https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


# A rainfall estimation ensemble method using a cGAN for SEVIRI geostationary satellite data

Selina Janner[1], Julius Polz[1,2], Luca Glawion[1], Harald Kunstmann[1,3], and Christian Chwala[1]

[1]Institute of Meteorology and Climate Research - Atmospheric Environmental Research (IMK-IFU), Campus Alpin, Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany
[2]Institute of Meteorology and Climate Research - Atmospheric Trace Gases and Remote Sensing (IMK-ASF), Karlsruhe Institute of Technology, Karlsruhe, Germany
[3]Institute of Geography, University of Augsburg, Augsburg, Germany

**Correspondence:** Selina Janner (selina.janner@kit.edu)

**Abstract.** Geostationary satellites (GEOs) provide real-time information on the atmosphere and can be used for quantitative precipitation estimation (QPE) with low latency. The relation between GEO observations and precipitation is complex, though, limiting QPE performance.

Here we present a conditional generative adversarial network (cGAN) to perform QPE using data from SEVIRI infrared and water vapor channels. The model is trained with precipitation maps derived from the German radar network (RADKLIM-YW).

The model is evaluated in Germany for a separate time period and compared with the operational GEO QPE product PDIR-Now. Our main findings are: 1. The cGAN produces more realistic precipitation structures and intensities (including high precipitation rates) and has a higher location accuracy then the benchmark product; 2. Ensemble predictions are spatially under dispersed, uncertainty in precipitation intensity is captured better than uncertainty in precipitation location. Ensemble spread and other limitations, such as spurious amplified errors at extreme intensities, do not substantially reduce the overall skill improvements over the benchmark product. Future work should focus on extending the framework across a other geographic domains to assess generalization under varying viewing geometries and precipitation regimes.

## 1 Introduction

Precipitation is a fundamental component of the hydrological cycle and the major driver for hazards such as floods and droughts. Accurate, high-resolution, near real-time precipitation information is essential for weather and hydrological models, fore/now-casting and early warning systems. Rain-gauge measurements and radar precipitation products are dedicated tools for Quantitative Precipitation Estimation (QPE), but their coverage is sparse or nonexistent in many regions of the world. In contrast, satellites provide extensive spatial coverage across political and economical borders. Due to their distance and top-down view, their usage for rainfall observations comes with certain challenges and limitations, though.

Satellite QPE can be based on data from low Earth orbiting satellites (LEOs) or geostationary satellites (GEOs). LEOs can host active or passive microwave sensors, which are directly built for QPE or provide indirect measurements sufficient for QPE. Their limitation is the infrequent revisit time due to their orbit making them unsuitable for real-time application for a highly

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


dynamic atmospheric process such as precipitation. The Global Precipitation Project (GPM; Hou et al. (2014)) combines many LEOs into one precipitation product, but the average revisit time in the tropics is still above one hour. GEOs offer a much higher temporal resolution , but their large orbital altitude prevents the use of active and passive microwave sensors for QPE and results in coarser spatial resolution. The Field of view of the SEVIRI instrument (Aminou et al., 1997) aboard the Meteosat Second Generation (MSG), which is used in this work, is centered around the prime meridian an encompasses all of Africa, most of Europe, parts of south America and western Asia.

SEVIRI scan strategy (south-north while spinning east-west) captures the same area every 15 minutes with an instantaneous measurement. Similar instruments are aboard other GEOs like MTG, GOES and HIMAWARI, making GEOs attractive for real-time applications across the globe.

However, deriving accurate rainfall estimates from GEO data is fundamentally difficult (Levizzani and Kidd, 2011). GEO sensors operating in the visible spectrum (VIS) and thermal infrared measure upwelling radiance near cloud top, not the precipitation processes occurring within and below the cloud column. Cold cloud-top brightness temperatures are a statistical proxy for deep convection, but the relationship has large variance: warm-topped clouds can produce significant rainfall without anomalously cold tops, while cold cirrus anvils can overlie precipitation-free areas (Roebeling and Holleman, 2009). Unlike microwave retrievals, which respond more directly to hydrometeors through the cloud column, GEO-QPE therefore relies inherently on statistical associations and spatial cloud patterns rather than point-scale physical signals. This makes the problem well-suited to data-driven approaches capable of exploiting large-scale contextual information, but it also means that irreducible uncertainty in the relation from cloud-top to surface-rainfall cannot be eliminated by any retrieval method.

A multitude of approaches have been developed to address this challenge, ranging from simple regression models (Roebeling and Holleman, 2009) to random forests (Kühnlein et al., 2014) and convolutional neural networks (Sadeghi et al., 2019; Nguyen et al., 2018b, 2020a, b; Pfreundschuh et al., 2022). PERSIANN-CCS (Hong et al., 2004) incorporates spatial context by segmenting adjacent pixels into cloud objects classified by temperature, size and texture; its successor PDIR-Now (Nguyen et al., 2020a) extends this with a temporal dimension and physics-informed corrections. Despite its advances, the PDIR logarithm yields spatially smooth outputs and systematically underestimates the intensity and spatial variability of extreme events. More generally, AI models with pixel-based training loss such as MSE (Lang et al. (2024), Lam et al. (2023)) tend to produce spatially smooth predictions, which can lead to an underestimation of extremes. The deep convolutional network Hydronn (Pfreundschuh et al., 2022) estimates per-pixel rainfall probability distributions, but requires heuristic post-processing for temporal accumulation.

This leaves a gap at the intersection of reproducing extreme precipitation amounts, while being consistent in space and time. Generative models like conditional generative adversarial networks (cGAN) are structurally better suited to this problem (Isola et al., 2017; Leinonen et al., 2021), though few studies have explored this direction for GEO-QPE. Hayatbini et al. (2019) demonstrated that cGAN training substantially improves precipitation estimation from GOES-16 multispectral imagery relative to the 2004 PERSIANN-CCS baseline. No generative GEO-QPE model has since been applied to SEVIRI, evaluated against a more recent operational baseline such as PDIR-Now, assessed for spatial structure skill, or extended to provide ensemble-based uncertainty quantification. Interest in generative approaches for precipitation retrieval is nonetheless growing: Guilloteau et al.

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


(2025) demonstrated a diffusion model conditioned on coincident GOES and passive microwave data, while Ward-Leikis et al. (2025) found cGANs competitive with diffusion models for precipitation downscaling at substantially lower computational cost.

This study asks whether a cGAN can produce skillful, spatiotemporally consistent precipitation estimates from SEVIRI observations. We further assess whether adversarial training improves not only detection and intensity metrics but also the spatial structure of predicted fields and thus a realistic distribution of rainfall values, including extreme values. We limit the study region to Germany were suitable reference data from the radar network is available to serve as a target dataset for the cGAN. We evaluate skill against PDIR-Now using both standard point-wise scores and spatial structure diagnostics, and provide ensemble uncertainty estimates.

## 2 Data and Methods

### 2.1 Data

Our cGAN uses SEVIRI level 1.5 data as input and is trained using the RADKLIM-YW radar product as reference. We use the PERSIANN PDIR-Now product as benchmark for comparison.

#### 2.1.1 Domain and temporal coverage

In this work, which focuses on the methodological development and evaluation of our GEO QPE probabilistic cGAN, we use only data from Germany where high-resolution (in time and space) high-quality weather radar products are available.

For training, data from May to November of 2021 is used, for testing we use April 2021 and for evaluation April to September of 2019.

#### 2.1.2 SEVIRI Level 1.5

For the condition (the input) of the cGAN we use geostationary satellite data from the Meteosat Second Generation (MSG) series. The Spinning Enhanced Visible and InfraRed Imager (SEVIRI) is a multispectral imaging instrument providing continuous observations of the full disc (i.e., providing continuous hemispheric-scale images rather than swath coverage) (Aminou et al., 1997). The 12 channels are scanned every 15 minutes in a south-north and east-west direction. In our work we use the 10.8 $\mu$m Infrared (IR) channel and the 7.3 $\mu$m watervapor (WV) channel. Both of these, contrary to visual light channels, measure brightness temperature continuously at day and night. The level 1.5 data includes correction of geometric and radiometric effects and is geolocated. The field of view is centered at 0° N/ 0° E, there the resolution is 3 km. We limit the extent to the domain of RADKLIM-YW while keeping the SEVIRI data on its native grid. To be able to use this as input for the cGAN we use a fixed min-max normalization on each channel. One example of the normalized brightness temperature is shown in Fig. 2 (a) and (d).

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


#### 2.1.3 RADKLIM-YW

As target/reference we use the RADKLIM-YW dataset from the German Weather Service (DWD) (Winterrath et al., 2018). RADKLIM-YW provides 5 min rainfall sums from quality-controlled, climatologically corrected and gauge-adjusted weather radar observations on a 1 km grid for Germany. We linearly interpolate RADKLIM-YW to the SEVIRI grid and only use time stamps at 10, 25, 40 and 55 min past the hour as those are closest to the scanning time of SEVIRI for our domain. Figure 2 (b) shows an example scene of precipitation from RADKLIM-YW. In Fig. 2(e) accumulated hourly precipitation rate is shown for comparison with the benchmark dataset that is only available at 1 h resolution.

#### 2.1.4 PDIR-Now

As a benchmark dataset we use PERSIANN Dynamic Infrared–Rain Rate (PDIR-Now) (Nguyen et al., 2020a). PDIR-Now is based on infrared channels from multiple geostationary satellites. The PDIR algorithm is based on PERSIANN-CCS (Hong et al., 2004). Similar to PERSIANN-CCS, it classifies cloud patches based on cloud-top temperature, texture, and size, and relates these cloud characteristics to rainfall rates. In contrast to PERSIANN-CCS, the PDIR algorithm dynamically adjusts the temperature–rainfall relationships using regional rainfall climatology. The operational PDIR-Now product further applies bias correction using LEO microwave precipitation estimates. PDIR-Now has improved regional precipitation patterns and rain/no-rain classification. It provides precipitation estimates with a latency of 15–60 minutes at 0.04° ( 4 km) spatial and hourly temporal resolution. We interpolated (nearest-neighbor) the data to the SEVIRI-grid in the RADKLIM-YW domain. In Fig. 2(g) an example scene is shown.

