DEFORMATION TRANSIENT ANALYSIS IN LONG TERM INSAR AND GPS TIME SERIES BY MEANS OF THE PRINCIPAL COMPONENT ANALYSIS (PCA): RESULTS FROM DPC-INGV S3 PROJECT

G. Pezzo¹, C. Tolomei¹, E. Serpelloni², A. Gualandi², S. Salvi¹, J.P. Merryman Boncori³, A. Lugari¹

¹ Istituto Nazionale di Geofisica e Vulcanologia, CNT, Roma, Italy
² Istituto Nazionale di Geofisica e Vulcanologia, CNT, Bologna, Italy
³ Sarmap SA, Purasca, Switzerland

Introduction. The aim of the DPC-INGV S3 project “Short term earthquake prediction and preparation” (2014-15) was exploring the possibility that multiparametric monitoring of a number of observables potentially related to the crustal strain field could represent a tool for better constraining seismic hazard in Italy. The basic approach followed in the project considers seismogenic process as “critical” process in the frame of the “self-organized criticality” model (e.g., Kagan, 1994; Main, 1995, 1996). In this model, the seismogenic process is “critical” in nature (e.g. Kagan, 1992; Turcotte, 1992) having as a basic dynamical element the fault interaction (e.g., Sacks and Rydelek, 1995; Castellaro and Mulargia, 2001). This interaction is responsible for the observed phenomenology, that is characterized by irregularly distributed bursts of energy releases (time-space clustering of earthquakes) and scale relationships (e.g. the ‘Gutenberg-Richter law’).

In this context, the basic goal of the project was the full exploitation of a huge amount of data to detect possible large scale/short term (months to weeks) transient strain field variations that could be the marker or the responsible of upcoming earthquakes. One of the key points is the space-time analysis of the surface deformations. Thus, our aim (Research Unit UR-01, Deformation transient analysis in long term InSAR and GPS time series) was the analysis of the mean ground velocity and the relative displacement time series, obtained from both InSAR and Continuous GPS (CGPS) data.

The use of radar satellite data and the differential interferometric technique allows to explore the processes and phenomena relative to the accumulation and subsequent release of crustal deformation during the seismic cycle at a high spatial resolution, while continuous GPS data guarantee high accuracies and precisions and higher temporal sampling with respect to InSAR observations. The integration of GPS and InSAR measurements, during both the SAR processing and post-processing steps, allows an overall improvement of InSAR accuracies, while allowing the exploration of systematic errors in InSAR data. The high temporal sampling of GPS and the wide area coverage of InSAR observations provide a solid observational space for monitoring both long-term crustal deformation and the occurrence of possible (shorter terms) deformation transients, for which we would expect some spatially coherent signals in the geodetic time-series.

Finally, the GPS and InSAR displacement time-series will be the input of a multivariate statistical analysis, which will start from the use of a Principal Component Analysis (PCA) method. We focus our analysis on a study area of major concern located in the Po plain, in particular in the Emilia region, affected by a seismic sequence during 2012.

Seismotectonic context. The Emilia region is located in the Po Plain, representing the E-W continental collisional boundary between the subducting Adria plate to the north, and the overriding northern Apennines block to the south (e.g. Picotti and Pazzaglia, 2008). Seismic sections (e.g. Boccaletti and Martelli, 2004) reveal that central Po Plain is characterized by an arcuate blind thrust system, where the related growth folds are located below the Plio-Quaternary sedimentary cover (Pieri and Groppi, 1981; Pieri, 1983; Consiglio Nazionale delle Ricerche, 1992; Burrato et al., 2003). The current tectonic activity of the frontal part of the northern Apennines accretionary wedge was proposed by many authors basing on geomorphological analysis, subsurface geology, seismicity, and present day stress field (Meletti et al., 2000; Valensise and Pantosti, 2001; Burrato et al. 2003; Scrocca et al., 2007; Boccaletti et al., 2010). Burrato et al. (2003) described active fold growth in this area observing the northward
migration of the Po River and other smaller watercourses during the Late Quaternary. Recent horizontal GPS measurements (Serpelloni et al., 2006; Devoti et al., 2011) suggest ~1 mm yr\(^{-1}\) of shortening, confirming the occurrence of active thrusting in the area.

