Authors' answers to Reviewer's comments

We would like to thank the Reviewer for the comments and suggestions that helped improving the quality of the manuscript. Below is an item-by-item reply to the Reviewer's comments and suggestions. The Reviewer's comments are reported in italic (1), the answers to the Reviewer are reported in blue (2), and the modifications to the manuscript are reported in green (3).

The aim of this paper is to assess the characteristics of a suite a ground-based (GHCN-Daily, PRISM and Stage IV) and space-borne (TMPA) QPE products over the CONUS over a 10-years period at various time scales. Rain gauge observations and PRISM are taken as a reference for the evaluation of remote sensed products at the annual, seasonal, and daily scales over River Forecast Centers (RFCs). It is shown that if all products present similar annual average accumulation over the CONUS, discrepancies appear at the RFC scale in particular with 3B42RT. The gauge correction in 3B42 often mitigates these biases compared to 3B42RT at the annual and seasonal scale except for the Western US. Extreme daily precipitation is shown to be challenging to retrieve for all remote sensing products.

The research area of using ground and satellite data to obtain good quality distributed rain estimations at the global scale recovers a real need for a number of applications ranging from climate analysis to prediction of floods.

The paper is clear and structured, but the first part is hard to follow with numerous statistical numbers and too few interpretation of the results. The novelty of the work needs to be better highlighted. The results are not always or poorly explained. Keys aspects in the comparison methodology are not well described. Recommendation is "major revision". I would recommend that the authors address carefully the points mentioned below for the manuscript to be ultimately accepted for publication.

1. Title: the authors should include the time scales they are considering (daily to annual). One important value of TMPA and Stage IV is that these datasets are available at sub-daily scale. This is not evaluated here.

We welcome the suggestion. The title will be changed into:

Evaluation of precipitation estimates over CONUS derived from satellite, radar, and rain gauge datasets at daily to annual scales (2002-2012)

2. As a basis for this analysis, the GHCN-Daily gauges are is used as a reference to remote sensed precipitation products. It is probably not suitable since some of these products (3B42, Stage IV) ingest gauge correction over the CONUS and are not mutually independent. This is particularly highlighted with some extremely high correlation coefficients (e.g. p.11496 l. 24, p.11501 l.29). In this context what is the contribution of this analysis, and can it be better explicated?

At the locations specified by the Reviewer (p. 11496, l.24, p. 11501, l. 29), the correlations observed between GHCN-D and PRISM are indeed mentioned. For PRISM, this high correlation was "*expected due to the fact that PRISM gridded precipitation estimates incorporate GHCN-D stations*". Earlier in the text, we also noted that (p. 11494, l. 1-4) that "*The PRISM precipitation estimates incorporate surface data observations from GHCN-D among others and the systematic comparison of point surface observations from GHCN-D and gridded estimates from PRISM will be performed as a consistency check.*" and that (p. 11494, l. 4-6) "*The PRISM precipitation estimates will be used as a baseline data set to evaluate remotely sensed precipitation products (Stage IV, 3B42, 3B42RT) at the annual and seasonal scale.*".

If we think, that for PRISM this point was sufficiently developed, for the other products that are being evaluated (3B42, Stage IV), we welcome the request for a better clarification.

For 3B42, the monthly gauge adjustment uses the precipitation gauge analysis from the Global Precipitation Climatology Centre (GPCC) (Huffman and Bolvin 2013). The gauge analysis used is the GPCC Monitoring Product at 1 deg grid resolution (Schneider et al., 2011). This specific analysis uses SYNOP (synoptic weather observation reports) and CLIMAT reports that are received near-real time from 7000-8000 automated stations worldwide. This number of 7000-8000 stations worldwide (GPCC Monitoring product) is to be compared with approximately over 20000 GHCN-D stations for CONUS alone (Menne et al. 2012). Furthermore, while GHCN-D incorporates surface observations from different sources (see Table 2 in Menne et al. 2012), we selected the subset from the US-Cooperative Observing network (US-COOP), which represented about 9000 stations. The US-COOP network includes first order stations (1600 manual and automatic synoptic stations) and stations from volunteer observers. We further reduced this number by selecting only the stations that reported at least 90% of the time during the period 2002-2012 (4075 stations). In that context, if it is possible that some of the first order automated synoptic stations included in GHCN-D are also used in the GPCC gauge analysis (SYNOP), most of the GHCN-D stations used for evaluation are not a part of the GPCC Monitoring Product used in the 3B42 adjustment (please note that GPCC proposes Full Data Reanalysis Products which might possibly incorporate more of the GHCN-D stations including the US-COOP stations from the volunteer observers).

