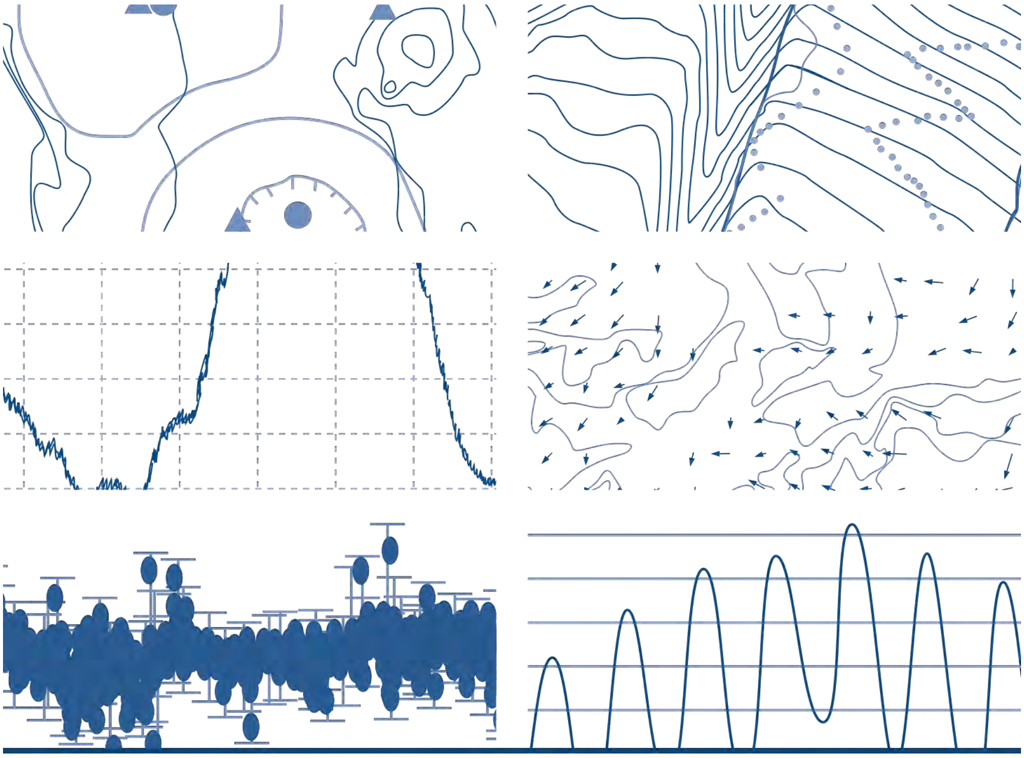
Icelandic
Meteorological Office

Veðurstofa
Íslands



# Probabilistic daily streamflow forecasts based on the combined use of a hydrological model and an analogue method

Philippe Crochet

Report

VÍ 2014-006

Icelandic
Meteorological Office

Veðurstofa Íslands

# Probabilistic daily streamflow forecasts based on the combined use of a hydrological model and an analogue method

Philippe Crochet, Icelandic Met Office

**Report**

Veðurstofa Íslands
Bústaðavegur 7–9
150 Reykjavík

+354 522 60 00
vedur@vedur.is

VÍ 2014-006
ISSN 1670-8261

Veðurstofa Íslands

Keypage

| Report no.: | Date: | ISSN: | Public ☒ Restricted ☐ |
|---|---|---|---|
| VÍ 2014-006 | December 2014 | 1670-8261 | Provision: |

| | |
|---|---|
| **Report title / including subtitle**<br>Probabilistic daily streamflow forecasts based on the combined use of a hydrological model and an analogue method | **No. of copies:** 20<br>**Pages:** 68<br>**Managing director:**<br>Jórunn Harðardóttir |
| **Author(s):**<br>Philippe Crochet | **Project manager:**<br>Philippe Crochet<br>**Project number:**<br>4812-0-0001 |
| **Project phase:** | **Case number:**<br>2013-326 |

**Report contracted for:**
Orkurannsóknasjóður

**Prepared in cooperation with:**

**Summary:**
A hydrological ensemble prediction system (HEPS) is developed by coupling the distributed hydrological model WaSiM-ETH with a meteorological ensemble prediction based on analogues. The analogue procedure is used to formulate objective precipitation and temperature ensemble forecasts with a horizontal resolution of 1 km and lead times of one to three days, which are then used as input to WaSiM-ETH. The system is evaluated on several catchments over a period of five years. Results indicate that the analogue method provides better precipitation and temperature forecasts than persistence and climate and produces reliable prediction intervals. The hydrological ensemble prediction system is capable of predicting flow discharge with reasonable accuracy and demonstrates a great potential for flow forecasting. The uncertainty of the hydrological predictions depends both on the uncertainty of the meteorological forecasts and of the hydrological modelling. A simple correction procedure improves the discharge predictions, but its use in an operational environment will be conditioned by the quality and availability of real-time flow observations.

| | |
|---|---|
| **Keywords:**<br>Iceland, weather forecast, streamflow forecast, analogue, WaSiM-ETH, ensemble | **Managing director's signature:**<br>Jórunn Harðardóttir<br>**Project manager's signature:**<br><br>**Reviewed by:**<br>Matthew Roberts, Guðrún Nína Petersen, SG |

# Contents

# 1 Introduction

The Icelandic Meteorological Office (IMO) is responsible for hydrological monitoring and issuance of flood warnings. In Iceland, floods are primarily of three different origins (Snorrason et al., 2012): (i) meteorological floods induced by rain, which are often combined with melting of snow, (ii) floods due to ice formation and release within river channels, and (iii) glacier outburst floods which originate from meltwater lakes at the edge of or beneath a glacier, or via geothermal or volcanic activity.

Floods of meteorological origin can occur in all parts of the country. They are mainly observed in autumn and during wintertime and springtime. Predicting such floods a few days ahead in a reliable manner is needed in order to ensure timely warnings and support civil protection services.

Recently, a data-driven method based on nearest neighbours was developed to forecast daily streamflow up to three days ahead, and tested on selected Icelandic catchments (Crochet, 2013). This method attempts to predict future streamflow on the basis of the comparison of current and past streamflow and meteorological observations. This technique has been used in streamflow forecasting by, for example, Karlsson & Yakowitz (1987), Galeati (1990) and Akbari et al. (2011). The advantage of this technique is its simplicity, as there is no need to describe the complexity of the hydrological processes through modelling, but its application is usually limited to the short-range. Although the results demonstrated a great potential for this method, its successful application in real-time will strongly depend on the quality and availability of streamflow observations, which can be poor or simply missing during periods of variable durations, e.g in wintertime when ice may form in rivers.

In order to continue the development of our short- and medium-range flood forecasting capability, the present report explores the possibility of coupling the distributed hydrological model WaSiM-ETH with meteorological forecasts. Coupled hydro-meteorological forecasting systems have proven to be effective tools to achieve short- and medium-range hydrological forecasts (Bartholmes & Todini, 2005; Addor et al., 2011; Smiatek et al., 2012, and many others). WaSiM-ETH, in particular, has been evaluated for this purpose (Jasper et al, 2002).

Short- and medium-range weather forecasts currently available at IMO are obtained from different numerical weather prediction (NWP) models (e.g. ECMWF, Hirlam, Harmonie). The European Centre for Medium Range Weather Forecasts (ECMWF) deterministic NWP model is a global model with a horizontal resolution of approximately 16 km and maximum lead time of 10 days. Hirlam (High Resolution Limited Area Model) is a short-range weather forecasting system for operational use jointly developed by consortium of European meteorological institutions (hirlam.org). IMO receives NWP products from the Danish Meteorological Institute at three horizontal resolutions, 15 km, 5 km and 3 km and maximum lead times of 60, 54, 48 h. Harmonie is developed jointly by the Hirlam community and Meteo-France and ALADIN and is locally operated by IMO to provide weather forecasts over Iceland with 2.5 km horizontal resolution and a maximum lead time of 48 h.

WaSiM-ETH has until now been used at IMO in water-resources and climate change research projects with a spatial resolution of 1 km and a daily temporal resolution (Jónsdóttir, 2008; Einarsson & Jónsson, 2010; Þórarinsdóttir, 2012; Crochet & Þórarinsdóttir, 2014). Operating

WaSiM-ETH with coarser horizontal resolution NWP models without prior downscaling may be problematic, especially in small and medium-sized mountainous catchments where precipitation and temperature exhibit strong gradients. This could lead to inconsistencies or biases in the hydrological simulations, especially if the model is calibrated and then operated with different meteorological inputs. Precipitation in particular is one of the most difficult variables to forecast with NWP models with reasonable accuracy, both in terms of magnitude and timing, especially flood-triggering precipitation in complex terrain. Also, in Iceland, a correct assessment of snow accumulation and snowmelt is crucial for obtaining correct simulations of runoff in winter and spring.

Sources of uncertainty in the hydro-meteorological forecast chain are many (Ramos et al., 2010). They include among others the corrections and downscaling procedure of the meteorological predictions, antecedent conditions of the system, observation networks (meteorological and hydrological) and model limitations to fully represent the hydrological processes. One of the most successful ways of dealing with uncertainty is the use of ensembles. Forecast uncertainty can be derived from the dispersion of the ensemble members. Ensemble weather predictions made by ECMWF (ECMWF-EPS) can be used as input to hydrological models to produce an ensemble of flow forecasts. Considering that their spatial resolution is even coarser than the one of the deterministic model, downscaling these EPS prior to use them as input to a hydrological model is recommended and has proven to increase the forecast skills (Renner et al., 2009).

An alternative approach for producing high-resolution ensemble meteorological forecasts is the analogue method. Analogue-based methods have been used in weather forecasting (Radinović, 1975; Kruizinga & Murphy, 1983; Van den Dool, 1989; Fraedrich et al., 2003) and climate downscaling (Zorita & Von Storch, 1999; Wetterhall et al., 2005) to extract local weather information, which can not be simulated by coarse-resolution NWP or climate models with sufficient accuracy. In the past decade in particular, this technique has been applied to produce probabilistic quantitative precipitation forecasts (PQPFs) (Obled et al., 2002; Hamill & Whitaker, 2006; Gibergans-Báguena & Llasat, 2007; Diomede et al., 2008; Marty et al., 2012). The first step of this technique is to compare a predicted meteorological situation to all past situations collected in an historical archive, according to some variables, which are supposed to be well predicted by NWP models, to extract the dates of the closest matches and to form an ensemble weather forecast with the associated meteorological observations on these dates, at locations of interest. The use of this technique requires long time-series of meteorological observations.

This report presents the development and evaluation of a hydrological ensemble prediction system (HEPS) at IMO by forcing the distributed hydrological model WaSiM-ETH with meteorological ensemble predictions based on analogues. The analogue method proposed here builds on the different approaches described above and takes advantage of high-resolution gridded daily precipitation and temperature datasets available for the whole of Iceland over a 49 year period. In Section 2, a brief description of WaSiM-ETH is provided and the principle of the analogue-based method is presented. Section 3 presents the data and catchments used in the study. Section 4 presents the evaluation of the analogue method for the prediction of precipitation and temperature, followed by an evaluation of the hydrological predictions on selected catchments. Some concluding remarks are made in Section 5.

# 2 Methods

## 2.1 WaSiM-ETH

WaSiM-ETH (Schulla & Jasper, 2007) is a processed based distributed hydrological model which has both physically based and conceptual model components. This model has been used at IMO in a number of studies. The model performs calculations per grid cell and the basin is subdivided into flow time zones for the routing of surface runoff. Several alternative modelling approaches are proposed to the user, depending on data availability. In order to define a trade-off between model complexity and observational uncertainty, the number of meteorological inputs was kept to a minimum in this study, i.e. precipitation and temperature. For this reason, the Hamon method was used to calculate the potential evapotranspiration (ETP). The empirical constants in the Hamon method were defined by comparison with Pennman-Monteith ETP. Snow accumulation and melt were modelled according to a simple temperature-index approach with various thresholds defining the rain/snow fraction and snowmelt. A simple bucket approach is used to model interception, with a leaf area index dependent storage capacity. Infiltration of water into the soil is modelled after a procedure proposed by Peschke (1987). The Richards-approach was used for modelling vertical water fluxes in the unsaturated zone, coupled dynamically with a 2-D groundwater flow model for base flow generation. The hydraulic properties of the soil are parameterised according to Van Genuchten (1976). A module solving the heat flux balance in the soil is available but had not yet been tested at the time of the study. For this reason, in order to account for frozen soil effects in winter, the soil of the entire catchment was frozen over an arbitrary period of six months (Nov-Apr), by placing an impermeable layer at the surface. The model is used at a spatial resolution of 1 km and a daily temporal resolution.

## 2.2 The analogue method

### 2.2.1 Principle and background

The analogue method assumes that if two meteorological situations are similar regarding large scale atmospheric circulation, they should also be similar with respect to local meteorological conditions. Local meteorological conditions depend on the synoptic situation but local features such as orography and surface properties play also an important role.

Considering a given meteorological situation characterised by some synoptic meteorological variables (observed or predicted), the search for past situations similar (analogue) to that synoptic situation and the associated local weather observed on these days, should give valuable information on how local meteorological conditions could develop. Hence, the method exploits the concept of analogy to formulate a probabilistic prediction of some local meteorological variables, conditional to that given synoptic situation. In other words, the analogue method capitalises on historical information collected at location of interest and can therefore be seen as an objective expert prediction system based on past experience.

This technique has been applied in weather forecasting (Radinović, 1975; Kruizinga & Murphy, 1983; Obled et al., 2002, Fraedrich et al., 2003; Hamill & Whitaker, 2006; Gibergans-Báguena & Llasat, 2007; Diomede et al., 2008; Marty et al., 2012) and climate downscaling (Zorita & Von Storch, 1999; Wetterhall et al., 2005) to indirectly extract local meteorological information from NWP or climate models, such as precipitation or temperature, which is not directly simulated with sufficient accuracy. These models on the other hand are suitable for predicting synoptic variables characterising atmospheric circulation.

The analogue method is applied here to make short- and medium-range predictions of daily precipitation and temperature given predicted synoptic meteorological situations provided by a NWP model. The first step of the technique is to compare a predicted meteorological situation at time $t$ to all past situations collected in a historical archive, according to certain atmospheric variables and some analogy criteria. Then, the dates of the $N$ best analogues, i.e. the $N$ closest matches to the target situation are extracted and sorted, and a meteorological ensemble prediction is formed, with the precipitation and temperature sample, observed on these days ($N$ members):

$$P(i,t) = P(u_i) \tag{1}$$

$$T_{2m}(i,t) = T_{2m}(u_i) \tag{2}$$

where $t = t_0 + D$ is the date of the target situation, $t_0$ the initial time, $D$ the forecast lead time ($D \geq 0$) and $u_i$ is the date of the $i^{th}$ analogue, $(i = 1,N)$.

In other words, the analogue method finds dates in the archives corresponding to situations close to the target one according to certain atmospheric variables and reshuffles observed precipitation and temperature series on these dates according to their degree of similarity (analogy) with the predicted meteorological situation. This sample reflects objectively our best knowledge of potential future precipitation and temperature outcome at location of interest, given the predicted synoptic meteorological situation and past experience.

In practise, the archive length is limited and it is not possible to compare two atmospheric states precisely but only through a limited subset of variables. Therefore, it is very unlikely that absolute similar situations can be obtained. This is especially true if the target meteorological situation is not observed but predicted with more or less accuracy. This is why an ensemble of analogues is extracted to reflect the uncertainty of the method given the available information. This ensemble is used to define a conditional probability distribution for the variables of interest and the target day, given the predicted meteorological situation.

### 2.2.2 Prerequisites

Several practical and methodological aspects need to be considered before applying the analogue method (see also Crochet, 2013):

- The selection of atmospheric variables (the predictors) describing the meteorological situation should be physically linked to the variable to be predicted (the predictant). These atmospheric variables have to be well predicted by the NWP model to be used. The historical archive in which analogue situations are searched for results from a trade-off between i) archive length ii) number of variables defining the system and iii) data homogeneity (Obled et al., 2002). The quality of the archive needs to be homogeneous throughout the entire period so that no bias is introduced when searching for analogues. Inhomogeneities could result from e.g. instrumental changes, network density and data processing techniques (Obled et al., 2002). A limitation of the method is that it cannot forecast a value lower or larger than recorded in the archive without some sorts of post-processing. Such a strategy is not included in the proposed method at this stage of development. For this reason, the historical archives need to be as long as possible so that a large variety of situations can be found. This is especially important for events whose return period may be longer than the archive length. Rare situations never observed in the past may prove difficult to predict.

- The size of the analogy domain needs to be adapted to the problem to be solved. As the domain size increases, it may prove difficult to find close analogues. According to Van den Dool (1989), simplifying assumptions need to be made. In particular, if we are only concerned with a limited area, it may be sufficient to restrict the size of the analogy domain around that area to find good analogues.

- The number of analogues to select must result from a compromise between sampling quality and decreasing degree of analogy, which also depends on the analogy criteria (Obled et al., 2002). The larger the number of selected analogues the better the sampling but the lower the analogy with the current situation. In particular, rare events whose return periods are longer than the archive length may prove difficult to sample properly even if the sample size is increased and systematic biases may be expected.

### 2.2.3 Predictors

Mean sea level pressure (*MSLP*), geopotential height (*Z*), specific humidity (q) and temperature (T) at different pressure levels are considered in this study to describe the meteorological situations at the synoptic scale and to identify weather analogues. The *MSLP* and geopotential height (*Z*) describe atmospheric circulation patterns and are expected to influence precipitation and temperature, through their associated wind regime. Humidity ($q$) and temperature ($T$) give additional information on the associated air masses.

### 2.2.4 Analogy criteria

Two criteria are used to select the weather analogues. Firstly, the similarity of atmospheric circulation patterns (*MSLP* and *Z* fields) between target and candidate situations is evaluated with the Teweles-Wobus ($S1$) skill score (Wilks, 1995). The $S1$ score compares the shape of two fields by considering their gradient at each grid point of the analogy domain:

$$S1(u) = 100\frac{\sum_{i=1}^{n-1}\sum_{j=1}^{m}|\Delta A_i - \Delta F_i| + \sum_{i=1}^{n}\sum_{j=1}^{m-1}|\Delta A_j - \Delta F_j|}{\sum_{i=1}^{n-1}\sum_{j=1}^{m} G_i + \sum_{i=1}^{n}\sum_{j=1}^{m-1} G_j} \tag{3}$$

with

$$\Delta A_i = A(i+1, j, u) - A(i, j, u) \tag{4}$$

$$\Delta F_i = F(i+1, j, t) - F(i, j, t) \tag{5}$$

$$\Delta A_j = A(i, j+1, u) - A(i, j, u) \tag{6}$$

$$\Delta F_j = F(i, j+1, t) - F(i, j, t) \tag{7}$$

and

$$G_i = max(|\Delta A_i|, |\Delta F_i|) \tag{8}$$

$$G_j = max(|\Delta A_j|, |\Delta F_j|) \tag{9}$$

where $F$ is the target (predicted) $MSLP$ or $Z$ field at time $t$, $A$ is the analyzed $MSLP$ or $Z$ field from the archive at time $u$ (i.e candidate analogue), $\Delta F_i$, $\Delta F_j$ are the predicted gradients in the east-west and south-north directions, $\Delta A_i$, $\Delta A_j$ the analyzed gradients in the same directions, around a given grid point (i,j), $G_i$, $G_j$ the maximum of these two gradients and $n,m$ the number of grid points in the east-west and south-north directions, respectively.

A perfect match between two situations leads to S1=0 and thus the smaller S1, the better. This score has been used by e.g. Obled et al. (2002), Wetterhall et al. (2005) and Marty et al. (2012) with either $MSLP$ or $Z$, to forecast precipitation. It was also applied to $MSLP$ fields by Woodcock (1980) for temperature forecasts. According to Woodcock (1980), the $S1$ skill score measures the similarity of the geostrophic winds between two situations and is an eminent suitable tool for selecting analogues for those weather elements largely controlled by wind regimes.

Secondly, the similarity with respect to humidity ($q$) and temperature ($T$) is evaluated by calculating the RMSE over all grid points (i,j) of the analogy domain. The smaller the RMSE, the better:

$$RMSE(u) = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(F(i,j,t) - A(i,j,u))^2}{nm}} \tag{10}$$

## 2.3 Hydrological ensemble prediction system (HEPS)

The hydrological ensemble prediction system (HEPS) proposed in this study is passively coupling WaSiM-ETH with the precipitation and temperature ensemble predictions obtained by the analogue method. The model is run off-line in conditions similar to an operational run. Two runs are made each day: i) a run in analysis (control) mode and ii) a run in forecast mode. In control mode, observed precipitation and temperature are used to force WaSiM-ETH at initial time $t_0$ and results are used to initialize the forecast mode. The model initialisation with observations is expected to prevent any long-term drift in the hydrological predictions that could result from accumulating error in the predicted meteorological inputs. In forecast mode, the meteorological ensemble prediction is used to force WaSiM-ETH. The model is run sequentially for every ensemble member, thus producing an ensemble of flow forecasts.

Considering a daily time step and a meteorological ensemble prediction made of $N$ members ($N$ analogues) at forecast range $t_0 + D$ days ($D \geq 1$), the number of hydrological forecasts (members) to be sequentially issued at lead time $t_0 + D$ is $M_D$=$N^D$. This means that each day, WaSiM-ETH should be run once in analysis mode and $M_D$ times in forecasting mode.

The different steps of HEPS up to $D$=3 days are as follows:

1. Import observed (or analyzed) gridded daily temperature and precipitation ending at $t_0$ ($D$=0).

2. Run WaSiM-ETH in analysis (control) mode with input meteorological data from step 1 and save all storage grids.

3. Import precipitation and temperature ensemble predictions issued at $t_0$ for lead time $D$=1 day ($N$ members).

4. Initialize WaSiM-ETH with storage grids from step 2 and run model with input meteorological data from step 3 ($M_1$=$N$ sequential runs).

5. Save all storage grids from step 4 ($M_1$=$N$ sets of grids).

6. Import precipitation and temperature ensemble predictions issued at $t_0$ for lead time $D$=2 days ($N$ members).

7. Initialize WaSiM-ETH with storage grids from step 5 and run model with input meteorological data from step 6 ($M_2$=$N^2$ sequential runs).

8. Save all storage grids from step 7 ($M_2$=$N^2$ sets of grids).

9. Import precipitation and temperature ensemble predictions issued at $t_0$ for lead time $D$=3 days ($N$ members).

10. Initialize WaSiM-ETH with storage grids from step 8 and run model with input meteorological data from step 9 ($M_3$=$N^3$ sequential runs).

11. Move to next day ($t_0$): initialize WaSiM-ETH with storage grids from step 2 and go to step 1.

## 2.4 Correction procedure

In order to reduce systematic biases in the flow forecasts, if any, a recursive error correction procedure which builds on the algorithm proposed by Boi (2004) is defined, assuming that discharge observations are available in real-time and are of good quality. The algorithm establishes a correction to be applied on predicted discharge issued at time $t_0$, based on the error between observed and analyzed discharge in the past $n$ days:

$$corr(t_0) = \sum_{j=0}^{n} 0.5^{j+1}(Q_{obs}(t_0 - j) - Q_{analysis}(t_0 - j)) \tag{11}$$

$$Q_{analysis_{corr}}(t_0) = Q_{analysis}(t_0) + corr(t_0) \tag{12}$$

$$Q_{corr}(i, t_0 + D) = Q_{DMO}(i, t_0 + D) + corr(t_0) \tag{13}$$

Where $Q_{obs}$ is the observed discharge, $Q_{analysis}$ is the simulated discharge in analysis (control) mode, $Q_{analysis_{corr}}$ the corrected analysis, $Q_{DMO}$ is the predicted discharge (direct model output), $Q_{corr}$ the predicted discharge after correction and $corr$ the correction term. The term 0.5 means that half the contribution to the correction term is given by the last correction term, and this contribution decays as we go back in time.

## 2.5 Probabilistic predictions

An empirical probability distribution function can be derived from the meteorological and hydrological ensemble predictions and several statistics extracted. A $100(1-p)\%$ prediction interval can be constructed with quantiles corresponding to the cumulative probabilities $p/2$ and $1-p/2$: $[F_{p/2}; F_{(1-p/2)}]$, where $F$ is precipitation ($P$), 2m temperature ($T_{2m}$) or discharge ($Q$) ensemble prediction. The probability (not) to exceed a threshold can also be estimated by calculating how many members are (not) exceeding this threshold.

