UTRECHT UNIVERSITY,

Applied Data Science

Nowcasting rainfall in the Netherlands

with a focus on extreme summer precipitation events

An analysis based on a Deep Generative Model of Rainfall

by

Karolina Mazurek

(student number: 7789831)

Supervisors: Derek Karssenberg (Utrecht University)

Thomas Berends (Nelen & Schuurmans)

July 2022

Abstract

Extreme summer precipitation often caused by convection is a phenomenon that can lead to flooding, but at the same time, it is challenging for forecasting methods. Among others, deep learning tools are being used to tackle the problem. The study concerns implementation of DeepMind's Deep Generative Model of Rainfall (DGMR) to data about rainfall in the Netherlands gathered in Nationale Regenradar. DGMR is a nowcasting method which allows to forecast precipitation within the lead time of 90 minutes, based on an input of data referring to 20 minutes. In the research, data for summer, extreme precipitation events was used as an input of the model. Thirteen such events were chosen in consultation with a meteorologist from Koninklijk Nederlands Meteorologisch Instituut (KNMI). The study's results show that the DGMR proves applicable to the NRR data and this analysis can be used as a proof of concept for further research. Two research questions were addressed: model's performance in nowcasting precipitation of convective or partly convective type, and, change of nowcast's accuracy with increasing lead time. The results of the study approve that convective rainfall is more difficult for the model to nowcast than a mixed type one. Additionally, as the lead time was increasing, a drop in nowcast accuracy was observed.

Contents

The code of the analysis is available at GitHub repository:

https://github.com/KAMazurek/Master-thesis-Nowcasting-rainfall-in-the-Netherlands-with-a-focus-on-extreme-summer-precipitation-e

The NRR data used in the analysis can be accessed via the link:

https://drive.google.com/drive/folders/1CtLOu6UQxzP0gk4BcyJtHiyZLNknM_AQ?usp=sharing

1. Introduction

1.1. Problem definition

Precipitation nowcasting is defined as forecasting of rainfall which is to happen within several hours ahead from the present (Ravuri et al., 2021; WMO, 2019). While some institutions, such as World Meteorological Organization set the boundary on zero to six hours (WMO, 2019), sometimes researchers keep the window as short as two hours which is the case for Google's DeepMind research concerning nowcasting precipitation in the UK (Ravuri et al., 2021). Nowcasting is crucial for efficient short-term decision-making in case of extreme precipitation events which can lead to flooding. As flooding with origins in precipitation is one of the most severe natural dangers worldwide (Westra et al., 2014; Breugem, 2020), tools for its forecasting are needed.

Additionally, extreme rainfall is especially likely to happen in summer when rain sometimes originates from convection (Wapstra, 2016; Breugem, 2020). It lasts for approximately for an hour, but it is intensive (Breugem, 2020). Convective rainfall is expected to be more sensitive to temperature increases than other types of precipitation and as the climate change progresses, this specific type of rainfall has an increasing share in extreme rainfall events (Moore et al., 2012; Gadian et al., 2017). At the same time, as a complex phenomenon which occurs locally and in short periods of time, convection is difficult to model (Becker et al., 2019). In a survey conducted in 2018 among centers which develop weather models, convective precipitation was indicated as the most relevant cause of systematic errors in whether forecasting (Reynolds et al., 2019).

Three approaches are used to tackle the problem: numerical weather prediction (NWP) methods (Sun et al., 2014), algorithms of convective storms (Muñoz, 2017) and models based on deep learning tools (Šaur, 2017). The main advantage of the third type of models is that they learn from historically observed data and no meteorological assumptions are required. Therefore, even a process with undefined rules can be approached this way (van der Kooij, 2021). The aforementioned Deep Generative Model of Rainfall (DGMR) developed by DeepMind for the UK

(Ravuri et al., 2021) is an example of this approach and the pretrained model together with necessary code is publicly available on GitHub (Deep Mind, 2021).

Since convective rainfall occurring in summer proves challenging to forecast and at the same time it is relevant for real-life applications, especially safety-related, this study aims to apply the new DeepMind's model to nowcasting precipitation events in the Netherlands. The goal is to find out how the model performs for extreme summer precipitation events and if there exists potential to use it in practice e.g., in flood early warning systems. In this study, an extreme event is defined as meeting threshold of at least 2.5 millimeters of rain per 5 minutes.

1.2. Research question

The main question to be answered in the study is: what is the ability of the DeepMind's model for nowcasting extreme summer precipitation in the Netherlands? It is followed by sub questions:

- How does the predictive power of the model decrease with lead time?

- How does the model perform in purely convective events nowcasting in comparison to its performance in partly convective (mixed) events nowcasting?

As convective precipitation is more sudden, and more difficult to forecast and non-convective types of precipitation for the existing models, (Reynolds et al., 2019) the hypothesis is that DeepMind's model will also achieve lower levels of accuracy in the case of this kind of rain than when it nowcasts rainfall of mixed kind. Additionally, it is expected that longer lead time will be related to decreasing accuracy of prediction.

2. Theoretical background

The nowcasting model developed by Google's DeepMind, Deep Generative Model of Rainfall (DGMR), is based on the concept of conditional generative adversarial network (GAN) (Ravuri et al., 2021). The idea behind this class of models is to train an additional neural network (*discriminator*) which indirectly participates in the training of the main neural network (*generator*). The networks compete: the generator gives prediction., e.g., generates an image as an output and the discriminator judges its performance (Adaloglou, 2020). In the case of the DGMR,

two discriminators are used reflecting the spatiotemporal nature of data needed for probabilistic weather forecasting. One is responsible for the spatial aspect and anther for the temporal one. The data set used by DeepMind for training and validation of the model consists of raster images originating from meteorological radars of Met Office, the United Kingdom's national weather service. The data covers the entire Great Britain with 1km x 1km grid resolution and new observations are available in 5 min timestamps.

Figure 1 – Input and output of the DGMR model (DeepMind, 2021).

The model operates on a time series represented by a number of raster images. Each of the images depicts values of rainfall in raster cells (in millimeters per 5 min). A series of images used as input refers to 4 timestamps (20 min) and an output is a series of 18 images (90 min) with forecasted values. A schematic description of this process is presented in figure 1.

The DeepMind team compared the performance of GDMR to 3 baselines. The first one, PySTEPS, is a widely used model implemented in Python. The second, UNet, is a model developed specifically for nowcasting. The third, axial attention model, is an implementation of another deep-learning-based method; MetNet. The authors of DGMR state that their model delivers forecasts with higher spatial and temporal accuracy. They indicate that also specifically in case of convective events, GDMR can better assess the spatial coverage of the event over a longer period without over-estimating the intensity (Ravuri et al., 2021).

3. Methods

3.1. Overview of the methodology

In the study data concerning extreme, summer precipitation events in the Netheralnds, from 2016 to 2022 was used. The events were chosen in consultation with a meteorologist from KNMI. The data related to the events underwent preprocessing step in which it was downloaded into Google Collab environment and converted into a tensor object of Tensorflow which is a maching learning library in Python. Then, the DGMR was run once for each of the events. The timestamp containing the maximum level of precipitation which occurred in the event was always placed as the last of nowcasted by the DGMR timestamps.

