Pair Selection Optimization for InSAR Time Series Processing

Delphine Smittarello[1], Nicolas F. d'Oreye[2], Maxime Jaspard[1], Dominique Derauw[3], and Sergey V Samsonov[4]

[1]European Center for Geodynamics and Seismology
[2]National Museum of Natural History of Luxembourg and European Center for Geodynamics and Seismology
[3]Centre Spatial de Liège and Universidad Nacional de Rio Negro
[4]Natural Resources Canada

November 23, 2022

Abstract

The ever-increasing amount of SAR data motivates the development of automatic processing chains to fully exploit the opportunities offered by these large databases.
The InSAR Mass processing Toolbox for Multidimensional time series (MasTer) is an optimized tool to automatically download SAR data, select the interferometric pairs, perform the interferometric mass processing, compute the geocoded deformation maps, invert and display the velocity maps and the 2D time series on a web page updated incrementally as soon as a new image is available.
New challenges relate to data management and processing load. We address them through methodological improvements dedicated to optimizing the InSAR pair selection.
The proposed algorithm narrows the classical selection based on the shortest temporal and spatial baselines thanks to a coherence proxy and balances the use of each image as Primary and Secondary images thanks to graph theory methods.
We apply the processing to three volcanic areas characterized with different climate, vegetation and deformation characteristics: the Virunga Volcanic Province (DR Congo), the Reunion Island (France) and the Domuyo and Laguna del Maule area (Chile-Argentina border).
Compared to pair selection based solely on baseline criteria, this new tool produces similar velocity maps while reducing the total number of computed differential InSAR interferograms by up to 75\%, which drastically reduces the computation time.
The optimization also allows to reduce the influence of DEM errors and atmospheric phase screen, which increase the signal-to-noise ratio of the inverted displacement time series.

Pair Selection Optimization for InSAR Time Series Processing

D. Smittarello[1], N. d'Oreye[1,2], M. Jaspard[1], D. Derauw[1,3,4], S. Samsonov[5]

[1]European Center for Geodynamics and Seismology, 19 rue Josy Welter, L-7256 Walferdange, Gd Duchy of Luxembourg
[2]National Museum of Natural History, Department of Geophysics/Astrophysics, Rue Josy Welter 19, L-7256 Walferdange, Grand-Duchy of Luxembourg
[3]Laboratorio de Estudio y Seguimiento de Volcanes Activos (LESVA),IIPG-Universidad Nacional de Rio Negro - CONICET, 1242 Av. General Roca, 8332, General Roca, Argentina
[4]Centre Spatial de Liège (CSL), Avenue du Pré Aily, B-4031 Angleur, Belgium
[5]Canada Center for Mapping and Earth Observation, Natural Resources Canada (NRCAN), 560 Rochester Street, Ottawa, ON K1A 0E4, Canada

Key Points:

- Proposed optimized pair selection improves InSAR time series quality while reducing computation time and memory requirements
- Proposed coherence proxy allows pair and image rejection without computing the interferograms
- Algorithm written in Python is used for automatic processing with MasTer toolbox and applicable to other time series software

Corresponding author: Delphine Smittarello, `smittare@phare.normalesup.org`

Abstract

The ever-increasing amount of SAR data motivates the development of automatic processing chains to fully exploit the opportunities offered by these large databases. The InSAR Mass processing Toolbox for Multidimensional time series (MasTer) is an optimized tool to automatically download SAR data, select the interferometric pairs, perform the interferometric mass processing, compute the geocoded deformation maps, invert and display the velocity maps and the 2D time series on a web page updated incrementally as soon as a new image is available. New challenges relate to data management and processing load. We address them through methodological improvements dedicated to optimizing the InSAR pair selection. The proposed algorithm narrows the classical selection based on the shortest temporal and spatial baselines thanks to a coherence proxy and balances the use of each image as Primary and Secondary images thanks to graph theory methods. We apply the processing to three volcanic areas characterized with different climate, vegetation and deformation characteristics: the Virunga Volcanic Province (DR Congo), the Reunion Island (France) and the Domuyo and Laguna del Maule area (Chile-Argentina border). Compared to pair selection based solely on baseline criteria, this new tool produces similar velocity maps while reducing the total number of computed differential InSAR interferograms by up to 75%, which drastically reduces the computation time. The optimization also allows to reduce the influence of DEM errors and atmospheric phase screen, which increase the signal-to-noise ratio of the inverted displacement time series.

Plain Language Summary

Development of satellite remote sensing greatly helps to mitigate natural hazard in remote or dangerous areas like volcano-tectonic regions or landslide-prone regions. In particular, Synthetic Aperture Radar Interferometry (InSAR) offers the possibility to measure ground surface displacements with millimeter resolution. Several methods exist to benefit from the large amount of data to perform time series of ground deformation with sub-centimeter resolution. However, the ever-increasing number of available images poses new challenges (e.g. to process the large amount of data, to manage large databases and to extract useful information in near-real time for operative purposes). MasTer is a fully automatic tool able to provide updated velocity maps and displacement time series resulting from the processing of satellites radar images, which are regularly acquired by space agencies. Hereby, we present a methodological development to speed up the processing and improve the signal-to-noise ratio of the obtained ground deformation time series. This is achieved by optimizing the InSAR pair selection. By also reducing the storage space and raw-memory requirements, it allows processing longer time series with the same computational infrastructure. The proposed algorithm, written in Python, is included in the MasTer toolbox, though it can easily be adapted for other time series software.

1 Introduction

Classic differential Radar Interferometry (InSAR) processing consists in estimating the ground surface displacements by measuring the phase offsets between two Synthetic Aperture Radar (SAR) signals acquired with the same imaging geometry at two different epochs (Massonnet & Feigl, 1998). Since the first interferogram computed in the 1990's, the diversity of SAR sensors and the amount of SAR data available has increased exponentially, which motivated the development of time series methods such as Persistent Scatterer Interferometry (PSI) (Ferretti et al., 2000, 2001) and Small Baseline Subset (SBAS) (Berardino et al., 2002).

PSI-like methods measure the displacement at stable point scatterers (usually man-made structures). To identify only the stable scatterers, such methods consider interferometric pairs with large perpendicular baselines. On the contrary, SBAS-like methods consist in selecting interferometric pairs satisfying small spatial and temporal baselines to derive a time series of displacement estimated on a given time interval. The small baseline criterion aims at optimizing the coherence on surfaces made of uniform scatterer distributions. To exploit the very large amount of data provided by space agencies, semi-automatic or automatic time series of ground deformation methods are developed either using PSI or SBAS techniques, or both, such as LiCSBAS (Morishita et al., 2020), MintPy (Yunjun et al., 2019), SNAP-StaMPS (Delgado Blasco et al., 2019), P-SBAS (Elefante et al., 2013; De Luca et al., 2015), or SqueeSAR (Ferretti et al., 2011).

Here we use MasTer toolbox, which is a fully automatic, unsupervised processing chain (Derauw et al., 2020; d'Oreye et al., 2021) based on the Multidimensional Small Baseline Subset (MSBAS) method (Samsonov & d'Oreye, 2012, 2017; Samsonov et al., 2017, 2020), which is a 2D/3D extension of the SBAS method. Thanks to a set of shell scripts, MasTer allows the automation from the Single Look Complex (SLC) images download up to the 2D decomposition of ground deformation time series (vertical and horizontal). Interferograms and deformation maps are computed using the MasTer Engine, which is an extension of the CSL InSAR Suite (CIS) Software (Derauw, 1999).

Considering N successive SAR acquisitions acquired in a given geometry, it is theoretically possible to form $\frac{N(N-1)}{2}$ interferometric pairs. However, as revisit time decreases (e.g. thanks to the use of SAR satellite constellations), or simply because N increases, computing all the pairs theoretically available quickly becomes a time-consuming and computationally heavy task. Moreover, when the spatial and temporal baselines increase, many low coherence pairs are useless. Seasonal effects may also affect numerous pairs. Tao et al. (2018) show that the quality and quantity of multi-temporal differential interferograms used to produce a deformation time series using the SBAS-based StaMPS method affect its accuracy, and that processing a larger number of interferograms does not always provide better results.

It is hence of prime importance to select the optimal list of interferometric pairs. The challenge is 1) to minimize the total number of interferograms to compute in order to maximize the processing efficiency, 2) to keep only the best-quality interferograms to improve the accuracy of the deformation measurement. Classical SBAS-like methods aim to retrieve the temporal evolution of the ground surface deformation from a selection of pairs based on a critical value of the perpendicular baseline in order to minimize the coherence loss due to spatial decorrelation. Additionally, the short temporal baseline selection minimizes the effect of the temporal decorrelation. However, results from such a selection might be degraded by possible seasonal fluctuation of coherence (e.g. rain, snow,...).

Several solutions were proposed to improve the pair selection. Yang et al. (2012); Pepe et al. (2015) propose an algorithm relying on Simulated Annealing to select a Delaunay triangulation in the temporal/perpendicular baseline plane that maximize a cost function based on the coherence values of interferometric pairs. The semi-automatic selection of optimum image pairs (SASOIP) method (Zhang et al., 2018; Wu et al., 2019) evaluates the coherence of point targets in a small feature region as a criterion to increase the quality of the selected pairs. Using also a coherence threshold criterion, Ishitsuka et al. (2016) propose to determine the optimal baseline on a pixel-by-pixel basis. This enables the use of a greater number of interferometric pairs for highly coherent pixels and a minimal number of interferometric pairs for noisy pixels. Such a procedure increases the number of pixels available for surface displacement mapping.

However, none of these methods is optimal to increase both the processing efficiency and the accuracy of the time series. They require the computation of the coherence map for every pair, including those that will not be selected, and/or they do not ensure that the selection avoids splitting the data into several subsets and creating gaps in the time-baseline plot.

In this study, we present a new algorithm that aims at optimizing the pair selection by limiting the total number of interferograms to be computed, restricting the processing to the interferograms with the best quality without decreasing the deformation time series accuracy, and by balancing the usage of each image as Primary and Secondary images to improve the signal-to-noise ratio. The algorithm is tested by comparing several inversions performed using the MSBAS method within the MasTer toolbox, though it can easily be adapted for other time series software.

The proposed method considers the time-baseline plot as a directed graph with the nodes representing the SAR acquisition dates and the arcs representing interferometric pairs. Using this formalism, we developed an algorithm to extract an optimized subgraph. In order to drive the pair selection, we weight the arcs using a coherence proxy based on a function of the spatial and temporal baselines. Such a proxy avoids computing the coherence map of all possible interferograms as it can be calibrated by computing the coherence maps for a subset of pairs (ideally longer that a seasonal cycle).

We tested the algorithm on three volcanic areas : the Domuyo-Laguna del Maule (DLM) area (Chile-Argentina boundary), the Reunion Island (France) and the Virunga Volcanic Province (VVP) area (Democratic Republic of Congo) (Fig. 1). It is important to note that the maximum temporal and spatial baselines (B_T and B_P respectively) used to build the time-baseline graphs to be optimized in these 3 cases studies must be considered as for demonstration purpose. They must not be taken as default values for all studies. Appropriate values of baselines must be chosen with care and depend on the type of target (ground cover and seasonal conditions), the type of expected signal to be measured, the satellite orbital characteristics, the potential errors such as the "fading signal" (Ansari et al., 2020), the computer resources available etc. It is the responsibility of the user to chose these appropriate baselines values carefully taking into account these criteria while maintaining as much as possible a gap-less time-baseline graph. We remind in the Supporting Information (Text S1) some basic concepts that could assist the user to select these baselines values.

2 Geological settings and InSAR data availability of 3 test-sites

The three areas are different in term of climate, expected signal of deformation, backscattering properties and temporal decorrelation characteristics. We mainly used data from the Sentinel–1 constellation available in the three areas. On the VVP, we also compare the algorithm performance with additional data from RADARSAT and COSMO-SkyMed satellites.

2.1 Domuyo and Laguna del Maule (Chili-Argentina)

Laguna del Maule (LDM) volcanic complex (Talca province, Chile) and Domuyo volcano (Neuquén province, Argentina) are located in the Southern Andes. On both volcanoes, large rates of surface uplift were recently detected by satellite geodetic measurements : up to 30 cm/yr on Laguna del Maule (Fournier et al., 2010; Feigl et al., 2014; Singer et al., 2014; Le Mével et al., 2015) and 15 cm/yr in LOS on Domuyo (Astort et al., 2019; Lundgren et al., 2020; Derauw et al., 2020). Although it is mostly a desert area, Fig. 2 show that the excellent InSAR coherence quickly decreases when the time delay between acquisitions increases. It also clearly reveals a seasonal effect with strong decorrelation during austral winter (June-Aug.) because of the important snow falls affecting these high altitude regions of the Andes while images from summer to summer recover a satisfying coherence level. Sentinel–1 acquires SAR images on a regular basis since October 2014. In this study we use 340 images acquired in IW mode along the ascend-

Table 1. InSAR data availability. The table contains, for each acquisition mode used in this study (col. 1), the total number of available images (N, col. 2), the time delay between two acquisitions (in days, col. 3), the period spanned by the data set (col.4), the temporal (B_T) and spatial (B_P) baseline criteria used in the original time series processing (col. 5) and the total number of interferometric pairs before (col. 6) and after optimization (col. 7).

Area Sat. Mode Orbit	Images (N)	δt (days)	Time span (yyyy-mm-dd:yyyy-mm-dd)	B_T/B_P ($days/m$)	Pairs B_T/B_P	Pairs B_T/B_P opt. 3 or opt. 3/4
Domuyo						
S1 IW A18 NT	140	24/12[a]	2014-10-30:2020-11-27	450/20	798	376/471
CT 0.24	99				231	181
MVCP 0.25	124				700	337
MVCP 0.30	96				385	240
MVCP 0.35	68				206	150
MVCP 0.40	59				174	124
S1 IW D83 NT	200	24/12/6[b]	2014-10-23:2020-11-26	450/20	1868	557/707
CT 0.24	153				782	358
MVCP 0.25	179				1718	500
MVCP 0.30	146				1219	400
MVCP 0.35	115				841	321
MVCP 0.40	95				625	261
Reunion						
S1 SM A144	125	12	2016-10-10:2020-10-30	50/50	291	275
S1 SM D151	119	12	2016-10-11:2020-12-01	50/50	241	235
S1 IW A144	122	12	2016-10-04:2020-11-24	70/70	412	334
S1 IW D151	120	12	2016-10-05:2020-11-25	70/70	385	313
VVP						
CSK A	459	4	2011-04-15:2020-05-21	200/200	2453	1248/1570
CSK D	516	4	2011-04-13:2020-05-18	200/200	2861	1439/1804
RS F2F D	76	72/24[c]	2010-03-28:2019-06-09	400/400	618	222/285
RS F21N D	42	24	2009-12-15:2014-04-17	400/400	277	122/156
RS UF A	77	24[d]	2012-03-03:2019-05-26	400/400	829	228/300
S1 IW A174	226	24/12/6[e]	2014-10-17:2020-12-02	400/20	2575	649/840
S1 IW D21	141	24/12[f]	2014-10-07:2020-11-22	400/20	826	383/481

Acronyms:
A = Ascending; D = Descending, SM = Strip Map; IW = Interferometric Wide Swath
S1 = Sentinel–1; CSK = CosmoSkyMed; RS = RADARSAT
F2F = Fine; F21N = Fine; UF = UltraFine
NT = No Threshold
CT = Coherence Threshold for pair rejection
MVCP = Minimum Value of Coherence Proxy for image rejection

[a] 24 days then 12 days since May 2017 then 6 days since Aug. 2020.
[b] 24 days then 12 days since Fev 2017 then 6 days since Dec. 2018.
[c] 72 days then 24 days since Dec. 2013.
[d] 24 days since Jul. 2013.
[e] 24 days then 12 days since Fev. 2017 then 6 days since Jun. 2018.
[f] 24 days then 12 days since Fev. 2017.

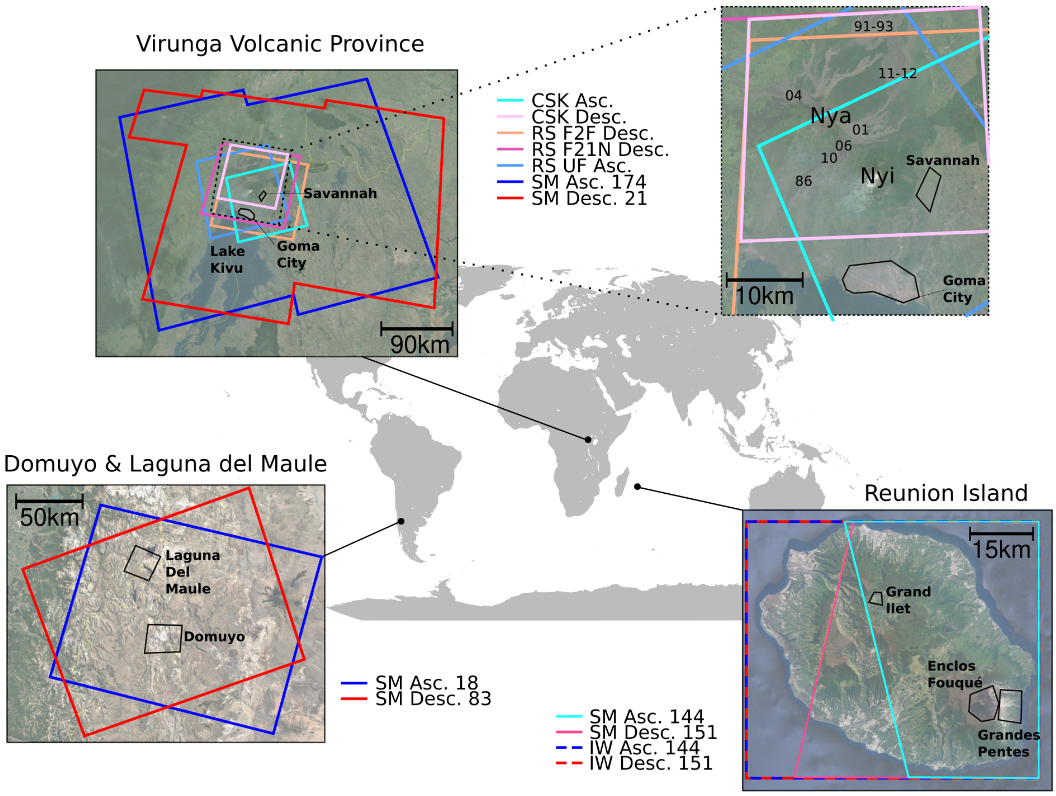

Figure 1. Location on a global view of the three studied areas : Reunion Island, Virunga Volcanic Province (VVP) and Domuyo and Laguna del Maule. InSAR footprints used for each acquisition mode are represented on Google Earth background. In addition, black polygons mark the regions of interest defined in each area and used for mean coherence estimates. Zoom in on the Nyamulagira (Nya) and Nyiragongo (Nyi) where numbers marks the lava flow emplacement year mentioned in main text and Fig. 4

ing orbit 18 and the descending orbit 83 (Table 1). Data span 6 years (2014/10/23–2020/11/27). Delay between two acquisitions has reduced from 24 days at the beginning of the mission to 12 days since February and May 2017 and to 6 days since December 2018 and August 2020 for descending and ascending orbits respectively.

2.2 La Reunion Island (France)

La Reunion is a French tropical island located in the Indian Ocean. This volcanic island is formed by the Piton des Neiges (PdN) Volcano to the northwest and by the Piton de la Fournaise (PdF) Volcano to the southeast. PdF volcano is a very active volcano with nearly one eruption every 10 months since 1985 (Roult et al., 2012). Its poorly vegetated lava field ensures a high level of coherence (Fig. S1), which offers favorable conditions for mapping the frequent and large co-eruptive displacements using InSAR related methods. PdN is a dormant and largely eroded volcano with three Circus : Cilaos, Mafate and Salazie. The tropical vegetation that covers large parts of the edifice quickly degrades the coherence (Fig. S1). Moreover, seasonal coherence fluctuation are related to the existence of rainy seasons common to the tropical climate. Several active landslides are identified in the Salazie Cirque (Delacourt et al., 2009; Raucoules et al., 2020). The Reunion island has been imaged by several SAR satellites since the early 2000's. Here we focus on the analysis of 479 Sentinel–1 SAR acquisitions spanning four years (2016/10/04–2020/11/01) and freely available in the frame of the European Copernicus Program. Images were acquired along the ascending orbit 144 and descending orbit 151 in both StripMap (SM) and Interferometric WideSwath (IW) modes (Table 1). Although each mode is revisited every 12 days, the alternation of acquisition in SM and IW modes leads to a revisit time of 6 days both in ascending and descending geometries. Acquisition of an ascending image precedes by ∼12 hours the descending acquisition. A dense GNSS network monitored by the Observatoire Volcanologique du Piton de la Fournaise (OVPF-IPGP) allows us to validate our InSAR deformation time series by comparing them to GNSS data from the PdF area.

2.3 The Virunga Volcanic Province (DRC)

The Virunga Volcanic Province (VVP) is located on the eastern border of the Democratic Republic of Congo, in the Kivu Basin region. Two active volcanoes Nyiragongo and Nyamulagira, both hosting an active lava lake at the time of writing, stand in the VVP at only a few kilometers from the cities of Goma (in DR Congo) and Gisenyi (in Rwanda). The population of that region is rapidly growing and reached ∼ 1.1M inhabitants in 2017 (Mairie de Goma, 2017; *PopulationData.net*, 2017). Because large parts of the VVP remains practically inaccessible due to the dense equatorial jungle and recurrent armed conflicts, InSAR is a very important tool to complement the permanent GNSS (Geirsson et al., 2017; Ji et al., 2017) and seismic (Oth et al., 2017) ground-based monitoring networks maintained by the Goma Volcano Observatory (GVO). However, the equatorial vegetation limits the coherence to mostly the recent bare lava flows and the urban areas. Here also, seasonal coherence fluctuations are related to the existence of rainy seasons. In this study we use 1565 SAR acquisitions from CosmoSkyMed (CSK), RADARSAT (RS) and Sentinel–1 (S1) satellites, spanning 11 years in total (2009/12/15–2020/11/22) (Table 1). Delay between acquisitions has progressively evolved from 72 days with the first RS acquisitions to 24, 12 and 6 days with S1 and even 4 days with CSK.

3 Methods

3.1 MasTer: a fully automatic processing chain

The MasTer toolbox consists of a set of shell scripts that coordinate the full automation of differential InSAR mass processing and feeding of an MSBAS processor that generates displacement time series in LOS and/or vertical and horizontal (east-west) directions. The whole processing chain, from SLC data downloading to results displaying in a dedicated web page incrementally updated for every new image (or updated orbit), is optimized and human-supervision free. In the following, we briefly describe the relevant steps and refer to Derauw et al. (2020); d'Oreye et al. (2021) for a detailed overview of the possibilities offered by MasTer. See also Supporting Information Text S2 for possible strategies when implementing the proposed optimization in an automatic mass processing chain.

3.1.1 InSAR Processing

Starting from a list of SLC images, MasTer toolbox aims at selecting a list of pairs for interferogram computation. The original selection is done using classic spatial (B_P) and temporal (B_T) baseline threshold criteria. Through a mass processing step, all interferograms from the list are computed using the MasTer Engine, a command line InSAR processor and tools written in C derived from the CSL InSAR Suite software (CIS) (Derauw, 1999). Interferometric products (amplitude, coherence, differential interferometric phase, unwrapped phase and deformation maps) are computed in a global-primary SAR geometry. The interferograms are filtered (Goldstein & Werner, 1998), unwrapped with SNAPHU (Chen & Zebker, 2002), interpolated to fill empty isolated pixel gaps surrounded by unwrapped values and deramped. At last, InSAR products are geocoded on a predefined grid similar for all satellites geometries (Derauw et al., 2020).

3.1.2 MSBAS

Deformation maps computed from all the selected interferometric pairs feed the Multidimensional Small Baseline Subset (MSBAS) processor (Samsonov & d'Oreye, 2012, 2017; Samsonov et al., 2020). MSBAS is an extension of the Small BAseline Subset method (Berardino et al., 2002) written in C++. Unlike SBAS, which aims to invert displacements acquired in a given imaging geometry along the LOS, MSBAS inverts simultaneously several data sets (i.e. different satellites and/or different LOS), to obtain east-west (EW) and vertical cumulative deformation as well as linear velocity maps. From these results, one can extract EW and vertical displacement time series for each pixel that remained coherent along the whole time span and in each acquisition geometry. Nowadays, SAR-satellite data acquisition is done following side-view geometry along sub-polar orbits, making InSAR measurements poorly sensitive to North-South displacements. The inversion of LOS displacements is thus generally restricted to the vertical and East-West components. The possibility to retrieve the full 3D decomposition is limited to some specific cases, such as slow-moving landslides or glacier flows, when the degree of freedom can be reduced by assuming that the displacements occur parallel to the surface (Samsonov et al., 2020) or in a few occasion when a large number of acquisitions from different looking-angle are available (Peltier et al., 2017). However, because this is restricted to rare applications, we deal here with the inversion in 2D corresponding to the vertical and East-West directions. This approximation is reasonable as long as the displacement in the North-South component is not significantly larger than the East-West nor the Vertical components (Samsonov & d'Oreye, 2012; Nobile et al., 2018).

Another important aspect of the MSBAS method is that it inverts all the deformation maps provided by an external InSAR processing. Because the InSAR processing is the most time-consuming task, this MSBAS capability makes it especially suitable for an incremental usage: as soon as a new image is added to the database, only the interferograms from the pairs of interest formed with that new image must be computed. These new deformation maps are then added to the previously computed deformation maps feeding the MSBAS processor. MSBAS inversion takes only few minutes to hours depending on the number of deformation maps and their spatial extension.

3.1.3 MSBAS with a coherence threshold

After InSAR Mass Processing but before MSBAS inversion, an optional automated procedure is available in MasTer to reject the interferometric pairs that do not satisfy a coherence test. This MasTer functionality allows to automatically reject the interferometric pairs affected by a mean coherence measured lower than a given threshold over a reference region. We use this option at DLM. Indeed, based on a comparison with published GNSS information, Derauw et al. (2020) show that the deformation at DLM is measured more accurately while considering a MSBAS processing where a coherence restriction is applied (CT processing). For all the pairs satisfying the B_P and B_T criteria, the average coherence is computed on a square area (see LDM region kml footprint in Fig. 1) known to be prone to snow cover in winter season. The processing discard from the MSBAS inversion all the pairs affected by a mean coherence computed on that reference region lower than a given threshold. In our case, selecting 0.24 as the threshold on the average coherence ensure discarding pairs with no or poor interferometric signal on the area of interest. This concerns typically pairs including images acquired during the austral winter.

