Article

A Clustering Approach for the Analysis of InSAR Time Series: Application to the Bandung Basin (Indonesia)

Michelle Rygus [1,*], Alessandro Novellino [2], Ekbal Hussain [2], Fifik Syafiudin [3], Heri Andreas [4] and Claudia Meisina [1]

1   Department of Earth and Environmental Sciences, University of Pavia, Via Adolfo Ferrata 1, 27100 Pavia, Italy
2   British Geological Survey, Keyworth, Nottingham NG12 5GG, UK
3   Geospatial Information Agency of Indonesia (Badan Informasi Geospasial), Jl. Ir. H. Juanda No. 193, Dago, Kecamatan Coblong, Kota Bandung 40135, Indonesia
4   Department of Geodesy and Geomatics Engineering, Institute of Technology Bandung, Jalan Ganesha 10, Bandung 40132, Indonesia
*   Correspondence: michelle.rygus01@universitadipavia.it

Abstract: Interferometric Synthetic Aperture (InSAR) time series measurements are widely used to monitor a variety of processes including subsidence, landslides, and volcanic activity. However, interpreting large InSAR datasets can be difficult due to the volume of data generated, requiring sophisticated signal-processing techniques to extract meaningful information. We propose a novel framework for interpreting the large number of ground displacement measurements derived from InSAR time series techniques using a three-step process: (1) dimensionality reduction of the displacement time series from an InSAR data stack; (2) clustering of the reduced dataset; and (3) detecting and quantifying accelerations and decelerations of deforming areas using a change detection method. The displacement rates, spatial variation, and the spatio-temporal nature of displacement accelerations and decelerations are used to investigate the physical behaviour of the deforming ground by linking the timing and location of changes in displacement rates to potential causal and triggering factors. We tested the method over the Bandung Basin in Indonesia using Sentinel-1 data processed with the small baseline subset InSAR time series technique. The results showed widespread subsidence in the central basin with rates up to 18.7 cm/yr. We identified 12 main clusters of subsidence, of which three covering a total area of 22 km^2 show accelerating subsidence, four clusters over 52 km^2 show a linear trend, and five show decelerating subsidence over an area of 22 km^2. This approach provides an objective way to monitor and interpret ground movements, and is a valuable tool for understanding the physical behaviour of large deforming areas.

Keywords: land subsidence; InSAR; time series analysis; clustering; Bandung

1. Introduction

Interferometric Synthetic Aperture Radar (InSAR) is a commonly used technique for measuring surface deformation with millimetric accuracy [1]. The output of InSAR processing is a collection of time series representing the ground displacements of Measurement Points (MPs) on the surface during the observed period. InSAR has been used for monitoring natural hazards such as landslides, earthquakes, volcanic activity, and ground subsidence [2], as well as for monitoring man-made structures such as dams, buildings [3], and bridges [4], and anthropogenic activities such as the pumping or injection of fluids. Under certain geological conditions, if groundwater abstraction exceeds groundwater recharge for a long duration over a large area, the subsurface can compact and result in land subsidence. In several areas around the world where the subsurface materials are composed of relatively recent alluvial, marine, or lacustrine deposits composed of alternating coarse-grained water-bearing strata with fine-grained compressible layers, land subsidence has been generated by unsustainably extracting groundwater from underlying

aquifers [5]. In such conditions, InSAR has been used as a tool for monitoring groundwater storage changes and analyzing land subsidence or uplift in large urban areas caused by groundwater extraction or injection [6–9].

InSAR analysis of a stack of synthetic aperture radar (SAR) images provides a one-dimensional projection of displacement along the Line of Sight (LOS) of a satellite; however, the actual surface motions resulting from subsurface deformation processes typically occurs in three dimensions (i.e., east-west, north-south, and up-down) [10]. Consequently, interpreting and conveying LOS measurements to stakeholders who are unfamiliar with the notion of a viewing geometry confined to a single dimension can be challenging [11]. As a solution, some studies have simply interpreted InSAR LOS solely as vertical deformation and assumed the horizontal component to be negligible [12,13], which could lead to inaccuracies in the interpretation of the data if a horizontal component exists in the original LOS dataset. To overcome this constraint, when displacement measurements from both the ascending and descending geometries are available, it has become increasingly common in InSAR studies to combine the two viewing geometries to extract the east-west and vertical components of motion [14–16]. Performing this task allows for better estimation of the magnitude and direction of surface motions [17].

Recently, the combination of an increase in the volume of SAR data, driven by the development of new satellite missions with high temporal coverage, and the improvements in InSAR processing and satellite data storage capabilities, has allowed scientists to move from using InSAR data from opportunistic science to routine monitoring [18]. These changes have led to an increase in the amount of information that can be extracted from these datasets and have resulted in a more comprehensive understanding of the evolution of the Earth's surface and subsurface processes. However, until a few years ago InSAR interpretation was mainly limited to analysis of the average displacement rates [19], but advances in innovative big data analysis methods change this and enable exploitation of the full displacement time series [20–22]

In the case of large InSAR datasets, traditional manual analysis is a complex and time-consuming process and more automated techniques are necessary. Previous work has used a variety of methods to help interpret InSAR time series, ranging from semi-automatic [21,23,24] and automatic statistical approaches [25], to the use of supervised [26,27] and unsupervised [20,28,29] machine learning algorithms. These studies have aggregated MPs based on their average velocities [27] but inevitably fail to identify non-linear movements. Exploiting the full displacement time series enables us to better explore the InSAR dataset and gain a much deeper understanding of the deformation kinematics: for example, how landslide velocity responds to rainfall [30]. Additionally, clustering has been extended to the spatial distribution of principal components extracted from InSAR time series datasets [6,31]. However, the Principal Component Analysis used to estimate the principal components assumes linearity in the data which may not be appropriate for analyzing large, complex InSAR datasets since non-linear data relationships would be neglected.

To overcome these limitations, this paper presents a novel InSAR data mining approach that exploits displacement time series to gain a more comprehensive understanding of the deformation patterns over a large area. We implemented an unsupervised clustering technique to identify patterns and features in the data that may not be visible by traditional clustering methods: our approach involves the use of Uniform Manifold Approximation and Projection (UMAP) [32] for dimensionality reduction followed by the Hierarchical Density-based Spatial Clustering of Applications with Noise (HDBSCAN) [33] algorithm. Our study expands on a recent application of this approach [29] in two ways: (1) combining both ascending and descending satellite observation geometries to retrieve the vertical and east-west displacement time series to get a more accurate estimation of the vertical movement, and (2) introducing a change detection method to determine when and in which manner the deformation trend changes over time to aid in further interpretation of the data. It is often the case that a single linear model cannot adequately describe the

evolution of displacement over time; instead, multiple linear/polynomial relationships applying to different time spans may be more appropriate [34–37].

We applied our time series data mining method to an InSAR data stack over the Bandung Basin in Indonesia where land subsidence is an ongoing phenomenon [38–42]. A deeper understanding behind dynamics of the ground movement is required to disentangle the contribution of triggering factors throughout the basin. Improved knowledge of the causal factors will support local authorities (e.g., Indonesia's Ministry of Public Works and Housing) in the management and prevention of the associated risks such as increasing flooding and damage to buildings, bridges, and roads [38,39]. The detailed abbreviations and definitions used in the paper are listed in Table 1.

Table 1. List of abbreviations used in this paper.

Abbreviation	Explanation
DBCV	Density-Based Clustering Validation
ESDM	Office of Mineral and Energy Resources
GNSS	Global Navigation Satellite System
HDBSCAN	Hierarchical Density-Based Clustering
InSAR	Interferometric Synthetic Aperture Radar
K-S test	Kolmogorov–Smirnov test
LOS	Line of Sight
MintPy	Miami INsar Time-series software in Python
MPs	Measuring Points
PCA	Principal Component Analysis
SAR	Synthetic Aperture Radar
UMAP	Uniform Manifold Approximation and Projection

2. Bandung Basin

As a rapidly growing urban region, the Bandung metropolis in West Java, Indonesia is home to over eight million inhabitants and is one of the most important centres for political, economic, and social activity in Indonesia [40]. In recent decades, this area has experienced a rapid increase in population and urban expansion which has been driven by growth of its industrial sector [40]. The city lies within the Bandung Basin which is a large intra-montane (Figure 1) basin encircled by a range of mountains and volcanic highlands covering an area of 2300 km^2 where the Citarum River along with its tributaries forms the main drainage system of the basin catchment. Deposits in the basin consist of a notably thick series of Holocene age lake deposits consisting of uncompacted claystone, siltstone, and sandstone belonging to the Kosambi Formation, Late Pleistocene-Holocene age volcanic breccias and tuffs of the Cibeureum Formation, and the Cikapundung Formation which consists of Early Pleistocene conglomerate and compact breccias, tuff, and andesitic lava [43]. The unconsolidated Holocene clayey lake deposits render the area prone to land subsidence, which previous InSAR studies attributed to excessive groundwater pumping from a deep, confined aquifer [41–44]. Additional factors contributing to the subsidence are the compaction of alluvium soil [45], settlement of highly compressible soils induced by surficial loads such as new residential settlements, or the development of industrial buildings [39,42].

Considering the dynamic setting of the Bandung area and the multitude of factors that can contribute to subsidence movements at various rates and patterns, this location offers a great opportunity to test our data analysis approach.
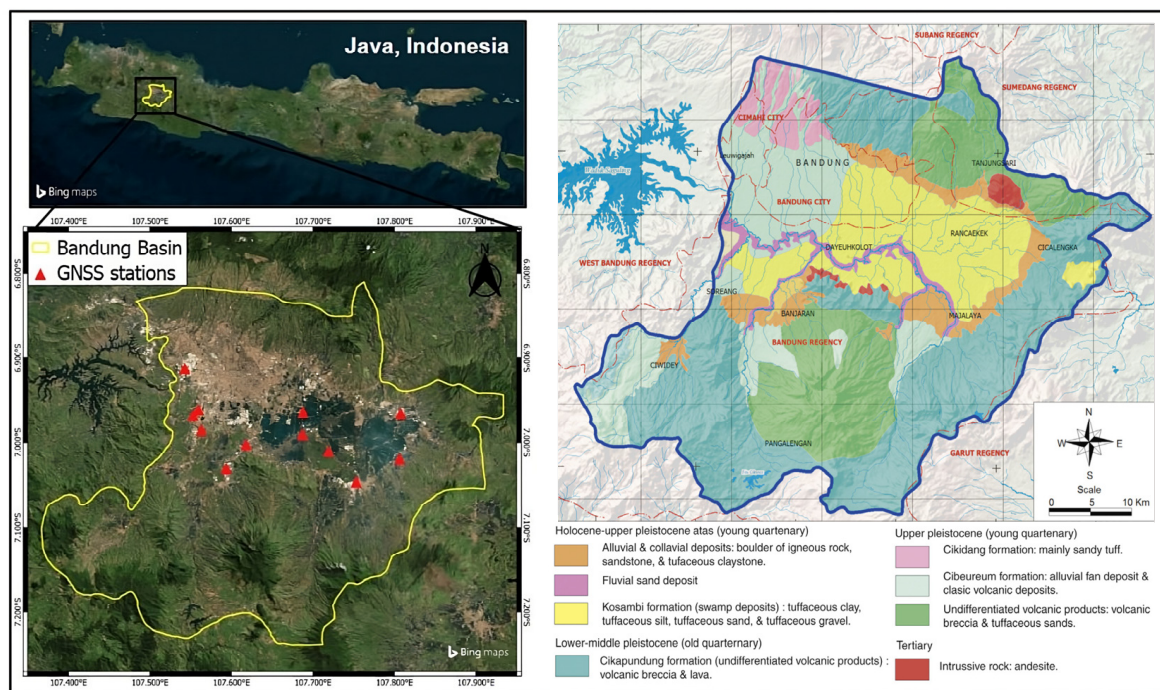
Figure 1. Location and aerial view (left), along with the geological map of the Bandung Basin (right—modified from [45]). The red triangles indicate the location of the Global Navigation Satellite System (GNSS) stations used for validation of the InSAR results.

3. InSAR Data

In this study, we processed 344 Sentinel-1A/B radar images acquired in C-band (wavelength 5.6 cm) between January 2015 and December 2020 over the Bandung Basin, Indonesia (Table 2). We produced interferograms with every two consecutive pairs of images using the Interferometric synthetic aperture radar Scientific Computing Environment (ISCE) software [46,47]. The resulting small baseline interferogram stack was processed using the Miami INsar Time-series software in Python (MintPy) [47] to produce line-of-sight (LOS) displacement time series for every pixel in the dataset, for both the ascending and descending satellite geometries. The reference point was selected in a mountainous area located in the north of the study area that is believed to be relatively stable in terms of displacement. To obtain reliable InSAR time-series datasets, a temporal coherence mask of 0.7 was applied to the resulting MPs.

