Politecnico di Torino

Master Degree on Petroleum and Mining Engineering
A.a. 2020/2021
Ottobre, 2021

Master Degree Thesis

Application of the clusterization data approach to the satellite measures of altimetric variation in the Emilia-Romagna Region (Italy)

Tutors:
Prof. Vera Rocca
Prof. Alfonso Capozzoli
Ing. Luisa Perini

Candidate:
Alberto Manuel García Navarro

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the Central University of Venezuela and the Polytechnic of Thurin for the magnificent academic opportunities, as well as to all the docents and personal that contributes to an integral part of society as it is the educational area.

To the Emilia-Romagna Region and Edison Stoccaggio for the data and their permission to be used for developing this thesis.

To my tutors in Italy, professors Vera Rocca and Alfonso Capozzoli, for their continuous support and always excellent guidance, without which this work couldn’t have been possible.

To all my professors at the Central University of Venezuela, in particular to my academic tutor professor Miguel Castillejo.

To Luisa Perini from the National Council for the Research in Bologna, for its visionary collaboration with the DIATI that gave start to this work.

To Roberto Chiosa for his invaluable support on the R programing field, corner stone of this thesis.

To Raffaele Stefanelli and Timur Gukov from Edison Stoccaggio, for their gently support regarding gas storage activity on the Ravenna area.

To Andrea Chahoud from the ARPAE Emilia-Romagna, for its kindly support in the area of water management plans for the Emilia-Romagna region.

To Nicola Santocchi from the Ministry of Economic Development, for its collaboration regarding the hydrocarbon activities data in Italy.

To my mother, father and brother.

To my companion Mattea, the fancy rat of my life.

To my big brothers in life Jan and Anthony.

To my friend and life counselor Paolo.

To my *fratello* Giuseppe, a.k.a. John.

To all my friends and colleagues that have the patience to bear with me and know when to give me a hand.

SUMMARY

LIST OF FIGURES

LIST OF TABLES

Abstract

The clusterization and decomposition techniques for time series have been around for more than 50 years, but is now that the computational power at disposition makes them an appealing strategy to deal with massive data sets comprehending millions of entries. Subsidence, as natural and anthropogenic originated phenomenon, constitutes the perfect sample to put them in practice since it requires a great number of measured points in order to properly characterize an entire region in terms of ground vertical movement.

The scope of the thesis is the definition and the implementation of a methodological approach for the identification and quantification of the area affected by subsidence and uplift due to underground gas storage (UGS) activities. Because of the peculiar seasonal and cyclical behavior of the anthropogenic events under analysis, the points affected by UGS could be identified by a similar special and temporal ground movement behavior. Furthermore, according to the amplitude of the movements, the point can be subdivided according to their 'vicinity' to the UGS, or, in other terms, according to the effects of UGS.

The developed methodological approach combines decomposition analysis of normalized time series data, partitive and hierarchical clusterization (4 and 8 classes each) for subsidence satellite InSAR measures, and GIS mapping. R as programing language is adopted for the implementation of the algorithm. The time series decomposition analysis allows identifying three important components: seasonal, trend and remainder components. The seasonality is the key element adopted for the analysis of the UGS effects, because of its seasonal and cyclical nature. The clusterization allows grouping, separating and categorizing different classes of movement on which points will respond in the same way, even though not physically connected. Therefore, points (and so the area) affected by UGS can be identified and grouped according to their seasonality: gas injection and consequent ground uplift during the summer period, gas production and consequent surface subsidence during the winter period. Furthermore, because of each cluster is characterized by mean values, also the areal distribution of the magnitude of vertical subsidence (i.e., subsidence maps) can be inferred.

The reliability of the developed approach was successfully tested on a UGS located in the Po Plain. InSAR data and position of the UGS concession (in GIS environment) represent the input data of the analysis; the results, in terms of areal influence of the subsidence/uplift movement was validated via the results of a previous research work (Codegone et al, 2016). It is worthy to be mention that the location UGS concession (available from public sources) does not necessarily correspond with the extension of subsidence area, the accurate determination of which is the scope of the present work.

The novel developed approach, not founded on previous works, provides a true new point of view to study the ground movement phenomenon opening also unexplored possibilities and new opportunities for applications in many other related fields. For example, the analysis of trend component shows great potential in detecting natural subsidence trend if a sufficient amount of data could be analyzed, both in terms of time scale and area of investigation.

Keywords: subsidence, clusterization, R, time series decomposition analysis, GIS, InSAR, SqueeSAR

1 INTRODUCTION

This thesis, as an experimental work on its spirit, looks for paving the road to a new generation of subsidence analysis in accordance with the development in the late 1990's and during the last 20 years of better tools to perform data-enriched analysis with lower data processing times.

Subsidence or uplifting, as expressions of land vertical movement have always been, and continue to be, of great concern especially in case of high urbanized area, like the Emilia-Romagna region. No matter the case: for habitational purposes, industrial development, transportation or agricultural activities, the change on altitude on the ground can involves serious consequence in terms of material damages and human lives. Since the mapping and ground measures by optical instrument, up to the most advanced satellite sensor measures (InSAR) they all have one goal in mind: to measure, as exactly as possible, how much and how fast surface changes its vertical position.

The Italy territory in general, and the high industrialized and urbanized Po Plain in particular, counts with a long history of subsidence analysis and engineering work in order to minimize its impact on structures and infrastructure safety.

As well recognized, the ground movement in the Po Plain is attributed to both anthropic and natural events. This thesis proposes to explore a new way to analyze the anthropic related behavior of subsidence by performing a multi approach analysis of a peculiar zone of the Ravenna area, on the Emilia-Romagna region. This initiative born as a collection of efforts and interest also from the ARPAE (regional level organism in charge of monitoring subsidence among other competencies) which expressed great interest on giving new insides and have a fresh perspective to analyze the massive amount of satellite processed data.

Cluster and decomposition analysis aren't exactly the latest of discoveries for analyzing data, they have been around during all the twentieth century, but is from a couple of decades up to our days that the computational power available at reasonable cost makes them appealing. Therefore, this thesis proposes to analyze the subsidence data, as time series retrieved from InSAR satellite measurements, by stripping out its seasonal components and clustering them. The focus is the identification of ground movements due to the underground gas storage activity, because of its well-defined seasonal and cyclical behavior. The scope of the developed algorithm is an early stage of subsidence pattern-recognition form measure points spatially distributed over selected analyzed areas. R as programming language and ArcMap as GIS software have been selected as data analysis tools

2 STATE OF THE ART

2.1 Subsidence

Subsidence refers to a surface point vertically moving in a descent trajectory on time over the Earth's surface in relation to the sea level. At the same time uplifting exist as an expression of a vertical ascending condition, both phenomena can arise from natural or anthropical causes (Fish Bureau USA, 1981).

The natural component it's mainly represented by geological phenomena (erosion, sedimentation, tectonic movements, post-glacial rebound…). On the other hand, an anthropic component is influenced by fluid injections or removal and structure constructions, as more important factors. Times scales move very differently one from another: natural components take from years to geologic eras; anthropic effects can be seen even monthly during a single year. Even more important for recognize between them are their scale of magnitude for the same time frame: while a geological component could reach few millimeters in a year (unless we consider volcanic events or violent earthquakes) man-created subsidence can reach from few to several centimeters, so taking into consideration the fact that this phenomena can be studied by superposition of effects, theoretically they can be distinguished when human populations are in the vicinities of an affected area.

Whereas the vertical component is maybe the clearest evidence of ground surface displacement, exist also a horizontal component one that can be of four types: uniform translation, rotation, extension or contraction (Land Information New Zealand [LINZ], 2015). Among the mechanism that drives them it is convenient to separate them on deep-seated: plate tectonic deformation, fault rupture, deep zones earthquakes, etc.; and shallow-seated: small-scale rockfall, large-scale earthquakes, anthropic induced landslides, tectonic lateral spreading and area-wide ground stretching; or anthropic liquefaction-induced ground oscillation, among the principals (Land Information New Zealand [LINZ], 2015). For the case of anthropic related subsidence, shallow seated events are far more important for the analysis. Nevertheless, when to take into consideration both components, vertical and horizontal, for analyzing ground surface movement highly depends on the area characterization and scale considered. Faulty or tectonic active areas (Guo et. al, 2018) like San Andreas Fault area in the United States, or known human-caused soil liquefaction zones for example (Caracas-La Guaira Viaduct 1 fall on 2006 in Venezuela for instance), cannot avoid a mixed analysis. When regarding to the scale only the needs of the study will tell when few millimeters are representative, maybe for a single house unit sitting over 100 m2 would be more sensitive than a survey ran over a small town of 100km2 (see Laouami, 2019 and LINZ, 2015).

Carmina and Donati (1999) on their analysis of the subsidence of the Po plain area, cite the method proposed by Pirazzoli and use this simple equation to analyze the components that determine, algebraically, the module of velocity for the vertical movement of a certain area:

$$Vtot = Vt + Vsl + Vc + Vpgr + Va \qquad (1)$$

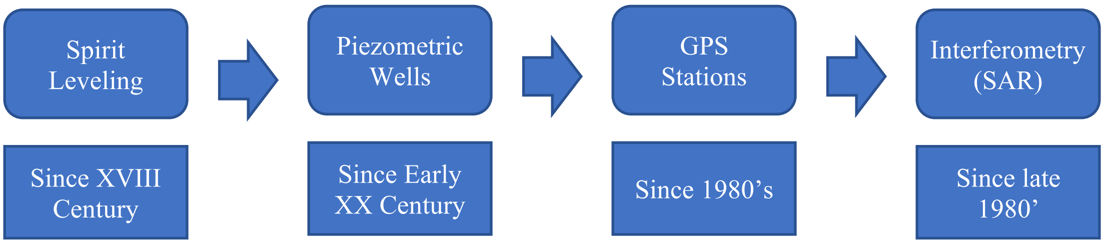
Where for natural components Vt represents the tectonic effect, Vsl and Vc are sedimentation and its compaction, $Vpgr$ the post glacial rebound related to the Pleistocene deglaciation. Va accounts for all anthropogenic components of vertical movement. Take into consideration that positive values will indicate uplifting values and negative subsidence ones, in the end the algebraic sum of all factors will determine the final effect seen by the area.

When referred to the oil industry it is important to note that compaction, and its associated effect of surface subsidence, are much more pronounced for shallow, unconsolidated reservoirs than for the deeper, more competent sandstones. It is therefore necessary to experimentally determine the compressibility of shallow reservoir sands in order to estimate to what degree compaction will enhance the hydrocarbon recovery, and also, to enable the prediction of the resulting surface subsidence, which can cause serious problems if the surface location of the field is adjacent to the sea or a lake. Usually, the deformation of unconsolidated sands is inelastic and leads to complications in relating laboratory measured compressibilities to the in-situ values in the reservoir (Dake, 1983).

This kind of subsidence analysis foresees mainly the following steps: data acquisition (using leveling surveys using benchmarks or satellite interferometry for instance), definition of the tensional and deformative behavior of the interested geological formation (seismic amplitude variation with angle/offset-AVA/AVO, analysis for example), surface 2D and 3D reconstruction models for analysis of subsidence areas; like linear, natural, cubic spline, thin-plate spline or ordinary kriging; and the performing of laboratory procedures like uniaxial or triaxial compressional test, as well as hydrostatic compression or oedometer test, aid to determine the stress-strain behavior of the soil. Regarding common analysis techniques to characterize the soil Mohr circle of stress and stress path plotting (load and unload) are of particular interest.

2.1.1 Historical review on anthropic subsidence evaluation

When talking about the tools used to measure subsidence, we found several ways to do it, from analog to digital solutions, as figures 1 and 2 show (consistently with the scope of our work mainly anthropic-related tools for measuring subsidence will be considered).



Spirit Leveling → Piezometric Wells → GPS Stations → Interferometry (SAR)

Since XVIII Century | Since Early XX Century | Since 1980's | Since late 1980'

Fig. 1 - General trend of subsidence measurement techniques evolution. Author.

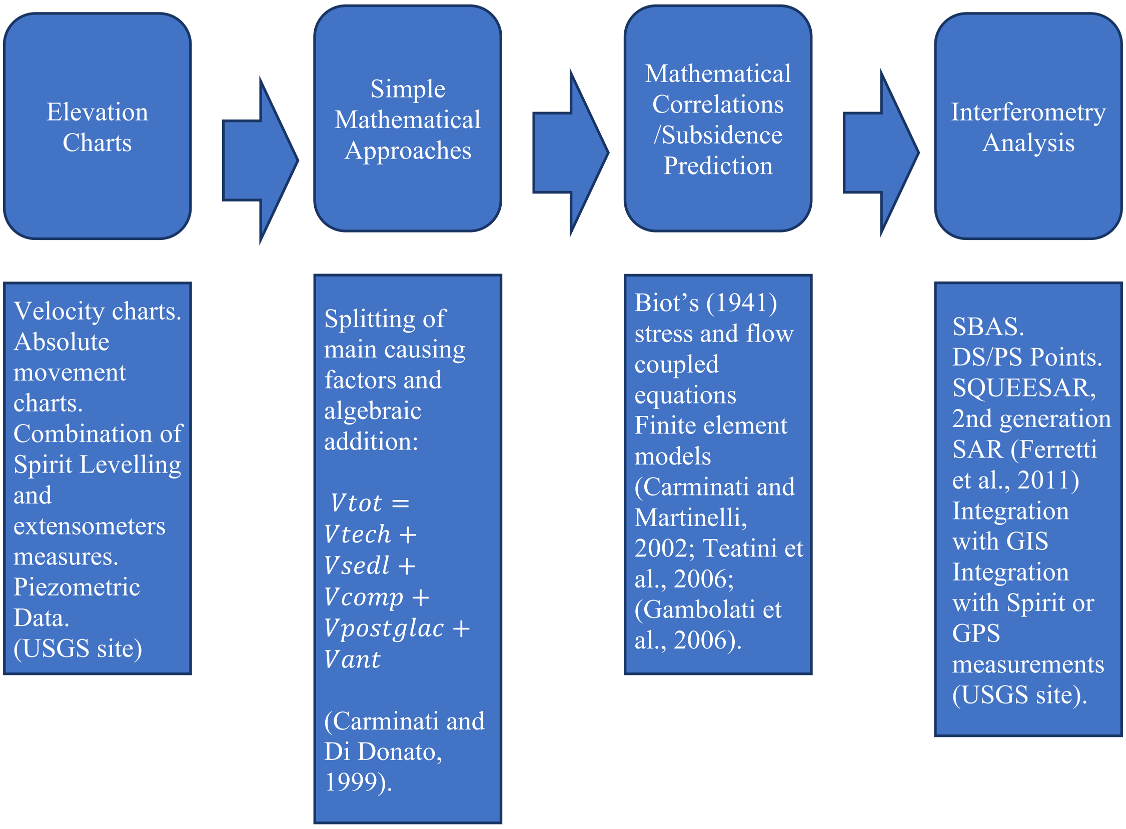

Elevation Charts
Simple Mathematical Approaches
Mathematical Correlations /Subsidence Prediction
Interferometry Analysis
Velocity charts. Absolute movement charts. Combination of Spirit Levelling and extensometers measures. Piezometric Data. (USGS site)
Splitting of main causing factors and algebraic addition:
$Vtot = Vtech + Vsedl + Vcomp + Vpostglac + Vant$
(Carminati and Di Donato, 1999).
Biot's (1941) stress and flow coupled equations Finite element models (Carminati and Martinelli, 2002; Teatini et al., 2006; (Gambolati et al., 2006).
SBAS. DS/PS Points. SQUEESAR, 2nd generation SAR (Ferretti et al., 2011) Integration with GIS Integration with Spirit or GPS measurements (USGS site).

Fig. 2 - General trend of the analysis tools evolution. Author.

2.1.2 Monitoring techniques

The conventional techniques according to Fergason, Rucker, and Panda (2015) include:

Conventional optical levelling techniques. Relational elevation differences between dedicated survey monuments should be determined through the use of differential digital level or conventional level techniques. When plausible, these measurements should be taken in relation to a stable bedrock benchmark, however it is often sufficient to observe relative differences without a direct tie to a stable point. Elevations measured by levelling should be expected to meet an accuracy of at least ±3mm and monument spacing needs to be planned to grant it.

Real-time kinematics (RTK). Are ordinarily the speediest and most efficient accessible strategies. Nonetheless, RTK GPS overview strategies don't have the degree of vertical precision as other study procedures. Heights estimated by RTK ought normal to meet a precision of basically ±21mm at 95% certainty proportion and landmark dispersing should be intended to accomplish that degree of exactness.

Static GPS observations. Have the capacity of acquiring level and vertical positions with moderately high precision and can be used to gauge both even and vertical deformation because of land subsidence and earth fissuring.

Photo-geologic analysis Classical low-sun-point (LSA). Aeronautical photography lineament examination to look for earth fissures was first portrayed exhaustively by Beckwith et al. (1991)

and can be used as a customary checking tool to evaluate changes in a subsiding landscape, but alone isn't adequate, and the lineament investigation should be trailed by a geographical surveillance of photograph lineaments and an earthbound quest for earth fissures. Traditional LSA-aerial photography has been generally supplanted by high-resolution digital aerial imagery. The utilization of this imagery has added benefits in terms of a lower cost and times. The essential impediments of this technique for earth fissure search and for use it as an observing tool comes from the coverage and quality of the remote sensing information utilized for the investigation and the state of the considered territory.

Geologic mapping/reconnaissance. Of capital importance for the monitoring program. The inspections should be performed by qualified personal on site looking for cracks, potholes, among others signs that may indicate earth fissuring. Visual inspections should be performed as contemporary as possible to the additional field measurements listed in the monitoring program. Primary limitations to the method are merely related to the ability to detects these anomalies on the ground. The vegetation can mask and limit the ability to see signs of cracking. Human activity, livestock and wild animals also limit the number of possible observations and accurate results
.
Tape-extensometers. Are useful on measuring horizontal displacement between two or more fixed points with a measurement precision of ±0.01mm. Tape extensometer arrays are part of monitoring system and are able to quantify the rate of deformation for a known earth fissure or to predict its occurrence on areas considered to have a high probability for future fissure formation. Quadrilateral arrays are more advanced tools to better understand deformation in greater detail and determine if shear deformation is taking place.

Rod extensometers. As the tape extensometers allows to register horizontal ground displacement at previously identified earth fissure locations by registering the change on position of the fixed rods set in place.

Borehole tiltmeters (BTMs). Helps on the monitor of ground deformation, in general when large infrastructure are presents and when large-scale failure is a likely to be seen. The device measures the angle of movement from the horizontal position in microradians (µrad) and deliver the data collected to electronic monitoring devices that are integrated to early warning system.

Time-domain reflectometry (TDR). Relies on a pulsed electromagnetic signal along a coupled coaxial cable to detect reflected changes derived from deformation. Both travel time and signal amplitude are measured. Position is determined by the travel time, and the severity of the deformation with the amplitude. Due to the fact that the amplitude of the signal is an approximate measure of deformation, the TDR instrumentation should be understood as a means to detect but not fully quantify deformation. The system is obsolete nowadays due to sensitivity problems related to the cable filling material.

InSAR (satellite-based synthetic aperture radar data). The principles of interferometry are described by Gabriel and Goldstein (1988). Interferometry has the ability to detect and quantify minute changes in the elevation of terrain by comparing phase variances of satellite-based measures of ground elevation, side-looking radar data between orbits of a similar trajectory. To detect the rate and distribution of subsidence occurring in an area over time InSAR imagery is used. Several InSAR-based methodologies exist and include traditional 2-pass InSAR, stacked 2-pass InSAR interpretations, and various persistent scatter (PS) InSAR data can be analyzed by direct observation of interferograms, cross-sectional presentation of vertical change, and interactive three-dimensional presentation with contouring. When assessing changes in the location or rate of regional ground deformation and to help refine characterization of bedrock/alluvial interfaces and/or general lithologic variations in the deep alluvium InSAR data analysis comes very handy (Rucker et al., 2008; Weeks and Panda, 2004). As the previous imaging techniques presented before, results from InSAR will depend on the terrain, ground and vegetation conditions, as well as anthropic activities

The approaching selection is generally based on several key according to Tomás et al. (2014) and Galloway and Burbey (2011), including:

- The cost, likely the most relevant parameter.
- The demanded accuracy and resolution, referred to the characteristic of the specific subsidence phenomenon of interest.
- The form of data (punctual, linear, spatially distributed) and measuring frequency (time between acquisitions), which are largely determined by the subsidence pattern (extent, rate, spatial and temporal variability).
- Weather conditions and land cover (urban, rock outcrops, forest, etc.).
- Flexibility of the selected method, mainly related to the possibility to selecting the time and location of the measurement acquisition, the data availability, as well as the time needed to perform the acquisition campaign.
- Geometry and kinematics of the subsidence phenomenon.

2.1.2.1 Interferometric analysis on the detail

The ARPAE 2016 and 2017 publications, related to the retrieval of subsidence measures for the Emilia-Romagna plain, provide an excellent theoretical base for the understanding of the satellite retrieve technology and would be the base for the development of this section.

Coherent radar systems and in particular the Synthetic Aperture Radar type (SAR) provides accurate measures between the satellite measuring sensor and the ground target, by registering the travel time of a signal emitted against it since an electromagnetic wave is emitted and the retrodifused signal is received back after reflecting on the target (see figs. 3 and 4). Due to the acquisition periodicity SAR data provides several measures of along the Line of sight of the satellite. When analyzed the different values for distance variation in movement of the ground targets can be detected. Differential Interferometry (DInSAR) is the most conventional tool for analysis an consist on comparing two different images of the same area, but with several factor causing lack of precision, that lead to the development by The Polytechnic of Milano of the Persistent Scatter Interferometry (PSI and PSInSAR algorithm) in order to take advantage of the stable radar fixed targets (Permanent Scatters) that keep their electromagnetic properties, making possible the reconstruction on time of average velocities and historical movement trend. In general PS correspond to buildings, rocks and significant objects in terms of visive impact.

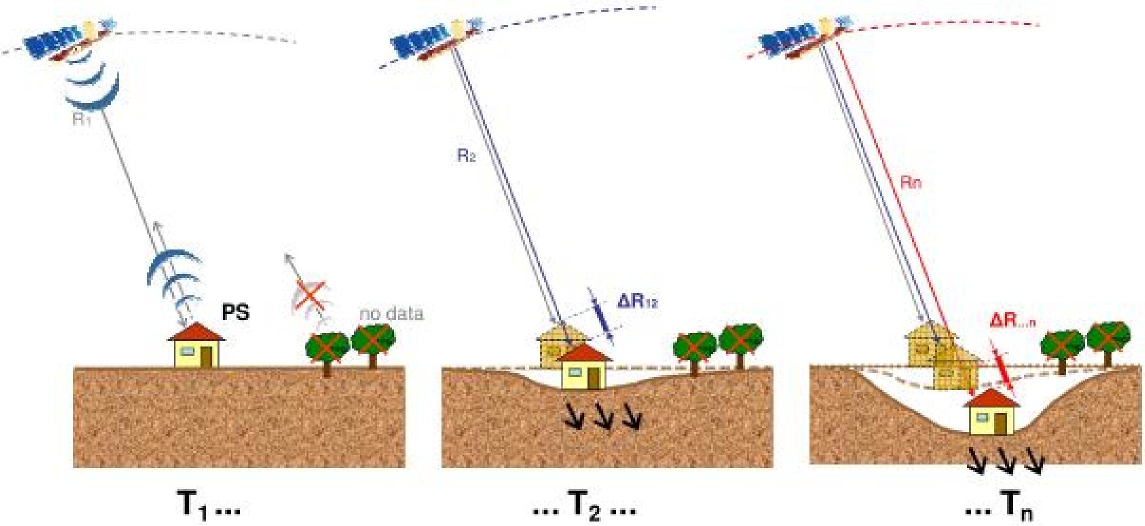


Fig. 3 - Basic principles: The base on the interferometric analysis relies on the comparison between two measures of distance made to the same target at different moments on time. ARPAE.

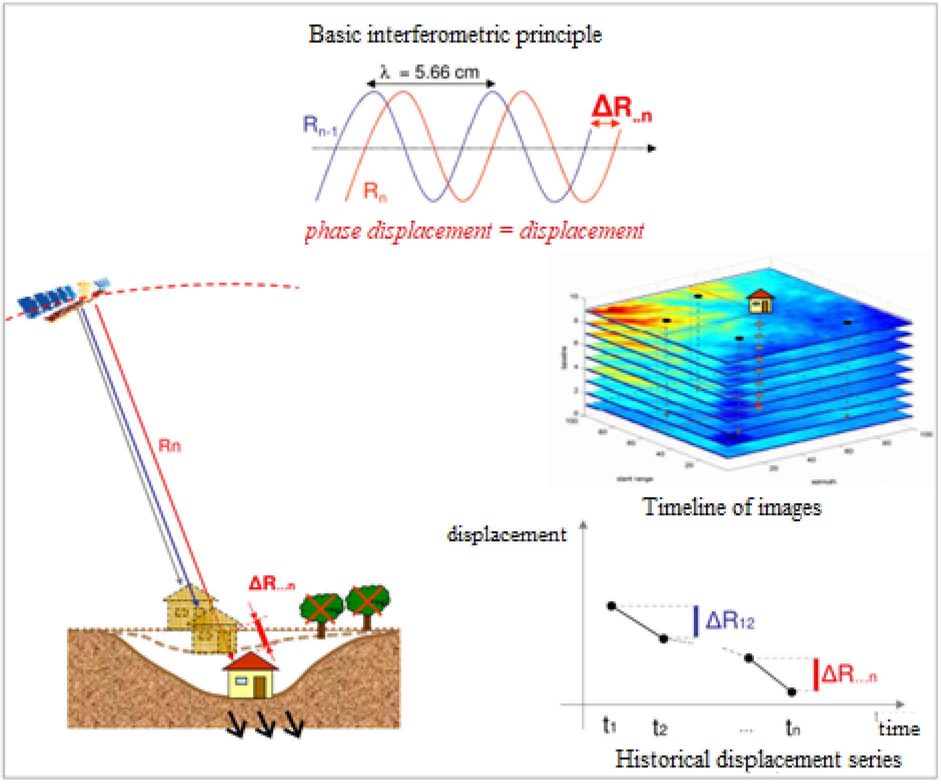

Basic interferometric principle
λ = 5.66 cm
Rn-1
Rn
ΔR..n
phase displacement = displacement
Rn
ΔR...n
Timeline of images
displacement
ΔR12
ΔR...n
t1
t2
...
tn
time
Historical displacement series

Fig. 4 - Schematic representation of the base interferometry principle for measuring the ground movement by measuring the phase difference of successive waves, as an extension of the principle for estimating historical displacement series of ground targets. ARPAE.

It is important to retake that during de satellite passage over a surface of land, the measure of the movement is referred to the projection of the present objects, due to the fact that the acquisitions are not taken perpendicular to the ground but tilted of some degrees (usually from 23 to 45) with respect to the vertical component (Line of Sight [LOS] fig. 5).

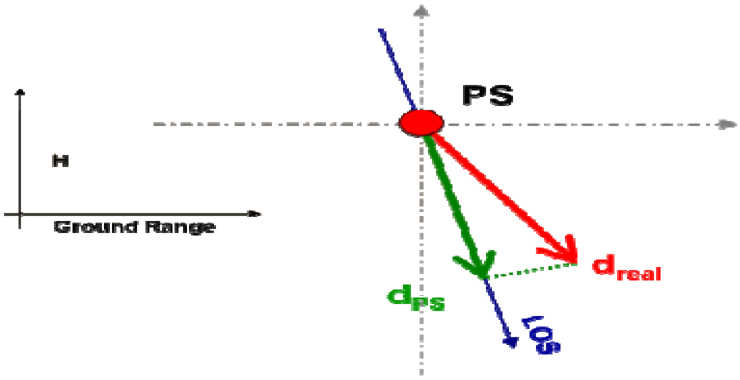


Fig. 5 - The retrieved movement is that of the projection of the real movement of the objects along the LOS direction. ARPAE.

In order to get the absolute values of movement (remember that all of the values result from comparing images and by doing so are relative) a master image is used as point of reference in time, and correlated to a zero-velocity radar target on ground (REF, reference point).

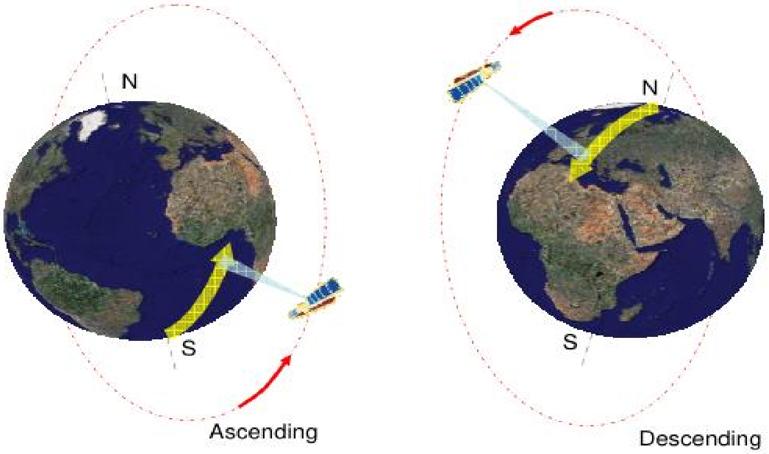


Fig. 6 - Ascending and descending acquisition geometries. ARPAE.

With regard to the acquisition Geometry, two modes are available: ascending (when the satellite passages occur from south to north) or descending (when it is done in North-south sense) see fig. 6, because of the geostationary condition of the satellite, the same area of the ground is capture by the sensors and two tracks are available. Both trajectories have the same traveling time and the incidence angle of the radiation emitted to measure the distance depends on the selected type of sensor. By using both of them it is possible to estimate the vectors of vertical movement velocity and horizontal on east-west direction for ground targets, these vectors form the absolute velocity vector that's nearly parallel to the descending LOS of the satellite. Moreover, different trajectories allow a 3D reconstruction of the movement vectors for a single area due to more possibilities of increment de MP and casting light over blind spots. It is worth to say that in the casa of a non-negligible horizontal velocity component, both tracks (ascending and descending) will behave differently for the same area (consult fig. 7).

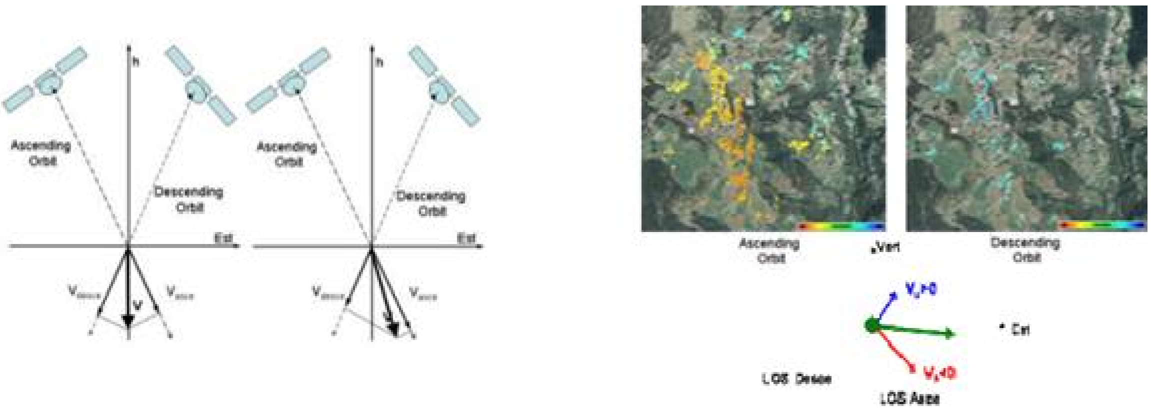


Fig. 7 - Different velocity tracks according to presence of horizontal movement. ARPAE.

To overcome the low-density target areas problem, the SqueeSAR algorithm was used by the ARPAE Emilia-Romagna, that gains point density through Distributed Scatterer or DS, that are identified and added to the analysis. Radar images can be composed of monopixel or multipixel objects on which the energy and the information on it behaves differently. PS are usually constituted by single pixels or a few ones interconnected. New DS are composed by several pixels that, as one, respond almost the same way when hit by the radar signal.

2.1.3 Diagnostic approaches for anthropogenic subsidence analysis

2.1.3.1 Terzaghi Principle of effective stress

All observable stress effects, such as compression, distortion, and changes in shearing resistance, are attributable solely to changes in effective stress (Terzaghi, 1936). Stresses applied to a saturated porous media are distributed to the solid skeleton and the pore fluid in equal measure. The former stresses cause skeletal deformations, which is why they are said to as effective. Because, by convention, stresses are positive when they are tensile and pressure is positive when it is compressive, the effective stress concept is expressed in index notation as:

$$\sigma ij = \sigma' ij - \alpha p \delta ij \qquad (2)$$

Where σij and $\sigma' ij$ are the components of the total and effective stress and p is the pore pressure. The symbol δij is Kronecker's delta, defined as $\delta ij = 1$ for i = j, and $\delta ij = 0$ for i 6= j. The parameter α is known as Biot's coefficient, added after Terzaghi's postulations. In the work of Terzaghi (1922), α was assumed to have a value of one, which is a reasonable assumption for soils but not always the case for rocks (Merxhani, 2016).

When a load is applied to soil, both the water in the pores and the solid grains carry it. Drainage (flow out of the soil) is caused by an increase in porewater pressure, and the load is transmitted to the solid grains. The permeability of the soil determines the rate of drainage. The stresses inside the solid granular fabric determine the soil's strength and compressibility (effective stresses) (University of the West of England).

2.1.3.2 Biot's coefficient α

Biot (1941) stated the coefficient of equation (3) as

$$\boldsymbol{\alpha = K/H} \tag{3}$$

Where K is the porous material's drained bulk modulus and 1/H is Biot's poroelastic expansion coefficient. It represents the change in bulk volume caused by a change in pore pressure while the stress remains constant.

2.1.3.3 Constitutive laws

The equations describing the response of a material or how a continuum deformable body reacts to changes of stress are called constitutive laws. Different models describe responses in terms of: elasticity, plasticity, viscous behavior and combinations of them. When analyzing the response of rocks and soils emerge a time-dependent response. The two principal coefficients are the Young modulus (E): as a measure of the stiffness of a body in uniaxial compression and Poisson modulus (ν) indicating the deformation ratio between lateral expansion or contraction to axial shortening or elongation.

The behavior of a given body is defined the behave "elastically" if a one-to-one correspondence between stresses and strains exists. It is plastic when gets permanently deformed after applying some stress, in this case the yield point (σ) is reached (maximum load-point before permanent deformation occurs). It is called elasto-plastic when the stress path follows different routes for the loading and unloading process indicating that a permanent deformation occurs but after releasing the applied load the material regains some of its original volume.

2.1.3.4 Analytic approaches to analyze soil subsidence

Geertsma (1973), determined that some or all of the following conditions are observed when subsidence takes place in zones above producing hydrocarbon reservoirs: 1. A reduction in reservoir pressure during the production period. 2. Production occurs across a large vertical interval. 3. contained in loose or weakly cemented rock contains the produced hydrocarbons, 4. The reservoirs have a rather small depth of burial.

Geertsma approaches reservoir compaction as a poro-elastic problem: the material is assumed to behave elastically and compacts as the pore pressure drops and the effective vertical stress increases.

The uniaxial compaction coefficient, characterizes de formation, and it is defined as the formation compaction per unit change in pore pressure reduction (Geertsma, 1973) is:

$$c_m = \frac{1}{m}\frac{dz}{dp}, or\ \varepsilon_z = c_m dp \tag{4}$$

It gets measured by lab test, usually assisted by an oedomenter-type cell for the case of loose sands and clays, very common in hydrocarbon reservoirs,

The author recognized three principal individual influences on reservoir-compaction behavior: the reduction in reservoir pressure, the vertical extent of the zone in which occurs the pore-pressure reduction, and the order of magnitude of the relevant deformation property of the reservoir rock. However, it does neglect time-dependent mechanisms, which can induce additional compaction, e.g., by grain cracking/failure or by dissolution.

In the end the approaches consent to predict with a certain of accuracy (depending on in-situ lithology, core samples quality…) the response of subsidence seen at the surface (ground displacement or strain) vs the compaction that will suffers a reservoir after fluid production and pressure changes (changes on effective stresses inside the pore.

2.1.3.5 Numeric approaches

Modern computational approaches include 3D FEM (Finite Element Method) rock mechanics model that allows to forecast the formation stress–strain behavior; and its corresponding effect in terms of subsidence evolution, induced by fluid pressure change; injection/production or aquifer support (Giani et al., 2017). This mathematical representations benefits from 3D model visualization and computational tools to evaluate subsidence as the ones described by the work of Lee et al. (2019).

2.2 Previous Studies on Ravenna and Emilia-Romagna region subsidence

Being a corner stone of this work the subsidence phenomena on the region, a deepen review is presented with a detailed description of the most relevant works, emphasizing their objectives and procedures, data used and a useful differentiation between anthropic and natural findings.

Study/Authors
"Survey Of Subsidence in The Emilian-Romagnola Plain - Period 2011-2016"
(*Rilievo Della Subsidenza Nella Pianura Emiliano-Romagnola*)
ARPA Emilia-Romagna
2018

Objectives/Performed tasks
To update the ground vertical movement values (subsidence and uplifting) through the SqueeSAR™ technique, for analyzing radar imagery acquired by RADARSAT-1 e RADARSAT-2 satellites and, for region lacking of coverage, from COSMO-SkyMed (CSK) belonging to the *Agenzia Spaziale Italiana* for the 2011-2016 period. Where elaborated 6 sites nominated: Piacenza, Parma, Bologna, Mirandola, Ravenna e Rimini, combining information of at least 15 GPS ground stations on the region.
Where verified and validated the interferometric data sets for producing diverse charts representing vertical ground movements for the above-mentioned period.

Methodology
On a first step (2016) an interferometric analysis was conducted for 6 elaboration areas, validating measuring points and estimating the mean velocities of the region for the 2011-2016 period and the corresponding time series of movement for each MP and DS. On a second stage was performed a check of the alignment for the elaboration sites at regional scale, followed by the calibration the interferometric data using 16 GPS permanent station present on the area. In order to individuate outliers and create the required charts for velocities and ground movement, statistics analysis through the usage of the SqueeSAR algorithm where performed.

Analyzed data
The studied area comprehends the plain territory of the Emilia Romagna region, total extension reaches over the 13.000 km2. An initial dataset of 1.974.150 MP divided on 6 elaboration sites, with their respective variograms plots (obtained by the kriging technique), was depurated into 1.912.781 MP once excluded anomalous outliers (noise and spurious measures) and individuated

real outliers clusters of anomalous real subsidence areas, with help of data acquired by GPS stations, cartographic data and high-resolution satellite imaginary.

Natural Related Findings
The method fully responded to the expected results. It could be further refined using dual geometry interferometric data.
Piacenza and Parma remain substantially stable, a fact highlighted in the previous period.
Ferrara and Ravenna lowering generally compatible with one natural subsidence. The latter continues the downward trend, substantially stable with maximum decreases of 2-3 mm / year except for particular areas.
Forlì-Cesena show average decreases of about 2 mm / year, down compared to the previous survey.
The coastline in its entirety has an average drop, relative to a strip of 5 km towards the hinterland, of about 3 mm / year, a further reduction compared to the previous period.

Anthropic Related Findings
Reggio reveals a diminishing on subsidence activity, even though the northern area of the capital presents up to 10 mm/year and the industrial area at Correggio's east maximums of 15mm/year.
In the province of Modena, a reduction in lowering is generally observed: except for areas at west of Carpi, maximums about 20 mm / year, south of Soliera, maximums about 25 mm / year, north of Bomporto values over 15 mm / year.
Bologna, presents a strong downsizing of the phenomenon, the reasons for which are linked mainly to the reduction of drinking water draining, concurrently with the entry into operation of the Reno-Setta shunt which has allowed greater use of surface water.
The centers of Sala Bolognese, Castello d'Argile, Venezzano and
Budrio with maximum speeds around 15 mm / year. The city of Bologna presents lowerings of a few mm/year up to a maximum of 5 mm / year.

Study/Authors
"Assessment of subsidence in the Ravenna area through an integrated InSAR / classical leveling approach"
(*Valutazione della subsidenza nell'area di Ravenna tramite un approccio integrato InSAR/livellazione classica*)
Massimo Fabris, Vladimiro Achilli, Sven Borgstrom, Mario Floris
2014

Objectives/Performed tasks
The extent of subsidence that takes place in the Ravenna municipality area is estimated by integrating the classic geometric leveling techniques with the more recent ones of SAR (Synthetic Aperture Radar) data analysis: Small BAseline Subset (SBAS). The area is of particular interest due to the presence of methane deposits whose exploitation has caused significant subsidence phenomena in the past: currently the downward rates, even if considerably reduced, are still higher than the natural ones and the connection with the phenomena related to 'average sea rise increases the hydraulic risk in the coastal area.

Methodology
For the geometric leveling data, 2 3D views were generated: one of the interpolations of the subsidence measured on the same benchmarks from 1982 to 1998 in correspondence with the "Ravenna Terra" field, and a second interpolation of the subsidence measured on the same benchmarks in the period 1999-2005 (data Arpa Emilia Romagna - Distart, University of Bologna) and superimposed on the planimetric dimensions of the fields. For the SAR data, an SBAS analysis was performed and speed maps of movements were generated, comparing the results obtained from the ERS and ENVISAT satellite data.

Analyzed data
The geometric leveling measurements used were carried out in 1982, 1986, 1992 and 1998 in the area of the "Ravenna Terra" field (Figure 1); 1999 and 2005 (surveys carried out by Arpa Emilia Romagna - Distart, University of Bologna) for the area of the entire Municipality (Arpa, 2006; Bonsignore, 2008), 84 ERS images and 96 ENVISAT images that overall cover the time interval from 1992 to 2010. The data analysis was carried out considering 5 different tracks.

Natural Related Findings
Natural subsidence effects were not analyzed.

Anthropic Related Findings
Much of the Adriatic coast has high subsidence values, especially near the "Porta Corsini Terra" field, the anthropogenic component linked to the extraction of gas probably has an important effect. The most critical area is that of the "Lido Dante", located near the "Angela-Angelina" field, with a lowering rate of about 1-2 cm / year.

Study/Authors
“Groundwater pumping and land subsidence in the Emilia-Romagna coastland, Italy: Modeling the past occurrence and the future trend”
P. Teatini, M. Ferronato, and G. Gambolati
2006

Objectives/Performed tasks
The main purposes of the paper are: (1) to ascertain whether subsidence still affecting the Emilia-Romagna coastland can be accounted for, at least partially, by the ongoing groundwater pumping; (2) to evaluate the environmental impact of existing water management plans in the Emilia-Romagna coastland over the next decade.

Methodology
(1) Implementation of advanced three-dimensional (3-D) finite element
(FE) nonlinear flow and poro-elastoplastic models;
(2) The model calibration and validation using observed ground water and land settlement records;
(3) The use of the validated models to estimate the expected land subsidence related to the planned withdrawal scenarios.

Analyzed data
The study area comprises the 10–15 km wide coastland of the Emilia-Romagna region extending between the Po River delta to the north and the Cesenatico municipality to the south. The 3-D FE grid is generated on the basis of the reconstructed lithostratigraphy of the Emilia-Romagna Coastland. Pumping tests performed in the study area over the period 1978-2002 using approximately 250 existing wells (results analyzed by the Theis-Jacob method). Estimates of the past groundwater withdrawals in the Emilia-Romagna coastland were available for the years 1976 and 2001.

Natural Related Findings
Natural subsidence effects were removed.

Anthropic Related Findings
Land settlement due to gas production is not easy to quantify. Its contribution can be qualitatively identified from the available records.
Groundwater withdrawal from the upper multiaquifer system is the main responsible for the anthropogenic land subsidence experienced by the coastland after WWII.

Land subsidence due to subsurface pumping should no longer be a serious problem over the next decades, at least in the central and northern part of the Emilia- Romagna coastland, for the planned scenario of groundwater use.

Study/Authors
"Multidisciplinary approach to the problem of subsidence in the Emilia-Romagna region" (*Approccio multi-disciplinare al problema della subsidenza nella regione Emilia-Romagna*)
Paolo Baldi, Nicola Cenni, Fabiana Loddo, Giovanni Martinelli, Marco Moro, Arianna Pesci, Michele Saroli, Salvatore Stramondo
2006

Objectives/Performed tasks
Using the DInSAR - SBAS interferometry technique, 52 images distributed over an area between the Bolognese Apennines and the Po River were processed. The temporal evolution of the piezometric heights of the wells present in the Region was also studied, with the aim to highlight areas potentially subject to subsidence comparable with DInSAR ground displacement data and with geodetic data (leveling and GPS data). In order to monitor the extent and magnitude of the subsidence phenomenon, a multitemporal analysis with SAR interferometry data was also carried out.

Methodology
Differential interferograms were calculated for the co-registered dataset. A phase unwrapping algorithm developed by Costantini and Rosen (1999) was applied to them and improved by developing a procedure that maximizes propagation in low coherence areas (Casu et al. 2006). To solve the resulting linear system applies hence a singular value decomposition or SVD. The long-term trend of the piezometric level has been estimated by means of a linear interpolation to least squares. The values represent the rate of lowering (-) or raising (+) of the water level in the wells compared to the average sea level.

Analyzed data
52 SAR images spread over an area between the Bolognese Apennines and the Po river. The SAR dataset was geometrically co-registered on a reference SAR image or "master image" (orbit 10769, 12 May 1997). The piezometry data analyzed refer to 105 piezometric wells, all characterized by an observation interval that goes from 1976 to 2004; in this period a number of measurements generally higher than 80 were performed. For each single well, an average of the annual observations was estimated in order to reduce any seasonal effects

Natural Related Findings
Periodic variations of the piezometric altitude were also highlighted, which can be associated with movements of a tectonic nature. Such evaluations are useful for a better interpretation of the temporal dynamics of the piezometric levels
The data show, in the vicinity of the foothills of the Bologna hills (anticline of M. Sabbiuno), minimum values of natural subsidence assessable in 2 - 4 mm / year.

Anthropic Related Findings
Proceeding radially from the foothills of the Bolognese hills towards the north and towards the Reno and Savena rivers there is an increase in the lowering of the soil attributable to greater fluid exploitation (mainly water), precisely where there is a greater increase in the thickness of the aquifer groups, and estimated at 50 - 60 mm / year.
The analysis of data from the piezometric monitoring network showed in Ferrara and Ravenna + 0.1-0.2 m / year, and Bologna, Modena, Reggio Emilia and Parma -0.5 m / year, where groundwater consumption is particularly high and the data leveling and GPS indicate a marked subsidence process.