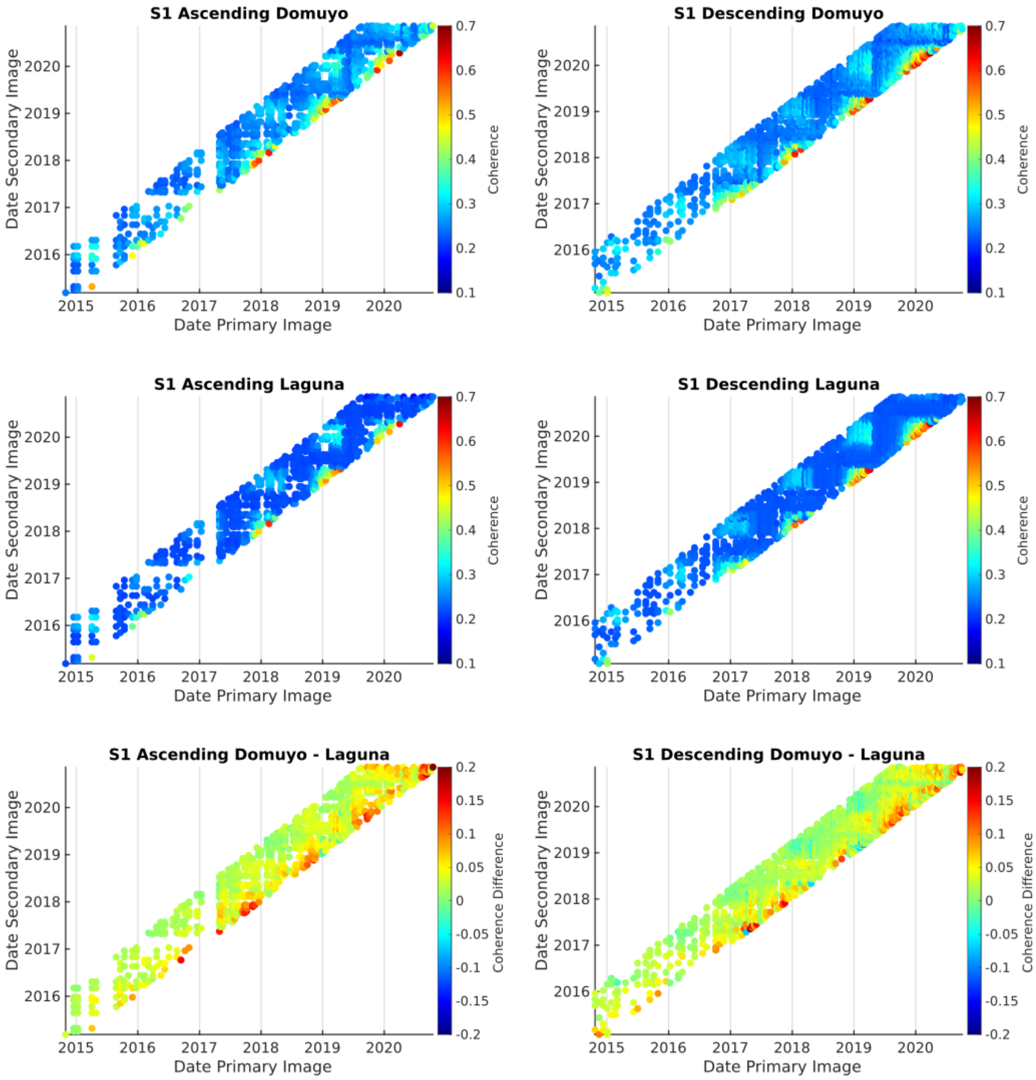

Figure 2. Mean coherence of Sentinel 1 ascending (left column) and descending (right column) interferograms as a function of the acquisition dates of the Primary and Secondary images. The mean coherence is computed either in the Domuyo area (first row) and at Laguna del Maule (second row). The third row represents the difference between first and second rows.

3.2 Coherence weighted graph for pairs selection optimization

In order to reduce the time and resources needed for the processing, special attention should be paid to the selection of interferometric pairs to be computed. A manual selection is no longer an option due to the amount of SAR data available. In the following, we first define the coherence proxy then explain how the selection algorithm makes use of this proxy to restrict the total number of pairs to process.

3.2.1 Coherence proxy definition

Distribution of coherence values being Gaussian, one could use a mean coherence value computed on a Region of Interest (ROI) to give a weight to each interferogram. However, computing the coherence for all the pairs is very expensive in terms of time and computer resources. A more efficient way is to define an easy-to-compute proxy of the coherence w for each pair of images. Note that if the distribution is not Gaussian, better statistical parameters describing the overall quality (such as the median, geometric mean or mode or mean value of the particular percentile) can be preferred to calibrate the coherence proxy.

The multitemporal coherence matrix can be used to characterize the dynamics of the coherence computed over a target area (Jacob et al., 2020). Temporal decorrelation and seasonal effects are the most important processes to be taken into account by the proxy, although it also includes the possibility to deal with spatial decorrelation by limiting the length of the spatial baselines. Thus, we defined the coherence proxy w as the weighted sum of three contributions (equation 1):

$$w = aw_1 + bw_2 + cw_3 \tag{1}$$

where w_1 accounts for the seasonal effect, w_2 accounts for the temporal decorrelation, w_3 accounts for the spatial decorrelation and a, b and c are the weighting coefficients whose determination is further explained below.

We define the seasonal contribution w_1 as:

$$w_1 = \left| \sin\left(\frac{DOY_P + (365 - DOY_{low})}{365} \times \pi\right) \times \sin\left(\frac{DOY_S + (365 - DOY_{low})}{365} \times \pi\right) \right|^{\alpha} \tag{2}$$

where DOY_P and DOY_S represent the Primary and Secondary image day of year, DOY_{low} and $\alpha \in [1;5]$ are calibration factors to be fixed by users. DOY_{low} in days represents the epoch of the year when coherence is the lowest and α accounts for the width of this low coherence period. Fig. S2-5 illustrate the behavior of w_1 when DOY_P, DOY_S, DOY_{low} and α vary.

To take into account the temporal and spatial decorrelation (Libert, 2018), w_2 and w_3 are defined respectively as :

$$w_2 = (M_{xc} - M_{nc})e^{-\beta|B_T|} + M_{nc} \tag{3}$$

$$w_3 = (M_{xc} - M_{nc})e^{-\gamma|B_P|} + M_{nc} \tag{4}$$

where B_T and B_P are the temporal and perpendicular baselines of each interferogram. M_{xc} and M_{nc} are the maximum and minimum expected values for the mean coherence in studied area and $\beta \in]0;1]$ and $\gamma \in]0;1]$ are calibration factors to be fixed by users. They account for the temporal and spatial decorrelation rates in the studied area. Fig. S6 illustrate the behavior of w_2 and w_3 when β and γ vary.

Weighting coefficients a, b and c are determined for each studied area and each acquisition mode as follow. Considering a calibration set of p pairs and coh(i) the mean value of the coherence computed on a region of interest for each pair $i \in [\![1;p]\!]$, we define the column vector $Coh = (coh(i))_{i\in[\![1;p]\!]}$. For each pair i of the calibration set, we compute the three contribution $w_1(i)$, $w_2(i)$ and $w_3(i)$. We define the three columns vectors $W_k = (w_k(i))_{(i,k)\in[\![1;p]\!]\times[\![1;3]\!]}$. We normalize according to Eq. 5:

$$W_k^*(i) = M_{nc} + \frac{W_k(i) - \min(W_k)}{\max(W_k) - \min(W_k)}(M_{xc} - M_{nc}); i \in [\![1;p]\!] \tag{5}$$

and we define the p-by-3 matrix $W = \begin{pmatrix} W_1^* & W_2^* & W_3^* \end{pmatrix}$. Considering A such that:

$$WA = Coh \tag{6}$$

$$A = \begin{pmatrix} a \\ b \\ c \end{pmatrix} = (W^T W)^{-1} W^T Coh \tag{7}$$

Table 2. Coherence proxy calibration. The table contains, for each acquisition mode used in this study (col. 1) and the Region of Interest (ROI, col. 2), the values of the calibration factors (DOY_{low}, α, β, γ,M_{xc} and M_{nc}, see section 3.2.1 for definition) and the weighting coefficients (a,b and c). R is the correlation coefficient between coherence data (mean value computed on the ROI) and the coherence proxy for all pairs used for calibration.

Area Sat. Mode Orbit	**ROI**	$\mathbf{DOY_{low}}$ (day)	α	β (day^{-1})	γ (m^{-1})	$\mathbf{M_{xc}}$	$\mathbf{M_{nc}}$	**a**	**b**	**c**	**R**
Domuyo											
S1 IW A18	Dom	230	1	0.0125	0.02	0.70	0.23	0.31	0.39	0.01	0.808
S1 IW A18	Lag	230	1	0.0125	0.02	0.70	0.23	0.32	0.41	0.05	0.815
S1 IW D83	Dom	230	1	0.0125	0.02	0.73	0.24	0.32	0.42	0.01	0.816
S1 IW D83	Lag	230	1	0.0125	0.02	0.73	0.24	0.30	0.43	0.00	0.861
Reunion											
S1 SM A144	GI	30	1	0.025	0.01	0.31	0.22	0.41	0.35	0.14	0.725
S1 SM D151	GI	30	1	0.025	0.01	0.37	0.23	0.28	0.40	0.14	0.627
S1 IW A144	GI	30	1	0.025	0.01	0.39	0.31	0.32	0.42	0.16	0.713
S1 IW D151	GI	30	1	0.025	0.01	0.39	0.30	0.34	0.33	0.27	0.721
VVP											
CSK A	Sav.	1	3	0.050	0.006	0.45	0.05	0.14	0.38	0.06	0.697
CSK D	Sav.	1	3	0.050	0.006	0.51	0.07	0.13	0.44	0.15	0.684
RS F2F D	Sav.	1	3	0.025	0.006	0.26	0.07	0.16	0.44	0.16	0.766
RS F21N D	Sav.	1	3	0.025	0.006	0.39	0.14	0.24	0.35	0.23	0.647
RS UF A	Sav.	1	3	0.025	0.006	0.30	0.06	0.20	0.35	0.16	0.706
S1 IW A174	Sav.	1	3	0.050	0.006	0.55	0.13	0.07	0.33	0.18	0.687
S1 IW D21	Sav.	1	1	0.025	0.006	0.45	0.13	0.17	0.25	0.18	0.557

where a, b and c are the weighting coefficients. Once this proxy is calibrated with a set of pairs, it is possible to associate a weight to each new pair without computing the coherence.

Fig. 3 shows how the calibration factors DOY_{low}, α, β and γ are chosen to adjust w_1, w_2 and w_3 to coherence data for S1 IW Desc mode in the DLM area. Calibration of others areas and modes are shown in Fig. S7-18. For each mode, the weighting coefficients a, b and c are inverted and results are shown in Table 2. The correlation coefficient R between the coherence data used for calibration and the coherence proxy being close to 1 shows that our proxy is generally able to discriminate between low and high coherence interferograms.

3.2.2 A graph formalism

We represent the time-baseline plot as a weighted directed graph where each SAR acquisition date is a node. Each Primary/Secondary images combination draws an arc. Each arc is given a weight w that is a proxy of the coherence. The optimization algorithm aims to remove the poorest-quality pairs from the graph in areas where many arcs are drawn while keeping all the pairs in areas poorly-connected. The optimization criterion k is defined as the maximum in-degree and out-degree of each node in the optimized graph. For a given node, these values correspond to the number of arcs entering and leaving the node respectively. This strategy aims to limit the total number of pairs computed in the mass processing to save time and memory, increasing the computation efficiency. It also allows to balance the use of each image as Primary and Secondary image which may reduce the influence on the final deformation time series of atmosphere artifacts.

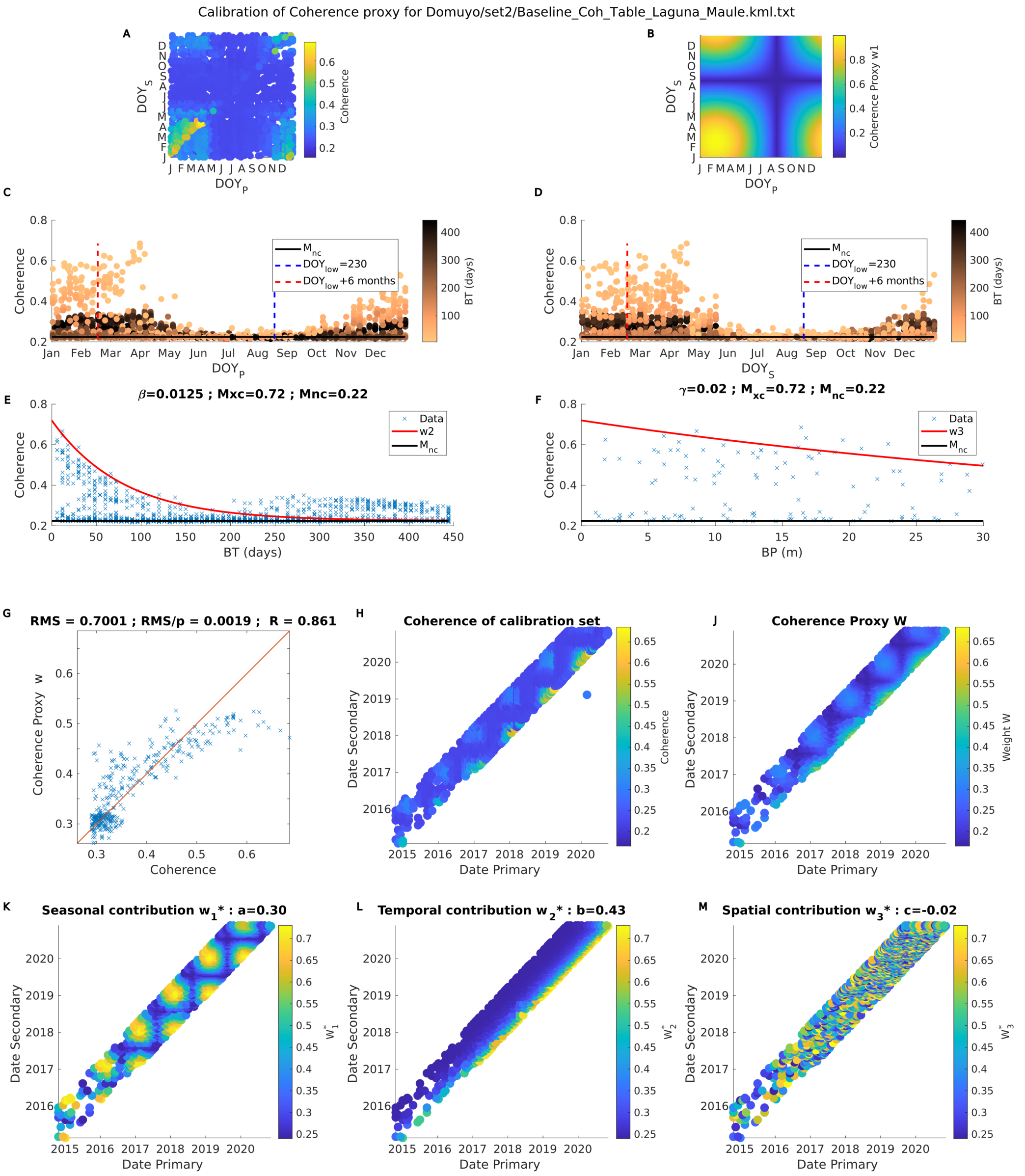

Figure 3. Calibration of the coherence proxy in the DLM area for S1 descending dataset (time span 2014-10-23:2020-11-26). A and B represent the mean coherence computed on the Laguna del Maule ROI and the seasonal contribution w_1 to the coherence proxy as a function of the primary and secondary image day of year of each pair. C and D represent the mean coherence computed on the Laguna del Maule ROI as a function of the Primary (Secondary) image Day of year respectively. Blue and red dash lines mark DOY_{low} and DOY_{low} + 6 months modulo 12 months, respectively. E and F represent the mean coherence computed on the Laguna del Maule ROI as a function of B_T for pairs with $B_P < 15$ m and as a function of B_P for pairs with $B_T < 25$ days, respectively. The red line represents the exponential decrease modeled by w_2 and w_3 in the coherence proxy. C to F, horizontal black line marks the minimum expected value of the coherence M_{nc}. G represents the coherence proxy w versus the mean coherence computed on the Laguna del Maule ROI for each pair of the calibration set. The red line marks the first bisector. H and J represent the mean coherence computed on the Laguna del Maule ROI for each pair of the calibration set and the coherence proxy w, respectively, as a function of the primary and secondary image acquisition date. K to M, the color scale represent the partial weights w_1, w_2 and w_3 (see text section 3.2.1), respectively. The value of the corresponding coefficients a, b, and c is indicated on top of each plot.

The algorithm starts with an initialization step followed by a straightforward process on all the nodes. During the initialization, the algorithm computes the in-degree and out-degree of each node. Then to each arc it associates a value v that is the number of arcs coming in to the same node, in other words the in-degree of the targeted node. Nodes are sorted in chronological order. In a for-loop, the algorithm verifies if the out-degree o of each node reaches the optimization criterion k. If not ($o \leq k$), it keeps all the arcs and steps to the following node. If yes ($o > k$), the algorithm clusters the out-arcs in two classes depending if v reaches the optimization criterion. All the arcs that don't satisfy the criterion ($v \leq k$) are kept. All others ($v > k$) are sorted by their weight. Then arcs with the smallest weights are removed until minimizing $|o-k|$. The values o and v are updated and the algorithm steps to the following node until the last one.

The optimization algorithm aims at selecting the most favorable pairs and limit the amount of interferograms to be processed by selecting a maximum number of $2k$ pairs for each image. Hence, if enough pairs are available, each image is taken k times as Primary and k times as Secondary image. Obviously, the lower k, the more restrictive the selection and hence the higher the gain in terms of computation load. Similarly, the larger the baseline criteria, the larger the number of pairs to process without optimization. Hence the efficiency of the optimization in terms of computation load depends on the initial baselines criteria and the choice of k. That means that with N images, the total number of pairs reaches a maximum of $(k \times N) - 2k$, where the subtrahend $2k$ takes into account the fact that the first and the last image of the data set can't be used as a Secondary and Primary image respectively. Without any pair selection criterion, the total number of pairs to compute with N images is $\frac{N(N-1)}{2}$ and considering a temporal baseline criterion (B_T), an order of magnitude of the number of possible pairs is $floor(\frac{B_T}{\delta t}) \times N$, where δt is the revisit time. Thus the optimization performance will increase with k much smaller than $floor(\frac{B_T}{\delta t})$.

This algorithm aims at keeping the graph connectivity and for a given image, even if all the pairs have a low coherence proxy value, the algorithm keeps $2k$ pairs. In order to force the rejection of these images before the InSAR Mass Processing, an additional option of the algorithm offers the possibility to reject some images if the coherence proxy doesn't reach a minimum value for at least one pair to form with that image. We call this the Minimum Value of Coherence Proxy for image rejection processing (MVCP processing). This option is tested on DLM for comparison with the coherence threshold restriction applied after InSAR Mass Processing.

4 Results

In this section, we look at results for the 3 targets then analyze the impact of our optimization.

4.1 Description of the displacements velocities measured with and without the optimization

Velocity maps reveal previously identified deformation patterns as well as more discrete features. On the VVP, the map reveals two main subsiding areas in the Nyamulagira volcanic field corresponding respectively to the 1991-1993 lava flow (Colclough, 2006, 2007; Toombs & Wadge, 2012) and the 2011-2012 lava flow (see profile BB' in Fig. 4)(Albino et al., 2015). A large area centered on the Nyamulagira crater also subsides at 10 mm.yr^{-1} and subsidence reaches locally 20 mm.yr^{-1} inside the Nyamulagira caldera. At last, local subsidence with a smaller amplitude (4-6 mm.yr^{-1}) is also noticeable at several locations in the lava flows formed by the 2004, 2006 and 2010 eruptions (Smets et al., 2010) (see profiles AA' and CC' in Fig. 4 and locations in Fig. 1). The DD' profile does not cross the Nyamulagira volcanic field. It rather extends across the Rift, from the city of Sake to the West up to the city of Goma and beyond to the east. A roughly 7 mm/yr subsidence is observed in a region undergoing significant magmatic CO2 degassing (Wauthier et al., 2018) located in the Southern tip of the uninhabited Virunga National Park (see DD' profile from 8.000 to 20.000 m in Fig. 4). Note that the DD' profile seems to be affected by offsets maybe related to residual plane or due to the fact that the zone is not connected to the volcanic part of the interferogram.

On Reunion Island, the velocity map reveals large ground displacement velocities on the Piton de la Fournaise volcano summit and its upper eastern flank (uplift 20cm/yr, eastward 28cm/yr). It maps also strong subsidence (15cm/yr) on the flanks where recent lava flows emplaced (Fig. 5). Those mean velocities do not reflect a continuous inflation but the cumulative effect of the 5 magma intrusions and 14 eruptions that occurred between October 2016 and November 2020 (*Communiqués OVPF*, 2016-2020). Each one of those events produced up to tens of cm of ground surface displacement in a few hours (Fig. 6). Outside of the volcano-related deformation detected on the poorly vegetated region of PdF, the MSBAS results reveals three other moving areas in the Salazie Cirque at the Piton des Neiges corresponding to three known landslides: Hell-Bourg (HB), Grand Ilet (GI) and Grand Ebouli (GE) (Delacourt et al., 2009; Belle et al., 2014; Raucoules et al., 2020).

In the DLM area, deforming areas identified in Fig. 7A are consistent with the previous results by Derauw et al. (2020). Circular patterns of uplift by several (tens of) cm are centered on Laguna del Maule and Domuyo volcano, respectively. Also, to the southeast of the Laguna del Maule Volcanic Complex, a frozen lava flow located between the Laguna Fea and the Laguna Negra subsides at about 2cm/yr (see profile AA' between 32 and 34 km on Fig. 7H) confirming the previous observation by Derauw et al. (2020).

4.2 Choice of optimization criterion k and efficiency

On the VVP area, we compare the velocity maps and some velocity profiles resulting from an MSBAS inversion with and without our optimized pair selection. The optimization of the pair selection is performed with $k = 3$ or $k = 4$ arcs entering and leaving each node (Fig. 4). Processing without optimization was performed using 3401 pairs. Optimized and not optimized processing produce very similar results nearly everywhere. Both optimizations also provided very similar results (see blue and green profiles in Fig. 4), although processing with $k = 3$ was performed with 1031 interferograms against 1320 for $k = 4$. Therefore, being more restrictive with $k = 3$ provides the best efficiency. See for instance the profiles AA' or CC' in Fig. 4 where the discrepancy is less than a mm/yr which is about 5 % of the maximum observed deformation. The main noticeable difference is observed along the BB' profile (Fig. 4) where the lava flow accumulated during the large volume eruption of 2011-2012 at Nyamulagira (Albino et al., 2015). The optimization reveals that the 2011-2012 lava flow compaction is about +4mm/yr faster than the value measured without the optimization (see section 5.2).

Fig. 5 confirms the good agreement between the processing without and with such an optimization in the high coherence region of the Piton de la Fournaise (see profiles AA' and BB' in Fig. 5), either for the combination of SM and IW acquisition modes or for each mode separately. Note, that velocity profile CC' crossing the Hell-Bourg landslide (Fig. 5) confirms the agreement between optimized and not optimized processing at least when IW and SM data are inverted together. However, contrary to what is observed on the Piton de la Fournaise area, velocities profile resulting from the processing of SM data only (purple on Fig. 5) or the IW data only (yellow Fig. 5) are slightly different. IW velocities appear underestimated in comparison to SM velocities (see section 5.3).

Fig. S19 and 8 compare respectively the baseline plots for La Reunion and for the VVP database. In the case of La Reunion, initial baseline criteria were 50 days and 50 m for images acquired in SM mode, and 70 days and 70 m for images acquired in IW mode.

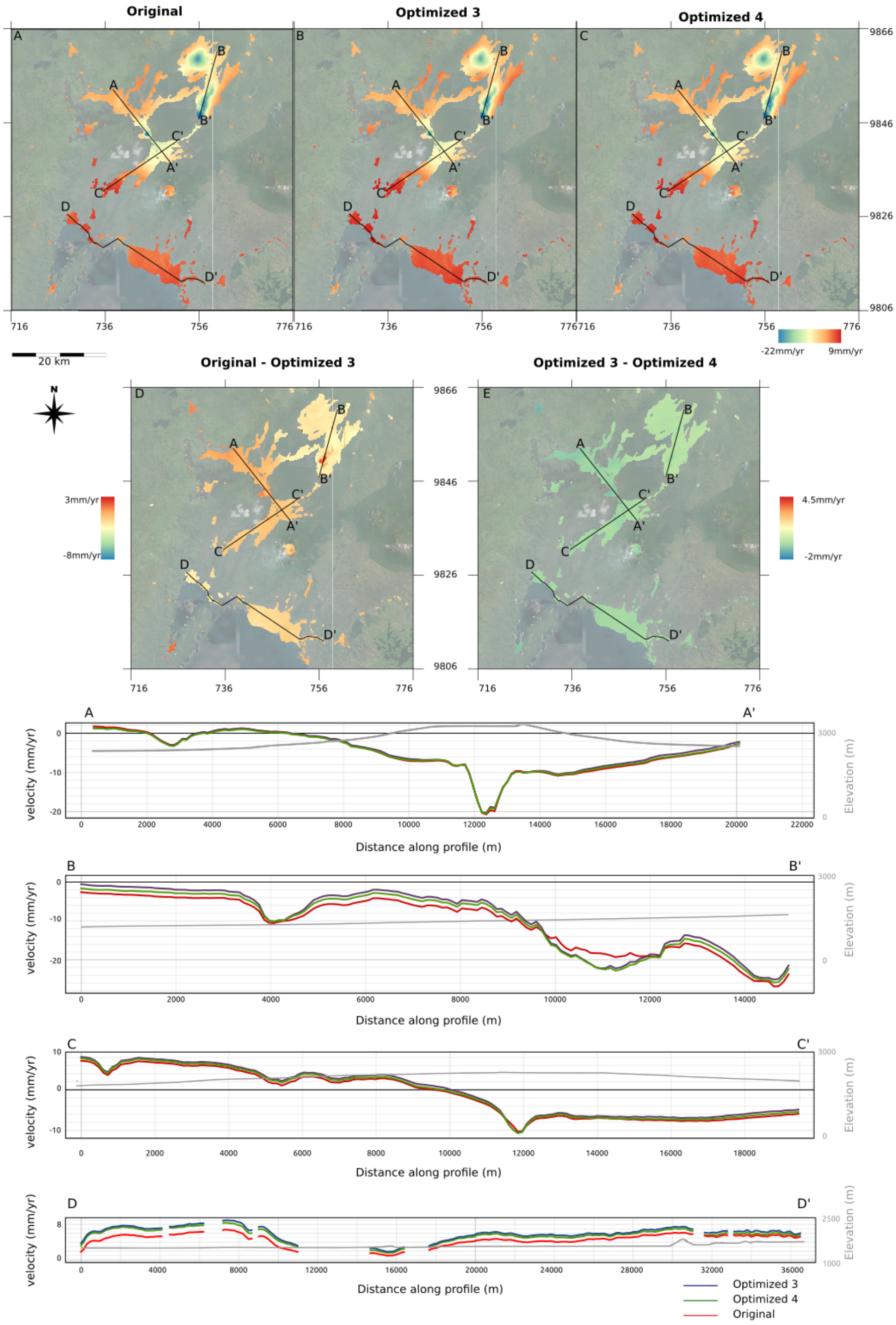

Figure 4. Vertical velocity on the VVP area from MSBAS inversion of Sentinel–1 ascending and descending data. A. Original processing without any optimization. B. and C. Optimized processing with $k = 3$ and $k = 4$, respectively. D. Difference between original (A.) and optimized (B.). E. Difference between optimized processing with $k = 3$ (B.) and $k = 4$ (C.). Velocity profiles AA', BB', CC' and DD' (bottom panel). Red, blue and green lines represent original and optimized with $k = 3$ and $k = 4$ processing results, grey lines show elevation profiles. - in AA' between 2.000 and 4.000 m = compaction of 2004 lava flow; between 11.000 and 13.000 m = subsidence in summit crater - in BB' between 3.000 and 5.000 m = compaction of 1967 and 1991-1993 lava flow; from 8.000 m to the B' = compaction of 2011-2012 lava flow - in CC' between 0 and 1.000 m = compaction of 1986 lava flow; from 11.000 m = compaction of 2006, 2010 and 2001 lava flows - in DD' between 8.000 and 20.000 m = subsidence in an areas undergoing significant magmatic CO_2 degassing (Wauthier et al., 2018).