### 2.2 Methods

In this work we introduce a conditional Generative Adversarial Network (cGAN) (Mirza and Osindero, 2014) for GEOS QPE. The trained cGAN generates rainfall maps for a specific condition (input), the SEVIRI data, and produces an ensemble of predictions. The quality of these predictions is measured against gauge adjusted weather radar data. In this section data pre-processing, model architecture and the metrics used for model selection and evaluation are described.

#### 2.2.1 Model Architecture and Training

Our model applies an adversarial training objective illustrated in Fig. 1. In adversarial training, a generator and a discriminator are trained jointly with opposing objectives. The generator receives two SEVIRI channels, the 7.3 μm water vapor (WV) and the 10.8 μm infrared (IR) channel, at two consecutive time steps and produces a precipitation estimate for the latter time step. The discriminator, used only during training, receives the SEVIRI data (latter time step) concatenated with either the generator's prediction or the RADKLIM-YW target, and classifies the precipitation map as real or fake. After each batch of 100 samples, the discriminator is updated with binary cross-entropy (BCE) loss, training it to better distinguish real from generated fields. The generator is updated with a weighted sum of BCE (encouraging it to fool the discriminator) and MSE (penalizing deviations from the target). BCE alone would produce realistic spatial patterns but poorly calibrated intensities;

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026

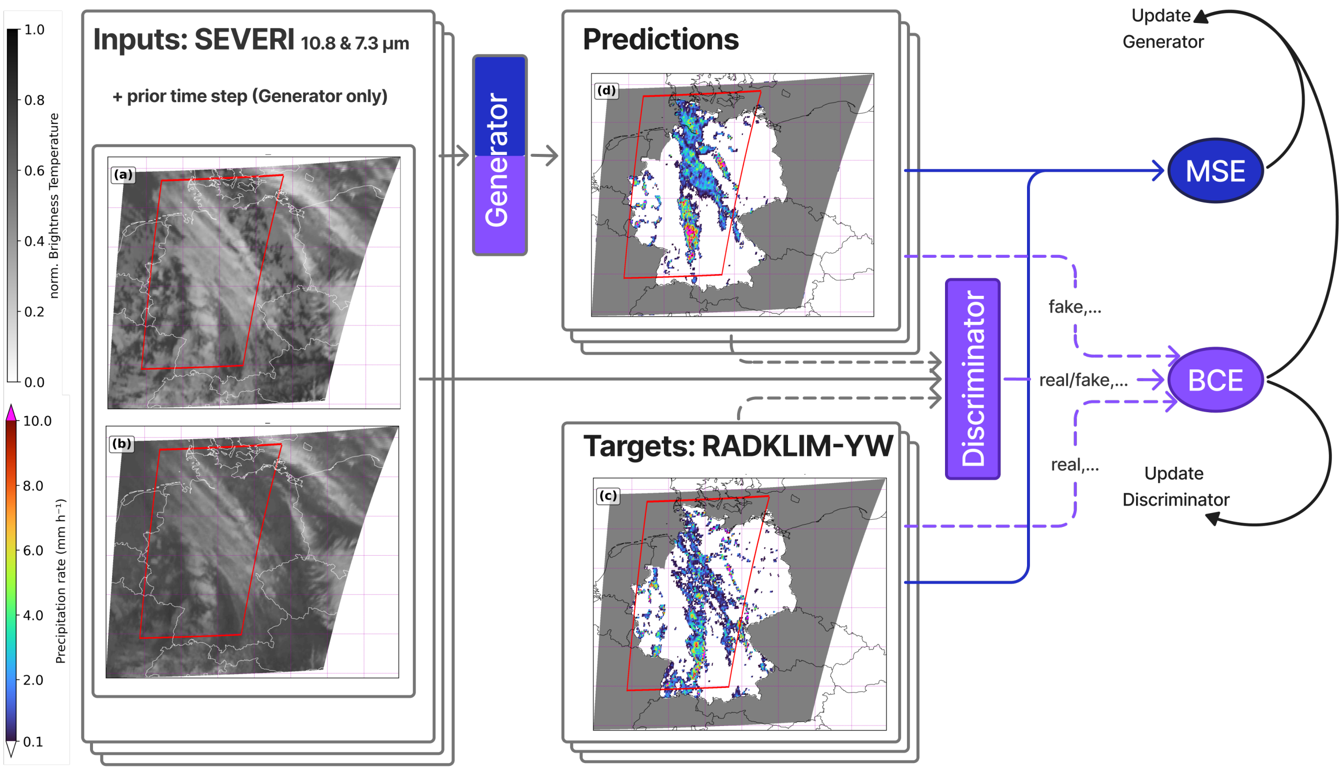




**Figure 1.** Schematic overview of cGAN training; with examples for SEVIRI (a,b), RADKLIM-YW (c) and cGAN (d) from the 27.04.2019 15:25 UTC (same as Fig. 2) as visualizations, importantly these is not a training images but red box shows the size and potential position of a training patch

MSE alone would produce well-calibrated but overly smooth fields. Achieving the right balance is critical: if the discriminator is too strong, it provides uninformative gradients to the generator, which then effectively ignores the BCE loss; if it is too weak, the generator can trivially fool it, risking mode collapse.

### The Generator

The generator is a two-level 3D U-Net Nguyen et al. (2018a) that maps a spatio-temporal SEVIRI input (two channels over two consecutive time steps) to a single predicted precipitation field at the same spatial resolution. Each encoder level applies a 3D ResNet block followed by dropout and 3D max-pooling, with skip connections taken directly post-dropout. The first level additionally incorporates a Squeeze-and-Excitation (SE) module Hu et al. (2018). Because the temporal dimension is fully collapsed after the first pooling step, subsequent pooling operates on the spatial dimensions only, so temporal resolution is reduced in a single step while spatial resolution is halved at each level. The bottleneck is a plain 3D ResNet block. Each decoder level upsamples via a resize-convolution (trilinear interpolation followed by a convolution), concatenates the corresponding encoder skip, and passes the result through a plain 3D ResNet block, mirroring the encoder structure. A final strided temporal

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


convolution collapses the time dimension, and a point wise convolution with LeakyReLU produces the single-channel output. Gal and Ghahramani (2016) showed that dropout can be interpreted as an approximate variational inference in a Bayesian neural network (Monte-Carlo dropout) and has been widely applied for cGANs (Isola et al. (2017), Brecht and Bihlo (2023) and Bihlo (2021)). Here, all dropout layers are active during inference, in order to run multiple forward passes of the same data to create an ensemble of predictions.

**The Discriminator**

The discriminator is a convolutional neural network used as a binary classifier of generated versus real precipitation scenes. Its input is formed by channel-concatenating SEVIRI WV and IR with either the generator output or the corresponding RADKLIM-YW reference. A series of strided convolutional blocks, each followed by batch normalization and LeakyReLU activation, progressively reduce the spatial and temporal resolution while increasing the number of feature channels. A point wise convolution then projects the feature maps to a single channel, and adaptive average pooling produces a logit per sample.

**Data Pre-processing**

RADKLIM-YW calibration is limited to german rain gauge network, to ensure high quality reference data we only consider grid points with radar availability that are contained entirely within Germany as valid. Masking is done with the *regionmask* python package Hauser et al. (2024).

For training, aligned patches (128 by 128 grid points) of RADKLIM-YW (5min rainfall amount) and SEVIRI (IR 10.8 $\mu$m, WV 7.3 $\mu$m brightness temperature) are used. To be able to train using data from the entire domain we allow up to 5000 masked RADKLIM-YW values in the randomly selected training patches. This way a large area of each patch can be used for loss calculation, while still including samples from the entire valid domain. The training samples are randomly selected and additionally flipped and/or rotated to increase sample diversity and prevent the model from overfitting to geographical and meteorological features (f.e. oceanic/continental climate regions or prevailing westerlies).

The fully convolutional generator can handle diverse input shapes. For testing and evaluation larger samples (160 by 280 grid points) are used that cover all of Germany (GER), the extent is shown in Figure 2.

The SEVIRI data is min-max normalized per channel shifting the range between zero and one. Radar precipitation is log-normalized (Eq. 1) shifting the heavily right skewed distribution to a more uniform distribution, suitable for ML training. Log-normalization is commonly used for precipitation (f.e. by Ayzel et al. (2020) and Liu et al. (2024)).

$$y_{\text{norm}} = \log_{10}(y+0.1) - \log_{10}(0.1) \quad (1)$$

### 2.2.2 Model selection

During training the generator is saved after every epoch. To account for the often noisy equilibrium-seeking rather than loss-minimizing adversarial training process the best model state is selected based on the mean FSS in the GER area during the

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


testing period (April 2021). FSS was successfully used for model state selection for a GAN with precipitation output in Glawion et al. (2025). The raw model output is the normalized 5min-rainfall amount. After denormalization, FSS is calculated for spatial scales of 1,3,9,15, and 45 grid points and thresholds $0.01\,\mathrm{mm}$, $0.1\,\mathrm{mm}$, and $1\,\mathrm{mm}$. Spatial scales include various neighborhood sizes in order to capture the spatial offsets from single grid point to supra-regional, with a focus on the small scale offsets. Thresholds were selected to range from the smallest precipitation resolved in the RADKLIM-YW data to heavy precipitation events. By including various scales and thresholds the model checkpoint with the largest mean FSS score is considered to have the most balanced skill and is used for subsequent evaluation.

#### 2.2.3 Metrics

This section introduces the metrics that were assessed to quantify the quality of the predicted rain fields. We considered the Fraction skill score (FSS) and the Radially averaged power spectrum density (RAPSD). The other used metrics are described in the Appendix A.

**Fraction Skill score (FSS)**

The FSS is a neighborhood based score. Compared to other pixel-wise metrics like MSE or PCC it can account for predictions being slightly shifted in space compared to the reference. We use the method from Antonio and Aitchison (2025), which calculates how well predicted and reference maps match for all samples at once. A FSS close to 1 indicates strong agreement, a score close to 0 signals weak/no agreement. The scale at which predictions are considered useful is the scale at which FSS reaches $FSS_{uniform}$ (Eq. 2, (Roberts and Lean, 2008)), and is dependent on the fractional coverage $f$ of the precipitation with the selected threshold.

$$FSS_{\mathrm{uniform}} = 0.5 + f/2 \tag{2}$$

We calculate the FSS for several window sizes ("scales") and thresholds (for evaluation period April - September 2019 $0.1\,\mathrm{mm}$, $1\,\mathrm{mm}$, and $10\,\mathrm{mm}$ at 1,3,9,15, and 45 grid points). When using the term mean FSS we refer to the mean FSS of all considered scales and thresholds. We use FSS both for model selection and evaluation of predictions.

