Ice sheet topography from a new CryoSat-2 SARIn data processing, and assessment by comparison to ICESat-2 over Antarctica

Supplementary materials

Part A: CLS level-2 processing chain description

As illustrated in the flow chart Fig. 1, the CLS level-2 processing chain includes three main operating steps: leading edge detection, retracking and interferometric phase resolution.

S1: Leading edge detection

An algorithm inspired from the Canny edge detector (Canny, 1986) was developed to identify the leading edges considered significant in the waveform. The algorithm was initially developed for 2D images, here it was adjusted here to make it suitable with altimetry acquisitions. Waveforms are processed separately as independent measurements. The leading edge detection algorithm includes the following steps:

- Noise reduction:

The SARIn waveform is normalized by its maximum energy, referenced as WF_{norm} in the following text. Speckle noise is smoothed with a Savitzky-Golay filter applied to the SARInM waveforms (Savitzky and Golay, 1964). The code source is from the python scipy library. The objective is to achieve a slight waveform smoothing, with limited distortions of the signal dynamics.

- Gradient calculation:

A waveform gradient is computed as follows:

$$WF_{grad}\,[n] \;=\; \tfrac{1}{2}\,(mean\,(WF_{\;smoothed}\,[n+dn]) \;-\; mean\,(WF_{\;smoothed}\,[n-\;dn]) \qquad [1]$$

With "n" representing the waveform simple index, and "dn" representing a window size with the aim of reducing the gradient sensitivity to speckle noise.f

- Gradient magnitude thresholding

Waveform samples with a gradient value higher than H_{thres} are kept and grouped together if they are adjacent. This provides a preliminary peak detection. Several peaks can be detected in a single waveform.

- Leading edge boundaries definition with double thresholding:

A second lower threshold is defined: $L_{thres.}$ The samples with a gradient value between L_{thres} & H_{thres} and connected/adjacent to the pre-defined peaks are integrated to these peaks. For each detected peak, the leading edge boundaries are defined as follows:

- Leading edge first sample "n_s" corresponds to the first sample of the peak
- Leading edge last sample "n_e" corresponds to the sample where waveform energy is maximum

- Leading edge quality control:

The leading edge is considered valid if its amplitude (LE_{amp}) and length (LE_{len}) are within expected range values, based on input thresholds (see below). With:

LE_{amp} = WF_{norm} [n_e] - min(WF_{norm} [n_s … n_e])

LE_{len} = n_e - n_s

Additional Comments:

The leading edge detection is controlled by 6 input parameters, set empirically:

- The window length used in the Savitzky-Golay filter *(set to **9** samples)*
- The polynomial order used in the Savitzky-Golay filter *(set to **3**)*
- The window length "dn" used in the gradient computation *(set to **5** samples)*
- The high threshold used in the peak detection "H_{thres}". The low threshold "L_{thres}" being equal to $\frac{Hthres}{4}$ *(H_{thres} set to **0.33**).*
- The threshold used to check the leading edge amplitude (*LE_{amp} must be higher than **0.33**)*
- The thresholds used to check the leading edge length (*LE_{len} must be between **5** and **50** samples*)

S2: Waveform retracking

The objective of the waveform retracking is to derive the altimeter range and interferometric phase difference from the waveform, coherence and phase difference signals.

- Phase and coherence vectors are smoothed with a running average window. The window size is set to 9 samples. For the phase vector, two operations are performed during the smoothing:
  - 2π jumps occurring within the average window are checked and corrected if necessary.
  - The phase noise level within the average window is controlled: if the STD of the phase values included in the window is above 0.5 radians, then no smoothed phase is recorded for the considered sample.

- Leading Edge maximum Coherence (LMC) retracker: Starting from the leading edge samples WF_{norm} [n_s…n_e], a subset of samples is defined, where WF_{norm} [n_s…n_e] > 0.5 * LE_{amp} + LE_{min_energy}, with LE_{amp} the amplitude of the leading edge as previously defined; LE_{min_energy} the minimum normalized energy of the waveform leading edge. Among this subset, the retracking point n_{rtk} is the sample with the highest smoothed coherence.

