecast, and therefore one or more shear turbulence diagnostic would be beneficial similar to Kim et al. (2015). It would also be a good next step to include the Ellrod3 turbulence diagnostic from Sharman & Pearson (2017), as they showed its improved performance over other turbulence diagnostics and could be an easy step to improving the forecast skill. Also the MOGREPS-G ensemble is designed to be time lagged to create a 24 member ensemble, and this should also be investigated in further work (Met Office, 2017). An alternative method for creating a probabilistic

forecast would be a multi-diagnostic approach, rather than the traditional ensemble members approach. Kim et al. (2018) showed that a multi-diagnostic approach using two numerical models (NOAA's Global Forecast System and Met Office's Unified Model) for creating a probability forecast had a far greater statistical performance than the current WAFC forecast and any single CAT diagnostic. This is another example of how probabilistic forecasting can improve forecast skill, but uses a different method to create it. It is also vital to add convective and mountain wave turbulence predictors in any further studies. This would then take into account Convectively Induced Turbulence (CIT) and Mountain Wave Turbulence (MWT) that are also leading causes of aviation turbulence and account for many injuries to passengers and crew. Also CIT and MWT predictors could benefit more than the Ellrod & Knapp (1992) TI1 forecast, from the multi-model ensemble technique, making a multi-model ensemble superior to a single-model ensemble.

Chapter 6

Multi-Diagnostic Multi-Model Ensemble

In Chapter 5 the idea of using a multi-model ensemble for aviation turbulence forecasting was introduced. However, because only one turbulence diagnostic is used not all the turbulence events are forecasted . The first trial only used the Ellrod & Knapp (1992) Turbulence Index 1 (Ellrod TI1) which can only forecast shear turbulence and not Mountain Wave Turbulence (MWT) or Convectively Induced Turbulence (CIT). Therefore this chapter combines multiple turbulence diagnostics and multiple ensemble forecasts to create a multi-diagnostic multi-model ensemble.

6.1 Methodology

The method used for this multi-diagnostic multi-model ensemble is the same as Chapter 5. The same observational data as Section 5.1 is used which is the Derived Equivalent Vertical Gust (DEVG) (Equation 5.1) which is an aircraft independent measure and is taken from a fleet of Boeing 747 and 777 aircraft. This trial also uses a full year of ensemble data between September 2016 and August 2017 from the Met Office Global and Regional Ensemble Prediction System (MOGREPS-G) and the European Centre for Medium Range Weather Forecasting (ECMWF) Ensemble Prediction System. The MOGREPS-G ensemble has 12 members and the ECMWF has 51 members, and the data used in this project is output to a horizontal resolution of 1° and 26 vertical levels (although only six are useful for aviation), rather than their native grid (which was used in Chapter 5). This is a slightly higher resolution than the World Area Forecast Centres gridded forecast of 1.25° and 7 vertical levels (ICAO, 2016). We have output data from 00 UTC for T+24, T+27, T+30, T+33 hours. This could impact the results as we will only be focusing half the day, limiting the number of flights and diurnal variations that might exist in the atmosphere. We

start by creating a probabilistic forecast for a simple equal weighting multi-diagnostic single-model ensemble for both the MOGREPS-G and ECMWF ensembles, as well as an optimised multi-diagnostic single-model ensemble. The optimisation works by first prescribing a set of initial weightings for each diagnostic and each threshold, and then running an iterative scheme to change the weightings to optimise the Relative Operating Characteristic (ROC) plot by maximising the Area Under the Curve (AUC). This optimisation method was also used by Gill & Buchanan (2014). It is important to note that this is a trial and error method (for the initial weightings) and therefore the AUC we have will not be the best the models can achieve. A multi-diagnostic multi-model ensemble will also be created with an equal weighting for all diagnostics, thresholds and ensembles, as well as an optimised ensemble combining all thresholds, diagnostics and ensembles to maximise the AUC.

The turbulence diagnostics used in this trial are taken from the Graphical Turbulence Guidance system 3 (GTG3) (Sharman & Pearson, 2017) as we use the Ellrod TI1, the Brown index (Brown, 1973), a mountain wave turbulence predictor (MWT12 from Sharman & Pearson (2017)) the Richardson number (Ri) and convective precipitation accumulation. The Ellrod TI1 turbulence diagnostic is the same used in Chapters 3, 4 and 5 and is defined as:

$$\text{TI1} = \text{DEF} \times \text{VWS} = \left[\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)^2\right]^{1/2} \times \left[\left(\frac{\partial u}{\partial z}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2\right]^{1/2} \quad (6.1)$$

where u is the horizontal wind velocity in the East-West direction, v is the horizontal wind velocity in the North-South direction, x is distance in the East-West direction, y is distance in the North-South direction and z is distance in the vertical. The Brown index was used as part of Chapter 4 and Gill (2014) and is useful because it includes absolute vorticity, vertical wind shear and deformation and is defined as:

$$\text{Brown} = \left[0.3\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} + f\right)^2 + \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right)^2\right]^{1/2} \quad (6.2)$$

where u is the horizontal wind velocity in the East-West direction, v is the horizontal wind velocity in the North-South direction, x is distance in the East-West direction, y is distance in the North-South direction and f is the coriolis frequency. The MWT predictor used in the project was MWT12 from Sharman & Pearson (2017) and was one that performed best in their trial over the United

States and is defined as:

$$\text{MWT} = d_s \times |TEMPG| \tag{6.3}$$

where d_s is a near surface diagnostic (low-level wind speed perpendicular to the ridgeline) and $TEMPG$ is the horizontal temperature gradient. The Richardson number was used as part of Chapter 3 and 4 and is defined as:

$$\text{Ri} = \frac{N^2}{(\partial U/\partial z)^2} = \frac{(g/\theta)(\partial\theta/\partial z)}{(\partial U/\partial z)^2}, \tag{6.4}$$

where N^2 is the Brunt–Väisälä frequency squared, U is horizontal wind speed, z is altitude, g is gravitational acceleration, and θ is potential temperature. For the convective precipitation accumulation we combined both the convective rain accumulation and convective snow accumulation to create one convective precipitation diagnostic. The convective data was on a higher resolution grid, so we used Iris (Met Office, 2013) to linearly regrid the data so that it had the same horizontal resolution as the other turbulence diagnostics.

Diagnostic	Units	Ensemble	Thr1	Thr2	Thr3	Thr4	Thr5
Ellrod T1 $\times 10^{-7}$	s^{-2}	MOGREPS-G	3	5	8	11	20
	s^{-2}	ECMWF	3	5	8	11	20
Brown $\times 10^{-5}$	s^{-1}	MOGREPS-G	12	15	20	25	30
	s^{-1}	ECMWF	12	15	20	25	30
MWT $\times 10^{-1}$	$K\,s^{-1}$	MOGREPS-G	1	5	10	15	20
	$K\,s^{-1}$	ECMWF	1	5	10	15	20
Richardson $\times 10^{-2}$		MOGREPS-G	5	10	20	50	100
		ECMWF	5	10	20	50	100
Convection $\times 10^{-2}$	$kg\,m^{-2}$	MOGREPS-G	1000	5000	10000	15000	20000
	m	ECMWF	1	5	10	15	20

TABLE 6.1: Table showing the five turbulence thresholds used for each of the turbulence diagnostics in this study

To create a probabilistic turbulence forecast, thresholds are set for each of the turbulence diagnostics and any time an ensemble member exceeds that threshold, it is classed as a turbulence forecast. The higher the number of ensemble members exceeding the threshold, the higher the probability of turbulence. The thresholds used for each diagnostic and ensemble are shown in Table 6.1 and are chosen so they divide the distribution of forecasted values into approximately even groups. The initial thresholds were selected based on the highest values of the turbulence diagnostic on 01/05/2016 (the first day of the trial), and then the ensemble forecast was created. Depending on if the

threshold was too high or too low it was then adjusted and the analysis run again until we had five suitable thresholds. We can see that the thresholds chosen are the same for both forecast centres, except for convective precipitation accumulation. This is because the thresholds needed for MOGREPS-G are much higher than for ECMWF because the units are different. The MOGREPS-G ensemble has units of 'kg m^{-2}' whereas ECMWF has units of 'm'. The two units are related however, because 1 kg m^{-2} is equivalent to 1 mm of precipitation (which is the same as 0.001 m). This therefore means the thresholds are actually the same but the units are slightly different.

The verification method is also the same as Section 5.3. We use the observational aircraft data $\pm$1.5 hours of the forecast time to see if the probability forecast can forecast the moderate-or-greater turbulence events. From these results we plot a ROC plot which shows forecast skill by plotting the hit rate (Equation 5.5) against the false alarm rate (Equation 5.6) (Jolliffe & Stephenson, 2012; Gill, 2016). The more skilful the forecast, the higher the AUC will be, and this will show a forecast that has found a good balance between forecasting as many hits as possible while minimising the number of false alarms. As discussed in Chapter 5, the forecast skill shown by a ROC plot is not the only way to show how useful a forecast is. The relative economic value (Richardson, 2000; Jolliffe & Stephenson, 2012) shows how valuable the forecast is for a given cost/loss ratio which will be user specific. If the forecast is more valuable for all cost loss ratios, known as sufficiency (Ehrendorfer & Murphy, 1988), any user would benefit from this model. Forecast reliability is also a way of understanding how well a forecast model performs (Jolliffe & Stephenson, 2012; Gill, 2016). By plotting the forecasted probability and the actual observed frequency, we can understand if we are over- or under-forecasting the turbulence events.

6.2 Single-Diagnostic Ensemble

To create a multi-diagnostic multi-model ensemble we must first create the probabilistic forecast for each of the turbulence diagnostics, before combining them. To do this we followed the same method as shown in Chapter 5 by creating a single-diagnostic single-model ensemble but this time for each of the five turbulence diagnostics. The probability forecast was created for both the MOGREPS-G ensemble and the ECMWF 51 member ensemble. Figure 6.1 is an example ROC plot of the Ellrod TI1 turbulence diagnostic for the ECMWF 51 member ensemble and MOGREPS-G ensemble using the first and therefore

lowest turbulence threshold. A direct comparison cannot be made with Figure 5.5 because the data periods are not identical and this trial has a coarser horizontal resolution (and as seen is Chapter 3 resolution has an impact on the location of turbulence forecasted). What we see is that both the ECMWF and MOGREPS-G have a less steep line than in Figure 5.5, which suggests that this trial has less skill than the previous trial. This is likely to be caused by the resolution of the forecast being much coarser in this trial and also the convective events have not been removed, so the Ellrod TI1 predictor will be less skilful. However, the similarities are the MOGREPS-G ensemble is slightly below the ECMWF line particularly at the higher false alarm rates. This would lead to a higher AUC for ECMWF than MOGREPS-G for the single-diagnostic single-model ensemble. The issue of the line being shorter for the MOGREPS-G ensemble is not of so much importance in this study and therefore a more direct comparison of AUC can be made.

We also created a combined equal weighted single-diagnostic multi-model ensemble. The AUC for each of the turbulence diagnostics and all five thresholds for the MOGREPS-G, ECMWF and the multi-model ensembles are shown in Table 6.2. As indicated in Figure 6.1 the ECMWF ensemble has a higher AUC for most of the diagnostics and thresholds. However, the MOGREPS-G is more skilful for some diagnostics, such as MWT. It is also interesting that the highest two thresholds for MWT have no skill, as the threshold is too high to capture any turbulence events. This is not a surprise because the MWT predictor will only be forecasting events over and around mountains. Therefore the number of events that could be forecasted are much lower than the other turbulence diagnostics that forecast, for example, around the jet stream. The multi-model ensemble is more skilful than either the ECMWF or MOGREPS-G ensemble for each predictor. This follows on from Chapter 5 and shows that having a multi-model ensemble is more skilful than a single model ensemble. However as in Chapter 5, we can't show it at the 95% confidence interval. Both the upper and lower confidence interval bounds for the multi-model ensemble are shown in Table 6.2 and there are only six occasions where the multi-model ensemble is significantly more skilful than MOGREPS-G, only one occasion for ECMWF and there are no occasions when it is significantly higher than both.

