Automated classification of A-DInSAR-based ground deformation by using random forest

Davide Festa, Nicola Casagli, Francesco Casu, Pierluigi Confuorto, Claudio De Luca, Matteo Del Soldato, Riccardo Lanari, Michele Manunta, Mariarosaria Manzo & Federico Raspini

Published online: 14 Oct 2022.

Submit your article to this journal

Article views: 375

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

🔓 OPEN ACCESS | Check for updates

Automated classification of A-DInSAR-based ground deformation by using random forest

Davide Festa[a], Nicola Casagli[a,b], Francesco Casu[c], Pierluigi Confuorto[a], Claudio De Luca[c], Matteo Del Soldato[a], Riccardo Lanari[c], Michele Manunta[c], Mariarosaria Manzo[c] and Federico Raspini[a]

[a]Department of Earth Sciences, University of Firenze, Firenze, Italy; [b]National Institute of Oceanography and Applied Geophysics - OGS, Sgonico (Trieste), Italy; [c]IREA-CNR, Napoli, Italy

ABSTRACT

Wide-area ground motion monitoring is nowadays achievable via advanced Differential Interferometry SAR (A-DInSAR) techniques which benefit from the availability of large sets of Copernicus Sentinel-1 images. However, it is of primary importance to implement automated solutions aimed at performing integrated analysis of large amounts of interferometric data. To effectively detect high-displacement areas and classify ground motion sources, here we explore the feasibility of a machine learning-based approach. This is achieved by applying the random forest (RF) technique to large-scale deformation maps spanning 2015–2018. Focusing on the northern part of Italy, we train the model to identify landslide, subsidence, and mining-related ground motion with which to construct a balanced training dataset. The presence of noisy signals and other sources of deformation is also tackled within the model construction. The proposed approach relies on the use of explanatory variables extracted from the A-DInSAR datasets and from freely accessible informative layers such as Digital Elevation Model (DEM), land cover maps, and geohazard inventories. In general, the model performance is very promising as we achieved an overall accuracy of 0.97, a true positive rate of 0.94 and an F1-Score of 0.93. The obtained outcomes demonstrate that such transferable and automated approach may constitute an asset for stake-holders in the framework of geohazards risk management.

1. Introduction

Starting from the last few decades, spaceborne Synthetic Aperture Radar (SAR) has been widely acknowledged as an active remote sensing tool enabling the accurate screening of the Earth's surface in all-weather, day or night (Bamler and Hartl 1998). Differential Interferometry SAR (DInSAR) and Advanced DInSAR (A-DInSAR) are powerful geodetic techniques that allow measuring accurate ground displacements from a pair of SAR images or multiple SAR images acquired over the same area in different times, respectively (Crosetto et al. 2016; Krishnakumar et al. 2021; Massonnet & Feigl, 1998). Recent developments in processing techniques coupled with ever-increasing technological progress in satellite-based SAR sensors have resulted in unprecedented quantities of monitoring data, which can no longer be investigated manually.

Ground motion caused by natural and anthropogenic sources recurrently affects critical assets and infrastructures resulting in multimillion euro direct and undirect damages (IPCC, 2022). In this context, the DInSAR and A-DInSAR techniques have been largely exploited to map, monitor, and model surface LOS (Line-Of-Sight) displacements due to geological hazardous events such as landslides (Antonielli et al. 2019; Del Soldato et al. 2018; Novellino et al. 2021; Rosi et al. 2017) land subsidence (Ezquerro et al. 2020; Osmanoğlu et al. 2011; Solari et al. 2018), volcanic and tectonic-related deformations (Atzori et al. 2019; De Luca et al. 2022; Weiss et al. 2020) as well as for man-induced displacements such as mining-related ground subsidence and uplift (Fan et al. 2021; Herrera et al. 2007; Ng et al. 2011; Samsonov, d'Oreye, and Smets 2013), settlement resulting from the building loads (Ciampalini et al. 2019), and engineering works (Milillo et al. 2018).

The launch of the European Space Agency (ESA) C-band Sentinel-1 (S1; Torres et al. 2012) satellites has laid the foundation for near real-time monitoring services and intensive mapping of ground deformation across the globe thanks to the enhanced revisiting

time (up to 6 days) and the free policy for data dissemination (Lanari et al. 2020; Showstack 2014). Moreover, the S1 satellites acquire data with a 250 km swath at 5 m by 20 m spatial resolution (single look), generating a large quantity of data. In this regard, several European countries have begun to implement or operate regional (Confuorto et al. 2021; Raspini et al. 2018) and national (Dehls et al. 2019; Kalia, Frei, and Lege 2017; Novellino et al. 2017; Papoutsis et al. 2020) ground motion services aimed at ground investigation, while continental-scale initiatives, such as the European Ground Motion Service (EGMS, https://land.copernicus. eu/pan-european /european-ground-motion-service), will provide deformation maps with a one year-update plan (Costantini et al. 2021; Crosetto et al. 2020).

At present, major efforts have been put into data processing methods and capacity, however the effective analysis and expert interpretation of large volumes of interferometric data constitute a serious limit to the prompt dissemination of information. Several millions of Measurement Points (MPs) are available for country-wide deformation maps; thus, the manual investigation of ground deforming areas might require very large time spans to be accomplished (Tomás et al. 2019). For istance, the Norwegian National Ground Motion Service provides Sentinel-1-related deformation data for more than 2 billion points in Norway (Bredal et al. 2019). Further, the PST-A project (*Piano Straordinario di Telerilevamento Ambientale*; Costantini et al. 2017) provided around 170 million points for the Italian territory. In this regard, different post-processing approaches to automatic detection of ground movements have been proposed, such as hot-spot-like methods (Bianchini et al. 2012; Lu et al. 2012) or clustering analysis (Barra et al. 2017; Festa et al. 2022; Montalti et al. 2019; Solari et al. 2019; Tomás et al. 2019). The first still requires a manual operation to investigate the origin of the resulting unstable hotspots, while the latter is restricted to thresholding in the spatial domain which can make the outcomes vulnerable to noisy data. Machine learning (ML) technologies have been successfully applied to wrapped interferograms for detecting volcanic deformation (Anantrasirichai et al. 2018), coal mining and tunneling subsidence (Anantrasirichai et al. 2021) and yet they fail to detect very localized deformation. Nava et al. (2022) prove effective to couple ML and SAR data for rapid landslide detection, however the temporal evolution and the estimation of the intensity of displacements are neglected.

In this article, we adopt an original approach for the implementation of the random forest (RF, Breiman 2001), a well-known and widely implemented ML technique, to automatically detect and classify ground deformations sources across Northern Italy using the spatial characteristics of A-DInSAR data. The significance of such a work relies on the necessity of periodically updating and assessing the geohazard activities detectable from continuously updated space-borne radar data. In remote sensing sciences, ML has been employed in several applications (Lary et al. 2016), such as land surface classification (Li et al. 2014), hyperspectral image classification (Camps-Valls et al. 2014), landslide susceptibility estimation (Catani et al. 2013). In this work, deformation maps resulting from the processing performed via the parallel computing version of the Small BAseline Subset (P-SBAS, Berardino et al. 2002; Casu et al. 2014; Manunta et al. 2019; Zinno et al. 2015) were used to retrieve MPs with outlying mean velocity values and spatially associate them to a set of predictors, while the time series of displacements were not considered. The most common morphometric features of the landscape, land cover, and geohazard inventories were taken into consideration as predictors in order to create a model for the purpose of classifying the origin of ground deformations in urban, non-urban, and forested environments. RF approach creates an ensemble of decision trees, established from randomly generated portions of the training data, whose votes are considered together to finalize a data-driven forest model. Our innovative idea is to adapt and apply RF at local scale to A-DInSAR measurement points. A major breakthrough regards the introduction of new explanatory variables in which can be obtained by the post-processing of P-SBAS data: an index expressing the motion trend in the East-West and vertical planes. Besides these, the method relies on other easily accessible input parameters extracted from a Digital Elevation Model (DEM) and from open access catalogs (e.g. landslide inventories). Therefore, we demonstrate through this work that is possible to adopt a simple ML-based classification scheme for assessing the source of active motions, which is a primary task in the management of large A-DInSAR datasets.

To the best of our knowledge, this is the first work presenting a simple implementation of the RF technique for the automatic recognition of the source of the motion bound to MPs with high deformation rates. The proposed novel approach acts at the pixel scale making this work a promising alternative considering the drawbacks of other ML classification methodologies (Anantrasirichai et al. 2021). The obtained data-driven model does not require any intervention by the user, and it is based only on the A-DInSAR dataset itself and on easily accessible information. The ascending and descending orbits deformation maps covering the Northern part of Italy and resulting from the exploitation of the whole S1 archive acquired from March 2015 to December 2018 were exploited. The procedure can also be tested in other geographical areas where DEM, land cover map, and geohazard inventories are available. The proposed methodology demonstrates the potential applicability of ML framework for the preliminary screening of newly updated A-DInSAR datasets covering very wide areas in continuity with the development of automated ground motion analysis may guarantee a valid support in land management and civil protection purposes.

2. Material and Methods

The flowchart of the presented methodology is synthesized in Figure 1. Input data and rationale of the procedure are explained in more detail in the following sections.

