

Wide-Area Supervised Classification of Ground Deformation Phenomena from European Ground Motion Service Products

Riccardo PALAMÀ^{1,*}, Anna BARRA¹, María CUEVAS-GONZÁLEZ¹, Kamila PAWŁUSZEK-FILIPIAK², José Antonio NAVARRO¹, Oriol MONSERRAT¹, Michele CROSETTO¹

¹ *Geomatics Research Unit, Centre Tecnològic de Telecomunicacions de Catalunya, Castelldefels, Barcelona, Spain, (rpalama@cttc.cat)*

² *Institute of Geodesy and Geoinformatics, Wrocław University of Environmental and Life Sciences, Wrocław, Poland*

**corresponding author*

Abstract

This work proposes a supervised classifier of ground motion (GM) phenomena using as main input the European Ground Motion Service (EGMS) datasets. The availability of such an extended dataset allows implementing wide area tools to detect and classify GM phenomena, that can be useful for potential users to evaluate hazard and mitigate risks. This work proposes a wide-area ground motion classifier (GMC) that categorizes areas affected by GM phenomena into three main classes, i.e. deep-seated gravitational slope deformation (DSGSD), landslides and subsidence. The implementation of the classifier is preceded by the identification of active deformation areas (ADAs) through the ADAfinder tool. The Extreme Gradient Boosting (XGB) technique was selected for this classification problem. The result of this work is a European map of ADAs, classified into the above-mentioned deformation classes.

Keywords: SAR Differential Interferometry (DInSAR), Ground Motion Classification, Extreme Gradient Boosting

Received: 9th December 2024. Revised: 20th February 2025. Accepted: 25th February 2025.

1 Introduction

Monitoring ground deformation represents a critical task for mitigating risks to infrastructure and the built environment, which has been tackled in recent years by different remote sensing techniques. Among the technologies employed, both spaceborne and terrestrial Synthetic Aperture Radar (SAR) imagery, processed through Persistent Scatterer SAR Interferometry (PSInSAR), has achieved significant performance (Crosetto et al. 2016; Ferretti, Prati, and Rocca 2001). PSInSAR measures the displacement of strong radar reflectors, i.e. persistent scatterers (PS), on the radar line-of-sight, by processing the SAR interferometric phases. This technique has evolved in the recent decades, achieving high performance both in terms of coverage and accuracy, through advanced approaches (Ferretti et al. 2011; Pepe et al. 2015), allowing the generation of regional and national datasets (Confuorto et al.

2023). PSInSAR was employed to produce a monitoring service of GM phenomena over the European territory, the European Ground Motion Service (EGMS). This wide-area dataset, which enables the creation of automated tools for detecting and classifying GM events, is the most recent addition to the product portfolio of the Copernicus Land Monitoring Service. The Service is funded by the European Commission in the frame of the Copernicus Programme and is implemented under the responsibility of the European Environment Agency (Crosetto et al. 2020). This work performs a wide-area ground motion classifier (GMC) that categorizes areas affected by GM phenomena into three main classes, i.e. slow-moving slope deformation phenomena (mostly represented by deep-seated gravitational slope deformation phenomena, DSGSD), landslides and subsidence. The Italian National Landslide Inventory is employed to derive DSGSD and landslide labels, the subsidence map of Emilia-Romagna region (Italy) and mining-

related subsidence in Poland for the subsidence label (Palamà et al. 2022; Pawluszek-Filipiak and Borkowski 2020). Specifically, the DSGSD phenomena consist of mountain slope deformation of steep, high mountain slopes, with slow movement rates (Ambrosi et al 2006, Hungr et al 2014, and has relevant differences with the rotational/translational landslide class – showing faster rates - that is also considered in this work. However, DSGSDs may evolve into faster phenomena, which often results in common properties between these two classes. It should be noted that volcanic and earthquake-related deformations are not considered in this work.

