High resolution forecasts of short duration extreme precipitation - forecasting quality and sensitivity to parameterization for two Norwegian cases

Master thesis in meteorology

Ingvild Thynes Villa

June 2021

University of Bergen
Faculty of Mathematics and Natural Sciences
Geophysical Institute

Abstract

Due to global climate change, the intensity and frequency of short duration extreme precipitation are expected to increase. Short duration extreme precipitation events in Norway are often convective precipitation happening during the summer. In this thesis, the short duration extreme precipitation events in Jølster (2019) and Nøtterøy (2020) were investigated. The goal was to examine what happened the day of the events and how well they were forecasted—also bringing in a discussion on how to do verification of rare locale extreme events. The verification revealed that the radar observation was deviating strongly from the point observations for both events, which impacted the verification result. The neighborhood verification for Jølster indicated that the overall forecast for the area around Jølster was good. However, the "nearest point" verification method showed that no high values were forecasted at Jølster. This method also showed that some of the ensemble members for Nøtterøy were very good. Simulation with the WRF model showed that changes in turbulence and microphysical schemes had a minor impact on the forecasted event. Indicating that both of the events were mainly large-scale driven.

Acknowledgements

I would like to thank my supervisor, Asgeir Sorteberg, who guided me through writing an MSc thesis during a global pandemic. With home office and remote communication, there where always time to answer questions and help me.

Also, a big thanks to my co-supervisor, Kjersti Konstali. Thanks for helping me running WRF (I would not have managed without you), answering stupid questions, and reading my thesis one time too many.

A huge thanks to my main squad at GFI; 2 hour-long lunch breaks and "halv dag" made the six years at GFI go by way too fast. It would never have been the same without you guys.

And, last but not least, thanks to my family for always supporting me. I would never have been able to combine soccer and a Master's degree without your help and full support.

Contents

1. Introduction

In the last 100 years, the annual precipitation in Norway has increased by 18% [Hanssen-Bauer et al., 2015]. We see the trend in all the seasons, but a bit less for the summer. Parts of this trend can be directly linked to the increase in temperature through the Clausius Clapeyron equation, but not everything. Running global climate models, simulating the climate change the last 150 years, only shows around 1/5 of the increase observed. This can either be explained by the fact that the model is not good enough simulating climate change sensitivity or that many observed changes are not related to human-made emission and natural influences (ex. solar radiation), but chaotic multidecadal variations [Sorteberg, 2014].

The positive trends are not only present in the total accumulated precipitation in Norway, both the frequency and intensity of short duration rainfall have increased at most Norwegian measuring sites in the last fifty years as well [Førland E. J., 2018]. In Norway, two main reasons are causing extreme precipitation. The first one is a low pressure coming in from the Atlantic ocean bringing heavy rainfall along the west coast. The second reason is heavy showers called convective precipitation, which mainly occurs during the summer [Michel et al.]. Predicting these extreme events becomes more and more crucial as the frequency and intensity increase. An increase in these events can potentially lead to an increase in flash floods' magnitude and frequency. There is a need for more knowledge to help society build up infrastructure and buildings prepared to face the challenges these weather events bring [Westra et al., 2014].

Norway is 1699 km long, from the northernmost to the southernmost point, and covers several climate zone. We have, in general, a warmer climate than other places at the same latitude. This is due to the heat brought by the North Atlantic current and the low pressures coming in from the Atlantic, bringing warm air northward [Seager et al., 2002]. The low pressure does not only transport heat but also moisture. The climate in Norway is quite moist, with much precipitation compared to the global average. The average amount of precipitation in Norway is about 50% higher than the earth's average.

Weather forecasting in the Nordic regions spans a wide range of phenomena and scales. Norway includes continental, maritime, and polar conditions, together with complex coastlines and varying topography on land. All of these factors imply local variations in weather. Regional variation benefits from a good description of small-scale phenomena and forcing [Müller et al., 2017].

Knowing how a changing climate is expected to lead to more frequent short duration extremes and the challenges with forecasting these events, I wanted to look closer into this topic. I ended

up investigating two different events of extreme precipitation in Norway. I assess how well they were forecasted, bringing in a discussion on verification methods of rare, extreme events. Running the Weather Research and Forecasting (WRF) model, I wanted to investigate the different events sensitivity to turbulence (planetary boundary layer) and cloud physics (microphysical) numerical scheme. Is the simulation of extreme events sensitive to the choice of numerical schemes? This will give us an impression of the leading mechanism of the events. If changes in turbulence schemes only give small changes in the event's outcome, we can assume that turbulence is not a primary driving mechanism for the event.

The different events happen in 2019 in Jølster and at Nøtterøy in 2020. Both events happened during summertime and can be considered short duration events. However, the synoptic situation was different, together with the extent in space and time for the various events. For the event in Jølster, a large amount of precipitation was observed over a large area and longer duration. At Jølster the event had dramatic consequences, leading to a casualty caused by a landslide. As the events get more and more intense due to climate change, these events can be expected to happen more frequently. Nøtterøy was a very local event, and the event in Nøtterøy gave a new Norwegian record in the amount of accumulated precipitation measured in one hour. The hour before and the hour after, almost no precipitation was measured, so this event had a very short duration.

The thesis is divided into eight chapters. The first chapter is the introduction. Chapter 2 including relevant theory about extreme precipitation, verification and stability parameters. Chapter 3 introduces the two different numerical models; WRF and AROME. Chapter 4 includes a description of radar and information about the lightning and ground observation used. Chapter 5 explains the method used to verify the forecasts. Chapter 6 includes all of the results. In chapter 7 the results are discussed, and chapter 8 brings it all together with a summary of my findings and some thoughts about future work.

2. Theory

2.1 Extreme precipitation

To form precipitation water vapor needs to condense into cloud droplets. Further condensation leads to growth of the cloud droplet turning them in to rain droplets. When these rain droplets are sufficiently big, gravity makes them fall towards the surface, resulting in precipitation. For the condensation to take place, the air needs to be saturated. There are two ways the air can become saturated, either by adding moisture or by cooling the air [Hakim, Gregory J. Patoux, 2017].

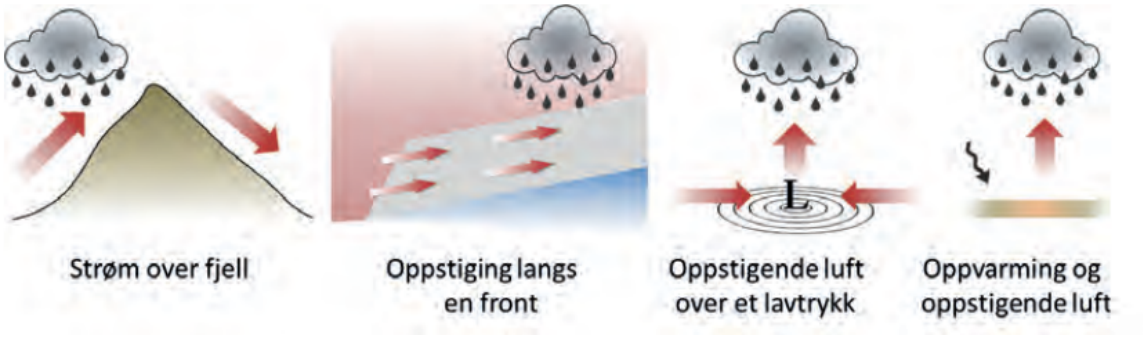

Figure 2.1: Precipitations can be provoked by the four different weather situations showed. First one show stream over a mountain, second show ascent along a front, third show ascent air over a low pressure and the fourth show ascent due to warming [Sorteberg et al., 2018].

In general, we can say that there is typically four different weather situations that lead to precipitation (Figure 2.1): Airflow over a mountain, ascent along a front, ascent over a low pressure and heating leading to ascent. Ascent causes the air to cool. When the air cools, its ability to hold moisture is reduced. This can lead to saturation of the air, and subsequent condensation. This leads to the formation of clouds, which again leads to precipitation [Sorteberg, 2014]. The amount of water vapor the air can hold before forming clouds is decided by the *Clausius-Clapeyron* equation. The equation gives the relation between temperature and saturation specific humidity. For typical air masses over Norway, the equation shows that a reduction in the temperature of 1K reduced the amount of water vapor the air can hold by 6-7% .

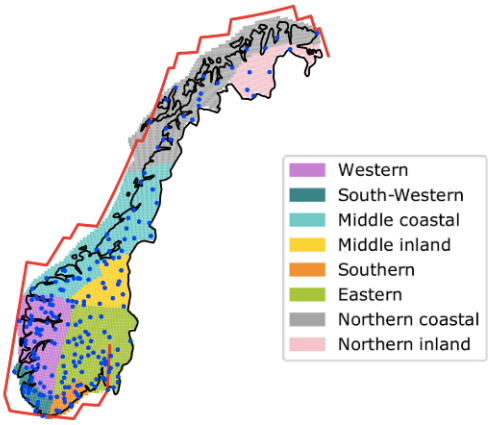

Figure 2.2: Overview of different regions in Norway [Michel et al.].

We can divide precipitation into two main types; dynamic and convective precipitation. Convective precipitation occurs when air rise vertically by convection. Dynamical precipitation occurs when air rises by a dynamical force. A dynamical force can for example be a front or a surface pressure system. While dynamic precipitation only needs lift and saturated air, convective precipitation needs instability in addition. Instability occurs when the parcel of air is lighter than its surroundings. Instability can happen either if a fluid is cooled from above or heated from below. The atmosphere is unstable when warm and light air is situated under colder and heavier air. If the air close to the ground is lighter than its surroundings, it will rise and cool, and condensation occurs. Formation of cumulus and cumulonimbus clouds can then find place. Convective precipitation is short-lived and often results in very local precipitation, which is the case when we have events of short duration extreme precipitation.

Convective storms are generated by moist air, lift and instability. Convective storms can be categorized into different groups. The simplest one are weak storms that grows and die within an hour, called single cells. They do not necessarily produce severe weather, but are important because they are the building block for the larger convective storms. For convective storms to become severe a significant increase of wind speed, or change in direction with height, is needed. This gives the storm system a tilt, which avoids the cancellation between the updraft and downdraft as they are in different locations. When the updraft and downdraft happens in different regions, this both reduces the water loading in the updraft and there will be a positive feedback between up and downdraft (Figure 2.3).

Short duration extreme precipitation is often damaging due to the intensity of the event. The intensity of the event is dependent on three different processes. The first process is how much of the water vapor is condensed. This depends on how fast the water vapor condensates and how much water vapor there is available. The second process is the ability to gain more moisture. To be able to continue to condensate, the air is dependent on moist air to be added into the process of forming precipitation. The third, and last process, is how efficient the formation of precipitation is.

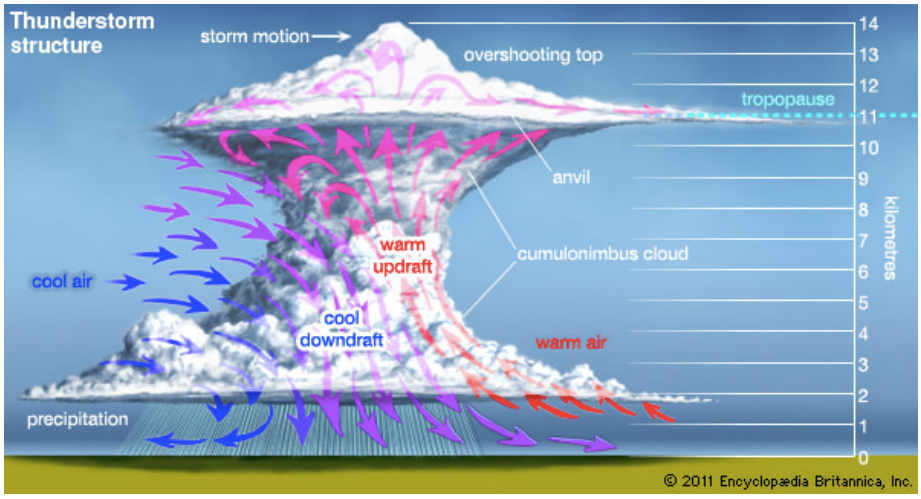

Figure 2.3: Illustration of a thunder storm [Krider, 2019].

For warm clouds, rain droplets are formed by a process known as collision and coalescence, forming droplets big enough for gravity to bring them down. For cold clouds, the Bergeron Findeisen process is forming droplets by growing ice crystals. When the ice crystals get large enough, they fall down and can then collide with other ice crystals, giving further growth. How efficient these processes are, depend on the microphysics in the clouds, how the droplet sizes are distributed in the clouds, and if there are clouds above. If there are clouds above with precipitation, the droplets can collide with the droplets in the lower clouds and bring them with them on their way down [Sorteberg et al., 2018].

In some cases, the updraft needed for the air close to the ground to rise is not present. We then have a conditionally unstable situation. The atmosphere is stable for a saturated air parcel, but unstable for dry air [Leffler, 2015]. This is due to differences in the wet and dry adiabatic lapse rate. We then need a release mechanism, for example topography, that forces the air to rise. Condensation and rainfall is then triggered. This is called a trigger mechanism. These trigger events can cause extreme amount of precipitation [Houze, 1997].

Figure 2.4 shows the distribution of the 60 minute and 360 minute 20 year return value (precipitation that statistically is predicted to occur one time every 20 year - *for more details see Section 2.3*) in Norway. There is a clear difference in where the extremes are expected to be highest for the two different time intervals. For the one hour return value South Norway stands out with values around 30 mm/h for a larger area. For the 6 hours return value the area with the highest expected return values are on the west coast in southern Norway. But we also observe that there is a minor area on the west coast of southern Norway with values up to 30 mm for one hour return value. Thunderstorms are not as common in Norway as further south in Europe , but during the summer months in the afternoon, especially in Agder, Telemark, and Østlandet it is common [Reitan, 2005].

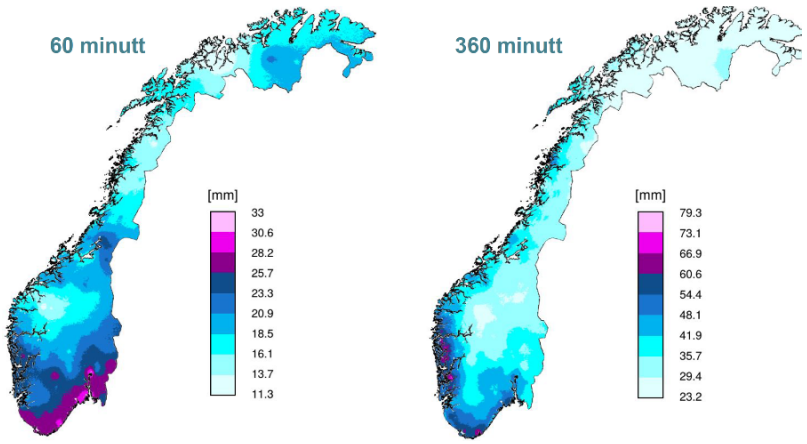

Figure 2.4: One hour (left) and six hours (right) accumulated precipitation with return period of 20 years (precipitation that statistical can occur one time every 20 year), based on available observation [Sorteberg et al., 2018].

2.2 Stability parameters

Stability parameters are a measure of the static stability in the atmosphere and are used in meteorology to quickly identify regions with a high probability for the development of thunderstorms or heavy rainfall. There are many different stability parameters, which all have different advantages and disadvantages. One common disadvantage for all of these stability parameters is discussed by Brooks [2009]; they do not take into account the local variations and different seasons. Knowledge about the area in consideration of stability parameters, is important.

When evaluating stability indices, it is important to take into consideration that no stability index functions perfectly well by itself. Best usage of the indices is by evaluating them together with other information. They are not able to fully describe the state of the atmosphere, and alone they can therefore give misleading results [Schultz, 1989]. They do not take into account the trigger mechanisms for lifting, which for conditionally unstable situation is very important. They only represent the potential and latent instability of a layer in the atmosphere [van Zomeren and van Delden, 2007]. However, this information is very useful together with other information and is therefore commonly used for detecting thunderstorms.

Convective available potential energy (CAPE) is the amount of buoyant energy available to produce strong updrafts. CAPE is a well known stability index and is widely used when evaluating the stability of the atmosphere [Ahrens, 2015]. When calculating CAPE we assume parcel theory and look at the difference between the temperature of the parcel and the environment (Equation (2.1)). A parcel of air is defined an imaginary small body of air. Parcel theory assumes that the air parcel retains its shape and general characteristics as it moves up (or down in the atmosphere). This assumption will have errors when calculating for the atmosphere but will work well in a core of a thunderstorm updraft. Inhibit convection (CIN) is, in many ways, the opposite of CAPE

(Equation 2.2). While CAPE represents the positive energy area, CIN represents the negative area. This is the amount of negative buoyant energy that is available to suppress the upward vertical acceleration. The larger the CIN, the stronger force is needed to lift the parcel to the level of free convection (LFC). If we have high values of CIN this can suppress convective development, despite high CAPE values [Pier, 2018]. Table 2.1 and 2.2 show what stabilities and probabilities of thunderstorm development to expect for different values.

$$CAPE = -R_d \int_{LFC}^{EL} T_{parcel} - T_{env} \; dln(p) \tag{2.1}$$

$$CIN = -R_d \int_{SFC}^{LFC} T_{parcel} - T_{env} \; dln(p) \tag{2.2}$$

CAPE	**Stability**
<0	Stable
0 to 1000	Marginally unstable
1000 to 2500	Moderately unstable
2500 to 3500	Very unstable
>3500-4000	Extremely unstable

Table 2.1: Stability categories associated with different values of CAPE.

CIN	**Probability of thunderstorm development**
<50	High changes of thunderstorm development.
50 to 200	Moderat changes of thunderstorm development.
>200	Risk of thunderstorm low.

Table 2.2: Stability categories associated with different values of CIN.

Total Totals (TT) index is another stability parameter commonly used. TT index consists of the Vertical Totals and the Cross Totals (Equation 2.3). The Vertical Totals are the difference between the temperature at 850 mb and the temperature at 500 mb. The Cross Totals are the difference in the dew point temperature at 850 mb and the temperature at 500 mb. This means that the TT index takes both the static stability at 850 mb (Vertical Totals) and the moisture at 850 mb (Cross Totals) into account. Table 2.3 show what the different TT indices indicate of probability for thunderstorm. One of the challenges with this index arise if the moisture is below the 850mb level. Then the index would not be representative of the instability in the atmosphere [Khole and Biswas, 2007].

$$TT_{index} = T(850mb) - T(500mb) + Td(850mb) - T(500mb) \tag{2.3}$$

K index also uses the dew point temperature at 850 mb and the temperature at 850 mb and 500 mb (Equation 2.4). K index includes the dew point depression at 700 mb, which is the difference

TT index	Thunderstorm Probability
<44	Thunderstorms not likely
44 to 50	Thunderstorms likely
51 to 52	Isolated severe thunderstorms
53 to 56	Widely scattered thunderstorms
56 to 60	Scattered severe thunderstorms more likely

Table 2.3: Probability of thunderstorm with different values of stability parameters TT index.

K index	Thunderstorm Probability
<30	Thunderstorm with heavy rain or severe weather possible
30 to 40	Better potential for thunderstorm with heavy rain
>= 40	Best potential for thunderstorms with very heavy rain

Table 2.4: Probability of thunderstorm with different values of stability parameters K index.

between the temperature and the dew point temperature. In general, the K index is a measure of thunderstorm potential. The higher the K the more likely for heavy rain. Since the formula include the dewpoint depression at 700 mb, dry air at the given level give values of low K. But, given moisture below 700 mb, unstable air, and a lifting mechanism, severe and strong thunderstorm with heavy rain can still occur [Cimini et al., 2015]. Table 2.4 give a overview of a general impression of what to expect for the different values of K.

$$K_{index} = T(850mb) + Td(850mb) - T(500mb) - DD(700mb) \tag{2.4}$$

2.3 IVF-values

Working with short duration precipitation, the so called IVF-value is a very useful value. IVF stands for precipitation intensity (I - *intensitet*), duration (V - *varighet*) and frequency (F - *frekvens/hyppighet*). The IVF value is often expressed with a return-value, estimating a value that can be expected to be exceeded one time during for example 5 years (meaning the "5-years" - return value). This value is very useful when evaluating how rare an event was.

The IVF values used in this thesis are from The Norwegian Climate Service Center. The statistics behind the IVF uses the measurements of precipitation with resolution of one minute from tipping bucket rain gauges for the stations with at least 10 years of history. These measurements are not good at measuring snow, therefore only the data from the summer months are used. The available measurements are used to calculate the return value. Interpolation is used to estimate the values for all of Norway, meaning IVF return value is available everywhere, not only in the location with ground measurements.

There are large uncertainty related to IVF-statistics. The uncertainty mainly originates from the

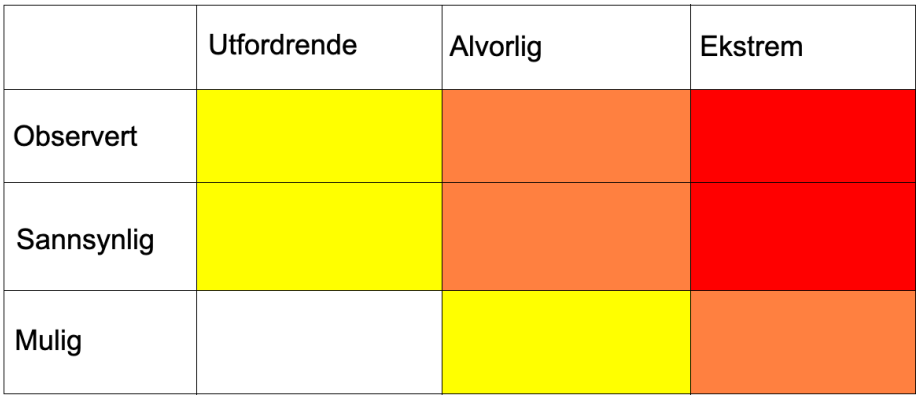

Figure 2.5: Meteorological institute hazard warning methodology. The matrix shows which color the event is categorized dependent on the observation, probability, and possibility of an event. There are three different levels shown; *Observed* has the possibility of 100%, *likely* means possibility >50% and *possible* means 35/40%-50%. For example, if there is a possibility of an extreme amount of precipitation, an orange warning is sent out. However, if it is likely for an extreme event to appear, red is sent out.

density of the ground measurements, the short available time series, methodology and climate change. There are many different statistical methods to calculate the IVF statistics. The decision made when choosing the basic data, type of extreme value distribution, and method for estimating the parameters of the selected distribution varies, and no "best" method is agreed upon. In Norway there are 90 tipping bucket rain gauges with time series longer than 10 years, meaning that there are large areas not covered. These short duration precipitation events are often very local, with low density of measurements the statistic can be very dependent of which events are observed. Another problem is related to increase of precipitation due to climate change. This should be taken into consideration, but because of the lack of data, data and timeseries from for example 1970 is used the same way as the ones from recent time, not considering climate change at all.

2.4 METs hazard warning methodology

During the summer of 2018 the Meteorological Institute (MET) started a new method of hazard warning, which includes warning for short time duration precipitation. This methodology is based on the standard CAP (Common Alerting Protocol), and is commonly used for nature hazards. This method involves warnings divided into colours. The colors are yellow, orange and red, combine degree of color and probability (Figure 2.5). Yellow indicates "be aware" and the weather is categorised as challenging, orange means "be prepared" and means server weather. The last one, and most serious is red, meaning "secure values" and indicates extreme situation [Meteorologisk Institutt, 2019].

Communicating the danger of extreme precipitation is a highly prioritised job at The Meteorological Institute. The Meteorological Institute stated in the report made after the event in Jølster that they have not established a methodology or a way of how to clearly communicate to the user when the weather is of an unpredictable character. The Meteorological Institute is working on a

better methodology that is better at updating the user continuously, close to the event and when the event is in progress [Meteorologisk Institutt, 2019].

2.5 Verification

2.5.1 What is a good forecast?

The question "What is a good forecast?" can seem like an easy question to answer but is in fact very complex. You could say that a good forecast is a forecast that predicts exactly what is going to happen. This defines a perfect forecast. The challenges come when the forecast is not perfect. Since the atmosphere is a chaotic system, a perfect forecast is basically impossible. Is it important that the precipitation is predicted in the right place, or is it more important that the average amount of precipitation is as accurate as possible? Or perhaps the most important part is that it covers the extreme events so people can be warned of possible danger. All these examples show some of the differences and difficulties that may lie in the assessment of what a good forecast is.

Murphy [1993] discuss the complexity of defining what a good forecast is and introduce three different types of goodness. How you define what a good forecast is, impact the result you are seeking when doing research on improving the forecast. In the paper, Murphy also discusses the challenges with establishing a well-defined goal when working on improving the performance of a forecast. The three different types of goodness are named consistency (type 1), quality (type 2), and value (type 3). Type 1, consistency, is based on the correspondence between the forecast and the forecaster's best judgment derived from her/his knowledge base. Type 2 compares the forecast conditions to the observed conditions at the valid time of the forecast. The last type of goodness (type 3) is when the forecast is used to increase economic and/or other benefits. The forecast is considered good when the input from the forecast is put into decision-making processes and, for example, given an economic growth.

In this thesis the verification used is the type 2 goodness. Looking at the correspondence between forecast and observation, meaning that a forecast of high quality has a close correspondence with the observations. The most traditional way is to do computation measures of the overall correspondence between forecasts and observations. The mean absolute error, skill scores, and mean-square error are examples of these measures. This type of verification can be done in different ways. This will be introduced in the next section.

