NTNU – Trondheim
Norwegian University of
Science and Technology

Climatology Cumulative Probability Regression

A Postprocessing Methodology Based on Climatology and Deterministic Forecasts, With a Case Study of Streamflow Forecasts at Osali

Johanne Borhaug

Teacher Education with Master of Science
Submission date: May 2014
Supervisor: Ingelin Steinsland, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

This study introduce a new postprocessing methodology for constructing probabilistic forecasts based on climatology and deterministic forecasts. The Climatology Cumulative Probability Regression (CCPR) methodology is based on transforming the climatology cumulative distribution function (cdf) to a new probabilistic forecast, where the transformation procedure is determined by the deterministic forecasts. We base the transformation on fitting a beta cdf on the scale of climatology cumulative probabilities (CCP-scale). The mean of the beta pdf is modelled by a logit link where the linear predictor have different forecasts as covariates. This methodology is flexible to include different forecasts and lead times. The methodology was tested for streamflow data at the catchment Osali in south western Norway for four different lead times in the period 01.09.2005-31.07.2009. In the case study, we applied the methodology where we successively added more deterministic forecasts into the model, starting with the hydrological forecast, adding the persistence forecast and finally adding the sliding window climatology forecast. When evaluating predictive performance using cross validation, the case study found that the inclusion of the persistence forecast is important for short lead times. When both the hydrological and the persistence forecast was included, the sliding window climatology forecast only added little extra predictive information.

Sammendrag

I dette studiet introduserer vi en ny postprosesserings metodologi for å konstruere probabilistiske varsel, basert på klimatologi og deterministiske varsel. Metodologien Climatology Cumulative Probability Regression (CCPR) er basert på en transformasjon av den klimatologiske kumulative fordelingsfunksjonen til et nytt probabilistisk varsel, der transformasjonen er bestemt av ulike deterministiske varsel. Transformasjonen baserer seg på å tilpasse en beta kumulativ fordelingsfunksjon på skalaen av klimatologiske kumulative sannsynligheter. Forventningsverdien i beta fordelingen er modellert gjennom en logit link, der den lineære prediktoren består av deterministiske varsel som kovariater. Metodologien er fleksibel til å inkludere både ulike varsel og ledetider. Vi testet metodologien for tilsigsdata ved nedslagsfeltet Osali i sør-vest Norge for fire ledetider i perioden 01.09.2005-31.07.2009. I case-studiet ble metodologien anvendt, der vi suksessivt la til flere deterministiske varsel i modellen. Først begynte vi med et hydrologisk varsel, deretter la vi til et persistensvarsel før vi til slutt la til et deterministisk varsel basert på klimatologien. Modellenes prediktive ytelse ble evaluert ved hjelp av kryssvalidering og case-studiet viste at det var viktig å inkludere persistensvarselet for korte ledetider. Når både det hydrologiske varselet og persistensvarselet var inkludert i modellen, viste det seg at det deterministiske varselet basert på klimatologi kun gav små forbedringer med hensyn på prediktiv ytelse.

Preface

This Masters's thesis is the product of my final semester as a student at the Natural Science with Teacher Education programme at NTNU. The work was carried out in the spring 2014.

First I would like to thank Oddbjørn Bruland and Robert von Hirsch from Statkraft for guidance and good discussions during my internship, summer 2013. They both made me interested in the field of hydrology and taught me a lot, which in turn motivated me when writing this Master's thesis. I would also like to thank Kolbjørn Engeland, Anders S. Lund and Henriette R. Tufte, respectively for providing data and figures, LaTeX- guidance and graphical-support. In addition a great thanks goes to the people at "Matteland", this year would not have been the same without you! Last, but definitely not least, I want to thank my supervisor, Professor Ingelin Steinsland for her excellent guidance throughout this last year of my studies. Our weakly meetings has encouraged me to work steady and your ideas and comments has contributed a lot to this thesis. Especially I would like to thank you for your patience and for always being positive.

Contents

1. Introduction

Water resources are a vital resource for society, but can also be viewed as a threat. Hydropower production gives rise to about 19% of all the electricity production in the world and is the largest renewable energy source contributing to electricity production. Norway has the highest yearly hydropower production in Europe, with Statkraft as a leading power company [Senter for fornybar energi, 2011]. Thus hydropower production is valuable to Norway in terms of wealth creation. However, handling of water resources also require a great responsibility in preventing environmental damages due to changes in the natural streamflow. Both to assess the risk of flooding and to optimize hydropower production, streamflow forecasts is an important tool. Statkraft use streamflow forecasts to make decisions concerning how much electricity they will produce and trade, both from day to day and when planning long-term. Their choices of actions are influenced by the weather conditions including temperature, precipitation and streamflow. These choices depend on many other variables as well, such as the energy price, the magazine state and production capacity [Engeland and Steinsland, 2014]. However, to be able to make beneficial decisions, good forecasts are needed.

Forecasts are often based on physical models, such as meteorological models for meteorological forecasts and hydrological models for hydrological forecasts. A forecast will always be associated with uncertainty due to e.g errors in input data, errors in the model parameters and imperfectness of the models [Engeland and Steinsland, 2014; Lewis, 2005]. Until the early 1990s deterministic forecasts, based on numerical weather prediction models, was the most common type of forecast [Gneiting and Raftery, 2005]. To be able to represent the uncertainty in a forecast, probabilistic forecasts are often preferred over deterministic forecasts. A probabilistic forecast is a forecast that takes the form of a predictive probability distribution function (pdf) or a predictive cumulative distribution function (cdf) [Gneiting et al., 2007]. The value of a probabilistic forecast lies in the ability of evaluating different actions according to the probabilities of different events. For this reason probabilistic forecasts has gained an increasing influence in many

applications such as economics, finance and meteorology [Ranjan and Gneiting, 2013].

To assess probabilistic forecasts, two types of qualities are often evaluated; calibration and sharpness. Calibration refers to the statistical consistency between the probabilistic forecast and the observation [Gneiting et al., 2007]. A forecast is calibrated if an event forecasted with probability p actually occur an amount of p on average [Gneiting et al., 2005]. PIT-histograms are often used to assess calibration of a probabilistic forecast. Sharpness refers to the concentration of the probabilistic forecast, with more concentrated as sharper. Hence sharpness is only a property of the forecast [Gneiting et al., 2007]. According to Gneiting et al. [2005] the aim of probabilistic forecasting is to maximize the sharpness of the predictive density subject to calibration. Put in other words; among calibrated forecasts, the best forecast is the sharpest, leading to shorter prediction intervals. Sharpness can be assessed by considering the average length of prediction intervals. In addition, to evaluate calibration and sharpness simultaneously, the continuous ranked probability score (CRPS) is often used. The CRPS is a proper scoring rule, which means that a forecaster achieves the best score by forecasting his true beliefs [Gneiting et al., 2005].

Followed by the position of probabilistic forecasts in many fields, different methods for creating probabilistic forecasts has been developed. Bayesian Model Averaging is a widely used postprocessing method that creates a predictive Gaussian density based on ensemble forecasts [Raftery et al., 2005]. An ensemble forecast is generated by several runs of a numerical weather prediction model, where the initial conditions and/or the numerical representation of the atmosphere differ [Gneiting and Raftery, 2005]. When different probabilistic forecasts are available, many methods has been developed to combine them and create a new and better forecast. Linear pooling is such a method that combines different probabilistic forecasts into a single combined forecast, where the idea is to weight each forecast based on their respective performance over a training period [Ranjan and Gneiting, 2013]. The concept behind linear pooling is easy to understand and sounds appealing, however Hora [2004] and Ranjan and Gneiting [2013] showed that if each of the individual predictive densities are calibrated, every nontrivial linear pool is uncalibrated. These results suggest that linear pools may be suboptimal, and it has been shown empirically that non-linear combination formulas can outperform linear methods [Ranjan and Gneiting, 2010; Allard et al., 2012; Ranjan and Gneiting, 2013]. One of the non-linear combination formulas is the beta-transformed linear pool (BLP). This method weights different predictive cumulative distribution functions and transforms the resulting cdf using a beta cdf [Ranjan and Gneiting, 2013].

In this Master's thesis we develop a, as far as we are aware of, new postprocessing methodology that captures the situation where you have historical observations, i.e the climatology, and different deterministic forecasts available and the aim is to create a probabilistic forecast. This postprocessing methodology can be useful in many different settings, as long as the situation described above holds. The climatology cdf is included in the methodology as a special case, namely if the deterministic forecasts do not contain any predictive information. The methodology is inspired by the BLP methodology, but use only one predictive cdf, namely the climatology and instead models the beta cdf based on the value of the different deterministic forecasts. The estimation of the parameters, used to model the beta cdf is done by minimum CRPS estimation. Minimum CPRS estimation is a procedure fitting the parameters of the model that yields the lowest CRPS-value over the training period [Gneiting et al., 2005].

To evaluate the postprocessing methodology we use streamflow data from the catchment Osali in south-western Norway, which is part of the Ulla-Førre hydropower complex. The entire complex was studied extensively by Engeland and Steinsland [2014], while this study can be viewed as an extension of their work. We apply the postprocessing methodology on daily forecasts provided up to 10 days before the validating observation. In our case study we consider the forecasts given $1, 2, 5$ and 10 days ahead of the validating observation. The main focus of this thesis is to develop a new postprocessing methodology to generate probabilistic forecasts and to evaluate the methodology.

The thesis is organized as follows: The next chapter introduce the data used in the case study including a brief exploratory analysis. Chapter 3 gives an overview of the background theory needed to develop the postprocessing methodology introduced in Chapter 4, and in addition present assessment methods to evaluate probabilistic forecasts. In Chapter 5 we present three models used in the case study based on the postprocessing methodology in Chapter 4. The results of the case study is given in Chapter 6, where we apply the postprocessing methodology to the data described in Chapter 2. We conclude the thesis with a discussion in Chapter 7, where we summarize the results and suggest possibilities for further work.

2. Data and case: Probabilistic forecast of streamflow at Osali

This chapter gives an overview of the data used in the case study performed in Chapter 6. Section 2.1 provides an introduction to the study area Osali. Section 2.2 introduce the observations and forecasts and provides a brief exploratory analysis of how the forecasts relates to the observed streamflow.

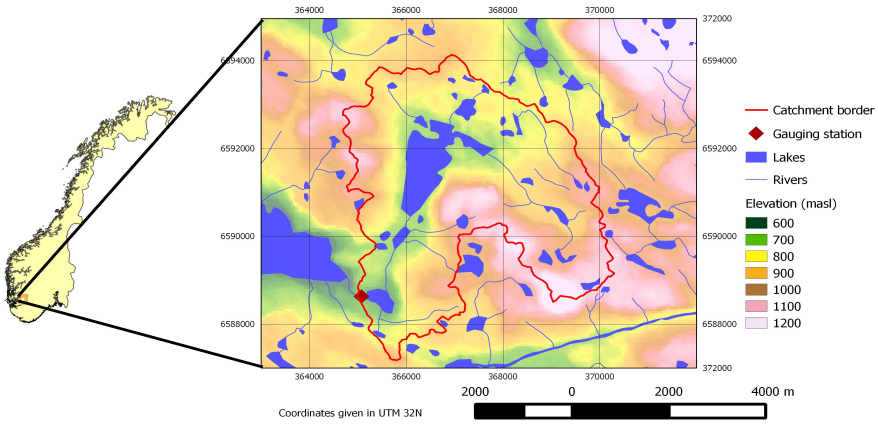


Figure 2.1: The case study region Osali, located in south-western Norway.

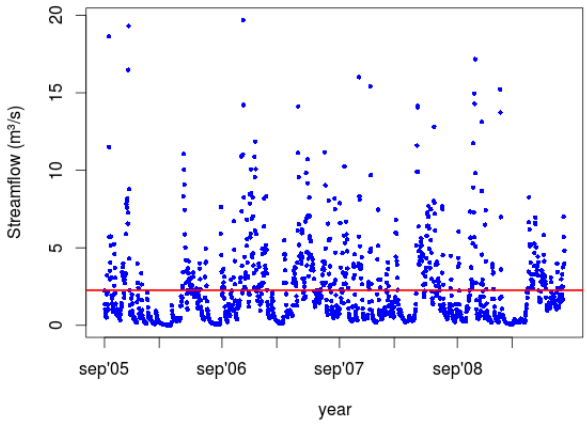


Figure 2.2: Daily observed streamflow at Osali in the period 01.09.2005-31.07.2009. The red line represents the mean observed streamflow which is 2.26 m^3/s.

2.1 The study area

In this study we consider the catchment Osali (23 km^2), which is part of the Ulla-Førre hydropower complex in south-western Norway (Figure 2.1) [Engeland and Steinsland, 2014]. This area has one of the largest hydropower reservoirs in Norway, Blåsjø (3.1 km^3). The complex consists of a system of several hydropower plants with an annual average production of 4.5 GWh. The average elevation of Osali is 890 meters above sea level, which is based on a digital elevation model with resolution 1×1 km^2 [Engeland and Steinsland, 2014]. For the period we consider in the case study (01.09.2005 - 31.07.2009), the mean temperature at Osali was 3.17 °C and the mean amount of precipitation per day was 9.15 mm.

2.2 Hydrological data and forecasts

Streamflow data for Osali was obtained from The Norwegian Water Resources and Energy Directorate for the period 01.01.1982-31.12.2012. Figure 2.2 shows the observed streamflow in the period 01.09.2005-31.07.2009, which is the period we make forecasts for in the case study. We denote the observations y_t, where $t = 1, \ldots, 1430$ is the number of the day in the forecast period 01.09.3005-31.07.2009. The observed values y_t seems to be seasonally stationary in that the mean is the same each year. Temperature and precipitation observations were

provided by the Norwegian Meteorological institute and Statkraft. Deterministic forecasts for 2 m temperature and precipitation were obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) [Engeland and Steinsland, 2014]. Through the next sections we present three different types of forecasts; the hydrological forecast, the persistence forecast and the sliding window climatology forecast. These forecasts are identical to the forecasts used by Engeland and Steinsland [2014], and further details can by found in their paper.

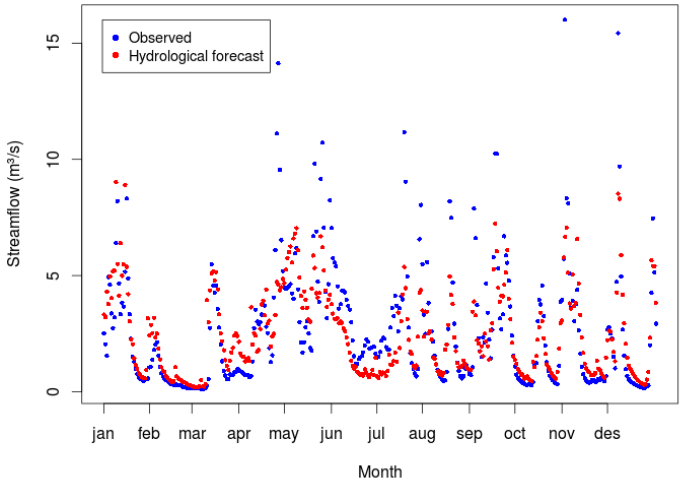


Figure 2.3: Daily observed streamflow y_t (blue dots) at Osali in the period 01.01.2007-31.12.2007. The red dots represent the hydrological forecast for lead time 1, $x^h_{t,1}$.

2.2.1 Hydrological forecast

The hydrological forecasts denoted $x^h_{t,l}$ for lead times $l = 0, 1, \ldots, 10$ days for each day t are available in the period 01.09.2005-31.07.2009. The forecast valid time is given by t and the issue time is $t - l$. Hence, the forecast $x_{t,l}$ is issued at time $t-l$, and the validating observation is y_t. The forecast for lead time 0 is obtained by running the hydrological model with observed temperature and precipitation as input data. In this study we only report results for lead time $1, 2, 5$ and 10. Figure 2.3 displays the observed streamflow y_t and the hydrological forecast for lead time 1, $x^h_{t,1}$, in 2007. We observe that the hydrological forecast follows the observations quite well, but especially the high streamflows are not captured by the hydrological forecast.

Figure 2.4 contains four boxplots of the difference between the observed streamflow and the hydrological forecast, for the lead times $l = 1, 2, 5$ and 10. We observe that the spread is larger for longer lead times, as expected. We also note that the median is a bit below zero for all lead times, hence more errors are negative than positive. However there are more large positive errors than negative errors for all lead times. Since streamflow is a strictly positive quantity, these results indicate that the hydrological forecast, in some cases, is unable to forecast the large streamflows at the right time.

