

Video-Based Convolutional Neural Networks Forecasting for Rainfall Forecasting

Andrew P. Barnes¹, Thomas R. Kjeldsen, and Nick McCullen

Abstract—This study presents a new methodology for improving forecasts of current monthly, regional precipitation using video-based convolutional neural networks (CNNs). Using 13 administrative regions of Great Britain as a case study, three CNN architectures are trained for each region to forecast monthly rainfall totals given forecast mean sea-level pressure and 2-m air temperature videos from the MetOffice GloSEA5 model and a benchmark rainfall data. The forecasts generated by the CNN and the GloSEA5 precipitation forecasts are both compared directly against a benchmark rainfall dataset for each of the regions. Following this, the CNN models are combined with the GloSEA5 forecasts to generate a new ensemble for each region which is then compared with the benchmark rainfall. The results show that the trained CNNs produce errors similar to the GloSEA5 model with RMSEs of 63 mm (single frame), 44 mm (slow fusion), and 37 mm (early fusion) compared with the GloSEA5 error of 33 mm. Regional variability remained consistent throughout the compared models. However, the CNN models all outperform GloSEA5 in the prediction of extreme events. Furthermore, treating the forecasts as an ensemble results in errors of 32 mm (CNN ensemble) and 31 mm (post-processing ensemble), both of which improve on the independent GloSEA5 forecasts.

Index Terms—Forecasting, meteorology, neural networks, rainfall.

I. INTRODUCTION

MODERN datasets allow efficient interpretation of large-scale (synoptic) conditions and their relevance to the description and prediction of local rainfall; for example, Richardson *et al.* [11] used a set of 30 weather patterns to explain regional rainfall variation in the U.K. based on the mean sea-level pressure (MSLP) patterns across the North Atlantic. The weather patterns used by Richardson *et al.* [11] were originally defined by Neal *et al.* [12] using MSLP data covering the North Atlantic to evaluate the performance of forecast models under different weather conditions. Despite the regional variation between MSLP patterns observed in the North Atlantic and the resulting impact on observed rainfall patterns across the U.K. shown by Richardson *et al.* [11], these approaches fail to consider temperature as a key variable. The relationship between temperature and rainfall is captured

by the thermodynamic Clausius–Clayperon relationship which can vary across a spatial domain as small as the U.K. [8]. Further studies also show a stark contrast in synoptic temperature conditions related to heavy or intense rainfall events [13], [22].

Weather patterns and synoptic conditions have also been used to directly predict future climatological variables. In a review by Pham *et al.* [19], neural networks were shown to be capable at predicting both daily and sub-daily rainfall values to within 10 mm across spatial domains varying from local station scales to state scales such as Florida. More advanced neural network structures have also been adopted to forecast rainfall. For example, Haidar and Verma [15] used a range of climatic indices such as minimum and maximum temperature, the Southern oscillation (SO) index, the North Atlantic oscillation (NAO) index, and many others with a convolutional neural network (CNN) architecture to predict rainfall in a specific Australian suburb (Innisfail). Haidar and Verma [15] were able to achieve lower root mean squared errors (RMSEs) than both the ACCESS-S1 hindcasts and a standard neural network architecture. More recently, approaches have used CNNs to interpret synoptic scale images comprising multiple meteorological variables. For example, Rasp and Thuerey [17] used geopotential height, temperature, wind speeds, and specific humidity among others at multiple preceding time steps to predict precipitation at lead times of 6 h and 1, 3, and 5 days within 2–3 mm on a 5.625° spatial resolution grid. Despite this, they highlight the importance of higher resolution models and conclude that larger networks generally offered higher accuracy scores. However, they used a single-architecture neural network [18].

