IMPROVING THE SPATIAL ACCURACY OF EXTREME TROPICAL CYCLONE RAINFALL IN ERA5 USING DEEP LEARNING

Guido Ascenso, Andrea Ficchì, Matteo Giuliani & Andrea Castelletti
Department of Electronics, Information, and Bioengineering
Politecnico di Milano
Milano, Italy
{firstname.lastname}@polimi.it

Leone Cavicchia & Enrico Scoccimarro
Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici
Bologna, Italy
{firstname.lastname}@cmcc.it

ABSTRACT

We propose a novel method for the bias adjustment and post-processing of gridded rainfall data products. Our method uses U-Net (a deep convolutional neural network) as a backbone, and a novel loss function given by the combination of a pixelwise bias component (Mean Absolute Error) and a spatial accuracy component (Fractions Skill Score). We evaluate the proposed approach by adjusting extreme rainfall from the popular ERA5 reanalysis dataset, using the multi-source observational dataset MSWEP as a target. We focus on a sample of extreme rainfall events induced by tropical cyclones and show that the proposed method significantly reduces both the MAE (by 16%) and FSS (by 53%) of ERA5.

1 INTRODUCTION

Tropical cyclones (TCs) are one of the costliest and deadliest natural hazards due to the combination of their strong winds and induced storm surges and heavy precipitation, which can cause devastating floods (Mendelsohn et al., 2012). Unfortunately, due to its high spatio-temporal variability, complex underlying physical processes, and lack of high-quality observations, precipitation is still one of the most challenging aspects of a TC to model (Zhao et al., 2022). However, as precipitation is a key forcing variable for hydrological processes acting across multiple space-time scales, accurate precipitation input is crucial for reliable hydrological simulations and forecasts which can be used to inform disaster risk management decisions.

A popular source of precipitation data is the ERA5 reanalysis dataset, frequently used as input to hydrological models when studying floods (Harrigan et al., 2020; Dullaart et al., 2020; Cantoni et al., 2022; Wanzala et al., 2022). Previous studies found that ERA5 systematically underestimates heavy precipitation events Bhattacharyya et al. (2022); Sun et al. (2022), and in particular (as shown later in this paper), the spatial distribution of TC-rainfall in ERA5 has large room for improvement and this is key to accurately identify TC landfall and inform decisions.

Here, we present a precipitation post-processing scheme based on U-Net, a popular deep-learning architecture (Ronneberger et al., 2015). Originally developed in the field of medical computer vision (Ronneberger et al., 2015), U-Net (Figure 1) is now ubiquitously used when there is a need to apply a transformation to gridded data without altering the input resolution. Rather than only adjusting the per-pixel precipitation values at each timestep of a given TC, we explicitly design our model to also adjust the spatial distribution of the precipitation; to the best of our knowledge, we are the first to do so. The key novelty of our model is a custom-made loss function, based on the combination of the Fractions Skill Score (FSS; (Roberts & Lean, 2008)) and Mean Absolute Error (MAE) metrics. We train and validate the model on over 200k time steps from global precipitation events induced by TCs. We show how a U-Net trained with our loss function can reduce the per-pixel MAE of ERA5 precipitation by nearly as much as a U-Net trained with only MAE as the loss function, while out-performing it significantly in terms of improved spatial patterns of precipitation. Finally, we discuss how the outputs of our model can be used for future research.

2 Data and methodology

2.1 Data

We downloaded the hourly global total precipitation fields of ERA5 from the Climate Data Store[1] at a resolution of 0.25° x 0.25°, between 1980 and 2020. We downloaded precipitation data over the same period also for the MSWEP dataset (a multi-source observational dataset blending gauge, satellite, and reanalysis data, and currently one of the most accurate precipitation datasets Sharifi et al. (2019)), which we used as the gold standard in our study (i.e., as the target for the deep learning model). As MSWEP comes with 3-hourly temporal and 0.1° x 0.1° spatial resolution, we regridded ERA5 onto MSWEP's grid using linear interpolation and aggregated it to a 3-hourly temporal resolution.

To locate TC centres, we used the best-track data from the International Best track Archive for Climate Stewardship (IBTrACS) project (version v04r00[2]), at a temporal resolution of 3h. For each timestep in IBTrACS (i.e., the location of a TC in time and space), we cropped ERA5 and MSWEP around a 500 km-radius box centred on that TC's location, yielding rainfall grids of dimensions 96 x 96 x 1 pixels for each dataset. Repeating this procedure for all timesteps in IBTrACS (except those marked as "spur" or "extratropical", as per Schreck et al. (2014)'s guidelines) resulted in 258,834 training pairs of (ERA5, MSWEP) grids, which we split into 165,654 (64%) for training, 41,413 (16%) for validation, and 51,767 (20%) for testing.