**Radially averaged power spectrum density (RAPSD)**

The RAPSD shows frequency dependent power of an 2D image. We use the pysteps (Pulkkinen et al., 2019) implementation, which applies a 2D Fourier-Transformation on the rainfall maps and then computes the radial average of the resulting power spectrum over all directions for each frequency. When referring to the RAPSD in this study, we refer to the mean RAPSD of all scenes in the considered time period. The RAPSD is useful for quantifying structures in precipitation fields. A systematic offset in magnitude can result from an overall intensity bias, while the slope of the spectrum reflects the relative contribution of small- versus large-scale patterns. Spikes and sudden drops in RAPSD point towards artifacts in the precipitation field.

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026

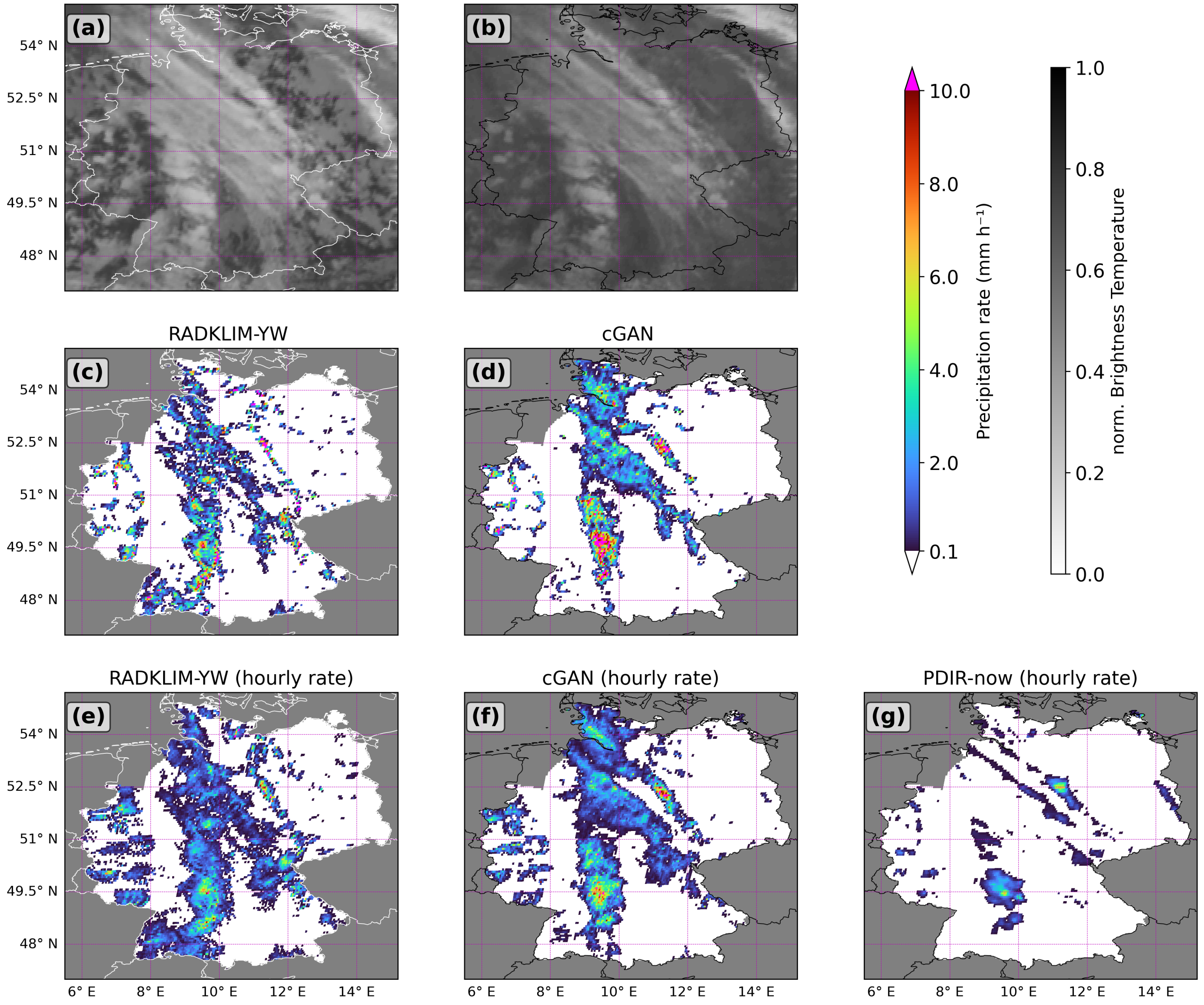




**Figure 2.** normalized SEVIRI IR (a) and WV (b) brightness temperature, RADKLIM-YW (c) and cGAN prediction (d) precipitation rate for 15:25 UTC 27.04.2019 and hourly precipitation rate 15-16 UTC for RADKLIM-YW (e), cGAN (f) and PDIR-Now (g)

**Table 1.** averages of MSE in $\mathrm{mm^2\,h^{-2}}$, MAE in $\mathrm{mm\,h^{-1}}$, PCC, Bias and totals of TP, TN, FN, FP and the POD, FARate, FARatio, Accuracy and Bias Ratio calculated from the totals; for Apr-Sep 2019 and an example scene in GER region calculated with precipitation rates of RADKLIM-YW and predictions from cGAN

| Metric | April to September | 27.4.19 19:25 UTC |
|---|---:|---:|
| TP | 7641550 | 2784 |
| TN | 269270972 | 37635 |
| FP | 9467565 | 3532 |
| FN | 7579901 | 1969 |
| POD | 0.5020 | 0.5857 |
| FARate | 0.0340 | 0.0858 |
| FARatio | 0.5534 | 0.5592 |
| CSI | 0.3095 | 0.3360 |
| Accuracy | 0.9420 | 0.8802 |
| Bias Ratio | 1.1240 | 1.3288 |
| MSE | 1.0028 | 3.1206 |
| MAE | 0.1308 | 0.6347 |
| PCC | 0.2489 | 0.3252 |
| Bias | -0.1665 | 0.4840 |

## 3 Results

### 3.1 cGAN Predictions

The cGAN predictions are available every 15 minutes in line with the SEVIRI scan frequency and reflect 5-minute RADKLIM-YW rainfall sums at the same timestep. At this sub-hourly temporal resolution, no comparable GEO QPE product is available.

The MAE compared to RADKLIM is approx. 0.13 and 0.63 for the April to September 2019 period and the example scene (Fig. 2) respectively. The critical success index (CSI) is approx. 0.31 and 0.34 (Table 1). These and most other metrics are spatially averaged for the valid RADKLIM-YW extent in Germany (and temporally for April to September column).

### 3.2 Hourly Accumulated Predictions Compared to PDIR-Now

In this section we evaluate our model results by comparing its hourly averages of precipitation rate from April to September 2019 to those of PDIR-now. The reference is the hourly precipitation rate calculated from RADKLIM-YW data. A video supplement (https://doi.org/10.5446/73585) visualizes the evolution of precipitation across multiple days. The video show hourly precipitation rates for RADKLIM-YW, PDIR-Now and the cGAN alongside one time slice from the two SEVIRI channels, used for model input, in the middle of each hour.

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026

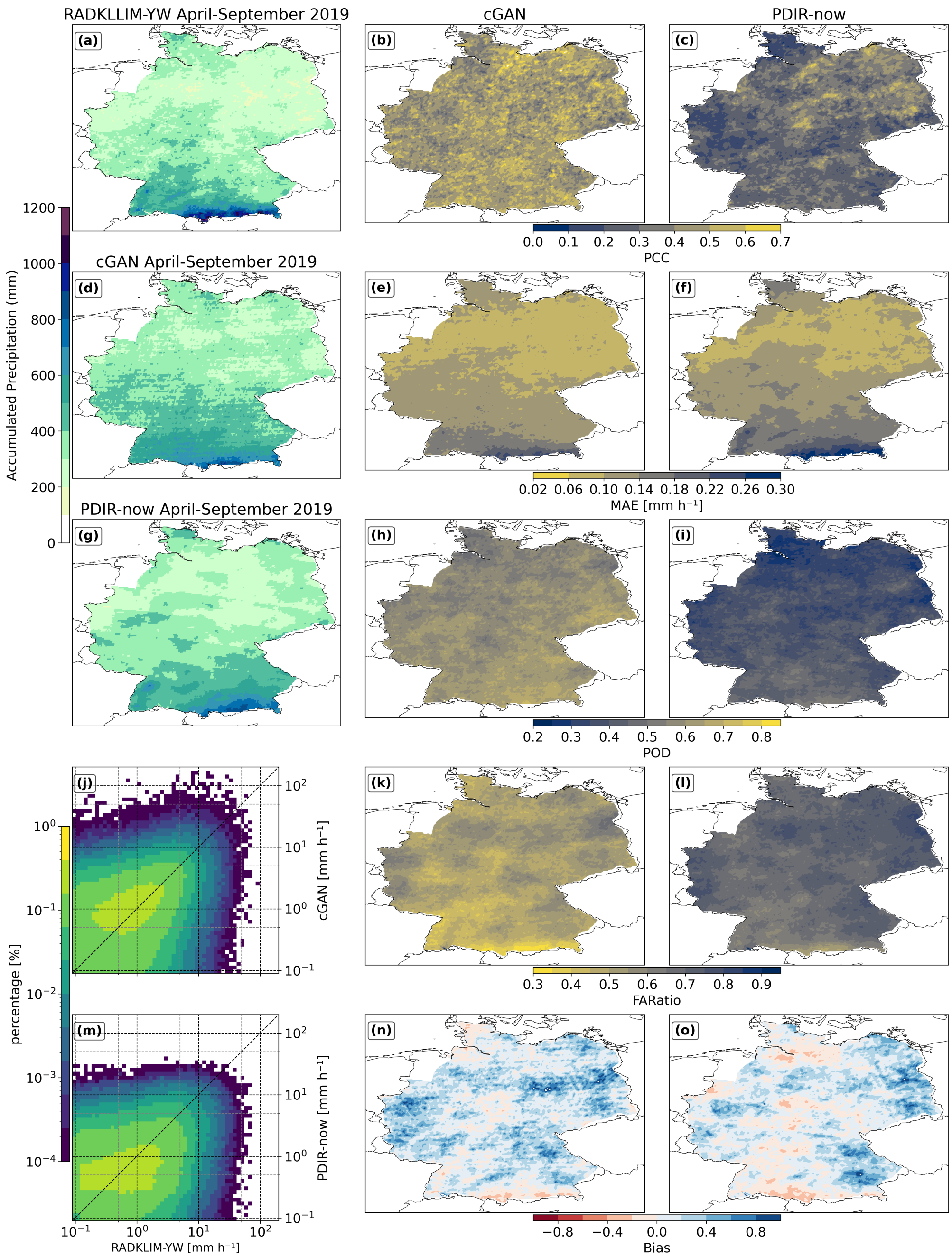




**Figure 3.** April-September 2019 evaluation of cGAN predictions and comparison to PDIR-now based on hourly precipitation [in $\mathrm{mm\,h^{-1}}$]; Accumulated Precipitation for RADKLIM-YW (a), cGAN (d) and PDIR-now (g); 2D histogram in percentage of values pairs for RADKLIM-YW and predictions from cGAN (j) and PDIR-now (m); Maps of mean PCC, MAE, POD, FA Ratio and Bias for cGAN (middle column) and PDIR-now (right column)

