

Response to reviewer comments on “**A probabilistic approach to removing cloud cover from MODIS Snow Cover Area products**” by V. López-Burgos et al., submitted to HESSD:

Dear Dr. Parajka,

Thank you for your comments and suggestions. Your suggestions encouraged us to think more deeply about the seasonal distribution of cloud removal power and its tradeoff with mapping accuracy for the algorithms we used. We had stored the information for daily and monthly cloud cover removal results and have therefore revised our paper to discuss this manner in more detail and compared it with the papers you suggested. Unfortunately, however, we no longer have access to the daily/monthly accuracy results so we were unable to include this in our revision.

We think that our monthly accuracy results' differences for the Terra/Aqua combination and Time Interpolation would be somewhat similar to those reported by *Parajka and Blöschl (2008)*, *Gao et al. (2010)* and *Xie et al. (2009)*. For the LWLR we think that the more closely related paper to use for estimating the seasonal mapping accuracy is *Parajka et al. (2010)* but with some differences. We have addressed this issue on the revised discussion section of our paper.

We have done our best to address your comments and suggestions. Please see our responses below. Our response to your general comments are included as part of our revised version of sections 5, 6 and 7 and is included below and our responses to your specific comments are enumerated afterwards.

The revised versions of Sections 5, 6 and 7 are reproduced below:

5 Results for each cloud removal method applied independently and in sequence

Each of the four methods was first tested separately to assess the effectiveness of its performance. Subsequently, a sequential approach was tested, in which the methods were applied in sequence. Two summary statistics are used to indicate the change achieved by implementation of each method:

$$\%Change = \left(\frac{\#Corrected(x_i,t) - \#Original(x_i,t)}{\#Original(x_i,t)} \right) \cdot 100 \quad (6)$$

$$\%Cover = \left(\frac{\#Pixels(x_i,t)}{\sum_{i=1}^6 \#Pixels(x_i,t)} \right) \cdot 100 \quad (7)$$

where x_i represents codes 1 through 6, and t represents the time period for which the statistic was calculated (e.g. year or month).

We first discuss the results of applying each method separately. The change in % cover achieved over the entire study time period by each method is shown in Table 1 for each category. Note that 39% of the image pixels were initially classified as cloudy.

In brief:

1. The Terra/Aqua combination method had a percent change of cloud cover, error and no-decision pixels of -23%, -97% and -3.7% respectively. This translates to a cloud cover change from 39% to 30%. These results are similar to the cloud removal achieved with the same method by Parajka and Blöschl (2008), Xie et al. (2009) and Gao et al. (2010). Furthermore, the two days missing in the Terra images time series were not missing in the Aqua images and so a more complete time series was achieved. In addition, the combination of MOD10A1 and MYD10A1 increased the snow covered pixels and the no snow pixels by 34.5% and 14.6% respectively (+1.55 and +8.13 snow and land cover). On a monthly basis, this approach works better during the Fall and Spring months versus the Winter months with a maximum cloudy pixels reduction in May (-36%) and a minimum cloud reduction in January (-16%) (Fig. 5). The monthly distribution of differences in cloud removal are similar to those of Parajka and Blöschl (2008), Xie et al. (2009), Gafurov and Bárdossy (2009) and Gao et al. (2010) while differing on the amount of cloud cover removed on each site. This might be due to differences in climate, topography, study area (km²) and cloud dynamics between the watersheds during the different time periods used.

