 doi:10.20944/preprints202206.0252.v1

Technical Note

An efficient global scale Sentinel-1 radar backscatter and interferometric processing system

Piyush S. Agram [1]*, Michael S. Warren [1], Matthew T. Calef [1] and Scott A. Arko [1]

[1] Descartes Labs, 1607 Paseo de Peralta, Suite B, Santa Fe, NM, 87501, USA
* Correspondence: piyush@descarteslabs.com

Abstract: We describe an efficient and cost effective data access mechanism for Sentinel-1 TOPS mode bursts. Our data access mechanism enables burst-based data access and processing, thereby eliminating ESA's Sentinel-1 SLC data packaging conventions as a bottleneck to large scale processing. Pipeline throughput is now determined by available compute resources and efficiency of the analysis algorithms. For targeted infrastructure monitoring studies, we are able to generate coregistered, geocoded stacks of SLCs for any AOI in the world in a few minutes. In addition, we describe our global scale radar backscatter and interferometric products and associated pipeline design decisions that ensure geolocation consistency across the suite of derived products from Sentinel-1 data. Finally, we discuss the benefits and limitations of working with geocoded SAR SLC data.

Keywords: InSAR; SAR; Sentinel-1

1. Introduction

Until a few years ago, transformation of Synthetic Aperture Radar (SAR) data to higher level products was considered a specialty of niche academic and engineering groups, largely due to limited availability of this type of data and associated restricted use licenses. The Sentinel-1 synthetic aperture radar (SAR) constellation [1], part of Europe's Copernicus Earth Observation programme, has significantly changed this situation. Data acquired by this constellation have been freely available for end users since October 2014, enabling local and global scale application users to reliably access and analyze SAR imagery over their areas of interest (AOIs). A number of wide-area applications have already been built on top of Sentinel-1's data archive, e.g [2–5]

While Sentinel-1 data are open and freely accessible, there are still challenges in efficiently using this dataset. The first is that the data are available via large-zip files that do not offer sufficiently fine-grained access. The second is that, once accessed, the data are presented in a non-Geographic Information System (GIS) compatible coordinate system that is not easily converted to more conventional coordinate systems. SAR processing tool suites, e.g. GMTSAR [6], perform most operations in this non-GIS coordinate system, and re-project results to a GIS coordinate system as a final step.

In our organization, Descartes Labs, we often build and deploy analytic models that use multiple sources of remote sensing data, and where GIS compatible coordinate systems provide a lingua franca. Products derived from SAR backscatter and interferometric SAR (InSAR) often play a crucial role in these models. Providing fine grained access to SAR data in a GIS compatible coordinate system makes it easier to build and deploy these models – essentially reducing SAR and InSAR data to "just another" data source.

To address this need for fine grained access to SAR and InSAR data in a GIS compatible coordinate system, we have developed a fast data access mechanism for Sentinel-1 Single Look Complex (SLC) datasets. This data access mechanism is efficient, and operates in a granular fashion similar to the way in which Cloud Optimized Geotiff's (COGs) are used for optical imagery. Additionally we have built a geocoding mechanism to faithfully represent SLC data on a GIS coordinate system. Powered by the data access and geocoding mechanisms, we have developed rapid global scale radar backscatter and interferometric processing pipelines. In this manuscript, we describe our data access mechanism, geocoder and the salient features of our processing pipelines.

Citation: Agram, P. S.; Warren, M. S.; Calef, M. T.; Arko, S. A. An efficient global scale Sentinel-1 radar backscatter and interferometric processing system. *Preprints* **2022**, *1*, 0. https://doi.org/

 doi:10.20944/preprints202206.0252.v1

This manuscript is organized as follows. Section 2 describes our fast Sentinel-1 SLC data access mechanism. Section 3 describes our geocoding approach and also discusses the benefits and limitations of using such a processing approach. Section 4 describes our global scale products and associated pipeline design decisions. In Section 5, we discuss the relevance to other SAR missions of the methods presented here.

2. Sentinel-1 data access

The Sentinel-1 mission images most land masses in the Interferometric Wide (IW) swath Terrain Observation with Progressive Spans (TOPS) imaging mode [1]. While the mission does use conventional stripmap (SM) and wave (WV) imaging modes over land, SM and WV collections represent a tiny fraction of total data volume acquired. In this manuscript, we only focus on data acquired in the most frequently used IW imaging mode. The concepts presented here for IW mode are also relevant for the less frequently used Extended Wide (EW) swath imaging mode imagery. Sentinel-1 IW mode Level-1 imagery is released in two flavors by ESA [7] - Single Look Complex (SLC) granules and Ground Range Detected (GRD) granules.

Because GRD data can be derived from SLC data, and because we must support both coherent and amplitude-based analyses, we prefer using SLC data over GRD data. The rest of the discussion will, therefore, focus on SLC data.

2.1. Global burst footprint map

One of the important features of Sentinel-1 SLC data is that the footprints corresponding to SLC granules are not fixed over time. However, the footprints of the contained bursts are stationary, thanks to the burst synchronization feature of the mission [1]. These burst footprints repeat almost exactly and can be used as a basis to organize temporal stacks of Sentinel-1 imagery [8]. This is similar to Landsat's Path-Row or Sentinel-2's tiling scheme.