Table 2. Information on the Sentinel-1 parameters and the InSAR results.

Parameters	SAR Satellite	
	S-1A&B *	S-1A&B *
Satellite orbit	Ascending	Descending
Track	98	149
Time span	4 January 2015–27 December 2020	7 January 2015–30 December 2020
Mean incidence angle (°)	48	43
Number of images	190	154
Number of acquisitions	153	146
Number of MPs	650,863	735,333
MP density (MP/km^2)	680	769

* Sentinel-1 B data collection commenced in September 2016.

4. Data Mining Methods

Our approach followed three main steps: (1) Combination of ascending and descending displacement time series to retrieve the vertical and east-west displacement time series;

(2) Time series clustering via UMAP and HDBSCAN; and (3) Cluster time series change detection using piecewise linear functions–PWLF (Figure 2).
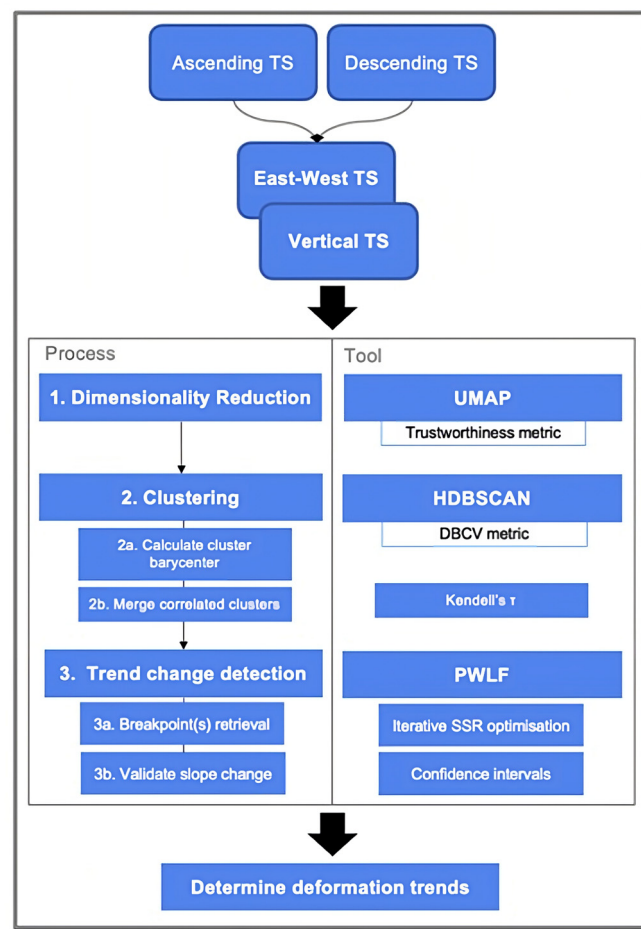

Figure 2. Workflow of the proposed approach.

4.1. Retrieval of Vertical and East-West Deformation Fields

For each satellite orbit, the retrieved deformation time series are one-dimensional projections onto the satellite's Line of Sight (LOS). When both ascending and descending InSAR datasets are available, combining the datasets is a common approach to retrieve the vertical and east-west deformation fields [48,49]. The combination is performed under the assumption that the contribution of north-south horizontal movements is negligible, which is a typical assumption in InSAR studies due to the poor sensitivity to this motion [50].

For Bandung, the LOS datasets were resampled both in time and space. Regarding the time interpolation, since the time series datasets retrieved from the ascending and descending satellite orbits were at different acquisition dates, consistent time steps were imposed by implementing a frequency conversion and resampling both time series datasets to equivalent lengths with a consistent 7-day time step. For the spatial interpolation, the point-wise datasets were resampled to a regular 30 m × 30 m grid. In the instance that both ascending and descending measurements were available at the same grid cell, the combination was performed using the known values of the LOS displacement in the ascending (Da) and descending (Dd) orbit within each grid cell i, the vertical component (Dv) and east-west component (De) were estimated using the following equation (adapted from [51]):

$$D_{vi} = \frac{E_{ai} \times D_{ai} - E_{ai} \times D_{di}}{E_{di} \times U_{ai} - E_{ai} \times U_{di}} \tag{1}$$

$$D_{ei} = \frac{E_{di} \times D_{ai} - E_{ai} \times D_{di}}{E_{di} \times U_{ai} - E_{ai} \times U_{di}} \tag{2}$$

where E, N, and U are the east–west, north–south, and vertical directional cosines that define the three-dimensional orientation of the satellite Sentinel-1 Interferometric Wide swath mode LOS. The directional cosines depend on the incidence angle θ, and the heading angle α of the satellite flight direction. At each location, the direction cosines are estimated as: $E = -\cos\alpha \times \sin\theta$ and $U = \cos\theta$ (Figure 3). For this work, an averaged value of θ across the study area for each orbit was used (Table 2).
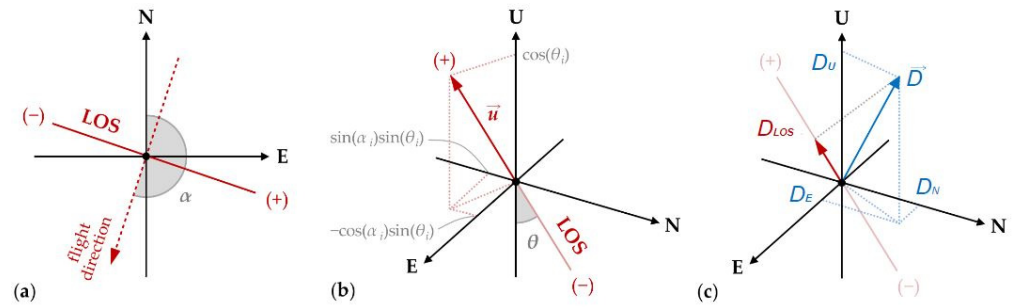

Figure 3. Representation of the satellite descending LOS (**a**) in the horizontal plane, (**b**) in 3D, and (**c**) the displacement vector projected along the LOS. Modified from [51].

4.2. Clustering Workflow

The following section details the use of UMAP [32] as a dimensionality reduction method to transform the interferometric data, and in doing so, increases the effectiveness of clustering implemented with the HDBSCAN [33] algorithm.

4.2.1. Dimensionality Reduction with UMAP

Dimensionality reduction is widely used in big data analytics to help analyze large, high-dimensional datasets [52]. UMAP is a non-linear, neighbor-based dimensionality reduction technique used for high-dimensional data [53]. The algorithm functions by extracting the relationship between data points in the original high-dimensional space (Figure 4), capturing the underlying structure of the data, and then using manifold learning to project the graph into a lower-dimensional space. UMAP maps the data points in the high-dimensional space onto a simpler, lower-dimensional space while preserving the relationships between them, thus capturing the global and local structure of the data [53].
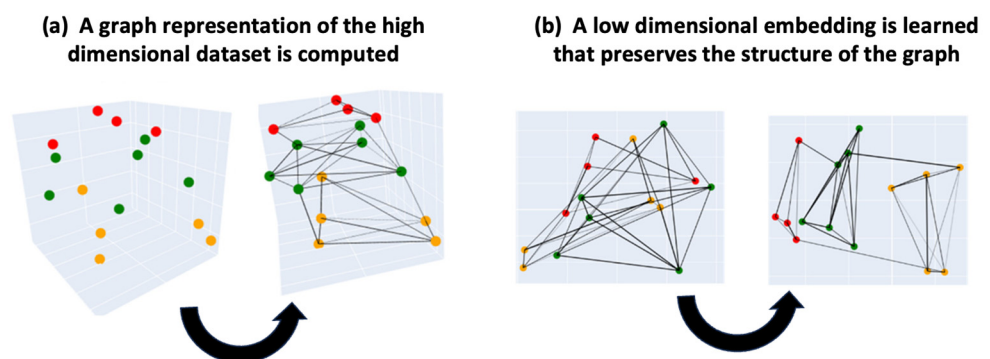

Figure 4. Visual overview of how UMAP functions: (**a**) initially, a graph representation of the high-dimensional input data is computed; (**b**) next, a low-dimensional embedding is learned so that the structure of the graph is preserved. Image modified from [54].

This dimensionality reduction step performed with UMAP is necessary to cluster the data effectively and support the density-based clustering algorithms, since standard clustering techniques on high-dimensional datasets often result in unsatisfying results due

to the high dimensionality [55]. When too many features are present, observations are harder to cluster, or each cluster can include heterogeneous features in the same class [32]. The UMAP algorithm was implemented in this study using the official Python package umap-learn [32]. Table 3 summarizes the hyperparameters and the evaluation metrics used in this study. The *min_dis* hyperparameter of UMAP sets the minimum distance apart that points are allowed to be in the low-dimensional representation, controlling how tightly the points are packed together in the resulting embedding. In this work, min_dis was set to zero to ensure that the MPs were packed together densely to encourage clearer separation of the clusters by the clustering algorithm.

Table 3. Hyperparameters and metrics of the machine learning algorithms used in this study.

ML Algorithm	Hyperparameter	Parameter Definition	Model Evaluation Metric
UMAP	n_neighbors	size of the local neighborhood	Trustworthiness
	min_dis	minimum distance between points	
HDBSCAN	min_samples	minimum number of neighbors to a core point	Density-Based Clustering Validation
	min_cluster_size	minimum size of a final cluster	

To validate the lower dimensional embedding output by UMAP, we exploited a metric called trustworthiness (Equation (14) [56]), which is a measure of how well the structure of the dataset is preserved after dimensionality reduction [57]. The value of trustworthiness ranges from zero to one and represents the degree to which the low-dimensional embedding preserves the pairwise distances of the high-dimensional data, with zero representing no preservation in the data and one reflecting total data conservation in the lower dimensional embedding. To evaluate the trustworthiness, we measured the distance of each sample in high-dimensional space against its closest neighbors using rank order, and evaluated the extent to which each rank changed in the low-dimensional embedding. We reduced the computational time and increased the performance by performing the trustworthiness validation on a representative random subset of the dataset (10% of the total MPs). We ensured equivalency between the subset and the original dataset using the Kolmogorov–Smirnov test (K–S test). The K–S test is a nonparametric goodness-of-fit test that determines whether two independent sample distributions are statistically different [57]. The test works under the null hypothesis (H_0) that two samples have the same distribution. We used a standard significance level of 0.05, which allowed us to accept the results within a 5% margin of error. When comparing the total InSAR dataset with a random subset, the subset was deemed as similar to the complete dataset and therefore, significantly representative, if the *p*-value was greater than 0.05. We implemented the K–S test in Python using the scipy.stats.ks_2samp function.

Using the nearest neighbor method, the UMAP-produced map is considered trustworthy if these neighbors are also placed close to the respective sample in the low-dimensional space. The trustworthiness scores range between zero and one, with higher scores indicating that more of the local structure of the original dataset is retained in the UMAP embedding. In order to achieve the best fitting embedding, the n_neighbors UMAP parameter was optimized using the randomized search method which selects and tests 100 random combination of hyperparameters. The n_neighbors hyperparameter controls how UMAP balances local versus global structure in the data by determining the number of nearest neighbors to consider while constructing the low-dimensional embedding. A larger value for n_neighbors results in a more global view of the data, while a smaller value results in a more local view. The embedding with the highest trustworthiness score was selected as the most reliable model to be used in the downstream clustering process.

4.2.2. Time Series Clustering with HDBSCAN

The input for this step is the low-dimensional embedding created by UMAP. We implemented the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [33] algorithm to detect and group together time series with similar displacement behavior, even though they may not be in close spatial proximity. The HDBSCAN algorithm was implemented in this study using the hdbscan Python package [58]. In order to achieve the best clustering outcome, the clustering algorithm was optimized by exploiting the relative_validity function of HDBSCAN. This is a metric known as the Density-Based Clustering Validation (DBCV), which works for density-based clustering algorithms since it takes noise into account and captures the shape property of clusters via densities instead of distances. To optimize the results, we tuned two HDBSCAN hyperparameters: min_samples and min_cluster_size. The min_samples parameter is used to control the number of samples in a cluster. The key concept in HDBSCAN is density reachability, which is a measure of how close a sample point is to a cluster. This parameter is used to control the density reachability threshold, and it affects the granularity of the clustering. A lower value will result in more clusters being identified, while a higher value will result in fewer clusters. The min_cluster_size parameter determines the minimum number of points that a dense region must contain for the algorithm to consider it a valid cluster. A low value may enable HDBSCAN to identify many small clusters, even if they only contain a few data points, while a high value may result in clusters that contain a large number of data points, which could miss smaller clusters. Multiple combinations of these two parameters were run via a random search of 100 different parameter combinations to find a result that generated the highest relative_validity score, while penalizing models with low coverage (the extent to which all the data points are assigned to a cluster or noise).

