Multidecadal Measurements of UTLS Gravity Waves Derived From Commercial Flight Data

C. J. Wright1 and T. P. Banyard1

1Centre for Space, Atmospheric and Oceanic Science, University of Bath, Bath, UK

Abstract. Gravity waves (GWs) are key drivers of atmospheric dynamics, with major impacts on climate and weather processes. However, they are challenging to measure in observational data, and as a result no large-area multidecadal GW time series yet exist. This has prevented us from quantifying the interactions between GWs and long-timescale climate processes. Here, we exploit temperatures measured by commercial aircraft since 1994 as part of the In-Service Aircraft for a Global Observing System (IAGOS) atmospheric chemistry research program to produce a novel 26-year time series of upper troposphere/lower stratosphere (UTLS) GW measurements across most of the Northern Hemisphere. We analyze 90,342 flight hours (76.2 million flight kilometers) of data, typically at a temporal resolution of seconds and with high temperature precision. We show that GW activity in the Northern Hemisphere UTLS is consistently strongest north of and poleward of the upper tropospheric jet. We also show that GW sources not typically observed in stratospheric data but assumed in model schemes, such as the Rocky Mountains, are visible at these altitudes, suggesting that wave momentum from these sources is deposited specifically between \( \sim 200 \) and \( 50 \) hPa. Our data show no significant impact of the Quasi-Biennial Oscillation, the Northern Annular Mode, or climate change. However, we do see strong evidence of links with the El Niño–Southern Oscillation, which modulates the measured GW signal by \( \sim 25\% \), and weak evidence of links with the 11-year solar cycle. These results have important implications for atmospheric process modeling and for understanding large-scale climate teleconnections.

Plain Language Summary. We use 26 years of temperature measurements, made by commercial aircraft flying routine passenger and freight routes, to measure and understand small-scale atmospheric waves (“gravity waves”). These waves play many vital roles in the atmosphere, and they are critically important to understanding and predicting weather and climate. However, due to their small size, they are extremely difficult to study at typical flight altitudes using any other method. Our data provide the longest large-area record of gravity waves from any dataset, with 76 million flight kilometers of measurements made since 1994. We see strong evidence that El Niño affects the behavior of these small waves over northern Europe and central Asia, many thousands of kilometers away from the Pacific Ocean. We also see weak evidence that the waves may be affected by the 11-year solar cycle. Our results provide a unique window on the interactions between small-scale atmospheric behavior and long–time scale climate dynamical processes and have important implications both for improving atmospheric models and for understanding long-distance climate teleconnections.

1. Introduction

In-Service Aircraft for a Global Observing System (IAGOS) is a European research infrastructure program designed to observe the upper troposphere and lower stratosphere (UTLS) using instruments fitted as an additional payload to commercial aircraft. IAGOS data are well validated and have been used in over 400 scientific studies to date. Including flights operated under the earlier MOZAIC and CARIBIC programs, the data set contains \( >60,000 \) flights, covering the period from 1994 onward.

While the primary focus of IAGOS is on atmospheric chemistry, the metadata recorded to contextualize these measurements are also of significant scientific use. These metadata provide a long-term high-resolution record of atmospheric temperatures and winds at small spatial scales, with high precision and good quality control. They have previously been used in a number of dynamical studies (Berkès et al., 2017; Callies et al., 2014, 2016; Cho & Lindborg, 2001; Li & Lindborg, 2018; Lindborg, 1999; Scott et al., 2001; Skamarock,
Here, we exploit these metadata to study geographically resolved small-scale atmospheric gravity waves (GWs) in temperature data. GWs are near-ubiquitous in the middle and upper atmosphere and play a vast range of vital dynamical roles, including controlling the stratospheric jets and the large-scale equator-to-pole and pole-to-pole circulations, driving chemical transport and ozone depletion, and acting to help stratospheric clouds form (Alexander et al., 2010, 2013; Fritts & Alexander, 2003; Geller et al., 2013; Murgatroyd & Singleton, 1961). However, due to their small physical scales (at most tens of kilometers and usually much less in the vertical, and tens to hundreds of kilometers in the horizontal), they are technically challenging to measure, and as a result few long GW time series exist.

The challenges of measuring GWs are particularly acute in the UTLS. Ground-based instruments and radiosonde balloons provide limited spatial coverage, particularly over the open ocean, while scientific aircraft and superpressure balloon campaigns are limited in number and coverage and do not take routine measurements. To add to this, the satellite methods that have revolutionized our stratospheric GW knowledge in the last two decades perform extremely poorly in the UTLS, due to their relatively coarse spatial resolution and the vertical structure of the temperature-tropopause itself, which heavily complicates use of profile-based detrending methods for extracting GW signatures from background variations. As such, IAGOS data can provide a unique long-term window on UTLS GWs, particularly at remote locations.

Callies et al. (2016) applied spectral decomposition methods to MOZAIC wind data from 2002–2010, demonstrating that the data contained signatures consistent with inertia GWs in the lower stratosphere. However, they did not identify such a relationship in the upper troposphere, a conclusion they ascribed to mesoscale variances in measured along-track wind components. Here, we use temperature data from a 1994–2019 superset of IAGOS data; these temperatures do not appear to share the along-track variance issues and contain clear GW signatures amenable to spectral analysis in both the UTLS.

In section 2 we describe the data used, and in section 3 the analyses performed. Section 4 then examines seasonal maps of the Northern Hemisphere, considering both directly measured geophysical variables (i.e., wind and temperature) and our analysis outputs. In section 5 we study time series for 18 selected regions, which are later tested against multiple climate indices in section 7. Section 6 then examines the relationship between GWs and the upper tropospheric jet, while section 8 discusses our results in the context of Rocky Mountain-driven GW forcing. Finally, we consider systematic biases and deficiencies in our data set in section 9, before summarizing and drawing conclusions in section 10.

We further include three appendices. Appendices A and B provide further technical detail on our methodological approach, while Appendix C contextualizes our results in relation to previous UTLS GW studies.

2. IAGOS Data

We use data from the IAGOS Data Portal (Boulanger et al., 2019), incorporating all final-version recorded flights from 1 August 1994, the start of the data set, to the end of 2019. A significant lag often exists between data being recorded and finalized, and accordingly from October 2017 onward our data set does not include all IAGOS flights. The proportion of such unprocessed data missing from our analysis scales approximately linearly with time, from ~5% in late 2017 to >95% in the last three months of 2019.

We specifically use aircraft-instrument-derived temperature data (“air_temp_AC”). These data are recorded by the aircraft’s standard instrumentation rather than the IAGOS package and thus may be less accurate, but are available for every flight. Using these data approximately doubles available data relative to temperatures measured by the instrument packages alone, with the increase largest in later years.

Since we detrend the data to identify GWs, absolute biases and drifts over long periods of flight are not important to our analysis, thus potentially justifying use of these data. To test this, Figure 1 compares “air_temp_AC” against three other IAGOS temperature products, computed over all level-flight measurements for which each variable-pair exists in January 2000, detrended and preprocessed as described below. We see close agreement, with correlations of >0.95 and linear fits. There is a bias between stagnation-temperature (i.e., temperatures at zero fluid speed) and directly measured temperature products; this bias is symmetric about zero perturbation and thus will lead to a reduction ~10% in our measured wave amplitudes relative to if stagnation products were used.

Other randomly tested months were analyzed in the same way, with equivalent results.
Figure 1. Comparison of perturbations in aircraft-measured temperature (“air_temp_AC”) against perturbations in (a) MOZAIC-measured temperature (“air_temp_PM”), (b) aircraft-measured stagnation temperature (“air_stag_temp_AC”), and (c) MOZAIC-measured stagnation temperature (“air_stag_temp_PM”) for corresponding measurements in each data set for the month of January 2000. Each panel shows a scatter-density plot of (vertical axis) “air_temp_AC” against (horizontal axis) each other product, with a bin size of 0.05 K on each axis. Close agreement is seen, justifying use of the aircraft-measured temperature data for our GW analyses.