2. The time interpolation method had a percent change of cloud cover, error and no-decision pixels of -43%, -95% and -46% respectively, with a cloud cover change from 39% to 22% while increasing the snow covered pixels and no snow pixels by 33% and 28% respectively (+1.51 and +7.25 snow and land cover). These results are similar to the 1-day temporal filter used by Parajka and Blöschl (2008) and significantly less than their 7-day temporal though the results are not strictly comparable since they applied the temporal filters to the already improved images by way of Terra/Aqua combination. The same applies to the cloud removal results obtained by Gao et al. (2010), which also used improved images as the inputs for their temporal filters. Xie et al. (2009) achieved a cloud cover reduction from 39.5% to 6.0% but the amount of days used for each image and the weighted mean function used are not clear. This method removed more bad pixels than Terra/Aqua but it does not give information on missing days. It also performed better during the Fall and Spring months than in the Winter months with a maximum cloud reduction in May (-84%) and a minimum in February (-17.6%). However, as shown on Figure 5, cloud reduction in November was not as good as in October and December due to higher percentage of cloud cover in November. As is the case with the Terra/Aqua combination, this method performed less strong during winter months due to cloud persistence beyond the temporal window used to improve the images.

3. The Nearest Neighbor spatial interpolation method is the least effective of the four methods. The percent change on cloud cover, error, and no decision pixels was only -5 %, -9% and -0.03% respectively with a cloud cover change from 39% to 37%. These results are similar to that of Parajka and Blöschl (2008). Their spatial interpolation method reduced the cloud cover from 52% to 46% although they applied this algorithm to the images after having improved them with the Terra/Aqua combination. Again, this method does not give information on missing days. Furthermore, this method does not show the seasonal difference in cloud reduction power of the three previous methods, its effect is almost constant throughout the time period.

4. The Locally Weighted Logistic Regression (LWLR) method achieved the highest effectiveness at the cloud covered and no decision pixels (-62% and -60% respectively), but was significantly less effective than either Terra/Aqua combination or time interpolation at reducing

the number of “error” pixels (-44%). This method also had the highest increase in snow covered pixels (96%) and no snow pixels (35%) and shows a similar seasonal pattern of cloud removal with the maximum cloud reduction in April (-90%) and a minimum in February again (-50%). The change in cloud cover was from 39% to 15%. Notwithstanding, it does not take care of the missing days problem either.

Figure 5a shows how the percentage of pixels classified as cloudy in the original MODIS SCA images varies for each month during the study period, from a low of 24% in April to a high of 69% in February. Figure 5b–d shows the %change in clouds, snow and land pixels achieved by each method for each month. Nearest neighbor spatial interpolation is consistently poor at cloud removal, while logistic regression is the most effective. For removing no decision pixels, the Terra/Aqua and time interpolation algorithms perform best. All methods increased the amount of snow covered pixels more than they increased the amount of land pixels, a reasonable result since the areas with the most consistent cloud cover throughout the study period are the areas of snow accumulation, mainly at higher elevations, and the period with more cloud cover is during the winter months when snow accumulates (Fig. 3).

The results above indicate that each method has different strengths and weaknesses. To synergistically exploit the strengths of all four methods, we next applied the methods in sequence (in the order Terra/Aqua combination, time interpolation, spatial interpolation, and locally weighted logistic regression), at each step retaining the results from the previous one. Table 1 shows that the sequential approach achieved a very high degree of cloud cover, error and no-decision pixel removal (-94 %, -99% and -64% respectively), while increasing the number of snow and land pixels by ~154% and ~54% respectively – a considerable increase in the amount of SCA indicated by the images. Figure 6 shows the progressive improvement obtained by sequential application of the methods in terms of percent cover (i.e. how the cloud/snow/no snow covers change with each step) and how this improvement distributes across months.

Overall, using the sequence the cloud cover was reduced from 39% to 30% using the Terra/Aqua combination, then to 14% with the Time Interpolation, to 13% with the Nearest Neighbor Spatial Interpolation and down to 2% with the LWLR. On a monthly basis the cloud cover was reduced to less than 10% for all months with the most cloudy month (February) having a final cloud cover of 9% and October, April and May having a final cloud cover of less than 1%. Figures 6 c and d look like a typical snow curve and an upside down snow curve. The increase in snow-covered pixels follows the expected monthly snow distribution except that more snow was added to the month of January than in February. This is likely due to the difficulty of removing clouds in February as it had a higher percentage of cloud cover. The increase in no snow pixels also shows the expected monthly distribution with the Fall and Spring months showing a higher increase. Moreover, only 5% of the days were left with a cloud cover greater than 10% (Not shown). These days are distributed between December and March with the majority being in February, the cloudiest month, and are grouped in doubles or triples (i.e. two or three consecutive days with heavy cloud cover). Therefore it makes sense that the algorithms were not able to reduce the cloud cover to less than 10%. Figure 7 shows the significant change obtained by the sequential method, for the MODIS SCA image of 18 February 2005. The changes achieved by the hybrid approach are clearly substantial, and in marked contrast to the changes achieved by each method applied independently.