## 2.6 Deterministic predictions

A deterministic prediction ($DF$) can be derived from the meteorological and hydrological ensemble predictions ($F$), considering different strategies, such as taking the best (first) ensemble member, the arithmetic mean or weighted mean, considering the rank of the ensemble member or the value of the analogy criteria:

$$DF(t_0+D) = \frac{\sum_{i=1}^{N} w_i F(i, t_0+D)}{\sum_{i=1}^{N} w_i} \tag{14}$$

where $N$ is the number of members in the ensemble and $w_i$ the weight of each member.

## 2.7 Forecast evaluation statistics

Several metrics were used to evaluate the forecast skills. Probabilistic forecasts were evaluated in terms of reliability, i.e. the agreement between forecast probability of an event and the mean observed frequency of that event (Renner et al., 2009). A probabilistic forecast is reliable if the observation (i.e. precipitation, temperature or discharge) lies in the empirical $100(1-p)\%$ prediction interval, $100(1-p)\%$ of the time. The ranked probability score (*RPS*) (Wilks, 1995) was also used to evaluate the probabilistic flow forecasts across $K$=10 classes of flow quantiles, covering all observed outcomes. The closer *RPS* is to 0, the better, and the worst possible *RPS* score is ($K$–1). The performances of the deterministic forecast were measured by the mean error (ME), root-mean squared error (RMSE) and Nash-Sutcliffe coefficient (Nash and Sutcliffe, 1970).

# 3 Data

## 3.1 Meteorological data

Predictors:

Mean sea level pressure (*MSLP*), geopotential height ($Z$), specific humidity (q) and temperature (T) at different pressure levels constitute the predictors describing the meteorological situations at synoptic scale, used to identify weather analogues. These fields are extracted twice daily (00UTC and 12UTC) on a 1° x 1° latitude-longitude grid from the ECMWF operational analysis and forecasts available for the period 2001–2006. The analogue meteorological situations are extracted from the ERA-40 reanalysis archive (Uppala et al., 2005) for the period 1958–2001.

Predictants:

Gridded series of daily 2m temperature ($T_{2m}$) and precipitation ($P$) valid from 00UTC to 00UTC are used to force WaSiM-ETH in control mode and to produce the analogue-based meteorological forecasts used to force WaSiM-ETH in forecast mode. These two datasets are available for the common period 1958–2006 with a horizontal resolution of 1 km. The gridded temperature dataset was obtained from the interpolation of temperature observations at meteorological stations (Crochet & Jóhannesson, 2011). The gridded precipitation dataset was obtained by downscaling ERA-40 precipitation with the LT-model (Crochet et al., 2007; Jóhannesson et al., 2007), whose parameters were optimised by comparison with ground measurements and glaciological data. This dataset is not strictly speaking an observational dataset but rather an analyzed dataset.

## 3.2 River basins

A set of five river basins of various types, for which daily discharge was available over several decades, was selected for this study (Figure 1). Table 1 presents information on the characteristics of each catchment. Rivers in Iceland are usually classified according to their source (Rist, 1990; Jónsdóttir et al., 2008), namely direct runoff (D), groundwater fed (L), glacial rivers (J) and whether they flow through lakes (S). The combination of letters indicates the origin of flow with the first letter indicating the primary origin. The hydrological regime of these catchments is influenced by rainfall in autumn and winter, snowmelt in spring (and glacier melt in summer).

Table 1. Characteristics of the considered watersheds. Letter combinations indicate the type of river, with the first letter indicating the primary type. Direct runoff river (D), presence of lakes (S), glacier-fed river (J), groundwater (L).

| Gauging station | vhm 19 | vhm 10 | vhm 26 | vhm 66 | vhm 64 |
|---|---|---|---|---|---|
| Name | Dynjandisá | Svartá | Sandá | Hvítá | Ölfusá |
| Type of river | D+L | D+L | D+L | D+J+L | L+D+J+S |
| Drainage area (km$^2$) | 42 | 397 | 267 | 1664 | 5687 |
| Mean altitude (m a.s.l.) | 555 | 535 | 391 | 664 | 480 |
| % Glacierized area | 0 | 0 | 0 | 20 | 12 |
| % Grassland | <1 | 43 | 25 | 10 | 25 |
| % Woodland | <1 | 0 | 0 | 2 | 2 |
| % Moss | 59 | 25 | 0 | 14 | 13 |
| % Little or no vegetation | 35 | 27 | 30 | 63 | 36 |
| % Wetland | 0 | 5 | 46 | 6 | 7 |
| % Lakes | 6 | <1 | <1 | <1 | 3 |

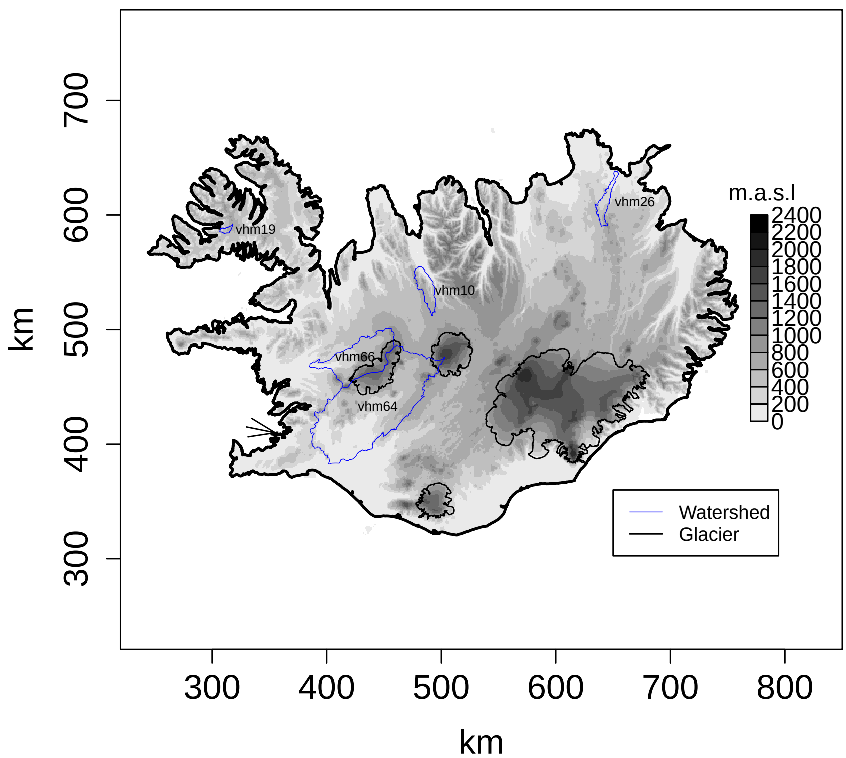



*Figure 1. Topography of Iceland and location of the watersheds considered here.*

# 4 Results

## 4.1 Analogue-based meteorological ensemble forecasts

In total, twenty different methods combining up to three different predictors were compared for the search of weather analogues (Table 2) and the prediction of daily 2m temperature and precipitation. Since daily precipitation, temperature and discharge are calculated from 00UTC to 00UTC, the ECMWF meteorological forecasts issued at initial time 00UTC are used to extract the predictors in forecast mode. Methods 1 to 19 are applied at 12UTC, i.e. centred on the 24h period for which temperature and precipitation are calculated. Method 20 is applied at both 00UTC and 12UTC, i.e. at the start and centre of the 24h period for which gridded daily surface temperature and precipitation are calculated.

The analogy domain around Iceland was chosen to be large enough to include areas with noticeable influence on the circulation patterns (see Figure 2). According to Van den Dool (1989), if we are only concerned by a limited area, it may be sufficient to restrict the size of the analogy domain around that area to find good analogues. With the proposed method, selected analogue days are the same for all catchments.

Seasonal effects were taken into account by defining a moving temporal window of $\pm$ K days centered on the target day, for the selection of weather analogues, so that in an archive made of $Y$ years, each target day was at most associated to $Y$ x (2K+1) potential analogues. By doing so, it is hoped that candidate situations will present similar characteristics in terms of solar energy, surface fluxes and other characteristics.

Because of time limitation, the analogy domain and temporal windows were not strictly speaking optimized. The following tests were performed:

- Analogy domain 1 (AD1): 60–70°N, 35–5°W.
- Analogy domain 2 (AD2): 55–75°N, 40–0°W.
- Temporal window 1 (TW1): $\pm$ 45-day window centered on the predicted calendar day.
- Temporal window 2 (TW2): $\pm$ 30-day window centered on the predicted calendar day.

For comparison, a $\pm$ 20-day temporal window was used by Kruizinga & Murphy (1983) to forecast temperature, a $\pm$ 45-day window was used by Hamill & Whitaker (2006) for precipitation forecast while Obled et al. (2002) used a $\pm$ 2-month window.

In order to concentrate the effort on the identification of the best predictor-predictant relationship, independently of the error affecting the ECMWF forecasts, the selection of the best method was made in perfect prognosis conditions, which means that the predictors are analyzed fields (D=0) and not forecast fields (D>0). The resulting predictors are then used for all lead times.

The data were split into two periods. The target (forecast or analyzed) period (01/09/2001–31/08/2006), and the archive period (01/01/1961–31/12/2000) used for the search of analogue days and their associated surface precipitation and temperature fields. For each day in the target period 2001 to 2006, the ECMWF operational analysis (forecast) was compared to the ERA-40 archive, according to the fields of predictors defined in Table 2, and the dates of the $N$ best analogues to target situation (either analysed or predicted at time $t$), noted.

In practise, each pre-selected candidate analogue day was sorted from best to worst, according to the analogy criteria defined in Section 2.2.4, and the $N$ best ones selected. When more than one pressure level was used for a given predictor (i.e. Z, cf. Table 2), the analogy criteria was calculated for each level and averaged, and each day ranked from best to worst. When different predictors were used together (i.e. Z, q and T, cf. Table 2), each candidate analogue day was first ranked according to $S1$, the 50 best analogues extracted, then resorted according to $RMSE$ and finally the $N$ best analogues selected ($N \leq 50$).

Daily temperature and precipitation observed on these sorted analogue days were then used to form the ensemble meteorological prediction for the target day which was compared to the observed weather on that day. Because of time limitation, the number of selected weather analogues was not strictly speaking optimised, and a comparison between $N$=25 and $N$=40 only was considered. For information, Kruizinga and Murphy (1983) used $N$=30 in their analogue-based temperature forecast, Obled et al. (2002) and Diomede et al. (2008) used $N$=50 in their analogue-based precipitation forecast, Gibergans-Báguena & Llasat (2007) used $N$=25, and Marty et al. (2012) used $N$=30 at $D$=0 up to $N$=60 at $D$=3 days.

The method was evaluated for catchment-averaged daily precipitation and temperature and not at individual grid points. The comparison between $N$=25 and $N$=40 indicated that results were similar regarding deterministic forecasts, so only results with $N$=25 are presented and used in the rest of the study.

Table 2. Predictors, analogy domains and temporal windows used to identify weather analogues.

| Predictors / Method | *MSLP* | $Z_{1000}$ | $Z_{850}$ | $Z_{700}$ | $Z_{500}$ | $q_{850}$ | $q_{700}$ | $T_{500}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | AD1 / TW1 | | | | | | | |
| 2 | AD2 / TW1 | | | | | | | |
| 3 | AD1 / TW2 | | | | | | | |
| 4 | AD2 / TW2 | | | | | | | |
| 5 | | | AD1 / TW1 | | | | | |
| 6 | | | AD2 / TW1 | | | | | |
| 7 | | | AD1 / TW2 | | | | | |
| 8 | | | AD2 / TW2 | | | | | |
| 9 | | | AD1 / TW1 | | AD1 / TW1 | | | |
| 10 | | | AD2 / TW1 | | AD2 / TW1 | | | |
| 11 | | | AD1 / TW2 | | AD1 / TW2 | | | |
| 12 | | | AD2 / TW2 | | AD2 / TW2 | | | |
| 13 | | AD1 / TW1 | | AD1 / TW1 | | | | |
| 14 | | AD2 / TW1 | | AD2 / TW1 | | | | |
| 15 | | AD1 / TW2 | | AD1 / TW2 | | | | |
| 16 | | AD2 / TW2 | | AD2 / TW2 | | | | |
| 17 | | | AD2 / TW2 | | AD2 / TW2 | AD1 / TW2 | | AD1 / TW2 |
| 18 | | | AD2 / TW1 | | AD2 / TW1 | AD1 / TW1 | | AD1 / TW1 |
| 19 | | | AD2 / TW1 | | | | AD1 / TW1 | |
| 20 | | AD1 / TW1 | | AD1 / TW1 | | | | |

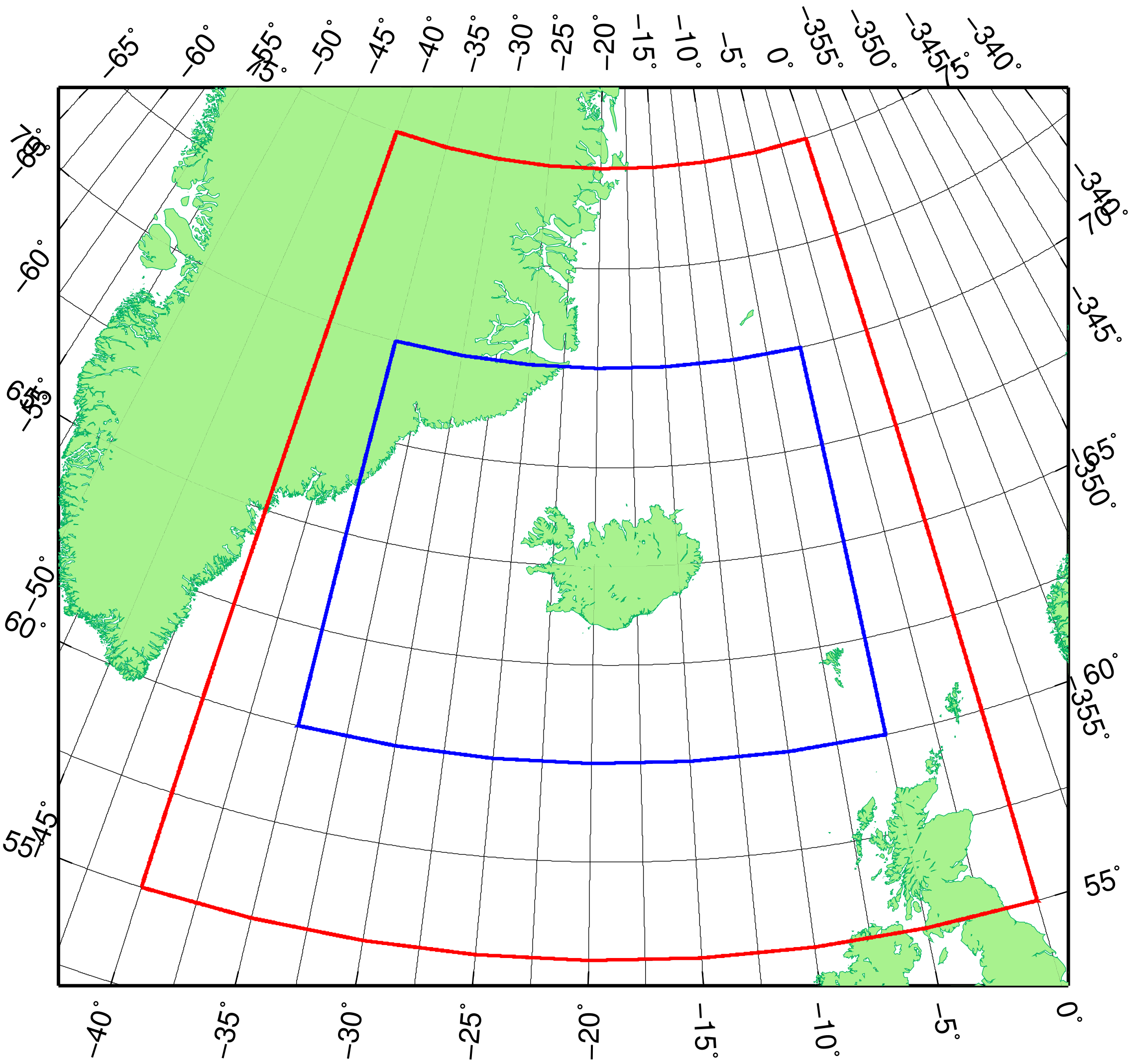



*Figure 2. Analogy domains AD1 (blue polygon) and AD2 (red polygon).*

Figures 3 and 4 present the performances of each analogue method (cf. Table 2), considering a deterministic prediction obtained by taking the weighted mean of the ensemble precipitation or temperature predictions. The weight was defined according to the rank of the analogue. Some preliminary results (not shown) indicated that the weighted mean seems to provide better deterministic predictions than the best ensemble member or the ensemble median. The analysis was performed by comparing observations and predictions, in terms of mean error (ME) and root mean squared error (RMSE). Each method was ranked from best to worst according to ME and RMSE. Overall, the best results are obtained with method No. 13. The best method according to temperature is No. 11 and the best one according to precipitation is No. 20. A closer inspection of the results indicates that the different methods are relatively comparable in terms of skills. It is also observed (not shown) that a systematic cold bias is affecting the temperature predictions in summer ($-2°C < ME < -1°C$), suggesting that in that season, large-scale atmospheric circulation may have less influence on local temperature variations than during the rest of the year. This bias could also be related to the fact that the decade 2001–2010 was warmer than the period 1961-2000 (Crochet & Jóhannesson, 2011). The second best method, No. 15, was selected rather than the best one, to produce the meteorological forecasts to be used as input to WaSiM-ETH, because temperature predictions are slightly less biased in summer with that method than with the best one (not shown). In principle, the temperature bias observed in summer could be corrected, but this was not considered in the present study.

The selected analogue method was then used to produce an ensemble of temperature and precipitation predictions for lead times $D$ of 1 to 3 days. In principle, $N$ should be increased with $D$ to compensate the increasing uncertainty associated to the predictors forecasted by the NWP model. This was not considered in the present study and the number of analogues ($N$) was kept constant with $D$. Deterministic predictions were derived from the ensemble mean, weighted by the rank of the ensemble member. Tables 3 and 4 summarise the results for all catchments, with that method. The deterministic predictions were also compared to two benchmark deterministic predictions, i) monthly climate in the period 1961–2000, i.e. $F(t_0+D) = E[A(M)]$, and ii) persistence, i.e. $F(t_0+D) = A(t_0)$, where $F$ is the forecast and $A$ the observation or analysis.

Figures 5 to 8 present the scatter plots of observed temperature and precipitation versus deterministic forecasts for two river catchments, vhm10 and vhm26. These catchments have been selected to evaluate the hydrological predictions. Appendices I and II present the results for the other catchments.

For temperature forecasts, the analogue method outperforms persistence and climate with respect to $RMSE$, but not $ME$ which is more biased than persistence. Analogue-based and climate-based temperature predictions have the tendency to be negatively biased. As mentioned earlier, this systematic bias is only observed in summer ($-2°C < ME < -1°C$). This result could suggest that in that season, large-scale atmospheric circulation may have less influence on local temperature variations than during the rest of the year, and/or that the climate in the period 1961-2000 was slightly colder than in the period 2001–2006. As expected, extreme events are difficult to predict, and temperature on warmest days is underestimated (too cold), while temperature on coldest days is overestimated (too warm). As explained earlier, this is related to the finite length of the archive, making difficult to predict rare events with return periods longer than the archive length and/or never observed in the archive. Persistence, on the other hand is not biased as expected, but the forecast skill for that method degrades rapidly with increasing lead time, as judged by the *RMSE* and scatter.

For precipitation forecasts, the analogue method outperforms persistence and climate with respect to *RMSE* and is equally unbiased. Heavy precipitation events are usually underestimated. As explained earlier, this underestimation is mostly related to the finite length of the archive, making difficult to predict rare events with return periods longer than the archive length, without some sort of post-processing.

As the lead time increases, so does the uncertainty of temperature and precipitation forecasts, as judged by *RMSE*, but only slightly for the analogue method. This result indirectly indicates that the ability of ECMWF to forecast geopotential height does not decrease much with the lead time, for the considered lead times. This is an advantage for the analogue method which in turn should benefit the hydrological forecasts.

Finally, results indicate that the prediction intervals derived from the ensemble precipitation and temperature forecasts are reliable for all catchments and all lead times, especially for temperature (see Figs. 9, 10 and Appendix III). This means that the observed temperature and precipitation lie in the empirical $100(1-p)\%$ prediction interval $100(1-p)\%$ of the time, on average. These results indicate that the number of selected analogues (*N*=25) is sufficiently large for all tested lead times with respect to reliability.

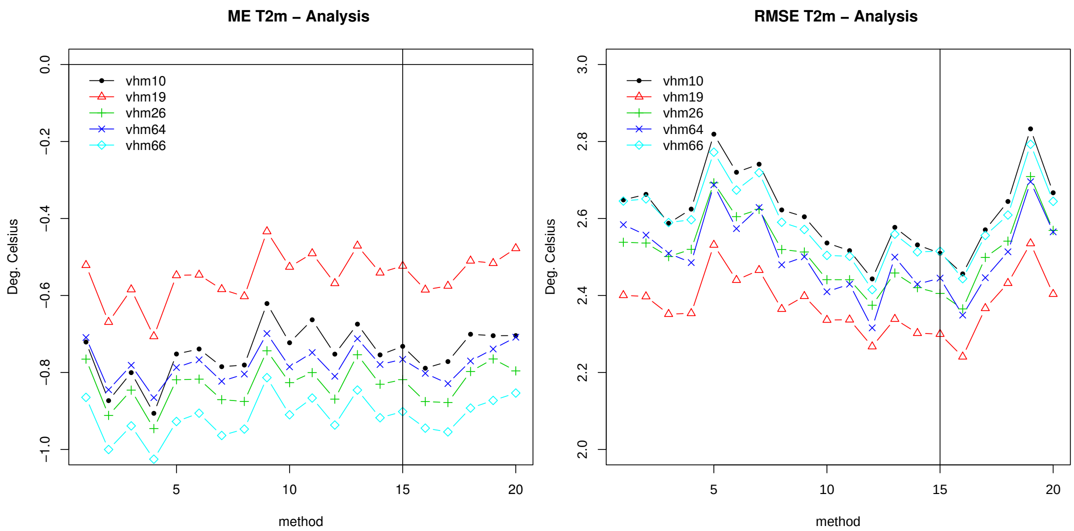



*Figure 3. Results of intercomparison of twenty different analogue methods. Mean error (left) and RMSE (right) for the analogue-based ensemble mean temperature forecast (°C). Vertical line indicates the selected method to perform the ensemble meteorological predictions.*

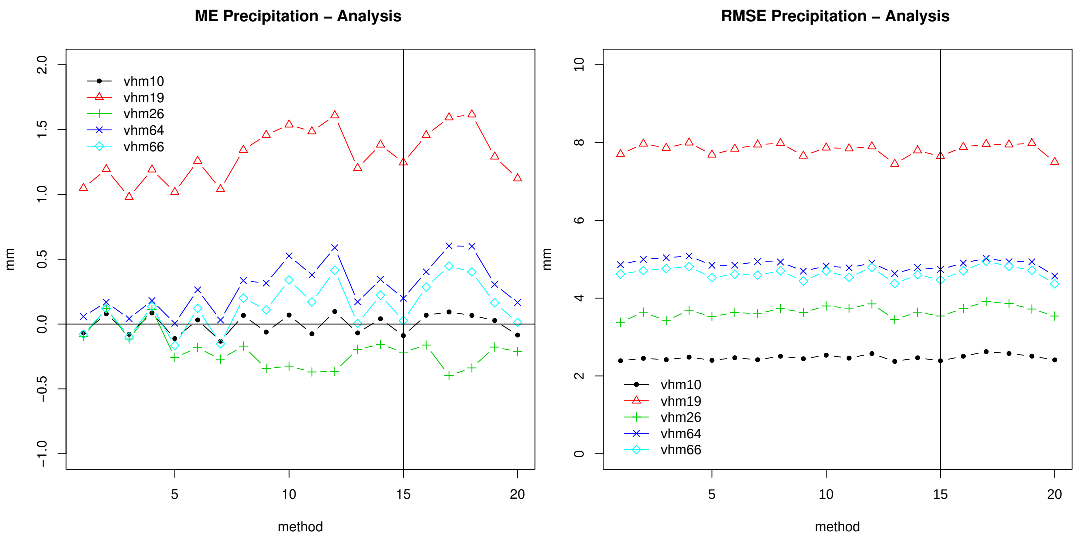



*Figure 4. As Fig. 3 but for precipitation (mm).*

Table 3. Results of the verification for the temperature forecasts, over the period 01/09/2001–31/08/2006 (method No. 15). The deterministic prediction is derived from the ensemble mean, weighted by the rank of the ensemble members. All units are in *Celsius*.