We built our own first version of Sentinel-1's global burst footprint map in August 2020 using the radar metadata in annotation files from all Sentinel-1 SLC imagery from Sep 2014 to July 2020. Subsequently, we adopted a naming convention (See Table 1) for individual burst footprints and for individual burst SLCs. Here, we refer to a burst footprint as an entry in our map corresponding to the geometric extent of an imaged area, and a burst SLC as an entry corresponding to particular SLC image for that burst. Note that the unique identifiers of individual burst SLCs includes the identifier of the corresponding footprint. Each of our Level-1 burst footprints have also been mapped to European Space Agency (ESA) burst polygons that were released in June 2022 [8]. This database of burst SLCs connects burst-level metadata, geometric extents for the burst and the ESA SLC granule from which the burst can be accessed.

The choice of the function of the ascending node time in our naming convention (see Table 1) need only be consistent for all labeling operations. Our burst databases are dynamic products that get updated automatically with new footprints and burst SLCs as imagery from ESA is ingested by our system. Once we had the initial footprint data, we ingested all historical SLC data and currently, keep up with incoming data from ESA. Our burst database can be searched with our identifiers as well as ESA identifers and we will continue labeling data using our mechanism even though ESA has updated its baseline processor [8,9] to include unique burst ids in the metadata since IPF 3.5.x, as this acts as an additional Quality Control (QC) check in our ingest pipelines.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2022 doi:10.20944/preprints202206.0252.v1

Table 1. Naming convention for Sentinel-1 burst footprints and individual bursts.

Product	Convention
Burst Footprint	TTT-AAAA-IIW-PP-VU
Individual Burst SLC	YYYY-MM-DD-SSS-TTT-AAAA-IIW-PP-VU-HASH
Field	**Description**
YYYY-MM-DD	Acquisition date
SSS	Mission name (e.g, S1A)
TTT	Track number (001-175)
AAAA	Function of time since ascending node
II	Imaging mode (e.g, IW)
W	Swath number (1-5)
PP	Polarization. Always VV for footprint
V	Looks side (R for Right, L for Left)
U	Pass direction (A for asc, D for desc)
HASH	4-char hash of the source SAFE granule

2.2. Rapid access to a single burst

Labeling all the global Sentinel-1 SLC data provides an indexing structure for our databases. While it let us organize our workflows, it does not speed up processing as we still needed to extract bursts from within TIFF files contained in SAFE zip archives. To address this, we adopted techniques used in the neuro-imaging community to build tools to enable random access within large zip archives (e.g., [10]). Note that this approach requires we scan the entire zip file once. We store the state of the zip decoder at various locations corresponding to individual bursts along with the radar metadata as part of our data ingest pipeline. With this setup, we are able to access radar metadata for any burst (in any polarization) without having to access large SAFE archives, and are able to decompress fragments of SAFE zip files client-side and quickly extract the associated imagery. This lookup mechanism can be used with any store of Sentinel-1 SLCs as long original zip files from ESA have not been modified. We are able to pull any burst ($\sim$ 80 MB compressed) from SLC SAFE zip files residing in our cloud storage buckets in 2-3 seconds and pull the same data from SLC SAFE zip files residing in NASA's Alaska Satellite Facility (ASF) archive in 6-7 seconds. This rapid data access mechanism lets us process backscatter and InSAR products on the live stream of SLC SAFE zup files from ESA very efficiently. This capability also allows us to rapidly test new analysis methods - especially interferometric and polarimetric methods at scale over large regions or time spans.

Rapid SLC data access has been identified as a bottleneck by various SAR user groups, including the Copernicus Land Monitoring Service (See Appendix F of [11]) and we believe the mechanism presented here is a scalable solution to address this. Some common use cases and benefits of our data access mechanism are listed here:

1. Processing tasks don't require nodes that can read an entire SLC SAFE file. This enables finer-grained workers and increased parallelism.
2. This rapid access allows us to run near-realtime pipelines that generate backscatter and InSAR products globally, with a latency of under a day.
3. For wide area analysis, this eliminates expensive and time-consuming data replication into shared file systems or large scratch disks and enables access to SLC data as part of the processing workflow without delay.
4. This enables access to co-pol imagery without having to transfer associated cross-pol imagery as well. This cuts down data transfer by half compared to approaches staging entire SAFE archives for interferometric applications.
5. This allows us to to predetermine burst footprints that don't contain land and completely avoid transferring that data for certain applications, further cutting down data transfer by almost 30%.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2022 doi:10.20944/preprints202206.0252.v1

Note that this data access mechanism will speed up any processing pipeline, irrespective of the manner in which the SLC data is used as it enables granular access in parallel to individual bursts.

3. Geocoded bursts

In the geospatial community, SAR datasets are notorious for their relatively complicated data distribution formats and need for custom processing tools to transform them into analysis-ready products. In this section, we describe in detail our geocoding process, which enables us to bring SAR and InSAR datasets into standard geospatial frameworks. We also provide analogies in terms of well known GIS concepts, in hope that this will assist in demystifying SAR data to users who are not familiar with this type of data and for future adopters of this technology.

