Sentinel-3 Performance improvement for ICE sheets

Product Validation Report

Scientific Exploitation of Operational Missions (SEOM)

Sentinel-3 SAR Altimetry

Study 4: Ice Sheets

Prepared by:	Malcolm McMillan	date: 1/4/2019
Approved by:	Andrew Shepherd	date: 1/4/2019
Accepted by ESA:	Jérôme Benveniste	date:

Distribution List

Jerome Benveniste	ESA
Marco Restano	ESA
Americo Ambrozio	ESA
Malcolm McMillan	University of Leeds
Andrew Shepherd	University of Leeds
Mònica Roca	isardSAT
Maria Jose Escorihuela	isardSAT
Roger Escola	isardSAT
Pierre Thibaut	CLS
Frédérique Rémy	LEGOS

Document Status:

Internal

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

AD	Applicable Documents
AIS	Antarctic Ice Sheet
AR	Analytical Retracker
ATM	Airborne Topographic Mapper
CLS	Collecte Localisation Satellites
CryoSat-2	CryoSat-2 satellite
DEM	Digital Elevation Model
ESA	European Space Agency
FBR	Full Bit Rate
GIS	Greenland Ice Sheet
FFT	Fast Fourier Transform
GLAS	Geoscience Laser Altimeter System
ITT	Invitation To Tender
KO	Kick off meeting
LEGOS	Laboratoire d'Etudes en Géophysique et Océanographie Spatiales
LiDAR	Light Detection And Ranging
LRM	Low Resolution Mode
NASA	National Aeronautics and Space Administration
NSIDC	National Snow and Ice Data Center
pLRM	Pseudo-LRM
pSAR	Pseudo-SAR
PVP	Product Validation Plan
PVR	Product Validation Report
RB	Requirements Baseline document
SAR	Synthetic Aperture Radar

SARin	Synthetic Aperture Radar interferometric
SEOM	Scientific Exploitation of Operational Missions
SoW	Statement Of Work
SPICE	Sentinel-3 Performance improvement for Ice sheets
SPOT	Satellite Pour l'Observation de la Terre
TPR	Threshold Peak Retracker
UL	University of Leeds
WP	Work Package

Applicable Documents

AD1	Scientific Exploitation of Operational Missions (SEOM). Sentinel-3 SAR Altimetry Statement of Work (SEOM S3-4SCI SAR Altimetry). Issue 1, 27/09/2014.
AD2	Special Conditions of Tender. Appendix 4 to AO/1-8080/14/I-BG.
AD3	SPICE Technical Proposal.
AD4	SPICE Implementation Proposal.
AD5	SPICE Requirements Baseline document, version 1, issued 19/11/2015.
AD6	CryoVal Land-Ice project. CryoVal Land Ice Technical Note. CS2-MSSL-RTEL-0001, October 2014.
AD7	SPICE Algorithm Theoretical Basis Document, version 1, issued 12/2/2018.

1. Introduction

Purpose

This document is the Project Validation Report (PVR) for the Sentinel-3 Performance improvement for ICE sheets (SPICE) proposal (AD3), which is a response to the European Space Agencies (ESA's) Sentinel 3 For Science – SAR Altimetry Studies (S3 4 SCI – SAR Altimetry Studies) Invitation To Tender (ITT), Ref. AO/1-8080/I-BG (AD1 and AD2). SPICE addresses the Study 4 theme related to Ice Sheets. The Project Validation Report has been written by the University of Leeds (UL), with contributions from isardSAT. UL as the prime contractor is the contact point for all communications regarding this document.

Address:

School of Earth and Environment,

Maths/Earth and Environment Building,

The University of Leeds,

Leeds,

LS2 9JT,

UK

Att: Malcolm McMillan (Science Lead)

Email: m.mcmillan@leeds.ac.uk

Telephone: + 44(0) 113 34 39085

Fax: +44 113 343 5259

ESA Bidder Code: 6000012896

Project Validation Report structure

The aim of the Project Validation Report is to document the results of the SPICE study validation activities. These have been carried out within the scope of WP4, whose objectives were to evaluate the generated datasets and methodologies. The remainder of the PVR document is structured into the following sections:

- **Section 2** – Description of validation sites.
- **Section 3** – Description of validation datasets.
- **Section 4** – Review of validation methods.
- **Section 5** – Results of validation activities.
- **Section 6** – Conclusions
- **Section** 7 – References.

2. Description of validation sites

This section outlines the principle validation sites used for the SPICE project, and reviews the scientific and technical justification for these choices. The principle study sites selected for this project were the Lake Vostok, Dome C and Spirit and sites in East Antarctica (Figure 1), and the Russell Glacier site on Greenland's western margin (Figure 2). The choice of these sites, as set out in the PVP, was governed by the following criteria:

(1) The availability of CryoSat-2 SAR acquisitions over land ice.
(2) The availability of high quality reference data for product validation.
(3) The study requirements to include sites in both Greenland and Antarctica.

The three Antarctic study sites were chosen because they were the focus of dedicated CryoSat-2 SAR campaigns in 2014 and were, additionally, covered by validation datasets. They therefore represented the only land ice SAR acquisitions made prior to the launch of Sentinel-3 where suitable validation data existed. In Greenland, no SAR data has been acquired by CryoSat-2 to date. Nonetheless, there remains a need to evaluate SAR performance over regions with different ice sheet characteristics, such as those found in the ablation zone of Greenland. We therefore selected a site within the SARIn mode mask, and as part of WP2 developed the capability to reprocess SARIn FBR via a single receive chain, to derive a pseudo-SAR (pSAR) product. For this analysis the Russell Glacier site in western Greenland was selected due to its proximity to Kangerlussuaq Airport, which has led to extensive surveying of this region by airborne campaigns and therefore offers a large amount of validation data. The periods of SAR data acquisition, where appropriate, are given in Table 1.

The four sites each offer specific characteristics that are beneficial to the aims of the SPICE study. Both Vostok and Dome C are located within the East Antarctic interior (Figure 1) and are characterised by relatively simple topography, low accumulation rates (Arthern et al., 2006) and an absence of surface melting. They therefore allow an evaluation of SPICE products in regions representative of a large part of the Antarctic interior region. The remaining Antarctic site, Spirit, lies in a region of steeper ice sheet topography. The dedicated SAR acquisitions across this site in 2014 provided the opportunity to assess the merits of different SAR processing methodologies in a region representative of the ice sheet margins. These areas are particularly important because they tend to exhibit the greatest changes in ice mass, yet their more complex terrain remains a challenge for conventional radar altimeters. The final site, Russell Glacier in Western Greenland, lies predominantly within the ablation zone of the ice sheet. It therefore experiences a range of atmospheric and snowpack conditions, including periods of surface melt and bare ice. As such it has enabled us to evaluate SAR performance in a region that is challenging for radar altimeters because of the significant changes in the scattering properties of the ice surface. Lying within the SARIn mode mask, and without dedicated SAR acquisitions, it has however been necessary to undertake the exploratory reprocessing of SARIn FBR to pLRM and pSAR L2 products.

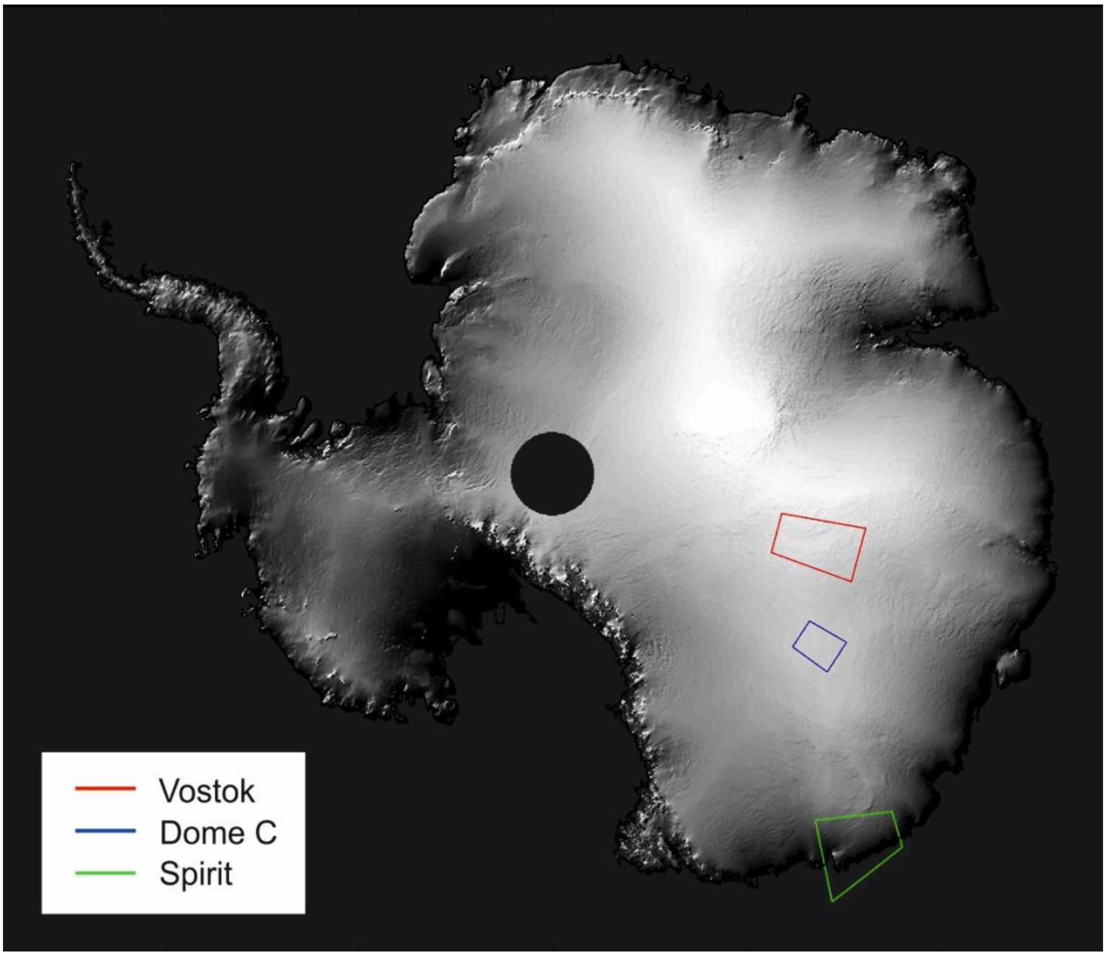


Figure 1. Antarctic validation sites.

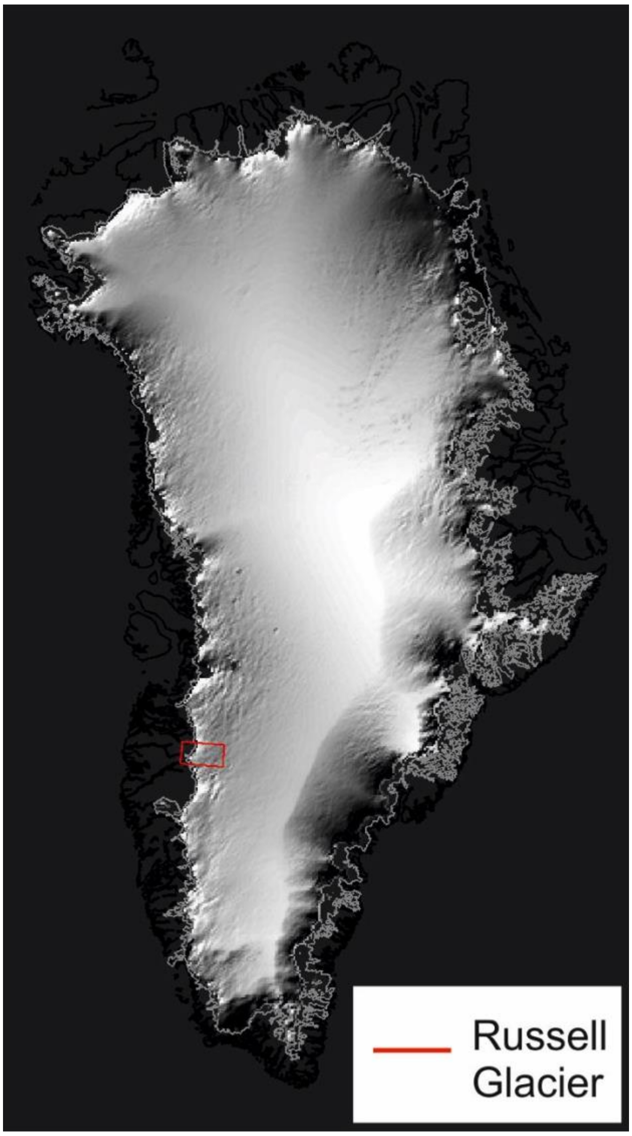


Figure 2. Greenland validation site.

