#Reviewer 1

We thank Reviewer#1 for their careful review and positive assessment of our manuscript. Please find below our answers to each comment.

General Comments

Overall this is a well written paper that adds some new insight into undercatch of the Thies tipping bucket gauge. My big comments is that there could be some more discussion. The Discussion and Conclusions sections are combined and it could be useful to separate the two. There are limited citations in the Discussion, yet adequate in the Introduction. The authors state on line 10-11 of page 3 "[t]hese results are used to identify areas within Spain where errors affecting snowfall accumulation are most significant." This is a very good application of the results to yield new important results. Revisit the literature on spatial precipitation patterns across Spain, etc., in the. Discussion, i.e., how these results could change those previous studies.

We added the following sentences in the discussion:

- This underestimation could affect previous studies of solid precipitation, especially if the period of time considered was associated with significant winter precipitation extremes (López-Moreno et al., 2011, Vicente-Serrano et al., 2011, Añel et al., 2014, Cortesi et al., 2014, Buisan et al., 2016)

- Wind speed was measured at the standard operational height of 10 m, instead of at gauge height. The main advantages are that measurements are less affected by obstacles, and that all stations in the operational network measure wind at this height. This allows for broader applicability of the derived transfer functions within the network. Previous work has shown only small improvements in the accuracy of results using the gauge height wind speed relative to using the 10 m wind speed (Kochendorfer et al. 2017)

- The main results of this study are also consistent with previous studies in which different gauges, including tipping buckets, were tested relative to the reference at different sites (Rasmussen et al., 2012, Wolff et al., 2014, Earle et al. 2016, Kochendorfer et al., 2017)

- One factor that was not included in the analysis was the impact of heating on the evaporation or sublimation of incident precipitation producing losses in the TPB accumulation, especially at low intensities (Savina et al., 2012, Zweifel and Sevruk, 2002). The tipping bucket used operationally and tested in Formigal was heated with only 49 W (in comparison with other models of tipping bucket with power greater than 100 W) and snowfalls at Formigal are usually characterized by high intensities. Based on this, we expect that, in addition to the impact on catch efficiency already included in transfer functions, longer delays due to melting could be expected. For

this reason the choice of 1 h and 3 h hour accumulation periods were preferable to shorter time periods.

We separated the discussion and conclusions sections, as suggested.

Something that could be considered is the calibration and evaluation of the equations.

More than 200 samples were used; is it possible to use a split sample approach to calibrate and evaluate the equations?

We have added the following in the results section:

To estimate the uncertainty of the proposed equations we split the 1 h and 3 h datasets each into two equal and independent datasets. One dataset was used to calculate the regression equations (114 events for 1 h period, 45 events for 3h period). The resulting equations were similar to those obtained using the entire dataset. The accuracy of the resultant regressions was then independently evaluated using the second sub-sample of each dataset (100 events for 1h period, 42 events for 3 h period). For the 1 h dataset the resultant RMSE was 0.13, and for the 3 h dataset it was 0.11. These values are acceptable given that the R^2 were between 0.6 and 0.7 showing that there was still some residual uncertainty in the regressions due to the variability and complexity of the relationship between the measurements from a TPB and a weighing gauge within a DFIR.

I have some thoughts about the project design, yet these are future consideration and do not affect how I read this paper. Repetition is good; it would be useful, at least as a backup to have a second and third Thies tiping bucket. Considering stating why there is only one Thies tiping bucket. Add a Hellman gauge as old Spanish observations were made with this configuration (p2, line 33). Does this include the Hellman shield?

We didn't include the Hellman gauge because the SPICE experiment was mainly focused on automated instruments. In addition, Formigal is an unattended site with no manual observations. The Hellman gauges used in Spain are unshielded.

Where the Hellman gauges run concurrently with the Thies tipping bucket? The Thies tipping bucket gauges are heated (p3, line 28), and the evapo-sublimation problems (see Goodison et al., 1998) should be discussed. The "[w]ind was measured at a standard height of 10 m ..." which is a AEMET standard (p4, line 2). However, an addition of a gauge height wind measurement could be useful. I have numerous comments to clarify the text, tables and figures.

