Weather and Climate Extremes 46 (2024) 100724

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Downscaling, bias correction, and spatial adjustment of extreme tropical cyclone rainfall in ERA5 using deep learning

Guido Ascenso [a,*], Andrea Ficchì [a], Matteo Giuliani [a], Enrico Scoccimarro [b], Andrea Castelletti [a]

[a] Politecnico di Milano, Department of Electronics, Information, and Bioengineering, Milan, 20133, Italy
[b] Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, 40127, Italy

ARTICLE INFO

Dataset link: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview, https://www.gloh2o.org/mswep/, https://www.ncei.noaa.gov/products/international-best-track-archive

Keywords:
Tropical cyclones
Extreme rainfall
Reanalysis
Bias correction
Deep learning

ABSTRACT

Hydrological models that are used to analyse flood risk induced by tropical cyclones often input ERA5 reanalysis data. However, ERA5 precipitation has large systematic biases, especially over heavy precipitation events like Tropical Cyclones, compromising its usefulness in such scenarios. Few studies to date have performed bias correction of ERA5 precipitation and none of them for extreme rainfall induced by tropical cyclones. Additionally, most existing works on bias adjustment focus on adjusting pixel-wise metrics of bias, such as the Mean Squared Error (MSE). However, it is equally important to ensure that the rainfall peaks are correctly located within the rainfall maps, especially if these maps are then used as input to hydrological models. In this paper, we describe a novel machine learning model that addresses both gaps, RA-U_{cmpd}, based on the popular U-Net model. The key novelty of RA-U_{cmpd} is its loss function, the *compound loss*, which optimizes both a pixel-wise bias metric (the MSE) and a spatial verification metric (a modified version of the Fractions Skill Score). Our results show how RA-U_{cmpd} improves ERA5 in almost all metrics by 3-28%—more than the other models we used for comparison which actually worsen the total rainfall bias of ERA5—at the cost of a slightly increased (3%) error on the magnitude of the peak. We analyse the behaviour of RA-U_{cmpd} by visualizing accumulated maps of four particularly wet tropical cyclones and by dividing our data according to the Saffir-Simpson scale and to whether they made landfall, and we perform an error analysis to understand under what conditions our model performs best.

1. Introduction

In the past 50 years, tropical cyclones (TCs) have killed close to 800,000 people and caused upwards of USD 1400 billion in economic losses from property damage, infrastructure destruction, agricultural losses, and business interruptions (World Meteorological Organization (WMO), 2023). Although sustained wind speed is typically used as a measure of the severity of TCs (National Hurricane Center (NHC), 2023; Bloemendaal et al., 2021), their torrential rainfall claims many more lives (Rappaport, 2014), mainly by triggering floods (Park et al., 2015, 2016; Bakkensen et al., 2018). Floods induced by TC rainfall occur not only in coastal areas hit during landfall, but also hundreds of kilometres inland, in the form of pluvial and river flooding (Villarini et al., 2011). The risk of rainfall-induced flooding increases as a TC weakens after landfall and moves more slowly, extending the period during which intense precipitation is discharged over a specific geographic area (Czajkowski et al., 2013; Villarini et al., 2014). Furthermore, climate change is expected to amplify TC-induced rainfall more than wind speed (Knutson et al., 2010; Walsh et al., 2016). Under a 2 °C global warming scenario, TC wind speeds are projected to increase by 1 to 10%, while TC rainfall rates are projected to increase by 6 to 22% (Knutson et al., 2020). However, due to the complex interplay of the underlying dynamic and thermodynamic processes, accurately modelling TC-induced rainfall remains challenging (Jiang et al., 2012; Scoccimarro et al., 2014; Zhang et al., 2018).

Reanalysis datasets, such as ERA5 (Hersbach et al., 2020) produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), are key resources for studying historical weather events like TCs. Indeed, precipitation fields from ERA5 are often used as inputs to hydrological models and risk assessment analyses of TC-induced flood and compound impacts (Harrigan et al., 2020; Dullaart et al., 2020; Cantoni et al., 2022; Sohrabi et al., 2023), as well as to provide the initial conditions for real-time operational hydrological forecasts and reforecasts (e.g., Harrigan et al. (2023)). However, previous studies have found that ERA5 systematically underestimates heavy precipitation events (Bhattacharyya et al., 2022; Sun et al., 2022) and have suggested that ERA5 rainfall data must be bias corrected before being

* Corresponding author.
E-mail address: guido.ascenso@polimi.it (G. Ascenso).

https://doi.org/10.1016/j.wace.2024.100724
Received 9 February 2024; Received in revised form 2 July 2024; Accepted 18 September 2024
Available online 30 September 2024

used in hydrological models (Sun et al., 2021; Ansari and Grossi, 2022). Additionally, TC rainfall often exhibits considerable spatial heterogeneity (Lin et al., 2015), with sharp gradients in precipitation patterns that may be difficult to capture accurately with the 31 km horizontal resolution of ERA5 (Smitha et al., 2018).

Although higher-quality gridded precipitation datasets exist (e.g. MSWEP, Beck et al. 2019b, and IMERG, Huffman et al. 2015), there are several reasons one might prefer a bias-adjusted version of ERA5 over other gridded datasets. In complex-terrain areas, which often correspond to TC-prone regions (e.g., Eastern Asia, Mozambique), the in-situ-based data usually suffer from sparse rain gauge measurements with low representativeness, which makes the interpolation unreliable (Jiang et al., 2021), while satellite-based precipitation products tend to have large uncertainties (being unable to reflect the topography-dependence of precipitation), as demonstrated in many studies (e.g., Hirpa et al., 2010). As reanalysis data produced by high-resolution atmospheric models can represent range-wide precipitation better than a collective network of rain gauges in some complex-terrain areas (Iseri et al., 2021), putting together the value of reanalysis and observations is urgently needed for hydrological and meteorological applications, like TC rainfall risk assessment. Also, reanalysis datasets like ERA5 can help complement observational products beyond complex-terrain areas, filling spatial and temporal gaps too (e.g., back extension for periods before satellite data were available). Moreover, when high frequency extreme concurrent events are considered, having a common dataset (ERA5 downscaled and bias-corrected), with the same grid, time frequency and temporal coverage for all the investigated fields (e.g. temperature and precipitation for heat waves and TC compound events) is fundamental. Finally, ERA5 is produced by the ECMWF Integrated Forecast System (IFS), which is the same model used for ECMWF's medium-range forecasts (HRES and ENS). Thus, for hydrological forecasting systems (e.g., GloFAS, see Alfieri et al., 2013) driven by ECMWF forecasts as inputs, it is better to use a consistent dataset, like the ERA5 reanalysis (original or adjusted) for initialization or historical simulation purposes (e.g., to obtain a climatology that is as consistent as possible with the forecasts).

Most studies on the bias correction of precipitation data focus on correcting the outputs of regional or global climate numerical models (Piani et al., 2010; Gudmundsson et al., 2012; Teutschbein and Seibert, 2012; Cannon et al., 2015; Ngai et al., 2017; Zhao et al., 2017; Luo et al., 2018; Ayugi et al., 2020; Mendez et al., 2020). Only a few works exist on the bias correction specifically of ERA5 precipitation data. Ansari and Grossi (Ansari and Grossi, 2022) corrected the large ERA5 precipitation biases found in the Upper Jhelum Basin in South Asia, comparing the performance of four established methods: linear scaling (Leander and Buishand, 2007), local scaling intensity (Teutschbein and Seibert, 2012), power transmission (Leander and Buishand, 2007), and distribution mapping (Sennikovs and Bethers, 2009). They found that all four methods (especially distribution mapping) were able to reduce ERA5 biases, especially in terms of the magnitude of rainfall. Still, none of them were able to reproduce the spatial patterns of the observed data for all indices of extreme precipitation considered. Probst and Mauser (2022) performed a linear bias correction of ERA5 precipitation in the Danube river basin, using different combinations of forcing datasets to perform the bias correction. They used bias-corrected and uncorrected ERA5 data as input for a hydrological model and found that using bias-corrected data reduced the model discharge error compared to the observed data from 26.8% to 4.7%. Voropay et al. (2021) also used a linear approach (Lenderink et al., 2007; Teutschbein and Seibert, 2012) to correct ERA5 precipitation biases over South Siberia, a region with a particularly sparse network of weather stations (Anisimov and Zhil'tsova, 2012). Their bias correction scheme reduced the Mean Absolute Error (MAE) of ERA5 precipitation—compared to rain gauge data—by 35%. Irwandi et al. (2023) used quantile mapping to perform bias correction of ERA5-Land (a higher-resolution version of ERA5 over land Muñoz-Sabater et al. (2021)) precipitation and temperature in the Lake Toba region of Indonesia. Their bias-corrected data reduced considerably the MAE for all months of the year (though the exact magnitude of the improvement was not reported). Finally, Cucchi et al. (2020) used the WATCH Forcing Data (WFD) methodology to bias adjust several ERA5 surface meteorological variables, including precipitation. However, their product has a spatial resolution of 0.5°, which is too low to resolve the complex structures and rainbands within tropical cyclones, and a monthly temporal resolution, which is too low for flood modelling.

With the recent surge in machine learning (ML) and artificial intelligence (AI) applications in climate science (Eccel et al., 2007; Zjavka, 2016; Yi et al., 2018; Cho et al., 2020; Chen et al., 2021; Han et al., 2021; Li et al., 2023; Zhang et al., 2023), some authors have started to use such models also for bias correction of climatological fields—again, mostly of outputs of numerical climate models (e.g., Wang and Tian, 2022; Tan et al., 2021; Kheir et al., 2023) and weather prediction models (Hess and Boers, 2022; Han et al., 2021; Bretherton et al., 2022). Only one paper has been published that used an ML model to bias correct ERA5 rainfall. Sun et al. (2022) used a random forest-based machine learning algorithm to correct ERA5 precipitation biases over the Hindu Kush–Karakoram–Himalayan mountainous region, using the Weather Research and Forecasting (WRF) model (Skamarock et al., 2008) as the target during the training of their algorithm. As the WRF data has a 9-km resolution, Sun et al. first downscaled ERA5 precipitation to match it; indeed, it is common for bias correction and downscaling to be performed concurrently or in cascade (Shrestha et al., 2017; Probst and Mauser, 2022). They tested their results on data from 940 gauges in 11 basins and found that their algorithm reduced precipitation biases by 10%–50%.

None of the works mentioned above specifically focus on TC-induced extreme rainfall, nor do they consider areas affected by TCs. Therefore, to our knowledge, there is a literature gap regarding how to correct ERA5 biases for TC-induced precipitation. In this paper, we bridge this gap using a variant of the popular deep learning architecture UNet (Ronneberger et al., 2015), called RA-UNet (Jin et al., 2020) (which we abbreviate as RA-U for simplicity). While the original UNet was developed for biomedical image segmentation (Ronneberger et al., 2015), variants of this powerful convolutional neural network are now used for image-to-image transformations across many fields (Baheti et al., 2020; Trebing et al., 2021; Hu et al., 2019; Moustafa et al., 2021), including downscaling and bias correction (Sharifi et al., 2019; Singh et al., 2021; Hess and Boers, 2022; Xiang et al., 2022; Le et al., 2023). Motivated by the potential use of ERA5 bias-adjusted rainfall maps as input to hydrological and hydrodynamic models, we set out to adjust not just pixel-level biases but also the spatial distribution of rainfall. We specifically aim to adjust the location of the rainfall peaks, which are especially biased in ERA5 and which have great value for hydrological models downstream (Smith et al., 2004; Zoccatelli et al., 2011). To this end, we have designed a custom loss function, which we call *compound loss*, combining a term to adjust pixel-wise biases (the Mean Squared Error, MSE) with a term to adjust spatial patterns: the Fractions Skill Score (FSS (Roberts and Lean, 2008)), a spatial verification metric frequently used in atmospheric science and hydrology (Zhao and Zhang, 2018; Skok, 2015; Skok and Roberts, 2016, 2018; Mittermaier, 2021; Necker et al., 2023). Most previous works on the bias adjustment of rainfall gridded products used pixel-wise metrics (most notably MSE) 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). By definition, pixel-wise metrics evaluate differences between two gridded datasets one pixel at a time and then compute the average over the entire grid, ignoring spatial patterns. This means that if the two grids are identical but offset by even one pixel, pixel-wise metrics could be very poor, not reflecting the actual similarity between the two grids (Gilleland et al., 2009). Furthermore,

pixel-wise metrics (especially MSE) discourage models from making predictions with sharp gradients, often resulting in predictions that are "blurred out" (Stengel et al., 2020; Hess and Boers, 2022; Lagerquist and Ebert-Uphoff, 2022). Using a spatial verification metric as a component of our compound function seeks to alleviate this problem. The results show that our model improves ERA5 in all metrics considered and performs better than other models we tested for comparison. This finding underscores the importance of explicitly adjusting the localization of the rainfall peaks, rather than just adjusting the bias at the pixel- or image-level, as is typically done. Finally, we discuss how the outputs of our model can be used for future research and how the work can be extended to other rainfall products—specifically, forecasts.

2. Methods

2.1. Data

We use hourly global ERA5 fields, retrieved from the Copernicus Climate Data Store[1] at a resolution of 0.25° × 0.25°, for the period 2001–2019. We chose not to include data prior to 2001 because preliminary analyses revealed that, prior to this year, the variance of the bias of ERA5 was too large and would likely have compromised the training of our model. This difference in bias before 2001 is probably attributable to the fact that fewer satellites were assimilated in reanalysis data sets prior to this period, as discussed in Diniz and Todling (2020). We use four fields from ERA5: total precipitation, total column of water, temperature at 850 hPa, and relative humidity at 850 hPa. The latter three were used as correction factors to help the model resolve the biases of the precipitation field and were selected based on previous literature (Sha et al., 2020; Hu et al., 2021, 2022; Ling et al., 2022). We use precipitation data over the same period also for the MSWEP dataset,[2] a multi-source observational dataset blending gauge, satellite, and reanalysis data (Beck et al., 2019b). Currently, MSWEP is one of the most accurate precipitation datasets (Sharifi et al., 2019), exhibiting better performance than other precipitation products in densely gauged and ungauged regions (Beck et al., 2017, 2019a). For instance, a recent study (Beck et al., 2019b) evaluated various gridded precipitation products at a common 3-h temporal and 0.1°spatial resolution with respect to the NCEP Stage-IV dataset, which merges data from 140 radars and ~5500 gauges over the continental US, and found that MSWEP had the highest correlation (74%) with these radar-gauge data than any of the other products (the second-best, IMERG Final V6, had a correlation of 58%). As MSWEP has a 0.1° × 0.1° spatial resolution, we regridded ERA5 onto MSWEP's grid using linear interpolation. Therefore, our task involves both bias adjustment and downscaling, for which UNet-like models are ideally suited and have been used in previous studies on precipitation (e.g., Hu et al., 2021, 2022; Ling et al., 2022; Hess and Boers, 2022). Furthermore, MSWEP is a 3-h aggregated product, so we computed a 3-h aggregate for ERA5 total precipitation and column of water and a 3-h average for ERA5 temperature and relative humidity to match it.

