Climatology of upper-tropospheric turbulence: capabilities and limitations of aircraft reports and ERA5 reanalysis diagnostics

Received: 15 April 2025 | Revised: 27 October 2025 | Accepted: 28 October 2025

DOI: 10.1002/qj.70073

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Climatology of upper-tropospheric turbulence: Capabilities and limitations of aircraft reports and ERA5 reanalysis diagnostics

Thorsten Kaluza[1] | **Paul D. Williams**[1] | **David M. Schultz**[2] | **Timothy P. Banyard**[2]

[1]Department of Meteorology, University of Reading, Reading, UK

[2]Centre for Atmospheric Science, Department of Earth and Environmental Sciences, University of Manchester, Manchester, UK

Correspondence
Thorsten Kaluza, Department of Meteorology, University of Reading, Reading, UK.
Email: t.p.kaluza@reading.ac.uk

Funding information
Natural Environment Research Council, Grant/Award Number: NE/W000997/1

Abstract

A climatology of upper-tropospheric turbulence is constructed from eight years of automated Aircraft Communications Addressing and Reporting System (ACARS) turbulence reports from commercial aircraft. Seasonal frequency maps for consistently reported turbulence intensities show contiguous planetary-scale occurrence frequency maxima over the North Atlantic and North Pacific winter storm tracks, and along tropical flight routes over both oceans. The analyzed turbulence intensities are encountered less often over North America, except for a widespread increase in the east during spring. Low Richardson numbers calculated from European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis Version 5 (ERA5) reanalysis fields and interpolated to report locations are diagnosed most frequently in regions where turbulence is observed most frequently. The frequency maps of the 99th percentile of the Richardson number largely mirror the large-scale patterns in the turbulence frequency maps. In comparison, the turbulence index (TI1) diagnoses increased potential for turbulence most frequently at high latitudes north of where turbulence is observed most frequently. For the analyzed 99th percentile thresholds of the diagnostics, the probability of detection (POD) is around 20%, despite the overprediction at average turbulence frequencies of 0.15%. Although the POD increases with higher diagnostic percentiles, this comes at the cost of an exponentially decreasing precision. Representative precision is achieved, if at all, only for a small fraction of the observed turbulence. Near high clouds, as indicated by low satellite infrared brightness temperatures, turbulence frequencies increase by a factor of more than five. However, a non-negligible POD is again only achieved at the expense of significant overprediction by the diagnostics. So, whereas EDR $\geq 0.18\ \mathrm{m}^{2/3} \cdot \mathrm{s}^{-1}$ is observed most frequently in regions where small Richardson numbers are diagnosed most frequently in ERA5, the rare occurrence of the analyzed turbulence intensity remains hidden in the diagnostic frequency maps.

KEYWORDS

aircraft turbulence reports, climatology, ERA5 turbulence diagnostics, upper troposphere

1 | INTRODUCTION

Atmospheric turbulence is commonly described as the cascading transfer of energy from larger- to smaller-scale eddies down to the molecular scale and dissipation through viscosity. This process redistributes heat, momentum, and chemical constituents within the atmosphere, modifying the evolution of the large-scale flow (e.g., Grams *et al.*, 2011; Martius *et al.*, 2010; Spreitzer *et al.*, 2019), the transport pathways of trace gases through mixing across transport barriers (e.g., Berthet *et al.*, 2007; Hoor *et al.*, 2004), and the radiative budget of the atmosphere (e.g., Forster & Shine, 1997; Riese *et al.*, 2012). Furthermore, turbulence is the leading cause of weather-related aircraft incidents (e.g., Gultepe *et al.*, 2019; Sharman *et al.*, 2012). As such, understanding when and where turbulence occurs in the atmosphere and how best to predict it are active research questions. Two approaches to answer these questions have been employed in past literature.

The first approach is to use observational data like the turbulence reports provided by commercial aircraft. The benefit of this approach is direct knowledge of the occurrence and magnitude of the turbulence. This approach is limited, however, first because reports are only available along flight routes and second because well-forecast turbulent regions or those easily spotted (e.g., convective storms) are undersampled by aircraft. For the upper troposphere, a few studies on the geographic distribution of turbulence that use aircraft reports do exist. For example, Wolff and Sharman (2008) examined over 2.3 million manual pilot reports of the perceived turbulence intensity from flights over the contiguous United States (CONUS). The manual pilot reports are not intended to provide research-quality turbulence information, and the derived turbulence maps should be interpreted as relative distributions and not true probabilities. They found relative maxima during winter for the aviation turbulence category moderate or greater (MOG) turbulence and regional maxima associated with mountain-wave turbulence, the jet stream, and convective storms. More recently, Ko *et al.* (2023) compiled MOG turbulence maps over the CONUS from more than 200 million automated reports of measured turbulence intensities. The turbulence maps show localized small-scale frequency maxima and pronounced regional variability, and although Ko *et al.* (2023) relate certain maxima to mountain waves and the jet-stream location, they conclude that more reports are necessary to derive robust MOG turbulence maps. The limitations of these few available studies highlight that, along with the continuously growing dataset of archived turbulence reports, more studies on the spatial and temporal distribution of upper-tropospheric turbulence are necessary.

The second approach to analyze the spatial and temporal distribution of turbulence is to use gridded datasets of meteorological quantities (such as from reanalyses, weather model output, or climate model output) to construct maps of turbulence diagnostics. Because turbulence as a microscale process is not resolved in the models, these diagnostics identify resolved flow features that indicate the potential for flow instabilities and turbulent breakdown. Different diagnostics are either associated with different flow instabilities or focus on varying aspects of the evolution of individual instabilities. Accordingly, no single diagnostic exists to identify all observed turbulence. Nevertheless, the easy availability of such gridded model datasets has led to a large number of studies examining the spatial and temporal distribution of turbulence diagnostics, in both the past and the future. For example, Jaeger and Sprenger (2007) examined 44 years of European Centre for Medium-Range Weather Forecasts (ECMWF) 40-year Re-Analysis (ERA40) data for the Northern Hemisphere and compiled frequency maps for turbulence diagnostics near the tropopause. They found that the diagnostics diagnose turbulence most frequently over the winter storm tracks, with varying location and extent of the maxima for individual diagnostics. A more recent study confirmed these results for a similar selection of diagnostics and on the 250-hPa pressure level, using 41 years of ECMWF Reanalysis Version 5 (ERA5) data (Lee *et al.*, 2023). A number of studies have examined the conditions near storm tracks in climate model output and have found the frequency of turbulence according to the diagnostics to increase in the future (e.g., Smith *et al.*, 2023; Williams & Joshi, 2013).

These two approaches to analyze the spatial and temporal distribution of turbulence complement each other in theory. In practice, the relation between a diagnostic climatology and a turbulence climatology depends on two characteristics of the analyzed diagnostic. The first characteristic is the fraction of turbulence that is attributed to the flow conditions described by the diagnostic. The second characteristic is the fraction of the flow conditions described by the diagnostic that is turbulent. Or to rephrase it in one question: "How much turbulence is diagnosed, and at what precision?" Several studies have quantified the predictive skill of the variety of known turbulence diagnostics using automated turbulence reports (e.g., Sharman *et al.*, 2006; Sharman & Pearson, 2017). These studies generally focus on turbulence diagnostics as a forecast tool and derive metrics that provide, among other things, a hierarchy of the predictive skill of the diagnostics. Although this hierarchy of the predictive skill helps guide the selection of the diagnostics used in turbulence diagnostic climatologies, the metrics used in

these studies are generally not applicable to interpret the climatologies.

We are unaware of any study that has analyzed a climatology of observed and diagnosed turbulence simultaneously along with the predictive skill of the selected diagnostics. We address this issue using a dataset of eight years of automated turbulence reports from commercial aircraft and compare it with two popular diagnostics calculated from ERA5 reanalyses (Hersbach *et al.*, 2020). We aim to show the geographic distribution of upper-tropospheric turbulence by season from the measurements and show how these diagnostics compare. We calculate the predictive skill of these diagnostics (how much turbulence is diagnosed and at what precision) and discuss the applicability of building climatologies of turbulence from diagnostics, with the insight that corresponding observations of turbulence provide.

Section 2 introduces the measurement and downlink procedure of the turbulence reports and the preselection steps applied to derive robust turbulence frequencies. This section furthermore introduces the ERA5 reanalysis and the turbulence diagnostics. Section 3 presents the seasonal frequency maps of the observed turbulence in comparison with the ERA5-derived diagnostics and discusses how they compare in the context of the predictive skill metrics. Section 4 revisits key aspects of the analysis using subsets of reports classified by cloud brightness temperature, aiming to determine whether the diagnostics perform differently under near-cloud and in-cloud conditions. Section 5 discusses the capabilities and limitations of the turbulence reports and the ERA5 diagnostics, as well as the methods applied to analyze and compare them. Our conclusions are summarized in Section 6.

2 | DATA AND METHODS

This study is based on automated Aircraft Communications Addressing and Reporting System (ACARS) turbulence reports from commercial aircraft from January 1, 2017–December 31, 2024. The turbulence reports are made available through the Meteorological Assimilation Data Ingest System (MADIS) database by the National Oceanic and Atmospheric Administration (NOAA). The start of the analyzed time period is set by a gap of archived ACARS reports in the MADIS database in 2016.

The cube root of the eddy dissipation rate $\mathrm{EDR} = \varepsilon^{1/3}$ (in $\mathrm{m}^{2/3} \cdot \mathrm{s}^{-1}$) has been established as the standard aviation turbulence intensity metric (Sharman & Pearson, 2017). For the automated turbulence reports and under the assumption of isotropy, EDR is derived from spectral analysis of the measured vertical wind (Sharman *et al.*, 2014). The measurements of the vertical wind are available at 8 Hz. Every 5 s, the on-board algorithm calculates the Fourier transform for a 10-s time series of the vertical wind and derives an EDR estimate from the power spectrum. This results in 12 EDR estimates each minute. From these, the system identifies both a median and the peak EDR value for each minute. In this study, we use the peak EDR value recorded each minute as the measure of turbulence.[1]

We apply four preselection steps to the turbulence reports downloaded from the MADIS archive. First, we disregard all reports below 8-km flight altitude to remove the bias of the localized probing of the middle and lower troposphere at airport locations. Second, we disregard all reports that have been flagged as questioned or rejected by the internal consistency checks that MADIS applies (such as an unrealistic aircraft speed over the distance between two consecutive reports). Third, we select only archived reports with a regular one-minute sampling period along individual flight tracks. Regular sampling is only given for flights for which MADIS fills flight segments without *routine* or *triggered* reports with $\mathrm{EDR} = 0\ \mathrm{m}^{2/3} \cdot \mathrm{s}^{-1}$ reports. Routine reports are downlinked either every 15 or every 20 minutes to minimize downlink cost. Triggered reports are additionally downlinked when the measured EDR exceeds certain thresholds (see next paragraph). Fourth, we exclude all reports located in the stratosphere, as explained further in the following paragraph.