2.5.2 Neighborhood verification

There are different ways to do verification of a weather forecast. Which method to use depends on what data is available, the length of the time series, the purpose of the verification, the resolution of the forecast, etc. The standard method for continuous and categorical verification statistics computed from point match-ups is often a good solution when doing verification. The method has been used for many years and is well understood. But there are some challenges related to this method. As the resolution of the models gets better, the models need to hit within a smaller range [Rossa et al., 2008]. The difference between expecting the model to hit within a few kilometers grid square (high resolution model) compared to for example 10km grid square (non-high resolution model) is large.

When looking at short duration extreme precipitation, we are often looking at convective precipitation. For a model to be able to simulate convection, it needs to have high resolution. It then follows that the forecast has to be more precise to be considered a good forecast. Traditional methods, like point verification as an example, would not give credit to extremes captured by a high resolution model that are just shifted a bit in time or space. But various neighborhood methods can, by looking in the neighborhood around each grid point for the events, give credit to the forecast for "almost" hitting with the prediction [Gilleland et al., 2009].

How important is it that the forecast hits perfect within a range of a few kilometers? As mentioned earlier, it depends on the purpose of the verification. Considering extreme precipitation, it is often hard to say exactly where the event will appear. Often unstable air is observed over a larger area but it is hard to know what will be the trigger mechanism, and therefore hard to know where the event will happen. But with high values of precipitation, exactly what grid square the event will happen might not be the most important thing. Knowing what area is in danger, to be able to send out a warning can be more important. Therefore, it can be useful to use neighborhood verification.

Neighborhood verification uses a spatial window (neighborhood) surrounding the observation point and/or the forecast. There are different ways of treating the points within the window. For example, you can include averaging (upscaling), thresholding, or generation of a probability density function [Rossa et al., 2008]. The thresholding approach is used for this thesis. Looking at extreme precipitation, it is of interest if the forecast is good at forecasting high values. The threshold is set for both the forecast and the observation. This gives the opportunity to classify every point into a hit, false alarm, correct negative, and missed event. By using the thresholding approach, we end up with a dichotomous forecast (See section 2.5.3).

There are different methods of defining the neighborhood. Figure 2.6 show three different methods of determining the neighborhood using the thresholding approach. Figure 2.6 (d), for example, define the neighborhood within a given area, only for the observed events and not for the forecasted event. A hit is recorded if the following two criteria is fulfilled; 1) the forecast value in the point is above or equal to the threshold, and 2) the observed precipitation at any given point in the neighborhood is at, or above the threshold. Comparing this to the method used in Figure 2.6 (e), the neighborhood is defined in the same way, but now also for the forecasted event [Schwartz, 2017] . A hit is recorded if both observation and forecast events occur anywhere in the neighborhood.

The frequency of the event can be very sensitive to both the size of the neighborhood, as well as the value of the threshold [Gilleland et al., 2009]. It can therefore be useful to consider a range of spatial resolutions and intensity thresholds.

2.5.3 Methods for dichotomous (yes/no) forecasts

A dichotomous forecast is a forecast saying either "no, the event will not happen" or "yes, it will happen". Using the thresholding example with extreme precipitation, the forecast either says "yes, it will rain more than or equal amount of the threshold" or "no, it will not rain more than

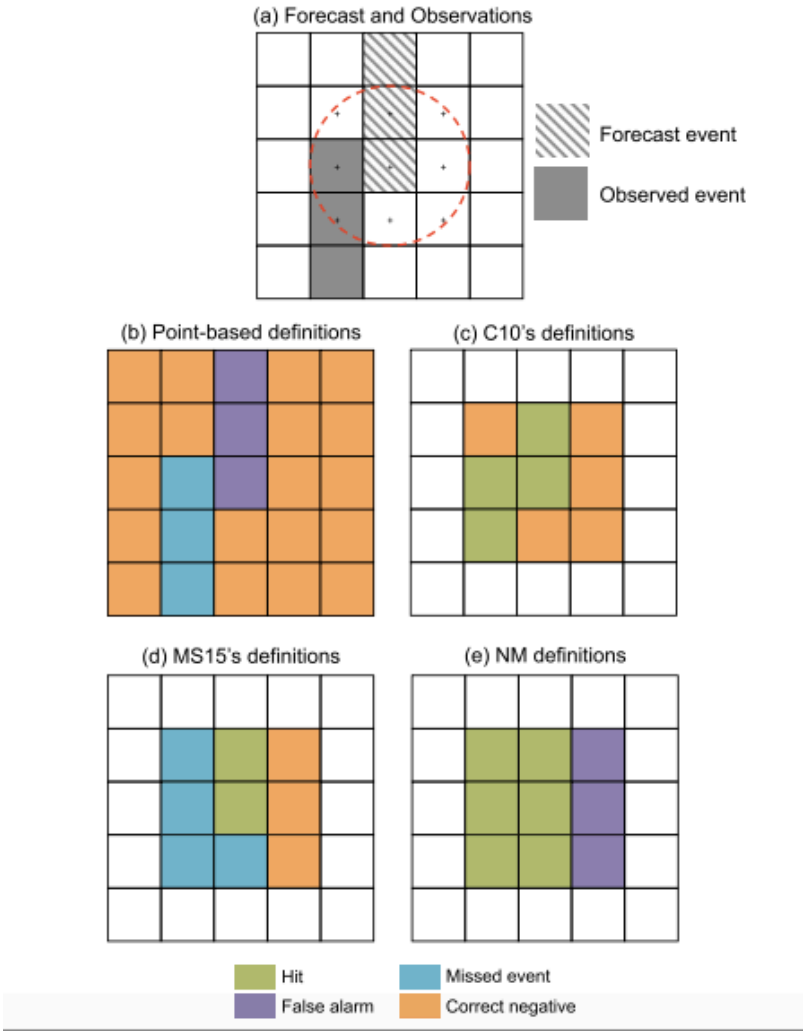

Figure 2.6: Hypothetical (a) forecast and corresponding observations, where forecast (observed) events have occurred in hatched (filled) grid boxes. Contingency table (Table 2.5) classification of the grid boxes based on (a) using (b) point verification, (c)-(e) different methods of neighborhood verification. [Schwartz, 2017]

		Observed		
		yes	no	Total
Forecasted	yes	hits (a)	false alarms (b)	forecast yes
	no	missed event (c)	correct negative (d)	forecast no
	Total	observed yes	observed no	Total

Table 2.5: Contingency table for categorical forecasts of a binary event. [Stephenson, 2000]

the threshold". We use a contingency table to verify these kinds of forecast (Table 2.5). By this definition, a perfect forecast only have hits and correct negatives, and no missed events or false alarms. By using the contingency table, one can easily calculate different categorical statistics [Stephenson, 2000].

Bias score give us an indication of whether the forecast tends to overcast or undercast. It measures the ratio of the frequency of forecast events to the frequency of observed events. The range goes from zero to infinity, where 1 indicates a perfect score. A BIAS below 1 indicates a tendency of undercasting, while a BIAS above 1 indicates overcasting [Stephenson, 2000]. The ones used in this thesis are introduced below.

$$BIAS = \frac{hits + false\_alarms}{hits + misses} \tag{2.5}$$

Hit rate, also named probability of detection, completely ignores false alarms but is sensitive to hits. It is good for rare events and can therefore be useful for extreme precipitation. It compares the ratio between hits and hits and misses. The range is from 0 to 1, where 1 is the perfect score. Values close to 1 indicate that the number of hits is much larger than the number of misses. Since it ignores false alarms, the hit rate should be used in conjunction with the false alarm ratio. [Stephenson, 2000]

$$Hit\_rate = \frac{hits}{hits + misses} \tag{2.6}$$

False alarm ratio (FAR) is sensitive to false alarms but completely ignores the misses. The range goes from 0 to 1, where 0 is the perfect score. FAR close to 0 means that the number of hits is much larger than the number of false alarms [Stephenson, 2000]. The ratio answers the question; what fraction of the "yes" events that were predicted did not occur?

$$FAR = \frac{falsealarms}{hits + falsealarms} \tag{2.7}$$

3. Numerical models

3.1 The Weather Research Forecast Model (WRF)

The Weather Research Forecast model (WRF) is a numerical weather prediction (NWP) and atmospheric simulations system that can generate atmospheric simulations using both real data (Observations or analyses) or idealized conditions. The WRF modeling system consists of four major programs: WRF pre-processing system (WPS), WRF data analysis (WRF-DA), Advanced Research WRF (ARW) solver, and Post-processing and visualization tools. For more detailed information than given in this section about WRF; see Appendix .1. The following information about WRF is from WRF modeling system User´s Guide [Wang et al., 2018]. For this thesis, WRF is used to simulate the two different extreme precipitation events using real data (analyses) and there sensitivity to changes in numerical schemes.

The model domain for the WRF was run for a area of 52°- 66°north and -6 °- 21 °east. The simulation starts at 00 UTC the day of the event and ends 00 UTC the day after. For all the simulations, the fifth generation of the European Center for Medium-Range Weather Forecast (ECMWF) reanalysis was used, *ERA5*. The runs where run with a time step of 8 second. Containing 51 vertical levels, with a grid size of 2km x 2km. The horizontal grid is defined by Lambert projection with center in 58.959°N and 7.599°E.

3.1.1 WRF physics options

There are different physics options available in the ARW, I will go into detail in two of them: The Planteray Boundary layer (PBL) and the Microphysics. Table 3.1 show an overview of the different WRF runs.

Planetary Boundary layer schemes

The part of the troposphere that is directly influenced by the Earth's surfaces is known as the planetary boundary layer (PBL). This layer responds to forcing from the surface with a timescale of an hour or less [Stensrud, 2007]. The purpose of the schemes is to distribute surface fluxes with boundary layer eddy fluxes and allow for PBL growth by entertainment. It is also referred to as the turbulence scheme. The scheme is responsible for the boundary layer and the vertical sub-grid-scale fluxes due to eddy transports in the whole atmospheric column [Skamarock et al., 2008]. For the thesis three different schemes are used. An overview and short description of the three different PBL schemes used in this study are given below.

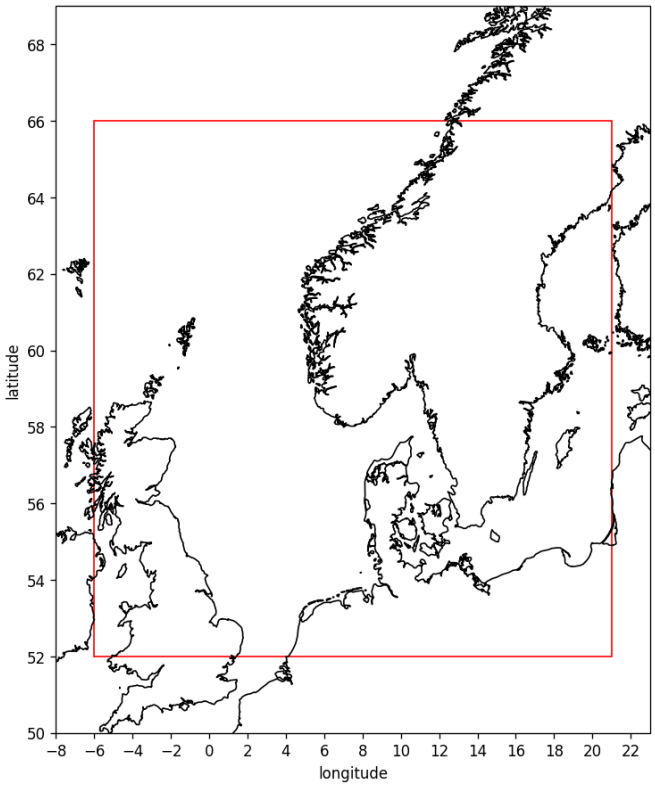

Figure 3.1: Model domain WRF show with red square (52°- 66°north and -6 °- 21 °east).

Yonsei Invierity (YSU) PBL (1) (Hong et al. [2006]) is the new generation of MRF PBL (Hong and Pan [1996]) which uses countergradient terms to represent fluxes due to non-local gradients. This new generation adds an explicit treatment of the entrainment layer at the PBL top. It is defined using a critical bulk Richardson number of zero (not 0.5 as before). Topographic drag effects and top-down mixing was added as an option or improved the last couple of years [Skamarock et al., 2008].

Mellor-Yamada-Janjic (MYJ) PBL (2) is a one-dimensional prognostic turbulent kinetic energy (TKE) scheme with local vertical mixing. The TKE production/ dissipation differential equation is solved iteratively [Skamarock et al., 2008].

Mellor-Yamada–Nakanishi–Niino (MYNN) (6) (Nakanishi and Niino [2009]) is a second-order turbulence closure scheme. It is based on the Mellor-Yamada (MY) PBL scheme but modified by revising key closure constants and introducing a new diagnostic equation for the turbulence master length scale [Huang and Peng, 2017].

These three schemes have some crucial differences and are therefore interesting to compare. The

Runs	mp_physics	bl_pbl_physics	sf_sfclay_physics
1	8	1	1
2	8	2	2
3	1	1	1
4	1	2	2
5	8	6	5
6	26	1	1

Table 3.1: An overview of the different WRF runs and which schemes correspond to which runs.

MYJ and the MYNN scheme have local mixing, while the YSU scheme has non-local mixing. MYJ and MYNN have a turbulent kinetic prediction, while the YSU has a more traditional K-closure.

Microphysics

The explicitly resolved cloud, water vapor, and precipitation processes are included in the microphysics option. The microphysics is treated as an adjustment process by directly updating the state variables. So it is carried out at the end of the time-step. The different schemes include different numbers of moisture variables. It is also referd to as the cloud physics. Some also include mixed phase and/or ice phase processes. An overview and short description of the three microphysics schemes used in this study are given below.

Kessler scheme (1) (Kessler [1969]) is the most simple microphysics scheme you get. This scheme includes no ice, which means no ice phase or mixed phase processes. It includes only warm clouds and takes water vapor, cloud water, and rain into account.

New Thompson et al. scheme (8) (Thompson et al. [2008]) is a much more advanced scheme then the Kessler. This scheme includes ice, snow, and graupel processes. This is very suitable for simulations with high resolution. It also consists of both ice phase and mixed phase processes. The scheme has been developed for especially mid-latitude convection, orographic, and snowfall [Skamarock et al., 2008].

WRF Singel-Moment 7-Class Microphysics Scheme (WSM7) (26) (Bae et al. [2019]) introduce a 7-class prognostic water substance. It includes, as for the 6-class, water vapor, clouds, rain, ice, snow, and graupel. Additionally, it introduces hail hydrometeor as the 7th prognostic water substance. Together with the 7-class microphysics, it predicts cloud condensation nuclei (CCN), number concentrations of clouds, and rain. (Hong et al. [2010]).

The main difference between the three different schemes is that the Kessler scheme only contains warm rain, while the New Thompson scheme and WSM7 include more advanced prognostic water substances. The Kessler scheme only includes idealized microphysics. WSM7 is different from the New Thompson with both the high number of water substances and the prediction of CCN, number concentration of clouds, and rain.

3.2 AROME

A Meteorological Cooperation on Operational Numerical Weather prediction (MetCoOp) version of the Météo-France Applications of Research to Operations at Mesoscale (AROME) model was, in 2014, put into operation by a cooperative effort of the Swedish and Norwegian meteorological services. This is the model used by the Meteorological Institute in Norway, and is in this thesis used to determine how well the events were forecasted. A short introduction of the model will be given in this chapter, but for the particular interested, a much more detailed description is given by Seity et al. [2011] and Müller et al. [2017].

The MetCoOp-AROME model covers Scandinavia and the Nordic sea with a 2.5 km horizontal resolution (Figure 3.2). The horizontal grid is defined by Lambert projection with center at 63.5°N and 15°E. The model has 65 vertical layers determined by a mass-based, terrain-following hybrid vertical discretization. Every 6 hour (00,06,12,18 UTC) a 66 hour forecast is produced. The model updates the atmospherical and land surface variables every 3 hours, used for data assimilation. Lateral and upper boundaries are from the ECMWF. Due to the delayed availability of ECMWF forecasts, the forecast used as boundaries is 3 to 6 h earlier than the actual forecast for the intermediate and main forecast cycles [Müller et al., 2017].

The cloud microphysics is base on the Kessler scheme, explained in section 3.1.1 and the three-class ice parameterization (ICE3) scheme. The Kessler scheme is used for the warm (liquid) processes, while the ICE3 scheme parameterizes the cold processes. ICE3 includes cloud ice, graupel, and snow, and more than 25 processes are parameterized by the scheme [Pinty and Jabouille, 1998]. In the MetCoOp-AROME, some modifications are made for the ICE3 scheme. This to correct for known errors, especially the T_{2m} winter bias, and to improve the clouds at low levels.

HARMONIE with RACMO Turbulence (HARATU) is the turbulence scheme used in AROME. HARATU is based on a scheme that was initially developed for a regional climate model called RACMO. The scheme uses a framework with prognostic equations for turbulent kinetic energy combined with a diagnostic length scale [Bengtsson et al., 2017].

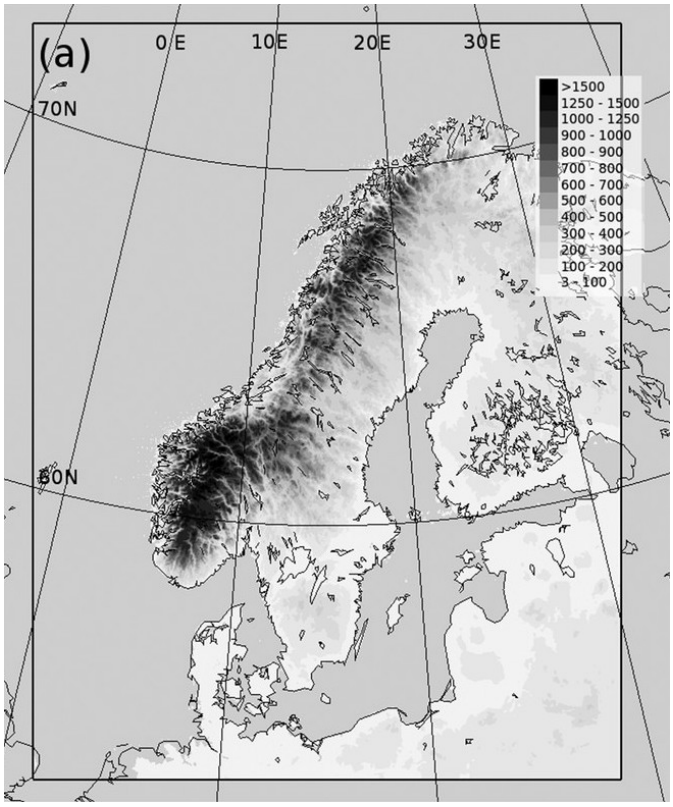

Figure 3.2: Figure showing entire model domain of AROME-MetCoOp and land topography (elevation [m]) [Müller et al., 2017].

4. Observational based data

4.1 Radar

To evaluate the events, radar data from MET was used. Radar stands for radio detection and ranging. For our use, we are trying to detect and ranging the precipitation. A radar sends microwaves with a given wavelength/frequency and measures the backscatter portion of this radiation. The signal returning includes many different types of information; the traveling time, the phase shift, the backscattered power, and the received radiation's polarization. This gives information about the type of hydrometeors, the amount of the hydrometeors, and wind speed through the Doppler effect [Reuder, 2019].

4.1.1 Scattering and radar theory

The Rayleigh scattering theory can be used for particles that are spherical and small compared to the wavelength (D<λ/16) given the following equation:

$$\sigma(D) = \frac{\pi^5}{\lambda^4} |\frac{\epsilon - 1}{\epsilon + 2}|^2 D^6 \tag{4.1}$$

σ is the backscatter cross-section, ϵ is the dielectricity constant, λ is the wavelength, and D is the diameter. Something to be aware of for this theory is that the diameter is raised to the power of 6. Meaning that the backscatter cross-section is very sensitive to, for example, error in the diameter.

The *radar reflectivity*, η, is the summation of all particles in a unit volume (1 m^3). It can be defines as radar reflectivity factor (Z):

$$Z = \frac{\sum D^6}{V_u} = \int_0^\infty N(D) D^6 dD \tag{4.2}$$

The equation is connected with the *radar equation* (Equation 4.3). The basic form of this equation is valid if we assume Rayleigh-theory and spherical water droplets. The practical use of the radar equation is typically written in logarithmic form, relating the reflectivity factor in decibels to the backscatter power in the decibel scale. r is the distance from the radar and P_r is the reflected power [Reuder, 2019].

$$z = 10logZ = 10logP_r + 10logr - 10logC' \tag{4.3}$$

Since we are interested in the *precipitation rate* (R), we need something related to the radar reflectivity. The semi-empirical or empirical Z-R ratio relates them together. There is no universal Z-R

ratio, and it can be hard to determine. The standard approach uses the Z-R relationship derived from Marshall and Palmer [1948] in the form of the power law, $Z = aR^b$. a and b depend on a number of different factors. This conversion from reflectivity to rain rate leads to both systematic and random significant error [Sivasubramaniam et al., 2018]. Discussion on this topic follows in the next subsection.

4.1.2 Errors and challenges in radar data

Different sources are causing an error in radar precipitation measurements. There are errors in both the reflectively measurements and the conversion from reflectively to precipitation rate. From the reflectively measurements, the error may originate from ground and sea cluster, beam blocking, anormal propagation, bright band, and non-meteorological echos (for example, a bird) [Elo, 2012].

Errors in the conversion from reflectivity to precipitation rate are linked up to the use of Z-R ratio. As mention earlier, there is no universal R-Z ratio. Precipitation occurs in the form of snow, rain, and a mixture of them in a cold climate. The different phases of the precipitation can cause problems. This can partly be solved using two sets of R-Z relations, one for rain and one for snow. However, both the European radar project OPERA and the Norwegian radars use only one R-Z ratio [Sivasubramaniam et al., 2018].

Another important point when looking at radar data is that we have to assume spherical droplets, which often is not the case. With a theory that is very sensitive to errors in diameter, this can have large consequences for the result.

4.2 Observation data

Observation data from https://seklima.met.no/observations/ is used, includeing observation and measurements that are available for the Meteorological Institute. The observations used have measurements for precipitation every hour. The different stations used are shown in Figure 4.1. The colored dots show stations that are used more frequently in this thesis than the rest. The red dot show Vassenden, which is around 20km from Jølster. The yellow dot show Kroken in Stryn. For Nøtterøy the stations at Vestskogen (blue) and Gjekstad (green) are shown.

4.3 Lightning data

The lightning data is from blitzortung.org. The data is used to detect convection and give an impression of how good the radar data is. Lightning detectors on the ground observe the lightning. There are 14 detectors on the mainland of Norway. These registers the very low frequency from the lightning. It will only be registered if more than one detector measures it. Using the information from the signal at different lightning detectors, the strength and location of the lightning can be calculated [Meteorologisk Institutt, 2017].

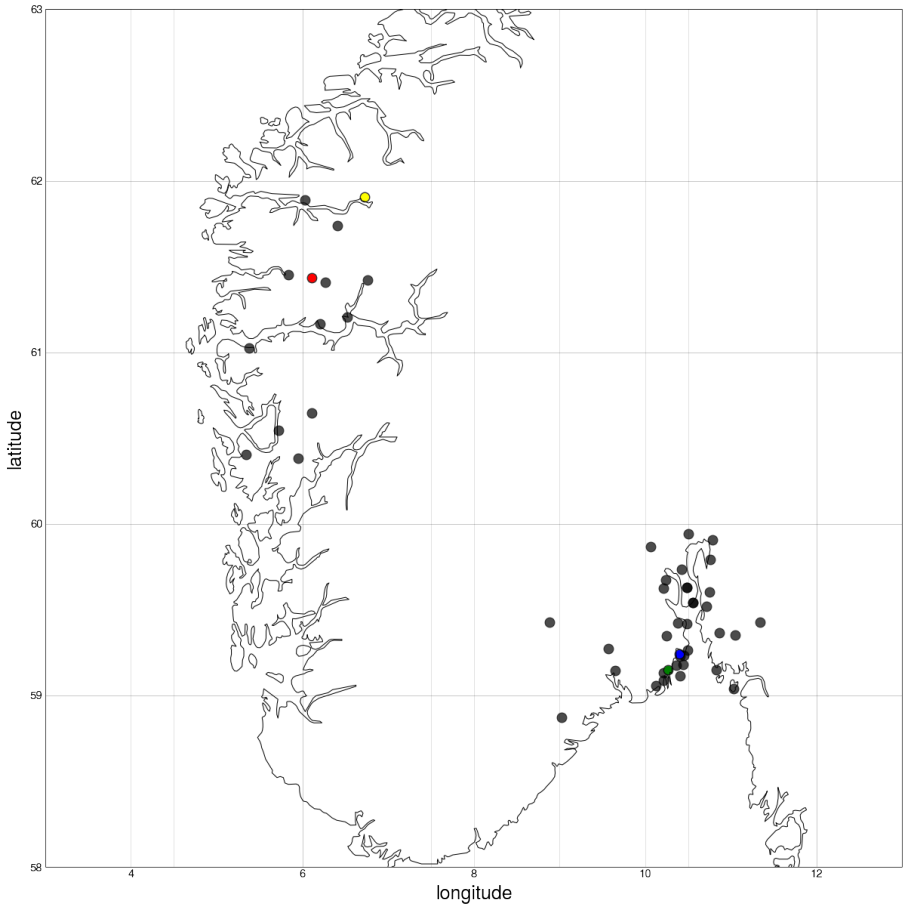

Figure 4.1: Map with an overview of the different ground measurements used. Some stations are shown with different colors: Vassenden (red), Kroken (yellow), Vestskogen (blue), and Gjekstad (green). These stations are used more frequently during the thesis.