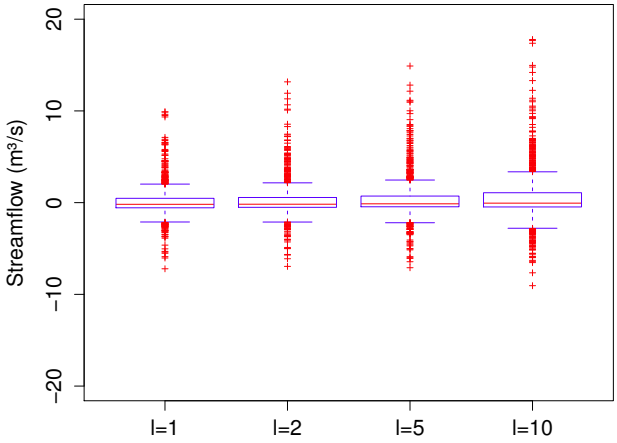


Figure 2.4: A boxplot of the error $\epsilon^h_{t,l} = y_t - x^h_{t,l}$, showing the variation of errors between observed streamflow and the hydrological forecast in the period 01.09.2005 − 31.07.2009 for the lead times $l = 1, 2, 5$ and 10. The blue box represents the 1st to 3rd quartile, the red line is the median and the red crosses represents the outliers.

2.2.2 Persistence forecast

A persistence forecast is to simply use the last available observation as your forecast, assuming that the streamflow remains constant. The persistence forecast denoted $x^p_{t,l}$ with issue time $t - l$ and valid time t is given by the last observed value at the issue time, namely $x^p_{t,l} = y_{t-l}$. As an example, if you want to make a persistence forecast of y_m at the time $t = m - 5$, the persistence forecast will be the observed streamflow at time $t = m - 5$, namely $x^p_{m,5} = y_{m-5}$.

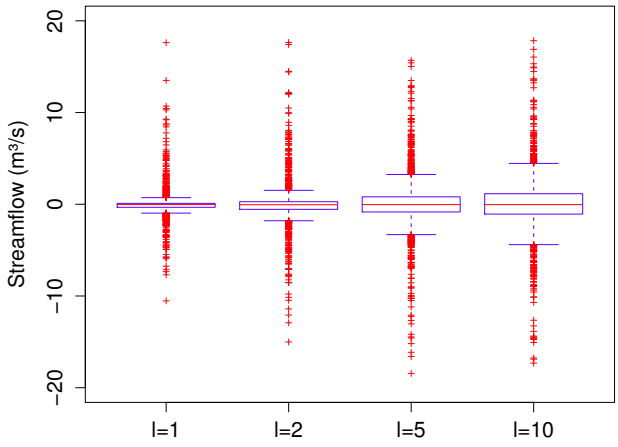


Figure 2.5: A boxplot of the error $\epsilon^p_{t,l} = y_t - x^p_{t,l}$, showing the variation of errors between observed streamflow and the persistence forecast in the period 01.09.2005 − 31.07.2009 for the lead times $l = 1, 2, 5$ and 10. The blue box represents the 1st to 3rd quartile, the red line is the median and the red crosses represents the outliers.

Figure 2.5, displays a boxplot of the difference between observed streamflow and persistence forecast for lead time $l = 1, 2, 5$ and 10. We observe that the spread increases with lead time, as it did for the hydrological forecast. For lead time 1 and 2, the boxes including the central 50% errors are shorter than for the hydrological forecast given in Figure 2.4 . This indicates that for shorter lead times, the persistence forecast can be more precise than the hydrological forecast. However the outliers for lead time 1 and 2 covers a wider interval for the persistence forecast than the hydrological forecast, which indicates that larger errors occur for the persistence forecast. We also note that the median is close to zero and centred in the boxes, with outliers fairly equally distributed above and below zero, hence the persistence forecast is unbiased.

2.2.3 Sliding window climatology forecast

The sliding window climatology forecast is constructed using the same approach as Engeland and Steinsland [2014], but we have a longer period of historical ob-

servations available. The sliding window climatology forecast x_t^c, with validating observation y_t, was constructed assigning each day of the year the median value of a 15-day window centred at the day of interest, based on the period 1982-2004. As an example, the sliding window climatology for January 13 is the median of the streamflow observations from January 6 to January 20 in the period 1982-2004. This deterministic forecast reflects the usual streamflow at the given day. Figure 2.7 shows the observed streamflow in 2007 and the sliding window climatology. As expected, the sliding window climatology has less variance than the observed values, because it is based on the median value over several years. We also observe that the period from May to July has the highest streamflows, which coincides with the major snow melting period at Osali [Engeland and Steinsland, 2014]. Figure 2.6 displays a boxplot of the error between the observed streamflow and the sliding window climatology forecast. The median is close to zero, but there is a positive skewness, as is also observed in Figure 2.7 where the sliding window climatology is more often below than above the observations.

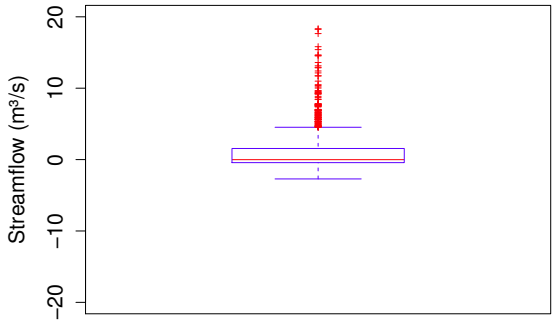


Figure 2.6: A boxplot of the error $\epsilon_t^c = y_t - x_t^c$, showing the variation of errors between observed streamflow and hydrological forecast in the period 01.09.2005 − 31.07.2009. The blue box represents the 1st to 3rd quartile, the red line is the median and the red crosses represents the outliers.

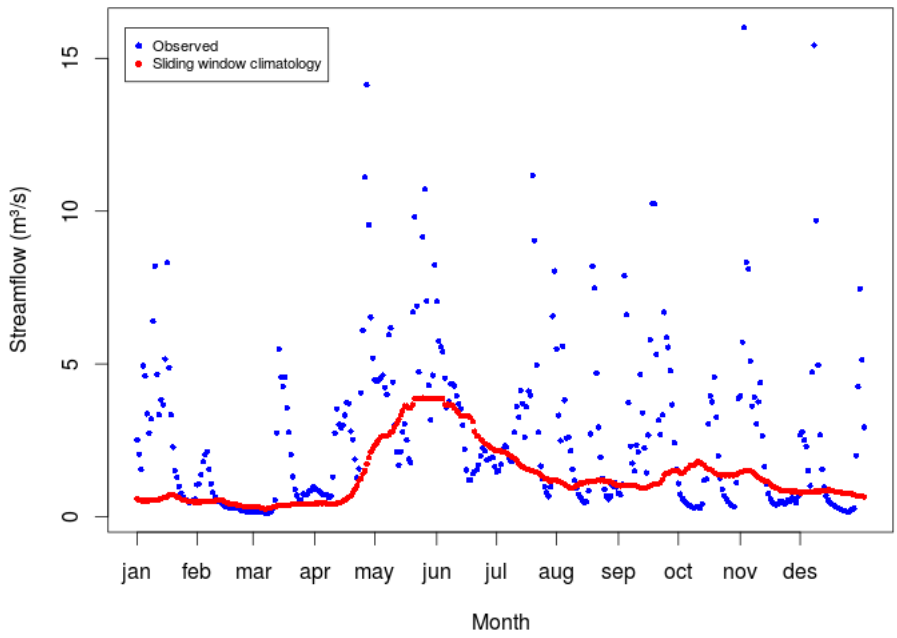


Figure 2.7: Observed streamflow and sliding window climatology in the period 01.01.2007-31.01.2007.

2.3 Lead times and vector of forecasts

The notation with lead times can be a bit cumbersome to keep track of for the different types of forecast. To clarify the notation used so far, which will also be used in Chapter 4 and 5, we introduce a vector of forecasts for a given valid time t and lead time l, which will be used to make a probabilistic forecast for y_t. We denote the vector including all forecasts by $\mathbf{x}_{\mathbf{t,l}}^{\mathbf{h,p,c}} = (x_{t,l}^h, x_{t,l}^p, x_t^c)$, where t is the forecast valid time, l is the lead time and the superscript h, p, c denotes that the hydrological forecast, persistence forecast and sliding window climatology forecast are included. Note that the hydrological forecast and the persistence forecast will depend on lead time. The sliding window climatology forecast however, will always forecast the median of the sliding window period centred at the day of

interest, y_t. As an example, consider the forecast given 1 September 2005, to predict the streamflow 6 September 2005. The forecast vector used to create a probabilistic forecast for y_6 is then $\mathbf{x}_{\mathbf{6,5}}^{\mathbf{h,p,c}} = (x_{6,5}^h, x_{6,5}^p, x_6^c)$. The hydrological forecast is then the forecast issued 1 September 2005 to predict the streamflow at 6 September 2005. The persistence forecast is the observed streamflow at September 1. The sliding window climatology forecast is the median of the streamflow observations from August 30 to September 13 in the period 1982-2004, hence centred at 6 September.

3. Background

This chapter present relevant background theory that we use in Chapter 4 to develop our post-processing methodology, or in Chapter 6 to evaluate the methodology. In section 3.1 we describe some quality measures of probabilistic forecasts and how to assess these measures. Section 3.2 provides a brief summary of the beta distribution and some of its properties. In this section we also consider a re-parametrization of the beta distribution in terms of its mean and a variance parameter. Section 3.3 describes the BLP-method, a method for producing calibrated probabilistic forecasts based on a beta transformation of predictive cumulative distribution functions. This method defines a framework that inspired the development of the CCPR-methodology presented in Chapter 4. The last section describes the logit link which is used to model a response that is restricted on the interval $(0, 1)$, which we in turn apply to develop our post-processing methodology for probabilistic forecasts.

3.1 Assessment methods

To evaluate different probabilistic forecasts it is necessary to have some criteria for goodness. Calibration and sharpness are often used as criteria describing goodness of the forecast. Calibration refers to the statistical consistency between the probabilistic forecasts and the observations and is thus a joint property of the observations and the forecast [Gneiting et al., 2007]. A forecast is calibrated if an event forecasted with probability p actually occur with a relative frequency of p [Gneiting et al., 2005]. For a probabilistic forecast, calibration is achieved if a p-percent prediction interval, in the long run, include p percent of the observations. One way of assessing calibration is by the probability integral transform (PIT) histogram, given in section 3.1.1 [Gneiting et al., 2007]. Sharpness refers to the spread of the probabilistic forecast, and is only a property of the forecast. Gneiting et al. [2005] states that the aim of probabilistic forecasting is to maximize the sharpness of the probabilistic forecast subject to calibration.

Hence, when selecting the best probabilistic forecast among calibrated forecasts, one would choose the probabilistic forecast with the lowest spread. Sharpness can be evaluated by comparing the length of prediction intervals [Gneiting et al., 2007]. In section 3.1.3 we introduce a scoring rule that assess both calibration and sharpness, namely the continuous ranked probability score (CRPS).

3.1.1 Assessing calibration

The probability integral transform (PIT) histogram is a method for assessing calibration [Gneiting et al., 2007; Ranjan and Gneiting, 2013]. Let $F(\cdot)$ denote a continuous predictive cumulative distribution function and y the observed value. The probability integral transform of y is defined as

$$z = F(y) \tag{3.1}$$

If Y is distributed according to F, then $Z \sim U(0,1)$, where $U(0,1)$ denotes the uniform distribution on the interval $(0,1)$ [Ranjan, 2009]. Thus to assess calibration, one can check the uniformity of a histogram of the z values, called a PIT histogram. The PIT histogram will be hump shaped, U-shaped, triangular shaped and resembling a $U(0,1)$ distribution for respectively an overdispersed, underdispersed, biased or calibrated forecast [Ranjan, 2009].

Figure 3.1 a) shows a simulation example, where the true generating process is $\mathcal{N}(10,1)$ illustrated by the black pdf. The overdispersed forecast, $\mathcal{N}(10,2)$, has a larger variance than the true generating process, which can be seen when comparing the green and the black pdf. An underdispersed forecast has a lower spread than the observed process, which is illustrated in the simulation example by a $\mathcal{N}(10,0.5)$ (purple) pdf. A biased forecast has a mean that differs from the mean of the process in interest, in this example given by a $\mathcal{N}(11,1)$ distribution (blue pdf). Figure 3.1 b) shows the PIT histograms for the four different cases. When $F(\cdot)$ from equation (3.1) is the cdf of the $\mathcal{N}(10,1)$ distribution, the PIT-histogram is close to uniform as given in the upper left corner. For the overdispersed forecast, the PIT-histogram is hump shaped, because the prediction intervals are too wide, resulting in an overweight of z-values centred around 0.5 The underdispersed forecast has a too low variance compared to the observation generating process. This leads to a U-shaped PIT-histogram, since many observations are too extreme compared to the cdf based on the $\mathcal{N}(10,0.5)$ distribution. The last example is the biased forecast, where the mean is switched from 10 to 11. From the PIT-histogram in Figure 3.1 b), we observe that there is a trend towards lower values of $F(y)$, since $F(\cdot)$ based on the $\mathcal{N}(11,1)$ distribution is shifted higher than the true observation generating process.

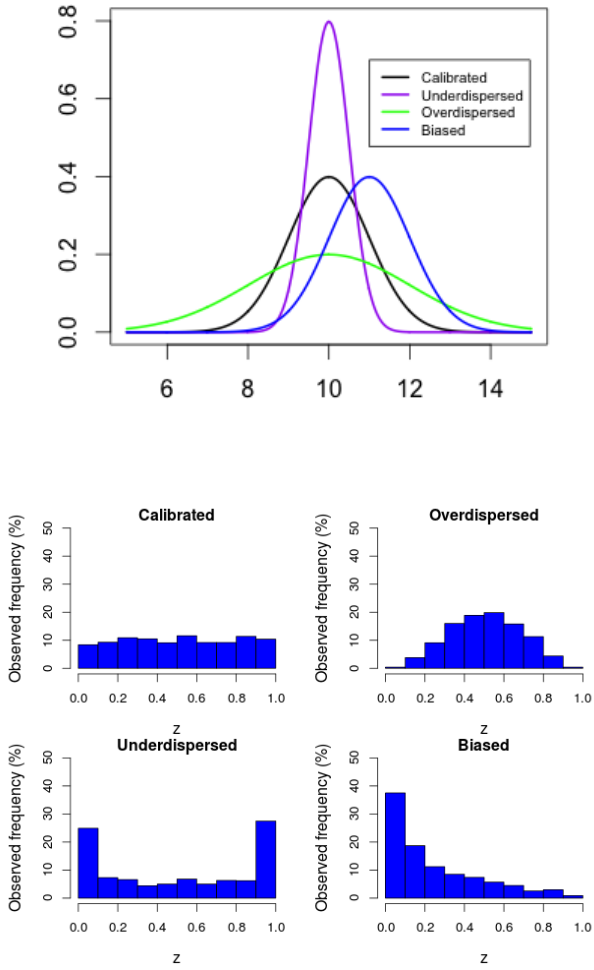


Figure 3.1: a) The predictive distributions used to illustrate PIT histograms. b) PIT histograms for a calibrated $\mathcal{N}(10, 1)$, an overdispersed $\mathcal{N}(10, 2)$, an underdispersed $\mathcal{N}(10, 0.5)$, and a biased $\mathcal{N}(11, 1)$ predictive distribution.

3.1.2 Assessing sharpness

Even though calibration is an important property for a good forecast, calibrated forecasts can be useless if they are not sharp. An example of this is the climatology, the forecast based on historical observations, which is calibrated but lack sharpness. Sharpness refers to the spread of the probabilistic forecast and is easily evaluated by comparing the average length of prediction intervals or boxplots of prediction intervals [Gneiting et al., 2007]. Short prediction intervals imply sharp forecasts.