3.1. Projection system of a SAR image

Sentinel-1 Level-1 SLC images are organized as a collection of bursts (Section 2). Each SLC burst, though not very obvious to users familiar with working on other types of geospatial datasets, has a well defined projection system. The primary challenge for Geographical Information System (GIS) software in working with SAR data is that this projection cannot be represented by a standard EPSG code [12] or a simple PROJ string [13]. Hypothetically, if we were to write a PROJ string [13] representation for a burst SLC, it would look something like:

```
+proj=zerodoppler
+side=right
+range=slant
+orbit=t1,x1,y1,z1,vx1,vy1,vz1;t2,x2,y2,z2,vx2,vy2,vz2,...
```

The *zero doppler* system indicates that pixels are arranged in an orthogonal grid with an axis that is tangential to the satellite's orbit - this is the system used by Sentinel-1 [7,9]. Most commonly used SAR datasets, such as those collected by Sentinel-1, the TerraSAR-X constellation, the COSMO-SkyMed constellation, ICEYE, Capella, RISAT etc, all use the zero doppler system [14]. The side argument indicates if the platform is looking left or right. The range argument indicates that the data is on a uniform slant range grid. Note that, while coordinates in a radar image are represented by two unique indices - row and column, the projection system is a 3D Coordinate Reference System (CRS), i.e., coordinate transformation depends on altitude of the point. In most cases, a reference height is assumed for the pixels in the radar image. Another reason our hypothetical projection system above cannot be treated as a standard system is because many points in physical 3D space (or standard CRS) map to the same radar pixel - when these points occur on the surface of the Earth, this effect is known as layover [15]. This can happen when, e.g., the SAR system is looking perpendicular to the side of a incline, and all points on the incline have the same range. However, any point in 3D physical space (or standard CRS) maps to a unique point in the SAR projection system. Note that the geolocation of points in the SAR image requires this orbit information, which we have included above as a semi-colon separated list of time-tagged Cartesian positions and velocities for simplicity. Often this information is read in from external orbit files or from metadata included with the radar imagery.

The key implication from this PROJ string analogy is that any change in the projection parameters, notably the orbit state vectors, above essentially represents a different coordinate system. A direct consequence of this is that every SAR image, even if acquired over the same AOI, inherently has a different projection system since satellite orbits do not repeat perfectly. One of the core features of SAR processing software is support for the transformation of points in standard CRS to SAR projection system and vice-versa.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2022 doi:10.20944/preprints202206.0252.v1

3.2. Geotransform of a SAR image

An affine geotransform [16] is a standardized way of associating map coordinates with pixels of a raster image. Continuing our effort to describe SAR data sets in a GIS-compatible manner, we can write the geotransform for a SAR SLC image that uses the projection system above as:

$$\begin{bmatrix} \text{slant range to first pixel in m} \\ \text{slant range pixel spacing in m} \\ 0 \\ \text{zero doppler time of first line in UTC} \\ 0 \\ \text{azimuth time interval in s} \end{bmatrix}$$

Many SAR constellations operate in near-polar orbits, and for such orbits the along-track time is roughly equivalent to the Y-coordinate or latitude and slant range from the satellite's orbit is roughly equivalent to the X-coordinate or longitude, if we were to compare this representation with geospatial datasets that are distributed in UTM or geographic coordinate systems. In case of GRD images, ground range replaces slant range range in the notation above and the interpretation presented here manuscript still applies.

3.3. Coregistration as reprojection

As described above, each SAR image is essentially distributed in its own projection. Hence, coregistering one SAR image to another acquired over the same AOI is essentially, in GIS terminology, a reprojection operation. A minor complication in this reprojection is the nature of SAR imaging itself. A SAR image represents a 2D representation of a 3D world - and here we are trying to coregister two different projections of the same underlying 3D world; this is akin to aligning photographs acquired from different perspectives. Consequently, aligning these images precisely requires a model of the 3D world, i.e, a digital elevation model (DEM) [17].

In traditional processing approaches, the projection system (also known as radar geometry) of one of the images from a stack of imagery (usually from same cluster of imaging geometries) is used as the reference system onto which one projects the other images, e.g., [18,19]. This approach involves a fair amount of bookkeeping and managing pixel-by-pixel transformations via a known coordinate system of the undelying DEM and is a core feature of SAR processing software.

In our approach, we use an UTM-based system as the common projection system for our Sentinel-1 products and we project every burst SLC to this common system. This approach is computationally efficient as we reduce the number of coordinate transforms and can use the DEM grid as our reference grid directly. This operation can be optimized to work efficiently on single CPUs and without heavy memory requirements [20]. Note also that the mapping from UTM coordinate to point in 3D space occurs by appending the elevation, as such it is only as smooth as the DEM itself. However, the mapping from a point in 3D space onto the SAR images coordinate system is as smooth as the orbit, and supports accurate interpolation. Once this process is complete we have geocoded imagery in a well known CRS, and we are able to use the same set of data manipulation tools that we have developed on optical and other geospatial datasets, and tuned over the years, with our SAR datasets [21]. In other words, we have reduced SAR geocoding to an enhanced 'warp' operation [16] with custom interpolators that have been designed to preserve phase and amplitude properties while accounting for any carriers on the data [22]. Note however, that any further interpolation of these data must respect the phase of the underlying data. These phases are nearly random and most interpolation schemes beyond down-sampling are not safe.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2022 doi:10.20944/preprints202206.0252.v1

3.4. Preserving signal fidelity

In accordance with standard signal processing principles, we oversample when geocoding the data to avoid artifacts due to aliasing particularly during interferometric processing. We geocode Sentinel-1 bursts to an aligned grid with 10m Northing x 2.5m Easting grid spacing even though the SLC data has a ground posting of ∼14m x ∼5m. Note that we oversample significantly more in easting (approx. slant range) to retain features in steep terrain to mimic the natural anisotropy of the range-doppler grid.

