



# Forecasting seasonal to sub-seasonal rainfall in Great Britain using convolutional-neural networks

Andrew Paul Barnes<sup>1</sup> · Nick McCullen<sup>1</sup> · Thomas Rodding Kjeldsen<sup>1</sup>

Received: 3 June 2021 / Accepted: 7 October 2022 / Published online: 19 November 2022  
© The Author(s) 2022

## Abstract

Traditional weather forecasting approaches use various numerical simulations and empirical models to produce a gridded estimate of rainfall, often spanning multiple regions but struggling to capture extreme events. The approach presented here combines modern meteorological forecasts from the ECMWF SEAS5 seasonal forecasts with convolutional neural networks (CNNs) to improve the forecasting of total monthly regional rainfall across Great Britain. The CNN is trained using mean sea-level pressure and 2-m air temperature forecasts from the ECMWF C3S service using three lead-times: 1 month, 3 months and 6 months. The training is supervised using the equivalent benchmark rainfall data provided by the CEH-GEAR (Centre for Ecology and Hydrology, gridded estimates of areal rainfall). Comparing the CNN to the ECMWF predictions shows the CNN out-performs the ECMWF across all three lead times. This is done using an unseen validation dataset and based on the root mean square error (RMSE) between the predicted rainfall values for each region and benchmark values from the CEH-GEAR dataset. The largest improvement is at a 1-month lead time where the CNN model scores a RMSE 6.89 mm lower than the ECMWF. However, these differences are exacerbated at the extremes with the CNN producing, at a 1-month lead time, RMSEs which are 28.19 mm lower than the corresponding predictions from the ECMWF. Following this, a sensitivity analysis shows the CNN model predicts increased rainfall values in the presence of a low sea-level pressure anomaly around Iceland, followed by a high sea-level pressure anomaly south of Greenland.

## 1 Introduction

Rainfall variation plays a significant role in the long-term investment strategy and typically takes two forms across Great Britain. First is the effect of droughts, or prolonged periods of lower-than expected rainfall, such as that following the wet summer of 2012, which resulted in farmers struggling to grow and harvest crops and a heightened risk of wildfires (Kendon et al. 2013). The second effect is flooding from high levels of precipitation, with two prominent examples of this being storms Ciara and Dennis from February 2020, which resulted in over £300 million worth of damage and costing the lives of five people (Griffith 2020). Therefore, it is important to increase our understanding of the processes controlling the temporal and spatial occurrences and magnitudes to help build societal resilience.

Previous research has demonstrated that the magnitude and spatial distribution of rainfall are tied to the patterns of atmospheric circulation (Utsumi et al. 2017; Baker et al. 2018; Gimeno et al. 2020). This is especially true during winter, when the precipitation variability over the British Isles is heavily influenced by the North Atlantic Oscillation (NAO), an index which is characterised by the pressure difference between the Azores and Iceland (Brown 2018). The NAO has been found to exert a strong influence on the storm track and strength of extra-tropical cyclones crossing the North Atlantic Ocean (Huntingford et al. 2014). These storms which originate over the North Atlantic are the main contributors to European rainfall (Gimeno et al. 2020). Other studies have identified links between NAO and flood levels (e.g. Hannaford and Marsh 2008; Macdonald et al. 2010).

Neal et al. (2016) created two classification models to investigate the synoptic atmospheric conditions leading to dry and wet periods, respectively, across the entire UK. The classification models grouped daily atmospheric conditions into either one of thirty or one of eight types depending on which model was being used. These models were developed by using daily sea-level pressure patterns across northern

✉ Andrew Paul Barnes  
apb41@bath.ac.uk

<sup>1</sup> Department of Architecture and Civil Engineering,  
University of Bath, Bath, UK

Europe and the North Atlantic and clustering them using a simulated annealing variant of the k-means clustering algorithm. Neal's models use the mean sea-level pressure anomalies across the North Atlantic to classify every day into one of thirty different circulation types. Subsequently, the set of 30 circulation types was reduced to a smaller set consisting of eight circulation types, the most prominent of which represent the positive and negative phases of the NAO. Using these eight patterns, Richardson et al. (2017) were able to show how certain patterns such as a negative NAO with a blocking pattern or anticyclonic conditions over the UK result in drier conditions than average. Furthermore, Richardson et al. (2017) also showed how a positive NAO pattern or a strong cyclonic pattern to the south west of the UK can result in wetter than average conditions. Similarly, Ummenhofer et al. (2017) used clustering techniques to group the precipitation rates across Europe into five types. The average sea-level pressure anomalies for each of these types were grouped together and showed strong relationships between the dominant pressure systems over the Arctic and Europe to the precipitation variability over the UK. None of the studies listed above appear to include temperature as a covariate despite the key relationship between temperature and rainfall via the Clausius-Clepyeron (CC) relation which states a warmer atmosphere can hold more water than a cooler one (Blenkinsop et al. 2015). The relationship between temperature and rainfall via the CC relation is evident but varies across a spatial domain even as small as the UK (Blenkinsop et al. 2015). Further studies also show a stark contrast in synoptic temperature conditions related to heavy or intense rainfall events (Allan et al. 2020).

Baker et al. (2018) used several linear regression models to explain up to 76% of regional rainfall variability in the UK using a series of selected sea-level pressure metrics. Richardson et al. (2020) further expanded the work on the original circulation pattern analysis of Richardson et al. (2017), to show that weather patterns can be used to inform both medium-range precipitation forecasting, highlighting that the use of the weather patterns in forecasting increased the accuracy of the drought forecast. Recently there has been a move towards the use of neural networks to predict future rainfall. A review by Pham et al. (2020) found that neural networks can generally predict daily and sub-daily rainfall to within 10 mm or less. Kumar et al. (2019) used a neural network to predict the monthly rainfall totals across several hydrologically homogeneous regions in India. However, the greatest success was reported by Haidar and Verma (2018), who developed a one-dimensional convolutional neural network (CNN) using climate variables (including, but not limited to, max temperature, min temperature, southern oscillation index, dipole mode index and interdecadal pacific oscillation) to predict monthly rainfall total in eastern Australia. They found that the resulting model showed

higher predictive performance than the recently released Australian Community Climate and Earth System (Hudson et al., 2017). Recently, CNNs have been shown as capable at predicting gridded precipitation; first, Larraondo et al. (2019) use geopotential height fields across the North Atlantic to predict total 3-hourly precipitation across Western Europe presenting more accurate results than alternative traditional machine learning approaches such as regression. In contrast, Rasp and Thuerey (2021) used a large number of input variables including geopotential height, temperature, wind speeds, specific humidity and 6-h accumulated precipitation to predict gridded rainfall estimates with cells  $5.625^\circ$ . Although both studies present promising results, neither provided a true sub-seasonal forecast. An image-based CNN would allow the interpretation of images, such as those used by Neal et al. (2016) but using the powerful inference offered by the CNN architecture shown by Haidar and Verma (2018), Larraondo et al. (2019) and Rasp and Thuerey (2021).

The present work fills this gap by producing a novel sub-seasonal rainfall forecasting methodology which uses image-based convolutional neural networks along with the meteorological forecasts from the ECMWF. The convolutional neural networks used are a novel deep learning approach, which combine both high-resolution sea-level pressure and 2-m air temperature forecast patterns across the North Atlantic to produce regional, monthly rainfall predictions. These models are then compared to the ECMWF SEAS5 precipitation forecasts (Johnson et al. 2019) using three different lead times (1 month, 3 months, and 6 months). To begin with, this study describes the datasets and pre-processing carried out (Section 2), before introducing CNNs in Section 3, detailing the architecture and training progress of the selected model in Section 3. The forecasts from the CNN are then compared with those from the SEAS5 system (Section 4), before making concluding remarks regarding the potential for CNNs to be used to predict regional rainfall.

