Temporal Super-Resolution, Ground Adjustment, and Advection Correction of Radar Rainfall Using 3-D-Convolutional Neural Networks

Julius Polz, Luca Glawion, Hiob Gebisso, Lukas Altenstrasser, Maximilian Graf, Harald Kunstmann, Stefanie Vogl, and Christian Chwala

***Abstract*— Weather radars are highly sophisticated tools for quantitative precipitation estimation (QPE) and provide observations with unmatched spatial representativeness. However, their indirect measurement of precipitation high above ground leads to strong systematic errors compared to direct rain gauge measurements. Additionally, the temporal undersampling from 5-min instantaneous radar measurements requires advection correction. We present ResRadNet, a 3-D-convolutional residual neural network approach, to reduce these errors and, at the same time, increase the temporal resolution of the radar rainfall fields by a 5-min short-range prediction of 1-min time-steps. The network is trained to process spatiotemporal sequences of radar rainfall estimates from a composite product derived from 17 C-band weather radars in Germany. In contrast to previous approaches, we present a method that emphasizes the generation of spatiotemporally consistent and advection-corrected country-wide rainfall maps. Our approach significantly increased the Pearson correlation coefficient (PCC) of the radar product (from 0.63 to 0.74) and decreased the root mean squared error (mse) by 22% when compared to 247 rain gauges at a 5-min resolution. An additional large-scale comparison to eight years of data from 1138 independent manual daily gauges confirmed that the improvement is robust and transferable to new locations. Overall, our study shows the benefits of using 3-D convolutional neural networks (CNNs) for weather radar rainfall estimation to provide 1-min, ground-adjusted, that is, bias-corrected with respect to on-ground sensors, and advection-corrected radar rainfall estimates.**

***Index Terms*— Convolutional neural network (CNN), deep learning, precipitation, residual neural network, weather radar.**

Manuscript received 11 March 2023; revised 24 October 2023 and 4 January 2024; accepted 26 February 2024. Date of publication 29 February 2024; date of current version 12 March 2024. This work was supported in part by German Research Foundation within German Research Foundation (DFG) Research Unit RealPEP under Grant CH 1785/1-2, in part by Helmholtz Association (HGF)-Innopool, and in part by the Federal Ministry of Education and Research under Grant 13N14826. *(Corresponding author: Julius Polz.)*

Julius Polz, Luca Glawion, and Christian Chwala are with the Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Campus Alpin, 82467 Garmisch-Partenkirchen, Germany (e-mail: julius.polz@kit.edu).

Hiob Gebisso, Lukas Altenstrasser, and Stefanie Vogl are with the Department of Informatics and Mathematics, Munich University of Applied Sciences, 80335 Munich, Germany.

Maximilian Graf is with the Chair of Regional Climate and Hydrology, Institute of Geography, University of Augsburg, 86159 Augsburg, Germany.

Harald Kunstmann is with the Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Campus Alpin, 82467 Garmisch-Partenkirchen, Germany, and also with the Chair of Regional Climate and Hydrology, Institute of Geography, University of Augsburg, 86159 Augsburg, Germany.

Digital Object Identifier 10.1109/TGRS.2024.3371577

I. Introduction

WEATHER radars are one of the most advanced tools for quantitative precipitation estimation (QPE) at high spatial and temporal resolutions. Their observations are not only essential for state agencies providing information for disaster or water management. They also became one of the most popular rainfall observation tools for citizens. The key to their success is the high spatial representativeness of the derived measurements. Even though they integrate over large volumes, depending on the distance from the radar [1], they provide several observations per km^2 which is significantly higher than rain gauge networks. For example, the average density of rain gauges is $1/330$ km^{-2} in Germany, which is high in a global comparison. However, the raw measurements of weather radars are subject to systematic and random errors due to their indirect measurement high above ground [2], [3]. Additionally, weather radar measurements are instantaneous, typically sampled every 5 min, which leads to a temporal undersampling. This becomes obvious when aggregating rain fields containing small fast-moving cells which leads to gaps in the rain field. As a solution, advection correction methods have been suggested [4], [5], [6].

A common approach to correct radar QPE biases is to merge radar and rain gauge data to provide optimal rainfall estimates near the ground. Most of them consider both radar and rain gauge data as sources of information to achieve a combined product. Others correct radar data to remove biases compared to rain gauge observations using a statistical approach [3], [7], [8]. However, scenario-dependent and advection-driven biases between radar and rain gauge data ask for a more dynamic approach like [9] propose. It is reasonable to assume that the spatiotemporal dynamics contained in the rainfall fields measured by weather radars can provide valuable information to correctly map radar rainfall estimates to the ground.

Deep neural networks have been used effectively for modeling dynamic physical systems. They provide a flexible and computationally efficient modeling framework that can outperform state-of-the-art physical models with comparatively small computational effort and at low latency [10]. The concept of using deep neural networks to derive improved reflectivity-rainfall (Z-R) retrieval for radar-derived QPE has been used recently by [11] who targeted hourly observations of 45 rain gauges from radar reflectivity at S-band using one radar.

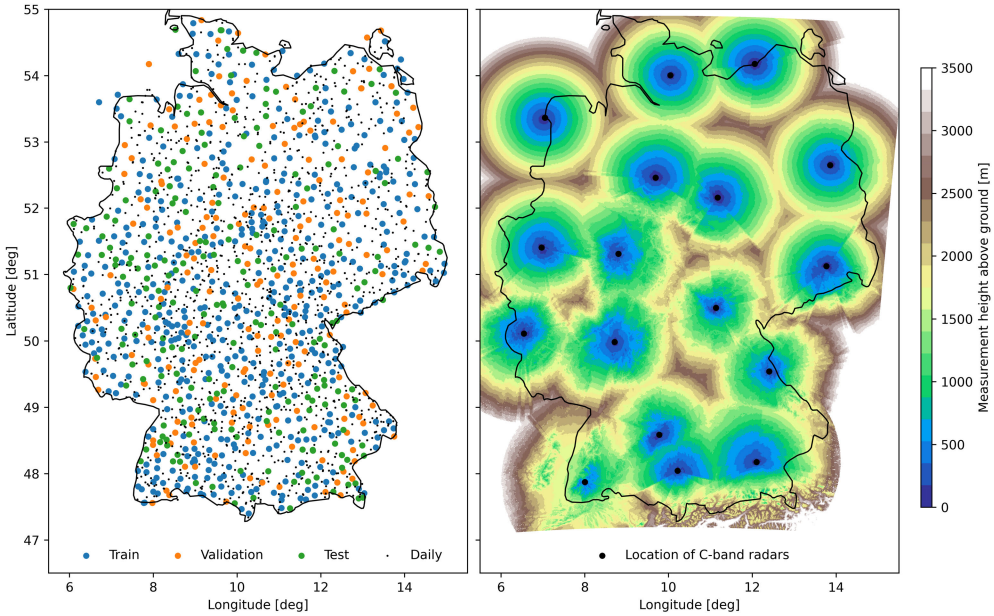

Fig. 1. (Left) Map of Germany showing the locations of the rain gauges used for the train, validation, and test split and (right) the minimal radar measurement height above ground of the composite derived from the 17 C-band radars. Training data are taken from 2020 and validation and test data are from 2021.

