Cloud Computing for Earth Surface Deformation Analysis via Spaceborne Radar Imaging: A Case Study

I. Zinno, *Member, IEEE*, L. Mossucca, S. Elefante, C. De Luca, V. Casola, O. Terzo, F. Casu, and R. Lanari, *Fellow, IEEE*

Abstract—We present a case study on the migration to a cloud computing environment of the advanced differential synthetic aperture radar interferometry (DInSAR) technique, referred to as Small BAseline Subset (SBAS), which is widely used for the investigation of Earth surface deformation phenomena. In particular, we focus on the SBAS parallel algorithmic solution, namely P-SBAS, that allows the production of mean deformation velocity maps and the corresponding displacement time-series from a temporal sequence of radar images by exploiting distributed computing architectures. The Cloud migration is carried out by encapsulating the overall P-SBAS application in virtual machines running on the cloud; moreover, the cloud resources provisioning and configuration phases are implemented in an automatic way. Such an approach allows us to preserve the P-SBAS parallelization strategy and to straightforwardly evaluate its performance within a cloud environment by comparing it with those achieved on a HPC in-house cluster. The results we present were achieved by using the Amazon Elastic Compute Cloud (EC2) of the Amazon Web Services (AWS) to process SAR datasets collected by the ENVISAT satellite and show that, thanks to the cloud resources availability and flexibility, large DInSAR data volumes can be processed through the P-SBAS algorithm in short time frames and at reduced costs. As a case study, the mean deformation velocity map of the southern California area has been generated by processing 172 ENVISAT images. By exploiting 32 EC2 instances this processing took less than 17 hours to complete, with a cost of USD 850. Considering the available PB-scale archives of SAR data and the upcoming huge SAR data flow relevant to the recently launched (April 2014) Sentinel-1A and the forthcoming Sentinel-1B satellites, the exploitation of cloud computing solutions is particularly relevant because of the possibility to provide cloud-based multi-user services allowing worldwide scientists to quickly process SAR data and to manage and access the achieved DInSAR results.

Index Terms—Cloud Computing, DInSAR, P-SBAS, Earth surface deformation, big data

1 Introduction

The investigation of Earth's surface deformation phenomena provides critical insights into several processes of great interest for science and society, especially from the perspective of further understanding the Earth System and the impact of the human activities. Indeed, deformation phenomena can be helpful for the comprehension of the geophysical dynamics dominating natural hazards such as earthquakes, volcanoes and landslide, and therefore they strongly support all those activities related to risk mitigation and prevention [1], [2], [3], [4], [5]. Furthermore, such information can be very useful also to detect, monitor and limit the effects of anthropogenic actions, such as ground-water exploitation [6], oil and gas extraction [7], gas capture and storage [8], [9], mining activities [10], tunnelling, building and management of dams and flood defences, and transportation [11], [12].

Earth observation techniques are very powerful for ground deformation estimation due to their characteristics of large spatial coverage and cost effectiveness. In this scenario, differential synthetic aperture radar interferometry (DInSAR) is regarded as one of the key methods for its ability to investigate surface deformation affecting large areas of the Earth with centimeter- to millimeter-level accuracy [13], [14], [15].

Among several DInSAR algorithms, a well-known approach is the technique referred to as Small BAseline Subset (SBAS) [16] that allows the computation of mean deformation velocity maps and the corresponding displacement time-series in different scenarios (volcanoes, tectonics, landslides, anthropogenic induced land motions) from a sequence of SAR acquisitions. An advanced parallel algorithmic solution for the SBAS approach, referred to as P-SBAS, which implements the complete DInSAR processing chain and is able to exploit distributed computing architectures, has been developed recently [17].

The P-SBAS capability of running on distributed systems, which encompass large, scalable computing resources, is particularly timely considering the current SAR remote sensing scenario that is characterized by the availability of massive amounts of data. Indeed, huge archives of SAR acquisitions that have been collected during the past 20 years are disposable. They comprise the long-term C-band ESA archives (e.g. ERS-1/2 and ENVISAT, a total of

- *I. Zinno, S. Elefante, C. De Luca, F. Casu, and R. Lanari are with the IREA, CNR, Napoli, Italy. E-mail: {zinno.i, elefante.s, casu.f, lanari.r}@irea.cnr.it.*
- *L. Mossucca and O. Terzo are with the ISMB, Istituto Superiore Mario Boella, Torino, Italy. E-mail: {mossucca, terzo}@ismb.it.*
- *C. De Luca and V. Casola are with the Department of Electrical Engineering and Information Technology, University of Naples Federico II, Napoli, Italy. E-mail: deluca.c@irea.cnr.it, valentina.casola@unina.it.*
- *S. Elefante is also with the Department of Geodesy and Geoinformation, Vienna University of Technology, Vienna, Austria.*

Manuscript received 13 Aug. 2014; revised 19 May 2015; accepted 25 May 2015. Date of publication 0 . 0000; date of current version 0 . 0000.
Recommended for acceptance by K. Keahey, I. Raicu, K. Chard, B. Nicolae.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCC.2015.2440267

approximately 60 TBytes on-line available through the Virtual Archive 4 at the time of writing [18]), as well as those provided by the X-band new generation SAR sensors, such as COSMO-SkyMed (CSK) that acquires about 600 GBytes of data every day [19], and TerraSAR-X (TSX) (about 90 GBytes of data per day [20]). Moreover, on April 2014, the SENTINEL-1A [21] SAR satellite operating within the framework of the COPERNICUS (formerly GMES) program of the European Union [22] has been launched and is expected to provide about 1.5 TBytes of raw data per day [23]. This sensor will be paired during 2016 with the SENTINEL-1B twin system that will cut in half the satellites revisit time from 12 to 6 days. Accordingly, a massive and ever increasing data flow that is acquired by a SAR mode that is specifically devoted to DInSAR applications [24] will be available soon and it will be released following a total free and open access policy.

These SAR data archives and the incoming data flow provided by the SENTINEL-1 constellation present a bottleneck problem for typical, limited 'in- house' or grid resources. In this scenario, a key role can be played by the Cloud Computing technologies that harness large collections of resources, both software and hardware, thus representing a promising solution to easily and rapidly manage and process large amounts of data, with important implications for scientific applications [25].

However, the deployment of scientific high performance computing (HPC) applications on cloud platforms is a challenging problem, which has been particularly highlighted recently. Indeed, the performance evaluation of scientific applications within cloud environments is currently an open issue and it is still under investigation to determine if, when and, most of all, how it is convenient to use the Cloud for HPC [26], [27], [28], [29]. This is a particularly actual issue when dealing with DInSAR data processing since several aspects have to be carefully taken into account to migrate to the Cloud. First of all, such scientific applications generally present very complex and multi-step workflows [16], [30], [31]. Depending on the specific algorithm, these workflows can have sequential or parallel implementations [17], [32]. This implies the adoption of diverse parallel programming techniques depending on the employed processing infrastructure. Therefore, the Cloud deployment requires a thorough analysis aimed at properly designing the computing architecture and the jobs scheduling among different Cloud machines. Moreover, typically, DInSAR algorithms are characterized by workflows presenting non-eliminable data dependencies between different steps [16], [30], [31]. Consequently, it is not possible to decompose the whole problem into totally independent tasks which can be initially parallel distributed and finally merged together. Accordingly, the adoption of a typical cloud dedicated paradigm, like the MapReduce framework, is not straightforward and not necessarily effective [33], [34], [35]. Finally, due to both the size of involved data, which can range from hundreds of GBytes up to tens of TBytes for the processing of a single interferometric dataset, destined to enlarge to PB-scale with Sentinel-1, and the nature of the operations which are performed on them, the majority of SAR algorithms are characterized by highly demanding requirements in terms of computing resources and hardware. Hence, the achievement of high computing performances on the Cloud for SAR applications is not a trivial problem [36].

In this work we present a case study based on the migration to the Amazon Web Services (AWS) public cloud environment of the P-SBAS DInSAR processing chain. The motivation to accomplish such a task is to specifically investigate the suitability of cloud computing technologies to effectively process large amounts (up to PBytes) of spaceborne SAR data on short timescales and at limited costs through the P-SBAS algorithm and, more generally, to provide a contribution to the cloud computing exploitation within the earth observation (EO) framework. Indeed, such an issue is currently at an early stage although some works have already been presented [37], [38].

To deploy the P-SBAS algorithm on the Cloud we adopted an approach that allowed us to preserve the intrinsic P-SBAS parallelization strategy and, therefore, to straightforwardly evaluate the Cloud exploitation benefits, in terms of achieved performances and steady costs, with a minimum migration effort. Indeed, both a processing time and a cost performance analysis have been carried out through a dedicated experimental study that has been conducted by exploiting an ENVISAT dataset of 64 SAR images acquired on the Napoli Bay area. In particular, we have first investigated the P-SBAS processing performances achieved by using different Cloud resource configurations within the AWS environment, when the number of exploited EC2 instances increases. Second, an estimation of the costs relevant to the experimental tests performed on the AWS cloud is carried out in order to evaluate the trade-off between processing times and costs of an interferometric computation depending on the employed resources.

Finally, the proposed P-SBAS cloud implementation has been tested by processing a large interferometric data stack composed of 172 SAR ENVISAT scenes in southern California. The considered dataset covers the Los Angeles metropolitan area, which includes a large variety of deformation phenomena, such as active seismic faults, aquifer seasonal exploitation and recharge, oil extraction and agricultural fields. In this case the input data size was about 80 GBytes while the overall data involved in the P-SBAS processing (input, intermediate and final products) were approximately 2 TBytes. The goal of such an experiment is to give insights on the increasingly relevant question of whether Cloud environments are suitable to massively process the existing and incoming large amount of SAR data in order to compute interferometric products within short time frames and at reduced costs.

2 P-SBAS Processing Chain Description

We provide in this section a concise description of the P-SBAS processing chain aimed at presenting its principal steps. In particular, each step is briefly described in terms of main tasks, implemented procedures and computational issues relevant to CPU usage, RAM capacity and I/O transfer requirements.