Study Site	SAR acquisition period	Latitude bounds	Longitude bounds
Vostok	24/11/2014 – 30/11/2014	79-75°S	100-110°E
Dome C	1/12/2014 – 7/12/2014	76-74°S	120-126°E
Spirit	17/11/2014 – 30/11/2014	66-69.5°S	135-147°E
Russell Glacier	none	67-67.5°N	50.5-48°W

Table 1. The spatial extent of each study site and the acquisition period of the CryoSat-2 baseline-B SAR data that were used in this study. Note that no SAR data has been acquired at Russell Glacier and instead we have processed SARIn FBR to form a pseudo-SAR product.

3. Description of validation datasets

This section describes the auxiliary validation datasets that have been used to evaluate the products generated during the SPICE project. Specifically, these data are independent ice sheet elevation measurements compiled from airborne and satellite platforms. These data are described in more detail below and also summarised in Table 2.

The principle validation datasets are comprised of airborne surface elevation measurements acquired by the Airborne Topographic Mapper (ATM) and Riegl Laser Altimeter instruments flown on-board NASA's Operation IceBridge campaigns (http://nsidc.org/icebridge/portal/). These datasets provide the most comprehensive airborne coverage of the polar ice sheets since 2009, with the high accuracy, spatial resolution and precision achievable with an airborne laser altimeter. The coverage of these data across each of the SPICE study sites is shown in Figure 3 and in Figure 4. To supplement these airborne data, we have also used ICESat satellite laser altimetry as an additional source of reference, to improve coverage in regions where IceBridge acquisitions are sparse. A brief description of each dataset is given below.

The Airborne Topographic Mapper (ATM) is an airborne scanning LIDAR developed by NASA to map ice surface elevation in the polar regions. Since 2009, it has been one of the principal instruments carried by NASA's Operation IceBridge. Elevation measurements are resampled to approximately 50 m along-track (varying with aircraft speed) and have a fixed 80 m across-track platelet at aircraft nadir. At a nominal operating altitude (500 to 750 m above the ice surface) the ATM elevation measurements have been estimated to achieve a horizontal accuracy of 74 cm, a horizontal precision of 14 cm, a vertical accuracy of 7 cm and a vertical precision of 3 cm (Martin et al., 2012).

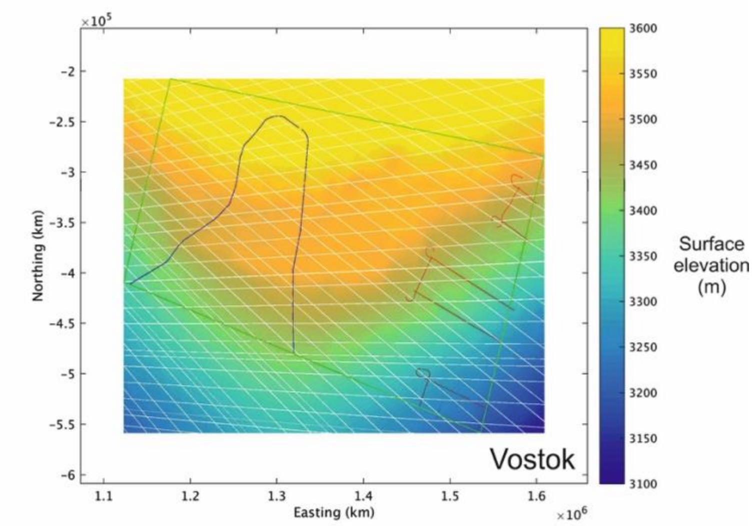
The Riegl Laser Altimeter is a Laser Altimeter System, also flown on selected Operation IceBridge campaigns in Antarctica. This instrument acquires elevation measurements with a range resolution of 2 mm and a ground footprint of 25 m along track by 1 meter across track. The reported error associated with these elevation measurements is 12cm (http://nsidc.org/icebridge/portal/).

The Geoscience Laser Altimeter System (GLAS) flown on-board the ICESat mission operated on a ~35-day campaign basis between 2003 and 2009, with approximately three campaigns acquired each year. The ground footprints are spaced at 172 m along-track and have a varying elliptical shape with average dimensions of approximately 50 x 95 m. GLAS has been shown to achieve a single shot elevation accuracy better than 0.05

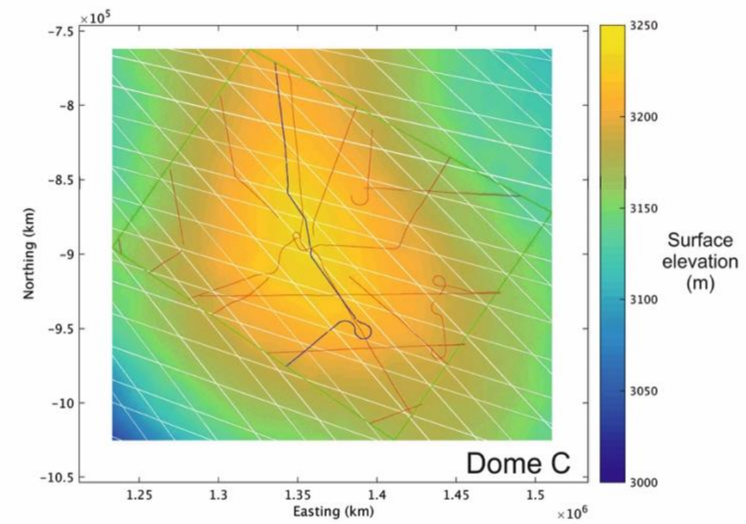
m under optimal conditions, although performance degrades over sloping terrain and under the presence of atmospheric forward scattering and detector saturation (Fricker et al., 2005). Data coverage is also adversely affected by the presence of clouds. The coverage achieved by ICESat at each of the SPICE study sites is shown in Figure 3 and in Figure 4.

Data Type	Location	Parameter	Acquisition Date (month/year)	Sensor	Data Provider	Availability Status
Airborne	Vostok	Ice surface elevation	11/2013	ATM	NASA, available online from nsidc.org	Archived at UL
Airborne	Vostok	Ice surface elevation	1/2009	Riegl	NASA, available online from nsidc.org	Archived at UL
Airborne	Dome C	Ice surface elevation	11/2013	ATM	NASA, available online from nsidc.org	Archived at UL
Airborne	Dome C	Ice surface elevation	1/2009 2/2009 12/2009 12/2011 12/2012	Riegl	NASA, available online from nsidc.org	Archived at UL
Airborne	Spirit	Ice surface elevation	1/2009 12/2009 1/2010 12/2010 11/2011 12/2011	Riegl	NASA, available online from nsidc.org	Archived at UL
Airborne	Russell	Ice surface elevation	4/2009 5/2009 5/2010 3/2011 4/2011 4/2012 5/2012 4/2013 4/2014 4/2015	ATM	NASA, available online from nsidc.org	Archived at UL
Satellite	All	Ice surface elevation	2009	ICESat	NASA	Archived at UL

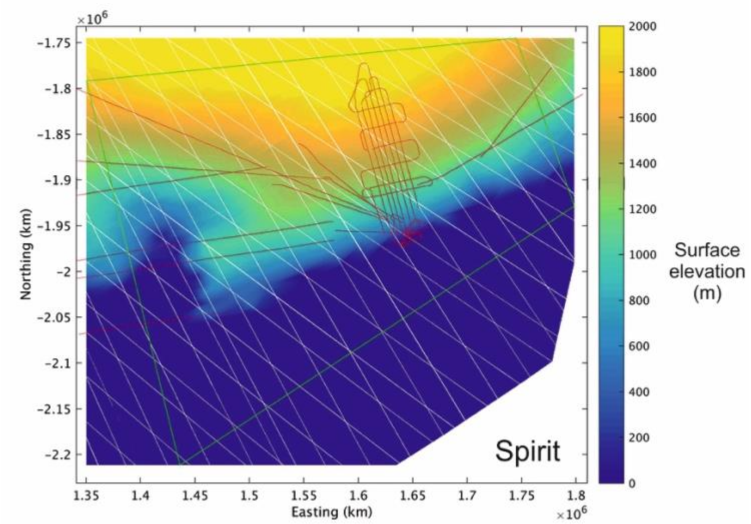
Table 2. Auxiliary data used within the SPICE study.



$\times 10^5$
Northing (km)
Easting (km)
$\times 10^6$
Vostok
Surface elevation (m)



$\times 10^5$
Northing (km)
Easting (km)
$\times 10^6$
Dome C
Surface elevation (m)



$\times 10^6$
Northing (km)
Easting (km)
$\times 10^6$
Spirit
Surface elevation (m)

Figure 3. Operation IceBridge ATM (blue) and Riegl (red) airborne laser altimetry flightlines, and ICESat (white) ground tracks over the SPICE Antarctic study sites. The green polygon marks the boundary of each site.

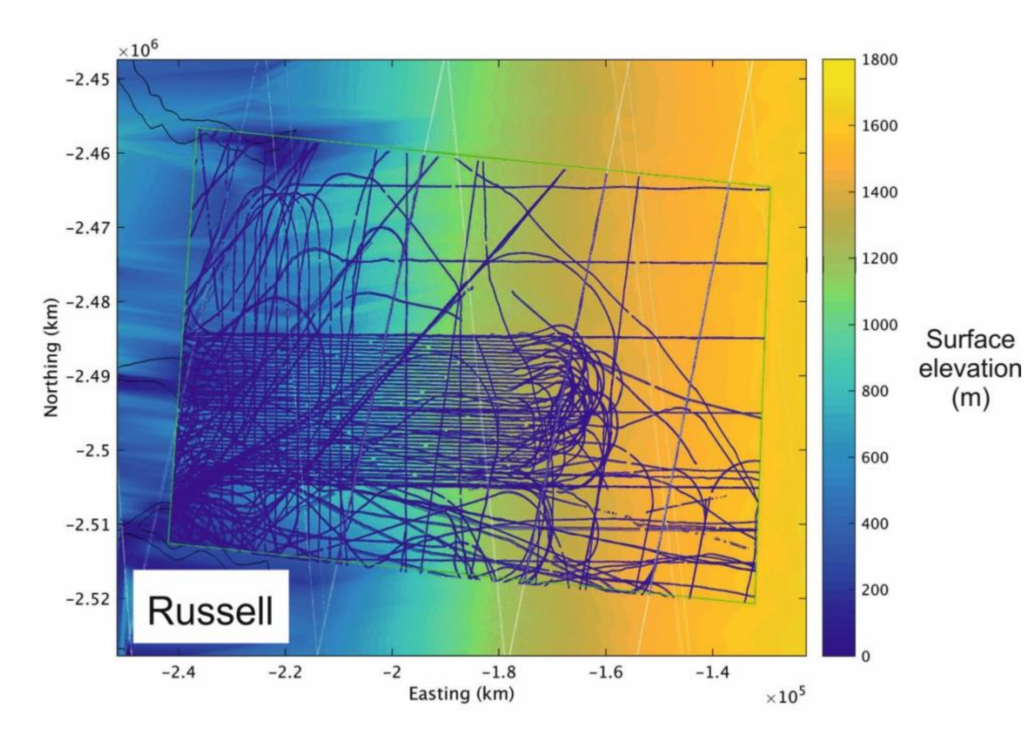


Figure 4. Operation IceBridge ATM (blue) airborne laser altimetry flightlines, and ICESat (white) ground tracks over the SPICE Greenland study site. The green polygon marks the boundary of the Russell Glacier site.