To illustrate this, we have taken the two thresholds with the highest AUC for each of the diagnostics (which is threshold 1 and 2 for all diagnostics) and plotted them in Figure 6.2. The Triangle icon is MOGREPS-G, the Diamond is ECMWF and the Circle is the combined multi-model ensemble. The 95% confidence interval for the multi-model ensemble are also included. We see

	Ellrod TI1				
	Thr1	Thr2	Thr3	Thr4	Thr5
ECMWF 51	0.7197	0.7032	0.6486	0.5978	0.5162
MOGREPS-G 12	0.7111	0.6905	0.6272	0.5854	0.5162
Multi-model 95% lower CI	0.7013	0.6973	0.6378	0.5923	0.5126
Multi-model	0.7244	0.7197	0.6601	0.6104	0.5210
Multi-model 95% upper CI	0.7477	0.7425	0.6837	0.6298	0.5306
	Brown Index				
	Thr1	Thr2	Thr3	Thr4	Thr5
ECMWF 51	0.6189	0.5781	0.5311	0.5170	0.5035
MOGREPS-G 12	0.5912	0.5595	0.5276	0.5125	0.5056
Multi-model 95% lower CI	0.6026	0.5643	0.5199	0.5103	0.5013
Multi-model	0.6244	0.5858	0.5328	0.5186	0.5054
Multi-model 95% upper CI	0.6476	0.6065	0.5454	0.5280	0.5105
	MWT				
	Thr1	Thr2	Thr3	Thr4	Thr5
ECMWF 51	0.5293	0.5038	0.5000	0.5000	0.5000
MOGREPS-G 12	0.5298	0.5055	0.5020	0.5000	0.5000
Multi-model 95% lower CI	0.5182	0.5015	0.4999	0.5000	0.5000
Multi-model	0.5302	0.5064	0.5019	0.5000	0.5000
Multi-model 95% upper CI	0.5438	0.5121	0.5051	0.5000	0.5000
	Richardson Number				
	Thr1	Thr2	Thr3	Thr4	Thr5
ECMWF 51	0.7694	0.7703	0.6833	0.5210	0.5040
MOGREPS-G 12	0.7694	0.7703	0.6730	0.5231	0.5040
Multi-model 95% lower CI	0.7596	0.7650	0.6813	0.5219	0.5009
Multi-model	0.7773	0.7864	0.7036	0.5326	0.5040
Multi-model 95% upper CI	0.7951	0.8060	0.7255	0.5445	0.5086
	Convective precip accumulation				
	Thr1	Thr2	Thr3	Thr4	Thr5
ECMWF 51	0.7142	0.7216	0.6994	0.6812	0.6480
MOGREPS-G 12	0.7126	0.7059	0.6855	0.6720	0.6611
Multi-model 95% lower CI	0.6968	0.6987	0.6833	0.6686	0.6534
Multi-model	0.7195	0.7222	0.7076	0.6924	0.6767
Multi-model 95% upper CI	0.7439	0.7461	0.7303	0.7154	0.6997

TABLE 6.2: Table showing the Area Under the ROC Curve (AUC) for five thresholds for each of the five turbulence diagnostics for the ECMWF 51-member ensemble, MOGREPS-G ensemble, combined multi-model ensemble 95% lower confidence interval, combined multi-model ensemble and combined multi-model ensemble 95% upper confidence interval. The data used has a forecast lead time between T+24 hours and T+33 hours between September 2016 and August 2017.

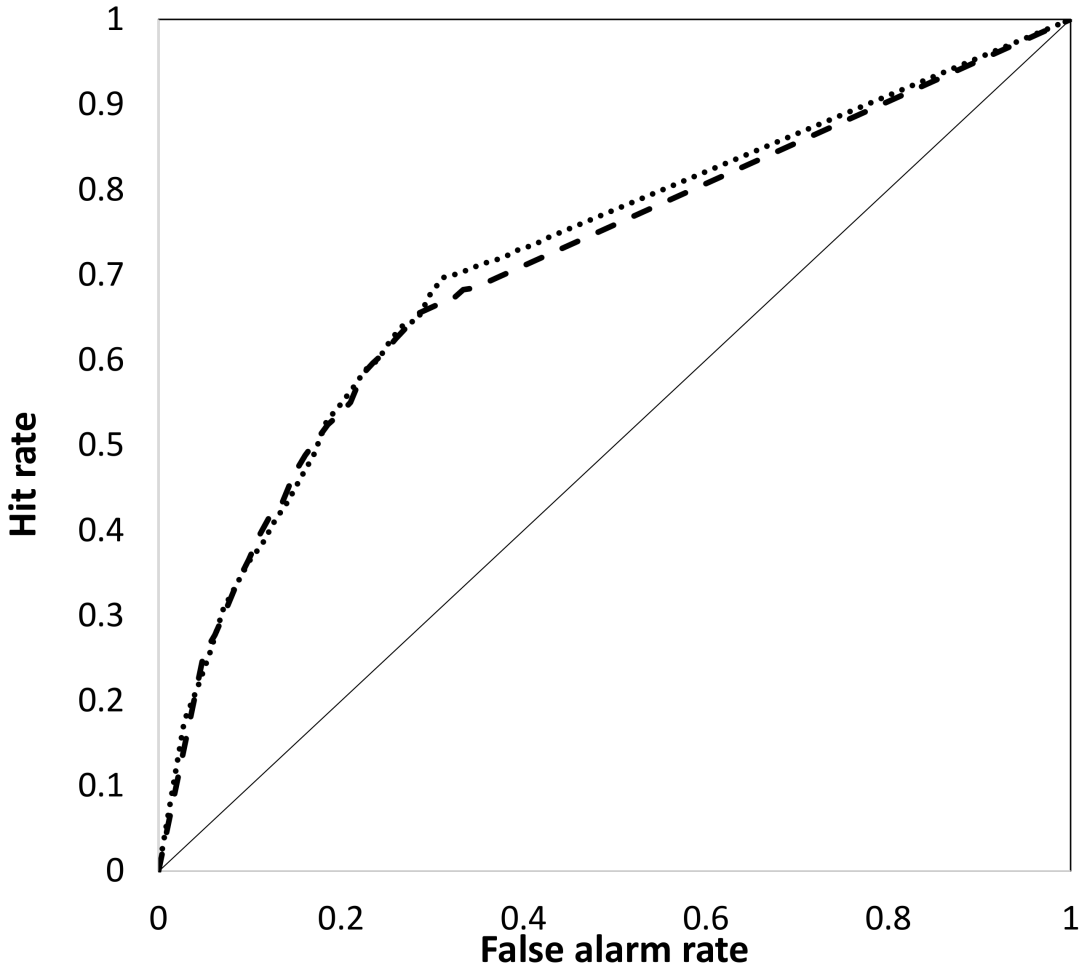

FIGURE 6.1: ROC plot of the global turbulence for threshold 1 of the Ellrod TI1 diagnostic for the MOGREPS-G ensemble (dash) and ECMWF 51-member ensemble (dot). The data used has a forecast lead time between T+24 hours and T+33 hours between September 2016 and August 2017.

that for the most part the ECMWF is more skilful than MOGREPS-G and there are only three occasions where the multi-model ensemble is significantly more skilful than either of the single model ensembles. Figure 6.2 also shows how well each of the different diagnostics perform, with the Richardson number producing the highest AUC and Ellrod TI1 and convective precipitation accumulation just behind. The Brown index and MWT however perform worse with much lower AUC, however including them may still have some benefit in a multi-diagnostic ensemble as they may be forecasting more extreme events that the other predictors could be missing. This is especially true of the MWT predictor as none of the the other predictors are targeting MWT and therefore without it, these events will be missed.

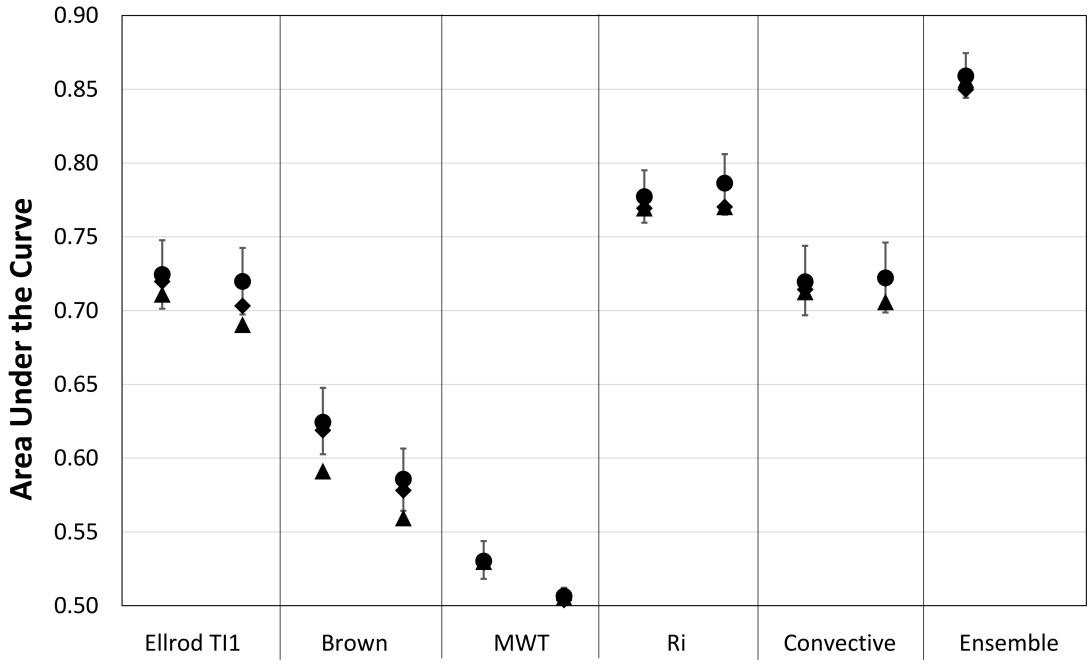

FIGURE 6.2: Plot showing the Area Under the ROC Curve (AUC) for the 2 thresholds with the highest AUC for 5 turbulence diagnostics from the MOGREPS-G ensemble (triangle), ECMWF 51-member ensemble (diamond) and combined multi-model ensemble (circle). The combined single-diagnostic multi-model ensemble has error bars showing the 95% confidence interval. For reference the combined equal weighted multi-diagnostic single-model ensemble and multi-diagnostic multi-model ensemble have also been included. The data used has a forecast lead time between T+24 hours and T+33 hours between September 2016 and August 2017.

6.3 Multi-Diagnostic Ensemble

The multi-diagnostic multi-model ensemble can be created in two ways, first by combining all thresholds, diagnostics and ensembles equally to create a simple super ensemble. The second is to combine them by optimising the diagnostics, thresholds and models used. In this study we create both the simple and optimised multi-diagnostic ensemble for the MOGREPS-G and ECMWF ensemble as well as a multi-diagnostic multi-model ensemble. The thresholds used in the multi-diagnostic multi-model ensemble are shown in Table 6.3. Some of the diagnostics have more weight (for example Richardson number and convective precipitation accumulation) and for some diagnostics, one model has a higher weight than the other (for example MOGREPS-G has more weight for the Brown index than ECMWF). The ROC plot for the optimised ensembles are shown in Figure 6.3. The first thing to note is it has a