The possibility of considering a well known CRS as a common projection system for stacking SAR imagery has only been possible due to the massive improvements in SAR metadata quality and auxiliary datasets in the last decade. Notably:

1. Our coordinate transforms rely on good DEMs and a global scale high quality DEM was not broadly available until the SRTM [23] data became available freely and publicly less than twenty years ago.
2. Quality of the satellite orbit data, which directly impact the projection system, has also significantly improved. With on-board GPS receivers these errors are on the order of a few cm for Sentinel-1 [24] and largely in the along-track direction. This is a massive improvement over the previous generation sensors where the errors were on the order of few m. This is equivalent to having more accurate projection information for generic geospatial data.
3. The accuracy of the geotransform in the metadata, especially in the along-track direction has also significantly improved in the last decade due to better clock synchronization. Working with previous generation sensors like ERS and Envisat, users would often encounter shifts in along-track / Northing direction during geocoding.
4. The accuracy of the slant range component of the geotransform depends on the two-way propagation delay through the atmosphere. Delay through the atmosphere exhibits a variance of a few 10s of cm [25,26], which is an order of magnitude smaller than Sentinel-1's range resolution and has minimal impact on this dataset. For readers well versed with SAR processing, we would like to note that if standardized slant range delay estimates due to the troposphere and ionosphere [27], or along track timing corrections derived from Enhanced Spectral Diversity (ESD) [28] were to be made systematically available on a global scale, it would be straightforward to incorporate these during geocoding in our pipelines.

We also note that, in well known CRSs, an equivalent pixel-by-pixel geometric transformation can always be determined for computing additional useful parameters like incidence angles, interferometric baselines etc without loss of precision.

In our global scale pipelines, we first calibrate the amplitudes of burst SLCs to correspond to thermal-noise correct $\sigma_{0,E}$ using lookup tables in the radar metadata before geocoding the imagery on an UTM grid, while carefully accounting for phase ramps in the data [22]. The phase of geocoded data is also modulated with the slant range inferred from a DEM [29] to make our geocoded data readily usable for interferometric applications.

3.5. Spatial Averaging

The presented approach also ensures consistent geocoding between suites of products derived from geocoded SLCs. SAR and InSAR workflows often involve a spatial averaging or a multi-looking step to help alleviate speckle effects. In traditional processing approaches, this spatial averaging is often performed in original radar coordinates (e.g, when generating GRDs from SLCs) and data is then resampled onto a uniform grid. This often leads to averaging of data from different segments on the ground and the differences can get further exaggerated in heterogeneous terrain by the geocoding operation, which is avoided with our approach.

This effect is understood within the InSAR community and one of the commonly used methods to alleviate this effect is to group acquisitions by satellite geometries and coregister every image in each group to a chosen reference image. This stacking approach significantly

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2022 doi:10.20944/preprints202206.0252.v1

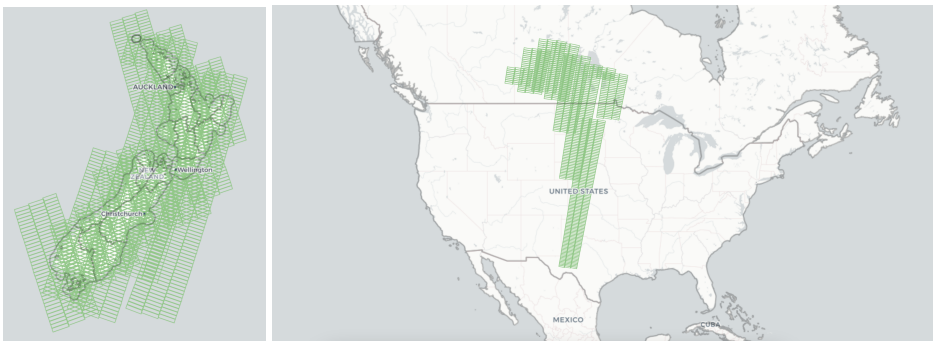

Figure 1. Examples from our global footprint database. (Left) Sentinel-1 burst footprints acquired over New Zealand till Jun 1, 2022. (Right) New previously un-imaged footprints between Dec 24, 2021 to Jun 1, 2022 over North America in response to Sentinel-1B outage in Dec 2022.

reduces phase closure artifacts in InSAR analysis compared to using interferograms that were generated pair-by-pair with different reference images. Note that even in this case, the same issue resurfaces when comparing averaged results across satellite tracks / geometry groups. In our approach, we always geocode the SAR image to a oversampled high resolution regular grid (10m Northing x 2.5m Easting) first and then average signals from the same segments on the ground, consistently to the extent afforded by the accuracy of the DEM used. This essentially allows us to generate high quality global backscatter (10m) and InSAR coherence (20m) datasets that are very well coregistered on the live Sentinel-1 datastream from ESA. We currently are unaware of any other complementary backscatter and InSAR dataset with similar coregistration properties.