The P-SBAS solution has been designed by carefully taking into account several aspects, such as data dependencies, task partitioning, inherent granularity, scheduling policy, load balancing and I/O. It is worth emphasizing that, due to the heterogeneous nature of the algorithms comprised within the SBAS processing chain, appropriate parallelization strategies have been adopted depending on the algorithmic structure of the particular processing step under evaluation [17], [38].

In order to take advantage of both multi-node and multi-core architectures, two-parallelization levels have been considered, i.e., at the process and the thread level. The former considers a coarse/medium granularity-based approach (mainly applied to the whole processing chain) while the latter relies on a fine-grained parallelization that has been implemented for the most computing-intensive processing operations [17].

The block diagram of the P-SBAS processing chain is shown in Fig. 1; note that in this scheme the steps depicted by red blocks represent the jobs that are parallel executed by simultaneously running on different computing units. Moreover, dashed line blocks depict the steps that exploit multi-core parallelism making use of multithreading programming techniques. Finally, the black blocks represent steps that are intrinsically sequential as they merge together information and data coming out from the previous steps' executions. To give a quantitative idea of the main computational characteristics of each step of the block diagram in Fig. 1 in terms of CPU and RAM usage, as well as I/O operations, we summarize in Table 1 the respective P-SBAS requirements relevant to a SAR dataset constituted by 64 ENVISAT images. This can be considered as a 'standard' dataset for the ENVISAT archives for what concerns its size, and therefore it has been used as a benchmark also in the course of the experimental analysis (see Section 4.1). It is worth noting that the P-SBAS steps' computational characteristics that are provided in Table 1 may significantly change when considering datasets acquired by different sensors, with a drastic increase in the size of data in the case of COSMO-SkyMed or SENTINEL-1 satellites.

Let us start our discussion by considering the SAR data focusing operation implemented by step A that consists of transforming the radar raw data into microwave images, often referred to as single look complex (SLC) images [15]. This step has a high computational burden because it mostly performs two-dimensional FFTs of complex data matrices on the order of GBytes [15]; moreover, it also has a remarkably significant data flow as it involves frequent I/O operations (see Table 1).

Step B performs the digital elevation model (DEM) conversion into SAR coordinates [39]. In this step data matrices typically on the order of tens of GBytes are processed, thus requiring an adequate amount of memory.

Within step C, the SAR image coregistration operation is carried out to reference all data to a unique radar geometry via an interpolation of all the focused images with respect to a reference one, referred to as the "master" image [40]. Also for this step the main limitation is the files size; indeed it requires intensive (hundreds of GBytes) I/O workload and significant (nearly tens of GBytes) RAM memory (see Table 1).

Step D performs the identification of the interferometric data pairs required for the subsequent co-registration refinement step E, in which possible residual sub-pixel shifts of the images are first evaluated and, afterwards, used to resample the images. This step is particularly I/O demanding (hundreds of GBytes to read/write) because it deals with full resolution SAR images.

Before moving to the interferogram generation, the evaluation of some parameters useful for the subsequent steps is carried out (block F of Fig. 1). Having both the co-registered images (steps C and E) and the DEM (as output of step B) referenced to a common radar geometry, the differential interferograms are generated through step G. These operations require a high memory RAM usage (tens of Gbytes) as well as intensive I/O workload (hundreds of GBytes) as they are carried out at the SAR images' full spatial resolution and in the complex domain. However, the final results are stored in low resolution mode following a complex spatial average (multi-look) operation [15]; this procedure mitigates the noise (decorrelation) affecting the DInSAR interferograms and drastically reduces the sizes of the final outputs, but not those of the intermediate products.

The modulo-2π restricted phase of each multi-look interferogram needs to be "unwrapped" to retrieve the original phase [15]. This procedure is carried out in step H and I by applying the extended minimum cost flow (EMCF) phase unwrapping (PhU) algorithm [41].

The phase unwrapping steps are the most demanding in terms of memory (tens of GBytes of RAM required) and CPU usage (it exploits multithreading programming) (see Table 1). Indeed, they deal with wrapped and unwrapped interferogram stacks represented by three-dimensional (3-D) matrices.

A pixel-based inversion of the unwrapped phases of the interferogram stack is, afterwards, carried out (step J) to retrieve the deformation time-series. Moreover, in step K, the estimation of possible residual phase artifacts affecting the interferograms, often referred to as "orbital phase patterns", is undertaken. Such phase ramps that are due to possible orbital inaccuracy, are subsequently removed from the wrapped interferograms and a final PhU step has to be performed on the "orbital error free" interferograms (second run of step H, I, and J of Fig. 1). Block L provides the final deformation time-series after executing the temporal coherence estimation [41], which provides information about the quality of the retrieved displacement measurements.

In the rest of this section we summarize the most critical issues concerning the porting of the parallel SBAS algorithm within distributed computing environments.

First of all, the P-SBAS approach requires the execution of different steps to be performed in a specific and fixed order. For instance, SAR image registration can occur only after the raw data focusing operation has been accomplished, or the unwrapping step can be executed only after the generation of the differential interferograms sequence. This rationale shows that the successively generated data and products of the SBAS processing chain are strongly interconnected. Moreover, in some steps the joint processing of the outputs generated by previous blocks need to be performed. Hence, a strategy that ensures the sharing of the middle chain products among the different steps is needed.

Second, many steps of the processing chain, such as the DEM conversion (step B), the interferogram generation (step G) and the phase unwrapping steps (H and I), are characterized by large RAM usage, as detailed in Table 1.

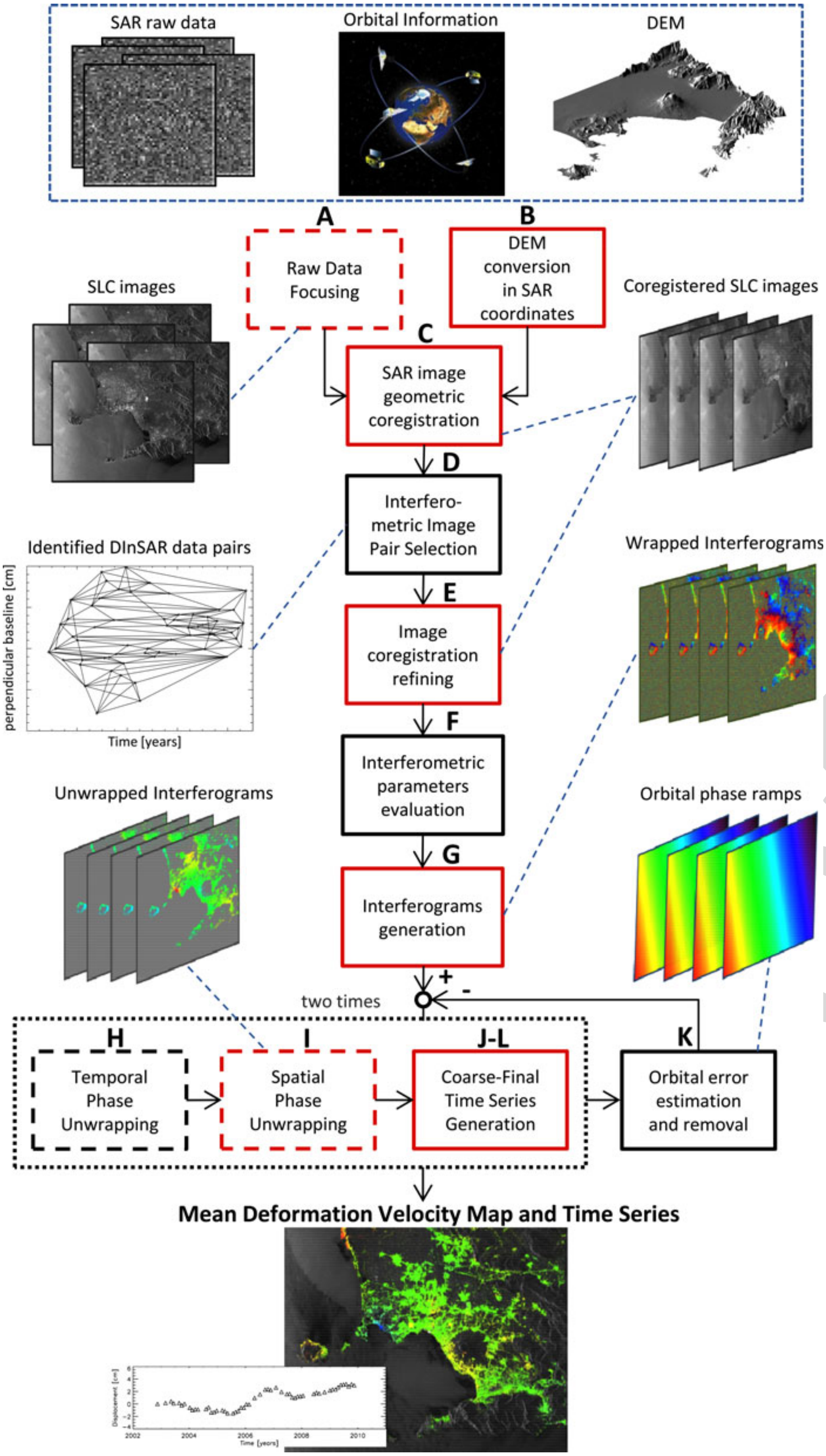

Fig. 1. P-SBAS workflow. Black and red blocks represent sequential and parallel (from a process-level perspective) processing steps, respectively. Dashed line blocks represent multi-threading programmed processing steps.

This issue places specific requirements on the computing instances that have to be provisioned to execute these steps.

Finally, very large amounts of data are involved in the SBAS processing - hundreds of GBytes for the considered standard dataset which can become TBytes if COSMO-SkyMed or SENTINEL-1 datasets are concerned - and, therefore, data transfer and I/O issues are two critical aspects to be carefully taken into account. Hence, dedicated storage capability and network facility need to be exploited.