| Forecast range | Analysis (*D*=0) | | D=1 day | | D=2 days | | D=3 days | |
|---|---|---|---|---|---|---|---|---|
| Statistics | ME | RMSE | ME | RMSE | ME | RMSE | ME | RMSE |
| vhm10 | –0.63 | 2.5 | –0.66 | 2.5 | –0.69 | 2.5 | –0.74 | 2.6 |
| vhm19 | –0.49 | 2.3 | –0.51 | 2.3 | –0.53 | 2.2 | –0.55 | 2.3 |
| vhm26 | –0.72 | 2.4 | –0.73 | 2.4 | –0.76 | 2.4 | –0.82 | 2.4 |
| vhm64 | –0.67 | 2.4 | –0.69 | 2.5 | –0.73 | 2.4 | –0.76 | 2.5 |
| vhm66 | –0.81 | 2.5 | –0.84 | 2.5 | –0.87 | 2.5 | –0.91 | 2.6 |

Table 4. As in Table 3 but for precipitation. All units are in *mm*.

| Forecast range | Analysis (*D*=0) | | D=1 day | | D=2 days | | D=3 days | |
|---|---|---|---|---|---|---|---|---|
| Statistics | ME | RMSE | ME | RMSE | ME | RMSE | ME | RMSE |
| vhm10 | –0.02 | 2.4 | 0 | 2.4 | 0.01 | 2.5 | –0.06 | 2.6 |
| vhm19 | 1.2 | 7.5 | 1.2 | 7.6 | 1.2 | 7.7 | 1.1 | 8.1 |
| vhm26 | –0.26 | 3.5 | –0.21 | 3.5 | –0.16 | 3.7 | –0.22 | 3.9 |
| vhm64 | 0.18 | 4.8 | 0.17 | 4.7 | 0.13 | 4.8 | 0.08 | 5.1 |
| vhm66 | 0.11 | 4.4 | 0.14 | 4.4 | 0.12 | 4.5 | 0.08 | 4.7 |

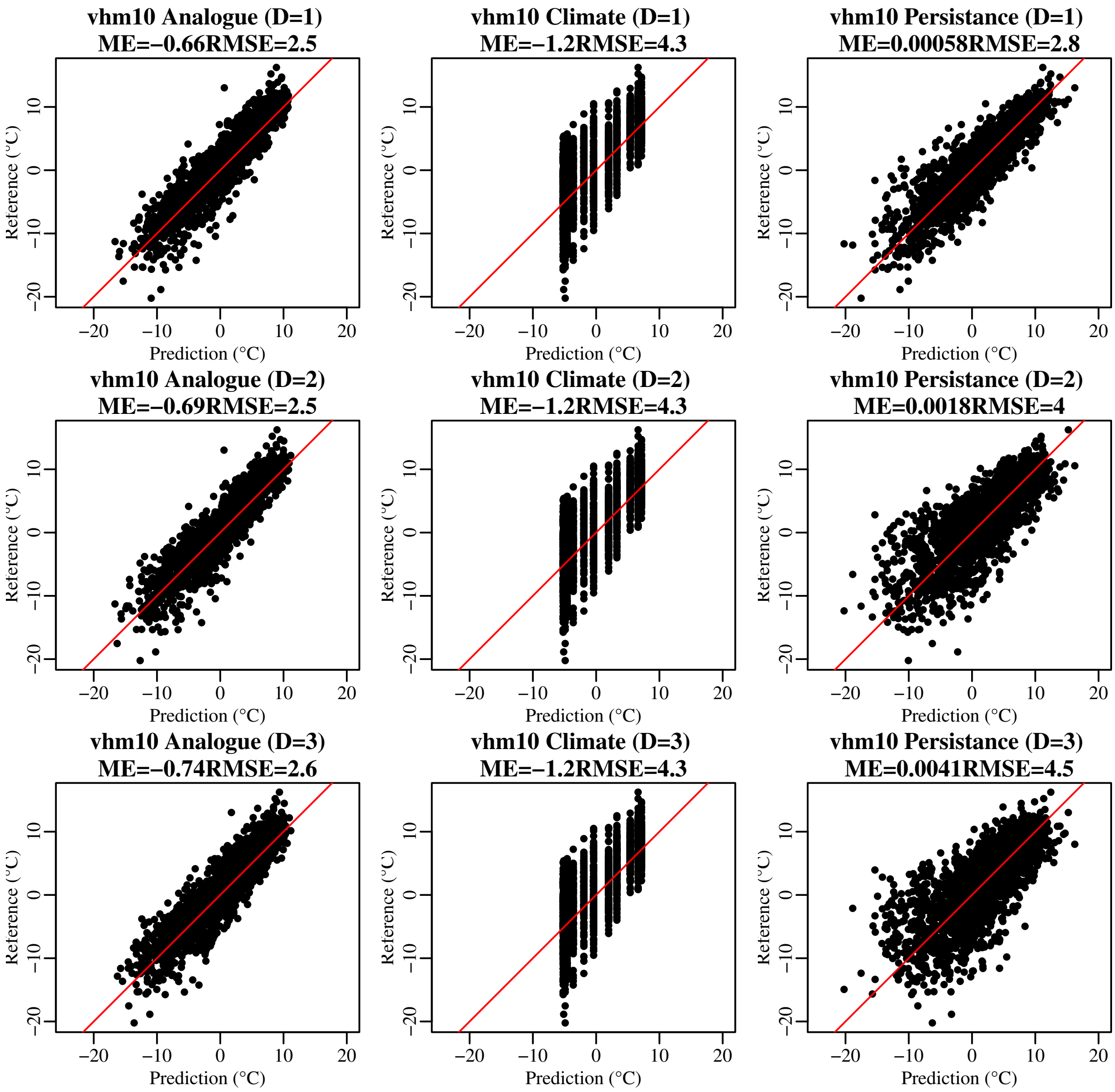



*Figure 5. Observed vs. deterministic temperature forecasts for vhm10 (catchment-averaged). Top (D=1 day), centre (D=2 days), bottom (D=3 days). Analogue method (left), climate (centre) and persistence (right). The 1:1 line corresponds to a perfect match. Mean error (ME) and root-mean squared error (RMSE) are also indicated.*

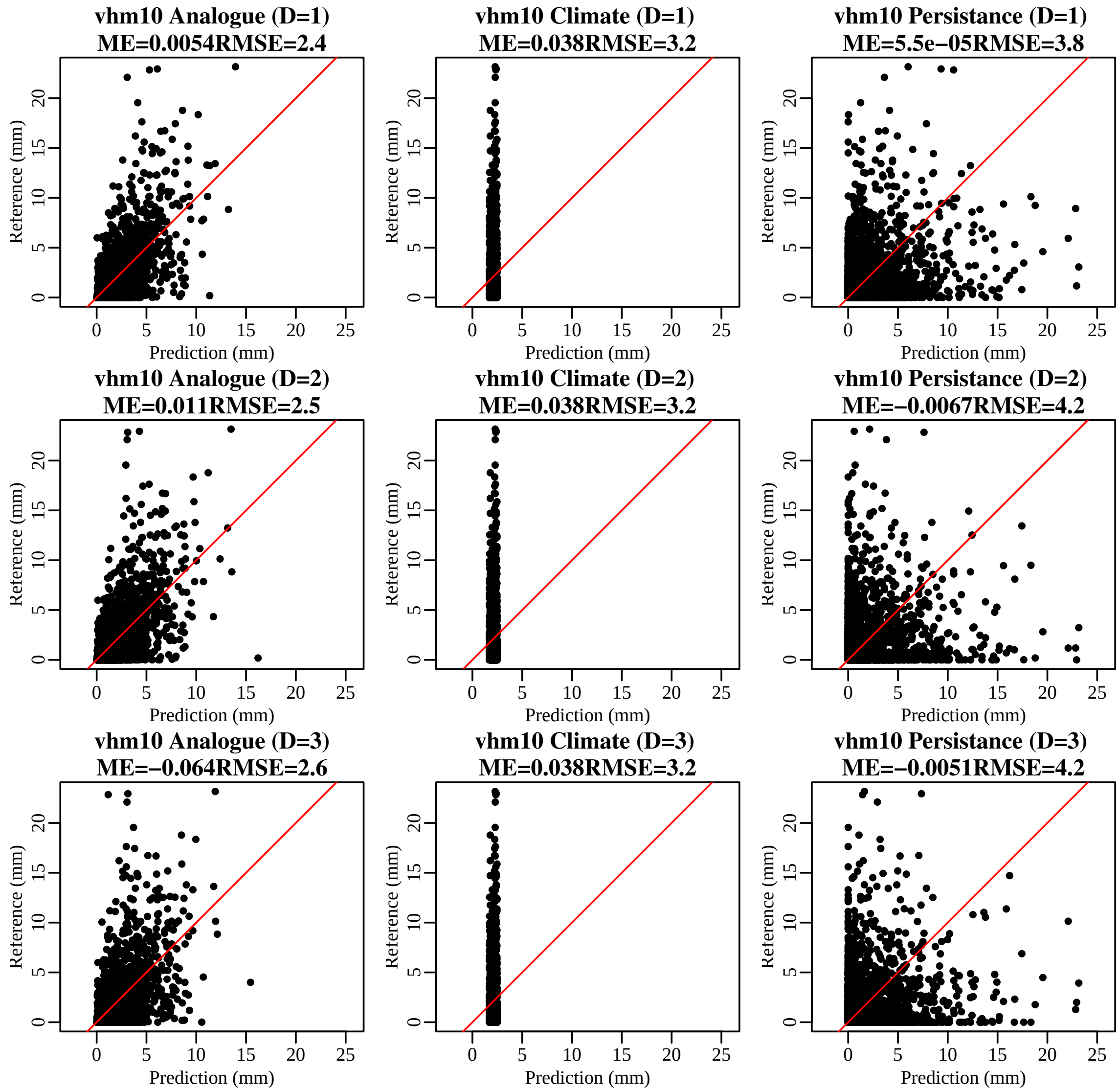



*Figure 6. As Fig. 5 but for precipitation.*

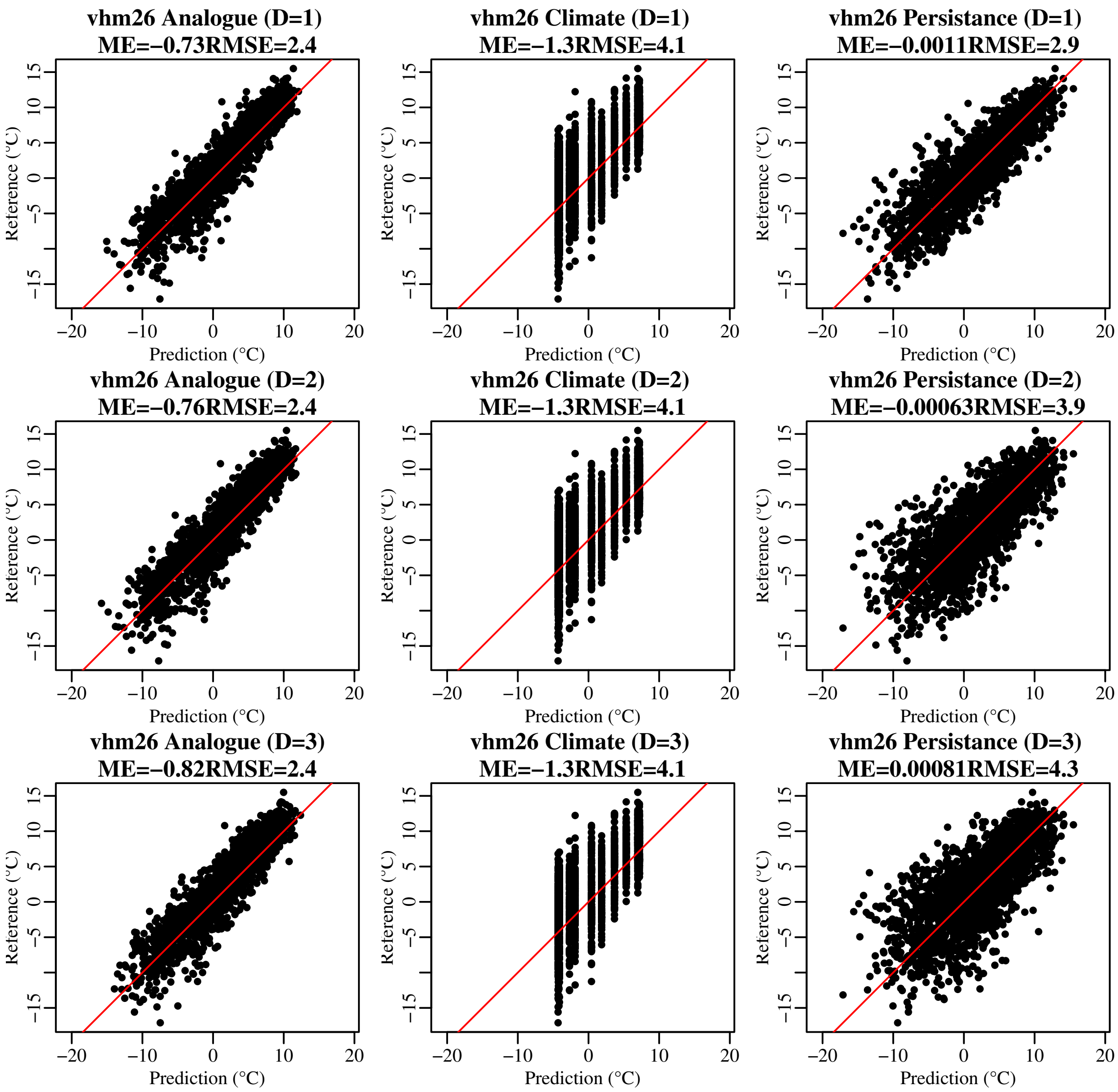



*Figure 7. Observed vs. deterministic temperature forecasts for vhm26 (catchment-averaged). Top (D=1 day), centre (D=2 days), bottom (D=3 days). Analogue method (left), climate (centre) and persistence (right). The 1:1 line corresponds to a perfect match. Mean error (ME) and root-mean squared error (RMSE) are also indicated.*

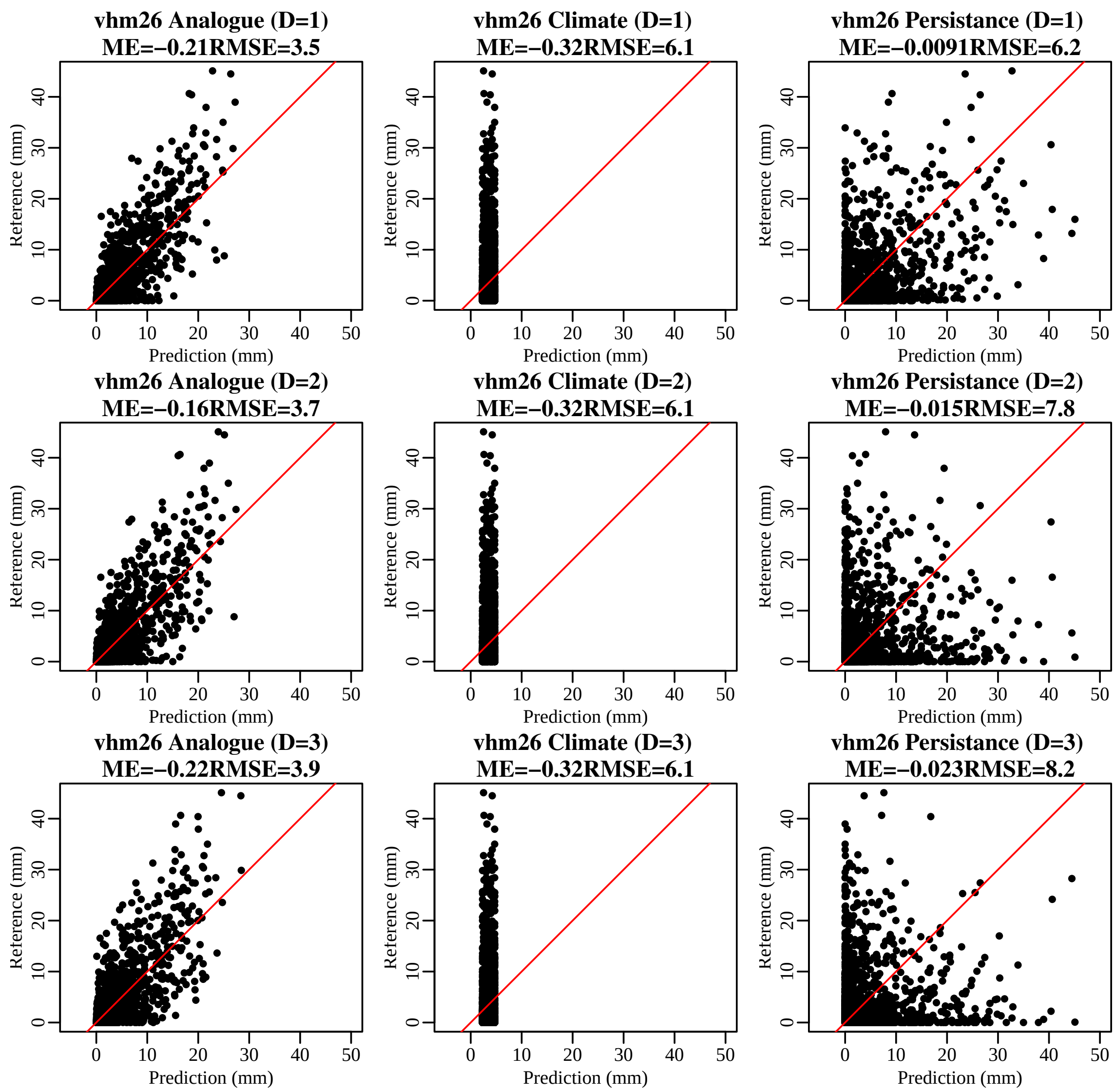



*Figure 8. As Fig. 7 but for precipitation.*

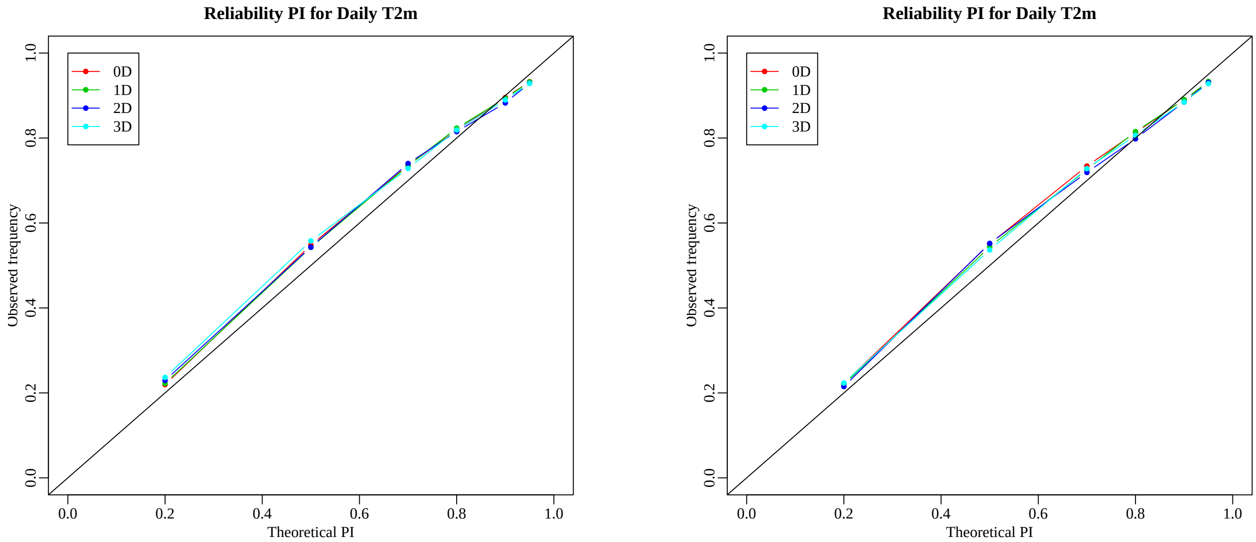



*Figure 9. Reliability diagrams for temperature forecasts: Observed relative frequency vs. theoretical prediction interval for vhm10 (left) and vhm26 (right). The 1:1 line corresponds to a perfect match.*

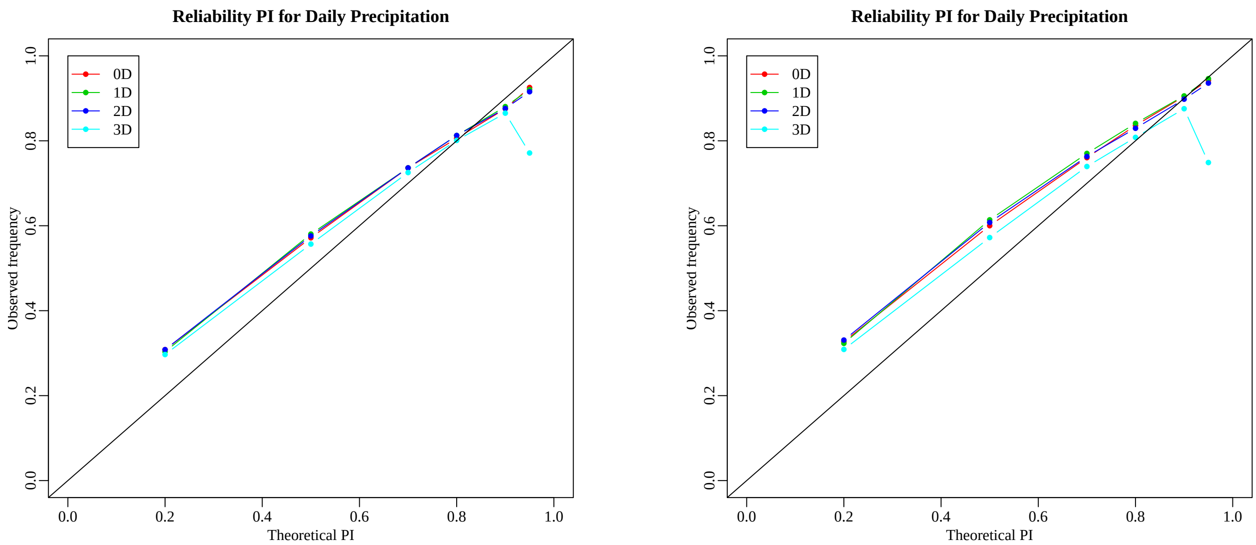



*Figure 10. As Fig. 9 but for precipitation.*

## 4.2 WaSiM-ETH model calibration and evaluation

Because of time limitation, two of the five studied catchments only were used for evaluating the ensemble flow forecasts (vhm10 and vhm26, cf. Table 1). As these catchments had already been analysed with WaSiM-ETH in previous published and unpublished studies, the resulting model setups and modeling strategy were partly adopted. In particular, for vhm26, results from Einarsson & Jónsson (2010) were used. A set of seven model parameters were calibrated: 1) recession constant of direct runoff, 2) drainage density, 3) soil percolation, 4) temperature threshold for beginning snowmelt, 5) temperature dependent melt factor, 6) fraction of snowmelt which is direct flow and 7) storage capacity of snow for water. The recession constant of interflow was arbitrarily set to twice the recession constant of direct runoff.