Study/Authors
“A century of land subsidence in Ravenna, Italy”
P. Teatini, M. Ferronato, G. Gambolati, W. Bertoni, M. Gonella
2004

Objectives/Performed tasks
Land subsidence in the Ravenna Municipality over the past century is reconstructed by homogenizing and implementing in a GIS environment, leveling measurements carried out since 1897. Leveling surveys were performed by the Italian Geographic Military Institute, the Ravenna Reclamation Authority and the Geological Service of the Ravenna Municipality on benchmark networks which were progressively refined as the anthropogenic component of the occurrence increased.

Methodology
GIS has been employed in the analysis of land subsidence in the Ravenna area with three main goals:
(1) georeferencing and homogenizing the available information; (2) interpolating and mapping the point wise records provided by leveling; (3) integrating the interpolation outcome over the different time periods to provide a complete analysis of the process during the last 100 years. Available subsidence records are homogenized into the Gauss-Boaga (Roma-40) Italian projection system, East Zone. The land sinking rates recorded over a number of time intervals covering the entire twentieth century are homogenized on a regular square grid by the use of kriging.

Analyzed data
Row data management and input in the GIS are operated by two different approaches: 1) historical data: Land subsidence contour maps published by Salvioni and Carbognin et al. for the 1897–1957 and 1949–1972 intervals, 2) contour lines have been digitized, georeferenced, and gridded into the selected reference grid and 3) recent data: Surveys performed by the Ravenna Municipality, homogenized and validated by the UNIBO (University of Bologna). For each monitoring period, the subsidence rates computed on the comparable benchmarks, are included in the GIS and interpolated by kriging on the selected grid.

Natural Related Findings
Cumulative land settlement, accounting for both the natural and the anthropogenic components, has achieved the value of 1.6 m from 1897 to 2002 in the industrial area located between the city and the seashore, with the coastland and the historical center settled by more than 1 m.

Anthropic Related Findings
Subsidence is primarily due to groundwater withdrawal from a well-developed multiaquifer system underlying the coastland and, subordinately, to gas extraction from deep reservoirs scattered through the area and still productive nowadays.
Land settlement, occurred at an average rate of about 5 mm/year until World War II, increased greatly up to an order of magnitude in the Ravenna industrial area.
New public aqueducts using surface water during the late 1970s and 1980s has reduced the subsurface water consumption and settlement rates to the pre-war values

Study/Authors
“Subsidence rates in the Po Plain, northern Italy: the relative impact of natural and anthropogenic causation”
E. Carminati, G. Martinelli
2002

Objectives/Performed tasks
Re-evaluation and re-interpretation of the most important data sets on subsidence in the central and southeastern parts of the Po Basin that have been collected over the past century. Evaluation of the relative importance of both artificial and natural causes of subsidence. Maps showing present-day vertical velocities for the eastern portion of the Po Plain were generated.

Methodology
Present-day subsidence rates were compared with the geodetic rates obtained by Arca and Beretta (1985) based on comparison of levelling data for the years 1897 and 1957, respectively. Subsidence rates were arithmetically summed these and noted that a sharp increase (up to 40–60 mm/year) had occurred in the second half of the century, in agreement with the hydrological observations.
SAR interferometry data was considered.

Analyzed data
Local government technical reports (Regione E-R, Idroser Agenzia,1995)
National government data (M.U.R.S.T., 1997b) data by Benedetti et al. (2000).
The Regione E-R technical reports for Modena and Ferrara for 1950 to 1995.
For the Bologna, Ravenna, Forlı and Riminithe data are from 1970 to 1995.
The data for the northern sectors (M.U.R.S.T., 1997b) extend back to 1942.
Most information is from 1950 to 1990.
Area surrounding Imola for the period 1970–1999.

Natural Related Findings
The main components of natural subsidence in the Po Plain are tectonic loading, sediment loading, sediment compaction (Sclater and Christie, 1980) and post-glacial rebound. An estimate of deglaciation induced subsidence was obtained from geodynamic
models.

Anthropic Related Findings
Withdrawal of deep, slow or non, rechargeable groundwaters is probably the most important cause of subsidence in the Po Basin. This has enhanced sediment compaction and, therefore, promoted subsidence around Bologna.
The difference between present-day subsidence and the long-term geological subsidence provides a measure of the anthropogenic influence in regional and local subsidence rates. Subsidence rates range from about 0 to 70 mm/year. Maximum subsidence rates (up to 60–70mm/year) occur in two general areas: the Po Delta, east of Rovigo, and the north zone of Bologna.

Study/Authors
“Separating natural and anthropogenic vertical movements in fast subsiding areas the Po plain (N. Italy) case”
E. Carminati and G. Di Donato
1999

Objectives/Performed tasks
For the Po Plain, was calculated a regional distribution of the total natural component of vertical velocity which is the result of long-term geological processes summed to the effects of the Pleistocene deglaciation. The present-day vertical velocity of an area, Vtot, with respect to the mean sea level is the sum of several factors (tectonic effects "Vt", load sedimentation "Vl", compaction Vc", post glacial rebound "Vpgr" and anthropogenic "Va" component. Va is composed by (Vcl) associated with the acceleration in sea-level rise due to anthropic climatic warming (Vloc) induced by local human activities.
Methodology

Anthropogenic components can be inferred by comparison between the total natural vertical velocities and vertical velocities obtained from other data sets (e.g., geodetic data). Vt,Vs and Vc are resolved by means of a backstripping procedure considering the effects of sediment compaction. The decompacted thickness of the post 1.43 Myr sediments is calculated using the exponential porosity-depth relation and the parameter for shaley-sand proposed by Sclater and Christie [1980]. Vtot, is calculated from the obtained original decompacted thicknesses.

Analyzed data
The stratigraphic data used to calculate the vertical velocities due to long term geological processes are the thicknesses of Quaternary sediments and mainly come from the 200 published on-land and off-shore wells (AGIP, 1959, 1977). Geodetic data [Arca and Beretta, 1985] provide vertical velocity values for the entire Po Plain. These geodetic velocities were obtained by Arca and Beretta [1985] with a revision and a comparison between two sets of data of fundamental altimetric network respectively referred to the periods 1877-1903 and 1950-1956.

Natural Related Findings
The main components of natural subsidence in the Po Plain are tectonic loading, sediment loading, sediment compaction (Sclater and Christie, 1980) and post-glacial rebound.
An estimate of deglaciation induced subsidence was obtained from geodynamic models.

Anthropic Related Findings
In the Po Plain several factors contribute or contributed to the local component: e.g., extraction of methane-bearing water from 1938 to 1964 in the Po delta; gas and oil extraction in a few localities [AGIP, 1959; Cassano et al., 1986]; diffuse water extraction for industrial, agricultural and civil use.

Study/Authors
“Mathematical Simulation of the Subsidence of Ravenna"
Giuseppe Gambolati, Giuseppe Ricceri, Werter Bertoni, Giovanni Brighenti and Enzo Vuillermin 1991.

Objectives/Performed tasks
To analyze mathematically land subsidence with a modeling approach comprehending 3D quasi finite element and linear models.

Methodology
Land subsidence is simulated with two separate models: 1) a 3D hydrologic model of subsurface flow on a regional scale followed by a 1D vertical consolidation model applied to the site where an accurate lithostratigraphic info soil is available, 2) a 3D calculation of the settlement over the Ravenna Terra field from beginning to end of its production life. Calibration against the historical records was done by trial and error acting upon the hydromechanical parameters of the groundwater system and using drawdown data at Ravenna. Subsidence due to water pumping and gas production is predicted by the vertical consolidation model and the Ravenna Terra reservoir model.

Analyzed data
Measures from 1949 to 1986 from geodetic surveys carried out by the Istituto Geografico Militare (IGM), the Direzione Generale del Catasto, the Consorzi di Bonifica of Ravenna and Lugo and the Municipality of Ravenna. Core analyses performed on the borehole RA1 core samples including extensive and intensive properties. Fluid production and data used as an input in the mathematical model provided by AGIP S.p.A. Ravenna Terra gas production and geological data. Piezometric registers for the Ravenna area period 1950-1985,
Natural Related Findings

The numerical analysis shown that gas fields located far offshore do not have significant influence on the ground movement inland.
A major impact on coastal environment could be seen if gas production comes from pools overlain by the Adriatic coastline.

Anthropic Related Findings
Quantitative results of the subsidence model applied to the Ravenna Terra gas field, can't be directly extrapolated to the other reservoirs currently under production in the Ravenna area since most of the parameters are typical of any specific field, an "ad hoc" model should be tailored.
Land subsidence at Ravenna is primarily the result of aquitard compaction caused by extensive groundwater withdrawals from the regional unconsolidated Quaternary aquifer-aquitard system that underlies the eastern Po River plain. Water pumpage is responsible for the generalized settlement that has affected the area (up to 1.30 m in Ravenna's industrial zone).
Gas production exerts a restricted but measurable influence on land sinking.
Numerical simulations of Subsidence due to groundwater overdraft over the period 1950-1985 proved to be consistent with the outcome from the geodetic survey.

2.3 Comments on subsidence study evolution

In order to get a general view of the subsidence phenomenon analysis a brief review is presented.

2.3.1 Data acquisition techniques

Initially cartographic maps and velocity charts obtained by classic spirit levelling, evolved into satellite SAR imagery and use of GPS permanent station to correlate and validate measure points. Nevertheless, classic geometric leveling techniques are often coupled with the more recent ones of SAR (InSAR, PSSAR), being these last ones the current standard for high precision ground movement measures, fact that emerges from the important amount of published works that make use of it.

2.3.2 Strategies and tools for data analysis:

- Finite elements nonlinear flow and poro-elastoplastic models.
- 3D finite element models.
- Small BAseline Subset (SBAS).
- Interferometric analysis (differential interferograms).
- GIS (Georeferenciate systems) that allows to correlate subsidence information and other relevant data into a mapped computational environment.
- The use of kriging as a tool to homogenize the subsidence/uplifting behavior over an area when the data spatial distribution allows it.
- 1D vertical consolidation model applied to sites where accurate lithostratigraphic information of the soil is available.
- 3D calculation reservoir/basin settlement.

Nowadays 3D modeling strategies, interferometric analysis and GIS environment lead the way of the subsidence analysis.

2.3.3 Main natural components identified for subsidence:

- Tectonic loading.
- Sediment loading.
- Sediment compaction.
- Post-glacial rebound.

The work of Carmina and Donati (1999) provides ranges for the Po Plain region of about 0,1 up to 2 mm/year for all of them combined.

2.3.4 Main anthropical components identified for natural subsidence:

Subsidence is primarily due to groundwater withdrawal from a well-developed multi-aquifer system underlying the coastland and, subordinately, to hydrocarbon extraction from reservoirs (shallows having greater impact on the phenomenon), this enhances sediment compaction. Subsidence rates range from about 0 to 70 mm/year for the Emilia-Romagna region (Carmina and Donati, 1999 and ARPAE 2018). Nevertheless, other activities as building or bridges construction on unconsolidated soils (clays or loose sands), large motor vehicles on highways as well as mining activities have the potential to make circumscribed areas to subside.

3 METHODOLOGY OF ANALYSIS

3.1 Time series analysis

A time series is a sequence of data points that occur in successive order over some period of time. This can be contrasted with cross-sectional data, which captures a point-in-time. (Hayes, 2021). Budget and sales forecasting on businesses, stock market analysis for understanding the index behavior on the market; and census analysis of population growth are common examples of its utility nowadays (National Institute of Standards and Technology, 2021). Oil and gas industry gains a lot from these approaches making them of capital importance when the time comes for presenting and assessing new developments, both in economic and technical terms. Pressure and production history can be categorized as well as time series, even though aren't often treated in that way.

Subsidence, as a natural or anthropogenic induced phenomenon is familiar with time series analysis, and numerous tools have been developed in order to frame and analyze its behavior during certain periods of time. In particular we will deepen the concept of "Seasonal Decomposition Analysis".

3.1.1 Classical Decomposition Analysis

The classical method of time series decomposition originated in the 1920s and was widely used until the 1950s (Hyndman and Athanasopoulos, 2021). It consists on interpret a data time series (a list of measurements correlated on time) as a superposition of different sub-series that allow to characterize the behavior of it. As starting point Hyndman and Athanasopoulos (2021) allow us to define three basilar concepts in order to identify a time series behavior of data:

Trend: Is represented by the behavior of data on time as a whole when increase or decrease its values. It can be also referred to as the change of direction when goes from an increasing to a decreasing trend.

Seasonal: It's referred to the change of the behavior and values of the data according to seasonal factors that repeats itself periodically. This can be observed for instance on the same time of different calendar years or same days of a week. As it said before it can be natural or anthropogenic in its origin.

Cyclic: When data instead of change its behavior (rise and fall) periodically, just does it at different moments on time and according to no defined frequency we talk about cycles. These are frequent on social and economic analysis with a duration usually longer than two years. Its change of magnitude and duration is very often greater than the seasonal one

It is of paramount importance to recognize when the time series date behaves cyclical or seasonal in order to select the proper tool for analysis.

3.1.2 Extracting Seasonality and Trend from Data

For our case of interest, seasonal behavior (subsidence-related), time series decomposition as mathematical procedure allows to split and analyze a unique time series of data into different time series. In particular we look for three subdivisions when talking about seasonal data:

Seasonal (*St*): The pattern that is repeated periodically in the time frame considered; it has a well-defined shape.
Trend (*Tt*): The magnitude of rise and fall of the overall group of point that conform the original series.
Reminder (*Rt*): The outliers or noise that represent an anomalous behavior. It could be useful when studying phenomena that behaves out of the norm for unknown reasons or to identify underlying factors previously not considered.

Moreover, there are two big approaches for decomposing the original time series into the above-mentioned series: Additive or Multiplicative. For the additive approach we add algebraically the values for each component being:

$$Yt \ = Tt \ + \ St \ + \ Rt \tag{5}$$

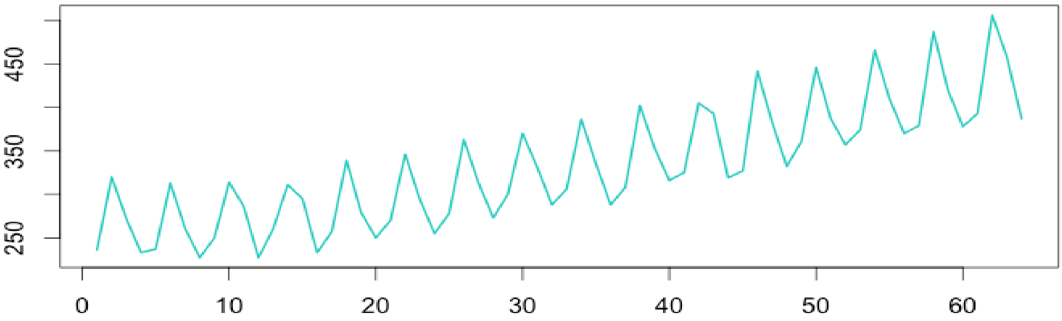


Fig. 8 - Example of additive time series behavior, note the stable amplitude of the series. Anomaly.io.

On the contrary in the multiplicative series:

$$Yt \ = Tt \times St \times Rt \tag{6}$$

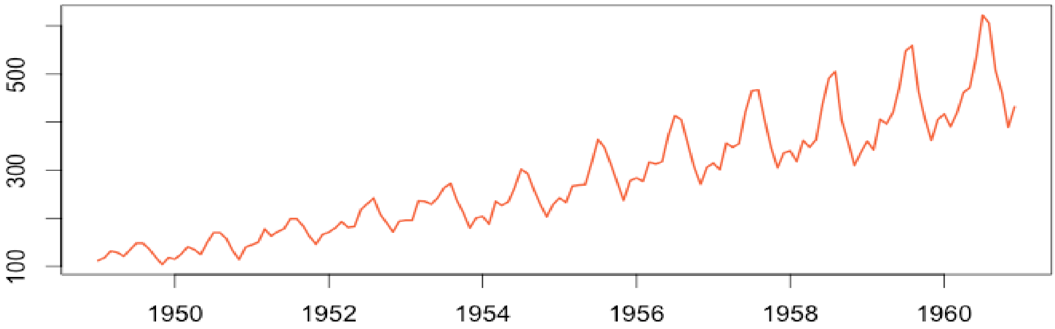


Fig. 9 - Example of multiplicative time series behavior, note the increasing amplitude of the series. Anomaly.io.

If the seasonal variation amplitude looks constant; in other words, it doesn't change when the time series value increases, we use the additive model. If the time series increases in magnitude, while the seasonal variation increases as well, we use the multiplicative model.

Another option to using a multiplicative decomposition is to initially transform the data until the variation in the series appears stable over time, then use an additive decomposition by applying a log transformation where:

$$Yt = St \times Tt \times Rt \tag{7}$$

is equivalent to

$$logYt = logSt + logTt + logRt \tag{8}$$

For the trend estimation, it is common practice to use linear, parabolic or polynomic functions usually possible to linearize and also easy to study and characterize (Ricci, 2005).

3.1.3 Moving average and smoothing

It forms the basis of many time series decomposition methods and is the first step to estimate the trend-cycle. A moving average of order m can be written as:

$$\widehat{\boldsymbol{T_t}} = \frac{\boldsymbol{1}}{\boldsymbol{m}} \sum_{\boldsymbol{j=-k}}^{\boldsymbol{k}} \boldsymbol{y_{t+j,}} \tag{9}$$

Were

$$m = 2k + 1 \tag{10}$$

That is, the estimate of the trend-cycle at time t is obtained by averaging values of the time series within k periods of t. Observations that are close to each other in time are also likely to be close in value. In this way, the average eliminates some randomness in the data, smoothing the trend-cycle component. This is an $m - MA$, meaning a moving average of m order.

3.1.4 Classical decomposition steps

As said before there are two forms of classical decomposition: an additive and a multiplicative decomposition. We will focus on a time series with seasonal period m (e.g., m=4 for quarterly data, m=12 for monthly data, m=7 for daily data with a weekly pattern).

In classical decomposition, it is assumed that the seasonal component is constant along the years that encompasses the time series. For multiplicative seasonality, the m values that form the seasonal component are sometimes called “seasonal indexes.” (Hyndman and Athanasopoulos, 2021))

3.1.4.1 Additive decomposition

Step 1: If m is an even number, the trend-cycle (Tt) component is computed by using a $2 \times m - MA$. If m is an odd number, the trend-cycle (Tt) component is computed by using $m - MA$.

Step 2: The detrended series is obtained: $Yt{-}Tt.$

Step 3: By averaging the detrended values for that season is then estimated the seasonal component. For example, with monthly data, the St for March is the average of all the detrended values of March in the data. Then seasonal component values are adjusted to ensure that they add to zero. St is given by stringing together these monthly values, and then replicating the sequence for each year of data.

Step 4: By subtracting the estimated seasonal and trend-cycle components the remainder component is calculated: $Rt = Yt - Tt - St$.

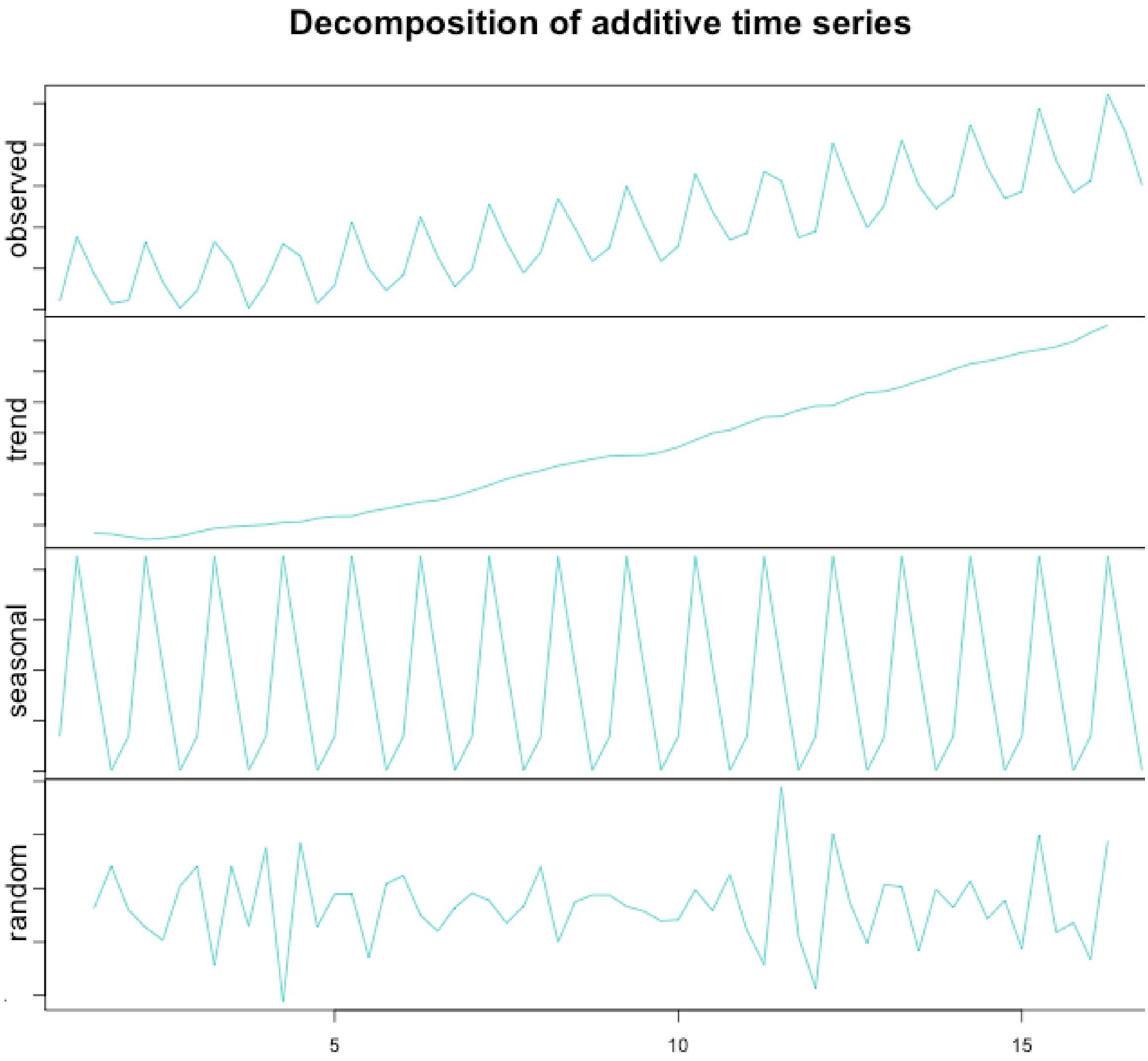


Fig. 10 - Example of fully decomposed additive series. Anomaly io.

3.1.4.2 Multiplicative decomposition

A classical multiplicative decomposition is similar, except that the subtractions are replaced by divisions.
Step 1: If m is an even number, the trend-cycle (Tt) component is computed by using a $2 \times m - MA$. If m is an odd number, the trend-cycle (Tt) component is computed by using $m - MA$.

Step 2: The detrended series is obtained: Yt/Tt.

Step 3: By averaging the detrended values for that season is then estimated the seasonal component. For example, with monthly data, the St for March is the average of all the detrended values of March in the data. Then seasonal component values are adjusted to ensure that they add to m. St is given by stringing together these monthly values, and then replicating the sequence for each year of data.

Step 4: By dividing the estimated seasonal and trend-cycle components the remainder component is calculated: $Rt = Yt/(Tt * St)$.

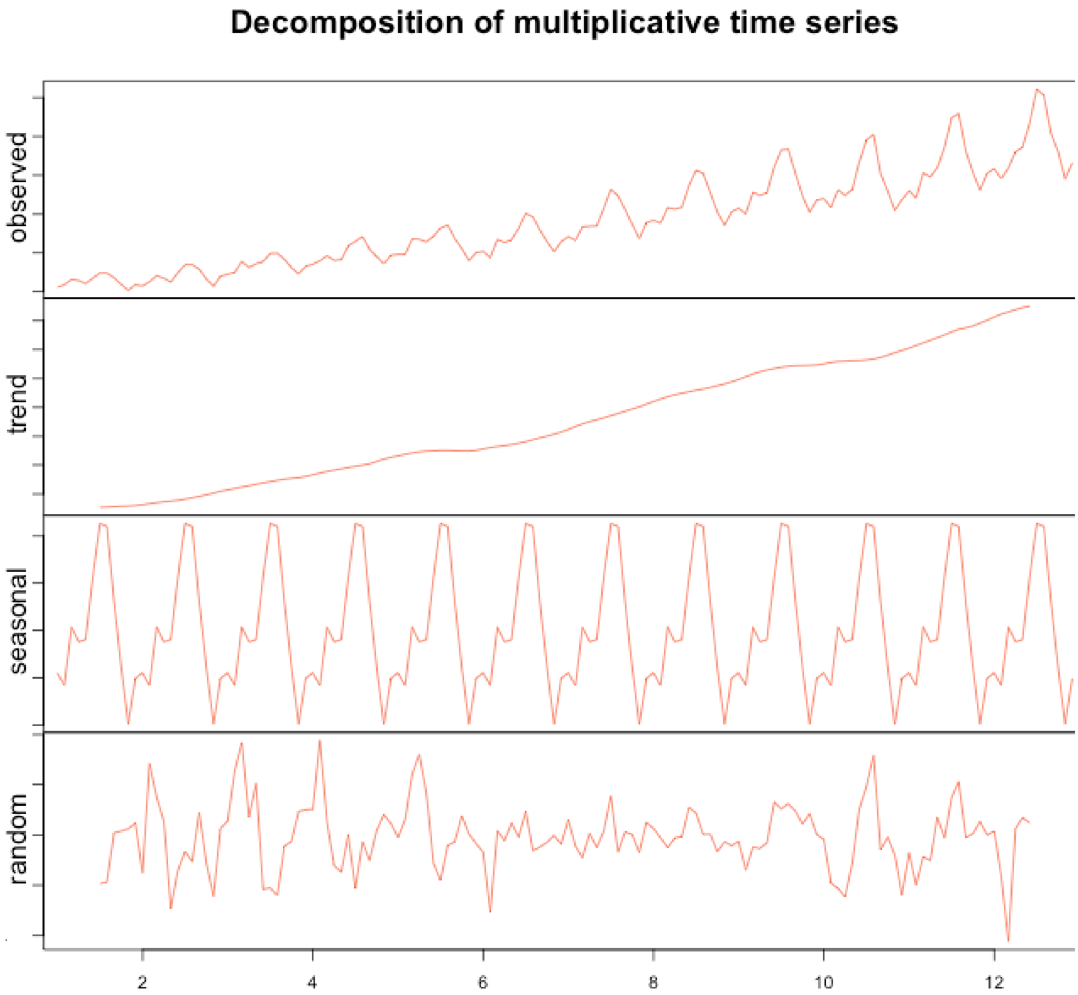


Fig. 11 - Example of fully decomposed multiplicative series. Anomaly io.

3.1.5 Analytic methods using R

Basic hypothesis for developing the decomposition (Ricci, 2007) can by described by the equation:

$$Yt = f(t) + Rt \qquad (11)$$

On which f(t) groups the seasonal and trend components ($Tt + St$), and Rt ~ NID (Nonlinear interaction decomposition) (0, σ2 [variance]), that means that all errors have a normal distribution behavior, are homoskedastic (σ2 constant) and independents among them. Programming language R allows to decompose the time series by using three different functions: decompose() and stl() on stat package, tsr() on ast package.

For the developing of calculations stl() function was selected that uses a Loess approach short for Local Regression, which is a non-parametric approach that fits multiple regressions in local neighborhood (Prabhakaran, 2021), which makes it a suitable candidate for smoothing any numerical vector.

In order to properly analyze and decompose a time series, data should be even distributed in time reason for which a proper interpolation could be considered. For the ongoing study "approxfun" R function is used as Interpolation Function (consulta appendix II), which returns a list of points linearly interpolated given data points, in order to obtain one single representative value for each considered period on the STL analysis. (R documentation). The stl() function is described on appendix I citing the software documentation.

3.2 Clusterization

3.2.1 Cluster

The definition of a cluster is imprecise and the best one relies on the nature of data and the desired results (Tan, Steinbach and Karpatne, 2018) In general it can be said that a cluster is nothing more than a collection of objects, that share some common property o characteristic which allows them to be organized in different classes or groups. Tan et at. (2018) gives a pretty detailed description of the cluster analysis tools, algorithms, types, strongness and weakness and it will be used to develop the next part of this chapter.

3.2.2 Clustering Analysis

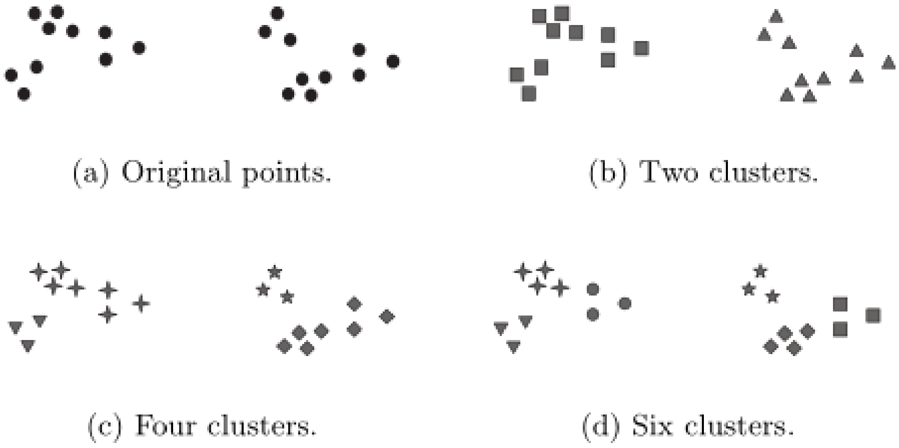
Cluster analysis can group data objects based on selected characteristics that describe both them and their relationships. The objective is that the objects within a group be similar (or related) to one another and different from (or unrelated to) the objects in other classes. The greater the homogeneity within a group and the greater the difference between groups, the better the clustering.

Cluster analysis partitions information into groups that are significant, valuable, or both. When significant gatherings are the objective, the cluster should catch the structure of the information. In any case, thus technique is utilized for data summarization to lessen the size of the set.

In the context of understanding data, clusters are potential classes and cluster analysis is the study of techniques for automatically finding classes (conceptually meaningful groups) of objects that share common characteristics

Some clustering techniques characterize each cluster in terms of a cluster prototype; i.e., a data object that is representative of the objects in the cluster that can be used as the basis for a number of additional data analysis or data processing techniques. Therefore, cluster analysis in this case study the techniques for finding the most representative cluster prototypes, and this are some examples:

Instead of applying the algorithm to the entire data set, it can be applied to a reduced data set consisting only of cluster prototypes. When a substantial reduction in the data size is desired, some loss of information is acceptable and many of the data objects are highly similar to one another data can be compressed by indexation of the prototypes presents on each cluster. Usually, cluster prototypes can optimize the finding of distance between groups or their "nearest neighbors", where the closeness of two clusters is estimated by the distance between their prototypes (see figure 12 for graphic clusterization examples).


(a) Original points.

(b) Two clusters.

(c) Four clusters.

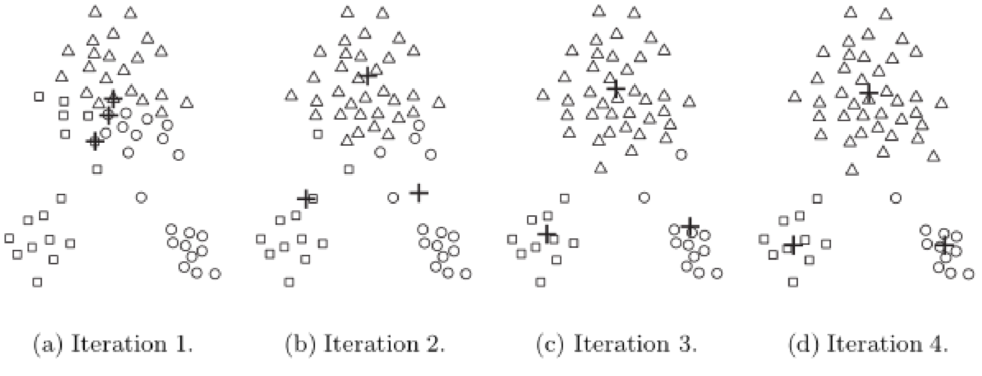
(d) Six clusters.

Fig. 12 - Example of clusters (b,c,and d) and partitional clustering (b-d). Tan et al. (2018).

3.2.3 Main Clustering Methods

We distinguish various types of clustering (or an entire collection of clusters): hierarchical (nested) versus partitional (unnested), exclusive versus overlapping versus fuzzy, and complete versus partial.

Hierarchical versus Partitional. The most commonly discussed distinction among different types of clustering is whether the set of clusters is nested or unnested, or in more traditional terminology, hierarchical or partitional. A partitional (unnested) clustering is simply a division of the set of data objects into non-overlapping subsets (clusters on which each data object is in exactly one subset) consult fig. 13.


(a) Iteration 1.

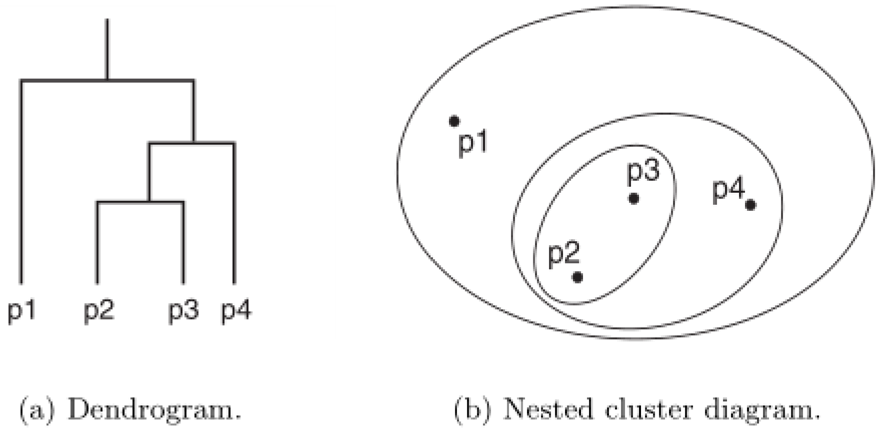
(b) Iteration 2.

(c) Iteration 3.

(d) Iteration 4.

Fig. 13 - K-means (exclusive) cluster type example. On the plot three centers are found after 4 iterations. Tan et al. (2018).

On the opposite hand when clusters can have subclusters is a hierarchical (nested) clustering (see fig. 14), which is a set of nested clusters that are organized as a tree. Each node (cluster) in the tree (except for the leaf nodes) is the union of its children (subclusters), and the root of the tree is the cluster containing all the objects.



(a) Dendrogram.

(b) Nested cluster diagram.

Fig. 14 - A hierarchical (overlapping) clustering of four points shown as a dendrogram and as nested clusters. Tan et al. (2018).

Exclusive versus Overlapping versus Fuzzy. If each object belongs to a single cluster we talk of an exclusive approach. In an overlapping or non-exclusive clustering, it's possible to reflect the fact that an object can simultaneously belong to more than one group (class). Fuzzy clustering, treats every object as part of every cluster with weights between 0 (strongly doesn't belong) and 1 (strongly belongs). As additional constraint the total sum of the weights for each object must be 1. Similarly, probabilistic clustering techniques compute the probability with which each point belongs to each cluster, and these probabilities must also sum to 1, avoiding the arbitrariness of assigning an object to only one cluster when it is close to several. In practice, a fuzzy or probabilistic clustering is often converted to an exclusive clustering by assigning each object to the cluster in which its weight or probability is highest.

Complete versus Partial. Many times, objects in the data set represent noise, outliers, or "uninteresting background" that can be convenient to avoid, so we refer to as partial approach. A complete clustering assigns every object to a cluster (consult fig. 15).

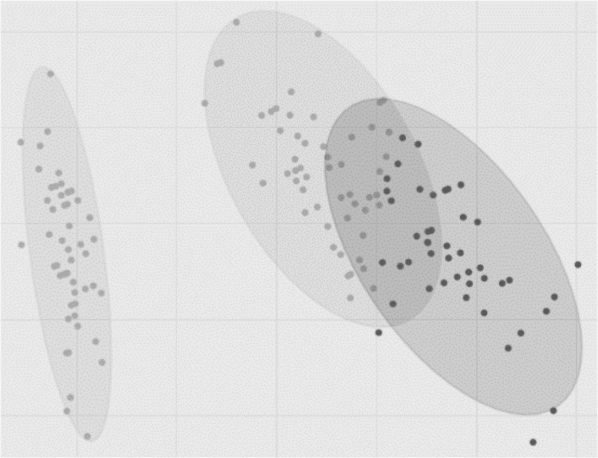


Fig. 15 – Fuzzy-partial clustering example, outliers can be observed outside the envelopes. Author.

3.2.4 Types of Clusters

Prototype-Based. A cluster in which each object is closer (more similar) to the prototype that defines it than to the prototype of any other cluster. For data with continuous attributes, the prototype of a cluster is often a centroid, e.g., the average (mean) of all the points inside the cluster. When, for example, the data has categorical attributes, the prototype is usually a medoid, e.g., the most representative point of a cluster. Very often, the prototype can be seen as the most central point, and in those cases, we commonly refer to prototype-based clusters as center-based clusters.

Graph-Based. then a cluster can be defined as a connected component in the case on which the data is represented as a graph, where the nodes are objects and the links represent connections among them.

Density-Based. A cluster consist of a dense region of objects that is surrounded by a region of low density, often employed when the clusters are sporadic or entwined, and with presence of noise and outliers.

Shared-Property (Conceptual Clusters) More generally we can also define a cluster as a set of objects that share some properties (physical or mathematical), such as temperature, density of measure points over maps for instance, or average net income, age of groups of individuals for example when dealing with social related data.

For our purposes two approaches will be detailed:

•K-means: prototype-based, partitional clustering technique that using centroids attempts to find a user-specified number of clusters (K).
• Agglomerative Hierarchical Clustering. “This approach refers to a collection of closely related clustering techniques that produce a hierarchical clustering by starting with each point as a singleton cluster and then repeatedly merging the two closest clusters until a single, all-encompassing cluster remains” (Tan et al. 2018).

3.3 K-means Clustering

With K-medoid constitutes the most important prototype-based clustering techniques that creates a one-level partitioning of the data objects. The method defines a prototype in terms of a centroid, in general the mean of a group of points, often applied to objects in a continuous n-dimensional space. K-medoid defines a prototype in terms of a medoid, on which we look for the most representative point for a group of points. A medoid by definition is an actual data point, while a centroid almost never corresponds to one present on the set.

3.3.1 The Basic K-means Algorithm

The K-means clustering technique is simple. K initial centroids are chosen, where K is a user-specified parameter, namely, the number of clusters desired. Each point is then assigned to the closest centroid, and each collection of points assigned to a centroid is a cluster. The centroid of each cluster is then updated based on the points assigned to the cluster. The assignment and update steps keep going until no point changes clusters, or in other words, until the centroids remain the same.

Basic K-means steps:
1: Select K points as initial centroids.
2: repeat
3: Form K clusters by assigning each point to its closest centroid.
4: Recompute the centroid of each cluster.
5: until Centroids do not change.

For several combinations of proximity functions and types of centroids, K-means always converges to a solution; (the process reaches a state in which no points are shifting from one cluster to another, and therefore, the centroids don't change). Since most of the convergence occurs in the early steps, however, the condition on line 5 of the algorithm is often replaced by a weaker condition, e.g., repeat until only 1% of the points change clusters.

The "kmeans" R function is used for this study: The data given by x are clustered by the k-means method, which aims to partition the points into k groups such that the sum of squares distance (?) from points to the assigned cluster centers is minimized. At the minimum, all cluster centers are at the mean of their Voronoi sets (the set of data points which are nearest to the cluster center). The algorithm of Hartigan and Wong (1979) is used by default.

The algorithm requires as input a matrix of M points in N dimensions and a matrix of K initial cluster centers in N dimensions. The number of points in cluster L is denoted by NC(L). D(I,L) is the Euclidean distance between point land cluster L. The general procedure is to search for a K-partition with locally optimal within-cluster sum of squares by moving points from one cluster to another. It seeks "local" optimal solutions such that no movement of a point from one cluster to another will reduce the within-cluster sum of squares. (Hartigan and Wong, 1979)

3.3.2 Strengths and Weaknesses

K-means is simple and can be used for a wide variety of data types. It is also very efficient, even though multiple runs are needed. However, is not suitable for all types of data like clusters of different sizes and densities, although it can typically find pure subclusters if a large enough number of clusters is specified. K-means also finds difficulties clustering data that contains outliers, case on which its detection and removal can improve results. In substance: K-means is restricted to data for which a center (centroid) is compatible with the type of analysis, K-medoid clustering instead, allows more flexibility but at greater computational cost.

3.4 Agglomerative Hierarchical Clustering

Two basic approaches exist for generating a hierarchical clustering:

Agglomerative: Start with the points as individual clusters and, at each step, merge the closest pair of clusters. This requires defining a notion of cluster proximity.
Divisive: Start with one, all-inclusive cluster and, at each step, split a cluster until only singleton clusters of individual points remain. In this case, we need to decide which cluster to split at each step and how to do the splitting.

A hierarchical clustering is often displayed graphically using a tree-like diagram called a dendrogram, which displays both the cluster-subcluster relationships and the order in which the clusters were merged (agglomerative view) or split (divisive view). For sets of two-dimensional points, a hierarchical clustering can also be graphically represented using a nested cluster diagram (see part (a) on fig. 14).

3.4.1 Basic Agglomerative Hierarchical Clustering Algorithm

The Basic agglomerative hierarchical clustering algorithm is represented by the following steps:

1: Compute the proximity matrix, if necessary.
2: repeat
3: Merge the closest two clusters
4: Update the proximity matrix to reflect the proximity between the new cluster and the original clusters.
5: until only one cluster remains.

The “hclust()” R function is used for this study along with the Ward’s method, which assumes that a cluster is represented by its centroid and measures the proximity between two clusters in terms of the increase in the SSE (sum of squared estimate of errors) that results from merging the two clusters. Ward’s method attempts to minimize the sum of the squared distances of points from their cluster centroids that are obtained from a Euclidean distance matrix (“dist()” function in R).

$$SSE = \sum_{i}^{K} = 1 \sum x\epsilon C_i dist^2(m_i, x) \tag{12}$$

On equation 1 x is a data point of a cluster C_i and m_i is its centroid. For each point, the error is the distance to the nearest centroid. To get SSE the errors are squared and added.

3.4.2 Defining Proximity between Clusters

The key operation of the hierarchical algorithm is the computation of the proximity between clusters, and it is the definition of cluster proximity that differentiates the various agglomerative hierarchical techniques. Cluster proximity is typically defined with a particular type of cluster in mind—. For instance, many agglomerative hierarchical clustering techniques, such as MIN, MAX, and Group Average, come from a graph-based view of clusters. MIN defines cluster proximity as the proximity between the closest two points that are in different clusters, or using graph terms, the shortest edge between two nodes in different subsets of nodes. On the opposite hand, MAX takes the proximity between the farthest two points in different clusters to be the cluster proximity, or using graph terms, the longest edge between two nodes in different subsets of nodes (check fig. 16). Another graph-based approach, the group average technique, defines cluster proximity to be the average pairwise proximities (average length of edges) of all pairs of points from different clusters.

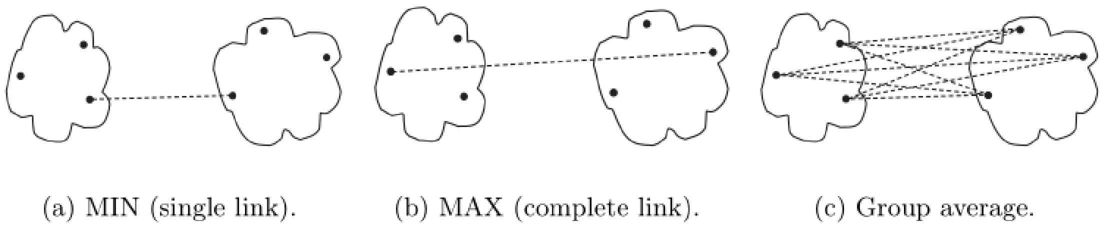


Fig. 16 - Graph-based definitions of cluster proximity. Tan et al. (2018).

An alternative technique, Ward's method (the one considered for this thesis), also assumes that a cluster is represented by its centroid, but it measures the proximity between two clusters in terms of the increase in the SSE that results from merging the two clusters. Like K-means, Ward's method objective is to minimize the sum of the squared distances of points from their cluster centroids. Ward's Method and Centroid Methods. For Ward's method, the proximity between two clusters is defined as the increase in the squared error that results when two clusters are merged. Thus, this method uses the same objective function as K-means clustering. Ward's method is often used as a robust method of initializing a K-means clustering, indicating that a local "minimize squared error" objective function does have a connection to a global "minimize squared error" objective function.

3.4.3 Strengths and Weaknesses

In general, hierarchical algorithms are typically used because the underlying application requires a hierarchy. However, they are expensive in terms of their computational and storage requirements. The notion of merging all data can also cause trouble for noisy, high-dimensional data, such as document data, problems that can be addressed to some degree by first partially clustering the data using another partitioning technique, such as K-means.

3.5 Clustering applications for the oil & gas sector

Olneva et al. (2018) clustered 1D, 2D, and 3D geological maps for West Siberian Petroleum Basin with seismic data (emphasizing the Achimov play with 207 fields discovered), developing two compatible approaches for analysis.

One approach, defined as "from general to particulars" (i.e., top-down), incorporated drilling data coming from 5000 existing wells with regional geology structures and tectonic elements, and available paleographic charts. The second one, defined as "from particulars to generals" (i.e. bottom-up), started from regional 2D seismic and geological data, aiming to create a training sample based on seismic geological patterns, counting on a regional database including more than 40.000 km2 of 3D data.

The top-down approach, involving the clustering algorithm K-means, allowed the update and the improvement of the regional clustering model of the play, in agreement with the adopted mathematical model and the local geology. The bottom-up approach, based on pattern recognitions, revealed its usefulness on generating a library of typical seismic images for specific events in the Achimov sequence, mostly associated with turbidites.

Cremashi, Shin and Subramani (2015) introduced a systematic approach to reduce this discrepancy using data clustering, model selection, and cluster identification techniques. Using

772 experimental data points (publicly available) and generating a database containing seven independent variables.

They estimated impact of each independent variable on threshold velocity. For each velocity type, these impacts were quantified using correlation coefficient between independent variables and the threshold velocity. The approach considered was kmean for data clustering, generating thirty-six different cluster sets; allowing them to measure and reduce the model prediction differences in threshold velocity estimations for production oil/gas pipelines design/operation. The performance tests revealed a significant reduction (several orders of magnitude). Results showed that the average of the error percentages between the predictions and experimental velocities are reduced up to 37%.