Figure 1a,b illustrates the seasonality of tropopause altitude alongside the constant mean cruise altitudes. In winter, aircraft probe more stratospheric air due to the lower tropopause heights. The vertical distribution of turbulence frequencies relative to the tropopause reveals a sharp gradient across the ERA5-derived lapse-rate tropopause (LRT, Figure 1c), resulting in a seasonal bias with fewer turbulence encounters in winter compared with summer. To minimize the impact of this tropopause seasonality and to focus on the regions where turbulence occurs most frequently, we limit the analysis to reports within the upper troposphere, excluding all stratospheric reports. For the dynamic tropopause, which is on average located below the LRT (Figure 1a,b), the turbulence frequency maximum shifts up to 1 km above the tropopause, indicating that turbulence often occurs in these regions of differing tropopause definitions. To minimize the influence of the tropopause seasonality, we proceed with the analysis using only reports below the LRT.

Figure 1c introduces $\mathrm{EDR} \geq 0.18\ \mathrm{m}^{2/3} \cdot \mathrm{s}^{-1}$ as the turbulence intensity threshold analyzed in this study. The turbulence intensity is selected based on the downlink thresholds for triggered reports. Reports are downlinked if the median EDR exceeded $0.06\ \mathrm{m}^{2/3} \cdot \mathrm{s}^{-1}$ over four of the previous six minutes, if the peak EDR exceeded $0.12\ \mathrm{m}^{2/3} \cdot \mathrm{s}^{-1}$ over three of the previous six minutes,

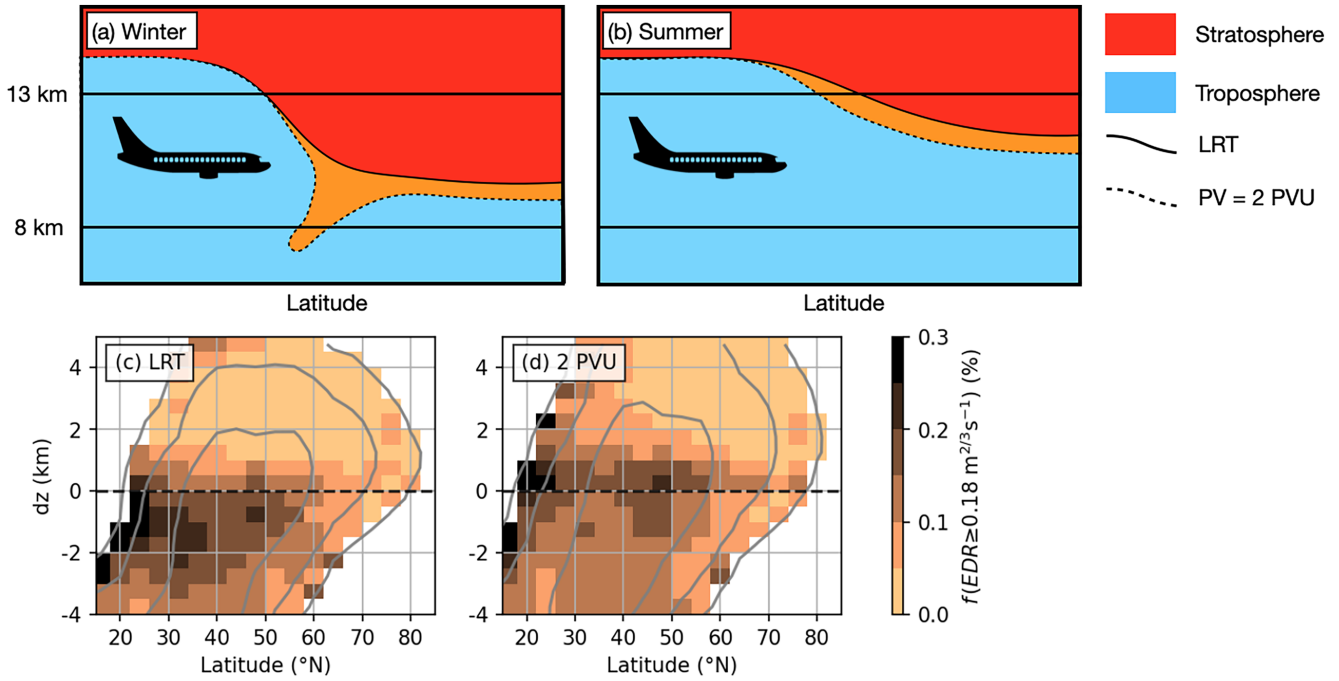

FIGURE 1 The upper panels show a schematic of the tropopause seasonality in a latitude–altitude cross-section. The cruise altitudes are indicated with horizontal lines. (a) The winter season with lower average tropopause altitudes compared with the summer season in (b). The PV = 2 PVU dynamic tropopause is indicated at, on average, lower altitudes compared with the lapse-rate tropopause (LRT), as well as a stratospheric intrusion of low PV air in winter. The lower panels show zonally averaged relative frequencies $f(\text{EDR} \geq 0.18\ \text{m}^{2/3} \cdot \text{s}^{-1})$ for all regularly sampled ACARS reports from the NOAA MADIS archive from January 1, 2017–December 31, 2024. The reports are binned at the local latitude and vertical distance from the local lapse rate tropopause in (c), and from the PV = 2 PVU dynamic tropopause in (d). Zonally averaged report counts of 10^4, 10^5, and 10^6 are indicated by the gray contour lines. [Colour figure can be viewed at wileyonlinelibrary.com]

or whenever the peak EDR exceeds $0.18\ \text{m}^{2/3} \cdot \text{s}^{-1}$ (WMO, 2013). Therefore, $\text{EDR} \geq 0.18\ \text{m}^{2/3} \cdot \text{s}^{-1}$ is the lowest turbulence intensity without a downlink bias. It falls between the aviation turbulence categories light or greater (LOG) turbulence, defined by $\text{EDR} \geq 0.15\ \text{m}^{2/3} \cdot \text{s}^{-1}$, and MOG turbulence, defined by $\text{EDR} \geq 0.22\ \text{m}^{2/3} \cdot \text{s}^{-1}$.

The second dataset we use is the ERA5 reanalysis (Hersbach *et al.*, 2020), to calculate two popular turbulence diagnostics and the LRT altitude (WMO, 1957). We use hourly ERA5 output fields on a regular 0.25° latitude–longitude grid and on the native vertical hybrid sigma–pressure level. Spatio-temporal matching of the LRT and the diagnostics from the two neighboring full-hour ERA5 fields with the turbulence reports is achieved through linear interpolation in the horizontal and vertical, as well as in time.

The first diagnostic we calculate from the gridded ERA5 fields is the Richardson number (Richardson, 1922):

$$\text{Ri} = N^2/S^2. \tag{1}$$

Static stability is defined as $N^2 = g/\theta \cdot (\partial\theta/\partial z)$, where g is Earth's gravitational acceleration, θ is the potential temperature, and z is the altitude. The squared vertical shear of the horizontal wind is defined as $S^2 = (\partial u/\partial z)^2 + (\partial v/\partial z)^2$, with the zonal and meridional wind components u and v. The Richardson number can be derived from stability analysis in linear wave theory, where $\text{Ri} \leq 1/4$ is an instability criterion. In the gridded ERA5 fields, the analytically derived gradient Richardson number becomes the bulk Richardson number, which can be interpreted to characterize the stability of the atmosphere as follows. First, small Richardson numbers can indicate wind shear that is strong enough to overcome the restoring buoyant forces in stably stratified flow, resulting in overturning of isentropes and generation of Kelvin–Helmholtz instability as a common source for turbulence in the free atmosphere (e.g., Dutton & Panofsky, 1970; Sharman *et al.*, 2012). Second, small or negative Richardson numbers can indicate weakly stable or unstable stratification (scaled by the wind shear), along with increased potential for moist convection and in-cloud turbulence.

The second diagnostic is the Turbulence Index 1 (TI1: Ellrod & Knapp, 1992):

$$\text{TI1} = \sqrt{S^2 \cdot (D_{\text{SH}}^2 + D_{\text{ST}}^2)}. \tag{2}$$

The shearing deformation is defined as $D_{\text{SH}} = \partial v/\partial x + \partial u/\partial y$ and the stretching deformation as $D_{\text{ST}} = \partial u/\partial x - \partial v/\partial y$ using the horizontal derivatives of the wind

components in zonal (x) and meridional (y) directions. TI1 is a semi-empirical diagnostic, which is commonly used in studies on atmospheric turbulence because of its high rank in the predictive-skill hierarchy of known diagnostics (e.g., Sharman & Pearson, 2017). Large TI1 values indicate strong wind shear, scaled by the flow deformation, which can increase horizontal temperature gradients and in turn increase wind shear according to the thermal wind relation (Ellrod & Knapp, 1992). The Richardson number and TI1 are related through the vertical shear of the horizontal wind as one component for instability in stably stratified flow. TI1 has a high predictive skill, although it neglects the influence of the stratification, implying that the strongest wind shears resolved in the models used in validation studies (e.g., Sharman & Pearson, 2017) frequently result in instability and turbulent breakdown of the flow. The relationship of the Richardson number and TI1, their varying characteristics, and their frequent appearance in studies on atmospheric turbulence motivated us to select them as example diagnostics for our study.

The third dataset we use consists of *Geostationary Operational Environmental Satellite 16* (*GOES-16*) Advanced Baseline Imager (ABI) satellite brightness-temperature measurements in infrared band 13 (10.3 μm). We use the full disk measurements at the native horizontal and temporal resolutions of 2 km and 10 minutes, respectively. Brightness temperatures are matched to the ACARS reports via a nearest-neighbor search in both space and time.