3.2. Characteristics of the data set

The data used in the study is extracted from a data warehouse of Nelen & Schuurmans using the Lizard API (Nationale Regenradar). The data set comes from Nationale Regenradar (NRR) which is also a product by Nelen & Schuurmans and consists of time series of raster images representing the amount of precipitation in millimeters per 5 minutes (Schuurmans et al., 2013). The NRR product is based on the raw radar images from the KNMI and calibrated using the rain gauges from the KNMI itself and the water authorities. The measurement interval is 5 minutes and the resolution of the raster is 1 km x 1 km, which is the same as in case of the data used to train the GDMR model. NRR is available in 4 time intervals of ascending quality: real time (available after 5 min), near-real time (after approximately 3 hours), another one available after 24 hours, and the final, cleaned version delivered after one month approximately. Two of these sets are used in the study: real time data (in Lizard API referred to as *Regen*) and the final, calibrated one (referred to as *NRR Realtime 5 min*).

Events analyzed for the study are selected by a meteorologist from KNMI. The events selected are cases of extreme summer rainfall, from June 2016 to June 2022 in the Netherlands. An event is considered as extreme if a warning was assigned to it by Koninklijk Nederlands Meteorologisch Instituut (KNMI). Table 1 presents basic characteristics of the chosen 13 events.

	Date	Warning	Rainfall type	Real-time data available
1	2016-06-02	Orange	mixed	No
2	2016-06-23	Orange	mixed	No
3	2017-08-29/30	Yellow	mixed	No
4	2017-09-11/12	Yellow	mixed	No
5	2018-05-29	Orange	convective	No
6	2019-06-05/06	Orange	convective	No
7	2020-08-11	Yellow	convective	No
8	2021-06-18	Orange	convective	Yes
9	2021-06-20	Yellow	mixed	Yes
10	2021-06-21	Yellow	mixed	Yes
11	2021-06-29	Orange	mixed	Yes
12	2021-08-21/22	Yellow	mixed	Yes
13	2022-06-05	Yellow	convective	Yes

Table 1 – Extreme summer precipitation events from May 2016 to June 2022 in the Netherlands.

For each of the events a time series from NRR is extracted, consisting of timestamps from the date of the beginning of the event at 8:00 to the day after the event at 8:00.

Two levels of warning exist, yellow and orange. The lowest level is assigned to heavy precipitation during thunderstorms when the intensity of the rainfall is above 30 mm/hour (2,5 mm/5 min) and an orange warning when it is above 50 mm/hour (5 mm/5 min). These thresholds were established by KNMI for rainfall events of completely or partially convective nature. Precipitation which is not convective occurs mostly in winter and other threshold levels apply to it.

An example of the events chosen for the analysis is presented in Figure 2. It shows the amount of rain observed within a day, based on measurements from 320 stations and interpolated over entirety of the Netherlands.

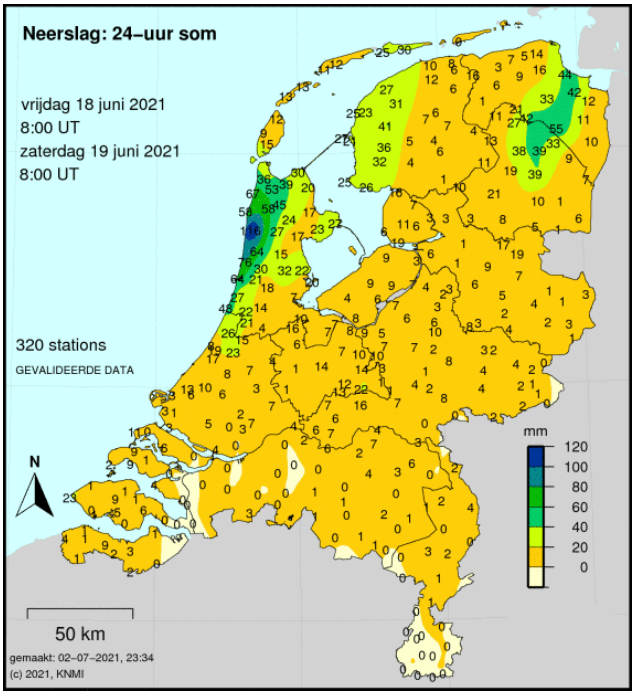


Figure 2 – Example event: 19th June 2021 8:00 – 20th June 2021 8:00 with the amounts of rainfall for that day (KNMI, 2021).

Additionally, rainfall type characteristic for each of the events was assigned by a meteorologist from KNMI. The distinction is relevant for the study and depicts if an event is purely convective or only partly convective. Rainfall of a not at all convective type is not included in the analysis, since it usually occurs in winter. Whenever possible, real-time data is analyzed and only in case it is not present in the Lizard API, the final, calibrated version is used instead. The reason behind this choice is to produce predictions which would be as close to a real-life situation as possible. A forecast used to determine whether e.g., safety measures related to possible flooding need to be undertaken, only a prediction based on the real-time dataset is available. At the same time, for seven events calibrated data is included and validation is done mostly on it.

3.3. Data preprocessing

Preprocessing of the data is depicted in Figure 3 below. The NRR data was initially extracted for each of the dates of events duration from 8:00 to 8:00 the day after. Then, for events with a duration of more than one day, time series were joined in such way that one event is always finally represented by one time series. At the next step, for each timestamp of each event, additional data

points of value 0 are added in order to obtain objects of shape corresponding to 1536 km x 1280 km area. This shape is one of two accepted by the model. An alternative approach would be to tile the data set into the shape corresponding to 256 km x 256 km which is not enough to include data covering the entire area of the Netherlands.

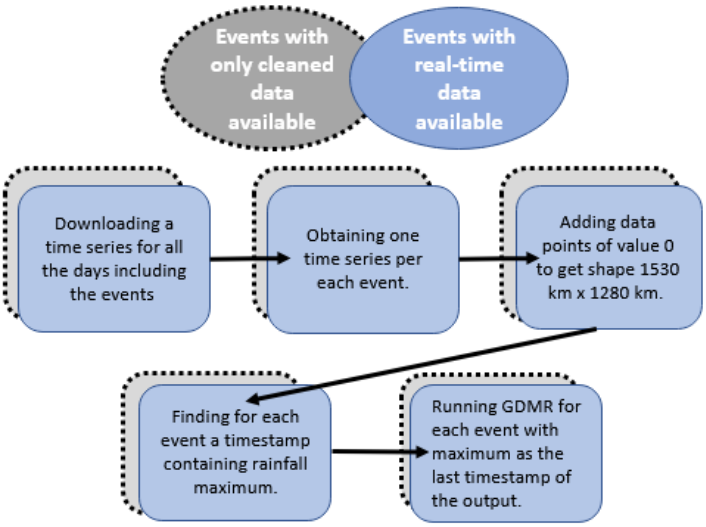


Figure 3 – Data preprocessing.