## 2 Data

Training the CNNs requires three key datasets: (1) the benchmark (observed) regional precipitation, (2) the forecasted mean-sea level pressure and (3) 2 m air temperature (2AT) patterns across the North Atlantic. Forecast precipitation is then also required to allow a comparison between the accuracy of the CNN model and that of the ECMWF SEAS5 model. This section begins by describing the extraction of the observed and forecasted regional rainfall before describing the process of extracting forecasted North Atlantic mean sea-level pressure (MSLP) and North Atlantic 2AT data. Finally, this section concludes by describing the separation of the dataset into training, testing and validation.

## 2.1 Regional precipitation

Regional, series of observed monthly precipitation totals (mm) covering the land surface of Great Britain are extracted from the CEH-GEAR (Tanguy et al. 2019) and the forecasted precipitation totals (mm) extracted from the ECMWF SEA5 (Johnson et al. 2019). The CEH-GEAR dataset represents the benchmark precipitation and is used to train the CNN model. The predictions made by the CNN are then compared to the precipitation hindcasts made by the ECMWF SEA5 service. Both datasets are used to produce average cumulative monthly rainfall between 1993 and 2017 for the 13 regions administrative regions of Great Britain (Fig. 1).

### 2.1.1 Benchmark regional rainfall

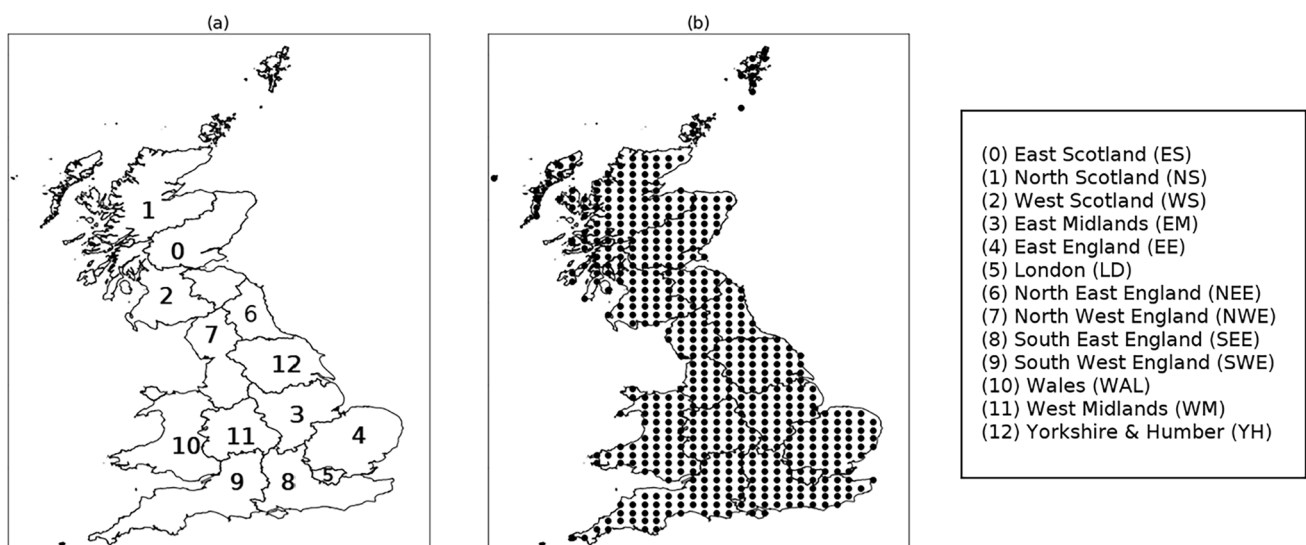
The CEH-GEAR dataset is a daily-mean rainfall dataset provided on a  $1 \text{ km} \times 1 \text{ km}$  grid covering the British Isles. The dataset spans the years 1950 to 2017; however, only data from 1993 to 2017 were used in this study due to the more limited range of the ECMWF hindcasts dataset as described in the Section 2.1.2. First, the daily CEH-GEAR data were aggregated to monthly values (mm). Next, the monthly rainfall data were further aggregated spatially to represent rainfall in each of the 13 regions. This was achieved by defining 598 points at 20 km intervals across the UK as shown in Fig. 1a. For each region, a representative monthly rainfall value is represented by the average monthly cumulative total of all points within the region's boundaries.

### 2.1.2 ECMWF SEA5 forecast regional rainfall

Forecast of bias-corrected rainfall data provided by the ECMWF SEA5 model was retrieved through the for three different lead times, 1 month, 3 months, and 6 months. As the ECMWF SEA5 results are provided as an ensemble of 52 model realisations, the ensemble mean is used. Rainfall data for each lead time variant is provided in cumulative monthly values (mm) at a global resolution of  $1^\circ \times 1^\circ$  cells; the data used in this study is bound between  $[100^\circ\text{W}, 70^\circ\text{N}]$  and  $[20^\circ\text{E}, 10^\circ\text{N}]$ . This data is then aggregated for each region to produce a total monthly rainfall value for each region for the months available. To do this for a given region and month, the month's rainfall data is first retrieved and is represented by a matrix of  $1^\circ \times 1^\circ$  cells. Next, each point within the selected region takes the value of the CEH-GEAR cell which contains it. Finally, the values for each point can be averaged to produce a total monthly rainfall value for the given region during the given month. Completing this operation for each month and each region produces a rainfall data matrix of size  $[13, 288]$  ( $[N\text{-Regions}, N\text{-Months}]$ ).

## 2.2 Meteorological data

Similarly to the rainfall data discussed in Section 2.1.2, the large-scale atmospheric dataset used in this study is provided at a  $1^\circ \times 1^\circ$  resolution and is obtained from the ECMWF SEA5 system and provided bias-corrected (Johnson et al. 2019). The variables of interest are the bias-corrected MSLP and bias-corrected 2AT. Both variables are provided globally and are forecasted using multiple lead times, with the MSLP provided in units of Pascal and the 2AT provided in Kelvin, as discussed above the ensemble mean is taken for



**Fig. 1** The 13 regions of Great Britain (a) and the corresponding 20 km by 20 km points (b) used to aggregate monthly rainfall totals (mm)

both the MSLP and 2AT patterns. As this study is focusing on the synoptic conditions in the North Atlantic Ocean, a subset of the data between  $-100^{\circ}$ – $20^{\circ}$  longitude and  $10^{\circ}$ – $70^{\circ}$  latitude was extracted for each of the three chosen lead times (1 month, 3 months and 6 months). For each lead time, each month is then represented by two matrices of size  $121 \times 61$ , one for the 2AT and the second for the MSLP with each entry representing a  $1^{\circ} \times 1^{\circ}$  spatial cell. The resulting dataset consists of 288 matrices for AT and 288 matrices for MSLP, one for each month covering the 24-year study period (1993–2017).

### 2.3 Data separation

The final step in data preparation is to distinguish the data used for training, testing and validation. The training data will be used to train and optimise the network, the testing data will be used in a hold-out testing scheme to ensure that the models are not overfitting, and finally, the validation data will be used to evaluate model performance after development. To begin, the validation dataset is set to contain the data from the years 2013 to 2017; then, the remaining months are split randomly between training (70%) and testing (30%).

## 3 Forecasting method

Three convolutional neural networks (defined below) were developed to forecast the monthly regional rainfall values across Great Britain using three different lead times; the

inputs and outputs of each CNN are shown in Fig. 2. This section introduces convolutional neural networks, their components and the architecture of the network followed by an outline of the training method including details on the data standardisation procedure used prior to training. CNNs are selected due to their capabilities of interpreting image data where traditional neural networks would fail, covering a wide variety of applications from tumour identification (Yang et al. 2019) to video classification (Karpathy et al. 2014).

