

A Comparison of Nowcasting Methods on the Italian Radar Mosaic

Valentina Gregori¹, Ferdinando De Tomasi¹, Gianluca Ferrari² and Andrea Chini²

Abstract—Nowcasting refers to high-resolution meteorological forecasting on a short period, with the aim to identify imminent critical situations. Indeed, numerical weather forecasting is normally not detailed enough and may take too long to calculate for this scenario. Nowcasting techniques typically analyze radar echoes and range from simple advection algorithms to more complex methods that attempt to model growth and decay of precipitation fields. These forecasts are highly relevant for a variety of industrial, agricultural, and leisure activities. Radarmeteo, a company providing meteorological services, uses nowcasting to assist the work of their forecasters and to provide customized solutions to business customers. To increase the value of their products, Radarmeteo was interested in 1) augmenting the usability of their nowcasts by shortening the computation times and 2) examine more recent methods to improve their quality. In this work, we compare four popular nowcasting algorithms with their current proprietary method. We selected 193 interesting events from a radar mosaic of the Italian peninsula and evaluate multiple metrics between the observed and forecast fields. The tested methods resulted generally superior to the existing proprietary algorithm, since they manage to reduce the number of missed events and false alarms, and to increase the spatial accuracy of the forecasts. Moreover, the chosen implementation of these methods reduced the computation time by one or two orders of magnitude.

Keywords—radar, nowcasting, motion field, reflectivity

I Introduction

In meteorology, nowcasting indicates forecasts for the next few hours (i.e., 0 h to 6 h) on a spatial resolution of a few kilometers. These forecasts are based on remote sensing of the atmospheric conditions via radar and, to lesser degree, via satellite. Although numerical weather prediction (NWP) produces more reliable forecasts on a longer time range, it is still lacking at smaller temporal and spatial scales among other reasons

due to its high computational complexity (Simonin et al. 2017; Pulkkinen et al. 2020). Nowcasting systems, on the other hand, use radar or satellite images as immediate snapshots of the state of the atmosphere and, by combining multiple of such snapshots, estimate where precipitation systems are likely to move in the near future.

One of the simplest such approaches is an extrapolation algorithm that consists in moving forward the current radar echoes for each forecast time. This is the Lagrangian persistence method, which works under the assumption that no growth or decay factors act along the advection path (Germann and Zawadzki 2002; Pulkkinen et al. 2020). This assumption is reasonable for large scale stratiform systems, but inadequate for convective situations. More sophisticated nowcasting approaches have been proposed in literature to overcome this limitation. For instance, the forecast skill can be improved by limiting the prediction horizon of small scale features. The underlying idea is that a precipitation field can be described as a hierarchy of structures (Lovejoy and Schertzer 1995) from small (< 100 m) to large (> 200 km) scales and that the lifetime of each such level is proportional to its scale (Schertzer et al. 1997). Nowcasting methods that use this idea do not attempt to forecast structures beyond their expected lifetime; the result is a reflectivity field that gradually becomes smoother as the forecast time increases (Seed 2003; Turner et al. 2004; Ruzanski and Chandrasekar 2011). One way of dealing with this smoothing process is to introduce some stochastic perturbations that represent initiation, growth and decay of precipitation fields (Pulkkinen et al. 2019b; Bowler et al. 2006; Pulkkinen et al. 2019a). In this way, rather than one fixed prediction, the model produces an ensemble of potential realizations of the future, which can be interpreted in probabilistic terms. Lastly, there are hybrid methods that attempt to have the best of two worlds by merging more long-term NWP with the extrapolation-based approach (Liguori and Rico-Ramirez 2012). In essence, these use the prediction obtained by the nowcasting algorithm in the very short term (e.g., one hour) and then gradually transition to a

¹Università del Salento, Piazza Tancredi 7, 73100, Lecce (LE)

²Radarmeteo Srl, Via IV Novembre 119, 35020, Due Carrare (PD)
valentina.gregori@studenti.unisalento.it 0499125902
Ferdinando.DeTomasi@le.infn.it
gianlucaferrari@radarmeteo.com
andreachini@radarmeteo.com

Length: 12 pages, 5159 words, 6 figures, and 2 tables.

NWP prediction in the subsequent 2 h to 6 h time range.

All these methods require to estimate the so called “motion field”: the velocity field that models the advection of the precipitation systems. Comprehensive reviews of the motion field algorithms were presented by Reyniers (2008) and Liguori and Rico-Ramirez (2014). A possible procedure consists in using the wind as estimated by NWP models (Pierce et al. 2004). Other methods, instead, exclusively use a sequence of radar images. For instance, the “area tracking algorithm” divides a radar image into a grid of boxes. Each box is then forced to match with the box in the subsequent radar image following with which it has the maximum cross correlation (Rinehart and Garvey 1978). This procedure showed good results in widespread and persistent rain situations. Instead, for thunderstorms and convective scenarios the “cell tracking methods” are more efficient (Liguori and Rico-Ramirez 2014). These approaches first select an individual radar echo and store its characteristics (e.g., reflectivity, position, occupied area, and so on). These features are then used to detect similar patterns in the following images (Li et al. 1995).

Besides the obvious public and academic interest in weather forecasting, the importance of weather for economic and social activities has also led to interest from the private sector. An example is represented by Radarmeteo, a private company in the Veneto region of Italy that provides a variety of tailored meteorological services to business customers. One of their services consists exactly in providing short-term forecasts and derived products. Their current nowcasting algorithm consists in moving the radar echoes along the field obtained by the wind predicted by global forecast system (GFS). Radarmeteo was interested in improving this algorithm along two main lines. First, the company wanted to augment the usefulness of the nowcasts by reducing the current computation times, as about 40 min were required to provide forecasts for the next 3 h at a 20 min interval. Moreover, since the method tended to overestimate the speed of the precipitation systems giving spatial mismatching between the forecasted and observed rain fields, Radarmeteo was interested in evaluating more recent nowcasting techniques to improve the quality of their forecasts.

To accomplish these tasks we selected four motion field estimators: Lucas-Kanade (LK) (Lucas and Kanade 1981), Proesmans (Proesmans et al. 1994), Dynamic and Adaptive Radar Tracking of Storms (DARTS) (Ruzanski et al. 2011), and Variational

Echo Tracking (VET) (Germann and Zawadzki 2002). Among the available nowcasting algorithms we decided to test three deterministic methods and one ensemble approach: Extrapolation (Germann and Zawadzki 2002), Spectral Prognosis (S-PROG) (Seed 2003), Autoregressive Nowcasting Vertically Integrated Liquid (ANVIL) (Pulkkinen et al. 2020), and Short-Term Ensemble Prediction System (STEPS) (Bowler et al. 2006). The aim of this work is to compare these recent algorithms with the current proprietary method of the company to understand whether the quality of the forecast can be improved. The implementations for all these algorithms are available in the open source Pysteps library (Pulkkinen et al. 2019b), a Python framework for precipitation nowcasting. We developed an importer plugin to load the proprietary radar mosaic of the Italian Peninsula within this framework and further scripts to execute the experimental evaluation. For our experiments, 193 precipitation events were selected in the period from September 2019 to September 2020. The forecasts obtained from all possible combinations of motion field estimator and nowcasting method were compared with the proprietary nowcasting method by means of several categorical, continuous, and spatial metrics.

