



**HAL**  
open science

## Using Machine Learning to Estimate Nonorographic Gravity Wave Characteristics at Source Levels

Mozhgan Amiramjadi, Riwal Plougonven, Ali R. Mohebalhojeh, Mohammad Mirzaei

► **To cite this version:**

Mozhgan Amiramjadi, Riwal Plougonven, Ali R. Mohebalhojeh, Mohammad Mirzaei. Using Machine Learning to Estimate Nonorographic Gravity Wave Characteristics at Source Levels. *Journal of the Atmospheric Sciences*, 2023, 80, pp.419-440. 10.1175/JAS-D-22-0021.1 . insu-04196539

**HAL Id: insu-04196539**

**<https://insu.hal.science/insu-04196539v1>**

Submitted on 5 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Using Machine Learning to Estimate Nonorographic Gravity Wave Characteristics at Source Levels

MOZHGAN AMIRAMJADI<sup>1</sup>,<sup>a</sup> RIWAL PLOUGONVEN,<sup>b</sup> ALI R. MOHEBALHOJEH,<sup>a</sup> AND MOHAMMAD MIRZAEI<sup>a</sup>

<sup>a</sup> Space Physics Department, Institute of Geophysics, University of Tehran, Tehran, Iran

<sup>b</sup> Laboratoire de Météorologie Dynamique/IPSL, Ecole polytechnique, Institut Polytechnique de Paris, Paris, France

(Manuscript received 29 January 2022, in final form 26 September 2022)

**ABSTRACT:** Machine learning (ML) provides a powerful tool for investigating the relationship between the large-scale flow and unresolved processes, which need to be parameterized in climate models. The current work explores the performance of the random forest regressor (RF) as a nonparametric model in the reconstruction of nonorographic gravity waves (GWs) over midlatitude oceanic areas. The ERA5 dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) model outputs is employed in its full resolution to derive GW variations in the lower stratosphere. Coarse-grained variables in a column-based configuration of the atmosphere are used to reconstruct the GWs variability at the target level. The first important outcome is the relative success in reconstructing the GW signal (coefficient of determination  $R^2 \approx 0.85$  for “E3” combination). The second outcome is that the most informative explanatory variable is the local background wind speed. This questions the traditional framework of gravity wave parameterizations, for which, at these heights, one would expect more sensitivity to sources below than to local flow. Finally, to test the efficiency of a relatively simple, parametric statistical model, the efficiency of linear regression was compared to that of random forests with a restricted set of only five explanatory variables. Results were poor. Increasing the number of input variables to 15 hardly changes the performance of the linear regression ( $R^2$  changes slightly from 0.18 to 0.21), while it leads to better results with the random forests ( $R^2$  increases from 0.29 to 0.37).

**KEYWORDS:** Gravity waves; Inertia-gravity waves; Machine learning

### 1. Introduction

Atmospheric gravity waves (GWs) constitute an essential driver of the middle-atmosphere circulation (Andrews et al. 1987; Fritts and Alexander 2003). Their wavelengths span a wide range of scales, from a few to thousands of kilometers in the horizontal, a portion of which will remain subgrid scale in general circulation models (GCMs) in the foreseeable future. While this requires the global effects of the GWs to be taken into account by means of parameterizations (Kim et al. 2003), the procedures employed introduce significant uncertainties in GCMs and climate projections. Any improvement of parameterizations needs to deal with certain problems, the first of which is lack of enough global observation of fine-scale motions in a wide range of heights, including remote

regions of the upper stratosphere and mesosphere. The second problem is that of the incomplete state of theoretical modeling of the generation mechanisms, especially for non-orographic GWs (Plougonven and Zhang 2014). The third problem is related to the challenges of validating parameterizations. To illustrate the latter problem, the role of GWs in the breakdown of the austral polar vortex in the spring can be considered. It has been shown that climate models lack sufficient GW drag in the stratosphere during springtime (McLandress et al. 2012), which contributes to persistent model biases (Butchart et al. 2010). Different modeling groups have chosen different strategies to enhance the forcing from GWs on the austral polar vortex, via changes to different parameterizations, enhancing the orographic GWs (Garcia et al. 2017) or the stratospheric forcing from the nonorographic GWs (de la Cámara et al. 2016; Garcia et al. 2017). However, the relative contributions of orographic and nonorographic waves to this “missing drag” remains a matter of debate. In other words, global observations of gravity waves in the middle atmosphere are still not precise enough to efficiently constrain parameterizations, and the criteria tied to improved climatology for the stratospheric circulation do not provide a sufficient constraint for validating different choices on gravity wave parameterizations (Plougonven et al. 2020).

<sup>1</sup> Denotes content that is immediately available upon publication as open access.

Amiramjadi's current affiliation: Leibniz Institute of Atmospheric Physics, University of Rostock, Kühlungsborn, Germany.

Corresponding author: Mozghan Amiramjadi, amiramjadi@iap-kborn.de

DOI: 10.1175/JAS-D-22-0021.1

© 2023 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

For the nonorographic waves, convection produces waves on a wide range of scales: the large-scale envelope of convection may contribute to large-scale GWs, while the convective cells lead to smaller-scale GWs with different characteristics (Fovell et al. 1992; Alexander et al. 1995; Lane et al. 2001). For waves excited by jets and fronts, there is still no complete theoretical framework to relate the characteristics and amplitudes of the emitted waves to the background flow (Plougonven and Zhang 2014). Observational case studies (e.g., Uccellini and Koch 1987) and idealized simulations (e.g., O'Sullivan and Dunkerton 1995) have also indicated that jet exit regions are favored locations for large-amplitude, subsynoptic GWs. The picture that emerges however may be more subtle because some recent idealized simulations highlight the whole region of the jet favorable for GW emission (Waite and Snyder 2009).

Parameterizations in climate models are required to represent the effects of processes that are not directly resolved on the flow that is resolved. In other words, they represent the effects of subgrid-scale processes on the large-scale flow. Traditional schemes are based on available physical considerations, and relate the subgrid-scale processes to resolvable coarse-grained variables, through explicit functions and relations. Although incorporating physical considerations as much as possible, these parameterization schemes retain empirical aspects and include parameters that remain poorly constrained (Alexander et al. 2010). Their parameters can therefore be used for *tuning* the models (Hourdin et al. 2017). Information that contributes to constrain parameterizations includes observations and high-resolution simulations. Local and global observational datasets will each have their own limitations. For instance, global observations derived from satellite observations retain limitations in resolution (e.g., Hindley et al. 2019). Processes like gravity waves, although unresolvable in climate models, are nonetheless *computable* as they result from the well-established Navier–Stokes equations in a stratified fluid. High-resolution simulations hence also provide valuable and reliable information to guide parameterizations (Stephan et al. 2019).

Parameterizations constitute a major source of uncertainty in climate and weather models, and modeling subgrid-scale phenomena such as convection and gravity waves from the knowledge of resolved flow, at a moderate computational cost, remains a challenge. New methods like machine learning (ML) and data assimilation provide novel approaches to increase the performance of parameterization schemes by learning from observations and targeted high-resolution simulations. This kind of systematic learning can be carried out either online by nesting from a local high-resolution domain to its parent domain in GCMs, or offline using pre-created libraries from training data.

Schneider et al. (2017) considered the subgrid-scale processes such as those in cloud turbulence as computable in the sense that their governing equations are known. While they cannot be resolved in climate models, they can be explicitly represented in high-resolution simulations. The computable processes are then made distinct from the noncomputables such as biogeochemical processes whose governing equations

are either poorly known or remain still unknown. They argued that the parameterization schemes that employ governing equations for computable processes can benefit from a more reliable training, because the knowledge of the governing equations allows to predict samples that may not exist in training observational datasets. The situation should be compared with that of noncomputable processes for which the only available information for training comes from the observational data.

Neural networks trained on a high-resolution simulation have been employed to parameterize convection by Gentine et al. (2018). They showed that their “cloud brain” skillfully predicts horizontal and vertical structure of convective zones including intertropical convergence zone and extratropical cyclones in middle and upper troposphere, while weaker results were obtained for boundary layer shallow convection. They explained this issue could arise from the inherently deterministic nature of neural network and the localized learning employed in contrast to a certain degree of memory in the parent numerical prediction model used for training (Pritchard et al. 2011). As another example of using ML in parameterization of convection, O’Gorman and Dwyer (2018) utilized random forests (Breiman 2001) in an idealized simulation. They demonstrate that random forests are capable of reconstructing the precipitation extremes without the need for special training on extremes. Implementation of ML scheme in a GCM showed remarkable results in case of global warming when it was trained with the current and warmer climate samples. Results were surprisingly good even when the scheme only learned from the warmer climate (O’Gorman and Dwyer 2018).

ML has been identified as a powerful tool for constructing new parameterizations, and the studies described above constitute first steps in that direction. Such approaches however depend on the availability of high-resolution simulations in the framework of the model used. Contrary to parameterizations based on an explicitly formulated simplification of the physics of the model [e.g., the scheme of Alexander and Dunkerton (1999), which has become used in several models], parameterizations built by ML emulating high-resolution simulations will be difficult to transfer. Due to these difficulties and to the opacity of the resulting parameterizations, it is very likely that traditional parameterization schemes, based on explicitly specified relations and physical principles, will remain key components for many years to come. The challenge with these physically based parameterizations is to better understand the processes to be represented, and ML methods can contribute to this objective, without aiming at replacing or emulating existing parameterizations. ML methods can indeed be used to probe the relationship between the large-scale flow and subgrid-scale processes, providing elements of answers to fundamental questions: for a given large-scale flow, how much uncertainty is there on the subgrid-scale process considered? In other words, what part of the subgrid-scale process is *deterministic*, and what part is *stochastic*? Which explanatory variables are most relevant, and contribute most to the estimation of the subgrid-scale process? For this purpose, we exploit the advantages of random forest

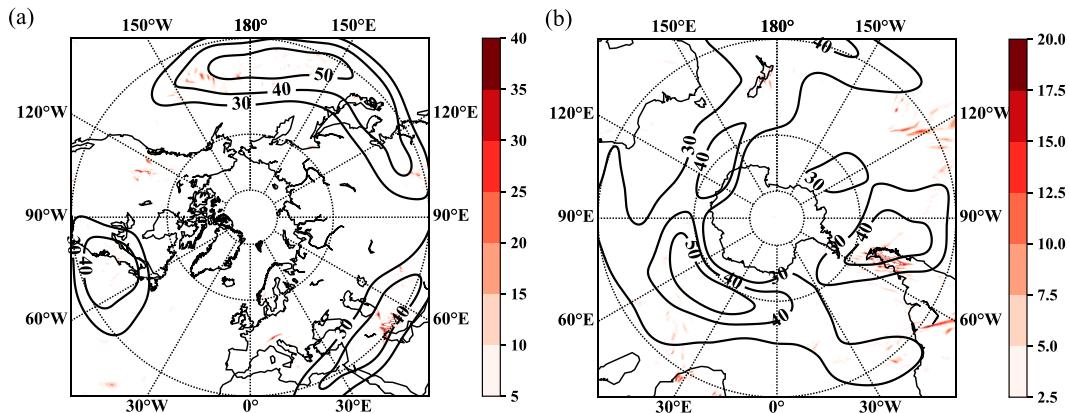


FIG. 1. Snapshots of absolute momentum flux (shaded; unit: mPa) and isotach (solid smoothed black line for the values greater than  $30 \text{ m s}^{-1}$ ) at 100 hPa for (a) the Northern Hemisphere at 1200 UTC 3 Jan 2019 and (b) the Southern Hemisphere at 1200 UTC 3 Aug 2019.

regressor to probe the relationship between the coarse-grained description of the flow in the troposphere and GW signals in the lowermost stratosphere. The main aim of present paper is to explore two insights for the development of GW parameterizations. First, to indicate how much of the GW signal can be reconstructed from the low-resolution resolvable flow. Second, to rank the variables used to train the model by order of relevance when the model presumes no relation between variables in advance and tries to learn from the data only.

The paper is organized as follows. The data used to feed the model and the methodology have been described in section 2. The model setup and information about GW signature and targets are discussed in section 3. The explanatory variables which carry the most information on GWs signatures are explored in section 4. A comparison between non-parametric and parametric model is presented in section 5, followed by the summary and perspectives in section 6.

## 2. Data and methodology

At least three recent studies have used machine learning to contribute to the investigation and improvement of GW parameterization. Two studies use ML to emulate an existing parameterization, and the other uses ML to capture the relation between the large-scale flow and the resulting GWs. These three studies are briefly reviewed below, with an emphasis on the explanatory variables that are retained to describe the large-scale flow. Espinosa et al. (2021) employed an artificial neural network (ANN) to emulate the process of GWs (Alexander and Dunkerton 1999) in an atmospheric model of intermediate complexity. They used a “parameterization” or an estimation of GWs effect on the large-scale flow due to lack of observations, computational limits, and controlling the degrees of freedom to avoid unwanted underfitting in the ANN. Five years of model output (1 year for training and 4 years for testing) including three components of wind, temperature, and surface pressure have been employed as input variables to estimate

zonal and meridional GW drag in their survey. Chantry et al. (2021) used three years of IFS model data to train their ANN. They considered the variables used to predict the velocity tendencies (horizontal components of wind and temperature profiles) and model level descriptors (the pressure, half-level pressure, and geopotential). As another example of the recent application of machine learning in GW parameterization, Matsuoka et al. (2020) implemented a deep convolutional neural network for evaluation of each component of the wind fluctuations ( $u'$ ,  $v'$ , and  $w'$ ) and then the calculation of gravity wave momentum fluxes at 100 hPa. In addition, they used ML to emulate an existing scheme. For this purpose, they trained their model using data from 29-yr reanalysis datasets of low-resolution zonal and meridional winds, temperature, and specific humidity at 300, 700, and 850 hPa over the Hokkaido region (Japan). Two-dimensional input arrays directly transferred to the three-dimensional wind fluctuations at the target level in the latter study, contrary to two other studies which considered columnar information to train the models.