The P-SBAS processing chain has been, until now, only extensively tested on a dedicated cluster that is located at the CNR-IREA premises where it has originally been

TABLE 1
P-SBAS Workflow Requirements for a Standard SAR Data Set Relevant to the Napoli Bay Area (64 ENVISAT Images)

Step	CPU usage	Maximum Resident Set Size (MB)	I/O Workload (MB)
Raw Data Focusing	53%	3,700	928,000
DEM Conversion	37,5%	9,400	83,000
SAR Image registration	14%	7,800	592,000
Interferometric Pair selection	12,5%	280	2,100
Image coregistration refining	22,5%	5,600	226,000
Interferometric parameters evaluation	34%	6,000	2,500
Interferograms generation	25%	13,000	686,000
Temporal phase unwrapping	72,5%	9,000	53,000
Spatial phase unwrapping	72,5%	22,000	80,000
Deformation and residual topography estimation	12,5%	5,800	2,160
Orbital error estimation and removal	19%	7,200	6,430
Displacement time series	15%	6,000	12,600

These values refer to a standard data set (64 ENVISAT images, see Section 4.1). The CPU usage percentage values refer to the percentage of CPU that each step got on average per single node. Note that the represented values have been normalized with respect to the maximum achievable CPU usage, which in our case is given by the full exploitation of eight cores per node. The I/O workload is calculated as the total of data read and written per step and is given in MegaBytes.

developed. The P-SBAS technique has already been broadly used to process the ENVISAT SAR data archives[1] but its rationale is also well suited to process the upcoming SENTINEL-1 data stream.

3 Cloud Solution

3.1 Cloud Migration

As known in the literature, there are different possibilities for migrating an existing application in a cloud computing environment, that comprise either adaptation or re-design of the application components. More specifically, according to [44], the following four major approaches can be identified:

- i Replace components with Cloud offerings;
- ii Partially migrate some of the application functionality to the Cloud;
- iii Migrate the whole software stack of the application to the Cloud;
- iv Cloudify the application.

Obviously, each of the above mentioned solutions provides advantages and dis-advantages and the choice is driven by the evaluation of the tradeoff among different aspects, including the migration effort, the infrastructure costs, and the overall performance (not only the total computation time).

In our case the goal was to evaluate the performances tha tare achieved within a Cloud environment, by directly comparing them to those obtained on a HPC cluster. Moreover, we aimed at investigating the costs associated with the Cloud exploitation for the P-SBAS processing in order to evaluate its convenience. Therefore, at this early stage, we opted for a migration approach that allows us to straight forwardly attain such aims by minimizing the deployment effort at the same time.

Before managing the actual migration of the P-SBAS application to cloud, a deep analysis of the workflow that takes into account both software and hardware requirements was carried out.

The P-SBAS algorithm has a complex workflow that is composed of several steps both sequential and parallel (see Fig. 1). The majority of the P-SBAS chain algorithms are developed in the Exelis Interactive Data Language (IDL) [45]: a scientific programming language that combines numerical and graphical abilities and is widely used by engineers and scientists developing algorithms for SAR and DInSAR data processing. IDL is a commercial software and therefore each machine running the application requires a licence.

As extensively described in Section 2, the P-SBAS rationale provides that in many steps of the chain the joint processing of the outputs generated by previous steps needs to be performed; hence, a common storage and an efficient management of data sharing between the processing nodes are required. For this purpose, Network File System (NFS) was adopted, which is a distributed file system protocol that allows exploiting a network to access to a remote storage for saving and updating files, as though they were on the local storage [46].

According to the specific P-SBAS requirements, we limited the possible migration choices to the last two approaches (iii and iv) and evaluated their respective benefits and drawbacks.

The iii) approach is the most common example of migration to the Cloud, where the application is easily encapsulated in virtual machines (VMs) that run on the Cloud. Such an approach implies the minimum migration effort and allows us to easily evaluate the benefits, in terms of processing times and costs, to exploit the Cloud for scientific applications. Indeed, the Cloud adoption could be a valid alternative with respect to the need to update/buy new clusters, which are very expensive, require maintainance and rather quickly become obsolete.

Furthermore, thanks to the use of VM, several benefits are gained, which are in order:

1) Security and Isolation: it runs services in cloud environment totally independent from each other;

1. The SBAS capability to process ERS data sequences has also been extensively assessed [42], [43].

2) Ease of administration and management: due to the common virtualization layer and the adoption of snapshots for both installation and configuration;
3) Disaster recovery: VM can be launched in a few minutes and furthermore, can be cloned and migrated to different locations;
4) High reliability and load balancing optimization.

The iv) approach could be interesting as, in addition to the previous ones, some benefits would be expected from the exploitation of software using specific cloud features (as for example Hadoop or Cloud orchestration software). However, although the adoption of the MapReduce-based framework seems to be promising in managing large clusters of data, results from the literature [33], [34] show that this is particularly true when it is possible to pre-process data to optimize both the mapping and reducing phases. Indeed, the mapping phase may require a number of steps that could be automatized through a workflow of jobs (e.g Hadoop workflow) but the data constraints and the scheduling policies should be properly tuned. In particular, concerning the interferometric processing, the data mapping phase is easy because of two main constraints. First of all it is not always possible to split the input data of the P-SBAS processing steps in smaller portions to be distributed to all available nodes because of the spatial dependancies. Second, as already mentioned, some intermediate steps of the chain require access to all data products produced by the previous steps.

Based on the above considerations, we estimated that the cloudification effort should require a separate analysis to better gain benefit from the cloud and this will be addressedin future work.

For all these reasons, the approach of migrating the whole P-SBAS software stack to the Cloud was adopted to port the P-SBAS processing chain to the Cloud environment.

3.2 Cloud Deployment

As a cloud computing infrastructure for the P-SBAS algorithm deployment we chose AWS Elastic Compute Cloud (EC2) because it is currently a feature-rich, stable and commercial public cloud [47]. AWS is a web service that allows users to easily configure and instantiate VM images. Moreover, AWS EC2 is based on the XEN para-virtualization technology [48], which also provides the capability of sizing instances based on the EC2 Compute Unit (ECU).

We implemented an Amazon virtual private cloud (VPC) that is a logically isolated section of the AWS in which resources can be launched in a completely defined virtual network. The easy customization of the network configuration allows the users to fully control the virtual networking environment through IP address range selection, subnet creation, route tables configuration, network gateways and multiple layers of security (security groups and network access control lists). Indeed, thanks to this customization, server linking and SAR data uploading have been configured with predefined rules to ensure a secure connection between end points and to allow only authorized users to share data.

Moreover, we configured a customized Amazon Machine Image (AMI) containing the operating system (Ubuntu distribution), the packages enabling an easy integration with AWS services (i.e., Amazon EC2 API and AMI tools), and the software needed for the P-SBAS processing, such as IDL. Note that this software can run with a license that is available on a local server (IDL licenses server) and we need that all AMIs in the Cloud securely communicate with this server to gather the proper license.

In order to automate as much as possible the provisioning and configuration of cloud resources, (e.g. automatically handle EC2 instances, network and storage facilities), we developed Linux Bash scripts implementing the AWS Command Line Interface (CLI) which is a unified tool provided by AWS to manage multiple services [47]. The goal of these scripts is twofold; on the one hand, we need to set up the configuration to run the P-SBAS algorithm, on the other hand, we want to connect, start and stop additional computing nodes (workers) in order to perform the parallel tasks of the algorithm on different dedicated machines.

More specifically, through the aforementioned Linux Bash scripts, one of the instances is configured as the master node and performs multiple tasks. In particular it handles the whole P-SBAS workflow by managing the scheduling of parallel jobs that will be executed by the worker nodes, it performs computation, and it acts as an NFS server (see Fig. 2). The implemented scripts, also allow the selection of the number and type of EC2 instances to provision. A dynamical use of the employed EC2 instances is accomplished by allowing to start/stop them according to the parallel/sequential parts of the algorithm and to terminate all the instances at the end of the processing. This solution allows us to optimize the EC2 instances exploitation, and therefore the associated costs, because they are maintained active only when operative, thus avoiding that they remain idling while other nodes are working. Also, the worker nodes configuration is automatically performed by connecting each of them, as soon as the instance is started, both to the master node through an NFS client as well as to the IDL licence server.

Moreover, the implemented Bash script automatically performs the creation of the selected storage volume. Such an external disk is specifically configured to support the P-SBAS I/O workload and it is mounted to the master node.

4 Experimental Analysis

In this section we present an experimental analysis aimed at investigating the P-SBAS processing performances that are achieved by using two different cloud resources configurations within the AWS environment (in terms of instance type and storage) when the number of exploited instances increases. The goal of our study is twofold; on the one hand, we evaluated the cloud processing elapsed times by also comparing them with the corresponding processing times obtained on a HPC cluster. On the other hand, we carried out an estimation of the costs relevant to the experimental tests that have been performed on the AWS cloud in order to evaluate the trade-off between elapsed times and costs of an interferometric SAR data processing, depending on the exploited resources. Such an analysis could be very useful to select the best configuration in terms of typology and amount of cloud resources to use according to the specific time/cost requirements.

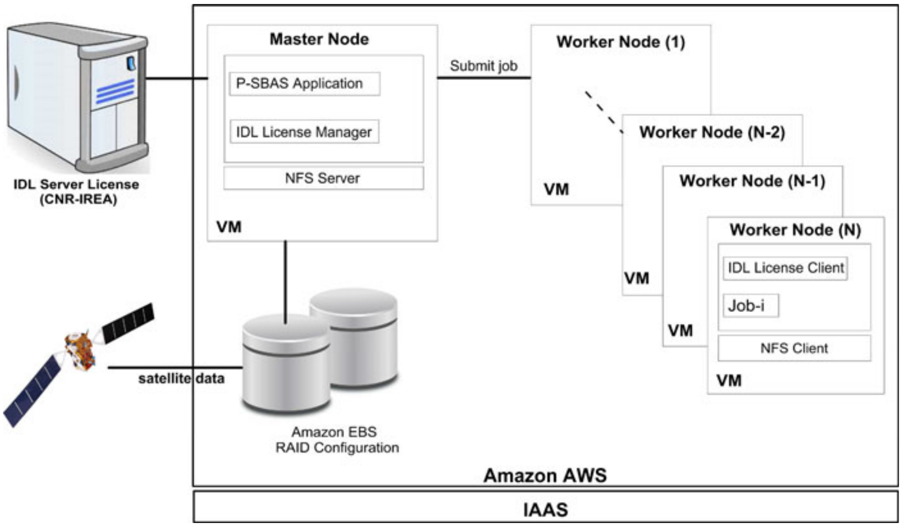

Fig. 2. Public cloud platform for DInSAR analysis; it consists of one master and N worker nodes as well as a common storage volume that contains the input data. All the components are located in the Amazon cloud.