4. Global scale Sentinel-1 products

In this section, we describe the various global scale products built on top of our fast Sentinel-1 data access (Section 2) and geocoding (Section 3) mechanisms. While we derive our global scale products from geocoded bursts, we do not archive these oversampled SLCs as our tools allow us to regenerate this on-demand as needed.

4.1. Global burst footprint database

The global footprint database is a dynamic product which we update when new previously un-imaged burst footprints are encountered during ingest (See Figure 1 for example). This global database also allows us to track common processing parameters like EPSG codes, bounding boxes, nominal values for minimum and maximum altitude in the footprint, etc, to simplify orchestration of processing jobs at scale.

4.2. Global burst SLC database

The global burst SLC database is a dynamic product which we update with every new incoming SLC image released by ESA. In addition to nominal metadata like source SAFE granule, swath number and burst number; we also establish zip-decoder state and lookup tables, which enable us to access individual bursts independently as described in Section 2. Each polarization is stored as an independent entry in this global database. Our geocoding pipelines use this database to process geocoded bursts and generate backscatter and interferometric products at global scale (see Figure 2).

4.3. Global calibrated $\sigma_{0,E}$ product

We process all V-transmit IW mode data acquired by Sentinel-1 to a 10m calibrated, thermal-noise corrected $\sigma_{0,E}$ product in decibel scale with 12-bit quantization and a dynamic range from -40 dB to 30 dB. Backscatter products are generated from the amplitude of the geocoded SLCs using a Gaussian kernel of resolution 10m [30] in Easting and decimated to an aligned grid of 10m. The imagery is tagged by burst ids and contain two bands - VV and VH. This product is accessible just like other satellite imagery datasets on our platform [21]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2022 doi:10.20944/preprints202206.0252.v1

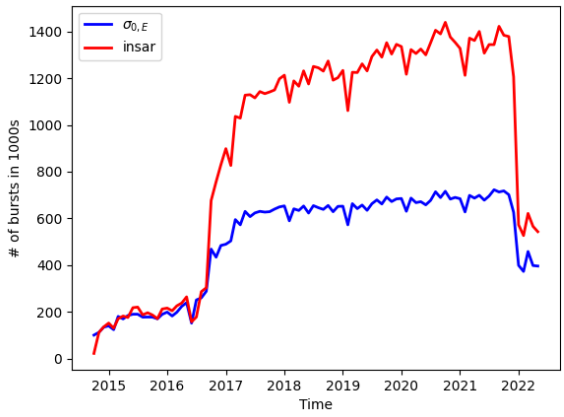

Figure 2. Monthly burst product counts in thousands in our global scale radar backscatter (blue) and interferogram (red) products.

Figure 3. Example of coherence mosaics from our global InSAR product. (Left) 6-day coherence mosaic from over Iceland from July 6-10, 2020. (Right-top) 12-day coherence mosaic from May 3-15, 2021 over Bangladesh. (Right-bottom) 24-day coherence mosaic from May 3-26, 2021 over Nile Delta in Egypt.

and is used in a number of large scale change detection pipelines, e.g., [31–35]. We can also process H-transmit data (mostly over Greenland and Antarctica), as well as EW mode data (mostly over sea ice) on request, or for experiments using the same mechanisms. Since we start from SLCs instead of GRD products, this global backscatter product does not have artifacts like noisy edges or missing lines as valid data regions are better labeled in SLC products and burst overlaps are preserved during processing.

We realize that Radiometric Terrain Correction (RTC) [36] is a topic of interest for backscatter products and we have developed a solution to transform our $\sigma_{0,E}$ to γ_0 using static correction factors. We plan to report our progress on this area in a future manuscript.

4.4. Global wrapped interferogram product

We process all VV-pol IW mode data acquired by Sentinel-1 over footprints over land to a 20m wrapped interferogram product. All interferometric pairs with temporal baseline of less than or equal to 24 days are automatically generated by our system. Interferograms generated by complex cross multiplication of geocoded bursts, filtered using a Gaussian kernel of resolution 50m [30] and decimated to an aligned grid of 20m. We do not apply any ESD corrections [28] as this has very little impact on interferometric coherence and the product should be considered equivalent to those generated with purely geometric coregistration strategies [37]. If desired, users can estimate the ESD offset using burst overlaps and apply a phase only correction [38] as part of their post-processing workflows.