We derive three kinds of seasonal frequency maps by sorting the upper-tropospheric reports into $4° \times 4°$ latitude–longitude bins and calculating the frequency of reports exceeding the thresholds of interest (EDR, Ri, TI1) relative to all reports in each bin. First, we derive vertically averaged seasonal turbulence frequency maps for EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$ from the regularly sampled upper-tropospheric ACARS turbulence reports from January 1, 2017–December 31, 2024. Second, we derive vertically averaged seasonal turbulence diagnostic frequency maps for the 99th percentile of the Richardson number at the report location, Ri ≤ 0.37, and for the 99th percentile of TI1 at the report location, TI1 $\geq 2.13 \times 10^{-6}\ s^{-1}$. The percentile method is a common approach in turbulence diagnostic climatologies, with the choice of the percentile being motivated either empirically or to mirror the frequency of the analyzed turbulence intensity. We select the 99th percentile as an example and discuss the relevance of this choice in Section 3.2. The third kind of seasonal frequency maps we derive are for Ri ≤ 0.37 and TI1 $\geq 2.13 \times 10^{-6}\ s^{-1}$ from all upper-tropospheric ERA5 grid points within the altitude range 11 ± 0.9 km, which corresponds to the mean and standard deviation of the vertical distribution of the analyzed upper-tropospheric reports. We exclusively use data from aircraft equipped with the National Center for Atmospheric Research (NCAR) EDR reporting algorithm, focusing on flight tracks with a regular one-minute sampling frequency from routine, triggered, and interpolated reports, and on consistently downlinked turbulence intensities with EDR $< 0.18\ m^{2/3} \cdot s^{-1}$. Consequently, the derived turbulence frequencies should not exhibit sampling bias.

3 | RESULTS

The following subsections present turbulence frequency maps for individual seasons in comparison with the according frequency maps of the 99th percentile of the Richardson number and TI1. We begin with the frequency maps for the Northern Hemisphere winter (December–January–February, DJF) in Section 3.1, followed by an analysis of the average and regional predictive skill of both diagnostics in Section 3.2. After the more detailed description of the approach for DJF, the comparison of the observed and diagnosed turbulence frequencies under consideration of the regional predictive skill is repeated for the remaining seasons in Section 3.3.

3.1 | Turbulence and diagnostic frequency maps for DJF

Figure 2a shows the turbulence frequency map for EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$ derived from all ACARS turbulence reports from flights during DJF and in the altitude range between 8 km and the LRT. The geographic coverage of the MADIS ACARS reports is limited to flights leaving or arriving at airports in the USA. EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$ occurs most frequently over large areas of the North Pacific and the western North Atlantic, whereas North America exhibits comparatively low turbulence frequencies across broad areas (Figure 2a). Additional turbulence frequency maxima occur over tropical oceans.

Figure 2b shows the DJF frequency map for the 99th percentile of the ERA5 Richardson number at the report location, Ri ≤ 0.37. The distribution of Ri ≤ 0.37 shows a general agreement with the distribution of EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$ (Figure 2a), capturing the key features identified in the turbulence map. Small Richardson numbers are diagnosed most frequently in regions where EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$ is observed most frequently. When examined in more detail, differences between observed and diagnosed turbulence become evident in the frequency maps (Figure 2a,b). For example, Ri ≤ 0.37 occurs frequently over northern South America and over the

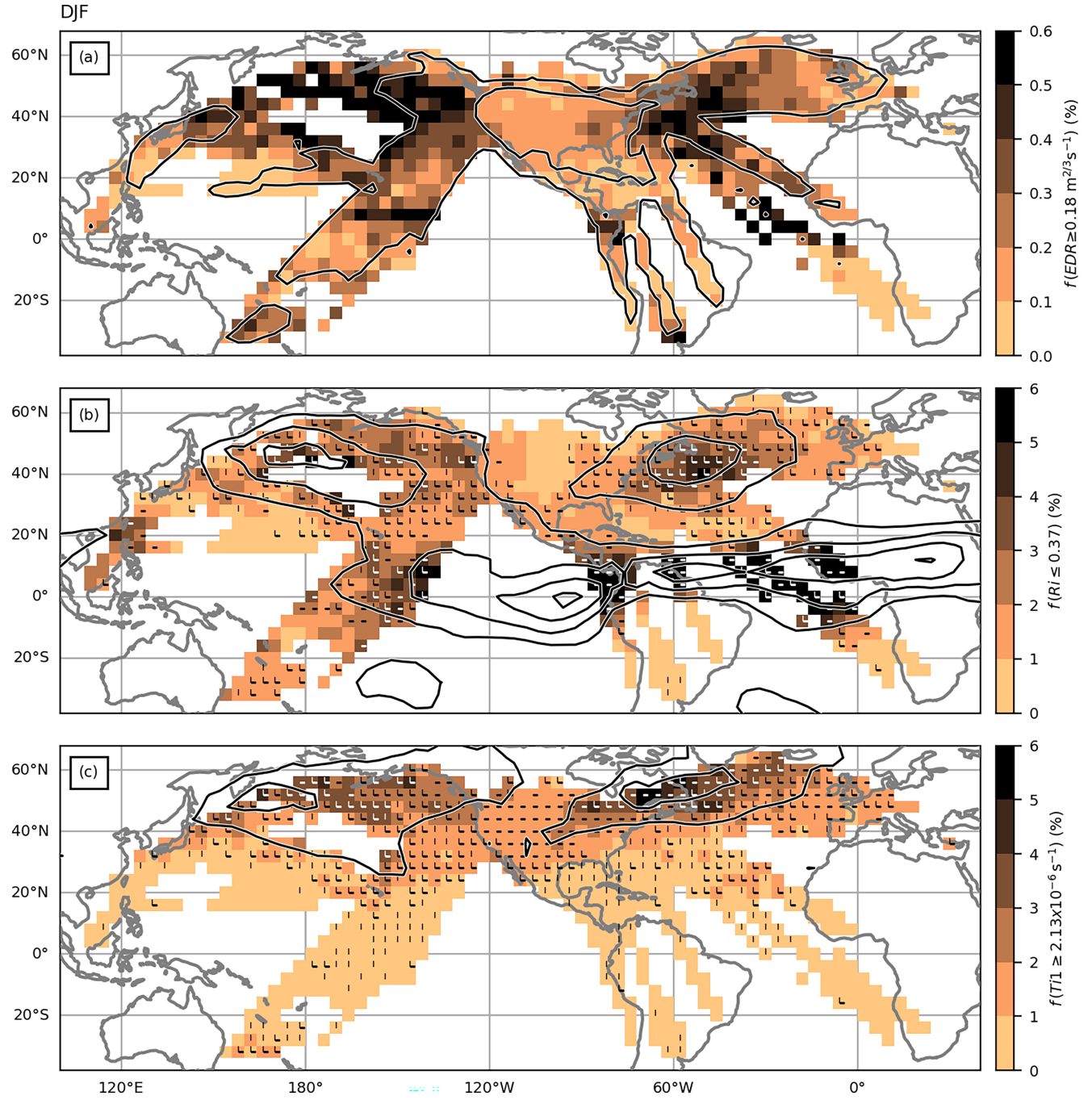

FIGURE 2 Relative frequency maps derived from upper-tropospheric ($8\ \text{km} \le z \le z_{\text{LRT}}$) turbulence reports during DJF for (a) EDR $\ge 0.18\ \text{m}^{2/3} \cdot \text{s}^{-1}$, (b) Ri ≤ 0.37, and (c) TI1 $\ge 2.13 \times 10^{-6}\ \text{s}^{-1}$. The display threshold is 5000 reports per bin. Black contour lines in (a) show report counts of 10^4, 10^5, and 10^6. Filled contours in (b) and (c) show the relative frequencies of the diagnostics at the report location, and black contour lines the relative frequencies (in steps of 3%) derived from all upper-tropospheric grid points in the ERA5 in the range 11 ± 0.9 km altitude. Vertical and horizontal line markers in (b) and (c) indicate above-average local diagnostic precision and probability of detection (POD), respectively, compared with the seasonal average (see also Figure 3e). [Colour figure can be viewed at wileyonlinelibrary.com]

Philippine Sea (Figure 2b), where EDR $\ge 0.18\ \text{m}^{2/3} \cdot \text{s}^{-1}$ is rarely observed (Figure 2a). Over Argentina and east of Australia, EDR $\ge 0.18\ \text{m}^{2/3} \cdot \text{s}^{-1}$ is observed comparatively frequently (Figure 2a) where Ri ≤ 0.37 exhibits no frequency maximum (Figure 2b). The relative frequency maxima of Ri ≤ 0.37 diagnosed along the flight tracks align with the frequency maxima of Ri ≤ 0.37 derived from all tropospheric ERA5 grid points in range 11 ± 0.9 km altitude (Figure 2b). The alignment of the frequency maxima suggests that the air masses sampled by the turbulence

reports are, to some extent, representative for the upper troposphere. For example, Ri $\leq$ 0.37 occurs over 6% of the time over the western North Atlantic, in regions where Ri $\leq$ 0.37 is also encountered over 4% of the time along flight tracks. The difference in the magnitude of the frequency maxima could arise from mitigation strategies in flight planning, where forecast regions of small Richardson numbers are avoided. Additionally, the interpolation of the Richardson number to the time and location of the reports smooths extreme values compared with those at ERA5 grid points, contributing to the observed discrepancies.

Figure 2c shows the DJF frequency map for the 99th percentile of TI1 at the report location, TI1 $\geq 2.13 \times 10^{-6}$ s^{-1}. Compared with the Richardson number, TI1 shows less agreement with the distribution of the observed turbulence (Figure 2a). The predominant features in the TI1 frequency map are the maxima over the North Pacific and the western North Atlantic (Figure 2c), which are shifted towards higher latitudes compared with the turbulence frequency maxima (Figure 2a). The northward shift of TI1 frequency maxima is apparent along the flight tracks and in the altitude range of 11 ± 0.9 km, indicating that the occurrence of TI1 $\geq 2.13 \times 10^{-6}$ s^{-1} along the flight tracks is generally representative for the distribution of TI1 $\geq 2.13 \times 10^{-6}$ s^{-1} in the upper troposphere.