Similar to the study of Matsuoka et al. (2020), we use output from a numerical weather prediction model to provide the target (the resolved GWs) and the coarse-grained description of the large-scale flow. In contrast to their study, we investigate nonorographic waves and use a global model. The dataset that has been used in the present survey is ERA5 from the European Centre for Medium-Range Weather Forecasts (ECMWF). The data are available hourly with a horizontal resolution of 31 km at 137 model levels. With such relatively high resolution, ERA5 is able to resolve and represent part of the spectrum of GWs reasonably well. In continuity with performance assessments of earlier analysis or reanalysis (Wu and Eckermann 2008; Preusse et al. 2014; Jewtoukoff et al. 2015), Amiramjadi et al. (2020) examined the realism of GW signature existing in ERA5. They showed that GWs in ERA5 data are in good spatiotemporal agreement with those observed in HIRDLS satellite in GRACILE dataset (Ern et al. 2018), but weaker than observations by a factor of about 3, which varies with time and location.

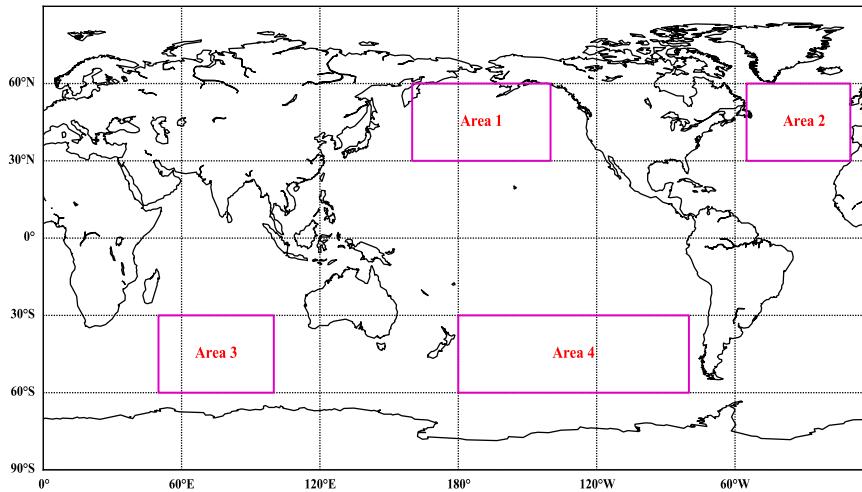


FIG. 2. Areas of study in the midlatitudes of the Northern and Southern Hemispheres over oceans.

From the ERA5 data, a lower-resolution dataset with  $2.5^\circ$  grid distance in both zonal and meridional directions is used to provide the explanatory variables. With the lower-resolution dataset, almost the whole spectrum of GWs remains unresolved. Additionally, a low-pass spectral filter with a cutoff zonal wavenumber of 22 has been applied to explanatory variables to remove possible signature of gravity waves (Sato et al. 2012, 2009; Amiramjadi et al. 2020). As explanatory variables we extract the temperature gradient, the fraction of cloud cover, the potential vorticity (PV) anomaly, the relative vorticity, vertical velocity and horizontal wind speed at 15 pressure levels from 1000 to 100 hPa. These variables have been chosen to represent GWs non-orographic sources (e.g., horizontal wind speed for jet stream, large-scale vertical velocity for convection, temperature gradient for fronts).

As target variables describing the GWs, we have chosen the local absolute momentum flux [ $m = \rho \sqrt{(u'w')^2 + (v'w')^2}$ ] and the standard deviation of two diagnostic variables of horizontal divergence ( $\sigma_d$ ) and vertical velocity ( $\sigma_w$ ), which are considered in the lower stratosphere at 100 hPa pressure level with high resolution ( $0.25^\circ \times 0.25^\circ$ ). These fluctuations are

obtained as the standard deviation of 900 grid squares (about  $750 \times 750 \text{ km}^2$ ) by a simple moving average from the high-resolution data (Fig. 1). To avoid the effect of mountain waves and focus on non-orographic GWs, the areas of study have been chosen over oceans far from the lands and significant topographic obstacles (Fig. 2). As the main configuration, one year from December 2018 to November 2019 in midlatitudes of the Northern and Southern Hemispheres has been chosen for testing in this study. For training, several different combinations of one or three years have been used comprising 2017, 2018, and 2020 (Table 1). These combinations provide past and future training datasets and varied geographical locations rather than testing data.

### 3. The best estimation of GWs signal from local knowledge of the large-scale flow

#### a. Model setup

The “random forest” algorithm (Breiman 2001; James et al. 2013), hereafter RF for reference, is employed as our non-parametric statistical model, for this study. It consists of an ensemble of decision trees, each making the best use of a

TABLE 1. Summary of different combinations for training and testing.

Combination	Training data	Testing data
E2019	One grid point in the main configuration (2019) using $K$ -fold method	The same grid point of training using $K$ -fold method
E3	One grid point in 2017, 2018, and 2020	The same grid point of training, but in 2019
A2017	All grid points in 2017 over the areas of study	Targets in 2019
A2018	All grid points in 2018 over the areas of study	Targets in 2019
A2020	All grid points in 2020 over the areas of study	Targets in 2019
A3	All grid points in 2017, 2018, and 2020 over the areas of study	Targets in 2019
B2	All grid points in 2017 to 2020 over area 2	Targets over areas 1 and 4
B3	All grid points in 2017 to 2020 over area 3	Targets over areas 1 and 4
B23	All grid points in 2017 to 2020 over areas 2 and 3	Targets over areas 1 and 4

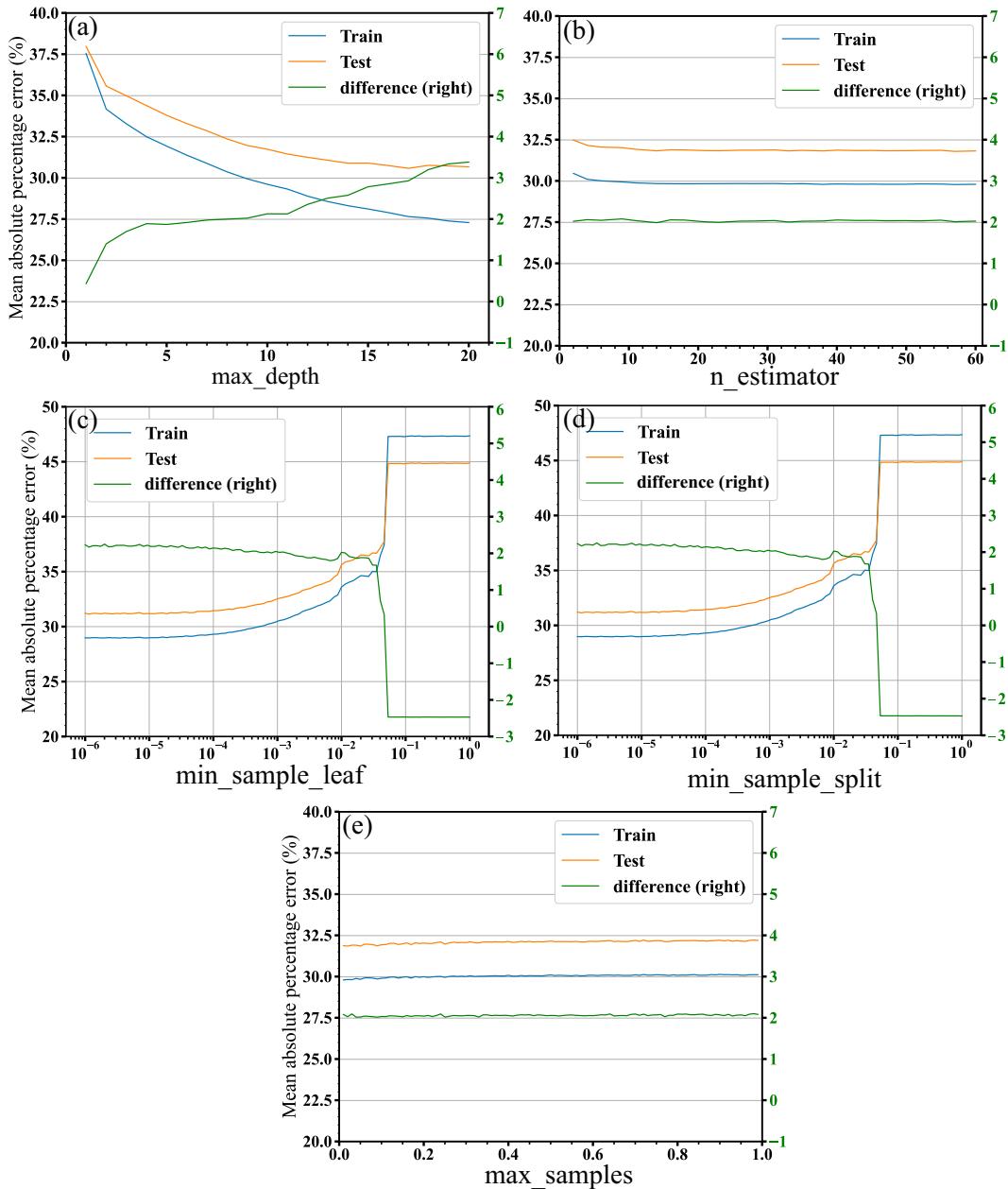


FIG. 3. Mean absolute percentage error (MAPE) depicted for vertical velocity as target. The results have been shown for (a) the tree depth (`max_depth`), (b) the number of trees in the forest (`n_estimators`), (c) the minimum number of samples that should be present in the leaf node after splitting a node (`min_sample_leaf`), (d) the minimum number of training targets in any given node in order to split it (`min_sample_split`), and (e) the fraction of the original dataset that is given to any individual trees (`max_sample`). The results for the MAPE values depicted for the test (orange line) and train (blue line) datasets and their differences (green line scaled on the right y axis).

random subsample of the input variables. The output is obtained as the average of individual outputs. Here, as the common practice in parameterization, the RF is considered as a column-based frame without any interaction with neighboring grid cells. To implement RF, use is made of the “random forest regressor” class from scikit-learn package version 0.23.1 in Python (Pedregosa et al. 2011). One of the attractive features

of this algorithm and this package is in providing information on the relative importance of explanatory variables. This provides precious insights into the relevance of a large number of variables for estimating gravity waves. Another encouraging practical advantage of RF is that unlike the other well-known artificial intelligence models (such as neural networks), the input variables do not need to be standardized.

TABLE 2. The 18 most relevant and 42nd and 46th explanatory variables ranked by “mean decrease impurity” method used for reconstruction of GW signal in the current study. Mentioned in parenthesis is the pressure level of Exp\_Var and “wind,” “grad\_x,” “grad\_y,” “grad\_t,” “W,” “Vo,” and “PV\_an,” respectively, stand for “horizontal wind speed,” “meridional,” “zonal,” and “absolute temperature gradient;” “vertical velocity,” “relative vorticity,” “anomaly of potential vorticity;” and “anomaly of potential vorticity.”

Target	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	16th	17th	18th	42nd	46th
m	wind(100)	wind(150)	Vo(100)	Vo(150)	W(250)	W(300)	W(200)	PV_an(100)	grad_x(250)	grad_y(100)	grad_t(1000)	PV_an(700)	PV_an(150)	grad_x(200)	PV_an(1000)	PV_an(200)	PV_an(400)	W(100)	grad_y(150)	PV_an(250)
w	wind(100)	wind(150)	W(200)	W(250)	Vo(200)	W(300)	grad_x(1000)	Vo(150)	PV_an(100)	grad_x(1000)	Vo(250)	grad_x(200)	PV_an(1000)	W(700)	grad_x(100)	W(600)	wind(750)	W(800)	grad_x(250)	grad_x(250)
d	wind(150)	W(300)	W(250)	grad_x(100)	W(600)	Vo(150)	grad_x(1000)	W(200)	Vo(200)	PV_an(150)	grad_x(200)	W(700)	grad_x(1000)	W(500)	wind(100)	grad_x(150)	W(400)	Vo(100)	W(850)	grad_x(400)

TABLE 3. As in Table 2, but using “mean decrease accuracy” in order to rank the explanatory variables.

