ABSTRACT: The way the large-scale flow determines the energy of the nonorographic mesoscale inertia–gravity waves (IGWs) is theoretically significant and practically useful for source parameterization of IGWs. The relations previously developed on the $f$ plane for tropospheric sources of IGWs including jets, fronts, and convection in terms of associated secondary circulations strength are generalized for application over the globe. A low-pass spatial filter with a cutoff zonal wavenumber of 22 is applied to separate the large-scale flow from the IGWs using the ERA5 data of ECMWF for the period 2016–19. A comparison with GRACILE data based on satellite observations of the middle stratosphere shows reasonable representation of IGWs in the ERA5 data despite underestimates by a factor of smaller than 3. The sum of the energies, which are mass-weighted integrals in the troposphere from the surface to 100 hPa, as given by the generalized relations is termed initial parameterized energy. The corresponding energy integral for the IGWs is termed the diagnosed energy. The connection between the parameterized and diagnosed IGW energies is explored with regression analysis for each season and six oceanic domains distributed over the globe covering the Northern and Southern Hemispheres and the tropics. While capturing the seasonal cycle, the domain area-average seasonal mean initial parameterized energy is weaker than the diagnosed energy by a factor of 3. The best performance in regression analysis is obtained by using a combination of power and exponential functions, which suggests evidence of exponential weakness.

KEYWORD: Gravity waves

1. Introduction

Inertia–gravity waves (IGWs) become dominant modes of motion at mesoscales of the atmosphere, i.e., at horizontal scales smaller than about 500 km (Callies et al. 2014; Zagar et al. 2017), thereby contributing to the loss of predictability in weather prediction (Judt 2018) and model uncertainty in climate prediction (Liu 2019). For current general circulation models (GCMs), the resolvable scales of atmospheric phenomena are on the order of 100 km. Resolving smaller-scale processes is inhibitive due to the computational costs of explicit representation of detailed physical processes at such scales. Therefore, subgrid-scale phenomena, such as IGWs must be parameterized in GCMs.

The first step in parameterization of IGWs is based on observational studies (Geller et al. 2013; Jewtoukoff et al. 2015; Ern et al. 2018) to identify the mechanism of their generation, propagation, and dissipation. Previous studies have shown that the most important energy sources of IGWs are orography, convection, and coupled jet and front systems (e.g., Plougonven and Zhang 2014), which are mainly located in the troposphere. Based on their sources, the IGWs are thus classified into “orographic” and “nonorographic” (e.g., Kim et al. 2003) and dealt with separately in IGW parameterization schemes as they have distinct characteristics and impacts (McLandress et al. 2013). Through their sources, the nonorographic IGWs are connected to the large-scale flow which is dominantly balanced. There has been a growing interest in exploiting the connection to improve the representation of amplitude in “source parameterization” of IGWs (Charron and Manzini 2002; Züllicke and Peters 2008; Richter et al. 2010; Mirzaei et al. 2014; de la Cámara and Lott 2015; Chun et al. 2019). The objective is to include the spatiotemporal variability of IGWs at the source level (de la Cámara and Lott 2015), which is expected to increase the realism of their representation and impact in the middle atmosphere. This becomes ever more important as the top of the atmospheric models is extended in the middle atmosphere, which is greatly affected by forcing and variability of the Rossby and gravity waves. The development of instruments (e.g., lidar, radar, and satellite imagery) that monitor the upper atmosphere layers improves our knowledge of wave interactions and can be helpful in upgrading the nonorographic wave drag schemes. Observations
confirm that a significant part of the gravity waves that exist in the middle atmosphere have nonorographic origins (Holt et al. 2017).

Convection frequently occurs in the tropical regions. The IGWs generated by convection can reach the middle atmosphere and impact the semianual and quasi-biennial oscillations (Dunkerton 1997; Baldwin et al. 2001; Müller et al. 2018). Accordingly, the schemes introduced for the convectively generated waves have seen significant development during the last decade. Moreover, in contrast to the Northern Hemisphere for which the biases were mainly reduced by implementation of mountain drag schemes, the parameterizations for the convectively generated IGWs proved specifically important in the Southern Hemisphere (Chan et al. 2001) which is mainly covered by oceans. In this regard, Bossuet et al. (1998) implemented a simple scheme which relates the gravity wave momentum fluxes to the precipitation flux as an index of convective activity in the model. Following Lindzen (1981), they applied the linear theories of vertical propagation and considered the convective gravity wave drag in the same way as the mountain waves. This implementation resulted in significant reduction of equatorial easterly bias and slowing down of the stratospheric winds. It also had a positive impact on the monsoon. Kershaw (1995) introduced another parameterization for the convectively generated gravity waves by relating the wave momentum flux to the intensity of convection and the stability of the air above the convective layer.

Contrary to the convective gravity waves, the parameterization of IGWs due to “jets” and “fronts” is still under consideration in GCMs. Plougonven and Zhang (2014) provided a detailed discussion of the importance, generation mechanisms, propagation, impact and parameterization of IGWs in the vicinity of jet–front systems. Charron and Manzini (2002) and Richter et al. (2010) have adopted the frontogenesis function in the midtroposphere to prognostic and identify the regions of wave generation. Through a combination of this approach, the implementation of other parameterization procedures for the convective and orographic gravity waves as well as increasing the spatial and temporal resolution of the numerical model, Richter et al. (2010) achieved a reduction in the cold pole bias and a more realistic result for the frequency of stratospheric sudden warming events. It should be noted that in Richter et al. (2010) the use of the frontogenesis function for parameterization of amplitude was limited in the sense that a fixed amplitude for the frontal gravity waves at the source level was employed.

Based on earlier theoretical works on gravity wave generation by sheared potential vorticity anomalies in idealized settings (Lott et al. 2010, 2012), a parameterization of amplitude was implemented by de la Cámara and Lott (2015) in Laboratoire de Météorologie Dynamique zoom (LMDz) GCM. This can be regarded as the first full implementation and test of a dynamically based source parameterization in GCMs. For the implementation of the analytically obtained emission formula of Lott et al. (2012), however, no procedure for verification is presented in de la Cámara and Lott (2015), except a qualitative comparison with the horizontal temperature gradient at 600 hPa level. Given the complexity and lack of a complete theory of IGW emission in real cases (Plougonven and Zhang 2014), a statistical analysis is essential for the verification and possibly rectification of any theoretical formula by the actual diagnosis of IGWs in the atmosphere. Focusing on a single generation mechanism due to linear vertical shear, no distinction is made in de la Cámara and Lott (2015) between the likely different contributions by jets and fronts in IGW emission. Recently, Chun et al. (2019) used the frontogenesis function and the residual of nonlinear balance equation (RNBE) to take account of gravity wave generation by jet–front systems. They follow the impact of these lower atmosphere sources on the middle-atmosphere momentum flux and find weak but systematic correlations in different regions. As shown by Yasuda et al. (2015), the RNBE formula is the same as the source term describing the IGWs emitted from the balanced flow. In analysis by Chun et al. (2019), the convective IGW generation is excluded and no discussion is given on the way the frontogenesis function and RNBE can be used for parameterization of amplitude. It is worth mentioning that in the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecast System (IFS), for nonorographic gravity waves, the amplitude at source level is set to a globally uniform value and adjusted by a prescribed relation for latitude and resolution (see section 5.3 in ECMWF 2018). This is a gross simplification of spatial variability and the complete neglect of temporal variability in the wave emission process.

Following earlier work by Zülicke and Peters (2008), a new approach to parameterization of amplitude or energy for IGWs was introduced by Mirzaei et al. (2014) which is referred to as MZMAP for brevity. The underlying principle in MZMAP is to relate the energy of IGWs to a properly defined measure of fastness for processes associated with jets, fronts, and convection. The ageostrophic Rossby number, relevant to cases where the dominant balance is geostrophy, is taken as the measure of fastness. For each source, with direct relation to the ageostrophic Rossby number, the strength of the secondary circulations is used to construct the parametric relations. In this way, the energy of generated IGWs is made proportional to the square of ageostrophic Rossby number. MZMAP developed and implemented the energy parameterization for the life cycle of midlatitude baroclinic waves in idealized numerical simulations with the Weather Research and Forecasting mesoscale model (WRF). In this regard, they considered a certain amount of energy for each source of nonorographic IGWs based on the diagnostics such as the Lagrangian wind speed deceleration in the exit region of the jets, the frontogenesis function, and the latent heat released in convective regions. The generation from the source has to be followed by the propagation in the stratosphere to eventual dissipation in the mesosphere. For the present study, we focus on IGW generation in the troposphere/lower stratosphere. This covers the most relevant sources of nonorographic IGWs but neglects vertical propagation.