- Altimeter range R_n is derived from the time delay computed at retracking point n_{rtk} and corrected from several geophysical effects: ionosphere and wet/dry troposphere delays, and effects due to solid Earth tide, pole tide and ocean loading tide

- Interferometric phase value is extracted from the smoothed signal at retracking point n_{rtk}

S3: Interferometric phase resolution

From the altimeter range and the interferometric phase difference, the measurement is geolocated and potential interferometric phase ambiguities are resolved:

- Look angle computation following Jensen (1999) and Gray et al. (2013) formulations. The look angle being the angle between the satellite-nadir vector and satellite-estimated POCA vector (in case of linear surface this angle is equivalent to the across-track slope):
  $\theta = \sin^{-1}\left(\frac{\lambda\ \Delta\varphi}{2\pi\ B}\right) - \gamma$

with λ the radar wavelength (~0.022084 m); B the antenna baseline (=1.1676 m); γ the platform roll angle available in the L1B product; $\Delta\varphi$ the interferometric phase difference at retracking point

- Satellite coordinates are shifted from lat/lon/altitude to Earth CEntered Frame (x_{sat},y_{sat},z_{sat}) coordinates. Determination of the along-track unit vector in the ECEF frame using the next measurement coordinates. From the along-track unit vector, the across-track unit vector can be derived.

- Measurement geolocation (x_{poca},y_{poca},z_{poca}) in the across-track plan with geometrical computations in the ECEF frame involving the altimeter range and the satellite look angle. Estimated POCA is shifted back to lat/lon/altitude coordinates.

- Phase ambiguities control: In the CryoSat-2 configuration phase ambiguities arise when the look angle exceeds ~0.54°. In such situation the POCA is displaced ~6,85 km away from nadir. To check for potential phase ambiguities, two controls are performed:
  - The theoretical across-track POCA location is derived from the reference DEM. If it is located more than 5 km away from nadir (conservative value), a flag ambiguity warning is raised.
  - If more than one peak is detected on the waveform, it most likely means that the measure is acquired over an irregular topography, with scatterers from different across-track locations contributing to the signal. In a conservative approach, a flag ambiguity warning is also raised to ensure there is no uncertainty on the actual POCA location.

- In case of ambiguity warning: Other SARIn geolocation solutions are computed by adding and subtracting 2π to the phase difference value. The elevation solution the closest to the DEM is kept, except if solutions close to each other are detected, separated by less than 15 m in elevation. In this case we consider there is a risk of wrong relocation, hence the measurement is discarded.

- Outliers rejection: A theoretical elevation is computed at POCA from the reference DEM, using a bi-linear interpolation on the DEM grid. If the bias with the SARIn estimation exceeds 50 m the measurement is discarded.

S4: Alternative waveform retracking algorithms, for analyses section 3.5

In the analyses presented section 3.5, the Leading Edge maximum Coherence (LMC) retracker previously described (A2) is substituted in the processing chain to a threshold and a gradient retrackers, following these specifications:

- **Threshold retracker:**
  - Identification of the location n_{th} where the waveform energy exceeds a Threshold Level (TL), between the first and last samples of the first detected leading edge (n_s & n_e). With $n_{th} = LE_{amp} \times TL + LE_{min_energy}$. Three thresholds have been tested : 0.25, 0.5 and 0.75.
  - The retracking point $\mathbf{n_{rtk}}$ is linearly interpolated between the adjacent samples to n_{th} position:

$$\boldsymbol{n_{rtk}} = n_s + \frac{TL - WF[n_{th}-1]}{WF[n_{th}] - WF[n_{th}-1]} \quad (1)$$

- **Gradient retracker:**
  - The gradient of the normalized and smoothed waveform WF_{norm} is derived from equation (1), with dn=1.
  - The retracking point $\mathbf{n}_{rtk}$ is located at the sample where the gradient is maximum between the first and last samples of the first detected leading edge (n_s & n_e).