2.1. A-DInSAR dataset post-processing elaboration

A-DInSAR techniques exploit multiple SAR images acquired over the same area to estimate changes in path length between the satellite and the ground surface (Casu et al. 2014). Among the different approaches for detecting surface deformation and for analyzing their space-time characteristics, the Small BAseline Subset (SBAS) technique uses small baselines to limit the spatial decorrelation, multi-looked data to reduce phase noise and a coherence-based selection criterion (Berardino et al. 2002). Recently, Casu et al. (2014) developed an efficient SBAS processing chain, namely P-SBAS, which simultaneously exploits a fast access to the S1 data archives, High-Performance Computing resources and external geodetic data (de Luca et al. 2017;
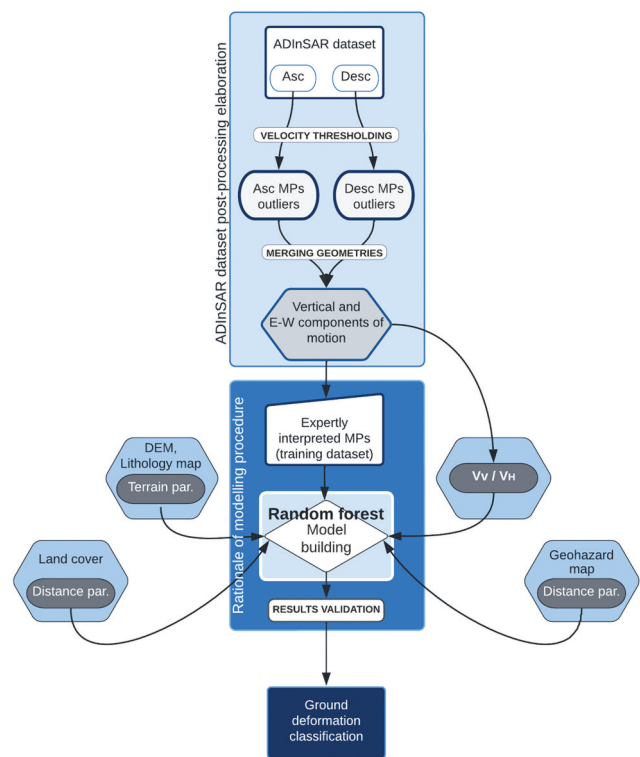

Figure 1. Workflow of the proposed methodology.

Lanari et al., 2020; Manunta et al. 2019). This approach, which is oriented toward large areas surface deformation analysis, generates a quantity of sparse and small MPs which can be fed to a RF classifier as point-like features.

S1 ascending and descending radar images acquired in C-band (wavelength 5.5 cm and incidence angle ranging from 33° to 43°) covering Northern Italy regions (with Tuscany region included), acquired from 23 March 2015 to 26 December 2018 were employed and processed by means of P-SBAS. An overlap between frames of almost 20% of the frame size in the azimuth direction has allowed an additional consistency check for the continuity preservation of the achieved results. The P-SBAS results are in the form of pixels with a grid spatial resolution of about 80 m x 80 m. The processing procedure includes appropriate interferogram phase filtering and accurate phase unwrapping procedures (De Luca et al., 2022; Pepe and Lanari 2006). Global navigation satellite system (GNSS) position time series are selected and exploited to filter out atmospheric phase screen and possible residual phase artifacts. The final average measurement accuracy is of about 5 mm on the single retrieved displacement time series (Lanari et al.

2020; Manunta et al. 2019). However, time series of displacement, despite being available for each MPs, are not employed for this work proof of concept since the presented approach is based on the mean velocity computed for each pixel by linearly fitting the time series. To include only very accurate points, only pixels holding a temporal coherence value greater than 0.9 were retained for this analysis. 2,749,737 velocity MPs were counted in the ascending orbit, while 2,594,041 MPs compose the descending velocity map. The overall MPs density ranges from 18.1 to 19.2 points per km^2 for the ascending and descending pass, respectively. The higher density and significantly cluster are in urban areas compared to forested and agricultural areas. The mean velocity value (μV) of both datasets is around zero (Figure 2), while the standard deviation (σ) of the velocity distribution stands at ±0.2 cm/year which usually corresponds to the displacement interval related to stable ground conditions (Catani, Canuti, and Casagli 2012; Raspini et al. 2018).

Given that the velocity values in the two datasets share approximately bell-shaped distributions (Figure 2), the standard deviation of the sample was used as a cutoff threshold for identifying MPs with relatively high deformation rates. In particular, any value out of the range defined by μV±3σ (i.e. ± 0.6 cm/year) can be considered as an outlier (Santos 2020) and therefore is retained. This criterion allowed to restrict the analysis to a total of ~145,000 unstable ground measurements (Figure 2; 76,464 MPs and 68,193 MPs from the ascending and descending passes, respectively), finally selected as targets of our work.

Furthermore, being a key-step for the analysis, the information from the two different LOS geometries were combined to resolve the E–W (V_H) and vertical (V_V) velocity components (Notti et al. 2014) by exploiting the S1 P-SBAS data architecture. Since the P-SBAS mean deformation velocity maps are divided following an 80 m x 80 m geographical regular grid (de Luca et al. 2017), the displacement vector decomposition
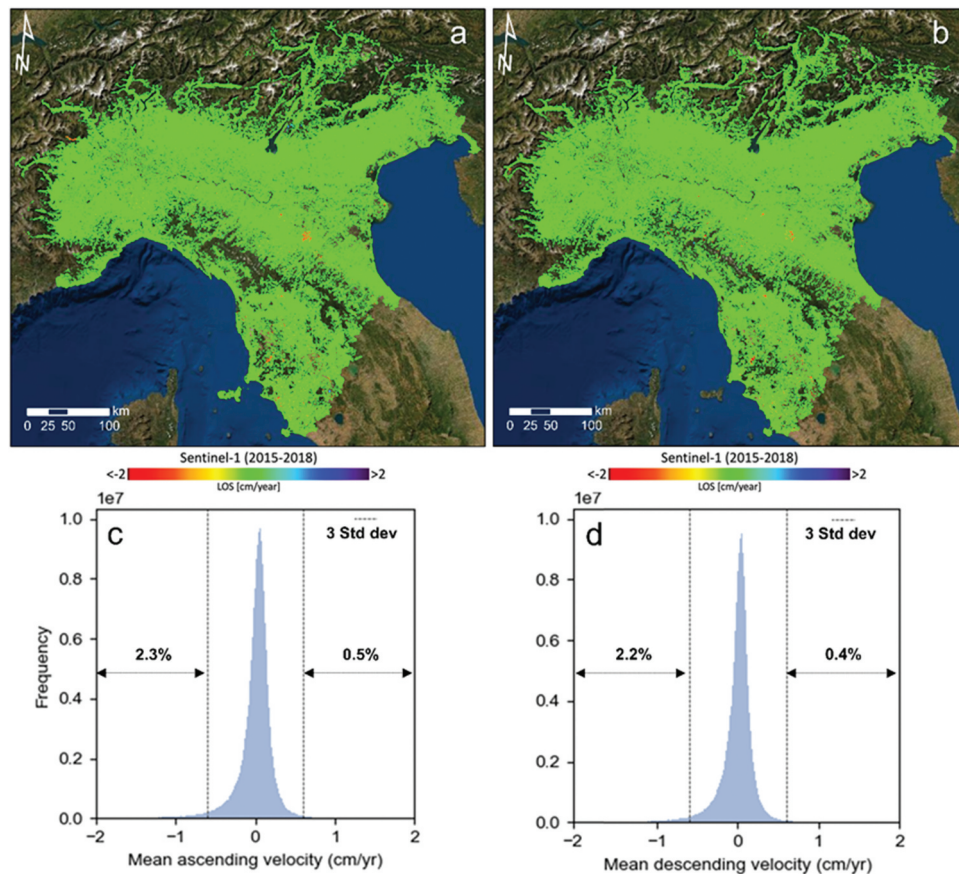
Figure 2. (a) Ascending and (b) descending velocity maps and related frequency distribution of the (c) ascending and (d) descending mean velocity (μV). The cutoff threshold, namely three times the standard deviation (3σ), is represented with the dashed line.

was computed for those pixels accounting for coincident LOS ascending and descending measurements. This can be done by solving the following equation (1) (Notti et al. 2014):

$$V_V = \frac{((V_D/\text{elos}_D) - (V_A/\text{elos}_A))}{(\text{hlos}_D/\text{elos}_D - \text{hlos}_A/\text{elos}_A)}; V_H$$
$$= \frac{((V_D/\text{hlos}_D) - (V_A/\text{hlos}_A))}{(\text{elos}_D/\text{hlos}_D - \text{elos}_A/\text{hlos}_A)} \qquad (1)$$

where V_A and V_D are the ascending and descending LOS velocity vectors; $\text{hlos}=\cos(\alpha)$; $\text{elos}=\cos(1.571 - \alpha)*\cos(4.712 - \theta)$; θ and α are the azimuth and the incidence angle, respectively.

Despite not being available for every MPs, the reprojection of the displacement rate on the E–W and vertical plane will constitute an important explanatory information in the proposed model-building framework (see Section 2.2.) As a result of the post-processing phase, the final dataset counts 20,212 MPs.

2.2. Model parameters

2.2.1. Terrain parameters
In this work a 10 m cell-size grid DEM, derived from TINITALY/01 project (Tarquini et al. 2007) released by Istituto Nazionale di Geofisica e Vulcanologia (INGV), was used. Some of the main DEM-derived indices largely employed in landslide susceptibility (Catani et al. 2013) are here used to infer a correlation between the spatial occurrence of MPs with outlying velocities and the motion trigger. Regarding our study case, we extracted the exact numerical and categorical value to every MPs centroid location by computing and resampling the following raster data into our original 80 m grid (Table 1):

- Elevation, which may be indicative of different altitudes tied to various environmental settings

(Ayalew, Yamagishi, and Ugawa 2004) and different types of surface deformations.
- Aspect identifies the downslope direction. The raster pixels are expressed in terms of compass direction that the surface faces at that location with respect to the cardinal direction.
- Slope describes the slope for each raster cell in degrees based on the elevation at each point.
- Curvature, which expressed the relative change in slope and is computed as the second derivative of the surface topography.
- Profile curvature, indicative of the rate at which the slope gradient changes toward the direction of maximum slope.
- Planar curvature, respect to the latter, is calculated orthogonally to the direction of the maximum slope.
- Topographic Wetness Index (TWI), which can be though as a steady state wetness index (2), where A is the upslope contributing area of each raster pixel and β is the slope angle.

$$ln\left(\frac{A}{\tan \beta}\right) \qquad (2)$$

The characteristics of the soil coverage may be considered as an important driving property to discriminate different causes of ground deformation. The diverse geomechanical and hydraulic properties reflected by the different units were considered. Their spatial distribution is well depicted in the 1:500 000 lithological vector map of Italy (www.isprambiente.gov.it/progetto-1250-ita). We grouped all the geological formations into eleven classes: (i) granular soils; (ii) debris and alluvial deposits; (iii) clays and compact marls; (iv) massive calcareous rocks; (v) igneous intrusive rocks; (vi) igneous effusive rocks; (vii) schistous rocks; (viii) metamorphic rocks; (ix)

Table 1. Explanatory parameters used for the definition of the model.