To locate TC centres, we use the best-track data from the International Best Track Archive for Climate Stewardship (IBTrACS) project (version v04r00[3]), which has a 3-h temporal resolution (Knapp et al., 2010). Following the guidelines by Schreck et al. (2014), we eliminated all records labelled as "spur" or "extratropical", as well as all those belonging to storms that never attained tropical cyclone status (wind speed ≥ 34 knots). As MSWEP data are aggregated forward in time (e.g., rainfall at 03:00 UTC refers to the rainfall accumulated between 03:00 and 05:59 UTC, included) and IBTrACS reports instantaneous data every three hours starting at 00:00, we shifted all data in IBTrACS forward by 1.5-hour using linear interpolation, so that each time step (i.e., row in IBTrACS) would be centred within the MSWEP accumulation window. For each time step in IBTrACS, we then cropped the ERA5 and MSWEP fields around a 1000 km-side box centred on that TC's location in time and space. This resulted in grids of dimensions 96 × 96 × 1 for MSWEP and 96 × 96 × 4 for ERA5. Repeating this procedure for all time steps in IBTrACS yielded 59,644 pairs of (ERA5, MSWEP) grids, which we split into 36,567 (62%) for training, 9142 (15%) for validation, and 13,935 (23%) for testing, ensuring that no TC was split among different sets. Fig. 1 illustrates a flowchart of this entire data pre-processing pipeline.

2.2. Quantifying spatial errors via the compound loss

Following the work by Lagerquist and Ebert-Uphoff (2022), we quantify the errors in the localization of the peaks of rainfall using a *spatial verification* metric frequently used in atmospheric science: the Fractions Skill Score (FSS; Roberts and Lean, 2008). The FSS is a metric that measures the similarity of the spatial patterns in two gridded products (usually, a model prediction against observations) and takes values between 0 (no match) and 1 (perfect match). The method to compute the FSS first applies an intensity threshold Q (here considered as a percentile of rainfall intensity, calculated separately for each image) to the input grids, turning them into binary maps with pixels of values 1 (0) if greater (smaller) than Q. Given these binary maps, fractional coverages of threshold exceedances can be computed for different sized neighbourhood areas; the FSS uses the fraction of pixels within a patch of size N that are positive in each map, and computes a skill score from the mean squared difference of this fraction between the observed and predicted grid over all possible patches. In other words, it measures the average overlap between N-sized patches of the two binary grids (Fig. 2). In the extreme case where N is equal to the size of the image, the FSS behaves similarly to the Dice score, which was used as the loss function for the original UNet (Ronneberger et al., 2015). In our implementation, we selected $N = 15$, following a coarse grid search in the range [9, 21].

To use the FSS as a loss function for a neural network, we adjust it in two ways: first, we invert it so that FSS = 1 indicates no match between the observed and predicted grids (thus yielding the greatest possible gradient); second, we make it differentiable by replacing the hard threshold Q by an *arctan* function of Q (Eq. (1)), followed by a Gaussian filter and a normalization into the range [0, 1]. Cumulatively, these steps bring the values in each grid close 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 as FSS′.

$$binary\ \ map = arctan(input\ \ grid - percentile(input\ \ grid, Q)) \tag{1}$$

Finally, we define the compound loss as follows:

$$L_{cmpd} = 0.7(FSS'_{Q=80} + FSS'_{Q=95} + FSS'_{Q=99}) + MSE \tag{2}$$

where the weights of the two terms (0.7 and 1) were found empirically by searching in the range [0.1, 1] for both, with 0.1 increments. In the compound loss, the FSS′ term is actually the sum of three values of FSS′ calculated for different percentile thresholds (80th, 95th and 99th percentile values for each image). This was done in an attempt to teach the model to adjust the localization of rainfall peaks of different intensities, rather than for a single threshold. The use of percentile thresholds in the FSS′ aims to discard the influence of any biases in rainfall intensities and specifically concentrate on spatial accuracy, while the MSE component is used for adjusting pixel-level biases.

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

[2] https://www.gloh2o.org/mswep/

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

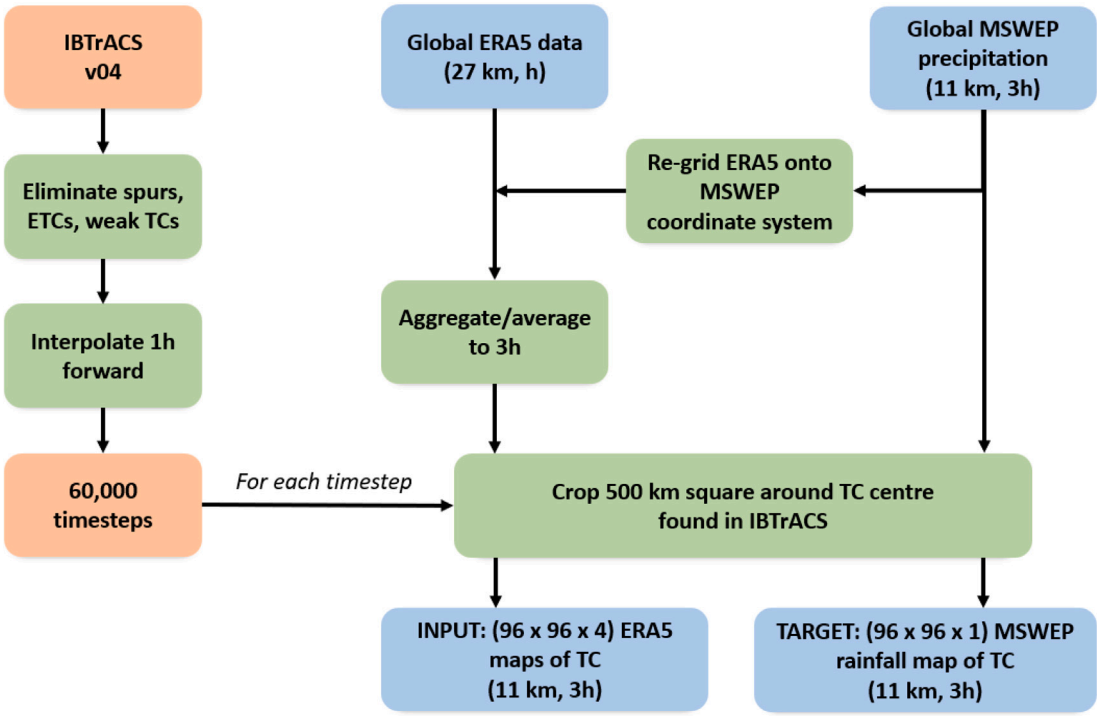

Fig. 1. Flowchart of the data pre-processing steps taken to prepare our dataset.

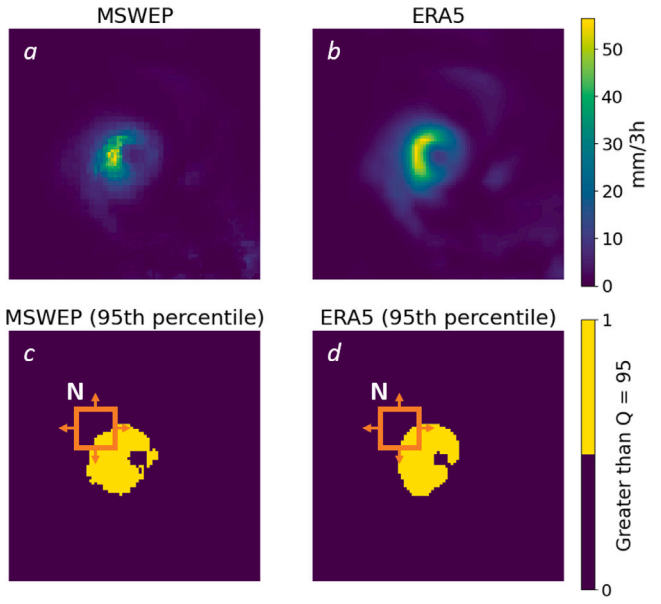

Fig. 2. An example of how the Fractions Skill Score (FSS) is computed. First, the two maps to be compared (panels *a* and *b*) are converted into binary maps (panels *c* and *d*) by setting all pixels above a threshold Q to 1 and all others to 0. Then, for each map, the method calculates the fraction of pixels above Q within a patch of size N (the orange rectangle in panels *c* and *d*). It then repeats this calculation by moving the patch N over all possible locations in the image (i.e., for every grid cell). The MSE for all fractions of all patches of the two images is then calculated and normalized with respect to a low-skill Roberts and Lean (2008), yielding the value of the FSS. We would expect the FSS to be close to 1 in the example in this image, as the two binary maps are quite similar.

2.3. Network implementation

In the original UNet (Ronneberger et al., 2015), 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 up-sampling 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 via so-called "skip connections" (which concatenate a copy of the encoder's block with the corresponding decoder block) to facilitate the transfer of spatial information to the semantically complex layers of the decoder.

The Residual Attention UNet (RA-UNet, or RA-U for short) (Jin et al., 2020) augments the original UNet in two ways: it replaces the convolutional blocks with residual blocks, and it adds attention modules along the skip connections (Fig. 3). Residual blocks enhance the learning process by enabling a direct flow of gradients through the network, mitigating the vanishing gradient problem commonly encountered in deep networks. Attention modules, conversely, augment the model's capability to focus on salient features within the input data. By incorporating a gating mechanism, these modules selectively emphasize important features and suppress less relevant ones, thus improving the feature representation power of the network. Integrating residual blocks and attention modules into the UNet architecture significantly elevates its performance, particularly in tasks requiring precise localization and feature delineation.

Contrary to the original implementation of RA-U, we chose to omit batch normalization and dropout, as our model showed no sign of overfitting (see the Results section). We trained our model using the Adam (Kingma and Ba, 2014) optimizer and early stopping (which triggered after about 30 epochs) on an NVIDIA A100 GPU. The hyper-parameters of the network are reported in Table 1. For an explanation of how these parameters affect the training of the network, the reader can refer to Kingma and Ba (2014) and Liu et al. (2022).

2.4. Performance evaluation

To disentangle the skill of the compound loss from the capacity of RA-U, we implemented also two other loss functions: a standard MSE, and another composite loss function proposed by Hess and Boers (2022), who also developed a UNet model for the bias adjustment of

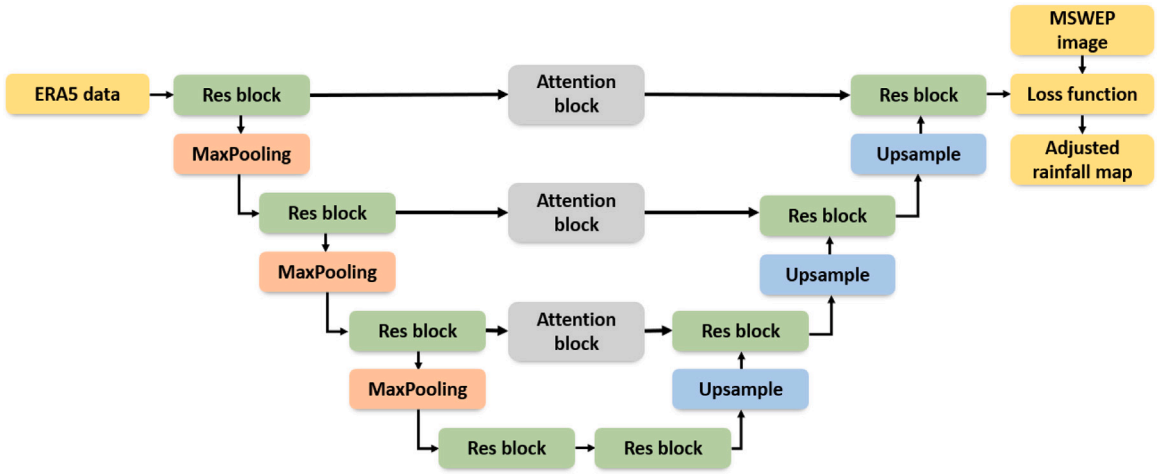

Fig. 3. Schematic representation of the architecture of RA-U, using four residual blocks in each branch. In our network three residual blocks are used per branch (see Table 1).

Table 1
Hyperparameters of our implementation of RA-U.

Hyperparameter	Value
Initial learning rate	0.0001
Adam β_1	0.9
Adam β_2	0.999
Batch size	32
Res blocks filters	[32, 64, 128, 256]

rainfall maps. Their loss, which we will refer to as L_{Hess}, is defined as:

$$L_{\text{Hess}} = \frac{\lambda}{N}\left(\sum_{i=1}^{N} w(y_i)(y_i - \hat{y}_i)^2\right) + (1-\lambda)\text{MS-SSIM}(y, \hat{y}) \tag{3}$$

where $\hat{y}_i$ and y_i are the predicted and target maps, λ weighs the two terms of the loss, MS-SSIM is the Multiscale Structural Similarity Index (Wang et al., 2003), i.e., a measure of the similarity of the patterns in two images (incorporating image details at different resolutions), and $w(y_i)$ is a function that weighs the MSE component, defined as:

$$w(y_i) = \min(\alpha e^{\beta y_i}, 1) \tag{4}$$

with parameters α and β controlling the slope of the exponential. In the original implementation, Hess and Boers used $\alpha = 0.007$, $\beta = 0.048$, and $\lambda = 0.158$. These values of α and β make the exponential saturate at around 100 mm/3 h of rainfall and assign very small weights (< 0.1) to rainfall pixels below 50 mm/3 h.

Although their formulas are different, L_{cmpd} and L_{Hess} share a key similarity: they both have a term to adjust local biases considering a pixel-wise metric and one to adjust of spatial patterns. However, whereas L_{cmpd} targets the spatial patterns only of peaks and high-rainfall regions, L_{Hess} attempts to correct the entire image. These similarities and subtle differences make L_{Hess} an excellent benchmark against which to evaluate the effectiveness of L_{cmpd}. Furthermore, Hess and Boers showed that their U-Net model trained with this loss function outperformed traditional statistical downscaling methods (quantile mapping, ridge regression) on global heavy precipitation events, obtaining an MSE of 1.368, compared to 1.473 for ridge regression and 2.071 for quantile mapping.

We also implemented L_{MSE} as a naive comparison, which we would expect both other losses to outperform, especially in terms of spatial accuracy. Therefore, we trained three identical RA-U models using the three different loss functions (RA-U$_{MSE}$, RA-U$_{Hess}$, RA-U$_{cmpd}$), and evaluated their performance on the test set using the following nine metrics (all considering MSWEP maps as the reference, and all reported as average values on the test set): (1, 2, 3) FSS′ for the three percentile thresholds Q used in L_{cmpd}, (4) MAE, (5) MSE, (6) spatial correlation, (7) absolute delta bias (i.e., the absolute value of the difference between total rainfall in the model output and the corresponding MSWEP maps), (8) percent delta bias (the same as absolute delta bias, but taking the signed percent error instead of the absolute value), (9) the distance between the highest peak (i.e., pixel with highest intensity) in the model output and the peak in the MSWEP map, and (10) and the Symmetric Mean Absolute Percentage Error (SMAPE), a metric based on relative errors, frequently used for forecast evaluation (Hewamalage et al., 2023) and corresponding to a modified version of MAPE, which is often used for rainfall forecast evaluation (Jolliffe and Stephenson, 2012; Ramos et al., 2021). We use SMAPE to describe by what percentage the peaks of ERA5 and RA-U differ from those of MSWEP. It is defined as:

$$SMAPE = \frac{\sum_{t=1}^{n} |F_t - A_t|}{\sum_{t=1}^{n} (A_t + F_t)} \tag{5}$$

where F_t and A_t are the forecast (in our case, the peak from a given RA-U or ERA5 map) and observed (in our case, the peak from a given MSWEP map) values at time t. Compared to other metrics like the root mean squared error (RMSE), SMAPE has useful properties like robustness to outliers (Hewamalage et al., 2023) and symmetry (i.e., it treats negative and positive errors equally, in the version reported in Eq. (5)); it is also expressed as a percentage, making it easier to interpret. However, if an ERA5/RA-UNet map had a peak of equal magnitude to that in the corresponding MSWEP image but in a very distant location in the map, this metric would be 0, erroneously. To prevent this, we define a new version of this metric, *nSMAPE*, using a neighbourhood approach. For a given MSWEP image, we locate the peak of rainfall. Then, in the corresponding ERA5/RA-U image, we draw a box of side n = 3 around that pixel (i.e., we select the 9 pixels around and including it), and compute SMAPE within that neighbourhood, thus discarding strongly displaced peaks.