3.2 | How much turbulence is diagnosed, and at what precision?

The DJF frequency map for the 99th percentile of the Richardson number (Figure 2b) broadly agrees with the turbulence frequency map for EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1} (Figure 2a), whereas the distribution of the 99th percentile of TI1 (Figure 2c) exhibits a distinct northward shift of the storm track maxima compared with the observed turbulence and the Richardson number. The alignment (or lack thereof) of the seasonal diagnostic frequency maxima with the seasonal turbulence frequency maxima does not provide information about the classification skill of the diagnostics (i.e., the question posed in the Introduction: "How much turbulence is diagnosed, and at what precision?").

Starting with the distribution of observed turbulence intensities (Figure 3a) and the threshold EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1}, a classification metric quantifies how well a diagnostic threshold separates reports with EDR < 0.18 m$^{2/3} \cdot$ s^{-1} from reports with EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1} (Figure 3b,c). A common way to illustrate the classification skill is to calculate the probability of detection (POD) and the probability of false detection (POFD) and plot POD–POFD pairs as a function of the diagnostic threshold as receiver operator characteristic (ROC) curves (Figure 3d). Figure 3d shows the ROC curves for the Richardson number and TI1, calculated from all upper-tropospheric EDR reports from flights during DJF and for EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1}. The ROC curve for TI1 is shifted towards the top-left corner relative to the ROC curve for the Richardson number, indicating that TI1 has a lower POFD at any given POD and therefore a higher classification skill. The POD of 19.3% for Ri $\leq$ 0.37 and 23.8% for TI1 $\geq 2.13 \times 10^{-6}$ s^{-1} answers the first question, "How much turbulence is diagnosed?" The precision, however, cannot be inferred from the ROC curves, because the precision depends on the class imbalance, defined by the fraction of reports that exceed EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1}, and the POD and the POFD are independent of the class imbalance.

In Figure 3e, the precision on the *y*-axis is plotted against the POD on the *x*-axis, thus including the class imbalance:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{\text{POD} \times \text{P}}{\text{POD} \times \text{P} + \text{POFD} \times \text{N}}, \quad (3)$$

considering the true positives (TP), false positives (FP), positives (P: EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1}), and negatives (N: EDR < 0.18 m$^{2/3} \cdot$ s^{-1}). TI1 has comparatively high precision at sharp thresholds with low POD (Figure 3e), indicating that the largest TI1 values resolved in ERA5 are frequently associated with EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1}. Compared with TI1, the Richardson number has a lower precision at sharp thresholds,[2] but it approaches TI1 with increasing POD, and both diagnostics converge towards the average class imbalance P/(P + N) = 0.0019 at the right-hand *y*-axis intercept, which corresponds to the average occurrence frequency of 0.19% during DJF for EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1}.

According to the EDR reports, 0.19% of the upper troposphere during DJF is turbulent with EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1}. The 99th percentiles of the Richardson number and TI1 each define 1% of the upper troposphere as turbulent. Overprediction is therefore inherent to the approach and increases further, because only a fraction of all observed EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1} is described by the 99th percentile of a diagnostic. Ri $\leq$ 0.37 with a POD of 19.3% exhibits a precision of 2.5% (Figure 3e), and TI1 $\geq 2.13 \times 10^{-6}$ s^{-1} with a POD of 23.8% exhibits a precision of 3.4%. The precision curves indicate that, at the given model resolution, the seasonal turbulence frequency distribution for EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1} cannot be inferred reliably from individual diagnostics. Strict diagnostic thresholds might offer comparatively high precision, but only for a small fraction of turbulence encounters; the majority may not align with the climatological features inferred from such thresholds. Lowering the threshold increases the POD, but

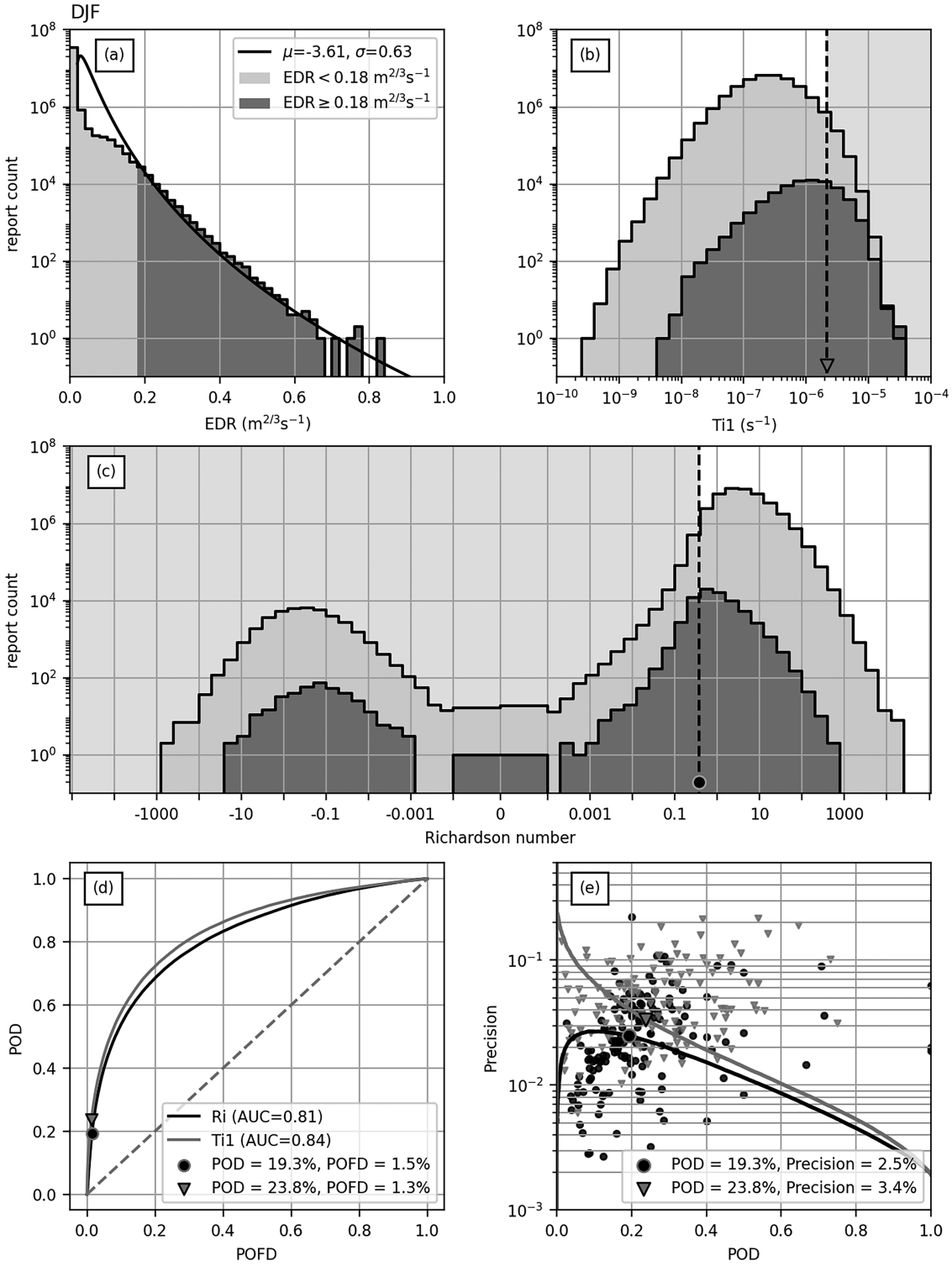

FIGURE 3 Classification skill of the ERA5-derived Richardson number and TI1 during DJF at the report location. The first three plots show the distribution of (a) the measured EDR, (b) the TI1 index at the report location, and (c) the Richardson number at the report location. Reports with EDR $< 0.18\ m^{2/3} \cdot s^{-1}$ are indicated in light gray shading and those with EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$ in dark gray shading. The black line in (a) represents the lognormal fit obtained using the method described by Sharman *et al.* (2014). For details on this method and a comparison of the fit with values reported in the literature, see the Appendix. The reports classified as turbulent by the 99th percentile thresholds of the diagnostics (dashed vertical lines) are highlighted by the shading in (b) and (c). The 99th percentile thresholds are indicated by the circle marker for Ri and the triangle maker for the TI1. Panel (d) shows the ROC curves for Richardson number (black) and TI1 (gray) and the markers as in (b) and (c). The precision curves are shown in (e) with lines, and markers as in (d). The small cirlcle and triangle markers indicate the POD–precision pairs of Ri ≤ 0.37 and TI1 $\geq 2.13 \times 10^{-6}\ s^{-1}$ for the individual $4° \times 4°$ bins in Figure 2.

at the cost of exponentially decreasing precision, thereby reducing the representativeness of the inferred climatological features.

Figure 3e is derived from all upper-tropospheric ACARS reports in the MADIS archive during DJF. About 50% of these reports are downlinked from flights over the contiguous USA (contour lines in Figure 2a), which therefore dominates the average predictive skill displayed in Figure 3e. The range of regional variability of the predictive skill for the 99th percentiles of TI1 and Richardson number is illustrated in Figure 3e, and above-average POD and precision is highlighted in Figure 2b,c. It is neither feasible nor necessarily reasonable to downscale classification metrics to arbitrarily small temporal and spatial scales. The way the metrics are presented here helps identify contiguous regions of enhanced or reduced classification skill on seasonal time-scales. The 99th percentile of the Richardson number shows above-average predictive skill in the storm tracks and locally along the tropical flight routes (Figure 2b), whereas POD and precision are both below average over large areas of the western USA. For TI1 $\geq 2.13 \times 10^{-6}$ s^{-1}, POD and precision are also enhanced in the storm tracks (Figure 2c). In the Tropics, turbulence is rarely associated with the selected TI1 threshold (POD below average). Over the USA, on the other hand, the precision exhibits below-average values. In summary, the inherent overprediction of diagnostic thresholds with non-negligible POD, the exponential trade-off between POD and precision, and the regional variability of both metrics highlight that the geographic distribution of EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1} cannot be inferred from the distribution of the ERA5-derived Richardson number and TI1.