3.1.3 Continous ranked probability score

In addition to the separate evaluation of calibration and sharpness, proper scoring rules are often used to assess both qualities simultaneously. A scoring rule is proper if the forecaster get the best score by predicting his true beliefs, although it is possible to achieve the same score by using another forecast [Gneiting and Raftery, 2007]. The continuous ranked probability score (CRPS) is a proper scoring rule that assess both calibration and sharpness in one measure [Gneiting et al., 2005]. It is negatively oriented, in that a smaller CRPS score indicates a better forecast. A perfect forecast results in a CRPS of 0. Let F be the cumulative distribution function of the probabilistic forecast and y_t the observation, in our case the observed streamflow at time t. The CRPS for each time point is defined as

$$CRPS_t(F, y_t) = \int_{-\infty}^{\infty} [F(\epsilon) - H(\epsilon - y_t)]^2 d\epsilon \tag{3.2}$$

where t is the issue time and H is the Heaviside function defined as

$$H(\epsilon - y_t) = \begin{cases} 0, & \epsilon < y_t \\ 1, & \epsilon \geq y_t \end{cases} \tag{3.3}$$

Figure 3.2 illustrates the $CRPS_t$ for a probabilistic forecast that is $\mathcal{N}(0, 1)$ when the verifying observation $y_t = 1$. Note that the CRPS does not equal the grey area, but the grey area visualize the value of $|F(\epsilon) - H(\epsilon - y_t)|$ for all y_t. Hence the $CRPS_t$ is the value of $F(\epsilon) - H(\epsilon - y_t)$ (indicated by the grey area) squared, integrated from $-\infty$ to ∞.

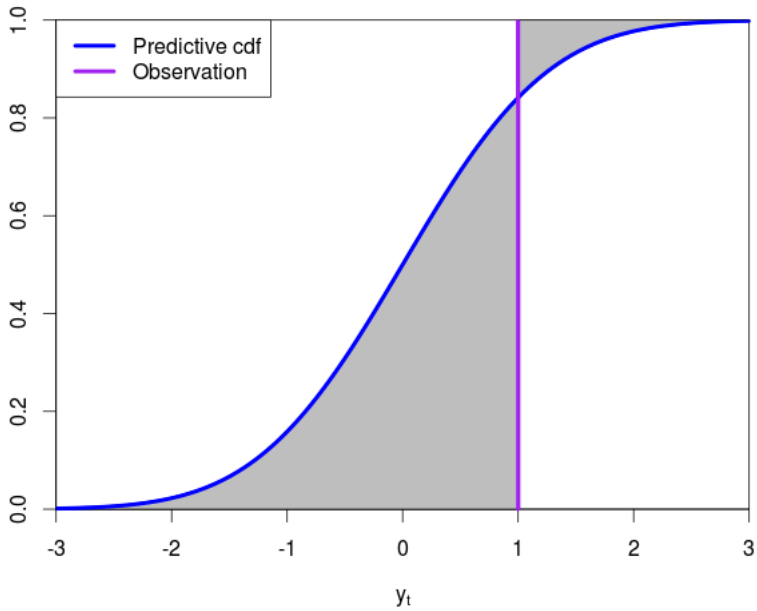


Figure 3.2: Illustration of $CRPS_t$ for a predictive cdf that is $\mathcal{N}(0,1)$ and the verifying observation of $y_t = 1$. The grey area represent the expression $|F(\epsilon) - H(\epsilon - y_t)|$, which squared and integrated from $-\infty$ to ∞ gives the CRPS.

To assess a probabilistic forecast, we average $CRPS_t$ over M forecasts, where M is the total number of observation-forecast pairs.

$$CRPS = \frac{1}{M} \sum_{t=1}^{M} CRPS_t \tag{3.4}$$

The mean absolute error (MAE)

$$MAE = \frac{1}{M} \sum_{t=1}^{M} |y_t - x_{t,l}| \tag{3.5}$$

is a scoring function commonly used to evaluate deterministic forecasts [Gneiting, 2011]. It measures the average difference between observations and forecasts. For deterministic forecasts, the CRPS reduces to the MAE, hence CRPS can be interpreted as a probabilistic generalization of the MAE [Jolliffe and Stephenson, 2012]. Both the CRPS and the MAE has the same unit as the forecasts and observations.

3.2 Beta distribution

A random variable $Y \in (0, 1)$ is beta distributed if its probability density function $f(y, \alpha, \beta)$ is

$$f(y; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1 - y)^{\beta-1} \tag{3.6}$$

where $\Gamma(\cdot)$ represents the gamma function and $\alpha, \beta > 0$ [Casella and Berger, 2002]. The expected value, μ and variance, σ^2 of the beta distribution are

$$\mu = \mathrm{E}[Y] = \frac{\alpha}{\alpha + \beta} \qquad \text{and} \qquad \sigma^2 = \mathrm{Var}[Y] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \tag{3.7}$$

In our post processing methodology it is convenient to re-parametrize the beta distribution by its mean μ and a variance parameter $\nu = \frac{1}{\alpha+\beta}$ [Kass and Raftery, 1995]. This leads to

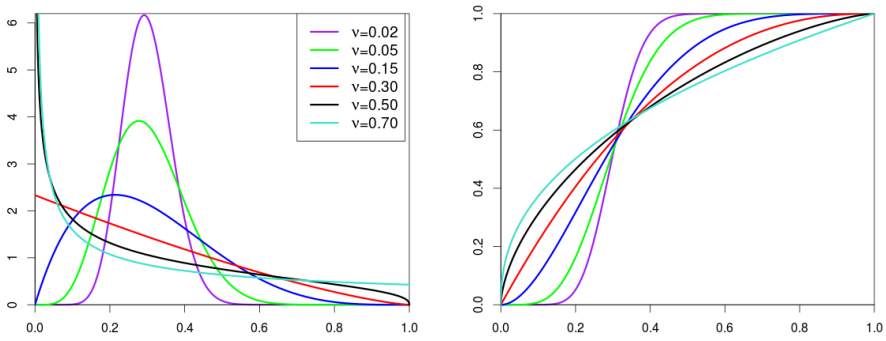
$$\alpha = \frac{\mu}{\nu} \qquad \beta = \frac{1 - \mu}{\nu} \tag{3.8}$$

$$\mathrm{E}[Y] = \mu \qquad \mathrm{Var}[Y] = \frac{\mu(1 - \mu)}{1 + \frac{1}{\nu}} \tag{3.9}$$

Figure 3.3 shows the beta distribution (pdf and cdf) with mean $\mu = 0.3$, for different values of the variance parameter ν. We observe that the variance increases with an increase in ν, as can also be seen from equation (3.9). Also it is clear that the shape can be quite different for small changes in ν, as we observe in the change from $\nu = 0.15$ to $\nu = 0.3$.

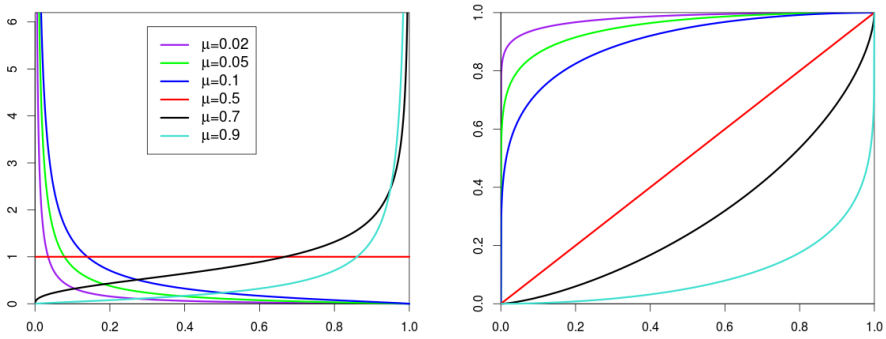
In Figure 3.4 we can observe how the beta distribution varies for different means μ, when the variance parameter $\nu = 0.5$. As a special case we observe that when $\mu = \nu = 0.5$ the beta distribution is the uniform distribution on the interval (0,1). For a given variance parameter ν, this choise of μ results in the

largest variance, as can be verified in Equation (3.9), where the variance increases symmetrically around $\mu = 0.5$.



(a) (b)

Figure 3.3: (a) Beta pdf and (b) Beta cdf for varying variance parameter ν when $\mu = 0.3$.



(a) (b)

Figure 3.4: (a)Beta pdf and (b)Beta cdf for varying mean μ when the variance parameter $\nu = 0.5$.

3.3 Beta-transformed linear pool

Our methodology for creating probabilistic forecasts, described in Chapter 4, is inspired by the work of Ranjan and Gneiting [2013], concerning beta-transformed linear pools. Linear pooling is a method that combines different probabilistic forecasts into a single combined forecast. The idea is to weight each forecast based on their respective performance over a training period. From now on we denote each of the c predictive densities by $f_i(y)$, $i = 1, \ldots, c$, and the corresponding predictive cumulative distribution functions $F_i(y)$.

The linear pool is defined as

$$f(y) = \sum_{i=1}^{c} \omega_i f_i(y) \tag{3.10}$$

where ω_i are weights that sum up to one and are determined based on the predictive performance of the forecast during a training period [Ranjan and Gneiting, 2013].

As mentioned in section 3.1, the purpose of aggregating probabilistic forecasts is to generate a probabilistic forecast that is as sharp as possible and calibrated. However Hora [2004] and Ranjan and Gneiting [2013] showed that if each of the individual probabilistic forecasts is calibrated, every nontrivial linear pool is uncalibrated. A trivial linear pool is the case where all weights are zero except one, which is 1, resulting in one of the original forecasts. These results suggest that linear pools may be suboptimal, and it has been shown empirically that non-linear combination formulas can outperform linear methods [Ranjan and Gneiting, 2010; Allard et al., 2012; Ranjan and Gneiting, 2013]. One of the non-linear combination formulas is the beta-transformed linear pool (BLP).

The BLP cumulative distribution function $F_{\alpha,\beta}(y)$ is defined as

$$F_{\alpha,\beta}(y) = B_{\alpha,\beta}\left(\sum_{i=1}^{c} \omega_i F_i(y)\right) \tag{3.11}$$

where $B_{\alpha,\beta}(\cdot)$ denotes the cumulative distribution function of the beta distribution with parameters $\alpha > 0$, $\beta > 0$, ω_i are non-negative weights summing to one and $F_i(y)$ for $i = 1, \ldots, c$ are the cumulative distribution functions for the different probabilistic forecasts [Ranjan and Gneiting, 2013].

3.4 Logit link

In the generalized linear model setting one aim is to fit the expected value of a random variable based on explanatory variables or covariates. Let the random variable be denoted Y_i and the actual outcome y_i for $i = 1, \ldots, n$. For a given set of p covariates x_{ij} the linear predictor η_i is defined as

$$\eta_i = \sum_{j=1}^{p} x_{ij}\beta_j \qquad i = 1, \ldots, n \tag{3.12}$$

A link function relates the linear predictor η_i to the expected value μ_i of Y_i [McCullagh and Nelder, 1989]. When $0 < \mu < 1$, as is the case for the beta distribution, a convenient link function is the logit given by

$$\eta_i = \text{logit}(\mu_i) = \log(\frac{\mu_i}{1-\mu_i}) \tag{3.13}$$

The advantage with the logit link is that it maps the interval $(0, 1)$ onto the whole real line [McCullagh and Nelder, 1989]. Since η_i can attain any value on the real line depending on x_{ij} and β_j while μ_i is constrained to the interval $(0, 1)$, it is essential to link η_i to the mean μ_i by the logit link. Solving equation (3.13) with respect to μ_i yields

$$\mu_i = \frac{1}{1 + e^{-\eta_i}} \tag{3.14}$$

From equation (3.14) we observe that μ_i is constrained on $(0, 1)$. Figure 3.5 illustrates the logit transformation between η and μ. We observe that as η increases, μ goes towards 1 and as η decreases μ goes towards 0.

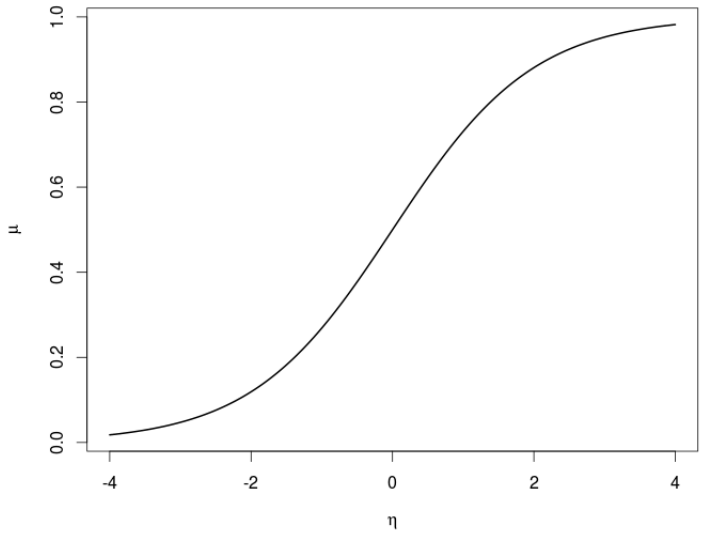


Figure 3.5: Plot of the logit link between η and μ.

4. Climatology cumulative probability regression

In this chapter we present our postprocessing methodology, which we refer to as the Climatology Cumulative Probability Regression (CCPR). As mentioned in the introduction, we are in the setting where we have available a climatology and different deterministic forecasts. Our aim is to use all the available information to obtain a probabilistic forecast. In Section 4.1, we introduce the ideas behind the CCPR. Section 4.2 defines the CCPR mathematically, and provides an illustration of the postprocessing methodology where we transform the climatology cdf, through a beta cdf, into a new probabilistic forecast. The construction of the climatology cdf, which is part of the method, is presented in Section 4.3. The last section describe how to model the mean and variance of the beta cdf used in the CCPR.

4.1 Overview of idea behind the CCPR method

The main objective of the CCPR methodology is to exploit the information present in the climatology combined with different deterministic forecasts. This section provides an overview of the ideas used to develop the CCPR methodology. We illustrate the ideas with a synthetic example, where we consider only the climatology and a hydrological forecast.

Let Y_t be the random variable that we want to make a probabilistic forecast for. The left plot in Figure 4.1 displays the probability distribution function (pdf) of Y_t based on the climatology, denoted $f_{clim}(y_t)$. We observe that most observations are close to zero, and that the probability of observing larger values than 10 is small. The right plot in Figure 4.1 shows the climatology cumulative distribution function (cdf) of Y_t, $F_{clim}(y_t)$. Recall that the cdf of a random variable Y_t describes the probability that Y_t is less than or equal to y_t. Hence

the range of a cdf is $[0, 1]$. We now introduce a scale of cumulative probabilities linked to the climatology cdf, which we denote the climatology cumulative probability scale (CCP-scale). The CCP-scale is a scale of cumulative probabilities, and hence it is ranging between 0 and 1. For each cumulative probability p on the CCP-scale, the associated value y_t is given by the inverse climatology cdf of p, namely $F^{-1}_{clim}(p) = y_t$. The y-axis in the plot of the climatology cdf in Figure 4.1, is the CCP-axis.

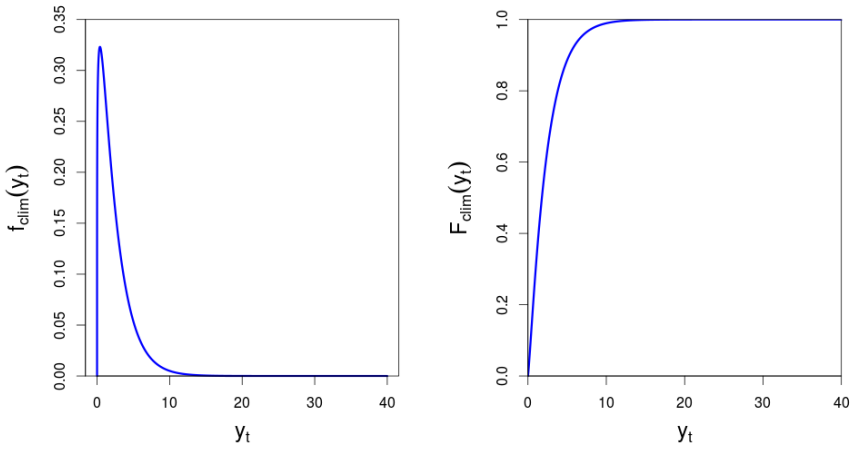


Figure 4.1: Pdf based on the climatology to the right and the corresponding cdf to the left.

