Received: 11 November 2024 | Revised: 14 April 2025 | Accepted: 20 May 2025

DOI: 10.1002/qj.5037

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

A climatology of trapped lee waves over Britain and Ireland obtained using deep learning on high-resolution model output

Jonathan Coney[1] | **Andrew N. Ross**[1] | **Leif Denby**[2] | **He Wang**[3] | **Simon Vosper**[4] | **Annelize van Niekerk**[5] | **Tom Dunstan**[4]

[1]Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, UK

[2]Danish Meteorological Institute, Copenhagen, Denmark

[3]Department of Computer Science, UCL Centre for Artificial Intelligence, London, UK

[4]Met Office, Exeter, UK

[5]European Centre for Medium-Range Weather Forecasts, Reading, UK

Correspondence
Jonathan Coney, National Oceanography Centre, Joseph Proudman Building, 6 Brownlow Street, Liverpool, L3 5DA, UK.
Email: jonathan.coney@noc.ac.uk

Present address
Jonathan Coney, National Oceanography Centre, Liverpool, UK

Funding information
Met Office; Natural Environment Research Council, Grant/Award Number: NE/S007458/1

ABSTRACT

This article presents a climatology of trapped lee waves over Britain and Ireland obtained through deep learning. Several deep-learning models trained to diagnose lee-wave occurrence, amplitude, wavelength, and orientation are applied to a 31-year high-resolution hindcast dataset covering 1982–2012, from UK Climate Projections (UKCP18) data, driven by ERA-Interim reanalysis data. Building on previous work to examine lee-wave characteristics over Britain and Ireland, this study applies a new technique to a much larger dataset than has been used in the past. There is little diurnal variability observed in the occurrence and characteristics of lee waves. Spatially, most lee waves occur over hilly regions, such as the Scottish Highlands, the Lake District and the Pennines in England, and North Wales. Seasonally, lee waves occur more in the winter months than in the summer. The link between synoptic weather patterns and lee waves is quantified, with more lee waves produced and a higher likelihood of higher amplitude waves under patterns with faster synoptic wind speeds, such as the positive phase of the North Atlantic Oscillation (NAO+). The mean orientation of waves is broadly in line with the synoptic wind direction, though with a large spread in some cases. High horizontal wind speeds aloft are a necessary but not sufficient indicator of high-amplitude lee waves. When other meteorological variables are used to predict the prevalence of lee waves using a random forest, the Scorer parameter is the most important for predicting the generation of lee waves alongside horizontal wind speed: there is less importance placed on the stability. This climatology provides a novel data-driven insight into the formation and propagation of lee waves over Britain and Ireland.

KEYWORDS

machine learning, mountain waves, trapped lee waves

1 | INTRODUCTION

Internal gravity waves (or mountain waves) are produced by the forced ascent of air over orography in a stratified atmosphere. Under the right conditions, these waves can be trapped in the lower troposphere rather than propagating up into the stratosphere or mesosphere, forming trapped lee waves (or lee waves for short). These lee waves occur over Britain and Ireland on a regular basis, propagating downwind from orography, despite the orography being relatively small in height and width compared with other mountain ranges globally. Lee waves with a high amplitude (vertical velocity $> 3\ \mathrm{m \cdot s^{-1}}$ is used at the Met Office for forecasting severe lee-wave activity: e.g., Vosper *et al.*, 2013) lead to an increased likelihood of rotor activity, that is, regions of strong turbulence downstream of the orography. Rotors are particularly hazardous to aircraft and high-sided road vehicles (Vosper *et al.*, 2013). In addition, lee waves are a source of horizontal momentum transport in the atmosphere, as well as wave drag (Bretherton, 1969; Shutts, 1992).

To diagnose conditions conducive for waves and identify whether wave trapping is likely, the Scorer parameter (Scorer, 1949) is often used. The Scorer parameter l at a height z is defined as

$$l^2(z) = \frac{N^2}{U^2} - \frac{1}{U}\frac{d^2U}{dz^2},$$

where $N(z)$ is the Brunt–Väisälä frequency and $U(z)$ is the horizontal wind speed. In general, the first term is the more important and the second (shear) term is only significant when the wind shear is large (Blockley & Lyons, 1994). Two-dimensional linear waves can propagate vertically if their horizontal wavenumber, k (set by the scale of the orography), is less than l. If the Scorer parameter decreases with height, then waves can become trapped in the layer where $k < l$ (Durran, 2003). If $k < l$ everywhere, then waves cannot be generated at all. Generation of trapped lee waves therefore relies on strengthening horizontal winds and/or decreasing stability with height. Vosper (2004) extended the approach to the effects of inversions on lee waves. In reality, lee waves are complex three-dimensional and nonlinear phenomena, which cannot be explained fully by simple linear theory.

The prevalence and characteristics of lee waves over Britain have been investigated before. Worthington (2006) explored the possibility of a diurnal cycle of lee waves over Britain using Very High Frequency (VHF) radar measurements from a site near Aberystwyth in Wales and satellite imagery over a region covering Wales, Ireland, and much of the Midlands and Northern England from 1990 to 2006. They found no evidence of a diurnal cycle in orientation or amplitude of waves, but there was a seasonal cycle in amplitude of waves over their area of observations. However, there may be a diurnal cycle of waves in other parts of the world: Ruff and Ólafsson (2019) attribute a small diurnal cycle in downslope windstorms in Iceland to the changing prevalence of gravity waves throughout the day.

Vosper *et al.* (2013) presented results from a three-year climatology of trapped lee waves, using model output from the (then operational at the Met Office) 3D Velocities Over Mountains (3DVOM) model. They found that waves were more likely to occur in the winter months than in the summer months, and large-amplitude waves (those more likely to be related to strong turbulence and rotors, with typical vertical velocity amplitudes $> 3\ \mathrm{m \cdot s^{-1}}$) occur more frequently in the Scottish Highlands than in North Wales or the Pennines. Since the 3DVOM model was retired, trapped lee-wave forecasting at the Met Office relies on the high-resolution numerical weather prediction (NWP) model output of the operational UKV model configuration, part of the Met Office Unified Model (MetUM: Tang *et al.*, 2013). In general, UKV predicts lee waves in good agreement with observations from aircraft campaigns, and better than the dry linear model 3DVOM did (Sheridan *et al.*, 2017). However, these cited improvements in NWP models have not been used to produce climatological information about lee waves to improve and expand on existing work about lee waves (which was produced using older datasets or models). Producing an updated climatology using a large dataset from a more sophisticated model enables investigation of lee-wave occurrence and characteristics in the current climate. In addition, by using future climate scenarios, any projected changes to lee waves can be analysed: for example, due to changing weather patterns.

The development and progress of machine-learning techniques in the past decade means that automated analysis of large datasets (such as NWP model output) can be performed efficiently. Machine learning is the process of training a model to perform a given task on some data (such as extracting features from data and making predictions based on these features) and assessing the model's performance on some different (unseen) data (Reichstein *et al.*, 2019). One example of this is deep learning, such as the use of deep convolutional neural networks, which extract patterns from large quantities of data through learning feature representation of data via convolutions and nonlinear transforms (Gu *et al.*, 2018). Convolutional neural networks are able to extract spatial patterns from two-dimensional data such as images (or NWP output on a pressure surface). Random forests are another example of a machine-learning technique, an ensemble of decision trees used to characterise data based on features

within the data (Breiman, 2001). Decision trees are a data classification and/or regression method where some input data are classified based on a sequence of rules: the data are split based on rules at "decision nodes", with the output of a tree (its class or value) produced at a "leaf node" (Kotsiantis, 2013). Random forests train multiple decision trees on a random subset of the training data, and the output of each decision tree is aggregated in the output of the random forest (typically by averaging each tree's probabilistic output, as in Pedregosa *et al.*, 2012).

Machine-learning models are increasingly being used for forecasting and investigating climatological trends, as well as in a wide range of other environmental applications. For example, Böhm *et al.* (2021) produced a climatology of fog occurrence over the Atacama desert using a neural network trained on satellite brightness temperatures. Weather fronts can be identified using machine-learning techniques, and their climatology and impact on precipitation events can be detected and evaluated (e.g., Justin *et al.*, 2023; Niebler *et al.*, 2022). Entire NWP models are being replicated by machine-learning models, with promising results (e.g., Lam *et al.*, 2023), but there is some concern over the ability of neural-network-based weather forecasting models to forecast extreme events well compared with traditional NWP models (Charlton-Perez *et al.*, 2024).

In previous work, the authors have trained machine-learning models to detect and characterise trapped lee waves from high-resolution NWP data (described in detail in Coney *et al.*, 2024). In that work, deep-learning models were trained to predict the location, wavelength, orientation, and amplitude of trapped lee waves over Britain and Ireland, from two-dimensional slices of UKV vertical velocity data on the 700-hPa surface, using the Python library `fastai` (Howard & Gugger, 2020). The model for segmenting (classifying each pixel in the input as either containing a lee wave or not) the input vertical velocity slice performed well against hand-labelled truth data, and the models trained to predict wave characteristics performed favourably against a spectral technique, the two-dimensional Stockwell transform (S-transform: e.g., Stockwell *et al.*, 1996; Hindley *et al.*, 2016), while also delivering a significant speed increase compared with the S-transform.

This article details the development of a climatology of trapped lee waves over Britain and Ireland, created by applying deep-learning models developed by Coney *et al.* (2024) to 31 years of high-resolution regional climate model output using UK Climate Projections (UKCP18) local data. The features in the climatology include the frequency of occurrence, wavelength, amplitude, and orientation of trapped lee waves, covering a much longer period of time and using a more sophisticated model than the work of Vosper *et al.* (2013). This article also investigates variations in lee-wave location, frequency, and characteristics, as a function of time of day, season, and weather regime. Finally, random forests are used to investigate relationships between lee waves and 40 other regional climate model variables in order to evaluate the important physical processes controlling the climatology. This methodology permits a statistical approach to examining the prevalence and characteristics of trapped lee waves over Britain and Ireland, and comparisons with expectations from linear theory.

Previous work (such as that by Worthington, 2006 and Vosper *et al.*, 2013) did not consider different synoptic weather conditions over Britain and Ireland, for example, the weather patterns identified by Neal *et al.* (2016). These weather patterns group similar synoptic weather conditions into regimes. They have been used to explore the relationship between synoptic weather and, for example, observed lightning activity (Wilkinson & Neal, 2021) and future flood risk under climate change (Perks *et al.*, 2023). An overview of the methods and the data used are given in Section 2, the results are presented and discussed in Section 3, and some concluding remarks are made in Section 4.

2 | METHODS

2.1 | Data

2.1.1 | Meteorological data

The high-resolution model data for the current climate were sourced from the UKCP18 Local data suite: 2.2-km resolution limited-area hindcasts using the MetUM (version 10.6) for the period 1982–2012 (Lowe *et al.*, 2018; Manning *et al.*, 2023). ERA-Interim reanalysis data (Dee *et al.*, 2011) were downscaled by a regional climate model (RCM) configuration of the MetUM covering Europe, and then this was used to drive a convection-permitting configuration of the MetUM over Britain and Ireland to produce the data used here (UKCP18 Local). The convection-permitting model has a horizontal resolution of 2.2 km and 70 vertical levels (Manning *et al.*, 2023). The setup is similar to the 1.5-km resolution UKV configuration used operationally by the Met Office. Using the UKCP18 Local data instead of the archive of operational UKV data meant that available data existed for a longer time period (31 years versus 8 years at the time of writing) and that the data were all produced with the same MetUM version. Both these mean that the UKCP18 data are more suitable for developing a climatology of lee waves than operational UKV model output.

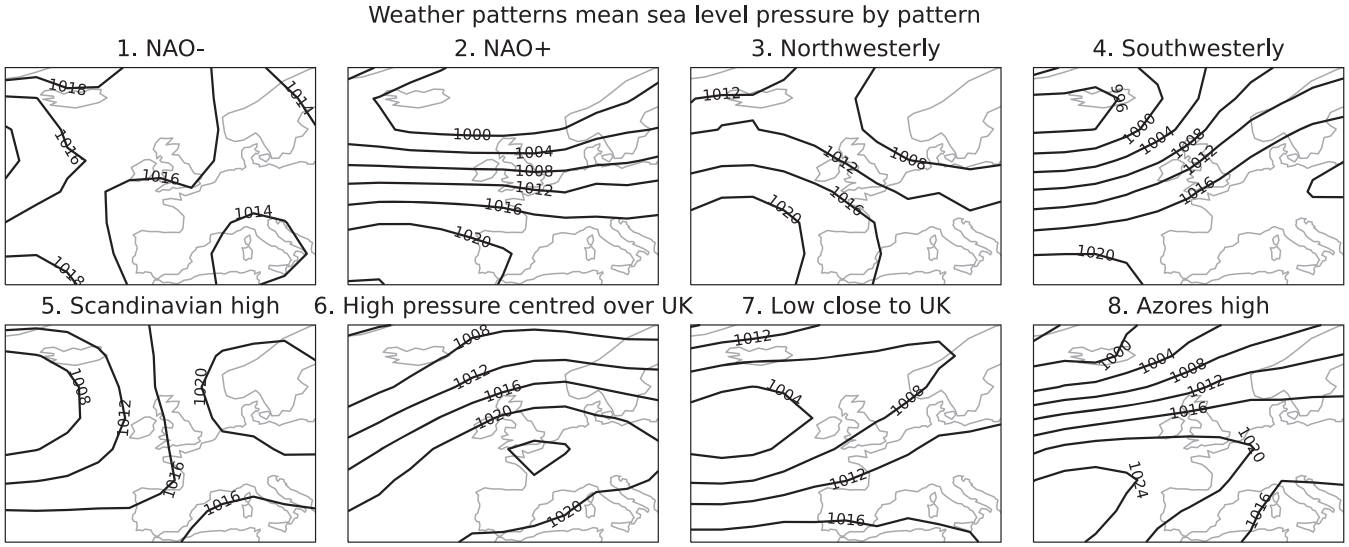

FIGURE 1 The eight broad weather patterns from Neal *et al.* (2016). Mean sea-level pressure (MSLP) contours are shown at 4-hPa intervals for each pattern, except NAO-, which has contours at 2-hPa intervals.

