Inverse Modeling of GPS Multipath for Snow Depth Estimation—Part II: Application and Validation

Felipe G. Nievinski and Kristine M. Larson

***Abstract*—GPS multipath reflectometry (GPS-MR) is a technique that uses geodetic quality GPS receivers to estimate snow depth. The accuracy and precision of GPS-MR retrievals are evaluated at three different sites: grasslands, alpine, and forested. The assessment yields a correlation of 0.98 and an rms error of 6–8 cm for observed snow depths of up to 2.5 m. GPS-MR underestimates *in situ* snow depth by 10%–15% at these three sites, although the validation methods do not measure the same footprint as GPS-MR.**

***Index Terms*—Artificial satellites, electromagnetic reflection, global positioning system, interferometers, multipath channels, radar remote sensing.**

I. Introduction

A NEW method to measure snow depth based on GPS multipath present in signal-to-noise ratio (SNR) observations was introduced in [1]. An improved forward/inverse methodology has been formulated in [2]. It capitalizes on knowledge about the antenna response and the physics of surface scattering to aid in retrieving the unknown snow conditions. Although it had been demonstrated with simulations [2], its application to actual measurements collected under real-world conditions poses challenges that must be overcome to prove the concept.

Here, we examine snow depth retrievals based on the GPS multipath algorithm and assess both the precision and accuracy of the method. Multiple metrics are developed to assess the quality of the results. The accuracy of the method is evaluated by comparing with *in situ* data over a multiyear period. Three field sites were chosen to highlight different limitations in the method, both in terms of terrain and forest cover.

We start with a general development applicable to all sites, in which intermediate results are explored in more detail. Then, we proceed to show the final snow depth time series at each of the three sites, validating them against independent *in situ* measurements.

Manuscript received June 16, 2013; revised November 12, 2013; accepted December 10, 2013. Date of publication January 24, 2014; date of current version May 22, 2014. This work was supported by the National Science Foundation (NSF) (EAR 0948957 and AGS 0935725), by the National Aeronautics and Space Administration (NASA) (NNX12AK21G), and by a CU interdisciplinary seed grant. Dr. Nievinski has been supported by a Capes/Fulbright Graduate Student Fellowship (1834/07-0) and a NASA Earth System Science Research Fellowship (NNX11AL50H). Some of this material is based on data, equipment, and engineering services provided by the Plate Boundary Observatory operated by UNAVCO for EarthScope (www.earthscope.org) and supported by the NSF (EAR-0350028 and EAR-0732947).

F. G. Nievinski is with the Departamento de Cartografia, Universidade Estadual Paulista "Júlio de Mesquita Filho," Presidente Prudente, SP 19060-900 Brazil (e-mail: fgnievinski@gmail.com).

K. M. Larson is with the Department of Aerospace Engineering Sciences, University of Colorado, Boulder, CO 80309 USA (e-mail: kristinem.larson@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2297688

II. General Development

In this section, we examine the satellite coverage, over time and space; the matching of model and measurements, in terms of observation residuals; the quality control (QC) procedures used to mitigate anomalous results; and combinations of estimates obtained from different satellites.

A. Satellite Coverage and Track Clustering

All GPS multipath reflectometry (GPS-MR) retrievals reported here are based on the newer GPS L2C signal. Of the ∼30 GPS satellites, 8–10 L2C satellites were available between 2009 and 2012 (8, 9, and 10 satellites at the end of 2009, 2010, and 2011, respectively). Satellite observations are partitioned into ascending and descending portions, yielding approximately 20 unique tracks per day at a site with good sky visibility. GPS orbits are highly repeatable in azimuth, with deviations at the few-degree range over a year, translating into ∼50–100-cm azimuthal displacement of the reflecting area (corresponding to the first Fresnel zone at 10°–15° elevation angle for a 2-m high antenna; see [3, Appendix]). This repeatability permits clustering individual retrievals by azimuth. It also allows the simplification that estimated snow-free ground reflector heights are fairly consistent from day to day, facilitating the isolation of the varying snow depth during the snow-covered period.

For a given track, its revisit time is also repeatable, amounting to practically one sidereal day. The deficit in time relative to a calendar day results in the track time of the day receding ∼4 min and 6 s every day [4]. This slow but steady accumulation eventually makes the time of day to return to its starting value after ∼one year. As all GPS satellites drift approximately at the same rate, the time between successive tracks remains nearly repeatable. Its reciprocal, the sampling rate, has a median equal to approximately one track per hour, with a low value of one track within 2 h and a high of one track within 15 min; both extremes occur every day, with low-rate idle periods interspersed with high-rate bursts. The time of the day reduced to a fixed day (e.g., January 1, 2000) could also be used to cluster tracks. Neighboring clusters that are close in azimuth and/or in reduced time of the day are expected to be more comparable, as they sample similar conditions and are subject to similar errors. We materialize this notion in terms of diagrams of track clusters, having azimuth and reduced time of day as horizontal

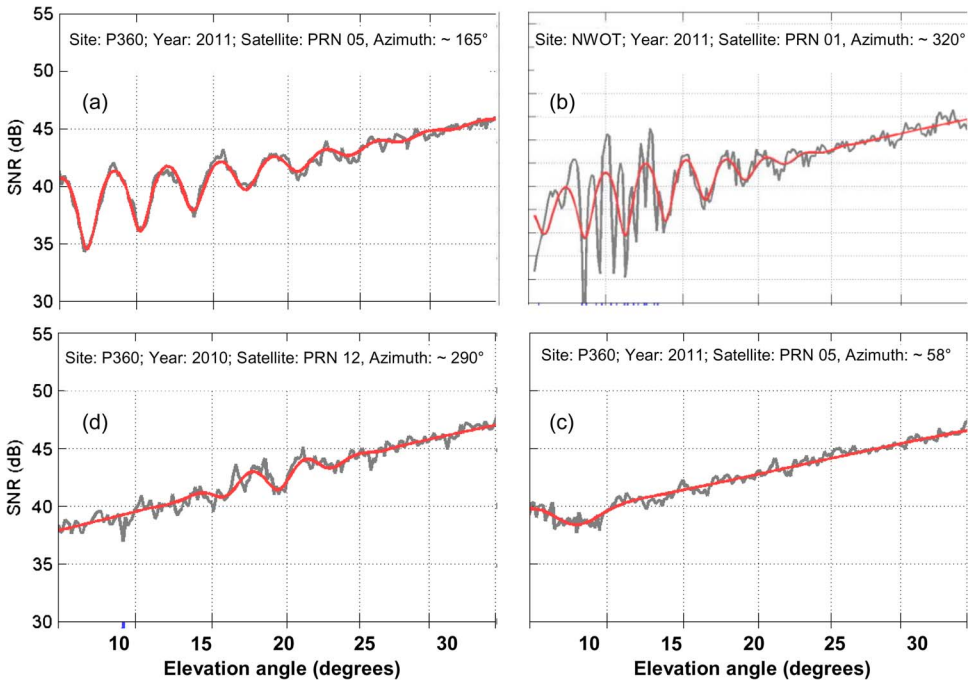

Fig. 1. Examples of observations. (a) Good fit. (b) Presence of secondary reflections. (c) Vanishing interference fringes. (d) Atypical interference fringes.

and vertical axes, respectively; it will be applied to each of the following three sites (Fig. 5).

