NTNU

Kunnskap for en bedre verden

Department of Mathematical Sciences

TMA4500 - Fordypningsprosjekt

Probabilistic Forecasting of Solar Radiation in Norway Using Bayesian Model Averaging for Ensemble Forecasts

Author:
André Egeli

December 2022

Abstract

In this study, we utilized Bayesian Model Averaging (BMA) to improve the accuracy of ensemble solar radiation forecasts for 15 locations in Norway. The forecast period was from January 1, 2020, to July 1, 2022, and the ensemble forecasts consisted of six members. We evaluated the performance of five different kernel distributions for use in the BMA method and assessed the probabilistic calibration using the probability integral transform (PIT) histogram and the predictive performance using the continuously ranked probability score (CRPS). We also evaluated the deterministic accuracy of the post-processed forecasts using the root-mean-squared error (RMSE). Our results showed that all of the kernel distributions were able to calibrate the forecasts, with the mixture of Beta distributions scaled by the physical domain of solar radiation performing the best in terms of improving the predictive distribution.

Table of Contents

1 Introduction

Solar radiation forecasting has not traditionally been a focus for the meteorological community in Norway. Global solar radiation maps, such as those provided by Solargis (sol, c) and GlobalSolarAtlas (sol, b), only cover the southern part of the country, and the Norwegian Meteorological Institute does not offer solar radiation forecasts to the public (yr.no). However, there is increasing interest in solar radiation forecasts in Norway due to the potential for solar energy to play a significant role in the country's energy mix. A study by Multiconsult suggests that solar energy could be as important for Norway as hydropower is currently (Cato, 2022). Accurate solar radiation forecasts are important for predicting the power output of solar panels, as there is a direct relationship between solar radiation and energy production (Ettah and Obiefuna, 2011). Solar forecasting also enables grid operators to balance energy generation and consumption and is considered a crucial aspect of integrating solar energy into power grids(Zwane et al., 2022).

Probabilistic forecasting involves assigning a probability distribution over potential outcomes to summarize future events. This complete set of probabilities is known as a probabilistic forecast (Probabilistic forecasting). The aim of probabilistic forecasting is to optimize the sharpness of the forecast probability density function (pdf) while ensuring calibration (Raftery et al., 2005). Calibration refers to the consistency between the forecast pdf and the observed outcomes. A forecast is considered calibrated if an event with probability p of occurring, actually occurs a proportion of p times on average. Sharpness refers to the spread of the forecast pdf. A forecast is considered sharp if the prediction intervals are narrower than those obtained using simple methods such as climatological data. The more concentrated the forecast, the sharper it is.

The physical property we consider is solar radiation, specifically Global Horizontal Irradiance (GHI). GHI is defined as the total solar radiation incident on a horizontal surface and is composed of direct normal radiation (DNI) and diffuse radiation (DHI), as well as ground-reflected radiation (ghi). GHI is predictable due to its strong dependence on the sun's angle in the sky. By considering the angle of the sun, the time of year, and accounting for diffuse and reflected radiation, it is possible to calculate the physical limits of solar radiation. (C. N. Long).

There are various methods for forecasting solar radiation, the suitability of which depends on the intended forecast lead time. Forecast lead time refers to the time between the issuance of a forecast and the occurrence of the predicted phenomena. Figure 1 presents a conceptual visualization of the forecast skill of each method as a function of lead time. At sub-hourly lead times, methods based on observations have a high skill as weather patterns change relatively little over short periods. There is a wealth of research on short-term solar radiation prediction using in-situ ground observations, including statistical methods such as auto-regressive or state-space models, as well as machine learning techniques like random forests and artificial neural networks (Benavides Cesar et al., 2022). When the forecast lead time is several hours, relying on observed values alone becomes unreliable, and other methods must be used to forecast solar radiation.

Geostationary satellites are a common method for measuring and predicting solar radiation. To predict solar radiation using geostationary satellites, a satellite cloud motion vector (CMV) is estimated from satellite images, and the future position of clouds is inferred to compute the solar radiation (sol, a). While the accuracy of these methods is lower than those based on observations, they have the advantage of large spatial and temporal coverage (Noia et al., 1993). However, geostationary satellites, which orbit the Earth in a circular geosynchronous orbit in the plane of the equator, always measure from the same angle, leading to decreased spatial resolution and increased errors at latitudes higher than 65 degrees (Babar, 2019). This makes the use of geostationary satellites to measure and predict solar radiation in Norway problematic and necessitates the use of alternative methods.

Numerical weather forecasting (NWP) is a method of predicting large-scale trends in the atmosphere using physical relationships. NWPs are often used to predict solar radiation over longer time horizons, but they can be affected by model biases and local weather conditions (sol, a). In northern Europe, the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Meteorological Cooperation on Operational Numeric Weather Prediction (MetCoOp) are two major providers of NWP. These forecasts often consist of multiple ensemble members, each generated
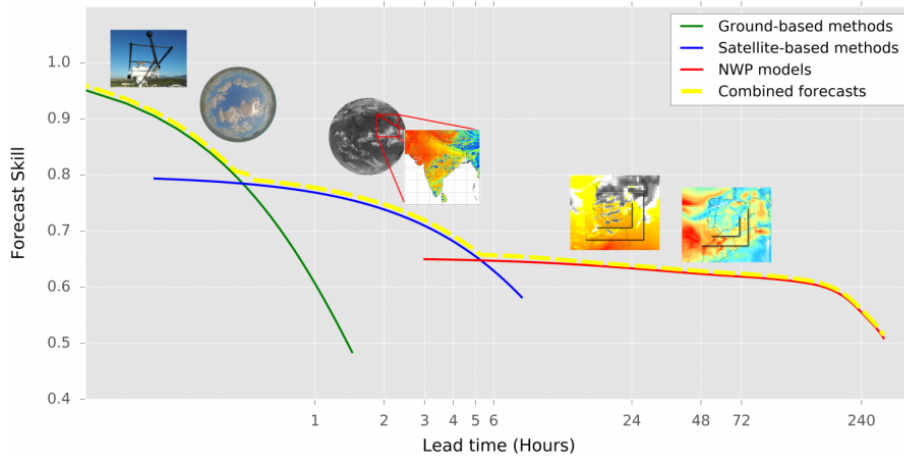
Figure 1: A conceptual plot of forecast skill compared to lead time for the different techniques that exist sol (d).

from different initial conditions. However, ensemble forecasts can be under-dispersive (meaning the actual observation falls outside the ensemble range) and biased (meaning they have systematic errors) (Gneiting et al., 2005). Statistical post-processing of ensemble forecasts is therefore essential for obtaining reliable forecasts, and there exist many different applicable methods. Examples include Bayesian Model Averaging (Gneiting et al., 2005), Model Output Statistics (Verzijlbergh et al., 2015), Quantile Regression models (Lauret et al., 2017) and finally, (Verbois et al., 2022), where 16 different machine learning methods for post-processing ensemble forecasts of solar radiation were compared.

In this study, Bayesian Model Averaging (BMA) is used to combine multiple ensemble forecasts into a probabilistic prediction. BMA treats each ensemble member as a statistical model and combines them through a weighted average. The first application of BMA in ensemble weather forecasting was in the post-processing of temperature forecasts, using a Gaussian distribution as the kernel (Gneiting et al., 2005). (Kleiven, 2017) took this idea further when post-processing ensemble forecasts for precipitation, and based the BMA kernel on a unit-scale Beta distribution, while also considering the climatology of precipitation. This type of statistical post-processing for ensemble forecasts is commonly used for temperature, precipitation, and wind speed, but is not a widely adopted method for solar radiation forecasting. However, solar radiation has some similarities to precipitation, given that the numeric values of both are non-negative and bounded by physical limits. We introduce two novel methods that truncate the Gaussian and Cauchy distributions by the physical domain of solar radiation and use them as kernels in the BMA context. We also utilize a Beta-distribution, scaled by the physical possible domain, as the kernel. To the best of our knowledge, these methods have not been previously explored in the literature.

To evaluate the probabilistic solar radiation forecasts generated by different methods, we focus on the measures of sharpness and calibration. The Probability Integral Transform (PIT) histogram is used to assess the calibration of the forecasts (Raftery et al., 2005). The PIT is a measure of the accuracy of a model's predictions by comparing the predicted probabilities of an event occurring with the actual frequency of that event. The PIT histogram is a graphical representation of the PIT values and can be used to identify areas where the model may need to be improved, as well as to compare the performance of different models and evaluate the reliability of their predictions. To quantitatively compare and rank the competing models, we use the Continuous Ranked Probability Score (CRPS). CRPS is a statistical measure that evaluates the accuracy of ensemble forecasts by measuring the difference between the predicted probabilities of an event occurring and the observed frequency of that event. It is therefore a method to assess both calibration and sharpness. The Continuous Ranked Probability Skill Score (CRPSS) is used to compare how well a model performs over a baseline model, where the baseline model is a simple model like the raw ensembles or using the climatology. The CRPSS then represents the increase, or decrease, in the accuracy of the post-processed models compared to the baseline model. Lastly, root-mean-squared error (RMSE)

is used to measure the deterministic accuracy of a point forecast.

The objective of this study is threefold. First, we aim to assess the quality of current NWPs for solar radiation in Norway. Second, we aim to generate calibrated probabilistic forecasts of solar radiation, which provide additional information about the uncertainty of the prediction. This is especially important in the context of solar forecasting, as prediction intervals are necessary to mitigate risk in grid operations (David and Lauret, 2018). Finally, we aim to evaluate and propose several methods for post-processing solar radiation ensemble forecasts.

In this study, we use ensemble forecasts from the MEPS and compare them to in-situ ground observations at 15 locations in Norway. In Chapter 2, we provide an overview of the theories relevant to solar radiation, probabilistic forecasting, and the models used in this study. In Chapter 3, we present and visualize the data and conduct a brief exploratory analysis of the ensemble members and errors. In Chapter 4, we describe the methodology behind the post-processing methods and provide an overview of how the models are generated. The results of the post-processed forecasts are presented in Chapter 5. Finally, in Chapter 6, we discuss the results and provide a conclusion.

2 Background

This Section provides the theoretical background for solar radiation, the models used, and the distributions applied in this study. We first introduce the physical quantity of solar radiation. Then, we present the theories behind probabilistic forecasting and the scoring rules used to evaluate the models. Finally, we outline the theory behind the post-processing models. For notation, we use x_{stm} to denote the ensemble number m at location s, forecasted for time t, and y_{st} to denote the corresponding observation at location s measured at time t. In this Section, we omit the notation for s and t as we primarily focus on a single forecast at a single location. However, some equations involve multiple training days, which are denoted by t.

2.1 Solar radiation

Solar position can be described using a horizontal coordinate system, which consists of two angles: the zenith angle (θ) and the azimuth angle (ϕ). The zenith angle represents the angle at which the sun shines down on the observer, with 0° being directly overhead, and the azimuth angle represents the direction from which the sun is shining, with 0° being from the north and 180° being from the south. Figure 2 shows the reference system for these angles. The NREL's SPA algorithm is used to calculate these angles for a given location and time (Reda and Andreas, 2004). Sunset and sunrise occur when the zenith angle passes the 90 degrees mark. Figure 3 illustrates how the angles change through a given day in relation to the measured GHI.
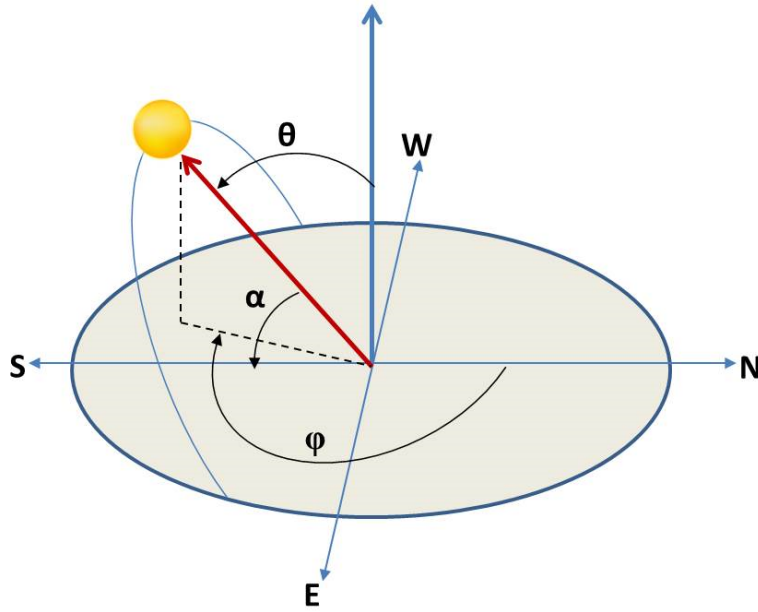

Figure 2: The figure illustrates the solar position relative to the observer.

The Baseline Surface Radiation Network (BSRN) has developed a toolkit to determine the physical and extremely rare limits of solar radiation based on the time of year, time of day, and latitude of the observation station (C. N. Long). These limits represent the maximum and minimum possible solar radiation accounting for all sources. The physical possible limit can be calculated using the following equation, where $-4W/m^2$ represents radiative cooling during the night, μ_0 is the zenith angle measured in radians, and S_a is the extraterrestrial radiation, which is the solar radiation outside the earth's atmosphere (on average, this is 1361 W/m^2). This limit is referred to as the physical possible limit (PPL).
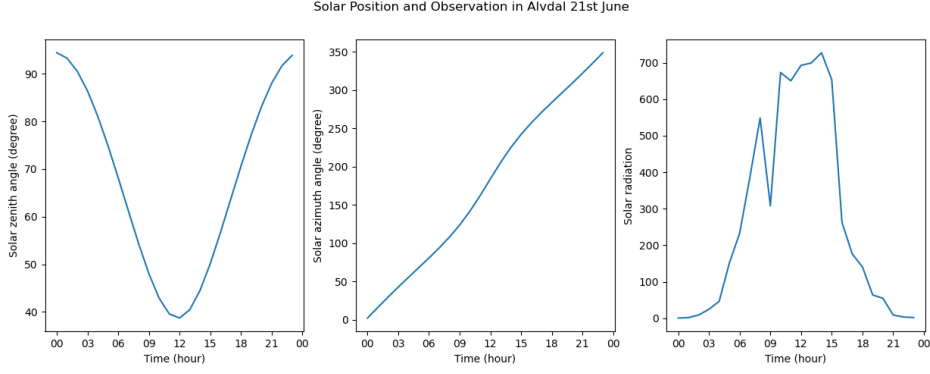
$$-4W/m^2 < GHI < S_a \cdot 1.5\mu_0^{1.2} + 100W/m^2$$

Figure 3: This figure shows the changes in the zenith and azimuth angles with respect to solar radiation over the course of a single day at Alvdal on June 21st.

The extremely rare limits are less frequently reached by observations and should be carefully examined.

$$-2W/m^2 < GHI < S_a \cdot 1.2\mu_0^{1.2} + 50W/m^2$$

The extremely rare limit is called ERL. These limits will be used frequently throughout the paper.

2.2 Probabilistic forecasting

In Section 1, it was stated that the goal of probabilistic forecasting is to maximize the sharpness of the forecast pdfs subject to calibration. To evaluate sharpness and calibration, we use PIT histograms and the CRPS score. The PIT histogram is a visual tool for assessing calibration, while the CRPS score is a proper scoring rule that considers both sharpness and calibration simultaneously. A proper scoring rule is a metric for which the expected score is maximized when the forecasted distribution is equal to the true distribution (Proper Scoring Rules). To assess the deterministic point forecasts we use RMSE, which is used to measure the difference between a predicted value and an observed value.

