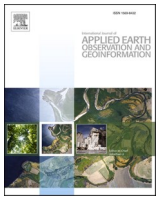
International Journal of Applied Earth Observation and Geoinformation 118 (2023) 103276

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Unsupervised detection of InSAR time series patterns based on PCA and K-means clustering

Davide Festa [a,*], Alessandro Novellino [b], Ekbal Hussain [b], Luke Bateson [b], Nicola Casagli [a,c], Pierluigi Confuorto [a], Matteo Del Soldato [a], Federico Raspini [a]

[a] Department of Earth Sciences, University of Firenze, Via La Pira 4, Firenze, Italy
[b] British Geological Survey, Keyworth, Nottinghamshire NG12 5GG, UK
[c] National Institute of Oceanography and Applied Geophysics – OGS, Borgo Grotta Gigante n. 42/c, Sgonico (Trieste), Italy

ARTICLE INFO

Keywords:
InSAR time series
PCA
Unsupervised learning
Data mining
Automated approach
EO data downstream

ABSTRACT

The need for implementing efficient value-adding tools able to optimise Earth Observation data usage, compels the scientific community to find innovative solutions for the downstream of Earth Observation information. In this paper we present an unsupervised and automated approach based on Principal Component Analysis (PCA) and K-means clustering to detect patterns of natural or anthropogenic ground deformation from Interferometric Synthetic Aperture Radar (InSAR) Time Series. For our proof-of-concept, we focus on the Valle d'Aosta region (Northwest Italy) where mass wasting processes frequently occurs, interacting with human activities and infrastructures. The large volumes of Sentinel-1 data produced allows for retrieving horizontal and vertical Time Series from multi-geometry data fusion of Line-of-Sight (LOS) InSAR measurements. The added benefit of combining ascending/descending InSAR data and interpolating displacements in time at different time steps is here explored prior to data dimensionality reduction and feature extraction through PCA. The retrieved principal components serve as a continuous solution for cluster membership indicators in the K-means clustering method, allowing to define spatially and temporally coherent displacement phenomena. The signal of the ground deformation clusters is then deconstructed into the underlying trend and seasonality components to enhance the interpretability of the classified satellite InSAR features. Using InSAR Time series data spanning 2014–2020, the proposed approach detects several slope movements and anthropogenic deformations with both linear and seasonal displacement behaviours. The results demonstrate the potential applicability of our transferable approach to the development of automated ground motion analysis systems.

1. Introduction

The large volumes of freely accessible Sentinel-1 (S-1) data acquired worldwide have fostered the deployment of Time Series Interferometric Synthetic Aperture Radar (TS-InSAR) to reliably monitor and detect ground deformation. Nowadays, space-borne radars can provide consistency in data collection, a short revisit time (up to 6 days before the S-1B anomaly dated to December 2021; ESA, 2022) and the Terrain Observation of Progressive Scans (TOPS) acquisition mode (De Zan & Monti Guarnieri, 2006). Consequently, InSAR techniques have reached a remarkable state of maturity enabling the screening of millimetre scale deformations over the Earth's surface (Raspini et al., 2018). Many well-established data processing pipelines have been developed in this framework. The flourishing of several advanced multi-temporal Differential SAR Interferometry (A-DInSAR) techniques such as Permanent Scatterers (PSInSAR™; Ferretti et al., 1999), Small baseline subset (SBAS; Berardino et al., 2002), SqueeSAR™ (Ferretti et al., 2011), coherent pixels technique (CPT; Blanco-Sánchez et al., 2008) and stable point network (SPN; Arnaud et al., 2003) have greatly prompted the use of space-borne data to gain a better understanding of natural hazards and anthropogenic deformations. Specifically, Persistent Scatterer Interferometry (PSI) techniques have been deployed for mapping and monitoring slow-moving landslides (Intrieri et al., 2018; Zhang et al., 2020), subsidence due to groundwater exploitation, loading or underground mining (Bateson et al., 2015; Ciampalini et al., 2019; Ezquerro et al., 2020), volcanic upheaval (de Luca et al., 2022) and multiple sources of motion (Festa et al., 2022).

The automation of large InSAR data post-processing (i.e., the

* Corresponding author at: Dipartimento di Scienze della Terra, Università degli Studi di Firenze, Via La Pira, 4, 50121 Florence, Italy.
E-mail address: davide.festa@unifi.it (D. Festa).

https://doi.org/10.1016/j.jag.2023.103276
Received 17 November 2022; Received in revised form 18 March 2023; Accepted 21 March 2023
Available online 29 March 2023

European Ground Motion Service; Crosetto et al., 2020) has become an essential task to make ground motion data readily understandable by a broad range of nonexpert stakeholders and end-users. So far, most of studies aiming to map and characterise ground motion mainly rely on the average displacement rate obtained from fitting a linear regression model to the entire TS-InSAR data (Festa et al., 2022; Lu et al., 2012; Novellino et al., 2019; Tomás et al., 2019). Existing spatial clustering methods (Tomás et al., 2019) propose to aggregate ground deforming reflectors without providing any information regarding the different motion trends coexisting within the same unstable area. Only few methods have been tested so far for detecting and classifying deformation signals from individual time series of ground deformation. Some approaches address TS-InSAR decomposition (Schlögl et al., 2021; Xiong et al., 2021), while others perform either manual (Cigna et al., 2011) or automated (Ansari et al., 2021; Berti et al., 2013; Chang & Hanssen, 2016) TS-InSAR-based methods for retrieving spatial and temporal trends in land motion. Cigna et al. (2011) relies on the manual interpretation of large A-DInSAR datasets, which is a time-consuming activity where the results may be affected by subjectivity. Berti et al. (2013) and Chang & Hanssen (2016) try to trace back TS-InSAR temporal displacement signals to a priori pattern models that would inhibit the scalability of the method in classifying unknown temporal patterns in large InSAR datasets. Also, the advantage of combining ascending/descending radar data is not addressed by Ansari et al. (2021). The above-mentioned gaps, together with the actual need for thoroughly managing large amount of InSAR datasets, are tackled within this manuscript.

For systematic large-scale mapping, the large volume of TS-InSAR information enclosed within millions of Measurement Points (MPs) calls for efficient data mining approaches to disclose the underlying displacement patterns. Therefore, we explore the applicability of an automated unsupervised learning approach by applying a statistical dimensionality-reduction method which includes a Principal Component Analysis (PCA; Jolliffe & Cadima, 2016) followed by K-means clustering (Wu et al., 2008) to detect non-stationary deforming trends at a regional scale. The novelties brought by this work, are the following:

- The proposed data mining approach is applied to the decomposed vertical and horizontal components displacement TS-InSAR rather than to the LOS values.
- The robustness of the clustering approach is evaluated against different TS-InSAR interpolation techniques.
- Every retrieved cluster is further decomposed into trend and seasonality components for a complete mapping of the temporal displacement behaviour. To the best of our knowledge, this approach has not been tested on PSI TS-InSAR before.

For our proof-of-concept, we focus on the Valle d'Aosta region which is among the regions with the highest mortality rate from landslides in Italy and where other identified sources of ground motion includes subsidence and mining-related displacements (Confuorto et al., 2021).

2. Study area

Valle d'Aosta is a region located in the north-western part of the Italian Alpine arc and is bordered by Switzerland to the North and by France to the West. The 3,262 km^2 area encompasses a mostly mountainous landscape where almost 50% of its territory has an elevation higher than 2,000 m.a.s.l.

As a result of Alpine orogenesis and Quaternary glaciation, Valle d'Aosta is a complex structural-geomorphological context (Dal Piaz et al., 2001) where a main central east–west valley oriented is framed by several north–south oriented sub-valleys, constituting an ideal scenario for the application of the InSAR technique (i.e., ground deformation detection along N-S direction it is poorly constrained by the satellite geometry). The post-glacial action (Holocene to present) has influenced the relief and the ongoing slope dynamics (Carraro & Giardino, 2004). Indeed, Valle d'Aosta is extensively characterised by gravitational mass movements such as shallow landslides, rockfalls, rock glaciers and deep-seated gravitational slope deformation (DSGSD) which sometimes involve entire mountain flanks.

3. Data

3.1. PSI datasets

The launch of the C-band S-1 constellation has allowed the acquisition of huge volumes of freely accessible radar images with an unprecedented temporal sampling over Europe. In this study, 256 and 261 S-1 scenes covering Valle d'Aosta were exploited from ascending and descending acquisition geometries, respectively (Table 1). The stacks of images, acquired with the Interferometric Wide (IW) swath mode by both S-1A and S-1B and spanning the time interval from 19 October 2014 to 12 May 2020, were processed by means of the SqueeSAR™ algorithm (Ferretti et al., 2011) as part of the Valle d'Aosta S-1 monitoring service (Crosetto et al., 2020). The type of deformation model used to process the InSAR data is linear. Images have been captured along satellite track n.88 in ascending orbit and along-track n.66 in descending orbit (Fig. 1). The SqueeSAR™ processing chain was able to provide 365,384 ascending and 367,263 descending displacement TS-InSAR subdivided between point-wise coherent scatterers (PS) and partially coherent distributed scatterers (DS), allowing a rather high MP density (ca. 56 MP/km^2). A coherence threshold of 0.75 has been applied to the deployed InSAR data.

3.2. Regional inventories

Considering the historical causes of motion in Valle d'Aosta, the datasets used in the validation of our procedure include:

- landslide inventory from the Italian Landslide Inventory project (IFFI; Trigila et al., 2007), where 2,188 landslides and 141 DSGSDs have been mapped until 2016 throughout the Valle d'Aosta territory (Fig. 2), covering more than 590 km^2 (ca. 18% of the entire region).
- Glaciers database, with 900 rock glaciers marked as active or relict landforms (SCT, n.d.). Besides the well know geohazards, other sources of land motion have been linked to anthropogenic surface displacements, namely subsidence due to groundwater over-exploitation and mining-related activities (Confuorto et al., 2021).
- Database of areas subject to past/recent quarrying and dump activities (Fig. 2) available online at SCT (n.d.).

4. Methods

The workflow of the proposed TS-InSAR data mining approach (Fig. 3) is performed through a Python-based code, which is freely available on Github at https://github.com/maybedave/InSAR-Time-Series-Clustering. The code consists of three main steps: 1) Spatial and temporal post-processing of PSI dataset to retrieve newly interpolated

Table 1
S-1 parameters and PSI results.