This approach tackles a supervised feature-based classification task, which is complicated by the presence of missing values, such as the absence of Persistent Scatterer points in ADA polygons for one of the two Sentinel-1 orbit trajectories. To address this, the Extreme Gradient Boosting (XGB) algorithm is employed, chosen for its ability to handle incomplete datasets and its strong performance in similar machine learning applications. XGB is an evolution of traditional decision trees, which sequentially adds in order to minimise a chosen loss function (Chen and Guestrin 2016). Once trained, the classification algorithms were tested on the test set and their performance compared. For the final deployment of a user-level product, the trained model is launched on the unlabelled dataset to produce a new global classified dataset. The importance of the classification features is also studied, in order to gain insight about the performed classification. The contributions of this work can be summarized as (i) the implementation of a classifier of ground deformation phenomena working for the whole European territory, combining PSInSAR (EGMS) data, DEM and Land-Cover, preceded by (ii) the systematic derivation of a training dataset for such classifier using labelling data available from landslide and subsidence inventories. The GM classification aims at providing interested users with a dataset of GM phenomena that can be used for an early identification of areas at risk. Furthermore, we propose a GM classifier framework with high generalizability, using input data sources (EGMS, DEM and Land cover maps), that are available over wide areas.

This work is organized as follows: section 2 illustrates the employed datasets, section 3 addresses the pre-processing stages, i.e. ADA extraction and training dataset preparation, whereas the classification algorithm is described in

section 4. The main results are illustrated in section 5, and conclusions are drawn in section 6.

2 Dataset description

The main input of the GMC consists of the EGMS Basic Product data, derived from Sentinel-1 data, collected from ascending and descending orbit trajectories. These maps are combined with secondary inputs, i.e. a European Digital Elevation Model (EuDEM) and the land cover map (Corine CLC). On the other hand, available inventories of landslide and subsidence phenomena were employed for the preparation of the training dataset. In this implementation, the training dataset was prepared considering GM phenomena in the Italian and Polish territories. The Italian National Landslide Inventory (IFFI) was used to label the DSGSD and landslide ADAs in Italy, whereas the subsidence ones were produced from the subsidence inventory of the Emilia-Romagna region (Bitelli et al. 2015; Trigila, Iadanza, and Spizzichino 2010).

On the other hand, most of the labelled ADAs in Poland concern subsidence phenomena induced by underground mining activity, which were collected from previous work, e.g. those concerning the Legnica-Glogow copper mining areas and the Silesian coal basin, and included within the subsidence training data. Furthermore, a little portion of landslide ADAs were labelled in Poland, mostly located in the Carpathian Mountains area.

3 Pre-processing

3.1 Extraction of Active Deformation Areas

The ADAfinder tool was developed (Barra et al. 2017; Ezquerro et al. 2020; Navarro et al. 2020) with the aim of easing the management, use and interpretation of PSInSAR results, consisting of an ADA detection algorithm based on few spatial and statistical parameters of the pixel displacement time series. The ADA detector first removes outliers and isolated PS points, then a velocity threshold is applied to eliminate points that are considered as stable. In this work we set the value of this threshold at 5 mm/year, considering that the average noise level for the velocity values of the EGMS Basic product is about 2 mm/year. Then, the detected points whose distance is lower than 40 m are grouped together into one polygon defining a new ADA. The final stage computes a quality

index (QI) for each detected ADA, with values ranging from 1 (reliable ADAs) to 4 (very noisy ADAs). The QI values are computed accounting for the spatio-temporal correlation properties of the displacement values associated with the points forming the ADA under analysis. In this work, only the ADAs whose QI is equal to 1 or 2 are considered. The ADAfinder output consists of polygons associated with the detected ADAs, together with their QI, and various relevant statistical parameters (mean, maximum and minimum displacement velocity, number of PS points), together with the displacement time series of the measurement points (MPs) within the polygon.

The ADAfinder tool was applied to each burst of the EGMS Basic product. The total processing time was about 48 hours. We observe that the bursts of the EGMS Basic product are associated with the burst of Sentinel-1 data, separately for the ascending and descending orbit trajectories. The detected ADAs in both trajectories were finally merged to generate a single database of detected ground deformation areas associated with the Sentinel-1 ascending and descending orbit trajectories.

