University of Bergen

Geophysical Institute

Validation of precipitation forecasts from the AROME numerical weather prediction model in Norway

Jostein Karlstrøm Aga

Master's thesis

Master's Specialisation:
Climate Dynamics

Supervisor:
Asgeir Sorteberg

November 2022

Preface

This thesis is written as part of the master programme in meteorology and oceanography at the University of Bergen, and was done through the year 2022.

I would like to thank my supervisor Asgeir Sorteberg for helping me through this MSc thesis, providing useful advice whenever I asked for it, and for always bringing a positive mindset to our meetings.

In addition, I want to thank my family for supporting me through my entire study period, and encouraging me to finally get this thesis over the finish line.

Abstract

High-quality precipitation forecasts are key to ensure the public and economic safety during severe precipitation events, and the process of validating these forecasts is a continuous effort. But with Norway's varied climate and landscape, this can pose a challenge. In this thesis, the precipitation forecast data from the post-processed AROME-MetCoOp model were validated against observational data over a period from 1. December 2019 - 31. April 2022 in these six locations: Bergen, Oslo, Trondheim, Tromsø, Kristiansand and Nesbyen.

The results were split into a climatology part and verification part. For climatology, Bergen and Tromsø forecasted way too little total precipitation, with the biggest deviation during summer for Bergen and winter for Tromsø. This was not due to a bias on mean precipitation amount in the model, but it could be due to the model underestimating the orographic enhancement. The model predicted a bit too much winter precipitation Oslo, Kristiansand and Nesbyen, which could be related to wind-induced undercatch of solid precipitation, although more research is needed.

Precipitation distribution seemed to be somewhat narrow overall, forecasting too many low-intensity precipitation events, but struggling to forecast enough extreme precipitation. For verification results, forecast quality remained fairly constant with increasing forecast lengths (up to +48h ahead), and improved slightly with longer accumulation lengths (also up to 48h). It looks like the model performs better when the (high) hourly variability gets averaged out.

All in all, Kristiansand was the best-performing location, while Tromsø saw the poorest results.

Contents

Preface **i**

Abstract **ii**

1 Introduction **1**

2 Theory **4**

2.1 Forecast Verification 4

2.1.1 What is a good forecast? 4

2.1.2 Spatial verification methods 6

2.1.3 Deterministic forecast verification 7

2.1.4 Dichotomous forecast verification 8

2.1.5 Probabilistic forecast verification 10

2.2 Precipitation formation 11

2.2.1 Frontal precipitation 12

2.2.2 Convective precipitation 13

2.2.3 Orographic precipitation 14

3 Methods **15**

3.1 Observations 15

3.1.1 Measuring precipitation at weather stations 15

3.1.2 Obtaining observational data 17

3.2 AROME-MetCoOp 21

3.2.1 Introduction to AROME 21

3.2.2 Obtaining post-processed AROME forecast data 24

3.3 Processing the data 25

3.3.1 Missing values 25

3.3.2 Precipitation thresholds 26

3.4 Custom-made forecasts 26
3.5 Climatology 27
3.5.1 Weather persistence 27
3.5.2 Precipitation distribution 29
3.5.3 Extreme precipitation 30
3.6 Dichotomous forecast verification 34

4 Results **36**
4.1 Climatology 36
4.1.1 Weather persistence 37
4.1.2 Precipitation distribution 40
4.1.3 Extreme precipitation 46
4.2 Dichotomous forecast verification 53
4.2.1 AROME, fcfix and fcday forecast verification 53
4.2.2 AROME and fcpersist forecast comparison 61

5 Discussion **64**
5.1 Climatology 64
5.1.1 Dataset properties 64
5.1.2 Seasonal weather duration 65
5.1.3 Variability in AROME precipitation distribution 66
5.1.4 Underestimation of AROME extreme precipitation 69
5.2 Forecast verification 70
5.3 Choosing the precipitation threshold 71
5.4 How trustworthy are observed data? 72

6 Conclusion **74**

Bibliography **75**

A Data **80**

Chapter 1

Introduction

Precipitation forecasts are very important in areas like hydropower companies, agriculture, weather warnings, and society in general, and it is therefore crucial to make them as accurate as possible (Køltzow et al., 2020). Over the past decades, forecast quality has greatly improved, and a 7-day forecast is now roughly as good as a 5-day forecast 20 years ago (Figure 1.1) This is mainly due to better models, more data available, better data assimilation (like 4D-Var), and more computational power (Bauer et al., 2015).

Norway spans about 1700 km from the northernmost to southernmost tip, covering several different climate zones. In addition to its long coastline, Norway's topography is varied, complex and highly irregular, with numerous fjords, mountain ranges, valleys but also some flatter areas. All these factors contribute to large local variations in weather, which poses an extra challenge for weather forecasts (including precipitation) in Norway (Müller et al., 2017).

Norway is also unusually warm compared to other places at the same latitude. The two main reasons for this are the Gulf Stream transporting warm water across the North Atlantic towards the Norwegian Sea and beyond, and low-pressure systems frequently bringing warm and moist air towards the Norwegian mainland. This makes Norway one of the wettest places in Europe, and much wetter than other regions at similar latitudes. (Seager et al., 2002; Villa, 2021).

Figure 1.2 illustrates pretty nicely the differences in annual precipitation amount across Norway. The map on the left uses the current climate normal (1991-2020), while the map on the right shows the relative difference from the old normal (1961-1990). The western coast receives the most precipitation, up to 3000-4000 mm in some places, while

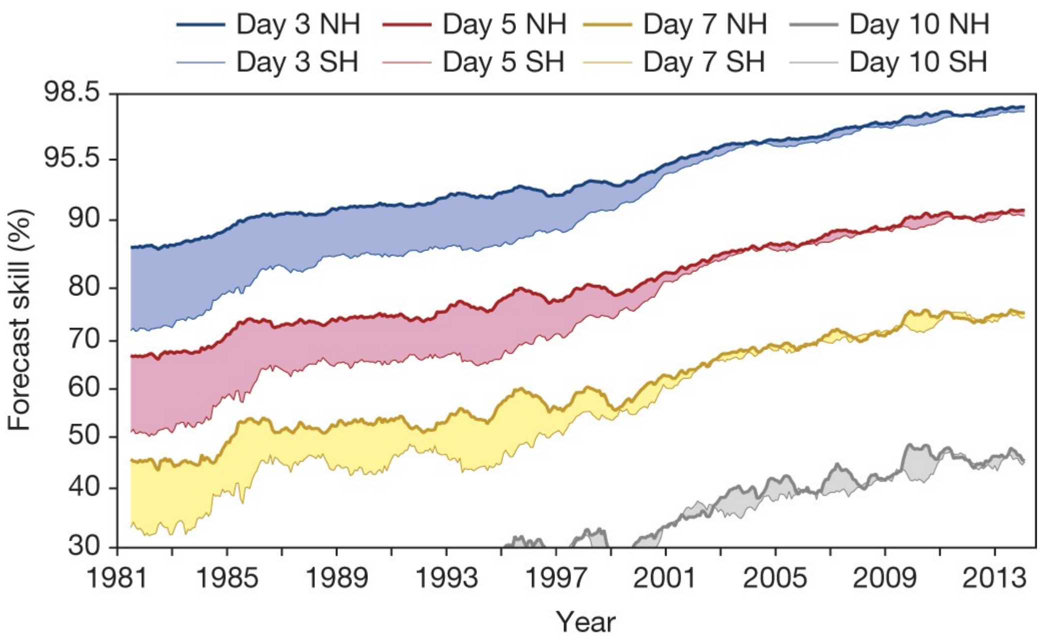


Figure 1.1: Forecast skill improvements (500 hPa height) through history for 3-10 day range forecasts over the extra-tropical northern (NH) and southern hemispheres (SH). From the early 2000s, the use of satellites measurements vastly improved the SH forecast skill (Bauer et al., 2015).

Nordland county is also quite wet. The driest places are the most inland parts of Eastern Norway and Finnmark county. It is clear Norway has gotten wetter overall since the last climate normal, and in fact the annual precipitation the last 100 years has increased by 18% (Hanssen-Bauer et al., 2015).

It is the Norwegian Meteorological Institute (MET Norway) who is responsible for providing weather forecasts in Norway and releasing them to the public. They currently use the regional high-resolution AROME-MetCoOp numerical weather prediction (NWP) model when forecasting within the Nordic region, and the AROME-Arctic model for Svalbard and the Norwegian mainland above the Arctic circle.

Publications on previous AROME-MetCoOp precipitation validations seem to mostly be confined to various MET Norway reports, which are written and published on a quarterly basis. Although their structure are more informational and less analytical than a typical science article, consisting mostly of a short written summary along with lots of figures and tables. Køltzow et al. (2020) did a study on verification of solid precipitation in Norway with focus on wind-induced undercatch (too low precipitation measurements than reality in windy and snowy conditions), and found this had a substantial impact on the verification results.

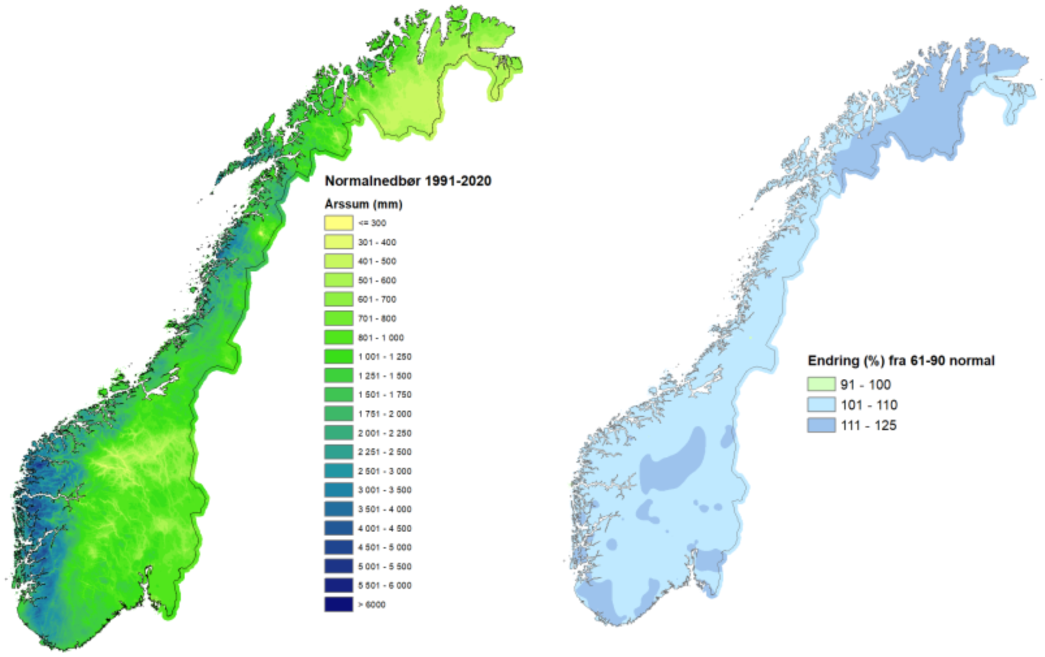


Figure 1.2: Annual precipitation in Norway from 1991-2020 (left) and the relative difference (%) from 1961-1990 (right). Figure from Norwegian Climate Service Center.

The main goal of this thesis is to validate the post-processed AROME-MetCoOp precipitation forecast against observed measurements from weather stations in these six locations: Bergen, Oslo, Trondheim, Tromsø, Kristiansand and Nesbyen. Data is taken from 1. December 2019 - 31. April 2022. The validation process will be divided into two parts, climatology and dichotomous forecast verification results. For climatology, the objectives are to find any anomalies in total precipitation amount (both overall and seasonal), precipitation distribution (how good is the forecast at predicting dry, light rain and heavy rain hours), as well as validating both the forecasted frequency and amount for extreme precipitation events. Verification results look at how well AROME-MetCoOp is able to correctly predict the precipitation at the right time, and this is done with various verification methods, forecast lengths, and accumulation lengths.

The thesis structure going forward is as follows: Chapter 2 contains theory of various forecast verifications methods, along with precipitation types. Chapter 3 lists all the methods used to obtain and process the forecasted and observed precipitation data. Chapter 4 contains the most important climatology and dichotomous forecast verification results. Chapter 5 discusses the main findings from these results (and Appendix results) in greater detail, and Chapter 6 conclude the work with a brief summary of the results at each location, and where the model performed the best overall.

Chapter 2

Theory

This chapter starts with defining what is considered a good forecast, before outlining the various methods of forecast verification. Finally, the main precipitation formation processes are listed.

2.1 Forecast Verification

2.1.1 What is a good forecast?

Before we want to validate a forecast, we need to establish what the desired outcome should be first. The understanding of what counts as a good forecast is not entirely clear, and the answer given depends on who you ask. A forecaster may say the goodness comes from similarities between the forecaster's judgement and the observations, while the user could be more concerned about whether or not the forecast leads to favourable outcomes of their decisions.

Murphy (1993) discussed three types of goodness: Consistency (type 1), quality (type 2) and value (type 3). High forecast *consistency* is achieved when there is correspondence between the forecast and the forecaster's best judgement derived from their knowledge base. As an example, a weather forecaster might intentionally overstate the seriousness of a storm (poor consistency) because they think people would otherwise under-prepare.

Table 2.1: Forecast quality attributes (Murphy, 1993; cawcr, 2015).

Attribute	**Description**
Accuracy	Level of agreement between the forecast and observations
Association	Strength of the linear relationship between the forecasts and observations
Bias	Correspondence between the mean forecast and mean observation
Discrimination 1	Correspondence between the conditional mean forecast and conditioning observation, averaged over all observations
Discrimination 2	Difference between the conditional mean forecast and unconditional mean forecasts, averaged over all observations
Reliability	Average agreement between the forecast values and the observed values
Resolution	Ability of the forecast to sort or resolve the set of events into subsets with different frequency distributions
Sharpness	Tendency of the forecast to predict extreme values
Skill	Relative accuracy of the forecast over some reference forecast
Uncertainty	Variability of the observations. The greater the uncertainty, the more difficult the forecast will tend to be

High forecast *quality* is achieved when the forecast matches the observed conditions at the time it was forecasted, and is probably the most intuitive goodness type. If the predicted storm turns out to be less severe than anticipated, then the forecast has poor quality.

High forecast *value* is achieved when the forecast allows the user to make the best decisions for increased economic, safety and/or other benefits. Although the storm turned out to be less serious, lives and properties were saved because people were more prepared than what they otherwise would have been (good value). A forecast can also be high quality but have little to no value. For instance, a forecast never predicting hurricanes to form would likely get close to 100% accuracy, but in the rare occasions where such an even do happen, the consequences could be devastating and this forecast would be of no value to the public.

In this thesis, it is the forecast quality that is validated, although forecast value is also of importance when choosing the locations. Murphy (1993) describes ten attributes that contribute to the quality of a forecast, listed in Table 2.1.

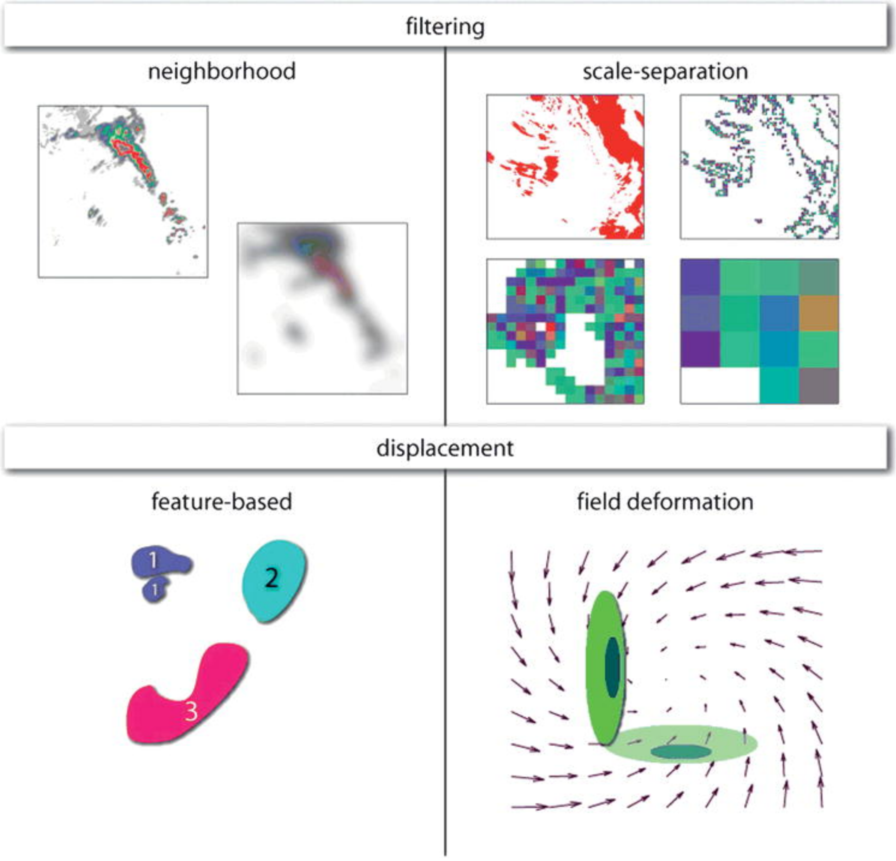


Figure 2.1: Schematic representations of the various spatial verification methods (except point verification) (Gilleland et al., 2009).

2.1.2 Spatial verification methods

When verifying a forecast, an important decision to make is whether the model grid point corresponding to an observation source (weather station, buoy etc.) should be influenced by nearby grid points, and in that case how this should be done.

Point verification is the simplest form for spatial verification, and is the method used in this thesis. Only one model grid point is validated against observations at a time, without considering nearby grid points. It is a well known and fairly straight-forward method to produce forecast verification results, although it is also generally more error-prone and may provide incomplete information about the forecast quality. If a forecast feature is displaced slightly in space but is otherwise correct, it can still yield poor verification results. Moreover, as horizontal model resolution increases over time, the model needs to hit within a smaller range (Gilleland et al., 2009).

Neighbourhood is a filter-based method where a smoothing filter is applied to the forecast field and sometimes also the observed field. The field is upscaled by averaging the values of neighbouring grid points within a certain radius of each other. The result is a smoothed version of the original field. Most filters than can be applied preserve peak values, which is important in capturing extreme event features.

Scale separation/decomposition is another filter-based method where each field is decomposed using some type of spatial bandpass filter, e.g. Fourier transforms. By isolating the various forecast features by scale (like large-scale frontal systems or small-scale convective showers), it is now easier to find which scale the main sources of error originate from, and assess the capability of the forecast to reproduce the observed scale structure in the observations.

Features-based approach identifies individual structures/features within a field and analyse these structures separately, finds the best matches of features across these fields, and compares these matched features based on various attributes (spatial displacement, orientation, size, average intensity, etc.)

Field deformation verification attempts to manipulate the forecast field to resemble the observed field in the best possible way (for instance by minimizing the accuracy or bias score difference). The resulting product is a vector field that describes which adjustments were made, that are then evaluated either diagnostically (why the field looks like it does) or analytically (Gilleland et al., 2009, 2010).

Schematics of these spatial verification methods can be seen in Figure 2.1.

2.1.3 Deterministic forecast verification

A forecast is said to be deterministic when there is only one possible solution, e.g. it will precipitate 2 mm tomorrow. If the forecast contains continuous variables (can be any value within a physically realistic range), the goal is to measure how the forecast values differ from the values of the observations.

Two common verification metrics in this category are Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

$$MAE = \frac{1}{N} \sum_{k=1}^{N} |F_k - O_k| \tag{2.1}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(F_k - O_k)^2} \tag{2.2}$$

F_k and O_k are forecast and observations, respectively, of a given pair number k with N total values. Their score ranges from 0 to ∞, with a perfect score of 0 (no deviation for any k). Both seek to find the average magnitude of the forecast errors, although neither of them indicate the direction of the deviations. Since the values are squared for RMSE, it is more sensitive to major errors than MAE, which can be very useful if large errors are particularly undesirable. RMSE will always be equal to or larger than MAE, and the relative difference can say something about if the forecast is dominated by small or larger errors.

2.1.4 Dichotomous forecast verification

A deterministic forecast can be reduced to a binary (dichotomous) forecast where there are only two possible outcomes for an event: Yes, it did happen, or no, it did not happen. The boundary between these is determined with a specific threshold value, for instance all hourly forecast and observed values above 0.1 mm/h counts as rain events, while those under it are considered no rain events. With two datasets (forecast and observation) as well as two outcomes (yes and no), we can set up a 2 x 2 contingency table which shows the occurrence of all possible outcomes, seen in Table 2.2.

Table 2.2: A 2 x 2 Contingency table.

		Observed	
		Yes	*No*
Forecast	*Yes*	Hit	False alarm
	No	Miss	Correct negative

This contingency table acts as a basis for several different verification scores. While each have their own strengths and weaknesses, they are able to give a quite coherent forecast verdict when combined. The ones presented below will be part of the dichotomous forecast verification results later on.

Accuracy is the overall fraction of forecasts that were correct. Score ranges from 0 to 1, with 1 as the perfect result. It is a simple measure of forecast quality, although it can be misleading in situations with very rare events, where correct negatives would completely dominate the contingency table outcome.

$$Accuracy = \frac{Hits + Correct\ negatives}{Total} \tag{2.3}$$

Bias frequency says something about the forecast frequency of "yes" events compared to the observed frequency of "yes" events. Score ranges from 0 to ∞, with 1 as the perfect result. This metric indicates whether the forecast has a tendency to overforecast (score > 1) or underforecast (score < 1) events, e.g. if the number of forecasted rain hours is higher or lower than the total observed rain hours. Bias does not measure how well the forecast events correlate with the observation events, only relative frequencies.

$$Bias\ frequency = \frac{Hits + False\ alarms}{Hits + Misses} \tag{2.4}$$

Hit rate gives the fraction of observed "yes" events that were correctly forecast. Score ranges from 0 to 1, with 1 as the perfect result. It is sensitive to hits, but ignore false alarms, and thus is best applied together with the false alarm ratio below. It is also quite good for rare events, since only observed yes events are considered.

$$Hit\ rate = \frac{Hits}{Hits + Misses} \tag{2.5}$$

False alarm ratio provides the fraction of observed "yes" events that actually did not occur. Score ranges from 0 to 1, with 0 as the perfect result. It is sensitive to false alarms, but ignore misses.

$$False\ alarm\ ratio = \frac{False\ alarms}{Hits + False\ alarms} \tag{2.6}$$

Success ratio represents the fraction of forecasted "yes" events that were correctly observed. Score ranges from 0 to 1, with 1 as the perfect result. Success ratio is also equal to 1 - False alarm ratio.

$$Success\ ratio = \frac{Hits}{Hits + False\ alarms} \tag{2.7}$$

2.1.5 Probabilistic forecast verification

The opposite of deterministic is a probabilistic forecast, which expresses the chance of an even occurring, e.g. there is a 40% chance it will precipitate at least 2 mm tomorrow. The effect of this can be two-fold: On one hand, a probabilistic forecast allows the forecaster to better express the inherent uncertainly all weather forecasts have. On the other hand, it is only probabilistic up until the time of the observation, which (in these settings) are always deterministic. It also requires longer periods of data to verify these forecast to make sure if the forecasted probabilities turned out right.

Brier score (BS) is the probabilistic equivalent of mean squared error, and seeks to find the magnitude of the probability forecast errors. It is therefore defined as

$$BS = \frac{1}{N}\sum_{k=1}^{N}(f^k - o^k)^2 \tag{2.8}$$

where f^k is the forecast probability, while o^k is observed. k is the number of the N total forecast event pairs. Score ranges from 0 to 1, where a perfect score is 0.

In our case, the post-processed AROME-MetCoOp forecast is deterministic, which means f^k and o^k can only either be 0 or 1. If both are 0 (correct negative), the bracket number is 0 for that pair number k, and the same is true if both are 1 (hit). This means the only non-zero options left are miss and false alarm, where the value inside the bracket equals one for that pair number. With this information, we can derive a simplified BS for dichotomous forecasts.

$$BS = \frac{Miss + False\ alarm}{Total} \tag{2.9}$$

Brier Skill Score (BSS) evaluates the relative skill of the forecast over that of climatology (BS_{ref}), in terms of predicting whether or not an event occurred, and is defined as

$$BSS = \frac{BS - BS_{ref}}{0 - BS_{ref}} = 1 - \frac{BS}{BS_{ref}} \tag{2.10}$$

Score ranges from $-\infty$ to 1, where a perfect score is 1. BSS = 0 indicates a forecast with no skill over the reference/climatology forecast (BS = BS_{ref}), and a negative BSS would tell it is better to trust the climatology than looking at the forecast. Climatology could for instance be the average observed temperature on a given day over the last 30 years (continuous), or the probability of rainfall on a given day based on observations over the last 30 years (probabilistic) (cawcr, 2015).

2.2 Precipitation formation

Precipitation formation can be divided into three categories: Frontal precipitation, convective precipitation and orographic precipitation. These are not mutually exclusive and may occur at the same time.

2.2.1 Frontal precipitation

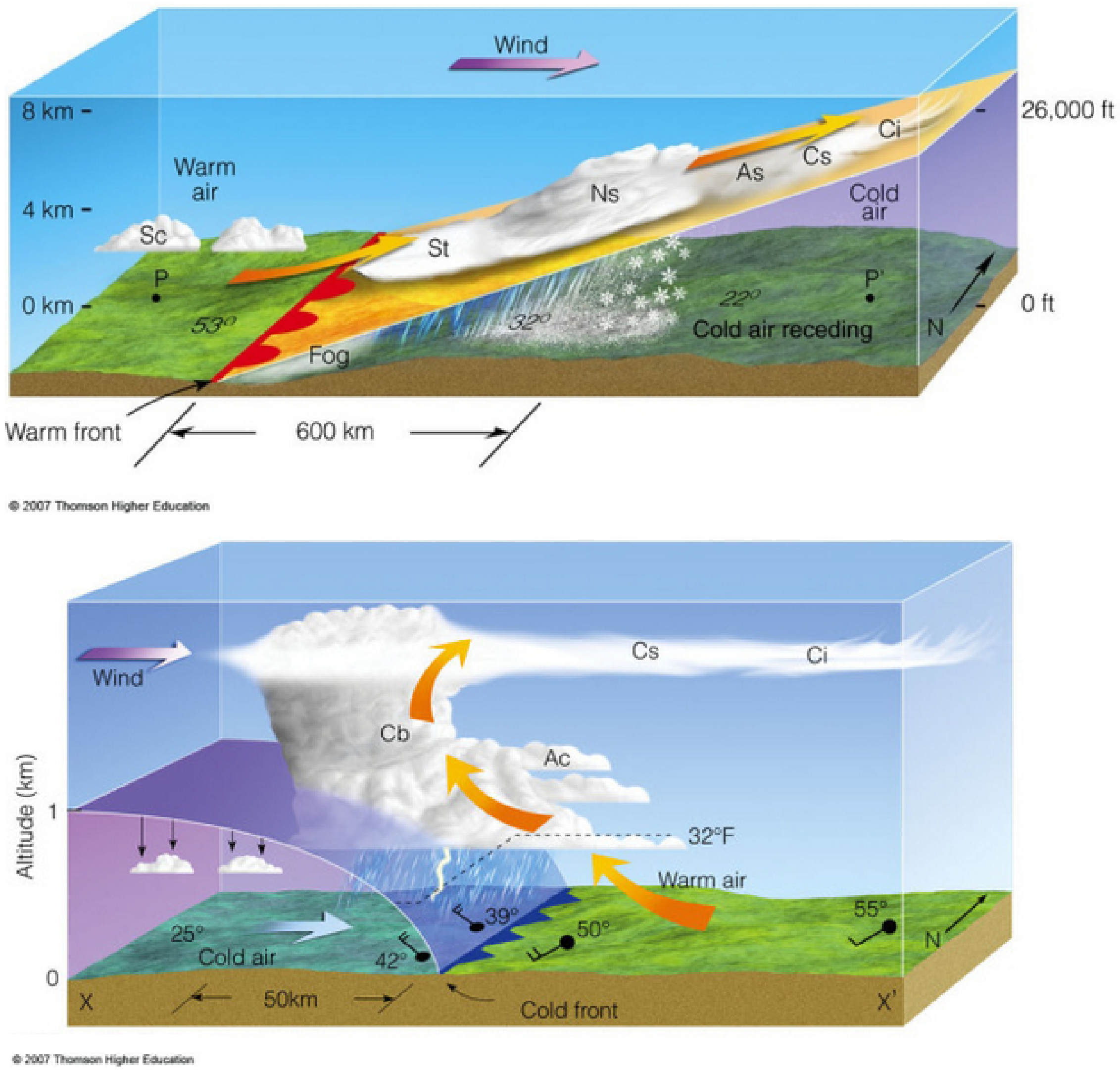


Figure 2.2: Illustrations of a warm front (top) and cold front (bottom) (Ahrens, 2014).

A front is the boundary between two air masses of different densities (temperature and humidity), and is associated with a moving cyclone (low pressure system). In the front of the cyclone, a warm front causes advancing warm and moist air to slowly rise above the retreating cold and dry air, since warm air is lighter than cold air. This causes relatively moderate and uniform precipitation in quite a large area ahead of the front.

Further back in the cyclone, there is a cold front advancing much quicker, where the cold air undercuts and displaces the warm air ahead. The resulting sharp inclination (slope) at the front usually leads to more intense precipitation, or even thunderstorms in extreme cases. If the cold front catches up to the warm front (which often happens), the end result is an occluded front (Rafferty, 2012). Figure 2.2 illustrates both a warm front (top) and cold front (bottom). Frontal precipitation is the most common precipitation type at our latitudes.

2.2.2 Convective precipitation

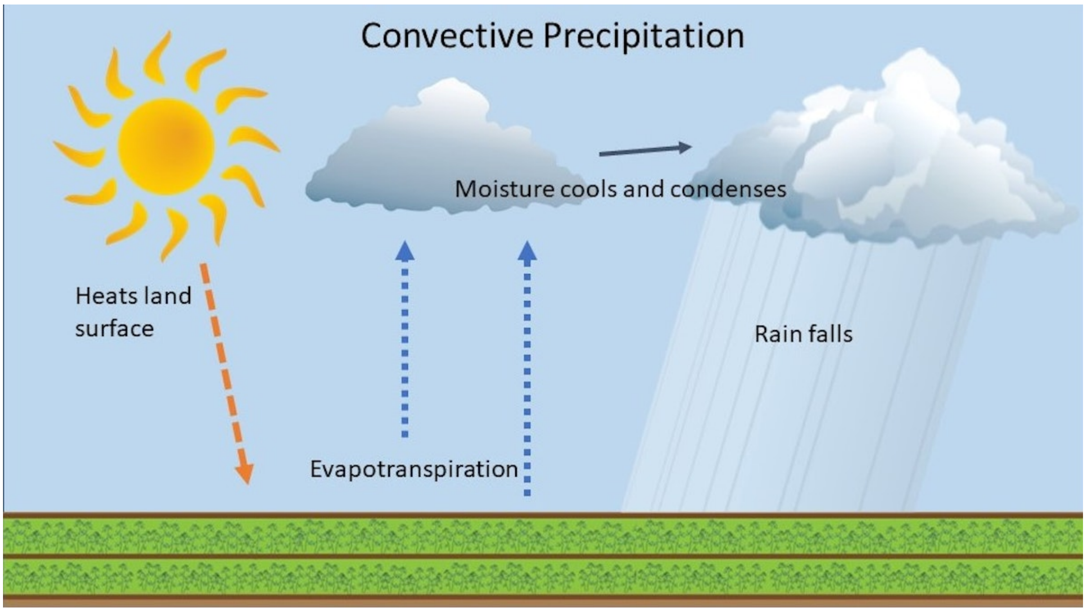


Figure 2.3: Illustration of convective precipitation (Stuart-Haëntjens, 2018).

Convective precipitation occurs when the surface is heated, causing a shallow layer of air above to warm up and rise (convection) due to higher buoyancy (Figure 2.3). If the heated air is moist and the vertical transport is strong enough, it will gradually cool down due to lower pressure, and the water vapour will condense, form clouds and eventually precipitate (Ahrens, 2014). These rain shower can be quite local and very intense, but usually dissolve after a few hours. In Norway, convective precipitation is most common during hot summer days in the afternoon, after the sun has heated up the ground for several hours. These conditions are most favourable in the eastern part of Norway. During winter, convective showers may also happen when very cold air is advected over relatively warm water and heated up.

2.2.3 Orographic precipitation

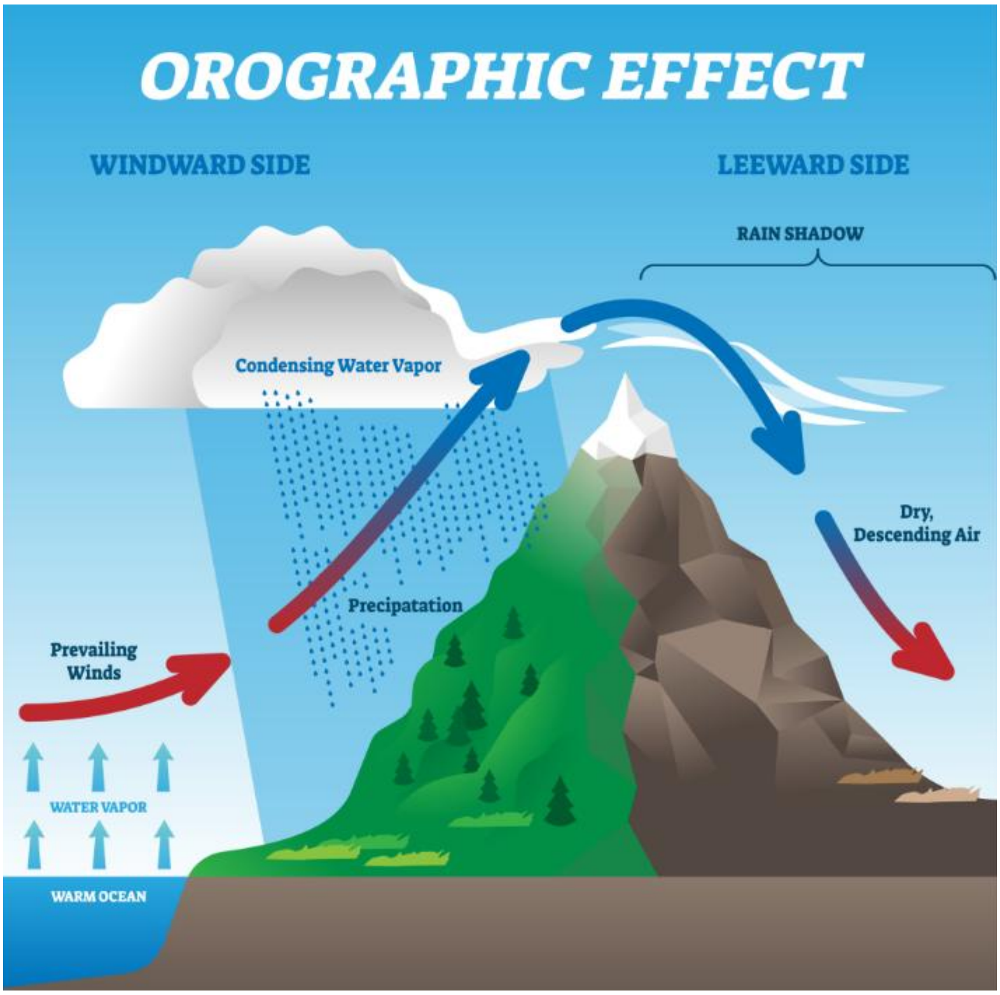


Figure 2.4: Illustration of orographic precipitation (Sahay, 2022).

Orographic precipitation happens when moist air is forced upwards by mountains, causing the air to cool and eventually form clouds (Figure 2.4). Precipitation usually falls on the windward side, although for lower mountain ranges it is possible for some precipitation to be carried over to the leeward side. With little moisture left, the air mass reaching the leeward side is very dry, causing it to heat up faster when descending than the rate it cooled on the way up. The leeward side is often called the rain shadow due to experiencing vastly less precipitation than the windward side (Ahrens, 2014). Orographic precipitation tends to occur during a low pressure and then acts as orographic enhancement of front precipitation. These conditions are very common in Western Norway, and is one of the main reasons why the precipitation map in Figure 1.2 looks the way it does.

Chapter 3

Methods

This chapter describes some general properties of a weather station and the various instruments used in precipitation measuring, the AROME-Met-CoOp numeric weather prediction model, how, where and which data from these sources were obtained, and how they were processed to output the results in this thesis. All data were downloaded and processed in MATLAB® R2021b.

For the record, the terms precipitation, rain, rainfall and similar will be written interchangeably for the sake of variety, although in all cases these always refer to precipitation as a whole, all types included.

3.1 Observations

3.1.1 Measuring precipitation at weather stations

There are roughly 200 professional weather stations in Norway, largely owned and operated by MET Norway (Nipen et al., 2020). While manual weather stations that required on-site personnel to operate and send in data were the norm in the past, more and more Automatic Weather Stations (AWS) have replaced them in later years. These are measuring precipitation every hour (and sometimes every minute), in contrary to manual stations where precipitation is usually measured 1-4 times per day (snl, 2020).

The use of private weather stations have gotten increasingly more attention lately, as they vastly outnumber MET Norway's own stations by a factor of about 50. Some of these stations are produced by Netatmo, a french company specialising in smart home devices, and where local weather data recorded from each individual station can be shared online and pooled together in their Weathermap service (MET Norway, 2018). Even though their data quality are highly variable due to not following the World Meteorological Organization (WMO) standards for weather observations (WMO, 2018) and lack of information about how they are maintained, their inclusion in weather forecasting have shown to improve overall forecast quality, especially for short-term temperature forecasts (Nipen et al., 2020). Bárdossy et al. (2021) found that using private weather stations (with sufficient quality control) could improve temporal precipitation interpolation.

