Enhancing land subsidence awareness via InSAR data and Deep Transformers

Diana Orlandi
Dept Information Engineering
University of Pisa
Pisa, Italy
diana.orlandi@phd.unipi.it

Federico A. Galatolo
Dept of Information Engineering
University of Pisa
Pisa, Italy
federico.galatolo@ing.unipi.it

Mario G. C. A. Cimino
Dept of Information Engineering
University of Pisa
Pisa, Italy
mario.cimino@unipi.it

Alessandro La Rosa
Dept of Earth Science
University of Pisa
Pisa, Italy
alessandro.larosa@dst.unipi.it

Carolina Pagli
Dept of Earth Science
University of Pisa
Pisa, Italy
carolina.pagli@unipi.it

Nicola Perilli
Dept. Civil and Industrial Eng.
University of Pisa
Pisa, Italy
nicola.perilli@unipi.it

***Abstract*—The increasing availability of Satellite technology for Earth observation enables the monitoring of land subsidence, achieving large-scale and long-term situation awareness for supporting various human activities. Nevertheless, even with the most-recent Interferometric Synthetic Aperture Radar (InSAR) technology, one of the main limitations is signal loss of coherence. This paper introduces a novel method and tool for increasing the spatial density of the surface motion samples. The method is based on Transformers, a machine learning architecture with dominant performance, low calibration cost and agnostic method. This paper covers development and experimentation on four-years surface subsidence (2017-2021) occurring in two Italian regions, Emilia-Romagna and Tuscany, due to ground-water over-pumping using Sentinel-1 data processed with P-SBAS (Parallel Small Baseline Subset) time-series analysis. Experimental results clearly show the potential of the approach. The developed system has been publicly released to guarantee its reproducibility and the scientific collaboration.**

Keywords—InSAR, Transformers, ground-water extraction, subsidence, situation awareness.

I. INTRODUCTION AND BACKGROUND

SAR (Synthetic Aperture Radar) remote sensing satellites for Earth observation are widely recognized as a key method for monitoring our planet. SAR satellites are capable of providing large-scale and long-term situational observations to a variety of surface processes. In the literature, *situational awareness* refers to the process of aggregating spatial-temporal variables and measurements to raise the abstraction (i.e., semantic) level of operational models, making them more adaptive to the global or long-term circumstances [1][2]. Let us consider land subsidence, that can occur both in a continuous time-progressive manner or as a sudden sinking of the ground surface and, as it is well known in literature, it can be caused by human activities (e.g. ground-water over-pumping, exploitation of underground reservoirs for oil and gas withdrawal, collapse of tailing dams) with a significant and irreversible impact on ecosystems [3]. Concerning the ground-water over-pumping, irrigation can be considered as one of the main purposes. However, an inadequate exploitation of underground water resources can lead to irreversible phenomena, such as the deceleration of the recharging time of an aquifer [4]. Thus, the monitoring of land subsidence can be considered as strategic for many stakeholders ensuring actions for environmental, socio-economic and technological interests, as well as for policy makers promoting sustainable practices in land and water management [4][5].

Conventional techniques of monitoring of land subsidence occurring in areas interested by ground-water extraction use a spatially limited sampling via piezometric levels [5], which is often restricted to a few small areas. Moreover, in many countries, such as Italy, most of the data collected is owned by singular municipalities or private users. Thus, despite anthropogenic water consumption being one of the main reasons of anthropogenic land subsidence, detailed information at a broad scale is often lacking [4]. In the last twenty years satellite Interferometric Synthetic Aperture Radar (InSAR) has emerged as a promising technology for studying and monitoring surface motions in different fields of the Geohazards, including subsidence in ground-water extraction. InSAR methods are capable of obtaining measures of surface displacements with sub-cm accuracy at an unprecedented level of spatial detail (tens of m pixel size over hundreds of km wide areas) and temporal resolution. Specifically, interferograms are generated by differencing SAR images taken at different times from the same orbital position. With Sentinel-1A/B satellites constellation, the revisit time can be as short as 6 days [5][6].

Nevertheless, even with the most-recent missions, one of the main limitations of InSAR is signal loss of coherence, largely due to changes in the surface conditions between the acquisitions. As a result, InSAR-derived surface displacement time-series are characterized by a low spatial density of point-like targets, especially in non-urban areas, where vegetation and cultivated fields can make the estimation of the surface deformation a task not easy to solve [5].