The epicentral area of the 2012 Emilia seismic sequence is characterized by low historical seismicity. Since at least the year 1000 A.D. this area has never experienced earthquakes of magnitude larger than ML ~ 4.7 [1574 Finale Emilia earthquake: Rovida et al. (2011)], whereas larger magnitude earthquakes occurred west and east of the 2012 epicentral area. The 2012 sequence shows two mainshocks occurred on May 20 and 29, 2012, of magnitude ML 5.9 and 5.8, respectively. Pezzo et al. (2013) investigated the related coseismic surface deformation using COSMO-SkyMed and Radarsat-1 satellite SAR images and GPS data: SAR and GPS data inversions suggested that the main activated faults were the Ferrara thrust (for the May 20 shock) and the Mirandola one (for the May 29 shock). They also observed that the maximum deformation measured by InSAR data corresponds to the top of the Mirandola and Ferrara folds: in the long-term, the growth of these folds cause the wide northward bend of the Po river course and the deviation of the Secchia and Panaro rivers, as suggested by Burrato et al. (2003).

An interesting finding of Pezzo et al. (2013) is an aseismic deformation, well modelled by slip on the eastern side of the frontal Mirandola thrust later activated during the May 29 mainshock, occurred in the time span between the two mainshocks.

**InSAR data.** We processed InSAR data sets for the X-Band (COSMO-SkyMed) and C-Band (ERS and Envisat) sensors to retrieve the largest possible spatial and temporal coverage. We used both ascending and descending COSMO-SkyMed image data sets, spanning the June 2012- December 2013 time interval, to study the post-seismic deformation, as well as ASAR and ERS image data sets, spanning from 1992 to 2010, for the inter-seismic period (Tab. 1). Using the multi-temporal InSAR technique named SBAS (Small BAseline Subset, Berardino et al., 2002), we obtained mean ground velocity maps and displacement time series for each coherent pixel in the temporal span of the processed images (Figs. 1 and 2). We use time series and mean ground velocity retrieved from CGPS to calibrate InSAR velocity maps and remove spurious signals during the SAR data processing. The CGPS displacement rates are used to validate the InSAR results.

Tab. 1 - Available SAR data set for COSMO-SkyMed image beams (Stripmap mode), ERS and ASAR sensors for both ascending and descending geometries, covering almost the entire Emilia Region.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Orbit</th>
<th>Track-Frame</th>
<th>Temporal Span</th>
<th>N° of images</th>
<th>N° of pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>COSMO-SkyMed</td>
<td>Ascending</td>
<td>05 west</td>
<td>8/7/12-28/12/13</td>
<td>26</td>
<td>86</td>
</tr>
<tr>
<td>COSMO-SkyMed</td>
<td>Ascending</td>
<td>05 east</td>
<td>9/7/12-17/11/13</td>
<td>22</td>
<td>70</td>
</tr>
<tr>
<td>COSMO-SkyMed</td>
<td>Ascending</td>
<td>04</td>
<td>13/7/12-23/12/13</td>
<td>25</td>
<td>88</td>
</tr>
<tr>
<td>COSMO-SkyMed</td>
<td>Ascending</td>
<td>03</td>
<td>30/5/12-14/8/13</td>
<td>17</td>
<td>48</td>
</tr>
<tr>
<td>COSMO-SkyMed</td>
<td>Ascending</td>
<td>10</td>
<td>4/7/12-14/12/13</td>
<td>34</td>
<td>128</td>
</tr>
<tr>
<td>COSMO-SkyMed</td>
<td>Ascending</td>
<td>11</td>
<td>28/7/12-26/12/13</td>
<td>39</td>
<td>251</td>
</tr>
<tr>
<td>COSMO-SkyMed</td>
<td>Descending</td>
<td>03</td>
<td>20/6/12-20/12/13</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td>COSMO-SkyMed</td>
<td>Descending</td>
<td>05</td>
<td>6/7/12-16/12/13</td>
<td>25</td>
<td>76</td>
</tr>
<tr>
<td>ERS</td>
<td>Descending</td>
<td>165-2705</td>
<td>30/9/92-5/6/99</td>
<td>36</td>
<td>134</td>
</tr>
<tr>
<td>ASAR</td>
<td>Descending</td>
<td>166-2705</td>
<td>25/1/03-21/8/10</td>
<td>33</td>
<td>139</td>
</tr>
</tbody>
</table>