S5: ESA PDGS level-2 flags used to discard bad quality data

The field "*flag_prod_status_20_ku*" contained in the ESA PDGS level-2 products includes different data flags. These data flags are used in the processing chain to discard bad quality measurements before the processing. In details the following flags are tested:

- *"block_degraded"*
- *"orbit_error"*
- *"sarin_rx_1_error"*
- *"sarin_rx_2_error"*
- *"mispointing_error"*
- *"delta_time_error"*
- *"sarin_bad_baseline"*
- *"sarin_bad_velocity"*
- *"sarin_side_redundant"*
- *"calibration_warning"*

Part B: Additional statistics relative to the Antarctica study (section 3)

The following numbers are the total amount of CryoSat-2 measurements analysed in the analysis presented section 3. They are acquired by the satellite during a full cycle: 369 days of acquisitions over the Antarctic ice sheet at 20 Hz rate (March 2018 to March 2019)

Total amount of CS-2 data:	19 309 429 measurements
After first selection:	18 528 779 measurements (95.96%)
After second selection (CLS data):	17 117 765 measurements (88.65%)
After second selection (PDGS data):	13 739 648 measurements (71.16%)

Note: *The percentage under bracket indicates the ratio of measurements relative to the total amount of CryoSat-2 20 Hz measurements*

Quality controls, first common selection

	First common selection to CLS & PDGS data	
Quality checks	**quality flags error**	**Low SNR**
% of discarded data	**0.00007 %**	**4.04 %**

Quality controls, second selection for each processor

Second selection specific to CLS data

Quality checks	No detected peak	Noisy phase	Geolocation uncertainty	Coherence < 0.7	Bias to DEM > 50m
% of discarded data	0.1 %	0.38 %	2.15 %	5.03 %	0.6 %

Second selection specific to PDGS data

Quality checks	Ambiguity warning flag	Elevation error flag	Coherence < 0.7	Bias to DEM > 50m
% of discarded data	12.58 %	0.33 %	14.77 %	4.71 %

Tables S1: Quality checks and selection applied to the CryoSat-2 SARInM elevations over the Antarctic ice sheet data set. In the first table the percentages are relative to the total amount of measurements. In the second table the percentages are relative to the amount of measurements after first selection. One measurement can be rejected by several quality controls (corrolary: the sum of the indicated percentages does not represent the sum of the total measurements discarded).

CryoSat-2 and ICESat-2 elevation differences statistics over the Antarctic ice sheet