This work is organized as follows. The algorithms that will be considered in our experiments will be explained in Section II. We continue with our experimental setup in Section III, after which we will present the results in Section IV. Finally, we conclude the paper in Section V.

II Methods

In the following we both describe the nowcasting algorithms and the motion field estimators used in this work.

II-A Nowcasting Algorithms

The algorithms included in our experiments are a standard baseline, three deterministic methods, and one ensemble method. They can be described as follows.

1) *Eulerian persistence*: this method produces a prediction for the future that is identical to the current observation (Germann and Zawadzki 2002). Therefore, the forecasted reflectivity field Ψ for each successive time is obtained by replicating the observed reflectivity field, such that

$$\Psi(t + \tau, \mathbf{x}) = \Psi(t, \mathbf{x}) , \quad (1)$$

where τ is the prediction interval. This method is often included as a trivial baseline.

2) *Extrapolation*: this method is known as Lagrangian persistence since the intensity of the reflectivity field remains constant in the Lagrangian coordinate frame (Germann and Zawadzki 2002). The temporal evolution of precipitation is solely driven by advection by means of a precalculated velocity vector $\mathbf{v} = \boldsymbol{\alpha}/\tau$, where $\boldsymbol{\alpha}$ is the displacement vector. It follows that the forecast reflectivity field can be written as

$$\Psi(t + \tau, \mathbf{x} + \boldsymbol{\alpha}) = \Psi(t, \mathbf{x}) . \quad (2)$$

Obviously, this method does not allow for growth and dissipation processes.

3) *S-PROG*: this is an advection-based nowcasting method that models the dissipation processes by avoiding tracking small scale systems beyond their lifetime (Seed 2003). The rain field is represented as a hierarchical set of features (i.e., multifractal framework); since large features evolve more slowly than small features, the former are propagated longer in time than the latter. Formally, the logarithm of the radar reflectivity is decomposed into a spectral cascade via the Fast Fourier Transform (FFT), so that each level of this cascade represents a feature of the original field at a given scale. A band-pass filter is then applied to select the appropriate frequencies. This procedure limits the propagation of small systems in time, which leads to spatial smoothing in the forecasts. The temporal evolution of a system at the j^{th} cascade level is then modeled by means of a second order autoregression

$$\hat{\Psi}_j(t, \mathbf{x}) = \phi_{j,1} \hat{\Psi}_j(t - \tau, \mathbf{x}) + \phi_{j,2} \hat{\Psi}_j(t - 2\tau, \mathbf{x}) , \quad (3)$$

where $j \in \{1, \dots, N\}$ is one of the N cascade levels, $\hat{\Psi}$ represents the Fourier component of the reflectivity field at the j^{th} level normalized by using the mean and standard deviation of that level, and $\phi_{j,1}$ and $\phi_{j,2}$ are the estimated autoregressive coefficients.

4) *ANVIL*: this approach (Pulkkinen et al. 2020) combines two previous methods, namely RadVil (Boudevillain et al. 2006) and S-PROG (Seed 2003). The first algorithm solves a mass balance equation applied to Vertically Integrated Liquid (VIL), which is a superficial density, taking into account a source, S , and a dissipation, D , term:

$$\frac{d\text{VIL}}{dt} = \frac{\partial \text{VIL}}{\partial t} - u \frac{\partial \text{VIL}}{\partial x} - v \frac{\partial \text{VIL}}{\partial y} = S(t) - D(t) . \quad (4)$$

Using only the rain-rate information and under certain simplifying assumptions, the previous equation can be solved numerically. However, this approach does not

consider the predictability limits of some precipitation systems. ANVIL attempts to solve this issue by adopting the approach proposed by S-PROG. Therefore, VIL is decomposed into a cascade of multiple scales, and unnecessary frequencies are removed. As before, an autoregressive model of the second order is used for the temporal evolution of each cascade level.

5) *STEPS*: this is an ensemble method that can be considered as an extension of the previous S-PROG algorithm (Seed 2003; Bowler et al. 2006). To deal with the smoothing process that affects S-PROG with an increasing forecast time, Equation 3 has been modified to include a noise term at each of the j cascade levels:

$$\hat{\Psi}_j(t, \mathbf{x}) = \phi_{j,1} \hat{\Psi}_j(t - \tau, \mathbf{x}) + \phi_{j,2} \hat{\Psi}_j(t - 2\tau, \mathbf{x}) + \epsilon_j(t, \mathbf{x}) \\ \forall j \in \{1, \dots, N\} . \quad (5)$$

The stochastic term $\epsilon_j(t, \mathbf{x})$ represents the uncertainty associated with the development of the precipitation systems. By considering several realizations of the noise term, an ensemble of forecasts can be created.

II-B Motion Field Estimators

All previous nowcasting algorithms except for Eulerian persistence are advection-based methods. This means that they model the movement of a precipitation system after a time τ via its velocity field $\mathbf{v} = (u, v)$. A possible approach consists in simply using the forecast of the wind provided by NWP. On the other hand, some more recent methods estimate the motion field directly from a sequence of radar ‘‘images’’. The following methods are built upon this idea.

1) *LK*: under the assumption that the reflectivity remains constant during the motion, this optical flow technique (Lucas and Kanade 1981) describes the spatio-temporal intensity changes in terms of the velocity field as

$$I_x u + I_y v + I_t = 0 , \quad (6)$$

where I_x , I_y , and I_t represent the variation of the intensity with respect to space and time. This equation is underdetermined when considering a single pixel, but it can be made overdetermined by assuming that it holds for a neighborhood of pixels instead. A unique velocity field is then found via the least squares criterion.

2) *Proesmans*: this method solves Equation 6 with the addition of a smoothness constraint that makes the problem solvable without recurring to least squares (Proesmans et al. 1994). The estimated velocity field is used to find a correspondence between the pixels

of the first and second images. If important changes in brightness are registered, the field is corrected by taking into account the image gradients.