5. Methods

5.1 Neighborhood verification

There are different ways of defining neighborhoods when using neighborhood verification. For example, you can have a neighborhood only in space, or in both time and space. It can be defined only for the observations or for both the observation and the forecast. In this thesis, I have used a definition called the "neighborhood maximum" (NM) approach (Figure 2.6 (e)). The neighborhood is defined for both the observation and the forecast [Schwartz, 2017]. This implies that it is a hit if both forecast and observed events occur anywhere in the neighborhood. This also implies that a correct negative is recorded when neither observed nor forecast events occur within the neighborhood (Table 5.1).

Neighborhood maximum definition				
	a (hit)	b (false alarm)	c (missed events)	d (correct negative)
Forecast condition	$F_k \geq$ q for some k$\in$ S_i	$F_k \geq$ q for some k$\in$ S_i	$F_k <$ q for some k$\in$ S_i	$F_k <$ q for some k$\in$ S_i
Observation condition	$O_k \geq$ q for some k$\in$ S_i	$O_k <$ q for some k$\in$ S_i	$O_k \geq$ q for some k$\in$ S_i	$O_k <$ q for some k$\in$ S_i

Table 5.1: Criteria for filling contingency table for the i'th grid point. S denotes the unique set of grid points within the neighborhood of i, q represents a precipitation accumulation event threshold, and F and O represent the forecasts and observations at i. q can either be an absolute threshold (e.g. 1.0 mm/h) or a percentile threshold (e.g. the 90th percentile) [Schwartz, 2017].

Since we define our neighborhood for both the forecast and the observation, we get more hits than the other neighborhood verification types. More hits can indicate a better forecast. But this also indicates that we will get more false alarms and missed events and fewer correct negative. It can therefore be hard to argue which definition one should use.

Table 5.1 does not include the definition of verification in time. When doing the different verification runs, I also included a neighborhood for time in some of them. This means that you allow your forecast and observation to be a bit inaccurate both in space and in time. If I define my neighborhood to be one hour back in time and one hour forward in time (giving a neighborhood of 3 hours), the forecast can show precipitation at t-1 while the observations show that the rain came

at t+1, but it would still be considered a hit.

Running the verification I was not only interested in verifying the forecast, but also to get an impression of the different result comparing point verification and neighborhood verification. Therefore, I ran a verification with two different thresholds, 2 year and 5 year return values for the location for every event. I did this for point verification, neighborhood verification with different sizes of the neighborhood and neighborhood verification including a time neighborhood. An overview of the different verification runs is shown in Table 5.2.

Overview of verification			
Verification type	**Threshold**	**Neighborhood space**	**Neighborhood time**
Point verification	2/5 year return value		
Neighborhood verification (space)	2/5 year return value	7 km	
Neighborhood verification (space)	2/5 year return value	12 km	
Neighborhood verification (space and time)	2/5 year return value	7 km	3h
Neighborhood verification (space and time)	2/5 year return value	12 km	5h
Neighborhood verification (space and time)	2/5 year return value	12 km	7h

Table 5.2: An overview of the different verification method used in the thesis. All runs are done for both 2- and 5-year return values as the threshold.

AROME produces a forecast 00,06,12 and 18 UTC during a day and creates a forecast 66 hours forward in time. To evaluate how the goodness of the forecast changes in terms of when it was produced, I ran verification for every prediction made from 12 UTC two days before the event to 00UTC the day of the event. For Jølster, the 18UTC forecast the day before the event was not usable, so verification for that time is missing.

The verification run is produced using AROME as the forecast and radar images were used as the observation. The radar data has a resolution of 1 km when reading in the data, while the AROME data has a resolution of 2.5 km. Since they have different resolutions, an interpolation had to be done to use the different verification methods. The interpolation method "nearest-neighbor interpolation" was used to interpolate the AROME forecast to 1 km resolution. This interpolation method makes the new grid and gives every point in the new grid the value to the nearest old grid point. The verification is done from and area from 3°E to 15°E, and from 55°N to 63 °N.

This verification method allows us to calculate categorical statistics. The result of the categorical statistics calculated for every event is from an area of 100km x 100km. The square is centered at the location of the event.

5.2 Point verification - closest point

Another verification method that can be useful when working with forecast verification of extremes is to look at the closest point forecasted above a given threshold. Suppose you, for example, have a ground observation with high values. In that case, you can validate the forecast by finding how far it is to the closest point with forecasted values similar to the observation. Point verification is introduced in earlier chapters, but then in terms of comparing every grid point of forecast to every grid point of an observation (in my case radar data). This closest point method is also a type of point verification, but this method is not dependent on the radar data in contrast to the method introduced earlier. This method only depends on the forecast validation, its resolution, and the threshold you set.

This method is used for both the AROME forecast and for the WRF runs. Every grid point within a circle of 1° radius is found. It then follows that all forecasted precipitation further than 1° away from the center is not considered a "hit". All the points within the circle above the given threshold are counted, and the distance to the different points is collected. This gives us information on the distance to the closest point, and the area covered where the forecast predicted above the threshold. Figure 5.1 is an illustration of how it is done. I used the IVF return values for 2 and 5 years as thresholds when verifying both AROME and WRF.

When calculating the closest point, the average distance of the three closest points is used. This is to make sure that it was not only one grid cell forecasting above the threshold. The resolution of the model also needs to be considered. If the closest distance is less than the model's resolution, it can be considered spot on. The better the resolution of the model is, the more accurate distance you will get.

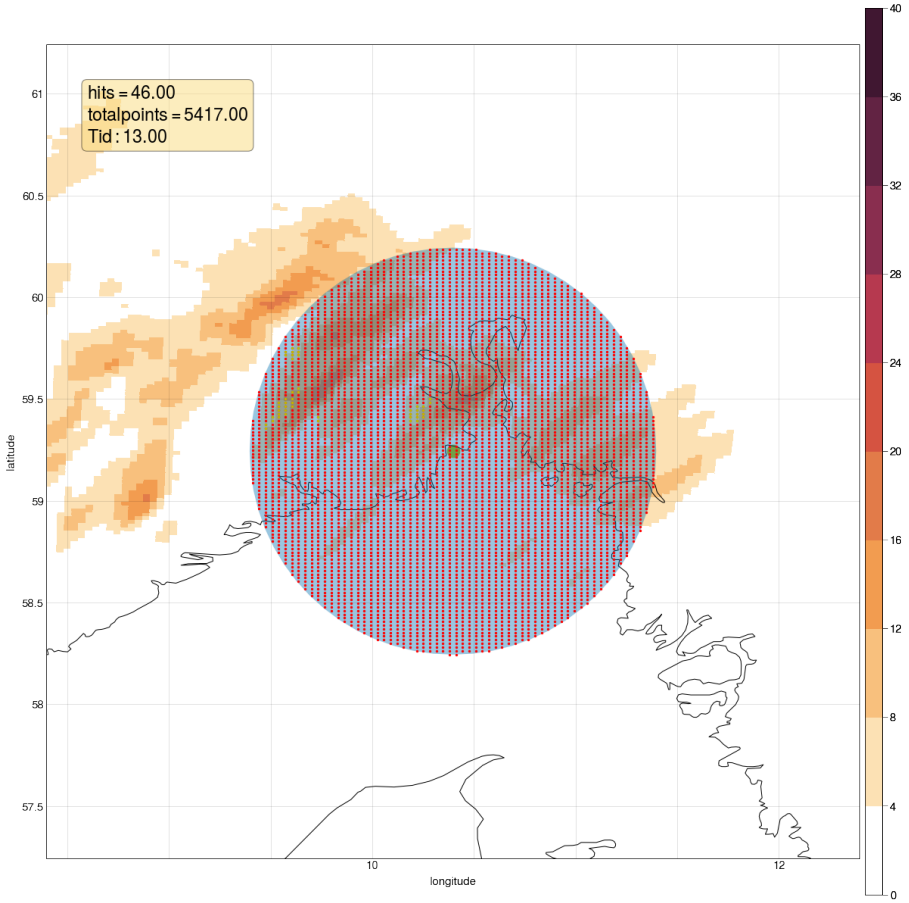

Figure 5.1: Illustration picture of how the point verification - "closest point" is done. Figure showing AROME forecast for Nøtterøy. The center of the circle is placed in Nøtterøy. The yellow dots show grid points with forecast above the threshold, while red dots show the grid points below the threshold.

6. Results

For this thesis, two different events of extreme precipitation in Norway were investigated. The first event happened in Jølster, located on the west coast of Norway during summer 2019. The second event happened in 2020 at Nøtterøy in southern Norway. The results from Jølster is presented first (Chapter 6.1), followed by the result from Nøtterøy (Chapter 6.2).

The two chapters are build up similarly, where the first section presents a description of the event based on radar data, observation, ERA5 reanalysis data, and AROME analysis data. The second section displays verification of the AROME forecast, studying how well the event were forecasted. This leads us to the topic of ensemble members and spatial correlation. The last part looks at the sensitivity of the events to turbulence (pbl) and cloud physics (microphysical) numerical scheme by presenting the result from the WRF runs.

6.1 Jølster 30.07.19

30 July 2019 started as a perfect summer day in Jølster, with sun, and temperatures almost reaching 27°C. But later that day an extreme event of short duration precipitation took place. At Vassenden, around 20km from Jølster, 33.1 mm/h was measured at 17UTC. This is a extremely high value, considering that the estimated 30 years IVF-value for Jølster is 21.6 mm/h. The event triggered several landslides, one which caused a casualty. From The Meteorological Institute only a yellow warning message was sent out, which is the first of three levels of danger. There were large amount of precipitation over extensive parts of the southwestern part of Norway. Due to the tragic consequences of the event located in Jølster, I found it natural to focus on this area. In this chapter, we will take a closer look at what happened that day.

6.1.1 Event description - synoptic and mesoscale analysis

On the 28th of July, we observe a low pressure of 1002 hPa coming in from the south, moving towards Great Britain (Figure 6.1). The low pressure moves further toward Great Britain and southern Norway and reaches 992 hPa at the 29th of July. A high pressure is located west of Finnmark in the Barents sea, with a central pressure of 1032 hPa. The situation of a low pressure over Great Britain and high pressure south of the northern parts of Norway gives easterlies winds in southern Norway. With a cyclonic rotation over Great Britain and a anti cyclone rotation at higher latitudes, the winds generated by the different pressure system at latitude of southern Norway are in the same direction. This strong pressure gradient generates strong winds. The temperature decreases over the west and southern part of Norway from 29th of July to the 30th of July. The

most intense precipitation came in the period 14-16 UTC the 30th of July. When referring to the timestep the event took place, I will refer to 15 UTC.

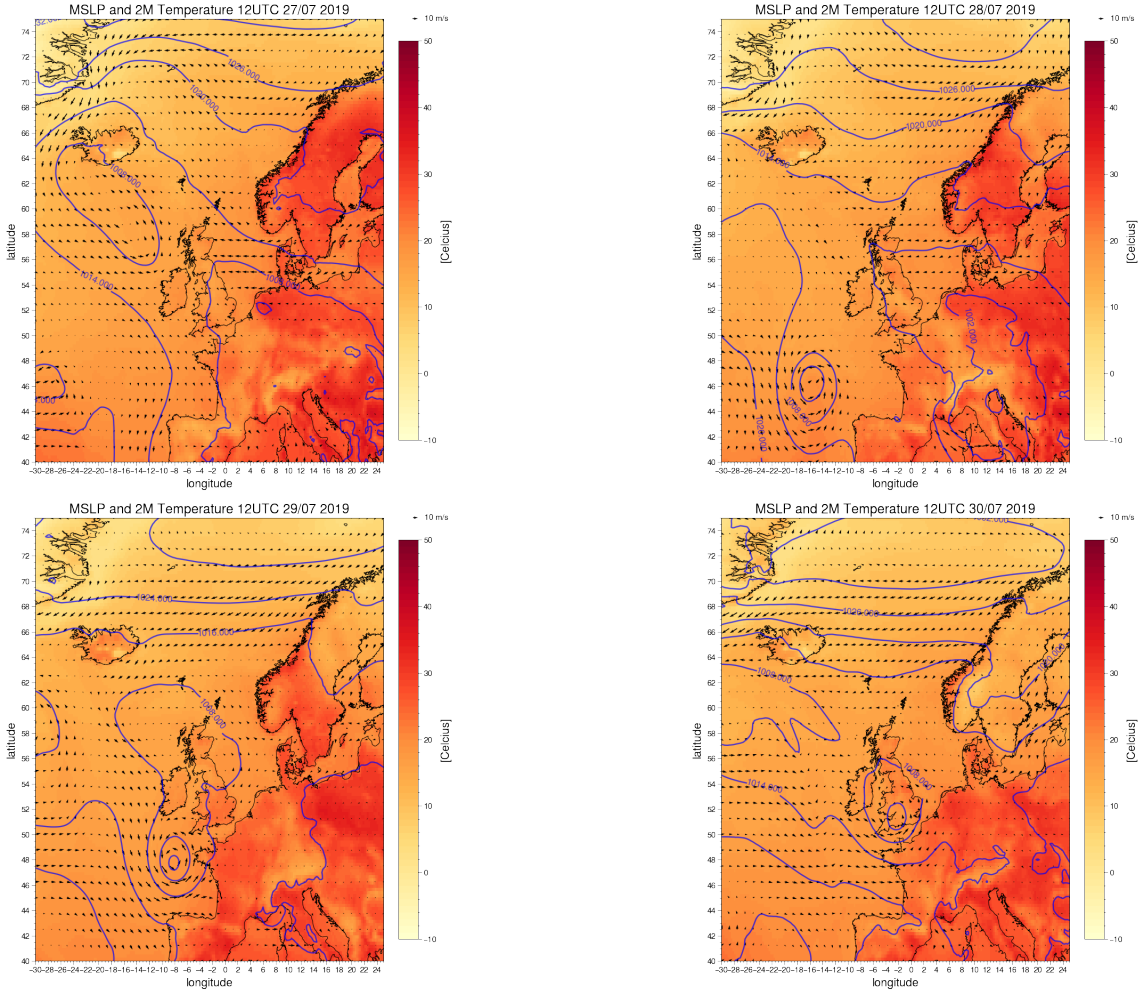

Figure 6.1: ERA5 reanalysis data showing 2 meter temperature (shaded) in Celcius, wind at 10m height in m/s, and the mean sea level pressure (contours) in hPa for the period before the event. The figure show 12UTC three days before the event (*upper left*), two days before the event (*upper right*), one day before the event (*lower left*) and the day of the event (*lower right*).

The local analyses show a decrease at around 0.013 g/kg in specific humidity and an increase of 37% in relative humidity from midnight on the 29th to midnight on the 31st (Figure 6.2). Relative humidity almost reach 100% 31th of July 00UTC. The increase in relative humidity indicates moist and/or cold air coming in over Jølster. The decrease in specific humidity illustrates that the air is not mainly moist, but cold. The cold air is brought by the winds generated from the high pressure (Figrue 6.1), bringing cold and dry air from the north. Comparing the specific humidity with the temperature, we see that they evolve identical, which substantiate that the air is mainly cold, and relatively dry. When cold air comes in over a relatively hot surface, the atmosphere becomes unstable. The temperature decreases by almost 8 degrees from the 29th to the 31st. All theses changes in the different parameters can be considered as distinct changes considering the duration of the changes.

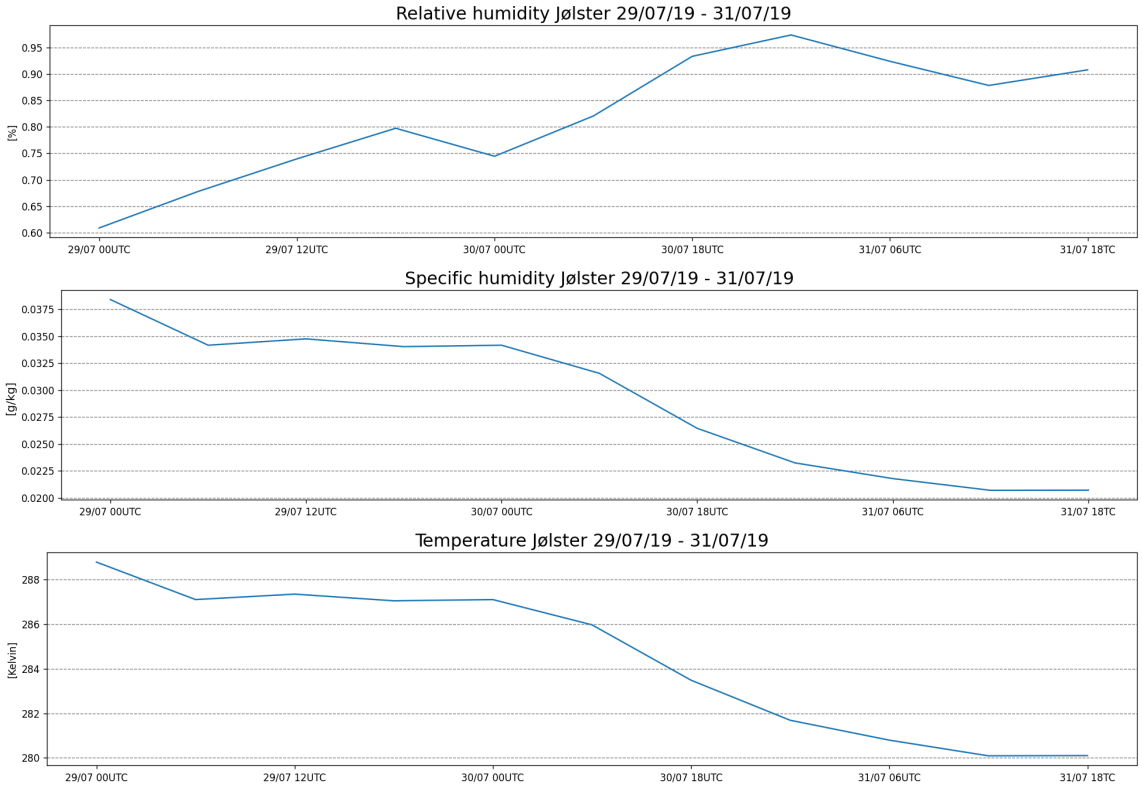

Figure 6.2: Evolution in time of relative humidity [%], specific humidity [g/kg] and temperature [K] for Jølster. Data presented is AROME analysis data at 850 hPa. The values shown are a mean of 50 km x 50 km area centered at Jølster. The event in Jølster is defined to be most intense in the period 14-16 UTC 30/07.

Stability indices

Convective available potential energy (CAPE) is a very commonly used stability indices for determining the intensity of deep convection, described in chapter 2.2. For the event at Jølster, there were generally high values of CAPE over large parts of western Norway (Figure 6.3). At 12UTC the 30th of July we observe values between 1000 - 2500 J/kg over a large area, indicating moderately unstable air, in additions to a smaller area with values >2500 J/kg which indicates very unstable air. The high values of CAPE decreases with time, presumably by the convective available energy being used to form precipitation. It is interesting to note that the CAPE values do not reveal anything in the large-scale analysis which indicates a higher likelihood for an extreme event in Jølster is greater than anywhere else on the west coast of Norway. At 12UTC we also observe low values of CIN at Jølster, indicating high chances of thunderstorm development.

Another index that is useful to look at when detecting area with risk of heavy precipitation is the K index (Chapter 2.2). Values over 30 indicates a high likelihood of thunderstorm activity [Galvin, 2016]. In the area of Jølster the values varies between 31 and 33. Values between 30 and 40 correspond to potential for thunderstorm with heavy rain (Table 2.4). Other areas on the west coast of Norway have significant higher values than the area of Jølster. The overall observation of the K index shows much of the same as the CAPE, with high values over a large part of western Norway (Figure 6.4). There is nothing indicating that Jølster is in larger risk of convective storms than other places on the west coast of Norway for this stability indices as well.

After evaluating the synoptic situation and the stability indices, the low- and high pressure positions which sets up easterlies over western Norway which brings cold air over western parts of Norway, creating an unstable situation. When cold air comes in over a warmer surface, it can create a conditionally unstable situation, and we then need a trigger mechanism to initiate the convection. These events are tough to forecast, which we will take a closer look at in chapter 6.1.2. In the next section, we will look at observations and radar data available for the time of the event.

Observations

For further investigation of the event stations in the area around Jølster, together with the lightning the day of the event and the data from the radar (Figure 6.2). Both maps show max hourly precipitation observed by the radar at every grid point on the 30th of July. Figure 6.2 (left) shows available ground measurements, as well as the hourly radar maximum in every grid point. The radar observes no high values at or close to Jølster. The highest value observed by the radar that day is 47.6 mm/h, and is located in the blob a bit south-west of Jølster. There is observed much precipitation over a relatively large area. In three hours, the station at Stryn measured 63.21 mm. At Vassenden, around 20km from Jølster, the measurement shows 33.1 mm in one hour at 14 UTC. After this, the station breaks down, maybe due to the heavy amount of precipitation. High values were observed over most parts of the west coast of Norway. At Fossmark in Hordaland, 33 mm in one hour was observed.

The map to the right shows both hourly radar maximum in every grid point and the lightning detected the day of the event. The lightning is plotted with a shading color to show at which time of the day the event happened. Lightning is a very good indication of where convection happened. In the period between 15 UTC and 19 UTC, over 6000 lightning strikes were observed at Vestlandet.

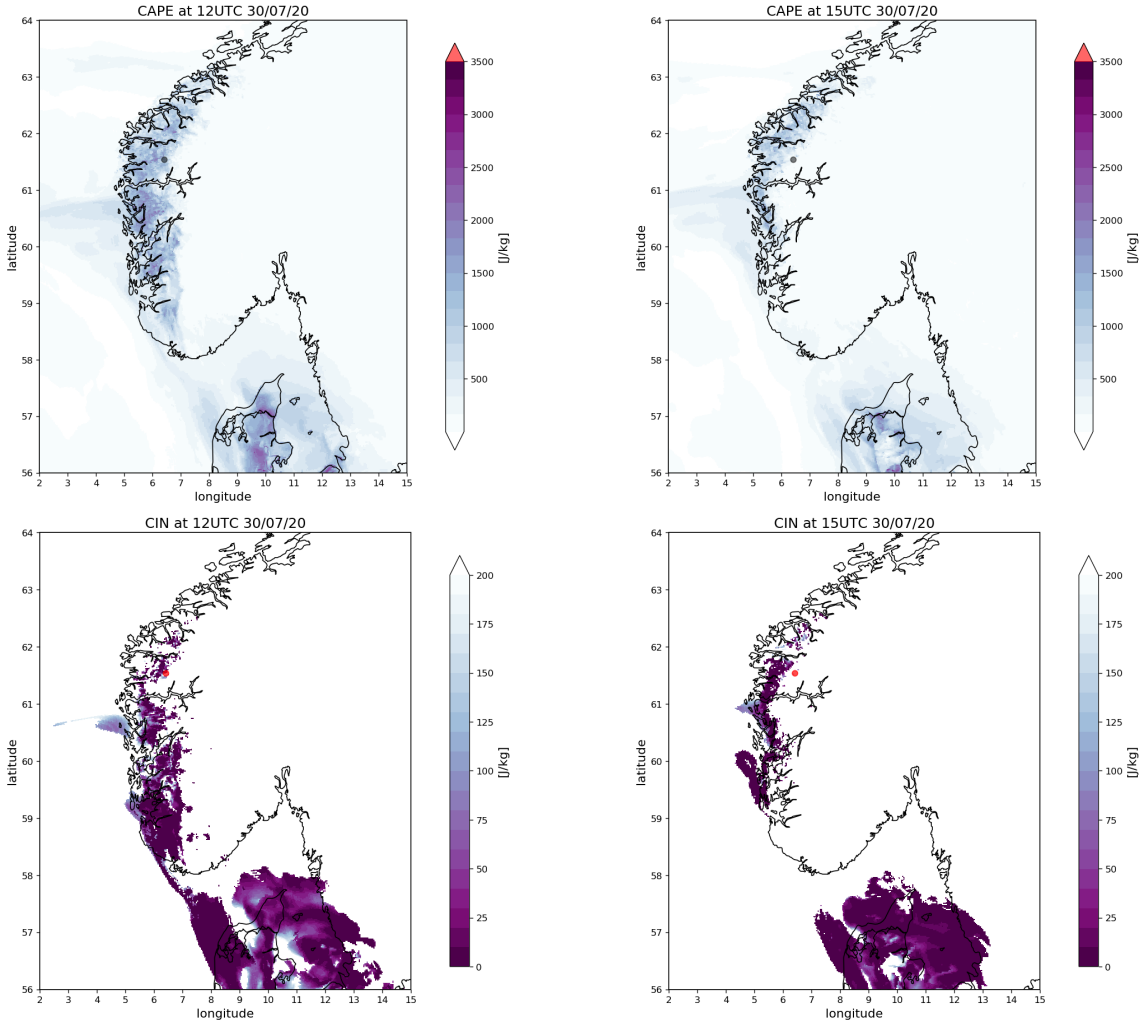

Figure 6.3: Stability indices CAPE [J/kg] from AROME forecast made 30/07 00UTC for different time steps at 30/07/19 *upper*) and for CIN [J/kg] (*lower*). The data is from 12UTC and 15UTC. The black (CAPE) and red (CIN) dot show the location of Jølster.

Lightning data together with radar can be used as a verification of the radar data. As discussed in the theory (Chapter 4.1.2) there are different challenges when doing radar observations, so the observations should not be trusted blindly. The lightning data indicates that there was convection in Jølster. In the area at the exact longitude as Jølster, and further east, between Sognefjorden and Innvikfjorden, the radar observes no precipitation. But the lightning shows that there where convection in the area. This indicates that maybe the topography disturbs the radar signal so that the radar does not observe precipitation in the given area.