Our approach for producing probabilistic forecasts is to fit a pdf on the CCP-scale based on different deterministic forecasts, and then transform the pdf back to the original scale. When considering pdf's on the CCP-scale, the interpretation is as follows: the pdf gives the probability that the observed value y_t falls in different intervals of the climatology cumulative distribution function, namely it is a pdf for $F_{clim}(y_t)$. As an example, imagine that we fit a uniform distribution ($U(0, 1)$), on the CCP-scale (Figure 4.2). This pdf implies that all cumulative probabilities in the climatology are mapped to their original values, i.e. $P(F_{clim}(y_t) \leq p) = p \; \forall \; p \; \in (0, 1)$. Hence, the result of fitting a uniform distribution on the CCP-scale, is that the postprocessed probabilistic forecast is the climatology.

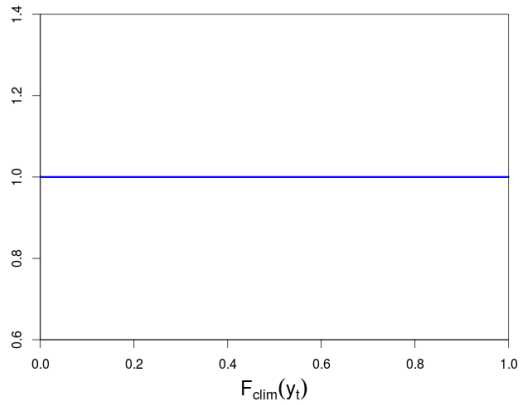


Figure 4.2: The climatology corresponds to the uniform distribution on the CCP-scale.

We proceed by introducing a hydrological forecast, $x_t^h = 5$. Figure 4.3 shows the pdf of the hydrological forecast on the CCP-scale. To be able to visualize the hydrological forecast, we choose a uniform density on the CCP-scale between $F_{clim}(4.7)$ and $F_{clim}(5.3)$. In the original scale this would result in a pdf with all density placed on the interval $y_t \in (4.7, 5.3)$, but not uniformly distributed because the climatology cdf is not linear.

To create a new probabilistic forecast, our idea is to fit a probability distribution function on the CCP-scale based on the hydrological forecasts, and then transform the resulting pdf back to the original scale. The left plot of Figure 4.4 shows an example of a pdf (purple pdf) on the CCP-scale based on the climatology (blue pdf) and the hydrological forecast pdf (red pdf). This pdf has less weight on the low values of $F_{clim}(y)$, indicating that the event of observing y_t where $F_{clim}(y_t)$ is small, has a low probability. The result of this is shown in the new pdf in original scale (purple pdf) in the right panel of Figure 4.4 where the probability for small values are much reduced compared to the climatology pdf (blue dotted pdf). Simultaneously, the density for larger values have increased compared to the climatology pdf, with a peak near the hydrological forecast. To summarize, the CCPR procedure begin with forecasts in the original scale, go through fitting a density on the CCP-scale and then transform the result back to the original scale.

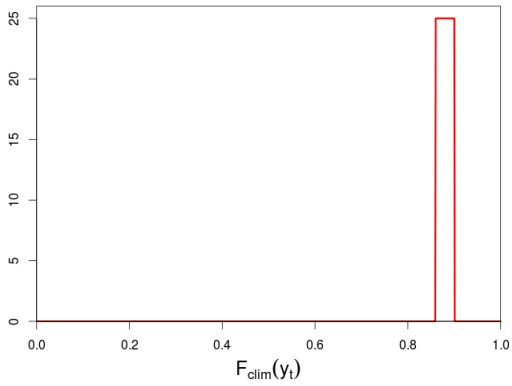


Figure 4.3: Pdf based on the hydrological forecast, $x_t^h = 5$, on the CCP-scale.

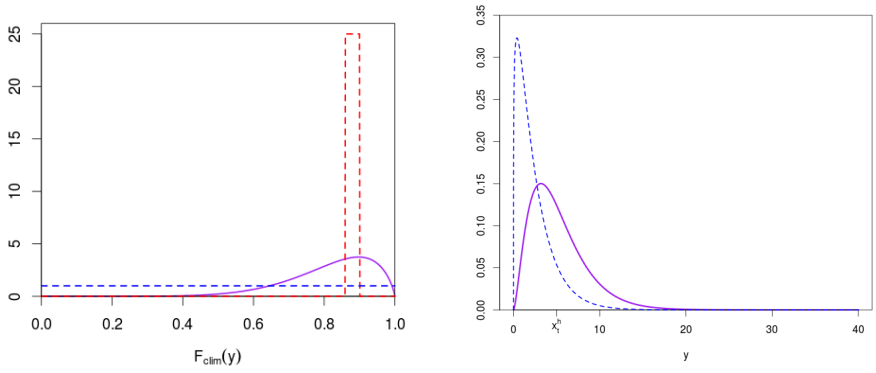


Figure 4.4: The left plot displays the fitted pdf (purple pdf) based on the hydrological pdf (dotted red) and the climatology pdf (dotted blue) on the CCP-scale. The right plot shows the postprocessed probabilistic forecast in the original scale (purple pdf), while the dashed blue pdf denotes the climatology pdf.

4.2 Climatology cumulative probability regression

As mentioned before, the CCPR method can be used in a setting where we have available the climatology and k different deterministic forecasts. We use the name climatology cumulative probability regression for our postprocessing methodology to emphasize that we use covariates (the deterministic forecasts) to fit a probability distribution function on the CCP-scale. Following the notation introduced in Section 2.3, we represent the forecasts with valid time t by a forecast vector $\mathbf{x_t}$. For simplicity, we suppress lead time from the notation. In this thesis we use a superscript with the abbreviations h, p and c respectively representing the hydrological, persistence and sliding window climatology forecast, to clarify which forecasts are present in the forecast vector. As an example $\mathbf{x_t^{h,p}}$ denotes the vector of hydrological and persistence forecast with valid time t, $\mathbf{x_t^{h,p}} = (x_t^h, x_t^p)$.

Further, we denote the climatology cdf by $F_{clim}(y_t)$. Section 4.3 present the construction of the climatology cdf based on the historical observations. As described in Section 4.1, the idea behind the CCPR-methodology is to fit a probability distribution on the CCP-scale, and transform the pdf back to the original scale. Since the CCP-scale has the range $(0, 1)$, we choose to fit a beta distribution on the CCP-scale, because it has the right range and, as we observed in Section 3.2, it can attain many different shapes including the uniform distribution. Hence our methodology include the climatology as a special case. This is an appealing property of the methodology, because if the forecasts are without predictive performance, the best forecast is the climatology since it at least is calibrated. Section 4.4 describe how to model the beta pdf based on the deterministic forecasts.

At a given time point t, the postprocessed probabilistic forecast for the validating observation y_t, given as a cdf, is denoted $F(y_t; \mathbf{x_t})$, emphasizing that the deterministic forecasts $\mathbf{x_t}$ determines the probabilistic forecast. We follow the notation used in the BLP-method described in Section 3.3, and defined in equation (3.11). In our framework we set $c = 1$ and let $F_1(y) = F_{clim}(y)$.

The CCPR postprocessing methodology can then be expressed as

$$F(y_t; \mathbf{x_t}) = B_{\mathbf{x_t}}(F_{clim}(y_t)) \tag{4.1}$$

where $\mathbf{x_t}$ denotes the vector of the explanatory variables with valid time t and

$B_{\mathbf{x_t}}(\cdot)$ denotes the beta cumulative distribution function determined by the vector of explanatory variables. From Equation (4.1) we note that the predictive cdf is obtained by transforming the empirical cdf $F_{clim}(y_t)$ by a beta cdf which is modelled by the deterministic forecasts $\mathbf{x_t}$. Notation wise it is easier working with cdf's in our methodology. However, the basic concept of the transformation procedure is based on fitting a pdf on the CCP-scale and then transforming the pdf back to the original scale.

An illustration of the procedure of transformation is given in Figure 4.5, which shows the transformation from the climatology cdf (pink cdf) to the postprocessed cdf forecast $F(y_t; \mathbf{x_t})$ (green cdf). The figure shows the transformation in detail for one point $(4, F_{clim}(4))$, while the procedure is the same for all points $(y_t, F_{clim}(y_t))$. The specific value of $\mathbf{x_t}$ at the given time point generates a beta cdf, $B_{\mathbf{x_t}}(\cdot)$ (grey cdf). Note that the beta cdf is a cdf on the CCP-scale, so the y-axis of the beta cdf is the left x-axis in Figure 4.5. To find the postprocessed cdf (green cdf), each point y_t is assigned the cdf value of $B_{\mathbf{x_t}}(F_{clim}(y_t))$. From Figure 4.5, we observe that before the transformation, the cdf value based on the climatology for $y = 4$ is approximately 0.8, indicating that the probability of observing a streamflow of 4m/s^3 or lower is close to 0.8. After the transformation however, the probability of observing a streamflow of at most 4m/s^3 is reduced to 0.2, $F(4; \mathbf{x_t^k}) = 0.2$. The reduction is due to the fact that the beta cdf has low values for cumulative probabilities up to around 0.8 on the CCP-scale. Note that this example is very extreme with a huge change between the climatology cdf and the postprocessed cdf, with large probabilities for observing high values of y_t.

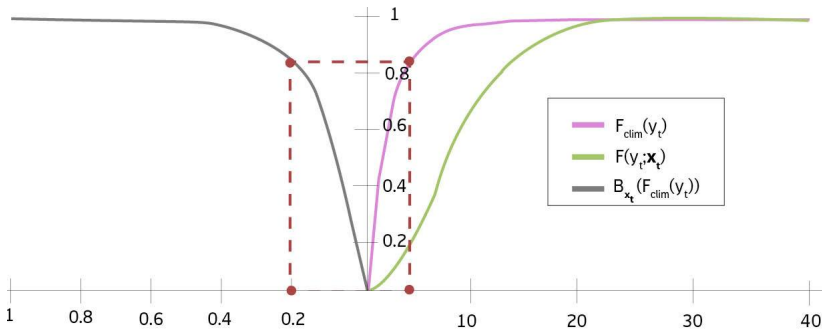


Figure 4.5: Illustration of the transformation procedure from the climatology cdf (pink cdf), through the beta cdf (grey cdf), resulting in the postprocessed probabilistic forecast $F(y; x_t)$ (green cdf).

4.3 Construction of climatology

The climatology cdf can be constructed in different ways. It is possible to fit a suitable distribution to the data, or extend our approach to make the climatology cdf a smooth curve. We have chose the climatology empirical cdf, $F_{clim}(y_t)$, to be a step function based on the historical observations, which jumps $\frac{i}{m}$ at observation values, where i is the number of observations at that value and m is the total number of historical observations. Figure 4.6 express the relationship between the historical observations and the climatology cdf. From the histogram of observations in Figure 4.6 a), we note that most of the observed streamflow is below $5\text{m}^3/\text{s}$. This trend is represented in the empirical cdf in Figure 4.6 b) which has a steep slope for low stream flows.

4.4 Modelling the mean parameter μ and the variance parameter ν of the beta cdf

To parametrize the beta distribution used in the transformation procedure, we use the re-parametrization by μ and ν as described in section 3.2. The advantage of using μ and ν is that these parameters has an easier interpretation than α and β which in turn ease the modelling. Because $\mu \in (0, 1)$, we model the mean by a logit link, given in section 3.4. The general method for modelling the mean is given by

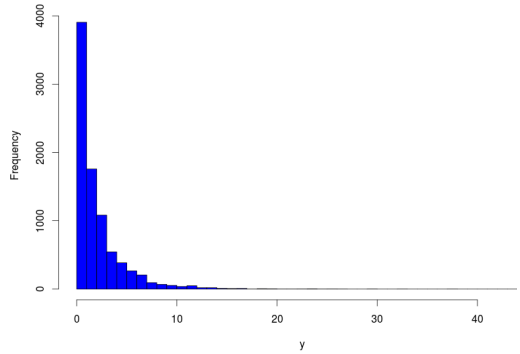
$$\eta_t = h(\mathbf{x_t}) \tag{4.2}$$

where $h(\cdot)$ is a function of the forecast vector $\mathbf{x_t}$, and hence depend on the value of the forecasts included. Equation (4.3) defines the logit link between η_t and μ_t.

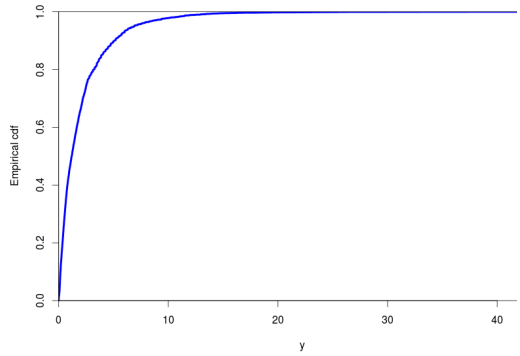
$$\mu_t = \frac{1}{1 + e^{-\eta_t}} \tag{4.3}$$

The variance parameter ν_t can also be modelled as a function of the deterministic forecasts, Equation (4.4). Since $\nu_t > 0$ it is important that $g(\mathbf{x_t}) > 0 \; \forall \; \mathbf{x_t}$. This restriction can be complied by different choices of $g(\cdot)$. In Chapter 5, we present the choices of $h(\cdot)$ and $g(\cdot)$ made in our case study.

$$\nu_t = g(\mathbf{x_t}) \tag{4.4}$$



(a)



(b)

Figure 4.6: a) Histogram of historical observations b) Climatology cdf based on historical observations.

5. Case study: Models and evaluation scheme

In this chapter we present our choice of $h(\cdot)$ and $g(\cdot)$, from Equation (4.2) and (4.4) for the case study. Equation (5.1) define our choice of $g(\cdot)$ for all models. This imply that the variance parameter does not vary based on the value of the forecast at a given time point t, hence we suppress valid time from the notation. However, we allow for the variance parameter ν to vary with lead time. For every model we fit the variance parameter ν for lead time $l = 1, 2, 5$ and 10, resulting in four different γ_ν for each model. Since γ_ν is squared, ν is always positive.

$$\nu_t = \nu = g(x_t) = \gamma_\nu^2 \tag{5.1}$$

We now introduce three different choices for modelling the mean, resulting in three models; Model 1, Model 2 and Model 3. For each model, we only present the choice of linear predictor η_t, since the mean μ_t is obtained by the logit link as given in Equation (4.3).