Karpathy *et al.* [16] presented four options for including a time dimension in the structure of a CNN for classification of videos. The first method takes a single frame (SF) representation of the video (SF), and the second takes two images at either end of the video, passing them through the network individually before combining them at the end (late fusion). The third approach passes the entire video through the network from start to finish [early fusion (EF)], and finally, the fourth approach passes time-defined subsets of the video through the network separately, slowly merging them through the network until they are eventually recombined at the end [slow fusion (SlowF)]. Their results highlight advantages of the different architectures; however, an ensemble approach combining all the architectures is found to be best in this domain.

This letter takes a different approach to current literature, rather than analyzing static images of the preceding conditions

Manuscript received August 24, 2021; revised December 20, 2021; accepted April 11, 2022. Date of publication April 14, 2022; date of current version April 29, 2022. The work of Andrew P. Barnes was supported by the National Productivity Investment Fund from the EPSRC under Grant EP/R512254/1. (Corresponding author: Andrew P. Barnes.)

The authors are with the Department of Architecture and Civil Engineering, University of Bath, Bath BA2 7AY, U.K. (e-mail: apb41@bath.ac.uk; trk23@bath.ac.uk; nm268@bath.ac.uk).

Digital Object Identifier 10.1109/LGRS.2022.3167456

(e.g., [12]), these conditions are characterized by a sequence of images (i.e., “a video”) representing the temporal evolution of the synoptic conditions. The meteorological videos are then interpreted by the video-based CNN variants proposed by Karpathy *et al.* [16]. A case study using the regional, monthly rainfall of the 13 regions of Great Britain are used in this study. Regional forecasting models are generated using each of the variants and are then combined into two ensembles, the first is a combination of the CNN variants as shown by Karpathy *et al.* [16] and the second includes forecasts from the MetOffice’s GloSea5 model [20]. These approaches are then compared and contrasted both against each other and against a benchmark rainfall dataset.

II. DATA

Two data sources are used to train and evaluate the CNN models. First, the Centre for Ecology and Hydrology Gridded Estimate of Areal Rainfall (CEH-GEAR) [21] is used to represent the benchmark rainfall totals, providing gridded monthly rainfall across Great Britain between 1890 and 2017. Next, from the MetOffice GloSea5 (MOS5) hindcasts [20], daily forecast MSLP and 2AT patterns are used to train the CNNs to predict the benchmark rainfall data from CEH-GEAR using a 1-mo lead time. Also, from the MOS5 dataset monthly rainfall totals are extracted for each region. However, due to the temporal limits of the MOS5 data (1994–2020) and the CEH-GEAR data (1890–2017), the temporal range of this study is limited within 1994–2017.

A. Rainfall

Great Britain has been split into 13 administrative regions as indicated by Fig. 1, and each of these regions is then further split into a set of points spaced equally at 30 km \times 30 km. The monthly rainfall total (millimeter) for a given region is calculated by averaging the total monthly rainfall from each point within the region. Both the CEH-GEAR and MOS5 datasets are provided in a gridded format, and as such each point assumes the value of the grid cell which contains the point. Following this, each region’s rainfall dataset is standardized as follows:

$$\hat{p}_{i,t} = \frac{p_{i,t} - \bar{p}_i}{\text{std}(p_i)} \quad (1)$$

where $p_{i,j}$ is a regional series of monthly rainfall totals (mm) where $i = \{0, 1, \dots, 12\}$ indicates the region being standardized and $t = \{1, \dots, 275\}$ indicates which time step (month in the series) is currently being calculated. Standardizing the regional rainfall in this way removes biases of regions with particularly high rainfall, for example, the North West of England compared with regions of low rainfall such as the South East of England. The result of this extraction and standardization are two matrices of rainfall events of size [275, 13] where 275 is the number of months available to forecast (275 instead of 276 because a 1-mo lead time is required which removes January 1994, as available data) and 13 is the number of regions.