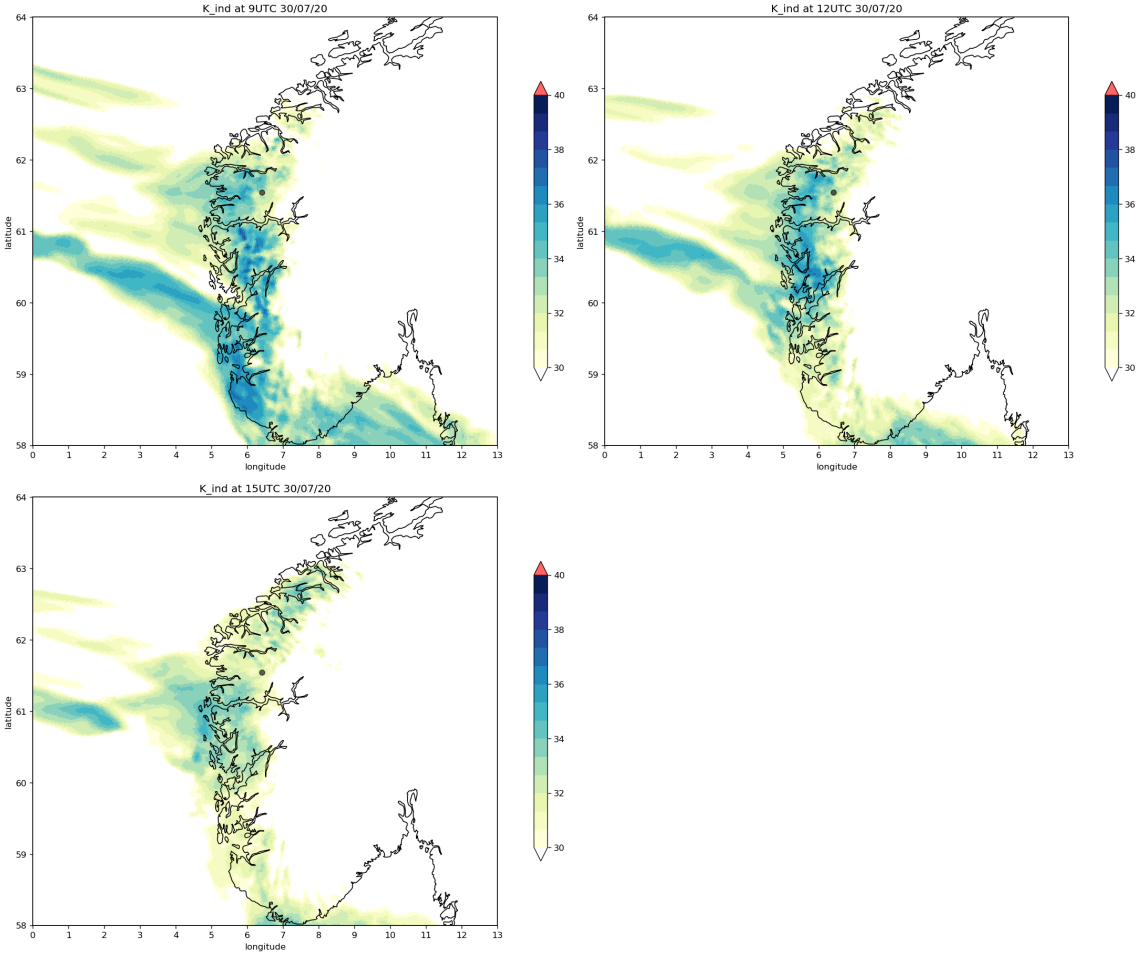

Figure 6.4: Stability indices K index from AROME analysis data for different time steps at 30/07/19. The black dot show location of Jølster.

6.1.2 Validation of the forecasts

For the event in Jølster only a yellow warning was sent from The Meteorological Institute for heavy rain showers, floods and landslides. The Meteorological Institute has, in a report published after this event, stated that they should have upgraded the warning to orange. In this section we will have a look at the forecasts made for the event, using both point and neighborhood verification to verify the forecast and look at the spread in ensemble members.

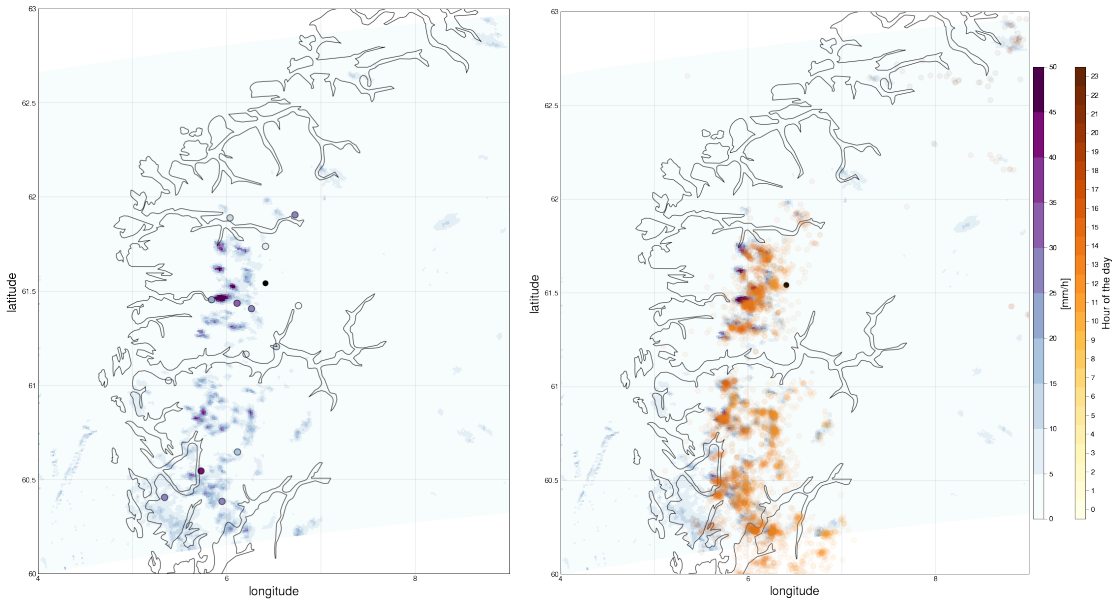

Figure 6.5: Map of Vestlandet showing hourly maximum precipitation [mm/h] observed by the radar in every grid point the 30th of July (shaded purple) together with max hourly ground observations (dot purple) in the map to the *left*. Both plotted in the same color shading. The plot to the *right* shows the same maximum precipitation observed by the radar in every grid point on the 30th of July (shaded purple) together with lightning (shaded orange) data from the same day. The lightning data has color shading, after which hour the lightning stroked. The black dot shows the location of Jøster.

Neighborhood verification

For this section point and neighborhood verification is presented, trying to get an impression of how well the event was forecasted. The section will also include both neighborhood in space and neighborhood in time and space. This is to get an impression of how the result changes when using the different methods. The different verification methods are run for all the available forecasts from AROME, from the one issued at 12 UTC the 28th (51 hours before the event) to the one issued at 00 UTC the day of the event (15 hours before the event). We will then also evaluate how the verification changes when we are getting closer to the event. The result is presented with different categorical statistics. For this the point and neighborhood verification, radar data is used as the observation. As threshold, the 2 and 5 year IVF return value for the location of the event is used.

The first categorical statistic we will be looking at for the event, is the *bias score/ frequency bias*. The frequency bias tells us whether the forecast tends to overcast (bias frequency > 1) or undercast (bias frequency < 1) the event. Overcasting means a high occurrence of false alarms, meaning that the forecast shows precipitation over the threshold, while the observation (in this case radar) does not show precipitation over the threshold. Undercasting is the opposite. The frequency bias show good results for all the different neighborhoods, the different dates and both thresholds (Figure 6.6). Most values are close to 1 when using neighborhood verification. The area of the neighborhood has a negligible impact on the result. But comparing the values for point verification

versus neighborhood verification we clearly see a positive effect of neighborhood verification. Point verification gives high values, indication overcasting. When using neighborhood verification the bias frequency is much closer to 1. As we observed when comparing the radar with observation and lightning data, the radar did not capture the event in Jølster, only the amount of precipitation further west. By expanding the neighborhood we include an area where the radar has captured precipitation, giving hit where the point verification would give a false alarm.

The next categorical statistics we will look at are hit rate (HIT) and false alarm rate (FAR). The hit rate is sensitive to hits but will ignore the false alarms. Hits are when both the forecast and the observation show values above threshold within the neighborhood, or in the point. False alarm is when the forecast has values above threshold within the neighborhood/in the point, but no values above threshold was observed. Many hits do not necessarily correspond to a good forecast. If a forecast also includes many false alarms, and especially when working with weather extremes, it might not be a good forecast. Therefore, the hit rate will be evaluated together with the false alarm (Figure 6.7). These two categorical statistics demonstrates well how a wider neighborhood provides better values. We can not expect the model to predict extremes at an exact time and place. However, there is a fine line between being very forgiving by having a wide neighborhood and losing too much temporal and spatial information.

With the largest neighborhood in time and space, the verification on the forecast made the day of the event (+15h) have a hit rate of almost 0.28. Meaning that almost 30% of the forecasted precipitation was a hit. But if we only consider neighborhood in space or point verification we have almost no hits. For the largest neighborhood 72% of the gridpoints where the forecast said "yes it will rain above the threshold", rain above the threshold was not observed by the radar. But keep in mind that the radar did not capture the event exactly at Jølster. For the false alarm ratio we observe very high numbers for the largest neighborhood. For these two categorical statistics, in contrast to bias frequency, the results vary dependent on the definition of the neighborhoods.

To sum up all of this statistics; the neighborhood verification show that the model had high precipitation in the area around Jølster. But the exact time and place were not necessarily correct. Comparing the forecast made at different times, there is not a clear pattern of the forecast being better closer to the event. We will do further investigation on the forecast by looking at ensemble members and using closest point verification.

Ensemble members

Spread in the ensemble members is expected, but how large the spread is can give an indication of how confident the weather forecast is. Figure 6.8 show time series of gridpoint accumulated precipitation for the different ensemble members for forecast made at 12UTC 29th of July (+27h) (*upper*) and 00UTC 30/07 (+15h) (*lower*) and the observations from Vassenden and Kroken. The observations are included mainly to compare and evaluate the ensemble members. Vassenden has only available data until 18UTC the 30th of July, the following 48h no data is available. Therefore Kroken is included, even though it is a bit far away from Jølster. Most of the ensemble members forecast precipitation at the time around 14 UTC the 30th of July, while two of the ensemble members forecast an earlier onset of precipitation. The range in intensity and amount is large. Four of the ensemble members forecast below 9 mm precipitation in total. For the forecast made

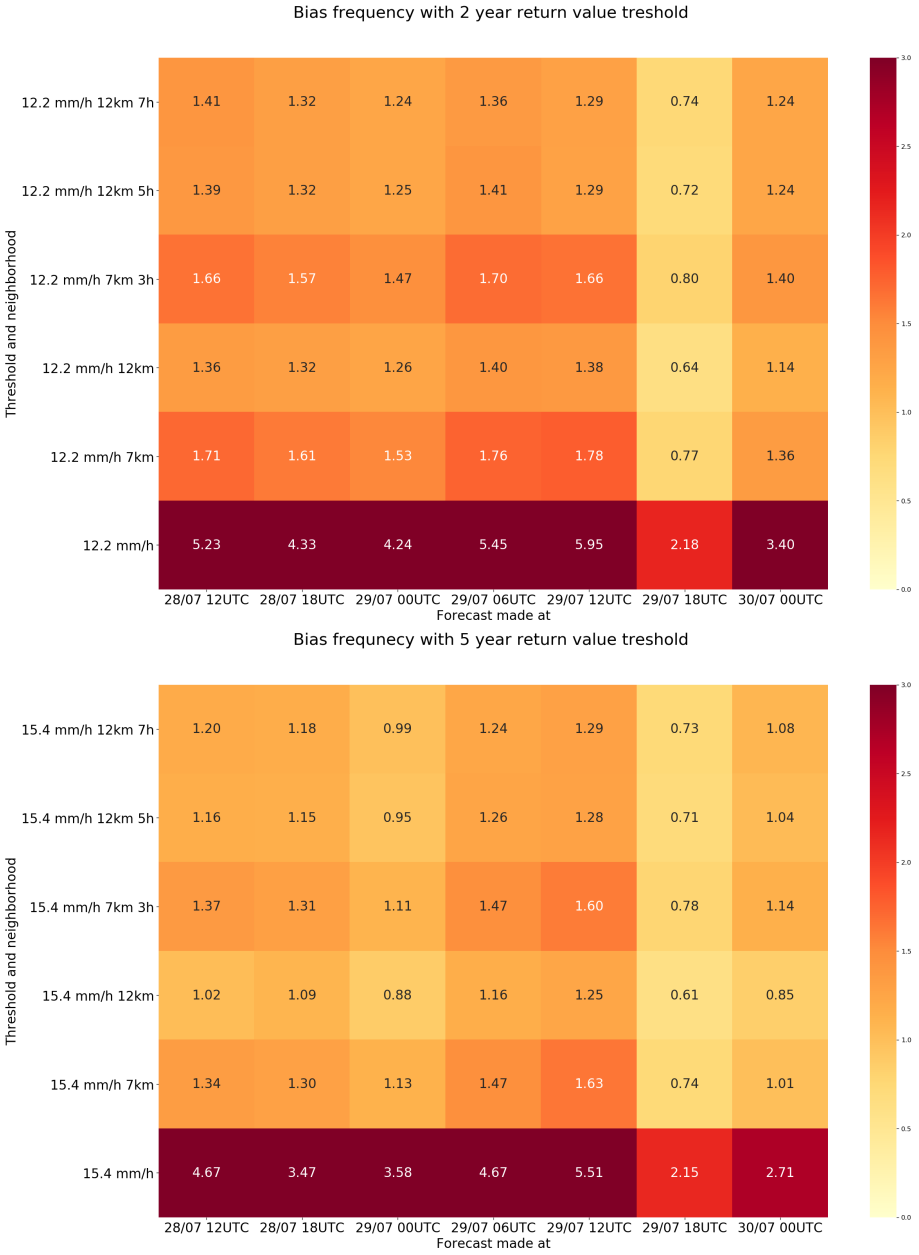

Figure 6.6: Frequency bias for Jølster with 12.2 mm/h (upper) and 15.4 mm/h (lower) threshold. The frequency bias for different verification methods and neighborhoods (y-axis) is shown. For example, '15.4 mm/h' means point verification with a threshold of 15.4 mm/h. '12.2 mm/h 7 km' mean neighborhood verification in space with a threshold of 12.2 mm/h and a spatial neighborhood of 7 km. The last type is a neighborhood in time and space: '15.4 mm/h 12km 5h' means a threshold of 15 mm/h with a neighborhood of 12 km range and 5 hours time range. The x-axis shows when the forecast is made.

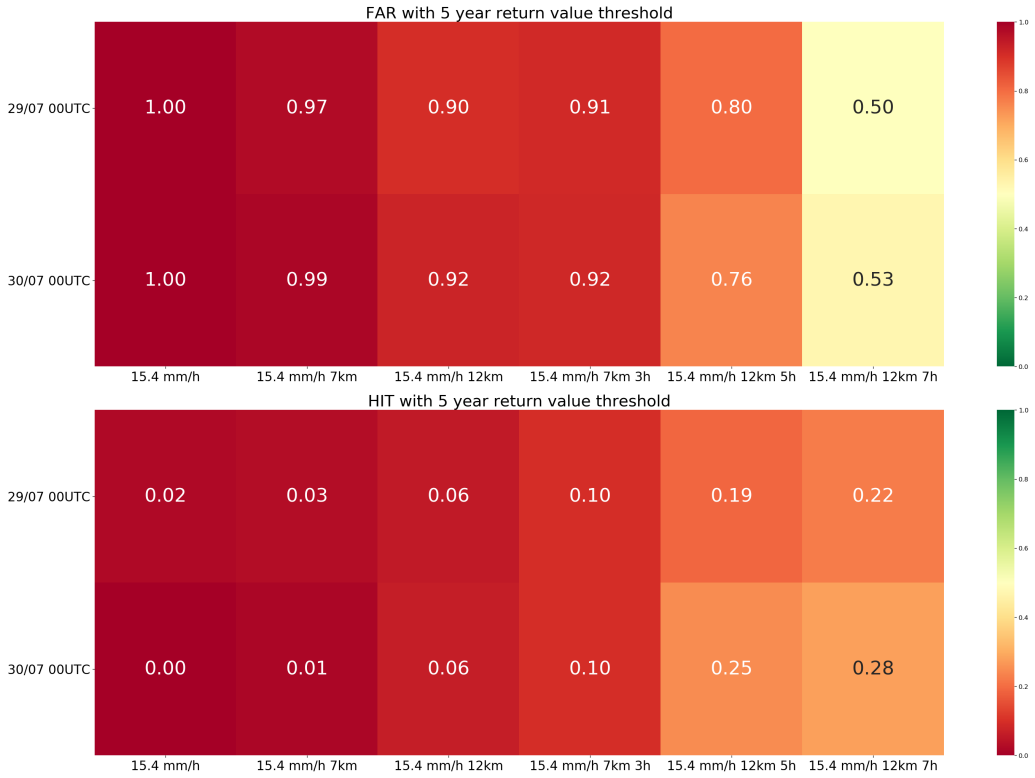

Figure 6.7: False alarm ratio (upper) and hit rate (lower) for Jølster with 15.4 mm/h threshold. Forecast made at 00 UTC the day before the event and the day of the event (y-axis). Including results from different verification methods (x-axis). The color scaling goes from 0 to 1. 'XX mm/h' indicating point verification with the given threshold. The others are neighborhood verification, with neighborhoods either in space ('XX mm/h Y km') or in both space and time ('XX mm/h Y km Z h').

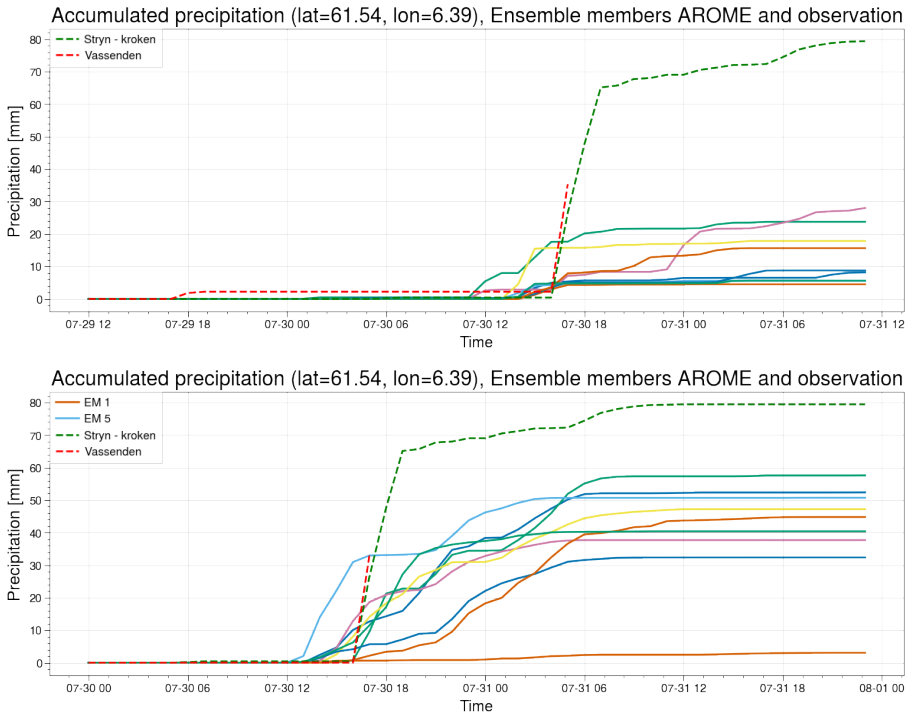

Figure 6.8: Time series of accumulated precipitation for all different ensemble members (*solid*) from the AROME forecast made at 12UTC 29/07 (*upper*) and 00UTC 30/07 (*lower*). The AROME accumulated precipitation is from Jølster (lat=61.54, lon=6.39), and the observation show data from Vassenden and Kroken (*dashed*). Kroken has no data for 48h after 18UTC 30/07.

midnight the day of the event, there is relatively good agreement between eight of nine ensemble members on the amount of accumulated precipitation on the 30th of July. But how intense the precipitation is (shown by the growth rate of the accumulated precipitation) varies greatly within the eight ensemble members. Some forecast around 30 mm in two hours, while another reaches 30 mm at midnight (12 hours). We observe that non of the ensemble members predict the amount and intensity that were observed at Kroken. But the ensemble members for the forecast made 00UTC the 30th of July (+15h) is much closer than 12UTC the 29th of July (+27h).

Not capturing the intensity of the event is a problem because the intensity is the damaging part for short duration events. Many of the ensemble members have the same total precipitation after 7 hours, but the intensity varies. It seems like the model has a problem predicting the extreme intensity. We therefor compare two different ensemble members, one with good forecasted intensity and one with bad forecasted intensity (Figure 6.9) from the forecast made midnight the day of the event. This is to investigate why strong intensity is triggered for some ensemble members while not for others.

Stability is the change of potential temperature with height ($d\theta/dz$). The stability refers to the

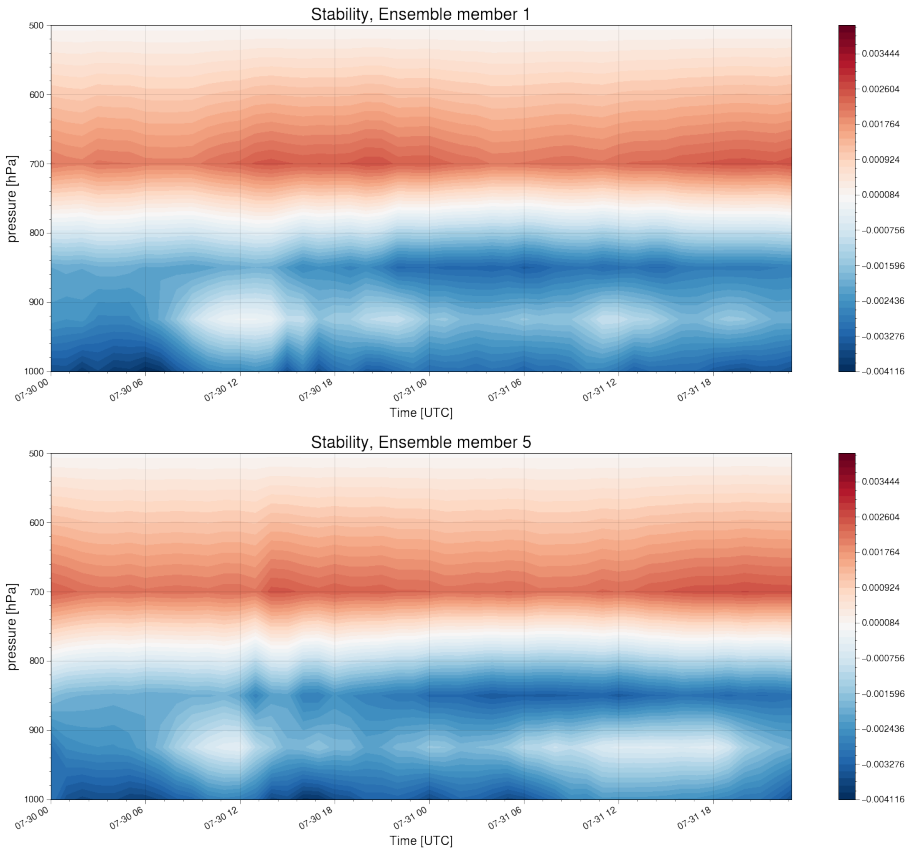

Figure 6.9: Evolution of vertical stability ($d\theta/dz$) for two different ensemble members. Time along the x-axis and pressure along the y-axis. Ensemble member 1 forecast little precipitation and has low intensity. Ensemble member 5 has the highest intensity of the ensemble members.

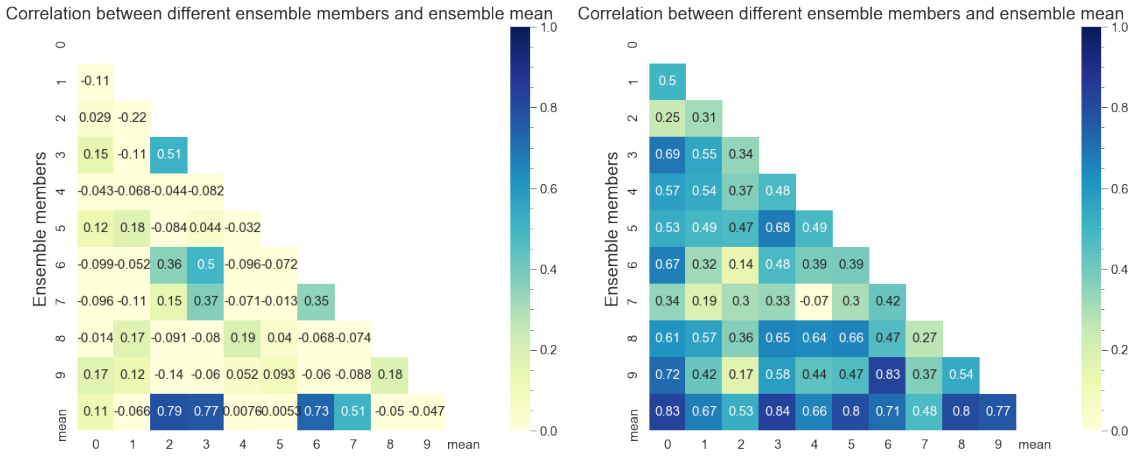

Figure 6.10: Spatial correlation between the different ensemble mean and the ensemble mean. The left figure shows the spatial correlation for the forecast made at 12UTC 29/07 (+27h) and the figure to the right for the forecast made at 00UTC 30/07 (+15h).

ability of the air to prevent vertical movement. For both the ensemble members an unstable layer is located at the ground, with a stable layer on top (Figure 6.9). Ensemble member 5 forecasts 30 mm in two hours. Ensemble member 1 forecasts 4 mm for the whole day. It is interesting how the neutral air (stability = 0) at 950 hPa builds up from 7UTC and toward the event for both the members. When the event happen, the air becomes unstable leading to a higher possibility of heavy convection. Comparing the two members, member 5 have more unstable air around 850 hPa at the time of the event and the event seem to have impact at higher altitude. For ensemble member 1 it seem like what happen at lower altitude has minor impact on the higher altitude.