3. Methods

3.1. Preprocessing of IAGOS Data

IAGOS data include ascents and descents as well as level cruising. The data also vary in resolution and occasionally include minor errors. As such, some preprocessing is required to provide consistent data for analysis. Our preprocessing chain, which is described in detail in Appendix A, produces data which varies to only a limited degree in height and travel direction, is linearly spaced with a uniform point separation of 1 km, and only contains continuous segments (hereafter “cruises”) of at least 1,000-km length.

After additional removal of out-of-band data (section 3.2), we are left with 76.2 million 1-km-spaced data points, distributed as shown in Figure 2 and representing 90,342 flight hours of data. Our data are biased toward the Northern Hemisphere, especially the northern Atlantic Ocean, and skew toward higher altitudes in later years. We divide the data into three pressure bands (>225, 225–205, and <205 hPa, respectively) to contextualize relative coverage as a function of height; these bands are chosen to split the data into three approximately equal subsets, rather than on the basis of physical differences, and are not used below.

Figures 3a–3d show the broad characteristics of the resulting data set. Temperatures and zonal winds show smooth distributions with positive skew, while meridional winds show a near-symmetric distribution focused around 0. Pressures show a spiky distribution (note the logarithmic ordinate), consistent with preferred flight levels.

3.2. GW Analysis

To identify GWs, we use a one-dimensional Stockwell transform (ST) method. Such methods have been widely used in previous GW research (e.g., M. J. Alexander et al., 2008; Hindley et al., 2015; Moss et al., 2016; Wright et al., 2010) and allow us to localize measured GWs in both location and wavenumber.

We first smooth each cruise with a 900-km second-order Savitzky-Golay filter and subtract this from the original data to produce perturbation series. Figure 6 of Hindley et al. (2015) shows the transfer function for this filter, which here provides good transmittance up to ~500 km, dropping off sharply above this. Callies et al. (2014) saw a change in horizontal dynamical scales at ~500 km, supporting this choice. We then spectrally analyze each cruise, using the ST implementation of Hindley et al. (2019). From this we estimate wave amplitude $A$ and along-flight horizontal wavenumber $k$ for the strongest signal present at each data point (see, e.g., M. J. Alexander et al., 2008). We restrict the analysis to wavelengths between 10 and 600 km.

Figures 3e and 3f show the resulting amplitudes and wavelengths. Both distributions are smooth, with amplitudes centered at 0.7 K and wavelengths peaking at 500 km. A maximum is seen in the lowest two wavelength bins at ~20 km, arising from small-scale noise aliased into the analyzed range. To prevent this spurious peak contaminating our results, we remove data with wavelengths below 25 km. We also remove wavelengths longer than 500 km, as they will be attenuated by the detrending filter and in any case are expected to be...
Figure 2. Data points available for our analysis after filtering, distributed in (a–d) time and (e–h) space and subdivided into (a and e) vertically below 225 hPa (b and f) 225–205 hPa (c and g) vertically above 205 hPa, and (d and h) all data. These height bands are chosen purely to provide a roughly even split and have no especial physical significance. Numerical values in (a)–(d) are in thousands of data points, with column (row) totals shown at the end of the column (row). Values in (e)–(h) are as a percentage of the global total in that height range, using 3° × 3° bins.

dynamically different. Finally, we remove points with measured wavelengths more than twice the distance between the measurement and the start or end of the cruise; this reduces edge-truncation effects in our results but will introduce a short-wavelength bias near the start and end of flights. This effect will be largest over Europe, where most flights commence or terminate. The removed data are distributed approximately evenly across the other distributions in Figure 3, and removing them does not affect their form.

Sensitivity tests suggest our results are only weakly affected by the choice of a 1-km interpolant or the definition of a cruise. The choice of detrending filter also has only a small effect: boxcar, Lanczos, and Gaussian filters were also tested, with very little difference for the boxcar or Lanczos and, for the Gaussian, a shift to slightly higher amplitudes and longer wavelengths but with otherwise similar results. However, our results are strongly affected by the 900-km filter length. In particular, tests with significantly longer filters produce large peaks in wavelength over the midpoints of typical long-haul flights, that is, the mid-Atlantic and central Asia, a bias which is avoided by using our smaller value.

3.3. Cluster Mapping

3.3.1. Hierarchical Cluster Analysis

Due to the uneven distribution of our measurements, balancing the limited data in many regions with dense coverage in others is nontrivial. Using a high-resolution grid leaves large gaps, while a low-resolution grid underrepresents available resolution, leading to the discard of large areas when we later impose a minimum number of measurements per grid box.

To balance these competing scales, we use an approach combining hierarchical clustering and statistical bootstrapping. This is based on the work of Wright et al. (2012, 2013) and produces maps with good spatial coverage over most of the Northern Hemisphere, while still exploiting the full resolution where measurement density is high. We carry out this analysis at the seasonal level, using combined data from all years for
Figure 3. Histograms of basic data properties, integrated across the entire data set. (a)–(d) show directly measured temperature, zonal wind speed, meridional wind speed and pressure. (e) and (f) show derived wave amplitudes and wavelengths.

boreal winter (December–February, DJF), spring (March–May, MAM), summer (June–August, JJA), and autumn (September–November, SON) respectively. Further details are provided in Appendix B.

3.3.2. Bootstrap Statistics
Due to the inhomogeneous data distribution, while these spatial clusters represent both large and small scales well, they contain highly uneven numbers of measurements, with a spread of 3 orders of magnitude in population (compared to 5 orders of magnitude using a regular 3° grid). While suitable for mapping, this may introduce spurious intercluster variations in the statistics we compute, namely, the median and Gini’s coefficient of concentration $G$ (used previously for GW studies in, e.g., Hindley et al., 2019; Kuchar et al., 2020; Plougonven et al., 2012; Wright et al., 2013). Means were also calculated for all clusters and showed spatially consistent results but with spike values in some clusters due to outliers.

To correct for these possible variations, we bootstrap our statistical calculations within each cluster. We first sample 500 sets of 1,000 randomly chosen measurements from each cluster, allowing replacement of the same value. We then compute our statistics independently for each set, recording the median value of this 1,000-point distribution. This median is used as our overall statistical estimate for the cluster properties. This allows us to compute comparable statistics for each cluster, which are not massively biased by their different populations. To test sensitivity, these numerical values (i.e., 500 and 1,000) were varied systematically across an order of magnitude in each direction, with only small changes in the final results.

3.3.3. Final Mapping
Finally, we bin the statistics onto a 0.5° × 0.5° grid to facilitate mapping. To do this, we iterate over each individual grid box, finding the cluster with the geographically closest center. We then assign the statistics computed for that cluster to the grid box and move on to the next. If the nearest cluster center is $>300$ km away, then we leave the grid box blank.

Figures 4a and 4b show maps of the cluster-areas generated for DJF using data from all heights. We see a dense patchwork over the North Atlantic and western Europe, with larger clusters in other regions. This is representative of the original data distribution (Figure 2). Figure 4c compares the area assigned to each cluster with the midlatitude area of a regular grid box (dashed lines); most clusters represent an area equivalent to a 2° grid or smaller, with almost all representing an area less than a 5° grid.
3.4. Tropopause Calculation

For context, we calculate the tropopause pressure level using daily ERA5 reanalysis output (Hersbach et al., 2020), at full vertical resolution but downsampled to a spatial resolution of 1.5° × 1.5°. For each grid point, we find the lowest altitude where (a) the vertical temperature lapse rate is <2 K km⁻¹ and (b) this remains true for at least 2 km above this point. We require the tropopause to lie between 700 and 10 hPa, linearly interpolating to find the transition below 2 K km⁻¹ as accurately as possible between model levels. Any grid points where these criteria are not met are bilinearly interpolated from surrounding grid points.