As our model treats input maps in isolation, without using information regarding the temporal evolution of a TC, we investigated if the model corrected not just the single maps, but also the overall climatology of TC-induced rainfall. To do so, we accumulated the rainfall of each TC over its entire lifetime, and averaged the accumulated maps of all TCs in the test set. We then calculated the differences between the average accumulated maps of MSWEP and the models tested to show the differences between the climatologies. We then repeated this analysis stratifying our data first based on the Saffir-Simpson scale of intensity, and then based on whether the TCs had made landfall or not.

Finally, also compared the accumulated maps of the models tested and MSWEP for some particularly wet TCs:

- Cyclone Gamede (2007), the second wettest tropical cyclone in history and the wettest in the data period used in this study.
- Hurricane Harvey (2017), the wettest tropical cyclone to ever hit the United States, which caused catastrophic flooding in Texas and Louisiana, killing over 100 people.
- Typhoon Morakot (2009), the deadliest TC to impact Taiwan in recorded history, killing over 600 people (Lee et al., 2011; Wang et al., 2022).
- Typhoon Lekima (2019), the third costliest TC in Chinese history.

2.5. Error analysis

Finally, we performed an error analysis to understand under which conditions RA-U$_{cmpd}$ improved or not ERA5 the most. For this purpose, we constructed a matrix with N columns and M rows, where columns 1 to N-8 are various features extracted from the data and the last 8 columns are the improvement of the metrics mentioned in the previous section (e.g., we define the improvement in MSE as *MSE(ERA5) - MSE(RA-U$_{cmpd}$)*); the M = 14,000 rows are the data points in the test set. Specifically, we considered the following features:

- SEASON: year in which the TC occurred (from IBTrACS).
- LAT: latitude of the TC (from IBTrACS).
- LON: longitude of the TC (from IBTrACS).
- WMO_WIND: wind speed of the TC (from IBTrACS).
- DIST2LAND: distance of the TC from land (from IBTrACS).
- STORM_SPEED: translational speed of the TC (from IBTrACS).
- rain_peak_diff: difference between the highest-valued pixel in ERA5 and MSWEP.
- rain_tot_diff: difference between the total precipitation in ERA5 and MSWEP.
- asymmetry_diff: difference in asymmetry between ERA5 and MSWEP. We quantified asymmetry for a given image by dividing the image into four quadrants, computing the total precipitation in each quadrant, and taking the asymmetry of the TC to be the standard deviation of the total precipitation in the four quadrants.
- area_diff: difference in area of rainfall between ERA5 and MSWEP, computed by counting all pixels in an image that are above 1 mm/3 h.
- area_heavy_diff: difference in area of heavy rainfall between ERA5 and MSWEP, computed by counting all pixels in an image that are above 30 mm/3 h.

Given this matrix, we constructed a Random Forest that estimates, separately for each metric of improvement (i.e., the last eight columns of the matrix), how much RA-U$_{cmpd}$ improved ERA5 by using the features of the TC as predictors. For each Random Forest thus trained, we recorded the performance in terms of the R^2 of its prediction. Then, for each metric, we computed the *feature importance* of each feature using permutations, a technique less prone to inflating the importance of numerical features than the impurity-based feature importance of random forest (Breiman, 2001). First, we re-trained the Random Forest by randomizing one feature column at a time and recording its R^2. Randomization was performed by randomizing the order in which the values of each data point (i.e., row of the matrix) appeared in the column, not by actually generating new values sampled from a random distribution. We repeated this process ten times per feature and averaged the results, yielding the average decrease in R^2 when a feature is randomized. This was taken to represent the importance of the feature for the given metric: the higher the value, the more the considered feature explains why RA-U$_{cmpd}$ improved ERA5 under this metric.

We then further explored a subset of feature importances that we found to be of particular interest. To do so, we plotted the partial dependence plots for these features, which show how the dependent variable (i.e., a given metric of improvement) varies as a function of a given feature.

3. Results and discussion

3.1. Training curves

The first key difference between the three networks tested can be seen in their training plots showing the evolution of the loss over the training period (Fig. 4): RA-U$_{cmpd}$ is more stable and less susceptible to overfitting. The amount of overfitting can be gauged by comparing the training and validation curves for each network: the further apart they are, the more overfitting there is, meaning the network performs better on images it has seen during training than on new ones. Indeed, RA-U$_{cmpd}$ shows negligible overfitting, indicating that the representation it learns can generalize well to images it has not seen before and that the network's capacity is fully exploited. This is not the case for RA-U$_{Hess}$ and RA-U$_{MSE}$, which show substantial overfitting and have unstable validation loss curves. This overfitting could be reduced by adding a "regularization" component to the loss function, which attempts to constrain the values of the parameters of the network to a desired range. However, there is no guarantee it would reach levels as low as for RA-U$_{cmpd}$. Furthermore, regularizing a neural network is time-consuming and difficult. Therefore, RA-U$_{cmpd}$ has the advantage over RA-U$_{Hess}$ and RA-U$_{MSE}$ of being useable off-the-shelf.

3.2. Network performance evaluation

The results in Table 2 and Fig. 5 show that RA-U$_{cmpd}$ improves ERA5 under all metrics by 3%–28% (statistically significant at $p<0.01$), except for nSMAPE, for which it is slightly worse: on average, the peaks in maps from RA-U$_{cmpd}$ have a 3% greater error than those from ERA5. Repeating the training of our network with multiple configurations of our loss function revealed that there is a tradeoff between FSS and absolute delta bias on one hand, and nSMAPE (and in fact any other metric of peak magnitude error, even with different values of n) on the other. Therefore, our network greatly reduces all other biases (and by a larger margin than the other benchmarks tested) at the cost of a slightly increased nSMAPE bias. Nevertheless, in many of the figures shown in the next sections RA-U$_{cmpd}$ appears to have a lower peak magnitude error than ERA5, especially when looking at accumulated maps of particularly wet TCs, confirming that the method is valid under extreme precipitation scenarios under all metrics considered here. Indeed, RA-U$_{cmpd}$ has lower nSMAPE than ERA5 for 41% of the samples in the test set.

A key takeaway from Table 2 and Fig. 5 is that RA-U$_{MSE}$ and RA-U$_{Hess}$ aggravate the percent and absolute delta bias of ERA5, whereas RA-U$_{cmpd}$ reduces it. RA-U$_{Hess}$ in particular worsens the absolute delta bias of 19% with respect to ERA5, corresponding to a marked increase in the total rainfall overestimation. This is also clearly visible from the accumulated plots (Fig. 6), in which RA-U$_{Hess}$ appears to introduce a strong bias in the eye region. A bias in a similar region is observed also for RA-U$_{MSE}$ and RA-U$_{cmpd}$, but of a much lower magnitude (Fig. 6). Indeed, RA-U$_{cmpd}$ (and to a similar degree RA-U$_{MSE}$) is able to adjust not only single maps in isolation, but the overall climatology of TC-induced precipitation, with a moderate bias only in the eye region. We attribute this large difference between RA-U$_{cmpd}$ and RA-U$_{Hess}$ to the fact that L_{Hess} uses a weighting function for the MSE component that assigns very small (< 0.1) weights to pixels below 50 mm/3 h. Considering that only about 1% of the images in our dataset have pixels above 50 mm/3 h, the MSE component of L_{Hess} effectively ignores most of the pixels in our dataset. Therefore, RA-U$_{Hess}$ optimizes primarily the MS-SSIM component of L_{Hess}, which is able to preserve and even correct spatial features (as highlighted in Fig. 8 and by its excellent performance in the spatial metrics reported in Table 2) but not large pixel-wise biases. By contrast, L_{cmpd} assigns equal weight to all pixels in the calculation of the MSE, and is therefore better able to resolve pixel-wise biases. However, this greater weight placed on the MSE component also means that its blurring effect, widely documented

Table 2
Metrics averaged on the test set, comparing ERA5 with the post-processed output of RA-U trained with the three losses considered. The best values for each metric are reported in bold.

Metric name	Units	ERA5	RA-U$_{MSE}$	RA-U$_{Hess}$	RA-U$_{cmpd}$
MAE	mm/3h, per pixel	1.34	1.18	1.20	**1.15**
MSE	(mm/3h)^2, per pixel	11.53	8.57	9.97	**8.27**
$FSS'_{Q=80}$	–	0.10	0.09	0.09	**0.08**
$FSS'_{Q=95}$	–	0.30	0.27	0.28	**0.25**
$FSS'_{Q=99}$	–	0.51	0.39	0.45	**0.37**
Spatial correlation	–	0.80	0.83	**0.86**	**0.86**
Percent delta bias	–	6.97	7.18	8.54	**4.08**
Absolute delta bias	mm/3h, per image	3134	3311	3726	**3040**
Highest peak distance	km	150	135	**131**	134
nSMAPE	%	36.8	39.0	34.5	39.6

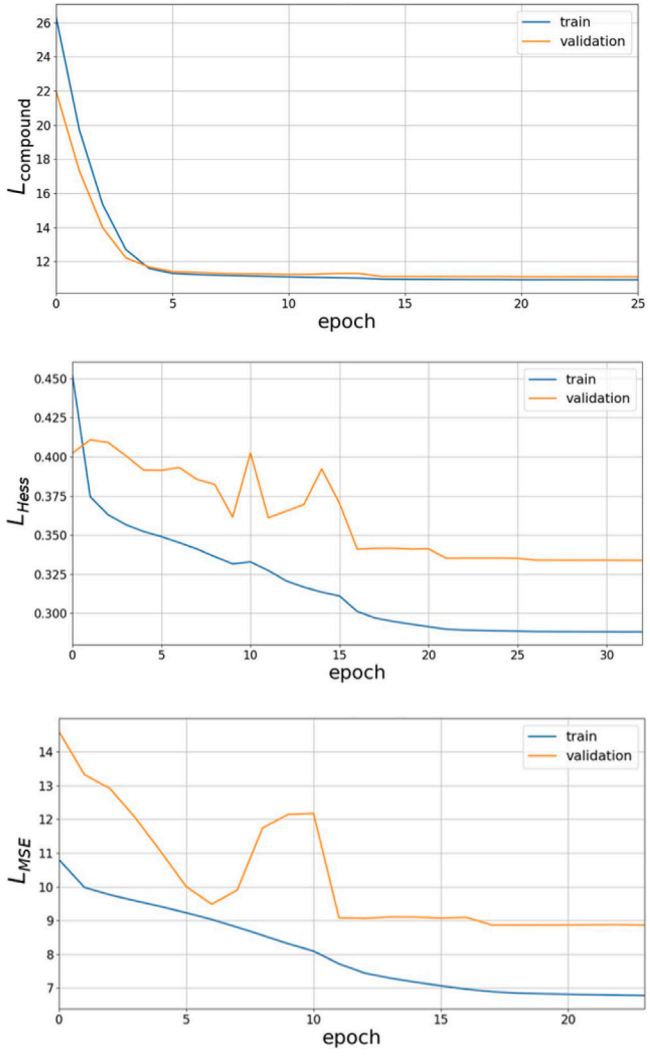

Fig. 4. Training curves for RA-U$_C$ (top), RA-U$_H$ (middle), and RA-U$_M$ (bottom), showing both training and validation loss curves over the training period, quantified as the number of epochs (i.e., times a model passes through the whole training dataset) until convergence. The ranges of the losses of the three plots are different due to the way the three losses are computed (Eqs. (2)–(3)). Furthermore, the number of epochs is different for each plot because we used early stopping during training: the training was forced to stop if the validation loss did not decrease for six consecutive epochs.

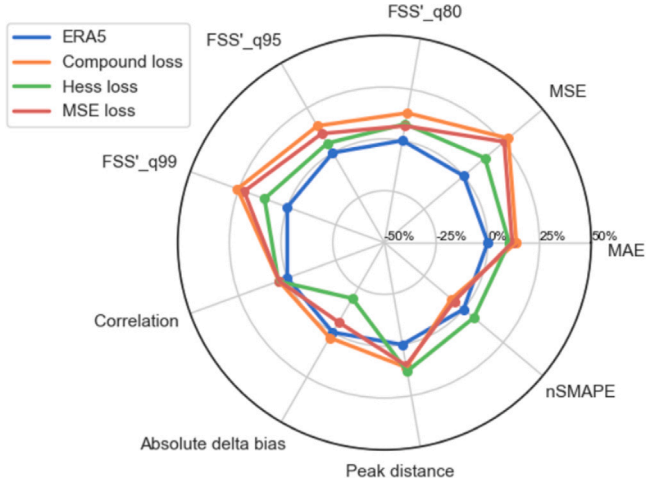

Fig. 5. Spider plot showing by what percentage each model improves ERA5 over each metric. The percent absolute delta bias metric is omitted here because a percentage of a percentage would be out of scale with the other axes; moreover, it is highly correlated with the absolute delta bias.

in the literature (e.g., Hess and Boers, 2022), is more pronounced for RA-U$_{cmpd}$ than for RA-U$_{Hess}$. This is particularly evident in Fig. 8, which shows that RA-U$_{cmpd}$ struggles to preserve spatial patterns when there are no well-defined peaks in ERA5. Indeed, RA-U$_{cmpd}$ seems to be most effective when ERA5 has peaks of rainfall that are clearly wrong in localization or intensity. In these cases, RA-U$_{cmpd}$ is able to adjust these biases, while RA-U$_{MSE}$ and RA-U$_{Hess}$ tend to aggravate them.

Another interesting finding from Table 2 is that RA-U$_{cmpd}$ has a lower MSE than RA-U$_{MSE}$. This may seem surprising, but is likely attributable to the fact that without a spatial component, L_{MSE} suffers fully from the blurring effect of the MSE. As it is easy for the network to learn that it just needs to blur out the inputs to reduce its loss, the model falls into a local minimum of sorts. This phenomenon is known as *shortcut learning* (Geirhos et al., 2020): it happens when a model finds a way to minimize its loss by doing something (in our case, automatically blurring the outputs) that was not the intended task (reducing pixel-wise and spatial biases). As RA-U$_{cmpd}$ considers also a spatial component in its loss function, it cannot take this shortcut and therefore needs to actually learn how to solve the true task. Thus, it achieves lower MSE by virtue of continuing training past the local minimum into which RA-U$_{cmpd}$ falls.