The imagery are tagged with the reference burst id and temporal baseline and contain two bands - coherence (see Figure 3) at 10-bit quantization and wrapped phase (see Figure 4) at 16-bit quantization. As with backscatter, this product is accessible just like other satellite imagery datasets on our platform and is used in a number of large scale change detection pipelines, e.g., [33–35]. This global product has already been used to build quick-look deformation time-series tools and will easily support a global scale volcano monitoring system. Note that users can always combine this interferometric product with the $\sigma_{0,E}$ product above to recreate the original interferogram and apply more spatial looks if needed for their application.

 doi:10.20944/preprints202206.0252.v1

Figure 4. Example of wrapped phase data from our global InSAR product. (Left) Track 33 12-day interferogram of M 6.6 northern Qinghai (USGS id: us7000g9zq) earthquake from Jan 7, 2022. (Right) Track 102 6-day interferogram of M 6.0 Thrapsanon (USGS id: us7000fes8) earthquake from Sep 27, 2021.

5. Discussion

The SLC data access mechanism described in this manuscript can be easily adopted for use with ScanSAR data from sensors like Envisat and ALOS-2, and also with stripmap data to enable quick access to sections of long strips. Access to any SAR data format in which imagery arrays are stored as contiguous chunks (e.g, CEOS, COSAR) can be sped up with this mechanism.

The consistency of its imaging modes, i.e., central frequency and bandwidth, is one of the key strengths of the Sentinel-1 mission observation plan, and allows us to use geocoded bursts efficiently. For missions that acquire imagery in a large number of modes over the same AOI, using geocoded SLCs for interferometric applications might required additional mode-specific preprocessing and design of nested regular grids to handle heterogeneity in imaging resolution. As mentioned in Section 3, we start with lowest level calibrated SAR data - Level-1 SLCs in our pipelines as opposed to raw data [29]. This is because every SAR instrument and associated calibration information (which is often not publicly shared) is unique and needs to be constantly monitored. Starting from SLC data, assuming that the data providers have calibrated the data to the best of their ability - both geometrically and radiometrically, allows us to focus on goal of building and maintaining large scale sensor-agnostic analytic pipelines. Our geocoding method treats the TOPS phase ramp [22] as an instance of an arbitrary phase addition. This allows to bring in targeted spotlight imagery from other SAR sensors like TerraSAR-X, Cosmo-SkyMed, ICEYE and Capella into our data system without significant changes.

In Section 3, we noted that atmospheric propagation delay can result in variation of few 10s of cm in slant range in a stack of SAR imagery. For very high resolution data (e.g spotlight mode) the magnitude of variation is comparable to image resolution, often leading to decorrelation in interferometric applications. In such cases, we can perform the same sort of cross-correlation based offset estimation to generate coregistered stacks as in the traditional approach. In fact, these offsets can be estimated with the geocoded imagery and with simple geometric transformations translated back to offsets in radar coordinates before geocoding the data again with the necessary corrections. We have validated this approach with data from TerraSAR-X, COSMO-SkyMed and ICEYE. In general, our approach works very well for X-band and C-band data which are a lot less impacted by the ionosphere than L-band data [39].

In geospatial analytics, different projection systems are better suited for different types of analysis and applications. For example, Equal Area or UTM CRSs are suited for data aggregation or field surveys whereas geographic (LatLong) CRSs are better suited for others. The same holds true for SAR data as well. Native radar geometries (like the zero doppler system) are best suited for spectral sub-banding and calibration operations; whereas transforming the data to well known CRS early in the processing chain allows us to leverage scalable data systems and tools developed for managing large volumes of satellite imagery and efficiently operationalize established SAR applications like backscatter and InSAR. We believe that using geocoded SLCs as a building block significantly lowers the entry barrier for remote sensing users to start using global scale SAR datasets and for use of SAR data in conjunction with other geospatial datasets in more widely used geospatial frameworks.

The global products that we have described in this manuscript should be considered as building blocks for wide area analytics, similar to [31–35], along with data from other satellite sensors like Landsat, Sentinel-2 etc. In the near future, we hope to build on these

 doi:10.20944/preprints202206.0252.v1

products and build fast pipelines for global deformation and change detection products on a global scale.

6. Conclusions

In this manuscript, we have described the design principles and implementation details of a global scale processing system for generating SAR backscatter and interferometric products that we have developed. The presented approach is efficient, cost-effective and highly scalable; and is suited for handling, in near realtime, large volumes of SAR data that are expected to be acquired by missions like Sentinel-1, NISAR and other commercial providers in the near future.

Author Contributions: Conceptualization, M.T.C., P.S.A. and M.S.W.; methodology,P.S.A, M.T.C., and M.S.W.; investigation, P.S.A.; data curation, P.S.A. and M.S.W.; software, P.S.A., M.S.W. and M.T.C.; validation, P.S.A., M.S.W. and M.T.C.; writing—original draft preparation, P.S.A., M.T.C., and M.S.W.; writing—review and editing, P.S.A., M.T.C., M.S.W. and S. S. A.; visualization, P.S.A. and M.T.C.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The single-look complex (SLC) Sentinel-1 imagery and associated metadata are available at the Alaska Satellite Facility's Vertex Portal here: https://search.asf.alaska.edu/#/.

Acknowledgments: The authors would like to thank Kelly Olsen, Alice Durieux, Ahmad Hotaji Malekshah and other members of Descartes Labs Applied Science group for providing useful feedback during pipeline development. We would also like to thank following members of Sentinel-1 mission team for patiently clarifying our numerous questions - Nuno Miranda, Guillaume Hajduch, Muriel Pinheiro and Antonio Valentino.