Below in figure 4, the area included in NRR data is presented together with a map of the Netherlands. The grey are is shows the spatial scope of NRR. There is also one exemplary timestamp with rainfall depicted with black and white spectrum. In this study, the DGMR is always run for this area. Additional 0 values added to obtain the shape required by the model and enlarging the NRR area, as explained above, were only used for running the model but removed before validation.

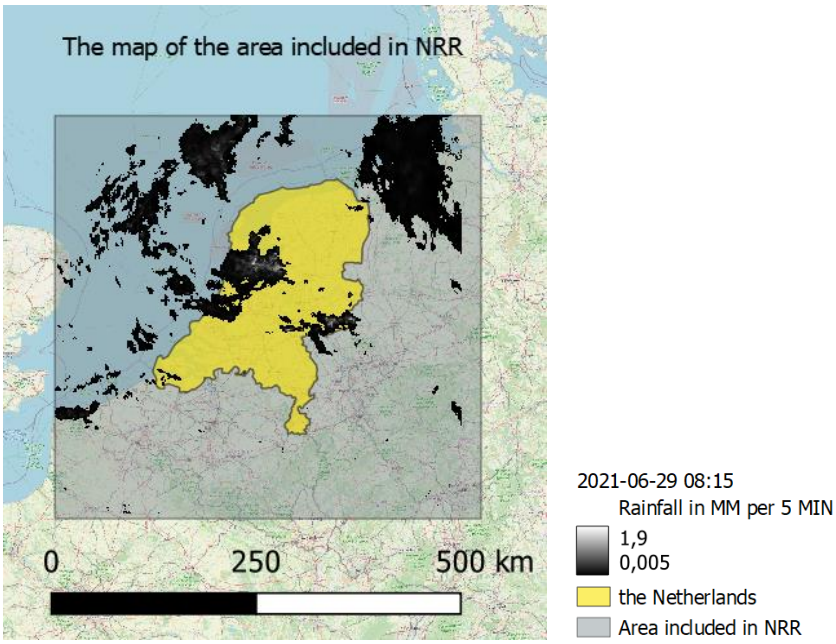


Figure 4 – the map of the area covered by NRR with an exemplary timestamp and the map of the Netherlands.

Since the time frame which can be processed by the model at once is 110 minutes it was necessary to choose only limited part of the events. The timestamp containing the maximum rainfall occurring in the event was always chosen as the last, eighteenth, timestamp in the output of the model. It means that for each event, the timestamps used as input for the model range from 110 to 90 minutes before the maximal values of precipitation per cell, per 5 minutes. Therefore, 22 timestamps are extracted from each event.

Real-time data contains disturbances, and they are not addressed in the preprocessing. Therefore, a gap in the quality of forecasts between the two needs to be assumed. However, real-time data is used whenever available because it is the only possible base for a current decision-making process and is valuable because of this potential real-life application. An example of problems existing in the real-time dataset can be observed in the event on June 18th, 2021 (figure 4). Colors depicting intensity of precipitation in millimeters (from dark green meaning no rain) suggest strong rainfall in the second picture. However, as visible in the series of pictures, the shape created by these colors shows that the colors do not depict real rainfall in this case. Solutions for removing this kind of error could have been addressed in postprocessing. For example, disruptions caused by windmills,

are sometimes eliminated with filters. However, this problem lies out of the scope of this research and real-time data was included without adjustments.

Precipitation event June 18th in the area included in NRR

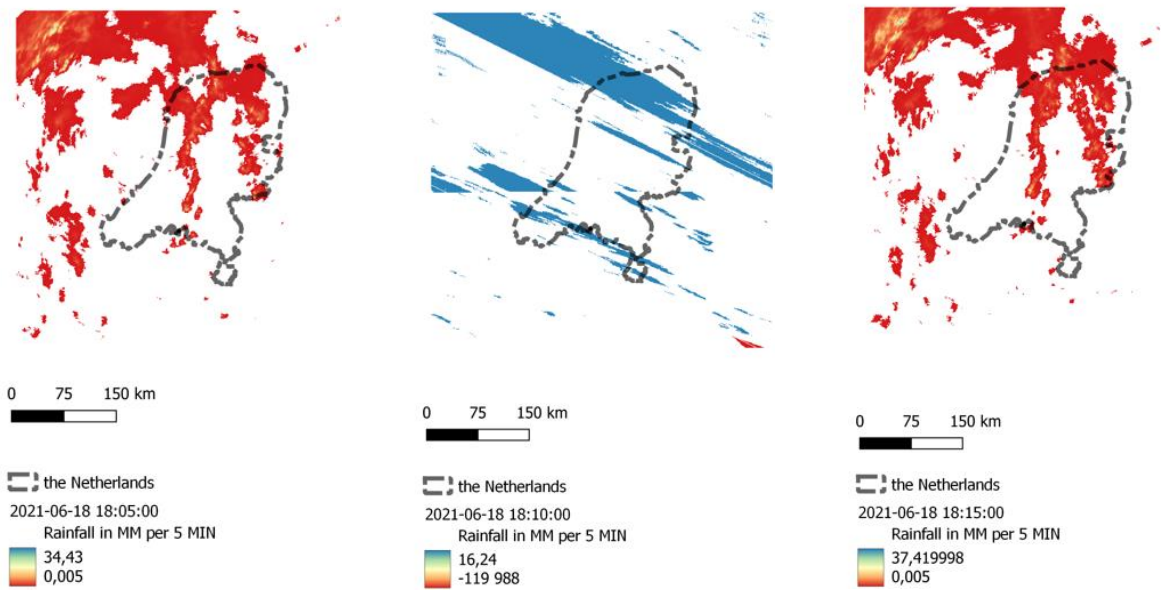


Figure 4 – An example of disruptions occurring in real time data is visible in the second of the pictures (extracted from the rainfall event on June 18th, 2021).

3.4. Application of DGMR to the NRR data

In this study the pretrained DGMR was used. No additional training was performed. All preprocessing steps, as well as running the model, were done with Python machine learning library Tensorflow in Google Collab environment. The model and Python functions for running it are provided in DeepMind's GitHub and were used in the study (DeepMind, 2021).

The model was run once for each of the precipitation events. The number of input and output timestamps is forced by the models' technical specification. Therefore, always 4 timestamps (20 minutes) were used as input and 18 timestamps (90 min) were generated as output.

The code of the analysis is available at GitHub repository:
https://github.com/KAMazurek/Master-thesis-Nowcasting-rainfall-in-the-Netherlands-with-a-focus-on-extreme-summer-precipitation-e

The NRR data used in the analysis can be accessed via the link:
https://drive.google.com/drive/folders/1CtLOu6UQxzP0gk4BcyJtHiyZLNknM_AQ?usp=sharing

3.5. Evaluation Metrics

3.5.1. MSE

Two metrics were used to evaluate the performance of the model: Mean Squared Error (MSE) and Critical Success Index (CSI). Both were also used in the study by DeepMind (Ravuri et al., 2021).