Pluviometres (commonly referred to as rain gauges) are the standard way of measuring precipitation intensity, which is the amount of precipitation measured over a certain period. For manual stations, the rain content is poured into a graduated cylinder and measured to give the total rainfall (in case of snow, the content is first melted then measured). For automatic stations, several different methods are currently used. One of them is weighing rain gauges, where the weight of the collected water is measured (typically by a weight cell or a string of vibrating wire) as a function of time and converted into rainfall depth. (AMS, 2012; snl, 2017).

A low-cost but more inaccurate alternative is tipping bucket rain gauges, where a funnel collects and channels the precipitation into a seesaw-like container. Once a pre-defined amount of precipitation falls, the lever tips, emptying the container in the process, and an electric signal is sent indicating a rainfall increase. However, these gauges are less accurate than weighing rain gauges, since rainfall may stop before the lever has tipped, and the stored water would then give the next rain period a "head start" by only requiring a tiny amount of rain to tip the lever and thus indicate more rain has fallen than the actual amount. During intense rain, some moisture may be lost between the time it takes from the lever to tip until a new bucket is ready to collect rain. They also struggle with measuring snowfall as it may just cover the funnel like a blanket, and installing a heater to melt the snow will lose too much moisture to evaporation for the measurement to be accurate (Groisman and Legates, 1994).

Disdrometres and hotplates precipitation gauges are more modern instruments which uses other techniques than collecting rainfall. Disdrometres measure drop size distribution and falling velocity of precipitation, which can be used to estimate kinetic energy of raindrops and thus their potential effect on soil erosion and pollution in surface water flows. That said, measurements tend to be more error-prone during heavy rainfall and when measuring very large droplets, as well as in windy conditions (ARM; Kathiravelu et al., 2016).

Hotplates consist of two thermally isolated aluminium plates (one facing upwards and one facing downwards) which are heated up with electricity to about $75°C$. Precipitation rate is estimated by calculating the power required to either melt snow, evaporate snow, or evaporate rain on the upward-facing plate, compensated for wind effects by subtracting out the power on the lower, downward-facing plate. Hotplates provide wind speed, temperature and precipitation intensity data every minute, making them ideal for real-time applications like aircraft de-icing and road weather conditions. They are also considered low-maintenance (albeit quite power hungry) and very accurate even in windy and snowy conditions. Although, they struggle a bit more in conditions with hail and graupel as the ice pellets tend to bounce off the plates before they can be melted (Rasmussen et al., 2011).

Images of various pluviometre types can be seen in Figure 3.1.

3.1.2 Obtaining observational data

Naturally, it is not feasible to perform a full-scale point verification for all Norwegian weather stations, meaning only the stations deemed the most relevant and impactful were picked based on a few criteria. First, places assumed to give the highest forecast *value* were looked at, i.e. the biggest cities in the country. Next, places with different geographic and climatic features (coastal, inland, mountain etc.) were desired, since some of the major cities like Oslo/Drammen, or Stavanger/Sandnes are very close to each other and experience mostly the same weather. And lastly, if there were more than one AWS within these regions, the one with the most complete time series was chosen so that the validation results were as accurate as possible.

As such, we chose six AWS based on these criteria, where all of them having weighing rain gauges installed. Figure 3.2 shows where these stations are located in Norway, and Table 3.1 lists their location data, how long they have been operating, and missing data (if any). Hourly observed precipitation data for the analysis period (1. December 2019 to 31. April 2022, or 882 days/21168 hours in total) were downloaded at each station from `https://seklima.met.no/`. Resolution of observed precipitation data was 0.1 $kg\ m^{-2}$ (equivalent to millimetre rain).

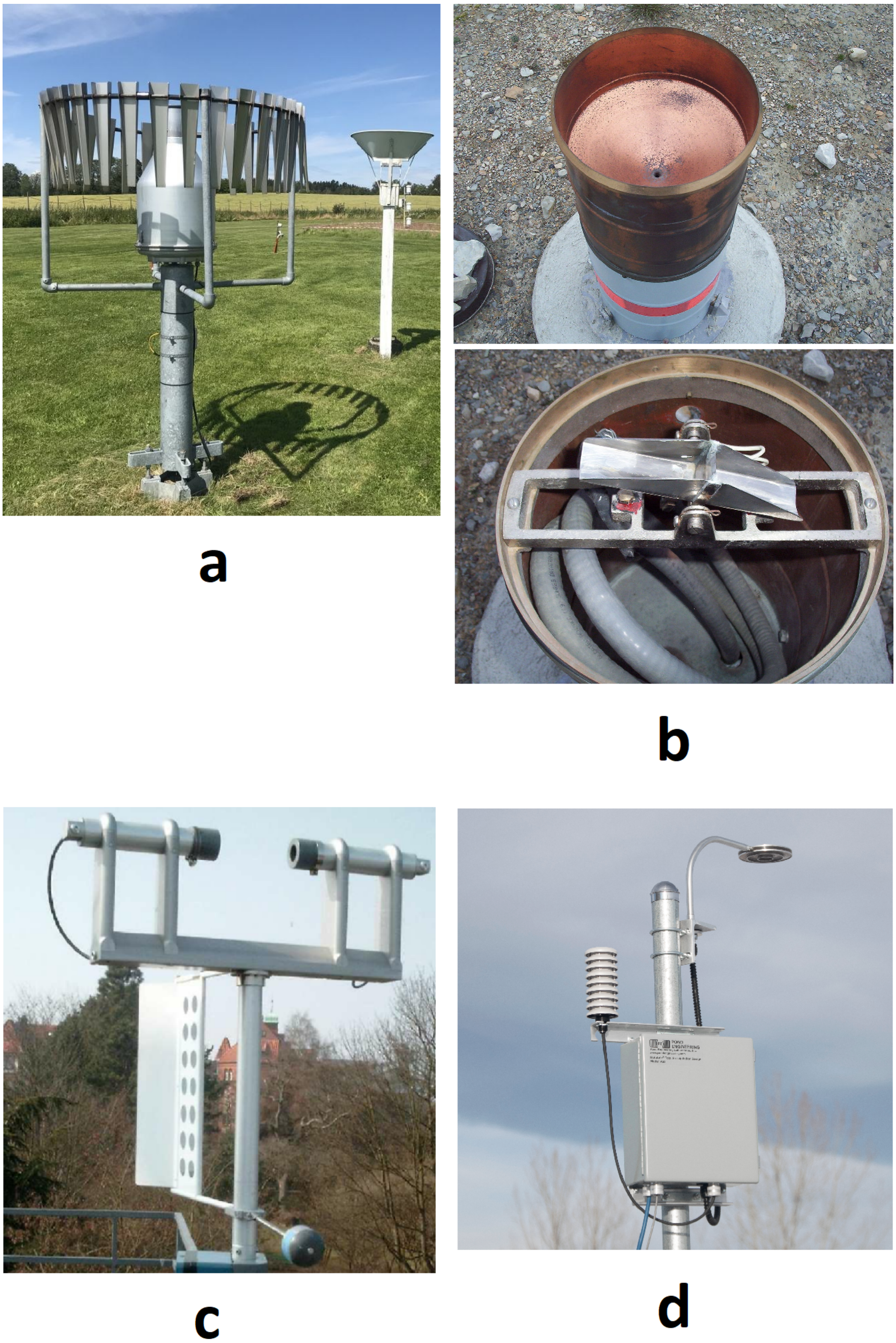
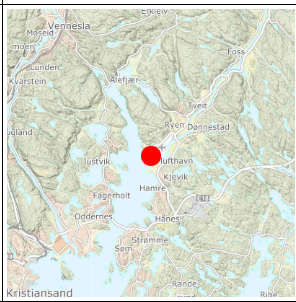
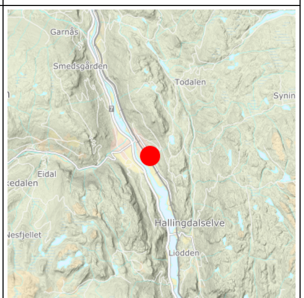


Figure 3.1: Images of different pluviometre types: a) Weighing rain gauge, b) Tipping bucket rain gauge (upper - exterior, lower - interior), c) Optical disdrometre, and d) Hotplate precipitation gauge.

Figure 3.2: Location of all six weather stations used.

Table 3.1: Local topography maps and meta data for the chosen weather stations.

	Bergen (Florida)	Oslo (Blindern)	Trondheim (Lade)
			
Model latitude	60.3824° N	59.9414° N	63.4465° N
Model longitude	5.3391° E	10.7216° E	10.4409° E
Model altitude	9 m	93 m	25 m
Station latitude	60.3830° N	59.9423° N	63.4428° N
Station longitude	5.3327° E	10.7200° E	10.4428° E
Station altitude	12 m	94 m	13 m
Operational since	1949	1931	2004
Missing observation values	0 of 21168 (0%)	0 of 21168 (0%)	461 of 21168 (2.18%) over 5 periods

	Tromsø (Vervarslinga)	Kristiansand (Kjevik)	Nesbyen (Todokk)
			
Model latitude	69.6566° N	58.2009° N	60.5640° N
Model longitude	18.9456° E	8.0731° E	9.1332° E
Model altitude	77 m	9 m	173 m
Station latitude	69.6537° N	58.2000° N	60.5670° N
Station longitude	18.9368° E	8.0767° E	9.1323° E
Station altitude	100 m	12 m	166 m
Operational since	1895	1939	2003
Missing observation values	7 of 21168 (0.03%) over 1 period	592 of 21168 (2.80%) over 3 periods	0 of 21168 (0%)

3.2 AROME-MetCoOp

3.2.1 Introduction to AROME

AROME-MetCoOp is a high-resolution NWP model operated by Meteorological Cooperation on Operational Numerical Weather Prediction (MetCoOp), and is based on Applications of Research to Operations at Mesoscale (AROME-France) model by Météo-France (Seity et al., 2011; Müller et al., 2017). MetCoOp started in 2010 as a collaborative effort between the Norwegian Meteorological Institute (MET Norway) and Swedish Meteorological and Hydrological Institute (SMHI), and since when AROME-MetCoOp became operational in 2014, the departments have benefited from operating, developing and monitoring the same weather model. Finnish Meteorological Institute (FMI) joined MetCoOp in 2017, and the Baltic countries Estonia, Lithuania and Latvia is set to join in 2022 (Kristiansen and Blaauboer, 2018; MET Norway).

AROME-MetCoOp domain covers the Nordic countries as well as the North Sea and Baltic Sea (see Figure 3.3) with a 2.5 x 2.5 km horizontal resolution and 65 vertical levels, where the vertical resolution decreases by height up until the vertical boundary layer located at roughly 33 km altitude. The model is forced by the lateral and upper boundary conditions of the large-scale European Center for Medium-Range Weather Forecast (ECMWF) model (Müller et al., 2017).

To improve scenarios less suited for deterministic forecasts, like resolving the stochastic nature of rapidly-growing convective cells (Müller et al., 2017), an ensemble version of AROME-MetCoOp named MEPS (MetCoOp Ensemble Prediction System, operative since 2016) is currently used. As of 2022, it contains 30 ensemble members; one control member (AROME-MetCoOp with unperturbed initial and boundary conditions), while the rest are perturbed members. The control member is updated every 3 hours, with a 66-hour forecast produced every main cycle (00, 06, 12, 18 UTC). The 3-hour intermediate forecasts (03, 09, 15, 21 UTC) are used for data assimilation for the following main cycle (Frogner et al., 2019; Homleid et al., 2021).

The raw forecasts are then post-processed (also known as the MET Nordic dataset), which takes a limited number of surface variables from MEPS and perform bias corrections based on real-time observations from Netatmo stations, WMO-stations from MET and FMI, non-WMO stations in Norway, and radar. Unlike MEPS, the output product is deterministic (only one solution), and it is also downsized to 1 km horizontal grid spacing.

MET Nordic forms the basis of operational forecasts delivered by MET Norway, like `https://www.yr.no` (MET Norway NWP Wiki).

The cloud microphysics is based on the Kessler scheme for warm (liquid) processes, and the three-class ice parametrization (ICE3) scheme for cold processes. ICE3 includes cloud ice, snow and graupel, and more than 25 processes are parametrized by the scheme (Pinty and Jabouille, 1998). AROME-MetCoOp contains some modifications to ICE3, mainly to account for its weaknesses during winter season, e.g. T_{2m} being too low due to low-level clouds decaying too quickly in cold conditions (Müller et al., 2017).

For simplicity, AROME-MetCoOp and by its extension MEPS will just be denoted as AROME going forward.

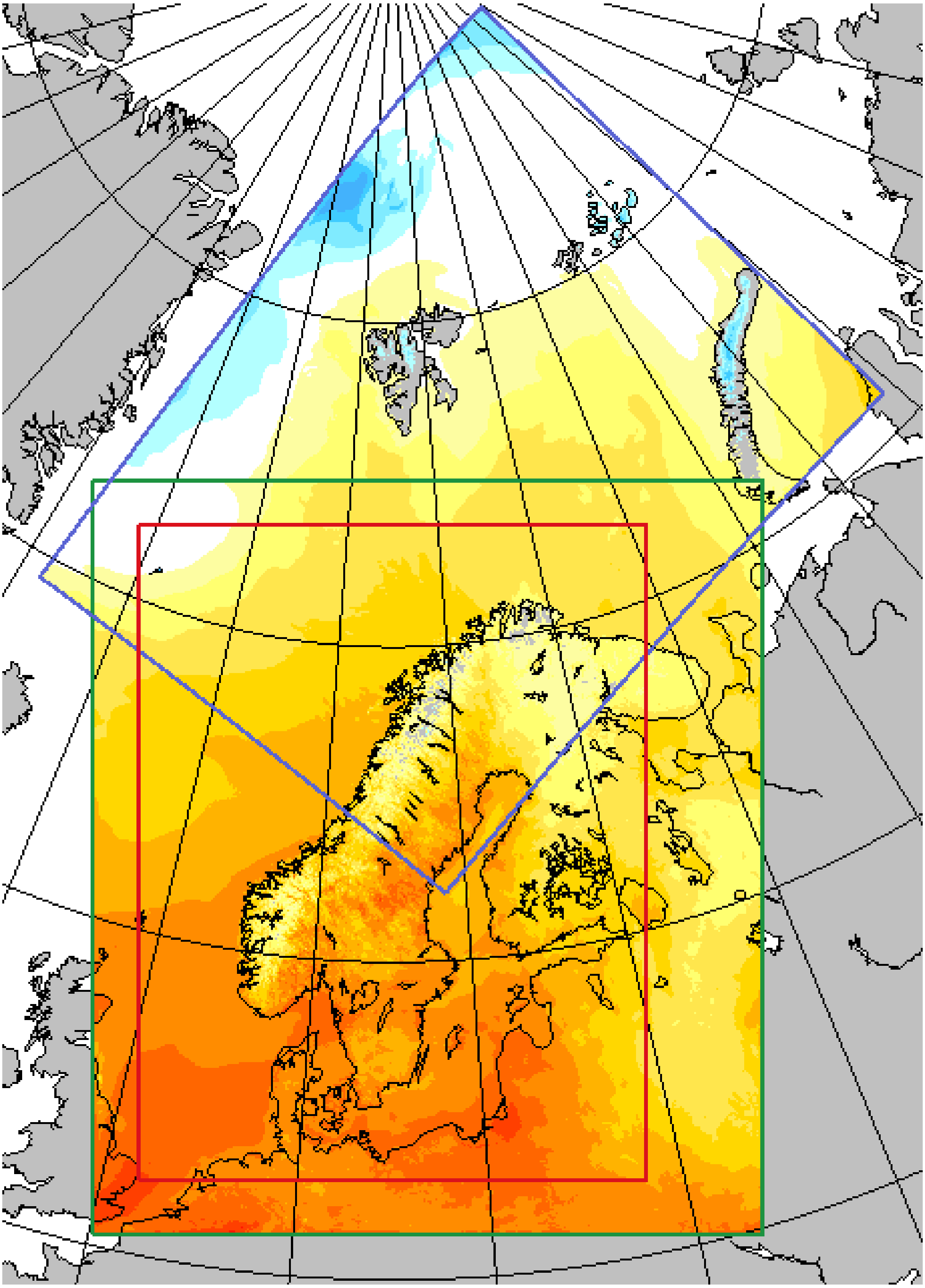


Figure 3.3: Model domain of MEPS (green square), post-processed (red square) and AROME-Arctic (blue square) (MET Norway NWP Wiki).

Table 3.2: Forecast hours and accumulated precipitation lengths used in the verification results. For example, forecast hour +4 from a 00:00 UTC run denotes the accumulated precipitation predicted from 03:00 until 04:00.

Accumulation length	**Forecast hours**
1 hour	+4h, +13h, +48h
6 hours	+4h-9h, +13h-18h, +25h-30h
24 hours	+4h-27h, +13h-36h, +25h-48h
48 hours	+1h-48h

3.2.2 Obtaining post-processed AROME forecast data

Post-processed hourly forecast precipitation data (from MEPS version **cy43**) were obtained from `https://thredds.met.no/thredds/catalog/metpparchive/catalog.html`, a public site ran by MET Norway with readily available data as NetCDF files. Then, we chose the closest available model grid point to each station's location. Resolution of forecast precipitation data was at least 0.0000001 $kg\ m^{-2}$, or 1 000 000 times higher than for observed data, which is way more than needed for all practical purposes.

While the main cycle of the AROME model is ran every 6 hours, only forecasts ran at 00 and 12 UTC were used here, reducing the size of downloaded forecast data by half. This resulted in 1764 unique forecasts considering an 882 days long analysis period (1. December 2019 to 31. April 2022). For climatology results, the first 12 forecast hours were used to ensure no overlapping forecast periods when comparing with observed data, while Table 3.2 shows the various forecast hours and accumulation lengths chosen for the dichotomous forecast verification. Forecast hours ranging from +49h to +66h often contained missing data, which was the main reason these times were discarded. Nevertheless, picking different accumulation lengths taken at different forecast hours is a key part to determine if the model quality changes with forecast length, and if verification results for individual precipitation hours differs from accumulated rainfall over several hours.

3.3 Processing the data

3.3.1 Missing values

Both forecasted and observed datasets were sometimes incomplete, and these missing values were treated as not-a-number (NaN) to make sure there were no "holes" in the datasets and all precipitation values appeared at their correct time stamps. NaN-values in observed data were displayed in Table 3.1, whereas 19 of the 1764 forecasts were missing from the MET database and thus filled with NaN. This was true for all six locations.

NaN values in general need to be dealt with properly to avoid flawed results. In this case, a "zero-tolerance policy" was conducted to be on the safe side: If any section of a dataset contained at least 1 NaN value, then the whole section was considered as NaN and ignored from the calculations. A section could for instance refer to daily precipitation in climatology (all 24 hourly values must be valid for that day to be accepted), or 48h accumulated precipitation in verification results. The exception was for monthly precipitation where NaN values were set to 0 instead to avoid entire months of data being overwritten and lost. While this resulted in some months losing rainfall to NaNs and underestimating the "real" precipitation amount, this applied both to forecasted and observed data, meaning the relative difference between the two were not affected that much.

In addition, all neighbouring observation values part of any given 12-hour cycle (01:00 to 12:00, and 13:00 to 00:00) that contained at least one NaN value were replaced with NaN. For example, the missing Tromsø data lasted from 09.02.2021 15:00 to 09.02.2021 21:00. Following the rule above, all data from 13:00 including 00:00 the next day were filled with NaN. This way, observed data can be considered as 1764 unique 12-hour sets where each set solely contains either real or NaN values, which was identical to how forecast data was structured except they contained 48-hour sets instead of 12-hour. It is not strictly necessary to do it like this, but having both datasets on compatible formats makes them easier to keep track of and compute the results later on. The downside is some more data were lost, though it can be argued the relative increase in NaN values was small enough to not have any noticeable effect.

- **Tromsø:** +5 NaN values, 12 in total
- **Kristiansand:** +31 NaN values, 492 in total
- **Trondheim:** +56 NaN values, 648 in total

- Bergen, Oslo and Nesbyen are unaffected since they did not contain any NaN values

Lastly, NaN values occurring at different times in forecast and observed data were overlapped. For instance, if forecast data contained NaNs at 3. February 2021 while observed data contained NaNs at 21. March 2020, both datasets would now contain NaN values at both dates. This is important both for climatology results (same number of data points when comparing rainfall amount) and forecast verification (guarantee that real values from one dataset are never compared against NaN values from the other).

3.3.2 Precipitation thresholds

In cases where we only evaluated whether it precipitated or not (binary outcome), all values ≥ 0.1 mm/hour (mm/h for short) or ≥ 0.1 mm/day were treated as precipitation, and otherwise as no precipitation. Since forecast data resolution is way higher than observed data, it is possible the accumulated forecast rainfall for a given duration indicates precipitation, even though none of the individual hours gathered enough rain to pass the threshold. Forecast values were only ever rounded down if hourly/daily precipitation was still below 0.1 mm (treated as no rain for statistical purposes), and only in situations where total precipitation amount was not being considered.

3.4 Custom-made forecasts

Part of the task in this thesis is to evaluate how well AROME performs against simple custom-made forecast, both originating from AROME data itself and recent observations. Three such forecasts were created:

- **fcfix:** Bias-corrected AROME forecast by taking the relative difference in forecasted and observed mean value of all data points throughout the analysis period (1. December 2019 - 31. April 2022). For example, if total forecasted precipitation was 20% lower than total observed precipitation, then each hourly forecast value was multiplied by a factor of $1/(1-0.2) = 1.25$. Used in climatology, extreme precipitation, and dichotomous forecast verification.

- **fcday:** Use yesterday's hourly observations to forecast today's weather for that respective hour of the day. For example, if it rained 0.7 mm yesterday from 12:00

to 13:00, then fcday will predict 0.7 mm precipitation today from 12:00 to 13:00. Only used in dichotomous forecast verification.

- **fcpersist:** Use last hour's observation and forecast that the weather will remain unchanged the next 12 hours. Updated twice a day (two 12-hour periods). For example, if it precipitated 1.6 mm from 23:00 to 00:00, fcpersist will forecast 1.6 mm rainfall every hour for the next 12 hours. Only used in dichotomous forecast verification.

These forecasts served different purposes depending on how they were created. fcday and fcpersist were made using available observations at the time, and acted as low-quality benchmark forecasts that AROME is expected to outperform overall. fcpersist was also used to examine how persistent the weather is at each location, and if the nowcasting (first few hours into the forecast) of fcpersist can match or even outperform AROME. Fcfix on the other hand is technically an improvement over AROME by removing the bias of forecasting too little/too much total precipitation. That said, the main goal here is not to create a better forecast than AROME, but to use fcfix as a tool to check if AROME suffers from any mean precipitation bias.

3.5 Climatology

This section outline the methods of deriving a location's climatology from forecasted and observed data over the analysis period.

3.5.1 Weather persistence

This part was based purely on observed data, hence the step of overlapping NaN values between forecast and observed data was not performed, as it would have lost some data for no good reason.

Weather persistence says something about how often the weather changes at a given location, i.e. weather stability. One way to approach this is to look at average rain weather duration and average dry weather duration. To find these, we need two variables: *total rain/dry hours*, and *total rain/dry weather periods*. The former was found simply by evaluating all hourly observed precipitation values against the ≥ 0.1 mm/h threshold value, and count the number of occurrences for each outcome. The latter was found when counting the amount of times it changed between rain and no rain between subsequent hours/days, marking the end of a rain/dry weather period and the start of a new. A rough schematic of this process can be seen in Figure 3.4. In this example, the average rain weather duration would be 3.5 hours (7 rain hours divided by 2 rain periods). Average rain/dry weather duration was calculated for hourly and daily precipitation data, both with the same method of procedure.

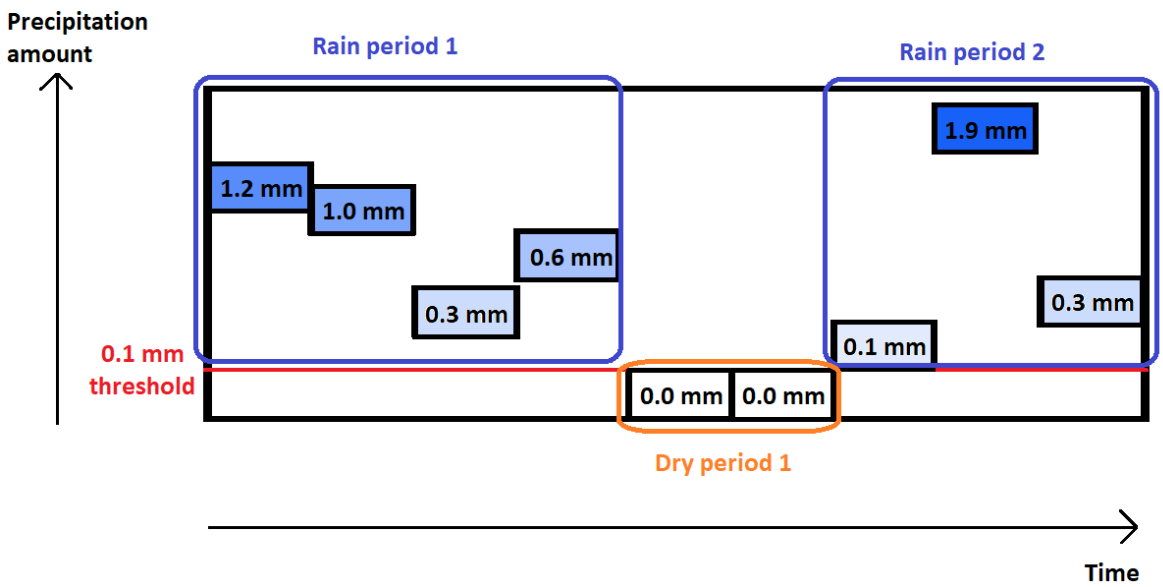


Figure 3.4: Schematic of how rain/dry weather periods were counted.

In some instances, periods may start or end rather abruptly due to limitations in the dataset. The first real value after an arbitrary long string of NaN values was always considered to be the start of a new period. This was done because the number of changes between dry/rainy weather (if any) is unknown to the observer, and to protect against NaN periods that may last for weeks from erroneously increasing the perceived average rain/dry weather duration. In our updated schematic (Figure 3.5), there is now an extra rain period and one less rain hour. Notice how the average rain weather duration has decreased from 3.5 hours to 2 hours just by substituting a rain hour with NaN.

When dividing the data into seasons, the first value for each season was also treated as a new period. E.g. even if it precipitated on both 31. August 2020 and 1. September 2020, they were considered as two separate precipitation periods since they belonged in two different seasons. One rather obvious downside were the calculated total rain/dry weather periods being slightly higher than "reality", with the most "split-up" scenarios like daily precipitation divided by seasons being impacted the most (potentially generating an extra period once every 90 data points or so). By adding up total rain rain/dry weather periods for each season and compare that number to total periods without any seasonal division, the increase came out to be roughly 2-5% for daily data, and >0.4% for hourly data.

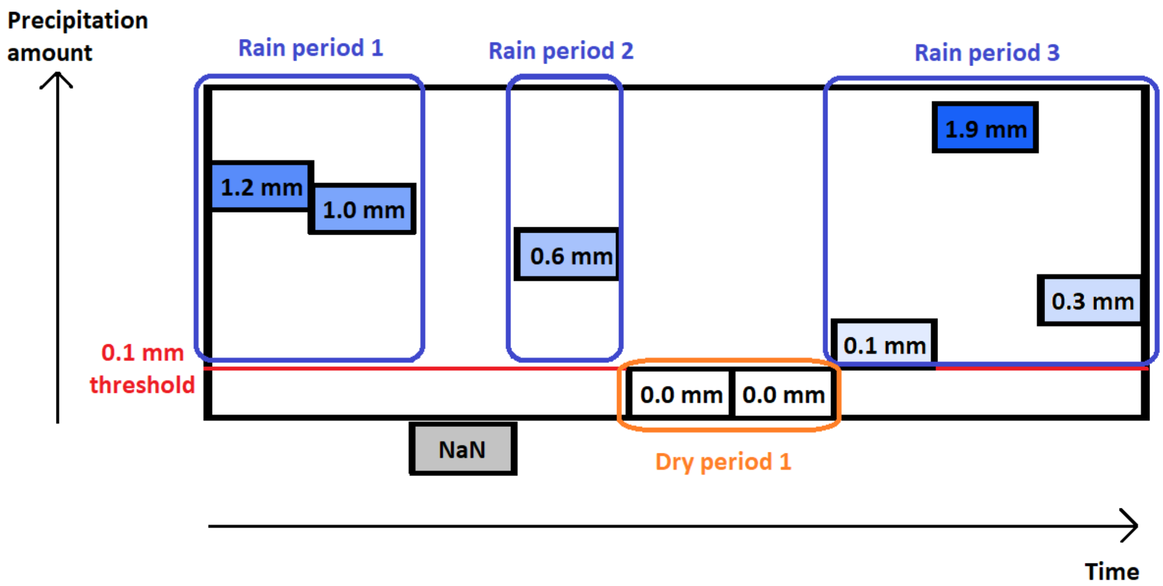


Figure 3.5: Schematic of how rain/dry weather periods were counted if NaN values are present.

3.5.2 Precipitation distribution

The next step was to look at the distribution of hourly and daily precipitation values to tell something about a location's precipitation pattern and whether it is dominated by no rain, light rain or heavy rain. This time, observed precipitation distribution were also compared against data from the AROME forecast and the custom-made fcfix forecast.

Precipitation values were sorted into various groups based on the quantity (frequency distribution). Table 3.3 shows the specific values and intervals used. Values less than 0.1 mm were treated as no rain, and therefore displayed as 0 mm. 0.1-1 mm/h for instance means $0.1 \leq$ mm/h ≤ 1, while all other intervals listed, e.g. "1-2 mm/h", stand for $1 <$ mm/h ≤ 2. NaN values were given their own separate section.

Other common statistical scores used were the mean value, precipitation intensity and percentiles. The mean was calculated as the average precipitation amount when including all non-NaN data points (all observed values, and the non-rounded values from the first 12 forecast hours from AROME and fcfix). Precipitation intensity is the average rainfall

Table 3.3: Specific numbers and intervals used for frequency distribution and percentile calculations

Hourly frequency distribution	**Daily frequency distribution**	**Percentiles**
0 mm/h	0 mm/day	25
0.1-1 mm/h	0.1-5 mm/day	50 (median)
1-2 mm/h	5-10 mm/day	75
2-3 mm/h	10-15 mm/day	90
3-4 mm/h	15-20 mm/day	99
4-5 mm/h	20-25 mm/day	
5-6 mm/h	25-30 mm/day	
6-7 mm/h	30-35 mm/day	
7-8 mm/h	35-40 mm/day	
>8 mm/h	40-45 mm/day	
	45-50 mm/day	
	>50 mm/day	

amount when we only include rain hours/days ($\geq$ 0.1 mm) and ignore the rest.

Percentile (also knows as k-th percentile) is the precipitation intensity value where a given percentage k of all rain hour/day values are less than or equal to. E.g. if there were 5176 rain hours in total at one location, the 90 percentile rainfall amount would be rank 517 of 5176 (rounded down), and thus higher than 90% of all rainfall values. Table 3.3 displays all percentiles used (identical for hourly and daily data).

3.5.3 Extreme precipitation

What counts as an extreme precipitation event can be a complex and subjective discussion where not only the actual rainfall amount matters, but also the time- and spatial scale of the event, intensity, return period/frequency, and any consequences for human life and the surrounding wildlife and infrastructure (Barlow et al., 2019). Nevertheless, the total measured precipitation should be seen in relation to the climatological normal (30 year moving period, currently lasts from 1991-2020) for that area in order to decide if it could qualify as an extreme event. For example, although 50 mm rain over 24 hours in Bergen is uncommon, it is hardly enough to be seen as an extreme event on its own. However, 50 mm precipitation in 24 hours inside the polar desert at Longyearbyen, Svalbard, could cause some truly devastating effects, because the frequency of such an event would be extremely low, and the local landscape and society would very likely not be able to withstand such amounts.

One definition used for extreme precipitation is all days above the 99.5 percentile of rain days from the current climatological normal for that area. With the ongoing climate changes, this threshold value could change over time, pushing the boundaries of what can be characterized as an extreme event. This is something that should be kept in mind when designing infrastructure in the future (Sorteberg, 2012).

Using the 99.5 percentile with our dataset would only return 4 out of 882 daily values, so to give ourselves a bit more data to work with we slackened this definition somewhat. Instead, the mean value of top 10%, 1% and 0.1% of all hours, and top 10%, 5% and 1% of all days were used to classify extreme precipitation events (including dry hours/days, but excluding NaN hours).

While comparing these results between observed and forecast (plus fcfix) would give a good indication on how accurately AROME can predict extreme precipitation AMOUNTS, they do not tell anything about how good it is at forecasting extreme events at the correct TIME. To do that, two new terms were introduced: Top x% shared extreme hours/days, and top x% shared extreme hours with 6 hour tolerance.

For the first term, the wettest top 1% and top 0.1% of all hourly precipitation values (top 5%/1% for daily values) were identified and ranked for both forecasted and observed data, along with their recorded time and date of occurrence. These times were then cross-checked between each dataset to find how many values they had in common within the top rankings. In other words, how many of the forecasted extreme precipitation values *also* turned out to appear among the top observed extremes? Any top x% ranking here would suffice to count as a hit, meaning if the 14th wettest forecast hour didn't end up being the 14th wettest observed hour, it would still count as a hit as long as both hours were within the wettest top x% of all hours. All hits were summed up and divided by the total data points within the top x% bracket to get a "hit rate" score ranging from 0 to 1, with 1 being perfect score. Figure 3.6 illustrates how this process was done.

Since these requirements of getting a hit are quite harsh (e.g. forecast extreme could be off by only one hour and it would still count as a miss), especially during convective-driven extreme precipitation events, a separate category with a 6 hour tolerance period were also added for hourly data. I.e. forecasting the extreme event up to 6 hours too early or too late would now be deemed as acceptable. A few rules needed to be set before these new calculations could be made. First, each forecast/observed value could only be used once. This was to prevent scenarios like forecasting extreme rainfall several hours in a row but only one of them were observed, scoring several hits in the process as they technically all were within the 6 hour tolerance period. Second, the closest matching hours were given priority in the following order: +0h (exact same time), +1h (forecasted one hour too late), -1h (forecasted one hour too early), +2h, -2h and so on up until -6h.

While neither of these methods are able to tell the whole story on their own, they should still complement each other well. If the model was extremely good at forecasting the correct amount, but never at the correct time, it would be very noticeable by reading the shared extreme hours results. Likewise, if the model always forecasted extreme events right when it should, but consistently predicted way too little precipitation, the mean extreme precipitation values would show a clear deviation.

Lastly, any seasonal variations in extreme precipitation events were looked at. We used two parameters, mean extreme precipitation amount for top 1% of hourly values (top 5% for daily values), and frequency of forecasted and observed extreme precipitation events (return values). Considering some seasons contained more data points than others, and extreme rainfall tend to be very season-dependant, using the above method of shared extreme days could run into some issues at drier locations without any notable extreme precipitation during certain seasons. The top x% ranks could get very "muddy" with lots of similar low-end values, making the hit rate requirements borderline unfair to overcome. Instead, a frequency-based metric was selected, as it functions independent of sample size.

Seasonal return values for extreme precipitation were calculated by taking the top 1%/5% of all hours/days and distribute them by which season they were recorded. This meant the list of extreme precipitation events by each season could have varying lengths, where the shorter lists could have some extra uncertainty due to low sample size.

Top 0.1%

	Forecast:		Observed:	
Rank	Amount:	Day:	Amount:	Day:
1	72.6573	523	84.6	468
2	64.5165	623	83.8	9
3	59.8462	525	67.4	395
4	58.1246	42	66.6	97
5	56.8432	9	63.9	176
6	53.1686	158	63.1	523
7	53.0987	740	62.4	741
8	52.7613	737	59.9	737
9	48.4682	642	58.0	768
10	47.3156	202	54.1	810
11	45.7215	420	51.6	69
12	41.8344	591	47.2	42

3 hits:	523	9	737
7 misses:			
Score:	**0.3**		

Figure 3.6: Schematic of how top x% shared extreme hours/days were calculated. In this hypothetical example, the 10 wettest days constitute all values above the top 0.1% threshold. Day means which day in the dataset the rainfall was recorded, with day 1 being 1. December 2019 and then counting upwards. Day 42 barely does not make the cut for observed data and thus ends up as a miss. Overall, with 3 hits and 7 misses, this location gets a score of 0.3. Also notice how the forecast consistently underestimates the extreme rainfall amount.

3.6 Dichotomous forecast verification

This section explains how the forecast verification methods outlined in Chapter 2.1 were used to calculate the results of dichotomous forecasts. Using the various scenarios listed in Table 3.2, a contingency table was created for each scenario (10 in total) and for each location. This was done for both AROME and fcday forecasts. Given that fcday only forecasts 24 hours into the future, only the +4h and +13h (1 hour accumulation length) as well as +4h-9h and +13h-18h (6 hour accumulation length) were valid.

To produce a contingency table, each individual precipitation forecast (AROME and custom-made) was compared against its respective observed precipitation value to determine if that forecast was a hit, miss, correct negative or false alarm based on the 0.1 mm threshold value (same for all accumulation lengths). Each outcome was added up and sorted into its own category, with NaNs as a 5th, unused category for all forecasts/observations containing at least one NaN value. This meant for longer accumulation lengths like 24h and 48h, some more data were lost as NaN since the longer the accumulation length is, the more likely it is to overlap with at least one NaN value.

