Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 11 (4): 610-620, 2011
ISSN 1812-5654 / DOI: 10.3923/jas.2011.610.620

A Clustering Approach for Studying Ground Deformation Trends in Campania Region through PS-InSAR™ Time Series Analysis

Gabriella Milone and Germana Scepi
Department of Mathematics and Statistics, University of Naples-Federico II, Italy

Abstract: The recent development of remote sensing techniques brings new possibilities of precise ground movement measurements. One of these possibilities is the PS-InSAR™ method. It allows detecting small and long period ground deformations on large areas. In particular, the PS-InSAR technique involves interferometric phase comparison of several radar images of the same scene taken at different times along the same orbit by satellite radar sensors. The ground deformations occurring in the Campania Region has been recently investigated by applying this interferometric technique in a peculiar project developed by Italian Ministry of Environment. The outcome consists in a very large and complex database that cannot be easily analysed with traditional tools. In order to improve the interpretability of this dataset and to support the study of the geologists of this project, we apply a clustering data mining approach on a sample area and identify homogeneous clusters of ground deformation velocity trends. We report in this paper the main results of this analysis, showing how the chosen clustering method allows a coherent geological interpretation of this large and scattered dataset.

Key words: Temporal data mining, temporal clustering, partitioning algorithms, PS-InSAR technique

INTRODUCTION

In this paper, the PS-InSAR™ data were used to study with a statistical approach, small, long lasting ground deformations which occurred in the Campania Region (Southern Italy) during the years 1992-2000.

Campania region has a complex geological structure. It is characterized by an intense urbanization and by the interplay of several geodynamic processes related to the presence of active volcanoes (Vesuvius, Phlegraean Fields and Ischia), seismogenic structures characterised by high magnitude earthquake (e.g., 6.9 Mw in 1980), widespread landslides and geological instability, sinkhole, subsidence and long- to short-term tectonic warping, which produce a complex ground deformation pattern. These factors make the studied region particularly threatened with geological risks.

The ground deformations occurring in the Campania Region has been investigated (Vilardo *et al*., 2009) in the context of the activity of PODIS Project (founded by European Union QCS 2000-2006 PON-ATAS) of MATTM (Environment Ministry) and Campania Region. In this project the landslide phenomenon was detected using the PS-InSAR™ method. PS-InSAR technique (Ferretti *et al*., 2000, 2001) is a dynamically developed branch of satellite radar interferometry. It exploits a set of dozens satellite SAR images in order to detect small ground deformations for large areas. PS-InSAR technique derives information only about ground movements for stable radar targets, so called Permanent Scatterers (PS) points. They correspond with rock outcrops or man-made features on the ground like buildings, bridges, etc.

The Permanent Scatterers Synthetic Aperture Radar Interferometry has been applied on ERS 1 and ERS 2 satellite radar images, whose availability for the study area is dated between June 1992 and December 2000 during the period of regular activity of the two satellites. ERS 1 has operated regularly from 25/7/1991 to 10/3/2000, ERS 2 has started regular acquisitions in May 1995 and it is still operational, but has no more been used for technical problems, in the PS-InSAR radar interferometry, from the beginning of 2001.

In order to improve the interpretability of this large temporal database formed by PS deformation time series, we apply a clustering data mining algorithm on a sample area related to the central sector of the Campania region, between the towns of Benevento and Avellino (Fig. 1). This area has been chosen as study area because it includes two large urban areas, it is characterised by a complex landscape (wide valleys and mountain ranges) and by high seismic and landslide risks. Our main aim is to classify the time series of measured displacements for identifying homogenous regions in which there is a significant time dependent component to the deformation field.

Corresponding Author: Gabriella Milone, Department of Mathematics and Statistics, University of Naples-Federico II, Italy

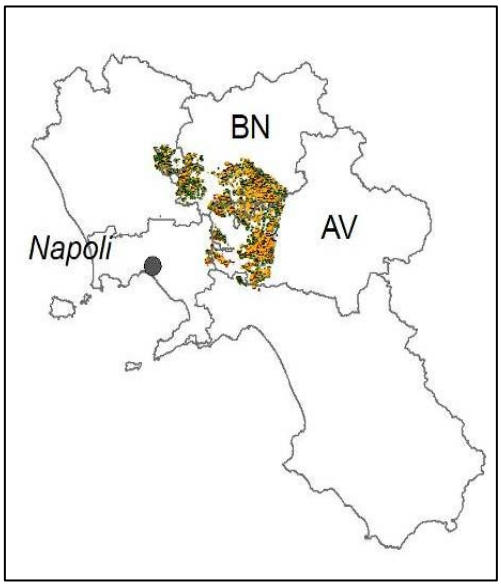

Fig. 1: Study area location (coloured dots) in the Campania region (Italy). (City legend: BN, Benevento; AV, Avellino)

Clustering is perhaps the most frequently used algorithm in the framework of Knowledge Discovery in Temporal Databases (KDTD) being useful in its own right as an exploratory technique. Many temporal data mining applications make use of clustering according to similarity and optimization of temporal set functions. However, clustering algorithms of the traditional type are often limited in dealing with large temporal data sets (Milone, 2008) and the proprieties (such as handling outliers, time and space complexity, interpretability of results and so on) of the different algorithms must be taken into account for the choice among the different alternatives (Scepi, 2009). A priori information on the structure of the data can help us as well.

On our data set, we apply a partitioning method the CLARA algorithm. Partitioning algorithms are the most commonly used algorithms in the context of data mining, where hierarchical algorithms show an high computational complexity. In this family of algorithms, we choose to apply a k-medoids algorithm, CLARA, because for its own proprieties it performs very well on our data set respect to others (such as for example k-means) and we discuss the differences. For the choice of the input number of cluster, we look at several statistical information based on the indices computed by CLARA but, for a suggestion, we also consider the results obtained implementing the k-means algorithm on our data set. A priori information on the geological structure of the area help us on the choice of the number of k also. We show how the clustering algorithm enriches the geological interpretability of the results and allows a coherent geological interpretation of a very large and scattered information, which cannot be easily analysed otherwise.