MSE is defined as:

$$MSE = \frac{\sum(y_i - \hat{y}_i)^2}{n} \quad ,$$

where y_i refers to an observed value, $\hat{y}_i$ to predicted value and n is a number of data points.MSE is used because, as a continuous measure, it shows accuracy of the rainfall intensity prediction and location accuracy at the same time (Ravuri et al., 2021). Location accuracy of MSE in this study is presented on an aggregated level (per timestamp, but not per event). Additionally, the level of MSE over different lead times provides information crucial to observe how predictive power of the model changes with lead time

MSE with 4 criteria of selection was measured. The criteria included:

1. Only values over 0 m,
1. Only values of at least 1.5 mm,
2. Only values of at least 2 mm (yellow warning),
3. Only values of at least 5 mm (orange warning).

It is aimed to prevent MSE from overestimating predictive power of the model by including cells with 0 or small values which are most of all cells even for the timestamps when extreme precipitation occurs.

The reason to exclude 0 is that once the goal is to focus on extreme precipitation, locations with no rainfall are not relevant and their prevalence could make evaluation metrics seem much better than they are for the locations with strong rainfalls. Moreover, the number of data points with 0 value was vastly increased during the preprocessing. It was necessary because the model can process only tensors of specified shape and the original shape of tensors created based on NRR data was smaller than the required one. Additionally, MSE only for data points with levels of rainfall meeting thresholds of KNMI's warnings was calculated. One more threshold was set 1.5 mm to present MSE for precipitation that occurs but does not meet the definition of an extreme event.

3.5.2. CSI

Critical Success Index (CSI) can be defined as follows:

$$CSI = \frac{TP}{TP + FP + FN}.$$

It is a metric based on confusion matrix, hence TP represents *true positive*, FP: *false positive* and FN: *false negative.* In case of CSI, true positive means that the nowcasted value and observed value are both at least as high as the threshold. False positive occurs when predicted value at least meets the threshold but observed value does not. False negative means that observed value is at least as high as the threshold but the nowcasted one is lower than threshold.

Since in case of this study CSI requires a transformation of continuous set of values into binary classification, a threshold is needed. Two thresholds were used in the research: 2mm and 2,5 mm (yellow warning). If the threshold is 2, CSI is supposed to inform about the quality of classifying data points as located below or above the value of 2. Threshold of 2.5 mm was chosen to verify the quality of classification whether precipitation will get to the level related to the first level of warning (yellow). Because there are relatively few data points meeting this criteria (therefore, the

classification might be easy), a lower threshold of 2 mm was selected additionally. CSI can be interpreted as the higher the value the better the classification.

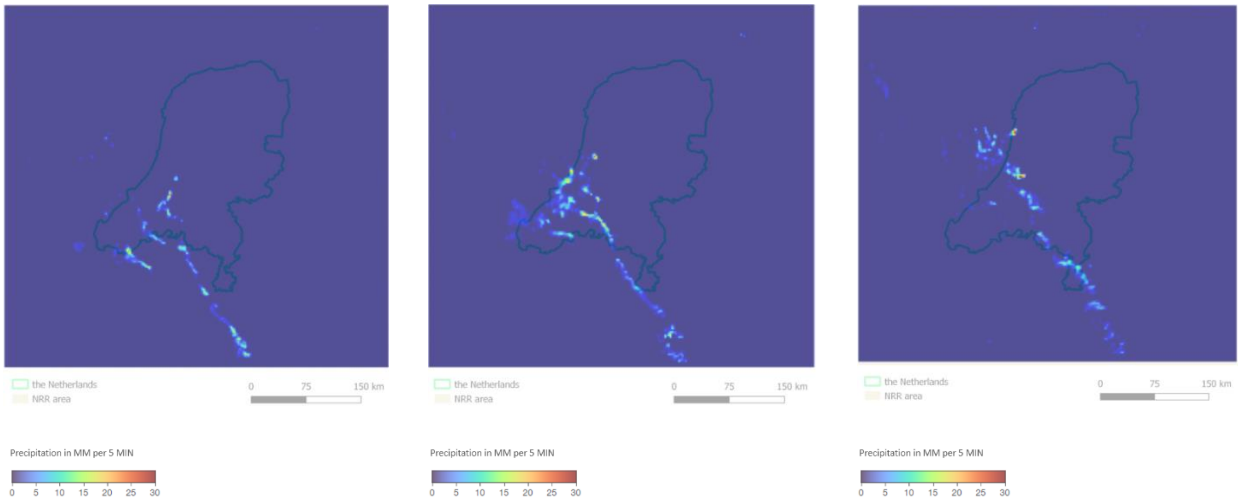
CSI is useful especially for evaluating the model's spatial accuracy which is potentially important in real-life applications (interventions in case of extreme precipitation). With a given threshold, in this case 2 mm and 2,5 mm, CSI evaluates the quality of binary classification performed by the model (Ravuri et al., 2021).

The difference between the model's performance in predicting convective and partly convective (mixed) precipitation is conducted for calibrated data only because there are substantially more cases of purely convective precipitation among events included in this data set. Both metrics are applied for this part of the analysis.

4. Results

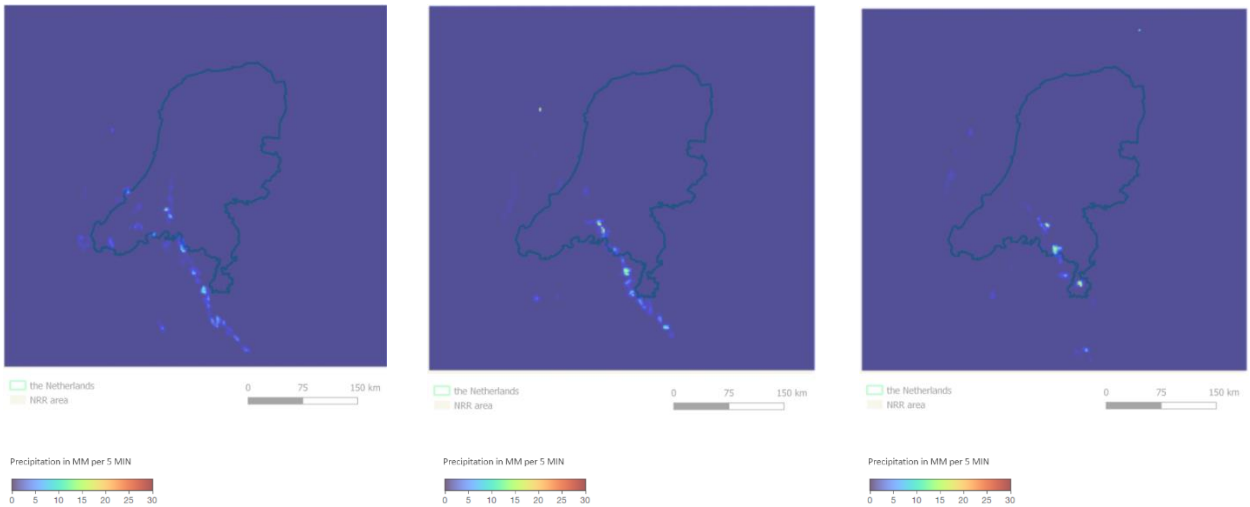
The DGMR was applied to data concerning each of the events, without further training. For each event's data 4 timestamps (20 minutes) were used as input of the model, a context, based on which the DGMR creates an output time series of 18 timestamps (90 minutes). These number of input and output timestamps are fixed by the DGMR. The model was run once for each event. The eighteenth timestamp was always the one where in the NRR the maximum values of rainfall during the event occurred (measured in millimeters per 5 minutes). For each of the events an animation presenting the nowcast and observed values can be printed.