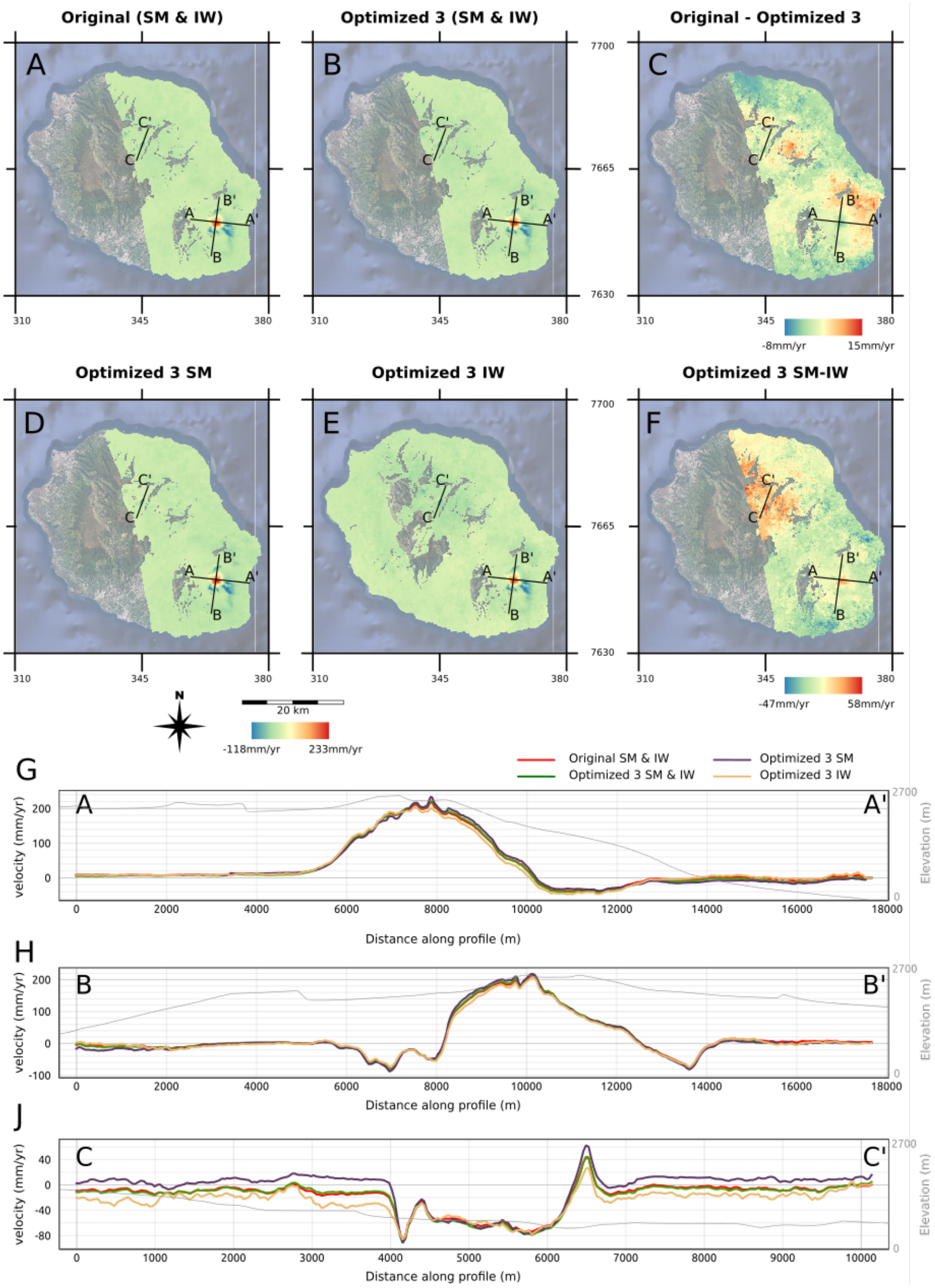

Figure 5. Vertical linear velocity on the Reunion Island area from MSBAS inversion of Sentinel–1 ascending and descending data in Stripmap (SM) and Wide Swath (IW) modes. Map A: Original processing without any optimization. Map B: Optimized processing with $k = 3$. Map C: Difference between original and optimized (Map A.- Map B.). Map D and E: Optimized processing with $k = 3$ restricted to acquisition in SM and IW modes respectively. Map F: Difference between optimized processing with $k = 3$ on SM acquisition mode and on IW acquisition mode (Map D. - Map E.). Graphs G, H and J (bottom panel) are velocity profiles AA', BB' on the Piton de la Fournaise and CC' on the Hell-Bourg landslide. Red and green lines represent original and optimized (with $k = 3$) processing results for combined acquisition modes. Purple and yellow lines represent optimized with $k = 3$ processing results for SM and IW modes respectively. The grey line represents the corresponding elevation profile

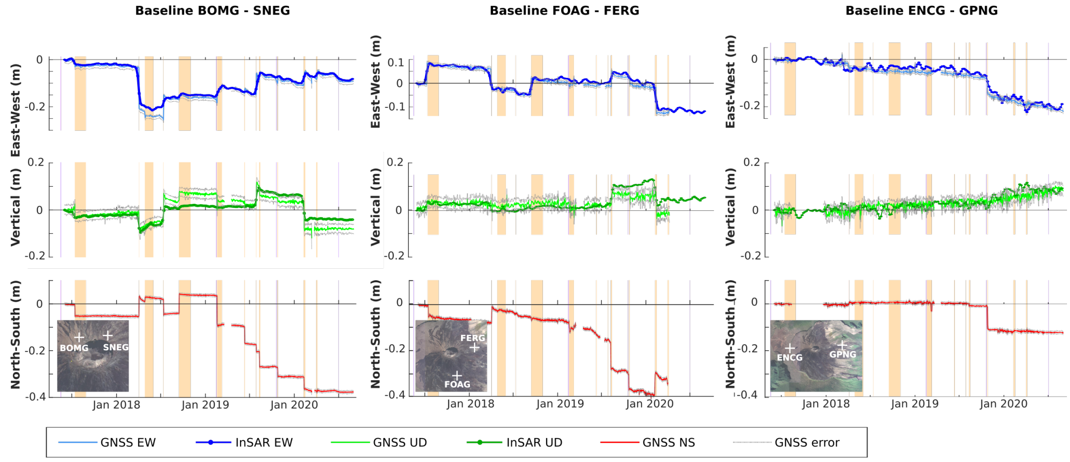

Figure 6. Differential time series for 3 pairs of pixels at Piton de la Fournaise volcano where GNSS stations are installed (BOMG and SNEG on the summit, FOAG and FERG at the base of the cone and ENCG and GPNG on the western border of Enclos Fouqué caldera and on the eastern flank respectively). MSBAS time series combine Wide Swath and StripMap S1 data. East-west, vertical and north-south components are blue, green and red lines, respectively. Orange and purple vertical bars mark the eruptions and magma intrusions. Grey dotted lines represent the GNSS data uncertainties.

Thanks to the short revisit time of Sentinel–1 on that region of the world, a new image is acquired in each mode every 12 days. With the 50 days temporal baseline criterion, only a small number of images were used more than 3 times as Primary or as Secondary images and the amount of pairs to be rejected by the optimization remains low. Only 5 and 3 % of the images were discarded for the SM mode in Ascending and Descending orbits (See two firsts rows in Fig. S19). For images acquired in IW mode, because the initial criteria was slightly higher, the benefit reaches 19 % for both Ascending and Descending orbits (See two lasts rows in Fig. S19).

In the case of the VVP, the results vary depending on the satellite (CSK, RADARSAT and S1) because of their orbital characteristics. The highest gain performance is obtained with S1 data acquired along the Ascending orbit. Because the revisit time was 24 days at the beginning of the Sentinel–1 mission, a temporal baseline as large as 400 days was necessary to ensure enough pairs in the baseline plot. However, when the revisit time was shortened down to 12 then 6 days, the same large baselines resulted in a huge amount of pairs considered for processing (2,575 pairs). The optimization with $k = 3$ discarded 75% of the pairs (see penultimate rows in Fig 8). A similar performance is achieved with RADARSAT data acquired in Ultra Fine (UF) mode. The optimization is less performative for CSK data with only 25% of discarded pairs while the revisit time is 4 days and the baseline criteria are 150 days, 150 m. The reason lies in the large orbital tube preferred by the Italian Space Agency, which is less efficient for SBAS-like methods as only a small number of interferograms respect the short spatial baseline criterion. Hence the optimization is particularly important for that type of orbital configuration as it will always ensure that the maximum number of pairs to compute will be $kN - 2k$ (where N is the number of images in a given mode) in the case of enough branches for each node and only two isolated nodes (i.e. the beginning and the end of the baseline plot). See Fig. S19 and 8 for more details about the baseline plots and the optimization for each data set considered.

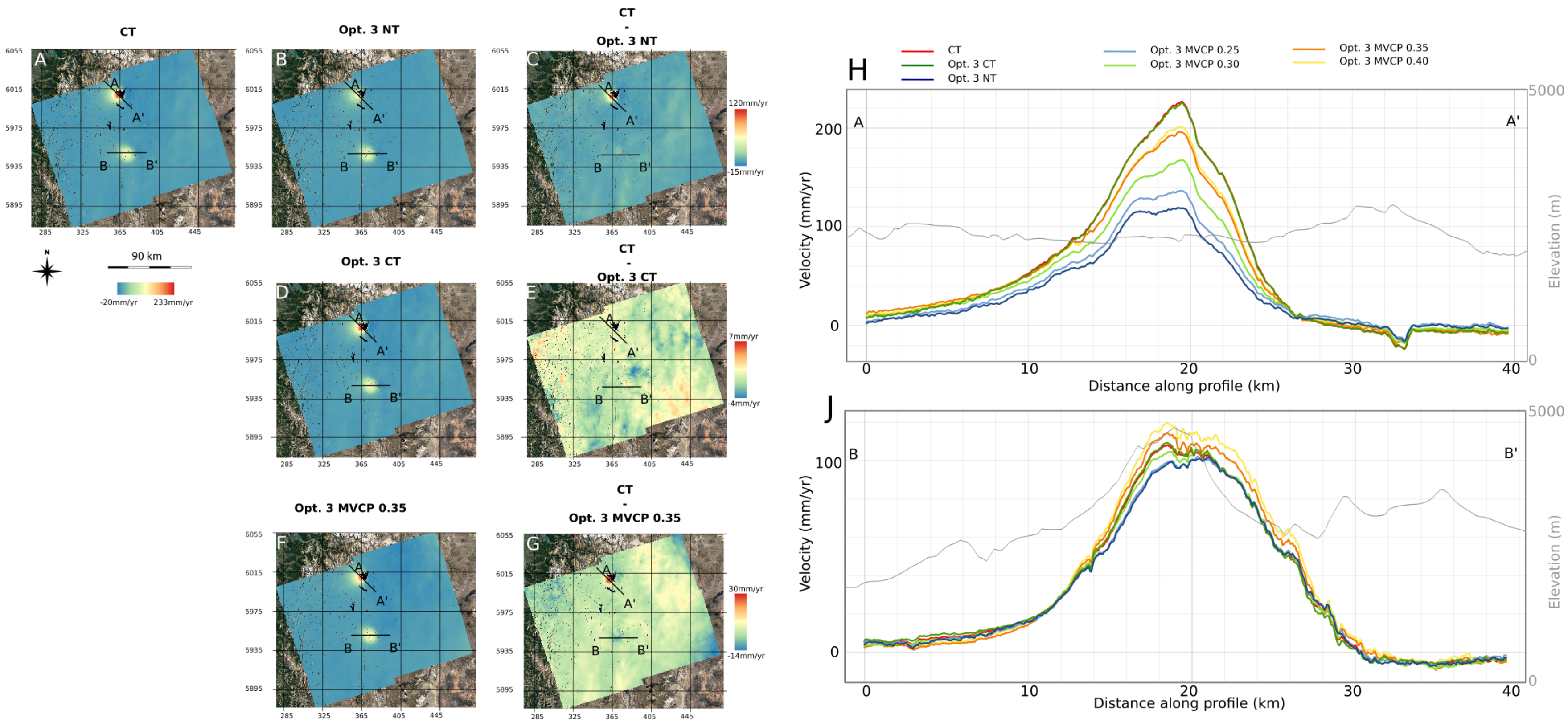

Figure 7. Velocity maps of Domuyo and Laguna del Maule area from an SBAS inversion of Sentinel–1 ascending data. The pair selection is done using (A) the original baseline criteria and an additional coherence threshold criterion at 0.24 (CT) ; (B) the optimized processing with $k = 3$ (NT); (D) the optimized processing with $k = 3$ and an additional coherence threshold criterion at 0.24 (Opt.3 CT); (F) the optimized processing with $k = 3$ and an additional coherence proxy minimum value criterion at 0.35 (Opt. 3 MVCP 0.35). (C) shows the difference between A and B. (E) shows the difference between A and D. (G) shows the difference between A and F. (H) and (J) display AA' (Laguna del Maule) and BB' (Domuyo) velocity profiles for all processing methods.

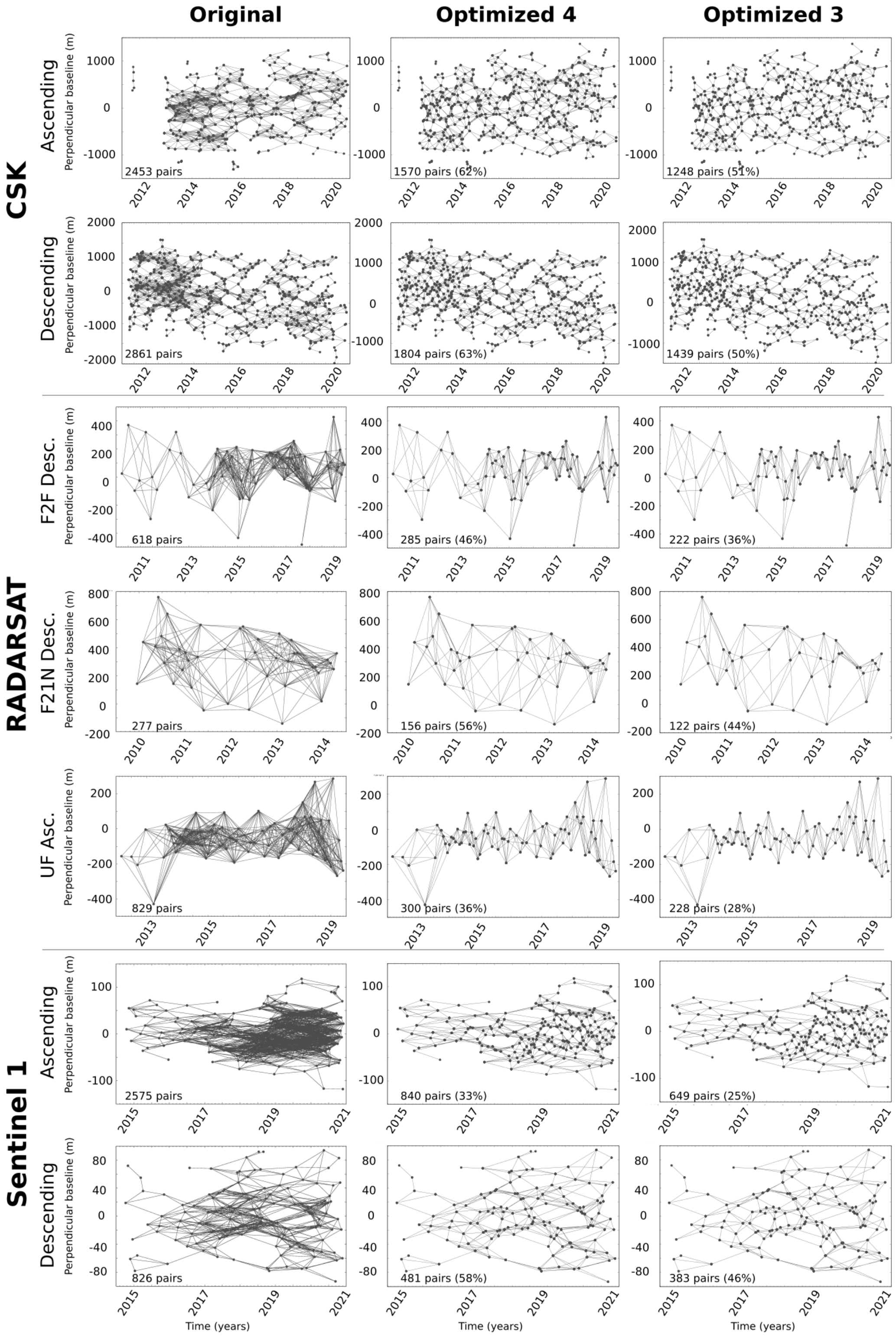

Figure 8. Original (column 1) and optimized baseline plots for each CSK, RADARSAT and S1 data set acquired on the VVP. The optimization is performed with $k = 4$ (column 2) and $k = 3$ (column 3). The total number of pairs used is indicated in the lower left corner of each baseline plot.The percentage of pairs kept after optimization with reference to the original processing is also indicated in the optimized baselines plots. Initial baselines criteria (B_P and B_T) are indicated on top of each baseline plot.

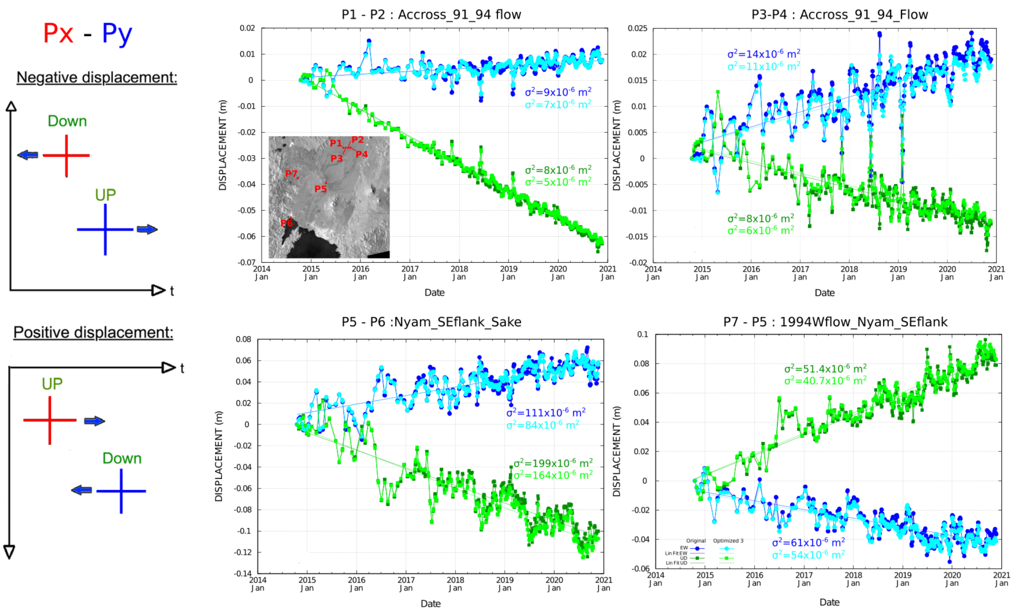

Figure 9. Double difference of vertical (green) and EW (blue) ground deformation (in m) between 4 pairs of pixels located on the flanks of Nyamulagira volcano computed with original processing (dark lines) and optimized processing (light lines). The inset is an amplitude image with the location of the seven pixels.

4.3 Signal to noise ratio improvement

To evaluate the benefit of the optimization and the balanced use of Primary and Secondary images on the time series quality, we compare the differential time series for several pairs of pixels computed with the original processing and with the optimized processing. Differential measurements between two pixels show the displacement of one pixel with respect to the other one. It has the advantage of providing an accurate reference for the movement. In addition, if the reference point is distal, though close enough to the deforming region, common systematic errors like those induced by the atmosphere that may affect both closely located pixels at the same time, will cancel out. Four east-west and vertical differential times series from the VVP are shown on Fig. 9 along with their linear fit. Pixels P1-P4 are located on the 1991-1993 lava flow from Nyamulagira. Pixels P5, P6 and P7 are located respectively on the flank of Nyamulagira near the 2010 eruptive site, at the base of the Western shoulder of the Rift on the 1994 lava flow and in the city of Sake, along the northern shore of Lake Kivu. Differential displacements between P1-P2 and P3-P4 confirm the well-known constant ongoing subsidence of the 1991-1993 lava flow (Samsonov & d'Oreye, 2012). Differential displacement of P5 with respect to P6 and P7 shows the constant subsidence of the 2010 eruptive site. A seasonal variation of nearly 4 cm in the vertical direction is clearly visible on the P5-P6 time series. Similar seasonal variation is also observed on the P7-P5 time series though with a much smaller amplitude (less than 1 cm). Whatever the area, time series that results from the optimized processing are very similar but less noisy than the original time series. The variance of residuals with respect to the linear fit is reduced by 15-30%. Particularly significant is the amplitude reduction of the outliers.

4.4 Validation with GNSS data

We showed with the VVP example that an optimization with $k = 3$ provides adequate results. In the VVP, a GNSS network exists (Geirsson et al., 2017; Ji et al., 2017) but stations are located in non-coherent areas meaning that it is impossible to compare both measurements.

The permanent GNSS monitoring network installed at PdF allows to cross check the MSBAS results with ground based measures. Fig. 6 compares 2D MSBAS and 3D GNSS differential ground deformation time series. The three pairs of MSBAS pixels are co-located with the six GNSS stations BOMG, SNEG, FOAG, FERG, ENCG and GPNG. During the 3 year time span of the compared periods (2017-06-01 : 2020-09-01), we note a very good agreement between both measurements. MSBAS results mainly remain within the GNSS uncertainty (grey lines). We notice however some offsets between GNSS and MSBAS at the occasion of large co-eruptive displacements. This is related to two limitations of the InSAR method. The major cause of discrepancy is the very strong spatial gradient that may occur during fast magma propagation toward the surface. Such a gradient cannot be unwrapped and the resulting deformation maps underestimate the displacements. The second cause of discrepancy come from the low sensitivity of InSAR to north-south displacement. Displacements are inverted in two dimensions assuming that north-south displacements are insignificant. This approximation was shown to be valid as long as the displacement in the north-south component is not significantly larger than in the other directions (Samsonov & d'Oreye, 2012). When this condition is not full-filled, MSBAS may overestimate the east-west and/or the vertical components. This is for instance illustrated with the FOAG-FERG baseline during the October 2019 eruption (Fig. 6). More details about comparison of MSBAS and GNSS data on PdF can be found in Samsonov et al. (2017) who compared a RADARSAT time series to GNSS displacements recorded at each station.

In the DLM region, we have no access to GNSS data but qualitative comparison was performed with published results in Derauw et al. (2020) and our optimized processing provides results very similar in a more efficient way.

4.5 A Coherence Threshold Restriction

We here compare the CT processing (Fig. 7A) with an optimized CT processing (Fig. 7D and E) and an optimized No Threshold (NT) processing (Fig. 7B and C). Derauw et al. (2020) demonstrated that processing without any coherence threshold underestimate the ground velocity at Laguna del Maule by about 48% and in the Domuyo area by 5-10%. The optimized NT processing gives underestimates (Fig. 7C) similar to the results of Derauw et al. (2020) without any optimization. This is due to the fact that even if the algorithm identifies well the low coherence pairs, it is forced to keep at least $2k$ pairs of images even if low weights are attributed to these pairs. Difference between the CT and the optimized CT shown in Fig. 7E are less than 5 mm/yr (i.e. less than 2.5% of the maximum deformation). Moreover, those differences correlate with the topography, where maximum values are on the summits while minimum values are in the valleys. Such a correlation with the topography suggests a reduction of the impact of the atmospheric noise due to the symmetric use of each image. However, the CT processing requires first the computation of a large number of coherence maps while the optimization aims at reducing the number of pairs to process. The results from the additional MVCP processing including an image rejection criterion using several values of the minimum coherence proxy value [from 0.25 to 0.40] are compared to the CT processing results (see Fig. 7F, H and J). Using such an *a priori* rejection criterion on images reduces the underestimates at LDM. Underestimation of the ground velocity is about 40% with $th = 0.25$ and reaches 12% with $th = 35$ and $th = 0.40$. An optimization with $k = 3$ in addition to the coherence restriction remains the better option. Velocity profiles at Laguna del Maule and Domuyo are very similar to the originals with only 23% and 20% of the pairs used in ascending and descending LOS respectively (Fig. 10). However this option requires the computation of all the coherence maps. The image rejection based on the coherence proxy does not fully retrieve the deformation measured with the original processing on LDM but provides an acceptable and efficient result for monitoring purposes.