Chen and Chandrasekar [12] applied a 2-D-convolutional neural network (CNN) to derive point scale rainfall from reflectivity and differential reflectivity in the vicinity of one radar, also at S-band. Vogl et al. [13] applied a neural network approach deriving rainfall estimates from reflectivity measured at X-band using one reference rain gauge and investigated the influence of different temporal aggregations on the neural network performance. Hassan et al. [14] applied classical machine learning algorithms to determine suitable physical retrieval algorithms in different scenarios. Moraux et al. [15] present a multi-modal approach for merging rain-gauge, satellite, and radar rainfall estimates using all three data sources as model input. In summary, these studies directly used radar reflectivity as input for their retrievals without increasing the temporal resolution. They provide a proof of concept that a learned Z-R retrieval can improve radar rainfall estimates at S- or X-band. The respective case studies involved either single radar stations, a low amount of reference stations, or a short time period. It remains unclear if an improved Z-R retrieval or an estimate of the spatiotemporal dynamics and a reduced spatial or temporal mismatch between radar and on-ground reference is responsible for the improvement. It is also unclear if radar rainfall estimates from a radar composite, which can include additional errors or discontinuities, can be used to consistently improve surface rainfall at high spatial and temporal resolutions (e.g., 1 km and 1 min) at a country-wide scale.

The research questions addressed by this study target these knowledge gaps.

1) Are 3-D-convolutional residual neural networks a suitable tool to increase the temporal radar resolution of 5 min by a short-term prediction of five 1-min time-steps?
2) Can biases between high-resolution C-band weather radar rainfall estimates and rain gauges on the ground be effectively reduced?
3) Does the method provide spatiotemporally consistent country-wide rainfall maps over Germany which can be used to correct advection-driven undersampling?

In summary, we aim to show that 3-D-convolutional residual neural networks are capable of simultaneous temporal super-resolution, ground adjustment, and advection correction of radar rainfall. Our evaluation is based on eight years of country-wide radar and rain gauge observations. For the sake of brevity, we denote our proposed method by ResRadNet.

II. Data and Methods

A. Data

All data used in this study was provided by the German meteorological service (DWD). We used two rain gauge and two radar datasets. The 1-min rain gauge data are freely available on DWD's opendata online archive (https://opendata.dwd.de/). The network operated Ott Pluvio^2 sensors until 2018. From 2018 to 2020 these rain gauges were replaced by rain[e] sensors from Lambrecht meteo. Both sensors are weighing gauges that provide the accumulated rainfall amount on a 1-min basis. While they have a different resolution (Pluvio^2 0.01 mm and rain[e] 0.001 mm) the quantization of the dataset remains the same with 0.01 mm. The average station density was one rain gauge per 330 km^2.

The daily rain gauge data is also freely available on DWD's opendata online archive [16]. The network consisted of Hellmann rain gauges with a manual readout of the accumulated rainfall amount of the last 24 h at 5:50 UTC. Therefore, the aggregation time was from 5:50 to 5:50 UTC on the following day. The quantization of the dataset was 0.1 mm.

For the period from 2013 to 2021, there were 1066 1-min and 2150 daily rain gauge stations available. Around 50% of the daily rain gauges were located at the same site as the 1-min rain gauges. In order to obtain an independent network of daily rain gauges we removed these stations from the analysis. The

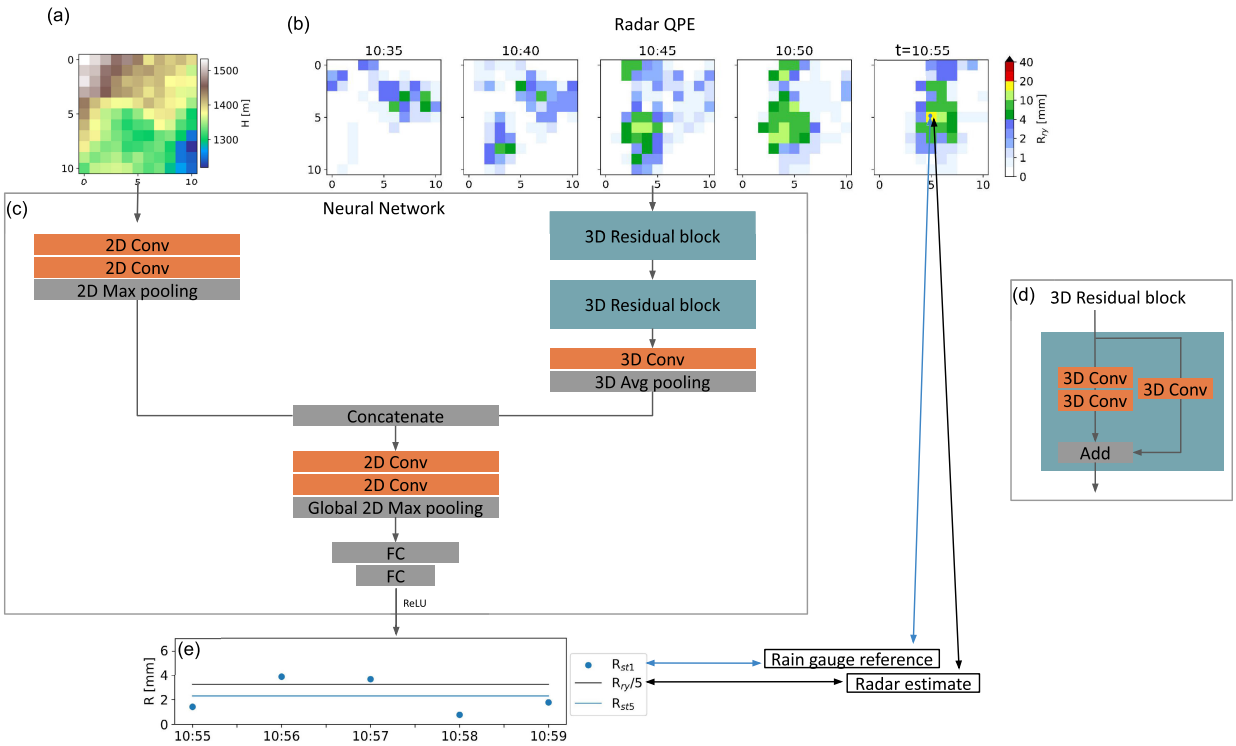

Fig. 2. Schematic overview of the ResRadNet architecture, the input, and the target data. To predict (e) one set of five 1-min rain gauge observations, the model uses (a) the minimal measurement height above ground, and (b) five 5-min radar time-steps with a spatial extent of 11 × 11 pixels (≡11 km), centered at the rain gauge location. The neural network (c) is using convolutional layers (orange) and (d) skip connections with added layers organized as residual blocks. The height and the rainfall information are processed by separate input branches and later concatenated before being processed by another convolutional block and a final fully connected part. Beneath the 1-min rain gauge observations R_{st1}, (e) also shows the 5-min average rainfall from the center radar pixel $R_{ry/5}$ and the rain gauge reference R_{st5}. ResRadNet was trained to predict R_{st1} from (a) and (b) and thus improve $R_{ry/5}$.

remaining 1138 locations are shown as black dots in Fig. 1. Note that not all stations were available for the full period due to maintenance, redistribution, and new installation. The number of missing gauges per year was less than 5%.

The radar products from DWD used in this study are RADOLAN-RY, RADOLAN-RW, and RADKLIM-YW. A detailed explanation of the operational routines of these three radar products from DWD, e.g., on the used three-part Z-R relation, the rain gauge adjustment or climatological corrections can be found in [17] for RADOLAN-RY and RADOLAN-RW and in [18] for RADKLIM-YW.

RADOLAN-RY is a quality-checked and attenuation-corrected composite of 17 weather radars in Germany, with a temporal resolution of 5 min and a spatial resolution of 1 × 1 km on a 900 × 900 km equally spaced grid. DWD derives the rainfall rate from a Z-R relationship based on measured reflectivity at C-band.

