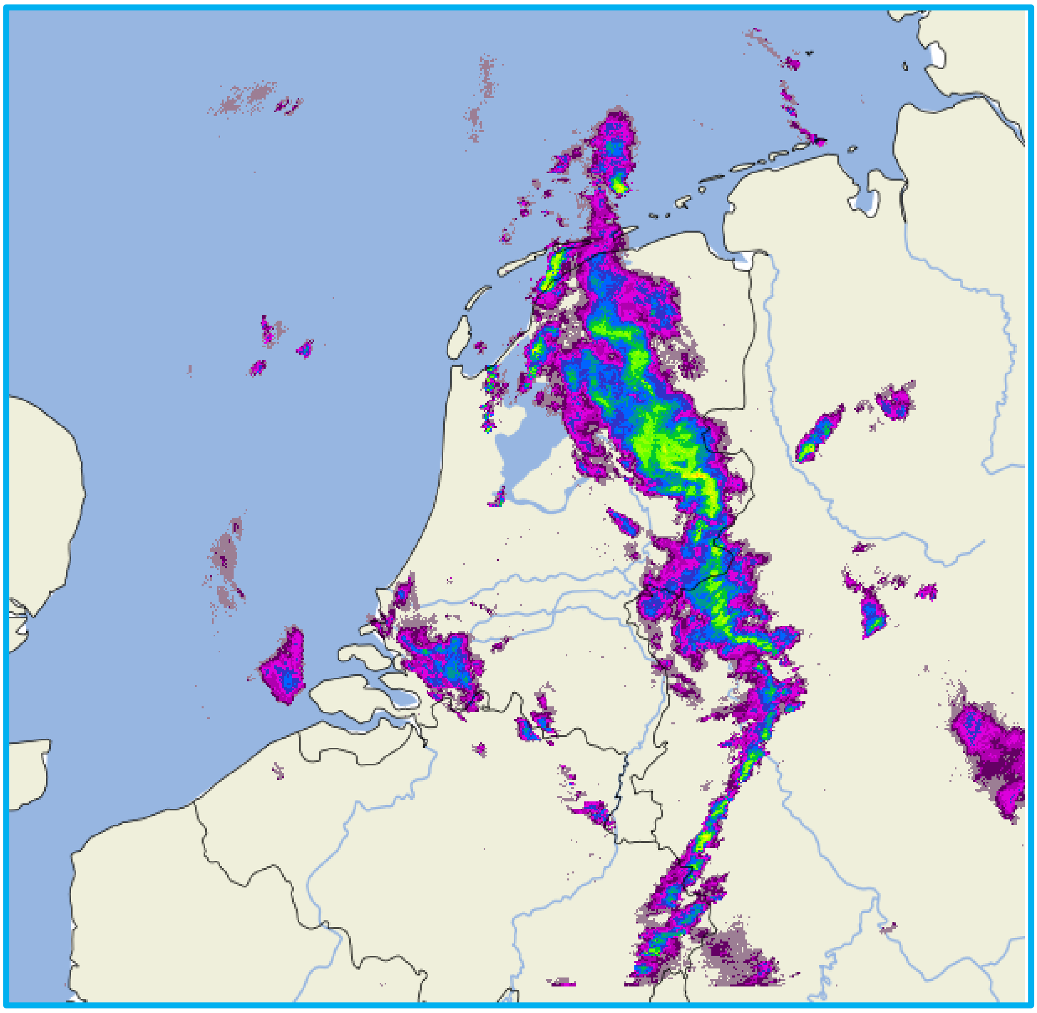
Improving Precipitation Forecasts with DGMR and KNMI's Harmonie using Linear Blending



Kevin Luo (2858975)
Supervisors: Arno Siebes (UU) & Joep Bosdijk (Weather Impact)
Second examinator: S. Mehrkanoon

Master Applied Data Science
June 28th, 2024

Weather Impact

Utrecht University

Abstract

Extreme precipitation events can lead to floodings, causing causalities and damage. Early flood warnings and preventative measures are essential. Precipitation forecasting can be done by nowcasting, which is accurate for short-term precipitation forecasts (lead times up to 2 hours). Weather Impact uses DGMR as their nowcasting model, which follows a deep learning approach. Numerical Weather Prediction models are used for long-term precipitation forecasts but require around 6 hours to generate accurate forecasts. This results in a time gab where no accurate precipitation forecasts can be generated, a crucial moment regarding making decisions for preventative measures. The focus of this thesis project is to determine the potential of linear blending with the DGMR output and KNMI's Harmonie data to close the time gab. This includes finding the optimal start and end time for the linear blending, using a training set containing data from May 24 to June 11, 2024. The optimal start and end time are 135 min and 465 min, respectively. These times are used on the test set, which covers May 15th and 16th, and June 18th, 2024. The quality of the precipitation forecast is evaluated based on a combination of the skill scores: probability of detection, false alarm rate and fractional skill score. Compared to the DGMR and Harmonie, linear blending showed some improvements for certain lead times and precipitation rate thresholds.

Contents

Abstract *2*

Contents 3

1. Introduction *4*

Research objective 5

Report structure 5

2. Background 6

KNMI radar data 6

Z-R relationship: from radar reflectivity to rainfall rate 7

DGMR 7

Lagrangian persistence 7

Linear blending and saliency 8

Evaluation metrics 8

3. Data *11*

Radar composites knmi data 11

Harmonie cy40 knmi data 11

Ethical and legal considerations 11

4. Methods *12*

General 12

Data loading 12

Data pre-processing 12

Event selection 13

Testing 14

5. Results and Discussion *15*

6. Conclusion *21*

7. Future research *22*

List of references *23*

Appendix I *27*

1. Introduction

Extreme precipitation events can cause disruptive floods. An example is extreme rainfall amounts in July 2021 that was caused by the mesoscale low-pressure system in northwestern and central Europe. This led to severe flooding, causing over 240 casualties and more than 25 billion USD in damage (AON, 2021; Imhoff et al., 2023; Koks et al., 2022; Kreienkamp et al., 202). These disruptive of floodings can be minimised if there is a timely anticipation of the flood, which is possible when there is a flood early warning system in place that is able to produce rainfall forecasts that are reliable and available within an acceptable timeframe (Imhoff et al., 2023).

Rainfall forecasts can be generated from nowcasting and numerical weather prediction (NWP) models. Traditional radar-based nowcasting methods use the latest radar observation to forecast into the future by statistically and heuristically extrapolation. These methods are more effective in shorter time spans. The nowcast methods have a higher predictive ability than NWP models over the initial hours due to the usage of high spatial and temporal resolutions of remotely sensed data. However, for longer timespans (beyond 2 hours in the future), the performance of radar-based nowcasts decreases due to the imperfections to determine the precipitation advancement (growth and decay) and uncertainty in assessing the storm motion fields (Imhoff et al., 2023; Turner et al., 2004).

At Weather Impact, Google's Deep Generative Model of Rain (DGMR) is used as nowcasting model. Unlike traditional nowcasting methods, which follows the same approach as numerical weather prediction (NWP) models by modelling precipitation through advection equations, DGMR uses a different methodology. Traditional nowcasting models use optical flow to estimate motion fields to estimate an advection forecast. In contrast, DGMR uses deep learning methods, going beyond the reliance on the advection equations. DGMR predicts precipitation rates at each grid location by training on extensive radar observations, which allows the DGMR to model complex, non-linear precipitation phenomena more effectively than traditional nowcasting models (Ravuri et al., 2021).

Weather radar data is suitable as input data for precipitation nowcasting due to its high spatial and temporal resolution. Despite recent advances in numerical weather prediction (NWP), nowcasting remains the primary approach, which typically outperforms NWP forecasts for the first few hours (around 2-5 hours), depending on the weather conditions, geographical location and NWP characteristics (Pulkkinen et al., 2019).

In NWP models, the atmosphere is divided into grid cells. For each grid cell, information like temperature, air pressure, humidity, wind, radiation etc. are kept track. These values are constantly changing, which are determined by solving complex mathematical equations about physical processes in the atmosphere. For a forecast, the initial values of the variables are determined for each grid cell from observations from weather satellites, ground stations, weather balloons or other sources (KNMI, n.d.). Nowadays, weather prediction has been transformed through the implementation of ensemble forecasts, which comprises multiple runs of NWP models. In each run, different initial conditions and/or equations that represents the atmosphere processes are used. This addresses the major source of forecast uncertainty (Gneiting & Raftery, 2005). However, NWP models are limited because of the model spin-up time and long process to forecast, which makes the forecasts less reliable in the initial hours. Most NWP models need several hours between model initialization and delivery at the end user. In addition, NWP models run on a too coarse temporal resolution for usage in flood early warning systems (Lin et al., 2005; Radhakrishnan & Chandrasekar, 2020; Roberts and Lean, 2008; Berenguer et al., 2012).

Blending combines the nowcast and NWP forecasts, which the blending output have the advantage of both techniques and can extend the nowcasting lead time with reliable forecasts (Radhakrishnan & Chandrasekar, 2020). This thesis project focuses on extending the nowcast lead time by blending the DGMR nowcasting and the NWP model forecasts based on precipitation events in The Netherlands, using linear blending.

Research objective

As mentioned before, extreme precipitation events can have big consequences. To minimise the damage, early warnings should be issued. Nowcasting models, like DGMR, can be used to forecast precipitation in the future but these models can only generate accurate forecasts up to 2 hours ahead. NWP models can generate accurate forecasts beyond the 2 hours maximum, however, NWP models need quite some time to be able to generate accurate forecasts. This means that the forecasts are not accurate for the first few hours. Blending should enable to combine these two models to fill in the gab. Weather Impact wants to know whether blending has any potential to improve their forecasts from the DGMR model. Therefore, the goal of this research is to find the answer the following research question:

How does linear blending of nowcast (DGMR) and NWP (KNMI Harmonie Cy40) impact the accuracy and reliability of precipitation forecasts beyond the lead time of 2 hours?