3) *DARTS*: this method calculates the flux of a precipitation system from a temporal sequence of radar images by solving the continuity equation in the frequency domain with an ordinary least squares (Ruzanski et al. 2011). The growth and decay terms are neglected in the equation.

4) *VET*: in this case the motion field is obtained by minimizing a cost function with two constraints: the residuals of the reflectivity conservation equation and a smoothing penalty function (Germann and Zawadzki 2002). The first requirement is used as a weak constraint, as reflectivity is not necessarily preserved in case of growth and dissipation processes.

III Experimental Setup

In the following we provide the details about the data employed in this work, the setting of the used algorithms, and the metrics selected for our experiments.

III-A Data

The radar mosaic used in our work aggregates the data from 26 radar stations located in the Italian territory and 14 radar stations in the neighboring nations. The radars in the Italian territory are operated by regional environmental protection agencies (Italian: Agenzia regionale per la protezione ambientale), the Italian Civil Protection Department, and the ENAV company, which is responsible for the provision of air traffic services. The foreign radars are operated by their respective national departments. Figure 1 shows the position of all radar stations. Where needed, the information provided by these agencies were converted in vertical maximum intensity (VMI) (i.e., the maximum reflectivity recorded in the column above each point of fixed longitude and latitude) and, lastly, merged in a shared reference system. The final mosaic extends from $6^{\circ}0'E$ and $48^{\circ}0'N$ to $19^{\circ}0'E$ and $36^{\circ}1'N$ with a resolution of 0.01° . The radar data are available at a time interval of 10 min.

We considered 193 precipitation events that were selected in the period from 15/09/2019 to 14/09/2020. For each hour of this time interval, we calculated the distribution of the precipitation cumulative referred to the previous three hours. If the 99th percentile of the distribution exceeds a threshold of 10 mm, the event is included in our analysis set. When two or more episodes fall in the same day we considered only



Fig. 1: Location of the 40 stations that provide the radar data to build the mosaic of the Italian peninsula.

the one with the highest cumulative rainfall. For each selected event we considered the data collected in the preceding 4 h at timesteps of 10 min. The data from the first hour were reserved for the estimation of the motion field. The first datum of the second hour is used in the experiments as “the last observed field”, whereas the final three hours leading up to the event were employed in the test phase to compare the observed rainfall with the forecast.

III-B Algorithms

To forecast the reflectivity field on the selected data we consider all the combinations between the four motion field estimators (i.e., LK, Proesmans, DARTS, and VET) and the four nowcasting algorithms (i.e., Extrapolation, S-PROG, ANVIL, and STEPS). For each event, the forecasts extend to the subsequent 3 h at an interval of 10 min. The implementation adopted for these methods is provided in the Pysteps library (Pulkkinen et al. 2019b).

The number of radar images required to build the motion field varies with the methods as follows:

- LK: current field + 2 previous fields;
- DARTS: current field + 5 previous fields;
- VET: current field + 2 previous fields;
- Proesmans: current field + 1 previous field.

For the STEPS nowcasting method we choose an ensemble of 20 members. The forecasts reported for

this algorithm in the following analysis come from the mean of the members.

The 16 combinations were compared with the Eulerian persistence baseline and with the proprietary algorithm of Radarmeteo. This is an extrapolation-based nowcasting method that extracts the local maxima from the input reflectivity field and other random points. This sub-sampled field is then moved in the direction of the wind as forecast by the GFS model. Lastly, an interpolation is made to reconstruct the final reflectivity field. Since the computation of this algorithm is time expensive, the predictions were calculated at steps of 20 min.

III-C Verification

Several metrics have been calculated to compare the skill of the nowcasting approaches, as done in related studies (Pulkkinen et al. 2020; Pulkkinen et al. 2019b; Mandapaka et al. 2012). Since a single validation technique does not provide a comprehensive picture of an algorithm’s performance, we employed three classes of verification metrics:

- continuous, which compare the entire spectrum of the reflectivity scale;
- categorical, which evaluate the ability to distinguish between different classes of precipitation;
- spatial, which evaluate the location of the precipitation.

For the continuous verification, we calculated the normalized mean squared error (NMSE):

$$\text{NMSE} = \frac{\sum_{i=1}^N (Z_i - \hat{Z}_i)^2}{\sum_{i=1}^N (Z_i + \hat{Z}_i)^2}, \quad (7)$$

where Z_i and \hat{Z}_i represent the reflectivity of the i^{th} pixel in the observed and forecast fields and N is the total number of pixels. We also evaluated the β coefficient, an estimation of the slope of the linear regression of the forecasts versus the observations, that is defined as:

$$\beta = \frac{s_{o,f}}{s_o^2}, \quad (8)$$

where $s_{o,f}$ is the covariance between the observed and forecast fields, and s_o^2 is the variance of the observed field. Ideally β should be equal to 1.

For the categorical verification, a reflectivity threshold Z_t is defined to make the problem binary. The pixels below this threshold ($Z < Z_t$) are classified as “non-rainy”, the pixels above the threshold ($Z \geq Z_t$) are considered as “rainy”. The possible outcomes when comparing such classification with the ground truth is

Observation	Forecast	
	$Z \geq Z_t$	$Z < Z_t$
$Z \geq Z_t$	hits (H)	misses (M)
$Z < Z_t$	false alarm (F)	correct negative (R)

TABLE 1: Contingency table in a binary classification problem.

reported in the contingency table (see Table 1). This table was then used to define the critical success index (CSI), probability of detection (POD), false alarm ratio (FAR), and false alarm rate (FA) as follows:

$$\begin{aligned} \text{CSI} &= H/(H + M + F), \\ \text{POD} &= H/(H + M), \\ \text{FAR} &= F/(H + F), \\ \text{FA} &= F/(R + F). \end{aligned} \quad (9)$$

The CSI is the ratio of correctly forecast rainy events over all the forecast rainy events plus misses, whereas the POD is the fraction of the rainy events that are correctly forecast. Both metrics range from 0 (i.e., no skill) to 1 (i.e., perfect skill). The FAR and FA indices are often confused: the former is the fraction of forecasts that are false alarms, the latter is the fraction of observed “non-rainy” events that are false alarms. Both ranges from 1 (i.e., no skill) to 0 (i.e., perfect skill).

The metrics presented to this point were all based on a pixel-to-pixel comparison between the forecast and observed fields. Since this approach is very restrictive, new indicators that take into account a neighborhood of pixels have been proposed in literature. In this study we evaluate the fractions skill score (FSS), a spatial verification score that compares the pixels inside a sliding window in the observed and forecast fields as follows:

$$\text{FSS} = 1 - \frac{\frac{1}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (P_o(i, j) - P_f(i, j))^2}{\frac{1}{n_x n_y} (\sum_{i=1}^{n_x} \sum_{j=1}^{n_y} P_o(i, j)^2 + \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} P_f(i, j)^2)}, \quad (10)$$

where n_x and n_y indicate the two sides of the reference area, and P_o and P_f are the fraction of pixels inside the area that exceeds a certain reflectivity threshold in the observation and forecast. This metric ranges from 0 (i.e., no skill) to 1 (i.e., perfect skill).