Observed values



T + 30 min T + 60 min T + 90 min

Nowcasted values



T + 30 min T + 60 min + 90 min

Figure 5 – Precipitation event 5th – 6th June 2019. Observed and nowcasted values.

The example above (figure 5) visualizes 3 chosen timestamps of an event which took place from 5th to 6th June 2019. It is a representative case of the whole sample and the comparison between

the observed and forecasted rainfall suggests that DGMR might give accurate and precise picture of the precipitation event.

4.1. Nowcast accuracy and lead time

The change in GDMR's predictive power over lead time was assessed with the level of MSE achieved by the model at timestamps. . Figure 6 presented below shows MSE plot for an exemplary event on 23rd of June, 2016.

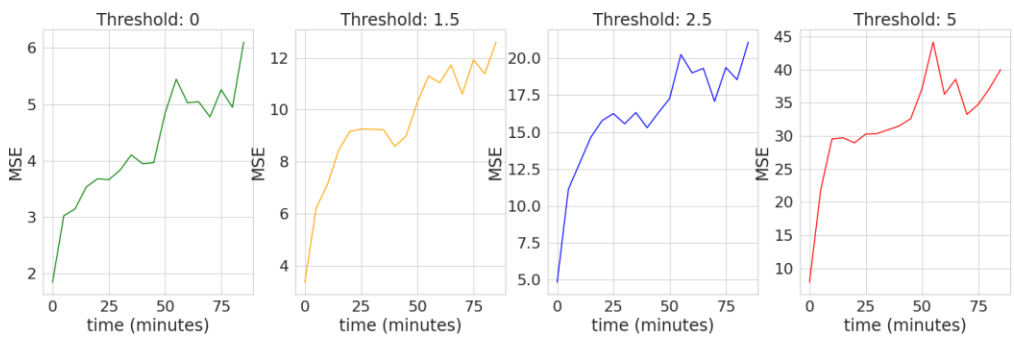


Figure 6 – MSE change over lead time for the event on 2016-06-23, for four sets of data points selected with different thresholds.

All line plots depicting this metric were printed for 12 events and are all placed in Appendix 1. Subplots present MSE calculated on different set of selected data points, accordingly to thresholds described in section 3.4.1. In the line plots in Appendix 1 it is visible that MSE tends to increase together with lead time which means that the quality of nowcasts decreases. The process is non-linear. The most sudden surge of MSE occurs typically between the first and fifth forecasted timestamp and anomalies such as MSE decreasing with time occur mostly in case of events for which real-time data was included instead of calibrated one

A pattern occurring in the data even more consistently is that the level of MSE increases together with the selection threshold. The more extreme the precipitation (and the fewer data points) the higher MSE. Nevertheless, trend of the MSE value over timestamps, indicated with the shape of the line, is always, for each event, consistent for all 4 thresholds.

4.2. Performance of the GDMR model in nowcasting convective and mixed events – comparison

In order to have a set of MSE values which can be compared between purely and partly convective events, only events where calibrated data was used is included. The choice is based on the fact that the numbers of convective and mixed events are more balanced in this data set (4 mixed and 3 convective).

Table 2 presents levels of MSE for the two types of event. MSE is calculated for each of the four thresholds.

Threshold	**Mean MSE for purely convective events**	**Mean MSE for mixed events**
Over 0	2.9	2.8
1.5	6.99	6.75
2.5	13.82	12.25
5	29.37	24.48

Table 2 – Comparison of MSE aggregated for purely convective events and partly convective events.

For each of the thresholds, the MSE value calculated for convective events is higher than for partly convective events. It means that the predictive power of GDMR is lower for convective events as expected.

Date	**CSI Threshold = 2**	**CSI Threshold = 2.5**	**Precipitation type**
2016-06-02	.66	.86	mixed
2016-06-23	.62	.82	mixed
2017-08-20/30	.71	.89	mixed
2017-09-11/12	.79	.91	mixed
2018-05-29	.65	.83	convective
2019-06-05/06	.57	.77	convective

2020-08-11	.55	.66	convective

Table 3 - Critical Success Index (CSI) computed per event, for all events where calibrated data was used.

Similar trend occurs when CSI levels for both types of precipitation is compared. Table 3 shows CSI calculated per event with 2 different thresholds indicated on section 3.4.2.

CSI levels reported in Table 3 show that on average CSI for convective events have lower value than for mixed events. Based on the table, mean MSE for partly convective events (threshold = 2) is 0.69, while mean MSE convective ones with the same threshold is 0.59. The result is similar for the threshold of yellow warning: mean MSE for partly convective events is 0.87 and for purely convective: 0.75.

Therefore, the analysis of both metrics, MSE and CSI, suggest that nowcasting for purely convective events is more challenging for the DGMR than nowcasting for partly convective ones.

5. Discussion & Conclusions

The research question addressed by this study concerns the abilities of the DeepMind's model in nowcasting extreme summer precipitation in the Netherlands. The answer is given through an analysis of the model's outputs with regard to two subquestions.

The first sub-question was how does the predictive power of the model decrease with the increasing lead time. To answer this question, plots of MSE levels for a number of timestamps were used. It allows to study how the accuracy of the prediction changes with lead time and it is relatively consistent over the events where calibrated data was used that the quality of the prediction drops significantly already within the first 20 minutes (5 timestamps) of prediction. It can serve as an indication for possible real-life applications where nowcasts for more than 20 minutes ahead should be considered less reliable than for the first 5 timestamps. However, for real-time data the MSE behaves less regularly over the timestamps. It indicates to one of the limitations of the study which is no access to real-time NRR data for some of the included events. Analysis on the calibrated data sets is less valuable from real-life perspective because at a time when an extreme event happens, only the non-calibrated version of the data exists. Thus, the

knowledge of the model's performance for a larger real-time, uncalibrated data set about would be potentially more applicable than an analysis of calibrated data.

Another limitation for lead time analysis concerns the number of data frames used as input of the model and the lead time of nowcast. The DGMR model which nowcasts only up to a lead time of 1.5 hour, allows to explore only a relatively short period of time that meets the definition of nowcasting which could be even up to 6 hours, according to some definitions (Schmid et al., 2019). The number of input and output data frames is forced by the DGMR model's technical specification, therefore the analysis of quality of nowcast with increasing lead time is possible only up to 90 minutes. However, there is no access provided to the functions where these variables are fixed, and therefore, it is not possible to extend the lead time.

However, the DGMR's performance can also be measured with CSI which is a method potentially more suitable for some decision-making applications than MSE. This kind of situation might occur e.g., in open-air events planning. If there is to be decided whether to cancel an event or even to evacuate participants, the information that is most needed is the accuracy of a model to classify whether a worrisome threshold of rainfall will be met or not. Such level might be set e.g., at the first, yellow, KNMI warning. The DGMR can potentially be useful in such a setting. Its CSI for calibrated data with 2.5 mm (yellow warning) as a threshold, in 5 out of 7 events was above 0.8. It can be interpreted as accuracy of over 80%. The outcome was reached for the nowcasts including the entirety of lead times. It is a promising result and in future research CSI for the DGMR could be tested also for a number of events including in real-time data.

