Article

Coupling a Neural Network with a Spatial Downscaling Procedure to Improve Probabilistic Nowcast for Urban Rain Radars

Marino Marrocu *[ID] and Luca Massidda [ID]

CRS4, Center for Advanced Studies, Research and Development in Sardinia, Loc. Piscina Manna ed. 1,
09050 Pula, Italy; luca.massidda@crs4.it
* Correspondence: marino.marrocu@crs4.it

Abstract: Rainfall forecasting plays a key role in mitigating environmental risks in urban areas, which are subject to increasing hydrogeological risk due to transformations in the urban landscape. We present a new technique for probabilistic precipitation nowcasting by generating an ensemble of equiprobable forecasts, which is especially useful for weather radars with limited spatial range, and that can be used operationally on devices with low computational capacity. The ensemble members are obtained by a novel stochastic noise generation process, consistent with the spatial scales not resolved by the prediction model, which allows continuous downscaling of the output of a deep generative neural network. Through an in-depth analysis of the results, for precipitation accumulated in the first hour, measured by all the most robust skill indicators, extended to an entire year of data at 5-min time resolution, we demonstrate that the proposed procedure is able to provide calibrated, reliable, and sharp ensemble rainfall forecasts with a quality comparable or superior to the state-of-the-art classical alternative optical flow technique. The ensemble generation procedure we propose is sufficiently general to be applied in principle to other deterministic architectures as well, thus enabling their generalization in probabilistic terms.

Keywords: probabilistic nowcast; meteorological radar data; deep learning; generative neural network; optical flow; spatial downscaling

1. Introduction

Precipitation nowcasting, following the definition of the World Meteorological Organization (WMO), is a high-resolution forecasting technique up to two hours in the future that is essential for weather-dependent decision-making processes in many sectors, such as emergency services, flood warning systems, air traffic control, and energy management to name a few [1].

Nowcasting systems based on the evolution of the estimated precipitation field by weather radar have been shown to be remarkably effective Quantitative Precipitation Nowcasting (QPN) tools. The most common classical nowcast methodologies are based on the concept of optical flow (OF) and make use of a sequence of images prior to the issue of the forecast to advect precipitation fields with radar-based wind estimates [2–4]. These methods suffer from increasing uncertainty as the forecast time increases, due to the fact that they lack a physical description of thunderstorm growth and decay, which is considered to be the major deficiency of such nowcasting systems [5]. One approach to mitigate this drawback is the fusion of Numerical Weather Prediction (NWP) data with radar data [6] or even the assimilation of radar data into NWP models [7]. A different approach for characterizing the uncertainties in the evolution of nowcasting has been introduced with *STEPS* [8], where the S-PROG scheme [9] produces a cascade of forecasts, the small-scale features of which are randomly generated consistently with the uncertainty expressed by the nowcasting model in the period preceding the forecast. *STEPS* has proved to generate

skillful rainfall forecast in a variety of rainfall conditions [10] and in addition to being currently used operationally has also become a de facto benchmark for any data-driven nowcast procedure and a study and development platform for the science community [11].

The potential offered by the highly nonlinear neural network approach to the problem of rainfall radar nowcasting, starting from the pioneering studies of the Hong-Kong Meteorological Observatory group [12,13], has also been the subject of widespread interest in recent years [14–18]. In [15], neural networks have been used to detect growth and decay patterns, achieving significant performance improvements in terms of RMSE compared with a forecast based on classical OF methods, but also highlighting the need for a probabilistic approach to quantify the uncertainty of the forecast. In [14], a U-Net type network architecture is proposed to implement a data driven rain radar nowcasting system, achieving good performance but highlighting the progressive alteration of the spectral characteristics of the solution in the forecast. In [16], the same U-Net architecture is enriched with attention modules obtaining good results and limiting the size of the model needed. In [19], rain radar images and wind forecasts are used in a deep learning model to transform the rainfall intensity regression problem into a classification problem in which the probability of exceeding a certain precipitation threshold is estimated. In [17], a deep architecture with a self attention mechanism is proposed for the realization of a probabilistic forecast, integrating ground-based radar information with satellite images and weather model forecasts obtaining a forecast over the entire territory of the US. Using significant computational resources, the probabilistic forecast is achieved by generating precipitation probabilities for different thresholds. An alternative approach for making probabilistic forecasts is proposed in [18], where the problem of loss of detail and spectra alteration is mitigated by using a generative type network to produce different scenarios from the learned distributions. More recently, in [20], by combining a discriminator and an attention module, the authors show that their NN architecture outperforms the current operational QPN with lead times till 3 h.

As also highlighted in [20], the forecasts issued by current deep learning systems express uncertainty with increasing timescales related to the increasing level of smoothness of the predicted precipitation fields. It is therefore not surprising that such forecasts may completely miss some small-scale weather phenomena that, as mentioned earlier, are critical to the operational value of the forecast itself [13]. Another common feature of all neural network-based methods is the huge computational resources required and the challenging replicability of the architecture solutions also because they are often not released in the public domain. OF solutions such as *STEPS* on the other hand exhibit a problem that cannot be eliminated when applied to small domains due to the advection of the information out of the domain as the prediction time increases.

The basic idea of this work is to verify the feasibility and effectiveness of an approach that, by overcoming these limitations, manages to fully exploit the unquestionable predictive potential demonstrated by the neural networks, and by the PredNet [21,22] in particular. The chosen operational model is based on the interpretation of the neural network prediction as an optimal estimate of the mean of an ensemble of nowcasts which spread in turn is representative of the associated uncertainty. The approach, which we believe can also be generalized to other nowcast systems based on deterministic neural networks other than PredNet, consists of injecting, during the extrapolation phase, stochastic noise at scales not explicitly resolved by the network.

This, as we shall see, makes it possible to generate an ensemble of forecasts that we verify are all in a good approximation equally likely, in an operational configuration similar to that of the *STEPS* procedure, which is thus the natural candidate to be used as a benchmark for testing its performance.

In the following, Section 2 presents data and methods used to obtain the results discussed in Section 3. Section 4 hosts the conclusions and a discussion of potential future developments.

2. Materials and Methods

This section describing Materials and Methods is organized as follows: the data used for simulation and verification are presented in Section 2.1, and Section 2.2 describes the neural network used in this analysis, while the procedure for its probabilistic exploitation and the *STEPS* methodology used as a benchmark are presented in Sections 2.3 and 2.4, respectively.

2.1. Data

Our motivation for undertaking this study was to test the feasibility of a short-range precipitation forecasting procedure that could be used to cover the nowcasting needs for a metropolitan area [23]. To this end, we used data measured from a meteorological radar located in the region of Osaka and Kyoto cities in Japan. These data, in the public domain and already used in [24], were collected from an online archive available on Yahoo (https://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/js/; the service is no more available from Wednesday, 6 April 2022), and cover three consecutive years, from 2015 to 2017. The spatial resolution is 1 km^2, the temporal frequency is 5′, and the domain, shown in Figure 1, covers an area of about 10^4 km^2. The climate of the area is classified as warm-temperate, with 1730 mm annual cumulation and particularly heavy rainfall during the summer season.
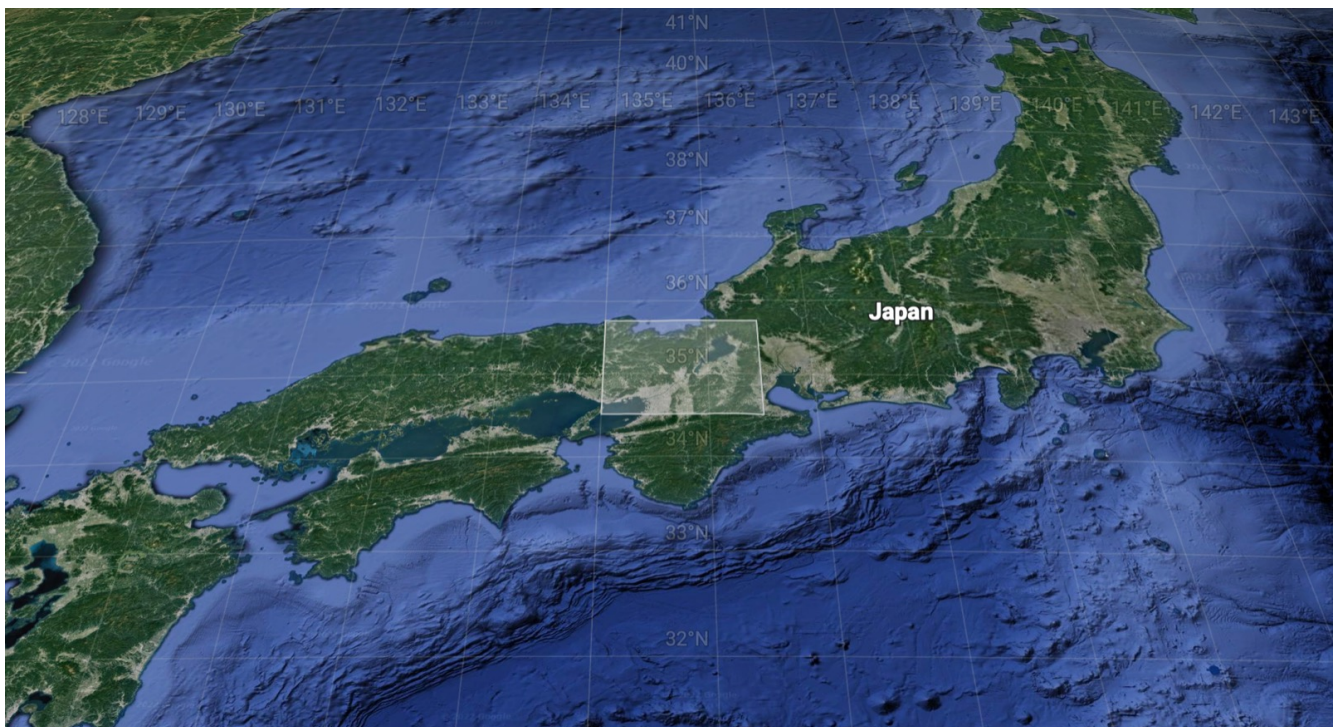

Figure 1. The map of Japan in which is highlighted the area that represents the domain covered by the weather radar whose data are used in this study.