RADOLAN-RW is an hourly aggregated and gauge-adjusted version of RADOLAN-RY. The adjustment consists of a weighted combination of multiplicative and additive factors derived from the comparison of the 1-min rain gauges described above with the radar grid at an hourly aggregation.

For RADKLIM-YW, the DWD uses the same radar and gauge data, as well as the daily rain gauges described above to derive a gauge-adjusted and climatologically corrected radar product at a temporal resolution of 5 min and a spatial resolution of 1 × 1 km on an equally spaced grid of 1100 × 900 km extending 100 km to the east and west compared to RADOLAN-RY. It should be noted that the adjustment weights are calculated at a 1-h resolution and remain constant for the 12 5-min timesteps within this hour. Thus, the 5-min observations still deviate from the rain gauges used for adjustment. The climatological corrections aim to reduce errors for example radar spokes from beam blockage and range-dependent underestimation. We used RADKLIM-YW as a visual reference for high-resolution maps.

The dataset covers the years 2001–2021, but we omitted the time period between 2001 and 2012 since the number of available rain gauges was much lower due to the ongoing set-up of the gauge network. To compare our results to an advection-corrected version of RADOLAN-RY we used the Lucas-Kanade algorithm [19] implemented in PYSTEPS version 1.7.1 [6]. We used a base resolution of 5 min and five intermediate timesteps. To compare 1-min neural network estimates to RADOLAN-RY we interpreted the 5-min resolution as a constant average during the 5 min [see $R_{ry}/5$ in Fig. 2(e)].

The minimal measurement height above ground (see Fig. 1) is derived from the terrain following the beam angle of the precipitation scan and is provided on the same grid as RADOLAN-RY. Where no terrain is blocking the beam an angle of 0.8° is used. In the radar composite, the minimum of the measurement heights of two overlapping radar scans is used. This usually gives the measurement height of the closest radar which also has the highest weight in the composite of

measurements. In some parts of Germany, the measurement height exceeds 2.5 km.

B. Data Pre-Processing

We train our model on sequences of RADOLAN-RY radar images with the objective to predict an assigned sequence of rain gauge measurements. The selection of radar sequences and the assigned rain gauge were defined as shown in Fig. 2: If a rain gauge Y was contained in the radar pixel $X_{t,i,j}$ at time t in minutes, then $X_{i,j}$ was associated with rain gauge measurements $(Y_{t'})_{t'\in\{t,\ldots,t+4\}}$. For grid calculations we used ωradlib [20]. As model input, the neighboring radar pixels $\{X_{k,l}\}_{k\in\{i-5,\ldots,i+5\},l\in\{j-5,\ldots,j+5\}}$ at times $t-20, t-15, \ldots, t$ were concatenated to form a sample of shape (5, 11, 11) [see Fig. 2(b)]. Additional model input was given by the minimal measurement height above ground from the 11 $\times$ 11 radar pixels [see Fig. 2(a)].

We excluded all samples where the radar input contained missing values or where all values were zero. We also excluded all samples where the associated rain gauge had a missing value, but no additional requirements were made for the rain gauge reference. Data were split into spatially and temporally separated train, validation, or test sets by a random selection of 581 stations for training, 238 for validation, and 247 for testing of the model. The locations of the selected rain gauges are shown in Fig. 1. For training only the year 2020 was considered and, if not indicated otherwise, the full year 2021 was used for both validation and testing. The training data were used to train the neural network and the validation data were used for model selection and bias correction as described below. The test dataset was an independent dataset only used for validation. The training dataset contained 8 773 000 samples in randomized order.

C. Model Architecture

The model architecture of ResRadNet consisted of a two-branch CNN with two 3-D-residual blocks in the radar input branch and a 2-D convolution block in the measurement height branch. A schematic overview is given in Fig. 2(c), a more precise description of the layers and connections is given in Table I, and code to re-build the model or retrieve the trained model is given in [21]. The model architecture was designed using the Keras API in Tensorflow version 2.7.0 [22] and optimized with the AMSgrad version of the Adam algorithm and a learning rate of 0.0001 [23], [24] using a mean squared error (mse) loss [see (3)]. The residual blocks consisted of two convolutional branches receiving the same input [see Fig. 2(d)]. The first uses two 3-D convolutional layers with a kernel size of 3 and ReLU activation. The second uses one 3-D convolutional layer with a kernel size of 1. They are followed by an added layer that combines the two branches. In general, such a design allows for the use of deeper networks to learn more complex features without running into gradient vanishing problems [25]. More specifically, as [26] describe this network design encourages the representation of features as a perturbation of the input. Therefore, in our case, the model is encouraged to learn to represent the optical flow of a field rather than to produce a precise representation of the field itself in every layer.

We trained the model with a batch size of 1000. After each epoch, the same 100 000 random validation samples were evaluated and the model was saved. We did not use the full validation set to speed up training time. After training, the model with the best validation loss was selected. Due to the skewed distribution of rainfall estimates, there was a multiplicative bias of the form $\overline{R_{\mathrm{st}}}/\overline{R_{\mathrm{nn}}} = \alpha \neq 1$, where R_{st} is the rain gauge reference, R_{nn} is the neural network prediction and where the $\overline{R}$ indicates the mean of the quantity. We decided to remove this bias by using a bias correction factor $R_{\mathrm{nn}} \mapsto \alpha R_{\mathrm{nn}}$ which is equivalent to a common mean field bias reduction [8]. We computed this factor using the validation dataset.

In inference mode, the final model was used like a spatio-temporal filter kernel that was applied to the neighborhood of every radar pixel. The computation time for a full radar image was 5 s per time step using an NVIDIA Tesla V100 32 GB GPU. The performance loss on smaller GPUs is negligible due to the small size of the input data. An approximate 20-s overhead for serialization and de-serialization of the samples could be omitted by parallel computation.

D. Evaluation

The derived QPE is evaluated for different temporal aggregations (1, 5 min, and daily) using a set of pixel-wise error metrics which are commonly used in the field [27]. The Pearson correlation coefficient (PCC) measuring the linear correlation between predicted values and ground truth is defined as

$$\mathrm{PCC} = \frac{\sum_n \left(R_{\mathrm{st}} - \overline{R_{\mathrm{st}}}\right)\left(R - \overline{R}\right)}{\sqrt{\sum_n \left(R_{\mathrm{st}} - \overline{R_{\mathrm{st}}}\right)^2 \sum_n \left(R - \overline{R}\right)^2}} \tag{1}$$

where R_{st} is the rain gauge reference and R is the radar (R_{ry}) or neural network (R_{nn}) prediction. Accordingly, the mse, (normalized) root mse (NRMSE), and normalized bias (NBIAS) are given by

$$\mathrm{mse} = \overline{(R_{\mathrm{st}} - R)^2} \tag{2}$$

$$\mathrm{NRMSE} = \frac{1}{\overline{R_{\mathrm{st}}}}\sqrt{\overline{(R_{\mathrm{st}} - R)^2}} \tag{3}$$

and

$$\mathrm{NBIAS} = \frac{\overline{R_{\mathrm{st}}} - \overline{R}}{\overline{R_{\mathrm{st}}}} * 100. \tag{4}$$

The NBIAS is independent of the temporal resolution and, therefore, also reflects the relative bias for the whole evaluation period. Let R^+ and R^+_{st} indicate if radar or rain gauges detect rainfall, i.e., let them be equal to 1 if the value of R (respectively, R_{st}) is larger than zero and zero otherwise. Then the mean detection error (MDE) is defined as the frequency of cases, where $R^+_{st} \neq R^+$. Additionally, we investigate the similarity, or rather maximal dissimilarity, between the two distributions using the Kolmogorov–Smirnov (KS) test. It is defined as the maximum deviation between the cumulative distribution of reference and predicted values.