Boesen, Haber and Hoverten (2021) introduced a new graph-Laplacian based semi-supervised clustering method. Very large datasets using a limited amounts of labelled data points where clustered and handled with relatively low computational power, using MATLAB as computational tool. Based on amplitude-versus-angle inversion parameters and borehole information (3D oil prospection) the clustering was performed. A light-running clustering algorithm was developed, ignoring petrophysical relationships and taking a data-driven approach, similar to the work of Yu et al. (2008) who used a fusion of genetic algorithms, simulated annealing and neural networks to predict oil reservoir properties.

A Life of Field dataset was clustered, containing a fault-block constrained central oil reservoir Another field dataset was clustered, characterized by a stratigraphic trapped channeling system. According to author's finding, reasonable results were obtained with simple assumptions and small amounts of computational resources. Also portrayed several ways to increase the accuracy and expand the applicability possibilities.

Yiping et al. (2021) Analyzed a weighted multi-view collaborative fuzzy C-means clustering algorithm improved using the double-layer-nested particle swarm optimization. Initially performances and risks, evaluation system of overseas oil and gas exploration projects with 20 indexes are. Then a membership matrix is obtained applying the clustering algorithm to classify exploration projects. Using a classification sensitivity coefficient, each index contribution is determined. Finally, the comprehensive score of each project is obtained, by the weighted sum of the score values of each index and then carried out the ranking and the investment priority of all projects.

The obtained results show that 24 exploration projects can be clustered into 4 classes, being 14 first-class exploration projects for preferred investment, 6 s-class ones maintained steadily, 3 third-class ones are retained, and 1 fourth-class one is excluded. In the case of the same class, investment priority is defined by the projects with a high comprehensive ranking. According to the authors, the proposed method can improve decision-making process for exploration investments.

Bhaskaran, Chennippan and Subramaniam (2020) developed a forecast model to predict future fault occurrences in order to sort out operational problem in the oil pipelines. The authors propose fault identification and prediction methods based on K-means clustering and time-series forecasting along with linear regression algorithm using multiple pressure data. Using a scaled-down experimental hardware lab setup resembling characteristics of onshore unburied pipeline in India a real-time validation of the approach is validated.

In the proposed work, crack and blockages are identified by taking pressure rise and pressure drop inferred from two cluster assignment. The obtained numerical results from K-means clustering get as outcome the maximum accumulated range of multiple pressures. Inspection process can benefit from it being able to estimate the normal and abnormal performance of oil transportation in a simple yet robust way by using this final cluster center data. When confronted to the historical pressure datasets The developed forecast model successfully predicts future fault occurrences rate with a validity of 91.9%. The models are intended to easy pipeline operator's jobs, reducing the need for complex computation processing to assess and predict the condition of deployed oil pipelines and prioritizing the planning of their inspection and rehabilitation programs.

4 THE EMILIA-ROMAGNA AREA

4.1 Geological setting

The Alps' and Northern Apennines' foredeep are represented by the Po Plain (Fig. 17). Active shortening, which is accommodated by folding and blind thrust faulting, characterizes the area (Pieri and Groppi1, 981; Oriand Friend, 1984; Massolietal, 2006; Toscani et al., 2009).

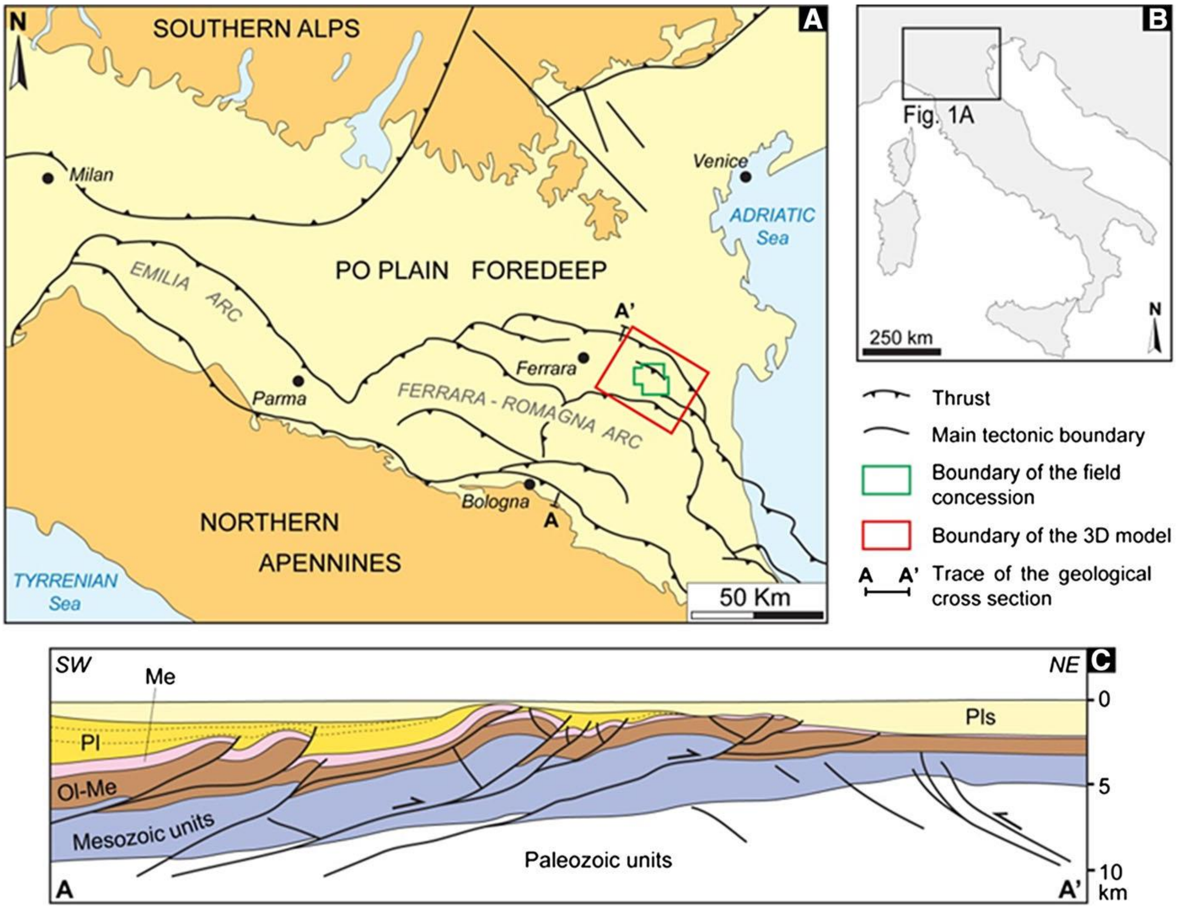


Fig. 17 - (A) Simplified structural map of the eastern Po Plain and surrounding regions (modified from Pieri and Groppi 1981; Toscani et al. 2009). (B) Location of Fig. 1A. (C) Geological cross section across the eastern Po Plain (modified from Toscani et al. 2009). Pls Pleistocene, Pl Pliocene, Me late Messinian, Ol-Me Oligocene-Early Messinian.

The current geological configuration is the result of the Northern Apennines' Oligocene to Neogene evolution which resulted in the development of asymmetric foredeeps with an outward migrating behavior (Ghielmi et al., 2010). Late Oligocene–Quaternary clastic sediments filled these gaps, which were gradually added (Ghielmi et al., 2010).

Compressional tectonics changed the passive margin behavior to foredeep basin during the Mesozoic to Quaternary period. Both the foredeep succession and its pre-Pliocene substratum were severe influenced by Pliocene tectonic deformation.

Transitional sub-marine fan-delta conglomerate (Boreca Conglomerate), Messinian evaporitic deposits (Gessoso-Solfifera Formation), Mesozoic shelf carbonate, (Gallare group); comprise the pre-Pliocene succession from shallow to deep (Agip 1982; Mancin et al. 2009).

The filling of the Po Plain foredeep basin by clastic sediments is documented on the Plio-Pleistocene sequence, consisting of Pliocene–Pleistocene transgressive marine clays (Santerno Formation) as well as turbiditic sequences of early-middle Pliocene age (Porto Corsini Formation) and middle-late Pliocene age (Porto Garibaldi Formation).

Progradation of the Po Plain has resulted in the creation of slope, shelfal, coastal, and deltaic clastic deposits since the middle Pleistocene (Carola and Ravenna Formations, Pleistocene), eventually leading filling at its maximum the foreland basin (Ghielmi et al., 2010).

Two overlapping pools hydraulically linked constitute the reservoir under consideration, which reaches an average depth of 1200 m ssl). Trap's trending is NW–SE, being an asymmetric anticline connected to a NW–SE striking a regional thrust by the NE. Sands and silty sands alternate in the formation.

The present aquifer that's connected to the reservoir goes NW and SE and sealing faults contain in the NE and SW directions. The Santerno Formation (clay formation) with an average thickness of 80 m ensures the reservoir's hydraulic seal.

4.2 Withdrawal activities on the region

4.2.1 UGS activity

Numerous hydrocarbon fields were discovered in the Emilia-Romagna area beginning in the early 1950s, several of which were later converted into and operated as underground gas storage systems (UGS). The italian DGS-UNMIG (*Direzione Generale per la Sicurezza - Ufficio Nazionale Minerario per gli Idrocarburi e le Georisorse*) provides abundant related information (fig. 18).

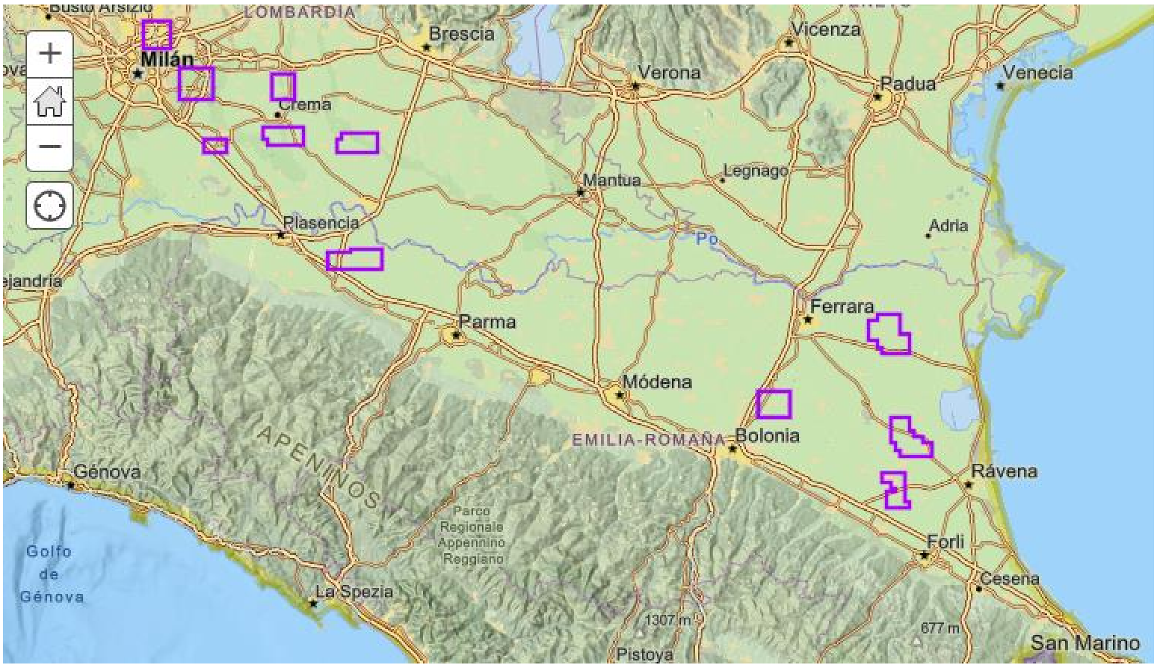


Fig. 18 - 2021 configuration of UGS camps for the Emilia-Romagna region. DGS-UNMIG.

4.2.2 Oil and gas extraction

Noteworthy also the extensive activity of fluid withdrawal referred to oil & gas concessions, as can be seen on figure 19.

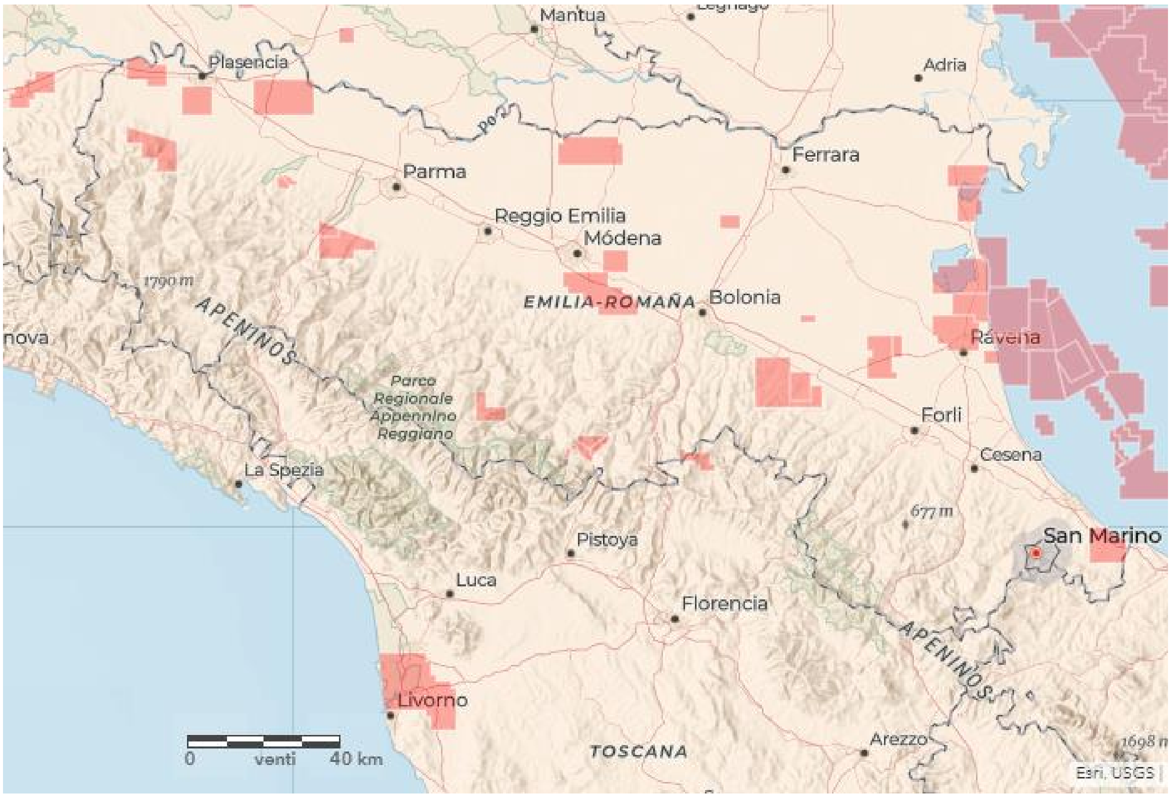


Fig. 19 - 2021 configuration of oil and gas camps (on red) for the Emilia-Romagna region. DGS-UNMIG.

4.2.3 Water extraction

Water withdrawal severely interest the area on which the authorities have supervised and imposed several restrictions and limits, in order to respect the recharging capacity of the aquitards and mitigate in this way its impact on the subsidence phenomenon, highly affected by this activity during the last 70 years (Carminati et al., 1999). The last plans for the water usage (*Piano Tutela delle Acque* 2005, *Piano di Gestione 2015* and *Progetto di Piano di Gestione 2021*) monitored intermittently the produced values for all the 205 "comuni" belonging to the Po Plain area as follows: for the year 2003 a consumption of 662Mm3 was registered, the period 2009-2011 experienced a small decrease reaching the 659Mm3/year; and the last numbers available show a significant diminishing with reduction up to 623Mm3/year of water extracted for the period 2015-2018.

As referred by the *Agenzia regionale per la prevenzione, l´ambiente e l´energia* from *Arpae Emilia-Romagna*, the assessment is composed by aggregation of assessments conducted separately for the different types of use of underground water (Civil-Aqueduct, Industrial, Irriguous and Zootechnical) which define the different estimation methodology. The civil uses of aqueducts and part of the industrials (about 50%) correspond to measured or registered data while the remaining part of the industrial withdrawal, irrigation and other uses refer instead to estimation methods. The quantities supplied are considered reliable at the provincial scale, while at the municipal level deviations of up to 50% are plausible, which may be even higher in the case of small volumes.

4.3 Technical-Historical synthetic description of the area

4.3.1 Major anthropic events related to fluid withdrawn in Emilia-Romagna Region

1951	Water rate extraction 200 x 10^6 m3/year (Carminati and Martinelli, 2002).
1950-56	Gas-bearing extraction operation setting (Carminati and Di Donato,1999)
1977	Water rate extraction 740 x 10^6 m3/year (Carminati and Martinelli, 2002).
1970-80	New public aqueducts using surface water reduced the aquifer overdraft (Teatini et al., 2006).
1992	Water rate extraction 710 x 10^6 m3/year (Teatini et al., 2006).
1999	Water rate extraction 703 x 10^6 m3/year (Carminati and Martinelli, 2002).

Studies in the southern part of the Po Plain (Emilia-Romagna Region) show that total recharge of local groundwater is 710 x 10^6 m3/year (Carminati and Martinelli, 2002).

4.3.2 Major monitoring related events

1930's	First evidence of the phenomenon date back to the associated with the Po Delta and linked to water pumping and land drainage projects (Gambolati et al., 2006).
1940	Earliest scarce and scattered data of average piezometry in the coastal aquifer (Teatini et al., 2006).
1950's	Spirit leveling was started in the eastern Po River plain by the Italian Geographic Military Institute (IGM), line (IGM16) (Teatini et al., 2006).
1976	Piezometric monitoring Network setting for the E-R region, managed by ARPA Emilia-Romagna (about 600 150-average-depth artesian wells distributed over the entire region) (Baldi, P et al., 2006).
1970's	The Geological Service of the Ravenna municipality instituted and surveyed every 4 to 6 years over a municipal benchmark network (Gambolati et al., 2006).
1996	Leveling line running along the Adriatic coastline established by the ARPA Emilia-Romagna (Teatini et al., 2006).
Late 1990's	ENI-E.&P. installed a number of borehole extensometers to measure the shallow aquifer system compaction along the coastland (Teatini et al., 2006).
1997-96	Setting of the Regional Monitoring Network managed by ARPA E-R. 60 GPS stations plus 2300 high precision geometric leveling markers (Teatini et al., 2006).

4.4 Dataset of ground movement data

The ARPAE 2016 and 2017 publications, related to the retrieval of subsidence measures for the Emilia-Romagna plain, provide an excellent base for the understanding of the available data acquisition and elaboration process and would be the base for the development of this section. In order to better portrait and to be as true as possible with their work which follows is a respectfully edited translation of the data elaboration processing conducted during the acquisition campaign or vertical ground movement from the Padana plain region.

The studied area comprehends mostly the plain territory belonging to the Emilia-Romagna region, with a total extension of about 13.000 km2 as can be seen on fig. 20.

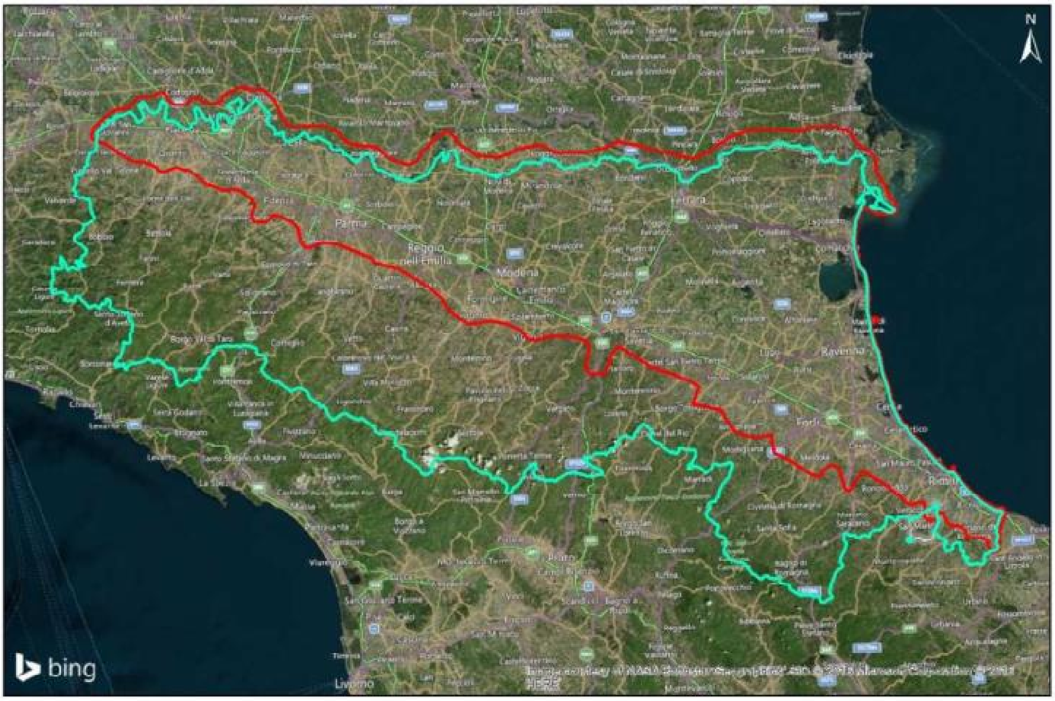


Fig. 20 - Extension of the gross studied area. ARPAE.

4.4.1 Major survey-related events (last 20 years) according to the ARPAE Emilia-Romagna

1999	• ARPA Survey for the entire network (leveling markers plus GPS stations).
2002	• ARPA Checking survey for only GPS stations.
2005-07	• First PSInSAR regional survey (about 11.000 km2) supported by nearly 50% of high precision geometric leveling markers.
2011-12	• Updating of PSInSAR regional measurements (about 11.000 km2) supported by 17 permanent GPS stations.
2016	• Last updating of PSInSAR regional measurements (about 11.000 km2) accounting for MS and PS points, supported by all available permanent GPS stations.

4.5 Satellite data acquisition and definition on the elaboration sites

4.5.1 Satellite data acquired

In order to ensure the continuity of the monitoring program (RADARSAT-1 and RADARSAT-2 satellites measures) the SqueeSAR algorithm, called stitching, allowed both data to be elaborate on the same data set.

This joint process requires the same acquisition geometry from both measure tracks (ascending or descending). On the opposite hand, it is vital that both radar objectives are seen by both satellites.

On figure 21 it is showed the standard scheme of a satellite acquisition phase on ascending modality (from south to north): (a) satellite sight line perpendicular from the ground forms an angle theta, (b) satellite orbit isn't perfectly oriented on north-south sense but it differs a sigma angle, (c) tridimensional view.

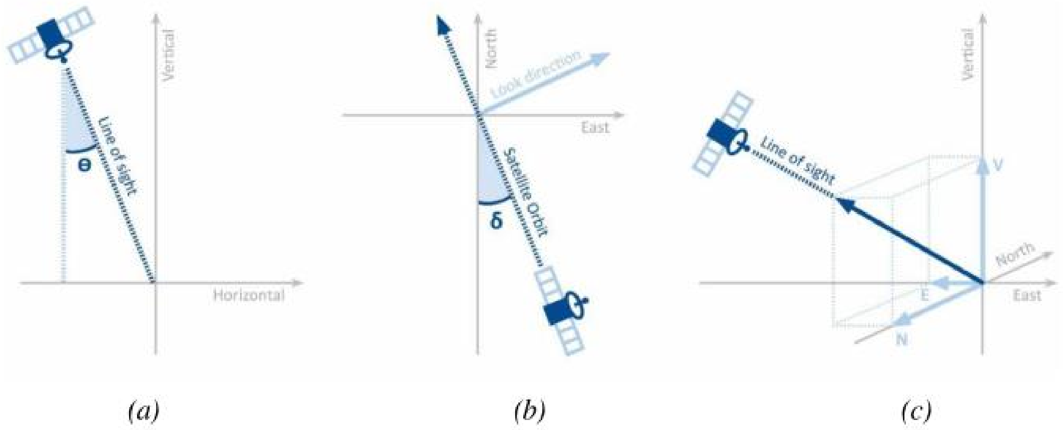


Fig. 21 - Standard scheme for ascending geometry acquisition. ARPAE.

For a limited area of the Parma area, the COSMO-SkyMed (CSK) from the Italian space agency cover the acquisitions, due to lack of coverage from RSAT2. On figure 22 is visualized the coverage available for the RSAT2 and CSK satellites regarding to the studied area.

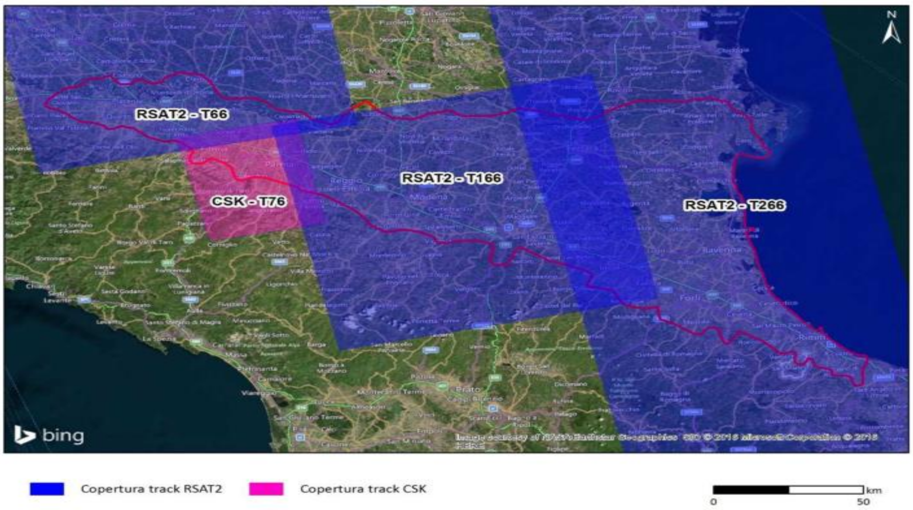


Fig. 22 - Coverage of RSAT and CSK available to the studied area. ARPAE.

4.5.2 Definition and preparation of the elaboration sites

Figure number 23 reports the extension of the elaboration sited based on satellite coverage and table 1 a summary related to the dataset.

Due to 2012 Mirandola seismic activity, it was necessary to singularly elaborate the site. In fact, the images acquired in correspondence with the time of the episodes (may 2012) were excessively affected by high values of movement for being useful at a technical level. It is for this reason that the period time considered for the Mirandola site was one year lower (see table 1).

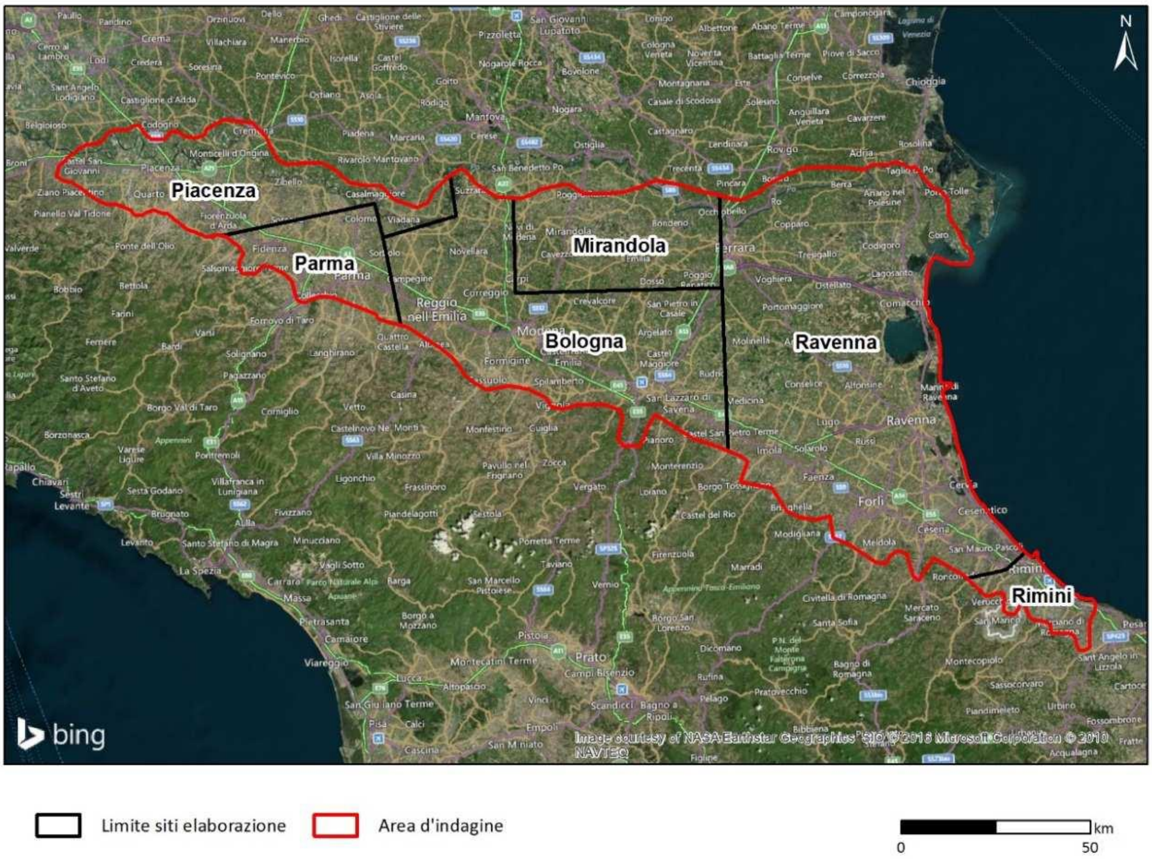

N
Piacenza
Parma
Mirandola
Bologna
Ravenna
Rimini
bing
Image courtesy of NASA Earthstar Geographics SIO © 2018 Microsoft Corporation © 2018 NAVTEQ
Limite siti elaborazione
Area d'indagine
km
0
50

Fig. 23 - Elaborated sites distribution over the Emilia-Romagna region. ARPAE.

Site	Satellites	Θ	Number of Images	Coverage period
Piacenza	RSAT1 – RSAT2	34,1°	73	24/05/2011 – 09/05/2016
Parma	CSK	34°	66	03/05/2011 – 14/04/2016
Bologna	RSAT1 – RSAT2	34°	69	07/05/2011 – 16/05/2016
Mirandola	RSAT1 – RSAT2	34,7°	55	06/06/2012 – 16/05/2016
Ravenna	RSAT1 – RSAT2	33°	75	14/05/2011 – 23/05/2016
Rimini	RSAT1 – RSAT2	34,7°	70	02/05/2011 – 23/05/2011

Table 1 - Characteristic of the elaborated sites (acquisitions). ARPAE.

4.6 Interferometric analysis

On 2011-12 Arpa (now Arpae) commissioned by the authorities of the Emilia Romagna region and in collaboration with the Dicam (*Dipartimento di Ingegneria Civile, Chimica, Ambientale e dei Materiali*) of the Bologna university, performed the acquisition campaign of the vertical ground movements from de padana plain for de Emilia-Romagna 1 region, using the SqueeSAR technique (PSInSAR of second generation) as interferometric analysis technique of satellite radar data. This work (2016) updated the knowledge existing for the period 20011-16, and continue to use the SqueeSAR technique but refined.

4.6.1 SqueeSAR analysis results for the elaboration sites

The performed analysis was able to generate an output of nearly 2 million measure points (MP) over the all interested area. It is worth to mention that the point distribution responds to the usage of the land. Maximum densities are observed on habited areas and where anthropic infrastructure is present, whereas on cultivated or vegetated areas the density is lower.

Worth of mentioning is the fact that SqueeSAR data are bond to two precision indexes: standard deviation and temporary coherence.

Standard deviation (V_STDEV), on the SqueeSAR analysis, is referred to the mean velocity of the measure points with regard to the reference point. It will depend, in general, of the physical distance among them (MP-REF) radiometric quality of the MP, number of elaborated images, the elapsed time for the acquisition and the temporal continuity of the campaign.

Temporal coherence (COHERENCE) as index reflects how the movement of the points follows a certain analytical model, which is to say a certain mathematical function. The SqueeSAR data series is usually confronted with polynomial and sinusoidal equations. Each temporal series for movement is confronted with the selected model, and gets as result the temporal coherence index, which runs between 0 (no coherence and not correspondence with the model at all) and 1 (maximum coherence/correspondence and perfect fit).

Table 2 synthetize the mean values of standard deviation and coherence obtained for the different elaborated sites. It is worth to mention that standard deviation values don’t go upper than 1mm/year, indicating high precision measures.

Site	V_STDEV [mm/year]	COHERENCE
Piacenza	0,33	0,83
Parma	0,36	0,76
Bologna	0,31	0,85
Mirandola	0,39	0,90
Ravenna	0,29	0,83
Rimini	0,27	0,88

Table 2 - Average velocity standard deviation and temporal coherence. ARPAE.

4.7 Ravenna Elaboration site

Over the Ravenna area 325.000 MP where identified, with an average density of 56 MP/Km2. Point distribution, presented by annual average velocity and its standard deviation is shown on figure 24.

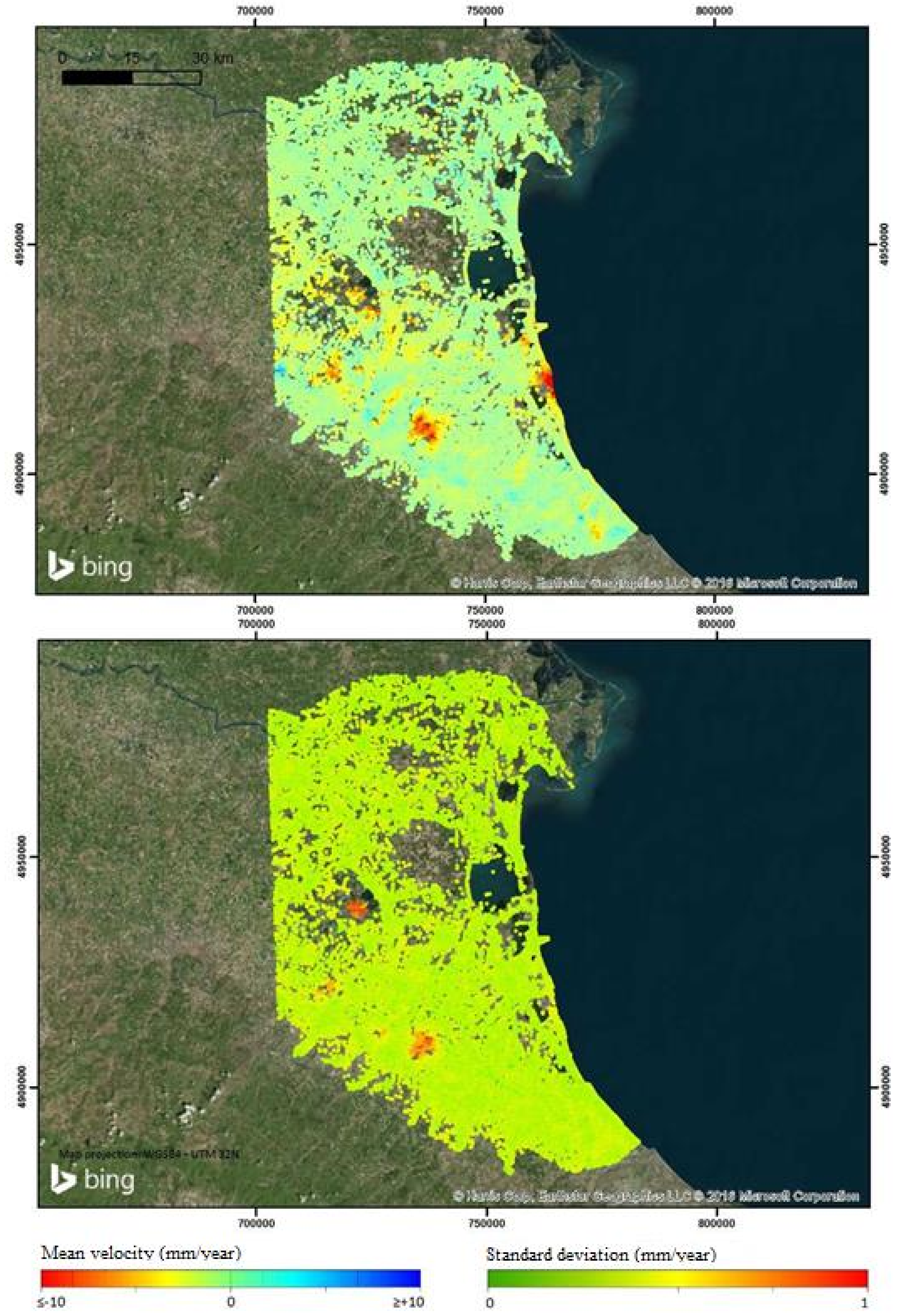

700000
750000
800000
4950000
4900000
0
15
30 km
bing
© Harris Corp, Earthstar Geographics LLC © 2016 Microsoft Corporation
Map projection: WGS84 - UTM 32N
Mean velocity (mm/year)
≤-10
0
≥+10
Standard deviation (mm/year)
0
1

Fig. 24 - Average annual velocity (on top) and correspondent standard deviation (on bottom) of MP for the Ravenna site after processing. ARPAE.

4.8 Analysis of vertical ground movement

The cartographied territory is presented below (figs. 25 and 26) using an isokinetic curve representation comprehending the entire plain region of the Emilia-Romagna zone, a small section from the north border with the Lombardia region and the east coastal line and the 100 m above sea level isoline, totalizing about 11.300km2. The movement analysis took into account the isokinetic lines cartography as well as the single PS/DS. In particular the last ones, has been used for highlighting the maximum velocity point that usually passes unobserved using the isolines. With respect to the previous cartographic analysis, this time 1.912.781 where available against 315.371, special mention it is worth to make that 1.285.490 points where only referred to the Parama elaborating zone, through the usage of COSMO-SkyMed high resolution imagery, due to the lack of coverage from the historical acquisition of the RADARSAT-2.

In order to facilitate the comparison at a cartographic level between both of the campaigns, similar maps where developed and are publicly available:

1.Velocity chart of vertical ground movement for the 2011-2016 period. Scale 1:250.000
2.Velocity variation chart of vertical ground movement from the 2006-2011 period to the 2011-2016 period. Scale 1:250.000.
3.Velocity chart of vertical ground movement for the 2011-2016 period. Bologna Province. Scale 1:100.000.
4.Velocity chart of vertical ground movement for the 2011-2016 period. Coastal zone. Scale 1:100.000.

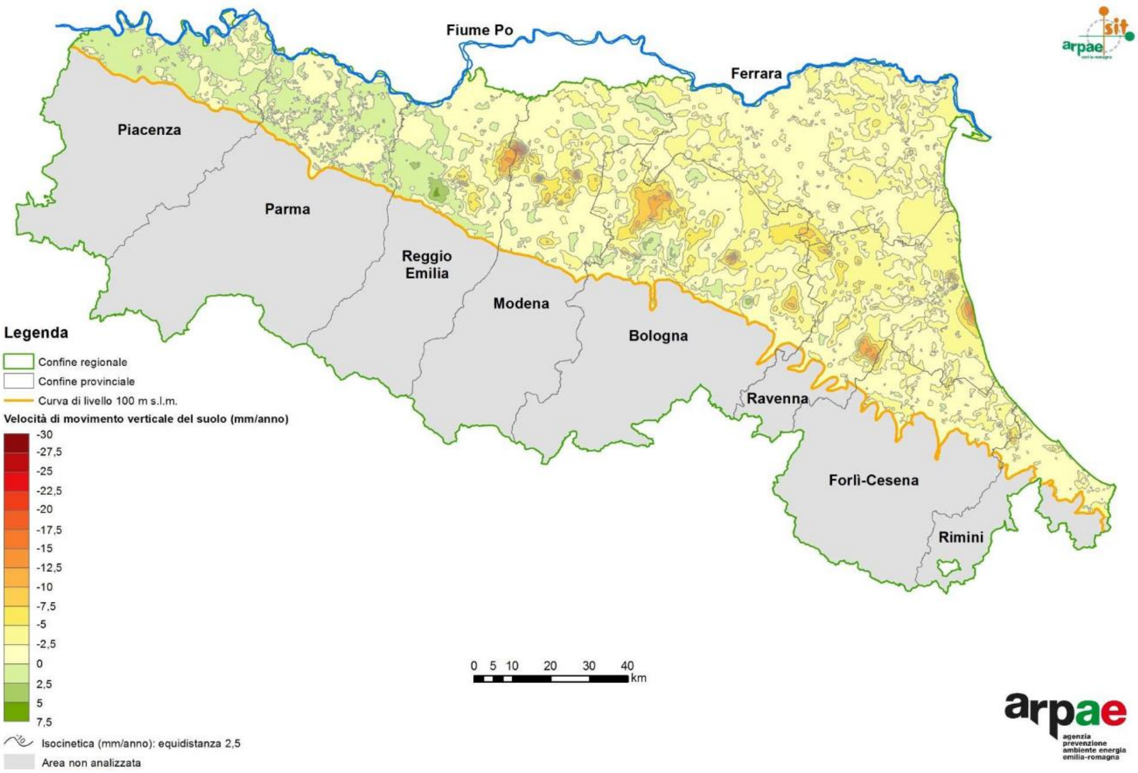


Fig. 25 – Original velocity chart for vertical ground movement velocity period 2011-2016. ARPAE.

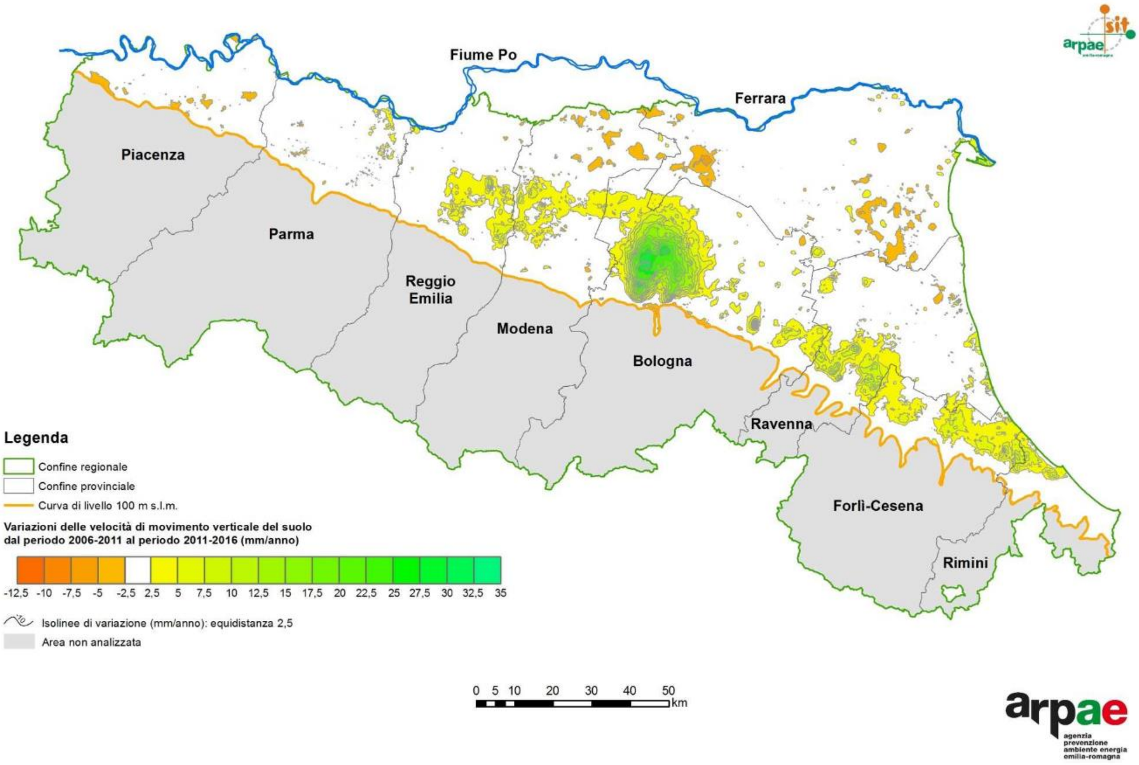


Fig. 26 – Original chart of vertical velocity variation for ground movement form periods analysis 2006-2011 an 2011-2016. ARPAE.

With respect to the precedent campaign (2006-2011), 79% of the territory doesn’t shows significative variations on trend, whereas a diminishing on the subsidence trend is notorious for 18% of the area.

On Table 3, foreach province, are reported the surface (km2 and %) in correspondence with the variation on velocity between the 2006-2011 and 2011-2016 periods grouped in three principal classes: the first one shows the surfaces on which the trend is negative, it means downward movement, the second are areas on which the variation oscillates between 0 +- 2.5 mm/year (mostly stable over time), and the third one is related areas on where there is a positive trend or uplifting phenomena.

Class of velocity variation (mm/year)	Surface (km2)																	
	PC	%	PR	%	RE	%	MO	%	BO	%	FE	%	RA	%	FC	%	RN	%
< -2,5	40	*5*	8	*1*	4	*1*	62	*5*	19	*1*	165	*6*	16	*1*				
-2,5 to 2,5	753	*95*	894	*96*	829	*81*	1044	*80*	1118	*60*	2404	*91*	1237	*81*	288	*49*	63	*20*
> 2,5	2		24	*3*	189	*18*	195	*15*	714	*39*	62	*3*	280	*18*	279	*51*	257	*80*
Total	795		926		1022		1301		1851		2631		1533		585		320	

Table 3 - Provincial surfaces subdivided by movement variation of velocity classes and relative percentage, in light blue the Ravenna area data (the minus represents a negative trend or a subsidence condition). ARPAE.

In tables 4 and 5 are expressed, for each province, the surfaces and relatives percentages, subdivide by relative movement velocity classes, respectively to the periods 2006-2011 and 2011-2016.

Surface (km2)

Class of velocity variation (mm/year)	PC	%	PR	%	RE	%	MO	%	BO	%	FE	%	RA	%	FC	%	RN	%
-35 to -30									22	*1*								
-30 to -25									55	*3*								
-25 to -20							1		84	*5*			11	*1*				
-20 to -15					5		5		121	*7*			25	*2*	4	*1*	1	
-15 to -10					27	*3*	36	*3*	11	*6*	11		53	*3*	20	*4*	7	*2*
-10 to -5					116	*11*	191	*15*	366	*20*	188	*8*	317	*21*	224	*38*	45	*14*
-5 to 0	46	*6*	354	*38*	567	*55*	949	*73*	1062	*57*	2150	*87*	1123	*73*	335	*57*	272	*84*
0 to 5	747	*94*	572	*62*	307	*30*	119	*9*	32	*2*	129	*5*	2					
5 to 10																		

Table 4 - Period 2006-2011: Provincial surfaces dived by movement class and relative percentage, in light blue the Ravenna area data. ARPAE.