MATERIALS AND METHODS

The processing technique of satellite radar images: The PS-InSAR™ (Permanent Scatterers Interferometry Synthetic Aperture Radar) was evolved by scientists from the Politecnico di Milano (POLIMI) in 1999 (Ferretti *et al*., 2001) and it is based on a multi-interferogram approach. In particular, this technique involves the interferometric phase comparison of several radar images of the same scene (a portion of the earth surface, wide 100x100 km) taken at different times along the same orbit by the satellite radar sensors. ERS 1 and ERS 2 satellites orbit at an elevation of 780 km and take on the same image every 35 days. The critical deformation rate is 2.8 cm in 35 days, or 0.8 mm in a day. Accuracy of PS-InSAR technique depends on the number of radar images used in processing. In order to obtain good results about 20 images are required as a minimum. Not all the acquired images are suitable for interferometric analysis, so that the time interval between two consecutive images can be longer.

The processing technique is based on the identification of the radar benchmarks, named Permanent Scatterers, which are stable natural reflectors (rock outcrops, buildings and urban structures), characterized by stable individual radar-bright and radar-phase over long temporal series of interferometric SAR images. This technique has proved useful at exploring slow movements of the earth surface, induced by natural and man-induced phenomena such as land subsidence, landslides, seismic faults, volcanic uplifts and tectonic deformations, with very high spatial resolution, both at a local and at a regional scale (Colasanti *et al*., 2003; Farina *et al*., 2006).

The final results are:

- A map of the PS identified in the image and their coordinates: latitude, longitude and precise elevation (accuracy on elevation better than 1 m); each PS is labelled by a code
- Average LOS deformation rate of every PS (expressed as average velocity with an accuracy usually between 0.1 and 1 mm year^{-1}, depending on the number of available interferograms and on the phase stability of each single PS)

Code	Lat	Lon	Vel	Coherence	19920424	19920529	19920807	19920911	19921
DD022	44,442198	11,355714	-0,90	0,83	3,3	4,3	0,1	6,5	
DD024	44,442278	11,355214	-2,00	0,62	2,4	6,5	6,3	0,1	
DD025	44,442278	11,355184	-2,10	0,62	6	23,1	6	-1,5	
DD029	44,442018	11,356654	-0,70	0,76	1	-0,4	-0,5	2,9	
DD030	44,442158	11,355714	0,50	0,83	-5,2	-2,7	-4,6	8,3	
DD032	44,441838	11,357554	-1,70	0,65	7,1	7,6	3,3	4	
DD033	44,441978	11,356694	-0,40	0,86	2,5	0,9	-0,7	4,3	
DD034	44,442088	11,355934	-0,50	0,85	3,6	4,1	-0,6	5,6	
DD035	44,442128	11,355704	-0,20	0,67	5,4	4,1	7,1	15,5	
DD036	44,442108	11,355834	-0,50	0,62	4	4,1	3	1,5	
DD041	44,442058	11,355904	-1,20	0,72	4,8	6,9	-0,1	8,1	
DD042	44,442098	11,355634	0,10	0,63	2,9	2,8	-3	2,3	
DD046	44,442068	11,355634	-0,10	0,85	3,5	3,8	-1,9	0,6	
DD050	44,442078	11,355284	-1,00	0,63	1,9	6,2	3,2	6,7	
DD052	44,441818	11,356744	-1,00	0,67	8,1	10,2	-5	3,3	
DD056	44,441958	11,355614	-0,20	0,82	2,9	1,3	-2	3,7	

Fig. 2: Database output table

- Displacement time series showing the relative (i.e. with respect to a unique reference image) LOS position of PS in correspondence of each SAR acquisition. The time series represent, therefore, the motion component of the PS in the direction of the Line of Sight (LOS) as a function of the time (accuracy on single measurements usually ranging from 1 to 3 mm)

As in all differential interferometry applications, the results are not absolute both in time and space, but the deformation data are referred to the master image (in time) and the results are computed with respect to a reference point of known elevation and motion (in space)

The traditional radar interpherometry can be affected by atmospheric noise, but the PS-InSAR™ technique is capable to highly reduce the effect of time decorrelation and atmospheric phase (Ferretti *et al*., 2001).

The PS database of our analysis: Our studied scene encompasses the area of Campania (Southern Italy) between the towns of Benevento and Avellino. We have selected the data referred to this area for a first sample study. The sample is representative of the different and more important geological structures existing in Campania and it is formed by high seismic and landslide risks zones. The availability of the data is from June 1992 to December 2000, period of regular activity of the two satellites.

The PS points, registered in this area, constitute the benchmarks of a geodetic network, already available on the territory, corresponding, in general, to rock outcrops and man-made objects characterized by various geometry and material, such as building roofs and terraces, lamps, statues, walls, belfries, manholes, sewage covers, landscaping, etc.

Therefore, our database is formed by 18.452 PS characterized by a coherence value higher than 0.80, with a time series made by 72 observations (Fig. 2). The average velocities in the direction of the Line of Sight (LOS) vary between + 5,90 and -15,96 mm year^{-1}. Each PS point is defined by the code, the coordinates (latitude = North, longitude = East), the average velocity of ground deformation expressed in mm/year (negative rates indicate subsidence, positive rates indicate uplift along LOS), the coherence value (reliability index of series) and the time-series of 72 displacement data. The distance is centred, for each point, with respect to the distance that the same point has with the satellite in the image termed as master.

The time series of the measured displacements (Fig. 3) allows us to identify regions in which there is a significant time dependent component to the deformation field and enables the analysis of the deformation evolution of the PS over time. In order to identify clusters of homogenous zones respect to the ground deformation velocity trends, a clustering data mining approach has been used.

Clustering algorithms: Clustering is the subject of active research in several fields such as statistics, pattern recognition and machine learning. Data mining adds to clustering the complications of very large datasets with very many attributes of different types. This imposes unique computational requirements on relevant clustering algorithms. A variety of algorithms have recently emerged that meet these requirements and have been successfully applied to real-life data mining problems (Berkhin, 2006).