For fcpersist data, this process was done slightly different. Since this forecast is only 12 hour long, a contingency table was made for each of the 12 forecast hours, in addition to two accumulated forecasts from +1h-6h and +7h-12h forecast hours (14 scenarios in total). This was done to more accurately analyse the hourly evolution of fcpersist forecast quality compared to AROME (which also received the same treatment). The 6 hour accumulation forecasts served to reduce the variance between each individual hour.

After all contingency table outcomes were summed up to get the total number of hits, misses, false alarms and correct negatives for each scenario (NaNs excluded), these were used to calculate the various verification methods explained in Chapter 2.1. In the case of Brier skill score, we needed to make a new contingency table for a "background" or "climatology" forecast to function as the reference Brier score (BS_{ref}) in order to compare its relative skill with AROME, fcday and fcpersist. One alternative would be to randomly guess the occurrence of rain and no rain, and in a binary forecast setting with only two possible outcomes, this reference forecast would be correct 50% of the time, resulting in a BS_{ref} score of 0.5.

Instead, the reference forecast was set to always predict the most common outcome of rain or no rain based on the observed data for a given scenario, thus lowering the BS_{ref}

score and making it harder to beat. Using the reduced version of Brier score in Equation 2.9, we get two possible solutions. The first solution is where the reference forecast always predicts rain, resulting in all misses converting to hits (observed rain), and all correct negatives would be changed to false alarms (no observed rain). With miss being an impossible outcome, the Brier score equation is reduced to Equation 3.1.

$$BS_1 = \frac{False\ alarm}{Total} \tag{3.1}$$

Likewise for the second solution, in scenarios where no rain was observed most often, the reference forecast would always predict dry weather. All hits/false alarms would be converted to misses/correct negatives respectively, and with no false alarms left, we get Equation 3.2.

$$BS_2 = \frac{Miss}{Total} \tag{3.2}$$

The chosen BS_{ref} was set to whichever the lowest value of BS_1 and BS_2 was at a given scenario, and is always between 0 and 0.5.

Chapter 4

Results

This chapter will present the most important results found in this thesis. Given the sheer amount of data involved and the numerous methods used, this chapter's figures will mostly be limited to hourly data, since with a sample size of over 20 000 hours they should generally be more accurate. Other results like for daily data will also be covered, though mostly by referring to tables and figures in the Appendix. For the most part, daily data results show the same pattern as hourly results.

The structure in this chapter is mostly the same as the last part of Methods chapter, starting with climatology which includes weather persistence, precipitation distribution and extreme precipitation, followed by dichotomous forecast verification results.

4.1 Climatology

Table 4.1 shows the total recorded precipitation for AROME forecast and observed data during the analysis period at each location, as well as their ratio. Keep in mind this is accumulated precipitation after NaN values from the two data sets were overlapped with each other (as explained in section 3.3.1), meaning the actual observed rainfall is a little bit lower. For example, Bergen (Florida) recorded 6386.6 mm, thus roughly 50 mm rain was removed to provide a fair basis of comparison with forecasted precipitation.

Table 4.1: Total recorded precipitation for AROME forecast (using first 12 forecast hours) and observed data from 1. December 2019 to 31. April 2022. The ratio listed on the right is the same ratio fcfix uses to bias-correct the mean forecast precipitation, and thus the total fcfix precipitation is equal to observed.

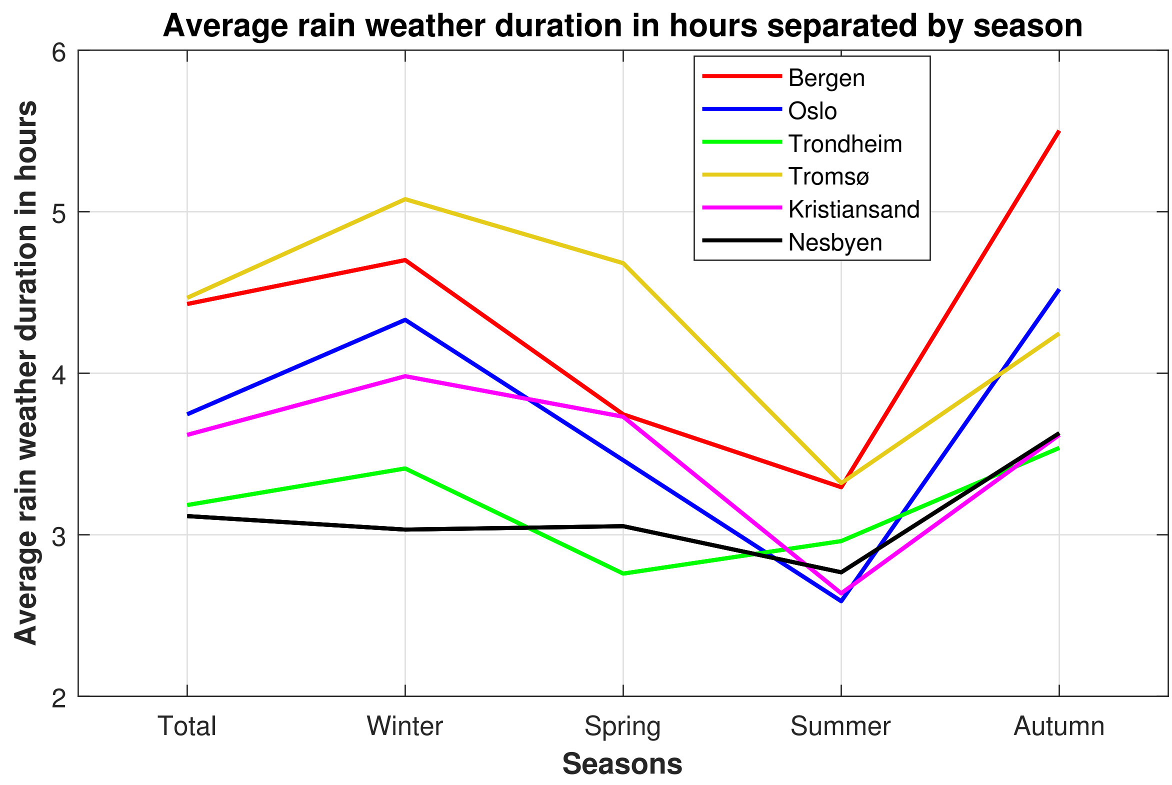
	Total forecasted precipitation	**Total observed precipitation**	**fcfix ratio**
Bergen	4924.3 mm	6336.0 mm	0.777
Oslo	2076.2 mm	1914.9 mm	1.084
Trondheim	2482.5 mm	2171.2 mm	1.143
Tromsø	2431.5 mm	3090.7 mm	0.787
Kristiansand	3242.3 mm	3219.6 mm	1.007
Nesbyen	1287.1 mm	1059.6 mm	1.214

The main talking point is the AROME model in Bergen and Tromsø seem to substantially underestimate the total rainfall, only recording 78% and 79% of observed rainfall, respectively. Especially given the wet climate in Bergen, the AROME model generating 1400 mm less rain during a 2.5 year timespan is quite remarkable. Oslo, Trondheim and Nesbyen on the other hand forecasts more precipitation that what was observed, but the total quantity is not too far off. Nesbyen has the biggest discrepancy where AROME predicts 21% too much rain, although Nesbyen is also by far the driest location here, thus even a relatively small difference in absolute precipitation could have a big impact on the ratio. Kristiansand almost gets a perfect match with only a 0.7% difference. This means AROME and fcfix forecast will by all intents and purposes act as the same forecast here and reflect as such in the figures below.

4.1.1 Weather persistence

Figure 4.1 presents the average rain/dry weather duration in hours for each location, both overall and seasonal. The numbers themselves are listed in Figure A.1. When looking at the overall duration, there seem to be three ”pairs” developing. The average rain weather duration in Bergen/Tromsø is 4.5 hours, a bit less than 4 hours at Oslo/Kristiansand, while Trondheim/Nesbyen experience the shortest rain periods, right above 3 hours on average.

When examining the seasonal variance, it is clear summer rainfall is generally very brief, while winter and especially autumn months experience longer sessions of precipitation, with 5.5 hours on average in Bergen as the highest. This feature fits well with the perception of autumn/winter being dominated by larger low-pressure systems precipitating



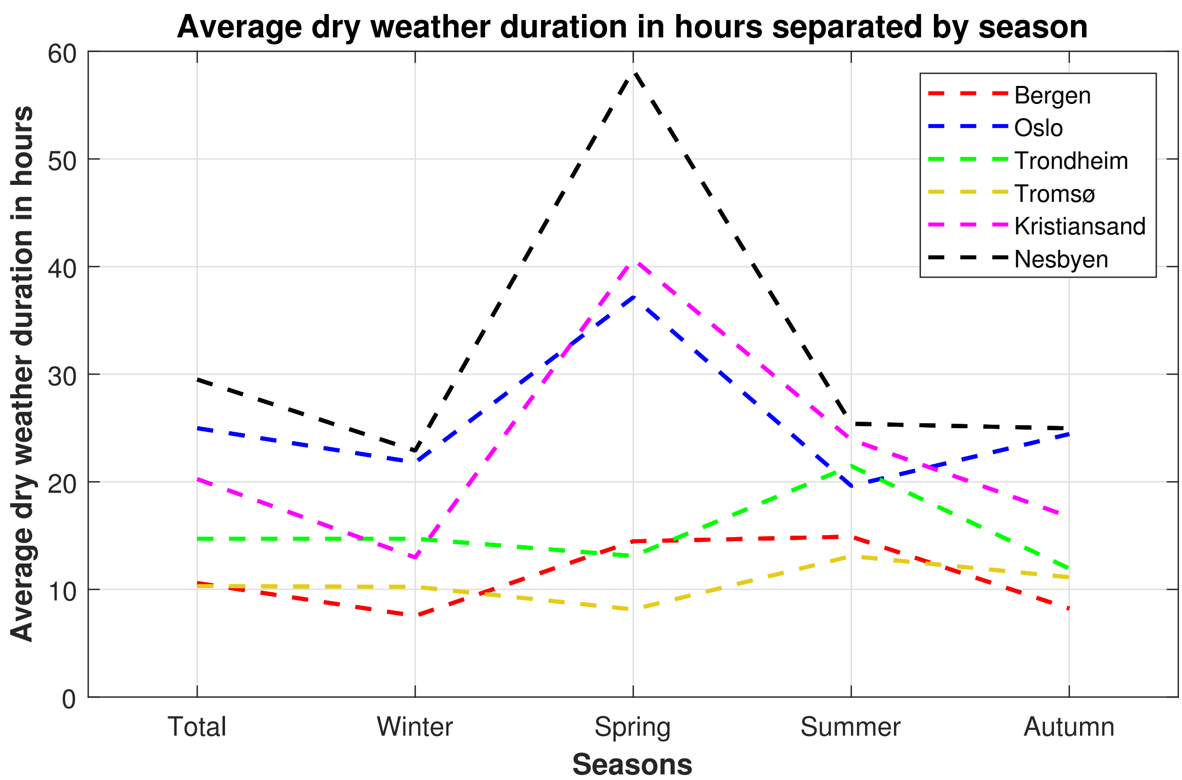


Figure 4.1: Average rain/dry weather duration in hours for each location, separated by season. Total is average of all seasons. Data from 1. December 2019 - 31. April 2022.

over a longer period of time, while summer rain often originates from short-lived convective systems. This is further backed up by the stations located in southern and eastern part of Norway (Oslo, Nesbyen and Kristiansand), where convective precipitation is most common, experiences the shortest rain weather durations during summer.

Since total rain/dry weather periods for a given location necessarily must be practically identical, the average rain/dry weather duration is also directly proportional to the number of rain/dry hours. Interestingly, there only seem to be a weak correlation between total precipitation amount and how long the average rain shower lasts. We can see Bergen and Tromsø get about the same number of rain hours, despite Tromsø receiving less than 50% of Bergen's total rainfall amount. Same story can be seen for Trondheim and Nesbyen, where Trondheim recorded twice as much rainfall even though the average rain weather duration is barely any longer than in Nesbyen.

The average dry weather duration do in some ways follows the opposite pattern of average rain weather duration. Note the scale on the y-axis is different. Nesbyen tops the overall duration, expecting about 30 hours of dry weather between each rain shower, while Tromsø and Bergen have to settle with only 10 hours on average. Spring sees a clear distinction between the south-eastern locations and the rest of Norway, where Nesbyen can anticipate almost 60 hours (2.5 days) of continuous dry weather on average before any significant rain is recorded. These three locations also experience the highest seasonal variation in general, whereas Bergen and Tromsø in particular see less variance in weather stability through the year.

This data was also calculated for daily weather duration periods (Figure A.2 in the Appendix), and the results mostly show the same features as for hourly data. Tromsø experiences the longest rain weather duration by far during spring and summer, with 6.2 days during spring as clear stand-out point, being 2.5 days longer than Trondheim in second place. Not surprisingly, Bergen sees the longest rain periods during autumn, where it rains (more than 0.1 mm) for almost 9 days straight on average followed by only 2 days of dry weather.

4.1.2 Precipitation distribution

Where Table 4.1 showed whether AROME predicted too little or too much rain over the whole analysis period, the figures in this subsection will look at where the deviation (if any) takes place when distributing by precipitation quantity, and if fcfix can fix this deviation.

When looking at hourly data (see Figure A.3 in the Appendix), the majority falls into the 0 mm category (ranging from 69% (Tromsø) to 90% (Nesbyen) of all observed hours). 0.1-1 mm sits in a clear second place (8% (Nesbyen) - 26% (Tromsø)), and generally speaking, the more intense rainfall, the rarer they become. For instance, it rains less than 1 mm/h in 98.8% of all non-NaN hours observed at Nesbyen, meanwhile the same number for Bergen on the other side of the scale is 90.9%.

For values > 5 mm/h (heavy rainfall), Bergen observes by far the most cases, with 125 hours in total. Kristiansand observes 63 hours, while the others range from roughly 10-30 hours. Tromsø is placed last with only 11 hours recorded, and none of them above 8 mm/h. Regarding Oslo, 10 out of 27 hours above 5 mm/h are also above 8 mm/h, a fraction much higher than for the other locations. This shows the occurrence of very heavy rainfall (even way higher than 8 mm/h) is more common relative to "moderately" heavy rainfall.

Figure 4.2 displays the ratio of total forecasted (AROME/fcfix) to observed precipitation hours for each quantity bracket, where a ratio of 1 is the preferred outcome. Starting in Bergen, AROME forecasts too many dry hours (+11%) and thus too few rain hours, and fcfix is not able to reduce the amount of dry hours in a significant way. The same can be seen in Tromsø, where the difference is 15%, or 2197 extra dry hours. Even with fcfix increasing AROME precipitation by 27% for every single hour, this was still only enough to convert 399 of the excess dry hours into rain hours.

Likewise, AROME severely under-predicts number of hours in the 0.1-1 mm bracket for Bergen (30% less) and Tromsø (37% less), and again fcfix is not able to correct this in any capacity, likely due to most of the extra hours gained from the 0 mm/h bracket were pushed further into the 1-2 mm/h bracket, resulting in almost no net gain. Kristiansand's AROME forecast on the other hand seems to perform very well, staying close to the black dashed line up until 6 mm/h, which accounts for >99.8% of all non-NaN hours.

Past the >4 mm/h brackets or so, the ratios start to become highly irregular, but most of this should probably be blamed on the low sample size at such high intensities. Bergen

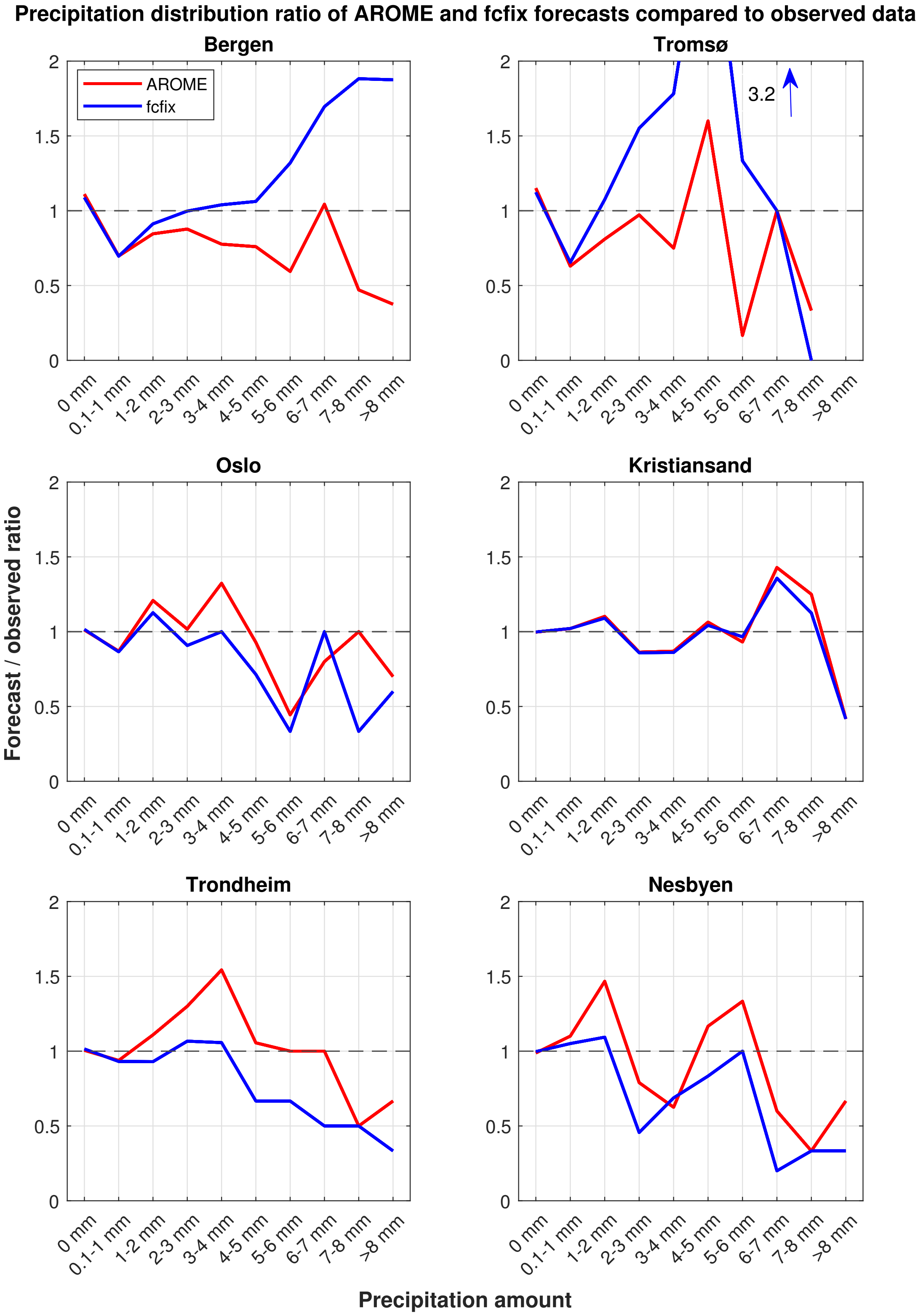


Figure 4.2: Ratio of hourly precipitation distribution amount between forecast (AROME (red lines) and fcfix (blue lines)) and observed data for each location. Black dashed line highlights a ratio of 1, which is the ideal outcome where both forecasted and observed data recorded the same number of hours within a specific quantity bracket. Tromsø has one value off-chart at 3.2 (blue arrow) to keep the y-range consistent. Data from 1. December 2019 - 31. April 2022. Forecast data taken from the first 12 forecast hours.

fcfix seem to overcompensate quite significantly and shifts way too many hours from lower brackets into the highest ones. This should also be quite reliable unlike for other locations since heavy rain in Bergen is a relatively common event. The general pattern for AROME during heavy rain seems to be a slight overall underestimation (although with some uncertainty), and that fcfix, despite with the total precipitation forecast bias removed, is not able to produce a reasonably more accurate precipitation distribution.

When looking at daily data (Figure A.4), the deviation seen between AROME and observed hourly data in Bergen is now reduced for daily data (279 vs. 264 days), but very much remains in Tromsø (286 vs. 242 days). AROME predicts way too few heavy rain days (> 30 mm/day) in Bergen, but unlike with hourly data, fcfix now manages to correct this underestimation quite nicely, and is all in all a very good precipitation distribution forecast here. A rather peculiar fact that really showcases the variety in heavy rain days across the country is that Bergen registers 31/46 of all observed days above 35 mm rain across *all* locations, where not a single day comes from Tromsø or Nesbyen.

Figure 4.3 shows how the AROME and fcfix hourly forecasts perform compared to observed data at various percentiles, as well as for overall precipitation intensity. The main difference here is dry hours are ignored completely, meaning this figure only says something about the distribution *whenever* it rains. Broadly speaking, the higher the rainfall amount is, the more trustworthy these ratios are, as dividing two small numbers with each other may lead to wildly varying results.

AROME precipitation intensity is in general slightly higher than observed, but not too far off. Kristiansand and Bergen have the highest observed intensity values with 1.04 mm/h and 1.02 mm/h, respectively (Figure A.5 in the Appendix). Tromsø receives the least amount of precipitation when raining with only 0.49 mm/h on average. This can be explained by being the location with the highest number of rain hours (dragging the average down), while only getting a moderate amount of precipitation overall compared to other locations.

25th percentile shows a very distinct spike almost everywhere, and many locations end up with a ratio approaching 2. This however can largely be attributed to the low resolution of observed data. As the Appendix results show, four of the six locations have a 25th percentile of 0.1 mm, while Kristiansand and Bergen have 0.2 mm. Since percentiles only look at a single value, it cannot know what the values around it are, and if the value

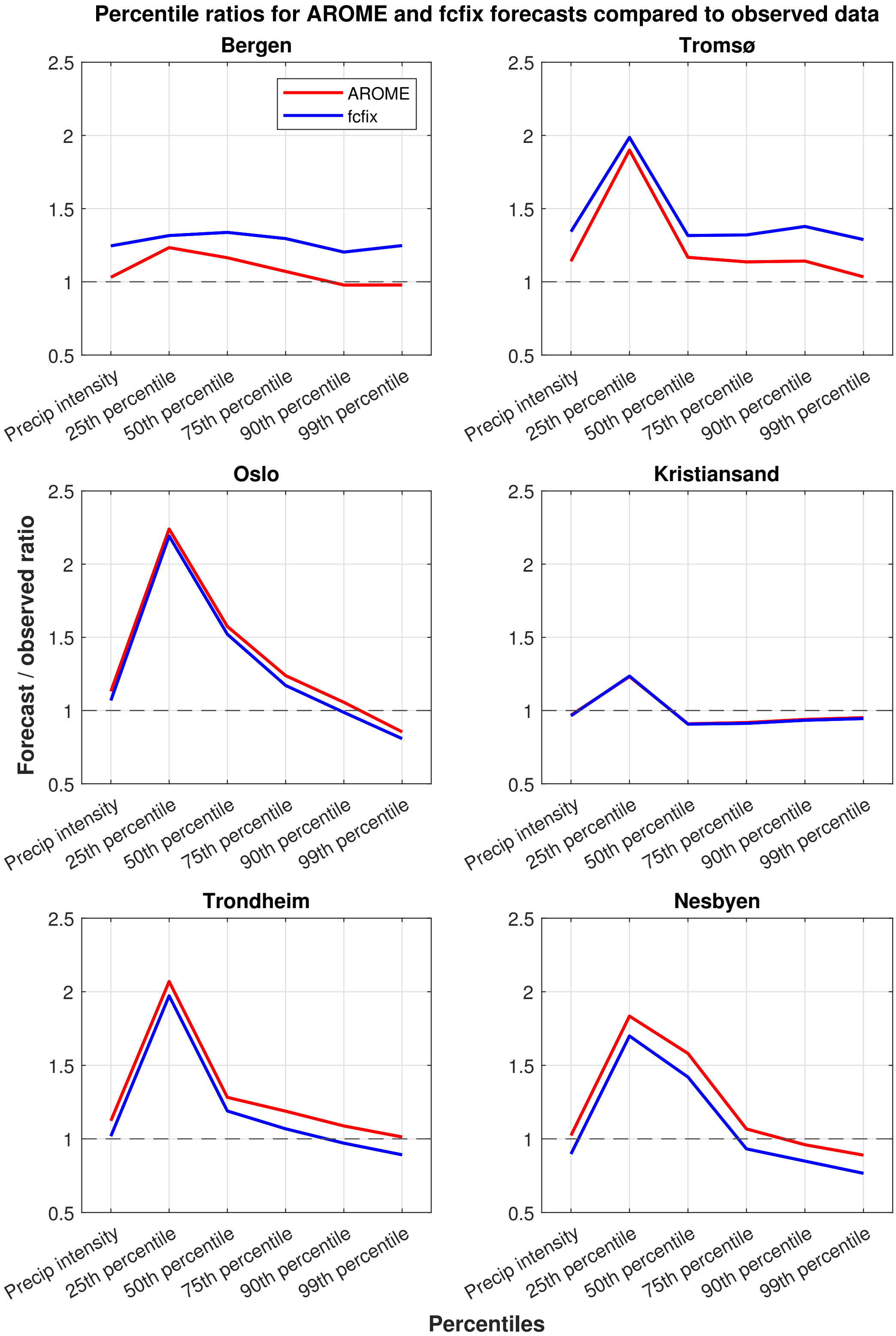


Figure 4.3: Ratio of hourly precipitation intensity and various percentiles between forecast (AROME (red lines) and fcfix (blue lines)) and observed data for each location. Black dashed line highlights a ratio of 1, which is the ideal outcome where both forecasted and observed data recorded the same rainfall amount for a given percentile. Data from 1. December 2019 - 31. April 2022. Forecast data taken from the first 12 forecast hours.

ranked above it (the 25.01th percentile if we want) happens to be 0.2 mm instead of 0.1 mm, this would obviously have a huge impact on the resulting ratio.

The general trend seems to be too high ratio at lower percentiles (forecast precipitation higher than observed), but it decreases towards higher percentiles and even goes below 1 at some locations. This would suggest the model predicts a bit too much precipitation during light rain, but at some locations may struggle to generate enough precipitation during more extreme events. In other words, the overall AROME precipitation distribution might be too narrow. Apart from the 25th percentile spike at Tromsø, the ratio development towards higher percentiles in Bergen and Tromsø does not see much of a change.

Fcfix results in general reflect the total precipitation ratios listed in Figure 4.1, where Bergen and Tromsø see a relatively large adjustment, while AROME and fcfix forecasts in Kristiansand once again show practically identical results. Fcfix adjusts the ratio in the right direction (towards 1) in Oslo, Trondheim and Nesbyen, but worsens the ratio in Bergen and Tromsø. This indicates the quantity distribution when raining is already quite good for AROME Bergen and Tromsø despite predicting significantly less rain overall, however AROME forecasting too few rain hours overall as seen above also plays a role here.

Looking at daily data in Figure A.6, precipitation intensity appears to follow the fcfix ratio pattern on total precipitation (i.e. too low values in Bergen and Tromsø). The AROME ratio for these two locations tend to be too low for percentiles, but unlike with hourly values, the extra precipitation in fcfix improves ratio closer to 1 rather than make it worse. For the other locations the pattern seem to repeat itself with high ratio at the 25th percentile (though not as spiky as for hourly data), before gradually decreasing when approaching higher percentiles.

Figure 4.4 shows the monthly precipitation ratios between AROME/fcfix and observed data, with Figure A.7 to A.9 displaying all the numbers. The singular high spikes during spring in Oslo, Nesbyen and Kristiansand are a result of unusual dry months where only a couple of millimetres difference can cause the ratio to sky-rocket. For instance, Oslo registered 3.5 mm precipitation during the whole of March 2022, and while the forecasted rainfall was only 8.5 mm too much, this nevertheless results in a very high ratio.

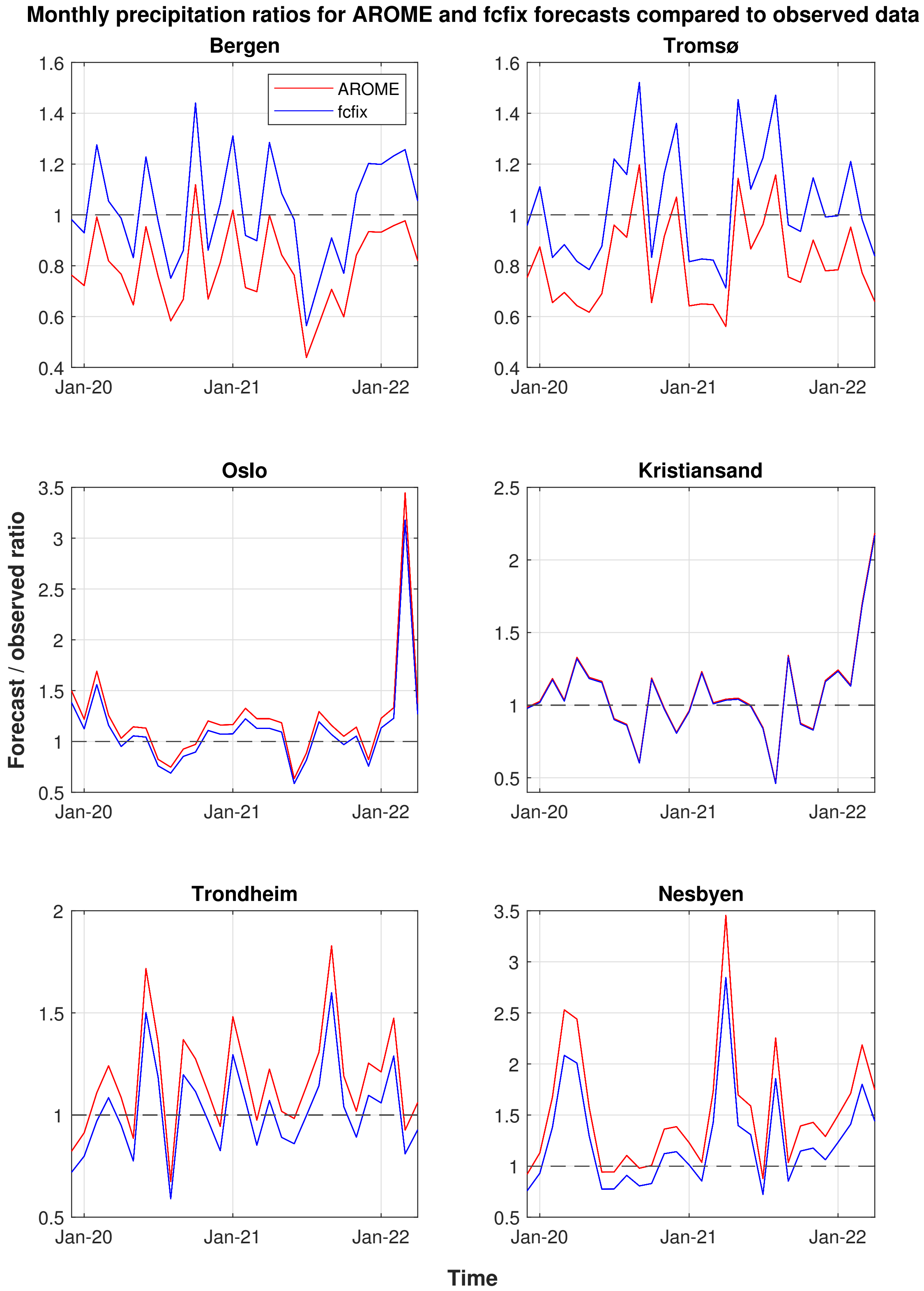


Figure 4.4: Monthly precipitation ratio between forecast (AROME (red lines) and fcfix (blue lines)) and observed data for each location. Black dashed line highlights a ratio of 1, which is the ideal outcome where both forecasted and observed data recorded the same rainfall amount for a given month. Note the y-axis range may vary. Data from 1. December 2019 - 31. April 2022. Forecast data taken from the first 12 forecast hours.

Even with some irregularities, Oslo and Kristiansand forecasts seem to perform the best, with Nesbyen not too far off given the majority of these months only recorded about 10-30 mm. Trondheim ratios alter quite rapidly between each month, and this station experiences enough months with rainfall surpassing ∼50 mm for the ratios to be relatively trustworthy. AROME forecasts too little precipitation in 27/29 months in Bergen and 25/29 months in Tromsø, which is to be expected with the model not forecasting a sufficient amount in the first place.

A recurring trend however is that AROME in most cases forecasts too much precipitation during very dry months (ratio approaches 2-4), but very rarely underestimates it. Another notable deviation is October 2021 in Bergen when the weather stations measures a record-high 647 mm rain, but the AROME model only manages to predict 387 mm, a difference of 260 mm or 40%.

As a side note, Trondheim (Lade) lost about 18 days in April 2020, and Kristiansand (Kjevik) lost about 10 days in October and November 2021 due to missing observed data. As explained in the Methods chapter, these values were also removed from forecast data to provide parity. While this means we do not know the "true" ratio for these months, they should nevertheless have very little effect on the wider picture.

Figure A.10 in the Appendix lists the seasonal precipitation ratios. AROME forecasts the least amount of rain during summer in Bergen, with only 66% of total observed rainfall, closely followed by autumn. This pattern is also prevalent for the south-eastern stations Oslo, Kristiansand and Nesbyen. In Tromsø however, AROME struggles the most during spring (66% of observed precipitation) and winter, while getting fairly close to the preferred ratio during summer and autumn. Trondheim's pattern share some similarities to Tromsø's, but with less variation and all seasons having a ratio above 1.

4.1.3 Extreme precipitation

This subsection discusses how the AROME and fcfix forecasts perform during extreme precipitation events, here defined as values within the top x% of all non-NaN data points. This involves if forecasts are able to both predict the correct precipitation amount, and do so at the correct time. Both of these parameters are particularly important to get right to maximise the forecast *value*. If forecasters fail to announce an extreme precipitation event due to the model missing it entirely, it may cause more serious damage than if local authorities could react and implement safety features ahead of time. Correspondingly, forecasting an extreme precipitation event either way too early or too late could catch the public by surprise and hinder society's ability to plan around the event.

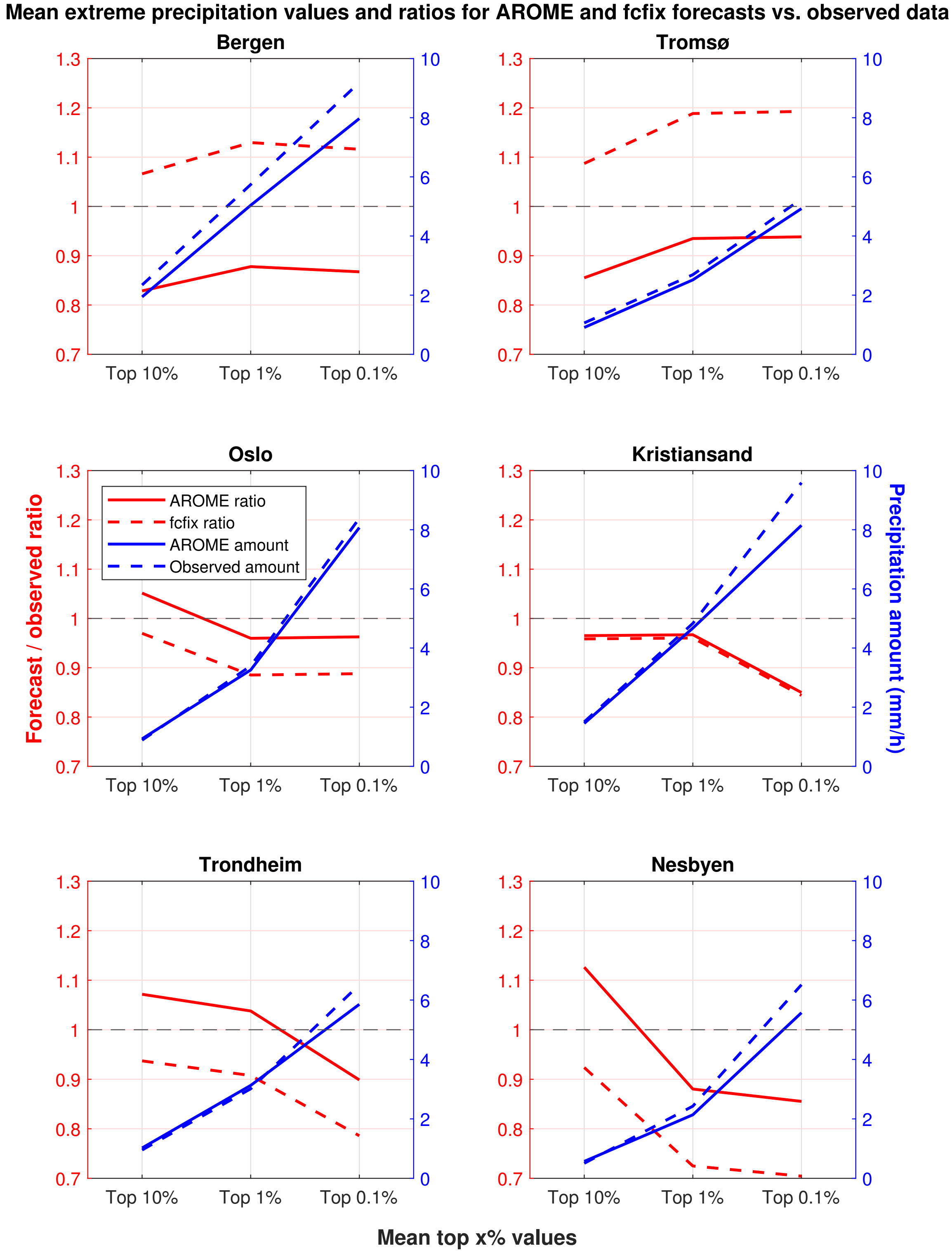


Figure 4.5: Hourly mean precipitation amount for top 10/1/0.1% of all values. Red lines show the ratio of said precipitation between AROME/fcfix forecasts and observed data, while blue lines display the actual precipitation amount for the various top x% categories for AROME forecast and observed data (no fcfix). Black dashed line highlights a ratio of 1, which is the ideal outcome where both forecasted and observed data ended up with the same hourly mean extreme precipitation amount. Data from 1. December 2019 - 31. April 2022. Forecast data taken from the first 12 forecast hours.