Class of movement (mm/year)	Surface (km2)																	
	PC	%	PR	%	RE	%	MO	%	BO	%	FE	%	RA	%	FC	%	RN	%
-35 to -30																		
-30 to -25																		
-25 to -20							1											
-20 to -15							4		1				2					
-15 to -10					15	*1*	11	*1*	24	*1*			27	*2*	1			
-10 to -5					46	*4*	99	*8*	238	*13*	31	*1*	120	*8*	16	*3*	4	*1*
-5 to 0	252	*32*	476	*51*	657	*64*	1093	*84*	1479	*80*	2588	*98*	1379	*90*	564	*96*	316	*99*
0 to 5	543	*68*	451	*49*	302	*30*	93	*7*	111	*6*	11		5		6	*1*	1	
5 to 10					2													

Table 5 - Period 2011-2016: Provincial surface divided by movement type and relative percentage, in light blue the Ravenna area data. ARPAE.

In figures 27 and 28 for each province are shown, using histograms, the divided surfaces by their relative velocity class of ground movement, respectively to the periods 2006-2011 and 2011-2016.

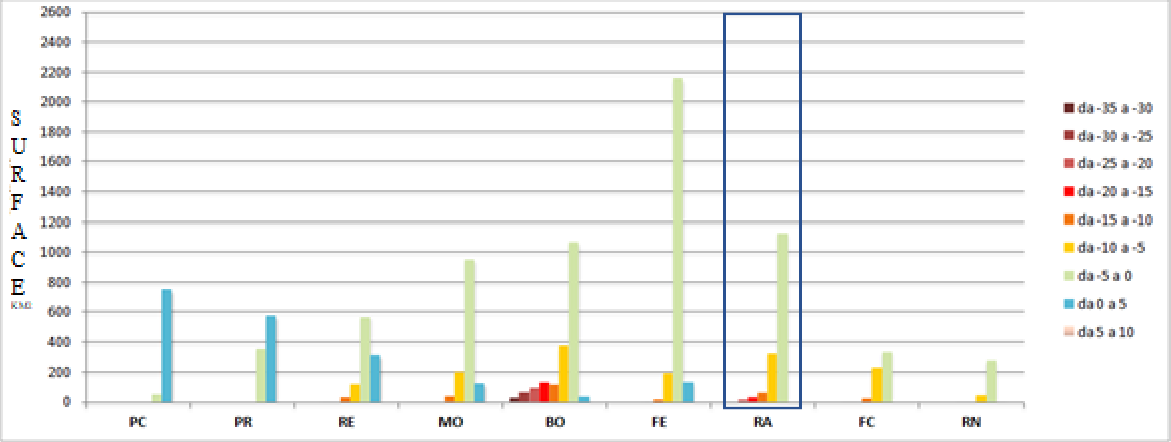


Fig. 27 - Period 2006-2001: Histograms for divided surfaces according to movement classes, inside the light blue rectangle the Ravenna area data (mm/year). ARPAE.

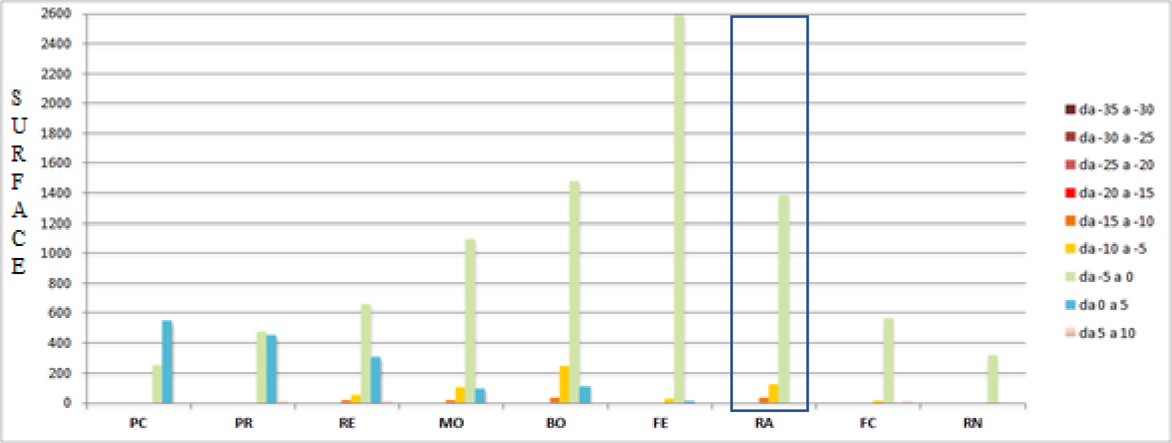


Fig. 28 - Period 2011-2016: Histograms for divided surfaces according to movement classes, inside the light blue rectangle the Ravenna area data (mm/year). ARPAE.

4.8.1 Ravenna province detailed analysis

On Ravenna province the general overview shows a reduction on the subsidence phenomenon with respect to the precedent analysis, the downward movements are of about 3 mm/year. Some historical subsidence areas as the Fiumi Uniti basin with maximums of 15 mm/year also in diminishing trend, and a vast area at Faenza's east in between rivers Lamone and Montone at highway level, with also 15 mm/year in average of downward values according to the Reda zone movements. It is worth of mention another small area regarding to the industrial zone at Conselice's north with maximums that can overpass the minus 15 mm per year. On the other hand, Ravenna city shows and stable downward trend within the 2-3 mm per year and compatible with natural subsidence trends (Carmina and Di Donato, 1999).

5 DATA ANALYSIS METHODOLOGY AND CASE STUDY

The work was performed at different stages. The first one comprehends the data acquisition on with research for the optimal procedures for dealing with the information were made in order to properly store, access and process it, extracting the maximum benefit. At a second stage proper mechanism to store data where defined. The third stage was the preprocessing of the accumulated information that leads to a final stage on which a proper case study was selected and conclusions stated.

Finally, the developed investigation methodology was applied and tested on a selected case study Diagram on fig 29 details the process.

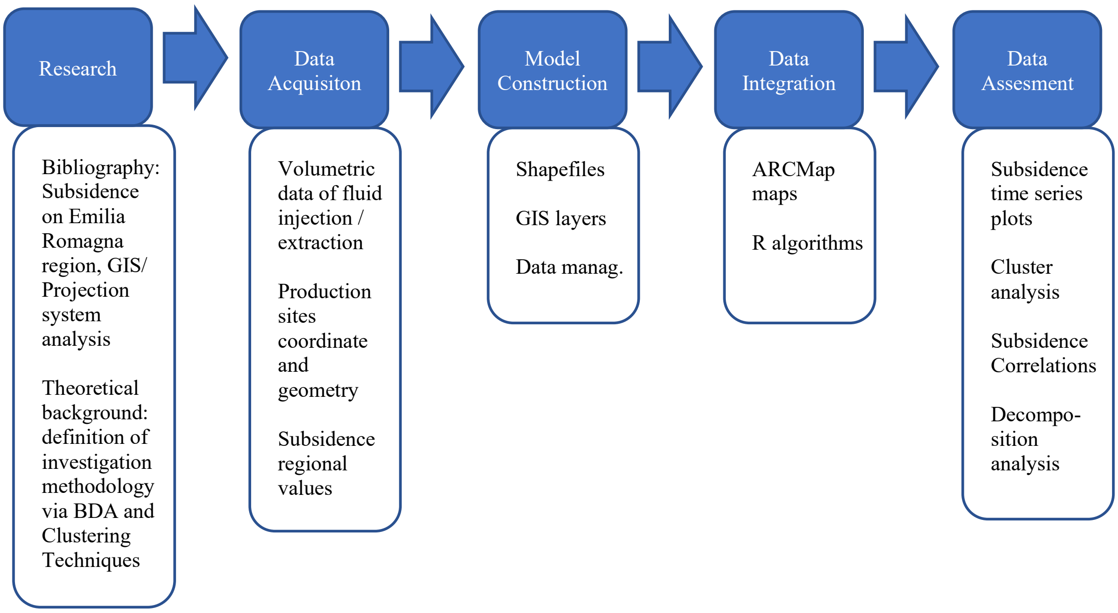

Research
Data Acquisiton
Model Construction
Data Integration
Data Assesment
Bibliography: Subsidence on Emilia Romagna region, GIS/ Projection system analysis
Theoretical background: definition of investigation methodology via BDA and Clustering Techniques
Volumetric data of fluid injection / extraction
Production sites coordinate and geometry
Subsidence regional values
Shapefiles
GIS layers
Data manag.
ARCMap maps
R algorithms
Subsidence time series plots
Cluster analysis
Subsidence Correlations
Decompo-sition analysis

Fig. 29 - Workflow. Author.

5.1 Databases for production-injection activities and ground movement data in the Emilia-Romagna region

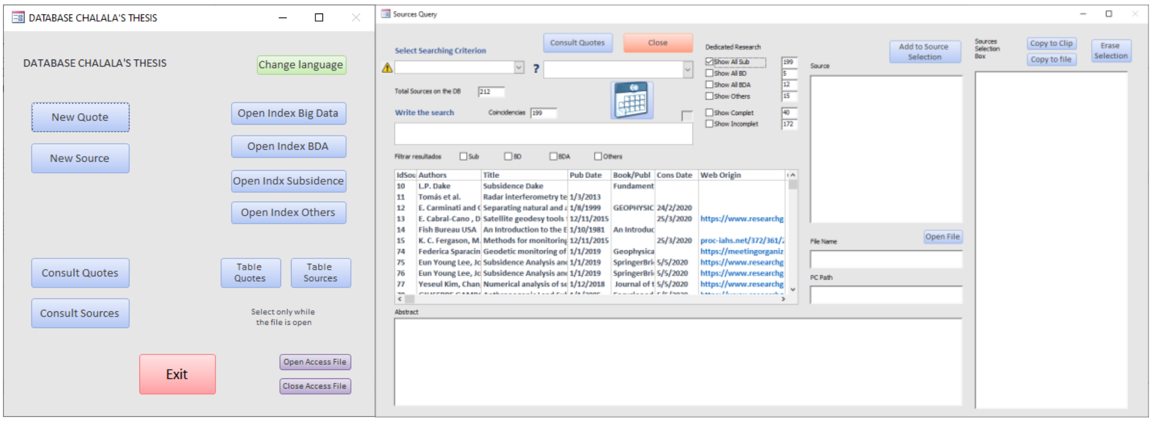
5.1.1 <u>Bibliographic research documentation management</u>

Several bibliographies were consulted in attention to the existing works on subsidence analysis for the area, as the same time volumetric data on data withdrawal and placement of the major withdrawal/injection activities (oil, gas and water). Data related to the subsidence of the Emilia - Romagna region was the starting point that leaded all the way along the develop of this work.

Since the research should be addressed over two fronts: useful data to work with and the suitable tools for processing it, an ad hoc access database fully portable and automated was developed from scratches. In the database all the information was properly classify according to its nature and field of interest:

- Subsidence (vertical ground movement anthropic caused mainly).
- Big Data (massive data volume analysis)
- Big Data Analytics (Clusterization and others).

The software allowed two entry modes: references and quotes, on which each quote would have a linked refence previously added. Several entries criteria were adopted: such as key words, author, citation, date or period of creation/modification/publication, among other well studied functionalities. Fig. 30 shows a glance about the developed tool.



Total Sources on the DB 212

Coincidencias 199

☑ Show All Sub 199
□ Show All BD 5
□ Show All BDA 12
□ Show Others 15
□ Show Complet 40
□ Show Incomplet 172

Filtrar resultados □ Sub □ BD □ BDA □ Others

IdSou	Authors	Title	Pub Date	Book/Publ	Cons Date	Web Origin
10	L.P. Dake	Subsidence Dake		Fundament		
11	Tomás et al.	Radar interferometry te	1/3/2013			
12	E. Carminati and (	Separating natural and a	1/8/1999	GEOPHYSIC	24/2/2020	
13	E. Cabral-Cano , D	Satellite geodesy tools f	12/11/2015		25/3/2020	https://www.researchg
14	Fish Bureau USA	An Introduction to the E	1/10/1981	An Introduc		
15	K. C. Fergason, M.	Methods for monitoring	12/11/2015		25/3/2020	proc-iahs.net/372/361/2
74	Federica Sparacin	Geodetic monitoring of	1/1/2019	Geophysica		https://meetingorganiz
75	Eun Young Lee, Jo	Subsidence Analysis and	1/1/2019	SpringerBri	5/5/2020	https://www.researchg
76	Eun Young Lee, Jo	Subsidence Analysis and	1/1/2019	SpringerBri	5/5/2020	https://www.researchg
77	Yeseul Kim, Chan	Numerical analysis of se	1/12/2018	Journal of t	5/5/2020	https://www.researchg

Fig. 30 - Two views of the database. At left the initial landing page, at right an example of a reference searching screen. Author.

5.1.2 Cartographic and volumetric data management

On the basis of the bibliography and theoretical background revision (DGS-UNMIG), the location of gas and oil field concession present on the area from the 1991 up to 2020 were individuated, and extracted from charts and the interactive DGS-UNMIG ArcGIS map. Successively the downloaded maps on pdf format were converted into images (png) ready to be inserted in a GIS environment on which the coordinates were extracted by contouring of each single polygon present on the map. According to the maps different projections were considered (Lambert_Conformal_Conic and Transverse_Mercator) with mainly GCS_Monte_Mario_Rome (EPSG: 4806), GCS_Monte_Mario (EPSG: 4265), RDN2008_UTM_zone_32N (EPSG: 7791) GCS_WGS_1984 (EPSG: 4326) as coordinate systems. All of them where respectively adjusted in order to converge into a single representation on ARCMap as plotting tool on which WGS_1984 was the default coordinate system selected. The specific field polygons were obtained from the technical reports made available, always by the DGS-UNMIG, via the VIDEPI project, a collection of technical reports containing information of non-active concessions in the country. A total of nearly 35 concessions and 40 fields were retrieved and their defining coordinates among other useful information was obtained.

Water withdrawal data was available only as cumulative production for the period 2003, 2006-2011 and 2018. Volumes are grouped at Comune level (data available from *Agenzia regionale per la prevenzione, l'ambiente e l'energia* from *Arpae Emilia-Romagna*).

Because of the investigate area in also interested by numerous Underground Gas Storage systems, the information about their position and about their historical production/injection data were also collected in the database. *ARPAE Emilia-Romagna* (subsidence analysis reports), DGS-UNMIG (hydrocarbon production bulletins, field charts and maps), *Ministero dell'Ambiente* (environmental permit documentation), *Dipartimento delle Acque della Regione Emilia-Romagna* (hydrological balances of produced water) and Edison Stoccaggio (subsidence report from a field under gas storage operation) where the main fonts to retrieve data.

Finally, the data types are described on the table below:

Data	Nature
Pumping volumes and pressures values:	• Numeric vs time: cubic meters with monthly frequency from 1991 to 2020.
Coordinates of the locations of interest for fluid extraction/injection sites:	• Latitude and longitude in degrees, locations from 1991 to 2020).
Coordinates of subsidence measurement points (MP) of interest:	• Latitude and longitude in degrees.
Subsidence values for the MP	• Quasi-monthly frequency from 2006 to 2011 (mm).

Table 6 - Available raw data for perform the analysis. Author.

The type of data was of several natures: Numeric vs time: Pumping volumes and Pressures values (cubic meters with monthly frequency). Coordinates of the locations of interest for fluid extraction/injection sites (latitude and longitude in decimal degrees). Coordinates of subsidence measurement points of interest (values of latitude and longitude), combined with them the lowering values of the ground with monthly frequency (mm). Time window from 1992 to 2016.

This huge amount of information related only to fluid extraction/injection sites (oil, gas and water), led to the need of developing a coordinated algorithm in order to produce the kml files as input to the GIS software which will convert them into shapefiles and finally as useful layers. Figure 31 gives a glance of the excel-based storage/preprocessing approach and interconnected files (the elements of the chart doesn't reflect any database related symbology).

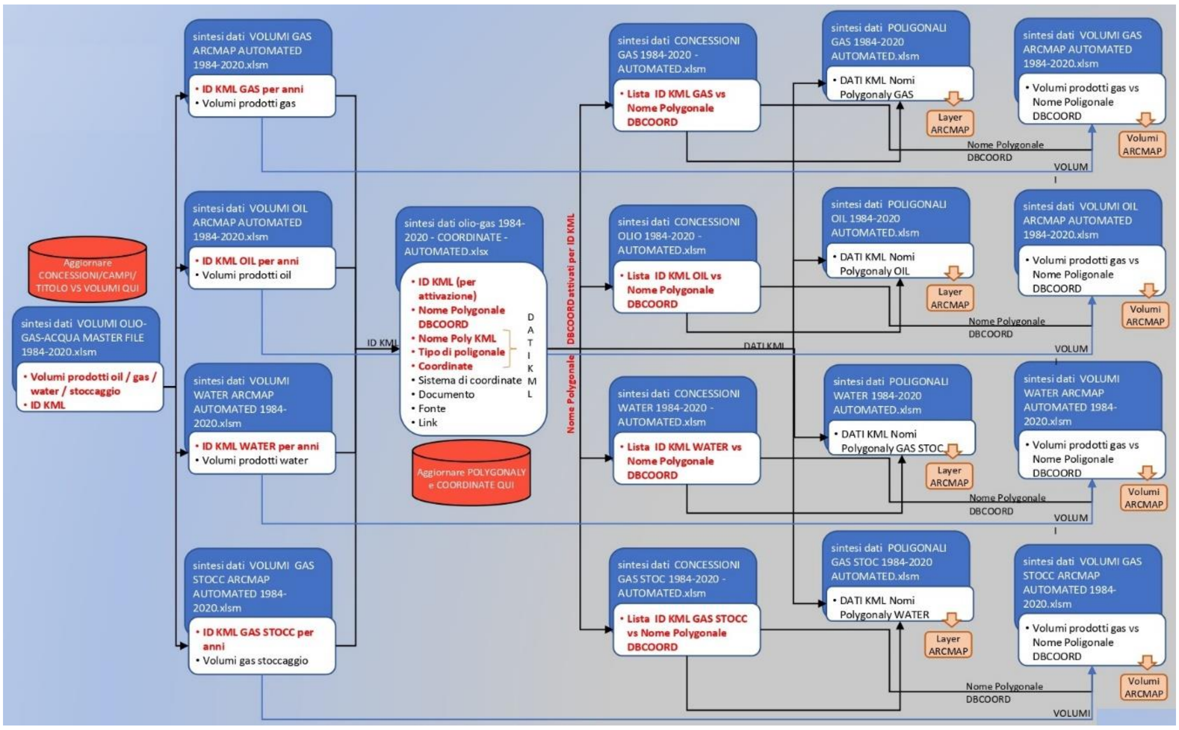

Aggiornare CONCESSIONI/CAMPI/ TITOLO VS VOLUMI QUI
sintesi dati VOLUMI OLIO-GAS-ACQUA MASTER FILE 1984-2020.xlsm
• Volumi prodotti oil / gas / water / stoccaggio
• ID KML
sintesi dati VOLUMI GAS ARCMAP AUTOMATED 1984-2020.xlsm
• ID KML GAS per anni
• Volumi prodotti gas
sintesi dati VOLUMI OIL ARCMAP AUTOMATED 1984-2020.xlsm
• ID KML OIL per anni
• Volumi prodotti oil
sintesi dati VOLUMI WATER ARCMAP AUTOMATED 1984-2020.xlsm
• ID KML WATER per anni
• Volumi prodotti water
sintesi dati VOLUMI GAS STOCC ARCMAP AUTOMATED 1984-2020.xlsm
• ID KML GAS STOCC per anni
• Volumi gas stoccaggio
ID KML
sintesi dati olio-gas 1984-2020 - COORDINATE - AUTOMATED.xlsx
• ID KML (per attivazione)
• Nome Polygonale DBCOORD
• Nome Poly KML
• Tipo di poligonale
• Coordinate
• Sistema di coordinate
• Documento
• Fonte
• Link
DATI KML
Aggiornare POLYGONALY e COORDINATE QUI
Nome Polygonale
DBCOORD attivati per ID KML
DATI KML
sintesi dati CONCESSIONI GAS 1984-2020 - AUTOMATED.xlsm
• Lista ID KML GAS vs Nome Polygonale DBCOORD
sintesi dati CONCESSIONI OLIO 1984-2020 - AUTOMATED.xlsm
• Lista ID KML OIL vs Nome Polygonale DBCOORD
sintesi dati CONCESSIONI WATER 1984-2020 - AUTOMATED.xlsm
• Lista ID KML WATER vs Nome Polygonale DBCOORD
sintesi dati CONCESSIONI GAS STOC 1984-2020 - AUTOMATED.xlsm
• Lista ID KML GAS STOCC vs Nome Polygonale DBCOORD
sintesi dati POLIGONALI GAS 1984-2020 AUTOMATED.xlsm
• DATI KML Nomi Polygonaly GAS
Layer ARCMAP
sintesi dati POLIGONALI OIL 1984-2020 AUTOMATED.xlsm
• DATI KML Nomi Polygonaly OIL
Layer ARCMAP
sintesi dati POLIGONALI WATER 1984-2020 AUTOMATED.xlsm
• DATI KML Nomi Polygonaly GAS STOC
Layer ARCMAP
sintesi dati POLIGONALI GAS STOC 1984-2020 AUTOMATED.xlsm
• DATI KML Nomi Polygonaly WATER
Layer ARCMAP
sintesi dati VOLUMI GAS ARCMAP AUTOMATED 1984-2020.xlsm
• Volumi prodotti gas vs Nome Poligonale DBCOORD
Volumi ARCMAP
sintesi dati VOLUMI OIL ARCMAP AUTOMATED 1984-2020.xlsm
• Volumi prodotti gas vs Nome Poligonale DBCOORD
Volumi ARCMAP
sintesi dati VOLUMI WATER ARCMAP AUTOMATED 1984-2020.xlsm
• Volumi prodotti gas vs Nome Poligonale DBCOORD
Volumi ARCMAP
sintesi dati VOLUMI GAS STOCC ARCMAP AUTOMATED 1984-2020.xlsm
• Volumi prodotti gas vs Nome Poligonale DBCOORD
Volumi ARCMAP
Nome Polygonale DBCOORD
VOLUMI

Fig. 31 - Excel workflow for cartographic and volumetric data management. Author.

Once the research stage was concluded the following information was available and useful to be analyzed:

GIS data	Oil & production records and field geoferenciated geometries for nearly 35 concessions (1991 up to 2016).
Volumetric data	Injection/production rates and volumes, as well as pressure for selected cases for three UGS fields, active or inactive, within the period 2002-2013 Pressure data for a few hydrocarbon production/storage fields
Subsidence data	A shapefile containing close to 325.000 ground vertical movement MP from the Ravenna area, for the period 2011-2016, on which three of the UGS fields reside. Vertical velocity time series of the area on the same previous shapefile

Table 7 - Available processed data for perform the analysis. Author.

5.2 Pilot test identification

The following criteria were established to select the more suitable candidates to be analyzed:

- Plotting of UGS polygonal and hydrocarbon production fields all over the Emilia Romagna region.
- Spatial and temporal intersection with the available subsidence dataset at the moment (Ravenna area).
- Pilot test identification for accuracy and validity of the approach.

The selected type analysis needed to be one on which a population of hundreds of thousands of points (even millions) could be compared and categorized in the shortest time possible, that trends could be extracted and the change of parameters of analysis could be performed without severely affect the availability of the available computational resources. It is to say that when looking for trends and recognizable patterns the best solution is represented by the UGS fields, where seasonal patterns of injection and production are established along with defined periods of the year for them to happen. In fact, UGS operations consist of periodical and seasonal injection (in summer time) and production (in winter time) of natural gas in underground formation, usually exploited gas reservoir then converted in UGS system. It is also possible to compare both pressure and volume extraction with values of vertical subsidence as many authors have already demonstrated (Codegone et al. 2016 for instance): when lithology is adequate, the changes on subsidence values for a filed have a direct and proportional correlation with the fluid movement patterns. Fig 32 exemplifies the seasonal behavior mentioned.

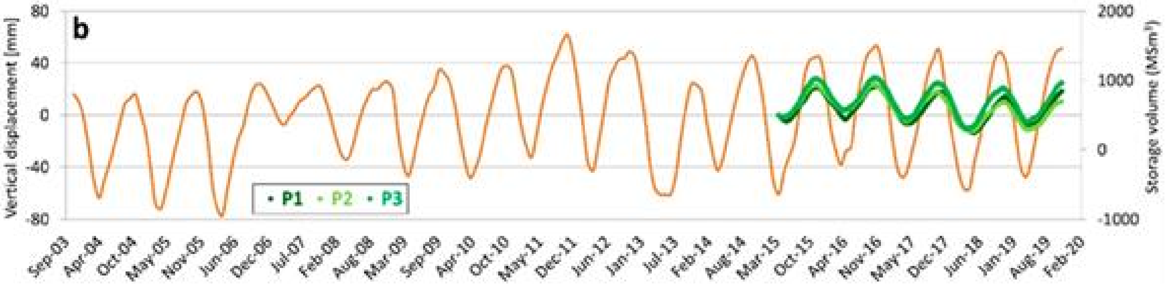


Fig. 32 - Seasonal behavior of an UGS data example of vertical movement, showing the work of Benetatos et al. (2020).

5.3 Data analysis methodology

Because of the peculiar features of the anthropogenic events under analysis (UGS), the points affected by UGS could be identified by a similar special and temporal ground movement behavior. Furthermore, according to the amplitude of the movements, the point can be subdivided according to their 'vicinity' to the UGS, or, in other terms, according to the effects of UGS.

As a consequence, the time series decomposition analysis and a subsequent clusterization represent the proposed data analysis methodology.

The time series decomposition analysis should allow to identify three important components:

- seasonal component, or quantitatively how varies the phenomenon;
- trend that represents how the overall patterns increase or decrease both in magnitude that in amplitude of the seasonality effect;
- random or reminder values that could indicates a possible way to characterize the area in terms of an on-going unavoidable subsidence (possibly natural and region-related).

The seasonality, individuates the gas storing effect on the vertical ground movement in agreement with the seasonal UGS operations concept: gas injection in the reservoir from March to September (uplifting) and production from October to February (subsidence). The trend can give information in relation with UGS activities changes in the injection or production pattern, and maybe also about other regional or local scaled phenomena, always in relation with the considered time span. For the considered case, natural phenomena (e.g., post glacial rebound, erosion or sediment deposition) are not able to significantly affect the subsidence or uplifting values, because they tend to work on a larger time scale (Carmina and Donati, 1999), unless violent episodes like earthquakes or great sediment displacement due to landslides occurs.

Due to the fact that the seasonal behavior analyzed is nearly constant in amplitude and periodicity and additive approach was taken. Fig 33 illustrates a classic additive decomposition example chart.

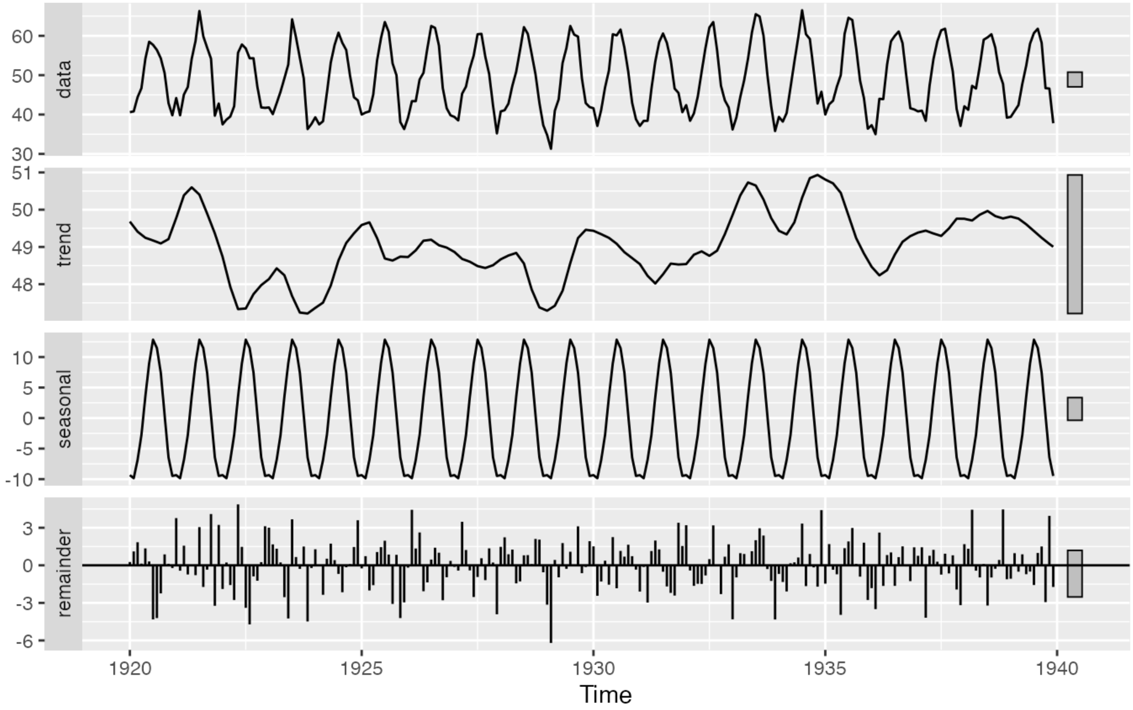


Fig. 33 - Example of decomposition analysis plots. Hyndman et al.

The clusterization, allows to group, separate and categorize different classes of movement on which points will respond in the same way, even though not physically connected. As a consequence, points (and so the area) affected by UGS can be identified and grouped. A further clusterization allowed the identification of different degree of influence of the investigated phenomenon.

The selected clusterization approaches were: hierarchical and k-mean as a means to determine which one will better represent the subsidence phenomena, both are prototype based, with the main difference that an hierarchical approach creates nested relation between clusters, making them subcluster up to a certain level, and the k-means approach partition data on well-defined and differentiated classes on which each class is represented by the distance of its center to the center of others, user defined, n-classes.

It is important to point out the superposition of effects (Carmina and Donati on 1991) between natural and anthropogenic components. Generally speaking, the effect of natural component on the seasonality can be negligible due to the difference in scale time of the two phenomena. Whereas, other anthropogenic phenomena are not so easily identifiable form the UGS activity. Water production is characterized by a seasonal activity as well: water withdrawal is more important during the spring and summer season while gas injection takes place, on the other hand during autumn and winter water withdrawal diminish leaving the responsibility of the fluid production almost entirely for the gas production of storage gas for heating purposes. Furthermore, other activities are time invariant and punctual: buildings and infrastructure construction, effects of highway or railway etc.

The data analysis method was tested on two different case study: one pilot test and one case study. The area of investigation was identified as well as the measure points (MP) available, and the main anthropogenic activity presents. Then, via R programming language, the following steps are performed: time series decomposition analysis, data clusterization and layer generation to be plotted over a map containing information of the selected MP.

It is noteworthy to remind that all the time series correspond to MP spread over the region, in this case the Ravenna area, and a proper integration between the cluster analysis and the GIS environment is vital for the adopted methodology. In this way, R (via R Studio) works as a link between the cartographic o subsidence data contained in the shapefiles (latitude, longitude and ground vertical movement on time) and the data analytics tools (the clustering techniques already available as programmed functions).

The script developed (entirely available on appendix III), counts with high versatility and flexibility, allowing easy changes on its structure, and permitting the user to customize the data analysis to be performed with a minimum of programming knowledge. Number of clusters, type of data to cluster (raw or normalized), GIS shapefile writing, excel output with the clusterization results, plotting of absolute or normalized data (even not clustering the same types) and interactive plots of the clusters over a map within the program running the script, are among the options to choose. In order to help with the post-processing analysis and to better organize the results, a name-based approach for the output files was developed. In each case the name of the file will contain (separated with hyphens or underscores): name of the analyzed area, clustering techniques used, number of clusters, type of clustered and plotted data, and some other relevant information (type of coordinate system or stage of the UGS process for instance).

5.4 Pilot Processing Test

The pilot test was conducted over the UGS Field 2 concession, which accounts for three storage clusters: Field A (628 MP), Field B (7 MP) and Field C (28 MP). The number of MP indicates the amount of measuring points from the Ravenna data set that resides at the inside of the field polygons that were previously retrieved and plotted over ArcMap. Fig 34 shows the selected study area, 35 and 36 detail the internal zones and 37 shows the ARPAE MP residing on Field C. Fig 38 show the injected/produced gas volumes for the 3 field: higher the volumes usually stronger the effects in terms of induced vertical movements on ground surface.

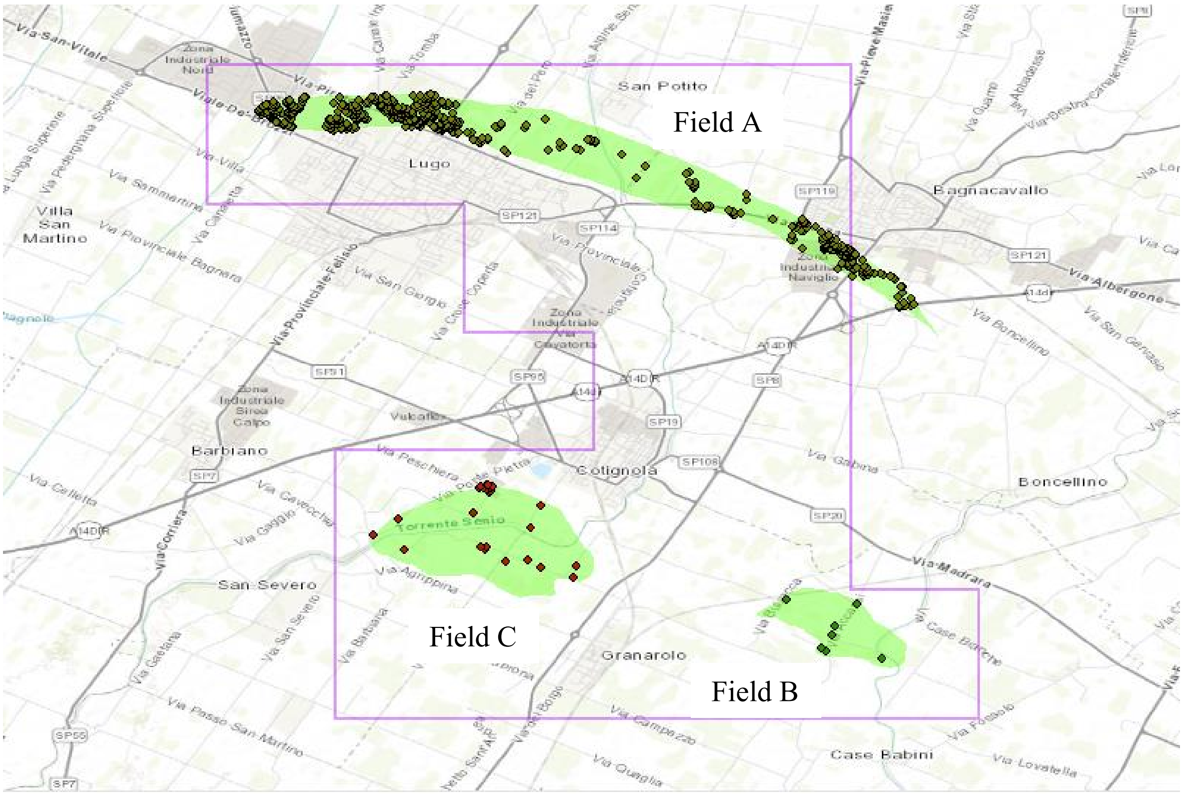


Fig. 34 - Field A (up), Field B (down right) e Field C (down left) clusters.

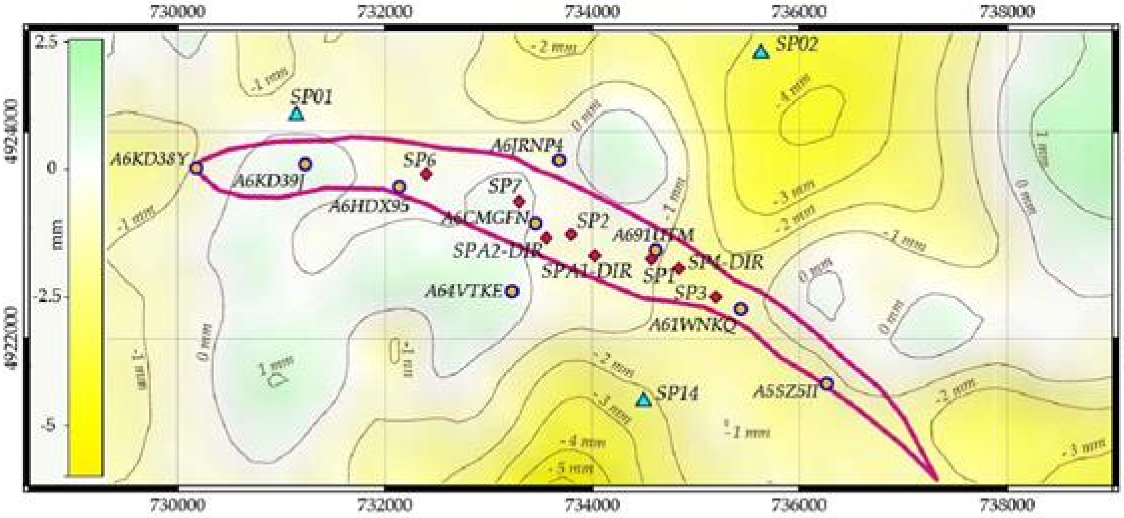


Fig. 35 - Field A polygon detail. The scale represents the year velocity of vertical ground movement from April to September 2020 (internal points do not correspond to the used MP's). Edison internal technical report.

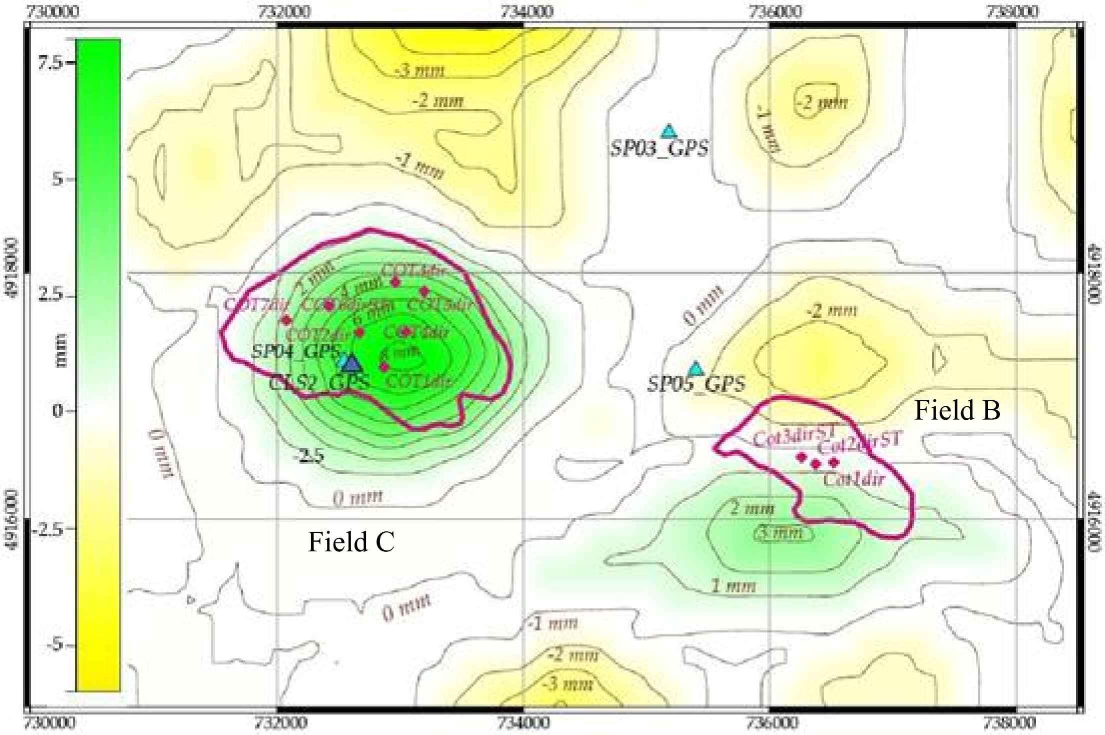


Fig. 36 - Fields A and B polygons detail. The scale represents the year velocity of vertical ground movement from April to September 2020 (internal points do not correspond to the used MP's). Edison internal technical report.

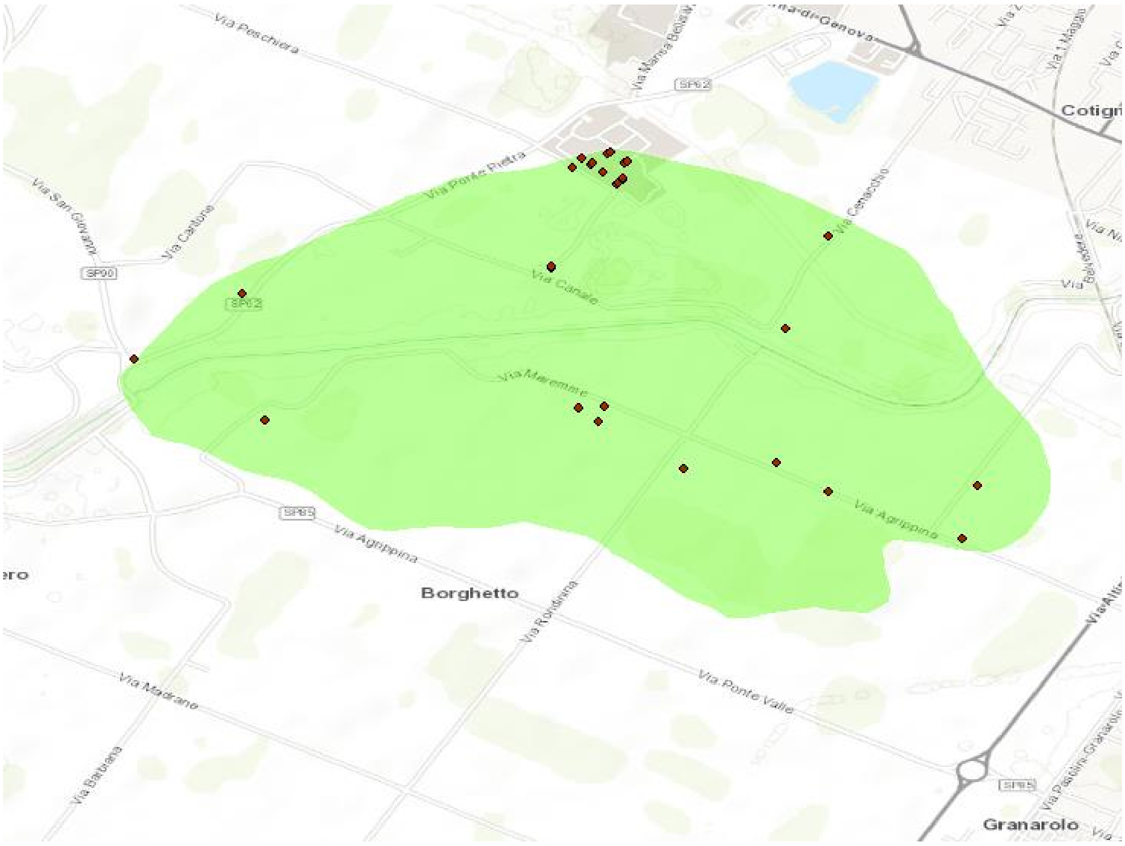


Fig. 37 - Field C MP distribution over the Ravenna dataset, GIS map view. Author.

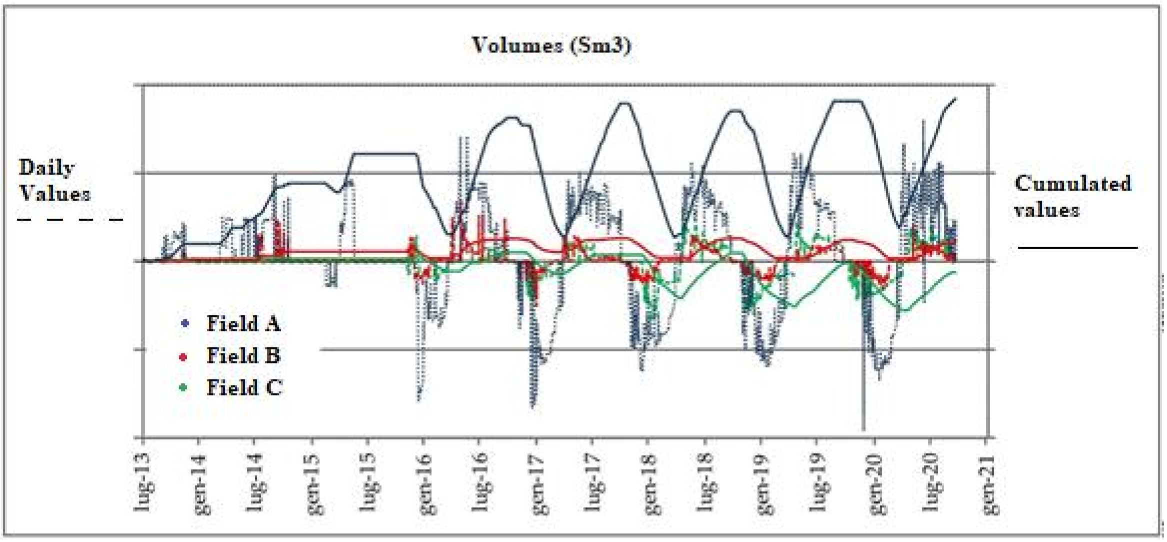


Fig. 38 - Injection/production trends Clusters Field A, Field B and Field C. Edison internal technical report.

Once determined and specified the more suitable MP, the approach followed this algorithm:

1. AREAL DATA SELECTION FROM ORIGINAL SAR DATASET: Here, all the vertical movement information regarding to the time series is saved for the selected MP.

2. RESAMPLING BY LINEAR INTERPOLATION. Since the seasonal decomposition requires one value by period of the month to be considered, a linear interpolation was performed to get a single characteristic value of vertical movement at the middle of the month. Not doing it affects the results of the decomposition because the function on R interprets as another consecutive period different measures of subsidence for the same month. On R the approxfun() linear function approximation from stats library was used.

3. TIME SERIES DECOMPOSITION: On this stage the decomposition analysis is performed for each time series belonging to every MP and getting a vector with all the information. On R the stl() function from stats library was used.