### 3.1 Network structure

Convolutional neural networks (CNNs) are a type of neural network that specialise in interpreting data that are in image form, which in the present case are images consisting of two matrices: the 2AT and the MSLP patterns. The resulting input matrix for each month is a 3D matrix of size  $[2 \times 121 \times 61]$ , the first  $[121 \times 61]$  slice containing the MSLP and the second containing the 2AT. During training, this matrix is fed through the layers in the CNN with each layer performing a unique matrix transformation; the layers used in the models for this study are detailed below.

The name (CNN) is derived from the networks' use of convolutional (i.e. a layer which performs many different operations on the input) layers. An example convolutional layer is provided in Fig. 3. Here, the layer accepts a  $[3 \times 3]$  matrix as its input matrix; it then passes a predefined kernel (mask) over the image in increments (also known as the step size) of one pixel. At each step, the kernel is multiplied by the pixels underneath it, then summing the resulting values

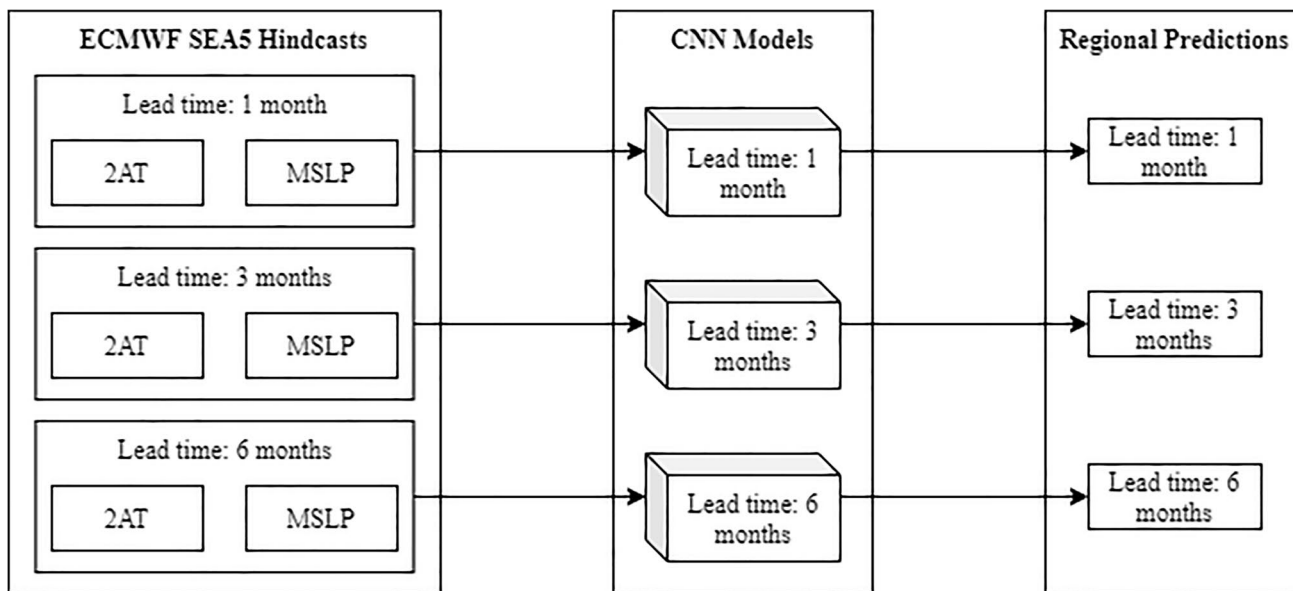
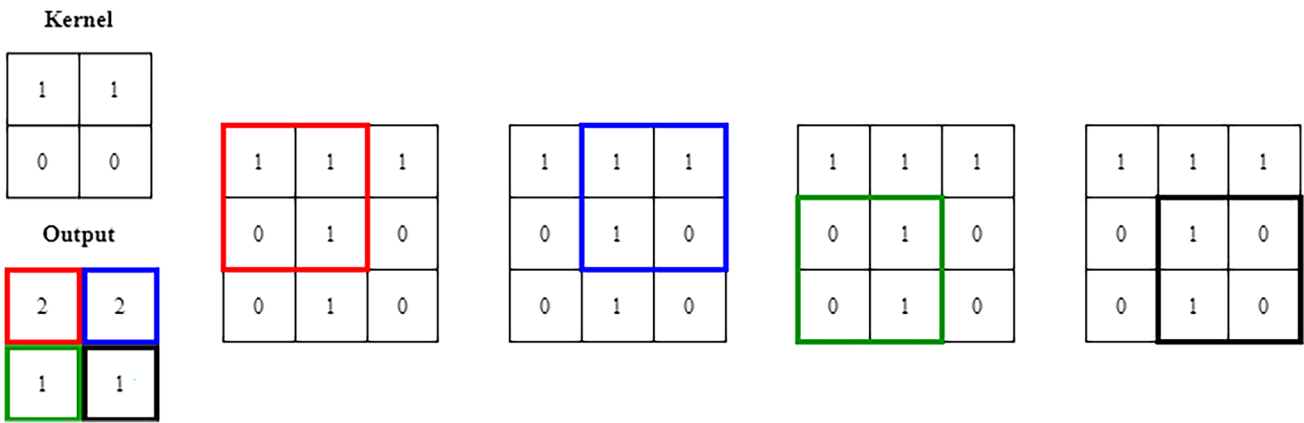


Fig. 2 The three forecast models showing their inputs and outputs

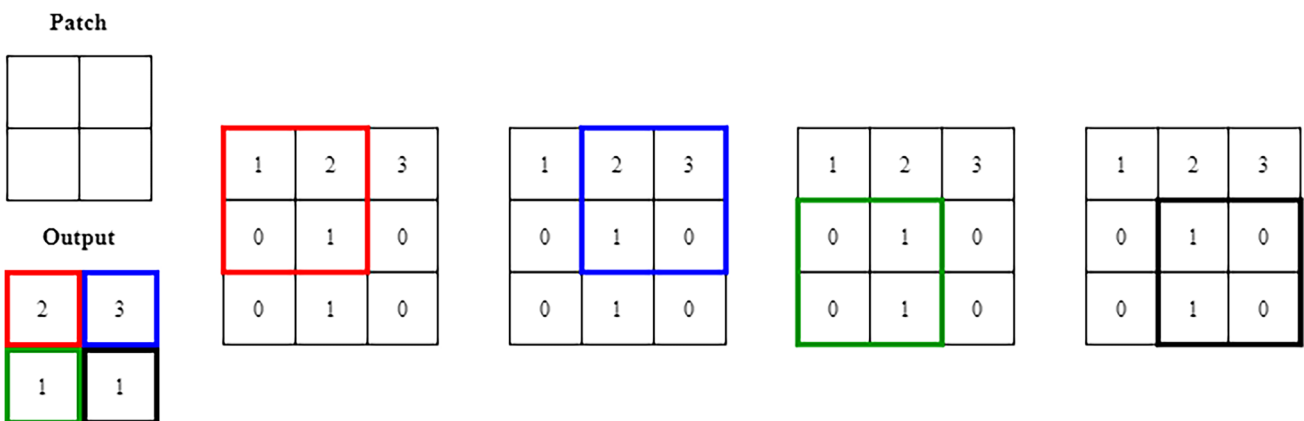


**Fig. 3** A convolution applied to a 3-by-3 image using a 2-by-2 kernel. The kernel is moved over the image by one pixel at a time, calculating the value of the subsection by multiplying the kernel by the pixels highlighted and summing the result

to produce the output of this step. The outputs of each step are put back together to form a new lower-resolution image. In this case, the kernel (defined here as a  $2 \times 2$  block with a row of two 1s above a row of two 0s, indicating the kernel is looking for a horizontal line) has identified a horizontal line at the top of the image and no horizontal lines below this. This can be seen in Fig. 3 where a 2 is produced for both the top left and top right hand corners indicating that both of these contain some or all of a horizontal line. In practice, it is common to use multiple kernels at each layer such that each layer can capture many different features. These result in an output matrix larger than that of the input matrix; for example, if three kernels of size  $[2 \times 2]$  were used on a  $[3 \times 3]$  matrix input, the resulting output would be  $3 \times [2 \times 2]$  (three kernels of size  $[2 \times 2]$ ).

