Hydrology and
Earth System
Sciences

Streamflow forecasts from WRF precipitation for flood early warning in mountain tropical areas

María Carolina Rogelis[1] **and Micha Werner**[1,2]

[1]UNESCO-IHE, P.O. Box 3015, 2601 DA Delft, the Netherlands
[2]Deltares, P.O. Box 177, 2600 MH Delft, the Netherlands

Correspondence: Maria Carolina Rogelis (c.rogelisprada@unesco-ihe.org)

Abstract. Numerical weather prediction (NWP) models are fundamental to extend forecast lead times beyond the concentration time of a watershed. Particularly for flash flood forecasting in tropical mountainous watersheds, forecast precipitation is required to provide timely warnings. This paper aims to assess the potential of NWP for flood early warning purposes, and the possible improvement that bias correction can provide, in a tropical mountainous area. The paper focuses on the comparison of streamflows obtained from the post-processed precipitation forecasts, particularly the comparison of ensemble forecasts and their potential in providing skilful flood forecasts. The Weather Research and Forecasting (WRF) model is used to produce precipitation forecasts that are post-processed and used to drive a hydrologic model. Discharge forecasts obtained from the hydrological model are used to assess the skill of the WRF model. The results show that post-processed WRF precipitation adds value to the flood early warning system when compared to zero-precipitation forecasts, although the precipitation forecast used in this analysis showed little added value when compared to climatology. However, the reduction of biases obtained from the post-processed ensembles show the potential of this method and model to provide usable precipitation forecasts in tropical mountainous watersheds. The need for more detailed evaluation of the WRF model in the study area is highlighted, particularly the identification of the most suitable parameterisation, due to the inability of the model to adequately represent the convective precipitation found in the study area.

1 Introduction

Numerical weather prediction (NWP) models are fundamental to extend flood forecast lead times beyond the concentration time of a watershed. The significant advances in NWP and computer power during the last decades have led to the generation of high-resolution precipitation forecasts at the catchment scale, and quantitative precipitation forecasts (QPFs) from high-resolution NWPs are increasingly used in flood forecasting systems as a result (Xuan et al., 2009). Particularly for flashy watersheds with short times of concentration, such as those typically found in tropical mountainous watersheds, forecast precipitation is required to provide sufficient forecast lead times in support of timely warnings.

Despite the significant advances, NWP results contain noise, are contaminated by model biases, are too coarse to adequately resolve all features such as convection, and are influenced by uncertainty inherent in the initial conditions (Colman et al., 2013). Furthermore, weather forecasting in tropical mountains is highly challenging. In the tropics, local and mesoscale effects are more dominant than synoptic influences (except for tropical cyclones). In addition, there are limitations in the availability of surface and upper air monitoring networks (Laing and Evans, 2010), which influences modelling initialisation (Cuo et al., 2011). Additionally, the formation and movement of deep convection and mesoscale convective systems is affected by orography (Colman et al., 2013).

According to Habets et al. (2004) the potential of NWP precipitation forecast to be used by hydrological models for flood forecasting is mainly affected by (i) localisation of the events, since an error of a few kilometres can lead to the precipitation being forecast in the wrong watershed, (ii) tim-

ing of the events, since the response of the basin depends on previous events and on the timing of the present event, and (iii) precipitation intensity. This is especially true in flash-flood-prone watersheds, typical of tropical mountainous areas where location errors that are considered small at meteorological level can lead to a flood event in a watershed being completely missed (Vincendon et al., 2011).

In order to address the uncertain nature of NWP, ensemble prediction systems (EPSs) have been developed (Demeritt et al., 2007). In contrast to deterministic systems that produce one prediction, EPSs produce a suite or ensemble of predictions to reflect uncertainty, providing the capability to transform predictions into a probability distribution function (Leutbecher and Palmer, 2008; Demeritt et al., 2007). As EPS rainfall predictions often exhibit greater skill than deterministic predictions, the hope is then that EPS products will increase the skill and time horizon of flood forecasts (Demeritt et al., 2010).

The use of NWP ensembles to drive flood forecasting systems has increased and is a relevant research topic (Cloke and Pappenberger, 2009). Currently many flood forecasting centres use EPSs for representing uncertainty, but most of these are in Europe, Canada, the US and Australia (Demeritt et al., 2007). Experience of this kind of system in developing countries is very limited (Fan et al., 2014).

Rainfall forecasts provided by NWP require post-processing to correct bias and to reliably quantify uncertainty (Robertson et al., 2013). Several approaches have been used to produce ensemble rainfall forecasts by post-processing raw NWP. Robertson et al. (2013) present a method that uses a simplified version of the Bayesian joint probability modelling approach to produce forecast probability distributions for individual locations and forecast lead times; Theis et al. (2005) propose a methodology based on the hypothesis that some probabilistic information about a precipitation forecast at a certain time and location can be derived from its spatio-temporal neighbourhood in the model precipitation field. A set of forecasts is extracted from the spatio-temporal neighbourhood of a point and used to derive a probabilistic forecast at the central point of the neighbourhood; Bremnes (2004) proposes a method to produce probabilistic forecasts in terms of quantiles from NWP output using probit regression and quantile regression; Clark et al. (2004) use a two-stage approach that includes logistic regression and ordinary least squares regression to generate precipitation and temperature ensembles. Probabilistic forecasts of precipitation are especially challenging since precipitation has a mixed discrete–continuous probability distribution (Cortinas et al., 2002). In this study, a simple two-stage approach is used to produce precipitation ensembles from WRF forecasts, based on probit regression and quantile regression.

The assessment of the value of forecasts is a crucial issue. Verification is essential for the understanding of the abilities, weakness and value of forecasts, which leads to improving the forecast system. Several scales are proposed in the literature to assess the quality of forecasts as well as skill scores aimed at quantifying the relative accuracy of a forecast with respect to a reference forecast (Wilks, 2011), providing an estimation of its added value. Added value is often more important than a measure of skill (Jolliffe and Stephenson, 2012).

The objective of this paper is to assess the potential of NWP for flood early warning purposes in tropical mountainous watersheds, and the possible improvement that bias correction can provide. The main contribution of the research is focused on the assessment of the potential of WRF post-processed precipitation forecasts in providing skilful flood forecasts in watersheds characterised by rough topography and a tropical environment, where experience with precipitation forecasts, particularly ensemble approaches, has been very limited.

The study area is the upper part of a high tropical mountainous watershed located in Bogotá (Colombia). This corresponds to a páramo, which is a tropical high-mountain ecosystem, characterised by soils with a high water storage capacity and high conductivity. The hydrologic behaviour of páramo is complex, and there are still major gaps in knowledge (Sevink, 2011; Reyes, 2014; Buytaert et al., 2005, 2006). Additionally, hydrometeorological data are scarce.

In Colombia, NWP models forecasts are provided by the Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), the national hydro-meteorological institute. The raw output of these models is deterministic. In this study, the operational WRF model that is used by IDEAM for daily weather forecasts is used to provide precipitation forecasts that are post-processed and used to drive a hydrologic model of the study area. The discharges obtained from the hydrological model are used to assess the skill of the WRF model. This research assesses the current operational setup, in the meteorological agency of Colombia, which is based on the consideration that the orography has a significant effect in the mesoscale processes and that convection-allowing models such as the WRF may have the potential to provide skilful forecasts.

The paper is structured as follows: Sect. 2 describes the methods and data used for the assessment of the potential of the WRF model. In Sect. 3 the results are presented, including the bias correction of precipitation and the verification of precipitation and discharge forecasts. Section 4 discusses the results. Finally, the conclusions are presented in Sect. 5.

2　Methods and data

2.1　Study area

The Tunjuelo River basin is located in the south of the city of Bogotá (see Fig. 1). Its area is approximately $380\,km^2$ and the upper basin corresponds to a páramo area. The upper basin is composed of three watersheds – Chisacá, Mugroso and Curubital (see Fig. 1c). These discharge into two
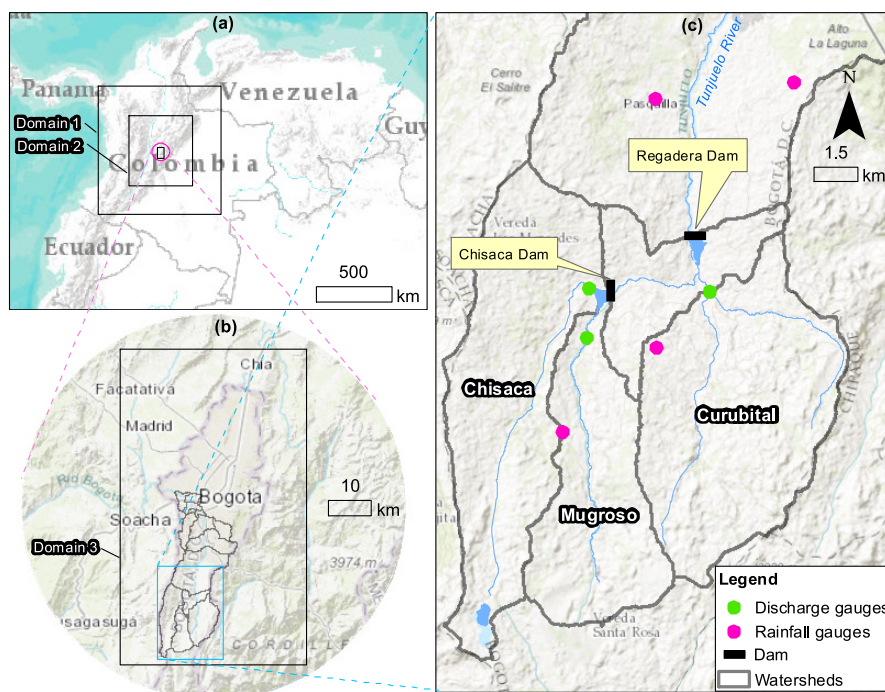
Figure 1. Study area. Service layer credits: Esri, DeLorme, NAVTEQ, TomTom, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, and the GIS User Community.

reservoirs (Chisacá and Regadera) with volumes of 3.3 and 6.7 Mm^3 respectively. The reservoirs are operated to supply 1.2 m^3 s^{-1} of water to the south of Bogotá. Flood waves in the urbanised lower basin are dominated by the discharge release of the two reservoirs.

The last major flood event in the basin occurred in 2002, causing the river to change its course to flow into two mining pits that since act as inline reservoirs. The peak release of the Regadera dam for that event reached 100 m^3 s^{-1} (Rogelis, 2006), which caused flooding downstream. The current flood warning criteria include warning levels set on the water levels in the Regadera Reservoir. Forecasting of the input discharges to this reservoir, as well as to the Chisacá reservoir, which is located immediately upstream, is crucial.

The upper basin of the Tunjuelo river has a unimodal precipitation regime (rainy season April–November) (Bernal et al., 2007). The largest discharges in the upper Tunjuelo basin occur during the months May, June and July (Rogelis, 2006).

The monitoring network is shown in Fig. 1. Two tipping bucket telemetric rain gauges currently operate in the upper Tunjuelo River basin and three discharge gauges are available in the three watersheds that discharge into the two reservoirs of the upper basin with hourly records.