Target	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	16th	17th	18th	42nd	46th
m	wind(100)	Vo(100)	Vo(150)	wind(150)	W(250)	W(200)	W(300)	grad_x(100)	PV_an(100)	grad_x(1000)	grad_x(250)	PV_an(700)	PV_an(150)	PV_an(1000)	grad_x(200)	grad_x(1000)	PV_an(400)	grad_x(150)	W(150)	W(100)
w	wind(100)	wind(150)	W(200)	W(250)	Vo(200)	Vo(150)	grad_x(1000)	W(300)	PV_an(100)	grad_x(1000)	PV_an(100)	grad_x(200)	grad_x(100)	Vo(250)	grad_x(150)	W(100)	wind(700)	PV_an(400)	W(700)	W(150)
d	wind(150)	grad_x(100)	W(250)	W(300)	Vo(200)	PV_an(150)	grad_x(1000)	grad_x(200)	W(200)	Vo(150)	wind(100)	grad_x(100)	PV_an(100)	grad_x(100)	Vo(100)	W(600)	W(700)	grad_x(150)	grad_x(150)	grad_x(100)

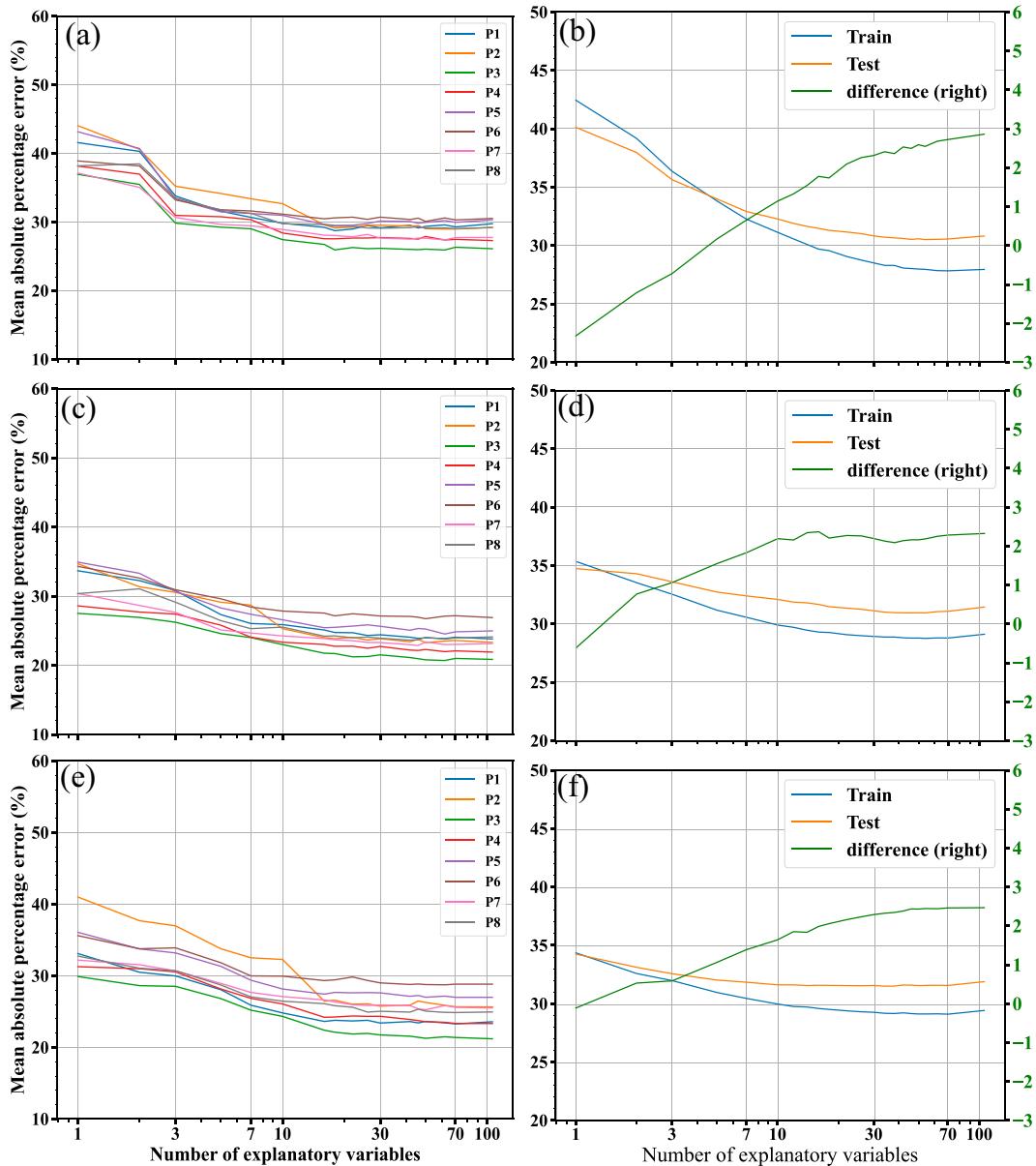


FIG. 4. MAPE depicted for different number of explanatory variables based on variable significance output of model for (a),(b) absolute momentum flux, (c),(d) standard deviation of horizontal divergence, and (e),(f) standard deviation of vertical velocity as targets in different runs. Results are plotted for (a),(c),(e) eight randomly chosen grid points and the test (orange line) and train (blue line) datasets and their differences [green line on the right y axis in (b), (d), and (f)].

While making stronger assumptions, parametric models employ a fixed number of parameters, are computationally faster and easier to implement in GCMs for their simpler structure. The performance depends on how accurately the parameters have been chosen. Linear regression is an example of a parametric statistical model. In contrast with parametric models, nonparametric models make fewer but more flexible assumptions about the data and consequently work more slowly. Here, we take advantage of Python to train the model only one time with the proper data and save it as a

preprepared library. At any time later, we can use such trained model to estimate the target in each call.

Here to set up our model, we need and use all data over the areas of study for four years (1378 grid points, including 7 explanatory variables over 15 pressure levels throughout the 1000 to 100 hPa layer per year, from December 2016 to November 2020) and  $K$ -fold cross validation, with  $K = 4$  to avoid the possible spatiotemporal correlation in data (without shuffling to split data into train/test sets). It implies that 75% of samples (6 035 640) is used to build up the trees and the model

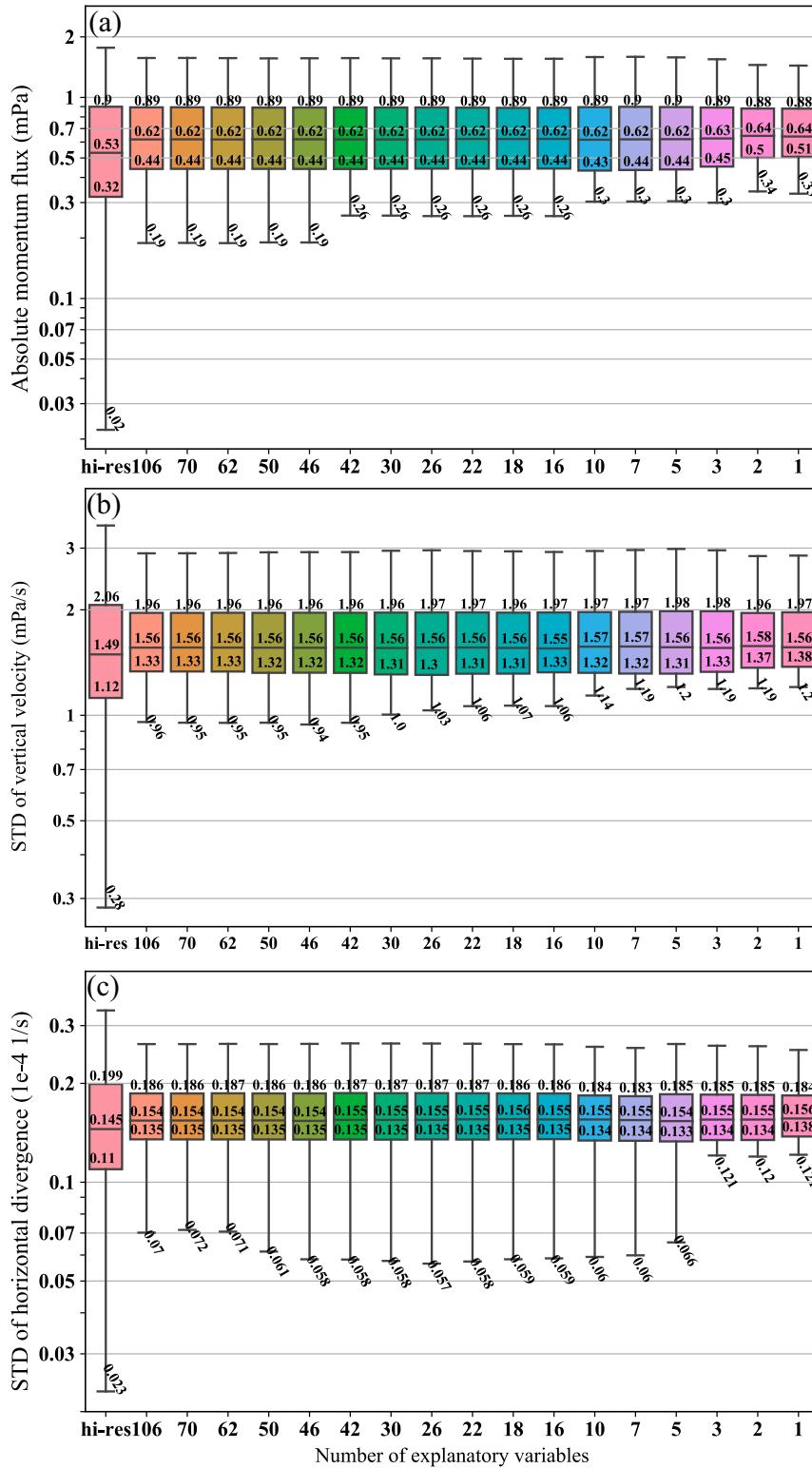


FIG. 5. Boxplots of (a) absolute momentum flux (mPa), (b) standard deviation of vertical velocity ( $\text{mPa s}^{-1}$ ), and (c) standard deviation of horizontal divergence ( $10^{-4} \text{ s}^{-1}$ ) based on high-resolution target data and reconstructed signal with different number of explanatory variables, mentioned at x axis, based on variable significance output of model for the combination A3. The first and third quartiles, median, and minimum are plotted on the boxes.

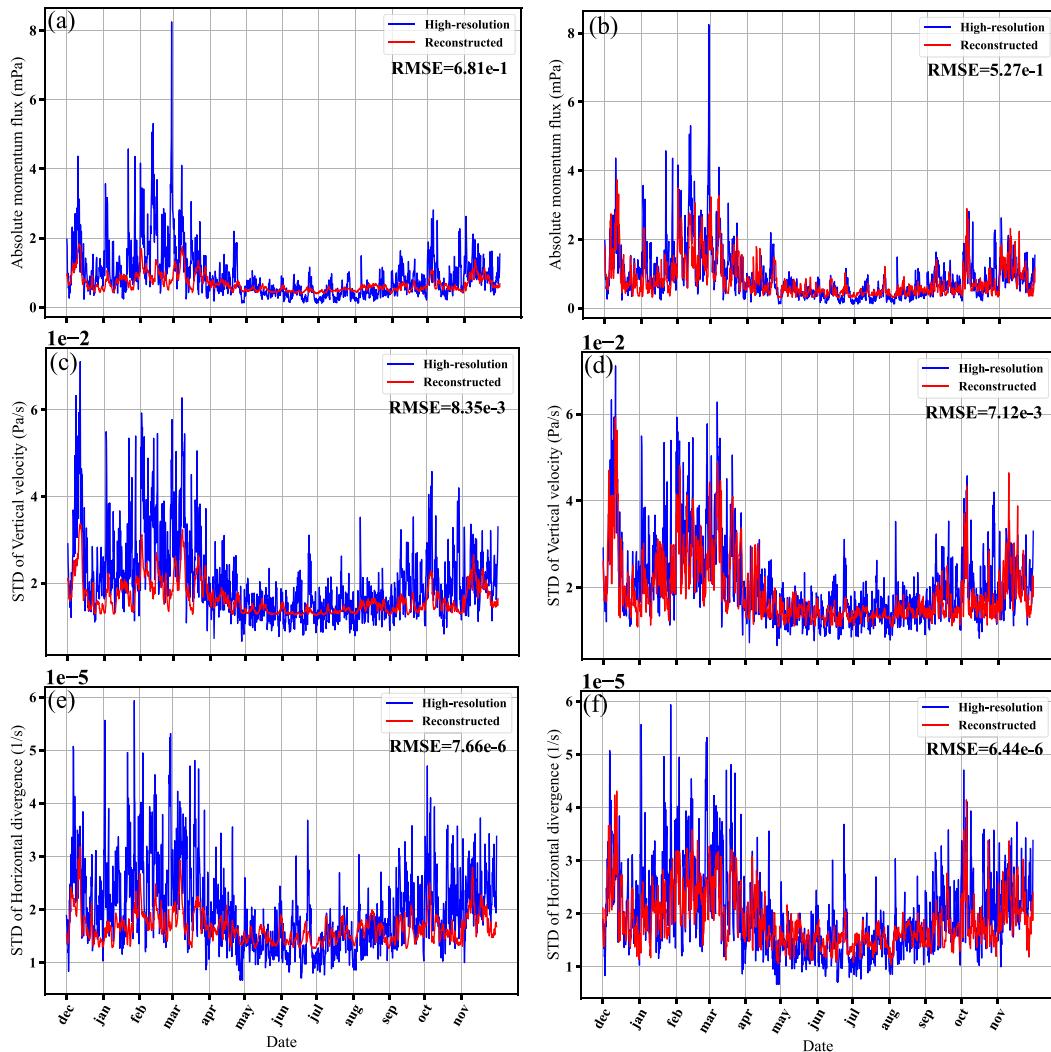


FIG. 6. Time series of (a),(b) absolute momentum flux (mPa), (c),(d) standard deviation of vertical velocity ( $\text{Pa s}^{-1}$ ), (e),(f) standard deviation of horizontal divergence ( $\text{s}^{-1}$ ) for the combination A3 and the gridpoint P3. Blue and red lines represent the high-resolution and reconstructed signals, respectively. (a),(c),(e) The signal reconstructed via Exp-Vars = 1, (b) Exp-Vars = 46, (d) Exp-Vars = 42, and (f) Exp-Vars = 26.

will be tested with the rest of data. When exploring one parameter, we keep the other parameters constant as their best values, knowing that these hyperparameters highly depend on the dataset. In other words, there generally is not a unique, optimal value for each parameter, but rather a range of values.