Using multiple kernels in a single layer then presents a problem of complexity, increasing the size of the data as it moves through the convolutional layer further increases the computational resources needed for processing. To

counteract this, the implementation of max-pooling layers are used, which also reduce the dimensionality of the data. Max-pooling layers look through a kernel but instead of multiplying the pixels under it; the patch simply takes the maximum pixel value and uses this to represent its output. This process is illustrated in Fig. 4 using a variation of the  $[3 \times 3]$  matrix present in Fig. 3. As the max pooling layers generally follow a convolutional layer the input, they receive is a multi-dimensional matrix, in the example described in the above paragraph the output provides  $3 \times [2 \times 2]$  matrices (one  $[2 \times 2]$  for each kernel). Each kernel's output is processed independently, such that the number of matrices outputted by max pooling layers stays the same as the number which are given to them. However, these matrices will have reduced dimensions. Take for example an output from a convolutional layer of  $5 \times [3 \times 3]$  matrices; feeding this into a max pooling layer as shown in Fig. 4 will produce  $5 \times [2 \times 2]$  matrices.



**Fig. 4** A max-pooling operation applied to a 3-by-3 image using a patch of size 2-by-2. As the patch moves across the image, it extracts the maximum pixel value at each timestep



Max-pooling layers enable the CNN to reduce the dimensionality of the data. Relating this back to the datasets described in Section 2, the input is a 3D matrix ([variables, latitude, longitude]=[2, 61, 121]), but the output is a vector of length 13 (a monthly rainfall value for each of the 13 regions), so a final set of layers is needed to convert the multi-dimensional outputs of the convolutional/max-pooling layers to a 1D matrix. To do this, a linear layer is used; an example of which is shown in Fig. 5. The linear layer consists of several nodes each of which is connected by a weighted edge to every pixel in the output of the layer before it (also referred to as fully connected layers). The output ( $y$ ) for all nodes in the linear layer can then be calculated as

$$y = xW^T + b, \tag{1}$$

where  $x$  is an array containing the output of the previous layer and the weights matrix  $W^T$  is of size  $[|x| \times |y|]$ , and  $b$  is a randomly initialised bias term (of size  $[|y| \times 1]$ ). The resulting vector  $y$  is then passed through an activation function,

which in this case is a rectified linear unit function on each element; this activation function applies an absolute function to the values, ensuring that all output values in  $y$  are positive.

The architecture of the CNNs produced for this experiment is shown in Fig. 6. It consists of two convolutional layers, two max pooling layers and two linear layers. Figure 6 also shows how the size of the data moving through the network is changing; in the first convolutional layer, for example, the original input image is split into  $16 \times [30 \times 60]$  matrices. This indicates that in the first convolutional layer, there are 16 randomly generated kernels (all kernels used are of size (2, 2)); the kernels are randomly generated to enable to identification of features which may not have been previously known. An activation layer then follows each convolutional layer using the ReLU function ( $\max(x, 0)$ ). These are then reduced in size through the first max-pooling layer before being fed into the second convolutional and max-pooling layers. Following these, the output is passed through two linear layers, the first containing 100 neurons and the second containing 13. This final linear layer of 13 neurons

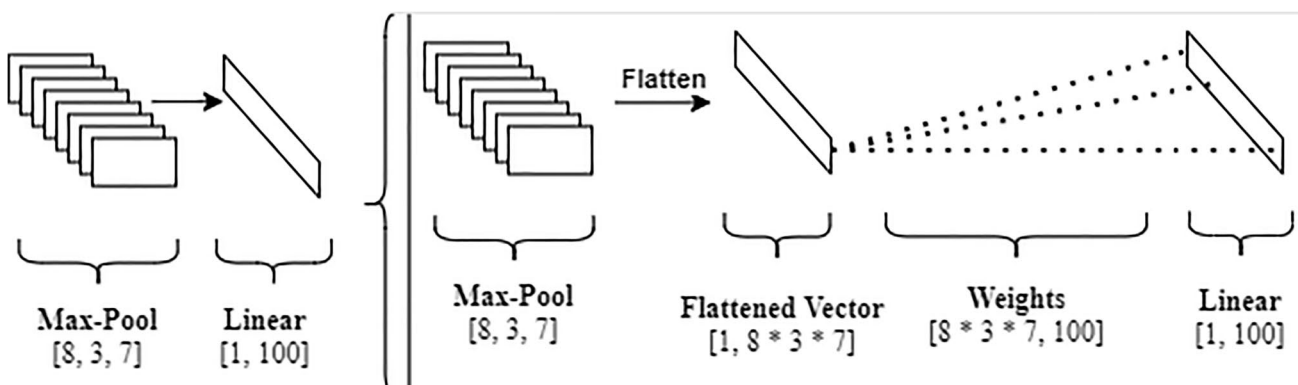


Fig. 5 A linear layer example following a max pooling layer. This example shows a max pooling layer with output dimensions [8, 3, 7] feeding into a linear layer of size [1, 100]. The max-pooling output is first flattened into a 1D vector. Each value in this vector is then con-

nected to every node in the linear layer via a weighted edge which results in a weights matrix of size  $[8 \times 3 \times 7, 100]$ . The linear layer then uses these weights and the flattened max-pooling vector to calculate the layer's output as described by Eq. 2

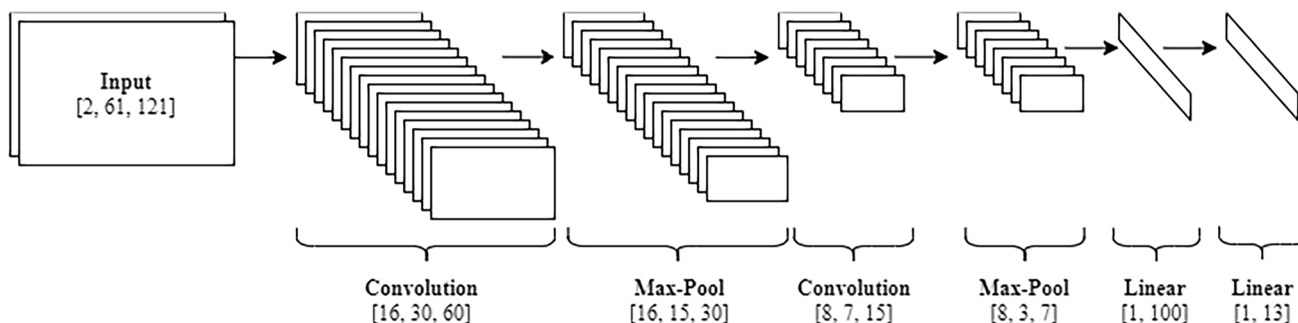


Fig. 6 Architecture of the convolutional neural network model. Each layer is provided with the number of dimensions of its output matrix. For example, the first convolutional layer has an output size of [16,

30, 60]; this also indicates the number of filters used in the layer (16). These are followed by max-pooling layers and finally two dense linear layers

outputs the monthly rainfall predictions for each of the 13 regions.