Experience usually shows that one calibrated parameter set may not yield equally good simulations of all parts of the observed hydrograph. In addition, using flow discharge at the catchment outlet only for calibrating the model, as in the present study, can lead to a problem of equifinality (Beven & Freer, 2001). This means that many different parameter sets can give similar model performance. In view of this problem, multi-objective calibration strategies have been proposed, such as multi-site calibration, multi-response or multi-variable calibration (see for instance Yapo et al., 1998; Gupta et al., 1998; Bergström et al., 2002; Engeland et al., 2006; Raj Shrestha & Rode, 2008; and many others). The multi-objective calibration is based on the assumption that a single objective function cannot adequately measure the important characteristics of the observed data and attempts to identify an ensemble of optimal solutions based on a trade-off between different objective functions. This means that there may exist more than one optimal solution as it is normally not possible to find a single parameter set minimising all objective functions. Here, a multi-response approach was adopted, considering the following performance criteria:

Nash-Sutcliffe coefficient (Nash & Sutcliffe, 1970) on daily discharge:

$$NASH = 1 - \frac{\sum_{i=1}^{n}(Q_{sim}(i) - Q_{obs}(i))^2}{\sum_{i=1}^{n}(Q_{obs}(i))^2 - \frac{1}{n}(\sum_{i=1}^{n} Q_{obs}(i))^2} \tag{15}$$

Logarithmic Nash-Sutcliffe coefficient on daily discharge:

$$LogNASH = 1 - \frac{\sum_{i=1}^{n}(log(Q_{sim}(i)) - log(Q_{obs}(i)))^2}{\sum_{i=1}^{n}(log(Q_{obs}(i)))^2 - \frac{1}{n}(\sum_{i=1}^{n} log(Q_{obs}(i)))^2} \tag{16}$$

Nash-Sutcliffe coefficient on mean daily discharge (seasonality):

$$NASH_{clim} = 1 - \frac{\sum_{i=1}^{365}(Q_{simclim}(i) - Q_{obsclim}(i))^2}{\sum_{i=1}^{365}(Q_{obsclim}(i))^2 - \frac{1}{365}(\sum_{i=1}^{365} Q_{obsclim}(i))^2} \tag{17}$$

Logarithmic Nash-Sutcliffe coefficient on mean daily discharge (seasonality):

$$LogNASH_{clim} = 1 - \frac{\sum_{k=1}^{365} (log(Q_{simclim}(k)) - log(Q_{obsclim}(k)))^2}{\sum_{k=1}^{365} (log(Q_{obsclim}(k)))^2 - \frac{1}{365} (\sum_{k=1}^{365} log(Q_{obsclim}(i)))^2} \tag{18}$$

Root-mean squared error on annual discharge:

$$RMSEA = \sqrt{\frac{\sum_{j=1}^{m} (\frac{1}{n(j)} \sum_{i=1}^{n(j)} Q_{sim}(i) - \frac{1}{n(j)} \sum_{i=1}^{n(j)} Q_{obs}(i))^2}{m}} \tag{19}$$

with $Q_{obs}$: observed discharge, $Q_{sim}$: simulated discharge, $n$: sample size, $m$: number of years, $n(j)$: number of days in year $j$, $Q_{obsclim}$: mean observed discharge on julian day $k$ ($1 \leq k \leq 365$), $Q_{simclim}$: mean simulated discharge on julian day $k$.

The choice of the performance criteria was guided by the objective to reasonably simulate the seasonality of daily streamflow and the annual water balance. Different performance criteria can be defined if the goal is to focus on floods for instance. However, this would require to increase the length of the calibration period so as to have a representative flood sample, which in turn would require to reduce the number of simulations because of computer limitations.

An ensemble of 500 parameter-sets was formed by randomly generating the values of each parameter from a uniform distribution. The corresponding 500 model runs were completed for the period Sept. 1990 to Aug. 1998 for vhm10 and 1995 to Aug. 2001 for vhm26, including a spin-up period of three years. Each model setup was then ranked from best (1) to worst (0) according to each of the above criteria. Then, the average rank was calculated and each model setup sorted accordingly. The best model setup was then selected and validated. The validation period was Sept. 1981 to Aug. 2005 for vhm10 and Sept. 1961 to Aug. 2005 for vhm26, including a spin-up period of five years. A precipitation correction by a factor 1.35 and 1.1 was applied to vhm10 and vhm26, respectively. A constant flow of 4 $m^3/s$ was added to the simulated streamflow for vhm26 to account for an external source of groundwater flow suspected to enter the catchment.

Results of the simulations in the calibration and validation periods are summarised in Tables 5 and 6 and Figures 11 to 14. Overall, WaSiM-ETH model simulations compare well with observations. In the detail, it proved difficult to adequately simulate low flow in winter, especially for vhm10, possibly because of uncertainties in the observed discharge in that season. The magnitude of winter floods induced by heavy rainfall on frozen ground was usually underestimated. Runoff resulting from spring snowmelt was generally well simulated, but some discrepancies were observed in some years in late spring or early summer, due to an over (under) estimation of the snowpack. The best model parameterisation identified in the calibration period was found to be acceptable in the validation phase and selected to conduct the hydrological predictions. By using one single parameter set, the uncertainty related to the model calibration will not be included in the evaluation of the hydrological predictions at this stage of the study.

Table 5. Model calibration/validation at gauge vhm10.

| Period / Statistics | Nash | LogNash |
|---|---|---|
| Calibration | 0.62 | 0.58 |
| Validation | 0.59 | 0.53 |

Table 6. Model calibration/validation at gauge vhm26.

| Period / Statistics | Nash | LogNash |
|---|---|---|
| Calibration | 0.76 | 0.63 |
| Validation | 0.74 | 0.67 |

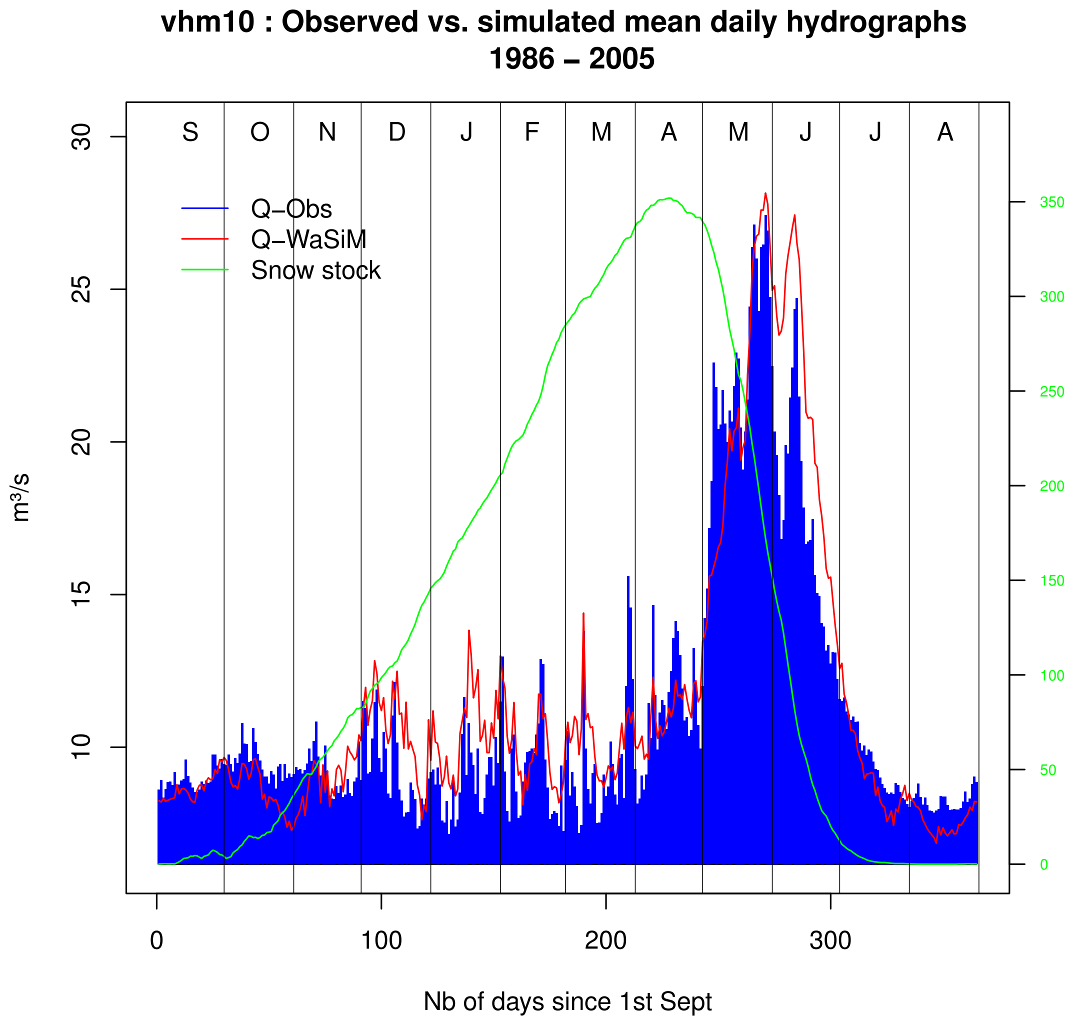



*Figure 11. WaSiM-ETH model validation at gauge vhm10. Discharge seasonality: observed (blue), best model simulation (red). Simulated catchment-averaged snowpack (green) (mm SWE).*

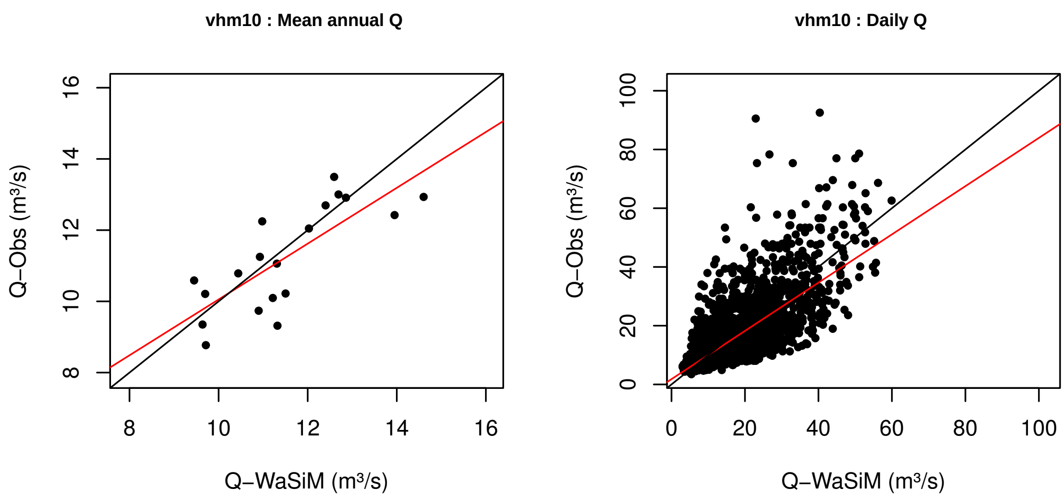



*Figure 12. WaSiM-ETH model validation at gauge vhm10. Observed versus simulated discharge. Annual discharge (left), daily discharge (right). The black line corresponds to the 1:1 line and the red line to the regression line.*

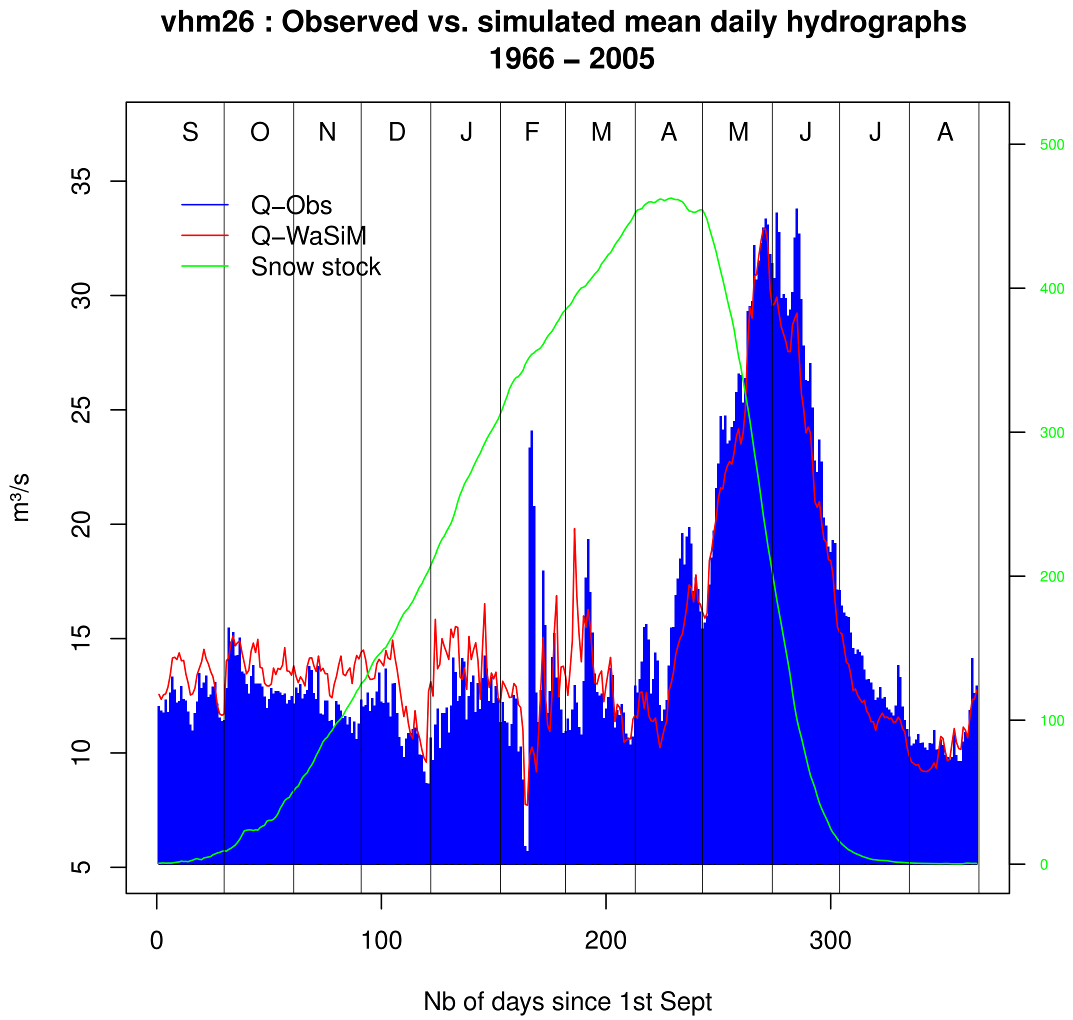
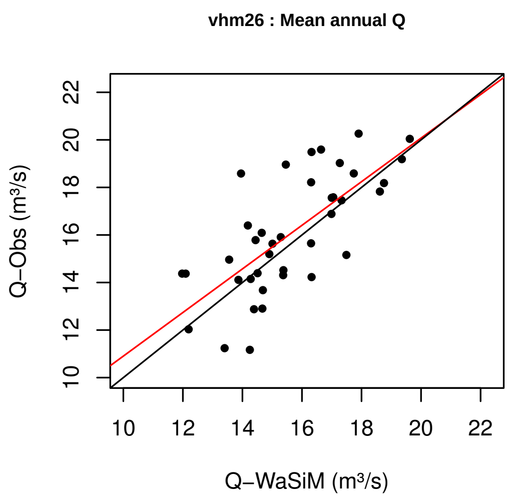



*Figure 13. WaSiM-ETH model validation at gauge vhm26. Discharge seasonality: observed (blue), best model simulation (red). Simulated catchment-averaged snowpack (green) (mm SWE).*



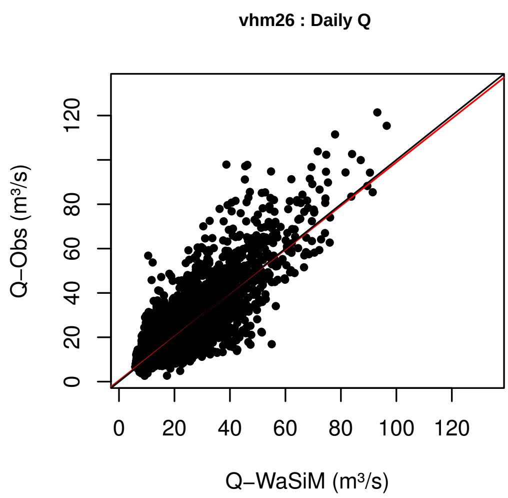




*Figure 14. WaSiM-ETH model validation at gauge vhm26. Observed versus simulated discharge. Annual discharge (left), daily discharge (right). The black line corresponds to the 1:1 line and the red line to the regression line.*

## 4.3 Hydrological predictions

The meteorological ensemble predictions obtained with the analogue method selected in Section 4.1 were used as input to WaSiM-ETH to produce hydrological ensemble predictions up to three days ahead. The precipitation correction defined in the calibration/validation procedure was applied, and for vhm26, an external source of water was considered by adding 4 $m^3$/s to the simulated discharge.

Considering that the meteorological ensemble prediction is made of $N$=25 members, this leads to a hydrological ensemble prediction made of $M_D = N^D$ members at lead time $t_0 + D$ days (see Section 2.3). This gives $M_1$=25 members at lead time $t = t_0 + 1$ day, $M_2$=625 members at lead time $t = t_0 + 2$ days and $M_3$=15625 members at lead time $t = t_0 + 3$ days. Because of computational limitations, the number of ensemble members was arbitrarily limited as follows. For vhm10, $M_1$=25 at $D$=1 day, $M_2$ =75 at $D$=2 days and $M_3$=150 at $D$=3 days. For vhm26, the number of members was slightly larger than for vhm10: $M_1$=25 at $D$=1 day, $M_2$ =100 at $D$=2 days and $M_3$=200 at $D$=3 days. Which means that each day, WaSiM-ETH is run once in analysis mode and 250 times in forecasting mode for vhm10 and 325 times for vhm26.

In order to analyse the contribution of different error sources and discriminate between meteorological uncertainty and hydrological uncertainty, three sets of verifications were conducted:

- The first one evaluates the errors stemming from the hydrological components of the model chain (model structure and its calibration) and from the meteorological and hydrological observations. In particular, it should be kept in mind that the meteorological observations are gridded and do not reflect perfectly the true meteorological situation. This evaluation is done by comparing the control (analysis) run ($Q_{analysis}$) against observed discharge ($Q_{obs}$).
- The second evaluation is used to verify the skills of the meteorological predictions and to highlight the propagation of the meteorological uncertainty into hydrological uncertainty. This is done by comparing the predicted discharge ($Q_{DMO}$) against the analysis ($Q_{analysis}$).
- The third evaluation is used to verify the overall chain (meteorological, hydrological and observational). This is done by comparing the predicted discharge before ($Q_{DMO}$) and after correction ($Q_{corr}$) against observed discharge ($Q_{obs}$).

Tables 7 to 13 and Figures 15 to 20 give an overview of the skills of the deterministic forecasts, considering i) the ensemble mean, ii) ensemble median and iii) the best (first) ensemble member. Figures 21 to 23 evaluates the reliability of the probabilistic forecasts and Table 14 presents the *RPS*. Finally, Figures 24 to 29 present a case study for the hydrological year Sept. 2005 to Aug. 2006.

### 4.3.1 Deterministic forecasts

- Evaluation of the control run ($Q_{analysis}$) against observed discharge ($Q_{obs}$) (Table 7):

  First, a comparison between Tables 5, 6 and 7 indicates that in average, discharge simulations are of poorer quality in the period 2001–2006 than in previous years for both catchments, especially for vhm10. This quality varies dramatically from year to year, as judged by the Nash-Sutcliffe coefficient. A closer inspection (not shown) indicates that these variations are mainly related to poor simulations during specific periods of variable length, especially in winter, for a number of reasons. Baseflow in winter is sometimes overestimated, which could be related to both calibration and observational problems, while peak discharge is usually underestimated in winter both because of model calibration problems and uncertainties in the meteorological inputs. Discharge can be overestimated during some periods in spring in relation to uncertainties in snow accumulation and melt which could be both related to a calibration problem in the snow model or uncertainties in the meteorological observations.

- Evaluation of the deterministic flow forecasts ($Q_{DMO}$) against control run ($Q_{analysis}$) (Tables 8/9 and Figures 15/16):

  According to the Nash-Sutcliffe coefficient, deterministic forecasts derived from the ensemble mean are slightly better than those derived from the ensemble median or the best ensemble member. The mean error indicates that the ensemble mean and best ensemble member provide an unbiased deterministic prediction while the ensemble median leads to a slight underestimation. The scatter plots indicate that the largest flows (and in particular flood peaks) are usually underestimated by both the ensemble mean and median, while the best ensemble member is not systematically biased with that respect. Finally, forecast uncertainty increases with lead time, slightly more rapidly for the best ensemble member than for the ensemble mean or median.

- Evaluation of the deterministic flow forecasts ($Q_{DMO}$ and $Q_{corr}$ ) against observed discharge ($Q_{obs}$) (Tables 10 to 13 and Figures 17 to 20):

  Uncertainties in the hydrological model and in the measurements have significant effects on forecast skill. For the uncorrected predictions ($Q_{DMO}$), the ensemble median provides the best deterministic forecasts, followed by the ensemble mean and then the best ensemble member. Surprisingly, the predictions ($D > 0$) are of slightly better quality than the control run ($D = 0$) for both catchments, when the ensemble mean or median is used to define the deterministic forecast. When the best ensemble member is used though, the predictions are poorer than the control run. These results are partly related to the relative good quality of the predicted meteorological inputs compared to the analysed ones. As mentioned earlier, interpolation errors may affect the gridded data. Gridded precipitation in particular is not derived from measurements but from an orographic precipitation model. In flood situations, large precipitation errors can lead to very poor flow predictions. The uncertainty associated to the analyzed precipitation can be large in these situations and the use of an ensemble meteorological forecast handles these situations better than a deterministic analysis. These results indicate that the mean of an ensemble prediction leads to better results than a deterministic prediction. The use of the correction procedure (cf.

Table 7. Results of the verification of the control simulation (analysis) vs. observed discharge at gauges vhm10 and vhm26, over the period 01/09/2001–31/08/2006 ($Q_{analysis}$ vs. $Q_{obs}$).

| Catchment | vhm10 | | vhm26 | |
|---|---|---|---|---|
| Period / Statistics | ME ($m^3/s$) | Nash | ME ($m^3/s$) | Nash |
| 01/09/2001–31/08/2006 | 0.98 | 0.18 | 1.9 | 0.41 |
| 01/09/2001–31/08/2002 | 1.4 | 0.29 | 1.8 | 0.66 |
| 01/09/2002–31/08/2003 | 0.72 | 0.48 | 3.3 | –0.72 |
| 01/09/2003–31/08/2004 | 1.45 | 0 | 1.6 | 0.11 |
| 01/09/2004–31/08/2005 | 2.1 | –0.72 | 1.7 | 0.14 |
| 01/09/2005–31/08/2006 | –0.8 | 0.55 | 1.1 | 0.68 |

Section 2.4) greatly improves the forecast skills at all lead times by reducing the bias and scatter.

Finally, for both DMO and corrected forecasts, the forecast skill decreases with increasing lead time in relation to the degradation of the quality on the meteorological predictions with lead time.