2.2.1 Probability Integral Transform histogram

The Probability Integral Transform (PIT) histogram is a widely used method for evaluating the calibration of forecasts. According to (Baran and Lerch, 2016), "the PIT is defined as the value of the predictive cumulative density function evaluated at the verifying observations." Calibrated probabilistic forecasts have a PIT histogram that is close to uniform, meaning that values modeled from a continuous distribution can be converted to random variables with a standard uniform distribution. If the random variable X has a cumulative density function F_X, we define the random variable $Y = F_X(X)$, and this should have a standard uniform distribution. Proof:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(F_X(X) \leq y) \\ &= P(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) \\ &= y \end{aligned}$$

The shape of the PIT histogram provides information about the calibration of the forecasts. When the predicted interval is too wide, the PIT histogram has a hump shape, which is known as
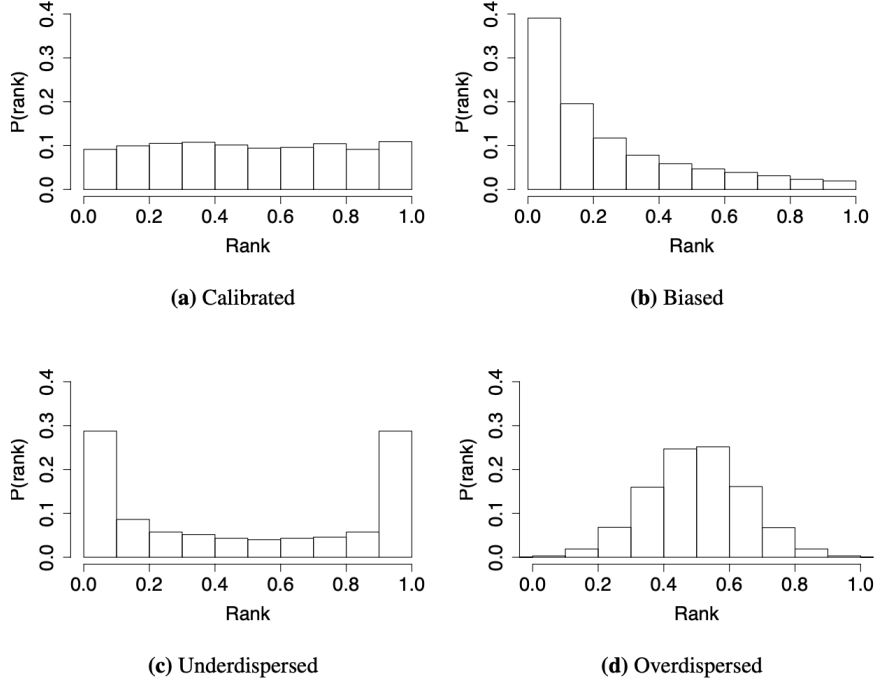
Figure 4: Figure of the different types of PIT-histograms. Kleiven (2017)

overdispersion. If the predicted interval is too narrow, the PIT histogram has a U-shape, known as underdispersion. Additionally, the histograms may be shifted to the left or right for biased forecasts. Figure 4 shows examples of these different cases.

2.2.2 Continuous Ranked Probability Score

Proper scoring rules are an important tool for evaluating probabilistic forecasts, as they maximize the expected score when the issued forecast is the true distribution of the quantity being forecasted (Gneiting and Raftery, 2007). The Continuous Ranked Probability Score (CRPS) is a widely used proper scoring rule for evaluating the accuracy of probabilistic forecasts (Zamo and Naveau, 2018). It is particularly useful for this purpose because it takes into account the entire distribution of predicted probabilities rather than just the point estimate, allowing the forecasted probabilities to be directly compared to the observations. CRPS is also often used in meteorology (Jordan et al., 2017). If F is the predictive cdf and y is the corresponding observation, CRPS is defined as

$$CRPS(F, y) = \int_{\mathbb{R}} (F(x) - 1(x \geq y))^2 dx, \tag{2.1}$$

The CRPS measures the difference between the predicted and observed cumulative distributions and is expressed in the same unit as the observed variable. It is a non-negative value, with smaller values indicating better performance. There are closed-form solutions for calculating the CRPS for many distributions, which can be found in (Zamo and Naveau, 2018). However, some of the distributions used in this study, including the Cauchy distribution and the scaled Beta distribution, do not have closed-form solutions for the CRPS. In these cases, we use simulations to approximate the score. For ensemble members or simulated samples from a distribution, the method for approximating the CRPS involves:

$$CRPS \approx \frac{1}{M} \sum_{m=1}^{M} (F_m(y) - [y \geq x_m])^2 \tag{2.2}$$

6

where M is the number of ensemble members or simulated samples, F_m is the cumulative density function for the mth member or sample, y is the observed value, and x_m is the predicted value for the mth member or sample.

When evaluating the probabilistic forecasts in a given location over N days with F_t being the predictive cdf of y_t, the mean CRPS is

$$CRPS = \frac{1}{N} \sum_{t=1}^{N} CRPS(F_t, y_t) \tag{2.3}$$

We also compare the performance of different models using the Continuous Rank Probability Skill Score (CRPSS). The CRPSS is defined as:

$$CRPSS = \left(1 - \frac{CRPS_i}{CRPS_0}\right) \cdot 100 \tag{2.4}$$

where $CRPS_0$ is the average CRPS for a baseline model, and $CRPS_i$ is the average CRPS for model i. The CRPSS indicates the percentage increase or decrease in accuracy compared to the baseline model. A value of 100 represents perfect skill compared to the baseline model, while a value of 0 indicates no skill compared to the baseline model. Negative values indicate negative skill compared to the baseline model (crp).

2.2.3 Root-Mean Squared Error

The Root-mean squared error (RMSE) is a measure of the accuracy of a point forecast. It is calculated as the square root of the mean squared error between the predicted value, $\hat{y}$, and the true value, y. In the context of solar radiation forecasting, if the measured solar radiation at time t is y_t and the corresponding point forecast is $\hat{y}_t$, the RMSE can be calculated over a set of N observations as follows:

$$RMSE = \sqrt{\frac{\sum_{t=i}^{N}(\hat{y}_t - y_t)^2}{N}} \tag{2.5}$$

RMSE is expressed in the same units as the predicted and true values, and a smaller value indicates a more accurate forecast.

2.3 Bayesian Model Averaging (BMA)

Bayesian model averaging is a common post-processing method of ensemble forecasting introduced by (Raftery et al., 2005). The method generates a probabilistic forecast by combining several pdf forecasts. In BMA, each ensemble member is treated as a statistical model with an associated pdf $g(y|x_m)$. The resulting BMA pdf is then a weighted average of the individual pdfs, with weights w_m satisfying $\sum_{m=1}^{M} w_m = 1$. Given an ensemble member x_m with an associated pdf $g(y|x_m)$, the resulting BMA takes the form

$$f(y|x_1, x_2, ..., x_M) = \sum_{m=i}^{M} w_m g(y|x_m), \tag{2.6}$$

The weights can be estimated based on a training period, or all ensemble members can be treated equally by setting $w_m = 1/M$ for all m. When using BMA for ensemble weather forecasting, it is common to assume that the ensemble members are exchangeable, meaning that they are considered equally likely scenarios of the future (Broecker and Kantz, 2011). In this case, the parameters for the distributions $g(y|x_m)$ are the same for all ensemble members.

The individual pdf $g(\cdot)$ is chosen to represent the quantity of interest. Following the original paper (Raftery et al., 2005), the quantity of interest was temperature, and a Gaussian kernel was chosen for use in BMA. The mean was represented as $\alpha + \beta x_m, m = 1., , , .M$ and standard deviation was σ. The distribution for an individual ensemble pdf using the Gaussian kernel and exchangeable members have the form

$$Y|x_m \sim \mathcal{N}(\alpha + \beta x_m, \sigma^2) \tag{2.7}$$

Given the observation y, ensemble members $m = 1, ..., M$, with distribution $g(\cdot)$, the BMA have the form

$$f(y|x_1, x_2, ..., x_M) = \frac{1}{M} \sum_{m=i}^{M} g(y|x_m), \tag{2.8}$$

To generalize the notation for all kernels, we denote the parameters $\vec{\theta}$. However, all the distributions we are using have three parameters, where two of them represent bias-correction parameters and the last represents the uncertainty within each model. We can generalize the estimators to $\vec{\theta} = (\alpha, \beta, \tau)$, and use this as notation in the rest of the Section. We estimate the bias-correction parameters α and β for all the models by linear regression, and the parameter representing the uncertainty, τ, with maximum likelihood estimation on the BMA distribution. If N is the number of training days from time $t - N$ to $t - 1$, M the number of ensembles, and $g(\cdot)$ represents the distribution of the random variable, linear regression corresponds to maximizing the log-likelihood of $M \cdot (N-1)$ random variables. The formula is

$$(\hat{\vec{\theta}}_{ML}) = \arg\max_{\vec{\theta}} \sum_{t=t-N}^{t-1} \frac{1}{M} \sum_{m=1}^{M} log(g(y_t|x_{tm}) \tag{2.9}$$

This estimation yields estimates for the bias-correction parameters $\hat{\alpha}$ and $\hat{\beta}$, and we need to optimize the uncertainty parameter on the BMA distribution. We fix the bias-correction parameters α and β, before we estimate the uncertainty parameter τ by the log-likelihood of the BMA probabilistic forecast. Using the same training days as in equation 2.9, this formula looks like

$$(\hat{\tau}_{ML}) = \arg\max_{\tau} \sum_{t=t-N}^{t-1} log\left(\frac{1}{M} \sum_{m=1}^{M} g(y_t|x_{tm})\right) \tag{2.10}$$

2.4   Cauchy Distribution

The Cauchy distribution is a probability distribution characterized by two parameters: the location parameter, x_0, which specifies the peak of the distribution, and the scale parameter, γ, which specifies the half-width at half-maximum, also known as the probable error (Cauchy distribution). The probability density function of the Cauchy distribution is given by:

$$f(x; x_0, \gamma) = \frac{1}{\pi} \left(\frac{\gamma}{(x - x_0)^2 + \gamma^2} \right) \tag{2.11}$$

2.5   Truncated Distribution

Truncated distributions are used in practical cases where the values are limited to a certain domain and the support of the original distribution is infinite. Solar radiation is in a certain domain, where it is restricted by 0 in the lower end and the physical possible limit, explained in section 2.1, for the upper end.

Suppose we have a probability density function f(x) found from BMA, with a corresponding cumulative density function F(x). The range after BMA can be infinite, but we want to constrain this domain such that a random variable X from this distribution can take the values $a \leq X \leq b$.

This turns into the pdf given by:

$$f(x | a \leq X \leq b) = \frac{g(x)}{F(b) - F(a)} \tag{2.12}$$

where $g(x) = f(x)$ for $x \in (a, b)$, and $g(x) = 0$ everywhere else.

This new distribution will have pdf equal to $\frac{g(x)}{F(b) - F(a)}$ and a cumulative density function (CDF) equal to $\frac{F(x) - F(a)}{F(b) - F(a)}$. A truncated distribution may also change the mean and variance. This is explained in later Sections.

2.6 Unphysical forecast

The share of unphysical values (U) is a measure of how much of the distribution is outside the physical possible limits of solar radiation. It is calculated by finding the ratio of the probability of the distribution being outside the limits to the total probability of the distribution. The formula to calculate U for a distribution with probability density function f(x) and cumulative density function F(x) bounded by the domain $x \in [a, b]$ is given by:

$$U = \frac{F(a) - F(b)}{1 - F(b) + F(a)} \tag{2.13}$$

This value is useful for understanding the sharpness of the post-processing method, as it indicates how much of the distribution is outside the physical possible limits. A smaller value of U indicates a more precise distribution, while a larger value indicates a less precise distribution.

2.7 Beta distribution

The Beta distribution is a probability distribution commonly used when the values of a random variable are restricted to the domain between 0 and 1. It is often used to model fractions or percentages (Abonazel et al., 2022). The pdf of a beta distribution for $0 \leq x \leq 1$ and shape parameters $\alpha, \beta > 0$ is given by

$$g(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha - 1} (1 - x)^{\beta - 1}, \tag{2.14}$$

where $\Gamma(\cdot)$ is the Gamma function and is defined as $\Gamma(n) = (n - 1)!$. To make the Beta distribution more intuitive, we can follow the idea behind Beta regression, and reparameterize the distribution in terms of the mean parameter μ and a precision parameter ϕ. This results in the following pdf

$$g(x; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1 - \mu)\phi)} y^{\mu\phi - 1} (1 - y)^{(1 - \mu)\phi - 1}, \tag{2.15}$$

where $\mu = \frac{\alpha}{\alpha + \beta}$ and $\phi = \alpha + \beta$. To transform the parameters back to the shape-parameters α and β is simple, as $\alpha = \mu\phi$ and $\beta = \phi(1 - \mu)$. The precision parameter, ϕ, represents the variability in the distribution, with larger values corresponding to smaller variance. A figure showing the Beta pdf and CDF for different values of μ and ϕ can be found in Figure 5.
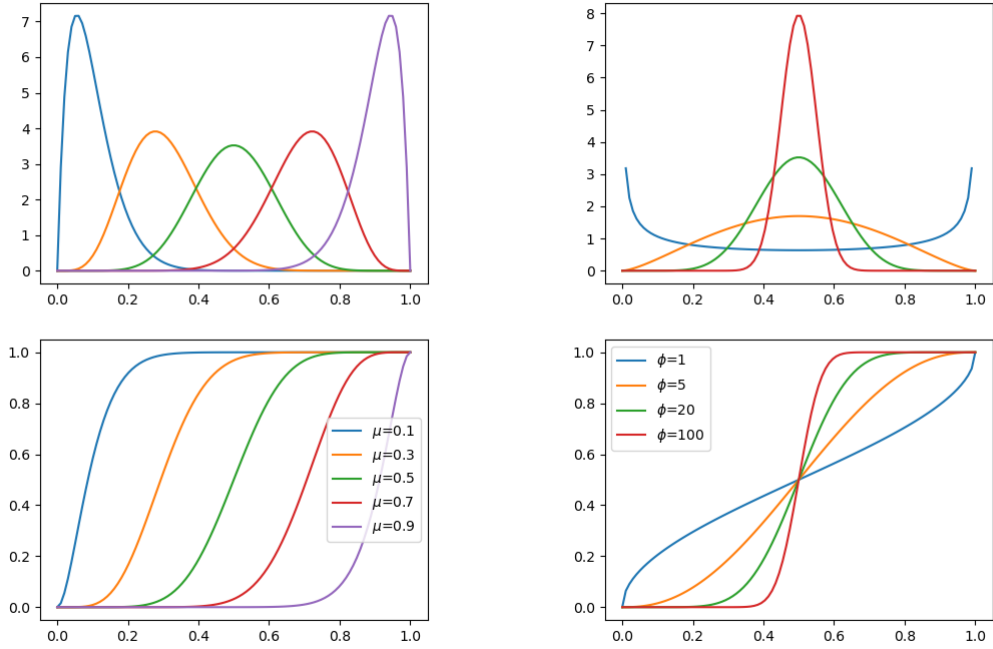
(a) ϕ fixed at 20   (b) μ fixed at 0.5

Figure 5: The upper and lower figures show the probability density function (pdf) and the cumulative distribution function (CDF) for a Beta distribution with different values for the parameters μ and ϕ. In the left figure, the parameter ϕ is fixed at 20 while the parameter μ has five different values. In the right figure, the parameter μ is fixed at 0.5 while the parameter ϕ has four different values. The pdf and CDF illustrate the shape of the Beta distribution for each combination of μ and ϕ values.

The mean and variance for a random variable X from this distribution are given as

$$E(X) = \mu$$
$$Var(X) = \frac{\mu(1-\mu)}{1+\phi}$$

We can see from the variance function that for larger values of ϕ with a fixed mean, the variance is smaller, and therefore ϕ is called a precision parameter. We can also see from Figure 5, that the Beta distribution is very flexible. Since μ represents the mean of the Beta distribution, it also must be bounded by [0,1]. To ensure that it is always inside the domain, we can represent μ as a function of other parameters by a linear predictor. We assume that μ can be written as

$$\eta = \alpha + \sum_{i=1}^{N} \beta_i x_i = g(\mu), \qquad (2.16)$$

for N explanatory variables x_i. The function $g(\cdot)$ represents a link function, and describes the relationship between the mean μ and the linear predictor η. The logit function is a natural choice in the regular Beta regression model to constrain the domain between [0,1].

In addition to the standard form of the Beta distribution, it is possible to scale the distribution to a different range by using a linear transformation. This can be done by substituting the variable

x in the standard form of the Beta distribution with a new variable y defined on the desired range [a,c] by using the transformation $y = x(c-a) + a$. The scaled Beta distribution is equal to the one using two parameters, scaled by the range (c-a). The scaled Beta distribution has a pdf

$$f(y; \mu, \phi, a, c) = \frac{f(x; \mu, \phi)}{c - a}, \tag{2.17}$$

where the variable x can be calculated using the transformation $x = \frac{y-a}{c-a}$, and y is the target of interest. The mean, μ, of the Beta distribution must also be bounded within the range [a, c]. To facilitate this, we can apply a linear transformation to μ, resulting in a new variable, μ_t, which represents the transformed μ bounded within the range [a, c]. This is achieved through the transformation $\mu_t = \mu(c-a) + a$. The mean, μ_t, for the scaled Beta distribution is found by

$$\mu_t = \mu(c - a) + a = logit^{-1}(\eta)(c - a) + a \tag{2.18}$$

3 Forecasts and observations of solar radiation in Norway

In this chapter, we provide an overview of the case study, including the format and locations, as well as the observations and forecasts. We present visual plots of both the observations and ensemble forecasts and discuss the pre-processing of the observations. Finally, we compare the ensemble forecasts with the observations.