This study was performed primarily to support current AEMET operations measurements, and the Hellman gauge and the Thies are collocated in only a few stations, with very low frequency of snowfalls.
In the context of SPICE, an adjustment for wind speed at gauge height is used. This winter season in Formigal we plan to install heated wind sensors at gauge height to measure the

difference between gauge height and 10 m wind speeds; however, due to our relatively high gauge height, we expect the wind speeds to be within about 10% of each other.
In the article below the authors stated that: "measurements did not actually indicate any advantages to using the estimated gauge height winds over the 10 m height winds"

Kochendorfer, J., Nitu, R., Wolff, M., Mekis, E., Rasmussen, R., Baker, B., Earle, M. E., Reverdin, A., Wong, K., Smith, C. D., Yang, D., Roulet, Y.-A., Buisan, S., Laine, T., Lee, G., Aceituno, J. L. C., Alastrué, J., Isaksen, K., Meyers, T., Brækkan, R., Landolt, S., Jachcik, A., and Poikonen, A.: Errors and adjustments for single-Alter shielded and unshielded weighing gauge precipitation measurements from WMO-SPICE, Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-684, in review, 2017

We have added a related sentence in the discussion section as commented previously.

In general the paper is well written, and as such I can make these detailed comments to ask for further clarification. The tables need some work to make them more understandable. The Table and Figure captions are too brief and need to be expanded to provide more insight. The Figures need some modification to make them easier to read and in some cases more intuitive. I suggest the use of double mass curves in several instances, such as Figure 5a, as they would provide more insight into event by event differences.

We have provided more detailed explanation in some captions and we have also modified several figures.

Specific Comments

page 1, lines 1-3: the title could be shortened to remove some of the little words. How about "Assessment of the snowfall accumulation underestimation by Spanish National Weather Service operational tipping bucket gauges" Perhaps the name of the gauge "Thies tipping bucket" could be added, as this is the first time I have seen this specific gauge being adjusted for undercatch.

We have observed, comparing with another heated Tipping Bucket from another manufacturer that the undercatch is quite similar and the main contributing factors are temperature and (especially) wind. It is likely that the results would be the same, especially for the areas of Spain where the undercatch of measured precipitation is higher. That is the main reason why we just put 'tipping bucket gauges' in the title. In addition, from the title "Assessment of the snowfall accumulation underestimation by Spanish National Weather Service operational tipping bucket gauges" and reading the article you can know the manufacturer. This way we avoid including a commercial model in the title.

We have modified the title as follows:

Assessment of snowfall accumulation underestimation by tipping bucket gauges in the Spanish operational network

page 1, lines 27-29: it is appropriate to give specifics such as catch ratios, but this seems awkward. Consider rewriting

We have rewritten the sentence as follows:

"At wind speeds of 1.5 m s-1 the tipping bucket recorded only 70% of the reference precipitation. At 3 m s-1, the amount of measured precipitation decreased to 50% of the reference, was even lower for temperatures colder than -2 ºC, and decreased to 20% or less for higher wind speeds"

page 1, lines 29-31: this sentence is unclear

We have rewritten the sentence as follows:

"The implications of precipitation underestimation for areas in Northern Spain are discussed within the context of the present analysis, by applying the transfer function developed at the Formigal-Sarrios, and results from previous studies"

p2, lines 2-3: the first sentence is not clear in terms of what "it" refers to. Is "it" the measurement or the accumulation?

We have rewritten the sentence as follows:

"Variability of snowfall accumulation strongly influences the ecology and hydrology of mountainous areas and cold regions, impacting economic activities including winter tourism, hydropower generation, floods and water supply for agriculture. For this reason an accurate measurement of snowfall accumulation is critical."

p2, line 9: "undercatch" from "wind-induced updrafts" is not quite correct. It is more than just updrafts.

We removed -induced updrafts

p2, line 16: "a secondary reference for solid precipitation" implies that there is a primary reference.

Yes. The primary reference as established by the first WMO intercomparison of snowfall measurements is a Tretyakov gauge sheltered by *bushes* (currently called the *R0 reference). We added the following reference* (Goodison et al. 1998)

p2, line 25: is there a specific “automatic gauge” in the centre of the DFAR?

Yes, the Geonor T200-B3 or the OTT Pluvio^2. We added to the text

p3, line 1: cite the map (see my comments on Figures 1 and 2 below)

This citation has been added.

p3, line 7 (and throughout): I don’t like the term “transfer” function. Mathematically this implies altering the pattern of the data, such as a Fourier transform. If this is a SPICE term, then disregard. If not, consider another term such as “equation” rather than “transfer function.”

Yes, it is a SPICE term, we could also use “adjustment”

p3, line 7: the estimation of “true snowfall amounts” is not actually the case here, as this work aims to derive an equation or set of equation to estimate the DFAR snowfall amount. Yang et al. (1993 Eastern Snow Conference) showed that true snowfall is often more than DFIR snowfall.

