

RESEARCH ARTICLE

Can a climate model successfully diagnose clear-air turbulence and its response to climate change?

Paul D. Williams^{ORCID} | Luke N. Storer

Department of Meteorology, University of Reading, Reading, UK

CorrespondenceP.D. Williams, Department of Meteorology, University of Reading, Earley Gate, Reading, RG6 6ET, UK
Email: p.d.williams@reading.ac.uk**Funding information**

Natural Environment Research Council, Grant/Award Number: NE/L002566

Abstract: Atmospheric turbulence causes the majority of weather-related aircraft accidents. Climate models project large increases in clear-air turbulence as the jet streams become more sheared in response to climate change. However, climate models have coarser resolutions than the numerical weather prediction models that are used to forecast clear-air turbulence operationally, raising questions about their suitability for this purpose. Here we provide the first rigorous demonstration that climate models are capable of successfully diagnosing clear-air turbulence and its response to climate change. We use an ensemble of seven clear-air turbulence diagnostics to compare 38 years of historic turbulence diagnosed from climate model simulations and high-resolution reanalysis data. We find that the differences in turbulence between the climate model and reanalysis data are much smaller than the spread between the diagnostics. When using a climate model to calculate the probabilities (and their temporal trends) of encountering clear-air turbulence of any strength, at any flight cruising level, and in any season, we find that most of the uncertainty stems from the turbulence diagnostics rather than the climate model. These results confirm the suitability of climate models for the task of producing future clear-air turbulence projections. The turbulence increases are generally larger when diagnosed from the reanalysis data than the climate model, suggesting that previous quantifications from climate models of the response of clear-air turbulence to climate change may be underestimates. Our results show that the key to reducing uncertainty in projections of future clear-air turbulence lies in improving the clear-air turbulence diagnostics rather than the climate models.

KEYWORDS

aviation, clear-air turbulence, climate change, climate models, jet streams, reanalysis data, uncertainty

1 | INTRODUCTION

Atmospheric turbulence is a significant hazard to the aviation sector, causing 71% of all weather-related aircraft accidents experienced by regularly scheduled air carriers

(Gultepe *et al.*, 2019). Turbulence adversely impacts both passenger safety and airline economics (Golding, 2002), and it also contributes to the fear of flying for nervous fliers (Sharman *et al.*, 2012). It has been estimated that the US air transport sector experiences around 1,000

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

turbulence events annually, injuring approximately 1,500 flight attendants and 560 passengers (Lindsey, 2000). Other studies broadly corroborate these injury rates, estimating at least 790 annual turbulence events for US air carriers, of which 568 are injury-producing events, resulting in 687 minor flight attendant injuries, 38 serious flight attendant injuries, 120 minor passenger injuries, and 17 serious passenger injuries (Kauffmann, 2002). Officially published turbulence injury rates are known to be underestimates, because reporting to the authorities is mandatory only for serious injuries – such as those requiring hospitalisation for more than 48 hr – and is optional for minor injuries (Sharman *et al.*, 2006). Injured flight attendants are often temporarily unable to work following a turbulence encounter, and their recuperation time leads to an estimated 7,000 lost working days annually for a typical airline (Sharman and Lane, 2016).

The main physical categories of aviation turbulence are reviewed by Storer *et al.*, (2019b). They include convectively induced turbulence (CIT; Uccellini and Koch, 1987; Koch and Dorian, 1988), mountain wave turbulence (MWT; Lilly, 1978), and clear-air turbulence (CAT; Chambers, 1955; Endlich, 1964; Dutton and Panofsky, 1970; Knox, 1997). CAT, which is defined as high-altitude aircraft bumpiness in regions devoid of significant cloudiness and away from thunderstorm activity, is the focus of the present study. About half of all passenger aircraft encounters with turbulence occur in cloudless skies (Cowen, 1998). An important source of CAT is vertical wind shear: changes in wind speed with altitude can create Kelvin–Helmholtz shear instabilities, which generate growing wave modes that eventually overturn and create turbulent eddies. In some cases, the shear instabilities are initiated by loss of atmospheric balance and the consequent generation of gravity waves (Williams *et al.*, 2003; 2005; 2008), which locally modify the shear and stratification (Palmer *et al.*, 1986; Dunkerton, 1997). In addition to its impacts on the aviation sector, shear-induced CAT in the upper troposphere and lower stratosphere has also been implicated in the break-up and subsequent crash landing of the Challenger Space Shuttle shortly after lift-off in 1986 (Uccellini *et al.*, 1986).

Although the atmosphere exhibits variability on a wide range of time-scales and length-scales (Williams *et al.*, 2017), aircraft respond only to a subset of turbulent eddies on length-scales of around 100 m to 1 km (Sharman *et al.*, 2006). These length-scales are usually smaller than the numerical grid resolution of atmospheric models (except in high-resolution case-studies; e.g., Lane *et al.*, 2012). Therefore, it is usually not possible to explicitly simulate aircraft-affecting turbulence. Instead, this subgrid-scale turbulence is diagnosed from the resolved flow. Some

CAT diagnostics that are widely used for this purpose include the Ellrod and Knapp (1992) indices, the Colson and Panofsky (1965) index, the Brown (1973) index, and the Lighthill–Ford index (Knox *et al.*, 2008; McCann *et al.*, 2012).

The turbulence diagnostics that are used to forecast CAT operationally are demonstrably skilful (Sharman *et al.*, 2006). However, they each have individual strengths and weaknesses, and no single diagnostic predictor is capable of forecasting all CAT events. For example, a diagnostic that is optimised for calculating CAT generated by unbalanced flow may miss CAT generated by frontogenesis. For this reason, Sharman *et al.*, (2006) combined several CAT diagnostics to create the Graphical Turbulence Guidance (GTG) system, which is used operationally. The individual CAT diagnostics contributing to GTG are optimally weighted when combined, to maximise the deterministic forecast skill. GTG has now been updated to forecast MWT in addition to CAT (Sharman and Pearson, 2017). Recently, probabilistic forecasts of aviation turbulence, generated using ensemble prediction systems, have been tested as a potentially more informative alternative to traditional deterministic forecasts (Gill and Buchanan, 2014; Storer *et al.*, 2019a; 2020).

Several recent studies have used an ensemble of turbulence diagnostics to understand how CAT has increased historically according to reanalysis datasets (Jaeger and Sprenger, 2007) and how it will increase in future according to climate change models (e.g., Williams and Joshi, 2013; Storer *et al.*, 2017; Williams, 2017). In particular, Williams and Joshi (2013) and Williams (2017) used a basket of 21 diagnostics to analyse future CAT in the North Atlantic flight corridor in winter at 200 hPa at the time of CO₂ doubling compared to pre-industrial concentrations. They found on average 59% more light CAT, 94% more moderate CAT, and 149% more severe CAT. Storer *et al.*, (2017) extended these studies to analyse future CAT globally in all seasons and at various altitudes, finding that some midlatitude regions are projected to experience statistically significant increases of several hundred per cent more turbulence by the period 2050–2080. For example, the North Atlantic, North America, and Europe are each projected to experience at least a doubling of severe CAT at 200 hPa.

Various direct and indirect lines of observational evidence suggest that CAT has already started to increase (Jaeger and Sprenger, 2007; Wolff and Sharman, 2008; Lee *et al.*, 2019). The CAT increases are consistent with the strengthened vertical wind shear in the midlatitude upper troposphere and lower stratosphere that is projected by climate models (Delcambre *et al.*, 2013; Lv *et al.*, 2021). Indeed, the North Atlantic jet stream has already become 15% more sheared since satellites began observing it in

1979, according to three different reanalysis datasets (Lee *et al.*, 2019). Other effects of climate change on aviation include modified flight routes and journey times resulting from changes to the jet stream wind speeds (Irvine *et al.*, 2016; Williams, 2016) and increased take-off weight restrictions resulting from hotter air on the runway (Coffel and Horton, 2015; Gratton *et al.*, 2020). Gratton *et al.*, (2022) gives a recent review.