S-1 parameters and PSI results		
SAR satellite	S-1A&B	S-1A&B
Satellite geometry	Ascending	Descending
Track	88	66
Sensor mode	IW	IW
LOS angle	38.59°	43.12°
Azimuth angle	8.87°	10.49°
No. of scenes	256	261
No. of MPs	365,384	367,263
Time spans	04–11-2014 to 12–05-2020	10–10-2014 to 11–05-2020

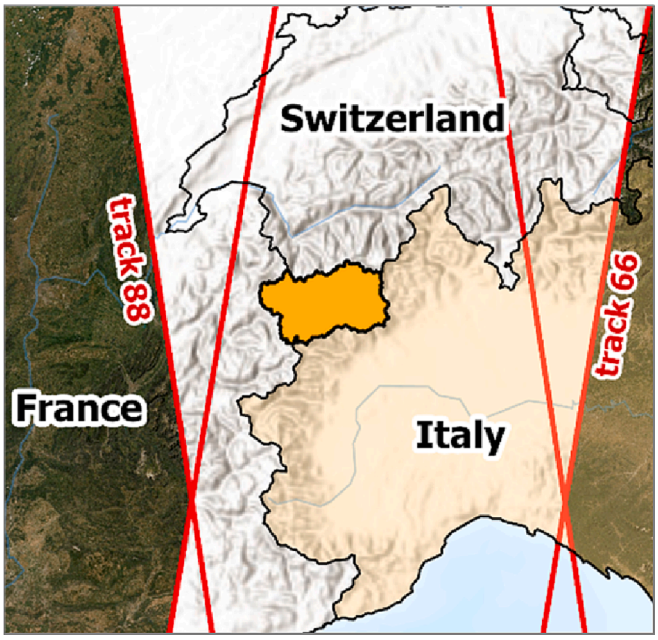


Fig. 1. S-1 satellite tracks along which the deployed radar images were captured.

vertical and horizontal displacement TS-InSAR; 2) PCA-based dimensionality reduction and features retrieval; 3) Unsupervised K-Means learning for TS-InSAR automated clustering and decomposition of cluster's centroids.

4.1. TS-InSAR post-processing

Multi-geometry InSAR data fusion aims to reproject and estimate the LOS observed displacement signal along vertical and horizontal directions based on the combination of measurements acquired from two or more different SAR satellite orbit acquisitions. Independent InSAR datasets can be merged on the (i) spatial and (ii) temporal level if LOS measurements are referred to the same location and period.

The accomplishment of the (i) spatial interpolation is performed by merging sparse MPs from overlapping geometries into spatially coherent domains. We propose a squared gridding of point-like data consisting of an adaptive scaling process able to enclose at least one MP from ascending and descending geometries. The process starts by assessing the extent of the InSAR datasets to produce each individual grid cell as a geometric polygon. The optimal size of the grid cell is automatically assessed as follows: considering an interval $n = [10..1000]$ the totality of cells C_n produced at the n grid size is evaluated against $Cad_n = Ca_n \cup Cd_n$, where Ca_n and Cd_n are the number of cells with at least one MP from the ascending and descending geometry, respectively. The optimal size of the grid cell corresponds to the smallest grid size (so to capture motion at higher resolution using more cells – C_n) but large enough to contain at least one ascending and descending MP. This size is evaluated by plotting the results of C_n/Cad_n against the n range of size values and by identifying the point of maximum curvature (Fig. 4) through the Kneed algorithm (Satopaa et al., 2011). Empty grid cells resulting from the absence of simultaneous data from both geometries are automatically deleted.

Also, (ii) interpolation in time is required to retrieve vertical and horizontal TS-InSAR since the image acquisition dates are usually different in each viewing geometry. The proposed rationale performs the fusion of LOS TS-InSAR data on a cell-by-cell basis on the interpolated grid, thus resulting in unique ascending and descending TS-InSAR for each grid squared polygon. After aggregating and averaging the displacement values per time step, the approach pipeline follows with the selection of the ascending and descending overlapping time window (oTw) as the new temporal frame for TS-InSAR resampling. Then, the new frequency conversion nf (1) is automatically computed as:

$$nf = oTw/(S_{asc} + S_{desc}) \quad (1)$$

where oTw unit is days, while S_{asc} and S_{desc} are the total numbers of ascending and descending geometry scenes. The interpolation of LOS velocities ($V_{asc}; V_{desc}$) to a new frequency enables the computation of

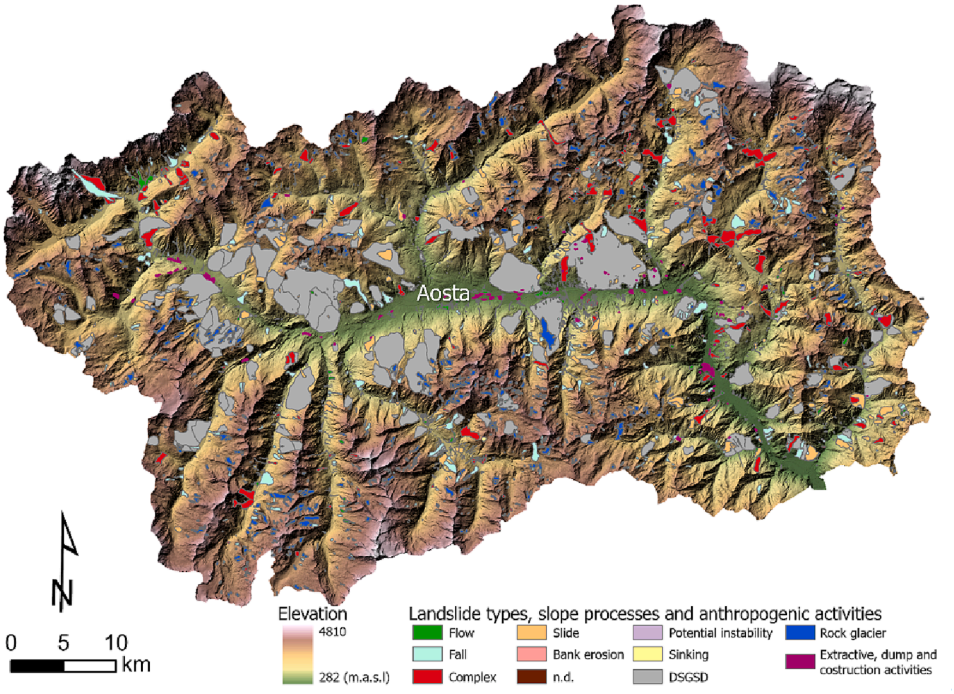


Fig. 2. Geomorphologic setting of the Valle d'Aosta region, with landslide types, slope processes and anthropogenic moving areas outlined according to the IFFI catalogue and to the regional databases (SCT, n.d.).

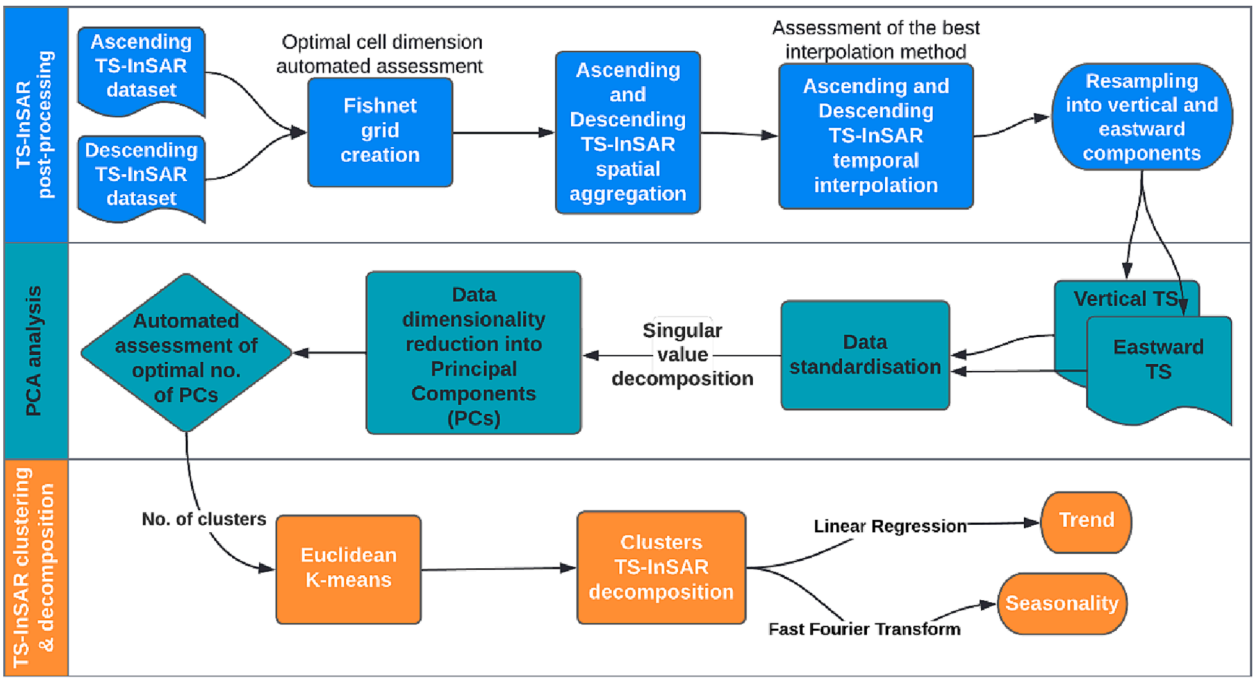


Fig. 3. Workflow of the proposed approach.

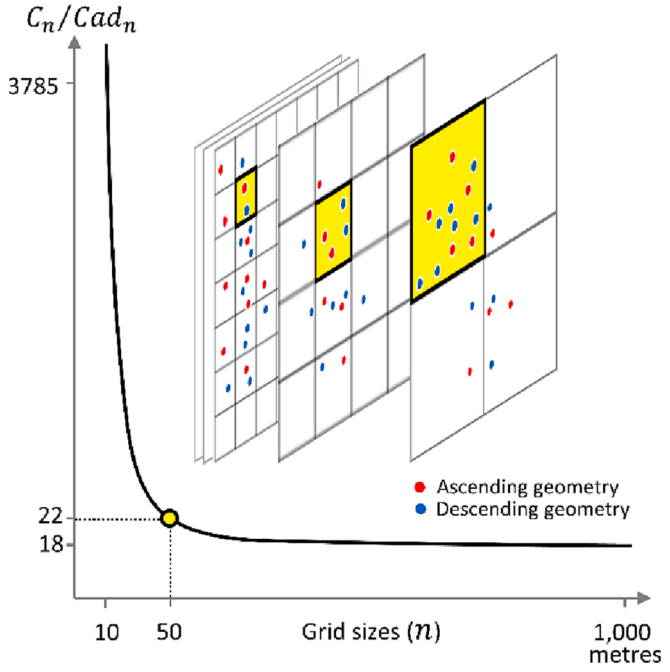


Fig. 4. Procedure for the identification of the optimal size of the grid for the spatial interpolation of MPs from the two LOS satellite geometries of acquisition.

vertical (V_V) and horizontal (V_H) (2) TS-InSAR on each time step by solving the following set of equations:

$$\begin{pmatrix} V_{asc} \\ V_{desc} \end{pmatrix} = \begin{pmatrix} -sin\theta_{asc}cos\alpha_{asc} & cos\theta_{asc} \\ -sin\theta_{desc}cos\alpha_{desc} & cos\theta_{desc} \end{pmatrix} \begin{pmatrix} V_H \\ V_V \end{pmatrix} \quad (2)$$

where θ and α are the satellite incidence and azimuth angle, respectively.