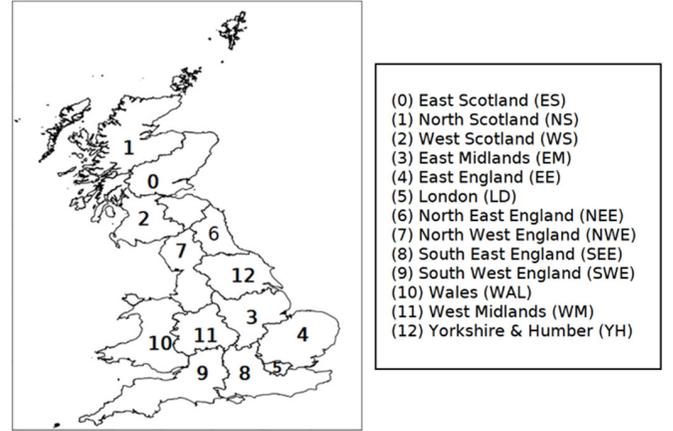


Fig. 1. 13 administrative regions of Great Britain.

B. Meteorological Data

Synoptic patterns of MSLP and 2AT were extracted from the MOS5 hindcasts [20], and these patterns were extracted for the middle 28 days of every month used in the study (February 1994 to December 2017) and covered a synoptic extent between [100 °W, 10 °N] and [20 °E, 70 °N]. The middle 28 days was chosen to represent most of the month while ensuring all months had the same amount of data available. The MOS5 data were available on a 2.5° \times 2.5° grid format, meaning each pattern is represented by a matrix [121, 61] in size. Each pattern was extracted from the final day of the preceding month; for example, patterns for days 1 February 1995 up to and including 28 February 1995 were extracted using forecasts ran on 31 January 1995.

Next, the resulting sets of MSLP and 2AT patterns were standardized separately as follows:

$$\bar{z}_{x,y}^{\text{var}} = \frac{z_{x,y}^{\text{var}} - \overline{z_{x,y}^{\text{var}}}}{\text{std}(z_{x,y}^{\text{var}})} \quad (2)$$

where $z_{x,y}^{\text{var}}$ represents the set cells at a given position where $x = \{1, \dots, 121\}$ and $y = \{1, \dots, 61\}$ for the variable given by $\text{var} = \{\text{MSLP}, 2\text{AT}\}$. The resulting two sets of matrices are of size [276, 121, 61, 28] ([number of months, longitude cells, latitude cells, number of days]) and are finally combined to give a single matrix of size [276, 121, 61, 28, 2] such that each dimension represents [number of months, longitude cells, latitude cells, number of days, variables].

C. Training, Testing, and Validation

To reduce overfitting and provide a fair comparison between the developed CNN models and MOS5 predictions, the dataset was split into training, testing, and validation. The training data were used to optimize the CNN, the testing data were used to ensure overfitting does not occur, and finally the validation dataset was used to compare the resulting CNN models with MOS5 predictions. To ensure seasonal consistency between both the training and validation datasets the validation dataset consisted of all data for years taken at a four-year interval (1997, 2001, 2005, 2009, 2013, and 2017) which equates to

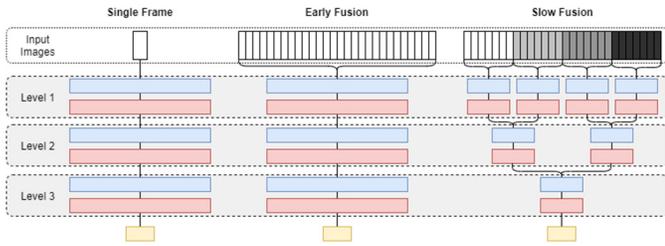


Fig. 2. Three CNN architectures for including a temporal dimension, adapted from Karpathy *et al.* [20] (2014). Blue boxes indicate convolutional layers, red boxes indicate max-pooling layers, and yellow boxes indicate fully connected linear layers.

26% of the total data, and the remaining years are allocated to training (74% of total data). During the training process, the training dataset was further split into training (70%) and testing (30%).

