Functional Data Analysis for Interferometric Synthetic Aperture Radar Data Post-Processing: The case of Santa Barbara mud volcano

Analisi di dati funzionali per il post-processamento di dati interferometrici: Il caso dei vulcanelli di Santa Barbara

Matteo Fontana, Alessandra Menafoglio, Francesca Cigna, Deodato Tapete

Abstract The systematic and widespread availability of cheap computing power and storage space, alongside the drastic improvement in data transmission and active sensor technologies, has triggered an increasing interest by scientific and government institutions in the use of satellite data. One of the most promising applications is to provide emergency management agencies with early warning systems for catastrophic natural events, in order to support decision making. The data used in this work are ground displacement records gathered via interferometric processing of satellite RADAR data. We provide a novel post-processing approach based on a Functional Data Analysis paradigm, and apply it to Santa Barbara mud volcano.
Abstract *La sistematica disponibilità di memoria e potenza di calcolo, insieme al drastico miglioramento nelle tecnologie di tramissione dati e dei sensori attivi, ha stimolato un interesse crescente da parte di istituzioni scientifiche e governative verso l'utilizzo di dati satellitari. Una delle applicazioni più promettenti è fornire sistemi di allerta precoce alle agenzie di gestione delle emergenze, per supportare i processi decisionali. I dati usati in questo studio sono stime di spostamento del terreno ottenute dall'elaborazione interferometrica di immagini RADAR satellitari. Viene proposto un innovativo metodo di post-processamento basato sul paradigma dell'analisi di dati funzionali, e la sua applicazione ai vulcanelli di Santa Barbara.*

Key words: Functional Data Analysis, InSAR, Remote Sensing, Conformal Prediction

---

Matteo Fontana
MOX - Department of Mathematics, Politecnico di Milano, Milan, Italy
e-mail: matteo.fontana@polimi.it

Alessandra Menafoglio
MOX - Department of Mathematics, Politecnico di Milano, Milan, Italy

Francesca Cigna
Italian Space Agency (ASI), Rome, Italy

Deodato Tapete
Italian Space Agency (ASI), Rome, Italy

1 Motivation

In the last years, a significant advance in data transmission technologies and availability of computing power, storage space and sensors, have triggered a boost in terms of possible applications of satellite imaging data. Among the wealth of data that can be acquired by passive or active sensor arrays on-board satellites, Synthetic Aperture Radar (SAR) images and their processing with Interferometric SAR (InSAR) methods are one of the most useful sources of information for natural hazard monitoring [10].

The statistical analysis of the output products of advanced InSAR processing is still at an early stage. We are witnessing several attempts at using standard statistical frameworks such as time-series analysis and geostatistics, as shown in the review by [2]. To the best of our knowledge, no attempts have been made yet to combine the dynamics in time and in space, or to employ more advanced statistical techniques to gather novel insights from these specific data.

Our aim is to provide a proof of concept for the application of advanced statistical techniques in this realm, tackling the issue of geological hazard monitoring and early-warning. We provide deeper explanations about the novelty of our approach with respect to the current state of the art in [6]. We exploit, as a test case, the event of a mud volcano eruption occurred in the village of Santa Barbara, in the eastern sector of the city of Caltanissetta, in Sicily (Italy). On the 11^{th} of August 2008 the area was affected by paroxysmal eruption that caused damage to urban infrastructure as far as 2 km from the main eruptive vent. A more detailed description of the event and its geological features can be found in [5] and [8].

2 The Application: InSAR data of Caltanissetta

We used 32 ENVISAT Advanced SAR scenes acquired along ascending track T172 between 12/10/2002 and 07/06/2008 (i.e. before the mud volcano erupted). These data are in C-band (5.6 cm wavelength, 5.3 GHz frequency) and characterized by a Line-Of-Sight (LOS) with ~23° look angle, VV co-polarization, ~20 m ground resolution and nominal site revisit of 35 days.