The vertical and horizontal components of motion are here retrieved assuming the N component of motion negligible.

The TS-InSAR temporal interpolation is performed by testing linear, quadratic and cubic methods. The objective is to find out the most accurate model able to fill data gaps in the new frequency, given that the aim is to preserve as much as possible the original non-stationary TS-InSAR data. To achieve that, we randomly select 100 time series from the dataset, eliminate every 2nd displacement measurement (i.e., the time steps with even index), interpolate the missing values, and evaluate the Root Mean Square Error (RMSE) (3) between the observed and forecasted values:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - o_i)^2} \quad (3)$$

where f is the forecasted value, o is the observed value and n is the total sample size. RMSE can be defined as the standard deviation of the residuals, therefore lower average values are associated to a better average model prediction. The lower mean RMSE value retrieved for 100 time series interpolated via linear, quadratic, and cubic methods is here considered as a proxy for choosing an individual fitting model.

4.2. Principal component analysis (PCA)

Given the large quantity of temporal displacement data embedded within TS-InSAR datasets, here we use PCA, a statistical technique which consists of data dimensionality reduction by balancing between information loss and optimal number of dimensions to retain (Jolliffe & Cadima, 2016; Prentice, 1982). In our case, we retrieve the Principal Components (PCs) variables accounting for the maximum amount of variance to characterise temporally and spatially changing deformation patterns from any TS-InSAR dataset without a priori constraints (Chaussard & Farr, 2019). Prior assuming that the sources are uncorrelated, the decomposition of a mixed deformation signal through the PCA results in a unique solution (Ebmeier, 2016).

The here performed PCA procedure (Pedregosa et al., 2011) is applied to multiple univariate time series (i.e., the TS-InSAR dataset) and can be divided into (i) data standardisation, (ii) data dimensionality reduction into PCs, (iii) automated assessment of the optimal number of PCs to retain. The aim of (i) is to standardise all the TS-InSAR to a comparable scale by removing the mean and scaling to unit variance, so

that each feature is equally considered and that PCA will not result in biased estimations. The second step (ii) involves linear dimensionality reduction using Singular Value Decomposition (Halko et al., 2011) of the data to project it to orthogonal components explaining a maximum amount of the variance. To achieve (iii), we adopt the Scree-plot criterion, where the eigenvalues are plotted against the number of components, and the optimal number of PCs is retrieved by looking for the "elbow" in the curve (section 4.1). This is done by automatically selecting the optimal number of PCs able to explain much of the variability of the TS-InSAR dataset.

4.3. TS-InSAR clustering and decomposition

The K-means algorithm is an inductive and unsupervised clustering method that is here adopted in the form of the standard Euclidean version to cluster unlabelled sparse TS-InSAR data into groups with similar characteristics (Pedregosa et al., 2011). Given two time series, $Q = q_1 \cdots q_n$ and $C = c_1 \cdots c_n$, their Euclidean distance D (4) is defined as:

$$D(Q, C) = \sqrt{\sum_{i=1}^{n} (q_i - c_i)^2} \tag{4}$$

K-means aims to separate n samples x in k clusters (c) of equal variances, each described by the mean μ_i (i.e., the centroid) of the TS samples in the cluster. The final clustering algorithm objective is to perform the minimisation of centroids inertia (5):

$$\sum_{i=0_{\mu_i \in c}}^{n} \min(||x_i - \mu_i||^2) \tag{5}$$

Inertia can be defined as a metric expressing how internally coherent clusters are, where lower values represent better model convergence.

K-means is a scalable approach towards very large samples and needs only a parameter to be initialised, namely the number of clusters, which is in our case automatically defined by the optimal number of PCs. Running PCA before k-means clustering has multiple advantages: it is the continuous solution of the cluster membership indicators in K-means clustering (Ding & He, 2004), it can reduce the problem of inflation of Euclidean spaces and boost the computations.

Every MP is grouped into a cluster of similar patterns. These patterns are retrieved by considering the barycentre (i.e., centroid) of the average shape of the grouped time series. To enhance the readability of the TS-InSAR clusters, the cluster centroid is further analysed to decompose the signal into trend and seasonality, where the first can be defined as the general direction of the overall data, while the seasonality is the repeating short-term cyclic component in the series. Finally:

- A least squares Linear Regression technique is applied to fit a linear model by minimising the residual sum of squares between the observed targets in the time series and the targets predicted by the linear approximation to retrieve the trend.
- The basic Fast Fourier Transform (FFT) is applied by computing the power spectrum (Priestley, 1981) to retrieve the seasonality. By looking at the frequencies corresponding to the peaks of spectral power with the highest magnitude, it is possible to assess the periodicity of a time-related series.

5. Results

5.1. Parameters calibration

Several fishnet configurations based on a range of selected grid size have been evaluated. The best grid size is automatically found at a value of 48 m, corresponding to a total of 2,480,665 cells subdividing a rectangular area of ca. 80 × 40 km enclosing the study area. Note that the grid size value results from the peculiar characteristics of the analysed TS-InSAR datasets (see section 4.1). On the other hand, *Cad*, namely the number of cells retaining both ascending and descending TS-InSAR, results in 93,385 cells, which constitutes the final number of features.

Due to the inherent temporal mismatch between the ascending and descending TS-InSAR datasets (256 and 261 scenes respectively), the proposed approach resamples the LOS measurements from the two different geometries into a new temporal frequency (*nf*), thus enabling LOS reprojection into vertical and horizontal components. The *nf*, computed as shown in section 4.1, is equal to 4 days considering the deployed test datasets, bringing the number of vertical and horizontal scenes to 504. Interpolation methods were tested by evaluating the prediction performance obtained with 100 random time series interpolated with linear, quadratic and cubic methods. Lower RMSE values are associated to a better average model prediction. In our case (Fig. 5), the linear interpolation method is preferred since it carries a mean RMSE value of 3.32, which is lower compared to the mean RMSE obtained by both quadratic and cubic interpolations (3.76 and 3.83 respectively). This result is indicative of the deployed data used within this work and it may differ based on different TS-InSAR datasets used as inputs.

The optimal number of PCs to retain is automatically selected through the Scree plot criterion, by plotting the eigenvalues on the y-axis and the number of factors on the x-axis. It always displays a downward curve, since the first component explains much of the variability, while the next components explain lower fractions of the overall variability. Finding the "elbow" in that curve relates to the number of factors retaining much of the TS-InSAR dataset information since they will constitute the cluster membership indicators in the K-means clustering. In our case, after performing the PCA analysis over the vertical and horizontal TS-InSAR datasets, the Kneedle algorithm selected 4 components to retain.

5.2. Automated TS-InSAR data mining

5.2.1. PCA-based K-means clustering

The standardised PCA scores for the reprocessed vertical and horizontal TS-InSAR dataset (Fig. 6) depict the underlying data structure. The two successive sets of the selected 4 orthogonal components explaining the maximum amount of the two datasets variance exhibit similar characteristics. The eigenvectors of the first PC (PC1) show that the direction of maximum variance (i.e., the eigenvector) is stretched parallel to the x-axis direction (representing the satellite acquisitions) showing a mostly stable and steady behaviour with displacement values slightly fluctuating around zero. PC1 (Table 2) demonstrates that most of the TS-InSAR signals covering the Valle d'Aosta region are stable (explained amount of variance stands at 76.55% and 74.14% for the horizontal and vertical dataset, respectively), which corresponds to an expected scenario when looking at a regional PSI dataset. PC2, PC3 and PC4 show a strong periodic fluctuation of the displacement signals which follows a clear yearly seasonal cycle. Regarding the horizontal dataset, PC2 accounts for 8.22% of the total variance (Table 2) and displays a seasonal deformation with a positive trend (i.e., the relative movement towards east). Concerning the vertical dataset, PC2 accounts for 7.91% of the total variance while displaying an overall negative trend, which translates into a downward displacement. PC3 and PC4 are indicative of displacement signals with similar amplitudes and similar trends oscillating around zero, however, they are out of phase over time in both analysed TS-InSAR datasets.

Accounting for the variance, the K-means clustering resulted in the classification of the analysed TS-InSAR dataset into 4 different types of clusters. The spatial distribution of the classified features, along with the representation of the centroid series (within the 10th and 90th percentiles) of each (i) horizontal and (ii) vertical TS-InSAR cluster and their relative percentage distributions are shown in Fig. 7.

Concerning (i) clusters, around 54% of the classified features belong to Cluster1h (stable westward; Fig. 7b). They show an overall uncorrelated behaviour, with random fluctuations of displacements around zero

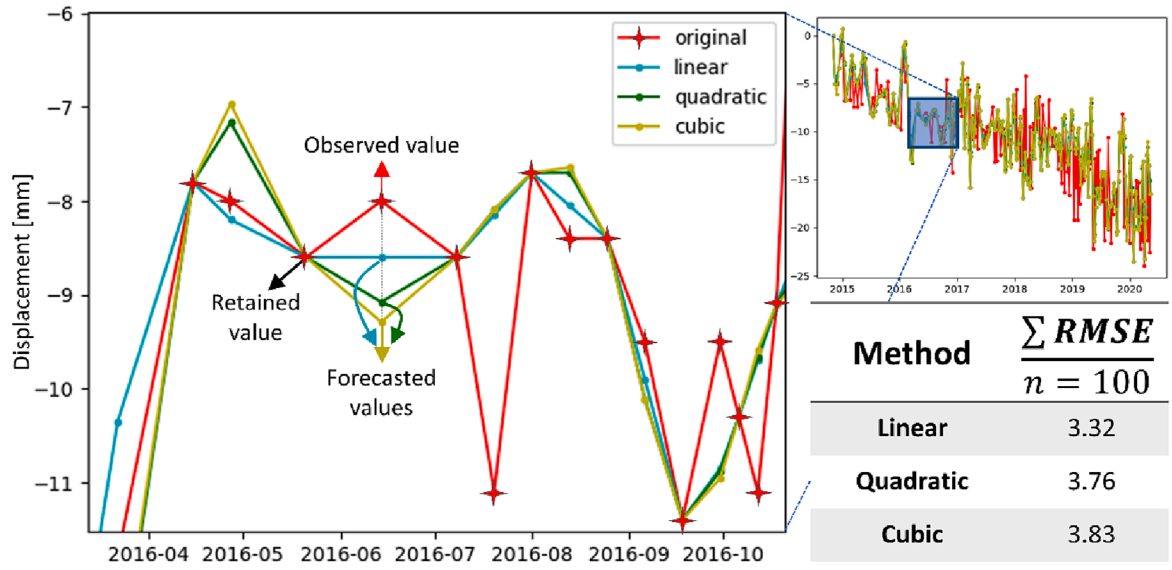


Fig. 5. Adopted procedure for the assessment of the interpolation method which best approximates the observed values, in the framework of TS-InSAR frequency up-sampling. The original time series (i.e., from the SqueeSAR dataset) are reprojected into new time series via linear, quadratic and cubic interpolation methods. We evaluate the mean RMSE of every method by interpolating 100 random time series from the ascending and descending datasets.