6 Evaluation of accuracy of the results

Finally, we assess the accuracy of the results by comparing the SCA images with data from four available SNOTEL sites located in the mountainous zones of the Salt River basin (Fig. 1). Although the SNOTEL sites are effectively point-scale measurements, their pixel locations were localized using ArcGIS 9.3 by converting the station's point shapefile to a raster with the same extent, resolution and projection of the images. This raster was then converted to an ASCII file for processing. Several evaluation statistics were calculated for (a) the original MOD10A1 and MYD10A1 images, (b) each cloud removal method applied separately, and (c) the sequential cloud removal approach.

For each SNOTEL site pixel, the hit, false alarm, miss and correct rejection rates were computed for each day of the available time series. A simple approach to evaluation would be to consider each SNOTEL site pixel to be snow covered if the corresponding SNOTEL station has recorded measurable SWE ($SWE > 0$). This assumption would imply that each station measurement, corresponding to an area of approximately $9m^2$ (snow pillow standard size is $3 \times 3m$), is representative of conditions across the entire $500 \times 500m$ ($250 km^2$) image pixel. However, this assumption can be poor for several reasons, an important one being that the MODIS sensor fails to map snow when snow depths are less than 4 cm (Hall and Riggs, 2007).

Based on an average snow density of $0.3621 g \cdot cm^{-3}$ (average snow density for one of our SNOTEL stations during WY 2005), 4cm of depth corresponds to approximately 1.4478 cm of SWE. Therefore, to establish a more accurate basis for evaluation, a sensitivity analysis was performed to find a threshold value of recorded SWE above which the pixel could be considered to be snow covered. To do this, the SNOTEL station raster was used as the observed ground truth and the MODIS SCA images as the modeled forecast. The conditional probabilities of hits (observed = snow/forecast = snow), false alarms (observed = no snow/forecast = snow), misses (observed = snow, forecast = no snow) and correct rejections (observed = no snow/forecast = no snow) were computed for 20 different threshold SWE values (from 0 to the maximum value of $SWE = 49.53$ cm recorded at the stations over the observation time period), and an optimal threshold ($SWE = 2.61$ cm) was selected that minimized the sum of the conditional probabilities of misses and false alarms while maximizing the sum of the conditional probabilities of hits and correct rejections.

The following evaluation statistics recommended by Wilks (1995) were then computed:

$$PC = \frac{a+d}{n} \quad (8)$$

$$TS = \frac{a}{a+b+c} \quad (9)$$

$$B = \frac{a+b}{a+c} \quad (10)$$

$$FAR = \frac{b}{a+b} \quad (11)$$

$$H = \frac{a}{a+c} \quad (12)$$

where a = number of hits, b = number of false alarms, c = number of misses, d = number of correct rejections, and $n = a+b+c+d$. The Proportion Correct (PC) and Threat Score (TS) are accuracy statistics, while B is a measure of bias, the False Alarm Ratio (FAR) is a measure of reliability and the Hit rate (H) is a measure of

discrimination.