The reflectivity thresholds chosen to calculate the categorical and spatial metrics are 15 dBZ, 25 dBZ, and 35 dBZ. The sides of the reference square considered for FSS are 2 pixels, 8 pixels, and 32 pixels, where 1 pixel is equal to about 1 km.

IV Results

The evaluation of the computation time required to obtain the forecasts is one of the relevant parameters to take into account in nowcasting since it determines the usefulness of the forecast itself. Moreover, it is equally important to determine the relative strengths of a specific method. In the following both points are addressed.

IV-A Computation Time

Table 2 reports the mean, maximum, and minimum computation times required by each algorithm to produce 3 h of forecasts at an interval of 10 min. Due to hardware heterogeneity the reported numbers are indicative. In the case of the proprietary method only a rough estimation of the mean computation time was available and it refers to forecasts calculated at steps of 20 min. We observe that the computation time required by the new algorithms is one or two orders of magnitude less than the time employed by the proprietary method. This is a relevant result since one of the goal of this work was to augment the usefulness of the forecasts by shortening the computation time with respect to the proprietary algorithm. Even more important, we note that all methods but STEPS can be computed before a new radar image arrives. Among the proposed methods, Extrapolation is obviously computationally less expensive since it only moves the reflectivity field forward in time without performing any further calculation. On the other hand, STEPS, being an ensemble method, is the slowest since it requires the computation of 20 realizations. Focusing on the computation of the motion field estimators, we note that DARTS, LK, and Proesmans show similar calculation time, whereas VET is more computationally expensive.

IV-B Continuous Verification

To understand the influence of the motion field estimator on the forecast skill, we compare the trend of the NMSE and the β coefficient with respect to the lead time for each algorithm in Figure 2. The bold lines represent the mean skill of the algorithms, averaged over the events, and the 25th and 75th percentiles are highlighted by the dashed lines. As expected, for each algorithm, NMSE and β respectively increase and decrease with the lead time, as result of the natural degeneration of the forecast. Focusing on each single nowcasting method, both indices do not reveal relevant differences when different motion fields are employed. Since LK and VET consistently show slightly better

Name	Time [s]		
	Mean	Max	Min
Extrapolation DARTS	36.59	47.71	29.14
Extrapolation LK	35.52	43.77	25.19
Extrapolation Proesmans	45.24	48.72	42.46
Extrapolation VET	129.79	178.89	87.68
S-PROG DARTS	51.45	54.58	48.92
S-PROG LK	53.81	58.03	44.78
S-PROG Proesmans	69.53	75.87	65.58
S-PROG VET	256.74	343.66	174.46
ANVIL DARTS	71.55	78.46	67.43
ANVIL LK	74.41	82.81	65.52
ANVIL Proesmans	88.37	97.44	85.57
ANVIL VET	274.90	360.22	194.27
STEPS 20 DARTS	642.38	1546.69	413.32
STEPS 20 LK	625.35	1427.66	400.19
STEPS 20 Proesmans	653.93	1458.51	431.95
STEPS 20 VET	726.85	1559.81	514.29
proprietary	2400.00	–	–

TABLE 2: Mean, maximum, and minimum computation times required by each algorithm on the selected 193 events. For the proprietary algorithm only an estimation of the mean time was available.

performance than the other methods, in the following analyses we only report the results obtained when VET is used.

To compare the skill of the nowcasting methods, we report in Figure 3 the mean trend of the continuous metrics as function of the lead time. The proposed methods outperform the proprietary algorithm at each lead time. ANVIL shows a superior mean NMSE up to a lead time of 70 min, after which it is outperformed by S-PROG and STEPS. The performance of ANVIL continues to degrade with longer lead times; after 120 min it performs worse than the simple Extrapolation and after 160 min its NMSE is higher than that of the Eulerian baseline. The comparison of the β score is far simpler: ANVIL shows the best skill, followed by S-PROG, STEPS, and Extrapolation.

This analysis highlights that, at least for the first hour of forecast, the method that approximates the observed reflectivity field most accurately is ANVIL. However, the results of β highlight that all algorithms underestimates the reflectivity field. This can be a side effect of the experimental protocol: since the last observed field is located at the beginning of a precipitation event, it is challenging to correctly forecast the following growth.