Spacial correlation - Ensemble members

Another way of evaluating the spread in ensemble members is to look at the pattern correlation between the different members, together with the ensemble mean. Pattern correlation show the linear correlation between two variables at corresponding location on two different maps. Pattern correlation give useful information on how the precipitation pattern from the different ensemble members agree with each other. We detect a large difference of agreements between the ensemble members for the two forecasts made at 00UTC the 30th of July (+15h) and at 12UTC the 29th of July (+27h) (Figure 6.10). The forecast made midnight the day of the event has much higher correlation than the one made +27h before. The closer the coefficient is to 1, the better the pattern of the two maps agrees. The forecast gets more confident moving towards the event, and be considered quite confident at midnight the day before the event.

Closest point verification

To further evaluate the goodness of the forecasts the ”closest point” method is used. This method works well for rare event and looks at the closest grid square forecasted above a given threshold in the area of the event. There are different ways of doing this. I have chosen to look at every grid

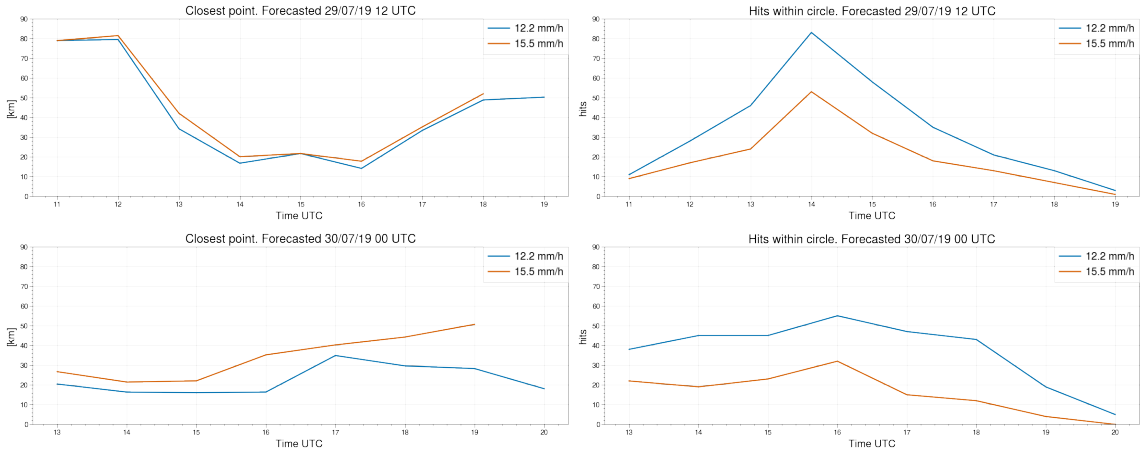

Figure 6.11: Average distance for the three closest point (*left*) and hits withing a circle of 1° radius (*right*). Time along the x-axis and distance (*left*) / numbers of hits (*right*) along the y-axis. The results are from the "closest point" verification method.

square in an area of a circle with radius of 1° (See section 5.2). The average distance of the three nearest gridsquares relative to Jølster and the amount of hits withing the circle is calculated. The closest point is only counted if it is withing the circle of 111 km (corresponding to 1° latitude) radius from Jølster. One of the advantages of this method is that it is not dependent on the radar data.

In general we observe that the distances gets smaller and the hits gets higher when we move towards the time of the events (14-16 UTC), and opposite when passing the time of the event and moving forward in time (Figure 6.11). For the lowest threshold, the 2 years return value, we observe that in the time period 14-16UTC the closest point are forecasted around 16km away from Jølster. The closest point we observe is at 16UTC for the forecast made 12UTC the day before the event and is 14.15 km away. The period of time with observed hits above threshold within the circle is more narrow for the forecast closer to the event. The forecast made 12UTC the 29th of July (+27h) is relatively good, considering the peak of hits, and the evolution of the closest distance. But when the event happen the distance is still a bit to far away, to say that it was a good forecast. This is interesting since the neighborhood verification showed that the overall forecast was relatively good, especially considering the bias frequency. In contrast, this method shows that the forecast never really forecast high precipitation amount at Jølster.

6.1.3 WRF simulation

Furthermore we evaluate the sensitivity of the event in Jølster to changes in the turbulence (pbl) and cloud physics (mp) schemes. If the events are not sensitive to changes in the given schemes, it can be a strong indication that the events are mainly driven by the large scale synoptic situation, and not by the turbulence or microphysics.

To compare the different runs the max hourly precipitation forecasted for every grid cell is shown (Figure 6.12 and 6.13). This detect if high values of precipitation were forecasted at, or close to,

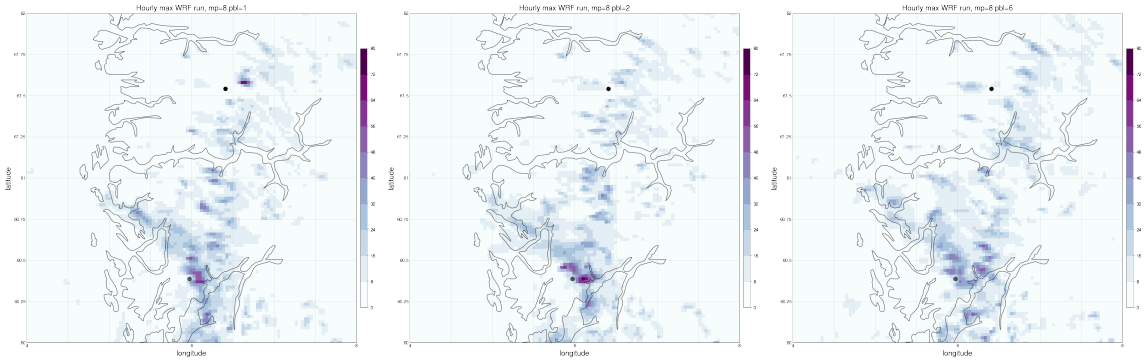

Figure 6.12: Maximum hourly precipitation for every grid cell for WRF runs with different planetary boundary layer schemes. The colorbar goes from 0 to 80 mm/h. Two black dots, one showing the location of Jølster (lat = 61.54) and one the location of Kvamskogen (lat = 60.38). *mp* meaning microphysical scheme and *pbl* meaning planetary boundary layer scheme.

Jølster the 30th of June. WRF was ran with three different turbulence (pbl) scheme (Figure 6.12) and three different microphysical schemes (Figure 6.13). This type of figures is an efficient and easy way to compare the different runs with focus on hourly extremes. At a first glance the different schemes do not seem to have a large impact on max hourly precipitation. No extreme values are predicted at Jølster. In the figure a black dot at Kvamskogen is also included. Kvamskogen also observed very high values, and for the WRF runs this event actually seem to be good forecasted. One run that deviate a bit from the other runs is the run with michropysical scheme 26 (Figure 6.13 *middle*). It has less precipitation around Kvamskogen and has a blob of heavy precipitation at same longitude as Jølster but 0.5° further south. No of the other schemes have this blob.

Observing the ensemble members from AROME, we found that the model seem to struggle capturing the intensity of the events. Some of the ensemble had total precipitation that were relatively high, but the intensity is too low. The changes in schemes seem to have almost no impact on the timing of precipitation (Figure 6.14). For both locations they all agree at what time the precipitation will start. For Jølster the amount of precipitation is small and the timing is far from the observed event. For a point a bit east for Kvamskogen, where high amount of precipitation was modeled, we observe that the model agrees with the observations. The amount of forecasted precipitation is much larger, and the timing is better. Some of the simulations have an intensity that is very close to the observed. Four out of five runs predict higher amount of accumulated precipitation for the time period then the observation in Kvamskogen.

Going into details of the different WRF runs, we see that run number 2 (mp=8 pbl=2) has a very similar intensity as the observation at 13UTC. While run number 5 (mp=8 pbl=6) is the one that differs the most from the observation regarding intensity and timing. Figure 6.15 show the stability and the vertical wind in the point west of Kvamskogen for the two different runs. We have ascent for both runs at the time of the event. For run number 5, the vertical velocity is stronger compared to run number 2, but the air moves upward over a more extended time period. There is a weak ascent in the three hours before the maximum vertical velocity appears for run number 5. These

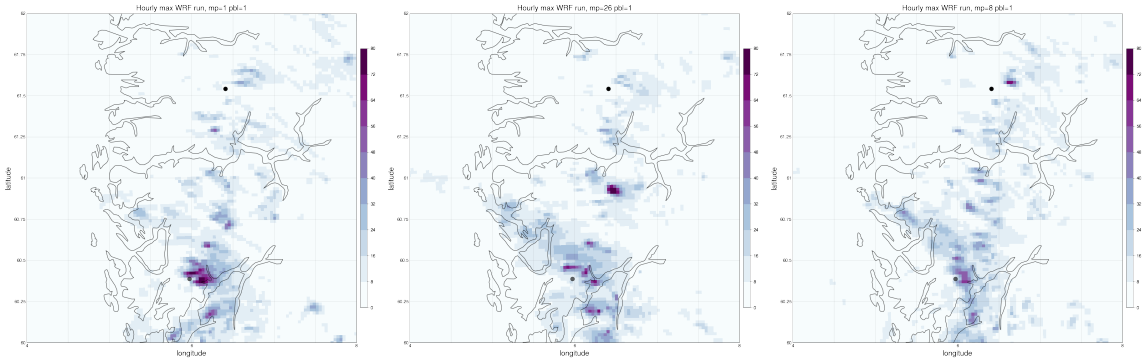

Figure 6.13: Maximum hourly precipitation for every grid cell for WRF runs with different microphysical schemes. The colorbar goes from 0 to 80 mm/h. Two black dots, one showing the location of (lat = 61.54) and one the location of Kvamskogen (lat = 60.38). *mp* meaning microphysical scheme and *pbl* meaning planetary boundary layer scheme .

three hours of weak ascent bring rain, and when the ascent reaches a maximum, at the time of the event, there is not as much moisture available as for run number 2. Therefore the intensity is stronger for run number 2. Both the runs have descent right after the ascent. This can reminiscent of a waveform passing Kvamskogen and can, for example, be caused by topography.

Verification - Closest point above threshold

One of the same verification methods as introduced when evaluating AROME is used to evaluate the different runs. The hits within a circle of 1°radius and average distance to the three closest points are calculated. We observe that the shape of the different runs is relatively similar for the two different thresholds (Figure 6.16). Hits within the circle increase as we move toward the time of the event and decrease afterward.

The closest point does not show very good results. No distance is below 10km for the 5 year return value, meaning nothing is forecasted above this threshold within 10km distance from Jølster. For the 2 year return value threshold, the closest distance is only below 10km two times for all the different runs and different time steps. These timesteps are at 21UTC and 22UTC, which is much later than the time of the event in Jølster.

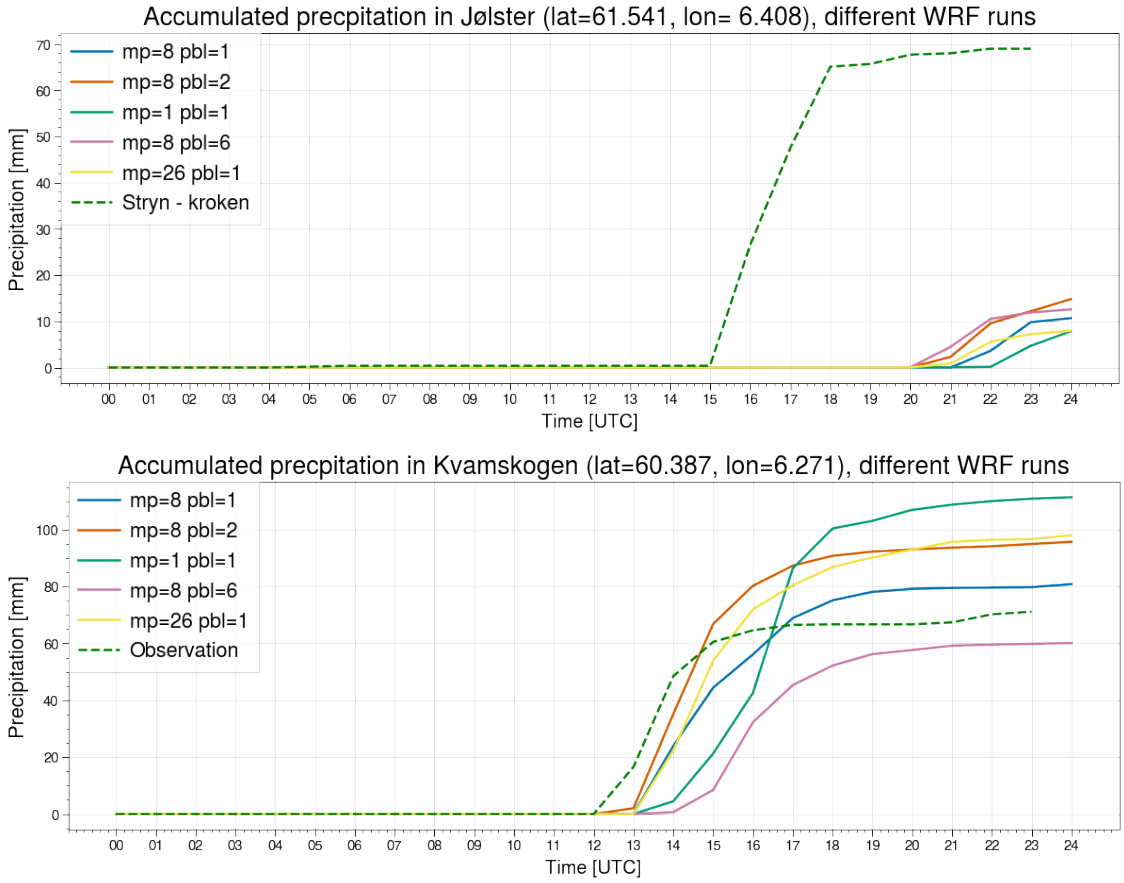

Figure 6.14: Accumulated precipitation, for Jølster and east of Kvamskogen, for different WRF runs and observation (dashed line). Since Vassenden has no available data after 18UTC, Stryk - Kroken is the observation for Jølster.

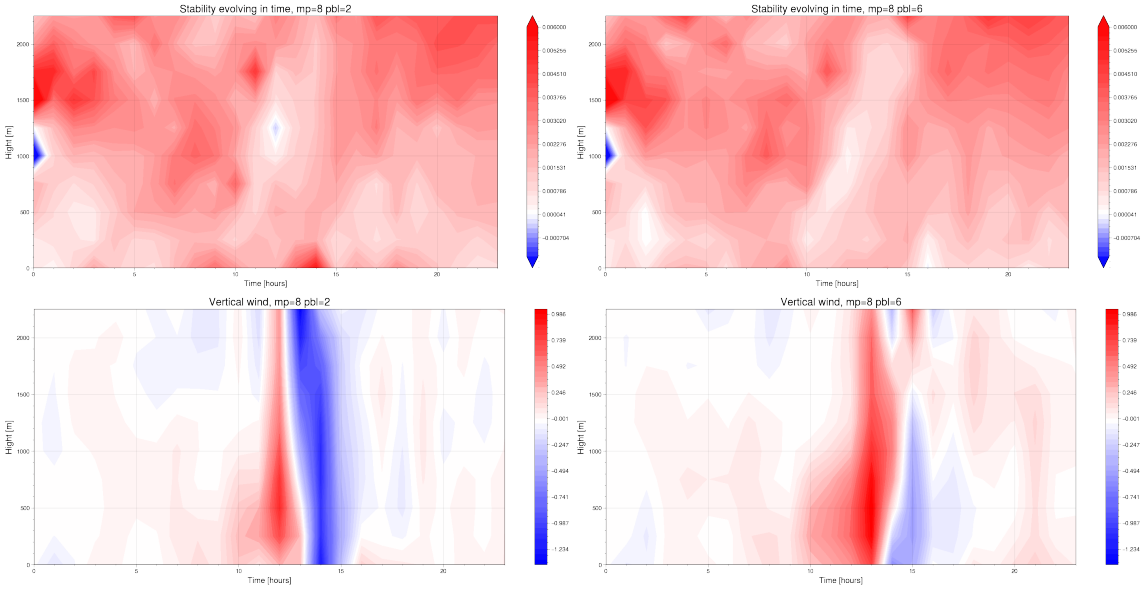

Figure 6.15: Vertical stability (*upper*) and vertical wind (*lower*) for two different WRF runs. Time along the x-axis and height [m] along the y-axis. WRF run number 2 (*left*) is with mp=8 and pbl=2. WRF run number 5 (*right*) is with mp=8 and pbl=6. *mp* meaning microphysical scheme and *pbl* meaning planetary boundary layer scheme.

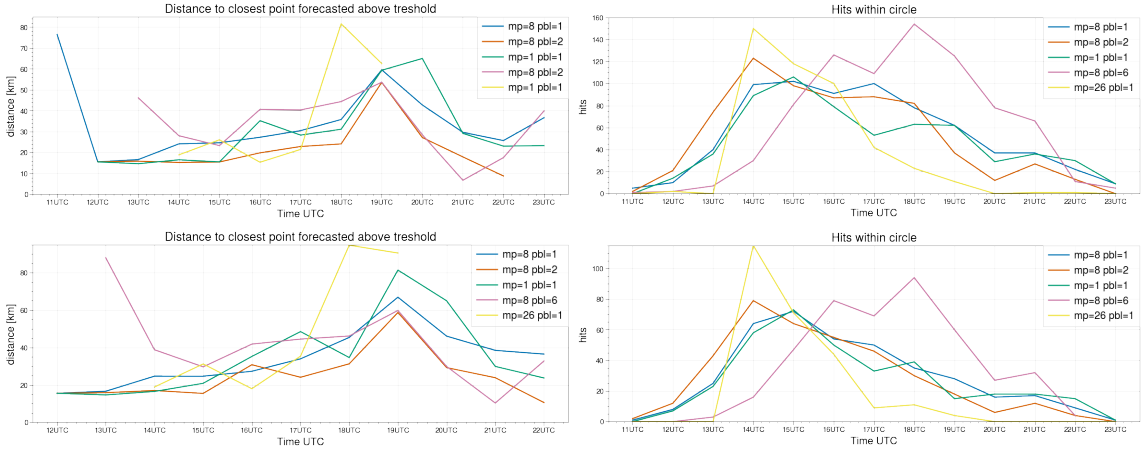

Figure 6.16: Average distance for the three closest point (*left*) and number of hits withing a circle of 1° radius (*right*) above a given threshold. The treshold is the 2 year return value (*upper*) and the 5 year return value (*lower*). Time along the x-axis and distance (*left*) / numbers of hits (*right*) along the y-axis. The results are from the "closest point" verification method.

6.2 Nøtterøy 21.08.20

The second event happened in the area of Nøtterøy on the 21st of August 2020. This was a short duration extreme precipitation event. A new official Norwegian record for the highest amount of accumulated precipitation measured over one hour was set during this event. This record was measured at METs station in Sandefjord - Gjekstad, and set to 63.6 mm/h. The hour before and after the event at 13UTC, this station measured 2.1 mm/h and 1.7 mm/h, implying that this was a very short duration event.

6.2.1 Event description - synoptic and mesoscale analysis

At the time around the event at Nøtterøy a low pressure comes from the Atlantic ocean towards southern Norway (Figure 6.17). The low pressure is 968 hPa at 12UTC the day before the event, while at the day of the event the center show 984 hPa. On the day of the event, the low pressure center is located south of Great Britain. The winds generated from the low pressure give Nøtterøy onshore winds at the date of the event.

For a more local investigation the mean sea level pressure (MSLP), the specific humidity at 850 hPa, and the temperature at 850 hPa at Nøtterøy from the AROME analysis is evaluated (Figure 6.18). The different parameters all show changes at the time of the event, the 21st of August at 13UTC. We note a drop of 6 hPa in MSLP from 6UTC - 16 UTC. This is not a remarkably large drop, but happens over a relatively short time. The specific humidity shows a peak at the time of the event, and increase by 0.0025 g/kg. The temperature increases by 2K just before the event, followed by a drop of 3 K at 15 UTC.

Stability indices

Evaluating the stability of the air masses for Nøtterøy we look at CAPE, CIN and the TT index. We observe high values of CAPE up to 3000 J/kg coming from the southwest, moving in toward Nøtterøy (Figure 6.19). This is considered as very unstable air. As the air moves towards land, and Nøtterøy, the stability increases (CAPE decreases). At 15UTC a blob of CAPE above 1000 J/kg is located west of Nøtterøy, and at the next time step, it seems to have passed Nøtterøy. Values over 1000 J/kg corresponds to moderately unstable air masses. Compared to the event at Jølster, CAPE around the area of the event stands out. The highest values of CAPE are located over Nøtterøy at 15UTC, while the event happened at 13UTC. Evaluating CIN (Figure 6.20), we observe very low values at Nøtterøy two hours before the event. At 13UTC the low values of CIN at Nøtterøy is gone. Low values of CIN indicate that little force is needed to lift the parcel to LFC.

For the TT index (Figure 6.21) we observe the same very unstable air as observed with CAPE at 6UTC right outside Mandal. Large areas show values above 55, indicating that scattered severe thunderstorms are likely. As for CAPE, the values decrease as the air moves towards land, and comes in as a band hitting Nøtterøy between 15UTC and 16UTC. The TT index shows relatively high values over a large part of southern Norway, even though the band hitting Nøtterøy stands out. With high values over a large area, it can be hard to define the area of risk.

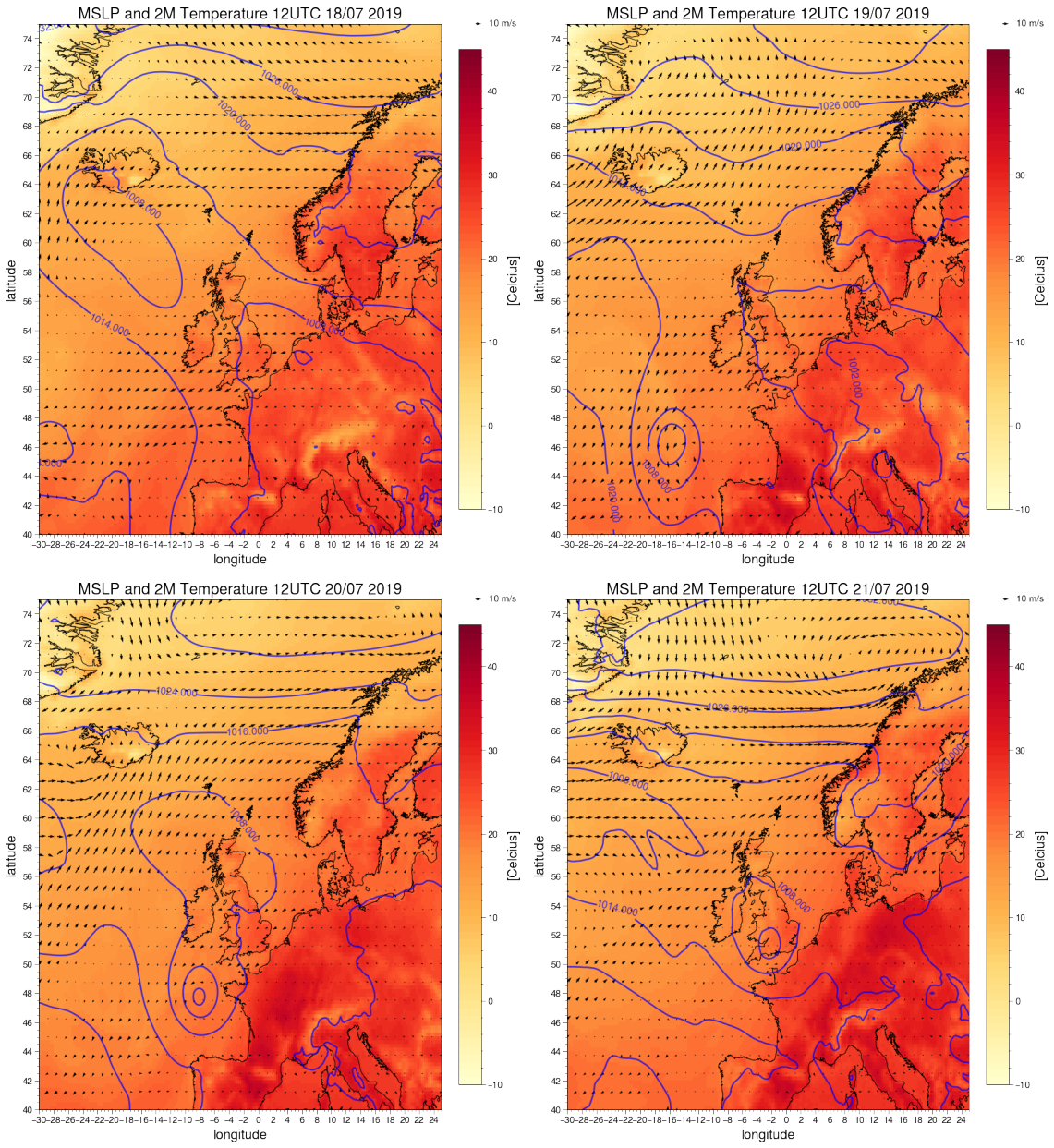

Figure 6.17: ERA5 reanalysis data showing 2 meter temperature (shaded) in Celcius, wind at 10m height in m/s, and the mean sea level pressure (contours) in hPa for the period before the event. The figure show 12UTC three days before the event (*upper left*), two days before the event (*upper right*), one day before the event (*lower left*) and the day of the event (*lower right*).