The parameter tuning results are almost the same for all three targets, so they are presented only for vertical velocity (Fig. 3).

Figure 3a shows the response of model to variation of tree depth. We test values for this parameter, controlled by scikit-learn’s “max depth,” ranging from 1 to 20 for shallow to deeper trees, respectively. It can be inferred that the model learns from the training dataset and predicts them perfectly but cannot generalize the learning to reconstruct the test data, because it overfits for large values of depth over training data.

Therefore, the trees with the depth value of 11 are chosen, where the test dataset reaches an acceptable accuracy before the model starts to overfit by following difference between test and train mean absolute percentage error (MAPE) curves (green curve in Fig. 3a).

A model with more trees can learn better but runs more slowly. Figure 3b shows the performance of the model for different numbers of trees. This parameter can be controlled by “n\_estimators” in scikit-learn package. MAPE reveals that the accuracy grows as the number of trees increases, but the accuracy levels off for the number of estimators greater than 14. The error approaches its minimum at number of trees equal to 35 and remains almost constant for the number of trees greater than 35 [ $\text{MAPE}(n=14) - \text{MAPE}(n=35) \approx 0.03$  and  $\text{MAPE}(n=60) - \text{MAPE}(n=35) \approx 0.03$ ]. It is found that the results are not very sensitive to the number of trees. As a

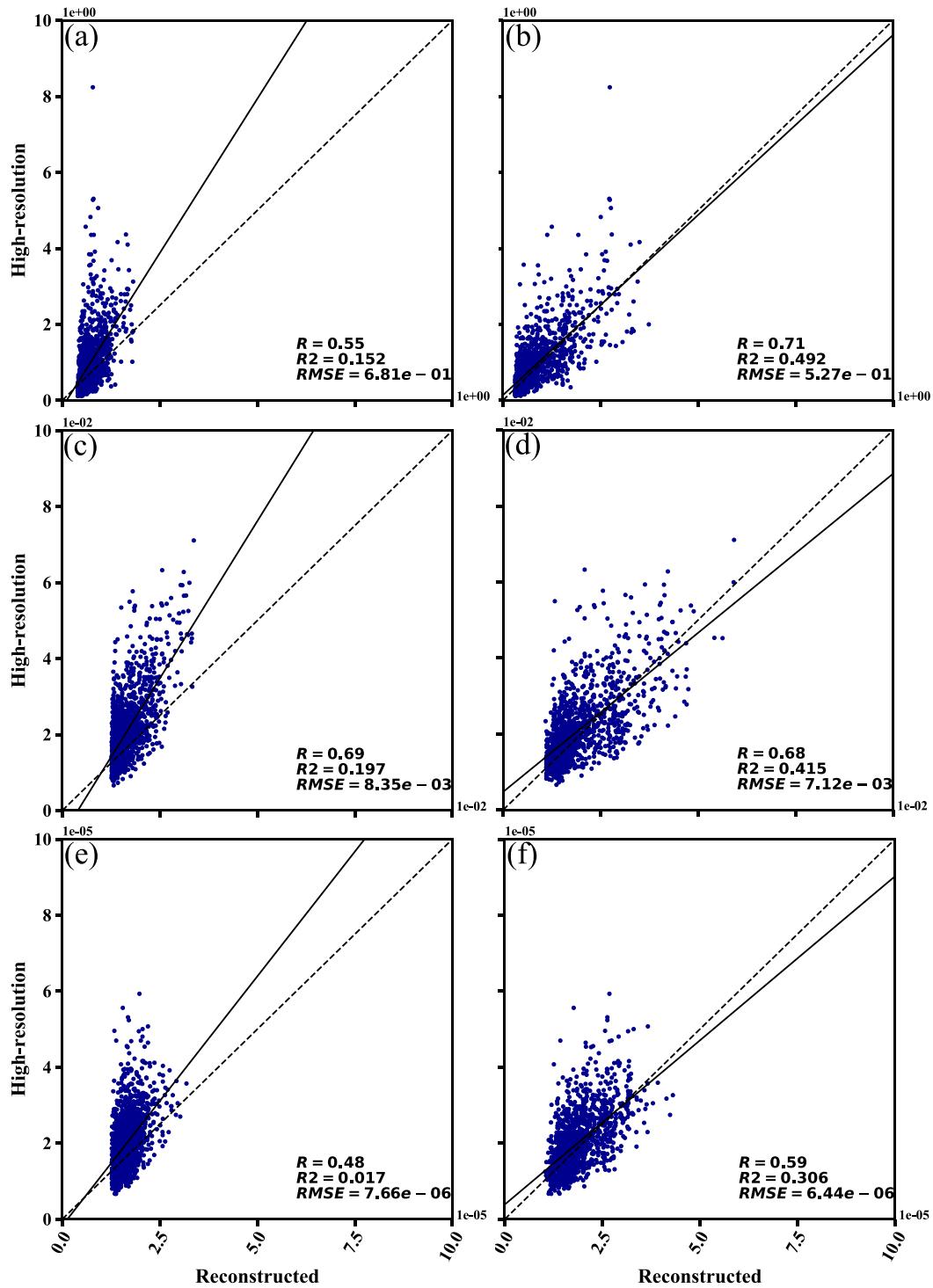


FIG. 7. Scatterplots of high-resolution vs reconstructed GW signals as presented in Fig. 6. Shown in the panels are correlation coefficient ( $R$ ), coefficient of determination ( $R^2$ ), and root-mean-square error (RMSE). The ground truth and regression lines are overlaid as dashed and solid lines, respectively.

compromise in the present study the number of trees retained was 35.

As the next parameter, we investigate the “minimum number of samples required to be at a leaf node,” the so-called `min_sample_leaf` in scikit-learn package, varying from 0.5 to  $5 \times 10^{-6}$  portion of samples. From Fig. 3c, we can clearly see that the error is great when the model stops splitting the nodes with the larger numbers of samples. The model learns better by letting it grow and have less samples at each terminal node. For trees having less than approximately 10 samples per terminal node ( $1.6 \times 10^{-6}$  portion of the training data), performance appears optimal. This is close to the default value of the model (1 sample).

Another parameter to explore is “the minimum number of samples” required to split an internal node. The parameter is varied from full to  $1 \times 10^{-6}$  portion of samples using “`min_samples_split`” in scikit-learn package. One can see that the model suffers from underfitting for larger numbers of samples, preventing the model to learn more (Fig. 3d). The MAPE decreases as the minimum number of samples need to split the nodes decreases. However, the model starts to saturate for the portion of samples smaller than  $2 \times 10^{-5}$  (approximately 120 samples). Any value below that (here, 113 samples) can be brought into play for this parameter.

We also tested fractions of bootstrapped data given to each decision tree from 1% to 100% (Fig. 3e) and can see that the results are quite insensitive to the “`max_samples`” parameter. Hence, we retain the default value for this parameter which is 100% of samples.

Then, the model has been run with a different number of explanatory variables (Exp-Vars) sorted by significance for each target (Table 2). The MAPE has been calculated and depicted for test and training sets of data as well as the eight randomly chosen grid points as will be described in section 4. The results obtained show that for all targets (Fig. 4), our statistical model learns better as the number of Exp-Vars increases. However, although the decrease in error is significant for the first five Exp-Vars, it then quickly saturates for number of Exp-Vars more than 40 for absolute momentum flux and vertical velocity, while the performance for horizontal divergence reaches its saturation point by 25 Exp-Vars. For more Exp-Vars, it is found that while additional Exp-Vars do not permit any improved performance, they lead to overfitting the model. It appears that a range of 25–50 Exp-Vars allows the best performances of the models for the three targets. Further analysis presented in section 3b leads us to retain 46, 42, and 26 Exp-Vars, respectively, for absolute momentum flux, vertical velocity, and horizontal divergence as targets.

*b. Different targets*

Here we explore the variability of different reconstructed targets compared to the high-resolution data for different numbers of Exp-Vars by plotting the boxplots for the combination A3 (Fig. 5). Again, one can see that the model learns better as the number of Exp-Vars increases. This improvement is largely due to better estimates of the interquartile range, while the minimum values still suffer from overestimation.

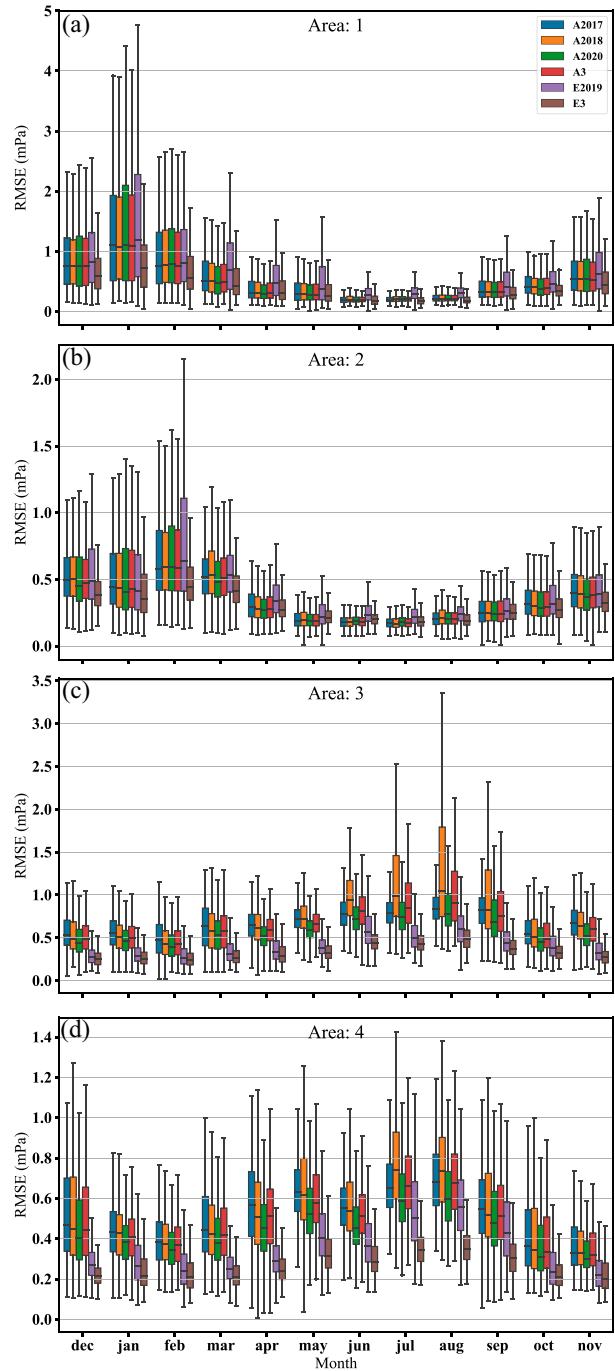


FIG. 8. Boxplots of root-mean-square error of absolute momentum flux (mPa) for December 2018–November 2019, over (a) area 1, (b) area 2, (c) area 3, and (d) area 4. Mentioned in the upper-right corner of (a) is the combination that the model is trained over (more details in Table 1).

However, the median attains its best estimation by  $\text{Exp-Vars} = 5$  ( $\text{Exp-Vars} = 16$  for vertical velocity), and the first quartile tries to find its best value, the maximum and the third quartile hardly vary for different numbers of Exp-Vars.