In the literature, many advanced processing algorithms have been proposed to improve InSAR data, e. g. to reduce the atmospheric effects, such as Multi-Temporal (MT-)InSAR, Permanent Scatterer Interferometry (PSI), Small Baseline Method (SBAS) [7]. In addition, many open-source time-series analysis software packages have been developed, e.g. for Sentinel-1 data SNAP, GMTSAR, STaMPS, GIANT, MintPy [7]. As a result, many data pipeline have been developed in this field, with different levels of efficiency in terms of parallel computing, and with different accuracy. With the increase of complexity of such approaches, an important design problem is to consider their management costs, in terms of parametric sensitivity (i.e., calibration cost), knowledge needed, reusability on multiple application domains, and so on. Consequently, researchers started addressing the problem via machine learning based methods, which are well-known in the literature for the dominant performance, the low calibration cost, and the agnostic method [8][9].

This paper presents a novel approach to overcome incoherence and enhance the spatial density of InSAR-derived surface displacement time-series. The approach is based on Transformers [10]. With respect to conventional deep learning models for sequential data, a transformer processes sequential input according to a self-attention mechanism, i.e., a weighting of the significance of an input of the sequence. As a result, parallelization during training is facilitated and made efficient for large dataset. In addition, traditional deep learning architectures, such as Long-Short Term Memory (LSTM) and Convolutional Neural Networks (CNN), exhibits the so-called vanishing / exploding gradient problem, which makes the design and calibration very costly [11][12]. In contrast, the transformer architecture employs exclusively attention building blocks, with very good results with far less engineering tuning time.

The proposed system has been developed and publicly released to guarantee its easy adoption and reproducibility [13]. Experimental studies, with a focus on Central Italy, in the regions of Emilia-Romagna and Tuscany, have been carried out. The proposed approach clearly shows promising results, effectiveness and efficiency.

The paper is structured as follows. Section II is devoted to materials and methods. Experimental results and discussion are covered by Section III. Finally, Section IV draws conclusions.

II. Materials and Methods

A. InSar dataset and pre-processing

In this paper, the utilized Single Look Complex (SLC) images, acquired in Interferometric Wide swath mode (IW), come from the European Space Agency (ESA) Sentinel-1A/B satellites (which work in the C-band in 12-day revisit cycle) in both Ascending and Descending tracks. The combined use of two acquisition geometries has been adopted in order to provide the best identification of the components of the surface displacement with cm-scale accuracy.

Concerning the area of study, it lies on track 117 for Ascending and on track 168 for Descending, with a total of 211 and 261 images analyzed, respectively (Table I). Thus, the number of interferograms processed for this study was 591 for the ascending and 740 for the descending tracks.

TABLE I. The Sentinel-1 InSar data ingestion

Region	*Period of time*	*Geometry*	*Path*
Emilia-Romagna	04/2017-11/2021	Ascending	117
Emilia-Romagna	06/2017-12/2021	Descending	168
Tuscany	04/2017-06/2021	Ascending	117
Tuscany	06/2017-12/2021	Descending	168

Data processing was performed through the ESA's Geohazards Exploitation Platform (GEP). The approach used, P-SBAS (Parallel Small BAseline Subset) Interferometry, is a processing chain for the generation of Earth deformation time-series and maps of the yearly mean velocity of the surface displacement [14]. It is an implementation of the SBAS approach, a well-established multi-temporal InSAR technique, and it revealed to be useful also in other fields of the Geohazards, such as the monitoring of coastal land subsidence and landslides phenomena [15]. For this study, among all the parameters to be set in the pre-processing step, it has been chosen 0.75 for the threshold of the coherence, VV (Vertical, Vertical) mode for the polarization and, finally, concerning the Digital Elevation Model (DEM), the one coming from the Shuttle Radar Topography Mission at 1 arcsec spatial resolution has been chosen (30 m/pixel).