Figure 4.5 is two-fold; the red graphs show the AROME/fcfix ratio, while the blue graphs display the actual hourly mean extreme precipitation amount (top 10/1/0.1%) for AROME and observed data (no fcfix). Keep in mind these results also include dry hours, unlike with percentiles.

Starting with AROME ratios, the general downward trend towards higher rainfall levels is prevalent here as it was in Figure 4.3, and all locations end up with a ratio below 1 for the top 0.1% bracket. Bergen and Tromsø are once again outliers here with a slightly increasing ratio with higher rainfall levels. Oslo appears to achieve the best results, while Bergen, Tromsø and partially Kristiansand underestimate hourly extreme precipitation for all brackets.

Fcfix does not seem to improve the extreme precipitation forecast, but rather make it worse. Bergen and especially Tromsø overshoots the black dashed line too much, and Oslo, Trondheim and Nesbyen all remove precipitation compared to AROME which as mentioned already struggles reaching very high precipitation values to some extent. This suggests the general mean bias for AROME is not the main issue here.

When analysing the mean precipitation amount, the desired outcome is the blue solid and dashed line overlapping each other, signifying no difference in AROME and observed results. Oslo, Tromsø and Trondheim performs the best here, while Kristiansand takes a somewhat surprising win with the highest top 0.1% mean observed precipitation amount ahead of Bergen and Oslo. Tromsø in particular experiences fairly low extreme values, together with by far the lowest 90th and 99th percentile values. This showcases a place where it rains quite often, but very rarely reaches values that can reasonably be defined as extreme precipitation events.

The steepness of the blue line and whether it is smooth or breaks midway, can say something about the distribution of moderately (top 10%) to very extreme events (top 0.1%), because each bracket step reduces data volume by the same amount (90%). In Bergen, the seemingly straight line indicates a very even distribution; lots of hours with heavy rain that sort themselves pretty nicely. Oslo on the other hand measures relatively low top 10/1% mean values, but a very high top 0.1% mean value. This implies the extreme rainfall here is very "top-heavy" and dominated by a small group of incredibly wet hours, which is also consistent with the results found for precipitation distribution.

For daily extreme precipitation (Figures A.13 and A.14), the main findings are AROME ratios for Kristiansand and Nesbyen are pretty much spot on, Bergen and Trondheim fcfix the same, and the downward ratio trend in hourly extreme data is essentially gone.

Figure 4.6 illustrates to which degree the AROME forecast is able to predict extreme precipitation events at the correct time, by looking at the top 1/0.1% shared extreme hours with observed data, both with and without a 6h tolerance window. In the top 1% bracket, Nesbyen achieves the most hits and thus the highest score (0.63 with tolerance and 0.39 without), closely followed by Kristiansand. In the no tolerance category, Tromsø obtains by far the lowest score with 0.12, only getting 26 hits out of 211 values, less than half the number of hits Trondheim in fifth place manages to get. Tromsø catches up somewhat when including the tolerance window, but is still placed last with 83/211 hits and a score of 0.39.

The top 0.1% bracket only consists of the 21 wettest hours recorded, and as such the requirements for getting a hit is tougher. Oslo and Kristiansand achieve 3 hits each as the two best performers in the no tolerance category, while Trondheim and Nesbyen (the winner of the top 1% bracket) did not get a single hit. With tolerance, Oslo's AROME forecast delivers the most accurate extreme precipitation forecast at the appropriate time. Tromsø performs quite well relative the top 1% category results, while Trondheim only manages to get a single hit even with 6h tolerance. Given it also gets the second worst score in top 1%, it can be argued AROME in Trondheim lacks some ability to forecast hourly extreme events at the right time.

Daily data for top 5% shared extreme days (Figures A.13 and A.14) show all locations perform very evenly, ranging from 25 to 32 out of 44 possible hits. For top 1% category, 5 of the 8 wettest days in Oslo share the same date for both AROME and observed data, followed by Bergen and Trondheim at 4 hits each. Tromsø only has a single day in common, and get the worst results for shared extreme days by also receiving the lowest score in the top 5% category.

Figure 4.7 shows the seasonal variation in top 1% hourly mean extreme precipitation amount for AROME and observed data (red lines), as well as return value of these extreme events (blue lines). When comparing AROME and observational mean precipitation, Trondheim displays the least deviation, while Bergen, Nesbyen and Kristiansand are not forecasting quite enough precipitation to match the observed amount regardless of season. If we focus on the seasons where extreme events appear most often (usually summer and autumn), it is rather clear the AROME model slightly underestimates the rainfall intensity pretty much everywhere.

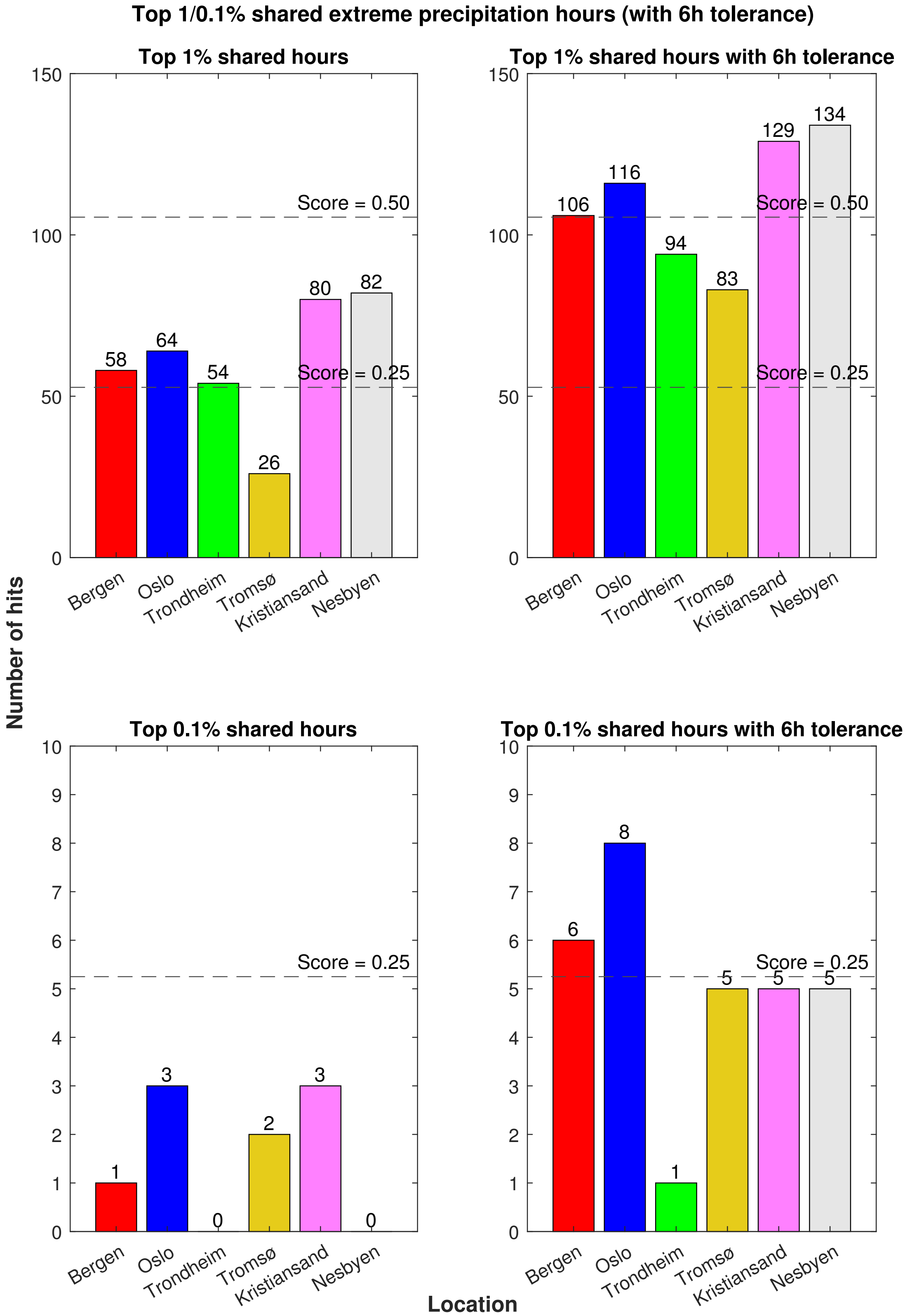


Figure 4.6: Top 1% (top) and 0.1% (bottom) shared extreme hours between AROME forecast and observed data, with a 6h tolerance window (right side) and without (left side). Bars show the number of hits for each location, and the black dashed lines indicate score thresholds of 0.25 and 0.5. The ideal score is 1 (all 211 hits for top 1% and all 21 hits for top 0.1%). For more details, see Figure 3.6. Data from 1. December 2019 - 31. April 2022. Forecast data taken from the first 12 forecast hours.

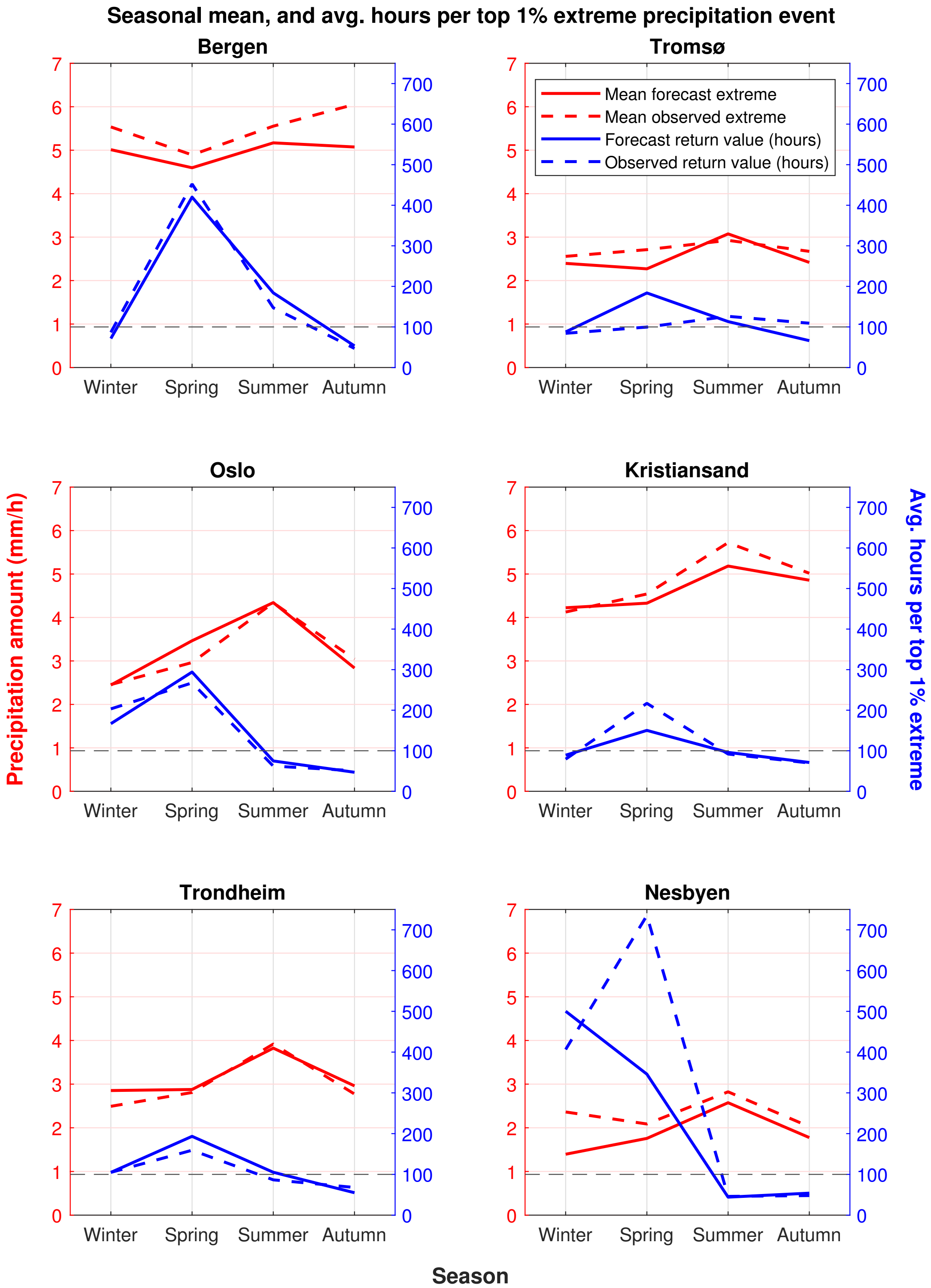


Figure 4.7: Seasonal mean for the top 1% extreme precipitation hours (red lines), and average number of hours (return value) between each occurrence of a top 1% extreme event (blue lines) for both AROME forecast and observed data. Overlapping red/blue solid and dashed line means forecast results match observed results, which is desirable. Black dashed line indicates a return value of 100 hours, which is the average when including all seasons. The higher the return value is, the more unlikely it is for an extreme precipitation event to happen during that season relative to other seasons. Data from 1. December 2019 - 31. April 2022. Forecast data taken from the first 12 forecast hours.

Summer months see the highest mean extreme precipitation, and this is especially notable in the south-eastern locations and Trondheim. This is likely the result of intense rain showers from convective-driven extreme rainfall events. Bergen and Tromsø see less seasonal variation, and Bergen is also the only location where observed extremes are most intense during autumn with 6 mm/h on average.

When studying the return values, an important feature is the relative difference between seasons, not necessarily the hourly return value in itself. Spring is by far the least likely season to experience an extreme precipitation event (highest number of hours elapsed on average per extreme event), and this is true for all locations.

Interestingly, observed values in Tromsø show almost no seasonal variation neither for mean values nor return values. From Figure A.10 we saw AROME heavily underestimating spring precipitation here, and the effects of this can be seen again as a clear deviation between AROME and observed results. Less forecasted spring precipitation overall leads to lower mean extreme values, which again lowers the number of spring hours making the top 1% cut, and ultimately increases the return value.

Nesbyen experiences the most drastic variations, where an observed extreme precipitation event is about 15 times less likely to happen during spring than during summer and autumn. While AROME argues a spring extreme hour should come about every 350 hours on average compared to 750 hours for observed data, this is another example where low sample size can exaggerate the results. That said, while only 8 extreme hours were observed during spring, the same number for AROME is 17 hours, which still can be considered a significant difference.

Figure A.16 from the Appendix shows daily results for top 5% values (44 days), and share many of the same main findings as for hourly data. For example, not a single extreme day was observed during spring in Nesbyen (which also records by far the lowest mean values of around 10-13 mm/day), and Bergen sees an extreme event (by our lenient definition) every 9 days during the autumn months (average return value is 20 days). The biggest difference is the prevalent hourly summer extreme events are all but gone from daily data, where most locations now experiencing a much more even distribution.

4.2 Dichotomous forecast verification

This subsection presents the forecast verification results as explained in Chapter 2.1 and 3.6. A total of nine different verification methods were used, and all results for the various forecast hours and accumulation lengths are listed in Figure A.23 and onwards in the Appendix.

Since there are *a lot* of numbers to be crunched, only the parameters deemed the most important are present in the figures below, and only for 1h and 6h accumulated precipitation. The other results will only be discussed very briefly, in particular mean square error (MAE) and root mean square error (RMSE) as the ratio between these can say something about the average size of each precipitation error. If the values are relatively close to each other (low ratio), the error are mostly small. If the values are far away from each other (high ratio), then there are considerable amount of large errors.

- Accuracy - fraction of forecasts that are correct, and a way to describe the forecast quality in itself.
- Bias frequency - ratio of forecasted rain events compared to observed rain events, and shows whether the forecast has any bias towards too few or too many rain events.
- Brier skill score - the relative skill of a forecast compared to the climatology, and takes into account the difficulty of forecasting the weather at each location. For instance, forecasting dry weather at a very dry location will most often be correct and yield a high accuracy score, but that does not mean that forecast has high skill.

4.2.1 AROME, fcfix and fcday forecast verification

Figures 4.8 and 4.9 show the accuracy (red), bias frequency (blue) and Brier skill score (yellow) for 1h accumulated precipitation at each location. There are three forecast lengths for AROME and fcfix, and two for fcday. There is also an accompanying table (Figure 4.10) where all numbers are listed.

The three south-eastern stations achieve the highest AROME +4h forecast accuracy with around 0.90, followed by Trondheim (0.84), Bergen (0.82), while Tromsø gets the lowest score with 0.79. Accuracy also seem to keep up fairly well as forecast lengths are increased, with only a small reduction in the range of 0.02 - 0.04 (no reduction in Trondheim).

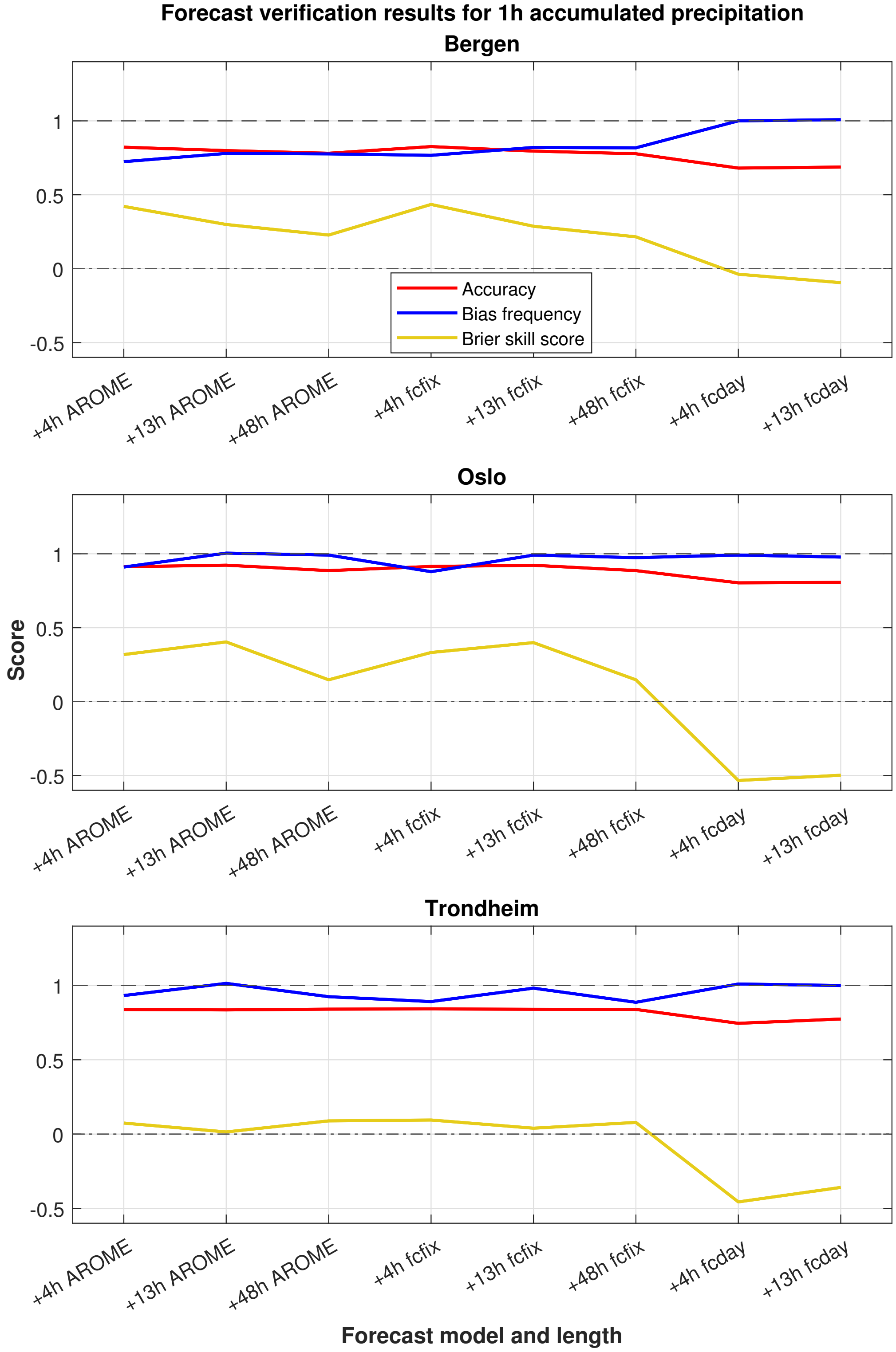


Figure 4.8: Accuracy, bias frequency and Brier skill score verification results for 1 hour accumulated precipitation at Bergen, Oslo and Trondheim. Forecast lengths were +4h, +13h and +48h, using forecast models AROME, fcfix and fcday (no +48h forecast). The ideal score for all parameters are 1, shown as a black dashed line. Brier skill scores below 0 (black dash-dotted line) signify that the forecast model quality is worse than the reference forecast based on climatology. Data from 1. December 2019 - 31. April 2022.

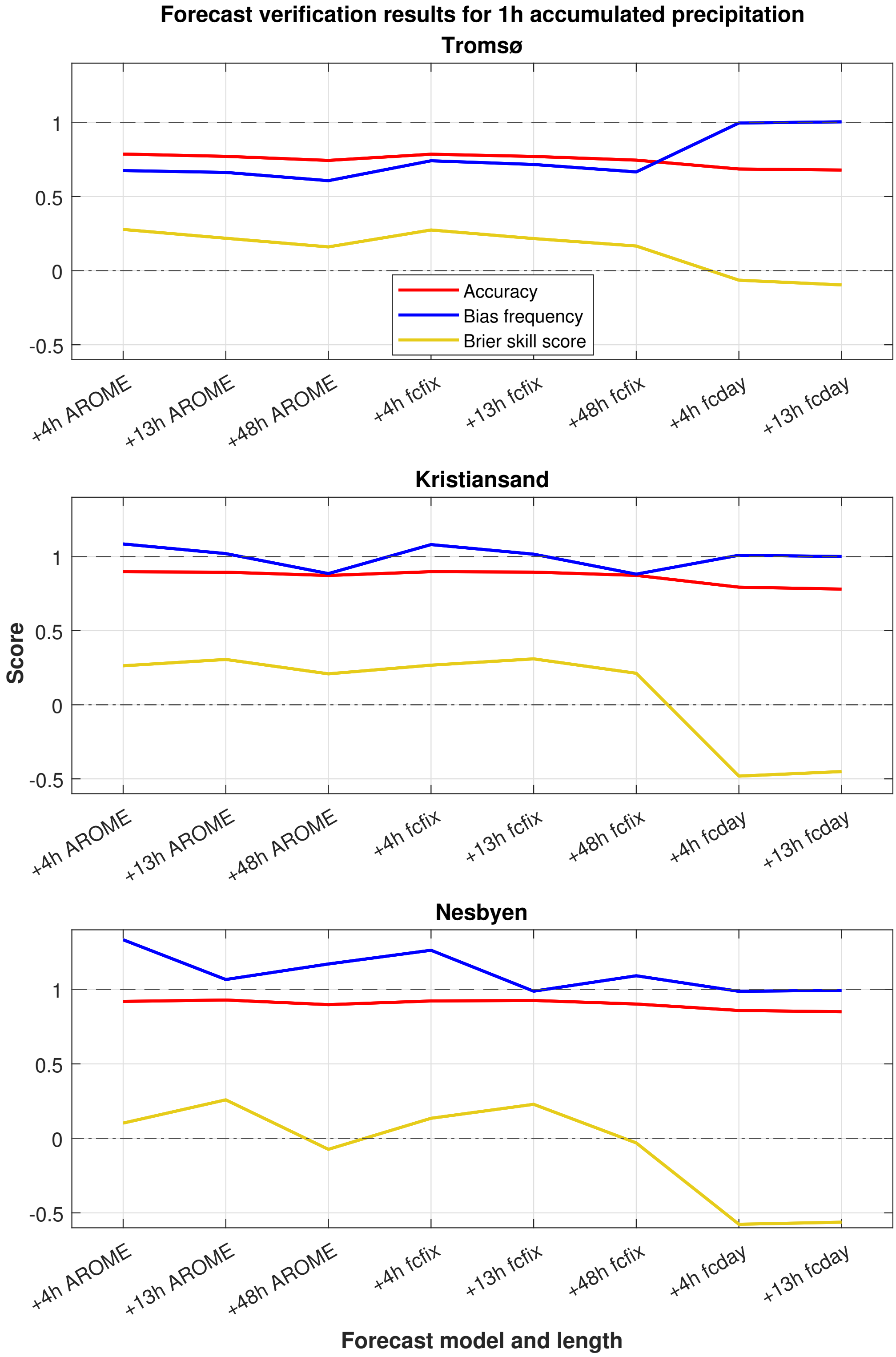


Figure 4.9: Accuracy, bias frequency and Brier skill score verification results for 1 hour accumulated precipitation at Tromsø, Kristiansand and Nesbyen. Forecast lengths were +4h, +13h and +48h, using forecast models AROME, fcfix and fcday (no +48h forecast). The ideal score for all parameters are 1, shown as a black dashed line. Brier skill scores below 0 (black dash-dotted line) signify that the forecast model quality is worse than the reference forecast based on climatology. Data from 1. December 2019 - 31. April 2022.

Do fcfix increase the accuracy? Not really. There are only tiny differences to be seen, and none of them are significant enough to warrant any attention. Fcday however reveals itself as a very poor forecast (as expected), with a clear drop in accuracy for all locations, with the biggest drop in Bergen (0.82 to 0.68).

Bias frequency results see some more fluctuations. Tromsø (0.68) and Bergen (0.72) have a notably low bias score for +4h AROME forecast, showing a clear underforecast tendency of rain events (negative bias). Nesbyen is the only location with a considerably high bias score (1.33), indicating false alarm events are more prevalent than miss events. When increasing the AROME forecast hours, the bias frequency tendency seems quite erratic depending on location. Bergen sees a slightly less negative bias, Oslo has no bias any more compared to +4h, Tromsø's negative bias actually amplifies a bit, and Kristiansand goes from a slight positive bias to a slight negative bias.

Fcfix bias share mostly the same patterns as AROME, without changing the numbers much. Nevertheless, it seems to remove a little bit of the negative bias in Bergen and Tromsø, which is likely related with the rather big increase in fcfix rainfall amount, and therefore it gains some extra false alarms and fewer misses (AROME dry hour values pushed over the 0.1 mm threshold). Fcday shows practically no bias whatsoever in any location.

Brier skill score in Bergen and Tromsø display a clear falling trend towards higher forecast hours, while the south-eastern stations (especially Nesbyen) show a rather interesting pattern of highest BSS from the +13h forecast (highest skill) and lowest from +48h. Trondheim's AROME forecast is barely more skilful than the reference climatology forecast, and Nesbyen's +48h forecast is actually slightly worse overall than the climatology forecast. Although to be fair, since Nesbyen is such a dry place, it makes the background forecast really hard to beat. One can essentially never predict rain, and based on these hourly data, it would produce an accuracy score as high as 0.90, which is higher than many AROME forecasts elsewhere.

Looking at MAE and RMSE (see Figure A.23), their values are mostly proportional to the total precipitation for each location, which makes senses since they look at the average forecast error per event. Their values and therefore the average errors increase slightly with longer forecast hours for about every location, which is to be expected. RMSE to MAE ratio in Oslo is somewhat higher than places like Trondheim and especially Tromsø. This indicates there are more severe forecast errors present here than in Tromsø, which sees more low-value errors as the norm. This coheres with previous results giving

1h accumulated precipitation

	AROME			fcfix			fcday	
BERGEN	+4h	+13h	+48h	+4h	+13h	+48h	+4h	+13h
Accuracy	0.82	0.80	0.78	0.83	0.80	0.78	0.68	0.69
Bias frequency	0.72	0.78	0.78	0.77	0.82	0.82	1.00	1.01
Brier skill score	0.42	0.30	0.23	0.43	0.29	0.22	-0.04	-0.09
OSLO	+4h	+13h	+48h	+4h	+13h	+48h	+4h	+13h
Accuracy	0.91	0.92	0.89	0.91	0.92	0.89	0.80	0.81
Bias frequency	0.91	1.00	0.99	0.88	0.99	0.97	0.99	0.98
Brier skill score	0.32	0.40	0.15	0.33	0.40	0.15	-0.53	-0.50
TRONDHEIM	+4h	+13h	+48h	+4h	+13h	+48h	+4h	+13h
Accuracy	0.84	0.84	0.84	0.84	0.84	0.84	0.75	0.77
Bias frequency	0.93	1.01	0.92	0.89	0.98	0.89	1.01	1.00
Brier skill score	0.07	0.01	0.09	0.09	0.04	0.08	-0.46	-0.36
TROMSØ	+4h	+13h	+48h	+4h	+13h	+48h	+4h	+13h
Accuracy	0.79	0.77	0.74	0.79	0.77	0.75	0.69	0.68
Bias frequency	0.68	0.66	0.61	0.74	0.72	0.67	1.00	1.00
Brier skill score	0.28	0.22	0.16	0.27	0.22	0.17	-0.06	-0.10
KRISTIANSAND	+4h	+13h	+48h	+4h	+13h	+48h	+4h	+13h
Accuracy	0.90	0.89	0.87	0.90	0.89	0.87	0.79	0.78
Bias frequency	1.08	1.02	0.88	1.08	1.02	0.88	1.01	1.00
Brier skill score	0.26	0.31	0.21	0.27	0.31	0.21	-0.48	-0.45
NESBYEN	+4h	+13h	+48h	+4h	+13h	+48h	+4h	+13h
Accuracy	0.92	0.93	0.90	0.92	0.93	0.90	0.86	0.85
Bias frequency	1.33	1.07	1.17	1.26	0.99	1.09	0.99	0.99
Brier skill score	0.10	0.26	-0.07	0.13	0.23	-0.03	-0.58	-0.56

6h accumulated precipitation

	AROME			fcfix			fcday	
BERGEN	+4-9h	+13-18h	+25-30h	+4-9h	+13-18h	+25-30h	+4-9h	+13-18h
Accuracy	0.84	0.85	0.83	0.84	0.85	0.84	0.67	0.68
Bias frequency	0.92	0.93	0.92	0.95	0.96	0.95	1.00	1.00
Brier skill score	0.67	0.69	0.65	0.68	0.69	0.66	0.31	0.33
OSLO	+4-9h	+13-18h	+25-30h	+4-9h	+13-18h	+25-30h	+4-9h	+13-18h
Accuracy	0.87	0.87	0.86	0.87	0.87	0.86	0.70	0.71
Bias frequency	0.91	0.91	0.93	0.89	0.89	0.93	1.00	1.00
Brier skill score	0.51	0.51	0.46	0.52	0.52	0.47	-0.14	-0.12
TRONDHEIM	+4-9h	+13-18h	+25-30h	+4-9h	+13-18h	+25-30h	+4-9h	+13-18h
Accuracy	0.82	0.82	0.82	0.83	0.82	0.82	0.66	0.68
Bias frequency	1.07	1.11	1.13	1.05	1.08	1.09	1.00	1.00
Brier skill score	0.51	0.46	0.48	0.51	0.47	0.48	0.05	0.07
TROMSØ	+4-9h	+13-18h	+25-30h	+4-9h	+13-18h	+25-30h	+4-9h	+13-18h
Accuracy	0.80	0.80	0.78	0.81	0.81	0.78	0.64	0.65
Bias frequency	0.85	0.84	0.83	0.88	0.88	0.86	1.00	1.00
Brier skill score	0.60	0.60	0.55	0.61	0.61	0.57	0.27	0.30
KRISTIANSAND	+4-9h	+13-18h	+25-30h	+4-9h	+13-18h	+25-30h	+4-9h	+13-18h
Accuracy	0.88	0.86	0.86	0.88	0.86	0.86	0.69	0.69
Bias frequency	1.06	1.07	1.04	1.06	1.07	1.04	1.00	1.00
Brier skill score	0.58	0.53	0.50	0.58	0.53	0.51	-0.08	-0.06
NESBYEN	+4h	+13h	+48h	+4h	+13h	+48h	+4h	+13h
Accuracy	0.89	0.88	0.87	0.89	0.88	0.87	0.73	0.74
Bias frequency	1.17	1.13	1.21	1.12	1.08	1.17	1.00	1.00
Brier skill score	0.47	0.41	0.33	0.47	0.43	0.33	-0.25	-0.28

Figure 4.10: Accuracy, bias frequency, and Brier skill score verification results (number format) for 1h and 6h accumulated precipitation at each location. Forecast lengths were +4h, +13h and +25/+48h, using forecast models AROME, fcfix and fcday (no +25h/+48h forecast). Data from 1. December 2019 - 31. April 2022.

the notion of Oslo, while relatively dry, has a clear extreme precipitation structure, in contrast to Tromsø where it tends to rain often but mostly with very low intensity.

Figure 4.11 and 4.12 also show the accuracy (red), bias frequency (blue) and BSS (yellow), but for 6h accumulated rainfall at each location. Overall, there is less accuracy variation between stations compared to 1h rainfall, ranging from 0.89 (Nesbyen) to 0.80 (Tromsø) for +4h AROME forecast, and the accuracy also seem to hold up slightly better with later forecast hours. This might be because having a 6h window reduces the variability somewhat when verifying the forecast, as each hourly value are of less importance compared to 1h verification where the model only has one attempt to get it right.

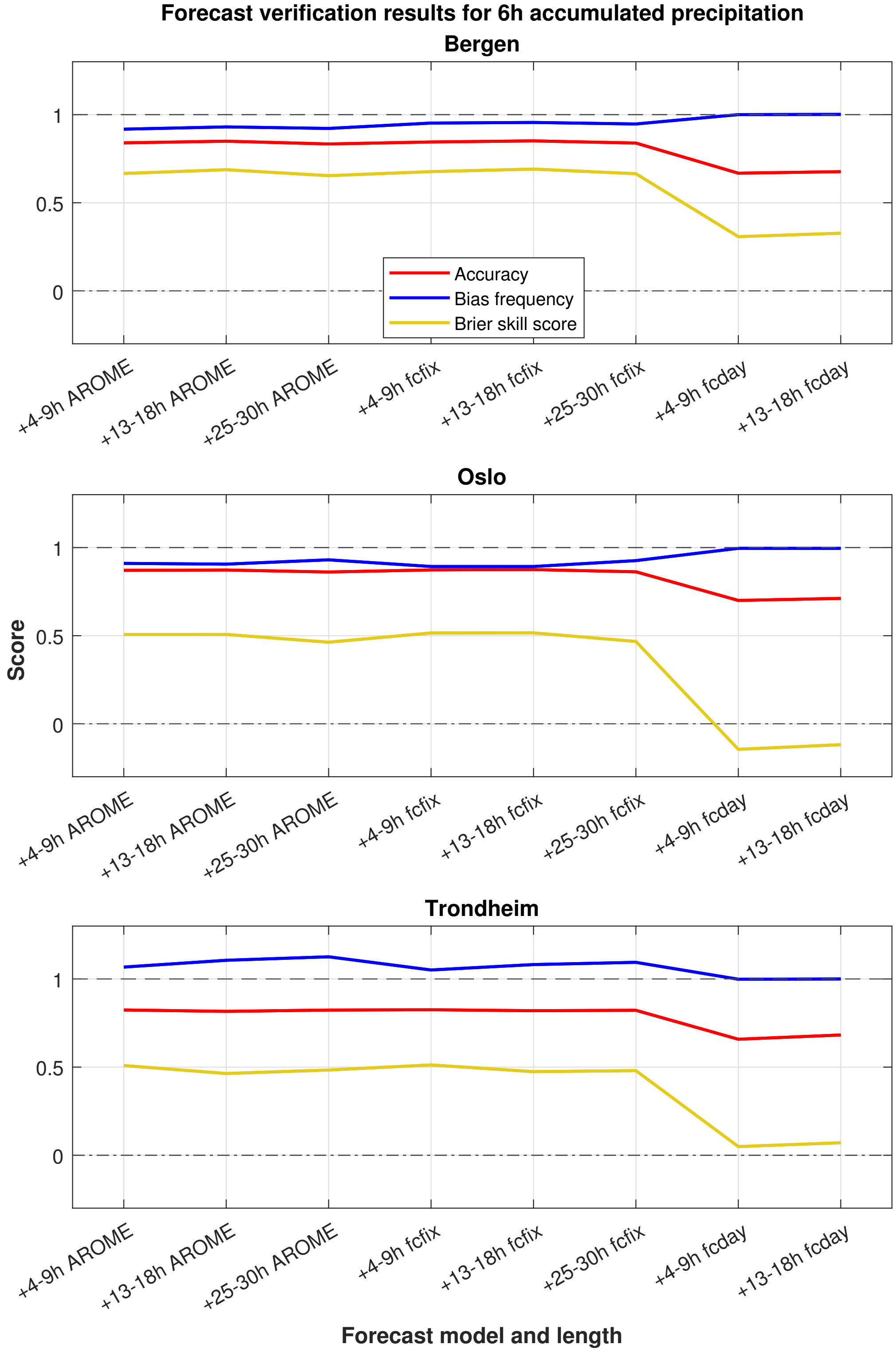


Figure 4.11: Accuracy, bias frequency and Brier skill score verification results for 6 hour accumulated precipitation at Bergen, Oslo and Trondheim. Forecast lengths were +4h, +13h and +25h, using forecast models AROME, fcfix and fcday (no +25h forecast). The ideal score for all parameters are 1, shown as a black dashed line. Brier skill scores below 0 (black dash-dotted line) signify that the forecast model quality is worse than the reference forecast based on climatology. Data from 1. December 2019 - 31. April 2022.