3.3 | Turbulence and diagnostic frequency maps for MAM, JJA, and SON

Figure 4 presents the turbulence frequency map for EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1} for March–April–May (MAM), June–July–August (JJA), and September–October–November (SON), alongside the 99th percentile of the Richardson

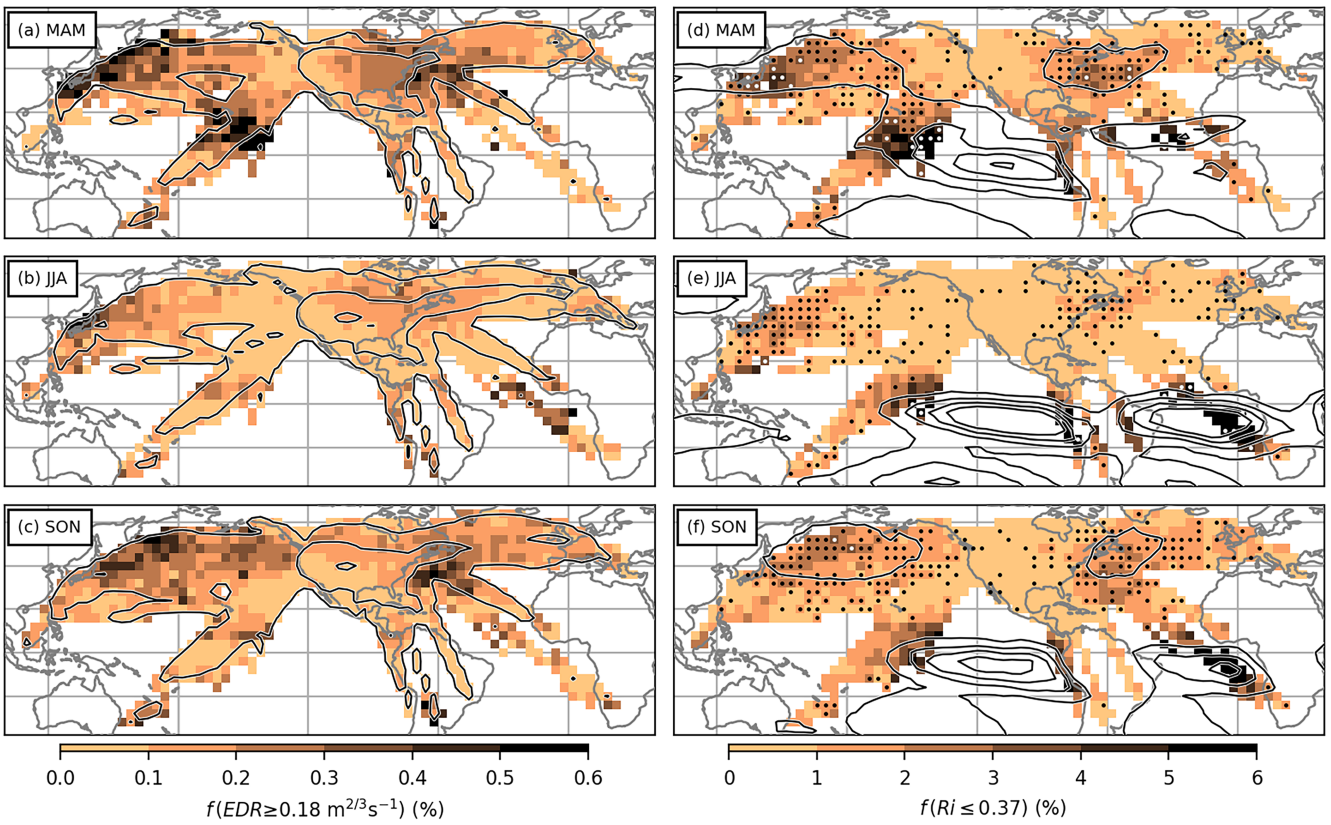

FIGURE 4 Relative frequency maps for EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1} from upper-tropospheric (8 km $\leq z \leq z_{LRT}$) turbulence reports, for (a) MAM, (b) JJA, and (c) SON. The display threshold is 5000 reports per bin. Black contour lines show report counts of 10^4, 10^5, and 10^6. The right-hand panels show relative frequency maps for Ri ≤ 0.37, for (d) MAM, (e) JJA, and (f) SON. Black contour lines show the relative frequencies (in steps of 3%) derived from all upper-tropospheric grid points in the ERA5 in altitude range 11 ± 0.9 km. Stippling highlights bins where POD and precision are enhanced simultaneously compared with the seasonal average (see also Figure 5d). [Colour figure can be viewed at wileyonlinelibrary.com]

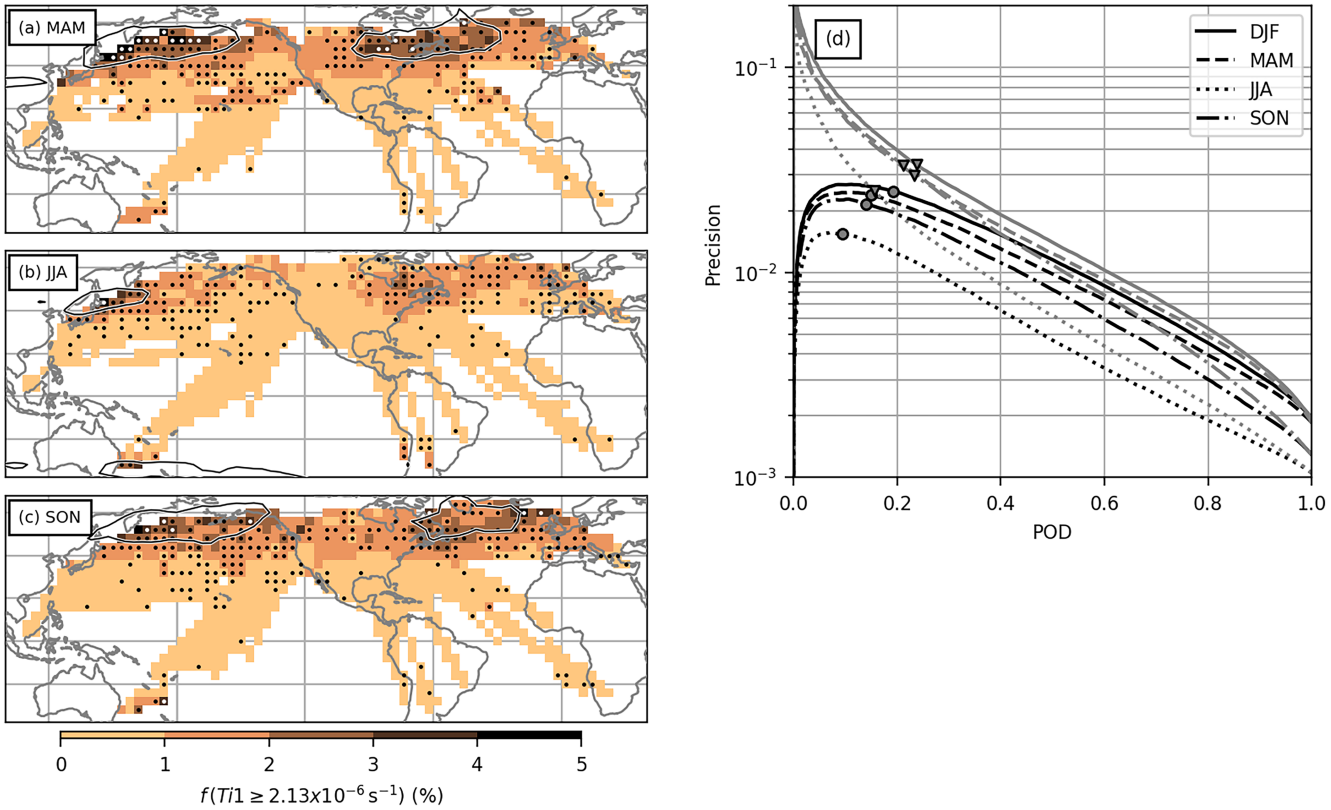

FIGURE 5 Relative frequency maps for TI1 $\geq 2.13 \times 10^{-6}$ s^{-1} at upper-tropospheric (8 km $\leq z \leq z_{LRT}$) report locations, for (a) MAM, (b) JJA, and (c) SON. The display threshold is 5000 reports per bin. Black contour lines show the relative frequencies (in steps of 3%) derived from all upper-tropospheric grid points in ERA5 in altitude range 11 ± 0.9 km. Stippling highlights bins where POD and precision are enhanced simultaneously compared with the seasonal average. (d) shows the precision curves as introduced in Section 3.2 and Figure 3e, for the Richardson number in black and TI1 in gray, for all seasons. The classification skill of the 99th percentiles Ri ≤ 0.32 and TI1 $\geq 2.13 \times 10^{-6}$ s^{-1} is indicated by the circle and triangle marker. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Centered pattern correlation of the turbulence frequency maps for EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1} and the frequency maps for Ri ≤ 0.37 and TI1 $\geq 2.13 \times 10^{-6}$ s^{-1}. The data are scaled by the cosine of latitude to account for the convergence of meridians in the latitude–longitude grid.

	Pattern correlation coefficient against f(EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1})			
	DJF	**MAM**	**JJA**	**SON**
f(Ri ≤ 0.37)	0.43	0.48	0.47	0.16
f(TI1 $\geq 2.13 \times 10^{-6}$ s^{-1})	0.38	0.31	0.30	0.42
f(Ri ≤ 0.37) north of 20°N	0.60	0.65	0.75	0.68
f(TI1 $\geq 2.13 \times 10^{-6}$ s^{-1}) north of 20°N	0.30	0.49	0.55	0.38

number Ri ≤ 0.37. The corresponding frequency maps for TI1 $\geq 2.13 \times 10^{-6}$ s^{-1} are shown in Figure 5. During MAM and compared with DJF, turbulence frequencies for EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1} decrease over the North Atlantic and the North Pacific, where the frequency maximum shifts northwestward (Figure 4a). Turbulence frequencies increase across large parts of eastern North America, which may be related to the seasonal occurrence of deep convection as a source of upper-tropospheric turbulence. In tropical latitudes, turbulence frequencies remain elevated over the Pacific but not over the Atlantic. The 99th percentile of the Richardson number, Ri ≤ 0.32, broadly captures the large-scale distribution of observed turbulence, similar to DJF. In contrast, f(TI1 $\geq 2.13 \times 10^{-6}$ s^{-1}) continues to peak at high latitudes over the oceans, with the maximum over the northwest Pacific partially aligning with the observed turbulence frequency maximum. Table 1 lists the centered pattern correlation coefficients

for each seasonal diagnostic map compared with the corresponding turbulence frequency map, providing a more quantitative assessment of their spatial agreement. The pattern correlation coefficient is generally higher for the Richardson number than for TI1, except in SON, where the signal in the Tropics dominates the Richardson number frequency map. For comparison, we also calculated the pattern correlation using only data north of 20°N.