3.2 Training dataset preparation

The following preparation step creates a uniformized GMC dataset of ADAs from the orbit-wise European ADA maps. Such dataset will be the basis to train and test ML classifiers, thus for each polygon it contains geographical data and specific classification features. The creation of the GMC dataset starts from the European ADA maps for the ascending and descending LOS. It should be noted that each orbit-wise ADA map contains the ADAs associated with all the tracks of one Sentinel-1 orbit trajectory. The intersecting ascending/descending ADAs are merged to form one ADA of the GMC dataset (which will be referred to as global ADA), whereas the ADAs that have no intersections are directly included in the GMC dataset. The global ADAs are then labelled using the ground-truth datasets described in Section 2, and collected into the GMC training dataset, whereas the remaining ADAs are collected into the unlabelled GMC dataset (unseen data). Assigning a GM class to the unlabelled ADAs is the final objective of this work. Following the creation of the dataset, classification features are computed for each global ADA. The EGMS MPs contained within each global ADA polygon are extracted and their features computed from the PS

point metrics in the EGMS dataset (i.e. velocity, acceleration, seasonality). In the case of no intersection, the global ADA polygon obtained from a single Sentinel-1 orbit is used to extract the PS points of the opposite orbit associated with that global ADA, and their metrics added to the feature vector. If the number of opposite-orbit points is smaller than an acceptable limit (set as 3), the feature values for that orbit are marked as missing values. This procedure, illustrated in Figure 1, allows combining, where possible, the information associated with both Sentinel-1 orbit trajectories. On the other hand, if one ADA is associated with only one orbit trajectory, part of its classification features will be missing, which is, however, tackled by using classification algorithms that cope with missing values. We observe that employing EGMS Basic product, derived from Sentinel-1 ascending and descending orbit data, allows exploiting the full resolution of Sentinel-1 SAR images, whose pixel spacing is about 4m and 14m in range and azimuth, respectively.

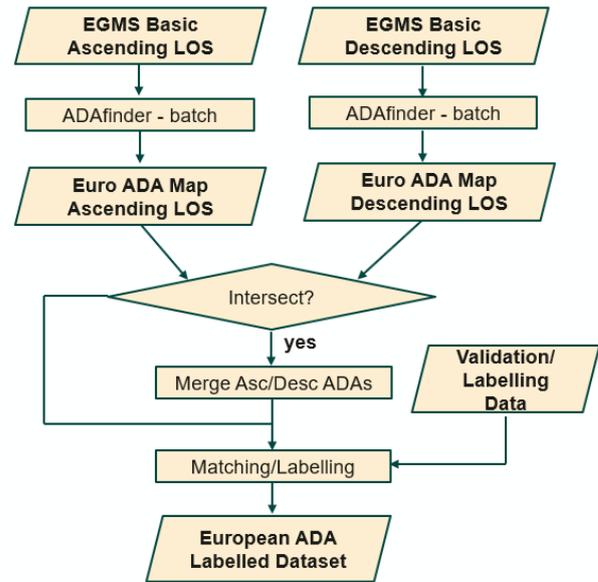


Figure 1. Scheme of the training dataset preparation routine

The feature vector structure is summarized in Table 1. The features associated with PS data are doubled to consider the ascending and descending Sentinel-1 orbit trajectories, yielding a feature vector of 21 features. The training dataset (Figure 2) contains 15898 ADAs, and the numbers of ADAs per class are as follows: 2169 (DSGSD), 6179 (landslide), 5460 (subsidence). This dataset will be split into training (80%), validation (10%) and test (10%) sets in the implementation and performance evaluation stages of the classifier (see section 5).

Table 1. Feature vector structure for the GMC

Feature Name	Description
<i>MnVel</i>	Mean velocity (PS)
<i>MnVStd</i>	Mean standard deviation of velocity (PS)
<i>MnAcc</i>	Mean acceleration (PS)
<i>MnAStd</i>	Mean standard deviation of acceleration (PS)
<i>MnSeas</i>	Mean seasonality (PS)
<i>MnDA</i>	Mean dispersion of amplitude (PS)
<i>MnTCoh</i>	Mean temporal coherence (PS)
<i>Mean_DEM</i>	Mean DEM value
<i>Std_DEM</i>	Standard deviation of DEM
<i>Mean_Slope</i>	Mean slope value
<i>Std_Slope</i>	Standard deviation of slope
<i>Mean_Aspect</i>	Mean value of the aspect angle
<i>Std_Aspect</i>	Standard deviation of the aspect angle values
<i>LandCover</i>	Dominant land cover/use class within the ADA