TABLE I

ARCHITECTURE OF RESRADNET. THE TABLE RESEMBLES THE MODEL SUMMARY GIVEN BY THE KERAS API OF TENSORFLOW VERSION 2.7.0. THE TOTAL NUMBER OF TRAINABLE PARAMETERS WAS 177 985

Layer	Type	Output Shape	Kernel size	Activation	nParam	Connected to	Note
input-1 (QPE)	(InputLayer)	(5, 11, 11,1)			0		
conv3D-1	(Conv3D)	(5, 11, 11, 32)	3x3x3	-	896	'input-1'	
relu1	(ReLU)	(5, 11, 11, 32)			0	'conv3D-1'	
conv3D-2	(Conv3D)	(5, 11, 11, 32)	3x3x3	-	27680	'relu-1'	
relu-2	(ReLU)	(5, 11, 11, 32)			0	'conv3D-2'	
conv3D-3	(Conv3D)	(5, 11, 11, 32)	1x1x1	-	64	'input-3'	
add-1	(Add)	(5, 11, 11, 32)			0	'relu-2','conv3D-3'	
conv3D-4	(Conv3D)	(5, 11, 11, 32)	3x3x3	-	27680	'add-1'	
relu-3	(ReLU)	(5, 11, 11, 32)			0	'conv3D-4'	
conv3D-5	(Conv3D)	(5, 11, 11, 32)	3x3x3	-	27680	'relu-3'	
relu-4	(ReLU)	(5, 11, 11, 32)			0	'conv3D-5'	
conv3D-6	(Conv3D)	(5, 11, 11, 32)	1x1x1	-	1056	'add-1'	
add-2	(Add)	(5, 11, 11, 32)			0	'relu-4', 'conv3D-6'	
conv3D-7	(Conv3D)	(3, 9, 9, 64)	3x3x3	ReLU	55360	'add-2'	valid pad
averagepool3D-1	(AveragePooling3D)	(1, 4, 4, 64)	2x2x2		0	'conv3D-7'	
reshape-1	(Reshape)	(4, 4, 64)			0	'averagepool3D-1'	
input-2 (Height)	(InputLayer)	(11, 11, 1)			0		
conv2D-1	(Conv2D)	(10, 10, 4)	2x2	-	20	'input-2'	valid pad
conv2D-2	(Conv2D)	(9, 9, 8)	2x2	ReLU	136	'conv2D-1'	valid pad
maxpool2D-1	(MaxPooling2D)	(4, 4, 8)	2x2		0	'conv2D-2'	
concatenate-1	(Concatenate)	(4, 4, 72)			0	'reshape-1','maxpool2D-2'	
conv2D-3	(Conv2D)	(4, 4, 64)	2x2	-	18496	'concatenate-1'	
conv2D-4	(Conv2D)	(4, 4, 32)	2x2	ReLU	8224	'conv2D-3'	
maxpool2D-3	(MaxPooling2D)	(2, 2, 32)	2x2		0	'conv2D-4'	
conv2D-5	(Conv2D)	(2, 2, 32)	2x2	-	4128	'maxpool2D-3'	
conv2D-6	(Conv2D)	(2, 2, 32)	2x2	ReLU	4128	'conv2D-5'	
globalmaxpool2D-1	(Global MaxPooling2D)	(32)			0	'conv2D-6'	
dense-1	(Dense)	(64)		ReLU	2112	'globalmaxpool2D-1'	
dense-2	(Dense)	(5)		ReLU	325	'dense-1'	

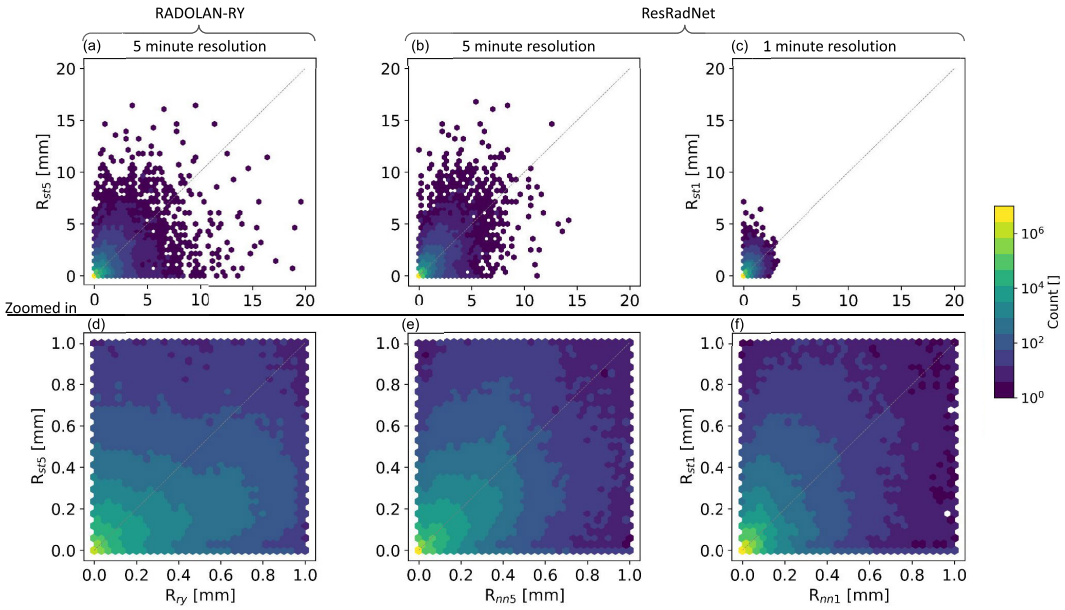

Fig. 3. Two-dimensional-histograms showing the relationship between RADOLAN-RY [R_{ry} in (a) and (d)] or ResRadNet [R_{nn} in (b), (c), (e), and (f)] and rain gauge (R_{st}) observations. For (a) and (b) temporal resolution is 5 min. For (c) it is 1-min. (d)–(f) Zoomed-in version of (a)–(c) with smaller bin widths to visualize what is happening for smaller rainfall amounts below 1 mm.

III. RESULTS

A. Model Selection and Comparison at 1-min Resolution

ResRadNet achieved the best mse on the 100 000 sample validation set after 15 epochs. For this validation data, the bias correction factor α was close to 1 with a value of 1.176 and we applied it to all predicted rainfall estimates presented hereafter. Therefore, the model achieved zero NBIAS, a PCC of 0.68, and an MDE of 0.16 when compared to the full 1-min validation set. A constant prediction of five 1-min values using the average RADOLAN-RY value achieved worse results with an NBIAS of −24.3%, a PCC of 0.57, and an MDE of 0.26. For the test data results at a 1-min resolution were similar with a slightly increased NBIAS (see Table II). The 2-D histogram comparing R_{nn1} and R_{st1} shown in Fig. 3(f) showed that missed extremes ($R_{nn1} \ll R_{st1}$) were more frequent than

false extremes ($R_{nn1} \gg R_{st1}$) considering the higher values of bins close to the x-axis.