The current study aims to generalize, apply, and assess the energy parameterization relations of MZMAP over the globe. The paper is organized as follows. Section 2 describes the data and methods with emphasis on the energetics relations and the modifications required in the global domain with respect to the midlatitude f-plane domain previously examined. The results
and statistical model evaluation are presented in section 3. This includes a comparison between the IGWs identified in the data with those given by satellite observations at middle stratosphere, and a review of wave sources as jets, fronts, and convection which enter the relations. It is followed by a comparison of parameterized and diagnosed energies, separated by season, region, and source. In particular, the objective here is to adjust the IGW sources to improve the representation of spatiotemporal variability by means of identifying suitable regression functions with advanced statistical measures. Summary and conclusions are given in section 4.

2. Formulation, data, and methodology

a. Main relations

A set of parametric relations for nonorographic gravity wave energy has been proposed by MZMAP in which the IGW energy is related to the ageostrophic motions. To characterize the wave sources, the cross-stream ageostrophic wind speed $u^a_\text{conv}$, the frontogenesis function $F_{\text{front}}$ (Miller 1948; Hoskins 1982), and the latent heat released during condensation $Q_{\text{conv}}$ (Emanuel et al. 1987) are used and the bulk relations for $e^\text{jet}$, $e^\text{front}$, and $e^\text{conv}$ as the specific source energies associated with, respectively, the jet, front, and convection are written as

$$e^\text{jet} = (u^\text{jet}_a)^2,$$

$$e^\text{front} = \left(\frac{g}{\theta_0} F_{\text{front}} L_{\text{front}}^z\right)^2,$$

$$e^\text{conv} = \left(\frac{1}{\theta_0 \partial \theta/\partial z} Q_{\text{conv}} L_{\text{conv}}^z\right)^2.$$

Here, $f$ is the Coriolis parameter, $g$ is the acceleration due to gravity, $\theta_0$ is a reference potential temperature, $L_{\text{front}}^z$ is the vertical scale of the front, and $L_{\text{conv}}^z$ and $L_{\text{jet}}^z$ are the vertical and horizontal scales of the convection, respectively. Guided by the numerical simulations in MZMAP and quasigeostrophic scaling, empirically valid asymptotics are set for these parameters: $\theta_0 = 300 \text{K}, \partial \theta/\partial z = 3.1 \text{K km}^{-1}, L_{\text{front}}^z = 2 \text{km}, L_{\text{conv}}^z = 100$, and $L_{\text{jet}}^z = 2 \text{km}$.

The concept of Lagrangian Rossby number $R_{\text{Ro}}$ is exploited in order to detect the ageostrophic motions that are sufficiently fast to generate unbalanced flow (see McIntyre 2009 for a thorough discussion). In the region of unbalanced flow, we have $R_{\text{Ro}} = u_a/(fL_\text{d}) = \omega_a/(2\pi f)$ with $\omega_a$ the IGW frequency attributed to the ageostrophic motions. As in MZMAP, $\omega_a/f$ is assumed greater than unity for the criterion of IGW generation. This criterion leads us to $R_{\text{Ro}} > 1/(2\pi) \approx 0.15$ with $L_\text{d}$ the scale of the fastest wind speed reduction in the jet exit region and thus

$$u_a > fL_\text{d}/2\pi.$$

Similar to $u^\text{jet}_a$, one can define the ageostrophic horizontal wind associated with front and convection by $u^\text{front}_a$ and $u^\text{conv}_a$, respectively. Using $e^\text{front} = (u^\text{front}_a)^2$ and $e^\text{conv} = (u^\text{conv}_a)^2$, and the criterion (2.4) for each of $u^\text{jet}_a$, $u^\text{front}_a$, and $u^\text{conv}_a$ in (2.1), (2.2), and (2.3), leads to the following thresholds:

$$u^\text{jet}_a = fL_\text{h}/2\pi,$$

$$F_{\text{front}} = \frac{1}{2\pi} \frac{\theta_0 f^2 N}{g},$$

$$Q_{\text{conv}} = \frac{1}{2\pi} \frac{\theta_0 f^2 N^2 f}{g} L_{\text{conv}}^z.$$

below which the corresponding source is assumed to be inactive. In (2.5) to (2.7), $u^\text{jet}_a$, $F_{\text{front}}$, and $Q_{\text{conv}}$ are the thresholds for, respectively, the ageostrophic wind speed associated with the jet, frontogenesis function and latent heating, $N = [(g/\theta_0) \partial \theta/\partial z]^{1/2}$ is the buoyancy frequency, and the quasigeostrophic scaling $L_{\text{front}}^z/L_{\text{h}}^z = f/N$ has been used. For diagnostics, $L_{\text{conv}}^z$ and $L_h$ are taken equal to 4 and 250 km, respectively, following the numerical simulations in MZMAP. The thresholds are calculated at each grid point for the ageostrophic wind speed and the latent heating. The threshold for the frontogenesis function is empirically set to $F_{\text{front}} = 0.15 \text{K (100 km)}^{-1} \text{h}^{-1}$. The total IGW energy attributed to the nonorographic sources is then considered as the sum of the energies from the three sources, that is

$$E_{\text{total}} = e^\text{jet} + e^\text{front} + e^\text{conv},$$

which constitutes our initial estimate for parameterized energy.

b. In the tropical regions

MZMAP applied their parameterization relations in a channel on the $f$ plane. On the global domain, the $f$-plane relations lose validity in the tropical region, here taken to be between 20$^\circ$S and 20$^\circ$N, as the Coriolis parameter approaches zero along the equator. Hence, to be applicable in the tropical region, the energetic relations should be modified. Given the lack of noticeable temperature gradient in the tropical region, the contribution of the frontal source for IGWs can be neglected. This has been confirmed by Charron and Manzini (2002) and Chun et al. (2019). The focus is then on the IGW sources due to jet and convection which will be discussed next.

1) PARAMETRIC RELATIONS FOR JET

MZMAP used the linear theory for their parameterization and split the actual flow into the geostrophic wind as the mean flow and the ageostrophic part as the perturbation which is related to IGWs. This decomposition is not strictly applicable in the tropical regions due to the breakdown of geostrophy and thus small Rossby number asymptotics. To extend the approach of MZMAP to the equatorial regions, the uniformly valid asymptotics developed by McWilliams (1985) spanning both the geostrophic and gradient wind scalings is exploited. In this asymptotics, for the parameter $\varepsilon$ that measures the ratio of horizontal divergence $\delta$ to relative vertical vorticity $\xi$, the scaling $\varepsilon = (L_d/L_\text{d})^{\text{Ro}} \max(1, \text{Ro})$ is obtained, where Ro is the Rossby number and $L_d = c/f$ is the Rossby deformation length, with $c$ a gravity wave phase speed. For the case with dominance of geostrophic balance examined by MZMAP, $\varepsilon$ reduces to Ro. In the equatorial region where gradient wind scaling replaces geostrophic scaling, $\varepsilon$ reduces to the square of the Froude
number, \(Fr = V/c\), with \(V\) the scale for horizontal velocity. Using modified expansion (Warn et al. 1995; Vallis 1996) for \(\zeta\), the master variable in the Froude number asymptotics, \(\delta\) first appears in order \(\epsilon\). In this way, the divergent wind plays in Froude number asymptotics the same role of ageostrophic wind which first appears in order \(Ro\) in Rossby number asymptotics. On this basis, the actual wind is decomposed into wind which first appears in order \(Ro\) in Rossby number as- by filtering data cannot cover the small-scale parameterized con- triction is thought to be responsible for a significant part of... scale convection. Therefore, the resulting divergent wind may also... generation of large shearing motion in the vertical di- direction is thought to be responsible for a significant part of small horizontal scale IGW activity (Mohabalhojeh and Dritschel 2004). For a better estimate of \(e^{\text{jst}}\) by giving more predominance to vertical shear, therefore the vertically averaged divergent wind is subtracted from \(v_d\) before using (2.9). For consistency the same procedure is applied to ageostrophic wind outside the tropical region. The resulting relations for jet thus become

\[ e^{\text{jst}} = |v_d - \mathbf{v}_d|^2, \]  
(2.10)

in the extratropical region and

\[ e^{\text{jst}} = |v_j - \mathbf{v}_j|^2, \]  
(2.11)

in the tropical region, where \(\mathbf{v}_d\) and \(\mathbf{v}_j\) denote the vertically averaged ageostrophic and divergent wind, respectively.