4. Validation methods

4.1 Review of Validation methods

The validation activities within the SPICE study served two principal aims. Firstly they were used during the inter-comparison and development phase of SPICE, to assess the impact of the various different candidate L1 and L2 processor options. Secondly, they were used in the final validation stage, once the optimal Phase 2 configuration had been identified, to (1) assess the extent to which the SPICE algorithms improved upon existing baseline solutions and (2) provide a final error characterization of the SPICE products.

To objectively assess the overall quality of the various datasets, it was necessary to compare them to high quality reference datasets, in order to determine their absolute accuracy. Reference datasets consisted of airborne and laser altimeter tracks of surface elevation, as described in Section 3. By combining these two sources of validation data we were able to achieve comprehensive coverage at all 4 study sites.

To perform the validation of the SPICE output datasets generated in WP's 2 & 3 required two principle steps. First, the altimeter elevation estimates produced in WP3 needed to be relocated to the point on the illuminated ice surface that was closest to the satellite (the so-called *Point of Closest Approach*). This step enables a like-for-like comparison between the satellite and airborne measurements, taking into account the different sizes of their beam footprints. For this process, we used an established method (Roemer et al., 2007) to relocate the altimeter measurements, which is described in more detail in Annex 1 of this document.

Secondly, we then compared the relocated altimeter measurements to nearby reference data. To do this, we firstly identified locations where SPICE and validation points were separated by no more than 100 metres from each other. For each SPICE data point that satisfied this criteria, we then selected the validation record that was closest to the altimeter measurement, to account for instances where multiple validation records were present within the 100 m search radius. We then corrected the reference data elevation to account for the fact that they were not exactly co-located with the satellite measurements, and that for non-flat surfaces there will consequently be an elevation difference purely due to the difference in the sampling position. This procedure was repeated for every validation pair, to generate a set of differences and associated statistics for each scenario. This validation process was then repeated for each reference dataset, for each study site, and for each SPICE dataset, to build up a comprehensive set of statistics for inter-comparison and analysis.

4.2 Design of validation process

The primary purpose of the validation activities was to objectively assess the accuracy of the Level-2 elevations, which themselves represent one of the fundamental geophysical altimetry products. Because validation is therefore performed at the L2 stage, inter-comparison of WP2 (L1 processing) options requires a choice regarding WP3 (L2 processing) configuration, and vice versa. The approach and rationale for our choice is as follows. First we inter-compare the different WP3 (L2) processing options, because during Phase 1 of the study it was clear that the first order challenge associated with improving SAR elevation measurements was that associated with robust retracking in areas of complex topography terrain. By tackling this step first, we therefore addressed the first order problem, which is otherwise likely to mask more subtle improvements or degradations due the various L1 processing configurations. Having identified the optimal L2 processing configuration, we then proceeded to evaluate the more subtle differences resulting from the different WP2 (DDP) evolutions.

4.3 Performance metrics

One of the principle aims of the validation activities was to provide a statistical justification for the selection of the optimum processing configuration. This required that all the different validation information (different sites, different reference datasets and so on) was integrated to arrive at a final conclusion and associated recommendations. At the outset of Phase 2, it was therefore necessary to define performance metrics that could be used to rank the different options during the inter-comparison activities. Based on experiences from Phase 1 of the study, it was clear that the elevation differences at each site tended not to be normally distributed. We therefore selected the following statistics to describe the distribution of elevation differences relative to the fiducial measurements, and then used these metrics to evaluative each SPICE dataset:

(1) The median elevation difference was used to estimate the overall elevation bias associated with each processing scenario.
(2) The Median Absolute Deviation (MAD) from the median was used to estimate the overall dispersion of elevation differences associated with each processing scenario.

To combine each of these metrics across all scenarios, we then computed the Root-Mean-Square (RMS) of each of (1) and (2), across the different study sites and validation datasets. Finally, to provide a single metric that could be used inter-compare, and rank, the different processing options, we computed the RMS of the bias (1) and dispersion (2) statistics.

Once the optimum Phase 2 configuration had been identified, we then compared the optimum Phase 2 solution to the Phase 1 baseline configuration. Furthermore, to assess the significance of any differences between Phase 1 and Phase 2 configurations, we performed significance testing on each paired set of Phase 1-Phase 2 elevation differences. This enabled us to assess whether there was a statistically significant difference between the two configurations, based on both the central values and the dispersion of the differences. For this, we selected a 5% significance threshold and tested using the non-parametric Mann Whitney U (Hollander et al., 2015) and Kolmogorov-Smirnov (Massey, 1951) tests for the central values and distribution, respectively. The Mann Whitney U test is a non-parametric test used to assess the equality of population medians, and is equivalent to the Wilcoxon rank sum test. It is chosen here because the set of elevation differences tends not to be normally distributed, and we have therefore used the median elevation difference as a measure of the relative bias between datasets. Likewise, the Kolmogorov-Smirnov test, which tests for significant differences between two population distributions, is used because it does not assume the data are normally distributed. A 5% threshold was chosen in accordance with standard statistical practice to ensure that the probability of mistakenly identifying a significant difference in the population, as an artefact of the sampling distribution, was below 5%.

5. Results of validation activities

This section describes the results arising from the validation activities conducted during WP4. In Sections 5.1-5.4 we address the inter-comparison of the various SPICE processing options, which forms the basis for our selection of our final Phase 2 processing configuration. Sections 5.5 and 5.6 then assess the relative change in performance between our baseline and final solutions. More specifically, in Section 5.1, we present the inter-comparison of Level-2 SAR processing configurations. In Section 5.2, we inter-compare the Level-1 SAR processing options. Section 5.3 covers the pLRM options inter-comparison, Section 5.4 provides an analysis of the impact of temporal changes in elevation on our inter-comparison, and Sections 5.5 and 5.6 describe the analysis of overall change in performance between baseline (Phase 1) and SPICE optimum (Phase 2) solutions, for SAR and pLRM, respectively.

5.1 Validation of SAR L2 processing options

During WP3, several novel L2 processing evolutions were implemented. This section describes the independent evaluation of these candidate L2 configurations, which forms the basis for the selection of our final Phase 2 product. We firstly tested the performance of two novel pre-retracking modules, which were designed to stabilise the retracking process, in instances where the waveform was complex and exhibited multiple peaks. These type of waveforms are common in regions of complex ice topography, such as around much of the ice sheet margin, where multiple distinct surface reflections are captured within the receive window. In summary, these pre-retracking modules consisted of:

1. Using an auxiliary DEM to seed the retracking process, by identifying the waveform peak closest to the nadir range.
2. Using Batch processing of sequences of multiple waveforms to ensure along-track consistency in the peak selected for retracking.

Furthermore, we also compared empirical and analytical retrackers, both with and without the pre-retracking modules described above, to assess which type of approach to retracking was better suited to SAR altimetry over ice sheets. The selected retrackers were (1) an empirical Threshold Peak Retracker (TPR), based upon

Phase 1 analysis and past studies such as CryoVal-LI [AD6], and (2) a new analytical retracker (AR) based upon the SAMOSA heritage. Further details relating to these algorithms are provided in the Algorithm Theoretical Basis Document [AD7].

In total, these new options provided 6 candidate SAR L2 configurations ({Normal, Batch, DEM} x {TPR, AR}). Each solution was compared to both validation datasets, at each of our Antarctic study sites ({IceBridge, ICESat} x {Dome C, Vostok, Spirit}), giving a total of 36 scenarios from which to compile our overall validation statistics. At this inter-comparison stage, we focused our analysis on the sites where we had actual SAR altimetry acquisitions, namely Dome C, Vostok and Spirit, because we wanted our conclusions to be as closely applicable to the Sentinel-3 configuration as possible. The results of the SAR L2 inter-comparison are presented in Table 3.

Spirit							
STATISTIC	PHASE 1	PHASE 2 CONFIGURATIONS					
	Normal	**Normal**		**DEM**		**Batch**	
	SAR TPR	**SAR TPR**	**SAR AR**	**SAR TPR**	**SAR AR**	**SAR TPR**	**SAR AR**
IceBridge no. of measurements	*102*	*102*	*102*	*85*	*84*	*102*	*102*
IceBridge median difference (m)	-2.95	-2.48	-2.60	-0.20	0.79	-2.43	-3.36
IceBridge median absolute deviation (m)	5.27	4.69	6.46	2.20	3.32	4.69	5.64
ICESat no. of measurements	*58*	*58*	*58*	*48*	*48*	*58*	*58*
ICESat median difference (m)	-1.75	-2.08	-2.88	0.40	-0.59	-1.64	-3.30
ICESat median absolute deviation (m)	5.68	6.05	5.64	4.10	2.49	5.60	6.49

Dome C							
STATISTIC	PHASE 1	PHASE 2 CONFIGURATIONS					
	Normal	**Normal**		**DEM**		**Batch**	
	SAR TPR	**SAR TPR**	**SAR AR**	**SAR TPR**	**SAR AR**	**SAR TPR**	**SAR AR**
IceBridge no. of measurements	*28*	*28*	*28*	*33*	*33*	*28*	*28*
IceBridge median difference (m)	-1.57	-1.04	-2.05	-0.86	-2.05	-1.04	-1.21
IceBridge median absolute deviation (m)	0.32	0.29	0.22	0.26	0.29	0.29	0.28
ICESat no. of measurements	*51*	*51*	*51*	*46*	*46*	*51*	*51*
ICESat median difference (m)	-2.02	-1.42	-2.54	-1.47	-2.63	-1.42	-1.65
ICESat median absolute deviation (m)	0.30	0.25	0.40	0.15	0.19	0.25	0.40

Vostok							
STATISTIC	PHASE 1	PHASE 2 CONFIGURATIONS					
	Normal	**Normal**		**DEM**		**Batch**	

	SAR TPR	SAR TPR	SAR AR	SAR TPR	SAR AR	SAR TPR	SAR AR
IceBridge no. of measurements	*25*	*25*	*25*	*23*	*23*	*25*	*25*
IceBridge median difference (m)	-1.46	-0.86	-1.92	-0.50	-1.42	-0.86	-1.46
IceBridge median absolute deviation (m)	0.19	0.21	0.24	0.73	0.94	0.21	0.37
ICESat no. of measurements	*228*	*228*	*228*	*237*	*237*	*228*	*228*
ICESat median difference (m)	-2.21	-1.66	-2.73	-1.39	-2.32	-1.66	-1.99
ICESat median absolute deviation (m)	0.30	0.24	0.35	0.35	0.53	0.24	0.39
Summary							
STATISTIC	PHASE 1	PHASE 2 CONFIGURATIONS					
	Normal	**Normal**		**DEM**		**Batch**	
	SAR TPR	**SAR TPR**	**SAR AR**	**SAR TPR**	**SAR AR**	**SAR TPR**	**SAR AR**
RMS of the median differences (m)	2.05	1.69	2.48	0.94	1.80	1.59	2.33
RMS of the median absolute deviations (m)	3.17	3.13	3.51	1.93	1.75	2.99	3.52
RMS of the median and median absolute deviations (m)	2.67	2.52	3.04	1.52	1.78	2.39	2.99

Table 3. Summary of statistics for the inter-comparison of the SAR L2 processing options. For each site and processing scenario, elevations were compared to both IceBridge airborne data and ICESat laser altimetry data. For each statistic (row), boxes highlighted in green indicate the scenario presenting the closest agreement to the reference datasets. To assimilate statistics across all sites and validation datasets, the RMS difference was taken for both the bias (median) and dispersion (median absolute deviation). In this calculation, no weighting was applied in accordance with the number of observations, in order to make sure that each option carried equal weight and that the summary statistics were not skewed towards sites with a higher number of measurements. Finally, to provide a single metric for the quality of each processing scenario, we computed the RMS of the median and median absolute deviation (final row).