B. Comparison at 5-min Resolution

The results of the test dataset at a 5-min resolution are shown in Table II. Both RADOLAN-RY's and ResRadNet's PCC improved compared to the 1-min results. The KS-test and MDE improved for the radar and worsened for the neural network. The RMSE increased in both cases and the NBIAS was independent of the resolution. Except for the KS-test the ResRadNet achieved better results than RADOLAN-RY.

Fig. 3 shows a scatter density comparison of the rain gauge and the radar or neural network predictions. The most obvious improvement of ResRadNet could be observed near the x- and y-axis of the plots [comparing panels Fig. 3(a) and (b)] and for the smaller values comparing panels Fig. 3(d) and (e). The panels showing ResRadNet data show a higher concentration of points along the diagonal and a lower concentration closer to the axes. While there is still a large uncertainty for extreme values, outliers close to both axes could be reduced.

Fig. 4 shows a long-term analysis using the test stations and the years 2013 to 2021. It confirms that for the 5-min resolution, ResRadNet performed better than the radar for all years and all scores. The average improvement was 0.2 for the PCC, 2 for the NRMSE, 20% for the NBIAS and 7% for the MDE. The year 2016 showed exceptionally poor PCC and NRMSE for RADOLAN-RY while ResRadNet seemed to be able to correct this issue.

C. Comparison to Daily Rain Gauge Data

The comparison to the independent set of daily rain gauge measurements confirmed the 5-min test results. ResRadNet improved the PCC with a value of 0.86 compared to 0.84 for RADOLAN-RY and decreased the RMSE from 3.28 to 2.65. The MDE and KS-Test were very similar for both products. The resolution-independent NBIAS was similar to the 1-min gauges despite the different station locations (see Table II). The long-term evaluation from 2013 to 2021 (excluding the training year 2020) showed that the improvement was very consistent (see Fig. 4). The 2021 scores were, again, the highest for both RADOLAN and ResRadNet and the improvement was smaller than in previous years. The NBIAS shows the same temporal dynamics as for the 5-min data but is slightly higher for both products.

As an additional reference, we compared the operational gauge adjustment routine of the DWD (RADOLAN-RW) and the daily rain gauge data, which was not used for this adjustment (see Fig. 4). It could be observed that, at this temporal resolution, the MDE is only marginally better than for RADOLAN-RY or ResRadNet. The NBIAS and NRMSE are lower than for ResRadNet and the PCC is higher with values around 0.9. An exception is the year 2016 when RADOLAN-RW suffers from a similar performance decrease as RADOLAN-RY and does not outperform ResRadNet.

D. Spatial and Temporal Coherence

The spatial and temporal coherence of the rainfall fields can be seen in the maps shown in Fig. 5 and the animation

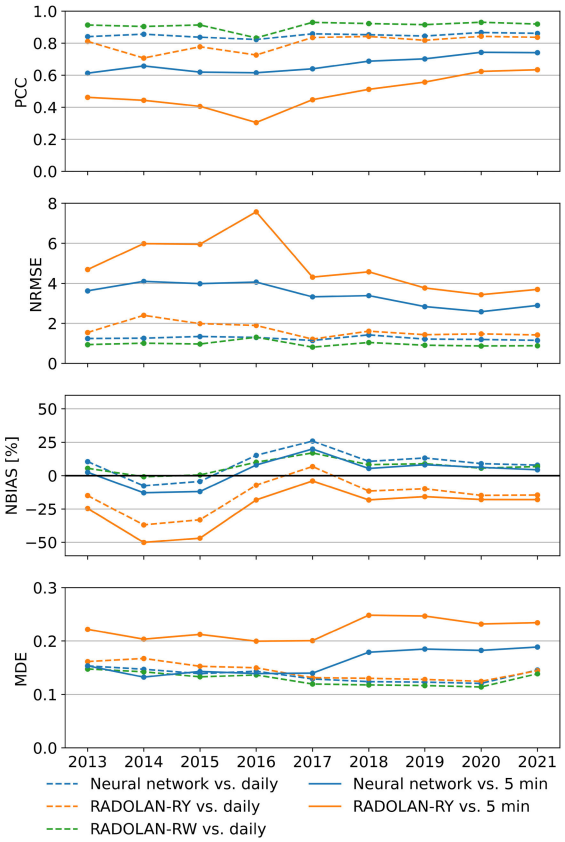

Fig. 4. Long-term performance comparison of (orange) RADOLAN-RY, (green) RADOLAN-RW, and (blue) ResRadNet. The PCC, NRMSE, NBIAS and MDE metrics are shown using the 247 5-min test (solid lines) and 1138 daily (dashed lines) rain gauges as a reference. RADOLAN-RW is not available at a 5-min resolution.

provided on Zenodo [28]. Since no radar reference at a 1-min resolution was available, the analysis of the 1-min rainfall maps from the neural network was done by visual inspection.

The maps showed that, compared to RADOLAN-RY and RADKLIM-YW, ResRadNet produced smoother structures with more gentle gradients. However, no additional structures with an artificial character like previously discovered in [29] were produced and the spatial distribution and connection of rain cells looked reasonable. Fig. 5 shows that the neural network decreased the overestimation (less red colors) in the southwest of Germany and around the upper two rain cells in the southeast without an increased underestimation (more blue colors). It even improved the severe misplacement of the small rain cell with the highest intensity close to the border of the study area in the southeast.

The temporal coherence of the neural network predictions was judged by visual inspection of the provided animation. It was similar to both radar products, that is, between 5-min time steps similar discontinuities could be observed. However, the 1-min neural network resolution removed these discontinuities at the presented scale creating a fluid motion.

To investigate if this fluid motion presented a plausible advection scheme we compared a 180-min aggregation of rainfall fields containing small, fast-moving cells. The products we used were RADOLAN-RY, RADKLIM-YW, an aggregation of every fifth neural network prediction (ResRadNet every

TABLE II
RESULTS FOR THE 1- AND 5-MIN RESOLUTION OF THE TEST DATA AND THE INDEPENDENT DAILY GAUGES

Resolution	1-Minute		5-Minute		Daily	
Model	RADOLAN-RY	ResRadNet	RADOLAN-RY	ResRadNet	RADOLAN-RY	ResRadNet
MDE	0.26	0.16	0.23	0.19	0.15	0.15
$NBIAS$	-18.0	4.4	-18.0	4.4	-15.0	7.5
PCC	0.57	0.67	0.63	0.74	0.84	0.86
$RMSE$	0.033	0.027	0.141	0.110	3.275	2.650
KS-Test	0.177	0.067	0.051	0.073	0.044	0.048

5 min) to simulate a temporal undersampling, an aggregation of all ResRadNet predictions, and the advection corrected RADOLAN-RY. The maps are shown in Fig. 6. The two most prominent observations were that the neural network was moving the center of mass of the rainfall field on the bottom toward the center of mass of RADKLIM-YW, which was gauge-adjusted. And, the discontinuities that are visible for RY, YW, and "ResRadNet every 5 min" are gone for the "ResRadNet every 1 min" and the advection corrected RADOLAN-RY and RADKLIM-YW predictions. The smoothing of the rain field and the attenuation of extreme values is less pronounced for the neural network than for the Lucas–Kanade advection correction.

E. Influence of Measurement Height Above Ground

To investigate the influence of the measurement height above ground we analyzed all time steps in the test dataset with a reference rainfall amount of at least 0.1 mm. The considered variables in those time steps were the minimal measurement height above ground and the absolute error between R_{nn1} and R_{st1}. Fig. 7 shows a linear fit to the two variables that minimizes the squared error. The slope of the linear model can be interpreted as a 3.7% increase in the mean absolute error per kilometer.