3.1 Overview

In this project, observations are retrieved from two different sources, while the forecasts are retrieved from one source. Observations are retrieved from LMT (landbruksmeterologisk tjeneste) and Eklima. These sources provide measurements of the total radiation hitting the ground in W/m^2. 42 randomly selected observation stations are chosen from the open catalogs from LMT and Eklima[1] (see Figure 6).
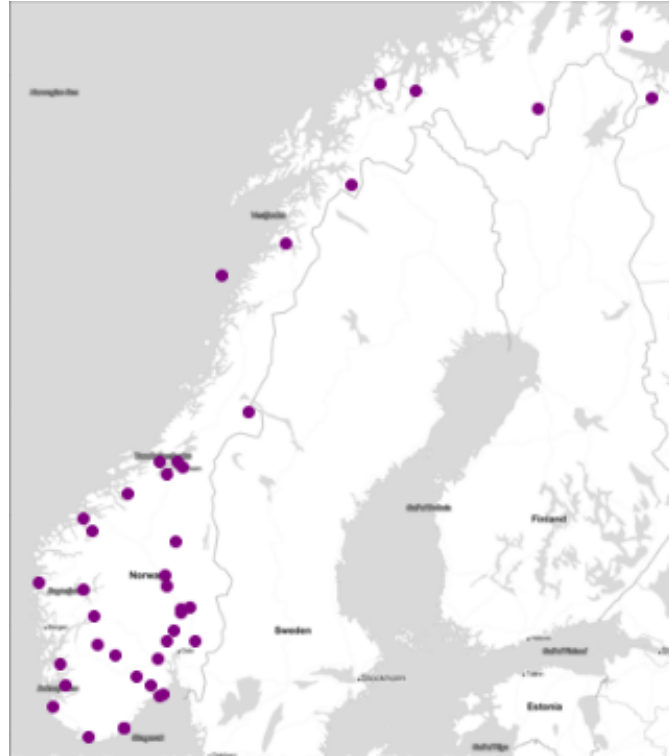
Figure 6: A Visualization of the 42 stations where the data have been retrieved from.

The forecasts in this project were provided by MET through their MEPS (MetoCoOp Ensemble Prediction System). The data can be retrieved from Thredds [2], the Norwegian meteorological institute's API for data retrieval. The ensembles are issued at 00.00 UTC every day, which in this study range from the 1st of January 2020 to the 1st of June 2022. The ensemble consists of $M = 6$ members and is made for $L = 54$ hours lead time. For notation, we let x_{siml} denote ensemble member m at location s, issue time i and lead time l, with the corresponding ground observation $y_{s(i+l)}$. However, in this study we are only considering lead time 12, so for simplicity, we set $t = i + l$, where l is 12 for all issue times. The notation simplifies, and we denote the ensembles x_{stm} with its corresponding observation y_{st}.

---

[1]Eklima: https://seklima.met.no, 18 stations
LMT: https://lmt.nibio.no, 23 stations.
[2]https://thredds.met.no/thredds/catalog.html

3.2 Observations

LMT is a project directed by Nibio, with the main task to obtain meteorological data for warning services and research from the most important agricultural and horticultural districts in Norway (lmt). The global horizontal irradiance is measured using pyranometers, which are devices that measure the amount of solar radiation on a horizontal surface in W/m^2. The instruments are either CM11 or CM3, and a picture of a CM11 pyranometer can be seen in Figure 7.

Figure 7: An image of a CM11 pyranometer from Kipp&Zonen.

Eklima is a service provided by the Norwegian climate service center (KSS) that collects and organizes climate and hydrological data for use in climate adaptation and research on the impacts of climate change on nature and society. The observations from Eklima are collected using the same method as the observations from LMT, as all the observation stations used in this study are owned by Nibio. The observations have an hourly resolution and can be used to visualize daily and seasonal trends. A histogram of all the observations for all locations is shown in Figure 8. These observations are heavily skewed towards zero, which can be attributed to the long nights and fewer sun hours in the northern part of Norway, particularly during the winter months. Most of the observations are also below 800 GHI
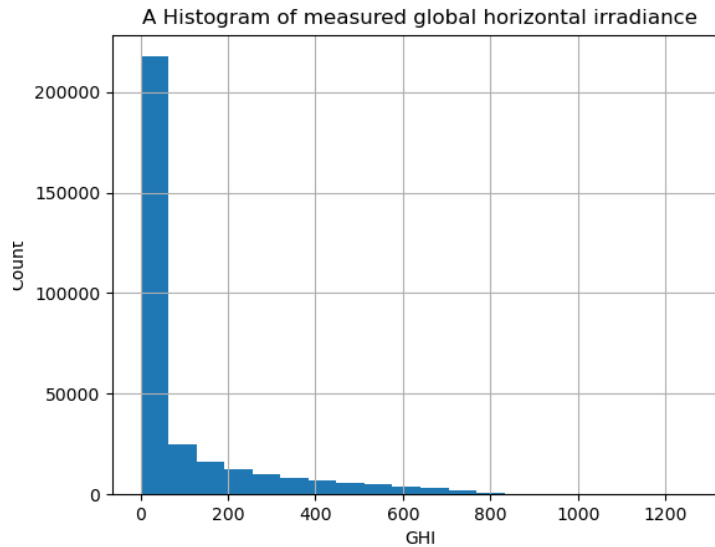
Figure 8: A histogram of measured global horizontal irradiance.

Figure 9 shows a scatter plot of the observations against time for Alvdal. The plot only includes observations from this location to make it clearer. The figure clearly illustrates the seasonal differences, with peaks in the middle of summer and lows in winter.
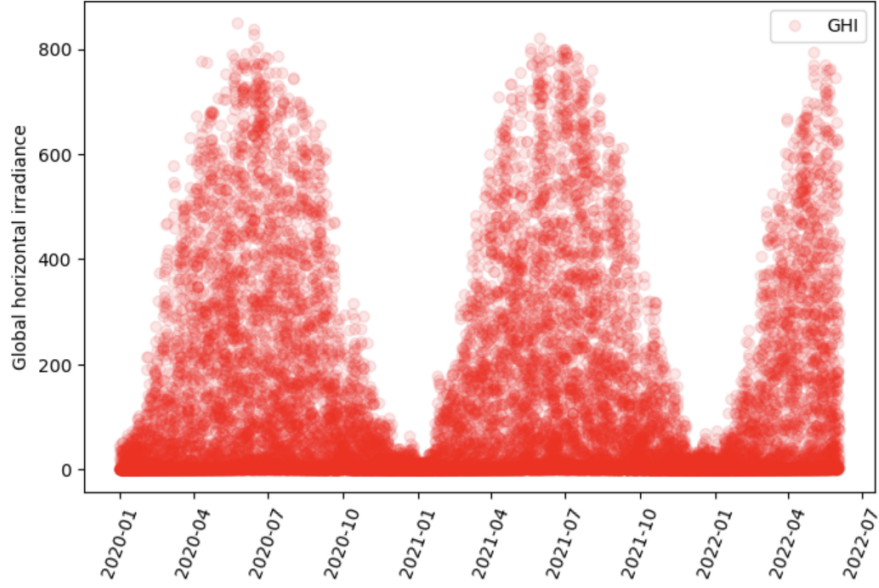
Figure 9: A plot showing GHI observations in Alvdal throughout the two and half years of observations.

Figure 10 illustrates the daily variations in global horizontal radiance, measured in W/m^2, at Alvdal. This is a 2D plot with the hour of the day on the y-axis and the date on the x-axis. The color of the plot represents the intensity of the radiation. The purple dashed line indicates the times of sunrise and sunset, and it is clear that there is no solar radiation between sunset and sunrise.
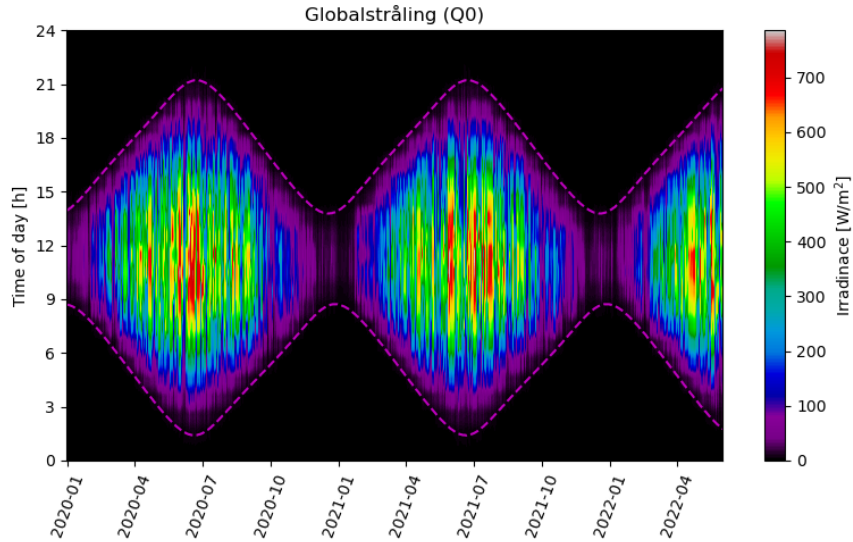
Figure 10: A 2D plot showing GHI measurement data in Alvdal. The colors represent the value of the measurement, and the purple dashed line represents sunset and sunrise.

Our dataset consists of 338820 rows of observations, and we conducted an analysis of all the data to identify any observations that fall outside the physically possible limits, as discussed in

14

Section 2.1. The results of this analysis are shown in Figure 11. From this figure, it is clear that some observations are not valid. In total, 33 observations fall outside the range of the physically possible limits and are removed, while 528 observations fall outside the extremely rare limits. This corresponds to 0.156 % of the observations and is a small enough number to be retained.
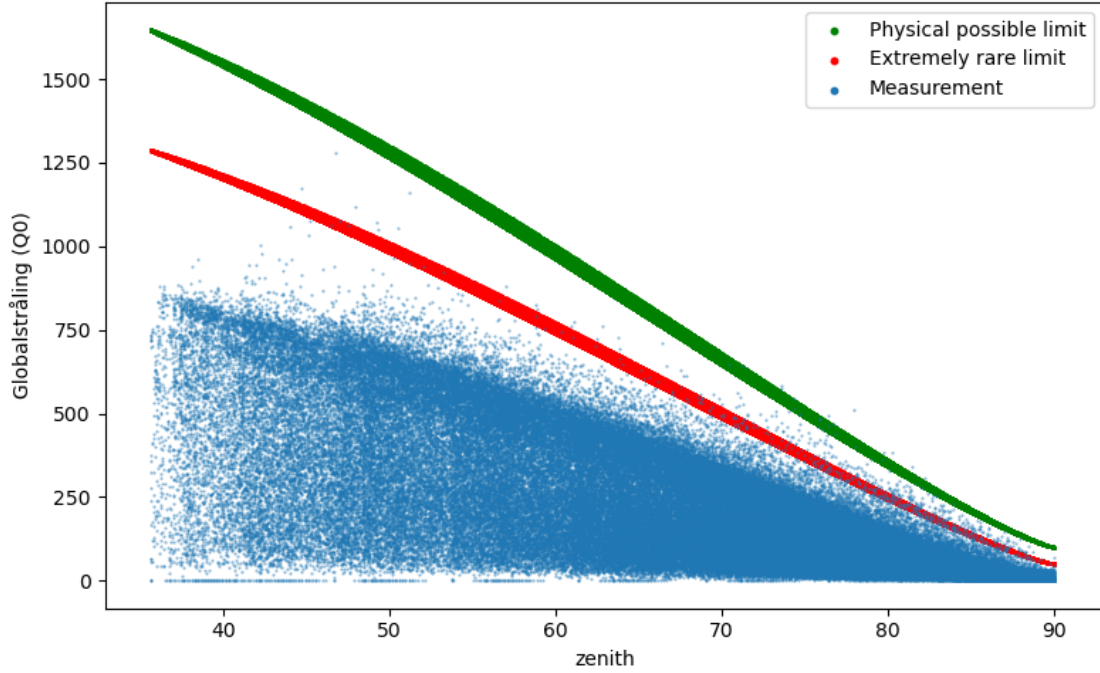
Figure 11: The BSRN check of physical and extremely rare limit for the observations measured by stations owned by Nibio. The blue dots are observations, while the red and green line is respectively the extremely rare and physical possible limit.

3.3 Ensemble forecasts

The ensemble forecast used in this study is obtained from MEPS, which is an implementation of the ensemble version of the HARMONIE-AROME model (Frogner et al., 2019). The HARMONIE-AROME model has two main radiation schemes: short-wave and long-wave. The short-wave scheme is responsible for modeling global horizontal irradiance, and the physical processes underlying these schemes are described in (ECMWF, 2016). The ensembles consist of one control ensemble and five members. The control ensemble is generated without any perturbations to the model, as described in (WMO, 2012).

Similar plots to those in Section 3.2 can be created for the ensemble forecasts, but many of them are not included due to their similarity. However, a plot showing the zenith angle compared to the control ensemble (Figure 12) is of interest. This plot demonstrates that the solar radiation forecasts are consistently below the physical possible limit, which is a notable finding.
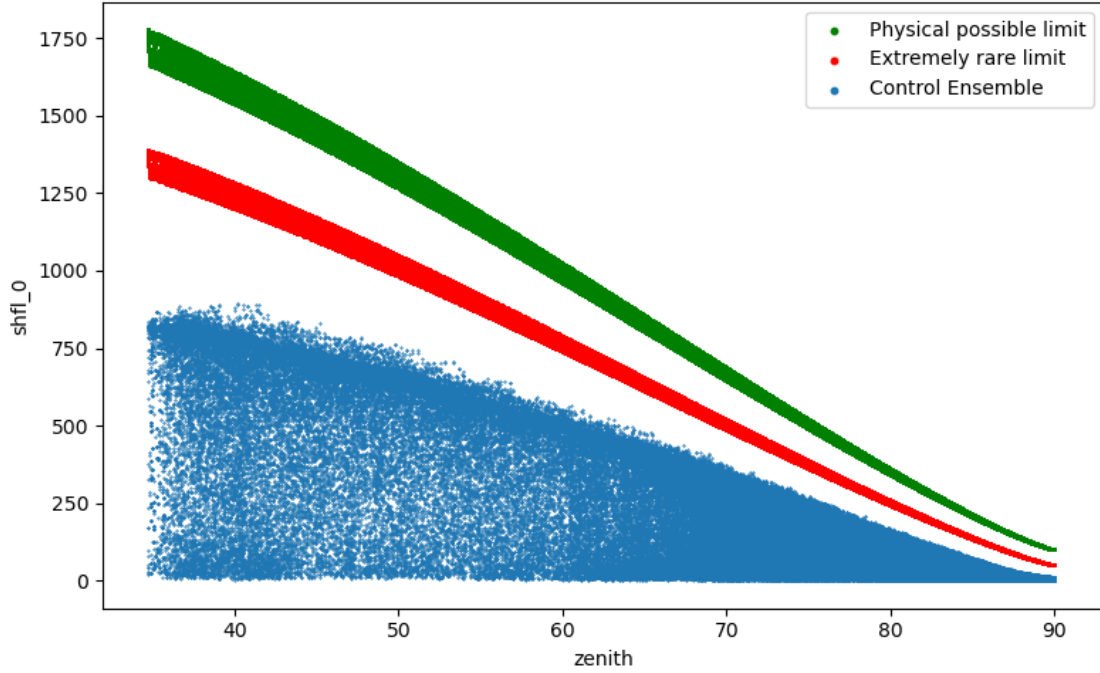
Figure 12: The BSRN check of physical and extremely rare limit for the control ensemble made by MEPS. The blue dots are forecasted values, while the red and green line is respectively the extremely rare and physical possible limit.

3.4 Pre-processing

Missing values are common in ground observations due to instrument malfunctions or poor measurements. Upon examination of the missing values in our dataset, we found that some locations had large portions of their observations missing. To address this, we removed locations with more than 5 missing values, resulting in a final dataset of 15 locations. A list of these locations and their metadata can be found in the appendix.