It is known that different climate models project a quantitatively similar CAT response to climate change (Storer *et al.*, 2017). This finding provides some reassurance that the projected CAT increases are robust. However, it has not yet been rigorously demonstrated that climate models are actually capable of successfully diagnosing CAT and its response to climate change. Specifically, the grid resolution of the simulated atmosphere is much coarser in climate models than in the numerical weather prediction models that are used to forecast CAT operationally. The relatively coarse grid may introduce a bias into the diagnosis of CAT (which is one of several unresolved processes in climate models; Williams, 2005). This possibility raises questions about the future CAT increases that are projected by climate models (although the underlying physical mechanism responsible for the increases is supported by other independent lines of evidence; Lee *et al.*, 2019).

The present study aims to confront the above issues, by comparing historic CAT diagnosed from climate model simulations and high-resolution reanalysis data. While some previous studies (Jaeger and Sprenger, 2007) have analysed CAT using reanalysis data, and other studies (Williams and Joshi, 2013; Storer *et al.*, 2017; Williams, 2017) have analysed CAT using climate models, what is currently missing is a direct quantitative comparison of historic CAT in reanalysis data and climate models. Such a comparison is essential to verify the suitability of climate models for CAT diagnosis and to demonstrate confidence in the future projections, and it is the main novelty of the present study. Comparisons between climate models and reanalysis data typically focus on comparing the large-scale features. However, successfully capturing the large scales does not guarantee that climate models correctly capture the scales near the numerical grid spacing, which are the relevant scales for diagnosing CAT.

The layout of the present study is as follows. Section 2 describes the methodology, including the details of the climate model simulations and reanalysis data used in the comparison. It also describes the CAT diagnostics computed, and how they are calibrated to various turbulence strength categories ranging from light to severe. Section 3 presents the results, first comparing the historic climatologies of the global CAT distribution between the climate model and reanalysis data, and then comparing the

response to historic climate change over four decades. Finally, Section 4 concludes the paper with a summary and discussion.

2 | METHODOLOGY

To diagnose historic CAT from a climate model, we use 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 is the same climate model used by Storer *et al.*, (2017), so the findings of the present study directly aid in the interpretation of previous results. The model has a horizontal grid spacing of 1.25° in latitude and 1.875° in longitude, giving an array of 192×144 grid boxes globally. We use 38 years (1968–2005) from the historical climate simulation, which is driven by previous climate forcings including CO_2 .

To diagnose historic CAT from reanalysis data, we use ERA-Interim from the European Centre for Medium-Range Weather Forecasts (ECMWF; Dee *et al.*, 2011). ERA-Interim has a horizontal grid spacing of 0.7° in both latitude and longitude, giving an array of 512×256 grid boxes globally. This resolution is substantially higher than the HadGEM2-ES climate model and is approaching (but does not reach) the typical resolutions of around 0.25° that are used to produce operational CAT forecasts (Sharman and Pearson, 2017). We use 38 years (1979–2016) from the reanalysis data. Note that this is not exactly the same period as the climate model (although it is the same duration) because ERA-Interim does not start until 1979 and the HadGEM2-ES historical simulation ends in 2005. Therefore, to make the time series as long as possible for comparison, the time periods are slightly offset. The CAT climatologies and trends over 38 years will be relatively insensitive to this offset, such that most of the difference between the climate model and reanalysis will be attributable to underlying differences between HadGEM2-ES and ERA-Interim rather than the offset.

As a sensitivity test to analyse whether different diagnoses of CAT between HadGEM2-ES and ERA-Interim are attributable to their different grid resolutions or to other model differences, part of our analysis uses Iris (Met Office, 2013) to re-grid the ERA-Interim data through linear interpolation to have the same horizontal grid and vertical levels as HadGEM2-ES before calculating the turbulence diagnostics. Therefore, when we present our results in the next section, the inclusion of this sensitivity analysis leads to a three-way comparison between “climate model”, “reanalysis”, and “reanalysis (climate grid)”.

From the 21 CAT diagnostics that have been used in previous climate change studies (Williams and Joshi, 2013;

Storer *et al.*, 2017; Williams, 2017), we choose to analyse a subset of seven whose diagnostic abilities are supported by strong observational evidence. Although dozens of CAT diagnostics have been proposed, when they are verified using turbulence data from aircraft only a subset are typically found to be skilful enough to be included in operational forecasting systems. The seven diagnostics used in the present study fall into this subset: six of them are used in the operational GTG forecasting system and are skilful (as measured by the area under the receiver operating characteristic curve) at diagnosing turbulence encounters reported by pilots (Sharman *et al.*, 2006). We include the wind speed multiplied by the directional wind shear as a seventh diagnostic, because aircraft observations have shown that CAT is associated with both large wind speeds and large directional wind shears (Endlich, 1964), so it is logical to take the product of the two as a CAT diagnostic. The use of this subset of diagnostics ensures that our results are as reliable and trustworthy as possible. The seven CAT diagnostics used here are the following. The Richardson number diagnoses regions of Kelvin–Helmholtz shear instability (e.g., Endlich, 1964; Dutton and Panofsky, 1970) and is defined by

$$Ri = -\frac{1}{\rho\theta} \frac{\partial\theta/\partial p}{(\partial\mathbf{u}/\partial p)^2}, \quad (1)$$

where ρ is density, p is pressure, θ is potential temperature, and $\mathbf{u} = (u, y)$ is the horizontal velocity vector. Because small values of Ri are associated with turbulence, here we use $-Ri$ as the diagnostic, so that larger (i.e., more positive) values are associated with turbulence. The Colson–Panofsky index has its origins in the turbulent kinetic energy (TKE) balance equation (Colson and Panofsky, 1965) and is defined by:

$$CP = (\Delta z)^2 \left(\frac{\partial\mathbf{u}}{\partial z} \right)^2 \left(1 - \frac{Ri}{Ri_{crit}} \right), \quad (2)$$

where z is altitude, Δz is the vertical grid spacing, and $Ri_{crit} = 0.5$ is a critical Richardson number. The frontogenesis function diagnoses regions where the magnitude of the horizontal potential temperature gradient increases following the flow (Sharman *et al.*, 2006) and is defined by:

$$F_\theta = \frac{D}{Dt} \left(\left| \frac{\partial\mathbf{u}}{\partial\theta} \right|^2 \right), \quad (3)$$

where t is time, D/Dt denotes the Lagrangian time derivative, and the partial derivative is taken at fixed horizontal position using potential temperature as a vertical coordinate. Ellrod's turbulence index (variant 1) is derived from a simplified frontogenesis function (Ellrod and Knapp,

1992) and is defined by:

$$TI1 = \left| \frac{\partial\mathbf{u}}{\partial z} \right| \sqrt{\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right)^2 + \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \right)^2}, \quad (4)$$

where x and y are longitude and latitude, respectively. The wind speed multiplied by the directional wind shear diagnoses regions of both high wind speed and large directional wind shear (Endlich, 1964) and is defined by:

$$\sqrt{u^2 + v^2} \left| \frac{\partial\psi}{\partial z} \right|, \quad (5)$$

where ψ is the horizontal wind direction. The magnitude of the residual of the nonlinear balance equation diagnoses regions of strong imbalance (e.g., Knox *et al.*, 2008; McCann *et al.*, 2012) and is defined by:

$$UBF = \left| -\nabla^2\Phi + 2J(u, v) + f\zeta - \beta u \right|, \quad (6)$$

where ∇^2 is the Laplacian operator, Φ is the geopotential, $J()$ is the Jacobian operator, f and β are the Coriolis parameter and its latitudinal derivative, respectively, and ζ is the vertical component of the vorticity. The North Carolina State University index (version 1) was inspired by a study of severe turbulence encounters (Kaplan *et al.*, 2005) and is defined by:

$$NCSU1 = \frac{1}{\max(Ri, 10^{-5})} \max \left(u \frac{\partial u}{\partial x} + v \frac{\partial v}{\partial y}, 0 \right) |\nabla\zeta|, \quad (7)$$

where $\max()$ is the maximum operator and ∇ is the horizontal gradient operator.