5 Discussion

5.1 Calibration of the coherence proxy

Because computing the coherence for each pair of images is a time consuming step, we designed a quick way to provide an *a priori* estimate. Perlock et al. (2008) following Zebker et al. (1992) model the coherence as a product of several terms. However, using such a definition will systematically discard long temporal baselines (i.e. > 1 yr), while in our case-study areas decorrelation is seasonal and some long pairs may be coherent enough to have a strong interest when computing long time series over years. Here we choose to define a coherence proxy as the weighted sum of a seasonal component, a temporal component and a spatial component. Doing so, a pair with a long temporal baseline may have a non-negligible weight if both other terms are favorable. This coherence proxy is an efficient way to select the best quality pairs among a list while knowing only the orbits and acquisition dates. However, the weighting coefficients require a calibration as they can differ as a function of the studied area and satellites used. For example, with Sentinel–1 data at Domuyo and Laguna del Maule, the seasonal decorrelation is very strong (Fig. 2). The temporal decorrelation is also important. However, the spatial decorrelation is minimized by the small orbital tube of Sentinel–1 that allows to keep only pairs that satisfy a very short baseline criterion (here we used $B_P = 20$ m). The weight given to the spatial component w_3 is smaller than the two others (Tab. 2). At the tropical Reunion Island, the seasonal and temporal components are also the most important even if the spatial component is a bit more significant (Tab. 2). This could be due to the use of a larger initial perpendicular baseline criteria ($B_P = 50$ or 70 m). In an equatorial area like the Virunga, the temporal decorrelation is the most important factor (Fig. 3). However, the smaller weight assigned to the seasonal component could be a consequence of the more complex periodicity of the climatic constants in the area with two rainy seasons whereas our seasonal model only takes into account a periodicity of 12 months with a symmetric period of low and high coherence.

The calibration of the proxy can be performed in several ways. We present here four methods, starting from the quickest and less accurate to the longest and most demanding in terms of computing resources. The first method is an empirical calibration by trial and error in assigning values to a, b and c parameters (see Eq. 1). The second method consists in the inversion of these parameters using the coherence measured from a manual selection of a small number of pairs with different baselines characteristics and spanning different seasons (see Eq. 7). The third method bases the inversion on the coarse coherence values obtained from the systematic computation of all the possible pairs satisfying the baseline criteria over at least two years. The coarse coherence is estimated by computing the interferometric processes using a very high multi-looking factor to speed up the computation. The LazInSAR tool available in MasTer toolbox is well designed for that purpose as it is a single command line C program requiring only the location of the Primary and Secondary SLC image as input parameters. The fourth method probably provides the best calibration but is also the most expensive in terms of computational load as it requires the systematic computation of the coherence at the full resolution of the final interferometic products, for all pairs satisfying the baseline criteria, over a significant time period. Two years seems a good compromise to detect seasonal oscillations.

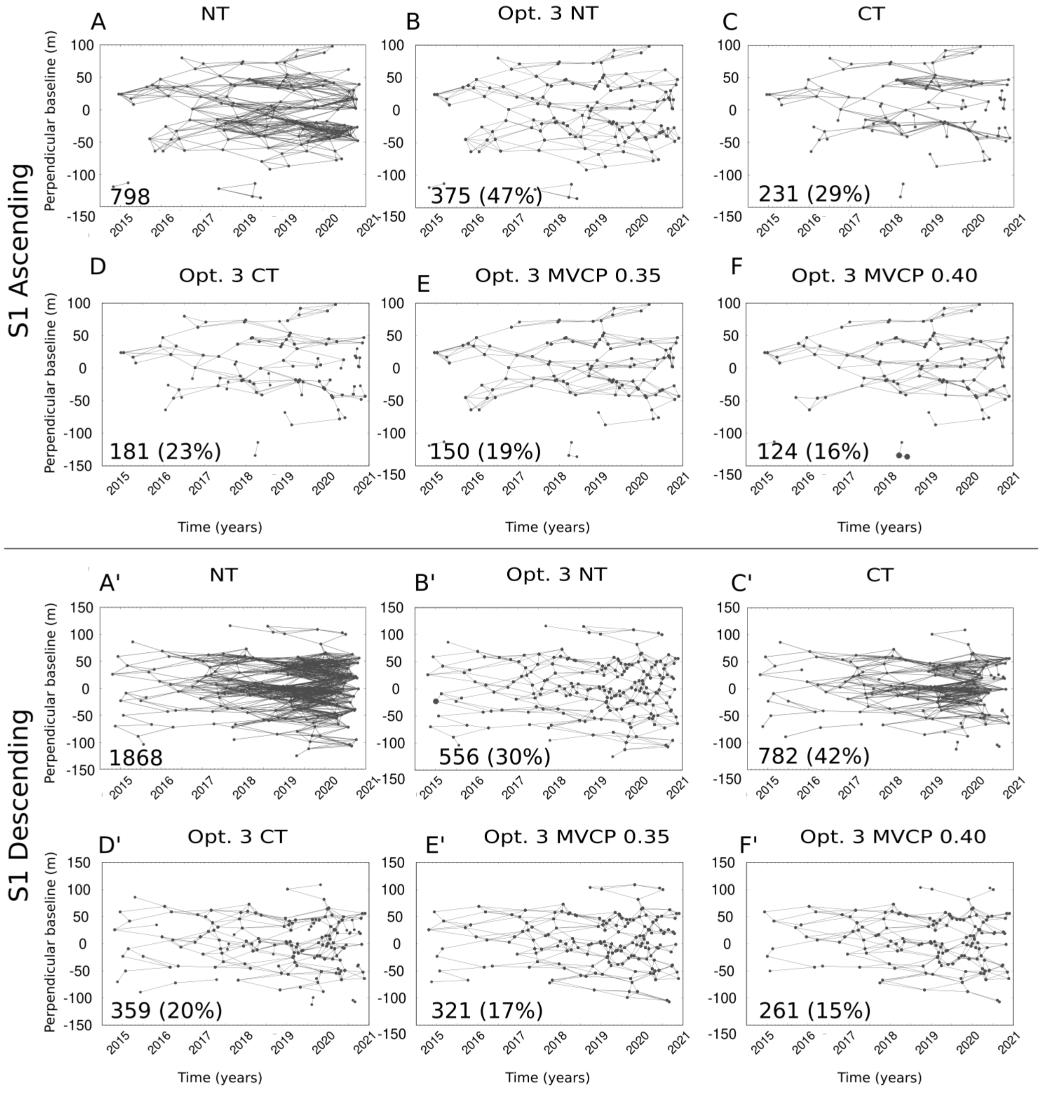

Figure 10. Baseline plots in the Domuyo-Laguna del Maule area for S1 Asc (A to F) and S1 Desc (A' to F') data. No threshold (NT) processing (A) and optimized ($k = 3$) NT processing (B) time baseline plots. Coherence threshold (CT) processing (C) and optimized ($k = 3$) CT processing time baseline plots, computed with a coherence threshold restriction at 0.24 on the LDM ROI. Optimized ($k = 3$) MVCP processing time baseline plot computed with a coherence proxy minimum value of 0.35 and 0.40, respectively (E and F). The total number of pairs used is indicated in the lower left corner of each baseline plot. The percentage of the number of pairs kept after optimization with reference to the original NT processing is also indicated. B_P=20 m and B_T=450 days.

Note that care must be taken if specific events occur during the test period used for the calibration. Some events may strongly modify temporarily or permanently the back-scattering properties of the ground and hence the coherence (e.g. wildfires, flooding, lava flow emplacements, landslides...). This could affect the assessment of the proxy and hence its performance to detect further high coherence interferograms. For instance, on La Reunion Island, after the burning of the vegetation in the Grand Brûlé area of Piton de la Fournaise volcano in January 2019, the average coherence increased by about 10% on the area (Fig. S1B). To the contrary, important lava flows occurring during frequent eruptions decrease the mean coherence. We therefore decided to exclude PdF from the ROI used to compute the mean coherence while calibrating the proxy.

5.2 On the optimization

In the original processing, among all possible pairs, only those satisfying the spatial and temporal baseline criteria (B_P,B_T) are processed. The more restrictive the criteria, the smaller is the number of interferograms to compute, store and integrate in the MSBAS processing. Inversely, larger baseline criteria require more computation time, data storage space and available raw-memory (MSBAS loads all the deformation maps in memory before performing the inversion). Nowadays, modern SAR satellites often fly in constellations, which shortens the revisit time. Moreover, when the orbital tube is maintained narrow enough (like for Sentinel–1), the number of interferograms that fit even small baselines criteria quickly increases. Using more restrictive criteria is not always an optimal solution, especially when seasonal decorrelation may affect the quality of the interferograms or when deformation is slow and dominated by the noise in short interferograms. Moreover, short temporal baselines (typically less than 3 months) can also induce a type of error recently identified and called the "fading signal" (Ansari et al., 2020). Studying that effect is out of the scope of the present paper and the optimization algorithm presented here does not prevent the "fading signal" when using short B_T criteria. However, the functionalities of the MasTer toolbox can offer a convenient tool to test and study that still-not-well-understood phenomenon.

The Laguna del Maule case study illustrates the benefit of keeping some long temporal baselines while shorter temporal baselines are discarded. Interferograms spanning one year from summer to summer may have better coherence than shorter interferograms due to the snow cover during austral winter, or during seasonal transitions. Keeping such long temporal baseline interferograms helps to retrieve the deformation over years. Also, by choosing a very restrictive spatial baseline, it may be difficult to keep enough connectivity in the baseline plot. If the database is split in several subsets, the MSBAS inversion is badly constrained which could produce artificial jumps or oscillations in the time series. Our algorithm solves this problem as it allows to keep baseline criteria large enough when necessary while rejecting redundant pairs to avoid the computation of a large number of unnecessarily interferograms.

By limiting the number of pairs to 6 with each image taken 3 times as Primary and 3 times as Secondary, the algorithm also produces a symmetric use of each image. Such a symmetry contributes to compensate atmospheric phase screens. From a mathematical point of view atmospheric artifacts can not be discriminated from a reversal deformations. An atmospheric artifact affecting an image will be seen as a spike in the ground deformation time series, which will not contribute much to the linear rate when there are many points in the time series. Nevertheless, to be able to resolve that atmospheric signal, one needs to have interferograms using the contaminated image as a Primary and others using it as a Secondary. In practice, the atmospheric noise gets mixed up with other sources of noise (e.g. orbital errors) and may not be fully compensated. The balanced use of image as Primary and Secondary ensures that we have symmetric network that resolves the atmospheric signal, which will be considered as much as a positive than a negative contribution. Experience shows that it contributes to significantly lower the variance in the time series (Fig. 9).

The influence of possible systematic DEM errors, which are inversely proportional to the height of ambiguity ha, is also reduced. On the one hand, the optimization reduces the total number of used interferograms and favors the selection of pairs with the smallest baselines B_P, i.e. the largest ha (Fig. S20). On the other hand, the selection process has statistically the same chance to keep pairs with positive or negative height of ambiguity, which means that DEM errors randomly add or subtract. We use the Nyamulagira 2011-2012 lava flow to quantitatively evaluate this DEM error reduction (see profile BB' in Fig. 4). Because the thick lava flow was not in place in 2000 when the SRTM DEM (Farr et al., 2007) used for the InSAR processing was computed, that flow produces on each interferometric pair a signal of d/ha fringes, where d is the thickness of the unmapped flow. The average absolute height of ambiguity increases from 2200 m in the original processing to 2300 m after optimization, and the mean number of interferograms computed with a given image drops from typically 15-20 to 6. Supposing that the DEM errors add (hence neglecting the sign of ha), the average thickness of the 2011-2012 lava flow being of the order of 15 m (Albino et al., 2015), the mean DEM error in deformation estimation hence drops from maximum 20/146th of a fringe (about 0.4cm) to 6/153th of a fringe (about 0.1cm). These values are however overestimated since they do not take into account the sign of ha. Finally, the pair selection also allows to reduce the time needed for computation, the space where data and products are stored and the raw-memory required to run the MSBAS inversion.

5.3 High resolution of SM mode better detect deformation in low coherent vegetated area.

On the Reunion Island, we processed separately and conjointly data from both acquisition modes available with Sentinel–1 (SM and IW). Coherence is much higher on the Piton de la Fournaise edifice (range [0.5-0.8]) than in the Cirque (range [0.2-0.35]) due to the vegetation (Fig. S1). No differences in the velocities measured using SM and IW modes are noticeable on the highly coherent Piton de la Fournaise edifice. However in the Salazie Cirque, SM detected deformation signal has a larger amplitude than IW (Fig. 5) and lateral variations are smoother. Considering that IW and SM have similar looking angle and revisit time, the main difference between both modes is the resolution. The higher resolution of SM images probably allows to observe coherent pixels surrounded by decorrelated pixels, which are seen in coarser resolution IW images as only a decorrelated region. Missing part of the signal, IW tends to underestimate the velocity compared to SM. The combined usage of IW and SM has the benefit of a higher temporal resolution (see section 2.2), but the velocity is somehow averaged between SM and IW velocities and thus underestimated.

5.4 A coherence threshold is required on the Laguna del Maule due to very specific conditions

At Laguna del Maule, a coherence threshold is required to reject poorly-coherent pairs. Otherwise the ground displacement may be underestimated by up to 47%. The strong loss of coherence in the area is due to snow-fall during the austral winter. However, snow-fall also affects the neighboring Domuyo volcano. Domuyo summit (4709 m) is higher than Laguna del Maule (2400 m), and we could expect more snow cover during the winter and hence stronger underestimation of the deformation. Surprisingly, the displacement is underestimated by only 5-10% in the Domuyo area, less than at Laguna Del Maule. Fig. 2 shows that the coherence loss is less important and that the coherence increase is faster during spring in the Domuyo area than at Laguna del Maule. Looking to average elevation profiles (Fig. S21), Domuyo Volcano is a mountain with steep slopes while the Laguna del Maule complex lie in a large flat depression hosting a lake.

The steep slopes and strong winds affecting the Domuyo are probably less favorable for snow accumulation, allowing some pixels to remain coherent. Our algorithm first selects the pairs that could be potentially removed while trying to keep connectivity between the graph nodes, then it removes the less favorable pairs using the coherence proxy criterion. If an image contains snow, all pairs formed with that image will have poor coherence. However, the algorithm is designed to keep at least 2k of them. Adding the possibility to remove an image if a minimum value on the coherence proxy is not reached for at least one pair improves the quality of the results and hence our ability to properly retrieve the deformation signal at Laguna del Maule. In this specific case, the coherence proxy is an acceptable solution if time or computer resources are limited in order to provide a quick solution for operational purposes. However, a proper *a posteriori* selection based on coherence (and not coherence proxy) values may be needed.

6 Conclusions

The amount of SAR data available thanks to shorter time delays between acquisitions results in new challenges in processing automatically and in near-real time long time series of EW and vertical deformation for volcano monitoring purposes. Adopting a graph point of view, we implemented a new pair selection tool to the automatic and unsupervised MasTer toolbox. This pair selection aims at limiting the number of pairs using each image and favors a symmetric use of each image. The selection criterion is a coherence proxy computed *a priori* from the orbital parameters and acquisition dates. When possible, using such a proxy avoids the long computation of many pairs. The optimization tool has been tested on three volcanic regions with very different characteristics: an equatorial forest in the Virunga, the tropical Reunion Island and the Domuyo and Laguna del Maule areas affected by snow seasonality. These tests show that the optimization improves the processing efficiency. The pair selection reduces by up to 75% the number of interferograms to compute. This means shorter computation time, as well as smaller data storage and raw-memory requirements. The algorithm proves to efficiently cope with periodic (annual) variations of coherence. Its performance can be further increased by implementing an additional coherence restriction to remove pairs affected by very strong decorrelation. This allows avoiding underestimating the deformation signal in some specific cases such as the regions of the Andes heavily affected by snow falls. The restriction to the best pairs and the symmetric use of each image as Primary and Secondary image also improve the time series quality, reducing by 15 to 30% the variance. In particular atmospheric noise is better compensated and the influence of DEM errors is minimized when using this optimization tool.

Acknowledgments

This research took advantage of continuous improvements of both the InSAR software and the MasTer implementation during projects, namely RESIST, MUZUBI and SM-MIP, principally funded by the Belgian Scientific Policy (BelSPo) and the luxembourgish Fond National de la Recherche (FNR). MSBAS software is freely available from http://insar.ca. The MasTer Engine (InSAR) software is now maintained and developped at LESVA (Laboratorio de Estudio y Seguimiento de Volcanes Activos) of the National University of Rio Negro (UNRN). Implementation of this automatized monitoring service was made possible thanks to the Copernicus Open Data Policy. Sentinel–1 data were obtained from ESA at https://schihub.copernicus.eu and it luxembourgish mirror https://www.collgs.lu/geocatalog.html. We are also grateful to Belspo and the GEO-GSNL, Virunga Supersite initiative, and Italian Space Agency (ASI) support for funding COSMO-SkyMed images and to the Canadian Space Agency for providing RADARSAT-2 data. All DEM used is this study were obtained using SRTM 1 Arc-Second Global (DOI number: /10.5066/F7PR7TFT) data freely available on https://earthexplorer.usgs.gov/. We thank the Observatoire Volcanologique du Piton de la Fournaise / Institut de Physique du Globe de Paris (OVPF/IPGP) for collecting and providing the GNSS time series used in this study. Raw GNSS data is available through the website: http://volobsis.ipgp.fr. This paper has benefited from interesting and helpful discussions with Léonard de Haro about the different ways to explore a graph and the subgraph extraction. We also thank T. Shreve for english reviewing of this manuscript We also would like to thank both the anonymous reviewers and the associated editor who helped to improve this manuscript.

References

Albino, F., Smets, B., d'Oreye, N., & Kervyn, F. (2015). High-resolution TanDEM-X DEM: An accurate method to estimate lava flow volumes at Nyamulagira Volcano (DR Congo). *Journal of Geophysical Research: Solid Earth*, *120*(6), 4189–4207.

Ansari, H., De Zan, F., & Parizzi, A. (2020). Study of systematic bias in measuring surface deformation with sar interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(2), 1285–1301.

Astort, A., Walter, T. R., Ruiz, F., Sagripanti, L., Nacif, A., Acosta, G., & Folguera, A. (2019). Unrest at Domuyo Volcano, Argentina, Detected by Geophysical and Geodetic Data and Morphometric Analysis. *Remote Sensing*, *11*(18), 2175.

Belle, P., Aunay, B., Bernardie, S., Grandjean, G., Ladouche, B., Mazué, R., & Join, J.-L. (2014). The application of an innovative inverse model for understanding and predicting landslide movements (Salazie cirque landslides, Reunion Island). *Landslides*, *11*(3), 343–355.

Berardino, P., Fornaro, G., Lanari, R., & Sansosti, E. (2002). A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Transactions on geoscience and remote sensing*, *40*(11), 2375–2383.

Chen, C. W., & Zebker, H. A. (2002). Phase unwrapping for large SAR interferograms: Statistical segmentation and generalized network models. *IEEE Transactions on Geoscience and Remote Sensing*, *40*(8), 1709–1719.

Colclough, S. (2006). Investigations of nyamuragira and nyiragongo volcanoes (democratic republic of the congo) using insar. In *Fringe 2005 workshop* (Vol. 610).

Colclough, S. (2007). *Investigations of nyamuragira and nyiragongo volcanoes (drc), using interferometric synthetic aperture radar* (Unpublished doctoral dissertation). University of Cambridge.

Communiqués ovpf. (2016-2020). Retrieved from `https://www.ipgp.fr/fr/ovpf/actualites-ovpf`

Delacourt, C., Raucoules, D., Le Mouélic, S., Carnec, C., Feurer, D., Allemand, P., & Cruchet, M. (2009). Observation of a large landslide on La Reunion Island using differential SAR interferometry (JERS and Radarsat) and correlation of optical (Spot5 and Aerial) images. *Sensors*, *9*(1), 616–630.

Delgado Blasco, J. M., Foumelis, M., Stewart, C., & Hooper, A. (2019). Measuring urban subsidence in the Rome metropolitan area (Italy) with Sentinel-1 SNAP-StaMPS persistent scatterer interferometry. *Remote Sensing*, *11*(2), 129.

De Luca, C., Cuccu, R., Elefante, S., Zinno, I., Manunta, M., Casola, V., … Casu, F. (2015). An on-demand web tool for the unsupervised retrieval of earth's surface deformation from sar data: The p-sbas service within the esa g-pod environment. *Remote Sensing*, *7*(11), 15630–15650.

Derauw, D. (1999). Phasimétrie par Radar à Synthèse d'Ouverture; théorie et applications. *Doctorate, Liege*.

Derauw, D., Jaspard, M., Caselli, A., Samsonov, S., et al. (2020). Ongoing automated ground deformation monitoring of Domuyo-Laguna del Maule area

(Argentina) using Sentinel-1 MSBAS time series: Methodology description and first observations for the period 2015–2020. *Journal of South American Earth Sciences*, *104*, 102850.

d'Oreye, N., Derauw, D., Samsonov, S., Jaspard, M., & Smittarello, D. (2021). Master: A full automatic multi-satellite insar mass processing tool for rapid incremental 2d ground deformation time series. In *2021 ieee international geoscience and remote sensing symposium igarss* (p. 1899-1902). doi: 10.1109/IGARSS47720.2021.9553615

Elefante, S., Imperatore, P., Zinno, I., Manunta, M., Mathot, E., Brito, F., … Casu, F. (2013). Sbas-dinsar time series generation on cloud computing platforms. In *2013 ieee international geoscience and remote sensing symposium-igarss* (pp. 274–277).

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., … others (2007). The shuttle radar topography mission. *Reviews of geophysics*, *45*(2).

Feigl, K. L., Le Mével, H., Tabrez Ali, S., Córdova, L., Andersen, N. L., DeMets, C., & Singer, B. S. (2014). Rapid uplift in Laguna del Maule volcanic field of the Andean Southern Volcanic zone (Chile) 2007–2012. *Geophysical Journal International*, *196*(2), 885–901.

Ferretti, A., Fumagalli, A., Novali, F., Prati, C., Rocca, F., & Rucci, A. (2011). A new algorithm for processing interferometric data-stacks: Squeesar. *IEEE transactions on geoscience and remote sensing*, *49*(9), 3460–3470.

Ferretti, A., Prati, C., & Rocca, F. (2000). Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry. *IEEE Transactions on geoscience and remote sensing*, *38*(5), 2202–2212.

Ferretti, A., Prati, C., & Rocca, F. (2001). Permanent scatterers in sar interferometry. *IEEE Transactions on geoscience and remote sensing*, *39*(1), 8–20.

Fournier, T., Pritchard, M., & Riddick, S. (2010). Duration, magnitude, and frequency of subaerial volcano deformation events: New results from Latin America using InSAR and a global synthesis. *Geochemistry, Geophysics, Geosystems*, *11*(1).

Geirsson, H., d'Oreye, N., Mashagiro, N., Syauswa, M., Celli, G., Kadufu, B., … Kervyn, F. (2017). Volcano-tectonic deformation in the Kivu Region, Central Africa: Results from six years of continuous GNSS observations of the Kivu Geodetic Network (KivuGNet). *Journal of African Earth Sciences*, *134*, 809–823.

Goldstein, R. M., & Werner, C. L. (1998). Radar interferogram filtering for geophysical applications. *Geophysical research letters*, *25*(21), 4035–4038.

Ishitsuka, K., Tsuji, T., & Matsuoka, T. (2016). Pixel-based interferometric pair selection in InSAR time-series analysis with baseline criteria. *Remote Sensing Letters*, *7*(7), 711–720.

Jacob, A. W., Vicente-Guijalba, F., Lopez-Martinez, C., Lopez-Sanchez, J. M., Litzinger, M., Kristen, H., … others (2020). Sentinel-1 insar coherence for land cover mapping: A comparison of multiple feature-based classifiers. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 535–552.

Ji, K. H., Stamps, D. S., Geirsson, H., Mashagiro, N., Syauswa, M., Kafudu, B., … d'Oreye, N. (2017). Deep magma accumulation at Nyamulagira volcano in 2011 detected by GNSS observations. *Journal of African Earth Sciences*, *134*, 824–830.

Le Mével, H., Feigl, K. L., Córdova, L., DeMets, C., & Lundgren, P. (2015). Evolution of unrest at Laguna del Maule volcanic field (Chile) from InSAR and GPS measurements, 2003 to 2014. *Geophysical Research Letters*, *42*(16), 6590–6598.

Libert, L. (2018). Towards operational use of combined Split-Band Interferometry and Multidimensional Small Baseline Subset: application to geohazard

monitoring in the Kivu region.
Lundgren, P., Girona, T., Bato, M. G., Realmuto, V. J., Samsonov, S., Cardona, C., ... Aivazis, M. (2020). The dynamics of large silicic systems from satellite remote sensing observations: the intriguing case of Domuyo volcano, Argentina. *Scientific reports*, *10*(1), 1–15.
Mairie de Goma. (2017). *Rapport annuel 2017.*
Massonnet, D., & Feigl, K. L. (1998). Radar interferometry and its application to changes in the Earth's surface. *Reviews of geophysics*, *36*(4), 441–500.
Morishita, Y., Lazecky, M., Wright, T. J., Weiss, J. R., Elliott, J. R., & Hooper, A. (2020). LiCSBAS: An open-source InSAR time series analysis package integrated with the LiCSAR automated Sentinel-1 InSAR processor. *Remote Sensing*, *12*(3), 424.
Nobile, A., Dille, A., Monsieurs, E., Basimike, J., Bibentyo, T. M., d'Oreye, N., ... Dewitte, O. (2018). Multi-temporal dinsar to characterise landslide ground deformations in a tropical urban environment: Focus on bukavu (dr congo). *Remote Sensing*, *10*(4), 626.
Oth, A., Barrière, J., d'Oreye, N., Mavonga, G., Subira, J., Mashagiro, N., ... others (2017). KivuSNet: The first dense broadband seismic network for the Kivu Rift region (western branch of East African Rift). *Seismological Research Letters*, *88*(1), 49–60.
Peltier, A., Froger, J.-L., Villeneuve, N., & Catry, T. (2017). Assessing the reliability and consistency of insar and gnss data for retrieving 3d-displacement rapid changes, the example of the 2015 piton de la fournaise eruptions. *Journal of Volcanology and Geothermal Research*, *344*, 106–120.
Pepe, A., Yang, Y., Manzo, M., & Lanari, R. (2015). Improved EMCF-SBAS processing chain based on advanced techniques for the noise-filtering and selection of small baseline multi-look DInSAR interferograms. *IEEE transactions on geoscience and remote sensing*, *53*(8), 4394–4417.
Perlock, P. A., González, P. J., Tiampo, K. F., Rodríguez-Velasco, G., Samsonov, S., & Fernández, J. (2008). Time evolution of deformation using time series of differential interferograms: Application to La Palma Island (Canary Islands). *Pure and applied geophysics*, *165*(8), 1531–1554.
Populationdata.net. (2017). Retrieved from `https://www.populationdata.net/pays/rwanda/`
Raucoules, D., de Michele, M., & Aunay, B. (2020). Landslide displacement mapping based on ALOS-2/PALSAR-2 data using image correlation techniques and SAR interferometry: application to the Hell-Bourg landslide (Salazie Circle, La Réunion Island). *Geocarto International*, *35*(2), 113–127.
Roult, G., Peltier, A., Taisne, B., Staudacher, T., Ferrazzini, V., Di Muro, A., et al. (2012). A new comprehensive classification of the Piton de la Fournaise activity spanning the 1985–2010 period. Search and analysis of short-term precursors from a broad-band seismological station. *Journal of Volcanology and Geothermal Research*, *241*, 78–104.
Samsonov, S., Dille, A., Dewitte, O., Kervyn, F., & d'Oreye, N. (2020). Satellite interferometry for mapping surface deformation time series in one, two and three dimensions: A new method illustrated on a slow-moving landslide. *Engineering Geology*, *266*, 105471.
Samsonov, S., & d'Oreye, N. (2012). Multidimensional time-series analysis of ground deformation from multiple InSAR data sets applied to Virunga Volcanic Province. *Geophysical Journal International*, *191*(3), 1095–1108.
Samsonov, S., & d'Oreye, N. (2017). Multidimensional small baseline subset (MSBAS) for two-dimensional deformation analysis: Case study Mexico City. *Canadian Journal of Remote Sensing*, *43*(4), 318–329.
Samsonov, S., Feng, W., Peltier, A., Geirsson, H., d'Oreye, N., & Tiampo, K. F. (2017). Multidimensional Small Baseline Subset (MSBAS) for volcano moni-

toring in two dimensions: Opportunities and challenges. Case study Piton de la Fournaise volcano. *Journal of Volcanology and Geothermal Research*, *344*, 121–138.