In addition to the version of ResRadNet that was presented above, we trained a second model in exactly the same way but excluded the height information by omitting the concatenate layer. The results show that the slope and intercept of the linear fit are higher for the model that does not use the height information. Here, the slope represents an increase of 4.6% of the MAE per kilometer. This model also showed a decreased PCC of 0.60 which was computed using the 1-min test data analogous to the first column of Table II where the model with the height information achieved a PCC of 0.67 and RADOLAN-RY achieved a PCC of 0.57.

IV. DISCUSSION

We evaluated the ability of a 3-D-CNN to produce spatially and temporally coherent rainfall fields at an increased temporal resolution while reducing biases to rain gauge measurements on the ground.

The spatial structure of the neural network images is smoother than for the unadjusted instantaneous radar rainfall measurements, but less smooth than the Lucas-Kanade advection corrected RADOLAN-RY product. The comparison to the rain gauges shows that the point-wise accuracy measured by PCC, RMSE, and MDE is much better for ResRadNet. We see two possible factors explaining this. On the one hand, a diffusive process due to the advection, collision, and coalescence of raindrops may lead to an actual diffusion of the distribution of rainfall on the ground. On the other hand, the location of the rain gauge in the radar pixel is not considered in this study, which leads to the neural network prediction as a maximum likely rain gauge measurement anywhere within the pixel. This may of course lead to a smoother distribution than the actual radar measurement.

Despite this slight imprecision of the rain gauge location, we could show that the added information about the minimal measurement height above ground significantly increases the model performance, thus increasing the confidence in a meaningful ground adjustment.

The temporal coherence of the neural network at a 5-min resolution is similar to RADOLAN-RY. However, the fluid motion created by the 1-min predictions indicates that the neural network acts as a reasonable optical flow estimator, even though it was trained to provide pixel-wise predictions and not spatiotemporally consistent fields. RADOLAN-RY achieved a better MDE at a 5-min resolution than at a 1-min resolution, while the opposite holds for the neural network. This also shows that the neural network is able to perform useful temporal interpolation.

The scores show in general that the neural network is able to significantly improve the rainfall estimates of RADOLAN-RY. Compared to [5] we found that our combined advection correction and ground adjustment method does impact and improve light rain estimation. The comparison at the daily resolution showed that the performance gain is a significant step toward the performance of the operational hourly gauge-adjusted product RADOLAN-RW while operating at a much higher temporal resolution. While RADKLIM-YW is gauge-adjusted and climatology-corrected and therefore assumed to be much better than the un-adjusted product, the improvement of the neural network toward RADKLIM-YW is larger than we would have anticipated. Our evaluation strategy using a set of independent rain gauges and a separate time period shows that ResRadNet is transferable to new locations with similar rainfall climatology.

A clear limitation of the neural network is that the uncertainty for extreme values cannot be reduced and that they are underestimated. However, as explained in Section III outliers close to the axis (see Fig. 4) could be reduced effectively. Another drawback is that the multiplicative correction factor was necessary to correct the NBIAS. We assume that both the extreme value underestimation and the correction factor are due to the heavily skewed distribution of precipitation.

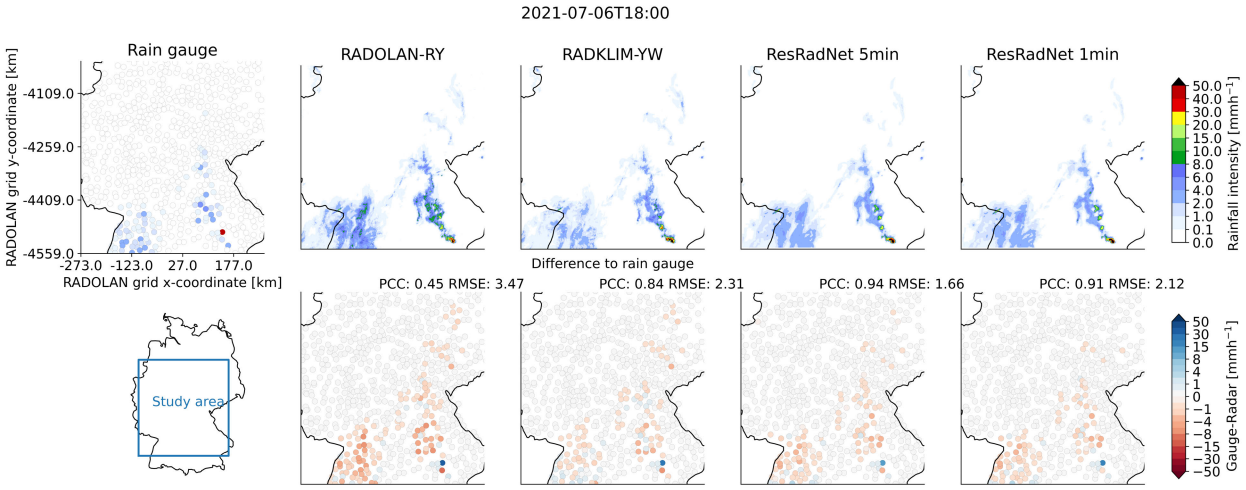

Fig. 5. Maps of rainfall intensity for 18:00 on July 6, 2021. The upper row shows the 5-min rain gauges, RADOLAN-RY, RADKLIM-YW, an aggregation of the neural network predictions from 18:00 to 18:05 (ResRadNet 5 min), and the single neural network prediction from 18:00 (ResRadNet 1 min). The bottom row shows the selected study area in Germany and the difference between the rain gauge value and the grid cell it is contained in using the product on top of the respective map. Red colors indicate an overestimation compared to the rain gauges and blue colors are an underestimation.

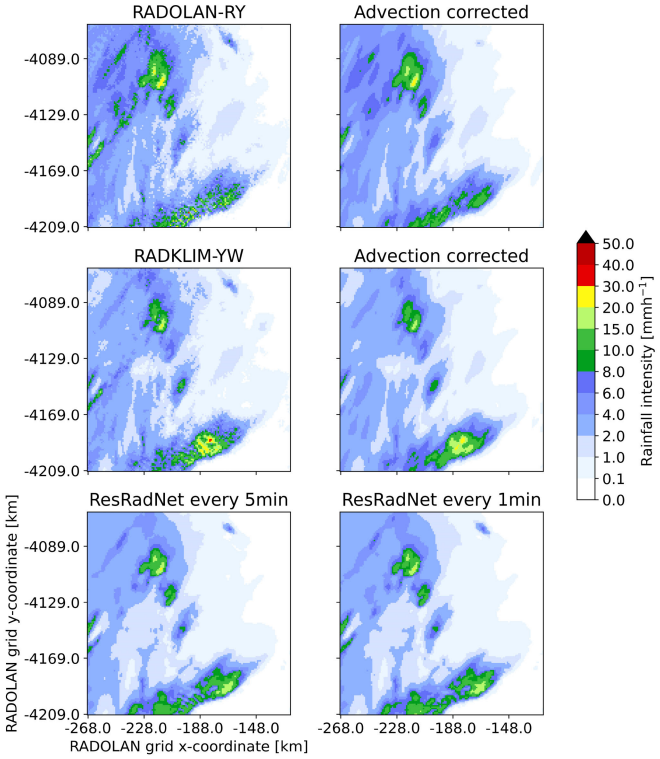

Fig. 6. Maps of rainfall sum between 16:00 and 19:00 on July 6, 2021. The products shown are from top to bottom: RADOLAN-RY, RADKLIM-YW, and the neural network predictions. The left column shows an aggregation of 5-min instantaneous measurements and the right column shows the advection corrected radar products and all 1-min time steps.