5.1 Model 1 - Hydrological forecast in beta-mean

The first model use the information contained in the hydrological forecast to model the mean of the beta cdf in Equation (4.1). To model η_t we use $F_{clim}(x_t^h)$ as a covariate, as seen in Equation (5.2). We choose this approach to distinguish the transformation done according to the differences in $F_{clim}(x_t^h)$ rather than in x_t^h. In this setting, the difference in the mean of the beta cdf between two days with hydrological forecast x_1^h and x_2^h will depend on how far apart the values are on the CCP-scale. As an example, consider the empirical cdf in Figure 4.6. A difference in $10\text{m}^3/\text{s}$ yields quite different values on the CCP-scale, depending on

what the forecasts are. If $x_1^h = 20$ and $x_2^h = 30$, the mean in Equation (4.3) will be more similar than if $x_1^h = 1$ and $x_2^h = 11$. Note that the mean model also distinguish between lead times, which means we fit 2 parameters to model the mean for each lead time in Model 1.

$$\eta_t = h(x_t^h) = \gamma_0 + \gamma_h F_{clim}(x_t^h) \tag{5.2}$$

The interpretation of the influence of the parameters in Equation (5.2) on the mean μ_t (Equation 4.3) can be a bit cumbersome, since the logit link is involved. First we note that γ_0 determines the mean μ_t when $F_{clim}(x_t^h) = 0$, namely $\mu_t = \frac{1}{1+e^{-\gamma_0}}$. Figure 5.1 illustrates this point in that the intersection with the y-axis is determined by γ_0. When comparing the graphs with identical values of γ_h but different values for γ_0, e.g the blue and orange graph, we observe that γ_0 also has an effect on the range of μ_t. In addition γ_h has a clear impact on the shape of the function of μ_t, as seen when comparing the orange and blue graph ($\gamma_h = 3$) or the black, purple and green graph ($\gamma_h = 1$). A poor forecast would lead to a small value of γ_h indicating that the beta distribution is quite similar for all forecasts x_t^h and resulting in similar probabilistic forecasts, i.e the climatology as forecast. If the hydrological forecast x_t^h has a good performance, one would expect a negative γ_0 and a large value of γ_h, leading to a vast variety in μ_t for different values of $F_{clim}(x_t^h)$

5.2 Model 2 - Persistence and hydrological forecast in beta-mean

Model 2 is an extension of Model 1, where the persistence forecast x_t^p is included in modelling the mean (Equation 5.3). As for Model 1, we observe that η_t and hence also μ_t varies with the value of the forecasts at time t. For each lead time, three parameters are fitted; γ_0, γ_h and γ_p. The interpretation of the parameters γ_0, γ_h and γ_p is the same as for Model 1. However it is worth noting that the effect of γ_h in Model 1 will be represented by both γ_h and γ_p in Model 2. The magnitude of γ_h and γ_p indicate predictive performance of the forecasts, with larger values for better performance.

$$\eta_t = h(\mathbf{x_t^{h,p}}) = \gamma_0 + \gamma_h F_{clim}(x_t^h) + \gamma_p F_{clim}(x_t^p) \tag{5.3}$$

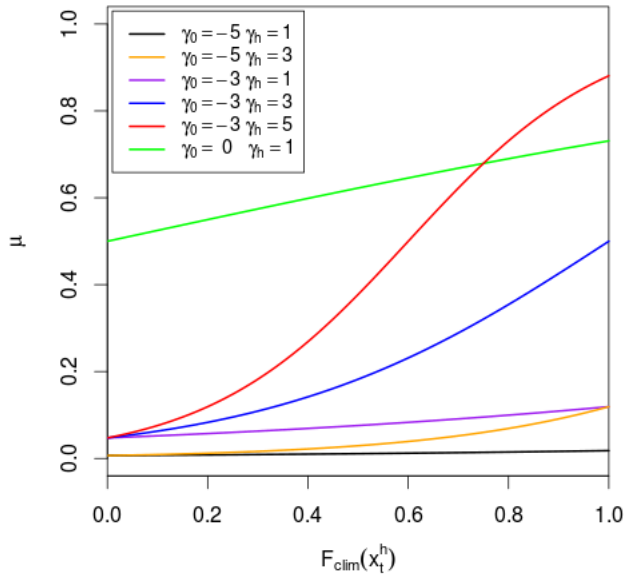


Figure 5.1: Mean parameter μ_t of the beta distribution as a function of $F_{clim}(x_t^h)$ for different values of γ_0 and γ_h

5.3 Model 3 - Sliding window climatology, persistence and hydrological forecast in beta-mean

Model 3 is a further extension of Model 2, where the sliding window climatology forecast is included in modelling the mean. The linear predictor η_t is modelled by

$$\eta_t = h(\mathbf{x_t^{h,p,c}}) = \gamma_0 + \gamma_h F_{clim}(x_t^h) + \gamma_p F_{clim}(x_t^p) + \gamma_c F_{clim}(x_t^c) \tag{5.4}$$

where x_t^c is the sliding window climatology given in Section 2.2.3, with validating observation y_t.

5.4 Model evaluation scheme

To evaluate the predictive performance of our models we use cross validation and divide the data into four periods as given in Table 5.1. When estimating the coefficients γ_k, we successively omit periods one by one and use the resulting parameters to make predictions for the period omitted. This results in a fourfold cross-validation scheme. The parameters are fitted using minimum CRPS estimation, which is fitting the parameters that yields the lowest CRPS [Gneiting et al., 2005]. This is done by using a optimizing routine in R. The procedure demands initial values to find an optimum. To validate stability of the optimization, we performed the analysis with different initial conditions, and the results were very similar in all cases. For each period in the cross validation scheme, we found the predictive CRPS based on the cross validation scheme and the CRPS reported in Chapter 6 is the average of the four predictive CRPS values. The same method is used to find average prediction intervals.

Period number	**Date**
Period 1	01.09.2005 - 31.08.2006
Period 2	01.09.2006 - 31.08.2007
Period 3	01.09.2007 - 31.08.2008
Period 4	01.09.2008 - 31.07.2009

Table 5.1: Periods used in the cross validation

6. Results from the Osali case study

This chapter present the results from the case study, described in Chapter 2 where we fitted the models given in Chapter 5. We first present the coefficients in the different models and consider the effects on the predictive densities in Section 6.1. Section 6.2 and 6.3 evaluate calibration and sharpness for the three models. All results are obtained by using the programme R.

6.1 Coefficients and probabilistic forecasts

This section present the coefficients that determines the beta transformation for the different models and consider some of the resulting probabilistic forecasts. First we present the coefficients present in all models, namely γ_0 and γ_ν. Then we proceed by exploring the results of the three models separately.

6.1.1 The constant parameter γ_0 and the variance parameter ν

Table 6.1 displays the estimated value of γ_0 for lead time $l = 1, 2, 5$ and 10 for Model 1, 2 and 3. All models have increasing values of γ_0 by increasing lead time l. This both indicate that the mean μ_t of the beta distribution has a higher starting point as a function of $F_{clim}(x_t)$ and that the size of the interval of possible values of μ_t decrease with increasing lead time (Figure 5.1). However the other coefficients of the models also contributes to the resulting model of μ_t.

γ_0	$l = 1$	$l = 2$	$l = 5$	$l = 10$
Model 1	-3.06	-2.91	-2.57	-1.50
Model 2	-2.98	-2.89	-2.63	-1.70
Model 3	-3.15	-3.12	-2.73	-1.87

Table 6.1: γ_0 for the different models and lead time $l = 1, 2, 5, 10$.

To investigate how the variance parameter varied for different lead times and between the models, a plot of ν for lead times $l = 1, 2, 5, 10$ for Model 1,2 and 3 was made (Figure 6.1). First we note that the variance parameter ν increases with increasing lead time in all models. This indicates that the probabilistic forecasts have a larger variance for longer lead times, as expected. Secondly we see that ν is always lowest for Model 3, indicating that for equal values of μ_t Model 3 yields the sharpest forecast. However, the difference in ν is small between Model 2 and 3 for all lead times, and between all models except for lead time 1 where Model 1 has a much higher value of ν than Model 2 and 3.

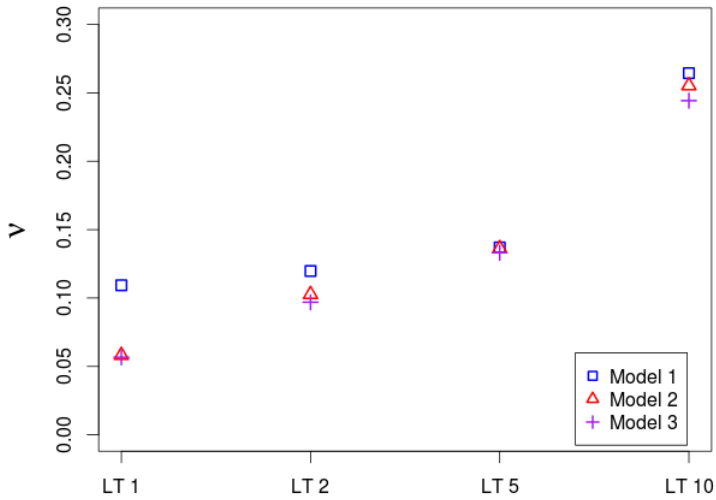


Figure 6.1: Estimated variance parameter ν in Model 1, 2 and 3, for the lead times $l = 1, 2, 5$ and 10.

Figure 6.2 illustrates the effect of the variance parameter on the beta pdf on the CCP-scale. The plot shows that for lead time 1 (black pdf's), there are clear

differences in the beta pdf for Model 1 (solid line) compared to Model 2 and 3 (dotted and dashed lines), with Model 2 and 3 as sharper. We also observe that the beta pdf's have an increasing variance with increasing lead time. Lead time 10 represents the most prominent difference in the shape of the beta pdf compared to the other lead times cf. the large difference in ν for lead time 10 compared to ν for the other lead times (Figure 6.1). For lead time 5 (red pdf's) and 10 (blue pdf's), there are almost indistinguishable differences in the beta pdf's between the different models, while for lead time 2 (purple pdf's), Model 1 (solid line) differs from Model 2 and 3 (dotted and dashed lines). An important aspect of this finding is that it is not only the absolute value of the differences in ν that has an influence on the resulting beta pdf's. In Figure 6.1, we observe that the difference between ν for lead time 2 and ν for lead time 10 for Model 1 and Model 3 are quite similar. However, the resulting beta pdf's differ more for lead time 2 than for lead time 10. This indicates that it is the relative difference in ν that results in varying beta pdf's, which in turn results in different probabilistic forecasts.

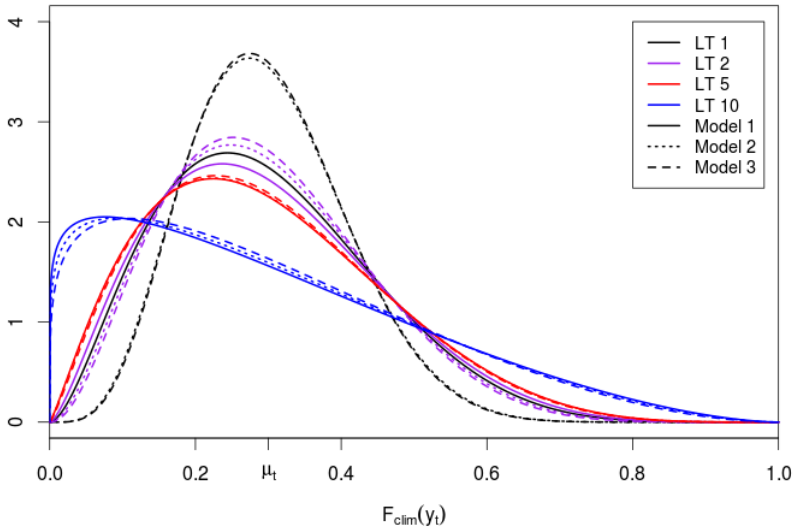


Figure 6.2: The effect of the variance parameters on the beta pdf on the CCP-scale, where the mean parameter $\mu_t = 0.3$. Note that the pdf for Model 2 (red dotted pdf) for lead time 5 follows the pdf for Model 1 (red, solid pdf), and hence is difficult to observe.

6.1.2 Model 1 - hydrological forecast

Figure 6.3 displays the estimated coefficient γ_h in Model 1 for lead time $l = 1, 2, 5$ and 10. We observe that γ_h decrease for longer lead times, indicating a decaying predictive ability of the hydrological forecast.

The effect of γ_h on the mean parameter μ_t of the beta distribution, is given in Figure 6.4 where μ_t is given as a function of $F_{clim}(x_t^h)$ for lead time $l = 1, 2, 5$ and 10. First we note that there are only minor differences between lead time $1, 2$ and 5, which all model the mean μ_t in the range from about 0.1 to 0.9. This indicates that the hydrological forecast has a high predictive ability for these lead times. For lead time 10 however, μ_t does not change as much as for the other lead times, indicating a poorer predictive performance of the hydrological forecast for lead time 10 compared to the other lead times.

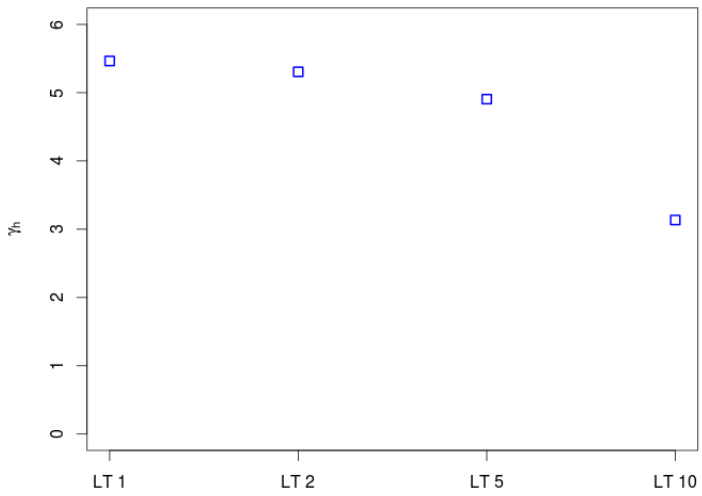


Figure 6.3: Estimated γ_h in Model 1 for the lead times $l = 1, 2, 5, 10$.

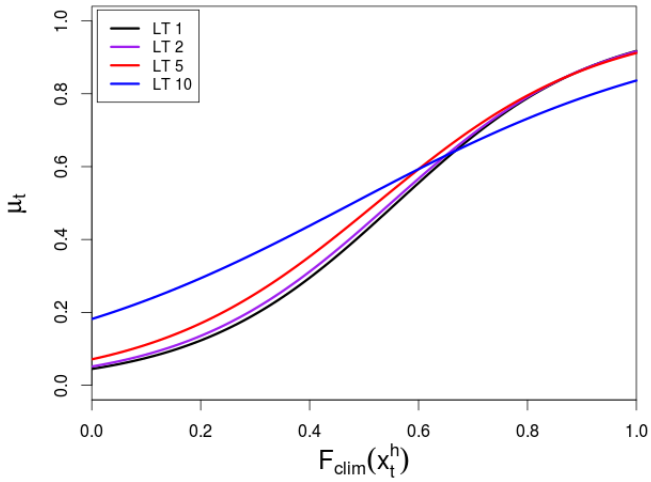


Figure 6.4: Mean of beta distribution, μ_t as a function of the cumulative distribution function of the hydrological forecast, $F_{clim}(x_t^h)$.

Figure 6.6 displays beta pdf's on the CCP-scale for Model 1 for different hydrological forecasts. In all cases the beta pdf for lead time 10 stands out among the others. It is worth noting that this effect does not only come from the difference in the mean value, but also from the large difference in the variance parameter ν for lead time 10 compared to the other lead times (Figure 6.2). In Figure 6.6 b), we observe that the spread is quite high. A reason for this is that the mean parameter contributes to the variance, as seen in Equation (3.9). For a given value of the variance parameter ν, the variance grows symmetrically with $\mu_t = 0.5$ as the maximum. Hence when μ_t is close to 0.5 the variance of the beta pdf is larger than when μ_t is close to zero or one.

Examples of the resulting probabilistic forecasts are given in Figure 6.7 and 6.8, for respectively lead time 1 and lead time 10. The probabilistic forecasts are illustrated by histograms, where we sample $n = 10000$ observations from the predictive cdf, $F(y_t; \mathbf{x_t^h})$. Since the climatology cdf is a step function, so is the predictive cdf and this is also reflected in the histograms. As an example, in figure 6.8 for $x_t^h = 5.1$ there are three small bins for $y \in (30, 45)$, while the weight ideally should be placed more uniformly on the interval. However, the histograms work well for illustrative purposes.

For both lead times, we observe that a lower forecast yields a sharper predictive density. At first glance this may seem contradictory to the results presented in the previous section, where the forecast of $x_t^h = 1.2$ produced the least sharp beta pdf on the CCP-scale. However, the probabilistic forecast is also largely based on the shape of the climatology cdf (Figure 4.6b). To illustrate the concept, we plot the central 95% prediction intervals on CCP-scale and original scale (Figure 6.5). On the CCP-scale, the prediction interval for the forecast $x_t^h = 1.2$ (purple points) is wider than for $x_t^h = 5.1$ (red points). However, when transformed to the original scale, the prediction interval for the low forecast (purple points) is notably shorter than for the high forecast (red points). The reason for this is the transformation procedure through the climatology cdf, which is much steeper for low CCP-values. Hence the pdf's on the CCP-scale only provide some information concerning the final probabilistic forecast.

When comparing the probabilistic forecast for lead time 1 (Figure 6.7) with lead time 10 (Figure 6.8) we see that the forecast for lead time 1 is much sharper (Note that the axis differs in all figures). This is due to the differences in beta pdf's on the CCP-scale (Figure 6.6) where lead time 10 always have the largest variance.

