Reply to referee #1

The referee acknowledges the usefulness of our work and finds the manuscript well presented and discussed. The main complaint is about the density of the paper, which is recommended to be reduced, particularly by shortening/condensing some sections. The referee also raises a comment on the argument to use basin-wide averages of precipitation products in our analysis. Below, we reply to this comment and provide more details on this key decision in our methodology.

We fully agree with his interpretation on the rationality behind the use of basin-wide averages. First, our work aims at performing a meteorological evaluation to understand better the potential use of satellite and model based precipitation products for hydrological applications in different basins of the target region. Hence, the 'spatial unit' for analysis was set at sub-catchment level. The second motivation is that by such spatial averages both products and gauge-based interpolated fields are assessed accounting regional differences in precipitation. We acknowledge that the spatial average implies a degree of arbitrarily as it is based on our 'ad hoc' definition of 'spatial homogeneity'. However, the spatial average is intuitive since it captures the most prominent precipitation features. We have delete the first phrase ("reduce random errors", "appropriate scale", etc) in section 2.4 (HESSD version) and rewritten the introductory sentence of this section to avoid misleading. The section 2.4 (HESSD version) has been re-numbered as 2.3.2 in the revised manuscript.

Concerning the shortening and editing of the manuscript, we have condensed the 'Result' and 'Discussion' sections emphasizing key messages. The shortening of these two sections resulted in a ratio of 0.94 between the revised manuscript and the HESSD version.

Reply to referee #2

The main concerns of the referee are on the 'structure' of the paper and that the 'aim and objectives are not clearly stated upfront'.
Certainly, the goal and motivation of our work arise from the increasing amount of studies that report the use of satellite products and NWP precipitation outputs for water resources applications. Particularly, for the target region, this trend has become more often because the well-known limitations in density of gauge-based rainfall (i.e. Arias-Hidalgo et al., 2013). In such scope, our work performs an evaluation to assess how well satellite-based and model-based products capture spatial and temporal features of precipitation among different basins. This is mandatory before any intake of precipitation products for hydrological studies given the variety of precipitation regimes and mechanisms. We clarify that performing such a tailored evaluation is the intended contribution of our work. We make clear emphasis of this statement throughout all the sections of the revised manuscript.
The referee also claims that 'a justification of the methodological steps is often lacking'. We reply in greater detail to the questions on methodology posted in the 'Specific Comments' and embedded in the 'Technical Corrections' in this response. They have been addressed in the revised manuscript *(See described changes in the revised manuscript).*

Why is not CMORPH part of the rainfall product evaluation?
The starting point of our methodology rests on the identification of what for the target region are promising satellite products and NWP precipitation outputs from which we can detect strengths and deficiencies in rainfall estimation. Among the satellite products the TRMM 3B42 was identified at the forefront. The work by Dinku et al. (2010) over a similar region (the Pacifc-

Andean basin and the Northern coast of Colombia) offered, to the best of our knowledge, a key reference due to the similarities on climatologic features. While comparing several precipitation products Dinku and co-authors reported contrasting results for the referred Colombian regions. For the northern dry littoral these authors found 'estimation skills particularly bad for CMORPH' and better for TRMM 3B42V6 at 10-daily accumulations. The converse was found over the wet western Pacific coast where CMORPH was slightly better especially at daily scale.
(See p. 9, section 2.3.3, L: 268-280)
Obviously, the generality of these results cannot be extrapolated to our study region; however, it provided the starting hint on the choice of the satellite product to evaluate. A second reason why CMORPH was no considered has to do with the length of the dataset. The production of the CMORPH product starts 2nd December of 2002. This has implications that go beyond the setting of a common period for comparison. Omitting the year 1998, one of the wettest years in the record due to El Niño anomalies would have resulted in a sampling error with implications in the comparison between precipitation products.

p. 417 L10-12: Were values outside the time period of interest also excluded from the dataset?
Yes, first the quality control was performed for the long-term records then the period of interest was restricted to the common range between the TRMM products and the WRF retrospective simulation. *(See p.6, section 2.3.1, L: 170-171 and p.7, section 2.3.2, L: 203-205)*

p. 418 L18: TRMM 3B42RT is mentioned but not considered in the comparison. Why?
The focus of our work is no real time applications rather it is a retrospective assessment of precipitation products. The TRMM 3B42 estimates supersede the 3B42RT estimates as each month of 3B42 is computed. The 3B42 processing is designed to maximize data quality, so 3B42 is strongly recommended for any research work not focused on real-time applications (Huffman and Bolvin, 2012). We have removed the reference to the 3B42RT in the introductory

sentence to avoid misleading. We better argue the choice of the 3B42 and upgrade the section 2.3.2 (HESSD version) with relevant literature on validating 3B42 V6 and/or comparing V6 and V7. *(See again p. 9 section 2.3.3, L: 268-280)*

p. 419 L8: Could you not produce a longer WRF simulation (e.g. 1998-2013) to provide say 15-years long time series for this comparison? The North Western Retrospective Simulation (OA-NOSA30) was produced by The Scientific Modelling Centre from Venezuela (CMC) and the National Institute of Meteorology and Hydrology from Ecuador (INAMHI) for the period January 1996 to December 2008 in the scope of the Observatorio Andino project implementation (Muñoz et al., 2010). Since then, the production of a longer simulation has not been carried out. In our study, we restrict to evaluate this product as available.