Singer, B. S., Andersen, N. L., Le Mével, H., Feigl, K. L., DeMets, C., Tikoff, B., ... others (2014). Dynamics of a large, restless, rhyolitic magma system at Laguna del Maule, southern Andes, Chile. *GSA today*, *24*(12), 4–10.

Smets, B., Wauthier, C., & d'Oreye, N. (2010). A new map of the lava flow field of nyamulagira (dr congo) from satellite imagery. *Journal of African Earth Sciences*, *58*(5), 778–786.

Tao, Q., Gao, T., Hu, L., & Wang, Z. (2018). Optimal selection and application analysis of multi-temporal differential interferogram series in StaMPS-based SBAS InSAR. *European Journal of Remote Sensing*, *51*(1), 1070–1086.

Toombs, A., & Wadge, G. (2012). Co-eruptive and inter-eruptive surface deformation measured by satellite radar interferometry at nyamuragira volcano, dr congo, 1996 to 2010. *Journal of Volcanology and Geothermal Research*, *245*, 98–122.

Wauthier, C., Smets, B., Hooper, A., Kervyn, F., & d'Oreye, N. (2018). Identification of subsiding areas undergoing significant magmatic carbon dioxide degassing, along the northern shore of Lake Kivu, East African Rift. *Journal of Volcanology and Geothermal Research*, *363*, 40–49.

Wu, H., Zhang, Y., Kang, Y., Lu, Z., & Cheng, X. (2019). Semi-automatic selection of optimum image pairs based on the interferometric coherence for time series SAR interferometry. *Remote Sensing Letters*, *10*(11), 1105–1112.

Yang, Y., Pepe, A., Manzo, M., & Lanari, R. (2012). An Effective Approach To Select The Interferometric SAR Data Pairs Based On Simulated Annealing. In *Fringe 2011* (Vol. 697, p. 106).

Yunjun, Z., Fattahi, H., & Amelung, F. (2019). Small baseline InSAR time series analysis: Unwrapping error correction and noise reduction. *Computers & Geosciences*, *133*, 104331.

Zebker, H. A., Villasenor, J., et al. (1992). Decorrelation in interferometric radar echoes. *IEEE Transactions on geoscience and remote sensing*, *30*(5), 950–959.

Zhang, Y., Kang, Y., et al. (2018). A New Automatic Selection of Optimum Interferometric Image Pairs in time Series Sar Interferometry. In *Igarss 2018-2018 ieee international geoscience and remote sensing symposium* (pp. 502–505).

Supporting Information for "Pair Selection Optimization for InSAR Time Series Processing"

D. Smittarello[1], N. d'Oreye[1,2], M. Jaspard[1], D. Derauw[1,3,4], S. Samsonov[5]

[1]European Center for Geodynamics and Seismology, 19 rue Josy Welter, L-7256 Walferdange, Gd Duchy of Luxembourg

[2]National Museum of Natural History, Department of Geophysics/Astrophysics, Rue Josy Welter 19, L-7256 Walferdange, Grand-Duchy of Luxembourg

[3]Laboratorio de Estudio y Seguimiento de Volcanes Activos (LESVA),IIPG-Universidad Nacional de Rio Negro - CONICET, 1242 Av. General Roca, 8332, General Roca, Argentina

[4]Centre Spatial de Liège (CSL), Avenue du Pré Aily, B-4031 Angleur, Belgium

[5]Canada Center for Mapping and Earth Observation, Natural Resources Canada (NRCAN), 560 Rochester Street, Ottawa, ON K1A 0E4, Canada

Contents of this file

1. Text S1 to S2
2. Figures S1 to S21

Introduction The supporting information presented in the following consist in a short text on the baseline criteria selection, on the possible optimization strategies, and a set of figures that illustrate how the coherence behave and how the coherence proxy defined in the main text is calibrated and adjusted to the different data sets used in the study (3 regions, 3 satellites and several acquisition modes).

Text S1. Baseline Criteria Selection The choice of B_T and B_P is critical and depends on the goal of the study. Their choice depends on several criteria that the user must carefully assess and adapt in function of e.g:

- The studied target (e.g. the possible DEM inaccuracies, the type of ground cover, the seasonal decorrelation, the type of expected deformation etc... , which are among the main drivers in the selection of B_T and B_P),
- The satellite (some satellites, such as CSK, have large orbital tubes, which do not favor short baselines between images close in time and hence force the usage or either long B_T or large B_P, or both)
- The computing resources available (the larger the maximum B_T and B_P, the larger the total number of compatible pairs to process, and hence the more space it takes on hard drive and the more it mobilizes the CPU's)
- The time of the processing (the larger the maximum B_T and B_P, the larger the total number of compatible pairs to process, and hence the longer time it takes)
- The type of application (systematic unsupervised processing for automatic monitoring or detailed study of a specific case)
- The speed of expected deformation (fast deformation requires short temporal baselines for avoiding aliasing or decorrelation due to ground cover reshaping, slow deformation requires long temporal baselines to maximize the signal to noise ratio)
- Other types of potential errors such as the fading signal (Ansari et al., 2020).

It is hence the responsibility of the user to find the appropriate values of B_T and B_P given its needs and taking into account these criteria while maintaining a gap-less Baselines Plot (interferometric graph connectivity).

Text S2. Optimization strategies

When the optimization is used in an incremental procedure (i.e. to be run each time a new image is made available), two strategies are possible. The strategies are described here for usage with MasTer, but it could be adapted for usage with other mass processing tools.

When a new image is made available, an original graph is computed (that is the graph of all pairs satisfying the B_T and B_P criteria). MasTer then lists among that graph all the pairs that do not exist in the database yet, that the pairs involving the new image. Here two possibilities exist:

1) MasTer computes the deformation maps for all these new compatible pairs so that they all exist in the data base whatever the optimization will select. Optimization then selects among that original graph a subset of pairs, and only these new deformation maps are added to the data base used to perform the MSBAS inversion. This speeds up the inversion process and reduces the noise as we demonstrated. However, it does not reduce the number of deformation maps to compute (because they are all computed), which is the most time-consuming tasks. However, experience shows that each new image requires about the same number of new pairs satisfying B_T and B_P criteria to be processed (as long as the orbital characteristics of the satellite does change). Hence, one knows in average the computer load and processing time a new image will require to update the time series. If that amount of processing load is manageable by the computer infrastructure, it is a convenient strategy as the user is ensured to have all the pairs processed that may be needed.

2) The other possibility consists in only computing the deformation maps for the pairs identified in the optimized graph, not in the original graph. This obviously reduces a lot the number of interferometric pairs to process, which is the most time consuming task. This is hence very convenient in the case of a one-shot study, e.g. a new study on a target area where a large number of images exists but no pairs were processed yet, and no update must be processed as soon as a new image will be made available. However, for the operation of an incremental monitoring ingesting every new image automatically, we can't exclude that a new image wouldn't suddenly destabilize the optimized graph. In such a case, the optimization would select a significantly different path through the original graph, which would in turn require the computation of a very large number of new interferometric pairs (including between old archives). In such a case, the time to process all the missing pairs may exceed by far the time usually required to compute only the small number of pairs involving only the new image, as in the case 1. In the worst case scenario, that processing time might exceed the time up to the delivery of the next image, which might hamper the use of that tool for monitoring purposes. This risk might be more important for instance when monitoring a highly coherent zone using large temporal baseline. For that reason, this second strategy is not recommended for automatic monitoring.

References

Ansari, H., De Zan, F., & Parizzi, A. (2020). Study of systematic bias in measuring surface deformation with sar interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(2), 1285–1301.

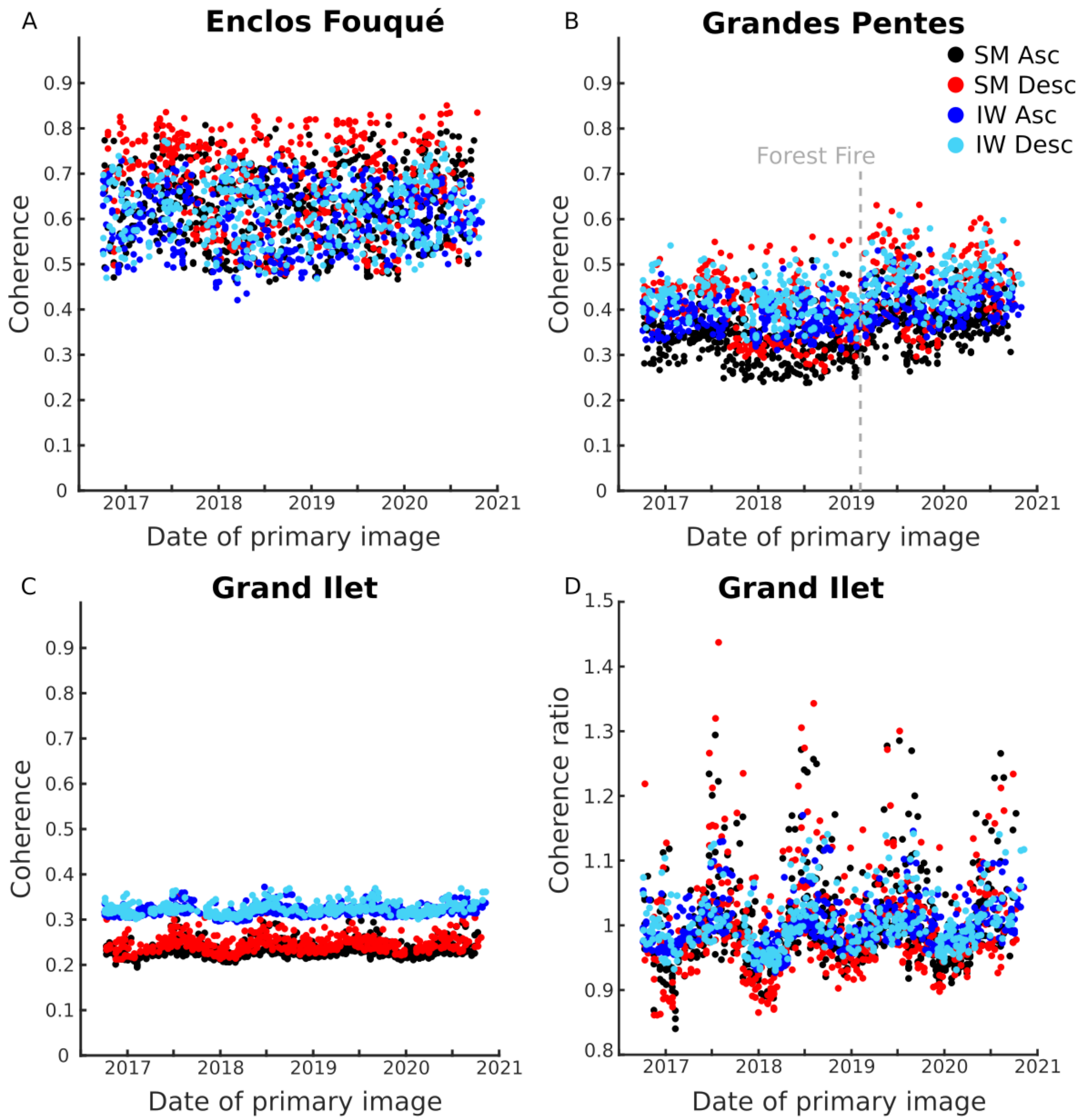

Figure S1. Characteristics of the coherence of the Reunion Island. Each dot represents the mean coherence for each pair of images computed on a ROI defined by a kml files framing the Enclos Fouqué (A), the lower eastern flank (Grandes Pentes (B)) and Grand Ilet (C). On panel D, each dot represents the ratio of the coherence for each pair to the mean value computed from all pairs on the Grand Ilet area.

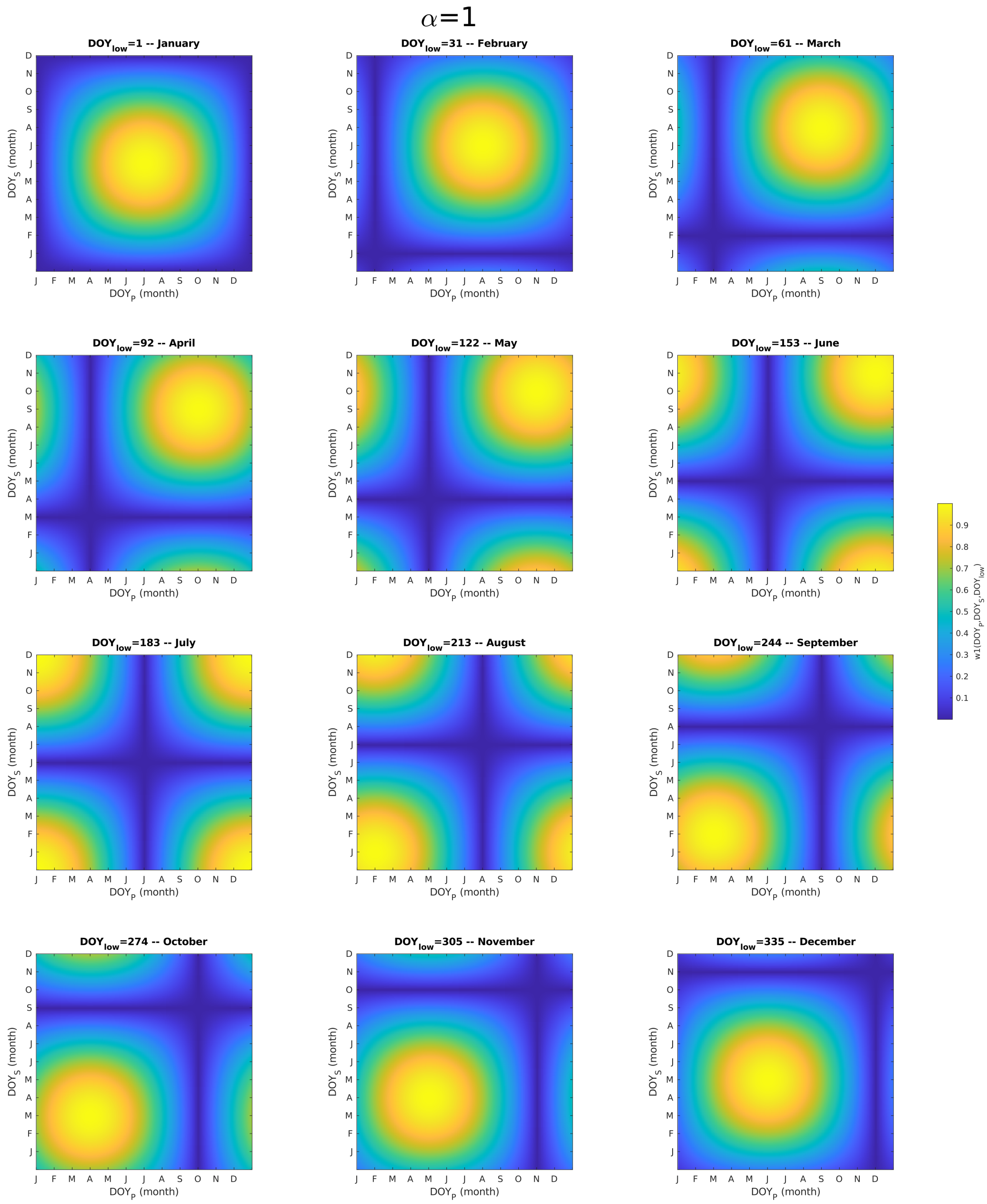

Figure S2. Evolution of the seasonal contribution $w1$ as a function of the Primary image day of year DOY_P and Secondary image day of year DOY_S for 12 values of the lowest coherence day of year DOY_{low} corresponding to the beginning of each month and the calibration factor $\alpha = 1$.

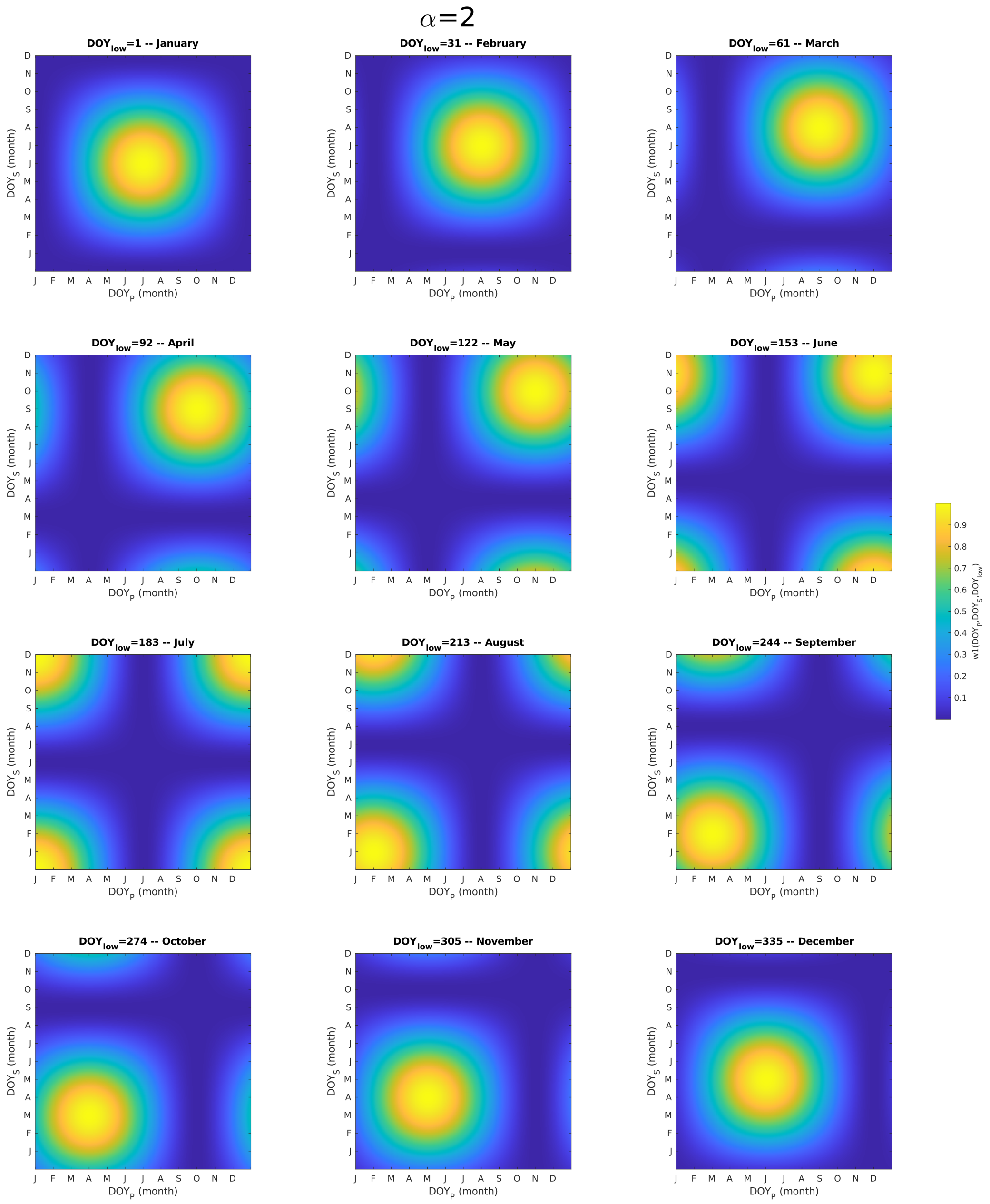

Figure S3. Evolution of the seasonal contribution $w1$ as a function of the Primary image day of year DOY_P and Secondary image day of year DOY_S for 12 values of the lowest coherence day of year DOY_{low} corresponding to the beginning of each month and the calibration factor $\alpha = 2$.

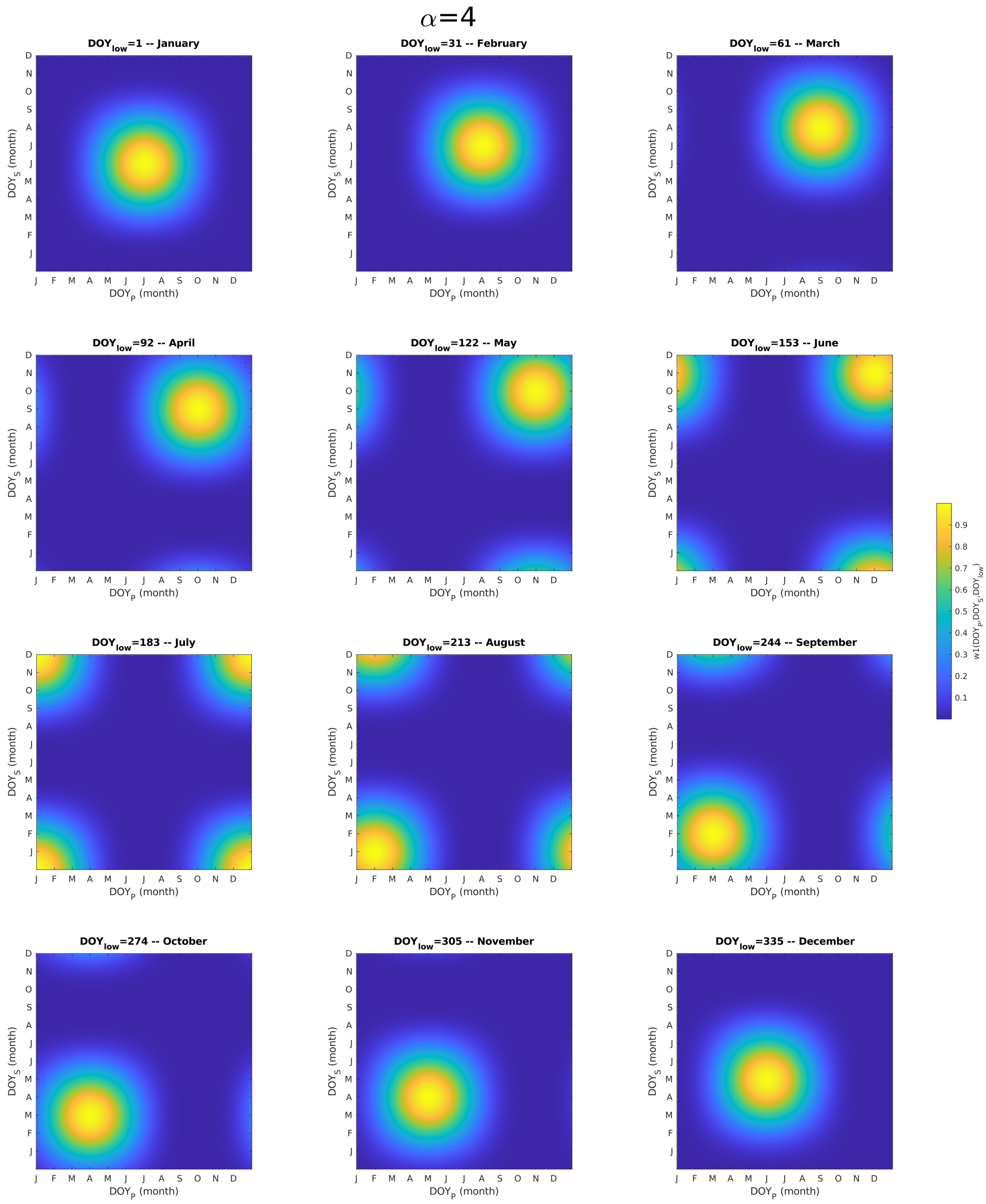

Figure S4. Evolution of the seasonal contribution $w1$ as a function of the Primary image day of year DOY_P and Secondary image day of year DOY_S for 12 values of the lowest coherence day of year DOY_{low} corresponding to the beginning of each month and the calibration factor $\alpha = 4$.