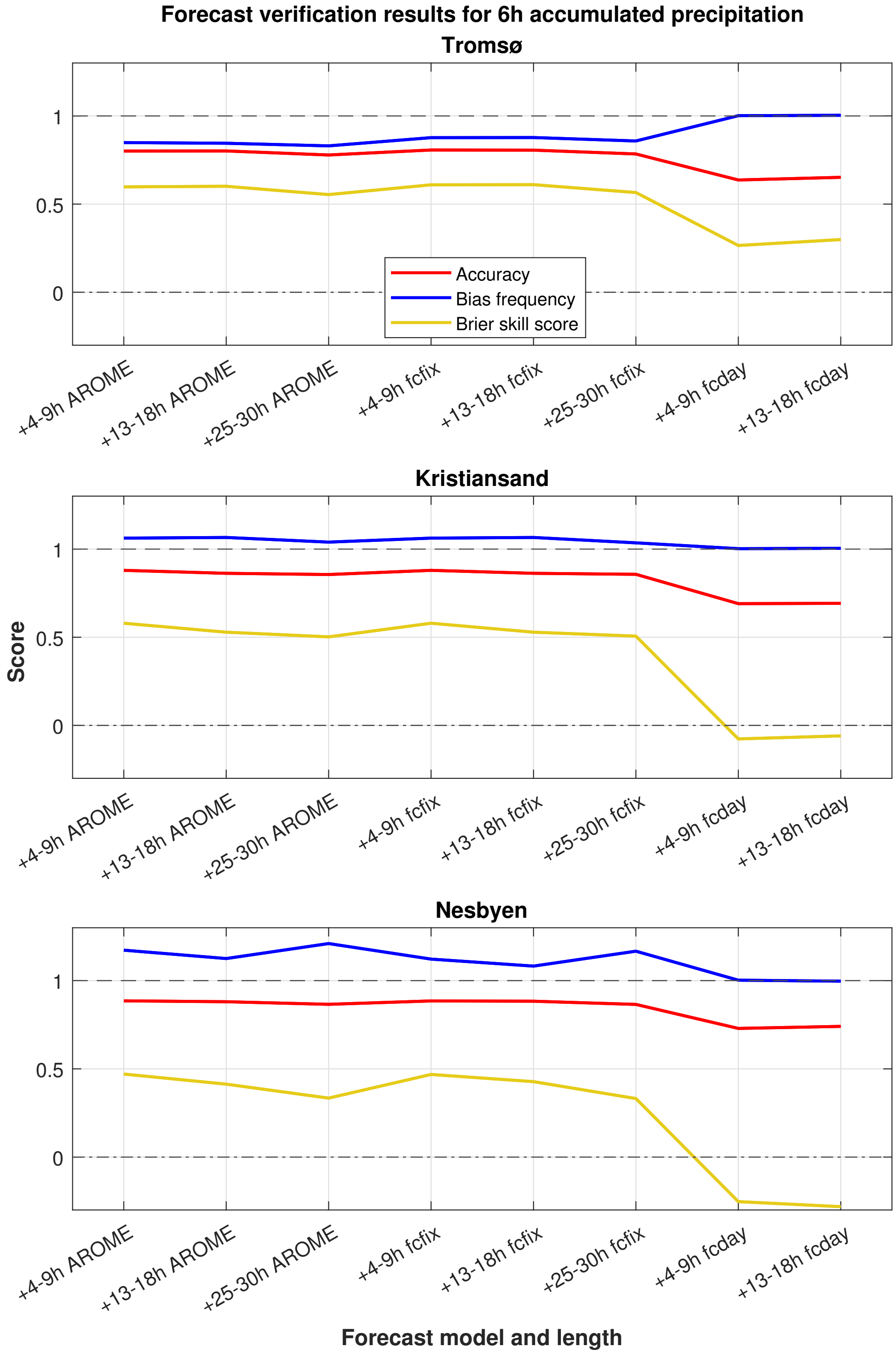


Figure 4.12: Accuracy, bias frequency and Brier skill score verification results for 6 hour accumulated precipitation at Tromsø, Kristiansand and Nesbyen. Forecast lengths were +4h, +13h and +25h, using forecast models AROME, fcfix and fcday (no +25h forecast). The ideal score for all parameters are 1, shown as a black dashed line. Brier skill scores below 0 (black dash-dotted line) signify that the forecast model quality is worse than the reference forecast based on climatology. Data from 1. December 2019 - 31. April 2022.

Furthermore, bias frequency also sees reduced variance across stations, i.e. values are closer to 1. Trondheim's overall negative bias for 1h precipitation has now changed to a slight positive bias. Bergen and Tromsø show considerably less negative bias; while number of misses and false alarms have not changed much, a sizeable chunk of correct negatives are now converted into hits, since it only needs to rain at least one of the six hours for that period to count as a rain event.

All in all, there are not a whole lot of differences for 6h verification results. Fcfix results are by all means identical to AROME, meaning adjusting the total mean precipitation essentially does nothing to improve AROME verification results. Increasing accumulation length does not help fcday the slightest either.

Brier skill scores on the other hand show a substantial increase across the board, however as the accuracy score highlights, this is not because the forecast in itself have improved much, but the background forecast (BS_{ref}) is now much easier to beat. When looking at Bergen and Tromsø, the distribution of observed rain/dry events for 6h accumulated rainfall is basically 50/50, thus it is like flipping a coin for the background forecast when it tries to predict the correct outcome. Even fcday with a BSS of around 0.3 has no issues beating it.

When studying 24h and 48h accumulation results (Figure A.25 and A.26), Bergen becomes the overall best performing station. This entails the highest accuracy score (0.89 for 24h and 0.92 for 48h), almost bias-free, very high hit rate and low false alarm ratio, best Brier score as well as a decent BSS. Kristiansand takes a close second place, helped among other things by a very respectable BSS (0.75 for 24h and 0.66 for 48h). Oslo and Nesbyen seem to perform the worst, although it should be stated none of the locations are showing particularly bad results.

In general, the longer the accumulation length is, the easier is it to forecast the correct outcome. For instance, roughly 75% of 48h accumulated forecasts in Bergen are hits, which after all only requires at least 0.1 mm rain in total for both data sets. And in cases of correct negatives (no rain in either forecast or observed data), the weather has to be stable enough for at least 48h straight to not precipitate, and the results show that AROME is generally quite good in these situations and avoids predicting any rainfall.

4.2.2 AROME and fcpersist forecast comparison

Figure 4.13 and 4.14 present the evolution of accuracy and Brier skill scores through the first 12 forecast hours for AROME and fcpersist forecasts. The purpose with these figures is pretty simple: Find how many hours into the forecast where the solid (AROME) and dashed (fcpersist) graphs cross each other. The result marks the turning point where AROME starts to outperform fcpersist.

To recap, fcpersist is the equivalent of watching the current weather outside and foresee it is going to stay exactly the same for the next 12 hours. And due to there being some slowness in the atmosphere, meaning it usually takes some time to switch from one weather condition to another, fcpersist is actually a really good forecast initially, but only for the first couple of hours before it falls off hard. As we already have seen with fcday, basing the future weather solely on a single forecast gets (not surprisingly) less and less accurate as time goes on.

After examining the figures and with the help from some interpolations, here are the number of hours for each location after which one should stop looking out the window and instead trust AROME to provide the best forecast.

- Bergen: 2.5 hours
- Oslo: 2 hours (2.5 for BSS, 1.5 for accuracy)
- Trondheim: 2 hours (although they follow each other closely until +4h)
- Tromsø: 3.5 hours
- Kristiansand: 1 hours (technically 0 hours since they are pretty much equal at +1h)
- Nesbyen: 1.5 hours

Another thing to notice is the modest score variability from hour to hour in the AROME forecast. As the figures (and verification data in general) in the previous subsection use very distinct forecast hours (+4h, +13h etc.), these figures show quite well how that might have an effect on whether the forecast is perceived as good or not, and therefore is something to be cautious about when analysing the results.

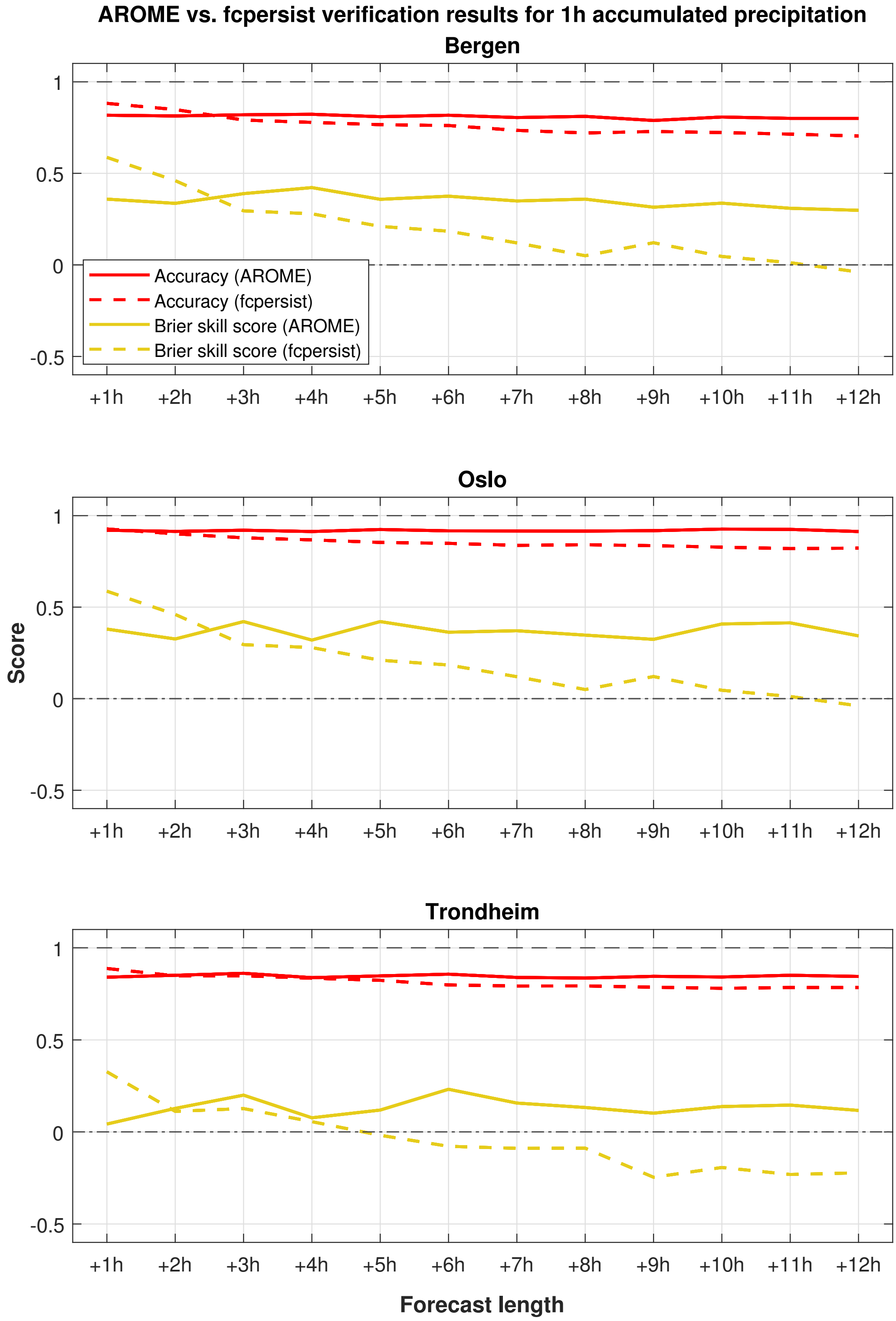


Figure 4.13: Accuracy and Brier skill score verification result comparison between AROME and fcpersist forecasts for 1 hour accumulated precipitation at Bergen, Oslo and Trondheim. The 12 first forecast hours were used for both forecasts. The ideal score for accuracy and Brier skill score is 1, shown as a black dashed line. Brier skill scores below 0 (black dash-dotted line) signify that the forecast model quality is worse than the reference forecast based on climatology. Data from 1. December 2019 - 31. April 2022.

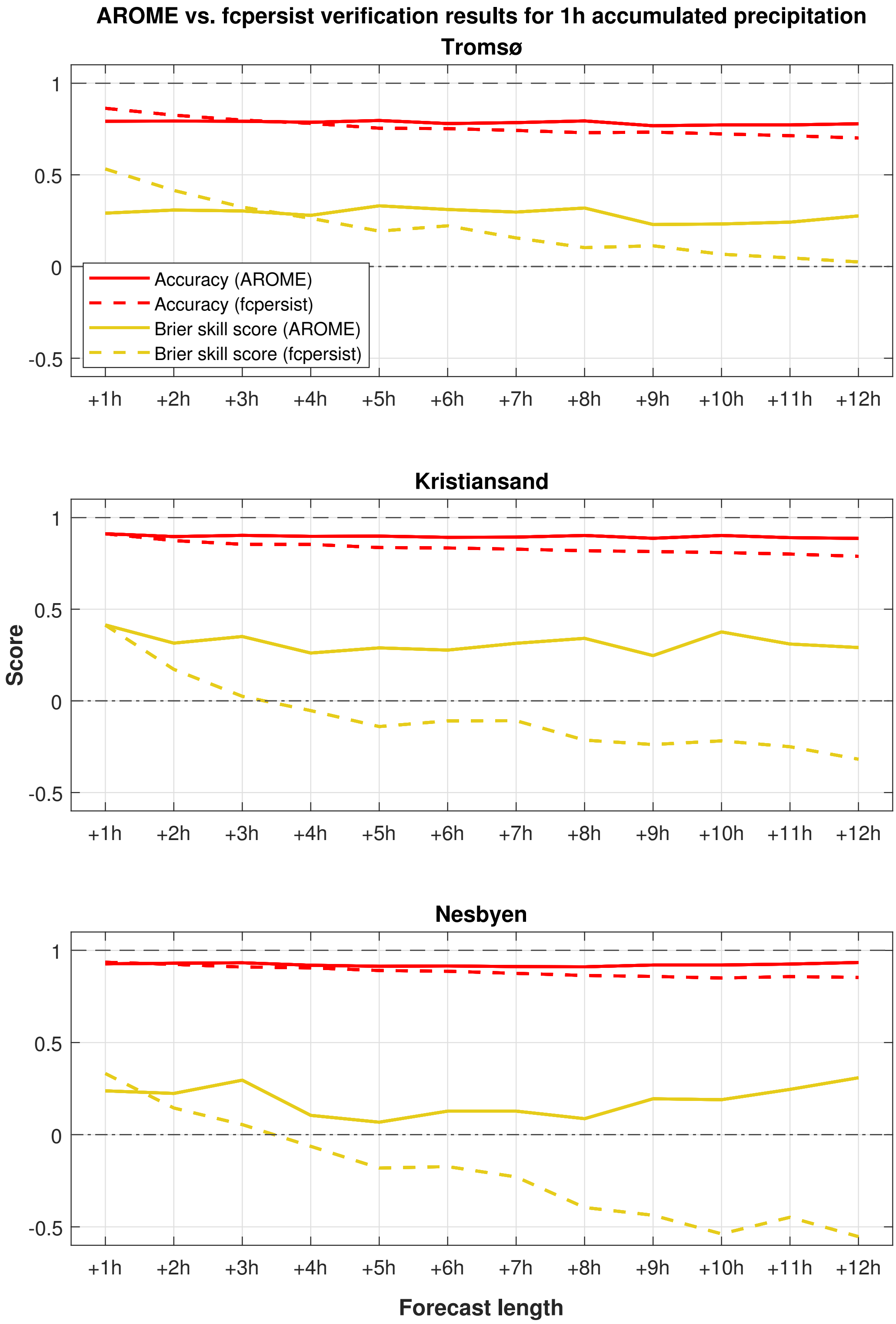


Figure 4.14: Accuracy and Brier skill score verification result comparison between AROME and fcpersist forecasts for 1 hour accumulated precipitation at Tromsø, Kristiansand and Nesbyen. The 12 first forecast hours were used for both forecasts. The ideal score for accuracy and Brier skill score is 1, shown as a black dashed line. Brier skill scores below 0 (black dash-dotted line) signify that the forecast model quality is worse than the reference forecast based on climatology. Data from 1. December 2019 - 31. April 2022.

Chapter 5

Discussion

5.1 Climatology

5.1.1 Dataset properties

All climatology results uses the first 12 forecast hours from AROME, because that is also the time span between each unique forecast starting point that was downloaded. This also means forecast hour +13-48h only appear for verification results. The model may take some time to "spin up" and generate the right atmospheric state, which could in theory affect the precipitation forecast the first couple of hours. Therefore, the total precipitation when using forecast hours +13-24h, +25-36h and +37-48h were also looked at.

No major deviation is found for any location, and the difference is mostly only a few percentage points, which is not enough to alter the results in any significant way. There also does not seem to be any clear precipitation amount trend for later forecast hours, and some locations for example even received the highest rainfall amount for middle forecast hours and a bit less for early and late hours.

The analysis period lasted 29 months, and while there is still a lot of data to process, it is not really a long period of time on a climatic time-scale. A question that could be asked is "Were these data representative for the climate at each location?" Did the datasets contain lots of unusual weather, events and similar that could have skewed the results in any way compared to what they would have been in different weather?

Honestly, it is not really meaningful to answer these questions, as the thesis's main objective is to verify AROME forecast *against* observations for this specific period. If the results here are poor, then chances are this would be the case anyway if we chose any other analysis period. If the model cannot reasonably handle all the various weather conditions we typically see in Norway, then it is simply not a good forecast. Another point is that AROME is updated and improved on a frequent basis, and starting the analysis period (given all data were available) several years earlier would involve data from older AROME versions. This would make it harder to objectively compare forecast results between earlier and recent years.

This thesis deals with point verification (as explained in Chapter 2.1), which only considers forecasts and observations for that specific place. Given some areas experience high local variation in weather, the results shown here does not really say anything about the forecast quality of the immediate surrounding areas. For instance, a forecasted extreme precipitation event might contain just the right amount of rainfall and appear at the correct time, but the forecast location was just a few kilometres off. This could produce some bad results since there was only a (small) spatial error, and the verification used here only considers temporal and quantity errors.

The addition of other variables like temperature, wind speed, wind direction and geopotential, as well as raw AROME forecast data could further help explaining the various forecast deviations, though that was outside the scope of this thesis. In that case, it would now be possible to determine in greater detail if the precipitation form (rain, snow, hail etc.) and wind speed/direction has any correlation with forecast quality, and perhaps identify certain conditions where AROME deviates the most from observed results. This is either way a potential topic for further research.

5.1.2 Seasonal weather duration

In Chapter 3.5.1 about weather persistence, there was a section explaining how rain/dry weather durations were counted when going from one season to another. The method used was to always treat the start of a new season as the beginning of a new period, no matter if the weather remained unchanged when switching into the new season. At first glance this seemed like an obvious weakness, and it was shown this approach did increase the number of total periods by 2-5% for daily data, and slightly lower the average rain/dry weather duration as a result.

As such, an alternative approach could be to only let the start date of a rain/dry weather period determine which season it belongs to, no matter how far into the next season it continues. However, this method could run into some issues with long spells of extremely stable weather. Theoretically, there could be a drought beginning in late February (considered a winter dry weather period) and extend far into spring. All these dry hours would still contribute to the average dry weather duration for spring, but not be considered as a spring dry weather period since the drought started in February. We could now end up with an extraordinary long winter dry weather period on paper even though most of it took place during spring, and few to none spring dry weather periods although these conditions had evidently been present.

To conclude, each method seems to have its advantages and disadvantages, but they both appear to have a potential weakness for stable weather (most applicable to Nesbyen) where the likelihood of a natural weather change right at a season boundary is very low, splitting that period in two.

Nevertheless, the general features shown in Figure 4.1 turn out as expected. Dry weather duration is highest during spring, (especially for the south-eastern stations), which from monthly data also sees the least amount of precipitation overall. Spring usually sees less low-pressure activity in the Atlantic Ocean due to lower latitudinal temperature gradient, and the air is still not warm enough for significant convective systems to form, leading to less overall rainfall in these months. Convective systems also depend on sufficient ground heating from the sun for hot air to rise up and form clouds, but the sun during spring is still fairly low on the horizon.

Nesbyen and Oslo are located in what is called the rainshadow of the major mountain ranges in Southern Norway. Any moist air coming from the west need to rise and pass over these first, and by doing so enough moisture is usually depleted from the air that it is no longer able to precipitate when arriving in the eastern part of Norway. Bergen, Trondheim and Tromsø on the other hand are all coastal cities facing the Atlantic, and are surrounded by mountains with potential for orographic enhancement.

5.1.3 Variability in AROME precipitation distribution

Bergen and Tromsø are the clear outliers when it comes to AROME forecast climatology. First, it does not predict enough rainfall with only 77% and 78% of observed rainfall. Interestingly, AROME forecasted 51 mm more total rainfall in Trondheim than Tromsø, yet the observations say the actual difference is over 919 mm in favour of Tromsø. The stark difference between Trondheim and Tromsø/Bergen is not immediately obvious, although Trondheim is situated a bit further inland where the surrounding (mountainous) coastline follows a southwest-northeast direction that might provide some extra rainshadow.

Second, AROME precipitation distribution shows way too many 0 mm hours, and way too few 0.1-1mm hours compared to what we see from observations. At the same time, *when* it rains, the percentile value ratios are honestly quite good, albeit a tad too high. Since the pool of AROME rain hours is much lower than the pool of observed hours, the 50th percentile value (median) could be ranked at number 2500/21000 for AROME and 3000/21000 for observed data. This works in favour of AROME for precipitation intensity ratio, because even with less rain overall, there are less rain hours to divide the precipitation between as well. This shows the under-prediction in total precipitation mostly happens independent of quantities.

AROME predicts less summer rainfall relative to other seasons, although it is hard to find a definitive pattern. For south-eastern stations, this is mostly due to too much winter/spring precipitation, not necessarily too little summer rain. One possible explanation could be AROME underestimating the rainshadow effect, and that less moisture crosses the mountains than anticipated. Tromsø struggles particularly with forecasting enough winter and spring precipitation, which normally falls as snow. This could indicate some issues with AROME winter moisture generation in the Arctic, like in situations where very cold air is heated above the relatively warm coastal sea, resulting in often short-lived convective winter precipitation. This could also explain why Nesbyen, which has an inland climate and thus does not experience these conditions, overestimates total winter/spring precipitation instead.

There are actually not that many hours from post-processed AROME forecast data that show exactly 0 mm, instead a fair share of them show trace amount of precipitation that does not pass the 0.1 mm/h threshold (this is observed in real life too). While these are included in the total precipitation calculations, they are nevertheless considered as dry hours elsewhere. It could seem that AROME precipitation distribution is too narrow when generating these hours with a trace of rainfall. In other words, not enough AROME forecast hours with very light rain passes the threshold value, while in reality the variation might be bigger and hence more values are likely to count as rain hours.

As mentioned, why this issue only seems to occur in Bergen/Tromsø and not anywhere else is unclear (Oslo shows minor signs of this too, but on a much smaller scale). Orographic effects not being fully resolved could be an option, where moisture advecting from the ocean is more affected by the local topography than first anticipated.

Another option is the air (when approaching a mountain range) not being forced upwards early enough due to the local high pressure anomaly that will form upstream. This would

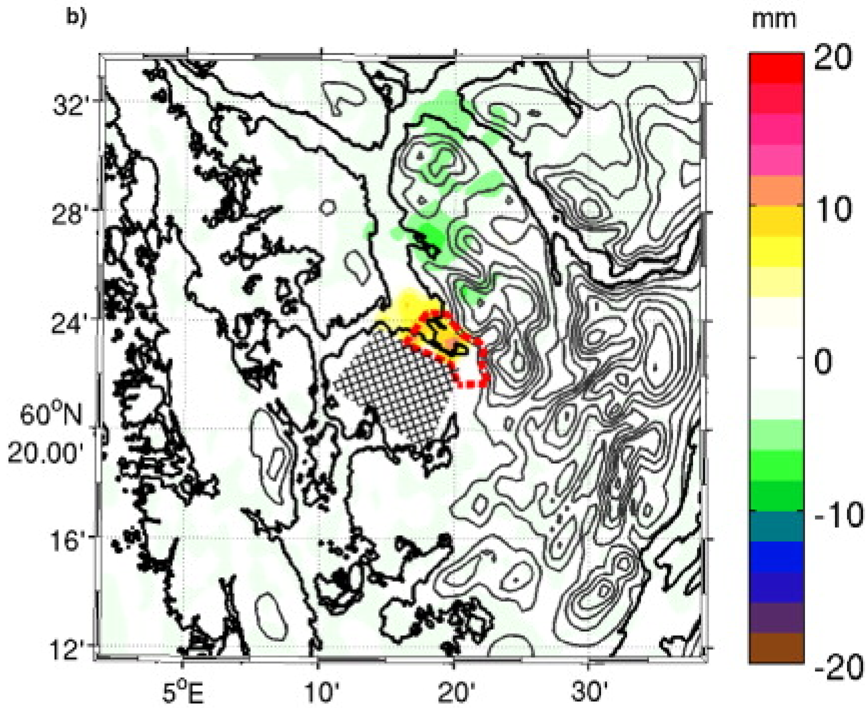


Figure 5.1: Net precipitation effect in Bergen city centre (red dotted area) by removing Løvstakken (black-gridded area) from the model. Numbers are control member (reality) minus no-mountain member, meaning positive values show a relative decrease in rainfall when removing Løvstakken.

shift the model precipitation field closer to the mountain than what the observations tell. Jonassen et al. (2013) showed that artificially removing Løvstakken (one of Bergen's seven mountain ranges, located west-southwest of the city centre) from their model would in fact increase the precipitation amount on the leeward-side. This suggests Løvstakken (with a summit of 477 m) creates a so-called spillover effect, where some of the extra precipitation generated by orographic enhancements on the windward side is carried over to the leeward-side (see Figure 5.1).

The purpose of fcfix is to check if removing the mean bias could improve AROME forecast quality, and for Bergen/Tromsø (where the fcfix ratio is highest), this does not seem to be the case. Fcfix is not able to do anything about the imbalance between dry and very light rain hours, and instead for Bergen adds way too many heavy rain and extreme hours and thus a too high mean extreme value. Having the same total precipitation amount as observed data, but barely any more rain hours than AROME, results as seen in too high percentile values.

Fcfix data for Bergen improves with daily results for precipitation distribution and percentiles, though some of this could be explained by the nature of mean values. The longer the time span is for each value, the better fcfix becomes. For total rainfall, it is a perfect match with observed results. For monthly data, it is still a clear improvement, but it cannot do (and is not designed to do) anything with the inherent monthly ratio variability of AROME. And for daily data, we see it make a really good precipitation distribution.

When considering the hourly trace rainfall from above, chances are high they will pass the 0.1 mm threshold when accumulated into 24 hours. Days that record less rainfall than 0.1 mm would very often consist of relatively stable weather that does not offer good opportunities for rainfall to form, which likely lowers the difficulty for AROME to predict the correct daily forecast. With hourly variability getting evened out, the increase in mean rainfall from fcfix becomes proportionally more important.

5.1.4 Underestimation of AROME extreme precipitation

As stated in the Results, the overall trend is the hourly forecast to observed ratio decreases as precipitation events get more extreme, i.e. less sharpness. This could indicate an issue with enough moisture generating and precipitating fast enough. The deviation is arguably highest for autumn months in Bergen, which again could be linked to an underestimation of orographic effects. AROME mean summer extreme values in Oslo and Trondheim on the other hand matches observations quite well, and these events are also fairly common with a return value below 100 hours which makes these results more reliable.

AROME also seems to improve a bit when the time interval is increased from hours to days (except for Bergen), for instance the downward ratio trend with more intense rainfall is mostly gone. Daily extremes are usually more variable, where not all 24 hours are just as intense as others, and it could very well be some dry hours in there as well. If AROME's resolution (the ability to distribute precipitation into different categories) is a bit too low overall, then increasing the time window could have the same effect as with trace precipitation where the average rainfall matters more, and where the hard-to-predict observed extreme peaks will be averaged out.

How distinct extreme precipitation values are for each location might explain some of the results in Figure 4.6. In other words, areas where top extremes are rare but intense, and not common but relatively moderate and thus blends inn with other medium rainfall hours. Tromsø is definitely the best example here of the latter category, with plenty of low-intensity rain hours but very few heavy rainfall hours. It is also the only Arctic location included with the coldest weather overall, and it is known from Clausius-Clapeyron's equation that colder air can hold less moisture than warmer air.

This should in theory make it harder to score hits for shared extreme hours if the boundary for what and what does not count as an extreme event is hard to distinguish between. Hence, only small variations in precipitation amount between forecast and observations could see the forecast value ranked at 100th place plummet down to 600th place for observed values and fall outside the top x% bracket. Tromsø basically sees no seasonal variation in extreme precipitation either, meaning the time window for when such an event may reasonably appear is spread out over the entire year, making it harder to pinpoint the exact hour of the extreme. These are all plausible reasons why it is not performing very well here.

5.2 Forecast verification

Overall forecast quality only sees a very small decrease for longer forecast hours, and some of it might be due to the inherent variability from hour to hour seen in Figure 4.13 and 4.14. How AROME would perform as a medium or long-range forecast would be interesting to witness, though the extra computational power needed for such a high-resolution grid would be immense.

Not surprisingly, it is the south-eastern stations that performs best for 1h accumulated precipitation, as less rain hours overall and relatively stable weather makes it easier to get hits and correct negatives. Kristiansand perform really well here given it is the second wettest place after all, with only a slightly worse accuracy and BSS than Oslo which only receives about 60% of the total rainfall. It also achieved the second highest hit rate (ability to forecast observed rain events), only behind Nesbyen.

Brier skill scores show the relative skill of a forecast compared to the climatology, and is directly related to the difficulty of forecasting at a location, in this case the ratio of rain and dry hours. Bergen earns the highest BSS (0.42) as it is a rather difficult place to forecast (high uncertainty), and although the climatology results are not particularly great, AROME delivers a reasonably good forecast verification result compared to the reference forecast. Tromsø's AROME forecast show about the same relative skill as Kristiansand, but this does not take into account how Kristiansand's climatology forecast is much harder to beat. Trondheim and Nesbyen's BSS scores are barely positive. For Nesbyen, this is sort of understandable consider how ”easy” it to get a high accuracy score there, whereas for Trondheim it is a rather poor result, as only Bergen and Tromsø got an easier BS_{ref} to beat.

Increased accumulation lengths (especially +24h and +48h) see an overall increase in accuracy and BSS, which again shows how hourly forecasts are by far the hardest to get right. Interestingly, Bergen, Tromsø and Trondheim are now among the top performers for accuracy, while Oslo and Nesbyen show a clear decrease. This has to do with how the ratio of rain/dry hours shift with longer accumulation length. Where dry places like Oslo and Nesbyen approach a 50/50 split (harder to get a high accuracy score and hit rate, and a low false alarm ratio), other stations start to see a clear majority of rain events. BSS takes this into account, and shows that AROME in Kristiansand delivers the most skilful forecast, with Tromsø in a definitive last place.

Fcday forecast was made as a low-quality challenger to AROME, and it is safe to say it has very little to contend with. It performs significantly worse than AROME in pretty much every verification metric, and shows that the weather in Norway is just too unstable to rely on yesterday's observations as the sole method to forecast the future weather. Much of the same can be said about fcpersist, however it also played an important role in quantifying after how long AROME becomes better than purely relying on current observations. Both fcday and fcpersist are based solely on observations, which is why they show no bias at all against observed data.

As with hourly climatology results, fcfix does not really change much regarding verification results, which shows the hourly forecast variability is far more important than removing the bias on mean precipitation.

5.3 Choosing the precipitation threshold

Depending on the precipitation distribution of your data, the chosen precipitation threshold value could be important for the results. This thesis uses a threshold value of 0.1 mm/h and 0.1 mm/day, but especially the daily limit is quite forgiving since 0.1 mm precipitation over 24 hours really is not much. For instance, thresholds like ≥ 0.2 mm/h and ≥ 1 mm/day are often used as well, the latter one is the definition of a wet day in WMO's guidelines on calculating climate normals (WMO, 2017). Increasing the threshold value adds an extra buffer and reduces some variability from the very light rain hours.

Tromsø would arguably be the most interesting location to look at if threshold values were increased to those above, since it experienced a very high amount of light rain hours and a very low fcfix ratio. To give an idea, Tromsø observed 6325 rain hours in total. 1807 of those (28.6%) were exactly 0.1 mm, while 1079 (17%) recorded 0.2 mm. This means close to half of all observed rain hours only recorded 0.2 mm or less. If the hourly precipitation threshold was $\geq$ 0.2 mm, the total number of rain hours would now be 2992 (AROME), 3376 (fcfix) and 4518 (observed), which is an AROME vs. observed rain hours ratio of 0.66. This is almost identical to the 0.1 mm/h ratio (0.65), showing that increasing the threshold would not have any significant impact here. If we do the same experiment for daily data with a new threshold of $\geq$ 1 mm, the new AROME vs. observed rain days ratio is 92.5%, slightly down from 93%.

Oslo also forecasts a bit too few rain hours as seen in Figure 4.3, but the new threshold values from above are now pretty close to 1, which could have some follow-up effects on percentiles. Either way, keeping the threshold at 0.1 mm allows us to better spot precipitation distribution anomalies, and provides more information about very light rain hours since less of them are "lost" as dry hours.

5.4 How trustworthy are observed data?

All precipitation observations have associated uncertainties with them (see Chapter 3.1.1), which makes it harder to assess the true forecast quality. Køltzow et al. (2020) found that wind-induced undercatch of solid precipitation in cold regions in Norway has a significant impact on verification results, and that the verification process ideally should be split between liquid and solid precipitation. A transfer function was applied to counteract the undercatch, and increase the observed measurements closer to the "true" value, although this process also brings its own uncertainties.

Since neither wind data nor temperature data is included in this thesis, it is not really possible to determine when it rains and when it snows. That said, an estimate can still be made by taking the general climate for each location into consideration. The most relevant data to look at is winter precipitation ratio between AROME and observed data (Figure A.10 in the Appendix), particularly in Tromsø, Oslo and Nesbyen.

As mentioned before, Tromsø forecasts way too little winter precipitation (ratio of 0.67), which gives the opposite effect of precipitation undercatch. This does not mean no undercatch is present here (if it is, the true ratio would be even lower), just that it is unlikely to be the main reason why AROME underperforms during winter. Oslo and Nesbyen looks more promising, as the winter ratios are among the highest for any season, and well above 1. Although neither of them are known for being very windy during winter months, it is definitely possible there could be some undercatch issues, but that remains as speculations for now.

Figure 5.2 is an example of a small section of observed hourly precipitation values in Tromsø. It shows two single low-intensity rain hours among lots of dry hours, and none of these hours contained any rainfall in the forecast. This peculiar pattern is found rather commonly in observed data, and it might look a bit strange for single hours with very light rain to appear seemingly out of nowhere before vanishing the next hour. This does not mean these observations are wrong, but it would be interesting to see if nearby stations on Tromsøya (e.g. at Langnes airport and Tromsø Holt) show a similar pattern. In fact, these rain hours do add up quite a bit over time, and are also the cause of many rain days.

6337	0
6338	0
6339	0
6340	0
6341	0
6342	0
6343	0
6344	0
6345	0.2000
6346	0
6347	0
6348	0
6349	0
6350	0
6351	0.1000
6352	0
6353	0
6354	0
6355	0
6356	0
6357	0
6358	0
6359	0
6360	0

Figure 5.2: A sequence of observed precipitation hours in Tromsø. It was not uncommon see one or two low-intensity rain hours right in the middle of longer dry weather periods.

Chapter 6

Conclusion

The goal of this thesis was to validate the AROME precipitation forecast at specific locations deemed to provide the best forecast value. Both post-processed AROME and observed precipitation data were downloaded, with an analysis period from 1. December 2019 to 31. April 2022, before being processed to produce a series of climatology and verification results. Based on these results and their discussion, here is a list ranking the overall AROME forecast quality from best to worst location, including their reasoning.

1. **Kristiansand -** Some really solid results for almost all parameters. Virtually no fcfix ratio deviation, and close to ideal precipitation distribution and percentile ratios for almost all values. Some monthly rainfall ratio disparity is present, but is still among the best performers. A bit too low top 0.1% mean extreme precipitation ratio overall (though this was the case everywhere) and during summer, but achieves one of the best results for shared extreme hours. It is also among the top accuracy and BSS scores performers for every accumulation length, which is really strong as the second wettest location. On par with fcpersist even after just one hour.

2. **Oslo -** Pretty good fcfix ratio with 1.08. Decent precipitation distribution ratio other than extreme values, though low percentile ratios are way too high and high percentiles ratios a bit too low. Reasonably good monthly ratio. Perhaps the best overall mean extreme precipitation ratio, and one of few locations that get the average summer extreme rainfall right. Third place for top 1% shared extreme hours, but wins the top 0.1% category. Solid verification results all around.

3. **Nesbyen -** Somewhat too high fcfix ratio, however it is also the driest location by far. Precipitation distribution ratio is rather variable, and with a downward

trend towards higher values. Decent percentile results, but with the same trend. Some very notable spikes for monthly ratio results, but this is mostly due to very low precipitation values. Too low mean values for the highest extremes (for all seasons), but pretty much spot on for daily mean extremes. Best hit rate of all stations for top 1% shared extreme hours. Highest accuracy for 1h accumulated precipitation, however BSS tells us the relative skill is not that high.