3.3. Notable cyclones

Having established that RA-U$_{cmpd}$ is the best performing of the models tested, we now show examples of how it performs for some notable TCs, of which we show the accumulated rainfall maps: Cyclone Gamede from 2007 (Fig. 9), Hurricane Harvey from 2017 (Fig. 10), Typhoon Morakot from 2009 (Fig. 11), and Typhoon Lekima from 2019 (Fig. 12). For Cyclone Gamede, RA-U$_{cmpd}$ displays the same small bias in the eye region that was observed also in Fig. 6, but it greatly reduces the large biases of ERA5 in terms of both total rainfall and localization of the peak. Indeed, the accumulated map of RA-U$_{cmpd}$ mimics that of MSWEP in having just one marked peak in the south-east portion

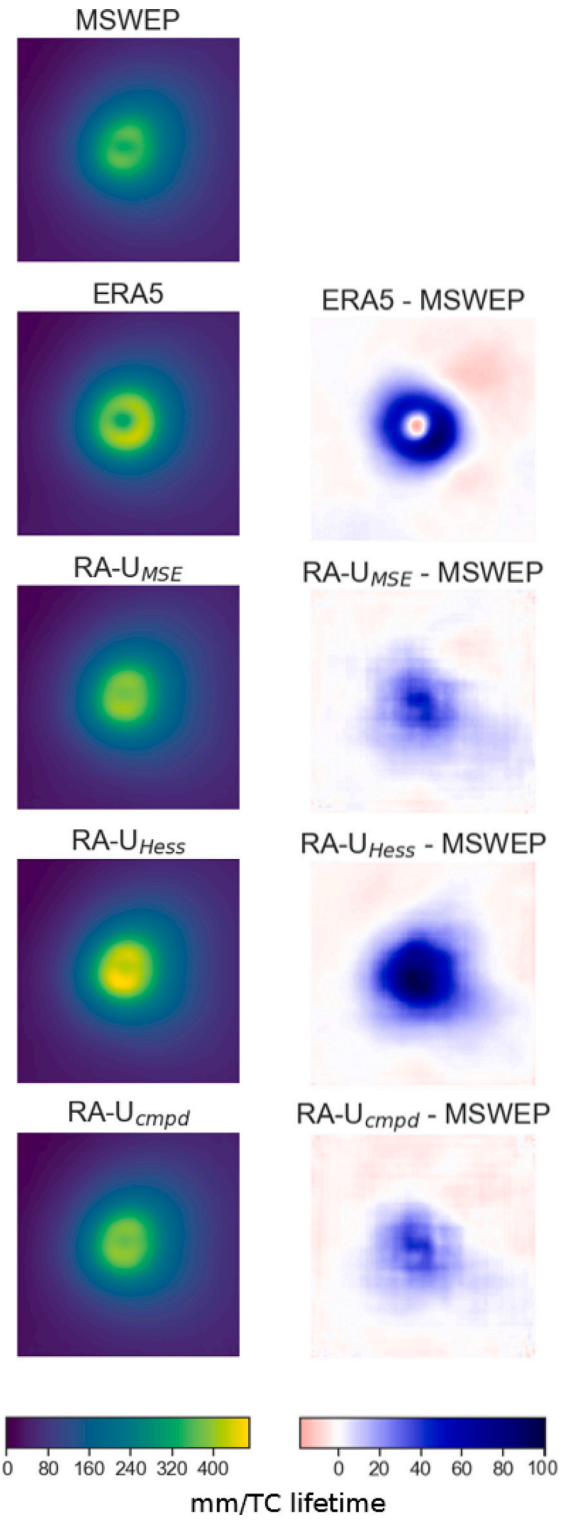

Fig. 6. Accumulated rainfall maps averaged over all TCs in the test set, and signed bias with respect to MSWEP.

of the eyewall, whereas ERA5 mistakenly assigns high values to the entire eyewall. RA-U$_{cmpd}$ is able to correct most of this spatial bias, preserving high precipitation only in the area where the observed peak is in MSWEP. For Hurricane Harvey (Fig. 10), the rainfall pattern is much less symmetrical than for Gamede. The maps of both ERA5 and RA-U$_{cmpd}$ reflect this, with biases that are not as clearly defined around or within the eye wall. Nevertheless, also for Harvey RA-U$_{cmpd}$ reduces the overall bias and manages to almost entirely remove one of the spurious peaks of rainfall, near the centre of the map. The accumulated maps for Cyclone Gamede and Hurricane Harvey support our previous analysis of the behaviour of RA-U$_{cmpd}$. For Cyclone Gamede, for which ERA5 has a strong bias along the eyewall and an exaggerated peak, RA-U$_{cmpd}$ reduces both types of bias well. For Hurricane Harvey, for which ERA5 greatly overestimates the peak and which does not have a well-defined structure, RA-U$_{cmpd}$ just tends to blur out the image, correcting some of the bias of ERA5 but to a lesser extent than for Cyclone Gamede. It should be noted that Cyclone Gamede was in the training set, while Hurricane Harvey was in the test set. However, given how RA-U$_{cmpd}$ has almost zero overfitting, we suspect this does not play a role here.

Finally, in ERA5 Typhoon Morakot has a strong negative bias in the eye region, whereas Typhoon Lekima has a strong positive bias. In both cases, RA-U$_{cmpd}$ is able to substantially reduce these biases, although, in the case of Typhoon Morakot, both ERA5 and RA-U$_{cmpd}$ are unable to reproduce the uniquely shaped 'striped' features near the eye.

3.4. Performance across saffir-simpson classes

To further investigate RA-U$_{cmpd}$'s performance, we stratify the results according to the Saffir-Simpson scale.[4] For the sake of simplicity, we focus on the following three macro-classes: equal to (or less than) Category 1, Category 2–3, and Category 4–5. Both the spider plot (Fig. 13) and the average rainfall maps for the three categories (Fig. 14) show how RA-U$_{cmpd}$ reduces ERA5 biases more for more intense TCs, especially in terms of absolute delta bias and peak localization. In fact, for Category 4–5 RA-U$_{cmpd}$ almost entirely removes the large ERA5 bias along the eyewall. A surprising result is that the positive bias RA-U$_{cmpd}$ displayed within the eye region for the accumulated maps (Figs. 6, 9) changes sign in the average map for Category 4–5 (which instead has a small bias around the eyewall) but is present for the other two categories. This may simply be due to the fact that ERA5 itself has a strong negative bias in the eye region for these TCs, which RA-U$_{cmpd}$ cannot fully resolve. Surprisingly, the same conclusions cannot be drawn for the nSMAPE metric (i.e., the error on the intensity of the peaks), which seems to deviate more (although by small amounts) compared to ERA5 for more intense TCs. However, although the relative performance of RA-U$_{cmpd}$ becomes worse, the absolute value of this bias is also smaller for these stronger TCs, and therefore likely not meaningful. Indeed, the previous section showed how for some very intense TCs (e.g., Fig. 9), RA-U$_{cmpd}$ is capable of reducing all biases of ERA5, including on the peak magnitude.

3.5. Performance for landfalling and non-landfalling TCs

Our results indicate that RA-U$_{cmpd}$ improves ERA5 less for landfalling TCs than for non-landfalling ones (Fig. 15). However, this is mainly because ERA5 better represents landfalling TCs than non-landfalling ones (cfr. the range of the bias in Figs. 16 and 17). Indeed, it is well documented that ERA5 precipitation has particularly severe biases over the tropical oceans (e.g., Hassler and Lauer, 2021; Lavers et al., 2022). This is particularly evident in the accumulated rainfall maps for non-landfalling TCs (Fig. 16), which show a marked bias for ERA5 near the eyewall and a less marked one for RA-U$_{cmpd}$ near the eye. The pattern changes substantially for landfalling TCs (Fig. 16), for which the structure is less symmetric—likely as a consequence of the TCs weakening as they make landfall (Zhao et al., 2021, 2022). Surprisingly, in this case RA-U$_{cmpd}$ does not have its usual bias in the eye region. Indeed, the absolute delta bias of RA-U$_{cmpd}$ is lower for landfalling TCs than for non-landfalling ones.

[4] https://www.noaa.gov/jetstream/tc-potential

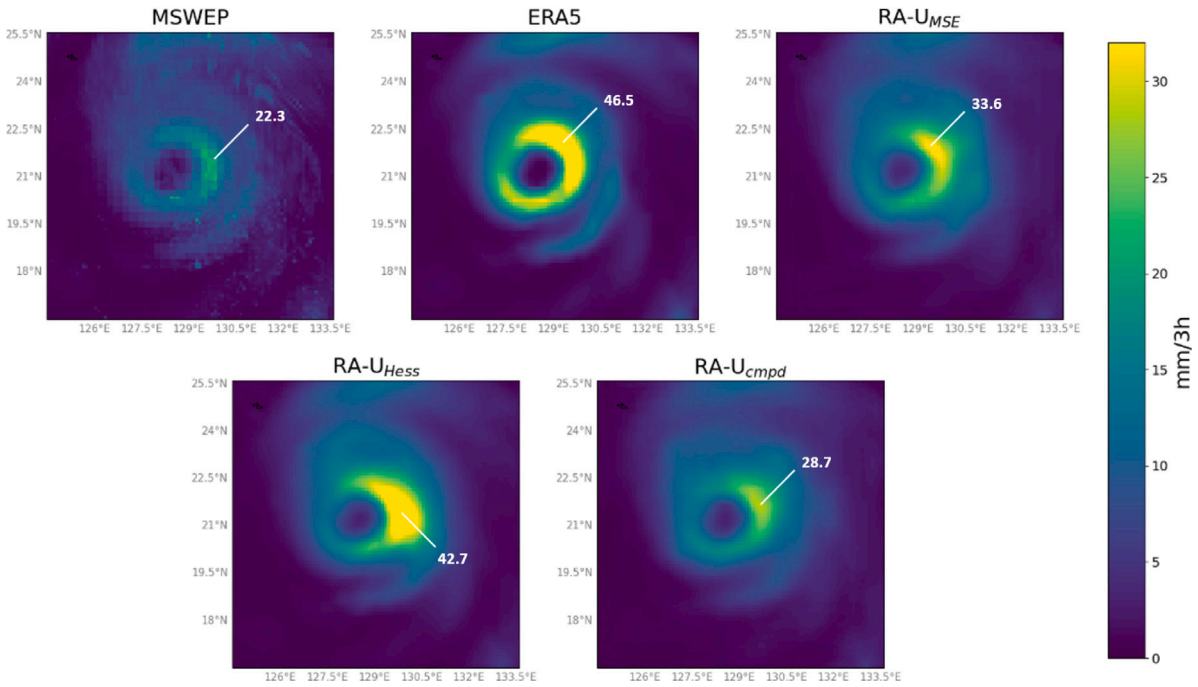

Fig. 7. A random rainfall map from the test set, on which RA-U$_{cmpd}$ performs better than the other two models by adjusting the bias of both total rainfall and peak localization, whereas the other two models aggravate both. The rainfall map is the snapshot of Typhoon Kong-rey (the most powerful TC of 2018) registered in IBTrACS at time 03/10/2018 07:30. In this plot, the colorbar is clipped to a value close to MSWEP's maximum; had we scaled it according to the maximum of all subplots (i.e., ERA5), the features in the MSWEP and RA-U$_{cmpd}$ maps would have been washed out. We report the unclipped figure in the Supplementary Materials(Fig. 20).

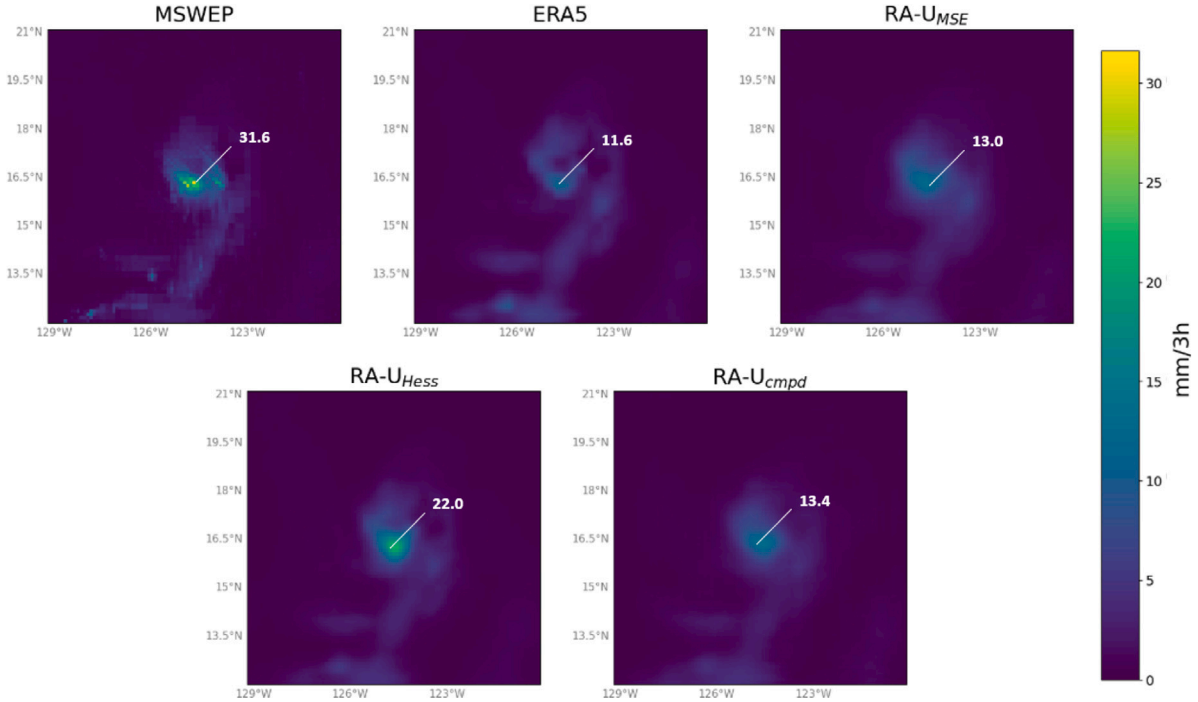

Fig. 8. A random rainfall map from the test set, in which all models perform relatively poorly. While RA-U$_{MSE}$ and RA-U$_{Hess}$ and able to somewhat preserve the pattern of rainfall, RA-U$_{cmpd}$ makes it slightly worse. The rainfall map is the snapshot of Hurricane Irwin registered in IBTrACS at time 30/07/2017 04:30.