For Stage IV, the in-situ data incorporation depends on each River Forecast Center (RFC). Briefly, Stage IV represents the final stage of the process that combines mosaicked estimates from the 12 RCFs (Lin and Mitchel 2005). Stage IV analysis are produced for hourly, and 6-hourly totals. The 6-hourly totals are combined into daily accumulation (12Z-12Z). The gauges used at the RFC level include available hourly rain gauges such as HADS (Hydrometeorological Automated Data System) gauges, ASOS (Automated Surface Observing System), and AWOS (Automated Airport Weather Stations) reports (Hou et al. 2014). Furthermore, some of the Western RFCs (California-Nevada: CNRFC, Northwestern: NWRFC) do not use radar estimates due to poor coverage over mountainous areas. The RFCs use an automated analysis of rain gauge observations (Mountain Mapper) that incorporates the gridded monthly precipitation climatology from PRISM (Parameter-elevation Regressions on Independent Slopes Model). Changes in gauge-adjustments procedures are always possible at the RFC level by using the best possible in-situ data observations. It is therefore possible that some GHCN-D stations could be incorporated in the bias-adjustment but it is not known for certain. What we do know is that the SERFC does not use any of the GHCN-D stations (Nelson et al. 2010). In that context, except PRISM that incorporates GHCN-D data including COOP stations, it is reasonable to assume that the HADS, ASOS, and AWOS observations used in the Stage IV bias-adjustment procedure are different from the GHCN-D stations used for the remotely sensed products evaluation.

In both cases (Stage IV, 3B42) it is virtually impossible to track down and identify the stations that are or aren't used in the bias-adjustment procedure. Therefore, a high correlation might or might not indicate that a particular or a group of station is included in that procedure. While we are fully aware, that this point might constitutes a limitation of the present study, there aren't many alternative at hands due to the lack of an alternate consistent long-term dense rain gauge network that could be used in the evaluation of the different precipitation products. However, being aware of this challenge, we performed an assessment of Stage IV using the US Climate Reference Network (USCRN). The USCRN consists in 114 surface stations for the contiguous US with about a decade (since 2002) worth of observations for the first installed stations (Diamond et al. 2013). This analysis is presented in Nelson et al. (2015). Results showed that for the unconditional rainfall rate (mm/day), comparison of Stage IV with USCRN in situ data displayed satisfying to strong statistic (R^2>0.85) for the different RFCs.

To summarize, the fraction of rain gauge intervening both in the evaluation (GHCN-D) and the adjustment procedure (3B42, Stage IV) appear limited and shouldn't prevent GHCN-D and the COOP subset used to be suitable to perform the radar/satellite products evaluation over the period 2002-2012.

To account for this point, we will add the following, in the datasets descriptions in section 2 (GHCN-D, Stage IV, 3B42). For GHCN-D (p. 11493, l. 10):

The GHCN-D datatset incorporates surface observations from different sources (see Table 2 in Menne et al. 2012). We selected the subset from the US-Cooperative Observing network (US-COOP), which represented about 9000 stations. The US-COOP network includes first order stations (1600 manual and automatic synoptic stations) and stations from volunteer observers.

For Stage IV (p. 11494, l. 17):

Stage IV represents the final stage of the process that combines mosaicked estimates from the 12 RCFs. The gauges used at the RFC level for bias-adjustment include available hourly rain gauges such as HADS (Hydrometeorological Automated Data System) gauges, ASOS (Automated Surface Observing System), and AWOS (Automated Airport Weather Stations) reports (Hou et al. 2014). Furthermore, some of the Western RFCs (Colorado Basin River: CBRFC, California-Nevada: CNRFC, Northwestern: NWRFC) do not use radar estimates due to poor coverage over mountainous areas. Those RFCs use an automated analysis of rain gauge observations (Mountain Mapper) that incorporates the gridded monthly precipitation climatology from PRISM. Although changes in gauge-adjustments procedures are always possible at the RFC level including the incorporation of the best available in-situ observations, it is reasonable to assume that the in-situ observations used in the Stage IV bias-adjustment procedure (HADS, ASOS, AWOS) are different from the US-COOP subset of GHCN-D used for the evaluation. For the Western RFCs however, the incorporation of the PRISM climatology that uses GHCN-D in-situ data will have to be kept in mind for any analysis.