4. **Trondheim -** Fcfix ratio a bit too high (1.14). Forecasts too many mid-intensity rain hours but to few extremes, although percentiles ratios are quite good apart from 25th. Highly variable monthly ratio. Quite good overall mean extreme precipitation ratio, albeit with the same downward trend as the others. Arguably the best forecast for seasonal mean extreme ratios, but quite poor score for shared extreme hours. Mediocre 1h accuracy, and BSS shows it is barely better than reference forecast. Pretty decent results for longer accumulation lengths.

5. **Bergen -** Way too low fcfix ratio given how wet Bergen is, and generally too few forecasted rain hours overall (and too many dry hours). Very good percentile ratios, but because the low total precipitation amount is offset by the low total rain hours. Monthly ratio is generally irregular and too low. Also too low mean extreme precipitation ratios, both overall and seasonal. Average performance in shared extreme hours. Low accuracy and bias frequency, but fairly high BSS. Both accuracy and bias improves greatly with longer accumulation lengths.

6. **Tromsø -** Very low fcfix ratio, way too few rain hours and likewise way too many dry hours. Precipitation distribution ratio for heavy rain hours is highly variable (although low sample size). Quite good percentile ratios apart from 25th. Decent mean extreme ratio, but AROME adds more seasonal extreme variation than observed. By far the worst results for top 1% shared extreme hours, but decent for top 0.1%. Very low accuracy and bias frequency, BSS is not that good either, and verification results remain quite poor also for longer accumulation lengths.

As for future work, including other variables like wind and temperature data could help getting an even more detailed forecast validation for more locations, and better quality control of observed data. This might also give more definitive answers to how and why AROME seem to underperform in some areas. Another point is to cross-check Tromsø observed data with other nearby stations, and by its extension figure out why AROME is not able to produce enough winter precipitation here.

Bibliography

Guidelines on the Calculation of Climate Normals. WMO - No. 1203. World Meteorological Organization, 2017. ISBN: 978-92-63-11203-3.

Guide to Instruments and Methods of Observation. WMO - No. 8. World Meteorological Organization, 2018. ISBN: 978-92-63-10008-5.

C Donald Ahrens. *Essentials of meteorology: an invitation to the atmosphere*. Cengage Learning, 7th edition, 2014.

American Meteorological Society. Weighing rain gauge. URL `https://glossary.ametsoc.org/wiki/Weighing_rain_gauge`. Accessed: 13.10.2022.

Atmospheric Radiation Measurement user facility. Disdrometer. URL `https://www.arm.gov/capabilities/instruments/disdrometer`. Accessed: 02.11.2022.

Mathew Barlow, William J. Gutowski, John R. Gyakum, Richard W. Katz, Young-Kwon Lim, Russ S. Schumacher, Michael F. Wehner, Laurie Agel, Michael Bosilovich, Allison Collow, Alexander Gershunov, Richard Grotjahn, Ruby Leung, Shawn Milrad, and Seung-Ki Min. North American extreme precipitation events and related large-scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends. *Climate Dynamics*, 53:6835 – 6875, 2019. doi: 10.1007/s00382-019-04958-z.

Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525:47 – 55, 2015. ISSN 7567. doi: 10.1038/nature14956.

A. Bárdossy, J. Seidel, and A. El Hachem. The use of personal weather station observations to improve precipitation estimation and interpolation. *Hydrology and Earth System Sciences*, 25(2):583–601, 2021. doi: 10.5194/hess-25-583-2021.

Collaboration for Australian Weather and Climate Research. Forecast Verification methods Across Time and Space Scales , 2015. URL `https://www.cawcr.gov.au/projects/verification/`. Accessed: 17.11.2022.

Inger-Lise Frogner, Andrew T. Singleton, Morten Ø. Køltzow, and Ulf Andrae. Convection-permitting ensembles: Challenges related to their design and use. *Quarterly Journal of the Royal Meteorological Society*, 145(S1):90–106, 2019. doi: 10.1002/qj.3525.

Eric Gilleland, David Ahijevych, Barbara G. Brown, Barbara Casati, and Elizabeth E. Ebert. Intercomparison of Spatial Forecast Verification Methods. *Weather and Forecasting*, 24(5):1416 – 1430, 2009. doi: 10.1175/2009WAF2222269.1.

Eric Gilleland, David A. Ahijevych, Barbara G. Brown, and Elizabeth E. Ebert. Verifying Forecasts Spatially. *Bulletin of the American Meteorological Society*, 91(10):1365 – 1376, 2010. doi: 10.1175/2010BAMS2819.1.

Pavel Ya. Groisman and David R. Legates. The Accuracy of United States Precipitation Data. *Bulletin of the American Meteorological Society*, 75(2):215 – 228, 1994. doi: 10.1175/1520-0477(1994)075⟨0215:TAOUSP⟩2.0.CO;2.

Inger Hanssen-Bauer, H Drange, EJ Førland, LA Roald, KY Børsheim, H Hisdal, D Lawrence, A Nesje, S Sandven, A Sorteberg, et al. Klima i Norge 2100. *Bakgrunnsmateriale til NOU Klimatilpassing., Norsk klimasenter, Oslo, Norway*, 2015.

Mariken Homleid, Gunnar Noer, Frank Thomas Tveter, and Lene Østvand. Verification of Operational Weather Prediction Models December 2020 to February 2021, 2021. URL `https://www.met.no/publikasjoner/met-info/met-info-2021/_/attachment/download/b07d2dd5-ed1e-45cc-90c7-04c36aaf33ad:2530a42612dcdda5ba916e17f1b8a2c758160ffe/Verification_report_202012-202102.pdf`.

Marius O. Jonassen, Haraldur Ólafsson, Aslaug S. Valved, Joachim Reuder, and Jan A. Olseth. Simulations of the Bergen orographic wind shelter. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1):19206, 2013. doi: 10.3402/tellusa.v65i0.19206.

Gopinath Kathiravelu, Terry Lucke, and Peter Nichols. Rain Drop Measurement Techniques: A Review. *Water*, 8(1), 2016. ISSN 2073-4441. doi: 10.3390/w8010029.

Jørn Kristiansen and Dick Blaauboer. Joining forces for the best short-range weather forecasts, 2018. URL `https://www.emetsoc.org/uwc-short-range-weather-forecasts/`. Accessed: 07.10.2022.

Morten Køltzow, Barbara Casati, Thomas Haiden, and Teresa Valkonen. Verification of Solid Precipitation Forecasts from Numerical Weather Prediction Models in Norway. *Weather and Forecasting*, 35(6):2279 – 2292, 2020. doi: 10.1175/WAF-D-20-0060.1.

MET Norway. MetCoOp. URL `https://www.met.no/en/projects/metcoop/`. Accessed: 07.10.2022.

MET Norway. Private værstasjoner gir bedre temperaturvarsler på Yr, 2018. URL `https://www.met.no/nyhetsarkiv/private-vaerstasjoner-gir-bedre-temperaturvarsler-pa-yr`. Accessed: 02.11.2022.

MET Norway NWP Wiki. NWPdocs/wiki. URL `https://github.com/metno/NWPdocs/wiki`. Accessed: 07.10.2022.

Allan H. Murphy. What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. *Weather and Forecasting*, 8(2):281 – 293, 1993. doi: 10.1175/1520-0434(1993)008⟨0281:WIAGFA⟩2.0.CO;2.

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*, 32(2):609 – 627, 2017. doi: 10.1175/WAF-D-16-0099.1.

Thomas N. Nipen, Ivar A. Seierstad, Cristian Lussana, Jørn Kristiansen, and Øystein Hov. Adopting Citizen Observations in Operational Weather Prediction. *Bulletin of the American Meteorological Society*, 101(1):E43 – E57, 2020. doi: 10.1175/BAMS-D-18-0237.1.

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, 1998. In Conf. on Cloud Physics, pages 217–220. Amer. Meteor. Soc Everett, WA.

John P. Rafferty. Front, 2012. URL `https://www.britannica.com/science/front-meteorology`. Accessed: 17.11.2022.

Roy M. Rasmussen, John Hallett, Rick Purcell, Scott D. Landolt, and Jeff Cole. The Hotplate Precipitation Gauge. *Journal of Atmospheric and Oceanic Technology*, 28(2):148 – 164, 2011. doi: 10.1175/2010JTECHA1375.1.

Ankita Sahay. Precipitation: Introduction, Types, Examples, Various Factors, 2022. URL `https://www.embibe.com/exams/precipitation/`. Accessed: 17.11.2022.

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):2563–2586, 2002. doi: 10.1256/qj.01.128.

Y. Seity, P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson. The AROME-France Convective-Scale Operational Model. *Monthly Weather Review*, 139(3):976 – 991, 2011. doi: 10.1175/2010MWR3425.1.

Asgeir Sorteberg. Flere dager med nedbør, 2012. URL `https://folk.uib.no/gbsag/Publications/Sorteberg_tograder_2012.pdf`. Accessed: 27.10.2022.

Store Norske Leksikon. Automatisk nedbørsmåler, a. URL `https://snl.no/automatisk_nedb%C3%B8rm%C3%A5ler`. Accessed: 13.10.2022.

Store Norske Leksikon. Manuell nedbørsmåler, b. URL `https://snl.no/manuell_nedb%C3%B8rm%C3%A5ler`. Accessed: 13.10.2022.

Ellen Stuart-Haëntjens. How farmers on the Great Plains are changing the local climate, 2018. URL `https://massivesci.com/articles/farms-climate-change-crop-rotation-rain-fallow/`. Accessed: 17.11.2022.

Ingvild Thynes Villa. High resolution forecasts of short duration extreme precipitation - forecasting quality and sensitivity to parameterization for two Norwegian cases. Master's thesis, The University of Bergen, 2021.

Appendix A

Data

BERGEN	Total	Winter	Spring	Summer	Autumn
Total rain hours	6240	2496	1206	794	1744
Total dry hours	14928	4007	4675	3622	2624
Total NaN hours	0	0	0	0	0
Total rain weather periods	1409	531	322	241	317
Total dry weather periods	1409	531	323	243	319
Avg. rain weather duration in hours	4.429	4.701	3.745	3.295	5.502
Avg. dry weather duration in hours	10.595	7.546	14.474	14.905	8.226

OSLO	Total	Winter	Spring	Summer	Autumn
Total rain hours	2757	1074	495	510	678
Total dry hours	18411	5429	5386	3906	3690
Total NaN hours	0	0	0	0	0
Total rain weather periods	736	248	143	197	150
Total dry weather periods	737	249	145	199	151
Avg. rain weather duration in hours	3.746	4.331	3.462	2.589	4.520
Avg. dry weather duration in hours	24.981	21.803	37.145	19.628	24.437

TRONDHEIM	Total	Winter	Spring	Summer	Autumn
Total rain hours	3668	1221	930	530	987
Total dry hours	17008	5282	4483	3886	3357
Total NaN hours	492	0	468	0	24
Total rain weather periods	1152	358	337	179	279
Total dry weather periods	1156	359	342	181	281
Avg. rain weather duration in hours	3.184	3.411	2.760	2.961	3.538
Avg. dry weather duration in hours	14.713	14.713	13.108	21.470	11.947

TROMSØ	Total	Winter	Spring	Summer	Autumn
Total rain hours	6401	2158	2144	893	1206
Total dry hours	14755	4333	3737	3523	3162
Total NaN hours	12	12	0	0	0
Total rain weather periods	1433	425	458	269	284
Total dry weather periods	1431	423	459	269	284
Avg. rain weather duration in hours	4.467	5.078	4.681	3.320	4.246
Avg. dry weather duration in hours	10.311	10.243	8.142	13.097	11.134

KRISTIANSAND	Total	Winter	Spring	Summer	Autumn
Total rain hours	3101	1521	485	422	673
Total dry hours	17419	4982	5372	3898	3167
Total NaN hours	648	0	24	96	528
Total rain weather periods	857	382	130	160	186
Total dry weather periods	860	384	132	163	189
Avg. rain weather duration in hours	3.618	3.982	3.731	2.638	3.618
Avg. dry weather duration in hours	20.255	12.974	40.697	23.914	16.757

NESBYEN	Total	Winter	Spring	Summer	Autumn
Total rain hours	2019	755	287	429	548
Total dry hours	19149	5748	5594	3987	3820
Total NaN hours	0	0	0	0	0
Total rain weather periods	648	249	94	155	151
Total dry weather periods	649	251	96	157	153
Avg. rain weather duration in hours	3.116	3.032	3.053	2.768	3.629
Avg. dry weather duration in hours	29.505	22.900	58.271	25.395	24.967

Figure A.1: Total rain/dry weather periods as well as average rain/dry weather duration in hours for observed data at each location.

BERGEN	Total	Winter	Spring	Summer	Autumn
Total rain days	614	208	147	111	148
Total dry days	268	63	98	73	34
Total NaN days	0	0	0	0	0
Total rain weather periods	112	27	42	28	17
Total dry weather periods	112	27	42	30	17
Avg. rain weather duration in days	5.482	7.704	3.500	3.964	8.706
Avg. dry weather duration in days	2.393	2.333	2.333	2.433	2.000

OSLO	Total	Winter	Spring	Summer	Autumn
Total rain days	422	141	85	97	99
Total dry days	460	130	160	87	83
Total NaN days	0	0	0	0	0
Total rain weather periods	146	45	39	34	30
Total dry weather periods	147	46	40	35	30
Avg. rain weather duration in days	2.890	3.133	2.179	2.853	3.300
Avg. dry weather duration in days	3.129	2.826	4.000	2.486	2.767

TRONDHEIM	Total	Winter	Spring	Summer	Autumn
Total rain days	510	160	131	96	123
Total dry days	350	111	94	88	57
Total NaN days	22	0	20	0	2
Total rain weather periods	126	40	35	33	22
Total dry weather periods	127	38	37	34	21
Avg. rain weather duration in days	4.048	4.000	3.743	2.909	5.591
Avg. dry weather duration in days	2.756	2.921	2.541	2.588	2.714

TROMSØ	Total	Winter	Spring	Summer	Autumn
Total rain days	637	191	188	128	130
Total dry days	244	79	57	56	52
Total NaN days	1	1	0	0	0
Total rain weather periods	107	34	30	26	23
Total dry weather periods	105	30	28	26	22
Avg. rain weather duration in days	5.953	5.618	6.267	4.923	5.652
Avg. dry weather duration in days	2.324	2.633	2.036	2.154	2.364

KRISTIANSAND	Total	Winter	Spring	Summer	Autumn
Total rain days	422	173	76	76	97
Total dry days	432	98	168	103	63
Total NaN days	28	0	1	5	22
Total rain weather periods	135	40	33	35	29
Total dry weather periods	134	41	33	37	27
Avg. rain weather duration in days	3.126	4.325	2.303	2.171	3.345
Avg. dry weather duration in days	3.224	2.390	5.091	2.784	2.333

NESBYEN	Total	Winter	Spring	Summer	Autumn
Total rain days	367	127	56	91	93
Total dry days	515	144	189	93	89
Total NaN days	0	0	0	0	0
Total rain weather periods	141	43	32	35	33
Total dry weather periods	142	45	33	36	33
Avg. rain weather duration in days	2.602	2.953	1.750	2.600	2.818
Avg. dry weather duration in days	3.627	3.200	5.727	2.583	2.697

Figure A.2: Total rain/dry weather periods as well as average rain/dry weather duration in days for observed data at each location.

	BERGEN		
	Forecast	Observed	fcfix
0 mm	16397	14770	16073
0.1-1 mm	2979	4273	2974
1-2 mm	838	991	904
2-3 mm	372	424	423
3-4 mm	177	228	237
4-5 mm	98	129	137
5-6 mm	41	69	91
6-7 mm	24	23	39
7-8 mm	8	17	32
>8 mm	6	16	30
# NaN values	228	228	228
Total hours	21168	21168	21168

	TROMSØ		
	Forecast	Observed	fcfix
0 mm	16788	14591	16389
0.1-1 mm	3531	5607	3669
1-2 mm	456	563	605
2-3 mm	104	107	166
3-4 mm	24	32	57
4-5 mm	8	5	16
5-6 mm	1	6	8
6-7 mm	2	2	2
7-8 mm	1	3	0
>8 mm	1	0	4
# NaN values	252	252	252
Total hours	21168	21168	21168

	OSLO		
	Forecast	Observed	fcfix
0 mm	18424	18218	18492
0.1-1 mm	1890	2173	1881
1-2 mm	429	355	400
2-3 mm	121	119	108
3-4 mm	45	34	34
4-5 mm	13	14	10
5-6 mm	4	9	3
6-7 mm	4	5	5
7-8 mm	3	3	1
>8 mm	7	10	6
# NaN values	228	228	228
Total hours	21168	21168	21168

	KRISTIANSAND		
	Forecast	Observed	fcfix
0 mm	17203	17251	17212
0.1-1 mm	2120	2077	2121
1-2 mm	572	519	566
2-3 mm	221	256	220
3-4 mm	100	115	99
4-5 mm	50	47	49
5-6 mm	27	29	28
6-7 mm	20	14	19
7-8 mm	10	8	9
>8 mm	5	12	5
# NaN values	840	840	840
Total hours	21168	21168	21168

	TRONDHEIM		
	Forecast	Observed	fcfix
0 mm	16918	16823	17067
0.1-1 mm	2838	3026	2818
1-2 mm	489	441	410
2-3 mm	117	90	96
3-4 mm	54	35	37
4-5 mm	19	18	12
5-6 mm	6	6	4
6-7 mm	4	4	2
7-8 mm	1	2	1
>8 mm	2	3	1
# NaN values	720	720	720
Total hours	21168	21168	21168

	NESBYEN		
	Forecast	Observed	fcfix
0 mm	18720	18945	18887
0.1-1 mm	1928	1752	1841
1-2 mm	220	150	164
2-3 mm	45	57	26
3-4 mm	10	16	11
4-5 mm	7	6	5
5-6 mm	4	3	3
6-7 mm	3	5	1
7-8 mm	1	3	1
>8 mm	2	3	1
# NaN values	228	228	228
Total hours	21168	21168	21168

Figure A.3: Distribution of hourly precipitation amount for AROME, observed and fcfix data at each location.

	BERGEN			TROMSØ		
BERGEN	Forecast	Observed	fcfix	Forecast	Observed	fcfix
0 mm	279	264	272	286	242	279
0.1-5 mm	309	275	275	398	397	357
5-10 mm	107	97	104	127	121	139
10-15 mm	64	76	73	30	63	52
15-20 mm	40	61	37	15	28	18
20-25 mm	30	33	28	5	8	11
25-30 mm	11	13	32	1	3	4
30-35 mm	8	15	13	0	1	2
35-40 mm	4	11	10	1	0	0
40-45 mm	6	6	4	0	0	1
45-50 mm	4	2	2	0	0	0
>50 mm	3	12	15	0	0	0
# NaN values	17	17	17	19	19	19
Total days	882	882	882	882	882	882

	OSLO			KRISTIANSAND		
OSLO	Forecast	Observed	fcfix	Forecast	Observed	fcfix
0 mm	488	453	491	404	423	404
0.1-5 mm	242	297	244	264	236	264
5-10 mm	67	56	70	63	79	64
10-15 mm	34	29	32	40	29	40
15-20 mm	16	9	13	26	20	25
20-25 mm	5	10	5	12	14	13
25-30 mm	4	7	4	8	18	7
30-35 mm	6	2	4	10	11	10
35-40 mm	2	1	1	5	5	5
40-45 mm	1	1	1	5	1	5
45-50 mm	0	0	0	2	3	2
>50 mm	0	0	0	0	0	0
# NaN values	17	17	17	43	43	43
Total days	882	882	882	882	882	882

	TRONDHEIM			NESBYEN		
TRONDHEIM	Forecast	Observed	fcfix	Forecast	Observed	fcfix
0 mm	324	343	330	479	507	489
0.1-5 mm	358	354	373	307	295	312
5-10 mm	98	91	88	53	36	48
10-15 mm	38	33	30	16	15	11
15-20 mm	11	12	12	5	5	3
20-25 mm	5	2	2	3	5	1
25-30 mm	2	3	5	1	1	0
30-35 mm	4	1	2	0	1	1
35-40 mm	2	3	0	0	0	0
40-45 mm	0	1	0	1	0	0
45-50 mm	0	0	0	0	0	0
>50 mm	1	0	1	0	0	0
# NaN values	39	39	39	17	17	17
Total days	882	882	882	882	882	882

Figure A.4: Distribution of daily precipitation amount for AROME, observed and fcfix data at each location.

BERGEN	Mean	Precip intensity	25 percentile	50 percentile	75 percentile	90 percentile	99 percentile	# of rain hours
Forecasted data	0.235	1.059	0.247	0.582	1.392	2.641	5.773	4543
Observed data	0.303	1.027	0.200	0.500	1.300	2.700	5.900	6170
Ratio	0.777	1.031	1.234	1.164	1.071	0.978	0.979	0.736
fcfix	0.303	1.279	0.263	0.669	1.684	3.249	7.361	4867

OSLO	Mean	Precip intensity	25 percentile	50 percentile	75 percentile	90 percentile	99 percentile	# of rain hours
Forecasted data	0.099	0.796	0.224	0.472	0.991	1.796	4.358	2516
Observed data	0.091	0.703	0.100	0.300	0.800	1.700	5.100	2722
Ratio	1.084	1.132	2.240	1.574	1.239	1.056	0.855	0.924
fcfix	0.091	0.752	0.219	0.456	0.937	1.679	4.121	2448

TRONDHEIM	Mean	Precip intensity	25 percentile	50 percentile	75 percentile	90 percentile	99 percentile	# of rain hours
Forecasted data	0.121	0.673	0.207	0.385	0.832	1.523	3.951	3530
Observed data	0.106	0.599	0.100	0.300	0.700	1.400	3.900	3625
Ratio	1.143	1.123	2.070	1.282	1.188	1.088	1.013	0.974
fcfix	0.106	0.610	0.197	0.357	0.748	1.359	3.478	3381

TROMSØ	Mean	Precip intensity	25 percentile	50 percentile	75 percentile	90 percentile	99 percentile	# of rain hours
Forecasted data	0.116	0.557	0.190	0.350	0.682	1.256	2.896	4128
Observed data	0.148	0.489	0.100	0.300	0.600	1.100	2.800	6325
Ratio	0.787	1.141	1.900	1.167	1.136	1.142	1.034	0.653
fcfix	0.148	0.656	0.199	0.395	0.792	1.516	3.608	4527

KRISTIANSAND	Mean	Precip intensity	25 percentile	50 percentile	75 percentile	90 percentile	99 percentile	# of rain hours
Forecasted data	0.159	1.013	0.246	0.546	1.285	2.442	6.090	3125
Observed data	0.158	1.046	0.200	0.600	1.400	2.600	6.400	3077
Ratio	1.007	0.968	1.232	0.910	0.918	0.939	0.951	1.016
fcfix	0.158	1.009	0.247	0.544	1.277	2.427	6.048	3116

NESBYEN	Mean	Precip intensity	25 percentile	50 percentile	75 percentile	90 percentile	99 percentile	# of rain hours
Forecasted data	0.061	0.544	0.183	0.316	0.641	1.152	3.647	2220
Observed data	0.051	0.531	0.100	0.200	0.600	1.200	4.100	1995
Ratio	1.214	1.024	1.834	1.581	1.068	0.960	0.890	1.113
fcfix	0.051	0.477	0.170	0.284	0.559	1.019	3.141	2053

Figure A.5: Mean, precipitation intensity and various percentiles for hourly AROME, observed and fcfix data at each location. Ratio is forecast divided by observed data.

BERGEN	Mean	Precip intensity	25 percentile	50 percentile	75 percentile	90 percentile	99 percentile	# of rain days
Forecasted data	5.644	8.323	1.439	4.475	11.480	21.137	49.599	586
Observed data	7.284	10.484	1.500	6.000	15.200	24.700	54.500	601
Ratio	0.775	0.794	0.960	0.746	0.755	0.856	0.910	0.975
fcfix	7.262	10.584	1.760	5.712	14.532	27.114	63.818	593

OSLO	Mean	Precip intensity	25 percentile	50 percentile	75 percentile	90 percentile	99 percentile	# of rain days
Forecasted data	2.380	5.466	0.711	2.858	7.245	14.080	35.843	377
Observed data	2.206	4.631	0.400	1.800	6.100	13.600	31.000	412
Ratio	1.079	1.180	1.777	1.588	1.188	1.035	1.156	0.915
fcfix	2.195	5.081	0.690	2.697	6.756	12.987	33.061	374

TRONDHEIM	Mean	Precip intensity	25 percentile	50 percentile	75 percentile	90 percentile	99 percentile	# of rain days
Forecasted data	2.914	4.705	0.929	2.705	6.250	11.489	31.911	519
Observed data	2.553	4.304	0.600	2.500	5.900	10.600	31.600	500
Ratio	1.141	1.093	1.549	1.082	1.059	1.084	1.010	1.038
fcfix	2.548	4.162	0.885	2.450	5.601	10.047	27.906	513

TROMSØ	Mean	Precip intensity	25 percentile	50 percentile	75 percentile	90 percentile	99 percentile	# of rain days
Forecasted data	2.790	4.174	0.947	2.669	5.756	9.738	21.321	577
Observed data	3.554	4.939	0.900	2.700	7.400	12.800	23.900	621
Ratio	0.785	0.845	1.053	0.989	0.778	0.761	0.892	0.929
fcfix	3.547	5.244	1.162	3.299	7.283	12.130	27.104	584

KRISTIANSAND	Mean	Precip intensity	25 percentile	50 percentile	75 percentile	90 percentile	99 percentile	# of rain days
Forecasted data	3.828	7.381	1.041	3.596	10.022	19.963	42.146	435
Observed data	3.804	7.672	0.900	3.500	9.900	22.500	44.000	416
Ratio	1.006	0.962	1.157	1.027	1.012	0.887	0.958	1.046
fcfix	3.802	7.330	1.034	3.571	9.953	19.826	41.857	435

NESBYEN	Mean	Precip intensity	25 percentile	50 percentile	75 percentile	90 percentile	99 percentile	# of rain days
Forecasted data	1.475	3.294	0.599	1.746	4.339	8.308	23.763	386
Observed data	1.212	2.928	0.300	1.300	3.400	7.200	24.700	358
Ratio	1.218	1.125	1.997	1.343	1.276	1.154	0.962	1.078
fcfix	1.215	2.783	0.531	1.473	3.648	6.902	19.570	376

Figure A.6: Mean, precipitation intensity and various percentiles for daily AROME, observed and fcfix data at each location. Ratio is forecast divided by observed data.

	BERGEN				OSLO			
	Forecasted data	Observed data	Ratio	fcfix	Forecasted data	Observed data	Ratio	fcfix
'Dec-2019'	281.244	368.700	0.763	361.870	85.402	57.000	1.498	78.772
'Jan-2020'	340.047	470.700	0.722	437.531	92.135	75.600	1.219	84.983
'Feb-2020'	328.154	330.900	0.992	422.229	65.768	38.900	1.691	60.663
'Mar-2020'	210.908	257.400	0.819	271.371	49.153	39.100	1.257	45.337
'Apr-2020'	132.134	172.300	0.767	170.014	36.069	35.000	1.031	33.269
'May-2020'	73.998	114.500	0.646	95.212	48.953	42.800	1.144	45.153
'Jun-2020'	44.188	46.300	0.954	56.856	135.029	119.400	1.131	124.547
'Jul-2020'	168.576	222.400	0.758	216.903	120.784	146.600	0.824	111.407
'Aug-2020'	100.162	171.700	0.583	128.876	40.769	54.600	0.747	37.604
'Sep-2020'	256.417	383.600	0.668	329.926	70.842	76.600	0.925	65.343
'Oct-2020'	216.153	193.100	1.119	278.120	200.181	206.200	0.971	184.641
'Nov-2020'	309.217	462.000	0.669	397.863	102.491	85.300	1.202	94.534
'Dec-2020'	168.484	207.300	0.813	216.785	161.535	139.000	1.162	148.995
'Jan-2021'	119.929	117.700	1.019	154.310	54.887	47.100	1.165	50.626
'Feb-2021'	81.734	114.400	0.714	105.166	37.920	28.600	1.326	34.976
'Mar-2021'	153.130	219.300	0.698	197.029	41.480	33.900	1.224	38.260
'Apr-2021'	87.785	87.900	0.999	112.951	16.771	13.700	1.224	15.469
'May-2021'	45.097	53.500	0.843	58.026	114.530	96.700	1.184	105.639
'Jun-2021'	83.193	109.100	0.763	107.043	39.565	62.300	0.635	36.493
'Jul-2021'	46.355	105.700	0.439	59.645	112.268	126.900	0.885	103.553
'Aug-2021'	38.127	66.700	0.572	49.057	17.981	13.900	1.294	16.585
'Sep-2021'	127.358	180.200	0.707	163.869	71.910	62.100	1.158	66.327
'Oct-2021'	387.612	647.600	0.599	498.732	143.081	136.200	1.051	131.974
'Nov-2021'	273.919	325.300	0.842	352.447	49.080	43.000	1.141	45.270
'Dec-2021'	131.847	141.100	0.934	169.645	27.010	32.900	0.821	24.913
'Jan-2022'	321.072	344.600	0.932	413.117	28.745	23.400	1.228	26.513
'Feb-2022'	292.219	305.100	0.958	375.993	86.440	64.900	1.332	79.730
'Mar-2022'	58.418	59.800	0.977	75.166	12.059	3.500	3.445	11.122
'Apr-2022'	46.826	57.100	0.820	60.250	13.339	9.700	1.375	12.303

Figure A.7: Monthly precipitation for AROME, observed and fcfix data at Bergen and Oslo. Ratio is forecast divided by observed data.

	TRONDHEIM				TROMSØ			
	Forecasted data	Observed data	Ratio	fcfix	Forecasted data	Observed data	Ratio	fcfix
'Dec-2019'	94.767	115.100	0.823	82.875	97.435	129.200	0.754	123.863
'Jan-2020'	110.587	121.100	0.913	96.711	133.153	152.400	0.874	169.269
'Feb-2020'	112.161	101.100	1.109	98.087	97.209	148.400	0.655	123.576
'Mar-2020'	56.692	45.700	1.241	49.578	129.929	187.000	0.695	165.170
'Apr-2020'	19.540	18.000	1.086	17.088	78.302	121.800	0.643	99.540
'May-2020'	57.781	65.200	0.886	50.531	48.398	78.400	0.617	61.525
'Jun-2020'	51.510	30.000	1.717	45.047	16.757	24.300	0.690	21.302
'Jul-2020'	112.456	83.000	1.355	98.345	74.284	77.400	0.960	94.433
'Aug-2020'	51.717	76.700	0.674	45.228	143.521	157.400	0.912	182.450
'Sep-2020'	132.384	96.700	1.369	115.772	175.440	146.600	1.197	223.027
'Oct-2020'	68.722	53.900	1.275	60.099	50.979	77.800	0.655	64.806
'Nov-2020'	93.515	84.100	1.112	81.781	111.124	121.400	0.915	141.265
'Dec-2020'	15.288	16.200	0.944	13.369	35.634	33.300	1.070	45.299
'Jan-2021'	34.963	23.600	1.481	30.576	28.821	44.900	0.642	36.639
'Feb-2021'	18.473	15.100	1.223	16.155	97.419	149.800	0.650	123.843
'Mar-2021'	123.034	126.200	0.975	107.596	107.380	166.000	0.647	136.506
'Apr-2021'	77.288	63.100	1.225	67.589	102.436	182.700	0.561	130.221
'May-2021'	33.913	33.300	1.018	29.658	18.186	15.900	1.144	23.119
'Jun-2021'	46.094	46.900	0.983	40.310	55.088	63.600	0.866	70.030
'Jul-2021'	63.850	56.000	1.140	55.838	84.182	87.400	0.963	107.016
'Aug-2021'	150.692	115.200	1.308	131.783	51.965	44.900	1.157	66.060
'Sep-2021'	130.674	71.500	1.828	114.277	72.377	95.800	0.756	92.008
'Oct-2021'	183.713	154.300	1.191	160.660	55.526	75.500	0.735	70.587
'Nov-2021'	159.635	156.600	1.019	139.603	83.106	92.200	0.901	105.647
'Dec-2021'	82.625	65.900	1.254	72.257	87.543	112.200	0.780	111.288
'Jan-2022'	160.138	132.200	1.211	140.043	157.927	201.400	0.784	200.764
'Feb-2022'	116.444	79.000	1.474	101.832	66.174	69.500	0.952	84.124
'Mar-2022'	63.916	69.000	0.926	55.896	117.726	152.400	0.772	149.658
'Apr-2022'	59.940	56.500	1.061	52.418	53.463	81.100	0.659	67.964

Figure A.8: Monthly precipitation for AROME, observed and fcfix data at Trondheim and Tromsø. Ratio is forecast divided by observed data.

	KRISTIANSAND				NESBYEN			
	Forecasted data	Observed data	Ratio	fcfix	Forecasted data	Observed data	Ratio	fcfix
'Dec-2019'	213.319	216.600	0.985	211.855	35.191	38.200	0.921	28.980
'Jan-2020'	194.658	189.800	1.026	193.322	32.960	29.200	1.129	27.144
'Feb-2020'	271.922	229.800	1.183	270.055	47.780	28.500	1.676	39.348
'Mar-2020'	127.201	122.900	1.035	126.328	26.317	10.400	2.530	21.673
'Apr-2020'	65.664	49.400	1.329	65.213	19.996	8.200	2.439	16.467
'May-2020'	30.264	25.400	1.191	30.056	24.076	15.300	1.574	19.827
'Jun-2020'	132.915	114.200	1.164	132.002	97.029	103.100	0.941	79.906
'Jul-2020'	156.234	172.300	0.907	155.162	59.914	63.600	0.942	49.341
'Aug-2020'	53.032	61.100	0.868	52.668	33.999	30.800	1.104	27.999
'Sep-2020'	78.378	129.400	0.606	77.840	51.878	53.000	0.979	42.723
'Oct-2020'	244.607	206.000	1.187	242.928	124.365	123.500	1.007	102.418
'Nov-2020'	178.870	182.500	0.980	177.642	54.611	40.100	1.362	44.973
'Dec-2020'	303.706	373.900	0.812	301.621	71.220	51.400	1.386	58.652
'Jan-2021'	74.957	78.000	0.961	74.443	28.133	22.900	1.229	23.168
'Feb-2021'	79.776	64.800	1.231	79.229	18.353	17.700	1.037	15.114
'Mar-2021'	78.638	77.400	1.016	78.099	20.623	11.900	1.733	16.984
'Apr-2021'	18.329	17.600	1.041	18.203	14.858	4.300	3.455	12.236
'May-2021'	175.856	167.800	1.048	174.649	58.700	34.600	1.697	48.341
'Jun-2021'	69.857	69.900	0.999	69.377	80.188	50.500	1.588	66.037
'Jul-2021'	57.544	68.100	0.845	57.149	114.464	130.400	0.878	94.264
'Aug-2021'	29.676	64.000	0.464	29.472	27.273	12.100	2.254	22.460
'Sep-2021'	66.086	49.200	1.343	65.632	50.006	48.300	1.035	41.181
'Oct-2021'	83.900	95.900	0.875	83.324	71.336	51.200	1.393	58.747
'Nov-2021'	55.958	67.100	0.834	55.574	41.140	28.800	1.428	33.880
'Dec-2021'	96.293	82.300	1.170	95.632	15.732	12.200	1.290	12.956
'Jan-2022'	96.163	77.400	1.242	95.502	10.627	7.100	1.497	8.752
'Feb-2022'	161.561	141.900	1.139	160.452	44.010	25.700	1.712	36.243
'Mar-2022'	26.148	15.400	1.698	25.968	4.153	1.900	2.186	3.421
'Apr-2022'	20.745	9.500	2.184	20.603	8.217	4.700	1.748	6.767

Figure A.9: Monthly precipitation for AROME, observed and fcfix data at Kristiansand and Nesbyen. Ratio is forecast divided by observed data.

BERGEN	Forecasted data	Observed data	Ratio	fcfix
Winter	2064.7	2400.5	0.860	2656.6
Spring	808.3	1021.8	0.791	1040.0
Summer	480.6	721.9	0.666	618.4
Autumn	1570.7	2191.8	0.717	2021.0

OSLO	Forecasted data	Observed data	Ratio	fcfix
Winter	639.8	507.4	1.261	590.2
Spring	332.4	274.4	1.211	306.6
Summer	466.4	523.7	0.891	430.2
Autumn	637.6	609.4	1.046	588.1

TRONDHEIM	Forecasted data	Observed data	Ratio	fcfix
Winter	745.4	669.3	1.114	651.9
Spring	492.1	477.0	1.032	430.4
Summer	476.3	407.8	1.168	416.6
Autumn	768.6	617.1	1.246	672.2

TROMSØ	Forecasted data	Observed data	Ratio	fcfix
Winter	801.3	1041.1	0.770	1018.7
Spring	655.8	985.3	0.666	833.7
Summer	425.8	455.0	0.936	541.3
Autumn	548.6	609.3	0.900	697.3

KRISTIANSAND	Forecasted data	Observed data	Ratio	fcfix
Winter	1492.4	1454.5	1.026	1482.1
Spring	542.8	485.4	1.118	539.1
Summer	499.3	549.6	0.908	495.8
Autumn	707.8	730.1	0.969	702.9

NESBYEN	Forecasted data	Observed data	Ratio	fcfix
Winter	304.0	232.9	1.305	250.4
Spring	176.9	91.3	1.938	145.7
Summer	412.9	390.5	1.057	340.0
Autumn	393.3	344.9	1.140	323.9

Figure A.10: Total seasonal precipitation for AROME, observed and fcfix data at each location. Ratio is forecast divided by observed data.