3.6. Error analysis

The error analysis results, reported in the form of a heatmap (Fig. 18), reveal interesting patterns of how important different features were for helping overcome the biases of ERA5. First, none of the Random Forests achieved a high R^2 on any of the metrics of improvement considered. The Random Forests achieved the highest values for MAE (0.46) and MSE (0.45). This lack of predictive skill is also reflected by the fact that, for most metrics, most (and often all) features have importance near or at zero. However, two features are frequently identified as important: the differences between ERA5 and MSWEP in terms of total precipitation and in terms of asymmetry. In fact, rain total difference seems to be so important to explain the improvement of the metrics that randomizing it would result in a model

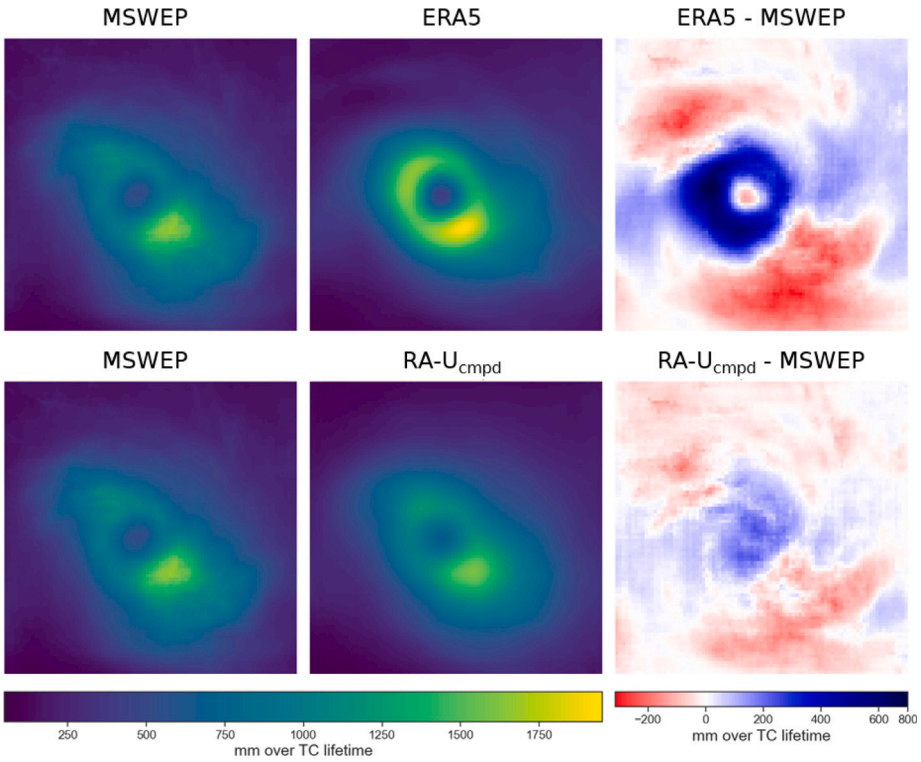

Fig. 9. Accumulated rainfall maps for Cyclone Gamede, averaged over all the timesteps of this TC present in our dataset.

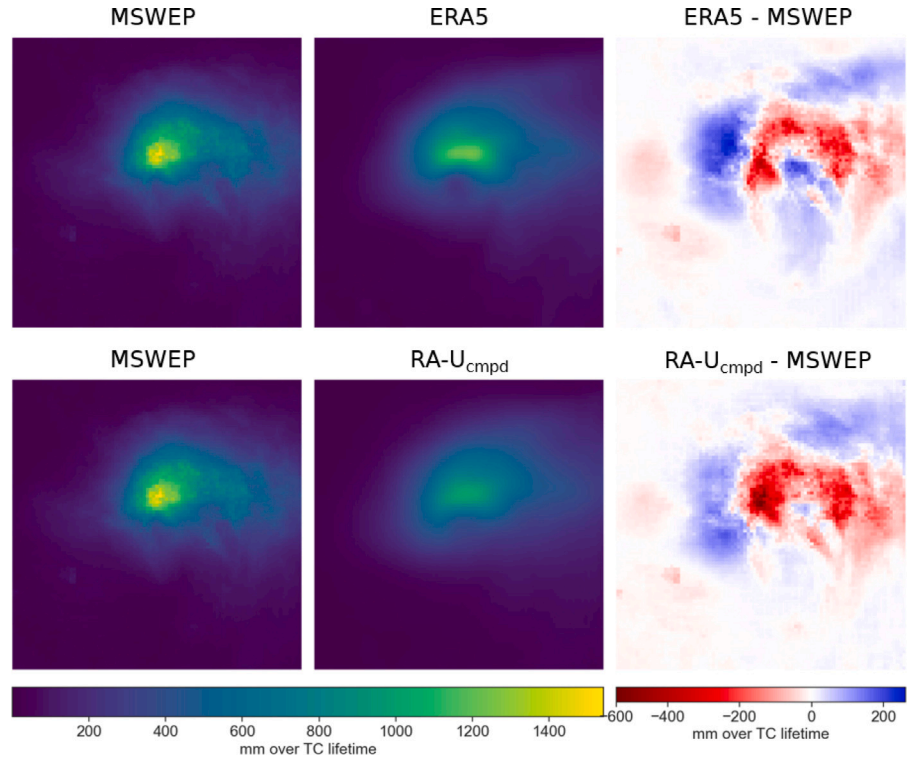

Fig. 10. Accumulated rainfall maps for Hurricane Harvey, averaged over all the timesteps of this TC present in our dataset.

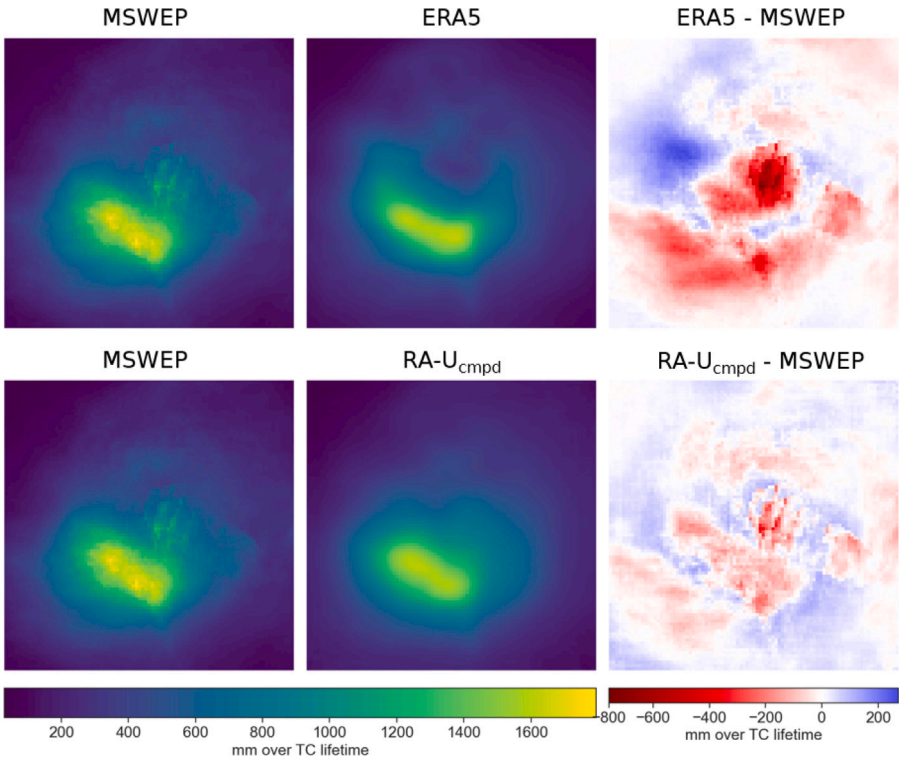

Fig. 11. Accumulated rainfall maps for Typhoon Morakot, averaged over all the timesteps of this TC present in our dataset.

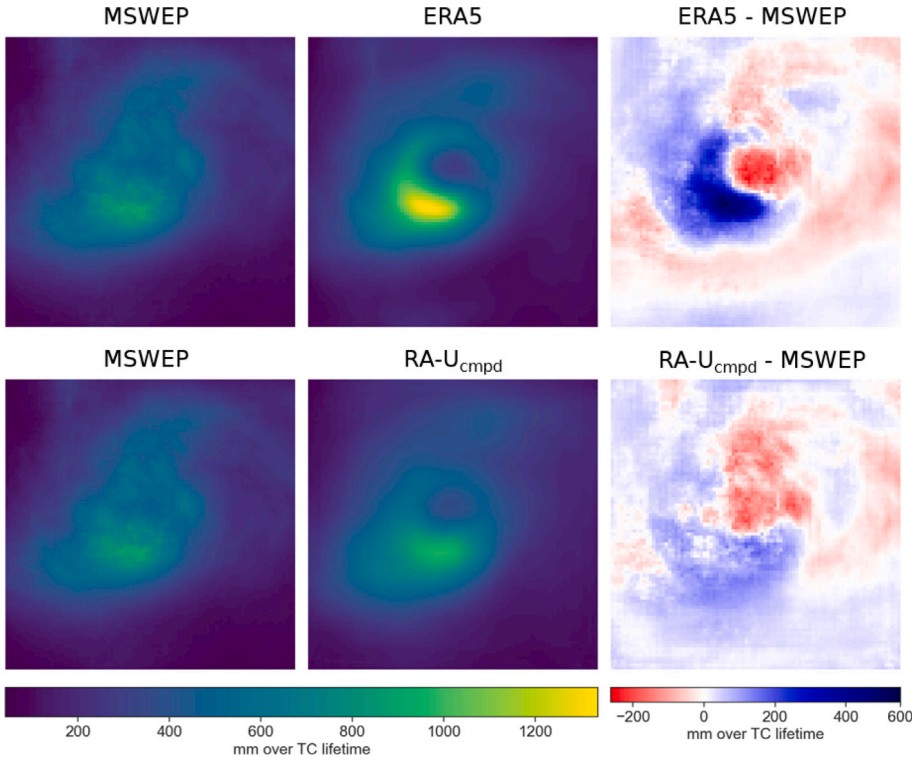

Fig. 12. Accumulated rainfall maps for Typhoon Lekima, averaged over all the timesteps of this TC present in our dataset.

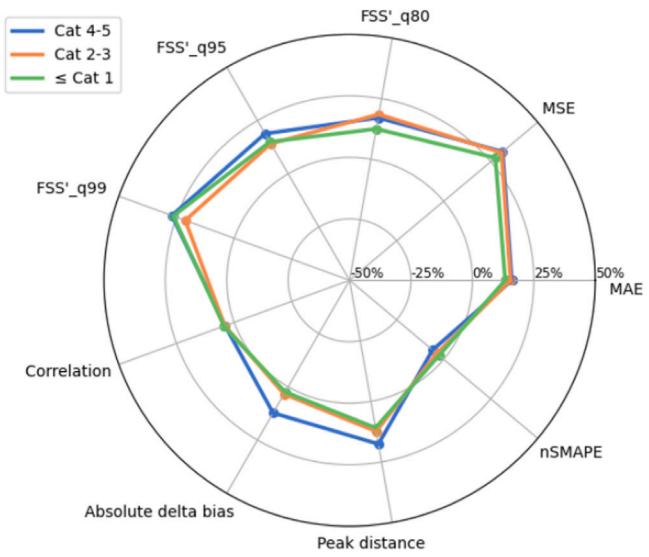

Fig. 13. Spider plot showing by what percentage RA-U$_{cmpd}$ improves ERA5 over each metric for different categories of TC intensity.

that performs worse than guessing at random (i.e., *feature importance* > 1). Also interesting is that the longitude seems important for improving MSE, but not for other metrics. Finally, rain peak difference appears to be important for reducing the pixel-wise biases of ERA5, in agreement with the observations derived from the previous results.

It would be impractical and redundant to inspect all 88 feature importances individually. Rather, we chose to explore further (Fig. 19) the importance of the most prominent features for the improvement in MSE (longitude, rain peak difference, total rain difference, and asymmetry difference), which are the ones that emerged as most interesting from the previous discussion. We do this exploration via partial dependence plots, which show how the improvement in MSE changes as the features change. For the longitude (Fig. 19, panel *a*) the curve drops around 120–150°West, which is in the middle of the Pacific ocean. Upon inspecting further the data points in this region, we discovered that they have significantly ($p<0.01$) different characteristics from those of the rest of the domain. In particular, TCs in this region overestimate less than average the total amount of rainfall (rain_tot_diff = 513 ± mm/image for TCs between 120–150°West and 1102 ± 100 mm/image for TCs in the rest of the domain) and underestimate less than average the area of rainfall (area_diff = 15 ± 20 pixels for TCs between 120–150°West and −152 ± 10 pixels for TCs in the rest of the domain). Therefore, given that these cyclones are better represented than average in ERA5, it is reasonable that the improvement brought by RA-U$_{cmpd}$ is only marginal. For rain peak difference (panel *b*) and asymmetry difference (panel *d*), the curve drops close to the 0 point on the *x*-axis. This is to be expected: if ERA5 has little to no bias, RA-U$_{cmpd}$ cannot reduce it by much. However, whereas the curve for asymmetry difference is relatively symmetric around $x = 0$, the one for rain peak difference is not, presenting a marked skew to the right instead. This confirms our previous analyses: when ERA5 has peaks of rainfall that are clearly much higher than they should be, RA-U$_{cmpd}$ can correct them effectively; on the other hand, it cannot correct well the opposite case of large underestimation of peaks. The same behaviour is observed for the curve for rain total difference (panel *c*).

3.7. Limitations and future work

There are a few avenues for future research to build upon our work. First, we only used four ERA5 variables as input. However, the large-scale ML models for climatology (e.g., PanguWeather (Bi et al., 2023), GraphCast (Lam et al., 2022)) that recently achieved massive success used several dozens, even hundreds of inputs, often considering several variables at more than thirty pressure levels. Had we done the same, we would have needed to substantially increase the size and complexity of RA-U$_{cmpd}$, which in turn would likely have caused overfitting, as our dataset only has 60,000 samples. Although regularization methods could be introduced that we did not use (dropout, batch normalization), it is unclear whether such a deep network could be trained at all on so few samples. Therefore, a likely better option is to use a feature selection algorithm to select which (and how many) variables to train the model. The selection algorithm would need to be able to account for the fact that there is a complex relationship between the optimal number of inputs, the size of the network, and the likelihood of overfitting due to the limited size of the dataset.

We also identified a trade-off between correctly localizing peaks of rainfall and estimating their magnitude. Our solution greatly reduces errors in the first at the cost of slightly increased errors in the second. However, while it is likely possible to apply post-processing algorithms to the outputs of our model to fix these biases, the same cannot be said for biases in peak localization, which are not so easily resolved.

Another interesting future research direction is applying RA-U$_{cmpd}$ to datasets different from ERA5. Most prominently, the capability of RA-U$_{cmpd}$ to adjust pixel-wise and spatial biases makes it an especially promising tool for enhancing TC-induced rainfall forecast products, such as the ECMWF's medium-range High-Resolution (HRES) and the Sub-seasonal to Seasonal (S2S) ensemble forecast products; in particular, we expect RA-U$_{cmpd}$ to be suitable to be applied to HRES and S2S as these operational forecasts are produced with the same ECMWF Integrated Forecasting System (IFS) that is also used to produce ERA5. The real-world implications of this are profound, as even minor inaccuracies in the spatial distribution of predicted rainfall can lead to significant errors in flood forecasting, potentially underestimating or overestimating the flood risk in different areas. Improving the predictions of where and how much rain will fall can inform more effective water management and disaster risk reduction strategies, such as reservoir management, anticipatory actions and flood defence planning.