Table 8. Results of the verification of the deterministic flow forecasts vs. control simulation at gauge vhm10, over the period 01/09/2001–31/08/2006 ($Q_{DMO}$ vs. $Q_{analysis}$).

| Forecast range | D=1 day | | D=2 days | | D=3 days | |
|---|---|---|---|---|---|---|
| Method / Statistics | ME ($m^3/s$) | Nash | ME ($m^3/s$) | Nash | ME ($m^3/s$) | Nash |
| Ensemble mean | 0 | 0.9 | –0.04 | 0.83 | –0.03 | 0.8 |
| Ensemble median | –0.38 | 0.88 | –0.4 | 0.82 | –0.36 | 0.79 |
| Best member | 0.07 | 0.8 | 0.03 | 0.71 | 0 | 0.65 |

Table 9. As Table 8 but at gauge vhm26.

| Forecast range | D=1 day | | D=2 days | | D=3 days | |
|---|---|---|---|---|---|---|
| Method / Statistics | ME ($m^3/s$) | Nash | ME ($m^3/s$) | Nash | ME ($m^3/s$) | Nash |
| Ensemble mean | –0.18 | 0.92 | –0.23 | 0.86 | –0.25 | 0.83 |
| Ensemble median | –0.65 | 0.91 | –0.64 | 0.86 | –0.61 | 0.82 |
| Best member | –0.03 | 0.87 | –0.09 | 0.79 | –0.2 | 0.73 |

Table 10. Results of the verification of the deterministic flow forecasts vs. observed discharge at gauge vhm10, over the period 01/09/2001–31/08/2006 ($Q_{DMO}$ vs. $Q_{obs}$).

| Forecast range | D=1 day | | D=2 days | | D=3 days | |
|---|---|---|---|---|---|---|
| Method / Statistics | ME ($m^3/s$) | Nash | ME ($m^3/s$) | Nash | ME ($m^3/s$) | Nash |
| Ensemble mean | 0.98 | 0.25 | 0.94 | 0.23 | 0.94 | 0.25 |
| Ensemble median | 0.59 | 0.28 | 0.58 | 0.26 | 0.62 | 0.27 |
| Best member | 1 | 0.08 | 1 | 0.06 | 0.98 | 0.04 |

Table 11. As Table 10 but at gauge vhm26.

| Forecast range | D=1 day | | D=2 days | | D=3 days | |
|---|---|---|---|---|---|---|
| Method / Statistics | ME ($m^3/s$) | Nash | ME ($m^3/s$) | Nash | ME ($m^3/s$) | Nash |
| Ensemble mean | 1.7 | 0.46 | 1.7 | 0.43 | 1.6 | 0.42 |
| Ensemble median | 1.2 | 0.48 | 1.3 | 0.45 | 1.3 | 0.43 |
| Best member | 1.9 | 0.35 | 1.8 | 0.33 | 1.7 | 0.32 |

Table 12. Results of the verification of the corrected deterministic flow forecasts vs. observed discharge at gauge vhm10, over the period 01/09/2001–31/08/2006 ($Q_{corr}$ vs. $Q_{obs}$).

| Forecast range | D=1 day | | D=2 days | | D=3 days | |
|---|---|---|---|---|---|---|
| Method / Statistics | ME ($m^3/s$) | Nash | ME ($m^3/s$) | Nash | ME ($m^3/s$) | Nash |
| Ensemble mean | 0 | 0.59 | –0.04 | 0.37 | –0.03 | 0.29 |
| Ensemble median | –0.39 | 0.57 | –0.4 | 0.35 | –0.36 | 0.27 |
| Best member | 0.06 | 0.41 | 0.03 | 0.18 | 0.01 | 0.08 |

Table 13. As Table 12 but at gauge vhm26.

| Forecast range | D=1 day | | D=2 days | | D=3 days | |
|---|---|---|---|---|---|---|
| Method / Statistics | ME ($m^3/s$) | Nash | ME ($m^3/s$) | Nash | ME ($m^3/s$) | Nash |
| Ensemble mean | –0.17 | 0.71 | –0.22 | 0.54 | –0.23 | 0.46 |
| Ensemble median | –0.64 | 0.7 | –0.63 | 0.54 | –0.6 | 0.44 |
| Best member | –0.02 | 0.63 | –0.08 | 0.45 | –0.18 | 0.36 |

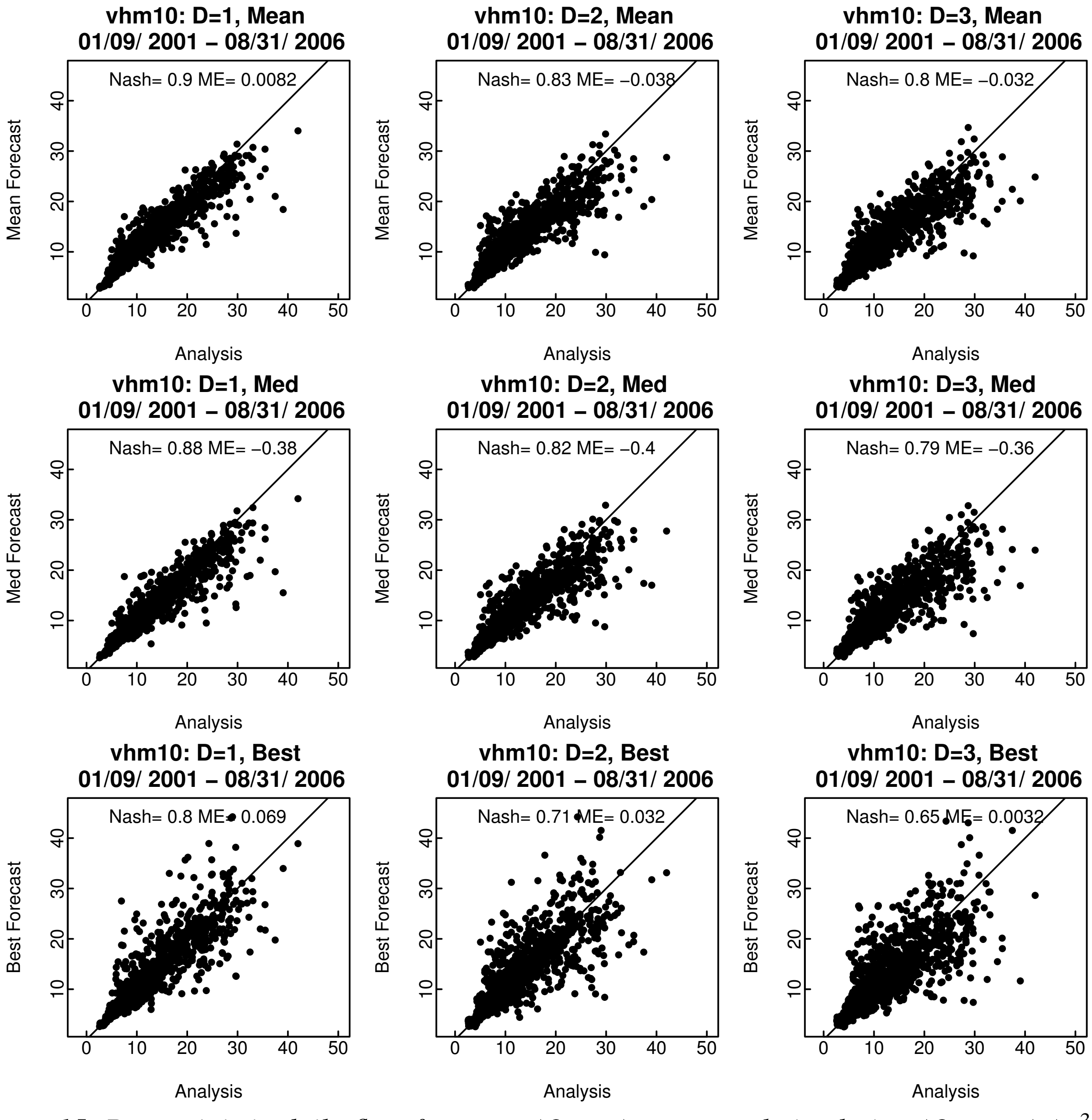



*Figure 15. Deterministic daily flow forecasts ($Q_{DMO}$) vs. control simulation ($Q_{analysis}$) ($m^3/s$) at gauge vhm10, and lead times of one to three days, considering the ensemble mean (top), median (middle) and best ensemble member (bottom). The 1:1 line represents a perfect match. Nash-Sutcliffe coefficient and mean error are also given.*

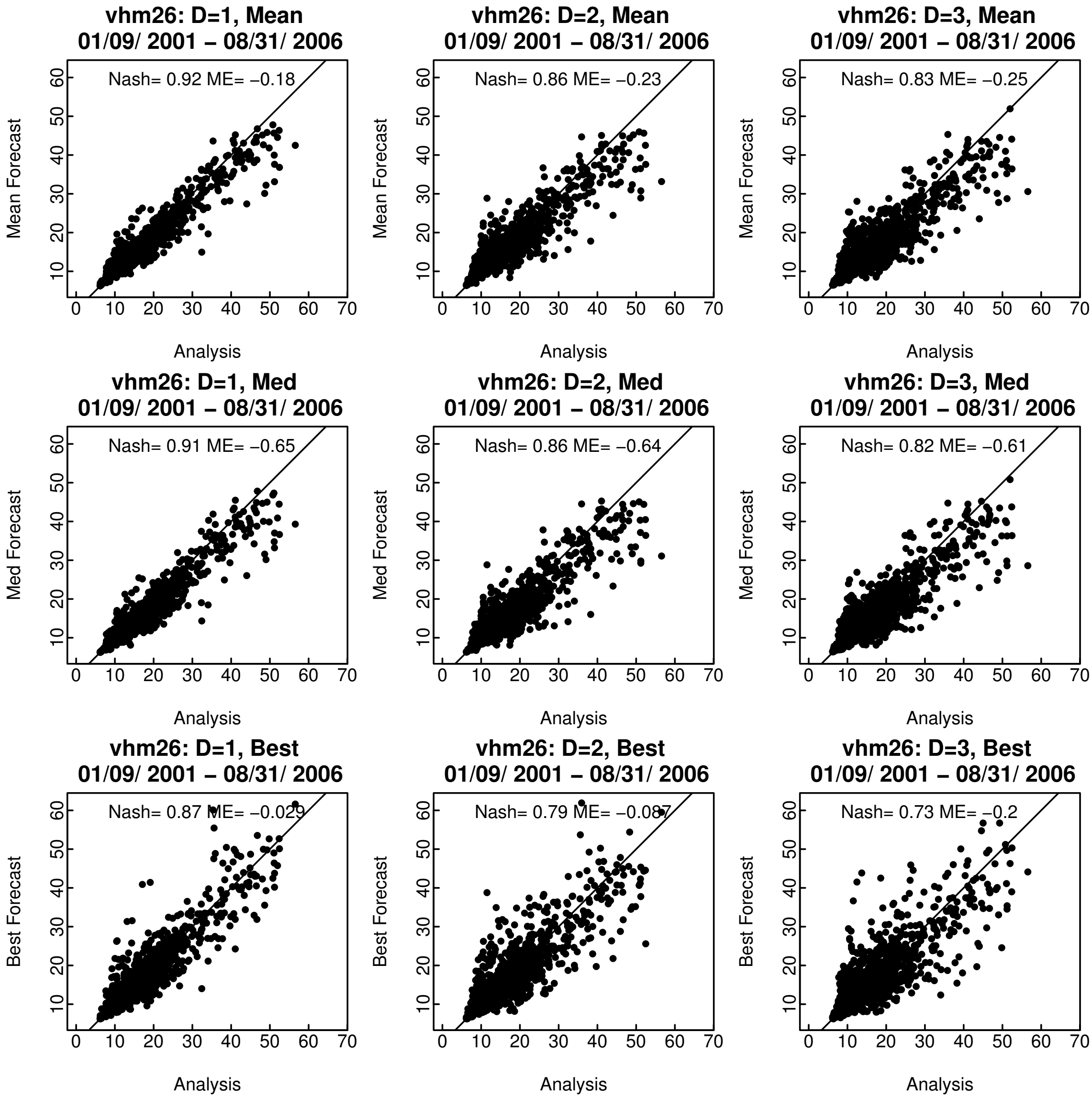



*Figure 16. As Fig. 15 but at gauge vhm26.*

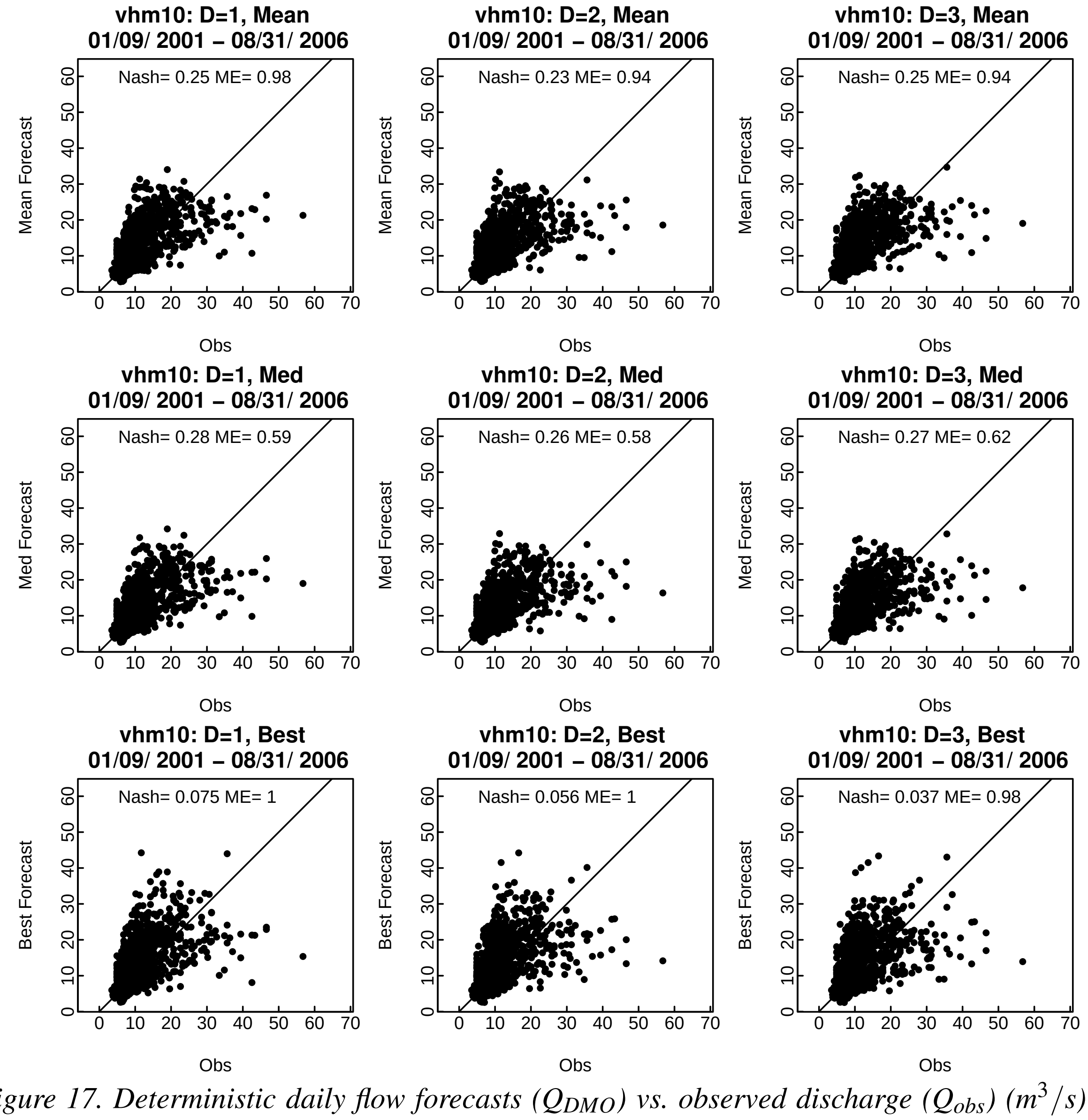



*Figure 17. Deterministic daily flow forecasts ($Q_{DMO}$) vs. observed discharge ($Q_{obs}$) ($m^3/s$) at gauge vhm10, and lead times of one to three days, considering the ensemble mean (top), median (middle) and best ensemble member (bottom). The 1:1 line represents a perfect match. Nash-Sutcliffe coefficient and mean error are also given.*

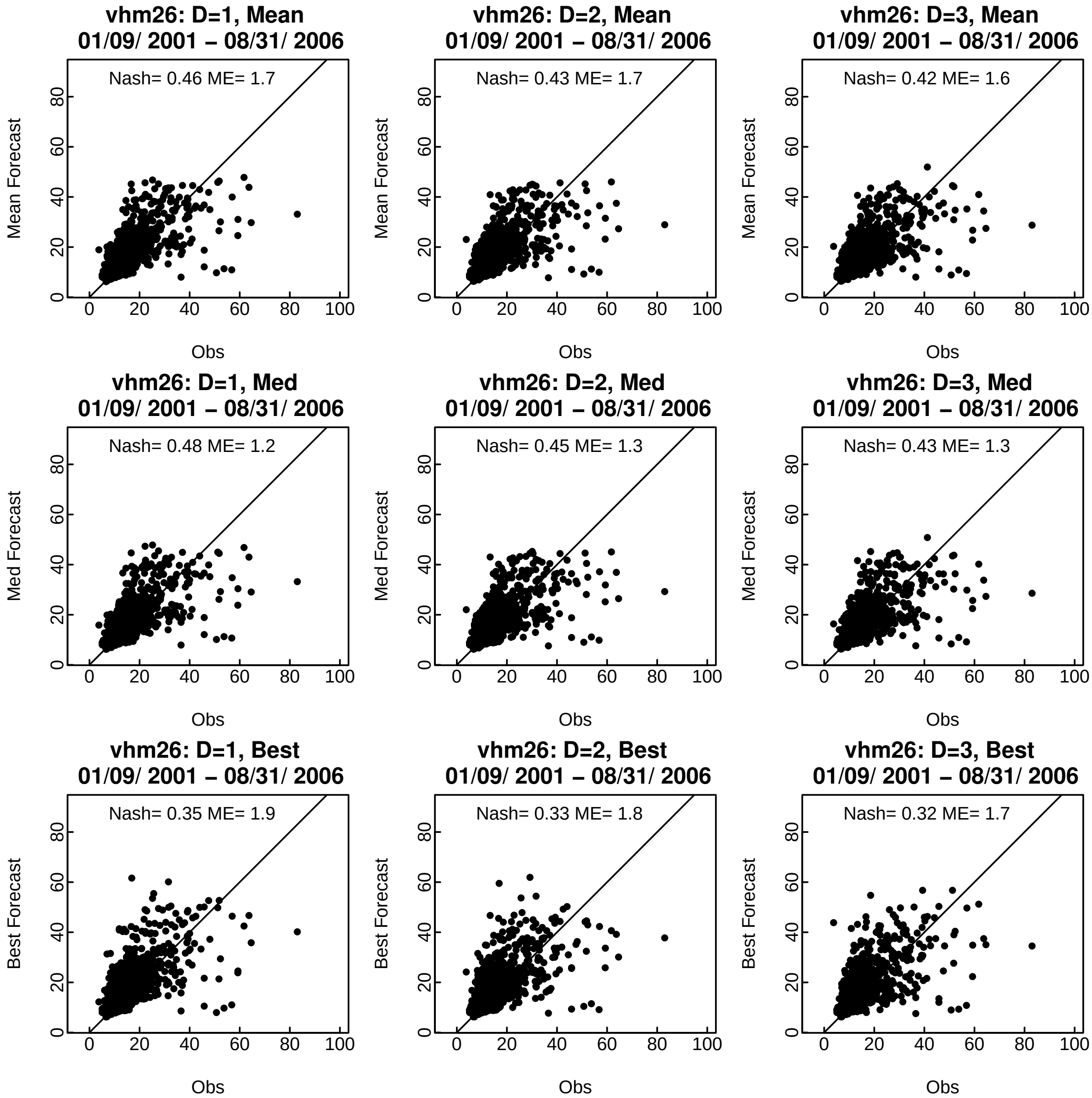



*Figure 18. As Fig. 17 but at gauge vhm26.*

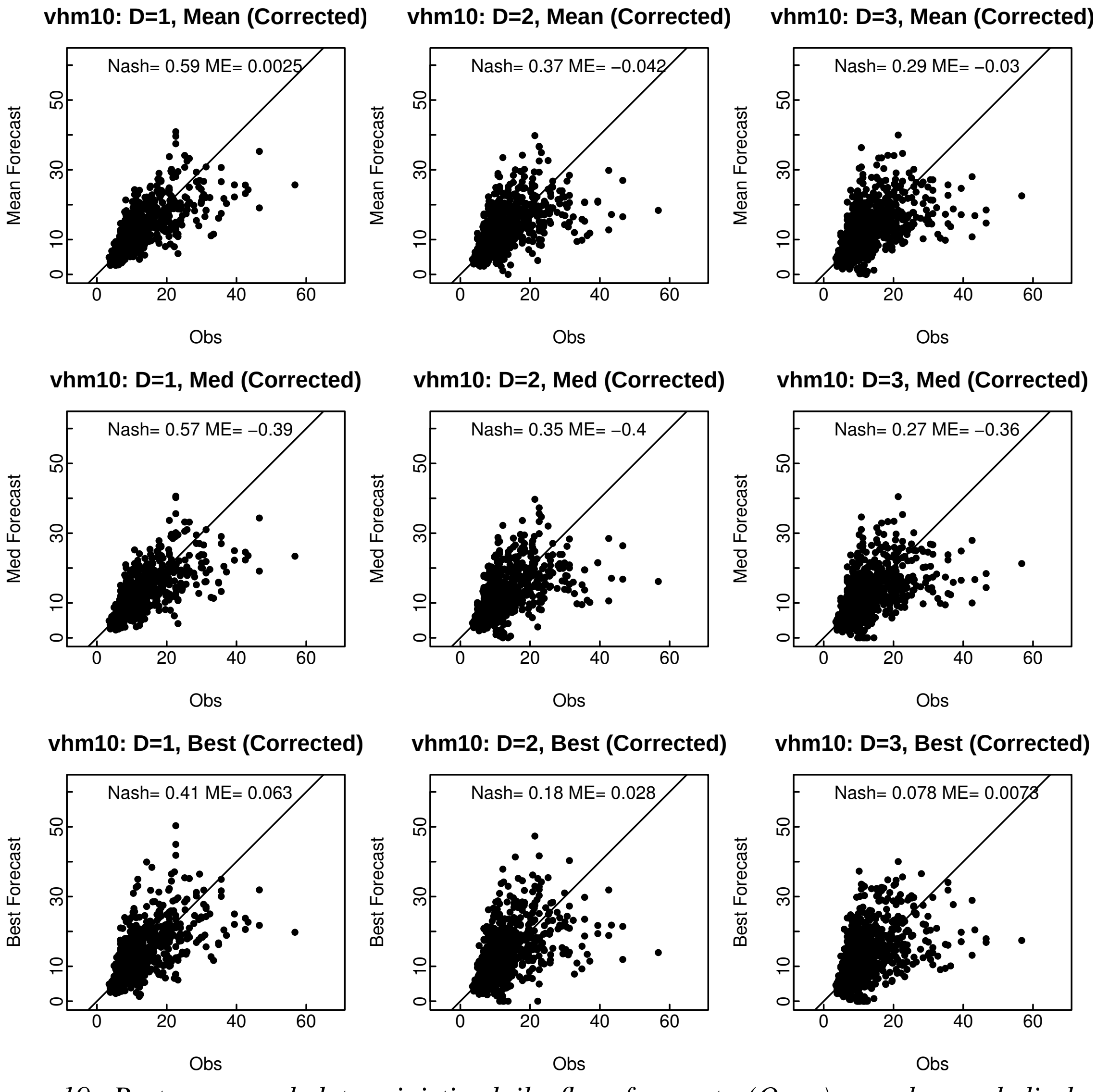



*Figure 19. Post-processed deterministic daily flow forecasts ($Q_{corr}$) vs. observed discharge ($Q_{obs}$) ($m^3/s$) at gauge vhm10, and lead times of one to three days, considering the ensemble mean (top), median (middle) and best ensemble member (bottom). The 1:1 line represents a perfect match. Nash-Sutcliffe coefficient and mean error are also given.*

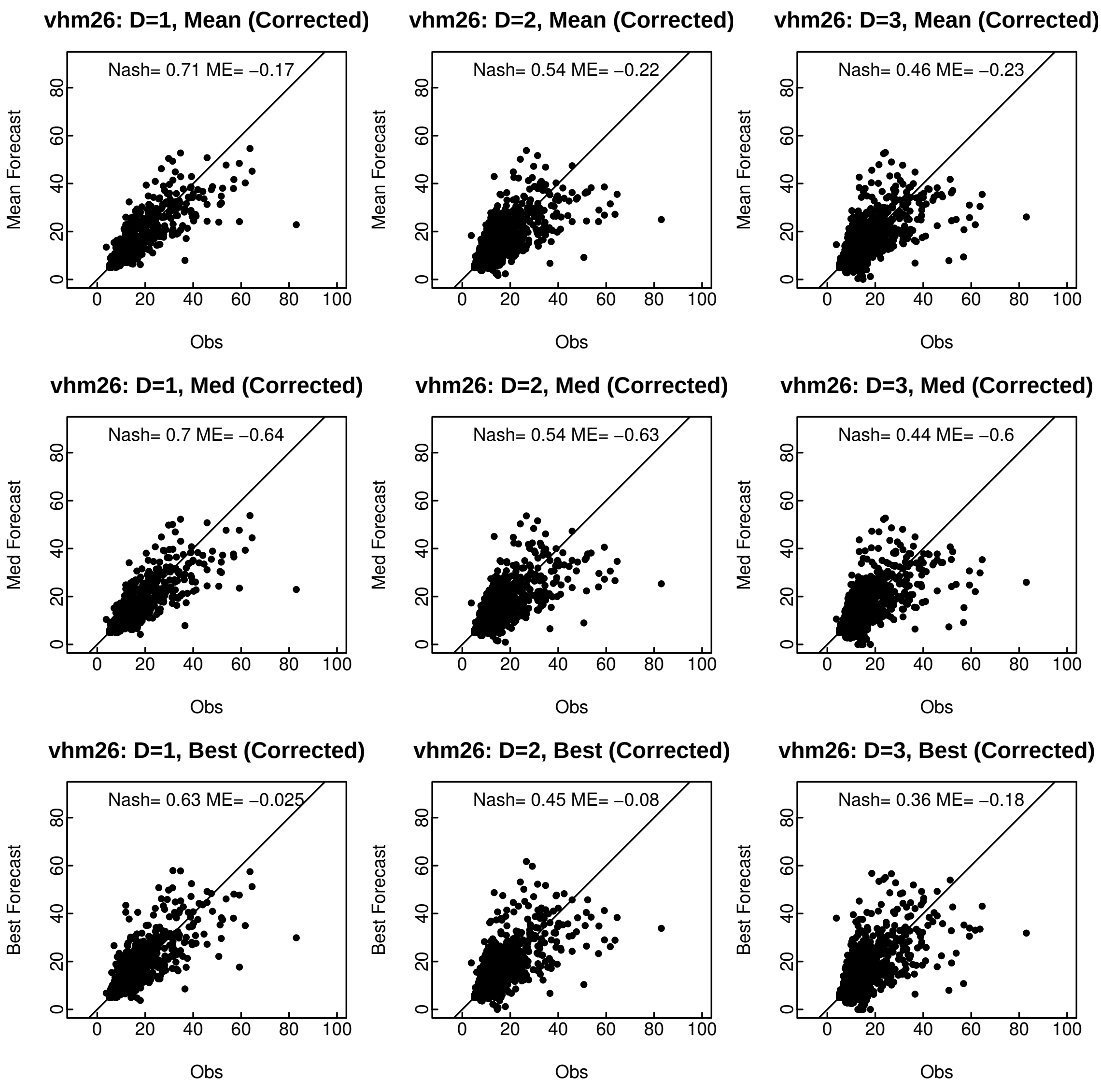



*Figure 20. As Fig. 19 but at gauge vhm26.*

Table 14. Probabilistic flow forecasts: rank probability score (*RPS*) at gauges vhm10 and vhm26, over the period 01/09/2001–31/08/2006.

| Catchment | vhm10 | | | vhm26 | | |
|---|---|---|---|---|---|---|
| Lead time | D=1 | D=2 | D=3 | D=1 | D=2 | D=3 |
| $Q_{DMO}$ vs. $Q_{analysis}$ | 0.27 | 0.45 | 0.56 | 0.22 | 0.34 | 0.44 |
| $Q_{DMO}$ vs. $Q_{obs}$ | 1.74 | 1.67 | 1.63 | 1.25 | 1.25 | 1.26 |
| $Q_{corr}$ vs. $Q_{obs}$ | 1 | 1.1 | 1.25 | 0.76 | 1.01 | 1.15 |

### 4.3.2 Probabilistic forecasts

- Evaluation of the probabilistic flow forecasts ($Q_{DMO}$) against control run ($Q_{analysis}$):

  The meteorological ensemble predictions (both temperature and precipitation) were found to be reliable and this reliability is properly transferred to the hydrological predictions, as judged by Fig. 21, where the evaluation is made against the control run. Although the prediction intervals are reliable for all lead times, the prediction uncertainty increases with lead time, as judged by the increasing value of the rank probability score (*RPS*), though the skill remains quite high (Table 14). These results indicate that the use of a sub-ensemble is acceptable (experimental $M_D \ll$ theoretical $M_D$). Note that the prediction interval calculated on a sub-ensemble made with the first 25 members only for all lead times ($M_1 = M_2 = M_3 = N$) is not reliable and strongly underestimated at $D$>1 day (not shown), indicating that it is necessary to define $M_D > N$ for $D$>1 day.