4. Cluster-Generated Maps

Figure 5 shows seasonal maps of median temperatures, zonal winds, and meridional winds between 200 and 250 hPa and the number of measurements assigned to each cluster. Figure 6 shows differences between these data and the ERA5 reanalysis over the same pressure range and period. At this altitude and over this large a height range and time period, we expect the reanalysis fields to reasonably approximate true values and therefore allow us to identify systematic biases created by our sampling pattern. Figure 7 shows GW amplitudes (top rows), G (middle rows), and along-track GW wavelengths (bottom rows) over the same pressure range. We focus on the region 20–85°N, 130°W to 140°E, due to good sampling in this region.

The maps (except Figures 5y–5B) have been smoothed using a 2-D Gaussian of full width at half maximum (FWHM) 2.5° × 2.5° to facilitate analysis of large-scale features. In the unsmoothed data, not shown, small-area orographic wave sources can be clearly identified, especially in the smaller clusters over Europe and the North Atlantic, and these will be studied in more detail in future work.

4.1. Temperature, Wind, and Measurement Density

4.1.1. Absolute Value Maps

Figure 5 shows directly measured variables. These data have been processed identically to the GW data discussed below, to help understand sampling biases.
Figure 5. Cluster-analysis derived maps of IAGOS-measured (a–d) temperature, (e–h) standard deviation of temperature, (i–l) zonal wind speed, (m–p) standard deviation of zonal wind speed, (q–t) meridional wind speed, (u–x) standard deviation of meridional wind speed, and (y–B) number of measurements, computed over the pressure range 150–350 hPa. Figure combines data from all years of IAGOS data. Solid gray shading indicates insufficient data density for analysis, according to the criteria outlined in section 3.3.
Temperatures (Figures 5a–5d) are low in DJF and high in JJA with a negative meridional gradient from tropics to pole. Standard deviations (Figures 5e–5h) are ~4% or less, with higher internal variability in DJF/MAM and lower internal variability in SON/JJA.

Zonal winds, Figures 5i–5l, are strongly positive over the western Atlantic all year and over northern Africa/southern Asia for most of the year, with weaker winds in JJA and stronger in DJF. Winds are moderately strong across most of western North America, especially compared to Eurasia poleward of ~40°N. Standard deviations (Figures 5m–5p) are high over the North Atlantic, consistent with subseasonal variability in jet location. Additional standard deviation peaks are seen east of Asia in MAM, over central Asia in JJA, and over the central United States in DJF and SON; all of these regions also exhibit high absolute values.

Meridional winds (Figures 5q–5t) are smaller than zonal, consistent with Figure 3c. Except for JJA, they are negative over western North American, Siberia, the edges of Scandinavia, and China and positive over the North Atlantic, the eastern coast of Russia, and to a lesser degree the Middle East and Iran. JJA exhibits a different morphology, with weaker and less spatially consistent winds over the North Atlantic, positive winds over the Eastern Mediterranean, and positive winds over Siberia. Standard deviations (Figures 5u–5x) are largest over the North Atlantic and (again, except JJA) North America.

Measurement density (Figures 5y–5B) is greatest on direct great-circle flight paths between Europe and major extra-European population centers, particularly in North America. There is a slight reduction over mainland Europe relative to surrounding locations due to our removal of edge-truncated waves in this region.

4.1.2. Biases Relative to Reanalysis
Figure 6 shows differences from gridded ERA5 output, averaged over the 200- to 250-hPa range. We note clearly that these differences do not necessarily represent any inaccuracy in either ERA5 or the IAGOS data—even assuming both perfect observations and a perfect model, large differences will be seen due to the very different spatial and temporal sampling of the two data sets. In general, pointwise differences between IAGOS and reanalysis temperatures and winds should be small (e.g., Berkes et al., 2017), and therefore, we assume differences represent sampling biases rather than measurement inaccuracies, primarily arising due to aircraft routing optimizations.

Regional temperature differences (Figures 6a–6d) are 5 K or less, except for a lightly sampled region at high latitudes in MAM. Temperatures have a consistent positive bias, except for the far east of Asia and high latitudes near Greenland. Meridional wind differences (Figures 6i–6l) are generally patchy, with a positive bias over the North Atlantic jet and the central eastern Atlantic off Africa coast and a negative bias over Scandinavia and the North Sea. Zonal winds (Figures 6e–6h) are the most different field, with patchy biases in most places but with a persistent and large positive bias in the North Atlantic jet stream. This bias is consistent with aircraft flying in the peak of the jet traveling eastward and avoiding the jet going westward, leading to a net positive bias when averaged.

4.2. GWs
GW parameters are shown in Figure 7. Amplitudes (Figures 7a–7d) are generally largest in DJF and MAM and lower in SON and JJA, but with numerous exceptions. They are consistently low over the mid-Atlantic and consistently high over the North Atlantic and most land poleward of ~40°N. These regions represent a mix of possible wave sources.

Local maxima over southeastern Greenland, western Mongolia, the Sikhote-Alin mountains of Eastern Russia, and the Rocky Mountains of North America are located close to meridionally aligned ridges, which would be expected to generate wave activity when exposed to strong near-zonal surface winds (e.g., Bacmeister, 1993). Gini's coefficient $G$ is generally higher than the surrounding area at these locations (except Sikhote-Alin)—this is consistent with orographic wave sources, which tend to be intermittent in nature (Wright et al., 2013).

Many of these orographic signatures are not routinely observed in lower stratospheric satellite data (e.g., Geller et al., 2013). This may suggest that wave momentum generated here is deposited specifically in the lowermost stratosphere. We further discuss the specific case of the Rockies in section 8.

The strongest clearly nonorographic feature in DJF and MAM is a band of high GW amplitude stretching east from Newfoundland. This band closely mirrors the North Atlantic winter storm track and is well resolved
Figure 6. Difference between ERA5 reanalysis and our measured (a–d) temperature, (e–h) zonal wind speed, and (i–l) meridional wind speed. Data are computed over the pressure range 150–350 hPa for the period August 1994 to December 2019 and are plotted such that a positive value indicates a positive bias in IAGOS relative to ERA5.

due to the high flight density over this region. Measured amplitudes in this band are likely to be skewed slightly high, due to the routing bias discussed above. This makes sense in the context of the data set: if GWs have the largest amplitudes near the jet, and flights benefit in fuel and travel time from flying nearby on eastward flight legs, then this will skew the measured value anywhere the jet is present during our data set high relative to other locations. This bias does not imply that there is not a maximum in this band but merely that we overestimate its relative magnitude. As such, this bias likely enhances the apparent importance of this nonorographic peak with respect to fixed-location orographic hot spots such as southeast Greenland.

$G$ (Figures 7e–7h) is generally highest around the Atlantic jet, consistent with the varying jet location. There is also higher $G$ over some orographic regions in winter, consistent with orographic source intermittency, and over large fractions of the observed area in general in JJA.

Wavelengths (Figures 7i–7l) are relatively short near the pole in all seasons. We believe this to be a bias arising from the sampling direction of the aircraft trajectories. Aside from this, some regional features can perhaps be seen, with short wavelengths over mountain ranges in DJF and MAM and (to a minor degree) in patches over the Atlantic Ocean corresponding to regions of high $G$. In general however, the known difficulties of accurately sampling GW wavelengths from randomly oriented 1-D or 2-D cuts through the atmosphere (e.g., Ern et al., 2004) make interpretation of these maps challenging.