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


**Table 2.** same as Table 1 but for Apr-Sep 2019 in GER region calculated from hourly precipitation rates RADKLIM-YW, cGAN and PDIR-Now

| Metric | cGAN | PDIR-Now |
|---|---|---|
| TP | 2976478 | 1985001 |
| TN | 63555602 | 62212626 |
| FP | 2900127 | 4243103 |
| FN | 2025728 | 3017205 |
| POD | 0.5950 | 0.3968 |
| FARate | 0.0436 | 0.0638 |
| FARatio | 0.4935 | 0.6813 |
| CSI | 0.3767 | 0.2147 |
| Accuracy | 0.9311 | 0.8984 |
| Bias Ratio | 1.1748 | 1.2451 |
| MSE | 0.3540 | 0.3898 |
| MAE | 0.1099 | 0.1271 |
| PCC | 0.4406 | 0.2858 |
| Bias | -0.1464 | -0.0871 |

### 3.2.1 Precipitation amounts and intensities

The accumulated precipitation totals for the April–September 2019 evaluation period are shown in Figure 3 (a: RADKLIM-YW, d: cGAN, g: PDIR-now). RADKLIM-YW totals range from above 900 mm in the alps to below 200 mm in parts of northern and eastern Germany. Both models (cGAN and PDIR-now) capture the north–south gradient, but compress its range: high totals are underestimated and low totals are overestimated. The cGAN mean bias over the domain is -0.1464, compared to -0.0871 for PDIR-now (Table 2), indicating that PDIR-now better preserves the domain-averaged total. Accumulated fields from the cGAN exhibit periodic grid artifacts. Their origin and the trade-off involved in retaining them are discussed in Section 4. PDIR-now instead shows faint diagonal line artifacts in the accumulated totals.

The mean pixel-wise metrics during the evaluation period show a different spatial distribution and pattern. The mean MSE and MAE of the cGAN are smaller and the PCC larger than those of PDIR-now (Table 2). Spatially, the cGAN's MAE advantage over PDIR-now is present across the entire domain and is most pronounced in the southern mountain ranges (Fig. 3, middle versus right column). Orography-following variations in skill are not visible for the PCC (Fig. 3 b,c), where only small-scale fluctuations can be observed for the cGAN. PDIR-now has more pronounced regional differences with large areas of low PCC values. In summary, the cGAN shows more spatially uniform predictive skill and higher pixel-wise accuracy.

The 2D histograms (Fig. 3 j, m) and the precipitation rate histogram (Fig. 4 c) reveal differences in the predicted intensity range. The cGAN produces precipitation intensities exceeding $100\,\mathrm{mm\,h^{-1}}$ (maxima are given in Fig. 4 c), its upper tail is

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026

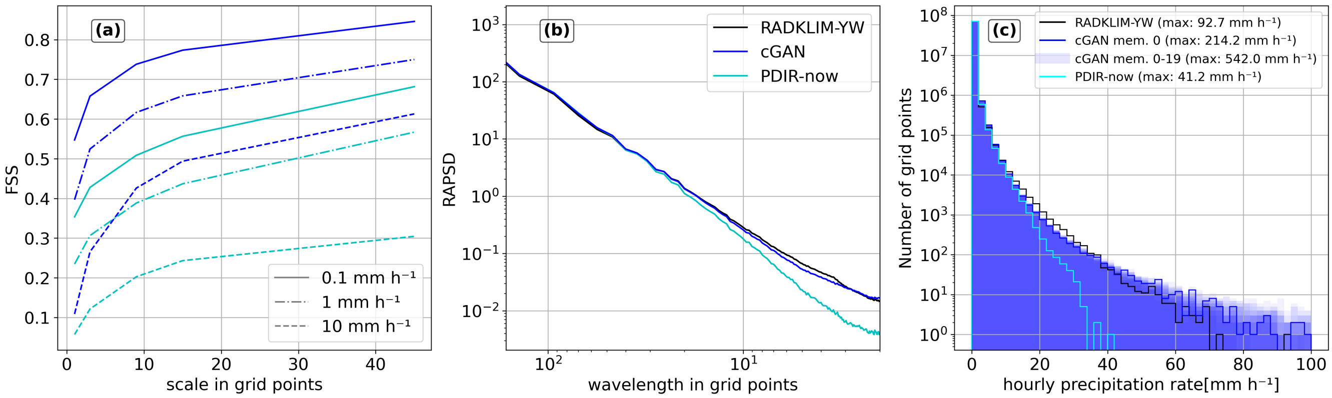




**Figure 4.** April-September 2019 skill evaluation of cGAN and PDIR-Now against RADKLIM-YW. (a) FSS as a function of spatial scale (grid spacing on average equals 3.3 km east/west to 6.2 km north/south) for three precipitation thresholds (0.1, 1, and 10 $\mathrm{mm\,h^{-1}}$) (b) Radially averaged power spectral density of hourly precipitation fields. (c) Histogram of hourly precipitation rates; the shaded blue band shows density of cGAN members 0-19.

comparable to the RADKLIM-YW distribution, while PDIR-now predictions are bounded at approximately $40\mathrm{mm\,h^{-1}}$. The cGAN's density in the 2D histogram (Fig. 3 j) runs closer to the diagonal across moderate to high observed rates, indicating that the cGAN more frequently assigns predicted rates consistent with the observed intensity. PDIR-now (Fig. 3 m), by contrast, more often assigns low predicted rates ($<1\mathrm{mm\,h^{-1}}$) even when observed rates are high. The cGAN's peak density is positioned closer to the diagonal than that of PDIR-now, consistent with its higher domain-mean PCC. Relative to RADKLIM-YW deviations in the precipitation distribution remain: underestimated counts in the 10–$36\mathrm{mm\,h^{-1}}$ and excess below this range and at the rare very high precipitation rates above (Fig. 4c).

The precipitation rate histogram for the 2 August 2019 18:40 UTC event (Fig. 5) shows cGAN Member 0 alongside Members 1-19 as a shaded band. The cGAN reproduces precipitation rates above $10\mathrm{mm\,h^{-1}}$, which are absent from PDIR-now predictions for this event (see maps in Fig. 5). Member 0 underestimates counts in the 10–$40\mathrm{mm\,h^{-1}}$ range relative to the reference and shows excess at rare very high values; spread across members in this range is examined further in Section 3.3.2. Distributional results in the second presented scene (Fig. 6) are similar though PDIR-now is able to reproduce this less extreme distribution to a higher degree.

### 3.2.2 Discrimination between rain and no rain

At a detection threshold of $0.1\mathrm{mm\,h^{-1}}$, the cGAN achieves higher domain-mean POD and CSI and lower FA Ratio and FA Rate than PDIR-now (Table 2). The spatial maps of POD and FA Ratio (Fig. 3 h,i,k,l) show that these differences are consistent across the domain rather than confined to particular regions, with the cGAN maintaining both higher POD and lower FA Ratio throughout Germany.

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


#### 3.2.3 Scale and threshold dependent precipitation estimation skill

The FSS quantifies skill over spatial neighborhoods rather than at individual grid points, making it less sensitive to small positional offsets between predicted and observed fields. We evaluate FSS at three intensity thresholds (0.1, 1, and $10\,\mathrm{mm\,h^{-1}}$) and a range of spatial scales (Fig. 4 a). The wet-grid-point fraction in the evaluation dataset is approximately 7 %-9 % (Table A1), giving $\mathrm{FSS}_{\mathrm{uniform}} \approx 0.54$ (Eq. 2). Since the percentage is smaller at higher thresholds, $\mathrm{FSS}_{\mathrm{uniform}}$ decreases accordingly. We therefore adopt $\mathrm{FSS} > 0.5$ as a skillfulness criterion throughout.

As expected, FSS increases with neighborhood size and decreases with threshold for both models. At 0.1 $\mathrm{mm\,h^{-1}}$ the cGAN achieves an FSS of 0.55–0.85 across all evaluated scales, while PDIR-now reaches $\mathrm{FSS} = 0.5$ only at a neighbourhood of 9 grid points (~30 km in east/west, ~56 km in north/south). At $1\,\mathrm{mm\,h^{-1}}$ the cGAN FSS remains above the values PDIR-now achieves at both 0.1 $\mathrm{mm\,h^{-1}}$ and 1 $\mathrm{mm\,h^{-1}}$. At 10 $\mathrm{mm\,h^{-1}}$ both models show reduced skill, but the cGAN exceeds $\mathrm{FSS} = 0.5$ at approximately 16 grid points (~53 km in east/west, ~99 km in north/south) while PDIR-now does not reach this threshold at any evaluated scale.

#### 3.2.4 Spatial structures of precipitation maps

The RAPSD evaluates how well each model reproduces the spatial variance of precipitation across a range of scales. Figure 4 b compares the mean RAPSD of the cGAN, PDIR-now, and RADKLIM-YW. Below wavelengths of 30 grid points, the RAPSD of PDIR-now falls substantially below that of RADKLIM-YW in both magnitude and slope, indicating a deficit of small-scale spatial variance, consistent with the qualitatively smoother appearance of PDIR-now fields in the example scenes (Fig. 2). The cGAN RAPSD remains closer to the RADKLIM-YW reference across this range.