Conflicts of Interest: The authors assert that they have no conflicts of interest.

References

1. Torres, R.; Snoeij, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; Potin, P.; Rommen, B.; Floury, N.; Brown, M.; et al. GMES Sentinel-1 mission. *Remote Sensing of Environment* **2012**, *120*, 9 – 24. The Sentinel Missions - New Opportunities for Science, https://doi.org/https://doi.org/10.1016/j.rse.2011.05.028.
2. Westerhoff, R.; Kleuskens, M.; Winsemius, H.; Huizinga, H.; Brakenridge, G.; Bishop, C. Automated global water mapping based on wide-swath orbital synthetic-aperture radar. *Hydrology and Earth System Sciences* **2013**, *17*, 651–663.
3. Bauer-Marschallinger, B.; Freeman, V.; Cao, S.; Paulik, C.; Schaufler, S.; Stachl, T.; Modanesi, S.; Massari, C.; Ciabatta, L.; Brocca, L.; et al. Toward global soil moisture monitoring with Sentinel-1: Harnessing assets and overcoming obstacles. *IEEE Transactions on Geoscience and Remote Sensing* **2018**, *57*, 520–539.
4. Dehls, J.F.; Larsen, Y.; Marinkovic, P.; Lauknes, T.R.; Stødle, D.; Moldestad, D.A. INSAR. No: A National Insar Deformation Mapping/Monitoring Service In Norway–From Concept To Operations. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019, pp. 5461–5464.
5. Ferretti, A.; Passera, E.; Capes, R. End-to-end implementation and operation of the European Ground Motion Service (EGMS): Algorithm Theoretical Basis Document. Technical Report EGMS-D3-ALG-SC1-2.0-006, European Environment Agency, 2021.
6. Sandwell, D.; Mellors, R.; Tong, X.; Wei, M.; Wessel, P. Open radar interferometry software for mapping surface Deformation. *Eos, Transactions American Geophysical Union* **2011**, *92*, 234–234, [https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2011EO280002]. https://doi.org/https://doi.org/10.1029/2011EO280002.
7. Hajduch, G.; Bourbigot, M.; Johnsen, H.; Piantanida, R. Sentinel-1 Level 1 Detailed Algorithm Definition. Technical Report S1-RS-MDA-52-7441, European Space Agency, 2022.
8. Sentinel-1 Mission Performance Cluster. Sentinel-1 Burst Id map, version 20220530. https://sar-mpc.eu/test-data-sets/, 2022. Generated by Sentinel-1 SAR MPC.
9. Piantanida, R.; Hajduch, G.; Poullaouec, J. Sentinel-1 Product Specification. Technical Report SEN-TN-52-7445, European Space Agency, 2022.
10. Rajna, Z.; Keskinarkaus, A.; Kiviniemi, V.; Seppänen, T. Speeding up the file access of large compressed nifti neuroimaging data. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2015, pp. 654–657.
11. Larsen, Y.; Marinkovic, P.; Dehls, J.; Bredal, M.; Bishop, C.; Jokulsson, G.; Lars-Petter, G.; Frauenfelder, R.; Salazar, S.; Voge, M.; et al. European Ground Motion Service (EGMS): specification and implementation plan (Version 1.01). Technical report, European Environment Agency, 2020.