We see no evidence of the strong subtropical GW maxima over the Caribbean, Africa, and Southeast Asia usually observed in stratospheric GW observations in JJA (e.g., Geller et al., 2013). We believe that, rather than these maxima not being present in the real atmosphere, this absence instead is due to aircraft routing biases. Specifically, if these maxima are to a significant extent driven by large-scale storm activity, as seems likely from theoretical, model, and satellite evidence, then planes will be routed around rather than through such activity. This will lead to a sharp reduction in relative magnitude in our data set. However, our available flight metadata do not provide sufficient information to test this hypothesis. The low amplitudes measured over the mid-Atlantic may also be related to such routing choices.
5. Regional Time Series

Figure 8 shows time series of GW amplitude for 18 selected regions, mapped at top right. Most of these regions have been selected visually from Figures 7a–7d as local maxima in at least one season, with J, M, and P added to fill remaining large spatial gaps in regions of good data coverage.

For each region, median GW amplitude is computed over a rolling 31-calendar-day window, composited over all years and over a geographic circle of radius 500 km indicated by the black circles surrounding the letters. This is done using the original flight-track data, rather than the cluster-analysis output.

Within each region, values above the annual median are shown in blue and below in red. The order of display (i.e., A–R.) is selected to emphasize systematic changes in seasonality: specifically, they are ordered by linear correlation with the Rocky Mountains (Region A), selected as (a) arbitrarily, it is the westernmost region and (b) subjectively, it has a clear seasonal cycle. Time series most similar to (A) are assigned earlier letters alphabetically. Sequences of alphabetically close letters on the map therefore indicate consistent GW-amplitude seasonality at supraregional geographical scales.

At bottom right, the spread of amplitudes in panels (A)–(R) is illustrated, allowing comparison between the scaled time series. For each region, a box-and-whisker plot is shown, highlighting the range between the 18th and 82nd percentiles (box, equivalent to the first standard deviation of a normal distribution) and between the 2.5th and 97.5th percentiles (whiskers, equivalent to the two standard deviation range).

Regions A–H show clear seasonal cycles with maxima in winter/early spring and minima in summer/early autumn. Regions I–P then show multiple minima and maxima, while Regions Q and R are largely in antiphase with A–H and peak in summer.

Spatial groupings with strong similarities in seasonal cycle and median amplitude can be seen in the data. These include a broad curve sweeping from the Rocky Mountains across the Atlantic to central Europe (A–E and G), a central Asian grouping (N–Q), and pairs in Greenland/Iceland (I and L) and the Middle East (F and H). These groups are distributed in zonal bands, with relatively short meridional distances leading to
much larger differences in seasonal cycle. For example, Regions I, J, L, and M lie spatially close to individual members of the A–E and G grouping but exhibit large differences from this group. These strong meridional gradients are consistent with the seasonal-level view seen in Figure 7 but are shown here to be applicable at the subseasonal level.

6. The Effect of the Upper Tropospheric Jet on Observed GWs

In section 4 we saw a clear pattern over the North Atlantic of large GW amplitudes to the north of the wind jets. In this section, we investigate this feature in more depth by examining the height dependence of GW amplitude and wind in the zonal mean.

Figures 9a–9d show zonal mean GW amplitudes, binned onto a grid of height 5 hPa and width 4° latitude. Mean tropopause pressure is indicated for context (thick black lines). This has been weighted spatially and temporally in the same way as the observations, although importantly that there is significant variability in this level. Figures 9e–9h show zonal mean zonal winds derived in the same way, while Figures 9j–9m show the zonal sum of individual measurements, peaking at around three million (10^6.5) data points per grid box. These figures have been adjusted for the uneven longitude distribution of the data by first gridding onto a 30° grid in longitude and then taking the zonal mean (or sum, for Figures 9j–9m). However, some bias is unavoidable and in particular the Pacific Ocean and the Southern Hemisphere away from routes between Europe and a small number of major airports are significantly underrepresented. Due to highly limited data coverage south of the Equator, we restrict the following discussion to the Northern Hemisphere; however, we note that based on the limited data available, it appears likely that most of our Northern Hemisphere conclusions are mirrored at least as far as ∼40°S.

We consider first zonal mean zonal winds, Figures 9e–9h. We see a clear jet spanning all measured heights in all four seasons, centered at ∼25°N in DJF and ∼40°N in JJA. The jet is strongest in DJF and weaker in

Figure 8. (top right) Map indicating sample regions selected as described in the text. Black circles surrounding each identifying letter show the area averaged. (a–r) Individual composite-year time series at each lettered location. Blue indicates above-median GW amplitudes, and red below. Ordinates show wave amplitudes in Kelvin; data have been averaged over a rolling 31-day window. Gray shading indicates the seasons plotted in Figure 7. “CMR Border” indicates the region at the junction of China, Mongolia, and Russia, between the Yablonovy (Russia) and Đà Xìng’ăn Lìng (China) mountain ranges. (bottom right) Box-and-whisker plots showing range of GW amplitudes measured in each region. Upper and lower horizontal lines show the 2.5th and 97.5th percentiles of the data, shaded boxes cover the region between the 18th and 82nd percentiles, and the black solid line indicates the median.
Figure 9. Seasonal plots of IAGOS-derived (a–d) zonal mean GW amplitude, (e–h) zonal mean zonal wind, and (j–m) zonal-sum number of measurements for (a, e, and j) SON, (b, f, and k) DJF, (c, g, and l) MAM, and (d, h, and m) JJA, as a function of (vertical axis) pressure level and (horizontal axis) latitude. Thick black line indicates mean tropospheric pressure level, derived from ERA5 data computed at IAGOS measurements locations.

JJA, consistent with the expected climatological state of the UTLS, and in general increases in speed with height across our data range.

We next examine zonal mean GW amplitudes, Figures 9a–9d. Larger GW amplitudes are clearly visible in the stratosphere, that is, above the thick black line, and these amplitudes generally increase with height (with minor exceptions). Maximal GW amplitudes at each latitude are seen slightly poleward of and above the jet in all four seasons. Amplitudes are generally lower in the troposphere and higher in the stratosphere. Outside the tropics they in general grow with altitude, with exceptions including some small patches at high latitudes in DJF and a disconnected low-altitude near-pole region and higher-altitude midlatitude peak in MAM. There is also evidence of lower-altitude maxima at around 40°N and 300 hPa in DJF and MAM. These lie within the jet but are disconnected from the larger regions of high amplitude in the stratosphere.

7. Climate Drivers of GW Variability

A unique benefit of these data is the combination of long duration and wide spatial coverage. Long-term records capable of resolving GW activity over such large spatial regions are rare, and we believe this to be the largest and longest such data set. As such, our data provide a unique opportunity to assess the role of long-term climate processes in driving or modulating GW activity.

In Figure 10 we compare GW amplitude time series from the 18 regions defined in Figure 8 to five climate indices—“Nino3.4,” representing the El Niño–Southern Oscillation (ENSO), “NAM,” representing
Figure 10. Analyses of possible gravity wave driving mechanisms. (i) and (ii) Pearson linear correlation coefficient between GW amplitudes for each region defined in Figure 8 and the El Nino 3.4 index (Nino3.4), Northern Annular Mode (NAM) index, 50-hPa QBO zonal winds (QBO-50), Total Solar Irradiance (TSI), and the Hadley Centre/Climate Research Unit surface temperature index (HadCRUT). (i) shows each region individually, and (ii) shows statistics for all regions combined. Box-and-whisker plots are as described in Figure 8; individual white symbols correspond to the points in panel (i). (iii) Results of a linear regression applied to each region based on the described indices, at a lag described by the y-axis of panel (iv) for each data set in months. Solid (semitransparent) symbols show significant (nonsignificant) results at the p = 0.01 level. (iv) As (iii) but for all regions combined. White (black) markers indicate significant (nonsignificant) results. (v) Temporal coverage for each region, at the monthly scale. Solid black indicates data coverage over the region that month. (vi) Region map, reproduced from Figure 8.

The Northern Annular Mode, "QBO-50," representing Quasi-Biennial Oscillation-driven equatorial wind speeds at 50 hPa, "TSI," representing the time-varying output of the Sun, and "HadCRUT," representing long-term changes in surface temperature (Coddington et al., 2015; Morice et al., 2012; Ogi et al., 2004; Trenberth & Stepaniak, 2001). Our HadCRUT4 time series is computed over the Northern Hemisphere only, since all regions lie here and relationships in the Southern Hemisphere may differ due to the very different relationship between GWs and the southern polar vortex.