The analysis presented in Table 3 shows that, overall, introducing DEM-based pre-retracking in combination with the TPR retracker yields SAR elevation measurement that are in overall closest agreement with our independent validation datasets. This configuration not only has the lowest overall RMS error (final row, Table 3), but is also ranked top in the greatest number of validation runs (best performing in 7 out of 12 runs, compared to 2 out of 12 for the next best configurations Normal+TPR and Batch+TPR).

The benefits of introducing DEM pre-retracking are particularly clear at the Spirit site, which has steeper and more complex topography. Here the absolute bias has been reduced from ~ 2 metres to less than 50 cm (TPR retracking) and the dispersion of the differences is typically one-half to one-third of the other pre-retracking configurations. Given that the pre-retracking modules were specifically designed to improve performance in complex regions such as Spirit, it is encouraging that our DEM approach has achieved this. Regarding the Batch method, further work is needed to understand and adapt this novel approach, and this will be addressed in more detail within the scientific roadmap.

Turning to the retracking inter-comparison, it is evident from the statistics that, in general, the empirical (TPR) retracker achieves a better accuracy than its analytical (AR) counterpart. The average RMS error of the TPR solutions is ~50 cm (~20%) lower than the AR retrievals (final row, Table 3), and the TPR outperforms the AR in 31 out of the 36 scenarios. The AR tends to perform less well at all sites, and based on the analysis presented here, there is no evidence that the relative performance between the AR and TPR solutions varies as a function of topographic complexity. Based on this evaluation, alongside the analysis of waveforms and retracking conducted in WP3, we conclude that even the current generation of analytical retrackers struggle to accommodate the complexity of ice sheet echoes. Although the AR can often be tuned manually to fit individual ice sheet waveforms, establishing a more universal set of parameters that is suitable for large-scale, automated processing remains a target for on-going research. Modelling the convolution of the response from often complex topography and the poorly known depth distribution of scattering elements is a challenging task, and one that will require further dedicated study.

Based upon this analysis of L2 SAR processing options, and the selection procedure identified in Section 4.3, we therefore choose the DEM+TPR as our final Phase 2 configuration. Next we use this configuration to inter-compare the performance of the various SAR L1 processing options.

5.2 Validation of SAR L1 processing options

Using the optimal Phase 2 Level-2 processing configuration outlined above, we performed an inter-comparison of the different Level-1 DDP options. As with our previous analysis, we use the Median and Median Absolute Deviation as evaluation metrics and compute the Root-Mean-Square to combine across different validation datasets and study sites. Because there were initially a large number of possible DDP options, we conducted a 2-stage evaluation, whereby all options were used to generate a preliminary L1b product, the various L1b were then examined during WP2 activities, and only credible candidate configurations were progressed through the L2 processing to undergo full independent evaluation. Table 4 presents the summary statistics for the different DDP options that progressed to the full evaluation stage, which were (1) to increase the Zero-Padding factor to 4 in the range FFT, (2) to apply a Hamming weighting, and (3) to perform the multi-looking without incorporating samples with a value of zero. In the case of the latter option, zero-valued samples can occur due to wrapping in the window after the geometrical corrections are applied, or if stack masking is used [AD5, AD7]. Also given as a baseline reference is the standard Phase 1 DDP options, namely a Zero-Padding factor of 1, no Hamming weighting and inclusion of zero's in the multi-looking, which was then combined with the optimal Phase 2 L2 processing.

Dome C				
STATISTIC	BASELINE DDP + PHASE 2 L2	PHASE 2 DDP OPTIONS + PHASE 2 L2		
		Zero-padding 4	**Hamming weighting**	**No multi-looking zeros**
IceBridge no. of measurements	*33*	*33*	*33*	*33*
IceBridge median difference (m)	-0.86	-0.61	-0.97	-0.87
IceBridge median absolute deviation (m)	0.26	0.22	0.32	0.26
ICESat no. of measurements	*46*	*46*	*46*	*46*
ICESat median difference (m)	-1.47	-1.17	-1.51	-1.47
ICESat median absolute deviation (m)	0.15	0.20	0.21	0.15

Vostok				
STATISTIC	BASELINE DDP + PHASE 2 L2	PHASE 2 DDP OPTIONS + PHASE 2 L2		
		Zero-padding 4	**Hamming weighting**	**No multi-looking zeros**
IceBridge no. of measurements	*23*	*23*	*23*	*23*
IceBridge median difference (m)	-0.50	-0.27	-0.50	-0.50
IceBridge median absolute deviation (m)	0.73	0.64	0.82	0.73
ICESat no. of measurements	*237*	*237*	*237*	*237*
ICESat median difference (m)	-1.39	-1.15	-1.46	-1.45
ICESat median absolute deviation (m)	0.35	0.34	0.36	0.39

Spirit				
STATISTIC	BASELINE DDP + PHASE 2 L2	PHASE 2 DDP OPTIONS + PHASE 2 L2		
		Zero-padding 4	**Hamming weighting**	**No multi-looking zeros**
IceBridge no. of measurements	*85*	*85*	*85*	*85*
IceBridge median difference (m)	-0.20	-0.11	0.03	-0.24
IceBridge median absolute deviation (m)	2.20	2.11	2.40	2.19

ICESat no. of measurements	*48*	*48*	*48*	*48*
ICESat median difference (m)	0.40	-0.55	0.78	-0.57
ICESat median absolute deviation (m)	4.10	3.89	4.60	3.78
Summary				
STATISTIC	BASELINE DDP + PHASE 2 L2	PHASE 2 DDP OPTIONS + PHASE 2 L2		
		Zero-padding 4	**Hamming weighting**	**No multi-looking zeros**
RMS of the median differences (m)	0.94	0.76	1.02	0.97
RMS of the median absolute deviations (m)	1.93	1.83	2.15	1.82
RMS of the median and median absolute deviations (m)	1.52	1.40	1.69	1.46

Table 4 Summary of statistics for the inter-comparison of L1 SAR processing options. For each site and processing scenario, elevations were compared to both IceBridge airborne data and ICESat laser altimetry data. For each statistic (row), boxes highlighted in green indicate the scenario presenting the closest agreement to the reference datasets. To assimilate statistics across the various sites and validation datasets, the RMS difference was taken for both the bias (median) and dispersion (median absolute deviation). In this calculation, no weighting was applied in accordance with the number of observations, in order to make sure that each option carried equal weight and that the summary statistics were not skewed towards sites with a higher number of measurements. Finally, to provide a single metric for the quality of each processing scenario, we computed the RMS of the median and median absolute deviation (final row).

Compared to the variation between L2 processing scenarios (Section 5.1), the differences between the various L1 options are much less pronounced. This confirms our preliminary Phase 1 analysis and further supports the decision to address first the L2 processing, before then turning to the L1 algorithm developments. In total, increasing the zero-padding factor to 4 yields the optimal solution in 8 out of the 12 different runs. The magnitude of the improvement is, however, not large and mostly manifests as a reduction in the elevation bias, rather than the dispersion of the differences (Table 4). In contrast, in almost all (10 out of 12) cases, introducing the Hamming weighting degrades the accuracy, in comparison to our baseline solution. The final option, which is to omit zero samples from the multi-looking, has minimal impact on accuracy, in comparison to our baseline scenario. Based upon this analysis and the statistics presented in Table 4, we therefore select zero-padding 4 as our final Phase 2 L1 processing configuration.

5.3 Validation of pLRM processing options

In addition to SAR algorithm evolution, the SPICE study has also developed a new pLRM processor, and associated L1b and L2 measurements, from the same SAR FBR data. These activities have been designed to investigate the possibility of generating a low resolution product, similar to the historical record, from a closed burst SAR mode of operation. Although the main focus of the validation activities, in line with the study objectives, has been to assess the performance of the various SAR configurations, we have additionally conducted an inter-comparison of pLRM measurements derived using a range of different processing configurations, in order to assess any associated impact on the quality of the L2 elevation observations. In terms of L2 processing, we inter-compare solutions derived using each of the pre-retracking modules, in combination with the TCOG retracker based on the analysis from Phase 1. We choose to focus our inter-comparison on assessing the impact of the different pre-retracking modules, as (1) these represent the most important novel L2 processing evolution of this study, and (2) we wish to determine whether the substantial improvements shown in the SAR analysis also hold for the lower resolution pLRM mode too.

In terms of the L1 processing, it is important to note that because pLRM does not involve SAR processing of coherent echoes, most of the L1 options are not applicable here. Nonetheless, the range FFT is still applied and so we are able to assess the impact of changing the Zero-Padding Factor in this calculation. As with the previous analysis in sections 5.1 and 5.2, we use the same metrics to perform this evaluation. The summary statistics resulting from this analysis are presented in Table 5.

Dome C						
	PHASE 1	PHASE 2 CONFIGURATIONS				
Zero-padding factor		**ZP4**	**ZP2**	**ZP4**	**ZP2**	**ZP4**
Pre-retracking	**Normal**	**Normal**	**Batch**	**Batch**	**DEM**	**DEM**
Retracking	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**
IceBridge no. of measurements	*25*	*25*	*25*	*25*	*28*	*28*
IceBridge median difference (m)	-1.04	-0.76	-1.04	-0.76	-1.50	-0.88

IceBridge median absolute deviation (m)	0.23	0.30	0.23	0.30	0.59	0.34
ICESat no. of measurements	*52*	*52*	*52*	*52*	*48*	*48*
ICESat median difference (m)	-1.57	-1.32	-1.57	-1.32	-1.75	-1.29
ICESat median absolute deviation (m)	0.36	0.37	0.36	0.37	0.38	0.32
Vostok						
	PHASE 1	PHASE 2 CONFIGURATIONS				
Zero-padding factor	**ZP2**	**ZP4**	**ZP2**	**ZP4**	**ZP2**	**ZP4**
Pre-retracking	**Normal**	**Normal**	**Batch**	**Batch**	**DEM**	**DEM**
Retracking	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**
IceBridge no. of measurements	*21*	*21*	*21*	*21*	*22*	*22*
IceBridge median difference (m)	-0.89	-0.63	-0.89	-0.63	-0.90	-0.68
IceBridge median absolute deviation (m)	0.26	0.18	0.26	0.18	0.55	0.49
ICESat no. of measurements	*216*	*216*	*216*	*216*	*232*	*232*
ICESat median difference (m)	-1.69	-1.44	-1.69	-1.44	-1.39	-0.99
ICESat median absolute deviation (m)	0.32	0.32	0.32	0.32	0.74	0.67
Spirit						
	PHASE 1	PHASE 2 CONFIGURATIONS				
Zero-padding factor	**ZP2**	**ZP4**	**ZP2**	**ZP4**	**ZP2**	**ZP4**
Pre-retracking	**Normal**	**Normal**	**Batch**	**Batch**	**DEM**	**DEM**
Retracking	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**
IceBridge no. of measurements	*124*	*124*	*124*	*124*	*72*	*72*
IceBridge median difference (m)	-2.33	-1.97	-1.95	-1.77	9.30	9.40
IceBridge median absolute deviation (m)	7.30	7.30	7.10	7.30	8.57	4.71
ICESat no. of measurements	*71*	*71*	*71*	*71*	*50*	*49*
ICESat median difference (m)	0.19	0.67	0.19	0.67	3.78	5.46