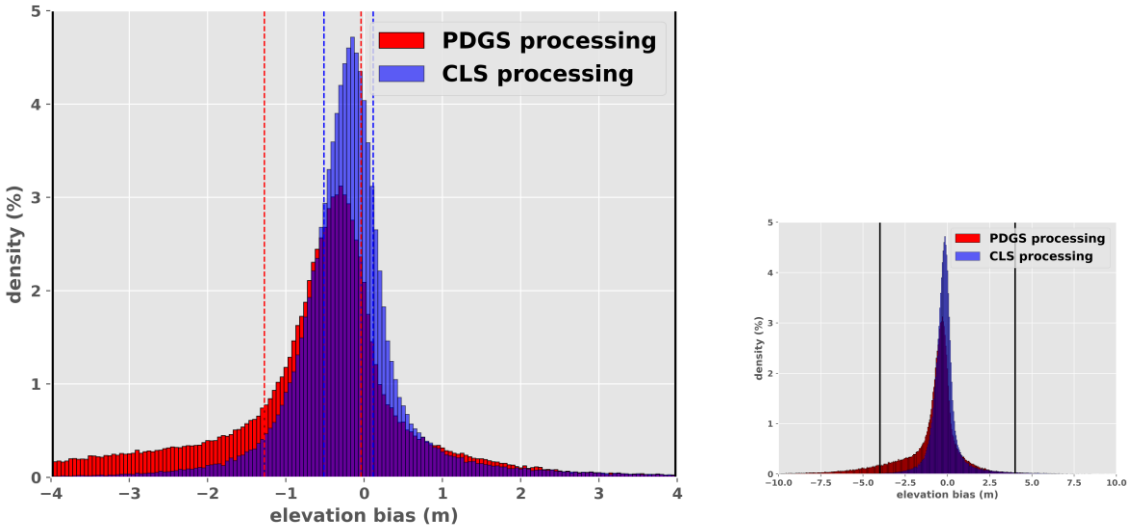


Figure S1: (right) Histogram of the elevation differences between CryoSat-2 and ICESat-2 space-time co-located measurements. CryoSat-2 SARIn mode elevations are derived from the PDGS level-2 products (red) and from the level-2 CLS processor (blue). Dotted vertical lines display the value of the first and third interquartile ranges for each data distribution. (left) same histogram with the elevation bias (x axis) extended to +/- 10 m.

	Number of co-located point	Median bias (cm)	Median Absolute Deviation (cm)	InterQuartile Range (cm)
CLS processor	249 823	-17.8	46.5	63
PDGS processor	209 400	-47.3	80.4	123.4

Table S2: Statistics on the elevation differences between CryoSat-2 and ICESat-2 space-time co-located measurements.

	% of points with |bias| > 1 m	% of points with |bias| > 2 m	% of points with |bias| > 5 m	% of points with |bias| > 10 m	% of points with |bias| > 25 m
CLS processor	15.86	6.25	1.70	0.51	0.08
PDGS processor	37.73	21.26	5.47	0.91	0.14

Table S3: Population of elevation differences outliers between CryoSat-2 and ICESat-2 space-time co-located measurements, at different elevation thresholds.

	After iterative 3-Sigma filter							
	Number of co-located pt	Median bias (cm)	Mean bias (cm)	STD (cm)	RMSE (cm)	MAD (cm)	IQR (cm)	% filtered data
CLS processor	226 379	-18	-20.8	48.1	52.5	41	55.8	9.38
PDGS processor	185 221	-41.3	-53	103	115.9	65.3	91.8	11.55

Table S4: Statistics on the elevation differences between CryoSat-2 and ICESat-2 space-time co-located measurements, after iterative 3-Sigma filtering.

CryoSat-2 and ICESat-2 elevation difference statistics as a function of the ground distance between nadir & CryoSat-2 SARIn assumed POCA