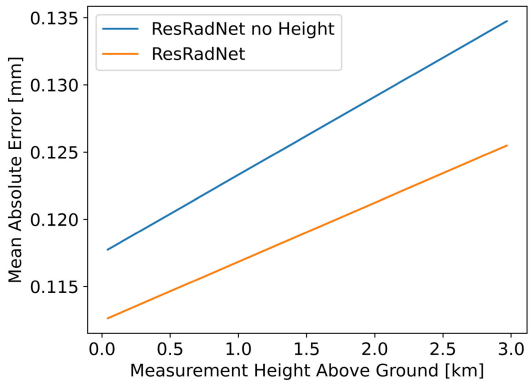

Fig. 7. Impact of measurement height above ground on rainfall estimates at a 5-min resolution. The lines show a linear fit to the absolute error between ResRadNet and the rain gauge reference in the test dataset, where the rainfall amount exceeded 0.1 mm. The orange line shows ResRadNet using the minimal measurement height above ground as described in the method section and the blue line shows a ResRadNet version where the height information is not given to the model (by omitting the concatenate layer).

In the progress of this work, we experimented with log transformations to reduce this problem, but without reasonable success. These drawbacks clearly need to be considered when using ResRadNet. However, the considerably improved scores that proved to be consistent over a period of eight years and the temporal super-resolution we achieved make our approach very valuable. Additionally, we were able to show that the model can be used to correct for advection-based temporal undersampling when aggregating multiple time steps with less smoothing of gradients than common advection correction techniques.

Finally, we emphasize the real-time applicability of our method: With the low latency of 5 s for the production of five 1-min time-steps of rainfall maps covering all of Germany, an application to a real radar system with measurements performed every 5 min is realistic. Additionally, the required data consisting of 25 min of available radar data and an estimate for the measurement height above ground should be available to potential users.

V. Conclusion

In this study, we evaluated the performance of a 3-D-convolutional residual neural network for simultaneous ground adjustment, advection correction, and temporal super-resolution of weather radar images. In an attempt to solve all three issues by training a single neural network, we were able to significantly increase the quality of the gridded country-wide 5-min radar product RADOLAN-RY.

We were able to show that 3-D-convolution in a residual network architecture is a suitable tool to increase the temporal radar resolution of 5 min by a short-term prediction of five 1-min time-steps. While our neural network ResRadNet is only trained to predict one pixel value at a time, it generates continuous predictions for neighboring pixels, resulting in spatially and temporally consistent rainfall fields. Using the raw RADOLAN-RY radar product as a baseline, the model was able to effectively reduce biases between country-wide C-band weather radar rainfall estimates and 247 1-min and 1138 daily rain gauges on the ground. By using a separate set of rain gauges for training and evaluation we demonstrated the transferability of the network to new locations. With plausible advection schemes and a 1-min resolution produced by the neural network, an exemplary case study showed that the model acts as a suitable optical flow estimator that can be used for advection correction. Despite the significant improvements that ResRadNet provides, we experienced a common issue with using neural networks for modeling precipitation. The heavily skewed distribution of rainfall leads to an underestimation of extremes and makes an additional mean field bias correction necessary. Logarithmic transformations of the input data did not yield the desired improvements. Higher potential for an accurate representation of extreme values may be given by probabilistic (ensemble) neural networks like generative adversarial networks. A potential point of view is that the current deterministic approach naturally favors the maximum likely predictions over extremes. As for this study, we point out that the produced rainfall fields may be understood as a well-performing maximum likely estimate. Future studies should aim to provide stochastic approaches to extremes.

APPENDIX

ANIMATION

The animation is available at Zenodo [28].

ACKNOWLEDGMENT

The authors would like to thank Tanja Winterrath and Elmar Weigel from DWD for the helpful discussions about the DWD radar and rain gauge data. They also like to thank Kai Mühlbauer and everyone else involved in the *ω*radlib project for providing great radar processing software.

REFERENCES

[1] V. Pejcic, P. Saavedra Garfias, K. Mühlbauer, S. Trömel, and C. Simmer, "Comparison between precipitation estimates of ground-based weather radar composites and GPM's DPR rainfall product over Germany," *Meteorologische Zeitschrift*, vol. 29, no. 6, pp. 451–466, Nov. 2020. [Online]. Available: https://www.schweizerbart.de/papers/metz/detail/29/94062/Comparison_between_precipitation_estimates_of_grou?af=crossref

[2] J. W. Wilson and E. A. Brandes, "Radar measurement of rainfall—A summary," *Bulletin Amer. Meteorological Soc.*, vol. 60, no. 9, pp. 1048–1060, Sep. 1979. [Online]. Available: https://journals.ametsoc.org/view/journals/bams/60/9/1520-0477_1979_060_1048_rmors_2_0_co_2.xml

[3] S. Pulkkinen, J. Koistinen, T. Kuitunen, and A.-M. Harri, "Probabilistic radar-gauge merging by multivariate spatiotemporal techniques," *J. Hydrol.*, vol. 542, pp. 662–678, Nov. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002216941630590X

[4] E. N. Anagnostou and W. F. Krajewski, "Real-time radar rainfall estimation. Part I: Algorithm formulation," *J. Atmos. Ocean. Technol.*, vol. 16, no. 2, pp. 189–197, Feb. 1999. [Online]. Available: https://journals.ametsoc.org/view/journals/atot/16/2/1520-0426_1999_016_0189_rtrrep_2_0_co_2.xml

[5] B.-C. Seo and W. F. Krajewski, "Correcting temporal sampling error in radar-rainfall: Effect of advection parameters and rain storm characteristics on the correction accuracy," *J. Hydrol.*, vol. 531, pp. 272–283, Dec. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002216941500267X

[6] S. Pulkkinen et al., "Pysteps: An open-source Python library for probabilistic precipitation nowcasting (v1.0)," *Geosci. Model Develop.*, vol. 12, no. 10, pp. 4185–4219, Oct. 2019. [Online]. Available: https://gmd.copernicus.org/articles/12/4185/2019/

[7] S. Vogl, P. Laux, W. Qiu, G. Mao, and H. Kunstmann, "Copula-based assimilation of radar and gauge information to derive bias-corrected precipitation fields," *Hydrol. Earth Syst. Sci.*, vol. 16, no. 7, pp. 2311–2328, Jul. 2012. [Online]. Available: https://hess.copernicus.org/articles/16/2311/2012/

[8] J. L. McKee and A. D. Binns, "A review of gauge–radar merging methods for quantitative precipitation estimation in hydrology," *Can. Water Resour. J./Revue Canadienne des Ressources Hydriques*, vol. 41, nos. 1–2, pp. 186–203, Apr. 2016, doi: 10.1080/07011784.2015.1064786.

[9] L.-P. Wang et al., "Enhancement of radar rainfall estimates for urban hydrology through optical flow temporal interpolation and Bayesian gauge-based adjustment," *J. Hydrol.*, vol. 531, pp. 408–426, Dec. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022169415003996

[10] J. Pathak et al., "FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators," 2022, *arXiv:2202.11214*.

[11] T. Yo, S. Su, J. Chu, C. Chang, and H. Kuo, "A deep learning approach to radar-based QPE," *Earth Space Sci.*, vol. 8, no. 3, Mar. 2021, Art. no. e2020EA001340, doi: 10.1029/2020ea001340.

[12] H. Chen and V. Chandrasekar, "Deep learning for surface precipitation estimation using multidimensional polarimetric radar measurements," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2021, pp. 359–362.