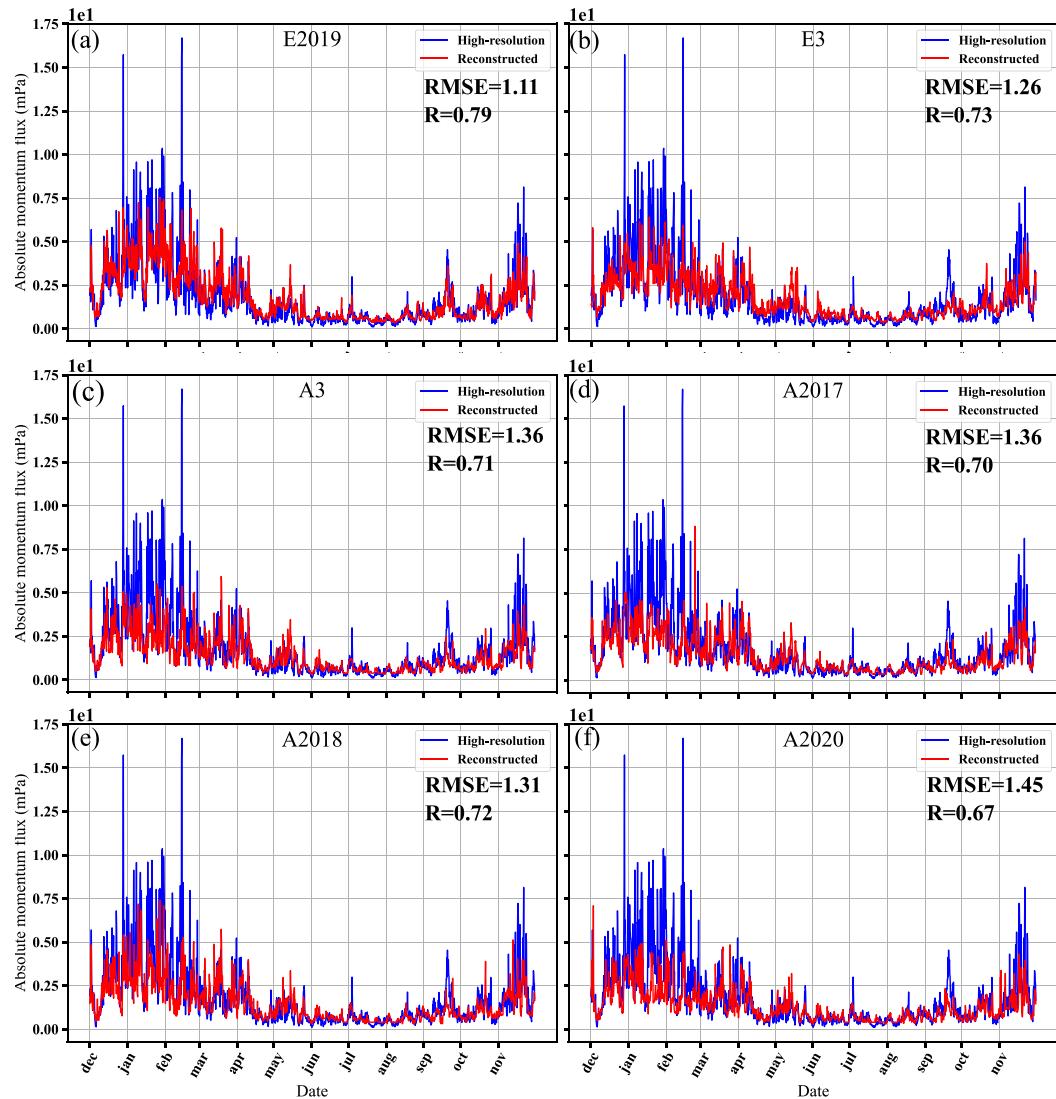


FIG. 9. Time series of absolute momentum flux (mPa) as the signal of gravity waves (blue line for the high-resolution and red line for the reconstructed data). The model has been trained through (a) E2019, (b) E3, (c) A3, (d) A2017, (e) A2018, and (f) A2020 and tested for P1.

The variability of reconstructed momentum flux barely changes from  $\text{Exp-Var} = 1$  to  $\text{Exp-Var} = 2$ , but comparing to Fig. 4 it can be inferred that using only one variable fails to provide sufficient information for a good reconstruction. The reconstruction improves notably from  $\text{Exp-Var} = 10$  to  $\text{Exp-Var} = 16$  and  $\text{Exp-Var} = 42$  to  $\text{Exp-Var} = 46$ , although the main elements of the boxplots do not change for more Exp-Vars (Figs. 4a and 5a). Though they cannot obtain their ideal values close to the elements of high-resolution data (specifically first quartile and minimum) even using all Exp-Vars, we choose  $\text{Exp-Var} = 46$  for our experiment considering the best overlapping with the high-resolution data variations focusing on third quartile, maximum, and upper outliers (not shown) as high gravity waves activity. Similar arguments hold for vertical velocity. One may consider the first 42 Exp-Vars in

Table 2 for reconstruction of GWs signature in the vertical velocity as the target which can represent relatively larger-scale GWs. On the other hand, for the reconstructed horizontal divergence, better results cannot be achieved by increasing the number of explanatory variables beyond  $\text{Exp-Var} = 26$ . Thus, this number of explanatory variables seems to be adequate for efficiency in reconstruction of GW signals in horizontal divergence. In addition, we can see a notable improvement in reconstruction of minimum values from  $\text{Exp-Var} = 3$  to  $\text{Exp-Var} = 5$ , while it fails to reconstruct minimum values even with all Exp-Vars.

Figures 6 and 7 show time series and scatterplots of reconstructed GW signals for P3 (Fig. 14 shows locations of eight randomly chosen grid points and will be discussed in section 4). One may notice the improvements from  $\text{Exp-Vars} = 1$  to the other numbers of Exp-Vars (different values for different

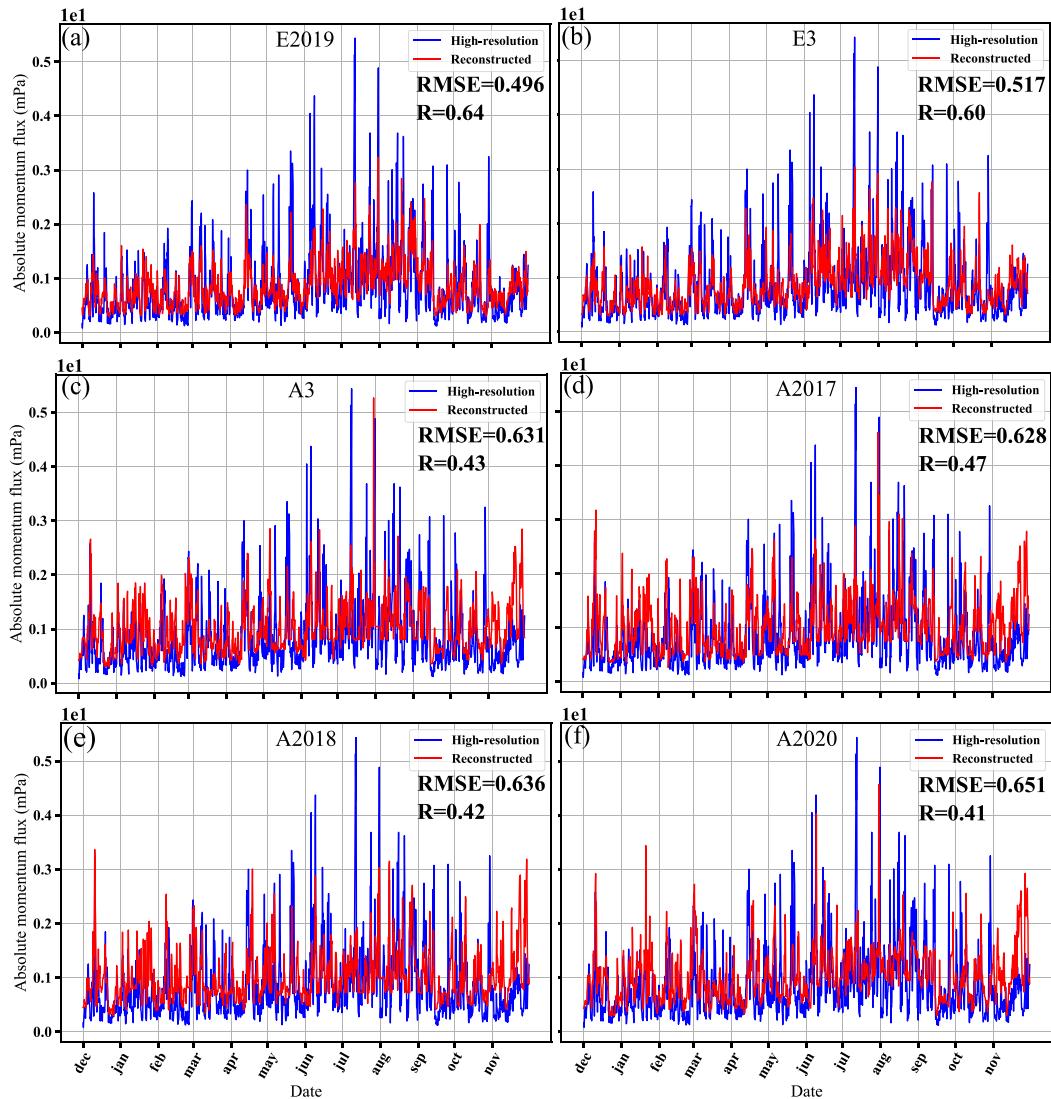


FIG. 10. As in Fig. 9, but for P6.

targets). It can be seen from the time series and statistical measures presented on the scatterplots that the model can capture the existence of gravity waves activity, not the variability nor the extremes, using only one explanatory variable, that is the upper-level wind speed. The model learns more and can capture the variabilities better when the number of Exp-Vars increases. We focus on absolute momentum flux for the rest of this study since it is the most interesting diagnostic quantity in gravity waves studies (Alexander et al. 2010).

*c. How universally the RF could reconstruct the GW signal*

To explore the performance of the model in the reconstruction of a specific sample using different training samples, we considered several different combinations of one or three years of data available in 2017, 2018, and 2020 (Table 1). The results are represented as the root-mean-square error (RMSE)

of absolute momentum flux in Fig. 8 separately plotted over four areas of study for the months of the year 2019. It is discernible that the error is smaller in the months during which the wave activity is small, notably in the Northern Hemisphere. In the Northern Hemisphere, performance of the model is nearly the same when the model learns from the past using “A2017” and “A2018,” from future using “A2020” and even a collection of them using “A3” attaining correlation coefficients of  $R \approx 0.65$  to 0.70. Training the model over one grid point and testing it over the same point via fourfold cross validation method (combination “E2019” in the Table 1) results in larger errors compared to other combinations, which is not surprising. For “E2019” combination, we exclude one-fourth of data at each grid point to test and use the rest to train the model. Unlike the other regression models, random forest regression model cannot extrapolate outside the range of data and then fails to estimate unseen data. When faced such unseen high (low) GWs activity in winter

(summer), the model assumes that the prediction could be close to the maximum (minimum) values of the training set in other seasons. Obviously, the average of ensemble predictions will not fall outside the highest and lowest values in the sample and leads to a notable error. Similarly, one may consider 3 years of data on one grid point for training and one year data for testing (combination “E3”). The error is significantly less than (or in some cases in the same range of) other combinations so that we even can achieve correlation coefficients of  $R \approx 0.93$  and  $0.88$  over areas 1 and 2, respectively. In this case, the model learns from the data which are similar to testing data in terms of timing, intensity, and loss of GWs activity. Weaker, but not inexact, results for the other combinations are likely related to different percentage participation of input components (section 4).

Additionally, to illustrate both the strong variability of the GW signal and the ability of the model to reproduce a significant part of this variability, the time series (Figs. 9 and 10) confirm that the model can identify high gravity wave activities but cannot fully reconstruct the intermittencies and extremes (e.g., wintertime over P1). From Figs. 8c and 8d we see the results for the Southern Hemisphere. Although the average errors are less than the values in the Northern Hemisphere, the errors are significantly high for the combination “A2018” in the wintertime. Except for the “E2019” and “E3,” which represent similar results as in the Northern Hemisphere ( $R \approx 0.60$  and  $0.88$ , respectively, for “E2019” and “E3”), there is no reasonable difference between other combinations in training the model ( $R \leq 0.50$ ). Therefore, one may simply conclude that increasing the training dataset by including more and varied geographical locations fails to be efficient everywhere.

We investigated this idea by training the model over the Atlantic and Indian Oceans (areas 2 and 3 with the same number of grid points, 247 points) and testing it over the northern and southern Pacific Ocean (areas 1 and 4). Figure 11 shows the RMSE of reconstructed momentum flux over areas 1 and 4 for 12 months (from December 2018 to November 2019). Relatively high RMSE in wintertime compared to the summertime of both hemispheres again indicates that the model can identify high gravity wave activities but cannot reconstruct the full variability. Furthermore, training and testing the model over the same hemisphere may lead to better results, specifically over the Southern Hemisphere. This result may suggest that the mechanism of nonorographic gravity waves generation and/or their propagation into the background flow slightly differs in the midlatitudes of the Northern and Southern Hemispheres.

#### 4. Which explanatory variables carry the most information?

As mentioned in section 2, several outputs from ERA5 data are used to train the model at tropospheric pressure levels from 1000 to 100 hPa in which the main GWs sources are located. In section 3b we explored the minimum number of explanatory variables that have the highest impact to optimize the model’s performance. In this section, we analyze the variables we utilized to train the model. Here, we use the advantage of random forest to rank the features fed to the model by

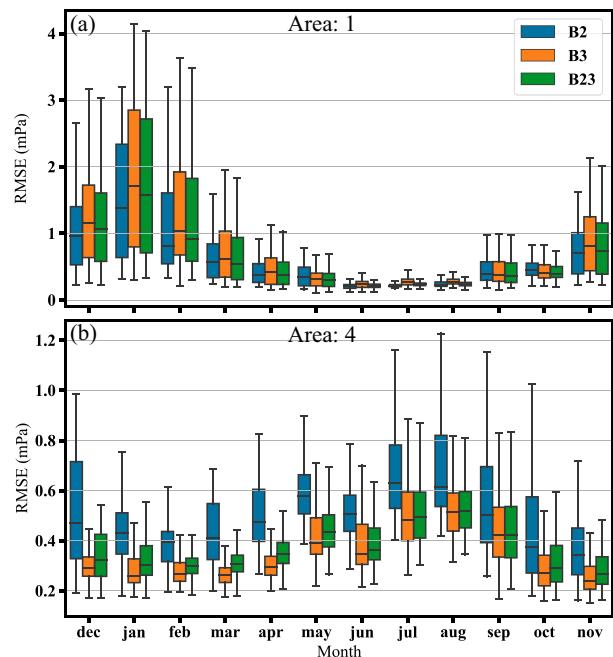


FIG. 11. As in Fig. 8, but over (a) area 1 and (b) area 4. Mentioned in the upper-right corner of (a) is the combination/area that the model is trained over (more details in Table 1).

employing all variables available for all grid points on four areas at all pressure levels in four years (from 2017 to 2020). The random forest algorithm in scikit-learn package provides information on the relevance of explanatory variables using two methods when it builds decision trees. The variables can be ranked based on the so-called mean decrease impurity method (MDI, for brevity). The model employs the “Gini index” as an impurity function. In this case, the ranking, which is the average indexes over all of decision trees in the forest, can be interpreted as “Gini importance.” As an alternative method, one may consider the “mean decrease accuracy” (MDA) which provides “permutation importance” scores in evaluating the relevance of input variables. The biggest advantage of MDI compared to MDA is in the speed of computation during the random forest training as MDA is computationally expensive (e.g., in this experiment 4–5 CPU times slower than MDI). However, more accurate results can be obtained using MDA and one should note that retraining the model is not needed at each modification of the dataset.