We compute the seven CAT diagnostics from both HadGEM2-ES and ERA-Interim using snapshots (not time averages) of the zonal and meridional wind and temperature fields every six hours over the 38 years. We compute the CAT diagnostics at two pressure altitudes: 200 hPa (corresponding to approximately 12 km or 39,000 ft) and 250 hPa (corresponding to approximately 10 km or 34,000 ft). The horizontal and vertical spatial derivatives are calculated using centred second-order finite differences. For each CAT diagnostic, we compute the probability of exceeding each of five turbulence strength thresholds every six hours, following the methodology of Williams (2017) and Storer *et al.*, (2017). The strength categories correspond to what would be experienced as light, light-to-moderate, moderate, moderate-to-severe, and severe turbulence on a large passenger aircraft. The thresholds are derived with reference to the cube-rooted eddy dissipation rate, which is proportional to the vertical acceleration of an aircraft experiencing turbulence (MacCready, 1964). The specific thresholds

for each diagnostic are dependent on the resolution and other characteristics of the input temperature and wind data, and therefore separate thresholds are calculated for the climate model and reanalysis data (and re-gridded reanalysis data). In each case, the global, annual probability distribution for each CAT diagnostic is computed using six-hourly data from the first 20 years of each dataset. The usage of 20 years for the calibration period is consistent with previous work (Williams and Joshi, 2013; Williams, 2017) but is arbitrary and the results are insensitive to this choice. Then the diagnostic thresholds are calculated by taking the top 0.1% (99.9–100%) of the probability distribution to be severe turbulence; the next 0.1% (99.8–99.9%) to be moderate-to-severe turbulence; the next 0.2% (99.6–99.8%) to be moderate turbulence; the next 0.5% (99.1–99.6%) to be light-to-moderate turbulence; and the next 2.1% (97–99.1%) to be light turbulence. These percentages are consistent with the log-normal distribution of the cube-rooted eddy dissipation rate, as measured by aircraft (Williams, 2017). Separate thresholds are calculated at 200 and 250 hPa. Having calculated the thresholds, a turbulence event occurs any time a diagnostic value exceeds a threshold.

3 | RESULTS

3.1 | Climatologies of the global CAT distribution

The global spatial distribution of historic CAT in December, January, and February (DJF) at 200 hPa is shown in Figure 1. Each map shows the probability of encountering light CAT at each point on the globe, as diagnosed from the full 38 years of each dataset. The corresponding maps for light-to-moderate, moderate, moderate-to-severe, and severe CAT (not shown) indicate similar spatial patterns but reduced probabilities, as expected, and similar results are obtained at 250 hPa (not shown). For most of the seven diagnostics, visual inspection of Figure 1 indicates good broad agreement (with absolute differences under 8%) of the spatial patterns and magnitudes between the probabilities diagnosed from the climate model and reanalysis data (whether re-gridded or not). The differences are larger for the Colson–Panofsky index and frontogenesis function, especially in the Tropics in the Indian Ocean and West Pacific Ocean. For some diagnostics, the meridional structure exhibits a single peak in the Tropics. The negative Richardson number falls into this category, because of the weak stratification in the Tropics at this altitude. For other diagnostics, the meridional structure exhibits two peaks in the midlatitudes of the Northern and Southern Hemispheres. Variant 1 of Ellrod's turbulence index falls into

this category, because of the large flow deformation and vertical wind shear in the vicinity of the jet streams.

To investigate the strength of the annual cycle, the corresponding analysis for June, July, and August (JJA) is shown in Figure 2. The spatial patterns and magnitudes of the probabilities are broadly similar to those for the opposite season (Figure 1). For the diagnostics whose meridional structure exhibits a single peak in the Tropics, such as the Colson–Panofsky index, the tropical peak generally moves from south of the Equator in DJF to north of the Equator in JJA, because the Hadley circulation moves towards the summer hemisphere (Dima and Wallace, 2003). For the diagnostics whose meridional structure exhibits two midlatitude peaks, such as variant 1 of Ellrod's turbulence index, the largest peak is typically in the Northern Hemisphere in DJF and in the Southern Hemisphere in JJA, because the subpolar jet streams are stronger in the winter hemisphere. The residual of the nonlinear balance equation behaves differently, being stronger in the Northern Hemisphere in both seasons and increasing in magnitude in the Northern Hemisphere from DJF to JJA. Unexpectedly, the frontogenesis function does not clearly exhibit the expected midlatitude peaks, especially in the climate model, perhaps because the resolution is too coarse to resolve frontogenesis. Similar results are found for the transitional seasons of March, April, and May (MAM) and September, October, and November (SON), with the peaks in intermediate positions, as expected (not shown).

Because there are seven different CAT diagnostics each applied to three different datasets, we have an ensemble of 21 different estimates of the global spatial distribution of the probability of encountering light CAT at 200 hPa in DJF (Figure 1) and JJA (Figure 2). We wish to quantify whether the main source of uncertainty within each 21-member ensemble stems from the use of multiple diagnostics or multiple datasets. In other words, are the main differences between the 21 maps in Figures 1 and 2 along the rows or down the columns? To answer this question, for each season we take each of the seven CAT diagnostics in turn, and at each latitude and longitude we compute the standard deviation of the probabilities across the three datasets, before averaging them globally to produce a global-mean standard deviation. This procedure yields a quantification of the global-mean inter-dataset uncertainty for each diagnostic. We also take each of the three datasets in turn, and at each latitude and longitude we compute the standard deviation of the probabilities across the seven CAT diagnostics, before again averaging them globally. This procedure yields a quantification of the global-mean inter-diagnostic uncertainty for each dataset.

The results of this analysis are shown in Table 1. Consistent with visual inspection of Figures 1 and 2, the