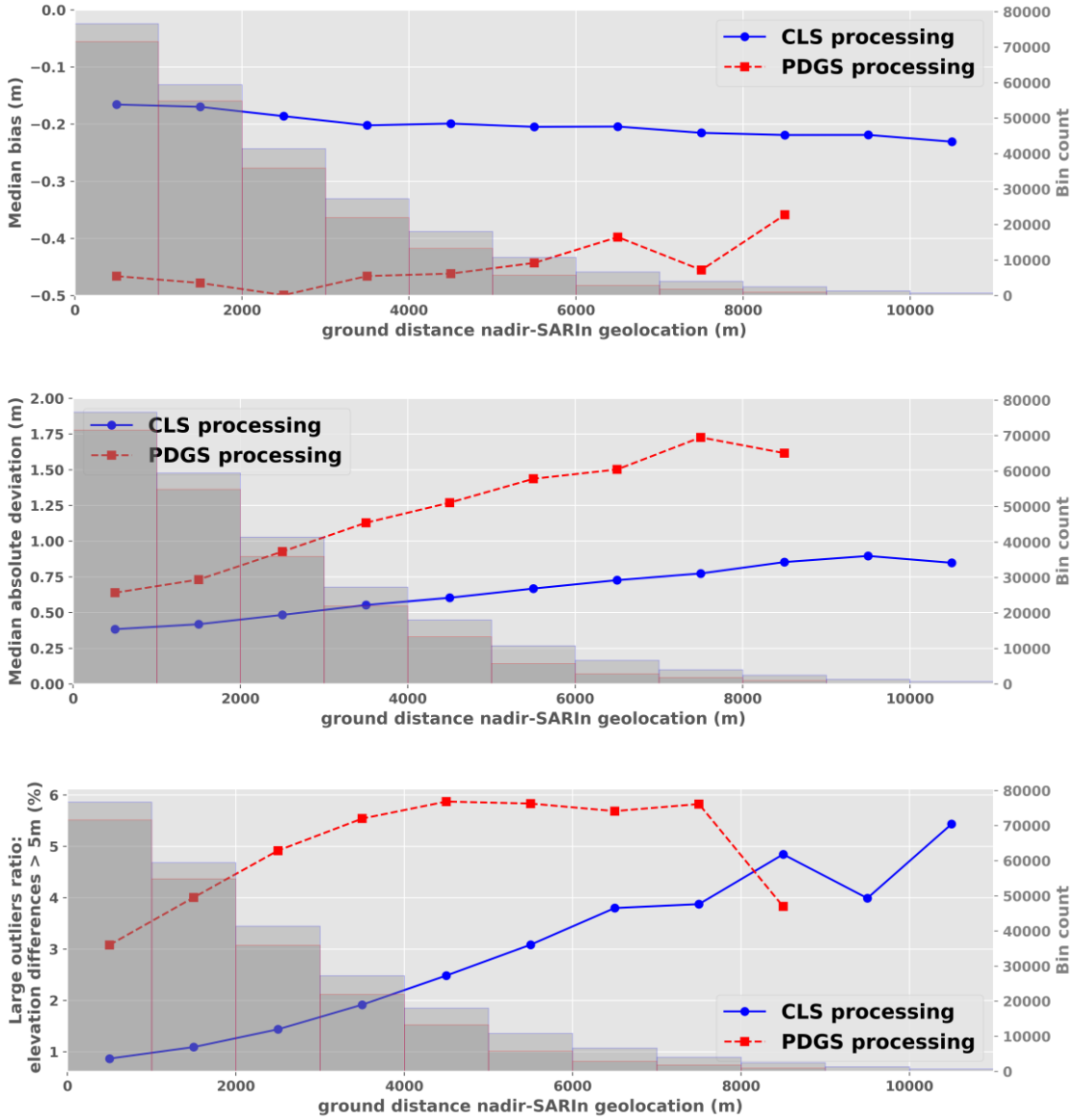


Figure S2: Median difference (top), median absolute deviation (middle) and outliers ratio (bottom) between CryoSat-2 SARIn mode and ICESat-2 ATL06 co-located elevations, as a function of the nadir-POCA distance. SARIn mode elevations are derived from the PDGS level-2 product (red) and from the level-2 CLS processor (blue). Elevations are computed as CryoSat-2 – ICESat-2. Gray bars display the number of co-located measurements for each surface slope interval.

CryoSat-2 and ICESat-2 elevation difference statistics, obtained with the CLS processor, with a threshold retracker instead of the coherence retracker