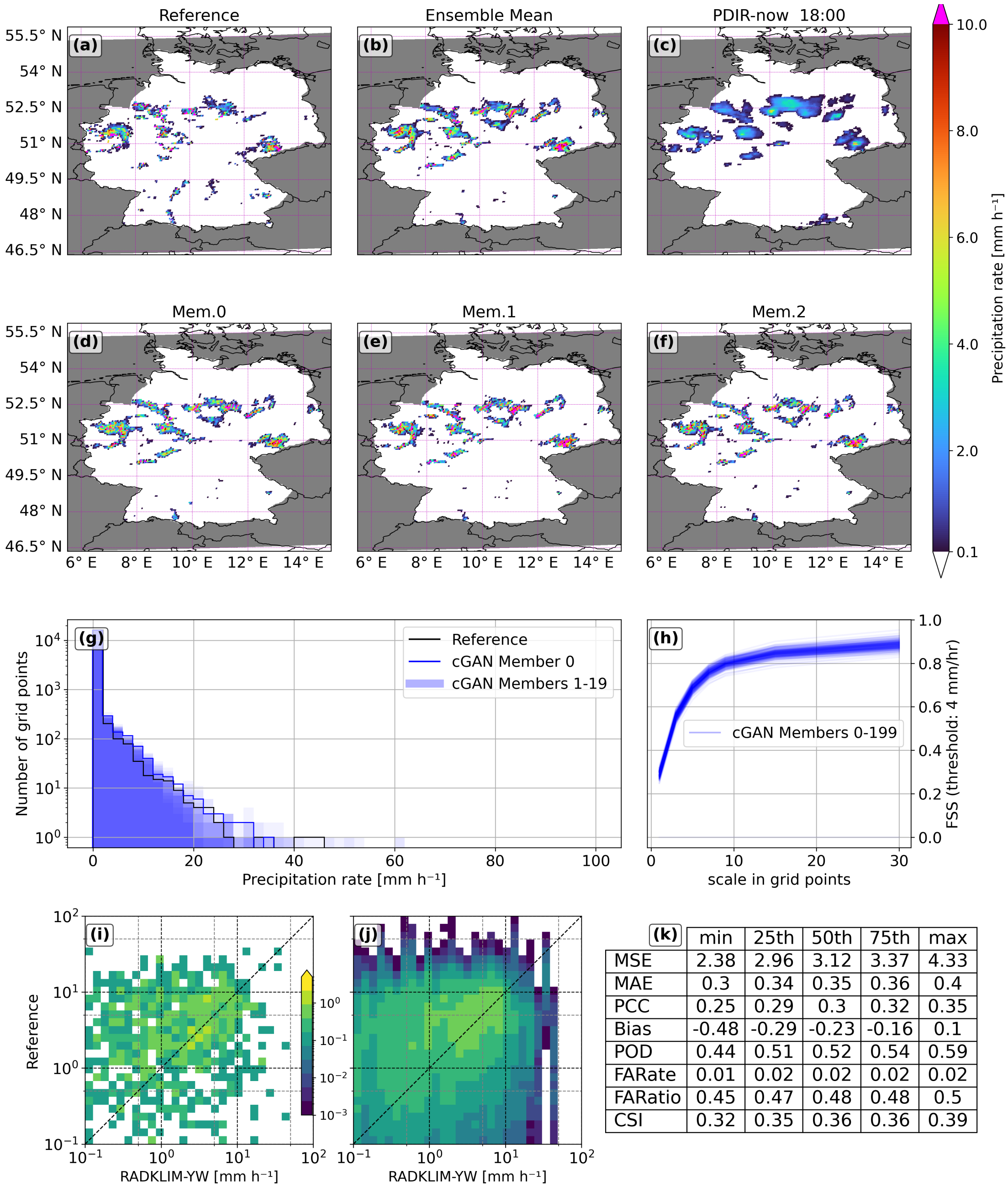
### 3.3 Ensemble Predictions

We evaluate a 200-member cGAN ensemble at 15-minute temporal resolution. The spread across ensemble members visible in the precipitation rate histogram (Fig. 4 c) is examined here in detail through full-domain scene analysis and local spatiotemporal transects. Since PDIR-now does not provide probabilistic output, no quantitative ensemble comparison is made, though PDIR-now fields are included in selected figures for visual reference.

#### 3.3.1 Example Ensemble Predictions of Full-domain Scenes

We first examine 2 August 2019 18:40 UTC (Fig. 5), a scene featuring scattered convective cells across Germany and a comma echo, which is a distinctive radar signature of a mesoscale convective system, in the western part of the domain. The cGAN reproduces the comma echo structure and matches the position of other precipitation areas well, reflected in a high FSS. Ensemble spread in precipitation location is small: the three depicted members show convective cells of similar shape in similar positions. Spread in precipitation amount is larger; bias ranges from negative to positive across members, POD ranges from 0.44 to 0.59, and for rates above $12\,\mathrm{mm\,h^{-1}}$ the reference counts lie within the 20-member ensemble range. Below that

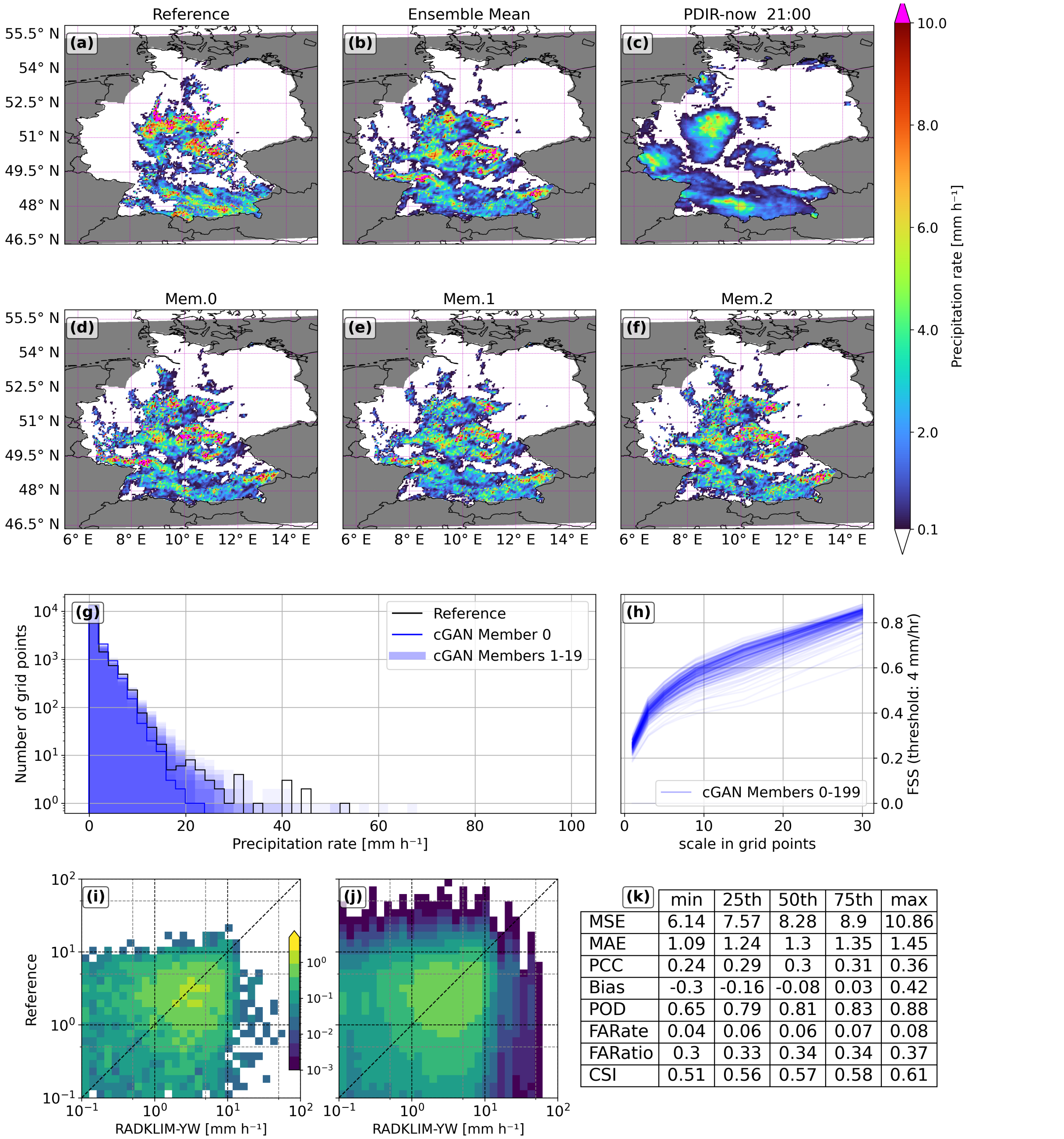
https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


| (k) | min | 25th | 50th | 75th | max |
|---|---|---|---|---|---|
| MSE | 2.38 | 2.96 | 3.12 | 3.37 | 4.33 |
| MAE | 0.3 | 0.34 | 0.35 | 0.36 | 0.4 |
| PCC | 0.25 | 0.29 | 0.3 | 0.32 | 0.35 |
| Bias | -0.48 | -0.29 | -0.23 | -0.16 | 0.1 |
| POD | 0.44 | 0.51 | 0.52 | 0.54 | 0.59 |
| FARate | 0.01 | 0.02 | 0.02 | 0.02 | 0.02 |
| FARatio | 0.45 | 0.47 | 0.48 | 0.48 | 0.5 |
| CSI | 0.32 | 0.35 | 0.36 | 0.36 | 0.39 |

**Figure 5.** Ensemble evaluation for 2 August 2019 18:40 UTC: precipitation rate maps for the reference (a), cGAN ensemble mean (b), PDIR-now (18-19 UTC) (c), and three individual members (d-f), precipitation rate histogram (g), FSS across all 200 members (h), 2D histograms for a single member (i) and all 200 members combined (j), and a table of skill metrics (minimum, maximum, 25th, 50th and 75th percentile of metrics across the 200 member ensemble) (k)

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


| | min | 25th | 50th | 75th | max |
|---|---|---|---|---|---|
| MSE | 6.14 | 7.57 | 8.28 | 8.9 | 10.86 |
| MAE | 1.09 | 1.24 | 1.3 | 1.35 | 1.45 |
| PCC | 0.24 | 0.29 | 0.3 | 0.31 | 0.36 |
| Bias | -0.3 | -0.16 | -0.08 | 0.03 | 0.42 |
| POD | 0.65 | 0.79 | 0.81 | 0.83 | 0.88 |
| FARate | 0.04 | 0.06 | 0.06 | 0.07 | 0.08 |
| FARatio | 0.3 | 0.33 | 0.34 | 0.34 | 0.37 |
| CSI | 0.51 | 0.56 | 0.57 | 0.58 | 0.61 |

**Figure 6.** Same as Figure 5, but for 20 May 2019 21:25 UTC (PDIR-Now 21-22 UTC.

threshold all members overestimate counts. For this scene the cGAN confidently locates precipitation but shows limited spatial diversity across members.