Traditionally, clustering techniques are divided into hierarchical and partitioning. Hierarchical clustering is further subdivided into agglomerative and divisive. The basics of hierarchical clustering include the Lance-Williams formula, the idea of conceptual clustering,

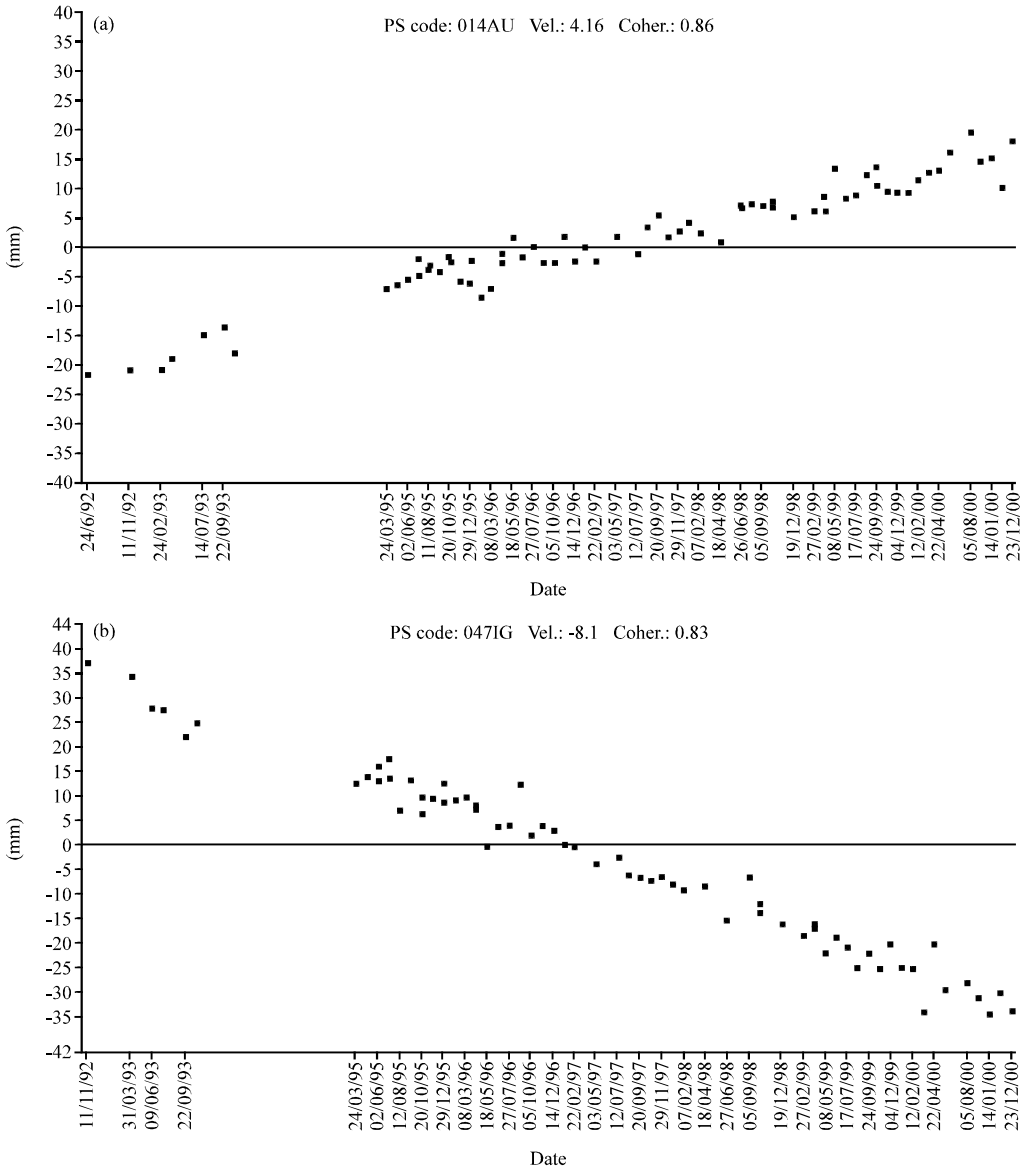

Fig. 3: PS's time series plottings: (a) uplifting trend and (b) subsidence trend

(now) classic algorithms, such that SLINK and COBWEB and also newer algorithms such as CURE and CHAMELEON.

The hierarchical algorithms are widely used for their flexibility (handling of any forms of similarity or distance) and applicability to any attribute types. One advantage of the hierarchical algorithms is that the number of clusters is not required to be provided as a parameter. Furthermore, they have legible results in terms of dendrograms. However, the quadratic computational complexity restricts their application to small data sets.

While the hierarchical algorithms build clusters gradually, the partitioning algorithms learn clusters directly. In doing so, they either try to discover clusters by iteratively relocating points between subsets, or try to identify clusters as areas densely populated with data. These algorithms are defined as Partitioning Relocation Methods. They are further categorized into probabilistic clustering (EM framework, algorithms SNOB, AUTOCLASS, MCLUST), k-medoids methods (algorithms PAM, CLARA, CLARANS and its extension) and k-means methods. Such methods concentrate on how well the points fit into their clusters and tend to build clusters of proper convex shapes.

Different applications in data mining have demonstrated that Partitioning Methods are very useful

in restricting very large universes. This leads us to consider these algorithms as the real data discovery tools and to choose our clustering method within this family of methods.

Partitioning algorithms: The optimization based partitioning algorithms typically represent clusters by a prototype. The objects are assigned to the cluster represented by the most similar prototype. An iterative control strategy is used to optimize the whole clustering such that the average or squared distances of the objects to its prototypes are minimized. These clustering algorithms are effective at determining a good clustering, if the clusters are of convex shape, similar size and density and if their number k can be reasonably estimated. Depending on the kind of prototypes, one can distinguish k-means, k-modes and k-medoids algorithms.

In k-means algorithm (MacQueen, 1967), the prototype, called the center, is the mean value of all objects belonging to a cluster. K-means clustering has been one of the popular clustering algorithm because it is one of the simplest unsupervised learning algorithms. It is considered an efficient algorithm for the time complexity. However, the result strongly depends on the initial guess of centroids and computed local optimum is known to be a far cry from the global one. Furthermore it requires several passes on the entire dataset, which can make it very expensive for large datasets as the dataset in our application and it is sensitive to the outliers and noises, often found in data relating to classified images. The k-medoids approach is more robust in this aspect.