Figure 12 shows the ranking of explanatory variables and highlights the importance of the pressure levels as well as Exp-Vars. A first notable results is the relevance of horizontal wind speed at the target level as the most relevant Exp-Vars, which is confirmed by both MDA and MDI. This agrees with our knowledge of the impact of mean flow on the existence and propagation of GWs in the lower stratosphere, e.g., 150 and 100 hPa (Plougonven et al. 2017). The low-resolution vertical velocity is the most informative explanatory variable in the middle to upper troposphere (200–400 hPa). And finally, the anomaly of potential vorticity conveys the most information in the lower to middle troposphere (400–700 hPa) on reconstruction

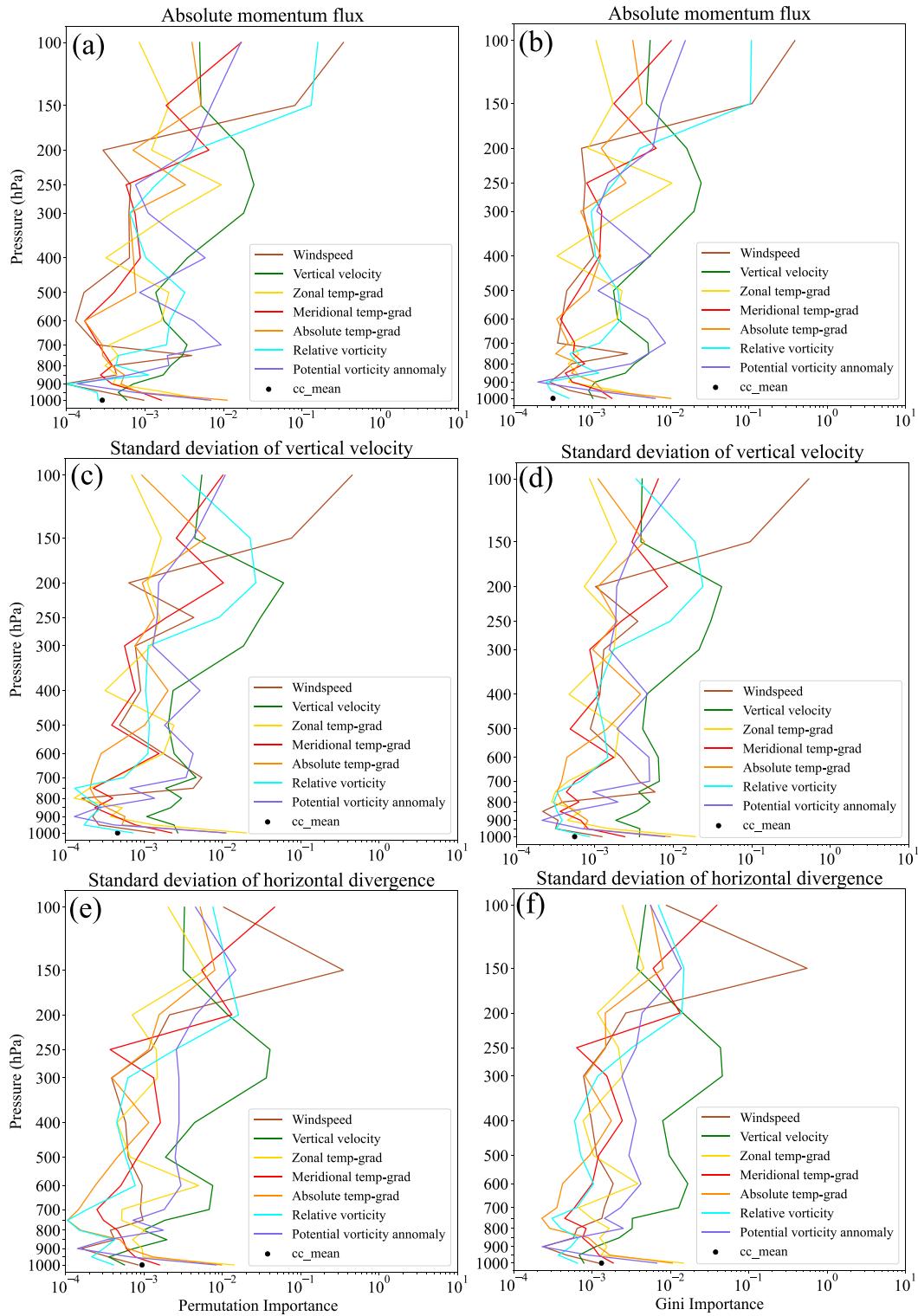


FIG. 12. Variable importance of explanatory variables in decision trees including wind speed (brown), vertical velocity (green), relative vorticity (aqua), meridional temperature gradient (red), zonal temperature gradient (yellow), absolute temperature gradient (orange), potential vorticity anomaly (purple), and vertical average of the fraction of cloud cover (black point represented at 1000 hPa). Shown are variables importance based on pressure levels using (a),(c),(e) mean decrease accuracy and (b),(d),(f) mean decrease impurity methods for different targets: (a),(b) absolute momentum flux, (c),(d) standard deviation of vertical velocity, and (e),(f) standard deviation of horizontal divergence.

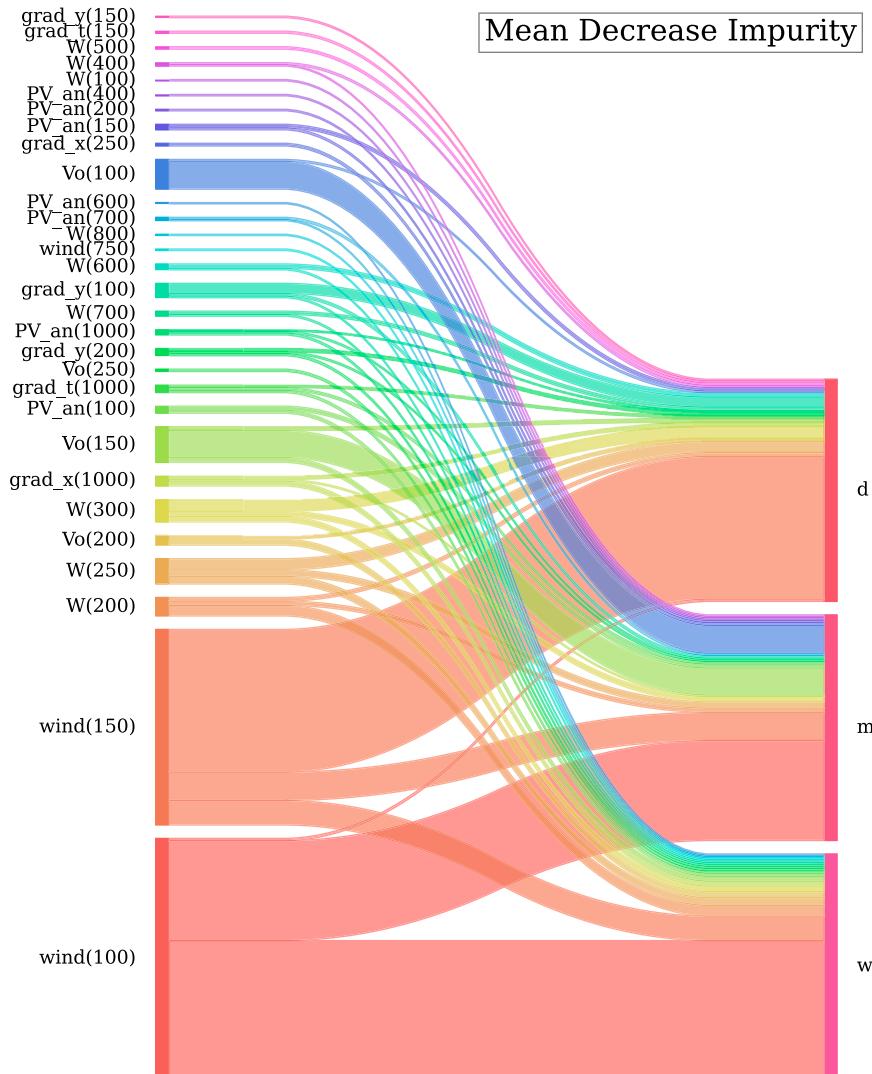


FIG. 13. Sankey diagram for the three targets showing the top 20 Exp-Vars sorted based on their importance by mean decrease impurity methods (see also Table 2).

of GWs at the target level. The significance of PV anomaly in these levels could be reminiscent of the vorticity anomalies in nonorographic GW parameterizations proposed by de la Cámara and Lott (2015) following the earlier theoretical works on GW generation by sheared PV (Lott et al. 2010, 2012). There are also some evidences of upper-level and surface fronts in some explanatory variables, but almost with less Gini scores comparing to other Exp-Vars (Fig. 13). Similar results and diagram can be obtained using the MDA method (Table 3). The results may slightly differ in different runs for the variables with approximately the same importance score. This is the outcome of the “random” inner nature of RF in building up the decision trees.

These results are almost the same for all three targets and both methods. Considering the momentum flux, we can find the importance of relative vorticity in the lower stratosphere in the top four ranks. Upper-tropospheric relative vorticity

has been placed in next ranks, then horizontal wind speed and low-resolution vertical velocity in reconstruction of high-resolution vertical velocity. As it is found that horizontal wind speed at the level of the target is a prominent source of information on the GW, the monthly mean horizontal wind speed at 100 hPa has been depicted in the summertime of both the Northern and Southern Hemispheres where the gravity waves activity is relatively high (Fig. 14). The location of the eight points P1 to P8 in Fig. 14 can explain dissimilar gravity wave activity for these points, consistent with their positions relative to the jet stream. P1 which displays the strongest values in GW momentum flux (see, e.g., Figs. 8 and 9) is located exactly in the exit region of the jet stream while P4 usually experiences rather calm conditions. The wind speed is approximately weaker in the Southern Hemisphere compared with wintertime of the Northern Hemisphere. Although P6 is located close to the jet streak, the wave activity is somewhat less in contrast