BERGEN	Mean Top 10% precip	Mean Top 1% precip	Mean Top 0.1% precip	Top forecast extreme	Top observed extreme
Forecasted data	1.942	5.039	7.971	14.735	3.450
Observed data	2.343	5.740	9.190	0.600	13.700
Ratio	0.829	0.878	0.867	24.559	0.252
fcfix	2.498	6.484	10.256	18.960	4.439
	Top 1% shared extreme hours	Top 1% shared extreme hours with ± 6h tolerance	Top 0.1% shared extreme hours	Top 0.1% shared extreme hours with ± 6h tolerance	
Total forecast hits	58	106	1	6	
Total data points	211	211	21	21	
Ratio	0.275	0.502	0.048	0.286	
fcfix hits	58	106	1	6	

OSLO	Mean Top 10% precip	Mean Top 1% precip	Mean Top 0.1% precip	Top forecast extreme	Top observed extreme
Forecasted data	0.930	3.258	8.072	26.677	0.176
Observed data	0.884	3.394	8.385	1.000	14.100
Ratio	1.051	0.960	0.963	26.677	0.013
fcfix	0.858	3.005	7.445	24.606	0.163
	Top 1% shared extreme hours	Top 1% shared extreme hours with ± 6h tolerance	Top 0.1% shared extreme hours	Top 0.1% shared extreme hours with ± 6h tolerance	
Total forecast hits	64	116	3	8	
Total data points	211	211	21	21	
Ratio	0.303	0.550	0.143	0.381	
fcfix hits	64	116	3	8	

TRONDHEIM	Mean Top 10% precip	Mean Top 1% precip	Mean Top 0.1% precip	Top forecast extreme	Top observed extreme
Forecasted data	1.022	3.127	5.855	9.414	0.000
Observed data	0.954	3.013	6.515	0.000	15.800
Ratio	1.072	1.038	0.899	Inf	0.000
fcfix	0.894	2.735	5.121	8.233	0.000
	Top 1% shared extreme hours	Top 1% shared extreme hours with ± 6h tolerance	Top 0.1% shared extreme hours	Top 0.1% shared extreme hours with ± 6h tolerance	
Total forecast hits	54	94	0	1	
Total data points	211	211	21	21	
Ratio	0.256	0.445	0.000	0.048	
fcfix hits	54	94	0	1	

Figure A.11: Hourly mean precipitation amount for top 10/1/0.1% of all non-NaN hours, in addition to top 1/0.1% shared extreme hours with and without 6h tolerance for AROME, observed and fcfix data at Bergen, Oslo and Trondheim. Ratio is forecast divided by observed data. Top forecast extreme is the highest forecasted rainfall amount in one hour (using the first 12 forecast hours), with the corresponding observed value for that hour below. Vice versa for top observed extreme.

TROMSØ	Mean Top 10% precip	Mean Top 1% precip	Mean Top 0.1% precip	Top forecast extreme	Top observed extreme
Forecasted data	0.907	2.517	4.926	11.592	0.550
Observed data	1.061	2.692	5.250	0.000	7.200
Ratio	0.855	0.935	0.938	Inf	0.076
fcfix	1.153	3.199	6.262	14.736	0.700
	Top 1% shared extreme hours	Top 1% shared extreme hours with ± 6h tolerance	Top 0.1% shared extreme hours	Top 0.1% shared extreme hours with ± 6h tolerance	
Total forecast hits	26	83	2	5	
Total data points	211	211	21	21	
Ratio	0.123	0.393	0.095	0.238	
fcfix hits	26	83	2	5	

KRISTIANSAND	Mean Top 10% precip	Mean Top 1% precip	Mean Top 0.1% precip	Top forecast extreme	Top observed extreme
Forecasted data	1.451	4.673	8.149	13.979	0.019
Observed data	1.503	4.833	9.590	1.300	14.600
Ratio	0.965	0.967	0.850	10.753	0.001
fcfix	1.441	4.641	8.093	13.883	0.019
	Top 1% shared extreme hours	Top 1% shared extreme hours with ± 6h tolerance	Top 0.1% shared extreme hours	Top 0.1% shared extreme hours with ± 6h tolerance	
Total forecast hits	80	129	3	5	
Total data points	211	211	21	21	
Ratio	0.379	0.611	0.143	0.238	
fcfix hits	80	129	3	5	

NESBYEN	Mean Top 10% precip	Mean Top 1% precip	Mean Top 0.1% precip	Top forecast extreme	Top observed extreme
Forecasted data	0.570	2.138	5.573	10.558	0.626
Observed data	0.506	2.429	6.515	0.700	12.100
Ratio	1.126	0.880	0.855	15.084	0.052
fcfix	0.467	1.761	4.589	8.695	0.516
	Top 1% shared extreme hours	Top 1% shared extreme hours with ± 6h tolerance	Top 0.1% shared extreme hours	Top 0.1% shared extreme hours with ± 6h tolerance	
Total forecast hits	82	134	0	5	
Total data points	211	211	21	21	
Ratio	0.389	0.635	0.000	0.238	
fcfix hits	82	134	0	5	

Figure A.12: Hourly mean precipitation amount for top 10/1/0.1% of all non-NaN hours, in addition to top 1/0.1% shared extreme hours with and without 6h tolerance for AROME, observed and fcfix data at Tromsø, Kristiansand and Nesbyen. Ratio is forecast divided by observed data. Top forecast extreme is the highest forecasted rainfall amount in one hour (using the first 12 forecast hours), with the corresponding observed value for that hour below. Vice versa for top observed extreme.

BERGEN	Mean Top 10% precip	Mean Top 5% precip	Mean Top 1% precip	Top forecast extreme	Top observed extreme
Forecasted data	27.908	35.271	52.903	69.912	59.648
Observed data	34.642	44.798	71.288	51.200	103.200
Ratio	0.806	0.787	0.742	1.365	0.578
fcfix	35.909	45.382	68.069	89.954	76.748
	Top 5% shared extreme days	Top 1% shared extreme days			
Total forecast hits	27	4			
Total data points	44	8			
Ratio	0.614	0.500			
fcfix hits	27	4			

OSLO	Mean Top 10% precip	Mean Top 5% precip	Mean Top 1% precip	Top forecast extreme	Top observed extreme
Forecasted data	15.794	21.293	35.331	44.792	20.048
Observed data	15.281	20.765	31.500	26.400	44.500
Ratio	1.034	1.025	1.122	1.697	0.451
fcfix	14.568	19.640	32.589	41.314	18.491
	Top 5% shared extreme days	Top 1% shared extreme days			
Total forecast hits	27	5			
Total data points	44	8			
Ratio	0.614	0.625			
fcfix hits	27	5			

TRONDHEIM	Mean Top 10% precip	Mean Top 5% precip	Mean Top 1% precip	Top forecast extreme	Top observed extreme
Forecasted data	15.286	20.288	37.052	66.129	38.563
Observed data	13.892	18.340	33.413	31.600	41.800
Ratio	1.100	1.106	1.109	2.093	0.923
fcfix	13.368	17.742	32.402	57.831	33.724
	Top 5% shared extreme days	Top 1% shared extreme days			
Total forecast hits	28	4			
Total data points	44	8			
Ratio	0.636	0.500			
fcfix hits	28	4			

Figure A.13: Daily mean precipitation amount for top 10/5/1% of all non-NaN days, in addition to top 5/1% shared extreme days for AROME, observed and fcfix data at Bergen, Oslo and Trondheim. Ratio is forecast divided by observed data. Top forecast extreme is the highest forecasted rainfall amount in one day (using the first 12 forecast hours), with the corresponding observed value for that day below. Vice versa for top observed extreme.

TROMSØ	Mean Top 10% precip	Mean Top 5% precip	Mean Top 1% precip	Top forecast extreme	Top observed extreme
Forecasted data	12.763	16.313	23.647	35.040	27.221
Observed data	15.864	19.184	25.450	20.400	30.300
Ratio	0.805	0.850	0.929	1.718	0.898
fcfix	16.225	20.738	30.061	44.545	34.605
	Top 5% shared extreme days	Top 1% shared extreme days			
Total forecast hits	25	1			
Total data points	44	8			
Ratio	0.568	0.125			
fcfix hits	25	1			

KRISTIANSAND	Mean Top 10% precip	Mean Top 5% precip	Mean Top 1% precip	Top forecast extreme	Top observed extreme
Forecasted data	23.742	31.585	43.365	48.866	48.026
Observed data	24.495	31.656	41.725	33.700	48.300
Ratio	0.969	0.998	1.039	1.450	0.994
fcfix	23.579	31.368	43.067	48.531	47.696
	Top 5% shared extreme days	Top 1% shared extreme days			
Total forecast hits	32	3			
Total data points	44	8			
Ratio	0.727	0.375			
fcfix hits	32	3			

NESBYEN	Mean Top 10% precip	Mean Top 5% precip	Mean Top 1% precip	Top forecast extreme	Top observed extreme
Forecasted data	9.576	13.032	23.913	40.561	29.890
Observed data	8.960	13.067	23.388	20.900	33.100
Ratio	1.069	0.997	1.022	1.941	0.903
fcfix	7.886	10.732	19.693	33.403	24.615
	Top 5% shared extreme days	Top 1% shared extreme days			
Total forecast hits	32	3			
Total data points	44	8			
Ratio	0.727	0.375			
fcfix hits	32	3			

Figure A.14: Daily mean precipitation amount for top 10/5/1% of all non-NaN days, in addition to top 5/1% shared extreme days for AROME, observed and fcfix data at Tromsø, Kristiansand and Nesbyen. Ratio is forecast divided by observed data. Top forecast extreme is the highest forecasted rainfall amount in one day (using the first 12 forecast hours), with the corresponding observed value for that day below. Vice versa for top observed extreme.

BERGEN	Mean top 1% forecasted extremes	Mean top 1% observed extremes	Avg. hours per top 1% forecast extreme	Avg. hours per top 1% observed extreme
Winter	5.012	5.536	71.462	86.707
Spring	4.596	4.892	420.071	452.385
Summer	5.169	5.553	184.000	147.200
Autumn	5.074	6.054	53.268	46.968

OSLO	Mean top 1% forecasted extremes	Mean top 1% observed extremes	Avg. hours per top 1% forecast extreme	Avg. hours per top 1% observed extreme
Winter	2.446	2.456	166.744	203.219
Spring	3.465	2.964	294.050	267.318
Summer	4.342	4.337	74.847	62.197
Autumn	2.837	3.043	46.968	50.791

TRONDHEIM	Mean top 1% forecasted extremes	Mean top 1% observed extremes	Avg. hours per top 1% forecast extreme	Avg. hours per top 1% observed extreme
Winter	2.853	2.492	104.887	104.887
Spring	2.877	2.809	193.321	159.206
Summer	3.824	3.918	105.143	86.588
Autumn	2.960	2.773	54.987	67.875

TROMSØ	Mean top 1% forecasted extremes	Mean top 1% observed extremes	Avg. hours per top 1% forecast extreme	Avg. hours per top 1% observed extreme
Winter	2.396	2.557	87.716	84.299
Spring	2.270	2.712	183.781	99.678
Summer	3.074	2.926	113.231	126.171
Autumn	2.418	2.673	66.182	109.200

KRISTIANSAND	Mean top 1% forecasted extremes	Mean top 1% observed extremes	Avg. hours per top 1% forecast extreme	Avg. hours per top 1% observed extreme
Winter	4.224	4.126	89.082	79.305
Spring	4.329	4.541	150.179	216.926
Summer	5.183	5.721	96.000	91.915
Autumn	4.858	5.018	71.111	69.818

NESBYEN	Mean top 1% forecasted extremes	Mean top 1% observed extremes	Avg. hours per top 1% forecast extreme	Avg. hours per top 1% observed extreme
Winter	1.395	2.363	500.231	406.438
Spring	1.757	2.088	345.941	735.125
Summer	2.574	2.824	44.160	46.000
Autumn	1.776	2.026	53.926	48.000

Figure A.15: Mean of the top 1% hourly forecasted/observed extremes, and average hours between each occurrence of a top 1% extreme precipitation event for forecast and observed data at each location. Data is divided into seasons.

BERGEN	Mean top 5% forecasted extremes	Mean top 5% observed extremes	Avg. days per top 5% forecast extreme	Avg. days per top 5% observed extreme
Winter	37.486	44.672	14.263	15.056
Spring	30.653	44.033	49.000	81.667
Summer	39.656	35.567	92.000	61.333
Autumn	33.075	45.710	10.111	9.100

OSLO	Mean top 5% forecasted extremes	Mean top 5% observed extremes	Avg. days per top 5% forecast extreme	Avg. days per top 5% observed extreme
Winter	20.343	19.600	22.583	33.875
Spring	17.137	16.960	35.000	49.000
Summer	23.943	21.500	16.727	12.267
Autumn	21.465	21.381	13.000	11.375

TRONDHEIM	Mean top 5% forecasted extremes	Mean top 5% observed extremes	Avg. days per top 5% forecast extreme	Avg. days per top 5% observed extreme
Winter	18.266	16.613	20.846	16.938
Spring	15.799	19.720	28.125	45.000
Summer	24.029	18.410	26.286	18.400
Autumn	21.526	18.846	11.250	13.846

TROMSØ	Mean top 5% forecasted extremes	Mean top 5% observed extremes	Avg. days per top 5% forecast extreme	Avg. days per top 5% observed extreme
Winter	15.068	19.779	15.882	19.286
Spring	14.830	19.114	24.500	17.500
Summer	20.984	18.183	36.800	30.667
Autumn	16.902	18.620	15.167	18.200

KRISTIANSAND	Mean top 5% forecasted extremes	Mean top 5% observed extremes	Avg. days per top 5% forecast extreme	Avg. days per top 5% observed extreme
Winter	30.539	29.576	13.550	12.905
Spring	33.067	34.438	24.400	30.500
Summer	31.971	29.129	35.800	25.571
Autumn	28.223	32.938	17.778	20.000

NESBYEN	Mean top 5% forecasted extremes	Mean top 5% observed extremes	Avg. days per top 5% forecast extreme	Avg. days per top 5% observed extreme
Winter	11.435	8.840	67.750	54.200
Spring	9.010	NaN	81.667	Inf
Summer	13.816	13.877	9.200	8.364
Autumn	12.899	12.865	10.706	10.706

Figure A.16: Mean of the top 5% daily forecasted/observed extremes, and average days between each occurrence of a top 5% extreme precipitation event for forecast and observed data at each location. Data is divided into seasons.

			1h accumulated precipitation					
BERGEN	+4h forecast	+4h fcfix	+4h fcday	+13h forecast	+13h fcfix	+13h fcday	+48h forecast	+48h fcfix
Hit	307	322	258	267	274	227	245	252
Miss	229	214	278	228	221	270	243	236
False Alarm	81	89	278	119	132	274	134	147
Correct negative	1123	1115	926	1111	1098	969	1099	1086
NaN values	19	19	19	34	34	19	38	38
OSLO	+4h forecast	+4h fcfix	+4h fcday	+13h forecast	+13h fcfix	+13h fcday	+48h forecast	+48h fcfix
Hit	137	135	51	157	155	54	131	129
Miss	86	88	172	66	68	171	99	101
False Alarm	66	61	170	67	66	166	97	95
Correct negative	1451	1456	1347	1435	1436	1349	1394	1396
NaN values	19	19	19	34	34	19	38	38
TRONDHEIM	+4h forecast	+4h fcfix	+4h fcday	+13h forecast	+13h fcfix	+13h fcday	+48h forecast	+48h fcfix
Hit	149	146	82	144	143	90	148	141
Miss	147	150	214	136	137	191	144	151
False Alarm	127	118	217	140	132	191	122	118
Correct negative	1276	1285	1180	1261	1269	1221	1258	1262
NaN values	60	60	66	78	78	66	87	87
TROMSØ	+4h forecast	+4h fcfix	+4h fcday	+13h forecast	+13h fcfix	+13h fcday	+48h forecast	+48h fcfix
Hit	245	261	239	222	235	231	201	218
Miss	269	253	274	282	269	278	323	306
False Alarm	102	120	272	112	126	280	117	131
Correct negative	1122	1104	952	1105	1091	948	1076	1062
NaN values	21	21	22	38	38	22	42	42
KRISTIANSAND	+4h forecast	+4h fcfix	+4h fcday	+13h forecast	+13h fcfix	+13h fcday	+48h forecast	+48h fcfix
Hit	159	159	62	169	169	70	147	147
Miss	77	77	173	86	86	185	122	122
False Alarm	97	96	175	91	90	185	91	90
Correct negative	1356	1357	1273	1326	1327	1243	1305	1306
NaN values	70	70	76	87	87	76	94	94
NESBYEN	+4h forecast	+4h fcfix	+4h fcday	+13h forecast	+13h fcfix	+13h fcday	+48h forecast	+48h fcfix
Hit	112	109	32	110	101	36	90	87
Miss	44	47	124	56	65	131	74	77
False Alarm	96	88	122	67	63	130	102	92
Correct negative	1488	1496	1462	1492	1496	1443	1455	1465
NaN values	19	19	19	34	34	19	38	38

Figure A.17: Contingency table results for 1h accumulated precipitation for AROME, fcfix and fcday against observed data.

			6h accumulated precipitation					
BERGEN	+4-9h forecast	+4-9h fcfix	+4-9h fcday	+13-18h forecast	+13-18h fcfix	+13-18h fcday	+25-30h forecast	+25-30h fcfix
Hit	661	680	546	673	685	556	653	668
Miss	174	155	289	159	147	281	176	161
False Alarm	105	115	289	101	110	282	111	117
Correct negative	800	790	616	792	783	621	783	777
NaN values	19	19	19	34	34	19	36	36
OSLO	+4-9h forecast	+4-9h fcfix	+4-9h fcday	+13-18h forecast	+13-18h fcfix	+13-18h fcday	+25-30h forecast	+25-30h fcfix
Hit	323	321	194	315	314	197	310	310
Miss	133	135	262	131	132	252	135	135
False Alarm	92	86	260	89	84	250	104	102
Correct negative	1192	1198	1024	1190	1195	1041	1174	1176
NaN values	19	19	19	34	34	19	36	36
TRONDHEIM	+4-9h forecast	+4-9h fcfix	+4-9h fcday	+13-18h forecast	+13-18h fcfix	+13-18h fcday	+25-30h forecast	+25-30h fcfix
Hit	480	476	319	452	448	310	461	451
Miss	129	133	290	124	128	269	112	122
False Alarm	170	164	289	185	175	269	184	176
Correct negative	920	926	795	920	930	845	919	927
NaN values	60	60	66	78	78	66	83	83
TROMSØ	+4-9h forecast	+4-9h fcfix	+4-9h fcday	+13-18h forecast	+13-18h fcfix	+13-18h fcday	+25-30h forecast	+25-30h fcfix
Hit	622	639	544	626	644	573	601	618
Miss	238	221	315	238	220	301	264	247
False Alarm	108	115	316	104	114	304	117	124
Correct negative	770	763	562	753	743	559	737	730
NaN values	21	21	22	38	38	22	40	40
KRISTIANSAND	+4-9h forecast	+4-9h fcfix	+4-9h fcday	+13-18h forecast	+13-18h fcfix	+13-18h fcday	+25-30h forecast	+25-30h fcfix
Hit	398	398	224	389	389	231	373	373
Miss	87	87	260	99	99	258	111	111
False Alarm	117	117	261	131	131	260	130	128
Correct negative	1087	1087	938	1053	1053	934	1054	1056
NaN values	70	70	76	87	87	76	91	91
NESBYEN	+4-9h forecast	+4-9h fcfix	+4-9h fcday	+13-18h forecast	+13-18h fcfix	+13-18h fcday	+25-30h forecast	+25-30h fcfix
Hit	309	299	141	270	265	126	268	260
Miss	67	77	235	81	86	226	79	87
False Alarm	132	123	236	125	115	225	152	145
Correct negative	1232	1241	1128	1249	1259	1163	1224	1231
NaN values	19	19	19	34	34	19	36	36

Figure A.18: Contingency table results for 6h accumulated precipitation for AROME, fcfix and fcday against observed data.

24h accumulated precipitation

BERGEN	+4-27h forecast	+4-27h fcfix	+13-36h forecast	+13-36h fcfix	+25-48h forecast	+25-48h fcfix
Hit	1075	1083	1074	1086	1058	1060
Miss	111	103	113	101	127	125
False Alarm	84	96	76	83	92	101
Correct negative	440	428	447	440	431	422
NaN values	49	49	49	49	51	51

OSLO	+4-27h forecast	+4-27h fcfix	+13-36h forecast	+13-36h fcfix	+25-48h forecast	+25-48h fcfix
Hit	640	637	644	642	644	642
Miss	177	180	171	173	170	172
False Alarm	99	94	123	121	121	116
Correct negative	794	799	772	774	773	778
NaN values	49	49	49	49	51	51

TRONDHEIM	+4-27h forecast	+4-27h fcfix	+13-36h forecast	+13-36h fcfix	+25-48h forecast	+25-48h fcfix
Hit	891	888	893	889	880	875
Miss	87	90	84	88	95	100
False Alarm	148	143	135	127	144	140
Correct negative	537	542	551	559	539	543
NaN values	96	96	96	96	101	101

TROMSØ	+4-27h forecast	+4-27h fcfix	+13-36h forecast	+13-36h fcfix	+25-48h forecast	+25-48h fcfix
Hit	1058	1068	1048	1060	1041	1047
Miss	162	152	175	163	181	175
False Alarm	75	82	76	85	83	95
Correct negative	409	402	405	396	397	385
NaN values	55	55	55	55	57	57

KRISTIANSAND	+4-27h forecast	+4-27h fcfix	+13-36h forecast	+13-36h fcfix	+25-48h forecast	+25-48h fcfix
Hit	760	760	742	742	725	725
Miss	79	79	100	100	115	115
False Alarm	117	117	128	126	124	121
Correct negative	699	699	685	687	687	690
NaN values	104	104	104	104	108	108

NESBYEN	+4-27h forecast	+4-27h fcfix	+13-36h forecast	+13-36h fcfix	+25-48h forecast	+25-48h fcfix
Hit	611	600	597	592	594	590
Miss	100	111	118	123	120	124
False Alarm	174	160	167	156	172	158
Correct negative	825	839	828	839	822	836
NaN values	49	49	49	49	51	51

Figure A.19: Contingency table results for 24h accumulated precipitation for AROME and fcfix against observed data.

48h accumulated precipitation

BERGEN	+1-48h forecast	+1-48h fcfix
Hit	1303	1305
Miss	84	82
False Alarm	54	62
Correct negative	254	246
NaN values	64	64

TROMSØ	+1-48h forecast	+1-48h fcfix
Hit	1287	1290
Miss	128	125
False Alarm	58	67
Correct negative	214	205
NaN values	72	72

OSLO	+1-48h forecast	+1-48h fcfix
Hit	893	888
Miss	194	199
False Alarm	105	99
Correct negative	503	509
NaN values	64	64

KRISTIANSAND	+1-48h forecast	+1-48h fcfix
Hit	979	979
Miss	100	100
False Alarm	89	89
Correct negative	470	470
NaN values	121	121

TRONDHEIM	+1-48h forecast	+1-48h fcfix
Hit	1150	1144
Miss	67	73
False Alarm	95	92
Correct negative	333	336
NaN values	114	114

NESBYEN	+1-48h forecast	+1-48h fcfix
Hit	865	860
Miss	117	122
False Alarm	165	156
Correct negative	548	557
NaN values	64	64

Figure A.20: Contingency table results for 48h accumulated precipitation for AROME and fcfix against observed data.

												1h accumulated precipitation												
	forecast												**fcpersist**											
BERGEN	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h
Hit	277	272	304	308	280	290	282	283	278	267	257	253	395	362	325	324	304	296	280	262	282	261	252	239
Miss	221	219	213	229	239	221	243	231	262	241	248	244	103	129	192	213	215	215	245	252	258	247	253	258
False Alarm	98	107	102	81	94	98	98	99	107	95	101	105	103	136	173	174	194	202	218	236	216	237	246	259
Correct negative	1148	1146	1125	1126	1131	1135	1121	1131	1097	1141	1138	1142	1143	1117	1054	1033	1031	1031	1001	994	988	999	993	988
NaN values	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19
OSLO	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h
Hit	148	140	150	137	147	145	143	149	141	151	155	156	165	142	131	112	103	99	91	91	80	75	72	77
Miss	77	85	91	86	82	85	90	78	73	67	71	75	60	83	110	111	126	131	142	136	134	143	154	154
False Alarm	63	66	49	66	51	61	57	70	71	62	61	77	68	91	102	121	130	134	142	142	153	158	161	156
Correct negative	1456	1453	1454	1455	1464	1453	1454	1447	1459	1464	1457	1436	1451	1428	1401	1400	1385	1380	1369	1375	1377	1368	1357	1357
NaN values	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19
TRONDHEIM	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h
Hit	146	165	176	149	166	185	178	177	169	170	164	159	197	167	169	159	148	138	136	136	115	120	116	117
Miss	137	126	118	148	128	132	145	145	123	142	133	140	86	124	125	138	146	179	187	186	177	192	181	182
False Alarm	134	127	117	127	131	111	128	134	140	127	120	124	104	134	132	142	153	163	165	165	186	181	185	184
Correct negative	1286	1285	1292	1279	1278	1275	1252	1247	1271	1264	1286	1280	1313	1275	1274	1261	1253	1220	1212	1213	1222	1207	1218	1217
NaN values	60	60	60	60	60	60	60	60	60	60	60	60	63	63	63	63	63	63	63	63	63	63	63	63
TROMSØ	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h
Hit	231	245	248	247	269	257	265	265	226	236	232	239	400	373	349	332	316	326	306	292	294	282	277	271
Miss	280	276	272	270	262	299	268	262	298	281	292	295	110	148	171	185	215	229	226	234	230	235	247	263
False Alarm	82	84	90	102	93	85	107	97	107	116	105	91	129	156	180	197	213	203	223	237	235	247	252	258
Correct negative	1149	1137	1132	1123	1118	1101	1102	1118	1111	1109	1113	1117	1102	1064	1041	1027	997	983	986	978	982	977	965	949
NaN values	21	21	21	21	21	21	21	21	21	21	21	21	22	22	22	22	22	22	22	22	22	22	22	22
KRISTIANSAND	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h
Hit	177	160	166	159	166	166	173	178	169	186	174	175	189	158	139	130	119	122	122	109	106	107	102	93
Miss	80	97	87	77	77	87	91	74	85	80	96	97	67	99	114	106	124	131	142	143	148	158	167	179
False Alarm	70	79	78	97	95	96	90	92	107	86	90	96	83	114	133	142	153	150	150	163	166	165	170	179
Correct negative	1366	1357	1362	1360	1355	1344	1339	1349	1332	1341	1333	1325	1351	1319	1304	1312	1294	1287	1276	1275	1270	1260	1251	1239
NaN values	70	70	70	70	70	70	70	70	70	70	70	70	73	73	73	73	73	73	73	73	73	73	73	73
NESBYEN	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h
Hit	110	99	114	112	108	107	111	101	104	102	112	113	109	93	86	77	67	67	61	48	44	36	43	36
Miss	57	56	52	44	53	61	65	69	67	68	59	51	58	62	80	79	94	101	115	122	127	134	128	128
False Alarm	70	65	65	96	97	86	88	86	71	70	70	63	54	70	77	86	96	96	102	115	119	127	120	127
Correct negative	1507	1524	1513	1492	1486	1490	1480	1488	1502	1504	1503	1517	1523	1519	1501	1502	1487	1480	1466	1459	1454	1447	1453	1453
NaN values	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19

Figure A.21: Contingency table results for 1h accumulated precipitation at each for the first 12 forecast hours, for AROME and fcpersist against observed data.

	6h accumulated precipitation			
	forecast		fcpersist	
BERGEN	+1-6h	+7-12h	+1-6h	+7-12h
Hit	687	647	467	401
Miss	151	207	371	453
False Alarm	92	97	30	96
Correct negative	811	790	873	791
NaN values	19	19	19	19
OSLO	+1-6h	+7-12h	+1-6h	+7-12h
Hit	327	336	193	135
Miss	122	97	256	298
False Alarm	85	98	40	98
Correct negative	1207	1210	1252	1210
NaN values	19	19	19	19
TRONDHEIM	+1-6h	+7-12h	+1-6h	+7-12h
Hit	467	501	272	218
Miss	113	114	308	397
False Alarm	176	152	28	82
Correct negative	944	933	1089	1000
NaN values	60	60	63	63
TROMSØ	+1-6h	+7-12h	+1-6h	+7-12h
Hit	631	658	488	400
Miss	245	211	387	468
False Alarm	98	104	38	126
Correct negative	765	766	825	744
NaN values	21	21	22	22
KRISTIANSAND	+1-6h	+7-12h	+1-6h	+7-12h
Hit	400	395	234	178
Miss	91	93	256	309
False Alarm	101	114	38	94
Correct negative	1098	1088	1159	1106
NaN values	70	70	73	73
NESBYEN	+1-6h	+7-12h	+1-6h	+7-12h
Hit	279	283	138	87
Miss	73	82	214	278
False Alarm	127	125	25	76
Correct negative	1262	1251	1364	1300
NaN values	19	19	19	19

Figure A.22: Contingency table results for 6h accumulated precipitation for AROME and fcpersist against observed data.

1h accumulated precipitation

BERGEN	**+4h forecast**	**+4h fcfix**	**+4h fcday**	**+13h forecast**	**+13h fcfix**	**+13h fcday**	**+48h forecast**	**+48h fcfix**
Accuracy	0.8218	0.8259	0.6805	0.7988	0.7954	0.6874	0.7809	0.7775
Bias frequency	0.7239	0.7668	1.0000	0.7798	0.8202	1.0080	0.7766	0.8176
Hit rate	0.5728	0.6007	0.4813	0.5394	0.5535	0.4567	0.5020	0.5164
False alarm ratio	0.2088	0.2165	0.5187	0.3083	0.3251	0.5469	0.3536	0.3684
Success ratio	0.7912	0.7835	0.4813	0.6917	0.6749	0.4531	0.6464	0.6316
Brier score	0.1782	0.1741	0.3195	0.2012	0.2046	0.3126	0.2191	0.2225
Brier skill score	0.4214	0.4347	-0.0373	0.2990	0.2871	-0.0945	0.2274	0.2154
MAE	0.2860	0.3173	0.4947	0.2970	0.3269	0.5303	0.3074	0.3437
RMSE	0.7492	0.8204	1.1422	0.8058	0.9002	1.2362	0.8546	0.9622

OSLO	**+4h forecast**	**+4h fcfix**	**+4h fcday**	**+13h forecast**	**+13h fcfix**	**+13h fcday**	**+48h forecast**	**+48h fcfix**
Accuracy	0.9126	0.9144	0.8034	0.9229	0.9223	0.8063	0.8861	0.8861
Bias frequency	0.9103	0.8789	0.9910	1.0045	0.9910	0.9778	0.9913	0.9739
Hit rate	0.6143	0.6054	0.2287	0.7040	0.6951	0.2400	0.5696	0.5609
False alarm ratio	0.3251	0.3112	0.7692	0.2991	0.2986	0.7545	0.4254	0.4241
Success ratio	0.6749	0.6888	0.2308	0.7009	0.7014	0.2455	0.5746	0.5759
Brier score	0.0874	0.0856	0.1966	0.0771	0.0777	0.1937	0.1139	0.1139
Brier skill score	0.3183	0.3323	-0.5335	0.4037	0.3991	-0.4981	0.1475	0.1475
MAE	0.0912	0.0885	0.1764	0.1011	0.0968	0.1407	0.1291	0.1229
RMSE	0.4110	0.4040	0.6431	0.4915	0.4604	0.4817	0.5668	0.5483

TRONDHEIM	**+4h forecast**	**+4h fcfix**	**+4h fcday**	**+13h forecast**	**+13h fcfix**	**+13h fcday**	**+48h forecast**	**+48h fcfix**
Accuracy	0.8387	0.8423	0.7454	0.8358	0.8400	0.7744	0.8409	0.8391
Bias frequency	0.9324	0.8919	1.0101	1.0143	0.9821	1.0000	0.9247	0.8870
Hit rate	0.5034	0.4932	0.2770	0.5143	0.5107	0.3203	0.5068	0.4829
False alarm ratio	0.4601	0.4470	0.7258	0.4930	0.4800	0.6797	0.4519	0.4556
Success ratio	0.5399	0.5530	0.2742	0.5070	0.5200	0.3203	0.5481	0.5444
Brier score	0.1613	0.1577	0.2546	0.1642	0.1600	0.2256	0.1591	0.1609
Brier skill score	0.0741	0.0947	-0.4565	0.0144	0.0396	-0.3590	0.0888	0.0785
MAE	0.1263	0.1181	0.1683	0.1212	0.1115	0.1411	0.1433	0.1338
RMSE	0.4229	0.3849	0.4796	0.3818	0.3462	0.3914	0.4486	0.4244

TROMSØ	**+4h forecast**	**+4h fcfix**	**+4h fcday**	**+13h forecast**	**+13h fcfix**	**+13h fcday**	**+48h forecast**	**+48h fcfix**
Accuracy	0.7865	0.7854	0.6857	0.7711	0.7705	0.6788	0.7437	0.7455
Bias frequency	0.6751	0.7412	0.9961	0.6627	0.7163	1.0039	0.6069	0.6660
Hit rate	0.4767	0.5078	0.4659	0.4405	0.4663	0.4538	0.3836	0.4160
False alarm ratio	0.2939	0.3150	0.5323	0.3353	0.3490	0.5479	0.3679	0.3754
Success ratio	0.7061	0.6850	0.4677	0.6647	0.6510	0.4521	0.6321	0.6246
Brier score	0.2135	0.2146	0.3143	0.2289	0.2295	0.3212	0.2563	0.2545
Brier skill score	0.2780	0.2743	-0.0643	0.2185	0.2165	-0.0962	0.1602	0.1661
MAE	0.1575	0.1755	0.2244	0.1596	0.1782	0.2208	0.1680	0.1895
RMSE	0.4033	0.4537	0.5266	0.4731	0.5458	0.5335	0.4739	0.5592

KRISTIANSAND	**+4h forecast**	**+4h fcfix**	**+4h fcday**	**+13h forecast**	**+13h fcfix**	**+13h fcday**	**+48h forecast**	**+48h fcfix**
Accuracy	0.8970	0.8976	0.7932	0.8941	0.8947	0.7802	0.8721	0.8727
Bias frequency	1.0847	1.0805	1.0085	1.0196	1.0157	1.0000	0.8848	0.8810
Hit rate	0.6737	0.6737	0.2638	0.6627	0.6627	0.2745	0.5465	0.5465
False alarm ratio	0.3789	0.3765	0.7384	0.3500	0.3475	0.7255	0.3824	0.3797
Success ratio	0.6211	0.6235	0.2616	0.6500	0.6525	0.2745	0.6176	0.6203
Brier score	0.1030	0.1024	0.2068	0.1059	0.1053	0.2198	0.1279	0.1273
Brier skill score	0.2627	0.2670	-0.4814	0.3056	0.3095	-0.4508	0.2085	0.2123
MAE	0.1395	0.1389	0.2386	0.1402	0.1399	0.2854	0.1677	0.1672
RMSE	0.5211	0.5182	0.7369	0.5184	0.5177	0.9041	0.5718	0.5700

NESBYEN	**+4h forecast**	**+4h fcfix**	**+4h fcday**	**+13h forecast**	**+13h fcfix**	**+13h fcday**	**+48h forecast**	**+48h fcfix**
Accuracy	0.9195	0.9224	0.8586	0.9287	0.9258	0.8500	0.8977	0.9018
Bias frequency	1.3333	1.2628	0.9872	1.0663	0.9880	0.9940	1.1707	1.0915
Hit rate	0.7179	0.6987	0.2051	0.6627	0.6084	0.2156	0.5488	0.5305
False alarm ratio	0.4615	0.4467	0.7922	0.3785	0.3841	0.7831	0.5313	0.5140
Success ratio	0.5385	0.5533	0.2078	0.6215	0.6159	0.2169	0.4688	0.4860
Brier score	0.0805	0.0776	0.1414	0.0713	0.0742	0.1500	0.1023	0.0982
Brier skill score	0.1026	0.1349	-0.5764	0.2588	0.2287	-0.5625	-0.0735	-0.0304
MAE	0.0673	0.0595	0.0818	0.0704	0.0654	0.1196	0.0875	0.0810
RMSE	0.3414	0.3085	0.3716	0.4349	0.4307	0.6744	0.4668	0.4476

Figure A.23: Forecast verification results for 1h accumulated precipitation for AROME, fcfix and fcday against observed data.