III. FORECASTING METHOD

In this section, the CNN variants used are followed by a description of the two ensemble approaches which are used for evaluation.

A. CNN Variants

To forecast regional monthly rainfall totals using the videos representing the forecast meteorological images, as described in Section II-B, three CNN architecture variants are trained for each region resulting in a total of 39 CNN models. Each of the three models will incorporate the time dimension (number of days) into the CNN differently and are based on the architectures proposed by Karpathy *et al.* [16] (2014).

First, an SF approach is adopted which takes a mean across the time dimension, creating a new matrix containing the average MSLP and 2AT patterns across the 28 days. This, in turn, reduces the number of dimensions from 4 ([longitude cells, latitude cells, number of days, variables]) to 3 ([longitude cells, latitude cells, variables]). This 3-D matrix is then interpreted as a static image and used to train the CNN. Next, an EF approach processes the entire 4-D matrix at once without any preprocessing. This involves the convolutional and max-pooling layers having 3-D filters with kernels covering latitude, longitude, and temporal dimensions. Finally, a SlowF approach splits the initial 4-D matrix along the temporal dimension into four equally sized (weekly) matrices of [longitude, latitude, $(28/4 = 7)$ 7, variables]. Each of these weekly matrices is then passed through the first level of the network individually before being recombined into two bi-weekly matrices by concatenating the results of the first across the temporal dimension. The “bi-weekly” matrices are then passed through a second level of the network, concatenated into a single matrix and then passed through a final third level of the network. A summary diagram of the architectures is shown in Fig. 2.

At each layer, kernels of size 2 were used such that the SF model uses kernels of size [2, 2] and the EF/SlowF models use kernels of size [2, 2, 2] to incorporate the temporal dimension convolutions. Both the SF and EF models were generated

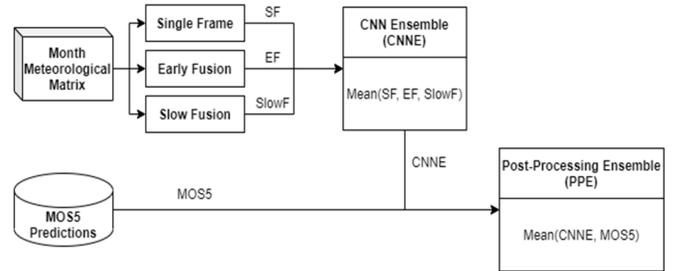


Fig. 3. Two ensemble approaches are presented. First, CNNE is calculated by taking the mean of all three CNN results for a given month. Then, the mean of CNNE and MOS5 prediction is taken as PPE.

using 128 filters in each layer, whereas the SlowF model uses 32, 64, and 128 filters in each respective layer. These sizes were chosen following several trials which varied the number of filters in each architecture.

Each of the CNN models was trained using a learning rate of 0.0001 and the Adam optimization method [22]. As mentioned in Section II-C, the training data are split into training and testing data with a split of 70% training to 30% testing. For each region and architecture variant combination, three training cycles were completed and the model with the lowest final test error was selected to represent the region and architecture combination, and this is to provide an accurate representation of the architecture/region selection which is not influenced by a poor training cycle.

B. Ensemble Forecasting

Regional forecasts of rainfall across the 13 regions were also made by combining the outputs from the CNN and MOS5 into two ensemble means. First, the CNN ensemble (CNNE) is defined as the mean of the predictions made by all three CNN architecture variants (SF, SlowF, and EF). Karpathy *et al.* [20] found that an average of the outputs improved accuracy compared with the individual CNN architectures alone; thus, this ensemble approach was adopted in this study. The second approach is a Post-processing ensemble (PPE) which is defined as the mean of the CNNE and MOS5 prediction. An overview of these ensembles for a given month is given in Fig. 3.