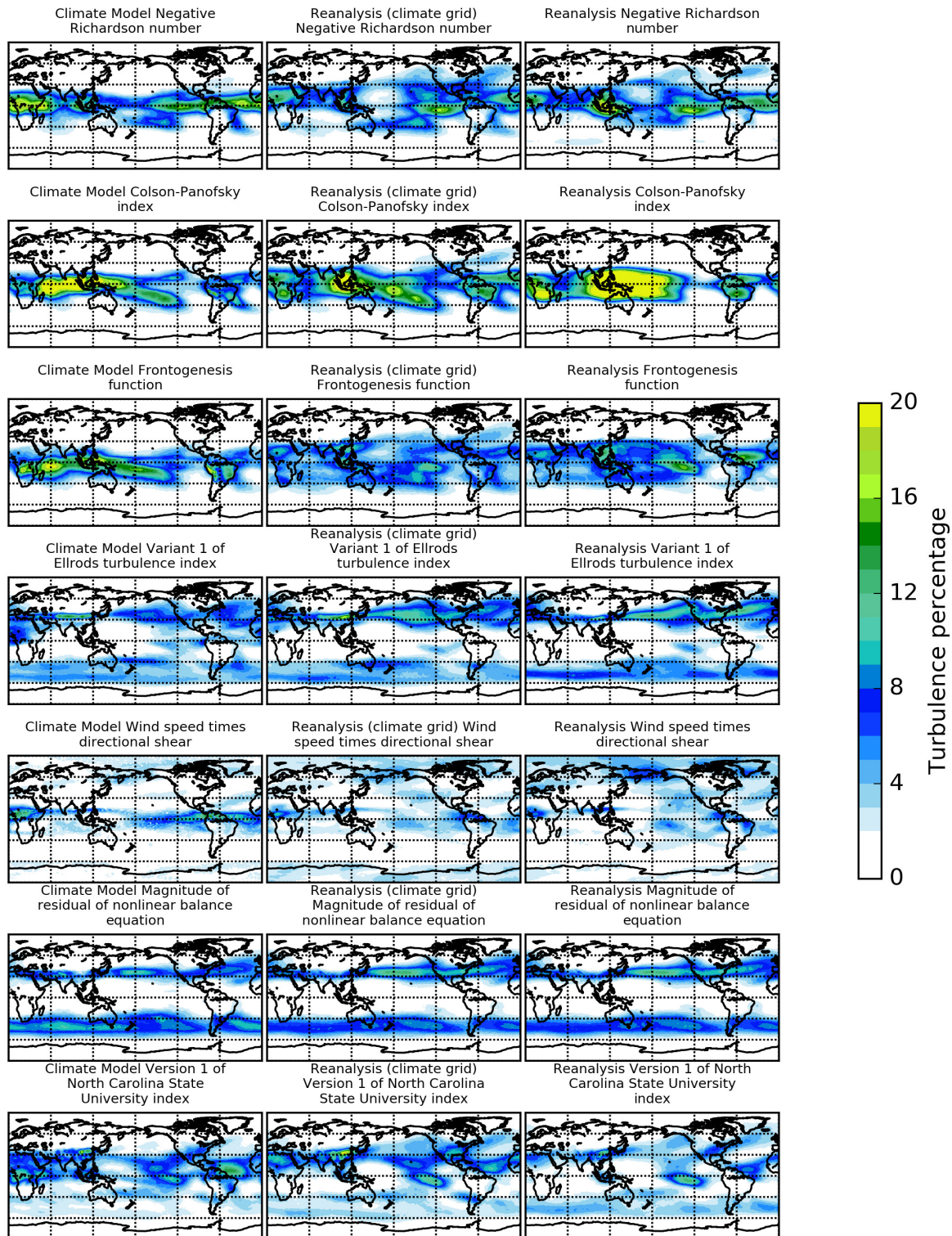


FIGURE 1 Global maps of the probability (%) of encountering light CAT at 200 hPa in DJF. Seven turbulence diagnostics are shown (one per row). The turbulence probabilities are calculated from 38 years of the HadGEM2-ES historical climate simulation (left column), 38 years of ERA-Interim reanalysis data after re-gridding to have the same resolution as HadGEM2-ES (middle column), and 38 years of ERA-Interim reanalysis data at its original resolution (right column) [Colour figure can be viewed at wileyonlinelibrary.com]

variability across the datasets for each diagnostic is typically much smaller than the variability across the diagnostics for each dataset. The majority of the uncertainty

within the 21-member ensemble therefore stems from the use of multiple diagnostics rather than the use of multiple datasets. In other words, the spread between the

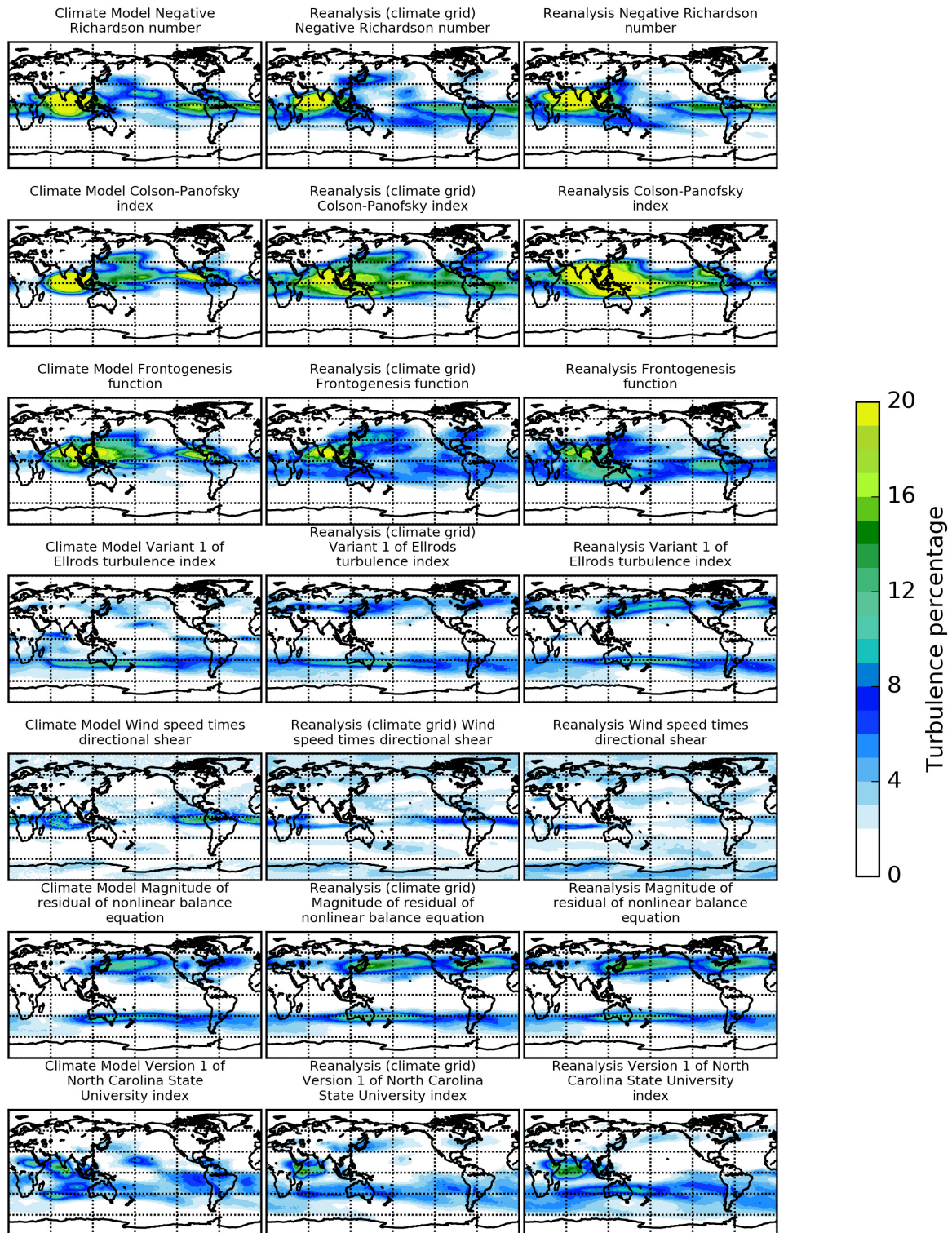


FIGURE 2 As Figure 1, but for JJA [Colour figure can be viewed at wileyonlinelibrary.com]

climate model and reanalysis data is much smaller than the spread between the diagnostics. The two reanalysis datasets are not strictly independent, because one is a coarse re-gridding of the other, but ensuring strict independence by excluding the reanalysis on the climate grid

from the standard deviation calculations in Table 1 does not change the overall findings. Close inspection of Table 1 reveals that there is generally slightly less uncertainty in the transitional seasons (MAM and SON) than the extremal seasons (DJF and JJA). There is generally slightly

TABLE 1 Breakdown of uncertainty sources for the probability of encountering light CAT at 200 hPa

Turbulence diagnostic	DJF	MAM	JJA	SON
Negative Richardson number	0.68%	0.61%	0.69%	0.50%
Colson–Panofsky index	1.00%	0.96%	1.06%	0.87%
Frontogenesis function	0.75%	0.62%	0.78%	0.63%
Variant 1 of Ellrod’s turbulence index	0.41%	0.36%	0.48%	0.33%
Wind speed × directional shear	0.38%	0.30%	0.40%	0.32%
Magnitude of residual of nonlinear balance equation	0.22%	0.20%	0.27%	0.24%
Version 1 of North Carolina State University index	0.53%	0.48%	0.50%	0.44%
Climate model	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%

Note: For each season (columns), the global-mean inter-dataset standard deviations (i.e., the variability across the three datasets for each diagnostic) are shown in the top seven rows, and the global-mean inter-diagnostic standard deviations (i.e., the variability across the seven diagnostics for each dataset) are shown in the bottom three rows.

TABLE 2 For 21 different estimates of the probability of encountering CAT, calculated from three datasets using seven diagnostics, this table shows the ratio of the average global-mean inter-diagnostic standard deviation to the average global-mean inter-dataset standard deviation, for each season, pressure altitude, and strength category

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

less agreement between the diagnostics for the reanalysis data at its original resolution than for either the climate model or the reanalysis data interpolated onto the climate model grid, suggesting that the use of higher-resolution input fields increases the inter-diagnostic uncertainty. The best agreement between the climate model and reanalysis data evidently occurs for the purely dynamical diagnostics (i.e., Ellrod’s turbulence index, wind speed × directional shear, and nonlinear balance equation residual), whereas the worst agreement occurs for the diagnostics explicitly involving a thermodynamic component (i.e., Richardson number, Colson–Panofsky index, frontogenesis function, and North Carolina State University index). The inter-diagnostic uncertainty would be reduced if only the dynamical (or only the thermodynamic) diagnostics were included in the ensemble, but we have included both categories to fairly reflect the diversity of CAT generation mechanisms.