ICESat median absolute deviation (m)	3.70	4.10	3.70	4.10	8.27	3.69
Summary						
	PHASE 1	PHASE 2 CONFIGURATIONS				
Zero-padding factor	**ZP2**	**ZP4**	**ZP2**	**ZP4**	**ZP2**	**ZP4**
Pre-retracking	**Normal**	**Normal**	**Batch**	**Batch**	**DEM**	**DEM**
Retracking	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**	**pLRM TCOG**
RMS of the median differences (m)	1.45	1.23	1.36	1.18	4.26	4.51
RMS of the median absolute deviations (m)	3.35	3.43	3.28	3.43	4.88	2.47
RMS of the median and median absolute deviations (m)	2.58	2.57	2.51	2.56	4.58	3.64

Table 5. Summary of statistics for the inter-comparison of pLRM processing options. For each site and processing scenario, elevations were compared to both IceBridge airborne data and ICESat laser altimetry data. For each statistic (row), boxes highlighted in green indicate the scenario presenting the closest agreement to the reference datasets. To assimilate statistics across the various sites and validation datasets, the RMS difference was taken for both the bias (median) and dispersion (median absolute deviation). In this calculation, no weighting was applied in accordance with the number of observations, in order to make sure that each option carried equal weight and that the summary statistics were not skewed towards sites with a higher number of measurements. Finally, to provide a single metric for the quality of each processing scenario, we computed the RMS of the median and median absolute deviation (final row).

In comparison to the SAR inter-comparison (Sections 5.1 and 5.2), we find that pLRM performance is much less sensitive to the choice of Level-1 and Level-2 processing configuration. Furthermore, there is no one single configuration that consistently outperforms the others across the different validation runs. Most notable is that, unlike with the SAR processing where the pre-retracking offered a significant improvement, for pLRM we do not find that this is the case. That the pre-retracking module has less impact on the pLRM than on the SAR is not surprising, given that each LRM sample integrates power over a larger area and it is therefore less likely that multiple distinct peaks will be resolved in the waveform. Furthermore, it is interesting to observe that the DEM pre-retracking module tends to degrade the pLRM solution, particularly at Spirit where there is more complex topography and a steep along-track slope. In this case, it is likely that the larger along-track beamwidth of the pLRM, combined with the reduced number of multi-peaked waveforms, means that the DEM-based method is less able to correctly identify a true nadir return. As a result the upslope migration of the echoing point is not adequately accounted for, and the elevations tend to be biased high. While the principal focus of the DEM method was SAR mode acquisitions, it may be possible in future to better adapt the approach for pLRM, and this will be discussed further in the scientific roadmap. Overall, the ZP2+Batch+TCOG performs marginally better according to our predefined assessment criteria, and so it is selected as our Phase 2 pLRM configuration.

5.4 Sensitivity of results to regional elevation change

Because the timestamp of the SPICE acquisitions (2014) and the validation data (2009-2013) were different, prior to making final conclusions we also evaluated the temporal rates of elevation change for each study site. For this, we used published elevation rates derived from CryoSat-2 for the period 2010-2013 inclusive (McMillan et al., 2014). At inland sites the median rate of elevation change was small, equating to +0.004 m/yr (Vostok) and +0.013 m/yr (Dome C). Compared to the overall magnitude of the elevation biases (Tables 3-5), the cumulative impact of these rates of elevation change are therefore minimal (typically 1-5%), and do not alter the best performing configurations identified above. At coastal sites the rates of elevation change are larger in magnitude and are negative (median values of -0.026 m/yr and -0.87 m/yr, at Spirit and Russell, respectively). The associated impact on our statistics at these sites is not uniform. For example, at the Spirit site, adjusting for elevation change reduces the bias of Phase 2 SAR measurements relative to IceBridge from -0.20 m to -0.10 m, whereas it increases the bias relative to ICESat from +0.40 m to +0.53 m. Nonetheless, for all but one of the inter-comparison runs, it does not change the scenario that is identified as the best performing configuration. The one exception occurs when comparing the different SAR options to ICESat at Spirit where, with an elevation correction applied, the DEM + AR L2 configuration returns a slightly lower magnitude of bias (-0.46 m) than the DEM + TPR configuration (+0.53 m). Even so, the difference between these configurations is not large, and it does not affect the overall RMS ranking used to choose the optimal processing configuration. We therefore conclude that our Phase 2 selections are not dependent upon whether an elevation rate correction is applied at our study sites. We do note that for future studies, however, if the inter-comparison was to be done at other regions, for example the Amundsen Sea where dynamically driven rates of elevation change are much higher, then it is likely that correcting for differences in time stamp between the datasets would indeed be necessary.

5.5 Evaluation of final SAR products

Based upon the analysis presented above, we define our optimal Phase 2 SAR processing configuration to consist of a Zero-Padding factor of 4 within the L1 DDP, the DEM pre-retracking module and the Threshold Peak Retracker. To synthesise all the experiments conducted above, we now present a comparison between the ESA L2 SAR dataset, our baseline Phase 1 configuration, which has a Zero-Padding Factor of 1 and no pre-retracking module, and our final Phase 2 product. Having now completed the options inter-comparison, we also extend our analysis to include the bespoke SPICE pSAR processing at the Russell Glacier site. Figure 5, **Figure 6** and Table 6 summarise the statistics for each of these processing chains across all sites, together with the change in performance between our baseline (Phase 1) and optimal (Phase 2) solution. Finally, we test for significance to identify whether the observed changes between Phase 1 and Phase 2 are significant at the 5 % level.

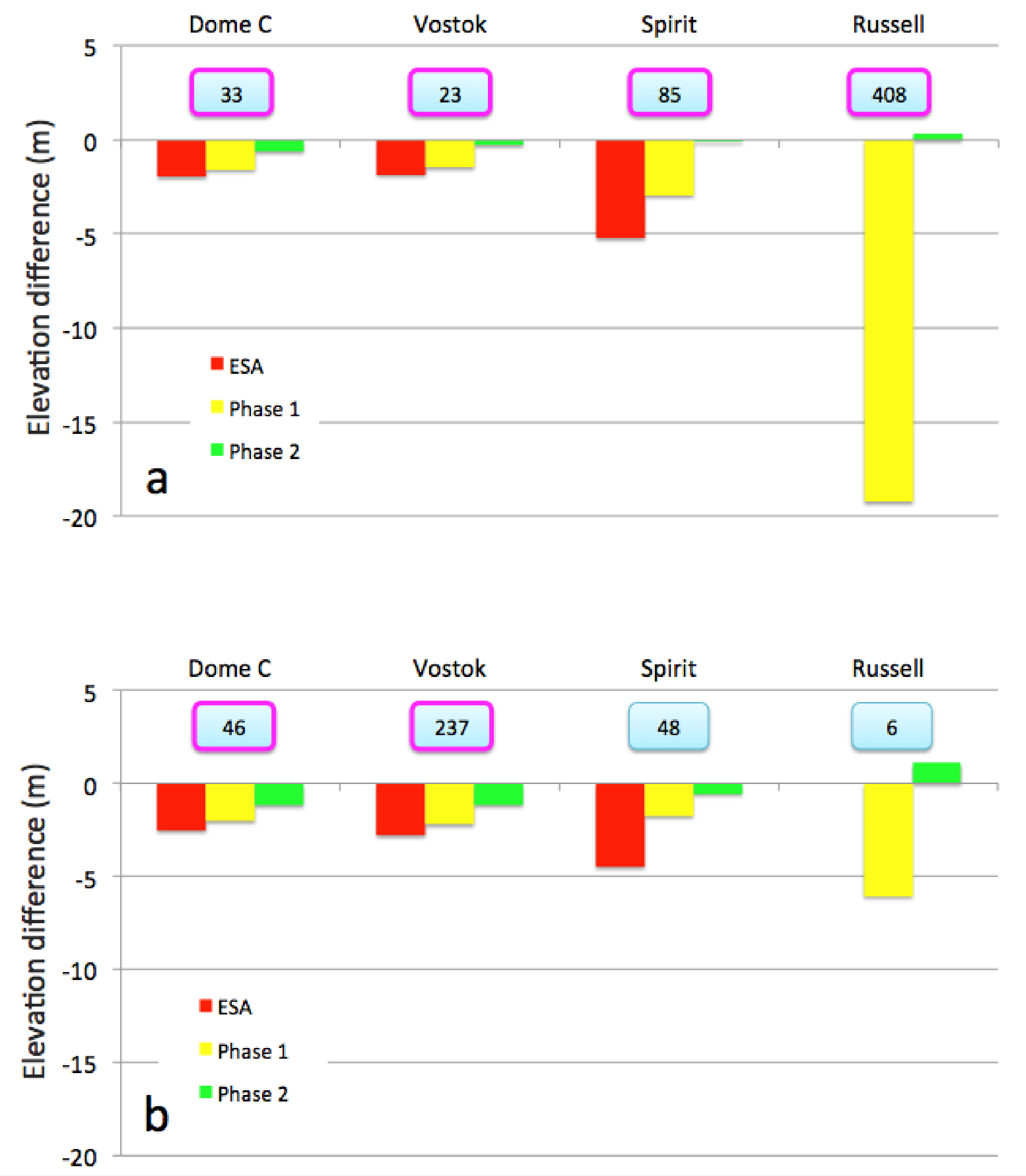


Figure 5. Elevation biases of ESA, SPICE Phase 1 and SPICE Phase 2 SAR data, relative to (a) IceBridge and (b) ICESat reference datasets. The cyan boxes indicate the number of validation pairs at each site, which were used in the inter-comparison. Boxes highlighted in magenta indicate a significant difference between Phase 1 and Phase 2 datasets, based on a 5% significance threshold. No ESA SAR data are present at the Russell Glacier site because this falls within the SARIn mode mask and the SARIn FBR data were converted to pSAR as part of the SPICE study.