Report structure

The outline of this thesis report starts with an introduction, which introduces the reader to the topic by outlining the problem statement and the research objective. Following the introduction, an overview of relevant background information is described. This chapter allows the reader to gain a deeper understanding of the subject matter (Chapter 2). Chapter 3 focuses on the data description. This includes information on the data sources and its relevance to this thesis project. An overview of the research process is described in Chapter 4. Here, the methods that are applied are described in detail, along with justifications for their use based on the research objective. The results of these methods are presented in Chapter 5. Then the results section is followed by Chapter 6, which presents the answer to the research question. At last, the thesis report is concluded with Chapter 7, addressing some limitations and recommendations for future research.

2. Background

KNMI radar data

The radar reflectivity information is obtained using rotating radar antenna on a tower in Herwijnen and the building of the Koninklijke Marine in Den Helder (see figure 1). The radar antenna emits pulses contiguously that are reflected by droplets that are encountered by the signal in the air (see figure 2). The antenna acts as both the transmitter and receiver. The signal that is reflected provides information about the precipitation and its geographical distribution over the area (KNMI, n.d.-b).

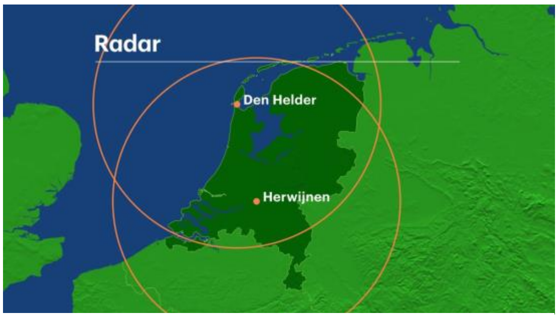


Figure 1. Location of the radars of the KNMI. From Buienradar (n.d.).

Figure 2. Radar antenna in Herwijnen. From: KNMI (n.d.-b.).

The radar data is used to estimate the movement of the precipitation in the next few hours by simply extrapolating the precipitation area on the map over time. This method works best for fronts and precipitation areas. The method is less accurate for developing showers, due the possibility of changes in the intensity. In some cases, developing showers dissipate as they move further inland, become larger, or new showers are developing. These sudden changes are often the cause for dangerous situations. Meteorologists use a large amount of weather data and model calculations to issue timely warnings for such cases (KNMI, n.d.-b).

The use of radar reflectivity data for precipitation forecasting has its limitations. For instance, the radar's range is limited, leading to somewhat distorted radar images at the edges of the map (greater distance from the radar antennas). Due to the curvature of the Earth, pulses at greater distances, like the areas around Northern France and Eastern England, are the reflected pulses of precipitation are at higher altitudes, around 6000 to 8000 meters in the atmosphere (instead of 1500 m) (KNMI, n.d.-b). See Figure 3.

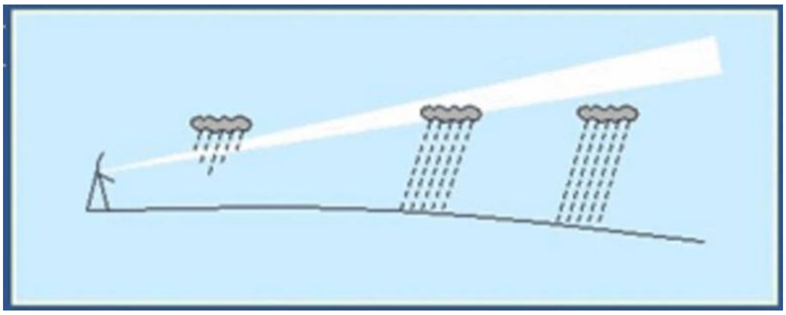


Figure 3. Visualization of the effect of the Earth's curvature on the radar images. From KNMI (n.d.-b).

Another limitation associated with the use of radar is that not all reflected pulses originate from precipitation. Under certain conditions, the pulses can follow a trajectory that is more curved than the curvature of the earth. This leads to situations where at certain distances from the radar, the pulses will hit the ground. This causes noise on the radar image (KNMI, n.d.-b).

Some radar echoes can be deflected by the atmosphere and reach the sea surface. This is called sea clutter. Under these conditions, sea waves can cause for reflected beams that are detected by the radar. This effect is also known as abnormal propagation and occurs when a temperature inversion is present (KNMI, n.d.-b). A temperature inversion is an abnormal behaviour of temperature in the troposphere, in which a layer of cool air at the surface is overlain by a layer of warmer air (North Dakota State University, 2018). In addition, there is ground clutter which are deflected radar echoes caused by plants, trees, buildings or other man-made structures (Hubbert et al., 2009).

Z-R relationship: from radar reflectivity to rainfall rate

Before the use of radar, rain gages were used to measure rainfall rates. However, a high ratio of rain gages per area was needed to assess rainfall rates accurately. In scattered convective situations, heavy rainfall can occur over parts of a river basin, and it is possible that these rainfalls are not captured by the few rain gages. This can lead to inaccurate forecasts.

These problems have been solved by advancement in radar technology, which measures reflectivity of rain droplets in the air. A good relationship between radar reflectivity (Z) and precipitation rate (R) was introduced. Z depends on the raindrop size distribution and the drop size (to the sixth power). R depends on the raindrop size distribution, the size of the drops (to the third power) and the fall velocity for a given drop size. The research on the Z-R relationship was continued, on the size distribution of raindrops as a function of rainfall rate. From this research, the following Z-R relationship is derived (Fournier, 1999):

$$Z = 200 \times \; R^{1.6} \tag{2-1}$$

Z = reflectivity factor in dBZ
R = rainfall rate in mm/h

DGMR

The DGMR algorithm is a conditional generative model that predicts N future radar fields using M past radar fields. DGMR uses radar observations of the previous 20 minutes (four consecutive observations) as input for a generator that allows for multiple precipitation forecasts. The output of DGMR are 18 frames (90 min ahead). The algorithm uses two loss functions and a regularization term to adjust parameters by comparing real radar observations to the forecasts generated by the algorithm. Furthermore, DGMR is trained on a large amount of data from precipitation events, which are extracted from the radar stream (Ravuri et al., 2021).

The first loss function is defined by a spatial discriminator, which is a convolutional neural network that attempts to distinguish real radar fields from generated fields. This ensures for spatial consistency and discouraging blurry forecasts. The second loss function is defined by a temporal discriminator, which is a three-dimensional convolutional neural network that attempts to distinguish observed and generated radar sequences, ensuring temporal consistency and penalizes jumpy forecasts. To improve accuracy, a regularization term is introduced that penalizes deviations at the grid cell resolution between the real radar sequences and the algorithms predictive mean, which is computed with multiple samples. The regularization term is essential to ensure location-accurate forecasts and improve performance (Ravuri et al., 2021).

Ravuri et al. (2021) also compared DGMR with Pysteps, which is a widely used precipitation nowcasting model. In this study, an event in eastern Scotland with intense showers over land was tested. It was seen that Pysteps had trouble maintaining these showers and overestimated the rainfall intensity over time, which was not observed in real life and did not cover the spatial extent of the rainfall sufficiently. By comparison, DGMR was able to produce more accurate forecasts. However, DGMR produced less accurate precipitation rates at lead times +90 min and at the edges of the radar range. In addition, the output of the models was assessed by meteorologists and the majority preferred the DGMR forecasts.

Lagrangian persistence

Traditional nowcasting models are extrapolation based, meaning that it assumes that over time the changes in precipitation can be captured by moving the radar echoes along a stationary motion field without any changes in intensity. This is known as Lagrangian persistence. For radar-based precipitation nowcasting, predictive uncertainty comes from errors in the estimation of the current state of the rainfall and motion fields (initial state errors) and limitations of the Lagrangian persistence as a model to predict the change in rainfall and motion fields (model errors) (Pulkkinen et al., 2019).

In the Lagrangian assumption, the main cause for model errors comes from the changes in precipitation in terms of initiation, growth, decay and termination processes that violate the steady-state assumption (Pulkkinen et al., 2019).

Most of the weather systems assume constant intensity in time. In reality, except for extensive and persistent systems, typical time scale for the growth and dissipation of rain fall is just about an hour, which makes the Lagrangian assumption invalid and therefore, the nowcasting model is not able to produce useful forecasts beyond 2 or 3 hours (Cheung et al., 2015).

Linear blending and saliency

From various methods to blend nowcasting with NWP model, linear blending is one of them. Linear blending essentially uses a weighted sum of the nowcast and the NWP model. The weighting function is determined by a linear function:

$$weight_{nwp} = \frac{t - start_{blending}}{end_{blending} - start_{blending}} \tag{2-2}$$

$$weight_{nowcast} = 1.0 - weight_{nwp} \tag{2-3}$$

Pysteps' linear blending uses a predefined start and end time for the blending. Before this start time, the resulting blended forecasts only uses the nowcasts. After the end time, the resulting blended forecasts only consist of the NWP forecasts. Between the start and end time, the weight for nowcast decreases from 1 to 0 linearly (equation 2-3), while the weight for NWP increases from 0 to 1 linearly (equation 2-2). The weights are then applied to every grid cell of every timestep (Pysteps, n.d.).

The linear blending can be further improved by implementing saliency, which is the intensity of each pixel. This method is based on a technique for merging multiple images. The saliency-based linear blending method preserves the grid cell intensities if they are high enough based on their ranked salience. Linear blending without saliency has a problem of producing underestimated blended values during the middle of the forecast. This problem can be minimised by using saliency. The salient-based linear blending determines two-dimensional weights based on the differences between normalized intensities from the nowcast and NWP forecast. These weights are determined for each forecast hour (Hwang et al., 2015).

The novelty of this method is the ability to preserve the values in both the nowcast and NWP forecasts if they exceed a certain threshold. For instance, if there are two convective cells in both nowcast and NWP forecasts in the middle of the forecast lead time, salient based linear blending tends to shrink the cells by applying different weights (higher weights for higher values and lower weights for lower values, which leads to preserving most of the high-valued pixels and eliminates many of the low-valued pixels) while regular linear blending cuts every value in half. Salient-based blending shows better results than regular linear blending (Hwang et al., 2015).