At the same time, MSE and its vulnerability to prevalence of particular values in the data set could be further studied as well. It can be done e.g., in a form of plotting histograms of differences between observed and nowcasted values. It could give an information about the distribution of nowcast's quality for different levels of rainfall. Then values with the largest differences, which may be the ones with low rainfall levels, could be excluded from the sample and the MSE could be measured for these with a level that matters in a particular kind of situations.

The second sub question analyzed in the study is how the model performs in purely convective events nowcasting in comparison to its performance in partly convective (mixed) events nowcasting. As mentioned in section 1.1., forecasting of convective events is a challenge for the existing methods. It also applies to the DGMR. The difference in model's accuracy for purely and

partly convective events can be noticed not only when a continuous measure such as MSE is applied. Also in binary classification based on the DGMR's nowcast, worse results are obtained for convective than for mixed type rainfall.

However, this information does not apply to real-life decision making such as crisis management. It can be applied in further research about nowcasting as an indication that the challenge still exists, and that there is a difference in the forecast accuracy also between purely and partly convective rain.

The main limitation in the analysis of the topic is scarcity of extreme events and even bigger ones of purely convective cases. Therefore, the data sets on precipitation are imbalanced. Evidently, a part of the definition of an extreme event is its rareness. Accordingly, a data set including an extreme event still consists mostly of data points representing non-extreme precipitation. It means scarcity of data about the events of interest and it might have an impact on the values of evaluation metrics.

Imbalance of extreme and non-extreme observations in the data sets could be addressed by the use of Generative Adversarial Networks which are models based on the same concept of two competing neural networks, as DGMR. They are commonly used to produce artificial data points of a type that is scarce in the original data set (Belderbos, 2021). This is one of possibilities for future studies.

Nevertheless, the DGMR proves applicable to the NRR data and the study can be used as a proof of concept for further research. As expected, the quality of the nowcast drops as the lead time increases. It occurs more consistently for events where calibrated data was used than for the ones analyzed based on real-time data. The second hypothesis, assuming that purely convective rainfall is more difficult for DGMR to be nowcasted than partly convective rainfall, is also correct. However, a comparison with other models for the same data, and running the model for a bigger data set of extreme events is needed. The limitations could possibly be addressed and overcome in future research concerning nowcasting.

6. Bibliography

6.1. Sources of data and the model

DeepMind (2021). *Skillful precipitation nowcasting using deep generative models of radar*. https://github.com/deepmind/deepmind-research/tree/master/nowcasting

Nationaleregenradar, https://nationaleregenradar.nl/

6.2. References

Adaloglou N. (2020). *Vanilla GAN (GANs in computer vision: Introduction to generative learning).* AI Summer. https://theaisummer.com/gan-computer-vision/#vanilla-gan-generative-adversarial-networks-2014

Becker T., Bechtold P., Sandu I. (2019). *Modelling convective precipitation – progress and challenges.*https://www.ecmwf.int/en/about/media-centre/science-blog/2019/modelling-convective-precipitation-progress-and-challenges

Breugem A.J., Wesseling J.G., Oostindie K., Ritsema C.J. (2020) *Meteorological aspects of heavy precipitation in relation to floods – An overview*, Earth-Science Reviews, Volume 204, https://doi.org/10.1016/j.earscirev.2020.103171.

DeepMind (2021). *Nowcasting the next hour of rain*. https://www.deepmind.com/blog/nowcasting-the-next-hour-of-rain

Gadian, A. & Blyth, Alan & Bruyere, Cindy & Burton, Ralph & Done, James & Groves, James & Holland, Greg & Mobbs, Stephen & Pozo, Jutta & Tye, Mari & Warner, James. (2017). *A case study of possible future summer convective precipitation over the UK and Europe from a regional climate projection: EUROPEAN CONVECTIVE PRECIPITATION.* International Journal of Climatology. https://doi.org/10.1002/joc.5336

Koninklijk Nederlands Meteorologisch Instituut (KNMI) (2021). https://www.knmi.nl/nederland-nu/klimatologie

Kooij E.C. (2021) Thesis TU Delft. *Nowcasting heavy precipitation in the Netherlands: a deep learning approach.*

Moore, B.J., Neiman, P.J., Ralph, F.M., Barthold, F.E., (2012). *Physical processes associated with heavy flooding rainfall in Nashville, Tennessee, and Vicinity during 1–2 May 2010: the role of an atmospheric river and mesoscale convective systems.* Mon. Weather Rev. 140, 358–378. https://doi.org/10.1175/MWR-D-11-00126.1

Muñoz C., Wang L.P., Willems P. (2018), *Enhanced object-based tracking algorithm for convective rain storms and cells*, Atmospheric Research, Volume 201, Pages 144-158. https://doi.org/10.1016/j.atmosres.2017.10.027.

Ravuri, S., Lenc, K., Willson, M. *et al.* (2021). *Skillful precipitation nowcasting using deep generative models of radar*. Nature 597**,** 672–677. https://doi.org/10.1038/s41586-021-03854-z

Reynolds C., Williams K., Zadra A. (2019). *WGNE Systematic Error Survey Results Summary 11 February 2019. https://www.wcrp-climate.org/JSC40/12.7b.%20WGNE_Systematic_Error_Survey_Results_20190211.pdf*

Šaur, D. (2017). *Forecasting of Convective Precipitation Through NWP Models and Algorithm of Storms Prediction*. Doi: 10.1007/978-3-319-57261-1_13.

Schmid, F., Wang, Y., Harou, A. (2017). *Guidelines for Nowcasting Techniques*. World Meteorological Organization. https://library.wmo.int/doc_num.php?explnum_id=3795

Schmid, F., Wang, Y., Harou, A. (2019). *Nowcasting Guidelines*. World Meteorological Organization. https://public.wmo.int/en/resources/bulletin/nowcasting-guidelines-%E2%80%93-summary

Schuurmans H., van Vossen J. (2013). *Nationale Regenradar toelichting operationele neerslagproducten*,NoorderzijlvestBA8186. https://nationaleregenradar.nl/pdfs/hoofdrapport_NRR_definitief.pdf

Sun, Juanzhen & Xue, Ming & Wilson, James & Ballard, Sue & Onvlee, Jeanette & Joe, Paul & Barker, M & Li, Ping-Wah & Golding, Brian & Xu, Mei & Pinto, James & Sun, & Wilson, & xu, Dongmei. (2013). *Use of NWP for Nowcasting Convective Precipitation: Recent Progress and Challenges*. Bulletin of the American Meteorological Society. 95. 409-426. https://doi.org/10.1175/BAMS-D-11-00263.1

Wapstra, L. (2016). Thesis Utrecht University, Faculty of Geosciences. *The performance of rainfall nowcasts from Nationale Regenradar for extreme-event rainfall and runoff prediction. A quantitative analysis using FEWS and 3Di.*

