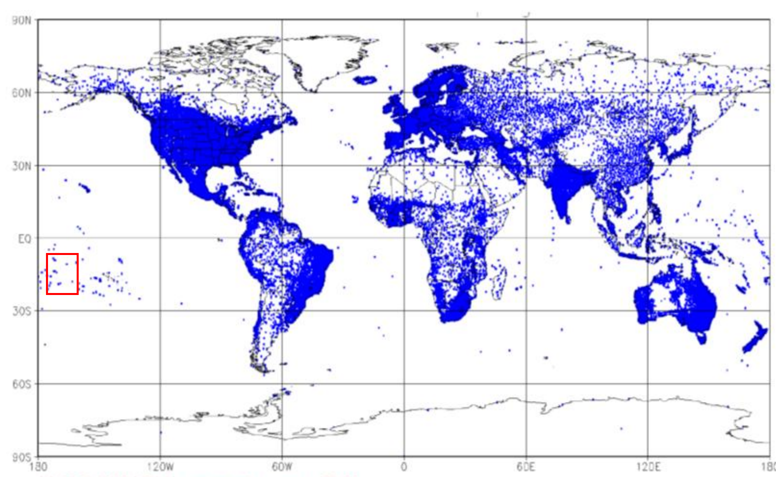
We would like to thank the referee for his useful comments. Please find bellow our answers.

1. P.3 L.22: item b: Does this dataset (TMPA-V7) include information from the rain gauges shown in Fig.1 (all of them, half of them, just a few of them, none of them?). The authors note that TMPA-v7 and GPM-IMERG are the most similar...can the authors discuss a bit why this is so? I assume they use much of the same satellite and rain gauge data...?

GPM is an international US/Japanese Earth science mission involving NASA and JAXA, respectively. The GPM mission improved and expanded on TRMM. GPM and TRMM provide precipitation data derived from different passive microwave (PMW) sources used in IMERG and TMPA, respectively [Huffman et al. 2015], including: Sounder for Atmospheric Profiling of Humidity in the Intertropics by Radiometry (SAPHIR), Advanced Technology Microwave Sounder (ATMS), Atmospheric Infrared Sounder (AIRS), Cross-Track Infrared Sounder (CRIS), and TRMM Combined Instrument (TCI) algorithms (2B31). They also include TRMM Microwave Image (TMI, data ended on 8 Apr 2015), GPM Microwave Imager (GMI), Advanced Microwave Scanning Radiometer for Earth Observing Systems (AMSR-E), Special Sensor Microwave Imager/Sounder (SSMIS), Microwave Humidity Sounder (MHS), Special Sensor Microwave Imager (SSM/I), Advanced Microwave Sounding Unit (AMSU), Operational Vertical Sounder (TOVS) and microwave-adjusted merged geo-infrared (IR).

TMPA 3B42 version 7 is obtained from the preprocessing of data provided by different satellite-based sensors between 1998 and April 2015, in both real and near-real time (TMPA 3b42 data are available at ftp://disc2.nascom.nasa.gov/data/TRMM/Gridded/3B42RT). The 3B42 algorithm (every three hours) combines precipitation estimates from TMI, AMSR, SSMIS, SSM/I, AMSU, MHS, TCI, *MetOp-B* and IR. After the preprocessing is complete, the 3-hourly multi-satellite estimations are summed for the month and combined with monthly rainfall obtained from Global Precipitation Climatology Centre (GPCC), which uses ground-based precipitation. The last step is to scale each 3-hourly rainfall estimate for the month to sum to the monthly value (for each pixel separately, 0.25-degree by 0.25-degree spatial resolution).

Both TMPA and GPM-IMERG adopt the Global Precipitation Climatology Centre (GPCC) monthly rain gauge analysis (Huffman et al. 2014). The Monitoring Product is represented on internationally exchanged meteorological data i.e. gauge observations from world-wide 6,000 to 7,000 stations (see next figure, *(*Schneider et al., 2014). The average gauge density is about 2 gauges per 2.5° by 2.5° lat/long grid box only. Building upon the figure of rainfall stations and lat/long grid *(*Schneider et al., 2015), it very probably that 105 rainfall stations used in our study were considered by GPCC calculations.



*Spatial distribution of monthly in-situ stations with a climatological precipitation normal, based on at least 10 years of data in GPCC data base (*Schneider et al., 2014)

Huffman, G.J.,Bolvin, D.T.,Nelkin, E.J.: Day 1 IMERG Final Run Release Notes; NASA Goddard Earth Sciences Data and Information Services Center: Greenbelt, MD, USA, 2015.