Evaluation metrics

The forecasting performance can be assessed using a contingency table that shows the number of forecasted event and the actual occurrences of the event. In this case, the event is precipitation that exceeds a certain threshold. This table creates four combinations of forecasts: events that was forecast and occurred (hit); events that were not forecast but occurred (miss); events that were forecast but did not occur (false alarm); and events that were not forecast and did not occur (correct negatives) (Table 1). Using this contingency table, various skill scores can be determined, such as the probability of detection (POD) and false alarm ratio (FAR).

Table 1: Contingency table for rainfall. From: Xue et al. (2017)

		Rainfall observed	
		Yes	No
Rainfall forecast	Yes	Hit	False alarm
	No	Miss	Correct negative

POD

The POD is a verification measure to assess the categorical forecast performance, which is the ratio between the forecast hits and number of observed events. The POD ranges from 0 to 1, with 1 as the perfect POD (Boneh et al. 2015). The POD is calculated as follows:

$$POD = \frac{Hits}{Hits + Miss} \tag{2-4}$$

FAR

The FAR is another verification measure to assess the categorical forecast performance, but the FAR is the ratio between the number of false alarms and number of forecast events. The FAR also ranges from 0 to 1 but 0 being the perfect FAR (Boneh et al. 2015). The FAR is determined as follows:

$$FAR = \frac{False\ alarms}{Hits + False\ alarms} \tag{2-5}$$

FSS

The fractions skill score (FSS) is a neighbourhood verification method to verify deterministic forecasts of binary events. The FSS also mitigates double penalty effects at different spatial scales by comparing neighbourhood fractions, which represent the relative proportion of binary events occurring within a specified area like precipitation exceeding a specified threshold. The FSS compares the spatial relationship between the forecast and observed neighbourhood fractions, measuring the improvement in skill beyond random chance (Necker et al., 2023). The purpose of this verification method is to assess how forecast skill varies with spatial scale in a way that can be understood by users intuitively and is directly applicable for postprocessing. The FSS has a range from 0 to 1, with 1 being the perfect FSS and 0 meaning no skill (Roberts & Lean, 2008).

The verification method starts with converting the data to binary fields using suitable thresholds. All grid cells exceeding the threshold gets a value of 1 and others a value of 0. Then, for every grid cell in the binary field the fraction of surrounding cells within a given square of length n that have a value of 1 is computed (figure 4). These quantities represent a measure for the spatial density in the binary field (Roberts & Lean, 2008).

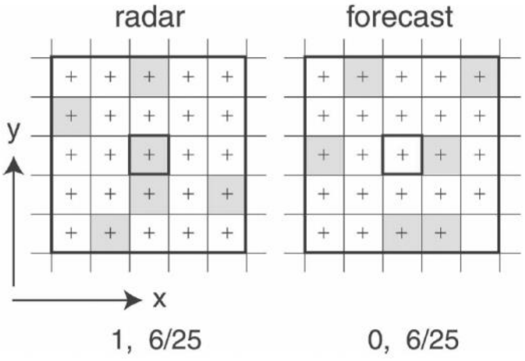


Figure 4. Visualization of the determination of the FSS. From Robert & Lean (2008).

The FSS is computed as follows:

$$FSS_n = \frac{MSE_n - MSE_{n,ref}}{MSE_{n,perfect} - MSE_{n,perfect}} = 1 - \frac{MSE_n}{MSE_{n,perfect}} \tag{2-6}$$

Where MSE_(n)perfect is 0 is the MSE of a perfect forecast for neighbourhood length n. MSE_ref is the reference used for each neighbourhood length n, which can be thought as the largest possible MSE that can be obtained from the forecast and observed fractions. Skill increases with spatial scale until there comes a point at which some desired level of skill has been reached (Figure 5). This is the smallest spatial scale at which output from the forecast model should be presented at. The scales should always exceed five grid cell lengths. At larger spatial scales, skill increases further but the information of the forecasts is limited by the additional smoothing (Roberts & Lean, 2008).

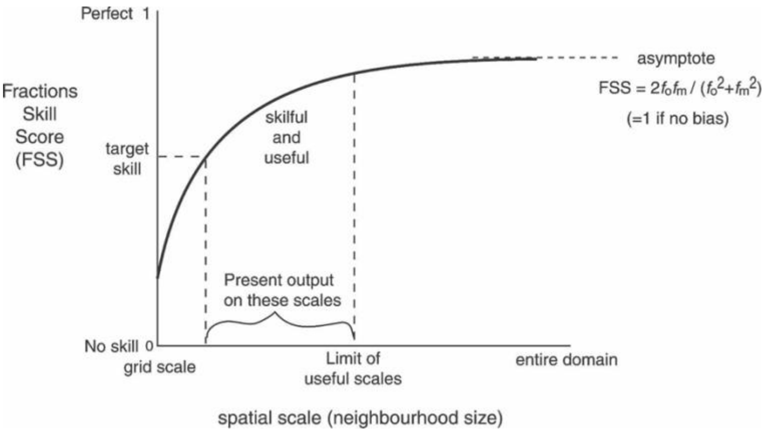


Figure 5. Schematic graph of skill against spatial scale. From Robert & Lean (2008).

3. Data

Radar composites knmi data

The radar reflectivity composites dataset includes gridded files that contains radar reflectivities (in dBZ) at an altitude of 1500 meters. The data covers The Netherlands and surrounding areas, with the following boundaries:

- North bound latitude: 55.97
- East bound longitude: 10.86
- South bound latitude: 48.9
- West bound longitude: 0.0

The radar reflectivity is measured by two radars, locating in Herwijnen and Den Helder. The radar reflectivity is recorded at a time interval of 5 minutes, with each pixel being 1,0 by 1,0 km (KNMI, 2015). This data is provided in h5 files.

Harmonie cy40 knmi data

The Harmonie-Arome Cy40 dataset includes information in gridded format of near surface and altitudes up to 300 m and a few pressure levels from the Harmonie-Arome Cy40 model. This dataset covers Europe, with the following boundaries:

- North bound latitude: 59.0
- East bound longitude: 16.0
- South bound latitude: 44.0
- West bound longitude: -8.0

The output frequency of this dataset is 1 hour, and each pixel is an area of 2,5 by 2,5 km. This dataset contains forecasts for the next 48 hours (KNMI, 2018; KNMI, n.d.-a). This dataset includes many variables, including rainfall intensities and accumulated rainfall amounts. Furthermore, the data is provided in grib files.

Initially, there was a plan to retrieve a full dataset of Harmonie. However, it was discovered that KNMI does not store any history of these data. Although access to the archive data of KNMI was possible, it came with costs for data delivery. Consequently, an alternative method was adopted, involving using Harmonie Arome Cy40 data. This dataset is made available by KNMI but is limited to only the last 10 files. To ensure the availability of this data, efforts were made to download the files daily to create an independent archive. However, this method for retrieving data remained challenging. Data for some days were missed due to oversight and some files were not even made available by KNMI. Consequently, our own archive does not contain data from continuous days. This limitation has an impact on selecting the events to test on, which will be further elaborated in Chapter 5.

Ethical and legal considerations

Ethical and legal considerations is not applicable for this thesis project as the data only consist of precipitation rates.

4. Methods

General

The programming is done on Google Colab, which is a free online service provided by Google. This platform offers access to GPU, which significantly improved the time needed to execute the DGMR model. Due to its accessibility, cost-effectiveness and computational power, Google Colab was a good choice to use for this thesis project.

The blending was done using the Python package *Pysteps*, which is a free and open-source Python framework for short-term ensemble prediction systems. Pysteps mainly focuses on probabilistic nowcasting of radar precipitation fields, but it has a wider range of uses.

The blending was done using Pysteps' *linear_blending.forecast.* However, the original code for this function was modified in such a way that the DGMR model can be used as the nowcast component. The original code of Pysteps' linear blending uses its own nowcasting model that is determined by motion fields. The modified version skips the step of determining the motion fields for the nowcast component and uses the DGMR output instead.

Linear blending using Pysteps was selected as the blending method for this thesis project due to the availability of a well-supported Python package and its big and active community. This made it possible to get some results within the given timeframe for this thesis project. In addition, it was suggested by Ruben Imhoff (one of the developers of Pysteps) that linear blending is a suitable starting point for testing and experimenting with blending methods.

Pysteps' *advection_correction* was also modified to meet the needs for this thesis project. This function was used to interpolate precipitation rates. This will be elaborated later. The modified version of this function gives as output an array of the original Harmonie dataset and the interpolated forecasts (between two original timepoints with an hour in between). The original version of *advection_correction* only provides a single timestamp.

Data loading

The data from the radar reflectivity dataset was downloaded through the API of KNMI. To download a whole dataset with multiple h5 files, an API key was obtained by emailing opendata@knmi.nl. Reading these h5 files was done by using the importer *knmi_hdf5*, following the same steps as the previous internship project (Bosdijk, n.d.).

As for the Harmonie Cy40 data, the Python package *Xarray* was used to read the grib files. Specifically*, open_dataset* was used to read the data. In addition, the parameter '181' was selected to get the data for rainfall and step type as 'instant' to get the rainfall intensities (KNMI, n.d.-a). The engine parameter was set to *cfgrib,* which was required to be able to read the grib files using *Xarray*.

The grib files of Harmonie Cy40 includes various variables, including the rainfall intensities and latitude and longitude. The rainfall data was selected by filtering the data to the variable 'unknown'. Using the documentation of Harmonie dataset, it was determined that this 'unknown' variable is the rainfall intensities (KNMI, n.d.-a).

Data pre-processing

Removing clutter

The raw radar data from KNMI needed to be prepared before using it as input for the DGMR model. Like mentioned before, the raw data contained noise that does not originate from precipitation. The removal of this noise was done by applying the Gabella filter. This was done by using the function *clutter.filter_gabella* from the Python package *wradlib*. A window size of 5 pixels was used, which was based on the data pre-processing steps from a previous internship project that this thesis continues (Bosdijk, n.d.).

Converting reflectivity to precipitation rate

The radar data from KNMI provides reflectivity values in dBZ, which needed to be converted to precipitation rates in mm/h. This can be done by applying equation 2-1, using Pysteps' function *conversion.to_rainrate*. This conversion was needed because precipitation rates in mm/h tells the amount of precipitation that will fall and that is critical for weather forecasting and flood warning measures.