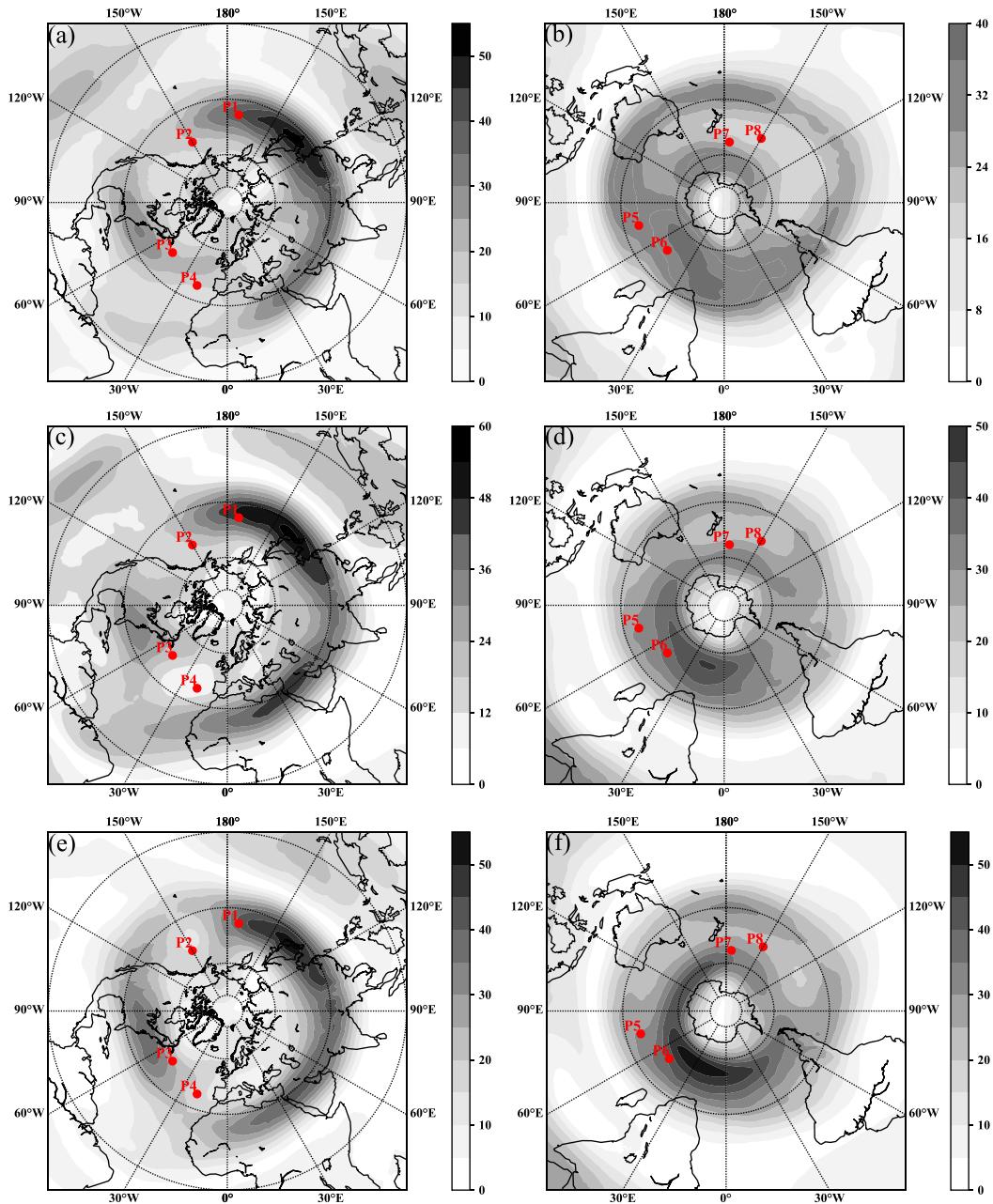


FIG. 14. Snapshots of monthly mean horizontal wind speed (shaded;  $m s^{-1}$ ) over (a),(c),(e) the Northern Hemisphere and (b),(d),(f) the Southern Hemisphere for (a) December 2018, (b) June 2019, (c) January 2019, (d) July 2019, (e) February 2019, and (f) August 2019. Location of eight randomly chosen grid points has been depicted with red dots.

with similar points in the Northern Hemisphere (Fig. 8). This may explain the notable overestimation at P6.

**5. Toward linear regression in reconstruction of gravity waves signals**

Nonparametric models such as random forest have the disadvantage that it is difficult to apply the obtained model

because of its more complicated structure. Parameterization schemes are classically based on explicit, analytic formulas, including heuristic or ad hoc ingredients. Here we explore the performance of a parametric model, namely, linear regression (or more precisely multilinear regression, hereafter LR for reference), to reconstruct the GW signal. To this end, first, we need to explore the existence of a linear relationship between target and Exp-Vars in advance. To have the same conditions

for both models, we surveyed the linearity of the Exp-Vars ranked by order of relevance as output of RF. It revealed that there may be a linear relationship between absolute momentum flux and wind speed, absolute temperature gradient, and PV anomaly with positive slopes (not shown). Because of the lack of linearity between the target and other Exp-Vars, we continued this experience with only three Exp-Vars, including wind speed, absolute temperature gradient, and PV anomaly on different pressure levels. Desired data are ranked regarding the relative importance using RF as presented in Table 4 (up to 20 variables). Again, we can see the relevance of wind speed at the level of the target as the most important Exp-Var in the reconstruction of the GW signal. The comparison between the performance of RF and LR has been carried out using the most relevant 5 and 15 variables (Table 4) for both statistical models using “A3,” as described in Table 1.

Input variables for the LR need additional data preparation. For better performance, it would be better to push the data closer to a normal distribution. One possible way for this purpose is to apply a transformation, a mathematical function, to change the distribution. The distributions with positive skewness can be transformed using either a logarithmic or a square root function. We employed the logarithmic function on momentum flux since it often has a lognormal distribution (Hertzog et al. 2012), while the application of the square root function leads to better results for the Exp-Vars. In addition to the above, we standardized Exp-Vars to vary in the [0, 1] range. Such data preparations did not significantly affect the results of the random forest regressor.

Figure 15 shows the time series for implementation of LR and RF in the reconstruction of the absolute momentum flux. The LR can identify low to medium GWs activity considering the less complexity (5 Exp-Vars) and computational costs (about 5–7 times faster than RF), while the RF model exhibits better performance in the case of high wave activity. Increasing the number of Exp-Vars from 5 to 15 in most cases degrades the performance of LR (Table 5), characteristically in the low GWs activity, while it increases the complexity of this model. Meanwhile, the performance of RF model improves with the same increase in the number of Exp-Vars.

### 6. Summary and perspective

Machine learning as a branch of artificial intelligence is being increasingly used into many aspects of atmospheric science. The model provided by ML can learn directly from the huge amount of data that are becoming available. This makes ML methods attractive in studying the complex systems for which the relation between various components is either unclear or unknown. This interesting feature of ML has motivated us to employ ML for investigating the particularly complicated problem of relation between nonorographic GWs and large-scale flow. To this aim, we used the ERA5 reanalysis dataset from the ECMWF since it can resolve and represent part of the spectrum of GWs nearly well with its relatively high resolution. A lower-resolution dataset derived from the full-resolution ERA5, through the troposphere to the lower stratosphere (from 1000 to 100 hPa), has been

TABLE 4. As in Table 2, but for “horizontal wind speed,” “absolute temperature gradient,” and “anomaly of potential vorticity.”

Target	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	16th	17th	18th	19th	20th
m	wind(100)	wind(150)	grad_ t(1000)	grad_ t(100)	PV_ an(400)	PV_ an(100)	PV_ an(150)	PV_ an(500)	PV_ an(1000)	PV_ an(700)	wind(1000)	grad_ t(150)	grad_ t(250)	grad_ t(300)	PV_ an(200)	wind(400)	grad_ t(200)	grad_ t(700)	grad_ t(400)	PV_ an(800)

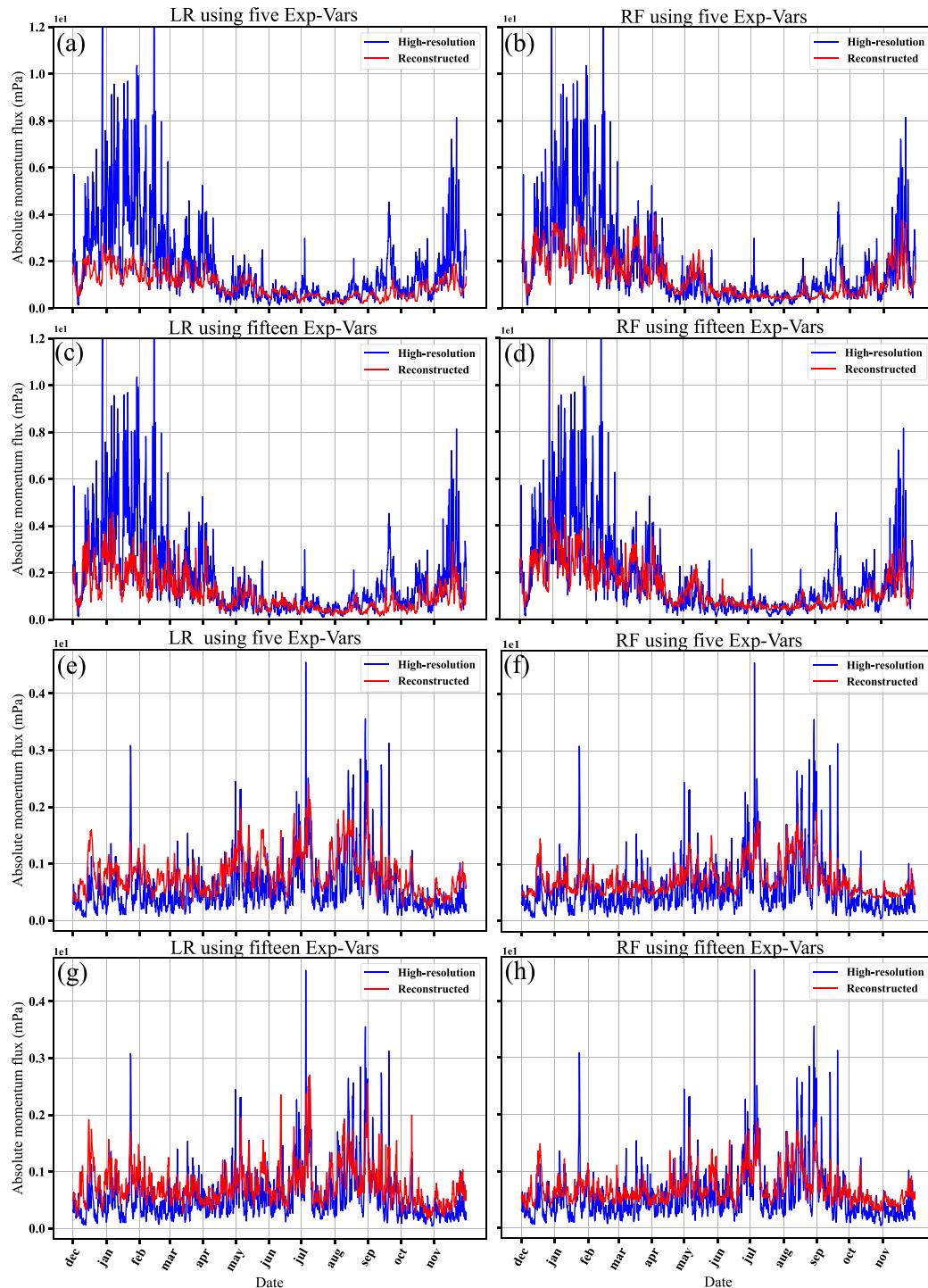


FIG. 15. Time series of high-resolution (blue line) and reconstructed (red line) absolute momentum flux (mPa) using (left) the LR model and (right) the RF model over (a)–(d) P1 and (e)–(h) P7. The signal is reconstructed using (a),(b),(e),(f) 5 Exp-Vars, and (c),(d),(g),(h) 15 Exp-Vars.

employed for explanatory variables in which almost the whole spectrum of GWs remains unresolved. The ERA5 dataset in full resolution has been chosen to extract the local momentum flux, and the standard deviation of two diagnostic variables of

horizontal divergence and vertical velocity at 100 hPa pressure level, as targets describing the GWs fluctuations. The full dataset including four years of data over the areas of study has been employed to select appropriately the model hyperparameters.

TABLE 5. Statistical measures ( $R$ : correlation coefficient;  $R^2$ : coefficient of determination; RMSE: root-mean-square error) for comparing the performance of the linear regression (LR) model and the random forest model (RF) in reconstruction of absolute momentum flux as the signal of GWs (with the unit of “mPa”) using 5 and 15 explanatory variables (Table 4) at 8 randomly chosen grid points in addition to all points (labeled as P1 to P8 and “all,” respectively) for 12 months (December 2018–November 2019).

Points	LR						RF					
	5			15			5			15		
	$R$	$R^2$	RMSE									
P1	0.62	0.074	1.75	0.62	0.192	1.63	0.62	0.285	1.54	0.64	0.326	1.49
P2	0.52	0.23	0.74	0.48	0.17	0.77	0.47	0.11	0.80	0.55	0.19	0.76
P3	0.56	0.24	0.65	0.53	0.20	0.66	0.55	0.29	0.63	0.58	0.31	0.62
P4	0.33	−0.16	0.42	0.35	−0.34	0.41	0.35	0.11	0.37	0.35	0.12	0.37
P5	0.30	−0.16	0.61	0.35	−0.30	0.65	0.31	−0.01	0.57	0.33	0.05	0.55
P6	0.41	0.15	0.59	0.40	0.01	0.64	0.40	−0.01	0.65	0.38	0.09	0.61
P7	0.62	0.05	0.45	0.59	0.14	0.43	0.64	0.30	0.39	0.63	0.31	0.39
P8	0.46	−0.55	0.44	0.45	0.09	0.36	0.49	0.13	0.35	0.45	−0.67	0.49
All	0.49	0.18	0.66	0.51	0.21	0.65	0.54	0.29	0.61	0.60	0.37	0.60

The tree depth was found to be an important parameter with significant impact on the results, while little sensitivity was found to other parameters such as the number of trees, and a range of parameter values was acceptable without affecting the best performance of the model.

The results of this experience with ML showed the positive and relative success of “random forest regressor” as a nonparametric procedure in predicting the variations of gravity wave related quantities. Among the explanatory variables tested, it turned out that the highest information is carried by the upper-level wind speed. This is consistent with the previous findings on the role of the upper-level jet as a main source of GWs and on the importance of lateral propagation into regions of strong winds. While the model can capture the existence of GWs activity using merely one explanatory variable, it fails on variability and the extremes of activity. In general, the timing of variations and peaks is reproduced rather accurately as indicated by correlation coefficients of above 0.65, but the peaks are often underestimated, such that coefficients of determination hardly reach 0.45. Lack of informative indicators for convective activity in the coarse-grained explanatory variables is the potential reason of this underestimation. By increasing the number of explanatory variables, the model can learn better and capture more variability. A good number of explanatory variables has been shown by tests to be between 25 and 50 variables. In the reconstructed signals, correlation coefficients above 0.65 are reached (0.93 in the case that we trained the model using 3 years of data at one grid point and tested at the same grid in another year and 0.7 in the case that the model learned from all grid points of 3 years, which is encouraging given the sharply variable target). Although the model skillfully captures much of the strong, short variations of the wave activity, it has failed to predict the variations of targets when the GW activity is weak. It should be noticed that the RF model cannot discover trends, as the model learns only from the training data and there is no pre-defined relation of explanatory variables and target. It means that the model cannot extrapolate the values falling outside the training dataset.