4. Summary and conclusions

While bias correction of ERA5 precipitation data has been relatively explored in the literature, bias correction of tropical cyclone-induced rainfall has not yet been addressed. Additionally, most existing works on bias adjustment focus on adjusting pixel-level or image-level metrics of bias by computing the average per-pixel or per-image difference in rainfall amounts compared to some reference. However, it is equally important to ensure that the rainfall peaks are correctly located within the rainfall maps, especially if these maps are then used as inputs to hydrological models. In this paper, we contribute a novel machine learning model that addresses both gaps, RA-U$_{cmpd}$. The key novelty of RA-U$_{cmpd}$ is its loss function, the *compound loss*, which optimizes both a pixel-wise metric (the MSE) and a spatial verification metric (the FSS′). The structure of the compound loss is similar to that of the loss proposed by Hess and Boers (Hess and Boers, 2022), i.e., RA-U$_{Hess}$, which we use as a comparison along with a simple MSE loss, but we introduce a new spatial verification component based on a modified version of the Fractions Skill Score and percentile thresholds. Our results indicate that RA-U$_{cmpd}$ clearly outperforms the other two models tested, and can effectively reduce the ERA5 bias in terms of amounts of rainfall and localization of rainfall peaks. This comparison underscores the importance of having an effective and general loss function to achieve the desired task; our compound loss reduces overfitting and allows generalization across various cases, thanks to a parameterization based on percentile thresholds, and avoiding weighting schemes of the bias component based on fixed values. By analysing the behaviour of RA-U$_{cmpd}$ for different types of events (including particularly wet

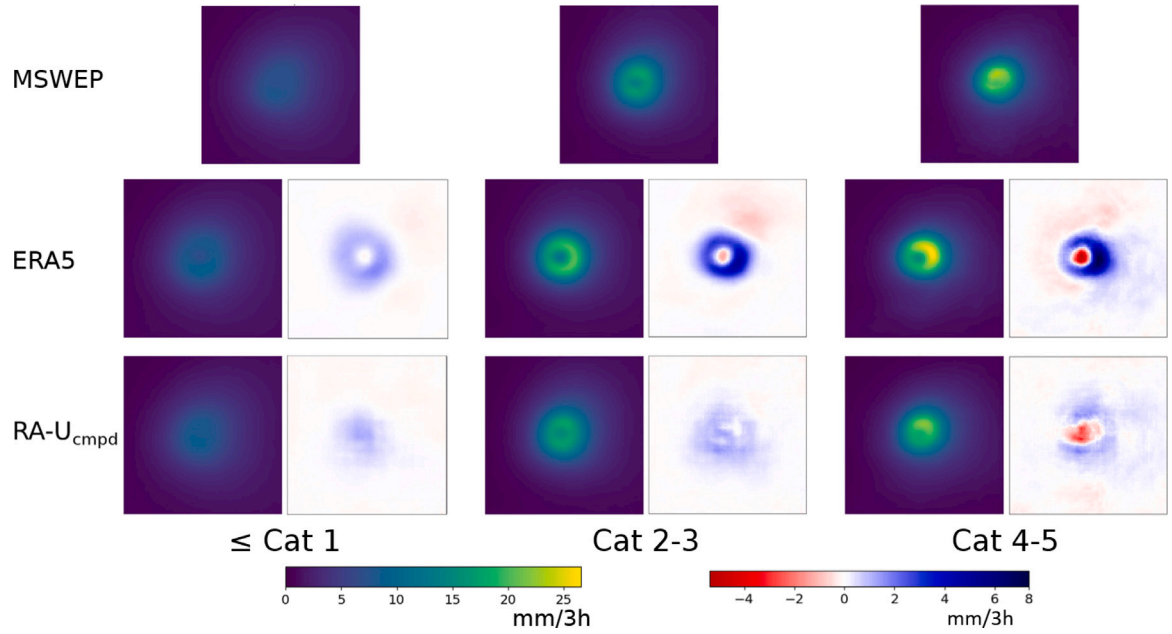

Fig. 14. Average rainfall maps of MSWEP, ERA5, and RA-U$_{cmpd}$ for different categories of TC intensity. Next to the ERA5 and RA-U$_{cmpd}$ plots are the error maps relative to the MSWEP maps (e.g., ERA5 - MSWEP). All maps are obtained by averaging over all instances of TCs belonging to a given category of intensity.

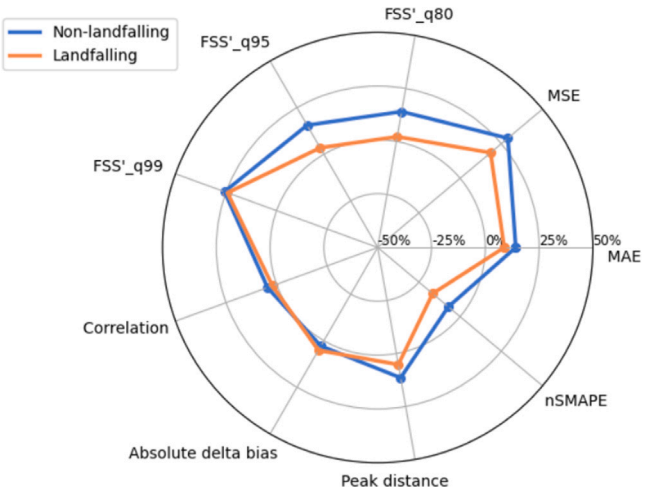

Fig. 15. Spider plot showing by what percentage RA-U$_{cmpd}$ improves ERA5 over each metric for different categories of TC intensity.

or intense tropical cyclones), we found that RA-U$_{cmpd}$ performs particularly well in cases where ERA5 has a clearly misplaced peak of rainfall and when the eye structure of the cyclone is well defined. This finding was corroborated by an error analysis, which indicated that RA-U$_{cmpd}$ improves ERA5 more when ERA5 has strong biases in terms of peak, total amount of rainfall and TC-rainfall shape asymmetry. Finally, we found that RA-U$_{cmpd}$ generally performs better for more intense cyclones (especially those of category 4–5 on the Saffir-Simpson scale), likely because they have more well-defined eyes. However, for these TCs the metric that quantified the error of the peak's magnitude was slightly worse than for weaker TCs. Future work should focus primarily on testing and tailoring our RA-U$_{cmpd}$ model to adjust biases of TC-induced rainfall forecasts and assess its value for early warning and anticipatory actions.

CRediT authorship contribution statement

Guido Ascenso: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Andrea Ficchì:** Writing – review & editing, Validation, Software, Methodology, Conceptualization. **Matteo Giuliani:** Writing – review & editing, Supervision. **Enrico Scoccimarro:** Writing – review & editing, Supervision, Resources. **Andrea Castelletti:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: This research was supported by the EU-funded Climate Intelligence (CLINT) project: [grant agreement ID: 101003876; DOI: 10.3030/101003876]. Andrea Ficchì also acknowledges support from the AXA Research Fund Fellowship on Coastal Livelihoods

Data availability

The raw data used during the study is freely available online:

- ERA5: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview
- MSWEP: https://www.gloh2o.org/mswep/
- IBTrACS: https://www.ncei.noaa.gov/products/international-best-track-archive

The processed data (including model outputs) used in the study will be made available online in the near future, as will be the code used to generate the models and loss functions described here.

Appendix. Supplementary materials

Unclipped version of Fig. 7

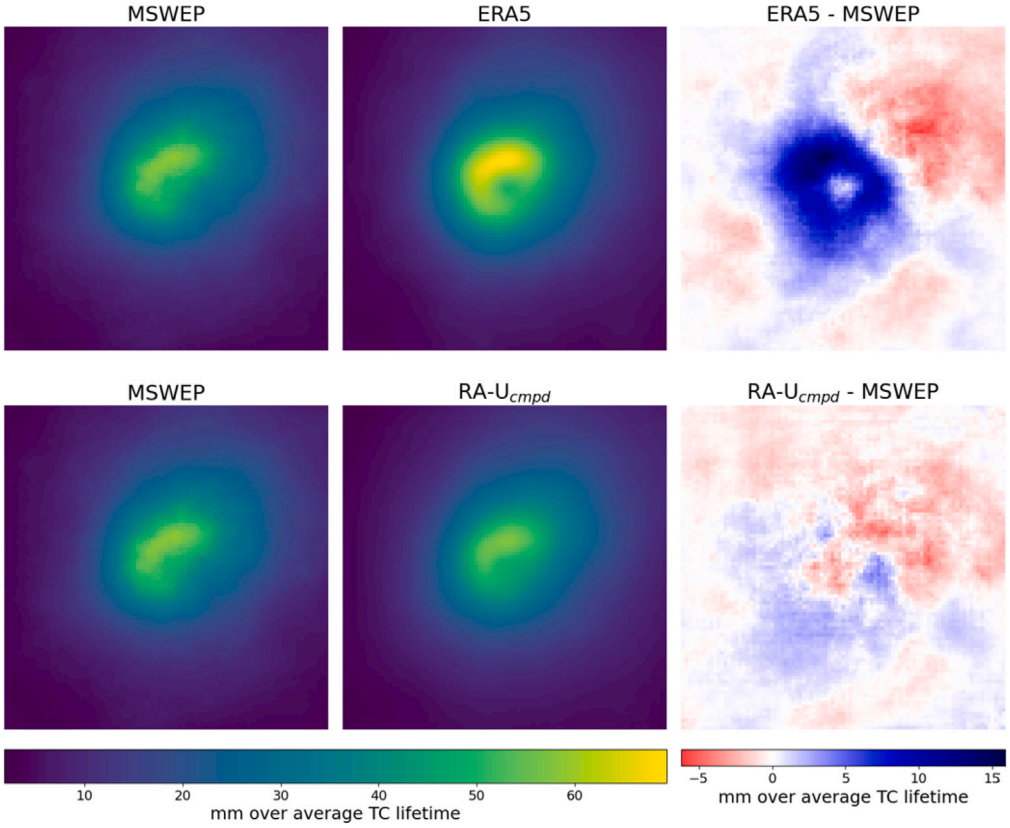

Fig. 16. Average rainfall maps of MSWEP, ERA5, and RA-U$_{cmpd}$ for landfalling TCs. Next to the ERA5 and RA-U$_{cmpd}$ plots are the error maps relative to the MSWEP maps (e.g., ERA5 - MSWEP). All maps are obtained by averaging over all samples for which IBTrACS reports DIST2LAND = 0.

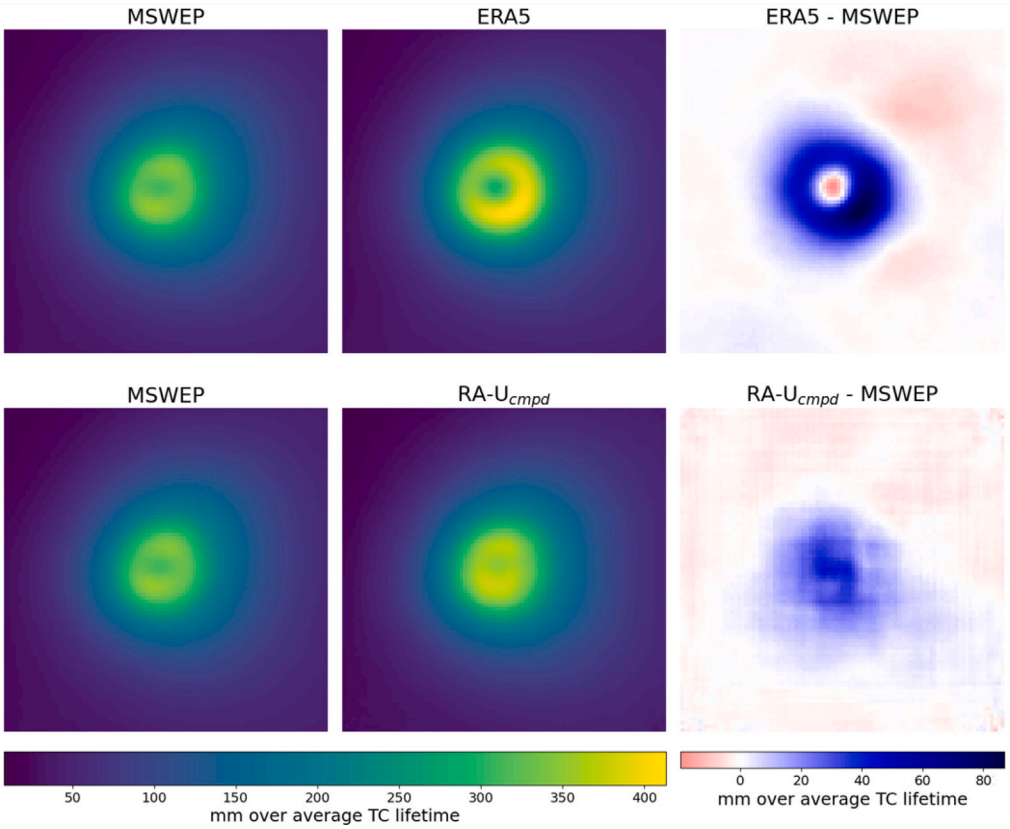

Fig. 17. Average rainfall maps of MSWEP, ERA5, and RA-U$_{cmpd}$ for non-landfalling TCs. Next to the ERA5 and RA-U$_{cmpd}$ plots are the error maps relative to the MSWEP maps (e.g., ERA5 - MSWEP). All maps are obtained by averaging over all samples for which IBTrACS reports DIST2LAND > 0.

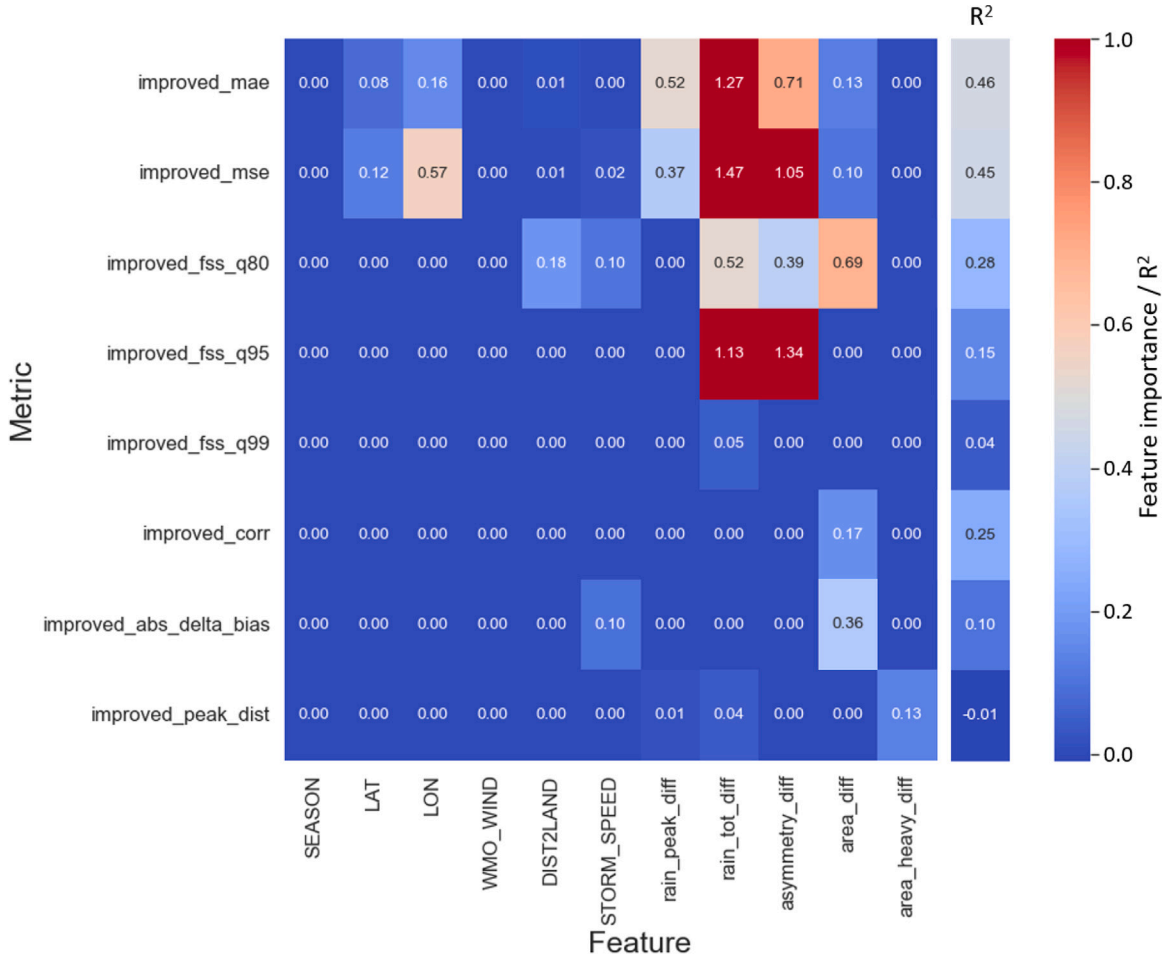

Fig. 18. Heatmap summarizing the results of the error analysis. Each column is a feature extracted from the data, and each row refers to a measure of how much RA-U$_{cmpd}$ improved ERA5. The value in each cell indicates how important that feature is for predicting (via Random Forest) the given metric. To the right of each row, we also report the R^2 value of the Random Forest trained for that metric. The colorbar was set to have a maximum value of 1 for clarity purposes and because the distinction between features that are exactly equal to or greater than one is irrelevant (both indicate that the Random Forest would have had $R^2 = 0$).