The resolution of the radar images obtained through an API is higher than the actual resolution of the data, so an upscaling of the image is performed as the first operation that results in the creation of square matrices of 96 × 96 elements. The data are a false-color representation of precipitation intensity. For each pixel, colors can then be mapped as a progressive index of rainfall intensity ranging from 0 to 14 as in Table 1 to which increasing values of precipitation rate, from 0 mm/h up to 100 mm/h, correspond.

Table 1. Lookup table between the RGB colors of the radar image representation, precipitation intensity, and the numerical value assigned to the class.

Class	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Color (RGB)	255	236	196	156	156	180	180	180	164	252	252	252	252	252	228
	255	254	254	234	218	198	254	234	218	254	234	218	190	158	158
	255	252	252	252	252	252	156	156	156	156	156	156	196	156	164
Rainfall (mm/h)	0	1	2	4	8	12	16	24	32	40	48	56	64	80	100

The array containing the color index (class) is normalized to a maximum unit value for processing in the neural network and is converted back to the rainfall intensity value during post-processing. The use of rainfall intensity or its processing by a logarithmic transformation proved to be less effective than the direct processing of the color scale index. The data were also filtered by removing all weather events in which the radar images were not complete, either because they were missing or because portions of the image had been lost.

2.2. PredNet

The radar image nowcasting problem can be considered a supervised learning problem, in which the parameters of a neural network are optimized to minimize an objective function that depends on the differences between the estimated precipitation image and the actual field. Among various network architectures applied to the weather nowcasting problem, we opted for the PredNet model. This paper is the evolution of a previous study [22] in which we analyzed the performance of the PredNet architecture, highlighting its excellent performance in merely deterministic terms, but also its limitations and the need for a revisiting in probabilistic terms. As specified, the architecture is borrowed from computer vision, likewise the optical flow methodology, which represents the state-of-the-art architecture in the field. It was chosen because of the originality of the implemented approach and because of its generality. Other more recent approaches have introduced improvements that are not directly applicable to the case of precipitation nowcasting because they are specific to the prediction of video frames, assuming that the moving objects are immutable in shape, or introducing the concept of image depth and perspective, clearly not applicable to our application.

The PredNet architecture consists of a deep, recursive convolutional network with connections from large to small scale (top-down) and vice versa. The approach leverages the experience of similar models based on recursive and convolutional networks for predicting video frames. It also draws particular inspiration from the concept of 'predictive coding' in the neuroscience literature, which hypothesizes that the brain continuously makes predictions by anticipating incoming stimuli.

The PredNet architecture employed is shown sketched in Figure 2 and consists of a ladder structure with a descending branch that generates the prediction of the precipitation field image at progressively more detailed scales and an ascending branch in which the prediction is compared with actual data. Further descriptive details of the architecture can be found in [22] to which we refer the reader for brevity and in Lotter, Kreiman, and Cox's article [21].

The training is conducted with a Gradient Descent procedure in which the network parameters are modified to minimise e_0, i.e., the Mean Absolute Error (MAE) between the calculated and the real image. Since we are interested in predicting not only the next frame, the loss function used in the training phase is obtained as the average MAE error of the 24 images $h_0 = \hat{X}_t$ generated by the network with respect to the real precipitation images $a_0 = X_t$:

$$Loss = \sum_{t=0}^{24} MAE(X_t, \hat{X}_t) \qquad (1)$$

where *MAE* is an operator that returns the average absolute error between the real radar image and the one estimated by the model; the sum is extended to both the 12 instants used for learning the current weather situation and the 12 instants of the forecast, in order to train the network for both the reconstruction of the radar image and the rainfall forecast.
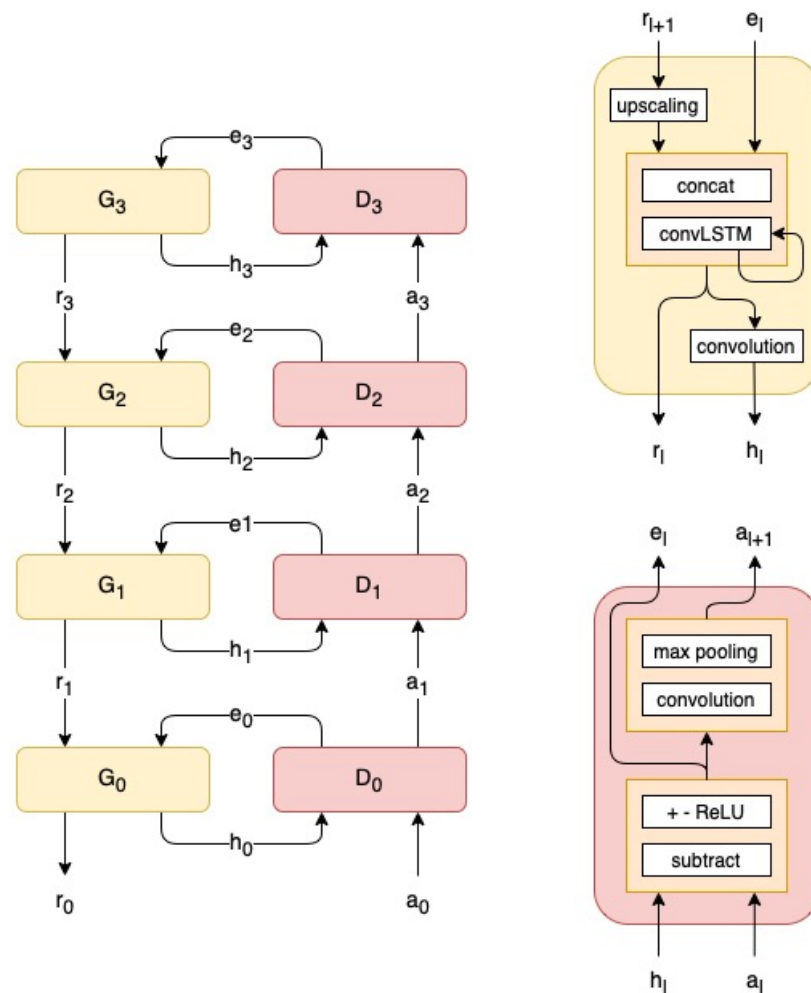

Figure 2. PredNet: on the left , the general architecture with the Discriminator and Generator blocks; on the right, the details of the two blocks.

The optimal weights of the neural network were obtained using a gradient descent procedure with the Adam optimiser. The neural network model and training procedures were implemented in Python [25] using the PyTorch library [26]. The calculation was performed on an NVIDIA Tesla K80 GPU with 12 GB of built-in RAM. The model will be hereafter referred to as *NN*.

The choice of the number of layers (4) and their dimension (see Table 2) is directly related to the size of the image to be processed (96, 96), and to the scale of weather phenomena that is reasonably conceivable to resolve given the field size and time horizon. The depth of the layers derives from a process of optimization of the achievable performance by limiting the size of the network so that it can be processed by a single GPU card, assuming a possible operational implementation in a small-scale city radar with reduced computational capabilities.

Table 2. PredNet tensors: size of the tensors for the different levels of the PredNet architecture expressed as (channels, width, height).

Level L	a_l	h_l	e_l	r_l
0	(1, 96, 96)	(1, 96, 96)	(2, 96, 96)	(1, 96, 96)
1	(16, 48, 48)	(16, 48, 48)	(32, 48, 48)	(16, 48, 48)
2	(32, 24, 24)	(32, 24, 24)	(64, 24, 24)	(32, 24, 24)
3	(64, 12, 12)	(64, 12, 12)	(128, 12, 12)	(64, 12, 12)

2.3. Proposed Methodology: Stochastic PredNet

In order to illustrate the procedure proposed in this paper, it is necessary to understand the way PredNet operates both in the phase preceding the emission of the forecast and in the prediction phase. During the hour preceding the forecast, PredNet 'digests' the 12 images at 5' intervals, starting from zero knowledge at instant zero and progressively reducing the differences with the measured input data, as illustrated in the Figure 3. This initial phase is similar to the data assimilation procedure that is part of the operational chain of a numerical weather service, and the duration of this pre-processing can be considered similar to the assimilation window. We will often refer to this phase in the following as 'assimilation'.
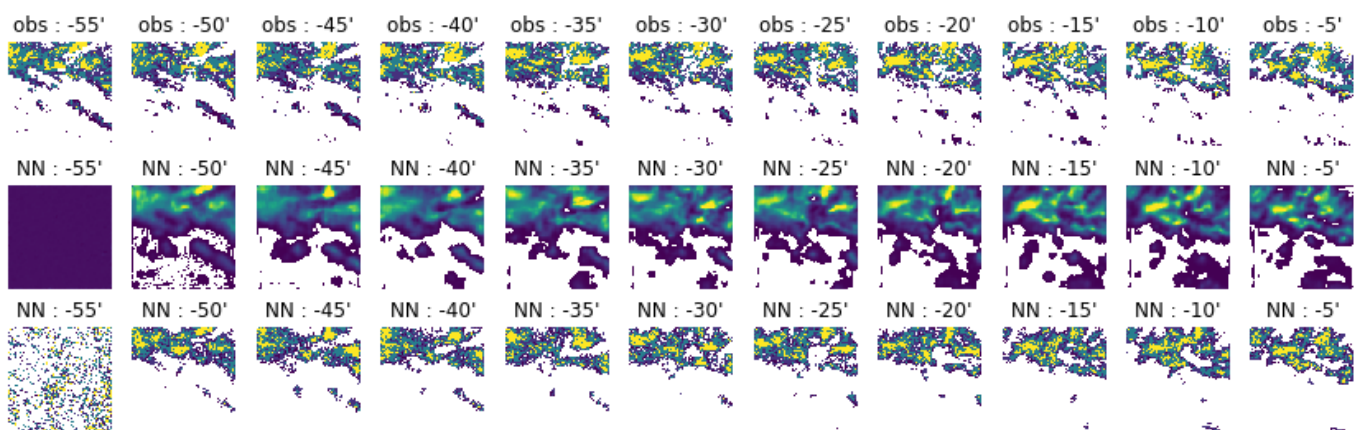

Figure 3. Sequence of 12 images at 5' intervals of the precipitation field during the hour preceding the nowcast. The first row shows the observed data (*obs*), the second row the estimated values by the PredNet (*NN*), and the third row one of the scenario fields obtained applying the spatial downscaling process described in the Section 2.3 to the corresponding fields of the second row.