The k-modes algorithm (Huang, 1997) extends the k-means paradigm to categorical domains. For k-medoids algorithms, the prototype, called the medoid, is the most centrally located object of a cluster. A versions of k-medoids methods is the algorithm PAM (Partitioning Around Medoids).

The PAM-algorithm (Kaufman and Rousseeuw, 1990) is based on the search for k representative objects or medoids among the objects of the dataset. These objects should represent the structure of the data. After finding a set of k medoids, k clusters are constructed by assigning each object to the nearest medoid. The goal is to find k representative objects which minimize the sum of the dissimilarities of the objects to their closest representative object.

Compared to k-means algorithm, PAM has the following features:

- It operates on the dissimilarity matrix of the given data set or when it is presented with an nxp data matrix, the algorithm first computes a dissimilarity matrix
- It is more robust, because it minimizes a sum of dissimilarities instead of a sum of squared Euclidean distances
- It provides a novel graphical display, the silhouette plot, which allows the user to select the optimal number of clusters

However, PAM lacks in scalability for very large databases and it present high time and space complexity.

Several fast partitioning-based clustering algorithms have been proposed in the literature, including Clustering LARge Applications (CLARA). The CLARA algorithm (Kaufman and Rousseeuw, 1990) is one of the popular clustering algorithms used today in data mining applications. This algorithm works on a randomly selected subset of the original data and produces near accurate results at a faster rate than other clustering algorithms. Compared to PAM, CLARA can deal with much larger data sets. It also tries to find k representative objects that are centrally located in the cluster. Internally, this is achieved by considering data subsets of fixed size, so that the overall computation time and storage requirements become linear in the total number of objects rather than quadratic. In PAM the collection of all pair-wise distances between objects is stored in the central memory, thereby consuming $O(n^2)$ memory space. Therefore PAM cannot be used for large values of n. To avoid this problem CLARA does not compute the entire dissimilarity matrix at a time.

Based on the above considerations, we choose this algorithm for our application and we observe that it fit very well to our dataset. We will shown in detail in the next section the algorithm chosen.

We underline that, in the context of large data applications, several other clustering algorithms have been proposed by Chih-Ping *et al.* (2003) for an empirical comparison among Fast Clustering Algorithms for Large Data Sets) like Clustering Large Applications based upon RANdomized Search (CLARANS), developed by Ng and Han (1994) and genetic algorithm based clustering methods (Estivill-Castro and Murray, 1997).

CLARA algorithm: CLARA is a combination of a sampling approach and the PAM algorithm. Instead of finding medoids, each of which is the most centrally located object in a cluster, for the entire data set, CLARA draws a sample from the data set and uses the PAM algorithm to select an optimal set of medoids from the sample. This is achieved by considering sub-datasets of fixed size, so that the time and storage requirements become linear rather than quadratic. Each sub-dataset is partitioned into k clusters using the same algorithm as in the PAM function. Once, k representative objects have

been selected from the sub-dataset, each object of the entire dataset is assigned to the nearest medoid. The sum of the dissimilarities of the objects to their closest medoid is used as a measure of the quality of the clustering. The sub-dataset, for which the sum is minimal, is retained. A further analysis is carried out on the final partition. Each sub-dataset is forced to contain the medoids obtained from the best sub-dataset until then. Randomly drawn objects are added to this set until the sample size has been reached.

Since, CLARA adopts a sampling approach, the quality of its clustering results depends greatly on the size of the sample. When the sample size is small, CLARA's efficiency comes at the cost of clustering quality.

CLARA can efficiently deal with large datasets (Chih-Ping *et al*., 2003). The clustering quality of CLARA is quite good, similar to the CLARANS algorithm, when the data size is more than 3000. In terms of execution time, as the data size increases, CLARA increases its execution time slightly and, eventually, outperforms other fast partitioning-based clustering algorithms. The insensitivity of CLARA to data size is attributed to its execution time being greatly influenced by a sample size that is a function of k and not of n (number of observations). As a sampling approach, CLARA is less susceptible to degree of cluster distinctness and level of data randomness, both in clustering quality and execution time.

However, when the number of clusters increases, the execution time of CLARA increases significantly and if the cluster sizes are asymmetric, the quality degrades.

For the properties just outlined, dealing with a small number of clusters, the CLARA algorithm performs very well on our dataset. We will show its powerful in the definition of coherent geological clusters.

We have used the following CLARA routine in MatLab™ language:

Required input arguments:

- x: Data matrix (rows = observations, columns = variables)
- kclus: No. of desired clusters
- vtype: Variable type vector (length equals number of variables)

Possible values are:

- Asymmetric binary variable (0/1)
- Nominal variable (includes symmetric binary)
- Ordinal variable
- Interval variable

Optional input arguments:

- metric: Metric to be used (default Euclidian (eucli) or mixed (mixed))

Possible values are:

- 'eucli' Euclidian (all interval variables)
- 'manha' Manhattan
- 'mixed' Mixed (not all interval variables)

We define:

- nsamp: Number of samples to be drawn from the dataset
- sampsize: Number of observations in each sample (should be higher than the number of clusters and lower than the number of observations)
- I/O:result=clara(x,kclus,vtype,'eucli',5,40+2*kclus)

The output of CLARA is a structure containing:

- result.dysobs: dissimilarities for each observation with the medoids
- result.metric: used metric
- result.number: number of observations
- result.idmed: Id of medoid observations
- result.ncluv: A vector with length equal to the number of observations, giving for each observation the number of the cluster to which it belongs
- result.obj: Objective function for the best subsample
- result.clusinf: Matrix, each row gives numerical information for one cluster. These are the cardinality of the cluster (number of observations), the maximal and average dissimilarity between the observations in the cluster and the cluster's medoid, the diameter of the cluster (maximal dissimilarity between two observations of the cluster) and the separation of the cluster (minimal dissimilarity between an observation of the cluster and an observation of another cluster)
- result.sylinf: Matrix based on the best subsample, with for each observation i of this subsample the cluster to which i belongs, as well as the neighbour cluster of I (the cluster, not containing i, for which the average dissimilarity between its observations and i is minimal) and the silhouette width of i

This function is part of LIBRA: the Matlab Library for Robust Analysis, available at: http://wis.kuleuven.be/stat/robust.html

RESULTS AND DISCUSSION

For a first empirical discussion, we have applied the k-means algorithm and the CLARA algorithm on our data set. As we don't know a priori the number of clusters, we have decided to evaluate the results obtained with different (k = 4, k = 5 and k = 6) number of clusters.