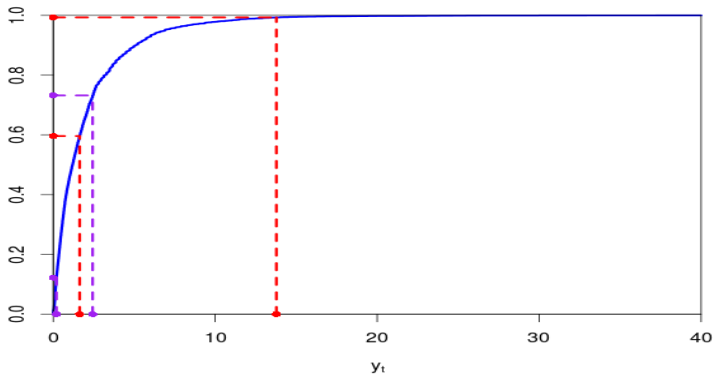
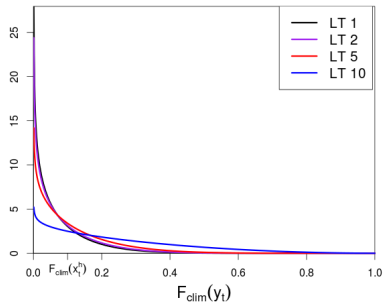
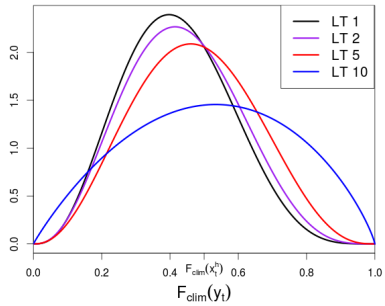


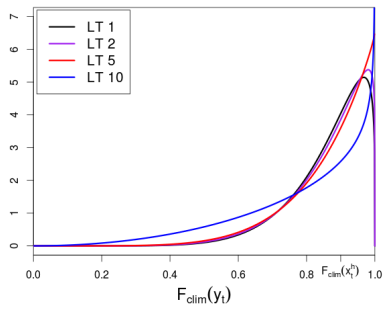
Figure 6.5: Central 95% prediction interval on the CCP-scale and in the original scale for a forecast of $x_t^h = 1.2$ (purple points) and $x_t^h = 5.1$ (red points), for lead time 1. The dashed lines demonstrates the transformation from the CCP-scale, through the climatology cdf, to the original scale.



(a)



(b)



(c)

Figure 6.6: Beta pdf's on the CCP-scale for a) A hydrological forecast of $x_t^h = 0.17$ ($F_{clim}(x_t^h) = 0.1$), b) A hydrological forecast of $x_t^h = 1.2$ ($F_{clim}(x_t^h) = 0.5$) and c) A hydrological forecast of $x_t^h = 5.1$ ($F_{clim}(x_t^h) = 0.9$)

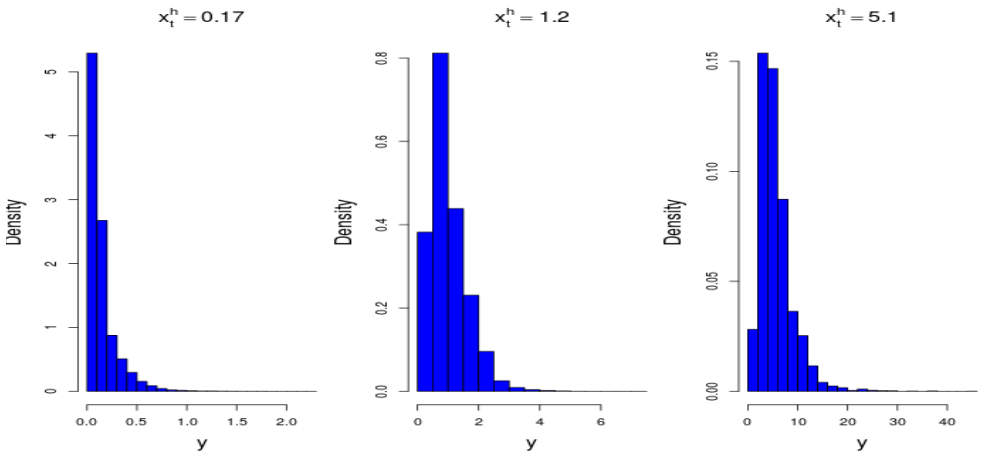


Figure 6.7: Post processed probabilistic forecast of y_t in Model 1 (Lead time 1) illustrated by histograms, for different hydrological forecasts $x_t^h = 0.1, x_t^h = 1.2$ and $x_t^h = 5.1$.

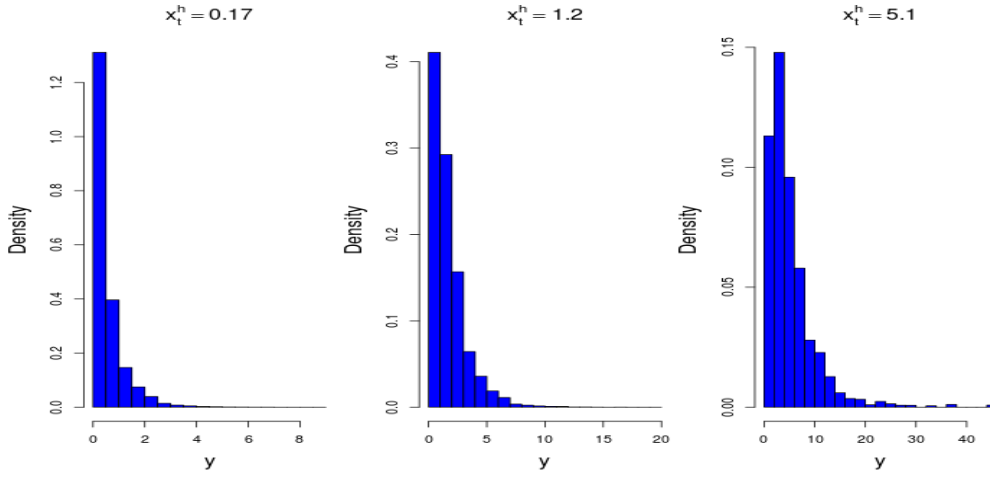


Figure 6.8: Post processed probabilistic forecast of y_t in Model 1 (Lead time 10) illustrated by histograms, for different hydrological forecasts $x_t^h = 0.1, x_t^h = 1.2$ and $x_t^h = 5.1$.

6.1.3 Model 2 - hydrological and persistence forecast

Figure 6.9 present the estimated coefficients for Model 2, where both the hydrological and persistence forecast are included in modelling the mean. First we note that γ_h is larger than γ_p for all lead times except $l = 1$. For lead time $l = 2, 5, 10$ the hydrological forecast dominates the mean of the beta pdf compared to the persistence forecast, while the opposite is true for lead time 1. Next we note that γ_p decrease with increasing lead time, except for a small increase between lead time 5 and 10, indicating that the value of the persistence forecast decrease with increasing lead time. This result is consistent with the results obtained by Engeland and Steinsland [2014], who showed that the persistence forecast contained much predictive information for shorter lead times. Figure 6.10 illustrate these trends exemplified by lead time 1 and 2, where the mean of the beta pdf μ_t is plotted as a function of $F_{clim}(x_t^p)$ for different values of the hydrological forecast. Comparing lead time 1 and 2, respectively solid and dotted lines, we observe that the mean varies considerably more with changing persistence forecast for lead time 1 than 2. However the value of γ_h puts restrictions on the range of μ_t also for lead time 1. This can be seen e.g when comparing the solid black and blue curve, where the maximum value of μ_t are respectively around 0.8 ($F_{clim}(x_t^h) = 0.9$) and 0.6 ($F_{clim}(x_t^h) = 0.1$).

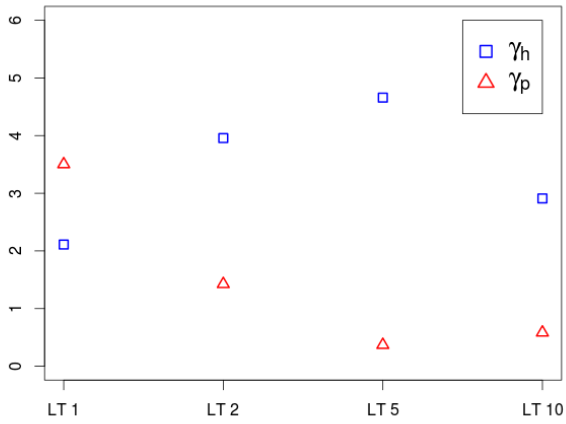


Figure 6.9: Estimated γ_h and γ_p in Model 2 for the lead times $l = 1, 2, 5, 10$.

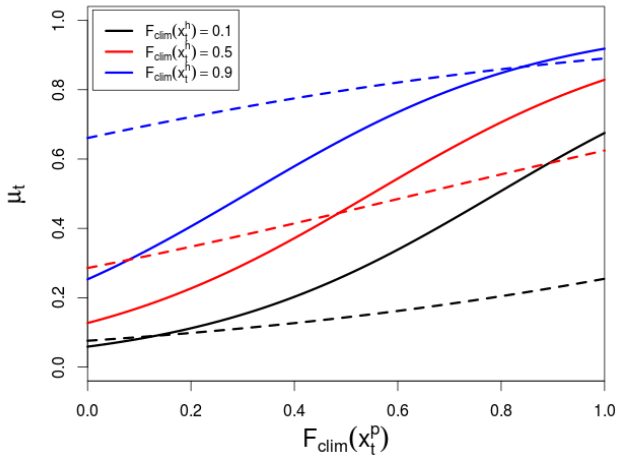


Figure 6.10: Mean of beta distribution, μ_t as a function of the cumulative distribution function of the persistence forecast, $F_{clim}(x_t^p)$, for different values of the hydrological forecast $F_{clim}(x_t^h)$. The solid lines represent lead time 1, while the dashed lines represent lead time 2.

In order to illustrate the parameters influence on the beta pdf, Figure 6.11 shows beta pdf's on the CCP-scale for lead time $1, 2, 5$ and 10 for different combinations of hydrological and persistence forecasts. The plot is organized as a matrix, where the row number determines the value of $F_{clim}(x_t^h)$ and the column number determines the value of $F_{clim}(x_t^p)$. As an example, the plot in the bottom-right corner displays the beta pdf's on the CCP scale for the given lead times when $F_{clim}(x_t^h) = 0.9$ and $F_{clim}(x_t^p) = 0.9$.

First we consider lead time 1, displayed as the black pdf. We observe that the pdf's differ more through each row than through each column. Hence the persistence forecast has a larger impact on the mean than the hydrological forecast for lead time 1. As an example, we compare column 1 where $F_{clim}(x_t^p) = 0.1$ with row 1 where $F_{clim}(x_t^h) = 0.1$. In column 1 the mean is always below 0.4 even though $F_{clim}(x_t^h)$ varies from 0.1 to 0.9. In the first row however, the mean increases more with an increase in $F_{clim}(x_t^p)$.

For lead time 2 (purple pdf), we observe an opposite trend of that for lead time 1. Now the pdf's in a given row are more equal, while the pdf's in each column differs more. This illustrate the effect of the change in γ_h and γ_p from lead time

1 to 2, where γ_h increases and γ_p decreases (Figure 6.9).

The trend where the hydrological forecast has the strongest effect on the mean of the beta pdf is enhanced from lead time 2 (purple pdf) to lead time 5 (red pdf). For lead time 5, the persistence forecast has almost no effect on the mean of the beta pdf, illustrated by very similar pdf's in each row. The beta pdf's in each column however, are very different with mean close to $F_{clim}(x_t^h)$.

The beta pdf's of lead time 10 (blue pdf) has a larger variance than the pdf's for the shorter lead times. We observe that also here, the hydrological forecast has a slight impact on the mean of the beta pdf's in that within a row the pdf's are very similar, while within a column the pdf's change more. However, the large variance causes all pdf's to be quite similar.

Figure 6.12 displays the post processed probabilistic forecast based on the beta pdf's in the bottom-left corner of Figure 6.11. It illustrates the relationship between the beta pdf's and the post processed probabilistic forecast for the lead times $l = 1, 2, 5$ and 10, when the hydrological forecast $x_t^h = 0.1$ ($F_{clim}(x_t^h) = 0.1$) and the persistence forecast $x_t^p = 5.1$ ($F_{clim}(x_t^h) = 0.9$). First we observe that for lead time 1, the probabilistic forecast has the lowest mean, and the mean seems to be closer to the persistence forecast than to the hydrological forecast. Hence the information obtained by the beta pdf in Figure 6.11, is transferred to the post processed probabilistic forecast. Next we observe that the probabilistic forecast for lead time 5 and 10 are quite similar, apart from the heavier right tail for lead time 10. This is also apparent in the difference between the beta pdf's in Figure 6.11, where the beta pdf for lead time 10 (blue pdf) has a heavier right tail than the beta pdf for lead time 5 (red pdf).

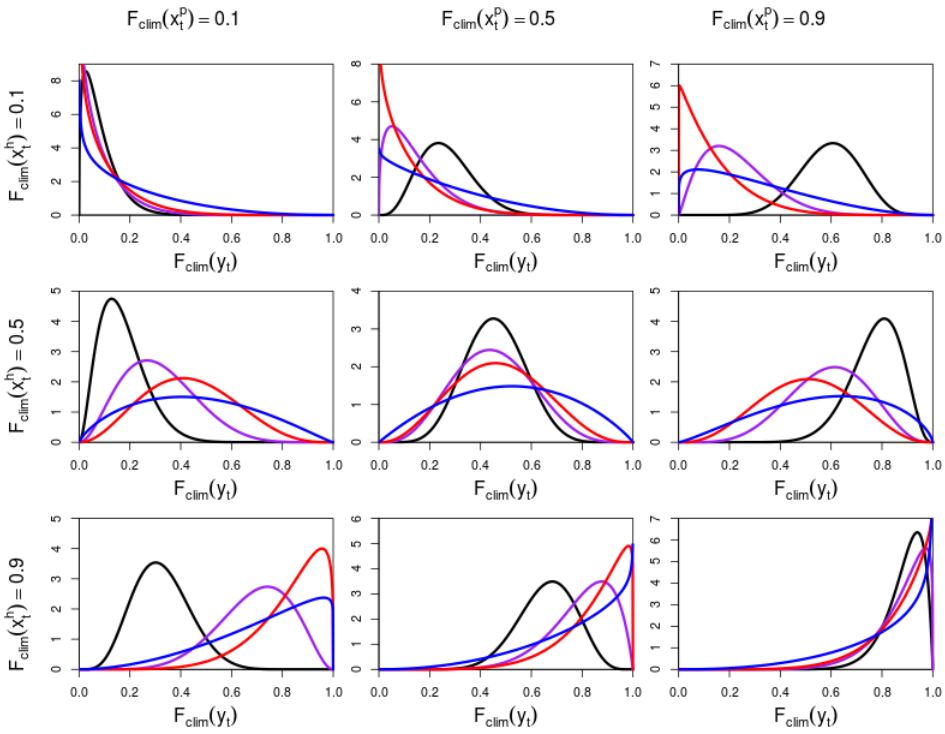


Figure 6.11: Beta pdf's on the CCP-scale for different combinations of the hydrological forecast x_t^h and the persistence forecast x_t^p and for different lead times $l = 1$ (black pdf), $l = 2$ (purple pdf), $l = 5$ (red pdf) and $l = 10$ (blue pdf).

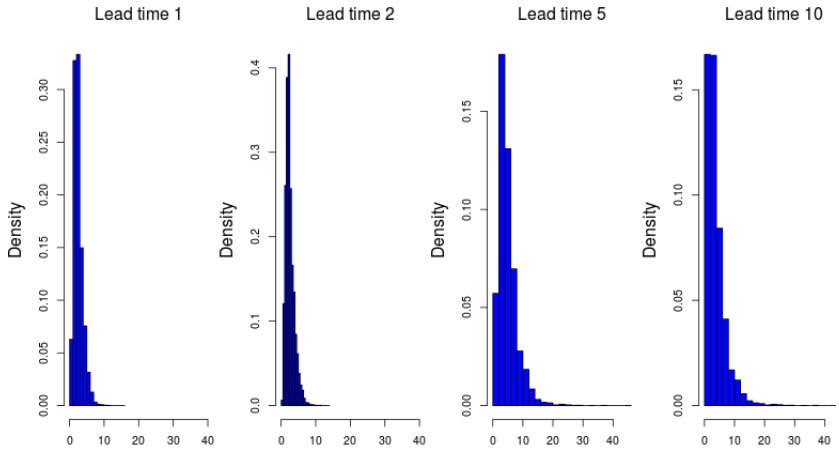


Figure 6.12: Post processed probabilistic forecast of y_t in Model 2 illustrated by histograms, for different lead times when the hydrological forecast $x_t^h = 0.17$ ($F_{clim}(x_t^h) = 0.1$) and the persistence forecast $x_t^p = 5.1$ ($F_{clim}(x_t^p) = 0.9$).