4 Ground-motion classifier algorithm

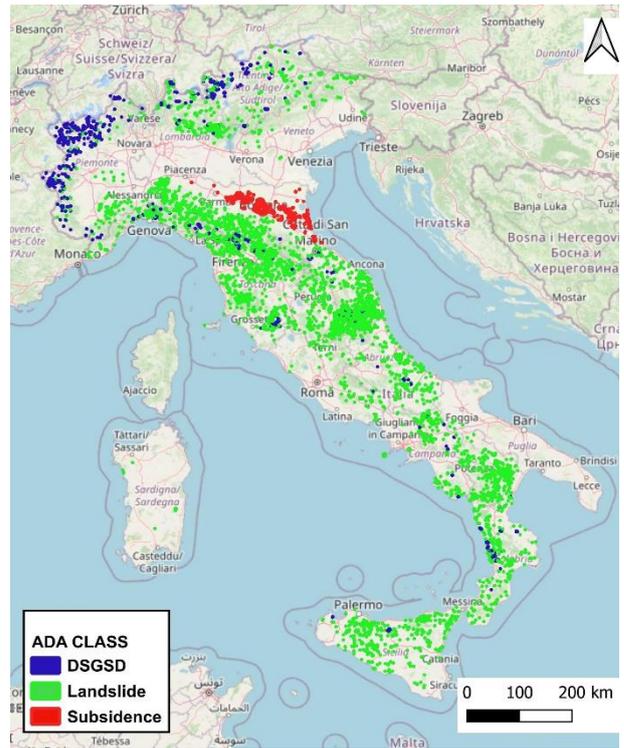
The GMC was implemented by the XGB algorithm, which is an ensemble technique that builds models sequentially, by progressively combining base learners. XGB uses decision trees as base learners and includes regularization terms to penalize the complexity of the model, avoid overfitting and ensure generalization (Chen and Guestrin 2016; Prokhorenkova et al. 2018). The XGB operation can be summarized by:

1. *Initialization*: set a simple initial model, which predicts the mean of the target variable.

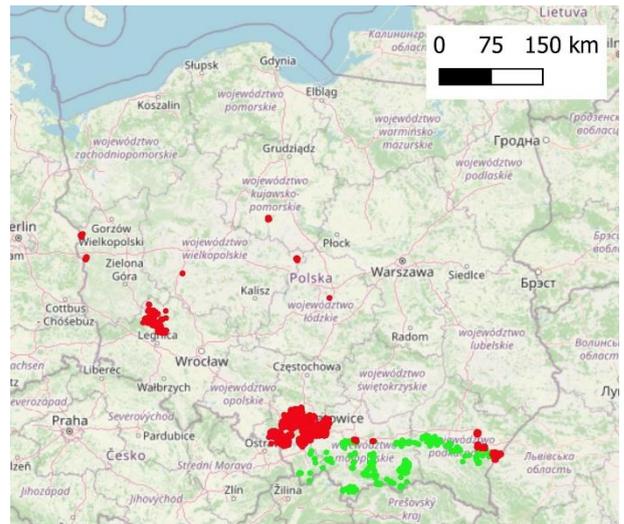
2. *Adding trees sequentially*, where the residuals are computed for each iteration. The residuals are represented by the gradients of the loss function with respect to the current predictions. A new tree that fits these residuals is added, predicting the gradient, with the aim of reducing the errors made by the current model.

3. *Learning objective and tree pruning*. The objective function to optimize is a combination of a loss function (*softmax* for multi-class classification) and both L1 and L2 regularization terms (respectively indicated as *alpha* and *lambda*). Tree pruning is performed by setting the *maximum depth* (i.e. the maximum number of levels of a tree) and the *minimum child weight* (i.e. the minimum sum of instance weights of a child).