B. Observations

Fig. 1 shows several representative examples of SNR observations. A typical good fit between measured and modeled is shown in Fig. 1(a), corresponding to the beginning of the snow season. Generally, the model/measurement fit is good when the scattering medium is homogeneous; it deteriorates as the medium becomes more heterogeneous, particularly with mixtures of soil, snow, and vegetation. Here, we discuss genuine physical effects as well as more mundane spurious instrumental issues that degrade the fit but do not necessarily cause a bias in snow depth estimates.

1) Secondary Reflections: Throughout, we have assumed the existence of a single specular reflection, which matches large planar surfaces. Finite and/or nonplanar surfaces represent a departure from this assumption; the first case transforms a simpler reflection into a more complicated diffraction phenomenon, while the second case might as well introduce secondary reflections, originating from disjoint surface regions. Interference fringes become convoluted with multiple superimposed beats [see Fig. 1(b)]. As long as there is a unique dominating reflection, the inversion will have no difficulty fitting it, as the extra reflections will remain approximately zero-mean.

2) Interferometric Power Effects: Random deviations of the actual surface with respect to its undulated approximation—called roughness or residual surface height—will affect the interferometric power P_{i}. SNR measurements will exhibit a diminishing number of significant interference fringes, compared to the measurement noise level [Fig. 1(c)]. This facilitates the model fit, but the reflector height parameter may become ill-determined—its estimates will be more uncertain. Changes in snow density also affect the fringe amplitude.

3) Direct Power Effects: Snow precipitation attenuates the satellite-to-ground radio link which affects SNR measurements through the direct power term. First, this shifts the SNR measurements up or down (in decibels); second, it tilts the trend tSNR as attenuation is elevation angle dependent; third, fringes in dSNR will change in amplitude because of the decrease in the coherent component of the direct power.

Partial obstructions can affect either or both direct P_{d} and interferometric powers P_{i}. In this case, SNR measurements, albeit corrupted, are still recorded. This situation is in contrast to complete blockages as caused by topography. The deposition of snow and the formation of a winter rime on the antenna are a particularly insidious type of obstruction, as their presence in the near-field of the antenna element can easily distort the gain pattern in a significant manner. In the far-field, trees are another important nuisance, so much so that their absence is held as a strong requirement for the proper functioning of multipath reflectometry [3].

4) Instrument-Related Issues: Satellite-specific direct power offsets and also long-term power drifts are to be expected as spacecraft age and modernized designs are launched. In addition, noise power depends on the state of conservation of receiver cables and on their physical temperature. Less subtle incidents are sudden ∼3-dB SNR steps, hypothesized to originate in the receiver switching between the L2C data and pilot subcodes, CM and CL [5].

C. QC (Intraclusters)

Anomalous conditions may result in measurement spikes, jumps, and short-lived rapidly varying fluctuations. For snow-depth-sensing purposes, it is necessary and sufficient to either

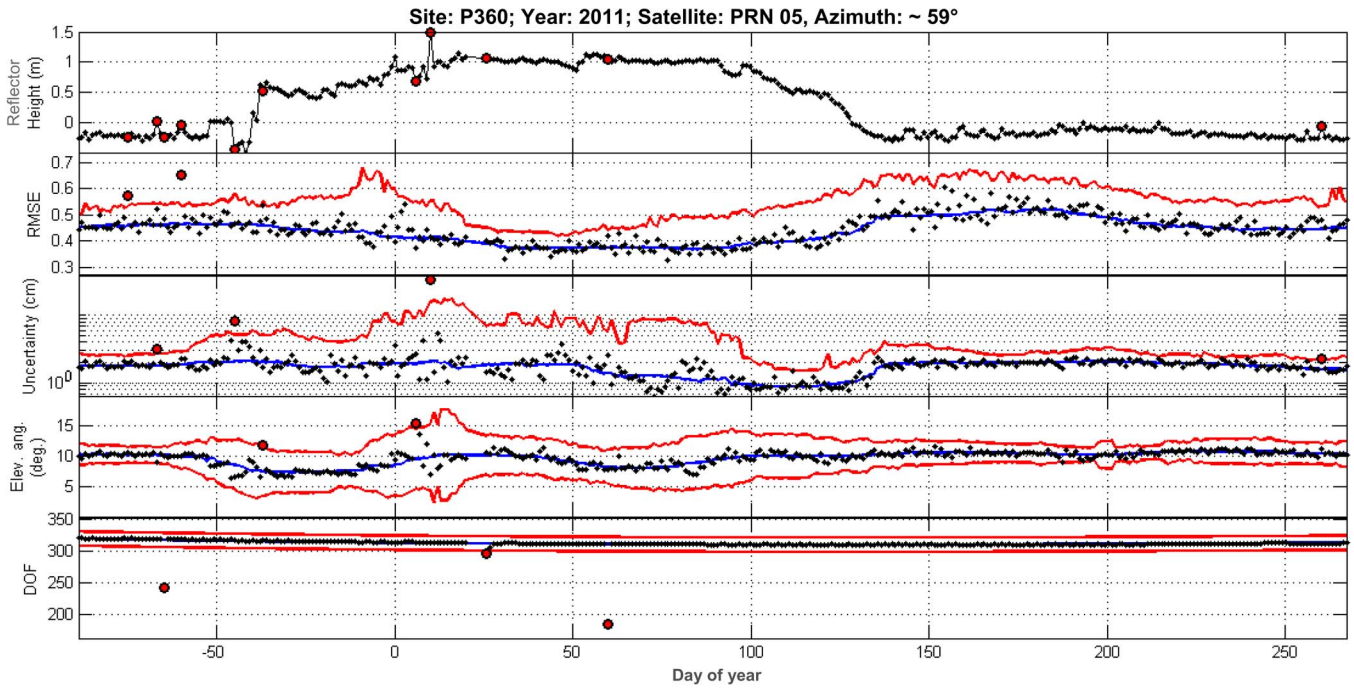

Fig. 2. Time series of QC tests for a single track cluster (satellite PRN 05, azimuth $\sim 59^\circ$, ascending), as observed at the grassland site (P360) in the water-year 2011. In each panel, black dots are independent day-to-day track retrieval statistics. Blue and red lines represent, respectively, the expected tendency and dispersion, both based on a 15-day moving average. The bounds are two sided for normally distributed variables (DOF and peak elevation angle) and one sided for χ^2-distributed variables (uncertainty and rmse). Red circles are retrievals deemed to be outliers, separately for each test in the lower four panels and in conjunction for all tests, in the top panel. RMSE is the same as $\hat{\sigma}_0$ and is expressed in observation units (decibels); DOF is a number count.

neutralize such measurement outliers through a statistically robust fit or detect unreliable fits and discard the problematic ones that could not otherwise be salvaged.