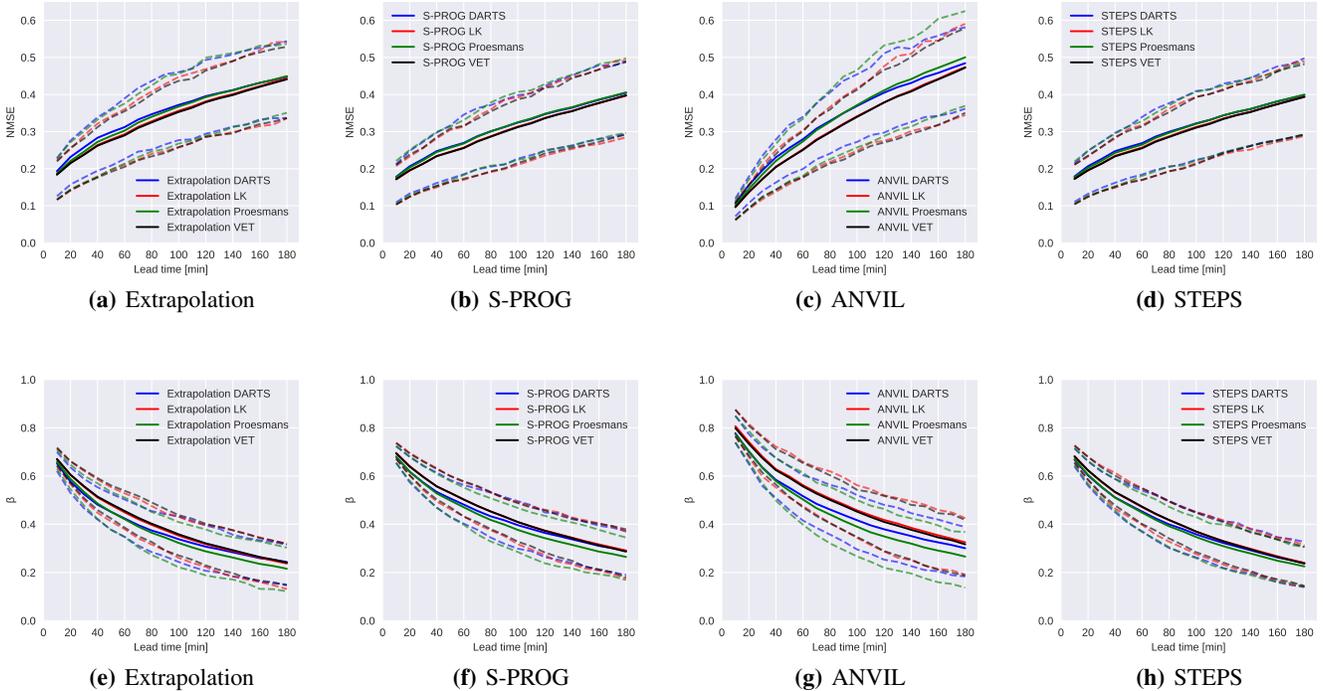


Fig. 2: normalized mean squared error (NMSE) (first row) and β (second row) obtained for (a) and (e) Extrapolation, (b) and (f) Spectral Prognosis (S-PROG), (c) and (g) Autoregressive Nowcasting Vertically Integrated Liquid (ANVIL), and (d) and (h) Short-Term Ensemble Prediction System (STEPS), each combined with the four available motion field methods. The continuous curves represents the mean of the scores over the 193 events, the dashed lines are the 25th and 75th percentiles.

Lastly, we note that although the approach followed by the proprietary method is conceptually similar to Extrapolation, the skills of these two algorithms are quite different. It is difficult to pinpoint the exact cause of this discrepancy, though a possible explanation may lie in the source of the motion field estimate. Indeed, it has been noted that image-based estimates can indirectly take orographic effects into account by empirically detecting a blockage of the precipitation systems from the sequence of radar images (Mandapaka et al. 2012).

IV-C Categorical Verification

For each reflectivity threshold mentioned in Section III-C we calculated the categorical metrics reported in Equation 9 for each event and each lead time. The performance of CSI, POD, and FAR, averaged over the events, is reported in Figure 4 for each nowcasting algorithm as a function of the lead time. The forecast skill of each algorithm decreases when either the lead time or the reflectivity threshold increase (Golding 1998). The former observation is attributed to the chaotic nature

of such systems, which naturally brings to a forecast worsening. The latter consideration depends on the fact that the areas that exceed a higher threshold are more localized.

Focusing on the single methods, the skill of the simple Extrapolation approach is generally better than the Eulerian baseline, though they perform similarly for high lead time or reflectivity thresholds. At 15 dBZ ANVIL shows a CSI that is only slightly better than the Eulerian baseline, thus indicating either a high quantity of missed events and/or false alarms. Both options are possible, since at 15 dBZ we observe a FAR and POD respectively higher and lower than the other methods, except for the proprietary algorithm. The high FAR is a symptom of the presence of false alarms, whereas a low POD highlights missed events. The situation improves for higher reflectivity thresholds where ANVIL starts to outperform the other methods showing better CSI and POD values, although we still register false alarms. On the other hand, at 15 dBZ and 25 dBZ the CSI of S-PROG and STEPS is similar at least up to a lead time

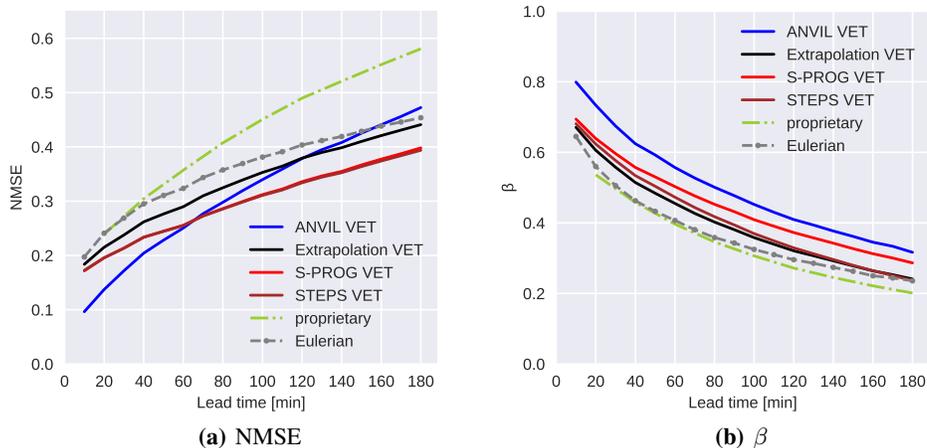


Fig. 3: For each nowcasting method the mean (a) normalized mean squared error (NMSE) and (b) β metrics, averaged over the 193 events, are reported.

of 40 min, later S-PROG keeps a better skill. However, looking at the other indices, STEPS reports a higher number of missed events, but also less false alarms than S-PROG. As the reflectivity threshold augments, both algorithms have more difficulty in the recognition of the precipitation systems.

We can reach similar conclusions by analyzing the behavior of POD versus FA reported in Figure 5 where we consider 20 min, 40 min, 60 min, and 120 min as lead time. Each point represents the mean value over the events and the horizontal and vertical lines depict the 25th and 75th percentiles for both indices. At 15 dBZ the best POD and FA are obtained by respectively S-PROG and STEPS. However, the inter-method variability is very low at least up to 60 min. The only exception is represented by the proprietary method, which consistently shows the highest quantity of missed events. The situation is similar at 25 dBZ, but the behavior of ANVIL and S-PROG becomes quite comparable. At 35 dBZ, instead, ANVIL shows the best POD, but also the highest FA.

Summarizing, on the basis of these categorical scores ANVIL outperforms the other methods at higher reflectivity thresholds at the cost of a higher number of false alarms. On the other hand, the light rain events are better recognized by S-PROG and STEPS. Despite these differences, the new methods generally outperform the proprietary algorithm.

IV-D Spatial Verification

For each reflectivity threshold and size of the spatial neighborhood (see Section III-C), we report in Figure 6 the mean FSS metric for each nowcasting algorithm, averaged over the events, as function of the lead time. As expected, the FSS consistently increases with the size of the neighborhood and decreases with the reflectivity threshold and the lead time.

At 15 dBZ and 25 dBZ the proposed methods always outperform the proprietary algorithm and the Eulerian baseline, at least up to 120 min. In line with the previous categorical analyses, at 15 dBZ S-PROG shows the best skill. However, as the reflectivity threshold increases, its performance decays: at 25 dBZ it is first comparable (at 2 pixels and 8 pixels) and later worse (at 32 pixels) than the ANVIL and Extrapolation methods, while at 35 dBZ it is outperformed by the Eulerian baseline. Also STEPS follows a similar trend, however it performs below the Eulerian baseline even at 25 dBZ up to a lead time of 120 min. On the other hand, ANVIL shows an increasing skill as the reflectivity threshold augments outperforming the other methods at 35 dBZ.