The Proportion Correct is a good measure of accuracy if the event (snow) and nonevent (no snow) occurred with equal frequency (i.e. 50/50). A completely accurate estimator will achieve a $PC = 1$ ($b = c = 0$) while a completely inaccurate estimator will have a $PC = 0$ ($a = d = 0$). This accuracy measure would be most useful during the winter months. During the Fall and Spring months snow cover is less frequent than no-snow cover, thus the Threat Score is a better accuracy statistic for those months since it is good for situations in which the event to be forecasted (snow occurrence in this case) occurs less frequently than nonoccurrence. An accurate estimator will achieve a $TS = 1$ ($b = c = 0$) while an inaccurate estimator will have a $TS = 0$ ($a = 0$). An unbiased estimator will achieve $B = 1$ ($b = c = 0$); $B > 1$ indicates that snow is estimated more often than observed, while $B < 1$ indicates that snow is estimated less often than observed. A good estimator will also achieve a False Alarm Ratio close to 0 (no false alarms) while FAR close to 1 indicates very poor performance (no hits). Finally, a good estimator will achieve a Hit Rate close to 1 (no misses).

The results are summarized in Table 2. We see that:

1. Both Terra and Aqua images have an overall high fraction of correctly classified snow-covered and no snow-covered pixels ($PC = 0.85, 0.82$) but tend to underestimate the occurrence of snow ($B = 0.85, 0.91$) therefore commission errors are low. The latter is shown by low FAR values (0.08, 0.15). The H values (0.78, 0.77) are consistent with the B values, both images underestimate the occurrence of snow although not too heavily. Overall, the Terra images have better accuracy than the Aqua images and both images show higher omission errors than commission. These results are consistent with those of Parajka and Blöschl (2008), Xie et al. (2009) and Gao et al. (2010) although they used other measures of accuracy.
2. The Terra/Aqua combination reduces the accuracy of the MOD10A1 image. This result is different from other studies where this step actually increased the images' accuracy (Parajka and Blöschl, 2008; Xie et al., 2009; and Gao et al., 2010) This might be due to differences in evaluation statistics used, the amount of stations with ground truth observations used to calculate the accuracies and their distribution throughout the watersheds, and/or the fact that the other studies included the entire Water Year including summer where mapping accuracies are high because there is no snow and therefore omission and commission errors are practically non-existent.
3. The time interpolation method achieves the best overall accuracy, and provides consistently better evaluation statistics – better even than those of the original images. This is likely due to the larger number of cloud free pixels, and therefore a higher amount of estimate-observation pairs used in computing the statistics. This idea assumes that the accuracy of the Terra images remains constant throughout the days, which is reasonable given that the watershed characteristics do not change from day to day (e.g. land cover, terrain, etc.) so that the only source of inaccuracy that can change from day to day is precisely the clouds that were removed.
4. The other three methods have varying results, with LWLR achieving the worst PC , TS , B and H statistics. The lower accuracy results for the LWLR might be caused by the size of the window used for the analysis, a smaller window has better accuracy results but removes less clouds (López-Burgos, 2010). The most appropriate tradeoff between these

two qualities can be chosen by the user based on the future applications of the final images. The lower accuracy of the LWLR may also be due to the thresholds used to decide if a pixel has snow or not. One way to improve this would be for the user to give more weight to the minimization of the sum of conditional probabilities of commission errors than omission errors (refer to section 4.4) since this would give more conservative results for snow cover. It is better to plan for less SWE in the watershed (underestimation) and find that there is more usable SWE than to plan for a higher amount of SWE (overestimation) and find out there was actually less. Development of a method that improves this step remains open.

5. The sequential approach is second best in terms of accuracy (PC = 0.85, TS = 0.74), matches the Terra/Aqua combination in terms of bias (0.89), is middling in terms of false alarms (0.10) and second in terms of hit rate ($H = 0.81$). Table 3 shows how the evaluation statistics changed with each step during the sequence application. As shown on the table, the first step (Terra/Aqua combination) worsens all the evaluation statistics except for the Hit Rate, which stayed the same. Following this step, the Time Interpolation was applied and the evaluation statistics improved to values even better than the original Terra image except for the FAR but still showed an improvement on this statistic over the Terra/Aqua combination. The Spatial Interpolation did not cause any significant changes and the LWLR worsened the evaluation statistics of the sequence but the final outcome is equal or better than the original Terra image. The final FAR statistic is higher than the original image and so is the Bias. This is most likely caused by an increase in the commission errors for the FAR and also a decrease in omission errors for the Bias. A way of improving or balancing this result was mentioned above.
6. Overall, all methods showed higher omission errors than commission errors. These results are similar to those of Parajka and Blöschl (2008), Xie et al. (2009) and Gao et al. (2010). In addition, the final result of the sequence has similar or higher accuracy than the original MODIS images, this result is also similar to the aforementioned papers with respect to the methods they used and final products.