Westra, S., Fowler, H.J., Evans, J.P., Alexander, L.V., Berg, P., Johnson, F., Kendon, E.J., Lenderink, G., Roberts, N.M., 2014. *Future changes to the intensity and frequency of short-duration extreme rainfall*. Rev. Geophys. 52, 522–555. https://doi.org/10.1002/2014RG000464

7. Appendix 1

7.1. Based on calibrated data

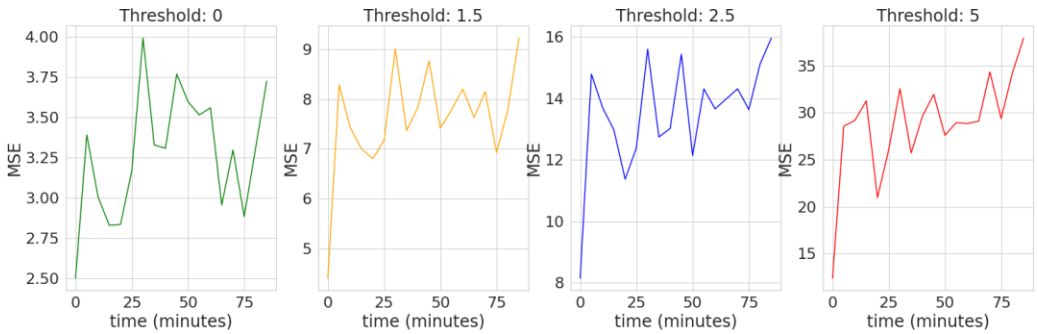


Figure 7 – MSE change over lead time for the event on 2016-06-02, for four sets of data points selected with different thresholds.

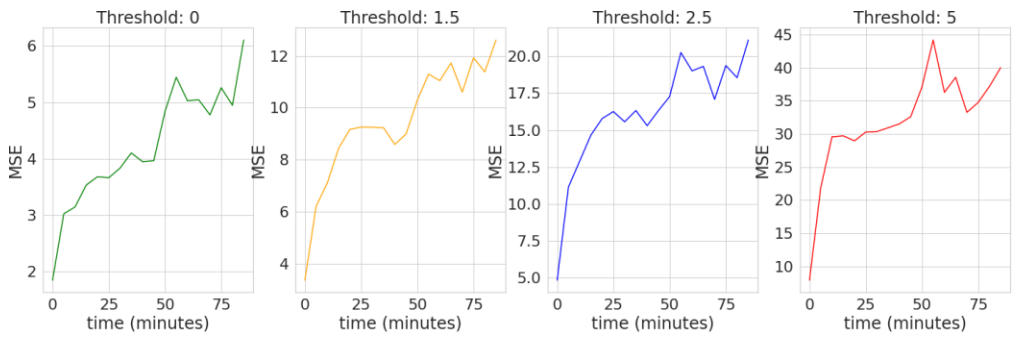


Figure 8 – MSE change over lead time for the event on 2016-06-23, for four sets of data points selected with different thresholds.

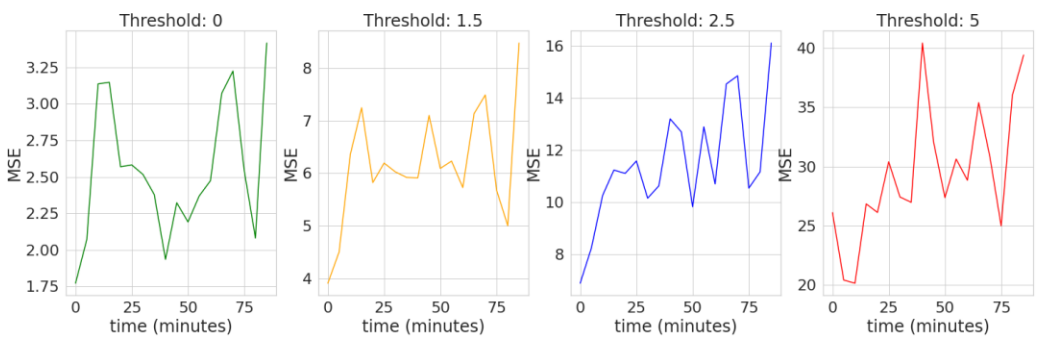


Figure 9 – MSE change over lead time for the event on 2017-08-29/30, for four sets of data points selected with different thresholds.

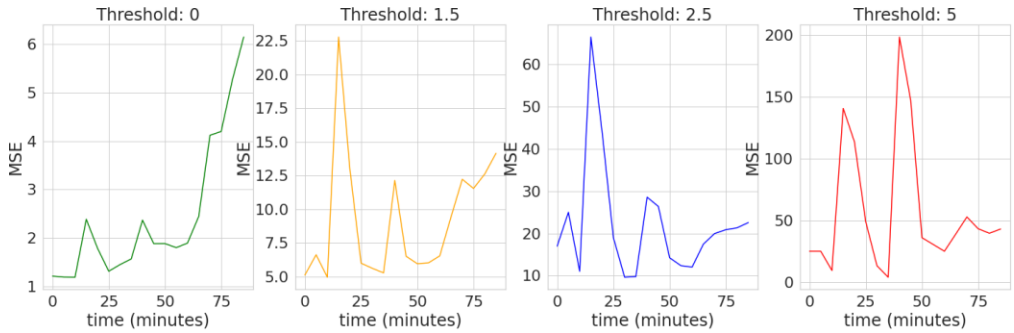


Figure 10 – MSE change over lead time for the event on 2017-09-11/12, for four sets of data points selected with different thresholds.

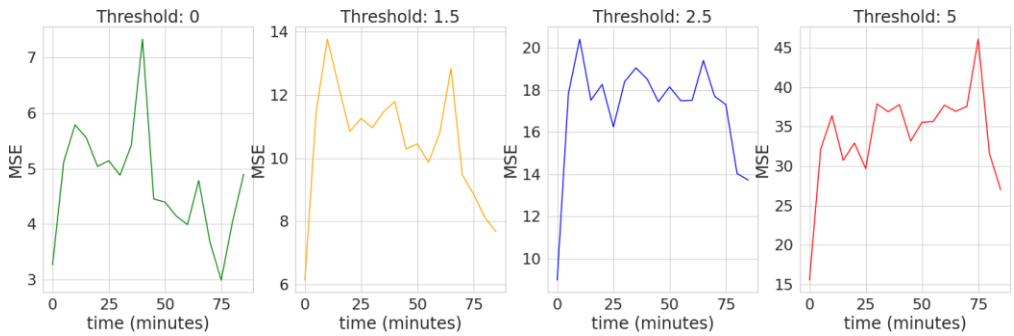


Figure 11 – MSE change over lead time for the event on 2018-05-29, for four sets of data points selected with different thresholds.