The pattern of divergent circulation in the tropical region exhibits strong divergent velocity, especially in June–August (JJA) over Indian Ocean and Africa (Trenberth et al. 2000). A large component of circulation in the tropics is rather than meridional and referred to as Walker circulation (Bjerknes 1969; Krishnamurti 1971). Therefore, to avoid overestimating the source energy due to the jet, the contribution of the sectorial zonal mean circulation in the divergent velocity over the tropical regions is removed in parameterization relations. The sectorial zonal mean refers to the zonal mean taken over each of the tropical regions introduced in section 2c.

Restricting motion to sufficiently short zonal scales for which one can filter the long Kelvin waves, it is possible to employ a time-scale separation between the fast and slow modes analog to that used for unbalanced flow based on the concept of Lagrangian Rossby number. It is not difficult to show that \(\beta L_z\) provides a useful frequency separation between the fast and slow modes in the equatorial region, using which and the criterion \(2\pi u^{\text{jst}}/\beta L_z > \beta L_z\) from (2.4) for the initiation of IGWs, the following threshold for divergent wind is obtained

\[ |v_{\text{jst}}| \geq \frac{1}{2\pi} (\beta NL_z)^{1/2} L_z^{1/2}, \]  
(2.12)

where \(\beta = df/dy\) is the Rossby or beta parameter and use has been made of the equatorial wave theory (Andrews et al. 1987) for the meridional length scale \(L_y = (NL_z/\beta)^{1/2}\).

2) PARAMETRIC RELATIONS FOR CONVECTION

To make the formula (2.3) applicable for the tropical region, we rewrite Eq. (B14) in MZMAP for the circulation induced by convection as

\[ \frac{g^2}{\beta} \frac{\partial^2 \psi_{\text{conv}}}{\partial z^2} + N^2 \frac{\partial^2 \psi_{\text{conv}}}{\partial x^2} = \frac{\partial Q}{\partial x}, \]  
(2.13)

in which \(x, y,\) and \(z\) are local Cartesian coordinates, \(\psi\) is the streamfunction, and \(Q = (g/\theta_b)Q_{\text{conv}}\) is the buoyancy forcing. Considering the length scales \(L_x, L_y,\) and \(L_z\) for, respectively, the zonal, meridional, and vertical directions, a straightforward scale analysis of Eq. (2.13) requires that \(L_z = (\beta/N)L_x L_y\) and

\[ u_{\text{conv}} = \frac{Q_{\text{conv}}}{\partial \theta_b/\partial z} L_x^{1/2} L_z^{-1/2}, \]  
(2.14)

for the scale of the zonal wind associated with the convectively induced circulation. The threshold for latent heating in tropical regions can be defined with the help of frequency separation between the fast and slow modes as carried out for Eq. (2.10), that is,

\[ Q_{\text{th}} = \frac{1}{2\pi} \frac{\theta_b}{g} \left( \beta L_z \right)^{1/2}. \]  
(2.15)

Alternatively, using the counterpart of (2.4) for the Froude number asymptotics of the equatorial regions, one can see from \(u > NL_z/(2\pi)\) that \(F_{\text{min}} = 1/(2\pi)\) is the minimum value of the Froude number for the activation of convective sources. Using (2.14), in turn this sets a minimum or threshold for diabatic heating, that is,

\[ Q_{\text{th}} = Fr_{\text{min}} \frac{\theta_b}{g} \frac{NL_z}{\beta L_z}, \]  
(2.16)

FIG. 1. Six areas over the oceans chosen in this study.
Consistent with the procedure that will be presented in section 2c for IGWs, here $F_{\text{res}}$ is set to 0.3.

A summary of the main relations, thresholds, and parameters involved in the energy parameterization is given in Tables S1 to S3 in the online supplemental material.

c. Data and methodology

To implement and assess the energy parameterization for regions across the globe, use is made of the ERA5 dataset, the fifth edition of high-resolution daily data from the ECMWF reanalysis. The data used are for four years from December 2015 to November 2019 at 6-hourly intervals. The reanalysis data are available hourly at horizontal resolution of 31 km on 137 model levels vertically, from the surface up to the pressure level of 0.01 hPa (around 80 km).

The IGWs in the ECMWF high-resolution data have contributions from both the parameterized and explicitly resolved parts. In this regard, one should note that the minimum resolvable IGWs in ERA5 are typically of the order of $6\Delta x \sim 180$ km horizontally and a few kilometers vertically. It means that the ERA5 data suffer from an underestimation of the gravity waves. Jewtoukoff et al. (2015) considered a factor of 5 to rescale the IGW momentum flux estimated by the ECMWF data with $0.125^\circ \times 0.125^\circ$ resolution in comparison with that from the observational balloons at about 20 km height.

A low-pass spectral filter is used to remove the fluctuations with zonal wavenumbers greater than 22 and thus the signature

![Fig. 2. The monthly mean potential energy (J kg$^{-1}$) of the IGWs identified using ERA5 data at 30 km height for March 2005 to February 2008. Panels show (a) January to (l) December.](image-url)
of IGWs (Sato et al. 2009, 2012). Such a choice corresponds to a horizontal length of about 1800 km in the tropics and 900 km in the midlatitudes. It cannot be ruled out that synoptic-scale balanced structures may be misrepresented as IGWs, but a further refined separation between balanced and unbalanced modes of motion is a topic of its own (Mirzaei et al. 2017) beyond the scope of the present work. For this reason, we use the spatial filter here and focus on seasonal and regional variability of the sources and the waves. Readers interested in the theoretical aspects of the spatial-scale separation underpinning our work may consult with Aspden and Vanneste (2010). The parameterization relations are applied to the low-pass-filtered data thus constructed. For temperature $T$, zonal velocity $u$, and meridional velocity $v$, the low-pass fields are subtracted from the actual fields to obtain the perturbations associated with IGWs, denoted respectively by $T^0$, $u^0$, and $v^0$, from which the specific IGW energy is diagnosed by

$$E = \frac{1}{2} \left( u^2 + v^2 + \frac{g^2}{N^2} \frac{T^2}{T^2} \right),$$

(2.17)

where $T$ denotes the low-pass temperature field. To adjust the parameterized IGW energy to the diagnosed IGW energy, a regression analysis is carried out. For regression, the linear, exponential, power and a combined exponential and power functions written as

$$E_{1,p} = aE_{0,p},$$

(2.18)

$$E_{2,p} = a \exp \left( -b/E_{0,p} \right),$$

(2.19)

FIG. 3. As in Fig. 2, but for the GRACILE data obtained by HIRDLS satellite instrument (Ern et al. 2018).
are assessed, where $E_{0,p}$ and $E_{i,p}(i = 1, \ldots, 4)$ are the initial parameterized energy from (2.8) and the four regression’s model parameterized energies, respectively. The form of regression functions ensures that the parameterized energies tend to zero in any case, corresponding to no wave for no source. The exponentially small forms (2.19) and (2.21) are motivated by the asymptotically weak radiation hypothesized by Vanneste and Yavneh (2004). The regression coefficients $a$, $b$, and $c$ are considered as geographical and seasonal factors which are estimated with respect to IGWs activity in different seasons (spring, summer, autumn, and winter) and domains (the Northern and Southern Hemispheres, and tropical regions).

In dealing with the complications arising from the horizontal propagation of gravity waves in the troposphere, the convection-based Froude number $Fr = V/(NL_{\text{conv}})$ is turned out to be useful. In flows passing a region of convection with vertical scale of $L_{\text{conv}}$, for $Fr \ll 1$, waves can propagate continuously upstream of the convection in analogy with airflow over the mountain (Kim et al. 2003; Durran 2015). As shown later, discarding regions with $Fr \leq Fr_{\text{mas}}$ acts to filter the smaller-amplitude waves that fill the domain and arise from horizontal propagation following generation from rather compact sources. In what follows, empirically $Fr_{\text{mas}}$ has been set to 0.3.