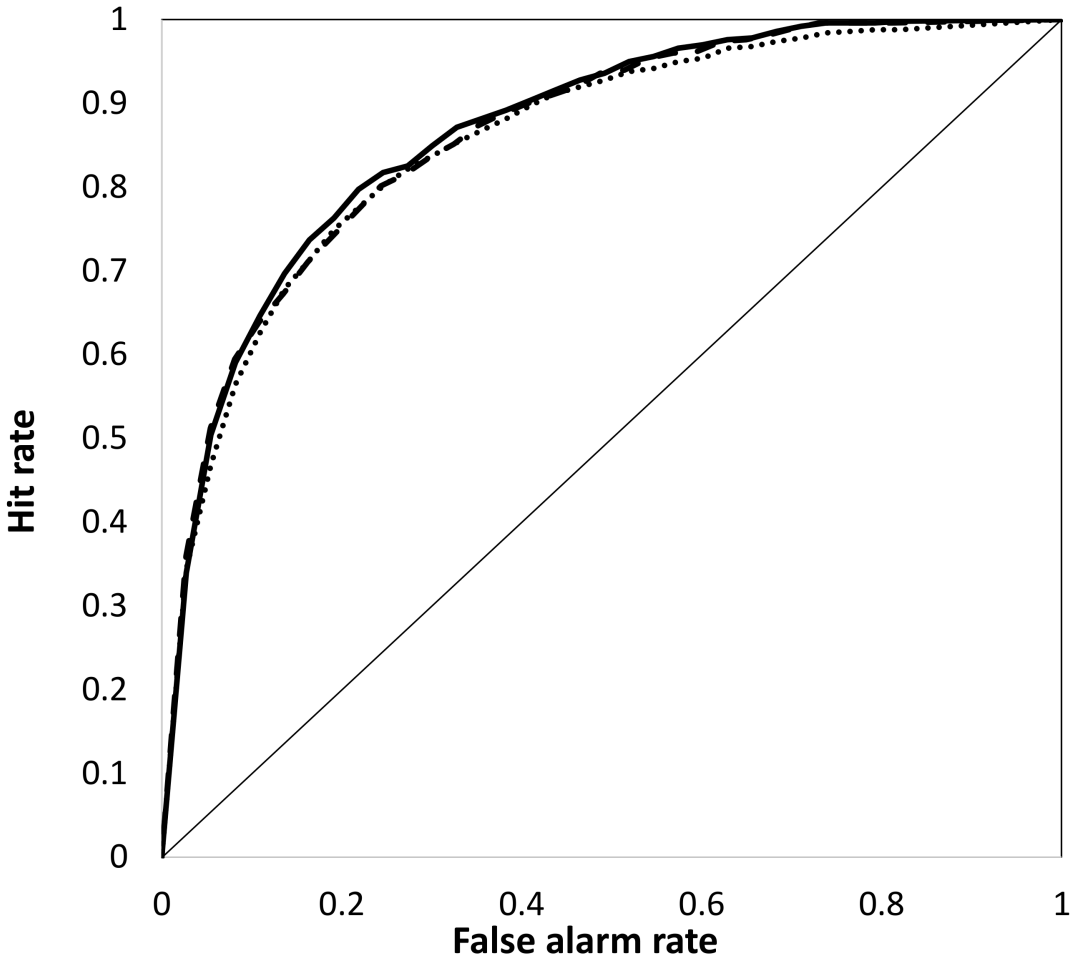

FIGURE 6.3: ROC plot of the global turbulence for a multi-diagnostic MOGREPS-G ensemble (dash), multi-diagnostic ECMWF 51-member ensemble (dot) and combined multi-diagnostic multi-model ensemble (solid). Five turbulence thresholds for each turbulence diagnostic are optimally combined to maximise the area under the ROC curve and uses a forecast lead time between T+24 hours and T+33 hours between September 2016 and August 2017.

much higher AUC than the single diagnostic ensembles shown in Figure 6.1. This shows that using multiple thresholds and diagnostics we can increase the forecast skill. It is also interesting that the MOGREPS-G ensemble has a higher skill than the ECMWF ensemble and at very low false alarm rates, MOGREPS-G is more skilful than the multi-model ensemble. This is a very interesting result and one that was not expected. In Section 6.2 we found that ECMWF has a consistently higher AUC for almost every diagnostic, yet when combining diagnostics we see that the MOGREPS-G ensemble is more skilful.

Diagnostic	Ensemble	Thr1	Thr2	Thr3	Thr4	Thr5
Ellrod T1	MOGREPS-G	0.1	0.1	0.1	0.4	0.1
	ECMWF	0.0	0.1	0.1	1.3	0.1
Brown	MOGREPS-G	0.0	0.1	0.1	0.0	0.9
	ECMWF	0.0	0.0	0.0	0.0	0.0
MWT	MOGREPS-G	0.1	0.0	0.0	0.0	0.0
	ECMWF	0.3	0.0	0.0	0.0	0.0
Richardson	MOGREPS-G	0.1	0.2	0.2	0.9	0.1
	ECMWF	0.1	0.1	0.1	0.3	0.1
Convection	MOGREPS-G	0.2	0.1	0.1	0.1	0.2
	ECMWF	0.0	0.1	0.2	0.1	0.1

TABLE 6.3: Table showing the weightings used of the MOGREPS-G and ECMWF 51 member ensemble to create the optimised multi-diagnostic multi-model ensemble.

For a single-diagnostic ensemble to be skilful, the forecast spread needs to be large enough to capture as many turbulence events as possible (hits) but also avoid too many false alarms. The ECMWF ensemble achieves this more successfully than the MOGREPS-G ensemble, which is why it has a higher AUC for almost all the diagnostics. However, if each of the diagnostics forecasts a turbulence event slightly differently, when combining ensembles, the area of forecasted turbulence will be increased. This can lead to more hits but there could also be more false alarms. The larger the ensemble spread of the individual diagnostics, the greater the ensemble spread of the multi-diagnostic ensemble. It appears in this case that although the spread for ECMWF works well for the individual diagnostics, when combining them, the number of false alarms starts to outweigh the number of hits and the resultant forecast skill is reduced. The MOGREPS-G ensemble however, has less forecast spread for the individual diagnostics, but then when combining them, the forecast spread

increases just the right amount and the forecast skill is then higher than the ECMWF ensemble.

	Combined equal	Combined optimised
ECMWF 51	0.8498	0.8555
MOGREPS-G 12	0.8564	0.8630
Multi-model 95% lower CI	0.8442	0.8530
Multi-model	0.8590	0.8679
Multi-model 95% upper CI	0.8745	0.8829

TABLE 6.4: Table showing the Area Under the ROC Curve (AUC) for the combined equal weighting multi-diagnostic ensemble and combined optimised multi-diagnostic ensemble for the ECMWF 51-member ensemble, MOGREPS-G ensemble, combined multi-model ensemble 95% lower confidence interval, combined multi-model ensemble and combined multi-model ensemble 95% upper confidence interval. The data used has a forecast lead time between T+24 hours and T+33 hours between September 2016 and August 2017.

Table 6.4 shows the AUC for the combined equal and combined optimised single-model and multi-model ensembles. The multi-model ensemble also has the 95% upper and lower confidence intervals and again we see that it is not significantly better than either of the two single-model ensembles. Figure 6.4 shows the results in Table 6.4 as a bar chart and we can clearly see that for both the combined equal and combined optimised ensemble, MOGREPS-G is more skilful. The optimised ensemble is also more skilful than the combined equal ensemble and shows the benefit of taking time to optimise the thresholds and diagnostics used.

Figure 6.5 is a value plot of the optimised multi-diagnostic ensemble for MOGREPS-G, ECMWF and multi-model ensemble. We see that it is the MOGREPS-G ensemble that has more value for all cost/loss ratios than the ECMWF ensemble which is the opposite to what was found in Figure 5.6 from Chapter 5. This follows on from the ROC plots where MOGREPS-G outperformed the ECMWF ensemble and again suggests that when combining the diagnostics, ECMWF has too large a forecast spread and therefore the balance between hits and false alarms is not quite optimised. What we also see is that the combined multi-model ensemble is not more valuable for all cost/loss ratios. This suggests that for some consumers it might be more beneficial to just have the MOGREPS-G ensemble and not include the ECMWF ensemble at all. However, the multi-model ensemble would add operational resilience and the

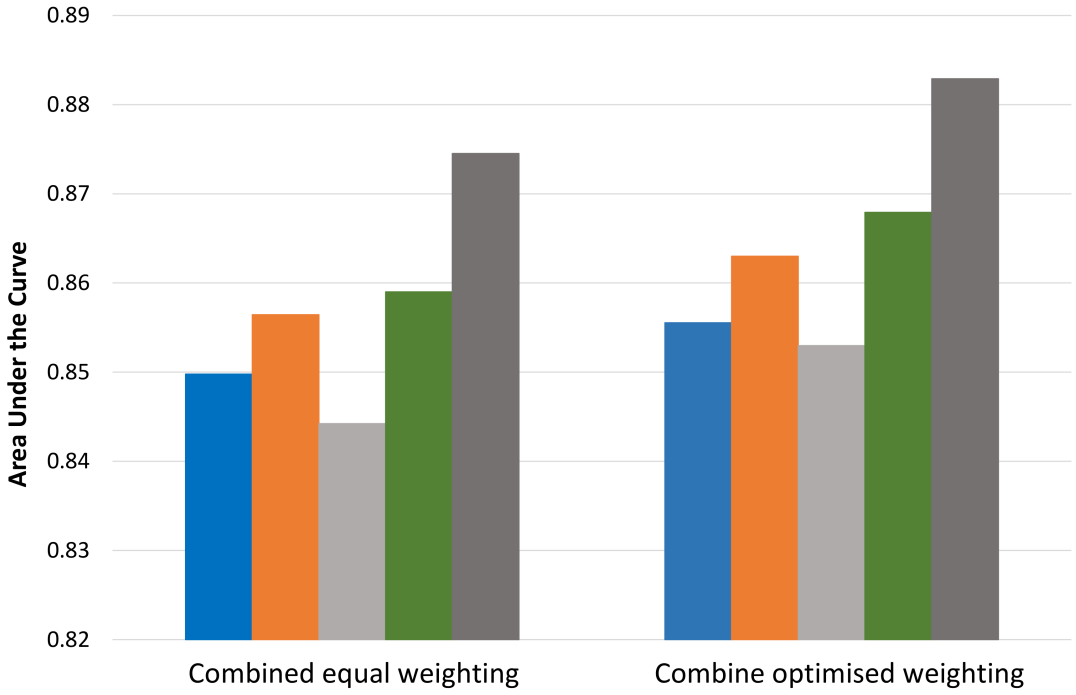

FIGURE 6.4: Bar chart showing the Area Under the ROC Curve (AUC) for the multi-diagnostic ECMWF 51-member ensemble (dark blue), multi-diagnostic MOGREPS-G ensemble (orange), combined multi-diagnostic multi-model ensemble 95% lower confidence interval (light grey), combined multi-diagnostic multi-model ensemble (green) and combined multi-diagnostic multi-model ensemble 95% upper confidence interval (dark grey). For the bar chart on the left, the five turbulence thresholds for each turbulence diagnostic are combined equally and on the right the five turbulence thresholds for each turbulence diagnostic are optimally combined to maximise the area under the ROC curve. The data used has a forecast lead time between T+24 hours and T+33 hours between September 2016 and August 2017.

multi-model ensemble is almost as valuable as the MOGREPS-G ensemble and would therefore be the more ideal option.

Forecast reliability is shown in Figure 6.6 and a calibration constant has been applied to the forecast probability. This constant is 1/150 for MOGREPS-G, 1/90 for ECMWF and 1/200 for the multi-model ensemble. What we find is there is a limit to how reliable the forecast can be. It appears that at a forecast probability above 0.25%, the frequency increases far faster than the forecasted probability. This is a problem with applying a linear calibration constant. For example, the highest actual forecast probability for the multi-diagnostic multi-model ensemble before calibration is 88%. However, the linear calibration

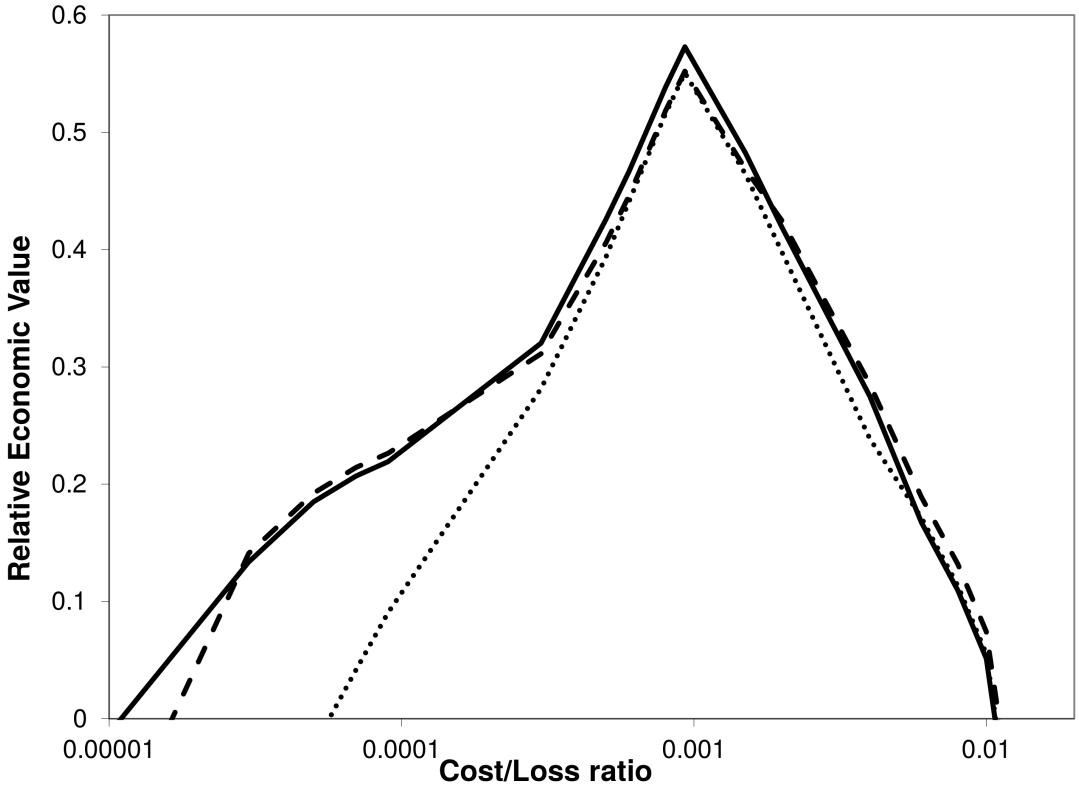

FIGURE 6.5: Value plot with a log scale x-axis of the global turbulence showing the forecast skill for a multi-diagnostic MOGREPS-G ensemble (dash), multi-diagnostic ECMWF 51-member ensemble (dot) and combined multi-diagnostic multi-model ensemble (solid). Five turbulence thresholds for each turbulence diagnostic are optimally combined to maximise the area under the ROC curve and uses a forecast lead time between T+24 hours and T+33 hours between September 2016 and August 2017.