In addition to missing values, the observations and forecasts were retrieved from sources using different timezones. The forecasts from MET were in Coordinated Universal Time (UTC), while the observations from LMT and Eklima were in Central European Time (CET), with Eklima not observing daylight saving time. The forecasts from MEPS were accumulated over the previous hour, while the observations from LMT were averaged values for the first hour after the given observation time. To ensure consistency, we converted all observations and forecasts to UTC.

Since measuring solar radiation at night is not useful, we only considered solar radiation at 12.00 UTC, which represents the average solar radiation between 11.00-12.00 UTC. There were also some missing values in the ensembles for some locations, so we only considered dates where all 15 locations had no missing values, resulting in a dataset with 775 rows of observations and forecasts for each location.

3.5 Forecast errors

The ensembles are issued at 00.00 each day and we are interested in how well they forecast the weather for the same day, 12 hours into the future. In this subsection, we investigate the errors of the control ensemble to identify any seasonal or locational biases. A forecast bias is a consistent difference between the observations and the forecasts, which can occur when the forecasts tend to be too high or too low.

To quantify the error, we use the following formula:

$$err = y_{st} - \hat{y_{st0}} \tag{3.1}$$

where y_{st} is the observation at location s at time t, and $\hat{y_{st0}}$ is the corresponding control ensemble forecast.

A boxplot of the errors for each location is shown in Figure 13. We can see that some locations exhibit clear biases, while others have close to zero bias. This may indicate that the forecast is more accurate in some locations than others. Of particular interest are the locations of Iskoras II, which has a positive bias, and Brunlanes and Hjelmeland, which have a negative bias. On the other hand, the forecasts at Holt appear to be very good, with approximately zero bias and few large errors.
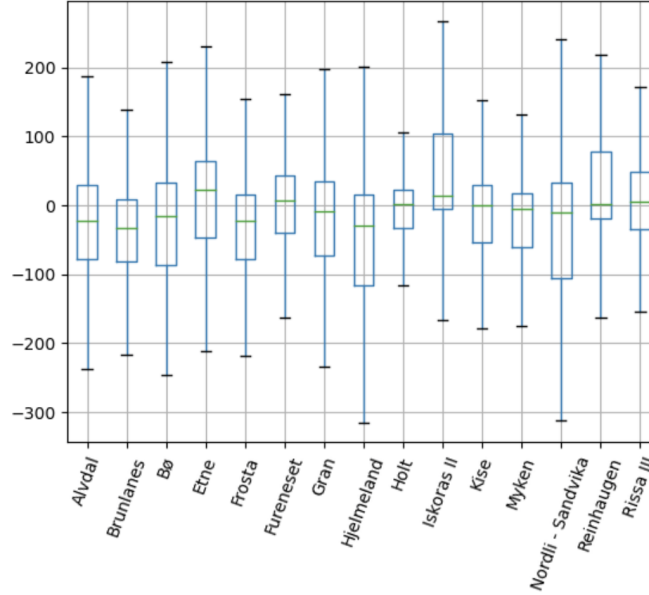
Figure 13: A boxplot without outliers, visualizing the average errors for the control ensemble for all the locations.

We can further split the errors into seasons to understand how the forecasts perform at different times of the year. The seasons are defined as follows: December, January, and February are winter, March, April, and May are spring, June, July and August are summer, and September, October, and November are autumn. It is clear from Figure 14 that the errors vary seasonally, with larger errors in the summer months and smaller errors in the winter months. In this figure, we can also see that the locations of Holt, Iskoras II, and Nordli-Sandvika have almost zero errors in the winter months, which is likely due to the low solar radiation observed in these months.
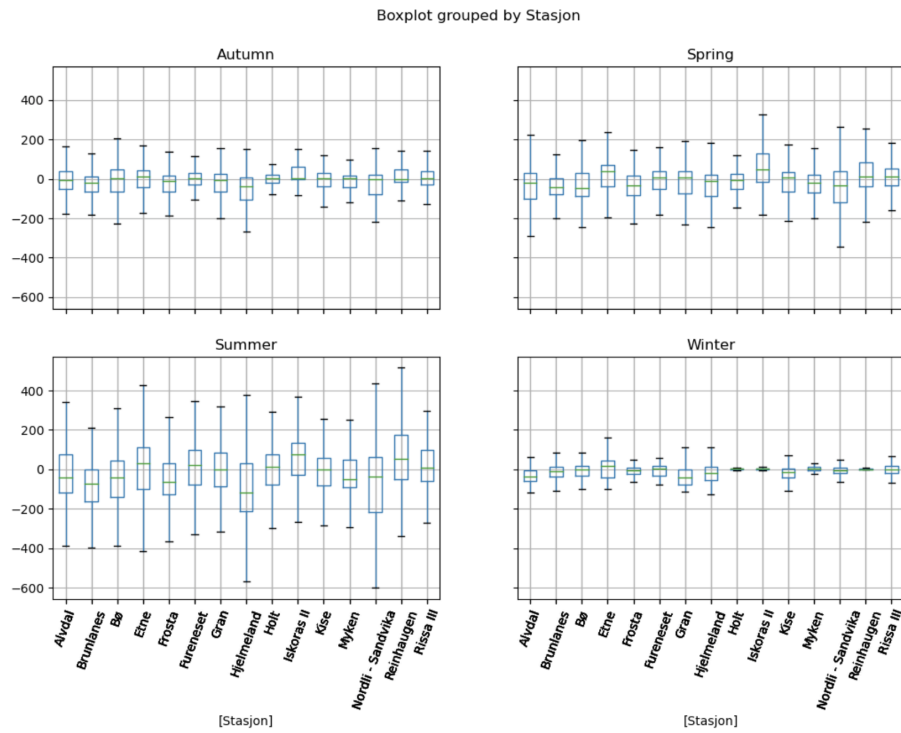
Figure 14: A boxplot without outliers, visualizing the average errors for the control ensemble for all the locations split into seasons.

In Figure 15, we visualize the bias by plotting a map of Norway with mark colors indicating positive (red) or negative (blue) bias. From this map, it appears that there is a positive bias in the north and a negative bias in the south.
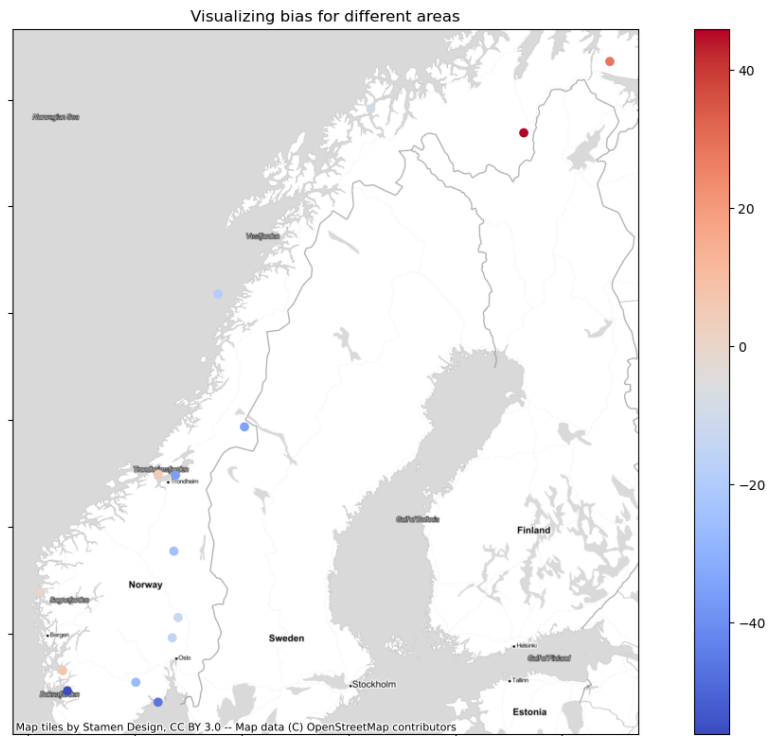
Figure 15: A map of Norway showing locational bias. The color represents the bias, and the dot is the location of the bias.

4 Methodology

In this section, we set up the post-processing methods based on the theory presented in Section 2. All post-processing methods are based on the BMA methodology. Since no previous research has used BMA for post-processing solar radiation, we test a range of different distributions. For simplicity, we suppress the location subscript s and present the models for a single location. Some of the methods are trained on a training period and are denoted with a subscript t, while others do not have this subscript. The ensemble number is denoted with m.

4.1 Raw Ensemble

We use the raw ensemble members as a baseline model, with all ensembles having equal weight. For the probability distribution function, we use the empirical distribution function, which is a step function that jumps to 1/m for each of the M ensembles. An example of this is shown in Figure 16.
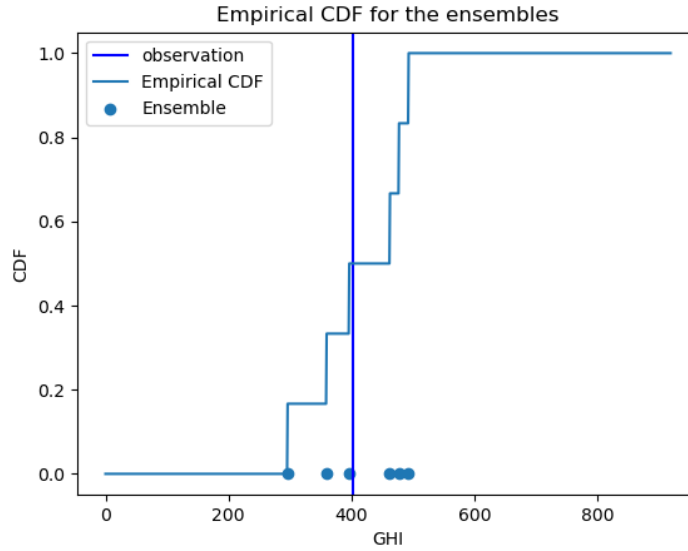
Figure 16: A figure demonstrating the empirical distribution function for one ensemble forecast. The blue vertical line is an observation, the blue dots are the ensembles and the lighter blue line is the empirical CDF.

By using the empirical distribution for the raw ensembles, we can calculate the PIT-histogram, CRPS, and RMSE metrics as described in Section 2.2. This model is used as the baseline for the CRPSS score calculated using Equation 2.4. The mean of the ensembles is used to calculate the RMSE.

4.2 BMA with Gaussian Kernel

The BMA methodology associates each member of an ensemble with a probability density function (pdf) $g(y|x_m)$. This pdf can be modeled as a normal distribution, with the form

$$y|x_m \sim \mathcal{N}(\alpha + \beta x_m, \sigma^2) \tag{4.1}$$

where α and β are the bias-correction parameters and σ is the uncertainty parameter. Given an observation y and corresponding ensemble members $x_m, m = 1, ..., M$, the BMA pdf has the form:

$$f(y|x_1, x_2, ..., x_M) = \frac{1}{M} \sum_{m=i}^{M} \mathcal{N}(\alpha + \beta x_m, \sigma^2) \qquad (4.2)$$

We then have to estimate the parameters α, β and σ, where the mean is modelled as $\alpha + \beta \cdot x_m, m = 1, ..., M$ and the standard deviation is σ. The bias-correction parameters are estimated by maximizing the log-likelihood of $M \cdot (N-1)$ independent Normal-distributed random variables

$$(\hat{\alpha}_{ML}, \hat{\beta}_{ML}) = \arg\max_{\alpha,\beta} \left(\sum_{t=1}^{N} \frac{1}{M} \sum_{m=1}^{M} log(f_{normal}(y_t|x_{tm})) \right) \qquad (4.3)$$

Once the bias-correction parameters α and β have been estimated, we fix them and estimate σ on the likelihood function of the BMA probabilistic forecast over the same training period:

$$\hat{\sigma}_{ML} = \arg\max_{\sigma} \prod_{t=1}^{N} \frac{1}{M} \sum_{m=1}^{M} f_{normal}(y_t|x_{tm}) \qquad (4.4)$$

An example of how this looks for a randomly chosen observation and its corresponding ensembles is shown in Figure 17. The blue line represents the observation, the black dots represent the individual ensembles, and the brown and black lines represent the individual ensemble pdfs and the BMA predictive pdf, respectively. In this specific example, there is a good spread between the ensembles, which results in a large uncertainty.
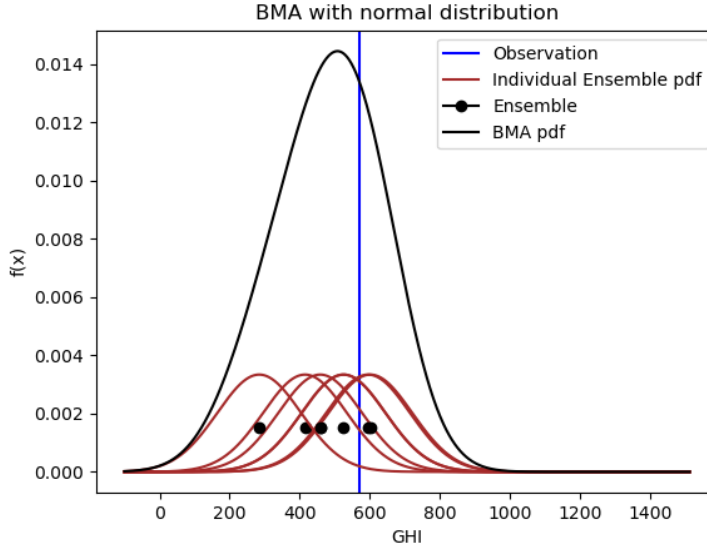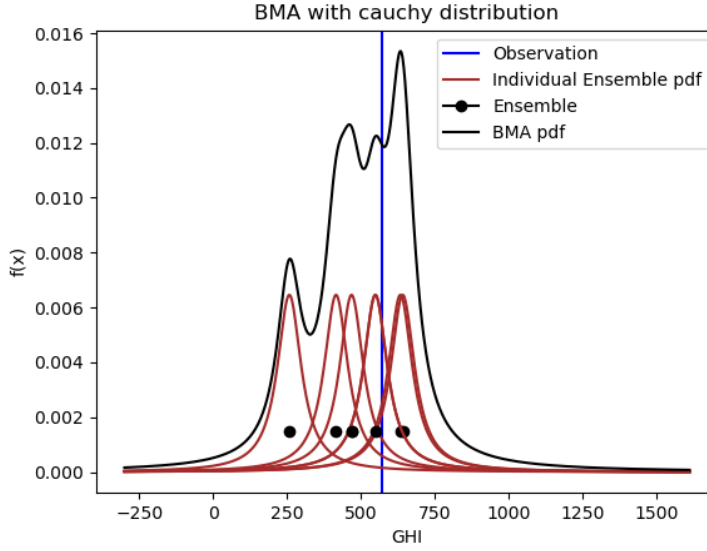

Figure 17: Example of how BMA looks like for the Gaussian distribution with six ensemble members. The brown lines show the probability density functions (pdfs) for each of the six ensemble members, while the black line represents the BMA predictive PDF. The blue vertical line indicates the observation, and the black dots represent the ensembles.

There exist a closed-form solution for a mixture of normal distributions. The CRPS formula for the mixture F given the observation y is

$$CRPS(F, y) = \sum_{m=1}^{M} w_i A(y - \mu_m, \sigma_i^2) - \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} w_i w_j A(\mu_i - \mu_j, \sigma_i^2 + \sigma_j^2)$$

$$A(\mu, \sigma^2) = \mu(2\Phi(\frac{\mu}{\sigma}) - 1) + 2\sigma\phi(\frac{\mu}{\sigma})$$

where $\mu_m = \alpha + \beta x_m$, $i = 1, ..., M$, and ϕ and Φ are respectively the pdf and CDF of a standard normal distribution. To calculate RMSE for the Gaussian BMA we use the mean of the distribution. The mean for BMA with exchangeable members and equal weights is given as

$$E[y|x_1, ..., x_m] = \frac{1}{M} \sum_{m=1}^{M} \alpha + \beta x_m \tag{4.5}$$

4.3 BMA with Cauchy Kernel

The probability density function that most closely resembles the errors is the Cauchy distribution. As shown in Figure 18, the Cauchy distribution provides a better fit for the errors, and therefore it is worth investigating how a Cauchy kernel with BMA performs.
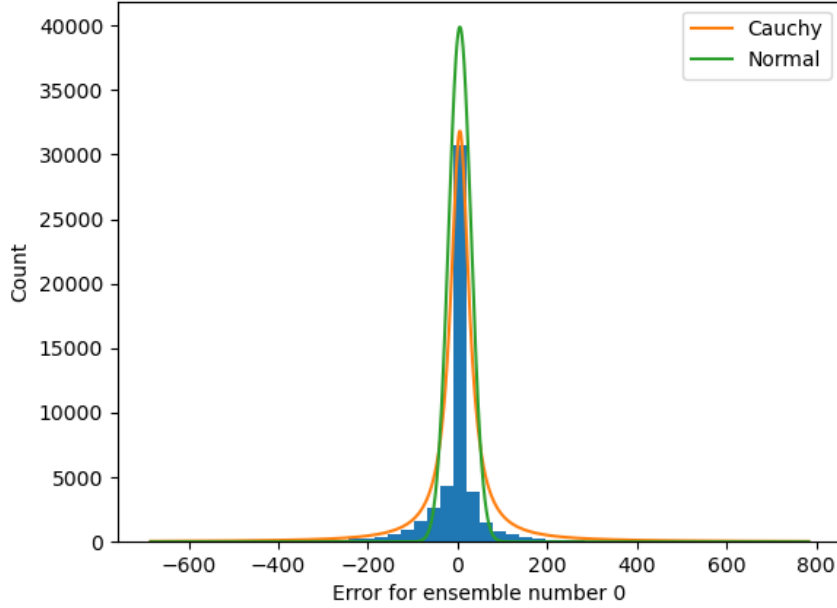
Figure 18: Figure showing the errors for the control ensemble compared to the observations.