6.1.4 Model 3 - hydrological, persistence and sliding window climatology forecast

Figure 6.13 displays the coefficients for Model 3, where the sliding window climatology forecast is added when modelling the mean compared to Model 2. The value of γ_h and γ_p are similar to the values of the coefficients in Model 2 (Figure 6.9). We observe that the sliding window climatology, represented by γ_c is larger than γ_p for lead time 5 and 10. This indicates that the sliding window climatology forecast is more important in modelling the mean for higher lead times, while the persistence forecast has a higher impact when modelling the mean for the shorter lead times. However we observe that both coefficients are much smaller than γ_h for all lead times, except γ_p for lead time 1, suggesting that the hydrological forecast has the largest impact in modelling the mean of the beta pdf for lead time $2, 5$ and 10. For lead time 1 however, the persistence forecast has the largest impact in modelling the mean.

Since the coefficients are quite similar for Model 2 and 3, and because the sliding window climatology forecast has such a small impact in modelling the mean, we do not interpret the parameters of this model as thoroughly as we did for Model 1 and 2. However, to understand the results concerning average width of prediction intervals presented in Section 6.3 for lead time 10, we provide a plot of the mean μ_t for Model 3 in Figure 6.14. The figure displays μ_t as a function

of the climatology cumulative probability of the persistence forecast, when the hydrological forecast $F_{clim}(x_t^h) = 0.9$, for different values of the sliding window climatology forecast for lead time 10. The mean for Model 2 as a function of the persistence forecast $F_{clim}(x_t^p)$ is also given (green graph). We observe that the mean for Model 2, which does not include the sliding window climatology, varies less than the mean in Model 3 has ability to do, due to changes in the sliding window climatology forecast. This can be seen in Figure 6.14, where the mean for Model 2, lie almost in the middle of the mean for different values of the sliding window climatology in Model 3. This causes Model 3 to be more flexible to get different values of the mean according to the combination of persistence forecast and sliding window climatology forecast. For low values of both the persistence and the sliding window climatology forecast, the mean in Model 3 will be less than the mean in Model 2. The opposite is true for large values of the persistence and sliding window climatology forecast.

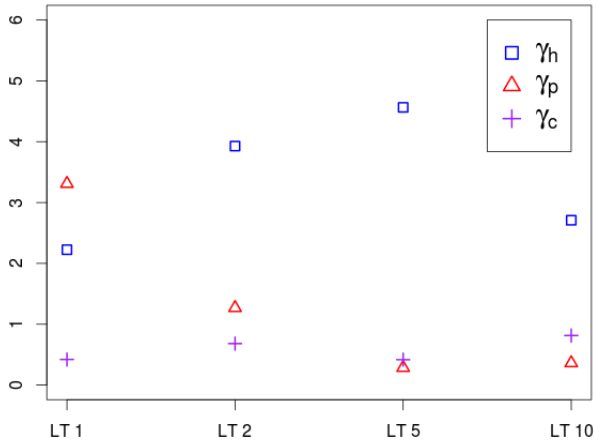


Figure 6.13: Estimated γ_h, γ_p and γ_c in Model 3 for the lead times $l = 1, 2, 5, 10$.

6.2 Assessing calibration

PIT-histograms provides a tool for assessing calibration [Gneiting et al., 2007]. Figure 6.15 displays PIT-histograms for lead time $1, 2, 5$ and 10 for Model 1, 2 and 3. In Figure 6.15 a) we observe that the PIT-histograms are close to uniform indicating a well calibrated predictive density for Model 1. Figure 6.15 b) and

c) are very similar and we observe a slightly hump-shaped PIT-histogram for lead times $l = 1$ and $l = 2$, indicating that these predictive densities are a bit overdispersed.

To investigate the calibration further, we divided the hydrological forecast into three categories; the 20% lowest forecasts (low flow), the 10% highest forecasts (high flow) and the remaining forecasts (medium flow) for each lead time. We then found the percentage of observations in different p% prediction intervals. Figure 6.16 displays the percentage of observations in different prediction intervals for all models and for lead time $l = 1, 2, 5, 10$, when the hydrological forecast is low. We observe that all models shows similar results within each lead time. Lead time 10 is close to perfectly calibrated for all models. For lead time $1, 2$ and 5 we observe that all models have a higher percentage of observation than the 25% to 75% prediction intervals. This indicates that these prediction intervals could be shorter.

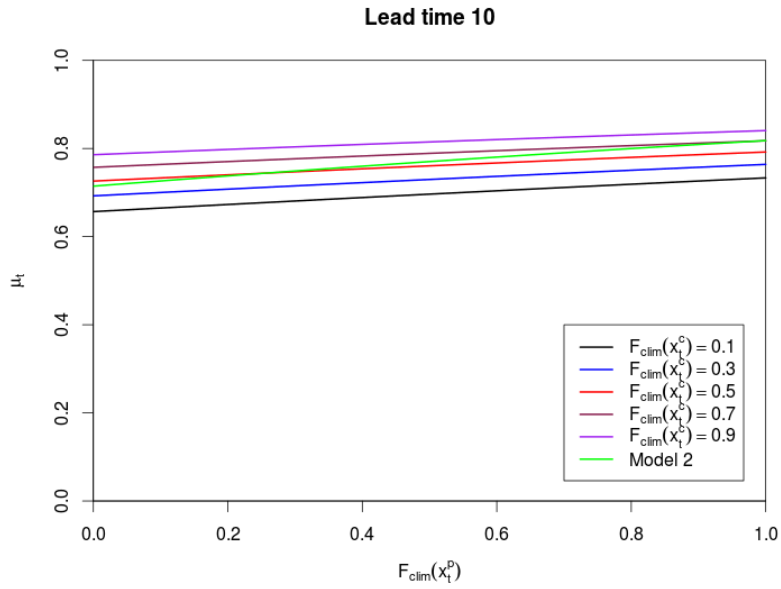
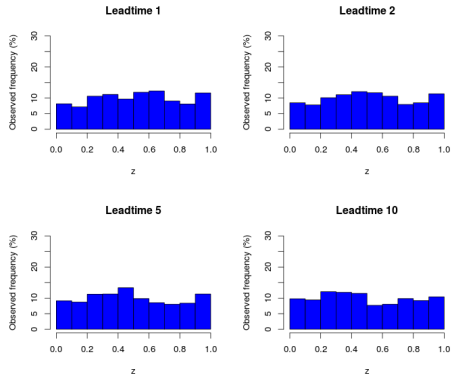
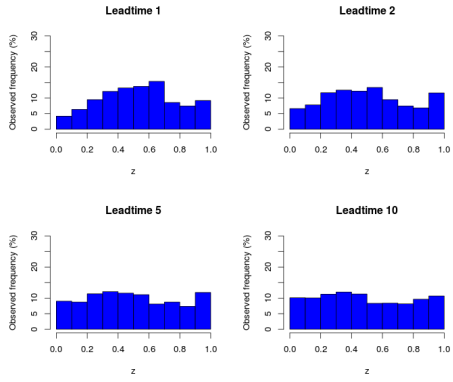


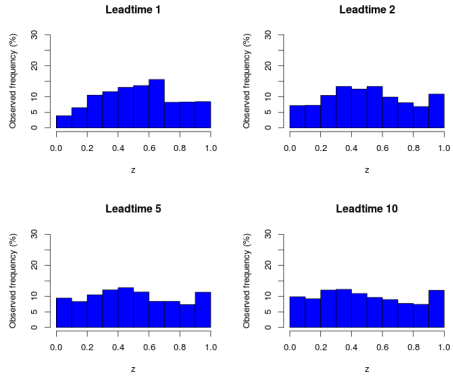
Figure 6.14: The mean of the beta distribution μ_t in Model 3 for lead time 10 as a function of the climatology cumulative probability of the persistence forecast $F_{clim}(x_t^p)$ when the hydrological forecast $F_{clim}(x_t^h) = 0.9$, for different values of the sliding window climatology $F_{clim}(x_t^c)$ and for Model 2.



(a)



(b)



(c)

Figure 6.15: PIT histograms for lead times $l = 1, 2, 5, 10$ for a) Model 1, b) Model 2 and c) Model 3.

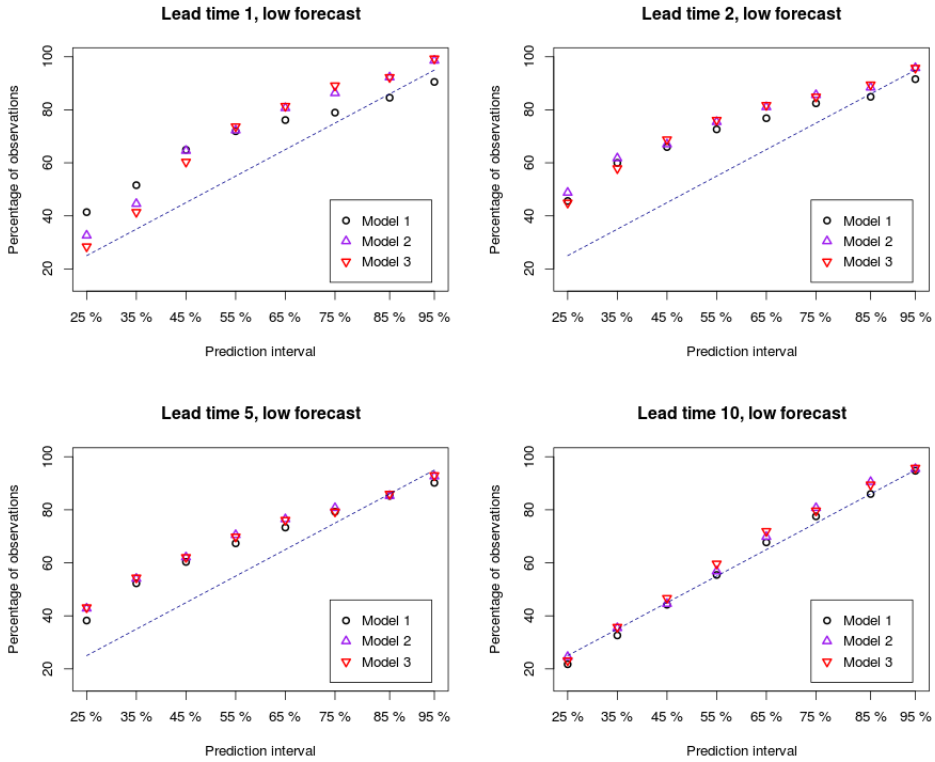


Figure 6.16: Percentage of observations for different prediction intervals when the hydrological forecast is low. The dashed line indicate where the values should lie for perfectly calibrated forecasts.

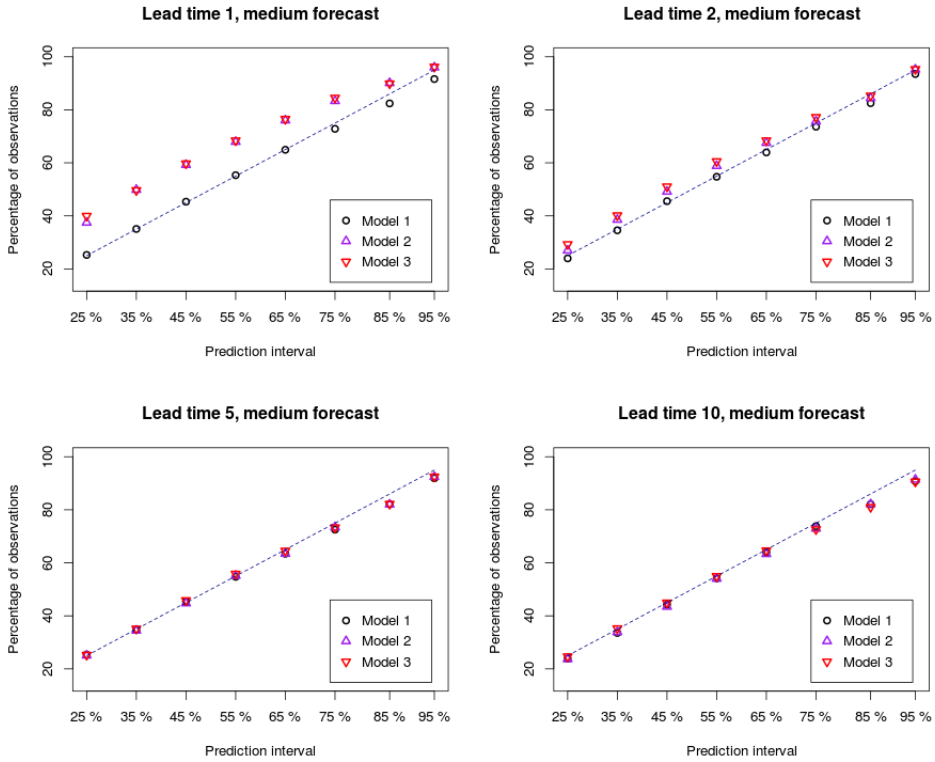


Figure 6.17: Percentage of observations for different prediction intervals when the hydrological forecast is medium. The dashed line indicate where the values should lie for perfectly calibrated forecasts.

The percent of observation in different prediction intervals for the medium hydrological forecasts of streamflow are presented in Figure 6.17. For lead time $2, 5$ and 10, we observe that all models yields similar results, and that all forecasts are close to calibrated. That the results are so similar for lead time 5 and 10 is due to the fact that both γ_p and γ_c are very small and ν is similar for for these lead times, yielding similar probabilistic forecasts for all models. For lead time 1, there is a difference between Model 1 and Model 2 and Model 3. Model 1 is close to calibrated, while Model 2 and Model 3 are overdispersed in that all points lie above the calibration line.

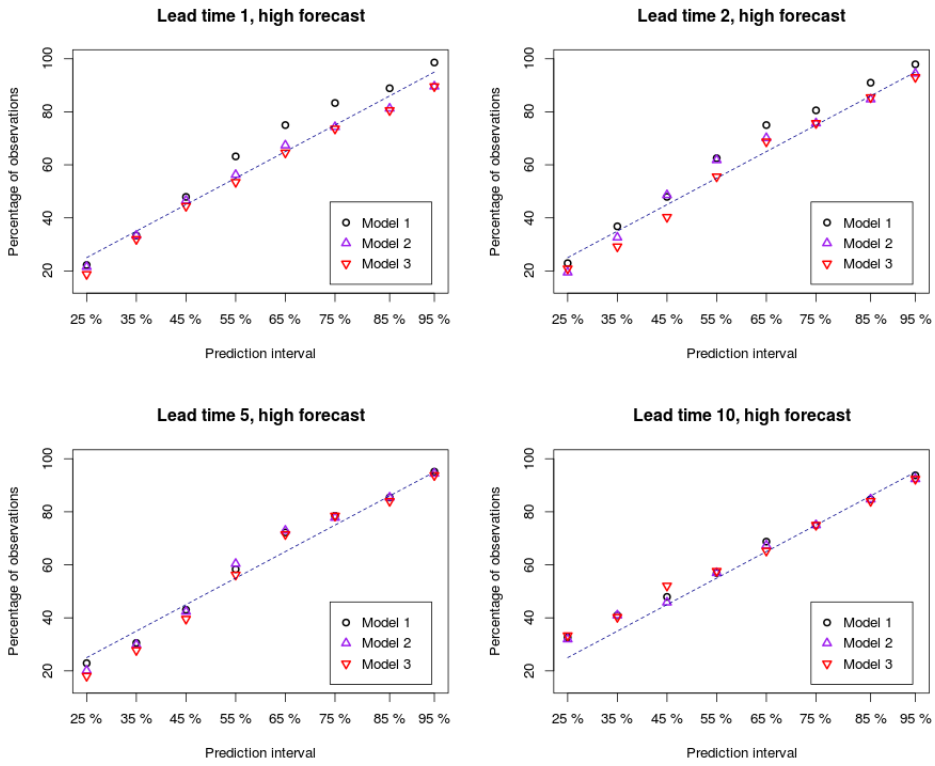


Figure 6.18: Percentage of observations for different prediction intervals when the hydrological forecast is high. The dashed line indicate where the values should lie for perfectly calibrated forecasts.

In Figure 6.18, we observe that for high hydrological forecasts for lead time 1, Model 2 and Model 3 seems to better calibrated than Model 1. For longer lead times the results are quite similar, and all models are close to calibrated. This indicates that the prediction intervals when the hydrological forecast is high seems to be of the right size.