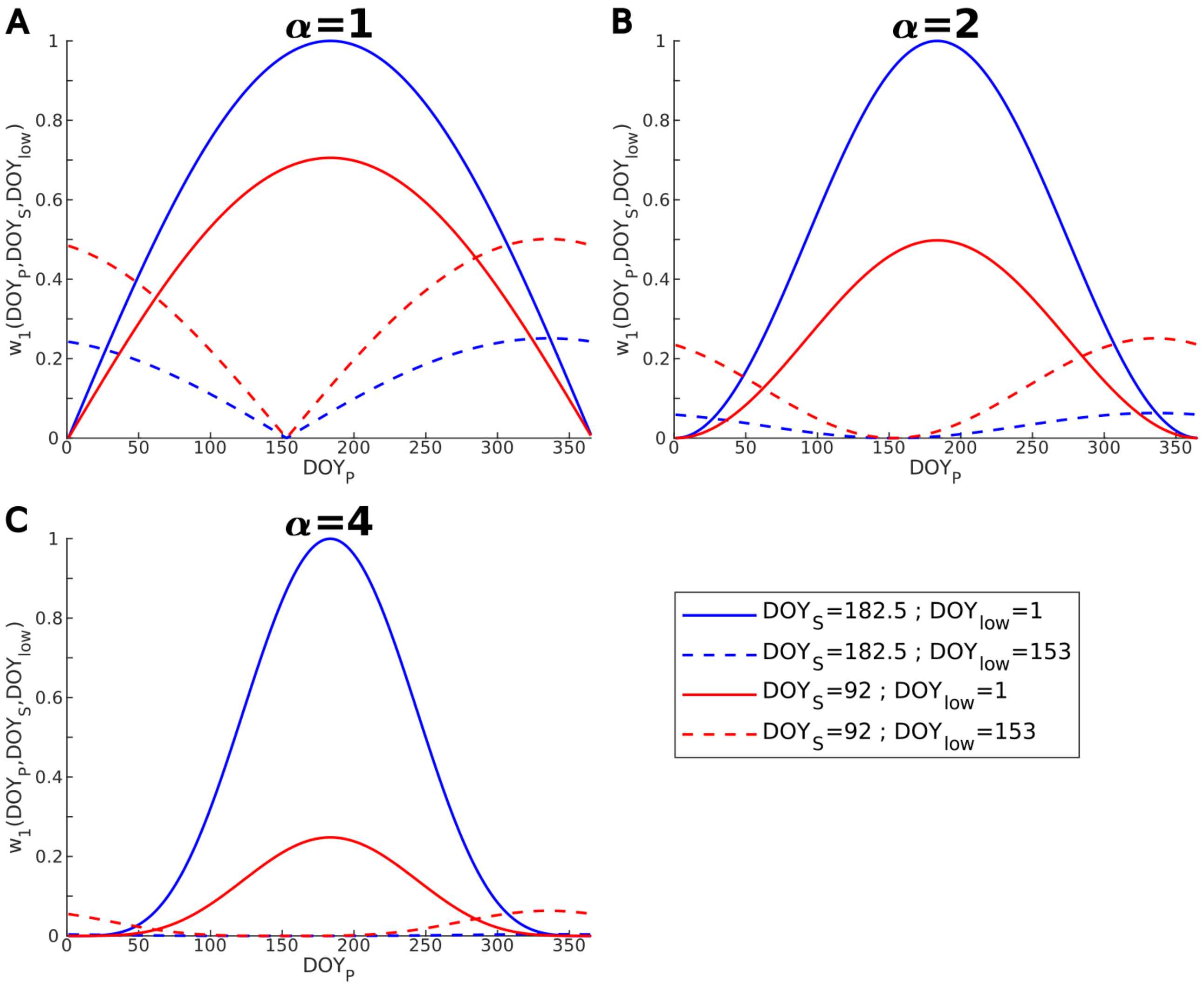

Figure S5. Evolution of the seasonal contribution $w1$ as a function of the Primary image day of year DOY_P for (A) $\alpha = 1$, (B) $\alpha = 2$ and (C) $\alpha = 4$. The Secondary image acquisition are fixed to April 2^{nd} ($DOY_S = 92$) in red or July 1^{st} ($DOY_S = 182.5$) in blue. The day of year with the lowest coherence is chosen an January 1^{st} ($DOY_{low} = 1$) in plain lines or June 2^{nd} ($DOY_{low} = 153$) in dashed lines.

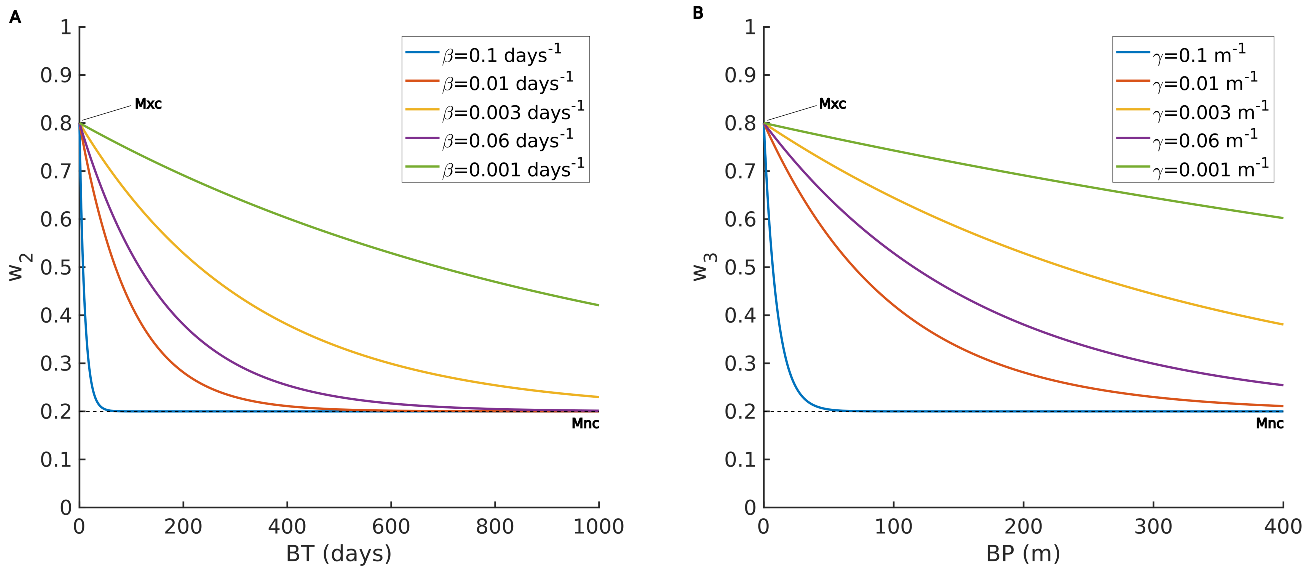

Figure S6. Evolution of temporal and spatial contributions to the coherence proxy. (A) Evolution of the temporal contribution w_2 as a function of the temporal baseline BT for several values of the calibration parameter β. (B) Evolution of the spatial contribution w_3 as a function of the perpendicular baseline BP for several values of the calibration parameter γ. M_{xc} and M_{nx} are the maximum and minimum values expected for mean the coherence in the region of interest.

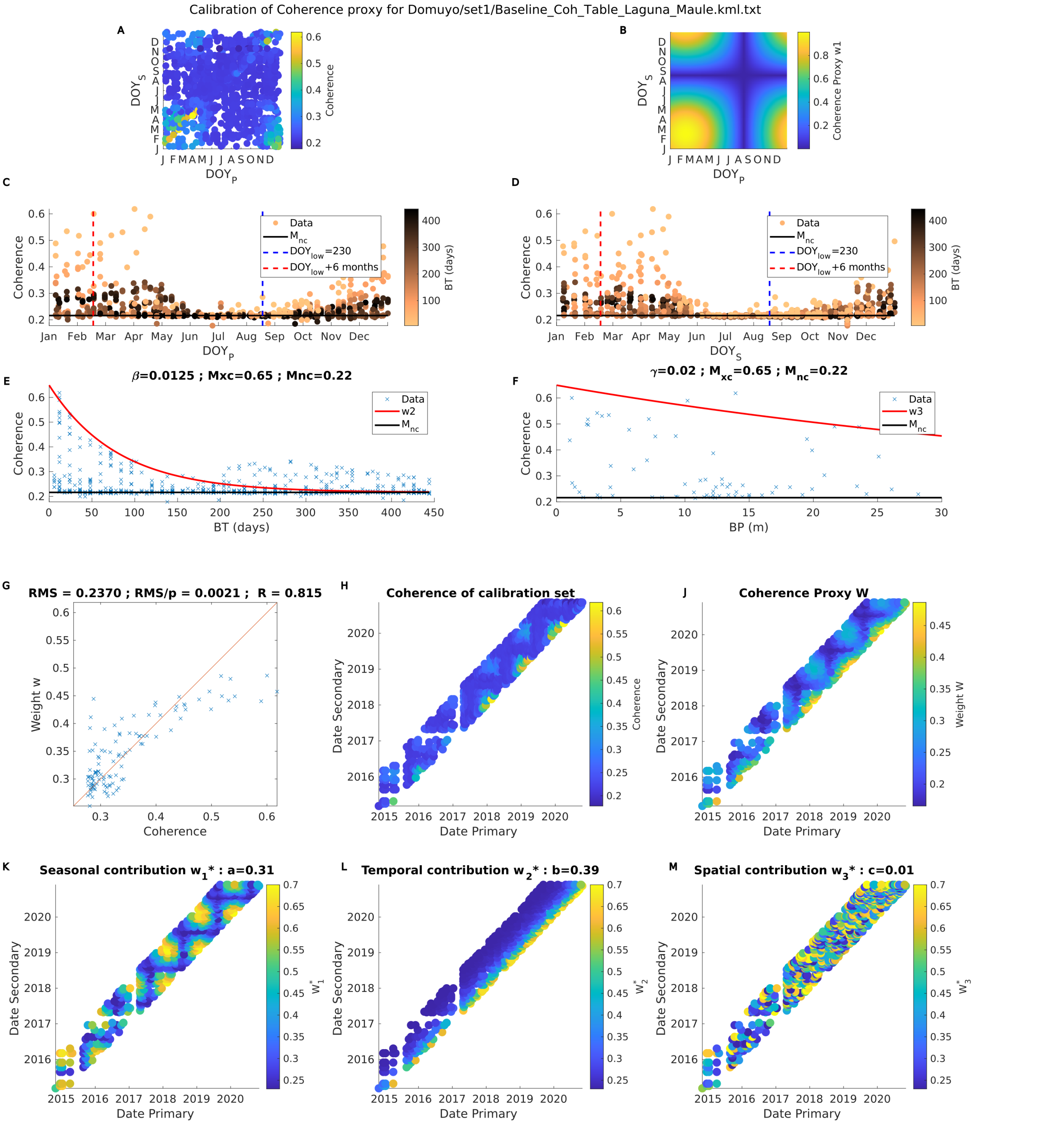

Figure S7. Calibration of the coherence proxy on the DLM area for S1 ascending dataset. A and B represent the mean coherence computed on the Laguna del Maule ROI and the seasonal contribution w_1 to the coherence proxy as a function of the primary and secondary image day of year of each pair. C and D represent the mean coherence computed on the Laguna del Maule ROI as a function of the Primary (Secondary) image Day of year respectively. Blue and red dash lines mark DOY_{low} and $DOY_{low} + 6$ months, respectively. E and F represent the mean coherence computed on the Laguna del Maule ROI as a function of BT for pairs with $BP < 15$ m and as a function of BP for pairs with $BT < 25$ days, respectively. Red line represent the exponential decrease modeled by w_2 and w_3 in the coherence proxy. C to F, horizontal black line marks the minimum expected value of the coherence M_{nc}. G represents the coherence proxy w versus the mean coherence computed on the Laguna del Maule ROI for each pair of the calibration set. The red line marks the first bisector. H and J represent the the mean coherence computed on the Laguna del Maule ROI for each pair of the calibration set and the coherence proxy w, respectively, as a function of the primary and secondary image acquisition date. K to M, the color scale represent the partial weights w_1, w_2 and w_3 (see text section 2.3.1.2), respectively. The value of the corresponding coefficients a, b, and c is indicated on top of each plot.

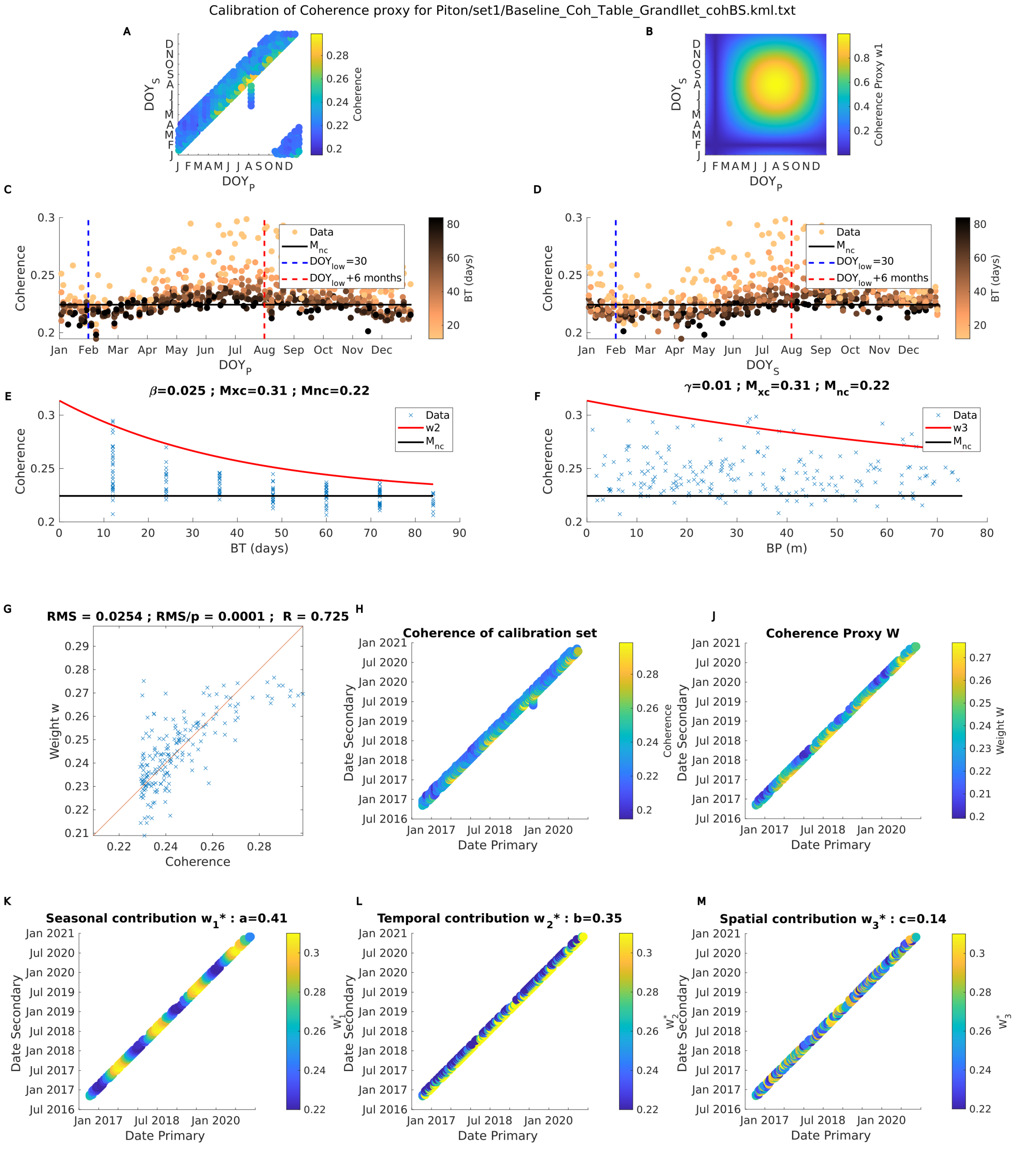

Figure S8. Calibration of the coherence proxy on the Reunion Island area for S1 SM ascending dataset. A and B represent the mean coherence computed on the Grand Ilet ROI and the seasonal contribution w_1 to the coherence proxy as a function of the primary and secondary image day of year of each pair. C and D represent the mean coherence computed on the Grand Ilet ROI as a function of the Primary (Secondary) image Day of year respectively. Blue and red dash lines mark DOY_{low} and $DOY_{low} + 6$ months, respectively. E and F represent the mean coherence computed on the Grand Ilet ROI as a function of BT for pairs with $BP < 15$ m and as a function of BP for pairs with $BT < 25$ days, respectively. Red line represent the exponential decrease modeled by w_2 and w_3 in the coherence proxy. C to F, horizontal black line marks the minimum expected value of the coherence M_{nc}. G represents the coherence proxy w versus the mean coherence computed on the Grand Ilet ROI for each pair of the calibration set. The red line marks the first bisector. H and J represent the the mean coherence computed on the Grand Ilet ROI for each pair of the calibration set and the coherence proxy w, respectively, as a function of the primary and secondary image acquisition date. K to M, the color scale represent the partial weights w_1, w_2 and w_3 (see text section 2.3.1.2), respectively. The value of the corresponding coefficients a, b, and c is indicated on top of each plot.

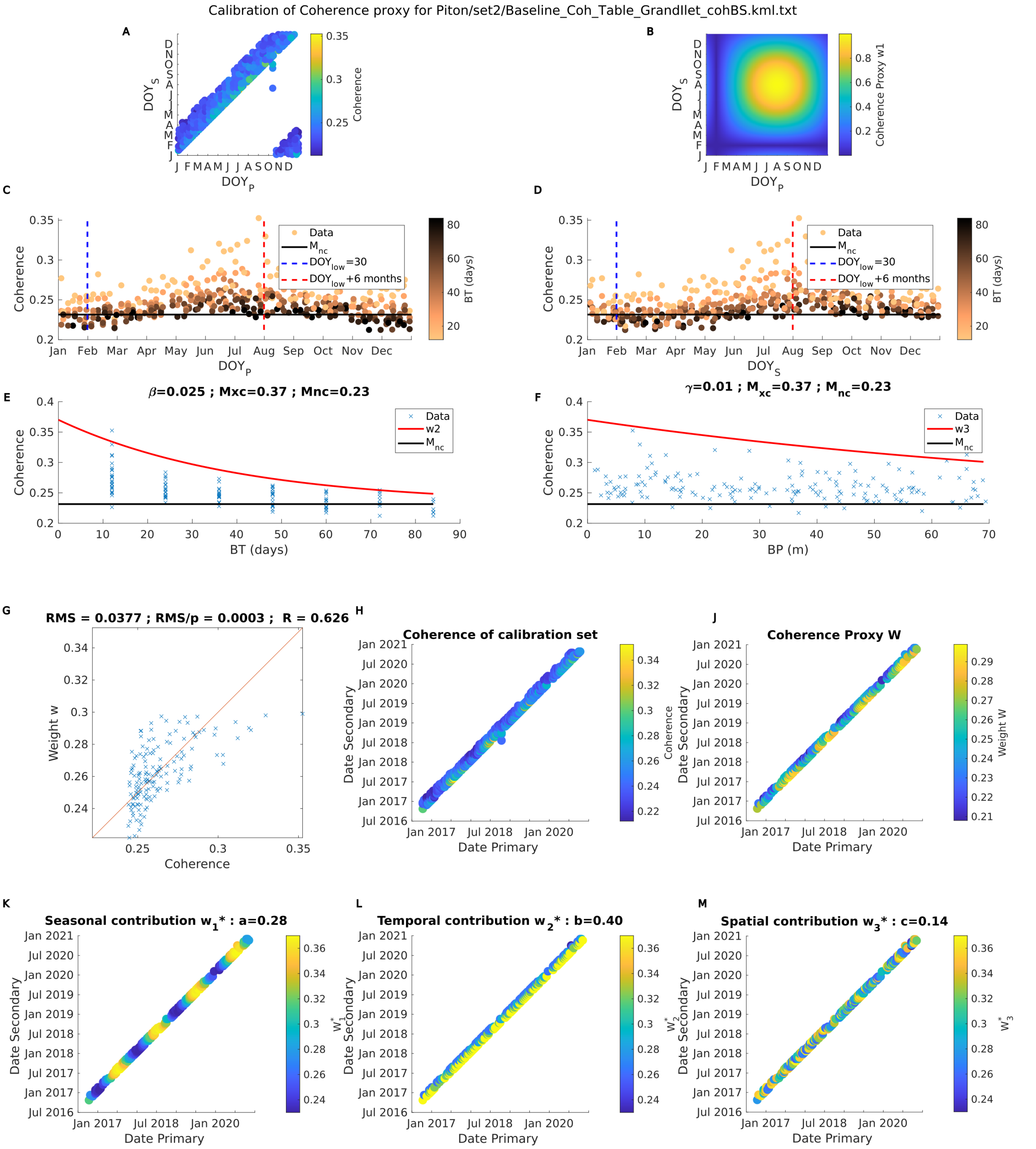

Figure S9. Calibration of the coherence proxy on the Reunion Island area for S1 SM descending dataset. A and B represent the mean coherence computed on the Grand Ilet ROI and the seasonal contribution w_1 to the coherence proxy as a function of the primary and secondary image day of year of each pair. C and D represent the mean coherence computed on the Grand Ilet ROI as a function of the Primary (Secondary) image Day of year respectively. Blue and red dash lines mark DOY_{low} and $DOY_{low} + 6$ months, respectively. E and F represent the mean coherence computed on the Grand Ilet ROI as a function of BT for pairs with $BP < 15$ m and as a function of BP for pairs with $BT < 25$ days, respectively. Red line represent the exponential decrease modeled by w_2 and w_3 in the coherence proxy. C to F, horizontal black line marks the minimum expected value of the coherence M_{nc}. G represents the coherence proxy w versus the mean coherence computed on the Grand Ilet ROI for each pair of the calibration set. The red line marks the first bisector. H and J represent the the mean coherence computed on the Grand Ilet ROI for each pair of the calibration set and the coherence proxy w, respectively, as a function of the primary and secondary image acquisition date. K to M, the color scale represent the partial weights w_1, w_2 and w_3 (see text section 2.3.1.2), respectively. The value of the corresponding coefficients a, b, and c is indicated on top of each plot.

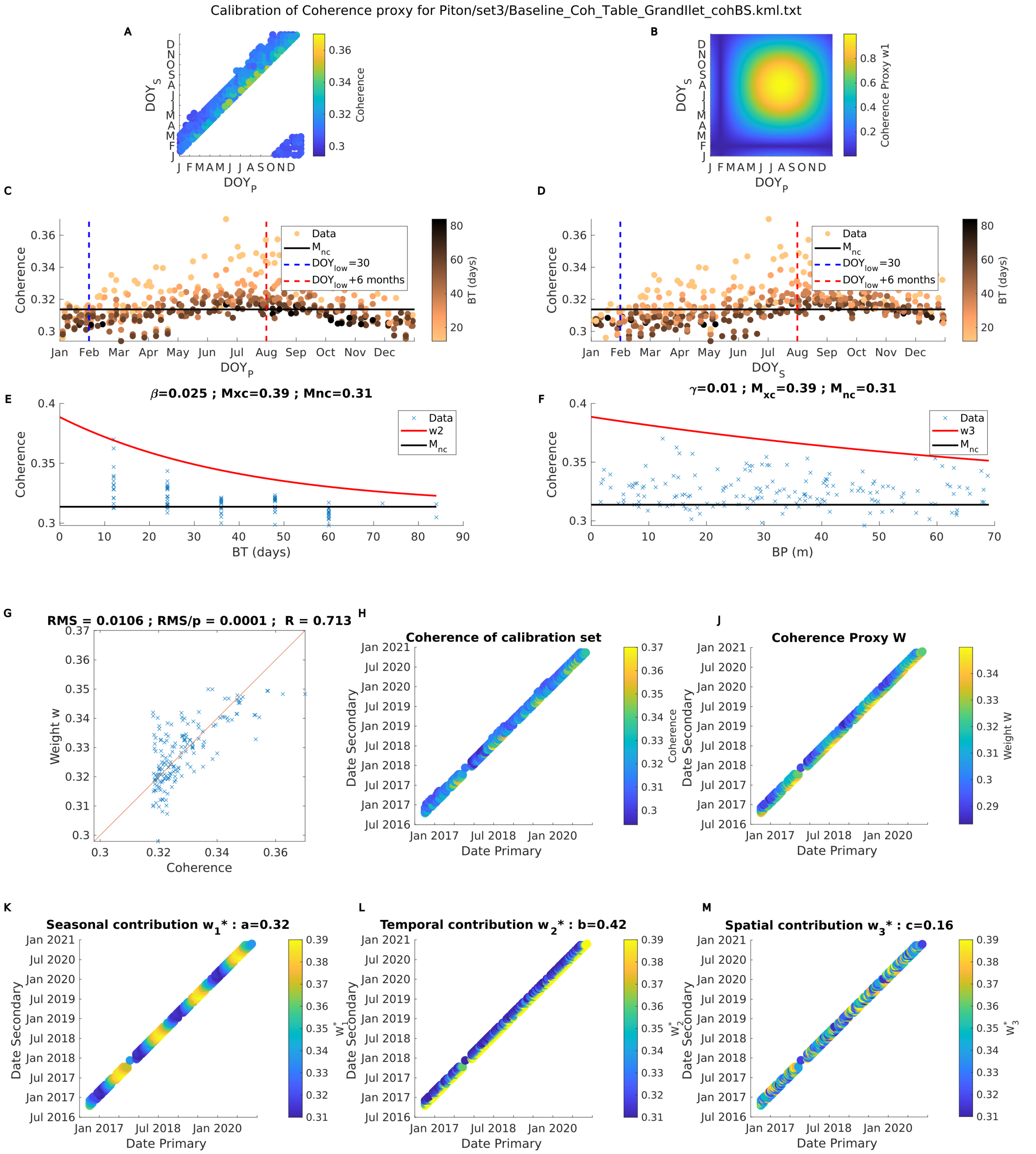

Figure S10. Calibration of the coherence proxy on the Reunion Island area for S1 IW ascending dataset. A and B represent the mean coherence computed on the Grand Ilet ROI and the seasonal contribution w_1 to the coherence proxy as a function of the primary and secondary image day of year of each pair. C and D represent the mean coherence computed on the Grand Ilet ROI as a function of the Primary (Secondary) image Day of year respectively. Blue and red dash lines mark DOY_{low} and $DOY_{low} + 6$ months, respectively. E and F represent the mean coherence computed on the Grand Ilet ROI as a function of BT for pairs with $BP < 15$ m and as a function of BP for pairs with $BT < 25$ days, respectively. Red line represent the exponential decrease modeled by w_2 and w_3 in the coherence proxy. C to F, horizontal black line marks the minimum expected value of the coherence M_{nc}. G represents the coherence proxy w versus the mean coherence computed on the Grand Ilet ROI for each pair of the calibration set. The red line marks the first bisector. H and J represent the the mean coherence computed on the Grand Ilet ROI for each pair of the calibration set and the coherence proxy w, respectively, as a function of the primary and secondary image acquisition date. K to M, the color scale represent the partial weights w_1, w_2 and w_3 (see text section 2.3.1.2), respectively. The value of the corresponding coefficients a, b, and c is indicated on top of each plot.