IV. RESULTS

This section first provides a comparison of the developed CNN models for each region against MOS5 predictions, followed by a discussion of the combined CNNE and PPE ensemble predictions. All results presented in this section refer to a comparison of the output from the models using a validation dataset as described in Section II-C; these data were kept separate from training and testing to ensure the networks had not been exposed to the validation data.

The average validation RMSEs for the CNN variants and MOS5 predictions across all regions are as follows: 63 mm (SF), 44 mm (SlowF), 37 mm (EF), and 33 mm (MOS5). The MOS5 outperforms all three CNN variants on the validation dataset across all regions. Despite this, the prediction patterns highlighted in Fig. 4 show that MOS5 predictions never exceeded 200 mm of rainfall, whereas all three CNN variants

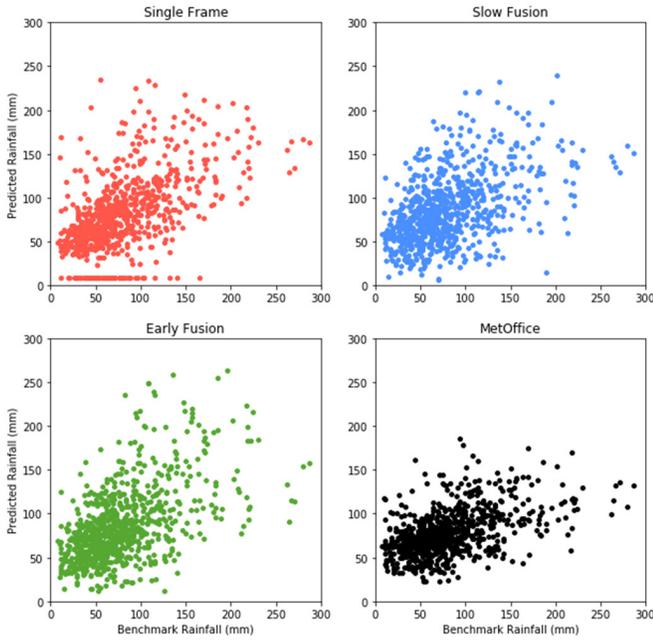


Fig. 4. Predicted rainfall against the benchmark rainfall value for all months in the validation dataset.

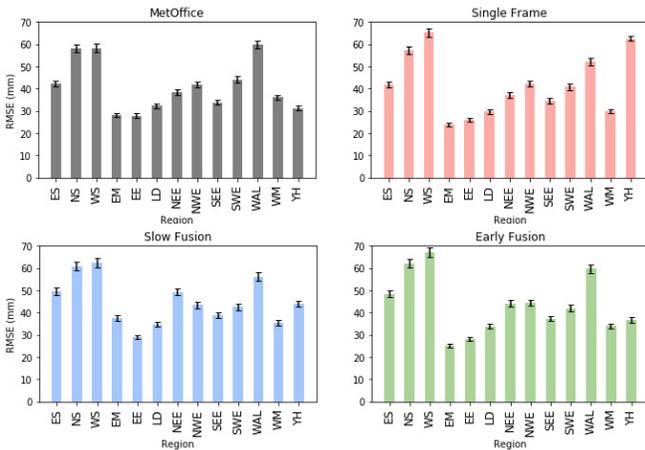


Fig. 5. Regional RMSEs for each model for the validation dataset; a 95% confidence interval of the errors is also presented in the error bars.

appear capable of doing so. This is especially true regarding the EF and SlowF variants which show strong positive correlations between predicted and benchmark rainfall, even though this comes with an increase in variance.

Fig. 5 shows the contribution to the RMSE values from across the regions. The results show that the regional bias of the models remains consistent with each model presenting a graph with broadly similar shapes with higher RMSE errors produced for regions known for higher levels of rainfall such as the three Scottish regions, Wales, and North East England. Notably, the SF variant producing a larger error for Yorkshire and Humber which relates to a subset of predictions of 1 mm in (Fig. 4 SF) indicates a lack of convergence of the CNN.