Preparing Harmonie Cy40

For the Harmonie Cy40 dataset, the data needed to be filtered to only The Netherlands due to its coverage of whole Europe. This was done by creating a boundary box with minimum longitude and latitude and maximum longitude and latitude. The same boundaries for the radar dataset were used to filter the Harmonie Cy40 dataset. Furthermore, the precipitation data from Harmonie Cy40 are in kg/m^{2}s and this need to be multiplied by 3600 to have it converted to kg/m^2 per hour.

Adjusting spatial resolution

As discussed in Chapter 3, the spatial resolution of the datasets KNMI radar and Harmonie Arome Cy40 did not align with each other, resulting in different dimensions for the input data. To address this problem, the spatial resolution of the radar composites dataset was down sampled to match the spatial resolution of the Harmonie Cy40 dataset. The downsampling was done by calculating the average of the grid cells, using *downscale_local_mean* from the Python package *skimage*. This method of downsampling was chosen for its simplicity.

Adjusting temporal resolution

Similarly, the temporal resolution of the two datasets did not match as well. The temporal resolution of the Harmonie Cy40 dataset was converted from hourly to every 15 minutes by interpolating using the motion fields. This step was done by using Pysteps' function *advection_correction*. The interpolation of the Harmonie Cy40 data to 15 minutes was selected as a balanced approach because it allowed for a sufficient temporal resolution for accurate precipitation modelling while avoiding the uncertainty of interpolating from 1 hour to 5 minutes. For the radar data, the conversion from 5 to 15 minutes was done by using data from every three timestamps, which corresponded to a total of 15 minutes.

Event selection

Due to the limited access to the Harmonie Cy40 dataset of KNMI, the blending was tested on days between May 15th, 2024, to June 11th, 2024. This time frame was chosen because it has the most consistent and continuous data available. For each day within this time frame, it was checked if both radar and Harmonie data from KNMI is available. If data for either dataset was missing on a given day, that given day was excluded from the testing. This check was crucial because the Harmonie data is needed to use it as the NWP component and input for the blending, while the radar data was needed to use it as the nowcast component and use it as the other input for the blending. In addition, the radar data was also needed to use as the actual observation to assess the quality of the blended forecasts.

The days from this period were split into a train and test set. The train set consisted of data from May 24 to June 11, 2024. The test set consisted of May 15th, 16th, 19th and 20th, and June 15th and 18th, 2024. The training set consisted of 80% of the data, while the test set consisted of 20%. The training and test sets were split in such a way that both sets included days with precipitation, with the test set consisting of mainly days with precipitation.

Testing

The blending was tested on 4 ideal moments of the days from the event selection. In an ideal scenario, the moment a forecast is made aligns with the availability of the Harmonie Cy40 data. The Harmonie data are updated every six hours at 00:00, 06:00, 12:00 and 18:00, with each dataset containing forecasts for the next 48 hours. For example, the Harmonie dataset at 06:00 on May 5th includes forecasts from 06:00 on May 5th to 06:00 on May 7th. However, the Harmonie data becomes available around five hours later.

Therefore, for the blending process, the fifth forecast of the Harmonie Cy40 was used and not the first forecast. The timeline goes as follows: at t = 0, a forecast is made and the DGMR model uses radar data from the previous 20 minutes (t = -20 min). At t = 0, the most recent Harmonie Cy40 data becomes available, which ensures that the blending uses the latest data (figure 6).

This approach assumed an ideal scenario where the most recent Harmonie data was used. However, in real-life, these conditions are not always met, and perfect alignment is not always possible. Testing the blending process under other circumstances was considered beyond the scope of this project.

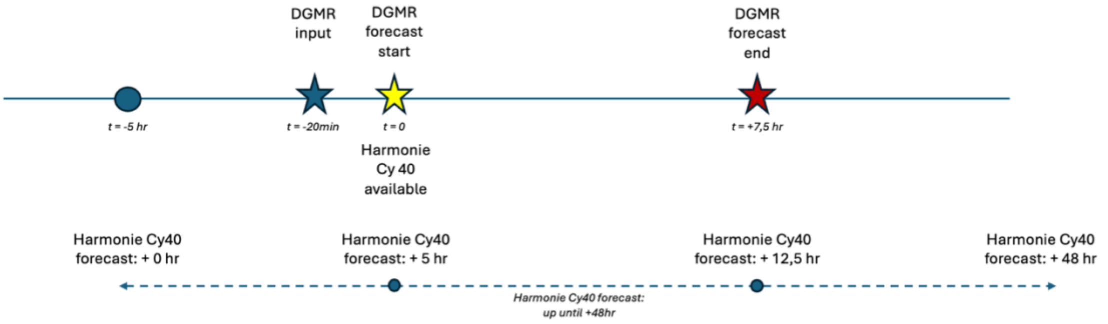

DGMR input
DGMR forecast start
DGMR forecast end
t = -5 hr
t = -20min
t = 0
t = +7,5 hr
Harmonie Cy 40 available
Harmonie Cy40 forecast: + 0 hr
Harmonie Cy40 forecast: + 5 hr
Harmonie Cy40 forecast: + 12,5 hr
Harmonie Cy40 forecast: + 48 hr
Harmonie Cy40 forecast: up until +48hr

Figure 6. Timeline of the blending process.

The blending was first tested on the train set (on the four ideal timestamps on a day). The purpose of this was to determine the optimal start and end time for the blending process. An exhaustive search approach was applied, meaning that every possible combination of start and end time (in steps of 15 minutes) were used for blending and the objective function was determined. The objective function consists of a total score of the POD, FAR and FSS, calculated as follows:

$$\sum_{r=1}^{R} \sum\nolimits_{th \in thresholds} (POD + 2 * FSS - FAR) \tag{4-1}$$

Where:

- R is the number of runs
- Thresholds are precipitation rates of 0.5, 2, 5, 10, 20 and 30 mm/h
- Where POD, FAR and FSS are calculated according to equation 2-4, 2-5 and 2-6 respectively
- If threshold >= 5, then $2 \cdot POD$ and $2 \cdot FAR$

The combination of start and end time with the highest score for this objective function is the optimal start and end time. Blending on the test set was done using these optimal start and end times. At last, an average of the POD, FAR and FSS were calculated of the days from the test set to get an overall estimation of the performance of the blending method.

In addition to the linear blending method, salient-based linear blending was also tested. However, this method was not considered with finding the optimal start and end time. This method was purely tested to see whether it showed any improvement compared to the regular linear blending method.

The actual Python code can be found on the Github repository (Luo, 2024).

5. Results and Discussion

In total, there were 496 possible combinations of start and end times for the blending. Blending with start time of 135 min and end time of 465 min, resulted in the highest objective value of 5642.14, using equation 4-1. This suggested that the blending preformed the best when the blending process started at t = 135 min and ended at t = 465 min (t = 0 is when the forecast is generated) for the training set. The lowest objective value that was achieved was 3404.23, which was for the combination of start time of 0 min and end time of 270 min. The optimal start and end time were used on the test set.

POD

Figure 7 shows a comparison between the linear blending (and salient) and DGMR of the average POD, and linear blending and Harmonie. In the figure, different precipitation rates thresholds are shown. The blended model had an initial POD score of 0.7. This POD score decreased gradually over a period of 7 hours. This indicated that the blended model was able to detect precipitation of 0.5 mm/h well but was able to detect less of the precipitation after 7 hours. From the figure, it can also be seen that the POD decreased for each threshold. For a threshold of 0.5 mm/h, the blended was able to detect more than DGMR and Harmonie for the most part. For thresholds larger than 0.5 mm/h, the blended did not necessarily detect more than DGMR and Harmonie. Figure 7 also shows that for larger thresholds (30 mm/h) had a POD score of around 0.0 during the period 7 hours. This indicated that no precipitation of 30 mm/h was detected. Overall, the POD decreases for all four models, indicating that less precipitation was detected with longer lead times. Appendix I also shows the POD scores for the models in separate plots.