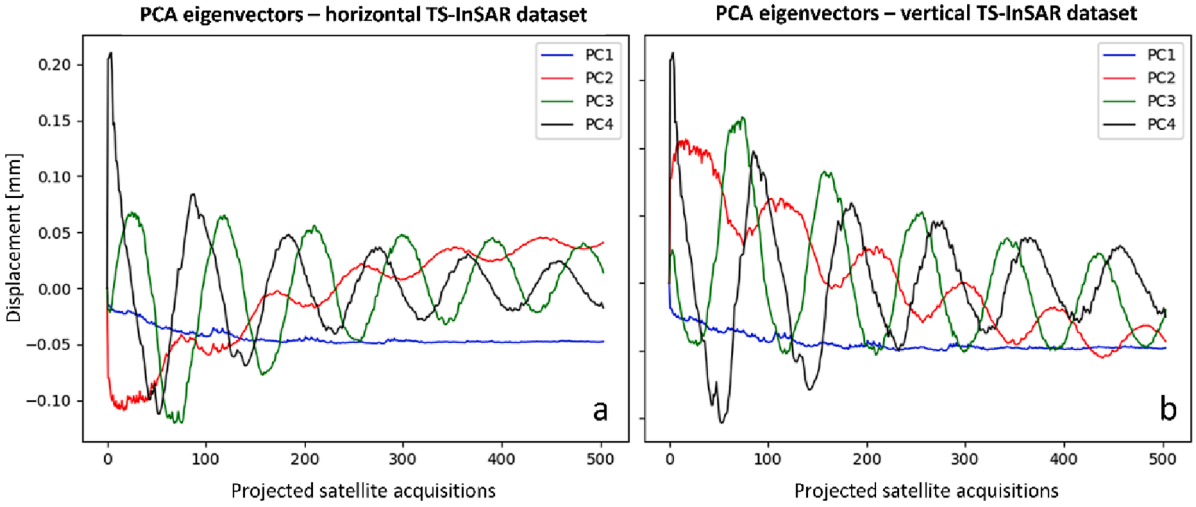


Fig. 6. Plots of the eigenvectors resulting from the standardised PCA conducted for the reprocessed a) horizontal and b) vertical TS-InSAR datasets.

Table 2
Eigenvalues scores resulting from the standardised PCA of the horizontal and vertical TS-InSAR datasets.

	Horizontal TS-InSAR dataset		Vertical TS-InSAR dataset	
PCs eigenvalues	Explained variance	% of total variance	Explained variance	% of total variance
PC1	385.07	76.55%	372.94	74.14%
PC2	41.35	8.22%	39.78	7.91%
PC3	12.59	2.50%	21.94	4.36%
PC4	3.74	0.74%	4.08	0.81%

indicating stable conditions and a barely noticeable average movement toward west. On the other hand, Cluster3h represents around 43% of features, denoting again a noisy series of displacement which however shows a general slightly positive trend, indicative of a relative eastward movement (stable eastward). Cluster2h and Cluster4h (only 1.8% and 1.6% of the overall features, respectively) can be assimilated if looking at the linear pattern and at the high rate of displacement (Fig. 7b), even though they show opposite motion directions: while Cluster2h exhibits a decisive negative trend (i.e., linear westward), Cluster4h exhibits a positive trend indicative of an eastward motion (i.e., linear eastward). The geographical distribution of both Cluster 1 h and Cluster3h appears uncorrelated, while Cluster2h and Cluster4h tends to be spatially aggregated on slopes with specific slope orientation, hence denoting a common process influencing the temporal behaviour of displacement of the ground, such as an active landslide.

By looking at (ii) clustering results, 50.5% and 39.8% of features are classified as Cluster1v (stable downward) and Cluster3v (stable upward), respectively (Fig. 7f). These two show several similarities with the previously exposed Cluster1h and Cluster3h, both in terms of series shape (Fig. 7e) and geographical distribution (Fig. 7d): they show mostly uncorrelated and noisy behaviour, with almost null displacement rates (within ± 5 mm), while they tend to be spatially scattered throughout the test area. On the other hand, Cluster2v (subsiding primary) and Cluster4v (subsiding secondary) are both less frequent within the classified features (0.8% and 8.9%, respectively): they exhibit a decisive negative trend (with Cluster2v holding higher rates of displacement)

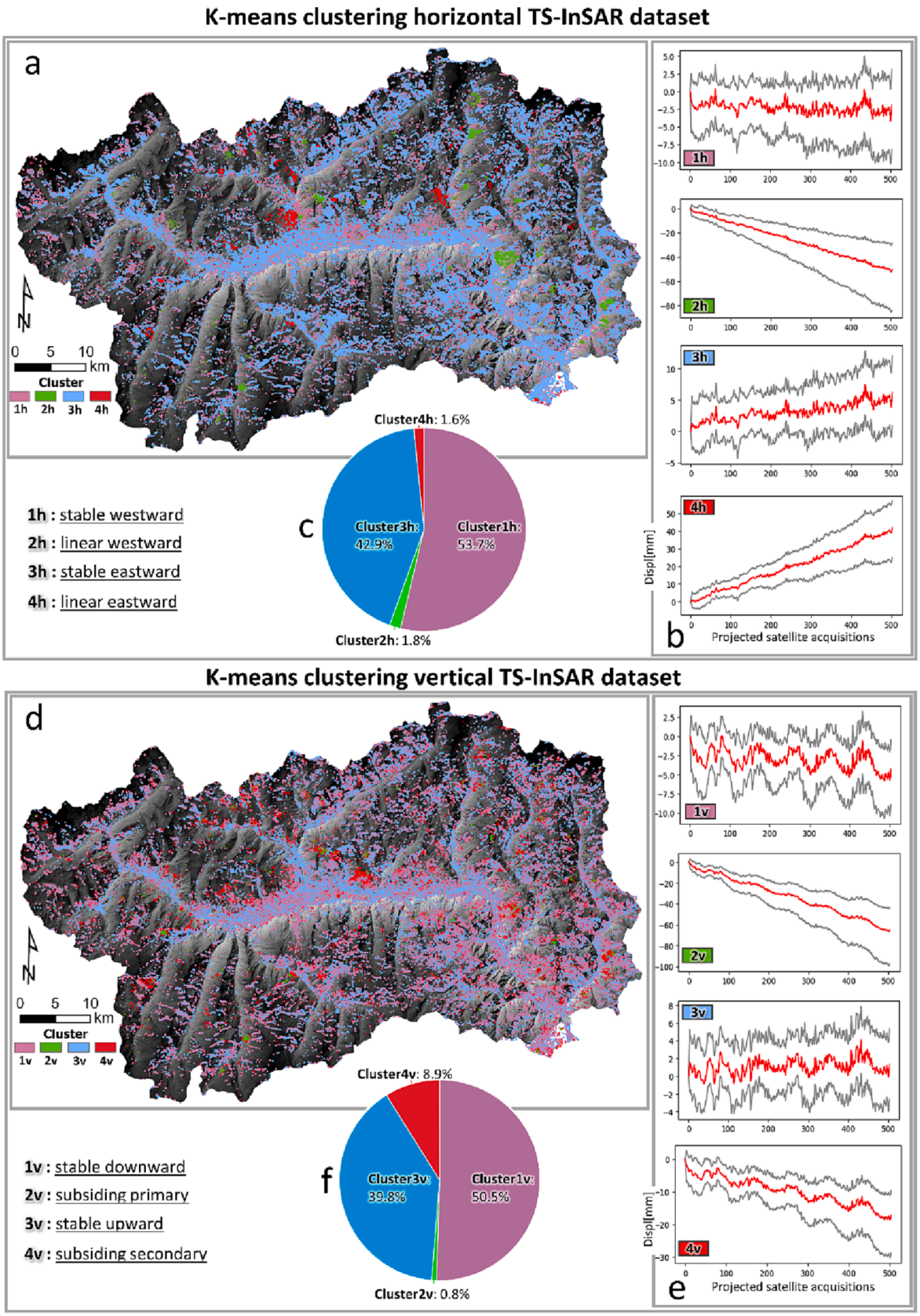


Fig. 7. Map of the geographical distribution of the a) horizontal and d) vertical features classified according to the K-mean clustering analysis, where the relative (b, e) clusters series are depicted by plotting the 10th, 90th percentile (grey series) and the centroid (red series). The cluster percentage distributions related to the horizontal and vertical TS-InSAR datasets are shown as (c, f) pie charts. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and tend to be mostly present on slopes. What is noticeable about the latter, is the repeating seasonal cycle markedly affecting the series.

Trend and seasonality component analysis of the retrieved cluster series are addressed more specifically in the next section (5.2.2).

5.2.2. *Cluster centroids series decomposition*

The last step of proposed approach concerns the decomposition of the clusters' series barycentre. Each centroid is fitted to a linear model to evaluate the underlying increasing or decreasing trend, while the potential seasonal component is obtained by converting the centroid into frequency domain. The outcomes of the twofold analysis are presented as follows: the visual plots of the horizontal and vertical centroid decomposition are depicted in Fig. 8 and Fig. 9 respectively, while the components are addressed quantitatively in Table 3. The spectral peaks obtained through the FFT are shown by plotting the x-axis covering the frequencies from zero to the Nyquist frequency (half the sampling rate), while the y-axis indicates the spectral power. Peaks corresponding to frequencies holding the largest power (amplitude) are shown in the tabular form, while with P_0 we indicate the peak with the lowest frequency (f_0) and with P_1 the peak carrying a greater frequency (f_1). The existence of an annual periodicity (i.e., seasonality) is eventually witnessed by the main peak approximating a frequency of $f \approx 365$ days^{-1}, while a peak associated to lower frequencies typically corresponds to the fundamental frequency (no predominant seasonality).