2.2 WRF model data and observed rainfall fields

The hydro-meteorological agency of Colombia, IDEAM, runs the Weather Research and Forecasting (WRF) model version WRFV3.1 for Bogotá, using the initial and boundary conditions provided by the GFS model (Ruiz, 2010). The WRF model has been used at IDEAM since 2007 to produce forecasts at national level, as well as in the Bogotá region at higher resolution (Arango and Ruiz, 2011).

The data set of WRF forecast used in this paper corresponds to 107 selected days when significant storms were recorded, during the period July 2009 to December 2002. For each of these days, deterministic forecasts with the WRF model were generated at 00:00, 06:00, 12:00 and 18:00 GMT. This choice of storms ignores days when the WRF model would have delivered false alarms, focusing exclusively on the days when high precipitation was effectively measured. This is due to operational reasons. The WRF forecasts have not been archived; therefore they had to be recreated, and to achieve a feasible setup of the experiment from a computational point of view, this choice had to be made.

The WRF model comprises three nested domains, centred in Bogotá. The coarsest domain covers most of the Colombian national territory with a spatial resolution of 15 km; the intermediate domain covers mainly the central and eastern Andean cordilleras with a spatial resolution of 5 km; and the finest domain covering Bogotá and the study area only has
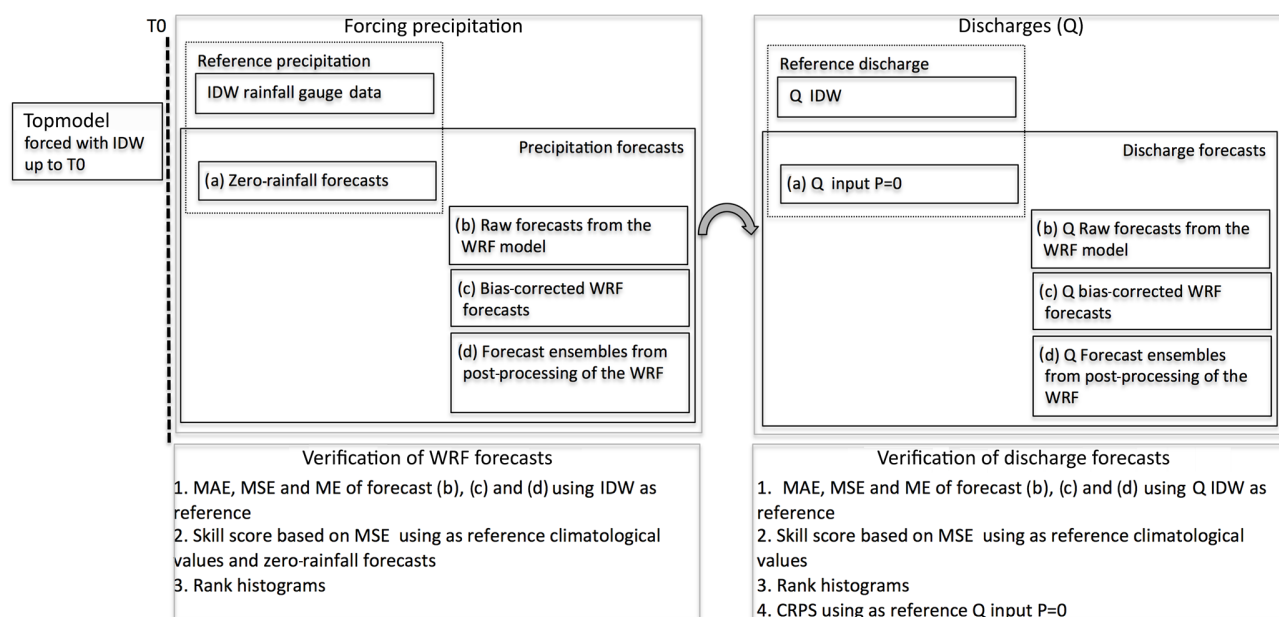
Figure 2. Overview of the methodology.

a spatial resolution of 1.67 km (Arango and Ruiz, 2011). The nested model domains have been set up using two-way communication (without smoothing). The parameterisation of the model corresponds to that used by IDEAM for its routine forecasts (Arango and Ruiz, 2011), using the Kain–Fritsch cumulus parameterisation scheme, except for the finest domain where convection is not parameterised but resolved. Microphysics is parameterised as the WRF Single-Moment 3-class scheme (WSM3). The lead time of the WRF model used operationally is 72 h, though in this paper only the first 48 h lead time was considered.

Observed hourly precipitation data were available from 18 tipping bucket rainfall gauges located in the city of Bogotá. These were used to produce rainfall fields through inverse distance weighing interpolation (IDW) for the entire area of the city that includes the study area.

Both WRF forecast rainfall fields and IDW rainfall fields were transformed to time series of mean average precipitation for the Chisacá, Mugroso and Curubital watersheds (see Fig. 1).

2.3 Methodology

The hydrology of the upper watersheds of the Tunjuelo river was modelled using the TOPMODEL (Beven and Kirkby, 1979). This was calibrated using as objective function the Kling and Gupta efficiency (Gupta et al., 2009) with the Shuffled Complex Evolution automatic search algorithm (Duan et al., 1992). The TOPMODEL was compared with two other hydrological models by Rogelis (2016), and found to provide a realistic representation of hydrological processes and good performance in the páramo watersheds for which

residence times in the range from 1 to 2 h were found. Furthermore, the preference of TOPMODEL for páramo areas has been reported previously by other authors (Buytaert and Beven, 2011). The hydrological models were forced with precipitation input obtained from inverse distance interpolation (IDW) of hourly rainfall obtained from gauges up to the time of start of the forecast (T0). From T0 the model was forced using four different rainfall forecasts, (a) zero-rainfall forecasts, (b) raw forecasts from the WRF model, (c) bias-corrected WRF forecasts, and (d) precipitation forecast ensembles obtained from the post-processing of the WRF model, based on probit regression and quantile regression. The precipitation forecasts obtained from each of the three nested domains were used to force the hydrological model and assess their difference in forecast skill.

Both precipitation forecasts and the corresponding streamflow forecasts were verified through the use of standard skill scores and rank histograms (see Fig. 2). The precipitation forecasts were verified through the use of (see Fig. 2): (1) the mean absolute error (MAE), mean square error (MSE) and mean error (ME) using the IDW precipitation as reference, and (2) through the skill score based on the MSE using as reference climatological values and the zero rainfall forecasts. In the case of streamflow forecasts obtained from the WRF model, these were verified through (1) the MAE, MSE and ME using as reference the discharge obtained from forcing the model with IDW precipitation, (2) the skill score based on MSE using as reference climatological values, and (3) the continuous ranked probability skill score (CRPSS) using as reference a streamflow time series generated using the precipitation forecast equal to zero to assess the added value of the WRF forecasts. The procedures for the production of the

different rainfall forecasts and the verification process are described in the following sections.

2.3.1 Generation of precipitation forecasts

The precipitation forecast to drive the models were generated under four strategies:

a. *Zero-rainfall forecasts.* After T0 values of zero precipitation are used to drive the TOPMODEL.

b. *Raw forecasts from the WRF.* The only post-processing of the WRF forecasts is the sampling of the grids to obtain the hourly mean areal precipitation for each watershed.

c. *Bias correction of WRF.* The time series of mean areal precipitation obtained from the WRF model are bias corrected through distribution-based scaling (DBS) (Yang et al., 2010). Mean areal precipitation was established using forecasts made for the 107 selected storm days as the biases during more extreme events is the prime interest. As reference values the observed time series of mean areal precipitation obtained from IDW are used, and the correction is carried out for each lead time of the WRF model. The DBS approach uses two steps: (1) correction of the percentage of wet time steps; a precipitation threshold of 0.1 mm (rainfall gauge accuracy) was used, below which a time step is considered to be dry, and (2) transformation of the remaining precipitation to match the observed frequency distribution. The gamma distribution is used to describe the probability distribution function of precipitation intensities given its ability to represent the asymmetrical and positively skewed distribution of precipitation. The distribution parameters are estimated using maximum likelihood estimation. To capture the main properties of normal precipitation, as well as of the extremes, the precipitation distribution is divided into two partitions separated by the 95th percentile. The resulting distribution corresponds to a double gamma distribution. The two sets of parameters are used to correct the WRF model outputs.

d. *Bias correction and generation of ensembles through post-processing of the rainfall forecast from the WRF model.* The reference and forecast time series were organised according to lead time and a two-stage post-processing model was applied to reflect the intermittent nature of rainfall (Rene et al., 2012; Clark et al., 2004). The first stage corresponds to the probit model (Scardovi, 2015; Kleiber et al., 2012; Bremnes, 2004) to simulate the probability of occurrence of precipitation. The second stage considers the probability distribution of the precipitation amount, given its occurrence, for which quantile regression was used. The mean precipi-

tation time series (observed hourly mean areal precipitation for each watershed) is first disaggregated into a time series of occurrence (1 = wet time step and 0 = dry time step) and precipitation amounts, which is established for the wet time steps only. The time series of occurrence is used as the response variable for the probit regression model. The probit model allows the probability of occurrence of precipitation to be estimated using the WRF hourly mean areal precipitation over each watershed as a predictor or secondary variable. The time series of precipitation amounts is used as the response variable for the quantile regression model. This allows different quantiles of the response variable to be estimated using the WRF hourly mean precipitation for each watershed as predictor or secondary variable.

In the probit regression model, given binary observations of precipitation occurrence $y_1, \ldots, y_n$, and $x_{i1}, \ldots, x_{ik}$ covariates associated with the ith response, the probability that $y_i = 1$, p_i, is written as (Albert, 2009):

$$p_i = P(y_i = 1) = \Phi(x_{i1}\beta_1 + \cdots + x_{ik}\beta_k), \tag{1}$$

where $\beta = (\beta_1, \ldots, \beta_k)$ is a vector of unknown regression coefficients and $\Phi()$ is the cumulative density function of a standard normal distribution. If we assume a uniform prior for β, then the posterior density is given by

$$g(\beta \mid y) \propto \prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{1-y_i}. \tag{2}$$

The binary response corresponds to the occurrence $y_i = 1$ or non-occurrence of rain $y_i = 0$. A latent variable Z_i is introduced in such a way that if Z_i is positive for $y_i = 1$ and Z_i is negative for $y_i = 0$. This latent variable is related to the k covariates by the normal regression model:

$$Z_i = x_{i1}\beta_1 + \cdots + x_{ik}\beta_k + \epsilon_i, \tag{3}$$

where $\epsilon_1, \ldots, \epsilon_n$ are a random sample from a standard normal distribution. Then,

$$P(y_i = 1) = P(Z_i > 0) = \Phi(x_{i1}\beta_1 + \cdots + x_{ik}\beta_k). \tag{4}$$

This can be considered as a missing data problem where there is a normal regression model of latent data $Z_1, \ldots, Z_n$ and the observed responses are missing or incomplete and can only be observed if $Z_i > 0(y_i = 1)$ or if $Z_i \leq 0(y_i = 0)$.