We first bin our data onto a monthly time scale over the 26 years of data, averaging in all cases over the 500-km radius areas shown in Figure 8 and reproduced in Figure 10vi for context. Figure 10vi (top right) shows the temporal coverage available after this binning—most regions are well covered over the full duration of the data set, with the possible exception of the Rocky Mountains (Region A).

Figures 10i–10iv then compare our GW amplitude time series to the five climate indices, using Pearson linear correlation (Figures 10i and 10ii) and multiple linear regression (Figures 10iii and 10iv). All time series, both GW and index, have been boxcar-smoothed by 3 months, except for the rapidly varying NAM and GW series which we compare to it.

7.1. Correlations

We first discuss linear correlations, Figures 10i and 10ii. Figure 10i shows correlations between each index and GW amplitudes within individual regions. Colored markers, described by the key at top, show linear correlation between GW amplitudes and each climate index in the region. Figure 10ii compresses these correlations into a single box-and-whisker plot for each index, with individual regions overlaid with markers.

We see extreme scatter, with little consistency between the results and significant interregion variability for correlations with every metric. Except for the relatively poorly measured Rocky Mountains (A) and one other point (the NAM and the Sikhote Alin, Region O), we see no correlations exceeding an absolute value of 0.25. There is perhaps a very slight positive skew to the set of correlations between GW amplitudes and...
TSI, but this is weak and nearly a quarter (5/24) of these correlations are negative. Daily data were also examined, with broadly similar results; these daily data were also lagged over the range −70 to +70 days, but no lag value was found to systematically improve the correlation with any index over a majority of regions.

7.2. Linear Regression

Figures 10iii and 10iv show the results of a multiple linear regression applied to the same combinations of GW measurements and indices. As with the correlations, the upper panel shows results for individual regions and the lower panel combined results.

To compute comparable regression coefficients, we normalize each index. For all except HadCRUT, we convert to relative ranges, that is, a mean of zero and range of one over the period August 1994 to December 2019. For HadCRUT, which is noncyclical, this conversion is less meaningful, and we keep the original units of Kelvin. This produces similar numerical values to the other comparisons since the full range is of order 1 K. Thus, for all indices except HadCRUT, a coefficient of +1 suggests that a change from the minimum index value over this period to the maximum leads to a 1 K increase in monthly mean GW amplitude, while for HadCRUT it implies that a 1 K increase in global mean surface temperature leads to the same change.

We repeat the regression at time lags from −11 to +11 months then sum the absolute regression coefficient for each index and maximize to identify the lag, which provides the best estimated fit for that index across all regions. This best fit lag is the one displayed in our figure, and the lag period in months is indicated in the y axis labels on Figures 10iv, with a positive value indicating that the GW signal leads the index and vice versa.

We encode the p value of the t statistic for each fit using color density in the upper panel. Solid-colored markers are significant at the p < 0.01 level, while semitransparent markers (the vast majority) are not. In the lower panel this information is encoded with marker color, where black markers are significant at the 0.01 level. A threshold of 0.01 is chosen due to the nature of the comparison. A typical threshold of 0.05 would imply that the null hypothesis was false in 1/20 cases just by chance; since we here make 90 (5 × 18) comparisons we would therefore expect 5 of these to be marked as significant regardless of the actual truth. By using a much lower threshold, we reduce this random effect. This test assumes independence of the measured amplitude series, which is an inherently poor assumption for geophysical data and becomes increasingly less likely to be the case as locations become spatially closer; therefore, these values must be considered skeptically.

This choice to use a single lag for each index over all regions is made to simplify the presented results, and it may well be the case that some regions have significant and physically meaningful effects at different lags to others; this may therefore be a fruitful pathway for future work.

The most interesting results are for ENSO, shown as blue diamonds. In 9 of our 18 regions we see a statistically significant link, with an increase in ENSO associated with a drop in GW amplitudes 7 months earlier. This reduction is of 0.15–0.4 K/cycle; that is, a maximal El Niño leads to GW amplitudes 25–60% higher than a minimum La Niña, with a 7-month teleconnective lag. All regions that do not exhibit this lag relationship, excluding the poorly temporally sampled Rockies, are not identified as statistically significant and also fall within two distinct geographic groups: Greenland/Iceland and Russia (recall from Figure 8 that most Russian regions also have a very different seasonal cycle to the majority of our data). These data therefore represent evidence of strong teleconnective links between the Pacific Ocean and GW generation, propagation and/or filtering in the troposphere and UTLS over North America, Europe, south central Asia, and the non-Arctic Atlantic Ocean.

Other indices produce less clear results but still present some interesting conclusions, although not as strong as for ENSO. Aside from ENSO, we see 10 statistically significant linkages—this is higher than would be expected by chance assuming independence, but since the atmospheric system is inherently highly linked needs to be treated with skepticism.

For the NAM, only one region shows a significant relationship, Afghanistan (H), at a lag of 2 months. This region is spatially distant from the NAM foci in the North Atlantic, Arctic, and Pacific, and a physical mechanism that could link the NAM to this region and no surrounding ones is not immediately obvious. Therefore, we are disinclined to believe this is a meaningful or real teleconnection. The lack of relationship between
our data and the NAM in general may be due to the time scales studied—our data are gridded at the monthly level, while the NAM can vary significantly at shorter time scales.

The QBO shows two significant results, but these are in antiphase with each other and spatially disparate. All other results are scattered across regression coefficient space, mostly at relatively small values, and it is thus difficult to conclude that these are truly meaningful signals. This may be because the QBO becomes important at heights above \( \sim 50 \) hPa, while our data are at much lower altitudes. Similarly, three regions show a significant relationship with HadCRUT, but again with both positive and negative significant points and with a range of regional values, most clustered near 0.

TSI is slightly more interesting. At a lag of 3 months, there are four significant points, all positive in sign. Ten of the 14 nonsignificant points are also positive, as are most correlations presented in Figure 10ii. Finally, a 3-month lag is a physically reasonable time for a teleconnective signal such as this to operate, and since solar input is global a mechanistic pathway to a particular region is not needed, as with the NAM. Therefore, although we have insufficient data to draw any firm conclusions, our results may present weak evidence of a positive link between the solar cycle and NH UTLS GWs, with solar maximum corresponding to increased GW amplitudes.

8. Implications of Strong GW Activity Over the Rockies

In our data (e.g., Figure 7), we see large amplitudes over the Rocky Mountains (Region A). However, similar activity is not usually observed in stratospheric satellite data, including measurements at altitudes as low as 20 km altitude (\( \sim 50 \) hPa) made using HIRDLS/Aura (e.g., Ern et al., 2018; Geller et al., 2013; Wright et al., 2015). While poorly temporally sampled in the early 2010s (Figure 10), the total number of measurements over this region in earlier years is sufficient to properly characterize the seasonal cycle (Figures 5y–5B), and thus, discussion of this mismatch is a meaningful question.

The specific comparison to HIRDLS is important. Over the Rockies, IAGOS aircraft usually travel near-meridionally, as long-haul flights to Europe dominate the data set. As such, their along-track vector is very similar to Aura, which has a polar orbit at 98\(^\circ\) inclination. Alternate HIRDLS profile pairs are spaced \( \sim 40 \) km apart at 20-km altitude (e.g., Figure 4b of Wright et al., 2015), and thus, there should be some degree of observational filter overlap between the shortest waves visible to HIRDLS and longer waves seen by IAGOS, which have mean wavelengths in this region of \( \sim 150–250 \) km with a standard deviation (not shown) of \( \sim 100 \) km.