B. The generation of the machine learning set

Given an area of observation, let us consider a surface subsidence time-series generated via InSAR data pre-processing. For each sampling time, the surface frame is made by geolocated samples of subsidence in cm/year. As a first step, a lattice made by 20k × 30k elements, with cell size of 100 m × 100 m, is superimposed to the frame. A sparse matrix is then generated from the lattice by taking, for each cell, the average subsidence. Finally, a cells mask is generated and applied by taking the intersection of all matrices: only cell values that are non-null in all matrices are considered as reliable subsidence values, whereas the other values are deleted.

Given the collection of masked sparse matrices $\{\boldsymbol{S}_t\}$ generated, the machine learning set is generated as follows. Select a random submatrix of 20×20 cells $\boldsymbol{S}_t^{\mathbf{20\times20}}$, corresponding to a 2km×2km surface. If more than a minimum density of 100 non-null values is available in the submatrix, then a submatrix vectorial record $\mathbf{v}$ is generated. To guarantee an upper limit to the batch size, if more than 300 non-null values are available (with respect to the total 400), only 300 of them are considered. A vectorial record is made by the sequence of coordinates (x_i, y_i) followed by the respective subsidence values z_i:

$$\mathbf{v} = [x_1, y_1, x_2, y_2, \ldots., x_i, y_i][z_1, z_2, \ldots z_i] \quad (1)$$

Finally, the machine learning set is made by the collection of vectorial records generated from the available matrices. The training and the testing sets are made by the 80% and 20% of randomly extracted vectorial records, respectively.

The machine learning task is a regression, i.e., to predict the value z_i by having the previous elements of the vectorial record:

$$x_1, y_1, x_2, y_2, \dots, x_i, y_i, z_1, \dots z_{i-1} \Longrightarrow z_i \quad (2)$$

C. The Transformer architectures

Given the machine learning task of Formula (2), different Transformer architectural models have been developed. For the sake of brevity, this section summarizes such models. The interested reader is referred to the relevant publication [10] and to our developed code [13] for detailed information.

Let us call *token* an element of the vectorial record. The task of Formula (2) is then to predict the next token of a sequence. A first approach called *Masked Language Modeling* (MLM) is to substitute part of the tokens with a *mask token* and to train the network to predict the correct value of the masked tokens given all the others. A Second approach called *Causal* (or *Autoregressive*) *Language Modeling* (CLM) is to forcefully mask all the subsequent tokens in the attention matrix and then to predict the next token given all the previous ones. Well-known examples of MLM and CLM are BERT (Bidirectional Encoder Representations from Transformers)[16] and GPT (Generative Pre-trained Transformer)[17], respectively. Since transformers operate in the space of large real-valued vectors, the first and last steps are called *input and output embeddings*, which convert the input and output tokens into vectors, respectively. Moreover, since a Transformer does not contain recurrence or convolution, to allow the model to make use of the order of the sequence, some information about the relative or absolute position of the tokens in the sequence is injected by the *positional encoding*. Fig. 1 shows a CLM architecture called Encoder-Decoder. Here, as an input embedding of the coordinates and of the subsidence value, linear projections $\mathbb{R}^2 \longrightarrow \mathbb{R}^{256}$ and $\mathbb{R} \longrightarrow \mathbb{R}^{256}$ are used, respectively. Similarly, as an output embedding, a linear projection $\mathbb{R}^{256} \longrightarrow \mathbb{R}$ provides the predicted subsidence. The positional encoding is implemented by summing to each element of the input embedding sine and cosine functions whose wavelengths form a geometric progression from 2π to $10000 \cdot 2\pi$. Multi-head attention consists of several attention layers running in parallel (6 layers in the proposed approach). An attention function is a mapping of a query and a set of key-values pairs to an output, where the query, keys, values and outputs are all vectors. The output is the weighted sum of the values $\boldsymbol{V}$, where the weight assigned to each value is computed by a compatibility function of the query $\boldsymbol{Q}$ with the corresponding key $\boldsymbol{K}$:

$$Attention(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = softmax(\boldsymbol{Q}\boldsymbol{K}^{\boldsymbol{T}} / \sqrt{d_k})\, \boldsymbol{V} \quad (3)$$

where $\boldsymbol{Q}$, $\boldsymbol{K}$ and $\boldsymbol{V}$ are matrices packing together multiple queries, keys of dimension d_k and values of dimension d_v, respectively. In addition to attention sub-layers, a fully connected feed-forward network is applied to each position separately and identically. In the structure of Encoder-Decoder attention layer, represented in Fig. 1, the encoder is made by a stack of 8 identical layers, each made by a multi-head self-attention sub-layer and a feed forward network. Moreover, the output of each sub-layer and a residual connection are normalized. Similarly, the Decoder is made by 8 identical layers, each having, with respect to the Encoder, a third sub-layer performing multi-head attention over the output of the encoder. Finally, two approaches have been investigated: Encoder-Encoder (trained via MLM) and Encoder-Decoder (trained via CLM).