Among the horizontal TS-InSAR clusters, the centroid marked as 1 h (Fig. 8) is the one depicting an overall stable trend (with the linear segment with the best fit holding a slope of approximately zero). 3 h and 4 h (i.e., stable eastward and linear eastward, respectively) carry a similarly positive trend, however, 4 h has a higher slope (i.e., higher deformation rate). Instead, 2 h (linear westward) carries a negative linear trend but with a slope magnitude comprised between the latter. The periodicity analysis highlighted that only 1 h and 3 h have a small trace of seasonality within their signals, since they both resulted in having P_1 corresponding to a frequency of 336 days^{-1}.

Regarding the vertical TS-InSAR clusters, centroids 1v and 3v depict trend with very low deformation rates (Fig. 9), the first being slightly negative (stable downward) and the second being slightly positive (stable upward; see slope values in Table 3). 2v holds a decisive negative linear trend (subsiding primary), with slope value of two orders greater than 1v and 3v. Centroid 4v (subsiding secondary) has still a negative trend but less pronounced than 2v. The periodicity analysis of the vertical clusters' centroids allows the retrieval of interesting results: 1v and 3v centroids have their main spectral peaks set to an annual cycle (P_0 corresponding to a frequency of 336 days^{-1}). 2v shows no seasonality at all (both P_0 and P_1 relate to lower frequencies), instead, centroid 4v holds a trace of seasonality having the peak P_1 set to an annual frequency (Table 3).

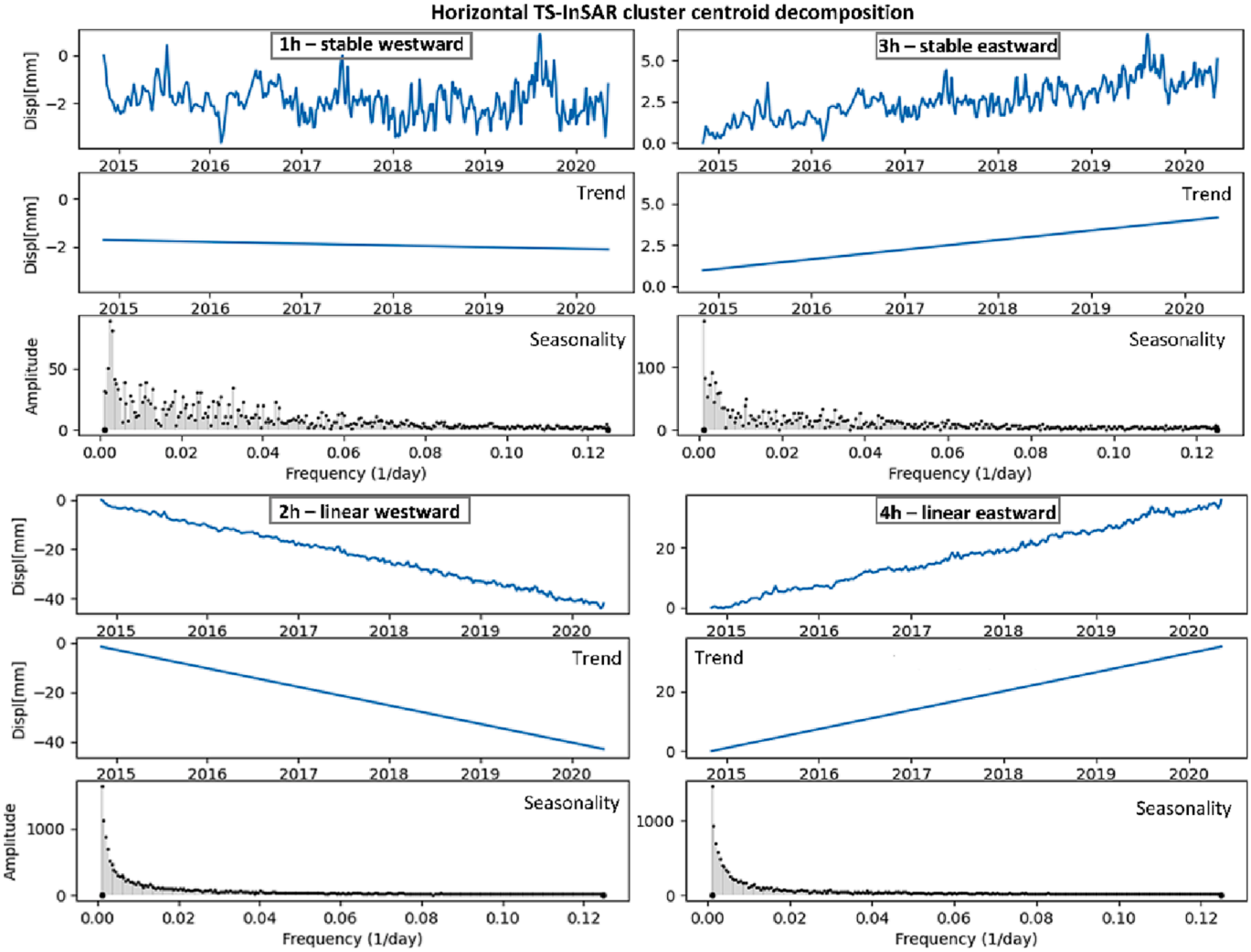


Fig. 8. Horizontal TS-InSAR cluster centroid decomposition into a trend (Linear Regression) and seasonality (power spectrum) components.

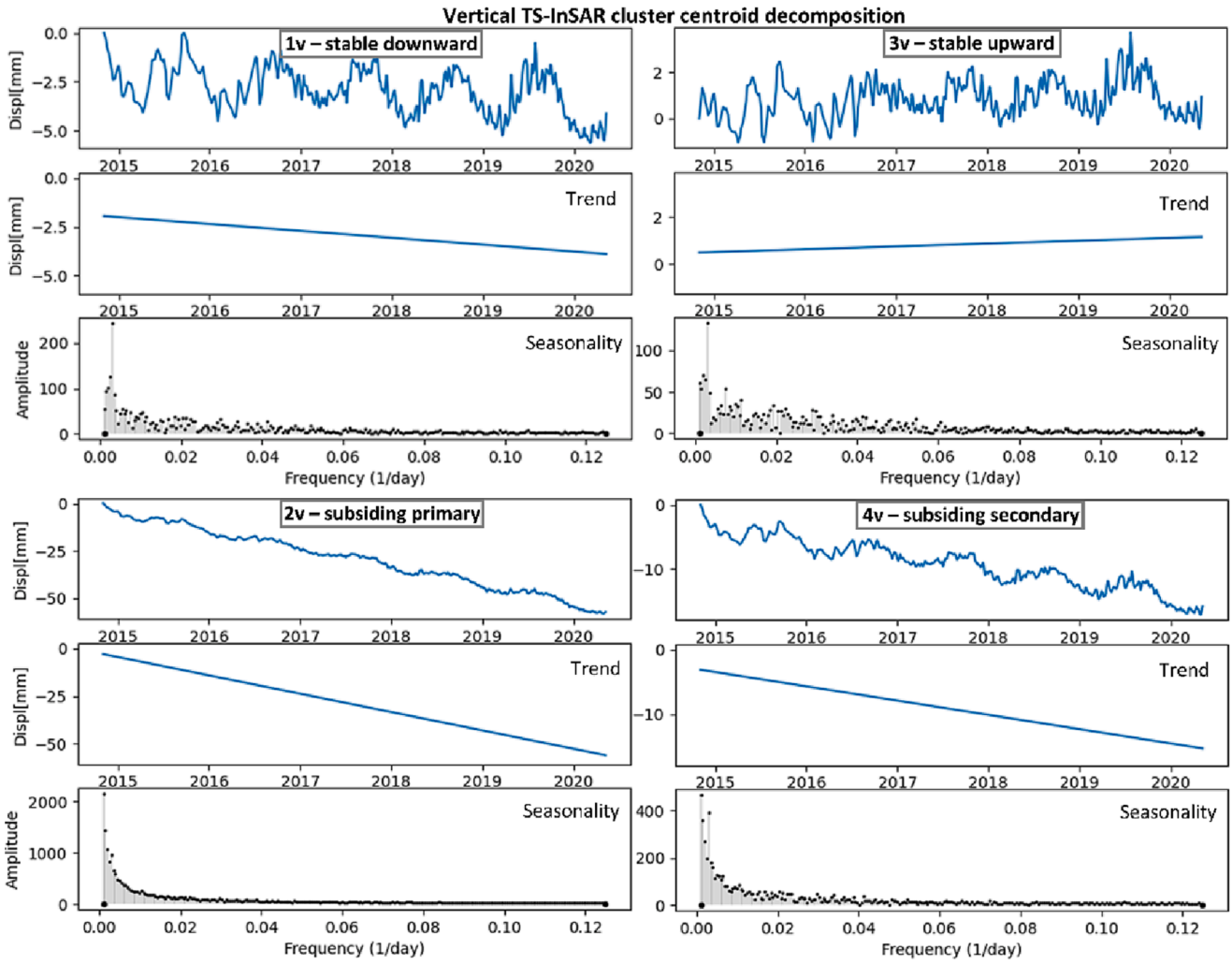


Fig. 9. Vertical TS-InSAR cluster centroid decomposition into trend (Linear Regression) and seasonality (power spectrum) components.

Table 3
Clusters centroids components obtained from the Ordinary least squares Linear Regression and Discrete Fourier Transform analysis. RMSE indicates the Root Mean Square Error estimated for the linear model. P_0 and P_1 denote the two main spectral peaks encountered, with P_0 being the peak with the highest spectral power.