Parameters	Type	Data Format	Scale/Resolution	Source
Elevation	Cont.	Grid	10 m	TINITALY/01
Aspect	Cont.	Grid	10 m	derived from TINITALY/01
Slope	Cont.	Grid	10 m	derived from TINITALY/01
Curvature	Cont.	Grid	10 m	derived from TINITALY/01
Profile curvature	Cont.	Grid	10 m	derived from TINITALY/01
Planar curvature	Cont.	Grid	10 m	derived from TINITALY/01
TWI	Cont.	Grid	10 m	derived from TINITALY/01
Lithology	Categ.	Shapefile (polygon)	1:500,000	ISPRA
Landslide distance	Cont.	Shapefile (polygon)	1:10,000	IFFI
Mining sites distance	Cont.	Shapefile (polygon)	1:100,000	CLC2018
Kvh	Cont.	Grid	80 m	derived from A-DInSAR dataset

Abbreviations: Cont. = Continuous Categ. = Categorical

evaporite rocks; (x) complex units with mainly pelitic layers; and (xi) complex units with mainly calcareous and arenaceous layers. In our rationale, every MP is labeled according to the lithology class within which the pixel's center falls.

2.2.2. Distance parameters

Proximity distances to important land motion predisposing factors can be of great help to calculate and to determine the kind of surface displacement (Table 1). For our analysis we used the polygons contained in the 1:10,000 scale Italian Landslide Inventory (*Inventario dei Fenomeni Franosi in Italia*, IFFI project; Trigila, Iadanza, and Guerrieri 2007), since it depicts the spatial distribution of mass movements for the Italian territory. The spatial diffusion of open-pit extraction sites and flooded mining pits were extracted from the III level classification of 2018 CORINE (https://land.copernicus.eu/pan-european/corine-land-cover) Land Cover map (CLC) at 1:100,000 scale provided by the Copernicus Land Monitoring Service. We computed the distance between every MPs centroid and the closest line segment to the polygon features in an equidistance projected coordinate system, where a distance of zero is here considered when a MP lies within a polygon.

2.2.3. K_{VH} parameter

In order to find out the deformation pattern which may be linked to a particular geological or anthropogenic phenomenon causing ground displacements, we derived a novel index (Table 1) based on the interferometric measures reprojected on the E–W and vertical plane (see Section 2.1). We introduce the variable K_{VH} (3) that is the ratio between V_V and V_H, namely the vertical (Up-Down) and the absolute horizontal (East-West) component values:

$$K_{VH} = V_V/|V_H| \qquad (3)$$

The use of K_{VH} is meant to introduce a strong attribute regarding the prevailing direction of the movement evidenced by the MP for which the index is calculated. A prevailing negative value of K_{VH} would suggest a downward motion, while a prevailing positive value of K_{VH} would indicate an upward motion. By definition, a value comprised between -1 and $+1$ would suggest a predominant horizontal movement (i.e. toward West or East).

2.3. Random forest

Random forest (RF) is an ensemble ML method which have become very popular in the past decade due to its excellent performance in handling non-parametric multidimensional classification and regression problems and in showing low chances of overfitting (Breiman 2001). RF is used to predict an outcome by involving the bootstrapped aggregation of several regression trees. A model is created based on a random selection of the training data from the input dataset, while the data not included are referred to as "out-of-bag" (OOB). The generation of a random subset of features ensures low correlation among decision trees, whose classification outcomes are aggregated to identify the most popular result. Therefore, the majority vote will yield the predicted class and the OOB sample is then used for cross-validation, finalizing the class prediction. Here, the common characteristics retrievable from the velocity maps plus other ancillary maps (i.e. numerical, categorical, and distance-based variables) were investigated as well as combined for explaining different causes of deformation tied to ground moving MPs. In other words, the procedure is concerned with predicting a categorical variable to point-like objects.

The method itself proves to be largely applicable in many scenarios. Several advantages are associated to RF: i) massive datasets and data heterogeneity can be dealt (Genuer et al. 2017), also allowing for the mixed use of categorical and numerical variables; ii) complex data fabric can be managed, and it is capable of accounting for nonlinear high-dimensional features which may interact with each other (Scornet, Biau, and Vert 2015); iii) it performs well in model interpretation and is fast to train. A further key benefit is the ability of RF to statistically evaluate variable importance to the model. Importance is calculated using Gini coefficients, that is how much the model's accuracy decreases when a given variable is excluded. This capability allows for studying the relative contribution of the different explanatory variables on the overall result.

A way to estimate the ensemble error is by computing the "out-of-bag error" (OOBE), which measures the prediction errors of RF through the comparison of the OOB predicted responses against the true responses. The overall performance of the model can be evaluated through diagnostic metrics aimed

at assessing how well features tied to a particular category are correctly predicted or miscategorized, by using TP, TN, FP, FN which are true and false positive and negatives, respectively. In this work, following the proper setting of the forest parameters, a balanced dataset for model training was tested and the classification performances are measured by means of Accuracy (4), True Positive Rate (TPR) (5) and F1-Score (7), which is the harmonic mean of TPR and Precision (6). These parameters are mathematically detailed below:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (4)$$

$$TPR = \frac{TP}{TP + FN} \quad (5)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (6)$$

$$F1 - Score = \frac{2 \cdot TPR \cdot Precision}{TPR + Precision} \quad (7)$$

2.4. Rationale of modeling procedure

To automatically classify the phenomena responsible for triggering the outlying MPs, we need to build the RF model by feeding it with ground truth examples of different classes of expertly interpreted MPs (Farina et al. 2006). In this article, we concentrate on slope deformation caused by landslides, subsidence (due to both natural or anthropogenic sources, such as groundwater pumping, natural compaction, or loading) and mining-associated activities (which might be linked to complex deformation patterns) as they are the most common cause of surface displacements in the test area. We also consider the outlying MPs arising from other unspecified triggering factors or inherent noisy signals which are still included in our filtered A-DInSAR dataset after the velocity thresholding. Real sites across Northern Italy which have been analyzed remotely within the framework of monitoring services (Confuorto et al. 2021) or through field investigations were exploited to build up the training dataset, while uncorrelated signals were included in the model by randomly picking up sparse MPs to reproduce noise and any other undefined ground deformation. As expected,

the manual classification of MPs resulted in an imbalanced dataset. Since ML methods for supervised classification tend to provide different performances for the samples in the most populated classes compared with the samples belonging to the least numerous classes (Johnson and Khoshgoftaar 2019), we adopted a simple data-leveling method to obtain a balanced training dataset. We performed random under-sampling for all the classes based on the number of elements of the least populated class, thus retaining a total of 2,184 features equally divided into four groups (Figure 3): (i) landslide (L); (ii) subsidence (S); (iii) mining-related displacement (M); (iv) noise or other phenomena (ND).

The validation dataset accounts for 15% of the training dataset (i.e. the OOB sample), while the test dataset is constituted by all the remaining MPs which could not be safely interpreted, namely 18,028 MPs.

We tuned the RF model settings in regard to the generation of the most accurate model. Different number of trees were here tested trying to minimize OOBE and building a robust model, where every decision tree is created using a random subset of the training data available. Decreasing the number of trees, the computation time of the model is reduced, though not constituting a major concern in our case given the dimensions of our training dataset. To make the ensemble of decision tree an effective predictor, a maximum number of randomly sampled variables for each tree is required, which is here considered empirically by looking at the square root of the total number of explanatory variables. On the other hand, the maximum tree depth (i.e. the maximum number of splits that will be made down a tree) is a data-driven parameter which is here automatically calculated based on the number of trees created and the number of variables included.

3. Results and discussion

3.1. Model tests results

Once fixed the RF model settings, we executed 50 runs on randomized portions of the training dataset, starting from a number of 250 trees and then moving on to 500 and 1000 trees. Considering the full parameters set, the mean and standard deviation of model accuracy obtained through the several runs were estimated and then the OOBE evaluated as the
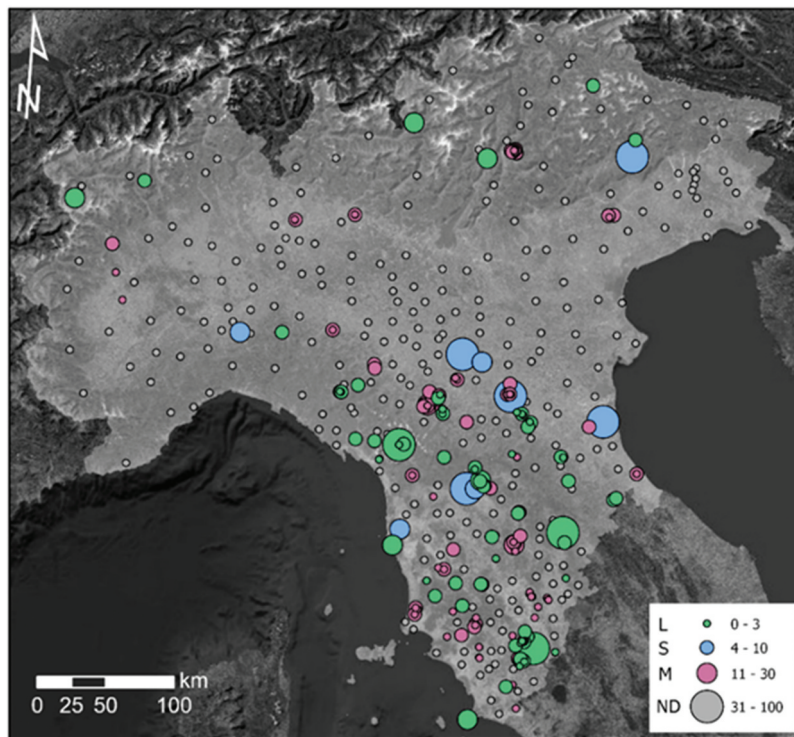
Figure 3. Map depicting the location of expertly interpreted outlying MPs constituting the balanced training dataset. MPs related to landslide (L), subsidence (S), mining-related displacements (M) and noise or other (ND) are shown by means of bubbles whose dimension is proportional to the number of aggregated features within the same discrete phenomenon.