5. *Shrinkage*. After adding a new tree, XGB weights the predictions (i.e. class scores) by a



(a)



(b)

Figure 2. GM Labeled dataset in Italy (a), Poland (b)

shrinkage factor (*learning rate*), such that subsequent trees make only small adjustments, which aims at improving the model robustness.

6. *Output*. The final prediction for one data sample is the sum of the initial prediction and the weighted scores of all the trees that are progressively added. The output scores for each class are transformed into probability values using the *softmax* function. XGB provides a measure of feature importance based on the number of times a feature is used to split the data across all the trees and the relative gain in the loss function attributed

to each split. This information summarizes the impact of each feature on the ground motion classification, which is intended to introduce the concept of explainable artificial intelligence (XAI) in the study of GM phenomena from SAR data. The main XGB hyperparameters (*learning rate*, *min child weight*, *maximum depth*, *alpha*, *lambda*) were set by a tuning routine based on a Bayesian optimization framework (Shahriari et al. 2016). The hyperparameter tuning launched 100 iterations, employing the accuracy as optimization metric, resulting in the following optimal hyperparameter values: *learning_rate* = 0.089, *max_depth* = 8, *min_child_weight* = 0, *alpha* = 0, *lambda* = 0.787.

5 Results

5.1 Test dataset

The GMC classifier was implemented on the Italy&Poland dataset, using three classes, i.e. landslide, DSGSD and subsidence (that is including also the underground mining ADAs). The training dataset was split into training (80%), validation (10%) and test (10%) sets. The results are computed on the test set and are shown in the confusion matrix (Table 2) and classification performance metrics (Table 3). The results show good performance, with a misclassification affecting DSGSD and landslide classes. Furthermore, we observe that for the landslide and subsidence classes the values of the metrics are always greater than 90%, with very high values associated to the subsidence class. This good performance seems to confirm that the hypothesis of the underground mining class being included into the subsidence class is correct. On the other hand, landslide and DSGSD classes have similar characteristics (occurring in areas with non-flat topography, i.e. mountainous and hilly environments) and they are often spatially close, as DSGSD are likely to evolve into landslides. This results in lower accuracy values for both classes and higher number of false positives.

Furthermore, with the aim of exploring the explainability of the classification algorithm, we have evaluated the feature importance (I_F) values, revealing that the more relevant features are DEM ($I_F \sim 22\%$), slope ($I_F \sim 20\%$), mean velocity of the PS displacement time series ($I_F \sim 18\%$ for both trajectories). In particular, DEM and slope features are better suited to identify the subsidence class, whereas discriminating between DSGSD and

landslide classes is enabled by the PS velocity and acceleration features.

Table 1. Confusion matrix – GM classification

	Landslide	DSGSD	Subsidence
Landslide	2478	86	49
DSGSD	234	425	1
Subsidence	15	1	4229

Table 2. Performance metrics – GM classification

GM class	precision	recall	F1-score
Landslide	0.91	0.95	0.93
DSGSD	0.83	0.64	0.73
Subsidence	0.99	0.99	0.99
Average	0.91	0.86	0.88

5.2 Deployment

The GMC model derived and trained as described in section 5.1 was deployed to the unlabeled ADAs over the whole European territory, producing a European map of ADAs classified into three classes of GM phenomena, i.e. DSGSD, landslide and subsidence, shown in Figure 3. We observe that subsidence ADAs are located in coastal and flat topography areas, such as the territory of northern France, Belgium, Netherlands, northern Germany, Poland, the Baltic countries and the United Kingdom. On the other hand, landslide ADAs are mostly present in areas with significant slopes, i.e. the Alpine arc, the Apennine mountains (Italy), Carpathian and Scandinavian mountains. Finally, DSGSD ADAs are mostly located in areas with very high slopes and altitudes (Alps and Pyrenees). The number of occurrences per class are as follows: 194,627 subsidence ADAs, 272,563 landslide ADAs, 17,432 DSGSD ADAs.

The GMC was implemented in Python language and launched on a 48 CPUs machine. The amount of time to train and test the GMC was about 3 hours.

6 Conclusions

This paper presents a ground motion classification technique applied to the European Ground Motion Service data, enriched by DEM, Slope and Land Cover data. The results show a distribution of landslide, subsidence and DSGSD ADAs that finds justification from the topography of the European territory. Future work will add further deformation classes to the GMC.