In evaluation of the UKCP18 Local output, Murphy *et al.* (2019) noted that observed summer precipitation generally agreed with the model output, but, for a case study of a heavy precipitation event, the amount of precipitation modelled was underestimated compared with National Climate Information Centre observations. Manning *et al.* (2023) found that, while UKCP18 Local data underestimated the frequency of occurrence of the strongest wind gusts ($> 32\ \mathrm{m \cdot s^{-1}}$) compared with point-based observations, they performed better than the ERA-Interim reanalysis at reproducing strong wind gusts. The UKCP18 Local suite also provides high-resolution model output for a future climate scenario, under the high-emissions Representative Concentration Pathway (RCP) 8.5 scenario (Kendon *et al.*, 2021). While these future climate data are not used in this article, subsequent work will explore changes in lee-wave activity in a changing climate using the same consistent framework and datasets.

As in Coney *et al.* (2024), vertical velocity on the 700-hPa pressure surface was used to identify lee waves from the model output, due to this pressure surface being above the orography, and captures lee waves sufficiently (as well as maintaining consistency with Vosper *et al.*, 2013). The data were available at three-hourly intervals at a 2.2-km horizontal resolution from January 1, 1982, 0300 UTC until December 30, 2012, 2100 UTC inclusive (Kendon *et al.*, 2021). The data were accessed through the Met Office Managed Archive Storage System (MASS) and analysed on the Natural Environment Research Council (NERC) computing facility JASMIN. In addition to the vertical velocity slices used by the deep-learning models, other variables were used to understand which conditions are important for the production of lee waves. These variables included the horizontal wind speed and direction, virtual potential temperature, and height on pressure surfaces from 925 hPa to 200 hPa. The model orography was used to calculate the two-dimensional standard deviation of the orography with a kernel of 5 × 5 grid points as a measure of local variability in the orography.

2.1.2 | Synoptic weather patterns

The daily mean sea-level pressure (MSLP) values over Western Europe have been clustered into similar weather patterns by using a "simulated annealing variant of *k*-means clustering" (Neal *et al.*, 2016). The 30 original patterns were grouped again into eight broad patterns, which are used in this article. These eight weather patterns are used here to examine the link between the prevalence and characteristics of trapped lee waves and the synoptic weather pattern. Figure 1 shows the MSLP for each of the patterns. Patterns 1 and 2 represent synoptic pressure regimes analogous to the two phases of the North Atlantic Oscillation (NAO). The negative NAO phase is associated with a flow-blocking and trough pattern (Benedict *et al.*, 2004), while a positive NAO is associated with strong westerly flow over Britain and Ireland (Washington & Palmer, 1999). The remaining six patterns all represent different locations of cyclonic and anticyclonic flow over northwestern Europe and the resultant synoptic wind directions over Britain and Ireland.

2.2 | Developing a climatology of lee waves

The trained deep-learning models from Coney *et al.* (2024) were applied to the 31 years worth of UKCP18 Local model data to classify each grid point as either a lee wave or non-lee wave, and to estimate wavelength, orientation (the direction of wave propagation, perpendicular to the wavefronts), and amplitude for lee waves over Britain and Ireland during 1982–2012. Specifically, the orientation as used in this article is the direction of the wavenumber vector associated with the trapped waves, and not the directions parallel to the wave crests and troughs. This vector has an underlying assumption that the trapped waves are monochromatic, while in reality trapped lee waves generated by three-dimensional orography are three-dimensional, as well as being a superposition of monochromatic waves of different wavelengths. Likewise, different measures of amplitude of lee waves exist. An advantage to using vertical velocity is that it remains consistent with previous studies of lee waves over Britain and Ireland (e.g., Vosper *et al.*, 2013), as well as being applicabile to forecasting and impacts of lee waves (such as for aviation). One more direct measure is the displacement of isentropic surfaces. Since $w = (U, V) \cdot \nabla_h(\zeta)$, where $\nabla_h(\zeta)$ is the slope of the isentropic surface, the vertical velocity w is necessarily proportional to the incoming wind (U, V).

The deep-learning models were trained to segment and characterise lee waves from UKV output (at a spatial resolution of 1.5 km regridded to 2 km), rather than UKCP18 Local data (at a spatial resolution of 2.2 km). It was expected that the deep-learning models would still be able to recognise wave patterns regardless of the resolution of the models, but there was some uncertainty as to what may be smoothed on NWP model output with a slightly coarser horizontal spatial resolution.

To verify that lee waves were still recognised by the deep-learning models regardless of resolution, the deep-learning models were applied to model output on data from models with spatial resolutions of 1.5 and 2.2 km, for data valid at the same times. Since there was no overlap in time between the UKV operational archive and UKCP18 Local data, data from Met Office Global and Regional Ensemble Prediction System (MOGREPS) (the operational Met Office ensemble forecasting system), which has the same spatial resolution as UKCP18 Local, were used instead. Vertical velocity data on the 700-hPa surface from both UKV and MOGREPS data, for February 1, 2021–January 30, 2022 (which excluded the training set for the deep-learning models), were used for this verification and regridded to 2 km. Figure 2 shows in the top row (Figure 2a,b) that lee waves can be recognised regardless by the segmentation model (black contour). Some finer detail in the vertical velocities is lost in the MOGREPS data (Figure 2b) compared with UKV (Figure 2a). This may lead to the slightly smaller frequency of occurrence observed in the MOGREPS data (Figure 2d) compared with UKV data (Figure 2c). Given the similarity between the MOGREPS and UKV data, it is reasonable to assess that the results are not too sensitive to the difference in resolution, setup, and post-processing between MOGREPS and UKV.

The segmentation masks and characteristics generated by the machine-learning models applied to the UKCP18 Local data were interrogated to investigate how lee waves vary with time of day and season and between synoptic weather patterns, to produce statistics for the prevalence and characteristics of lee waves geographically, and overall, for Britain and Ireland.

2.3 | Relationship of lee waves with other meteorological variables

In order to examine which meteorological variables were important for the generation of lee waves, six sub-regions (shown in Figure 3) with high topography were identified for analysis, as these showed a high frequency of occurrence of lee waves. To aid comparison with previous studies, these regions were deliberately chosen to be similar to the 3DVOM forecast domains used in Sheridan *et al.* (2017), but with the region over North Wales being extended southwards and the inclusion of a new region over the southwest of Ireland. The lee waves produced by the segmentation ML model were compared with other variables from the UKCP18 data within these regions, using random-forest models to decide which variables to concentrate on, from a larger set of variables. Random forests are machine-learning models that classify data based on a sequence of rules in decision trees, each trained on a subset of the data (as introduced in Section 1). Random-forest models were used because they have fewer parameters to tune than deep-learning models and perform well on data with small sample sizes (Biau & Scornet, 2016).

To examine the physical drivers of lee waves, a set of variables from the UKCP18 Local data likely to be correlated to wave activity was chosen. These variables included the horizontal wind speed and virtual potential temperature on the 200-, 300-, 500-, 650-, 750-, 850-, and 925-hPa pressure surfaces. From these, the Brunt–Väisälä frequency, the Scorer parameter, and the difference in wind direction between pressure levels were calculated. Also chosen were the 10-m wind speed, the weather pattern, and the height of the orography and its local standard deviation (to capture the local variability in orography). This gave 40 variables in total.

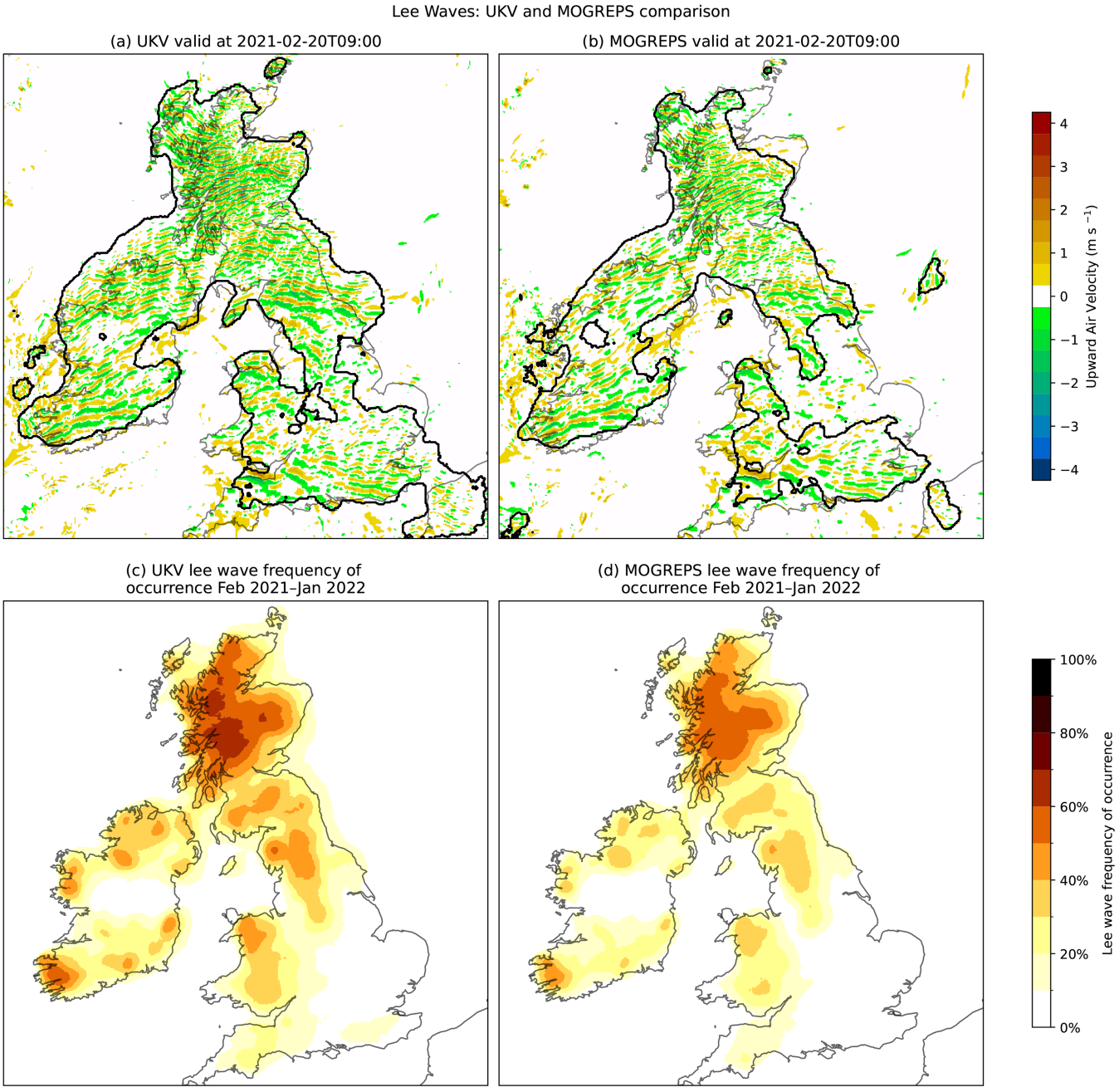

FIGURE 2 Resolution sensitivity between NWP models with spatial resolutions of 1.5 km (UKV) and 2.2 km (MOGREPS). (a) UKV and (b) MOGREPS output, respectively, for vertical velocities on the 700-hPa pressure surface as coloured contours, with the black contour showing the region identified by the deep-learning model as containing wave pixels. The frequency of occurrence of lee waves from February 2021–January 2022 inclusive for these two model datasets is shown in panels (c) and (d), respectively. [Colour figure can be viewed at wileyonlinelibrary.com]

The Brunt–Väisälä frequency at x hPa was calculated as $N(x) = \sqrt{\frac{g}{\theta_v}\frac{d\theta_v}{dz}}$, where $d\theta_v/dz$ was calculated using the surfaces above and below x hPa and θ_v at x hPa by interpolating the data from the surfaces above and below x hPa. This produced mid-pressure-level values of N, on the 887-, 825-, 775-, 725-, and 675-hPa surfaces. The Scorer parameter was calculated as $l = N/U$, with the smaller (and harder to calculate accurately, given data were only available on a limited number of pressure levels) shear term neglected (Blockley & Lyons, 1994). However, the influence of changing wind direction was still explored, by including the change in wind direction between each pressure surface as separate inputs. Values of the Scorer parameter were produced on the same pressure surfaces as the Brunt–Väisälä frequency by interpolating U onto the same surfaces as the Brunt–Väisälä frequency

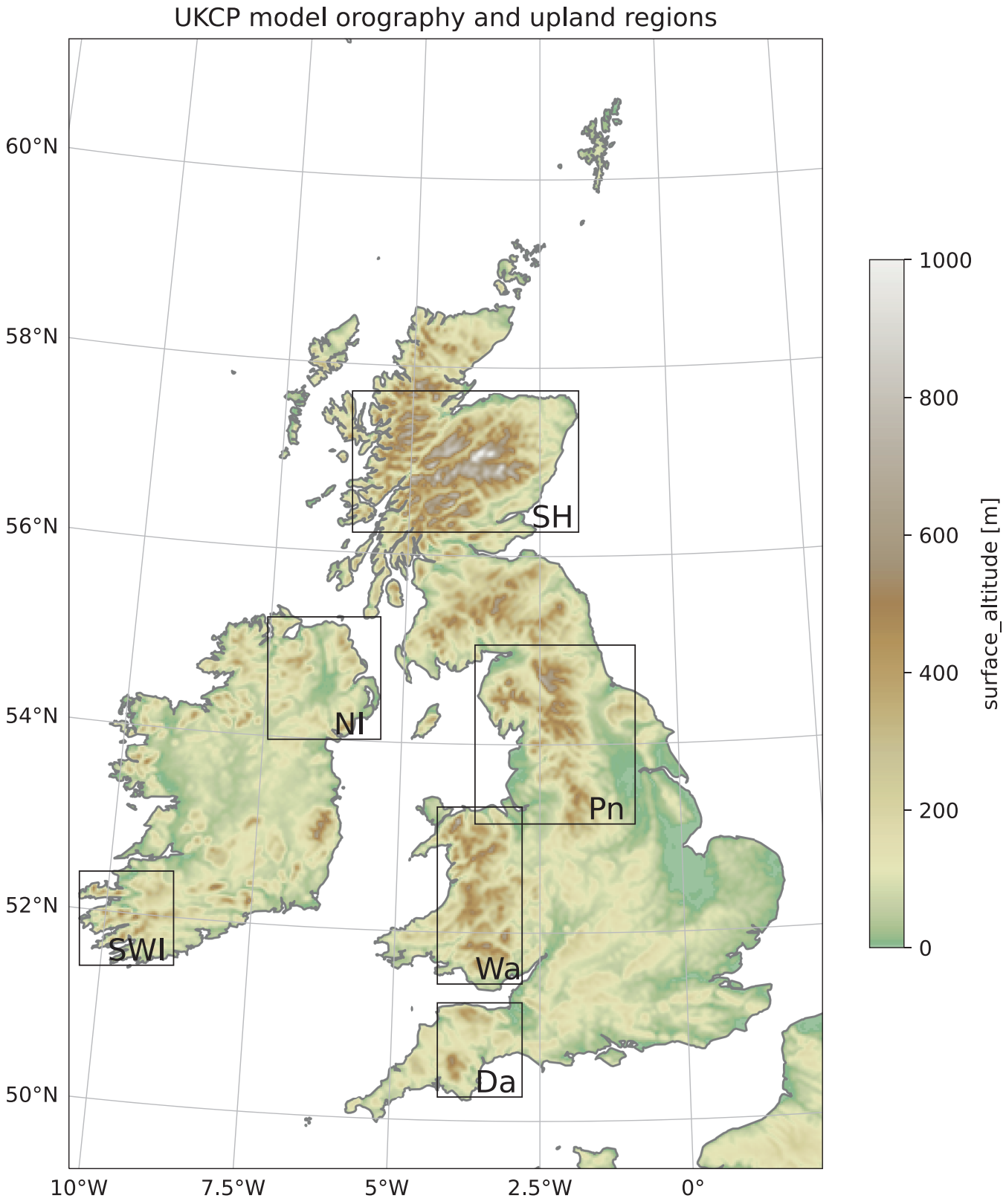

FIGURE 3 UKCP18 Local model orography over Britain and Ireland. Black bounding boxes indicate upland regions used in later analysis and are labelled for identification in the text: SH: Scottish Highlands; NI: Northern Ireland; SWI: Southwest Ireland; Pn: Pennines; Wa: Wales; Da: Dartmoor. [Colour figure can be viewed at wileyonlinelibrary.com]

(Blockley & Lyons, 1994). The change in the Scorer parameter from 887 to 675 hPa (spanning the 700-hPa pressure surface on which lee waves were identified) was also calculated.