(Please add Table 3 to the list of tables and correct the PC and B values for the Sequential method from PC= 0.84 to PC = 0.85 and B = 0.90 to B= 0.89)

Table 3. Evaluation statistics for each step of the sequence

Step	PC	TS	B	FAR	H
Unaltered Terra	0.85	0.73	0.85	0.08	0.78
Unaltered Aqua	0.82	0.68	0.91	0.15	0.77
T/A combination	0.83	0.71	0.89	0.12	0.78
Time Interpolation	0.86	0.76	0.92	0.10	0.83
Spatial Inter.	0.86	0.76	0.91	0.10	0.82
LWLR/End of Sequence	0.85	0.74	0.89	0.10	0.81

7 Discussion and conclusions

7.1 Differences in performance between the cloud removal algorithms

The cloud removal algorithms examined in this study gave differing results with regards to the number of cloudy pixels removed. In terms of number of cloudy pixels removed, LWLR was best, followed by time interpolation, and Terra/Aqua combination in that order. Spatial interpolation provided very little cloud removal performance. When removing clouds, all methods added more snow pixels than land pixels, a reasonable result since clouds tend to concentrate at higher elevations where more snow accumulates and the time period used only includes months when there is snow presence. The methods also share the same seasonal patterns of cloud removal and distribution of added snow-covered and no snow-covered pixels. Overall, the sequential approach achieved a 94% reduction of cloud covered pixels (from an overall cloud cover of 38.7683% to 2.4084 %), which can be considered to be very successful. The results would, of course, be even better if we did not include images from days that were completely covered by clouds and for which the algorithms had little effect. Furthermore, the sequential approach reduced the cloud cover to less than 10% for all months with a maximum of 9% in February followed by 4% in January and a minimum of 0.15% in April, showed similar seasonal trends as the methods applied separately and left only 5% of the images with more than 10% cloud cover.

It is likely that LWLR is able to remove more cloudy pixels than any other method because it uses information from a window of pixels of area $\sim 775 \text{ km}^2$ around each cloudy pixel ($\sim 7\%$ of the total area of the watershed). This gives the method an advantage over the other algorithms, which only draw upon information at the same pixel or in the 8 pixels neighborhood having an area of 0.86 km^2 . Because the Terra/Aqua combination and Time Interpolation methods depend on the dispersion of clouds during the time step, cloud cover that persists for several days can confound the methods. In contrast, even if cloud cover persists on the same area for several consecutive days, the LWLR method can use information from non-cloudy pixels at lower elevations or high elevations elsewhere in the watershed.

On the other hand, LWLR is unable to remove cloud cover on days when clouds cover all or much of the watershed. In such situations, the Terra/Aqua combination and time interpolation methods can be effective, provided the clouds do not persist for too many hours or days. Further, the Terra/Aqua combination helps to complete the time series, a positive and useful aspect that none of the other algorithms possess. Taken together, it therefore makes logical sense that a sequential combination of these algorithms provides synergistic effects. Overall, spatial interpolation provides only minor improvements, and its abilities are partially replicated by the LWLR approach. Elimination of this method from the sequence will reduce processing time and have very little overall impact on the results.