In this study, six domains or areas are considered over the oceans far away from the significant orographies to avoid the mountain waves. The input for the source energy parameterizations is derived from wind, temperature and humidity data. Being not available in ERA5 dataset, latent heating was estimated as described in appendix A. The parameterized and diagnosed IGW energies are calculated for each grid box and then integrated vertically from 1000 to 100 hPa using $\int Edp/dg$ to derive the IGW energy integral in the tropospheric column. Then, for the six domains shown in Fig. 1, the parameterized and diagnosed IGW energies are analyzed statistically using the scatterplots, regression to assess the geographical–seasonal factors, and various error measures (appendix B).

d. de la Cámara and Lott (2015) jet–front energy parameterization

While the focus is on the development and assessment of the MZMAP relations, it is useful to make a comparison with the other available source parameterization for jets and fronts presented in de la Cámara and Lott (2015). Based on the analysis of gravity wave emission by vertically sheared potential vorticity anomalies (Lott et al. 2010, 2012), in de la Cámara and Lott (2015) the following expression for the magnitude of Eliassen–Palm (EP) flux has been applied for the jet–front contribution to nonorographic IGWs:

$$\left| F^{(c)} \right| = G \frac{(\Delta z)^2}{4f} \rho_0 N(z) q^2 \exp(-\pi |N/U_z|),$$

in which $F^{(c)}$ is the vertical component of EP flux, $G$ is an order one tunable parameter, $\Delta z$ the depth of potential vorticity anomaly $q$, $\rho_0$ is the background density, and $U_z$ is the vertical shear of horizontal wind. To apply (2.22), de la Cámara and Lott (2015) take $q \approx \rho_0 \frac{1}{\vert \Delta z \vert} \partial \theta_0 / \partial z$. For comparison, we need to turn the estimate for EP flux to an estimate for energy. To this end, we use the property $F^{(c)} = c_r^{(c)} A$ with $c_r^{(c)}$ the vertical component of group velocity and $A$ the wave activity density, take the relation $A = -k \rho_0 \epsilon / \omega$ [see Eq. (4A.12) in Andrews et al. 1987], and exploit dispersion relation for midfrequency gravity waves $\omega = -Nk/m$. Here $\omega$ is the intrinsic frequency, and $k$ and $m$ are, respectively the zonal and vertical wavenumbers. In this way, we arrive at $|F^{(c)}| = \left( \frac{k \rho_0 \epsilon}{\vert \Delta z \vert} \right) \sim (U_z/L_o) \rho_0 \epsilon \sim 10^{-2} \rho_0 \epsilon$. The energy estimates for the jet and font sources are then obtained as $e^{\text{jet}} + e^{\text{front}} = 10^2 \left| F^{(c)} \right| / \rho_0$, which together with $e^{\text{conv}}$ form the basis for regression model assessment as carried out for the extended MZMAP relations. In applying (2.22), $G$ is taken equal to 3 as in de la Cámara and Lott (2015) and $\Delta z$ is set to grid distance in vertical integral of energy.

3. Results

a. Comparison with stratospheric observations

Before presenting the main body of results, let us first demonstrate the realism of IGWs deduced from the ERA5 data by a comparison with the global climatology of waves in the HIRDLs subset of GRACILE data (Ern et al. 2018) obtained through satellite observations of the stratosphere and mesosphere. Compared with the 31 km average horizontal grid distance in the ERA5 data, the horizontal resolution of the HIRDLs data is determined by the 90 km distance between two consecutive scans of the instrument (Ern et al. 2018; Meyer et al. 2018). For this, Figs. 2 and 3 are presented showing the monthly mean values of the specific potential energy of IGWs at 30 km height by the filtered ERA5 and the GRACILE datasets over the period March 2005 to January 2008 for which the satellite observations are available. Throughout
the seasonal cycle, qualitative agreement is seen between the two wave fields in spatial distribution of wave activity associated with the main mountain ranges, the midlatitude storm tracks, the intertropical convergence zone and areas of monsoon. For a quantitative assessment, shown in Fig. 4 are the annual cycles of the monthly mean values of the specific potential energy for the areas 1 and 2 [Southern Hemisphere (SH)], 3 and 4 [tropical region (TR)], 5 and 6 [Northern Hemisphere (NH)] (Fig. 1). Compared to the GRACILE data, the signature of IGWs in the ERA5 data is weaker by a factor of about 2 in the SH, 1.2 to 2.8 depending on season in the NH, and 1.4 in the TR regions. The difference between the two datasets is particularly marked around the peak of activity in the SH and NH regions. The ERA5 data also exhibit significantly less annual variation of IGW activity especially in the NH. This is opposite to the large seasonal variation that has been reported for climate models in Geller et al. (2013) (see their Fig. 2). Such discrepancy is likely due to the impact of orographic gravity.

Fig. 5. The local values of (a) diagnosed and (b) initial parameterized IGW energy in units of m$^2$ s$^{-2}$. (c) Horizontal wind speed for values greater than 50, with a 5 m s$^{-1}$ contour interval and $|\mathbf{v}_j|$ (color shaded) over the area 5 at the 250 hPa level for 0600 UTC 8 Feb 2016.

Fig. 6. The local values of (a) diagnosed and (b) initial parameterized IGW energy in units of m$^2$ s$^{-2}$. (c) Frontogenesis function [gray shaded, unit: K (100 km)$^{-1}$ h$^{-1}$] and potential temperature (red dashed contours, unit: K) over the area 1 at the 350 hPa level for 0000 UTC 17 Aug 2016.
waves in their global analyses, which are not considered here. Given the relation $|P(z)| = 2\rho_0 k/m |\epsilon_{pot}|$ between the absolute momentum flux and the specific potential energy averaged over one wave cycle ($\epsilon_{pot}$) (see also relations (9) and (11) in Ern et al. 2018, and discussions there), one may also infer the information on the absolute momentum flux by considering $\rho_0 \approx 1.8 \times 10^{-5} \text{kg m}^{-2}$ at $z = 30 \text{km}$ and the ratio $|k/m| \approx 10^{-2}$.

Despite the absence of a global observational dataset similar to GRACILE for the troposphere, from the link between the stratospheric waves and the tropospheric sources in terms of the spatial distribution of energies, one can infer a corresponding overall agreement of the wave sources between the ERA5 data and observations in the troposphere.

b. Typical cases

At each pressure level, the parameterization relations have been implemented in order to estimate the energy related to each source of IGWs in the troposphere and make comparison with the diagnosed IGW energy. Figures 5–7 present cases of the diagnosed ($E$) and the initial parameterized energy ($E_{0,p}$) over the areas 5, 1, and 4, respectively. Also shown in these

![Figure 7](image-url)
figures are the snapshots of meteorological fields related to the likely sources of gravity waves over the area of study at a certain pressure level. Referring to Fig. 5, it can be seen that the broad features of the diagnosed IGW energy are captured by the parameterization of the IGW source due to jet, considering the exit region and the curved area of the jet stream at the 250 hPa level with the associated cross-stream ageostrophic wind over the pacific ocean at 0600 UTC 8 February 2016. In the case shown in Fig. 6, where the connection to the upper-level front is evident, the initial parameterization relations suffer from a gross underestimation of the diagnosed energy. For comparison, shown in Figs. S1 and S2 of the online supplemental material are the local diagnosed and initial parameterized energies using the MZMAP and de la Cámara and Lott (2015) relations for the cases of Figs. 5 and 6, respectively. Due to differences in formulation and the fact that the procedure of de la Cámara and Lott (2015) makes no distinction between jets and fronts, the maxima of local parameterized energies deviate substantially from the sources defined in MZMAP and thus differ in location and magnitude. Such differences make it difficult to judge the relative performance of MZMAP versus de la Cámara and Lott (2015) parametric relations based on single events.

For the convection, one can find a good agreement in magnitude and location between $E$ and $E_{0,p}$ in Figs. 7a and 7b, and between $E_{0,p}$ and the distribution of latent heat release at the 500 hPa level presented in Fig. 7c. It should be noted that the apparent source around 4°S and between 150° and 140°W due to jet, where $|v|$ exceeds the threshold of 1.3 m s$^{-1}$ is discarded by the Froude number filter described above.

### c. Source distributions

Both of the calculated energies have been taken care of by quality controls. As discussed in section 2, the Froude number is taken into account to deal with the horizontal propagation of gravity waves in the troposphere. Figure 8 shows how applying the criterion of Fr $\leq$ 0.3 can significantly reduce the adverse effect of fluctuations coming from horizontal wave propagation in the troposphere. As an indicator of IGWs, shown in Fig. 8c is the field of horizontal divergence at the 350 hPa level.