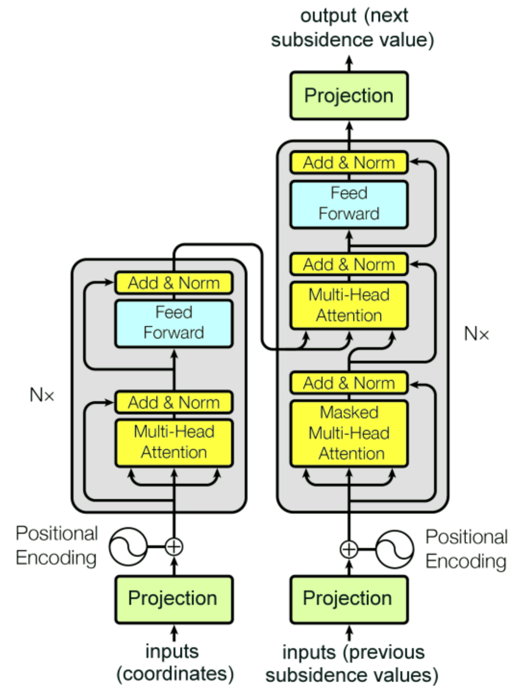


Fig. 1. Architecture of the Encoder-Decoder Transformer (adapted from [10])

In addition, a different approach has been also developed, based on *Vision Transformers* (ViT)[18]. ViTs are purposely designed for computer vision tasks, as an alternative to CNNs. The idea is to break down images in a series of patches, that are then transformed into vectors, and considered as words in a conventional transformer. In the approach developed in this paper, the DEM of the area under observation has been added to the other inputs, as an image representing the elevation data. Two approaches are possible: ViT-Encoder (trained via MLM) and ViT-Decoder (trained via CLM).

III. Experimental results and discussion

To show the effectiveness of the proposed approach, experimental studies have been carried out with different input dimensions. In particular, since the performance of the various Transformers architectures are similar, for the sake of simplicity in the following only Encoder-Decoder will be considered as a representative. The transformer has been compared with KNN regression, a non-parametric method that predict the output value by using the average (weighted by distance) among the neighborhoods. The size of the neighborhood has been set to the 25% of the training set, using cross-validation, as the size that minimizes error. Fig. 2 and Fig. 3 show the Mean Absolute Error (MAE) and the R-squared coefficient (i.e., the determination coefficient) on the testing set, for different input dimensions. It is clear that Transformers achieves promising results, up to a

MAE of .26 cm (MSE of 0.17), versus a MAE of .42 cm (MSE of 0.44) with KNN.

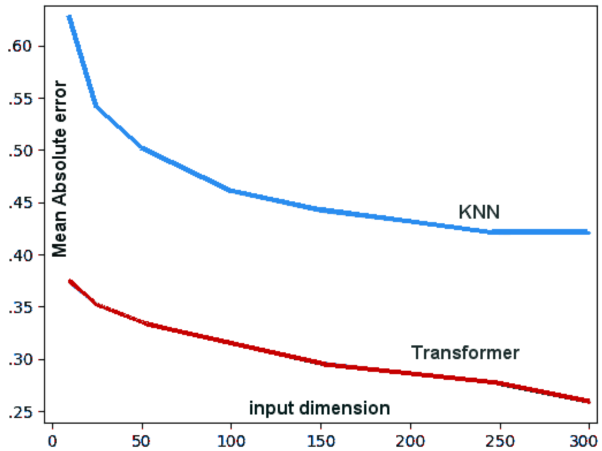


Fig. 2. Mean Absolute error of subsidence (in cm) on testing set for different points: KNN vs Transformer.