### 3.2 Training

As two separate variables with different scales and units are used for the input, they are standardised separately, such that both the 2AT matrix and mean sea level pressure were standardised by the mean and standard deviation of each of their individual matrix elements. Once standardised, these images were recombined into the  $[2 \times 61 \times 121]$  matrices required for training. Next, to reduce the potential for overfitting, the data are split randomly into two groups; a training and a validation dataset. The training dataset was used for training the weights of the CNN, whereas the validation data was used to compare the model predictions to those of the ECMWF system. Hence, of the data available, the months between 2013 and 2017 were retained for validation purposes and the remaining months between 1993 and 2012 were used for training. To train the weights of the CNN, the Adam optimisation (Kingma and Ba 2015) method is used along with the root mean-squared error (MSE) defined as

$$RMSE(x, y) = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}} \tag{2}$$

Finally, to ensure that the models are optimal, a stopping condition is used through a hold-out testing scheme; on each epoch, the testing data is used to calculate a testing loss. If after 5 epochs the testing loss does not reduce, the training scheme will revert the parameters to the epoch with the lowest testing error.

## 4 Validation results

This section details first a comparison between the forecasting capability of the trained CNN model developed in this study and the existing ECMWF model using only the validation dataset which was not used during training and then breaks down the CNN model to identify what features have been identified in the large-scale data as being most influential for informing the predictions (Section 4.2).

### 4.1 Model comparisons

To ensure a fair comparison, only validation data is used to compare the ECMWF and the CNN models. Figure 7 shows both the CNN and ECMWF predictions against the benchmark rainfall for the validation dataset. For all three lead times, the CNN shows wider spread in its predictions, especially when predicting the lower rainfall values. Decreasing the lead time from 6 to 3 months appears to make little

difference to the distribution of the predictions. However, at a 1 month lead time, the ECMWF predictions present a smaller range of predicted rainfall which comes at the cost of constant over prediction of the lower rainfall values and under-prediction of the higher rainfall values. Contrary to this, the CNN model yields a larger variation in the predictions of very low and very high rainfall amounts but appear to be less biased.

As shown in Table 1, the 3-month variants of both models perform the best, with the lowest RMSE scores. This trend is found in both the validation RMSE scores and the overall dataset RMSE scores. In all lead times and in both dataset variants, the CNN outperforms the ECMWF model; this is to be expected in the overall dataset (training, testing and validation) as the CNN was specifically trained using the data being evaluated. However, this does not explain the increased accuracy the CNN has in the validation dataset which was not available during training.

Further investigating the bias in the models towards the central range of rainfall values, Fig. 8 shows a summation of the residuals (differences between the benchmark and predicted rainfall) against the benchmark rainfall values for a 3-month lead time. This illustrates the strong under prediction of the higher rainfall values in both the CNN and ECMWF predictions; moreover, this highlights a weaker bias in the CNN model with approximately half the final cumulative residual compared to the ECMWF predictions. Following this, a weaker bias can be seen in the rainfall values leading up to 100 mm with a general trend of over prediction in both models, of which the CNNs bias appears stronger. A similar trend is seen for both 1-month and 6-month lead time predictions.

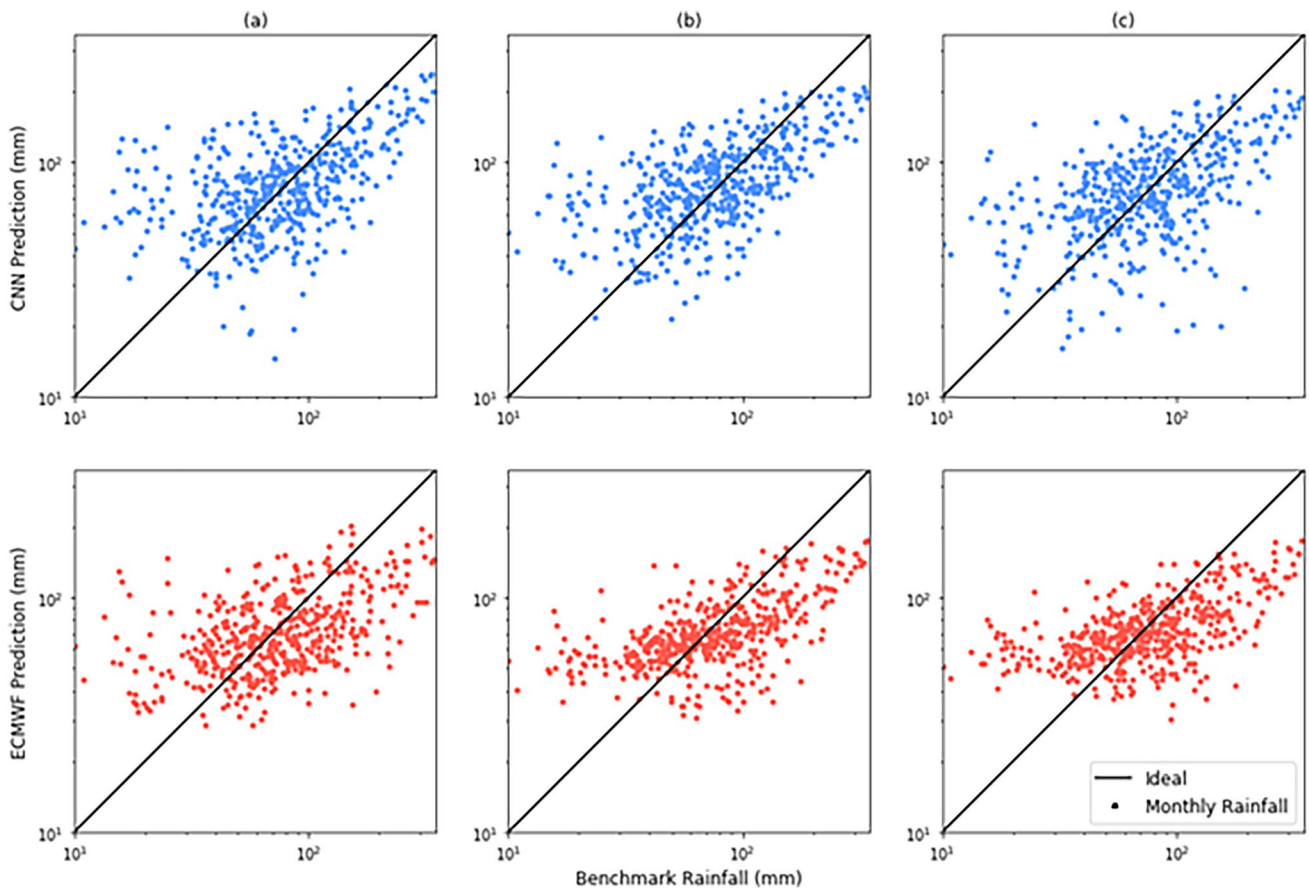
Identification of any regional or seasonal bias in the predictions is investigated based on proportional errors which are generated for each of the 13 regions across Great Britain and for each of the 12 months of the year. To do this, the mean-absolute error (MAE) is used to calculate the differences between the two series; the general equation for the MAE is defined as

$$MAE(x, y) = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \tag{3}$$

Subsequently, the MAE can then be used to calculate the proportional regional error (PRE) for a given region  $i$ :

$$PRE_i = \frac{MAE(p_i, \hat{p}_i)}{\bar{p}_i} \tag{4}$$

Here, the mean absolute error between the benchmark rainfall series ( $p_i$ ) and predicted rainfall series ( $\hat{p}_i$ ) is divided by the mean benchmark rainfall ( $\bar{p}_i$ ) for the given region  $i$ . This indicates how far away the predictions are based on the



**Fig. 7** Validation rainfall predictions against benchmark rainfall for both the CNN and ECMWF methods in all three lead times: **a** 1 month, **b** 3 months and **c** 6 months