For further analysis, we calculated the Euclidean barycenter within each cluster, which is simply the arithmetic mean for each time step, minimizing the summed Euclidean distance. Similar clusters were then merged by calculating the Kendall rank correlation τ [59] between pairs of extracted cluster barycenter time series. The Kendall rank correlation coefficient τ measure is used to determine the degree of association between variables based on the ranks of the data. Kendall's τ is obtained with the following equation:

$$\tau = \frac{P - Q}{\sqrt{(P + Q + T)\,(P + Q + U)}} \tag{3}$$

where P is the number of concordant pairs, Q the number of discordant pairs, T the number of ties in the first data series ranking, and U the number of ties in the second series ranking. The two series can be considered as statistically correlated when the null hypothesis (H_0) of an absence of association is rejected. In this work, we considered a significance level of 0.05 and rejected the null hypothesis when the p-value was lower than 0.05, thereby concluding that the time series were correlated. Additionally, a τ coefficient higher than 0.9 was selected to merge highly correlated cluster time series. T was computed in Python using the scipy.stats.kendalltau function.

4.3. Cluster Gradient Change Detection

We identified changes in the displacement rates through application of a piecewise-linear function model to each cluster barycenter using the PWLF Python package [60]. PWLF fits continuous piecewise linear functions to a set of data points. The package requires the number of breakpoints to be specified in advance, meaning that the number of linear segments in the piecewise linear function is fixed. Multiple linear segments were fitted to the cumulative deformation time series of each cluster centroid, with each linear segment representing a constant rate. The specific time at which the slope (i.e., velocity) of the linear segment changes is identified as a breakpoint, and these breakpoints represent the times when an acceleration or deceleration of displacement occurs.

Fitting the cluster barycenters to piecewise linear functions was based on an iterative least squares procedure. The number of breakpoints was a discrete parameter used in the modelling process with the optimal value obtained by minimizing the sum of the squared residuals for each piecewise function. The number of breakpoints within each cluster barycenter time series was unknown at the beginning of the modelling procedure. To determine the optimal number of breakpoints, the model was initialized with zero breakpoints, and then incrementally increased for each model run. The iteration ceased when the sum of the squared residuals of the fitted piecewise model no longer achieved an improvement greater than 15%, which was selected to avoid overfitting of the model. The breakpoint selection was then checked by computing the confidence interval of adjacent slopes to check that significant differences occurred on either side of the detected breakpoint, confirming that a change in displacement velocity did occur. Assuming that the parameters followed a normal distribution we estimate the 95% confidence interval of a particular slope as ± 1.96 times the standard error of the slope estimated by the model [61]. A significant change in slope indicated a deceleration or acceleration point that was identified when the confidence intervals of two consecutive slopes did not overlap.

The ©dentification of changes in the deformation provides valuable insights into the land movement dynamics of Bandung that has not been analyzed so far. Investigating these displacement changes along with ancillary information such as groundwater extraction/injection variations, precipitation changes, and changes in land use, can reveal insights into the causes of the observed deformation.

5. Results

This section presents the results obtained from application of the proposed time series clustering and change detection procedure for detecting locations with similar deformation behaviour and identifying the timing of when gradient changes occur over Bandung.

5.1. Decomposition of Vertical and Horizontal Displacement Fields

The LOS deformation velocity maps derived from the InSAR processing in the Bandung Basin are displayed in Figure 5. Similar velocity rates were detected within the study area by the different orbits, ranging from -11.8 cm/y to 7.8 cm/y from the ascending orbit and -10.3 cm/y to 5.6 cm/y from the descending orbit.

The vertical and horizontal displacements derived following the methodology described in Section 4.1 are shown in Figure 5c,d. The ground displacement was predominately in the vertical direction, which is comparable to results from a previous Global Navigation Satellite System (GNSS) campaign [41], and shows a similar deformation distribution to the LOS velocity maps. The main subsidence features were located within the central portion of the basin with observed rates as high as 16.2 cm/y. Uplift is perceived at rates up to 4.8 cm/y at the northern and southern boundaries of the study area. The estimated horizontal displacements (Figure 5d) were within the order of the noise of the InSAR signal, indicating that there was no significant horizontal motion with respect to the reference point. The small horizontal displacements, along with the focus of the work (characterizing the subsidence process), are the reasons why we focus solely on the vertical component of movement.

The high uplift rates (up to 4.8 cm/y) can be attributed to the location of the reference point used during the InSAR processing (shown in Figure 5a,b) which was initially thought to be located in a stable area (no GNSS data were made available to us at the time of InSAR processing). Instead, it is apparent that this area is actually underdoing subsidence. Since all of the MP displacement measurements are referenced to this point, the movement experienced by the reference point will be reflected in the entire displacement time series dataset and as a result, MPs that are subsiding at slower rates than the reference point will appear to be uplifting, and areas of subsidence will be underestimated.
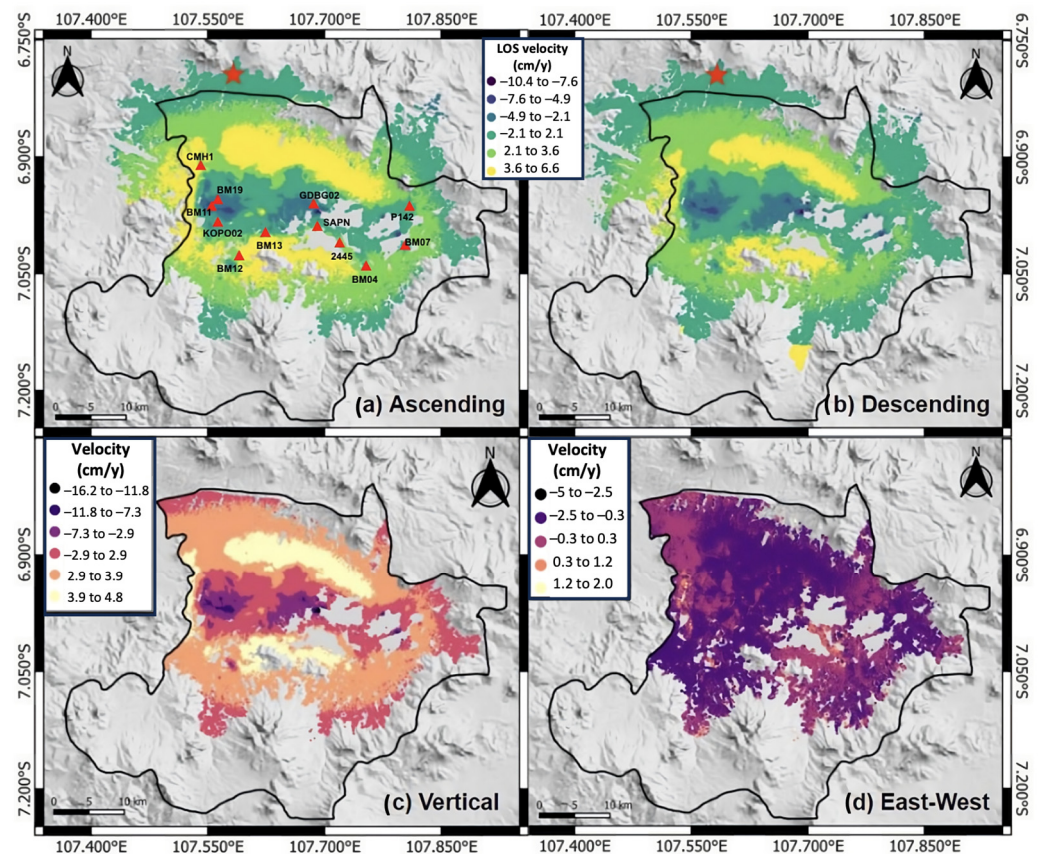
Figure 5. InSAR displacement rates along the (**a**) LOS ascending orbit, (**b**) LOS descending orbit, (**c**) vertical V$_U$ direction, and (**d**) east-west V$_E$ direction. Positive values represent the areas moving toward the sensor along the LOS while negative values indicate areas moving away from the satellite along the LOS. The location of the reference point is indicated by the red star in (**a,b**), the GNSS station locations are given by the red triangles in (**a**), and the black line outlines the extent of the Bandung Basin. The base map is a hillshade map produced from the 5 m resolution National Digital Elevation Model (Seamless DEM dan Batimetri Nasional—DEMNAS).

To correct for this reference issue, a continuous GNSS station was located within the study area (the location is shown by the red triangle in Figure 6) and compared with the mean deformation time series of the MPs within a 100 m buffer of the station. The analysis with the GNSS data reveals a discrepancy between the displacement measurements, with the InSAR displacements undergoing an apparent constant uplift while the GNSS measurements show slight subsidence at a rate of −0.41 cm/y (Figure S1 in the Supplementary Material).

The following steps are implemented to re-reference the displacement measurements using the GNSS measurements in order to correct for the bias imposed by the original reference point:

1. Both the GNSS and InSAR displacement time series are resampled using a 7 day moving average so that the dates of the displacement measurements from both sources coincide.
2. Linear regression is used to approximate the displacement time series of both the GNSS and the InSAR measurements.
3. The difference between the linear models is subtracted from the InSAR time series measurements of all MPs to reference the displacement dataset to this point.
4. We note that the GNSS reference location is also subsiding at a rate of −0.41 cm/y, so the broader subsidence pattern remains underestimated by this amount. Thus, 0.41 cm/y is subtracted from the InSAR vertical displacement rates.
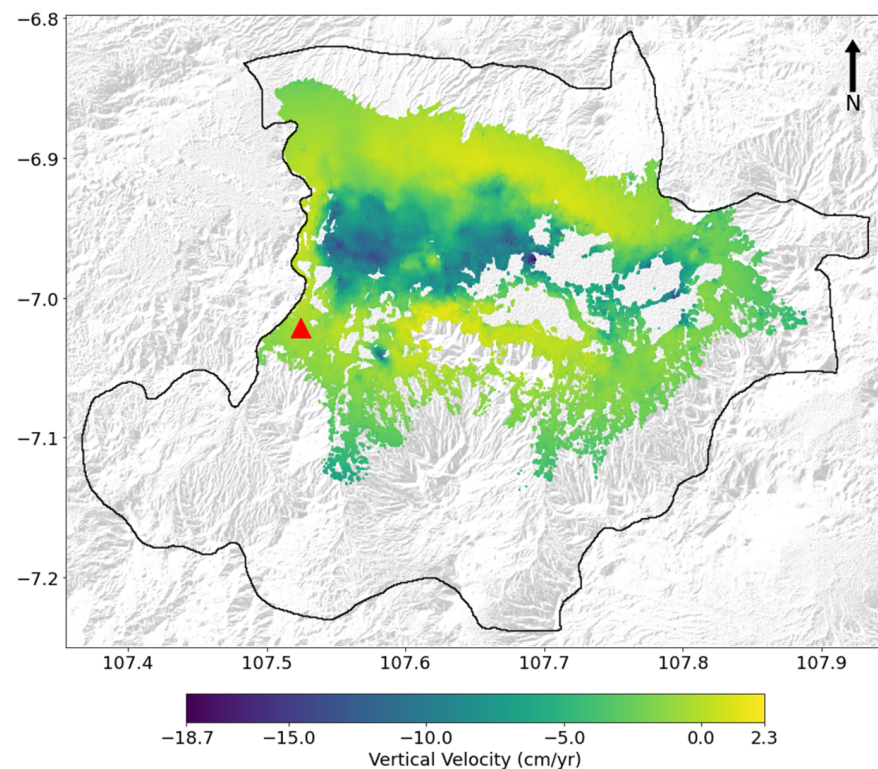
Figure 6. InSAR vertical displacement rates following reference correction using measurements from a continuous GNSS station (geographic location indicated by the red triangle). The black line outlines the extent of the Bandung basin. The base map is a hillshade map produced from the 5 m resolution National Digital Elevation Model (Seamless DEM dan Batimetri Nasional—DEMNAS).