We have considered some main indicators: the variance within and the variance between (for k-means); the distance between each object and its medoid and the separation of the clusters (for CLARA). Furthermore we have computed the Calinski and Harabasz index for the different classification obtained by the k-means algorithm and the average silhouette statistic (Kaufman and Rousseeuw (1990) for the different results obtained by CLARA (Scarso, 2008).

The Calinski and Harabasz index is defined as:

$$c_k = \frac{trB_k/k-1}{trW_k/n-k}$$

where, $tr(B_k)$ is the trace of the cluster matrix between the groups and $tr(W_k)$ is the trace of the cluster matrix within the groups. The optimal number of cluster is the value of k that maximizes c_k . For the k-means clustering on our dataset this value (computed by considering k varying from 2 to 6) is 4. The silhouette statistic for each observation i is defined as:

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

where, a_i is the average dissimilarity between the observation i and all the observations of the cluster to which i belongs and b_i is the minimum of all the average dissimilarities between the observation i and all the observations of each other cluster.

The CLARA algorithms furnishes the average silhouette that is the average of all the s_i. The optimal number of cluster is the value of k that maximizes this quantity. On our dataset this value is 4. Therefore, for both the algorithms, these indices suggest 4 as the optimal number of clusters. The information resulting from the reality of the ground confirms this hypothesis, even if interesting information could be found in the classification of PS in 5 and 6 groups. Therefore we decide to consider 4 as the input number of classes.

For the reasons expressed above, we consider more accurate the CLARA results and we show and describe the groups (Table 1) obtained by the CLARA algorithm.

The clustering analysis of the time series related to 18.452 PS allowed to recognize four different types of ground deformation trends (Fig. 4a-d and Table 2). In analyzing the geological causes of ground deformation through the PS data, other type of variables (economic, demographical) have not been considered because they are not linked to the studied phenomena.

Table 1: Statistical characteristics of the 4 classes

Cluster	N	N (%)	A	B	C
1	10823	59	35.01	23.83	1.78
2	1564	8	35.47	26.24	1.80
3	2988	16	268.92	32.87	13.67
4	3077	17	108.61	27.68	5.52

N = No. of cluster; A = Maximum similarities from median, B = Diameter of cluster, C = Separation of clusters

Table 2: Ground deformation characteristics of the 4 classes

Class	Trend	Average LOS velocity
Class 1	Stability or very light subsidence	-1.52 to +0.89 mm year^{-1}
Class 2	Slower and variable subsidence	+0.44 to -2.52 mm year^{-1}
Class 3	Faster subsidence	-0.93 to -15.96 mm year^{-1}
Class 4	Uplift	-0.47 to + 5.90 mm year^{-1}

In particular class 1 shows a trend characterized by stability or by a very light subsidence, with average LOS velocity values from -1.52 to + 0.89 mm year^{-1}. Two classes describe a subsidence trend characterised by different rates: slower and variable for class 2 (0.44 to -2.52 mm year^{-1}) and faster for class 3 (-0.93 to -15.96 mm year^{-1}). Class 4 shows an uplift trend, with average LOS velocity values from -0.47 to + 5.90 mm year^{-1}.

Geological interpretation: The clustering analysis of the time series allows a coherent geological interpretation of a very large and scattered dataset, which cannot be easily analyzed otherwise.

The clustered PS are spatially defined by coordinates, so that the interpretation of their deformation trend can be referred to a map showing the spatial distribution of the different ground deformation trend. The cluster individuation and the spatial mapping of the deformation trends is a basic data in the geological interpretation of the active ground deformation of the studied territory and in the definition of the natural risk (landsliding, subsidence, seismicity, morphotectonics). As a matter of fact the recognized ground deformation trends are the results of combined morphological, tectonic and anthropic processes, acting on the territory at different scale and intensity and the clustering method combined with the interferometric technique allows the evaluation of their rates and location with a very high precision.

The studied scene encompasses the central sector of Campania between Benevento and Avellino towns. The deformation trend of clustered PS can be mapped showing the spatial distribution of the different ground deformation trend in the study area.

In Fig. 5 the distribution of PS clusters is showed with reference to the main tectonic elements. The PS of classes 2 and 3 characterizes eastern sector (river Sabato valley) of study area showing variable rates of subsidence

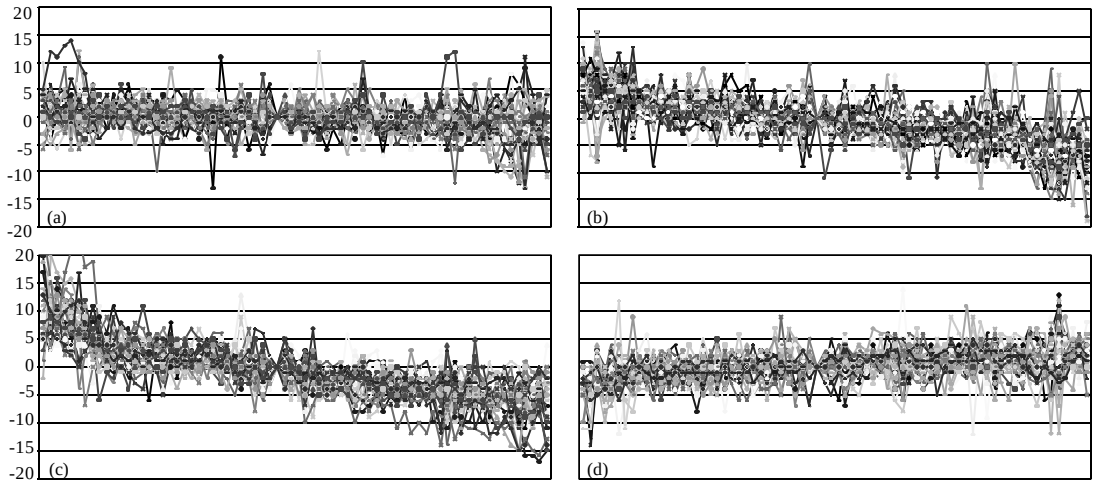

Fig. 4: Clustering classes trend (x axis = number of observations; y axis deformation in mm). (a) Class 1, (b) Class 2, (c) Class 3 and (d) Class 4