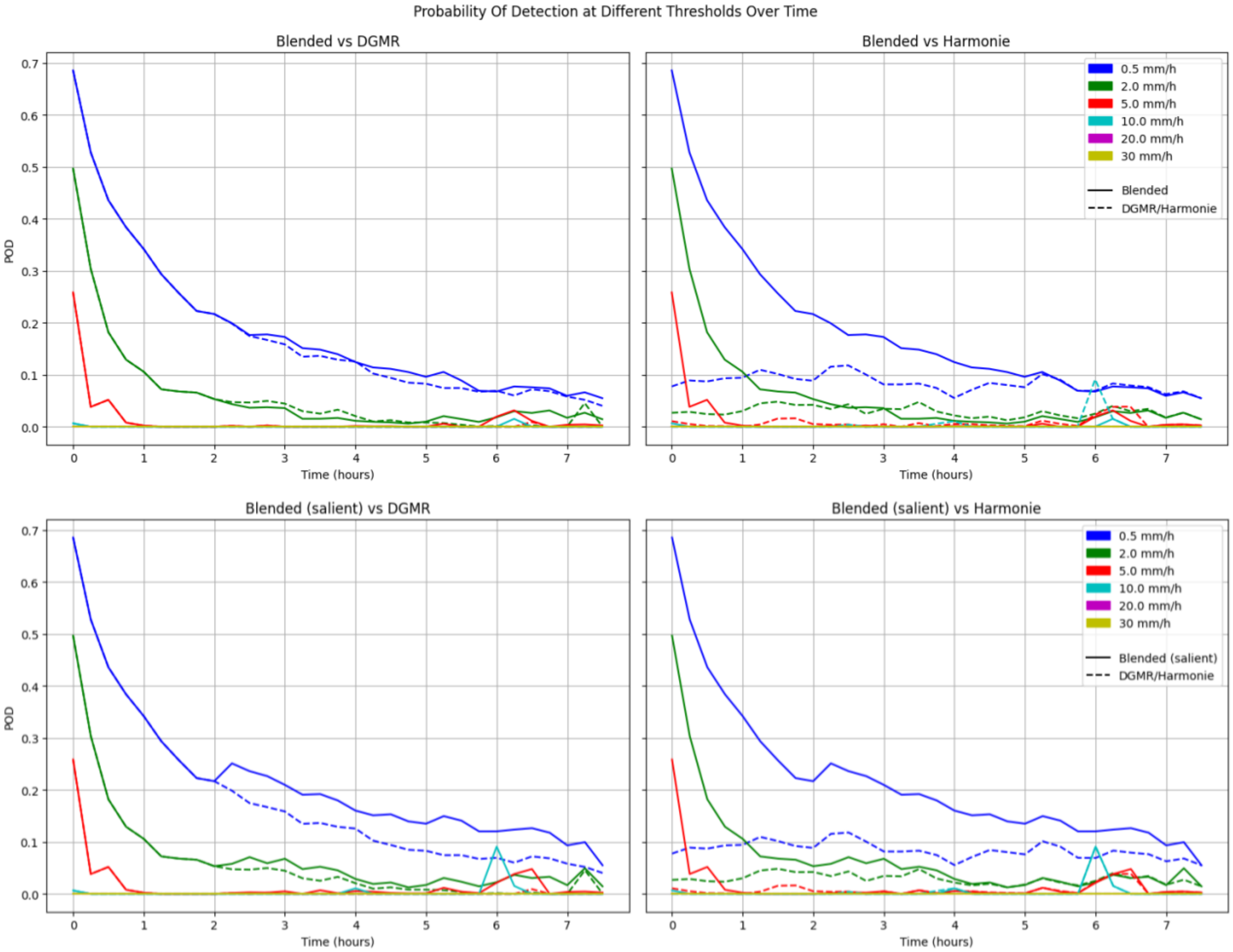


Figure 7. Average POD comparison between the blended and DGMR/Harmonie for different precipitation thresholds over a period of 7 hours.

It was also noticed that the POD scores for Harmonie were noticeable low, with a maximum of 0.1. This might be the reason why the blended models did not perform better in some cases. Overall, there was a slight improvement for low-intensity precipitation events when using linear blending.

It should be noted that the DGMR was modified to generate forecasts for lead times up to 7,5 hours. Initially, DGMR generates only 18 frames (90 min). The modified version used the last four frames (20 min) of the generated frames as input for the forecast of the next 18 frames and continued until 7,5 hours is reached. This means that the errors made by the modified DGMR could increase significantly. That might be another reason why the POD decreased significantly over time.

FAR

Figure 8 shows the differences of the average FAR between the blended, salient blended, DGMR and Harmonie for each precipitation rate threshold. A high FAR score means that there are a lot of false alarms. For all four models, lower thresholds had higher FAR scores than higher thresholds. This indicated that the blended model overpredicted low-intensity precipitation events. Overall, Harmonie preformed the worst with high FAR scores for every threshold and lead time. The overall best-performing model is the DGMR.

The blended models had a lower FAR for high thresholds, indicating that it was able to handle high-intensity precipitation better (compared to Harmonie). It was seen that the Harmonie had noticeable high FAR scores (around 1.0), this might be the cause for high FAR scores for the blended models as well since Harmonie is one of the input data for these models. It was also noticed that the Harmonie tend to overestimate precipitation rates when visualizing the data on a map of The Netherlands. The blended models were able to get lower FAR scores, and this might be due to DGMR, which balanced out the high FAR scores of the Harmonie. However, the FAR scores for all models fluctuated, which indicated inconsistency in forecasting performance. The fluctuations in FAR scores might be explained by the limited amount of data that was available for this thesis project.

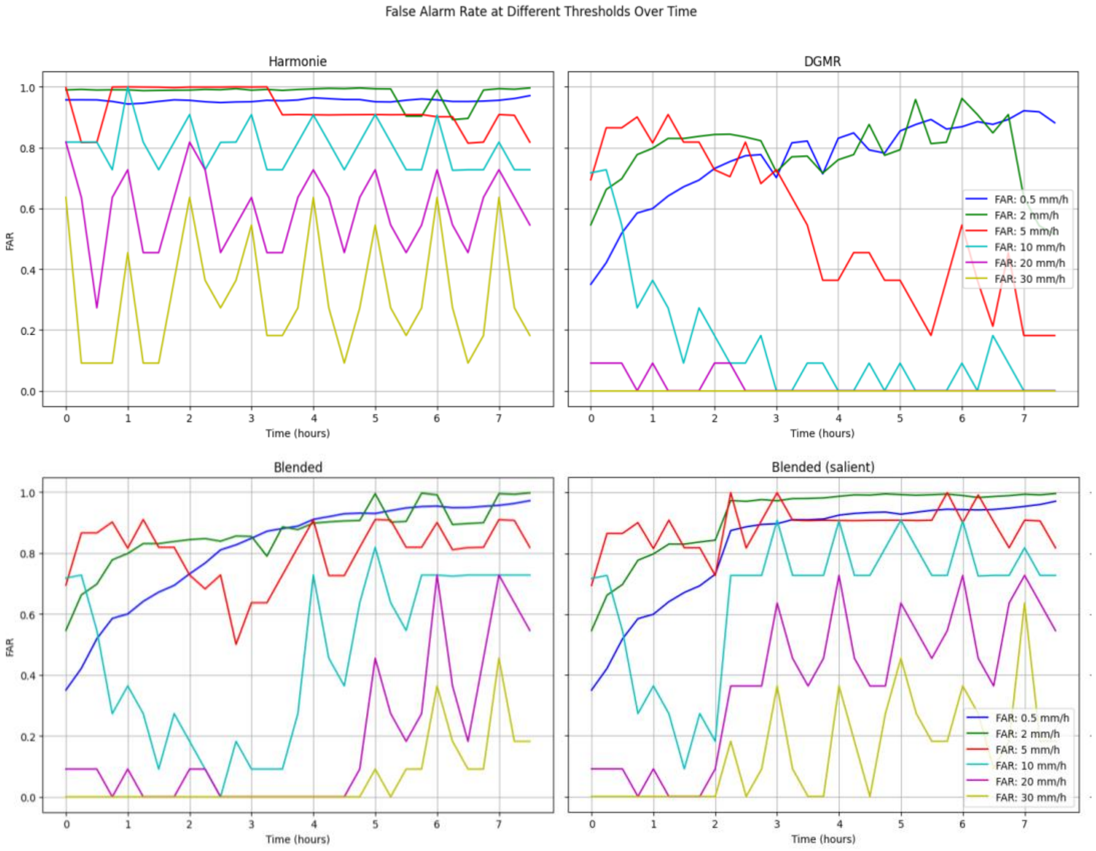


Figure 8. Average FAR comparison between the blended and DGMR/Harmonie for different thresholds over a period of 7 hours.

FSS

Figure 9 shows a comparison of the FSS for the blended models, DGMR and Harmonie. The larger the green areas are, the better the forecasts are. The FSS describes until what lead time the model can achieve a good forecast at each scale and for each precipitation rate threshold. The best preforming model was the DGMR, while the other three models showed similar FSS patterns. The blended models did show some improvement compared to the Harmonie. However, the blended models were not able to get good forecasts for high intensities (around 0.0), most likely because the blended models did not manage to detect the high intensity precipitation events at all (for high intensities: POD around 0.0). The blended models did manage to get good forecasts for lower intensities, however, DGMR still outperformed the models.

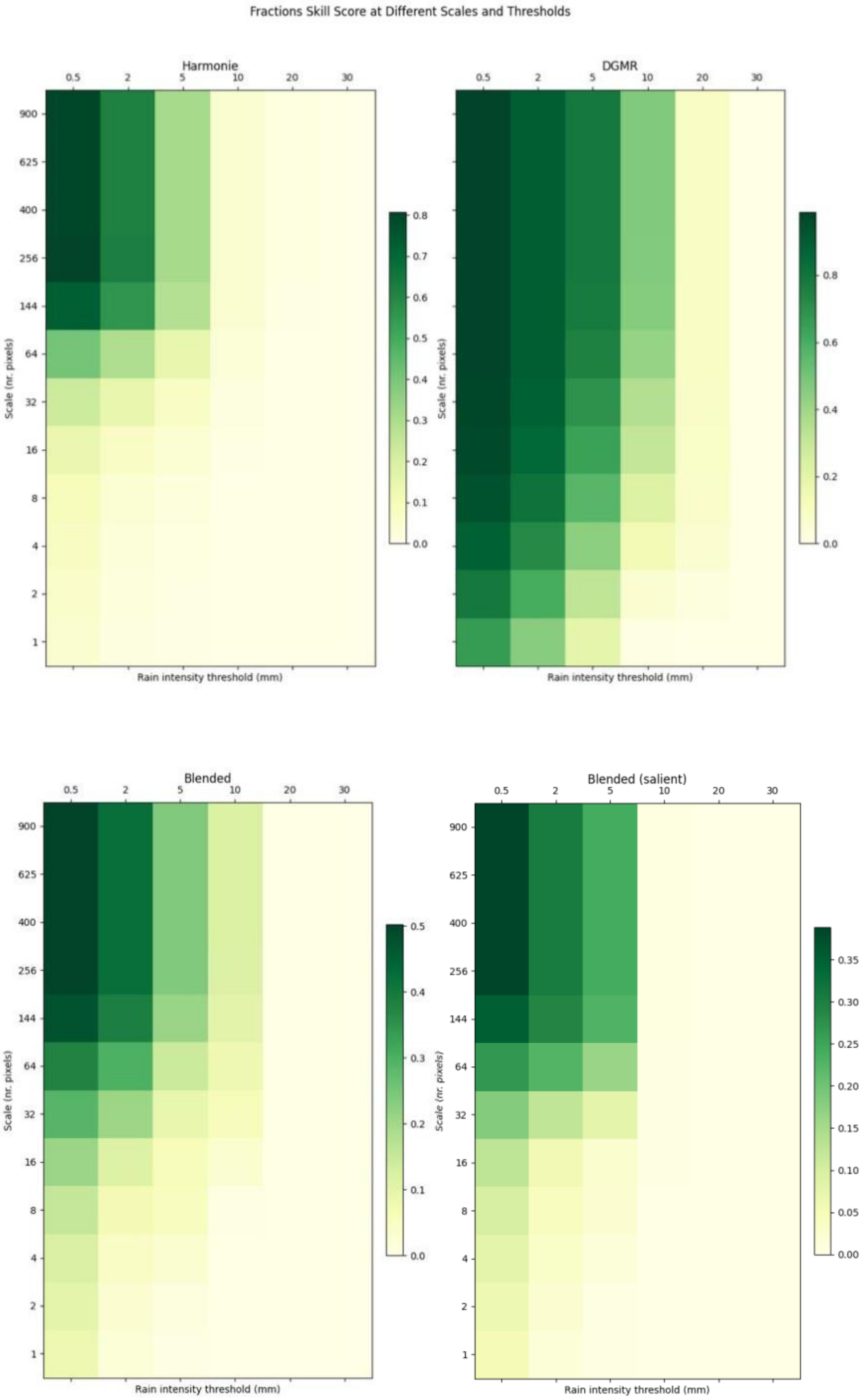


Figure 9. Average FSS comparison of the models for different thresholds and at different scales.