Table 2. Mean and standard deviation (Std dev) of accuracy computed for 250, 300 and 1000 trees. Out-of-bag errors (OOBE) displayed for every predicting class and related to the most accurate model.

Number of trees	250	500	1000
Mean accuracy (50 runs)	9.2e-1	9.3e-1	9.3e-1
Std dev accuracy (50 runs)	1.5e-2	1.6e-2	1.6e-2
Best model accuracy	9.5e-1	9.6e-1	9.6e-1
	Best model OOBE	Best model OOBE	Best model OOBE
Landslide (L)	5.2%	4.7%	4.3%
Subsidence (S)	5.9%	5.1%	4.5%
Mining-related disp. (M)	2.0%	2.0%	1.9%
Noise or other (ND)	21.3%	20.5%	20.0%

percentage of incorrect classifications for every predicting class by looking at the most accurate model (Table 2).

Results shows that the accuracy values obtained with different number of trees are very similar to each other, with mean values ranging from 9.2e-1 to 9.3e-1 and stable standard deviation values spanning from 1.5e-2 to 1.6e-2. Comparing the OOBE from the best models, a slight decrease of misclassification in the OOB sample was noted, probably correlated with the rise of the number of trees. The latter parameter does not seem to significantly affect the computation time,

which is comparable in all cases. We can assert that the performances obtained by applying the classification to the training dataset do not considerably enhance across several runs and by varying the number of trees, thus evidencing the practical implementation of a RF method in this kind of approach.

3.2. Variable importance and model validation

The best model achieved by setting 1000 trees has taken as a representative to explore its outcomes and characteristics through diagnostic metrics. We evaluated the importance of model variables by looking at the Gini coefficients calculated for every deployed explanatory training variable.

The distance parameter related to CLC mining sites is the main driving variable holding a median value of importance equal to 6.3, followed by K_{VH} with a value of 6.0 (Figure 4). These parameters show similarly higher values if compared to the remnants: In fact, together they account for a total 27% of overall importance in the model, thus remarking their breakthrough role for catching the statistical evidence for discriminating high-velocity MPs in different classes. It
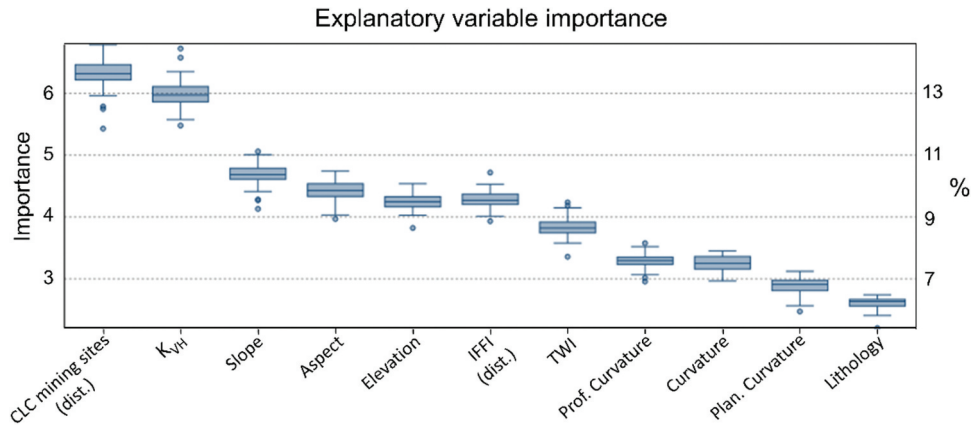
Figure 4. Boxplot related to the importance of every explanatory variable computed as the sum of the Gini coefficients from all the trees. The values are also reported as the percentage of the total sum of Gini coefficients.

can be inferred that the first parameter can be safely correlated with the classification of true mining-related displacement, while K_{VH} may be relevant to distinguish prevailing vertical surface movements (subsidence) from phenomena responsible for mainly horizontal motions (landslide). Slope, aspect, elevation and distance to IFFI landslides constitute a set of fairly relevant parameters with median importance values comprised between 4.7 and 4.3. Apart from TWI, the last set of parameters composed by profile curvature, total curvature, plan curvature and lithology accounts for a minor responsibility in the splitting process of a decision tree (Figure 4), with the local differences in soil/rock composition described by the lithological map being marginally useful to classify surface displacements (2.6 median importance).

The validation dataset mean confusion matrix indicated as the percentage of predicted samples with respect to the ground truth labeled samples, and a list of performance metrics achieved for every class are reported in Table 3. In general, classification indices are very good considering all classes (arithmetic mean of indices values), with an overall accuracy of 9.7e-1,

a TPR of 9.4e-1 and a value of F1-score equal to 9.3e-1. Class M (mining-related displacement) has the highest indices among all the other classes, while classes L (landslide) and S (subsidence) show comparable high performances with both recording relatively smaller F1-score values (9.5e-1 and 9.2e-1, respectively). On the other hand, MPs are more likely to be incorrectly labeled as ND (Noise or other) as witnessed by the relatively lower performance rates achieved (Table 3). In particular, the percentage of actual positives ND which are correctly identified (i.e. the TPR) is equal to 8.1e-1. This can occur considering that class ND gathers together high-velocity MPs with multiple deformation sources, including noisy and sparse signals which could not be related to a discrete and clearly discernible phenomenon. Given that, the lack of peculiar features makes their classification more difficult to the RF according to the parameters considered. This is confirmed also by the results reported in the confusion matrix (Table 3). Class ND is more often confused with class S and secondly with class L. Among the other classes, S is never confused with L and vice versa, thus demonstrating that the

Table 3. Validation dataset mean confusion matrix and performance metrics reported for every class.

Class	L	S	M	ND	
		Validation dataset: classification			
Landslide (L)	95.7%	0.0%	1.2%	3.1%	75-100%
Subsidence (S)	0.0%	95.5%	1.2%	3.3%	50-75%
Mining-related disp. (M)	0.7%	0.0%	98.1%	1.2%	25-50%
Noise or other (ND)	5.4%	12.2%	2.4%	80.0%	0-25%

Diagnostic metrics	Per class				Overall
Accuracy	9.7e-1	9.6e-1	1.0	9.4e-1	9.7e-1
TPR(True Positive Rate)	9.5e-1	9.8e-1	1.0	8.1e-1	9.4e-1
F1-score	9.5e-1	9.2e-1	9.9e-1	8.7e-1	9.3e-1

deployed parameters provide strong statistical attributes to discern subsidence from landslides. Mining-related displacements are confused in 2.4% of the cases with ND and only in 1.2% of the cases with both L and S.

3.3. Northern Italy predicted outcomes

After the training and validation phases, we used the best RF model to predict the source of displacements recorded through the A-DInSAR deformation map of Northern Italy. The classification concerned a total number of 18,028 high deformation MPs obtained from the training of 2,184 features expertly classified according to well-known cases of landslide, subsidence, and mining deformation phenomena. The geographical distribution of both outcomes and training data are depicted in Figure 5.

The proposed method detected that L and S outlying MPs are found in different morphometric and physic environments, reflecting the different geological processes. 2,712 MPs, which were labeled as L,

are located in correspondence of mountainous areas (mainly across the Northern Appennines), while 8,255 S points group themselves in the main plains (e.g. the Po Plain) and the main valley floors (e.g. alpine valleys) of the study area. We estimate a mean slope value of 16.1° for L features (with a standard deviation equal to ±9.4°), while for S-classified MPs we retrieve a mean slope of 0.5° and ±1.3° standard deviation thus confirming that slope constitutes an important parameter in our model. Another strong discriminant between L and S relates to elevation: we find mean values of 55.8 ± 66.6 m a.s.l. for S points and 528.4 ± 320.6 m a.s.l. for L points.