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


The second example, 20 May 2019 21:25 UTC Fig. 6), features a different meteorological regime with a large area of light-to-moderate precipitation and embedded convective cells. Both the cGAN and PDIR-now (shown for visual comparison) produce a region of false-positive precipitation in the western part of the domain, indicating a shared limitation of GEO-based QPE for this scene. Ensemble spread is greater than in the August case and the FSS is lower.

All three depicted members reproduce the false-positive western region, but differ in the location and intensity of precipitation within the larger contiguous fields. Member 1 shows a notably smaller heavy-precipitation area than members 0 and 2. The ensemble mean is spatially smoother than any individual member, reflecting positional offsets between members. In the histogram, the reference counts lie within the ensemble range for most precipitation bins. In the 2D histogram upper-right quadrant though member 0 alone underestimates the highest reference rates (cGAN $5\,\mathrm{mm\,h^{-1}}$ or below where the reference exceeds this). Considering 200 members, some predicted rates reach or exceed the reference maximum. Skill metrics vary across the ensemble (Fig. 6 table): MAE ranges from 0.40 to 0.53, bias changes sign between members, and at a scale of 9 grid points and threshold of $4\,\mathrm{mm\,h^{-1}}$ the majority of members exceed $\mathrm{FSS} = 0.5$, though some do not.

### 3.3.2 Spatiotemporal Development of Ensemble Predictions of local precipitation events

The preceding sections analyzed ensemble predictions over the full domain. Here we focus on the local scale to assess how the ensemble captures the spatiotemporal evolution of individual rainfall events.

Figure 7 zooms in on one convective cell from the larger map (Fig. 5). The cGAN reproduces the comma-echo structure, with members showing small positional and timing offsets. A spatial transect through the head of the comma echo shows that members place the precipitation core slightly east or west of the radar-observed position; all members underestimate the transect maximum. The time series at a fixed point shows ensemble precipitation arriving approximately $15\,\mathrm{min}$ later than in RADKLIM-YW. By 19:10 UTC the radar reference falls within the ensemble range. Ensemble spread in precipitation amount is notable, while spread in location and timing remains small.

Figure 8 examines a line-precipitation event on the afternoon of 26 April 2019. A northwest–southeast-oriented precipitation band near the North Sea moves eastward.

At the selected latitude transect all members locate the maximum at or adjacent to the radar-observed position. Most members underestimate the rainfall amount, only the highest predicted value reaches the reference precipitation rate. At the selected location most members predict heavy precipitation $15\,\mathrm{min}$ to late.

Across both cases, ensemble spread in precipitation amount exceeds spread in location and timing.

## 4 Discussion

### 4.1 Adversarial Training, Precipitation Realism, and Spatial Artifacts

PDIR-Now is a physics-informed operational precipitation retrieval that combines multispectral satellite observations with calibrated empirical relationships. It is, thus, a well-established baseline against which a purely data-driven model might not

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026

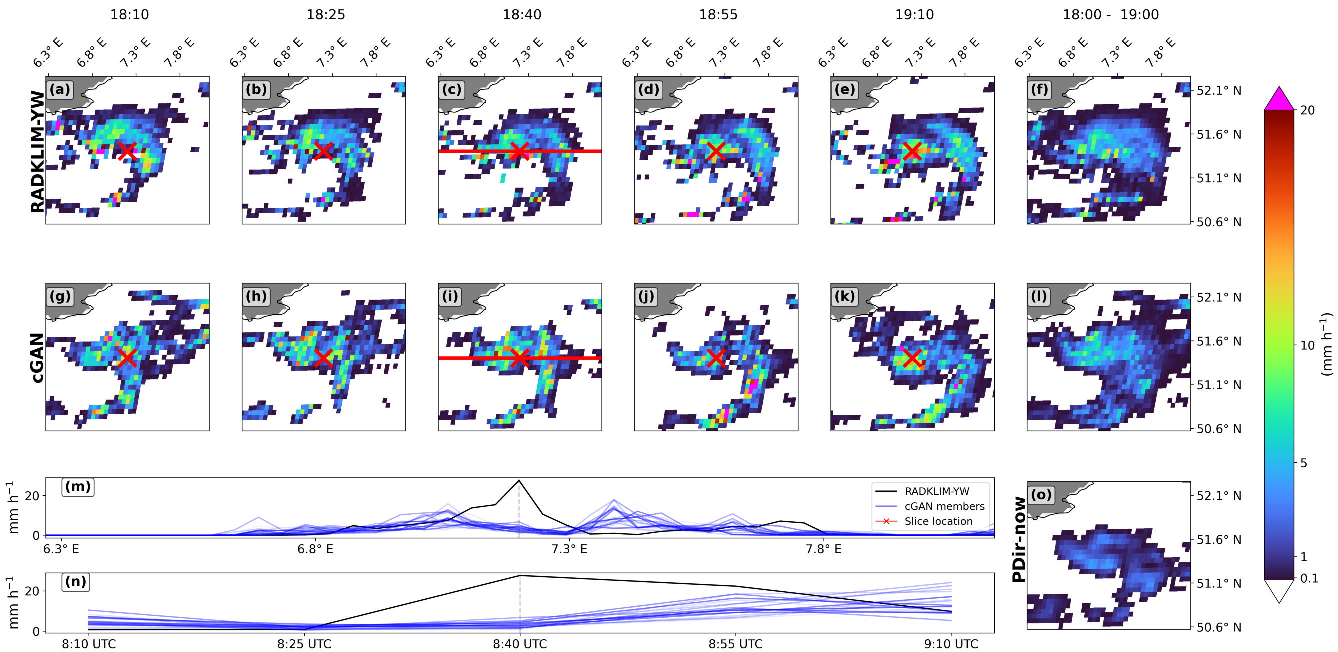




**Figure 7.** Spatiotemporal analysis of an event on 2 August 2019 in the west of Germany; RADKLIM-YW precipitation at five 15 min time steps and the hourly accumulation 18-19 UTC (a-f); corresponding cGAN ensemble member 0 predictions and hourly accumulation (g-l); the PDIR-Now hourly accumulation (o); spatial transect along the latitude of the marked cross at 18:40 UTC (m) and time series at the marked point over the full hour (n); black lines show RADKLIM-YW, blue lines show 20 individual cGAN ensemble members.

be expected to compete. The cGAN nonetheless outperforms it consistently across spatial structure, intensity distribution, and location skill metrics. These improvements can be attributed to adversarial training. A generator trained solely with pixel-wise loss functions such as MSE tends to produce spatially smooth outputs because minimizing average error is achieved by predicting the conditional mean of the training distribution (Isola et al., 2017; Leinonen et al., 2021). The discriminator counteracts this by penalizing outputs that lack the small-scale variability characteristic of real precipitation fields, driving the generator to recover spectral energy at fine spatial scales and to improve location skill. This mechanism explains the closer agreement with the observed power spectral density (RAPSD) and the improved reproduction of the intensity distribution, including high precipitation rates that PDIR-Now consistently underestimates.

Log-normalization of the training targets is used to encourage the model to generate more extreme precipitation rates, but the same transformation amplifies prediction errors upon back-transformation: a fixed error in log space corresponds to an exponentially larger deviation in physical units at high intensities leading to spurious large values.

Grid artifacts in accumulated precipitation fields are a known side-effect of generative adversarial training. Their severity is partly mitigated by using resize-convolution upsampling in the decoder rather than transposed convolution, which is prone to producing checkerboard patterns through uneven gradient overlap (Odena et al., 2016). The artifacts are not fully eliminated,

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026

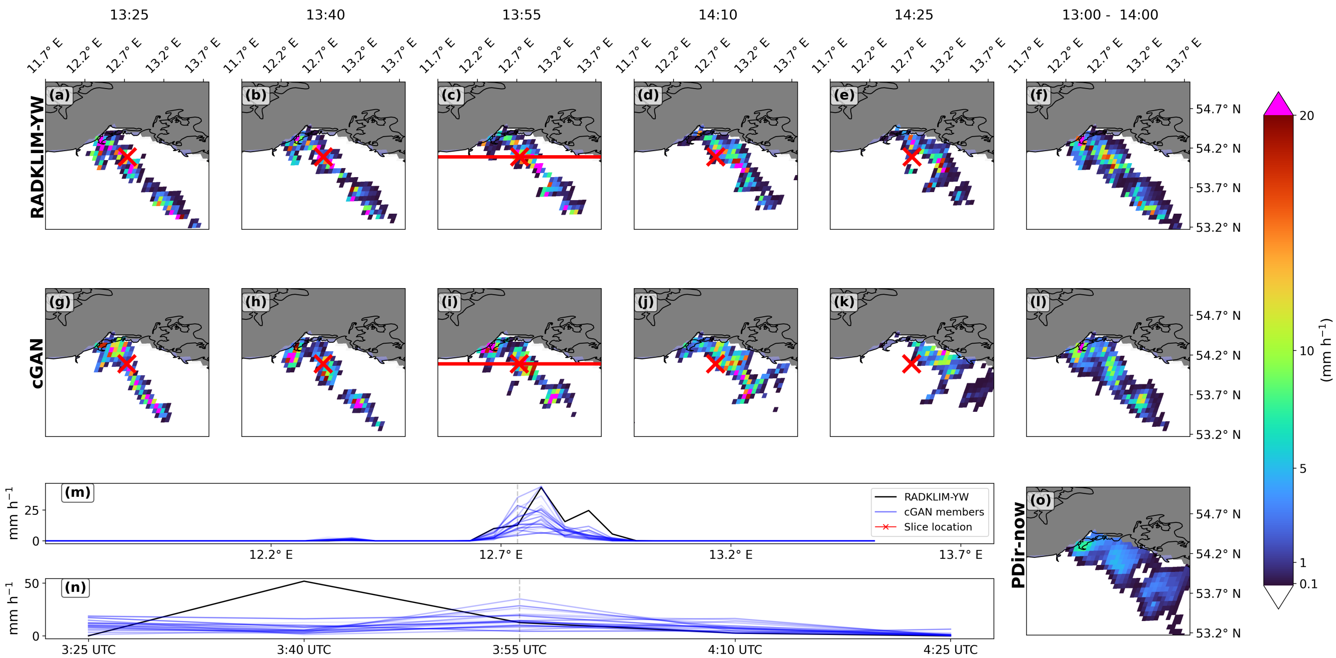




**Figure 8.** Same as for Fig. 7 but for an event on the 26 April 2019 in northern Germany

however, and further suppression attempts degraded overall predictive skill more than retaining them. They are therefore accepted as a known limitation, but only relevant to applications that accumulate predictions over multiple time steps. PDIR-Now shows artifacts in the accumulated precipitation in the accumulated maps in the form of larger-scale and slightly curved stripes. Based on their geometry they could maybe stem from the scan pattern of SEVIRI in combination with a regridding used during the production of PDIR-Now.