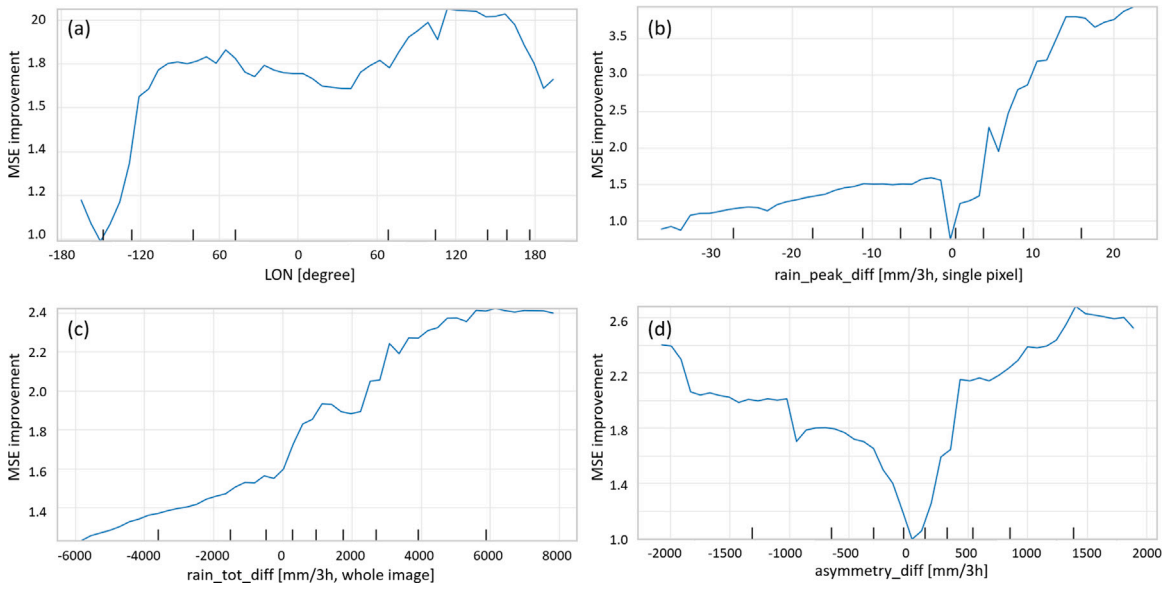

Fig. 19. Partial dependence plots for the improvement in MSE, focusing on four features: longitude (a), rain_peak_diff (b), rain_tot_diff (c), and asymmetry_diff (d).

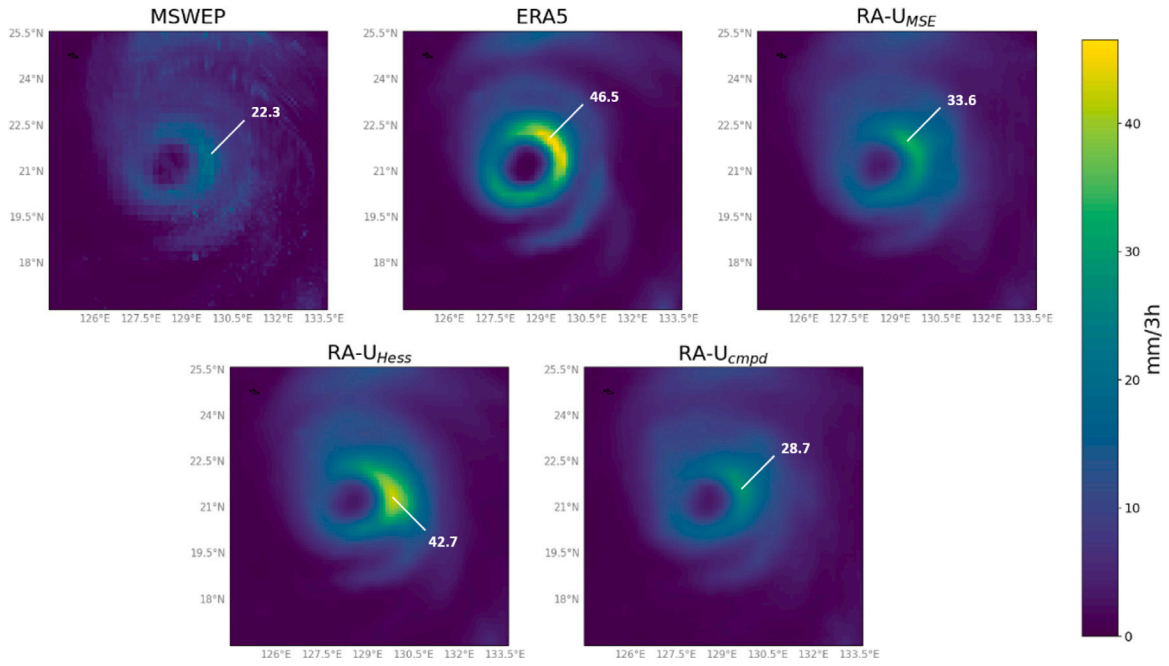

Fig. 20. A random rainfall map from the test set, on which RA-U$_{cmpd}$ performs better than the other two models by adjusting the bias of both total rainfall and peak localization, whereas the other two models aggravate both. The rainfall map is the snapshot of Typhoon Kong-rey (the most powerful TC of 2018) registered in IBTrACS at time 03/10/2018 07:30.

References

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., Pappenberger, F., 2013. Glofas–global ensemble streamflow forecasting and flood early warning. Hydrol. Earth Syst. Sci. 17 (3), 1161–1175.

Anisimov, O., Zhil'tsova, E., 2012. Climate change estimates for the regions of Russia in the 20th century and in the beginning of the 21st century based on the observational data. Russ. Meteorol. Hydrol. 37 (6), 421–429.

Ansari, R., Grossi, G., 2022. Performance evaluation of raw and bias-corrected ERA5 precipitation data with respect to extreme precipitation analysis: case study in upper jhelum basin, south Asia. Theor. Appl. Climatol. 150 (3–4), 1409–1424.

Ayugi, B., Tan, G., Ruoyun, N., Babaousmail, H., Ojara, M., Wido, H., Mumo, L., Ngoma, N.H., Nooni, I.K., Ongoma, V., 2020. Quantile mapping bias correction on rossby centre regional climate models for precipitation analysis over Kenya, east africa. Water 12 (3), 801.

Baheti, B., Innani, S., Gajre, S., Talbar, S., 2020. Eff-unet: A novel architecture for semantic segmentation in unstructured environment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 358–359.

Bakkensen, L.A., Park, D.-S.R., Sarkar, R.S.R., 2018. Climate costs of tropical cyclone losses also depend on rain. Environ. Res. Lett. 13 (7), 074034.

Beck, H.E., Pan, M., Roy, T., Weedon, G.P., Pappenberger, F., Van Dijk, A.I., Huffman, G.J., Adler, R.F., Wood, E.F., 2019a. Daily evaluation of 26 precipitation datasets using stage-iv gauge-radar data for the CONUS. Hydrol. Earth Syst. Sci. 23 (1), 207–224.

Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A.I., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J., Wood, E.F., 2017. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrol. Earth Syst. Sci. 21 (12), 6201–6217.

Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., Van Dijk, A.I., McVicar, T.R., Adler, R.F., 2019b. MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment. Bull. Am. Meteorol. Soc. 100 (3), 473–500.

Bhattacharyya, S., Sreekesh, S., King, A., 2022. Characteristics of extreme rainfall in different gridded datasets over India during 1983–2015. Atmos. Res. 267, 105930.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., Tian, Q., 2023. Accurate medium-range global weather forecasting with 3D neural networks. Nature 619 (7970), 533–538.

Bloemendaal, N., de Moel, H., Mol, J.M., Bosma, P.R., Polen, A.N., Collins, J.M., 2021. Adequately reflecting the severity of tropical cyclones using the new tropical cyclone severity scale. Environ. Res. Lett. 16 (1), 014048.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Bretherton, C.S., Henn, B., Kwa, A., Brenowitz, N.D., Watt-Meyer, O., McGibbon, J., Perkins, W.A., Clark, S.K., Harris, L., 2022. Correcting coarse-grid weather and climate models by machine learning from global storm-resolving simulations. J. Adv. Modelling Earth Syst. 14 (2), e2021MS002794.

Cannon, A.J., Sobie, S.R., Murdock, T.Q., 2015. Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes?. J. Clim. 28 (17), 6938–6959.

Cantoni, E., Tramblay, Y., Grimaldi, S., Salamon, P., Dakhlaoui, H., Dezetter, A., Thiemig, V., 2022. Hydrological performance of the ERA5 reanalysis for flood modeling in Tunisia with the LISFLOOD and GR4j models. J. Hydrol. Reg. Stud. 42, 101169.

Chen, H., Sun, L., Cifelli, R., Xie, P., 2021. Deep learning for bias correction of satellite retrievals of orographic precipitation. IEEE Trans. Geosci. Remote Sens. 60, 1–11.

Cho, D., Yoo, C., Im, J., Cha, D.-H., 2020. Comparative assessment of various machine learning-based bias correction methods for numerical weather prediction model forecasts of extreme air temperatures in urban areas. Earth Space Sci. 7 (4), e2019EA000740.

Cucchi, M., Weedon, G.P., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., Hersbach, H., Buontempo, C., 2020. WFDE5: bias-adjusted ERA5 reanalysis data for impact studies. Earth Syst. Sci. Data 12 (3), 2097–2120.

Czajkowski, J., Villarini, G., Michel-Kerjan, E., Smith, J.A., 2013. Determining tropical cyclone inland flooding loss on a large scale through a new flood peak ratio-based methodology. Environ. Res. Lett. 8 (4), 044056.

Diniz, F.L., Todling, R., 2020. Assessing the impact of observations in a multi-year reanalysis. Q. J. R. Meteorol. Soc. 146 (727), 724–747.

Dullaart, J.C., Muis, S., Bloemendaal, N., Aerts, J.C., 2020. Advancing global storm surge modelling using the new ERA5 climate reanalysis. Clim. Dyn. 54, 1007–1021.

Eccel, E., Ghielmi, L., Granitto, P., Barbiero, R., Grazzini, F., Cesari, D., 2007. Prediction of minimum temperatures in an alpine region by linear and non-linear post-processing of meteorological models. Nonlinear Process. Geophys. 14 (3), 211–222.

Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., Wichmann, F.A., 2020. Shortcut learning in deep neural networks. Nat. Mach. Intell. 2 (11), 665–673.

Gilleland, E., Ahijevych, D., Brown, B.G., Casati, B., Ebert, E.E., 2009. Intercomparison of spatial forecast verification methods. Weather Forecast. 24 (5), 1416–1430.

Gudmundsson, L., Bremnes, J.B., Haugen, J.E., Engen-Skaugen, T., 2012. Downscaling RCM precipitation to the station scale using statistical transformations–a comparison of methods. Hydrol. Earth Syst. Sci. 16 (9), 3383–3390.

Han, L., Chen, M., Chen, K., Chen, H., Zhang, Y., Lu, B., Song, L., Qin, R., 2021. A deep learning method for bias correction of ECMWF 24–240 h forecasts. Adv. Atmospheric Sci. 38 (9), 1444–1459.

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., Pappenberger, F., 2020. Glofas-ERA5 operational global river discharge reanalysis 1979–present. Earth Syst. Sci. Data 12 (3), 2043–2060.

Harrigan, S., Zsoter, E., Cloke, H., Salamon, P., Prudhomme, C., 2023. Daily ensemble river discharge reforecasts and real-time forecasts from the operational global flood awareness system. Hydrol. Earth Syst. Sci. 27 (1), 1–19.

Hassler, B., Lauer, A., 2021. Comparison of reanalysis and observational precipitation datasets including ERA5 and WFDE5. Atmosphere 12 (11), 1462.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146 (730), 1999–2049.

Hess, P., Boers, N., 2022. Deep learning for improving numerical weather prediction of heavy rainfall. J. Adv. Modelling Earth Syst. 14 (3), e2021MS002765.

Hewamalage, H., Ackermann, K., Bergmeir, C., 2023. Forecast evaluation for data scientists: common pitfalls and best practices. Data Min. Knowl. Discov. 37 (2), 788–832.

Hirpa, F.A., Gebremichael, M., Hopson, T., 2010. Evaluation of high-resolution satellite precipitation products over very complex terrain in ethiopia. J. Appl. Meteorol. Climatol. 49 (5), 1044–1051.

Hu, X., Naiel, M.A., Wong, A., Lamm, M., Fieguth, P., 2019. Runet: A robust unet architecture for image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.

Hu, Y.-F., Yin, F.-K., Zhang, W.-M., 2021. Deep learning-based precipitation bias correction approach for yin–he global spectral model. Meteorol. Appl. 28 (5), e2032.

Hu, Y.-F., Yin, F.-K., Zhang, W.-M., Deng, K.-F., 2022. A hybrid fusion precipitation bias correction approach for yin-he global spectral model. Meteorol. Appl. 29 (5), e2097.

Irwandi, H., Rosid, M.S., Mart, T., 2023. Effects of climate change on temperature and precipitation in the lake toba region, Indonesia, based on ERA5-land data with quantile mapping bias correction. Sci. Rep. 13 (1), 2542.

Iseri, Y., Diaz, A.J., Trinh, T., Kavvas, M.L., Ishida, K., Anderson, M.L., Ohara, N., Snider, E.D., 2021. Dynamical downscaling of global reanalysis data for high-resolution spatial modeling of snow accumulation/melting at the central/southern sierra nevada watersheds. J. Hydrol. 598, 126445.

Jiang, Y., Yang, K., Shao, C., Zhou, X., Zhao, L., Chen, Y., Wu, H., 2021. A downscaling approach for constructing high-resolution precipitation dataset over the tibetan plateau from ERA5 reanalysis. Atmos. Res. 256, 105574.

Jiang, X., Zhao, M., Waliser, D.E., 2012. Modulation of tropical cyclones over the eastern Pacific by the intraseasonal variability simulated in an AGCM. J. Clim. 25 (19), 6524–6538.