Example: high precipitation on June 10th, 2024, at 11:00

Figure 10 shows the precipitation rates of visualized on the map of The Netherlands for June 10th, at 11:00, where the low intensities are indicated with blue and high intensities with yellow or red. This figure compares the input: DGMR and Harmonie, the output: linear and salient blending and the ground truth. As was expected, the DGMR started good and was similar to the actual observation because it used the radar data. It was seen that the precipitation areas decreased over a period of 7 hours for the DGMR. This can be explained by that DGMR is not able to create new clouds, it only uses what is in the radar data of the last 20 minutes. It was also noticed that Harmonie overestimated the precipitation rates on this day. However, in this case, it somewhat balanced out the decay of the precipitation area of the DGMR. A full version with frames of every 15 minutes is available on Github (Luo, 2024).

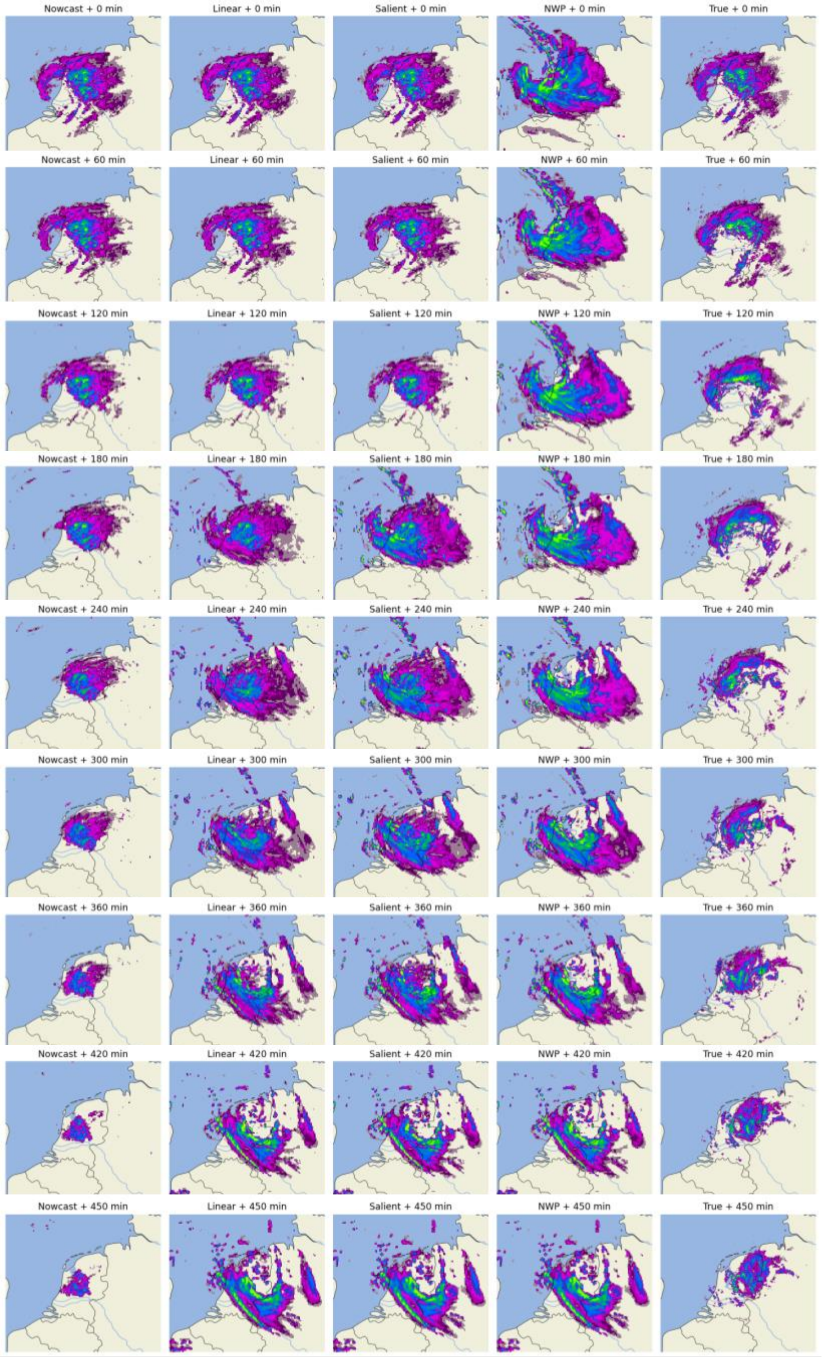


Figure 10. Visualization of the inputs (DGMR and Harmonie), the blended (salient) and the ground truth per hour.

The overestimation problem of Harmonie data was also seen for other days. From Figure 10, it can also be seen that Harmonie did not predict the precipitation at the correct location, which was also observed for some of the other days. This might be caused by the accuracy or the lower spatial resolution of the Harmonie data.

Figure 11 shows the comparison of the POD scores for the blended, Harmonie and DGMR for June 10th, 2024, at 11:00. The POD scores decreased gradually for every threshold, except for the threshold of 0.5 mm/h. For the threshold of 0.5 mm/h of the blended models, there was a fluctuation for the longer lead times. This was caused by the fluctuation of the DGMR (between lead time 4 and 6 hours). This led to a higher POD for the blended models. It was seen that there were some interrupted POD lines, or some lines were missing, indicating that the model were not able to detect some of the precipitation at all (figure 11).

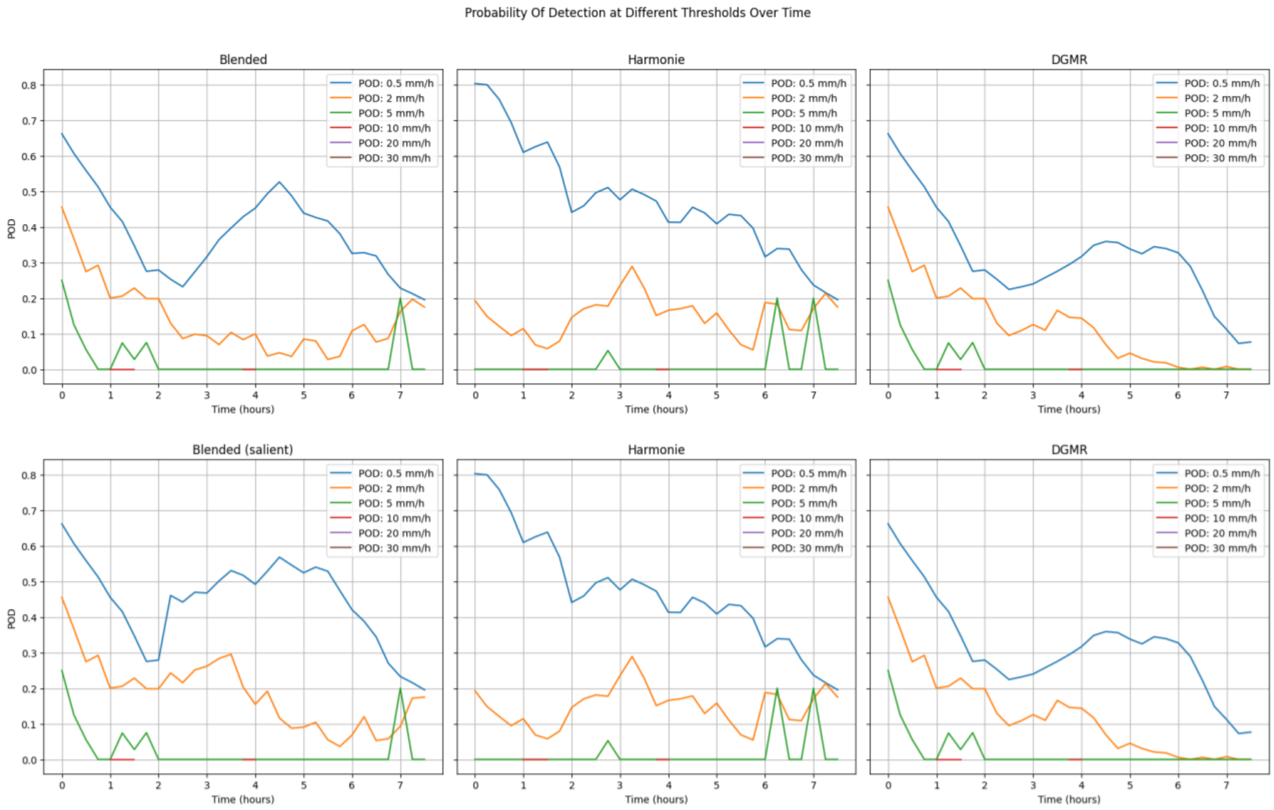


Figure 11. POD scores for June 10th, 2024.

Figure 12 shows the FAR of June 10th, 2024, at 11:00. It was seen that the FAR scores for the blended models started low and increased over the period of 7 hours. This might be due to the low FAR of the DGMR for the first few hours. During this time, the weight for the DGMR is high so it has more influence on the blending. The high FAR scores at the end might be due to the high FAR scores of the Harmonie, which has more influence for the longer lead times. The high FAR scores for the Harmonie were caused by the tendency of overestimating the precipitation (figure 10). Overall, the blended models' performance was in between the Harmonie and DGMR.