4.1 Dataset and P-SBAS Results

The experimental analysis has been carried out by exploiting an interferometric SAR data stack acquired over the Napoli bay area, a densely urbanized zone located in Southern Italy and including the active caldera of Campi Flegrei, the Vesuvius volcano, and the city of Napoli. In particular, 64 ENVISAT SAR data acquired from ascending orbits, spanning the 2002-2010 time interval and approximately covering an area of 100×100 km^2 (corresponding to an ENVISAT frame), have been considered. The overall input data, comprising the SAR raw data sequence, the DEM files and the orbital information come to approximately 20 GBytes. In Fig. 3 we show the mean deformation velocity map in the radar line of sight (LOS), which is one of the DInSAR results generated through the P-SBAS processing. It is represented in the radar slant-range coordinates and afterwards superimposed on a SAR amplitude image of the investigated area. The estimated mean deformation velocity has been only computed in the areas in which the measurement accuracy is not affected by decorrelation noise (coherent pixels). In particular, it is worth noting that in Fig. 3a a significant deformation pattern corresponding to the area of the Campi Flegrei caldera is clearly shown. Moreover, the computation of the temporal evolution of the detected deformation has also been carried out for each coherent pixel of the scene. For instance, the chronological sequence of the computed displacement of a specific pixel (located in the maximum deforming area of the Campi Flegrei caldera) is plotted in Fig. 3b. These results are in accordance with ground truth measurements [42], [43]. The maximum storage capacity required by this dataset processing is of about 450 GBytes.

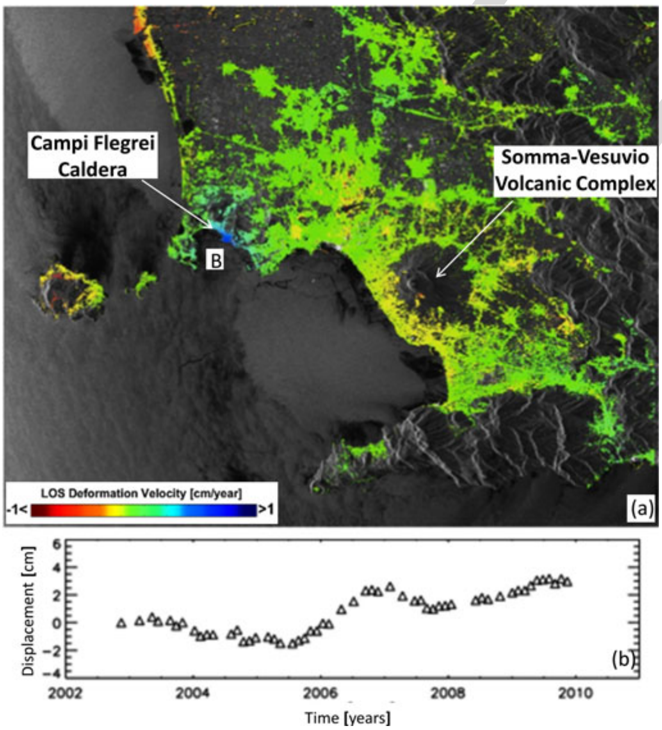

Fig. 3. (a) Mean deformation velocity map of the Napoli Bay area, in line of sight (LOS), which has been generated via the P-SBAS processing on Amazon Cloud. (b) Plot of the retrieved displacement time series relevant to a pixel located in the area of maximum deformation of the Campi Flegrei Caldera.

4.2 P-SBAS Performance Analysis within AWS Cloud

The Napoli Bay area dataset has been chosen because it was already used as a benchmark for the evaluation of the P-SBAS parallel performances on a HPC cluster [17]. Hereafter, we recall the major results that were obtained in order to introduce the analysis presented in this paper, which has been carried out on the AWS cloud. In [17] we thoroughly studied the scalable performances of the overall P-SBAS algorithm by evaluating its speedup with respect to the number of computing elements. As a computational platform, the HPC cluster that is located at the CNR-IREA premises was used. From this analysis it turned out that the major source of inefficiency for the P-SBAS parallel performances was ascribable to the presence of a non-negligible intrinsic processing sequential fraction, which, even if it is liable to a further reduction, remains essentially non-eliminable because of the complex nature of the P-SBAS algorithm. Apart from that factor which was quantitatively

TABLE 2
CNR-IREA Cluster Nodes and Amazon Instance Type Characteristics

	CNR-IREA cluster nodes	m2.4.xlarge	c3.8.xlarge
Processor	Intel Xeon E5—2670	Intel Xeon X5550	Intel Xeon E5—2680
cores	8 (used)*	8	8 (used)*
RAM	384 GB	68 GB	60 GB
Network	Infiniband(56 Gb/s)	1 Gb/s	10 Gb/s
NFS storage bandwidth	300 MB/s	128 MB/s	256 MB/s

* *Note that although the cluster nodes are equipped with 16 cores and the c3.8xlarge instances with 32 cores, only eight have been used for the experimental tests to achieve a proper comparison between the different platforms.*

TABLE 3
P-SBAS Processing Times on the Private CNR Cluster and on the AWS Cloud (for Two Different Resources Configurations) Relevant to the Napoli Bay Area Dataset

Nodes Number	CNR Cluster processing time (hours)	m2.4xlarge processing time (hours)	c3.8xlarge processing time (hours)
1	41	54	40.5
2	23	31	22
4	14	18.5	13.5
8	9	12.5	8.5
16	7	-	6.5

taken into account by computing the relevant P-SBAS Amdahl's law [49], [50], the speedups that were achieved, by exploiting from 1 up to 16 computing nodes, showed to be definitely satisfactory (with a maximum deviation from the Amdahl's law of about 18 percent in correspondence with 16 nodes).

The CNR-IREA cluster used for the scalability analysis consists of 16 nodes, each equipped with 2 CPUs (eight-core 2.6 GHz Intel Xeon E5-2670) and 384 GBytes of RAM (see Table 2). The cluster has a storage shared among different nodes that is implemented through NFS and employs a 56 Gbit/s InfiniBand interconnection. In particular, each processing node is equipped with a direct attached storage (DAS) system in a RAID 5 configuration that ensures a disk access bandwidth of approximately 300 MBytes/s.

The aim of this experimental section is to investigate the P-SBAS performances, in terms of processing times, when the number of computing nodes increases, within the AWS cloud by using different resource configurations and comparing them to the corresponding performances achieved on the CNR-IREA cluster. We carried out such an evaluation by exploiting, among the available EC2 instances, those which have similar characteristics to the used cluster nodes and which satisfy the P-SBAS computing requirements to process the selected data-set. Note that, from here on, we refer to an EC2 instance as to a single computing node. In accordance with our objective, we took both the m2.4.xlarge[2] and c3.8.xlarge instances into consideration because they have both processors that are comparable to the cluster nodes ones, as well as enough RAM to run P-SBAS without incurring page-faulting; they are, however, characterized by different network bandwidths (see Table 2). Actually, the network performance is a key factor to guarantee a good scalability; indeed, one of the major critical issues of the P-SBAS algorithm is the very large amount of data (in terms of inputs, intermediate products and final results) that are read/written during the overall processing into the common NFS storage. This implies that, when the number of exploited instances increases, the number of concurrent processes, which simultaneously read or write on the common storage volume, increases as well, thus remarkably raising the I/O workload. In this case, the network bandwidth can become a bottleneck. Furthermore, for the same reason, we carefully evaluated the Amazon elastic block store (EBS) volume type to be used as a common NFS storage.

More specifically, as a first test bed we exploited the m2.4xlarge instances for both master and worker nodes and we used, as a storage disk, one "provisioned IOPS" (SSD) EBS volume which guarantees 4,000 IOPS (I/O operation per second), i.e., a disk access bandwidth of about 128 MBytes/s (see Table 2).

In Table 3 the elapsed times relevant to the P-SBAS processing carried out by exploiting from 1 up to 8 nodes with this configuration are presented, together with the corresponding times related to the P-SBAS runs which have been performed on the CNR-IREA cluster. The experimental results achieved on the cloud with the m2.4xlarge instances showed computing performances worse than those obtained on the cluster. This is ascribable, in a small part, to the m2.4xlarge processor that is slightly worse than the cluster nodes one, but even more so it is ascribable to the bottleneck due to the network/disk bandwidth which already begins saturating at four nodes in correspondence to the P-SBAS steps that are heavier in terms of I/O workload and becomes fully saturated at eight nodes. Hence, it does not make sense to carry out the test with 16 nodes by exploiting this configuration.