- Evaluation of the probabilistic flow forecasts ($Q_{DMO}$ and $Q_{corr}$ ) against observed discharge ($Q_{obs}$):

  The evaluation stresses the incapacity of the hydrological ensemble predictions ($Q_{DMO}$) to provide reliable flow prediction intervals (cf. Fig. 22). The strong underestimation of the prediction interval observed when comparing $Q_{DMO}$ to $Q_{obs}$ is partly related to the hydrological modelling uncertainty and partly due to the uncertainty of the observed discharge. It is observed that the underestimation is more pronounced in winter than in other seasons (not shown). The *RPS* remains relatively unchanged for all lead times.

  Assuming that the observed discharge is of reasonable quality, the correction procedure ($Q_{corr}$) greatly improves the reliability of the prediction interval without eliminating the problem of underestimation completely (cf. Fig. 23). The *RPS* is also improved but as expected, the uncertainty increases with lead time (Table 14).

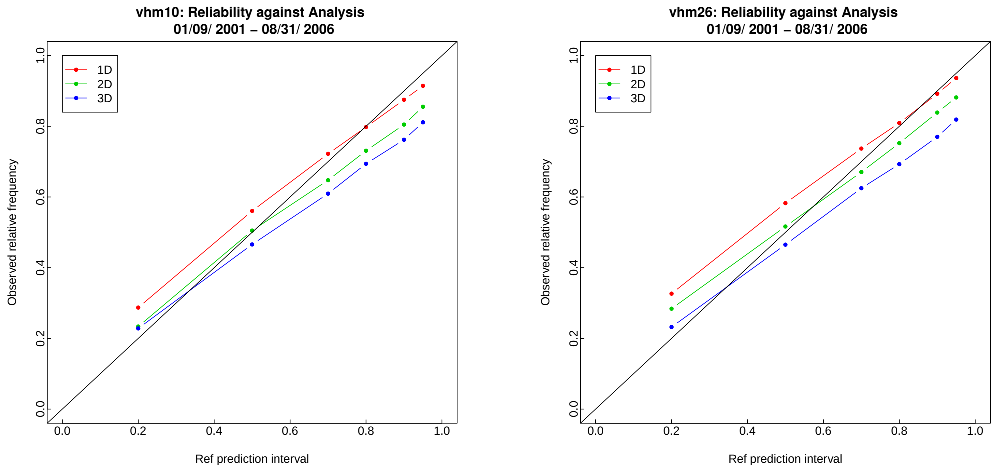



*Figure 21. Reliability diagram for the prediction interval ($Q_{DMO}$). Observed relative frequency vs. theoretical prediction interval. Verification against analysis ($Q_{analysis}$). Left-panel (vhm10), right-panel (vhm26). The 1:1 line represents a perfect match.*

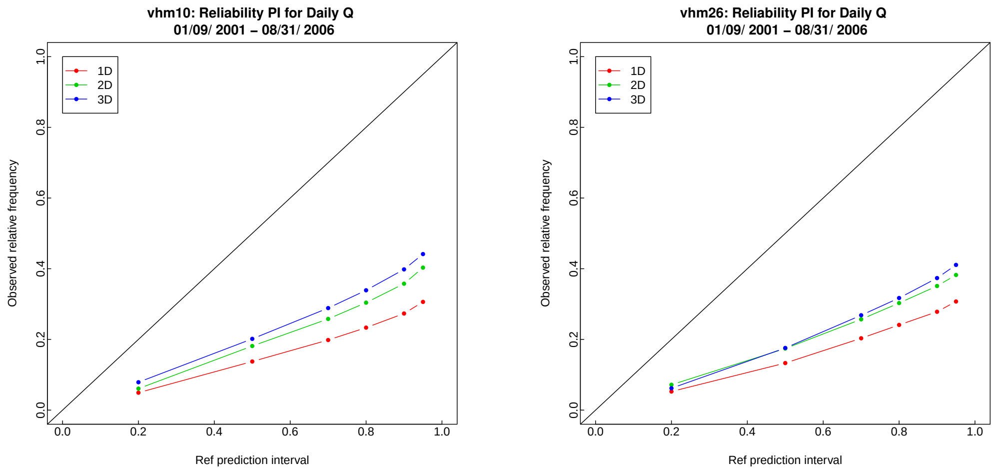



*Figure 22. As Fig. 21 but verification against observed discharge ($Q_{obs}$). Left-panel (vhm10), right-panel (vhm26).*

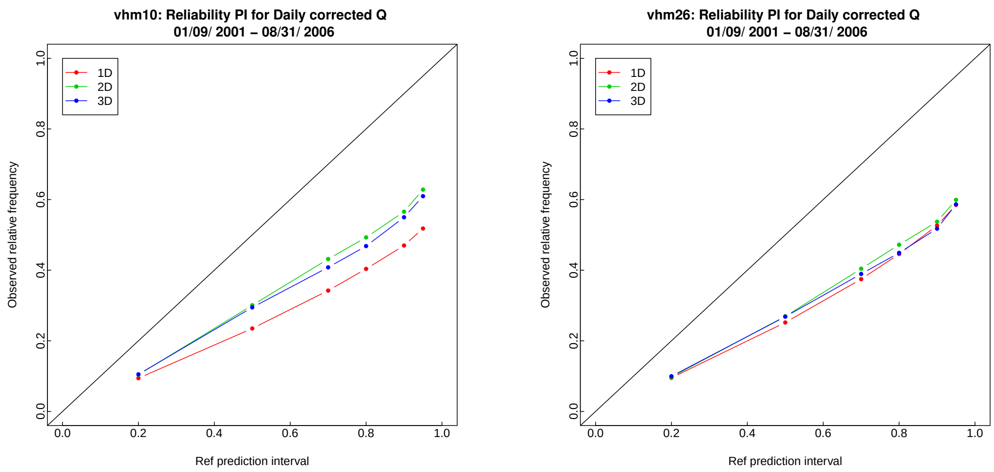



*Figure 23. Reliability diagram for the prediction interval after post-processing the ensemble prediction ($Q_{corr}$). Observed relative frequency vs. theoretical prediction interval. Verification against observed discharge ($Q_{obs}$). Left-panel (vhm10), right-panel (vhm26). The 1:1 line represents a perfect match.*

### 4.3.3 Case study

Figures 24 to 29 present a case study of flow forecast for the hydrological year Sept. 2005 to Aug. 2006. No major flood was observed in that period ($Q_{obs} < Q(T = 5years)$). The control runs simulate discharge with reasonable quality for both catchments (cf. Table 7). Winter flow peaks are usually underestimated while spring floods are well simulated. The correction procedure, improves the simulations in the control run. The occurrence and timing of all major hydrological events is usually well predicted two days in advance. The observed flow is usually located within the ensemble spread but the magnitude of flow peaks is often underestimated by the ensemble mean or median. The dispersion of this ensemble is related to the prediction uncertainty which increases during flood events. Some systematic errors observed in the control run, and related to model errors, propagate into the predictions, such as the flow underestimation in Sept. 2005 and June 2006 at gauge vhm10, and the flow overestimation in June 2006 at gauge vhm26. The correction procedures improves flow predictions by reducing the bias of the ensemble and deterministic forecasts.

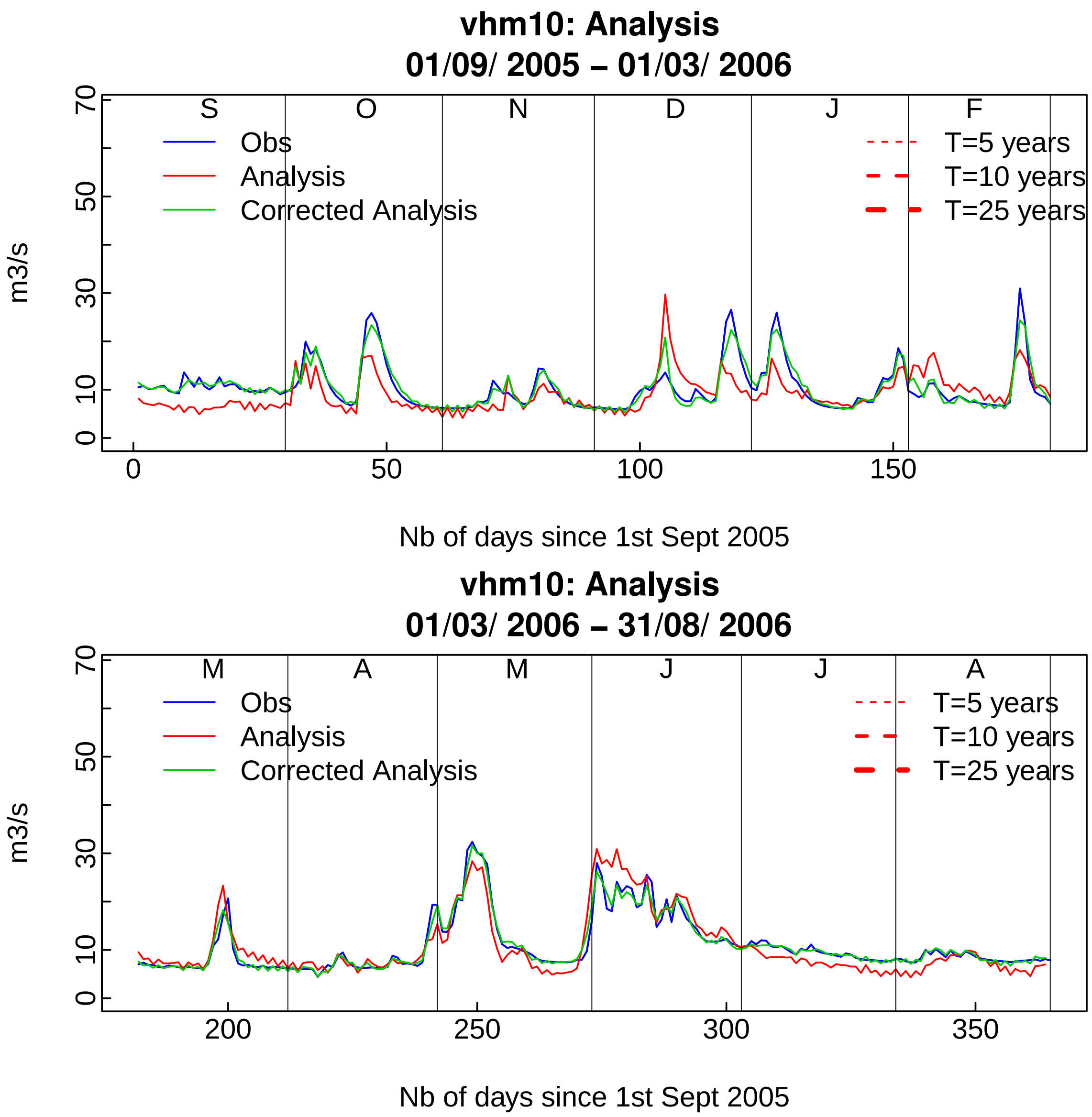



*Figure 24. Observed vs. simulated daily discharge in analysis mode ($Q_{analysis}$) at gauge vhm10, in the hydrological year Sept. 2005–Aug. 2006.*

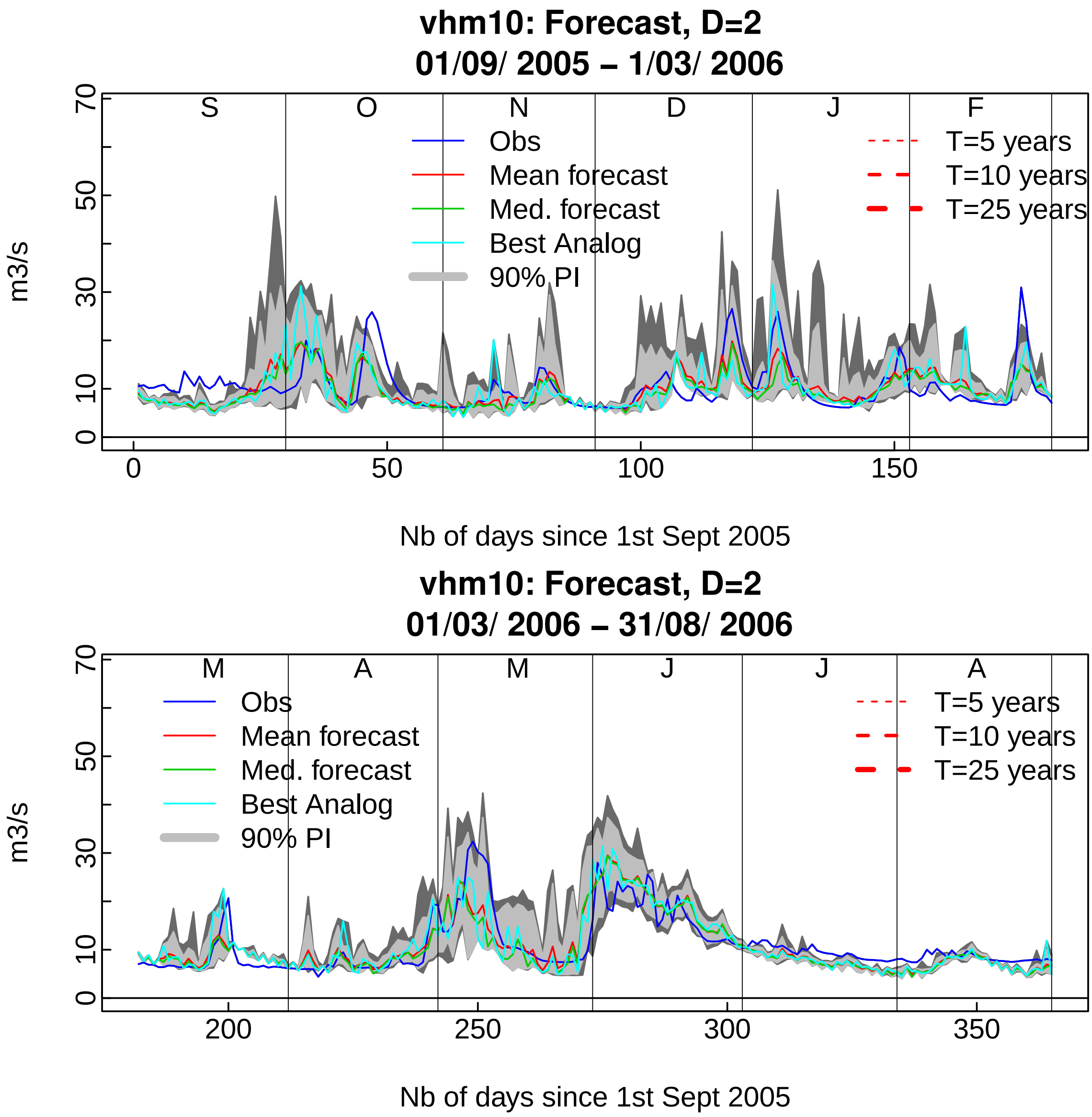



*Figure 25. Ensemble daily flow predictions ($Q_{DMO}$) at gauge vhm10 for a lead time of two days (D=2), in the hydrological year Sept. 2005–Aug. 2006. The dark grey band corresponds to the dispersion of the ensemble prediction and the light grey band represents the 90% prediction interval.*

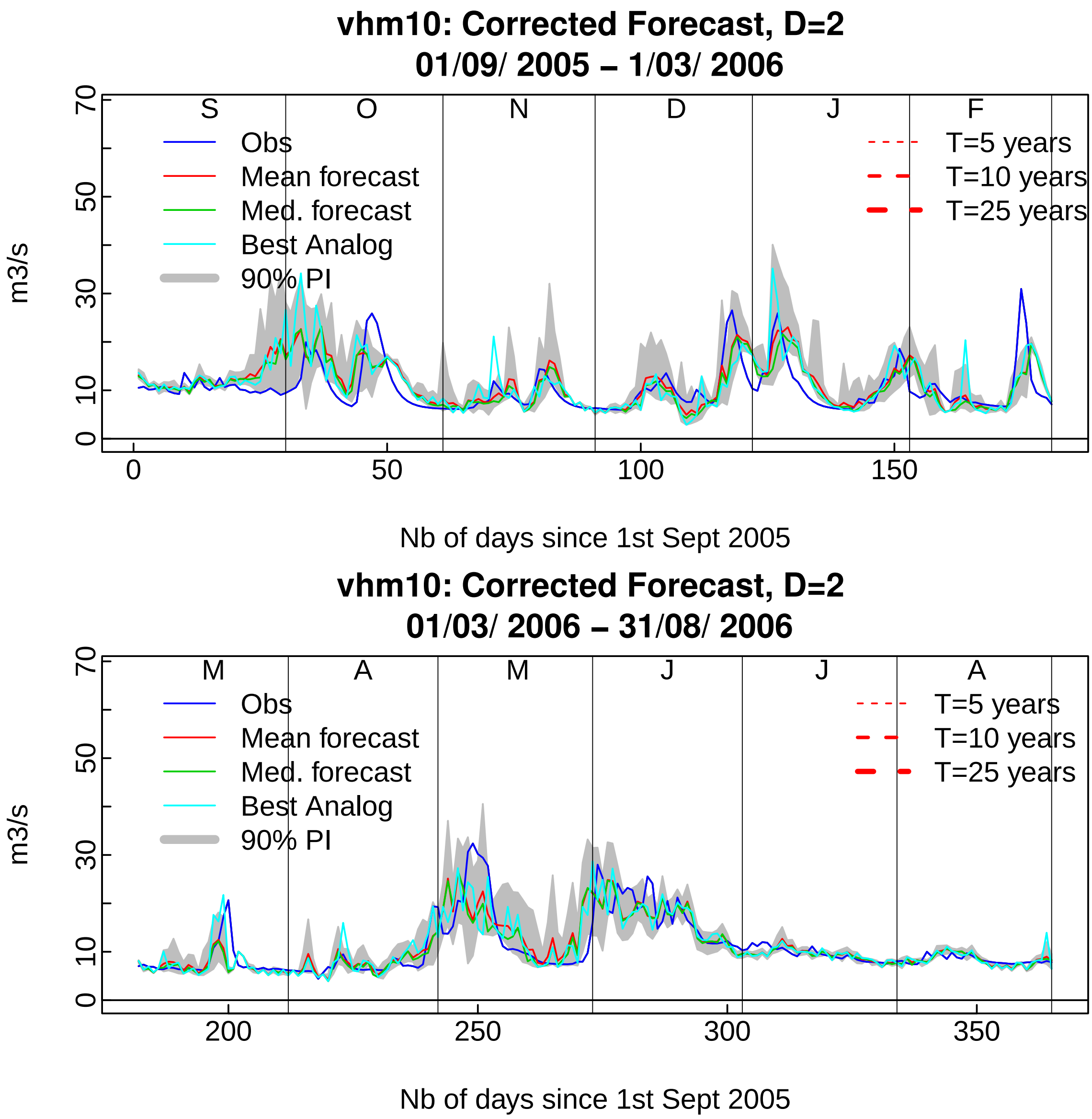



*Figure 26. As Fig. 25 after post-processing the ensemble predictions ($Q_{corr}$).*

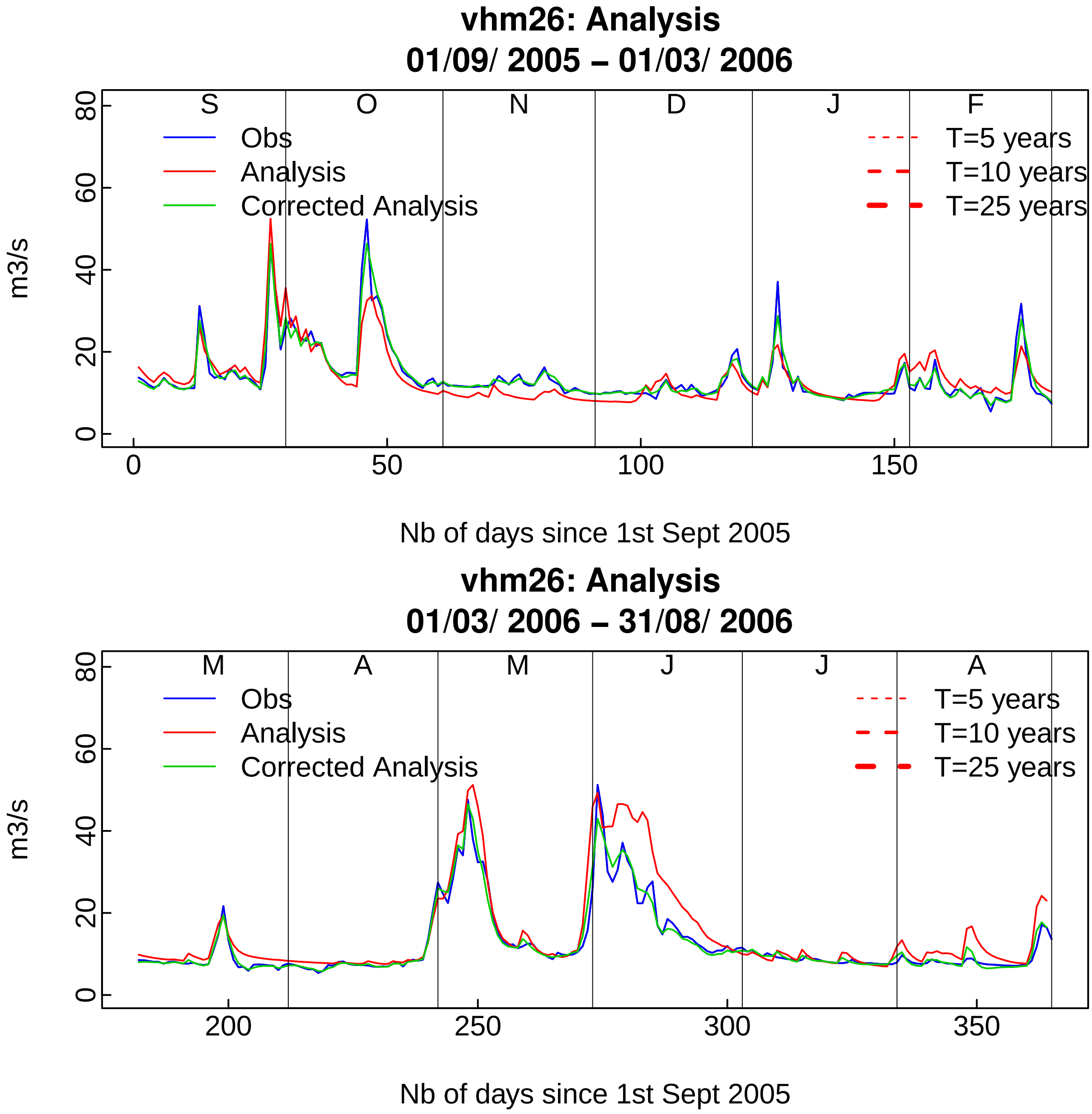



*Figure 27. Observed vs. simulated daily discharge in analysis mode ($Q_{analysis}$) at gauge vhm26, in the hydrological year Sept. 2005–Aug. 2006.*