(1/200) brings it down to 0.44% which is much lower than the observed frequency of 1.18%. The linear constant is suitable for most of the probability thresholds, with only the highest two for each ensemble resulting in an under-forecast. It could be possible to apply a non-linear constant, however for this study we have kept it simple with the linear constant.

6.4 Reduced Size Multi-Diagnostic Ensemble

As in Chapter 5, it is important to understand how the two ensembles compare to each other with the same ensemble size. The ECMWF ensemble has 51 members, so we reduce this by selecting the first 12 members and produce a bias free sub sample to directly compare with the 12 member MOGREPS-G

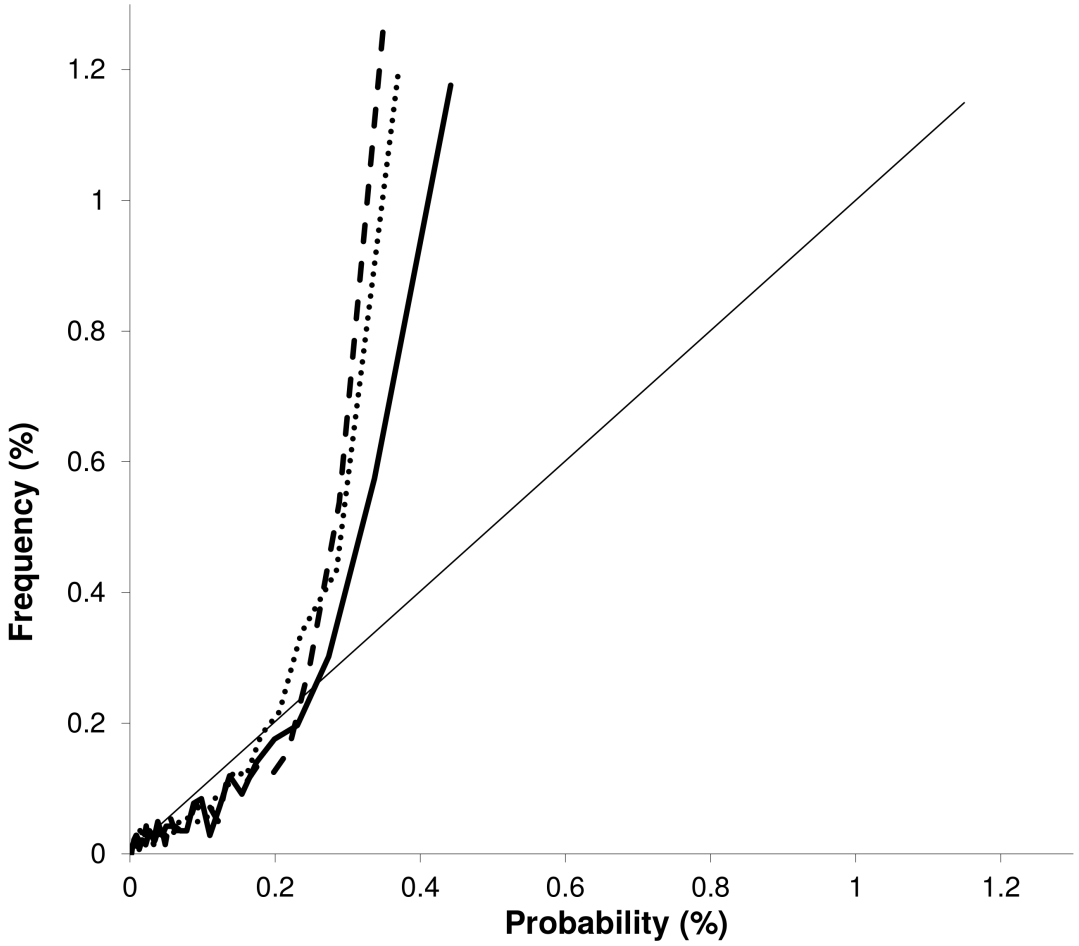

FIGURE 6.6: Reliability diagram for a multi-diagnostic MOGREPS-G ensemble (dash), multi-diagnostic ECMWF 51-member ensemble (dot) and combined multi-diagnostic multi-model ensemble (solid). Five turbulence thresholds for each turbulence diagnostic are optimally combined to maximise the Area under the ROC Curve and uses a forecast lead time between T+24 hours and T+33 hours between September 2016 and August 2017.

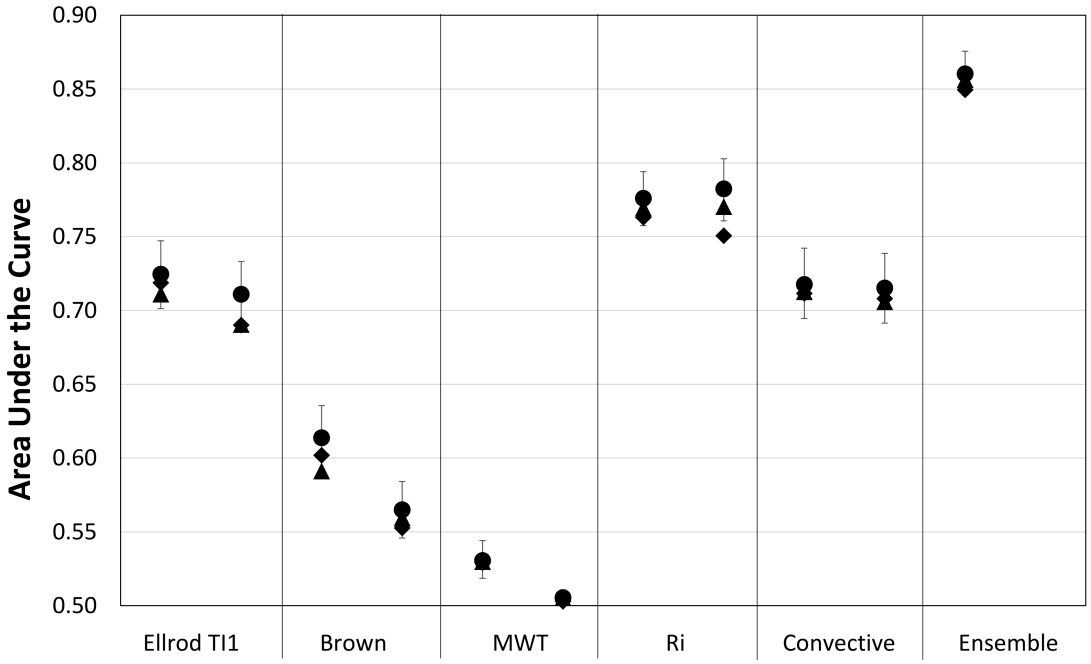

FIGURE 6.7: Plot showing the Area Under the ROC Curve (AUC) for the 2 thresholds with the highest AUC for five turbulence diagnostics from the MOGREPS-G ensemble (triangle), ECMWF 12-member ensemble (diamond) and combined multi-model ensemble (circle). The combined multi-model ensemble has error bars showing the 95% confidence interval. For reference the combined equal weighted multi-diagnostic single-model ensemble and multi-diagnostic multi-model ensemble have also been included. The data used has a forecast lead time between T+24 hours and T+33 hours between September 2016 and August 2017.

ensemble. The AUC for the first two thresholds for each turbulence diagnostic is shown in Figure 6.7. The circle is the multi-model ensemble with the 95% confidence intervals displayed, the MOGREPS-G ensemble (triangle) and ECMWF 12-member ensemble (diamond) are also shown. As in Figure 6.2 the multi-model ensemble is more skilful than both of the single model ensembles and for the most part the ECMWF ensemble is slightly more skilful than the MOGREPS-G ensemble. It is interesting to note that the 12 member ensemble seems to be slightly less skilful than the 51 member ECMWF ensemble, which is what we found in Chapter 5, but here we find it is consistent across all turbulence diagnostics.

Figure 6.8 is a bar chart showing the combined multi-diagnostic ensemble for the ECMWF 12-member ensemble, MOGREPS-G ensemble and combine multi-model ensemble with its upper and lower 95% confidence interval.

Diagnostic	Ensemble	Thr1	Thr2	Thr3	Thr4	Thr5
Ellrod T1	MOGREPS-G	0.1	0.1	0.1	0.8	0.2
	ECMWF	0.0	0.1	0.1	0.5	0.1
Brown	MOGREPS-G	0.0	0.0	0.2	0.0	0.0
	ECMWF	0.0	0.0	0.0	0.0	0.0
MWT	MOGREPS-G	0.1	0.0	0.0	0.0	0.0
	ECMWF	0.4	0.0	0.0	0.0	0.0
Richardson	MOGREPS-G	0.1	0.1	0.1	1.5	0.1
	ECMWF	0.0	0.1	0.2	0.1	0.1
Convection	MOGREPS-G	0.1	0.1	0.1	0.1	0.1
	ECMWF	0.0	0.1	0.2	0.1	0.1

TABLE 6.5: Table showing the weightings used of the MOGREPS-G and ECMWF 12 member ensemble to create the optimised multi-diagnostic multi-model ensemble.

The weightings used in the optimised multi-diagnostic multi-model ensemble are shown in Table 6.5. We find that the multi-model ensemble with the 12-member ECMWF ensemble is more skilful than the multi-model ensemble that uses the 51-member ensemble (from Figure 6.4). This agrees with Section 6.3 that a smaller ensemble spread for each individual diagnostic will give an overall better performance when combined in a multi-diagnostic ensemble. This is why the MOGREPS-G ensemble is more skilful than the ECMWF ensemble.

To investigate this further, Figure 6.9 is a plot showing the AUC for the ECMWF 12-member ensemble (circle) with its 95% confidence interval displayed, the ECMWF 51-member ensemble (diamond) and the MOGREPS-G ensemble (triangle) for the individual diagnostics. As expected, the 51-member ECMWF ensemble is more skilful than the 12 member ensemble and in some cases is significantly better. The MOGREPS-G ensemble has a similar forecast skill to the ECMWF 12-member ensemble and therefore the reduced number of members results in a lower AUC for the individual diagnostics. However, Figure 6.10 shows the equal combined multi-diagnostic ECMWF 51 member ensemble is slightly more skilful than the 12 member ensemble, but when optimised, the 12-member ensemble is much more skilful. It is also interesting that the MOGREPS-G ensemble is more skilful than either of the ECMWF multi-diagnostic ensembles.