4. TIME SERIES NORMALIZATION. Z-SCORE technique was used in R, after finding min-max normalization type (from 0 to 1) not satisfying. Normalization converts numeric values in the dataset to a common scale without distorting the ranges of values or losing information. In the present case study, it allows a better model of the seasonal phenomenon on a pseudo-sinusoidal behavior, because it keeps the original data's broad distribution and ratios while maintaining values within a scale applied to all the population (Microsoft documentation, 2019)

5. CLUSTERIZATION (HCLUST E KMEAN): The cluster functions separate in classes, according to the behavior of the seasonal component, the different time series of all the MP. The functions aim and seek for a repeatable pattern among the time series that can represent and group them in a number that was previously defined by the user (n-classes defined, n-prototype patterns defined). On R hclust() and kmeans() from stats library.

6. TIME SERIES PLOTTING (SEASONAL, TREND, REM). The centroid of each time series cluster was plotted reflecting the hierarchical and k-means clusterization analysis for diverse n-cluster. The graphs were compared with production time frame, assumed from October to march, and injection time frame, assumed form march to October. Functions summarise() and ggplot() from dyplr and ggplot2 libraries where used on R.

7. CLUSTER MAPPING (ARCMAP). Taking advantage of R robustness, a shapefile (with the required coordinate system) was generated and plotted as a layer on ArcMap, facilitating in that way the geographical correlation analysis between the clustering processing and the position of the MP. On R writeOGR() function from rgdal library made it possible.

5.4.1 Dummy test:

For the UGS Field 1, a dummy test was carried out to test the beta clustering algorithm. Three reservoir polygons were available: Field A, Field B and Field C, for which information regarding the trend of volumes (daily and cumulative) are available (fig. 38, Edison Stoccaggio technical reports). As evinced from figure 34, the available number of MP is very dissimilar among them, but the behavior of the reported subsidence is more intense for the Field C, which almost six-folded the other two areas increasing the chances to observe the phenomenon more clearly. In this sense, this field can be seen as a good compromise between subsidence phenomenon presence (due possibly to anthropic reasons, gas stocking in this context) and the number of points to be clustered.

The decision of selecting one or both types of clustering techniques (hierarchical and partitive) relies on the fact that we found ourselves in an experimental zone where, among other clustering strategies (e.g., dbscan or fuzzy) these seem to be more suitable to cluster a seasonal pattern, because both of them are prototype-based and the algorithm will find the more similar curves to group. It is expected that the areas containing MP interested by the same, or at least similar, local subsidence behavior will be grouped consequently. We decided to kept open both ways (hierarchical and partitive) in order to enlarge the possibilities offered by both of the approaches and later drop conclusions about their performances. The first one will allow to make relation between cluster (is a nested approach on which each cluster is a subcluster of another bigger that can contain it). The second will split them and create completely independent classes. At this point no solid criteria emerged for selecting one or other so we decided to maintain both, also because the computational resources allowed to run them in reasonable times in the same algorithm making possible output comparisons.

Should be noted that for all the produced plots (clusters results figs. 39 and 41) the UGS stage of “production/extraction” of gas is named *estrazione* (represented by a red line during from October to March) and de “storing/injection” period *iniezione* (represented by a blue line during from April to September). It is also presented on figures 40 and 42 their geographical representation.

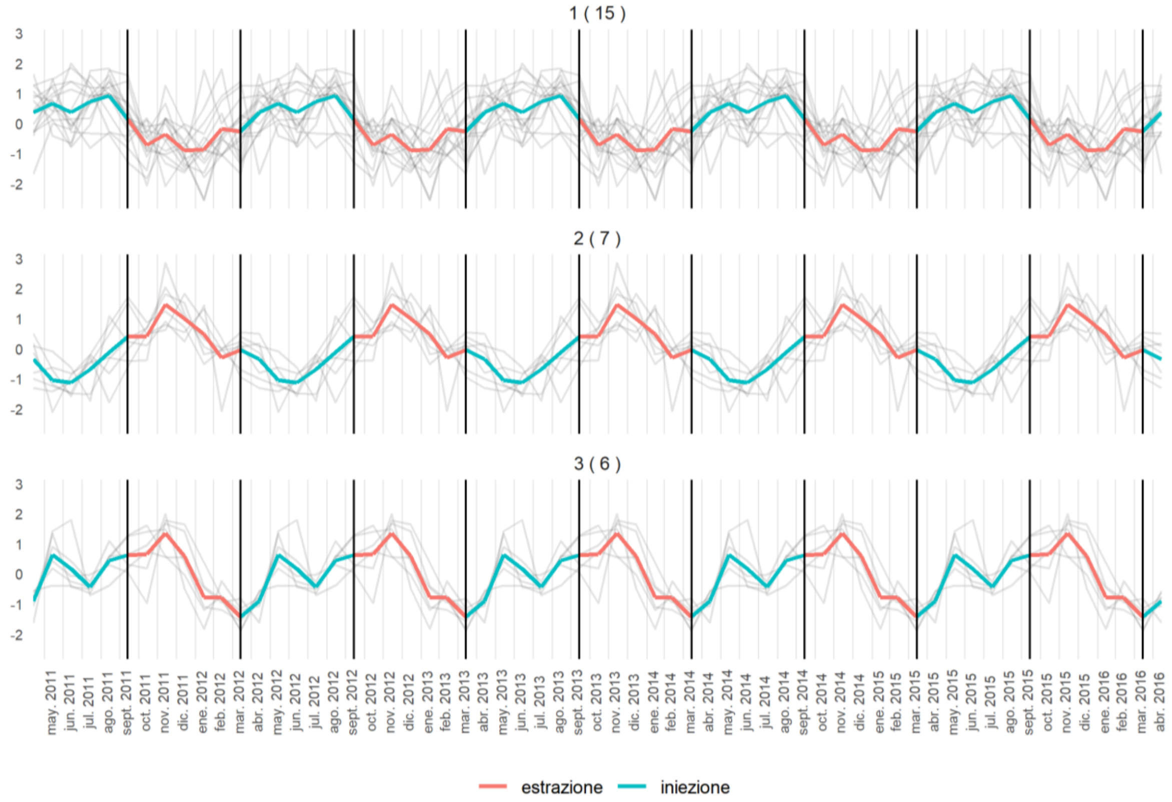


Fig. 39 - Field C area, Z normalized values were clustered using Hierarchical and plotted its seasonal component for the decomposition analysis. Author.

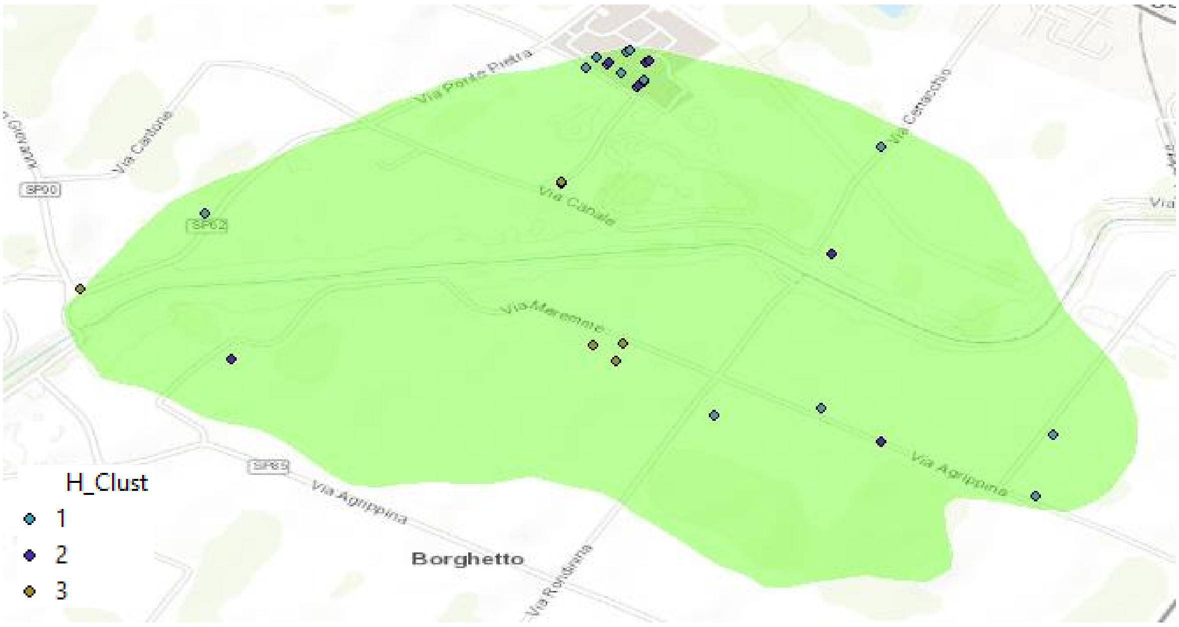


Fig. 40 - Field C area, Hierarchical clustering results over the map. Colors correspond to seasonal plot. Author.

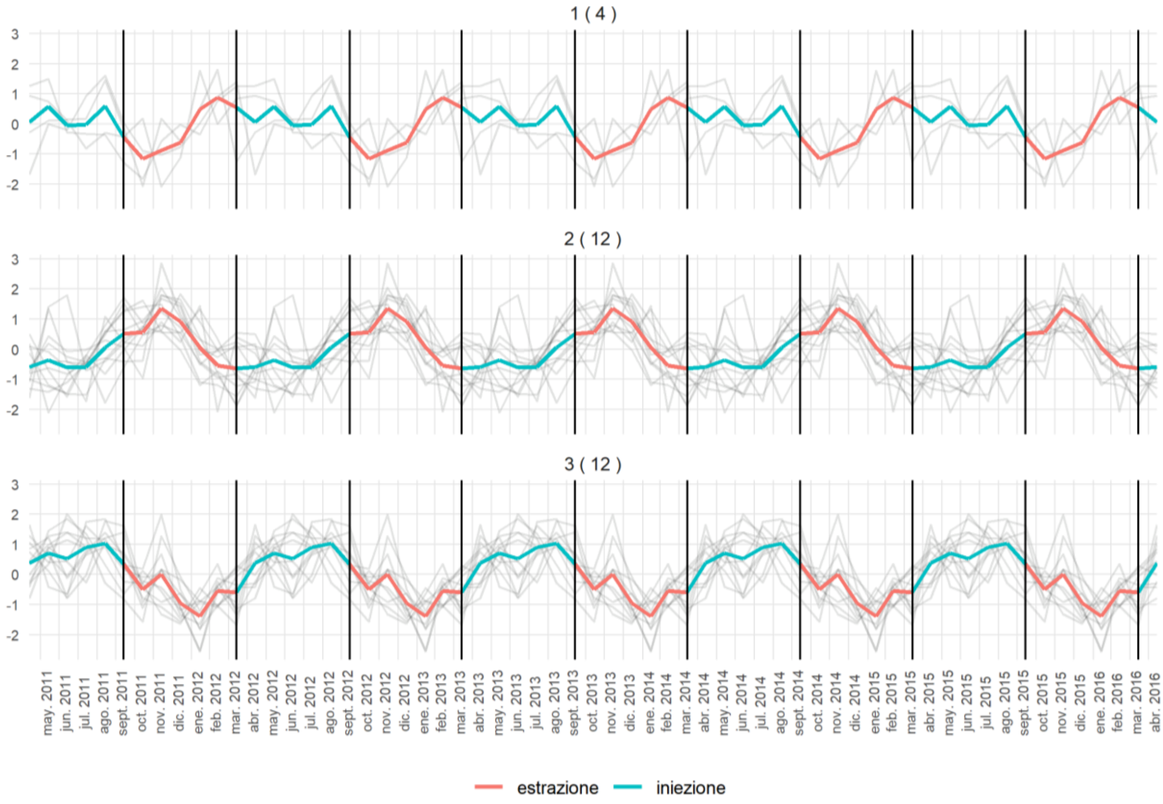


Fig. 41 - Field C area, Z normalized values were clustered using K-means and plotted its seasonal component for the decomposition analysis. Author.

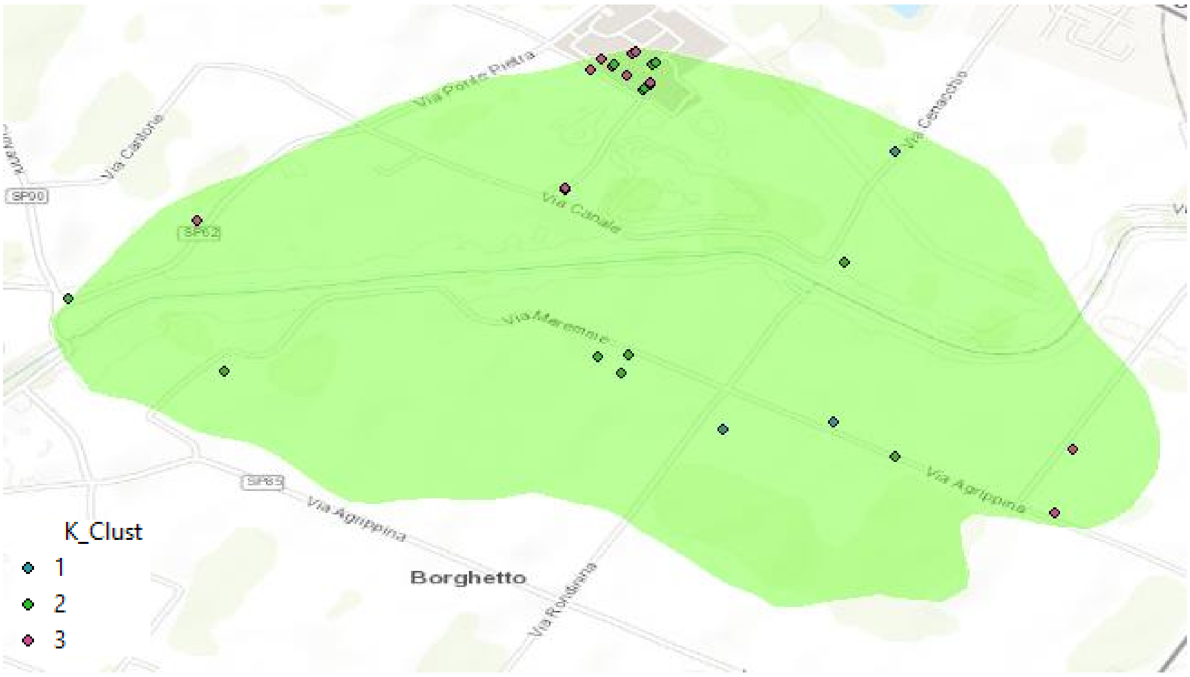


Fig. 42 - Field C area, K-means clustering results over the map. Colors correspond to seasonal plot. Author.

Table 8 resumes significant data for the test:

Area Selected	Field C, from UGS Field 2
Number of clusters analyzed	3,5,6 - 5 and 6 discarded for excess of dispersion
Type of clustering	K-means and Hierarchical
Hierarchical clustering output	Class 1 (15 MP) matching subsidence behavior
K-means clustering output	Class 3 (12 MP) matching subsidence behavior

Table 8 - Dummy test initializing data and results. Author.

As can be seen on figures 39 and 41 the resulting classes for both of the clustering approaches are similar in quantity, spatial location and sinusoidal pattern, main difference is related to the shape of the series non in agreement with the subsidence effect. There's a cluster of points recognized by both algorithms at the north area of the map (figures 40 and 42), that can be seen as a validation of their suitability for correctly recognize the behavior of the vertical ground movement time series. These results, consistent and coherent with what we have expected, allowed to consider the approach, and the script, valid to initiate the analysis over bigger areas.

5.5 X-Field Case Study

The X-FIELD is nowadays operated as UGS, it is located in northern Italy, in the Ravenna area, within a 100 km2 onshore concession (fig. 43 see UGS Field 2).

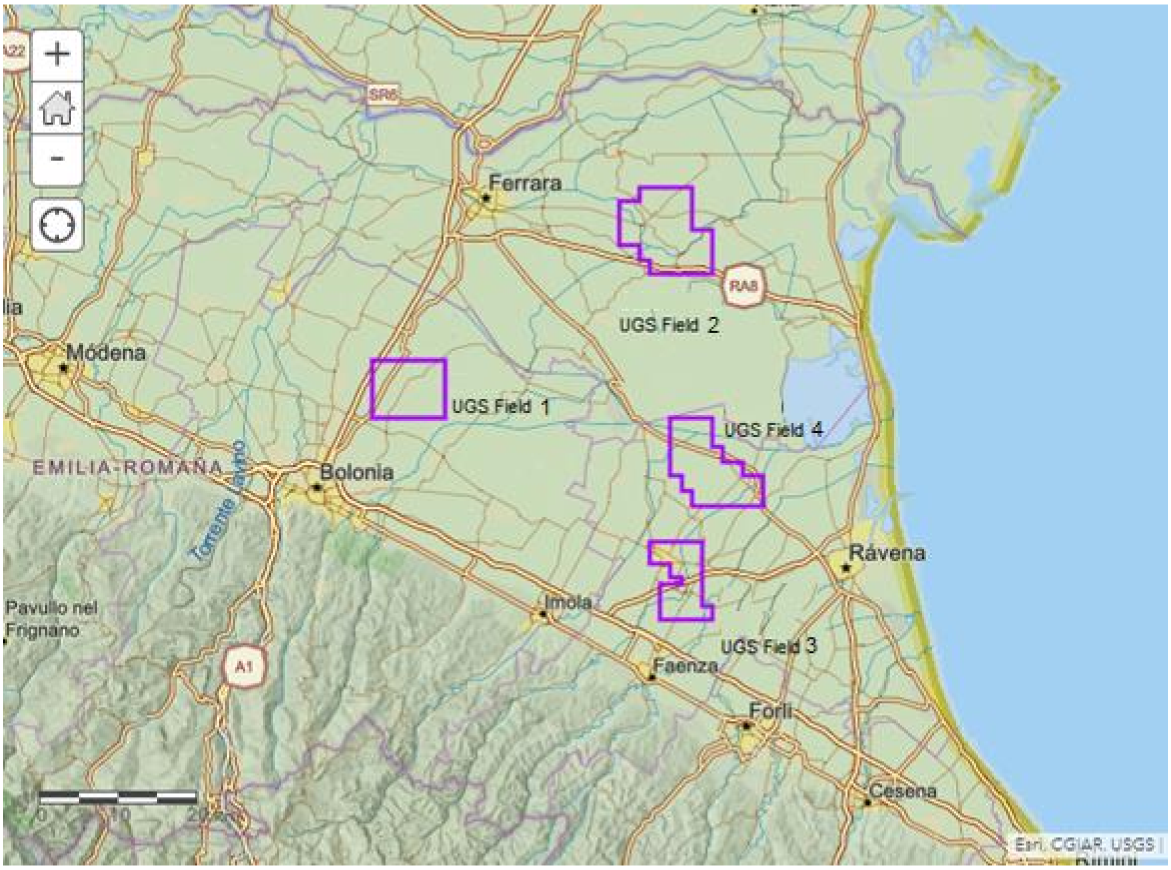


Fig. 43 - Detail on UGS fields on the Ravenna area, special attention to the UGS Field 2, upper right area of the map. DGS-UNMIG

Discovered in the early 1960s, after 20 years of primary production reservoir pressure had dropped by approximately 90 barsa. The field was converted into a natural gas storage facility in the 1980s. It is currently operated as UGS at a maximum injection pressure equal to the initial pressure. The storage schedule is seasonal, with injections taking place from April to October and withdrawals taking place from November to March (Codegone et al., 2016).

The field was selected for a further test of the data analysis methodology because:

- The area is covered by InSAR data; the available measure point density is satisfactory;
- No gas production in the same area and in the same period is present;
- Water extraction of the area is not taken into consideration and can be further analyzed the superposition effects on a successive study. Water extraction of the area is among the lowest values registered according to the regional authorities reports (*Dipartimento delle Acque della Regione Emilia-Romagna).*
- The UGS is a long term, well-known activity, object of technical paper (Codegone et. al, 2016): information about the area of influence of UGS in terms of ground movement is available and these maps are used to verify the results of the proposed data analysis methodology.

5.5.1 Data set

Initially a full dataset comprehensive of subsidence InSAR measurement from all the Ravenna area was available, about 325.000 measuring point between permanent and distributed scatters (see fig. 44). In order to find a good compromise of scale and efficiency of the available resources the area was later scaled at various levels: regional, local and areal, up to 16.144 points, 793 and 324 MP respectively, as can be seen on fig 2. These points corresponding to the period 2006-2011 are already elaborated in order to supreme noise and interferences from the satellite passage through the area of interest.

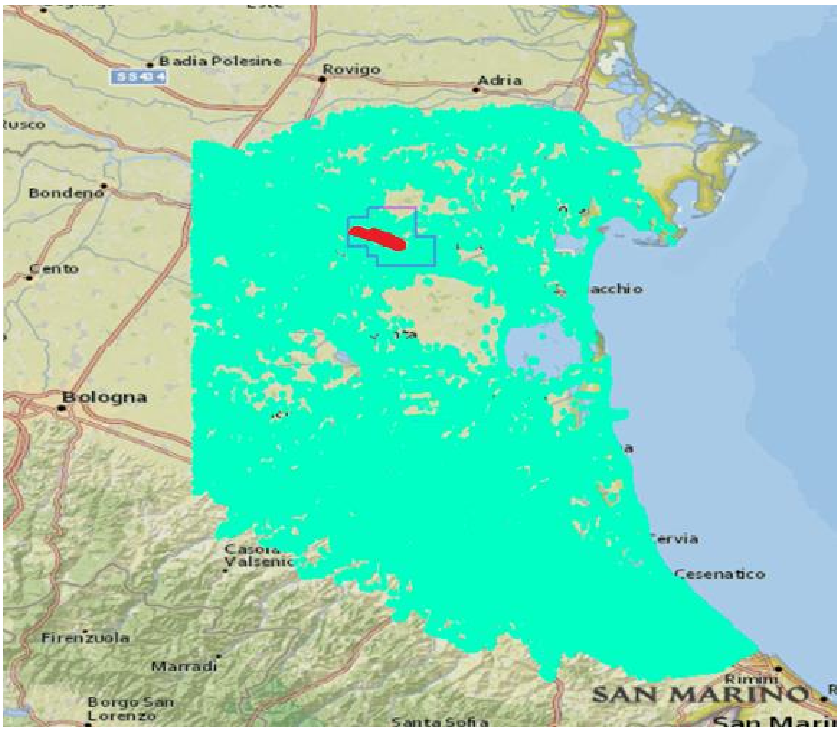


Fig. 44 - Location of the X-Field (on red) over the Ravenna dataset. DGS-UNMIG – ARPAE.

The analysis was carried out by defining the three zones as following: Regional X-Field (urban areas surrounding the UGS Field 2, fig 45), Local X-Field (defined by UGS Field 2 concession's reservoir boundaries, fig. 46) and Areal X-Field (polygons defined by the isolines of the Codegone's work for subsidence, figs. 48, 49 and 50). Fig 45 portraits the areal distribution of the MP along the Ravenna dataset, at all three levels.

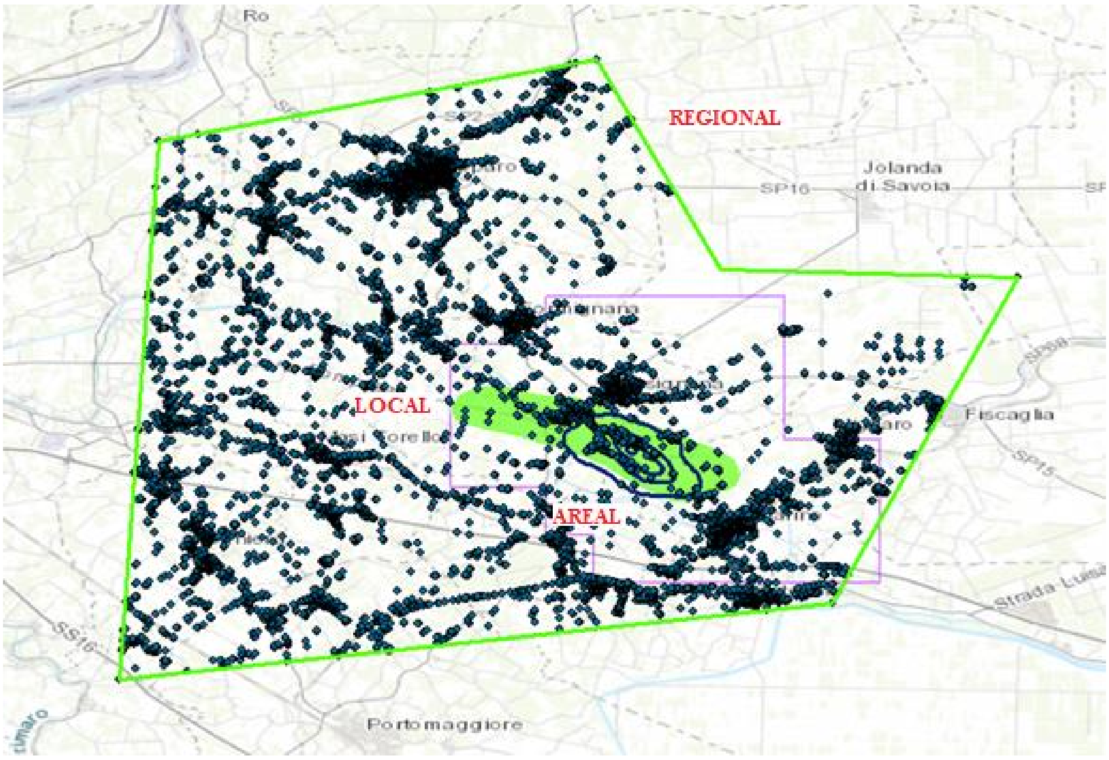


Fig. 45 - Regional analysis polygon (outer on green) comprehending 16.144 MP, including the X-Field among other urban and industrial centers. Local level is represented by the internal green bean-shaped area (the X-Field reservoir area); inside of it the polygon in blue, for the Areal level, represented by the isolines of vertical displacement developed by Codegone (2016). Purple polygon indicates the UGS Field 2 concession's administrative delimitation for merely informative purposes. DGS-UNMIG, ARPAE and Author.

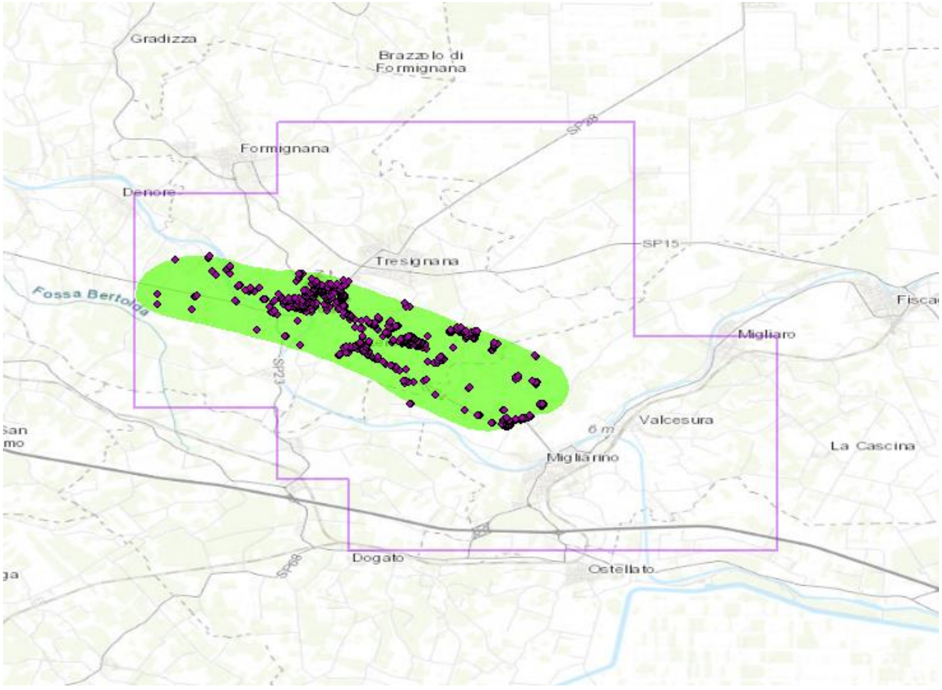


Fig. 46 – X-Field (Local level) zoom showing the 793 MP available. DGS-UNMIG – ARPAE.

The work of Codegone's analyzes the subsidence of the field and retrieves two important ground movement behaviors, corresponding to injection and production behavior of the subsidence. As a result, fig. 48 shows two maps with iso-variation lines for the vertical ground displacement for a selected area characterizing both periods, in agreement with seasonal behavior of the pressure change due to the gas storage operations. In fig. 47 the overall vertical displacement of two measure points (related to the Edison technical report, not to the ARPAE MP) inside the field are confronted against the pressure change on the reservoir on time.

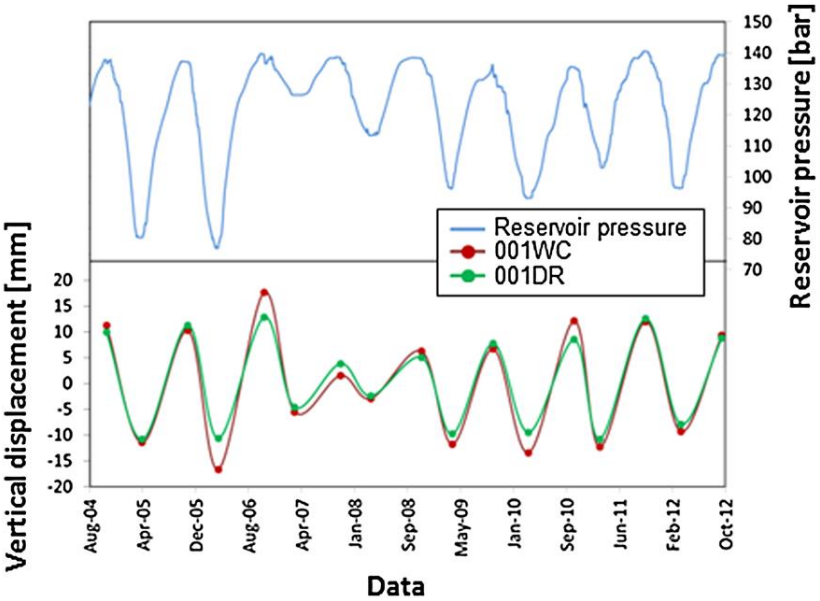


Fig. 47 - X-Field pressure and vertical displacement data behavior, previously analyzed from Codegone et. al. (2016).

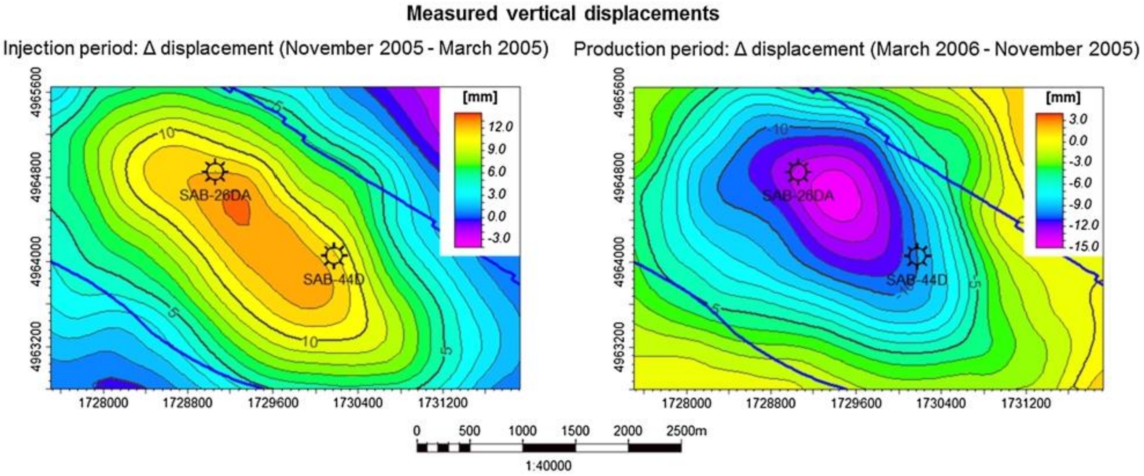


Fig. 48 - Injection/Production with identified iso-variation lines for Injection and Production subsidence related periods (Codegone et al, 2016), base for the refined area selection.

For our case the injection-season period polygon was selected as refinement grid (Areal X-Field) due to the presence of a greater amount of MP, as figures 49 and 50 show.

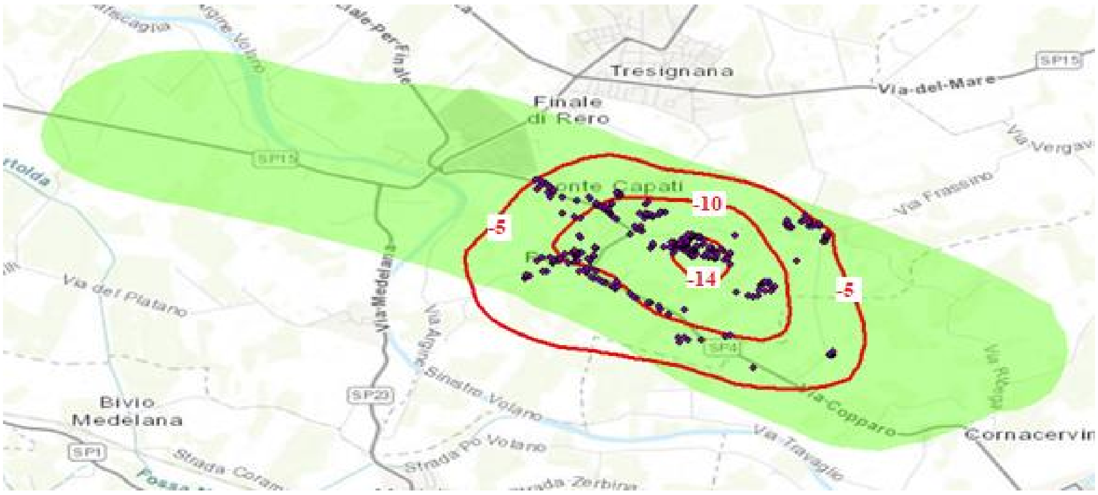


Fig. 49 - X-Field area zoom showing 302 MP inside the Production subsided related polygon area (in red). DGS-UNMIG, ARPAE and Codegone.

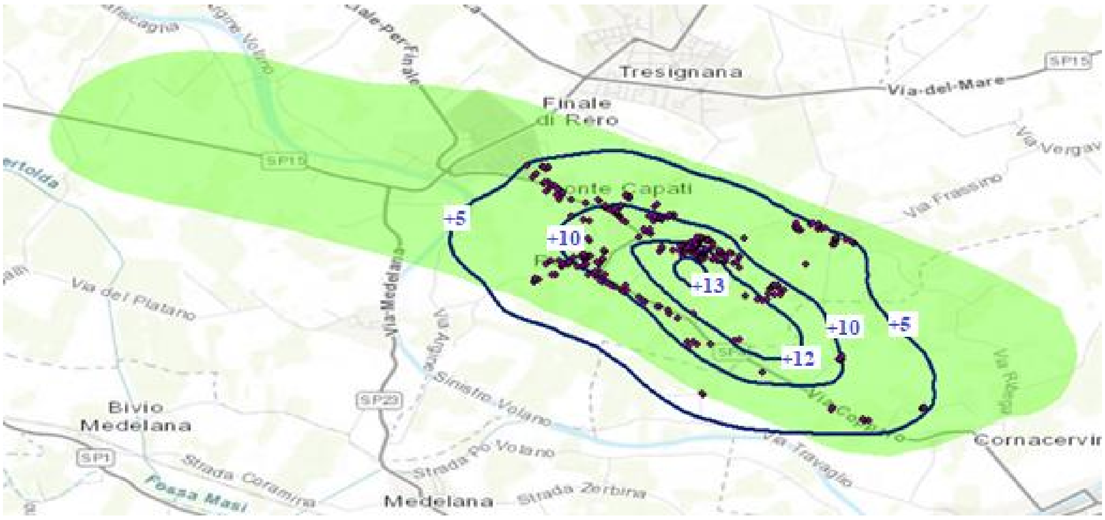


Fig. 50 - X-Field area zoom showing 324 MP inside the Injection subsided related polygon area (in blue). DGS-UNMIG, ARPAE and Codegone.

5.5.2 Workflow

Investigated dataset:
Spatial subdivision: macro "Regional X-Field", meso "Local X-Field" and micro "Areal-X-Field".
Time frame: the time series started on May 2011 and ended on April 2016.
Time sample: a total of 75 raw measure of subsidence on time were resampled into 64 in order to ensure the correct functioning of the decomposition analysis algorithm on R.

Data analysis methodology:
Decomposition: the 3 main component (seasonal, trend and remainder) were defined. The attention was focus primarily to the seasonal component because it allows a clear identification of the phenomenon. In addition, the trend component could add very important features not only of the phenomenon but also in terms of superposition of effect. Finally, the remainder component stands beside and beyond the focus of the present work.
Clusterization: each dataset was tested via both hierarchical and k-mean clustering approaches, considering 4 and 8 cluster classes analysis, based on the behavior of the seasonal component from the decomposition analysis (the time series were initially linearly-resampled and then z-normalized before being clustered).

6 RESULTS

For each dataset, time series cluster plots for both hierarchical and k-mean clustering approaches, mapping of the corresponding clusters over its GIS-related polygon and pie charts synthetizing the points distribution. Both approaches, Kmean and Hierarchical, where always used attending for further selection of a preferred one or to set the rules that defines the choice on each particular case. The plots and clusters mapped always correspond to z-normalized clustered (resampled vertical movement values) and plotted data; in case of the areal seasonal analysis section, both, z-normalized and absolute resampled values were plots (but always clustered the z-normalized). In the areal X-Field section an illustrative clustering and plotting of raw sampled date was performed. No raw data was used for clustering purposes.

6.1 Regional Level Seasonal analysis

6.1.1 4 Classes clustering results (z-normalized values plotted):

Seasonal component time series clustering charts

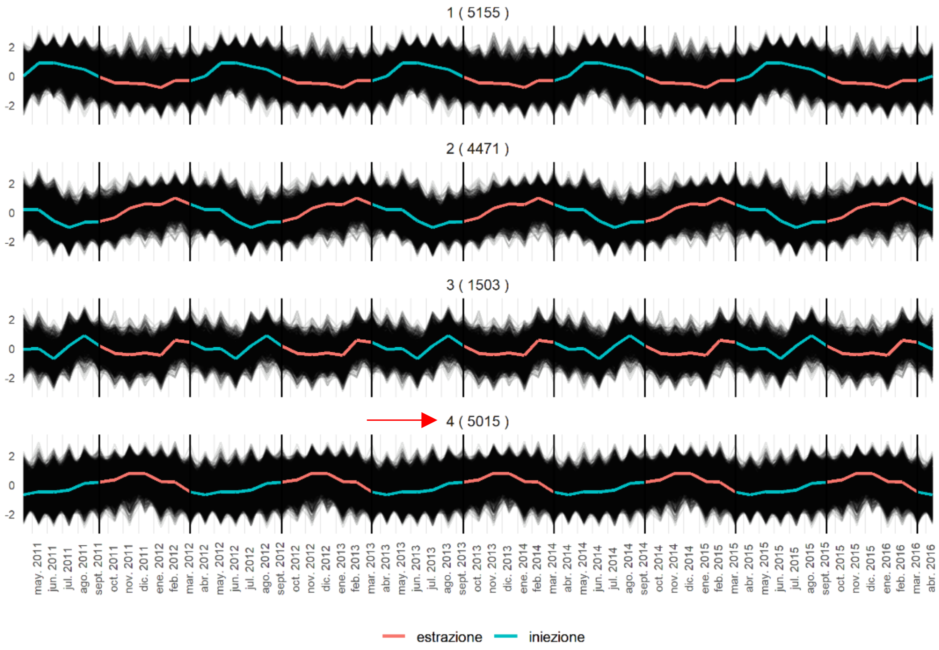


Fig. 51 - Regional Level (16.144 MP) - Hierarchical Clustering (4 Classes) seasonal component, z-normalized values clustered and plotted. Author.

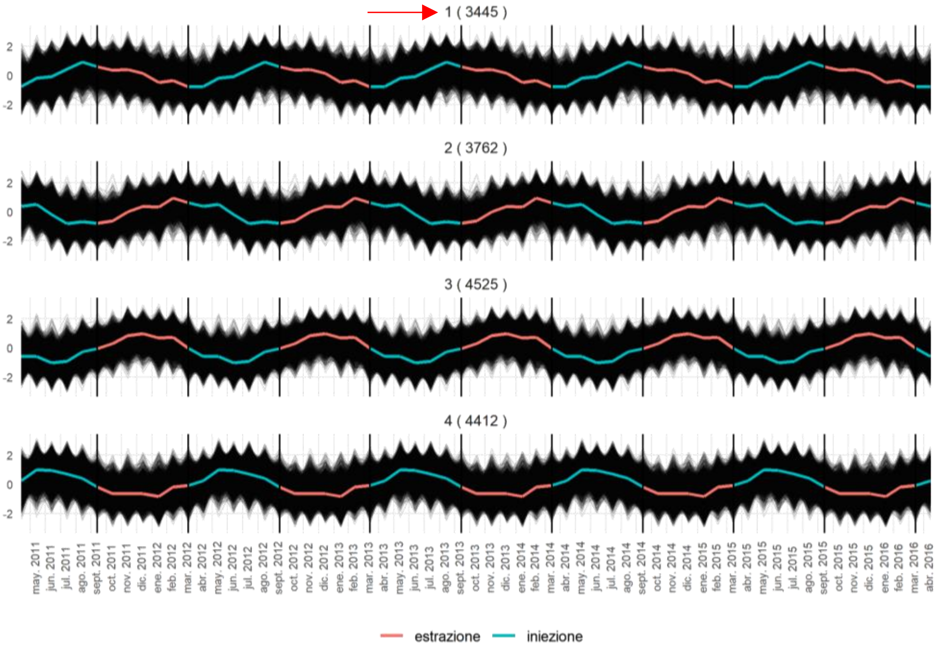


Fig. 52 - Regional Level (16.144 MP) - Kmean Clustering (4 Classes) seasonal component, z-normalized values clustered and plotted. Author.

MP distribution pie chart by clustering type

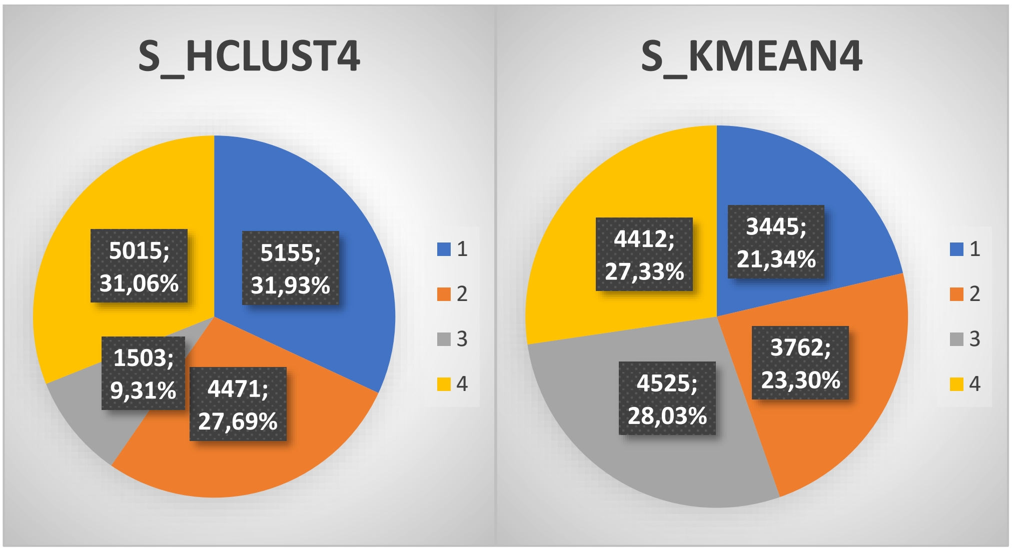


Fig. 53 - Regional X-Field, 4 Classes Hierarchical and Kmean Clustering (normalized data), seasonal component, MP distribution and relative percentages for each class. Author.

GIS visualization Hierarchical clustering

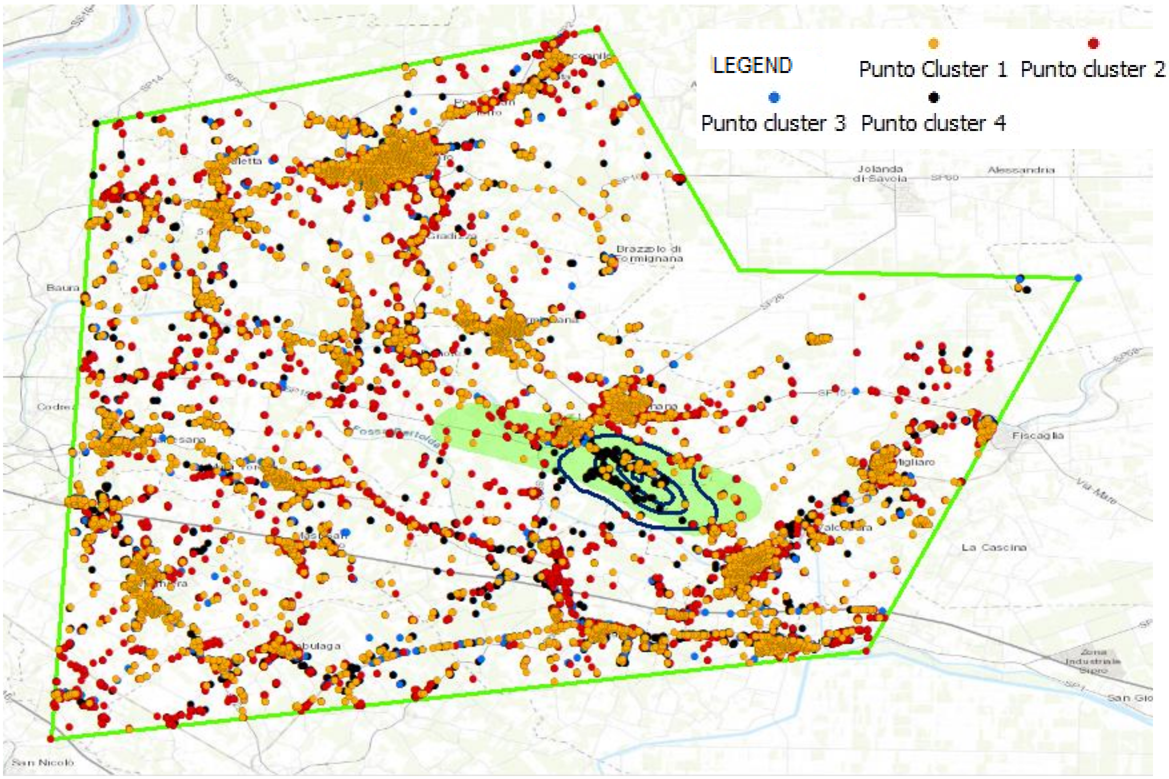


Fig. 54 - Regional X-Field, 4 Classes Hierarchical Clustering, seasonal component. Author.