2.2 Quantifying spatial errors

Previous works on the correction of rainfall gridded products used pixelwise metrics (e.g., Mean Absolute Error and Mean Squared Error) to guide the training of models and evaluate their performance (Le et al., 2020; Sadeghi et al., 2020; Hu et al., 2021; Han et al., 2021). However, pixelwise metrics encourage models to avoid predictions with sharp gradients, resulting in predictions that are "blurred out" (Stengel et al., 2020). Furthermore, if the predictions are perfect in terms of pattern and intensity but slightly offset (e.g., by even just one pixel), pixelwise metrics will be very poor, thus penalising excessively non significant errors (Gilleland et al., 2009).

Following the work by Lagerquist & Ebert-Uphoff (2022), we propose instead to use a *spatial verification* metric (i.e., a metric that quantifies the similarity of the spatial patterns in two gridded products) frequently used in atmospheric science: the Fraction Skill Score (FSS; (Roberts & Lean, 2008)), which takes values between 0 (no match) and 1 (perfect match). The FSS first applies an intensity threshold Q (which here we consider to be a percentile of rainfall intensity) to the input grids, turning them into binary maps, with pixels of values 1 (0) if greater (smaller) than Q. Given these binary maps, the FSS then calculates the fraction of pixels within a patch of size N that are positive, and computes the Mean Square Error between observed and predicted grids (over all patches) . Following a grid search for $N \in [3, 15, 27]$ and $Q \in [80, 95, 99]$, we selected N = 15 and Q = 99 as the combination of parameters that maximised the ability of the metric to measure the similarity of spatial patterns while minimising the pixelwise errors (computed via MAE). Therefore, our implementation of the FSS quantifies the similarity of patterns of rainfall in 15 x 15 pixels patches for the 1% most intense rainfall in the input grids.

To use the FSS as a loss function for a neural network, we adjusted it in two ways: (1) we inverted it, so that FSS = 1 indicates no match between the observed and predicted grids (thus yielding the greatest possible gradient); (2) we made it differentiable by replacing the hard threshold (Q) exceedance function by an *arctan* function of Q (Equation 1), followed by a Gaussian filter and a normalisation into the range [0, 1]. Cumulatively, these steps brought the values in each grid close

[1] https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview

[2] https://www.ncdc.noaa.gov/ibtracs/

Table 1: U-Net hyperparameters and implementation details

Name	**Value**	**Name**	**Value**
# blocks	3	# filters	[8, 16, 32, 64]
filter size	3 x 3	batch size	16
epochs	200	dropout	none
optimiser	RMSprop	lr	1e-4

to binary without applying a non-differentiable hard threshold. For the rest of the paper, we will refer to this modified version of the FSS using the notation FSS$'$.

$$binary\_map = arctan(input\_grid - percentile(input\_grid, Q = 99)) \tag{1}$$

2.3 Deep learning model

To perform the adjustment of ERA5 gridded precipitation, we used the U-Net deep convolutional neural network. In U-Net, inputs are first encoded via a series of convolutional and max pooling layers (which reduce the spatial resolution and increase the semantic meaning of the information), and then decoded via a series of convolutional and upsampling layers (which restore the spatial resolution while maintaining high-level semantic information). Blocks of the encoder and decoder that are at the same depth are then connected to facilitate the transfer of spatial information to the semantically complex layers of the decoder.

The hyperparameters and implementation details of the network are reported in Table 1. We trained two versions of U-Net: (i) U-Net$_{MAE}$, using MAE as a loss function; (ii) U-Net$_{comp}$, using a compound loss defined as:

$$compound\_loss = \alpha_0 MAE + \alpha_1 FSS' \tag{2}$$

where the optimal values for α_0 and α_1 were found to be 1 and 0.75, respectively, via tuning on the validation set.