Schneider, U., Becker, A., Finger, P. et al. GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. Theor Appl Climatol (2014) 115: 15. doi:10.1007/s00704-013-0860-x

2. P.3 L.5: Two questions related to details: could the observed rainfall have been interpolated to a 0.25x0.25 grid for better comparison with the TMPA data sets? 0.15x0.15 for the GPM-IMERG dataset? And in regions of mountainous terrain, oftentimes consideration of the altitude-rainfall gradient relationship is critical for spatially distributing rainfall onto a grid. Was this information included in the interpolation? Do the authors think this effect is important in this region?

Initially, observed rainfall was interpolated for 0.5°*0.5°, 0.25°*0.25° and 0.15°*0.15° grids (without significant changes in the rainfall analysis or hydrological model performance). Nonetheless, not to lose more detailed spatial information a 0.10°*0.10° grid over Andean regions (in this paper greater than 1500 masl) was selected for the interpolation process, since there are more rainfall stations close to each other over Andean regions than Amazon regions (in this paper, region lower than 1500 masl).

3. P.4 L.9: MGB uses the rainfall aggregated to a daily time step. So, it seems that one of the main possible advantages of the GPM-IMERG (30 minute time step) datatset compared to the TMPA-V7 (3 hourly) for computing runoff generation by MGB is lost, especially since the convective nature of the rainfall is likely better resolved, in theory at least, using a 30 minute time step. Is there anything in MGB that can take advantage of the diurnal temporal distribution of the rainfall? If not, I think the authors should at least comment that hydrological models using sub-diurnal time steps might have larger differences between the TMPA and GPM-IMERG datasets owing to their different temporal resolutions...Can the authors comment on this?

Thank you for your comment

The improvements in MGB-IPH software do not currently include topics about sub-daily time step (https://www.ufrgs.br/hge/mgb-iph/).

In general, the performance of the model when using GPM-IMERG datasets indicates that these data are useful for estimating observed streamflows in Andean-Amazonian regions (Ucayali basin, southern regions of the Peruvian and Ecuadorian Amazon Basin). These results are similar to those obtained from TMPA V7 estimates by Zubieta et al. [2015] for the 2003-2009 period. Streamflows obtained from GPM-IMERG, TMPA V7, TMPA RT datasets show the same spatial pattern as those obtained by using PLU, (low and high performance in the northern and southern regions of the ABPE, respectively). The ability to represent seasonal streamflows in the southern region using these four precipitation datasets is validated with statistical evaluation.

It is important to note that advantages of GPM-IMERG compared to the TMPA-V7, such as the temporal resolution (30 minutes against 3 hours, respectively), for estimating streamflows have not yet been fully analyzed. The use of sub-daily rainfall data is potentially interesting to simulate discharge variability in the Andean rivers, where short convective rainfall episodes are more relevant for hydrological variability. In this study, precipitation and streamflows were analyzed at the daily time step. Further flash flood modeling at smaller scales would be able to evidence the effects of sub-diurnal differences between datasets

Zubieta, R.,Geritana, A., Espinoza, J.C. and Lavado W.: Impacts of Satellite-based Precipitation Datasets on Rainfall-Runoff Modeling of the Western Amazon Basin of Peru and Ecuador, Journal of Hydrology, doi:10.1016/j.jhydrol.2015.06.064, 2015.

4. P.5 L.1: How were these thresholds selected? Are they based on some sort of statistical analysis?

The Amazon basin of Perú and Ecuador can present different rainfall regimes (Espinoza et al., 2009; Laraque et al., 2007). Rainfall thresholds for the initiation of events such as landslides or floods can be variable in space and time, for example, extreme rainfall (amount) in one Andean region may be normal in Amazon region. In this study, those thresholds are obtained from frequency analysis (percentiles 5, 20, 60, 90, 95).

Espinoza, J.C, Ronchail, J., Guyot, J.L., Cochonneau, G., Filizola, N.P., Lavado, C., De Oliveira, E., Pombosa, R., Vauchel P.: Spatio-Temporal rainfall variability in the Amazon basin countries (Brazil, Peru, Bolivia, Colombia and Ecuador), I.J. of Climatology 29(11): 1574–1594, doi:10.1002/joc.1791, 2009.