In order to overcome the above mentioned limitation, as a second test bed we used the c3.8xlarge instances which have a 10 Gbit/s network bandwidth and, as a storage, two "provisioned IOPS" (SSD) EBS volumes in a RAID 0 configuration which allows achieving a total throughput of 8,000 IOPS (I/O operation per second) that translates in a disk

2. Note that the m2 instances family at the time of writing has been replaced by the r3 instances family.

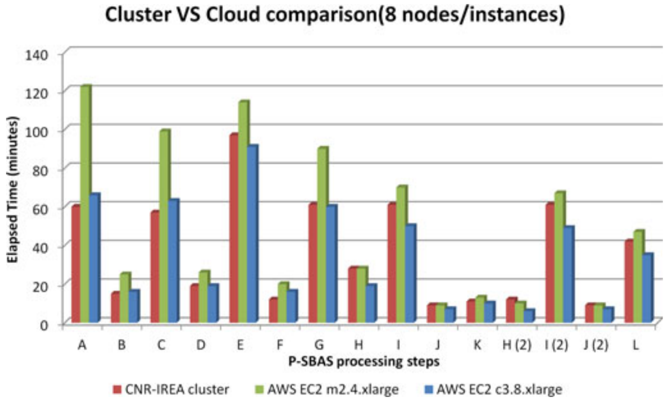

Fig. 4. P-SBAS steps processing times relevant to the CNR-IREA cluster run (red bars), the AWS-m2.4xlarge run (green bars) and the AWS-c3.8xlarge run (blue bars) by exploiting eight nodes/instances. Note that the H, I and J steps are performed twice as explained in Section 2.

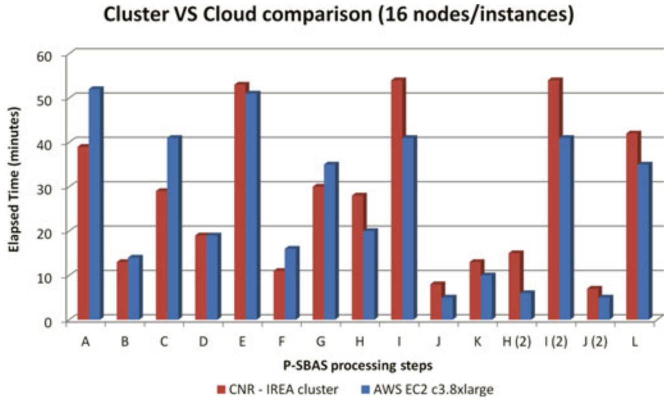

Fig. 5. P-SBAS steps processing times relevant to CNR-IREA cluster run (blue bars) and the AWS-c3.8xlarge run (red bars) by exploiting 16 nodes/instances. Note that the H, I, and J steps are performed twice as explained in Section 2.

access bandwidth of about 256 MBytes/s [47] (see Table 2). The fourth column of Table 3 represents the elapsed times relevant to the P-SBAS processing carried out by exploiting from 1 up to 16 nodes with the c3.8xlarge configuration. It is evident that in this case the performances that are achieved on cloud are definitely comparable to those obtained on the CNR-IREA cluster. Hence, we can conclude that by exploiting the proposed configuration, the cloud environment does not introduce further overhead to the P-SBAS parallel implementation, thus allowing us to take full advantage of P-SBAS scalable performances.

The histograms in Figs. 4 and 5 show in detail the computing times for each step of the P-SBAS processing chain relevant to the runs with 8 and 16 nodes respectively. In particular, in Fig. 4 the performances achieved by exploiting the cluster, the m2.4xlarge and the c3.8xlarge configurations are compared, while in Fig. 5 those relevant to the cluster and the c3.8xlarge are juxtaposed. It is evident from Fig. 4 that the m2.4xlarge configuration always presents the worst performances, but it is worth noting that the largest elapsed times are found in correspondence with the most I/O demanding P-SBAS processing steps, i.e., A, C and G steps. Actually, these are also the only steps in which the CNR-IREA cluster performances are better than the c3.8xlarge configuration ones, due to the cluster's larger network/disk bandwidth. This behavior is accentuated for the 16 nodes run depicted in Fig. 5 where the I/O workload is doubled, with respect to the eight nodes run, in correspondence with the P-SBAS parallel steps. In this case, at the A, C, and G steps of the P-SBAS processing chain, the disk access bandwidth of the RAID storage volume becomes saturated.

The obtained results suggest that, concerning P-SBAS elaborations of data-sets similar to the one that we are considering, exploiting a greater number of nodes would not be of use; while, in the perspective of processing larger datasets which require parallelization on more nodes, a storage volume with a higher I/O throughput can be used by a RAID configuration with more than two EBS volumes (up to now, it is possible to implement a RAID 0 configuration with up to eight volumes).

On the contrary, concerning the most CPU demanding steps of P-SBAS, i.e., I and H, Fig. 5 clearly shows that the times elapsed by exploiting the CNR-IREA cluster are greater than those obtained with the c3.8xlarge configuration. This is due to the slightly better processor of c3.8xlarge instances with respect to the one with the cluster nodes (see Table 2).

4.3 Cost Analysis

As a counterpart to the time performances analysis, we hereafter present the study of the costs relevant to the P-SBAS runs which have been carried out on the AWS cloud by exploiting the two different resources configurations discussed in the previous paragraph. Such an analysis is very useful to evaluate the trade-off between times and costs of an interferometric SAR data processing depending on the exploited AWS computing resources and would allow a scientific user to select the best configuration according to its specific time/cost requirements.

The overall cost of a P-SBAS run is given by the sum of the processing cost, i.e., the cost of the exploited EC2 instances, and the storage cost.

The AWS on-demand EC2 instances are charged for hourly use, therefore the price depends on the hourly cost of the selected instance and on the time of use. Note that the EC2 instances hourly cost can change depending on the region from which instances are rented. In our case the presented costs are relevant to the EU Ireland instances. The provisioned IOPS EBS volumes, instead, are charged by both the provisioned amount of GBytes per month, which is invariant for the same processing, as well as the provisioned IOPS multiplied by the percentage of days you provision for the month.[3]

Table 4 shows the costs relevant to the test that have been carried out and presented in the previous paragraph. In particular the m2.4xlarge hourly cost is 0.980 USD, while the c3.8xlarge cost is 1.680 USD. Moreover, the storage volume is charged by 0.138 USD per GBytes-month of provisioned storage and 0.072 USD per provisioned IOPS-month.

The costs presented in Table 4 have been extracted from the monthly AWS invoice (VAT excluded). However, they can be easily estimated by exploiting the following formula:

$$\begin{aligned} C_{TOT} = C_{Instance} * \left[T_{tot}\big|_{hours} + (N-1)*T_{par}\big|_{hours}\right] \\ + C^{GB}_{Storage} + C^{IOPS}_{Storage}; \end{aligned} \tag{1}$$

3. Note that the EBS volumes pricing changes for different volumes type; for example, standard EBS volumes are only charged by monthly GBytes of provisioned storage.

TABLE 4
AWS Instances and Storage Costs Relevant to the P-SBAS Runs which are Shown in Table 3

Nodes	m2.4xlarge cost (USD)	Single EBS provisioned IOPS cost (USD)	m2.4xlarge overall cost (USD)	c3.8xlarge cost (USD)	RAID configuration cost (USD)	c3.8xlarge overall cost (USD)
1	53	34,8	87,8	69	44,4	113,4
2	60	25,2	85,2	72,4	25,2	97,6
4	72	15,6	87,6	89	25,2	114,2
8	95	15,6	110,6	109	25,2	134,2
16	-	-	-	163	25,2	188,2

where:

$$C^{GB}_{Storage} = C_{monthly\ GB} * GB_{provisioned}, \tag{2}$$

$$C^{IOPS}_{Storage} = C_{monthly\ GB} * IOPS_{provisioned} * \frac{T_{tot}|_{days}}{30}. \tag{3}$$

In the above formulas the symbol C indicates costs, T_{tot} and T_{par} stand for the P-SBAS overall and parallel execution times, respectively, N represents the number of exploited nodes and the notations $(.)|_{hours}$or$(.)|_{days}$ indicate times, which have to be rounded up to hours or days.

In Fig. 6 we show the graphs representing the times/costs trade-off of the experimental tests, which have been carried out on AWS cloud and presented in the previous paragraph. In particular, the P-SBAS performances, depending on the number of exploited EC2 instances, are represented in the time-cost plane for the m2.4xlarge configuration (red) and for the c3.8xlarge configuration (blue). The graphs of Fig. 6 allow an understanding of the best AWS resources configuration and the best number of nodes to exploit in order to process an ENVISAT standard dataset according to the specific time/cost requirements.

In our case, since processing time reduction is a priority to provide results as soon as possible, the choice has been focused on the eight nodes exploitation with either the m2.4xlarge or the c3.8xlarge configuration. The latter configuration presents a 21 percent costs increment and a 33 percent times decrease with respect to the m2.4xlarge one, therefore for the case study discussed in the following section we opted for the eight nodes run with the c3.8xlarge instances and the RAID configured storage.

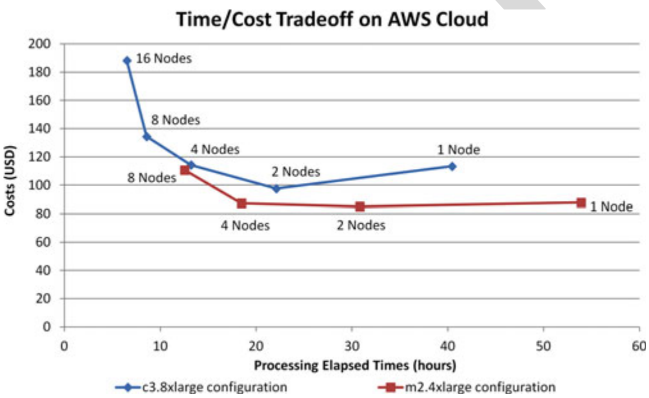

Fig. 6. Trade-off between the elapsed times (horizontal axis) and the overall costs (vertical axis) relevant to the P-SBAS processing of the Napoli Bay area dataset, changing the number of exploited EC2 instances. The red line refers to the configuration with the m2.4xlarge instances and a single EBS volume storage; the blue line refers to the configuration with the c3.8xlarge instances and RAID configured storage with 2 EBS volumes.

It is worth noting that we do not include in this analysis the cost relevant to both the initial phase of cloud resources configuration and deployment, that is, essentially, the storage creation, and data upload as well as the final P-SBAS output download, since they are not really significant amounting to 5 percent of the overall cost. However, for the sake of completeness, in Table 5 the times and costs of all the steps required by the initial deployment phase and data transfer are listed. Moreover, we do not take into account the costs of IDL licenses that are provided by the CNR-IREA institution.