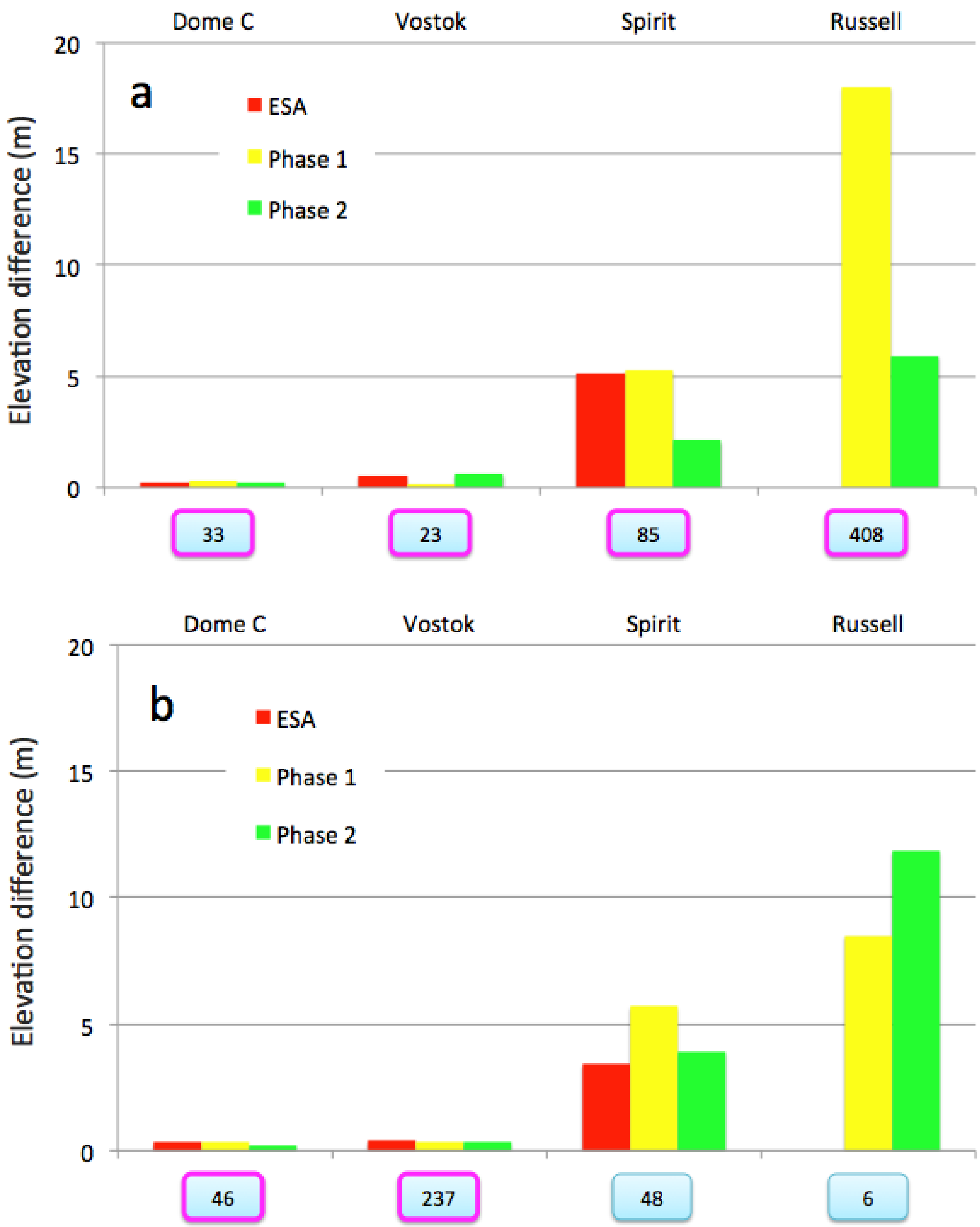


Figure 6. The dispersion of ESA, SPICE Phase 1 and SPICE Phase 2 SAR elevation differences, relative to (a) IceBridge and (b) ICESat reference datasets. The cyan boxes indicate the number of validation pairs at each site, which were used in the inter-comparison. Boxes highlighted in magenta indicate a significant difference between Phase 1 and Phase 2 datasets, based on a 5% significance threshold. No ESA SAR data are present at the Russell Glacier site because this falls within the SARIn mode mask and the SARIn FBR data were converted to pSAR as part of the SPICE study.

Spirit					
	ESA L2	PHASE 1	PHASE 2	PHASE 1 to PHASE 2 change (m)	PHASE 1 to PHASE 2 change (%)
IceBridge median difference (m)	-5.22	-2.95	-0.11	-2.84	-96
IceBridge median absolute deviation (m)	5.10	5.27	2.11	-3.17	-60
ICESat median difference (m)	-4.49	-1.75	-0.55	-1.20	-69
ICESat median absolute deviation (m)	3.40	5.68	3.89	-1.80	-32
Dome C					
	ESA L2	PHASE 1	PHASE 2	PHASE 1 to PHASE 2 change (m)	PHASE 1 to PHASE 2 change (%)
IceBridge median difference (m)	-1.96	-1.57	-0.61	-0.96	-61
IceBridge median absolute deviation (m)	0.26	0.32	0.22	-0.10	-31
ICESat median difference (m)	-2.53	-2.02	-1.17	-0.85	-42
ICESat median absolute deviation (m)	0.30	0.30	0.20	-0.10	-33
Vostok					
	ESA L2	PHASE 1	PHASE 2	PHASE 1 to PHASE 2 change (m)	PHASE 1 to PHASE 2 change (%)
IceBridge median difference (m)	-1.84	-1.46	-0.27	-1.20	-82
IceBridge median absolute deviation (m)	0.54	0.19	0.64	0.44	229
ICESat median difference (m)	-2.79	-2.21	-1.15	-1.06	-48
ICESat median absolute deviation (m)	0.38	0.30	0.34	0.04	13
Russell					

	ESA L2	PHASE 1	PHASE 2	PHASE 1 to PHASE 2 change (m)	PHASE 1 to PHASE 2 change (%)
IceBridge median difference (m)	-	-19.25	0.30	-18.95	-98
IceBridge median absolute deviation (m)	-	18.00	5.93	-12.07	-67
ICESat median difference (m)	-	-6.05	1.10	-4.94	-82
ICESat median absolute deviation (m)	-	8.50	11.82	3.32	39

Table 6. Comparison between ESA, SPICE Phase 1 and SPICE Phase 2 SAR processing scenarios. Green shading indicates a statistically significant improvement between Phase 1 and Phase 2, turquoise shading indicates a non-significant difference between Phase 1 and Phase 2, and red shading indicates a significant reduction in the accuracy between Phase 1 and Phase 2. Significance is measured at the 5% level using the Mann Whitney U and Kolmogorov-Smirnov tests for the central values and the distributions, respectively. No ESA SAR data are available at the Russell Glacier site as acquisitions are in SARIn mode only. SPICE solutions at Russell Glacier are a pseudo-SAR product, which was processed from SARIn FBR using capability developed as part of this study.

The results of the inter-comparison between the ESA product and SPICE Phase 1 and Phase 2 baselines indicate a widespread improvement in elevation accuracy delivered by our Phase 2 solution. In summary, 10 out of 16 validation runs show a significant improvement from Phase 1 to Phase 2, 4 out of 16 runs show no significant change, and 2 out of 10 validation runs show a significantly poorer performance. Of these latter two instances of reduced accuracy, one is based upon our smallest validation dataset (Vostok, IceBridge, 23 measurements) and one shows only a very small magnitude (4 cm), albeit significant, change. Ideally, future airborne acquisitions in this region would enable a more comprehensive validation of differences at these sites. Broadly speaking, the greatest magnitude, and percentage, improvements are found at the high relief coastal sites of Spirit and Russell Glacier. At these sites, relative improvements in accuracy are typically 60%-100%, suggesting that the pre-retracking module offers the greatest benefit in areas of complex topography with multi-peaked waveforms. With our Phase 2 processing, the bias at these sites is reduced to ~ 1 metre or less, delivering an order of magnitude improvement compared to the ESA and Phase 1 datasets.

Turning to the instances of insignificant changes, we find that these occur at our low latitude coastal sites, Spirit and Russell Glacier, and specifically where the comparison is made relative to the ICESat reference dataset. The combination of complex topography and low latitude results in small validation datasets (due to the wide ICESat track spacing at low latitudes, **Figure 3** and Figure 4) and highly variable elevation differences (due to the complex topography), and means that although there are relatively large differences between the Phase 1 and Phase 2 datasets, they are not statistically significant.

Table 7 provides a final high level summary of the statistics for the ESA and SPICE SAR datasets, showing the typical (mean) values for the (1) bias, (2) dispersion, and (3) integrated RMS quality metric, averaged across all sites. In Table 7, we report the mean values for each statistic, so as to provide a single metric that integrates the performance at both the inland and coastal sites, in order to summarise the continent-wide performance of each of the different products. According to all statistics, the Phase 2 processing configuration improves upon both the Phase 1 and ESA solutions. The greatest change is evident in the reduction of the bias, where Phase 2 provides a 68% (135 cm) improvement relative to Phase 1. This means that the Phase 2 data is tracking much closer to the true surface (the air-snow interface), which the laser validation datasets range to. At 64 cm, the mean bias of the Phase 2 product is now well below 1 metre, which is encouraging given that it incorporates SAR measurements over both inland and complex coastal terrain. This provides clear evidence that sub-metre accuracy is possible with a non-interferometric SAR altimeter over all ice sheet surfaces. The overall dispersion of the differences is also substantially improved with the Phase 2 processing, by 39% and 26% from the Phase 1 and ESA datasets respectively, indicating that a better consistency of measurements is achievable with our processing evolutions. In summary, by combining both statistics, we find that the Phase 2 processing has been able to halve the error relative to our reference datasets.

Summary statistics for SAR data					
	ESA L2	PHASE 1	PHASE 2	PHASE 1 to PHASE 2 change (m)	PHASE 1 to PHASE 2 change (%)
Mean absolute bias across all sites (m)	3.14	1.99	0.64	-1.35	-68
Mean dispersion across all sites (m)	1.66	2.01	1.23	-0.78	-39
RMS of the mean bias and dispersion (m)	2.51	2.00	0.98	-1.02	-51

Table 7. Final summary of ESA, Phase 1 and Phase 2 SAR processing scenarios for Antarctic study sites. Note that Russell Glacier is not included in this summary table because no ESA L2 SAR data is available for the comparison.

5.6 Evaluation of final pLRM products

Finally, we turn to a similar comparison of the pLRM products. Based upon the analysis presented in Section 5.3, we define our optimal Phase 2 pLRM processing configuration to be that which uses a Zero-Padding factor of 2 within the L1 DDP, the Batch pre-retracking module and the Threshold Centre of Gravity Retracker. In this section, we evaluate the changes in accuracy at all sites between our Phase 1 and Phase 2 solutions. As before, we test for significance to identify whether the observed changes between Phase 1 and Phase 2 are significant at the 5 % level. Figure 7, Figure 8 and Table 8 summarise the results of this analysis. No ESA L2 solution is presented as pLRM processing is not currently included within the ground segment.