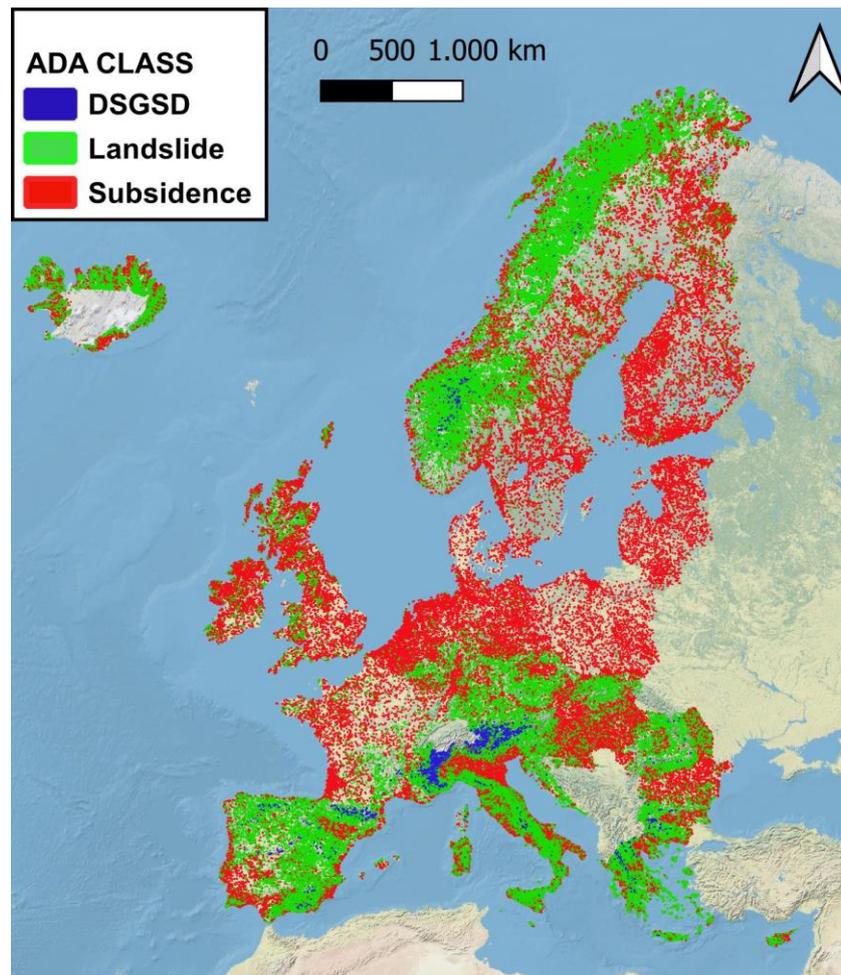


Figure 3 . Ground motion classification results over the European territory (covered by EGMS), as result of the deployment of the ground motion classifier to unseen data, i.e. unlabelled active deformation areas.

Acknowledgements

The authors would like to thank the European Environment Agency for providing the EGMS data. This work has been supported by the European Space Agency under the Living Planet Fellowship Scheme.

References

- Ambrosi, C., and Crosta, G.B. (2006), Large sacking along major tectonic features in the Central Italian Alps, *Engineering Geology*, Volume 83, Issues 1–3, 2006, pp. 183-200,
- Barra, A., Solari, L., Béjar-Pizarro, M., Monserrat, O., Bianchini, S., Herrera, G., Crosetto, M., Sarro, R., González-Alonso, E., Mateos, R.M., Ligüerzana, S., López, C., Sandro Moretti, S. (2017). “A Methodology to Detect and Update Active Deformation Areas Based on Sentinel-1 SAR Images”, *Remote Sensing*, 9(10):1002.
- Bitelli, G., Bonsignore F., Pellegrino, I., Vittuari, L., (2015). “Evolution of the Techniques for Subsidence Monitoring at Regional Scale: The Case of Emilia-Romagna Region (Italy).” *Proceedings of IAHS* 372:315–21.
- Chen, T., and Guestrin, C. (2016). “XGBoost: A Scalable Tree Boosting System.” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 13-17-August-2016:785–94.
- Confuorto, P., Casagli, N., Casu, F., De Luca, C., Del Soldato, M., Festa, D., Lanari, R., Manzo, M., Onorato, O., Raspini, F. (2023). “Sentinel-1 P-SBAS Data for the Update of the State of Activity of National Landslide Inventory Maps”, *Landslides* 20(5):1083–97.
- Crosetto, M, Monserrat, O., Cuevas-González, M., Devanthery, N., Crippa, B., (2016). “Persistent Scatterer Interferometry: A Review.” *ISPRS*