5 Large Area Analysis: The Southern California Experiment

In this section we present the processing of a large interferometric data stack relevant to a portion of southern California that also includes the Los Angeles metropolitan area, in which a large variety of phenomena that induce ground displacements occur, such as active seismic faults, aquifer seasonal extraction and recharge, oil extraction, and agricultural fields. In particular, 172 ENVISAT SAR data acquired from ascending orbits, spanning the 2004-2010 time interval and covering an area of approximately 200×200 km^2 (corresponding to four slightly overlapped ENVISAT frames), have been considered. In this case the overall input data come to approximately 80 GBytes.

The goal of such an experiment is to exploit the results that have been achieved in the previous section to process the aforementioned SAR dataset by properly adjusting cloud computing resources according to the P-SBAS requirements, thus proving that cloud environments allow dealing with large SAR data interferometric processing within short time scales and at limited costs. In particular, we simultaneously processed four ENVISAT frames through the deployment of four identical independent pools of resources interconnected by different subnets. The test bed encompasses the whole 32 instances, eight for the processing of each ENVISAT frame, in the same configuration employed for the first set of experiments (one master node acting both as worker node and NFS server and seven worker nodes) and four storage volumes in a RAID 0 configuration, each one of 500 GBytes.

In Table 6, times and costs related to each dataset processing are represented. Note that the elapsed times are different because they essentially depend on both the number of SAR dataset images and, above all, on the scene's electromagnetic characteristics that have a great impact on the number of the coherent pixels. In this case the overall data processing took less than 17 hours, with a total cost of $821

TABLE 5
Additional Costs Relevant to the P-SBAS Initial Cloud Deployment Phase and Data Transfer Relevant to the Napoli Bay Area Dataset

Task	Time	Cost (USD)
Create the Amazon EBS volumes for your array	30 seconds	-
Attach the Amazon EBS volumes to the instance that you want to host the array	10 seconds	-
EBS volumes pre-warming	2 hours	7.4
Initialize the logical RAID device from the newly attached Amazon EBS volumes	10 seconds	-
Create a file system on your RAID device	10 seconds	-
Create a mount point for your RAID array	5 seconds	-
Upload the data	.5 hours	1.85
Download final output	10 seconds	-

USD. The costs that are shown in Table 6 include processing cost, storage cost, and additional deployment and data transfer cost.

In Fig. 7 the retrieved LOS mean deformation velocity map of the overall area covered by the ENVISAT frames is showed. This image has been realized by first geocoding the four deformation velocity maps relevant to each frame. They have been afterwards properly combined and superimposed on Google Earth. Such a simple post-processing phase has been carried out a posteriori.

It is worth noting the various deformation patterns that are clearly shown in Fig. 7. For some of these the temporal evolution of the measured displacements is represented through the time series plot included in Fig. 7. In particular, Fig. 7b is relevant to the maximum deformation zone of the Santa Ana basin while Fig. 7c is relevant to the injection of fluids in the Downey area. Fig. 7d is relevant to the deformation surrounding the Pomona fault while Fig. 7e represents the land subsidence movement in the Lancaster zone that is very likely caused by water withdrawal.

6 Conclusions

In this paper, a case study concerning the migration to the AWS cloud computing environment of the advanced DInSAR algorithm referred to as P-SBAS, which generates interferometric products useful for the investigation of Earth surface deformation phenomena, was presented and discussed. The cloud deployment was carried out by encapsulating the overall P-SBAS application in virtual machines running on the cloud; moreover, the cloud resources provisioning and configuration phases were implemented in an automatic way. Such an approach allowed us to preserve the intrinsic P-SBAS parallelization strategy and, therefore, to straightforwardly evaluate its performances within a cloud environment by comparing them with those achieved on an HPC in-house cluster.

An experimental framework including two case studies was presented to envisage an effective roadmap for DInSAR processing of large data volumes in cloud computing environments.

The first case study concerned the processing of the complete sequence of 64 SAR images of an ENVISAT frame (about 100×100 km^2) acquired on the Naples Bay area (southern Italy), which has already been used as a benchmark for interferometric analysis. The experimental results showed that, within the AWS cloud, the P-SBAS processing presents very good performance in terms of processing elapsed times when the number of exploited nodes increases. Moreover, a thorough analysis of the P-SBAS processing costs when the type and the number of the exploited AWS cloud resources change was accomplished. Such a study allows the evaluation of the more suitable configuration in accordance to the specific time/cost requirements which, for the considered experiment, proved to be the one set up by the c3.8xlarge instances and the storage in a RAID configuration.

By exploiting the experimental results obtained in the first case study, we carried out the second experimental analysis of processing of a large interferometric dataset consisting of four ENVISAT frames, for a total of 172 images covering an area located in southern California of about 200×200 km^2, including the entire Los Angeles metropolitan zone. More specifically, this analysis was performed by employing 32 EC2 c3.8xlarge instances; the overall P-SBAS processing took less than 17 hours with a cost of about USD 850. The goal of such an experiment was to show that cloud computing environments are suitable for rapid

TABLE 6
P-SBAS Processing Times and Costs within the AWS Cloud Relevant to the Large Dataset of the Southern California Area

Frame numbers	Data set (#images)	c3.8xlarge instances	Elapsed time(hours)	Cost* (USD)
1	47	8	16.7	242
2	44	8	15.5	227
3	43	8	11	192
4	38	8	10.5	192
Overall	**172**	**32**	**16.7**	**853**

**Note that the represented costs include both the instances and storage prices.*

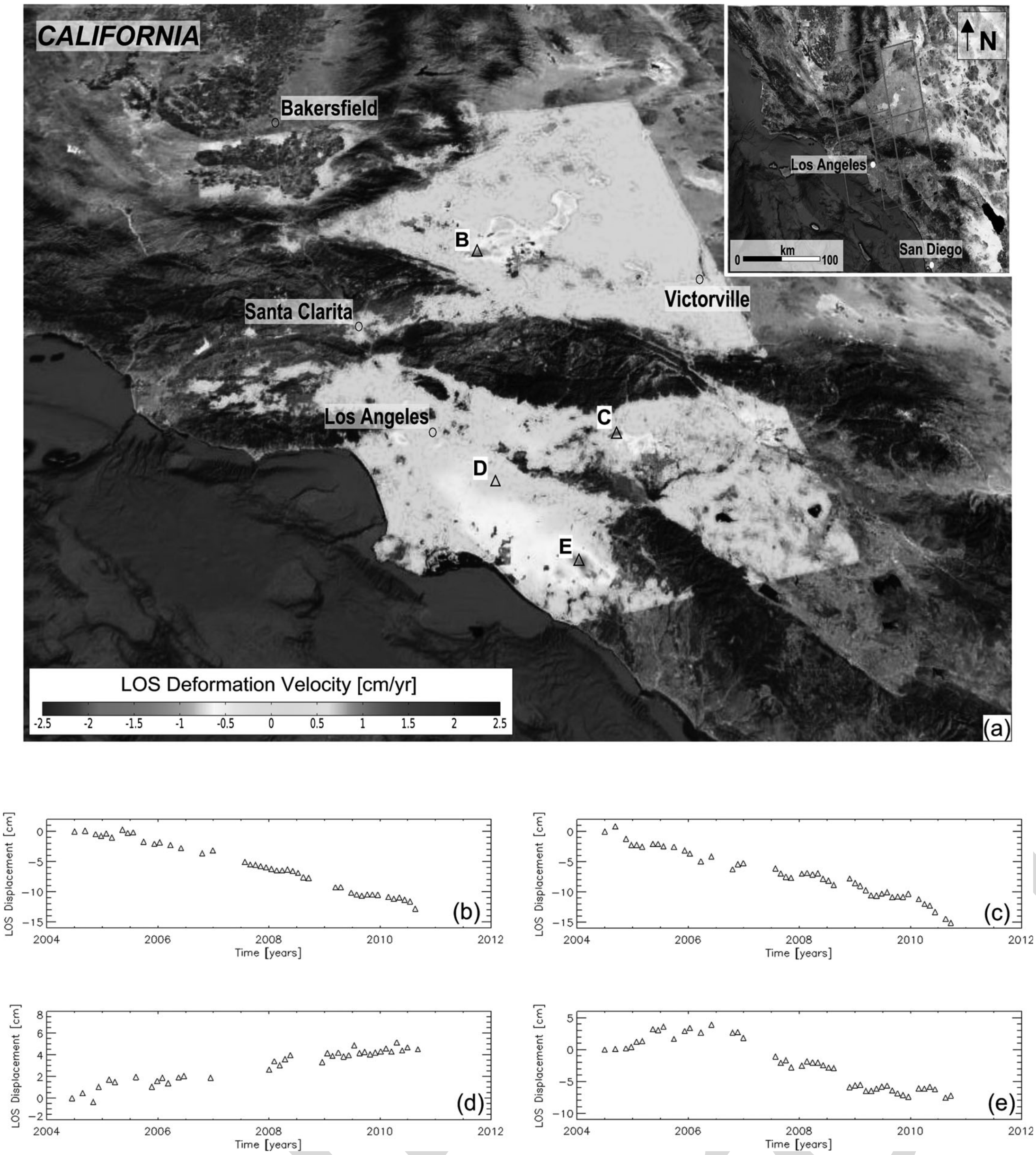

Fig. 7. Mean line of sight (LOS) deformation velocity map on the investigated region of southern California. The plots of the retrieved deformation time series are reported in (b)-(c)-(d)-(e) for four selected zones identified in (a) as B-C-D-E, respectively. Note also that in the inset of (a) it is shown the location and the extension of the four ENVISAT frames exploited for this analysis (see Table 6).

interferometric processing of large data volumes at limited costs. This implies that surface deformation analyses at a country and even continental scale are feasable with a potential great impact on the ongoing development of initiatives such as the case of the european plate observing system (EPOS) [51].