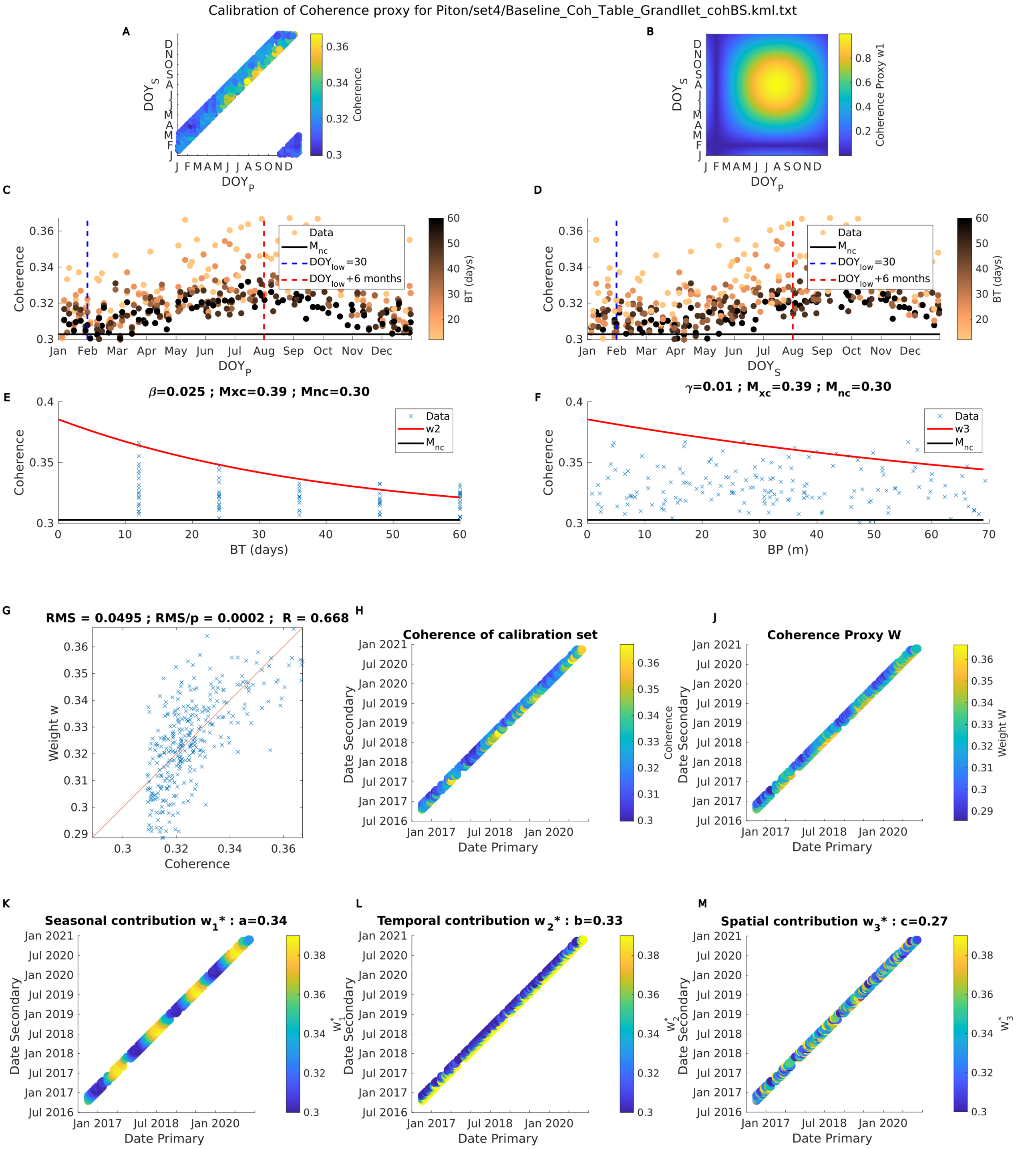

Figure S11. Calibration of the coherence proxy on the Reunion Island area for S1 IW descending dataset. A and B represent the mean coherence computed on the Grand Ilet ROI and the seasonal contribution w_1 to the coherence proxy as a function of the primary and secondary image day of year of each pair. C and D represent the mean coherence computed on the Grand Ilet ROI as a function of the Primary (Secondary) image Day of year respectively. Blue and red dash lines mark DOY_{low} and $DOY_{low} + 6$ months, respectively. E and F represent the mean coherence computed on the Grand Ilet ROI as a function of BT for pairs with $BP < 15$ m and as a function of BP for pairs with $BT < 25$ days, respectively. Red line represent the exponential decrease modeled by w_2 and w_3 in the coherence proxy. C to F, horizontal black line marks the minimum expected value of the coherence M_{nc}. G represents the coherence proxy w versus the mean coherence computed on the Grand Ilet ROI for each pair of the calibration set. The red line marks the first bisector. H and J represent the the mean coherence computed on the Grand Ilet ROI for each pair of the calibration set and the coherence proxy w, respectively, as a function of the primary and secondary image acquisition date. K to M, the color scale represent the partial weights w_1, w_2 and w_3 (see text section 2.3.1.2), respectively. The value of the corresponding coefficients a, b, and c is indicated on top of each plot.

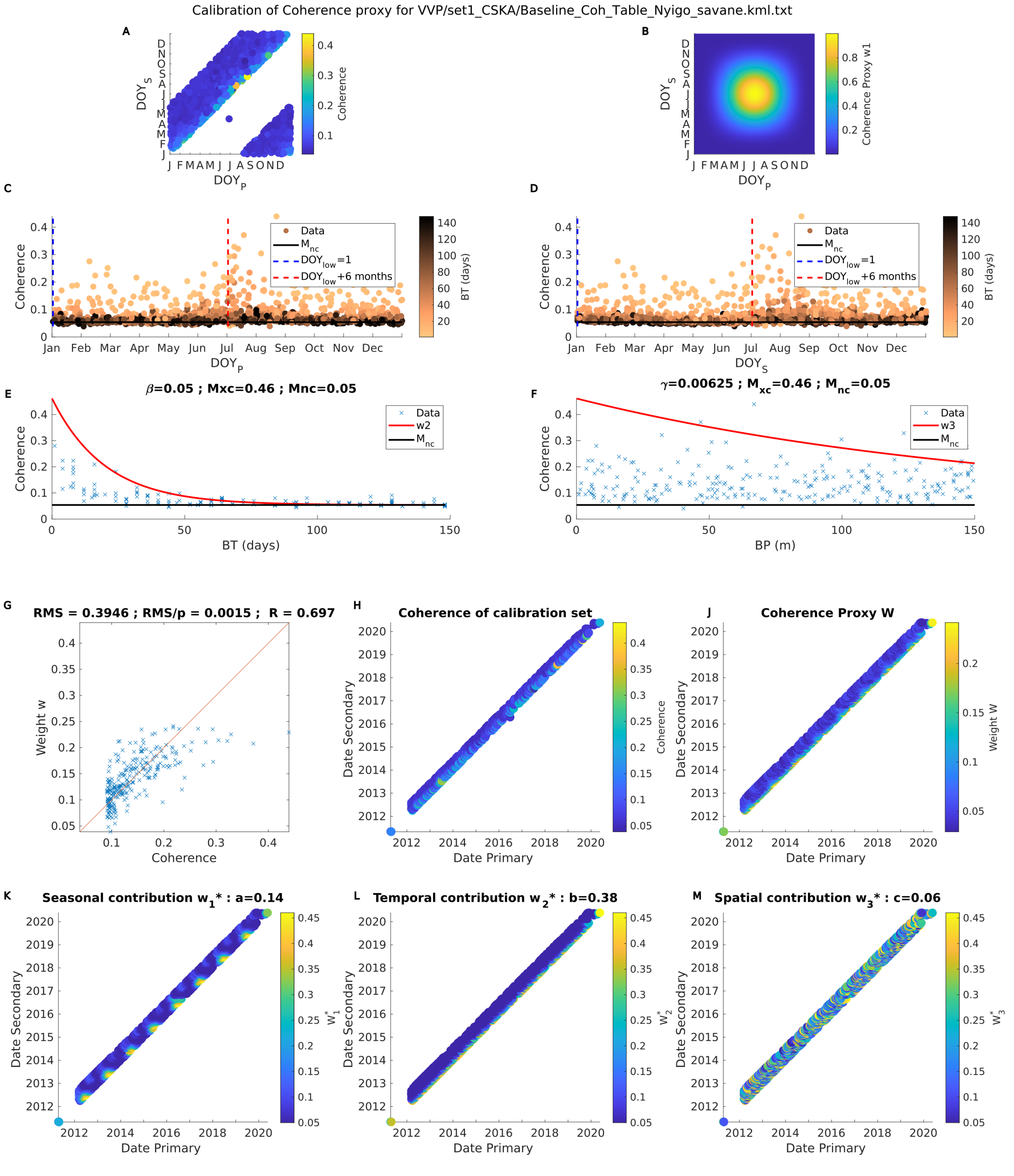

Figure S12. Calibration of the coherence proxy on the VVP area for CSK ascending dataset. A and B represent the mean coherence computed on the Savannah ROI and the seasonal contribution w_1 to the coherence proxy as a function of the primary and secondary image day of year of each pair. C and D represent the mean coherence computed on the Savannah ROI as a function of the Primary (Secondary) image Day of year respectively. Blue and red dash lines mark DOY_{low} and $DOY_{low} + 6$ months, respectively. E and F represent the mean coherence computed on the Savannah ROI as a function of BT for pairs with $BP < 15$ m and as a function of BP for pairs with $BT < 25$ days, respectively. Red line represent the exponential decrease modeled by w_2 and w_3 in the coherence proxy. C to F, horizontal black line marks the minimum expected value of the coherence M_{nc}. G represents the coherence proxy w versus the mean coherence computed on the Savannah ROI for each pair of the calibration set. The red line marks the first bisector. H and J represent the the mean coherence computed on the Savannah ROI for each pair of the calibration set and the coherence proxy w, respectively, as a function of the primary and secondary image acquisition date. K to M, the color scale represent the partial weights w_1, w_2 and w_3 (see text section 2.3.1.2), respectively. The value of the corresponding coefficients a, b, and c is indicated on top of each plot.

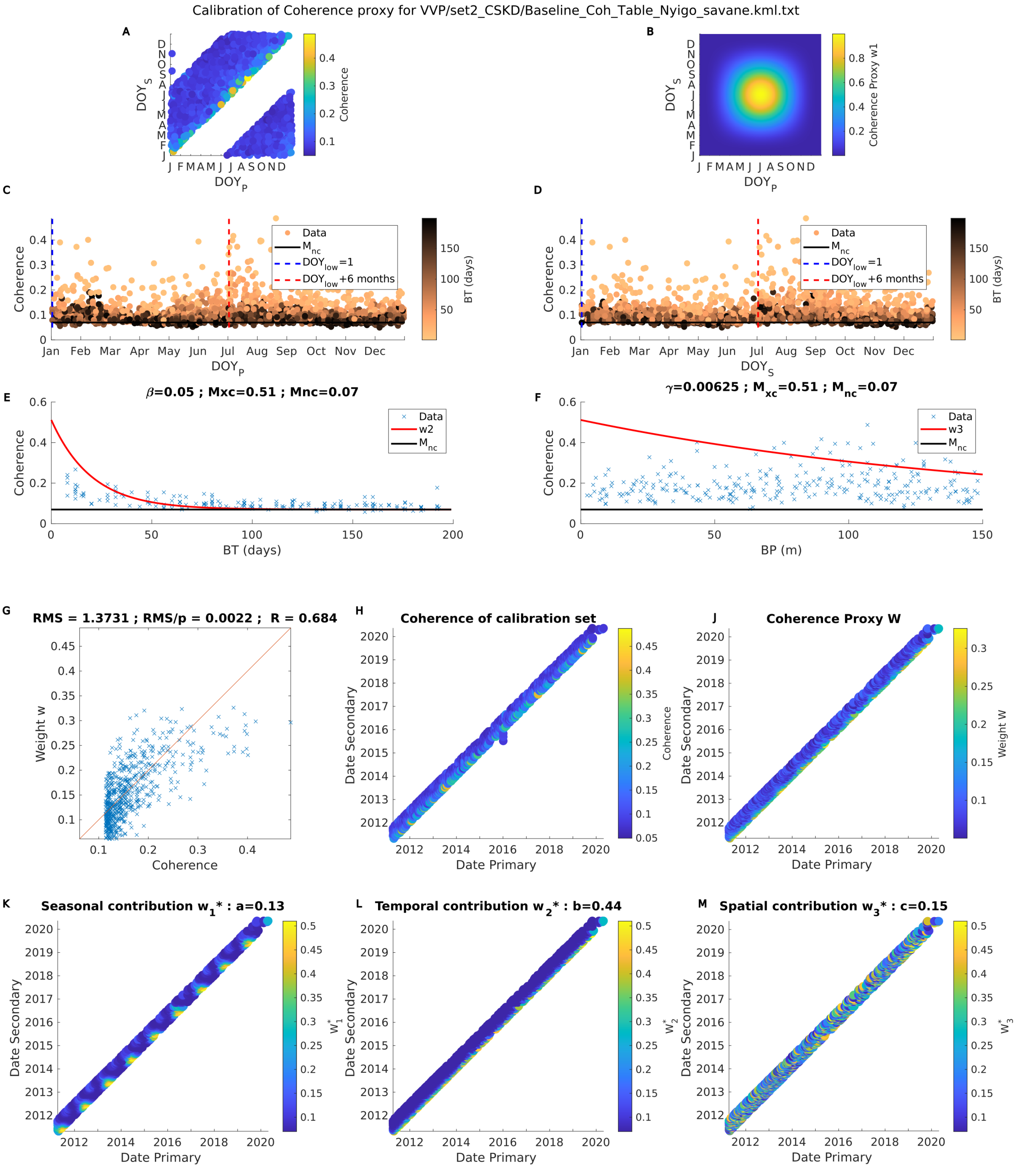

Figure S13. Calibration of the coherence proxy on the VVP area for CSK descending dataset. A and B represent the mean coherence computed on the Savannah ROI and the seasonal contribution w_1 to the coherence proxy as a function of the primary and secondary image day of year of each pair. C and D represent the mean coherence computed on the Savannah ROI as a function of the Primary (Secondary) image Day of year respectively. Blue and red dash lines mark DOY_{low} and $DOY_{low}+6$ months, respectively. E and F represent the mean coherence computed on the Savannah ROI as a function of BT for pairs with $BP < 15$ m and as a function of BP for pairs with $BT < 25$ days, respectively. Red line represent the exponential decrease modeled by w_2 and w_3 in the coherence proxy. C to F, horizontal black line marks the minimum expected value of the coherence M_{nc}. G represents the coherence proxy w versus the mean coherence computed on the Savannah ROI for each pair of the calibration set. The red line marks the first bisector. H and J represent the the mean coherence computed on the Savannah ROI for each pair of the calibration set and the coherence proxy w, respectively, as a function of the primary and secondary image acquisition date. K to M, the color scale represent the partial weights w_1, w_2 and w_3 (see text section 2.3.1.2), respectively. The value of the corresponding coefficients a, b, and c is indicated on top of each plot.

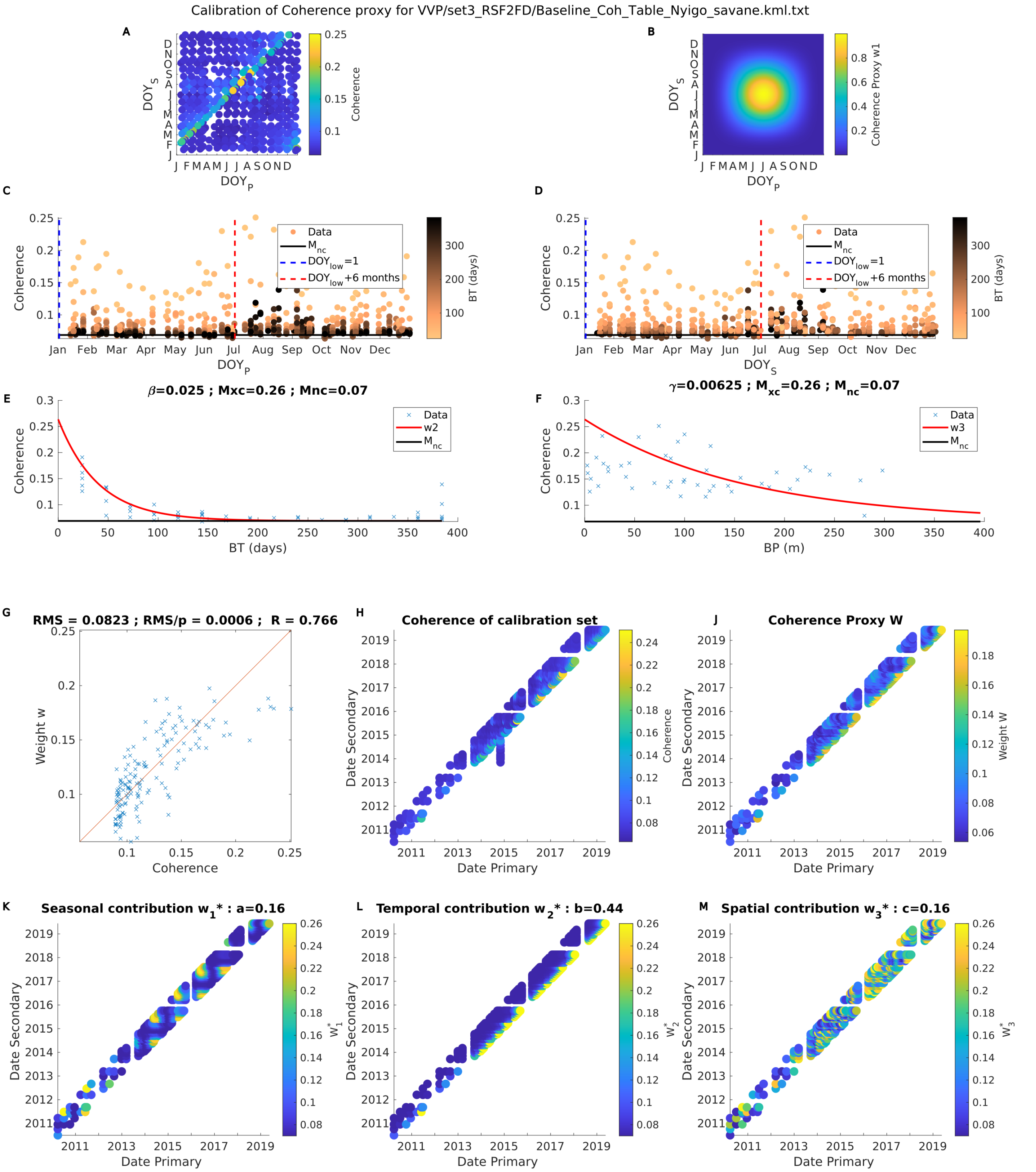

Figure S14. Calibration of the coherence proxy on the VVP area for RS F2F descending dataset. A and B represent the mean coherence computed on the Savannah ROI and the seasonal contribution w_1 to the coherence proxy as a function of the primary and secondary image day of year of each pair. C and D represent the mean coherence computed on the Savannah ROI as a function of the Primary (Secondary) image Day of year respectively. Blue and red dash lines mark DOY_{low} and $DOY_{low} + 6$ months, respectively. E and F represent the mean coherence computed on the Savannah ROI as a function of BT for pairs with $BP < 15$ m and as a function of BP for pairs with $BT < 25$ days, respectively. Red line represent the exponential decrease modeled by w_2 and w_3 in the coherence proxy. C to F, horizontal black line marks the minimum expected value of the coherence M_{nc}. G represents the coherence proxy w versus the mean coherence computed on the Savannah ROI for each pair of the calibration set. The red line marks the first bisector. H and J represent the the mean coherence computed on the Savannah ROI for each pair of the calibration set and the coherence proxy w, respectively, as a function of the primary and secondary image acquisition date. K to M, the color scale represent the partial weights w_1, w_2 and w_3 (see text section 2.3.1.2), respectively. The value of the corresponding coefficients a, b, and c is indicated on top of each plot.

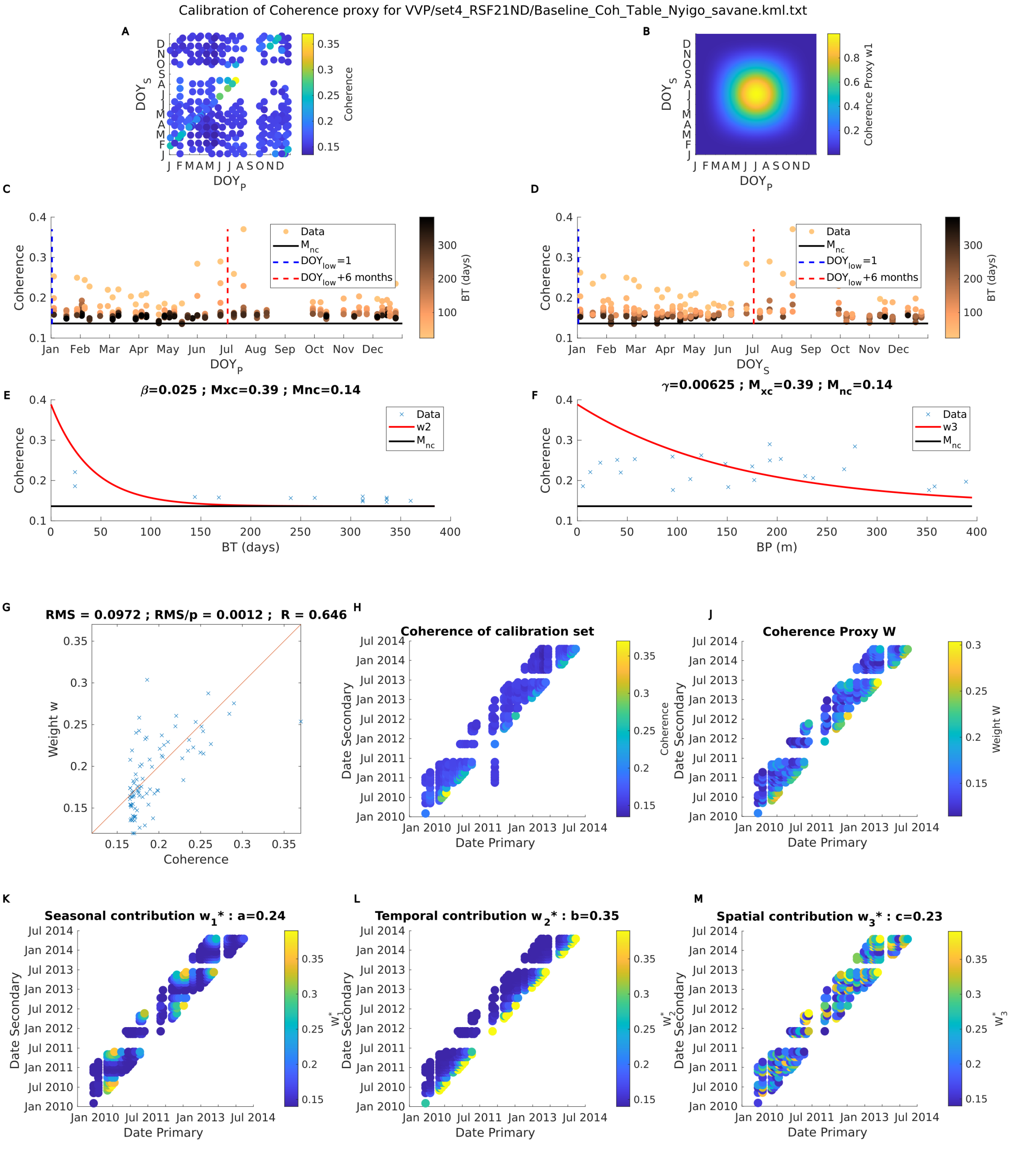

Figure S15. Calibration of the coherence proxy on the VVP area for RS F21N descending dataset. A and B represent the mean coherence computed on the Savannah ROI and the seasonal contribution w_1 to the coherence proxy as a function of the primary and secondary image day of year of each pair. C and D represent the mean coherence computed on the Savannah ROI as a function of the Primary (Secondary) image Day of year respectively. Blue and red dash lines mark DOY_{low} and $DOY_{low} + 6$ months, respectively. E and F represent the mean coherence computed on the Savannah ROI as a function of BT for pairs with $BP < 15$ m and as a function of BP for pairs with $BT < 25$ days, respectively. Red line represent the exponential decrease modeled by w_2 and w_3 in the coherence proxy. C to F, horizontal black line marks the minimum expected value of the coherence M_{nc}. G represents the coherence proxy w versus the mean coherence computed on the Savannah ROI for each pair of the calibration set. The red line marks the first bisector. H and J represent the the mean coherence computed on the Savannah ROI for each pair of the calibration set and the coherence proxy w, respectively, as a function of the primary and secondary image acquisition date. K to M, the color scale represent the partial weights w_1, w_2 and w_3 (see text section 2.3.1.2), respectively. The value of the corresponding coefficients a, b, and c is indicated on top of each plot.

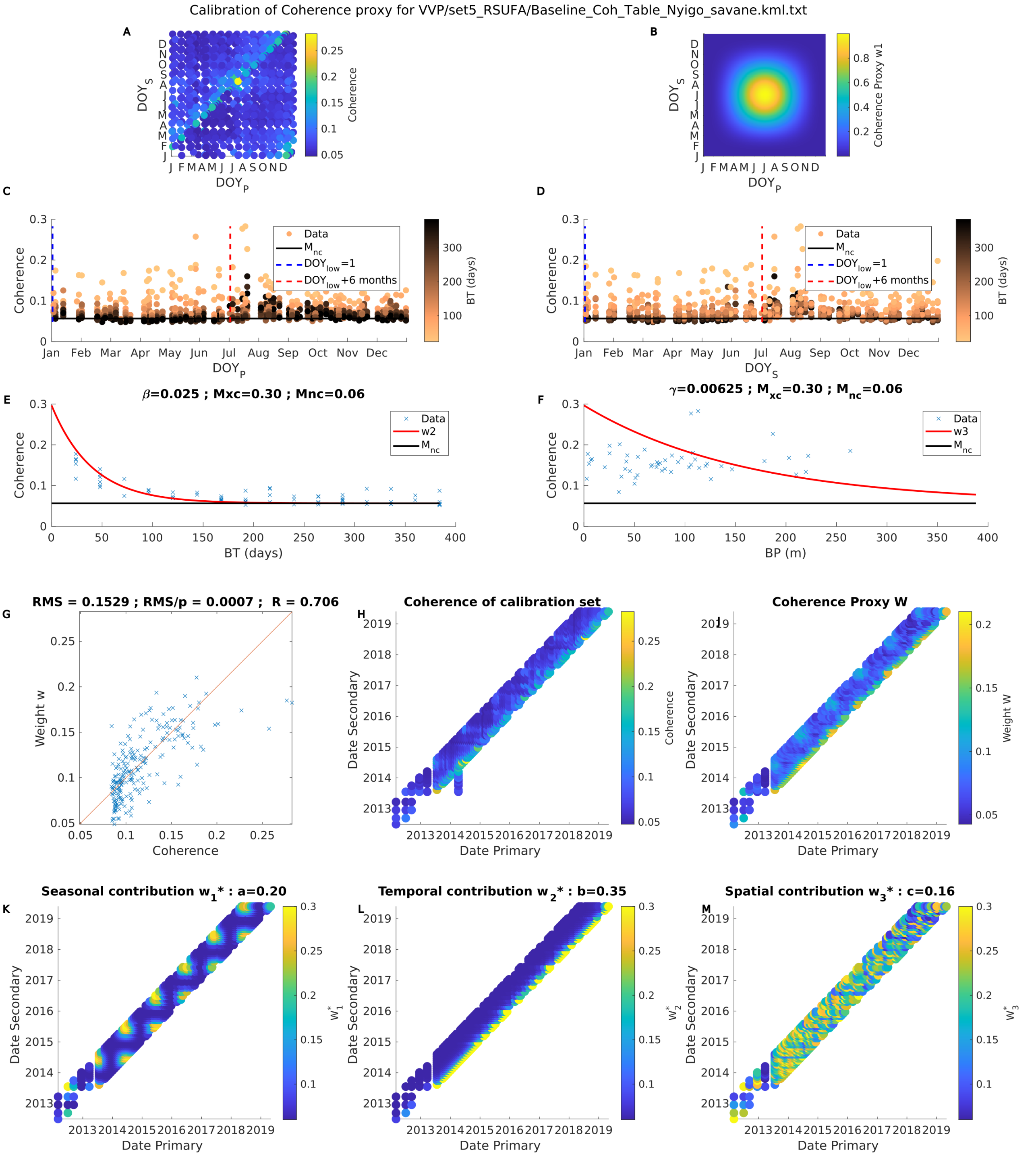

Figure S16. Calibration of the coherence proxy on the VVP area for RS UF ascending dataset. A and B represent the mean coherence computed on the Savannah ROI and the seasonal contribution w_1 to the coherence proxy as a function of the primary and secondary image day of year of each pair. C and D represent the mean coherence computed on the Savannah ROI as a function of the Primary (Secondary) image Day of year respectively. Blue and red dash lines mark DOY_{low} and $DOY_{low} + 6$ months, respectively. E and F represent the mean coherence computed on the Savannah ROI as a function of BT for pairs with $BP < 15$ m and as a function of BP for pairs with $BT < 25$ days, respectively. Red line represent the exponential decrease modeled by w_2 and w_3 in the coherence proxy. C to F, horizontal black line marks the minimum expected value of the coherence M_{nc}. G represents the coherence proxy w versus the mean coherence computed on the Savannah ROI for each pair of the calibration set. The red line marks the first bisector. H and J represent the the mean coherence computed on the Savannah ROI for each pair of the calibration set and the coherence proxy w, respectively, as a function of the primary and secondary image acquisition date. K to M, the color scale represent the partial weights w_1, w_2 and w_3 (see text section 2.3.1.2), respectively. The value of the corresponding coefficients a, b, and c is indicated on top of each plot.