Laraque, A., Ronchail, J., Cochonneau, G., Pombosa, R., Guyot, J.L.: Heterogeneous distribution of rainfall and discharge regimes in the Ecuadorian Amazon basin, Journal of Hydrometeorology 8: 1364–1381, doi:10.1175/2007JHM784.1, 2007.

5. P.5 L.25: The text "overestimate observations"...should likely be modified to something like "produce overestimates compared to observations".

Thank you so much, that would be considered:

Total annual rainfall over the ABPE during the selected period is shown in Figs. 1c-f, using all four precipitation products. The satellite-based datasets (GPM-IMERG, TMPA V7 and TMPA RT) produce overestimates compared to observations (PLU) during this period (by 11.1%, 15.7% and 27.7 %, respectively).

6. P.6 L.10-11: The authors have reported that the satellite-based datasets underestimate the dry and wet season rainfall much more for the Huallaga basin compared to the Ucayali basin: do the authors have any insights as to why this is?

The Huallaga basin is not predominantly an Amazon region as is the Ucayali basin. The Andean region (higher than 1500 msal) is more present in the Huallaga basin (51%) than Ucayali basin (39%). The limitations to represent adequate rainfall from the satellite-based precipitation can be due to the strong spatial variability of rainfall in the Amazon-Andes region. Our finding about predominant underestimation in relation to observed rainfall along the Huallaga basin is consistent with others research developed over Andean regions of Peru (Condom et al, 2010), Bolivia (Scheel et al., 2011) and Ecuador (Zulkafli et al., 2014).

Condom, T., Rau, P., Espinoza, J.C. 2011. Correction of the TRMM 3B43 monthly precipitation data over the mountainous areas of Peru during the period 1998–2007. Hydrological Processes.DOI: 10.1002/hyp.7949.

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

Zulkafli, Z., Buytaert, W., Onof, C., Manf, B., Tarnavsky, E., Lavado, W., and Guyot, J. L. 2014. A Comparative Performance Analysis of TRMM 3B42 (TMPA) Versions 6 and 7 for Hydrological Applications over Andean–Amazon River Basins. J. Hydrometeor., 15, 581–592, DOI: 10.1175/JHM-D-13-094.1

7. P.6 L.18-20: While I am not surprised that detection of light events was difficult, why do the products have such difficulty predicting strong rainfall events? I might have (perhaps erroneously or naively I admit) assumed that such events might be better detected. Can the authors comment on this? For example, are the strongest events occurring in high altitude/mountainous regions which are more difficult to detect? Is the smoothing to daily averages related to this problem?

Thank you for the suggestion, these paragraphs would be considered in the manuscript

High or extreme precipitation events can be variable in space and time, rainfall amount for extreme events in one Andean mountain may be normal in Amazon region. This limitation to represent adequate rainfall from the satellite-based precipitation can be due to the strong spatial variability of rainfall in the Amazon-Andes region. Indeed, the AB is distinguished by complex rainfall spatial distribution from the interactions between topography and large-scale humidity transport [Espinoza et al., 2015].

Assessment of rainfall estimates (GPM-IMERG, TMPA V7 and TMPA RT) with respect to PLU have been also perfomed using the Heidke Skill Score (HSS). HSS is a measure of skill in predictions, classified as below normal, near-normal and above-normal (Wilks, 1995). The assesment from HSS is based on the number of correctly predicted data where the category with the largest probability turns out to be correct. As reflected in the formula: $HSS = \frac{C-E}{N-E}$, where C is the number of correct predictions, E is the number of correct predictions expected by chance and N is the total number of predictions. HSS = 1 refers to a perfect prediction, HSS = 0 shows no skill and HSS < 0, indicates that a prediction is worse than a random prediction.

The HSS spatial distribution estimated from daily precipitation using each satellite dataset (GPM-IMERG, TMPA V7 and TMPA RT) and PLU was calculated using thresholds (0.1, 1, 5, 10 and 20 mm/day) as a reference prediction (Fig. S3a-c). In general, for the daily scale, the HSS score varies between 0 and 0.4, indicating low skill. The mean HSS for GPM-IMERG shows a moderate HSS score of around 0.4 in the Northern region (Fig. S3a). The lowest HSS values (lower than 0.2) for GPM-IMERG are mainly located in the Andean regions, where there are more rainfall stations than in the Amazonian regions. This could be due to strong spatial variability, which is characterized by rainfall decrease with altitude and by the leeward or windward position of the stations (Espinoza et al, 2009). Low scores are also observed in more scattered areas along the ABPE when TMPA V7 and TMPA RT are analyzed (lower than 0.15). Nevertheless, this relationship is slightly improved in the northern region of the Ucayali basin (~0.2).