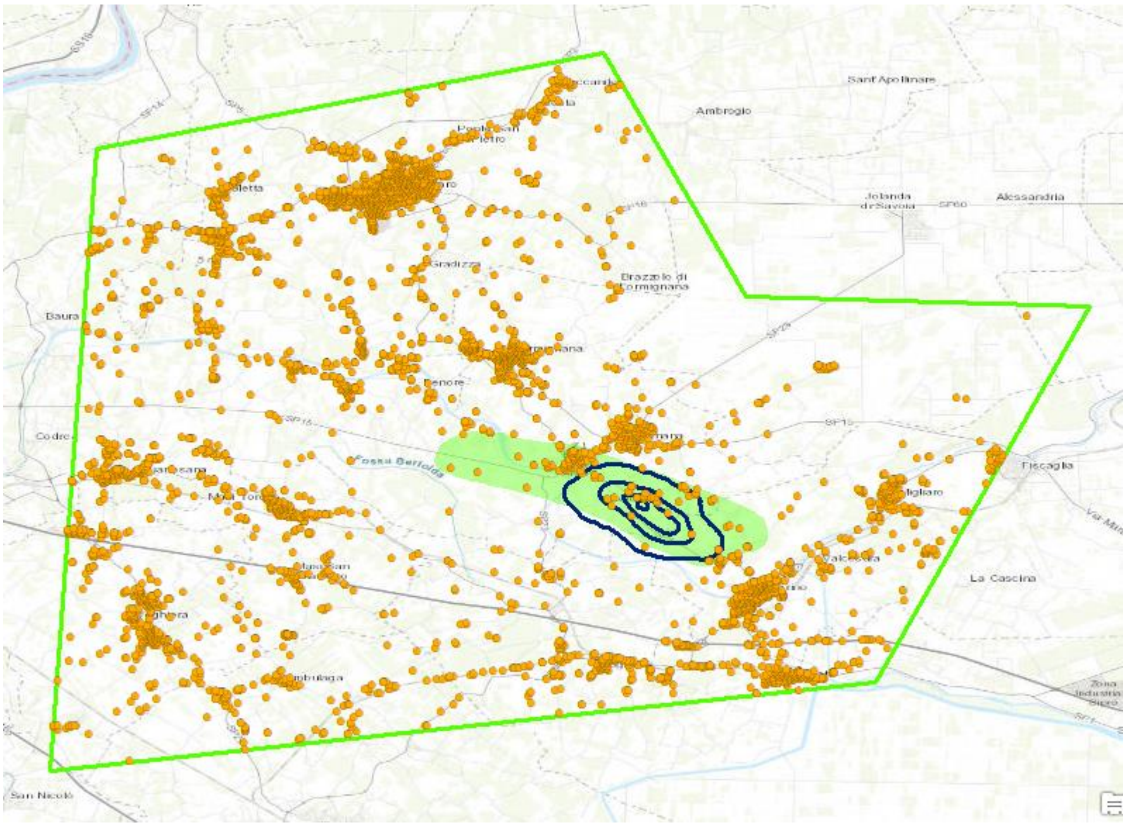


Fig. 55 - CLASS 1 Regional X-Field, 4 Classes Hierarchical Clustering, seasonal component. Author.

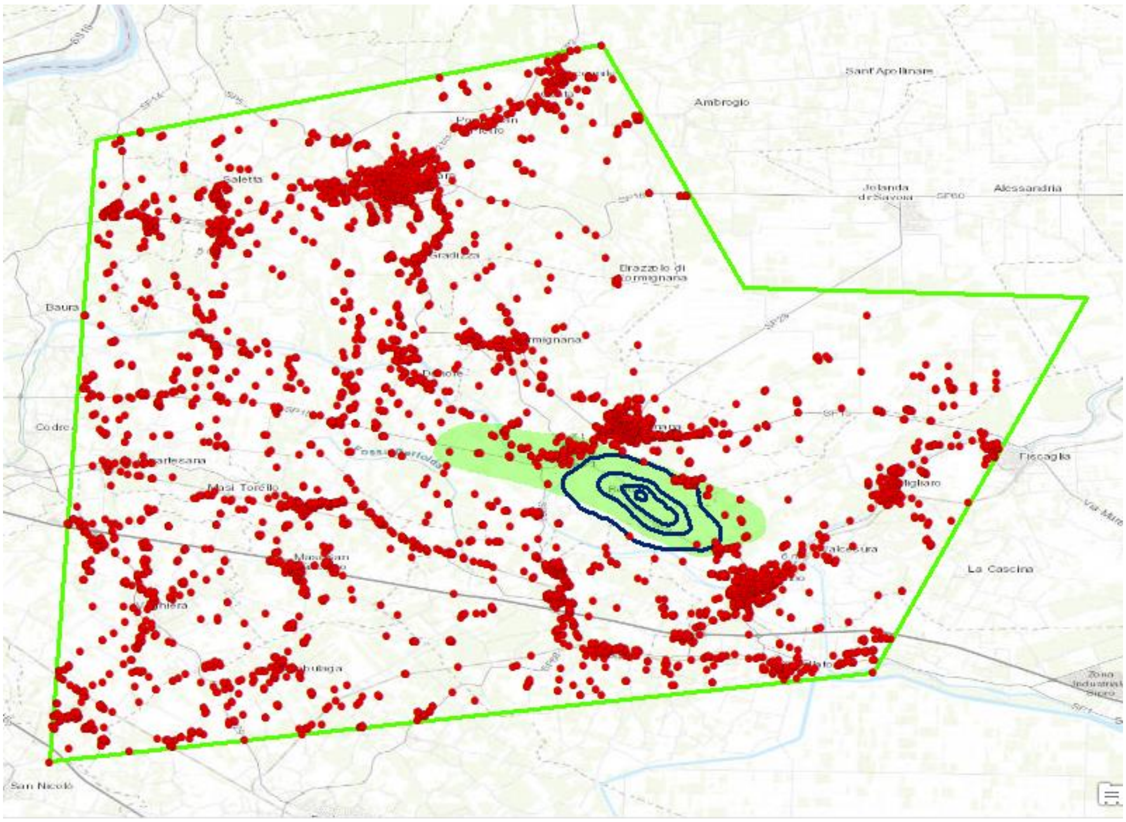


Fig. 56 - CLASS 2 Regional X-Field, 4 Classes Hierarchical Clustering, seasonal component. Author.

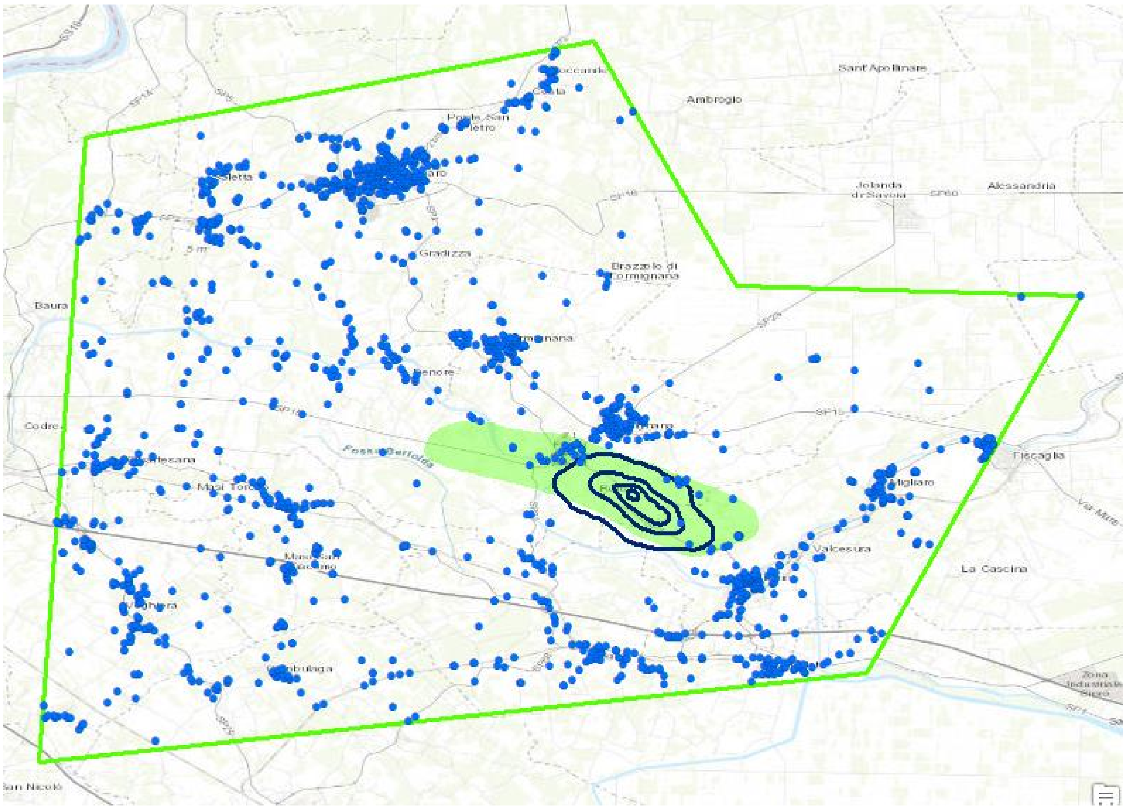


Fig. 57 - CLASS 3 Regional X-Field, 4 Classes Hierarchical Clustering, seasonal component. Author.

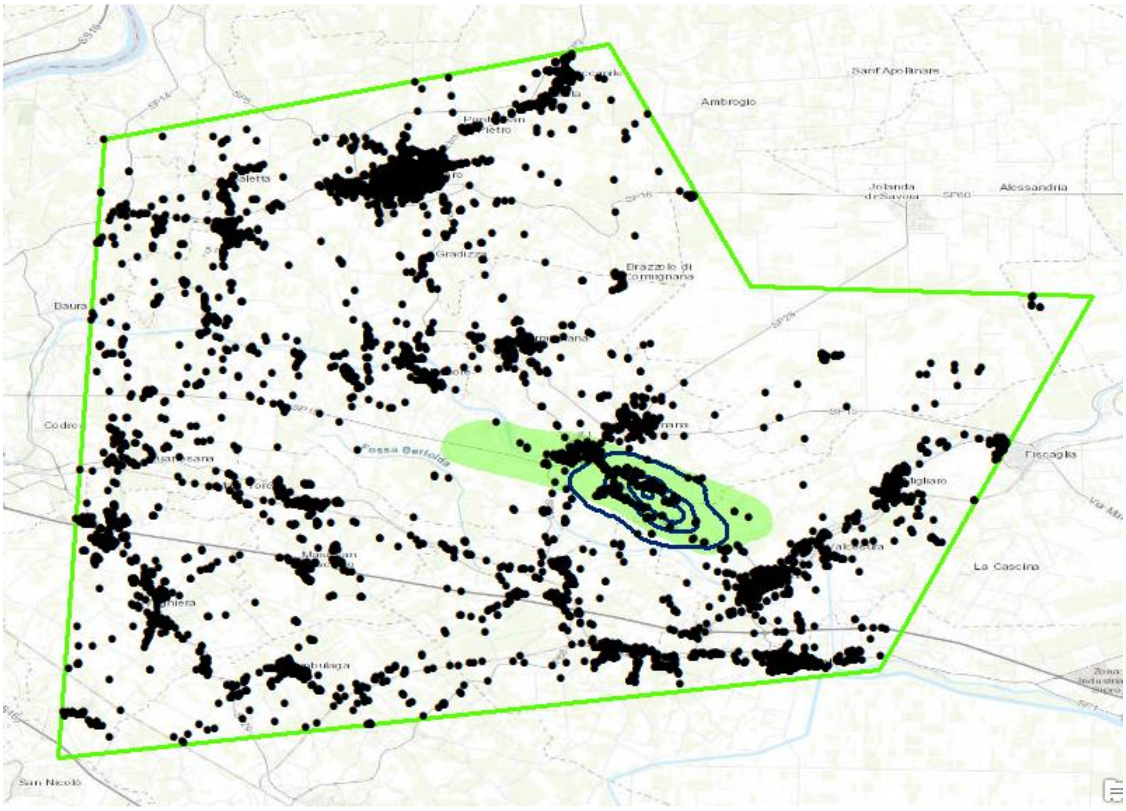


Fig. 58 - CLASS 4 Regional X-Field, 4 Classes Hierarchical Clustering, seasonal component. Author.

GIS visualization Kmean clustering

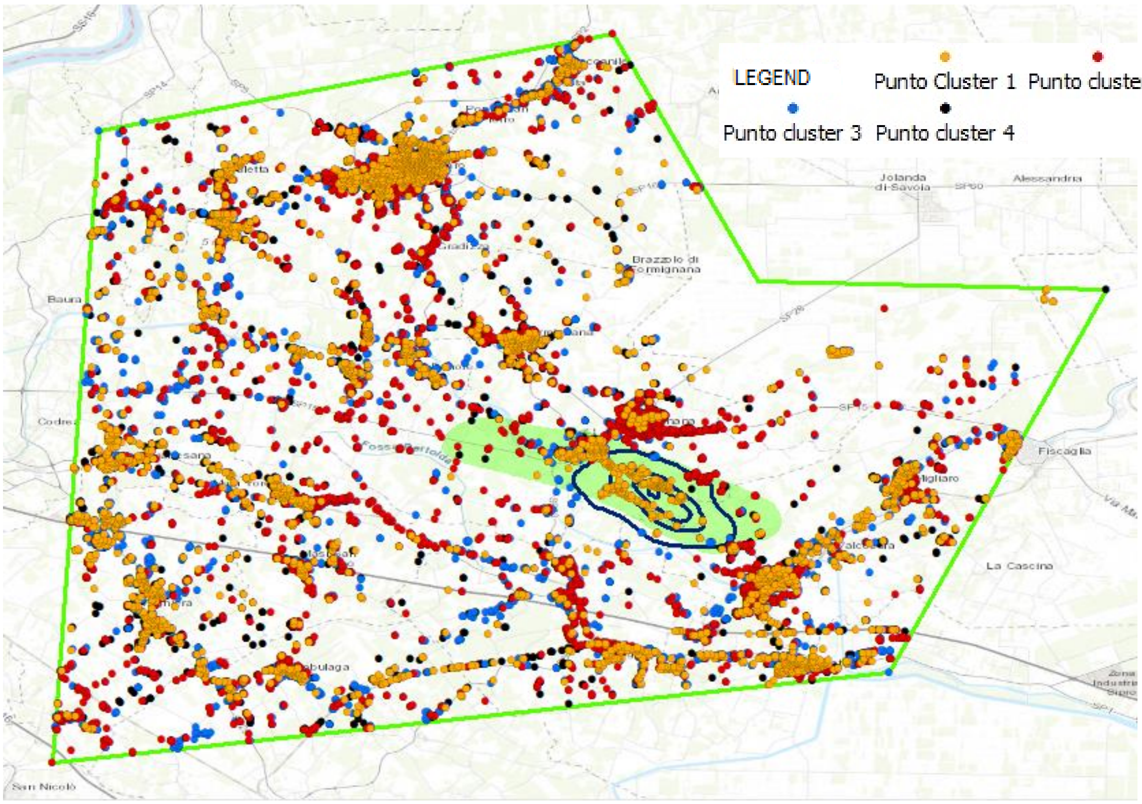


Fig. 59 - Regional X-Field, 4 Classes Kmean Clustering, seasonal component. Author.

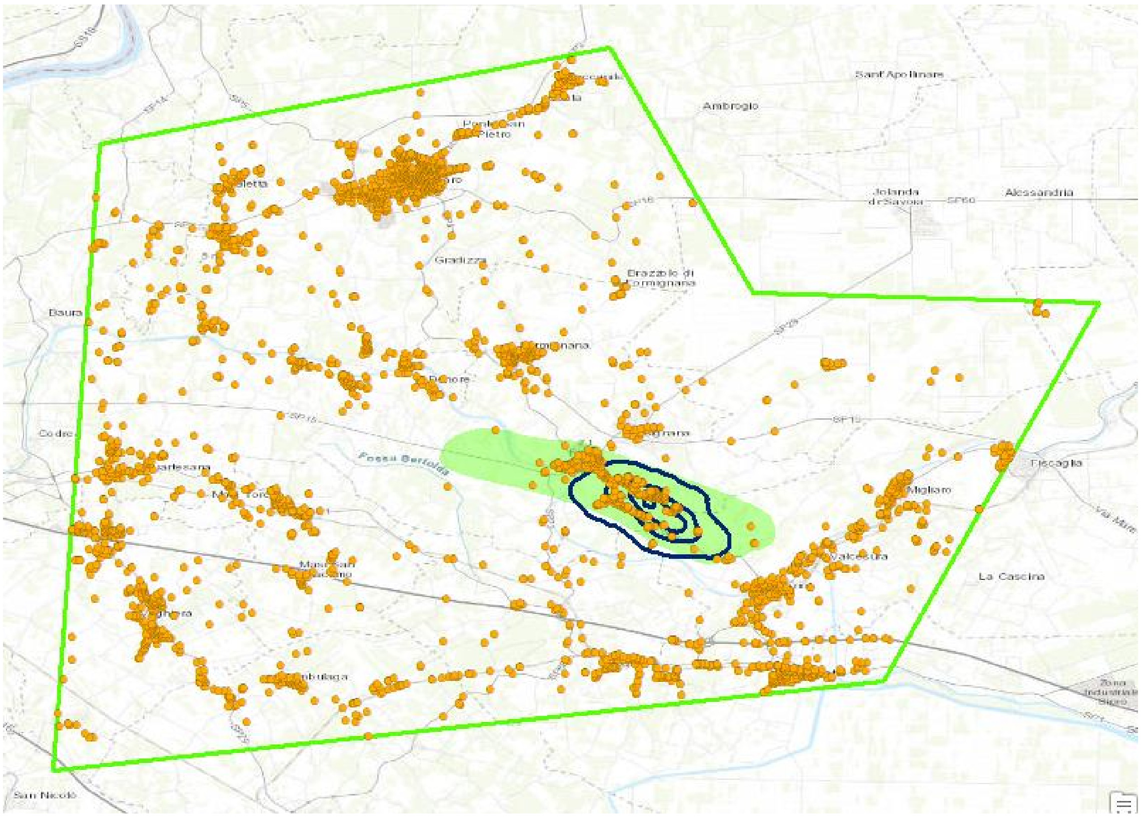


Fig. 60 - CLASS 1 Regional X-Field, 4 Classes Kmean Clustering, seasonal component. Author.

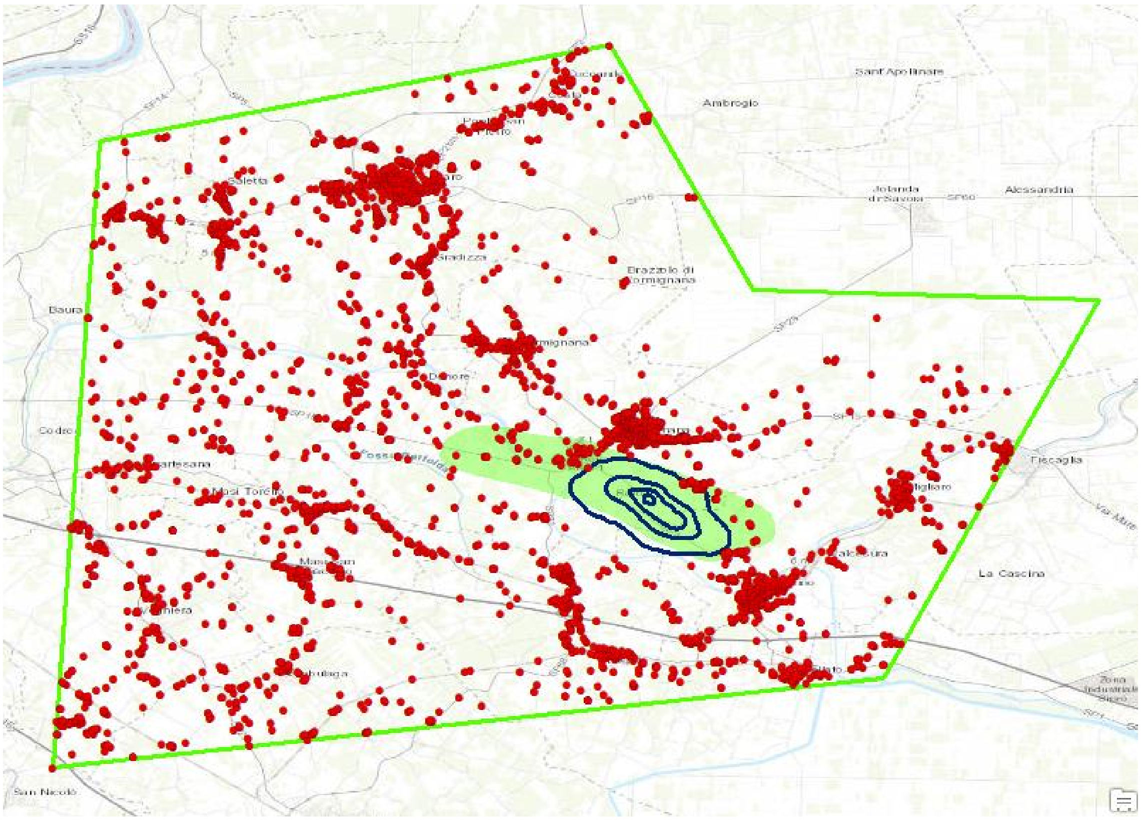


Fig. 61 - CLASS 2 Regional X-Field, 4 Classes Kmean Clustering, seasonal component. Author.

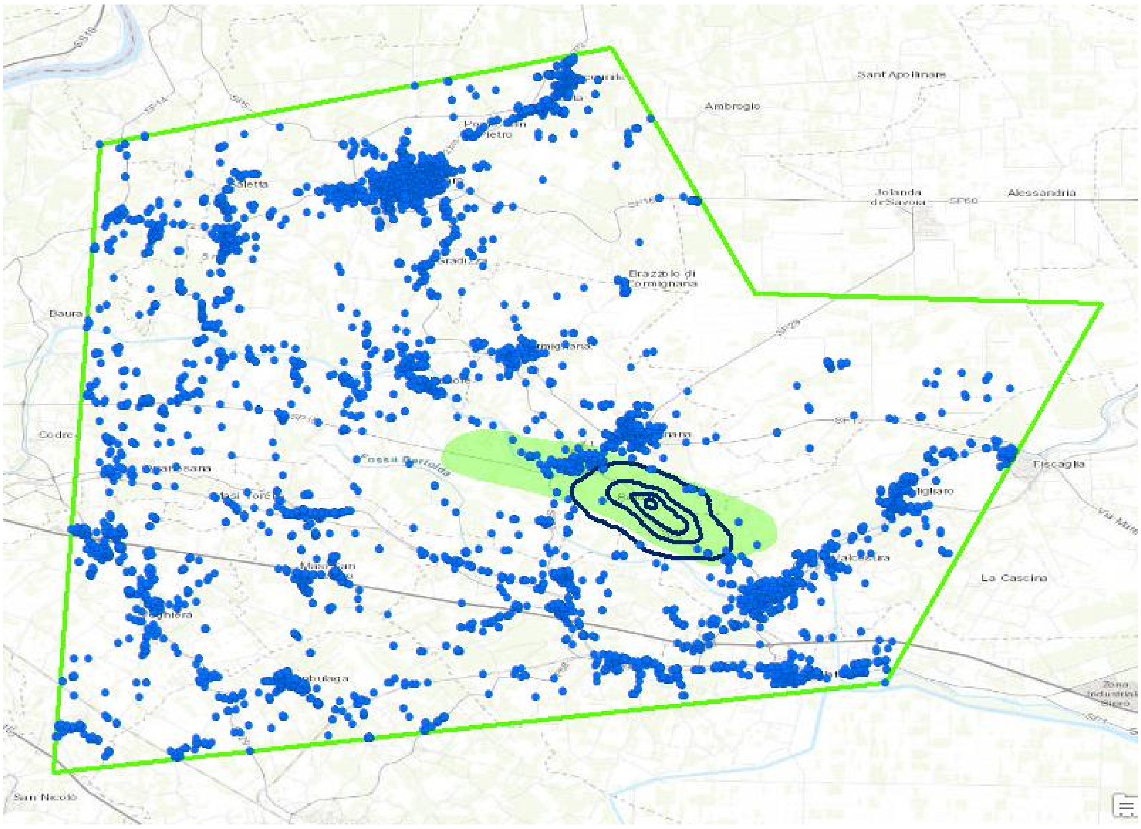


Fig. 62 - CLASS 3 Regional X-Field, 4 Classes Kmean Clustering, seasonal component. Author.

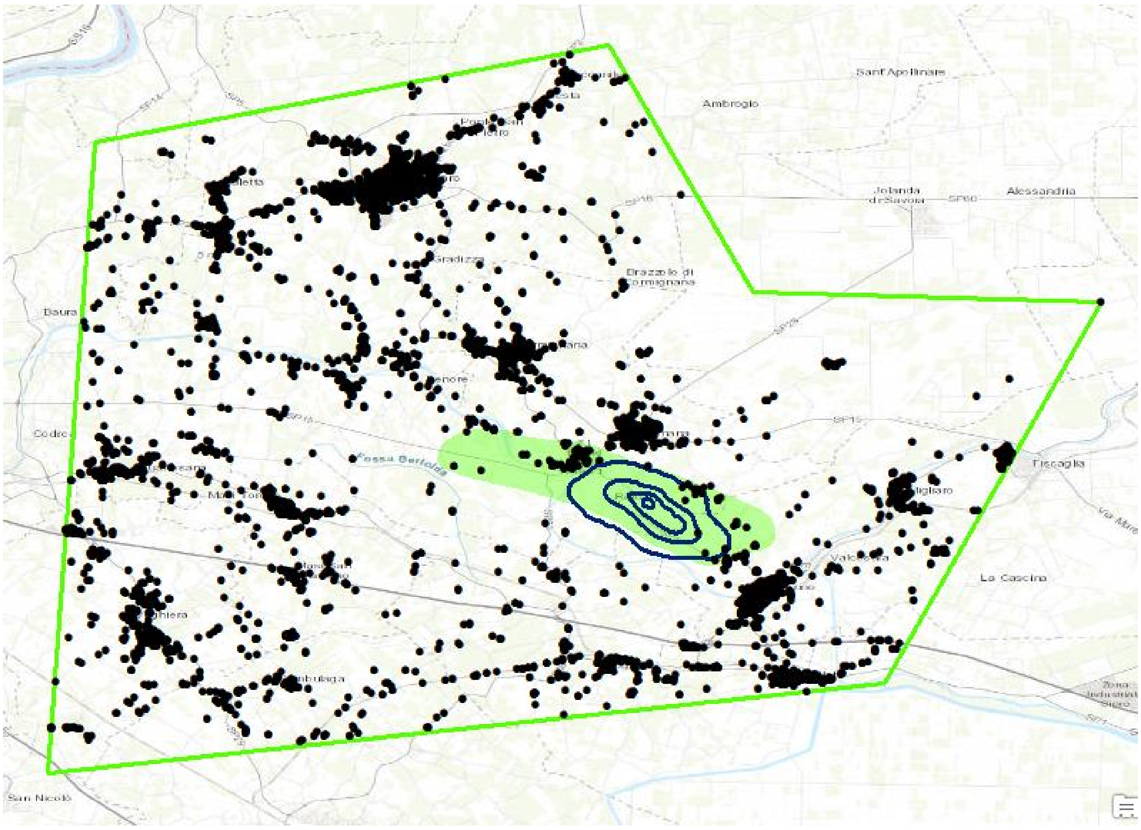


Fig. 63 - CLASS 4 Regional X-Field, 4 Classes Kmean Clustering, seasonal component. Author.

6.1.2 8 Classes clustering results (z-normalized values plotted):

Seasonal component time series clustering charts

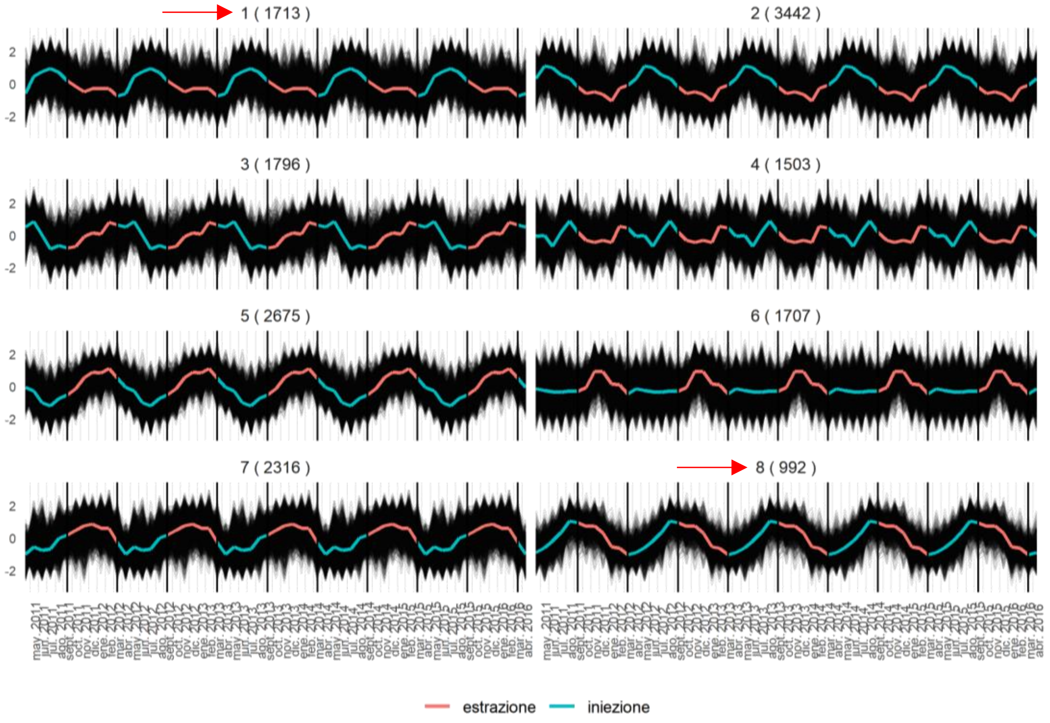


Fig. 64 - Regional Level (16.144 MP) - Hierarchical Clustering (8 Classes) seasonal component, z-normalized values clustered and plotted. Author.

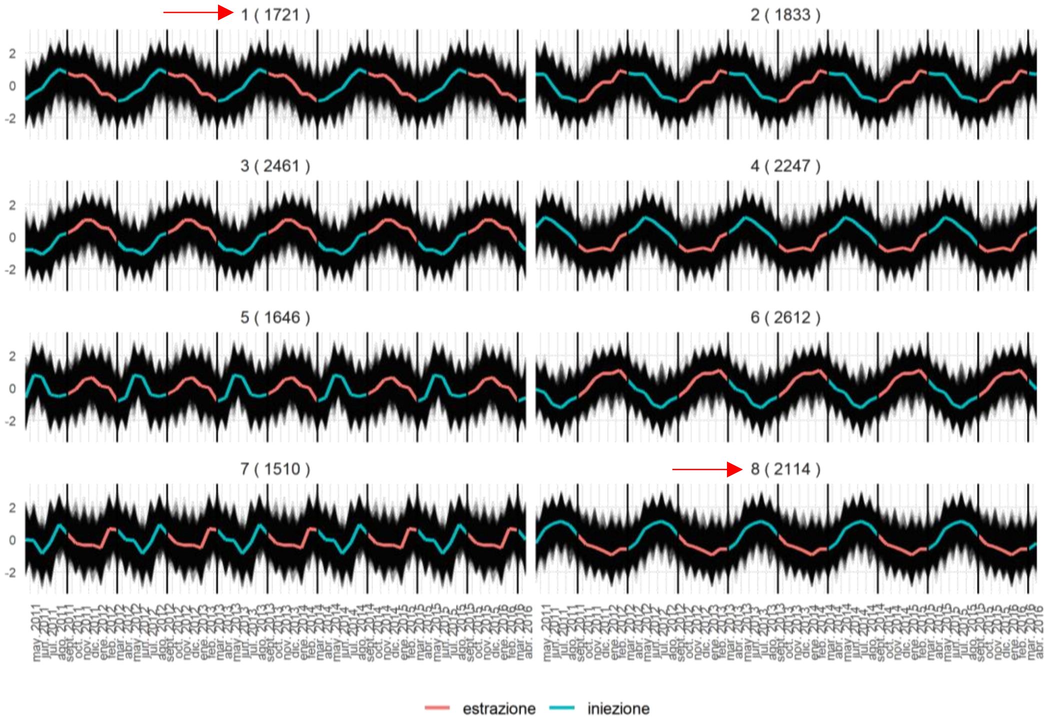


Fig. 65 - Regional Level (16.144 MP) - Kmean Clustering (8 Classes) seasonal component, z-normalized values clustered and plotted. Author.

MP distribution pie chart by clustering type

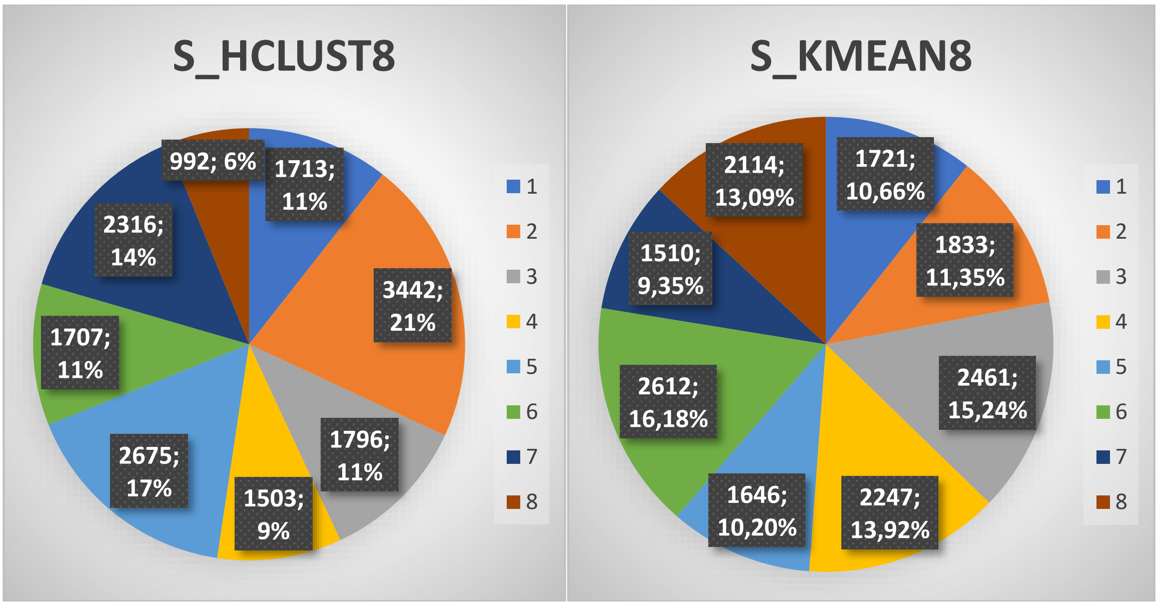


Fig. 66 - Regional X-Field, 8 Classes Hierarchical and Kmean Clustering (normalized data), seasonal component, MP distribution and relative percentages for each class. Author.

GIS visualization Hierarchical clustering

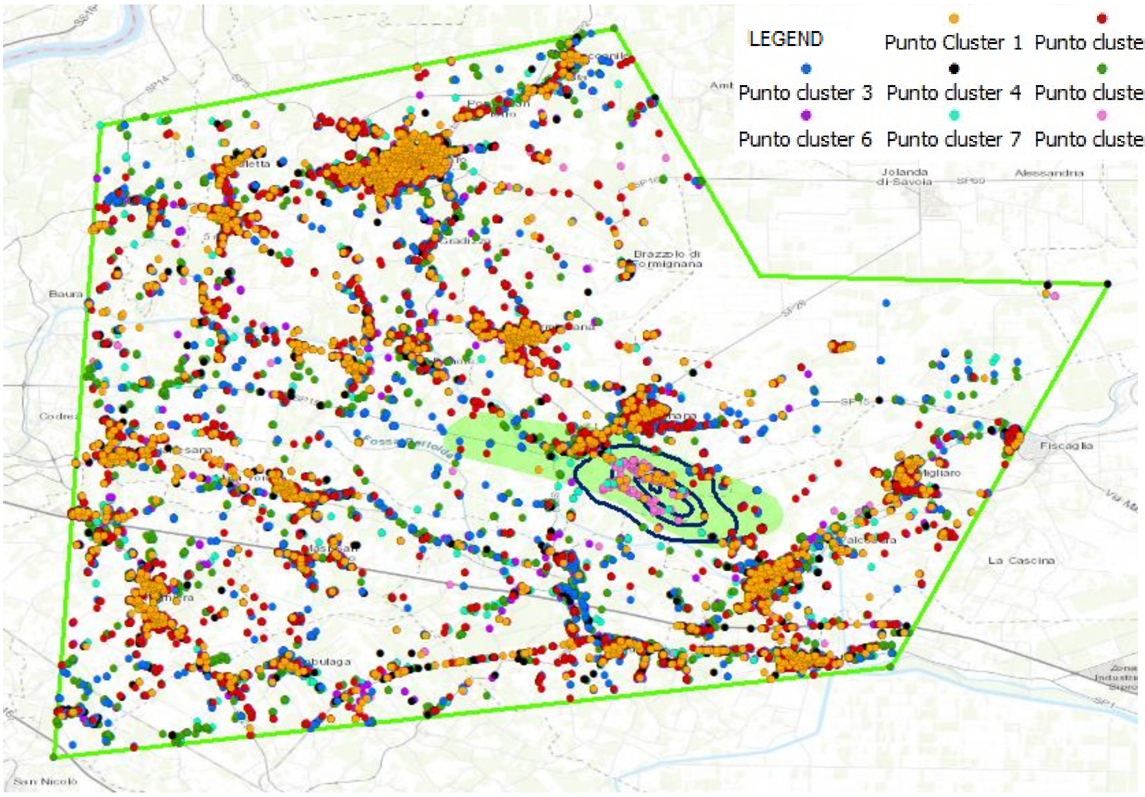


Fig. 67 - Regional X-Field, 8 Classes Hierarchical Clustering, seasonal component. Author.

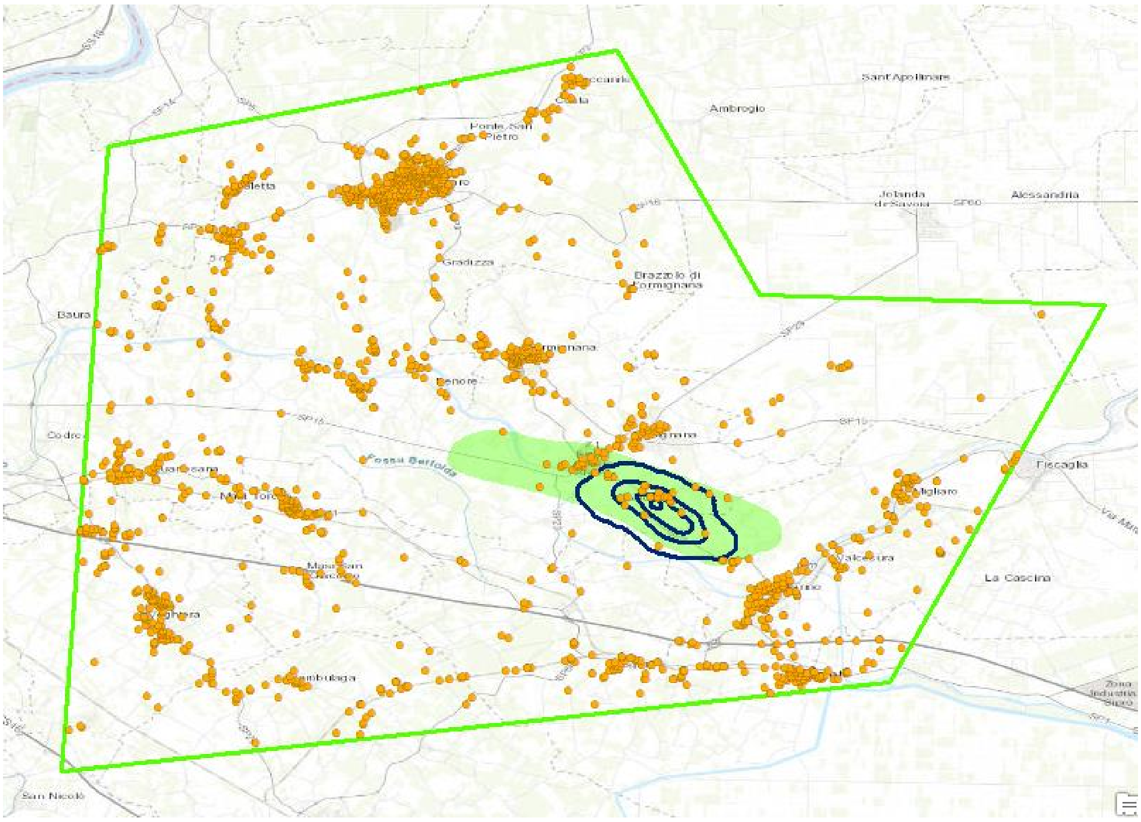


Fig. 68 - CLASS 1 Regional X-Field, 8 Classes Hierarchical Clustering, seasonal component. Author.

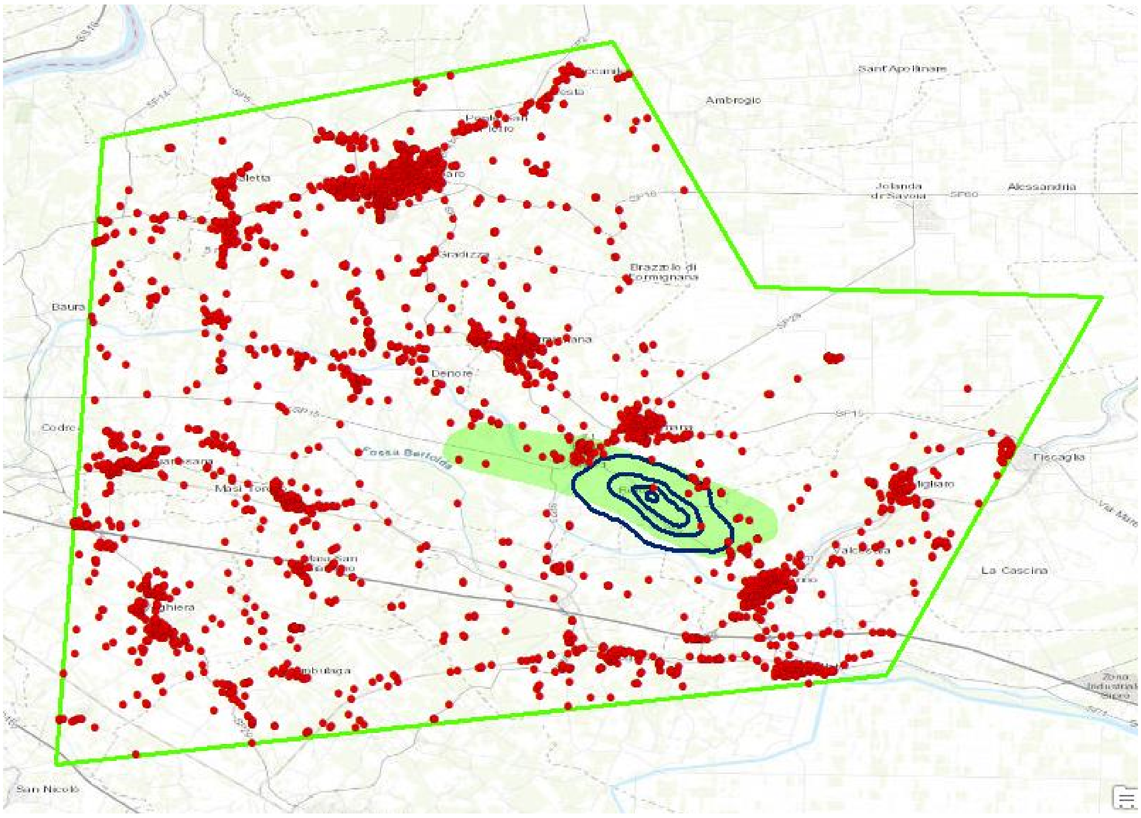


Fig. 69 - CLASS 2 Regional X-Field, 8 Classes Hierarchical Clustering, seasonal component. Author.

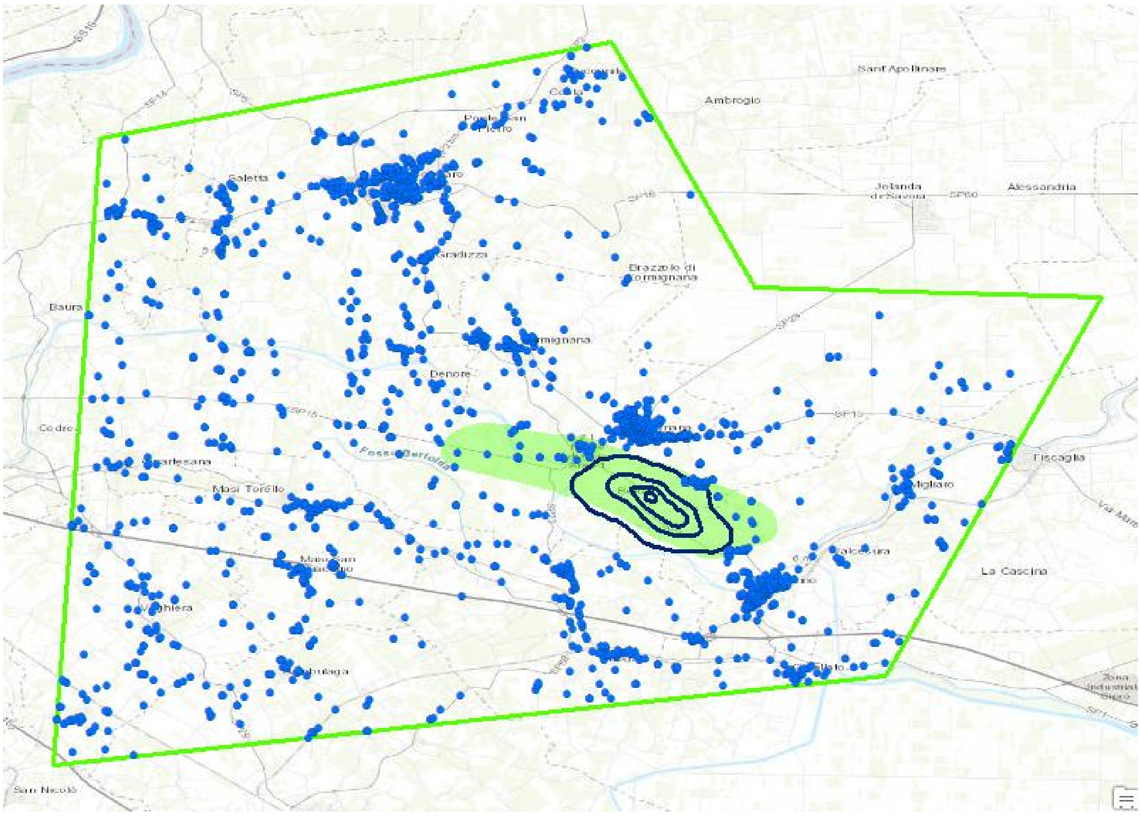


Fig. 70 - CLASS 3 Regional X-Field, 8 Classes Hierarchical Clustering, seasonal component. Author.