The domain average values of the initial parameterized energy integral, i.e., vertically integrated energy per unit area, associated with each of the IGW sources due to jet, front and convection have been presented in Fig. 9 for the SH, TR, and NH domains in months of June and November 2016. Based on the estimates provided, jet is the dominant source in each of SH, TR, and NH. The distribution of IGW source due to jet in TR can be understood in light of the results in Trenberth et al. (2000) who analyzed the pattern of global divergent circulation and defined two dominant modes. With a simple structure, the first mode covers the whole troposphere and explains about 60% of the annual cycle variance of divergent mass circulation. Being relatively shallow (roughly from 750 to 350 hPa), the second mode accounts for 20% of the variance and corresponds to a heat trough (Ramage 1971). A strong divergent velocity is present over the Indian Ocean during JJA to DJF (December–February), which is related to the first mode and oriented toward the winter hemisphere. The second mode also includes strong divergent velocity over Africa all around the year and Indian Ocean during JJA. Therefore, very large values are estimated for $E_{0,p}$ in relation to jets in the tropical regions especially during the months of May–November over the Indian Ocean. The jet-related energy over the tropical regions is significantly reduced after removing the sectorial zonal mean divergent wind (Figs. 9c,d). The final result agrees qualitatively with the monthly mean jet stream at the 250 hPa in June and November (Figs. 10a,b). The jet-generated IGWs in each hemisphere are stronger in winter than in summer. Surprisingly, in November 2016 the waves in the SH summer are stronger than those in the NH winter (Fig. 9b). But this corresponds to a strong subtropical jet in the SH which is of similar magnitude as the NH jet (Fig. 10b). The localization of imbalance-generated IGWs in the winterly subtropical jets has also been detected with

---

**FIG. 8.** The local values of the diagnosed IGW energy in units of m$^2$ s$^{-2}$ for (a) before and (b) after applying the Froude number filter in order to account for horizontal wave propagation. (c) Horizontal divergence ($10^5$ s$^{-1}$) at the 350 hPa level for 0000 UTC 22 Aug 2016.
vorticity anomalies (de la Cámara and Lott 2015) while NBE residuals appeared at selected spots (Chun et al. 2019). For such an analysis, in addition to the mathematical formulas, details of data treatment including filtering are important. Because we have used low-pass-filtered data as forcing, we expect our results to be robust, in the sense of having small sensitivity to the parameters, with regard to the large-scale origin of imbalance radiation.

The patchy structure of the IGW source due to front with its rather weak maxima (Figs. 10c,d), which are comparable to the results presented in Fig. 1 of Charron and Manzini (2002) and Chun et al. (2019), is responsible for the weakness of \( f_{\text{front}} \). The monthly and geographical variation of the convection source can be understood using the vertically averaged estimates of latent heat released by condensation presented in Figs. 10e and 10f. The signatures of intertropical convergence zone (ITCZ) and the Atlantic and Pacific storm tracks in the two hemispheres (Adler et al. 2003; Kalisch et al. 2016) are manifest. It is likely that the peaks of convective heating are decreased due to the low-pass filter used. The presence of sharper maxima in the storm tracks, particularly in early winter, may explain the larger values of the IGW source due to convection in the NH and SH over the TR. There is some indication that latent heat release may also be associated with midlatitude orography (see Fig. 12 in Molod et al. 2015), but we cannot rule out methodological reasons here due to the way latent heating is estimated indirectly as described in appendix A [for an alternative method see Haghighatnasab et al. (2020)].

The seasonal mean parameterized and diagnosed IGW energies are shown in Fig. 11. Although the initial parameterized energies are about a factor of 3 too small, their correlation with the diagnosed energy (0.79) is significant and explains about 63% of variance. Additionally, a seasonal cycle is detected in the extratropical regions with a maximum in winter and a minimum in summer. This means that a relevant part of the physical forcing is included in the initial parameterized energy. In the monthly mean, the final parameterized energies coming from the MZMAP relations and de la Cámara and Lott (2015) procedure closely agree and give a fairly accurate representation of the annual cycle in the diagnosed energies (Fig. 12). In comparing the two initial parameterized energies in Fig. 12, one should note the effect of the tunable parameter \( G \) in (2.22) without which the estimates by de la Cámara and Lott (2015) procedure are about a factor of 3 smaller than those by the MZMAP. The seasonal variability of tropospheric wave sources as shown with the diagnosed energy integral in Fig. 12.
amounts to 11% in the SH and 13% in the NH which are small but notable. A question of fundamental importance arises here on the relative significance of the nonorographic IGW variability at the tropospheric source versus the effect of filtering by the background wind and damping during vertical propagation (Plougonven et al. 2017) on the seasonal variability of IGWs higher in the stratosphere and mesosphere. Addressing this question requires implementing the full life cycle of IGWs from tropospheric generation to dissipation in the middle atmosphere within a GCM.

The authors are aware of possible systematic errors in either the underestimation of parameterized convective forcing or the overestimation of gravity waves from high-pass-filtered data. Further, details of vertical wave propagation are required for an accurate translation of the tropospheric energy integrals to the stratospheric momentum fluxes. Nevertheless, it is worth making comparison with the momentum flux estimates in Fig. 4 at 30 km height. Considering the flux–energy relation, an estimate can be made for our peak height-integrated density-weighted energy, that is

\[
E(8 \text{ km}) = \int_0^h \rho \omega dz \sim 160 \text{kJ m}^{-2}
\]

from which

\[
|F^c(8 \text{ km})| \sim 0.2 \text{Pa}
\]

is inferred. Taking into account the damping of IGWs with height during vertical propagation, a gross estimate for the momentum flux at \( z = 30 \text{ km} \) can be made by considering a constant vertical damping length \( h \) such that

\[
|F^c(30 \text{ km})| \sim |F^c(8 \text{ km})| \exp(-22/h). \]

For a damping length of 5 km, one can obtain

\[
|F^c(30 \text{ km})| \sim 2.5 \text{mPa}
\]

consistent with the estimates in Fig. 4. This required damping height is within the range of [2.6, 10] km obtained based on the satellite data in Geller et al. (2013) and the ECMWF-IFS operational analysis in Schoon and Zülicke (2018).

\[d. \text{ Statistical analysis}\]

To determine the geographical and seasonal factors of generation, using the scatterplots of \( E \) against \( E_{0,p} \) with daily

\[\text{FIG. 10. The vertically averaged monthly mean of (a),(b) horizontal wind speed (gray shaded, for values greater than 12 m s}^{-1}, \text{ with a 3 m s}^{-1}\text{ interval); (c),(d) the frontogenesis function [gray shaded, for positive values, with a 0.01 K (100 km)}^{-1}\text{ h}^{-1}\text{ interval]; and (e),(f) the latent heat released during condensation (gray shaded, for positive values, with a 0.02 K h}^{-1}\text{ interval] in (a),(c),(e) June and (b),(d),(f) November 2016.}\]
data, the four regression functions introduced in section 2c are considered. The results of the regression analysis carried out using these four functions have been depicted in Figs. 13a and 13b for January over the area 5 and June over the area 3, respectively. As one can see, the simple linear function fails to explain the relation between the diagnosed and initial parameterized energies and thus fails to serve as a regression model. Similar statement can be made when $e_{\text{jet}} + e_{\text{front}}$ is calculated using de la Cámara and Lott (2015) procedure as detailed in section 2d. Based on the root-mean-square error (RMSE) as the primary goodness of fit measure, the power function obtains the best scores among the functions examined. To make the regression’s model consistent with exponential asymptotics (Vanneste 2008) in the limit when the IGW sources due to jets and fronts are parameterized using de la Cámara and Lott (2015) procedure. The dashed and solid lines are for the SH and NH regions, respectively. The final parameterized energy integrals are those coming from regression using the combined function in (2.21). The inset in the lower-left-hand corner is to magnify the scatterplot for the tropical regions.

![Figure 11](image1.png)

**Fig. 11.** Scatterplot of the diagnosed vs the initial (red) and the final (black) parameterized energy integrals for the averages taken over the Northern Hemisphere, Southern Hemisphere, and the tropical regions denoted by, respectively, N, S, and T as well as the four seasons denoted by DJF, MAM, JJA, and SON. The final parameterized energies are those coming from regression using the combined function in (2.21). The inset in the lower-left-hand corner is to magnify the scatterplot for the tropical regions.

Having obtained the coefficients of regression, the statistical performance measures (appendix B) are used to evaluate the regression models. Tables 1 and 2 present the evaluation of four regression models related to Figs. 13a and 13b. The corresponding results for de la Cámara and Lott (2015) procedure are given in Table S4 for January.