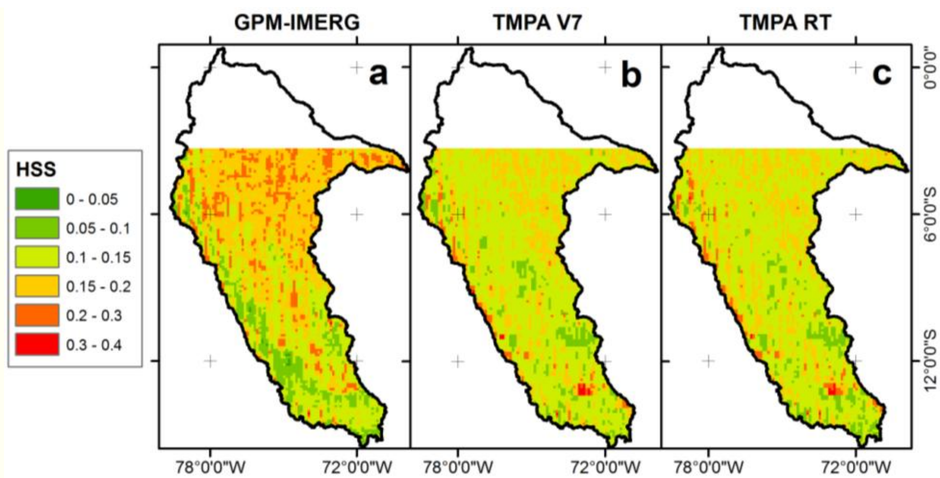


Fig. S3. Spatial variability of the Heidke Skill Score from a) GPM-IMERG, b) TMPA V7 and c) TMPA RT against PLU ground observation, period from 2014 to 2015.

Espinoza, J. C., Chavez, S, Ronchail, J., Junquas, C., Takahashi, K. and Lavado, W., 2015. Rainfall hotspots over the southern tropical Andes: Spatial distribution, rainfall intensity,and relations with large-scale atmospheric circulation. Water Resour.Res. 51, 3456-3475.

8. P.7 L.3: The calibration is glossed over a bit it seems: what set of parameters were calibrated? Were parameters calibrated separately for each precipitation dataset?. Also, no information is given on the quality or calibration of the MGB evaporation. Is it significant compared to the rainfall? Are there non-negligible compensating errors (evaporation bias might offset rainfall or discharge errors/biases)? A short discussion is needed.

To optimize the simulation of streamflows from precipitation datasets, different parameter sets were assigned to each basin in the ABPE during calibration. Analysis by sub-basin is more reliable than assigning the same parameter set to the entire basin [Zubieta et al., 2015]. Based on sensitivity analysis of the MGB-IPH model [Collischonn et al., 2007] six parameters were selected for calibration: Wm_{i} (mm), b_{i} (–), $Kint$ ($\mathrm{mm.d^{-1}}$), $Kbas_{\mathrm{i}}$ ($\mathrm{mm.d^{-1}}$), CS_{i} (–) and CI_{i} (–), where Wm represents water retained in the soil, which influences the evaporation process over time; $Kint$ and $Kbas$ control the amount of water in cases in which subsurface soil and groundwater, respectively, are saturated; and CS and CI allow for adjustment of retention time of flows [Collischonn et al., 2007]. To determine optimal parameters, an automatic calibration process was used in order to reduce the domain extent; a previous manual adjustment of the values was performed. To ensure impartiality, parameter sets were calibrated separately for each precipitation dataset. for each parameter value, different domains were considered initially, in which a first value determined by manual calibration was defined as the relative centroid for each domain. The MOCOM-UA multi-criteria global optimization algorithm [Yapo et al., 1998] was then used to find optimal solutions for six parameters. This process results in an effective and efficient search on the Pareto optimum space [Boyle et al., 2000]. To analyze the impacts on the calibrated parameters, average parameters were calculated for precipitation datasets and HRU (Table 4).