A careful analysis of the Figure 3 through comparison of the measured data, shown in the first column, and that estimated by the PredNet in the hour before the issuing of the nowcast shows that the difference between the 'assimilated' observations and the observations is, at least visually, minimal. A deeper analysis shows that the substantial difference lies in the resolved scales. In fact, as Figure 4 shows, the resolved power spectrum shows a marked loss of power from scales of the order of 16 km. This behaviour is not peculiar to the example shown, but represents a systematic feature for all the events studied and is caused by the minimisation process of the neural network's optimisation function. This weakness allows us, nevertheless, to address the problem of the uncertainty in the assimilation phase. Since the internal representation of the network is unable to resolve scales below a certain limit, we can add to the spectral representation of the fields a stochastic component in this range of scales. This enables for obtaining a set of scenarios, all indistinguishable from the point of view of the scales resolved by the neural network, representing an ensemble of initial conditions for the subsequent prediction phase.

The procedure used to inject stochastic noise at scales not explicitly resolved by the neural network takes inspiration from the dowsncaling procedure called *rainfarm* [27].

In detail, let us denote by X_t and $\hat{X}_t$ the observed and reconstructed images at n instants preceding the nowcast emission, and by $\overline{|P(X_t)_K|^2}$ and $\overline{|P(\hat{X}_t)_K|^2}$ the power spectra, as a function of wave number K, averaged over these n events. We then calculate the associated Fourier spectrum, defined minus a random multiplicative component of unitary norm:

$$P(X_t)_K^{rand} = \sqrt{\overline{|P(X_t)_K|^2}} \cdot e^{\Phi_K} \tag{2}$$

where Φ_K is a uniformly distributed random phase. Applying the inverse Fourier transform, F_T^{-1}:

$$g(r) = F_T^{-1}(P(X_t)_K^{rand}) \tag{3}$$

we obtain a random field $g(r)$ defined in the spatial domain, which by construction is Gaussian distributed. After normalisation of its distribution, an exponential transform is applied:

$$H(r) = e^{g_N(r)} \tag{4}$$

to obtain a field that will therefore have a log-normal distribution. We can think about $H(r)$ as the realisation of a random precipitation field with the same spatial correlation as the field X_t from which it originates, since, in a weak sense and minus a random phase of unitary norm in the Fourier transform, it has the same power spectrum.
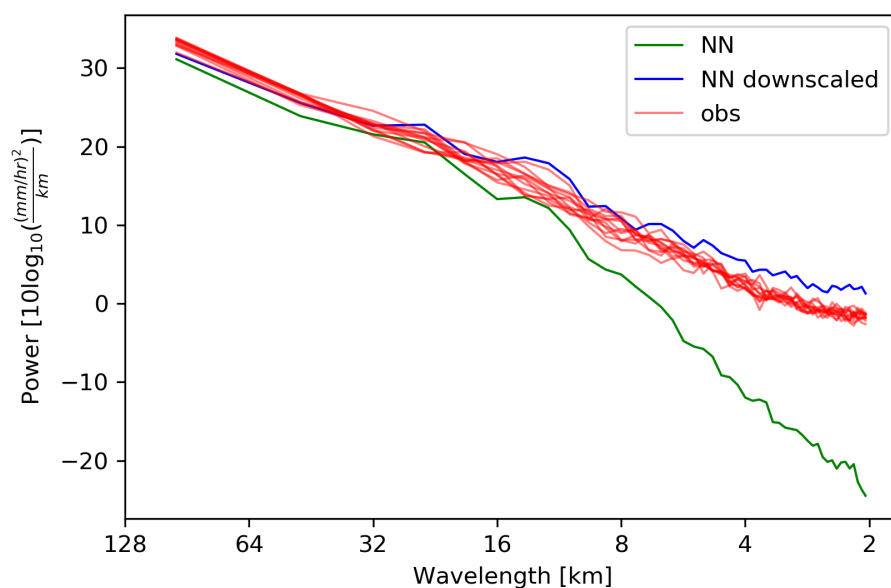

Figure 4. In red, the power spectrum for the 12 images at 5′ intervals of the precipitation field in the hour preceding the nowcast; in green, the power spectrum of the PredNet estimate for the image preceding the nowcast; in blue, the spectra of the field obtained by downscaling the *NN* output field using the procedure described in the Section 2.3.

Applying the same procedure starting from the average spectrum derived from the reconstructed images $\hat{X}_t$ and denoting by $R(r)$ the random field obtained by using the same random phases Φ used for the generation of $H(r)$, we define the field:

$$m(r) = \frac{H(r)}{R(r)} \tag{5}$$

$m(r)$ represents a dimensionless random field with spatial correlation similar to that of the fields from which it is derived, always bounded because $R(r) > 0$, with approximately unitary values proportional to the ratio between a random version of the field with the same level of details of measured data X_t and the corresponding lower resolution field $\hat{X}_t$. This field, used as a multiplier for the output of the neural network output $\hat{X}_t$, is able to inject

random noise at scales not resolved by the network and with the correct spatial correlation. The effect of the multiplicative process in the 'assimilation' phase can be evaluated by the comparison of the power spectra evaluated before (green lines) and after downscaling (blue lines) and those of the measured data (red lines), in Figure 4, and visually by comparison of the 3 rows of 12 fields observed, "assimilated" by *NN* and downscaled in Figure 3.

During the extrapolation stage, the network proceeds recursively by predicting the next step based on the previous radar images, with 12 time steps of 5′ to make a one-hour forecast, and using also the forecasts at the previous instants as input for all steps after the first.

As a result, the problem of unresolved scales becomes more and more apparent, as shown in Figure 5, where the green lines for the spectra of the predicted field at the 12 time steps show a progressive loss of spatial detail that spreads eventually over all scales. To overcome this issue, we devised to operate in the same way as for the assimilation step, dividing the extrapolation process into 5′ time steps and adding at each step a stochastic component coherent with the scales not explicitly resolved by PredNet via a multiplicative process with coefficients $m(r)$ defined in Equation (5). In this procedure, the only free parameter is the number n of timesteps used to compute the average power spectra $\overline{|P(X_t)_K|^2}$. Taking $n = 1$, i.e., without averaging, one will generate fields with a high but unrealistic level of detail. Conversely, excessively large values of n will have the effect of making the downscaling procedure ineffective. Empirically, it has been found to be optimal choosing $n = 3$. This value is used both in the definition of the perturbation scenarios (assimilation) and in the downscaling phase during the forecast. The effect of the multiplicative process on the unperturbed fields can be visually appreciated by comparison with the fields in the three groups of stamps (*obs*, *NN* and *m0* to *m9*) in Figure 6 (see the caption for explanatory details).
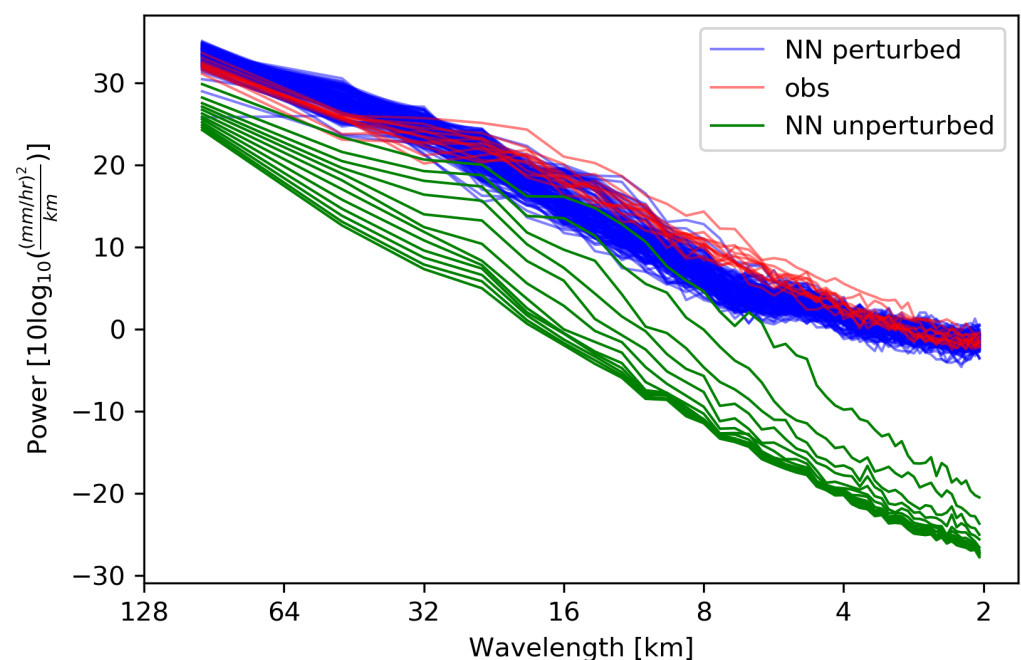