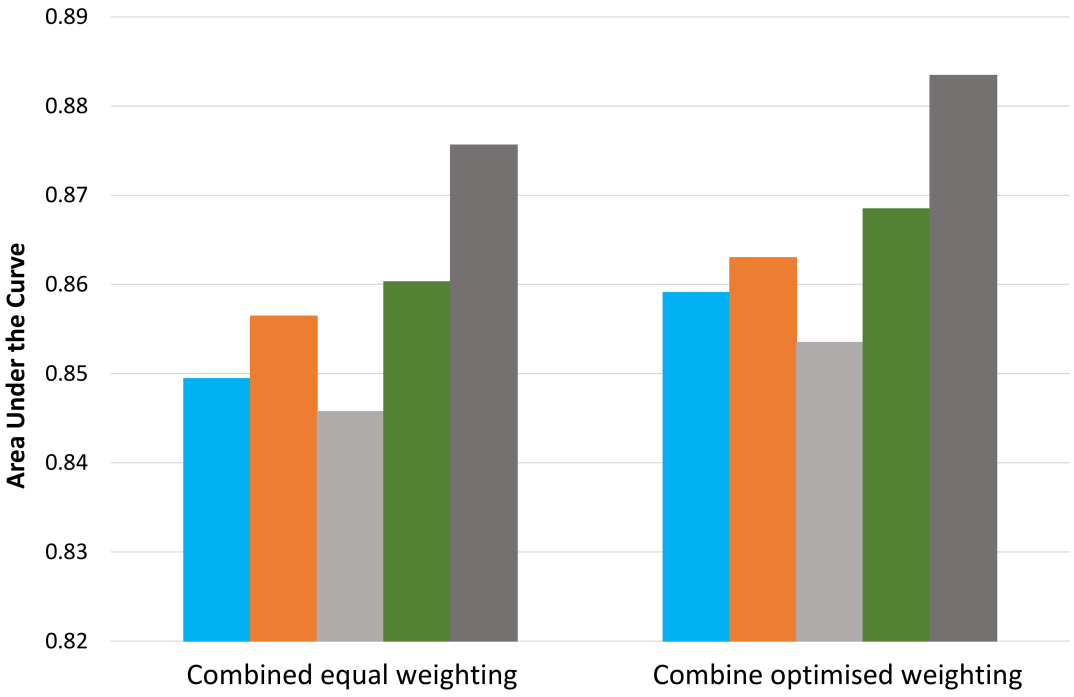

FIGURE 6.8: Bar chart showing the Area Under the ROC Curve (AUC) for the multi-diagnostic ECMWF 12-member ensemble (light blue), multi-diagnostic MOGREPS-G ensemble (orange), combined multi-diagnostic multi-model ensemble 95% lower confidence interval (light grey), combined multi-diagnostic multi-model ensemble (green) and combined multi-diagnostic multi-model ensemble 95% upper confidence interval (dark grey). For the bar chart on the left, the five turbulence thresholds for each turbulence diagnostic are combined equally. On the right of the bar chart, the five turbulence thresholds for each turbulence diagnostic are optimally combined to maximise the area under the ROC curve. The data used has a forecast lead time between T+24 hours and T+33 hours between September 2016 and August 2017.

6.5 Conclusions and Further Work

This study has expanded on the work in Chapter 5 by investigating the forecast skill of a multi-diagnostic multi-model ensemble for aviation turbulence. As in Chapter 5, we created probabilistic forecasts however this study created them for five turbulence diagnostics (Ellrod TI1, Brown Index, mountain wave turbulence, Richardson number and convective precipitation accumulation) and two ensembles: the Met Office Global and Regional Ensemble (MOGREPS-G) and the European Centre for Medium Range Weather Forecasting (ECMWF) ensemble prediction system. By combining the ensemble predictors a multi-diagnostic ensemble can be created. Then combining the two ensembles, a

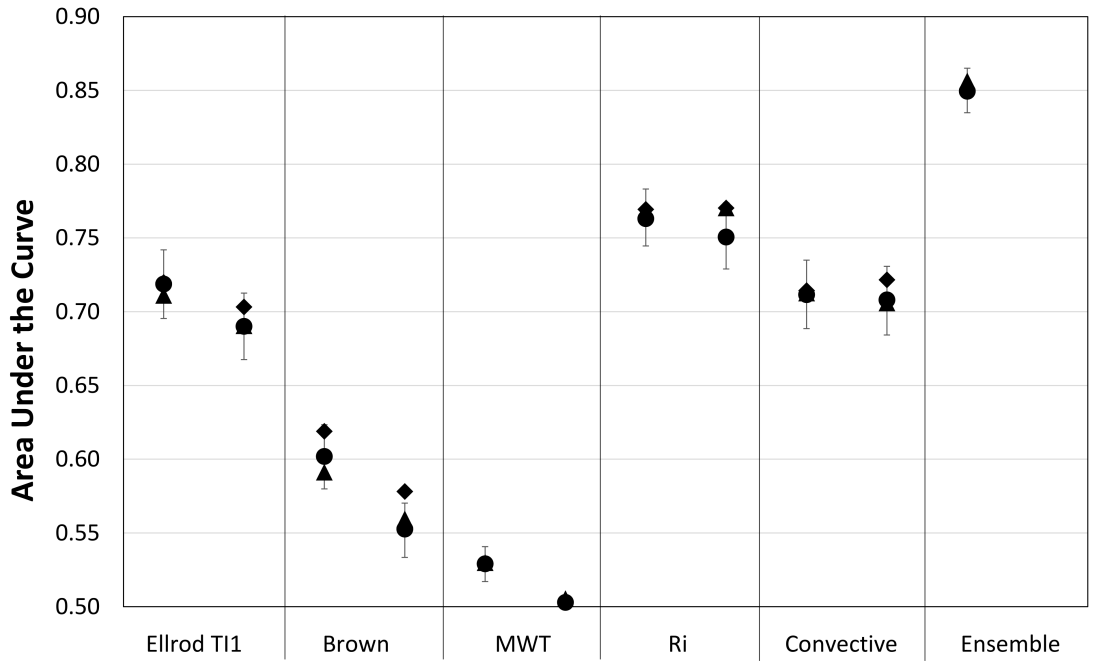

FIGURE 6.9: Plot showing the Area Under the ROC Curve (AUC) for the 2 thresholds with the highest AUC for five turbulence diagnostics from the MOGREPS-G ensemble (triangle), ECMWF 51-member ensemble (diamond) and ECMWF 12-member ensemble (circle). The ECMWF 12-member ensemble has error bars showing the 95% confidence interval. For reference the combined equal weighted multi-diagnostic single-model ensemble have also been included. The data used has a forecast lead time between T+24 hours and T+33 hours between September 2016 and August 2017.

multi-diagnostic multi-model ensemble can be created. The trial ran from September 2016 to August 2017 and used the 00 UTC forecast run and foretasted T+24, T+27, T+30 and T+33 hours. The forecast was verified against a fleet of Boeing 747 and 777 aircraft.

The results in this study agreed with Chapter 5 that the ECMWF 51 member ensemble was more skilful than the MOGREPS-G ensemble for the individual diagnostics. The multi-model ensemble was also more skilful than either of the single model ensembles (but not significantly for most examples). When combining the predictors, the multi-diagnostic ensemble was more skilful for the MOGREPS-G ensemble than the ECMWF ensemble for both the equal combined and optimised ensemble. Again, the multi-diagnostic multi-model ensemble was more skilful than the two single model ensembles. The relative economic value of the multi-diagnostic ensemble was also plotted and the MOGREPS-G ensemble was more valuable than the ECMWF ensemble,

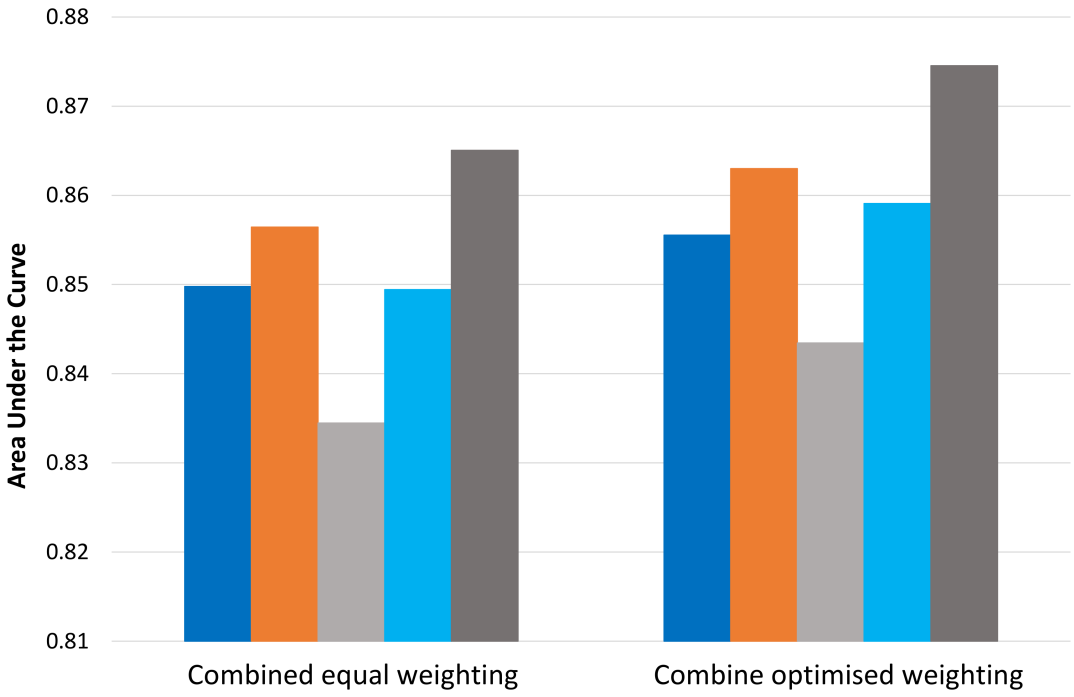

FIGURE 6.10: Bar chart showing the Area Under the ROC Curve (AUC) for the multi-diagnostic 51-member ensemble (dark blue), multi-diagnostic MOGREPS-G ensemble (orange), multi-diagnostic ECMWF 12-member ensemble 95% lower confidence interval (light grey), multi-diagnostic ECMWF 12-member ensemble (light blue) and multi-diagnostic ECMWF 12-member ensemble 95% upper confidence interval (dark grey). For the bar chart on the left, the five turbulence thresholds for each turbulence diagnostic are combined equally and on the right the five turbulence thresholds for each turbulence diagnostic are optimally combined to maximise the area under the ROC curve. The data used has a forecast lead time between T+24 hours and T+33 hours between September 2016 and August 2017.

and in some cases was also more valuable than the multi-model ensemble. The forecast reliability was also plotted and all ensembles have similar reliability, however at 0.25% forecast probability the ensembles start under-forecasting the events. To overcome this it could be possible to apply a nonlinear calibration, however we have kept a linear calibration in this study as it is suitable for most of the forecast probabilities.

Since the MOGREPS-G ensemble was more skilful than the ECMWF ensemble it was important to understand why, and therefore we created a 12 member ECMWF ensemble and ran the verification again. It was found that the 12 member ensemble for the individual diagnostics was less skilful than the 51 member ensemble (as in Chapter 5). When combined into the optimised

multi-diagnostic ensemble, the 12 member ensemble was more skilful. This therefore indicates a smaller ensemble spread for the individual diagnostics within a multi-diagnostic ensemble is important for optimising operationally in the future. Since the 12-member ECMWF ensemble provides a more skilful forecast than the 51-member ensemble, it could reduce computation costs for turbulence forecasting. Only 12 members would need to be calculated, saving memory space and computational time. It is also worth exploring the impact of ensemble size on a multi-diagnostic multi-model ensemble further to see if there is an optimum number of ensemble members to maximise forecast skill. Also the MOGREPS-G ensemble is designed to be time lagged to create a 24-member ensemble (Met Office, 2017), therefore investigating how well a 24-member MOGREPS-G ensemble performs against the 12-member ensemble would also be worth further study. Not only increasing the ensemble size, but adding a third ensemble would also be worth investigating. The Global Ensemble Forecast System (GEFS) from the National Centers for Environmental Prediction (NCEP) could provide more skill by adding its own strengths and weaknesses, but would need verifying before being made operational.