The key to QC lies in grouping results into statistically homogeneous units, having measurements collected under comparable conditions. In our case, azimuth-clustered tracks are the natural starting unit. Secondarily, we must account for genuine temporal variations in the tendency of results, i.e., from beginning to peak to the end of the snow season. The detection of anomalous results further requires an estimate of the statistical dispersion to be expected. Considering that the sample is contaminated with outliers, robust estimators—running median instead of the running mean and median absolute deviation over the standard deviation—are called for, if the first- and second-order statistical moments are to be representative. Given estimates of the nonstationary tendency and dispersion, a tolerance interval can then be constructed such that it bounds, for example, a 99% proportion of the valid results with 95% confidence level. We also desire QC to be judicious, or else, too many valid estimates will be lost. Notice that, in the present intracluster QC, we compare an individual estimate to the expected performance of the track cluster to which it belongs; later, we complement QC with an intercluster comparison of each cluster's own expected performance.

Based on our practical experience, no single statistic detects all of the outliers. In the sections that follow, we present four particular statistics that we have found to be useful. The accompanying time series of estimated reflector heights is shown in Fig. 2 (top panel); these raw reflector heights will be transformed into snow depth in the next section. Even in raw form, the snow season (between days -50 and 150) can be clearly distinguished by the dramatic increase in reflector height retrievals (only the negative bias with respect to the *a priori* value is shown). Several snow precipitation events are discernible as sudden rises in reflector height, followed by a slower decay indicating snow settling. At the end of the season, melting occurs until bare ground eventually becomes exposed. Over the remainder of the year, reflector heights vary just a few centimeters at this site, which indicates that our *a priori* value for the height of the antenna above the ground was accurate and that the site is devoid of vegetation with large amounts of water content.

Statistics that involve a sum of squared values are expected to follow a chi-squared distribution. To accommodate this characteristic, we handle such values in logarithm form, restoring normality. For these second-order statistics, a single-tailed tolerance bound was found more appropriate than the double-tailed intervals applied to first-order statistics.

1) DOF: The simplest statistic is the degree of freedom (DOF), essentially the number of observations per track (modulo a constant number of parameters). The time series shown in Fig. 2 (bottom panel) is stationary, so a global low-order polynomial (instead of a running average) suffices for this statistic. It does a good job detecting problems related to data outages.

2) Goodness of Fit: We use the scaled root-mean-square error (rmse, denoted $\hat{\sigma}_0$ in Part I [2, Sec. V-A]) to test for goodness-of-fit, i.e., how well measurements can be explained adjusting the unknown values for the parameters postulated in equation (24) of Part I [2]. Fig. 2 (second panel from the top) shows a time series of $\hat{\sigma}_0$ and its tolerance bound. Tracks near days -75 reject the null hypothesis of statistical equivalence against the long-term nonstationary average; thus, they are

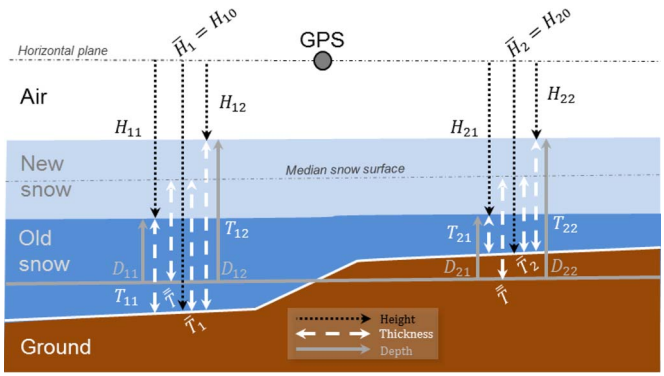

Fig. 3. Snow reflector height H, snow thickness T, and snow depth D as different vertical coordinates; for two track clusters (left and right) and three days (denoted 0, 1, and 2).

deemed outliers. Notice that we neutralized the risk posed by outliers in distorting the tolerance intervals.

3) Reflector Height Uncertainty: Sometimes the fit is good, but the reflector height uncertainty $\hat{\sigma}_{\widehat{H}_B}$ is bad as the multipath modulation is gone—there are not many oscillations greater than the measurement noise level [Fig. 1(c)]. Such cases are missed by the previous statistics but are detected by this one (middle panel in Fig. 2). High uncertainty is more common during heavy snowfalls.

4) Peak Elevation Angle: The peak elevation angle (defined in Part I [2, Sec. VI-B]) behaves much like a random variable, as it is determined by a multitude of factors. We can thus form a tolerance interval to detect outliers. This statistic was found to perform especially well in cases that were particularly challenging for the previous statistics. For example, some fits yielded small residuals and produced unsuspecting reflector height uncertainties, but the result corresponded to a substantially different peak elevation angle. This was the case not just because the SNR oscillations that are typically found in SNR turned out to be missing on that day. The aggravating factor is the presence of oscillations at atypical elevation angles. Despite being well fit and well determined, they are generated by different reflecting conditions, compared to those found on most other days [see Fig. 1(d)].

D. Combinations (Interclusters)

In the previous section, we dealt with results on a cluster-by-cluster basis. Now, we are interested in combining multiple clusters. The main purpose is to average out random noise. Noise mitigation aims at not only coping with measurement errors but also compensating for model deficiencies, to the extent that they are not in common across different clusters.

1) Vertical Datum: Before we combine different clusters, we have to address their long-term differences. Recall that total reflector height $H = H_A - H_B$ is the difference between its *a priori* value H_A and the estimated bias H_B; it is positive downward, reckoned from the horizontal plane containing the receiving antenna. For easier comparison, we denote H_{ij} the ith day and jth cluster estimate; $\mathcal{W}$ will denote the set of winter days, indexed by i (see Fig. 3). The initial situation is that snow surface heights H_{ij} will be greater downhill and smaller uphill; we amend it as follows. The clusterwise uncertainty-weighted median of estimated reflector heights H_{ij} (evaluated in the snow-free period $i \notin \mathcal{W}$) is taken as the ground height, valid for all days on a cluster-by-cluster basis

$$\overline{H}_j = \underset{i \notin \mathcal{W}}{\text{med}}\, H_{ij}. \tag{1}$$

Subtracting ground heights $\overline{H}_j$ from their respective snow surface heights H_{ij} results in snow thickness values

$$T_{ij} = H_{ij} - \overline{H}_j \tag{2}$$

which is a completely physically unambiguous quantity. Snow thickness is more comparable than snow heights across varying-azimuth track clusters. However, snow tends to fill in ground depressions, so thickness exhibits variability caused by the underlying ground surface, even when the overlying snow surface is relatively uniform. Further cluster homogeneity can be achieved by accounting for the temporally permanent although spatially nonuniform component of snow thickness. This is achieved by defining snow depth as

$$D_{ij} = T_{ij} - \delta\overline{T}_j \tag{3}$$

where the median thickness deviation $\delta\overline{T}_j = \overline{T}_j - \overline{\overline{T}}$ removes the clusterwise median thickness $\overline{T}_j = \text{med}_{i \in \mathcal{W}} T_{ij}$ (evaluated in the snow-covered period $i \in \mathcal{W}$) and restores the site-wide median thickness $\overline{\overline{T}} = \text{med}\, \overline{T}_j$. In effect, the vertical datum is established having the shape of the median snow surface and lowering its mean level closest to the ground level (although without conforming to the ground topography). As the deviations $\delta\overline{T}_j$ sum up to zero, average depth retains the same physical meaning as average thickness, albeit showing a smaller intercluster variability.