Figure S2 in the Supplementary Material shows the improvement in the corrected InSAR time series when compared to the continuous GNSS measurements. This procedure was applied to all of the MPs within the study area and the resulting vertical displacement rate map is shown in Figure 6. The re-referenced displacement pattern remains similar to the un-corrected map (Figure 5c); however, the range of velocity values have changed and now reflect more realistic displacement estimates with a maximum subsidence rate of −18.7 cm/y and an uplift rate of 2.3 cm/y, which is more or less stable given that the standard deviation for the re-referenced InSAR velocity results is 3.1 cm/y. These values are more in line with previous InSAR results [62–64].

We validated the reference corrected vertical InSAR measurements with additional GNSS observations from 12 stations located throughout the basin (the locations are shown in Figure 5a and the data are provided in Table S1 in the Supplementary Material) produced during a campaign that was carried out from 2016 to 2019 by students from the Institute of Technology Bandung. The correlation between the GNSS and InSAR velocities are shown in the scatterplots in Figure 7 with the comparison before the reference correction on the left and the analysis with the corrected values on the right. The correlation coefficient improved from 0.78 to 0.84 by implementing the reference correction, and the issue of the InSAR data underestimating the subsidence rates was mitigated.
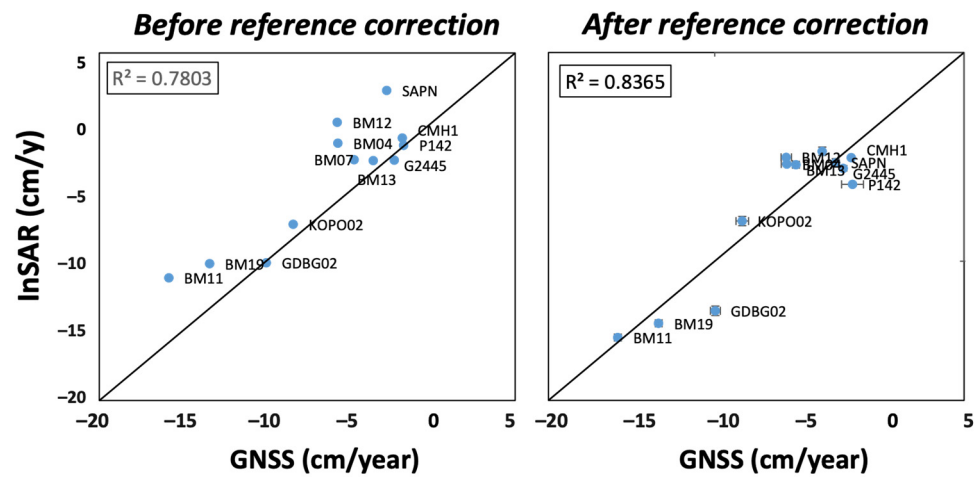
Figure 7. Comparisons of the GNSS and InSAR vertical velocities calculated for each GNSS site before (**left**) and after (**right**) the reference correction. InSAR velocities are calculated by averaging the velocities of the MPs located within a 100 m buffer of each GNSS site. The corresponding coefficient of determination is labeled in the plot. The geographical locations of the 12 GNSS stations are shown in Figure 5a.

5.2. Time Series Clustering

5.2.1. UMAP Dimensionality Reduction

Extracting features using the UMAP algorithm allowed for the transformation of the high-dimensional displacement time series dataset into a low-dimensional latent space. The input to the UMAP algorithm was a matrix in which the columns corresponded to the time and the rows represented the point-wise vertical displacement time series. The input data consisted of 312 dimensions, corresponding to the length of the time series, which are treated as independent samples by UMAP. Following transformation, the time series dataset was reduced to two dimensions, reflecting the number of informative features. The optimal value of the n_neighbors parameter used for the dimension reduction was selected as the value at which the maximum trustworthiness score plateaued, which was at a value of 100 (Figure 8). The high trustworthiness score of 0.994 indicates that the local structure of the original dataset was well preserved in the low-dimensional embedding.
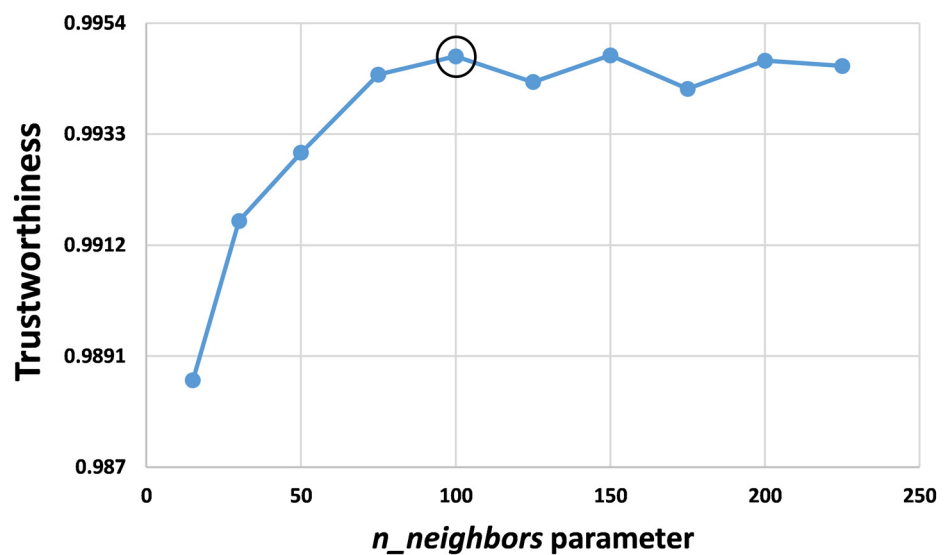

Figure 8. UMAP trustworthiness analysis to determine the optimal value of n_neighbors. The black circle indicates the value used in the optimized UMAP model for the vertical displacement dataset.

5.2.2. HDBSCAN Time Series Clustering

The inputs for this step were the low-dimensional features learned by UMAP. The UMAP algorithm placed the data in relative proximity to one another based on their temporal similarity, and HDBSCAN clustered these points based on their relative densities. Following the hyperparameter optimisation process described in Section 4.2.2, the tuned min_samples and min_cluster_size hyperparameters used in the final HDBSCAN model were 80 and 350 respectively, which resulted in 37 initial clusters.

To inspect the result, we mapped the geographical coordinates of each MP time series along with the cluster label assigned by HDBSCAN. Initially, HDBSCAN successfully clustered the southern and western portions of the basin but identified the central and northern portions as a single cluster, as shown by the purple colour in Figure 9a. The displacement time series for this cluster (Figure 9a) showed a high level of variability amongst the clustered MPs that needed to be further decomposed. In order to extract this information, we recursively applied the HDBSCAN algorithm to this single cluster, allowing the algorithm to distinguish features within this group. In total, 64 clusters were extracted from the vertical displacement dataset after this re-clustering procedure. The spatial distribution of the extracted clusters is shown in Figure 9b.

After extracting the clusters using HDBSCAN, we computed the barycenter displacement time series for each cluster and merged clusters that had correlated barycenters. This process resulted in a total of 36 vertical displacement clusters. Figure 9c shows the spatial distribution of the resulting clusters. The distribution of the extracted clusters shows that a small number of clusters dominate movement in the northern and southern borders of the basin, while the central portion of the basin, with a larger number of clusters, experiences more variation in deformation over time.

Table 4 lists the properties of the extracted clusters that demonstrated subsidence behaviour. Cluster 18 and cluster 25 had the highest values of cluster-averaged cumulative displacement (-21.7 cm and -22.3 cm, respectively) and were assigned the majority of subsiding MPs (61% between the two clusters).

Table 4. Properties of the main subsidence clusters extracted from the vertical interferometric dataset.

Cluster N°	N° of MPs	Relative N° of Subsiding MPs (%)	Area (km²)	Average Cumulative Displacement (cm)	Time of Detected Change
18	15,968	41	35.3	-21.7	NA
25	7759	20	18.2	-22.3	April 2017
19	4585	12	12.1	-0.5	NA
14	3972	10	11.3	-3.5	September 2017
23	1211	3	3.7	-17.1	January 2020
36	1135	3	3.0	-1.2	NA
17	1101	3	2.7	-16.3	February 2017 & September 2018
12	776	2	2.1	-1.2	July 2017
21	766	2	2.6	-9.5	August 2016
24	707	2	2.0	-10.5	January 2020
20	482	1	1.7	-9.8	NA
34	246	<1	0.9	-7.0	NA

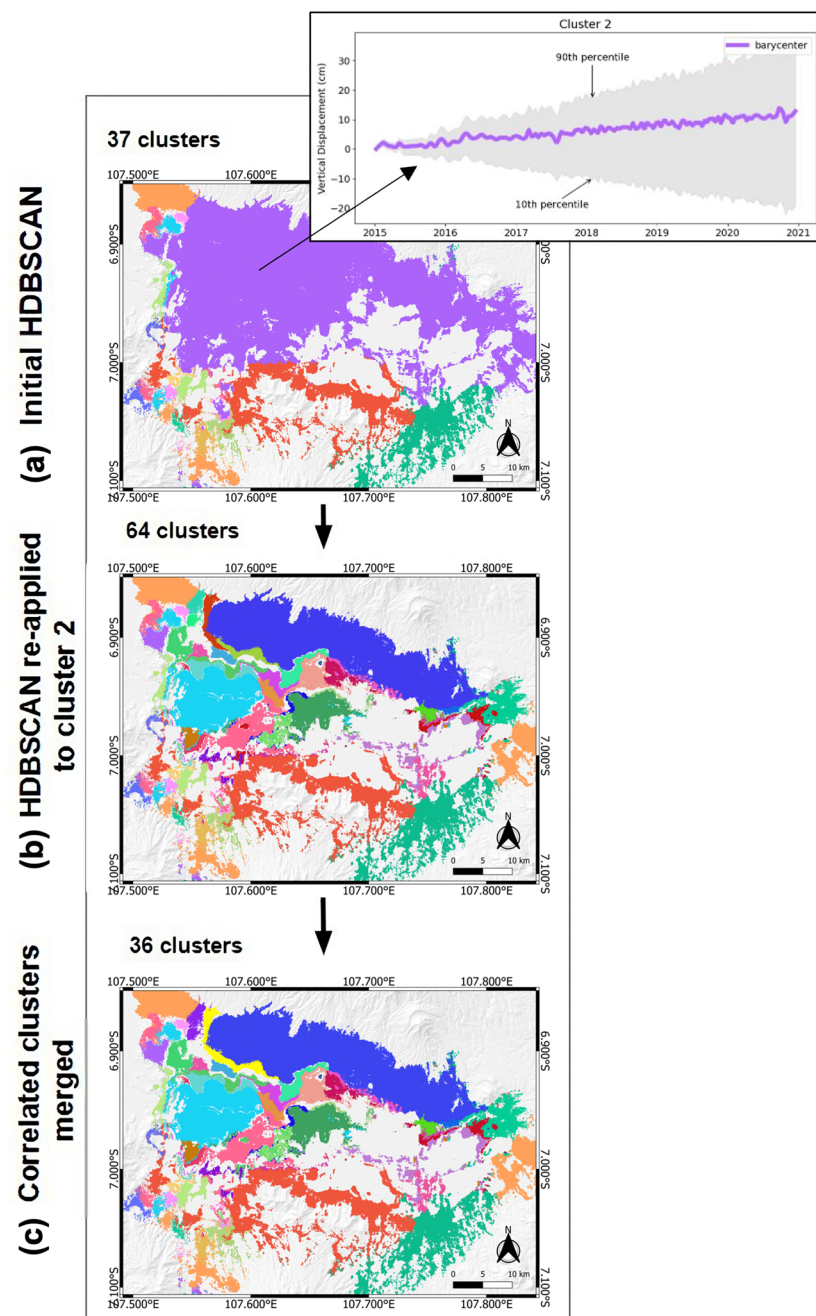NA: Not Applicable—no change point detected.

Figure 9. Map of the cluster distribution, with each color representing a different cluster, for the vertical interferometric displacement dataset: (**a**) Resulting from the initial application of HDBSCAN; (**b**) Resulting from the re-application of HDBSCAN clustering to the purple cluster in (**a**); (**c**) Resulting from merging correlated clusters.

5.3. Cluster Gradient Change Detection

Instead of simply estimating the average velocity for each cluster over the InSAR period by fitting a linear function to the whole displacement time series, we fit a piecewise linear model to identify sudden changes in the deformation rate or breaks in the trend. If different trends in the deformation behaviour occur due to variabilities in the triggering factors over time, then piecewise linear regression is an effective means to detect these changes [65]. The adopted strategy has allowed us to efficiently identify the underlying displacement patterns and detect when deformation changes occurred in an InSAR dataset.