Chapter 7

Conclusions

Turbulence is a major hazard to aviation, causing injuries to passengers and crew, costing the industry at least millions of dollars every year (Sharman & Lane, 2016). Recent research suggests Clear-Air Turbulence (CAT) will increase in the future with climate change (Williams & Joshi, 2013; Williams, 2017). These studies that investigate the impact of climate change on CAT used climate models, however there was previously a gap in the literature investigating how well climate models predict the location of CAT or how well they forecast the changes of CAT with climate change. The first part of this thesis attempted to answer this question by comparing the global distribution of CAT from the climate models to the location of CAT in reanalysis data. The reanalysis model used in this study was the ERA-Interim reanalysis data set and the climate model was the HadGEM2-ES model with historical greenhouse gas emissions. Both models have different horizontal resolutions and therefore the ERA-Interim data set (which has a higher resolution) was regridded onto the same grid as the climate model. The results showed that the climate model can predict the location of CAT on a global scale, however the resolution does impact the location. The spatial distribution from the climate model was more similar to the reanalysis data on the same grid than the reanalysis data on its native grid. This is important and further work is needed to fully understand how the resolution impacts CAT forecasting. The main result from this study was, there is more uncertainty in choosing which turbulence diagnostic to use than which numerical model was used. This therefore suggests the climate model can forecast the location of CAT.

We continued this study by also comparing the climate change response of the reanalysis data and the climate model data. We split both data sets into two 19 year periods and calculated the percentage change of CAT between the two periods. The main result again indicated there is more uncertainty in which turbulence diagnostic to use rather than which model was used. We also found that the global response in the climate model was much weaker than the

reanalysis data, suggesting that any estimates in a global climate change study could be an underestimate.

The previous climate change research only focused on the North Atlantic, therefore in Chapter 4 we investigate the response of CAT to climate change for the entire world. The study used the same HadGEM2-ES model as in Chapter 3, however the atmospheric forcing used was pre-industrial control (picontrol) and the Intergovernmental Panel on Climate Change (IPCC) Representative Concentration Pathway 8.5 (RCP8.5) (Flato et al., 2013). By comparing two 30 year periods (one 30 year period from picontrol and 2050–2080 from RCP8.5) we can see how the global climate system responds to climate change. 20 turbulence diagnostics were used and all showed an increase in CAT with climate change. By averaging all turbulence diagnostics we showed that CAT would increase in all seasons, at multiple flight levels and across all turbulence strength categories (light to severe). The global average increase in CAT masks some of the regional changes as the response is not uniform around the entire world. Therefore regional increases are also found and in some cases turbulence is set to more than double, with severe turbulence nearly tripling in the North Atlantic. As discussed in Chapter 3, the HadGEM2-ES model underestimates the global response of CAT to climate change and as a result these could also be underestimates. This therefore indicates a more urgent need to improve turbulence forecasting.

Ensemble forecasting has been shown to increase forecast skill (Gill & Buchanan, 2014; Buchanan, 2016), we therefore build on this work by following other areas of meteorology in multi-model ensemble forecasting, such as tropical cyclone forecasting (Krishnamurti et al., 2000; Vitart, 2006; Titley & Stretton, 2016). Here we combine the Met Office Global and Regional Ensemble Prediction System (MOGREPS-G) and the European Centre for Medium Range Weather Forecasting (ECMWF) ensemble prediction system. As an initial test the turbulence diagnostic Ellrod & Knapp (1992) Turbulence Index 1 (Ellrod TI1) was used and found that the ECMWF ensemble was more skilful in producing an ensemble forecast than the MOGREPS-G ensemble, however the multi-model ensemble was more skilful than either of the two single-model ensembles. Although the multi-model ensemble was more skilful, it was not significant at the 95% confidence interval. It was found though, that the relative economic value of the forecast was higher for all cost/loss ratios, which is known as sufficiency (Ehrendorfer & Murphy, 1988). This indicates that for any given user, no mater how much importance they have on maximising hits and minimising false alarms, the multi-model ensemble has greater value.

In addition the multi-model ensemble was as reliable as the single model ensemble. The multi-model ensemble also gives more operational resilience and creating one authoritative forecast whilst maintaining skill and reliability and increasing value, which would be particularly important in operational use in the future by the WAFCs.

Ensemble forecasting is one method being developed for turbulence forecasting. The other is multi-diagnostic forecasting, such as the Graphical Turbulence Guidance System (GTG) with the latest version GTG3 including MWT as well as shear turbulence predictors (Sharman & Pearson, 2017). Both methods have their own benefits to improving turbulence forecasting and therefore we combine them to make a multi-diagnostic multi-model ensemble in Chapter 6. We combine the Ellrod TI1, Brown index (Brown, 1973), mountain wave turbulence predictor (MWT12 from Sharman & Pearson (2017)), Richardson number and convective precipitation accumulation. Before combining them into one ensemble, we first run them as single-diagnostic single-model ensembles and a single-diagnostic multi-model ensemble similar to Chapter 5. The results agree with Chapter 5 that the ECWMF is more skilful than MOGREPS-G and a 12 member ECMWF ensemble is less skilful than the ECMWF 51 member ensemble. We also found that the multi-model ensemble is more skilful than the single model ensembles but not significantly at the 95% confidence interval. We then combined the diagnostics equally for all thresholds as well as optimising them by weighting each threshold and diagnostic to maximise the Area Under the Curve (AUC) of a Relative Operating Characteristic (ROC) plot. After doing this the MOGREPS-G ensemble was more skilful than the ECMWF ensemble, and when optimised, the ECMWF 12 member ensemble was more skilful than the ECMWF 51 member ensemble. This suggests that a smaller spread for the individual diagnostics, when combined, gives a much better spread and therefore forecast skill. It was also found that the multi-diagnostic multi-model ensemble was more skilful than either of the single model ensembles. However, the multi-diagnostic multi-model ensemble containing the ECMWF 12 member ensemble was again more skilful than the one containing the ECMWF 51 member ensemble. This is an important result because it suggests that it is better to have fewer ensemble members when creating a multi-diagnostic multi-model ensemble for turbulence forecasting. This therefore reduces the number of members needed to be processed and stored, saving computational time and money.

During this project we have answered some important questions in aviation turbulence research, such as discovering that CAT will increase with climate change globally in all seasons, flight levels and turbulence strength categories. We have also found that multi-model ensemble forecasting does improve forecast skill, although not significantly, and a smaller ensemble spread in a multi-diagnostic ensemble gives a better forecast skill. However we also raised some questions that would need to be answered in further study. For the reanalysis work, the impact of inter-annual and decadal variability such as the Atlantic Multi-decadal Oscillation (AMO), El Niño Southern Oscillation (ENSO) and North Atlantic Oscillation (NAO) requires further study. Also the new reanalysis data set ERA5 is being produced, which would provide an updated data set with a higher spatial resolution, and data at 3 hourly intervals (Hersbach & Dee, 2016), which would be worth using to investigate how CAT has changed in the climate system over the last 40 years.

The climate change work also needs to be extended to include other Representative Concentration Pathways such as RCP6 and RCP4.5. Also the Coupled Model Intercomparison Project 6 (CMIP6) will be available in a few years, which could provide more climate models that we can run the climate change analysis on, creating a more robust analysis and result. We also know that CAT is changing with climate change, however we do not know how that increase will come about. It is important to understand whether the number of CAT events will increase in the future or whether the same number of turbulence events will occur, but each event will increase in size. Both responses would have different impacts on the industry. For example more turbulence events could increase the number of diversions or actions being taken by pilots and flight planners, but a larger turbulence area would subject the airframe to longer periods of stress. Both would have impacts, and these would need to be investigated.

The ensemble forecasting also has some interesting areas of further study, the main one is investigating if there is an optimum number of ensemble members to maximise forecast skill, because fewer ECMWF ensemble members increased the forecast skill. Therefore, there could be an optimum number of members that produce a suitable forecast skill that, when combined, maximises the number of hits while minimising the number of false alarms. It would also be interesting to see if a time lagged MOGREPS-G ensemble (producing a 24 member ensemble) would provide a higher forecast skill. A third ensemble could also be added to create a three model multi-model ensemble, which could increase the forecast skill further.

Aviation turbulence is likely to be more important in the future, with a changing climate and an ever-increasing demand for air travel. Hopefully this project has answered some fundamental questions around aviation turbulence and climate change and has laid the foundation to introduce a new way of forecasting aviation turbulence.

Appendix A

Turbulence Equations

Negative Richardson number

$$Ri = R_s T_m \left(\frac{\frac{T_l}{p_l^{.286}} - \frac{T_u}{p_u^{.286}}}{\frac{T_m}{p_m^{.286}}} \right) \left(\frac{\frac{dp}{p}}{(du)^2 + (dv)^2} \right) \tag{A.1}$$

Magnitude of vertical shear of horizontal wind

$$\frac{(du)^2 + (dv)^2)^{0.5}}{\frac{R_s}{g} T_m \left(\frac{dp}{p_m} \right)} \tag{A.2}$$

Colson-Panofsky index

$$\left(\frac{3600}{1852} \right)^2 \left((du)^2 + (dv)^2 \right) \left(\frac{1 - Ri}{0.5} \right) \tag{A.3}$$

Frontogenesis function

$$\frac{du}{d\theta} = \frac{du}{T_u \left(\frac{1E5}{p_u} \right)^{.286} - T_l \left(\frac{1E5}{p_l} \right)^{.286}} \tag{A.4}$$

$$\frac{dv}{d\theta} = \frac{dv}{T_u \left(\frac{1E5}{p_u} \right)^{.286} - T_l \left(\frac{1E5}{p_l} \right)^{.286}} \tag{A.5}$$

$$\begin{aligned} \text{Frontogenesis function} = \frac{du}{d\theta} \left(\frac{du}{d\theta} \frac{du}{dx} + \frac{dv}{d\theta} \frac{du}{dy} \right) \\ + \frac{dv}{d\theta} \left(\frac{du}{d\theta} \frac{dv}{dx} + \frac{dv}{d\theta} \frac{dv}{dy} \right) \end{aligned} \tag{A.6}$$

Brown index

$$f = 2\Omega \sin(\phi) \tag{A.7}$$

$$\text{Brown index} = 0.3\left(\frac{dv}{dx} - \frac{du}{dy} + f\right)^2 + \left(\left(\frac{dv}{dx} + \frac{du}{dy}\right)^2 + \left(\frac{du}{dx} - \frac{dv}{dy}\right)^2\right)^{0.5} \tag{A.8}$$

Brown energy dissipation rate

$$\frac{1}{24}\left(0.3\left(\frac{dv}{dx} - \frac{du}{dy} + f\right)^2\right) + \frac{1}{24}\left(\left(\left(\frac{dv}{dx} + \frac{du}{dy}\right)^2 + \left(\frac{du}{dx} - \frac{dv}{dy}\right)^2\right)^{0.5}\left(du^2 + dv^2\right)\right) \tag{A.9}$$

Variant 1 of Ellrod"s turbulence index

$$\left(\frac{\left(du^2 + dv^2\right)^{0.5}}{\frac{R_s}{g}T\frac{dp}{p}}\right)\left(\left(\frac{dv}{dx} + \frac{du}{dy}\right)^2 + \left(\frac{du}{dx} - \frac{dv}{dy}\right)^2\right)^{0.5} \tag{A.10}$$

Variant 2 of Ellrod"s turbulence index

$$\left(\frac{\left(du^2 + dv^2\right)^{0.5}}{\frac{R_s}{g}T\frac{dp}{p}}\right)\left(\left(\left(\frac{dv}{dx} + \frac{du}{dy}\right)^2 + \left(\frac{du}{dx} - \frac{dv}{dy}\right)^2\right)^{0.5} - \frac{du}{dx} - \frac{dv}{dy}\right) \tag{A.11}$$

Flow deformation

$$\left(\left(\frac{dv}{dx} + \frac{du}{dy}\right)^2 + \left(\frac{du}{dx} - \frac{dv}{dy}\right)^2\right)^{0.5} \tag{A.12}$$