2) Averaging: The averaging of snow depths collected for different track clusters employs the inversion uncertainties to obtain a preliminary running weighted median D_k, calculated at a given spacing $k = 1, 2, \ldots$ (e.g., daily postings), with overlapping windows or not. The preliminary postfit residuals $\delta D_{ij} = D_{ij} - D_k$ then go through their own averaging—necessarily employing a wider averaging window (e.g., monthly)—which produces scaling factors for the original uncertainties. The running weighted median is then repeated, producing final averages. The variance factors reflect the fact that some clusters are better than others, vis-à-vis obstructions, etc.

Thus, the final GPS estimates of snow depth follow from an averaging of all available tracks, whose individual snow depth values were previously estimated independently. A new average is produced twice daily, utilizing the surrounding 1–2 days of data (depending on the data density), i.e., 12-h posting spacing and 24-h moving window width. The averaging interval must be an integer number of days, so as to minimize the possibility of snow depth artifacts caused by variations in the observation geometry, which repeats daily.

In the following site results (Figs. 6, 8, and 10), a dark-gray band denotes the 95% confidence interval for the average, which follows from the averaged least-squares uncertainty, scaled by the rms of intercluster residuals and then expanded based on Student's t-distribution for the number of tracks

TABLE I
SITE SUMMARY; LAST TWO COLUMNS REFER TO ITS SEPARATION FROM NEAREST SNOTEL STATION

Code	Environment	Latitude (degrees)	Longitude (degrees)	Altitude (m)	Horizontal (km)	Vertical (m)
P360	Grassland	44.317852	-111.450677	1857.9	~ 12	-60
RN86	Forested	41.864856	-111.502514	2590.7	0.350	~ 0
NWOT	Alpine	40.055387	-105.590527	3522.5	4	+380

TABLE II
IN SITU DATA QUANTITY; REPLICATION INDICATES THE NUMBER OF VALUES SAMPLED PER EPOCH

Code	Duration	Interval	# Epochs	Type	Replication
P360	6 mon.	~ 6 h	500	webcam	1
RN86	7 mon.	~ 2-3 week	9	manual	20-150
NWOT	3 yr.	~ 2-3 week	60	manual	1

available. The light-gray band denotes the 95% simultaneous prediction interval for a random observation, utilized to detect and reject outliers. The relationship between the two statistics is analogous to the standard error of the mean $\sigma/\sqrt{n}$ and the standard deviation σ in an n-element sample. Individual GPS tracks that passed QC are shown as gray dots.

III. SITE-SPECIFIC RESULTS

GPS-MR snow depth retrieval is now further explored at three stations (Table I) and over a longer period (up to 3 years). Throughout, we assess the performance of the GPS against independent nearly colocated *in situ* measurements (Table II). We also compare the GPS estimates to the nearest SNOTEL station [6]. Although not colocated with GPS, SNOTEL data—widely used for operational snow monitoring in the U.S.—are important because they provide accurate information on the timing of snowfall events.

A. Forested Site (RN86)

GPS site RN86 was installed at the T. W. Daniel Experimental Forest in fall 2011, so we show results for only one water-year (which is the period starting from October 1 to September 30 of the following year, encompassing the northern-hemisphere winter; a water-year is designated by the calendar year in which it ends). Topographical slopes range from 2.5° to 6.5° (at the 2-m spatial scale), with an average of $\sim 5^\circ$ within 50-m radius around the GPS antenna. RN86 was specifically built to study the impact of trees on GPS snow depth retrievals (Fig. 4). Ground crews manually collected *in situ* measurements around the GPS antenna approximately every other week starting in November 2011. Measurements are made every 1–2 m from the GPS up to 25–30 m. In the second half of the year, the sampling protocol was changed to azimuths of 0° (N), 45° (NE), 135° (SE), 180° (S), 225° (SW), and 315° (NW). With these data, it is possible to obtain *in situ* average estimates, with their own uncertainties (based on the number of measurements), which allows a more meaningful comparison.

Fig. 4. Aerial view of the forested site (RN86) around the GPS antenna (marked with a circle).

The diagram of track clusters (Fig. 5) indicates a reduced visibility at the current site, compared to other sites. Also, the cluster uncertainty is not as favorable, as indicated by the small-radius circles. Notice that clusters are concentrated due south, with only two clusters located within $\pm 90^\circ$ of north. Therefore, the GPS average snow depth is not necessarily representative of the azimuthally symmetric component of the snow depth. In the presence of an azimuthal asymmetry in the snow distribution around the antenna, the GPS average is expected to be biased toward the environmental conditions prevalent in the southern quadrant. To rule out the possibility of an azimuthal artifact in the comparisons, we have utilized only the *in situ* data collected along the SE/S/SW azimuths.

The comparison shows a generally excellent agreement between GPS and *in situ* data (Fig. 6). The first four and the last one *in situ* data points were collected with coarser spacing and/or smaller azimuthal coverage, which may be partially responsible for different performance in the first and second halves of the snow season. The correlation between GPS and *in situ* snow depths at RN86 amounts to 0.990, indicating a very strong linear relationship. Inspecting the individual differences at each of nine visits (Table III), we find that all are within the corresponding uncertainty, which is somewhat large given the propagation of GPS and *in situ* uncertainties. The GPS uncertainties are generally slightly smaller than the *in situ* ones, with a few exceptions when the number of usable tracks