7.2 Differences in evaluation results between the cloud removal algorithms

Time interpolation provides the best overall accuracy with consistently better evaluation statistics (better than that of the original images). LWLR removes the largest amount of cloud cover but has worse evaluation performance, due inherently to the extrapolation of information from surrounding pixels. This may be aggravated by the large window size used. Terra/Aqua combination also reduces accuracy even though the observations are from the same day and the instruments should give similar results. One reason for this may be that the MODIS instrument

aboard Aqua has degraded data quality in band 6 (70% of the band 6 detectors have been identified as non-functional), pertaining to the short-wave infrared portion of the electromagnetic spectrum ($\sim 1.6 \mu\text{m}$) where snow surfaces have low reflectance, and used in the computation of Normal Difference Snow Index (NDSI) used for SCA computations. As a consequence, band 6 was substituted with band 7 in an Aqua-specific algorithm to map snow. The 2005 accuracy of the Aqua MYD10A1 images had not yet been assessed (Hall and Riggs, 2007) and no papers were found that do so.

Overall, sequential application of the methods achieved accuracy similar to that of original MOD10A1 image, with improvements in certain areas. Further, the Threat Score increased (which means that more pixels were correctly classified as snow), the Bias came closer to one (meaning that in general there were relatively more hits than misses and false alarms and under-estimation of snow events was reduced), and the Hit Rate moved closer to one (meaning that more pixels were correctly classified as snow).

7.3 Conclusions

The sequential cloud removal approach has the potential to be applied successfully at other locations. Whereas the Terra/Aqua combination, Time Interpolation and spatial interpolation methods have been previously tested in watersheds having different topographical and climate characteristics (Parajka and Bloßsch, 2008; Gafurov and Bardossy, 2009; Xie et al., 2009; and Gao et al., 2010), the LWLR Probability of snow method has not previously been used for cloud removal and SCA estimation and is the only method found so far that uses probabilities in the rules used to remove clouds. It should therefore be subjected to more extensive testing (e.g. how to choose the better thresholds to get binary results). It may also be interesting to examine whether the additional use of slope information and topographic shading as explanatory variables would prove helpful.

In addition, in order to make better comparisons with results of other papers, it is necessary to address the mapping accuracies of the methods on a monthly basis to find out the tradeoffs between cloud removal and mapping accuracies during transitional and stable periods. This step was not included in the original version of this study (López-Burgos, 2010) and therefore could not be added to this paper. However, the seasonal differences in mapping accuracies of the Terra/Aqua combination and Time Interpolation are expected to be similar to the ones found by Parajka and Bloßsch (2008) and Gao et al. (2010). In general, mapping accuracies for these methods are higher during the snow-covered periods and lower for the snow accumulation and snow ablation periods. This is due to the sudden change of snow-cover from one day to the other during the transition periods. These two other studies also found that increasing the temporal window used in the time interpolation/filter reduces the accuracy of the images. This would be even more important during the transitional periods.

The most closely related study that we could use to give an estimate of the seasonal mapping accuracy of the LWLR is that of Parajka et al. (2010) who used a regional snow-line method for estimating the snow cover from MODIS during cloud cover. This method would be similar because it also uses snow cover information and elevation from cloud free pixels to decide if there is snow or not on the cloudy pixel. However, this method does not take into consideration

the aspect of the pixels, and the topography and snow regime in Arizona and Austria are different, so the mapping accuracies would not be strictly comparable (Parajka et al., 2010). It does give a general idea of how the seasonal mapping accuracies of LWLR would look like.

Our future work will include the assimilation of corrected SCA images into a distributed hydrologic model to reduce the uncertainty of streamflow forecasts. McGuire et al. (2005) and Andreadis and Lettenmaier (2006) have demonstrated that although assimilation of MOD10A1 images into the variable infiltration capacity model can provide favorable results, cloud cover reduces the usefulness of this data and requires the data to be assimilated in a non-continuous manner. Since removal of cloud cover can result in considerably larger estimates of snow, application of the algorithm developed in this work should help to simplify the assimilation process while improving the model estimates of various hydrological states and fluxes.