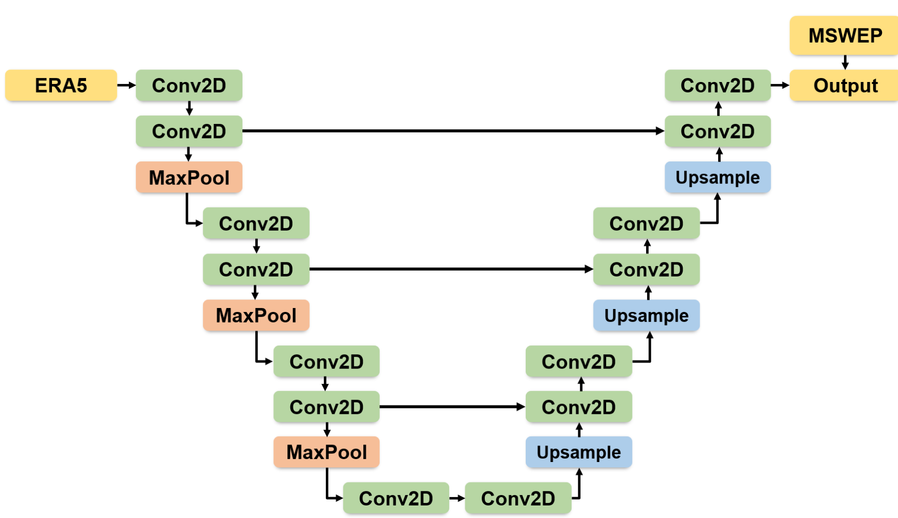

Figure 1: Sketch of U-Net's architecture as implemented in our experiments.

Table 2: Test set results

	ERA5	**U-Net$_{MAE}$**		**U-Net$_{comp}$**		
	Value	Value	% under ERA5	Value	% under ERA5	% under U-Net$_{MAE}$
MAE	1.170	0.951	**18.536**	0.974	**16.557**	**-2.430**
FSS′ (Q=80)	0.093	0.089	4.104	0.089	4.212	0.113
FSS′ (Q=95)	0.291	0.287	1.408	0.236	19.128	17.973
FSS′ (Q=99)	0.498	0.398	**19.936**	0.233	**53.215**	**41.566**

3 Results and Discussion

We report four metrics to evaluate the performance of our models (Table 2): MAE, to verify that the rainfall is also adjusted pixelwise (local biases); and three implementations of FSS′ with three intensity thresholds (80%, 95%, and 99% percentiles), to verify if the rainfall spatial patterns are adjusted only at the intensity threshold used in the loss function (99%) or also at lower intensities.

Results show that both U-Net$_{MAE}$ and U-Net$_{comp}$ reduce the MAE and FSS′ (Q=99). However, U-Net$_{comp}$, while retaining MAE performance comparable to U-Net$_{MAE}$, improves the FSS′ for Q=99 more substantially. Furthermore, while U-Net$_{MAE}$ improves the FSS′ only for Q=99, U-Net$_{comp}$ also significantly improves FSS′ for Q=95, showing a greater propensity for more general spatial pattern adjustment. We also show (Figure 2) that these improvements in spatial accuracy are more visible for U-Net$_{comp}$ than for U-Net$_{MAE}$.

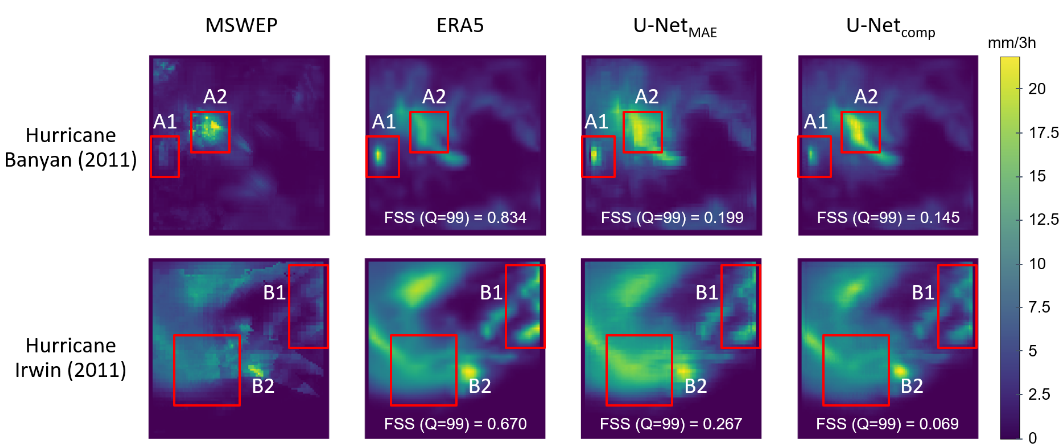

Figure 2: Examples of how U-Net$_{comp}$ corrects spatial patterns better than U-Net$_{MAE}$, which seems to rather blur out the input. Boxes A1-A2 and B1-B2 highlight parts of the grid in which the improvement of U-Net$_{comp}$ over U-Net$_{MAE}$ is especially noticeable.

4 Conclusions and future steps

In this paper, we presented a novel approach for the adjustment of gridded rainfall products which specifically aims to improve both the spatial patterns and intensity biases. Our method relies on a custom loss function, made by the combination of the MAE and FSS loss function, used to train a U-Net deep convolutional network. Compared to an equivalent network trained with a classic MAE loss, our method achieved comparable pixelwise bias adjustment and greatly superior spatial pattern adjustment.