For 3B42 and 3B42RT (p. 11495, l. 14):

The gauge analysis used is the GPCC Monitoring Product at 1 deg grid resolution (Schneider et al., 2010, 2011). This specific analysis uses SYNOP (synoptic weather observation reports) and CLIMAT reports that are received near-real time from 7000-8000 automated stations worldwide. While it is possible that some of the first order automated synoptic stations included in GHCN-D are also used in the GPCC gauge analysis (SYNOP), most of the US-COOP subset of the GHCN-D stations used for evaluation are not a part of the GPCC Monitoring Product used in the 3B42. Being virtually impossible to track down and identify the automated stations that are or aren't used in the bias-adjustment procedure for 3B42, we are confident that this number remaining relatively low will not compromise the independent assessment of the 3B42 dataset.

References:

Diamond, H. J., Karl, T. R., Palecki, M. A., Baker, B. C., Bell, J. E., Leeper, R. D., Easterling, D. R., Lawrimore, J. H., Meyers, T. P., Helfert, M. R., Goodge, G., and Thorne, P. W.: U.S. Climate Reference Network after one decade of operations: status and assessment. *Bulletin of the American Meteorological Society*, 94, 424-498, 2013.

Hou, D., Charles, M., Luo, Y., Toth, Z., Zhu, Y., Krzysztofowicz, R., Lin, Y., Xie, P., Seo, D.-J., Pena, M., and Cui, B.: Climatology-Calibrated Precipitation Analysis at Fine Scales: Statistical Adjustment of Stage IV toward CPC Gauge-Based Analysis. *Journal of Hydrometeorology*, 15, 2542-2557, 2014.

Huffman, G. J., and Bolvin, D. T.: TRMM and Other Data Precipitation Data Set Documentation, Lab. for Atmos., NASA Goddard Space Flight Cent. and Sci. Syst. and Appl. Inc., 2013. [Available at: ftp://precip.gsfc.nasa.gov/pub/trmmdocs/3B42_3B43_doc.pdf, accessed 28 Jan. 2015].

Nelson, B. R., Seo, D.-J., and Kim, D.: Multisensor Precipitation Reanalysis. *Journal of Hydrometeorology*, 11, 666-682, 2014.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., and Ziese, M.: GPCC Monitoring Product: Near Real-Time Monthly Land-Surface Precipitation from Rain-Gauges based on SYNOP and CLIMAT data. DOI: 10.5676/DWD_GPCC/MP_M_V4_100; http://dx.doi.org/10.5676/DWD_GPCC/MP_M_V4_100, 2011.

Schneider, U., Becker, A., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: Global Precipitation Analysis Products of the GPCC. Global Precipitation Climatology Centre, DWD, 12 pp., 2010. [Available at ftp://ftp-anon.dwd.de/pub/data/gpcc/PDF/GPCC_intro_products_2008.pdf, accessed 28 Jan. 2015]

3. "Seldom studies that deal with the long-term assessment of precipitation products (annual or multi-annual basis) are available in the scientific literature" (p. 11492 ll.9-11): could you please cite some studies? E.g. Chen, S., et al. (2013), Evaluation of the successive V6 and V7 TRMM multisatellite precipitation analysis over the Continental United States, Water Resour. Res., 49, 8174–8186, doi:10.1002/2012WR012795.

The reference will be added.

Chen, S., Hong, Y., Gourley, J. J., Huffman, G. J., Tian, Y., Cao, Q., Yong, B., Kirstetter, P.-E., Hu, J., Hardy, J., Li, Z., Khan, S. I., and Xue, X.: Evaluation of the successive V6 and V7

TRMM multisatellite precipitation analysis over the Continental United States, *Water Resour. Res.*, 49, 8174–8186, doi:10.1002/2012WR012795, 2013.

4. The impact of the resolution of these various products on the comparison results needs more discussion. The limitations of point rain gauge measurements for evaluation of area-averaged precipitation estimates have been documented for years by Ciach and Krajewski (1999), Ciach et al, (2003), Habib et al (2004) to cite a few. This issue is acknowledged in recent studies using gauges to evaluate satellite precipitation estimates (e.g. Kirstetter et al. 2013).

References:

Ciach, G.J., Habib, E., and Krajewski, W.F., 2003. Zero-covariance hypothesis in the error variance separation method of radar rainfall verification. Adv. Water Resour., 26, 573–580.

Ciach, G.J., and Krajewski, W.F., 1999. On the estimation of rainfall error variance. Adv. Water Resour., 2, 585–595.