Although the presented analysis is fully focused on the Amazon cloud exploitation, the proposed solution is also suitable to be deployed on other cloud environments. This is, for instance, the case of the Helix Nebula-the Science Cloud initiative for the provisioning of a sustainable scientific cloud [52] which is currently under development.

We can conclude that this study supports the possibility to provide, in the future, the P-SBAS processing as a multi-user service through the use of cloud computing platforms. The development of such a type of service can be of great benefit to provide a response to the current SAR remote sensing scenario that is characterized by the huge and ever increasing availability of data and by a large number of potential P-SBAS users represented by research institutions, governmental bodies, and companies. Moreover, the cloud environments extensive availability of computing resources perfectly matches with the perspective of processing, also on a continuous basis, the huge SAR data flow which is going to be provided by the recently launched SENTINEL-1A satellite that is characterized by a very large spatial coverage (250×250 km^2) and a very high revisit time (12 days). In this case, the joint exploitation of the P-SBAS processing and of cloud platforms could allow us to investigate Earth's surface deformation at nearly global scale, with continuously up-to-date results within reduced time frames.

ACKNOWLEDGMENTS

This work was supported in part by the Italian Ministry of University and Research (MIUR) under the project "Progetto Bandiera RITMARE," and in part by the Italian Civil Defence Department (DPC) of the Prime Minister's Office. This work has been carried out through the I-AMICA (Infrastructure of High Technology for Environmental and Climate Monitoring—PONa3_00363) project of Structural improvement financed under the National Operational Programme (NOP) for "Research and Competitiveness 2007-2013," cofounded with European Regional Development Fund (ERDF) and National resources. The ENVISAT SAR data have been provided by the European Space Agency through the Virtual Archive 4. The DEM of the investigated zone was acquired through the SRTM archive. The authors would like to thank S. Guarino, F. Parisi, and M.C. Rasulo for their technical support. The authors are also grateful to Dr. Paul Lundgren who proofread the English text.

REFERENCES

[1] D. Massonnet, M. Rossi, C. Carmona, F. Adragna, G. Peltzer, K. Feigl, and T. Rabaute, "The displacement field of the Landers earthquake mapped by radar interferometry," *Nature*, vol. 364, no. 6433, pp. 138– 142, Jul. 1993.

[2] K. Mogi, "Relations between the eruptions of various volcanoes and the deformations of the ground surfaces around them," *Bull. Earthquake Res. Instit.*, vol. 36, pp. 99–134, 1958.

[3] M.A. Chinnery, "The deformation of the ground around surface faults," *Bull. Seismol. Soc. Am.*, vol. 51, pp. 355–372, 1961.

[4] W.E. Farrell, "Deformation of the Earth by surface loads," *Rev. Geophys.*, vol. 10, no. 3, pp. 761–797, Aug. 1972.

[5] Y. Okada, "Surface deformation due to shear and tensile faults in a half-space," *Bull. Seismol. Soc. Am.*, vol. 75, no. 4, pp. 1135–1154, Aug. 1985.

[6] P. A. Hsieh, "Deformation-induced changes in hydraulic head during ground-water withdrawal," *Groundwater*, vol. 34, pp. 1082–1089, 1996.

[7] R. F. Yerkes and R. O. Castle. (1969). Surface deformation associated with oil and gas field operations in the United States. *Land Subsidence.* [Online], pp. 55–66. Available: pubs.rsc.org

[8] F. M. Orr, Jr., "CO 2 capture and storage: Are we ready?," *Energy Environ. Sci.*, vol. 2, pp. 449–458, Mar. 2009.

[9] S. Rackley. (2009). *Carbon Capture and Storage.* [Online]. Available: books.google.com.

[10] C. Q. C. S. M. Shixiang and H. Dade, "Numerical simulation study of ground deformation in underground mining," *Metal Mine*, 6, 005, 2004.

[11] M. P. O' Reilly and B. M. New. (1982). New Settlements above tunnels in the United Kingdom-their magnitude and prediction. *Publ.: Instit. Mining. Metallurgy* [Online]. Available: trid.trb.org

[12] R.W. Clough and R. J. Woodward, "Analysis of embankment stresses and deformations," *J. Soil Mechanics Found. Div.*, vol. 93, pp. 529–549, 1967.

[13] R. Burgmann, P. A. Rosen, and E. J. Fielding, "Synthetic aperture radar interferometry to measure Earth's surface topography and its deformation," *Annu. Rev. Earth Planet. Sci.*, vol. 28, pp. 169–209, May 2000.

[14] D. Massonnet and K. L. Feigl, "Radar interferometry and its application to changes in the Earth's surface," *Rev. Geophys.*, vol. 36, pp. 441–500, 1998.

[15] G. Franceschetti and R. Lanari, *Synthetic Aperture Radar Processing*, Boca Raton, FL, USA: CRC, Mar. 1999.

[16] P. Berardino, G. Fornaro, R. Lanari, and E. Sansosti, "A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2375–2383, Nov. 2002.

[17] F. Casu, S. Elefante, P. Imperatore, I. Zinno, M. Manunta, C. De Luca, and R. Lanari, "SBAS-DInSAR parallel processing for deformation time-series computation," *IEEE J. Select. Topics Appl. Earth Observations Remote Sens.*, vol. 7, no. 8, pp. 3285,3296, Aug. 2014.

[18] Geohazard Supersites and Natural Laboratories. [Online]. Available: http://eo-virtual-archive4.esa.int/, 2006-2013.

[19] E-GEOS, an ASI/Telespazio Company. [Online]. Available: http://www.e-geos.it/products/cosmo.html, 2015.

[20] AIRBUS Defence and Space. [Online]. Available: http://www.geo-airbusds.com/terrasar-x

[21] S. Salvi, S. Stramondo, G. J. Funning, A. Ferretti, F. Sarti and A. Mouratidis, "The sentinel-1 mission for the improvement of the scientific understanding and the operational monitoring of the seismic cycle," *Remote Sens. Environ.*, vol. 120, pp. 164–174, May 2012.

[22] Copernicus - The European Earth Observation Programme. [Online]. Available : http://www.copernicus.eu, 2015.

[23] ESA - Sentinel Online. [Online]. Available: https://sentinel.esa.int/web/sentinel/home, 2000-2015.

[24] A. Rucci, A. Ferretti, A. MontiGuarnieri and F. Rocca, "Sentinel 1 SAR interferometry applications: The outlook for sub millimeter measurements," *Remote Sens. Environ.*, vol. 120, pp. 156–163, May 2012.

[25] Y. Demchenko, Z. Zhao, P. Grosso, A. Wibisono, and C. De Laat, "Addressing big data challenges for scientific data infrastructure system and network engineering," *IEEE 4th Int. Conf. Cloud Comput. Technol. Sci.*, 2012, pp. 614–617.

[26] A. Gupta, P. Faraboschi, F. Gioachin, L.V. Kale, R. Kaufmann, B.-S. Lee, V. March, D. Milojicic, and C. H. Suen, "Evaluating and improving the performance and scheduling of HPC applications in cloud," *IEEE Trans. Cloud Comput*, Doi: 10.1109/TCC.2014.2339858.

[27] A. Gupta and D. Milojicic, "Evaluation of HPC applications on cloud," in *Proc. IEEE 6th Open Cirrus Summit*, Oct. 2011, pp. 22–26, Doi: 10.1109/OCS.2011.10.

[28] I. Sadooghi, J. H. Martin, T. Li, K. Brandstatter, Y. Zhao, K. Maheshwari, T. P. Pi. de LacerdaRuivo, S. Timm, G. Garzoglio, and G I. Raicu, "Understanding the performance and potential of cloud computing for scientific applications," *IEEE Trans. Cloud Comput.*, Doi: 10.1109/TCC.2015.2404821.

[29] E. Roloff, M. Diener, A. Carissimi, and P. O. A. Navaux, "High performance computing in the cloud: deployment, performance and cost efficiency," in *Proc. IEEE 4th Int. Conf. Cloud Comput. Technol. Sci.*, 3–6 Dec. 2012, pp. 371–378.

[30] A. Ferretti, C. Prati, and F. Rocca, "Permanent scatterers in SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 1, pp. 8–20, Jan. 2001.

[31] A. Hooper, H. Zebker, P. Segall, and B. M. Kampes, "A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers," *Geophys. Res. Lett.*, vol. 31, no. 23, p. L23611, Dec. 2004.

[32] M. Kseneman, D. Gleich, and A. Chowdhury, "Despeckling synthetic aperture radar images with cloud computing using graphics processing units," in *Proc. 5th Int. Conf. Pervasive Comput. Appl.*, 1–3 Dec. 2010, pp. 195–200.

[33] S. Ibrahim, H. Jin, L. Lu, L. Qi, S. Wu, and X. Shi "Evaluating MapReduce on virtual machines: the HadoopCase," in *Proc. 1st Int. Conf. Cloud Comput.*, 2009, pp 519–528.

[34] E. Molina-Estolano, M. Gokhale, C. Maltzahn, J. May, J. Bent, and S. Brandt, "Mixing Hadoop and HPC workloads on parallel filesystems," in *Proc. 4th Annu. Workshop .Petascale Data Storage*, 2009, pp. 1–5.

[35] J. J. Rehr, F. D. Vila, J. P. Gardner, L. Svec, and M. Prange "Scientific computing in the cloud," *IEEE Comput. Sci. Eng.*, vol. 12, no. 3, pp. 34–43, May-Jun. 2010

[36] S. Hardman, A. Riofrio, K. Shams, D. Freeborn, P. Springer, and B. Chafin, "Enabling earth science through cloud computing," in *Proc. Aerospace Conf.*, 3–10 Mar. 2012, pp. 1–8.

[37] P. Rosen, K. Shams, E. Gurrola, B. George, and D. Knight, "InSAR scientific computing environment on the cloud," in *Proc. Am. Geophys. Union Conf.*, San Francisco, CA, USA, 3–7 Dec., Vol. 1, p. 1508, 2012.