During JJA, the average frequency $f(\text{EDR} \geq 0.18\ \text{m}^{2/3} \cdot \text{s}^{-1})$ decreases to 0.1% (right-hand side y-axis intercept in Figure 5d). This decrease is also visible in Figure 4b, with reduced turbulence frequencies across all regions except for tropical latitudes over the Atlantic. The $f(\text{Ri} \leq 0.37)$ map captures the overall reduction, aside from the persistent maxima in the Tropics (Figure 4e). TI1 similarly reflects the overall decrease in the extratropics (Figure 5b). Although the Richardson number might be assumed to have better skill than TI1 during Northern Hemisphere summer, the precision curves (Figure 5d) indicate a noticeable drop in predictive skill for both diagnostics compared with the other seasons, with TI1 still outperforming the Richardson number on average. In SON, turbulence frequencies increase again, particularly along the storm tracks. This seasonal change is reflected by both the Richardson number and TI1, with TI1 exhibiting its frequency maximum at higher latitudes.

4 | DOES THE RICHARDSON NUMBER OUTPERFORM TI1 IN NEAR-CLOUD AND IN-CLOUD ENVIRONMENTS?

To assess whether the Richardson number or TI1 perform differently in near-cloud or in-cloud conditions compared with clear-air conditions, in terms of both predictive skill and their alignment with turbulence frequency maps, we repeated the analysis after dividing the dataset according to cloud-top brightness temperatures from satellite observations.

The nearest-neighbor pixels in the *GOES-16* ABI full-disk channel-13 (10.3-μm) measurements were assigned to the ACARS reports, at the native spatial grid spacing of 2 km and time intervals of 10 minutes. *GOES-16* data are available from January 1, 2020. Figure 6a shows that turbulence with EDR $\geq 0.18\ \text{m}^{2/3} \cdot \text{s}^{-1}$ occurs more

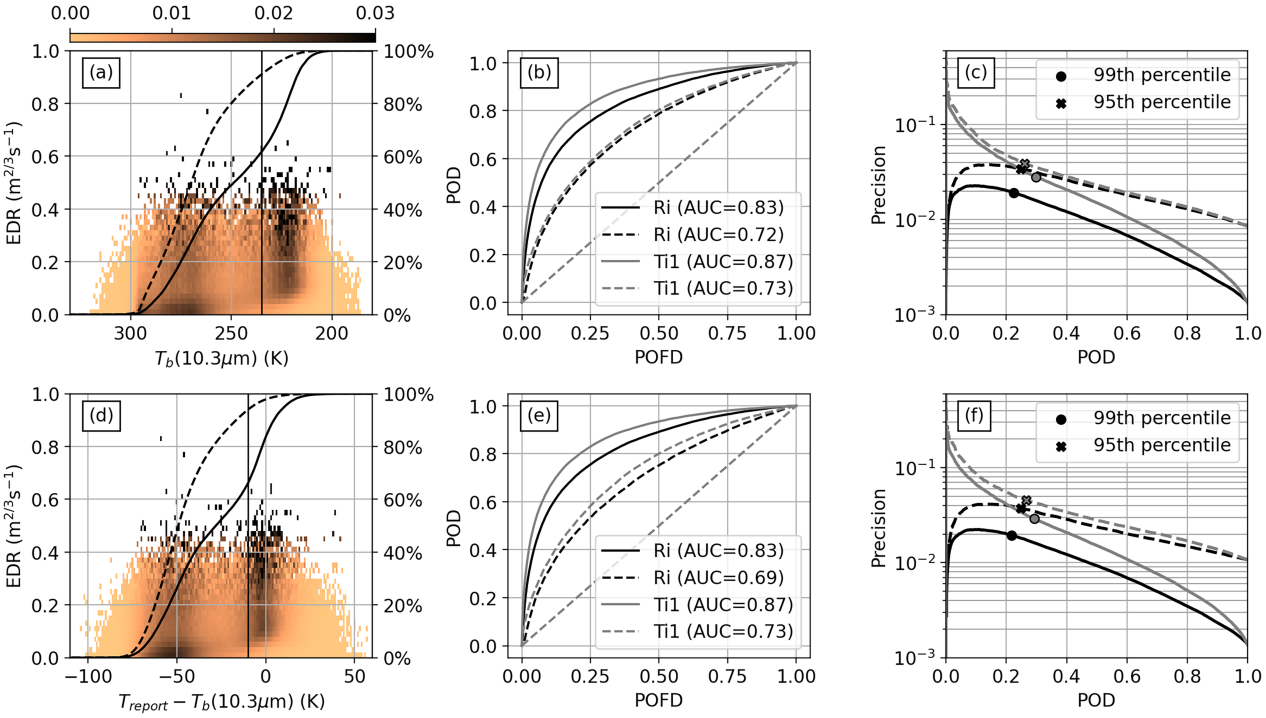

FIGURE 6 Classification skill of subsets of all upper tropospheric reports ($8\ \text{km} \leq z \leq z_{\text{LRT}}\text{km}$) during DJF. (a) Two-dimensional histogram of the *GOES* ABI 10.3-μm brightness temperature T_b (in K) at the report location and the measured EDR (in $\text{m}^{2/3} \cdot \text{s}^{-1}$). The histogram is normalized by the report count in each EDR bin. The black solid line shows the cumulative frequency from high to low temperatures for reports with EDR $\geq 0.18\ \text{m}^{2/3} \cdot \text{s}^{-1}$, and the black dashed line shows the cumulative frequency for all for reports. The vertical line shows the $T_\text{b} = 235$ K threshold. (b) The ROC curves for the Richardson number (black) and TI1 (gray), for reports with $T_\text{b} > 235$ K (solid lines) and reports with $T_\text{b} \leq 235$ K (dashed lines). (c) The according precision curves. The bottom row plots (d), (e), and (f) show the same information for the brightness-temperature difference $T_\text{report} - T_\text{b}(10.3\ \mu\text{m})$ and the threshold of −10 K. [Colour figure can be viewed at wileyonlinelibrary.com]

frequently at low brightness temperatures. The reports are split based on the empirical brightness temperature threshold of 235 K to separate the turbulence frequency maximum associated with high cloud tops from the background flow. The cumulative frequencies indicate that $T_b \leq 235$ K encompasses less than 10% of all reports and almost 40% of reports with EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$. The ROC curves (Figure 6b) indicate a lower skill for reports with $T_b \leq 235$ K, but the lower area under the curve is due first of all to the different class imbalance. The right-hand side *y*-axis intercept in Figure 6c shows an average frequency for EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$ larger than 0.8% for $T_b \leq 235$ K, compared with about 0.15% for $T_b > 235$ K. The precision is enhanced for $T_b \leq 235$ K compared with $T_b > 235$ K, particularly for the Richardson number, which would agree with the assumption that Ri performs better (relative to TI1) in near-cloud and in-cloud conditions. TI1, however, still outperforms the Richardson number, which would agree with the assumption that instabilities related to mesoscale-resolved wind shear (for example, in outflow regions of warm conveyor belts and embedded convection) contribute to near-cloud and in-cloud turbulence, particularly in the winter storm-track turbulence frequency maxima. Another possible approach to split the reports is based on the difference between the temperature at the report location and the brightness temperature (Figure 6d–f). A turbulence frequency maximum is evident for temperature differences around 0 K, where the reports are located close to the cloud deck. The separation of the reports based on the temperature difference threshold of −10 K achieves results comparable with the previous approach, which was based solely on the brightness temperature.

Figure 7 compares the DJF turbulence and diagnostic frequency maps for reports with $T_b > 235$ K with those with $T_b \leq 235$ K. For turbulence in the absence of high