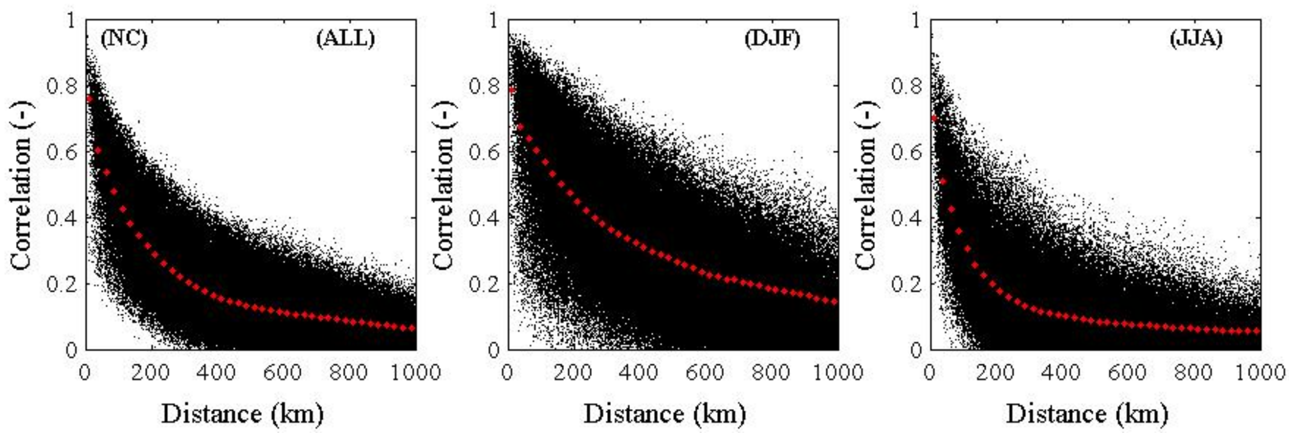
Ciach, G.J., Krajewski, W.F., and Villarini, G., 2007. Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data. J. Hydrometeorol., 8, 1325–1347.

Kirstetter, P.-E., Viltard, N. and Gosset, M. (2013), An error model for instantaneous satellite rainfall estimates: evaluation of BRAIN-TMI over West Africa. Q.J.R. Meteorol. Soc., 139: 894–911. doi: 10.1002/qj.1964

The difference in spatial resolution of the different products makes the comparison between sensors of different resolution delicate. Numerous studies have investigated the limitations of using rain gauge point measurements to evaluate area-averaged rainfall retrieved from sensors with coarser spatial resolution (Ciach and Krajewski 1999, Ciach et al. 2003, Habib et al. 2004, Ciach et al. 2007). Those differences can be due to the small-scale variability of rainfall averaged over the radar pixel and to deferring resolution between the gauge and the radar (Ciach and Krajewski 1999). For shorter accumulation period (from 5-min to several hours), the random sampling errors due to resolution differences can be dominant (Ciach and Krajewski 1999). With increasing accumulation period, the sampling errors decrease and are significantly reduced at the daily scale and beyond.

The comparisons presented in the paper at the annual (Figs 3-4) and seasonal scales (Figs 6-8) correspond to annual and seasonal accumulation over the 11-year period considered. Therefore, they correspond to long-term precipitation accumulation where the impact of the differing resolution is not significant. Previous studies have shown that for longer accumulation period (10-days) the correlation distance, that is the maximum distance between stations beyond which the correlations become insignificant, is of the order of several hundred kilometers (Gutowski et al. 2003). Those distances are several order of magnitude greater than the sensors spatial resolution.

However, when looking at the daily accumulations (section 5), the impact of the different resolutions needs to be considered. To investigate the impact of the resolution of each sensor, we computed for each RFC the interstation correlation of daily precipitation events. For each station, we selected the daily values greater than zero and computed the correlations regardless of the daily values of the other stations for the different RFCs. The correlations were computed for the annual and seasonal scales (DJF, JJA). The average correlation as a function of the distance was averaged over bins of 25-km. The figure below displays an example for the NCRFC, which represents a median situation in term of average rain rates (2.42mm/day). The red dotted lines represent the average values obtained for 25-km intervals.



The table below reports the values of the correlation distance (or distance beyond which the correlations become insignificant and defined as the distance corresponding to an average correlation of 1/e ≈ 0.37).

RFC	Average correlation distance (km) (distance at 1/e)		
	YEA	DJF	JJA
ABRFC	110	260	60
CBRFC	*60*	*140*	*30*
CNRFC	180	210	130
LMRFC	160	230	40
MARFC	210	380	80
MBRFC	110	210	60
NCRFC	130	310	80
NERFC	230	380	130
NWRFC	130	160	80
OHRFC	160	310	60
SERFC	110	210	40
WGRFC	80	210	60

We note that the correlation distance varies greatly from RFC to RFC and from season to season. At the daily scale considered, the averaged correlation distance is important and is greater than the spatial resolution of the radar (4-km) and the satellite (25-km). The longest correlation distances are found during winter compatible with the fact that cold precipitation is characterized by more widespread stratiform precipitation while warm season precipitation is characterized by more localized convective precipitation events. This seasonal dependency is comparable with results obtained by others over the central US and other time scales (Gutowski et al. 2003, Ciach et al. 2007). The CBRFC (Colorado Basin River RFC) is the one that displays the shorter correlation distance comparable with the satellite footprint (25-km), while the NERFC (Northeast RFC) displays the longer correlation distances.