Figure 5. In red, the power spectrum of the 12 images at 5′ intervals of the precipitation field during the nowcast hour; in green, the power spectrum related to the nowcast of the unperturbed PredNet; in blue, the power spectrum for the nowcast of 10 perturbed members.
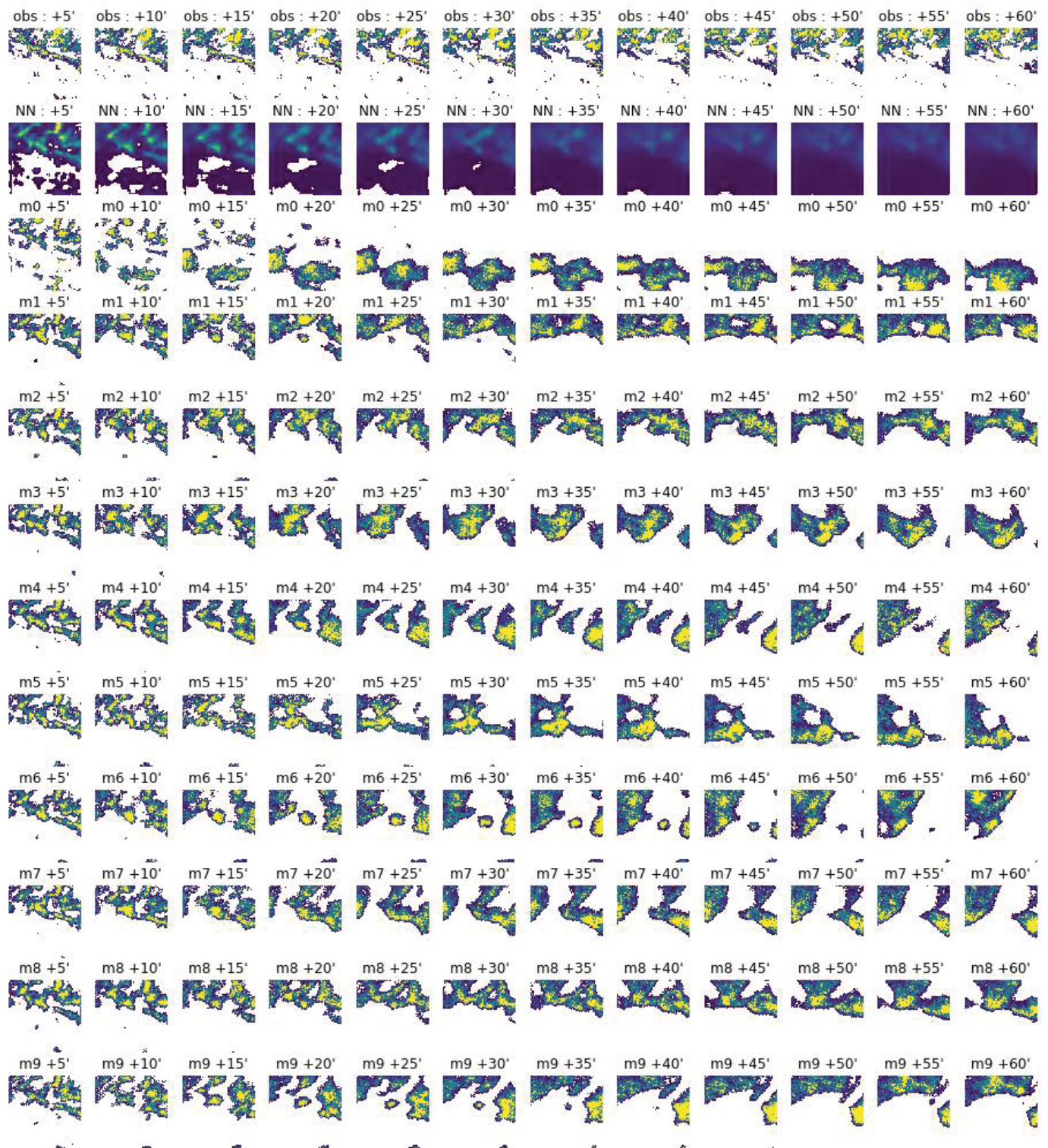
Figure 6. Sequence of 12 images at 5′ time intervals of the precipitation field for the nowcast phase. The first row (*obs*) shows the observed data, the second the deterministic unperturbed nowcast (*NN*) and the next 10 rows (*m0* to *m9*) the nowcast of 10 perturbed members.

In conclusion and, schematically, the proposed procedure can be summarised as follows (see pseudocode representation in Algorithm 1):

1. A set of equally probable scenarios *S* is generated in the hour before the forecast is issued, injecting into the neural network's estimate *D* of the measured data a noise

component consistent with the difference in spatial detail between the measured data and those "assimilated" by the network itself, using the multipliers N defined in Equation (5);

2. With the same method, at each nowcast time step f ahead of 5′, a stochastic component, consistent with the uncertainty evaluated under similar conditions in the period immediately preceding its emission, is added to the prediction;

3. We impose that the mean value of the ensemble members is equal to the deterministic unperturbed prediction of PredNet (*em_adjust* precedure), and that the Cumulative Distribution Function (*cdf_adjust* procedure) of the predicted field is equal to that of the last observed field;

4. Starting from step 2, the procedure is iterated for all 12 time steps of 5′ each, and for each ensemble member, obtaining NN, the one-hour forecast for the chosen number M of scenarios.

Algorithm 1: Pseudocode of the stochastic PredNet algorithm

Input : $O_{[-11:0]}$ `// 12 radar images of the past hour`

Output: $NN_{[1:M,1:12]}$ `// ensemble of M forecasts, 1 hour long`

`// PredNet 12 time steps deterministic nowcast` $D_{[1:12]}$

$D_{[-11:12]} = PredNet(O_{[-11:0]}, steps = 12)$

    `// for each time step in the past hour an ensemble of` M
    `scenarios` S `is built adding noise` N `at scales not resolved by`
    `the deterministic PredNet nowcast` D

for $f = -11$ *to* 0 **do**
    for $m = 1$ *to* M **do**
        $S_{m,f} = D_f \cdot N_{m,f}$
        `// cdf is adjusted to that of the observed field`
        $S_{m,f} = cdf\_adjust(S_{m,f}, O_f)$

    `// evolve perturbed scenarios` S `with PredNet`

for $f = 1$ *to* 12 **do**
    `// evolve each scenario independently`
    for $m = 1$ *to* M **do**
        $NN_{[m,-11:f]} = PredNet(S_{[-11:f-1]}, steps = 1)$
    for $m = 1$ *to* M **do**
        `// ensemble mean of NN is adjusted to the nowcast D`
        $NN_{m,f} = em\_adjust(NN_{m,f}, D_f)$
        `// noise is added compatible with unresolved scales`
        $NN_{m,f} = D_f \cdot NN_{m,f}$
        `// cdf is adjusted to that of the last observed field`
        $NN_{m,f} = cdf\_adjust(NN_{m,f}, O_0)$
        `// last NN is added as predictor to the scenario` m
        $S_{m,f} = NN_{m,f}$

The constraint of the ensemble mean, at step 3 of the described procedure, binds the ensemble mean solution to remain identical to the deterministic prediction. Without this constraint, each of the previously determined scenarios, with its associated stochastic component and thus uncertainty, can evolve independently of the others and can thus in principle diverge, and in any case be completely different, from the optimal solution obtained from the unperturbed deterministic forecast (see an example in the second row of images in Figure 6).

Theoretical support of the proposed methodology rests on the conjecture that the deterministic forecast, which we have already verified, in [22], gives better results than any other state-of-the-art procedure in terms of RMSE, indirectly proving that the solution can be considered optimal on average. By constructing around the deterministic solution an ensemble that has a variance (spread) that on average is directly proportional to the RMSE of the deterministic forecast (which we have assumed to coincide with the ensemble mean), we are implicitly verifying the plausibility of the initial conjecture. This is precisely what is discussed in Section 3.1 of the article.

2.4. The Benchmark Nowcast Technique: STEPS

One of the most effective approaches that can provide realistic and high-quality predictions that is widely used in operational contexts is provided by the method called *STEPS* [8]. See also [10], where more recent developments are also discussed.

STEPS is available in the public domain through a Python library called *pySTEPS* [28]. *pySTEPS* is used by a large community, and in addition to being used as an operational tool for nowcasting, it is used as a development, benchmarking, and study platform for implementing new techniques. For the estimation of the advection field, *pySTEPS* implements three methodologies: a local type based on the classical Lucas-Kanade (LK) technique [29,30], a global variational type (VET) [31,32], and a spectral type (DARTS) [33]. Extrapolation by advection is performed with a backward-in-time semi-Lagrangian scheme [32]. The most sophisticated technique in the *pySTEPS* library, used here as a benchmark, is called *STEPS* and takes into account the dependence of the predictability of precipitation phenomena for different scales. In practice, using an approach called SPROG [9], after a logarithmic transformation, a decomposition of the initial field into a multiplicative cascade of fields at decreasing spatial scales is performed, and each of these spatial fields is advanced with an advection obtained consistently during the period preceding the nowcast. Field decomposition at different scales can be achieved using techniques of varying degrees of complexity, such as the discrete cosine transform or wavelets. In this case, an FFT was used in which the different portions of the spectrum are weighted by Gaussian functions [11] and then, via inverse FFT, returned to the spatial domain. The additional uncertainty due to the fact that growth, decay, creation, and/or ultimate disappearance of precipitating cells may occur during the period of validity of the prediction, along with uncertainties related to the different spatial scales, have been taken into account through stochastic noise built on the spatial and temporal variability, observed in the period prior to the emission, which is added to each partial field.

For implementation details, see [28] and the related bibliography. This approach makes it viable to obtain not a single deterministic prediction, but a set of equally probable predictions (ensemble) and thus also to quantify their spatio-temporal uncertainty. The results obtained with this methodology will be indicated with *STEPS* in the following.

3. Results

The performances of the proposed method were evaluated for the year 2017. Data of 2015 and 2016 were used for training and validation of the proposed *NN* method. The performance indicators were evaluated against the precipitation value, at the time of forecast verification, estimated by the radar itself and expressed in mm/h. Of course, they do not represent a direct measure of error with respect to actual rainfall. The implicit assumption is that systematic errors in rainfall estimation will be offset by equivalent biases in the forecast. Ultimately, the error in the radar forecast is expressed in units transformed to mm/h.

The indicators used to verify the global quality of the probabilistic forecast are the Continuous Ranked Probability Score (CRPS) and the Rank Histogram (RH), and the class of indicators related to exceeding a fixed precipitation threshold: the Reliability Diagram (RD), the Receiver Operating Characteristic (ROC) curve, and the associated Area Under the Curve (AUC) [34].