Ok. We removed “true”.

p3, lines 8-10: reword “[t]he wind speed during snowfall events is included in this analysis to help determine the potential impact of wind-induced undercatch on Spanish snowfall measurements,” to be more specific, i.e., wind speed (and temperature) data were used ...

We have rewritten the sentence as follows:

“Wind speed and temperature data during snowfall events were used in this analysis to help determine the potential impact of wind-induced undercatch on Spanish snowfall measurements.”

p 3, line 16: “sub-alpine environment” is below the treeline, i.e., among the tree. This is likely the case here, but stating that the area “consist[s] of a mixture of bare ground and only very low grasses,” makes us think it is in the alpine. If the site is in an opening, state as such. Please clarify.

We have rewritten the sentence as follows:

This is an alpine environment consisting of a mixture of bare ground and only very low grasses

p3, line 17: consider showing a wind rose to illustrate that "[t]he prevailing winds are from the northwest all year round," since this affects wind speeds (lines 18-21) through upwind fetch lengths, etc.

I removed the sentence because snowfalls can be produced under very different weather conditions on this area (N, NW, W, SW and S) depending on prevailing atmospheric patterns during the corresponding winter season. So the sentence could be confusing.

p3, lines 23, 27, 28: web links may be useful for manufacturers?

This is a good suggestion, but we believe that is not needed. It is straightforward to find the relevant manufacturer websites, and in this way, we can avoid issues related to broken or outdated links.

p3, line 30: the "non real time" output is not explained well.

We have rewritten the sentence as follows:

"As described in Pluvio^2 manual, the NRT output was used because in this mode the instrument collects fine precipitation using an integrated 0.2 mm per hour threshold, making it more comparable to the Thies tipping bucket gauge, whose minimum resolution is the same 0.2 mm."

We have also included "Pluvio^2 manual" to show where more information can be found.

p4, line 8: "1 minute data" which data? all variables or just wind speed and temperature?

We have rewritten the sentence as follows:

To assure high data quality, the quality control procedures removed all capping events and filtered out periods (1h and 3h) where less than 90% of the 1 minute precipitation *precipitation?* data was available.

"Integrated data was delivered every 1-minute for all instruments, the sampling frequency was different depending on the instrument, according to WMO guidelines. All the data were recorded using two Campbell CR1000 data loggers. "

p4, lines 11, 30, 35, etc.: It would be informative to show where the AEMET operational gauges were in Figure 2.

We have added this in Figure 2.

p4, lines 26-27: provide a citation for “these gauges have been progressively replaced.” Consider adding a time period for this replacement.

There is no possible citation. This is an internal progress in a National Weather Service based on a combination of factors: new stations, retirement of volunteer observers, etc.

p4, line 32-33: “snowfall events were defined as precipitation events that occurred when the average maximum temperature was below 0_C.” What is the basis for this assumption? Snow can fall at air temperatures warmer than 0_C (e.g., Fassnacht et al., 2013 IAHS 360, 65-70).

Of course, and that is the case of Formigal. Please check Figure 4 where the disdrometer showed that between 0°C and 2 °C, most of precipitation was in the form of snow. Selecting only cases where average maximum temperature was colder than 0°C, we were confident that all events were snow events. It was also a useful threshold to derive, in the second part of the article, a transfer function for the operational network. In most of our operational stations we don’t have disdrometers so the only variable for discrimination the type of snow was temperature.

We have rewritten this paragraph as follows:

“Automatic weather stations were not equipped with disdrometers, and thus we used temperature data to select snowfall events. For the purpose of this analysis, snowfall events were defined as precipitation events that occurred when the average maximum temperature was colder than 0ºC and the total accumulation was greater than 0 mm during a 1 h time period. We consider these criteria to be adequate for the scope of this work, despite the fact that mixed precipitation can be observed at temperatures colder than 0ºC and that snow can fall at air temperatures warmer than 0ºC (e.g., Fassnacht et al., 2013 IAHS 360, 65-70)”

p5, lines 6-7: be sure to clarify that differences in snowfall accumulation” yield less precipitation at the other gauges (SA, UN, TPB) than at the DFAR.

OK. We have modified the figures.

p5, line 11: the line “which agree to within 90 - 100%” is unclear.

We have rewritten the sentence as follows:

In both situations, there is good agreement between both instruments DFAR and LPM which agree to within 10% (Figure 3b) and 1% (Figure 3a).

p5, line 13: “The deviations in accumulations are most likely related to the wind-induced undercatch” is a weak statement. Consider rewriting.