One random forest was trained for each region in Figure 3 using these variables, with the aim of predicting whether or not lee waves were present. Random forests were used because of the lack of normalisation or augmentations required on the input data: these models typically train more quickly than neural networks and, due to their internal modelling through decision boundaries, allow the interrogation of which variables were important in the prediction (e.g., Lundberg *et al.*, 2020). The "decision nodes" in each decision tree are based on the data: for example, horizontal wind speed being greater than some threshold.

The influence of each of these variables on the random forest's prediction was measured using SHapley Additive exPlanations (SHAP) values (Lundberg & Lee, 2017). Shapley values are used in game theory to estimate the contribution of different players in a game to the game's outcome. SHAP extends the concept of Shapley values to machine learning in order to explain each feature's contribution to the model output. Here, the version of SHAP for tree-based models (TreeExplainer: Lundberg *et al.*, 2020) was used. The selected meteorological variables are not all independent, and using SHAP values to visualise the importance of the variables to the prediction, instead of other importance methods such as the permutation feature importance, means that the correlation between variables in the data is accounted for in the game theoretic approach; however, SHAP values can still be spread between features that are correlated (Chen *et al.*, 2023).

The correlation between the amplitude of lee waves and the horizontal wind speed aloft was also examined. The local maximum lee-wave amplitude and local maximum 750-hPa wind speeds were aggregated for 8×8 pixel regions (17.6 km $\times$ 17.6 km) for part of the climatology period (1982–1987 inclusive). It is acknowledged that a fuller dataset would have been ideal, but the reduced dataset of horizontal wind speeds still consisted of some 12,000 files and is assumed to be sufficiently robust for analysis here to represent the distribution of horizontal wind speeds for the present-day climate. A reduced dataset was used in order to reduce the time taken to obtain the additional data and conserve disk space and processing time, as opposed to using the full dataset. Aggregating the data into 8×8 pixel squares meant that small local variations in wind speed and amplitude were removed, allowing a fairer comparison between broader-scale wind speeds and amplitudes. The crest of a wave and therefore the peak amplitude should occur within these squares in all cases apart from the longest wavelength lee waves, without smoothing out too much fine detail for short-wavelength lee waves.

3 | RESULTS

3.1 | Frequency of occurrence of lee waves

Figure 4 shows the seasonal frequency of occurrence of lee waves over Britain and Ireland. Here, seasons refer to meteorological seasons so, for example, spring is March, April, and May. The general location of lee waves changes little between seasons, with lee waves occurring more frequently over hilly areas (Scotland, Northern England, Wales, Southwest Ireland). The winter months contain the highest frequency of occurrence of lee waves and the summer months the lowest. For example, over the Highlands of Scotland, lee waves occur 60%–70% of the time during winter in the model forecasts, compared with 40%–50% of the time during the summer months. This is consistent with Vosper *et al.* (2013), where lee waves were forecast 57% of the time over the Grampians in Northern Scotland and they noted that lee waves are less common in the summer months than in the winter.

Figure 5 shows a histogram relating the UKCP18 Local model orography and the frequency of lee waves per pixel (normalised by column). The standard deviation σ of the model orography for the 5×5 pixel box surrounding each grid cell was calculated (as a measure of the local variability of orography: one would expect fewer lee waves over a plateau than a region of variable orography) and plotted against the grid cell's frequency of wave occurrence during the period of interest. There is a correlation (least-squares $R^2 = 0.711$ and Spearman rank $\rho = 0.662$) between the local orography variability, σ, and the frequency of occurrence of waves. The spread within the distribution shown in the histogram is likely due to lee waves propagating well downstream past the orography, and the influence of meteorology, which will vary across the country.

3.2 | Diurnal effects

Figure 6 shows the frequency of waves for different times of day in the UKCP18 data. Plots are only shown every six hours, since there is no discernible diurnal variation in the frequency of occurrence of lee waves over the entirety of Britain and Ireland. Figure 7 shows the probability density functions (PDFs) for the frequency of wave occurrence and characteristics of lee waves for each time of day, and

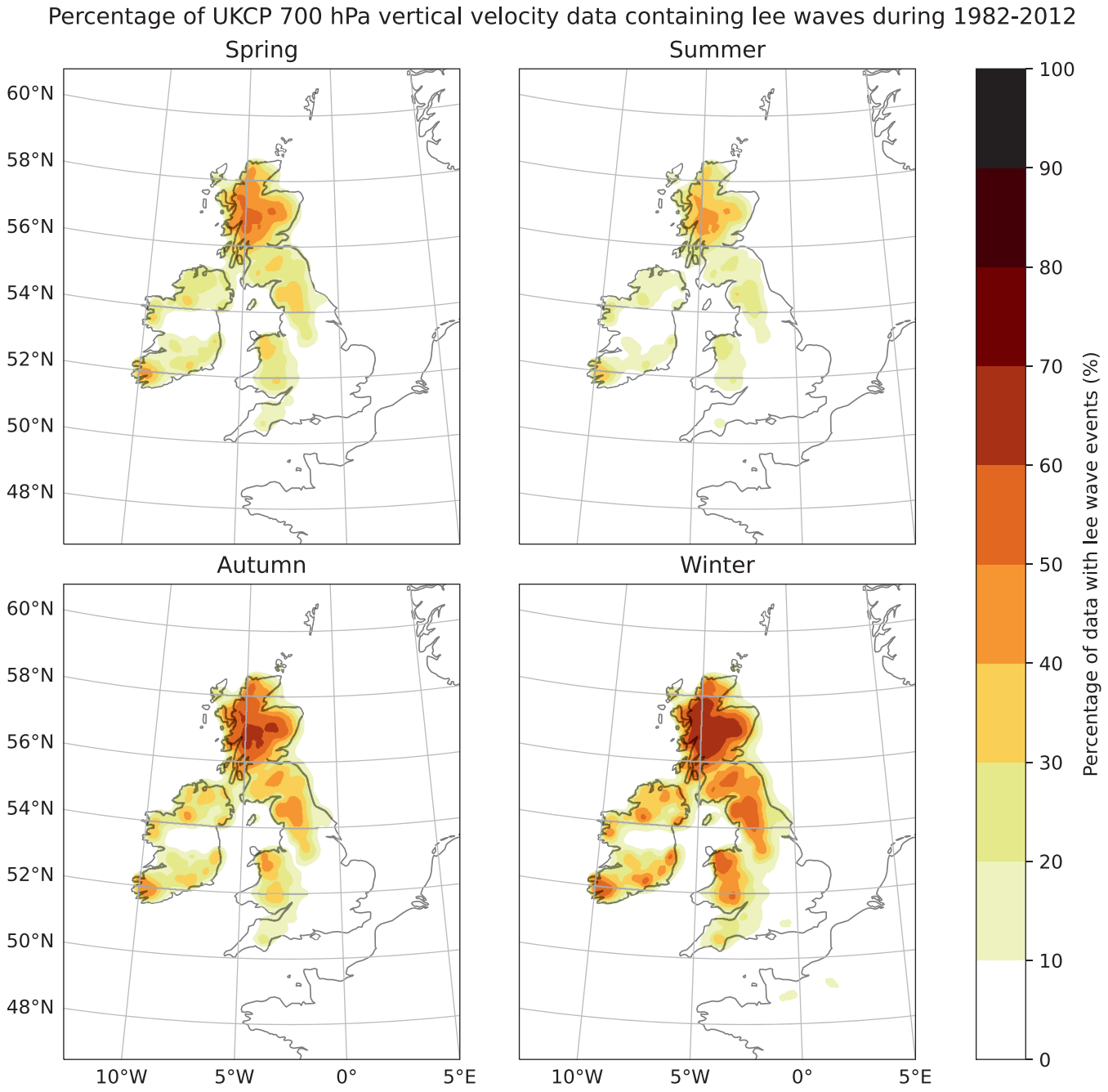

FIGURE 4 Maps showing the frequency of occurrence of lee waves over the UK by meteorological season. [Colour figure can be viewed at wileyonlinelibrary.com]

confirms that there is very little diurnal variation in either the frequency of occurrence or the wave characteristics in the data.

Within the range of amplitudes observed in Figure 7b, from < 0.1 to 5 $m \cdot s^{-1}$, the lack of a diurnal change except at very low probabilities and high amplitudes suggests that there is little diurnal change in lee-wave amplitude during the day. Likewise for the wavelength of waves (Figure 7c), most observed lee waves have a wavelength in the range 10–20 km but no diurnal change. Most lee waves are either northeast/southwest or northwest/southeast aligned (Figure 7d), but there is not an observed diurnal cycle of lee-wave orientations. Hence, there is no evidence to suggest a diurnal cycle in lee-wave frequency of occurrence or characteristics over Britain and Ireland, similar to the conclusions of Worthington (2006), who found no diurnal cycle of mountain-wave amplitude, albeit using a different method and a more localised dataset. These results imply that the properties of lee waves (at least, those propagating above the orography on the 700-hPa pressure

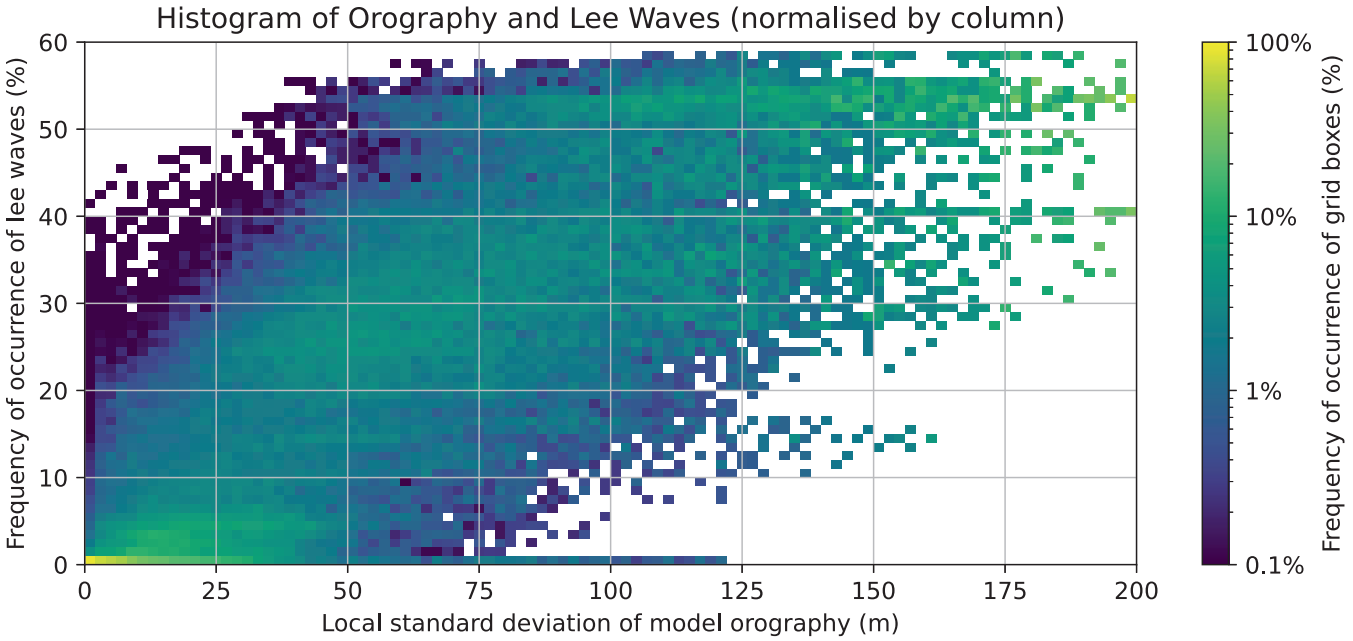

FIGURE 5 Histogram of the UKCP18 model orography and the frequency of waves per pixel. This plot is normalised by column (so the sum of each column is 100%). The colour bar represents the occurrence of grid boxes with a certain frequency of occurrence of lee waves given the local standard deviation of orography. [Colour figure can be viewed at wileyonlinelibrary.com]

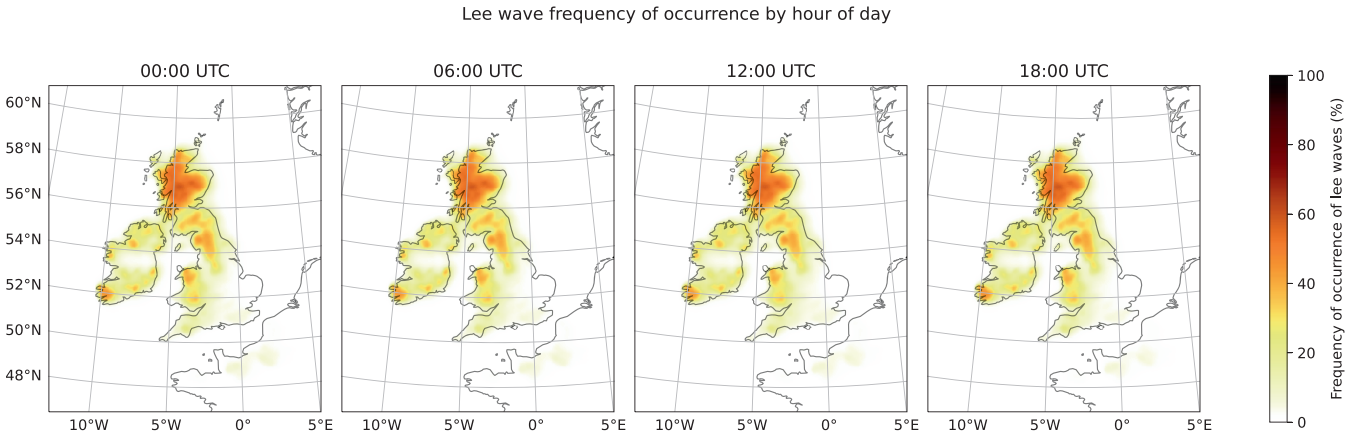

FIGURE 6 Map plots showing the frequency of occurrence of lee waves by hour of the day in the UKCP18 data over the climatology period 1982–2012 (the missing four panels for 0300, 0900, 1500, and 2100 UTC look identical to the four shown here). [Colour figure can be viewed at wileyonlinelibrary.com]

surface) are insensitive to boundary-layer structure, which would be a leading cause in any discernable diurnal cycle if it were observable in the UKCP18 Local data. It is still possible that an analysis of waves trapped closer to the surface (and the boundary layer) may show more evidence of a diurnal cycle.