The good performance of Extrapolation at 32 pixels can be attributed to the fact that simple advection is sufficient to model the field at big scale. Instead, at smaller scales, the terms of growth and decay, which we found in the other algorithms, become more relevant to model the precipitation systems. A possible explanation for the loss of skill of S-PROG and STEPS at 35 dBZ was hinted at previously; for this threshold the precipitation areas are localized and, by nature, these

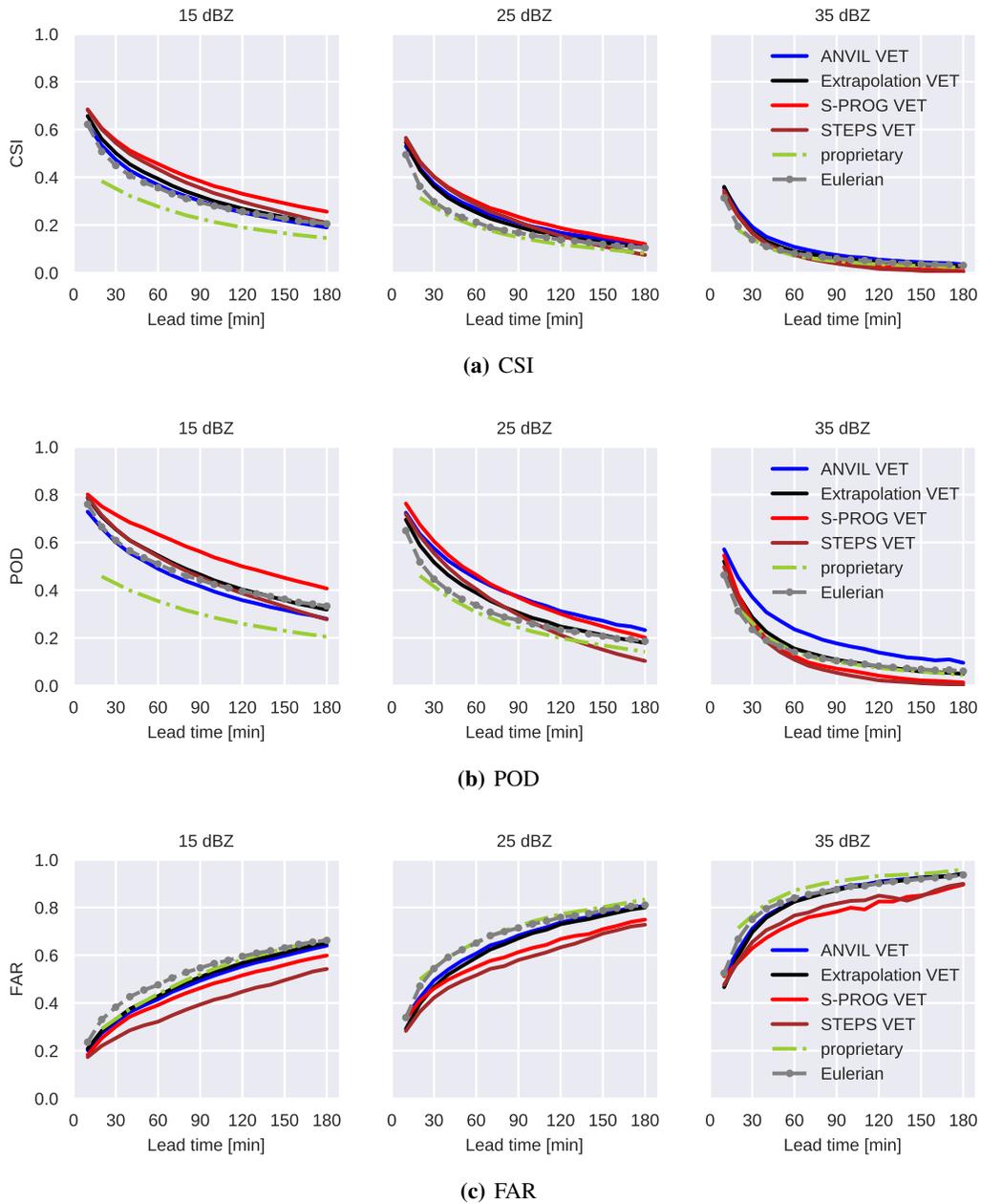


Fig. 4: For each nowcasting method, the mean (a) critical success index (CSI), (b) probability of detection (POD), and (c) false alarm ratio (FAR) metrics, averaged over the 193 events, are reported.

algorithms tend to cut these “fine details” as the lead time increases.

V Conclusions

The aim of this work was to improve the nowcasting product of Radarmeteo, an Italian company that provides meteorological services, by identifying possible algorithms that can 1) augment the usefulness of the forecasts by shortening the computation times, and 2) improve the quality of the forecasts. To address

these challenges we compared four popular algorithms with the proprietary method currently in use and a straightforward baseline method. Since the proposed algorithms are advection-based, we also selected four motion field estimators, and we combined them with the four nowcasting approaches. Both for the motion field estimators and the nowcasting algorithms we adopted the implementation provided in Pysteps, an open source library for nowcasting. For the tests, we selected 193 interesting events in the last year from the