A comparison has also been conducted between the performance of “random forest regressor” as a nonparametric model and linear regression (LR) as a simple parametric model. Here, the LR model demonstrates a modest performance in the application of machine learning for GW signal reconstruction, considering the lesser complexity, lower computational cost as well as accuracy and intermittency of the reconstructed signal. On the other hand, the performance of the RF model elucidates its positive perspective in GWs studies where the relation of explanatory variables with the targets is almost nonlinear and uncertain.

As a perspective, it will be interesting to explore the same procedure but with the GW signature obtained from observation instead of simulation data as the target. The results of such a study would be useful for determining the relation of GWs properties to the sources and could also be compared with the estimates for GWs from the current parameterization schemes. In such a setup, it would also be possible to employ the estimates for GWs from parameterizations as explanatory variables for the machine learning model trained to reconstruct observed GWs: the informative value of the output of the parameterization scheme could be revealing.

*Acknowledgments.* Authors wish to thank Drs. Albert Hertzog and Alvaro de la Cámara for helpful comments and interesting discussions. MA is indebted to École Polytechnique de Paris for providing support for the visit and the work environment during the year-long stay for this study. MA, ARM, and MM thank the University of Tehran for providing support during this study. RP is supported by ANR Project BOOST3R ANR-17-CE01-0016, and by the Strateole-2 project, with support from Centre National d’Etudes Spatiales (CNES). The studies of MA 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).

*Data availability statement.* The ERA5 data have been retrieved from <https://doi.org/10.24381/cds.bd0915c6>.

## REFERENCES

- Alexander, M. J., and T. J. Dunkerton, 1999: A spectral parameterization of mean-flow forcing due to breaking gravity waves. *J. Atmos. Sci.*, **56**, 4167–4182, [https://doi.org/10.1175/1520-0469\(1999\)056<4167:ASPOMF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056<4167:ASPOMF>2.0.CO;2).
- , J. R. Holton, and D. R. Durran, 1995: The gravity wave response above deep convection in a squall line simulation. *J. Atmos. Sci.*, **52**, 2212–2226, [https://doi.org/10.1175/1520-0469\(1995\)052<2212:TGW RAD>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<2212:TGW RAD>2.0.CO;2).
- , and Coauthors, 2010: Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models. *Quart. J. Roy. Meteor. Soc.*, **136**, 1103–1124, <https://doi.org/10.1002/qj.637>.
- Amiramjadi, M., A. R. Mohebalhojeh, M. Mirzaei, C. Zülicke, and R. Plougonven, 2020: The spatiotemporal variability of nonorographic gravity wave energy and relation to its source functions. *Mon. Wea. Rev.*, **148**, 4837–4857, <https://doi.org/10.1175/MWR-D-20-0195.1>.
- Andrews, D. G., J. R. Holton, and C. B. Leovy, 1987: *Middle Atmosphere Dynamics*. International Geophysics Series, Vol. 40, Academic Press, 489 pp.
- Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, <https://doi.org/10.1023/A:1010933404324>.
- Butchart, N., and Coauthors, 2010: Chemistry–climate model simulations of twenty-first century stratospheric climate and circulation changes. *J. Climate*, **23**, 5349–5374, <https://doi.org/10.1175/2010JCLI3404.1>.
- Chantry, M., S. Hatfield, P. Dueben, I. Polichtchouk, and T. Palmer, 2021: Machine learning emulation of gravity wave drag in numerical weather forecasting. *J. Adv. Model. Earth Syst.*, **13**, e2021MS002477, <https://doi.org/10.1029/2021MS002477>.
- de la Cámara, A., and F. Lott, 2015: A parameterization of gravity waves emitted by fronts and jets. *Geophys. Res. Lett.*, **42**, 2071–2078, <https://doi.org/10.1002/2015GL063298>.
- , —, V. Jewtoukoff, R. Plougonven, and A. Hertzog, 2016: On the gravity wave forcing during the southern stratospheric final warming in LMDZ. *J. Atmos. Sci.*, **73**, 3213–3226, <https://doi.org/10.1175/JAS-D-15-0377.1>.
- Ern, M., Q. T. Trinh, P. Preusse, J. C. Gille, M. G. Mlynarczyk, J. M. Russell III, and M. Riese, 2018: GRACILE: A comprehensive climatology of atmospheric gravity wave parameters based on satellite limb soundings. *Earth Syst. Sci. Data*, **10**, 857–892, <https://doi.org/10.5194/essd-10-857-2018>.
- Espinosa, Z., A. Sheshadri, G. Cain, E. Gerber, and K. Dalla-Santa, 2021: Machine learning emulation of parameterized gravity wave momentum fluxes in an atmospheric global climate model. *EGU General Assembly*, Online, EGU, EGU21-1398.
- Fovell, R., D. Durran, and J. Holton, 1992: Numerical simulations of convectively generated stratospheric gravity waves. *J. Atmos. Sci.*, **49**, 1427–1442, [https://doi.org/10.1175/1520-0469\(1992\)049<1427:NSOCGS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1992)049<1427:NSOCGS>2.0.CO;2).
- Fritts, D. C., and M. J. Alexander, 2003: Gravity wave dynamics and effects in the middle atmosphere. *Rev. Geophys.*, **41**, 1003, <https://doi.org/10.1029/2001RG000106>.
- García, R. R., A. K. Smith, D. E. Kinnison, Á. de la Cámara, and D. J. Murphy, 2017: Modification of the gravity wave parameterization in the Whole Atmosphere Community Climate Model: Motivation and results. *J. Atmos. Sci.*, **74**, 275–291, <https://doi.org/10.1175/JAS-D-16-0104.1>.
- Gentine, P., M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis, 2018: Could machine learning break the convection parameterization deadlock? *Geophys. Res. Lett.*, **45**, 5742–5751, <https://doi.org/10.1029/2018GL078202>.
- Hertzog, A., M. J. Alexander, and R. Plougonven, 2012: On the intermittency of gravity wave momentum flux in the stratosphere. *J. Atmos. Sci.*, **69**, 3433–3448, <https://doi.org/10.1175/JAS-D-12-09.1>.
- Hindley, N. P., C. J. Wright, N. D. Smith, L. Hoffmann, L. A. Holt, M. J. Alexander, T. Moffat-Griffin, and N. J. Mitchell, 2019: Gravity waves in the winter stratosphere over the Southern Ocean: High-resolution satellite observations and 3-D spectral analysis. *Atmos. Chem. Phys.*, **19**, 15377–15414, <https://doi.org/10.5194/acp-19-15377-2019>.
- Hourdin, F., and Coauthors, 2017: The art and science of climate model tuning. *Bull. Amer. Meteor. Soc.*, **98**, 589–602, <https://doi.org/10.1175/BAMS-D-15-00135.1>.
- James, G., D. Witten, T. Hastie, and R. Tibshirani, 2013: *An Introduction to Statistical Learning*. Vol. 112. Springer, 616 pp.
- Jewtoukoff, V., A. Hertzog, R. Plougonven, A. Cámara, and F. Lott, 2015: Comparison of gravity waves in the Southern Hemisphere derived from balloon observations and the ECMWF analyses. *J. Atmos. Sci.*, **72**, 3449–3468, <https://doi.org/10.1175/JAS-D-14-0324.1>.
- Kim, Y.-J., S. D. Eckermann, and H.-Y. Chun, 2003: An overview of the past, present and future of gravity-wave drag parameterization for numerical climate and weather prediction models. *Atmos.–Ocean*, **41**, 65–98, <https://doi.org/10.3137/ao.410105>.
- Lane, T. P., M. J. Reeder, and T. L. Clark, 2001: Numerical modeling of gravity wave generation by deep tropical convection. *J. Atmos. Sci.*, **58**, 1249–1274, [https://doi.org/10.1175/1520-0469\(2001\)058<1249:NMOGWG>2.0.CO;2](https://doi.org/10.1175/1520-0469(2001)058<1249:NMOGWG>2.0.CO;2).
- Lott, F., R. Plougonven, and J. Vanneste, 2010: Gravity waves generated by sheared potential vorticity anomalies. *J. Atmos. Sci.*, **67**, 157–170, <https://doi.org/10.1175/2009JAS3134.1>.
- , —, and —, 2012: Gravity waves generated by sheared three-dimensional potential vorticity anomalies. *J. Atmos. Sci.*, **69**, 2134–2151, <https://doi.org/10.1175/JAS-D-11-0296.1>.
- Matsuoka, D., S. Watanabe, K. Sato, S. Kawazoe, W. Yu, and S. Easterbrook, 2020: Application of deep learning to estimate atmospheric gravity wave parameters in reanalysis data sets. *Geophys. Res. Lett.*, **47**, e2020GL089436, <https://doi.org/10.1029/2020GL089436>.
- McLandress, C., T. G. Shepherd, S. Polavarapu, and S. R. Beagley, 2012: Is missing orographic gravity wave drag near 60°S the cause of the stratospheric zonal wind biases in chemistry–climate models? *J. Atmos. Sci.*, **69**, 802–818, <https://doi.org/10.1175/JAS-D-11-0159.1>.
- O’Gorman, P. A., and J. G. Dwyer, 2018: Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *J. Adv. Model. Earth Syst.*, **10**, 2548–2563, <https://doi.org/10.1029/2018MS001351>.
- O’Sullivan, D., and T. J. Dunkerton, 1995: Generation of inertia–gravity waves in a simulated life cycle of baroclinic instability. *J. Atmos. Sci.*, **52**, 3695–3716, [https://doi.org/10.1175/1520-0469\(1995\)052<3695:GOI WIA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<3695:GOI WIA>2.0.CO;2).
- Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830, <https://jmlr.org/papers/v12/pedregosa11a.html>.
- Plougonven, R., and F. Zhang, 2014: Internal gravity waves from atmospheric jets and fronts. *Rev. Geophys.*, **52**, 33–76, <https://doi.org/10.1002/2012RG000419>.

- , V. Jewtoukoff, A. Cámara, F. Lott, and A. Hertzog, 2017: On the relation between gravity waves and wind speed in the lower stratosphere over the Southern Ocean. *J. Atmos. Sci.*, **74**, 1075–1093, <https://doi.org/10.1175/JAS-D-16-0096.1>.
- , A. de la Cámara, A. Hertzog, and F. Lott, 2020: How does knowledge of atmospheric gravity waves guide their parameterizations? *Quart. J. Roy. Meteor. Soc.*, **146**, 1529–1543, <https://doi.org/10.1002/qj.3732>.
- Preusse, P., M. Ern, P. Bechtold, S. Eckermann, S. Kalisch, Q. Trinh, and M. Riese, 2014: Characteristics of gravity waves resolved by ECMWF. *Atmos. Chem. Phys.*, **14**, 10 483–10 508, <https://doi.org/10.5194/acp-14-10483-2014>.
- Pritchard, M. S., M. W. Moncrieff, and R. C. Somerville, 2011: Orographic propagating precipitation systems over the United States in a global climate model with embedded explicit convection. *J. Atmos. Sci.*, **68**, 1821–1840, <https://doi.org/10.1175/2011JAS3699.1>.
- Sato, K., S. Watanabe, Y. Kawatani, Y. Tomikawa, K. Miyazaki, and M. Takahashi, 2009: On the origins of mesospheric gravity waves. *Geophys. Res. Lett.*, **36**, L19801, <https://doi.org/10.1029/2009GL039908>.
- , S. Tateno, S. Watanabe, and Y. Kawatani, 2012: Gravity wave characteristics in the Southern Hemisphere revealed by a high-resolution middle-atmosphere general circulation model. *J. Atmos. Sci.*, **69**, 1378–1396, <https://doi.org/10.1175/JAS-D-11-0101.1>.
- Schneider, T., S. Lan, A. Stuart, and J. Teixeira, 2017: Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophys. Res. Lett.*, **44**, 12–396, <https://doi.org/10.1002/2017GL076101>.
- Stephan, C. C., C. Strube, D. Klocke, M. Ern, L. Hoffmann, P. Preusse, and H. Schmidt, 2019: Intercomparison of gravity waves in global convection-permitting models. *J. Atmos. Sci.*, **76**, 2739–2759, <https://doi.org/10.1175/JAS-D-19-0040.1>.
- Uccellini, L. W., and S. E. Koch, 1987: The synoptic setting and possible energy sources for mesoscale wave disturbances. *Mon. Wea. Rev.*, **115**, 721–729, [https://doi.org/10.1175/1520-0493\(1987\)115<0721:TSSAPE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<0721:TSSAPE>2.0.CO;2).
- Waite, M. L., and C. Snyder, 2009: The mesoscale kinetic energy spectrum of a baroclinic life cycle. *J. Atmos. Sci.*, **66**, 883–901, <https://doi.org/10.1175/2008JAS2829.1>.
- Wu, D. L., and S. D. Eckermann, 2008: Global gravity wave variances from *Aura* MLS: Characteristics and interpretation. *J. Atmos. Sci.*, **65**, 3695–3718, <https://doi.org/10.1175/2008JAS2489.1>.