3.3 | Weather patterns

To investigate the relationship between the occurrence of lee waves and the synoptic weather pattern (introduced in Section 2.1.2), Figure 8 shows the frequency of occurrence of lee waves geographically by weather pattern. The weather patterns most conducive to lee waves being generated are the NAO+ and Southwesterly patterns. The Scottish Highlands receive the most lee waves under all weather patterns, but the trend is not consistent across regions: the Pennines receive a similar amount of waves in NAO+ conditions, but comparatively less so under the Azores high pattern. While the spread of NAO- weather patterns is fairly constant between seasons, NAO+ conditions, which have a high frequency of occurrence of lee waves, occur more often during the winter than in the

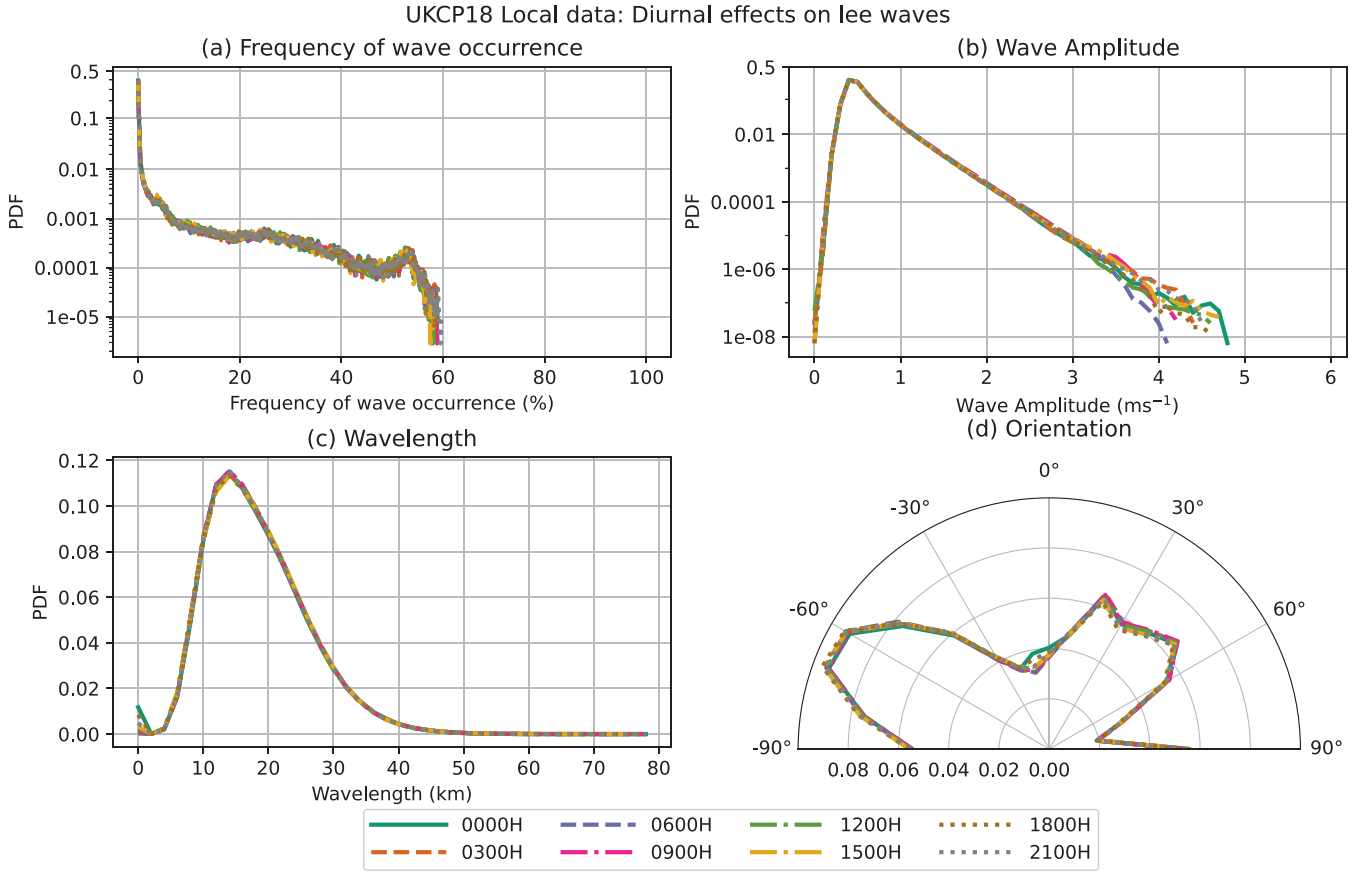

FIGURE 7 Probability density Functions (PDFs) of wave frequency of occurrence and characteristics by time of day. [Colour figure can be viewed at wileyonlinelibrary.com]

summer. This may explain the seasonal changes of lee waves in Figure 4: more wave-conducive weather occurs during the winter than in the summer. Under Scandinavian high conditions, lee waves occur infrequently over most of Britain and Ireland except for Scotland and a small patch in southwest Ireland.

Figure 9 shows an overview of the frequency of occurrence and characteristics of lee waves under different synoptic weather patterns. Each subplot shows a PDF of the weather pattern's lee-wave frequency of occurrence or characteristic. The following subsections will consider these characteristics in more detail.

3.3.1 | Amplitude

Figure 9b shows that the amplitude of lee waves is distributed similarly between weather patterns. The largest amplitude lee waves occur under northwesterly and southwesterly flows; however, the changes in amplitude between weather patterns are small compared with the actual mean amplitudes. The vast majority of lee waves have amplitudes $< 1\ \mathrm{m \cdot s^{-1}}$, and the largest amplitudes in the data are $\approx 5\ \mathrm{m \cdot s^{-1}}$. However, the true wave amplitude in vertical velocity is likely larger (e.g., Wildmann *et al.*, 2021). Figure 10 shows the 95th percentile amplitude for each of the weather patterns. No weather pattern shows a distribution of 95th percentile amplitudes noticeably stronger than any other. The regions where the strongest amplitudes occur tend to be places with higher orography, such as the Scottish Highlands (under all weather patterns) and the Pennines and Wales (particularly for the NAO+ pattern).

Table 1 shows how often lee waves of different amplitudes occur in the regions outlined in Figure 3. Overall, across Britain and Ireland there are lee waves with amplitudes stronger than $1\ \mathrm{m \cdot s^{-1}}$ more than half the time. While they are rare in general, the strongest lee waves (with amplitudes in excess of $3\ \mathrm{m \cdot s^{-1}}$, which are those more likely to be related to rotors and strong turbulence events: Vosper *et al.*, 2013) occur most frequently in the Scottish Highlands, where lee waves with amplitudes exceeding $3\ \mathrm{m \cdot s^{-1}}$ occur 0.3% of the time (equating to 0.5% of lee waves events in the region). The Pennines have the next most frequent occurrence of high-amplitude lee waves, at 0.04% of the time, almost ten times less likely

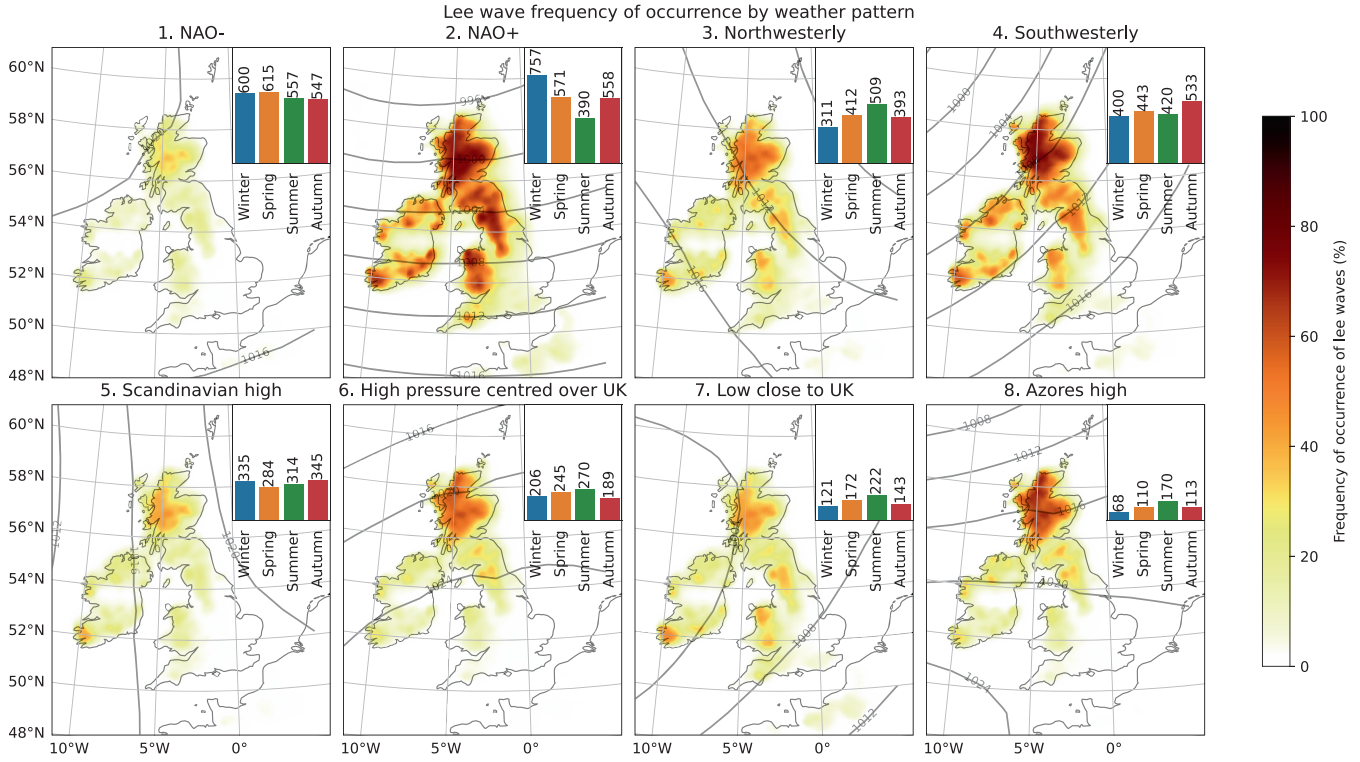

FIGURE 8 Lee-wave frequency of occurrence under each weather pattern (Neal *et al.*, 2016) from the UKCP18 model data over the climatology period 1982–2012 (0% indicates that lee waves never occur in a given weather pattern, and 100% indicates that lee waves always occur under that pattern). Average mean sea-level pressure contours are plotted for each pattern. The insets show, for each season, the total number of days on which the respective weather pattern occurred during the climatology period 1982–2012. [Colour figure can be viewed at wileyonlinelibrary.com]

than in the Scottish Highlands region. High-amplitude lee waves are rarest in Northern Ireland and Dartmoor. The occurrence of lee waves at stronger amplitudes decreases rapidly: for example, lee waves with amplitudes stronger than 2 $m \cdot s^{-1}$ occur 10.0% of the time overall, which decreases to 2.89% for 2.5 $m \cdot s^{-1}$ and further to 0.624% of the time for amplitudes stronger than 3 $m \cdot s^{-1}$.

3.3.2 | Wavelength

Figure 9c shows the PDFs of the wavelengths of lee waves under the different weather patterns. Lee waves under most regimes (all except NAO+ and Southwesterly) have a modal wavelength in the range 10–15 km. Lee waves produced under Southwesterly flow have a modal wavelength of approximately 15 km, while the longest modal wavelengths of approximately 20 km occur under NAO+ conditions.