Using the BMA methodology, we wish to model the conditional forecast $g(y|x_m)$ as a Cauchy distribution of the form:

$$y|x_m \sim \mathcal{C}(\alpha + \beta x_m, \gamma) \tag{4.6}$$

where the location parameter is expressed as $\alpha + \beta x_m$, $m = 1, ..., M$. To estimate the parameters, we use the same methodology described in Section 4.2, and estimate the bias-correction parameters α and β through linear regression and the uncertainty parameter γ through maximizing the likelihood function of the BMA function. The linear regression estimate corresponds to maximizing the

likelihood function of $M \cdot (N - 1)$ independent Cauchy distributed variables, where M is the number of ensembles and N is the number of training days. The function looks like

$$(\hat{\alpha}_{ML}, \hat{\beta}_{ML}) = \arg\max_{\alpha,\beta} \left(\sum_{t=1}^{N} \frac{1}{M} \sum_{m=1}^{M} log(f_{Cauchy}(y_t|x_{tm})) \right) \tag{4.7}$$

After estimating the bias-correction parameters α and β, we fix them and estimate γ on the likelihood function of the BMA probabilistic forecast using the same training period:

$$\hat{\gamma}_{ML} = \arg\max_{\gamma} \prod_{t=1}^{N} \frac{1}{M} \sum_{m=1}^{M} f_{Cauchy}(y_t|x_{tm}) \tag{4.8}$$

An example of the resulting distribution for a randomly chosen observation with its corresponding ensembles and predictive pdfs is shown in Figure 19.
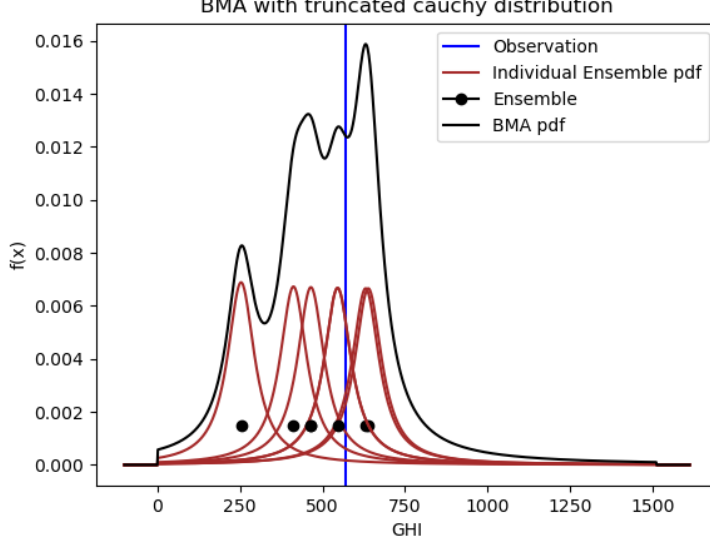

Figure 19: Example of how BMA looks like for the Cauchy distribution with six ensemble members. The brown lines show the probability density functions (pdfs) for each of the six ensemble members, while the black line represents the BMA predictive PDF. The blue vertical line indicates the observation, and the black dots represent the ensembles.

Cauchy distribution has no closed-form solution for calculating CRPS, and we need to estimate it using simulation-based CRPS explained in Equation 2.2. For calculating the RMSE we need a point forecast, and we choose the median of the Cauchy BMA since the Cauchy distribution has no defined mean. The median of the BMA is defined as

$$Median[y|x_1, ..., x_m] = \frac{1}{M} \sum_{m=1}^{M} \alpha + \beta x_m \tag{4.9}$$

4.4 BMA with Truncated Normal Kernel

As explained in Section 2.5, some parts of the BMA distributions might fall outside the physical range of solar radiation defined by 0 in the lower limit and PPL in the upper, where PPL is explained in Section 2.1. We truncate the Gaussian distribution to have support only in the

physical possible domain of solar radiation and denote this new domain [0, PPL]. A truncated Gaussian distribution with a random variable $a \leq x \leq b$ has pdf given by

$$f(x; \mu, \sigma, a, b) = \frac{1}{\sigma} \frac{\phi(\frac{x-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}, \quad (4.10)$$

and 0 if it is outside the domain [a,b]. We denote the truncated Gaussian distribution as $\mathcal{N}_T$, and set $\mu_m = \alpha + \beta x_m, m = 1, ..., M$. We now associate the conditional forecast $g(y|x_m), m = 1, ..., M$ by a truncated Normal distribution on the form

$$y|x_m \sim N_T(\alpha + \beta x_m, \sigma^2, 0, PPL) \quad (4.11)$$

To estimate the parameters of the truncated Gaussian distribution, we use the same inference methods as for the Gaussian Bayesian model averaging (BMA) described in Section 4.2. Specifically, we use Equations 4.3 and 4.4 to estimate the parameters, taking into account the truncation of the distribution by using the form of the truncated Gaussian distribution given in Equation 4.11. The resulting BMA has the form

$$f(y|x_1, x_2, ..., x_M) = \frac{1}{M} \sum_{m=i}^{M} \mathcal{N}_{\mathcal{T}}(\alpha + \beta x_m, \sigma^2, 0, PPL) \quad (4.12)$$

An example of the BMA for the truncated Gaussian distribution with six ensemble members is shown in Figure 20.
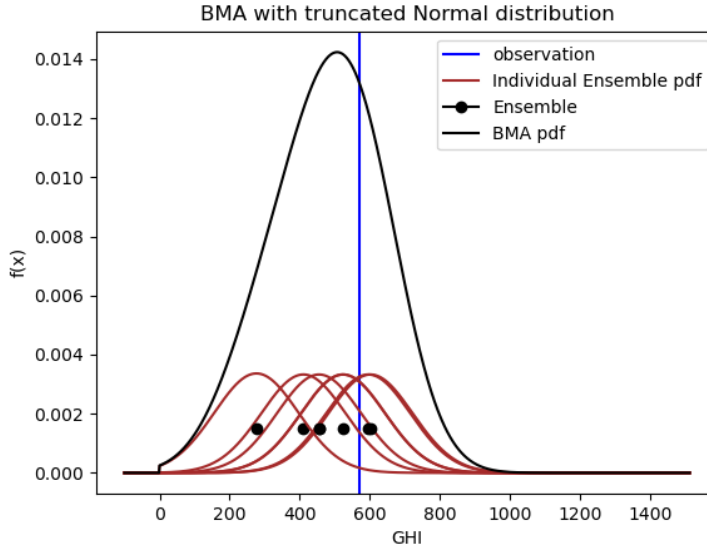
Figure 20: Example of how BMA looks like for the truncated Gaussian distribution with six ensemble members. The brown lines show the probability density functions (pdfs) for each of the six ensemble members, while the black line represents the BMA predictive PDF. The blue vertical line indicates the observation, and the black dots represent the ensembles.

To calculate the CRPS for the truncated Gaussian BMA, we use the simulation-based method described in Equation 2.2, as there is no closed-form solution for a mixture of truncated Gaussian distributions. The mean of the truncated Gaussian BMA with lower limit a and upper limit b is given by the equation

$$E[y|x_1, ..., x_m] = \frac{1}{M} \sum_{m=1}^{M} \mu_m + \frac{\phi(\frac{a-\mu_m}{\sigma}) - \phi(\frac{b-\mu_m}{\sigma})}{\Phi(\frac{b-\mu_m}{\sigma}) - \Phi(\frac{a-\mu_m}{\sigma})} \sigma, \quad (4.13)$$

where $\mu_m = \alpha + \beta x_m, m = 1, ..., M$, ϕ, and Φ is respectively the pdf and CDF of the standard normal distribution, a is 0 and b is PPL. The mean of the truncated Gaussian BMA is used to calculate the RMSE of this method.

4.5 BMA with Truncated Cauchy Kernel

The Cauchy distribution has heavier tails than the normal distribution. We can see from Figure 19 that a truncated Cauchy distribution might represent the values of solar radiation better, as large parts of the distribution are outside the physical domain. A continuous variable x is distributed by a truncated Cauchy distribution where a is the lower limit and b is the upper limit if the pdf is given by

$$g(x; x_0, \gamma, a, b) = \frac{\gamma}{c[\gamma^2 + (x - x_0)^2]}, \qquad a \leq x \leq b \tag{4.14}$$

where $c = tan^{-1}(b - x_0)/\gamma - tan^{-1}(a - x_0)\gamma$, x_0 is the location parameter and γ the scale parameter. We express the location parameter $x_0 = \alpha + \beta x_m, m = 1, ..., M$, and truncate the distribution to be within [a,b], where a is 0 and b is PPL defined in section 2.1. We denote the truncated Cauchy distribution as C_T. The conditional forecast $g(y|x_m)$ have the form

$$y|x_m \sim C_T(\alpha + \beta x_m, \gamma, 0, PPL) \tag{4.15}$$

To estimate the parameters of the truncated Cauchy distribution, we use the same method as described in Section 4.3. Specifically, we use Equations 4.7 and 4.8 to estimate the parameters, taking into account the truncation of the distribution by using the form of the truncated Cauchy distribution given in Equation 4.15. The resulting BMA is then given by:

$$f(y|x_1, x_2, ..., x_M) = \frac{1}{M} \sum_{m=i}^{M} C_{\mathcal{T}}(\alpha + \beta x_m, \gamma, 0, PPL) \tag{4.16}$$

An example of the BMA for the truncated Cauchy distribution with six ensemble members is shown in Figure 21.

Figure 21: Example of how BMA looks like for the truncated Cauchy distribution with six ensemble members. The brown lines show the probability density functions (pdfs) for each of the six ensemble members, while the black line represents the BMA predictive PDF. The blue vertical line indicates the observation, and the black dots represent the ensembles.

To calculate the CRPS for the truncated Cauchy BMA, we use the simulation-based method described in Equation 2.2, as there is no closed-form solution for the CRPS of truncated Cauchy distributions. The truncated Cauchy distribution has, unlike the Cauchy distribution, a defined mean. We can take the integral:

$$\frac{\gamma}{c} \int_a^b \frac{x dx}{\gamma^2 + (x - x_0)^2} \tag{4.17}$$

to get

$$E(x) = x_0 + \frac{\gamma}{2c} \left\{ log(1 + ((b - x_0)/\gamma)^2) - log(1 - ((a - x_0)/\gamma)^2) \right\} \tag{4.18}$$

The mean of the truncated Cauchy BMA is used to calculate the RMSE of this method. The mean for the BMA with exchangeable members and equal weights is given by

$$E[y|x_1, ..., x_m] = \frac{1}{M} \sum_{m=1}^{M} E(x_m), \tag{4.19}$$

where E(x) is defined in Equation 4.18, and $x_0 = \alpha + \beta x_m, m = 1, ..., M$.

4.6   BMA with scaled Beta Kernel

We use the four-parameter Beta distribution, which has a naturally restricted domain bounded by [a,c]. We express the domain of the scaled Beta distribution as $[0, PPL]$, where PPL is the physical possible limit of solar radiation, as defined in Section 2.1. We can then associate the conditional forecast $g(y|x_m), m = 1., , , .M$ by a scaled Beta distribution on the form

$$y|x_m \sim Beta(\mu_m, \phi, 0, PPL) \tag{4.20}$$

$$\mu_m = logit^{-1}(\alpha + \beta x_m) \cdot PPL \tag{4.21}$$

We first estimate the bias-correction parameters α and β using Beta regression. This is the same as maximizing the log-likelihood of $M \cdot (N-1)$ independent scaled Beta-distributed random variables, given by

$$(\hat{\alpha}_{ML}, \hat{\beta}_{ML}) = \arg \max_{\alpha, \beta} \left(\sum_{t=1}^{N} \frac{1}{M} \sum_{m=1}^{M} log(f_{beta}(y_t|x_{tm})) \right) \tag{4.22}$$

Next, we fix the bias-correction parameters α and β and estimate the precision parameter ϕ by maximizing the likelihood function of the BMA probabilistic forecast over the same training period

$$\hat{\phi}_{ML} = \arg \max_{\phi} \prod_{t=1}^{N} \frac{1}{M} \sum_{m=1}^{M} f_{beta}(y_t|x_{tm}) \tag{4.23}$$

The resulting Beta-BMA is then given by

$$f(y|x_1, x_2, ..., x_M) = \frac{1}{M} \sum_{m=1}^{M} f_{beta}(\mu_m, \phi, 0, PPL) \tag{4.24}$$
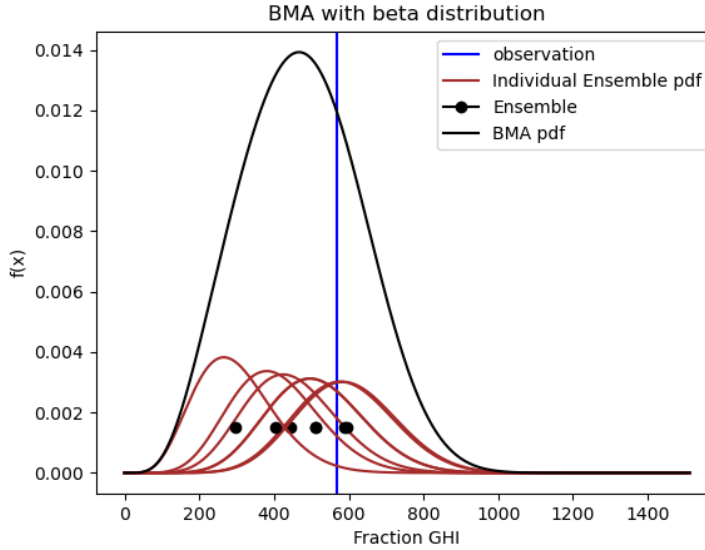
Figure 22: Example of how BMA looks like for the scaled Beta distribution with six ensemble members. The brown lines show the probability density functions (pdfs) for each of the six ensemble members, while the black line represents the BMA predictive PDF. The blue vertical line indicates the observation, and the black dots represent the ensembles.

The mean of the Beta-BMA with exchangeable members and equal weights is given by:

$$E[y|x_1, ..., x_m] = \frac{1}{M} \sum_{m=1}^{M} \mu_m, \tag{4.25}$$

where $\mu_m = logit^{-1}(\alpha + \beta x_m) \cdot PPL, m = 1, ..., M$.

Since there is no closed-form solution for the transformed Beta distribution for calculating the CRPS, we use a simulation-based method, as defined in Equation 2.2, to estimate the CRPS. To calculate RMSE we use the mean, defined in equation 4.25. A figure of the Beta-BMA for a random observation with the corresponding ensembles and its predictive probability density functions is shown in Figure 22.

4.7   Sliding window

To estimate the parameters in all the models, we use the sliding window method. The parameters α and β are estimated using linear regression on a sliding window training period of size N days. After fixing these parameters, we estimate the uncertainty parameter τ on the likelihood function of the BMA probabilistic forecast using the same training period. The training period for each validation day i consists of the N days preceding it, ranging from day $t = i - N$ to day $t = i - 1$.

5 Results

The goal of this section is to evaluate and compare the performance of the models introduced in Section 4. We begin by presenting a case study to illustrate the behavior of the mixtures at different times and locations. We then present the results of the pure ensembles as a baseline and evaluate the calibration of the different models. We also compare the CRPS and RMSE metrics and assess the percentage of unphysical values in the Cauchy and Gaussian distributions. Finally, we investigate how the parameters of the different models vary over time for the locations in which they were trained.