The performance study was performed on about 2500 precipitation one-hour long events (equivalent to more than 100 rainy days) in 2017 and corresponding *NN* and *STEPS* nowcasts were compared systematically to verify the effectiveness of the proposed methodology. Here, we define as an event, any occurrence of rain that happened during the year that allows the nowcast procedure to be used. In practice, this means that, in the three images before the forecast, the surface area of the domain covered by precipitation (Wet Area Ratio) is at least 10%. A single "event" is one hour long (the maximum nowcast time), and it is non overlapping with the following one. This choice incorporates into the performance evaluation also events of weak intensity that are usually discarded in the reviews of similar benchmark studies found in the literature. The analysis of a full year's worth of data allows the effectiveness of the proposed approach to be verified in the most statistically robust way, leaving it to further study to optimize the learning process for specific high-intensity events. The usefulness of the nowcast of precipitation using radar imagery is related to its use in operational contexts to estimate, for example, hydrologic risk following intense precipitation events. From this point of view, rather than the forecast at a given horizon of the instantaneous rain rate, one is interested in the cumulative value of precipitation. For this reason, verification was made on the cumulative forecast precipitation in the first hour of the forecast and only in some cases, as specified in the text, for that accumulated in the first half hour. For ease of reading, we will divide the discussion of the results into specific sections, one for each of the indicators mentioned. We begin by first discussing a key characteristic that an ensemble must possess in order to describe the uncertainty of the particular event being predicted: it must be able to adjust its variability between members (spread) in such a way to be proportional to the forecast error. This attribute of an ensemble is known as the spread–skill relationship, and ultimately represents an a posteriori verification of the plausibility of the assumptions made to construct our ensemble.

3.1. Spread–Skill

In [35], Palmer showed that a perfect ensemble must possess on average the characteristic that its spread, evaluated as the square root of the ensemble variance, must be identical to the root mean square error (RMSE) with respect to the ensemble mean. Figure 7 shows the scatter plot of the two variables for all the test cases and the 12 times of forecast and the correlation coefficient between the two time series. As can be seen, the spread is well-correlated with the RMSE. A more careful analysis shows that its values are slightly but systematically lower, especially for higher RMSEs (see also average values in the legend of the figure). This means that on average the NN ensemble is slightly underdispersive. This issue, which is also common to the *STEPS* ensemble, is nonetheless of minor importance because, as we will see from the analysis conducted later on the rank histograms in Section 3.3, the ensemble's overall characteristics in terms of spread are very good.

3.2. CRPS

The CRPS is a generalization of the mean absolute error for a deterministic forecast. For a continuous variable the instantaneous, $CRPS_t$ is defined as the quadratic difference between the forecast CDF and the empirical CDF of the observed value. When the forecast CDF is known only through an ensemble, different estimators of the instantaneous CRPS can be written. Following the indication in [36], we choose to use the expression:

$$CRPS_t = \frac{1}{M} \sum_{m=1}^{M} |X_t - \hat{X}_t^m| - \frac{1}{2M^2} \sum_{m,n=1}^{M} |\hat{X}_t^m - \hat{X}_t^n| \tag{6}$$

where $M = 100$ is the number of ensemble members, X_t is the observed radar image at time t and $\hat{X}_t^m$ the *m.th* ensemble nowcast image verifying at the same time t.
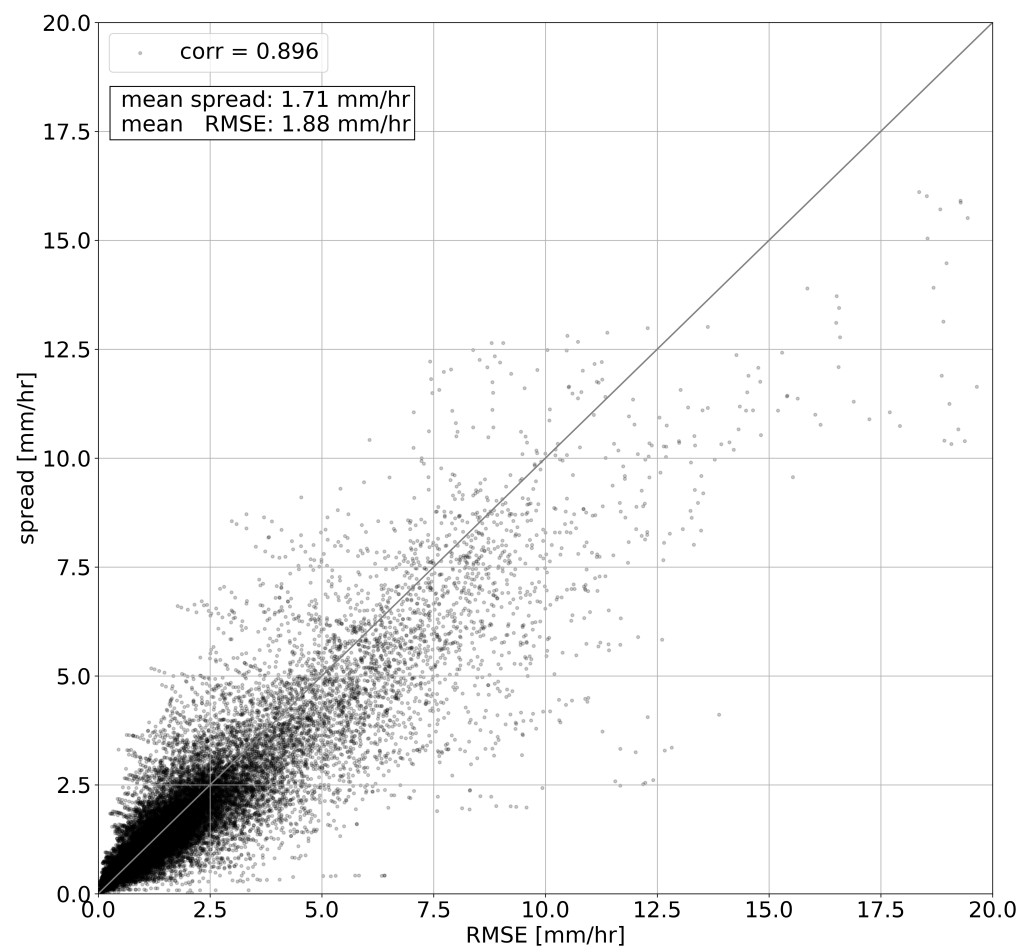
Figure 7. Scatter plot of spatial average of the ensemble mean root mean square error against ensemble spread for all the test cases and the 12 times of forecast. The labels show the correlation coefficient between the two time series and their respective mean values.

The first type of probabilistic verification was performed by comparison of the CRPS, averaged over all of the domains, for the one-hour accumulated precipitation evaluated for all the events. Results are shown in Figure 8. Despite the completely different approach, it is interesting to note that the two methods have, on average, very similar CRPS values. However, the full-year 2017 average CRPS of *NN* (0.54 mm/h) is slightly better than that of *STEPS* (0.56 mm/h) as shown in the legend within the figure. This proves a slightly better overall performance of the probabilistic forecast by *NN* than by *STEPS*. It should also be noted that this CRPS value for *NN* is evaluated for the entire domain, while that for *STEPS* is clearly limited to the domain of existence of the *STEPS* forecast only.

3.3. Rank Histogram

The observed values should be uniformly distributed among the ensemble members if they have equal probability, as hypothesized. The rank histogram, showing the frequencies at which the verifying observation falls in each of the 101 bins defined by the 100 ensemble member, should then be flat in the ideal case. It is interesting to compare the rank histograms for the two methodologies obtained by accumulating the prediction–observation comparisons over all the events studied so as to obtain an idea of the average spread and thus how reliable are the related ensemble forecasts. Due to the oft-mentioned limitations of the *STEPS* procedure in relation to the size of the domain covered by the radar, and to highlight the trend of this indicator as a function of forecast time in Figure 9, we first show the rank histogram for *NN* and *STEPS* for the value of the cumulative precipitation in the first 30′ of the forecast. Since the performance of each of the methods can vary substan-

tially depending on the event, the distribution for the test cases studied using box-plot, and whiskers delimiting the 5% and 95% quantiles, was shown in the figure. From the analysis and comparison of the two figures, it can be seen that, on average, for the all the events the rank for both methods gives indications of good quality for probabilistic prediction even at such a high level of detail in probability space. Careful analysis confirms a slight under-dispersion bias for both methods. *STEPS* also shows a slight bias with a tendency to overestimate the frequency of the medium high intensity events.
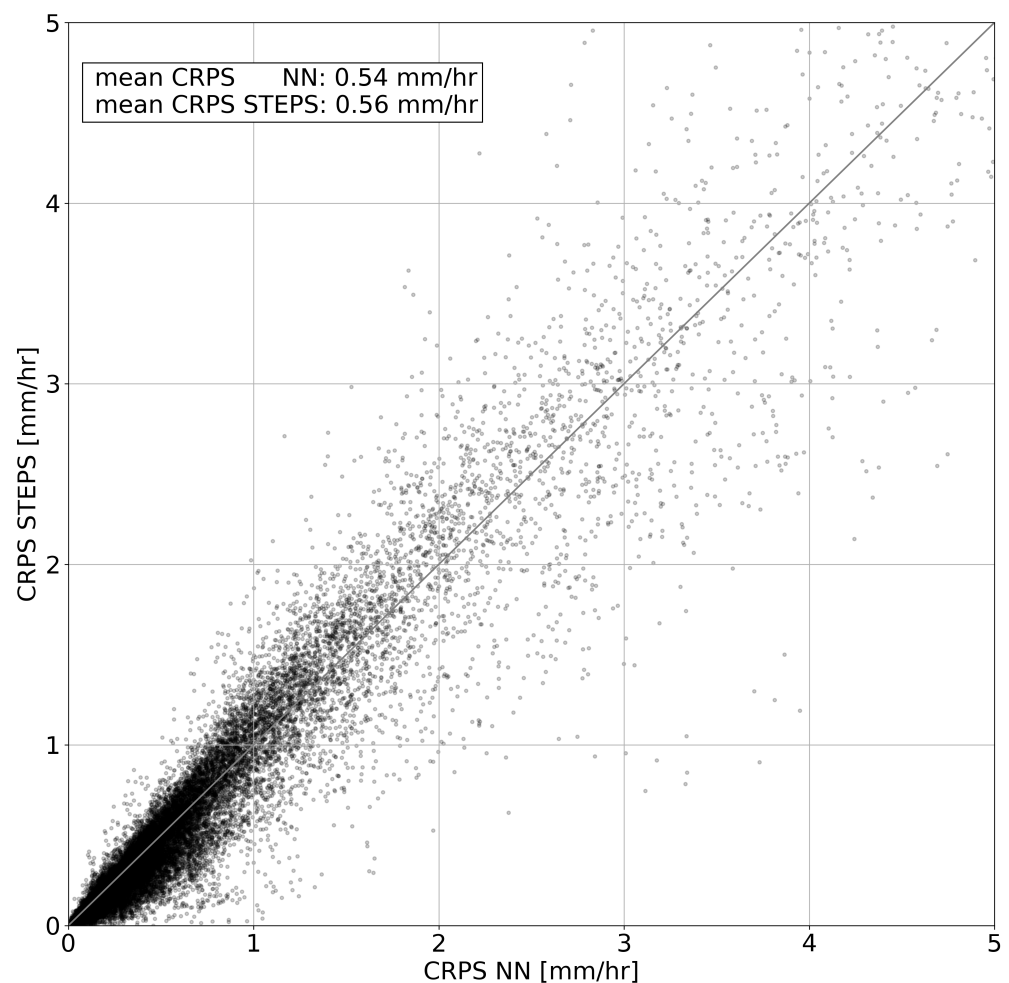
Figure 8. Plot of the *STEPS* against the *NN* CRPS for the nowcast of the one hour cumulated rain for all the precipitation events of 2017.