In order to avoid distributional assumptions, quantile regression (QR) will be used to describe the probability of the amounts of precipitation given its occurrence (Bremnes, 2004). Let $r_1, \ldots, r_{n*}$ denote observed precipitation amounts of cases with observed precipitation above a given lower threshold and $z_1, \ldots, z_{n*}$ corresponding predictor values where $z = (z, \ldots, z)^T$. For linear quantile functions (Bremnes, 2004),

$$q_\theta(z_i; \beta) = \beta_0 + \sum_{k=1}^{K} \beta_k z_{ik}, \tag{5}$$

an estimate of the $q_\theta(z_i; \beta), 0 < \beta < 1$ quantile, is obtained by solving the following minimisation problem with respect to β:

$$\arg\min_\beta \sum_{i=1}^{n*} p_\theta (r_i - q_\theta(z_i; \beta)), \tag{6}$$

where the function $q_\beta(*)$, is defined in terms of the absolute deviation of residuals (u) by

$$p_\theta(u) = \begin{cases} u\theta & \text{if } u \geq 0 \\ u(\theta - 1) & \text{otherwise.} \end{cases} \tag{7}$$

If several quantiles are of interest, the minimisation must be repeated for each quantile. A potential problem with using QR for the derivation of multiple conditional quantiles is that quantiles may cross, yielding predictive distributions that are not monotonically increasing, as a function of increasing quantiles (López López et al., 2014). In the present research study, the technique proposed by Muggeo et al. (2013) and implemented in the package quantregGrowth developed in the R environment (R Development Core Team, 2010) is used. This technique estimates nonparametric growth charts via quantile regression. Quantile curves are estimated via B-splines with a quadratic penalty on the spline coefficient differences, and non-crossing and monotonicity restrictions are set to obtain plausible estimates.

Two configurations were tested: (a) quantile regression applied to the raw precipitation data, and (b) quantile regression on the data transformed into normal domain through normal quantile transformation (NQT) (Bogner et al., 2012). For configuration (b) the time series of observed precipitation and forecast precipitation are transformed into the normal domain. After the derivation of the quantiles, the variables are back-transformed into original space. The rationale for using the transformation is that the joint distribution of transformed time series appears to be more linear, and can thus be better described by linear conditional quantiles (López López et al., 2014).

Back-transformation is, however, problematic if the quantiles of interest lie outside of the range of the empirical distribution of the untransformed variable in original space. To address this issue linear extrapolation in the tails of the distribution was used (López López et al., 2014).

To generate the probabilistic forecast of the occurrence of precipitation, the latent variable in the probit model is first sampled. The random value u is sampled from a uniform distribution. If $Z_i < u$ no precipitation occurs, if $Z_i \geq u$ precipitation is set to occur and the amount is computed with the quantile regression model. Given the occurrence of precipitation, the quantile regression model corresponding to the lead time under consideration is used to obtain the quantiles of the conditional distribution. The inverse cumulative function is approximated by a cubic spline and subsequently 50 uniform random numbers in the range 0–1 are generated and used to sample corresponding precipitation values from the inverse cumulative function.

2.3.2 Verification of forecasts

Deterministic forecasts of precipitation and discharge were verified using the mean absolute error (MAE), the mean squared error (MSE), the mean error (ME) and the skill score (SS) based on the MSE, using as reference forecasts provided by climatological values (subscript Clim in Eqs. 11 and 12) (Wilks, 2011) and zero forecasts (subscript 0 in Eqs. 13 and 14). The corresponding formulas are shown in Eqs. (8)–(14), where (y_k, o_k) is the kth of n pairs of forecast and observed values at a particular lead time. In the case of discharge, the discharge forecast obtained from the hydrological model driven with a precipitation input equal to zero, was used as reference. This reference was chosen given that this is the only choice when precipitation forecasts are not available, providing a method to assess the added value of WRF in comparison with current operation:

$$\text{MAE} = \frac{1}{n} \sum_{k=1}^{n} |y_k - o_k|, \tag{8}$$

$$\text{MSE} = \frac{1}{n} \sum_{k=1}^{n} (y_k - o_k)^2, \tag{9}$$

$$\text{ME} = \frac{1}{n} \sum_{k=1}^{n} (y_k - o_k) = \bar{y} - \bar{o}, \tag{10}$$

$$\text{MSE}_{\text{Clim}} = \frac{1}{n} \sum_{k=1}^{n} (\bar{o} - o_k)^2, \tag{11}$$

$$\text{SS}_{\text{Clim}} = \frac{\text{MSE} - \text{MSE}_{\text{Clim}}}{0 - \text{MSE}_{\text{Clim}}} = 1 - \frac{\text{MSE}}{\text{MSE}_{\text{Clim}}}, \tag{12}$$

$$\text{MSE}_0 = \frac{1}{n} \sum_{k=1}^{n} (0 - o_k)^2, \tag{13}$$

$$\text{SS}_0 = \frac{\text{MSE} - \text{MSE}_0}{0 - \text{MSE}_0} = 1 - \frac{\text{MSE}}{\text{MSE}_0}. \tag{14}$$

In order to assess the performance of the probit model constructed to model the occurrence of precipitation from the WRF model, the ROC (relative operating characteristic) diagram was used. This is a discrimination-based graphical forecast verification display (Wilks, 2011). For perfect forecasts the ROC curve consists of two line segments coincident with the left boundary and the upper boundary of the ROC diagram. At the other extreme of forecast performance, random forecasts consistent with the sample climatological probabilities, the ROC curve will consist of the 45° diagonal connecting the points (0, 0) and (1, 1). ROC curves for real forecasts generally fall between these two extremes, lying above and to the left of the 45° diagonal. Forecasts with better discrimination exhibit ROC curves approaching the upper-left corner of the ROC diagram more closely, whereas forecasts
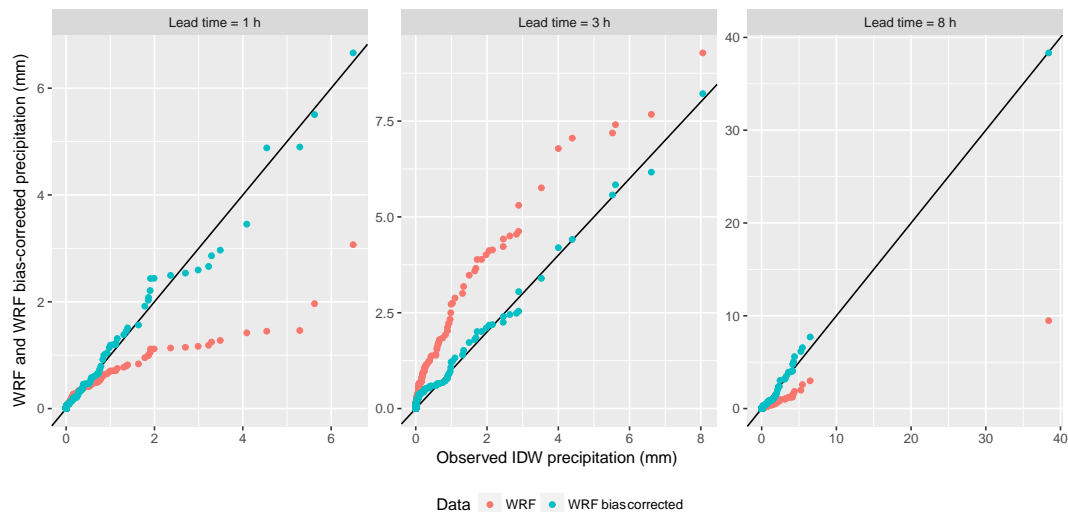
Figure 3. Q–Q plots (quantile–quantile plot, plot of the quantiles of the first data set against the quantiles of the second data set) for the Mugroso watershed comparing observed precipitation with WRF and WRF bias corrected precipitation.

with very little ability to discriminate the event exhibit ROC curves very close to the diagonal (Wilks, 2011). To summarise the ROC diagram, the area under the ROC curve was used (Wilks, 2011).

Regarding the quantile regression models, the pseudo-R-square measure (Koenker and Machado, 1999) was used to compare the models with normal quantile transformation and the ones based on raw data. The pseudo-R-square measure was proposed by Koenker and Machado (1999) as a goodness-of-fit indicator for quantile regression by comparing the sum of squared distances for the model of interest with the sum of squared distances between the observed and the fitted values that would be obtained if only the intercept term is included in the model (Hao and Naiman, 2007).

Ensemble forecasts of both precipitation and discharge were verified using rank histograms. Rank histograms make it possible to evaluate whether a collection of ensemble forecasts satisfy the consistency condition (Wilks, 2011). Rank histograms are constructed by accumulating the number of cases over space and time when the verifying analysis falls in any of $m+1$ intervals, where each of the $m+1$ intervals is defined by an ordered series of m ensemble members, including the two open-ended intervals (Jolliffe and Stephenson, 2012). Reliable or statistically consistent ensemble forecasts lead to a rank histogram that is close to flat (Jolliffe and Stephenson, 2012).

In order to assess the probabilistic forecasts of discharge obtained from the WRF precipitation ensembles, the CRPSS was used. The CRPSS is given by (Hersbach, 2000; Alfieri et al., 2014)

$$\mathrm{CRPSS} = \frac{\overline{\mathrm{CRPS_{ref}}} - \overline{\mathrm{CRPS_{forecast}}}}{\overline{\mathrm{CRPS_{ref}}}}, \tag{15}$$

where

$$\mathrm{CRPS} = \int\limits_{\infty}^{-\infty} \left[F(y) - F_0(y) \right]^2 \mathrm{d}y \tag{16}$$

and

$$F_0(y) = \begin{cases} 0, & y < \text{observed value} \\ 1, & y \geq \text{observed value.} \end{cases} \tag{17}$$

$F(y)$ is a stepwise cumulative distribution function (cdf) of the ensemble of each considered forecast. CRPSS ranges between 1 and $-\infty$. A CRPSS of 1 indicates a perfect forecast. Forecast ensembles are only valuable when CRPSS > 0, this is when the forecasts perform better than the reference. The discharge obtained from forecast precipitation equal to zero will be used as reference. Since this is a deterministic forecast, the CRPS$_\mathrm{ref}$ corresponds to the mean absolute error of all forecast and observed pairs (Hersbach, 2000).

3 Results

3.1 Bias correction of precipitation forecasts through DBS

Figure 3 shows the quantile–quantile plots (Q–Q plots) comparing the WRF precipitation and the bias-corrected WRF precipitation with the observed precipitation obtained from IDW, for the 107 selected days of significant precipitation. The figure shows results for three lead times for the Mugroso watershed, as examples of the 12 that were analysed, results for the other two watersheds were similar and not shown here for brevity; 12 h was chosen as a first estimate of the usable lead time for the forecast, since NWP models describing
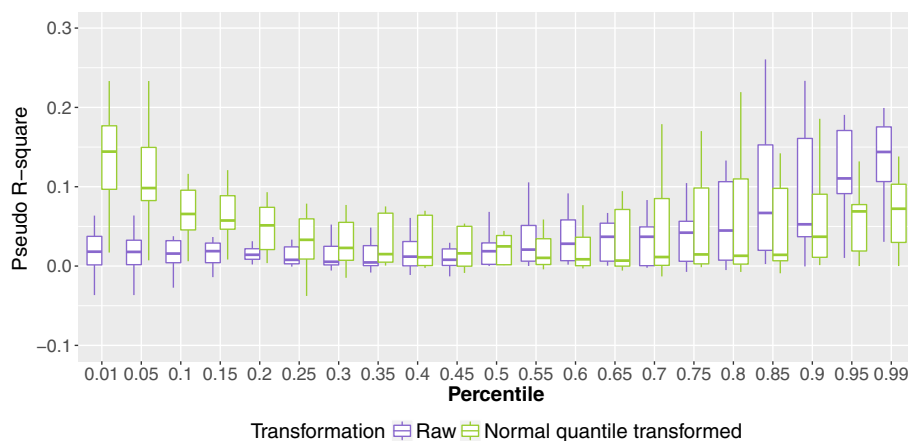
Figure 4. Pseudo-*R* square for the Mugroso watershed for the coarsest domain and lead times up to 12 h.

local–regional convective systems and orographic processes have been observed to significantly decrease their skill at lead times beyond 12–48 h (Cuo et al., 2011). The raw WRF and IDW time series were obtained from the corresponding rainfall fields as the mean areal precipitation calculated in the three watersheds of analysis by sampling the raster layers with the polygons of the watersheds.