### 4.2 Spatiotemporal Consistency

Spatial consistency of cGAN predictions arises partly by construction. The generator produces complete two-dimensional precipitation maps in a single forward pass, with all output pixels conditioned on the same SEVIRI input fields. The convolutional kernels enforce spatial dependencies across the output, implicitly discouraging physically disconnected structures at neighboring pixels (see case studies). It is also enforced by the discriminator which sees the full spatial context, but does not judge temporal consistency.

Temporal consistency is encouraged by consecutive samples sharing one of the two SEVIRI time steps as input, so the atmospheric state represented in adjacent forecasts overlaps substantially. This constrains the temporal evolution even without an explicit recurrent architecture. Our results show that the 15-min brightness temperature differences encode a useful cloud development signal that the model demonstrably exploits. Further extending the two-time-step input window could improve performance and reduce phase offsets of the kind observed in the transect time series of both case studies. In this application,

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


however, a larger temporal context is especially expensive: the high spatial resolution and large domain size mean that each additional time step substantially increases memory and compute requirements, making T=2 a practical compromise for single GPU inference capabilities.

The observational geometry induces irreducible positional offsets between cGAN predictions and RADKLIM-YW: SEVIRI measures upwelling radiance near cloud top, while RADKLIM-YW reports near-surface precipitation. The horizontal displacement between the cloud-top signal and the surface precipitation footprint introduces positional uncertainty independent of model skill.

The existing GEO-based probabilistic QPE approach Hydronn (Pfreundschuh et al., 2022)) produces per-pixel precipitation probability distributions but does not generate an ensemble of spatiotemporally consistent precipitation rate maps. In contrast, our cGAN framework produces an ensemble of spatially coherent precipitation fields in a single forward pass and exhibits emergent temporal consistency through the overlapping two-step SEVIRI conditioning.

### 4.3 Ensemble Spread and Underdispersion

That MC-Dropout generates meaningfully diverse ensemble members without a dedicated ensemble training objective is a notable property of the architecture. Spread in precipitation amount is sufficient to bracket the radar reference in many cases.

Spread in precipitation location is, however, consistently too small across scenes and members, and the overall ensemble is underdispersive. This reflects two separable aspects of the training design. First, the MSE and adversarial losses optimize a single deterministic prediction rather than the diversity of a predicted distribution, so there is no explicit incentive for members to explore different plausible precipitation positions. Second, the SEVIRI input provides a strong deterministic conditioning signal that persists across members despite varying dropout masks, constraining the range of outputs the generator explores.

The contrast between the August and May scenes offers evidence for the latter mechanism. The sharp convective SEVIRI signature in the August scene is associated with low ensemble spread and high location skill. The more diffuse large-scale pattern in the May scene produces greater spread across members. If SEVIRI conditioning is the dominant source of underdispersion, weaker atmospheric signals should yield larger spread, which is consistent with this observation.

The two regimes of underdispersion carry different operational implications: amount uncertainty is directly usable in risk-based decision frameworks, whereas the near-deterministic location prediction adds limited probabilistic value beyond the ensemble mean.

## 5 Conclusions

Geostationary satellites offer continuous, global-scale observation of the atmosphere, making them well suited to real-time precipitation monitoring in regions where radar and rain-gauge networks are absent or sparse. Translating their cloud-top radiance measurements into accurate surface precipitation remains challenging due to the complex and non-linear relationship between the two quantities. We find that a conditional GAN trained on SEVIRI observations outperforms PDIR-Now, a physics-informed operational retrieval, across spatial structure, intensity distribution, and location skill metrics.

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


We show that adversarial training recovers fine-scale spectral energy that pixel-wise losses suppress, producing precipitation fields with improved spatial structure and intensity representation, including at high precipitation rates. Spatiotemporal consistency follows from the convolutional architecture and the overlapping SEVIRI input between consecutive predictions. Remaining limitations include grid artifacts in accumulated fields, missing spatial variance of ensemble members and spurious amplified errors at the extreme intensity tail.

For our methodological developments and analyses, we decided to train and evaluate the model only on the Germany domain, which provides a long and consistent radar reference dataset, minimizing the impact of heterogeneity of reference data. For generalization across the full MSG disk the viewing geometry of SEVIRI and the diversity of precipitation regimes represent sources of uncertainty . Obtaining consistent precipitation reference data across multiple countries and radar networks remains challenging, though. Thus, a relevant future direction of work will be extending cGAN-based QPE across multiple radar networks and GEO domains to allow application for real-time data from the geostationary ring with the GOES, Meteosat, and Himawari satellites.

*Code and data availability.* The datasets used as input, reference and benchmark are all publicly available: SEVIRI Level 1.5 Image Data (Aminou et al., 1997) at the EUMETSAT data store, PDIR-Now (Nguyen et al., 2020a) at the UCI CHRS Data Portal and RADKLIM-YW at the german weather services (DWD) open data portal (Winterrath et al., 2018). The trained cGAN, alongside some data samples and code is available at https://doi.org/10.5281/zenodo.20624016.

*Video supplement.* A supplementary video showing an example scene is available at https://doi.org/10.5446/73585.

## Appendix A: Hourly precipitation rates: percentiles

Table A1 shows hourly precipitation rates for certain percentiles of wet grid points and the percentage/total of wet grid points.

## Appendix B: wet - dry binary metrics

These metrics consider only wether a prediction and reference are wet of dry. They can all be calculated from the true positives (TP; both predictions and reference are wet), false positives (FP: prediction is wet while reference is dry), true negatives (TN: both predictions and reference are dry) or false negatives (FN: prediction is dry while reference is wet).

The probability of detection (POD, Eq. B1) measures the fraction of precipitation events that are correctly identified.

$$POD = TP/(TP+FN) \tag{B1}$$

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


**Table A1.**

| | Reference | cGAN Member 0 | PDIR-Now |
|---|---|---|---|
| 50 PCTL | 0.5101 | 0.6082 | 0.5491 |
| 75 PCTL | 1.3013 | 1.3994 | 1.1659 |
| 90 PCTL | 2.7719 | 2.6829 | 2.2517 |
| 99 PCTL | 8.3562 | 7.3467 | 6.5821 |
| 99.9 PCTL | 18.9348 | 15.9286 | 12.8637 |
| 99.99 PCTL | 33.2284 | 33.9981 | 19.7038 |
| 99.999 PCTL | 52.3572 | 68.9577 | 28.246 |
| 99.9999 PCTL | 68.3877 | 145.7594 | 32.9114 |
| Percentage of wet pixels | 7.0 | 8.22 | 8.72 |
| Wet pixel total | 5002206 | 5876605 | 6228104 |

Percentiles of wet ($>0.1\,\mathrm{mm\,h^{-1}}$) grid point timestep pairs in $\mathrm{mm\,h^{-1}}$, from the April–September 2019 evaluation period at the hourly resolution. The 99.9999th percentile corresponds to approximately 5 observations.

The false alarm ratio (FARatio, Eq. B2) is the the share of wrongly predicted precipitation among all predictions of precipitation.

$$FARatio = FP/(FP+TP) \tag{B2}$$

The similarly named false alarm rate (FARate, Eq. B3) is the share of wrongly predicted precipitation compared to the count of dry reference values.

$$FARate = FP/(FP+TN) \tag{B3}$$

The accuracy (Eq. B4) measures the share of correctly identified wet or dry predictions.

$$accuracy = (TP+TN)/(TP+TN+FP+FN) \tag{B4}$$

The bias ratio (BR, Eq. B5) refers to the ratio of the amount of wet predictions and wet reference values.

$$BR = (TP+FP)/(TP+FN) \tag{B5}$$

The critical success index (CSI, Eq. B6) compared correctly identified precipitation (TP) to the amount of wet predictions and wet reference values, a perfect score of 1 indicates no FP or FN predictions.

$$CSI = TP/(TP+FP+FN) \tag{B6}$$

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


## Appendix C: other metrics

The Pearson correlation coefficient (PCC, Eq. C1) measures the linear relationship between predictions $p$ and reference values $r$. A value of 1 indicates perfect positive correlation, 0 no correlation and -1 perfect negative correlation.

$$PCC = \frac{\sum_{i=1}^{N}(p_i - \bar{p})(r_i - \bar{r})}{\sqrt{\sum_{i=1}^{N}(p_i - \bar{p})^2}\sqrt{\sum_{i=1}^{N}(r_i - \bar{r})^2}} \tag{C1}$$

The mean absolute error (MAE Eq. C2) measures the average absolute difference between predictions and reference values.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|p_i - r_i| \tag{C2}$$

The mean squared error (MSE, Eq. C3) measures the average squared difference between predictions and reference values, giving more weight to large errors.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(p_i - r_i)^2 \tag{C3}$$

The bias (Eq. C4) measures the relative mean difference between predictions and reference values and indicates systematic over- or underestimation.

$$Bias = \frac{\sum_{i=1}^{N} r_i - \sum_{i=1}^{N} p_i}{\sum_{i=1}^{N} r_i} \tag{C4}$$

*Author contributions.* SJ: Methodology, Data curation, Software, Visualization, Manuscript first draft/editing. JP: Methodology, Manuscript editing. LG: Methodology, Manuscript editing. HK: Supervision. CC: Funding acquisition, Supervision, Methodology, Manuscript editing

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Project 541482432 ("MERGOSAT").

https://doi.org/10.5194/egusphere-2026-4000
Preprint. Discussion started: 19 August 2026


## References

Aminou, D. M. A., Jacquet, B., and Pasternak, F.: Characteristics of the Meteosat Second Generation (MSG) Radiometer/Imager: SEVIRI, in: Sensors, Systems, and Next-Generation Satellites, vol. 3221 of *Sensors, Systems, and Next-Generation Satellites*, pp. 19–31, SPIE, London, UK, https://doi.org/10.1117/12.298084, 1997.