The above discussion will be synthetized in the revised manuscript. We will add (p. 11496, l. 20):

To compare the different estimates for the annual average precipitation, we make the assumption that each rain gauge represents with sufficient accuracy the area-averaged rainfall over the native resolution of the different products evaluated: PRISM and Stage IV (4x4-km^2) and 3B42 and 3B42RT (4x4-km^2). While, there are well known limitations of using rain gauge point measurements to evaluate area-averaged rainfall retrieved from sensors with coarser spatial resolution (Ciach and Krajewski 1999, Ciach et al. 2003, Habib et al. 2004, Ciach et al. 2007), the random sampling errors due to differing resolutions are mostly dominant at the sub-daily scales (Ciach and Krajewski 1999). For accumulation period of several days the correlation distance (maximum distance between stations beyond which the correlations become insignificant), is of the order of several hundred kilometers (Gutowski et al. 2003). Those distances are several order of magnitude greater than the sensors spatial resolution.

For the daily precipitation section (section 5), we will add (p. 11508, l. 13):

We will assume that the rain gauge is representative of the grid-averaged rainfall for Stage IV. The computation of the interstation correlation for daily events indicated that the correlation distance was greater than the 4-km spatial resolution of the radar (not shown).

We will also add the references mentioned (Ciach and Krajewski 1999, Ciach et al. 2003, 2007, Gutowski et al. 2003, Habib et al. 2004). See also answer to Reviewer #1.

References:

Gutowski Jr., W. J., Decker, S. G., Donavon, R. A., Pan, Z., Arritt, R. W, and Takle E. S.: Temporal–Spatial Scales of Observed and Simulated Precipitation in Central U.S. Climate. *Journal of Climate*, 16, 3841-3847, 2003.

Habib, E., Ciach, G. J, and Krajewski, W. F.: A method for filtering out raingauge representativeness errors from the verification distributions of radar and raingauge rainfall. *Adv. Water Resour.*, 27, 967–980, 2004.

5. p. 11494 ll.21-22: "The reader will find a more detailed description of the Stage IV precipitation estimates generation from the RFC level and up to the final mosaicked product as well as related artifacts and uncertainties in Nelson et al. (2014)". The present paper is dealing with such artifacts and uncertainties, so would you mind remind them even briefly here?

The following will be added (p. 11494, l. 19):

The reader will find a more detailed description of the Stage IV precipitation retrievals from the RFC level and up to the final mosaicked product in Nelson et al. (2015). In addition to radar only reflectivity scanning and processing (beam blockage, hot and cold biases, bright-band contamination, anomalous propagation, cone of silence), the final mosaicked estimates present biases that are visible in the long-term averages. The fact that not all the RFCs use the same *precipitation estimation algorithm generates radar-to-radar and RFC-to-RFC discontinuities (Nelson et al. 2015).*

6. Could the authors provide some explanations for the seasonal bias adjustment on 3B42 described on p.11500 ll.22-26?

We realize that the sentence was confusing. In this case, the "bias-adjustment" is referring to the differences between 3B42RT (multi-satellite product) and the 3B42 (satellite-gauge product) and the incorporation of the monthly gage analysis (see the cited reference: Huffman and Bolvin 2013). Therefore, the section p. 11500 l. 22 to p. 11501 l. 3 will be modified into:

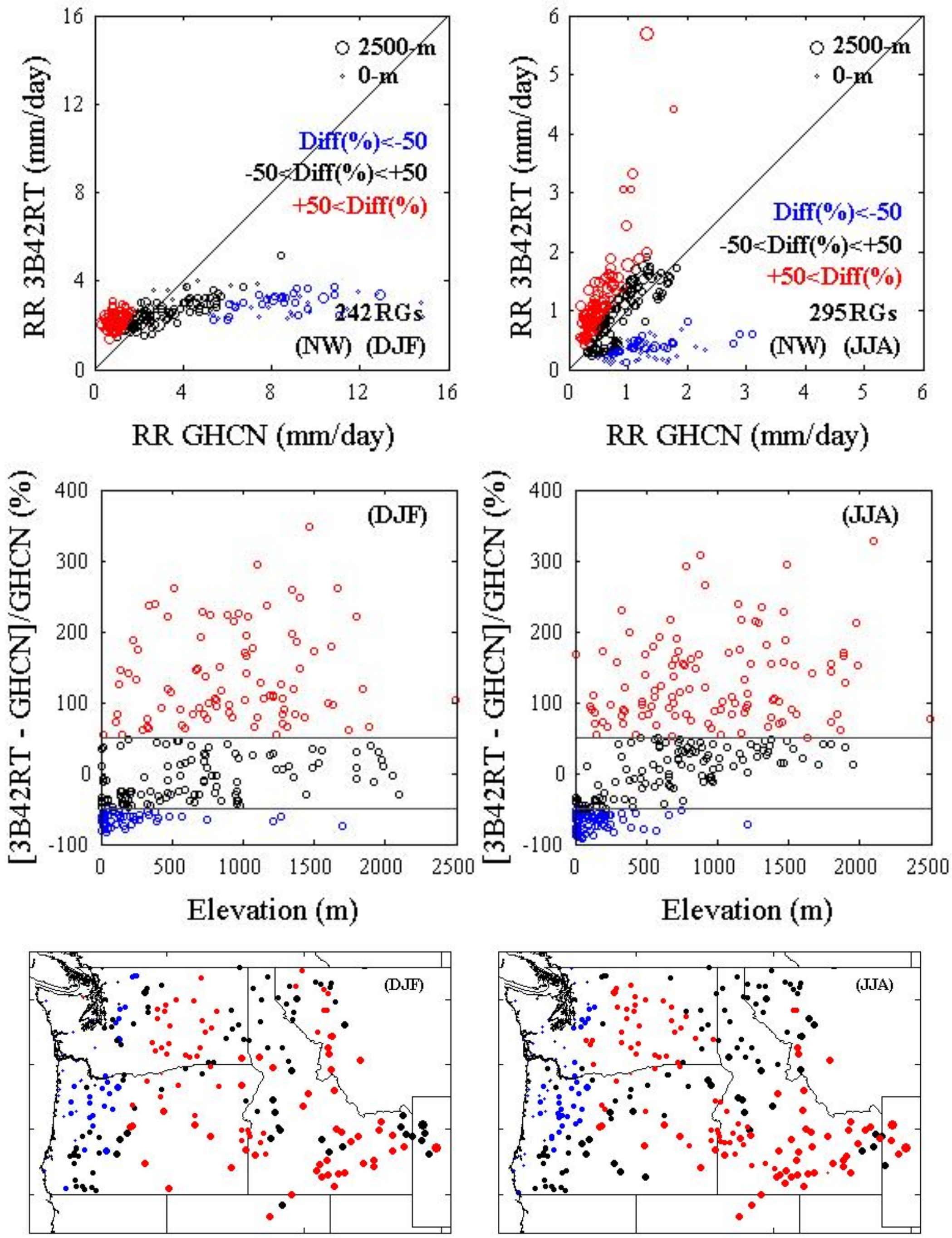
For winter, the bias-adjusted 3B42 precipitation estimates are lower than 3B42RT (3B42<3B42RT) over the Rockies (CB), over the highest latitudes along the US/Canadian border (NC, MB, NW), and East of the Mississippi (LM, SE). Conversely, the 3B42 estimates are found higher than the near-real time 3B42RT (3B42>3B42RT) along the West coast from Northern California up to the Pacific Northwest (NW, CN). For summer, the 3B42 estimates are found very significantly lower than 3B42RT (3B42<3B42RT) over the Midwest (MB, NC, AB). The rain gauge adjustment performed retrospectively corrects the possible overestimation of summertime convection by PMW sensors that mistake sub-cloud evaporation for precipitation (Dinku et al. 2010, 2011, Ochoa et al. 2014). Similarly, for the area of the Lower Mississippi domain (LM) located East of the Mississippi, 3B42 estimates are lower than 3B42RT.

7. Sections 3 and 4 mainly describe the results. Explanations for the observed discrepancies between the products, when provided, are simply listed randomly as in p.11502 ll. 24-29. More constructive and structured analysis and interpretation is necessary. What contribution does these sections bring to the state of knowledge?

We have modified substantially the sections 3 and 4 for clarity purposes and their respective contribution is better emphasized in the revised version. The revised sections included answers from comments #4, #6, #7, #9, #10, and #11.

8. On p. 11502 ll. 19-20: "The real-time 3B42RT displays moderate positive biases when compared to GHCN-D (+2.4%) and PRISM (+0.4%)". The scatterplot on Fig. 7b for 3B42RT shows much more discrepancies that compensate each other. Why is that?

Indeed, we noticed this point. While the bias-adjusted version 3B42 display points homogeneously scattered along the diagonal y=x, the real time 3B42RT presents more discrepancy (strong under/over estimation) that compensate (Q-Q plot along x=y).