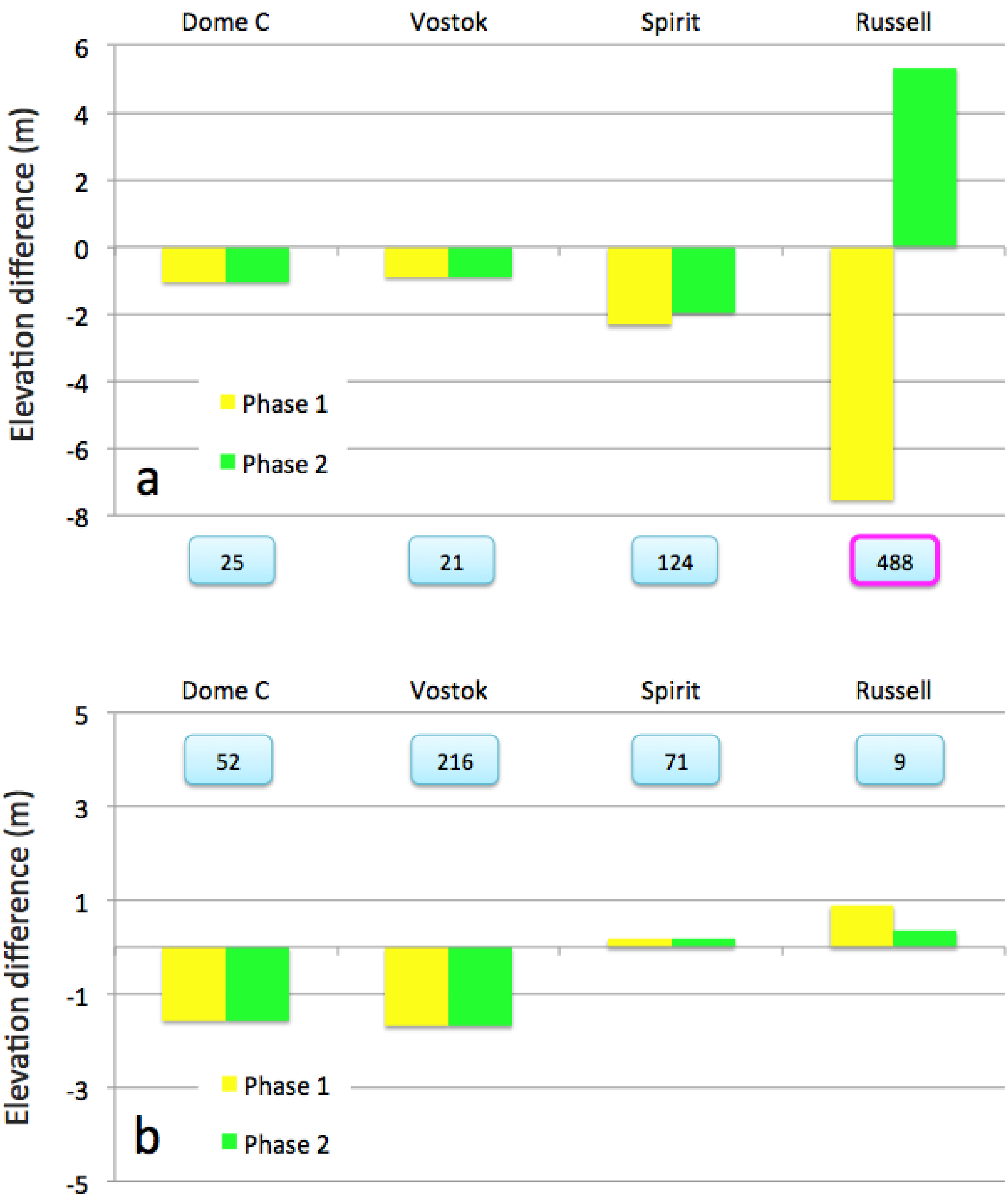


Figure 7. Elevation biases of ESA, SPICE Phase 1 and SPICE Phase 2 pLRM data, relative to (a) IceBridge and (b) ICESat reference datasets. The cyan boxes indicate the number of validation pairs at each site, which were used in the inter-comparison. Boxes highlighted in magenta indicate a significant difference between Phase 1 and Phase 2 datasets, based on a 5% significance threshold. No ESA data are present because generation of pLRM does not form part of the current CryoSat-2 ground segment.

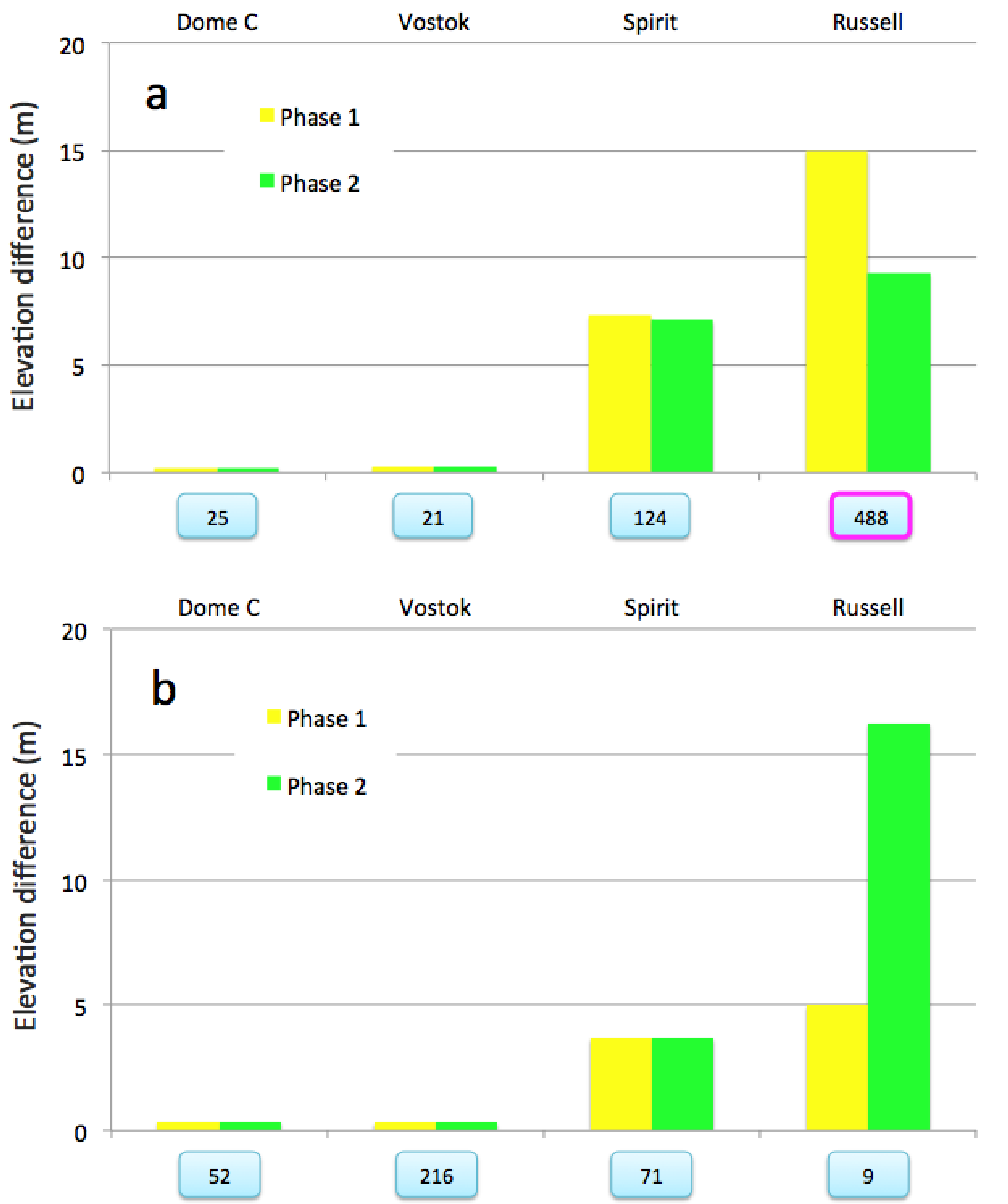


Figure 8. The dispersion of ESA, SPICE Phase 1 and SPICE Phase 2 pLRM elevation differences, relative to (a) IceBridge and (b) ICESat reference datasets. The cyan boxes indicate the number of validation pairs at each site, which were used in the inter-comparison. Boxes highlighted in magenta indicate a significant difference between Phase 1 and Phase 2 datasets, based on a 5% significance threshold. No ESA data are present because generation of pLRM does not form part of the current CryoSat-2 ground segment.

Spirit				
	PHASE 1	PHASE 2	PHASE 1 to PHASE 2 change (m)	PHASE 1 to PHASE 2 change (%)
IceBridge median difference (m)	-2.33	-1.95	-0.38	-16.12
IceBridge median absolute deviation (m)	7.30	7.10	-0.20	-2.74
ICESat median difference (m)	0.19	0.19	0.00	0.00
ICESat median absolute deviation (m)	3.70	3.70	0.00	0.00

Dome C				
	PHASE 1	PHASE 2	PHASE 1 to PHASE 2 change (m)	PHASE 1 to PHASE 2 change (%)
IceBridge median difference (m)	-1.04	-1.04	0.00	0.00
IceBridge median absolute deviation (m)	0.23	0.23	0.00	0.00
ICESat median difference (m)	-1.57	-1.57	0.00	0.00
ICESat median absolute deviation (m)	0.36	0.36	0.00	0.00

Vostok				
	PHASE 1	PHASE 2	PHASE 1 to PHASE 2 change (m)	PHASE 1 to PHASE 2 change (%)
IceBridge median difference (m)	-0.89	-0.89	0.00	0.00
IceBridge median absolute deviation (m)	0.26	0.26	0.00	0.00
ICESat median difference (m)	-1.69	-1.69	0.00	0.00
ICESat median absolute deviation (m)	0.32	0.32	0.00	0.00

Russell				
	PHASE 1	PHASE 2	PHASE 1 to PHASE 2 change (m)	PHASE 1 to PHASE 2 change (%)
IceBridge median difference (m)	-7.52	5.34	-2.18	-28.95

IceBridge median absolute deviation (m)	15.00	9.30	-5.70	-38.03
ICESat median difference (m)	0.90	0.36	-0.53	-59.55
ICESat median absolute deviation (m)	5.00	16.23	11.23	224.68

Table 8. Comparison between SPICE Phase 1 and Phase 2 pLRM processing scenarios. An ESA solution is not included because pLRM does not form part of the CryoSat-2 operational ground segment. Green shading indicates a statistically significant improvement between Phase 1 and Phase 2 and turquoise shading indicates a non-significant change. Significance is measured at the 5% level using the Mann Whitney U and Kolmogorov-Smirnov tests for the central values and distributions, respectively.

From this analysis (Table 8), it can be seen that the pLRM L1 and L2 options have made no measurable difference to the data quality at the inland sites of Dome C and Vostok, based upon the validation datasets we have available to us. Differences are evident at the coastal sites of Spirit and Russell Glacier, although they are mostly statistically insignificant at the 5% level. The only validation run that does indicate a statistically significant change is at the Russell Glacier site, when the data are compared to the large IceBridge reference dataset. Here we find that the Phase 2 configuration is able to improve the accuracy of the solution by ~ 30% - 40%. This suggests that the more sophisticated processing implemented in Phase 2 may yield benefits to pLRM in complex coastal areas, although more extensive validation datasets are required to understand whether this is indeed generalizable to other areas. Table 9 provides a final summary of the pLRM comparison statistics, showing the change in (1) bias, (2) dispersion, and (3) the integrated RMS quality metric, between the SPICE Phase 1 and Phase 2 datasets. These summary statistics show the small overall changes between the two pLRM processing configurations, which for all metrics are less than 10 cm (5%). On average our pLRM measurements have a bias of ~1.2 metres relative to the reference data, and a dispersion of ~2 metres.

Summary statistics for pLRM data				
	PHASE 1	PHASE 2	PHASE 1 to PHASE 2 change (m)	PHASE 1 to PHASE 2 change (%)
Mean absolute bias across all sites (m)	1.28	1.22	-0.06	-4.87
Mean dispersion across all sites (m)	2.03	2.00	-0.03	-1.64
RMS of the mean bias and dispersion (m)	1.70	1.65	-0.04	-2.55

Table 9. Final summary of SPICE Phase 1 and Phase 2 pLRM processing scenarios for Antarctic study sites. Note that Russell Glacier is not included in this summary table so that statistics are directly comparable to the SAR results presented in Table 7.

Finally, we can compare the Phase 2 SAR and pLRM summary statistics (Table 7 and Table 9), to provide a first assessment of the relative performance of the two modes of operation. According to all 3 summary metrics, our Phase 2 SAR configuration provides elevation measurements that are more closely aligned to the reference data, as compared to our pLRM measurements. The average improvement offered by SAR is ~ 40% - 50%, both in terms of reducing the overall bias and the dispersion of the elevation differences.

6. Summary & Conclusions

This report has detailed the validation activities of the SPICE study, which were conducted within WP4. These validation activities had two principle aims:

(1) To provide a quantitative and objectives means to inter-compare the different SPICE processing configurations developed within WP2 and WP3.
(2) To determine the extent to which the optimal SPICE configuration, as established during (1), improves upon the existing state-of-the-art SAR products, as defined by the ESA L2 and SPICE Phase 1 processing.