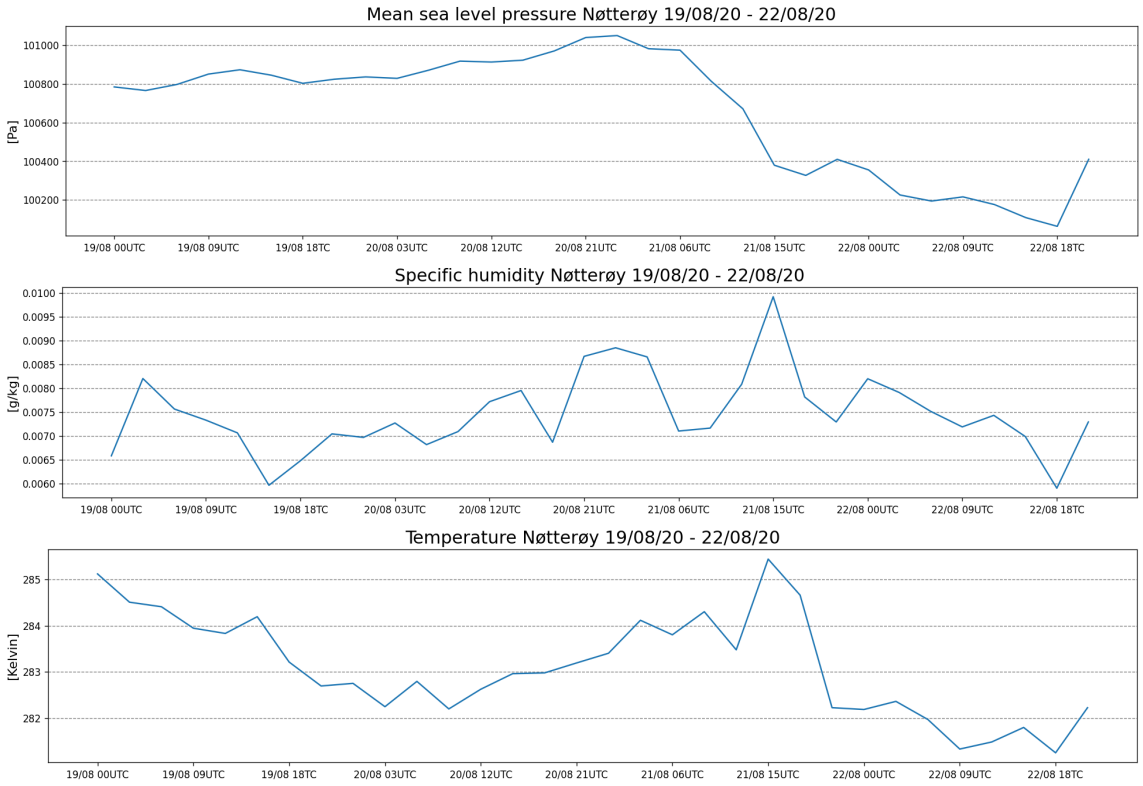

Figure 6.18: Evolution in time of relative humidity [%], specific humidity [g/kg] and temperature [K] for Nøtterøy. Data presented is AROME analysis data at 850 hPa. The values shown are a mean of 50 km x 50 km area centered at Nøtterøy. The event in Nøtterøy happened at 13 UTC 21/08.

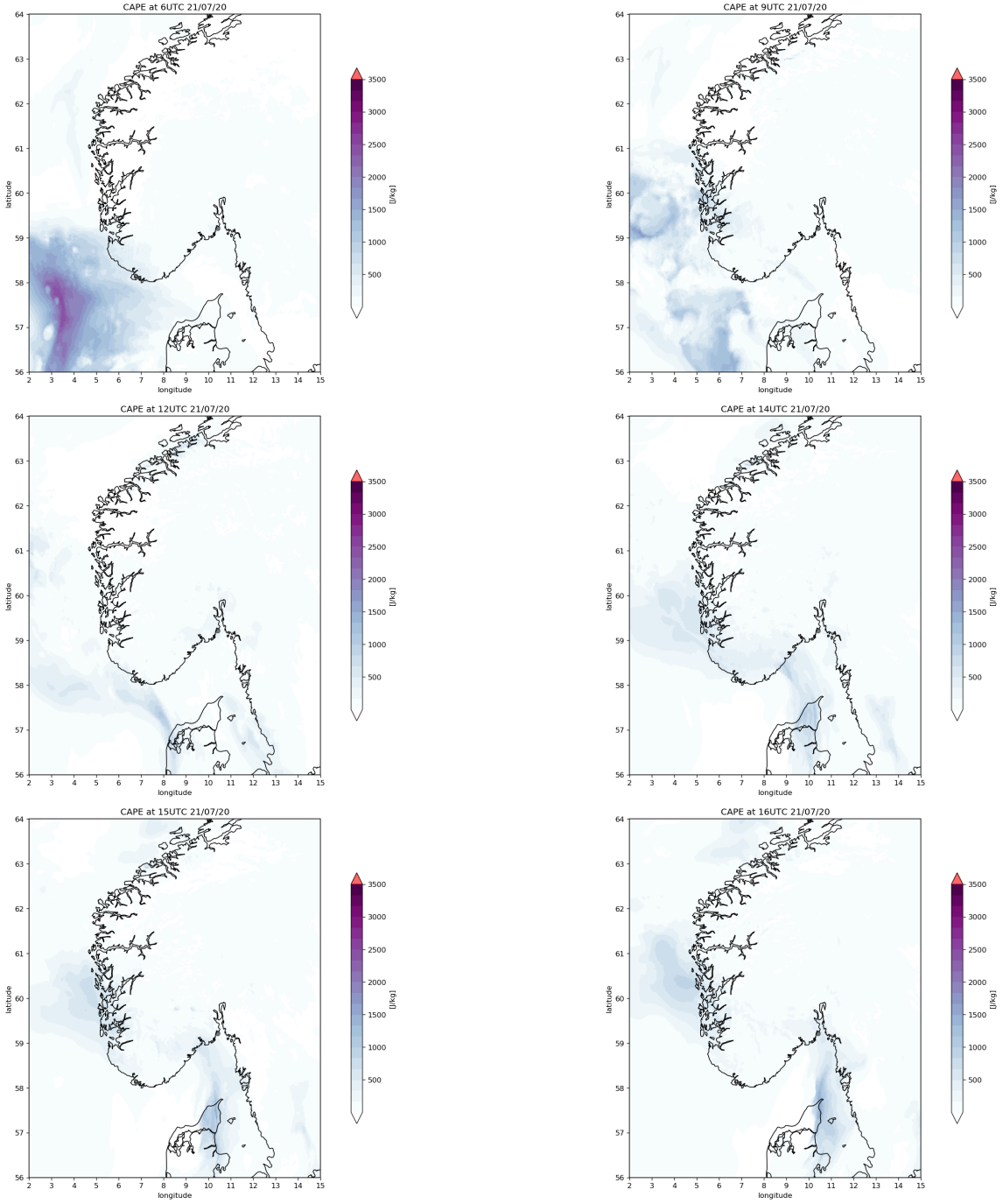

Figure 6.19: Stability indices CAPE [J/kg] from AROME analysis data for different time steps at 21/08/20. The black dot show location of Nøtterøy.

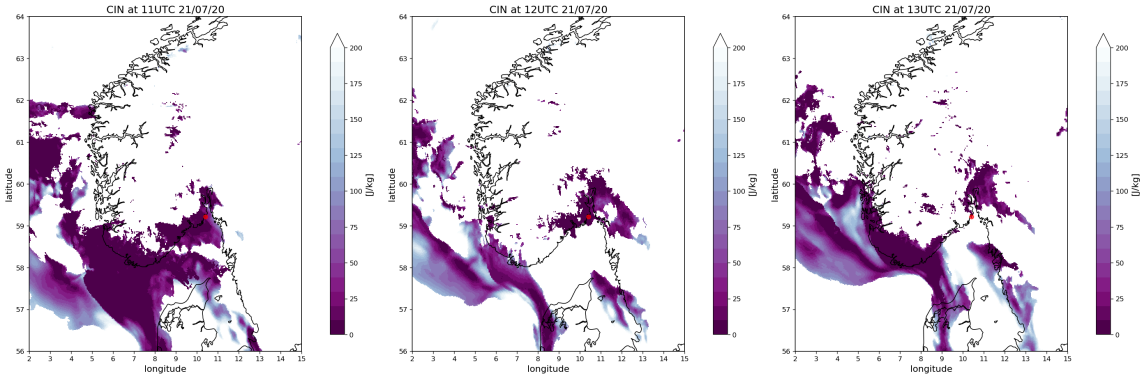

Figure 6.20: Stability indices CIN [J/kg] from AROME analysis data for different time steps at 21/08/20. The red dot show location of Nøtterøy.

Observations

On the 20th of August at 11:49 UTC, the day before the event, a yellow warning for torrential rain was sent out from The Meteorological Institute. On the 21st of August at 17:32 UTC, during the event, it was upgraded to orange. The max observed precipitation from the radar, shown in Figure 6.22, does not indicate a record-breaking extreme event in the area around Nøtterøy. Figure 6.22 show max observed hourly precipitation for every grid point (both maps), observations (left map), and lightning strikes the day of the event (right map). The ground observations do not correspond to the radar observations. The radar seems to capture the location of the event, but not the intensity. The radar shows a maximum of 19.5 mm/h at 13UTC, while the ground measurements are up to three times as large. It is conceivable that this can be an error caused when converting the reflectively to precipitation rate and/or the assumptions that the droplets are spherical and of a given size distribution.

The ground observations show how local this event was. There is a cluster of high precipitation measured around Nøtterøy, but not far away there are stations observing between 0 to 4 mm of precipitation in one hour. When observing the lightning strikes we detect the same; the band over Nøtterøy is narrow.

As mention earlier, a new record for observed precipitation for one hour was noted this day. At 13UTC in Sandefjord 63.3 mm/h was measured, and at Nøtterøy 58.2 mm/h was measured. The 200 years IVF-value for Nøtterøy is 38.2 mm/h. This highlights how extreme this event was.

6.2.2 Validation of the forecasts

We will in this section look closer at the forecasts made by AROME in the days before the event. The Meteorological Institute sent out a yellow warning for heavy rain showers the day before the event and upgraded it to orange during the event. Point verification, neighborhood verification in space, neighborhood verification in space and time and results from the ”closest point” method are presented in the following section. The different ensemble members are also evaluated in this section. All this is done to try to get an impression of how good the event was forecasted.

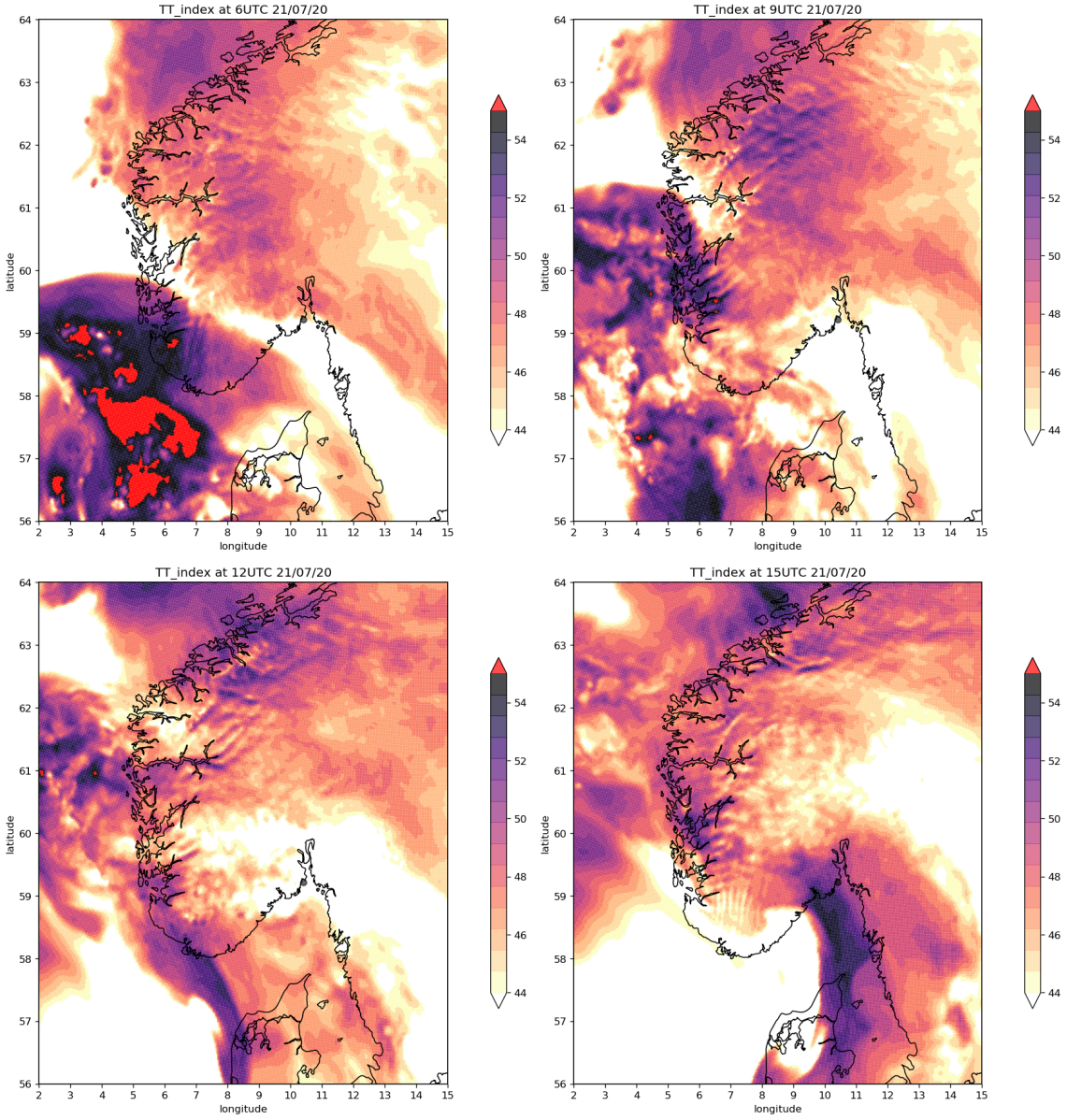

Figure 6.21: Stability indices TT index from AROME analysis data for different time steps at 21/08/20. The black dot show location of Nøtterøy.

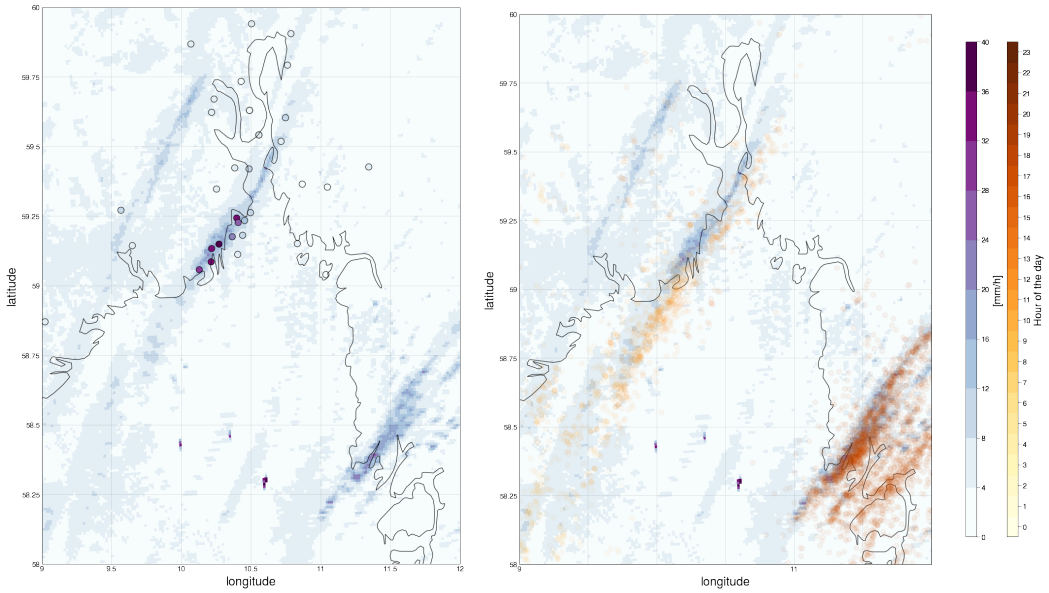

Figure 6.22: Map of Southern Norway showing maximum precipitation observed by the radar in every grid point the 21st of August (shaded purple) together with ground observations (dot purple) in the map to the *left*. Both plotted in the same color shading. The plot to the *right* shows the same maximum precipitation observed by the radar in every grid point on the 21st of August (shaded purple) together with lightning (shaded orange) data from the same day. The lightning data has color shading, after which hour the lightning stroked.

Neighborhood verification

In this section we present the results from point and neighborhood verification (explained in section 2.5.2). The verification is preformed for all available forecasts from AROME for the given event and with different types of neighborhood to get an impression on how changes in the definition of the neighborhood impact the results. Radar is used as the observation and the thresholds used are the 2 and 5 year return value for Nøtterøy. The result is evaluated by looking at categorical statistic.

There is a large spread in the bias frequency for Nøtterøy (Figure 6.23). The values show overall a very bad results. It is interesting to look at the forecast values made 12UTC (+25h) and 18UTC (+19h) the day before the event (20th of August). The forecast made at 12UTC (+25 h) and 18UTC (+19h) show very high values, especially for the point verification for both thresholds. High values indicated tendency of overcasting. Only studying the numbers from the neighborhood verification gives us the impression that the forecast made at 12UTC and 18UTC the day before the event is bad and overcasted. However, we need to take into account that even though the radar captures the location of the event, it does not capture the intensity. Figure 6.24 show a map of forecasted precipitation for the forecast made 12UTC (+25h) and 18UTC (+19h) the time of the event. The forecast show high forecasted precipitation over Nøtterøy, even though non of them forecasted the extremes that were observed (Figure 6.24). Comparing the forecast to the ground observations of over 60 mm/h and considering the location, the forecast does not seem to be overcasting.

Due to the error caused by bad radar data, no further categorical statistics is presented. Overall, considering the categorical statistics the event seem to be not well forecasted. Using radar as the truth when the radar is not capturing the intensity, the bias frequency get high when the forecast predict high values of precipitation. We will therefore look closer at the forecast using other methods. The ensemble members and the verification method called "closest point" method (Section 5.2) is included in the following sections.

Ensemble members

When predicting extremes it is interesting to evaluate the different ensemble members. This can give an indication of how confident the model is and show possible outcomes of the event that the ensemble mean does not capture.

There is an agreement between the different ensemble members that there will be precipitation between 8 UTC and 13UTC 21/08 (Figure 6.25). The different ensemble members agree that there will be almost no precipitation before and after the period. But the spread in the ensemble members regarding the amount of precipitation within the time range is large. For the forecast made at 12UTC on the 20th of August (+25h) (Figure 6.25 *upper*) on ensemble member forecast over 50 mm in three hours at the given point, and another one almost 40 mm in two hours. On the other end of the scale one of the ensemble members has an accumulated amount of only 5 mm for the whole period. The large spread of the ensemble members apply to the forecast made at 00 UTC on the 21st of August (+13h) (Figure 6.25 *lower*) as well. There are large discrepancies between the ensemble members and the two observations in Sandefjord and Nøtterøy. 50mm in three hours is is a very good prediction, but the intensity is still lower than observed.

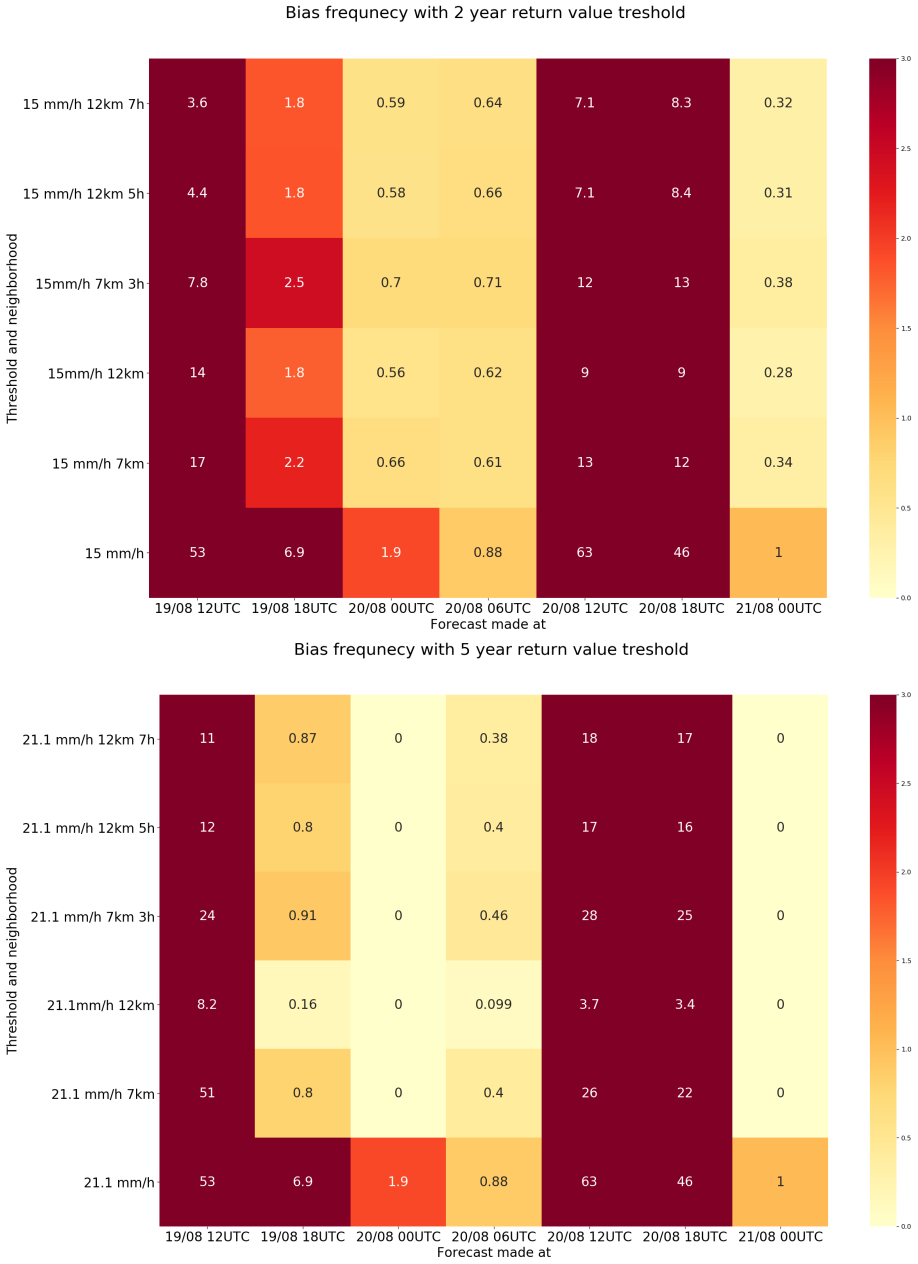

Figure 6.23: Bias frequency for Nøtterøy with 15 mm/h (upper) and 21.1 mm/h (lower) threshold. The frequency bias for different verification methods and neighborhood (y-axis) is shown. For example, '21.1 mm/h' means point verification with a threshold of 21.1 mm/h. '15 mm/h 7 km' mean neighborhood verification in space with a threshold of 15 mm/h and a spatial neighborhood of 7 km. The last type is a neighborhood in time and space: '21.1 mm/h 12km 5h' means a threshold of 21.1 mm/h with a neighborhood of 12 km range and 5 hours time range. The x-axis shows when the forecast is made.

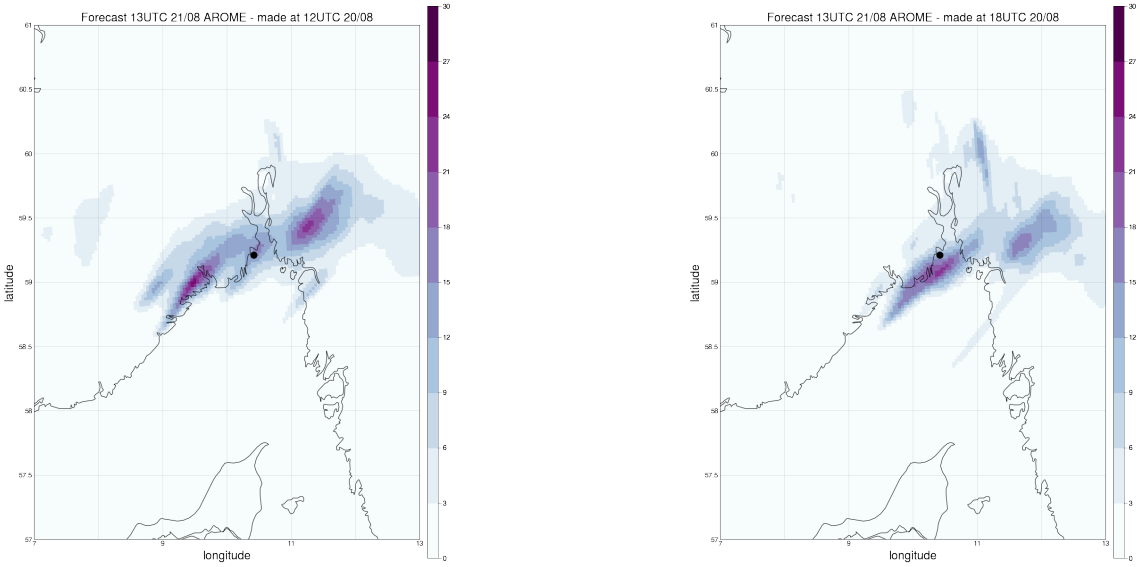

Figure 6.24: Forecasted precipitation for Nøtterøy (black dot) the time of the event (13UTC 21/08) for two forecasts made at different times. Left shows the forecast made at 12UTC 20/08, and right shows the forecast made at 18UTC 20/08.