The Q–Q plots show that both overestimation and underestimation of precipitations occurs. Underestimation is very noticeable for lead times of 1 h for values in the upper percentiles as shown in Fig. 3. In general for lead times between 2 and 5 h, overestimation of high precipitation values occurs, for larger lead times underestimation occurs and in the case of the lead time of 8 h a storm is entirely missed by the WRF model (see Fig. 3, Leadtime = 8 h).

3.2 Quantile regression model

Figure 4 shows the pseudo-*R* square for all quantile regression models for the Mugroso watershed and for the coarsest domain (resolution of 15 km), the results are similar for the other watersheds and domains. The *R*-square values are pooled for lead times up to 12 h. The figure shows that the values of the pseudo-*R* square up to the 50th percentile are higher when the normal quantile transformation is used, thus indicating a better fit of the quantile regression model. For higher percentiles the pseudo-*R* square of the models with normal quantile transformation is in approximately the same range of the raw data or lower. This means that for percentiles above the 50th percentile the fit of the quantile regressions is better for the raw data than for the normal transformed data.

3.3 Verification of deterministic precipitation forecasts and ensemble mean

Figure 5 shows the performance measures shown in Eqs. (8)–(14). The performance of all the WRF domains are shown for the three watersheds (Chisacá, Mugroso and Curubital).

From top to bottom the figure shows the ME, the MAE and the MSE. The lowest two figures for each watershed show the skill score based on MSE using climatology as a reference, and the skill score using zero precipitation as a reference.

The ME of the raw WRF forecasts shows that in the three watersheds for lead times between 1 and 5 h the precipitation amounts are overpredicted (positive ME), while after 5 h the ME shifts to negative values, indicating underprediction. The bias-corrected WRF forecasts show values of ME very close to zero, which could be expected given the procedure applied (see Fig. 5). In the case of the mean of the WRF ensemble without transformation, the ME is close to zero for lead times up to 6 h. Beyond this lead time there is a tendency to underpredict precipitation. The behaviour of the ME for the mean of the ensemble with normal quantile transformation is very similar to that when the transformation is not applied.

The highest values of MAE (see Fig. 5) are obtained from the DBS bias-corrected forecast for lead times larger than 6 h, while the raw WRF forecasts exhibit the highest values at the shorter lead times. In the case of the MSE, the highest values are obtained from the bias-corrected WRF forecasts at lead times of 2 and 8 h. The increase in MSE when bias correction is applied is mainly attributed to the influence of a single (observed) high-precipitation event that was not forecast by the WRF model. These intense precipitation events being missed in the forecasts cause all WRF values higher than the 95th percentile to be significantly increased through the bias correction procedure, increasing their square difference with the observed value. These extremes are mainly present at lead times of 2 and 8 h due to extreme precipitation events in this area typically occurring in the mid-afternoon, some 2 to 8 h after the start of the 18:00 and 12:00 GTM WRF forecast runs (the time zone of the area is UTC − 5).

For both MSE and MAE, the best performance is obtained from the ensemble mean of the WRF forecasts. This is the case both with and without normal quantile transformation (see Fig. 5), which provide very similar performance. Re-
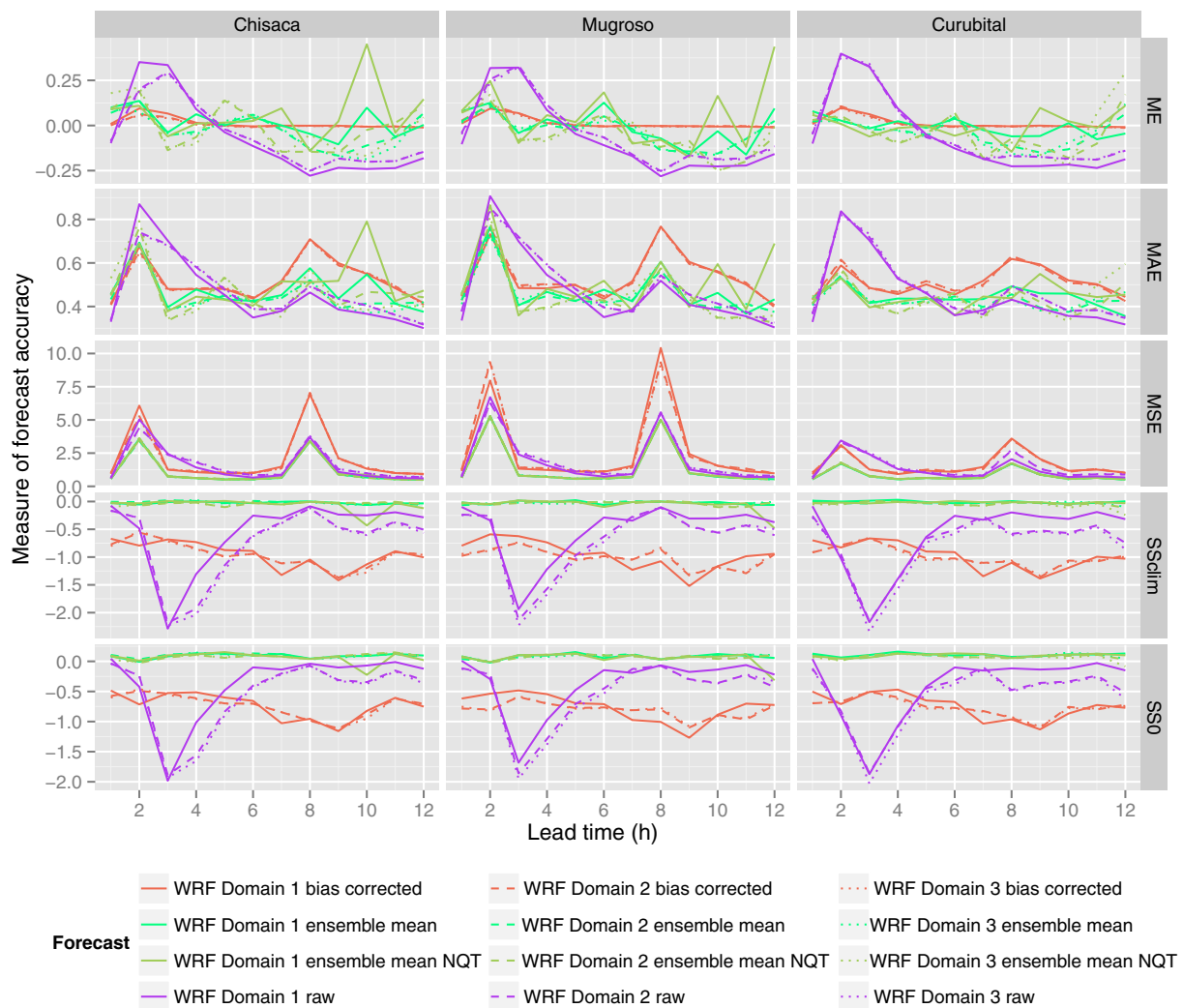
Figure 5. Accuracy measures for deterministic precipitation and ensemble mean obtained from the WRF model. Domain 1, domain 2 and domain 3 correspond respectively to the domains with resolutions 15, 5 and 1.67 km.

garding the skill score based on the MSE using the climatology as reference (see Eqs. 11 and 12) negative values are obtained for all forecasts, with the bias-corrected WRF and the raw WRF forecasts showing the worst skill. However, in the case of the ensemble mean, both for the normal quantile-transformed and raw ensembles, the values are very close to zero, albeit still negative. When zero precipitation is used as reference, the behaviour of the score is very similar to SS$_{Clim}$. However, most values for the ensemble mean forecast are positive, both with raw and with normal quantile-transformed data, with a maximum value of 0.16. This shows that the ensemble mean forecasts provide the best results in comparison to the other forecasts. In the case of the climatology being used as reference, they are approximately as good as the reference and compared to zero precipitation as reference; they provide an improvement up to 16 % over the reference.