The results of the calibration process indicate that overestimation by TMPA RT compared to observed rainfall (PLU), GPM-IMERG and TMPA V7 (Fig. 2a) in several months is consistent with a mean increase in Wm (+53%, +6%, +15% respectively), along with a predominantly mean decrease in $Kbas$ (-18%, -39% and -16% respectively) and $Kint$ (-25%, -15%, +2%) to achieve water balance (Table 4). Meanwhile, the overestimation by PLU compared to GPM-IMERG, TMPA V7 and TMPA RT (Fig. 3a) is consistent with a mean increase in Wm (+33%, +38%, +34% respectively), along with a mean decrease in $Kbas$ (-30%, -28% and -38% respectively) and $Kint$ (-17%, -16%, -17%) to achieve water balance (Table 4).

Table 3. Model parameters subjected to the process of automatic calibration for the Peruvian and Ecuadorian Amazon basin.

Parameter	HRU	Hydrological process	First guess	Domain
Wm(mm)	Shrubs, agricultural areas/not deep soils	Water storage on the HRU	200	50-1200
	Shrubs, agricultural areas/deep soils		400	50-1200
	Forest/not deep soils		350	50-1200
	Forest/deep soils		600	50-1200
	Pasture/not deep soils		120	50-1200
	Pasture/deep soils		240	50-1200
Kint(mm/d)	Shrubs, agricultural areas/not deep soils	Sub - surface flow	80	50-150
	Shrubs, agricultural areas/deep soils		90	50-150
	Forest/not deep soils		100	50-150
	Forest/deep soils		120	50-150
	Pasture/not deep soils		70	50-150
	Pasture/deep soils		80	50-150
Kbas(mm/d)	Shrubs, agricultural areas/not deep soils	Groundwater flow	30	10 - 100
	Shrubs, agricultural areas/deep soils		50	10 - 100
	Forest/not deep soils		70	10 - 100
	Forest/deep soils		80	10 - 100
	Pasture/not deep soils		55	10 - 100
	Pasture/deep soils		70	10 - 100
CS	All	Surface flow	15	0.35 - 40
CI(-)	All	Sub-surface flow	120	1 - 200

b(-)	All	Variable infiltration curve	0.12	0.01 - 2

Table 4. Values of the model mean parameters used in the Ucayali and Huallaga basins for each rainfall datasets for the 2014-2015 period.

		UCAYALI BASIN				HUALLAGA BASIN			
Parameter	HRU	PLU	GPM-IMERG	TMPA V7	TMPA RT	PLU	GPM-IMERG	TMPA V7	TMPA RT
Wm(mm)	Shrubs, agricultural areas/not deep soils	268	351	294	373	100	60	65	60
	Shrubs, agricultural areas/deep soils	340	472	503	597	132	102	96	99
	Forest/not deep soils	300	408	273	344	130	101	99	96
	Forest/deep soils	422	453	445	435	250	203	180	209
	Pasture/not deep soils	144	350	261	321	101	60	66	59
	Pasture/deep soils	196	400	454	496	150	120	116	121
Kint	Shrubs, agricultural areas/not deep soils	141	216	151	151	190	161	163	152
(mm/d)	Shrubs, agricultural areas/deep soils	180	236	156	163	220	189	195	198
	Forest/not deep soils	198	123	107	108	103	162	155	160
	Forest/deep soils	200	134	108	113	120	208	199	220
	Pasture/not deep soils	150	110	119	122	121	160	151	150
	Pasture/deep soils	180	113	126	128	132	193	201	190
Kbas	Shrubs, agricultural areas/not deep soils	103	121	89	93	55	70	72	80
(mm/d)	Shrubs, agricultural areas/deep soils	113	123	100	103	61	90	94	100
	Forest/not deep soils	53	134	59	53	44	70	69	80
	Forest/deep soils	62	25	69	62	63	90	88	100
	Pasture/not deep soils	64	112	66	64	46	70	76	80
	Pasture/deep soils	74	113	71	71	63	90	66	100
CS	All	18	16	17	17	2.6	2.4	2.6	2.5
CI(-)	All	112	111	118	111	111	133	135	132
b(-)	All	0.13	0.17	0.15	0.12	0.12	0.15	0.14	0.14

Yapo, P.O., Gupta, H.V., Sorooshian, S.:Multi-objective global optimization for hydrologic models. Journal of Hydrology 204, 83–97, 1998.

Boyle, D.P., Gupta, H.V., Sorooshian, S.: Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods. Water Resources Research 36 (12), 3663–3674, 2000.