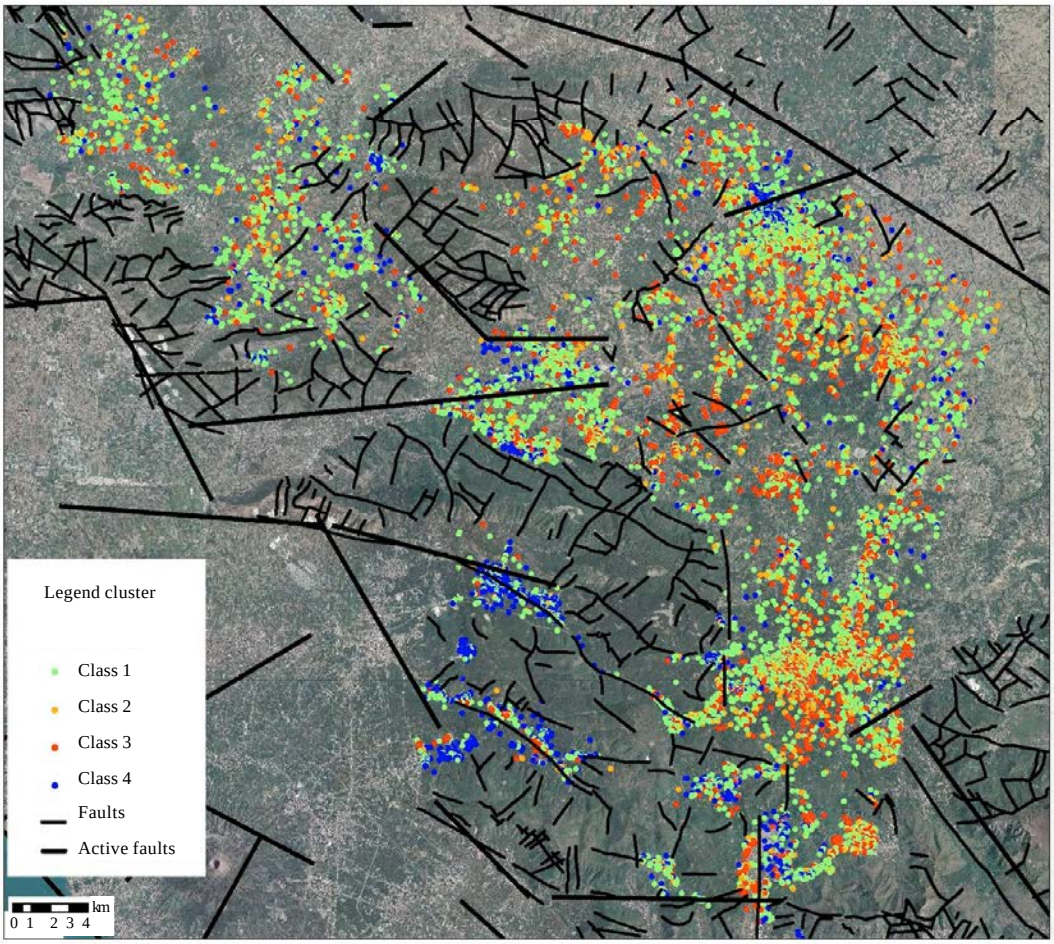

Fig. 5: Distribution of PS classes in the study area

along LOS. Mainly class 1 PS characterizes the South-Western sector (Monte Partenio) of the study area, showing rates of uplifting along LOS.

Two maps of the Avellino city (Fig. 6) and Benevento city (Fig. 7) are proposed in order to show the results of the application in a study of ground deformation in urban areas.

The Avellino complex ground deformation trends are proposed in Fig. 6. On the top map of Fig. 6, the spatial distribution of the PS velocity is given and it shows a scattered distribution of negative (yellow dots) and null (green dots) values in almost all the area. A biggest density of negative values is present only in the SE sector, while only some PS with positive values is present in the NW sector. On the bottom map of Fig. 6, the spatial distribution of PS ground deformation trends obtained by clustering analysis is showed; in this map the different ground deformation trends existing in the central sector (blue dots = uplift) and in the eastern and western sectors (red and orange dots = subsidence) of the city are more clearly recognized. The boundary between the three sectors is transitional and could be related to the presence