InSAR processing was carried out with the Small Baseline Subset (SBAS) technique developed by [1] and parallelized by [3]. The output dataset consists of $n = 1735$ coherent targets, distributed across an area of 150 km^2. For each target, the annual LOS velocity over the monitoring period, LOS displacement time series, temporal coherence, and elevation above the reference ellipsoid were estimated. The position of the targets can be seen in Fig. 1, while a plot of the corresponding temporal dynamics is shown in Fig. 2a. The 2002-2005 ground deformation scenario in Caltanissetta was described by [13]. Previous semi-automated analysis of a different InSAR dataset was carried out in [5], where the computation of Deviation Indices was proposed to identify trend changes in InSAR time series (see also [12]). We present the main findings about the mud volcano area in [6].

3 A Functional Approach for InSAR data post-processing

Each displacement series estimated via InSAR consists of a set of discrete, time-indexed evaluations of a continuous trajectory in time, with some degree of smoothness given by the physics of the deformation phenomenon. We can also assume some measurement error to be present, due to the particular nature of the measuring and processing technique. Moreover, in this specific experimental setting, derivatives of such continuous trajectories carry a lot of information: the first derivative represents the velocity of the measured displacement, while the second derivative is the acceleration profile and, via a multiplicative constant, force values. These specific and nonstandard data features call for the use of statistical methods able to correctly include and model them into the analysis process.

A very good candidate to do so is Functional Data Analysis (FDA) [9], the field of statistics which uses, as the unit of the analysis, one or more continuous functions over a domain, either univariate or multivariate. The first step in every FDA pipeline is the extraction of continuous functions from the discrete, longitudinal data points. In our case this was performed via a smooth B-spline basis. Further details about the procedure, and the choice of the smoothing parameter can be found in [6], while a plot of the smoothed data is shown in Fig. 2b. After the data smoothing, we performed a functional Principal Component Analysis (PCA) to assess and quantify the variability of the displacement curves. A functional clustering based on the K-Mean Alignment (KMA) procedure [11] was also performed, to identify a group structure

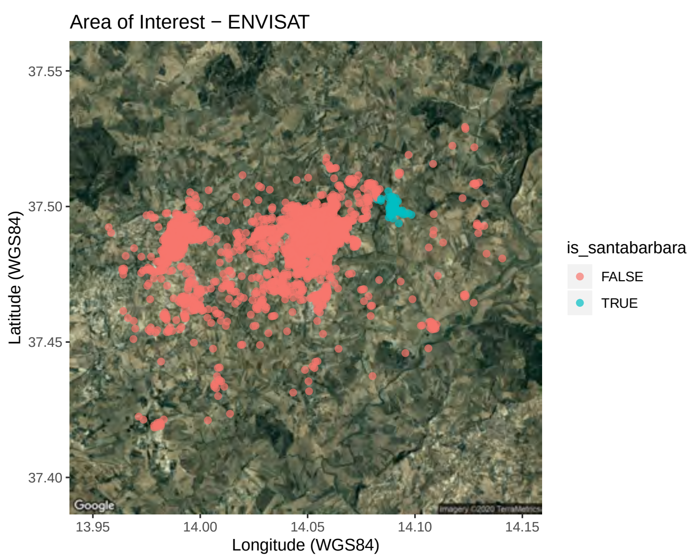
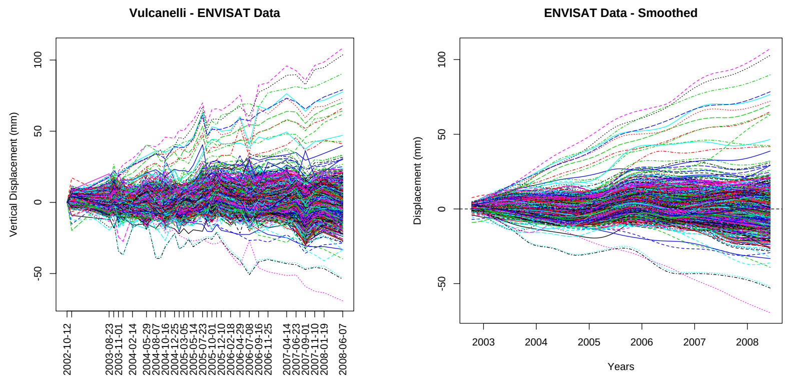