Next, the ensemble means are generated as described in Section III-B. Fig. 6 shows the validation dataset

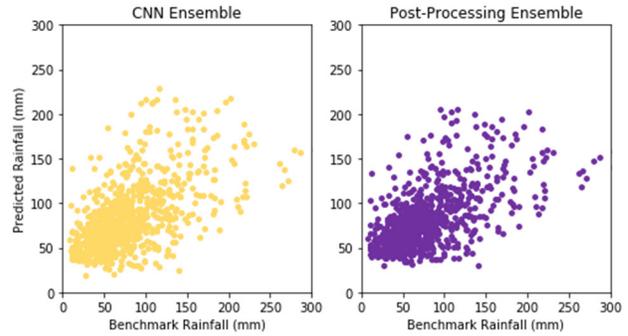


Fig. 6. Ensemble predictions for the validation dataset compared with the benchmark monthly rainfall totals (mm). (Left) CNNE and (Right) PPE.

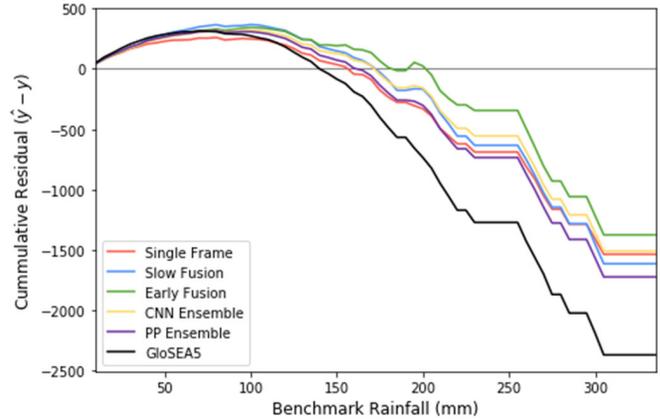


Fig. 7. Cumulative residual for all forecasting methods including the MetOffice’s GloSea5 model.

predictions for CNNE (left) and PPE (right); the variation in both the plots highlight the similarity between the two prediction models. The RMSEs for CNNE and PPE are 32 and 31 mm, respectively, across all regions; comparing this with the 33-mm RMSE found for MOS5, these results indicate that an ensemble approach with the CNN variants has the potential to provide more accurate rainfall forecasts. These results also further confirm those of Karpathy *et al.* [20] who also concluded that an ensemble approach improved their results.

Finally, Fig. 7. shows the cumulative residual of each CNN and ensemble variant alongside GloSea5 for rainfall values increasing from 0 to 350 mm. The cumulative residual is calculated by aggregating the residuals of all predictions through the benchmark rainfall domain in intervals of 5 mm. This figure illustrates that although the individual models appear less accurate at a high level, they do make significant improvement in the under-prediction of heavy rainfall events shown by the GloSea5 model. This is an important result as periods of high rainfall are often responsible for flooding.

V. CONCLUSION

This letter introduces a new technique for sub-seasonal rainfall forecasting using video-based CNNs. Regional CNNs were constructed using three different architecture variants and

trained using forecast daily MSLP and 2-m air temperature patterns with a lead time of one month. The forecast images occur throughout a given month and are used to predict the regional rainfall. Following this, two ensemble models were produced, one of which was used to take the mean of the CNN variants and the second took the mean of the CNN variants and the MOS5 model. All models were then compared with MOS5 predictions with the following findings.

- 1) Individually, the CNN variants were able to provide regional rainfall predictions based on forecast MSLP and 2AT patterns.
- 2) No regional bias was found between CNN variants. Higher errors were found in regions with higher levels of rainfall; however, this was to be expected due to the magnitude of the events.
- 3) The ensemble models both produced RMSEs lower than MOS5 predictions.
- 4) All CNN and ensemble approaches showed increased accuracy for heavy rainfall events in comparison to the MOS5 model.