We used PWLF to identify a total of 12 barycenter cluster time series in the vertical dataset, 11 of which had a single breakpoint and one (cluster 17) which had two breakpoints, corresponding to three main subsidence patterns: (1) constant linear subsidence, (2) subsidence with an acceleration, and (3) subsidence with a deceleration. Figure 10 shows the spatial distribution of the identified subsidence trends and the percentage distribution of the subsiding MPs. The timings of the identified breakpoints for each cluster are given in Table 4.
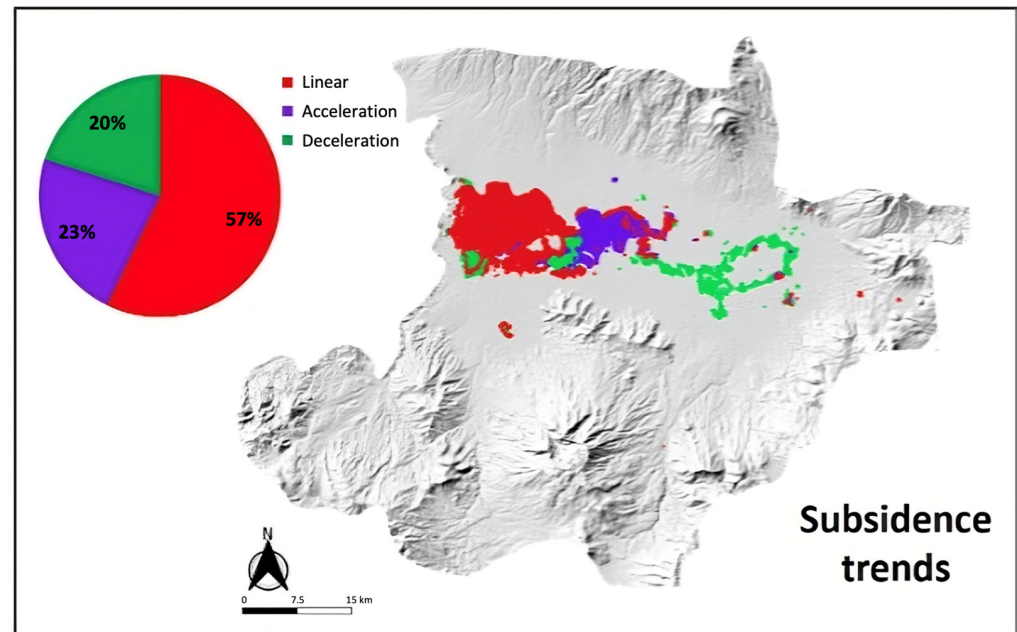
Figure 10. Map of the spatial distribution of the subsidence trend clusters extracted from the vertical displacement dataset.

The areas of subsidence are predominantly confined to the central portion of the basin. Additionally, a localized subsiding area in the southwest of the study area was evident. A linear trend dominates the subsidence signal across the basin, encompassing 57% (22,170 MPs) of the total subsiding MPs, followed by an accelerated subsidence signal (23%; 8771 MPs), and a decelerating subsidence trend (20%; 7767 MPs).

The cluster displacement time series with the fitted piecewise models belonging to each of the identified subsidence trend groups are shown in Figure 11.

Clusters 18 and 20 exhibit a strong, linear subsidence pattern (Figure 11a). According to [44], there are two plausible explanations for the observed constant subsidence: (1) It is possible that water extraction primarily occurs from confined aquifers, which are less influenced by seasonal recharge; (2) During the rainy season, water extraction rates may be higher, compensating for the recharge in the shallow, unconfined aquifer. However, since stress changes in unconfined aquifers are lower, it is unlikely that substantial subsidence rates, such as those observed in clusters 18 and 20, would result from compaction of the unconfined aquifer. For this reason, the authors concluded that industrial water usage from confined aquifers was likely the primary factor contributing to subsidence in the urban area of Bandung [44]. This inference is further corroborated by the fact that only the deep wells used for industrial purposes tap into the groundwater resources of the confined aquifer underlying Bandung [66]. The lower rates of subsidence in clusters 19 and 36 (Figure 11a) could reflect a difference in the type of groundwater usage. In these locations, the industrial land coverage is less significant, and is instead dominated by residential areas. Thus, it is possible that the lower but constant subsidence rates reflects the lack of industrial water extraction in these areas and instead could be representative of residential extraction, which coincides with the findings of [67] who investigated the subsidence phenomenon in Bandung at different spatial scales. In Bandung, extraction from the shallow aquifer

for household purposes is entirely permissible and unconstrained [66]. Unfortunately, there is no monitoring program in place to track the amount of domestic groundwater extracted [66], and therefore it is not currently feasible to quantitatively assess the role that residential shallow groundwater extraction has on the subsidence in these areas.
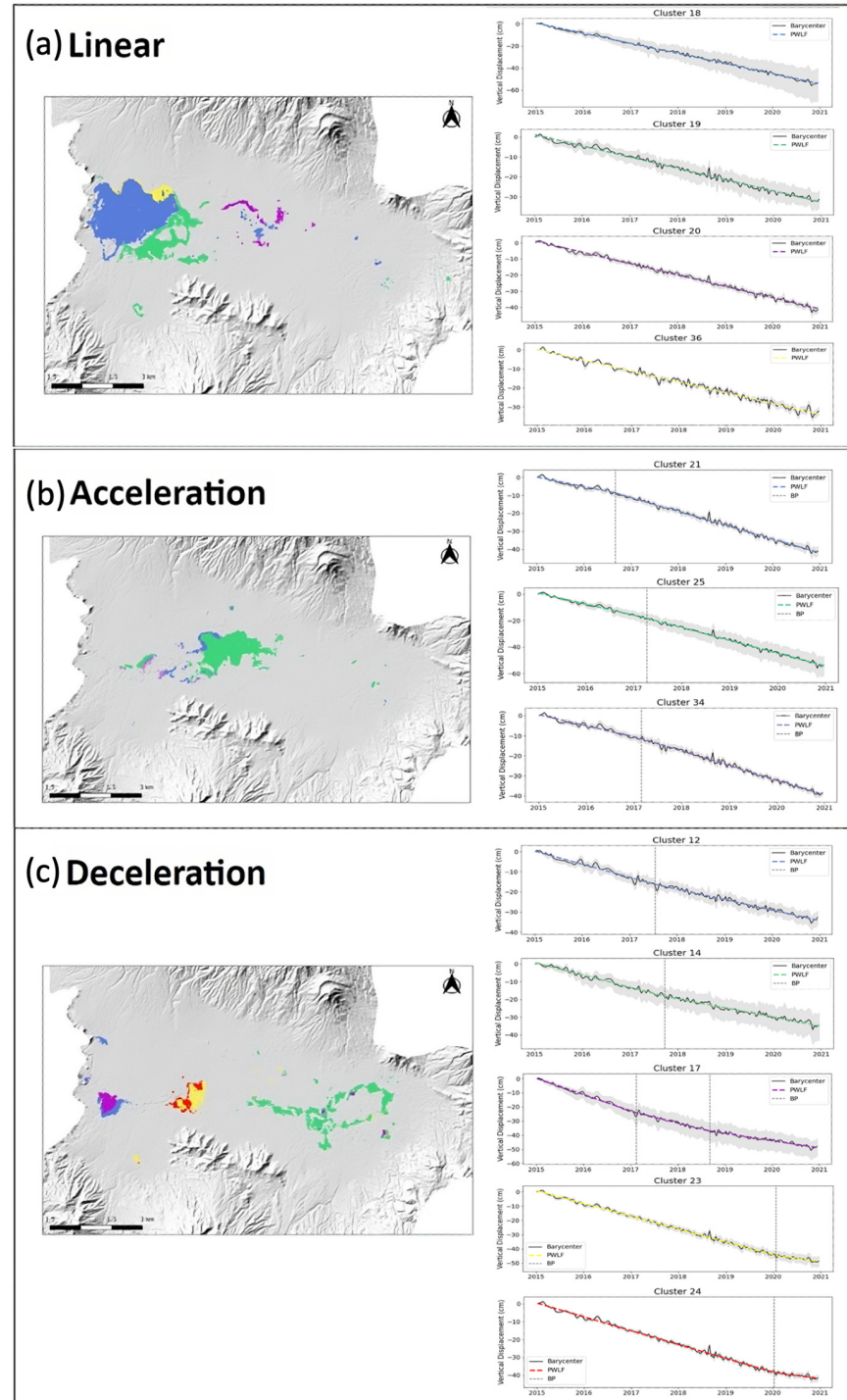

Figure 11. Map of the cluster location (left) and displacement time series (right) for each of the identified subsidence trends: (**a**) linear, (**b**) acceleration, (**c**) deceleration. The black line represents the cluster barycenter, the shaded grey area represents the 10% and 90% quantiles of MP time series within the cluster, the colored lines reflect the linear segments of the fitted piecewise model, the vertical dashed grey line indicates where a breakpoint is identified (right) and the corresponding color is mapped to show their spatial distribution (left).

Cluster 17, belonging to the decelerating subsidence group, represented the area with the highest subsidence rates over the studied period. The spatial distribution of this cluster is highlighted in Figure 12a and the displacement time series is shown in Figure 12c. Analysis of the cluster barycenter reveals an initial subsidence rate of −6.2 cm/y in January 2015, which decreases to −3.0 cm/y at the beginning of 2017, and further decelerates down to zero towards the end of 2018. Satellite imagery from Google Earth (Figure 12b) shows the presence of a large industrial complex surrounded by residential buildings in this area prior to 2010. Additional buildings were constructed during 2010–2015, followed by even more buildings and a highway during the InSAR observation period (2015–2020). This high-subsidence zone is in the western portion of the basin, within the Kosambi Formation where borehole logs in this area indicated that the clay thickness ranges from 47 m to 87 m [68].
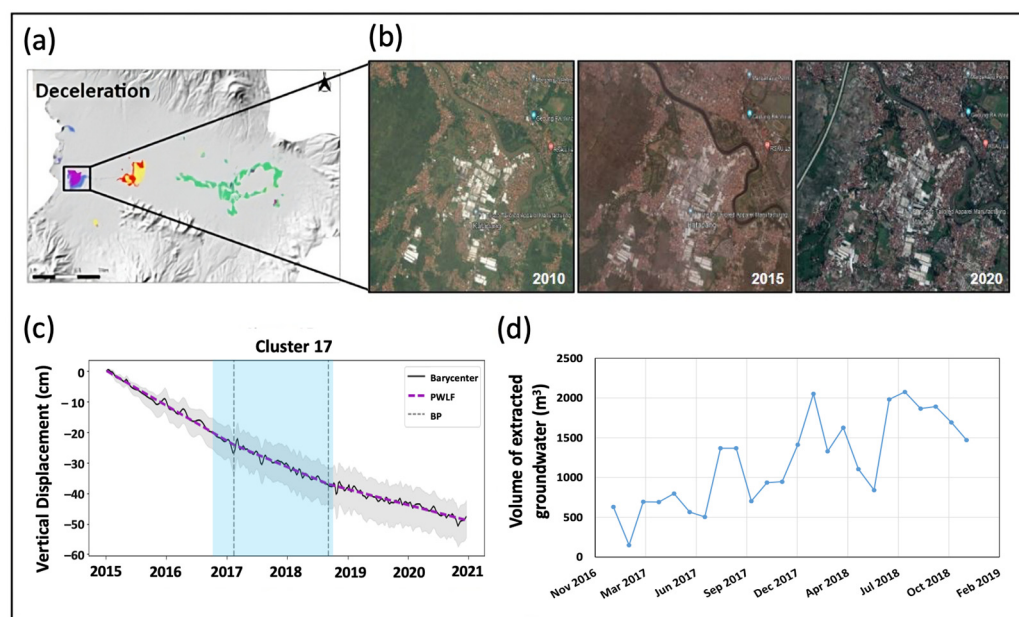

Figure 12. (**a**) Map of the decelerating subsidence clusters highlighting the location of cluster 17 with the highest average subsidence rate. (**b**) Satellite imagery (from Google Earth, www.google.com/earth, accessed on 8 February 2023) of the cluster 17 area from 2010, 2015, and 2020. (**c**) Cluster 17 displacement time series where the solid black line represents the cluster barycenter, the shaded grey area represents the 10% and 90% quantiles of MP time series within the cluster, the purple line represents the linear segments of the fitted piecewise model, and the black dotted lines indicate the detected breakpoints. The shaded blue area represents the time period for which measurements of groundwater pumping volumes are available. (**d**) Average volume of groundwater extracted from nearby pumping wells.