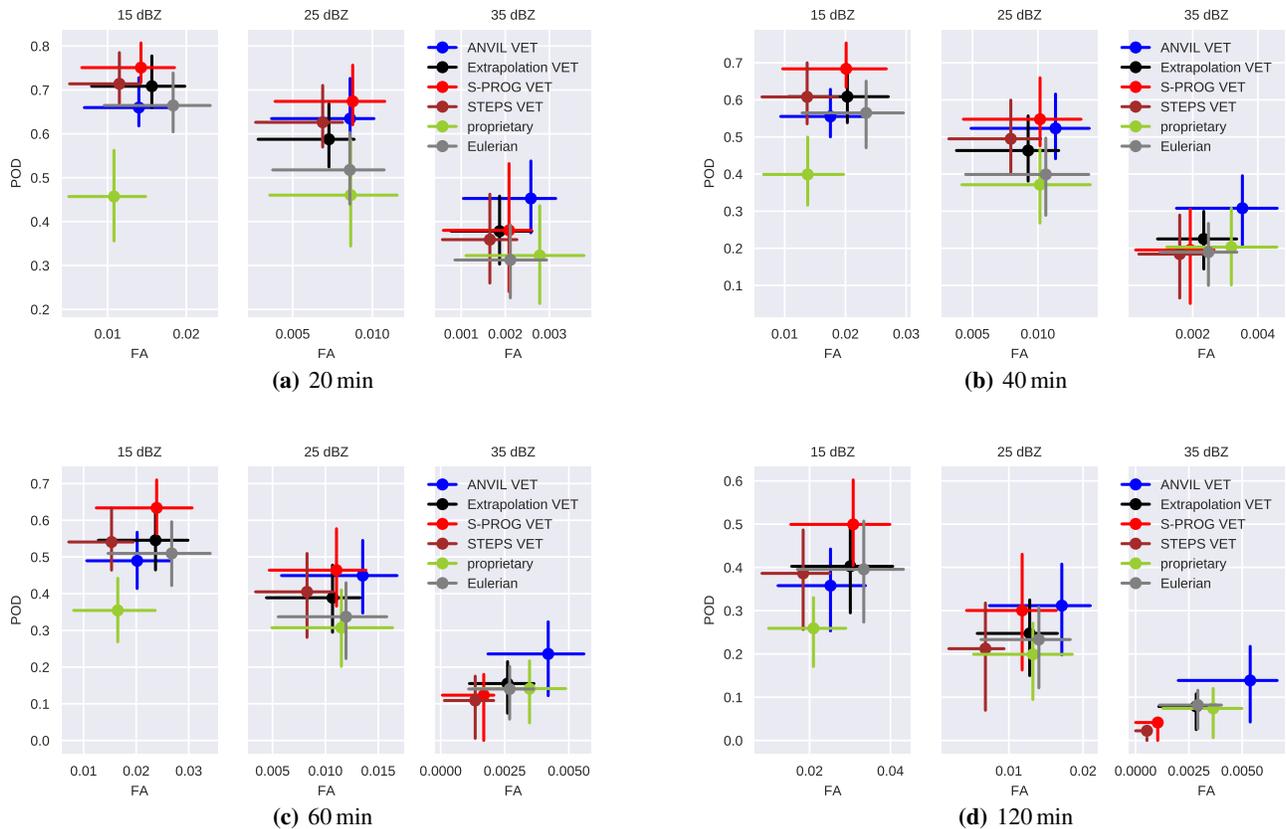


Fig. 5: For each nowcasting method we report the behavior of probability of detection (POD) versus false alarm rate (FA) at (a) 20 min, (b) 40 min, (c) 60 min, and (d) 120 min as lead time. Each point represents the mean value, averaged over the 193 events, and the horizontal and vertical lines depict the 25th and 75th percentiles for both indices.

Italian radar mosaic; for each event we calculated 3 h of forecasts at steps of 10 min.

To compare the skill of the algorithms we adopted several verification metrics: continuous, to perform a full-spectrum comparison between the intensities of observed and forecast fields; categorical, to evaluate the ability of the methods to distinguish between different classes of precipitation; and spatial, to estimate the precision of the location of the precipitation. On the basis of these evaluations we concluded that all proposed algorithms represent an improvement with respect to the existing proprietary method from different perspectives: better representativeness of the reflectivity, a reduction of the number of missed events and false alarms, and a gain in the spatial accuracy of the forecasts. S-PROG and STEPS showed good skills, both from a categorical and a spatial perspective, for events characterized by low reflectivity fields, whereas ANVIL performed better at higher reflectivity, although this comes at the cost

of an increase in false alarms. Lastly from the results of the continuous metrics it emerges that, at least for the first 70 min, the methods that better approximate the observed reflectivity field is ANVIL. Instead, the choice of the motion field estimator did not appear to influence the overall performance of the nowcasting methods. Finally, the chosen implementation of the tested methods allow to reduce the computation times by one or two orders of magnitude.