5.1 Sliding Window

To determine the optimal number of training days for the models, we performed a test using the average CRPS as a metric and the Beta-distributed BMA as the model. The results of this test are shown in Figure 23, which plots the average CRPS against the number of training days in the location of Alvdal. According to the "elbow" method, which is a heuristic for determining the point at which diminishing returns are no longer worth the additional computational cost, it appears that the optimal number of training days lies between 15 and 25. Based on these results, we chose to use a sliding window of 20 days to estimate the parameters for all models.
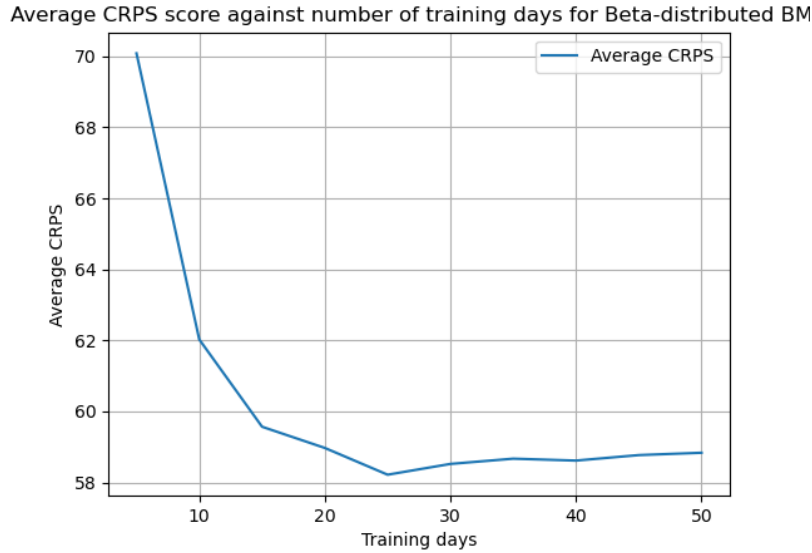
Figure 23: This figure shows the relationship between the average CRPS and the number of training days used to estimate the parameters for a Beta-distributed BMA model in Alvdal. The x-axis represents the number of training days, while the y-axis represents the average CRPS.

5.2 Case study: Differences in Time and space

The purpose of this subsection is to demonstrate the performance of the BMA model in different seasons and locations in Norway through specific case studies. Three cases have been selected to showcase the model's performance, including Alvdal in late February, Hjelmeland in late December, and Myken in the midst of summer. Alvdal is a municipality located in the southeastern part of Innlandet county, near the border with Hedmark county. Myken is an island municipality located in the central part of Møre og Romsdal county, near the border with Trøndelag county. Hjelmeland is a municipality located in the southwestern part of Rogaland county, along the coast of the North Sea. The probability density function (pdf) and the cumulative density functions (CDF) of the BMA models will be presented for these cases.

5.2.1 Alvdal 25th February 2020

The results from Alvdal on February 25th, 2020 are shown in Figure 24. In this specific location and time, the shapes of the BMA predictive probability density functions (pdfs) produced by the different models appear to be similar. The modes of the pdfs are also located close to each other. However, the Cauchy mixtures have a higher peak but also a heavier tail compared to the other models. The truncation applied to the Gaussian and Cauchy mixtures are also evident in the figure, as there is a discontinuity at 0 and at the upper limit of 634 GHI.
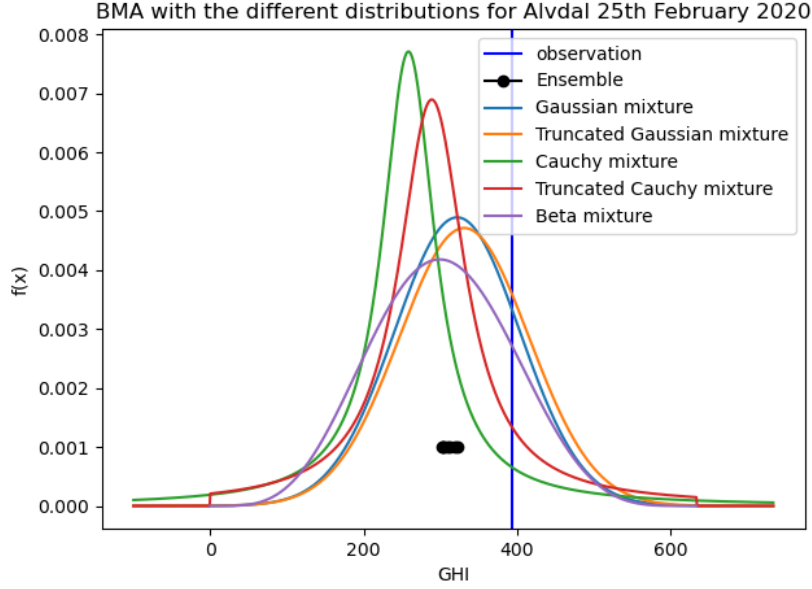
(a) pdf for the BMA of different distributions.
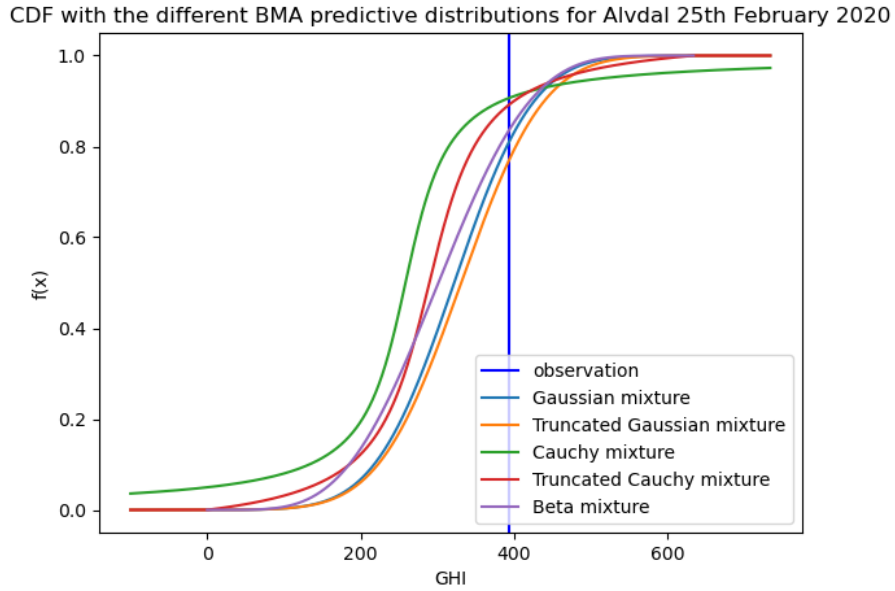
(b) Corresponding CDF for the BMA of different distributions.

Figure 24: A figure showing the results for Alvdal, trained from February 5th to February 24th, 2020, and evaluated on February 25th, 2020. The figure displays the BMA predictive probability density functions (pdfs) and the corresponding cumulative density functions (CDFs) for different models.

5.2.2 Hjelmeland 24th December

The results from Hjelmeland on Christmas Eve are shown in Figure 25. The figure shows that all the ensembles are close to the observation of 33, and none of the models perform poorly. However, the truncated Gaussian distribution has a strange shape compared to the others. In addition, both the Cauchy and Gaussian mixtures have parts of their distribution outside the physical possible domain.
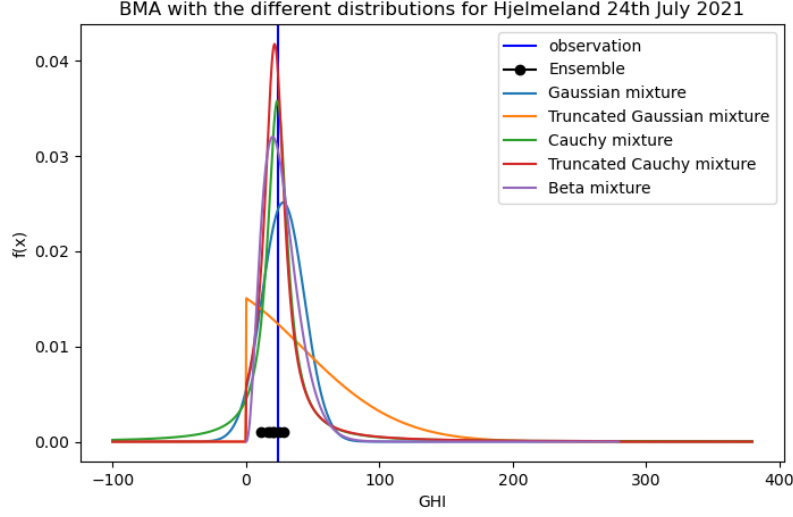
(a) pdf for the BMA of different distributions.
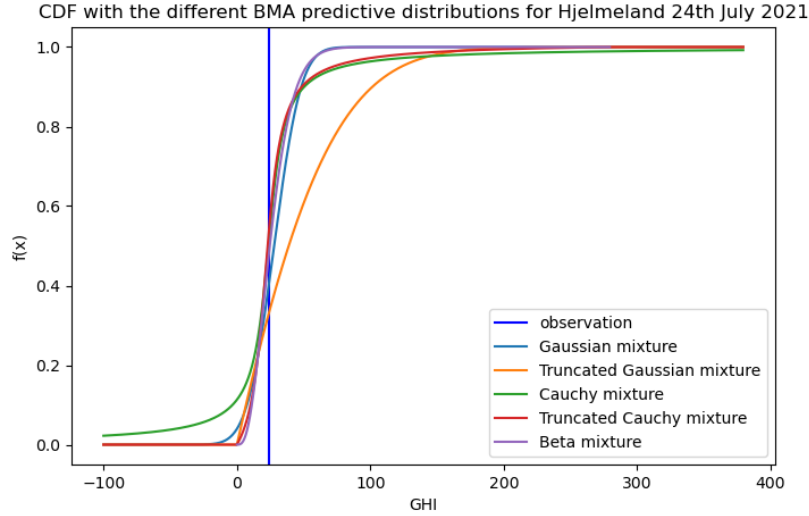
(b) Corresponding CDF for the BMA of different distributions.

Figure 25: A Figure showing the results from Hjelmeland, trained from December 4th to December 23rd and evaluated on December 24th, 2021. The figure displays the BMA predictive probability density functions (pdfs) and the corresponding cumulative density functions (CDFs) for different models.

5.2.3 Myken 2nd July

The results from Myken on July 2nd can be seen in Figure 26, where the observation has a value of 402 GHI. In this case, the post-processing has shifted the BMA predictive distribution to the left of the actual ensembles, indicating that the models have been bias-corrected. Additionally, all of the predictive distributions are in close agreement with the observation, indicating that the models are performing well for this specific case.
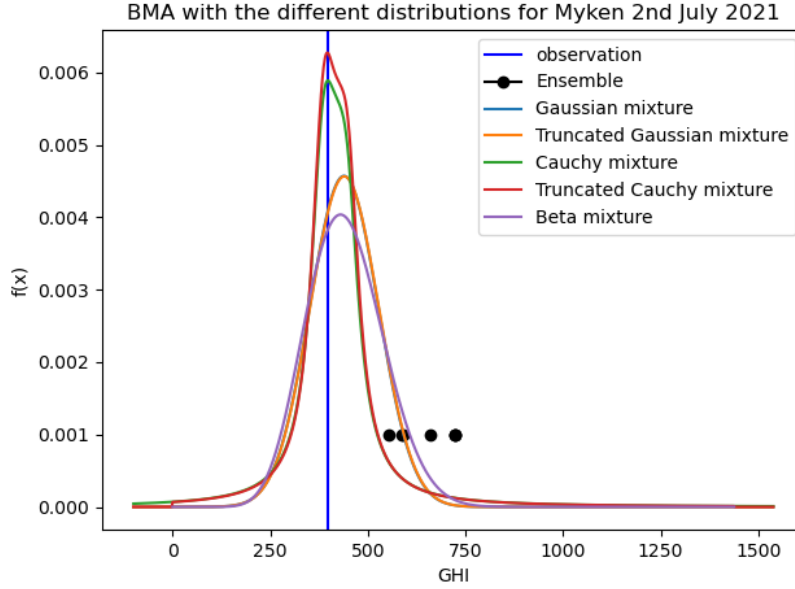
(a) pdf for the BMA of different distributions.
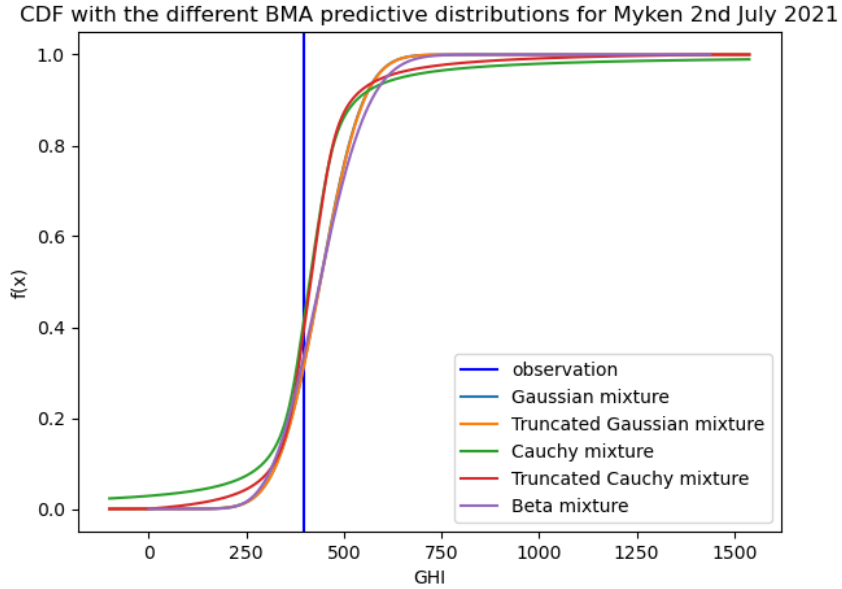
(b) Corresponding CDF for the BMA of different distributions.

Figure 26: A Figure showing the results from Myken, trained from June 12th to July 1st and evaluated on July 2nd, 2021. The figure displays the BMA predictive probability density functions (pdfs) and the corresponding cumulative density functions (CDFs) for different models.

5.3   Raw Ensemble

The results from using the raw ensembles as the baseline forecast can be seen in Figure 27. It is clear from the plot that the raw ensembles are underdispersed, as the shape of the plot resembles the c-plot in Figure 4. This indicates that the raw ensembles are uncalibrated, as previously noted in Section 2.2.
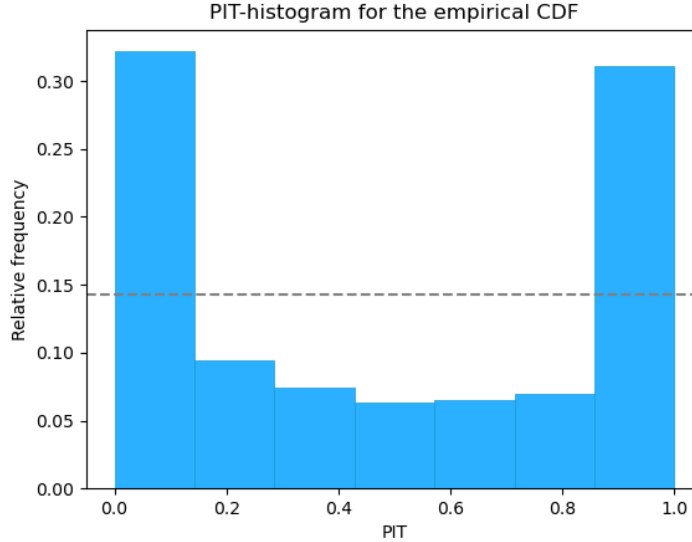
Figure 27: A figure showing a PIT-histogram obtained by using the raw ensembles as a forecast and modeling them using an empirical distribution function.