Vertical Relative Vorticity squared

$$\left(\frac{dv}{dx} - \frac{du}{dy}\right)^2 \tag{A.13}$$

Horizontal temperature Gradient

$$\left(\frac{dT^2}{dx} + \frac{dT^2}{dy}\right)^{0.5} \tag{A.14}$$

Wind speed

$$\left(u^2 + v^2\right)^{0.5} \tag{A.15}$$

Wind speed times directional shear

$$\left(u^2+v^2\right)^{0.5}\left|\frac{\arctan\left(\frac{v_u}{u_u}\right)-\arctan\left(\frac{v_l}{u_l}\right)}{\frac{R_s}{g}T\frac{dp}{p}}\right| \tag{A.16}$$

Flow deformation times wind speed

$$\left(\left(\frac{dv}{dx}+\frac{du}{dy}\right)^2+\left(\frac{du}{dx}-\frac{dv}{dy}\right)^2\right)^{0.5}\left(u^2+v^2\right)^{0.5} \tag{A.17}$$

Flow deformation times vertical temperature gradient

$$\frac{|dt|}{\frac{R_s}{g}T\frac{dp}{p}}\left(\left(\frac{dv}{dx}+\frac{du}{dy}\right)^2+\left(\frac{du}{dx}-\frac{dv}{dy}\right)^2\right)^{0.5} \tag{A.18}$$

Magnitude of residual of nonlinear balance equation

$$2\left|\frac{du}{dx}\frac{dv}{dy}-\frac{du}{dy}\frac{dv}{dx}\right| \tag{A.19}$$

Magnitude of horizontal divergence

$$\left|\frac{du}{dx}+\frac{dv}{dy}\right| \tag{A.20}$$

Negative absolute vorticity advection

$$\beta=\frac{2\Omega\cos(\phi)}{R_E} \tag{A.21}$$

$$\xi=\frac{dv}{dy}-\frac{du}{dy} \tag{A.22}$$

$$\frac{dQ}{dx}=\frac{d\xi}{dx} \tag{A.23}$$

$$\frac{dQ}{dy}=\frac{d\xi}{dy} \tag{A.24}$$

$$\text{Negative absolute vorticity advection}=-u\frac{dQ}{dx}-v\left(\beta+\frac{dQ}{dy}\right) \tag{A.25}$$

Magnitude of relative vorticity advection

$$\left|u\frac{dQ}{dx}+v\frac{dQ}{dy}\right| \tag{A.26}$$

Version 1 of North Carolina State University index

$$\frac{u\frac{du}{dx} + v\frac{dv}{dy}}{Ri\left(\frac{dQ}{dx}^2 + \frac{dQ}{dy}^2\right)^{0.5}} \tag{A.27}$$

Potential Vorticity

$$-g\zeta_a\frac{d\theta}{dp} \tag{A.28}$$

Bibliography

Atger, F. (1999). The skill of ensemble prediction systems. *Monthly Weather Review*, **127**(9), 1941–1953.

Bedka, K., Brunner, J., Dworak, R., Feltz, W., Otkin, J., & Greenwald, T. (2010). Objective satellite-based detection of overshooting tops using infrared window channel brightness temperature gradients. *Journal of Applied Meteorology and Climatology*, **49**(2), 181–202.

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, **129**(3), 420–436.

Bowler, N. E., Arribas, A., Mylne, K. R., Robertson, K. B., & Beare, S. E. (2008). The MOGREPS short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, **134**(632), 703–722.

Brown, R. (1973). New indices to locate clear-air turbulence. *Meteor. Mag*, **102**, 347–361.

Buchanan, P. (2016). Aviation Turbulence Ensemble Techniques. In *Aviation Turbulence*, pages 285–296. Springer.

Buizza, R. & Palmer, T. N. (1998). Impact of ensemble size on ensemble prediction. *Monthly Weather Review*, **126**(9), 2503–2518.

Chambers, E. (1955). Clear air turbulence and civil jet operations. *The Aeronautical Journal*, **59**(537), 613–628.

Clark, T. L. & Peltier, W. (1977). On the evolution and stability of finite-amplitude mountain waves. *Journal of the Atmospheric Sciences*, **34**(11), 1715–1730.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W., Johns, T., Krinner, G., et al. (2013). Long-term climate change: projections, commitments and irreversibility. Cambridge University Press.

Colson, D. & Panofsky, H. (1965). An index of clear air turbulence. *Quarterly Journal of the Royal Meteorological Society*, **91**(390), 507–513.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137**(656), 553–597.

Delcambre, S. C., Lorenz, D. J., Vimont, D. J., & Martin, J. E. (2013). Diagnosing Northern Hemisphere jet portrayal in 17 CMIP3 global climate models: Twenty-first-century projections. *Journal of Climate*, **26**(14), 4930–4946.

Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., Cooke, W. F., Dixon, K. W., Dunne, J., Dunne, K., et al. (2006). GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *Journal of Climate*, **19**(5), 643–674.

Dima, I. M. & Wallace, J. M. (2003). On the seasonality of the Hadley cell. *Journal of the Atmospheric Sciences*, **60**(12), 1522–1527.

Dutton, J. A. & Panofsky, H. A. (1970). Clear air turbulence: A mystery may be unfolding. *Science*, **167**(3920), 937–944.

Dutton, M. (1980). Probability Forecasts of Clear-Air Turbulence based on Numerical-Model Output. *Meteorological Magazine*, **109**(1299), 293–306.

Ehrendorfer, M. & Murphy, A. H. (1988). Comparative evaluation of weather forecasting systems: Sufficiency, quality, and accuracy. *Monthly Weather Review*, **116**(9), 1757–1770.

Ellrod, G. P. & Knapp, D. I. (1992). An objective clear-air turbulence forecasting technique: Verification and operational use. *Weather and Forecasting*, **7**(1), 150–165.

Epifanio, C. C. & Qian, T. (2008). Wave–turbulence interactions in a breaking mountain wave. *Journal of the Atmospheric Sciences*, **65**(10), 3139–3158.

FAA (2017a). Air Traffic Plans and Publications. https://www.faa.gov/air_traffic/publications (Accessed 09/02/2018).

FAA (2017b). Turbulence: Staying Safe. https://www.faa.gov/travelers/fly_safe/turbulence/ (Accessed 09/02/2018).

Fahey, T. H. (1993). Northwest Airlines atmospheric hazards advisory and avoidance system. In *International Conference on Aviation Weather Systems, 5 th, Vienna, VA*, pages 409–413.

Fels, S., Mahlman, J., Schwarzkopf, M., & Sinclair, R. (1980). Stratospheric sensitivity to perturbations in ozone and carbon dioxide: Radiative and dynamical response. *Journal of the Atmospheric Sciences*, **37**(10), 2265–2297.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W. J., Cox, P., Driouech, F., Emori, S., Eyring, V., et al. (2013). Evaluation of Climate Models. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Climate Change 2013*, **5**, 741–866.

Ford, R. (1994). Gravity wave radiation from vortex trains in rotating shallow water. *Journal of Fluid Mechanics*, **281**, 81–118.

Francis, P. N. & Batstone, C. (2013). *Developing a satellite product to identify severe convective storms hazardous to aviation.* Satellite Applications Technical Memo 11.

Fritts, D. C., Isler, J. R., & Andreassen, Ø. (1994). Gravity wave breaking in two and three dimensions: 2. Three-dimensional evolution and instability structure. *Journal of Geophysical Research: Atmospheres*, **99**(D4), 8109–8123.

Fritts, D. C., Garten, J. F., & Andreassen, Ø. (1996). Wave breaking and transition to turbulence in stratified shear flows. *Journal of the Atmospheric Sciences*, **53**(8), 1057–1085.

Gill, P. G. (2014). Objective verification of World Area Forecast Centre clear air turbulence forecasts. *Meteorological Applications*, **21**(1), 3–11.

Gill, P. G. (2016). Aviation Turbulence Forecast Verification. In *Aviation Turbulence*, pages 261–283. Springer.

Gill, P. G. & Buchanan, P. (2014). An ensemble based turbulence forecasting system. *Meteorological Applications*, **21**(1), 12–19.

Gill, P. G. & Stirling, A. J. (2013). Including convection in global turbulence forecasts. *Meteorological Applications*, **20**(1), 107–114.

Gnanadesikan, A., Dixon, K. W., Griffies, S. M., Balaji, V., Barreiro, M., Beesley, J. A., Cooke, W. F., Delworth, T. L., Gerdes, R., Harrison, M. J., et al. (2006). GFDL's CM2 global coupled climate models. Part II: The baseline ocean simulation. *Journal of climate*, **19**(5), 675–697.

Haverdings, H. & Chan, P. W. (2010). Quick access recorder data analysis software for windshear and turbulence studies. *Journal of Aircraft*, **47**(4), 1443–1447.

Hersbach & Dee (2016). ERA5 reanalysis is in production. ECMWF Newsletter. https://www.ecmwf.int/en/newsletter/147/news/era5-reanalysis-production.

Hoskins, B. J., McIntyre, M., & Robertson, A. W. (1985). On the use and significance of isentropic potential vorticity maps. *Quarterly Journal of the Royal Meteorological Society*, **111**(470), 877–946.

ICAO (2012). Guidance on the Harmonized WAFS Grids for Cumulonimbus Cloud, Icing and Turbulence Forecasts. https://www.icao.int/safety/meteorology/WAFSOPSG/Guidance Material/Forms/AllItems.aspx.

ICAO (2016). Guidance on the Harmonized WAFS Grids for Cumulonimbus Cloud, Icing and Turbulence forecasts. https://www.icao.int/airnavigation/METP/MOGWAFS Reference Documents/WAFS_HazardGridUserGuide.pdf.

Jaeger, E. & Sprenger, M. (2007). A Northern Hemispheric climatology of indices for clear air turbulence in the tropopause region derived from ERA40 reanalysis data. *Journal of Geophysical Research: Atmospheres*, **112**(D20).

Jolliffe, I. T. & Stephenson, D. B. (2012). *Forecast verification: a practitioner's guide in atmospheric science*. John Wiley & Sons.

Jones, C., Hughes, J., Bellouin, N., Hardiman, S., Jones, G., Knight, J., Liddicoat, S., O'Connor, F., Andres, R. J., Bell, C., et al. (2011). The HadGEM2-ES implementation of CMIP5 centennial simulations. *Geoscientific Model Development*, **4**(3), 543–570.

Kane, T., Brown, B., & Bruintjes, R. (1998). Characteristics of pilot reports of icing. In *Preprints: 14th Conference on Probability and Statistics*, pages 11–16.

Kauffmann, P. (2002). The business case for turbulence sensing systems in the US air transport sector. *Journal of Air Transport Management*, **8**(2), 99–107.

Kim, J.-H. & Chun, H.-Y. (2011). Statistics and possible sources of aviation turbulence over South Korea. *Journal of Applied Meteorology and Climatology*, **50**(2), 311–324.

Kim, J.-H., Chan, W. N., Sridhar, B., & Sharman, R. D. (2015). Combined winds and turbulence prediction system for automated air-traffic management applications. *Journal of Applied Meteorology and Climatology*, **54**(4), 766–784.

Kim, J.-H., Chan, W. N., Sridhar, B., Sharman, R. D., Williams, P. D., & Strahan, M. (2016). Impact of the North Atlantic Oscillation on transatlantic flight routes and clear-air turbulence. *Journal of Applied Meteorology and Climatology*, **55**(3), 763–771.

Kim, J.-H., Sharman, R., Strahan, M., Scheck, J. W., Bartholomew, C., Cheung, J. C., Buchanan, P., & Gait, N. (2018). Improvements in Non-Convective Aviation Turbulence Prediction for the World Area Forecast System (WAFS). *Bulletin of the American Meteorological Society*, (2018).

Kim, S.-H., Chun, H.-Y., & Chan, P. W. (2017). Comparison of Turbulence Indicators Obtained from In Situ Flight Data. *Journal of Applied Meteorology and Climatology*, **56**(6), 1609–1623.

Knox, J. A. (1997). Possible mechanisms of clear-air turbulence in strongly anticyclonic flows. *Monthly Weather Review*, **125**(6), 1251–1259.

Knox, J. A., McCann, D. W., & Williams, P. D. (2008). Application of the Lighthill–Ford theory of spontaneous imbalance to clear-air turbulence forecasting. *Journal of the Atmospheric Sciences*, **65**(10), 3292–3304.

Krishnamurti, T. N., Kishtawal, C., Zhang, Z., LaRow, T., Bachiochi, D., Williford, E., Gadgil, S., & Surendran, S. (2000). Multimodel ensemble forecasts for weather and seasonal climate. *Journal of Climate*, **13**(23), 4196–4216.