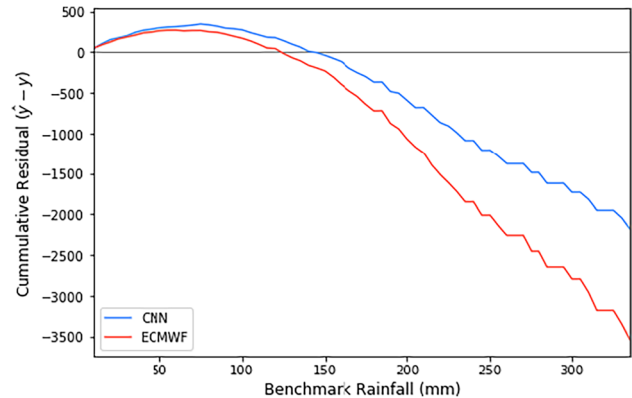
**Table 1** The RMSE scores for both the CNN and ECMWF models covering all three lead times given the validation dataset and the overall dataset are given

Model	Lead time	RMSE	
		Validation	Overall
CNN	1	46.47	38.09
	3	43.27	37.99
	6	48.53	39.02
ECMWF	1	53.36	47.50
	3	48.65	45.86
	6	49.60	52.26

region’s average rainfall value. Similarly, the proportional monthly error (PME) can be calculated as

$$PME_m = \frac{MAE(p_m, \hat{p}_m)}{\bar{p}_m} \tag{5}$$

As with Eq. 3, the difference between the benchmark series ( $p_m$ ) and predicted rainfall series ( $\hat{p}_m$ ) is divided

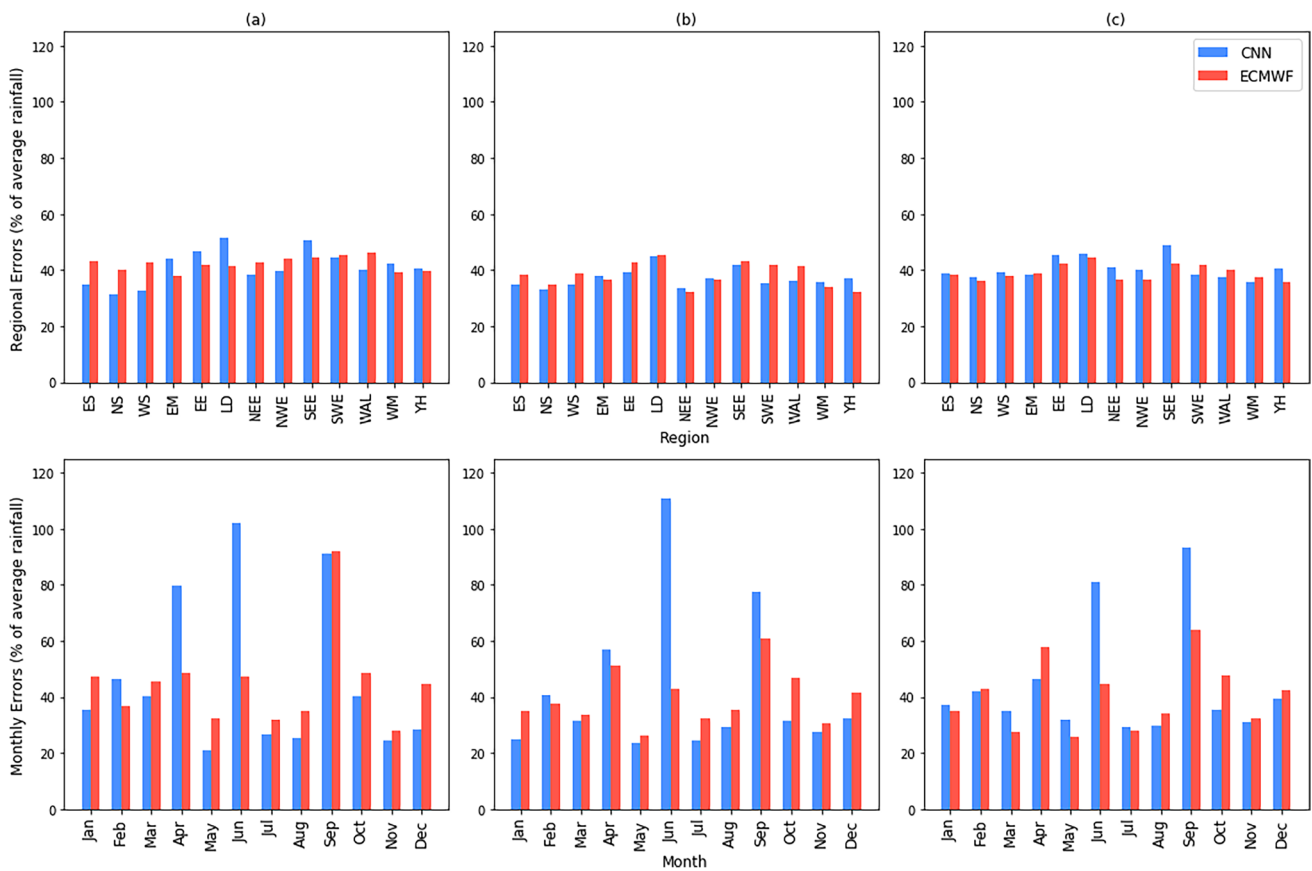


**Fig. 8** Cumulative residual of both methods using a three-month lead time against the actual regional rainfall

by the mean benchmark rainfall ( $\bar{p}_m$ ) for month  $m$  where  $m = 1, \dots, 12$ . Figure 9 (top row) shows the proportional error for each region and each month of the year.

The values of PREs are below 52% for all lead times and all regions; despite this, at a 1-month lead time, there





**Fig. 9** Regional and monthly rainfall areas as a proportion of the average rainfall for the given month or region. This is given for both methods and all three lead times: **a** 1 month, **b** 3 months and **c** 6 months

is a disparity between the CNN and the ECMWF predictions. The CNN errors at a 1-month lead time are highest for regions in the south east of Great Britain (London, South East England and East of England), whereas the ECMWF errors are higher for the west of Great Britain (Wales, South West England and North West England). This can be attributed to the known rainfall gradient across the UK, where western regions are likely to receive higher levels of rainfall and geographic effects produce a rain-shadow over the east (Mayes and Wheeler 2013). As described above and shown in Fig. 8, the ECMWF model produces larger errors for the higher rainfall values which are those occurring more in the west and north west regions, whereas the CNN provides more accurate predictions for the higher rainfall values which according to Fig. 9 comes at the cost of predicting east and south eastern rainfall. However, due to the different drivers of rainfall across the UK (Baker et al. 2018), this error could be representing the CNN identifying one rainfall driver and discounting the others.

The bottom row of Fig. 9 shows values of PME and highlights significant bias in the forecasted precipitation between each month of the year. First, the CNN across all three lead times produces high errors for April, June and September, whereas

the ECMWF forecasts have the largest proportional errors for April, September and October. However, the magnitude of these errors is different, with the CNN mis-predicting June rainfall by 100% of June's average rainfall, whereas the ECMWF forecasts only mis-predict June by 47.5%. September is shown to be difficult to predict with both models producing an error of 91.1% of the average September rainfall.

## 4.2 Sensitivity analysis

The benefit of using a convolutional neural network is that the weights between each layer can be used to identify the pixels of the input matrix which attribute the most to a given region's rainfall prediction. To do this, the integrated gradients method is employed (Sundararajan et al. 2017). Integrated gradients determine the importance of each pixel in the original image for, in this case, increasing the predicted rainfall value. For example, a negative attribution value indicates a pixel lowered the resulting rainfall prediction, whereas a positive attribution increases the resulting rainfall value. Figure 10 shows the normalised attribution of each MSLP (a) and 2AT (b) pixel for both North West England (top) and South East England (bottom).