Subsidence at this location may result from natural compaction due to the consolidation of young, soft sediments, as well as the added load of new buildings (i.e., settlement of highly compressible soil). Groundwater extraction may have also played a role in the observed subsidence. Pumping volume measurements of some nearby industrial pumping wells obtained from the Office of Mineral and Energy Resources (ESDM) of West Java Province show an increase in the volume of groundwater extracted during 2017 until August 2018 (Figure 12d). The volume of groundwater extracted showed a declining trend for the remainder of 2018. Unfortunately, no measurements were made during 2019, nor were groundwater levels monitored in this area. Therefore, the breakpoint detected in June 2019 marking a deceleration in subsidence cannot with certainty be attributed to lower rates of groundwater pumping. However, it is possible that decreased pumping from late-2018 and into 2019 could explain the stable ground conditions observed at this time.

Clusters 21 and 25 identified a large area of accelerating subsidence in the middle portion of the basin (Figure 11b). The average subsidence rates within these clusters increased from 1.6 cm/y to 4.7 cm/y, with the change occurring in late 2016. The land cover in this area was mostly residential with several large industrial complexes. From analyzing aerial photographs, we found no major land use change during the studied period that could explain the acceleration in subsidence. The accelerated rate might be due to increased groundwater usage. Unfortunately, no data on the groundwater levels or pumping volumes were available from this time to compare the water table variation with the observed changes in deformation behavior.

In contrast, the eastern portion of the basin principally shows a deceleration subsidence pattern. Unfortunately, this is a predominantly agricultural area where decorrelation due to vegetation and seasonal inundation degrades the interferometric coherence. The available MPs in this area are principally represented by cluster 14 (Figure 11c). The MPs within this cluster show an averaged subsidence rate of -1.72 cm/y from January 2015 until September 2017, after which the ground reached a stabilized condition. The coherent MPs coincide with residential settlements. The lack of industrialization could explain the low displacement rates at this location since groundwater is not heavily extracted by any industry and several sources [69,70] acknowledge that the use of groundwater as a source of irrigation is not a common practice in Bandung. Instead, the initial low rate of subsidence observed in this area could be associated with consolidation due to surficial loading of the residential settlements. It is possible that the initial subsidence started prior to 2015, and that the InSAR results used in this study spanning the period 2015 to 2020 captured the secondary consolidation phase where the subsidence rate declined and the ground began to stabilize.

6. Discussion

InSAR data derived from Sentinel-1 Synthetic Aperture Radar (SAR) images enables the extraction of ground movement measurements over a large area such as the Bandung Basin. However, interpreting a large number of points, each associated with a unique time series, is not intuitive. Machine-learning techniques can be a more effective way to quickly interpret the vast amount of information contained within large InSAR datasets. In this work, we resolved the vertical component of displacement by combining the ascending and descending satellite orbits, allowing for a more accurate determination of the actual displacement of the ground surface. The outlined data mining approach can also be applied to single LOS ground deformation time series when both viewing geometries are not available for combination. However, interpreting the extracted clusters must consider that the one-dimensional LOS displacement may not necessarily capture the full magnitude of the motion.

Dimensionality reduction is often performed on big data sets to extract informative features, which can serve as a pre-processing step to help downstream clustering algorithms perform better. Previous InSAR studies have used Principal Component Analysis (PCA) as a data reduction method to identify the dominant patterns of deformation by finding the directions of maximum variance in the data [71,72]. However, PCA is a linear decomposition method, which limits its usefulness in complex domains where non-linear structure exists. UMAP is an improvement over PCA because it allows for non-linear relationships between the high and low dimensions and so the data can be represented in a lower-dimensional space without losing too much information and without assuming any linearity. Additionally, if the dataset is too large and complex, PCA results may lead to misinterpretation [73]. Therefore, UMAP is more suitable as a pre-processing step when clustering large, complex InSAR datasets. Additionally, UMAP scales well with large datasets [29] and is able to find the best clusterable embedding given that the algorithm is able to preserve both the local and global structure of data [74].

Traditional clustering algorithms, such as K-means, may perform poorly during exploratory data analysis tasks as they are often designed to partition data into a prede-

termined number of clusters based on a given similarity measure. However, the number of clusters and the appropriate similarity measures are not always clear, and may need to be iteratively refined. These algorithms face common challenges, such as difficulties in selecting parameters, lack of robustness to data noise, and reliance on assumptions about cluster distributions [33]. To address these issues, this work is one of the first applications of a hierarchical clustering algorithm. We chose the HDBSCAN algorithm to cluster the InSAR time series data due to: (1) its efficient performance with large datasets, (2) robustness to data noise [75], an issue common with InSAR data, and (3) its ability to handle clusters of variable densities where traditional clustering algorithms that assume uniform density within clusters (e.g., k-means clustering) struggle. HSBSCAN is a density-based algorithm that identifies densely packed regions of data, so it does not require that every data point is assigned to a cluster. Data points not assigned to a cluster are considered as outliers or noise. HDBSCAN does not need the number of clusters specified beforehand, and is able to discover clusters of varying densities, whereas K-means will only work well with clusters that have similar densities. Therefore, HDBSCAN is more robust to noise and outliers, whereas K-means clustering forces all MPs into a pre-defined number of clusters. The procedure implemented in this study effectively reduces the noise level and increases the reliability of the deformation data. Ultimately, HDBSCAN can adapt to the structure of the data, making it a useful tool for uncovering hidden patterns and relationships.

Although we did not impose the spatial position of each MP during this analysis, we assumed that points in close proximity experienced the same geological condition and/or triggering factor and thus would move similarly. Indeed, when the location of each cluster was also considered (Figure 11), it was evident that HDBSCAN successfully identified and grouped nearby MPs solely based on their similar displacement profiles. In the temporal domain, the displacement time series associated with the subsidence clusters showed that the MPs within each cluster were aggregated according to their displacement magnitude and movement direction (Figure 11). We interpreted the clusters by computing the barycenter displacement time series, through which we fit a piecewise linear function to evaluate trend variations (accelerations or decelerations) in the deformation evolution.

The proposed method identified breakpoints in the deformation behaviour in the Bandung Basin that highlight considerable differences in the rates of vertical ground deformation and also in their spatial patterns. A total of four linear clusters over an area of 20 km^2, three accelerating clusters (8 km^2), and five decelerating displacement clusters (7 km^2) were identified within the subsidence trends. The subsidence is confined to the central basin where Late Quaternary aged alluvium and fluvial sediments are widespread and thick. Linear displacement trends were observed in certain locations, indicating continuous underlying processes. In the south-western portion of Bandung city, subsidence followed a dominantly linear pattern, with cluster 18 encompassing the greatest number of subsiding MPs (42%) over an approximate 40 km^2 area. The cumulative subsidence for this cluster was one of the highest (21.7 cm) over the studied period. The average subsidence rate of MPs within this cluster was −4.9 cm/y, with a maximum rate of −10 cm/y. This coincides with previous studies, which observed an extensive subsidence feature encompassing the area of Margaarish Town, Bandung City with displacement rates of −10 cm/yr to −2 cm/yr between 2014 to 2017 [62]. Another study, which used Sentinel-1 and ALOS-2 images acquired from September 2014 to July 2017, found that the mean velocity of this subsidence centre was −11.2 cm/y, covering an area of 14.3 km^2 [67]. The authors showed that the land use in this area was predominately residential (77%), followed by industrial (14%), and concluded that subsidence was primarily induced by groundwater pumping for residential and industrial usages. The consistency observed with our displacement rates and recent studies highlights the occurrence of subsidence in this area mainly associated with excessive groundwater withdrawal for residential and industrial purposes [67].

This study allows us to infer that changes in subsidence rates identified by decomposing the cluster barycenters with piecewise functions may be induced by groundwater

abstraction within industrial areas, or the consolidation of soft soils due to infrastructural loading. Previous studies on subsidence processes in the Bandung Basin using InSAR analysis either relied on the average annual deformation [42,62,76] or analyzed a small number of InSAR MPs to examine the displacement behaviour over time [44,70]. Both methods imply linear movement, which we have shown to be an inaccurate assumption. It is often the case that a single linear model cannot adequately describe the evolution of displacement over time; instead, multiple linear relationships applying to different time spans may be more accurate. Therefore, instead of simply estimating the average velocity for each cluster over the InSAR period by fitting a linear function to the displacement time series, we fit a piecewise linear model to identify sudden changes in the deformation rate or breaks in the trend. Assuming that different trends in the deformation behaviour occur due to variabilities in the triggering factors over time, piecewise linear regression is an effective means to detect these changes [77].

Our method overcomes the issues of traditional InSAR analysis methods by considering all of the available MPs, providing a more comprehensive way to analyze the InSAR data and detect changes in the displacement behaviour for each MP scattered throughout the basin. This provides us with both local and large-scale information to gain further insight into the processes driving subsidence which is particularly important in a data-scarce area such as Bandung where information on possible triggering factors is limited both spatially and temporally. For example, we were able to identify an area, represented by clusters 23 and 24 (Figure 11c), where a change occurred in the deformation pattern. Within these clusters, subsidence was initially detected at an average rate between 3–4 cm/y until January 2020 when the ground movement became stabilised. We infer that this change in deformation behaviour may have been a result of less groundwater being pumped from the underlying aquifer in this area towards the end of 2019. Through application of the proposed method, an indication of the groundwater usage is inferred, which otherwise would not have been possible since groundwater levels and abstraction volumes were unavailable in this area.

Cluster 17 (Figure 12) is an area of known subsidence where previous InSAR measurements revealed a localized region of subsidence that coincided with an industrial complex [44]. Vertical displacement time series from this area showed rapid subsidence at a rate of −10.8 cm/y with a generally linear subsidence trend from 2007 to 2009, which the authors attributed to groundwater extraction for industrial purposes [44]. However, in this study, the displacement time series of cluster 17 showed a change in the displacement behaviour at this location (Figure 13). An initial linear subsidence rate coinciding with the earlier results was followed by a period of deceleration ending in stabilization of the ground motion at this location. This important change in displacement behaviour would have gone undetected by simply analyzing the linear velocity map, which highlights the importance of leveraging the displacement time series data when interpreting InSAR results.

Subsidence in Bandung has commonly been attributed to over-extraction of groundwater from a confined aquifer [44]. However, we identified a possible additional contribution to subsidence, particularly in the eastern portion of the basin. In recent years, the urban expansion in this region in the form of newly developed residential areas and major roads may have resulted in additional loading of the ground, leading to compaction of the compressible Holocene sediments. During the initial phase of monitoring, moderate and decelerating subsidence rates were detected in this region (i.e., secondary consolidation), followed by eventual stabilization of movement during the final year of monitoring, implying the end of the consolidation process.
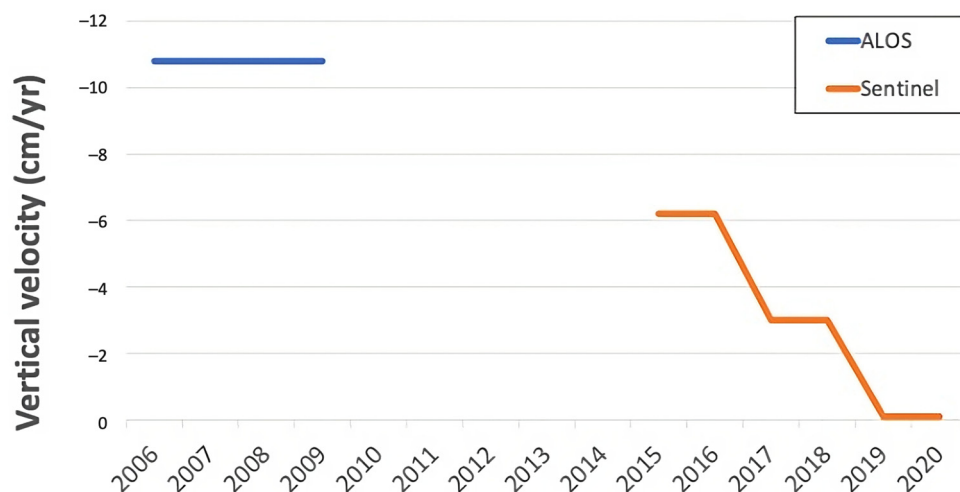
Figure 13. Evolution of the vertical displacement velocity from 2007 to 2020 at the location of cluster 17.