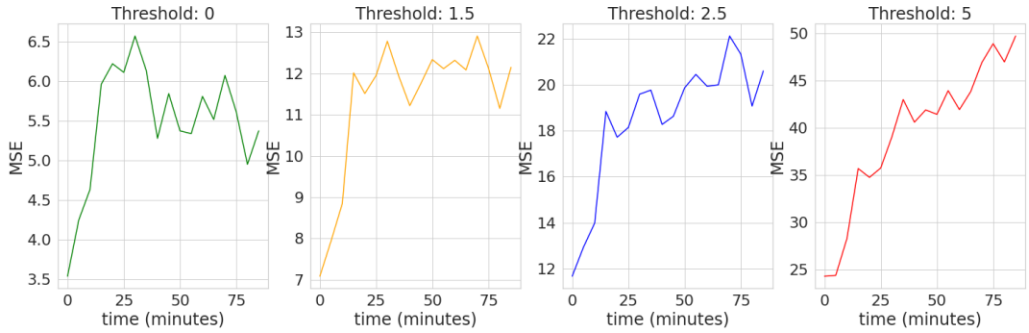


Figure 12 – MSE change over lead time for the event on 2019-06-05/06, for four sets of data points selected with different thresholds.

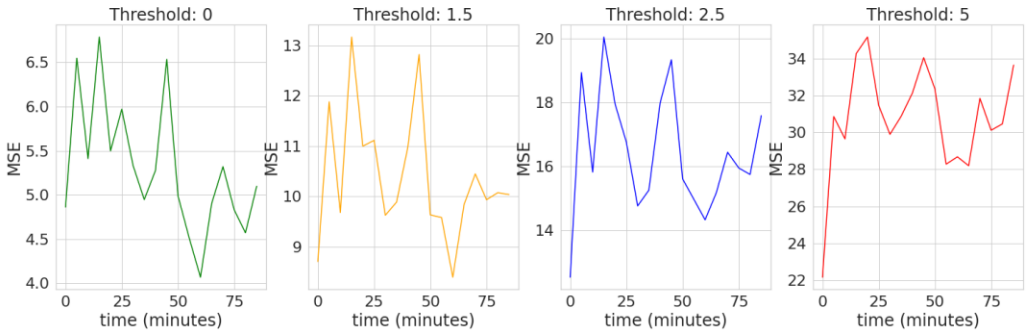


Figure 13 – MSE change over lead time for the event on 2020-08-11, for four sets of data points selected with different thresholds.

7.2. Based on real-time data

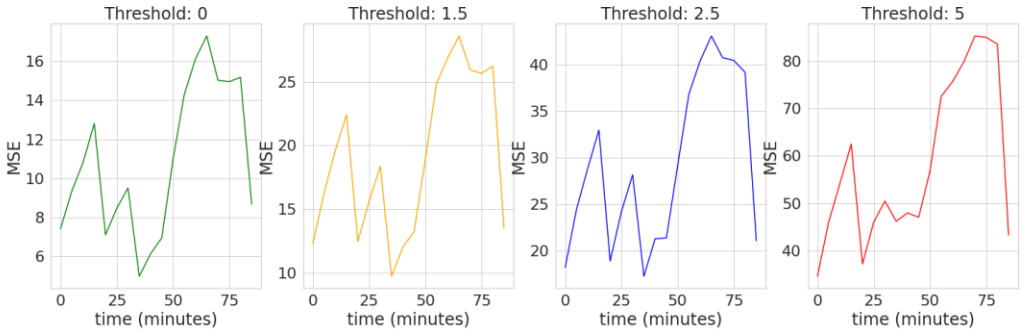


Figure 14 – MSE change over lead time for the event on 2021-06-18, for four sets of data points selected with different thresholds.

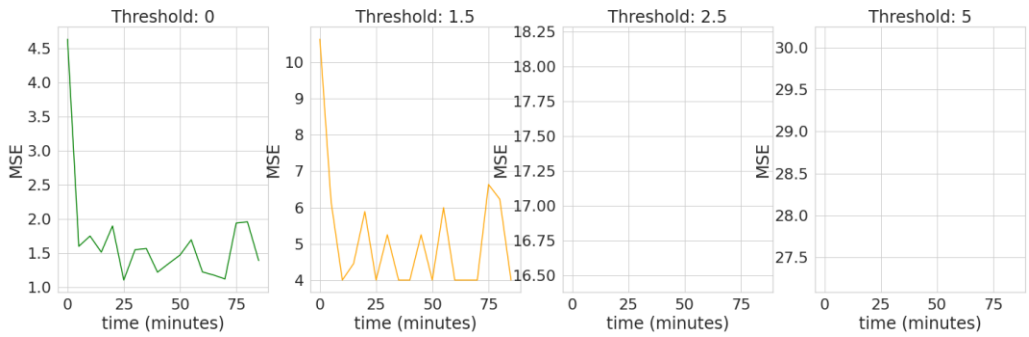


Figure 15 – MSE change over lead time for the event on 2021-06-20, for four sets of data points selected with different thresholds.

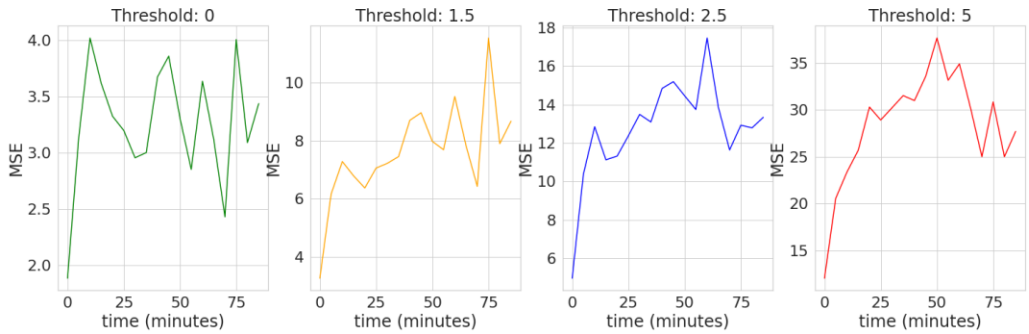


Figure 16 – MSE change over lead time for the event on 2021-08-21/22, for four sets of data points selected with different thresholds.

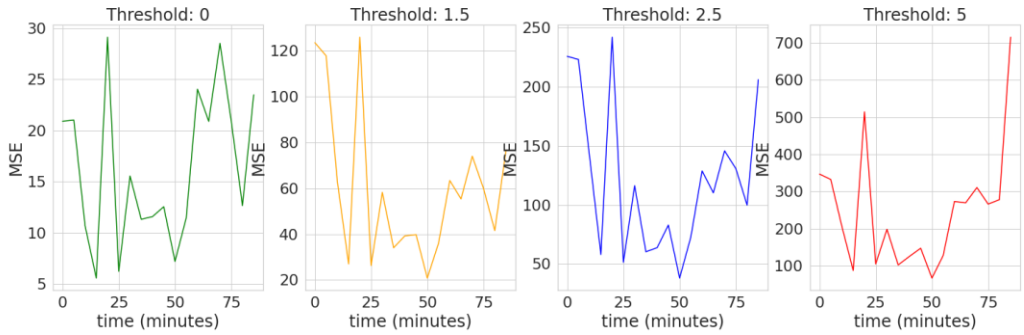


Figure 17 – MSE change over lead time for the event on 2022-06-05, for four sets of data points selected with different thresholds.