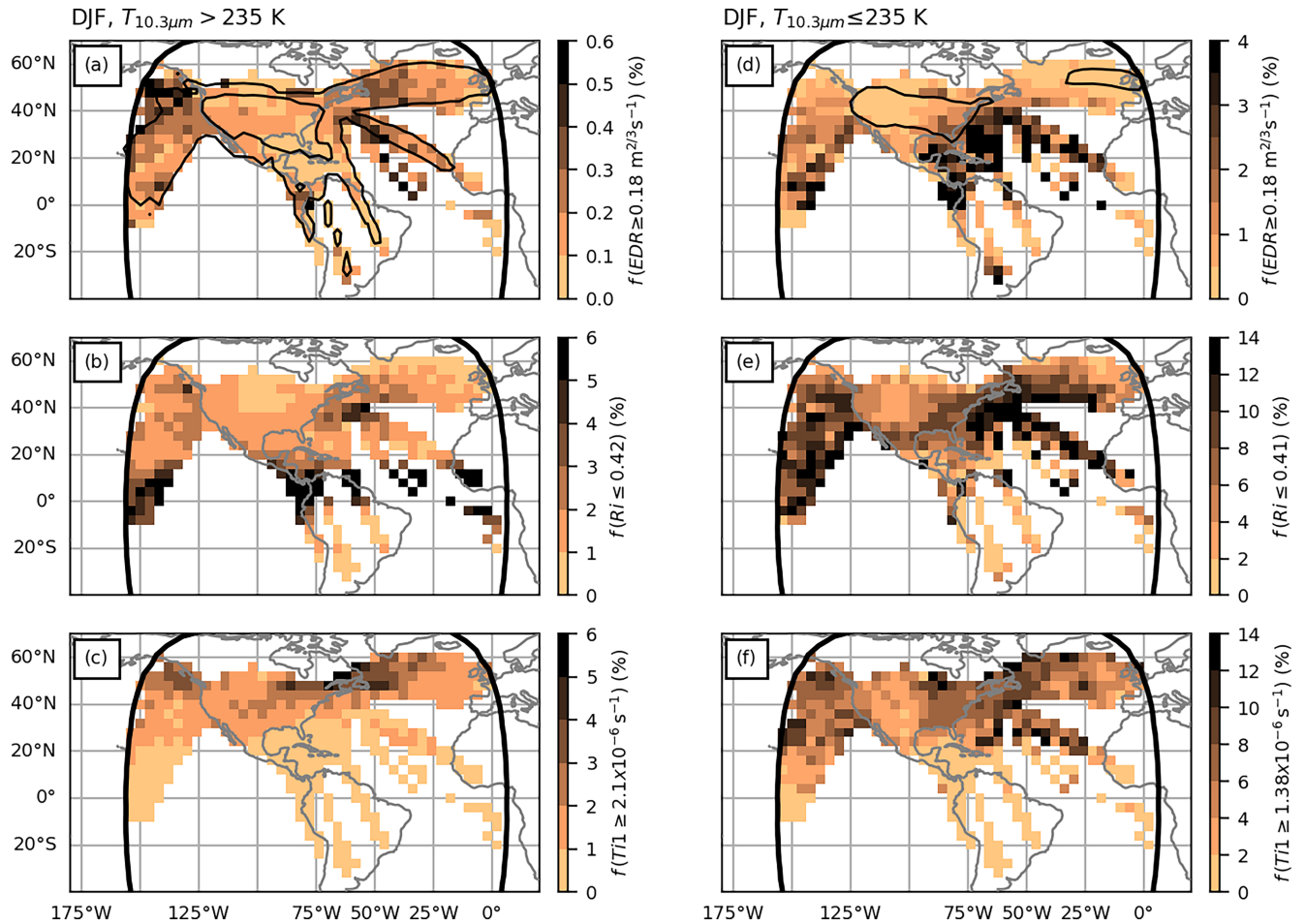

FIGURE 7 Relative frequency maps for turbulence reports between 8 km and z_{LRT} during DJF. Left panels for reports with *GOES 16* ABI channel-13 brightness temperature $T_b > 235$ K, for (a) EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$, (b) Ri ≤ 0.42, and (c) TI1 $\geq 2.10 \times 10^{-6}\ s^{-1}$. Right panels for reports with brightness temperature $T_b \leq 235$ K, for (d) EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$, (e) Ri ≤ 0.41, and (f) TI1 $\geq 1.38 \times 10^{-6}\ s^{-1}$. Black contour lines in (a) and (d) show report counts of 10^4, 10^5, and 10^6. The display threshold for $T_b > 235$ K and $T_b \leq 235$ K is 5×10^3 reports in sum from both subsets. [Colour figure can be viewed at wileyonlinelibrary.com]

clouds ($T_b > 235$ K), the observed turbulence frequencies (Figure 7a) are generally enhanced over the oceans compared with over land. The DJF frequency map for the 99th percentile of the Richardson number (Ri ≤ 0.42, Figure 7b) exhibits pronounced maxima in the Tropics, but shows a weaker maximum over the North Atlantic compared with the frequency distribution for all reports (Figure 2b). The frequency map for the 99th percentile of TI1 (TI1 $\geq 2.10 \times 10^{-6}$ s^{-1}, Figure 7c) remains largely unchanged relative to Figure 2c.

In contrast, turbulence in the vicinity of high clouds ($T_b \leq 235$ K) occurs most frequently at latitudes south of 40°N over the North Atlantic and along tropical flight routes (Figure 7d). In agreement with Figure 6c and compared with $T_b > 235$ K, the report count decreases, whereas turbulence frequencies increase. High clouds are associated with overall reduced Richardson numbers in ERA5, which is reflected by the fact that the 95th percentile of the Richardson number for $T_b \leq 235$ K (Ri $\leq$ 0.41) is comparable with the 99th percentile threshold used for $T_b > 235$ K (Ri ≤ 0.42). The enhanced turbulence frequencies also lead to a comparable classification skill between Ri ≤ 0.42 for $T_b > 235$ K and Ri ≤ 0.41 for reports with $T_b \leq 235$ K (Figure 6c), which is why we select the 95th percentile to analyze reports near high clouds. Frequencies of Ri ≤ 0.41 are enhanced where turbulence is observed most frequently (Figure 7d,e) but also extend into other regions, such as the central North Atlantic and large areas of the North Pacific (within the *GOES* field of view). Similarly, the occurrence frequencies for the 95th percentile of TI1 (TI1 $\geq 1.38 \times 10^{-6}$ s^{-1}) are enhanced over large portions of both the North Atlantic and the North Pacific.

5 | DISCUSSION

The planetary-scale frequency maxima for EDR $\geq$ 0.18 m$^{2/3} \cdot$ s^{-1} and their seasonality (Figures 2a and 4a–c) appear to align with the underlying large-scale dynamics of the atmospheric circulation, such as baroclinic wave activity over winter storm tracks, deep convection for the springtime continental maximum over eastern North America, and the location of the Intertropical Convergence Zone (ITCZ). Although deep convection and strongly sheared flow in baroclinic waves are sources of upper-tropospheric turbulence, this study highlights that caution is advisable when inferring the distribution of the infrequent occurrence of EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1} from frequently occurring flow features that characterize a geographic region.

Turbulence with EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1} in the upper troposphere is likely more common than the turbulence reports suggest, as airlines actively seek to avoid turbulence. Turbulence avoidance relies on onboard and ground-based radar to identify convection, on pilot and automated reports of turbulence encountered from other aircraft, and on turbulence forecasts from weather models, such as the Graphical Turbulence Guidance (GTG: Sharman & Pearson, 2017). Regional variations in observational coverage and differences in the proportion of easily avoidable turbulence could influence not only the overall turbulence frequency scale but also the internal consistency of the turbulence frequency maps (Figures 2a and 4a–c). Turbulence is more likely to be avoided when it is well forecast (e.g., by the GTG), visually identifiable (e.g., convective storms), or spatially limited (e.g., terrain-induced turbulence near mountain ranges). However, it remains uncertain to what extent the effect of turbulence avoidance on the representativeness of turbulence reports can be quantified. The agreement between the turbulence diagnostics along the flight tracks and at all upper-tropospheric grid points in ERA5 can only be a rough proxy, considering the identified limitations of the diagnostics for representing EDR ≥ 0.18 m$^{2/3} \cdot$ s^{-1}.

For the automated turbulence reports, EDR is derived solely from measurements of the vertical wind component. This approach assumes that the measured vertical velocity fluctuations are representative of the entire turbulent spectrum, which is only valid under the assumption of isotropy. However, atmospheric turbulence is not generally isotropic, particularly in stably stratified or strongly sheared flows, so EDR estimates may be biased toward certain flow structures or instability types.

The separation of reports based on cloud brightness temperature highlights interesting differences, most notably the increased turbulence frequencies near high clouds and the more widespread geographic distribution of the 95th percentile of TI1 compared with where TI1 maxima generally occur. However, in terms of classification skill, both the Richardson number and TI1 achieve a non-negligible POD only at thresholds associated with considerable overprediction, meaning that the observed turbulence distribution cannot be inferred directly from the diagnostic frequency maps.

Given the limited turbulence-report climatologies available in the literature, we demonstrate that these reports can be used to generate seasonal turbulence-frequency maps, provided that the report-downlink procedure and irregular sampling of the less turbulent stratosphere are accounted for. Accordingly, we present a turbulence climatology for one EDR threshold and the upper troposphere, and we emphasize that the effect of turbulence avoidance is not easily quantified. In light of the variety of available turbulence diagnostic

climatologies, we show that the interpretation of seasonal ERA5 turbulence diagnostic frequency maps with regards to EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$ is challenging at least. The interpretation of the diagnostic maps is difficult because of the infrequent occurrence of EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$, the exponential trade-off between POD and precision for individual diagnostics, and the regional variability of these metrics. The separation of reports based on cloud brightness temperature reveals that, under low brightness temperatures, turbulence frequencies increase by more than a factor of five, background-flow Richardson numbers are lower, and both the turbulence and diagnostic frequency maps exhibit markedly different geographic distributions. However, the turbulence distribution still cannot be inferred from the diagnostic distribution, as the sharp trade-off between POD and precision remains.

6 | SUMMARY

Seasonal frequency maps of upper-tropospheric turbulence have been derived from aircraft reports and compared with seasonal frequency maps of the 99th percentile of the Richardson number and TI1 in the ERA5 reanalysis, under consideration of the predictive skill of the two diagnostics. The turbulence frequency maps are based on eight years (January 1, 2017–December 31, 2024) of automated EDR turbulence reports from commercial aircraft, provided by the NOAA MADIS archive. Internal sampling biases have been taken into account by only considering archived reports with a consistent one-minute sampling frequency along flight tracks, and by deriving turbulence frequencies for EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$ as the lowest turbulence intensity in the reports without a downlink bias.

The seasonal turbulence frequency maps for EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$ show contiguous planetary-scale frequency maxima, as far as the geographic coverage of the US-based airlines that contribute to the MADIS archive allows. In the midlatitudes, the analyzed turbulence intensity is observed most frequently over the North Atlantic and North Pacific storm tracks during winter. North America shows comparatively low turbulence frequencies throughout the year, except for a widespread increase in the east during spring. At tropical latitudes, turbulence frequency maxima are observed over the Pacific and the Atlantic, varying in latitude and magnitude throughout the year.

The seasonal frequency maps of the 99th percentile of the Richardson number largely mirror the regional and seasonal patterns observed in the turbulence distribution, whereas the 99th percentile of TI1 predicts turbulence predominantly north of the observed winter storm-track turbulence frequency maxima. Overprediction is inherent to the 99th percentile approach, as EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$ occurs only 0.15% of the time according to the turbulence reports. Despite the overprediction, only about 20% of all observed turbulence is associated with each of the diagnostics at the given thresholds. Adjustment of the diagnostic threshold can increase the POD, but at the cost of exponentially decreasing precision. Representative precision is therefore only given for small fractions of the observed turbulence, if at all. The precision and the POD of both diagnostics vary regionally, which complicates the interpretation of the diagnostic frequency maps further. A separation of reports based on *GOES-16* satellite infrared brightness temperature shows that turbulence frequencies near high clouds increase by more than a factor of five, with a distinct maximum south of 40°N over the North Atlantic. However, a non-negligible POD is again only achieved at the expense of significant overprediction by the diagnostics. So, while EDR $\geq 0.18\ m^{2/3} \cdot s^{-1}$ is generally observed most frequently in regions where small Richardson numbers are diagnosed most frequently in ERA5, the rare occurrence of the analyzed turbulence intensity remains hidden in the diagnostic frequency maps.