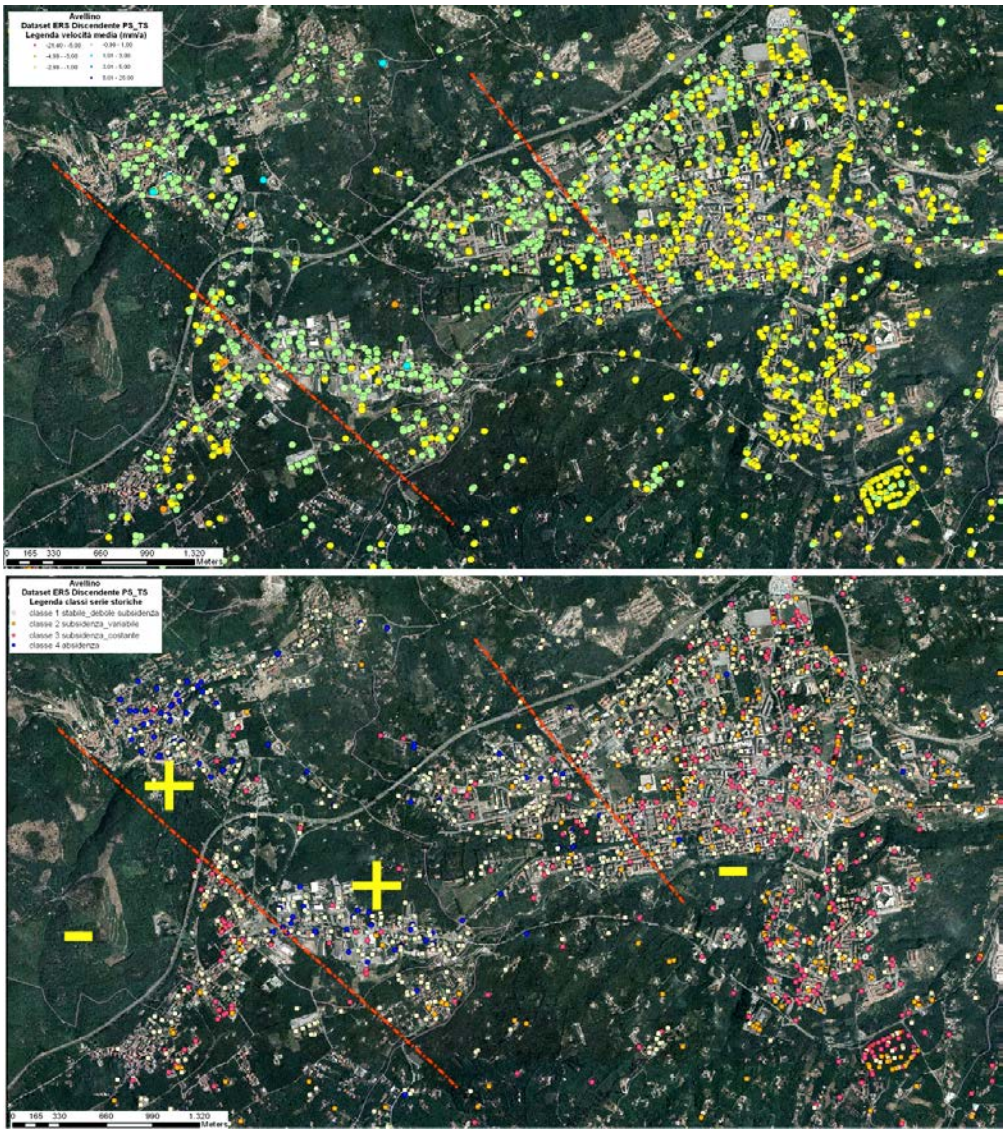

Fig. 6: Maps referred to Avellino town area: average LOS velocity of PS on the top and cluster classes distribution on the bottom; the red lines mark the sectors with different ground deformation trend (+: Uplifting; -: Subsidence)

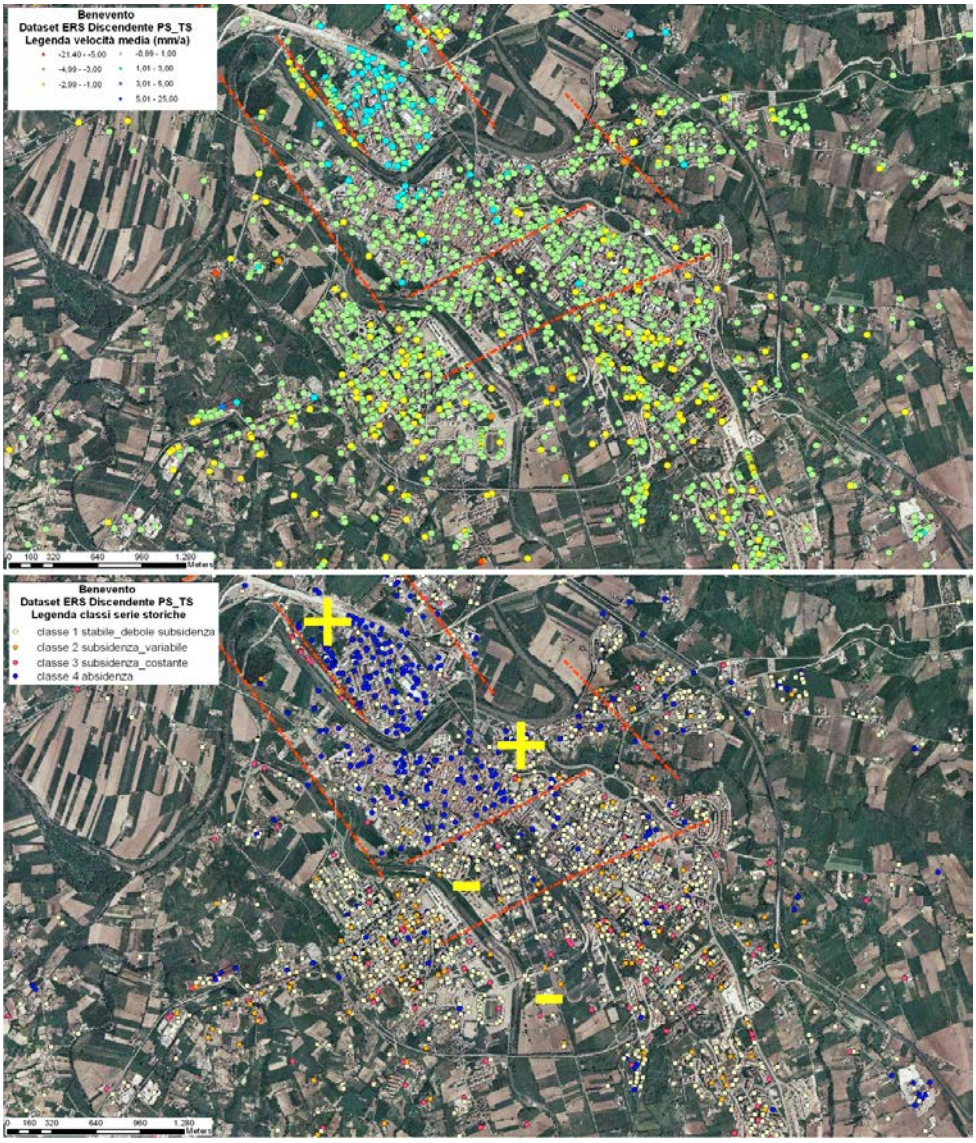

Fig. 7: Maps referred to Benevento city area: average LOS velocity of PS on the top and cluster classes distribution on the bottom; the red lines mark the sectors with different ground deformation trend (+: Uplifting; -: Subsidence)

of tectonic structures (i.e., faults) and to the anthropic activity (water pumping).