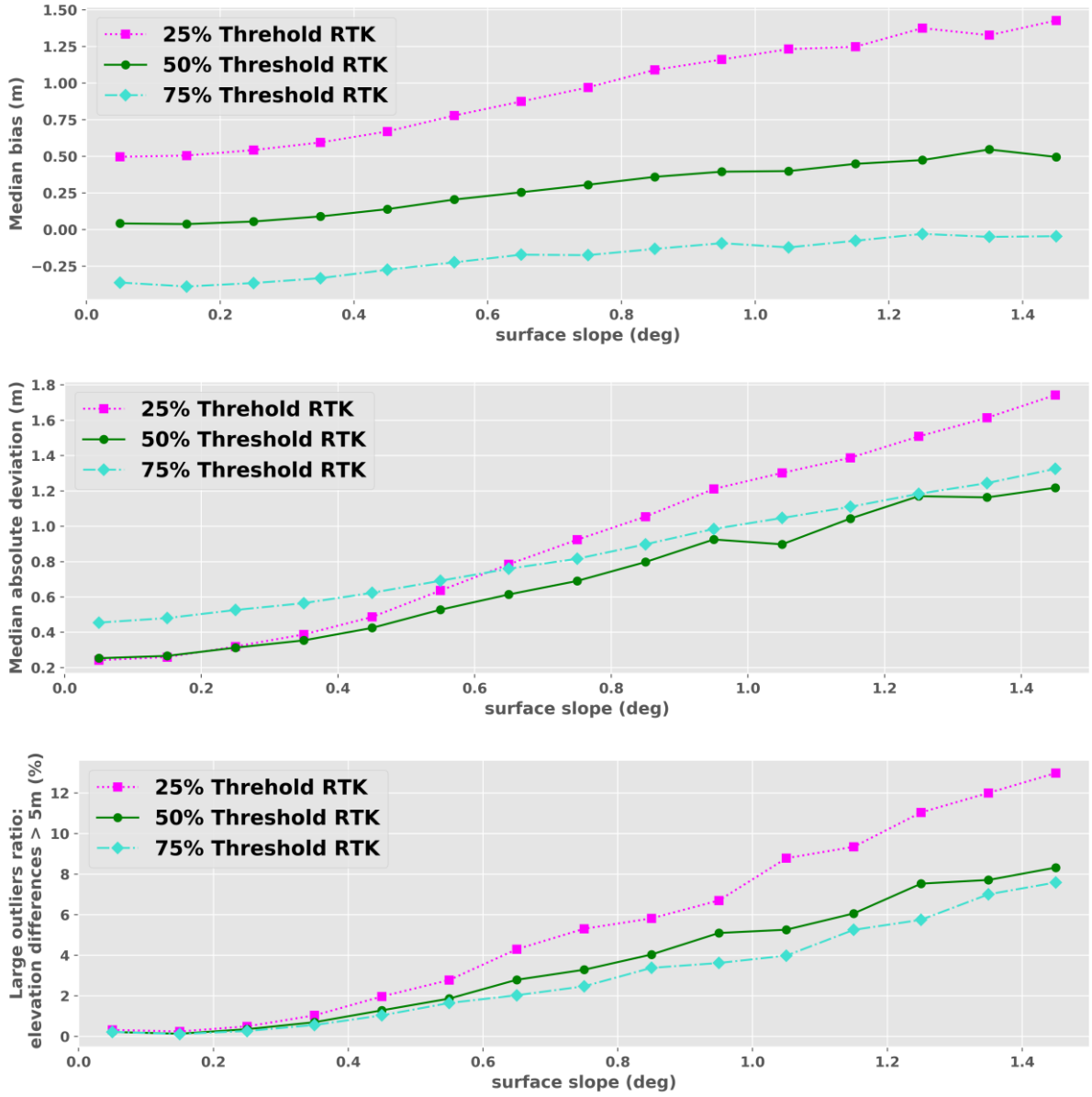


Figure S3: Median difference (top), median absolute deviation (middle) and outliers ratio (bottom) between CryoSat-2 SARIn mode and ICESat-2 ATL06 co-located elevations, as a function of the surface slope. SARIn mode elevations are derived from the CLS processor with a threshold retracker operating with 25 % (magenta), 50 % (green) and 75 % (turquoise) threshold values. Elevations are computed as CryoSat-2 – ICESat-2.

CryoSat-2 and ICESat-2: gridded statistics of the elevation differences, obtained with the 3 retrackers implemented in the CLS processor and analysed section 3.5