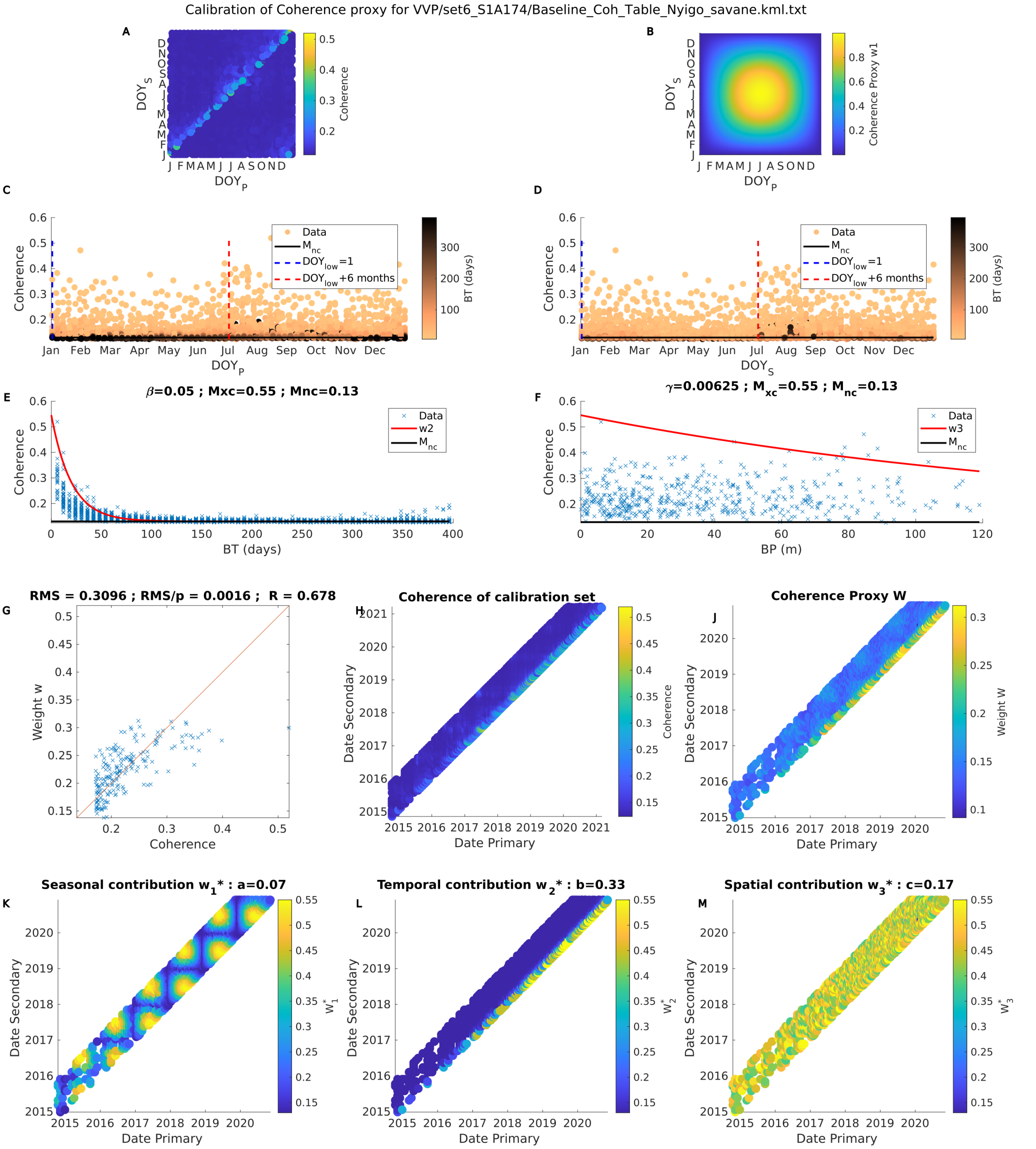

Figure S17. Calibration of the coherence proxy on the VVP area for S1 ascending dataset. A and B represent the mean coherence computed on the Savannah ROI and the seasonal contribution w_1 to the coherence proxy as a function of the primary and secondary image day of year of each pair. C and D represent the mean coherence computed on the Savannah ROI as a function of the Primary (Secondary) image Day of year respectively. Blue and red dash lines mark DOY_{low} and $DOY_{low} + 6$ months, respectively. E and F represent the mean coherence computed on the Savannah ROI as a function of BT for pairs with $BP < 15$ m and as a function of BP for pairs with $BT < 25$ days, respectively. Red line represent the exponential decrease modeled by w_2 and w_3 in the coherence proxy. C to F, horizontal black line marks the minimum expected value of the coherence M_{nc}. G represents the coherence proxy w versus the mean coherence computed on the Savannah ROI for each pair of the calibration set. The red line marks the first bisector. H and J represent the the mean coherence computed on the Savannah ROI for each pair of the calibration set and the coherence proxy w, respectively, as a function of the primary and secondary image acquisition date. K to M, the color scale represent the partial weights w_1, w_2 and w_3 (see text section 2.3.1.2), respectively. The value of the corresponding coefficients a, b, and c is indicated on top of each plot.

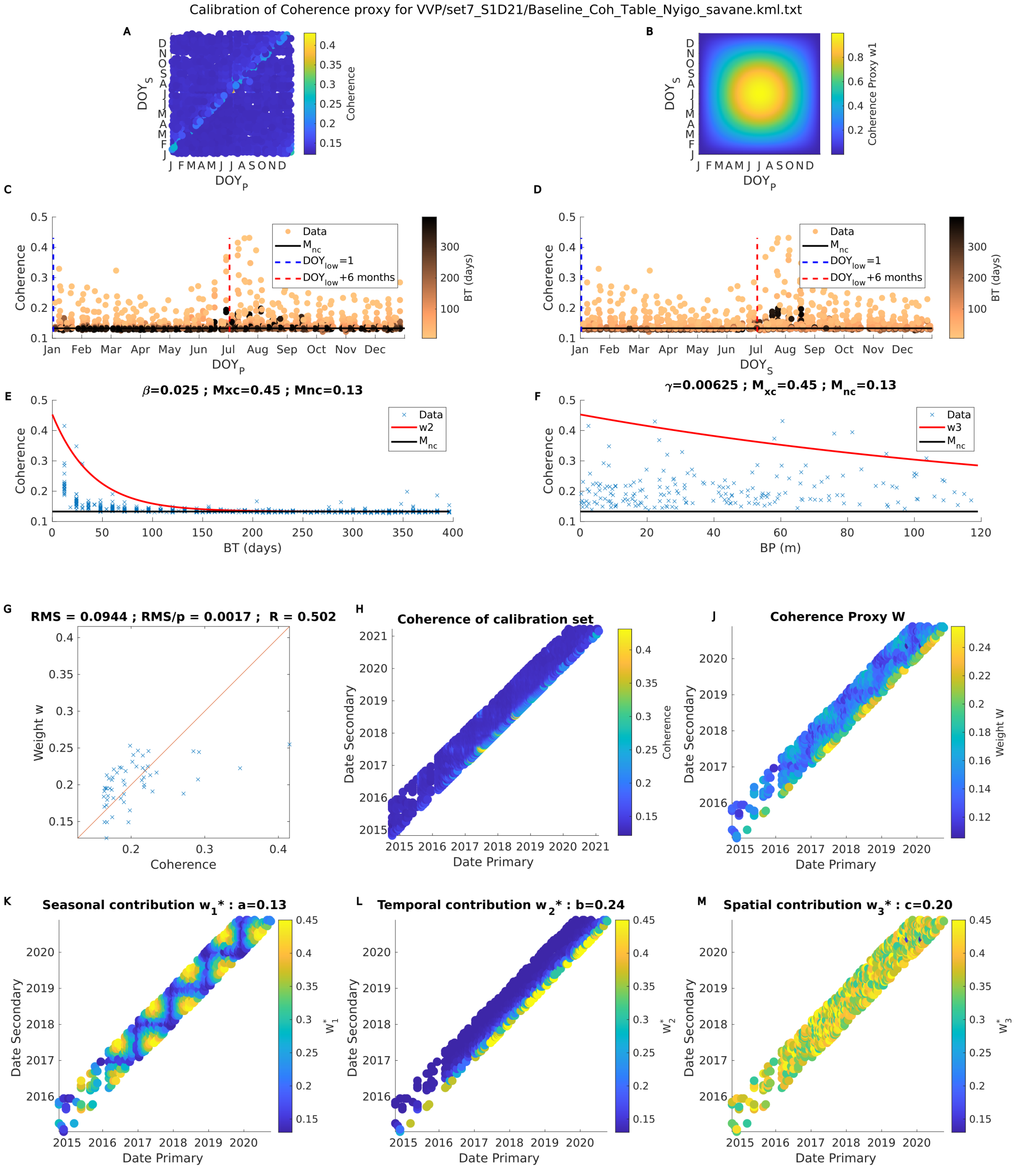

Figure S18. Calibration of the coherence proxy on the VVP area for S1 descending dataset. A and B represent the mean coherence computed on the Savannah ROI and the seasonal contribution w_1 to the coherence proxy as a function of the primary and secondary image day of year of each pair. C and D represent the mean coherence computed on the Savannah ROI as a function of the Primary (Secondary) image Day of year respectively. Blue and red dash lines mark DOY_{low} and $DOY_{low} + 6$ months, respectively. E and F represent the mean coherence computed on the Savannah ROI as a function of BT for pairs with $BP < 15$ m and as a function of BP for pairs with $BT < 25$ days, respectively. Red line represent the exponential decrease modeled by w_2 and w_3 in the coherence proxy. C to F, horizontal black line marks the minimum expected value of the coherence M_{nc}. G represents the coherence proxy w versus the mean coherence computed on the Savannah ROI for each pair of the calibration set. The red line marks the first bisector. H and J represent the the mean coherence computed on the Savannah ROI for each pair of the calibration set and the coherence proxy w, respectively, as a function of the primary and secondary image acquisition date. K to M, the color scale represent the partial weights w_1, w_2 and w_3 (see text section 2.3.1.2), respectively. The value of the corresponding coefficients a, b, and c is indicated on top of each plot.

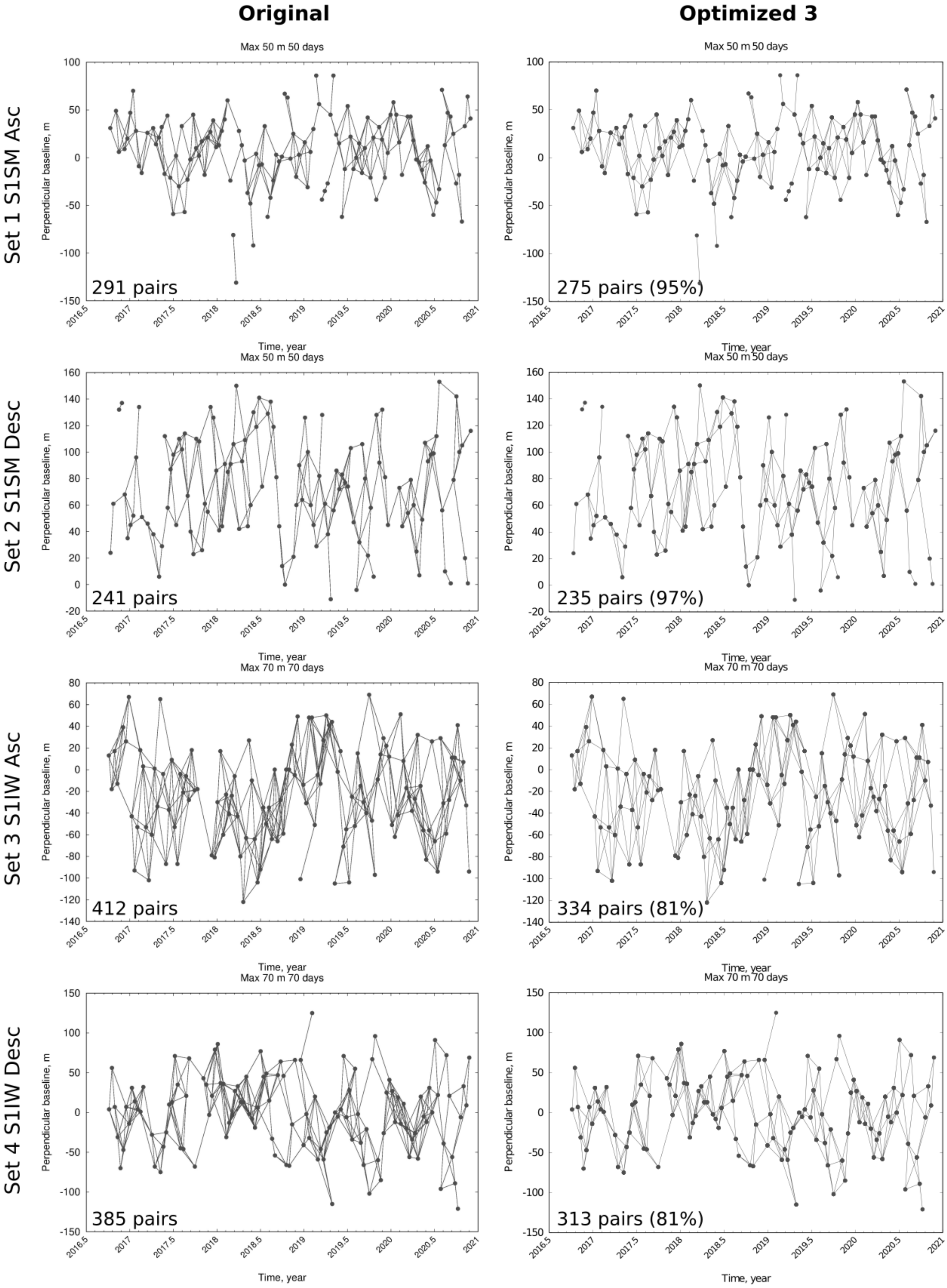

Figure S19. Original and optimized baseline plots for each S1 data set acquired on la Reunion Island. The optimization is performed with $k = 3$. The total number of pairs used is indicated in the lower left corner of each baseline plot. The percentage of pairs kept after optimization with reference to the original processing is also indicated in the optimized baselines plots. Initial baselines criteria (BP and BT) are indicated on top of each baseline plot.

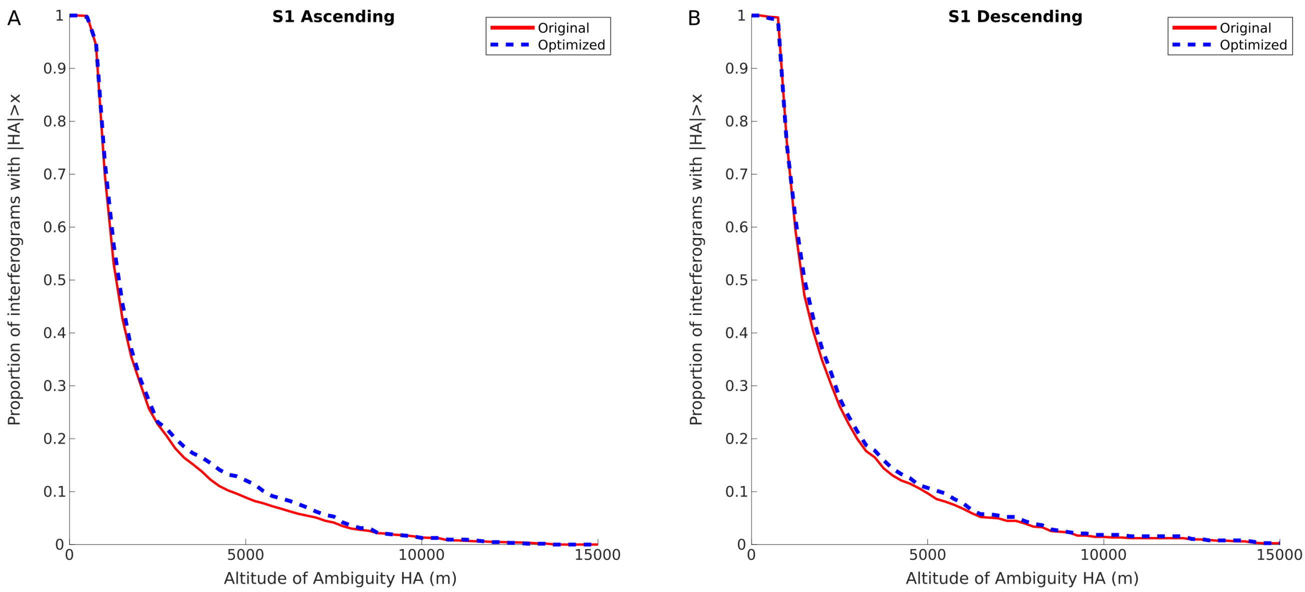

Figure S20. Proportion of interferograms in the original data set (plain red curve) and in the optimized dataset (dashed blue curve) which altitude of ambiguity is larger than a threshold as a function of the threshold. A. Sentinel 1 Ascending dataset on the VVP. B. S1 Descending dataset on the VVP.

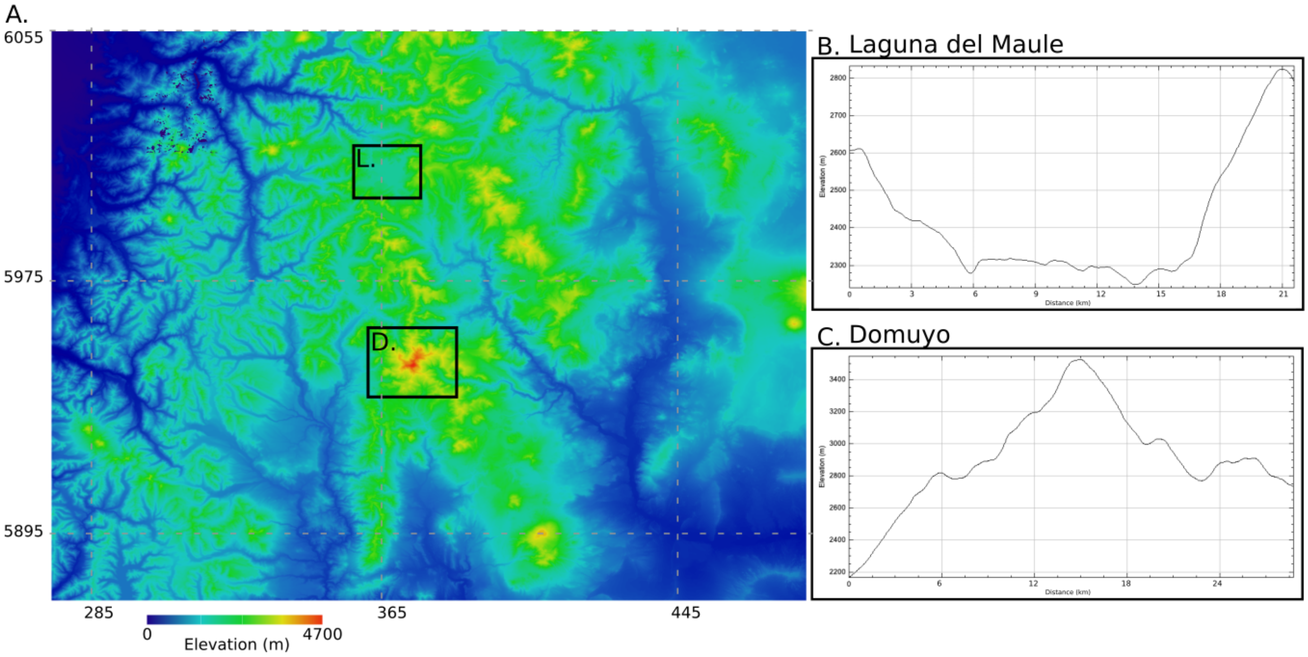

Figure S21. (A) Digital Elevation Model (SRTM 30 m) of the Domuyo and Laguna del Maule region. (B) and (C) Average East-West elevation profiles computed on the areas marked by black rectangles in (A) : the Laguna del Maule (rectangle L.) and Domuyo (rectangle D.) respectively.

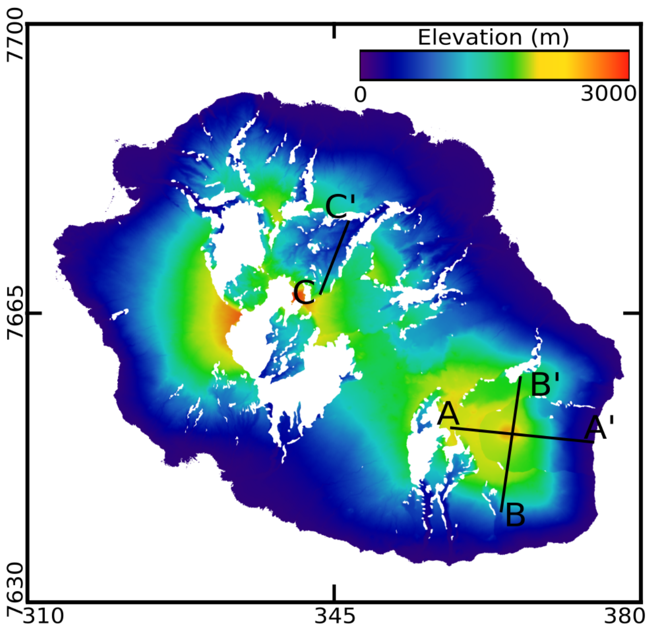

Figure S22. Digital Elevation Model (SRTM 30 m) of the Reunion Island.

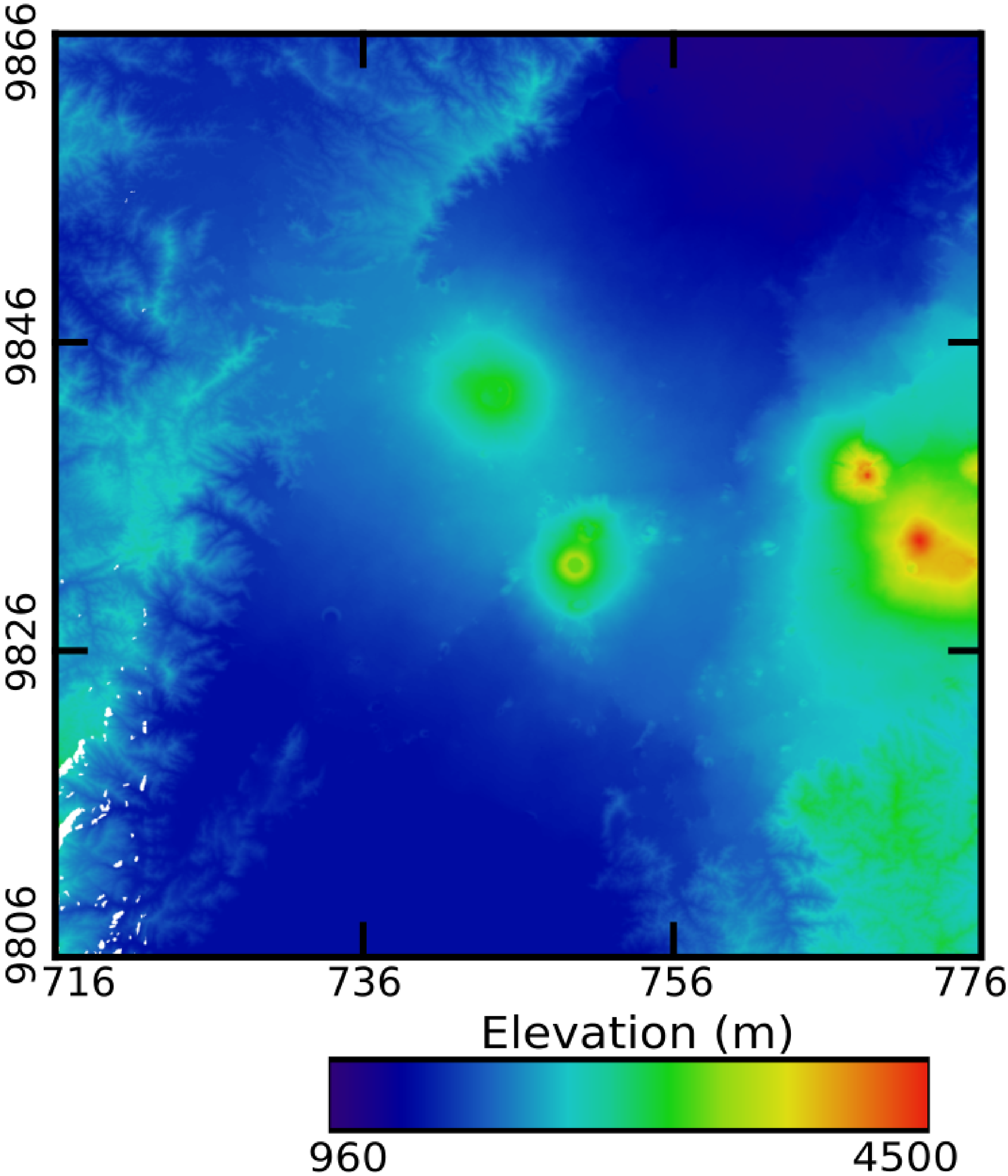

Figure S23. Digital Elevation Model (SRTM 30 m) of VVP.