The deformation behaviour in the Bandung Basin is complex and influenced by various factors. It is clear that high-quality and reliable ancillary data such as the lithological condition of the subsurface, measurements of groundwater levels, and land use changes are essential when interpreting the displacement time series. Further in-depth analysis is required to better understand the subsidence-triggering factors related to groundwater extraction. This can be achieved by integrating InSAR data with numerical groundwater and geomechanical models, which is part of our ongoing research and out of the scope of this paper.

A possible limitation of the developed method is defining what "relevant" changes are. The PWLF technique was shown to be a simple yet robust way to characterize the principal changes that occurred in the analyzed cluster time series, which typically corresponded with a notable change in the displacement velocity. However, changes that occur gradually over time rather than abruptly could go undetected using this change detection method. When fitting the PWLF model to the clustered time series, a 15% improvement of the sum of squared residuals of the resulting model was chosen in order to capture the overall displacement behaviour, but also to avoid overfitting the model. This number may need to be adjusted depending on the displacement behaviour (i.e., seasonal behaviour or slowly deforming areas) or the quality of the InSAR dataset (i.e., noisy datasets) to detect relevant deformation changes.

7. Conclusions

In this study, we developed an innovative method to characterize InSAR displacement time series by clustering the time series, and then automatically analyzing changes in those timeseries. Our proposed workflow uses unsupervised data mining techniques to efficiently extract meaningful insights from large InSAR datasets exploiting the full displacement time histories. The method applies data-dimensionality reduction with UMAP and clustering with HDBSCAN, and can automatically detect changes in InSAR time series by applying the piecewise linear function analysis. This work represents the first peer-reviewed application of a UMAP + HDBSCAN pipeline to analyze InSAR time series. Most existing InSAR studies generally identify and classify ground motion based on time-averaged displacement trends. Even the most recent clustering methods on InSAR time series are not able to capture both the spatial and temporal changes of ground motion and require the users to manually identify the number of clusters and remove noisy signals. Compared to existing InSAR clustering approaches, our proposed method reduced the noise in the dataset, and automatically identified the number of clusters through the use of the novel application of UMAP + HDBSCAN.

Our InSAR analysis over the Bandung Basin revealed a maximum subsidence velocity of -18.7 cm/y averaged over the observation period between January 2015 to December 2020. We applied our clustering method to this dataset and identified deforming areas with various patterns of accelerations and decelerations. We show that the clustering approach is able to detect and cluster time series into groups with similar temporal behavior. We identified 12 clusters of subsidence, of which three experienced accelerating subsidence over an area of 22 km^2, four linear clusters covering a total of area of 52 km^2, and five clusters of decelerating subsidence over an area of 22 km^2. The regions with the most rapidly accelerating subsidence experienced a rate change of -0.03 cm/y to -2.40 cm/y (cluster 21) and -3.1 cm/y to -4.5 cm/y (cluster 25) during the observation period of January 2015 to December 2020, while the regions with the most rapidly decelerating subsidence experienced a rate change of -5.81 cm/y to -0.55 cm/y (cluster 17) and -3.72 cm/y to 0.62 cm/y (cluster 23) during the same observation time. The maximum total subsidence of 18 cm occurred in cluster 18 over an area of approximately 35 km^2. If such subsidence is predominantly due to groundwater extraction, then our measurements have important implications for groundwater sustainability and long-term aquifer sustainability within the Bandung Basin.

Our approach improves the identification of ground displacement changes by providing an objective and comprehensive analysis of surface dynamics over a large area. By precisely determining the temporal and spatial variations of displacement changes, we significantly advance the understanding of large deforming regions which can aid in disentangling deformation signals arising from various triggering factors. Unlike existing InSAR clustering methods, our approach minimizes user involvement and harnesses the power of open-source Python packages, making it easily adaptable to different regions for studying various phenomena. This uniqueness allows us to systematically investigate ground movements at both local and regional scales, ensuring the replicability and reliability of our analysis. This increased insight empowers us to interpret ground motion patterns more effectively and identify their root causes. The successful implementation of our novel approach in the urban region of Bandung demonstrates its potential as an automated technique capable of supporting risk assessments for a wide range of geological hazards caused by both natural events and human activities.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Erten, E.; Reigber, A.; Hellwich, O. Generation of Three-Dimensional Deformation Maps from InSAR Data Using Spectral Diversity Techniques. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 388–394. [CrossRef]
2. Bernardi, M.S.; Africa, P.C.; de Falco, C.; Formaggia, L.; Menafoglio, A.; Vantini, S. On the Use of Interferometric Synthetic Aperture Radar Data for Monitoring and Forecasting Natural Hazards. *Math. Geosci.* **2021**, *53*, 1781–1812. [CrossRef]
3. Ma, P.; Zheng, Y.; Zhang, Z.; Wu, Z.; Yu, C. Building Risk Monitoring and Prediction Using Integrated Multi-Temporal InSAR and Numerical Modeling Techniques. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *114*, 103076. [CrossRef]
4. Venmans, A.A.M.; op de Kelder, M.; de Jong, J.; Korff, M.; Houtepen, M. Reliability of InSAR Satellite Monitoring of Buildings near Inner City Quay Walls. *Proc. IAHS* **2020**, *382*, 195–199. [CrossRef]
5. Castellazzi, P.; Garfias, J.; Martel, R. Assessing the Efficiency of Mitigation Measures to Reduce Groundwater Depletion and Related Land Subsidence in Querétaro (Central Mexico) from Decadal InSAR Observations. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *105*, 102632. [CrossRef]
6. Bonì, R.; Bosino, A.; Meisina, C.; Novellino, A.; Bateson, L.; McCormack, H. A Methodology to Detect and Characterize Uplift Phenomena in Urban Areas Using Sentinel-1 Data. *Remote Sens.* **2018**, *10*, 607. [CrossRef]
7. Delgado Blasco, J.M.; Foumelis, M.; Stewart, C.; Hooper, A. Measuring Urban Subsidence in the Rome Metropolitan Area (Italy) with Sentinel-1 SNAP-StaMPS Persistent Scatterer Interferometry. *Remote Sens.* **2019**, *11*, 129. [CrossRef]
8. Parker, A.L.; Pigois, J.-P.; Filmer, M.S.; Featherstone, W.E.; Timms, N.E.; Penna, N.T. Land Uplift Linked to Managed Aquifer Recharge in the Perth Basin, Australia. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *105*, 102637. [CrossRef]
9. Cigna, F.; Tapete, D. Urban Growth and Land Subsidence: Multi-Decadal Investigation Using Human Settlement Data and Satellite InSAR in Morelia, Mexico. *Sci. Total Environ.* **2022**, *811*, 152211. [CrossRef]
10. Aslan, G.; Foumelis, M.; Raucoules, D.; Michele, M.D.; Bernardie, S.; Cakir, Z. Landslide mapping and monitoring using persistent scatterer interferometry (PSI) technique in the french alps. *Remote Sens.* **2020**, *12*, 1305. [CrossRef]
11. Fuhrmann, T.; Garthwaite, M.C. Resolving Three-Dimensional Surface Motion with InSAR: Constraints from Multi-Geometry Data Fusion. *Remote Sens.* **2019**, *11*, 241. [CrossRef]
12. Zhou, L.; Guo, J.; Hu, J.; Li, J.; Xu, Y.; Pan, Y.; Shi, M. Wuhan Surface Subsidence Analysis in 2015–2016 Based on Sentinel-1A Data by SBAS-InSAR. *Remote Sens.* **2017**, *9*, 982. [CrossRef]
13. Zheng, M.; Deng, K.; Fan, H.; Du, S. Monitoring and Analysis of Surface Deformation in Mining Area Based on InSAR and GRACE. *Remote Sens.* **2018**, *10*, 1392. [CrossRef]
14. Cigna, F.; Tapete, D. Present-day land subsidence rates, surface faulting hazard and risk in Mexico City with 2014–2020 Sentinel-1 IW InSAR. *Remote Sens. Environ.* **2021**, *253*, 112161. [CrossRef]
15. Cigna, F.; Tapete, D. Satellite InSAR survey of structurally-controlled land subsidence due to groundwater exploitation in the Aguascalientes Valley, Mexico. *Remote Sens. Environ.* **2021**, *254*, 112254. [CrossRef]
16. Aslan, G.; Cakir, Z.; Lasserre, C.; Renard, F. Investigating Subsidence in the Bursa Plain, Turkey, Using Ascending and Descending Sentinel-1 Satellite Data. *Remote Sens.* **2019**, *11*, 85. [CrossRef]
17. Hanssen, R. *Radar Interferometry Data Interpretation and Error Analysis*; Springer: Dordrecht, The Netherlands, 2021. [CrossRef]
18. Biggs, J.; Wright, T.J. How Satellite InSAR Has Grown from Opportunistic Science to Routine Monitoring over the Last Decade. *Nat. Commun.* **2020**, *11*, 3863. [CrossRef]
19. Hussain, E.; Novellino, A.; Jordan, C.; Bateson, L. Offline-Online Change Detection for Sentinel-1 InSAR Time Series. *Remote Sens.* **2021**, *13*, 1656. [CrossRef]
20. Festa, D.; Novellino, A.; Hussain, E.; Bateson, L.; Casagli, N.; Confuorto, P.; Del Soldato, M.; Raspini, F. Unsupervised Detection of InSAR Time Series Patterns Based on PCA and K-Means Clustering. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *118*, 103276. [CrossRef]
21. Chang, L.; Hanssen, R.F. A Probabilistic Approach for InSAR Time-Series Postprocessing. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 421–430. [CrossRef]
22. van de Kerkhof, B.; Pankratius, V.; Chang, L.; van Swol, R.; Hanssen, R.F. Individual Scatterer Model Learning for Satellite Interferometry. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 1273–1280. [CrossRef]
23. Cigna, F.; Tapete, D.; Casagli, N. Semi-Automated Extraction of Deviation Indexes (DI) from Satellite Persistent Scatterers Time Series: Tests on Sedimentary Volcanism and Tectonically-Induced Motions. *Nonlinear Process. Geophys.* **2012**, *19*, 643–655. [CrossRef]
24. Tapete, D.; Casagli, N. Testing Computational Methods to Identify Deformation Trends in RADARSAT Persistent Scatterers Time Series for Structural Assessment of Archaeological Heritage. In *Computational Science and Its Applications—ICCSA 2013*; Murgante, B., Misra, S., Carlini, M., Torre, C.M., Nguyen, H.-Q., Taniar, D., Apduhan, B.O., Gervasi, O., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; pp. 693–707. [CrossRef]
25. Berti, M.; Corsini, A.; Franceschini, S.; Iannacone, J.P. Automated Classification of Persistent Scatterers Interferometry Time Series. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 1945–1958. [CrossRef]
26. Mirmazloumi, S.M.; Gambin, A.F.; Palamà, R.; Crosetto, M.; Wassie, Y.; Navarro, J.A.; Barra, A.; Monserrat, O. Supervised Machine Learning Algorithms for Ground Motion Time Series Classification from InSAR Data. *Remote Sens.* **2022**, *14*, 3821. [CrossRef]

27. Kulshrestha, A.; Chang, L.; Stein, A. Use of LSTM for Sinkhole-Related Anomaly Detection and Classification of InSAR Deformation Time Series. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4559–4570. [CrossRef]

28. Khalili, M.A.; Voosoghi, B.; Guerriero, L.; Haji-Aghajany, S.; Calcaterra, D.; Di Martire, D. Mapping of Mean Deformation Rates Based on APS-Corrected InSAR Data Using Unsupervised Clustering Algorithms. *Remote Sens.* **2023**, *15*, 529. [CrossRef]

29. Ansari, H.; Ruβwurm, M.; Ali, M.; Montazeri, S.; Parizzi, A.; Zhu, X.X. InSAR Displacement Time Series Mining: A Machine Learning Approach. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 3301–3304. [CrossRef]

30. Handwerger, A.L.; Fielding, E.J.; Sangha, S.S.; Bekaert, D.P.S. Landslide Sensitivity and Response to Precipitation Changes in Wet and Dry Climates. *Geophys. Res. Lett.* **2022**, *49*, e2022GL099499. [CrossRef]

31. Bonì, R.; Meisina, C.; Poggio, L.; Fontana, A.; Tessari, G.; Riccardi, P.; Floris, M. Ground Motion Areas Detection (GMA-D): An Innovative Approach to Identify Ground Deformation Areas Using the SAR-Based Displacement Time Series. *Proc. IAHS* **2020**, *382*, 277–284. [CrossRef]

32. McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **2018**, *3*, 861. [CrossRef]

33. McInnes, L.; Healy, J.; Astels, S. Hdbscan: Hierarchical Density Based Clustering. *J. Open Source Softw.* **2017**, *2*, 205. [CrossRef]