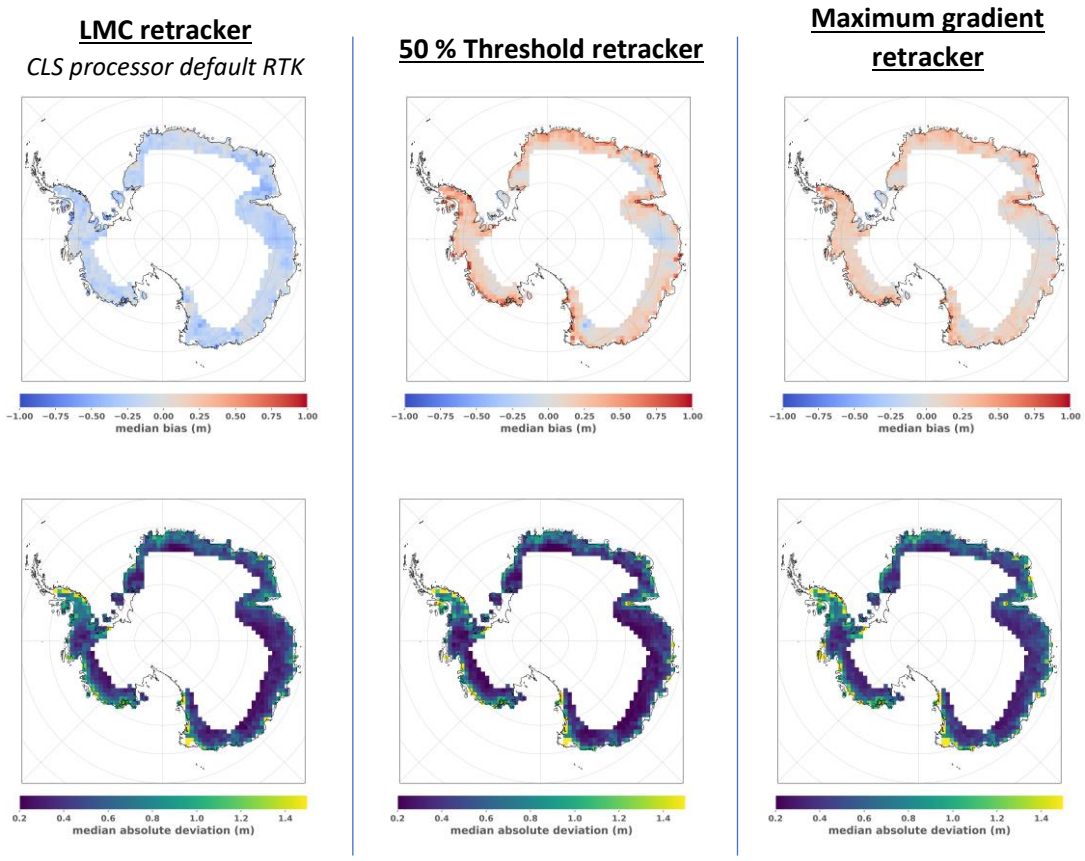


Figure S4: Gridded statistics of the elevation differences between CryoSat-2 SARIn and ICESat-2 ATL06 co-located elevations: median difference (top), median absolute deviation (bottom). The statics are obtained with the CLS processor with the Leading edge Maximum Coherence retracker (left), a 50 % threshold retracker (center) and a maximum gradient retracker (right). Elevations are computed as CryoSat-2 – ICESat-2. Grid resolution is 80 km.

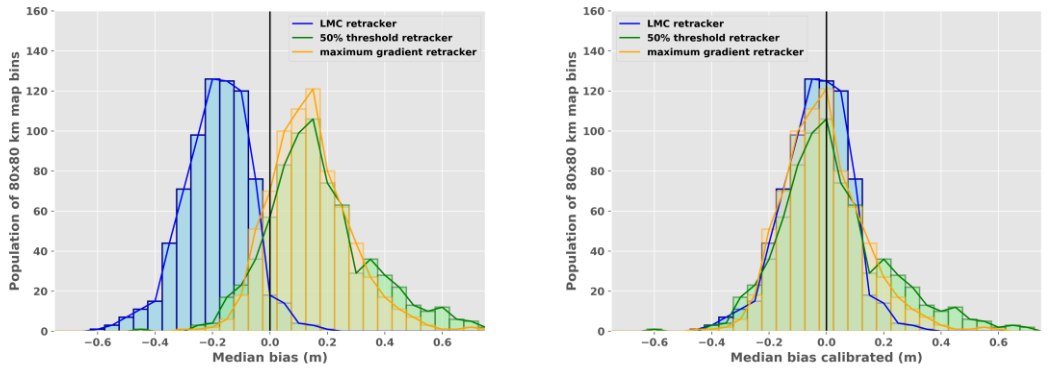


Figure S5: (left) Histograms of the 80x80 km gridded bins displayed in the top panels of Fig. B4, and representing the median elevation difference between CryoSat-2 SARIn and ICESat-2 co-located elevations (right) same histograms but shifted to "0" to be better compared.