ASAR and ERS mean velocity maps (1992-2010) show deformation patterns (Fig. 1) ascribable to different natural and human phenomena. We observe some strong subsidence patterns (red colours) mainly due to the intensive water pumping carried out in the Po plain. Subsiding patterns are mainly located in correspondence with the maximum thickness of soft
Along the entire Apennines front, it is possible to observe a relative (with respect to the Po Plain) uplift pattern. Moreover, comparing the space distribution of positive values (Fig. 1) with the main tectonic features represented by some buried growing anticlines (not shown in the text for reasons of space), we observe a rough correspondence: positive (uplift) mean ground velocity are located in correspondence of the culminations of the anticlines, in particular along the Mirandola anticline, located below the MO04 GPS station. This could be the result of two phenomena: the sediment compaction, that is higher where the sediment is thicker, and the tectonic growth of the buried active anticlines.

Fig. 1 – ENVISAT-ASAR descending mean velocity map along with two examples of displacement time series (1992-2010), at the top and bottom of the figure. The measurements are in line of sight (LoS): positive values (blue colours) are approaching to the sensor; vice versa for negative ones (red colours). Red stars indicate the main (Mw ≥ 5.0) seismic events; magnitudes are expressed as Mw. Black triangles represents the CGPS.
Post-seismic deformation maps from Cosmo-SkyMed (Fig. 2) show positive values (approaching to the sensor) in the 2012 epicentral area, reaching about 20 mm yr\(^{-1}\) in both ascending and descending geometries. As for the ERS and ASAR maps, there is a rough correspondence between the growing anticline and the positive velocity values, and vice versa; in particular this is visible in correspondence of the 2012 epicentres, represented by the red stars in Fig. 2. Thus, there are some similar patterns between pre- and post-seismic maps. Observed local subsidence is interpreted as due to fluid migration and compaction of the Po plain sediments. Post-seismic time series show quasi-exponential afterslip deformation patterns in the epicentral area for both the ascending and the descending maps (Fig. 2).
CGPS data. Concerning the continuous GPS data analysis, the spatially correlated Common Mode Error (CME) is one of the major noise sources in GPS position time series (Wdowinski et al., 1997; Dong et al., 2006) and it has been recently recognized and accounted by Serpelloni et al. (2013) to improve the accuracies of GPS measurements of vertical ground motion rates. In order to obtain an improvement in the daily repeatability of GPS positions, and an improvement in the signal-to-noise ratio, we adopted a PCA (Principal Component Analysis) method to estimate and remove the spatially correlated Common Mode Noise error in GPS time-series, following the procedure described by Dong et al. (2006) and by Serpelloni et al. (2013). This step required to enlarge the number of GPS stations included in the analysis, because the CME in the Euro-Mediterranean area has been described as non-spatially uniform (Serpelloni et al., 2013). Thus, we used state-of-the-art GPS data processing methods to derive a uniform and homogeneous GPS solution over the Euro-Mediterranean area, analyzing data from >2500 continuous GPS stations for the 1994/01/01 – 2015/05/26 time span. Additionally, in order to avoid the cancelation of interesting tectonic signals from the study area, the CME was estimated using a wider network, including tectonically stable regions in central Europe. We performed this analysis on residual time-series of high-quality GPS stations, after removing the linear trend, jumps and seasonal signals, excluding from the analysis the investigated area and all sites with clear non-linear or noisy data.