5.4   Calibration

To evaluate calibration, we use the probability integral transform (PIT) histogram. The Gaussian mixture is well-calibrated, as can be seen in Figure 28. Both the Gaussian mixture and the truncated Gaussian mixture have similar shapes but are mirrored. The Gaussian BMA is biased toward the upper end of the cumulative distribution function (CDF), while the truncated Gaussian BMA is biased toward the lower end of the CDF. For the Cauchy mixtures, shown in Figure 29, the Cauchy BMA appears to be slightly underdispersed, while the truncated Cauchy BMA is both underdispersed and biased towards the left. The Beta mixture, shown in Figure 30, has a similar shape to the truncated Cauchy distribution, with a bias towards the left. However, the rest of the PIT histogram appears to be almost uniform. The leftward bias in the Beta mixture may be due to the zero inflation of the observations, as the lower bound of the predictive distribution is 0, and therefore more of the observations lie in the lower end of the CDF.
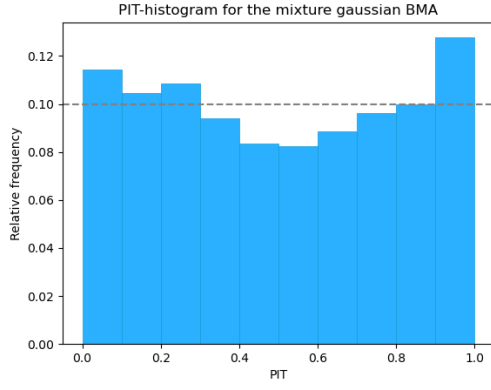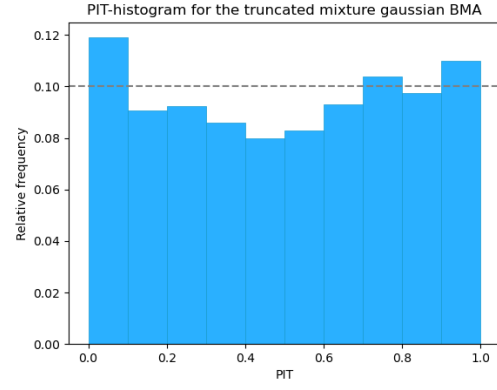
(a) Gaussian mixture   (b) Truncated gaussian mixture

Figure 28: PIT histograms for the Gaussian mixture and the truncated Gaussian mixture.
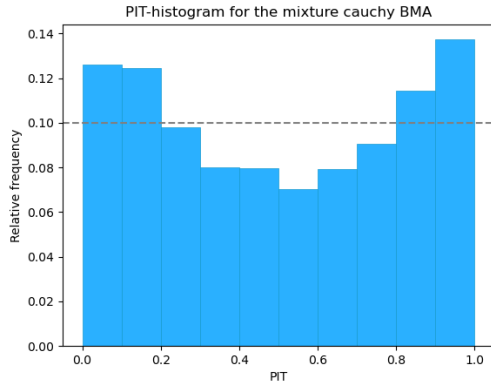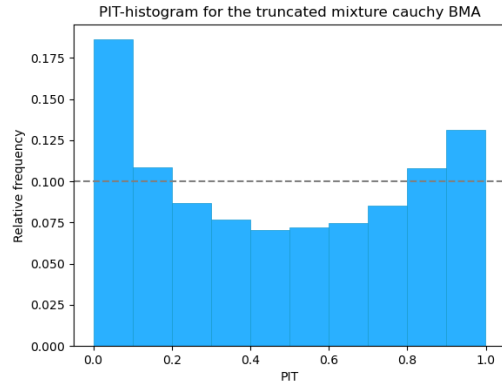
(a) Cauchy mixture   (b) Truncated Cauchy mixture

Figure 29: PIT histograms for the Cauchy mixture and the truncated Cauchy mixture.
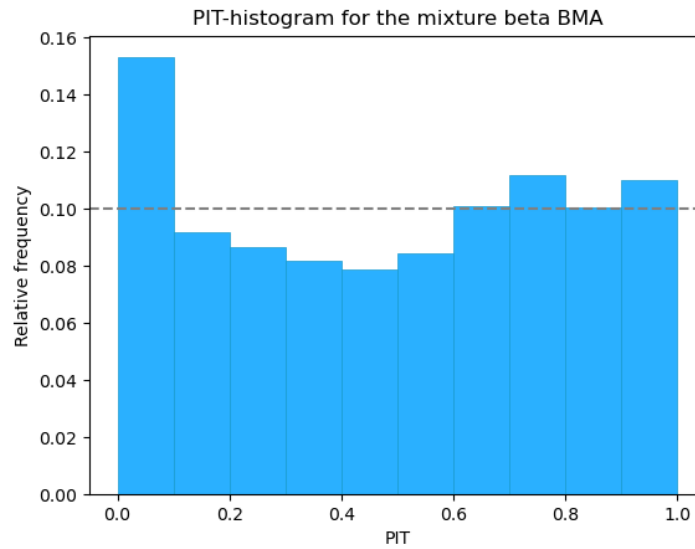
Figure 30: PIT histogram for the Beta BMA mixture.

5.5 Scoring metrics

5.5.1 CRPS and CRPSS

To further assess the performance of the different models, we plotted the average CRPS and CRPSS scores for each location. Figure 31 shows the average CRPS scores for the different models, while Figure 32 shows the average CRPSS scores. It can be seen that all models perform better than the baseline model, except for two locations: Holt and Iskoras II. These locations have a high degree of zero inflation during the winter months, and the truncated Cauchy distribution was not able to produce good predictive distributions. In general, it is difficult to determine which model performs the best based on the plots, but the Beta and Gaussian mixtures seem to perform consistently well across all locations.
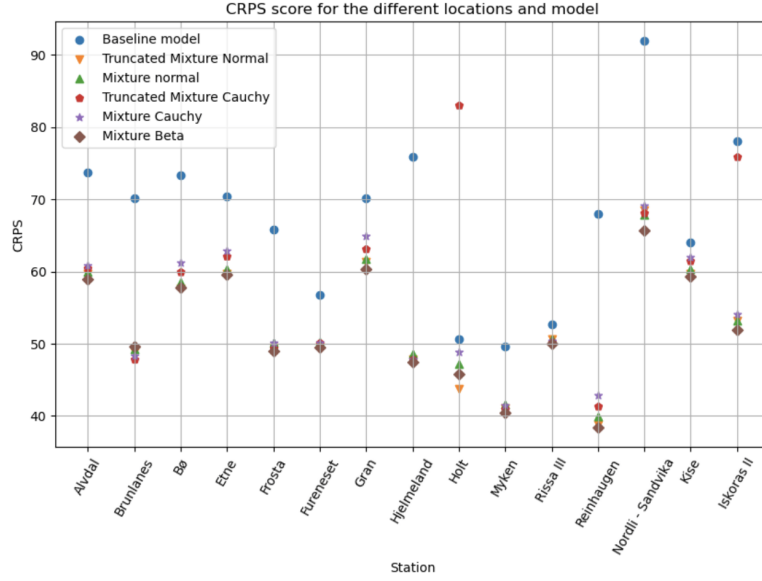
Figure 31: Average CRPS for the different models and locations.
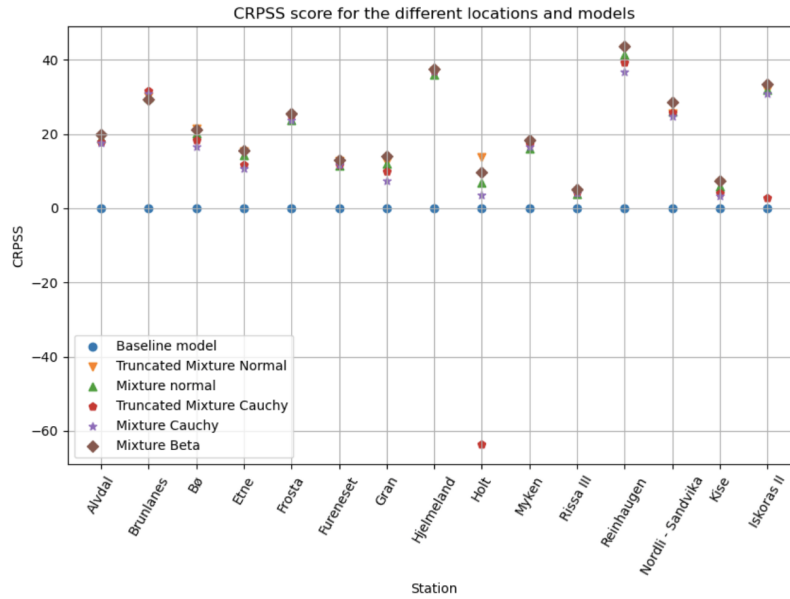
Figure 32: Average CRPSS for the different models and locations.

To further evaluate the performance of the different models, we present a table showing the average CRPSS scores for each location in Figure 33. The model with the best performance in a location is highlighted in green, while the model with the worst performance in a location is highlighted in red. From the table, it is clear that the Beta BMA model performs the best, with the highest CRPSS score in 12 out of 15 locations. The truncated versions of the distributions also tend to perform better than the non-truncated versions. It is interesting to note that the truncated Cauchy mixture performs particularly poorly in Holt, with a CRPSS score that is 63 % worse than the baseline model. However, in Brunlanes, the same model performs almost 32 % better than the baseline model. This variability in performance may be due to numerical instability.

	Baseline model	Truncated Mixture Normal	Mixture normal	Truncated Mixture Cauchy	Mixture Cauchy	Mixture Beta
Alvdal	0.00	19.07	18.78	17.94	17.52	20.04
Brunlanes	0.00	30.26	29.82	31.75	31.02	29.27
Bø	0.00	21.47	20.23	18.26	16.61	21.24
Etne	0.00	15.32	14.32	11.80	10.78	15.50
Frosta	0.00	24.77	23.86	24.94	23.77	25.54
Fureneset	0.00	12.70	11.40	11.65	11.63	12.93
Gran	0.00	12.53	11.89	9.98	7.40	14.01
Hjelmeland	0.00	37.00	35.99	37.02	36.98	37.48
Holt	0.00	13.70	6.87	-63.70	3.67	9.60
Myken	0.00	17.22	16.13	17.49	16.61	18.51
Rissa III	0.00	3.85	3.81	4.39	4.16	5.05
Reinhaugen	0.00	43.04	41.35	39.28	36.90	43.59
Nordli - Sandvika	0.00	25.52	26.16	25.88	24.81	28.51
Kise	0.00	6.84	5.80	4.05	3.26	7.28
Iskoras II	0.00	31.86	31.84	2.79	30.81	33.47

Figure 33: This figure shows a comparison of the CRPSS (a measure of performance relative to the baseline model) for different models in different locations. The model with the best performance in a location is highlighted in green, while the model with the worst performance in a location is highlighted in red. The table allows for easy comparison of the performance of the different models in each location.

5.5.2 RMSE

The results of the RMSE values for the different models and locations are shown in a table in Figure 34. It is notable that the baseline model performs worse in some locations, but also performs better in others. The main conclusion from this table is that the mixture models tend to improve the calibration of the predictive distributions, but do not always improve the deterministic accuracy.

	Baseline model	Truncated Mixture Normal	Mixture normal	Truncated Mixture Cauchy	Mixture Cauchy	Mixture Beta
Alvdal	122.84	117.04	116.76	117.81	118.47	116.04
Brunlanes	122.00	98.08	97.60	99.62	99.76	98.79
Bø	122.16	111.48	112.25	117.25	119.53	111.22
Etne	116.33	116.25	116.93	121.20	123.92	116.26
Frosta	113.77	98.99	99.05	100.32	101.36	98.35
Fureneset	106.11	102.90	103.54	105.41	105.38	102.98
Gran	117.93	118.86	118.78	122.21	124.93	117.54
Hjelmeland	128.15	93.86	94.01	94.95	94.12	93.61
Holt	96.87	98.18	99.37	175.97	104.61	98.38
Myken	89.45	82.86	83.24	83.30	83.86	81.95
Rissa III	97.17	101.35	101.11	102.71	102.33	100.44
Reinhaugen	109.30	83.12	84.08	84.64	88.65	82.23
Nordli - Sandvika	159.69	141.79	141.38	143.36	147.23	140.65
Kise	116.36	116.50	116.59	120.82	122.09	115.79
Iskoras II	123.23	109.63	109.30	149.21	110.68	108.25

Figure 34: This table shows the root mean square error (RMSE) values for different models in different locations. The model that had the best performance in a particular location is highlighted in green, while the model that had the worst performance is highlighted in red. The table demonstrates that the mixture models do not always improve the deterministic accuracy.

5.6 Share of distribution outside physical domain

In this section, we investigate the percentage of the distribution that falls outside the physical domain for the Gaussian and Cauchy mixture models. These models were not designed to be within the physical domain, but it is still important for the probabilistic forecast to have as much of the distribution inside the bounds. The Beta regression model and the truncated mixture models do not have any mass of their distribution outside the physical domain, so they are not considered in this analysis.

We use equation 2.13 to calculate the percentage of the distribution that is outside the physical domain for the Gaussian and Cauchy mixture models. In Figure 35, we present a visualization of the percentage of unphysical distribution for the Gaussian mixture model. This plot shows the average percentage of unphysical distribution across the 15 locations on a given day, with the blue-shaded line representing the standard deviation of the percentage across the locations. We can see that the percentage of unphysical distribution reaches as high as 20 % at times, with a standard deviation of around 10 % higher. This tends to occur during the winter months, possibly due to the low value of solar radiation in these months. As an example, Figure 25 shows a high portion of the distribution falling outside the physical domain when the measured solar radiation is close to 0.
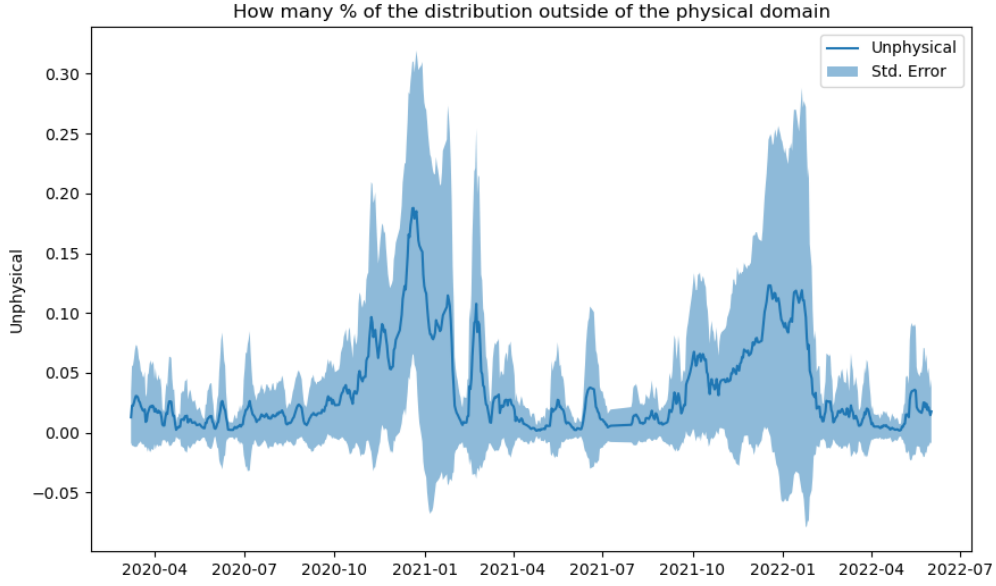
Figure 35: A figure showing the percentage of the probability distribution that is outside the physical domain for the Gaussian mixture model. The blue line represents the mean across all locations, while the blue-shaded region represents the standard error.
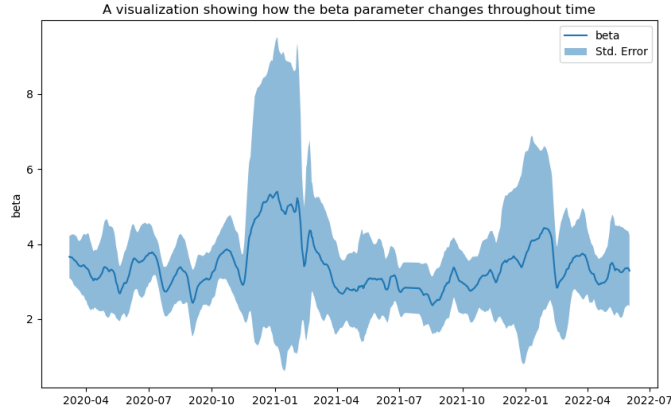
In Figure 36, we present a visualization of the percentage of unphysical distribution for the Cauchy mixture model. This plot shows that the Cauchy mixture model has a higher percentage of unphysical distribution, averaging around 5-10 % with a maximum of 20 %. This may suggest that neither a Gaussian mixture nor a Cauchy mixture is the best fit for the post-processing of solar radiation.
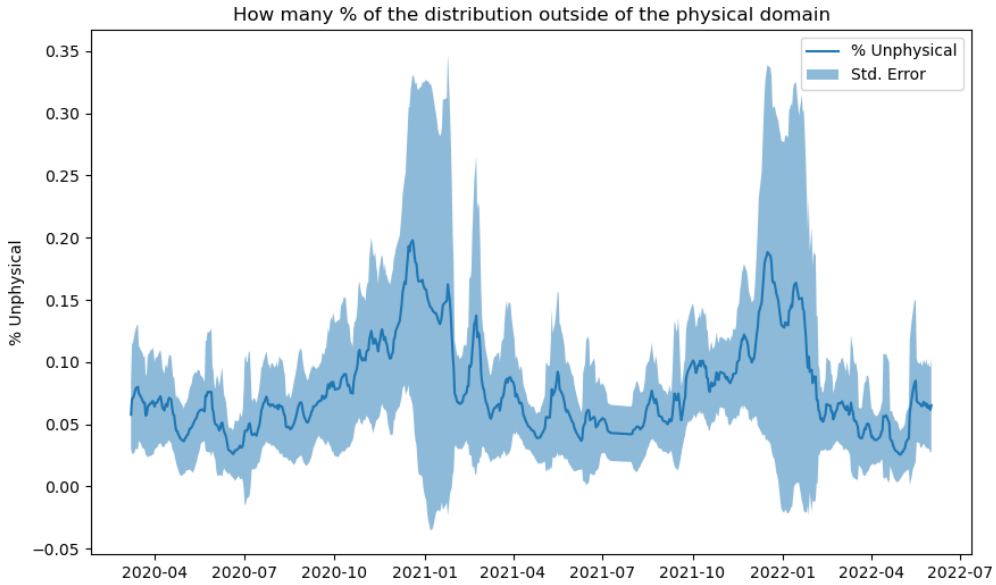
Figure 36: A figure showing the percentage of the probability distribution that is outside the physical domain for the Cauchy mixture model. The blue line represents the mean across all locations, while the blue-shaded region represents the standard error.