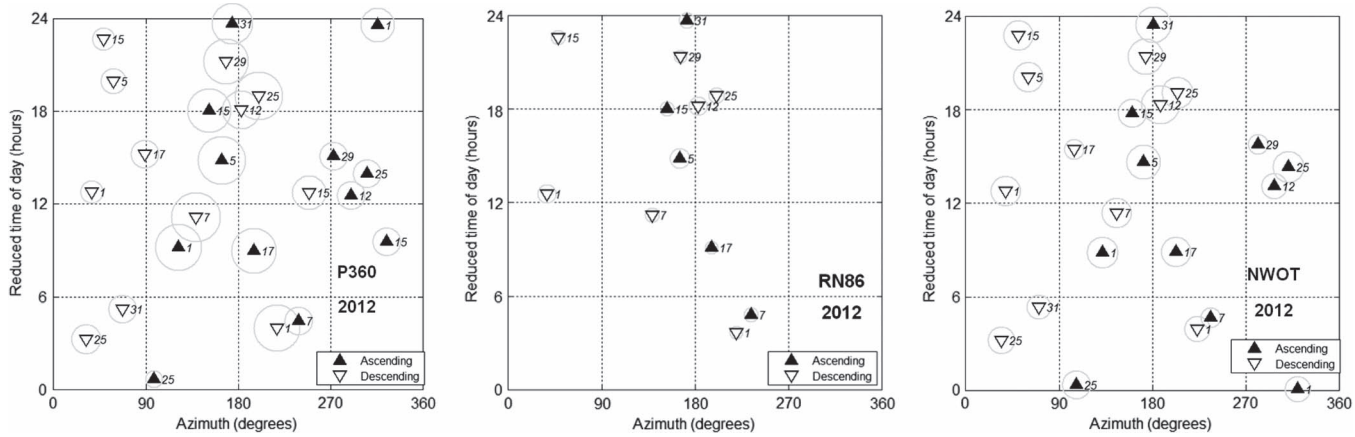

Fig. 5. Diagram of satellite track clusters available for at least 50% of the water-year 2012 at the grassland site (P360, left), forested side (RN86, middle), and alpine site (NWOT, right); italic labels denote satellite PRN number.

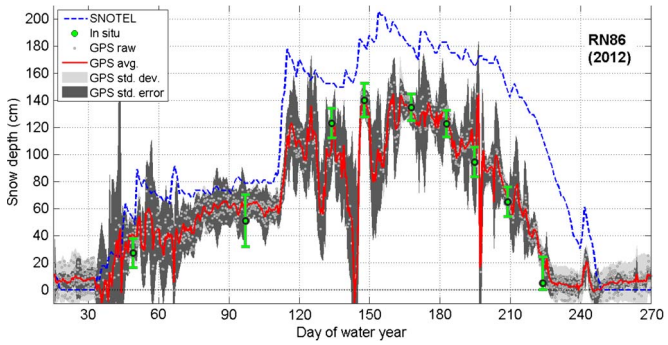

Fig. 6. Snow depth measurement at the forested site (RN86) for the water-year 2012; see text for discussion and description of details; the SNOTEL site is ∼350 m distant and at the same altitude.

TABLE III
SNOW DEPTH AT THE FORESTED SITE (RN86) FOR EACH SAMPLED DAY. THE 95% UNCERTAINTY IS GIVEN AFTER ±. COLUMN "DOY-W" DENOTES DAY OF WATER-YEAR, AND "NUM. OBS." IS THE NUMBER OF INDEPENDENT *IN SITU* OBSERVATIONS (UNKNOWN NUMBERS ARE INDICATED BY ∗, IN WHICH CASE THE *IN SITU* UNCERTAINTY IS TAKEN AS THE MEDIAN VALUE)

Year	Month	Day	DOY-W	GPS (cm)	*In situ* (cm)	Diff. (cm)	Num. obs.
2011	11	18	49	40 ± 11	27 ± 11	13 ± 15	4
2012	1	5	97	60 ± 18	51 ± 19	9 ± 26	*
2012	2	10	134	107 ± 22	123 ± 11	-16 ± 24	10
2012	2	24	148	138 ± 10	140 ± 12	-2 ± 16	15
2012	3	15	168	129 ± 8	135 ± 10	-6 ± 12	14
2012	3	30	183	129 ± 14	123 ± 10	7 ± 17	74
2012	4	11	195	100 ± 43	94 ± 11	6 ± 45	76
2012	4	25	209	68 ± 10	65 ± 11	4 ± 15	74
2012	5	10	224	20 ± 10	5 ± 19	15 ± 21	*

decreases. It is important to bear in mind, though, that there is a tradeoff between uncertainty and temporal representativeness, in that wider averaging windows (Section II-D2) yield smaller uncertainties but also smooth out potentially genuine rapidly changing snow depth variations.

Carrying out a regression between *in situ* and GPS values, the rms of snow depth residuals improves from 9.6 to 3.4 cm. The regression intercept and slope (with corresponding 95% uncertainties) amount to 15.4 ± 9.1 cm and 0.858 ± 0.09 m/m, respectively. According to these statistics, the null hypotheses of zero intercept and unity slope are rejected at the 95% confidence level. This implies that, at this location, GPS snow depth

Fig. 7. Ground conditions in the vicinity of the GPS antenna at the grassland site (P360).

estimates exhibit both additive and multiplicative biases. The latter is proportional to snow depth itself, which means that, compared to an ideal one-to-one relationship, GPS is found to underestimate *in situ* snow depth at this site by 14% ± 9%, albeit the uncertainty is somewhat large.

The SNOTEL sensors are exceptionally close to the GPS at this site, ∼350 m horizontally with negligible vertical separation. However, the former is located within trees, while the latter is located in the periphery of the forest and senses the reflections scattered from an open field. Therefore, only the timing of snowfall events agrees well, not the amount of snow. Although forest density is generally negatively correlated with snow depth [7], exceptions are not uncommon [8], [9], especially in localized clearings exposed to intense solar radiation, where shading of the snow by the trees reduces ablation.

B. Grassland Site (P360)

P360 (Fig. 7) is one of 1100 GPS stations that make up the EarthScope Plate Boundary Observatory (http://pbo.unavco.org). It was installed with the purpose of studying crustal deformation in the western U.S. The typical setup is a 2-m tall metal tripod drilled into bedrock. At the apex rests a choke-ring antenna (boresight facing zenith), both housed within a radome.

P360 is located in an open field; visibility to the ground is unobstructed. The nearest trees are ∼200 m away due west;

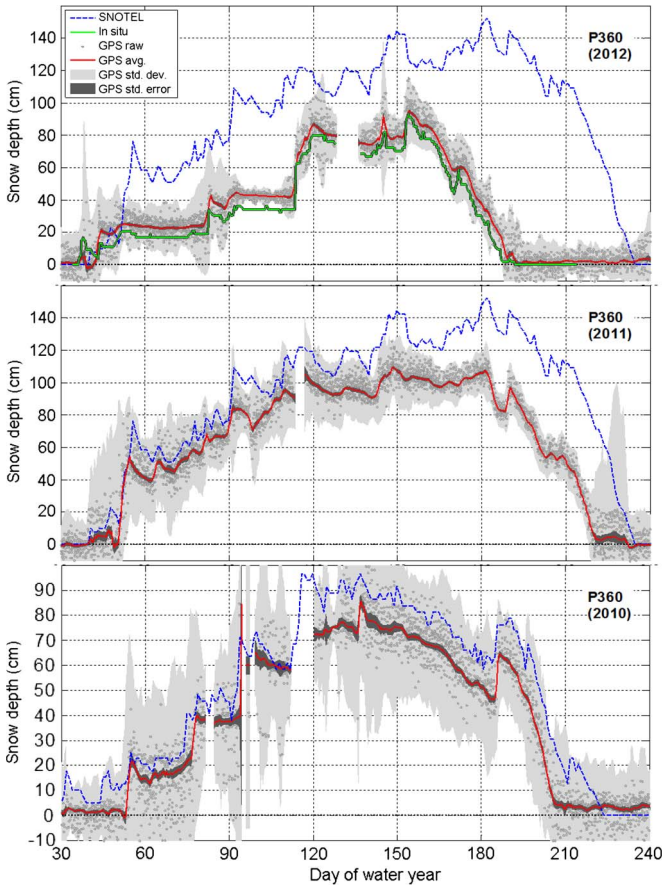

Fig. 8. Snow depth measurement at the grassland site (P360) for three water-years; the SNOTEL site is $\sim 12 \pm 4$ km distant and 60-m higher in altitude.