However, HIRDLS data show neither enhanced GW activity (e.g., Ern et al., 2018; Geller et al., 2013; Wright et al., 2015) over the Rockies, nor the intermittency characteristic of orographic wave sources in the stratosphere (Wright et al., 2013), even when filtered to identify only the shortest accessible length scales (Figure 7 of Wright et al., 2015). Assuming the waves seen here are orographic and therefore propagate mostly vertically, which seems likely given their geographic location and relatively large amplitudes, this suggests three possibilities: (a) that the waves break between \( \sim 200 \) and \( 50 \) hPa (\( \sim 10- \) to 20-km altitude), (b) that the waves propagate into the stratosphere but only have extremely short horizontal wavelengths and are thus invisible to HIRDLS, or (c) that the waves are invisible to HIRDLS due either to their phase fronts being aligned along the instrument boresight (e.g., Figure 1 of Wright & Hindley, 2018) or due to having vertical wavelengths \( < \sim 2–3 \) km.

Option (b) is unlikely as observations with AIRS/Aqua, which measures short-horizontal long-vertical wavelength waves, also show no significant stratospheric wave activity over the Rockies (Hoffmann et al., 2013). This restricts the possible wavelengths of any waves propagating into the stratosphere to small scales in both the vertical and horizontal. Also, our median horizontal wavelength is in the HIRDLS observational filter, so at least some signal would be expected to be visible unless there was a dramatic shift in horizontal wavelength between the two heights.

Option (c) is also unlikely as the HIRDLS boresight, which points at 47\(^\circ\) off the orbital track, will be aligned at a high angle to the zonally aligned waves the relative orientations of the winds and mountain ridge in this region would imply (Bacmeister, 1993), while short vertical wavelengths would preclude sharp vertical ascent.
It is therefore likely that the waves we observe here break in the UTLS, driving the winds specifically in this geographic and height region. This contrasts with many models (Geller et al., 2013), which propagate significant quantities of wave momentum flux into the stratosphere over the Rockies.

9. Discussion

The data used have significant, but known, biases and deficiencies. The biggest such deficiency is that the data are one-dimensional, lacking vertical and cross-track spatial information. This prevents us from estimating momentum fluxes and other important GW properties and complicates interpretation of our results. While we do not do so here, the combination of horizontal wavelength and wind speed can in principle be used to estimate vertical wavelength provided very significant assumptions are made (e.g., M. J. Alexander & Grimsdell, 2013), and this may be useful for future work. The lack of cross-track information is a problem common to stratospheric limb sounding studies of GWs, which is harder to adjust for, and therefore, we have in general avoided discussion of horizontal wavelengths, except in section 8 where the relative orientations of the measurements platforms being compared are known.

Biases include flight routing around storms, which may explain the lack of subtropical amplitude maxima in our summer data, a bias toward shorter wavelengths at the start of our routes due to spectral edge truncation, limited spatiotemporal coverage outside major flight paths, and a preference for flights to fly in the jet eastward and away from the jet westward leading to a likely positive skew over the North Atlantic in measured wave amplitudes. These biases are discussed in the text where appropriate. With the exception of the storm routing bias and to a lesser degree the jet-flight bias, these are likely to be systematic across the data set rather than seasonally varying and thus should not significantly affect our time series analyses for most regions.

Appendix C discusses our results in the context of previous UTLS GW studies, which in general agree well.

10. Summary and Conclusions

We have analyzed 26 years of UTLS GW measurements derived from commercial flight data, primarily over the Northern Hemisphere. This data set is uniquely long for a large area data set suitable for GW research, allowing us to study important effects inaccessible with other tools.

We show regionally varying seasonal cycles, with regional similarities primarily in the zonal direction. Our results include strong GW signals associated with orographic sources, which are not seen in stratospheric satellite data. The strongest and most consistent GW amplitudes are seen above and to the north of the upper tropospheric jet and over orographic hotspots such as southeast Greenland, East Asia, and the Rocky Mountains.

Analysis of the relationship between our long-timeseries data and multiple climate-system indices shows no significant relationship with the QBO, climate change, or the NAM. There may be weak evidence of a link with solar output 3 months earlier, although this is not statistically significant in most regions.

We do, however, see statistically significant evidence in all studied regions other than Russia and the Arctic of a 7-month lagged link with El Niño, with positive-phase ENSO leading to an increase in GW amplitudes in these regions of order 25%. A qualitative relationship between ENSO and UTLS GWs was previously suggested by Wang and Geller (2003) on the basis on 4 years of U.S. radiosonde data as a possible explanation for differences between their annual distributions, but here we advance significantly upon this earlier work by (a) demonstrating the link over a much larger area, (b) showing that it is statistically significant, and (c) and quantifying the change it induces in the wavefield.

Appendix A: Data Preprocessing

We first apply all error flags, removing data entirely if any variable we subsequently use is flagged as bad. After this, some records contain either negative time steps or repeated latitude/longitude pairs. For these records, we reorder the data to ascend monotonically in time and replace duplicated pairs via linear interpolation.

We next split the data into sections of straight near-level flight, removing any parts where height or direction changes quickly. These include ascents and descents, which we do not use. We label these straight stable
sections as “cruises,” which we define as having barometric altitude changes of <100 m and directional changes of <45° within any rolling 15-min time window. The first check ensures we measure a consistent physical regime, while the latter check is to avoid potentially measuring a laminar feature twice at different angles and misidentifying it as a wave. Each cruise is subsequently analyzed independently.

We then linearly interpolate the data to a 1-km along-track resolution. This is representative of the original data, which have a mean spacing of 1.09 km (median 1.01 km, standard deviation 0.50 km). We discard any cruises with discontinuities greater than 20 km, linearly interpolating over gaps smaller than this. Finally, we remove cruises which, after these steps, are <1,000 km long.

Appendix B: Cluster Analysis

To carry out our hierarchical cluster analysis, we first produce a list of latitude/longitude pairs associated with each data point. To make subsequent calculations computationally tractable, we assume that points within the same 0.2° × 0.2° box represent the same location and deduplicate at this level of precision (this does not affect our results, as the data values are restored once the grid is produced). This reduced data set still represents the data distribution at scales well above this and is therefore suitable for cluster generation, since we later impose a minimum area of 0.5° × 0.5°.

We next apply a $k$-means algorithm to the geolocation data, dividing the data into a fixed number of spatially compact clusters. For each season, we do this 10,000 times, selecting the iteration with the lowest total distance between measurements and cluster centers. We use a Euclidian distance metric in latitude/longitude space—this is strictly incorrect for the Earth but in practice introduces only a small error at the scale of our clusters and provides significant runtime benefits.

We initially define 4,000 clusters; this is arbitrary and is chosen as the highest round number to which we can apply our subsequent bootstrapping analysis in reasonable computer time (hours) and memory (tens of gigabytes) on the system used. Systematically varying this number suggests that using 4,000 clusters gives statistically similar results to maps produced using lower values but providing finer geographic resolution. Tests using regular latitude-longitude grids instead of the cluster analysis also produce similar results, but without geographically variable resolution.

We then reassign each measurement we removed in the deduplication step to the geographically closest cluster. Measurements >300 km from a cluster center are discarded. Finally, we require at least 500 measurements within each cluster, discarding those with fewer and reassigning their data to the next-nearest cluster, while retaining the 300-km limit. This reduces the number of clusters used to slightly below 4,000—for example, Figure 4 actually contains 3,764 clusters. The reassignment of discarded data to new cluster centers has a negligible effect on our results in practice but is included for completeness, as otherwise this step would discard useful data unnecessarily.

Appendix C: Validation

The height range, spatiotemporal coverage, and in situ nature of our data set are unusual and also have several inherent systematic biases. Therefore, it is important to validate our results against other long-term UTLS GW time series.