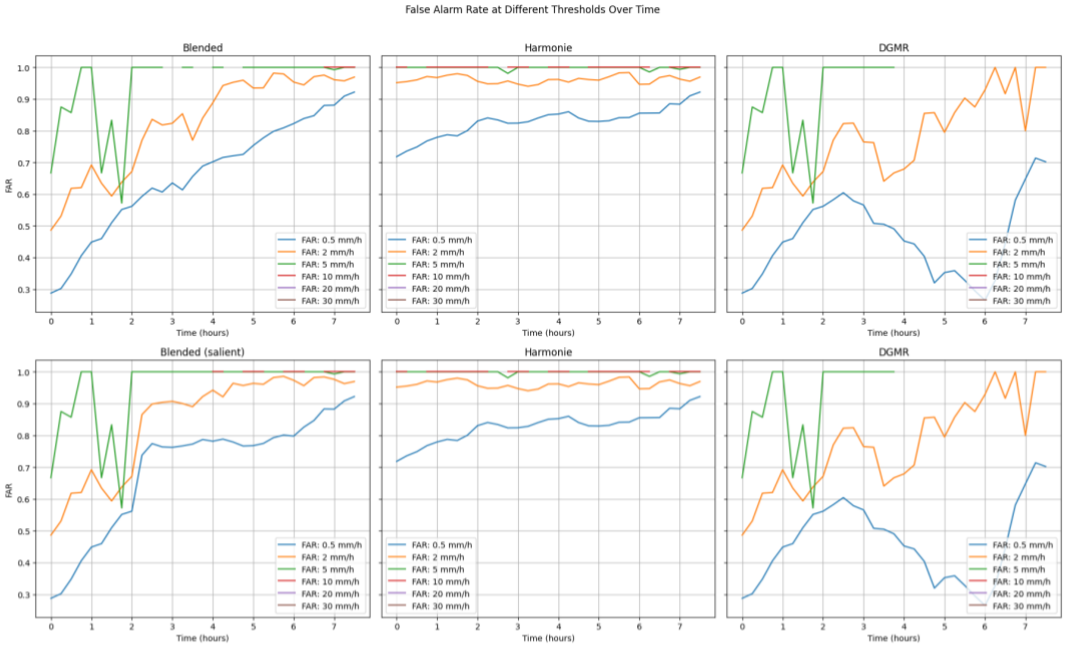


Figure 12. FAR scores for June 10th, 2024, at 11:00.

Figure 13 shows the FSS of June 10th, 2024, at 11:00. The blended models showed similar FSS pattern. The blended models were able to generate good forecasts for low intensities but were not able generate forecasts for high intensities at all. This can be seen at the POD scores as well. This might be due to the DGMR, which showed similar pattern.

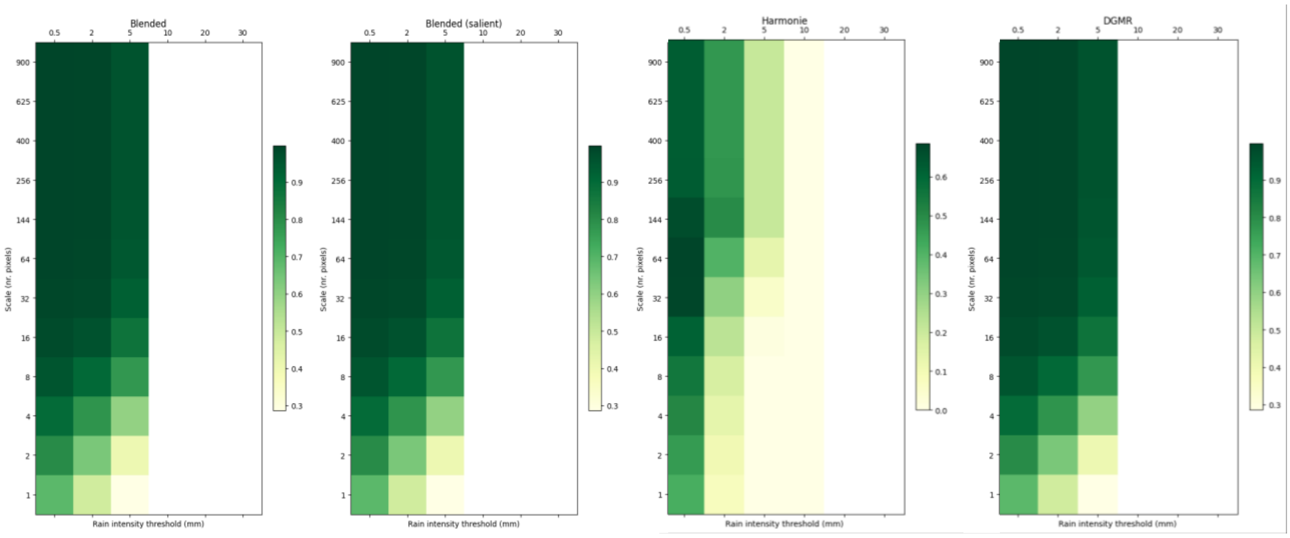


Figure 13. FSS scores for June 10th, 2024 (11:00)

6. Conclusion

This thesis project focused on determining the potential of linear blending for extending the precipitation forecast horizon of Weather Impact's DGMR model and finding the answer for the following research question:

How does linear blending of nowcast (DGMR) and NWP (KNMI Harmonie Cy40) impact the accuracy and reliability of precipitation forecasts beyond the lead time of 2 hours?

Linear blending showed some improvements in detecting low intensity precipitation events but struggled detecting high intensity precipitation events. This method also tended to overestimate precipitation rates.

This research question was answered by using Pysteps' linear blending with the output of DGMR and KNMI's Harmonie Arome Cy40 data. The blending method showed the best performance when using start time of 135 min and end time of 465 min for the blending process from the training set, which was determined using an exhaustive search approach. These values were used for the test set and were evaluated on the average of the probability of detection (POD), false alarm ratio (FAR) and fractional skill score (FSS). It was seen that the blending method was mainly influenced by the Harmonie data, which had the tendency to overestimate the precipitation rates, especially for low precipitation events. This explained the high FAR scores of the blending. The salient blending method showed similar performance as linear blending.

7. Future research

The main limitation of this thesis project was that the Harmonie dataset only covered a period of one month. For more comprehensive analysis and better findings, it is essential to use data that spans at least one year because the precipitation forecast needs to be accurate for the whole year and not only for the summer season. Seasonal variation in precipitation might be a factor that needs to be considered. This would improve the reliability of the results.

Another limitation of this thesis project was access to the KNMI's Harmonie data. Due to missing credentials as students, it was not possible to use the data. Weather Impact does have a license for Harmonie data that can have higher spatial resolution. This might improve the blending results and therefore, future research should consider using this other Harmonie data.

This thesis project mainly focused on linear blending without saliency. The salient based linear blending was not used in the step of finding the optimal start and end time. However, research have shown that this method has potential and therefore, it is worth continuing working on this part as well.

List of references

AON. (2021) *Global catastrophe recap: July 2021*. Tech. rep. London, United Kingdom: AON. URL: http://thoughtleadership.aon.com/Documents/20211008_analytics%-if-july-global-recap.pdf.

Berenguer, M., Surcel, M., Zawadzki, I., Xue, M., & Kong, F. (2012). The Diurnal Cycle of Precipitation from Continental Radar Mosaics and Numerical Weather Prediction Models. Part II: Intercomparison among Numerical Models and with Nowcasting. *Monthly Weather Review*, *140*(8), 2689–2705. https://doi.org/10.1175/mwr-d-11-00181.1

Boneh, T., Weymouth, G. T., Newham, P., Potts, R., Bally, J., Nicholson, A. E., & Korb, K. B. (2015). Fog forecasting for Melbourne Airport using a Bayesian decision network. *Weather and Forecasting*, *30*(5), 1218–1233. https://doi.org/10.1175/waf-d-15-0005.1

Bosdijk, J. (n.d.). *Radar-based nowcasting with deep learning* [Internship project]. Wageningen University.

Bowler, N. E., Pierce, C., & Seed, A. (2006). STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Quarterly Journal of the Royal Meteorological Society*, *132*(620), 2127–2155. https://doi.org/10.1256/qj.04.100

Buienradar. (n.d.). *Waarom de Buienradar niet altijd klopt*. https://www.buienradar.nl/nederland/weerbericht/blog/hoe-werkt-een-radar-regen-buienradar-uitleg-waarom-werkt-niet

Cheung, P., Li, P. W., & Wong, W. K. (2015, January). Blending of extrapolated radar reflectivity with simulated reflectivity from NWP for a seamless significant convection forecast up to 6 hours. *Hong Kong Observatory*. Retrieved May 14, 2024, from https://docs.com-swirls.org/latest/publications.html

Crane, R. K. (1990). Space-time structure of rain rate fields. *Journal of Geophysical Research*, *95*(D3), 2011–2020. https://doi.org/10.1029/jd095id03p02011

Fournier, J. D. (1999). Reflectivity rainfall rate relationships in operational meteorology. *National Weather Service*. https://www.weather.gov/tae/research-zrpaper

Gneiting, T., & Raftery, A. E. (2005). Weather Forecasting with Ensemble Methods. *Science*, *310*(5746), 248–249. https://doi.org/10.1126/science.1115255

Henrichs, Y., Lenderink, G., Kaspar, F., Nilson, E., Otto, F.E.L., Ragone, F., Seneviratne, S.I., Singh, R.K., Skålevåg, A., Termonia, P., Thalheimer, L., van Aalst, M., Van den Bergh, J., Van de Vyver, H., Vannitsem, S., van Oldenborgh, G.J., Van Schaeybroeck, B., Vautard, R., Vonk, D. and Wanders, N. (2021) *Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021*. Oxford, United Kingdom: World Weather Atribution. URL: http://hdl.handle.net/1854/LU-8732135.