visibility to the sky is also excellent. The ground is mostly flat, with topography deviating no more than 3.5 m above and below the mean horizontal plane within a 100-m radius of the antenna. Topographical slopes range from 0.7° to 1.7° (at the 2-m spatial scale), with an average of $\sim 1°$ within 50-m radius around the GPS antenna. At the submeter scale, the terrain is rugged, with exposed rocks and littered with loose cobbles. Land cover classification is grasslands. There is a watercourse 200-m away due NE-W.

At P360, we show three years of estimated snow depth. Collection started at the time when L2C tracking was enabled in the receiver. Throughout this period, there are sonic snow depth measurements available from a SNOTEL station. However, this SNOTEL site is 60 m higher in altitude and at a distance of 12 ± 4 km. In the third year, there are up to four times daily *in situ* validation data from a pole colocated with the GPS antenna; tick marks can be read from the photographs with a precision of 3 cm. Fig. 8 shows results separately for each water-year; in the second year, satellite PRN 25 had been launched, and for the third water-year, satellite PRN 01 was also launched, each new satellite adding potentially four new track clusters. This is not to say that all are equally good; the diagram in Fig. 5 includes circles that are proportional to the statistical weight of each cluster (i.e., larger circles indicate more important clusters).

Fig. 8 shows that, generally, the timing of snowfall events is comparable for the GPS and SNOTEL sensors, although the amount of snow is not. A salient temporal feature that is well captured by both GPS and SNOTEL is the sharp transition between accumulation and settling that happens when precipitation stops and snow depth starts to drop.

For water-year 2012, *in situ* measurements become available, and we find a markedly improved agreement in terms of absolute amounts of snow depth, in contrast to the SNOTEL comparison. Still, this type of *in situ* data is not expected to be exactly comparable with the GPS estimates. The pole data stem from one-time readings at a fixed location with no coverage area, so it lacks statistical replication necessary to quantify the variability of snow depth. Clearly, the pole readings are not estimates of the mean, and as such, they are not expected to fall within the GPS confidence bounds for the mean. Furthermore, its spatial footprint is orders of magnitude smaller than the GPS, so the GPS prediction bounds also do not apply.

The GPS/*in situ* bias that remains constant throughout time could be ascribed to an unchanging spatial trend in snow deposition, likely controlled by the underlying ground topography is this open field environment. The periods of improved GPS/*in situ* agreement could be explained based on the sample randomization offered by temporal variability, even if sampling takes place at a fixed location. During quiet periods, the GPS/*in situ* discrepancy exhibits serial autocorrelation, which complicates their regression analysis, as the frequent although nonreplicated *in situ* measurements are not statistically independent. To overcome this difficulty, we form monthly *in situ* averages. The GPS/*in situ* regression has intercept statistically significant (5.9 ± 1.5 cm), while the regression slope (0.959 ± 0.05 m/m) is not: The deviation from a one-to-one relationship (-0.04 m/m), albeit indicating a 4% underestimating on the part of GPS compared to *in situ*, is smaller than its 95% confidence interval (0.05 m/m).

During the snow-free period, we find that reflector height does not remain exactly zero. Variations occur mainly when the scattering medium is transitioning, from snow to slush/mud and eventually grass-covered soil. This issue is both a challenge and an opportunity. On the one hand, it poses the risk of being mistaken for snow depth events. In fact, the identification of the site-overall (i.e., noncluster specific) zero-level or bare-ground reflector height is perhaps the weakest link in the whole GPS processing chain of snow depth retrieval, as it relies on only a few data points—we compute it as the fifth percentile of site-average reflector heights over the snow-free period. On the other hand, such observations attest to the prospects of using GPS reflector heights for monitoring environmental targets other than snow, such as vegetation biomass [10]. If successful, the estimation of nonsnow targets would contribute to guaranteeing that snow depth remains nonnegative.

C. Alpine Site (NWOT)

We finally consider the data collected at the Niwot Ridge Long-term Ecological Research site in CO, USA. Topographical slopes range from 2° to 7° (at the 2-m spatial scale), with average of $\sim 5°$ within 50-m radius around the GPS antenna. At an altitude of 3500 m, it is located in a saddle-like mountaintop, in an alpine tundra environment. A 3-m tall continuously

Fig. 9. Ground conditions in the vicinity of the GPS antenna at the alpine site (NWOT).

operating GPS system was established there in 2009 (Fig. 9). Poles are staked at 50-m intervals, making up a 120-by-400 m Cartesian grid at which snow depth is measured manually using a snow sampling tube, approximately every two weeks; we use *in situ* data collected at the pole nearest to the GPS antenna (shown in Fig. 9). The ground at the present site is not as planar as in the previous two sites, but visibility to the sky is good, with no trees and only minor topographical obstructions, predominantly due east and west. Indeed, the diagram of track clusters in Fig. 5 shows a nearly uniform azimuthal distribution.

The nearest SNOTEL location is more than 4 km away and ∼380 m lower in altitude. As is typical for SNOTEL stations, this one is found among trees, whereas the GPS is above the tree line. Therefore, their comparison comes with caveats. Similar as that for the forested site, the comparison against colocated *in situ* data is more favorable. Results for the pole colocated with the GPS antenna are in good agreement for the first two years, less so in the third year (Fig. 10).

In the last year, the peak snow depth is much smaller than in previous years, and the performance of the GPS deteriorates. This is partially because the amount of snow, subject to redistribution by the wind, is not sufficient so as to fill in the ground depressions during most of the season. Therefore, the multiple GPS track clusters are not as comparable as in previous years, when the intercluster variability leveled out as the air/snow interface became more planar than the snow/ground interface. Furthermore, the spatial variability also affects the *in situ* data, as the single pole colocated with the GPS is not as representative of the area sampled by the GPS. The remaining poles are more than 50 m from the GPS, so their potential contribution is questionable under such low-snow conditions.