**Fig. 10** Normalised attribution values for increasing rainfall predictions in both NW England (top) and SE England (bottom)

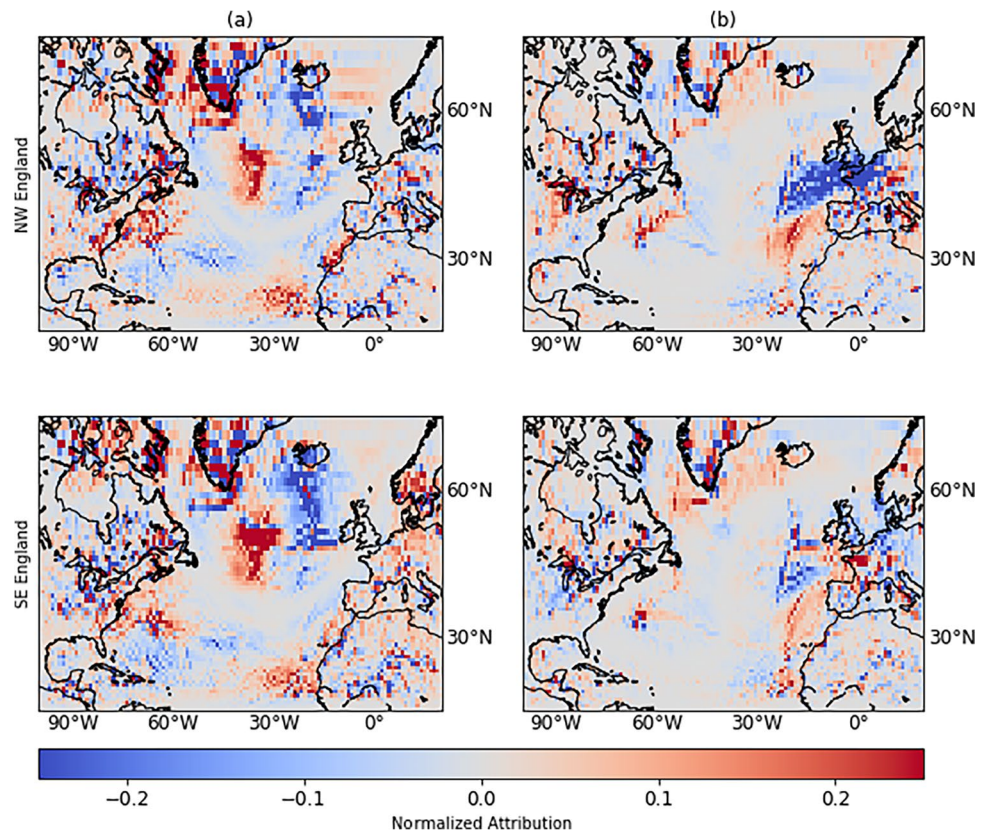


Figure 10a shows three dominant areas of interest for both NW and SE England; firstly, a strong negative attribution can be seen south west of Iceland, stretching towards the south west of the UK. This negative attribution will increase the magnitude of the rainfall prediction for both NW and SE England if the meteorological patterns contain a negative MSLP anomaly. Should the region instead contain a positive MSLP anomaly the rainfall values will decrease. In contrast, the second and third areas of interest are shown by the positive attribution values to the east of Newfoundland and surrounding Cape Verde, west of Africa. These regions of positive attribution will amplify the rainfall predictions if they contain positive MSLP anomalies and decrease the prediction if they contain negative MSLP anomalies.

The amplification of rainfall values due to a low-pressure anomaly to the west of the British Isles as shown in Fig. 10a agrees with the findings of previous studies (Richardson et al. 2017; Ummenhofer et al. 2017; Baker et al. 2018; Richardson et al. 2020). However, Fig. 10a also shows an extra region of higher pressure, south of Greenland, which further amplifies rainfall magnitude. This extra high-pressure anomaly could aid in the creation of a deep pressure gradient, increasing wind speeds between the centres and further enhancing mixing of the cold polar air from the North and the warm moist air from the south.

This mixing of cold and warm air is further shown by Fig. 10b which presents a combination of both positive and negative 2AT attributions around the UK. All lead times present a band of negative 2AT attributions stretching from the North Atlantic Ocean in a north easterly direction across the south of Great Britain towards Sweden and Norway. This band of negative 2AT is surrounded by mostly positive 2AT attributions and indicates a strong 2AT gradient in the area. However, for increasing the prediction for SE England, the attributions show a less distinct characterisation between the air masses indicating strong mixing of the cold and warm air masses.

Furthermore, the MSLP positive attributions around Cape Verde could indicate that the models are also enhancing rainfall when the MSLP anomaly differences between the north and mid-Atlantic Ocean are highest. This agrees with the findings of Baker et al. (2018), who also show that a pressure difference between a point between Scotland and Iceland and a point near Africa correlates strongly to rainfall in the UK. These MSLP differences appear to simulate the North Atlantic Oscillation index (NAO) which determines the MSLP difference between Iceland and the Azores. A heightened NAO is known to correspond to increased rainfall in the UK as highlighted by Brown (2018).

## 5 Conclusion

This paper has shown for the first time that convolutional neural networks (CNN) can be used as a tool for enhancing monthly, regional rainfall forecasting in the UK through the interpretation of forecasted mean sea-level pressure and 2-m air temperature patterns. Three CNN models were trained using monthly, regional rainfall from the CEH-GEAR dataset and MSLP and 2AT anomaly patterns from the ECMWF monthly hindcasts. Each model was trained using a different lead time—either 1 month, 3 months, or 6 months. A validation dataset was then used to compare the predicted regional rainfall values with those of the ECMWF precipitation hindcasts given at the same lead times. The CNN models were then analysed using an integrated gradients technique to explore how it made its predictions. The key results are as follows:

1. The CNNs provided more accurate regional monthly rainfall totals than the ECMWF SEAS5 model across all three lead times.
  - The CNN predicts the validation regional monthly rainfall totals with a lower RMSE than the ECMWF SEAS5 model for all three lead times. The RMSE improvements are as follows: 6.89 (1 month), 5.38 (3 months) and 1.07 (6 months).
  - Predicting the entire dataset including the training, testing and validation data series the CNN models continue to outperform the ECMWF SEAS5 predictions by the following RMSE differences for each lead time: 8.96 (1 month), 7.87 (3 months) and 13.24 (6 months).
  - The CNN model's residuals indicate the CNN has higher accuracies when predicting the heaviest rainfall events compared with the ECMWF SEAS5 model.
2. The CNN models show spatial and seasonal bias in its predictions.
  - The CNNs provide the most accurate results for northern and western regions of Great Britain but do not perform as well for the eastern and south-eastern regions.
  - The CNNs perform well for some months of the year (December, July, August, November and May) but produce very large errors for the months of June, September and April (over double the average rainfall for the month of June).
3. Sensitivity analysis was performed to identify how the CNNs were making their predictions, resulting in the following findings:

- A strong negative MSLP anomaly to the west of Great Britain is a key feature relating to increased rainfall predictions. However, a positive MSLP anomaly to the west of Great Britain will decrease the amount of rainfall predicted.
- A positive MSLP anomaly to the east of Newfoundland but to the west of the negative MSLP anomaly identified in point 1, and another positive MSLP anomaly over Cape Verde, will also enhance rainfall magnitudes over Great Britain. If these regions were to contain negative MSLP anomalies instead, the rainfall predictions would be reduced.
- A combination of positive and negative 2AT anomalies over the British Isles will also increase the monthly rainfall total predictions made by the CNN models.

While the study presented here was limited on the available data, future work could focus on the application of this forecasting method to provide higher-resolution forecast in both time and space dimensions. There is also now the opportunity to explore other meteorological variables using CNNs to identify where these variables may matter using an integrated gradients analysis. Another potential option is to modify the method presented in this paper to output a probabilistic forecast instead of a point-forecast which could form a future body of methods capable of improving both rainfall forecasting and our understanding of the processes which lead to rainfall.