To summarise how much larger the inter-diagnostic variability is than the inter-dataset variability, we

calculate the ratio of the global-mean inter-diagnostic standard deviation averaged over the three datasets (i.e., averaged over the bottom three rows in Table 1) to the global-mean inter-dataset standard deviation averaged over the seven diagnostics (i.e., averaged over the top seven rows in Table 1). These ratios are shown in Table 2 for each season, pressure altitude, and strength category. In all cases, the inter-diagnostic variability is around 2–3 times larger than the inter-dataset variability. Therefore, when using a climate model to calculate the probability of encountering turbulence of any strength, at any flight cruising level, and in any season, most of the uncertainty stems from the turbulence diagnostics rather than the climate model. The ratio of inter-diagnostic to inter-dataset variability generally decreases as the turbulence strength increases, such that the inter-diagnostic variability is typically nearly three times as large as the inter-dataset variability for light turbulence, but only around twice as large for severe turbulence. Nevertheless, diagnostic uncertainty dominates over model

uncertainty, for all turbulence strengths from light to severe.

3.2 | Response to climate change

We next perform a similar analysis for long-term trends in the global spatial distribution of historic CAT. After dividing the 38 years for each dataset into two equal periods of 19 years each, we count the number of turbulence events (i.e., threshold exceedances) at each latitude and longitude in each period, and we then find the percentage change from the first period to the second. The climate will have changed between the two periods, potentially increasing the frequency of occurrence of turbulence. However, given the relatively short time periods under consideration, we expect these increases to be smaller than those that have been found to result from doubling the CO₂ concentration (e.g., 149% more severe CAT over the North Atlantic at 200 hPa in winter; Williams, 2017) or from comparing the period 2050–2080 to pre-industrial conditions (e.g., 181.4% more severe CAT over the North Atlantic at 200 hPa annually; Storer *et al.*, 2017). Nevertheless, we note that other studies have previously identified and analysed CAT trends over relatively short periods (e.g., 44 years; Jaeger and Sprenger, 2007).

The global spatial distribution of the change in historic CAT between the two periods at 200 hPa in DJF and JJA is shown in Figures 3 and 4, respectively. Each map shows the relative (not absolute) change over time in the probability of encountering light CAT at each point on the globe, as diagnosed by comparing the first and second halves of each dataset. The corresponding maps at 250 hPa, and for the other turbulence strengths, are similar (not shown). For each of the seven diagnostics, visual inspection of Figures 3 and 4 indicates lower levels of agreement between the climate model and reanalysis data than was the case for the basic climatological maps in Figures 1 and 2. In all cases, there is a global-mean increase in turbulence over time as the climate changes, consistent with previous studies (Jaeger and Sprenger, 2007; Williams and Joshi, 2013; Storer *et al.*, 2017; Williams, 2017). However, the magnitude of the increase tends to be larger in the reanalysis data than the climate model. For each of the seven diagnostics, the two reanalysis maps (both on the original reanalysis grid and after re-gridding to match the climate model grid) are remarkably similar, indicating that the cause of the stronger turbulence increase in the reanalysis dataset compared to the climate model is not the finer grid resolution.

Having taken 38 years of historic data and computed an ensemble of 21 different estimates of the changing probability of encountering light CAT at 200 hPa in DJF

(Figure 3) and JJA (Figure 4), we wish to quantify whether the main source of uncertainty in the estimates stems from the use of multiple diagnostics or multiple datasets. We follow the same approach taken in Section 3.1. The global-mean inter-dataset uncertainty for each diagnostic and global-mean inter-diagnostic uncertainty for each dataset are shown in Table 3. The variability across the datasets for each diagnostic is again typically smaller than the variability across the diagnostics for each dataset, but only slightly smaller this time (compared to the corresponding breakdown in Table 1) because of the pronounced difference between the climate model and reanalysis data. The best agreement between the climate model and reanalysis data evidently occurs for the North Carolina State University index and the worst agreement occurs for the Colson–Panofsky index, consistent with visual inspection of Figures 3 and 4. There is no strong seasonality.

The ratio of the inter-diagnostic variability to the inter-dataset variability globally is summarised in Table 4 for each season, pressure altitude, and strength category. In all cases, the inter-diagnostic variability is around 40–60% larger than the inter-dataset variability. Therefore, when using a climate model to calculate the impacts of climate change on the probability of encountering turbulence of any strength, at any flight cruising level, and in any season, most of the uncertainty stems from the turbulence diagnostics rather than the climate model. The ratios are systematically smaller in Table 4 than in Table 2, however. The ratio of inter-diagnostic to inter-dataset variability generally slightly increases as the turbulence strength increases from light to severe, and also as the altitude increases from a pressure level of 250 to 200 hPa.

The turbulence percentage changes presented in Figures 3 and 4 are quantified in Table 5, averaged over the globe, the year, and the seven CAT diagnostics. Each entry in the table is positive, indicating more turbulence in each strength category, at each altitude, and in both the climate model and reanalysis data. The percentage increases in the reanalysis data (whether re-gridded or not) are typically 4–5 times as large as those in the climate model. There is no obvious systematic difference between the changes in the two reanalysis datasets (both on the original reanalysis grid and after re-gridding to match the climate model grid), as noted previously based on visual inspection of Figures 3 and 4, but now confirmed quantitatively. The relative increases are systematically larger at 200 hPa than 250 hPa in all three datasets, consistent with previous findings (Storer *et al.*, 2017). The relative increases also generally increase when moving through the turbulence strength categories from light to severe in the reanalysis data, but this trend appears to be absent (or even opposite) in the climate model.