6.3 Assessing sharpness

To evaluate the sharpness of the predictive densities based on the three models, we plot the average 95% prediction intervals for lead time $l = 1, 2, 5$ and 10 (Figure

6.19). We observe that all prediction intervals increase with increasing lead time. The prediction intervals for Model 2 and 3 are very similar for all lead times, except for lead time $l = 10$, where the model with the sliding window climatology included (Model 3) has a bit shorter average prediction interval width. When comparing the model without the persistence forecast included (Model 1) we see that the difference in average prediction interval width is largest for the two first lead times. This indicates that the persistence forecast is important to include for the shorter lead times.

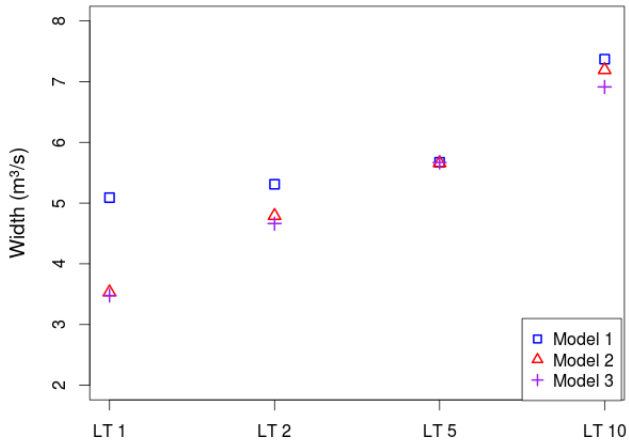


Figure 6.19: Average width of the 95% prediction intervals for lead time $l = 1, 2, 5, 10$ for Model 1,2 and 3.

To get a further impression of how the prediction intervals varied for different streamflows, we use the same division of hydrological forecasts as for the percentage of observations in different prediction intervals (low, medium and high flows). A boxplot of the prediction interval width for the different forecasts and models are given in Figure 6.20. First we note that the median width of prediction intervals increase with increasing forecasts for all models. This is due to the concept explained for Model 1 in Figure 6.5, in that larger forecasts results in a beta pdf with higher mean, which in turn results in larger prediction intervals because of the shape of the climatology cdf. Next, we observe that in general the median also increase with increasing lead time, indicating that the probabilistic forecasts are sharper for shorter lead times. When comparing the models, we observe that Model 2 and 3 yield similar results for all lead times. The major

difference between Model 1 and the two other models is for lead time 1 and 2. For lead time 1, there seems to be a reduction in 95% prediction interval width for medium and high flows between Model 1 and Model 2 and 3. For lead time 2, the main reduction in prediction intervals between Model 1 and Model 2 and 3 is for the high flows. The results indicate that the persistence forecast need to be included for lead time 1 and 2, but the sliding window climatology forecast adds little information. The boxplots in Figure 6.20, for lead time 5 and 10 are quite similar for all flows, except for high flows for lead time 10. This can be explained by the coefficients impact on the mean of the beta distribution. As we saw in Figure 6.14, the mean for Model 3 could be both lower and higher than the mean for Model 2, when the hydrological forecast was high ($F_{clim}(x_t^h) = 0.9$). A lower mean results in shorter prediction intervals because of the shape of the climatology cdf, as explained in Figure 6.5. A higher mean of the beta pdf, however results in larger prediction intervals. Hence, the larger spread in the prediction intervals for high flows for lead time 10 for Model 3 compared to Model 2 is caused by the sliding window climatology forecasts impact when modelling the mean of the beta distribution. For low values of the sliding window climatology forecast, the mean of the beta distribution is less than the mean for Model 2, resulting in shorter prediction intervals. Large values of the sliding window climatology results in a larger mean of the beta distribution in Model 3 compared to Model 2, yielding larger prediction intervals.

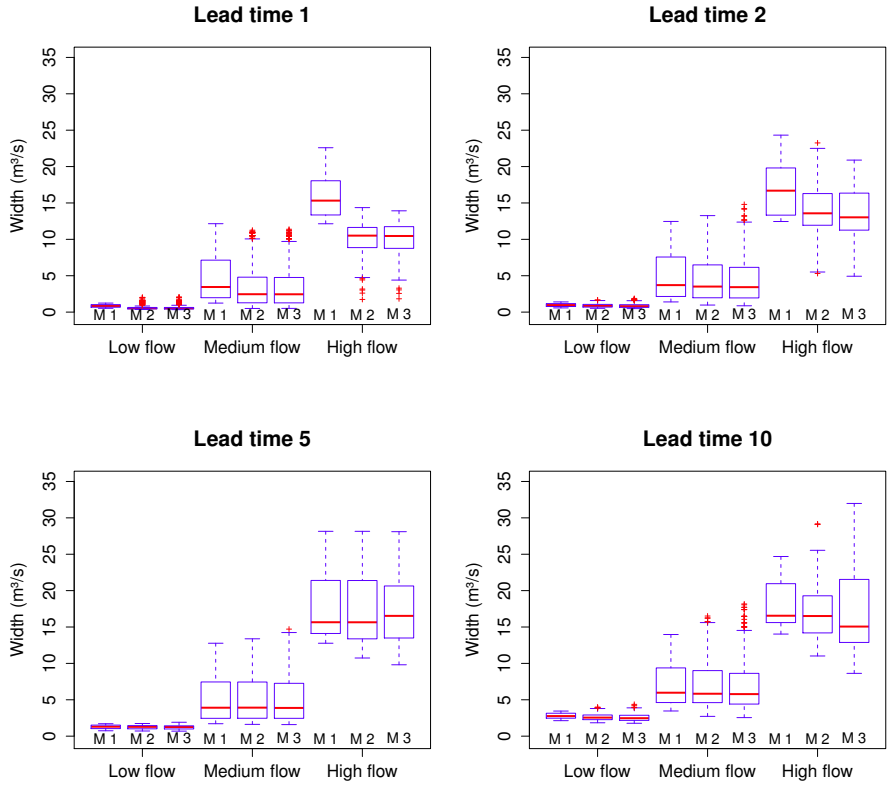


Figure 6.20: Boxplot of prediction intervals (m^3/s) for all models, divided into the 20% lowest (Low flow), the 10% highest (High flow) and the remaining hydrological forecasts (Medium flow) for lead time $l = 1, 2, 5, 10$. The blue box represents the 1st to 3rd quartile, the red line is the median and the red crosses represents the outliers.

CRPS-values was obtained by using the cross-validation scheme given in section 5.4. The CRPS for all models, including pure deterministic forecasts (i.e the mean absolute error) and the best model from the case study performed by Engeland and Steinsland [2014], refered here to as Model E&S (2014) for lead time $l = 1, 2, 5, 10$ are given in Table 6.2. First we note that the CRPS for Model 2 and 3 are very similar for all lead times. For lead time 5 and 10, all models has similar CRPS values. For the first two lead times, Model 1 has a higher CRPS value than Model 2 and 3, indicating that Model 2 and 3 has better predictive

abilities than Model 1 for these lead times. Next we note that our models outperform the deterministic forecasts for all lead times. Compared to the results of Engeland and Steinsland [2014], our postprocessing models yields similar CRPS-values. The main difference is for lead time 1, where Model E&S (2014) has the smallest CRPS, and for lead time 5 where Model 1 has the smallest CRPS. Model 2 and 3 have essentially the same CRPS for lead time 5 as Model 1.

CRPS	**l = 1**	**l = 2**	**l = 5**	**l = 10**
Hydrological forecast	0.94	0.99	1.09	1.44
Persistence forecast	0.73	1.23	1.79	2.10
Sliding window climatology	1.52	1.52	1.52	1.52
Model 1	0.67	0.70	**0.77**	1.03
Model 2	0.52	0.67	0.78	1.03
Model 3	0.52	**0.66**	0.78	1.03
Model E&S (2014)	**0.45**	0.67	0.84	**1.01**

Table 6.2: CRPS for streamflow forecasts for lead time $l = 1, 2, 5$ and 10. The bold number in each column denotes the lowest CRPS-value for the given lead time.

7. Discussion and conclusion

In this master thesis we have developed a, as far as we are aware of, new post processing methodology for the situation where the climatology and different deterministic forecasts are available. The CCPR-methodology is based on fitting a probabilistic forecast on the CCP-scale, and then transform the resulting probabilistic forecast back to the original scale. Since the methodology is new, the major focus of this thesis was to present the methodology and explore the methodology for a given case.

We have applied the CCPR-methodology to construct probabilistic forecasts for the streamflow at Osali for four lead times; $l = 1, 2, 5$ and 10. To investigate the impact of the different deterministic forecasts, we considered three models; Model 1 which included only the hydrological forecast, Model 2 with both hydrological and persistence forecast and Model 3 which in addition included the sliding window climatology forecast. The parameters were fitted by minimum CPRS estimation for the entire period.

In order to investigate the interpretation of the parameters, we performed an analysis with focus on the parameters impact in modelling the beta pdf on the CCP-scale. Through the analysis we experienced the difficulties in a straight forward interpretation of the parameters. This was both due to the use of a logit link when modelling the mean and that the beta distribution is parametrized by two parameters; μ and ν which interacts in determining the shape of the beta distribution. For instance, the variance of the beta distribution both vary with the variance parameter ν, but also with the mean parameter μ. In addition the post processed probabilistic forecast is also largely affected by the shape of the climatology cdf. Hence we found that these aspects of the model complicates a straight forward interpretation of the estimated parameters.

However, we discovered some useful strategies to obtain information from the parameters. First we note that smaller values of the variance parameter ν yield

sharper forecasts. Next we found that the constant γ_0, present in all models for modelling the mean μ_t of the beta pdf, determines the value of μ_t when $F_{clim}(\mathbf{x_t}) = 0$. A negative value, which we obtained in all models, indicates that when the climatology cdf value of the forecast $F_{clim}(\mathbf{x_t})$ is low, the mean of the beta distribution is also low. Hence, when a low streamflow is forecasted the post-processed probabilistic forecast also has high probabilities for low streamflows. To obtain a probabilistic forecast that clearly distinguishes between different values of the deterministic forecasts, the range of μ_t should be as large as possible. This is achieved by a combination of a low value for γ_0 and large values of γ_k, where the magnitude of γ_k corresponds to the predictive performance of forecast k. In the case study we observed that the magnitude of γ_k varied with lead time, indicating that the predictive performance varies with lead time. For Model 1, γ_h decreased with increasing lead time, leading to less influence of the hydrological forecast with increasing lead time. For models with more than one deterministic forecast included, the interpretation is a bit more cumbersome, but for each lead time the magnitude of γ_k indicates predictive performance. In the case study for both Model 2 and 3, we observed that for lead time 1, the persistence forecast had more impact in modelling the mean of the beta distribution than the hydrological forecast. For lead time $2, 5$ and 10 the results were opposite, in that the hydrological forecast influenced the mean more with increasing lead time compared to the persistence forecast. This is due to the fact that the hydrological forecast has more predictive information compared to the persistence forecast for longer lead times.

To evaluate the predictive performance of the models, we used a cross-validation scheme where we successively omitted one year of data when estimating the parameters and then used these parameters to evaluate the omitted year. Calibration was assessed through PIT-histograms. The evaluation showed that Model 1 was best calibrated, while Model 2 and 3 gave similar results and seemed to be a bit overdispersed for lead time 1 and 2. Further evaluation of percentage of observations in different prediction intervals showed that the overdispersion was due to too wide prediction intervals for low and medium hydrological forecasts of streamflow . To assess sharpness, average width of 95% prediction intervals in the three models was compared. Model 2 and 3 yield similar results for all lead times, while model 1 had, compared to Model 2 and 3, larger average width of prediction intervals for lead time 1 and to some extent for lead time 2 and 10. Further inspection of the prediction intervals, divided according to the magnitude of the hydrological forecast, showed that the main difference between Model 1 compared to Model 2 and 3 was for medium and high flows for lead time 1, and high flows for lead time 2. Since the results for Model 2 and 3 are very similar, the sliding window climatology does not seem to provide much additional

information to the model. Hence, based on the overall analysis we conclude that Model 2 is the best out of the three models, in that it is better than Model 1 for lead time 1 and 2, and Model 3 does not provide much predictive performance compared to the increase in parameters.

Our post processing methodology has only been tested on streamflow data from Osali. However, we believe that the framework can be used in other settings as well, where the climatology and deterministic forecasts are available. An advantage of the model is that we do not need to make assumptions for the underlying distribution of the data since we base the post processed probabilistic forecast on the climatology and the deterministic forecasts through the transformation procedure. The CCPR-methodology has the climatology as a special case, namely if the forecasts have no predictive information. This is an appealing property of the method, because the climatology is the best we can do in absence of other forecasts. On the other hand, the CCPR-methodology does not include the perfect forecast as a special case. However, we do not consider this to be a major disadvantage, since the need of a probabilistic forecast diminish in the situation of a perfect forecast. As mentioned already, a disadvantage of the model lies in the challenges faced when interpreting the parameters. However the main purpose of the CCPR-methodology is to create calibrated and sharp probabilistic forecasts, not to evaluate how the deterministic forecasts affects the final probabilistic forecast. Hence, if good results are achieved by using the CCPR-methodology, our main goal is reached.

There are several directions where the CCPR-methodology can be further developed. First we acknowledge that we have only explored a small amount of the possible ways to model the mean and the variance parameter of the beta distribution. An interesting extension of the models used in this case study is to also model the variance parameter ν, in a way that allows the variance of the beta distribution to vary. One suggestion is to use the spread of the sliding window climatology to model the variance parameter. Based on the analysis of calibration where the degree of calibration varied with the magnitude of the hydrological forecasts, it is seems to be a potential gain to model the variance parameter according to the hydrological forecasts. From the analysis of the parameters influence on the mean of the beta distribution, we observed that both very high and low means was never obtained. This could be the result of using the logit link, and hence further research could focus on different possibilities for modelling the mean. It could also be of interest to develop the climatology cdf to be continuous in stead of the step function we chose. Different smoothing techniques could be applied to generate the climatology cdf. We would also recommend to develop the climatology cdf to have a flattened probability distribution on the tails. In

our case study all forecasts and observations was larger than the smallest observation and smaller than the highest observation in the climatology, hence we did not need to focus on the tail distributions. To extend the model further, other probability distributions than the beta distribution could possibly be better when fitting a pdf on the CCP-scale. In addition, it could be interesting to investigate if the CCPR-methodology can be applied when a probabilistic forecast is already available. In this setting the climatology cdf could possibly be replaced by the probabilistic forecast, and the transformation procedure can be carried out in the same way as given in this thesis, based on some deterministic forecasts.

As a conclusion, we have introduced a new framework for creating probabilistic forecasts based on climatology and deterministic forecasts. The procedure have been applied on a set of streamflow data, yielding good results. However, the methodology should both be tested, and possibly further developed, in other situations as well to see if the procedure in general is a useful approach for making calibrated and sharp probabilistic forecasts.

Bibliography

Allard, D., Comunian, A., and Renard, P. (2012). Probability aggregation methods in geoscience. *Mathematical Geosciences*, 44:545–581.

Casella, G. and Berger, R. L. (2002). *Statistical inference*. Duxbury press.

Engeland, K. and Steinsland, I. (2014). Probabilistic postprocessing models for flow forecasts for a system of catchments and several lead times. *Water resources research*, 50:182–197.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268.

Gneiting, T. and Raftery, A. E. (2005). Weather forecasting with ensemble methods. *Science*, 310:248–249.

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

Gneiting, T., Raftery, A. E., III, A. H. W., and Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. *Monthly Weather Review*, 133(5):1098–1118.

Hora, S. C. (2004). Probability judgements for continous quantities: Linear combinations and calibration. *Management Science*, 50:597–604.

Jolliffe, I. T. and Stephenson, D. B. (2012). *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons Ltd.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association: Series B (Statistical Methodology)*, 90(430):773–795.

Lewis, J. M. (2005). Roots of ensemble forecasting. *Monthly Weather Review*, 133:1865–1885.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models. 2nd ed.* Chapman & Hall/CRC.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005). Using bayesian model averaging to calibrate forecast ensembles. *Monthly weather review*, 133(5):1155–1174.

Ranjan, R. (2009). Combining and evaluating probabilistic forecasts. *Ph.D. dissertation, University of Washington, Seattle*.

Ranjan, R. and Gneiting, T. (2010). Combining probability forecasts. *Journal of the Royal Statistical Society: Series B*, 72:71–91.

Ranjan, R. and Gneiting, T. (2013). Combining predictive distributions. *Electronic Journal of Statistics*, 7:7147–1782.

Senter for fornybar energi (2011). Vannkraft. `http://www.sffe.no/?page_id=50`. [Online; accessed 19-May-2014].