Other changes to the manuscript:

1. **The title has been changed to:** *“Reducing cloud obscuration of MODIS Snow Cover Area products by combining spatio-temporal techniques with a probability of snow approach.”*
2. **Study Area: Change to SI Units:** Section 3.1 should now read:

The Upper Salt basin (Fig. 1), having a drainage area of $\sim 11152.5 \text{ km}^2$, is a major source of surface water for the Phoenix Metropolitan Area. On an average annual basis, the precipitation varies from 381 to 1193.8 mm, runoff is $\sim 24.92 \text{ cms}$, minimum temperature varies from -13.9 to 3.9°C , and maximum temperature varies from 17.2 to 39.4°C (OCSOSU 2006a,b). The elevation ranges between 674m and 3472m a.s.l., and land cover types are primarily ponderosa pine (65 %), chaparral (26 %), pinyon pine-juniper (10 %), and desert grassland (Rinne, 1975). Winter precipitation is of paramount importance, since snowmelt can account for up to 85% of the usable water (Hawkins, 2006), and so snow accumulation and ablation are very important to water resources management.

The local water and power utility, the Salt River Project (SRP), currently relies on sporadic helicopter flights to verify snow cover extent. In addition, daily snowpack data from four SNOTEL locations at higher elevations in the eastern part of the watershed (where snow water storage is greatest; Fig. 1) are used for making streamflow forecasts, but this can lead to erroneous underestimates of streamflow since the point measurements are not representative of the areal pattern of snow accumulation. Consequently, the utility is interested in using remotely sensed SCA imagery to improve the accuracy of their forecasts.

3. **Section 3.3:** We mentioned the COOP data because they are widely used in the US for snow studies and are located at lower elevations in the watershed but we could not use them in our evaluation process because they seemed suspicious. We leave it to the discretion of the editor if this section should be removed or not.
4. **Section 4.4 should now read:**

The locally weighted logistic regression (LWLR) method uses relationships between the spatial and topographic attributes of pixels surrounding a cloudy pixel to estimate the “Probability of Snow Occurrence” (PSO, Eq. 1). This method is adapted from Clark and Slater (2006), who used precipitation observations at sparsely located meteorological stations, and spatial maps of elevation, latitude and longitude, to estimate daily precipitation totals across complex terrain in Western Colorado. Here, we used elevation and aspect as the explanatory variables.

To estimate PSO at a cloudy pixel, the LWLR method weights the information from neighboring pixels inversely with distance, and fits the data to a logistic curve. The following equations are used to calculate the PSO of a cloudy pixel ($\text{PSO}_{\text{icloud}}$):

$$\text{PSO}_{\text{icloud}} = \frac{1}{1 + \exp(-\mathbf{Z}_{\text{icloud}}\hat{\boldsymbol{\beta}})} \quad (1)$$

$$\hat{\boldsymbol{\beta}}_{\text{new}} = \hat{\boldsymbol{\beta}}_{\text{old}} + (\hat{\mathbf{X}}^T \hat{\mathbf{W}} \hat{\mathbf{V}} \hat{\mathbf{X}})^{-1} \hat{\mathbf{X}}^T \hat{\mathbf{W}} (\mathbf{Y}' - \boldsymbol{\pi}) \quad (2)$$

$$\pi = \frac{1}{1 + \exp(-\hat{\mathbf{X}}_{\text{ipix}} \hat{\boldsymbol{\beta}}_{\text{old}})} \quad (3)$$

$$\hat{\mathbf{V}} = \text{diag} \left(\hat{v}_{\text{ipix}, \text{ipix}} = \pi_{\text{ipix}} [1 - \pi_{\text{ipix}}]^T \right) \quad (4)$$

$$\hat{\mathbf{W}} = \text{diag} \left(\hat{w}_{\text{ipix}, \text{ipix}} = \left[1 - \left(\frac{d_{\text{ipix}}}{\text{MAXD}} \right)^3 \right]^3 \right) \quad (5)$$