34. Jiang, H.; Balz, T.; Cigna, F.; Tapete, D. Land Subsidence in Wuhan Revealed Using a Non-Linear PSInSAR Approach with Long Time Series of COSMO-SkyMed SAR Data. *Remote Sens.* **2021**, *13*, 1256. [CrossRef]

35. Schlögl, M.; Gutjahr, K.; Fuchs, S. The Challenge to Use Multi-Temporal InSAR for Landslide Early Warning. *Nat. Hazards* **2022**, *112*, 2913–2919. [CrossRef]

36. Bovenga, F.; Refice, A.; Pasquariello, G.; Nutricato, R.; Nitti, D. Identification and Analysis of Nonlinear Trends in InSAR Displacement Time Series. In *Microwave Remote Sensing: Data Processing and Applications*; SPIE: Bellingham, WA, USA, 2021; Volume 11861, p. 118610G. [CrossRef]

37. Park, S.-W.; Hong, S.-H. Nonlinear Modeling of Subsidence From a Decade of InSAR Time Series. *Geophys. Res. Lett.* **2021**, *48*, e2020GL090970. [CrossRef]

38. Abidin, H.Z.; Gumilar, I.; Andreas, H.; Murdohardono, D.; Fukuda, Y. On Causes and Impacts of Land Subsidence in Bandung Basin, Indonesia. *Environ. Earth Sci.* **2013**, *68*, 1545–1553. [CrossRef]

39. Gumilar, I.; Sidiq, T.; Meilano, I.; Bramanto, B.; Pambudi, G. Extensive Investigation of the Land Subsidence Impressions on Gedebage District, Bandung, Indonesia. IOP Conf. *Ser. Earth Environ. Sci.* **2021**, *873*, 012044. [CrossRef]

40. Tarigan, A.K.M.; Sagala, S.; Samsura, D.A.A.; Fiisabiilillah, D.F.; Simarmata, H.A.; Nababan, M. Bandung City, Indonesia. *Cities* **2016**, *50*, 100–110. [CrossRef]

41. Abidin, H.Z.; Andreas, H.; Gamal, M.; Wirakusumah, A.D.; Darmawan, D.; Deguchi, T.; Maruyama, Y. Land Subsidence Characteristics of the Bandung Basin, Indonesia, as Estimated from GPS and InSAR. *J. Appl. Geod.* **2008**, *2*, 167–177. [CrossRef]

42. Gumilar, I.; Abidin, H.Z.; Hutasoit, L.M.; Hakim, D.M.; Sidiq, T.P.; Andreas, H. Land Subsidence in Bandung Basin and Its Possible Caused Factors. *Procedia Earth Planet. Sci.* **2015**, *12*, 47–62. [CrossRef]

43. Widodo, J.; Naryanto, H.S.; Wisyanto; Hidayat, N.; Putra, A.P.; Izumi, Y.; Perissin, D.; Sri Sumantyo, J.T. Land Subsidence Assessment of Bandung City, Indonesia in Geological Perspective, Based on Interferometric SAR Using C-Band Data. In Proceedings of the 2021 Photonics & Electromagnetics Research Symposium (PIERS), Hangzhou, China, 21–25 November 2021; pp. 2377–2381. [CrossRef]

44. Chaussard, E.; Amelung, F.; Abidin, H.; Hong, S.-H. Sinking Cities in Indonesia: ALOS PALSAR Detects Rapid Subsidence Due to Groundwater and Gas Extraction. *Remote Sens. Environ.* **2013**, *128*, 150–161. [CrossRef]

45. Tirtomihardjo, H. Chapter 10—Groundwater Environment in Bandung, Indonesia. In *Groundwater Environment in Asian Cities*; Shrestha, S., Pandey, V.P., Shivakoti, B.R., Thatikonda, S., Eds.; Butterworth-Heinemann: Oxford, UK, 2016; pp. 193–228, Reprinted from Groundwater Environment in Asian Cities, Haryadi Tirtomihardjo, Chapter 10—Groundwater Environment in Bandung, Indonesia, 193-228, Copyright (2016), with permission from Elsevier. [CrossRef]

46. Rosen, P.A.; Gurrola, E.; Sacco, G.F.; Zebker, H. The InSAR Scientific Computing Environment. In Proceedings of the EUSAR 2012, 9th European Conference on Synthetic Aperture Radar, Nuremberg, Germany, 23–26 April 2012; pp. 730–733.

47. ISCE2: Interferometric synthetic aperture radar Scientific Computing Environment (ISCE), v2. Available online: https://github.com/isce-framework/isce2 (accessed on 1 June 2021).

48. Yunjun, Z.; Fattahi, H.; Amelung, F. Small Baseline InSAR Time Series Analysis: Unwrapping Error Correction and Noise Reduction. *Comput. Geosci.* **2019**, *133*, 104331. [CrossRef]

49. Manzo, M.; Ricciardi, G.P.; Casu, F.; Ventura, G.; Zeni, G.; Borgström, S.; Berardino, P.; Del Gaudio, C.; Lanari, R. Surface Deformation Analysis in the Ischia Island (Italy) Based on Spaceborne Radar Interferometry. *J. Volcanol. Geotherm. Res.* **2006**, *151*, 399–416. [CrossRef]

50. Notti, D.; Herrera, G.; Bianchini, S.; Meisina, C.; García-Davalillo, J.C.; Zucca, F. A Methodology for Improving Landslide PSI Data Analysis. *Int. J. Remote Sens.* **2014**, *35*, 2186–2214. [CrossRef]

51. Wright, T.J.; Parsons, B.E.; Lu, Z. Toward Mapping Surface Deformation in Three Dimensions Using InSAR. *Geophys. Res. Lett.* **2004**, *31*. [CrossRef]

52. Cigna, F.; Esquivel Ramírez, R.; Tapete, D. Accuracy of Sentinel-1 PSI and SBAS InSAR Displacement Velocities against GNSS and Geodetic Leveling Monitoring Data. *Remote Sens.* **2021**, *13*, 4800. [CrossRef]

53. Rani, R.; Khurana, M.; Kumar, A.; Kumar, N. Big Data Dimensionality Reduction Techniques in IoT: Review, Applications and Open Research Challenges. *Clust. Comput.* **2022**, *25*, 4027–4049. [CrossRef]

54. Lovrić, M.; Đuričić, T.; Tran, H.T.N.; Hussain, H.; Lacić, E.; Rasmussen, M.A.; Kern, R. Should We Embed in Chemistry? A Comparison of Unsupervised Transfer Learning with PCA, UMAP, and VAE on Molecular Fingerprints. *Pharmaceuticals* **2021**, *14*, 758. [CrossRef] [PubMed]

55. Keogh, E.; Mueen, A. Curse of Dimensionality. In *Encyclopedia of Machine Learning and Data Mining*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2017; pp. 314–315. [CrossRef]

56. Venna, J.; Kaski, S. Local Multidimensional Scaling. *Neural Netw.* **2006**, *19*, 889–899. [CrossRef]

57. Sainburgh, T.; McInnes, L.; Gentner, T.Q. Parametric UMAP Embeddings for Representation and Semisupervised Learning. *Neural Comput.* **2021**, *33*, 2881–2907. [CrossRef]

58. Kolmogorov–Smirnov Test. In *The Concise Encyclopedia of Statistics*; Springer: New York, NY, USA, 2008; pp. 283–287. [CrossRef]

59. Available online: https://github.com/scikit-learn-contrib/hdbscan (accessed on 13 July 2022).

60. Kendall, M.G. A New Measure of Rank Correlation. *Biometrika* **1938**, *30*, 81–93. [CrossRef]

61. Jekel, C.; Venter, G. Pwlf: A Python Library for Fitting 1D Continuous Piecewise Linear Functions; 2019. Available online: https://github.com/cjekel/piecewise_linear_fit_py (accessed on 29 August 2022).

62. Urgilez Vinueza, A.; Handwerger, A.L.; Bakker, M.; Bogaard, T. A New Method to Detect Changes in Displacement Rates of Slow-Moving Landslides Using InSAR Time Series. *Landslides* **2022**, *19*, 2233–2247. [CrossRef]

63. Luo, Q.; Li, J.; Zhang, Y. Monitoring Subsidence over the Planned Jakarta–Bandung (Indonesia) High-Speed Railway Using Sentinel-1 Multi-Temporal InSAR Data. *Remote Sens.* **2022**, *14*, 4138. [CrossRef]

64. Tolomei, C.; Lugari, A.; Salvi, S. Bandung (Indonesia) Area InSAR Mean Velocity Maps [Data Set]. Zenodus. 2016. Available online: https://zenodo.org/record/49676 (accessed on 6 July 2023).

65. Prasetyo, Y.; Tetuko, J.; Ismullah, I.H.; Abidin, H.Z.; Wikantika, K. Data optimization in Permanent Scatterer Interferometric Synthetic Aperture Radar (PS-INSAR) technique for land subsidence estimation. *INA-Rxiv* **2013**. [CrossRef]

66. Duan, L.; Gong, H.; Chen, B.; Zhou, C.; Lei, K.; Gao, M.; Yu, H.; Cao, Q.; Cao, J. An Improved Multi-Sensor MTI Time-Series Fusion Method to Monitor the Subsidence of Beijing Subway Network during the Past 15 Years. *Remote Sens.* **2020**, *12*, 2125. [CrossRef]

67. Rusli, S.R.; Weerts, A.H.; Taufiq, A.; Bense, V.F. Estimating water balance components and their uncertainty bounds in highly groundwater-dependent and data-scarce area: An example for the Upper Citarum basin. *J. Hydrol. Reg. Stud.* **2021**, *37*, 100911. [CrossRef]

68. Du, Z.; Ge, L.; Ng, A.H.-M.; Zhu, Q.; Yang, X.; Li, L. Correlating the Subsidence Pattern and Land Use in Bandung, Indonesia with Both Sentinel-1/2 and ALOS-2 Satellite Images. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *67*, 54–68. [CrossRef]

69. Ar Rahiem, M.M. Development of an Interactive WebGIS Platform for the Visualization of Hydrogeological Information, Bandung Basin, Indonesia. Master's Thesis, TU Darmstadt, Darmstadt, Germany, 2021.

70. Ohgaki, S.; Takizawa, S.; Herath, G.; Kataoka, Y.; Hara, K.; Kathiwada, N.R.; Moon, H.-J. *The State of the Groundwater; Sustainable Groundwater Management in Asian Cities*; Institute for Global Environmental Strategies: Hayama, Japan, 2006; pp. 12–21. Available online: https://www.jstor.org/stable/resrep00865.12 (accessed on 15 March 2023).

71. Sidiq, T.; Gumilar, I.; Abidin, H.Z. Land Subsidence Induced by Agriculture Activity in Bandung, West Java Indonesia. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *389*, 012024. [CrossRef]

72. Bonì, R.; Pilla, G.; Meisina, C. Methodology for Detection and Interpretation of Ground Motion Areas with the A-DInSAR Time Series Analysis. *Remote Sens.* **2016**, *8*, 686. [CrossRef]

73. Cohen-waeber, J.; Bürgmann, R.; Chaussard, E.; Giannico, C.; Ferretti, A. Spatiotemporal Patterns of Precipitation-modulated Landslide Deformation from Independent Component Analysis of InSar Time Series. *Geophys. Res. Lett.* **2018**, *45*, 1878–1887. [CrossRef]

74. Karimzadeh, S.; Matsuoka, M. Ground Displacement in East Azerbaijan Province, Iran, Revealed by L-band and C-band InSAR Analyses. *Sensors* **2020**, *20*, 6913. [CrossRef]

75. Allaoui, M.; Kherfi, M.L.; Cheriet, A. Considerably Improving Clustering Algorithms Using UMAP Dimensionality Reduction Technique: A Comparative Study. In *Image and Signal Processing. ICISP 2020*; El Moataz, A., Mammass, D., Mansouri, A., Nouboud, F., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2020; Volume 12119. [CrossRef]

76. McInnes, L.; Healy, J. Accelerated Hierarchical Density Clustering. In Proceedings of the 2017 IEEE International Conference on Data Mining Workshops (ICDMW), New Orleans, LA, USA, 18–21 November 2017; pp. 33–42. [CrossRef]

77. Ge, L.; Ng, A.H.-M.; Li, X.; Abidin, H.Z.; Gumilar, I. Land Subsidence Characteristics of Bandung Basin as Revealed by ENVISAT ASAR and ALOS PALSAR Interferometry. *Remote Sens. Environ.* **2014**, *154*, 46–60. [CrossRef]