On the other hand, 773 MPs are labeled as mining-related displacements (M) while 8,472 are classified as noise or other phenomena (ND). Most of the M features (i.e. 81% of the totality) appears to cluster within the boundaries of the CLC mining sites, while the remnants are situated in direct proximity (ranging from 0 to 576 meters) since the main driving parameter of our model is the distance from the mining locations. The other morphometric parameters such as slope, elevation,
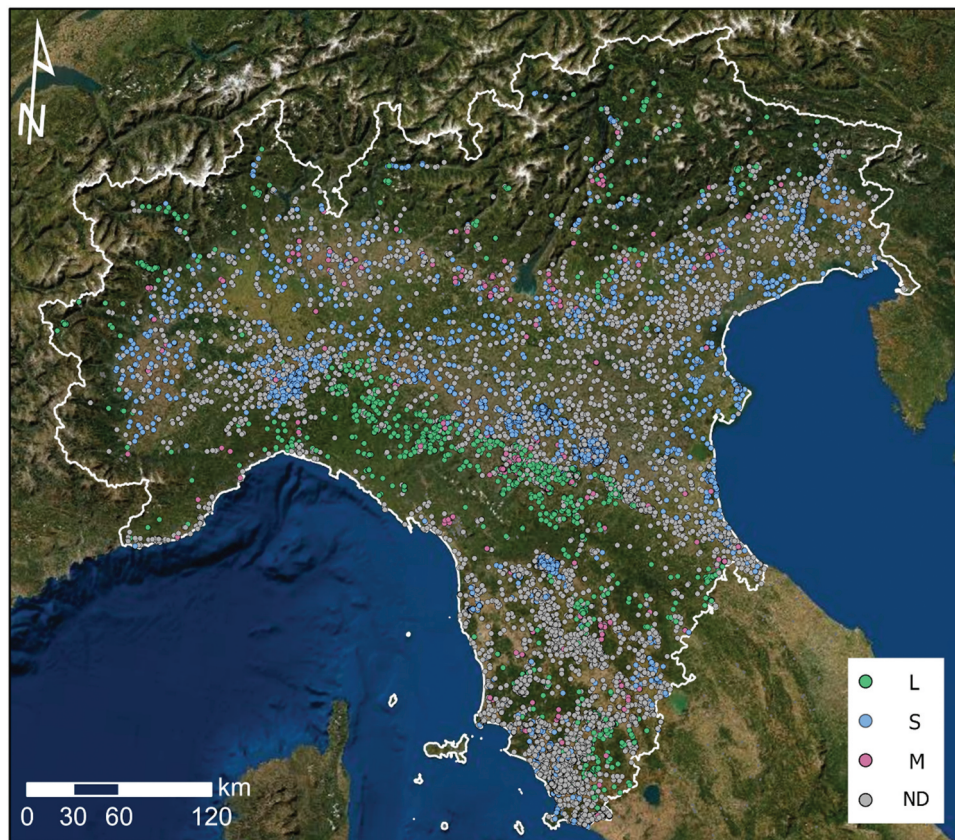
Figure 5. Northern Italy classification results of DInSAR-based MPs with high velocities according to 4 classes of deformation source: landslide, subsidence, mining-related displacement and noise or other phenomena.

aspect, and TWI do not show meaningful statistical distributions for M predicted features since mining sites are not spatially restricted to a particular physiographic environment. Concerning MPs labeled as ND, they hold, by definition, largely dispersed values among all the parameters.

The classification outcomes at local scale for several sites across the study area are shown in Figure 6.

The method detects that the majority of outlying MPs in the metropolitan area of Bologna (Emilia Romagna Region) refer to a well-known subsidence phenomenon (Figure 6a) with a long-term evolution due to local anthropogenic activities (i.e. overexploitation of the aquifers) and natural causes, such as tectonics (Stramondo et al. 2007). In the same area, several MPs linked to mining-related deformations are

classified in correspondence of 3 open-pit extraction sites of inert material which run stretches along the Reno River. In general, we find good agreement between the manual classification and the predicted classification, even if in few MPs in the buffer zone around mining areas appears misclassified and were marked as ND. This can be ascribed to a misleading combination of explanatory variable values that RF fail to interpret. On the other hand, in Figure 6b a case near Oriano hamlet (Emilia Romagna Region) is depicted, where outlying signals within a small area are correctly correlated to landslide and to mining areas which were previously cataloged as such. Also, our method proves to be able to detect L classified MPs which do not fall inside the IFFI landslide inventory. As evidence of this, in Figure 6c we show the
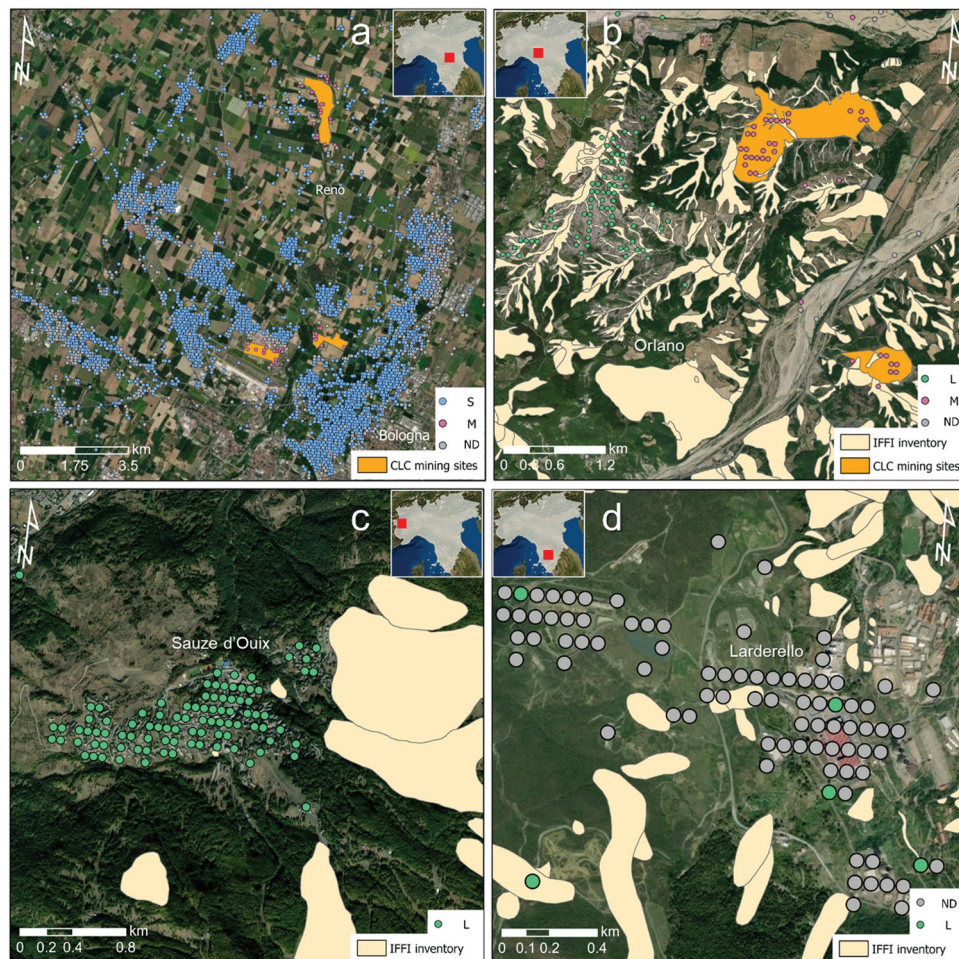
Figure 6. Classification results displayed for different sites across the study area. (a) Subsidence phenomena and extraction sites in Bologna city and surroundings (Emilia Romagna Region); (b) landslide phenomena and mining-related displacements in the Oriano hamlet (Emilia Romagna Region) in correspondence of IFFI and CLC inventories; (c) correctly interpreted landslide phenomena in Sauze d'Oulx (Piedmont Region) which has not been previously reported in the IFFI inventory; (d) surface deformations related to the geothermal activity in Larderello (Tuscany Region) where most of the outlying MPs have been correctly classified as noise or other sources of displacement.

case of the Sauze d'Ouilx hamlet (Piedmont Region) which is recognized to be affected by a large deep-seated gravitational slope deformation (DGSD; Fioraso et al. 2020). In this context, the proposed classification method could provide an aid in updating geohazard inventories.

Another interesting capability shown by our model concerns the indirect classification of source of deformations which were not considered in the training dataset. The evidence is provided by the case of Larderello geothermal field, which is settled in southern Tuscany region (Figure 6d). Almost every MPs is being correctly detected as ND (i.e. noise or other source of deformations) since the origin of the deformation is not related to any L, S or M phenomena. Therefore, here the combined statistical instances of the explanatory parameters have proven to be effective in not confusing ND with S, despite the area suffering from severe vertical movements which have persisted in the last decades (Botteghi et al. 2012).

3.4. Impact of variable KVH

We also analyzed the impact of the newly introduced K_{VH} variable in our modeling procedure. Being K_{VH}

a proxy for estimating the deformation pattern on the vertical and East-West directions, we expect that it allows constituting a solid explanatory index in the RF classification procedure. To this end, we performed the two-tailed Mann Whitney U test (Fay and Proschan 2010) for checking a significant difference between the K_{VH} distributions of the predicted MPs (Figure 7). We tested the four classes non-normal distributions pairwise for a level of significance (α) equal to 0.05, conducting 6 tests for every combination. The U statistic with normal approximation is then used to compute the probability p-value and evaluated against the selected level of significance: if p is less than α, the null hypothesis of no correlation (H_0: $class_1 = class_2$) is rejected and we accept that the difference between $class_1$ and $class_2$ populations is big enough to be statistically significant. Tests outcome confirms that each class distribution (Figure 7) substantially differs from the others since we obtain a p-value equal to 0.5e-2 when testing M predictions against ND predictions, while we obtain a p-value of zero for all the other test combinations. In the end, we demonstrate that in the proposed methodology the use of K_{VH} is fundamental for discerning between different sources of deformations detected through A-DInSAR data.
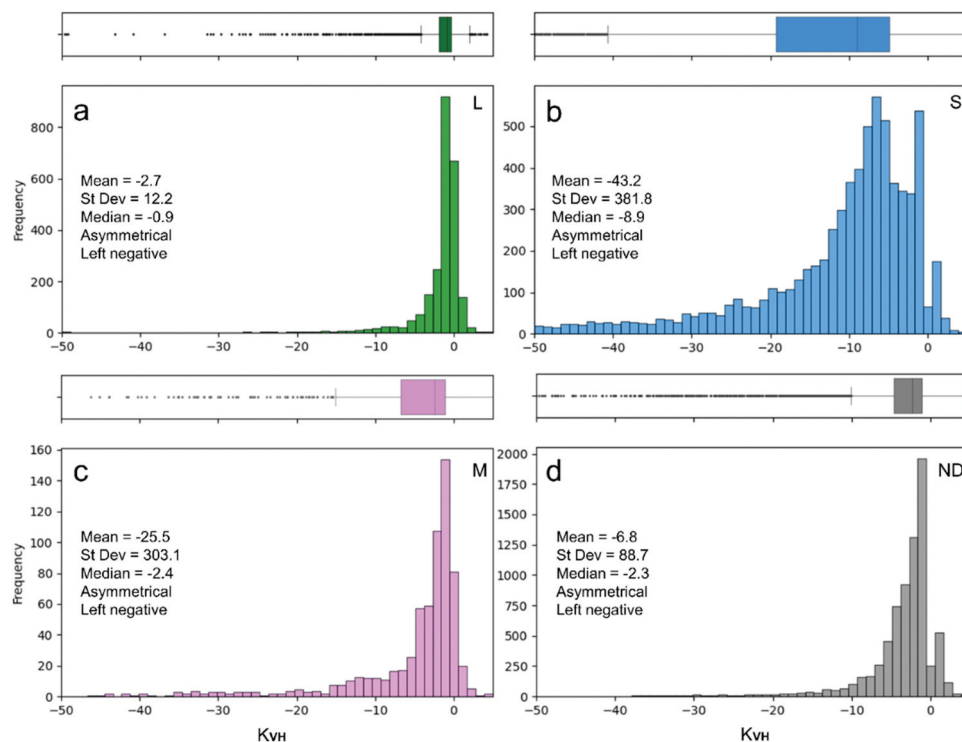
Figure 7. Histogram and plotbox of the KVH distributions related to predicted (a) landslide MPs, (b) subsidence MPs, (c) mining-related MPs and (d) noise or other MPs.

