HESSD

Interactive
comment

Interactive comment on "Streamflow forecasts from WRF precipitation for flood early warning in mountain tropical areas" *by* María Carolina Rogelis and Micha Werner

María Carolina Rogelis and Micha Werner

c.rogeliz@gmail.com

Received and published: 23 November 2017

Response to Referee No 1

We would like to thank the anonymous referee for an extensive review of our

manuscript and for providing us with helpful and constructive comments. For ease of reading we have copied the reviewer comments, as well as our response. We have also attached a track changes version of the manuscript, using as original version the one including the modifications suggested by the Editor.

GENERAL COMMENTS

Overall, convincing case that finer spatial resolution and bias correction does not improve forecast skill of heavy precipitation in regions of tropical convection. However, statistical post processing to create a probabilistic forecast does enhance the ability to forecast a probability of precipitation at short lead times. Given this, I recommend more discussion from a physical perspective of how the probit model method derives it's skill. It seems that this is the first stage (probit model) creates the biggest value in post-processing, rather than the second stage (quantile regression); it may be worth evaluating that.
RESPONSE:
We thank the reviewer for this suggestion. We extended the discussion in section 4.1 (Evaluating precipitation forecasts from the WRF model) as follows:
The AROC of the occurrence of precipitation shows that values higher than 0.5 are obtained in all watersheds for lead times 10 up to 5 hours, 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 hours, in the range of 0.63-0.67. This means that there is some skill in the probit model to forecast occurrence of precipitation for lead times up to 5 hours, 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 hours 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.

SPECIFIC COMMENTS

1. **p4 l20: I infer that you do not use a NWP ensemble (either based on GPS or WRF), but a single "deterministic" GPS-WRF forecast. That should be emphasized here for clarity.**
   RESPONSE: We thank the reviewer for this contribution. The clarification was added in section 2.2, as follows:

   The dataset of WRF forecast used in this paper corresponds to 107 selected days when significant storms were recorded, during the period July 2009-December 2002. For each of these days, deterministic forecasts with the WRF model were generated at 00:00 GTM, 06:00 GTM, 12:00 GTM and 18:00 GTM.

2. **p4 l35: For clarity, reflect the language on p2 l32 here to keep clear the distinction between DBS and probit post processing.**
   RESPONSE: The paragraph was modified as follows:

   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; d) and precipitation forecast ensembles obtained from the post processing of the WRF model, based on probit regression and quantile regression.

3. **p5 l10: The word "pre-processing" is unclear here, as it suggests *before* running WRF; do you mean post-processing?**
RESPONSE: We thank the reviewer for pointing out this mistake. The paragraph was modified as follows:

(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.

4. **p5 l13: As reference values the *observed* (?) time series of mean areal precip**
RESPONSE: The word *observed* was included, as follows:

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.

5. **p5 l20: Need more detail here: what periods were used in WRF and OBS to generate the gamma functions? Just the 107 storm days? This is relevant later to the extent the same periods are used for training and evaluation.**
RESPONSE: A clarification was added as follows:

(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.

6. **p5 l25: A more qualitative and physically based description needed here that does not depend on the reader fully understanding the statistical method. Is information from the neighborhood used for probability to form an ensemble? Below, what are the x covariates physically?**

C4

RESPONSE:
The explanation on the bias correction and generation of ensembles through post processing of the rainfall forecast from the WRF model, was improved as follows:

(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 precipitation 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.

7. **p8 l10: Why include SSclim and SS0 if not discussed?**
   RESPONSE:
   We believe that SSclim and SS0 are very important in the results and discussion sections. The results of these two parameters are presented and explained in section 3.4, -in the case of precipitation-, and in section 3.5 in the case of discharge. The behaviour of SSclim and SS0 supports part of the discussion

in section 4.1. In section 4.2 SSclim supports the discussion on lead times, at which the added value is totally lost in comparison with the climatology. In order to clarify the importance of these parameters, section 4.1 was complemented as follows:

The WRF data used in this study shows 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 SSclim) and providing only a modest improvement in comparison to the zero precipitation forecast (as shown by most values of SS0 being positive, albeit low).

8. **p9 l21: Computed from the 107 selected storm days? Are the CDFs for the same days as used to train the bias correction?**
RESPONSE:
The CDFs were changed to Q-Q plots for clarity, attending a comment from the Editor suggesting a clearer figure. A clarification was added that the figures correspond to the 107 selected storms. Additionally, in the methodology, a clarification was included in section 3.1 and 2.3.1. The modifications are as follows:

Section 3.1:
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.

In addition, a clarification was made in section 2.3.1 in response also to comment 5.

9. **p10: delete empty section 3.3.**
RESPONSE:

We thank the reviewer for pointing out this mistake. Empty section 3.3 was deleted.

10. **p10 l5 *shown* in Eqn 3 to Eqn 14**
RESPONSE:
We thank the reviewer for pointing out this mistake. The word was corrected.

11. **11. p12 l18: I recommend deleting fig 10 since it is barely discussed and expanding discussion of Fig 11.**
RESPONSE:
We believe that Figure 10 (Figure 11 in the attached version of the manuscript) is important for our discussion. The figure is presented in section 3.5. However, in section 4.2 (Evaluating discharge forecast the rank histograms are discussed), the figure is not mentioned explicitly, we believe that this may be the reason why the importance of the figure is not clear. We included a reference to the figure in section 4.2, as follows:

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 Figure 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 is 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 Figure 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.

Regarding Figure 11 (Figure 12 in the attached version of the manuscript) the discussion is included in section 4.1. However, the figure is also not mentioned explicitly in the text of the discussion. The text in the discussion was complemented as follows:

The results of the CRPSS (see Figure 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 hours. 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. E.g. 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.

12. **12. p13 l22: Fig 12 doesn't show this. . . Is there a missing figure?**
RESPONSE:
We agree with the reviewer that the figure showing IDW precipitation vs WRF raw precipitation does not show the reduction of differences between observations and bias corrected WRF precipitation. Figure 12 was eliminated (this change was made attending a suggestion of the Editor) and was replaced in this paragraph by Figure 3 (please see attached version of the manuscript) that shows a Q-Q plot of the observed IDW precipitation vs WRF and WRF bias corrected precipitation, showing clearly the effect of the bias correction.