Figure 11 shows the mean wavelength of lee waves over Britain and Ireland by weather pattern. The longest wavelengths occur in general in NAO+ conditions, to the east of Ireland and on the lee (east) side of the Pennines. The shortest wavelengths tend to occur under NAO- conditions, though with some longer wavelengths in Scotland. Apart from the NAO+ case (and, to a lesser extent, Southwesterly conditions), the mean wavelengths remain generally similar between patterns. The distributions of wavelengths in Figure 9c are broadly similar for all patterns except NAO+ and Southwesterly. According to theory, the wavelength is related to the Scorer parameter (e.g., World Meteorological Organization, 1993). However, outside the NAO+ and Southwesterly patterns, there is little change in the wavelengths produced overall. The NAO+ and Southwesterly patterns have generally faster wind speeds U than the other patterns, therefore decreasing the Scorer parameter $l = N/U$ during these patterns (assuming the stability N stays similar), and hence this results in longer wavelengths through the formula $\lambda = 2\pi/l$. With that said, the spatial pattern showing longer wavelengths on the lee side of orography suggests that the orography may affect the wavelength (particularly noticeable during NAO+ conditions, over the Pennines for instance). This could be because the length of the orography controls whether or not waves are produced at their theoretical wavelength. In addition, the horizontal spatial resolution of 2.2 km means that the minimum wavelength that can be resolved accurately within the data

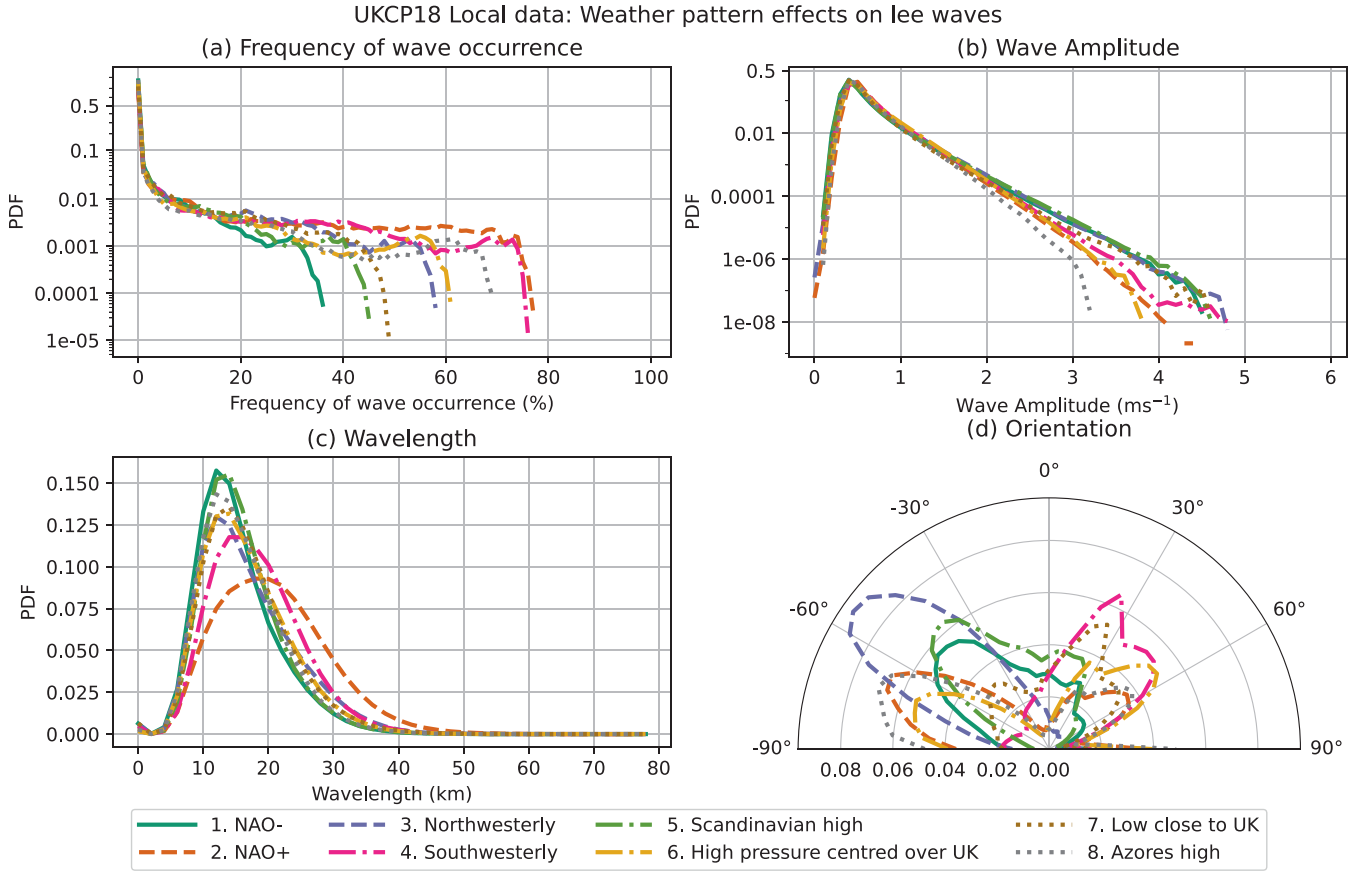

FIGURE 9 Probability density functions (PDFs) of lee-wave frequency of occurrence, amplitude, wavelength, and orientation under different synoptic weather patterns. [Colour figure can be viewed at wileyonlinelibrary.com]

is about 13 km (six grid points). Shorter wavelengths may be resolved (from 4.4 km, or two grid points, or longer), but possibly with decreased amplitude and uncertain wavelength. All except the NAO+ pattern have modal wavelengths of approximately 13 km.

3.3.3 | Orientation

Figure 9d shows that there is a different orientation pattern depending on the weather pattern. Some patterns have a unimodal distribution, such as the Southwesterly or Northwesterly patterns, generally in line with the predominant wind direction (lee waves in Northwesterly conditions tend to be northwest/southeast aligned). Others, such as the NAO+ and Low close to UK patterns, have bimodal orientation distributions. Since the eight patterns are themselves groupings of 30 more varied ones, there could be different pressure regimes grouped together into one weather pattern affecting the orientation of lee waves, resulting in bimodal wave orientations under certain patterns.

Figure 12 shows the mean orientation of lee waves for each weather pattern over Britain and Ireland, with MSLP contours overlaid for the respective pattern. Lee waves are broadly aligned with the MSLP contours, notably so in Northwesterly and Southwesterly conditions. However, there is a large spread in the orientation predictions, shown by the red arrows (indicating $\sigma > 45°$) in Figure 12. The least variation in wave direction occurs under the Northwesterly and Southwesterly regimes (probably because these are the only patterns with a specific wind direction), and the most variation in the Low close to the UK, which inherently will have a wide range of wind directions associated with it depending on the exact position of the low.

3.4 | Correlation between lee waves and other NWP variables

To explore the physical processes likely to be controlling lee-wave generation, random-forest models are used to investigate how well lee waves can be predicted from

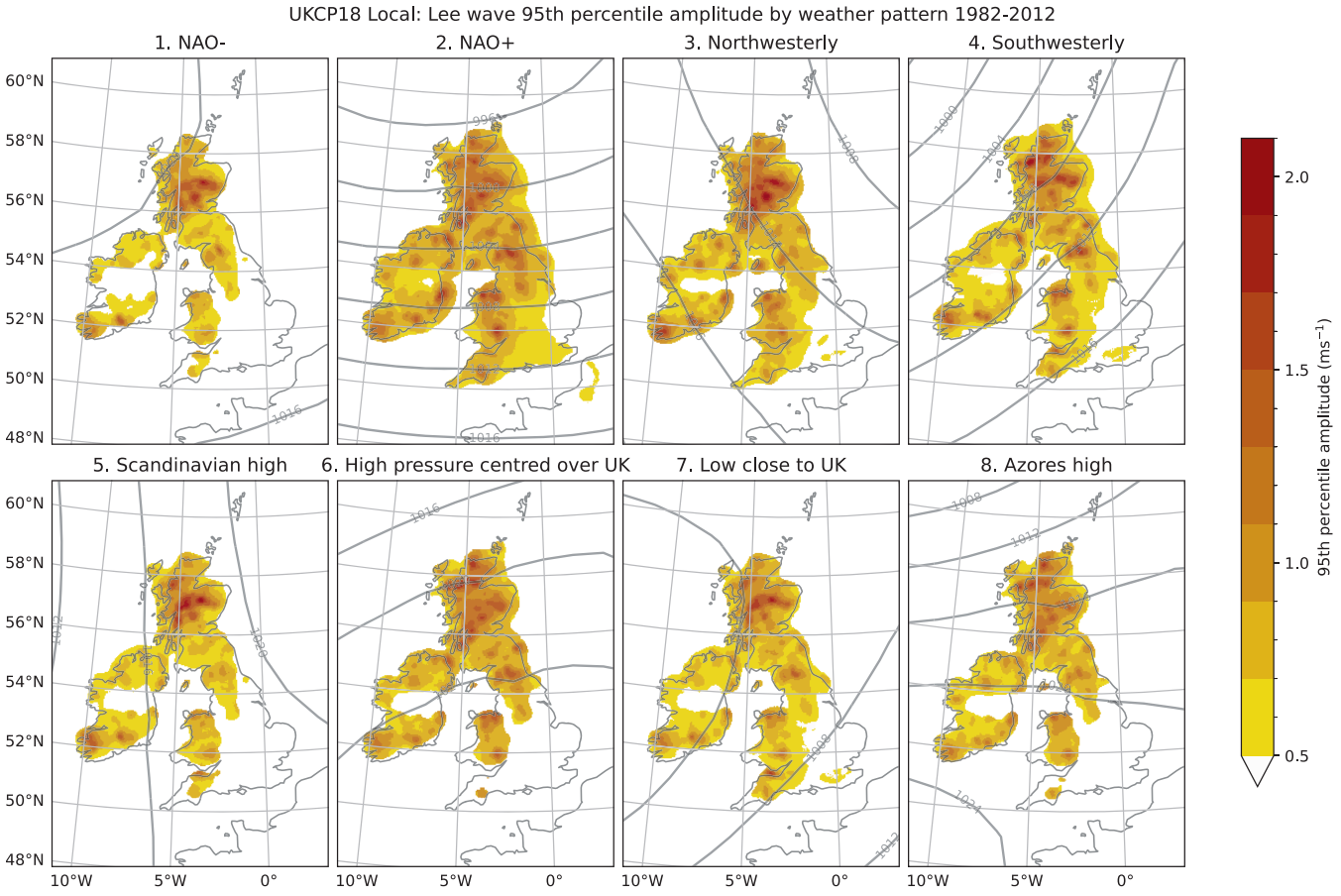

FIGURE 10 Map showing the 95th percentile lee-wave amplitude for each weather pattern. Areas with lee-wave occurrence < 5% of the time within the weather pattern are masked as 0. Note the compressed colour bar compared with other plots showing vertical velocity. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Table showing how often lee waves of different amplitudes occur somewhere within each region. This includes times when there is no wave activity in a region. The number in brackets excludes times where there is no wave activity in a region.

	Percentage of time with maximum wave amplitude			
Region	**$\geqslant$ 1 m · s^{-1}**	**$\geqslant$ 2 m · s^{-1}**	**$\geqslant$ 2.5 m · s^{-1}**	**$\geqslant$ 3 m · s^{-1}**
Scottish Highlands	34.7% (49.9%)	5.4% (7.7%)	1.66% (2.38%)	0.383% (0.551%)
Northern Ireland	11.3% (25.5%)	0.4% (0.9%)	0.06% (0.14%)	0.006% (0.012%)
SW Ireland	14.6% (28.3%)	0.8% (1.6%)	0.18% (0.36%)	0.023% (0.045%)
Pennines	20.1% (37.4%)	1.4% (2.6%)	0.28% (0.53%)	0.041% (0.076%)
Dartmoor	5.6% (18.9%)	0.2% (0.8%)	0.04% (0.12%)	0.006% (0.019%)
Wales	18.9% (38.3%)	1.4% (2.9%)	0.23% (0.46%)	0.025% (0.052%)
Overall	51.0% (59.7%)	10.0% (11.7%)	2.89% (3.39%)	0.624% (0.732%)

other NWP variables. A random forest is trained on a set of meteorological variables deemed to have a potential influence on the generation of lee waves in each of the regions shown in Figure 3 (one random forest per region): the horizontal wind speed on different pressure surfaces; the directional wind shear between surfaces; the virtual potential temperature θ_v; the Brunt–Väisälä frequency; a measure of the Scorer parameter (only the first term—see Section 2.3); the orography and its local standard deviation; the half-sine of the month of the year (precisely $\sin\left[\pi(m-1)/12\right]$, where m is the month of the year: $m = 1$ is January and so on); and the weather pattern. These

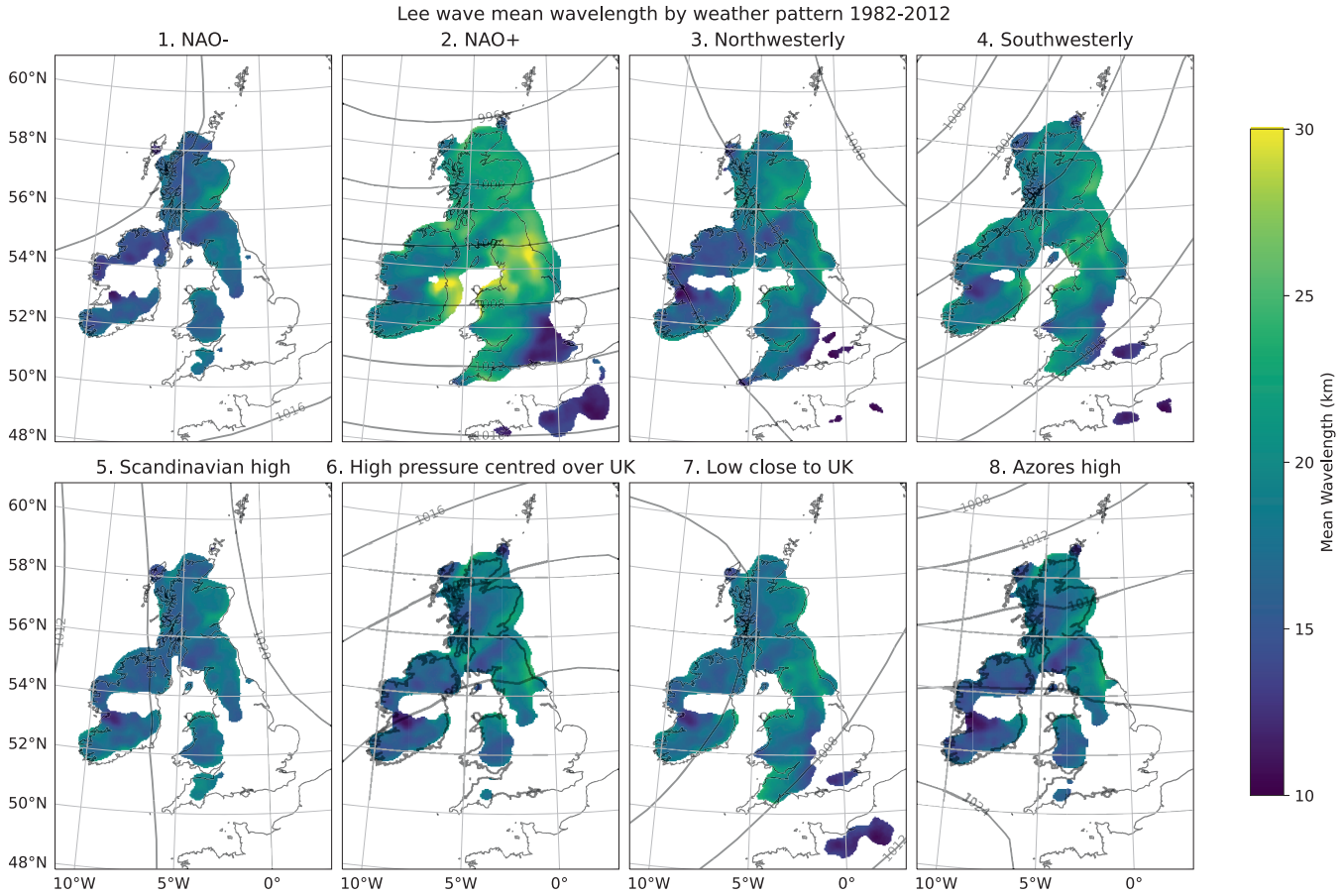

FIGURE 11 Map showing the mean wavelength by weather pattern. [Colour figure can be viewed at wileyonlinelibrary.com]

variables were chosen because of their involvement in the Scorer parameter (the stability, horizontal wind speed, turning effects) or the generation of lee waves seen earlier in this section (the month of the year, weather pattern, and orography). The relative importance of each of these is calculated using SHAP values, and results are shown by region in Figure 13a and as a normalised average over the regions in Figure 13b.