Jin, Q., Meng, Z., Sun, C., Cui, H., Su, R., 2020. RA-unet: A hybrid deep attention-aware network to extract liver and tumor in CT scans. Front. Bioeng. Biotechnol. 8, 605132.

Jolliffe, I.T., Stephenson, D.B., 2012. Forecast verification: a practitioner's guide in atmospheric science. John Wiley & Sons.

Kheir, A.M., Elnashar, A., Mosad, A., Govind, A., 2023. An improved deep learning procedure for statistical downscaling of climate data. Heliyon 9 (7).

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Knapp, K.R., Kruk, M.C., Levinson, D.H., Diamond, H.J., Neumann, C.J., 2010. The international best track archive for climate stewardship (ibtracs) unifying tropical cyclone data. Bull. Am. Meteorol. Soc. 91 (3), 363–376.

Knutson, T., Camargo, S.J., Chan, J.C., Emanuel, K., Ho, C.-H., Kossin, J., Mohapatra, M., Satoh, M., Sugi, M., Walsh, K., et al., 2020. Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. Bull. Am. Meteorol. Soc. 101 (3), E303–E322.

Knutson, T.R., McBride, J.L., Chan, J., Emanuel, K., Holland, G., Landsea, C., Held, I., Kossin, J.P., Srivastava, A., Sugi, M., 2010. Tropical cyclones and climate change. Nat. Geosci. 3 (3), 157–163.

Lagerquist, R., Ebert-Uphoff, I., 2022. Can we integrate spatial verification methods into neural network loss functions for atmospheric science?. Artif. Intell. Earth Syst. 1 (4), e220021.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Pritzel, A., Ravuri, S., Ewalds, T., Alet, F., Eaton-Rosen, Z., et al., 2022. GraphCast: Learning skillful medium-range global weather forecasting. arXiv preprint arXiv:2212.12794.

Lavers, D.A., Simmons, A., Vamborg, F., Rodwell, M.J., 2022. An evaluation of ERA5 precipitation for climate monitoring. Q. J. R. Meteorol. Soc. 148 (748), 3152–3165.

Le, X.-H., Lee, G., Jung, K., An, H.-u., Lee, S., Jung, Y., 2020. Application of convolutional neural network for spatiotemporal bias correction of daily satellite-based precipitation. Remote Sens. 12 (17), 2731.

Le, X.-H., Van, L.N., Nguyen, D.H., Nguyen, G.V., Jung, S., Lee, G., 2023. Comparison of bias-corrected multisatellite precipitation products by deep learning framework. Int. J. Appl. Earth Obs. Geoinf. 116, 103177.

Leander, R., Buishand, T.A., 2007. Resampling of regional climate model output for the simulation of extreme river flows. J. Hydrol. 332 (3–4), 487–496.

Lee, C.-S., Wu, C.-C., Wang, T.-C.C., Elsberry, R.L., 2011. Advances in understanding the "perfect monsoon-influenced typhoon": Summary from international conference on typhoon morakot (2009). Asia-Pacific J. Atmospheric Sci. 47, 213–222.

Lenderink, G., Buishand, A., Van Deursen, W., 2007. Estimates of future discharges of the river rhine using two scenario methodologies: direct versus delta approach. Hydrol. Earth Syst. Sci. 11 (3), 1145–1159.

Li, H., Zhang, Y., Lei, H., Hao, X., 2023. Machine learning-based bias correction of precipitation measurements at high altitude. Remote Sens. 15 (8), 2180.

Lin, Y., Zhao, M., Zhang, M., 2015. Tropical cyclone rainfall area controlled by relative sea surface temperature. Nature Commun. 6 (1), 6591.

Ling, F., Li, Y., Luo, J.-J., Zhong, X., Wang, Z., 2022. Two deep learning-based bias-correction pathways improve summer precipitation prediction over China. Environ. Res. Lett. 17 (12), 124025.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022. A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11976–11986.

Luo, M., Liu, T., Meng, F., Duan, Y., Frankl, A., Bao, A., De Maeyer, P., 2018. Comparing bias correction methods used in downscaling precipitation and temperature from regional climate models: a case study from the kaidu river basin in western China. Water 10 (8), 1046.

Mendez, M., Maathuis, B., Hein-Griggs, D., Alvarado-Gamboa, L.-F., 2020. Performance evaluation of bias correction methods for climate change monthly precipitation projections over costa rica. Water 12 (2), 482.

Mittermaier, M., 2021. A "meta" analysis of the fractions skill score: The limiting case and implications for aggregation. Mon. Weather Rev. 149 (10), 3491–3504.

Moustafa, M.S., Mohamed, S.A., Ahmed, S., Nasr, A.H., 2021. Hyperspectral change detection based on modification of unet neural networks. J. Appl. Remote Sens. 15 (2), 028505.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al., 2021. ERA5-land: A state-of-the-art global reanalysis dataset for land applications. Earth Syst. cience data 13 (9), 4349–4383.

National Hurricane Center (NHC), 2023. Saffir-simpson hurricane wind scale. https://www.nhc.noaa.gov/aboutsshws.php, (Accessed July 19, 2023).

Necker, T., Wolfgruber, L., Kugler, L., Weissmann, M., Dorninger, M., Serafin, S., 2023. The fractions skill score for ensemble forecast verification. Preprints, URL https://www.authorea.com/users/650600/articles/659203-the-fractions-skill-score-for-ensemble-forecast-verification.

Ngai, S.T., Tangang, F., Juneng, L., 2017. Bias correction of global and regional simulated daily precipitation and surface mean temperature over southeast Asia using quantile mapping method. Glob. Planet. Change 149, 79–90.

Park, D.-S.R., Ho, C.-H., Kim, J., Kang, K., Nam, C.C., 2016. Highlighting socioeconomic damages caused by weakened tropical cyclones in the Republic of Korea. Nat. Hazards 82, 1301–1315.

Park, D.-S.R., Ho, C.-H., Nam, C.C., Kim, H.-S., 2015. Evidence of reduced vulnerability to tropical cyclones in the Republic of Korea. Environ. Res. Lett. 10 (5), 054003.

Piani, C., Weedon, G., Best, M., Gomes, S., Viterbo, P., Hagemann, S., Haerter, J., 2010. Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. J. Hydrol. 395 (3–4), 199–215.

Probst, E., Mauser, W., 2022. Evaluation of ERA5 and WFDE5 forcing data for hydrological modelling and the impact of bias correction with regional climatologies: A case study in the danube river basin. J. Hydrol. Reg. Stud. 40, 101023.

Ramos, A.M., Roca, R., Soares, P.M., Wilson, A.M., Trigo, R.M., Ralph, F.M., 2021. Uncertainty in different precipitation products in the case of two atmospheric river events. Environ. Res. Lett. 16 (4), 045012.

Rappaport, E.N., 2014. Fatalities in the United States from atlantic tropical cyclones: New data and interpretation. Bull. Am. Meteorol. Soc. 95 (3), 341–346.

Roberts, N.M., Lean, H.W., 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Mon. Weather Rev. 136 (1), 78–97.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Sadeghi, M., Nguyen, P., Hsu, K., Sorooshian, S., 2020. Improving near real-time precipitation estimation using a U-net convolutional neural network and geographical information. Environ. Model. Softw. 134, 104856.

Schreck, C.J., Knapp, K.R., Kossin, J.P., 2014. The impact of best track discrepancies on global tropical cyclone climatologies using ibtracs. Mon. Weather Rev. 142 (10), 3881–3899.

Scoccimarro, E., Gualdi, S., Villarini, G., Vecchi, G.A., Zhao, M., Walsh, K., Navarra, A., 2014. Intense precipitation events associated with landfalling tropical cyclones in response to a warmer climate and increased CO 2. J. Clim. 27 (12), 4642–4654.

Sennikovs, J., Bethers, U., 2009. Statistical downscaling method of regional climate model results for hydrological modelling. In: Proceedings of the 18th World IMacS/MODSIM Congress. pp. 3962–3968.

Sha, Y., Gagne II, D.J., West, G., Stull, R., 2020. Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part II: Daily precipitation. J. Appl. Meteorol. Climatol. 59 (12), 2075–2092.

Sharifi, E., Eitzinger, J., Dorigo, W., 2019. Performance of the state-of-the-art gridded precipitation products over mountainous terrain: A regional study over Austria. Remote Sens. 11 (17), 2018.

Shrestha, M., Acharya, S.C., Shrestha, P.K., 2017. Bias correction of climate models for hydrological modelling–are simple methods still useful?. Meteorol. Appl. 24 (3), 531–539.

Singh, M., Kumar, B., Rao, S., Gill, S.S., Chattopadhyay, R., Nanjundiah, R.S., Niyogi, D., 2021. Deep learning for improved global precipitation in numerical weather prediction systems. arXiv preprint arXiv:2106.12045.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Duda, M.G., Huang, X.-Y., Wang, W., Powers, J.G., et al., 2008. A description of the advanced research WRF version 3. NCAR technical note 475, 113.

Skok, G., 2015. Analysis of fraction skill score properties for a displaced rainband in a rectangular domain. Meteorol. Appl. 22 (3), 477–484.

Skok, G., Roberts, N., 2016. Analysis of fractions skill score properties for random precipitation fields and ECMWF forecasts. Q. J. R. Meteorol. Soc. 142 (700), 2599–2610.

Skok, G., Roberts, N., 2018. Estimating the displacement in precipitation forecasts using the fractions skill score. Q. J. R. Meteorol. Soc. 144 (711), 414–425.

Smith, M.B., Koren, V.I., Zhang, Z., Reed, S.M., Pan, J.-J., Moreda, F., 2004. Runoff response to spatial variability in precipitation: an analysis of observed data. J. Hydrol. 298 (1–4), 267–286.

Smitha, P., Narasimhan, B., Sudheer, K., Annamalai, H., 2018. An improved bias correction method of daily rainfall data using a sliding window technique for climate change impact assessment. J. Hydrol. 556, 100–118.

Sohrabi, M., Moftakhari, H., Moradkhani, H., 2023. Efficient tropical cyclone scenario selection based on cumulative likelihood of potential impacts. Earth's Future 11 (10), e2023EF003731.

Stengel, K., Glaws, A., Hettinger, D., King, R.N., 2020. Adversarial super-resolution of climatological wind and solar data. Proc. Natl. Acad. Sci. 117 (29), 16805–16815.

Sun, H., Su, F., Yao, T., He, Z., Tang, G., Huang, J., Zheng, B., Meng, F., Ou, T., Chen, D., 2021. General overestimation of ERA5 precipitation in flow simulations for high mountain Asia basins. Environ. Res. Commun. 3 (12), 121003.

Sun, H., Yao, T., Su, F., He, Z., Tang, G., Li, N., Zheng, B., Huang, J., Meng, F., Ou, T., et al., 2022. Corrected ERA5 precipitation by machine learning significantly improved flow simulations for the third pole basins. J. Hydrometeorol. 23 (10), 1663–1679.

Tan, J., Chen, S., Lee, C.-Y., Dong, G., Hu, W., Wang, J., 2021. Projected changes of typhoon intensity in a regional climate model: Development of a machine learning bias correction scheme. Int. J. Climatol. 41 (4), 2749–2764.

Teutschbein, C., Seibert, J., 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. J. Hydrol. 456, 12–29.

Trebing, K., Stanczyk, T., Mehrkanoon, S., 2021. Smaat-unet: Precipitation nowcasting using a small attention-unet architecture. Pattern Recognit. Lett. 145, 178–186.

Villarini, G., Goska, R., Smith, J.A., Vecchi, G.A., 2014. North atlantic tropical cyclones and US flooding. Bull. Am. Meteorol. Soc. 95 (9), 1381–1388.

Villarini, G., Smith, J.A., Baeck, M.L., Marchok, T., Vecchi, G.A., 2011. Characterization of rainfall distribution and flooding associated with US landfalling tropical cyclones: Analyses of hurricanes Frances, ivan, and jeanne (2004). J. Geophys. Res.: Atmos. 116 (D23).

Voropay, N., Ryazanova, A., Dyukarev, E., 2021. High-resolution bias-corrected precipitation data over south siberia, Russia. Atmos. Res. 254, 105528.

Walsh, K.J., McBride, J.L., Klotzbach, P.J., Balachandran, S., Camargo, S.J., Holland, G., Knutson, T.R., Kossin, J.P., Lee, T.-c., Sobel, A., et al., 2016. Tropical cyclones and climate change. Wiley Interdiscip. Rev. Clim. Change 7 (1), 65–89.

Wang, C.-C., Kuo, H.-C., Chen, Y.-H., Chen, S.-H., Tsuboki, K., 2022. A decade after typhoon morakot (2009): What have we learned about its physics and predictability?. Weather Forecast. 37 (12), 2161–2181.

Wang, Z., Simoncelli, E.P., Bovik, A.C., 2003. Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. 2, Ieee, pp. 1398–1402.

Wang, F., Tian, D., 2022. On deep learning-based bias correction and downscaling of multiple climate models simulations. Clim. Dyn. 59 (11–12), 3451–3468.

World Meteorological Organization (WMO), 2023. Tropical cyclones. https://public.wmo.int/en/our-mandate/focus-areas/natural-hazards-and-disaster-risk-reduction/tropical-cyclones, (Accessed July 19 2023).

Xiang, L., Guan, J., Xiang, J., Zhang, L., Zhang, F., 2022. Spatiotemporal model based on transformer for bias correction and temporal downscaling of forecasts. Front. Environ. Sci. 10, 1039764.

Yi, C., Shin, Y., Roh, J.-W., 2018. Development of an urban high-resolution air temperature forecast system for local weather information services based on statistical downscaling. Atmosphere 9 (5), 164.

Zhang, W., Jiang, Y., Dong, J., Song, X., Pang, R., Guoan, B., Yu, H., 2023. A deep learning method for real-time bias correction of wind field forecasts in the western north Pacific. Atmos. Res. 284, 106586.

Zhang, W., Villarini, G., Vecchi, G.A., Smith, J.A., 2018. Urbanization exacerbated the rainfall and flooding caused by hurricane harvey in houston. Nature 563 (7731), 384–388.

Zhao, T., Bennett, J.C., Wang, Q., Schepen, A., Wood, A.W., Robertson, D.E., Ramos, M.-H., 2017. How suitable is quantile mapping for postprocessing GCM precipitation forecasts?. J. Clim. 30 (9), 3185–3196.

Zhao, D.-J., Xu, H.-X., Yu, Y.-B., Chen, L.-S., 2022. Identification of synoptic patterns for extreme rainfall events associated with landfalling typhoons in China during 1960–2020. Adv. Clim. Chang. Res. 13 (5), 651–665.

Zhao, D., Yu, Y., Chen, L., 2021. Impact of the monsoonal surge on extreme rainfall of landfalling tropical cyclones. Adv. Atmospheric Sci. 38, 771–784.

Zhao, B., Zhang, B., 2018. Assessing hourly precipitation forecast skill with the fractions skill score. J. Meteorol. Res. 32 (1), 135–145.

Zjavka, L., 2016. Numerical weather prediction revisions using the locally trained differential polynomial network. Expert Syst. Appl. 44, 265–274.

Zoccatelli, D., Borga, M., Viglione, A., Chirico, G.B., Blöschl, G., 2011. Spatial moments of catchment rainfall: rainfall spatial organisation, basin morphology, and flood response. Hydrol. Earth Syst. Sci. 15 (12), 3767–3783.