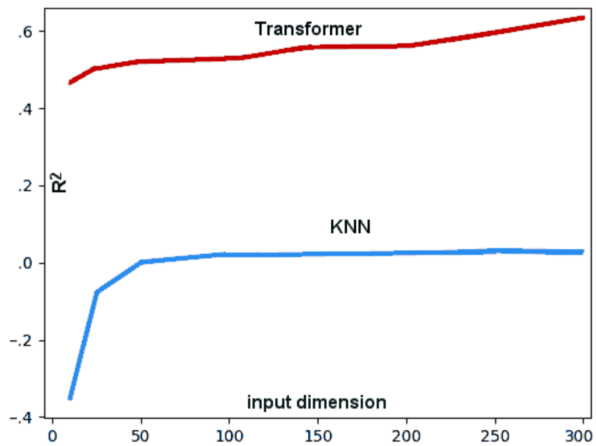


Fig. 3. R-square on testing set: KNN vs Transformer.

To better show the output provided by the method, a scenario is also illustrated. Fig. 4 and Fig. 5 are maps of the yearly mean velocity of the surface displacement (cm/year) in both ascending and descending tracks, respectively. As it was mentioned before, the combined use of both ascending and descending geometries can lead to a better identification of the surface displacement components. In this case, the maps show deformation signals corresponding to some of the main cities in Central Italy, such as Bologna, Modena, Reggio-Emilia and Pistoia, areas where for both tracks the deformation pattern is a circular range increase (ground motion away from the satellite in the Line Of Sight (LOS)) and thus consistent with ground subsidence. Other local deformations occur in a few limited areas on the Apennines mountain range, likely caused by landslides. Generally, the level of coherence is good in urban areas but at some locations loss of coherence still occurs.

In particular, in Fig. 6, regarding the municipality of Carpi (Emilia Romagna region) the density of the coherence points is very low, because a lot of deformation signal is lost, in both Ascending (a) and Descending (b) tracks, leading to a poor interpretation of the results. On the other hand, in Fig. 6 (c)(d) the additional surface motion samples released by the model on the neighborhood points sensibly increases the interpretability of the results in both Ascending and Descending geometries. Indeed, the enhanced density of the surface displacement time-series may help to detecting the areas affected by subsidence, from both quantitative and qualitative point of views. As a final outcome, Fig. 7 shows the subsidence of a single point occurring in the same area, over the time period of analysis, for both ascending (a) and descending (b) tracks.

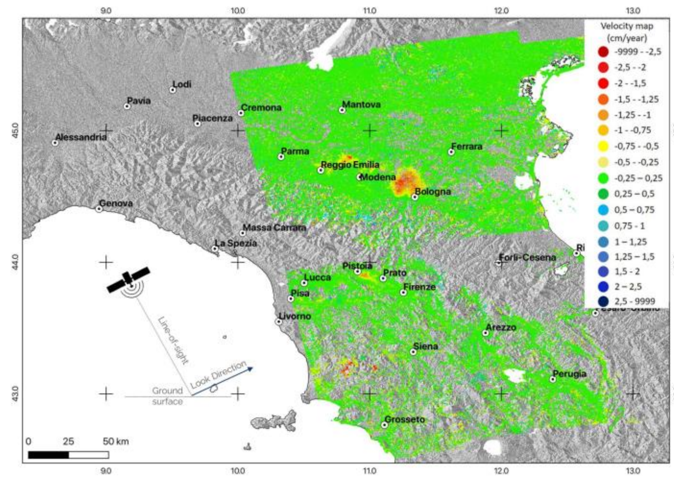


Fig. 4. Map of the yearly mean velocity of the surface displacement (cm/year) in the LOS direction across Central Italy from Ascending Sentinel-1 (04/2017-11/2021). Background by Copernicus Land Service DEM, of 25 m/pixel resolution.