We foresee two directions in which this work could be developed further: one that uses a more complex backbone for the prediction (i.e., a deeper/newer U-Net architecture), and one that applies the proposed method to gridded rainfall forecasts, to understand if improved spatial patterns of forecast rainfall can be useful for flood preparedness and emergency response applications.

Acknowledgments

Paper under double-blind review.

References

Suman Bhattacharyya, S Sreekesh, and Andrew King. Characteristics of extreme rainfall in different gridded datasets over india during 1983–2015. *Atmospheric Research*, 267:105930, 2022.

Elia Cantoni, Yves Tramblay, Stefania Grimaldi, Peter Salamon, Hamouda Dakhlaoui, Alain Dezetter, and Vera Thiemig. Hydrological performance of the era5 reanalysis for flood modeling in tunisia with the lisflood and gr4j models. *Journal of Hydrology: Regional Studies*, 42: 101169, 2022.

Job CM Dullaart, Sanne Muis, Nadia Bloemendaal, and Jeroen CJH Aerts. Advancing global storm surge modelling using the new era5 climate reanalysis. *Climate Dynamics*, 54:1007–1021, 2020.

Eric Gilleland, David Ahijevych, Barbara G Brown, Barbara Casati, and Elizabeth E Ebert. Intercomparison of spatial forecast verification methods. *Weather and forecasting*, 24(5):1416–1430, 2009.

Lei Han, Mingxuan Chen, Kangkai Chen, Haonan Chen, Yanbiao Zhang, Bing Lu, Linye Song, and Rui Qin. A deep learning method for bias correction of ecmwf 24–240 h forecasts. *Advances in Atmospheric Sciences*, 38(9):1444–1459, 2021.

Shaun Harrigan, Ervin Zsoter, Lorenzo Alfieri, Christel Prudhomme, Peter Salamon, Fredrik Wetterhall, Christopher Barnard, Hannah Cloke, and Florian Pappenberger. Glofas-era5 operational global river discharge reanalysis 1979–present. *Earth System Science Data*, 12(3):2043–2060, 2020.

Yi-Fan Hu, Fu-Kang Yin, and Wei-Min Zhang. Deep learning-based precipitation bias correction approach for yin–he global spectral model. *Meteorological Applications*, 28(5):e2032, 2021.

Ryan Lagerquist and Imme Ebert-Uphoff. Can we integrate spatial verification methods into neural network loss functions for atmospheric science? *Artificial Intelligence for the Earth Systems*, 1 (4):e220021, 2022.

Xuan-Hien Le, Giha Lee, Kwansue Jung, Hyun-uk An, Seungsoo Lee, and Younghun Jung. Application of convolutional neural network for spatiotemporal bias correction of daily satellite-based precipitation. *Remote Sensing*, 12(17):2731, 2020.

Robert Mendelsohn, Kerry Emanuel, Shun Chonabayashi, and Laura Bakkensen. The impact of climate change on global tropical cyclone damage. *Nature climate change*, 2(3):205–209, 2012.

Nigel M Roberts and Humphrey W Lean. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, 136(1):78–97, 2008.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Mojtaba Sadeghi, Phu Nguyen, Kuolin Hsu, and Soroosh Sorooshian. Improving near real-time precipitation estimation using a u-net convolutional neural network and geographical information. *Environmental Modelling & Software*, 134:104856, 2020.

Carl J Schreck, Kenneth R Knapp, and James P Kossin. The impact of best track discrepancies on global tropical cyclone climatologies using ibtracs. *Monthly Weather Review*, 142(10):3881–3899, 2014.

Ehsan Sharifi, Josef Eitzinger, and Wouter Dorigo. Performance of the state-of-the-art gridded precipitation products over mountainous terrain: A regional study over austria. *Remote Sensing*, 11(17):2018, 2019.

Karen Stengel, Andrew Glaws, Dylan Hettinger, and Ryan N King. Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117(29): 16805–16815, 2020.

He Sun, Tandong Yao, Fengge Su, Zhihua He, Guoqiang Tang, Ning Li, Bowen Zheng, Jingheng Huang, Fanchong Meng, Tinghai Ou, et al. Corrected era5 precipitation by machine learning significantly improved flow simulations for the third pole basins. *Journal of Hydrometeorology*, 23(10):1663–1679, 2022.

Maureen A Wanzala, Andrea Ficchi, Hannah L Cloke, Elisabeth M Stephens, Heou M Badjana, and David A Lavers. Assessment of global reanalysis precipitation for hydrological modelling in data-scarce regions: A case study of kenya. *Journal of Hydrology: Regional Studies*, 41:101105, 2022.

You Zhao, Chao Liu, Di Di, Ziqiang Ma, and Shihao Tang. High-resolution typhoon precipitation integrations using satellite infrared observations and multisource data. *Atmospheric Measurement Techniques*, 15(9):2791–2805, 2022.