C1. U.S. Radiosondes

Wang and Geller (2003) (WG03) examined GW energy densities derived from high-resolution radiosondes over the United States for the period 1998–2001 in two height bands—a tropospheric (2–8.9 km) band below and a stratospheric (∼18–24.9 km) band above our analysis (∼10- to 12-km altitude). In their tropospheric band (their Figure 5), WG03 saw strong GW activity over the Rockies and western United States throughout the year, over the northwestern United States in DJF, and low activity otherwise. In the stratospheric band (their Figure 6), they saw only weak GW activity over the Rockies (except for a slight enhancement over Wyoming in DJF) but significant activity in the southern and eastern United States in DJF, in the south only in MAM, and over Oklahoma in SON.

Our maps (Figure 7) share many features with these results. In DJF, we see good agreement, with strong GW activity over the U.S. Rockies and as much as our data cover of the eastern seaboard and southeastern United
States, consistent with a mixture of the two layers of WG03. In MAM, we see GW activity in the western United States consistent with their tropospheric band but do not see major activity over the southern United States (although our data are truncated near this region, and data density is often poor in the areas our data do cover). In JJA/SON our results are less consistent with WG03, with good agreement in most places but lacking their strong signal over the Rockies.

C2. North Atlantic Radiosondes
Plougonven et al. (2003) describe GW measurements from 224 radiosondes launched from ships in the North Atlantic (30–70°N, 50–0°W) during the FASTEX campaign in January/February 1997. They focused on short vertical wavelength features (<5 km), which they studied in the context of the jet up to 20 km. They saw GW activity centered on the jet axis in both the stratosphere and troposphere, peaking on the north side in the stratosphere, with a nominal maximum 300 km from jet center but with extremely large scatter (their Figures 2 and 5). Our peak GW activity is consistently to the north side of the jet but slightly further away, with a maximum between 5° and 10° north of the jet core. However, given the very high degree of scatter in the data of Plougonven et al. (2003), these conclusions are not inconsistent.

C3. Greenland Aircraft Turbulence Encounters
Lane et al. (2009) describe statistics of aircraft turbulence encounters over Greenland, derived from aircraft instrument reports. These events primarily arise due to interactions between GWs emitted from Greenland’s orography and background directional wind shear. They see a near-annual cycle in these reports, with peaks in November and January and a trough in May. The winter peaks and May trough are consistent with our data (Figure 8i), but we also see a secondary maximum in August/September, which they do not.

C4. Aberystwyth MST Radar
Vaughan and Worthington (2007) studied inertia-GWs using 8 years of data from the MST Radar in Aberystwyth, Wales (within our Region D). They observed a winter peak in occurrence rate. While these results carry the important caveat that they explicitly aimed to exclude the orographic waves that likely form a significant proportion of our data in this region, their results were again consistent with ours.

C5. European Satellite Cloud Imagery
Cruette (1976) studied lee waves patterns in satellite cloud imagery from the midtroposphere in western Europe over 3 years of data. She saw a strong seasonal pattern over an area including the United Kingdom (Region D) and Western Alps (edge of our Region G), with a peak in winter and trough in summer, again consistent with our results.

C6. CHAMP Estimates of Russian Lower Stratospheric GW Seasonality
Independent estimates of UTLS GW seasonality over Russia and central Asia (Regions N-R) have proven challenging to locate in the literature. This is a major issue, because the observed GW seasonality in our data in these regions is opposite to most other regions and to a more general assumption that GW amplitudes will be on average higher in winter due to the nature of most source mechanisms.

Some very limited evidence to support the observed seasonality is seen in Namboothiri et al. (2008), where CHAMP-derived GW variances at 20-km altitude for 2002 exhibit larger values in JJA than DJF by a factor of 4, but from a low base. However, this evidence is from an experimental instrument type over only a short duration, and longer-term measurements at 30-km altitude from the more modern HIRDLS and SABER instruments (Ern et al., 2018) do not show this, peaking instead in January.

We note clearly that the observed seasonality could be correct. A plausible mechanism, for example, would be nonorographic generation from summer storms, which, deep in the continent, will not be modulated strongly by the North Atlantic jet stream. The annual cycle of rainfall in this region is consistent with this mechanism, and filtering, horizontal propagation, or observational filter differences could easily explain the lack of observed signal at 30 km. However, given the lack of validation, it is important to remain skeptical of these specific results until further evidence is available.
IAGOS data are available from the IAGOS Data Portal. The software produced to generate the data and figures is available from Zenodo (under https://doi.org/10.5281/zenodo.4018148).

Acknowledgments
We would like to thank Manoj Joshi (UEA) for useful comments at the 2020 UK Climate Dynamics Meeting on possible mid-European wave sources and interregion comparisons, Paul Williams (Reading) for useful comments via email on aircraft routing around storms, and Neil Hindley (Bath) for providing the improved S-Transform code and for several useful comments relating to our data analysis pipeline. C. J. Wright is funded by Royal Society University Fellowship UF160545 and by NERC Grants NE/R001391/1 and NE/S00985X/1. T. P. Banyard is funded by an EPSRC Doctoral Training Account awarded to the University of Bath and by Royal Society Grant RG/E/EA/180217.

References

Alexander, M. J., Geller, M., McLandress, C., Polavarapu, S., Preusse, P., Sassi, F., et al. (2010). Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models. *Quarterly Journal of the Royal Meteorological Society*, 136, 1103–1124. https://doi.org/10.1002/qj.637

Alexander, M. J., Gille, J., Cavanaugh, C., Coffey, M., Craig, C., Eden, T., et al. (2008). Global estimates of gravity wave momentum flux from High Resolution Dynamics Limb Sounder observations. *Journal of Geophysical Research*, 113, D15S18. https://doi.org/10.1029/2007JD008807

Alexander, M. J., & Grimsdell, A. W. (2013). Seasonal cycle of orographic gravity wave occurrence above small islands in the Southern Hemisphere: Implications for effects on the general circulation. *Journal of Geophysical Research: Atmospheres*, 118, 11,589–11,599. https://doi.org/10.1002/2013JD020256

Alexander, S. P., Ekeloov, A. R., McDonald, A. J., & Pitts, M. C. (2013). Quantifying the role of orographic gravity waves on polar stratospheric cloud occurrence in the Antarctic. *Journal of Geophysical Research: Atmospheres*, 118, 11,493–11,507. https://doi.org/10.1002/2013JD020122

Bacmeister, J. T. (1993). Mountain-wave drag in the stratosphere and mesosphere inferred from observed winds and a simple mountain-wave parameterization scheme. *Journal of the Atmospheric Sciences*, 50(3), 377–399. https://doi.org/10.1175/1520-0469(1993)050<0377:MWDITS>2.0.CO;2

Berkes, F., Neis, P., Schultz, M. G., Bundke, U., Rohs, S., Smit, H. G. J., et al. (2017). In situ temperature measurements in the upper troposphere and lowermost stratosphere from 2 decades of IAGOS long-term routine observation. *Atmospheric Chemistry and Physics*, 17(20), 12,495–12,508. https://doi.org/10.5194/acp-17-12495-2017

Boulanger, D., Thouret, V., & Petzold, A. (2019). *IAGOS Data Portal*. Aercis. https://doi.org/10.25326/20

Callies, J., Bühler, O., & Ferrari, R. (2016). The dynamics of mesoscale winds in the upper troposphere and lower stratosphere. *Journal of the Atmospheric Sciences*, 73(12), 4853–4872. https://doi.org/10.1175/jas-d-16-0108.1

Callies, J., Ferrari, R., & Bühler, O. (2014). Transition from geostrophic turbulence to inertia-gravity waves in the atmospheric energy spectrum. *Proceedings of the National Academy of Sciences*, 11(48), 17,033–17,038. https://doi.org/10.1073/pnas.1407721112

Cho, J. Y. N., & Lindborg, E. (2001). Horizontal velocity structure functions in the upper troposphere and lower stratosphere. I. Observations. *Journal of Geophysical Research*, 106(D10), 10,223–10,232. https://doi.org/10.1029/2000JD900814