[13] S. Vogl, P. Laux, J. Bialas, and C. Reifenberger, "Modelling precipitation intensities from X-band radar measurements using artificial neural networks—A feasibility study for the bavarian oberland region," *Water*, vol. 14, no. 3, p. 276, Jan. 2022. [Online]. Available: https://www.mdpi.com/2073-4441/14/3/276

[14] D. Hassan, G. A. Isaac, P. A. Taylor, and D. Michelson, "Optimizing radar-based rainfall estimation using machine learning models," *Remote Sens.*, vol. 14, no. 20, p. 5188, Oct. 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/20/5188

[15] A. Moraux, S. Dewitte, B. Cornelis, and A. Munteanu, "A deep learning multimodal method for precipitation estimation," *Remote Sens.*, vol. 13, no. 16, p. 3278, Aug. 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/16/3278

[16] DWD. (2021). *DWD Climate Data Center (CDC): Historical Daily Precipitation Observations for Germany V21.3*. [Online]. Available: https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/more_precip/historical/DESCRIPTION_obsgermany_climate_daily_more_precip_historical_en.pdf

[17] H. Bartels et al. (2004). *Routineverfahren Zur Online-aneichung Der Radarniederschlagsdaten Mit Hilfe Von Automatischen Bodenniederschlagsstationen (Ombrometer)*. [Online]. Available: https://www.dwd.de/DE/leistungen/radolan/radolan_info/abschlussbericht_pdf.pdf?__blob=publicationFile&v=2

[18] T. Winterrath et al. (2017). *Erstellung Einer Radargestutzten Niederschlagsklimatologie*. [Online]. Available: https://opendata.dwd.de/climate_environment/GPCC/radarklimatologie/Dokumente/Endbericht_Radarklimatologie_final.pdf

[19] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Arti. Intell. (IJCAI)*. San Francisco, CA, USA: Morgan Kaufmann, 1981, pp. 674–679.

[20] K. Mühlbauer et al., "Wradlib/wradlib: Wradlib v1.16.2," Aug. 2022. [Online]. Available: https://zenodo.org/record/7016711

[21] J. Polz. (Mar. 2023). *Jpolz/ResRadNet: Example ResRadNet Architecture and Trained Model*. [Online]. Available: https://zenodo.org/record/7705457

[22] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*.
[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[24] S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of Adam and beyond," 2019, *arXiv:1904.09237*.
[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.
[26] M. M. Bronstein, J. Bruna, T. Cohen, and P. Velič ković, "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges," 2021, *arXiv:2104.13478*.
[27] J.-Y. Chen, S. Trömel, A. Ryzhkov, and C. Simmer, "Assessing the benefits of specific attenuation for quantitative precipitation estimation with a C-band radar network," *J. Hydrometeorol.*, vol. 22, pp. 2617–2631, Aug. 2021. [Online]. Available: https://journals.ametsoc.org/view/journals/hydr/22/10/JHM-D-20-0299.1.xml
[28] J. Polz, "Supplementary animation for resradnet," Mar. 2023, doi: 10.5281/zenodo.7723004.
[29] L. Glawion, J. Polz, H. Kunstmann, B. Fersch, and C. Chwala, "spateGAN: Spatio-temporal downscaling of rainfall fields using a cGAN approach," *Earth Space Sci.*, vol. 10, no. 10, 2023, Art. no. e2023EA002906, doi: 10.1029/2023EA002906.

Julius Polz received the B.Sc. and M.Sc. degrees in mathematics from Ludwig-Maximilians-University, Munich, Germany, in 2018. He is currently pursuing the Ph.D. degree in climate and environmental sciences with the Faculty of Applied Informatics, University of Augsburg, Augsburg, Germany.

He is a Doctoral Researcher with the DFG Research Unit RealPEP, Karlsruhe Institute of Technology (KIT) Campus Alpin, Garmisch-Partenkirchen, Germany. His research interests include machine learning and precipitation remote sensing with commercial microwave links and weather radars.

Luca Glawion received the B.Sc. degree in geography and the M.Sc. degree in climate and environmental sciences from the University of Augsburg, Augsburg, Germany, in 2021. He is currently pursuing the Ph.D. degree in environmental and climate sciences with the Faculty of Applied Informatics, University of Augsburg.

He is a Doctoral Researcher with HGF Innopool Project SCENIC, Karlsruhe Institute of Technology (KIT) Campus Alpin, Garmisch-Partenkirchen, Germany. His research interests include machine learning and precipitation downscaling.

Hiob Gebisso received the B.Sc. degree in mathematics from Saint Louis University, Saint Louis, MO, USA, in 2017, and the M.Sc. degree in stochastic engineering from Munich University of Applied Sciences, Munich, Germany, in 2022. For the master's degree, he conducted research in the field of machine learning based adjustment of weather radars for quantitative precipitation estimation with Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany.

He is currently a Data Scientist with Arvato Systems, GmbH, Gerlingen, Germany, where he is in charge of designing and developing forecasting and pricing models for ebooks.

Lukas Altenstrasser received the B.Sc. degree in business informatics from Munich University of Applied Sciences, Munich, Germany, in 2022, where he is currently pursuing the M.Sc. degree.

He completed his training as an IT specialist for systems integration with Retarus, GmbH, Gerlingen, Germany, in 2017. In 2021, he was an Intern with Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany, where he conducted research in the field of machine learning based adjustment of weather radars.

Maximilian Graf received the B.Sc. degree in geography from the Catholic University of Eichstätt-Ingolstadt, Eichstätt, Germany, in 2014, and the M.Sc. degree in climate and environmental sciences from the University of Augsburg, Augsburg, Germany, in 2018, where he is currently pursuing the Ph.D. degree in climate and environmental sciences with the Faculty of Applied Informatics.

He is also with the BMBF Project, HoWa-PRO. His research interests include improving and merging rainfall estimates from opportunistic and traditional rainfall sensors.

Harald Kunstmann received the Diploma degree in physics from the University of Heidelberg, Heidelberg, Germany, in 1994, and the Ph.D. degree in environmental natural sciences from ETH Zürich, Zürich, Switzerland, in 1998.

He is currently a Professor and the Chair for regional climate and hydrology with the University of Augsburg in joint appointment with Karlsruhe Institute of Technology (KIT), Campus Alpin, Garmisch-Partenkirchen, Germany. His research comprises hydrological climate change impacts, high-resolution coupled atmosphere-hydrology modeling, geostatistical merging of hydrometeorological variables, and quantification of spatio-temporal precipitation variability by attenuation analysis of commercial microwave links.

Stefanie Vogl received the Diploma degree in mathematics from the University of Regensburg, Regensburg, Germany, in 2004, and the Dr.rer.nat. degree in physics from Ludwig-Maximilians-University, Munich, Germany, in 2009.

She is currently a Professor of statistical learning and artificial intelligence with the Department of Computer Science and Mathematics, Munich University of Applied Sciences, Munich. She works on the development and application of neural network architectures and learning algorithms with focus on time series.

Christian Chwala received the Diploma degree in physics from the University of Regensburg, Regensburg, Germany, in 2008, and the Dr.rer.nat. degree in climate and environmental science from the University of Augsburg, Augsburg, Germany, in 2015.

His work focuses on the usage of opportunistic rainfall sensors and merging of weather radar data with ground-based observations for which he also applies data-driven methods. He currently leads several research projects in this field at Karlsruhe Institute of Technology (KIT), Campus Alpin, Garmisch-Partenkirchen, Germany.