	Linear Regression			Discrete Fourier Transform			
Cluster centroids	**Slope**	**Intercept**	**RMSE**	**Spectral peaks**	**Amplitude**	**Frequency**	**Period (days)**
1h	−0.08e-2	−1.72	0.67	$\mathbf{P_0}$	88.92	0.24e-2	403
				$\mathbf{P_1}$	81.18	0.30e-2	336
2h	−0.80e-1	−1.52	0.61	$\mathbf{P_0}$	1637.68	0.99e-3	1008
				$\mathbf{P_1}$	1126.94	0.15e-2	672
3h	0.64e-2	0.97	0.64	$\mathbf{P_0}$	173.15	0.99e-3	1008
				$\mathbf{P_1}$	91.13	0.30e-2	336
4h	0.69e-1	−0.8e-1	0.84	$\mathbf{P_0}$	1458.41	0.99e-3	1008
				$\mathbf{P_1}$	940.21	0.15e-2	672
1v	−0.39e-2	−1.94	1.03	$\mathbf{P_0}$	242.59	0.30e-2	336
				$\mathbf{P_1}$	126.21	0.24e-2	403
2v	−0.11	−2.81	1.61	$\mathbf{P_0}$	2143.65	0.99e-3	1008
				$\mathbf{P_1}$	1443.22	0.15e-2	672
3v	0.13e-2	0.49	0.76	$\mathbf{P_0}$	132.62	0.30e-2	336
				$\mathbf{P_1}$	69.78	0.20e-2	504
4v	−0.02	−3.08	1.18	$\mathbf{P_0}$	465.70	0.99e-3	1008
				$\mathbf{P_1}$	391.62	0.30e-2	336

5.3. Comparison of clustering results with regional inventories

The spatial distribution of the retrieved clusters is compared to the regional inventories listed in Section 3.2. It should be highlighted that the presented unsupervised approach intends to disclose hidden patterns within InSAR data, while instability-prone areas cannot be regarded as reliable ground truth labels, given the unknown state of activity. Cluster2h and Cluster4h (i.e., linear westward and linear eastward) features tends to aggregate within mapped landslides (38.2% and 24.6% of total features, respectively) and within mapped DSGSDs (55% and 70.2%,

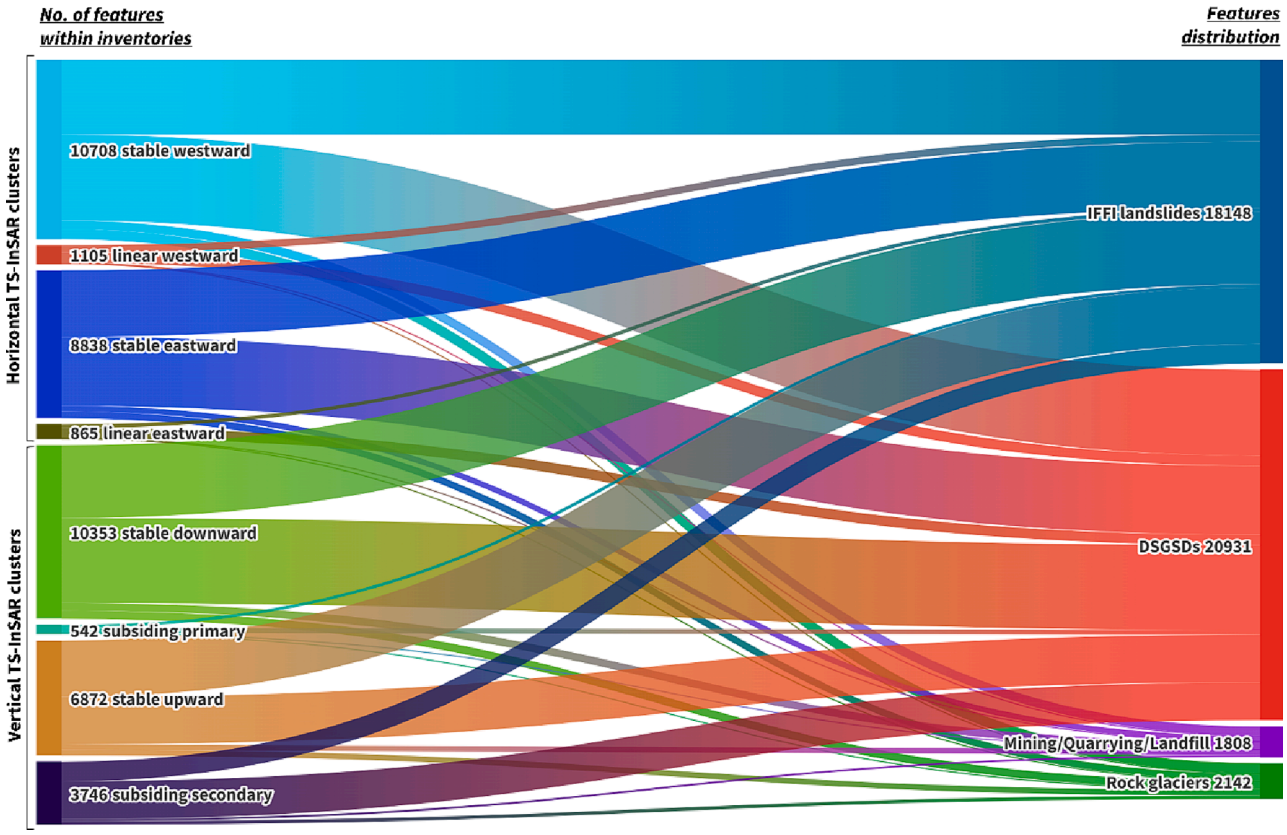


Fig. 10. Sankey diagram for the spatial comparison between the classified features of the horizontal and vertical TS-InSAR datasets and the pre-existing regional inventories relatable to possibly active source of ground motion.

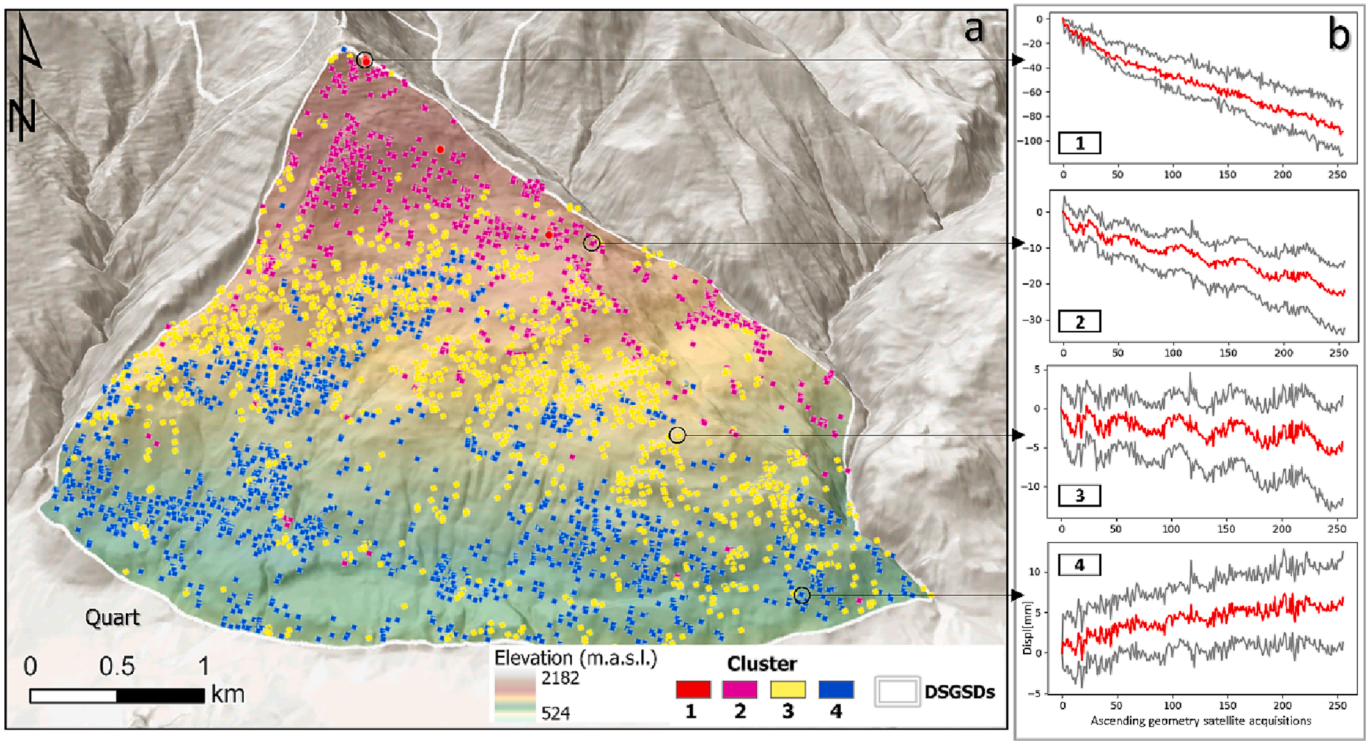


Fig. 11. Different temporal displacement behaviours are noticeable within a DSGSD phenomenon according to different types of (a) ascending geometry TS-InSAR clusters. The (b) cluster plots and the spatial distribution of the relative features allow an insightful understanding of the kinematic behaviour of the DSGSD.

respectively; Fig. 10). Comparable numbers are obtained for the same inventories when considering Cluster2v and Cluster4v (i.e., subsiding primary and subsiding secondary, respectively): they show the highest rates of displacement between the retrieved vertical TS-InSAR clusters. In general, it can be stated that both vertical and horizontal motions can be equally recognised within DSGSDs.

The mining/quarrying extraction sites alongside the rock glacier inventory have generally low grade of spatial intersection with the classified features. This can be linked to their relatively low areal extension, if compared to landslides and DSGSDs, which are higher in number (2,329 landslides and DSGSDs vs 86 extraction sites and 941 rock glaciers) and cover a greater areal extension throughout Valle d'Aosta (591 km^2 vs 71 km^2). For the same reason, the density of features is low for landslides and DSGSDs and higher for mining and quarrying sites. However, it is evident a relative higher intersection of Cluster2v with the rock glacier inventory: this can be regarded as evidence of the active state of movement of some of this kind of phenomena within Valle d'Aosta.

The presented approach can detect and characterise different motion patterns when looking at local unstable areas. As proof of the above, in Fig. 11 we report the clustering results applied to the ascending TS-InSAR dataset covering the extent of a DSGSD near the town of Quart. The spatial segregation of different motion patterns is coherent with respect to the downhill slope direction and the main morphological units of the observed phenomenon. Indeed, 4 main different cluster groups are indicative of a variation of displacement rate and temporal behaviour which become evident across the DSGSD longitudinal axis. TS-InSAR features residing in the scarp area prove to have the highest rates of displacement and a mostly linear trend (Fig. 11). The detected most active zones fade into features punctuating the main body of DSGSD (Fig. 11b), indicative of a less severe downward displacement trend. Lastly, the foot of the DSGSD is interested by a progressive shift from negative to positive (i.e., upward) displacements (Fig. 11). This can be related to the peculiar geometry of the failure surface. Within this framework, it should be mentioned that several studies conducted over Valle d'Aosta and the Alpine chain (Paranunzio et al., 2019) report that the frequency of not-ordinary climatic conditions, which follow a seasonal trend, is majorly related to slope failures events.

6. Discussion

The proposed unsupervised data mining approach can be regarded as a key contribution to the optimal management of big InSAR data, for the benefit of expert and non-expert users, for public and private scenarios. As a result, our data-driven analysis can greatly improve the interpretability and dissemination of InSAR-based information.