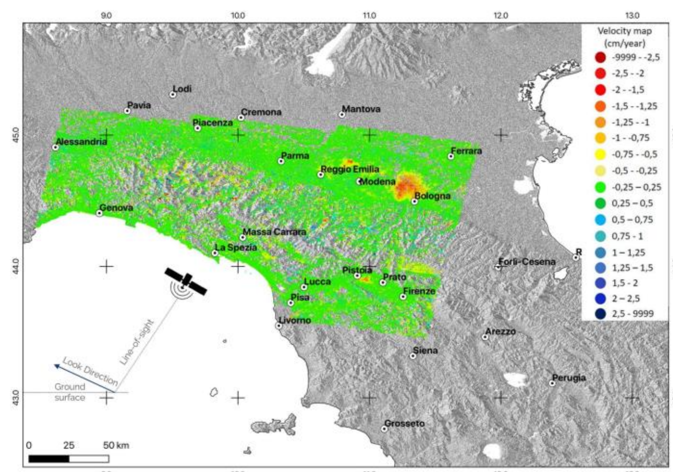
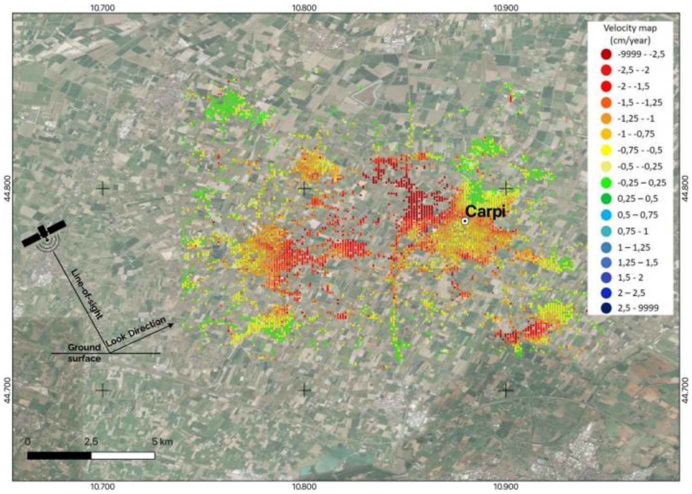


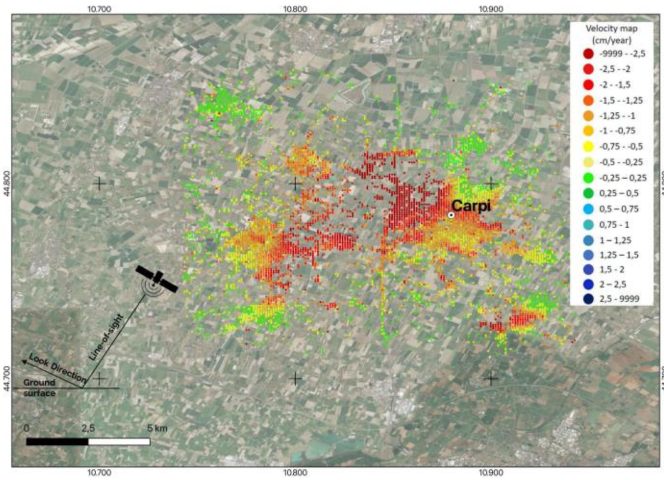
Fig. 5. Map of the yearly mean velocity of the surface displacement (cm/year) in the LOS direction across Central Italy from Descending Sentinel-1 (06/2017-12/2021). Background by Copernicus Land Servie DEM, 25 m/pixel resolution.

IV. Conclusions

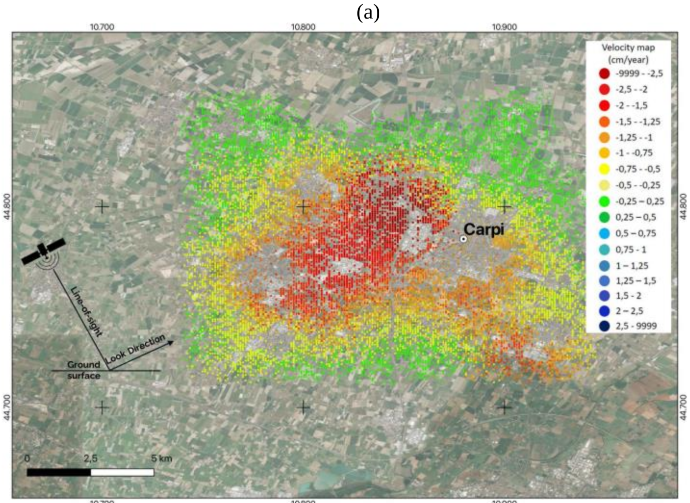
Monitoring land subsidence can be considered strategic for many stakeholders ensuring actions for environmental, socio-economic and technological interests, as well as for policy makers promoting sustainable practices in water management. In addition to traditional ground sensors, InSAR technology can enable an effective situational awareness supposed to provide a