12. Nicolai, R.; Simensen, G. The New EPSG Geodetic Parameter Registry. In Proceedings of the 70th EAGE Conference and Exhibition incorporating SPE EUROPEC 2008. European Association of Geoscientists & Engineers, 2008, pp. cp–40.
13. PROJ contributors. *PROJ coordinate transformation software library*. Open Source Geospatial Foundation, 2022. https://doi.org/10.5281/zenodo.5884394.
14. Small, D.; Schubert, A. Guide to ASAR geocoding. *ESA-ESRIN Technical Note RSL-ASAR-GC-AD* **2008**, *1*, 36.
15. Kropatsch, W.G.; Strobl, D. The generation of SAR layover and shadow maps from digital elevation models. *IEEE Transactions on Geoscience and Remote Sensing* **1990**, *28*, 98–107.
16. Warmerdam, F. The geospatial data abstraction library. In *Open source approaches in spatial data handling*; Springer, 2008; pp. 87–104.
17. Sansosti, E.; Berardino, P.; Manunta, M.; Serafino, F.; Fornaro, G. Geometrical SAR image registration. *IEEE Transactions on Geoscience and Remote Sensing* **2006**, *44*, 2861–2870.
18. Ferretti, A.; Prati, C.; Rocca, F. Permanent scatterers in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* **2001**, *39*, 8–20.
19. Berardino, P.; Fornaro, G.; Lanari, R.; Sansosti, E. A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Transactions on Geoscience and Remote Sensing* **2002**, *40*, 2375–2383.
20. Eineder, M. Efficient simulation of SAR interferograms of large areas and of rugged terrain. *IEEE Transactions on Geoscience and Remote Sensing* **2003**, *41*, 1415–1427.
21. Beneke, C.M.; Skillman, S.; Warren, M.S.; Kelton, T.; Brumby, S.P.; Chartrand, R.; Mathis, M. A Platform for Scalable Satellite and Geospatial Data Analysis. In Proceedings of the AGU Fall Meeting Abstracts, 2017, Vol. 2017, pp. IN32C–04.
22. Miranda, N. Definition of the TOPS SLC deramping function for products generated by the S-1 IPF. Technical Report COPE-GSEG-EOPG-TN-14-0025, European Space Agency, 2017.
23. Van Zyl, J.J. The Shuttle Radar Topography Mission (SRTM): a breakthrough in remote sensing of topography. *Acta Astronautica* **2001**, *48*, 559–565.
24. Fernández, M.; Peter, H.; Arnold, D.; Duan, B.; Simons, W.; Wermuth, M.; Hackel, S.; Fernández, J.; Jäggi, A.; Hugentobler, U.; et al. Copernicus Sentinel–1 POD Reprocessing Campaign. *Advances in Space Research* **2022**.
25. Eineder, M.; Minet, C.; Steigenberger, P.; Cong, X.; Fritz, T. Imaging geodesy—Toward centimeter-level ranging accuracy with TerraSAR-X. *IEEE Transactions on Geoscience and Remote Sensing* **2010**, *49*, 661–671.
26. Gisinger, C.; Schubert, A.; Breit, H.; Garthwaite, M.; Balss, U.; Willberg, M.; Small, D.; Eineder, M.; Miranda, N. In-depth verification of Sentinel-1 and TerraSAR-X geolocation accuracy using the Australian corner reflector array. *IEEE Transactions on Geoscience and Remote Sensing* **2020**, *59*, 1154–1181.
27. Gisinger, C.; Nagler, T.; Libert, L.; Marinkovic, P.; Larsen, Y.; Miranda, N.; Valentino, A.; Breit, H.; Suchandt, S.; Balss, U.; et al. Evaluation of ESA's Extended Timing Annotation Dataset (ETAD) for Sentinel-1-First Results for Ice Velocity Monitoring and InSAR Applications in Greenland, Iceland and Norway. In Proceedings of the FRINGE 2021. European Space Agency, 2021.
28. Prats-Iraola, P.; Scheiber, R.; Marotti, L.; Wollstadt, S.; Reigber, A. TOPS interferometry with TerraSAR-X. *IEEE Transactions on geoscience and remote sensing* **2012**, *50*, 3179–3188.
29. Zebker, H.A. User-Friendly InSAR Data Products: Fast and Simple Timeseries Processing. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 2122–2126.
30. Sandwell, D.T.; Price, E.J. Phase gradient approach to stacking interferograms. *Journal of Geophysical Research: Solid Earth* **1998**, *103*, 30183–30204.
31. Kontgis, C.; Warren, M.S.; Skillman, S.W.; Chartrand, R.; Moody, D.I. Leveraging Sentinel-1 time-series data for mapping agricultural land cover and land use in the tropics. In Proceedings of the 2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp). IEEE, 2017, pp. 1–4.
32. Kontgis, C.; Survila, K. Analysis of lowland rice across Asia. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2018, pp. 2058–2061.
33. Durieux, A.M.; Calef, M.T.; Arko, S.; Chartrand, R.; Kontgis, C.; Keisler, R.; Warren, M.S. Monitoring forest disturbance using change detection on synthetic aperture radar imagery. In Proceedings of the Applications of Machine Learning. International Society for Optics and Photonics, 2019, Vol. 11139, p. 1113916.
34. Durieux, A.M.; Ren, C.X.; Calef, M.T.; Chartrand, R.; Warren, M.S. Budd: Multi-modal bayesian updating deforestation detections. In Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2020, pp. 6638–6641.
35. Durieux, A.M.; Rustowicz, R.; Sharma, N.; Schatz, J.; Calef, M.T.; Ren, C.X. Expanding SAR-based probabilistic deforestation detections using deep learning. In Proceedings of the Applications of Machine Learning 2021. International Society for Optics and Photonics, 2021, Vol. 11843, p. 1184307.
36. Truckenbrodt, J.; Freemantle, T.; Williams, C.; Jones, T.; Small, D.; Dubois, C.; Thiel, C.; Rossi, C.; Syriou, A.; Giuliani, G. Towards Sentinel-1 SAR analysis-ready data: A best practices assessment on preparing backscatter data for the cube. *Data* **2019**, *4*, 93.
37. Xu, X.; Sandwell, D.T.; Tymofyeyeva, E.; González-Ortega, A.; Tong, X. Tectonic and anthropogenic deformation at the Cerro Prieto geothermal step-over revealed by Sentinel-1A InSAR. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *55*, 5284–5292.
38. De Luca, C.; Bonano, M.; Casu, F.; Manunta, M.; Manzo, M.; Onorato, G.; Zinno, I.; Lanari, R. The parallel SBAS-DInSAR processing chain for the generation of national scale sentinel-1 deformation time-series. *Procedia computer science* **2018**, *138*, 326–331.

 doi:10.20944/preprints202206.0252.v1

39. Yunjun, Z.; Fattahi, H.; Pi, X.; Rosen, P.; Simons, M.; Agram, P.; Aoki, Y. Range Geolocation Accuracy of C-/L-Band SAR and its Implications for Operational Stack Coregistration. *IEEE Transactions on Geoscience and Remote Sensing* **2022**, *60*, 1–19.