Lane, T. P. & Sharman, R. D. (2008). Some influences of background flow conditions on the generation of turbulence due to gravity wave breaking above deep convection. *Journal of Applied Meteorology and Climatology*, **47**(11), 2777–2796.

Lane, T. P., Sharman, R. D., Clark, T. L., & Hsu, H.-M. (2003). An investigation of turbulence generation mechanisms above deep convection. *Journal of the Atmospheric Sciences*, **60**(10), 1297–1321.

Lane, T. P., Sharman, R. D., Trier, S. B., Fovell, R. G., & Williams, J. K. (2012). Recent advances in the understanding of near-cloud turbulence. *Bulletin of the American Meteorological Society*, **93**(4), 499–515.

Lighthill, M. J. (1952). On sound generated aerodynamically. I. General theory. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 211, pages 564–587.

MacCready Jr, P. B. (1964). Standardization of gustiness values from aircraft. *Journal of Applied Meteorology*, **3**(4), 439–449.

Mancuso, R. & Endlich, R. (1966). Clear air turbulence frequency as a function of wind shear and deformation. *Monthly Weather Review*, **94**(9), 581–585.

Marlton, G. J. (2016). *On the development, characterisation and applications of a balloon-borne atmospheric turbulence sensor*. Ph.D. thesis, University of Reading.

McCann, D. W., Knox, J. A., & Williams, P. D. (2012). An improvement in clear-air turbulence forecasting based on spontaneous imbalance theory: the ULTURB algorithm. *Meteorological Applications*, **19**(1), 71–78.

Mecikalski, J. R., Berendes, T. A., Feltz, W. F., Bedka, K. M., Bedka, S. T., Murray, J. J., Wimmers, A. J., Minnis, P., Johnson, D. B., Haggerty, J., et al. (2007). Aviation applications for satellite-based observations of cloud properties, convection initiation, in-flight icing, turbulence, and volcanic ash. *Bulletin of the American Meteorological Society*, **88**(10), 1589–1607.

Meneguz, E., Wells, H., & Turp, D. (2016). An automated system to quantify aircraft encounters with convectively induced turbulence over Europe and the Northeast Atlantic. *Journal of Applied Meteorology and Climatology*, **55**(5), 1077–1089.

Met Office (2010 - 2013). *Iris: A Python library for analysing and visualising meteorological and oceanographic data sets*. Exeter, Devon, v1.2 edition.

Met Office (2017). MOGREPS-G guide to data. https://www.metoffice.gov.uk/services/data-provision/big-data-drive/wholesale/categories/mogrepsg-user-guide (Accessed 17/08/2018).

Met Office (2018). WAFC London Performance Indicators. https://www.metoffice.gov.uk/public/weather/aviation-wafc/ (Accessed 08/08/2018).

Miles, J. W. & Howard, L. N. (1964). Note on a heterogeneous shear flow. *Journal of Fluid Mechanics*, **20**(2), 331–336.

Molteni, F., Buizza, R., Palmer, T. N., & Petroliagis, T. (1996). The ECMWF ensemble prediction system: Methodology and validation. *Quarterly journal of the Royal Meteorological Society*, **122**(529), 73–119.

Nastrom, G. D. & Fritts, D. C. (1992). Sources of mesoscale variability of gravity waves. Part I: Topographic excitation. *Journal of the Atmospheric Sciences*, **49**(2), 101–110.

Nunez, R. (2018). Integrating GOES-16 Satellite into Convective Porosity Determination at CWSU Houston. In *Sixth Aviation, Range, and Aerospace Meteorology Special Symposium, Austin, Tx., American Meteorological Society*.

Owens, R. G. & Hewson, T. D. (2018). ECMWF Forecast User Guide. Reading: ECMWF. doi: 10.21957/m1cs7h.

Park, Y.-Y., Buizza, R., & Leutbecher, M. (2008). TIGGE: Preliminary results on comparing and combining ensembles. *Quarterly Journal of the Royal Meteorological Society*, **134**(637), 2029–2050.

Price, C. & Rind, D. (1994). Possible implications of global climate change on global lightning distributions and frequencies. *Journal of Geophysical Research: Atmospheres*, **99**(D5), 10823–10831.

Reeve, N. & Toumi, R. (1999). Lightning activity as an indicator of climate change. *Quarterly Journal of the Royal Meteorological Society*, **125**(555), 893–903.

Richardson, D. S. (2000). Skill and relative economic value of the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, **126**(563), 649–667.

Richardson, D. S. (2001). Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quarterly Journal of the Royal Meteorological Society*, **127**(577), 2473–2489.

Roach, W. & Dixon, R. (1970). A note on the paper 'on the influence of synoptic development on the production of high level turbulence'. *Quarterly Journal of the Royal Meteorological Society*, **96**(410), 758–760.

Schmetz, J., Tjemkes, S., Gube, M., & Van de Berg, L. (1997). Monitoring deep convection and convective overshooting with METEOSAT. *Advances in Space Research*, **19**(3), 433–441.

Schwartz, B. (1996). The quantitative use of PIREPs in developing aviation weather guidance products. *Weather and Forecasting*, **11**(3), 372–384.

Search Technology (2000). A pilot-centered turbulence assessment and monitoring system, TAMS, Phase II Final Report. Norcross, GA, 2000.

Shapiro, M. (1976). The role of turbulent heat flux in the generation of potential vorticity in the vicinity of upper-level jet stream systems. *Monthly Weather Review*, **104**(7), 892–906.

Shapiro, M. (1978). Further evidence of the mesoscale and turbulent structure of upper level jet stream–frontal zone systems. *Monthly Weather Review*, **106**(8), 1100–1111.

Shapiro, M. (1980). Turbulent mixing within tropopause folds as a mechanism for the exchange of chemical constituents between the stratosphere and troposphere. *Journal of the Atmospheric Sciences*, **37**(5), 994–1004.

Sharman, R. & Lane, T. (2016). *Aviation Turbulence: Processes, Detection, Prediction*. Springer.

Sharman, R. & Pearson, J. (2017). Prediction of energy dissipation rates for aviation turbulence. Part I: Forecasting nonconvective turbulence. *Journal of Applied Meteorology and Climatology*, **56**(2), 317–337.

Sharman, R., Tebaldi, C., Wiener, G., & Wolff, J. (2006). An integrated approach to mid-and upper-level turbulence forecasting. *Weather and Forecasting*, **21**(3), 268–287.

Sharman, R., Trier, S., Lane, T., & Doyle, J. (2012). Sources and dynamics of turbulence in the upper troposphere and lower stratosphere: A review. *Geophysical Research Letters*, **39**(12).

Sharman, R., Cornman, L., Meymaris, G., Pearson, J., & Farrar, T. (2014). Description and derived climatologies of automated in situ eddy-dissipation-rate reports of atmospheric turbulence. *Journal of Applied Meteorology and Climatology*, **53**(6), 1416–1432.

Shutts, G. (1998). Stationary gravity-wave structure in flows with directional wind shear. *Quarterly Journal of the Royal Meteorological Society*, **124**(549), 1421–1442.

Shutts, G. & Gadian, A. (1999). Numerical simulations of orographic gravity waves in flows which back with height. *Quarterly Journal of the Royal Meteorological Society*, **125**(559), 2743–2765.

Storer, L., Williams, P., & Joshi, M. (2017). Global Response of Clear-Air Turbulence to Climate Change. *Geophysical Research Letters*, **44**(19), 9976–9984.

Storer, L. N., Williams, P. D., & Gill, P. G. (2018). Aviation turbulence: Dynamics, forecasting, and response to climate change. *Pure and Applied Geophysics*, pages 1–15. In Press.

Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T. M., Hewson, T. D., Keller, J. H., Matsueda, M., Methven, J., Pappenberger, F., et al. (2016). The TIGGE project and its achievements. *Bulletin of the American Meteorological Society*, **97**(1), 49–67.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, **93**(4), 485–498.

Tebaldi, C., Nychka, D., Brown, B. G., & Sharman, R. (2002). Flexible discriminant techniques for forecasting clear-air turbulence. *Environmetrics*, **13**, 859–878.

Tenenbaum, J. (1991). Jet stream winds: Comparisons of analyses with independent aircraft data over southwest Asia. *Weather and Forecasting*, **6**(3), 320–336.

Titley, H. & Stretton, R. (2016). Tropical Cyclone Ensemble Forecasting at the Met Office: Upgrades to the MOGREPS Model and TC Products, and an Evaluation of the Benefit of Multi-model Ensembles . In *Preprints: 32nd Conference on Hurricanes and Tropical Meteorology*.

Trier, S. B. & Sharman, R. D. (2009). Convection-permitting simulations of the environment supporting widespread turbulence within the upper-level outflow of a mesoscale convective system. *Monthly Weather Review*, **137**(6), 1972–1990.

Trier, S. B., Sharman, R. D., & Lane, T. P. (2012). Influences of moist convection on a cold-season outbreak of clear-air turbulence (CAT). *Monthly Weather Review*, **140**(8), 2477–2496.

Truscott, B. (2000). EUMETNET AMDAR AAA AMDAR Software Developments—Technical Specification. *Doc. Ref. E_AMDAR/TSC/003. Met Office: Exeter, UK*.

Turner, J. (1999). Development of a mountain wave turbulence prediction scheme for civil aviation. *Met Office Forecasting Research Technical Report*, (265).

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., et al. (2011). The representative concentration pathways: an overview. *Climatic Change*, **109**(1-2), 5.

Vitart, F. (2006). Seasonal forecasting of tropical storm frequency using a multi-model ensemble. *Quarterly Journal of the Royal Meteorological Society*, **132**(615), 647–666.

Vrancken, P., Wirth, M., Ehret, G., Barny, H., Rondeau, P., & Veerman, H. (2016). Airborne forward-pointing UV Rayleigh lidar for remote clear air turbulence detection: system design and performance. *Applied Optics*, **55**(32), 9314–9328.

Walters, D., Wood, N., Vosper, S., & Milton, S. (2014). ENDGame: A new dynamical core for seamless atmospheric prediction. *Met Office documentation. can be consulted at http://www. metoffice. gov. uk/media/pdf/s/h/ENDGameGOVSci v2. 0. pdf*.

Williams, P. D. (2016). Transatlantic flight times and climate change. *Environmental Research Letters*, **11**(2), 024008.

Williams, P. D. (2017). Increased light, moderate, and severe clear-air turbulence in response to climate change. *Advances in Atmospheric Sciences*, **34**(5), 576–586.

Williams, P. D. & Joshi, M. J. (2013). Intensification of winter transatlantic aviation turbulence in response to climate change. *Nature Climate Change*, **3**, 644–648.

Williams, P. D., Read, P., & Haine, T. (2003). Spontaneous generation and impact of inertia-gravity waves in a stratified, two-layer shear flow. *Geophysical Research Letters*, **30**(24).

Williams, P. D., Haine, T. W., & Read, P. L. (2005). On the generation mechanisms of short-scale unbalanced modes in rotating two-layer flows with vertical shear. *Journal of Fluid Mechanics*, **528**, 1–22.

Williams, P. D., Haine, T. W., & Read, P. L. (2008). Inertia–gravity waves emitted from balanced flow: Observations, properties, and consequences. *Journal of the Atmospheric Sciences*, **65**(11), 3543–3556.

Wimmers, A., Griffin, S. M., Bachmeier, A. S., Gerth, J., & Lindstrom, S. (2018). Resolving Gravity Waves with Himawari-8 and GOES-16 Imagery at the New Limit of Resolution and the Application to Aircraft-Scale Turbulence. In *Sixth Aviation, Range, and Aerospace Meteorology Special Symposium, Austin, Tx., American Meteorological Society*.

WMO (2003). Aircraft Meteorological Data Relay (AMDAR) Reference Manual.

Wolff, J. & Sharman, R. (2008). Climatology of upper-level turbulence over the contiguous united states. *Journal of Applied Meteorology and Climatology*, **47**(8), 2198–2214.

Wurtele, M., Sharman, R., & Datta, A. (1996). Atmospheric lee waves. *Annual Review of Fluid Mechanics*, **28**(1), 429–476.

Ziehmann, C. (2000). Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models. *Tellus A*, **52**(3), 280–299.