Given that a single extreme storm which was entirely missed by the WRF forecast has a significant influence in the results shown in Fig. 5, (resulting in the poor performance at lead times of 2 and 8 h), the accuracy metrics were recalculated from the data set, but now excluding this storm. The results are shown in Fig. 6 only for the MSE, since the other metrics are less affected by the exclusion of the extreme storm. In Fig. 6 the reduction of the MSE is highly noticeable in all the forecasts (see Fig. 5 in comparison with Fig. 6), and the lowest values are exhibited by the ensemble mean of both raw and normal quantile-transformed data, which show very similar behaviour. This analysis clearly shows the sensitivity of performance measures based on squared residuals, such as MSE, to a single event with high precipitation being missed in the WRF forecasts.
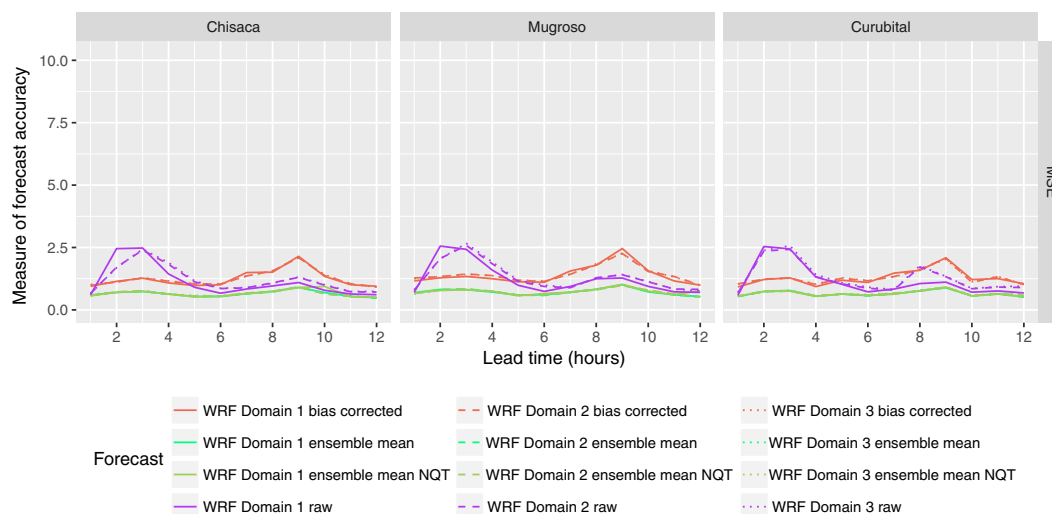
Figure 6. Accuracy measures for deterministic precipitation and ensemble mean obtained from the WRF model without the highest precipitation missed by the WRF model. Domain 1, domain 2 and domain 3 correspond respectively to the domains with resolutions 15, 5 and 1.67 km. The legend is the same as in Fig. 5.
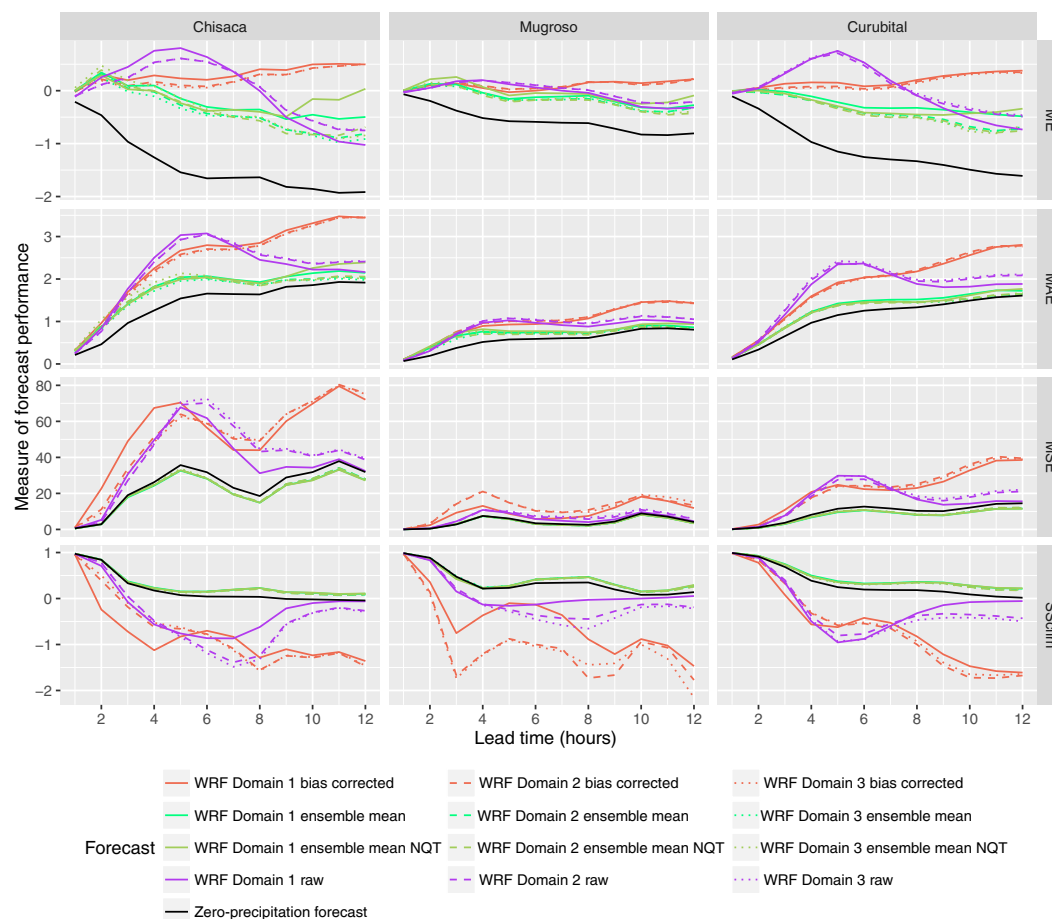

Figure 7. Performance measures for deterministic discharge and ensemble mean forecasts. Domain 1, domain 2 and domain 3 correspond respectively to the domains with resolutions 15, 5 and 1.67 km.
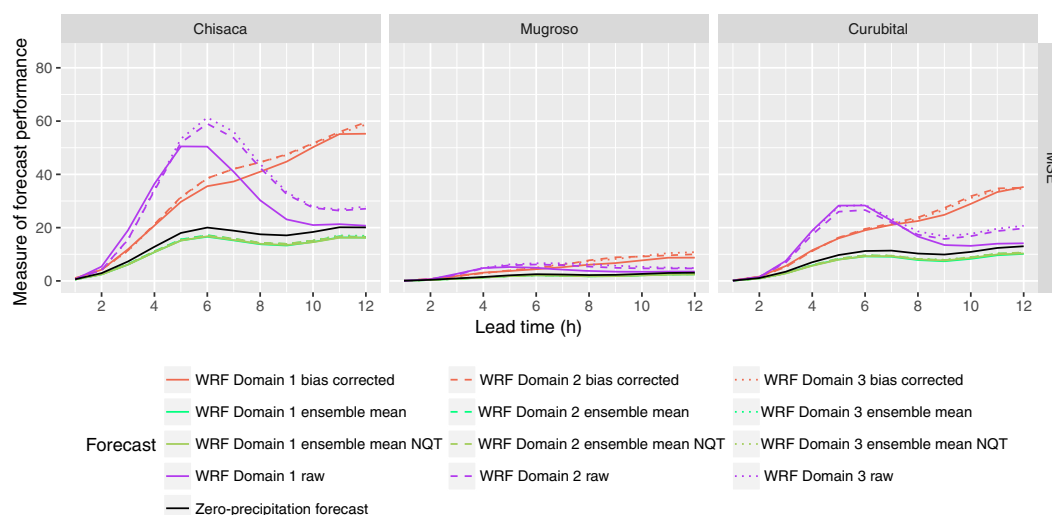
Figure 8. Performance measures for deterministic discharge and ensemble mean forecasts obtained excluding the storm with the highest precipitation missed by the WRF model. Domain 1, domain 2 and domain 3 correspond respectively to the domains with resolutions 15, 5 and 1.67 km. The legend is the same as in Fig. 7.
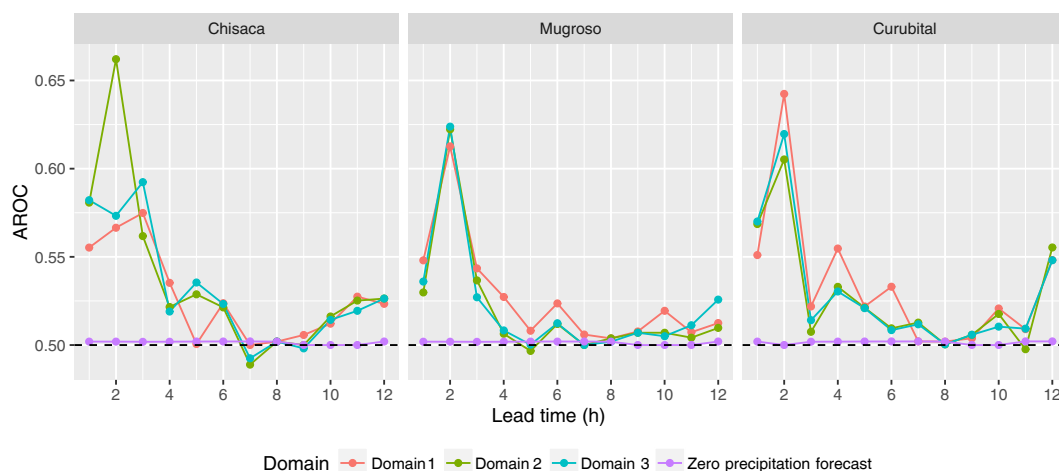

Figure 9. Area under the ROC curve (AROC) for the forecast of occurrence of precipitation with the probit model. Domain 1, domain 2 and domain 3 correspond respectively to the domains with resolutions 15, 5 and 1.67 km.

3.4 Verification of deterministic discharge forecasts and ensemble mean

The results for discharge using the precipitation forecasts are shown in Fig. 7. The structure of the figure is similar to that of Fig. 5, although the skill score is now determined using the climatology as reference, as the zero-precipitation forecast is considered as one of the forecasts. The ME shows that the raw WRF forecasts overestimate the discharges up to a lead time of some 8 h, while for larger lead times underestimation occurs. When the bias-corrected WRF forecasts are used, the ME is close to zero, slowly increasing with lead time. The use of the mean of the ensembles with and without normal quantile transformation produces ME close to zero

up to lead times of 4 h, beyond which underestimation occurs (negative values of ME). In terms of the MSE the mean of the ensemble with and without normal quantile transformation produce the smallest values for all lead times, which is reflected in the SS_{Clim}, where they produce the highest values, although there is little improvement in the skill over that obtained from using a zero-precipitation series as forecast.

Figure 8 shows the MSE calculated using the stream flow simulations, but now excluding the storm with the highest precipitation that was missed by the WRF model. Only the MSE is shown, since the other metrics are less affected by the exclusion of the extreme storm. This shows a significant improvement in comparison with Fig. 7, particularly for the DBS bias-corrected data.

From the results shown in Figs. 5–8, the difference in performance for the three domains of the WRF model is not significant, both in the precipitation and in the discharge forecasts.

3.5 Verification of probabilistic forecasts

Figure 9 shows the area under the ROC curve (AROC), showing the skill of the forecasts to discriminate between the occurrence and non-occurrence of precipitation. Scores are shown for the three watersheds (Chisacá, Mugroso and Curubital in Fig. 1); for the three domains of the WRF model for lead times up to 12 h; and for a precipitation forecast of zero. The behaviour of the AROC is similar in the three watersheds and three domains, with the highest values found in the first 2 h, dropping to a value close to 0.5 after a lead time of about 5 h. The AROC values for a forecast of zero precipitation is very close to 0.5 for all lead times, and thus is almost equal to random forecasts consistent with the sample climatological probabilities (Wilks, 2011). This shows that the forecasts have some skill at short lead times in comparison to both zero-precipitation forecasts and random forecasts, although that skill is limited.

The rank histograms of the precipitation ensembles are shown in Fig. 10. The behaviour for all domains and all lead times is similar, therefore the figure only shows the lead time of 1 h. The rank histogram is approximately uniform from rank 1 to rank 5, and from rank 6 to rank 10 there is an increase in the frequency of the IDW precipitation values falling in these higher ranks. Similar behaviour in all rank histograms was observed when the storm with the highest precipitation missed by the WRF model was excluded. This suggests that the WRF ensemble is slightly underforecasting.

Figure 11 shows the rank histograms for the discharges obtained from the precipitation ensemble using the TOP-MODEL. The rank histograms show approximate uniformity mainly for lead times up to 3 h. For longer lead times a slight underdispersion is observed. The figure shows lead times up to 6 h since for longer lead times the behaviour is similar to that of the lead time of 6 h.

Figure 12 shows the CRPS for the discharge ensembles, the MAE for the forecast produced by using zero as precipitation forecast and the corresponding CRPSS. The CRPSS exhibits values in the range 10.5–21 % (see Fig. 12) with most values in a range of 14–16 %. The same behaviour is observed in the three watersheds. The comparison between normal quantile-transformed quantile regressions and the quantile regressions with raw data again shows that the difference between the two is negligible.