Antonio, B. and Aitchison, L.: How to Derive Skill from the Fractions Skill Score, Monthly Weather Review, 153, 1021–1033, https://doi.org/10.1175/MWR-D-24-0120.1, 2025.

Ayzel, G., Scheffer, T., and Heistermann, M.: RainNet v1.0: a convolutional neural network for radar-based precipitation nowcasting, Geoscientific Model Development, 13, 2631–2644, https://doi.org/10.5194/gmd-13-2631-2020, 2020.

Bihlo, A.: A Generative Adversarial Network Approach to (Ensemble) Weather Prediction, Neural Networks, 139, 1–16, https://doi.org/10.1016/j.neunet.2021.02.003, 2021.

Brecht, R. and Bihlo, A.: Computing the Ensemble Spread From Deterministic Weather Predictions Using Conditional Generative Adversarial Networks, Geophysical Research Letters, 50, e2022GL101 452, https://doi.org/10.1029/2022GL101452, 2023.

Gal, Y. and Ghahramani, Z.: Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, in: Proceedings of The 33rd International Conference on Machine Learning, edited by Balcan, M. F. and Weinberger, K. Q., vol. 48 of *Proceedings of Machine Learning Research*, pp. 1050–1059, PMLR, New York, New York, USA, https://proceedings.mlr.press/v48/gal16.html, 2016.

Glawion, L., Polz, J., Kunstmann, H., Fersch, B., and Chwala, C.: Global spatio-temporal ERA5 precipitation downscaling to km and sub-hourly scale using generative AI, npj Climate and Atmospheric Science, 8, https://doi.org/10.1038/s41612-025-01103-y, 2025.

Guilloteau, C., Kerrigan, G., Nelson, K., Migliorini, G., Smyth, P., Li, R., and Foufoula-Georgiou, E.: A Generative Diffusion Model for Probabilistic Ensembles of Precipitation Maps Conditioned on Multisensor Satellite Observations, IEEE Transactions on Geoscience and Remote Sensing, 63, 1–15, https://doi.org/10.1109/TGRS.2025.3548518, 2025.

Hauser, M., Spring, A., Busecke, J., van Driel, M., Lorenz, R., and readthedocs assistant: regionmask/regionmask: version 0.12.1, https://doi.org/10.5281/zenodo.10849860, 2024.

Hayatbini, N., Kong, B., Hsu, K.-l., Nguyen, P., Sorooshian, S., Stephens, G., Fowlkes, C., Nemani, R., and Ganguly, S.: Conditional Generative Adversarial Networks (cGANs) for Near Real-Time Precipitation Estimation from Multispectral GOES-16 Satellite Imageries—PERSIANN-cGAN, Remote Sensing, 11, https://doi.org/10.3390/rs11192193, 2019.

Hong, Y., Hsu, K.-L., Sorooshian, S., and Gao, X.: Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System, Journal of Applied Meteorology and Climatology, 43, 1834–1853, https://doi.org/10.1175/JAM2173.1, 2004.

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T.: The Global Precipitation Measurement Mission, Bulletin of the American Meteorological Society, 95, 701–722, https://doi.org/10.1175/BAMS-D-13-00164.1, 2014.

Hu, J., Shen, L., and Sun, G.: Squeeze-and-Excitation Networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018), pp. 7132–7141, Salt Lake City, UT, USA, https://doi.org/10.1109/CVPR.2018.00745, 2018.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967–5976, IEEE, Honolulu, HI, USA, ISBN 978-1-5386-0457-1, https://doi.org/10.1109/CVPR.2017.632, 2017.

Kühnlein, M., Appelhans, T., Thies, B., and Nauss, T.: Improving the Accuracy of Rainfall Rates from Optical Satellite Sensors with Machine Learning — A Random Forests-Based Approach Applied to MSG SEVIRI, Remote Sensing of Environment, 141, 129–143, https://doi.org/10.1016/j.rse.2013.10.026, 2014.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P.: GraphCast: Learning skillful medium-range global weather forecasting, https://arxiv.org/abs/2212.12794, 2023.

Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., Raoult, B., Clare, M. C. A., Lessig, C., Maier-Gerber, M., Magnusson, L., Bouallègue, Z. B., Nemesio, A. P., Dueben, P. D., Brown, A., Pappenberger, F., and Rabier, F.: AIFS – ECMWF's data-driven forecasting system, https://arxiv.org/abs/2406.01465, 2024.

Leinonen, J., Nerini, D., and Berne, A.: Stochastic Super-Resolution for Downscaling Time-Evolving Atmospheric Fields With a Generative Adversarial Network, IEEE Transactions on Geoscience and Remote Sensing, 59, 7211–7223, https://doi.org/10.1109/TGRS.2020.3032790, 2021.

Levizzani, V. and Kidd, C.: Status of satellite precipitation retrievals, Hydrology and Earth System Sciences, 15, 1109–1116, https://doi.org/10.5194/hess-15-1109-2011, 2011.

Liu, M., Zuo, J., Tan, J., and Liu, D.: Comparing and Optimizing Four Machine Learning Approaches to Radar-Based Quantitative Precipitation Estimation, Remote Sensing, 16, https://doi.org/10.3390/rs16244713, 2024.

Mirza, M. and Osindero, S.: Conditional Generative Adversarial Nets, https://arxiv.org/abs/1411.1784, 2014.

Nguyen, H.-G., Pica, A., Maeder, P., Schalenbourg, A., Peroni, M., Hrbacek, J., Weber, D. C., Cuadra, M. B., and Sznitman, R.: Ocular Structures Segmentation from Multi-sequences MRI Using 3D Unet with Fully Connected CRFs, in: Computational Pathology and Ophthalmic Medical Image Analysis, edited by Stoyanov, D., Taylor, Z., Ciompi, F., Xu, Y., Martel, A., Maier-Hein, L., Rajpoot, N., van der Laak, J., Veta, M., McKenna, S., Snead, D., Trucco, E., Garvin, M. K., Chen, X. J., and Bogunovic, H., pp. 167–175, Springer International Publishing, Cham, Germany, ISBN 978-3-030-00949-6, 2018a.

Nguyen, P., Ombadi, M., Sorooshian, S., Hsu, K., AghaKouchak, A., Braithwaite, D., Ashouri, H., and Thorstensen, A. R.: The PERSIANN Family of Global Satellite Precipitation Data: A Review and Evaluation of Products, Hydrology and Earth System Sciences, 22, 5801–5816, https://doi.org/10.5194/hess-22-5801-2018, 2018b.

Nguyen, P., Ombadi, M., Gorooh, V. A., Shearer, E. J., Sadeghi, M., Sorooshian, S., Hsu, K., Bolvin, D., and Ralph, M. F.: PERSIANN Dynamic Infrared–Rain Rate (PDIR-Now): A Near-Real-Time, Quasi-Global Satellite Precipitation Dataset, Journal of Hydrometeorology, 21, 2893–2906, https://doi.org/10.1175/JHM-D-20-0177.1, 2020a.

Nguyen, P., Shearer, E. J., Ombadi, M., Gorooh, V. A., Hsu, K., Sorooshian, S., Logan, W. S., and Ralph, M.: PERSIANN Dynamic Infrared–Rain Rate Model (PDIR) for High-Resolution, Real-Time Satellite Precipitation Estimation, Bulletin of the American Meteorological Society, 101, E286–E302, https://doi.org/10.1175/BAMS-D-19-0118.1, 2020b.

Odena, A., Dumoulin, V., and Olah, C.: Deconvolution and Checkerboard Artifacts, Distill, 1, https://doi.org/10.23915/distill.00003, 2016.

Pfreundschuh, S., Ingemarsson, I., Eriksson, P., Vila, D. A., and Calheiros, A. J. P.: An Improved Near-Real-Time Precipitation Retrieval for Brazil, Atmospheric Measurement Techniques, 15, 6907–6933, https://doi.org/10.5194/amt-15-6907-2022, 2022.

Pulkkinen, S., Nerini, D., Pérez Hortal, A. A., Velasco-Forero, C., Seed, A., Germann, U., and Foresti, L.: Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1.0), Geoscientific Model Development, 12, 4185–4219, https://doi.org/10.5194/gmd-12-4185-2019, 2019.

Roberts, N. M. and Lean, H. W.: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events, Monthly Weather Review, 136, 78–97, https://doi.org/10.1175/2007MWR2123.1, 2008.

Roebeling, R. A. and Holleman, I.: SEVIRI Rainfall Retrieval and Validation Using Weather Radar Observations, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/10.1029/2009JD012102, 2009.

Sadeghi, M., Asanjan, A. A., Faridzad, M., Nguyen, P., Hsu, K., Sorooshian, S., and Braithwaite, D.: PERSIANN-CNN: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks–Convolutional Neural Networks, Journal of Hydrometeorology, 20, 2273–2289, https://doi.org/10.1175/JHM-D-19-0110.1, 2019.

Ward-Leikis, B., Rampal, N., Koh, Y. S., Gibson, P. B., Liu, H.-Y., Kitsios, V., Meyers, T., Adie, J., Juntao, Y., and Sherwood, S. C.: An intercomparison of generative machine learning methods for downscaling precipitation at fine spatial scales, https://doi.org/10.48550/arXiv.2512.13987, 2025.

Winterrath, T., Brendel, C., Hafer, M., Junghänel, T., Klameth, A., Lengfeld, K., Walawender, E., Weigl, E., and Becker, A.: Radar Climatology (RADKLIM) Version 2017.002; Gridded Precipitation Data for Germany: Radar-based Quasi Gauge-Adjusted Five-Minute Precipitation Rate (YW), https://doi.org/10.5676/DWD/RADKLIM_YW_V2017.002, 2018.