3.5. Spatial pattern analysis

The spatial distribution and aggregation of the high velocity MPs partly reflect the parent dataset distribution. Moreover, our data should be inherently grouped in space being this a constraint of A-DInSAR technique (MPs usually aggregate around reflecting elements that satellites can easily see; Lu et al. 2012). For this reason, we evaluated the Average Nearest Neighbor (ANN) for every predicted class, and we compared them to the overall distribution. In our case, ANN is based on the averaging of all the distances found between each MP of the same class and its nearest neighbor. The ANN ratio is the observed average distance divided by the expected average distance of a hypothetical random distribution with an equal number of points covering the same total area. If the observed average distance is less than the expected for a random distribution, the MPs class being analyzed is considered clustered, or conversely, it is considered dispersed. As can be seen in Table 4, all the classes exhibit ANN ratios largely greater than 1, denoting a trend toward dispersion with likelihood that the dispersed pattern could be result of random chance (z-scores largely greater than the critical value) less than 1%. This could be ascribed to the MPs density which has been greatly lowered during the post-processing operation (see section 2.1) while preserving the underlying surface area. On the other hand, if we compare the ratio between predicted classes ANN and the overall ANN related to whole outlying MPs we can retrieve interesting remarks: M, L and S classes appear to be more clustered in space than the overall distribution, with ANN ratios being around half of the overall ANN (Table 4). Instead, MPs belonging to ND class appear to be slightly more dispersed (value of ANN equal to the overall ANN multiplied by 1.1). As a result, we can deduce that the proposed method confirms to be helpful in detecting spatially aggregated features associated to discrete deforming phenomena, and besides, it proves reliable in identifying noisy deformation signals.

3.6. Discussion

Detecting and characterizing ground movements affecting urban and extra-urban territories is a crucial task in the framework of risk assessment and management conducted by entitled governmental departments and specialized bodies. Being the big streams of SAR data a concrete reality nowadays, an effort must be stepped up to explore efficient and well-functioning implements to help authorities making strategic decisions in terms of both public and private asset protection.

In this paper, we address the need to find a transferable approach to safely analyze and exploit large volumes of interferometric data in the framework of screening surface displacements related to both natural and anthropogenic sources. We prove the efficiency of the presented methodology aimed at the automated detection of pixel scale ground deformations, departing from A-DInSAR deformation maps covering very wide areas and composed of millions of MPs. As mentioned by Tomás et al. (2019), the manual inspection of ground deforming phenomena across large portions of territory can be very time-consuming. Based on the time spent on building the training dataset used within this work, we estimate that the expert interpretation of high velocity MPs covering Northern Italy might take up to several weeks. Conversely, it only took around 8 minutes to implement our ML approach on an Intel Core i7-8700 at 3.20 GHz (6 cores, 12 threads, 12 MB cache, 32 GB RAM, 500 GB SSD disk and Windows 10, 64 bits), considering both the training and the predicting operations conducted over 2,184 and 20,212 MPs, respectively.

Some limitations are linked with the usage of the presented methodology. Besides the largely acknowledged drawbacks of interferometry techniques (e.g. impossibility to capture fast movements, difficulties in retrieving coherent data from vegetated areas), the quality of the input A-DInSAR dataset and of the ancillary information is crucial for the correct

Table 4. Computed Average Nearest Neighbor ratio (ANN) for every predicted class and for the overall spatial distribution of outlying MPs, where also every Class ANN is evaluated against the overall ANN.

	Observed mean dist. (m)	Expected mean dist. (m)	ANN ratio	Class ANN/Overall ANN
Overall	512.1	1.3	394.0	-
L	752.2	3.6	209.0	0.5
S	491.5	2.1	234.0	0.6
M	981.0	6.9	142.2	0.4
ND	937.8	2.1	446.6	1.1

extrapolation of statistically relevant value intervals with regards to the deployed parameters used for the RF model. In our case study, the coexistence of different kind of deforming phenomena within the same area occasionally leads to a mismatch in the screening process of the source of displacement (Figure 6). In particular, mining areas might show typical complex displacement patterns which can replicate subsiding or slope movements, especially if restricting our analysis to satellite radar data only. This translates into the very high importance value of the distance parameter related to CLC mining sites, which demonstrates to be a real defining factor in the proposed model. Moreover, since ND is a broad ensemble containing different sources of unstable radar signals, our method understandably registers weaker predicting performances for that class. In this respect, it should be pointed out that the present approach is liable to misclassify MPs if the built model is not trained with a proper number of features or an adequate selection of explanatory variables. Moreover, the method requires the decomposition of the LOS velocities into the vertical and horizontal components, which can be accomplished only where both ascending and descending measurements are available. This implies a decreasing of the scene coverage, thus resulting in the decrease of the number of classifiable targets. Nevertheless, the method proves to be robust in overall correctly classifying the retrieved classes, demonstrating that by only using satellite data, it is possible to achieve a quick, automated, and reliable screening of the source of deformation of unstable radar targets.

Future developments will consist of detecting more specific types of sources responsible for ground deformations visible by spaceborne radar sensors, such as tectonic, volcanic-related displacements and other man-induced ground motions. In order to accomplish that, and for the purpose of increasing the accuracy of the assessments, new indices and parameters can be designed and implemented within recently developed machine learning algorithms. In particular, in the next future it will be assessed the feasibility of advanced data mining methods based on A-DInSAR time series analysis for a comprehensive and precise space-time recognition of deformational phenomena whose rate of displacement change over time.

4. Conclusions

In this work, we present an innovative machine learning (ML) methodology for the automatic detection and classification interferometric-based instability phenomena and its application to a regional scale level dataset. This data-driven approach relies on the transferable use of random forest (RF) applied to sparse Measurement Points (MPs) derived from P-SBAS (Parallel Small Baseline Subset) Advanced Differential Interferometric technique (A-DInSAR) and on input parameters which can be easily extracted from land cover maps, geohazard inventories and DEM. The results show great prediction performances for a four-class overall classification with a F1-score equal to 9.3e-1, a True Positive Rate (TPR) equal to 9.4e-1 and an Accuracy value of 9.7e-1. Visually inspected and interpreted MPs are used as geostatistical instances for modeling and exploring the spatial distribution of different ground deformation processes.

This method can be employed and reproduced with any point-like features for which are available Line-Of-Sight measures from both the ascending and descending interferometric passes and in other geographical areas where DEM, geohazard catalogs and land cover maps area available. Our approach is tested using the velocity map generated by P-SBAS processing of S1 archive dated between March 2015 and December 2018 and successfully detect several types of deformation occurring across northern Italy. In the framework of land management, the ever-increasing availability and quality of space borne radar data with the combined use of time series analysis will enable even more accurate hazard modeling applications.

Acknowledgements

Disclosure statement

Funding

Data availability statement

The data that support the findings of this study are available under reasonable request to the authors.

References

Anantrasirichai, N., J. Biggs, F. Albino, P. Hill, and D. Bull. 2018. "Application of Machine Learning to Classification of Volcanic Deformation in Routinely Generated InSAR Data." *Journal of Geophysical Research: Solid Earth* 123 (8): 6592–6606. doi:10.1029/2018JB015911.

Anantrasirichai, N., J. Biggs, K. Kelevitz, Z. Sadeghi, T. Wright, J. Thompson, A. M. Achim, and D. Bull. 2021. "Detecting Ground Deformation in the Built Environment Using Sparse Satellite InSAR Data with a Convolutional Neural Network." *IEEE Transactions on Geoscience and Remote Sensing* 59 (4): 2940–2950. doi:10.1109/TGRS.2020.3018315.

Antonielli, B., P. Mazzanti, A. Rocca, F. Bozzano, and L. Dei Cas. 2019. "A-DInSAR Performance for Updating Landslide Inventory in Mountain Areas: An Example from Lombardy Region (Italy)." *Geosciences* 9 (9): 364. doi:10.3390/geosciences9090364.

Atzori, S., A. Antonioli, C. Tolomei, V. de Novellis, C. de Luca, and F. Monterroso. 2019. "InSAR full-resolution Analysis of the 2017–2018 M>6 Earthquakes in Mexico." *Remote Sensing of Environment* 234: 111461. doi:10.1016/j.rse.2019.111461.