To get a better understanding of what triggers the different ensemble members to predict intense precipitation, we compare the vertical stability of the ensemble member with highest total precipitation and intensity to the ensemble member with lowest. Figure 6.26 show the evolution of the vertical stability in time. The stability refers to the ability of the air to prevent vertical movement, and is calculated by looking at the change in potential temperature with height ($d\theta/dz$). Ensemble member 5 from the forecast made 00UTC the 21st of August (+13h) has the highest amount of accumulated precipitation, while ensemble member 7 has the least.

Figure 6.25 (*lower*) shows that the event is predicted to happen around 11UTC. For both members we have a situation of convective instability, whit an unstable layer at the ground (Figure 6.26). For ensemble members 5 the air at the ground gets more unstable around the time of when the event is predicted, followed by a build up of stable air. At higher altitude the layer of neutral air has a drop right in altitude before the event. After the unstable air is formed, the neutral layer higher up is lifted. This is not the case for ensemble member 7. Ensemble member 7 has stable air at the ground around the time when precipitation is predicted, and the higher altitude seem to be less effected by the ground. This can be a result of the convection being stronger for EM5, than EM7. Comparing to the event of Jølster, the layer of neutral air is at higher altitude for Nøtterøy than for Jølster.

Spacial correlation - ensemble members

Classifying the spacial correlation between the different ensemble members and the ensemble mean also say something about how confident the forecast model is. Figure 6.27 shows much of the same

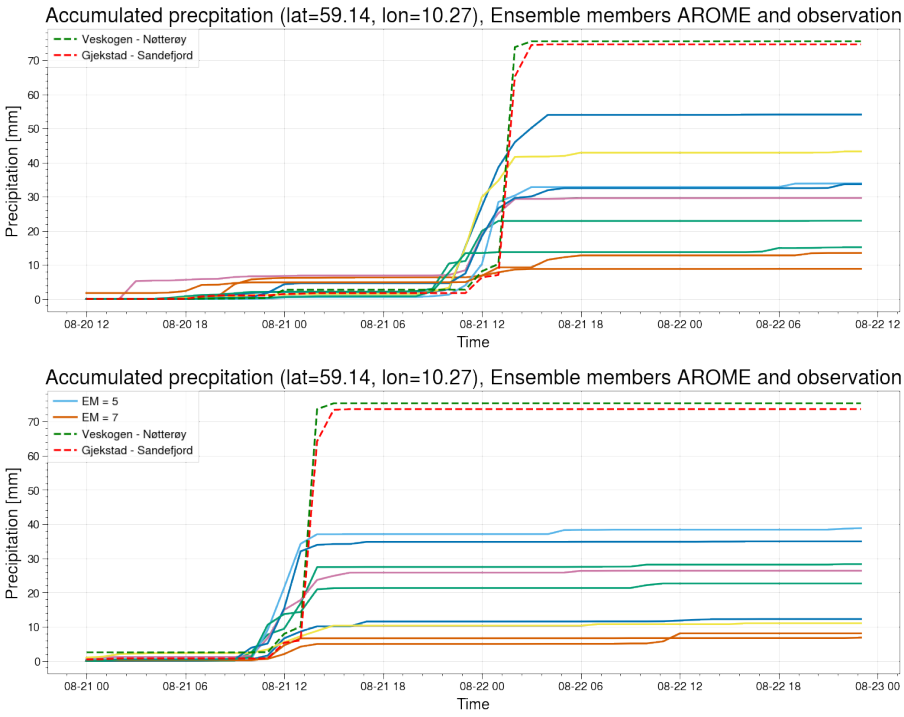

Figure 6.25: Time series of accumulated precipitation for all different ensemble members from the AROME forecast made at 12UTC 20/08 (*upper*) and 00UTC 21/08 (*lower*). The AROME accumulated precipitation is from Nøtterøy (lat=59.14, lon=10.27), and the observation show data from Vestskogen and Gjekstad.

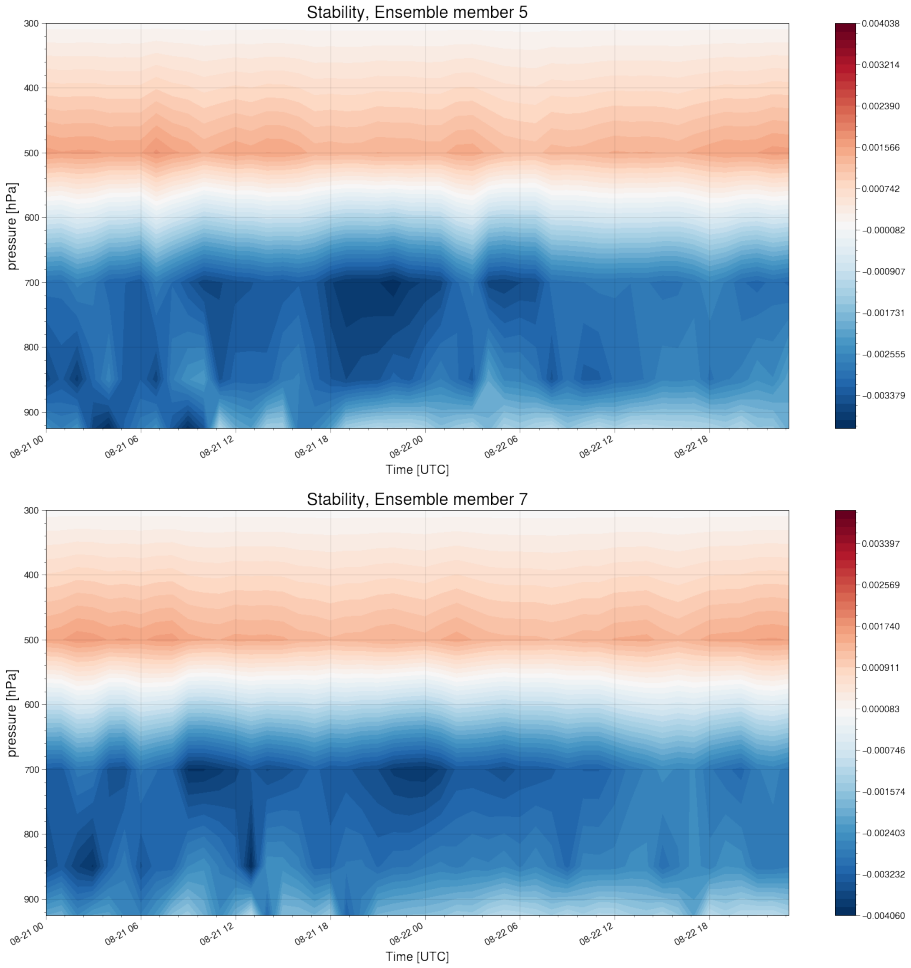

Figure 6.26: Evolution of vertical stability ($d\theta/dz$) for two different ensemble members. Time along the x-axis and pressure along the y-axis. Ensemble member 5 forecast little precipitation and has low intensity. Ensemble member 7 has the highest intensity of the ensemble members.

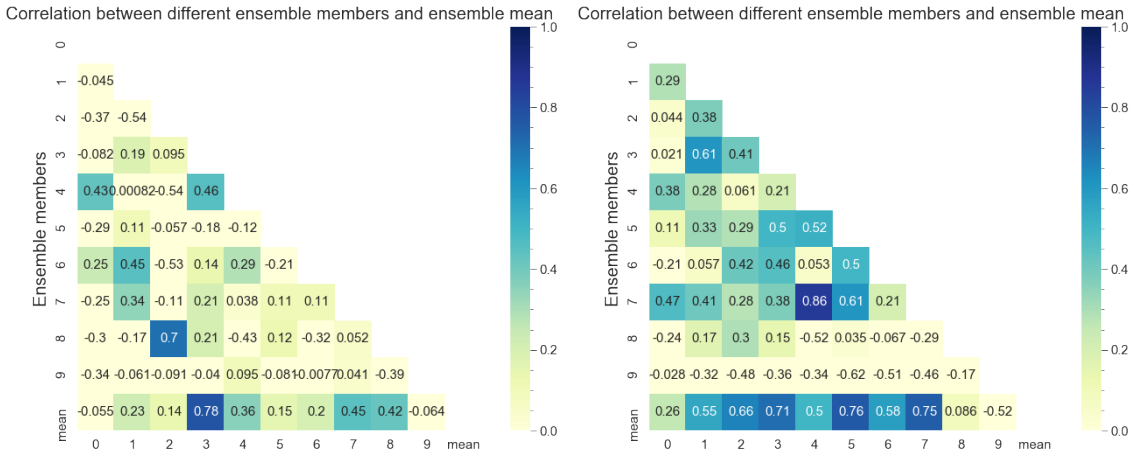

Figure 6.27: Spatial correlation between the different ensemble mean and the ensemble mean. The left figure shows the spatial correlation for the forecast made at 12UTC 29/08 (+25h) and the figure to the right for the forecast made at 00UTC 21/08 (+13h).

as Figure 6.25, the ensemble members agree more for the forecast closest to the event indicating that the model is more confident. But compared to the observations, some of the ensemble members from the forecast made +25h is closer to observation than for the forecast made +13h. There is no use of a confident model if the model is wrong.

Closest point verification

When working with extremes it can be interesting to find the closest grid square with forecasted value above a given threshold. As discussed before in this section, the neighborhood verification gives an indication that the event was not well forecasted from AROME. However, the neighborhood verification is dependent on a good radar estimate, when using radar as the truth. Looking at the closest point is not dependent on anything other than the forecast we are validating, the resolution of the model and the threshold you set.

Table 6.1 and 6.2 show the timesteps that gave hit within a circle of 111 km (correspond to 1° latitude) for the forecaste made at 12UTC on the 20th of August (+25h). We did the same analysis for the forecast from at 00UTC on the 21st of August (+13h), but this gave zero hits for all timesteps. It seems like the forecast made the day before the event at 12UTC is better than the forecast made the day of the event at 00 UTC. In general we expect the forecast to get better as we move toward the event. If the forecast is made too close to the event the result will be disturbed by so called spin up time. That is why no forecast closer to the event than midnight the day of the event is included in the thesis. However, +13h should be sufficient, and the event should not be affected by the spin up time.

The result from the "nearest point" verification method for a threshold of 15 mm/h (Table 6.1) at 13UTC shows good results, with a point 7 km away. With a resolution of 2.5 km this means a miss of 3 grids point, which can be considered good. Since this was a short duration event, we

Nearest point with a 2 years return value threshold				
Time	**11 UTC**	**12 UTC**	**13 UTC**	**14 UTC**
Hits within circle	479	569	141	5
Average distance for 3 nearest point	16,69	13,83	7,09	22,03

Table 6.1: Amount of hits and the average distance to the closest point within a circle of 1° forecast above a threshold. The circle is centered in Nøtterøy, and the threshold is set to the 2 year return value. The distance is the average distance of the three closest points above threshold.

Nearest point with a 5 years return values threshold				
Time	**11 UTC**	**12 UTC**	**13 UTC**	**14 UTC**
Hits within circle	101	100	21	0
Average distance for 3 nearest point	25,35	21,23	92,38	-

Table 6.2: Amount of hits and the average distance to the closest point within a circle of 1° forecast above a threshold. The circle is centered in Nøtterøy, and the threshold is set to the 5 year return value. The distance is the average distance of the three closest points above threshold.

want high values of precipitation close to Nøtterøy only at 13UTC. How the distance decrease as we move towards the time of the event for the 2 year return value, and increase after the event is positive. For the one with the 5 years return value it does not hit that well at the time of the event. An hour before it is 21,23 km away from Nøtterøy.

The "closest point" verification method can also be used for the ensemble members. Table 6.3 show the closest distance forecasted above the 2 years return value (15 mm/h) at the day of the event for all the ensemble members for the forecast made at 12UTC on the 20th of August (+25h). 5 out of 8 ensemble members have values above 15 mm/h less than 20 km away from Nøtterøy within the time period 10-18 UTC. 2 out of 8 members hit Nøtterøy at 13 UTC (time of the event), the distance is below the resolution of the AROME model. Finally, in the time period 10UTC-18UTC all members have values above the threshold within 40km radius.

6.2.3 WRF simulation

As for Jølster, we evaluate the event's sensitivity to changes in the numerical schemes by comparing different WRF runs. To evaluate the impact of the different schemes on the hourly precipitation maximum, the hourly maximum forecasted in every grid cell for the different runs are plotted

Ensemble members	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Time UTC	15	16	14	18	13	13	13	10	15
Closest point [km]	34,68	16,60	26,01	38,62	2,13	1,32	19,81	33,38	12,84

Table 6.3: The timestep and distance of the closest hit above the threshold for the different ensemble members. The closest point is defined as the average of the three nearest points. If the nearest hit is more than 1° away from Nøtterøy it is not considered.

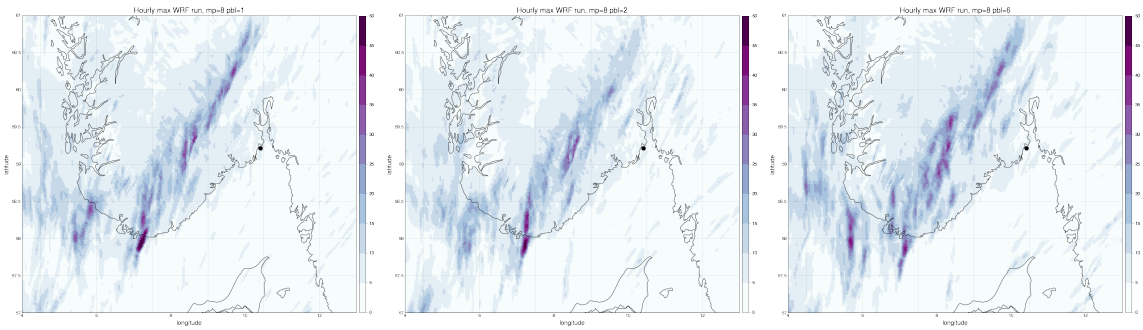

Figure 6.28: Maximum hourly precipitation for grid cell for WRF runs with different planetary boundary layer schemes. The black dot is showing the location of Nøtterøy. *mp* meaning microphysic scheme and *pbl* meaning planetary boundary layer scheme.

(Figure 6.28 and 6.29). All runs have a band of high values going over southern Norway. 50 mm in one hour is forecasted in the area outside Mandal in many of the runs, but no runs forecast hourly extreme values of precipitation at Nøtterøy. The change in cloud physics (microphysical) and turbulence (planetary boundary layer) scheme have a minor impact on the event. There are some difference in where the highest hourly extreme values are located, but the overall event are very similar.

The max hourly precipitation does not give an indication on the timing of the event. The accumulated precipitation show how the different runs agrees with timing, intensity and total accumulated predipitation for Nøtterøy (Figure 6.30). The runs agree on the timing, as well as the intensity from 10-12 UTC. The total amount of precipitation varies from around 11mm to 17.6 mm, considerably less compared to the observations at Nøtterøy. The model is not able to capture either the intensity, nor the amount of precipitation for the event.

Verification - closest point above threshold

For further verification of the different WRF runs the "closest point" method is used. Figure 6.31 show hits and distance to the closest point for the 2 year return value (*upper*) and hits within a circle for 5 year return value (*lower*). Due to the extremely low hits within circle for the ones with threshold of the 5 year return values, the closest distance is not included. For the 2 year return values only one run has hits for more than 2 timestep (run number 6 (mp=26, pbl=1)), and stands a bit out from the other. But with short time duration events we do not necessarily want hits at many timesteps, as it is suppose to be local and of short duration. We want hits at the time of the event.

Two of the runs peak in number of hits at 13UTC, the time of the event. Run number 1 (mp=1, pbl=1) has 72 hits, and has hits only at 13UTC. But the distance is 46 km from Nøtterøy. Run number 6 (mp=26, pbl=1) also peaks, and has hits both before and after the event. If the distance would decrease before the time of the event and increase afterwards this would be considered good,

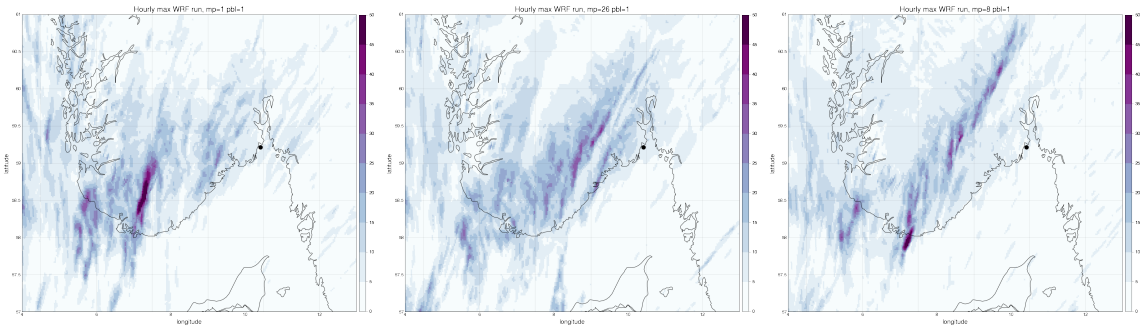

Figure 6.29: Maximum hourly precipitation for every grid cell for the WRF runs with different microphysical schemes. The black dot is showing the location of Nøtterøy. *mp* meaning microphysic scheme and *pbl* meaning planetary boundary layer scheme.

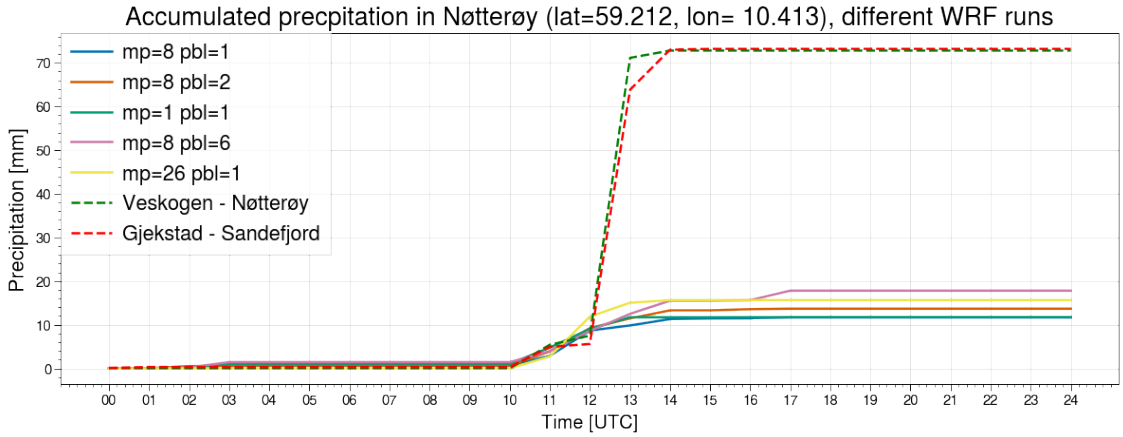

Figure 6.30: Accumulated precipitation for Nøtterøy for different WRF runs and observations (dashed line). As observation, Gjekstad and Vestskogen is used.

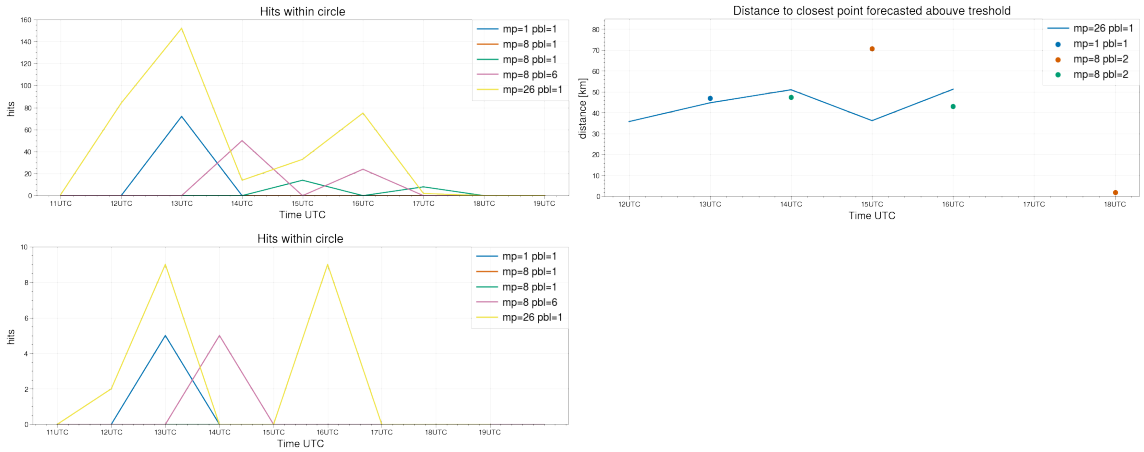

Figure 6.31: Number of hits above 2 (*upper left*) and 5 (*lower left*) year return value within a circle with a radius of 1°and average distance to three closest point forecasted above 2 year return value (*upper right*). The x-axis show time and the y-axis show number of hits (*left*)/ distance [km] (*right*).

but this is not the case. The closest distance for this run is above 35km for all timesteps.

7. Discussion

The events' similarities and differences, focusing on the impact this had on the different results, will be discussed in the following chapter. Verification of these type of events will also be discussed, in light of the results of the various verification methods.

The two different events have some similarities and some clear differences. The event in Nøtterøy is much more local and of shorter duration than the event in Jølster. For Jølster precipitation was widespread over large parts of Western Norway. For both events, the air is advected over the location by a synoptic weather system creating an unstable situation triggered at the given locations. For Jølster, cold air is advected in over a hot surface, creating unstable air. For Nøtterøy, the unstable air seems not to be created at the location but advected in toward Nøtterøy from the ocean at southwest and then triggered at Nøtterøy.

Only the Jølster event had good enough radar data to evaluate the advantages of using neighborhood verification. The radar did not capture the event at Jølster, only the precipitation around Jølster. Jølster had no ground measurements, so to validate the radar observation I comparing the radar to the lightning data. There was observed a lot of lightning trikes at Jølster, indicating convection and precipitation. The radar estimates for Nøtterøy was so bad that the verification results became misleading. The verification implied overcasting when the forecast was relatively good compared to the point observations, and good results when almost no precipitation was forecasted. That the radar did not capture the intensity of the event at Nøtterøy was a bit surprising. Issues with the radar not capturing precipitation due to the topography in the fjords of Norway, like for Jølster, is a well-known problem. However, this should not be an issue in Nøtterøy. Located at the coast, on a hot summer day, with minor disturbance from topography the conditions seemed suitable for good observations with a radar. But, as mentioned earlier, this might be due to the converting from reflectively to precipitation rate and/or the assumptions that the droplets are spherical and of a given size distribution. When we do not have good observations of these extreme events, it limits research on these events and the possibility of better understanding.

A results of the events being different, regarding the area covered and the duration of the event, is that bad radar data will impact the result of the neighborhood verification more significantly for Nøtterøy due to the how extremely local and short duration it was. For Jølster, there were high amount of precipitation over a large part of Western Norway. The radar observed high values around Jølster, but not at Jølster. The neighborhood verification method still gave good results. By choosing a neighborhood for both observation and forecast, the verification gave good results by using the observations around Jølster. The frequency bias showed numbers close to 1 indicating

no over- or undercasting. It should be mentioned that this type of neighborhood can be misleading, neighborhood for both the observation and the forecast can providing credit to grid squares where it should not. As mention earlier; there is a fine line between being very forgiving by having a wide neighborhood and losing too much temporal and spatial information. But for Jølster we know that heavy precipitation occurred, even though the radar did not capture it. Therefore it is nice to use the observed values from the radar around Jølster, to give credit to forecasted values at and around Jølster. For this event, we see some of the advantage of using neighborhood verification.

The method "nearest point" very clearly showed the disadvantage of neighborhood verification for these type of events. The results from neighborhood verification for Jølster showed relatively good results, especially for the frequency bias. We got the impression that the forecast was relatively good. This is probably the correct impression looking at the overall forecast for the area chosen around Jølster. However, when looking at the result from the "nearest point" method, we see that the forecast never predicted precipitation above the threshold closer to Jølster than 14 km away. The extreme values at Jølster were never forecasted, even though the overall forecast was good.

Verification of local short duration extremes can be difficult. As discussed, the type of event you are verifying and what you are using the result for is essential for how you do the verification. But also, the access to observations is shown to be a problem when validating local short duration precipitation events. Using methods, as the "nearest point" method, which is not dependent on how certain the observation is, is a solution. This method only gives information about the nearest point and hits within the area and nothing about the overall forecast, such as the frequency of overcasting or undercasting.