- Journal of Photogrammetry and Remote Sensing* 115:78–89.
- Crosetto, M., Solari, L., Mróz, M., Balasis-Levinsen, J., Casagli, N., Frei, M., Oyen, A., Anders Moldestad, D., Bateson, L., Guerrieri, L., Comerci, V., Steen Andersen, H., (2020). “The Evolution of Wide-Area DInSAR: From Regional and National Services to the European Ground Motion Service”, *Remote Sensing*, 12(12):2043.
- Ezquerro, P., Tomás, R., Béjar-Pizarro, M., Fernández-Merodo, J.A., Guardiola-Albert, C., Staller, A., Sánchez-Sobrino, J. A., Herrera, G., (2020). “Improving Multi-Technique Monitoring Using Sentinel-1 and Cosmo-SkyMed Data and Upgrading Groundwater Model Capabilities.” *Science of the Total Environment* 703:1–33.
- Ferretti, A., Fumagalli, A., Novali, F., Prati, C., Rocca, F., Rucci, A., (2011), “A New Algorithm for Processing Interferometric Data-Stacks: SqueeSAR”, *IEEE Transactions on Geoscience and Remote Sensing* 49(9):3460–70.
- Ferretti, A., Prati, C., and Rocca, F., (2001). “Permanent Scatterers in SAR Interferometry.” *IEEE Transactions on Geoscience and Remote Sensing* 39(1):8–20.
- Hungr, O., Leroueil, S. and Picarelli, L., (2014), The Varnes classification of landslide types, an update, *Landslides*, 11, 167–194.
- Navarro, J. A., Tomás, R., Barra, A., Pagán, J. I., Reyes-Carmona, C., Solari, L., Lopez Vinielles J., Falco, S., Crosetto, M., (2020). “ADAtools: Automatic Detection and Classification of Active Deformation Areas from PSI Displacement Maps”, *ISPRS International Journal of Geo-Information* 9(10):1–26.
- Palamà, R., Crosetto, M., Rapinski, J., Barra, A., Cuevas-González, M., Monserrat, O., Crippa, B., Kotulak, N., Mróz, M., Mleczko, M., (2022). “A Multi-Temporal Small Baseline Interferometry Procedure Applied to Mining-Induced Deformation Monitoring”, *Remote Sensing* 14(9).
- Pawluszek-Filipiak, K., Borkowski, A., (2020). “Integration of DInSAR and SBAS Techniques to Determine Mining-Related Deformations Using Sentinel-1 Data: The Case Study of Rydułtowy Mine in Poland”, *Remote Sensing* 12(2,242).
- Pepe, A., Yang Y., Manzo, M., Lanari, R., (2015). “Improved EMCF-SBAS Processing Chain Based on Advanced Techniques for the Noise-Filtering and Selection of Small Baseline Multi-Look DInSAR Interferograms”, *IEEE Transactions on Geoscience and Remote Sensing* 53(8):4394–4417.
- Prokhorenkova, L., Gusev G., Vorobev, A., Dorogush A.V., Gulin, A. (2018). “Catboost: Unbiased Boosting with Categorical Features.” *Advances in Neural Information Processing Systems*, 2018-December (Section 4):6638–48.
- Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., De Freitas, N. (2016). “Taking the Human out of the Loop: A Review of Bayesian Optimization.” *Proceedings of the IEEE* 104(1):148–75.
- Trigila, A., Iadanza, C., Spizzichino, D., (2010), “Quality Assessment of the Italian Landslide Inventory Using GIS Processing.” *Landslides* 7(4):455–70.