We have rewritten the sentence as follows:

Deviations in accumulations from a given gauge are caused mainly by wind-induced undercatch and precipitation type (i.e. dry snow or wet snow).

p5, line 17: change the word "traces" here, as this has a precipitation implication (P <precision). Also consider the occurrence of snow (see Fassnacht et al., 2001, Journal of Hydrology 253, 148-168, Figure 1).

We have rewritten the sentence as follows:

Results showed that for precipitation events at temperatures colder than 0 ºC, precipitation occurred primarily as snow, with only a few cases of mixed precipitation.

p5, lines 17, 18, p8line18, etc.: don't use "above", "below" or "near" to refer to temperatures, especially not 0_C, as these words have altitudinal implications. use warmer than (above), colder than (below), or at approximately (near).

OK. Done.

p5, lines 18-19: "the threshold temperature of 0 _C is suitable for classifying the precipitation as snow and not rain for the site." Be careful with such statements, as the discrimination of rain versus snow can be difficult (e.g., Harder and Pomeroy, 2014, doi: 10.1002/hyp.10214) as is further stated.

We have rewritten the sentence as follows:

"The number of cases where snow was detected at temperatures warmer than 0 ºC was still very high, which indicates that the threshold temperature of 0 ºC was suitable, **based on disdrometer data**, for classifying the precipitation as snow and not rain for the site."

We added the following paragraph to the discussion.

"Despite the difficulty of discriminating rain from snow (Harder and Pomeroy, 2014) the upper threshold temperature of 0 ºC was suitable for classifying the precipitation as snow. This was also supported by the high number of snow occurrences detected at temperatures warmer than 0 ºC and its consistency with previous work (Fassnacht et al., 2013).

p5, line 26: I would delete the word "ratio"

OK. Done.

p6, line 18-23: this paragraph is unclear. I don't quite understand what was done here to derive the 9.5%

The paragraph has been reworded:

The value of 9.5% was determined by calculating the correlation between the hourly TPB measurements and the DFAR measurements for all "melting factors" between 0 and 30%, and then creating a correlogram (Figure 7). A peak in the correlation was associated with a melting factor of 9.5%, where 9.5% of the Thies precipitation from a given hour was assumed to have melted in the following hour.

p7, line 4: change from "From this moment on" to "Hereinafter"
OK. Done.

p7, lines 14-16: any relation with elevation and the presence of a canopy? It seems to go with the valley vs. plateau discussion earlier

Of course in some stations it could be an additional factor. However, according to WMO guidelines, they should be located far from obstacles such as trees.

p7, line 23: state the time period "for all snowfall events"

We have rewritten the sentence as follows:

Using the derived transfer function (Equation 2, Table 3), the average catch ratio for each station was calculated for all 1 h snowfall events (Figure 11).

p8, line 10: "where" should be "were"

OK. Done.

p8, lines 11-13: reword "However this was not necessarily because the wind speed was lower in these mountainous areas during snowfall events, as the measurement stations are generally located in the bottom of the valleys where they are less affected by the wind." This sentence is unclear

Yes it is confusing. I removed the sentence.

p8, lines 15-16: the phrase "higher losses due to undercatch" doesn't seem correct here. Consider saying something like "more total undercatch"

OK. Done.

Table 2 is difficult to read. What is on the row heading (DFAR?) and what is on the column heading (TPB)? Is the top table for 1 h and the bottom for 3 h accumulation periods? I suggest spelling out TPB in the figure caption to help the reader. DFAR is assumed as a derivative of DFIR, so doesn't need to be spelled out.

We have modified as follows:

Table 2: Contingency tables of cases detected by each instrument and the sum of accumulation not detected for 1 h (top table) and 3 h (bottom table) accumulation periods. In the YES/YES case the precipitation amount is that measured by the reference.

1 h period	DFAR YES	DFAR NO
TPB YES	238 (400.33 mm)	11 (2.6 mm)
TPB NO	156 (45.11 mm)	

3h period	DFAR YES	DFAR NO
TPB YES	96 (385.58 mm)	6 (1.6 mm)
TPB NO	41 (15.62 mm)	

Table 3: the caption is not informative. Explain that each step adds a new variable to transfer functions, and that the last equation is the application to correct the data. Reformat the table. Other statistics would be useful, such as RMSE or the Nash-Sutcliffe coefficient of efficiency. Consider including the number of data points used in each equation or set of equations.