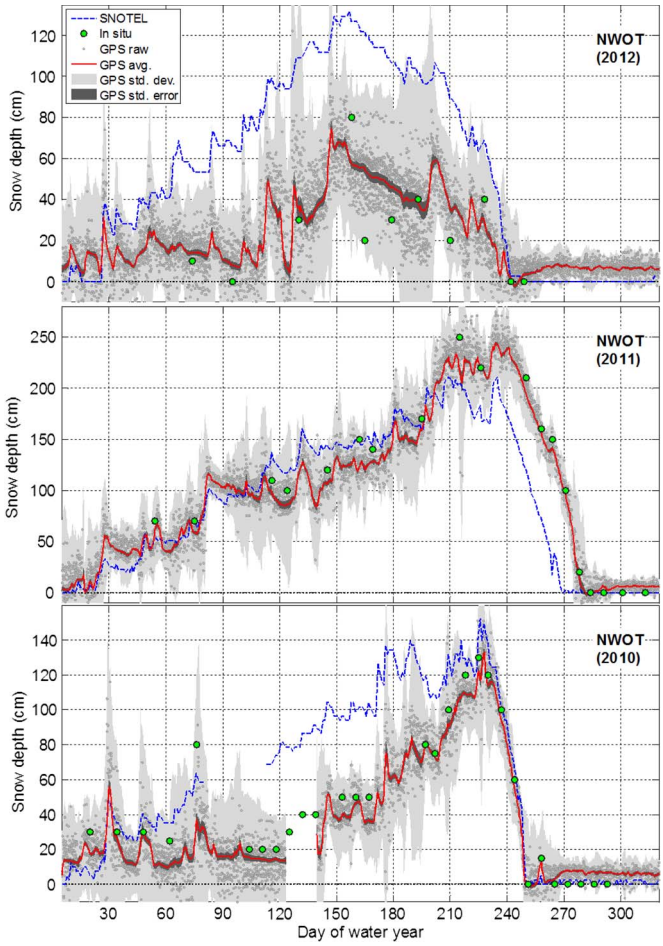

Fig. 10. Snow depth measurement at the alpine site (NWOT) for three water-years; the SNOTEL site is ∼4 km distant and ∼380-m lower in altitude.

The interannual variations in snow depth at the NWOT site are more drastic than at P360. Indeed, it exhibits a fivefold difference in peak snow depth, from ∼0.5 m in 2011–2012 to ∼2.5 m in 2010–2011. The timing of the end of the season varies by more than a month over this three-year period. The exact beginning of the season is less clear as the snow that accumulates from the initial precipitation events can be totally dissipated if the snowpack is not replenished with more frequent and vigorous snowfalls.

Fig. 11 shows a scatterplot of the GPS versus *in situ* snow depth for the three-year period at NWOT. The correlation is 0.980, which indicates a very strong linear relationship. Carrying out a regression between GPS and *in situ*, the rms of residuals improves from 10.7 to 7.8 cm. Notice that residuals are more dispersed at smaller snow depth values, especially when considered in proportion or relative to the *in situ* values (Fig. 11, bottom panel). The intercept 2.1 ± 2.8 cm is not statistically significant vis-à-vis its 95% confidence interval. The regression slope 0.89 ± 0.03 m/m indicates a small although statistically significant deviation from unity (−0.11 m/m), corroborating a similar finding first detected in the forested site, albeit now by a wider margin—four times smaller uncertainty, owing to the three times longer time series. Therefore, at this site, GPS estimates are ∼10% lower than *in situ* snow depth, although their footprints are not overlapping.

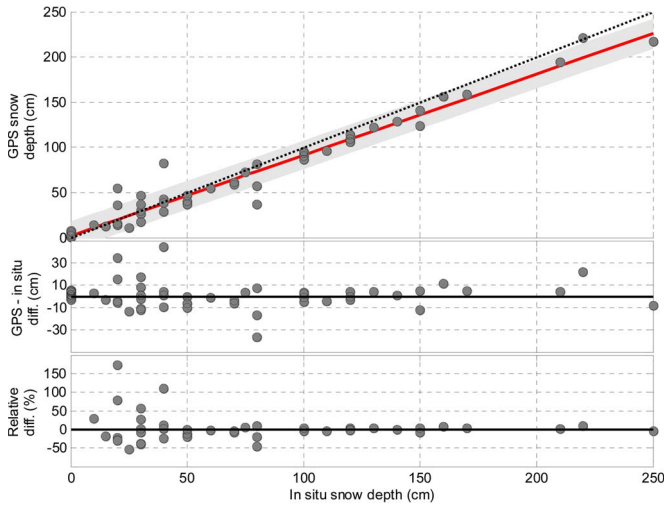

Fig. 11. Scatterplot of GPS versus *in situ* snow depth for all three years at the alpine site (NWOT). A simple linear regression is shown in red, with its 95% observation prediction interval shown as a light-gray band. Postfit residuals (with respect to the red line) are shown in the middle panel; the bottom panel shows residuals normalized by the *in situ* snow depth value. The ideal 1:1 diagonal is shown as a dotted black line for comparison.

TABLE IV
REGRESSION COEFFICIENTS, WITH *IN SITU* SNOW DEPTH TAKEN AS THE EXPLANATORY OR INDEPENDENT VARIABLE AND GPS SNOW DEPTH AS THE RESPONSE OR DEPENDENT VARIABLE

Station	Intercept (cm)	Slope (m/m)
P360	5.9 ± 1.5	1−0.04 ± 0.05
RN86	15.4 ± 9.1	1−0.14 ± 0.09
NWOT	2.1 ± 2.8	1−0.11 ± 0.03

At NWOT, we had extra *in situ* measurements, so we checked whether the GPS/*in situ* snow depth discrepancies depend on snow density and found a negligible correlation: coefficient value of −0.18. We also tried augmenting the regression, using *in situ* density in conjunction with *in situ* depth, thus estimating a total of three regression coefficients: a GPS/*in situ* additive bias (in meters), a GPS depth versus *in situ* depth multiplicative bias (in meters per meter), and a GPS depth versus *in situ* density linear coefficient (in meters per kilogram cubic meter). Compared to the two-coefficient regression, the introduction of density left the multiplicative bias unaltered: changes smaller than 1% for coefficient as well as for its uncertainty.

IV. CONCLUSION

We have demonstrated a statistical inverse model for estimating snow depth based on GPS multipath present in SNR observations. The model performance was assessed against independent *in situ* measurements and found to validate the GPS estimates to within the limitations of both GPS and *in situ* measurement errors after the characterization of systematic errors. The assessment yields a correlation of 0.98 and an rms error of 6–8 cm for observed snow depths of up to 2.5 m, with the GPS underestimating *in situ* snow depth by ∼5%–15% (Table IV). This latter finding highlights the necessity to assess effects currently neglected as recommended in [2].