The Scorer parameter between the 925- and 850-hPa surfaces (i.e., at 887 hPa) has the largest SHAP importance value in the Scottish Highlands, Wales, and the Pennines. A bulk measure of the Scorer parameter has the highest importance value in Southwest Ireland. In Northern Ireland the 925-hPa horizontal wind speed is the most important, while in Dartmoor the 750-hPa wind speed is the most important. The Scorer parameter at 887 hPa has at least the third highest SHAP value in every region examined here. The wind speed with the larger SHAP values differs depending on the region: for example, regions with generally bigger orography, such as the Scottish Highlands and Wales, have higher wind-speed SHAP values at 850 hPa, while regions with lower orography such as Northern and Southwest Ireland favour wind speeds at 925 hPa and 10 m respectively. The Brunt–Väisälä frequencies at 825 and 775 hPa have higher SHAP values than on other pressure surfaces. The SHAP values of the Brunt–Väisälä frequencies are more consistent (but also lower) between surfaces and regions than the horizontal wind speeds. The virtual potential temperature and changes in the wind direction have lower SHAP values than the horizontal wind speeds below 500 hPa.

While SHAP is not a perfect measure for understanding the importance of different features, due to it sharing credit between co-dependent variables, it does attempt to address this more than other methods such as permutation importance. The high importance of the Scorer parameter, particularly below the 700-hPa level at which waves are being identified, corresponds with our physical understanding of the processes controlling wave trapping, adding further support to the utility of SHAP for analysing these results. For a wave to propagate up to 700 hPa would require the Scorer parameter to be larger than the wavenumber magnitude below this level.

To unpick this further, distributions of the variables with the six highest SHAP values were obtained and plotted, along with the lee-wave occurrence as a function of these variables, for each region. These are shown in Figure 14. In every case, the occurrence of lee waves does

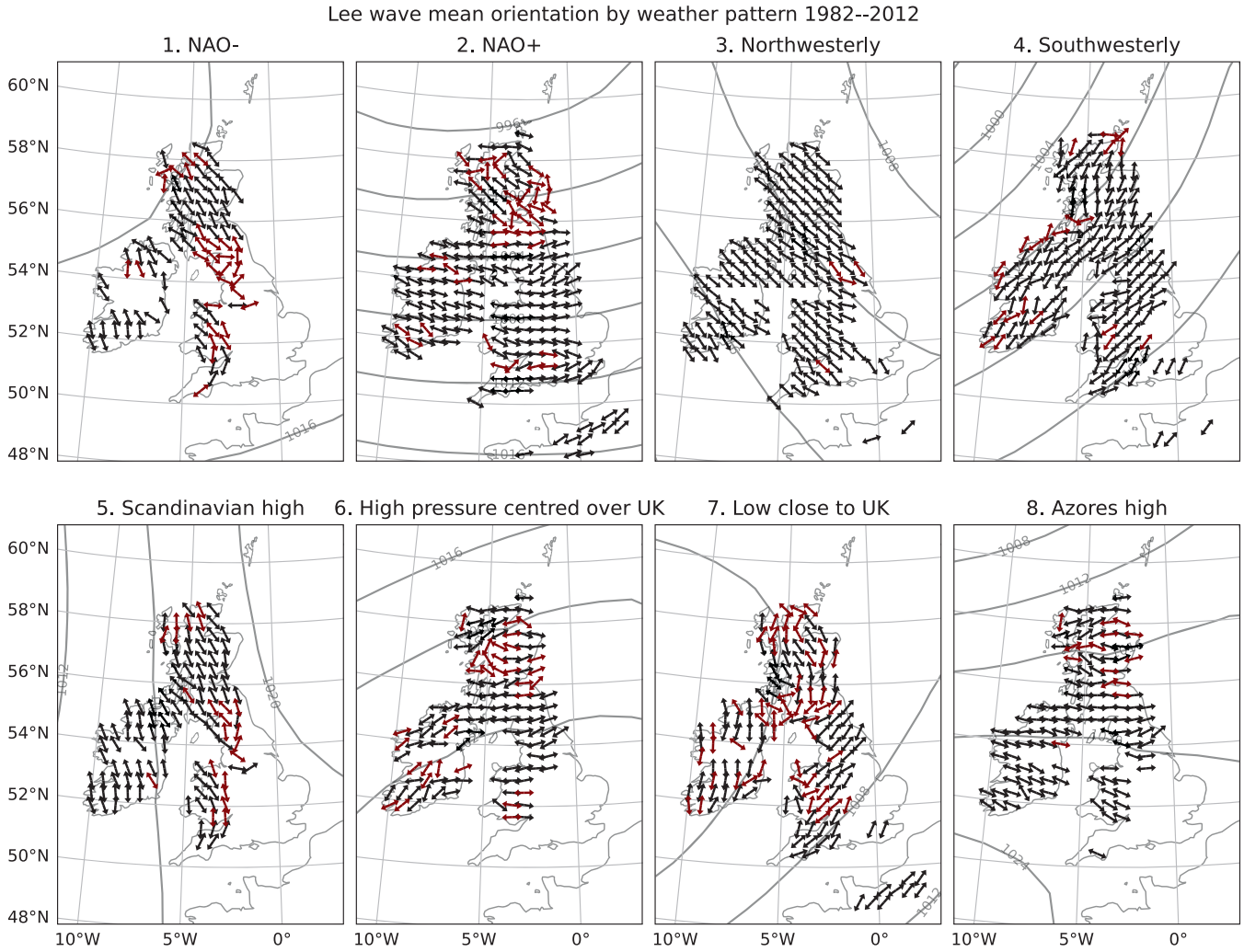

FIGURE 12 The mean orientation for each weather pattern. Arrows in maroon (on the web; lighter coloured in print) indicate that the standard deviation, σ, of the orientation at this point is > 45°, while black arrows indicate $\sigma \leq 45°$. [Colour figure can be viewed at wileyonlinelibrary.com]

not reach 100% in any region outside the very tails of the distributions, and there is a notable difference between regions. This may be due to the regions (shown in Figure 3) not being fully covered by topography. However, additional tests (not shown) restricted to pixels where the model orography was > 100 m only partly reduced the differences between regions. These results suggest that a combination of conditions are important for lee-wave occurrence in different regions, and lee-wave occurrence cannot be predicted by one variable alone. However, the differences in lee-wave occurrence in different regions do relate to the orography in the regions: there are fewer lee waves in Dartmoor and Southwest Ireland, which have relatively small hills compared with the Scottish Highlands, where there are much larger hills and a higher occurrence of lee waves.

Different measures of the Scorer parameter (Figure 14a–c) show a relatively steep gradient (particularly the bulk measure in Figure 14a) around 10^{-3} m^{-1}. The distribution of bulk Scorer parameters that support trapped lee-wave generation (the lower panel in Figure 14a) lies in the range 10^{-3} to 3×10^{-4} m^{-1}. The formula $\lambda = 2\pi/l$ (World Meteorological Organization, 1993) used to determine the wavelengths λ that these values of the Scorer parameter correspond to suggests that wavelengths of lee waves in the range 6–21 km are supported by the model, approximately in line with the range of wavelengths shown in Figures 7c and 9c. Values of the Scorer parameter higher than 10^{-3} m^{-1} would correspond approximately to wavelengths shorter than 5 km, which cannot be resolved in the model anyway. The drop off in lee-wave occurrence after this point may suggest that any lee waves that would be physically present may not be resolved in the model; however, these small-wavelength waves are likely to be of small amplitude anyway. The horizontal wind speed at 850 hPa (Figure 14d) and 925 hPa (Figure 14e) also shows a cut-off wind speed below which waves do not occur, again with regional differences. There is a marked difference in lee-wave occurrence in

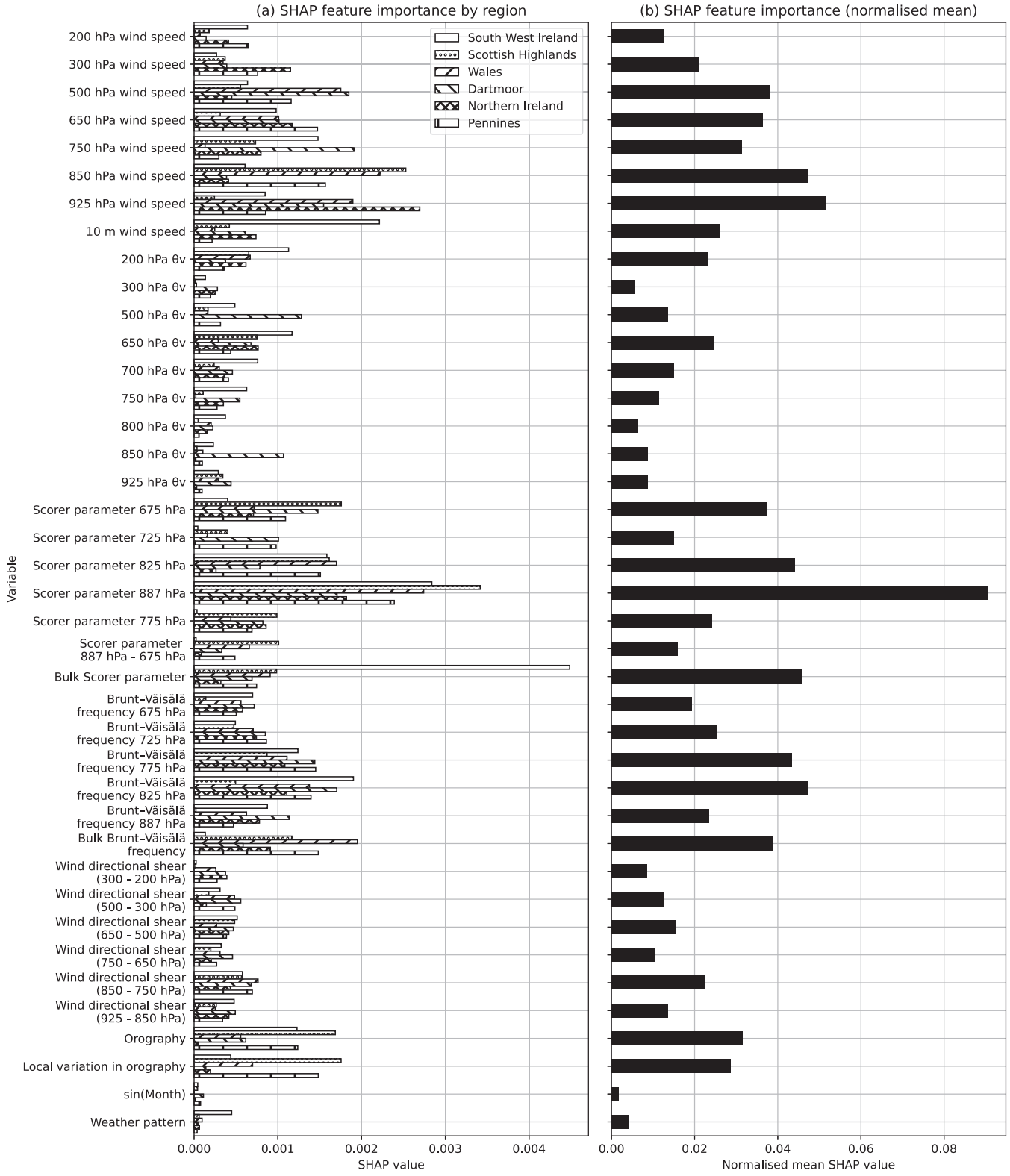

FIGURE 13 (a) SHAP values for each of the inputs into the random forest to predict wave generation, by region of Britain and Ireland. (b) Each region's SHAP values are normalised so that the sum of the SHAP values for each region is 1, and then the means of these are plotted.