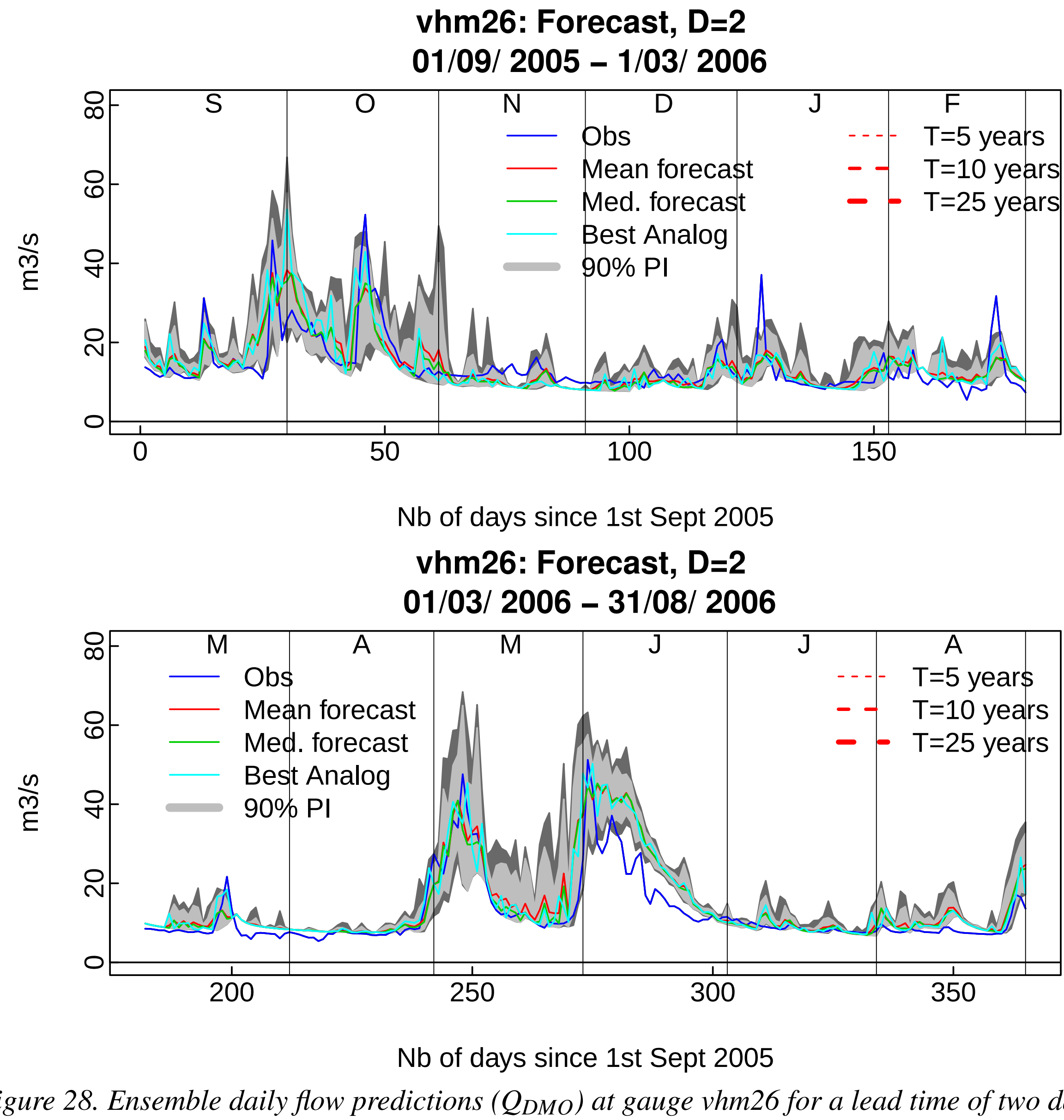



*Figure 28. Ensemble daily flow predictions ($Q_{DMO}$) at gauge vhm26 for a lead time of two days (D=2), in the hydrological year Sept. 2005–Aug. 2006. The dark grey band corresponds to the dispersion of the ensemble prediction and the light grey band represents the 90% prediction interval.*

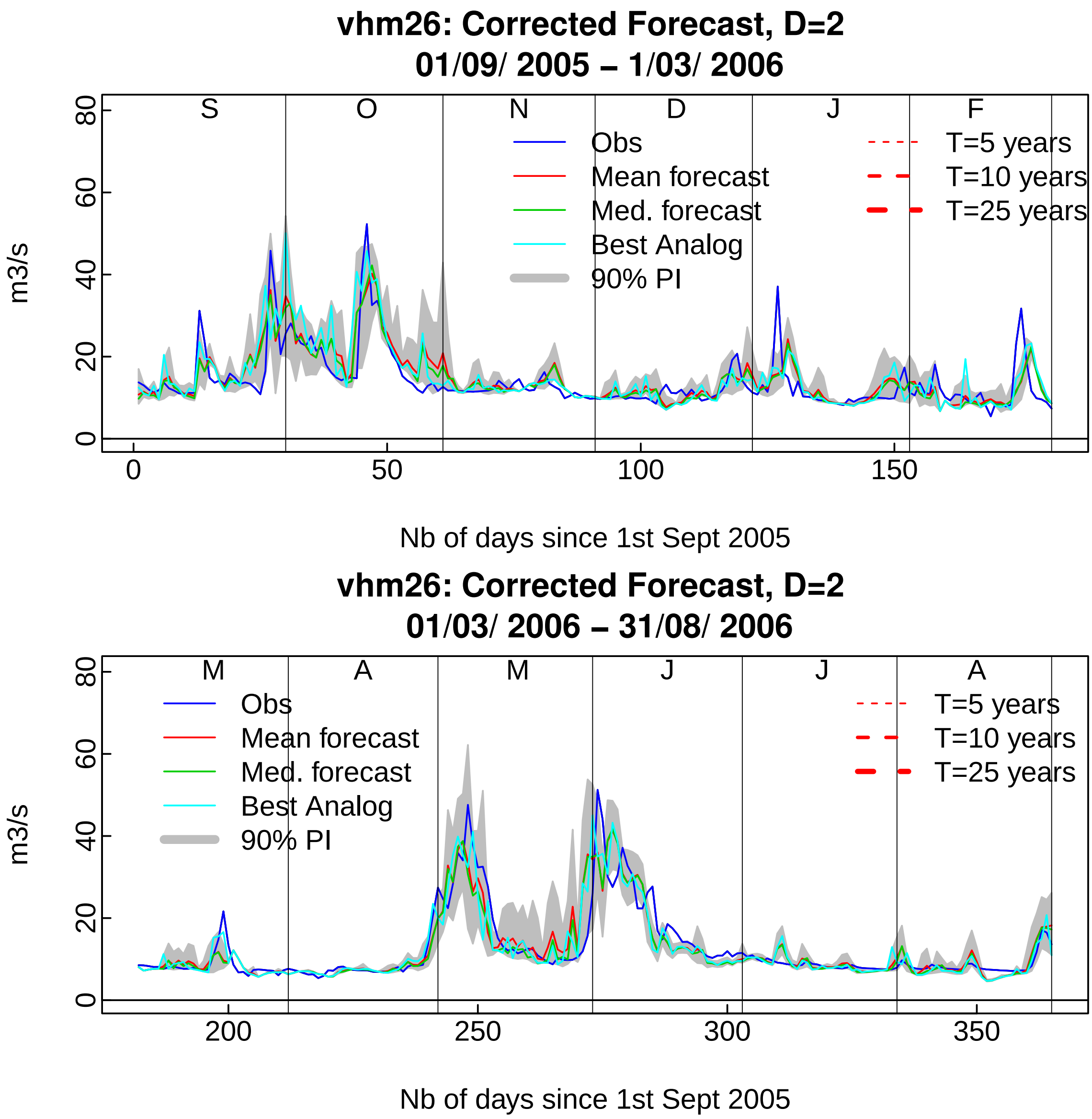



*Figure 29. As Fig. 28 after post-processing the ensemble predictions ($Q_{corr}$).*

# 5 Summary

A hydrological ensemble prediction system (HEPS) was proposed and evaluated on two river catchments located in Northern Iceland. The system is based on the coupling of the distributed hydrological model WaSiM-ETH with a meteorological ensemble prediction based on analogues. The principle of the analogue method is to search past days in a historical archive, most similar to a predicted synoptic meteorological situation and extract the associated precipitation and temperature observed on these days, at selected catchments, which are then used to force WaSiM-ETH.

Different algorithms were developed and tested off-line, considering various predictors and combination of predictors for the selection of weather analogues. The relationship between atmospheric circulation patterns in the vicinity of Iceland and local precipitation and temperature was demonstrated. This relationship was used to issue high-resolution short- and medium-range meteorological forecasts, given the synoptic meteorological situation predicted by ECMWF. The analogue method proposed here proved to be a useful, yet simple tool for producing probabilistic and deterministic meteorological forecasts. The method capitalises on historical information collected on the catchment and can therefore be seen as an objective expert system based on past knowledge and experience. In the present case, the interest of the method lies on an existing dataset of high-resolution gridded precipitation and temperature, available over more than 45 years, and developed in a previous research at IMO. This dataset will be extended up to present in the future, to allow the operational use of the method.

Daily temperature and precipitation predictions for lead times of up to three days were then used as input to WaSiM-ETH to produce an ensemble hydrological prediction. Ensemble hydrological predictions offer an added value compared to deterministic forecasts as both probabilistic and deterministic information can be extracted from the ensemble flow forecast. Another advantage of the method is that the same meteorological data used to calibrate/validate WaSiM-ETH are used to issue the meteorological forecasts, making the system consistent and robust, even in the presence of a bias in the precipitation and temperature data. Correcting the streamflow forecasts with observed discharge was observed to improve the forecast skills but the use of this procedure in real-time will be conditioned by the quality of the observed discharge. In practise, some rivers may be affected by icing conditions in winter, disrupting the water-level measurements and therefore the validity of the rating curves and the conversion into discharge.

An evaluation of the method for longer lead-times, up to ten days, should be considered in the future. More work is also needed to evaluate the skills of the forecasts in more details, for different discharge intervals and for flood forecast. The value of this method should be compared to a deterministic prediction obtained by coupling WaSiM-ETH with precipitation and temperature forecasts directly produced by NWP models. These two approaches should be treated as complementary. The uncertainty related to the hydrological model calibration was not included in the present system. In the future, the possibility to combine the ensemble weather forecasts with an ensemble of model parameterisations for WaSiM-ETH should be considered. Finally, further refinements of the analogue method itself could be investigated, such as a better optimisation of the analogy domain, the moving temporal window, the number of selected analogues and inclusion of other predictors.

# 6 Acknowledgements

This study was supported by the Landsvirkjun's Energy Research Fund. The author is grateful to Bolli Pálmason for preparing the ECMWF data used in this study and Tinna Þórarinsdóttir for her assistance with the setting of WaSiM-ETH for the studied catchments.

# 7 References


Addor, N., Jaun, S., Fundel, F., & Zappa, M. (2011). An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios. Hydrol. Earth Syst. Sci., 15, 2327–2347.

Akbari, M., Van Overloop, P. J. & Afshar, A. (2011). Clustered K nearest neighbour algorithm for daily inflow forecasting. Water Resour. Manage. 25, 1341–1357.

Bartholmes, J., & Todini, E. (2005). Coupling meteorological and hydrological models for flood forecasting. Hydrol. Earth Syst. Sci., , 9(4), 333–346.

Bergström, S., Lindström, G., & Pettersson, A. (2002). Multi-variable parameter estimation to increase confidence in hydrological modelling. Hydrol. Process. 16, 413–421.

Beven, K., & Freer, J. (2001). Equifinality, data assimilation and data uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol., 249, 11–29.

Boi, P. (2004). A statistical method for forecasting extreme daily temperatures using ECMWF 2-m temperatures and ground station measurements. Meteorol. Appl. 11, 245–251.

Crochet, P., Jóhannesson, T., Jónsson, T., Sigurðsson, O., Björnsson, H., Pálsson, F. & Barstad, I. (2007). Estimating the spatial distribution of precipitation in Iceland using a linear model of orographic precipitation. J. Hydrometeorol., 8, 1285–1306.

Crochet, P. & Jóhannesson, T. (2011). A dataset of daily temperature in Iceland for the period 1949–2010. Jökull, 61, 1–17.

Crochet, P. (2013). Probabilistic daily streamflow forecast using an analogue sorting method. Icelandic Met. Office report 2013–008,37pp.

Crochet, P. & Þórarinsdóttir, T. (2014). Flood frequency estimation for ungauged catchments in Iceland by combined hydrological modelling and regional frequency analysis. Icelandic Met. Office report, 2014–001, 40pp.

Diomede T., Nerozzi, F., Paccagnella, T. & Todini, E. (2008). The use of meteorological analogues to account for LAM QPF uncertainty. Hydrol. Earth Syst. Sci., 12, 141–157.

Einarsson, B. & Jónsson, S. (2010). The effect of climate change on runoff from two watersheds in Iceland. Icelandic Meteorol. Office, Rep. 2010-016. Available at http://www.vedur.is/-media/ces/2010_016.pdf

Engeland, K., Braud, I., Gottschalk, L., & Leblois, E. (2006). Multi-objective regional modelling. J. Hydrol. 327, 339–351.

Fraedrich K., Raible, C. C. & Sielmann, F. (2003). Analog ensemble forecasts of tropical cyclone tracks in the Australian region. Wea. Forecasting, 18, 3–11.

Galeati, G. (1990). A comparison of parametric and non-parametric methods for runoff forecasting. Hydrol. Sci. J., 35, 1, 79–94.

Gibergans-Báguena, & Llasat, M.C. (2007). Improvement of the analog forecasting method by using local thermodynamic data. Application to autumn precipitation in Catalonia. Atmos. Res., 86, 173–193.

Gupta, H.V., Sorooshian, S., & Yapo, P.O. (1998). Towards improved calibration of hydrologic models, multiple and non commensurable measures of information. Water Resour. Res. 34(4), 751–763.

Hamill, T. M. & Whitaker, J. S. (2006). Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application. Mon. Wea. Rev., 134, 3209–3229.

Jasper, K., Gurtz, J., & Lang, H. (2002). Advanced flood forecasting in Alpine watersheds by coupling meteorological observations and forecasts with distributed hydrological model. J. Hydrol., 267, 40–52.

Jóhannesson, T., Aðalgeirsdóttir, G., Björnsson, H., Crochet, P., Elíasson, E. B., Guðmundsson, S., Jónsdóttir, J. F., Ólafsson, H., Pálsson, F., Rögnvaldsson, Ó., Sigurðsson, O., Snorrason, Á., Blöndal Sveinsson, Ó. G. & Thorsteinsson Th. (2007). *Effect of climate change on hydrology and hydro-resources in Iceland.* National Energy Authority, Rep. OS-2007/011, 91 pp.

Jónsdóttir, J. F. (2008). A runoff map based on numerically simulated precipitation and a projection of future runoff in Iceland. Hydrol. Sci. J. 53(1), 100–111.

Jónsdóttir, J. F., Uvo, C. B. & Clarke, R. T. (2008). Trend analysis in Icelandic discharge, temperature and precipitation series by parametric methods. *Hydrology Res.* 39(5–6), 425–436.

Karlsson, M. & Yakowitz, S. (1987). Nearest-neighbor methods for non parametric rainfall-runoff forecasting. Water Resour. Res., 23 (7), 1308–1330.

Kruizinga, S. & Murphy, A.H. (1983). Use of an analogue procedure to formulate objective probabilistic temperature forecasts in the Netherlands. Mon. Wea. Rev., 111, 2244–2254.

Marty, R., Zin, I., Obled, C., Bontron, G. & Djerboua, A. (2012). Toward real-time daily PQPF by an analog sorting approach: application to flash-flood catchments. Journal of Applied Meteorology and Climatology, 51, 505–520.

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models I: A discussion of principles. J. Hydrol., 10, 282–290.

Obled, C., Bontron, G. & Garcon, R. (2002). Quantitative precipitation forecasts: a statistical adaptation of model outputs through an analogues sorting approach. Atmos. Res., 63, 303–324.

Peschke, G. (1987). Soil moisture and runoff components from a physically founded approach. Acta hydrophysica, 31 (3/4), 191–2005.

Raj Shrestha, R., & Rode, M. (2008). Multi-objective calibration and fuzzy preference selection of a distributed hydrological model. Environmental Modelling & Software, 23, 1384–1395.

Ramos, M.H., Mathevet, T., Thielen, J. & Pappenberger, F. (2010). Communicating uncertainty in hydro-meteorological forecasts: mission impossible? Meteorol. Appl. 17, 223–235.

Radinović, D. (1975). An analogue method for weather forecasting using the 500/1000 mb relative topography. Mon. Wea. Rev., 103, 639–649.

Renner, M., Werner, M.G.F., Rademacher, S., & Sprokkereef, E. (2009). Verification of ensemble flow forecasts for the river Rhine. J. Hydrol. 376, 463–475.

Rist, S. (1990). *Vatns er þörf.* Bókaútgáfa Menningarsjóðs, 248 pp. (In Icelandic).

Schulla, J. & Jasper, K. (2007). Model Description WaSiM-ETH. Technical report, Zürich, 181 pp. Available at http://www.wasim.ch

Smiatek, G., Kunstmann, H., & Werhahn, J. (2012). Implementation and performance analysis of a high resolution coupled numerical weather and river runoff prediction model system for the Alpine catchment. Environnemental Modelling & Software, 38, 231–243.

Snorrason, Á., Einarsson, B., Pagneux, E., Harðardóttir, J., Roberts, M. J., Sigurðsson, O., Þórarinsson, Ó., Crochet, P., Jóhannesson, T. & Þorsteinsson, T. (2012). Floods in Iceland (Ch. 13). In: "Changes in flood risk in Europe". IAHS Special publication 10. Edited by Z.W. Kundzewicz.

Uppala, S.M. & co-authors (2005). The ERA-40 re-analysis. Quart. J. Roy. Meteor. Soc., 131, 2961–3012.

Van den Dool, H. M. (1989). A new look at weather forecasting through analogues. Mon. Wea. Rev., 117, 2230–2247.

Van Genuchten, M.T. (1976). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sciences Society of American Journal, 44(5), 892–898.

Wetterhall, F., Halldin, S. & Xu, C.Y. (2005). Statistical precipitation downscaling in central Sweden with the analogue method. Journal of Hydrology, 306, 174–190.

Wilks, D. S. (1995). Statistical methods in the atmospheric sciences. Academic Press, San Diego, California, 467 pp.

Woodcock, F. (1980). On the use of analogues to improve regression forecasts. Mon. Wea. Rev., 108, 292–297.

Yapo, P.M., Gupta, H.V., & Sorooshian, S. (1998). Multi-objective global optimisation for hydrologic models. J. Hydrol., 204, 83–87.

Zorita, E. & von Storch, H. (1999). The analog method as a simple downscaling technique: Comparison with more complicated methods. J. Climate, 12, 2474–2489.

Þórarinsdóttir, T. (2012). Development of a methodology for estimation of technical hydropower potential in Iceland using high resolution hydrological modelling. University of Iceland. 106 pp.

## Appendix I

## Observed versus deterministic temperature forecasts over the period 01/09/2001–31/08/2006.

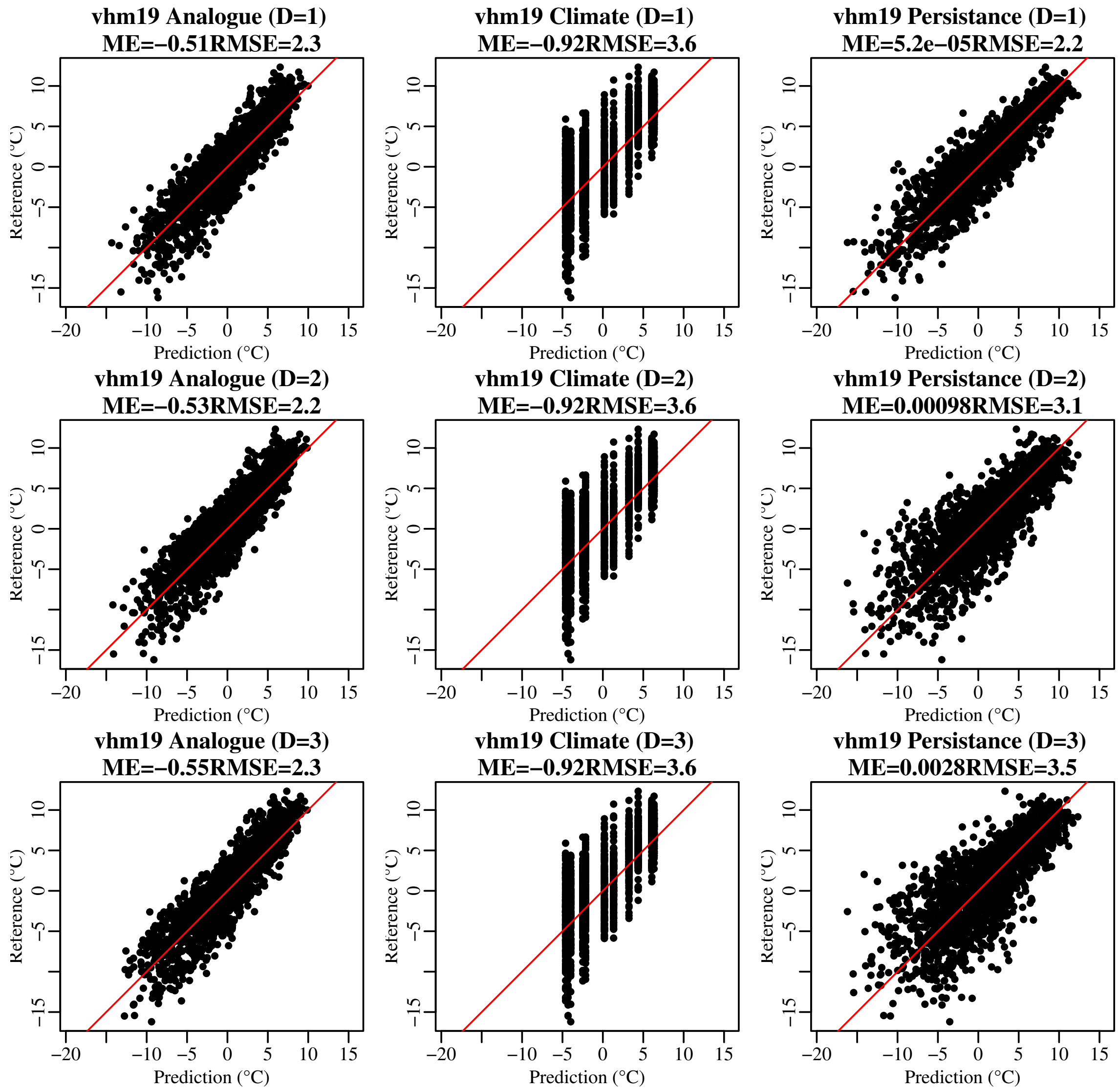



*Figure I.1 Observed versus deterministic temperature forecasts for vhm19. Top (D=1 day), centre (D=2 days), bottom (D=3 days). Analogue method (left), climate (centre) and persistence (right). Solid 1:1 line represents a perfect match.*