The strength of the approach is evidenced in Fig. 12, depicting unstable conditions within a landfill site, in the municipality of Aosta. The deformation map containing both ascending and descending TS-InSAR S-1 MPs (Fig. 12a) sets the starting point of our analysis, while the clustered features retrieved for the vertical component of motion (Fig. 12b) represent one part of a twofold outcome of the data mining approach. The procedure ensures an unprecedented improvement and readability of large amounts of radar data: 1) the analysis of the earth surface motion can be evaluated according to the vertical and horizontal directions, allowing for an enhanced interpretability of ground displacements; 2) the large quantity of targets is compressed into few data-driven clusters based on the similar temporal displacement behaviour, hence enabling to rapidly define common deformation processes; 3) the detected process is further analysed in terms of the main temporal trend of deformation and the eventual seasonality component. The temporal characterisation of clustered InSAR MPs represents a step forward respect to previous methods concerning the detection of unstable areas (Festa et al., 2022; Tomás et al., 2019). The analysed features demonstrate to be spatially clustered in accordance with the varying morphometry of the territory and in coherence with the described surface motion phenomenon.

A well-known limitation of conventional InSAR techniques is the low sensitivity to North-South (N-S) deformations. Since modern spaceborne radar systems move along near-polar orbits, there is no diversity in the viewing geometries to accurately estimate N-S displacements. Therefore, N-S displacement time series are here neglected. Aside from the intrinsic limitations of InSAR, some other constraints may arise from the proposed approach. For instance, the MPs aggregation into grid cells might cause a significant drop in terms of spatial resolution (Fig. 12b) hindering the correct detection of very small-scale surface movements. Also, the approach does not consider any weight related to the quantity of MPs from the two geometries and contained within the cell. Moreover, it should be considered that the absence of regular ascending and descending acquisitions constitutes a major obstacle for the applicability of the proposed approach. Anyway, whether only a single geometry of acquisition is available, our approach can be thought as still valid to reduce a TS-InSAR dataset into its PCs, to evaluate the temporal clusters and the related temporal components, as demonstrated in Fig. 11.

We point out that the series averaging within each grid cell and the rationale of temporal interpolation can possibly make an impact in correctly reproducing the motion trend along the vertical and horizontal components. Since many sources of errors due to radar acquisitions are responsible for the generally high noise component in the TS-InSAR signals, a low resampling rate (downsampling) might result in a general positive effect of smoothing errors from the series data. On the other hand, an excessive loss of data might cause an oversimplification that can be detrimental to the accuracy of the analysis. In this regard, the TS-InSAR dataset quality also plays a role: discontinuity and unbalance in radar data acquisitions might bring to a misleading reconstruction of the TS-InSAR components of motion, regardless of the adopted interpolation method (see section 5.1). This, rather than affecting the clustering performances, might affect the trend and seasonality components estimation. In view of the above-mentioned, S-1 archives coupled with a balanced resampling rate (see section 4.1) have shown to be appropriate for time-dependent series data mining, given the temporal consistency in data collection.

The restriction of the trend and seasonality analysis to the cluster centroid only, despite partially defeaturing the cluster series characterisation, requires a much lower computational effort while still constituting a valuable operation for the readability of the data-driven TS-InSAR clusters. Although the here presented approach is thought to extract, enrich, and summarise meaningful information from large scale InSAR data in an optimised way, more detailed results can be obtained by either using a reduced InSAR dataset as input (for small-scale analysis) or by deliberately increasing the number of Principal Components (i.e., number of clusters) to retain. In this framework, the procedure can also be useful for the analysis of TS-InSAR datasets covering the extent of a single basin, allowing for an extremely detailed evaluation of hardly noticeable motion trends occurring at a local scale.

The presented method can be further developed by integrating Independent Component Analysis (ICA) to retrieve the additive subcomponents of PCA-based results (i.e., to characterize the temporal behaviours of vertical and horizontal TS-InSAR clusters).

7. Conclusions

The presented unsupervised data mining for the automatic recognition of complex surface motion trends makes it possible to optimally interpret large-scale displacement TS-InSAR datasets, providing data-driven results for territorial planning operations and civil protection purposes.

The core of the adopted algorithm relies on a statistical data dimensionality reduction technique (i.e., PCA) prior to K-Means for the automated spatial clustering of TS-InSAR data based on similar temporal displacement behavior of the analysed radar targets. The latter are fed to the algorithm prior a reprojection of the LOS displacement series along

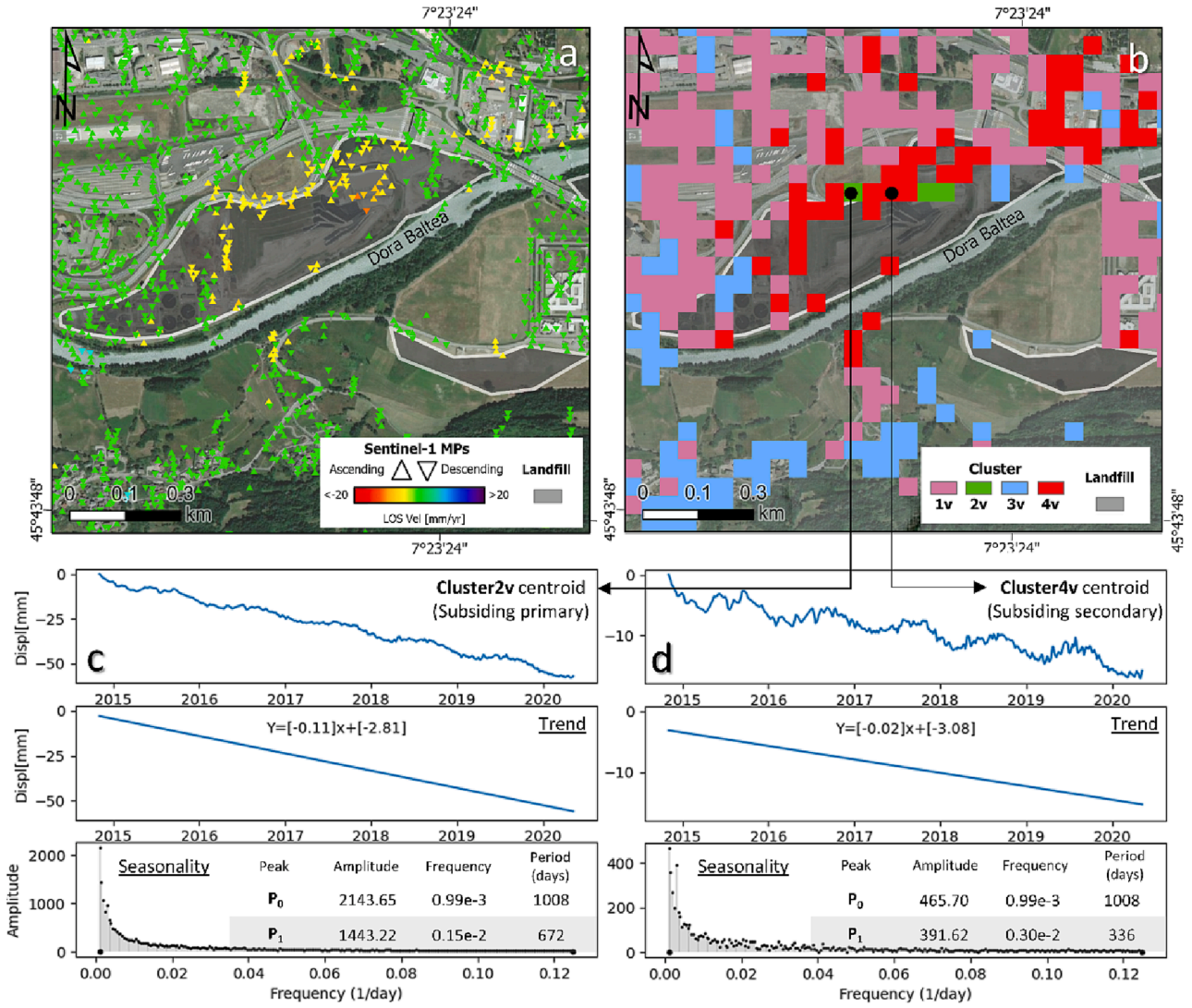


Fig. 12. (a) Mean LOS velocity of ascending and descending MPs on a landfill site located in the Aosta municipality. (b) Clustering results of the reprocessed vertical TS-InSAR dataset where the clustered features point out an ongoing subsiding process with a mostly linear trend and high rate of displacement (c) and with a more pronounced seasonal pattern while having a lower displacement rate (d).

the vertical and horizontal directions, where the applied data fusion rationale is robust about the spatial and temporal dimensions. Ultimately, the trend and seasonality components are retrieved to enhance the readability of the classified features.

The application of the approach to the regional territory of Valle d'Aosta offered a valuable example of automation strategy for the provision of support to risk-assessment against geohazards, both natural and anthropogenic. To promote the use of the approach in the framework of the increasing availability of space borne radar data (e.g., the European Ground Motion Service; EGMS, 2022), the adopted code is freely available on Github at https://github.com/maybedave/InSAR-Time-Series-Clustering.

In the near future, the current approach will be further developed through data fusion and deep learning-based solutions to feed and develop forecasting models to be used in decision-making strategies and early warning systems.

CRediT authorship contribution statement

Davide Festa: Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft, Visualization, Validation. **Alessandro Novellino:** Methodology, Investigation, Writing – review & editing. **Ekbal Hussain:** Methodology, Writing – review & editing. **Luke Bateson:** Writing – review & editing, Supervision, Project administration. **Nicola Casagli:** Supervision, Project administration. **Pierluigi Confuorto:** Validation, Writing – review & editing. **Matteo Del Soldato:** Writing – review & editing. **Federico Raspini:** Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the link to my code and sample data within the manuscript (Methods and conclusion).

Acknowledgements

The research was partly funded through the ALMEO project of the BGS International NC programme 'Geoscience to tackle Global

Environmental Challenges', NERC reference NE/X006255/1 and the grant 'Championing INnovation In The DIgital ENvironment', NERC reference NE/S016287/1.

References

Ansari, H., Ruβwurm, M., Ali, M., Montazeri, S., Parizzi, A., Zhu, X.X., 2021. InSAR Displacement Time Series Mining: A Machine Learning Approach. IEEE International Geoscience and Remote Sensing Symposium IGARSS 2021, 3301–3304. https://doi.org/10.1109/IGARSS47720.2021.9553465.

Arnaud, A., Adam, N., Hanssen, R., Inglada, J., Duro, J., Closa, J., & Eineder, M. (2003). ASAR ERS interferometric phase continuity. IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477), 2, 1133-1135 vol.2. Doi: https://doi.org/10.1109/IGARSS.2003.1294035.

Bateson, L., Cigna, F., Boon, D., Sowter, A., 2015. The application of the Intermittent SBAS (ISBAS) InSAR method to the South Wales Coalfield, UK. Int. J. Appl. Earth Obs. Geoinf. 34, 249–257. https://doi.org/10.1016/j.jag.2014.08.018.

Berardino, P., Fornaro, G., Lanari, R., Sansosti, E., 2002. A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. IEEE Trans. Geosci. Remote Sens. 40 (11), 2375–2383. https://doi.org/10.1109/TGRS.2002.803792.