Collischonn, W., Allasia, D.G., Silva, B.C., Tucci, C.E.M. : The MGB-IPH model for large-scale rainfall-runoff modeling, J. Hydrol. Sci. 52, 878–895, doi: 10.1623/hysj.52.5.878, 2007.

9. P.9 L.8-9: The authors state that seasonal streamflows in the southern region are well modeled using the satellite datasets, and indeed the results support this conclusion. But in the northern part of the Western Amazon basin/region, the results seem to indicate that satellite products are not useful for obtaining streamflows from hydrological modeling: so this implies that further progress is still required.. It think this should also be stated in the conclusions.

Thank you so much, this paragraph would be included

In general, the performance of the model when using the GPM-IMERG dataset indicates that these data are useful for estimating observed streamflows in Andean-Amazonian regions (Ucayali basin, southern regions of the Peruvian and Ecuadorian Amazon Basin). These results are similar to those obtained from TMPA V7 estimates by Zubieta et al. [2015] for the 2003-2009 period. Streamflows obtained from the GPM-IMERG, TMPA V7 and TMPA RT datasets show the same spatial pattern as those obtained by using PLU (low and high performance in the northern and southern regions of the ABPE, respectively). The ability to represent seasonal streamflows in the southern region using these four precipitation datasets is validated with statistical evaluation. Low performance of the model identified in the northern region is mainly related to the lack of adequate rainfall estimates, because it is consistent with estimated streamflows, so this implies that further progress is still required in satellite estimates of rainfall.

Zubieta, R.,Geritana, A., Espinoza, J.C. and Lavado W.: Impacts of Satellite-based Precipitation Datasets on Rainfall-Runoff Modeling of the Western Amazon Basin of Peru and Ecuador, Journal of Hydrology, doi:10.1016/j.jhydrol.2015.06.064, 2015.

10. P.9 L2.: A 20% detection rate seems low. Can the authors put this into some sort of context for the reader (e.g. is 20% indeed a reasonable value in this region, or would one hope to have 50%? or a higher value?). Also, what are the implications of these statistics and their impact on the MGB model simulations? Can a lower, say 10% detection rate, be assumed to be able to produce reasonable Nash scores? Or perhaps there is no clear relationship between these scores and the modeled discharge quality? This is not clear. Can the authors comment on this?

Thank you, a paragraph would be included in manuscript

Analysis of rain events from pixel value comparing PLU and estimated daily rainfall (GPM-IMERG, TMPA V7 and TMPA RT) suggests a low capacity for detection. This does not imply that they are not useful for hydrological modeling, because rain events not correctly detected for a region or a day could be correctly detected on another day or in nearby regions, compensating for the estimation of rainfall amount over large regions.

11. P10 L.15-19: In Fig.4d, it is seen that the poorest Nash scores are in the northern part of the region shown: looking at Fig.4a (or Fig.1b), we see that there are relatively few observations in this region. But, this is where one would hope to benefit the most from a satellite product, but it seems this is not the case. In L.15-19 it is stated that such products hold promise for operational applications in data sparse regions, but it doesn't seem to be the case here? Can the authors comment a bit more or perhaps modify this slightly (seemingly) over-optimistic text?

I am sorry, you are right. The comment would be modified

Their usefulness in Andean-Amazon basins and their applicability as input to hydrological models have been evaluated recently by comparing modeled and observed datasets. Results indicate that these datasets could be used for operational applications in some Andean-Amazon *regions* [Zulkafli et al., 2014; Zubieta et al., 2015]. However, hydrological modeling using satellite-based precipitation data does not yield successful results in equatorial regions. This is mainly because of inadequate satellite estimates, because streamflows resulting from hydrological modeling using observed rainfall show acceptable performance in the Napo River basin in the equatorial region [Zubieta et al., 2015].

Zubieta, R.,Geritana, A., Espinoza, J.C. and Lavado W.: Impacts of Satellite-based Precipitation Datasets on Rainfall-Runoff Modeling of the Western Amazon Basin of Peru and Ecuador, Journal of Hydrology, doi:10.1016/j.jhydrol.2015.06.064, 2015.

Zulkafli, Z., Buytaert, W., Onof, C., Manf, B., Tarnavsky, E., Lavado, W., and Guyot, J. L.: A Comparative Performance Analysis of TRMM 3B42 (TMPA) Versions 6 and 7 for Hydrological Applications over Andean–Amazon River Basins, J.Hydrometeor., 15, 581–592, doi: 10.1175/JHM-D-13-094.1, 2014.