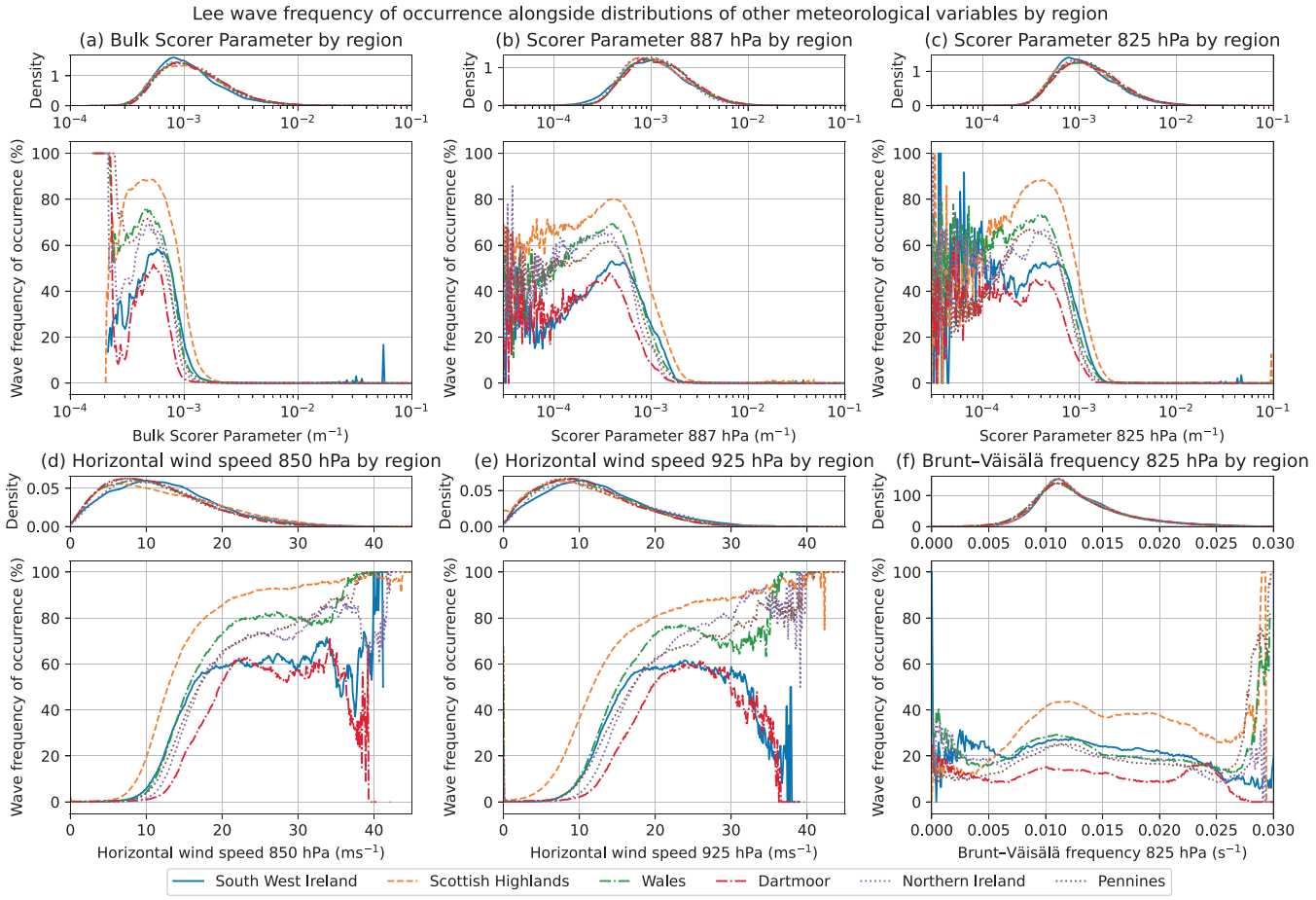

FIGURE 14 Distributions showing the relationship between the six variables with the highest SHAP values and the frequency of lee-wave occurrence (bottom of each pair of subplots). The PDF of each variable is also shown by region (top). [Colour figure can be viewed at wileyonlinelibrary.com]

Southwest Ireland and Dartmoor compared with the other regions for horizontal wind speeds greater than 30 m · s^{-1} (Figure 14e,f). This could be due to the lower orography in these regions, meaning forcing conditions for lee-wave generation are triggered less often. At lower horizontal wind speeds, flow is more likely to be blocked, meaning that horizontal wind speeds in these regions need to reach a certain threshold for waves to be generated. There is no such clear threshold or large change in the frequency of occurrence of lee waves as a function of the Brunt–Väisälä frequency at 825 hPa in Figure 14f, suggesting that it is the horizontal wind speed rather than the Brunt–Väisälä frequency that tends to control differences in the Scorer parameter between different locations and times, and hence controls lee-wave occurrence in Britain and Ireland. This is likely due to much greater spatio-temporal variability in wind speeds than stability over Britain and Ireland.

Simple analytical models have shown the importance of the Scorer parameter for the development of lee waves and this analysis confirms that by showing that lower-level values of l^2 seem to be the most important for lee-wave generation, particularly below the pressure surface that the climatology was created on (700 hPa). This makes sense, as the waves observed at 700 hPa will have propagated up from the surface and so are most affected by the wind and stability profile below 700 hPa. Further, it is variations in the horizontal wind speeds rather than in stability that are driving variations in the Scorer parameter. While the Scorer parameter used here neglects the curvature term, the relatively low importance of the wind shear compared with $\frac{N}{U}$ would suggest that the measure of the Scorer parameter used here is sufficient on its own.

3.5 | Relationship between wave amplitude and horizontal wind speed

Figure 15 shows a comparison between the amplitude of lee waves and the 750-hPa wind speeds for 1982–1987, for coarse-grained data (8 × 8 pixel regions as described in Section 2.3). This shorter subset of the full climatology data

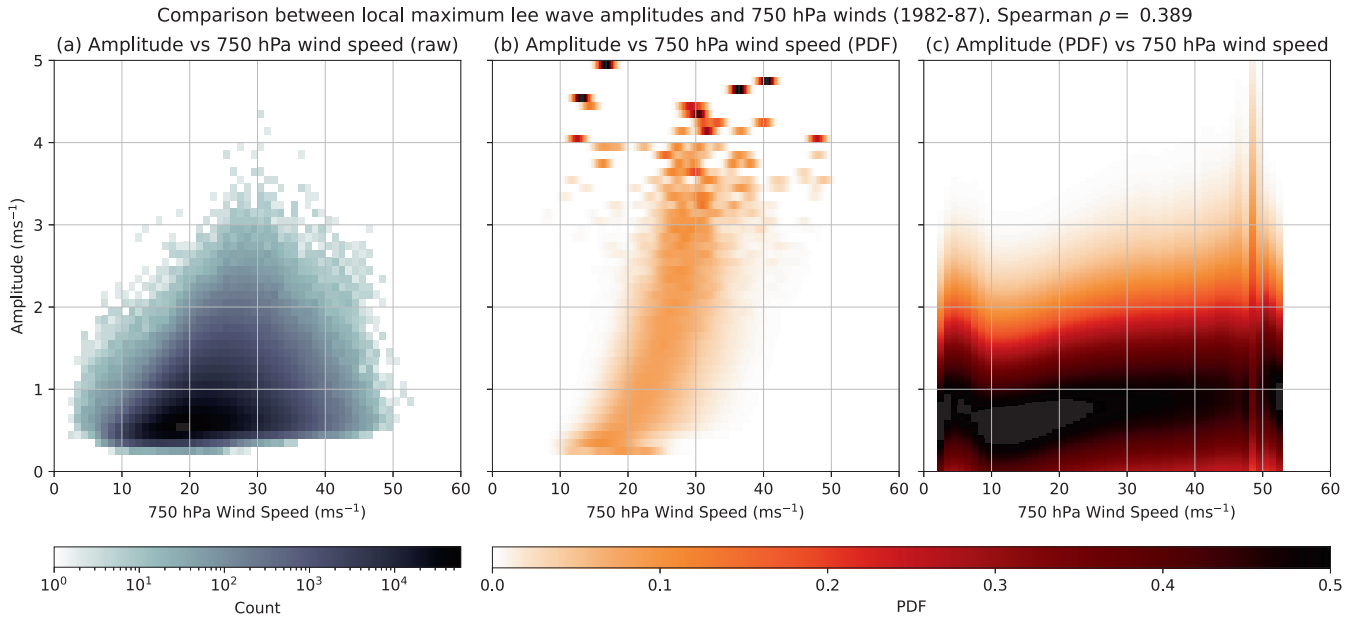

FIGURE 15 Histograms to compare the local (17.6 km × 17.6 km, or 8 × 8 grid points) maximal amplitude and 750-hPa wind speed for 1982–1987 over Britain and Ireland. (a) The raw data. Panels (b) and (c) are normalised to remove the effect of the large occurrence of low-amplitude waves and low wind speeds: (b) by amplitude (each row shows a probability density function (PDF) of 750-hPa wind speeds for the corresponding amplitude) and (c) by wind speed (each column shows a PDF of amplitudes for the corresponding 750-hPa wind speed). [Colour figure can be viewed at wileyonlinelibrary.com]

was obtained to save on disk space and time consumed extracting data from the archive. The 750-hPa surface was chosen, as it is one of the closest available pressure surfaces to the data used to predict the amplitudes (700 hPa). However, repeating the analysis using the horizontal wind speeds on the 650-hPa surface or the 10-m wind speeds instead produced similar histograms.

The raw data shown in Figure 15a is dominated by the high number of cases of small-amplitude lee waves ($< 1\ \mathrm{m \cdot s^{-1}}$), with a large spread of surface wind speeds at lower amplitudes. The Spearman rank correlation coefficient is $\rho = 0.389$, which does not show a strong correlation between the amplitude of lee waves and the 750-hPa wind speed, suggesting that there is no strong linear relationship between lee-wave amplitude and horizontal wind speeds, with the underlying orography and potentially the occurrence of flow blocking also playing a significant role. However, when normalising by amplitude in Figure 15b, there is a positive trend where stronger lee waves have stronger 750-hPa wind speeds associated with them. For example, the modal 750-hPa wind speed for lee waves with an amplitude of $1\ \mathrm{m \cdot s^{-1}}$ is $\approx 22\ \mathrm{m \cdot s^{-1}}$, and the modal 750-hPa wind speed for lee waves with an amplitude of $3\ \mathrm{m \cdot s^{-1}}$ is $\approx 30\ \mathrm{m \cdot s^{-1}}$. However, when normalising by the 750-hPa wind speed in Figure 15c, there are similar distributions of wave amplitudes for each wind speed. This shows that high wind speeds are necessary for strong-amplitude lee waves, but certainly not sufficient, with the majority of high wind-speed events still corresponding to low-amplitude lee waves.

4 | SUMMARY AND CONCLUSION

This article presents a new climatology of lee waves, produced using machine-learning techniques trained to detect and characterise lee waves using NWP model output. This study builds on previous work to investigate lee waves over Britain and Ireland (such as Worthington, 2006; Vosper *et al.*, 2013) in several ways. The climatology covers a longer period (1982–2012) than Worthington (2006) and Vosper *et al.* (2013), with full coverage over the entirety of Britain and Ireland. This study also uses data from a more sophisticated NWP model than the dry, linear 3DVOM model used in Vosper *et al.* (2013). The machine-learning techniques used to develop the climatology of lee waves detect and characterise lee waves resolved by the NWP model, rather than using peak wave amplitudes. While the findings in this article support the conclusions of previous studies, this work goes further: for example, investigating the relationship between lee waves, their characteristics, and weather patterns, as well as exploring which underlying meteorological variables are driving the variability in lee waves.

This work verifies some of the findings from Vosper *et al.* (2013) For example, both studies found the following:

lee waves are relatively common in the Scottish Highlands, the north of England/Pennines, and North Wales; lee waves occur more often in the winter months than the summer (Figure 4); amplitudes stronger than 3 $m \cdot s^{-1}$ occur rarely (Figure 9b); and the strongest-amplitude waves occur during westerly flow, which corresponds to the strongest-amplitude waves produced in NAO+ and Southwesterly regimes (Figure 9b). In addition, both pieces of work suggest that the amplitude of lee waves increases with increasing horizontal wind speed (Figure 15).

This study did not find any evidence within the UKCP18 data to suggest a diurnal cycle in trapped lee waves or their characteristics, consistent with the findings of Worthington (2006). The analysis presented here expands on that from Worthington (2006) to cover a much longer time period (1982–2012) than they used, over a larger region (the entirety of Britain and Ireland compared with a region centred around central Wales and one VHF radar). This lack of a diurnal cycle of lee waves is not necessarily replicated elsewhere in the world: Ruff and Ólafsson (2019) attribute observed diurnal changes in downslope wind storms to gravity waves. It may also be that any diurnal cycle may be too small to be captured by the NWP model, although the lee waves simulated in the output of the MetUM used here are generally in good agreement with observations, as shown by Sheridan *et al.* (2017). Another possibility is that the impacts of the diurnal cycle are most significant in a shallow layer near the surface and so they only impact on the shortest wavelengths of gravity waves, which are not resolved in this model dataset.

Lee waves occur more often in the autumn and winter months than the spring and summer months: this is likely due to the prevalence of weather patterns more conducive to strong winds in the winter than the summer. The NAO+ and Southwesterly weather patterns, which have some of the highest prevalence of lee waves across the country (Figure 8), occur more often in autumn and winter (40% of days) than in spring and summer (33% of days) during the climatology period 1982–2012. The weather patterns have more influence on lee waves than any diurnal effects, with relatively few lee waves being produced during NAO- and Scandinavian high conditions, compared with NAO+ or Southwesterly conditions. In general, the frequency of occurrence of lee waves is determined by the orography: there is correlation (Figure 5) between the local orography and whether waves occur. However, the geographic pattern of wave occurrence changes depending on the weather pattern. The spatial difference in lee waves under weather patterns is likely due to the relative strength of winds experienced in different locations; for example, the greater occurrence of lee waves over Scotland compared with (for example) Wales in Figure 8 under Azores high conditions and Southwesterly or NAO+ conditions is likely due to Scotland experiencing higher wind speeds under Azores high conditions than Wales, whilst there are more equal wind speeds between the two under Southwesterly or NAO+ regimes.

There is some variation in wave amplitude depending on the weather pattern: while the amplitude of most waves is less than 1 $m \cdot s^{-1}$ (Figure 9b), the mean amplitudes in NAO+ conditions are stronger than the mean of the complete dataset over most of Britain and Ireland. Wavelengths are longest under NAO+ conditions, with Southwesterly conditions exhibiting a similar distribution of wavelengths, with a larger spread than the other patterns. Wavelengths under all the other regimes present a smaller distribution of wavelengths (Figure 9c), but with some differences as to where waves are located (Figure 11). For example, under NAO- conditions, lee waves tend to be confined to Scotland, portions of Ireland, and the west coast of England and Wales. In contrast, lee waves in other regimes (such as Northwesterly, Low close to the UK, and Azores high) are more likely to propagate further, for example, over the Pennines and towards the eastern coast of Britain. The orientation of waves is broadly in line with the synoptic wind direction (Figure 12), though with some large spread within the data depending on the regime, but this is likely due to the multimodal distribution of the orientation under some regimes. The Northwesterly and Southwesterly patterns have waves oriented northwest/southeast and southwest/northeast, respectively. In Figure 9d, the orientation of lee waves under these regimes has a unimodal distribution, while the other patterns tend to exhibit more bimodal behaviour (such as that seen for lee waves in NAO+ conditions).