p. 420 L14: How many GTS stations were excluded from the analysis? And where they are situated (in case that had any effect on the TRMM 3B42 products?
Three GTS stations were identified in our dataset and excluded. The location of all GTS stations (five) in the PAEP region are presented in *Figure 1b* of the revised manuscript.

Whether they were used in the monthly gridded rain gauge analysis to correct the final TRMM product is hardly to asses given the difficulties to identify the dates when they were incorporated in the GTS. Therefore, we exclude all of them to assure the independence of the validation dataset.

p. 424 L28: By what criteria does KED outperform other interpolation techniques?

Table 2 shows that KED performed better in all statistics (Correlation, MSE and Performance). *(See p. 12 section 3.1, L: 362-367)*

p. 430 L7: Have these studies also compared TRMM 3B42 rainfall products, or WRF model outputs as well?
Cheng and Steenburgh (2005) and Ruiz et al. (2010) performed evaluations of WRF over North and South American region respectively. Habib et al. (2009) and Scheel et al. (2011) conducted TRMM evaluations over North American regions and the Central Andes in South America. *(See p.16, section 3.4, L: 498-508)*

p. 434 L18: What is the authors' definition for 'aceptable skills'?
We acknowledge that the term 'acceptable' in the context may become ambiguous. We replace it by 'the best' skills.
(See p.20, section 5, L: 20)

Figure 2: What are the units for the 'uncertainty'?
Total and KED uncertainty are normalized to their maximum values. Both are expressed in percentage. Further, KED uncertainty is scaled as to represent the proportional contribution to the Total uncertainty.
(See Figure 2)

Finally the referee asks for a shortening of the manuscript and thorough revising for clarity and concision. We have condensed sections wherever possible. The shortening resulted in a ratio of 0.97 between the revised manuscript and the HESSD version.

References:

Arias-Hidalgo, M., Bhattacharya, B., Mynett, A. E., and van Griensven, A.: Experiences in using the TMPA-3B42R satellite data to complement rain gauge measurements in the Ecuadorian coastal foothills, Hydrol. Earth Syst. Sci., 17, 2905-2915, doi:10.5194/hess-17-2905-2013, 2013.

Cheng, W. Y. Y. and Steenburgh, W. J.: Evaluation of Surface Sensible Weather Forecasts by the WRF and the Eta Models over the Western United States, Weather Forecast., 20, 812–821, doi:10.1175/WAF885.1, 2005

Habib, E., Henschke, A. and Adler, R. F.: Evaluation of TMPA satellite-based research and real-time rainfall estimates during six tropical-related heavy rainfall events over Louisiana, USA, Atmospheric Research, 94(3), 373–388, doi:10.1016/j.atmosres.2009.06.015, 2009.

Huffman, G. J. and Bolvin D. T.: TRMM and other data precipitation data set documentation (last access: 1 April 2013), 2012.

Muñoz, A., Lopez, P., Velasquez, R., Monterrey, L., Leon, G., Ruiz, F., Recalde, C., Cadena, J., Mejia, R., Paredes, M., Bazo, J., Reyes, C., Carrasco, G., Castellon, Y., Villarroel, C., Quintana, J., and Urdaneta, A.: An Environmental Watch System for the Andean countries: El Observatorio Andino, 91, 20, available at: http://arxiv.org/abs/1006.0926, 2010.

Ruiz, J. J., Saulo, C. and Nogués-Paegle, J.: WRF Model Sensitivity to Choice of Parameterization over South America: Validation against Surface Variables, Monthly Weather Review, 138(8), 3342–3355, doi:10.1175/2010MWR3358.1, 2010.

Scheel, M. L. M., Rohrer, M., Huggel, C., Santos Villar, D., Silvestre, E. and Huffman, G. J.: Evaluation of TRMM Multi-satellite Precipitation Analysis (TMPA) performance in the Central Andes region and its dependency on spatial and temporal resolution, Hydrology and Earth System Sciences, 15(8), 2649–2663, doi:10.5194/hess-15-2649-2011, 2011.