Comparing the spread of the ensemble members at the event of Nøtterøy with the spread of the event in Jølster, we examine that for Nøtterøy the spread is larger and more uniformly distributed within the range. For Jølster there is a significant difference between maximum and minimum, but mainly because one of the ensemble members stands out from the rest. But there is no need for a confident model (meaning low spread in ensemble members) if the model is fare away from the ground observation. When evaluating the ensemble members, we saw how the model could not forecast the intensity of the events. For Jølster, were one of the ensemble members forecast 30 mm in two hours, we have an observation in Vassenden measuring 30 mm in one hour before it break down. At Vassenden 65 mm is observed in two hours. It can seem strict to say that when the forecast show 30 mm in two hours and the ground observation show 30 mm in one hour, it is not good enough. But this is often the difference between a heavy precipitation event and an extreme event. How dangerous the short duration event is, depends on the intensity. If we would have 30 mm in two hours, the landslide that leads to a casualty in Jølster may never have happened. The same is for Nøtterøy, where one ensemble member forecasted 50 mm in three hours. This is a very high amount of precipitation and warns that a high amount of precipitation can find place, but one of the ground observations measured over 60mm in one hour. Meaning that the difference between forecasted and observed intensity is very high.

None of the runs with WRF done for this thesis forecasted heavy precipitation at the location investigated. Therefore, it is natural to discuss whether you can say something about the driving mechanism when the actual event is not modeled. The results from the runs for both the events

were overall that the changes in parameterization of the cloud physics and turbulence had a minor impact on the event, both for the timing and the total accumulated amount. Therefore, it is tempting to conclude that the events were not mainly driven by turbulence or cloud physics and more likely to be large-scale driven. But as mention before, the actual event is not forecasted, so how can changing the schemes and observing the forecast say something about the event? Even though the event's exact location is not recreated, the large scale situation and large scale event are still modeled, and we can therefore say that the changes do not impact the event.

8. Conclusion

The goal of the thesis was to understand what happened at the events in Nøtterøy and Jølster and to verify the forecast made for the events. To conclude, I will give a short summary of the main results discussed in this thesis:

- For both the events the synoptic situation lead to unstable air over a large area. For Jølster the easterlies brought cold air in over Norway, creating unstable air over large parts of the west coast of southern Norway. A triggering mechanism triggered an extreme amount of precipitation locally. None of the stability parameters indicated that Jølster was at more risk than the rest of the west coast of southern Norway. These types of events are hard to predict since you triggering mechanism is unknown. This was shown when verifying the forecasts. For Nøtterøy, a low pressure gave onshore winds, bringing unstable air in over land. The stability parameter showed more clearly that the unstable air would hit Nøtterøy, compared to the event in Jølster. Some of the ensemble members for Nøtterøy showed good forecasts. The forecast was further away from reality as we got closer to the event.

- The radar estimates did not capture any of the events in a good way and were shown to be unreliable. Comparing the lightning data to the radar data we saw how convection found place in Jølster, even though the radar did not capture any high amount of precipitation. In Jølster, the radar did not capture the precipitation probably due to topography. In Nøtterøy the intensity was not captured, possibly due to the assumptions done when converting the signal to precipitation amount. This has a significant impact on the verification.

- The sensitivity to the radar goodness of the observation regarding neighborhood verification was shown very clearly in this thesis, especially for Nøtterøy. It worked better for Jølster and we observed that the size of the neighborhood did not have such a significant effect on the bias frequency, but had a more substantial impact on the hit rate and false alarm ratio. The neighborhood verification showed that the forecast for the area chosen around Jølster was good in terms of frequency bias, with values close to 1 indicating no over- or undercasting.

- The ”closest point” method showed how some of the ensemble members for Nøtterøy were very good. However, it also showed how the forecast for Jølster never really hit Jølster, even though the neighborhood verification shows that the forecast was good. This leads to the conclusion that verifying local short duration events is difficult, and that statistics from multiple verification methods are needed to give a complete assessment of the goodness of forecast.

- The models seem to have a problem predicting the precipitation intensity. The heavy convection is not strong enough. The only place where we found that the WRF model was able to model the same intensity observed was a bit east of Kvamskogen.
- Evaluating the WRF runs we saw that the changes in schemes seem to have little impact on the timing of precipitation or on the total accumulated precipitation. This indicates that the primary driving mechanism is not the microsphysic or the turbulence.

As further work, it would be interesting to do the same procedure of verification for an event where the radar captures the event better than for the two events I examined. Also, to look even closer into the difference between the ensemble members that forecast the intensity well and the one that did not. The lack of available data provided challenges for further investigating on the ensemble members. The WRF models also provide many opportunities for further investigation of the events. It would be interesting to look at how the events would be affected by a temperature increase, giving helpful information on how the events are affected by climate change. Another opportunity would be to smoothing the topography out to investigate the impact of topography on the events.

Appendices

.1 WRF theroy

.1.1 WPS

WPS is the WRF model processing system. It consists of thee programs, all of which contribute to preparing input to the *real* program for real-data systems. The gridded data (Figure 1) is handled by the program *ungrib*, which extracts the meteorological fields from GRIB-formatted files. *Geogrid* is the second program in WPS. This program defines the model domains and interpolates statistic geographical data to the grid using the WRF Terrestrial Data (Figure 1). The last program, *metgrid*, use extracted meteorological field from ungrib and horizontally interpolates it to the model grid defined by geogrid. The outcome of geogrid is then sent to real [Wang et al., 2018].

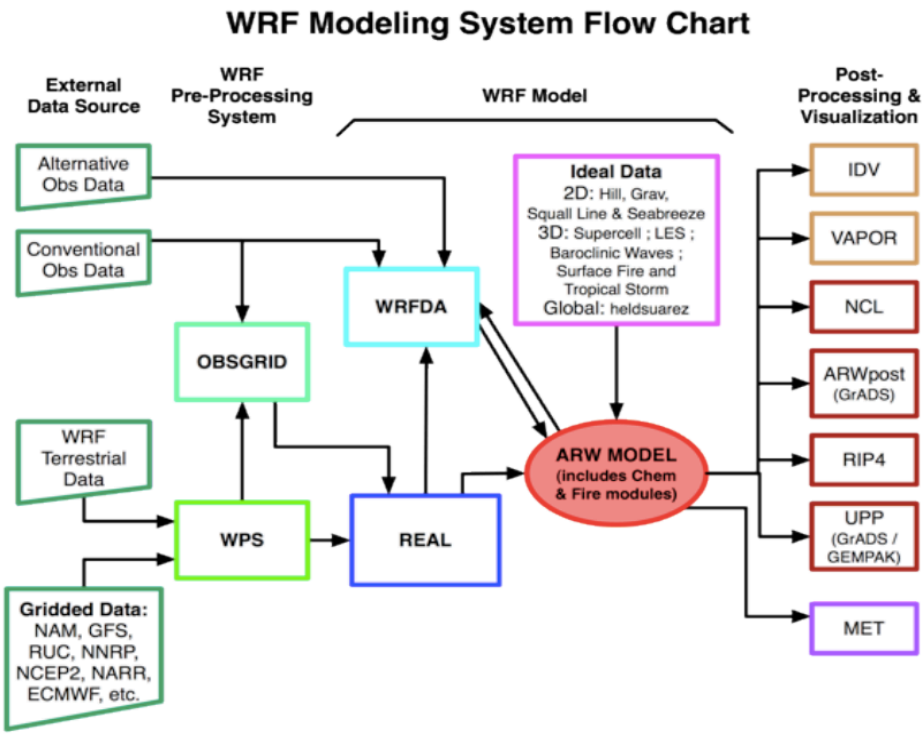

Figure 1: A illustration showing the WRF Modeling System [Wang et al., 2018].

Boundary Conditions

ARW provides many different lateral boundary conditions, and they are handled in the run-time options in the Fortran namelist file. For real-data cases, the lateral boundary conditions are split into two zones: the relaxation zone and the specified zone. The WPS contributes the specified zone with information. It contributes with information about humidity, potential temperature,

perturbations field, and the horizontal wind components for all four sides of the coarse domain. In the relaxation zone, the model is relaxed or nudged toward the large scale field. You can also include smaller domains inside the model domain you choose for your model, which we call nested domains. This domain uses the information from the original domain and not WPS. So you run the model for your model domain at a time step, and then this information is used to run for the nested domain.

Time Discretization

A time-split integration scheme is being used in the ARW solver. There is a wide range of frequencies in the atmosphere, and we split the frequency into a high- and low-frequency mode. For the high-frequency, the acoustic modes are integrated over smaller time steps to maintain numerical stability. For low-frequency, the modes are integrated by using the third-order Runge-Kutta (RK3) time integration scheme. The RK3 integrates a set of ordinary differential equations by using a predictor-corrector formulation [Skamarock et al., 2008].

Spatial Discretization

Using a numerical model to recreate the atmosphere's situation, we need to discretize it into fitting a grid. There are many different numerical grids. The ARW solver uses an Arakawa C grid staggering for the spatial discretization. An illustration of the grid is shown in Figure 2. The grid is staggered, meaning that some variables, in this case, the velocities, are staggered on-half grid length from some other variables, the thermodynamic variables. The θ is located at the mass point, and u,v, and w are defined as u points, v points, and w points. For the diagnostic variables in this model, the pressure p and inverse density α, are computed in the mass point. The length of the grid, Δx and Δy, are constant in the formulations of the model, while the vertical grid length $\Delta\eta$ is not, it is specified in the initialization.

.1.2 Governing equations

The ARW solver integrated the compressible, non-hydrostatic Euler equations. The vertical coordinate used to formulate the equation is the so-called terrain-following hydrostatic-pressure. The vertical coordinated is denoted by η and is called hybrid sigma-pressure vertical coordinate and was described by Park et al. [2013].

$$p_d = B(\eta)(p_s - p_t) + [\eta - B(\eta)](p_0 - p_t) + p_t \tag{1}$$

In equations 1 the p_0 is the reference sea-level pressure. B(η) is a third-order polynomial defining how to weight between the terrain-following sigma coordinates and just the pure pressure coordinate. p_d is the hydrostatic component of the pressure for dry air, and p_t and p_s is referring to values of p_d along the top and the surface boundaries.

The vertical coordinate metric is defines as

$$\mu_d = \partial p_d / \partial \eta = B_\eta(\eta)(p_s - p_t) + [1 - B_\eta(\eta)](p_0 - p_t) \tag{2}$$

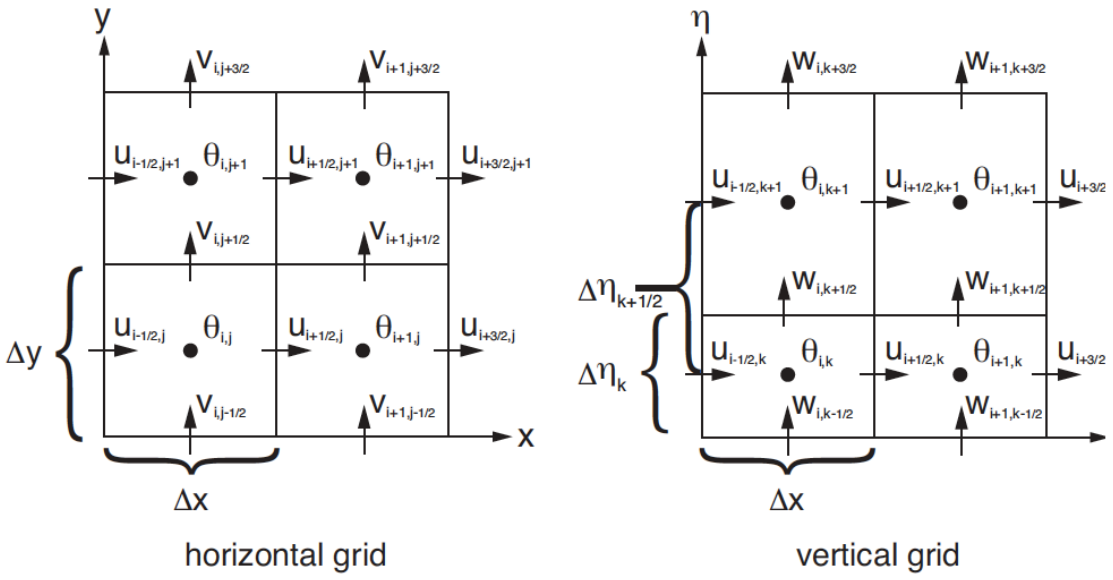

Figure 2: Horizontal and vertical grids of the ARW solver for the Arakawa C grid staggereing [Skamarock et al., 2008].

The flux form for the prognostic variable are defined as followed:

$$\mathbf{V} = \mu_d \mathbf{v} = (U, V, W), \Omega = \mu_d \omega, \Theta = \mu_d \theta_m, Q_m = \mu_d q_m. \tag{3}$$

where the covariant velocities in the horizontal and vertical direction is denoted by v=(u,v,w) and the contravariant "vertical" velocity is denoted $\omega = \dot{\eta}$. θ_m is the moist potential temperature, and q_m represents the mixing ratio of moisture variables. With these defined variables, the flux-form Euler equation can be defined:

$$\partial_t U + (\nabla \cdot \mathbf{V} u) + \mu_d \alpha \partial_x p + (\alpha/\alpha_d)\partial_\eta p \partial_x \phi = F_U \tag{4}$$

$$\partial_t V + (\nabla \cdot \mathbf{V} v) + \mu_d \alpha \partial_y p + (\alpha/\alpha_d)\partial_\eta p \partial_y \phi = F_V \tag{5}$$

$$\partial_t W + (\nabla \cdot \mathbf{V} w) + g[(\alpha/\alpha_d)\partial_\eta p - \mu_d] = F_W \tag{6}$$

$$\partial_t \Theta_m + (\nabla \cdot \mathbf{V} \theta_m) = F_{\Theta_m} \tag{7}$$

$$\partial_t \mu_d + (\nabla \cdot \mathbf{V}) = 0 \tag{8}$$

$$\partial_t \phi + {\mu_d}^{-1}[(\nabla \cdot \mathbf{V}) - gW] = 0 \tag{9}$$

$$\partial_t Q_m + (\nabla \cdot \mathbf{V} q_m) = F_{Q_m} \tag{10}$$

including the diagnostic equation for dry hydrostatic pressure

$$\partial_\eta \phi = -\alpha_d \mu_d \tag{11}$$

and the diagnostic relation for pressure with both dry air and water vapour

$$p = p_0 (\frac{R_d \theta_m}{p_0 \alpha_d})^\gamma. \tag{12}$$

In these equations, α is the inverse density taking into account the full parcel density. α_d is the inverse density for dry air.

The subscriptiom x,y and η denote differentiation in (4)-(10),

$$\nabla \cdot \mathbf{V}_a = \partial_x (Ua) + \partial_y (Va) + \partial_\eta (\Omega a) \tag{13}$$

and,

$$\mathbf{V} \cdot \nabla_a = U \partial_x a + V \partial_y a + \Omega \partial_\eta a, \tag{14}$$

where a represent a given variable. γ is the ratio between the heat capacity for dry air with constant volume and constant pressure. R_d is the gas constant for dry air, and p_0 is a reference surface pressure. The terms on the right-hand-side represent forcing terms. These terms are due to the model physics, turbulent mixing, spherical projections, and the earth's rotation [Skamarock et al., 2008].

References

C. Donald Ahrens. *Essentials of Meteorology - An Invitation to the Atmosphere*. Cengage Learning, 2015.

Soo Ya Bae, Song You Hong, and Wei Kuo Tao. Development of a Single-Moment Cloud Microphysics Scheme with Prognostic Hail for the Weather Research and Forecasting (WRF) Model. *Asia-Pacific Journal of Atmospheric Sciences*, 55(2):233–245, 2019. ISSN 19767951. doi: 10.1007/s13143-018-0066-3.

Lisa Bengtsson, Ulf Andrae, Trygve Aspelien, Yurii Batrak, Javier Calvo, Wim de Rooy, Emily Gleeson, Bent Hansen-Sass, Mariken Homleid, Mariano Hortal, et al. The harmonie–arome model configuration in the aladin–hirlam nwp system. *Monthly Weather Review*, 145(5):1919–1935, 2017.

Harold E. Brooks. Proximity soundings for severe convection for Europe and the United States from reanalysis data. *Atmospheric Research*, 93(1-3):546–553, jul 2009. ISSN 01698095. doi: 10.1016/j.atmosres.2008.10.005.

D Cimini, M Nelson, J Güldner, and R Ware. Forecast indices from a ground-based microwave radiometer for operational meteorology. *Atmos. Meas. Tech*, 8:315–333, 2015. doi: 10.5194/amt-8-315-2015. URL `www.atmos-meas-tech.net/8/315/2015/`.

Christoffer A Elo. Correcting and quantifying radar data. (2), 2012.

Dyrrdal A. V. Førland E. J. Climatic changes in short duration extreme precipitation and rapid onset flooding-implications for design values. Technical report, 2018.

J. F. P. Galvin. *An introduction to the metoeorology and climate of the tropics*. 2016.

Eric Gilleland, David Ahijevych, Barbara G. Brown, Barbara Casati, and Elizabeth E. Ebert. Intercomparison of spatial forecast verification methods. *Weather and Forecasting*, 2009. ISSN 08828156. doi: 10.1175/2009WAF2222269.1.

Jerome Hakim, Gregory J. Patoux. *Weather - a concise introduction*. 2017.

I Hanssen-Bauer, E J Førland, I Haddeland, H Hisdal, S Mayer, A Nesje, J E Ø Nilsen, S Sandven, A B Sandø, A Sorteberg, and B Ådlandsvik. Klima i Norge 2100 Kunnskapsgrunnlag for klimatilpasning oppdatert i 2015. Technical report, 2015. URL `www.miljodirektoratet.no/20804`.

Song You Hong and Hua Lu Pan. Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Monthly Weather Review*, 1996. ISSN 00270644. doi: 10.1175/1520-0493(1996)124<2322:NBLVDI>2.0.CO;2.

Song You Hong, Yign Noh, and Jimy Dudhia. A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, 2006. ISSN 00270644. doi: 10.1175/MWR3199.1.

Song-You Hong, Kyo-Sun Sunny Lim, Yong-Hee Lee, Jong-Chul Ha, Hyung-Woo Kim, Sook-Jeong Ham, and Jimy Dudhia. Evaluation of the WRF Double-Moment 6-Class Microphysics Scheme for Precipitating Convection. *Advances in Meteorology*, 2010, 2010. doi: 10.1155/2010/707253.

RA Houze. Stratiform precipitation in regions of convection. *Bulletin of the American Meteorological Society*, pages 2179–2195, 1997.

Yi Huang and Xindong Peng. Improvement of the Mellor-Yamada-Nakanishi-Niino Planetary Boundary-Layer Scheme Based on Observational Data in China. *Boundary-Layer Meteorology*, 162:171–188, 2017. doi: 10.1007/s10546-016-0187-0.

Edwin Kessler. *On the Distribution and Continuity of Water Substance in Atmospheric Circulations*. 1969. doi: 10.1007/978-1-935704-36-2.

Medha Khole and H R Biswas. Role of total-totals stability index in forecasting of thunderstorm/non-thunderstorm days over Kolkata during pre-monsoon season. Technical report, 2007.

E. Philip. Krider. "Thunderstorm", 2019. URL `https://www.britannica.com/science/thunderstorm`.

Ingela Leffler. Predicting Aviation Hazards during Convective Events Förutsäga farligt flygväder i samband. 2015.

J. S. Marshall and W. Mc K. Palmer. THE DISTRIBUTION OF RAINDROPS WITH SIZE. *Journal of Meteorology*, 1948. ISSN 0095-9634. doi: 10.1175/1520-0469(1948)005<0165:tdorws>2.0.co;2.

Meteorologisk Institutt. Obs-view Info om lynobservasjoner i Norge, 2017. URL `https://lyn.met.no/info.html`.

Meteorologisk Institutt. Intense byger med store konsekvenser i Sogn og Fjordane 30. juli 2019. page 42, 2019.

Clio Michel, Asgeir Sorteberg, Sabine Eckhardt, and Chris Weijenborg. Extreme precipitation and atmospheric rivers in.

Malte Müller, Mariken Homleid, Karl Ivar Ivarsson, Morten A.Ø. Køltzow, Magnus Lindskog, Knut Helge Midtbø, Ulf Andrae, Trygve Aspelien, Lars Berggren, Dag Bjørge, Per Dahlgren, Jørn Kristiansen, Roger Randriamampianina, Martin Ridal, and Ole Vignes. AROME-MetCoOp: A nordic convective-scale operational weather prediction model. *Weather and Forecasting*, 2017. ISSN 15200434. doi: 10.1175/WAF-D-16-0099.1.

A. H. Murphy. What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather & Forecasting*, (2):281–293, jun 1993. ISSN 0882-8156. doi: 10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

Mikio Nakanishi and Hiroshi Niino. Development of an improved turbulence closure model for the atmospheric boundary layer. *Journal of the Meteorological Society of Japan*, 87(5):895–912, 2009. ISSN 00261165. doi: 10.2151/jmsj.87.895.

Sang Hun Park, William C. Skamarock, Joseph B. Klemp, Laura D. Fowler, and Michael G. Duda. Evaluation of global atmospheric solvers using extensions of the jablonowski and williamson baroclinic wave test case. *Monthly Weather Review*, 2013. ISSN 00270644. doi: 10.1175/MWR-D-12-00096.1.

Heather Pier. Convective inhibition (cin): Definition role in forecasting, 2018. URL `study.com/academy/lesson/convective-inhibition-cin-definition-role-in-forecasting.html`.

Jean-Pierre Pinty and Patrick Jabouille. A mixed-phase cloud parameterization for use in mesoscale non-hydrostatic model: simulations of a squall line and of orographic precipitations. In *Conf. on Cloud Physics*, pages 217–220. Amer. Meteor. Soc Everett, WA, 1998.

Harald Reitan. *Været i Norge*. 2005.

J Reuder. Physical meteorology : Part II : Cloud physics. 2019.

Andrea Rossa, Pertti Nurmi, and Elizabeth Ebert. Overview of methods for the verification of quantitative precipitation forecasts. *Precipitation: Advances in Measurement, Estimation and Prediction*, pages 419–452, 2008. doi: 10.1007/978-3-540-77655-$0_1$6.

Paul Schultz. Relationships of Several Stability Indices to Convective Weather Events in Northeast Colorado. *Weather and Forecasting*, 4(1):73–80, 1989.

Craig S. Schwartz. A Comparison of Methods Used to Populate Neighborhood-Based Contingency Tables for High-Resolution Forecast Verification. *Weather and Forecasting*, 32(2):733–741, 2017. ISSN 0882-8156. doi: 10.1175/waf-d-16-0187.1.

R. Seager, D. S. Battisti, J. Yin, N. Gordon, N. Naik, A. C. Clement, and M. A. Cane. Is the Gulf Stream responsible for Europe's mild winters? *Quarterly Journal of the Royal Meteorological Society*, 128(586 PART B):2563–2586, oct 2002. ISSN 00359009. doi: 10.1256/qj.01.128.

Y. Seity, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson. The AROME-France convective-scale operational model. *Monthly Weather Review*, 2011. ISSN 00270644. doi: 10.1175/2010MWR3425.1.

Kuganesan Sivasubramaniam, Ashish Sharma, and Knut Alfredsen. Estimating radar precipitation in cold climates: The role of air temperature within a non-parametric framework. *Hydrology and Earth System Sciences*, 22(12):6533–6546, 2018. ISSN 16077938. doi: 10.5194/hess-22-6533-2018.

W C Skamarock, J B Klemp, J B Dudhia, D O Gill, D M Barker, M G Duda, X.-Y. Huang, W. Wang, and J G Powers. A description of the Advanced Research WRF Version 3, NCAR Technical Note TN-475+STR. *Technical Report*, (June):113, 2008. ISSN 1477870X.

Asgeir Sorteberg. Nedbør i Norge siden 1900. *Naturen*, 138(06):221–231, 2014. ISSN 0028-0887. URL `http://www.idunn.no/natur/2014/06/nedboer_i_norge_siden_1900`.

Asgeir Sorteberg, Stephanie Mayer, and Anita Verpe Dyrrdal. Ekstremnedbør i et klima i forandring. *Naturen*, 142(06):246–251, 2018. ISSN 0028-0887. doi: 10.18261/issn.1504-3118-2018-06-04.

David J. Stensrud. *Parameterization Scheme - Keys to Understanding Numerical Weather Prediction Models*. Cambrige University Press, 2007.

David B. Stephenson. Use of the 'odds ratio' for diagnosing forecast skill. *Weather and Forecasting*, 15 (2):221–232, 2000. ISSN 08828156. doi: 10.1175/1520-0434(2000)015<0221:UOTORF>2.0.CO;2.

Gregory Thompson, Paul R. Field, Roy M. Rasmussen, and William D. Hall. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review*, 2008. ISSN 00270644. doi: 10.1175/2008MWR2387.1.

Jeroen van Zomeren and Aarnout van Delden. Vertically integrated moisture flux convergence as a predictor of thunderstorms. *Atmospheric Research*, 83(2-4 SPEC. ISS.):435–445, feb 2007. ISSN 01698095. doi: 10.1016/j.atmosres.2005.08.015.

Wei Wang, Cindy Bruyère, Michael Duda, Jimy Dudhia, Dave Gill, Michael Kavulich, Kelly Werner, Ming Chen, Hui-Chuan Lin, John Michalakes, Syed Rizvi, Xin Zhang, Judith Berner, Domingo Munoz-Esparza, Brian Reen, Soyoung Ha, and Kate Fossell. Advanced Research WRF (ARW) Version 4.0 modeling system User's Guide. (June):464, 2018.

S Westra, H J Fowler, J P Evans, L V Alexander, P Berg, F Johnson, E J Kendon, G Lenderink, and N M Roberts. Reviews of Geophysics. pages 1–34, 2014. doi: 10.1002/2014RG000464.Received.