4 Discussion

4.1 Evaluating precipitation forecasts from the WRF model

The Q–Q plots of the raw WRF forecasts and the precipitation obtained from IDW interpolation shows that the WRF model tends to underpredict precipitation for the first hour of lead time and to overpredict precipitation for larger lead times. The overpredicting behaviour has been observed in other tropical areas (Mourre et al., 2016). While the details of the implementation of the WRF model are out of the scope of this study, possible causes of the precipitation error found by other authors include; errors in the lateral boundary conditions (Ochoa et al., 2014); poor representation of the topography (Ochoa et al., 2014); and choice of convective treatment, microphysics and planetary boundary layer (Jankov et al., 2005). However, there are individual cases where the WRF model did not detect the occurrence of high precipitation. Missing or underpredicted forecasts of intensive precipitation have also been observed by other authors (Kryza et al., 2013; Liu et al., 2015), which may suggest the need for an improvement in model parameterisation and data assimilation.

No significant differences among three domains of the WRF model were found in the behaviour of the scores for the tests carried out in this study. The two-way communication between the nested domains implies that precipitation volumes are aligned between the different resolutions, with the small differences found likely being due to differing ratios between the WRF model grid cell sizes and the sizes of the watersheds. Other applications of NWP (e.g. Roberts et al., 2009) show that finer resolutions are capable of producing more accurate predictions, and that physics configuration, resolution and initial conditions highly influence the WRF model performance (Kryza et al., 2013). The similarity of results regardless of resolution found in this study may be also related to deficiencies in the parameterisation of the model, or to the inability to sufficiently resolve the topography. A more detailed review of the WRF model would be required to reveal possible deficiencies, and to suggest improvements. Furthermore, this analysis has only been focused on the forecast of high precipitation when this has been observed, which provides insight into the skill of the WRF model to detect events that can lead to floods. However, the behaviour of the WRF model under other conditions such as those that may lead to false alarms is planned for future research, which may reveal other potentialities and deficiencies to be taken into account in the improvement of the model.

The WRF model has been shown to be highly sensitive to the parameterisation of cumulus and microphysical processes (Remesan et al., 2014; Rama Rao et al., 2012). Parameter sensitivity is generally dependent on local conditions (Di et al., 2014), and the best configuration varies with time and rainfall threshold (Jankov et al., 2005). While it was our
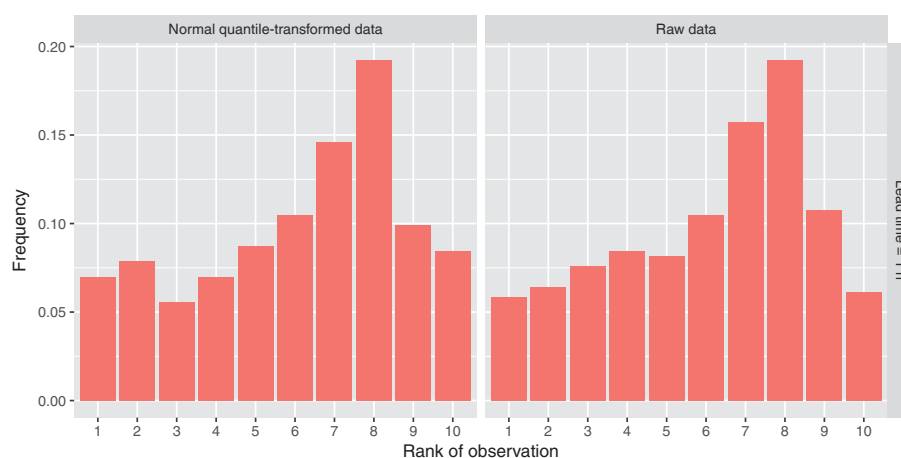
Figure 10. Rank histograms for the WRF ensemble for the Mugroso watershed for the finest domain (resolution of 1.67 km) and lead time of 1 h.
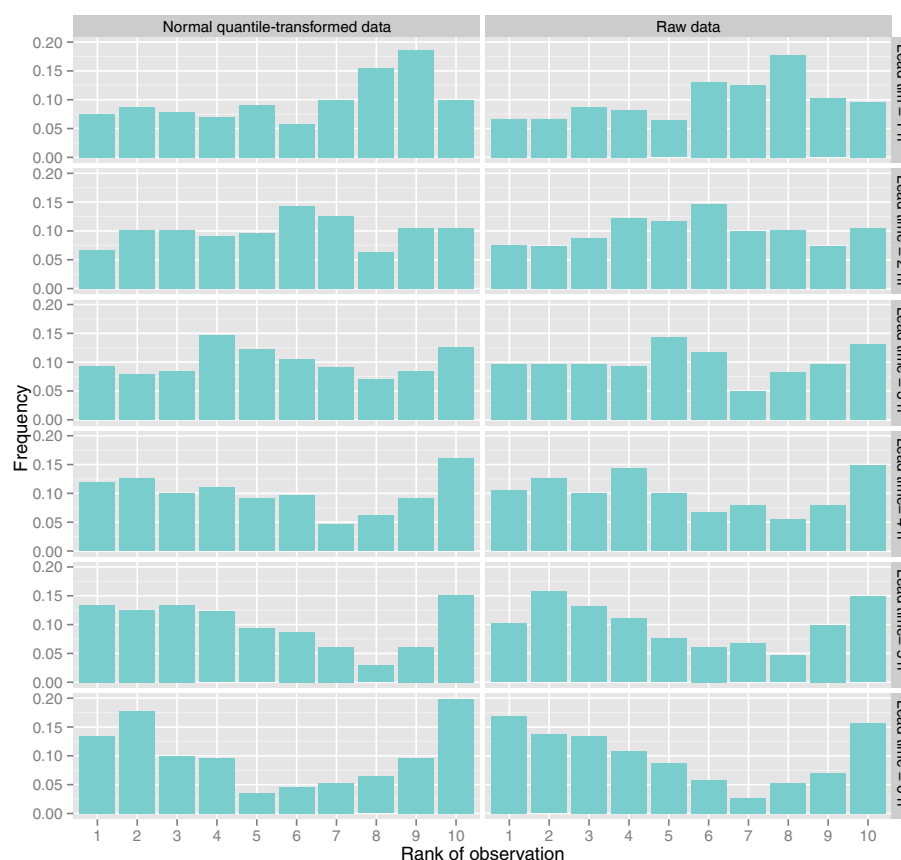

Figure 11. Rank histograms for the Q ensemble using as reference time series the discharge simulated with the TOPMODEL.

objective to test the configuration of the WRF model as it is used operationally by IDEAM and optimising the parameterisation of the WRF model is outside the scope of the present paper, further work could test various model parameterisations, assessing their sensitivity and searching for an optimal or a set of optimal configurations that help to improve pa-

rameterisation (in the case of the coarser model domains) and in particular resolving (in the case of the finest model domain) convective precipitation as frequently occurs in the case study area. Observing systems such as radars, gauges and satellites for data assimilation in NWP analysis could
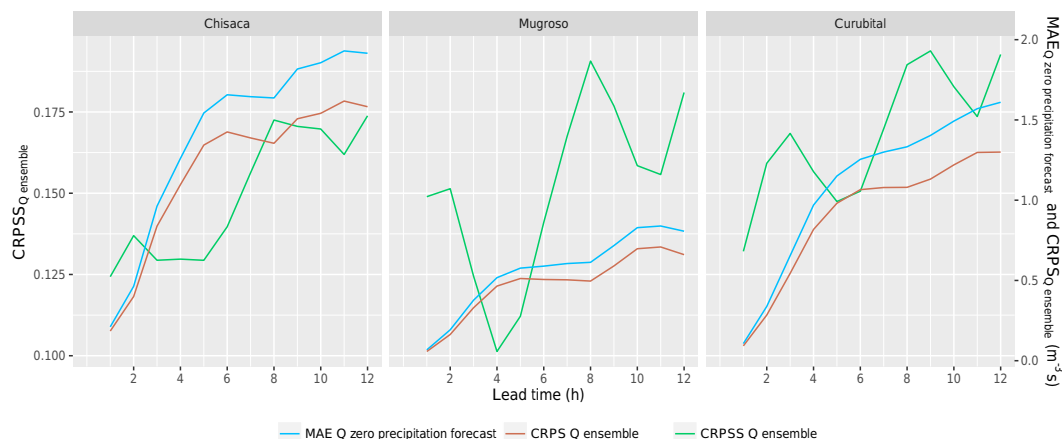
Figure 12. CRPSS for the discharge ensembles.

further improve the NWP forecasts (Rossa et al., 2011; Cuo et al., 2011; Yucel et al., 2015; Liu et al., 2015).

The application of the DBS approach has been reported previously to considerably reduce the differences in rainfall frequency between observations and forecasts (Yang et al., 2010). This reduction of differences was found in this study, as shown in Fig. 3. However, where high precipitation values were observed but not forecast in a single extreme event, the bias correction had the effect of reducing overall performance, particularly when measured by the MSE, since this is more sensitive to large residuals. This behaviour suggests that the DBS bias correction can introduce undesirable effects, which limit its effectiveness (Ehret et al., 2012). Similar behaviour was observed with the mean of the ensemble, although with less impact on the performance measures in comparison to the DBS approach. A significant difference between the ensemble means obtained from raw and normal quantile-transformed data were not found.

These biases are the result of the inability of the WRF model to adequately represent convective precipitation, failing to correctly predict when large events occur, or where it occurs. This has also been observed by other authors. For instance, Verkade et al. (2013) found that post-processing does not improve on all qualities at all lead times and at all levels of the verifying observations. The cause is rooted in the impossibility of the post-processing approaches to replace adequate model representation of physical processes (Haerter et al., 2011). Haerter et al. (2011) found that the conditions on climate model data to make the application of statistical bias correction schemes reasonable are that a realistic representation of the physical processes involved must be ensured; and that the quantitative discrepancies between the modelled and observed probability density function of the quantity at hand must be constant in time. Similarly, Chen et al. (2013) emphasise the impossibility any bias correction method being successful if there is no coherence between simulated and observed precipitation. The WRF data used

in this study show limitations in fulfilling both these conditions, which leads to the post-processed precipitation being no more skilful than the sample climatology (as shown by negative values of SS_{Clim}) and providing only a modest improvement in comparison to the zero-precipitation forecast (as shown by most values of SS0 being positive, albeit low). However, an important result of this analysis is that despite the limitations, the ensemble mean outperforms the DBS bias correction and seems to be less sensitive to the presence of intense precipitation events that are not forecast by the WRF model. Post-processing the deterministic forecasts to generate the ensembles is clearly more sophisticated than just correcting the mean bias (Robertson et al., 2013). During the process, different distributions are used for each raw WRF forecast, which results in an improvement of skill that outperforms the bias correction through DBS. The magnitude of the bias of the ensemble mean is nearly always smaller than the raw forecasts, as well as of the DBS bias-corrected forecasts, particularly for lead times up to 6 h.

Regarding the verification of the precipitation ensembles, the rank histograms show that the IDW precipitation too often falls between ranks 6 and 10. This reflects overdispersion of the ensemble towards low values. This produces an underforecasting bias in the ensemble. The same behaviour of the rank histograms is observed in the ensembles obtained from quantile regression, both when the normal quantile transformation is used in the precipitation data and when raw data are used. This means that the transformation of the values does not improve the consistency of the ensemble.