A perfect model will have $(MG, VG, R) \rightarrow 1.0$, $(FB, NMSE) \rightarrow 0.0$, and $\text{fac2} \rightarrow 100\%$, where MG and VG are the geometric mean bias and variance, respectively; $R$ is the correlation coefficient; FB is the fractional bias; NMSE is the normalized mean square error; and fac2 is the factor of 2 as defined in appendix B. Following the suggestion by Kumar et al. (1993), the performance of a model can be regarded acceptable if

$$\text{NMSE} \leq 0.5, \quad -0.5 \leq \text{FB} \leq 0.5, \quad \text{fac2} \geq 80.$$  \hspace{1cm} (3.1)

The two additional criteria introduced by Ahuja and Kumar (1996) to test model reliability

$$0.75 \leq (MG, VG) \leq 1.25,$$  \hspace{1cm} (3.1)

have also been considered as well as the indices of agreement $d$ and $d_i$ (Willmott 1981; Willmott et al. 2012), which describe the relative covariability of $E_{\text{ip}}(i = 1, \ldots, 4)$ and $E$.

Regarding maximum values of explained variance (EV) and minimum values of RMSE, the power, exponential and combined functions attain the best scores in model evaluation of the investigated six areas (Tables 1, 2). That is, the steep descent at small values seems to be an essential property of the fit. Here, for its asymptotic consistency, the combined function is adopted as the regression model (Table 3 and Table S5) and used to estimate the geographical and seasonal coefficients of generation. The scatterplots of $E_{\text{ip}}$ and $E$, and the related combined function governing the data are presented in Figs. 14–16 for the NH, TR, and SH, respectively. By numerical experiment, in all but a few cases, the exponent $c$ in the combined function is set to 0.25. The only exception is the Northern Hemisphere in summer for which $c$ is set to zero. The estimated geographical and seasonal coefficients are summarized in Table 4 for the generalized MZMAP and Table S6 for de la Cámara and Lott (2015) procedure. Overall, the variation in the regression coefficients $a$, $b$, and $c$ of the combined function seems small. To some extent, the same can be said for the

![Figure 12](image2.png)

**Fig. 12.** Annual cycle of the diagnosed (thick black), initial MZMAP (thick red), and final MZMAP (thin red) energy integrals. Also shown are the initial (thick blue) and final (thin blue) energy integrals when the IGW sources due to jets and fronts are parameterized using de la Cámara and Lott (2015) procedure. The dashed and solid lines are for the SH and NH regions, respectively. The final parameterized energy integrals are those coming from regression using the combined function in (2.21).
variation with region. These considerations may point to a universal fit, i.e., one set \((a, b, c)\) for all data irrespective of season and region (see Fig. S3 in the supplementary material). The universal fit performs with an explained variance \((EV = 0.13)\), which is in the upper range of the separate seasonal/regional fits with \(EV = (0.01, \ldots, 0.18)\), even though the tropical regions behave differently. To check robustness, a cross validation has also been undertaken in which the data have been grouped in six permutations formed by taking two years of data to fit and the other two years to test. The resulting mean and standard deviation are given for statistical regression in Fig. S4 and performance indices of \(EV, NMSE\) and \(RMSE\) in Table S7 of the supplementary material. In agreement with the 4-yr statistical analysis presented, the \(NMSE\) and \(RMSE\) measures are sufficiently definite particularly in the SH and NH regions. Indeed the statistical indices resulting from the fit to four years of data as provided with Table 3 are in 33 of 36 cases within the one standard deviation ranges given in Table S7—hence the fits can be considered as robust.

Most of the extratropical fits result in increasing energies below 150 kJ m\(^{-2}\) and decreasing above 200 kJ m\(^{-2}\). Hence, most of the initial energies are below 100 kJ m\(^{-2}\) and are thus corrected to higher values. There is one exception (NH-JJA), which has lower initial energies than the other extratropical cases and for which the combined function realizes a stronger upward-correction by a factor of 10 (see Fig. 11, black and red “JJA-N”). The fits for the tropical cases increase initial energies below 20 kJ m\(^{-2}\) and decrease them above. The exception of the low-latitude fits is TR-DJF—in Fig. 15a it appears to predict smaller initial energies than the others. However, finally it realizes a correction by the factor of 2 (see Fig. 11, black and red “DJF-T”). Hence, the initial energy predictions are too small by factors up to 10. These discrepancies are corrected with the seasonal fits of the combined function to daily data. The steep descent of the combined function for small energies results in a large upward-correction of small initial energies. This can also be confirmed by examining the frequency and probability distributions for the initial parameterized, parameterized, and diagnosed energies provided in Figs. S5–S7 of the online supplemental material for, respectively, the NH, TR, and SH domains. Generally, the regression by combined function rectifies the mean of distribution but still needs improvement in terms of representing the full range of variability in the diagnosed energies. Further

<table>
<thead>
<tr>
<th>Fitted function</th>
<th>(\mathbf{R})</th>
<th>(\mathbf{EV})</th>
<th>(\mathbf{MB})</th>
<th>(\mathbf{FB})</th>
<th>(\mathbf{NMSE})</th>
<th>(\mathbf{MG})</th>
<th>(\mathbf{VG})</th>
<th>(\mathbf{d_c})</th>
<th>(\mathbf{d})</th>
<th>fac2</th>
<th>(\mathbf{RMSE})</th>
</tr>
</thead>
<tbody>
<tr>
<td>SH Linear</td>
<td>0.347</td>
<td>0.000</td>
<td>(-19.598)</td>
<td>(-0.152)</td>
<td>0.179</td>
<td>0.805</td>
<td>1.226</td>
<td>0.048</td>
<td>0.534</td>
<td>88.105</td>
<td>55.218</td>
</tr>
<tr>
<td>Power</td>
<td>0.362</td>
<td>0.131</td>
<td>0.004</td>
<td>0.000</td>
<td>0.040</td>
<td>1.022</td>
<td>1.046</td>
<td>0.535</td>
<td>0.466</td>
<td>99.798</td>
<td>27.876</td>
</tr>
<tr>
<td>Exponential</td>
<td>(-0.012)</td>
<td>0.131</td>
<td>30.604</td>
<td>0.199</td>
<td>0.078</td>
<td>1.251</td>
<td>1.106</td>
<td>0.259</td>
<td>0.445</td>
<td>96.573</td>
<td>27.880</td>
</tr>
<tr>
<td>Combined</td>
<td>0.362</td>
<td>0.131</td>
<td>2.073</td>
<td>0.015</td>
<td>0.040</td>
<td>1.037</td>
<td>1.047</td>
<td>0.534</td>
<td>0.451</td>
<td>99.698</td>
<td>27.932</td>
</tr>
<tr>
<td>Tropical Linear</td>
<td>0.056</td>
<td>0.000</td>
<td>(-23.885)</td>
<td>(-0.176)</td>
<td>0.235</td>
<td>0.784</td>
<td>1.307</td>
<td>(-0.239)</td>
<td>0.321</td>
<td>81.048</td>
<td>65.463</td>
</tr>
<tr>
<td>Power</td>
<td>0.012</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.033</td>
<td>1.019</td>
<td>1.041</td>
<td>0.500</td>
<td>0.017</td>
<td>99.194</td>
<td>27.011</td>
</tr>
<tr>
<td>Exponential</td>
<td>0.023</td>
<td>0.002</td>
<td>3.144</td>
<td>0.021</td>
<td>0.033</td>
<td>1.041</td>
<td>1.043</td>
<td>0.502</td>
<td>0.151</td>
<td>99.194</td>
<td>26.984</td>
</tr>
<tr>
<td>Combined</td>
<td>0.011</td>
<td>0.000</td>
<td>1.568</td>
<td>0.011</td>
<td>0.033</td>
<td>1.030</td>
<td>1.042</td>
<td>0.502</td>
<td>0.079</td>
<td>99.194</td>
<td>27.027</td>
</tr>
<tr>
<td>NH Linear</td>
<td>0.374</td>
<td>0.000</td>
<td>(-35.622)</td>
<td>(-0.218)</td>
<td>0.239</td>
<td>0.681</td>
<td>1.712</td>
<td>(-0.220)</td>
<td>0.462</td>
<td>75.504</td>
<td>80.883</td>
</tr>
<tr>
<td>Power</td>
<td>0.348</td>
<td>0.121</td>
<td>(-0.013)</td>
<td>0.000</td>
<td>0.027</td>
<td>1.014</td>
<td>1.028</td>
<td>0.549</td>
<td>0.450</td>
<td>100.000</td>
<td>29.956</td>
</tr>
<tr>
<td>Exponential</td>
<td>(-0.019)</td>
<td>0.061</td>
<td>8.908</td>
<td>0.048</td>
<td>0.032</td>
<td>1.066</td>
<td>1.036</td>
<td>0.475</td>
<td>0.292</td>
<td>99.899</td>
<td>30.964</td>
</tr>
<tr>
<td>Combined</td>
<td>0.348</td>
<td>0.120</td>
<td>1.374</td>
<td>0.008</td>
<td>0.027</td>
<td>1.022</td>
<td>1.029</td>
<td>0.545</td>
<td>0.428</td>
<td>100.000</td>
<td>29.975</td>
</tr>
</tbody>
</table>
work is planned for future to increase the range of variability by adding stochastic noise to the otherwise deterministic procedure for energy parameterization in the manner that has become standard practice for the physics parameterization in weather and climate models (Berner et al. 2017). Overall, the results of statistical measures of model evaluation are very satisfactory over a wide range of environmental conditions.