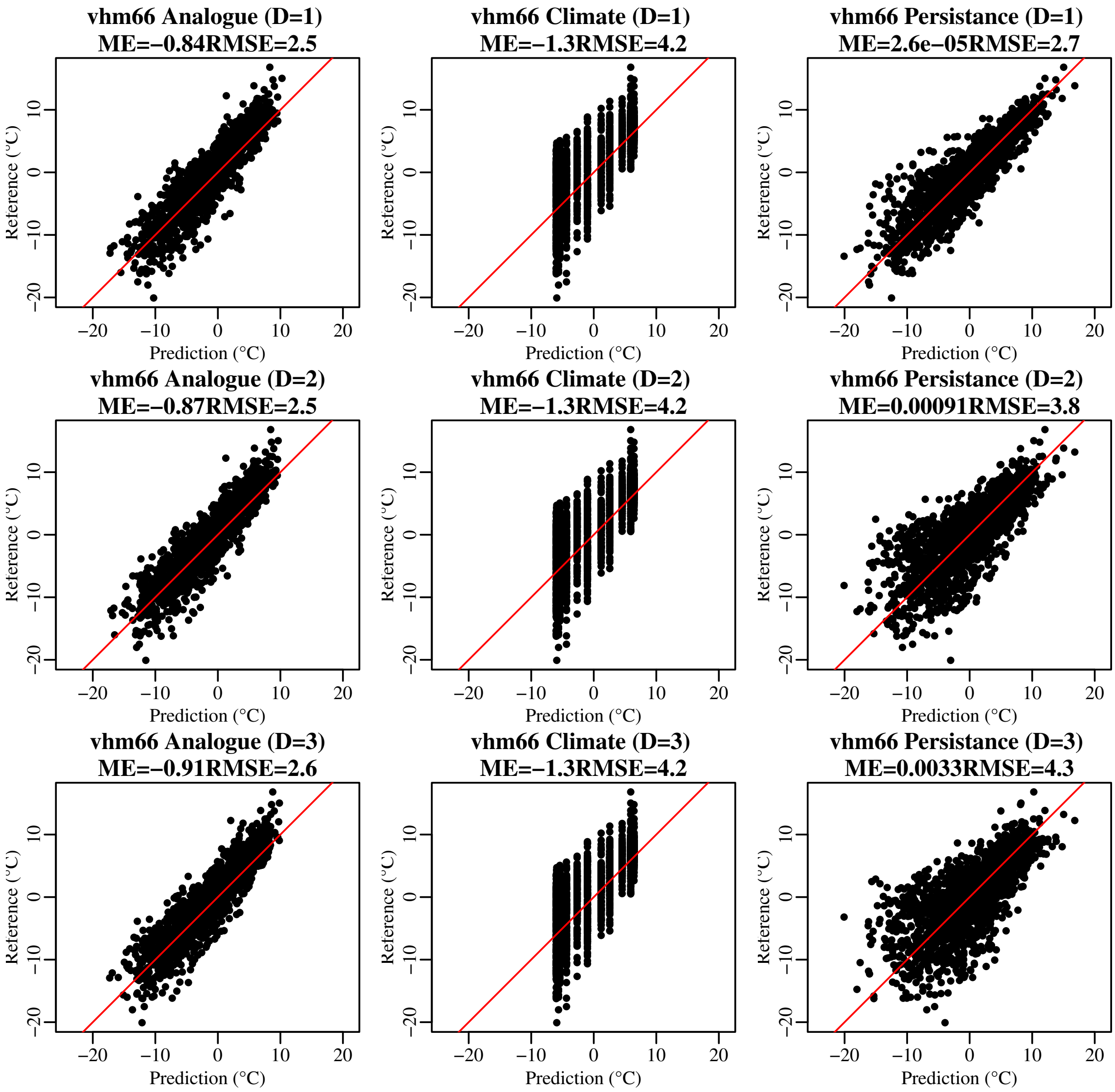



*Figure I.2 Observed versus deterministic temperature forecasts for vhm66. Top (D=1 day), centre (D=2 days), bottom (D=3 days). Analogue method (left), climate (centre) and persistence (right). Solid 1:1 line represents a perfect match.*

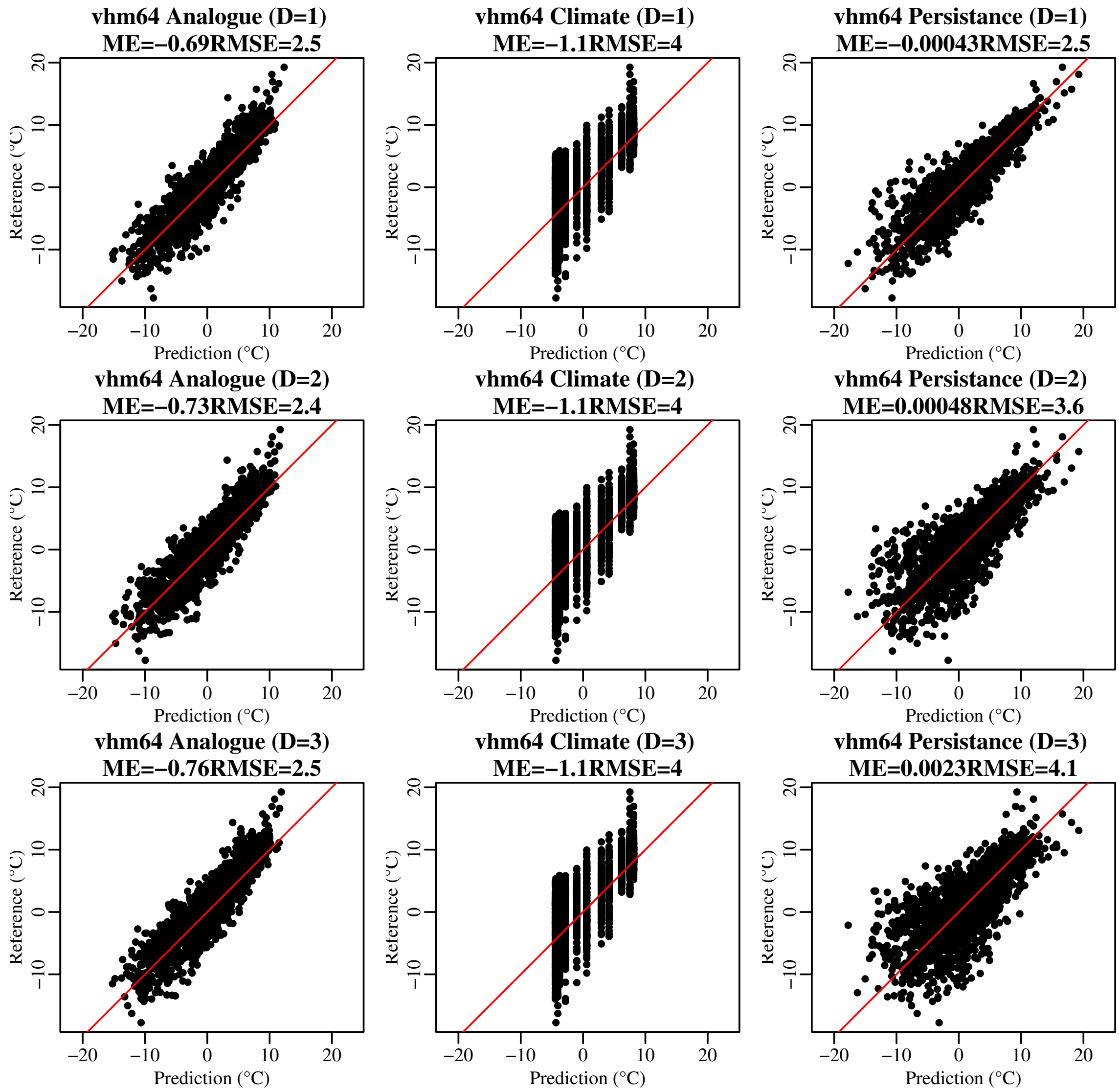



*Figure I.3 Observed versus deterministic temperature forecasts for vhm64. Top (D=1 day), centre (D=2 days), bottom (D=3 days). Analogue method (left), climate (centre) and persistence (right). Solid 1:1 line represents a perfect match.*

## Appendix II

## Observed versus deterministic precipitation forecasts over the period 01/09/2001–31/08/2006.

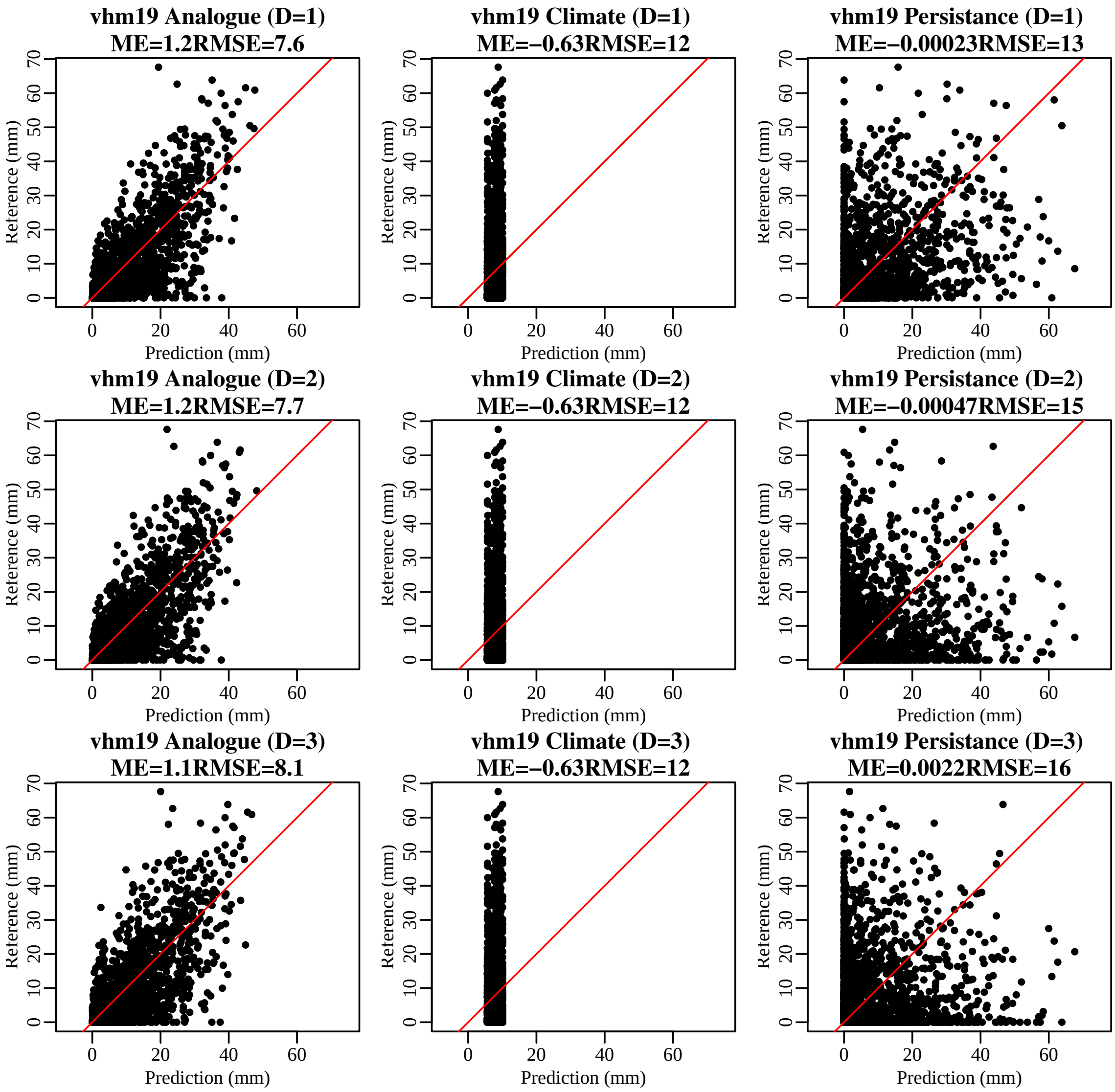



*Figure II.1 Observed versus deterministic precipitation forecasts for vhm19. Top (D=1 day), centre (D=2 days), bottom (D=3 days). Analogue method (left), climate (centre) and persistence (right). Solid 1:1 line represents a perfect match.*

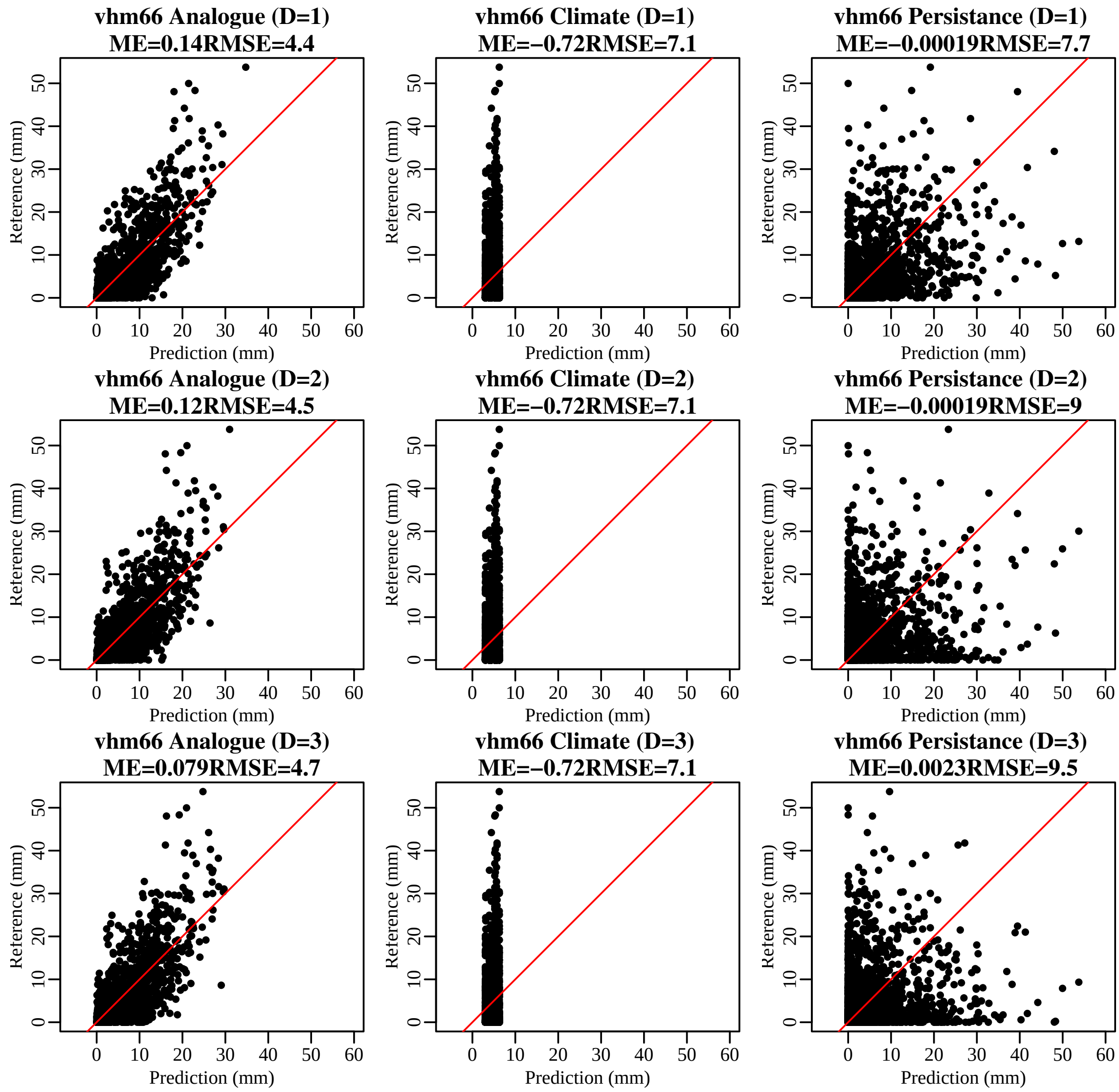



*Figure II.2 Observed versus deterministic precipitation forecasts for vhm66. Top (D=1 day), centre (D=2 days), bottom (D=3 days). Analogue method (left), climate (centre) and persistence (right). Solid 1:1 line represents a perfect match.*

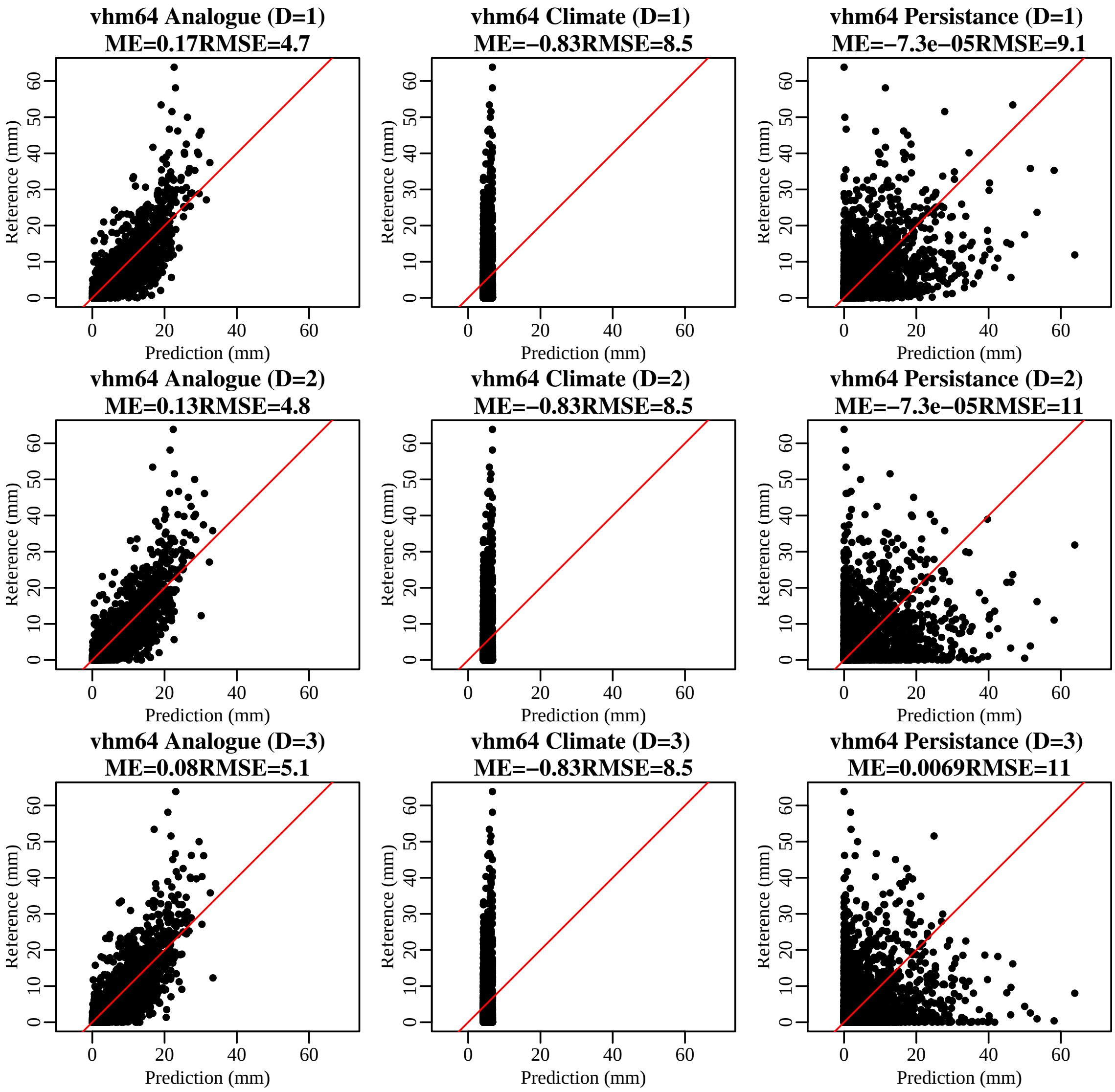



*Figure II.3 Observed versus deterministic precipitation forecasts for vhm64. Top (D=1 day), centre (D=2 days), bottom (D=3 days). Analogue method (left), climate (centre) and persistence (right). Solid 1:1 line represents a perfect match.*

## Appendix III

## Reliability diagrams for the prediction intervals of temperature and precipitation over the period 01/09/2001–31/08/2006.

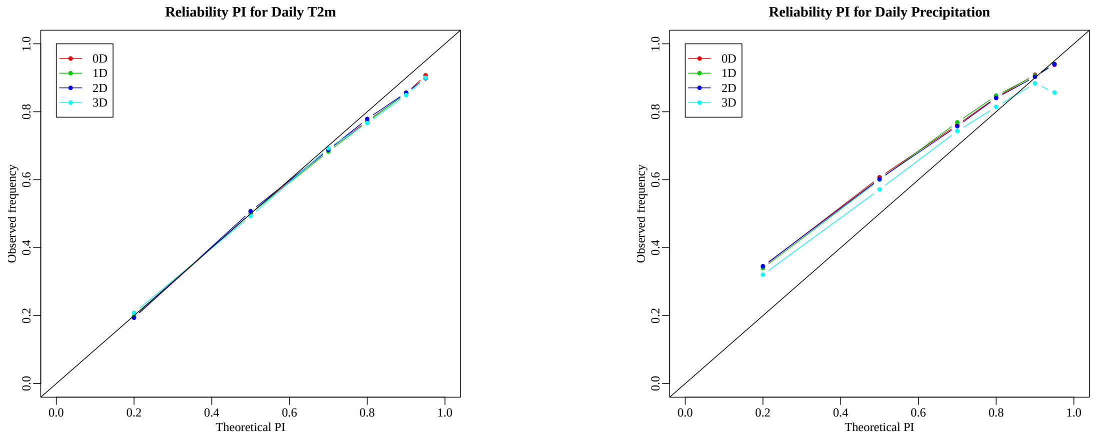



*Figure III.1 Reliability diagrams for temperature (left) and precipitation (right) forecasts for vhm19: Observed relative frequency vs. theoretical prediction interval.*

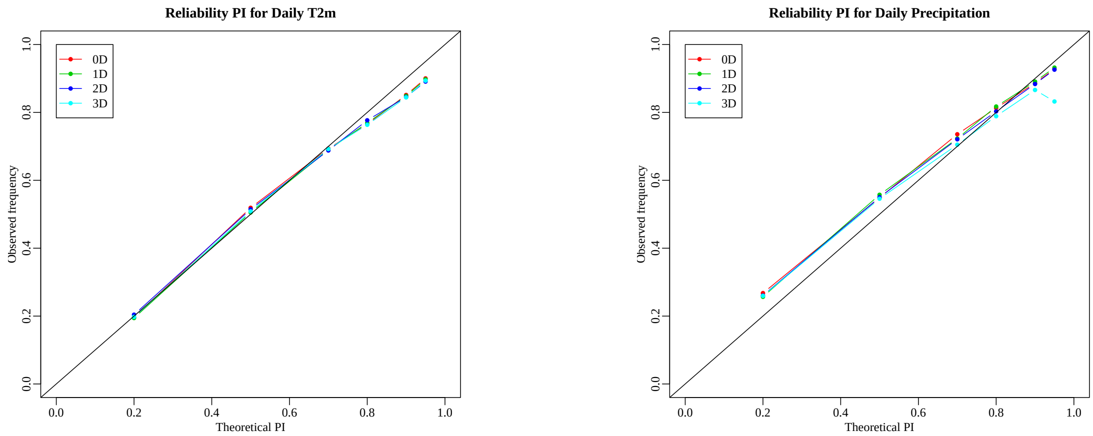



*Figure III.2 As Fig. III.1 but for vhm66.*

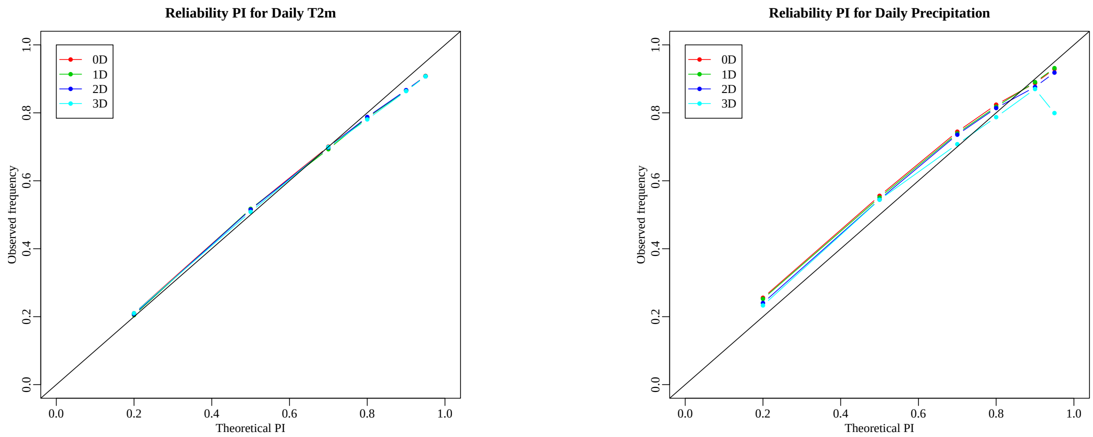



*Figure III.3 As Fig. III.1 but for vhm64.*