We have modified the tables and equations as follows:

Table 3. Derived transfer functions, where each step adds a new variable. Top table is for 1 h period and bottom table is for 3 h period. The value of the coefficient of determination (R^2) increases with more variables and longer periods. The number of data points used in the analysis were 214 and 87, for 1 h period and 3 h period respectively. The R^2 value increases with more variables and longer periods. Equation 4 and 8 were used to correct the data as explained in section 3.1.

Variables: CR=Catch Ratio, T=Temperature (°C), W=Wind speed (m/s), Acc=Accumulation (mm)

1 h Transfer functions	
CR=0.87*exp(-0.198*W)	R^2=0.49
CR=1.01*exp(0.077*T-0.176W)	R^2=0.57
CR=0.925*exp(0.069*T-0.176*W+0.078*Acc)	R^2=0.60
True accumulation (1h)= Acc/CR – 0.095*Acc/CR+ 0.095*Acc(previous hour)	

3 h Transfer functions	
CR=0.84*exp(-0.234*W)	R^2=0.52
CR=1.04*exp(0.094*T-0.201W)	R^2=0.60
CR=0.892*exp(0.067*T-0.212*W+0.049*Acc)	R^2=0.65
True accumulation (3h) = Acc/CR	

Figure 1 and 2: can these be combined? Perhaps just put the Formigal-Sarrios star on the map of Europe-North Africa. The different mountain ranges in Spain are not relevant to this paper, as shown. Since part of the objective of this paper (p3, line 6-7) is to "demonstrate the importance of accurate snowfall measurements within this network," considering showing were these operational gauges are in the various mountain systems of Spain (Figure 8-11). "Altitude" is height above the ground; I suggest using the term elevation in Figure 2 (see p4, line 13).

We have also modified the figure based on comments.

When we are discussing about the spatial distribution of areas with higher/lower undercatch in Spain we need to locate mountain ranges and know their names.

Figure 2: The star is not in the correct location. "Formigal-Sarrios" is about at the top of the "P" in Pyrenees on the map. It should be at latitude 42.57 and longitude -0.62.

OK. This has been modified.

Figure 3a and b: the wind speed is difficult to see in the plots. Consider putting this in a separate stacked graph, with at least twice the scale in the y-direction. Consider putting a different (heavier?) line type for 0_C. A set of double mass curves compared back to the DFAR may also be informative.

We have modified the figures to make them easier to understand.

Figure 4: change the x-axis from (-2,0) and (0,2) to "-2 to 0" and "0 to +2" I can't tell sleet from snow. Consider a different colour scheme, such as red for rain and blue for snow with green to yellow in between.

OK Done. We have modified the figures and change colors based on these recommendations

Figure 5a: what are the units on the y-axis? Done (mm)

Consider using two (1h & 3h) double mass curves (cumulative precip vs. cumulative precip) with DFAR on the x-axis.

We have modified the figure based on this recommendation.

Figure 6: change (-2,0) to "-2 to 0" etc. It is difficult to distinguish "-2 to 0" and "<-6." Consider a different color for "<-6"

OK. All colors have been changed.

Figure 7: define "melting factor" in the caption.

OK. Done. We have modified as follows:

"Correlation between the hourly TPB measurements and the DFAR measurements for different melting factors, where melting factor is the percentage of the Thies tipping bucket precipitation from a given hour melted in the following hour"

Figures 8 and 9: also show m/s to be scientific - km/h is to be operational (p7, lines 4-5)

We consider that in this case too we can use km/h because are the units used in operational network. We have already used m/s for the scientific section of the article. It is a choice that doesn't affect the results.
In the text is explained as follows:

Hereafter, we will use the units of km/h for wind speed because they are used in the operational network and can facilitate the comprehension of the results.

Figure 8, use a), b) and c)

OK. Done.

Figure 8, 9 and 11: can you use different color schemes? For Figures 8 and 9, they are different units and these colors could be confusing (maybe use greens ad greys). Repeating these colors in Figure 11 make it further confusing. In Figure 11, I would use yellow, orange and red to indicate a scale of ok to poor (e.g., caution, warning, alert). In Figure 10, red, white and blue are used. As these are only three colors, they are not as repetitive, but the legend is

in the wrong order and red is used for the coldest temperatures while blue is used for the warmest. This is counter-intuitive.

We have modified the colors based on these recommendations.

Figure 11: is it possible to add “error” or "uncertainty" as a secondary map?

It would be difficult and in some way it could be confusing and probably out of the scope of the objective of this section which is more related to illustrate areas of Spain where, due to the location of stations, the undercatch of precipitation is higher during snowfall conditions. This suggestion could be very useful for a future work dealing with the applications of the results of this work