While the statistical model performance has been shown only for the cutoff zonal wavenumber $s = 22$ for the separation of the mesoscale IGWs from the large-scale vortical flows, a question arises on the degree of sensitivity to the choice of $s$. To examine that, following the suggestion by Zagar et al. (2017) on the use of $s = 35$ for the separation, which was made based on a linear normal mode analysis of the ECMWF interim reanalysis and operational 2014–16 analysis data, a comparison has been carried out between the statistical performance of regression with the combined function as applied to the data for January and June 2016 and each of $s = 22$ and $s = 35$ values (see Tables S8, S9).

Generally, the expected lower variability in mesoscale activity leads to increase in statistical model performance when the higher value of $s$ is used. This can be judged by, for example, the reductions in RMSE of more than 50% by going to $s = 35$. But the behavior in terms of variations with season and region remain unchanged.

4. Summary and conclusions

The parametric relations for nonorographic inertia–gravity wave energy introduced by MZMAP were generalized for applications in global domains. In addition to having theoretical significance, the parametric relations for energy may be useful as part of a source parameterization for IGWs. MZMAP presented the relations with emphasis on the energy generated in the vicinity of IGW sources in the troposphere and examined the relations on a midlatitude $f$ plane. In the absence of a complete theory explaining wave emission by jets and fronts in real complex flows, MZMAP resort to the unifying principle of relating IGWs energies to the strength of secondary circulations associated with the tropospheric sources. The rationale is the relation between the fastness of the circulation induced by the source and the intensity of wave emission. By definition, the “parameterized” energy quantifies the wave source strength in terms of the imbalance in large-scale flow. For the validation of the parametric relations, correspondingly, the

<table>
<thead>
<tr>
<th>Season</th>
<th>RH</th>
<th>EV</th>
<th>MB</th>
<th>FB</th>
<th>NMSE</th>
<th>MG</th>
<th>VG</th>
<th>$d_r$</th>
<th>$d$</th>
<th>fac2</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SH</td>
<td>0.303</td>
<td>0.090</td>
<td>−1.888</td>
<td>−0.083</td>
<td>0.057</td>
<td>0.943</td>
<td>1.056</td>
<td>0.487</td>
<td>0.422</td>
<td>99.861</td>
<td>31.303</td>
</tr>
<tr>
<td></td>
<td>0.391</td>
<td>0.152</td>
<td>2.067</td>
<td>0.014</td>
<td>0.033</td>
<td>1.033</td>
<td>1.037</td>
<td>0.535</td>
<td>0.481</td>
<td>99.691</td>
<td>25.936</td>
</tr>
<tr>
<td></td>
<td>0.385</td>
<td>0.148</td>
<td>1.779</td>
<td>0.012</td>
<td>0.027</td>
<td>1.026</td>
<td>1.028</td>
<td>0.537</td>
<td>0.487</td>
<td>100.000</td>
<td>24.806</td>
</tr>
<tr>
<td></td>
<td>0.342</td>
<td>0.115</td>
<td>−1.032</td>
<td>−0.008</td>
<td>0.026</td>
<td>1.006</td>
<td>1.027</td>
<td>0.533</td>
<td>0.410</td>
<td>100.000</td>
<td>21.540</td>
</tr>
<tr>
<td>Tropical</td>
<td>0.122</td>
<td>0.011</td>
<td>1.651</td>
<td>0.029</td>
<td>0.146</td>
<td>1.126</td>
<td>1.249</td>
<td>0.500</td>
<td>0.131</td>
<td>88.057</td>
<td>22.015</td>
</tr>
<tr>
<td></td>
<td>0.256</td>
<td>0.060</td>
<td>1.844</td>
<td>0.038</td>
<td>0.169</td>
<td>1.168</td>
<td>1.336</td>
<td>0.519</td>
<td>0.274</td>
<td>82.515</td>
<td>20.007</td>
</tr>
<tr>
<td></td>
<td>0.208</td>
<td>0.037</td>
<td>1.117</td>
<td>0.023</td>
<td>0.092</td>
<td>1.081</td>
<td>1.148</td>
<td>0.512</td>
<td>0.351</td>
<td>92.105</td>
<td>15.157</td>
</tr>
<tr>
<td></td>
<td>0.139</td>
<td>0.007</td>
<td>1.509</td>
<td>0.034</td>
<td>0.115</td>
<td>1.107</td>
<td>1.185</td>
<td>0.508</td>
<td>0.302</td>
<td>89.349</td>
<td>14.806</td>
</tr>
<tr>
<td>NH</td>
<td>0.394</td>
<td>0.154</td>
<td>1.516</td>
<td>0.008</td>
<td>0.030</td>
<td>1.025</td>
<td>1.033</td>
<td>0.550</td>
<td>0.482</td>
<td>99.826</td>
<td>30.997</td>
</tr>
<tr>
<td></td>
<td>0.308</td>
<td>0.088</td>
<td>3.860</td>
<td>0.024</td>
<td>0.028</td>
<td>1.040</td>
<td>1.033</td>
<td>0.523</td>
<td>0.338</td>
<td>99.588</td>
<td>26.994</td>
</tr>
<tr>
<td></td>
<td>0.231</td>
<td>0.053</td>
<td>−0.673</td>
<td>−0.005</td>
<td>0.046</td>
<td>1.024</td>
<td>1.067</td>
<td>0.519</td>
<td>0.272</td>
<td>96.728</td>
<td>30.154</td>
</tr>
<tr>
<td></td>
<td>0.440</td>
<td>0.181</td>
<td>3.112</td>
<td>0.018</td>
<td>0.039</td>
<td>1.043</td>
<td>1.048</td>
<td>0.544</td>
<td>0.478</td>
<td>99.212</td>
<td>33.398</td>
</tr>
</tbody>
</table>
“diagnosed” energy is defined to represent the mesoscale IGW activity.

The ERA5 dataset from ECMWF was used for the period December 2015 to November 2019 at 6-hourly intervals from 1000 to 100 hPa. The large-scale flow was separated from the mesoscale IGW activity using a low-pass spatial filter acting on zonal wavenumbers with a cutoff value of 22. The realism of the identified IGWs was assessed by comparing the potential energy of the IGWs between the ERA5 and the HIRDLS subset of GRACILE data (Ern et al. 2018) at 30 km height. Overall, the waves in the ERA5 data are weaker than those in the GRACILE data by factors of less than 3.0, which vary with region and season. The low-passed filtered data representing the large-scale flow were then used in the parametric relations and subtracted from the actual data to obtain the IGW activity. The mass-weighted tropospheric integral over the layer [1000, 100] hPa of the specific energy resulting from the generalized relations was called initial parameterized energy. The corresponding mass-weighted integral of the mesoscale IGW activity defines the diagnosed energy.

The problem with horizontal wave propagation in the troposphere, leading to dislocations between the sources and the waves induced, has been partially addressed using an empirical Froude number–based criterion. To avoid mountain waves, only areas over the oceans in the tropics, the Northern and Southern Hemispheres were considered. The parametric relations constructed led to a considerable amount of excess energy in tropical region especially over the Indian Ocean in the months between May and November. Since the zonal mean circulation over each sector or tropical region considered is not relevant as an IGW energy source, to address the issue, the sectorial zonal mean part of the divergent wind was removed from the relations.