When predicting the prevalence of lee waves based on other meteorological variables, the Scorer parameter, horizontal wind speed, and Brunt–Väisälä frequency came out as being more important (having larger SHAP values) than the directional wind shear, virtual potential temperature, or month of the year (despite there being a seasonal cycle of lee waves). There is a height dependence, with surfaces around and below where the waves were detected (700 hPa) being more important than those well above (compare the SHAP values for horizontal wind speeds at 850 and 925 hPa with those at 200 and 300 hPa in Figure 13). This makes physical sense, as waves propagate up from the surface before being trapped at a particular level. Levels below the trapping level will influence the propagation/trapping of waves directly, while levels above will not. However, there may still be a relationship (albeit weaker) between higher levels and wave activity, as the different levels in an atmospheric profile are often correlated. While there are variations in lee-wave activity with the

month of the year and weather pattern, as seen in Figures 4 and 8, these are likely due to variations in the distributions of the underlying physical variables such as wind speed seasonally and between weather regime. The SHAP values suggest that there is a stronger correlation between the underlying physical variables than with month of the year or weather pattern. For example, the weather patterns drive the synoptic conditions, which result in more conducive conditions for lee waves, but it is the actual values of the Scorer parameter and wind speeds on a particular day that will determine whether or not waves occur.

The relationship with lee-wave occurrence is shown for the Scorer parameter and the horizontal wind speed, but not the Brunt–Väisälä frequency, in Figure 14, suggesting that it is the horizontal wind speed rather than the stability that tends to control variations in the Scorer parameter. While both variables contribute to the Scorer parameter, the greater variability in the wind speed likely means this dominates over stability. However, there is a variation in occurrence of lee waves between regions for these variables. The observed cut-offs in the distributions of the Scorer parameter correspond to physically plausible wavelengths for the trapped lee waves.

This demonstrates the importance of the stability and horizontal wind speeds for the generation of lee waves: through the Scorer parameter as shown in Figure 13 and the relationship between amplitude and horizontal wind speeds in Figure 15. However, despite a dominance of smaller amplitudes in the data, there is correlation between the 750-hPa wind speeds and the amplitude of the lee waves at 700 hPa, when the data are normalised by wave amplitude, but no such similar correlation when normalising by the 750-hPa wind speed in Figure 15. This suggests that strong wind speeds are a necessary but not sufficient indicator of high-amplitude lee waves.

The machine-learning-derived climatology described in this article presents an opportunity to investigate the correlation between the physics of lee waves and other meteorological phenomena, for example, to aid in developing parametrisation schemes: lee waves are important for momentum transport, as well as surface impacts such as rotors (Bretherton, 1969). The work supports the usefulness of simple concepts such as the Scorer parameter for predicting trapped lee waves even in real-world three-dimensional examples beyond the idealised linear framework in which they were developed. The authors hope that this work will serve as a useful demonstration of the benefits and limitations of linear theory for lee waves in the real world. Further work will include investigating how the distribution and characteristics of trapped lee waves change over Britain and Ireland in a future climate scenario. This will show how lee waves over Britain and Ireland, and the associated hazards to aircraft and road transport, may change in a warming climate.

ACKNOWLEDGEMENTS

The authors thank Robert Neal (Met Office) and Jonathan Wilkinson (Met Office, now ECMWF) for their assistance and for providing the weather-pattern data, which were used extensively in this article. Thanks also to Steve Derbyshire, Peter Sheridan, and Callum Dinnett of the Met Office orography group for their many helpful discussions along the way. The authors also thank John Marsham (Leeds) and Corwin Wright (Bath) for their helpful comments and suggestions.

J. Coney was supported in this work by the Leeds–York–Hull Natural Environment Research Council (NERC) Doctoral Training Partnership (DTP) Panorama under grant NE/S007458/1, and a CASE award from the Met Office. The authors would also like to thank the three anonymous reviewers for their comments and suggestions which improved the paper.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All the meteorological data (including those from UKCP18) were obtained from the Met Office Managed Archive Storage System (MASS), accessed through the JASMIN service. The weather-pattern data were provided by Jonathan Wilkinson (Neal *et al.*, 2016; Wilkinson & Neal, 2021). The lee-wave climatology data were obtained by applying the machine-learning models (code available from Coney, 2023) to the meteorological data from MASS.

ORCID

Andrew N. Ross https://orcid.org/0000-0002-8631-3512
Annelize van Niekerk https://orcid.org/0000-0002-4706-1295

REFERENCES

Benedict, J.J., Lee, S. & Feldstein, S.B. (2004) Synoptic view of the North Atlantic Oscillation. *Journal of the Atmospheric Sciences*, 61, 121–144. Available from: https://doi.org/10.1175/1520-0469(2004)061<0121:SVOTNA>2.0.CO;2

Biau, G. & Scornet, E. (2016) A random forest guided tour. *TEST*, 25, 197–227. Available from: https://doi.org/10.1007/s11749-016-0481-7

Blockley, J.A. & Lyons, T.J. (1994) Airflow over a two-dimensional escarpment. III: nonhydrostatic flow. *Quarterly Journal of the Royal Meteorological Society*, 120, 79–109. Available from: https://doi.org/10.1002/qj.49712051507

Böhm, C., Schween, J.H., Reyers, M., Maier, B., Löhnert, U. & Crewell, S. (2021) Towards a climatology of fog frequency in the Atacama Desert via multi-spectral satellite data and machine

learning techniques. *Journal of Applied Meteorology and Climatology*, 60, 1149–1169. Available from: https://doi.org/10.1175/JAMC-D-20-0208.1

Breiman, L. (2001) Random forests. *Machine Learning*, 45, 5–32. Available from: https://doi.org/10.1023/A:1010933404324

Bretherton, F.P. (1969) Momentum transport by gravity waves. *Quarterly Journal of the Royal Meteorological Society*, 95, 213–243. Available from: https://doi.org/10.1002/qj.49709540402

Charlton-Perez, A.J., Dacre, H.F., Driscoll, S., Gray, S.L., Harvey, B., Harvey, N.J. et al. (2024) Do AI models produce better weather forecasts than physics-based models? a quantitative evaluation case study of Storm Ciarán. *Npj Climate and Atmospheric Science*, 7, 93. Available from: https://doi.org/10.1038/s41612-024-00638-w

Chen, H., Covert, I.C., Lundberg, S.M. & Lee, S.-I. (2023) Algorithms to estimate Shapley value feature attributions. *Nature Machine Intelligence*, 5, 590–601. Available from: https://doi.org/10.1038/s42256-023-00657-x

Coney, J. (2023) jdconey/LeeWaveNet: initial release. software. https://doi.org/10.5281/zenodo.8193019

Coney, J., Denby, L., Ross, A.N., Wang, H., Vosper, S., van Niekerk, A. et al. (2024) Identifying and characterising trapped lee waves using deep learning techniques. *Quarterly Journal of the Royal Meteorological Society*, 150, 213–231. Available from: https://doi.org/10.1002/qj.4592

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S. et al. (2011) The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597. Available from: https://doi.org/10.1002/qj.828

Durran, D. (2003) Lee waves and mountain waves. In: Holton, J.R., Curry, J.A. & Pyle, J.A. (Eds.) *Encyclopedia of atmospheric sciences*. Burlington: Elsevier, pp. 1161–1169.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B. et al. (2018) Recent advances in convolutional neural networks. *Pattern Recognition*, 77, 354–377. Available from: https://doi.org/10.1016/j.patcog.2017.10.013

Hindley, N.P., Smith, N.D., Wright, C.J., Rees, D.A.S. & Mitchell, N.J. (2016) A two-dimensional Stockwell transform for gravity wave analysis of AIRS measurements. *Atmospheric Measurement Techniques*, 9, 2545–2565. Available from: https://doi.org/10.5194/amt-9-2545-2016

Howard, J. & Gugger, S. (2020) Fastai: a layered API for deep learning. *Information*, 11, 108. Available from: https://doi.org/10.3390/info11020108

Justin, A.D., Willingham, C., McGovern, A. & Allen, J.T. (2023) Toward Operational Real-time Identification of Frontal Boundaries Using Machine Learning. *Artificial Intelligence for the Earth Systems*, 2, 1–56. Available from: https://doi.org/10.1175/aies-d-22-0052.1

Kendon, E., Short, C., Pope, J., Chan, S., Wilkinson, J., Tucker, S. et al. (2021) Update to UKCP Local (2.2km) projections, Science Report, Met Office Hadley Centre. Technical report July, Met Office, Exeter, Devon. https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/ukcp18_local_update_report_2021.pdf

Kotsiantis, S.B. (2013) Decision trees: a recent overview. *Artificial Intelligence Review*, 39, 261–283. Available from: https://doi.org/10.1007/s10462-011-9272-4

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F. et al. (2023) Learning skillful medium-range global weather forecasting. *Science*, 382, 1416–1421. Available from: https://doi.org/10.1126/science.adi2336

Lowe, J.A., Bernie, D., Bett, P.E., Bricheno, L., Brown, S.J., Calvert, D. et al. (2018) *UKCP18 science overview report*. Exeter, UK: Met Office Hadley Centre, pp. 1–73. Available from: https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Overview-report.pdf

Lundberg, S.M. & Lee, S.-I. (2017) A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. et al. (Eds.) *Advances in Neural Information Processing Systems*, Vol. 30. New York: Curran Associates, Inc, pp. 4765–4774.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B. et al. (2020) From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2, 56–67. Available from: https://doi.org/10.1038/s42256-019-0138-9

Manning, C., Kendon, E.J., Fowler, H.J. & Roberts, N.M. (2023) Projected increase in windstorm severity and contribution from sting jets over the UK and Ireland. *Weather and Climate Extremes*, 40, 100562. Available from: https://doi.org/10.1016/j.wace.2023.100562

Murphy, J., Harris, G., Sexton, D., Kendon, E., Bett, P., Clark, R. et al. (2019) UKCP18 land projections: science report. Technical report March 2019, Met Office, Exeter. https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Land-report.pdf

Neal, R., Fereday, D., Crocker, R. & Comer, R.E. (2016) A flexible approach to defining weather patterns and their application in weather forecasting over Europe. *Meteorological Applications*, 23, 389–400. Available from: https://doi.org/10.1002/met.1563

Niebler, S., Miltenberger, A., Schmidt, B. & Spichtinger, P. (2022) Automated detection and classification of synoptic-scale fronts from atmospheric data grids. *Weather and Climate Dynamics*, 3, 113–137. Available from: https://doi.org/10.5194/wcd-3-113-2022

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. et al. (2012) Scikit-learn: machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830. Available from: https://doi.org/10.48550/arXiv.1201.0490

Perks, R.J., Bernie, D., Lowe, J. & Neal, R. (2023) The influence of future weather pattern changes and projected sea-level rise on coastal flood impacts around the UK. *Climatic Change*, 176, 1–21. Available from: https://doi.org/10.1007/s10584-023-03496-2

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. et al. (2019) Deep learning and process understanding for data-driven Earth system science. *Nature*, 566, 195–204. Available from: https://doi.org/10.1038/s41586-019-0912-1

Ruff, F. & Ólafsson, H. (2019) Analysis of observed rapid increases in surface wind speed. *Quarterly Journal of the Royal Meteorological Society*, 145, 28–39. Available from: https://doi.org/10.1002/qj.3377

Scorer, R.S. (1949) Theory of waves in the lee of mountains. *Quarterly Journal of the Royal Meteorological Society*, 75, 41–56. Available from: https://doi.org/10.1002/qj.49707532308

Sheridan, P., Vosper, S. & Brown, P. (2017) Mountain Waves in High Resolution Forecast Models: Automated Diagnostics of Wave Severity and Impact on Surface Winds. *Atmosphere*, 8, 24. Available from: https://doi.org/10.3390/atmos8010024

Shutts, G. (1992) Observations and Numerical Model Simulation of a Partialy Trapped Lee Wave over the Welsh Mountains. *Monthly Weather Review*, 120, 2056–2066. Available from: https://doi.org/10.1175/1520-0493(1992)120<2056:OANMSO>2.0.CO;2

Stockwell, R., Mansinha, L. & Lowe, R. (1996) Localization of the complex spectrum: the S transform. *IEEE Transactions on Signal Processing*, 44, 998–1001. Available from: https://doi.org/10.1109/78.492555

Tang, Y., Lean, H.W. & Bornemann, J. (2013) The benefits of the Met Office variable resolution NWP model for forecasting convection. *Meteorological Applications*, 20, 417–426. Available from: https://doi.org/10.1002/met.1300

Vosper, S.B. (2004) Inversion effects on mountain lee waves. *Quarterly Journal of the Royal Meteorological Society*, 130, 1723–1748. Available from: https://doi.org/10.1256/qj.03.63

Vosper, S.B., Wells, H., Sinclair, J.A. & Sheridan, P.F. (2013) A climatology of lee waves over the UK derived from model forecasts. *Meteorological Applications*, 20, 466–481. Available from: https://doi.org/10.1002/met.1311

Washington, R. & Palmer, M. (1999) The North Atlantic Oscillation. *Geography Review*, 13, 2–5. Available from: https://doi.org/10.1017/cbo9781139034135.008

Wildmann, N., Eckert, R., Dörnbrack, A., Gisinger, S., Rapp, M., Ohlmann, K. et al. (2021) In situ measurements of wind and turbulence by a motor glider in the andes. *Journal of Atmospheric and Oceanic Technology*, 38, 921–935. Available from: https://doi.org/10.1175/JTECH-D-20-0137.1

Wilkinson, J.M. & Neal, R. (2021) Exploring relationships between weather patterns and observed lightning activity for Britain and Ireland. *Quarterly Journal of the Royal Meteorological Society*, 147, 2772–2795. Available from: https://doi.org/10.1002/qj.4099

World Meteorological Organization. (1993) *Handbook of meteorological forecasting for soaring flight. Technical report WMO – No. 495*. Geneva: World Meteorological Organization.

Worthington, R.M. (2006) Diurnal variation of mountain waves. *Annales Geophysicae*, 24, 2891–2900. Available from: https://doi.org/10.5194/angeo-24-2891-2006

How to cite this article: Coney, J., Ross, A.N., Denby, L., Wang, H., Vosper, S., van Niekerk, A. *et al*. (2025) A climatology of trapped lee waves over Britain and Ireland obtained using deep learning on high-resolution model output. *Quarterly Journal of the Royal Meteorological Society*, e5037. Available from: https://doi.org/10.1002/qj.5037