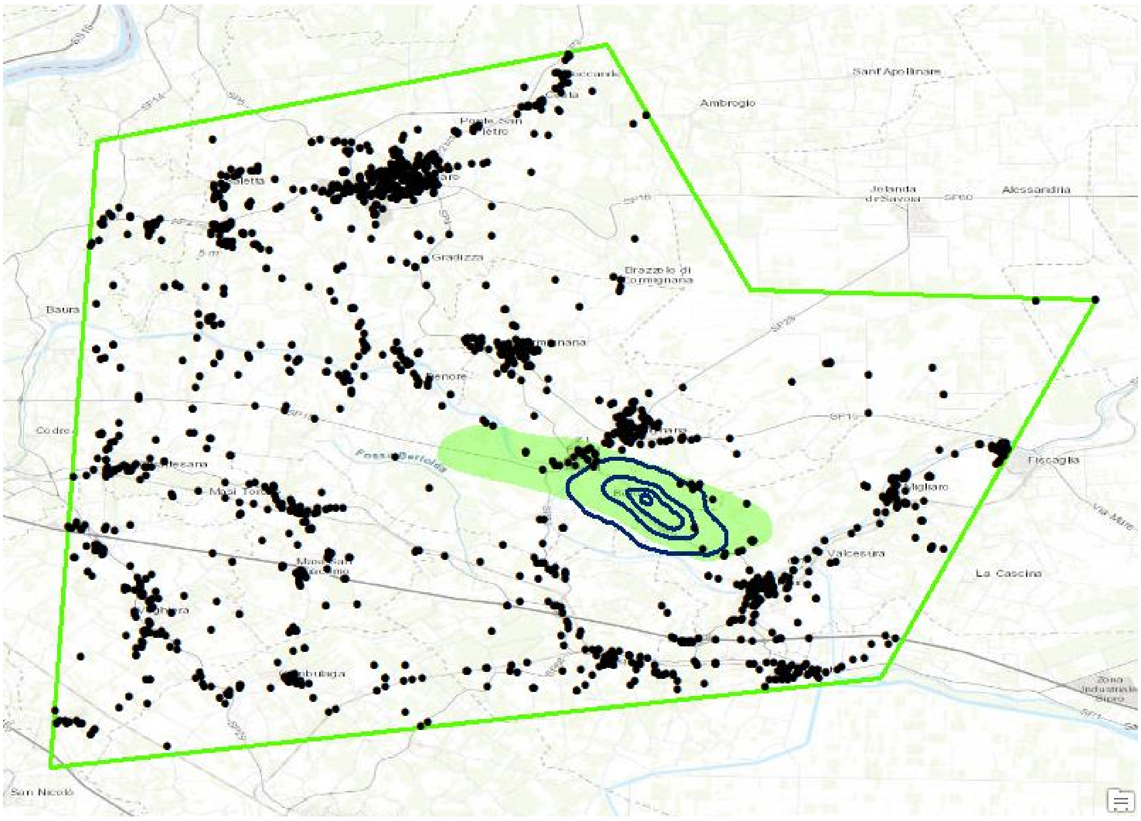


Fig. 71 - CLASS 4 Regional X-Field, 8 Classes Hierarchical Clustering, seasonal component. Author.

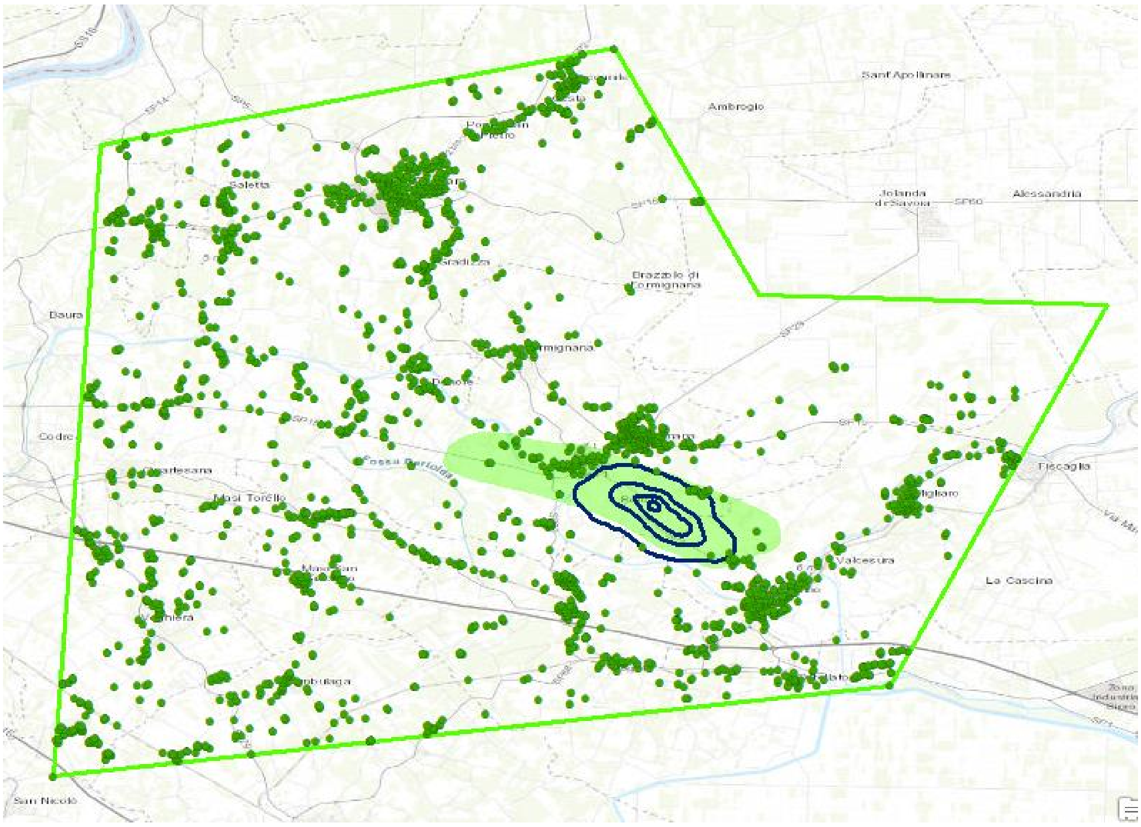


Fig. 72 - CLASS 5 Regional X-Field, 8 Classes Hierarchical Clustering, seasonal component. Author.

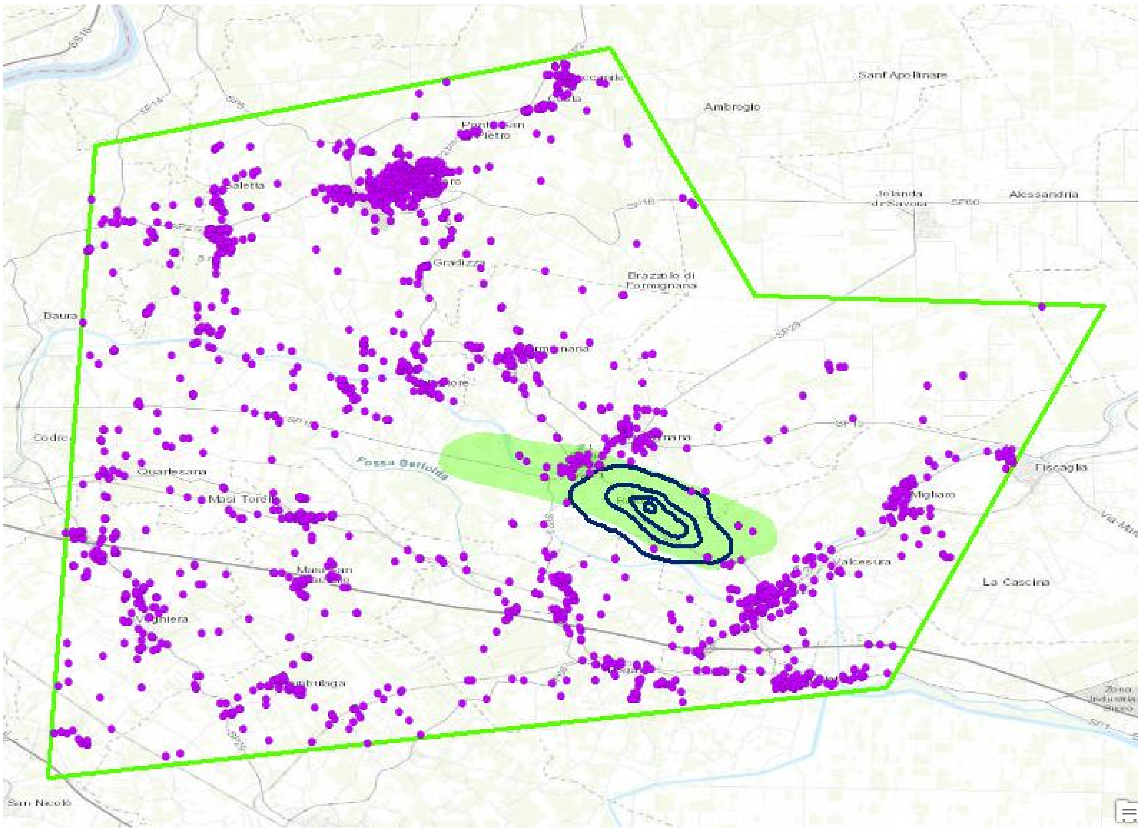


Fig. 73 - CLASS 6 Regional X-Field, 8 Classes Hierarchical Clustering, seasonal component. Author.

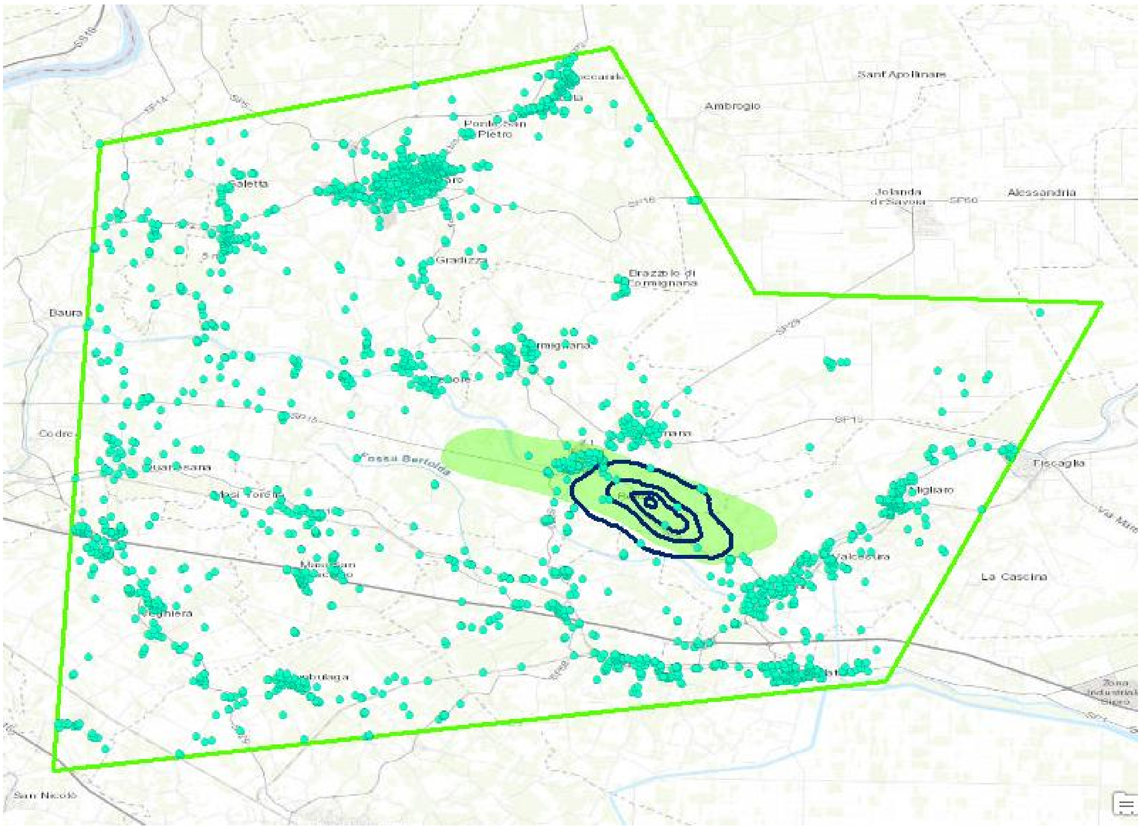


Fig. 74 - CLASS 7 Regional X-Field, 8 Classes Hierarchical Clustering, seasonal component. Author.

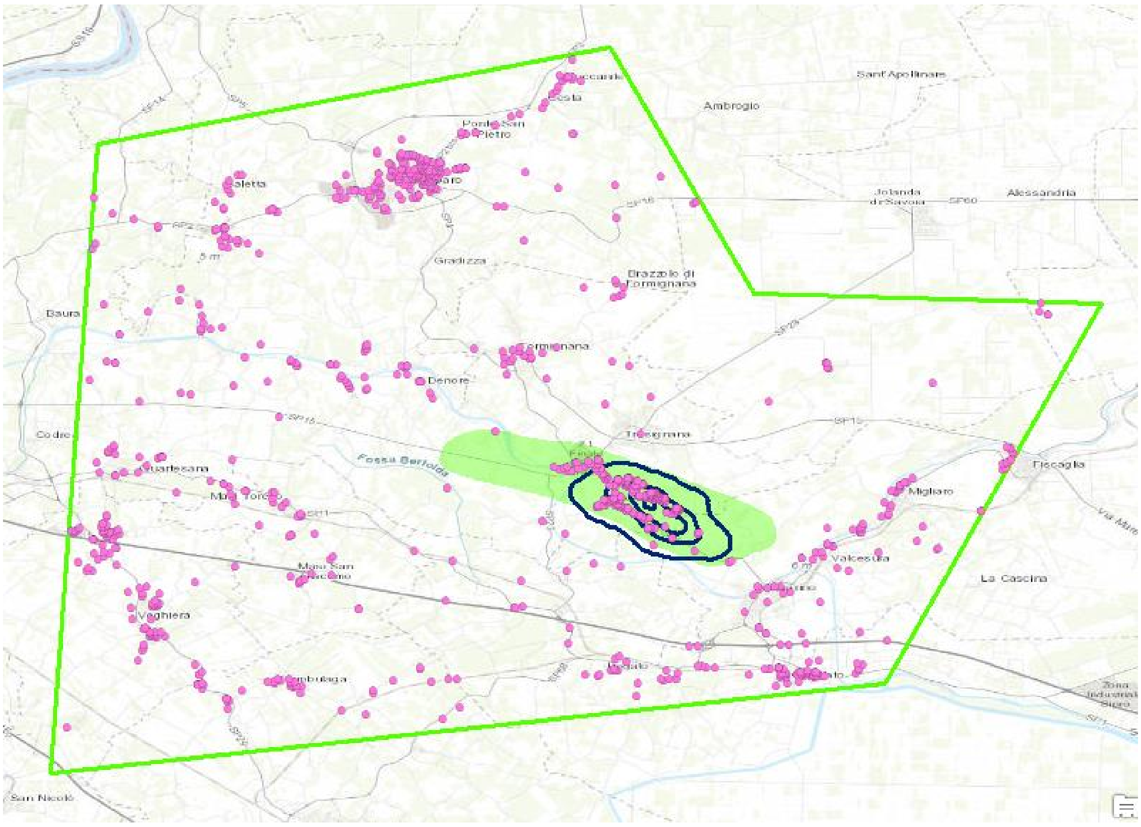


Fig. 75 - CLASS 8 Regional X-Field, 8 Classes Hierarchical Clustering, seasonal component. Author.

GIS visualization Kmean clustering

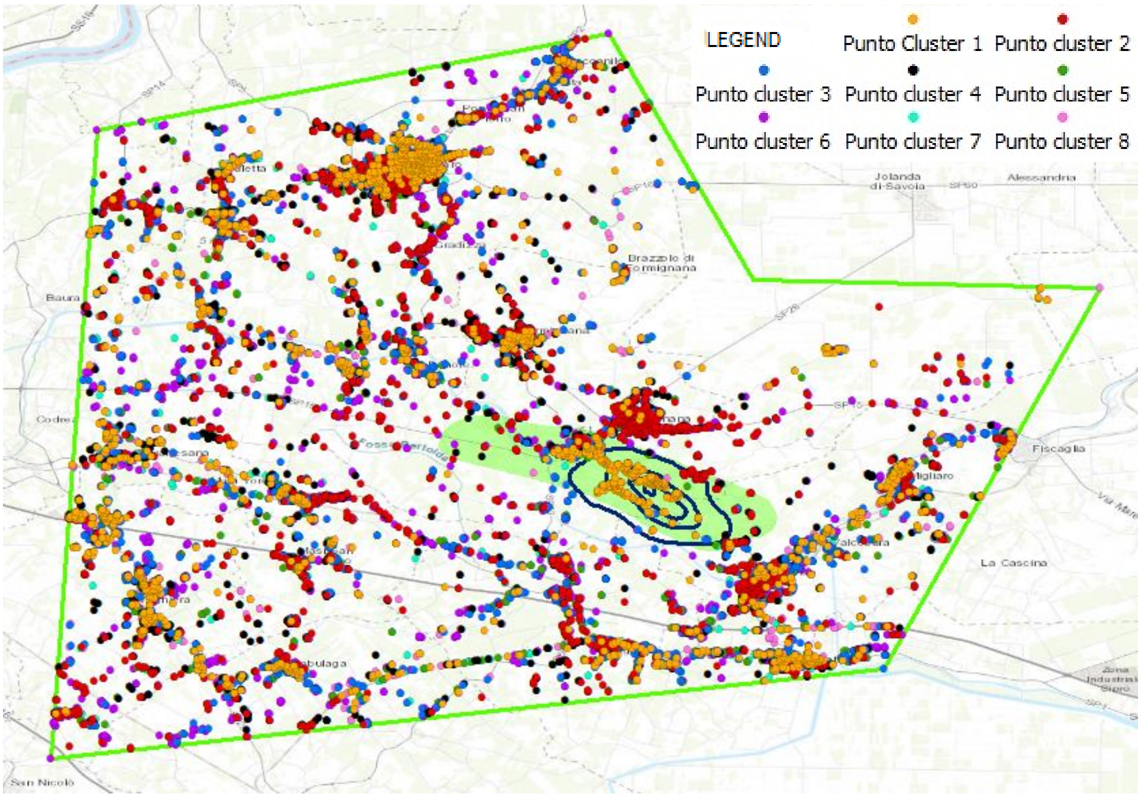


Fig. 76 - Regional X-Field, 8 Classes Kmean Clustering, seasonal component. Author.

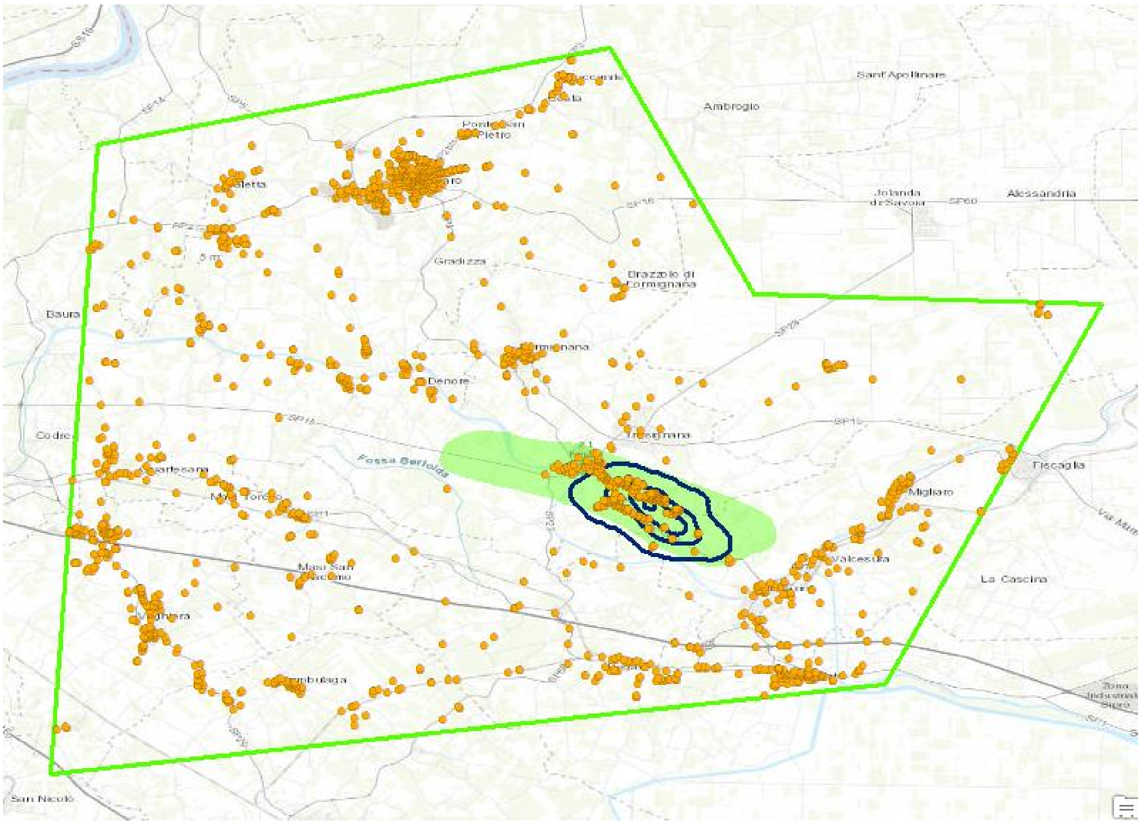


Fig. 77 - CLASS 1 Regional X-Field, 8 Classes Kmean Clustering, seasonal component. Author.

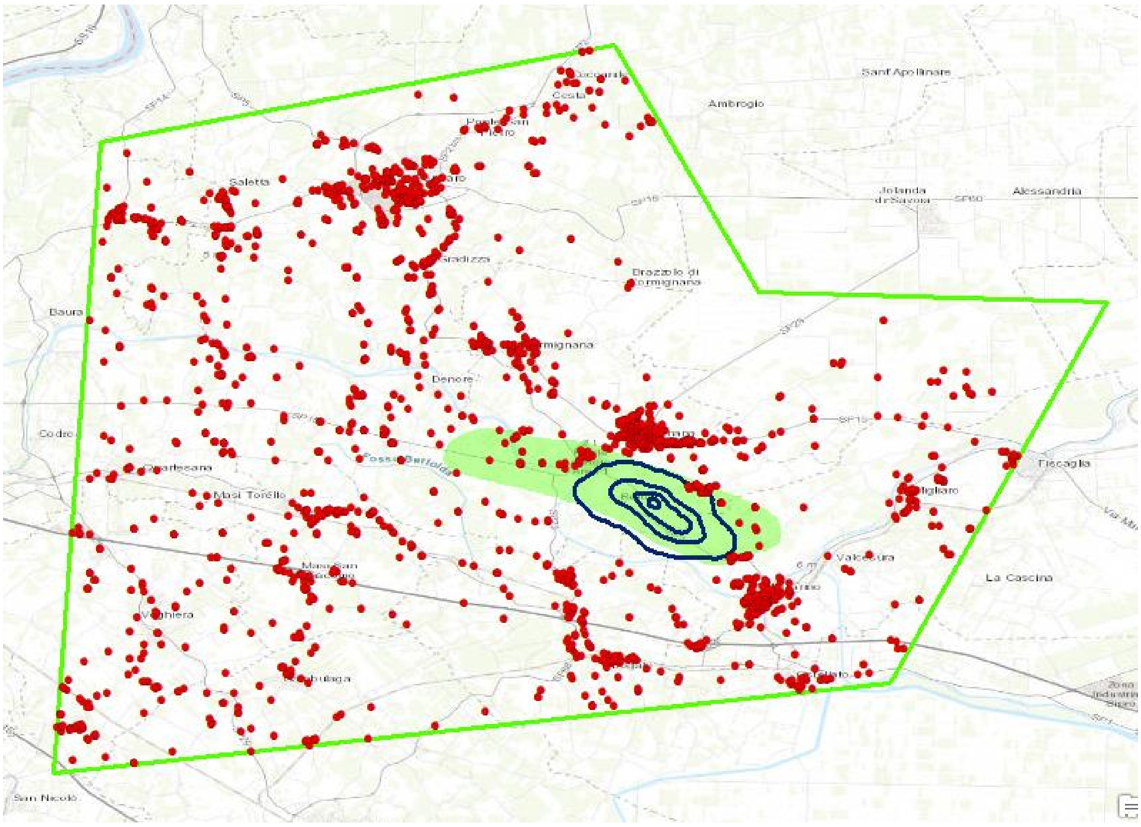


Fig. 78 - CLASS 2 Regional X-Field, 8 Classes Kmean Clustering, seasonal component. Author.

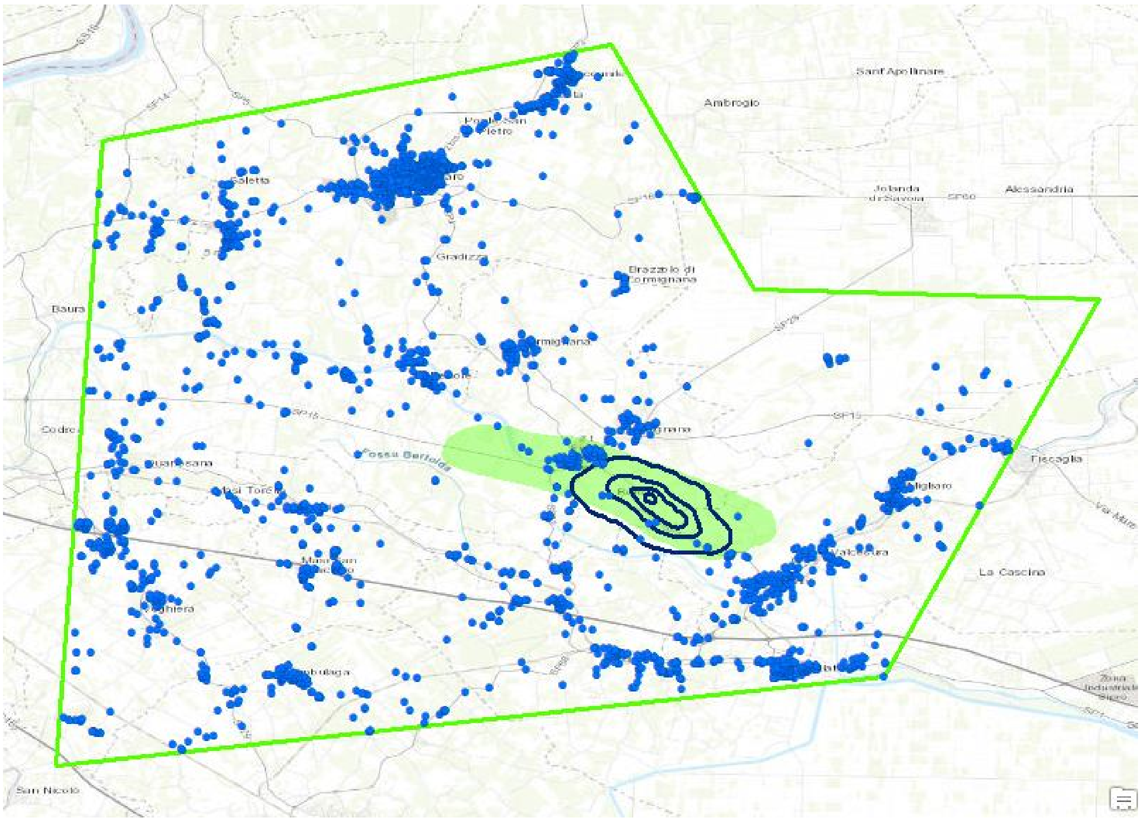


Fig. 79 - CLASS 3 Regional X-Field, 8 Classes Kmean Clustering, seasonal component. Author.

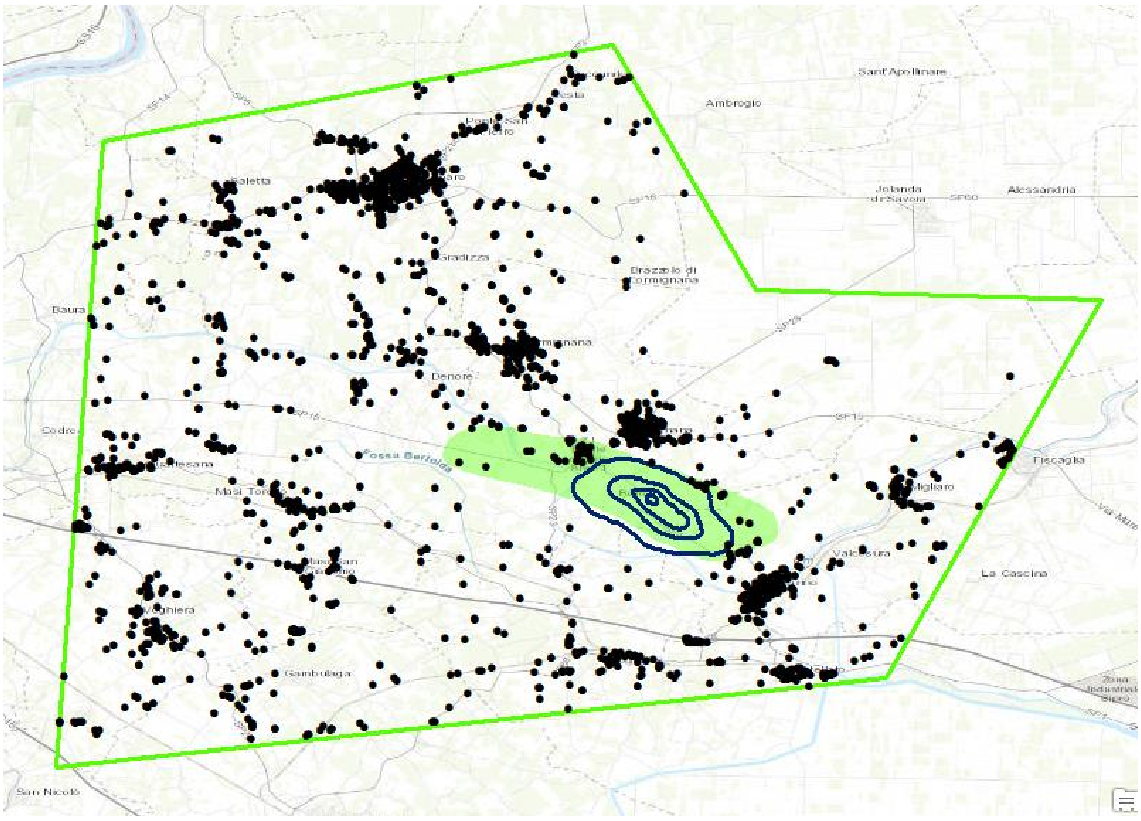


Fig. 80 - CLASS 4 Regional X-Field, 8 Classes Kmean Clustering, seasonal component. Author.

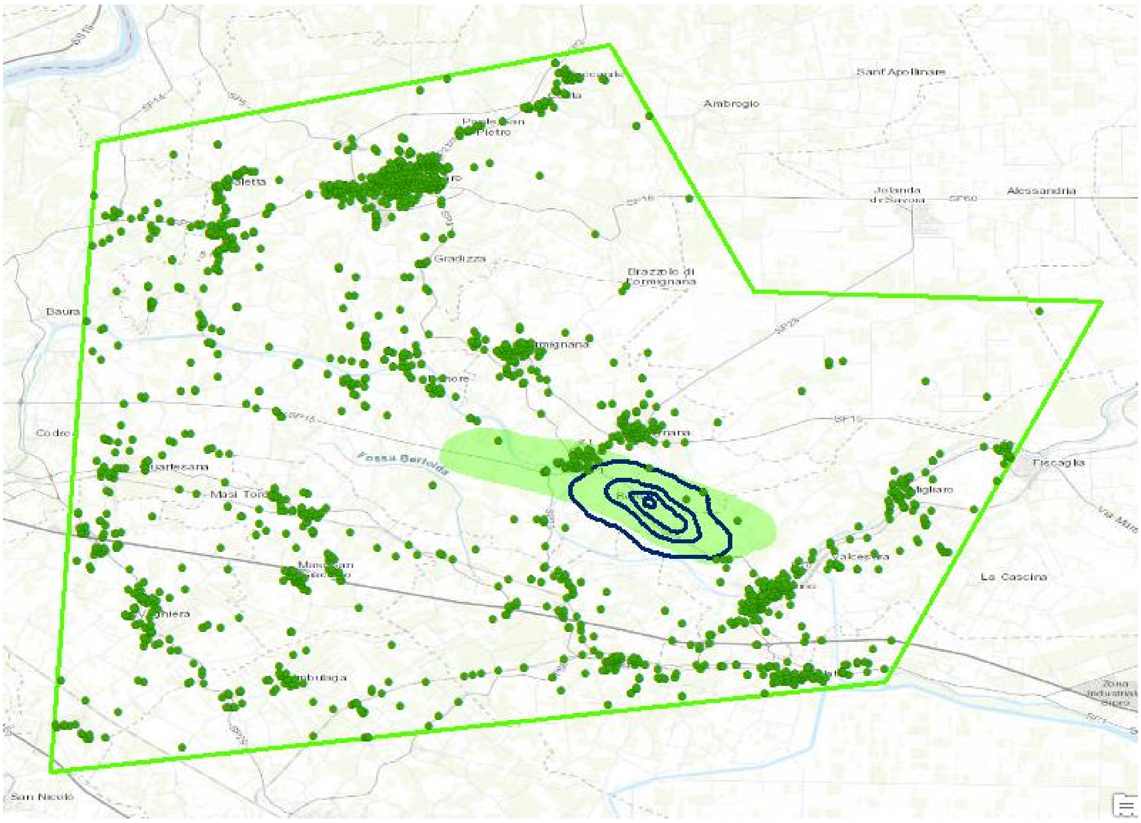


Fig. 81 - CLASS 5 Regional X-Field, 8 Classes Kmean Clustering, seasonal component. Author.

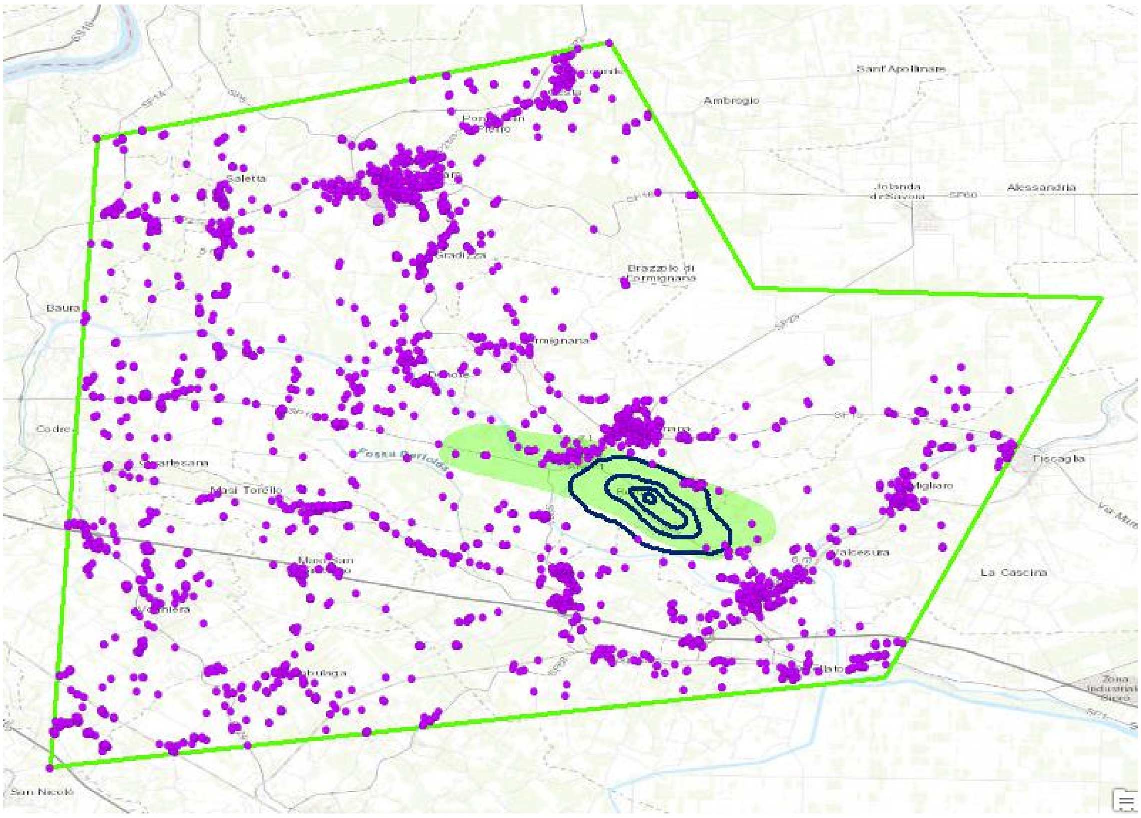


Fig. 82 - CLASS 6 Regional X-Field, 8 Classes Kmean Clustering, seasonal component. Author.

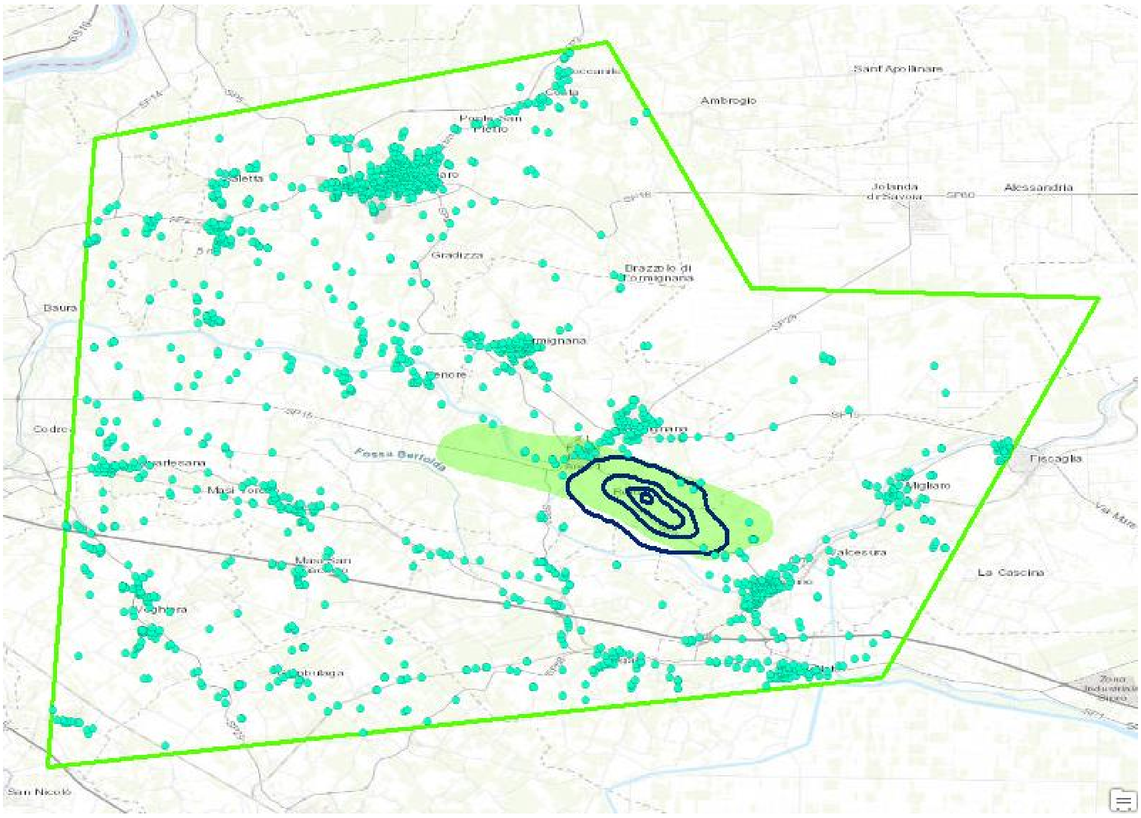


Fig. 83 - CLASS 7 Regional X-Field, 8 Classes Kmean Clustering, seasonal component. Author.

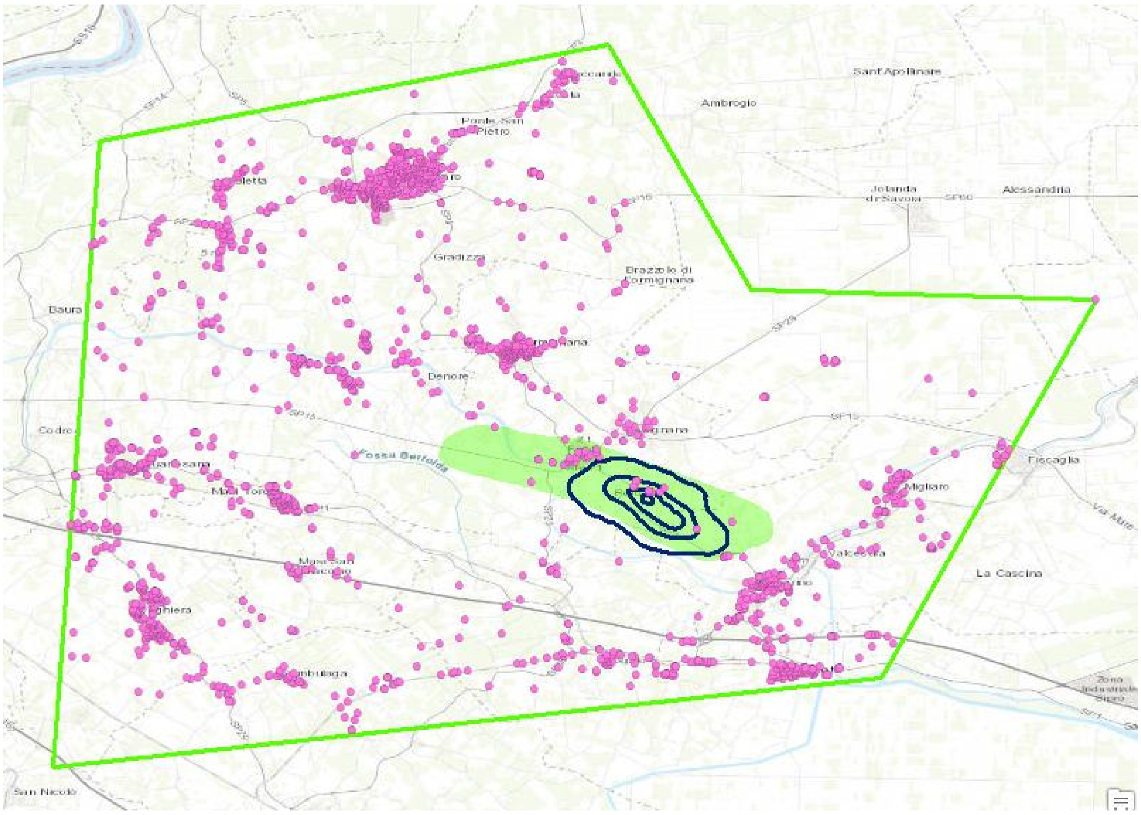


Fig. 84 - CLASS 8 Regional X-Field, 8 Classes Kmean Clustering, seasonal component. Author.

6.2 Local level Seasonal analysis

6.2.1 4 Classes clustering results (z-normalized values plotted):

Seasonal component time series clustering charts

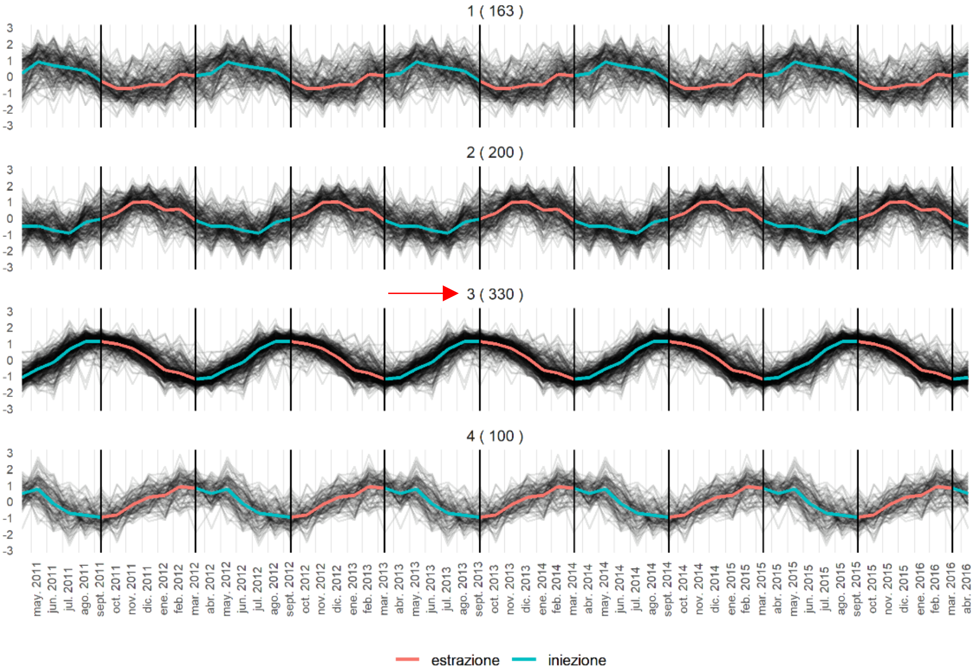


Fig. 85 - Local Level (793 MP) - Hierarchical Clustering (4 Classes) seasonal component, z-normalized values clustered and plotted. Author.