Berti, M., Corsini, A., Franceschini, S., Iannacone, J.P., 2013. Automated classification of Persistent Scatterers Interferometry time series. Nat. Hazards Earth Syst. Sci. 13 (8), 1945–1958. https://doi.org/10.5194/nhess-13-1945-2013.

Blanco-Sánchez, P., Mallorquí, J.J., Duque, S., Monells, D., 2008. The Coherent Pixels Technique (CPT): An Advanced DInSAR Technique for Nonlinear Deformation Monitoring. Pure Appl. Geophys. 165 (6), 1167–1193. https://doi.org/10.1007/s00024-008-0352-6.

Carraro, F., Giardino, M., 2004. Quaternary glaciations in the western Italian Alps - a review. In: Ehlers, J., Gibbard, P.L. (Eds.), Quaternary Glaciations Extent and Chronology, Vol. 2. Elsevier, pp. 201–208. https://doi.org/10.1016/S1571-0866(04)80071-X.

Chang, L., Hanssen, R.F., 2016. A Probabilistic Approach for InSAR Time-Series Postprocessing. IEEE Trans. Geosci. Remote Sens. 54 (1), 421–430. https://doi.org/10.1109/TGRS.2015.2459037.

Chaussard, E., Farr, T.G., 2019. A New Method for Isolating Elastic From Inelastic Deformation in Aquifer Systems: Application to the San Joaquin Valley, CA. Geophys. Res. Lett. 46 (19), 10800–10809. https://doi.org/10.1029/2019GL084418.

Ciampalini, A., Solari, L., Giannecchini, R., Galanti, Y., Moretti, S., 2019. Evaluation of subsidence induced by long-lasting buildings load using InSAR technique and geotechnical data: The case study of a Freight Terminal (Tuscany, Italy). Int. J. Appl. Earth Obs. Geoinf. 82, 101925 https://doi.org/10.1016/j.jag.2019.101925.

Cigna, F., del Ventisette, C., Liguori, V., Casagli, N., 2011. Advanced radar-interpretation of InSAR time series for mapping and characterization of geological processes. Nat. Hazards Earth Syst. Sci. 11 (3), 865–881. https://doi.org/10.5194/nhess-11-865-2011.

Confuorto, P., del Soldato, M., Solari, L., Festa, D., Bianchini, S., Raspini, F., Casagli, N., 2021. Sentinel-1-based monitoring services at regional scale in Italy: State of the art and main findings. Int. J. Appl. Earth Obs. Geoinf. 102, 102448 https://doi.org/10.1016/j.jag.2021.102448.

Crosetto, M., Solari, L., Balasis-Levinsen, J., Casagli, N., Frei, M., Oyen, A., Moldestad, D. A., 2020. Ground deformation monitoring at continental scale: the european ground motion service. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-293-2020.

Dal Piaz, G., Cortiana, G., del Moro, A., Martin, S., Pennacchioni, G., Tartarotti, P., 2001. Tertiary age and paleostructural inferences of the eclogitic imprint in the Austroalpine outliers and Zermatt-Saas ophiolite, western Alps. Int. J. Earth Sci. 90 (3), 668–684. https://doi.org/10.1007/s005310000177.

de Luca, C., Valerio, E., Giudicepietro, F., Macedonio, G., Casu, F., Lanari, R., 2022. Pre- and Co-Eruptive Analysis of the September 2021 Eruption at Cumbre Vieja Volcano (La Palma, Canary Islands) Through DInSAR Measurements and Analytical Modeling. Geophys. Res. Lett. 49 (7) https://doi.org/10.1029/2021GL097293.

de Zan, F., Monti Guarnieri, A., 2006. TOPSAR: Terrain Observation by Progressive Scans. IEEE Trans. Geosci. Remote Sens. 44 (9), 2352–2360. https://doi.org/10.1109/TGRS.2006.873853.

Ding, C., He, X., 2004. K-means clustering via principal component analysis. Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004, 225–232. https://www.scopus.com/inward/record.uri?eid=2-s2.0-14344257496&partnerID=40&md5=09310cc192cf973a0d4a7c2fcbc9e39f.

Ebmeier, S.K., 2016. Application of independent component analysis to multitemporal InSAR data with volcanic case studies. J. Geophys. Res. Solid Earth 121 (12), 8970–8986. https://doi.org/10.1002/2016JB013765.

European Ground Motion Service (EGMS). (2022). https://land.copernicus.eu/pan-european/european-ground-motion-service. (Accessed 18 January 2023).

European Space Agency (ESA). (2022, August 3). https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Mission_ends_for_Copernicus_Sentinel-1B_satellite. (Accessed 18 January 2023).

Ezquerro, P., del Soldato, M., Solari, L., Tomás, R., Raspini, F., Ceccatelli, M., Fernández-Merodo, J.A., Casagli, N., Herrera, G., 2020. Vulnerability Assessment of Buildings due to Land Subsidence Using InSAR Data in the Ancient Historical City of Pistoia (Italy). Sensors 20 (10). https://doi.org/10.3390/s20102749.

Ferretti, A., Prati, C., & Rocca, F. (1999). Permanent scatterers in SAR interferometry. IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293), 3, 1528-1530 vol.3. Doi: https://doi.org/10.1109/IGARSS.1999.772008.

Ferretti, A., Fumagalli, A., Novali, F., Prati, C., Rocca, F., Rucci, A., 2011. A New Algorithm for Processing Interferometric Data-Stacks: SqueeSAR. IEEE Trans. Geosci. Remote Sens. 49 (9), 3460–3470. https://doi.org/10.1109/TGRS.2011.2124465.

Festa, D., Bonano, M., Casagli, N., Confuorto, P., de Luca, C., del Soldato, M., Lanari, R., Lu, P., Manunta, M., Manzo, M., Onorato, G., Raspini, F., Zinno, I., Casu, F., 2022. Nation-wide mapping and classification of ground deformation phenomena through the spatial clustering of P-SBAS InSAR measurements: Italy case study. ISPRS J. Photogramm. Remote Sens. 189, 1–22. https://doi.org/10.1016/j.isprsjprs.2022.04.022.

Halko, N., Martinsson, P.G., Tropp, J.A., 2011. Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. SIAM Rev. 53 (2), 217–288. https://doi.org/10.1137/090771806.

Intrieri, E., Raspini, F., Fumagalli, A., Lu, P., del Conte, S., Farina, P., Allievi, J., Ferretti, A., Casagli, N., 2018. The Maoxian landslide as seen from space: detecting precursors of failure with Sentinel-1 data. Landslides 15 (1), 123–133. https://doi.org/10.1007/s10346-017-0915-7.

Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 374 (2065), 20150202. https://doi.org/10.1098/rsta.2015.0202.

Lu, P., Casagli, N., Catani, F., Tofani, V., 2012. Persistent Scatterers Interferometry Hotspot and Cluster Analysis (PSI-HCA) for detection of extremely slow-moving landslides. Int. J. Remote Sens. 33 (2), 466–489. https://doi.org/10.1080/01431161.2010.536185.

Novellino, A., Terrington, R., Christodoulou, V., Smith, H., & Bateson, L. (2019). Ground motion and stratum thickness comparison in Tower Hamlets, London.

Paranunzio, R., Chiarle, M., Laio, F., Nigrelli, G., Turconi, L., Luino, F., 2019. New insights in the relation between climate and slope failures at high-elevation sites. Theor. Appl. Climatol. 137 (3), 1765–1784. https://doi.org/10.1007/s00704-018-2673-4.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-Learn: Machine Learning in Python. J. Mach. Learn. Res. 12 (null), 2825–2830.

Prentice, I.C., 1982. Multivariate methods for the presentation and analysis of data. Palaeohydrological changes in the temperate zone in the last 15 000 years. IGCP 158 B. Lake and mire environments. Project guide III.

Priestley, M.B., 1981. Spectral Analysis and Time Series. Academic Press Inc.

Raspini, F., Bianchini, S., Ciampalini, A., del Soldato, M., Solari, L., Novali, F., del Conte, S., Rucci, A., Ferretti, A., Casagli, N., 2018. Continuous, semi-automatic monitoring of ground deformation using Sentinel-1 satellites. Sci. Rep. 8 (1), 7253. https://doi.org/10.1038/s41598-018-25369-w.

Satopaa, V., Albrecht, J., Irwin, D., Raghavan, B., 2011. Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior. In: 2011 31st International Conference on Distributed Computing Systems Workshops, pp. 166–171. https://doi.org/10.1109/ICDCSW.2011.20.

Schlögl, M., Widhalm, B., Avian, M., 2021. Comprehensive time-series analysis of bridge deformation using differential satellite radar interferometry based on Sentinel-1. ISPRS J. Photogramm. Remote Sens. 172, 132–146. https://doi.org/10.1016/j.isprsjprs.2020.12.001.

Sistema delle Conoscenze Territoriali (SCT). (n.d.). Geoportale. https://geoportale.regione.vda.it/. (Accessed 18 January 2023).

Tomás, R., Pagán, J.I., Navarro, J.A., Cano, M., Pastor, J.L., Riquelme, A., Cuevas-González, M., Crosetto, M., Barra, A., Monserrat, O., Lopez-Sanchez, J.M., Ramón, A., Ivorra, S., del Soldato, M., Solari, L., Bianchini, S., Raspini, F., Novali, F., Ferretti, A., Casagli, N., 2019. Semi-Automatic Identification and Pre-Screening of Geological-Geotechnical Deformational Processes Using Persistent Scatterer Interferometry Datasets. Remote Sens. (Basel) 11 (14). https://doi.org/10.3390/rs11141675.

Trigila, A., Iadanza, C., Guerrieri, L. (2007). The IFFI project (Italian landslide inventory): Methodology and results. In: Guidelines for Mapping Areas at Risk of Landslides in Europe, edited by: Hervás, J., ISPRA, Rome, Italy.

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.-H., Steinbach, M., Hand, D.J., Steinberg, D., 2008. Top 10 algorithms in data mining. Knowl. Inf. Syst. 14 (1), 1–37. https://doi.org/10.1007/s10115-007-0114-2.

Xiong, S., Wang, C., Qin, X., Zhang, B., Li, Q., 2021. Time-Series Analysis on Persistent Scatter-Interferometric Synthetic Aperture Radar (PS-InSAR) Derived Displacements of the Hong Kong-Zhuhai-Macao Bridge (HZMB) from Sentinel-1A Observations. Remote Sens. (Basel) 13 (4). https://doi.org/10.3390/rs13040546.

Zhang, Y., Meng, X.M., Dijkstra, T.A., Jordan, C.J., Chen, G., Zeng, R.Q., Novellino, A., 2020. Forecasting the magnitude of potential landslides based on InSAR techniques. Remote Sens. Environ. 241, 111738 https://doi.org/10.1016/j.rse.2020.111738.