In the appendix, a list of stations is provided along with the average percentage of the distribution that falls outside the physical domain. This information is useful in assessing the performance of the Cauchy and Gaussian mixture models, which are not designed to limit the distribution to the physical domain.

5.7 Parameter changes throughout time

The aim of this section is to examine how the parameters of the different models vary over time at the locations where they were trained. First, we present a plot of the change in the σ parameter of the Gaussian mixture model over time in Figure 37. This plot is generated using a 7-day moving average and displays the mean and standard error across all locations.
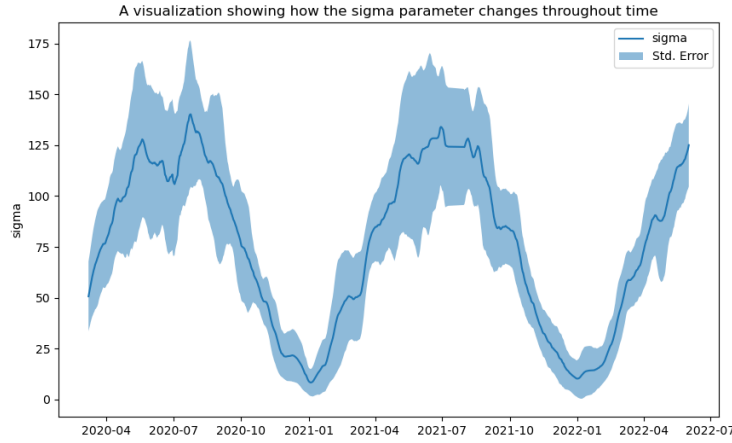
Figure 37: A visualization of how the σ parameter changes through time for the Gaussian mixture. The blue line represents the mean across all locations, while the blue-shaded region represents the standard error.

Next, we present plots showing the changes in the α (Figure 38), β (Figure 39), and ϕ (Figure 40) parameters of the Beta mixture model over time. These plots are also generated using a 7-day moving average and display the mean and standard error across all locations.
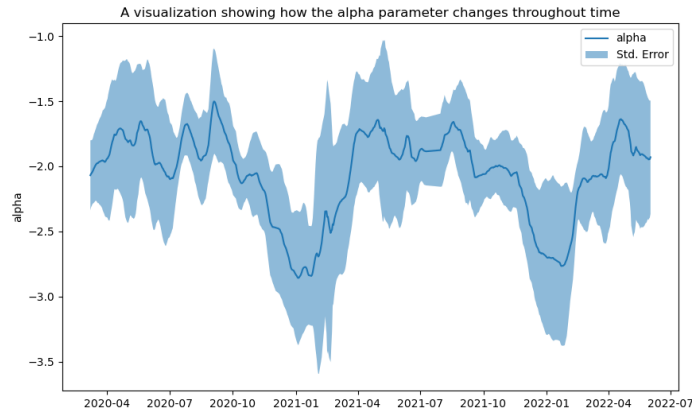
Figure 38: A visualization of how the α parameter changes through time for the Beta mixture. The blue line represents the mean across all locations, while the blue-shaded region represents the standard error.

Figure 39: A visualization of how the β parameter changes through time for the Beta mixture. The blue line represents the mean across all locations, while the blue-shaded region represents the standard error.
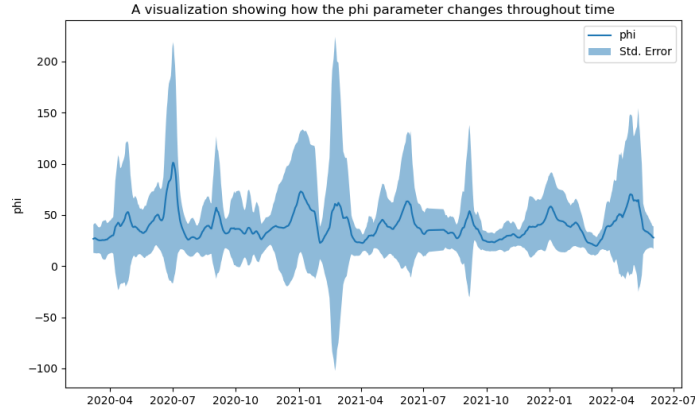
Figure 40: A visualization of how the ϕ parameter changes through time for the Beta mixture. The blue line represents the mean across all locations, while the blue-shaded region represents the standard error.

To further evaluate the generalizability of the method, we analyzed the temporal variations of the parameters in the Gaussian and Beta mixtures. The σ parameter in the Gaussian mixture showed consistent seasonal trends with low standard error, suggesting that it may be possible to use this method in other locations. In contrast, the parameters of the Beta mixture showed higher variations and were more difficult to interpret, indicating that it may be more challenging to use this method in other locations. Overall, these results suggest that the Gaussian mixture model may be more suitable for generalization to other locations, while the Beta mixture model may be more specific to the locations used in this study.

6 Discussion and Conclusion

In this study, five methods for post-processing ensemble forecasts of solar radiation were developed and evaluated based on the Bayesian Model Averaging (BMA) methodology. The models were trained and validated at 15 locations in Norway for a single lead time, using a sliding window of 20 training days to account for seasonal variations. It was found that the BMA using a scaled Beta distribution as a kernel performed best in terms of the mean Continuous Ranked Probability Score (CRPS) in almost all locations, and that the truncated versions of the Cauchy BMA and Gaussian BMA also performed well in most locations. The models constraining the predictive distributions within the physical domain, namely the truncated Normal, truncated Cauchy, and Beta distributions, performed better overall.

In the original BMA methodology, weights for each probability density function (PDF) are based on a weighted average of how well each forecast performed during a training period. In this study, however, all ensemble members were given equal weights, which is a valid approach if the ensemble members are interchangeable. It is worth considering the potential benefits of estimating the weights during the training period, as the control ensemble had a better performance than the other members.

For inference, maximum likelihood estimation was used to estimate the model parameters. An alternative method that could be explored in future research is CRPS minimization, as it aligns with the goal of minimizing the CRPS for the predictive distributions.

While the models developed in this study showed good performance for the locations and time period analyzed, further research is needed to examine their generalizability to other locations and times.

Bibliography

CRPSS. https://confluence.ecmwf.int/display/FUG/12.B+Statistical+Concepts+-+Probabilistic+ Data#id-12.BStatisticalConceptsProbabilisticData-ContinuousRankedProbabilitySkillScore(CRPSS). Accessed: 2022-12-20.

Global horizontal irradiance (ghi). https://www.homerenergy.com/products/pro/docs/3.11/global_ horizontal_irradiance_ghi.html. Accessed: 2022-12-01.

Om lmt. https://lmt.nibio.no. Accessed: 2022-11-27.

What is the value of accurate solar forecasting for utility-scale pv plants? https://www. solaranywhere.com/2017/accurate-solar-forecasting-value/, a. Accessed: 2022-11-26.

Solaratlas. https://globalsolaratlas.info/mapc=11.523088,8.4375,3, b. Accessed: 2022-09-20.

Solargis. https://solargis.com/maps-and-gis-data/download/norwa, c. Accessed: 2022-09-20.

How solargis is improving accuracy of solar power forecasts. https://solargis.com/blog/ best-practices/improving-accuracy-of-solar-power-forecasts, d. Accessed: 2022-11-26.

Mohamed R Abonazel, Zakariya Yahya Algamal, Fuad A Awwad, and Ibrahim M Taha. A new two-parameter estimator for beta regression model: method, simulation, and application. *Front. Appl. Math. Stat. 7: 780322. doi: 10.3389/fams*, 2022.

Bilal Babar. Solar resource assessment at high latitude regions. 2019.

Sándor Baran and Sebastian Lerch. Mixture emos model for calibrating ensemble forecasts of wind speed. *Environmetrics*, 27(2):116–130, 2016.

Llinet Benavides Cesar, Rodrigo Amaro e Silva, Miguel Ángel Manso Callejo, and Calimanut-Ionut Cira. Review on spatio-temporal solar forecasting methods driven by in situ measurements or their combination with satellite and numerical weather prediction (nwp) estimates. *Energies*, 15 (12):4341, 2022.

Jochen Broecker and Holger Kantz. The concept of exchangeability in ensemble forecasting. *Nonlinear Processes in Geophysics*, 18(1):1–5, 2011.

E. G. Dutton C. N. Long. Bsrn global network recommended qc tests, v2.0.

A. Mørk Cato. Sol kan bli like stort som vannkraft i norge. 2022.

Cauchy distribution. Cauchy distribution — Wikipedia, the free encyclopedia. https://en.wikipedia. org/wiki/Cauchy_distribution. Accessed: 2022-12-03.

Mathieu David and Philippe Lauret. Solar radiation probabilistic forecasting. *Wind Field and Solar Radiation Characterization and Forecasting*, pages 201–227, 2018.

ECMWF. Part iv: Physical processes. *IFS Documentation CY41R2. ECMWF*, 2016.

Emmanuel B Ettah and Josiah Obiefuna. The relationship between solar radiation and the efficiency of solar panels in port harcourt, nigeria. *International Journal of Applied Science and Technology*, 1.4, 2011.

Inger-Lise Frogner, Ulf Andrae, Jelena Bojarova, Alfons Callado, PAU Escribà, Henrik Feddersen, Alan Hally, Janne Kauhanen, Roger Randriamampianina, Andrew Singleton, et al. Harmoneps—the harmonie ensemble prediction system. *Weather and Forecasting*, 34(6):1909–1937, 2019.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

Tilmann Gneiting, Adrian E Raftery, Anton H Westveld, and Tom Goldman. Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. *Monthly Weather Review*, 133(5):1098–1118, 2005.

Alexander Jordan, Fabian Krüger, and Sebastian Lerch. Evaluating probabilistic forecasts with scoringrules. *arXiv preprint arXiv:1709.04743*, 2017.

Andreas Kleiven. Bayesian model averaging using varying coefficient regression and climatology cumulative probability regression-a case study of postprocessing hydrological ensembles from osali. Master's thesis, NTNU, 2017.

Philippe Lauret, Mathieu David, and Hugo TC Pedro. Probabilistic solar forecasting using quantile regression models. *energies*, 10(10):1591, 2017.

M Noia, CF Ratto, and R Festa. Solar irradiance estimation from geostationary satellite data: I. statistical models. *Solar Energy*, 51(6):449–456, 1993.

Probabilistic forecasting. Probabilistic forecasting — Wikipedia, the free encyclopedia. `https://statisticaloddsandends.wordpress.com/2021/03/27/what-is-a-proper-scoring-rule/`. Accessed: 2022-12-03.

Proper Scoring Rules. What is a proper scoring rule? `https://en.wikipedia.org/wiki/Probabilistic_forecasting`. Accessed: 2022-12-03.

Adrian E Raftery, Tilmann Gneiting, Fadoua Balabdaoui, and Michael Polakowski. Using bayesian model averaging to calibrate forecast ensembles. *Monthly weather review*, 133(5):1155–1174, 2005.

Ibrahim Reda and Afshin Andreas. Solar position algorithm for solar radiation applications. *Solar energy*, 76(5):577–589, 2004.

Hadrien Verbois, Yves-Marie Saint-Drenan, Alexandre Thiery, and Philippe Blanc. Statistical learning for nwp post-processing: A benchmark for solar irradiance forecasting. *Solar Energy*, 238:132–149, 2022.

Remco A Verzijlbergh, Petra W Heijnen, Stephan R de Roode, Alexander Los, and Harm JJ Jonker. Improved model output statistics of numerical weather prediction based irradiance forecasts for solar power applications. *Solar Energy*, 118:634–645, 2015.

WMO. Guidelines on ensemble prediction systems and forecasting. *World Meteorological Organization Weather Climate and Water*, 1091, 2012.

Michaël Zamo and Philippe Naveau. Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts. *Mathematical Geosciences*, 50(2):209–234, 2018.

Noshipo Zwane, Henerica Tazvinga, Christina Botai, Miriam Murambadoro, Joel Botai, Jaco D Wit, Brighton Mabasa, Siphamanadla Daniel, and Tafadzwanashe Mabhaudi. A bibliometric analysis of solar energy forecasting studies in africa. *Energies*, 15, 2022.

Appendix

A   All locations

Navn	Latitude	Longitude	Altitude	Used in project
Alvdal	62.10944	10.62687	478	X
Apelsvoll	60.70024	10.86952	262	
Brunlanes	58.9878	9.96187	20	X
Bø	59.4175	9.02859	105	X
Etne	59.6625	5.95383	8	X
Frosta	63.56502	10.69298	18	X
Fureneset	61.29272	5.04428	12	X
Fåvang	61.45822	10.1872	184	
Gausdal	61.22468	10.25878	375	
Gjerpen	59.22684	9.57805	41	
Gran	60.35575	10.55906	245	X
Hjelmeland	59.22995	6.14992	43	X
Hokksund	59.76152	9.89166	15	
Holt	69.65381	18.90946	12	X
Hønefoss	60.14032	10.2661	126	
Ilseng	60.80264	11.20298	182	
Kise	60.77324	10.80569	129	X
Kvithamar	63.48795	10.87994	28	
Landvik	58.340071	8.522554	10	
Særheim	58.7605	5.6508	87	
Midtstova	60.6563	7.2755	1162	
Njøs	61.1792	6.8608	45	
Hansbu	60.0825	7.4247	1160	
Årnes	60.1268	11.3933	160	
Nordli - Sandvika	64.4595	13.5978	420	X
Skjetlein	63.3403	10.2973	48	
Rissa III	63.5858	9.9705	23	X
Lebergsfjellet	62.5158	6.8717	625	
Surnadal - Sylte	62.9853	8.6895	5	
Linge	62.288	7.2173	34	
Møsstrand II	59.8397	8.1785	977	
Myken	66.7628	12.486	17	X
Tjølling	59.0467	10.125	19	
Losistua	68.1905	17.7905	740	
Valnesfjord	67.2763	15.1018	20	
Lyngdal	58.134	7.0452	6	
Landvik	58.34	8.5225	6	
Nordnesfjellet	69.5575895	20.4152984	697	
Pasvik - Svanvik	69.4552	30.041	27	
Reinhaugen	70.3357	28.9648	470	X
Iskoras II	69.3003	25.346	591	X

B   Share of distribution outside unphysical domain

Location	Gaussian mixture	Cauchy mixture
Alvdal	1.5 %	7.4 %
Brunlanes	2.4 %	4.7 %
Bø	3.2 %	7.1 %
Etne	1.7 %	5.8 %
Frosta	2.6 %	7.0 %
Fureneset	2.5 %	5.8 %
Gran	3.0 %	7.6 %
Hjelmeland	2.3 %	5.9 %
Holt	4.5 %	9.1 %
Myken	2.5 %	6.7 %
Rissa III	2.7 %	6.9 %
Reinhaugen	6.6 %	12.3 %
Nordli - Sandvika	5.6 %	9.8 %
Kise	2.6 %	6.6 %
Iskoras II	7.1 %	12.7 %

List of Figures

List of Tables