The AROC of the occurrence of precipitation shows that values higher than 0.5 are obtained in all watersheds for lead times up to 5 h, indicating that the probit model has the ability to discriminate between the occurrence or non-occurrence of precipitation up to these lead times. In all domains and watersheds, the highest values of AROC are obtained for lead times between 2 and 3 h, in the range of 0.63–0.67. This means that there is some skill in the probit model to fore-

cast occurrence of precipitation for lead times up to 5 h, even if the skill of the amount of precipitation is low. For longer times the probit model does not have any skill, consistent with other studies where the performance significantly reduces after a few hours (Liu et al., 2015). This behaviour of the probit model implies that for lead times up to 5 h the probability of occurrence of precipitation can be modelled from the WRF precipitation with some skill, meaning that the higher the WRF forecast precipitation value the higher the probability of occurrence of precipitation. However, even though the lead times at which the model still has skill are limited, this does imply that the current ability of NWP in tropical mountainous areas can provide a limited extension of lead time beyond the concentration time of small, flashy, watersheds.

4.2 Evaluating discharge forecast

The comparison of Figs. 7 and 8 shows that the biases in the precipitation forecasts are reflected in the biases of the discharge. When the single event with high precipitation that is not forecast by the WRF is excluded from the analysis, a reduction of bias in the forecast precipitation is observed, as well as a reduction in the bias of the forecast discharge. This is consistent with the findings of other studies where the errors in bias-corrected precipitation lead to amplified errors in modelled runoff (Teng et al., 2015), since in this analysis the increase in errors in the bias-corrected time series amplify the errors in the simulated discharges.

The forecasts generated with the post-processed ensemble precipitation forecasts exhibit better skill than the deterministic forecasts (based on raw, bias corrected and zero precipitation). Higher skill of the ensemble mean in comparison with deterministic forecasts has also been found in other flood forecasting systems driven by the WRF model (Calvetti et al., 2014) and other NWP models (Vincendon et al., 2011).

In terms of the lead time at which the coupled WRF-TOPMODEL forecasts provide value in these watersheds, a first limit is provided by the skill of the WRF model to forecast the occurrence of precipitation up to 5 h (see Fig. 9). A limit is also observed in the skill scores of the discharge forecast in Figs. 7 and 8. Beyond the lead time of 2 h, which is approximately the time of concentration of the watersheds, the skill (SS_{Clim} in Figs. 7 and 8) starts to drop significantly. For the discharge obtained from raw and DBS bias-corrected precipitation, 3 and 4 h are the respective lead times at which the added value is totally lost in comparison with the climatology (SS_{Clim} in Figs. 7 and 8 reaches zero). In the case of the ensemble means, the skill also decreases progressively after 2 h, with some value compared to climatology up to a lead time of some 6 h. Beyond this lead time, the values remain approximately constant in a range of 0.2 to 0.5 for the three watersheds. This behaviour indicates that the use of the ensemble mean provides some skill up to a lead time of 6 h. This is 3 h longer than the lead time provided by the raw WRF forecast or the DBS bias-corrected WRF forecast. A lead time of 6 h is consistent with other flood forecasting systems in small mountainous catchments driven by NWP models (Verkade and Werner, 2011).

The hydrological model produces ensemble results with rank histograms that do not reflect the overprediction of the precipitation ensemble, with approximately uniform rank histograms mainly for the first lead times (see Fig. 11). Approximately the same shape of the rank histograms is observed for ensembles obtained from quantile regression when no transformation is used in the precipitation data and when raw data are used. The minor influence of the transformation is consistent in all the performance assessments. This may be due to the relatively low improvement of the fit of the quantile regressions when the normal quantile transformation is used. According to Fig. 4, the most significant improvement is found in the low percentiles (up to the 25th percentile). For higher percentiles, the improvement is not significant and for some percentiles degradation of goodness of fit occurs. As floods occur for higher rainfall percentiles, the similarity in behaviour is logical.

The results of the CRPSS (see Fig. 12) show that forcing the hydrological model with WRF ensembles improves the forecasts in a range of 8.5–22 % in comparison with a forecast produced with the hydrological model forced with zero precipitation for lead times between 1 and 12 h. The positive values of the CRPSS imply added value to the forecasts, albeit modest. The CRPSS obtained from the ensembles is comparable to the CRPSS values found in other areas with other NWP models, albeit in the low range. For example, Robertson et al. (2013) found a CRPSS of 37 % on average for post-processed ensembles in Australia, where rainfall is predominantly produced by large-scale synoptic systems that are better predicted by NWP models. Therefore, given the high complexity of the meteorological conditions of the study area, and despite the relatively poor skill of the WRF model in predicting precipitation amounts, the WRF model does show promise at producing a benefit in its use for flood forecasting compared to not using precipitation forecasts. This is likely due to the skill found in the probit model in predicting the occurrence of rainfall.

Besides improvements in parameterisation and data assimilation, as described previously, bias correction of streamflow could provide a further skill improvement (Yuan and Wood, 2012).

5 Conclusions

This paper presents the assessment of WRF forecasts produced in a tropical high-mountain watershed in Bogotá, Colombia. The WRF forecasts were used to force a hydrological model of the study area. Simulated discharges were used to assess the value of the WRF forecasts for flood early warning. Although the skill of the forecasts developed is lim-

ited, results show that the streamflow forecasts obtained from a hydrological model forced by post-processed WRF precipitation do add value to the flood early warning system, particularly when compared to zero-precipitation forecasts.

The WRF model for the study area provides forecasts that are found to overpredict precipitation for the shorter lead times (up to 6 h), and that tend to fail to forecast high-intensity precipitation events. This behaviour may be due to parameterisation deficiencies, errors in boundary conditions and/or poor representation of the topography. A more detailed evaluation of the WRF model in this study area is recommended, including the assessing of other convective and microphysics schemes to identify the most suitable parameterisation. The use of satellite and soon to be available radar data may further help improve performance.

Bias correction of the WRF forecast precipitation, through distribution-based scaling (DBS) was shown to significantly reduce the differences in rainfall amounts between observations and forecasts. However, the DBS-corrected forecasts were found to be sensitive to failure of the WRF model to predict a small number of extreme events, resulting in a reduction of overall performance. This sensitivity was also observed with the ensemble mean forecasts, although for this post-processing technique the impact of these few missed events on the performance was lower when compared with DBS. This undesirable behaviour is the result of the inability of the WRF model to adequately represent convective precipitation, which cannot be corrected through simple post-processing.

Despite the limitations in the WRF forecasts, the ensemble mean was found to outperform the DBS bias correction, as well as being less sensitive to the presence of highly intense precipitation that are not correctly forecast by the WRF model. Although none of the precipitation forecasts that were assessed in this analysis showed added value when compared to climatology, the reduction of biases obtained from the ensembles show the potential of this method and model to provide usable precipitation forecasts. Although the skill of the quantitative precipitation models was found to be low, the probit model, which was applied to forecast the probability of occurrence of precipitation, based on the WRF forecast, showed that the WRF model does have some skill at short lead times (up to 5 h) in comparison to both zero-precipitation forecasts and random forecasts, although that skill is limited.

Increases in precipitation biases are reflected in the discharge forecasts. However, discharge forecasts generated with the post-processed ensembles did exhibit higher skill than deterministic forecasts (based either on raw, bias corrected or zero precipitation). Furthermore, the quality of these forecasts is better than that which could be obtained using zero precipitation as input to the hydrological model. The potentially usable lead times for discharge forecasts obtained from the WRF model found in this analysis (5–6 h) are in the range of lead times found in other studies with other

NWP models. Despite the fact that the added value of the WRF model forecasts is modest, the results found do show there is some potential for increasing forecast skill in areas of high meteorological and topographic complexity, as well as how these can be improved.

References

Albert, J.: Bayesian Computation with R, 2nd Edn., Springer-Verlag, New York, 2009.

Alfieri, L., Pappenberger, F., Wetterhall, F., Haiden, T., Richardson, D., and Salamon, P.: Evaluation of ensemble streamflow predictions in Europe, J. Hydrol., 517, 913–922, 2014.

Arango, C. and Ruiz, J. F.: Implementación del modelo WRF para la sabana de Bogotá, Tech. rep., Instituto de Hidrología, Meteorología y Estudios Ambientales, Bogotá, available at: http://www.ideam.gov.co/documents/21021/21132/Modelo_WRF_Bogota.pdf/f1d34638-e9f8-4689-b5f4-31957c231c46 (last access: 20 November 2015), 2011.

Bernal, G., Rosero, M., Cadena, M., Montealegre, J., and Sanabria, F.: Estudio de la Caracterización Climática de Bogotá y cuenca alta del Río Tunjuelo, Tech. rep., Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM – Fondo de Prevención y Atención de Emergencias FOPAE, Bogotá, 2007.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant, Hydrol. Sci. B., 24, 43–69, https://doi.org/10.1080/02626667909491834, 1979.

Bogner, K., Pappenberger, F., and Cloke, H. L.: Technical Note: The normal quantile transformation and its application in a flood

forecasting system, Hydrol. Earth Syst. Sci., 16, 1085–1094, https://doi.org/10.5194/hess-16-1085-2012, 2012.

Bremnes, J. B.: Probabilistic forecasts of precipitation in terms of quantiles using NWP model output, Mon. Weather Rev., 132, 338–347, 2004.

Buytaert, W. and Beven, K.: Models as multiple working hypotheses: hydrological simulation of tropical alpine wetlands, Hydrol. Process., 25, 1784–1799, https://doi.org/10.1002/hyp.7936, 2011.

Buytaert, W., Wyseure, G., De Bièvre, B., and Deckers, J.: The effect of land-use changes on the hydrological behaviour of Histic Andosols in south Ecuador, Hydrol. Process., 19, 3985–3997, https://doi.org/10.1002/hyp.5867, 2005.

Buytaert, W., Célleri, R., De Bièvre, B., Cisneros, F., Wyseure, G., Deckers, J., and Hofstede, R.: Human impact on the hydrology of the Andean páramos, Earth-Sci. Rev., 79, 53–72, https://doi.org/10.1016/j.earscirev.2006.06.002, 2006.

Calvetti, L., José, A., and Filho, P.: Ensemble hydrometeorological forecasts using WRF hourly QPF and TopModel for a middle watershed, Adv. Meteorol., 2014, 484120, https://doi.org/10.1155/2014/484120, 2014.

Chen, J., Brissette, F. P., Chaumont, D., and Braun, M.: Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America, Water Resour. Res., 49, 4187–4205, https://doi.org/10.1002/wrcr.20331, 2013.

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., and Wilby, R.: The Schaake Shuffle: a method for reconstructing space–time variability in forecasted precipitation and temperature fields, J. Hydrometeorol., 5, 243–262, 2004.

Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: a review, J. Hydrol., 375, 613–626, https://doi.org/10.1016/j.jhydrol.2009.06.005, 2009.

Colman, B., Cook, K., and Snyder, B.: Numerical weather prediction and weather forecasting in complex terrain, in: Mountain Weather Research and Forecasting Recent Progress and Current Challenges, chap. 11, edited by: Chow, F. K., De Wekker, S. F., and Snyder, B. J., Springer, Dordrecht, 2013.