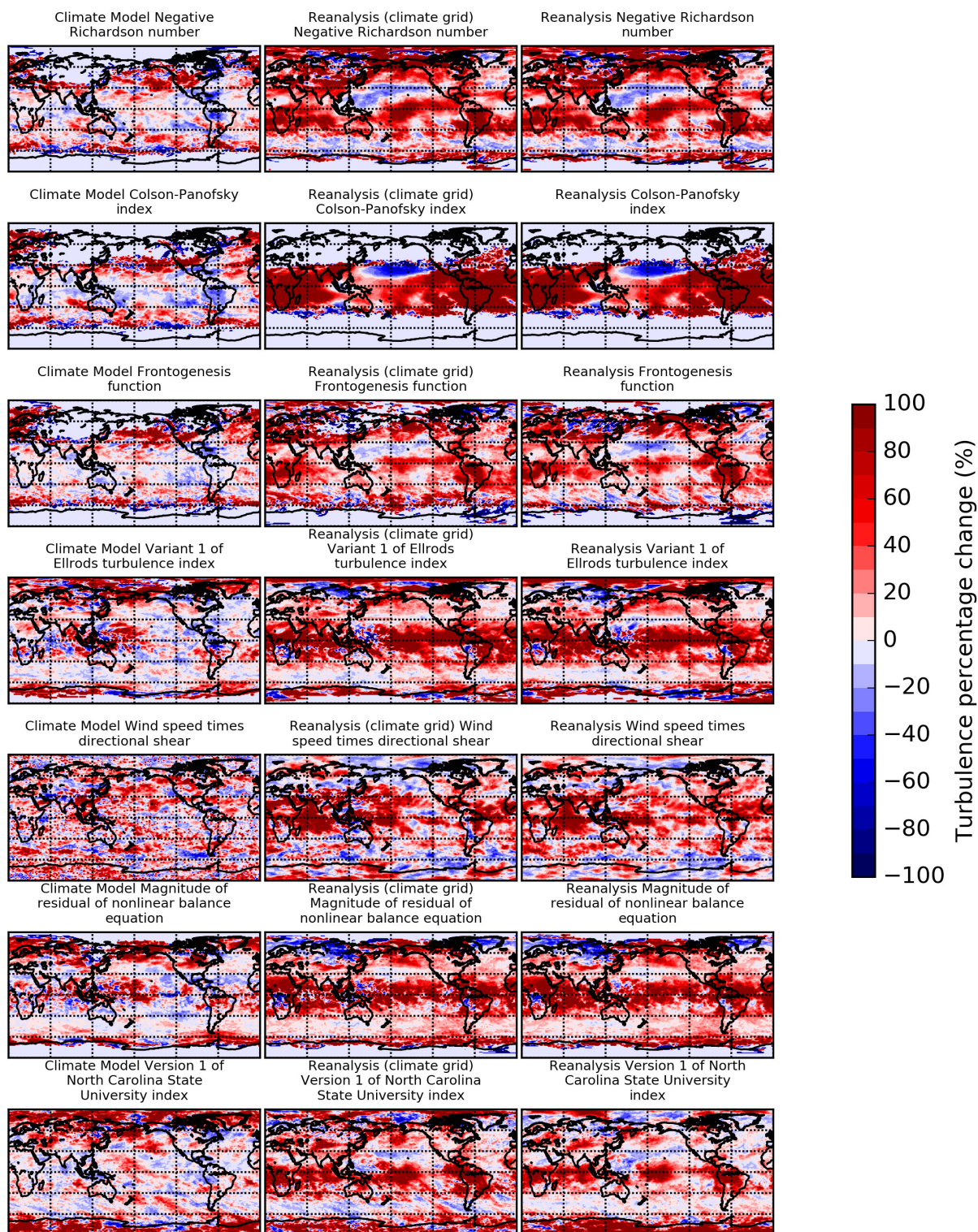


FIGURE 3 Global maps of the change over time in the probability (%) of encountering light CAT at 200 hPa in DJF. Seven turbulence diagnostics are shown (one per row). The changes are calculated from 38 years of the HadGEM2-ES historical climate simulation (left column), 38 years of ERA-Interim reanalysis data after re-gridding to have the same resolution as HadGEM2-ES (middle column), and 38 years of ERA-Interim reanalysis data at its original resolution (right column). In each case, the change in probability is calculated by comparing the second half of the dataset to the first half, and is expressed as a relative (not absolute) change [Colour figure can be viewed at wileyonlinelibrary.com]

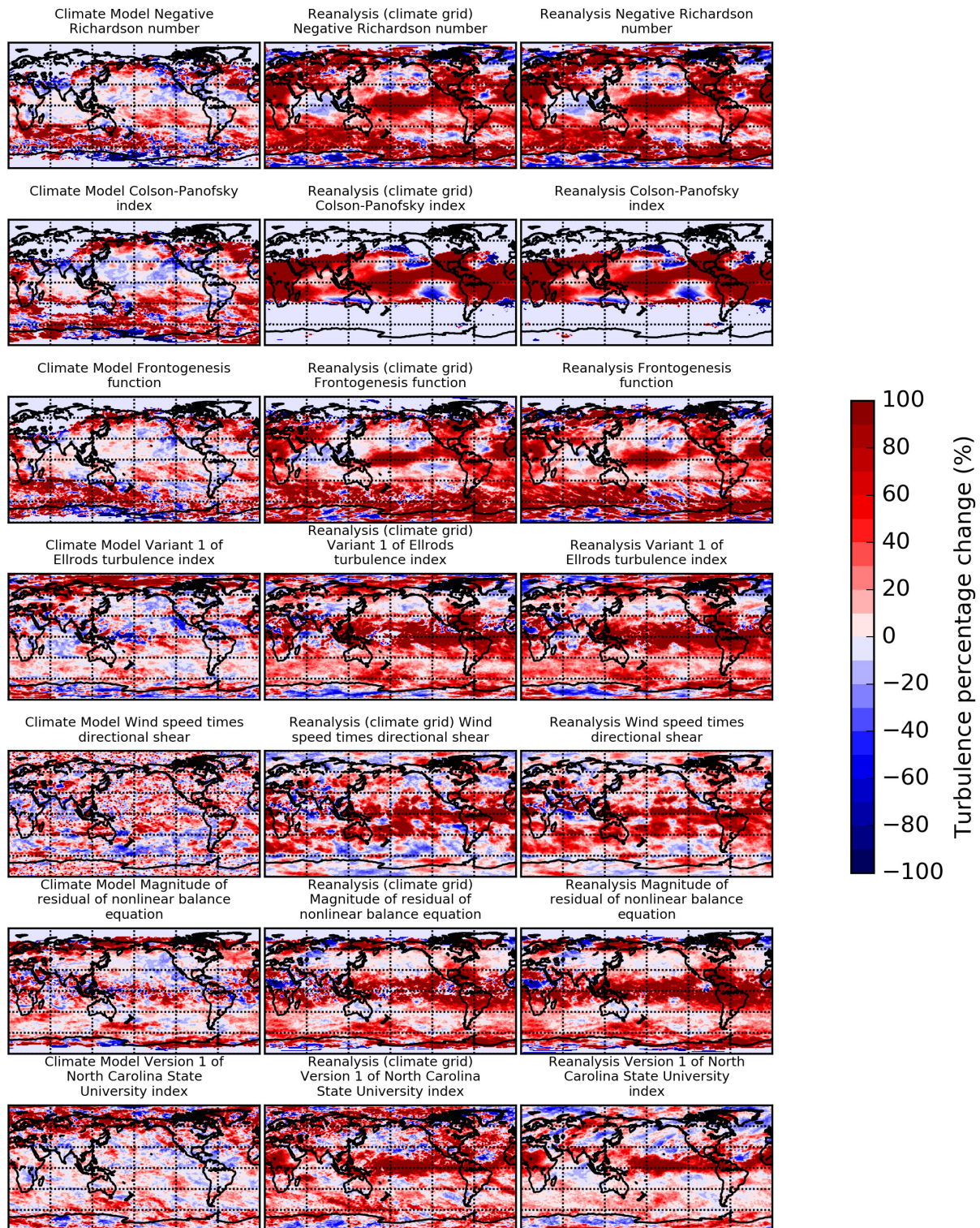


FIGURE 4 As Figure 3, but for JJA [Colour figure can be viewed at wileyonlinelibrary.com]

4 | SUMMARY AND DISCUSSION

This study has provided the first rigorous answer to the question of whether climate models are capable of successfully diagnosing CAT and its response to climate change.

We have found relatively good agreement between historic CAT diagnosed from climate model simulations and high-resolution reanalysis data, where “relatively good agreement” means that the climate model and reanalysis data are closer together than the diagnostics are. This

TABLE 3 Breakdown of uncertainty sources for the change over time in the probability of encountering light CAT at 200 hPa

Turbulence diagnostic	DJF	MAM	JJA	SON
Negative Richardson number	20.92%	20.63%	23.33%	26.49%
Colson–Panofsky index	32.12%	43.91%	38.47%	46.82%
Frontogenesis function	23.19%	27.59%	30.00%	30.16%
Variant 1 of Ellrod’s turbulence index	25.10%	30.16%	29.75%	32.46%
Wind speed × directional shear	27.29%	21.88%	27.34%	22.65%
Magnitude of residual of nonlinear balance equation	23.61%	27.17%	32.57%	29.20%
Version 1 of North Carolina State University index	15.38%	14.49%	16.63%	15.07%
Climate model	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%

Note: For each season (columns), the global-mean inter-dataset standard deviations (i.e., the variability across the three datasets for each diagnostic) are shown in the top seven rows, and the global-mean inter-diagnostic standard deviations (i.e., the variability across the seven diagnostics for each dataset) are shown in the bottom three rows.