These findings highlight the applicability of video-based CNNs to rainfall forecasting. This study was limited by the quantity of data used; however, the results show the use of an ensemble of different CNN architectures could provide invaluable post-processing to traditional numerical weather prediction models, especially focusing on improving the prediction of the most extreme events. To improve the models further, a cross-validation approach could be taken to increase the amount of data used during training, and to further increase the amount of data available data augmentation techniques could be applied to create a synthetic dataset, providing more training examples.

ACKNOWLEDGMENT

All data used in this study are freely available. The Centre for Ecology and Hydrology's CEH-GEAR dataset is available from their portal (<https://catalog.ceh.ac.uk/documents/ee9ab43d-a4fe-4e73-afd5-cd4fc4c82556>). The meteorological and gridded rainfall data extracted from the MOS5 model are available through ECMWF's Copernicus service (<https://cds.climate.copernicus.eu/cdsapp#!/dataset/seasonal-monthly-single-levels>).

REFERENCES

- [1] *State of the Global Climate 2020*, World Meteorol. Org., Cleveland, OH, USA, 2021.
- [2] N. Utsumi, H. Kim, S. Kanae, and T. Oki, "Relative contributions of weather systems to mean and extreme global precipitation," *J. Geophys. Res. Atmos.*, vol. 122, no. 1, pp. 152–167, 2017, doi: [10.1002/2016JD025222](https://doi.org/10.1002/2016JD025222).
- [3] P. Hu, Q. Zhang, P. Shu, B. Chen, and J. Fang, "Flood-induced mortality across the globe: Spatiotemporal pattern and influencing factors," *Sci. Total Environ.*, vol. 643, pp. 171–182, Dec. 2018, doi: [10.1016/j.scitotenv.2018.06.197](https://doi.org/10.1016/j.scitotenv.2018.06.197).
- [4] M. G. Donat, A. L. Lowry, L. V. Alexander, P. A. O'Gorman, and N. Maher, "More extreme precipitation in the world's dry and wet regions," *Nature Climate Change*, vol. 6, pp. 508–513, May 2016, doi: [10.1038/nclimate2941](https://doi.org/10.1038/nclimate2941).
- [5] S. Westra, L. V. Alexander, and F. W. Zwiers, "Global increasing trends in annual maximum daily precipitation," *J. Clim.*, vol. 26, no. 11, pp. 2904–2918, 2013.
- [6] S. Westra *et al.*, "Future changes to the intensity and frequency of short-duration extreme rainfall," *Rev. Geophys.*, vol. 52, no. 3, pp. 522–555, Sep. 2014, doi: [10.1002/2014RG000464](https://doi.org/10.1002/2014RG000464).
- [7] N. Utsumi, S. Seto, S. Kanae, E. E. Maeda, and T. Oki, "Does higher surface temperature intensify extreme precipitation?" *Geophys. Res. Lett.*, vol. 38, no. 16, pp. 1–5, Aug. 2011, doi: [10.1029/2011GL048426](https://doi.org/10.1029/2011GL048426).
- [8] S. Blenkinsop, S. C. Chan, E. J. Kendon, N. M. Roberts, and H. J. Fowler, "Temperature influences on intense U.K. Hourly precipitation and dependency on large-scale circulation," *Env. Res. Lett.*, vol. 10, no. 5, pp. 1748–9326, 2015.
- [9] L. Gimeno *et al.*, "Recent progress on the sources of continental precipitation as revealed by moisture transport analysis," *Earth-Sci. Rev.*, vol. 201, Feb. 2020, Art. no. 103070, doi: [10.1016/j.earscirev.2019.103070](https://doi.org/10.1016/j.earscirev.2019.103070).