**Author contribution** All three authors were involved with the conception of the idea. A.P.B extracted the data and developed both the code and models used in the analysis. The results were reviewed and analysed by A.P.B under the supervision of T.R.K and N.M. All authors were involved in the development of the manuscript.

**Funding** The first author's contributions were supported by the National Productivity Investment Fund from EPSRC, grant number EP/R512254/1.

**Data availability** All data used in this study is freely available. The Centre of Ecology and Hydrology's CEH-GEAR dataset is available from their portal (<https://catalogue.ceh.ac.uk/documents/ee9ab43d-a4fe-4e73-afd5-cd4fc4c82556>). The meteorological and gridded rainfall data extracted from the ECMWF's SEAS5 model is available through their Copernicus service (<https://cds.climate.copernicus.eu/cdsapp#!/dataset/seasonal-monthly-single-levels>).

**Code availability** Available on request.

### Declarations

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.



**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Allan RP, Blenkinsop S, Fowler HJ, Champion AJ (2020) Atmospheric precursors for intense summer rainfall over the United Kingdom. *Int J Climatol* 40:3849–3867. <https://doi.org/10.1002/joc.6431>
- Baker LH, Shaffrey LC, Scaife AA (2018) Improved seasonal prediction of UK regional precipitation using atmospheric circulation. *Int J Climatol* 38:e437–e453. <https://doi.org/10.1002/joc.5382>
- Blenkinsop S, Chan SC, Kendon EJ, Roberts NM, Fowler HJ (2015) Temperature influences on intense UK hourly precipitation and dependency on large-scale circulation. *Environ Res Lett* 10(5):054021. <https://doi.org/10.1088/1748-9326/10/5/054021>
- Brown SJ (2018) The drivers of variability in UK extreme rainfall. *Int J Climatol* 38(51):119–130. <https://doi.org/10.1002/joc.5356>
- Gimeno L, Vazquez M, Eiras-Barca J, Sori R, Stojanovic M, Algarra I, Nieto R, Ramos AM, Duran-Quesada AM, Dominguez F (2020) Recent progress on the sources of continental precipitation as revealed by moisture transport analysis. *Earth Sci Rev* 201. <https://doi.org/10.1016/j.earscirev.2019.103070>
- Griffith H (2020) A month of storms: Ciara, Dennis and atmospheric rivers. <https://hepex.inrae.fr/atm-rivers-ciara-dennis/>. Accessed 02 Jan 2022
- Haidar A, Verma B (2018) Monthly Rainfall forecasting using one-dimensional deep convolutional neural network. *IEEE Access* 6:69053–69063. <https://doi.org/10.1109/ACCESS.2018.2880044>
- Hannaford J, Marsh TJ (2008) High-flow and flood trends in a network of undisturbed catchments in the UK. *Int J Climatol* 28(10):1325–1338. <https://doi.org/10.1002/joc.1643>
- Hudson D, Alves O, Hendon HH, Lim E-P, Liu G, Luo J-J, MacLachlan C, Marshall AG, Shi L, Wang G, Wedd R, Young G, Zhao M, Zhou X (2017) ACCESS-S1 The new Bureau of Meteorology multi-week to seasonal prediction system. *J South Hemisphere Earth Syst Sci* 67(3):132–159. <https://doi.org/10.22499/3.6703.001>
- Huntingford C, Marsh T, Scaife AA, Kendon EJ, Hannaford J, Kay AL, Lockwood M, Prudhomme C, Reynard NS, Parry S, Lowe JA, Screen JA, Ward HC, Roberts M, Stott PA, Bell VA, Bailey M, Jenkins A, Legg T, ... Allen MR (2014) Potential influences on the United Kingdom's floods of winter 2013/14. *Nat Clim Change* 4(9):769–777. <https://doi.org/10.1038/nclimate2314>
- Johnson SJ, Stockdale TN, Ferranti L, Balmaseda MA, Molteni F, Magnusson L, Tietsche S et al (2019) SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Dev* 12(3):1087–1117
- Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L (2014) Large-scale video classification with convolutional neural networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, Columbus, OH, USA, pp 1725–1732. <https://doi.org/10.1109/CVPR.2014.223>
- Kendon M, Marsh T, Parry S (2013) The 2010–2012 drought in England and Wales. *Weather* 68(4):88–95. <https://doi.org/10.1002/wea.2101>
- Kingma DP, Ba JL (2015) Adam: a method for stochastic optimization. *Proceedings of the International Conference on Learning Representations*, San Diego, USA
- Kumar D, Singh A, Samui P, Jha RK (2019) Forecasting monthly precipitation using sequential modelling. *Hydrol Sci J* 64(6):690–700. <https://doi.org/10.1080/02626667.2019.1595624>
- Larraondo PR, Renzullo LJ, Inza I, Lozano JA (2019) A data-driven approach to precipitation parameterizations using convolutional encoder-decoder neural networks. <http://arxiv.org/abs/1903.10274>
- Macdonald N, Phillips ID, Mayle G (2010) Spatial and temporal variability of flood seasonality in Wales. *Hydrol Process* 24(13):1806–1820. <https://doi.org/10.1002/hyp.7618>
- Mayer J, Wheeler D (2013) Regional weather and climates of the British Isles - Part 1: Introduction. *Weather* 68:3–8. <https://doi.org/10.1002/wea.2041>
- Neal R, Fereday D, Crocker R, Comer RE (2016) A flexible approach to defining weather patterns and their application in weather forecasting over Europe. *Meteorol Appl* 23(3):389–400. <https://doi.org/10.1002/met.1563>
- Pham BT, Le LM, Le T, Bui KT, Le VM, Ly H, Prakash I (2020) Development of advanced artificial intelligence models for daily rainfall prediction. *Atmos Res* 237. <https://doi.org/10.1016/j.atmosres.2020.104845>
- Rasp, S. & Thuerey, N. (2021). Data-driven medium-range weather prediction with a Resnet pretrained on climate simulations: a new model for WeatherBench. *J Adv Model Earth Syst* 13:e2020MS002405. <https://doi.org/10.1029/2020MS002405>
- Richardson D, Fowler HJ, Kilsby CG, Neal R (2017) A new precipitation and drought climatology based on weather patterns. *Int J Climatol* <https://doi.org/10.1002/joc.5199>
- Richardson D, Fowler HJ, Kilsby CG, Neal R, Dankers R (2020) Improving sub-seasonal forecast skill of meteorological drought: a weather pattern approach. *Nat Hazards Earth Syst* 20:107–124
- Sundararajan M, Taly A, Yan Q (2017) Axiomatic attribution for deep networks. *Proceedings of the 34th International conference on machine learning*, 70:3319–3328. <https://doi.org/10.5555/3305890.3306024>
- Tanguy M, Dixon H, Prosdocimi I, Morri DG, Kelle V DJ (2019) Gridded estimates of daily and monthly areal rainfall for the United Kingdom (1890–2017) [CEH-GEAR]. NERC Environmental Information Data Centre. [dataset] <https://doi.org/10.5285/ee9ab43d-a4fe-4e73-afd5-cd4fc4c82556>
- Ummenhofer CC, Seo H, Kwon Y, Parfitt R, Brands S, Joyce TM (2017) Emerging European winter precipitation pattern linked to atmospheric circulation changes over the North Atlantic region in recent decades. *Geophys Res Lett* 44:8557–8566. <https://doi.org/10.1002/2017GL074188>
- Utsumi N, Kim H, Kanae S, Oki T (2017) Relative contributions of weather systems to mean and extreme global precipitation. *J Geophys Res Atmos* 122(1):152–167. <https://doi.org/10.1002/2016JD025222>
- Yang A, Yang X, Wu W, Liu, H, Zhuansun Y (2019) Research on feature extraction of tumor image based on convolutional neural network. *IEEE Access* 7:24204–24213. <https://doi.org/10.1109/ACCESS.2019.2897131>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.