Summarising the results of our inter-comparison, we find that for the high resolution SAR chain, the optimal processing configuration involves utilising a zero-padding factor of 4 in the L1 processing, a pre-retracking module based on a DEM, and a Threshold Peak Retracker. This configuration produces elevations that are more closely aligned with our validation datasets than alternative L1 (zero-padding factor of 2, Hamming weighting, multi-looking without zeros) and L2 (Batch processing pre-retracking module, Analytical Retracker) options. For the low resolution pLRM chain, the best performing option uses a zero-padding factor of 2, the Batch pre-retracking module and a Threshold Centre of Gravity Retracker, although the differences between the different configurations are typically small. Based upon this inter-comparison, these optimal SAR and pLRM choices form our final Phase 2 processing chain.

Turning next to the comparison between Phase 1 and Phase 2 datasets, we find that our SAR Phase 2 configuration generally improves the accuracy of elevation measurements relative to Phase 1. The improvement is typically 40%-70%, depending upon the metric, and statistically significant at the 5% confidence level. The improvement between Phase 1 and Phase 2 is greatest at the ice margin sites of Spirit and Russell, where the coastal ice topography is steeper and more irregular, leading to a higher proportion of complex multi-peaked waveforms. In contrast, the changes between the Phase 1 and Phase 2 pLRM configurations are much smaller, and almost entirely statistically insignificant. We believe that the reason for these differences is the different characteristics of the SAR and pLRM waveforms, due to the different surface area covered by each range cell. It is well-established that the narrower along-track beam width of the SAR mode of operation produces more sharply-peaked waveforms, and is therefore more likely to lead to a greater number of multi-peaked SAR waveforms. These multi-peaked waveforms are more amenable to the Batch and DEM pre-retracking modules developed within our Phase 2 activities. A worthwhile future activity would be to look more widely at SAR and pLRM ice sheet waveform morphology beyond the SPICE study regions, to systematically assess the extent to which they differ in their shape and characteristics across the continent. Sentinel-3, which provides SAR and pLRM L2 routinely as part of its ground segment, offers a ready opportunity to do this.

In terms of absolute accuracy, the analysis conducted within this WP suggests the following. For our final Phase 2 dataset, the bias of SAR measurements relative to the reference data is typically of the order 0.1 – 1 metre. Across the 4 study sites, there is no clear relationship between the bias and the topographic complexity of the region. The dispersion of the differences is on average ~1.2 metres, and of the order 0.1 – 1 metre at inland sites with relatively simple topography, increasing to 1 – 10 metres at more complex coastal sites. The increasing dispersion of the differences with topographic complexity reflects the challenges of reliably retracking multi-peaked waveforms, and of correctly locating the point of closest approach.

Across our study sites, the SPICE Phase 2 pLRM data has a bias generally of the order 0.1 – 1 metre, although the average of 1.2 metres is approximately double the corresponding SAR value of 0.64 m. Like SAR, there is again no clear relationship between the bias and the topographic complexity of the region. The dispersion of the differences is on average 2 metres, again roughly double the corresponding SAR statistic. As with our SAR data, the dispersion of the pLRM differences also tends to increase with topographic complexity, from the order of 0.1 – 1 metres at inland sites, to predominantly ~10 metres at more complex coastal sites.

In conclusion, therefore, our validation activities indicate that the novel algorithm development undertaken within the SPICE project has successfully improved the accuracy of the SAR altimetry measurements over ice sheets. Furthermore, our analysis provides the first independent error characterization of SAR and pLRM elevation measurements over ice sheets. The novel pre-retracking modules show particular promise to improve elevation retrievals in complex ice margin zones, and we believe should be further explored and developed in future studies. While the new SPICE pLRM processor that we have developed within the project cannot match the performance of its high resolution SAR counterpart, it does nonetheless achieve a level of accuracy that makes the data suitable for geophysical interpretation. As such, it represents a valuable demonstration and validation of the technique, which will be important for inter-calibrating historical low resolution and current high resolution satellite altimeter missions. Looking to the future, we hope that the validation activities reported here will act as a benchmark for assessments of Sentinel-3 SAR and pLRM acquisitions.

7. References

Arthern, R. J., Winebrenner, D. P., & Vaughan, D. G. (2006). Antarctic snow accumulation mapped using polarization of 4.3-cm wavelength microwave emission. *Journal of Geophysical Research-Atmospheres*, *111*(D6).

Fricker, H. A., Borsa, A., Minster, B., Carabajal, C., Quinn, K., & Bills, B. (2005). Assessment of ICESat performance at the Salar de Uyuni, Bolivia. *Geophysical Research Letters*, *32*(21). https://doi.org/10.1029/2005GL023423

Hollander, M. A., Wolfe, D., & Chicken, E. (2015). *Nonparametric Statistical Methods* (Third Edit). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119196037.ch1

Martin, C. F., Krabill, W. B., Manizade, S. S., Russell, R. L., Sonntag, J. G., Swift, R. N., & Yungel, J. K. (2012). Airborne Topographic Mapper Calibration Procedures and Accuracy Assessment. *NASA Technical Report NASA/TM\u20132012-215891, Goddard Space Flight Center, Greenbelt, Maryland 20771. Available Online at: Http://Ntrs.Nasa.Gov/Archive/Nasa/Casi.Ntrs.Nasa.Gov/20120008479.Pdf.*

Massey, F. J. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, *46*(253).

McMillan, M., Shepherd, A., Sundal, A., Briggs, K., Muir, A., Ridout, A., et al. (2014). Increased ice losses from Antarctica detected by CryoSat-2. *Geophysical Research Letters*, *41*(11). https://doi.org/10.1002/2014GL060111

Roemer, S., Legrésy, B., Horwath, M., & Dietrich, R. (2007). Refined analysis of radar altimetry data applied to the region of the subglacial Lake Vostok/Antarctica. *Remote Sensing of Environment*, *106*(3), 269–284. https://doi.org/10.1016/j.rse.2006.02.026

Slater, T., Shepherd, A., McMillan, M., Muir, A., Gilbert, L., Hogg, A. E., et al. (2018). A new digital elevation model of Antarctica derived from CryoSat-2 altimetry. *The Cryosphere*, *12*(4), 1551–1562. https://doi.org/10.5194/tc-12-1551-2018

8. Annex 1. Echo Relocation Methodology

The echo relocation step is used during the validation process to adjust the altimeter elevation measurements (the output of WP3) to the Point of Closest Approach, prior to comparison to the airborne data. This is an important step in the validation procedure because it provides a more consistent inter-comparison between the two datasets, which takes into account the different sizes of the radar and laser beams. It is important to note, however, that this step is not required when our new DEM pre-retracking module is invoked, because this module uses an additional step to identify the nadir return within the beam footprint, and therefore circumvents the need for relocation. It is, however needed for the validation of our conventional processing and Batch pre-retracking options.

Within this study, we use an established procedure for echo relocation, based upon the methodology of Roemer *et al.* (2007). This method has been shown to outperform conventional *slope-correction* methods, because it accounts for non-linear surface topography within the beam footprint (Roemer *et al.*, 2007). The processing steps of the method are summarised in Figure 9. This method uses an auxiliary Digital Elevation Model (DEM) to identify the point on the ice surface illuminated by the radar beam that is closest to the satellite, the so-called *Point of Closest Approach*. In this study, we use a DEM based upon 7 years of CryoSat-2 data (Slater et al., 2018) for this purpose. Having identified the Point of Closest Approach, the altimeter elevation measurement is then translated from its nadir ground track position to this new location. In practical terms, this means calculating and adjusting the latitude, longitude and elevation of the original measurement, based upon the coordinates of the identified Point of Closest Approach. A summary of the procedure is given in Figure 9, and a full description of the method can be found within Roemer *et al.* (2007).

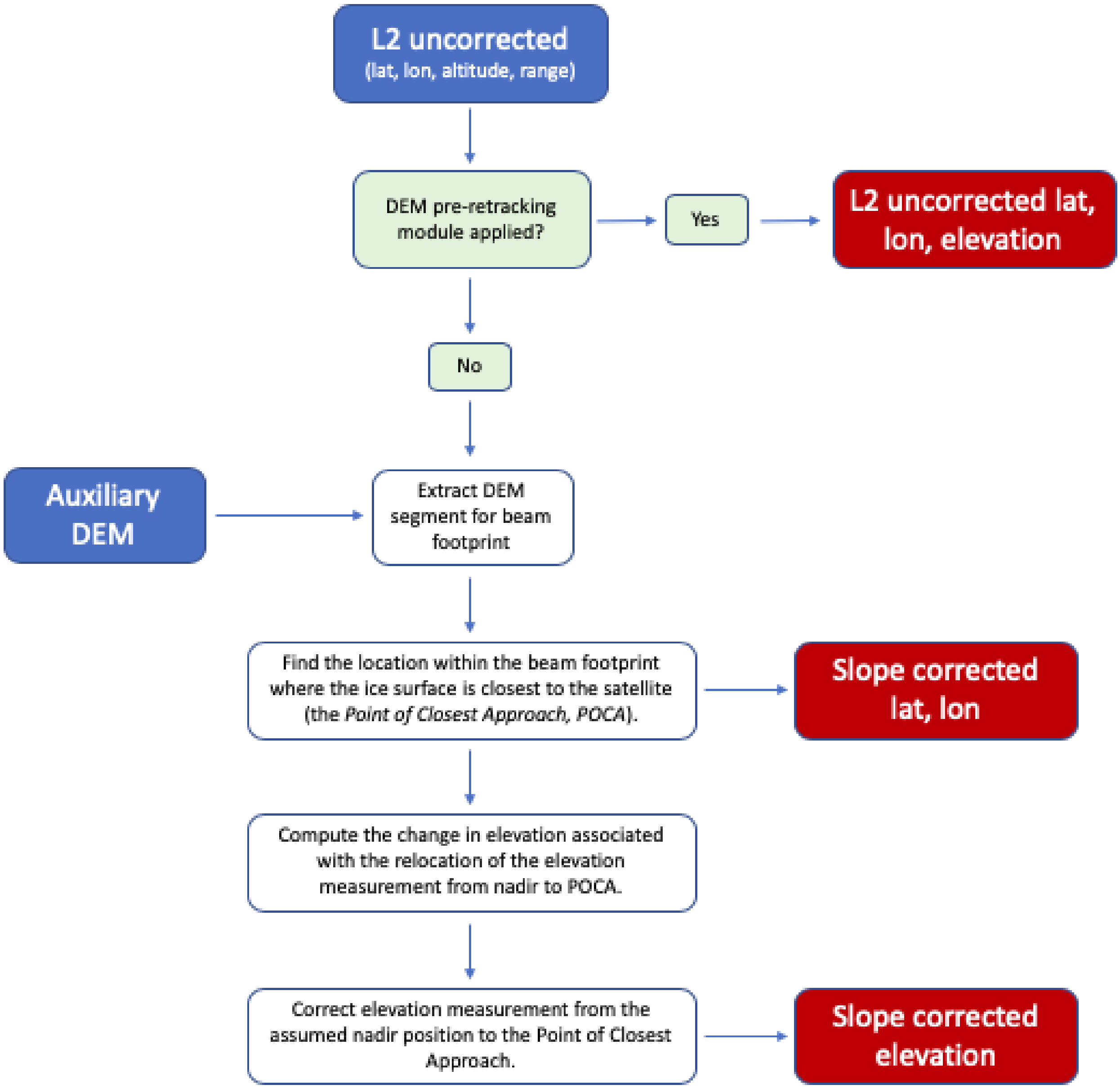


Figure 9. Block diagram showing the input data, processing steps, and output fields for the echo relocation. Blue blocks indicate input data, green blocks indicate decisions, white blocks indicate processing steps, and red blocks indicate output fields. Full details of the method are given in Roemer *et al*. (2007).