TABLE 4 For 21 different estimates of the change over time in the probability of encountering CAT, calculated from three datasets using seven diagnostics, this table shows the ratio of the average global-mean inter-diagnostic standard deviation to the average global-mean inter-dataset standard deviation, for each season, pressure altitude, and strength category

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 5 Global-mean annual-mean change over time in the probability of encountering five strengths of CAT at two pressure levels

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%

Note: The changes are averaged over all seven CAT diagnostics. The changes are calculated from 38 years of the HadGEM2-ES historical climate simulation (“climate model”), 38 years of ERA-Interim reanalysis data after re-gridding to have the same resolution as HadGEM2-ES (“reanalysis (climate grid)”), and 38 years of ERA-Interim reanalysis data at its original resolution (“reanalysis”). The change in probability is calculated by comparing the second half of the dataset to the first half, and is expressed as a relative (not absolute) change.

is a stringent test for a climate model, because to pass it the model has to capture both the planetary-scale and grid-scale features correctly. Our analysis proceeded by calculating seven CAT diagnostics from three datasets

every six hours over 38 years, to analyse five turbulence strengths at two pressure altitudes in four seasons. The spread between the climate model and reanalysis data was generally found to be much smaller than the spread

between the diagnostics. Therefore, when using a climate model to calculate the probabilities (and their temporal trends) of encountering turbulence of any strength, at any flight cruising level, and in any season, most of the uncertainty stems from the turbulence diagnostics rather than the climate model. It is well established that climate models project increases in CAT in response to climate change, and the present study confirms the suitability of climate models for this task.

One of our main findings is that the turbulence increases are generally larger when diagnosed from the reanalysis data than the climate model. This result suggests that previous quantifications from climate models of the response of CAT to climate change (Williams, 2017; Williams and Joshi, 2013; Storer *et al.*, 2017) may be underestimates. The time offset between the climate model (1968–2005) and reanalysis data (1979–2016) could account for part of these differences if the changes in CAT are nonlinear in time. The weaker global response to climate change in the climate model than the reanalysis data stems partly (but not fully) from the climate model incorrectly capturing the changes in the Tropics (assuming the reanalysis can be regarded as truth). However, the main area of interest for CAT is within and around the mid-latitude jet streams, so this apparent model bias may be largely inconsequential. On the other hand, we note that there are considerable inconsistencies in Figures 3 and 4 between the climate model and reanalysis data in the mid-latitudes, so biases beyond the Tropics may nevertheless be impacting the resulting projections.

Our results show that the limiting factor that is preventing the reduction of uncertainty in projections of future CAT is not the climate models, but the CAT diagnostics. Therefore, to reduce these uncertainties, further research is needed to improve and refine the diagnostics. The development of new CAT diagnostics is ongoing and welcome, but may actually increase inter-diagnostic uncertainty if the new diagnostic lies outside the spread of the existing ensemble. On the other hand, increased inter-diagnostic spread may lead to better reliability in operational CAT forecasts. What would perhaps be more useful is further insights into the circumstances in which each diagnostic does (and does not) add useful information to the diagnosis of CAT, so that a diagnostic may be down-weighted or eliminated from the ensemble on those occasions when it is merely adding noise. Gaining these insights will likely require further improvements to our fundamental understanding of the sources and dynamics of CAT.

Repeating the present study using additional climate models and reanalysis datasets should be a priority for future work. However, given the stark differences in spread between the diagnostics and gridded datasets found in

our study, it seems unlikely that our qualitative findings and conclusions will be affected by the inclusion of new datasets. Furthermore, it has been shown that different climate models behave similarly in terms of their quantitative response of CAT to climate change (Storer *et al.*, 2017), providing some reassurance that our findings are likely to be robust and that our conclusions are unlikely to be modified by the inclusion of additional climate models. Separately analysing the inter-diagnostic uncertainty arising from the dynamical and thermodynamic diagnostics should also be a priority, as should using other comparison metrics to evaluate the CAT diagnostics and the climate models. Other avenues for future research include an assessment of the response of the other forms of turbulence apart from CAT (i.e., CIT and MWT) to climate change. In particular, it is plausible that CIT could evolve as climate change increases deep convection, particularly in the Tropics (Price and Rind, 1994; Reeve and Toumi, 1999). Also currently lacking is an understanding of the impacts on CAT of some of the modes of natural variability, such as the El Niño Southern Oscillation (ENSO). For example, it is possible (but currently unproven) that the observed spike in pilots encountering turbulence over the USA in winter 1997–1998 was caused by a strong El Niño event (Wolff and Sharman, 2008). We call for future work to explore these areas.


AUTHOR CONTRIBUTIONS

Paul D. Williams: conceptualization; methodology; project administration; software; supervision; writing – original draft; writing – review and editing. **Luke N. Storer:** formal analysis; investigation; software; visualization.

ACKNOWLEDGEMENT

LNS acknowledges support through a PhD studentship from the UK Natural Environment Research Council (NERC) SCENARIO Doctoral Training Partnership (reference NE/L002566).

ORCID

Paul D. Williams  <https://orcid.org/0000-0002-9713-9820>

REFERENCES

- Brown, R. (1973) New indices to locate clear-air turbulence. *Meteorological Magazine*, 102, 347–361.
- Chambers, E. (1955) Clear-air turbulence and civil jet operations. *The Aeronautical Journal*, 59, 613–628.
- Coffel, E. and Horton, R. (2015) Climate change and the impact of extreme temperatures on aviation. *Weather, Climate, and Society*, 7, 94–102.