Ayalew, L., H. Yamagishi, and N. Ugawa. 2004. "Landslide Susceptibility Mapping Using GIS-based Weighted Linear Combination, the Case in Tsugawa Area of Agano River, Niigata Prefecture, Japan." *Landslides* 1 (1): 73–81. doi:10.1007/s10346-003-0006-9.

Bamler, R., and P. Hartl. 1998. "Synthetic Aperture Radar Interferometry." *Inverse Problems* 14 (4): R1–R54. doi:10.1088/0266-5611/14/4/001.

Barra, A., L. Solari, M. Béjar-Pizarro, O. Monserrat, S. Bianchini, G. Herrera, M. Crosetto, et al. 2017. "A Methodology to Detect and Update Active Deformation Areas Based on Sentinel-1 SAR Images." *Remote Sensing* 9 (10): 1002. doi:10.3390/rs9101002.

Berardino, P., G. Fornaro, R. Lanari, and E. Sansosti. 2002. "A New Algorithm for Surface Deformation Monitoring Based on Small Baseline Differential SAR Interferograms." *IEEE Transactions on Geoscience and Remote Sensing* 40 (11): 2375–2383. doi:10.1109/TGRS.2002.803792.

Bianchini, S., F. Cigna, G. Righini, C. Proietti, and N. Casagli. 2012. "Landslide HotSpot Mapping by Means of Persistent Scatterer Interferometry." *Environmental Earth Sciences* 67 (4): 1155–1172. doi:10.1007/s12665-012-1559-5.

Botteghi, S., C. Ventisette, D. Montanari, A. Manzella, S. Moretti, D. Dipartimento, D. Scienze, and A. Terra. 2012. *Contribution of Persistent Scatterer Interferometry (Psi) to Map Surface Displacement in the Travale – Radicondoli Geothermal Area (Tuscany, Italy)*. San Francisco, California: Institute of Geosciences and Earth Resources.

Bredal, M., J. Dehls, Y. Larsen, P. Marinkovic, T. ~. R. Lauknes, D. Stødle, and D. ~. A. Moldestad. 2019. "The Norwegian National Ground Motion Service (Insar.no): Service Evolution." *AGU Fall Meeting Abstracts* 2019: G13C–0559.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

Camps-Valls, G., D. Tuia, L. Bruzzone, and J. A. Benediktsson. 2014. "Advances in Hyperspectral Image Classification: Earth Monitoring with Statistical Learning Methods." *IEEE Signal Processing Magazine* 31 (1): 45–54. doi:10.1109/MSP.2013.2279179.

Casu, F., S. Elefante, P. Imperatore, I. Zinno, M. Manunta, C. de Luca, and R. Lanari. 2014. "SBAS-DInSAR Parallel Processing for Deformation Time-Series Computation." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (8): 3285–3296. doi:10.1109/JSTARS.2014.2322671.

Catani, F., P. Canuti, and N. Casagli (2012). "The Use of Radar Interferometry in Landslide Monitoring." *Paper Presented at 1st Meeting OfCold Region Landslides Network and 1st Symposium on Land-Slides in Cold Region, Harbin, China.*

Catani, F., D. Lagomarsino, S. Segoni, and V. Tofani. 2013. "Landslide Susceptibility Estimation by Random Forests Technique: Sensitivity and Scaling Issues." *Natural Hazards and Earth System Sciences* 13 (11): 2815–2831. doi:10.5194/nhess-13-2815-2013.

Ciampalini, A., L. Solari, R. Giannecchini, Y. Galanti, and S. Moretti. 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)." *International Journal of Applied Earth Observation and Geoinformation* 82: 101925. doi:10.1016/j.jag.2019.101925.

Confuorto, P., M. Del Soldato, L. Solari, D. Festa, S. Bianchini, F. Raspini, and N. Casagli. 2021. "Sentinel-1-based Monitoring Services at Regional Scale in Italy: State of the Art and Main Findings." *International Journal of Applied Earth Observation and Geoinformation* 102: 102448. doi:10.1016/j.jag.2021.102448.

Costantini, M., A. Ferretti, F. Minati, S. Falco, F. Trillo, D. Colombo, F. Novali, et al. 2017. "Analysis of Surface Deformations over the Whole Italian Territory by Interferometric Processing of ERS, Envisat and

COSMO-SkyMed Radar Data." *Remote Sensing of Environment* 202: 250–275. doi:10.1016/j.rse.2017.07.017.

Costantini, M., F. Minati, F. Trillo, A. Ferretti, F. Novali, E. Passera, J. Dehls, et al. (2021). "European Ground Motion Service (EGMS)." *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium*, 3293–3296. doi:10.1109/IGARSS47720.2021.9553562

Crosetto, M., O. Monserrat, M. Cuevas-González, N. Devanthéry, and B. Crippa. 2016. "Persistent Scatterer Interferometry: A Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 115: 78–89. doi:10.1016/j.isprsjprs.2015.10.011.

Crosetto, M., L. Solari, J. Balasis-Levinsen, N. Casagli, M. Frei, A. Oyen, and D. A. Moldestad. 2020. "Ground Deformation monitoring at continental scale: The European Ground Motion Service." *The International Archives of The Photogrammetry, Remote Sensing and Spatial Information Sciences* 2020: 293–298. doi:10.5194/isprs-archives-XLIII-B3-2020-293-2020.

Dehls, J. F., Y. Larsen, P. Marinkovic, T. R. Lauknes, D. Stødle, and D. A. Moldestad (2019). "INSAR.No: A National Insar Deformation Mapping/Monitoring Service In Norway – From Concept To Operations." *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan*, 5461–5464. doi: 10.1109/IGARSS.2019.8898614

Del Soldato, M., A. Riquelme, S. Bianchini, R. Tomàs, D. Di Martire, P. de Vita, S. Moretti, and D. Calcaterra. 2018. "Multisource Data Integration to Investigate One Century of Evolution for the Agnone Landslide (Molise, Southern Italy)." *Landslides* 15 (11): 2113–2128. doi:10.1007/s10346-018-1015-z.

de Luca, C., F. Casu, M. Manunta, G. Onorato, and R. Lanari. 2022. "Comments on "Study of Systematic Bias in Measuring Surface Deformation with SAR Interferometry." *IEEE Transactions on Geoscience and Remote Sensing* 60: 1–5. doi:10.1109/TGRS.2021.3103037.

de Luca, C., I. Zinno, M. Manunta, R. Lanari, and F. Casu. 2017. "Large Areas Surface Deformation Analysis through a Cloud Computing P-SBAS Approach for Massive Processing of DInSAR Time Series." *Remote Sensing of Environment* 202: 3–17. doi:10.1016/j.rse.2017.05.022.

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

Fan, H., L. Wang, B. Wen, and S. Du. 2021. "A New Model for three-dimensional Deformation Extraction with Single-track InSAR Based on Mining Subsidence Characteristics." *International Journal of Applied Earth Observation and Geoinformation* 94: 102223. doi:10.1016/j.jag.2020.102223.

Farina, P., D. Colombo, A. Fumagalli, F. Marks, and S. Moretti. 2006. "Permanent Scatterers for Landslide Investigations: Outcomes from the ESA-SLAM Project." *Engineering Geology* 88 (3): 200–217. doi:10.1016/j.enggeo.2006.09.007.

Fay, M. P., and M. A. Proschan. 2010. "Wilcoxon-Mann-Whitney or t-test? on Assumptions for Hypothesis Tests and Multiple Interpretations of Decision Rules." *Statistics Surveys* 4(none. doi:10.1214/09-SS051.

Festa, D., M. Bonano, N. Casagli, P. Confuorto, C. de Luca, M. Del Soldato, R. Lanari, et al. 2022. "Nation-wide Mapping and Classification of Ground Deformation Phenomena through the Spatial Clustering of P-SBAS InSAR Measurements: Italy Case Study." *ISPRS Journal of Photogrammetry and Remote Sensing* 189: 1–22. doi:10.1016/j.isprsjprs.2022.04.022.

Fioraso, G., M. Tararbra, N. Negro, and I. Quaternario. 2020. "La deformazione gravitativa profonda di Sauze d'Oulx (Alpi Occidentali): Analisi dei meccanismi defromativi in relazione all'evoluzione trado quaternaria dell'alta Valle di Susa." *Italian Journal of Quaterny Sciences* 23 (1): 37–54.

Genuer, R., J.-M. Poggi, C. Tuleau-Malot, and N. Villa-Vialaneix. 2017. "Random Forests for Big Data." *Big Data Research* 9: 28–46. doi:10.1016/j.bdr.2017.07.003.

Herrera, G., R. Tomás, J. M. Lopez-Sanchez, J. Delgado, J. J. Mallorqui, S. Duque, and J. Mulas. 2007. "Advanced DInSAR Analysis on Mining Areas: La Union Case Study (Murcia, SE Spain)." *Engineering Geology* 90 (3): 148–159. doi:10.1016/j.enggeo.2007.01.001.

Johnson, J. M., and T. M. Khoshgoftaar. 2019. "Survey on Deep Learning with Class Imbalance." *Journal of Big Data* 6 (1): 27. doi:10.1186/s40537-019-0192-5.

Kalia, A. C., M. Frei, and T. Lege. 2017. "A Copernicus downstream-service for the Nationwide Monitoring of Surface Displacements in Germany." *Remote Sensing of Environment* 202: 234–249. doi:10.1016/j.rse.2017.05.015.

Krishnakumar, V., Z. Qiu, O. Monserrat, A. Barra, J. López-Vinielles, C. Reyes-Carmona, Q. Gao, et al. 2021. "Sentinel-1 A-DInSAR Approaches to Map and Monitor Ground Displacements." *Remote Sensing* 13(6). doi:10.3390/rs13061120.