Analysis of the results showed an overall good agreement in magnitude and location between the initial parameterized and diagnosed energies. At first, the initial parameterized energies were compared with the diagnosed energies on the basis of seasonal regional means (see Fig. 11, black labels). Although the parameterized energies were too small, they correlated well with the diagnosis. A part of the explained variance (about 63%) is found to be due to the seasonal cycle in the extratropics. As an attempt to adjust the wave sources, assuming that our diagnostics of IGWs from ERA5 may represent realistic estimates of wave activity near the sources in the troposphere, regression analysis was carried out with daily data using suitable candidate functions. It was found that the power function can serve as the regression model. The theoretical relation between the initial parameterized energies and ageostrophic Rossby number provides motivation to use exponential
asymptotics. To this end, the power function was combined with a proper exponential function to take control over the behavior at very small values of initial parameterized energy. The regression analysis was performed separately for the four seasons and over different domains in the tropics, the Northern and Southern Hemispheres. Having considered the coefficients of regression as the seasonally and geographically dependent coefficients of generation, the final parameterized energy was then taken to be that from the adopted regression model. The effect of the fitted combined parameterization resulted in a correction of the seasonal mean values (see Fig. 11, red labels). The daily scatterplots (Figs. 13–16) are summarized in Table 3. The correlation between parameterized and diagnosed energies were below 0.5 and explained daily variances of below 18%. This seemingly low performance is not unexpected considering the high spatiotemporal variability of IGWs, dislocations between sources and waves due to wave propagation, and the imperfect separation between unbalanced IGWs and balanced vortical flows. To judge the performance, one should consider that ignoring intraseasonal variability by setting a fixed value for wave energies in each season leads to zero values for both correlation and explained variance between parameterized and diagnosed energies.

The statistical evaluation demonstrated that the regression with the combined function can substantially improve the representation of IGW variability at the source level. This holds true also for the parameterization of IGW sources due to jets and fronts using the formulation based on sheared potential vorticity anomalies. With regression using the combined function, the extended MZMAP relations and de la Cámara and Lott (2015) lead to equally reliable estimates for IGW activity. The improvement by regression models is important considering that most current global models grossly simplify the variability of IGWs and thus their energies at the source level. Further, the steepest descent for weak radiation was found to be important for an effective correction of IGW energy parameterization using ERA5 data. We conclude that such nonlinearities are essential during the gravity wave generation process, though the present data may not be sufficient to validate the specific form of exponentially small radiation suggested by Vanneste and Yavneh (2004). Another shortcoming of the present work is the use of spatial filtering to identify mesoscale gravity waves. This can be remedied by combining spatial filtering with a proper wave–vortex decomposition such as that carried out in Mirzaei et al. (2017) on the $f$ plane where diabatic forcing can also be consistently estimated by generalizing the method proposed in Haghighatnasab et al. (2020). The extent to which such combined procedure may change the performance of the model remains to be seen. Another issue for future research is the proper inclusion of tropical waves.

FIG. 15. As in Fig. 14, but over areas 3 and 4 (tropical).
Shared by the two methods of the extended MZMAP and de la Cámara and Lott (2015) examined here, the seemingly low statistical performance, as seen for example by the explained variance, seems to be a general property coming from the uncertainties and high variability in the diagnosed energies. It remains to be seen if any significant improvement can be made on the statistical performance by reducing the uncertainties in the diagnosed energies and/or involving the techniques of machine leaning and stochastic parameterization.

**Acknowledgments.** We thank the University of Tehran for providing support during this research. The studies of CZ were partly funded by the Deutsche Forschungsgemeinschaft through Grant ZU 120/2-2 for the research unit FOR 1898 (Multiscale dynamics of gravity waves/Spontaneous Imbalance). We thank the three anonymous reviewers for their constructive comments and helpful suggestions.

**Data availability statement.** The ERA5 data have been retrieved from https://doi.org/10.24381/cds.bd0915c6, the GRACILE data have been obtained from https://doi.org/10.1594/PANGAEA.879658, and the data processed for this paper are available from https://doi.org/10.22000/347.

**APPENDIX A**

### Condensational Heating

To estimate diabatic heating $Q$, the form for latent heating given by Emanuel et al. (1987) is used which is written in the pressure vertical coordinate as

$$Q = \omega \left( \frac{\partial \theta}{\partial p} + \Gamma_d \frac{\partial \theta}{\partial e} + \Gamma_m \frac{\partial \theta}{\partial \theta} \right),$$

(A.1)

where $\omega$ is the pressure vertical velocity, $\theta_e$ is the equivalent potential temperature, and $\Gamma_d$ and $\Gamma_m$ denote the dry and moist adiabatic lapse rates, respectively.
APPENDIX B
Statistical Performance Measures

For completeness, a brief explanation of the performance measures used for model evaluation is provided, with $O$ denoting the “observation” for the diagnosed energy and $P$ as the model “prediction” for the parameterized energy. “Pearson’s correlation coefficient” or commonly called “the correlation coefficient” or $R$ estimates the dependence or association of the two variables:

$$R = \frac{(O - \overline{O})(P - \overline{P})}{(P - \overline{P})^{1/2}(O - \overline{O})^{1/2}}, \quad (B.1)$$

where $\overline{O}$ and $\overline{P}$ are the mean values of the diagnosed and parameterized energies, respectively. The meaning and range of variation for $R$ is standard and not repeated here for brevity. Explained variance (EV) score, as the name implies, measures the ratio between variance of error and variance of observations. Alternatively, this score measures how well a model can explain variations in the dataset. EV is defined by

$$EV = 1 - \frac{(O - P)^{(O - P)}}{(O - \overline{O})^2}, \quad (B.2)$$

The highest value of EV that a model can attain is 1.0. Model bias is the mean error of the model predictions and the reference values:

$$MB = \overline{P} - \overline{O}, \quad (B.3)$$

which will be zero for a perfect model. Fractional bias is the normalized MB defined by

$$FB = \frac{P - \overline{O}}{0.5(P + \overline{O})}, \quad (B.4)$$

which varies between $-2$ and 2 and will be zero for a perfect model. Positive (negative) values indicate the model overestimation (underestimation). Normalized mean square error (NMSE) shows the scatter in the entire dataset:

$$NMSE = \frac{(P - O)^2}{\overline{P}\overline{O}}, \quad (B.5)$$

Smaller values of NMSE denote better model performance. The geometric mean bias is defined by

$$MG = \exp(\overline{P} - \overline{O}), \quad (B.6)$$

This measure is useful when the ratio $O/P$ (or its inverse) is on the order of 10100 and/or greater. It will take 1 for a perfect model, which does not mean necessarily that the observations and the model predictions are exactly the same. The concern is their orders. Values greater (less) than unity yield the model overestimation (underestimation). Defined by

$$VG = \exp(\ln P - \ln O)^2, \quad (B.7)$$

The geometric mean variance measures the variance of the datasets independently from the absolute values of the data. The indices of agreement $d$ (Willmott 1981) and $d_r$ (Willmott et al. 2012) are defined by

$$d = 1 - \frac{(P - O)^2}{(|P - O| + |O - \overline{O}|)^2}, \quad (B.8)$$

and

$$d_r = \begin{cases} \frac{1 - |P - O|}{c|O - \overline{O}|}, & \text{when } |P - O| \leq c|O - \overline{O}|, \\ \frac{|O - \overline{O}|}{c|P - O|} - 1, & \text{when } |P - O| > c|O - \overline{O}|, \end{cases} \quad (B.9)$$

with $c = 2$. The range of values for $d$ is $[0, 1]$ with $d = 1$ for a perfect model and $d = 0$ indicating no significant agreement between the model predictions and the observations. The refined index of agreement $d_r$ ranges between $-1$ (poor model estimation) and 1 (a perfect model). Defined by the percentage of the model predictions within a factor of 2 of the observations, that is,

$$\frac{1}{2} \leq \frac{P}{O} \leq 2, \quad (B.10)$$

where the factor of 2 (fac2) takes 100% for a perfect model. As a standard measure defined by

$$RMSE = \sqrt{\overline{(P - O)^2}}, \quad (B.11)$$

with zero value for a perfect model, it determines the distribution of data (pairs of model prediction and observation) around the line of the best fit.

REFERENCES


Ahuja, S., and A. Kumar, 1996: Evaluation of MESOPUFF-II-SOx transport and deposition in the Great Lakes region. AWMA Specialty Conf. on Atmospheric Deposition to the Great Lakes, VIP-72, AWMA, 283–299.


Ramage, C. S., 1971: 