Figure 10 show the cumulative rank histogram for *pySTEPS* and *NN* for the prediction of cumulative rainfall in the first hour of the forecast. As can be seen, both methodologies have on average a class distribution that is close to ideal although the rank histogram of *NN* is slightly more uniform.
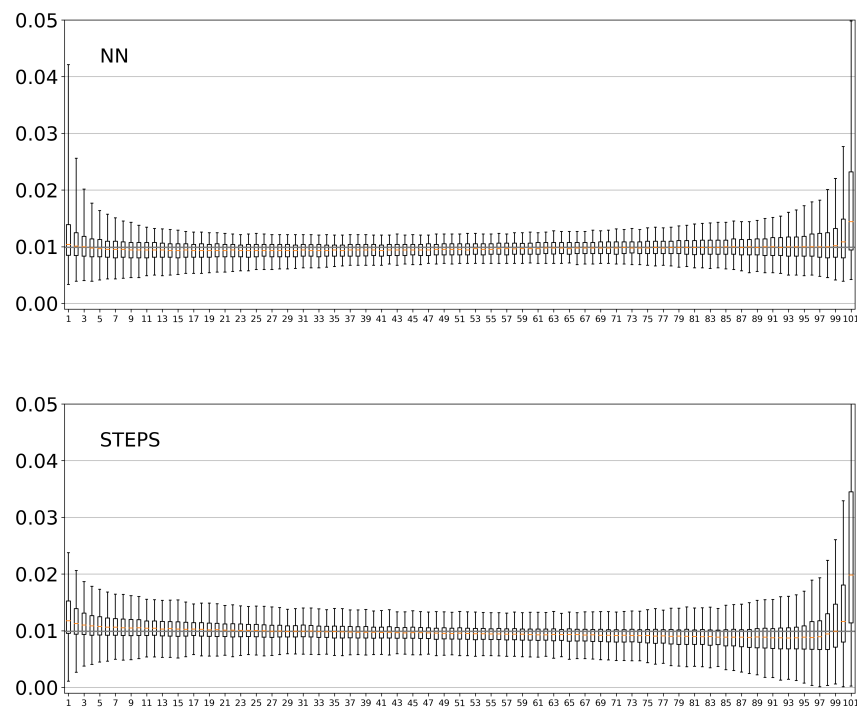
Figure 9. Boxplot of the rank histogram for the 100-member ensemble of half-hourly cumulative precipitation forecasts by the *NN* method and for the *STEPS* method for all rainfall events in 2017. The reach of the whiskers beyond the first and third quartiles are set to the 5% and 95% percentiles.
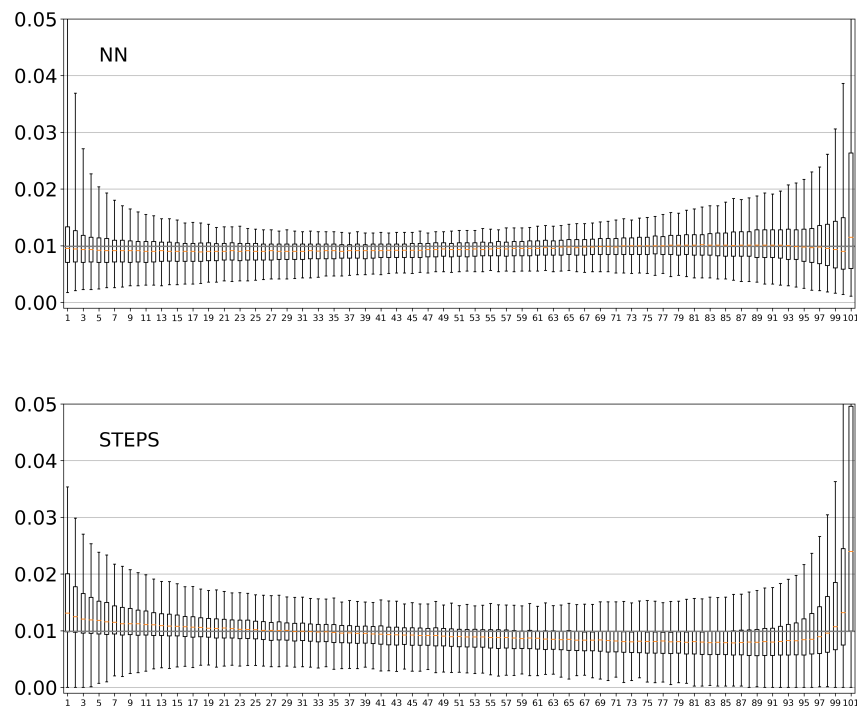

Figure 10. As for the previous figure for the accumulated precipitation in the first hour.

3.4. Reliability Diagram

The evaluation of performance related to the ability to predict the exceeding of given thresholds for the value of cumulative precipitation is also of considerable interest. Figure 11 shows the Reliability Diagram (RD) for the exceeding of the 0.1 mm/h threshold for cumulative precipitation in the first hour of nowcast, which we can take as an indication of

the ability to discriminate rainy areas from non-rainy areas. The shaded areas depict the area where the forecast has predictive value from a probabilistic point of view while the lower subplots show the numerosity with which each of the probability classes is predicted (probability histograms). Comparison with the plots on the left for *NN* and those on the right for *STEPS* reveals perhaps a slight prevalence of *STEPS* that remains very close to the ideal behavior represented by the dashed diagonal. From the respective probability histograms, it is important to note that intermediate probabilities are predicted much less frequently than those for extreme values (note the logarithmic scale for N), which is a clear indication that the ensembles are very selective in assigning high or low probabilities and thus effective (this characteristic is called "sharpness"). The same plots are shown in Figure 12 for exceeding gradually increasing precipitation thresholds: 1 mm/h, 5 mm/h, and 20 mm/h. These threshold values can be considered representative for the detection of low, medium and high intensity events. It can be seen that the diagram for *NN* and the associated probability histogram give a reliable forecast indication on all thresholds considered staying always well within the shaded area. The noisiness of the lines of the diagram for a 20 mm/h threshold is to be related not only to the greater difficulty in predicting such events, which occur with very low probability, but also to the small number of the sample against which the verification can be carried out. Note also the greater sharpness of NN when compared with *STEPS*, which "tries" to predict these extremely rare events even with a high probability.
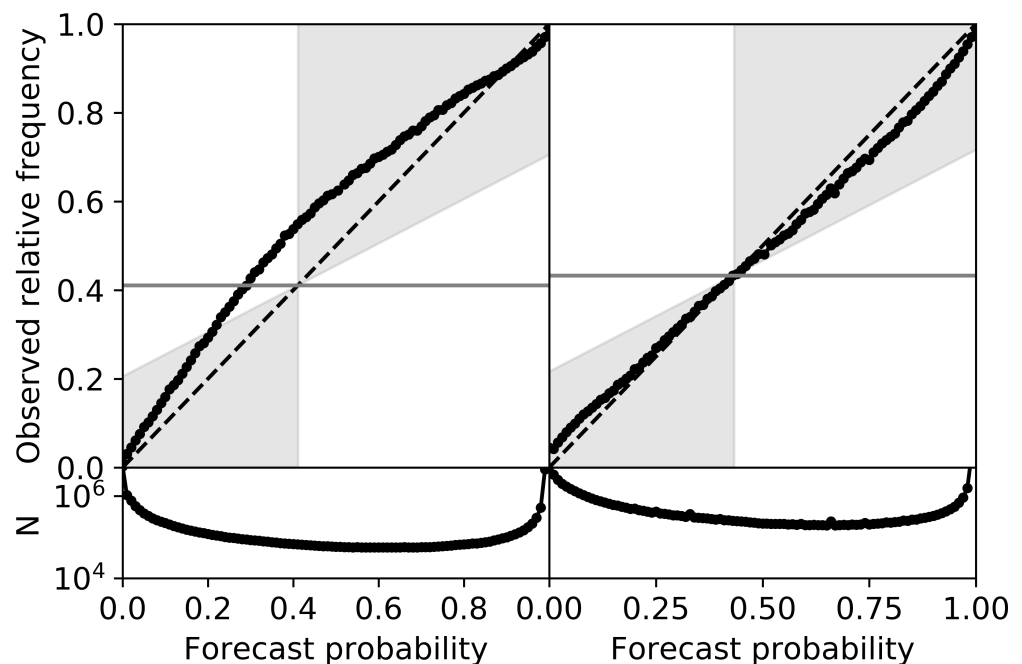