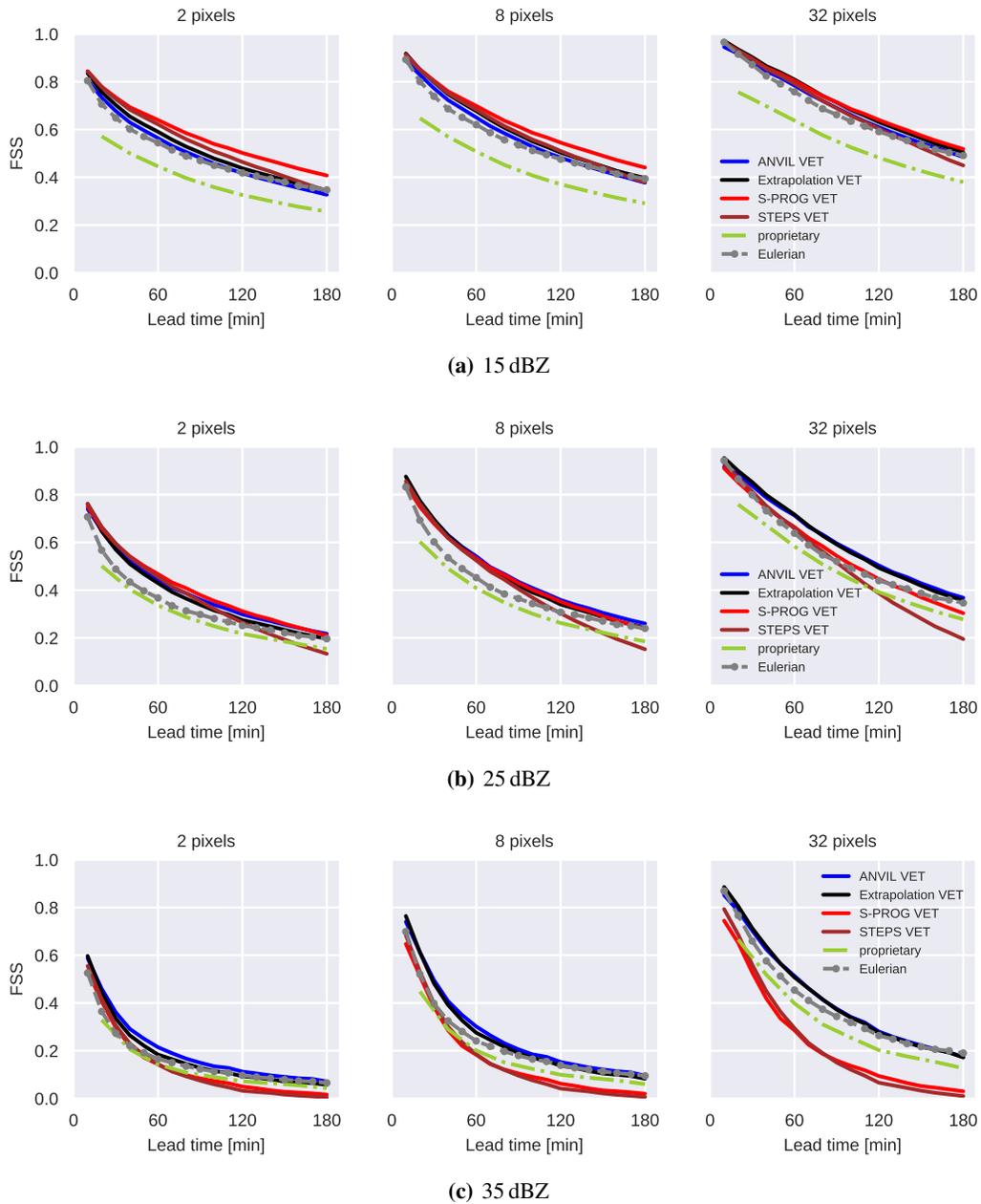


Fig. 6: Mean fractions skill score (FSS), averaged over the 193 events, calculated for each nowcasting method by setting as threshold (a) 15 dBZ, (b) 25 dBZ, and (c) 35 dBZ. In each case three spatial extensions have been considered.

References

- [1] Brice Boudevillain, Hervé Andrieu, and Nadine Chaumerliac. “Evaluation of RadVil, a radar-based very short-term rainfall forecasting model”. In: *Journal of Hydrometeorology* 7.1 (2006), pp. 178–189.
- [2] Neill E. Bowler, Clive E. Pierce, and Alan W. Seed. “STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP”. In: *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography* 132.620 (2006), pp. 2127–2155.
- [3] Urs Germann and Isztar Zawadzki. “Scale-dependence of the predictability of precipitation from continental radar images. Part I: Descrip-

- tion of the methodology”. In: *Monthly Weather Review* 130.12 (2002), pp. 2859–2873.
- [4] B. W. Golding. “Nimrod: A system for generating automated very short range forecasts”. In: *Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling* 5.1 (1998), pp. 1–16.
- [5] L. Li, W. Schmid, and J. Joss. “Nowcasting of motion and growth of precipitation with radar over a complex orography”. In: *Journal of applied meteorology* 34.6 (1995), pp. 1286–1300.
- [6] Sara Liguori and Miguel Angel Rico-Ramirez. “Quantitative assessment of short-term rainfall forecasts from radar nowcasts and MM5 forecasts”. In: *Hydrological Processes* 26.25 (2012), pp. 3842–3857.
- [7] Sara Liguori and Miguel Angel Rico-Ramirez. “A review of current approaches to radar-based quantitative precipitation forecasts”. In: *International Journal of River Basin Management* 12.4 (2014), pp. 391–402.
- [8] S. Lovejoy and D. Schertzer. “Multifractals and rain”. In: *New uncertainty concepts in hydrology and water resources* 270 (1995), pp. 61–103.
- [9] Bruce D. Lucas and Takeo Kanade. “An iterative image registration technique with an application to stereo vision”. In: (1981), 674–679.
- [10] Pradeep V. Mandapaka, Urs Germann, Luca Panziera, and Alessandro Hering. “Can Lagrangian extrapolation of radar fields be used for precipitation nowcasting over complex Alpine orography?” In: *Weather and Forecasting* 27.1 (2012), pp. 28–49.
- [11] C. E. Pierce, E. Ebert, A. W. Seed, M. Sleight, C. G. Collier, N. I. Fox, N. Donaldson, J. W. Wilson, R. Roberts, and C. K. Mueller. “The nowcasting of precipitation during Sydney 2000: an appraisal of the QPF algorithms”. In: *Weather and Forecasting* 19.1 (2004), pp. 7–21.
- [12] Marc Proesmans, Luc Van Gool, Eric Pauwels, and André Oosterlinck. “Determination of optical flow and its discontinuities using non-linear diffusion”. In: *European Conference on Computer Vision*. Springer. 1994, pp. 294–304.
- [13] Seppo Pulkkinen, V. Chandrasekar, and Ari-Matti Harri. “Stochastic Spectral Method for Radar-Based Probabilistic Precipitation Nowcasting”. In: *Journal of Atmospheric and Oceanic Technology* 36.6 (2019), pp. 971–985.
- [14] Seppo Pulkkinen, V. Chandrasekar, Annakaisa von Lerber, and Ari-Matti Harri. “Nowcasting of Convective Rainfall Using Volumetric Radar Observations”. In: *IEEE Transactions on Geoscience and Remote Sensing* (2020).
- [15] Seppo Pulkkinen, Daniele Nerini, Andrés A Pérez Hortal, Carlos Velasco-Forero, Alan Seed, Urs Germann, and Loris Foresti. “Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1. 0)”. In: *Geoscientific Model Development* 12.10 (2019), pp. 4185–4219.
- [16] Maarten Reyniers. *Quantitative precipitation forecasts based on radar observations: Principles, algorithms and operational systems*. Institut Royal Météorologique de Belgique Brussel, Belgium, 2008.
- [17] R. E. Rinehart and E. T. Garvey. “Three-dimensional storm motion detection by conventional weather radar”. In: *Nature* 273.5660 (1978), pp. 287–289.
- [18] Evan Ruzanski and V. Chandrasekar. “Scale filtering for improved nowcasting performance in a high-resolution X-band radar network”. In: *IEEE transactions on geoscience and remote sensing* 49.6 (2011), pp. 2296–2307.
- [19] Evan Ruzanski, V. Chandrasekar, and Yanting Wang. “The CASA Nowcasting System”. In: *Journal of Atmospheric and Oceanic Technology* 28.5 (May 2011), pp. 640–655.
- [20] D. Schertzer, S. Lovejoy, F. Schmitt, Y. Chigirinskaya, and D. Marsan. “Multifractal cascade dynamics and turbulent intermittency”. In: *Fractals* 5.03 (1997), pp. 427–471.
- [21] A. W. Seed. “A dynamic and spatial scaling approach to advection forecasting”. In: *Journal of Applied Meteorology* 42.3 (2003), pp. 381–388.
- [22] David Simonin, Clive Pierce, Nigel Roberts, Susan P. Ballard, and Zhihong Li. “Performance of Met Office hourly cycling NWP-based nowcasting for precipitation forecasts”. In: *Quarterly Journal of the Royal Meteorological Society* 143.708 (2017), pp. 2862–2873.
- [23] B. J. Turner, I. Zawadzki, and U. Germann. “Predictability of Precipitation from Continental Radar Images. Part III: Operational Nowcasting Implementation (MAPLE)”. In: *Journal of Applied Meteorology* 43.2 (Feb. 2004), pp. 231–248.