The Benevento city differential ground deformation trends are explained in Fig. 7. The top map shows the spatial distribution of the PS velocity is given with a gradual change in velocity values from positive (blue dots) to null (green dots) and negative (yellow dots) values passing from NW (top left) to SE (bottom right), with a central sector characterized by transitional features. The bottom map presents the spatial distribution of PS ground deformation trends obtained by clustering analysis; the different ground deformation trends existing in the Northern sector (blue dots = uplift) and in the Southern sector (red and orange dots = subsidence) of the city are clearly recognized. The boundary between the two sectors seems very sharp without transition feature and could be related to the presence of a buried active faults.

CONCLUSIONS

The cluster individuation and the spatial definition of the deformation trends is a basic data in the geological

interpretation of the active ground deformation of the studied territory and in the definition of the relevant natural risks.

The clustering analysis of the interferometric radar satellite data allows the evaluation of the rates and location of ground deformation processes with a very high accuracy. In particular, the examples described in Fig. 6 and 7 show the power of the clustering analysis in extracting important data in the analysis of geological processes by PS-InSAR processing database.

The time series clustering allowed us to identify:

- The regions in which there is a significant time dependent component to the deformation field
- The different geological causes of a deformation trend in a group of spatially coherent ground points

This study can be seen as a first, pilot study on the phenomenon. Currently, we are working on the analysis of the data set referred to the whole Campania region. Furthermore, we are studying the problem of searching the best algorithm for clustering this dataset. An innovative approach in this framework, based on Self Organizing Maps, has been proposed by Romano and Scepi (2006) but it has tested only on a sample of the 18.452 PS time series. We are working on a better specification of this algorithm.

ACKNOWLEDGMENTS

The PS database has been implemented under the TELLUS project, which has been developed in the framework of the PODIS project (Progetto Operativo Difesa Suolo) of the Ministero dell'Ambiente e per la Tutela del Territorio e del Mare (MATTM) of Italy and Regione Campania and has been funded by the European Union QCS 2000-2006 PON-ATAS. The authors thanks a lot Fabio Matano for his helpful comments in the geological interpretation of data and two anonymous referees for their helpful review.

REFERENCES

Berkhin, P., 2006. A Survey of Clustering Data Mining Techniques. In: Grouping Multidimensional Data Recent Advances in Clustering, Nicholas, K. and Teboulle (Eds.). Springer, New Mexico, pp: 25-71.

Chih-Ping, W., L.Yen-Hsien and H. Che-Ming, 2003. Empirical comparison of fast partitioning-based clustering algorithms for large data sets. Expert Syst. Applications, 24: 351-363.

Colasanti, C., A. Ferretti, C. Prati and F. Rocca, 2003. Monitoring landslides and tectonic motion with the Permanent Scatterers technique. Eng. Geol., 68: 3-14.

Estivill-Castro, V. and A.T. Murray, 1997. Spatial clustering for data mining with generic algorithms. Queensland University of Technology, Faculty of Information Management, Technical Report FIT-TR-97-10.

Farina, P., D. Colombo, A. Fumagalli, F. Marks and S. Moretti, 2006. Permanent Scatterers for landslide investigations: Outcomes from the ESA-SLAM project. Eng. Geol., 88: 200-217.

Ferretti, A., C. Prati and F. Rocca, 2000. Nonlinear subsidence rate estimation using Permanent Scatterers in differential SAR Interferometry. IEEE Trans. Geoscience Remote Sensing, 38: 2202-2212.

Ferretti, A., C. Prati and F. Rocca, 2001. Permanent scatters in SAR interferometry. IEEE Trans. Geoscience Remote Sensing, 39: 8-20.

Huang, Z., 1997. Clustering large data sets with mixed numeric and categorical values. Proceedings of the 1st Pacific-Asia Conference on Knowledge Discovery and Data Mining, Feb. 23-24, Singapore, pp: 21-34.

Kaufman, L. and P.J. Rousseeuw, 1990. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York, ISBN: 9780471735786.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Jan. 17-20, Berkeley, CA., pp: 281-297.

Milone, G., 2008. Temporal data mining: Tecniche e algoritmi di clustering. Ph.D. Thesis, Statistics, Università Federico II di Napoli.

Ng, R. and J. Han, 1994. Efficient and effective clustering method for spatial data mining. Proceedings of International Conference on Very Large Databases, Sept. 12-15, Santiago, Chile, pp: 144-155.

Romano, E. and G. Scepi, 2006. Integrating time alignment and self organizing maps for classifying curves. Proceedings of the Knowledge Extraction and Modeling, IASC-INTERFACE, IFCS, Workshop, Sept. 4-6, Anacapri, pp: 1-5.

Scarso, N., 2008. Stima del numero di cluster. Ph.D. Thesis, Statistics and Informatics, Università di Padova, Facolta di Scienze Statistiche.

Scepi, G., 2009. Clustering Algorithms for Large Temporal Data Set. In: Data Analysis and Classification, Francesco, P., C.N. Lauro and J.G. Michael (Eds.). Springer, Heidelberg, ISBN: 978-3-642-03738-2.

Vilardo, G., G. Ventura, C. Terranova, F. Matano and S. Nardò, 2009. Ground deformation due to tectonic, hydrothermal, gravity, hydrogeological, and anthropic processes in the Campania Region (Southern Italy) from Permanent Scatterers Synthetic Aperture Radar Interferometry. Remote Sensing Environ., 113: 197-212.