- [10] K. Emanuel, "Tropical cyclones," *Annu. Rev. Earth Planet. Sci.*, vol. 31, pp. 75–104, Feb. 2003, doi: [10.1146/annurev.earth.31.100901.141259](https://doi.org/10.1146/annurev.earth.31.100901.141259).
- [11] D. Richardson, H. J. Fowler, C. G. Kilsby, and R. Neal, "A new precipitation and drought climatology based on weather patterns," *Int. J. Climatol.*, vol. 38, no. 2, pp. 630–648, 2017, doi: [10.1002/joc.5199](https://doi.org/10.1002/joc.5199).
- [12] R. Neal, D. Fereday, R. Crocker, and R. E. Comer, "A flexible approach to defining weather patterns and their application in weather forecasting over Europe," *Meteorol. Appl.*, vol. 23, no. 3, pp. 389–400, 2016, doi: [10.1002/met.1563](https://doi.org/10.1002/met.1563).
- [13] R. P. Allan, S. Blenkinsop, H. J. Fowler, and A. J. Champion, "Atmospheric precursors for intense summer rainfall over the United Kingdom," *Int. J. Climatol.*, vol. 40, no. 8, pp. 3849–3867, 2019, doi: [10.1002/joc.6431](https://doi.org/10.1002/joc.6431).
- [14] D. Richardson, H. J. Fowler, C. G. Kilsby, R. Neal, and R. Dankers, "Improving sub-seasonal forecast skill of meteorological drought: A weather pattern approach," *Natural Hazards Earth Syst. Sci.*, vol. 20, pp. 107–124, Jan. 2020, doi: [10.5194/nhess-20-107-2020](https://doi.org/10.5194/nhess-20-107-2020).
- [15] A. Haidar and B. Verma, "Monthly rainfall forecasting using one-dimensional deep convolutional neural network," *IEEE Access*, vol. 6, pp. 69053–69063, 2018, doi: [10.1109/ACCESS.2018.2880044](https://doi.org/10.1109/ACCESS.2018.2880044).
- [16] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proc. CVPR*, 2014, pp. 1725–1732.
- [17] S. Rasp and N. Thuerey, "Data-driven medium-range weather prediction with a resnet pretrained on climate simulations: A new model for weatherbench," *J. Adv. Model. Earth Syst.*, vol. 13, no. 2, 2021, Art. no. e2020MS002405, doi: [10.1029/2020MS002405](https://doi.org/10.1029/2020MS002405).
- [18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.
- [19] B. T. Pham *et al.*, "Development of advanced artificial intelligence models for daily rainfall prediction," *Atmos. Res.*, vol. 237, Jun. 2020, Art. no. 104845, doi: [10.1016/j.atmosres.2020.104845](https://doi.org/10.1016/j.atmosres.2020.104845).
- [20] C. MacLachlan *et al.*, "Description of GloSea5: The met office high resolution seasonal forecast system," *Quart. J. Roy Meteorol. Soc.*, vol. 141, no. 689, pp. 1072–1084, 2015, doi: [10.1002/qj.2396](https://doi.org/10.1002/qj.2396).
- [21] M. Tanguy, H. Dixon, I. Prodocimi, D. G. Morris, and V. D. J. Keller, "Gridded estimates of daily and monthly areal rainfall for the United Kingdom [CEH-GEAR]," *NERC Environ. Inf. Data Centre*, pp. 1890–2017, Nov. 2019, doi: [10.5285/ee9ab43d-a4fe-4e73-afd5-cd4fc4c82556](https://doi.org/10.5285/ee9ab43d-a4fe-4e73-afd5-cd4fc4c82556).
- [22] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–15.
- [23] A. P. Barnes, C. Svensson, and T. R. Kjeldsen, "North Atlantic air pressure and temperature conditions associated with heavy rainfall in Great Britain," *Int. J. Climatol.*, vol. 42, no. 5, pp. 3190–3207, Oct. 2021, doi: [10.1002/joc.7414](https://doi.org/10.1002/joc.7414).