Figure 11. Reliability Diagram for exceeding the cumulative precipitation threshold of 0.1 mm/h in the first half hour of the forecast. On the right is the plot for *NN* and on the left is the plot for *STEPS*. The lower subplots represent, on a logarithmic scale, the numerosity with which each forecast probability class was issued. The gray horizontal line corresponds to the frequency with which the event is observed (climate), while the shaded areas highlight the area where the ensembles have good reliability and thus predictive value from a probabilistic point of view.
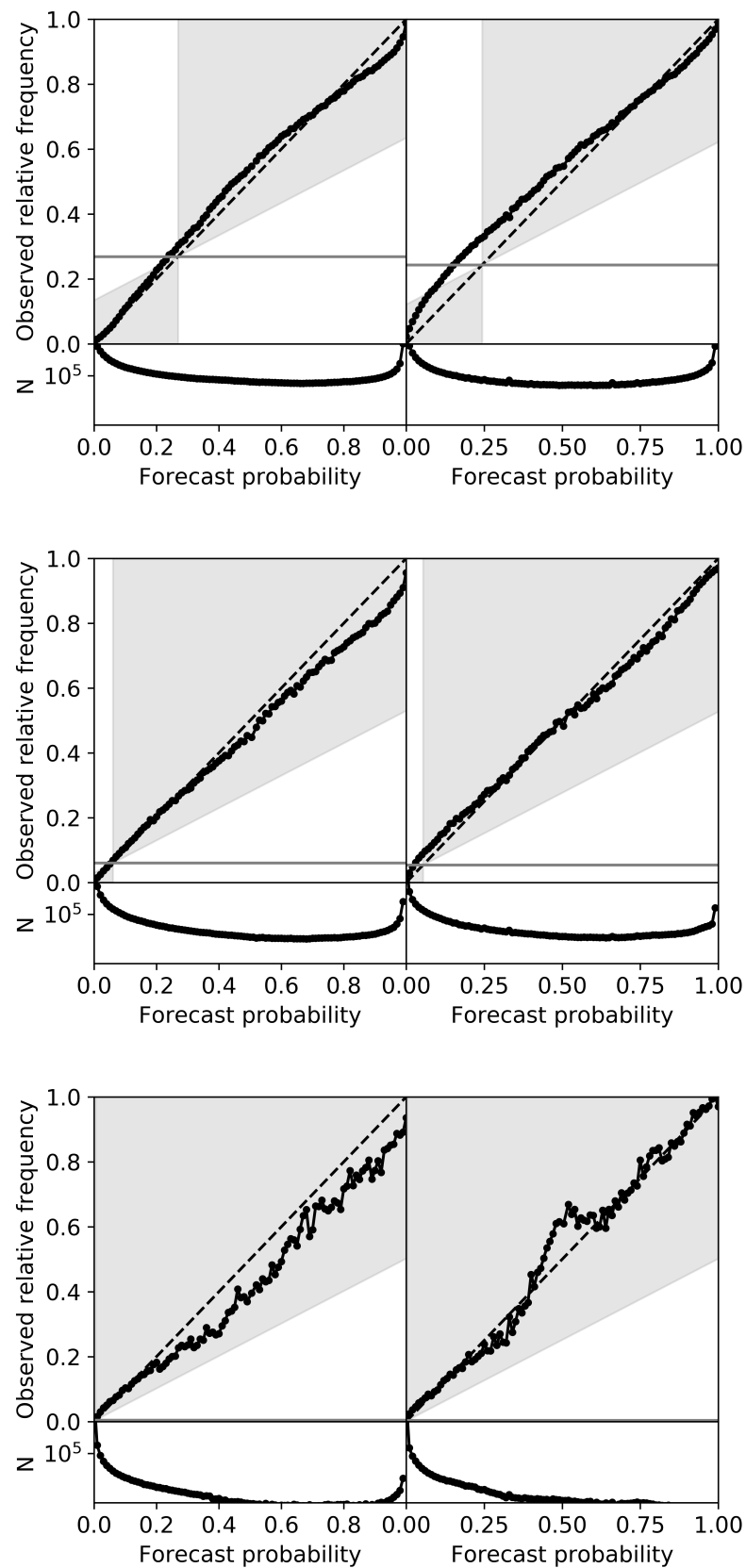
Figure 12. Reliability diagram for *NN* (to the right) and *STEPS* (to the left), for exceeding precipitation thresholds of: 1 mm/h (fisrt row), 5 mm/h (second row) and 20 mm/h (third row).

3.5. ROC Curve

A further indication about the quality of the probabilistic forecast comes from the ROC curve, which expresses a measure of the ability to characterize the forecast correctly in terms of expected probability. The graph is obtained by plotting the frequencies of forecast success (POD) versus false alarm frequencies (POFD) and can be used, for example, to identify the probability threshold that offers the best trade-off between costs and benefits. Figure 13 shows the plot for *NN* and *STEPS* for exceeding the cumulative precipitation threshold of 1 mm/h in the first hour of forecast. As can be seen *NN* is much closer to the ideal behavior (POD = 1 and POFD = 0) than *STEPS*, again indicating a better quality of the *NN* nowcast. An overall measure of forecast quality is expressed by the area under curve (AUC), which in the ideal case should be equal to 1. AUC values less than 0.5 (area under the dashed diagonal) indicate zero probabilistic forecast utility. Table 3 shows comparisons of AUC values for *NN* and *STEPS* for different cumulative thresholds.
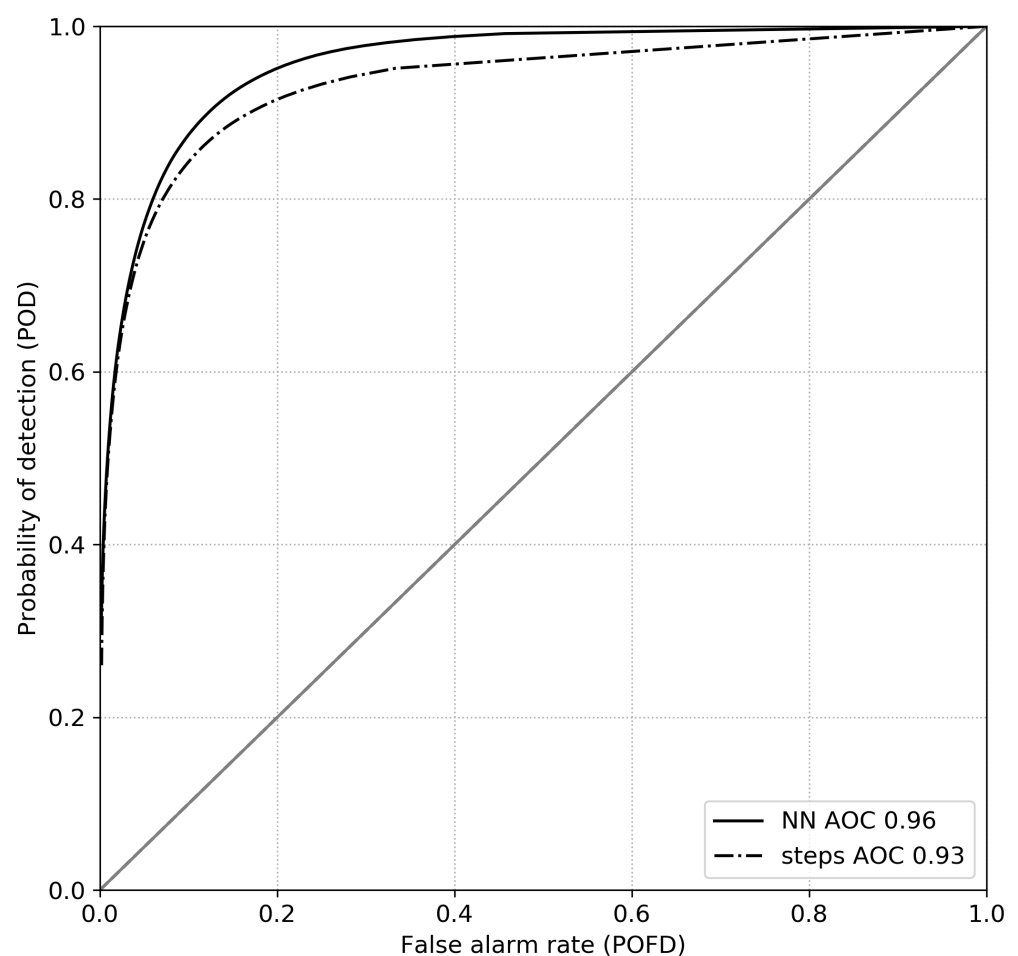

Figure 13. ROC curve for *NN* and *STEPS* for exceeding the cumulative precipitation threshold of 1 mm/h in the first hour of nowcast.

The variability of the AUC estimate for *STEPS* as the rainfall threshold increases, which in theory should be monotonically decreasing, is most likely to be related to the reduction in the nowcast area of *STEPS*. On the other hand, the *NN* procedure is designed precisely to circumvent this drawback, and in any case it consistently shows higher AUC values than those of *STEPS*.

Table 3. Area under the ROC AUC curve for *NN* and *STEPS* for exceeding different cumulative precipitation thresholds in the first hour of the forecast.

Tresh [mm/h]	NN	STEPS
1	0.96	0.93
2	0.96	0.93
5	0.96	0.94
10	0.95	0.94
16	0.94	0.93
20	0.93	0.91

4. Discussion

This paper is the natural continuation of [22] in which the same PredNet neural network used here was applied to the same data to verify the performance in deterministic terms of the prediction up to one hour of the precipitation field estimated from the images of a weather radar.

The proposed method uses the aforementioned network in a configuration capable of providing a set of forecasts all equally likelihood, through which the uncertainty of the predictor can be characterised in probabilistic terms.

The procedure has proven to succeed in providing a probabilistic forecast of a quality comparable or superior to that obtainable through *STEPS* even for a radar with a limited spatial coverage, such as a small urban radar, and thus capable of providing the decision maker with a powerful forecasting tool to mitigate risks from severe precipitation events.

Summarising the major achievements of this work, we have the following:

- We proposed a physical-statistical interpretation of the result obtained from the application of the PredNet network for the nowcasting of radar images, as the ensemble mean of a probabilistic forecast;
- This interpretation was tested by constructing around the unperturbed forecast an ensemble of scenarios obtained by adding noise, with the correct spatial correlation, compatible with the scales not explicitly resolved by the PredNet;
- The detailed analysis made on extended data set leads us to conclude that the method we proposed has performance superior or equal to the probabilistic *STEPS* method, which represents a well recognized benchmark;
- The proposed solution, based on a generative model, succeeds at providing an overall good quality probabilistic prediction for the entire domain covered by the radar and thus solves the problem that occurs when implementing a prediction procedure based on optical flow for a radar of limited spatial range;
- For a domain of small size, the extrapolation phase of the procedure can be performed with limited computational resources, allowing it to be performed in low-cost edge-computing devices that can be included in the same radar apparatus.