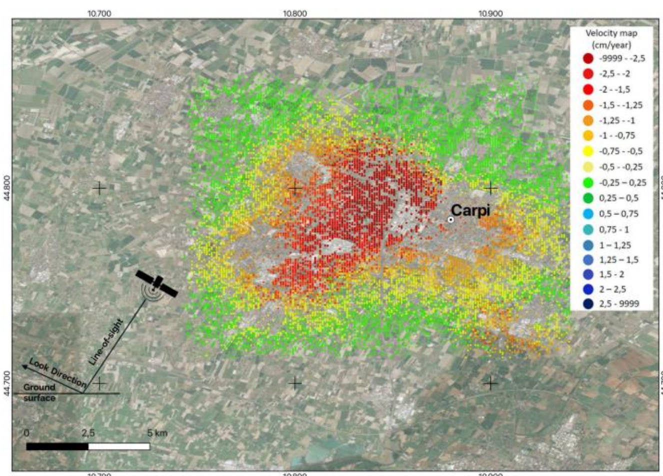
(a)



(b)

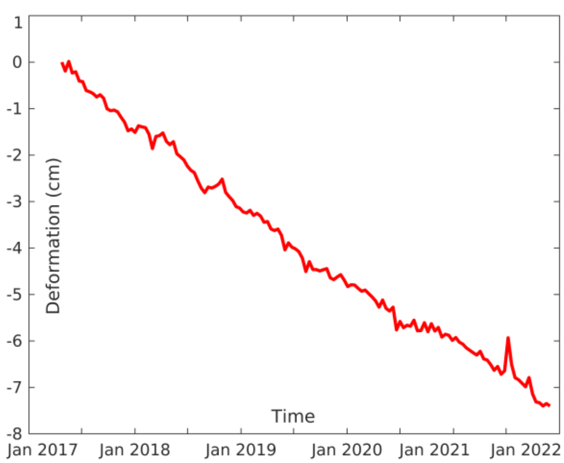


(c)

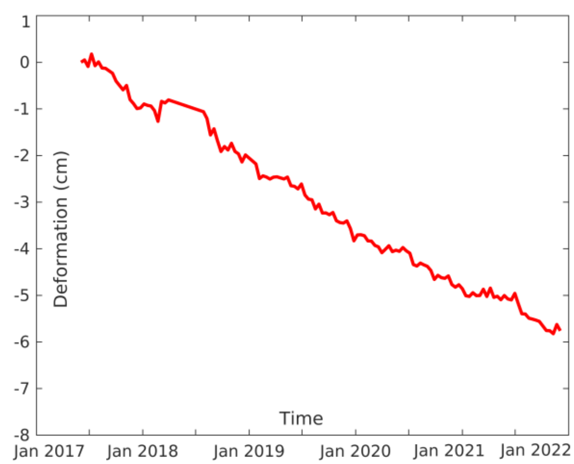


(d)

Fig. 6. Scenario of pattern of surface subsidence occurring in the western side of Carpi (Italy), in the ascending (1/06/2019) and descending (18/05/2020) tracks: (a)(b) original data, (c)(d) additional data generated by the proposed model.



(a)



(b)

Fig. 7. Time-series of a single point within one of the most subsidence-affected areas in Carpi, occurring over the time period of analysis, both in Ascending (a) and Descending (b) tracks.

high density of surface motion samples. However, one of the main limitations of this technique is represented by the interferometric coherence. To overcome this problem, this study shows that a solution based on Transformers can achieve promising results in terms of reconstruction of missing samples. Future works will focus on more experimentation and validation of the proposed method to show its effectiveness.

Acknowledgment

This work has been supported by the Italian Ministry of University and Research (MUR), formerly Ministry of Education, University and Research (MIUR), in the framework of the CrossLab project (Departments of Excellence), and under the National Operational Programme on Research and Innovation 2014-2020. This work is supported by the ESA sponsorship to the Geohazards Exploitation Platform ID. 67274 to CP, all InSAR processing was conducted on the G-TEP platform.