Coddington, O., Lean, J. L., Lindholm, D., Pilewskie, P., Snow, M., & NOAA CDR Program (2015). NOAA Climate Data Record (CDR) of Total Solar Irradiance (TSI). NRLTSI Version 2. NOAA National Centers for Environmental Information. Retrieved from https://data.nodc.noaa.gov/cgi-bin/iso?id=NOAA:CDDA000828

Crueu, D. (1976). Experimental study of mountain lee waves by means of satellite photographs and aircraft measurements. *Tellus*, 28(6), 498–523. https://doi.org/10.1340/tellusa.v28i6.11318

Ern, M., Preusse, P., Alexander, M. J., & Warner, C. D. (2004). Absolute values of gravity wave momentum flux derived from satellite data. *Journal of Geophysical Research*, 109, D20103. https://doi.org/10.1029/2004JD004752

Ern, M., Trinh, Q. T., Preusse, P., Gille, J. C., Mlynczak, M. G., Russell, J. M. III, & Riese, M. (2018). GRACLE: A comprehensive climatology of atmospheric gravity wave parameters based on satellite limb soundings. *Earth System Science Data*, 10(2), 857–882. https://doi.org/10.5194/essd-10-857-2018

Pitts, D. C., & Alexander, M. J. (2003). Gravity wave dynamics and effects in the middle atmosphere. *Reviews of Geophysics*, 41(1), 1003. https://doi.org/10.1029/2001RG000106

Geller, M. A., Alexander, M. J., Love, P. T., Bacmeister, J., Ern, M., Hertzog, A., et al. (2013). A comparison between gravity wave momentum fluxes in observations and climate models. *Journal of Climate*, 26(17), 6383–6405. https://doi.org/10.1175/jcli-d-12-0054.1

Herschbach, 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. https://doi.org/10.1002/qj.3803

Hindley, N. P., Wright, C. J., Smith, N. D., Hoffmann, L., Holt, L. A., Alexander, M. J., et al. (2019). Gravity waves in the winter stratosphere over the Southern Ocean: High-resolution satellite observations and 3-D spectral analysis. *Atmospheric Chemistry and Physics*, 19(24), 15,377–15,414. https://doi.org/10.5194/acp-19-15377-2019

Hindley, N. P., Wright, C. J., Smith, N. D., & Mitchell, N. J. (2015). The southern stratospheric gravity wave hot spot: Individual waves and their momentum fluxes measured by COSMIC GPS-RO. *Atmospheric Chemistry and Physics*, 15(14), 7797–7818. https://doi.org/10.5194/acp-15-7797-2015

Hoffmann, L., Xue, X., & Alexander, M. J. (2013). A global view of stratospheric gravity wave hotspots located with Atmospheric Infrared Sounder observations. *Journal of Geophysical Research: Atmospheres*, 118, 416–434. https://doi.org/10.1029/2012JD018658

Kuchar, A., Sacha, P., Eichinger, R., Jacobi, C., Pisot, P., & Rieder, H. E. (2020). On the intermittency of orographic gravity wave hotspots and their importance for middle atmosphere dynamics. *Weather and Climate Dynamics Discussions*. https://doi.org/10.5194/wcd-2020-21

Lane, T. P., Doyle, J. D., Sharman, R. D., Shapiro, M. A., & Watson, C. D. (2009). Statistics and dynamics of aircraft encounters of turbulence over Greenland. *Monthly Weather Review*, 137(8), 2687–2702. https://doi.org/10.1175/2009mwr2878.1

Li, Q., & Lindborg, E. (2018). Weakly or strongly nonlinear mesoscale dynamics close to the tropopause? *Journal of the Atmospheric Sciences*, 75(4), 1215–1229. https://doi.org/10.1175/jas-d-17-0063.1

Lindborg, E. (1999). Can the atmospheric kinetic energy spectrum be explained by two-dimensional turbulence? *Journal of Fluid Mechanics*, 388, 259–288. https://doi.org/10.1017/s0022112099004851

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research*, 117, D08101. https://doi.org/10.1029/2011JD017187

Moss, A. C., Wright, C. J., & Mitchell, N. J. (2016). Does the Madden-Julian oscillation modulate stratospheric gravity waves? *Geophysical Research Letters*, 43, 3973–3981. https://doi.org/10.1002/2016GL068498
Murgatroyd, R. J., & Singleton, F. (1961). Possible meridional circulations in the stratosphere and mesosphere. Quarterly Journal of the Royal Meteorological Society, 87(372), 125–135. https://doi.org/10.1002/qj.4970873720

Namboothiri, S. P., Jiang, J. H., Kishore, P., Igarashi, K., Ao, C. O., & Romans, L. J. (2008). CHAMP observations of global gravity wave fields in the troposphere and stratosphere. Journal of Geophysical Research, 113, D07102. https://doi.org/10.1029/2007JD008912

Ogi, M., Yamazaki, K., & Tachibana, Y. (2004). The summertime annular mode in the Northern Hemisphere and its linkage to the winter mode. Journal of Geophysical Research, 109, D20114. https://doi.org/10.1029/2004JD004314

Plougonven, R., Hertzog, A., & Guez, L. (2012). Gravity waves over Antarctica and the Southern Ocean: Consistent momentum fluxes in mesoscale simulations and stratospheric balloon observations. Quarterly Journal of the Royal Meteorological Society, 139(670), 101–118. https://doi.org/10.1002/qj.1965

Plougonven, R., Teitelbaum, H., & Zeitlin, V. (2003). Inertia gravity wave generation by the tropospheric midlatitude jet as given by the Fronts and Atlantic Storm-Track Experiment radio soundings. Journal of Geophysical Research, 108(D21), 4686. https://doi.org/10.1029/2003JD003535

Scott, R. K., Cammas, J.-P., Mascart, P., & Stolle, C. (2001). Stratospheric filamentation into the upper tropical troposphere. Journal of Geophysical Research, 106(D11), 11,835–11,848. https://doi.org/10.1029/2001JD900049

Trenberth, K. E., & Stepaniak, D. P. (2001). Indices of El Niño evolution. Journal of Climate, 14(8), 1697–1701. https://doi.org/10.1175/1520-0442(2001)014<1697:LIOENO>2.0.CO;2

Vaughan, G., & Worthington, R. M. (2007). Inertia-gravity waves observed by the UK MST radar. Quarterly Journal of the Royal Meteorological Society, 133(S2), 179–188. https://doi.org/10.1002/qj.142

Wright, C. J., & Hindley, N. P. (2018). How well do stratospheric reanalyses reproduce high-resolution satellite temperature measurements? Atmospheric Chemistry and Physics, 18(18), 13,703–13,731. https://doi.org/10.5194/acp-18-13703-2018

Wright, C. J., Osprey, S. M., Barnett, J. J., Gray, L. J., & Gille, J. C. (2010). High Resolution Dynamics Limb Sounder measurements of gravity wave activity in the 2006 arctic stratosphere. Journal of Geophysical Research, 115, D02105. https://doi.org/10.1029/2009JD011858

Wright, C. J., Osprey, S. M., & Gille, J. C. (2015). Global distributions of overlapping gravity waves in HIRDLS data. Atmospheric Chemistry and Physics, 15(14), 8459–8477. https://doi.org/10.5194/acp-15-8459-2015

Wright, C. J., Scott, R. B., Arbic, B. K., & Furnival, D. F. (2012). Bottom dissipation of subinertial currents at the Atlantic zonal boundaries. Journal of Geophysical Research, 117, C03049. https://doi.org/10.1029/2012JC007702

Wright, C. J., Scott, R. B., Furnival, D., Ailliot, P., & Vermet, F. (2013). Global observations of ocean-bottom subinertial current dissipation. Journal of Physical Oceanography, 43(2), 402–417. https://doi.org/10.1175/jpo-d-12-082.1