The main motivation that inspired us to conduct this study was to verify the feasibility of a nowcasting procedure for a rain radar installed in the city of Cagliari (Sardinia, Italy) [37] for which we do not have till now sufficient data to make a specific in-depth study. Of course, the results discussed are not directly exportable to another radar or another geographic area. However, there is also no valid reason to assume that the procedure could be applied with similar success to other radars and other geographic areas by repeating the network learning phase and using the same operating procedure. Of course, in the future, this assumption will have to be explicitly verified for the specific case.

Since training of the network was done using a general "event" choice criteria, it is very likely, therefore, that the forecast performance for intense events could be improved, even significantly, by using a more specific "event" choice criteria during the training phase.

Comparison with multiple alternative methods, which is lacking here, is essential to verify the quality of the results of a study. In this specific case, however, we were interested in the development of an open source solution to the probabilistic nowcasting

problem where, to our knowledge, the only one, and moreover unanimously recognized by the meteorological community as one of the best, is indeed *STEPS*. Another aspect that characterized the proposed method is the capability to be processed by a single GPU card, assuming a possible operational implementation in a city radar operating on a small scale and with reduced computational capabilities. The other probabilistic and neural network-based nowcasting procedures discussed in the introduction are not available, to our knowledge, in the public domain and moreover rely on the use of significant computational resources.

To our knowledge, this operational approach and the underlying interpretive idea are a completely new development and may represent a small step forward from the state of the art both in the field of operational procedures for radar nowcasting, especially for small radars, and in the interpretive aspect of the results obtainable through the use of a generative neural network.

References

1. Schmid, F.; Wang, Y.; Harou, A. Nowcasting Guidelines—A Summary. *Bull. Nº* **2019**, *68*, 2.
2. Ayzel, G.; Heistermann, M.; Winterrath, T. Optical flow models as an open benchmark for radar-based precipitation nowcasting (rainymotion v0. 1). *Geosci. Model Dev.* **2019**, *12*, 1387–1402. [CrossRef]
3. Dixon, M.; Wiener, G. TITAN: Thunderstorm identification, tracking, analysis, and nowcasting—A radar-based methodology. *J. Atmos. Ocean. Technol.* **1993**, *10*, 785–797. [CrossRef]
4. Chang, P.L.; Zhang, J.; Tang, Y.S.; Tang, L.; Lin, P.F.; Langston, C.; Kaney, B.; Chen, C.R.; Howard, K. An operational multi-radar multi-sensor QPE system in Taiwan. *Bull. Am. Meteorol. Soc.* **2021**, *102*, E555–E577. [CrossRef]
5. Smith, K.; Austin, G. Nowcasting precipitation—A proposal for a way forward. *J. Hydrol.* **2000**, *239*, 34–45. [CrossRef]
6. Chung, K.S.; Yao, I.A. Improving radar echo Lagrangian extrapolation nowcasting by blending numerical model wind information: Statistical performance of 16 typhoon cases. *Mon. Weather Rev.* **2020**, *148*, 1099–1120. [CrossRef]
7. Radhakrishnan, C.; Chandrasekar, V. CASA prediction system over dallas–fort worth urban network: Blending of nowcasting and high-resolution numerical weather prediction model. *J. Atmos. Ocean. Technol.* **2020**, *37*, 211–228. [CrossRef]
8. Bowler, N.E.; Pierce, C.E.; Seed, A.W. STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Q. J. R. Meteorol. Soc.* **2006**, *132*, 2127–2155. [CrossRef]
9. Seed, A. A dynamic and spatial scaling approach to advection forecasting. *J. Appl. Meteorol.* **2003**, *42*, 381–388. [CrossRef]
10. Velasco-Forero, C.; Pudashine, J.; Curtis, M.; Seed, A. *STEPS3–ADV–VERIFICATION REPORT*; National Library of Australia Cataloguing-in-Publication Entry: Melbourne, VIC, Australia, 2020. [CrossRef]
11. Pulkkinen, S.; Chandrasekar, V.; Harri, A.M. Fully spectral method for radar-based precipitation nowcasting. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 1369–1382. [CrossRef]
12. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.c. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 802–810.

13. Shi, X.; Gao, Z.; Lausen, L.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Deep learning for precipitation nowcasting: A benchmark and a new model. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5617–5627.

14. Ayzel, G.; Scheffer, T.; Heistermann, M. RainNet v1. 0: A convolutional neural network for radar-based precipitation nowcasting. *Geosci. Model Dev.* **2020**, *13*, 2631–2644. [CrossRef]

15. Foresti, L.; Sideris, I.V.; Nerini, D.; Beusch, L.; Germann, U. Using a 10-year radar archive for nowcasting precipitation growth and decay: A probabilistic machine learning approach. *Weather Forecast.* **2019**, *34*, 1547–1569. [CrossRef]

16. Trebing, K.; Stanczyk, T.; Mehrkanoon, S. SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture. *Pattern Recognit. Lett.* **2021**, *145*, 178–186. [CrossRef]

17. Sønderby, C.K.; Espeholt, L.; Heek, J.; Dehghani, M.; Oliver, A.; Salimans, T.; Agrawal, S.; Hickey, J.; Kalchbrenner, N. MetNet: A Neural Weather Model for Precipitation Forecasting. *arXiv* **2020**, arXiv:2003.12140.

18. Ravuri, S.; Lenc, K.; Willson, M.; Kangin, D.; Lam, R.; Mirowski, P.; Fitzsimons, M.; Athanassiadou, M.; Kashem, S.; Madge, S.; et al. Skilful precipitation nowcasting using deep generative models of radar. *Nature* **2021**, *597*, 672–677. [CrossRef] [PubMed]

19. Bouget, V.; Béréziat, D.; Brajard, J.; Charantonis, A.; Filoche, A. Fusion of rain radar images and wind forecasts in a deep learning model applied to rain nowcasting. *arXiv* **2020**, arXiv:2012.05015.

20. Ashesh, A.; Chang, C.T.; Chen, B.F.; Lin, H.T.; Chen, B.; Huang, T.S. Accurate and Clear Quantitative Precipitation Nowcasting Based on a Deep Learning Model with Consecutive Attention and Rain-Map Discrimination. *Artif. Intell. Earth Syst.* **2022**, *1*, e210005. [CrossRef]

21. Lotter, W.; Kreiman, G.; Cox, D. Deep predictive coding networks for video prediction and unsupervised learning. *arXiv* **2016**, arXiv:1605.08104.

22. Marrocu, M.; Massidda, L. Performance comparison between deep learning and optical flow-based techniques for nowcast precipitation from radar images. *Forecasting* **2020**, *2*, 194–210. [CrossRef]

23. Farris, S.; Seoni, A.; Ruggiu, D.; Bertoldo, S.; Perona, G.; Allegretti, M.; Marrocu, M.; Badas, M.; Viola, F.; Deidda, R. First analyses of rainfall patterns retrieved by a newly installed X-band radar over the Metropolitan area of Cagliari (Sardinia, Italy). In *Rainfall Monitoring, Modelling and Forecasting in Urban Environment. UrbanRain18: 11th International Workshop on Precipitation in Urban Areas. Conference Proceedings*; ETH Zurich, Institute of Environmental Engineering: Pontresina, Switzerland, 2019; pp. 39–40.

24. Sato, R.; Kashima, H.; Yamamoto, T. Short-Term Precipitation Prediction with Skip-Connected PredNet. In Proceedings of the International Conference on Artificial Neural Networks, Rhodes, Greece, 4–7 October 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 373–382.

25. Van Rossum, G.; Drake, F.L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, USA, 2009.

26. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*; Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: Vancouver, BC, Canada, 2019; pp. 8024–8035.

27. Rebora, N.; Ferraris, L.; Hardenberg, J.; Provenzale, A. RainFARM: Rainfall Downscaling by a Filtered Autoregressive Model. *J. Hydrometeorol.* **2006**, *7*, 724–738. [CrossRef]

28. Pulkkinen, S.; Nerini, D.; Pérez Hortal, A.A.; Velasco-Forero, C.; Seed, A.; Germann, U.; Foresti, L. Pysteps: An open-source Python library for probabilistic precipitation nowcasting (v1. 0). *Geosci. Model Dev.* **2019**, *12*, 4185–4219. [CrossRef]

29. Lucas, B.D.; Kanade, T. An iterative image registration technique with an application to stereo vision. In Proceedings of the Seventh International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada, 24–28 August 1981; pp. 674–679.

30. Bouguet, J.Y. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. *Intel Corp.* **2001**, *5*, 4.

31. Laroche, S.; Zawadzki, I. Retrievals of horizontal winds from single-Doppler clear-air data by methods of cross correlation and variational analysis. *J. Atmos. Ocean. Technol.* **1995**, *12*, 721–738. [CrossRef]

32. Germann, U.; Zawadzki, I. Scale-dependence of the predictability of precipitation from continental radar images. Part I: Description of the methodology. *Mon. Weather Rev.* **2002**, *130*, 2859–2873. [CrossRef]

33. Ruzanski, E.; Chandrasekar, V.; Wang, Y. The CASA nowcasting system. *J. Atmos. Ocean. Technol.* **2011**, *28*, 640–655. [CrossRef]

34. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*; Elsevier Academic Press: Amsterdam, The Netherlands; Boston, MA, USA, 2011.

35. Palmer, T.; Buizza, R.; Hagedorn, R.; Lawrence, A.; Leutbecher, M.; Smith, L. Ensemble prediction: A pedagogical perspective. *ECMWF Newsl.* **2006**, *106*, 10–17.

36. Zamo, M.; Naveau, P. Estimation of the Continuous Ranked Probability Score with Limited Information and Applications to Ensemble Weather Forecasts. *Math. Geosci.* **2018**, *50*, 209–234. [CrossRef]

37. Deidda, R.; Farris, S.; Badas, M.G.; Marrocu, M.; Massidda, L.; Seoni, A.; Urru, S.; Viola, F. A downscaling framework for precipitation nowcasting by merging radar retrievals at different scales and resolutions. In Proceedings of the EGU General Assembly 2021, Virtual, 19–30 April 2021.