The raw GPS data, in the form of daily, 30-seconds sampling, RINEX (Receiver INdependent EXchange) files, were analyzed with the GAMIT/GLOBK software. Stations velocities were obtained following a three-step procedure, consisting in: 1) the phase data reduction, 2) the combinations of solutions and realization of the reference frame, and 3) the time-series analysis.

In the first step, we used daily GPS phase observations to estimate site position, adjustments to satellite orbital parameters (EOPs) and time-variable piecewise linear zenith and horizontal gradient tropospheric delay parameters by means of the GAMIT (V10.5) software, applying loose constraints to the geodetic parameters, using IGS final orbits. GPS phase data are weighted according to an elevation-angle dependent error model, using an iterative analysis procedure whereby the elevation dependence is determined from the observed scatter of phase residuals. We applied the Ocean-loading and a pole-tide correction model FES2004. We used the Vienna Mapping Function (VMF1) for modeling the tropospheric delay and the Global Pressure and Temperature 2 (GPT2) model. We also accounted for non-tidal atmospheric loading in the phase data processing step, and for 2nd/3rd order ionosphere effects. We used the IGS absolute antenna phase center model for both satellite and ground-based antennas. In the second step we used the ST_FILTER program of the QOCA software (http://qoca.jpl.nasa.gov) to combine the regional daily loosely constrained solutions obtained in Step 1 and simultaneously realize a global reference frame by applying generalized constraints. Specifically, we defined the reference frame by minimizing the velocities of the IGB08 global core stations (https://igscb.jpl.nasa.gov/network/refframe_core.html), while estimating a seven-parameter transformation with respect to the IGS realization of the ITRF2008 NNR frame, named IGB08 (https://igscb.jpl.nasa.gov). In the third step, we analysed the position time-series, defined in the IGB08 reference frame. Changes in stations positions were modelled using standard functional model.

Position outliers were cleaned from the time-series adopting a post-fit RMS criterion (values larger than 3 times the post-fit Weighted Root Mean Square, WRMS, were discharged). A few stations recorded the coseismic offsets and postseismic transients due to the May 2012 Emilia earthquake sequence. For those sites, we modeled two coseismic offsets and an exponential decay function, with a 30-day decay constant. Non-tectonic jumps, mainly due to changes in the stations equipment, were defined from the analysis of station log-files, when available, and from visual inspections of the time series. Outliers were cleaned adopting a post-fit RMS criterion, in particular, we discarded values larger than 3 times the post-fit Weighted Root Mean Square (WRMS).
Here the time-series analysis is not performed with the goal of estimating station velocities, but with the goal of generating several types of “residual” time-series, where with the term “residuals” we intend the time-series obtained after removing a certain model from the raw data-set. In particular, we first generated the full residual time-series. These residual time-series ideally would contain only noise, and among the most important sources of time-correlated noise, the so-called Common Mode Error is the most important one. Thus, as previously mentioned, we estimated the CME using a Principal Component Analysis (PCA) technique on the residual time-series, and using the PCA outputs to apply a space-time filtering on the time-series, improving daily repeatability, accuracies and precisions of the velocity estimates. We generated two different sets of filtered (CME removed) time-series, for two different sets of stations. Two areas were identified, for which we generated two different sets of filtered (CME-removed) time-series: Area1) encompassing the Alpine and northern Apennines, Po Plain area, for which we generated de-trended time-series, where only the non-tectonic signals (i.e., instrumental offsets) have been removed, so the time-series contains the seasonal terms, and Area2) corresponding to the area where InSAR displacement time-series are available, for which we generated de-trended residual time-series, where also the seasonal terms are removed.

Fig. 3 – The first three (and most representative) principal components of the PCA method (from left to right). U represents the spatial response with respect to the temporal trend (V), as V is the response of the function in time. Concerning the first and second components, one-year-lasting signals are extracted and a seasonal component prevails, while for the third component a cyclical multi-annual basis signal is present.
It is worth noting that, in order to reduce the percentage of missing data when forming the PCA analysis, we have restricted the time-span processed with PCA to the 2006-2012 time interval, before the beginning of the May 2012 Emilia earthquake sequence and when most of the CGPS sites were operative. Moreover, we have excluded all GPS stations with a number of epochs <50% of the total expected.