where $\mathbf{Z}_{\text{icloud}}$ is a vector of elevation and aspect information for the cloudy pixels indexed as icloud, \mathbf{X} is vector of elevation and aspect information for the non-cloudy pixels, $\boldsymbol{\beta}$ is a vector of parameters, \mathbf{Y}' is a 0–1 vector indicating snow occurrence or not on the non-cloudy pixels, \mathbf{W} is a diagonal matrix of weights to be assigned to each non-cloudy pixel, \mathbf{V} is a diagonal matrix of the variance associated with the estimate of snow occurrence at each non-cloudy pixel, π_{ipix} indicates the PSO at each non-cloudy pixel, d_{ipix} is the distance from a non-cloudy pixel to the cloudy pixel, and **MAXD** is a coefficient specifying the window size used around the cloudy pixel (see Clark and Slater (2006) and Loader (1999) for details regarding implementation).

In regression analysis one makes use of one or more known independent explanatory variables \mathbf{X} to predict the value of an unknown dependent variable \mathbf{Y} . In Logistic regression the model fitted to the data is a logistic function or logistic curve (Eq. 1) that represents the probability of occurrence of a ‘categorical’ (0,1) variable \mathbf{Y} given \mathbf{X} , and the regression coefficients are estimated using maximum likelihood estimation via an iterative process (Eqs 2, 3 and 4). Note that although \mathbf{X} can take any value on $[-\infty, \infty]$, the value of \mathbf{Y} is confined to the range $[0, 1]$; in our case \mathbf{Y} represents the probability of snow occurrence. The regression is computed using information from the pixels that are not obscured by clouds, and then used to estimate probability of snow for the cloudy pixels in the image. In addition, the regression is computed and applied in a locally weighted, manner, meaning that pixels closer to the cloudy pixel have more weight than pixels those that are further. The weights are calculated via Eq. 3

For this study, we tested different values for the window size **MAXD** (from 5 to 45 pixels) for their ability to provide statistically robust results, reduce significant numbers of cloudy pixels, and require reasonable computational time. Reliability was evaluated using the Brier Score (BS) verification statistic (see Wilks, 2005) and Clark and Slater (2006) computed from the joint distribution of the LWLR forecast probabilities and the observed snow/land pixels in the image. Although the best results (not shown) were

obtained using $MAXD = 45$, we instead chose $MAXD = 30$ due to the relatively high performance achieved using only 1/3 rd of the computer time (22 h as opposed to 67 h on a 2.66 GHz Dual-Core Intel Zeron computer).

Note that LWLR does not, by itself, automatically reclassify cloudy pixels, but only provides an estimate of the probability that the pixel is actually snow or land. Therefore a minimum probability threshold must be selected to convert the probability to a binary outcome. To select this threshold, we conducted a sensitivity analysis by varying the threshold from 0 to 1 in steps of 0.025, and chose the threshold value that minimized the sum of the conditional probabilities of commission and omission errors computed over all non-cloudy pixels inside the window (for details see López-Burgos, 2010). The threshold was separately selected for each cloudy pixel to which LWLR was applied.

Figure 4 shows an example of the transition from original to corrected MOD10A1 image (26 November 2004), along with maps of the estimated PSO and the values of the thresholds selected.

5. **Fig 2:** We agree we should have used SI units. However, we currently cannot regenerate this graph as the first author no longer has access to the program and data used. 1 inch = 2.54 cm. Cloud cover accounts for ~ 39 % of the MOD10A1 pixels inside the watershed during the time period. This number is highest during January and February with ~50% and ~ 69% cloud cover respectively. Fig. 2 shows how during the months with the highest amount of snow, MOD10A1 shows almost now snow cover on the ground or significantly less than expected for several days.
6. **Fig. 3, 4, 7:** As mentioned above, we are unable to regenerate the figures. The subject matter is Snow Cover Area and therefore we think the equal area Sinusoidal projection is appropriate to show the results.

Again, thank you for your comments. They have helped us analyze in more depth the results we obtained.

Sincerely

Viviana López-Burgos