References

[1] Pulella, A.; Sica, F (2021). Situational Awareness of Large Infrastructures Using Remote Sensing: The Rome–Fiumicino Airport during the COVID-19 Lockdown. Remote Sens. 2021, 13, 299. https://doi.org/10.3390/rs13020299.

[2] A Ciaramella, Cimino M G C A, B Lazzerini, F Marcelloni, "Using Context History to Personalize a Resource Recommender via a Genetic Algorithm", Proceedings of the Tenth International Conference on Intelligent Systems Design and Applications, 2010 (ISDA '10), 965-970 (2010), doi: 10.1109/ISDA.2010.5687064.

[3] Galloway, D. L., and Burbey, T. J., (2011). Regional land subsidence accompanying groundwater extraction. Hydrogeology Journal, https://doi.org/10.1007/s10040-011-0775-5.

[4] Herrera-García, G., Ezquerro, P., Tomás, R., Béjar-Pizarro, M., López-Vinielles, J., Rossi, M., ... & Ye, S. (2021). Mapping the global threat of land subsidence. Science, 371(6524), 34-36.

[5] Béjar-Pizarro M, Guardiola-Albert C, García-Cárdenas RP, Herrera G, Barra A, López Molina A, Tessitore S, Staller A, Ortega-Becerril JA, García-García RP. Interpolation of GPS and Geological Data Using InSAR Deformation Maps: Method and Application to Land Subsidence in the Alto Guadalentín Aquifer (SE Spain). Remote Sensing. 2016; 8(11):965. https://doi.org/10.3390/rs8110965

[6] European Spatial Agency, Radar vision for Copernicus, www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1, acessed Feb 25, 2021.

[7] Duan W, Zhang H, Wang C, Tang Y. Multi-Temporal InSAR Parallel Processing for Sentinel-1 Large-Scale Surface Deformation Mapping. Remote Sensing. 2020; 12(22):3749. https://doi.org/10.3390/rs12223749.

[8] Xinyao Sun, Aaron Zimmer, Subhayan Mukherjee, Navaneeth Kamballur Kottayil, Parwant Ghuman, Irene Cheng, DeepInSAR: A Deep Learning Framework for SAR Interferometric Phase Restoration and Coherence Estimation (2019). *arXiv:1909.03120v2*.

[9] Rouet-Leduc, B., Jolivet, R., Dalaison, M. et al. Autonomous extraction of millimeter-scale deformation in InSAR time series using deep learning. Nat Commun 12, 6480 (2021). https://doi.org/10.1038/s41467-021-26254-3

[10] Vaswani, A., Shazeer, N. *et al.* (2017). Attention is all you need. NIPS - Advances in neural information processing systems, 30.

[11] F A Galatolo, M G C A Cimino, G Vaglini, "Formal Derivation of Mesh Neural Networks with Their Forward-Only Gradient Propagation", Neural Processing Letters, 53:1963-1978 (2021), doi: 10.1007/s11063-021-10490-1

[12] F A Galatolo, M G C A Cimino, G Vaglini, "Using stigmergy as a computational memory in the design of recurrent neural networks", Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods, 1:830-836 (2019), doi: 10.5220/0007581508300836

[13] F. A. Galatolo, Github repository, Cogsima2022, https://github.com/galatolofederico/cogsima2022, 2022.

[14] De Luca, C., Cuccu, R., Elefante, S., Zinno, I., Manunta, M., Casola, V., Rivolta, G., Lanaro, R. and Casu, F. (2015) An On-Demand Web Tool for the Unsupervised Retrieval of Earth’s Surface Deformation from SAR Data: The P-SBAS Service within the ESA G-POD Environment. Remote Sensing Journal, https://doi.org/10.3390/rs71115630

[15] Cigna, F., and Tapete, D. (2021) Sentinel-1 Big Data Processing with P-SBAS inSAR in the Geohazards Exploitation Platform: An Experiment on Coastal Land Subsidence and Landslides in Italy Remote Sensing Journal, https://doi.org/10.3390/rs13050885

[16] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2019, *arXiv:1810.04805v2*.

[17] T.B. Brown *et al.*, Language Models are Few-Shot Learners, 2020, *arXiv:2005.14165v4*.

[18] Dosovitskiy A. *et al.*, An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, 2021, *arXiv:2010.11929v2*.