ACKNOWLEDGEMENTS

This study was made possible in part due to the data made available to the National Oceanic and Atmospheric Administration by the following commercial airlines: American, Delta, Federal Express, Northwest, United, and United Parcel Service. This work used JASMIN, the UK's collaborative data analysis environment (https://jasmin.ac.uk, last accessed July 31, 2025). We used brightness temperature data from the Advanced Baseline Imager (ABI) on board the *GOES-16* satellite, provided by the NOAA National Environmental Satellite, Data, and Information Service (NESDIS).

CONFLICT OF INTEREST STATEMENT

The authors report no conflicts of interest.

DATA AVAILABILITY STATEMENT

ECMWF's ERA5 data have been accessed and processed from the NERC CEDA archive. ERA5 data can be accessed freely from https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5. ACARS EDR turbulence reports from the public NOAA MADIS archive are available at https://madis.ncep.noaa.gov (last accessed July 31, 2025). NOAA *GOES-16* was accessed in July 2025 from https://registry.opendata.aws/noaa-goes. The processed data that support the findings of this study are available from the corresponding author upon request.

ENDNOTES

[1]In the literature on the turbulence reporting algorithm, a report sometimes refers to a set of one-minute EDR measurements that are downlinked in a bundle.

[2]The negative Richardson numbers are not critical for the initial offset from TI1, aside from very low POD. The logarithmic scale in Figure 3c can be deceiving here.

ORCID

Thorsten Kaluza https://orcid.org/0000-0003-3700-389X
Paul D. Williams https://orcid.org/0000-0002-9713-9820
David M. Schultz https://orcid.org/0000-0003-1558-6975

REFERENCES

Berthet, G., Esler, J.G. & Haynes, P.H. (2007) A Lagrangian perspective of the tropopause and the ventilation of the lowermost stratosphere. *Journal of Geophysical Research: Atmospheres*, 112, D18102. Available from: https://doi.org/10.1029/2006JD008295

Dutton, J.A. & Panofsky, H.A. (1970) Clear air turbulence: a mystery may be unfolding. *Science*, 167, 937–944. Available from: https://doi.org/10.1126/science.167.3920.937

Ellrod, G.P. & Knapp, D.I. (1992) An objective clear-air turbulence forecasting technique: verification and operational use. *Weather and Forecasting*, 7, 150–165. Available from: https://doi.org/10.1175/1520-0434(1992)007<0150:AOCATF>2.0.CO;2

Forster, P. & Shine, K.P. (1997) Radiative forcing and temperature trends from stratospheric ozone changes. *Journal of Geophysical Research: Atmospheres*, 102, 10841–10855. Available from: https://doi.org/10.1029/96JD03510

Grams, C.M., Wernli, H., Böttcher, M., Čampa, J., Corsmeier, U., Jones, S.C. et al. (2011) The key role of diabatic processes in modifying the upper-tropospheric wave guide: a North Atlantic case-study. *Quarterly Journal of the Royal Meteorological Society*, 137, 2174–2193. Available from: https://doi.org/10.1002/qj.891

Gultepe, I., Sharman, R., Williams, P.D., Zhou, B., Ellrod, G., Minnis, P. et al. (2019) A review of high impact weather for aviation meteorology. *Pure and Applied Geophysics*, 176, 1869–1921. Available from: https://doi.org/10.1007/s00024-019-02168-6

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049. Available from: https://doi.org/10.1002/qj.3803

Hoor, P., Gurk, C., Brunner, D., Hegglin, M.I., Wernli, H. & Fischer, H. (2004) Seasonality and extent of extratropical TST derived from in-situ CO measurements during SPURT. *Atmospheric Chemistry and Physics*, 4, 1427–1442. Available from: https://doi.org/10.5194/acp-4-1427-2004

Jaeger, E.B. & 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), 2006JD008189. Available from: https://doi.org/10.1029/2006JD008189

Ko, H.C., Chun, H.Y., Sharman, R.D. & Kim, J.H. (2023) Comparison of eddy dissipation rate estimated from operational radiosonde and commercial aircraft observations in the United States. *Journal of Geophysical Research: Atmospheres*, 128, e2023JD039352. Available from: https://doi.org/10.1029/2023JD039352

Lee, J.H., Kim, J.-H., Sharman, R.D., Kim, J. & Son, S.-W. (2023) Climatology of clear-air turbulence in upper troposphere and lower stratosphere in the Northern Hemisphere using ERA5 reanalysis data. *Journal of Geophysical Research: Atmospheres*, 128, e2022JD037679. Available from: https://doi.org/10.1029/2022JD037679

Martius, O., Schwierz, C. & Davies, H.C. (2010) Tropopause-level waveguides. *Journal of the Atmospheric Sciences*, 67, 866–879. Available from: https://doi.org/10.1175/2009JAS2995.1

Richardson, L.F. (1922) *Weather prediction by numerical process*. Cambridge: Cambridge University Press. Available from: https://archive.org/details/weatherpredictio00richrich

Riese, M., Ploeger, F., Rap, A., Vogel, B., Konopka, P., Dameris, M. et al. (2012) Impact of uncertainties in atmospheric mixing on simulated UTLS composition and related radiative effects. *Journal of Geophysical Research: Atmospheres*, 117, D16305. Available from: https://doi.org/10.1029/2012JD017751

Sharman, R.D. & Pearson, J.M. (2017) Prediction of energy dissipation rates for aviation turbulence. Part I: forecasting nonconvective turbulence. *Journal of Applied Meteorology and Climatology*, 56, 317–337. Available from: https://doi.org/10.1175/JAMC-D-16-0205.1

Sharman, R.D., Tebaldi, C., Wiener, G. & Wolff, J. (2006) An integrated approach to mid- and upper-level turbulence forecasting. *Weather and Forecasting*, 21, 268–287. Available from: https://doi.org/10.1175/WAF924.1

Sharman, R.D., Trier, S.B., Lane, T.P. & Doyle, J.D. (2012) Sources and dynamics of turbulence in the upper troposphere and lower stratosphere: a review. *Geophysical Research Letters*, 39(12), 2012GL051996. Available from: https://doi.org/10.1029/2012GL051996

Sharman, R.D., Cornman, L.B., 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, 1416–1432. Available from: https://doi.org/10.1175/JAMC-D-13-0329.1

Smith, I.H., Williams, P.D. & Schiemann, R. (2023) Clear-air turbulence trends over the North Atlantic in high-resolution climate models. *Climate Dynamics*, 61, 3063–3079. Available from: https://doi.org/10.1007/s00382-023-06694-x

Spreitzer, E., Attinger, R., Boettcher, M., Forbes, R., Wernli, H. & Joos, H. (2019) Modification of potential vorticity near the tropopause by nonconservative processes in the ECMWF model. *Journal of the Atmospheric Sciences*, 76, 1709–1726. Available from: https://doi.org/10.1175/JAS-D-18-0295.1

Williams, P.D. & Joshi, M.M. (2013) Intensification of winter transatlantic aviation turbulence in response to climate change. *Nature Climate Change*, 3, 644–648. Available from: https://doi.org/10.1038/nclimate1866

WMO. (1957) Meteorology – a three-dimensional science: second session of the commission for aerology. *World Meteorological Organization Bulletin*, 4, 134–138. Available from: https://www.scribd.com/document/601621440/1957-WMO-bulletin-vi-4

WMO. (2013) *Instruments and observing methods report no. 114: amdar onboard software functional requirements specification instruments and observing methods. technical report*. Geneva: World Meteorological Organization. Available from: https://library.wmo.int/idurl/4/50064

Wolff, J.K. & Sharman, R.D. (2008) Climatology of upper-level turbulence over the contiguous United States. *Journal of Applied Meteorology and Climatology*, 47, 2198–2214. Available from: https://doi.org/10.1175/2008JAMC1799.1

How to cite this article: Kaluza, T., Williams, P.D., Schultz, D.M. & Banyard, T.P. (2025) Climatology of upper-tropospheric turbulence: Capabilities and limitations of aircraft reports and ERA5 reanalysis diagnostics. *Quarterly Journal of the Royal Meteorological Society*, e70073. Available from: https://doi.org/10.1002/qj.70073

APPENDIX A. LOGNORMAL FIT FOR THE EDR DISTRIBUTION

As a quality and consistency check, we have derived a lognormal fit for the EDR distribution from all upper tropospheric ACARS reports from January 1, 2017–December 31, 2024 (Figure 3a), following the method described in Sharman *et al.* (2014). We have adjusted the EDR range for the optimization of the fit parameter from 0.10–0.64 $m^{2/3} \cdot s^{-1}$ to 0.18–0.64 $m^{2/3} \cdot s^{-1}$. EDR values between 0.02 and 0.16 $m^{2/3} \cdot s^{-1}$ are underrepresented because of the reporting technique, as only turbulence intensities EDR $\geq$ 0.18 $m^{2/3} \cdot s^{-1}$ are downlinked consistently. The underestimation of low EDR is evident in Figure 3a. A similar fall-off for low EDR values is evident in the EDR probability distribution functions in Sharman *et al.* (2014) and Ko *et al.* (2023).

The lognormal fit parameters in this study ($\mu = -3.61$, $\sigma = 0.63$) agree more with the United Airlines distribution in Sharman *et al.* (2014) ($\mu = -3.83$, $\sigma = 0.79$) compared with the Delta Airlines distribution ($\mu = -2.85$, $\sigma = 0.57$). This agreement is expected, as we only use flight data with a regular one-minute sampling frequency from regular, triggered, and interpolated reports. Together with the exclusion of underestimated low EDR values during the lognormal fit optimization, this approach is comparable with the regular one-minute United Airlines reporting technique.