[38] S. Elefante, P. Imperatore, I. Zinno, M. Manunta, E. Mathot, F. Brito, J. Farres, W. Lengert, R. LanariandF. Casu, "SBAS-DInSAR time series generation on cloud computing platforms," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 21–26 Jul. 2013, pp. 274–277.

[39] J.C. Curlander and R. McDonough, *Synthetic Aperture Radar-System and Signal Processing*. Hoboken, NJ, USA: Wiley, 1991.

[40] E. Sansosti, P. Berardino, M. Manunta, F. Serafino, and G. Fornaro, "Geometrical SAR image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2861–2870, Oct. 2006.

[41] A. Pepe and R. Lanari, “On the extension of the minimum cost flow algorithm for phase unwrapping of multitemporal differential SAR interferograms,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2374–2383, Sep. 2006.
[42] A. Pepe, E. Sansosti, P. Berardino, and R. Lanari, “On the Generation of ERS/ENVISAT DInSAR Time-Series Via the SBAS Technique,” *IEEEGeosci. Remote Sens. Lett.*, vol. 2, no. 3, pp. 265–269, 2005.
[43] M. Bonano, M. Manunta, M. Marsella, and R. Lanari, “Long-term ERS/ENVISAT deformation time-series generation at full spatial resolution via the extended SBAS technique,” *Int. J. Remote Sens.*, vol. 33, pp. 4756–4783, 2012.
[44] V. Andrikopoulos, T. Binz, F. Leymann, and S. Strauch, “How to adapt applications for the cloud environment,” *Computing*, vol. 95, no. 6, pp. 493–535, 2013.
[45] Exelis Visual Information Solution. [Online]. Available: http://www.exelisvis.com, 2015.
[46] R. Sandberg, D. Goldberg, S. Kleiman, D. Walsh and B. Lyon, “Design and implementation of the SUN network filesystem,” in *Proc. USENIX Conf.*, 1985.
[47] Amazon Web Services. [Online]. Available: http://aws.amazon.com/, 2015.
[48] P. Barham, B. Dragovic, K. Fraser, S. Hand, Steven, T. Harris, H. Alex, R.Neugebauer, I. Pratt, Ian, and A. Warfield, “Xen and the art of virtualization,” *SIGOPSOper. Syst. Rev.*, vol. 37, pp. 164–177, Dec. 2003.
[49] G. Hager and G. Wellein. *Introduction to High Performance Computing for Scientists and Engineers*. Boca Raton, FL: CRC Press, 2010.
[50] H. El-Rewini and M. Abd-El-Barr, *Advanced Computer Architecture and Parallel Processing*. Hoboken, NJ, USA: Wiley, 2005.
[51] EPOS-European Plate Observing System. [Online]. Available: www.epos-eu.org, 2012–2012.
[52] HELIX NEBULA - The Science Cloud. [Online]. Available: http://www.helix-nebula.eu., 2014.

Ivana Zinno received the Laurea degree (summa cum laude) in telecommunication engineering and the PhD degree in electronic and telecommunication engineering both from the University of Naples Federico II, Naples, in 2008 and 2011, respectively. In 2011, she received a grant from the University of Naples to be spent at the Department of Electronic and Telecommunication Engineering for research in the field of remote sensing. Since January 2012, she has been a research fellow at IREA (Istituto per il Rilevamento Elettromagnetico dell’Ambiente)-CNR (National Research Council), Naples, mainly working on the development of advanced Differential SAR Interferometry (DInSAR) techniques for deformation time series generation by also exploiting parallel computing platforms. Her main research interests include the field of microwave remote sensing; in particular they concern differential SAR interferometry applications for the monitoring of surface displacements and information retrieval from SAR data by exploiting fractal models and techniques. She is a member of the IEEE.

Lorenzo Mossucca received the Laurea degree in computer engineering from the Polytechnic of Turin. From 2007, he works as a researcher at Istituto Superiore Mario Boella (ISMB) in the research unit Infrastructures and Systems for Advanced Computing (IS4AC). His research interests include studies on distributed databases, distributed infrastructures, Grid and Cloud Computing. In the last years, he is focusing his research on migration of scientific applications to cloud, in particular in bioinformatics and earth sciences fields. He has authored and coauthored about 30 international journal, book chapter, and conference papers, he joins in Technical Program Committee and as a reviewer in many international conferences and editor of the book *Cloud Computing with e-science applications* (CRC Press, 2015).

Stefano Elefante received the Laurea degree (summa cum laude) from the University of Naples, Napoli, Italy and the PhD degree in aerospace engineering from the University of Glasgow, Scotland, United Kingdom, in 1997 and 2001, respectively. He holds a PgCert in financial engineering from the Columbia University, New York, NY, and in applied statistics from the Birkbeck College, University of London, London, United Kingdom, since 2008 and 2011, respectively. He has more than 10 years of diverse professional experience in academia and industry, during which he has been involved in different scientific fields from statistical genetics and financial mathematics to aerospace technologies. From 2005 to 2009, he was a system analyst and research engineer with Boeing Research & Technology Europe, Madrid, Spain, where he conducted extensive research in the field of air traffic management developing and implementing aerospace systems for improving airport and airspace efficiency. Since 2011, he has been holding a research position with IREA-CNR, Napoli, Italy. His research interests include investigating innovative mathematical methodologies for remote sensing applications. He is currently developing novel parallel algorithms for synthetic 852 aperture radar interferometry (InSAR) within cluster, grid and cloud computing environments.

Claudio De Luca received the Laurea degree (110/110) in telecommunication engineering from the University of Naples “Federico II”, Naples, in 2012. He is currently working towards the PhD degree in computer and automatic engineering with the Department of Electrical Engineering and Information Technology of the University of Naples “Federico II”. His main research interests include field of cloud computing solution for intensive processing of remote sensing data, development of advanced algorithm for Sentinel-1 SAR, and InSAR data processing.

Valentina Casola received the PhD degree in computer engineering from the Second University of Naples in 2004. She is an assistant professor at the Department of Electrical Engineering and Information Technology of the University of Naples Federico II, Italy. She has published more than 70 papers in journals, conference proceedings, and books. Her research activities are both theoretical and experimental and are focused on security methodologies to design and evaluate distributed systems, including cyber physical infrastructures, cloud systems, and web services. These activities are led in cooperation with academic institutions and industrial partners within national and international projects.

Olivier Terzo received degree in electrical engineering technology and industrial informatics at the University Institute of Nancy, France, the MSc degree in computer engineering, and the PhD degree in electronic engineering and communications from the Polytechnic of Turin, Italy. He is a senior researcher at the Istituto Superiore Mario Boella (ISMB), from 2004 to 2009, he worked in the e-security laboratory, mainly with a focus on P2P protocols, encryption on embedded devices, security of routing protocols, and activities on grid computing infrastructures, and from 2010 to 2013, he was the head of the Research Unit Infrastructure Systems for Advanced Computing (IS4AC) at ISMB. Since 2013, he is the head of Research Area: Advanced Computing & Electromagnetics (ACE), dedicated on the study and implementation of computing infrastructure based on virtual grid and cloud computing and to the realization of theoretical and experimental activities of antennas, electromagnetic compatibility, and applied electromagnetics. His research interests include hybrid private and public cloud distributed infrastructure, grid and virtual grid, mainly activities are on applications integration in cloud environments. He published about 60 papers in conferences, journals, and book chapters and he is an editor of the book *Cloud Computing with e-Science Applications* (CRC Press, 2015).

Francesco Casu received the laurea degree (summa cum laude) and the PhD degree in electronic engineering from the University of Cagliari, Cagliari, Italy, in 2003 and 2009, respectively. Since 2003, he has been with the IREA-CNR (Napoli, Italy) where he currently holds a permanent researcher position. He was a visiting scientist with the University of Texas at Austin in 2004, the Jet Propulsion Laboratory, Pasadena in 2005, and the Department of Geophysics at the Stanford University in 2009. His main research interests include the DInSAR field, in the multi-pass interferometry (particularly concerning the improvement of the SBAS-DInSAR algorithm) and in the SBAS-DInSAR measurement assessment, with particular emphasis on novel generation satellite constellations such as COSMO-SkyMed, TerraSAR-X, and Sentinel-1. More recently, he has been involved in the development of DInSAR algorithms for unsupervised processing of huge SAR data archives by exploiting high performance computing platforms, such as GRID and cloud computing ones. Moreover, he acts as a reviewer of several peer-reviewed international journals.

Riccardo Lanari (M'91–SM'01-FM'13) graduated in electronic engineering (summa cum laude) from the University of Napoli, Federico II, Napoli, in 1989. In the same year, following a short experience at ITALTEL SISTEMI SPA, he joined IRECE and after Istituto per il Rilevamento Elettromagnetico dell'Ambiente (IREA), a Research Institute of the Italian National Research Council (CNR), Napoli, where, since November 2011, he was the institute director. He has lectured in several national and foreign universities and research centers. He was an adjunct professor of electrical communication with the l'Università del Sannio, Benevento, Italy, from 2000 to 2003, and from 2000 to 2008, he was the lecturer of the synthetic aperture radar (SAR) module course of the International Master in Airbone Photogrammetry and Remote Sensing offered by the Institute of Geomatics, Barcelona, Spain. He was a visiting scientist at different foreign research institutes, including the Institute of Space and Astronautical Science, Japan, in 1993; German Aerospace Research Establishment (DLR), Germany, in 1991 and 1994; and Jet Propulsion Laboratory, Pasadena, CA, in 1997, 2004, and 2008. His main research activities are in the SAR data processing field as well as in SAR interferometry techniques; on this topic, he is the holder of two patents, and he has authored or coauthored 80 international journal papers and the book *Synthetic Aperture Radar Processing* (1999, CRC Press). He is a distinguished speaker of the Geoscience and Remote Sensing Society of IEEE, and he has served as the chairman and as a technical program committee member at several international conferences. Moreover, he acts as a reviewer of several peer-reviewed international journals. He received a NASA recognition and a group award for the technical developments related to the Shuttle Radar Topography Mission. He is a fellow member of the IEEE.