The fit of SNR observations described in Part I [2] provided parameter estimates and their covariance matrix, as well as observation residuals, for each satellite track. In this paper, we have analyzed the resulting parameters and residuals. We have examined a few representative fits, illustrating and discussing the origin of a variety of good and bad conditions, such as measurement noise, well- versus poorly determined reflector heights, instrument-related issues, etc. Then, we have discussed a methodology to quality control (QC) these estimates based on track clusters; the thousands of tracks retrieved in a year can thus be analyzed in terms of only 10-20 units. We have introduced a specially designed diagram as a convenient summary of the track clusters available in a site. Such a repeatable sensing configuration allowed us to compare tracks belonging to the same cluster with the purpose of detecting and rejecting anomalous conditions. This principle leads to a number of strategies for QC of results, which are needed for operational use of GPS snow sensing. Site-wide averages are then compiled after different clusters are homogenized. This entailed accounting for genuine azimuthal asymmetry in the distribution of snow around the antenna and also dealing with issues such as assigning statistical weights to varying quality track clusters.

The suitability of GPS SNR measurements for snow monitoring was found to be heavily influenced by the site conditions; this lends weight to the finding of [3] that clearance to the satellite line-of-sight as well as to the ground is a strong requirement for GPS-MR. Therefore, the quality of retrievals may vary enormously over different azimuths at the same site. Furthermore, for a high-quality track cluster, a few percent of the individual retrievals might be discarded for a variety of reasons, such as receiver failures and heavy snowfall. QC is therefore mandatory for operational exploitation of GPS-MR.

A multitest QC strategy—including goodness-of-fit, reflector height uncertainty, peak elevation angle, and statistical DOF—was found to work best, as no single test detected all outliers. Combining multiple track clusters further improved the precision of GPS retrievals, yielding site-wide averages that captured remarkably well the temporal dynamics of snow accumulation and ablation, including sudden changes associated with new snow. Statistically robust methods—e.g., median instead of the mean—were adequate in achieving a reasonable level of processing automation and dispense with frequent manual intervention. Continuity of the time series as new satellites were launched every year indicates no obvious satellite-dependent biases; this stability is paramount for future utilization of GPS-MR results in climate studies.

Turning attention to aspects that would require more care in the future, our treatment of the azimuthal asymmetry exhibited by snow depth was admittedly cursory, in the sense that we only tried to minimize its impact on the site-wide averages by making clusters more comparable. This treatment worked well when the amount of snow was enough to fill in the ground depressions. However, when the amount of snow was insufficient to make the air/snow surface more planar than the snow/ground surface (alpine site, 2011–2012), the treatment failed to improve the dispersion around the site average.

Finally, further investigation is needed for the definition of the bare soil reflector height. The challenge is that, although

we can measure reflector heights precisely, we cannot unambiguously attribute an individual track estimate or even a site-wide average to a specific target, i.e., to distinguish between snow versus vegetation versus soil moisture changes manifested in reflector height. We have relied on the temporal dynamics of reflector heights along with reasonable assumptions about the snow behavior and optionally ancillary information (photographs, temperature records, climatic expectations, etc.) to determine the snow-covered period. This strategy worked very well for large amounts of snow, but it becomes less reliable for smaller amounts. As a rule of thumb, a 10-cm reflector height change would be a reasonable cutoff value for distinguishing snow, based on the behavior expected from other targets. Therefore, this issue is more serious for ephemeral snow sites, but it remains relevant for all sites.

ACKNOWLEDGMENT

The authors would like to thank J. Carlisle from Utah State University, Logan, UT, USA, for collecting field data at RN86, M. Williams from the University of Colorado Boulder, Boulder, CO, USA, for providing field and logistical support for NWOT through the NSF-funded Niwot LTER project, J. Normandeau from UNAVCO for providing additional oversight for the installation of the validation pole and camera at P360, PBO for providing the site photographs, and Google Earth for the satellite images. The receiver used at NWOT was leant to the project by Trimble Navigation, with engineering and archiving support provided by UNAVO. The SNOTEL data shown in this paper were retrieved from http://www.wcc.nrcs.usda.gov/nwcc/.

REFERENCES

[1] K. M. Larson, E. D. Gutmann, V. U. Zavorotny, J. J. Braun, M. W. Williams, and F. G. Nievinski, "Can we measure snow depth with GPS receivers?" *Geophys. Res. Lett.*, vol. 36, no. 17, pp. L17502-1–L17502-5, Sep. 2009.

[2] F. G. Nievinski and K. M. Larson, "Inverse modeling of GPS multipath for snow depth estimation—Part I: Formulation and simulations," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6555–6563, Oct. 2014.

[3] K. M. Larson and F. G. Nievinski, "GPS snow sensing: Results from the EarthScope Plate Boundary Observatory," *GPS Solut.*, vol. 17, no. 1, pp. 41–52, Jan. 2013.

[4] D. C. Agnew and K. M. Larson, "Finding the repeat times of the GPS constellation," *GPS Solut.*, vol. 11, no. 1, pp. 71–76, Jan. 2007.

[5] R. D. Fontana, W. Cheung, P. M. Novak, and T. A. Stansell, "The new L2 civil signal," in *Proc. ION GPS*, 2001, pp. 617–631.

[6] M. C. Serreze, M. P. Clark, R. L. Armstrong, D. A. McGinnis, and R. S. Pulwarty, "Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data," *Water Resour. Res.*, vol. 35, no. 7, pp. 2145–2160, Jul. 1999.

[7] D. M. Gray and D. H. Male, *Handbook of Snow—Principles, Processes, Management and Use*. New York, NY, USA: Pergamon, 1981, p. 776.

[8] W. Veatch, P. D. Brooks, J. R. Gustafson, and N. P. Molotch, "Quantifying the effects of forest canopy cover on net snow accumulation at a continental, mid-latitude site," *Ecohydrology*, vol. 2, no. 2, pp. 115–128, Jun. 2009.

[9] J. I. López-Moreno and J. Latron, "Influence of canopy density on snow distribution in a temperate mountain range," *Hydrol. Process.*, vol. 22, no. 1, pp. 117–126, Jan. 2008.

[10] E. E. Small, K. M. Larson, and J. J. Braun, "Sensing vegetation growth with reflected GPS signals," *Geophys. Res. Lett.*, vol. 37, no. 12, pp. L12401-1–L12401-5, Jun. 2010.

Felipe G. Nievinski received the B.E. degree in geomatics from Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre-RS, Brazil, in 2005, the M.Sc.E. degree in geodesy from the University of New Brunswick, Fredericton, NB, Canada, in 2009, and the Ph.D. degree in aerospace engineering sciences from the University of Colorado Boulder, Boulder, CO, USA, in 2013.

He is a Postdoctoral Researcher with Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), Presidente Prudente, Brazil, where he works in the field of GPS multipath reflectometry.

Kristine M. Larson received the B.A. degree in engineering sciences from Harvard University, Cambridge, MA, USA, in 1985 and the Ph.D. degree in geophysics from the Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA, USA, in 1990.

She is a Professor of aerospace engineering sciences with the University of Colorado Boulder, Boulder, CO, USA. Her current research focuses on GPS reflections.