6h accumulated precipitation

BERGEN	+4-9h forecast	+4-9h fcfix	+4-9h fcday	+13-18h forecast	+13-18h fcfix	+13-18h fcday	+25-30h forecast	+25-30h fcfix
Accuracy	0.8397	0.8448	0.6678	0.8493	0.8510	0.6764	0.8334	0.8387
Bias frequency	0.9174	0.9521	1.0000	0.9303	0.9555	1.0012	0.9216	0.9469
Hit rate	0.7916	0.8144	0.6539	0.8089	0.8233	0.6643	0.7877	0.8058
False alarm ratio	0.1371	0.1447	0.3461	0.1305	0.1384	0.3365	0.1453	0.1490
Success ratio	0.8629	0.8553	0.6539	0.8695	0.8616	0.6635	0.8547	0.8510
Brier score	0.1603	0.1552	0.3322	0.1507	0.1490	0.3236	0.1666	0.1613
Brier skill score	0.6660	0.6766	0.3078	0.6875	0.6911	0.3272	0.6537	0.6647
MAE	1.1157	1.2207	2.5229	1.1944	1.2726	2.5641	1.2354	1.3448
RMSE	2.4757	2.6543	5.2403	2.6714	2.7590	5.5050	2.7856	2.9599

OSLO	+4-9h forecast	+4-9h fcfix	+4-9h fcday	+13-18h forecast	+13-18h fcfix	+13-18h fcday	+25-30h forecast	+25-30h fcfix
Accuracy	0.8707	0.8730	0.7000	0.8725	0.8748	0.7115	0.8613	0.8624
Bias frequency	0.9101	0.8925	0.9956	0.9058	0.8924	0.9955	0.9303	0.9258
Hit rate	0.7083	0.7039	0.4254	0.7063	0.7040	0.4388	0.6966	0.6966
False alarm ratio	0.2217	0.2113	0.5727	0.2203	0.2111	0.5593	0.2512	0.2476
Success ratio	0.7783	0.7887	0.4273	0.7797	0.7889	0.4407	0.7488	0.7524
Brier score	0.1293	0.1270	0.3000	0.1275	0.1252	0.2885	0.1387	0.1376
Brier skill score	0.5067	0.5155	-0.1446	0.5070	0.5159	-0.1182	0.4630	0.4673
MAE	0.3726	0.3527	0.8701	0.4405	0.4189	0.8646	0.4753	0.4539
RMSE	1.1991	1.1513	2.3164	1.5763	1.5050	2.4609	1.7764	1.6834

TRONDHEIM	+4-9h forecast	+4-9h fcfix	+4-9h fcday	+13-18h forecast	+13-18h fcfix	+13-18h fcday	+25-30h forecast	+25-30h fcfix
Accuracy	0.8240	0.8252	0.6580	0.8162	0.8198	0.6822	0.8234	0.8222
Bias frequency	1.0673	1.0509	0.9984	1.1059	1.0816	1.0000	1.1257	1.0942
Hit rate	0.7882	0.7816	0.5238	0.7847	0.7778	0.5354	0.8045	0.7871
False alarm ratio	0.2615	0.2563	0.4753	0.2904	0.2809	0.4646	0.2853	0.2807
Success ratio	0.7385	0.7438	0.5247	0.7096	0.7191	0.5354	0.7147	0.7193
Brier score	0.1760	0.1748	0.3420	0.1838	0.1802	0.3178	0.1766	0.1778
Brier skill score	0.5089	0.5123	0.0492	0.4637	0.4742	0.0708	0.4835	0.4800
MAE	0.5643	0.5211	1.0376	0.5593	0.5163	0.8985	0.5541	0.5169
RMSE	1.5521	1.4390	2.5116	1.3804	1.2632	2.1465	1.2941	1.2022

TROMSØ	+4-9h forecast	+4-9h fcfix	+4-9h fcday	+13-18h forecast	+13-18h fcfix	+13-18h fcday	+25-30h forecast	+25-30h fcfix
Accuracy	0.8009	0.8067	0.6367	0.8013	0.8059	0.6517	0.7784	0.7842
Bias frequency	0.8488	0.8767	1.0012	0.8449	0.8773	1.0034	0.8301	0.8578
Hit rate	0.7233	0.7430	0.6333	0.7245	0.7454	0.6556	0.6948	0.7145
False alarm ratio	0.1479	0.1525	0.3674	0.1425	0.1504	0.3466	0.1630	0.1671
Success ratio	0.8521	0.8475	0.6326	0.8575	0.8496	0.6534	0.8370	0.8329
Brier score	0.1991	0.1933	0.3633	0.1987	0.1941	0.3483	0.2216	0.2158
Brier skill score	0.5976	0.6093	0.2653	0.6010	0.6102	0.2989	0.5539	0.5656
MAE	0.6302	0.6866	1.1346	0.6640	0.7301	1.1268	0.7225	0.7961
RMSE	1.3014	1.4461	2.1140	1.4330	1.5763	2.2038	1.5706	1.7383

KRISTIANSAND	+4-9h forecast	+4-9h fcfix	+4-9h fcday	+13-18h forecast	+13-18h fcfix	+13-18h fcday	+25-30h forecast	+25-30h fcfix
Accuracy	0.8792	0.8792	0.6904	0.8624	0.8624	0.6922	0.8555	0.8567
Bias frequency	1.0619	1.0619	1.0021	1.0656	1.0656	1.0041	1.0393	1.0351
Hit rate	0.8206	0.8206	0.4628	0.7971	0.7971	0.4724	0.7707	0.7707
False alarm ratio	0.2272	0.2272	0.5381	0.2519	0.2519	0.5295	0.2584	0.2555
Success ratio	0.7728	0.7728	0.4619	0.7481	0.7481	0.4705	0.7416	0.7445
Brier score	0.1208	0.1208	0.3096	0.1376	0.1376	0.3078	0.1445	0.1433
Brier skill score	0.5794	0.5794	-0.0765	0.5286	0.5286	-0.0592	0.5021	0.5062
MAE	0.5581	0.5562	1.5195	0.5938	0.5919	1.5212	0.7215	0.7191
RMSE	1.6984	1.6933	3.9720	1.7265	1.7231	3.9041	2.2528	2.2440

NESBYEN	+4-9h forecast	+4-9h fcfix	+4-9h fcday	+13-18h forecast	+13-18h fcfix	+13-18h fcday	+25-30h forecast	+25-30h fcfix
Accuracy	0.8856	0.8851	0.7293	0.8806	0.8835	0.7408	0.8659	0.8654
Bias frequency	1.1729	1.1223	1.0027	1.1254	1.0826	0.9972	1.2104	1.1671
Hit rate	0.8218	0.7952	0.3750	0.7692	0.7550	0.3580	0.7723	0.7493
False alarm ratio	0.2993	0.2915	0.6260	0.3165	0.3026	0.6410	0.3619	0.3580
Success ratio	0.7007	0.7085	0.3740	0.6835	0.6974	0.3590	0.6381	0.6420
Brier score	0.1144	0.1149	0.2707	0.1194	0.1165	0.2592	0.1341	0.1346
Brier skill score	0.4706	0.4683	-0.2527	0.4133	0.4275	-0.2813	0.3342	0.3317
MAE	0.2571	0.2236	0.4491	0.2881	0.2555	0.5084	0.3089	0.2789
RMSE	0.8126	0.7277	1.3290	1.0590	0.9696	1.6939	1.0451	1.0066

Figure A.24: Forecast verification results for 6h accumulated precipitation for AROME, fcfix and fcday against observed data.

24h accumulated precipitation

BERGEN	+4-27h forecast	+4-27h fcfix	+13-36h forecast	+13-36h fcfix	+25-48h forecast	+25-48h fcfix
Accuracy	0.8860	0.8836	0.8895	0.8924	0.8718	0.8677
Bias frequency	0.9772	0.9941	0.9688	0.9848	0.9705	0.9797
Hit rate	0.9064	0.9132	0.9048	0.9149	0.8928	0.8945
False alarm ratio	0.0725	0.0814	0.0661	0.0710	0.0800	0.0870
Success ratio	0.9275	0.9186	0.9339	0.9290	0.9200	0.9130
Brier score	0.1140	0.1164	0.1105	0.1076	0.1282	0.1323
Brier skill score	0.6279	0.6201	0.6387	0.6481	0.5813	0.5679
MAE	3.3447	3.4420	3.4779	3.6344	3.5843	3.6949
RMSE	6.4460	6.3837	6.7029	6.7173	6.9803	7.0906

OSLO	+4-27h forecast	+4-27h fcfix	+13-36h forecast	+13-36h fcfix	+25-48h forecast	+25-48h fcfix
Accuracy	0.8386	0.8398	0.8281	0.8281	0.8296	0.8314
Bias frequency	0.9045	0.8947	0.9411	0.9362	0.9398	0.9312
Hit rate	0.7834	0.7797	0.7902	0.7877	0.7912	0.7887
False alarm ratio	0.1340	0.1286	0.1604	0.1586	0.1582	0.1530
Success ratio	0.8660	0.8714	0.8396	0.8414	0.8418	0.8470
Brier score	0.1614	0.1602	0.1719	0.1719	0.1704	0.1686
Brier skill score	0.6622	0.6647	0.6393	0.6393	0.6425	0.6462
MAE	1.1725	1.0958	1.2875	1.2072	1.3850	1.3118
RMSE	2.9952	2.8351	3.3207	3.1710	3.4196	3.2851

TRONDHEIM	+4-27h forecast	+4-27h fcfix	+13-36h forecast	+13-36h fcfix	+25-48h forecast	+25-48h fcfix
Accuracy	0.8587	0.8599	0.8683	0.8707	0.8559	0.8552
Bias frequency	1.0624	1.0542	1.0522	1.0399	1.0503	1.0410
Hit rate	0.9110	0.9080	0.9140	0.9099	0.9026	0.8974
False alarm ratio	0.1424	0.1387	0.1313	0.1250	0.1406	0.1379
Success ratio	0.8576	0.8613	0.8687	0.8750	0.8594	0.8621
Brier score	0.1413	0.1401	0.1317	0.1293	0.1441	0.1448
Brier skill score	0.6570	0.6599	0.6807	0.6865	0.6502	0.6485
MAE	1.5359	1.4153	1.5828	1.4655	1.7139	1.5962
RMSE	3.1713	2.9385	3.1692	2.9902	3.3483	3.1597

TROMSØ	+4-27h forecast	+4-27h fcfix	+13-36h forecast	+13-36h fcfix	+25-48h forecast	+25-48h fcfix
Accuracy	0.8609	0.8627	0.8527	0.8545	0.8449	0.8414
Bias frequency	0.9287	0.9426	0.9191	0.9362	0.9198	0.9345
Hit rate	0.8672	0.8754	0.8569	0.8667	0.8519	0.8568
False alarm ratio	0.0662	0.0713	0.0676	0.0742	0.0738	0.0832
Success ratio	0.9338	0.9287	0.9324	0.9258	0.9262	0.9168
Brier score	0.1391	0.1373	0.1473	0.1455	0.1551	0.1586
Brier skill score	0.5102	0.5165	0.4782	0.4846	0.4500	0.4376
MAE	1.8459	1.9186	1.9732	2.0412	2.1217	2.2623
RMSE	3.2807	3.4204	3.5421	3.7061	3.8356	4.0957

KRISTIANSAND	+4-27h forecast	+4-27h fcfix	+13-36h forecast	+13-36h fcfix	+25-48h forecast	+25-48h fcfix
Accuracy	0.8816	0.8816	0.8622	0.8634	0.8552	0.8571
Bias frequency	1.0453	1.0453	1.0333	1.0309	1.0107	1.0071
Hit rate	0.9058	0.9058	0.8812	0.8812	0.8631	0.8631
False alarm ratio	0.1334	0.1334	0.1471	0.1452	0.1461	0.1430
Success ratio	0.8666	0.8666	0.8529	0.8548	0.8539	0.8570
Brier score	0.1184	0.1184	0.1378	0.1366	0.1448	0.1429
Brier skill score	0.7599	0.7599	0.7195	0.7219	0.7052	0.7091
MAE	1.6441	1.6390	1.8414	1.8371	2.0596	2.0539
RMSE	3.6846	3.6761	4.1396	4.1281	4.5631	4.5524

NESBYEN	+4-27h forecast	+4-27h fcfix	+13-36h forecast	+13-36h fcfix	+25-48h forecast	+25-48h fcfix
Accuracy	0.8398	0.8415	0.8333	0.8368	0.8290	0.8349
Bias frequency	1.1041	1.0689	1.0685	1.0462	1.0728	1.0476
Hit rate	0.8594	0.8439	0.8350	0.8280	0.8319	0.8263
False alarm ratio	0.2217	0.2105	0.2186	0.2086	0.2245	0.2112
Success ratio	0.7783	0.7895	0.7814	0.7914	0.7755	0.7888
Brier score	0.1602	0.1585	0.1667	0.1632	0.1710	0.1651
Brier skill score	0.6147	0.6188	0.6013	0.6097	0.5909	0.6050
MAE	0.8070	0.7216	0.8570	0.7749	0.8919	0.8068
RMSE	1.9596	1.8385	2.1111	1.9705	2.1926	2.0389

Figure A.25: Forecast verification results for 24h accumulated precipitation for AROME and fcfix against observed data.

48h accumulated precipitation

BERGEN	+1-48h forecast	+1-48h fcfix
Accuracy	0.9186	0.9150
Bias frequency	0.9784	0.9856
Hit rate	0.9394	0.9409
False alarm ratio	0.0398	0.0454
Success ratio	0.9602	0.9546
Brier score	0.0814	0.0850
Brier skill score	0.5520	0.5322
MAE	5.8163	5.7844
RMSE	10.2717	9.7218

TROMSØ	+1-48h forecast	+1-48h fcfix
Accuracy	0.8897	0.8862
Bias frequency	0.9505	0.9590
Hit rate	0.9095	0.9117
False alarm ratio	0.0431	0.0494
Success ratio	0.9569	0.9506
Brier score	0.1103	0.1138
Brier skill score	0.3158	0.2940
MAE	3.2995	3.3327
RMSE	5.3788	5.4488

OSLO	+1-48h forecast	+1-48h fcfix
Accuracy	0.8236	0.8242
Bias frequency	0.9181	0.9080
Hit rate	0.8215	0.8169
False alarm ratio	0.1052	0.1003
Success ratio	0.8948	0.8997
Brier score	0.1764	0.1758
Brier skill score	0.5082	0.5099
MAE	2.0967	1.9610
RMSE	4.4497	4.2551

KRISTIANSAND	+1-48h forecast	+1-48h fcfix
Accuracy	0.8846	0.8846
Bias frequency	0.9898	0.9898
Hit rate	0.9073	0.9073
False alarm ratio	0.0833	0.0833
Success ratio	0.9167	0.9167
Brier score	0.1154	0.1154
Brier skill score	0.6619	0.6619
MAE	2.9517	2.9442
RMSE	5.7004	5.6865

TRONDHEIM	+1-48h forecast	+1-48h fcfix
Accuracy	0.9015	0.8997
Bias frequency	1.0230	1.0156
Hit rate	0.9449	0.9400
False alarm ratio	0.0763	0.0744
Success ratio	0.9237	0.9256
Brier score	0.0985	0.1003
Brier skill score	0.6214	0.6145
MAE	2.6571	2.4581
RMSE	4.8028	4.4721

NESBYEN	+1-48h forecast	+1-48h fcfix
Accuracy	0.8336	0.8360
Bias frequency	1.0489	1.0346
Hit rate	0.8809	0.8758
False alarm ratio	0.1602	0.1535
Success ratio	0.8398	0.8465
Brier score	0.1664	0.1640
Brier skill score	0.6044	0.6101
MAE	1.4713	1.3011
RMSE	3.1394	2.8493

Figure A.26: Forecast verification results for 48h accumulated precipitation for AROME and fcfix against observed data.

1h accumulated precipitation

forecast

BERGEN	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h
Accuracy	0.8171	0.8131	0.8194	0.8222	0.8091	0.8171	0.8045	0.8108	0.7884	0.8073	0.7999	0.7999
Bias frequency	0.7530	0.7719	0.7853	0.7244	0.7206	0.7593	0.7238	0.7432	0.7130	0.7126	0.7089	0.7203
Hit rate	0.5562	0.5540	0.5880	0.5736	0.5395	0.5675	0.5371	0.5506	0.5148	0.5256	0.5089	0.5091
False alarm ratio	0.2613	0.2823	0.2512	0.2082	0.2513	0.2526	0.2579	0.2592	0.2779	0.2624	0.2821	0.2933
Success ratio	0.7387	0.7177	0.7488	0.7918	0.7487	0.7474	0.7421	0.7408	0.7221	0.7376	0.7179	0.7067
Brier score	0.1830	0.1870	0.1810	0.1780	0.1910	0.1830	0.1960	0.1890	0.2120	0.1930	0.2000	0.2000
Brier skill score	0.3590	0.3360	0.3890	0.4220	0.3580	0.3750	0.3490	0.3590	0.3150	0.3370	0.3090	0.2980
MAE	0.2684	0.2622	0.2753	0.2856	0.2695	0.2776	0.2731	0.2683	0.2798	0.2567	0.2723	0.2822
RMSE	0.7849	0.7058	0.7649	0.7484	0.7157	0.7187	0.7196	0.7307	0.7648	0.6800	0.7142	0.8017

OSLO	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h
Accuracy	0.9197	0.9134	0.9197	0.9128	0.9237	0.9163	0.9157	0.9151	0.9174	0.9260	0.9243	0.9128
Bias frequency	0.9378	0.9156	0.8257	0.9103	0.8646	0.8957	0.8584	0.9648	0.9907	0.9771	0.9558	1.0087
Hit rate	0.6578	0.6222	0.6224	0.6143	0.6419	0.6304	0.6137	0.6564	0.6589	0.6927	0.6858	0.6753
False alarm ratio	0.2986	0.3204	0.2462	0.3251	0.2576	0.2961	0.2850	0.3196	0.3349	0.2911	0.2824	0.3305
Success ratio	0.7014	0.6796	0.7538	0.6749	0.7424	0.7039	0.7150	0.6804	0.6651	0.7089	0.7176	0.6695
Brier score	0.0800	0.0870	0.0800	0.0870	0.0760	0.0840	0.0840	0.0850	0.0830	0.0740	0.0760	0.0870
Brier skill score	0.3800	0.3260	0.4210	0.3200	0.4210	0.3630	0.3710	0.3470	0.3240	0.4080	0.4140	0.3430
MAE	0.0907	0.0870	0.0882	0.0910	0.0932	0.0931	0.0988	0.0851	0.0919	0.0955	0.1040	0.0966
RMSE	0.4081	0.3811	0.4213	0.4106	0.4374	0.3804	0.5232	0.4931	0.3924	0.6935	0.5014	0.4386

TRONDHEIM	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h
Accuracy	0.8409	0.8514	0.8620	0.8385	0.8479	0.8573	0.8397	0.8362	0.8456	0.8420	0.8514	0.8450
Bias frequency	0.9894	1.0034	0.9966	0.9293	1.0102	0.9338	0.9474	0.9658	1.0582	0.9519	0.9562	0.9465
Hit rate	0.5159	0.5670	0.5986	0.5017	0.5646	0.5836	0.5511	0.5497	0.5788	0.5449	0.5522	0.5318
False alarm ratio	0.4786	0.4349	0.3993	0.4601	0.4411	0.3750	0.4183	0.4309	0.4531	0.4276	0.4225	0.4382
Success ratio	0.5214	0.5651	0.6007	0.5399	0.5589	0.6250	0.5817	0.5691	0.5469	0.5724	0.5775	0.5618
Brier score	0.1590	0.1490	0.1380	0.1610	0.1520	0.1430	0.1600	0.1640	0.1540	0.1580	0.1490	0.1550
Brier skill score	0.0430	0.1280	0.2000	0.0770	0.1190	0.2320	0.1570	0.1330	0.1020	0.1380	0.1460	0.1170
MAE	0.1146	0.1128	0.1188	0.1263	0.1343	0.1265	0.1410	0.1582	0.1368	0.1352	0.1229	0.1316
RMSE	0.3524	0.3628	0.3659	0.4225	0.3970	0.4335	0.4469	0.6409	0.4617	0.4471	0.4092	0.4249

TROMSØ	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h
Accuracy	0.7922	0.7933	0.7922	0.7865	0.7962	0.7796	0.7847	0.7939	0.7675	0.7721	0.7721	0.7784
Bias frequency	0.6125	0.6315	0.6500	0.6750	0.6817	0.6151	0.6979	0.6869	0.6355	0.6809	0.6431	0.6180
Hit rate	0.4521	0.4702	0.4769	0.4778	0.5066	0.4622	0.4972	0.5028	0.4313	0.4565	0.4427	0.4476
False alarm ratio	0.2620	0.2553	0.2663	0.2923	0.2569	0.2485	0.2876	0.2680	0.3213	0.3295	0.3116	0.2758
Success ratio	0.7380	0.7447	0.7337	0.7077	0.7431	0.7515	0.7124	0.7320	0.6787	0.6705	0.6884	0.7242
Brier score	0.2080	0.2070	0.2080	0.2140	0.2040	0.2200	0.2150	0.2060	0.2320	0.2280	0.2280	0.2220
Brier skill score	0.2910	0.3080	0.3030	0.2790	0.3310	0.3110	0.2970	0.3190	0.2290	0.2320	0.2420	0.2760
MAE	0.1459	0.1479	0.1469	0.1576	0.1593	0.1538	0.1495	0.1594	0.1441	0.1564	0.1472	0.1486
RMSE	0.4333	0.4327	0.4229	0.4032	0.4252	0.3805	0.3628	0.4661	0.3652	0.3904	0.3582	0.3944

KRISTIANSAND	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h
Accuracy	0.9114	0.8960	0.9025	0.8972	0.8984	0.8919	0.8931	0.9019	0.8866	0.9019	0.8901	0.8860
Bias frequency	0.9611	0.9300	0.9644	1.0847	1.0741	1.0356	0.9962	1.0714	1.0866	1.0226	0.9778	0.9963
Hit rate	0.6887	0.6226	0.6561	0.6737	0.6831	0.6561	0.6553	0.7063	0.6654	0.6992	0.6444	0.6434
False alarm ratio	0.2834	0.3305	0.3197	0.3789	0.3640	0.3664	0.3422	0.3407	0.3877	0.3162	0.3409	0.3542
Success ratio	0.7166	0.6695	0.6803	0.6211	0.6360	0.6336	0.6578	0.6593	0.6123	0.6838	0.6591	0.6458
Brier score	0.0890	0.1040	0.0970	0.1030	0.1020	0.1080	0.1070	0.0980	0.1130	0.0980	0.1100	0.1140
Brier skill score	0.4140	0.3150	0.3510	0.2610	0.2890	0.2770	0.3140	0.3410	0.2470	0.3760	0.3100	0.2910
MAE	0.1353	0.1452	0.1574	0.1392	0.1306	0.1349	0.1413	0.1509	0.1489	0.1501	0.1456	0.1656
RMSE	0.5239	0.5398	0.6302	0.5205	0.5046	0.5031	0.5828	0.6302	0.5159	0.5478	0.5223	0.5772

NESBYEN	+1h	+2h	+3h	+4h	+5h	+6h	+7h	+8h	+9h	+10h	+11h	+12h
Accuracy	0.9272	0.9306	0.9329	0.9197	0.9140	0.9157	0.9123	0.9111	0.9209	0.9209	0.9260	0.9346
Bias frequency	1.0778	1.0581	1.0783	1.3333	1.2733	1.1488	1.1307	1.1000	1.0234	1.0118	1.0643	1.0732
Hit rate	0.6587	0.6387	0.6867	0.7179	0.6708	0.6369	0.6307	0.5941	0.6082	0.6000	0.6550	0.6890
False alarm ratio	0.3889	0.3963	0.3631	0.4615	0.4732	0.4456	0.4422	0.4599	0.4057	0.4070	0.3846	0.3580
Success ratio	0.6111	0.6037	0.6369	0.5385	0.5268	0.5544	0.5578	0.5401	0.5943	0.5930	0.6154	0.6420
Brier score	0.0730	0.0690	0.0670	0.0800	0.0860	0.0840	0.0880	0.0890	0.0790	0.0790	0.0740	0.0650
Brier skill score	0.2380	0.2240	0.2960	0.1050	0.0680	0.1280	0.1280	0.0870	0.1950	0.1900	0.2460	0.3090
MAE	0.0685	0.0635	0.0632	0.0671	0.0626	0.0549	0.0564	0.0535	0.0589	0.0631	0.0581	0.0623
RMSE	0.4588	0.3622	0.3675	0.3410	0.2766	0.2264	0.2292	0.2168	0.2548	0.3950	0.3229	0.3756

Figure A.27: Forecast verification results for 1h accumulated precipitation at each for the first 12 forecast hours, for AROME against observed data.

				1h accumulated precipitation								
						fcpersist						
BERGEN	**+1h**	**+2h**	**+3h**	**+4h**	**+5h**	**+6h**	**+7h**	**+8h**	**+9h**	**+10h**	**+11h**	**+12h**
Accuracy	0.8819	0.8481	0.7907	0.7781	0.7655	0.7609	0.7345	0.7202	0.7282	0.7225	0.7139	0.7036
Bias frequency	1.0000	1.0143	0.9632	0.9274	0.9595	0.9746	0.9486	0.9689	0.9222	0.9803	0.9861	1.0020
Hit rate	0.7932	0.7373	0.6286	0.6034	0.5857	0.5793	0.5333	0.5097	0.5222	0.5138	0.4990	0.4809
False alarm ratio	0.2068	0.2731	0.3474	0.3494	0.3896	0.4056	0.4378	0.4739	0.4337	0.4759	0.4940	0.5201
Success ratio	0.7932	0.7269	0.6526	0.6506	0.6104	0.5944	0.5622	0.5261	0.5663	0.5241	0.5060	0.4799
Brier score	0.1180	0.1520	0.2090	0.2220	0.2350	0.2390	0.2650	0.2800	0.2720	0.2780	0.2860	0.2960
Brier skill score	0.5870	0.4600	0.2950	0.2790	0.2100	0.1840	0.1200	0.0500	0.1210	0.0460	0.0120	-0.0390
MAE	0.2514	0.3023	0.3511	0.3826	0.4149	0.4118	0.4281	0.4318	0.4442	0.4364	0.4522	0.4632
RMSE	0.7327	0.8760	0.9639	1.0049	1.0475	1.0538	1.0894	1.0668	1.1251	1.0811	1.1108	1.1670
OSLO	**+1h**	**+2h**	**+3h**	**+4h**	**+5h**	**+6h**	**+7h**	**+8h**	**+9h**	**+10h**	**+11h**	**+12h**
Accuracy	0.9266	0.9002	0.8784	0.8670	0.8532	0.8481	0.8372	0.8406	0.8354	0.8274	0.8194	0.8222
Bias frequency	1.0356	1.0356	0.9668	1.0448	1.0175	1.0130	1.0000	1.0264	1.0888	1.0688	1.0310	1.0087
Hit rate	0.7333	0.6311	0.5436	0.5022	0.4498	0.4304	0.3906	0.4009	0.3738	0.3440	0.3186	0.3333
False alarm ratio	0.2918	0.3906	0.4378	0.5193	0.5579	0.5751	0.6094	0.6094	0.6567	0.6781	0.6910	0.6695
Success ratio	0.7082	0.6094	0.5622	0.4807	0.4421	0.4249	0.3906	0.3906	0.3433	0.3219	0.3090	0.3305
Brier score	0.0730	0.1000	0.1220	0.1330	0.1470	0.1520	0.1630	0.1590	0.1650	0.1730	0.1810	0.1780
Brier skill score	0.4340	0.2250	0.1170	-0.0400	-0.1200	-0.1520	-0.2200	-0.2210	-0.3450	-0.3840	-0.3970	-0.3430
MAE	0.0876	0.1108	0.1292	0.1303	0.1295	0.1425	0.1548	0.1486	0.1533	0.1641	0.1682	0.1765
RMSE	0.4425	0.5133	0.5552	0.5674	0.5288	0.5804	0.6601	0.6902	0.6199	0.6395	0.7081	0.7095
TRONDHEIM	**+1h**	**+2h**	**+3h**	**+4h**	**+5h**	**+6h**	**+7h**	**+8h**	**+9h**	**+10h**	**+11h**	**+12h**
Accuracy	0.8882	0.8482	0.8488	0.8353	0.8241	0.7988	0.7929	0.7935	0.7865	0.7806	0.7847	0.7847
Bias frequency	1.0636	1.0344	1.0238	1.0135	1.0238	0.9495	0.9319	0.9348	1.0308	0.9647	1.0135	1.0067
Hit rate	0.6961	0.5739	0.5748	0.5354	0.5034	0.4353	0.4211	0.4224	0.3938	0.3846	0.3906	0.3913
False alarm ratio	0.3455	0.4452	0.4385	0.4718	0.5083	0.5415	0.5482	0.5482	0.6179	0.6013	0.6146	0.6113
Success ratio	0.6545	0.5548	0.5615	0.5282	0.4917	0.4585	0.4518	0.4518	0.3821	0.3987	0.3854	0.3887
Brier score	0.1120	0.1520	0.1510	0.1650	0.1760	0.2010	0.2070	0.2060	0.2140	0.2190	0.2150	0.2150
Brier skill score	0.3270	0.1120	0.1270	0.0560	-0.0180	-0.0780	-0.0890	-0.0880	-0.2460	-0.1930	-0.2310	-0.2220
MAE	0.0929	0.1367	0.1413	0.1423	0.1605	0.1618	0.1781	0.1884	0.1805	0.1834	0.1771	0.1775
RMSE	0.3258	0.4748	0.4650	0.4440	0.5021	0.4785	0.5520	0.6706	0.5847	0.5502	0.5466	0.5568
TROMSØ	**+1h**	**+2h**	**+3h**	**+4h**	**+5h**	**+6h**	**+7h**	**+8h**	**+9h**	**+10h**	**+11h**	**+12h**
Accuracy	0.8627	0.8254	0.7984	0.7806	0.7542	0.7519	0.7421	0.7295	0.7329	0.7231	0.7134	0.7007
Bias frequency	1.0373	1.0154	1.0173	1.0232	0.9962	0.9532	0.9944	1.0057	1.0095	1.0232	1.0095	0.9906
Hit rate	0.7843	0.7159	0.6712	0.6422	0.5951	0.5874	0.5752	0.5551	0.5611	0.5455	0.5286	0.5075
False alarm ratio	0.2439	0.2949	0.3403	0.3724	0.4026	0.3837	0.4216	0.4480	0.4442	0.4669	0.4764	0.4877
Success ratio	0.7561	0.7051	0.6597	0.6276	0.5974	0.6163	0.5784	0.5520	0.5558	0.5331	0.5236	0.5123
Brier score	0.1370	0.1750	0.2020	0.2190	0.2460	0.2480	0.2580	0.2710	0.2670	0.2770	0.2870	0.2990
Brier skill score	0.5320	0.4150	0.3240	0.2630	0.1930	0.2220	0.1560	0.1030	0.1130	0.0670	0.0470	0.0250
MAE	0.1315	0.1491	0.1569	0.1733	0.1819	0.1859	0.1994	0.1985	0.1969	0.1990	0.1952	0.2071
RMSE	0.3921	0.4214	0.4237	0.4526	0.4870	0.4606	0.5002	0.4906	0.5082	0.5086	0.4821	0.5188
KRISTIANSAND	**+1h**	**+2h**	**+3h**	**+4h**	**+5h**	**+6h**	**+7h**	**+8h**	**+9h**	**+10h**	**+11h**	**+12h**
Accuracy	0.9112	0.8740	0.8538	0.8533	0.8361	0.8337	0.8272	0.8189	0.8142	0.8089	0.8006	0.7882
Bias frequency	1.0625	1.0584	1.0751	1.1525	1.1193	1.0751	1.0303	1.0794	1.0709	1.0264	1.0112	1.0000
Hit rate	0.7383	0.6148	0.5494	0.5508	0.4897	0.4822	0.4621	0.4325	0.4173	0.4038	0.3792	0.3419
False alarm ratio	0.3051	0.4191	0.4890	0.5221	0.5625	0.5515	0.5515	0.5993	0.6103	0.6066	0.6250	0.6581
Success ratio	0.6949	0.5809	0.5110	0.4779	0.4375	0.4485	0.4485	0.4007	0.3897	0.3934	0.3750	0.3419
Brier score	0.0890	0.1260	0.1460	0.1470	0.1640	0.1660	0.1730	0.1810	0.1860	0.1910	0.1990	0.2120
Brier skill score	0.4130	0.1720	0.0250	-0.0530	-0.1400	-0.1090	-0.1080	-0.2140	-0.2380	-0.2180	-0.2500	-0.3180
MAE	0.1292	0.1679	0.2072	0.2033	0.2266	0.2327	0.2399	0.2683	0.2534	0.2603	0.2663	0.2742
RMSE	0.5210	0.6012	0.7584	0.6739	0.7787	0.7688	0.7733	0.9047	0.7922	0.8281	0.8455	0.8056
NESBYEN	**+1h**	**+2h**	**+3h**	**+4h**	**+5h**	**+6h**	**+7h**	**+8h**	**+9h**	**+10h**	**+11h**	**+12h**
Accuracy	0.9358	0.9243	0.9100	0.9054	0.8911	0.8870	0.8756	0.8641	0.8589	0.8503	0.8578	0.8538
Bias frequency	0.9760	1.0516	0.9819	1.0449	1.0124	0.9702	0.9261	0.9588	0.9532	0.9588	0.9532	0.9939
Hit rate	0.6527	0.6000	0.5181	0.4936	0.4161	0.3988	0.3466	0.2824	0.2573	0.2118	0.2515	0.2195
False alarm ratio	0.3313	0.4294	0.4724	0.5276	0.5890	0.5890	0.6258	0.7055	0.7301	0.7791	0.7362	0.7791
Success ratio	0.6687	0.5706	0.5276	0.4724	0.4110	0.4110	0.3742	0.2945	0.2699	0.2209	0.2638	0.2209
Brier score	0.0640	0.0760	0.0900	0.0950	0.1090	0.1130	0.1240	0.1360	0.1410	0.1500	0.1420	0.1460
Brier skill score	0.3320	0.1450	0.0550	-0.0630	-0.1810	-0.1730	-0.2290	-0.3950	-0.4370	-0.5380	-0.4480	-0.5530
MAE	0.0685	0.0739	0.0832	0.0829	0.0788	0.0890	0.0937	0.0905	0.0929	0.0987	0.0987	0.1109
RMSE	0.4848	0.4522	0.4710	0.4743	0.4352	0.4719	0.4819	0.4683	0.4732	0.5130	0.5164	0.5923

Figure A.28: Forecast verification results for 1h accumulated precipitation at each for the first 12 forecast hours, for fcpersist against observed data.

6h accumulated precipitation

	forecast		*fcpersist*	
BERGEN	**+1-6h**	**+7-12h**	**+1-6h**	**+7-12h**
Accuracy	0.8604	0.8254	0.7697	0.6847
Bias frequency	0.9296	0.8712	0.5931	0.5820
Hit rate	0.8198	0.7576	0.5573	0.4696
False alarm ratio	0.1181	0.1304	0.0604	0.1932
Success ratio	0.8819	0.8696	0.9396	0.8068
Brier score	0.1400	0.1750	0.2300	0.3150
Brier skill score	0.7090	0.6430	0.5220	0.3580
MAE	1.1367	1.1396	1.8605	2.4856
RMSE	2.5809	2.6158	4.7825	5.9185

	forecast		*fcpersist*	
TROMSØ	**+1-6h**	**+7-12h**	**+1-6h**	**+7-12h**
Accuracy	0.8028	0.8189	0.7555	0.6582
Bias frequency	0.8322	0.8769	0.6011	0.6060
Hit rate	0.7203	0.7572	0.5577	0.4608
False alarm ratio	0.1344	0.1365	0.0722	0.2395
Success ratio	0.8656	0.8635	0.9278	0.7605
Brier score	0.1970	0.1810	0.2450	0.3420
Brier skill score	0.6030	0.6380	0.5070	0.3150
MAE	0.6364	0.6255	0.8203	1.0920
RMSE	1.3825	1.2714	2.0821	2.6022

	forecast		*fcpersist*	
OSLO	**+1-6h**	**+7-12h**	**+1-6h**	**+7-12h**
Accuracy	0.8811	0.8880	0.8300	0.7725
Bias frequency	0.9176	1.0023	0.5189	0.5381
Hit rate	0.7283	0.7760	0.4298	0.3118
False alarm ratio	0.2063	0.2258	0.1717	0.4206
Success ratio	0.7937	0.7742	0.8283	0.5794
Brier score	0.1190	0.1120	0.1700	0.2270
Brier skill score	0.5390	0.5500	0.3410	0.0870
MAE	0.3965	0.3831	0.6615	0.9160
RMSE	1.3738	1.3011	2.7557	3.4557

	forecast		*fcpersist*	
KRISTIANSAND	**+1-6h**	**+7-12h**	**+1-6h**	**+7-12h**
Accuracy	0.8864	0.8775	0.8257	0.7611
Bias frequency	1.0204	1.0430	0.5551	0.5585
Hit rate	0.8147	0.8094	0.4776	0.3655
False alarm ratio	0.2016	0.2240	0.1397	0.3456
Success ratio	0.7984	0.7760	0.8603	0.6544
Brier score	0.1140	0.1220	0.1740	0.2390
Brier skill score	0.6080	0.5780	0.4010	0.1720
MAE	0.5576	0.5667	1.0296	1.4972
RMSE	1.7189	1.6375	3.3472	4.2194

	forecast		*fcpersist*	
TRONDHEIM	**+1-6h**	**+7-12h**	**+1-6h**	**+7-12h**
Accuracy	0.8300	0.8435	0.8020	0.7177
Bias frequency	1.1086	1.0618	0.5172	0.4878
Hit rate	0.8052	0.8146	0.4690	0.3545
False alarm ratio	0.2737	0.2328	0.0933	0.2733
Success ratio	0.7263	0.7672	0.9067	0.7267
Brier score	0.1700	0.1560	0.1980	0.2820
Brier skill score	0.5020	0.5690	0.4210	0.2220
MAE	0.4803	0.5671	0.7561	1.0269
RMSE	1.2237	1.6236	2.2674	2.8772

	forecast		*fcpersist*	
NESBYEN	**+1-6h**	**+7-12h**	**+1-6h**	**+7-12h**
Accuracy	0.8851	0.8811	0.8627	0.7967
Bias frequency	1.1534	1.1178	0.4631	0.4466
Hit rate	0.7926	0.7753	0.3920	0.2384
False alarm ratio	0.3128	0.3064	0.1534	0.4663
Success ratio	0.6872	0.6936	0.8466	0.5337
Brier score	0.1150	0.1190	0.1370	0.2030
Brier skill score	0.4310	0.4320	0.3220	0.0310
MAE	0.2747	0.2540	0.4222	0.5666
RMSE	1.1081	1.0012	2.3838	2.7293

Figure A.29: Forecast verification results for 6h accumulated precipitation for AROME and fcpersist against observed data.