**Results.** An interesting analysis method of space-time distributed data is the Principal Component Analysis (PCA) method. In brief, PCA compresses data by finding an intrinsic orthogonal coordinate system for the observation data that preserves the greatest possible amount of variance (‘spread’ of the distribution). PCA projects the data from a high-dimensional space (data or sensor space) onto a low-dimensional space, and allows detecting signals in the geodetic time series that show a spatial coherence higher than a desired threshold. In the considered case study, the first PC clearly represents a long-term linear motion, mainly associated with the subsidence of areas located in proximity of cities like Bologna (PC1 in Fig. 3). Such a signal is dominating the time series. For some frames, the second PC shows a random temporal evolution, with a spatial pattern mainly focused around the edges of the frame. This result may represent the influence of the ground control points for the whole data set. Finally, the third PC (for some frame sorted as second, and switched with the previous one) shows a temporal evolution described by two linear functions, intersecting around the epoch 2003. Again, the spatial region mainly affected by this PC is close to subsiding areas. This peculiar temporal evolution is likely reflecting the corrections performed in order to unify the ERS and ASAR data sets in a single time series database. It is worth noting that the second PC pattern could depend on the choice of the Ground Control Points in the InSAR filtering step. In fact, these points are generally selected over the frame border (not including areas affected by known subsidence or uplift phenomena) to better estimate orbital ramps and tropospheric artefacts.

**Conclusions.** We processed a large number of SAR images both for the inter (pre)-seismic phase that for the post-seismic phase. Similarly the available CGPS time series has been elaborated, subtracting the CME and seasonal trends. The link between SAR retrieved displacement time series (1992-2010) with respect to the GPSs, provides results concerning the main deformation patterns. These are characterized by strong subsidence, mainly related to the underground water pumping. In addition, some feeble tectonic signals can be observed, principally localized in correspondence of the Mirandola anticline. We then performed the Principal Component Analysis (PCA) to both GPS and InSAR displacement time-series, attempting to discriminate the different contributions to the observed ground movements. The PC Analysis of the GPS time series shows how the main contribution to the displacement can be explained by a seasonal signal with an annual frequency, and by a multi-annual seasonal signal. The PC Analysis of the InSAR time series shows that the principal contributions to the displacement can be explained by a linear trend (major part of the variance) and a seasonal signal (more evident for the ASAR case).

In this test area, although there are some difficulties to calibrate InSAR and GPS finely, both techniques provide interesting data concerning ground deformation patterns in time and space. However, the tectonic signals are very low with respect to the subsiding and noise ones and their interpretation could be very arduous. For the same reasons, we think that the PCA analysis is not able to highlight trends or patterns related to tectonics phenomena in this area as the involved values are very small and probably masked by larger signals.

**Acknowledgments.** Results carried out using COSMO-SkyMed® PRODUCTS, © ASI (Italian Space Agency) - provided under license of ASI in the framework of the S3 Project “Short term earthquake prediction and preparation” (DPC-INGV, 2013-14). The Envisat images are provided by ESA (European Space Agency) under the CAT.1P 5605. This study has benefited from funding provided by the Italian Presidenza del Consiglio dei Ministri – Dipartimento della Protezione Civile (DPC). This paper does not necessarily represent DPC official opinion and policies. The InSAR processing are carried out using the Sarscape© software (Sarmap, CH). We thank Dr. Riccardi P. and Dr. Cantone A. (Sarmap, CH) for the fruitful collaboration and useful suggestions during the SAR data processing.
References


Consiglio Nazionale delle Ricerche (1992). Structural model of Italy and gravity map, Progetto Finalizzato Geodinamica, Quaderni de La Ricerca Scientifica 114, 3.


Valensise, G. and D. Pantosti (Editors; 2001). Database of potential sources for earthquakes larger than 5.5 in Italy, Ann. Geofisc. 44, no. 4, 180, with CD-ROM, INGV.