Top) Scatterplot for the seasonal rain-rate for 3B42RT for winter (left) and summer (right) over Northwest (NW) displaying underestimation (blue) and overestimation (red) between 3B42RT and GHCN-D. Medium) Differences between 3B42RT and GHCN with respect to the elevation for DJF (left) and JJA (right). Bottom) Location of the points within the NWRFC for DJF (left) and JJA (right).

Results show that the 3B42RT pixels that display a strong underestimation (<-50%) with respect to GHCN are located West of the Cascades mountain range and for low to moderate elevation (h<500m) regardless of the season. Comparatively, the 3B42RT pixels that display a strong overestimation (>50%) are found East of the Cascades mountain range regardless of the elevation (200m<h<2500m). While, the average rain rate remains relatively constant East of the Cascades mountain range (R≈2mm/day), the seasonal differences are more important West of the Cascades with average rain rates less than 2mm/day in Summer to be compared with more than 5mm/day in Winter (and up to 15mm/day). Those difference being important (Diff. <-50%) regardless of the season, they illustrate the difficulties for satellite to capture orographic precipitation (DJF, JJA) and cold season precipitation (DJF). While the bias-adjustment (3B42) provides a good agreement with surface observation for summer (Fig. 7b), the bias-adjustment for winter (Fig. 7a) is more delicate due to the possible presence of cold precipitation.

To account for this point, we will add in the text (p. 11502, l. 19):

The real-time 3B42RT displays small biases when compared to GHCN-D (+2.4%) and PRISM (+0.4%). However, rather than the indication of a good performance, results show that the locations with overestimation are compensated by those with underestimations as can be seen with the Q-Q plot aligning along the diagonal. A closer look indicates that the 3B42RT pixels displaying the strongest underestimation (<-50%) with respect to GHCN-D are located West of the Cascades mountain range for low to moderate elevation (<500m) (not shown). Comparatively, the pixels that display the strongest overestimation (>50%) are found East of the Cascades regardless of the elevation. However, while the average rain rate remains relatively constant East of the Cascades throughout the year (R≈2mm/day), the seasonal differences are more important West of the Cascades with average rain rates of less than 2mm/day in Summer to be compared with more than 5mm/day in Winter. The important underestimation by 3B42RT West of the Cascades regardless of the season, illustrates the difficulties for satellite to capture orographic and cold season precipitation. We also note that despite the bias-adjustment, underestimation remains for 3B42 in winter due to uncertainties related to cold season precipitation measurements mentioned earlier or by rain gauge locations that cannot fully capture orographic effects that can be observed over distances smaller than the satellite resolution (Prat and Barros 2010a).

9. On p.11500 ll.26: "For summer, important negative bias adjustment (3B42<3B42RT) is found over the Midwest (MB, NC, AB) and corrects for the overestimation of summertime convection by PMW sensors that mistake sub-cloud evaporation for precipitation". Is there any citation for this statement?

Yes, there are few references documenting this point. In particular:

Dinku, T., Ruiz, F., Connor, S. J., and Ceccato, P.: Validation and Intercomparison of Satellite Rainfall Estimates over Colombia, *J. Appl. Meteorol. Climatol.*, 49, 1004–1014, doi:10.1175/2009JAMC2260.1, 2010.

Dinku, T., Ceccato, P., and Connor, S. J.: Challenges of satellite rainfall estimation over mountainous and arid parts of east Africa, *Int. J. Remote Sens.*, 30, 5965–5979, doi:10.1080/01431161.2010.499381, 2011.

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, 2014.

The references will be added in the text and in the list of references (see also comment #6).

10. The gauge stations have notorious issues to quantify precipitation during the Winter season, caused e.g. by icing, underestimation related to snow drift, etc. This propagates to precipitation products making use of this information (Stage IV, 3B42). The authors need to address this point in order to provide critical insight on the precipitation estimates over CONUS.

Further discussion on rain gauge uncertainties in quantifying cold season precipitation will be added in the revised version. Several studies have documented the uncertainties in measuring solid precipitation with rain gauges. Among the rain gauges that are used either in the adjustment procedure of precipitations products or in the evaluation procedure of remotely sensed products, we find different types of sensors including heated automated tipping bucket rain gauges (ASOS, AWOS, HADS) or procedures relying on human intervention (GHCN-D). For solid precipitation, an exhaustive assessment of those uncertainties can be found in Goodison et al. (1998), while other publication summarize uncertainties related to point measurements of precipitation (Sevruk, et al. 2009, McMillan et al. 2012 among others). Among the systematic errors in measuring frozen precipitation, are evaporation, chimney effect, wind field deformation, wetting losses (McMillan et al. 2012), or possible delayed tips due to snow melting in the funnel (Prat and Barros 2010a) among others. Wind effects can lead to underestimation of frozen precipitation by 50% or more (Groisman et al. 1999). A recent publication (Leeper et al. 2015) provides a good overview of the methodology used to quantify winter precipitation and the related uncertainties for the GHCN-D dataset and the COOP stations subset.