Cortinas, J. V. J., Brill, K. F., and Baldwin, M. E.: Probabilistic forecasts of precipitation type, in: 16th Conference on Probability and Statistics in the Atmospheric Sciences, 12–17 January 2002, Orlando, Florida, 2002.

Cuo, L., Pagano, T. C., and Wang, Q. J.: A review of quantitative precipitation forecasts and their use in short- to medium-range streamflow forecasting, J. Hydrometeorol., 12, 713–728, https://doi.org/10.1175/2011JHM1347.1, 2011.

Demeritt, D., Cloke, H., Pappenberger, F., Thielen, J., Bartholmes, J., and Ramos, M. H.: Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting, Environ. Hazards, 7, 115–127, 2007.

Demeritt, D., Nobert, Ś., Cloke, H., and Pappenberg, F.: Challenges in communicating and using ensembles in operational flood forecasting, Meteorol. Appl., 17, 209–222, 2010.

Di, Z., Qingyun, D., Wei, G., Chen, W., Yanjun, G., Jiping, Q., Jianduo, L., Chiyuan, M., Aizhong, Y., and Charles, T.: Assessing WRF model parameter sensitivity: A case study with 5 day summer precipitation forecasting in the Greater Beijing Area, Geophys. Res. Lett., 42, 579–587, https://doi.org/10.1002/2014GL061623, 2014.

Duan, Q.-Y. Y., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall–runoff models, Water Resour. Res., 28, 1015–1031, https://doi.org/10.1029/91WR02985, 1992.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?", Hydrol. Earth Syst. Sci., 16, 3391–3404, https://doi.org/10.5194/hess-16-3391-2012, 2012.

Fan, F. M., Collischonn, W., Meller, A., and Botelho, L. C. M.: Ensemble streamflow forecasting experiments in a tropical basin: the São Francisco river case study, J. Hydrol., 519, 2906–2919, https://doi.org/10.1016/j.jhydrol.2014.04.038, 2014.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Habets, F., LeMoigne, P., and Noilhan, J.: On the utility of operational precipitation forecasts to served as input for streamflow forecasting, J. Hydrol., 293, 270–288, 2004.

Haerter, J. O., Hagemann, S., Moseley, C., and Piani, C.: Climate model bias correction and the role of timescales, Hydrol. Earth Syst. Sci., 15, 1065–1079, https://doi.org/10.5194/hess-15-1065-2011, 2011.

Hao, L. and Naiman, D.: Quantile Regression, SAGE Publications, Thousand Oaks, California, 2007.

Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather Forecast., 15, 559–570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2, 2000.

Jankov, I., Gallus, W. A., Segal, M., Shaw, B., and Koch, S. E.: The impact of different WRF model physical parameterizations and their interactions on warm season MCS rainfall, Weather Forecast., 20, 1048–1060, https://doi.org/10.1175/WAF888.1, 2005.

Jolliffe, I. and Stephenson, D.: Forecast Verification: A Practitioner's Guide in Atmospheric Science, John Wiley & Sons, Ltd, West Sussex, England, 2012.

Kleiber, W., Katz, R. W., and Rajagopalan, B.: Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes, Water Resour. Res., 48, 1–17, 2012.

Koenker, R. and Machado, J. A. F.: Goodness of fit and related inference processes for quantile regression. (Statistical Data Included), J. Am. Stat. Assoc., 94, 1–22, 1999.

Kryza, M., Werner, M., Wałaszek, K., and Dore, A. J.: Application and evaluation of the WRF model for high-resolution forecasting of rainfall – a case study of SW Poland, Meteorol. Z., 22, 595–601, 2013.

Laing, A. and Evans, J.-L.: Introduction to Tropical Meteorology, The Comet Program, Boulder, USA, 2010.

Leutbecher, M. and Palmer, T. N.: Ensemble forecasting, J. Comput. Phys., 227, 3515–3539, 2008.

Liu, J., Wang, J., Pan, S., Tang, K., Li, C., and Han, D.: A real-time flood forecasting system with dual updating of the NWP rainfall and the river flow, Nat. Hazards, 77, 1161–1182, https://doi.org/10.1007/s11069-015-1643-8, 2015.

López López, P., Verkade, J. S., Weerts, A. H., and Solomatine, D. P.: Alternative configurations of quantile regression for estimating predictive uncertainty in water level forecasts for the upper Severn River: a comparison, Hydrol. Earth Syst. Sci., 18, 3411–3428, https://doi.org/10.5194/hess-18-3411-2014, 2014.

Mourre, L., Condom, T., Junquas, C., Lebel, T., Sicart, J., Figueroa, R., and Cochachin, A.: Spatio-temporal assessment of WRF, TRMM and in situ precipitation data in a tropical mountain environment (Cordillera Blanca, Peru), Hydrol. Earth Syst. Sci., 20, 125–141, https://doi.org/10.5194/hess-20-125-2016, 2016.

Muggeo, V. M. R., Sciandra, M., Tomasello, A., and Calvo, S.: Estimating growth charts via nonparametric quantile regression: a practical framework with application in ecology, Environ. Ecol. Stat., 20, 519–531, 2013.

Ochoa, A., Pineda, L., Crespo, P., and Willems, P.: Evaluation of TRMM 3B42 precipitation estimates and WRF retrospective precipitation simulation over the Pacific–Andean region of Ecuador and Peru, Hydrol. Earth Syst. Sci., 18, 3179–3193, https://doi.org/10.5194/hess-18-3179-2014, 2014.

Rama Rao, Y. V., Alves, L., Seulall, B., Mitchell, Z., Samaroo, K., and Cummings, G.: Evaluation of the weather research and forecasting (WRF) model over Guyana, Nat. Hazards, 61, 1243–1261, 2012.

R Development Core Team: A language and environment for statistical computing, Tech. rep., R Foundation for Statistical Computing, Vienna, Austria, available at: http://www.R-project.org (last access: 10 April 2015), 2010.

Remesan, R., Bellerby, T., Holman, I., and Frostick, L.: WRF model sensitivity to choice of parameterization: a study of the "York Flood 1999", Theor. Appl. Climatol., 122, 229–247, https://doi.org/10.1007/s00704-014-1282-0, 2014.

Rene, J.-R., Madsen, H., and Mark, O.: Probabilistic forecasting for urban water management: a case study, in: 9th International Conference on Urban Drainage Modelling, 4–6 September 2012, Belgrade, Serbia, 1–11, 2012.

Reyes, O.: Utilización de modelos hidrológicos para la determinación de cuencas en ecosistemas de páramo, Revista Ambiental Agua, Aire y Suelo, 56–65, available at: http://revistas.unipamplona.edu.co/ojs_viceinves/index.php/RA/article/view/432 (last access: 22 August 2015), 2014.

Roberts, N. M., Cole, S. J., Forbes, R. M., Moore, R. J., and Boswellc, D.: Use of high-resolution NWP rainfall and river flow forecasts for advance warning of the Carlisle flood, north-west England, Meteorol. Appl., 16, 23–34, 2009.

Robertson, D. E., Shrestha, D. L., and Wang, Q. J.: Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting, Hydrol. Earth Syst. Sci., 17, 3587–3603, https://doi.org/10.5194/hess-17-3587-2013, 2013.

Rogelis, M. C.: Sistema de alerta temprana del río Tunjuelo 2006, Tech. rep., Fondo de Prevención y Atención de Emergencias, Bogotá, 2006.

Rogelis, M. C.: Operational flood forecasting, warning and response for multi-scale flood risks in developing cities, PhD thesis, Delft University of Technology and of the Academic Board of the UNESCO-IHE Institute for Water Education, Leiden, the Netherlands, 2016.

Rossa, A., Liechti, K., Zappa, M., Bruen, M., Germann, U., Haase, G., Keil, C., and Krahe, P.: The COST 731 Action: a review on uncertainty propagation in advanced hydrometeorological forecast systems, Atmos. Res., 100, 150–167, https://doi.org/10.1016/j.atmosres.2010.11.016, 2011.

Ruiz, J. F.: ¿Como Interpretar Los Modelos De Pronóstico Del Estado Del Tiempo?, Tech. rep., Instituto de Hidrología, Meteorología y Estudios Ambientales, Bogotá, available at: http://bart.ideam.gov.co/wrfideam/GUIA_MODELOS.pdf (last access: 15 September 2014), 2010.

Scardovi, E.: Rainfall spatial predictions: a two-part model and its assessment, PhD thesis, University of Bologna, Bologna, available at: http://amsdottorato.unibo.it/6744/1/scardovi_elena_tesi.pdf (last access: 10 January 2016), 2015.

Sevink, J.: Páramo Andino Project Hydrology workshop in Merida, Venezuela, 2011.

Teng, J., Potter, N. J., Chiew, F. H. S., Zhang, L., Wang, B., Vaze, J., and Evans, J. P.: How does bias correction of regional climate model precipitation affect modelled runoff?, Hydrol. Earth Syst. Sci., 19, 711–728, https://doi.org/10.5194/hess-19-711-2015, 2015.

Theis, S. E., Hense, A., and Damrath, U.: Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach, Meteorol. Appl., 12, 257–268, https://doi.org/10.1017/S1350482705001763, 2005.

Verkade, J. S. and Werner, M. G. F.: Estimating the benefits of single value and probability forecasting for flood warning, Hydrol. Earth Syst. Sci., 15, 3751–3765, https://doi.org/10.5194/hess-15-3751-2011, 2011.

Verkade, J. S., Brown, J. D., Reggiani, P., and Weerts, A. H.: Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales, J. Hydrol., 501, 73–91, https://doi.org/10.1016/j.jhydrol.2013.07.039, 2013.

Vincendon, B., Ducrocq, V., Nuissier, O., and Vié, B.: Perturbation of convection-permitting NWP forecasts for flash-flood ensemble forecasting, Nat. Hazards Earth Syst. Sci., 11, 1529–1544, https://doi.org/10.5194/nhess-11-1529-2011, 2011.

Wilks, D. S.: Forecast Verification, in: vol. 100, 2nd Edn., Elsevier, San Diego, California, USA, 2011.

Xuan, Y., Cluckie, I. D., and Wang, Y.: Uncertainty analysis of hydrological ensemble forecasts in a distributed model utilising short-range rainfall prediction, Hydrol. Earth Syst. Sci., 13, 293–303, https://doi.org/10.5194/hess-13-293-2009, 2009.

Yang, W., Andréasson, J., Phil Graham, L., Olsson, J., Rosberg, J., and Wetterhall, F.: Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies, Hydrol. Res., 41, 211–229, 2010.

Yuan, X. and Wood, E. F.: Downscaling precipitation or bias-correcting streamflow? Some implications for coupled general circulation model (CGCM)-based ensemble seasonal hydrologic forecast, Water Resour. Res., 48, 1–7, 2012.

Yucel, I., Onen, a., Yilmaz, K. K., and Gochis, D. J.: Calibration and evaluation of a flood forecasting system: Utility of numerical weather prediction model, data assimilation and satellite-based rainfall, J. Hydrol., 523, 49–66, https://doi.org/10.1016/j.jhydrol.2015.01.042, 2015.