Lanari, R., M. Bonano, F. Casu, C. D. Luca, M. Manunta, M. Manzo, G. Onorato, and I. Zinno. 2020. "Automatic Generation of Sentinel-1 Continental Scale DInSAR Deformation Time Series through an Extended P-SBAS Processing Pipeline in a Cloud Computing Environment." *Remote Sensing* 12(18). doi:10.3390/rs12182961.

Lary, D. J., A. H. Alavi, A. H. Gandomi, and A. L. Walker. 2016. "Machine Learning in Geosciences and Remote Sensing." *Geoscience Frontiers* 7 (1): 3–10. doi:10.1016/j.gsf.2015.07.003.

Li, C., J. Wang, L. Wang, L. Hu, and P. Gong. 2014. "Comparison of Classification Algorithms and Training Sample Sizes in Urban Land Classification with Landsat Thematic Mapper Imagery." *Remote Sensing* 6 (2): 964–983. doi:10.3390/rs6020964.

Lu, P., N. Casagli, F. Catani, and V. Tofani. 2012. "Persistent Scatterers Interferometry Hotspot and Cluster Analysis (PSI-HCA) for Detection of Extremely slow-moving Landslides." *International Journal of Remote Sensing* 33 (2): 466–489. doi:10.1080/01431161.2010.536185.

Manunta, M., C. de Luca, I. Zinno, F. Casu, M. Manzo, M. Bonano, A. Fusco, et al. 2019. "The Parallel SBAS Approach for

Sentinel-1 Interferometric Wide Swath Deformation Time-Series Generation: Algorithm Description and Products Quality Assessment." *IEEE Transactions on Geoscience and Remote Sensing* 57 (9): 6259–6281. doi:10. 1109/TGRS.2019.2904912.

Massonnet, D., and K. Feigl. 1998. "Radar Interferometry and Its Application to Changes in the Earth's Surface." *Rev. Geophys* 36: 441–500. doi:10.1029/97RG03139.

Milillo, P., G. Giardina, M. J. DeJong, D. Perissin, and G. Milillo. 2018. "Multi-Temporal InSAR Structural Damage Assessment: The London Crossrail Case Study." *Remote Sensing* 10 (2). doi:10.3390/rs10020287.

Montalti, R., L. Solari, S. Bianchini, M. D. Soldato, F. Raspini, and N. Casagli. 2019. "A Sentinel-1-based Clustering Analysis for geo-hazards Mitigation at Regional Scale: A Case Study in Central Italy." *Geomatics, Natural Hazards and Risk* 10 (1): 2257–2275. doi:10.1080/19475705.2019.1690058.

Nava, L., K. Bhuyan, S. R. Meena, O. Monserrat, and F. Catani. 2022. "Rapid Mapping of Landslides on SAR Data by Attention U-Net." *Remote Sensing*, *14*(6) doi:10.3390/ rs14061449.

Ng, A. H.-M., L. Ge, K. Zhang, H.-C. Chang, X. Li, C. Rizos, and M. Omura. 2011. "Deformation Mapping in Three Dimensions for Underground Mining Using InSAR – Southern Highland Coalfield in New South Wales, Australia." *International Journal of Remote Sensing* 32 (22): 7227–7256. doi:10.1080/01431161.2010.519741.

Notti, D., G. Herrera, S. Bianchini, C. Meisina, J. C. García-Davalillo, and F. Zucca. 2014. "A Methodology for Improving Landslide PSI Data Analysis." *International Journal of Remote Sensing* 35 (6): 2186–2214. doi:10.1080/ 01431161.2014.889864.

Novellino, A., M. Cesarano, P. Cappelletti, D. Di Martire, M. Di Napoli, M. Ramondini, A. Sowter, and D. Calcaterra. 2021. "Slow-moving Landslide Risk Assessment Combining Machine Learning and InSAR Techniques." *CATENA* 203: 105317. doi:10.1016/j.catena.2021.105317.

Novellino, A., F. Cigna, M. Brahmi, A. Sowter, L. Bateson, and S. Marsh. 2017. "Assessing the Feasibility of a National InSAR Ground Deformation Map of Great Britain with Sentinel-1." *Geosciences* 7 (2). doi:10.3390/geosciences7020019.

Osmanoğlu, B., T. H. Dixon, S. Wdowinski, E. Cabral-Cano, and Y. Jiang. 2011. "Mexico City Subsidence Observed with Persistent Scatterer InSAR." *International Journal of Applied Earth Observation and Geoinformation* 13 (1): 1–12. doi:10. 1016/j.jag.2010.05.009.

Papoutsis, I., C. Kontoes, S. Alatza, A. Apostolakis, and C. Loupasakis. 2020. "InSAR Greece with Parallelized Persistent Scatterer Interferometry: A National Ground Motion Service for Big Copernicus Sentinel-1 Data." *Remote Sensing*, *12*(19) doi:10.3390/rs12193207.

Pepe, A., and R. Lanari. 2006. "On the Extension of the Minimum Cost Flow Algorithm for Phase Unwrapping of Multitemporal Differential SAR Interferograms." *IEEE Transactions on Geoscience and Remote Sensing* 44 (9): 2374–2383. doi:10.1109/TGRS.2006.873207.

IPCC. 2022. Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-OPörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056. doi:10.1017/9781009325844

Raspini, F., S. Bianchini, A. Ciampalini, M. Del Soldato, L. Solari, F. Novali, S. Del Conte, A. Rucci, A. Ferretti, and N. Casagli. 2018. "Continuous, semi-automatic Monitoring of Ground Deformation Using Sentinel-1 Satellites." *Scientific Reports* 8 (1): 7253. doi:10.1038/s41598-018-25369-w.

Rosi, A., V. Tofani, L. Tanteri, C. Tacconi Stefanelli, A. Agostini, F. Catani, and N. Casagli. 2017. "The New Landslide Inventory of Tuscany (Italy) Updated with PS-InSAR: Geomorphological Features and Landslide Distribution." *Landslides* 15. doi:10. 1007/s10346-017-0861-4.

Samsonov, S., N. d'Oreye, and B. Smets. 2013. "Ground Deformation Associated with post-mining Activity at the French–German Border Revealed by Novel InSAR Time Series Method." *International Journal of Applied Earth Observation and Geoinformation* 23: 142–154. doi:10.1016/j. jag.2012.12.008.

Santos, F. 2020. "Modern Methods for Old Data: An Overview of Some Robust Methods for Outliers Detection with Applications in Osteology." *Journal of Archaeological Science: Reports* 32: 102423. doi:10.1016/j.jasrep.2020.102423.

Scornet, E., G. Biau, and J.-P. Vert. 2015. "Consistency of Random Forests." *The Annals of Statistics* 43 (4): 1716–1741. doi:10.1214/15-AOS1321.

Showstack, R. 2014. "Sentinel Satellites Initiate New Era in Earth Observation." *Eos Trans. AGU* 95 (26): 239.

Solari, L., M. Del Soldato, S. Bianchini, A. Ciampalini, P. Ezquerro, R. Montalti, F. Raspini, and S. Moretti. 2018. "From ERS 1/2 to Sentinel-1: Subsidence Monitoring in Italy in the Last Two Decades." *Frontiers in Earth Science* 6. doi:10.3389/feart.2018. 00149.

Solari, L., M. D. Soldato, R. Montalti, S. Bianchini, F. Raspini, P. Thuegaz, D. Bertolo, V. Tofani, and N. Casagli. 2019. "A Sentinel-1 Based hot-spot Analysis: Landslide Mapping in north-western Italy." *International Journal of Remote Sensing* 40 (20): 7898–7921. doi:10.1080/01431161.2019. 1607612.

Stramondo, S., M. Saroli, C. Tolomei, M. Moro, F. Doumaz, A. Pesci, F. Loddo, P. Baldi, and E. Boschi. 2007. "Surface Movements in Bologna (Po Plain — Italy) Detected by Multitemporal DInSAR." *Remote Sensing of Environment* 110 (3): 304–316. doi:10.1016/j.rse.2007.02.023.

Tarquini, S., I. Isola, M. Favalli, and A. Battistini. 2007. *TINITALY, a Digital Elevation Model of Italy with a 10 Meters Cell Size (Version 1.0)*. Istituto Nazionale Di Geofisica e Vulcanologia (INGV). *[Data set]*.

Tomás, R., J. I. Pagán, J. A. Navarro, M. Cano, J. L. Pastor, A. Riquelme, M. Cuevas-González, et al. 2019. "Semi-Automatic Identification and Pre-Screening of Geological–

Geotechnical Deformational Processes Using Persistent Scatterer Interferometry Datasets." *Remote Sensing* 11 (14). doi:10.3390/rs11141675.

Torres, R., P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, et al. 2012. "GMES Sentinel-1 Mission." *Remote Sensing of Environment* 120: 9–24. doi:10.1016/j.rse.2011.05.028.

Trigila, A., C. Iadanza, and L. Guerrieri. 2007. "The IFFI Project (Italian Landslide Inventory): Methodology and Results." In *Guidelines for Mapping Areas at Risk of Landslides in Europe*, edited by J. Hervás, 15-18. Rome, Italy: ISPRA.

Weiss, J. R., R. J. Walters, Y. Morishita, T. J. Wright, M. Lazecky, H. Wang, E. Hussain, et al. 2020. "High-Resolution Surface Velocities and Strain for Anatolia from Sentinel-1 InSAR and GNSS Data." *Geophysical Research Letters* 47 (17): e2020GL087376. doi:10.1029/2020GL087376.

Zinno, I., S. Elefante, L. Mossucca, C. de Luca, M. Manunta, O. Terzo, R. Lanari, and F. Casu. 2015. "A First Assessment of the P-SBAS DInSAR Algorithm Performances within A Cloud Computing Environment." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (10): 4675–4686. doi:10.1109/JSTARS.2015.2426054.