We added in the text (p. 11501, l. 16):

In addition to fundamental limitations in radar and satellite measurement for snow and mixed precipitation events, additional uncertainties are introduced from in-situ data that are used either in the adjustment (HADS, ASOS, AWOS) or in the evaluation (GHCN-D) of remotely sensed products. Among the systematic errors in measuring frozen precipitation, are evaporation, chimney effect, wind field deformation, wetting losses, delayed tips due to snow melting in the funnel, or uncertainties due to human intervention in the measurement procedure (Goodison et al. 1998, Groisman et al. 1999, Sevruk et al. 2009, Prat and Barros 2010b, McMillan et al. 2012, Leeper et al. 2015). Although, this point is beyond the scope of this study, we note that the differences observed between remotely sensed and in-situ data for the higher latitude RFCs (CB, MB, NC, NE, NW, OH) are within the range of that observed for the other RFCs experiencing cold precipitation less frequently (-4.7% to -14.8% vs. -1.5% to -18.0% for Stage IV and -31.1% to +16.7% vs. -38.0% to +16.7% for 3B42) (Table 3).

References:

Goodison, B. E., Louie, P. Y. T., and Yang, D.: WMO solid precipitation measurement intercomparison. WMO Instruments and Observing Methods Rep., WMO/TD-872, 212 pp., 1998.

Groisman, P. Y., Peck, E. L., and Quayle, R. G.: Intercomparison of Recording and Standard Nonrecording U.S. Gauges , *J. Atmos. Ocean. Tech.*, 16, 602-609, 1999.

Leeper, R. D., Rennie, J., and Palecki, M. A.: Observational Perspectives from U.S. Climate Reference Network (USCRN) and Cooperative Observer Program (COOP) Network: Temperature and Precipitation Comparison, *J. Atmos. Ocean. Tech.*, In press, 2015.

McMillan, H., Krueger, T., and Freer, J.: Benchmarking observational uncertainties for hydrology: rainfall, river discharge, and water quality, *Hydrol. Process.*, 26, 4078-4111, 2012.

Prat, O. P., and Barros, A. P.: Ground observations to characterize the spatial gradients and vertical structure of orographic precipitation - Experiments in the inner region of the Great Smoky Mountains, *J. Hydrol.*, 391, 143-158, doi: 10.1016/j.jhydrol.2010.07.013, 2010b.

Sevruk, B., Ondras, M., and Chvila, B.: The WMO precipitation measurement intercomparison, *Atmos. Res.*, 192, 376-380, 2009.

11. p. 11503 ll.27-29: Any explanation why the correction is insufficient in mountainous areas?

The main reason is the difficulty in capturing orographic precipitation over complex terrain. Important orographic rainfall enhancement can typically be observed over distances smaller than the satellite resolution (Prat and Barros, 2010b). Another reason is the difficulty in measuring cold precipitation both by the sensor and the rain gauge (see previous comment). Finally, the locations of the rain gauges

We will add (p. 11503, l. 27):

Overall, the bias-adjusted 3B42 performed very well over the Great Plains (MB, NC) to correct for the overestimation of summertime convection (Table 3) with comparable results than Stage IV (Table 3). Important differences remained however for low daily rainfall (< 1mm/day) and for the western RFCs during wintertime mostly due to the difficulty in capturing orographic precipitation and uncertainties in retrieving cold precipitation by the satellite (Chen et al. 2013, Huffman and Bolvin 2013) but also and by the rain gauges used in the bias-adjustment and the evaluation (Goodison et al. 1998, Groisman et al. 1999, Leeper et al. 2015).

Reference:

Prat, O. P., and Barros, A. P.: Assessing satellite-based precipitation estimates in the Southern Appalachian mountains using rain gauges and TRMM PR, *Adv. Geosci.*, 25, 143-153, 2010b.

12. On p. 11505 ll.8-10. Chen et al. (2013) mention this issue. Chen, S., et al. (2013), Evaluation of the successive V6 and V7 TRMM multisatellite precipitation analysis over the Continental United States, Water Resour. Res., 49, 8174–8186, doi:10.1002/2012WR012795.

We thank the Reviewer for this suggestion. We added the reference at the end of the revised section 4.