Hubbert, J. C., Dixon, M., Ellis, S. M., & Meymaris, G. (2009). Weather Radar Ground Clutter. Part I: Identification, Modeling, and Simulation. *Journal of Atmospheric and Oceanic Technology*, *26*(7), 1165–1180. https://doi.org/10.1175/2009jtecha1159.1

Hwang, Y. J., Clark, A. J., Lakshmanan, V., & Koch, S. E. (2015). Improved Nowcasts by Blending Extrapolation and Model Forecasts. *Weather and Forecasting*, *30*(5), 1201–1217. https://doi.org/10.1175/waf-d-15-0057.1

Imhoff, R., De Cruz, L., Dewettinck, W., Brauer, C., Uijlenhoet, R., Van Heeringen, K., Velasco-Forero, C., Nerini, D., Van Ginderachter, M., & Weerts, A. (2023). Scale-dependent blending of ensemble rainfall nowcasts and numerical weather prediction in the open-source pysteps library. *Quarterly Journal of the Royal Meteorological Society*, *149*(753), 1335–1364. https://doi.org/10.1002/qj.4461

Kilambi, A., and I. Zawadzki, 2005: An evaluation of ensembles based upon MAPLE precipitation nowcasts and NWP precipitation forecasts. *32nd Conf. on Radar Meteorology*, Albuquerque, NM, Amer. Meteor. Soc., P3R.4, http://ams.confex.com/ams/pdfpapers/96982.pdf.

KNMI. (n.d.). *KNMI - weermodellen*. Retrieved May 6, 2024, from https://www.knmi.nl/kennis-en-datacentrum/uitleg/weermodellen

Koks, E., Van Ginkel, K., Van Marle, M. and Lemnitzer, A. (2022). Brief Communication: Critical Infrastructure impacts of the 2021 mid-July western European flood event. *Natural Hazards and Earth System Sciences Discussions*, 22, 3831–3838.

Koninklijk Nederlands Meteorologisch Instituut. (2015). *Precipitation - radar 5 minute reflectivity composites over the Netherlands - KNMI Data Platform* [Dataset]. https://dataplatform.knmi.nl/dataset/radar-reflectivity-composites-2-0

Koninklijk Nederlands Meteorologisch Instituut. (2018). *Weather model - HARMONIE-AROME Cy40 forecasts Europe, near surface, boundary layer and pressure level parameters - KNMI Data Platform* [Dataset]. https://dataplatform.knmi.nl/dataset/harmonie-arome-cy40-p3-0-2

Koninklijk Nederlands Meteorologisch Instituut. (n.d.-a). *Documentatie harmonie*. Open Data | KNMI Dataplatform. Retrieved June 5, 2024, from https://www.knmidata.nl/open-data/harmonie

Koninklijk Nederlands Meteorologisch Instituut. (n.d.-b). *KNMI - Hoe werkt de neerslagradar van het KNMI?* https://www.knmi.nl/kennis-en-datacentrum/achtergrond/neerslagradar-knmi-uitgebreide-toelichting

Kreienkamp, F., Philip, S.Y., Tradowsky, J.S., Kew, S.F., Lorenz, P., Arrighi, J., Belleflamme, A., Bettmann, T., Caluwaerts, S., Chan, S.C., Ciavarella, A., De Cruz, L., de Vries, H., Demuth, N., Ferrone, A., Fischer, R.M., Fowler, H.J., Goergen, K., Heinrich, D.,

Lin, C. A., Vasić, S., Kilambi, A., Turner, B., & Zawadzki, I. (2005). Precipitation forecast skill of numerical weather prediction models and radar nowcasts. *Geophysical Research Letters*, *32*(14). https://doi.org/10.1029/2005gl023451

Luo, K. (2024). ADS Thesis Project. Github. https://github.com/kevinluo97/ADS-Thesis-Project.git

Necker, T., Wolfgruber, L., Kugler, L., Weißmann, M., Dorninger, M., & Serafin, S. (2023). The fractions skill score for ensemble forecast verification. *Authorea (Authorea)*. https://doi.org/10.22541/au.169169008.89657659/v1

North Dakota State University. (2018, January). *Air temperature inversions Causes, characteristics and potential effects on pesticide spray drift*. NDSU Agriculture. https://www.ndsu.edu/agriculture/extension/publications/air-temperature-inversions-causes-characteristics-and-potential-effects

Pulkkinen, S., Nerini, D., Hortal, A. a. P., Velasco-Forero, C., Seed, A., Germann, U., & Foresti, L. (2019). Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1.0). *Geoscientific Model Development*, *12*(10), 4185–4219. https://doi.org/10.5194/gmd-12-4185-2019

Pysteps. (n.d.). Linear Blending [Python code]. Retrieved from https://github.com/pySTEPS/pysteps/blob/master/pysteps/blending/linear_blending.py

Radhakrishnan, C., & Chandrasekar, V. (2020). CASA Prediction System over Dallas–Fort Worth Urban Network: Blending of Nowcasting and High-Resolution Numerical Weather Prediction Model.

Journal of Atmospheric and Oceanic Technology, *37*(2), 211–228. https://doi.org/10.1175/jtech-d-18-0192.1

Ravuri, S. V., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., Madge, S., Prudden, R., Mandhane, A., Clark, A., Brock, A., Simonyan, K., Hadsell, R., Robinson, N. H., Clancy, E., Arribas, A., & Mohamed, S. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature*, *597*(7878), 672–677. https://doi.org/10.1038/s41586-021-03854-z

Roberts, N., & Lean, H. (2008). Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Monthly Weather Review*, *136*(1), 78–97. https://doi.org/10.1175/2007mwr2123.1

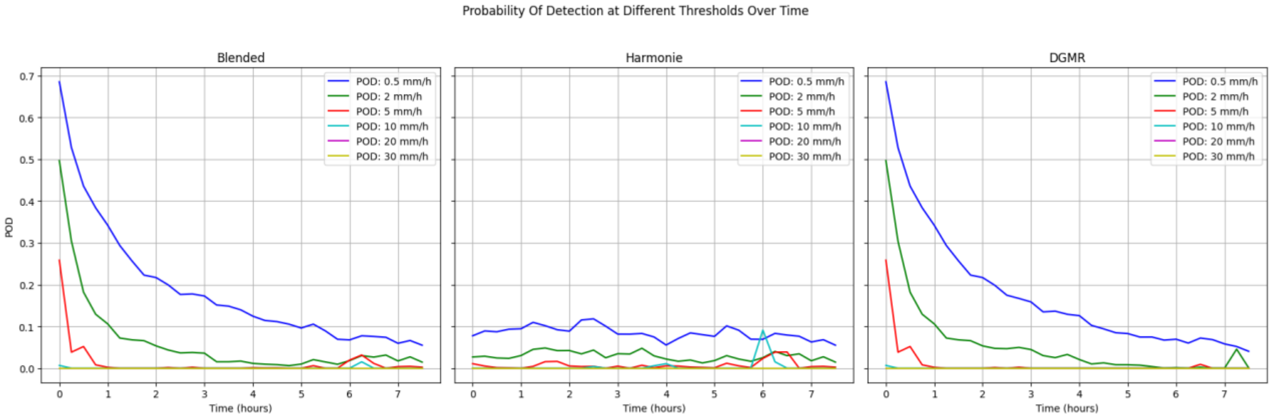
Seed, A., Pierce, C., & Norman, K. (2013). Formulation and evaluation of a scale decomposition-based stochastic precipitation nowcast scheme. *Water Resources Research*, *49*(10), 6624–6641. https://doi.org/10.1002/wrcr.20536

Turner, B., Zawadzki, I., & Germann, U. (2004). Predictability of Precipitation from Continental Radar Images. Part III: Operational Nowcasting Implementation (MAPLE). *Journal of Applied Meteorology*, *43*(2), 231–248. https://doi.org/10.1175/1520-0450(2004)043

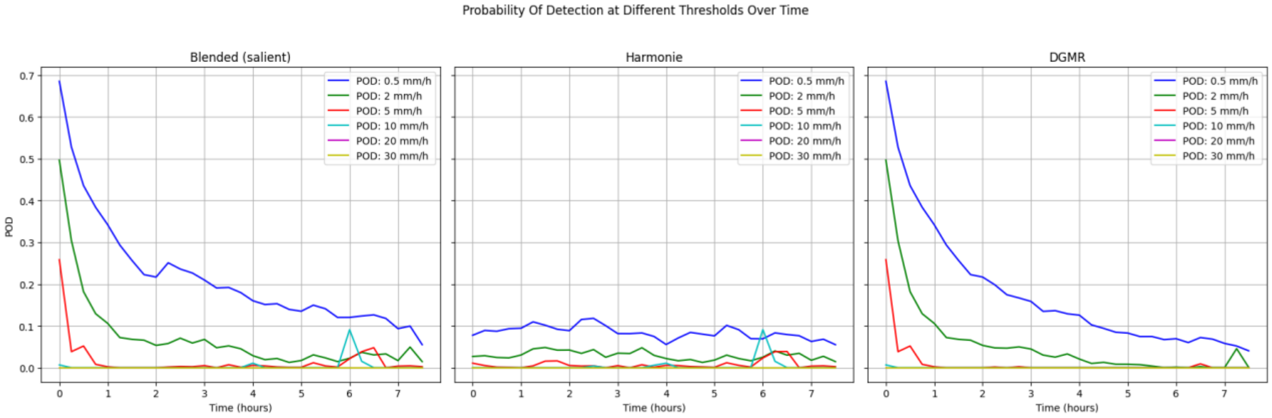
Xue, T., Xu, J., Guan, Z., Chen, H., Chiu, L. S., & Shao, M. (2017). An assessment of the impact of ATMS and CrIS data assimilation on precipitation prediction over the Tibetan Plateau. *Atmospheric Measurement Techniques*, *10*(7), 2517–2531. https://doi.org/10.5194/amt-10-2517-2017

Appendix I

Blended



Salient blended


Probability Of Detection at Different Thresholds Over Time
Blended (salient)
Harmonie
DGMR
POD
Time (hours)
POD: 0.5 mm/h
POD: 2 mm/h
POD: 5 mm/h
POD: 10 mm/h
POD: 20 mm/h
POD: 30 mm/h