- Colson, D. and Panofsky, H. (1965) An index of clear-air turbulence. *Quarterly Journal of the Royal Meteorological Society*, 91, 507–513.
- Cowen, R. (1998) Clearing the air about turbulence: a fearful flier's foray. *Weatherwise*, 51, 24–28.
- Dee, D.P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597.
- Delcambre, S.C., Lorenz, D.J., Vimont, D.J. and Martin, J.E. (2013) Diagnosing Northern Hemisphere jet portrayal in 17 CMIP3 global climate models: twenty-first-century projections. *Journal of Climate*, 26, 4930–4946.
- Dima, I.M. and Wallace, J.M. (2003) On the seasonality of the Hadley cell. *Journal of the Atmospheric Sciences*, 60, 1522–1527.
- Dunkerton, T.J. (1997) Shear instability of internal inertia–gravity waves. *Journal of the Atmospheric Sciences*, 54, 1628–1641.
- Dutton, J.A. and Panofsky, H.A. (1970) Clear-air turbulence: a mystery may be unfolding. *Science*, 167, 937–944.
- Ellrod, G.P. and Knapp, D.I. (1992) An objective clear-air turbulence forecasting technique: verification and operational use. *Weather and Forecasting*, 7, 150–165.
- Endlich, R.M. (1964) The mesoscale structure of some regions of clear-air turbulence. *Journal of Applied Meteorology*, 3, 261–276.
- Gill, P.G. and Buchanan, P. (2014) An ensemble-based turbulence forecasting system. *Meteorological Applications*, 21, 12–19.
- Golding, W. (2002) Turbulence and its impact on commercial aviation. *Journal of Aviation/Aerospace Education and Research*, 11, 19–29.
- Gratton, G., Padhra, A., Rapsomanikis, S. and Williams, P.D. (2020) The impacts of climate change on Greek airports. *Climatic Change*, 160, 219–231.
- Gratton, G.B., Williams, P.D., Padhra, A. and Rapsomanikis, S. (2022) Reviewing the impacts of climate change on air transport operations. *The Aeronautical Journal*, 126, 209–221.
- Gultepe, I., Sharman, R., Williams, P.D., Zhou, B., Ellrod, G., Minnis, P., Trier, S., Griffin, S., Yum, S.S., Gharabaghi, B., Feltz, W., Temimi, M., Pu, Z., Storer, L.N., Kneringer, P., Weston, M.J., Chuang, H.-y., Thobois, L., Dimri, A.P., Dietz, S.J., Franca, G.B., Almeida, M.V. and Neto, F.L.A. (2019) A review of high-impact weather for aviation meteorology. *Pure and Applied Geophysics*, 176, 1869–1921.
- Irvine, E.A., Shine, K.P. and Stringer, M.A. (2016) What are the implications of climate change for trans-Atlantic aircraft routing and flight time?. *Transportation Research Part D: Transport and Environment*, 47, 44–53.
- Jaeger, E. and 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). <https://doi.org/10.1029/2006JD008189>.
- Jones, C.D., Hughes, J.K., Bellouin, N., Hardiman, S.C., Jones, G.S., Knight, J., Liddicoat, S., O'Connor, F.M., Andres, R.J., Bell, C., Boo, K.-O., Bozzo, A., Butchart, N., Cadule, P., Corbin, K.D., Doutriaux-Boucher, M., Friedlingstein, P., Gornall, J., Gray, L.J., Halloran, P.R., Hurtt, G., Ingram, W.J., Lamarque, J.-F., Law, R.M., Meinshausen, M., Osprey, S., Palin, E.J., Parsons Chini, L., Raddatz, T., Sanderson, M.G., Sellar, A.A., Schurer, A., Valdes, P., Wood, N., Woodward, S., Yoshioka, M. and Zerroukat, M. (2011) The HadGEM2-ES implementation of CMIP5 centennial simulations. *Geoscientific Model Development*, 4, 543–570.
- Kaplan, M., Huffman, A., Lux, K., Charney, J., Riordan, A. and Lin, Y.-L. (2005) Characterizing the severe turbulence environments associated with commercial aviation accidents. Part 1: a 44-case study synoptic observational analyses. *Meteorology and Atmospheric Physics*, 88, 129–153.
- Kauffmann, P. (2002) The business case for turbulence sensing systems in the US air transport sector. *Journal of Air Transport Management*, 8, 99–107.
- Knox, J.A. (1997) Possible mechanisms of clear-air turbulence in strongly anticyclonic flows. *Monthly Weather Review*, 125, 1251–1259.
- Knox, J.A., McCann, D.W. and Williams, P.D. (2008) Application of the Lighthill–Ford theory of spontaneous imbalance to clear-air turbulence forecasting. *Journal of the Atmospheric Sciences*, 65, 3292–3304.
- Koch, S.E. and Dorian, P.B. (1988) A mesoscale gravity wave event observed during CCOPE. Part III: wave environment and probable source mechanisms. *Monthly Weather Review*, 116, 2570–2592.
- Lane, T.P., Sharman, R.D., Trier, S.B., Fovell, R.G. and Williams, J.K. (2012) Recent advances in the understanding of near-cloud turbulence. *Bulletin of the American Meteorological Society*, 93, 499–515.
- Lee, S.H., Williams, P.D. and Frame, T.H.A. (2019) Increased shear in the North Atlantic upper-level jet stream over the past four decades. *Nature*, 572, 639–642.
- Lilly, D.K. (1978) A severe downslope windstorm and aircraft turbulence event induced by a mountain wave. *Journal of the Atmospheric Sciences*, 35, 59–77.
- Lindsey, C. (2000) *A baseline of turbulence impacts on commercial air carrier operations*. Bellevue, WA: Northwest Research Associates. Report No. NWRA-CR-00-R210, prepared for Honeywell Commercial Electronics Systems.
- Lv, Y., Guo, J., Li, J., Han, Y., Xu, H., Guo, X., Cao, L. and Gao, W. (2021) Increased turbulence in the Eurasian upper-level jet stream in winter: past and future. *Earth and Space Science*, 8(2).
- MacCready, P.B.Jr. (1964) Standardization of gustiness values from aircraft. *Journal of Applied Meteorology*, 3, 439–449.
- McCann, D.W., Knox, J.A. and Williams, P.D. (2012) An improvement in clear-air turbulence forecasting based on spontaneous imbalance theory: the ULTURB algorithm. *Meteorological Applications*, 19, 71–78.
- Met Office (2013) *Iris: A Python library for analysing and visualising meteorological and oceanographic data sets v1.2*. Exeter, UK. <http://scitools.org.uk/>.
- Palmer, T.N., Shutts, G.J. and Swinbank, R. (1986) Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parametrization. *Quarterly Journal of the Royal Meteorological Society*, 112, 1001–1039.
- Price, C. and Rind, D. (1994) Possible implications of global climate change on global lightning distributions and frequencies. *Journal of Geophysical Research: Atmospheres*, 99, 10823–10831.

- Reeve, N. and Toumi, R. (1999) Lightning activity as an indicator of climate change. *Quarterly Journal of the Royal Meteorological Society*, 125, 893–903.
- Sharman, R. and Lane, T.P. (Eds.) (2016) *Aviation Turbulence: Processes, Detection, Prediction*. Cham, Switzerland: Springer.
- Sharman, R. and Pearson, J. (2017) Prediction of energy dissipation rates for aviation turbulence. Part I: forecasting non-convective turbulence. *Journal of Applied Meteorology and Climatology*, 56, 317–337.
- Sharman, R., Tebaldi, C., Wiener, G. and Wolff, J. (2006) An integrated approach to mid-and upper-level turbulence forecasting. *Weather and Forecasting*, 21, 268–287.
- Sharman, R.D., Trier, S.B., Lane, T.P. and Doyle, J.D. (2012) Sources and dynamics of turbulence in the upper troposphere and lower stratosphere: a review. *Geophysical Research Letters*, 39(12). <https://doi.org/10.1029/2012GL051996>.
- Storer, L.N., Williams, P.D. and Joshi, M.M. (2017) Global response of clear-air turbulence to climate change. *Geophysical Research Letters*, 44, 9976–9984.
- Storer, L.N., Gill, P.G. and Williams, P.D. (2019a) Multi-model ensemble predictions of aviation turbulence. *Meteorological Applications*, 26, 416–428.
- Storer, L.N., Williams, P.D. and Gill, P.G. (2019b) Aviation turbulence: dynamics, forecasting, and response to climate change. *Pure and Applied Geophysics*, 176, 2081–2095.
- Storer, L.N., Gill, P.G. and Williams, P.D. (2020) Multi-diagnostic multi-model ensemble forecasts of aviation turbulence. *Meteorological Applications*, 27(1). <https://doi.org/10.1002/met.1885>.
- Taylor, K.E., Stouffer, R.J. and Meehl, G.A. (2012) An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93, 485–498.
- Uccellini, L.W. and Koch, S.E. (1987) The synoptic setting and possible energy sources for mesoscale wave disturbances. *Monthly Weather Review*, 115, 721–729.
- Uccellini, L.W., Petersen, R.A., Keyser, D., Kocin, P.J., des Jardins, M., Brill, K.F. and Aune, R. (1986) A report on the upper-level wind conditions preceding and during the shuttle *Challenger* (STS 51L) explosion. *Bulletin of the American Meteorological Society*, 67, 1248–1265.
- Williams, P.D. (2005) Modelling climate change: the role of unresolved processes. *Philosophical Transactions of the Royal Society A*, 363, 2931–2946.
- Williams, P.D. (2016) Transatlantic flight times and climate change. *Environmental Research Letters*, 11(2). <https://doi.org/10.1088/1748-9326/11/2/024008>.
- Williams, P.D. (2017) Increased light, moderate, and severe clear-air turbulence in response to climate change. *Advances in Atmospheric Sciences*, 34, 576–586.
- Williams, P.D. and Joshi, M.M. (2013) Intensification of winter transatlantic aviation turbulence in response to climate change. *Nature Climate Change*, 3, 644–648.
- Williams, P.D., Read, P.L. and Haine, T.W.N. (2003) Spontaneous generation and impact of inertia-gravity waves in a stratified, two-layer shear flow. *Geophysical Research Letters*, 30(24). <https://doi.org/10.1029/2003GL018498>.
- Williams, P.D., Haine, T.W.N. and 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.N. and Read, P.L. (2008) Inertia-gravity waves emitted from balanced flow: observations, properties, and consequences. *Journal of the Atmospheric Sciences*, 65, 3543–3556.
- Williams, P.D., Alexander, M.J., Barnes, E.A., Butler, A.H., Davies, H.C., Garfinkel, C.I., Kushnir, Y., Lane, T.P., Lundquist, J.K., Martius, O., Maue, R.N., Peltier, W.R., Sato, K., Scaife, A.A. and Zhang, C. (2017) A census of atmospheric variability from seconds to decades. *Geophysical Research Letters*, 44, 11201–11211.
- Wolff, J.K. and Sharman, R.D. (2008) Climatology of upper-level turbulence over the contiguous United States. *Journal of Applied Meteorology and Climatology*, 47, 2198–2214.

How to cite this article: Williams, P.D. & Storer, L.N. (2022) Can a climate model successfully diagnose clear-air turbulence and its response to climate change?. *Quarterly Journal of the Royal Meteorological Society*, 148(744), 1424–1438. Available from: <https://doi.org/10.1002/qj.4270>