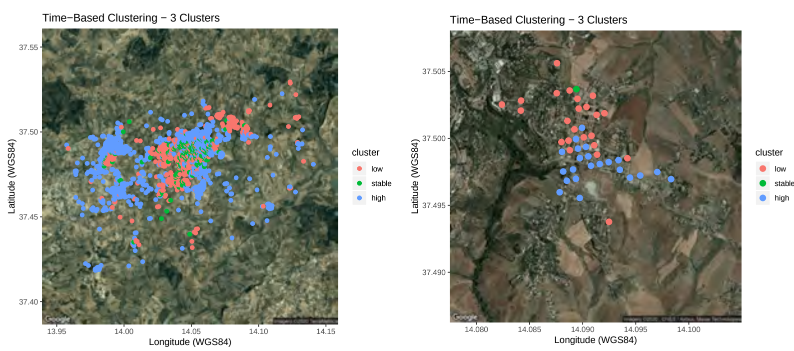
Fig. 1: Map of the geographical position of the target points on a satellite image of the area of Caltanissetta (Italy), with indication of Santa Barbara village.



(a) Unsmoothed Case (b) Smoothed Case

Fig. 2: Plot of the time dynamics of displacement for the target points in the smoothed and unsmoothed case.

in the data. In Fig. 3 we present a spatialization of the clusters obtained via the KMA procedure, with a zoom on the Santa Barbara area. Further details about the results of the exploratory analysis can be found in [6].



(a) Map of the whole area (b) Zoom over Santa Barbara

Fig. 3: A map of the target points over the city of Caltanissetta, where the color depends on the cluster that was assigned to the corresponding displacement curve by the KMA procedure, with a zoom on the Santa Barbara area

After the exploratory analysis, we focused our attention on the issue of set forecasting for these complex data objects, that is defining a prediction region for an infinite-variate data object such as a function. Providing meaningful forecasts with associated uncertainty for a given data object is, in general, a key problem from a theoretical and applied perspective. If we frame the early warning problem as identifying a time instant at which the time series departs significantly from its expected value, it appears immediately evident how an appropriate set forecast method represents a key stepping stone to realize any kind of outlier identification system.

The simplest and possibly the most natural way to forecast a time series of data objects is to use an autoregressive model. Let δ_t be the displacement surface measured at time t, which is here assumed to be a square-integrable function defined over bidimensional Euclidean space S, $\delta_t \in L^2(S)$. Modeling the surface at $t+1$ with a concurrent functional autoregressive (FAR) model of order 1 $FAR(1)$, we can write

$$\delta_{t+1}(s) = \Phi(s)\delta_t(s) + \varepsilon(s) \tag{1}$$

where $s \in S$ and $\Phi \in L^\infty(S)$. The main issue in this setting is the substantial lack of usable and interpretable methods for interval forecasting. A very powerful approach to tackle this kind of problems is represented by Conformal Prediction (CP) [14]. CP was developed in the late 1990s in the Machine Learning community as a method to provide interval forecast for support vector machines. Since then, it has been extended in several ways, and used as a framework to provide distribution-free prediction sets in the scalar case, and very recently also in the functional one [7].

To provide meaningful forecasts for the concurrent FAR(1) model in Equation 1, we extend the framework in [7] in two directions. The first one is moving from the functional univariate case to the bivariate one, required in this application. The second, and most challenging one, is to move from the case of iid observations (described in [7]) to setting of data with (temporal) dependence. This is performed by adapting the ideas described in [4] for the scalar case to the case of functions whose domain is bidimensional. The complete mathematical description of the method, alongside a detailed analysis of the Santa Barbara test case can be found in [6].

4 Conclusions

Motivated by the need to test advanced statistical methodologies to analyse displacement series obtained from satellite InSAR techniques, we illustrate a Functional Data Analysis framework for the analysis of displacement data related with a mud volcano eruption occurred in 2008 close to the city of Caltanissetta, Italy. Aiming to develop an early warning system, we describe an extension (fully shown in [6]) of a newly developed forecasting technique based on Conformal Prediction to the case of dependent functional observations defined over a multivariate domain.

5 Acknowledgements

This work was carried out in the framework of the ASI-POLIMI "Attività di Ricerca e Innovazione" project, grant agreement n.2018-5-HH.0.

References

1. Berardino P., Fornaro G., Lanari R., Sansosti E. 2002. A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. IEEE Trans. Geosci. Remote, 40, 2375-2383, doi:10.1109/TGRS.2002.803792
2. Bernardi M.S., Africa P.C., de Falco C., Formaggia L., Menafoglio A., Vantini S. 2020+. On the use of Interfeometric Synthetic Aperture Radar Data for Monitoring and Forecasting Natural Hazards. Manuscript, Politecnico di Milano.
3. Casu F., Elefante S., Imperatore P., Zinno I., Manunta M., De Luca C., Lanari R. 2014. SBAS-DInSAR Parallel Processing for Deformation Time-Series Computation. IEEE J. Sel. Top. Appl., 7, 3285-3296, doi:10.1109/JSTARS.2014.2322671
4. Chernozhukov V., Wüthrich K., Yinchu Z. 2018. Exact and Robust Conformal Inference Methods for Predictive Machine Learning with Dependent Data. Proc. 31st Conference On Learning Theory, in PMLR 75:732-749.
5. Cigna F., Tapete D., Casagli N. 2012. Semi-automated extraction of Deviation Indexes (DI) from satellite Persistent Scatterers time series: tests on sedimentary volcanism and tectonically-induced motions. Nonlin. Processes Geophys., 19, 643-655, doi:10.5194/npg-19-643-2012
6. Fontana M., Bernardi M., Cigna F., Menafoglio A., Tapete D., Vantini S. 2020+. Interferometric Synthetic Aperture Radar Data Post-Processing: A Functional Data Analysis Perspective. Manuscript, Politecnico di Milano.
7. Fontana M., Vantini S., Tavoni M., Gammerman A. 2020. A Conformal Approach for Distribution-Free Prediction of Functional Data. In: Aneiros G. et al. (eds) Functional and High Dimensional Statistics and Related Fields. Contribution to Statistics. Springer. Forthcoming.
8. Madonia P., Grassa F., Cangemi M., Musumeci C. 2011. Geomorphological and geochemical characterization of the 11 August 2008 mud volcano eruption at S. Barbara village (Sicily, Italy) and its possible relationship with seismic activity. Nat. Hazards Earth Syst. Sci., 11, 1545—1557, doi:10.5194/nhess-11-1545-2011
9. Ramsay J.O., Silverman B.W. 2005. Functional Data Analysis. Second edition. Springer Series in Statistics. New York, NY: Springer.
10. Rosen P.A., Hensley S., Joughin I.R., Li F.K., Rodríguez E., Goldstein R.M. 2000. Synthetic Aperture Radar Interferometry. Proc. IEEE, 88, 3, doi:10.1109/5.838084
11. Sangalli L.M., Secchi P., Vantini S., Vitelli V. 2010. 'K-Mean Alignment for Curve Clustering'. Computational Statistics & Data Analysis 54 (5): 1219–33. https://doi.org/10.1016/j.csda.2009.12.008
12. Tapete D., Casagli N. 2013. Testing computational methods to identify deformation trends in RADARSAT Persistent Scatterers time series for structural assessment of archaeological heritage. In: Murgante B. et al. (eds) Computational Science and Its Applications, ICCSA 2013. Lecture Notes in Computer Science, 7972, 693-707, doi:10.1007/978-3-642-39643-4_50
13. Vallone P., Giammarinaro M.S., Crosetto M., Agudo M., Biescas E. 2008. Ground motion phenomena in Caltanissetta (Italy) investigated by InSAR and geological data integration. Engineering Geology, 98, 3-4, 144-155, doi:10.1016/j.enggeo.2008.02.004
14. Vovk, V., Gammerman A., Shafer G. 2005. Algorithmic Learning in a Random World. New York, NY: Springer.