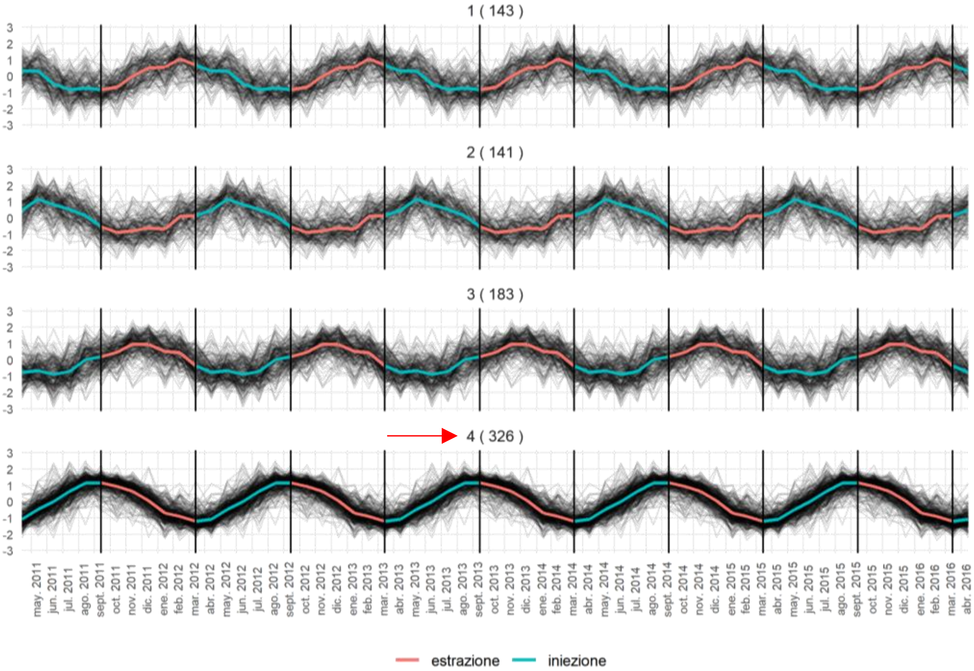


Fig. 86 - Local Level (793 MP) - Kmean Clustering (4 Classes) seasonal component, z-normalized values clustered and plotted. Author.

MP distribution pie charts by clustering type

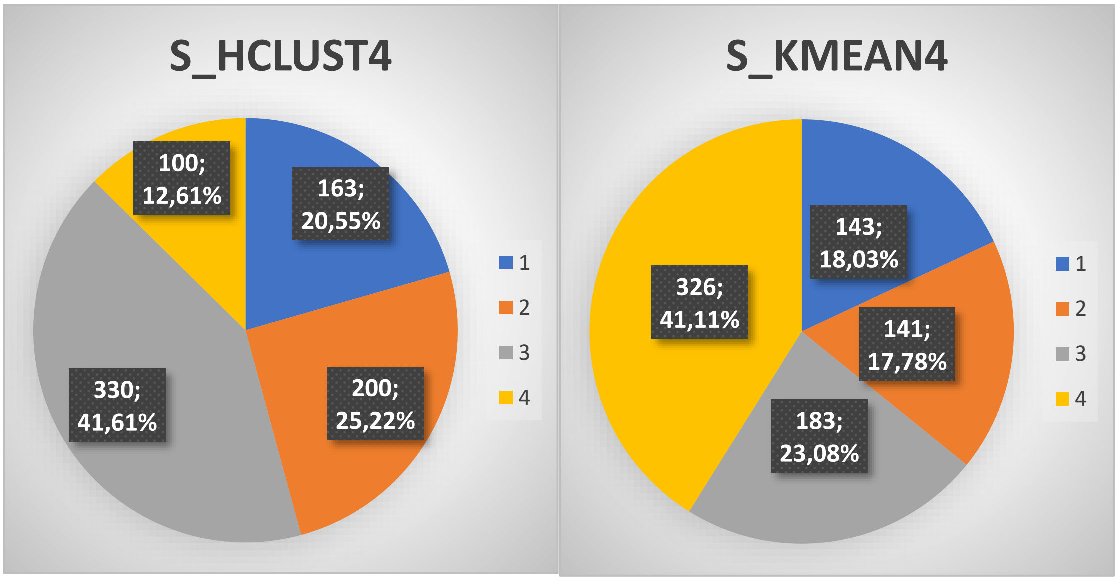


Fig. 87 - Local X-Field, 4 Classes Hierarchical and Kmean Clustering, seasonal component (normalized data), MP distribution and relative percentages for each class. Author.

GIS visualization Hierarchical clustering

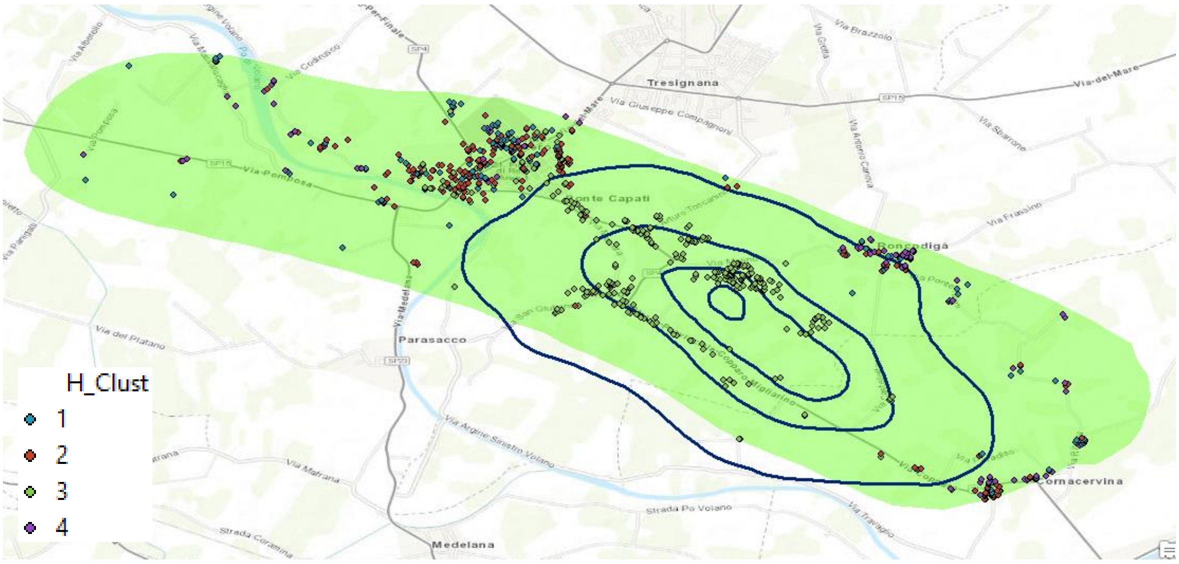


Fig. 88 - Local X-Field, 4 Classes Hierarchical Clustering, seasonal component. Author.

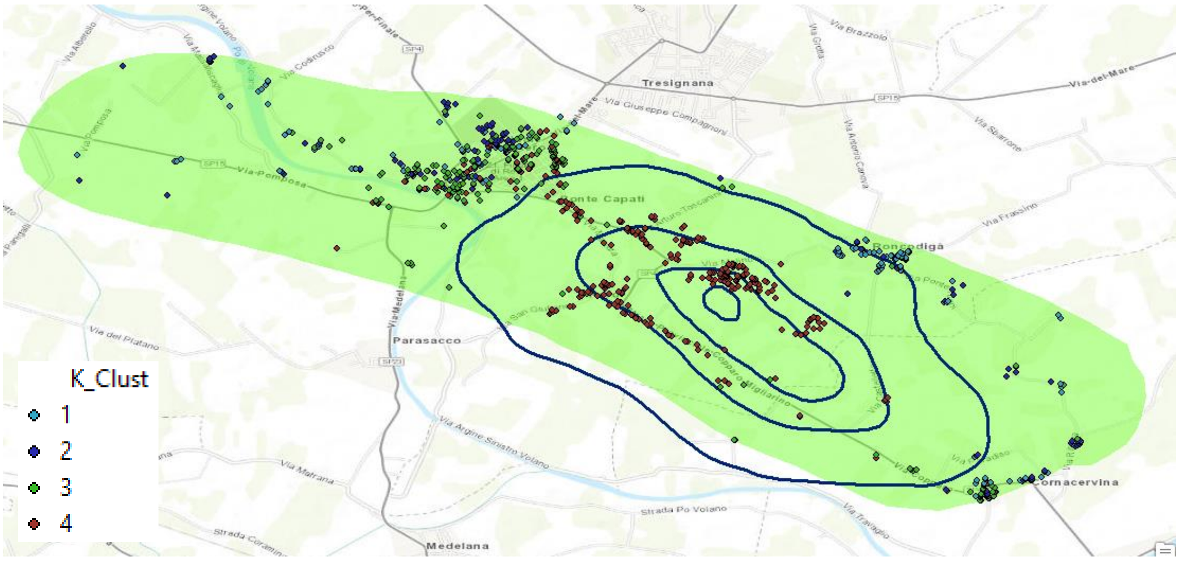


Fig. 89 - Local X-Field, 4 Classes Kmean Clustering, seasonal component. Author.

6.2.2 8 Classes clustering results (z-normalized values plotted):

Seasonal component time series clustering charts

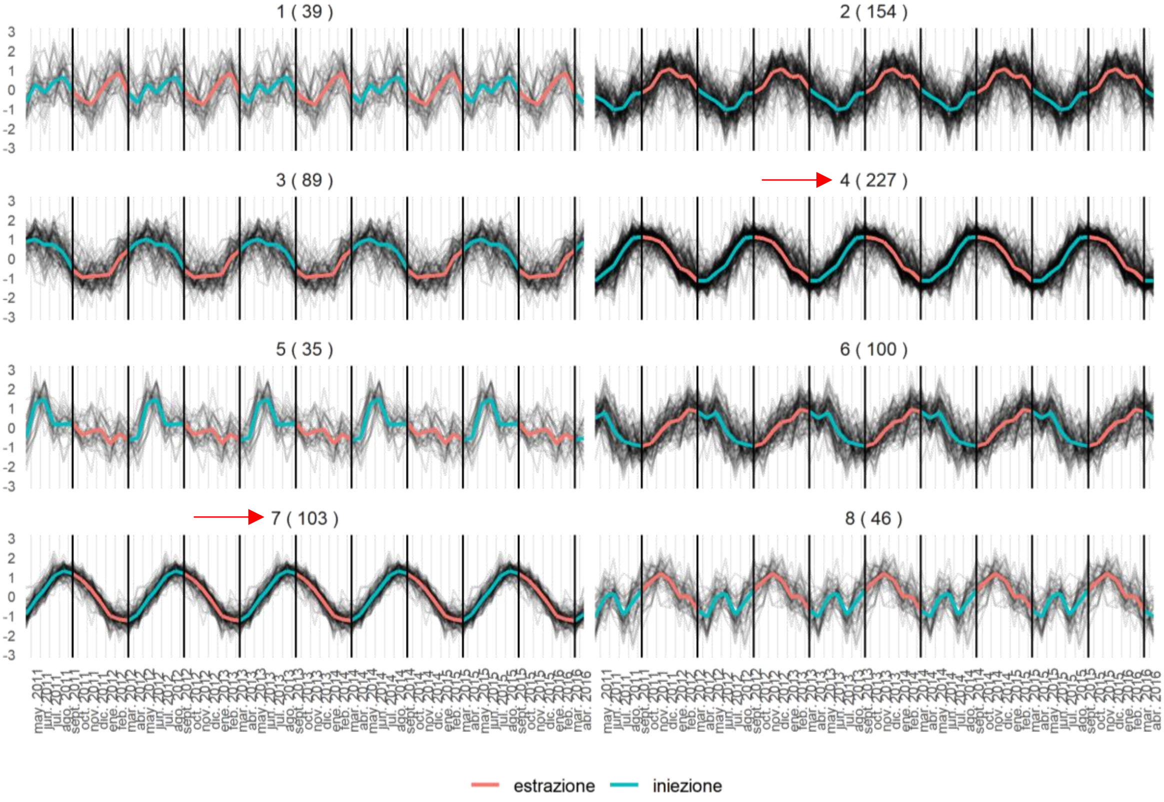


Fig. 90 - Local Level (793 MP) - Hierarchical Clustering (8 Classes) seasonal component, z-normalized values clustered and plotted. Author.

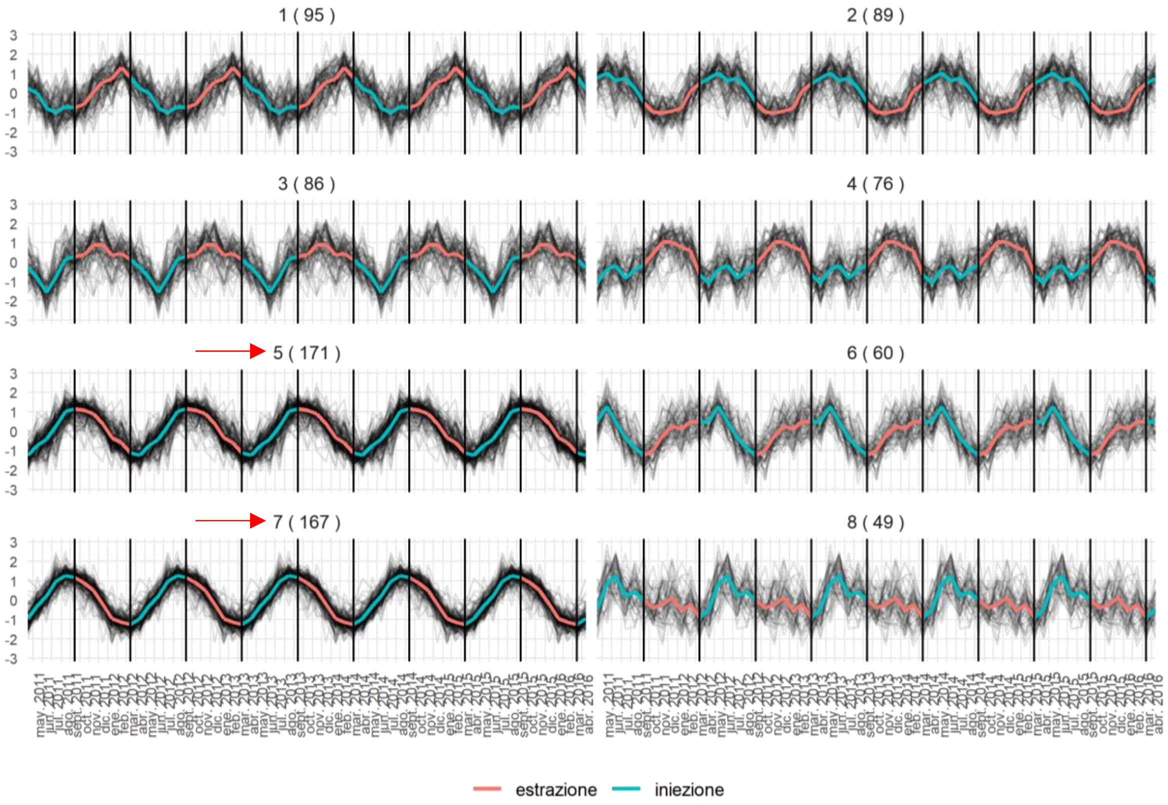


Fig. 91 - Local Level (793 MP) - Kmean Clustering (8 Classes) seasonal component, z-normalized values clustered and plotted. Author.

MP distribution pie charts by clustering type

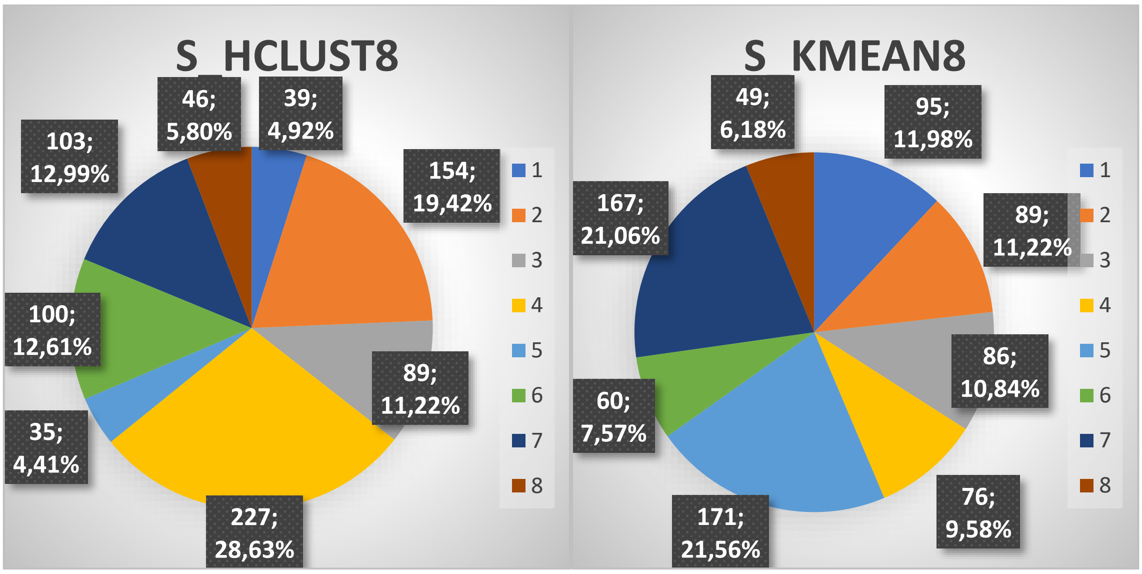


Fig. 92 - Local X-Field, 8 Classes Hierarchical and Kmean Clustering (normalized data), seasonal component, MP distribution and relative percentages for each class. Author.

GIS visualization Hierarchical clustering

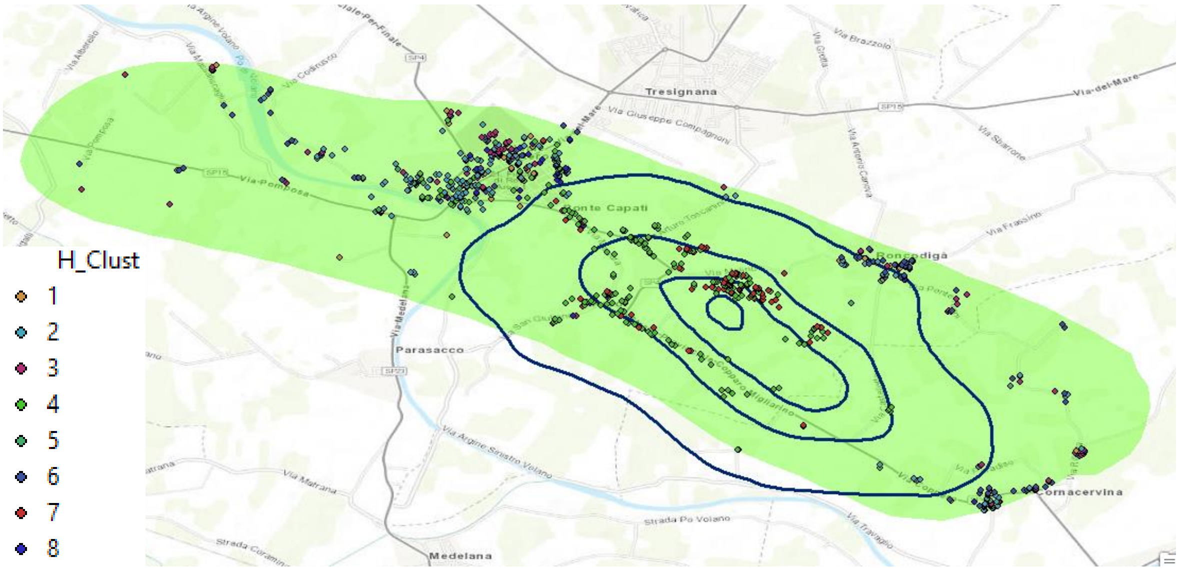


Fig. 93 - Local X-Field, 8 Classes Hierarchical Clustering, seasonal component. Author.

GIS visualization Kmean clustering

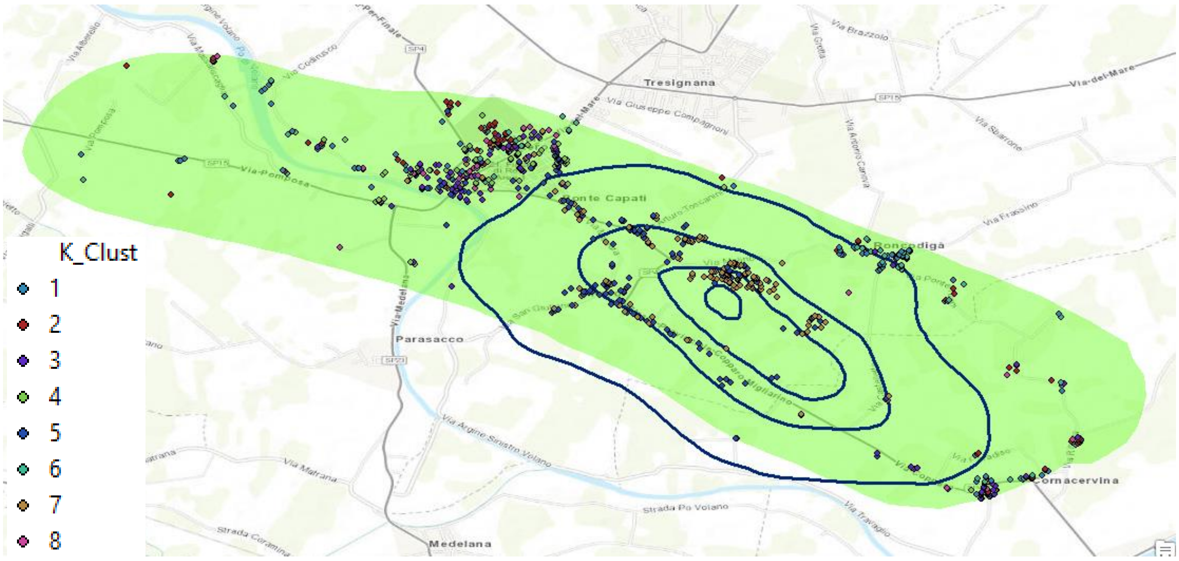


Fig. 94 - Local X-Field, 8 Classes Kmean Clustering, seasonal component. Author.

6.3 Areal level Seasonal analysis

6.3.1 4 Classes clustering results (z-normalized values plotted):

Seasonal component time series clustering charts

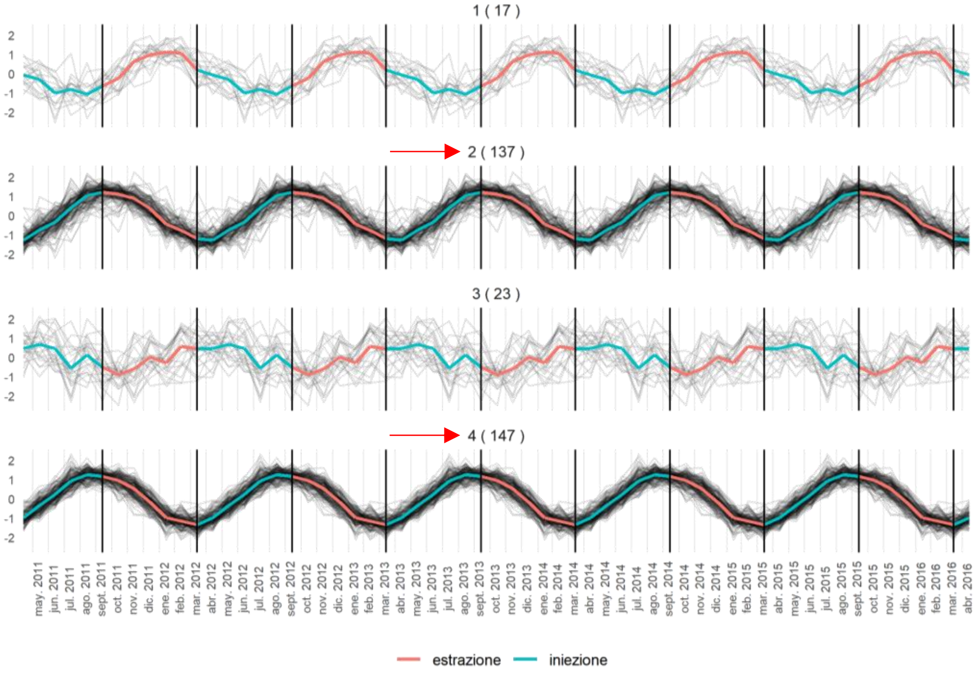


Fig. 95 - Areal Level (324 MP) - Hierarchical Clustering (4 Classes) seasonal component, z-normalized values clustered and plotted. Author.

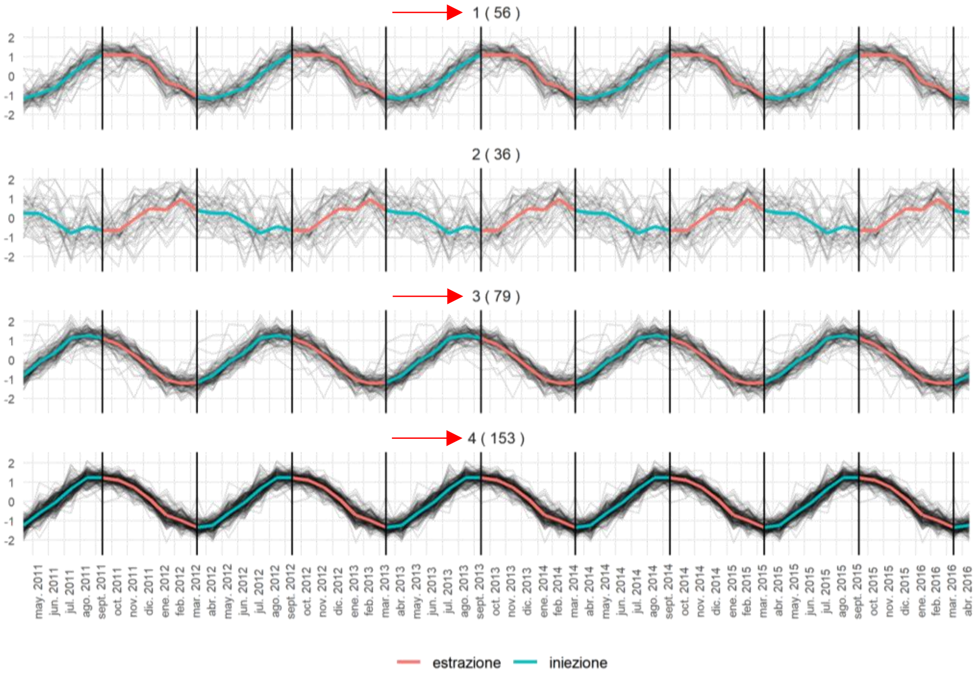


Fig. 96 - Areal Level (324 MP) - Kmean Clustering (4 Classes) seasonal component, z-normalized values clustered and plotted. Author.

MP distribution pie charts by clustering type

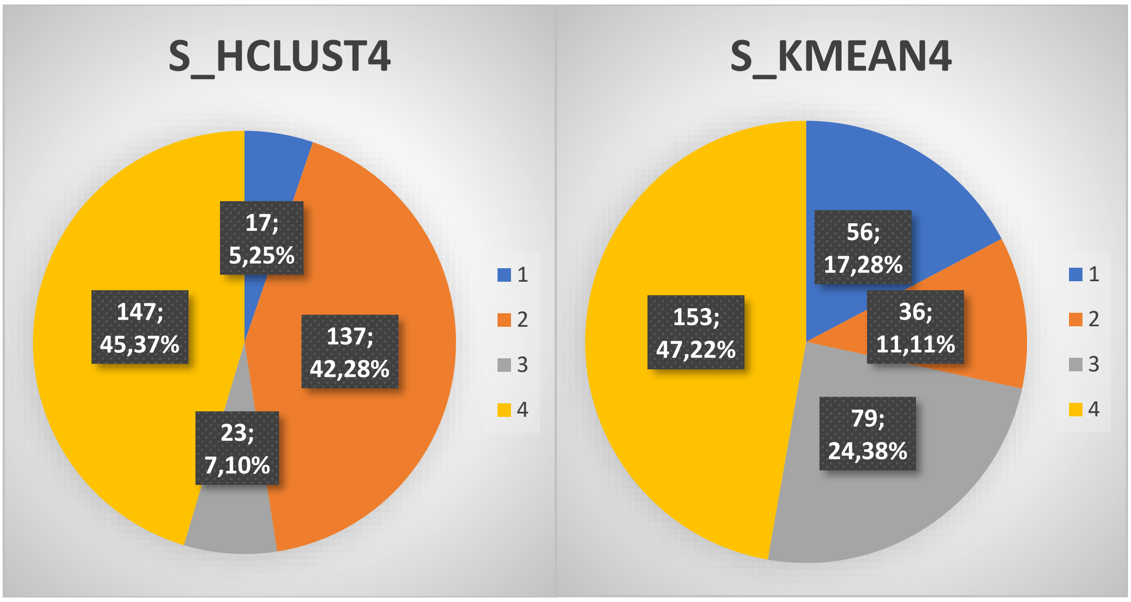


Fig. 97 - Areal X-Field, 4 Classes Hierarchical and Kmean Clustering (normalized data), seasonal component, MP distribution and relative percentages for each class. Author.

6.3.2 4 Classes clustering results (absolute values plotted):

Seasonal component time series clustering charts

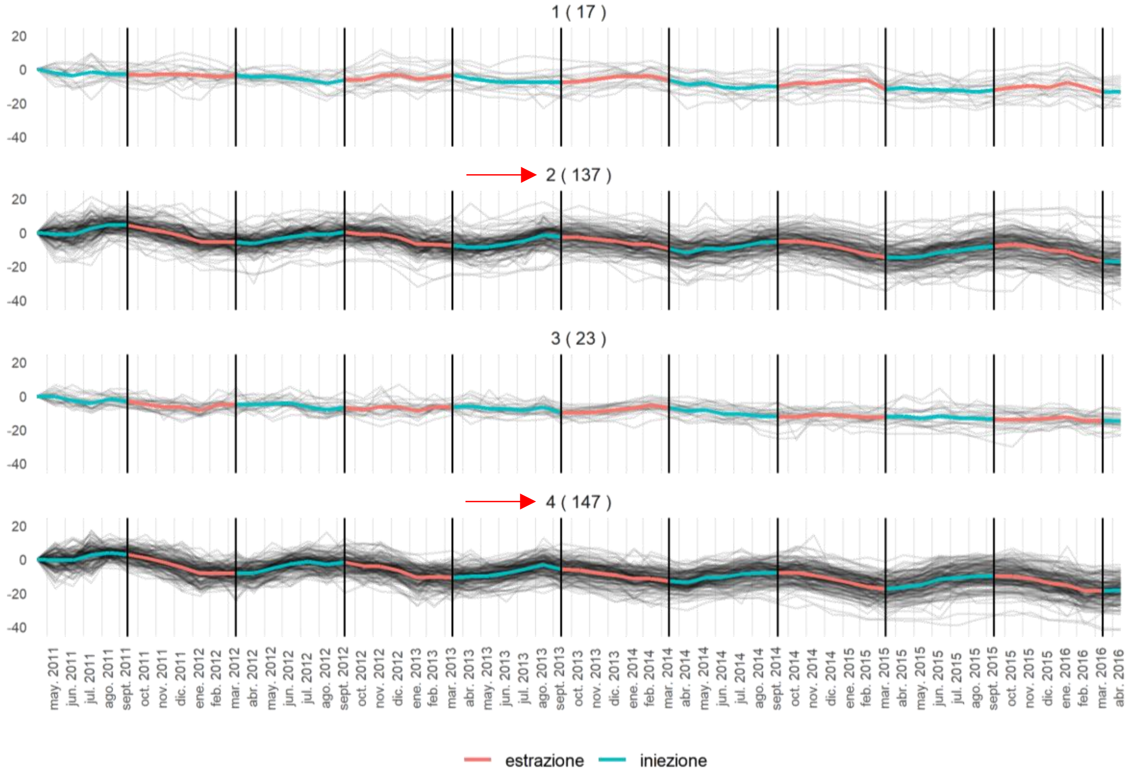


Fig. 98 - Areal Level (324 MP) - Hierarchical Clustering (4 Classes) z-normalized values clustered (by seasonal component) and resampled absolute vertical movement values plotted. Author.

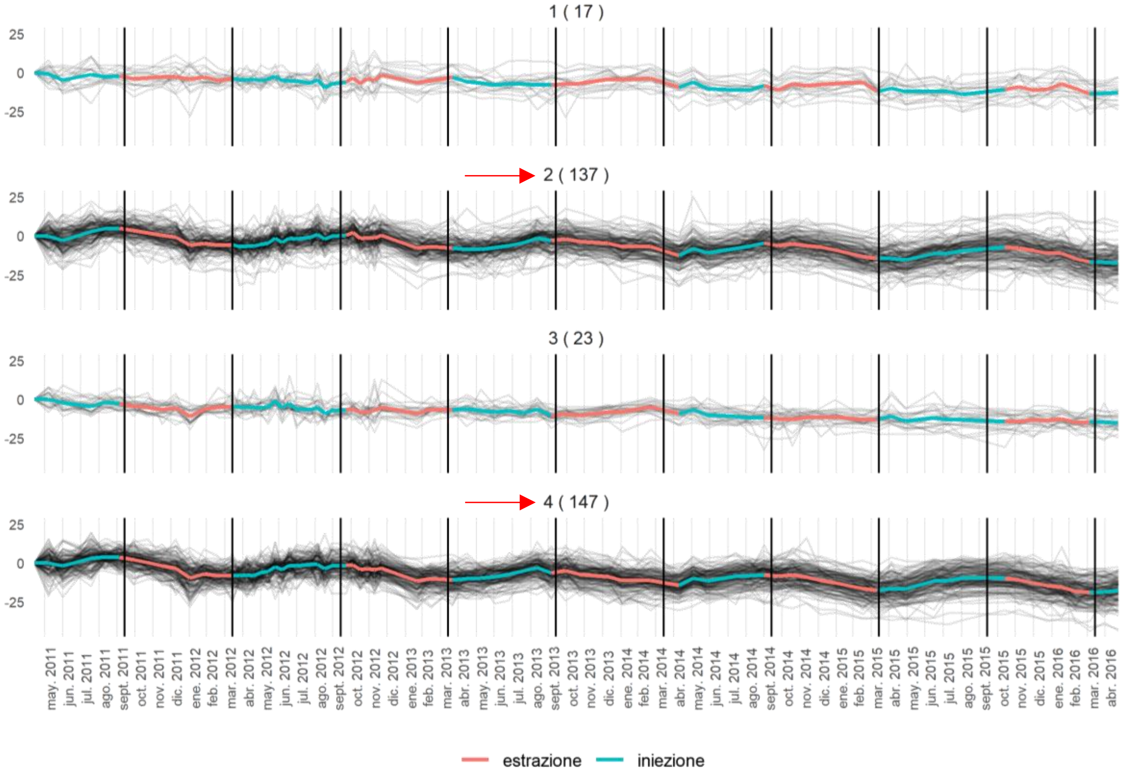


Fig. 99 - Areal Level (324 MP) - Hierarchical Clustering (4 Classes) z-normalized values clustered (by seasonal component) and absolute vertical movement values plotted. Author.

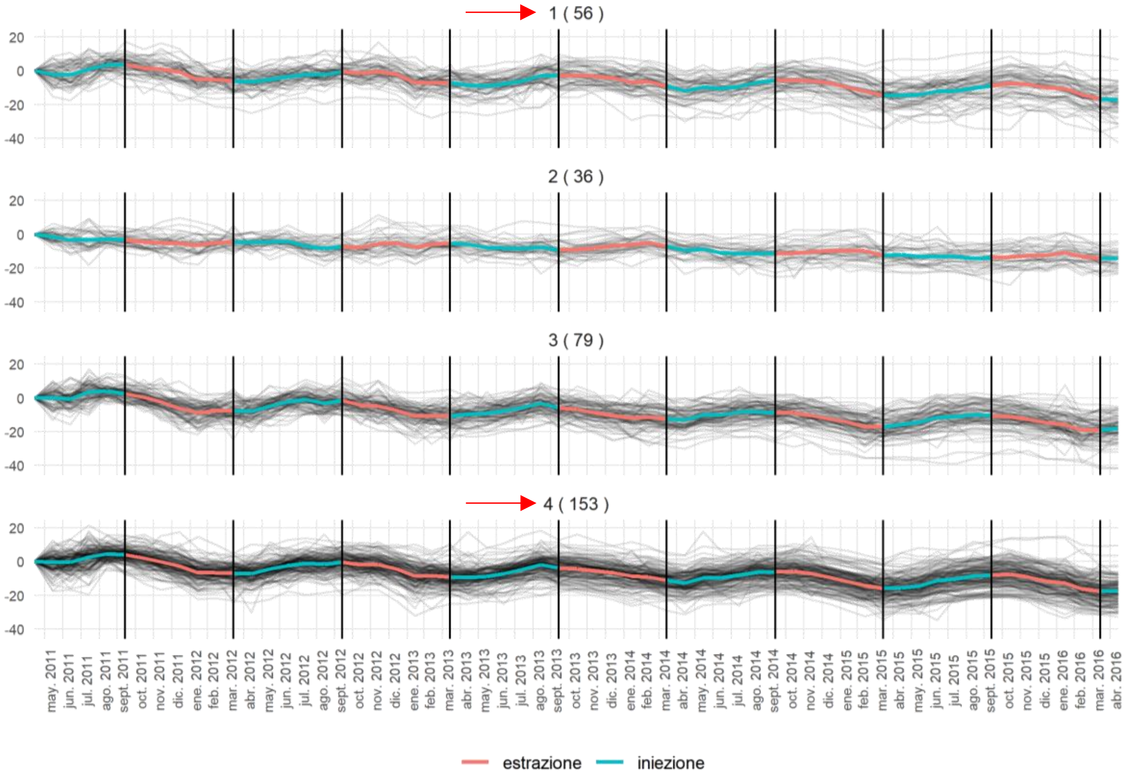


Fig. 100 - Areal Level (324 MP) - Kmean Clustering (4 Classes) z-normalized values clustered (by seasonal component) and resampled absolute vertical movement values plotted. Author.

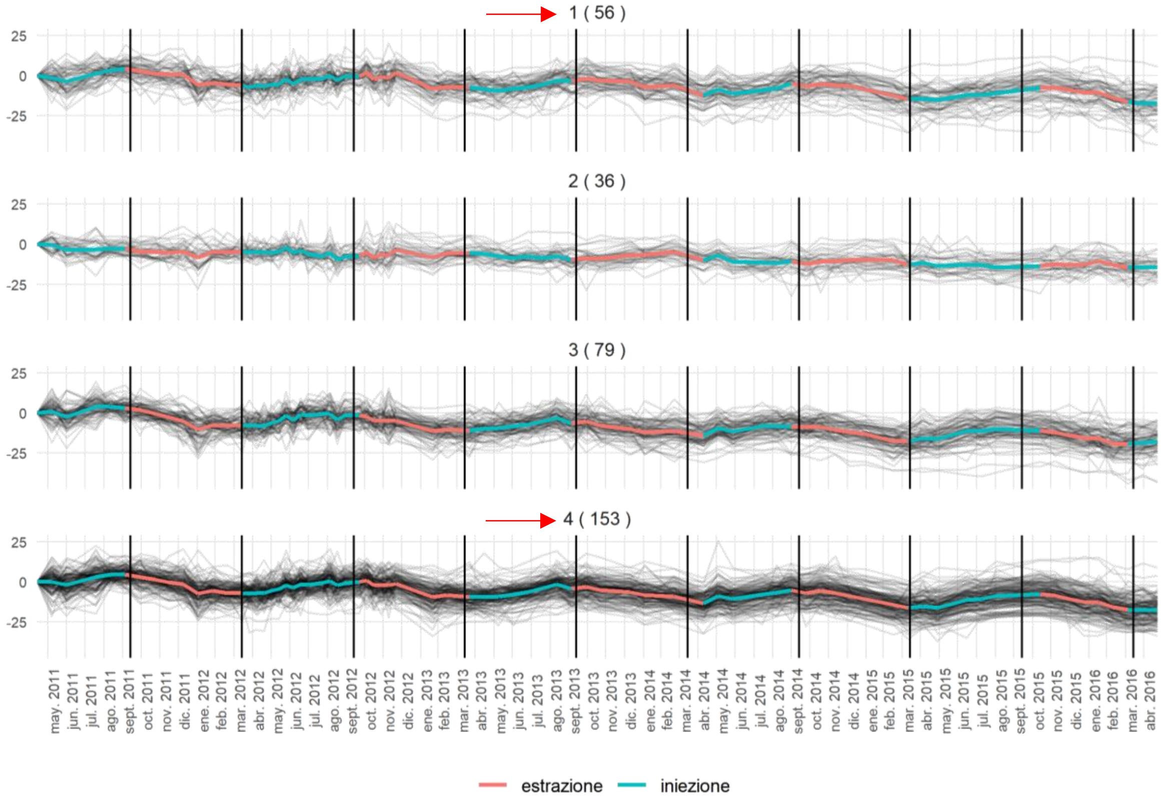


Fig. 101 - Areal Level (324 MP) - Kmean Clustering (4 Classes) z-normalized values clustered (by seasonal component) and absolute vertical movement values plotted. Author.

GIS visualization Hierarchical clustering

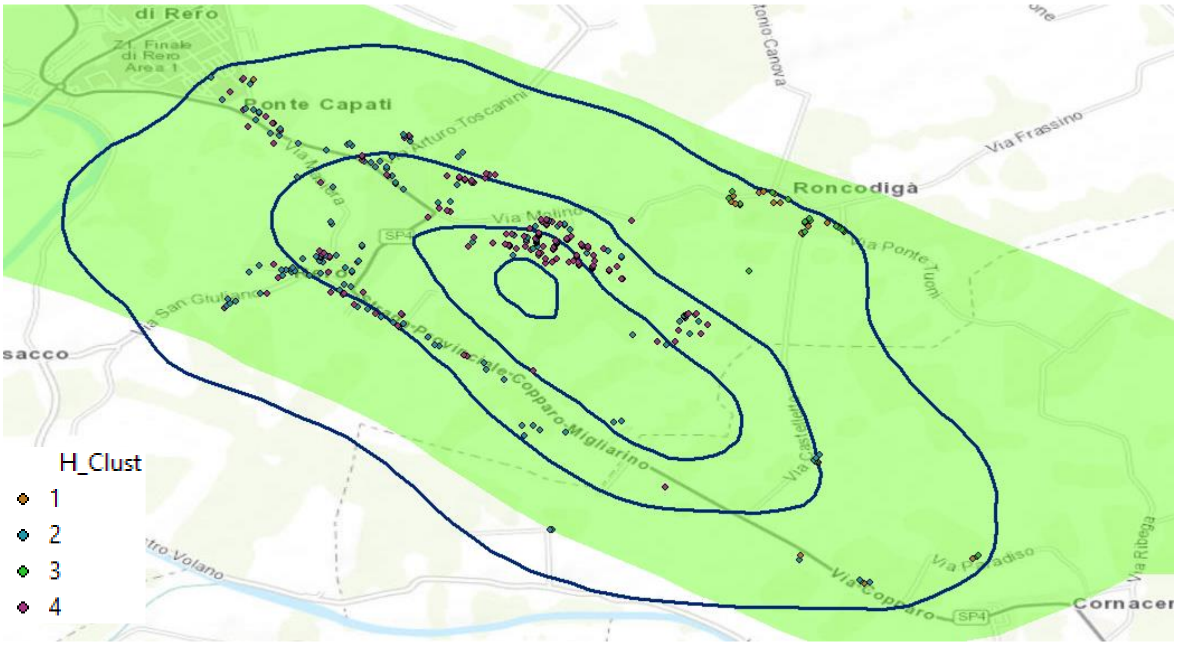


Fig. 102 - Local X-Field, 4 Classes Hierarchical Clustering, seasonal component, z-normalized values clustered and plotted. Author.

GIS visualization Kmean clustering

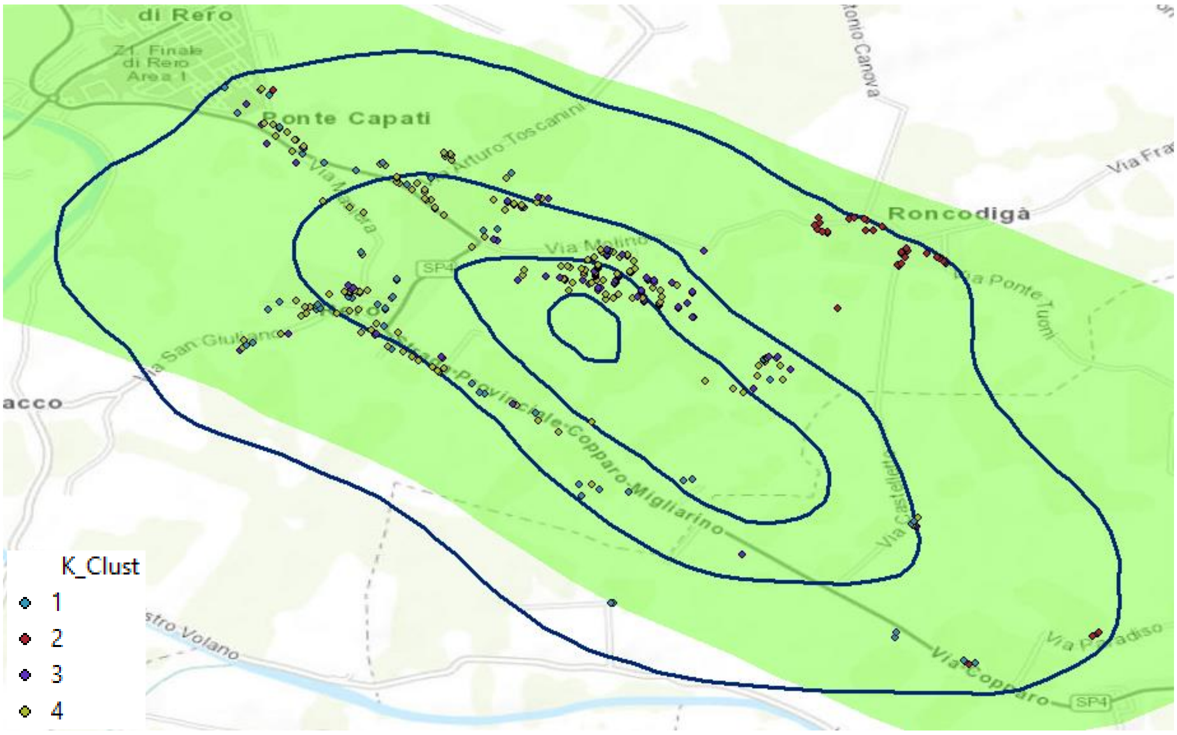


Fig. 103 - Local X-Field, 4 Classes Kmean Clustering, seasonal component, z-normalized values clustered and plotted. Author.

6.3.4 8 Classes clustering results (z-normalized values plotted):

Seasonal component time series clustering charts

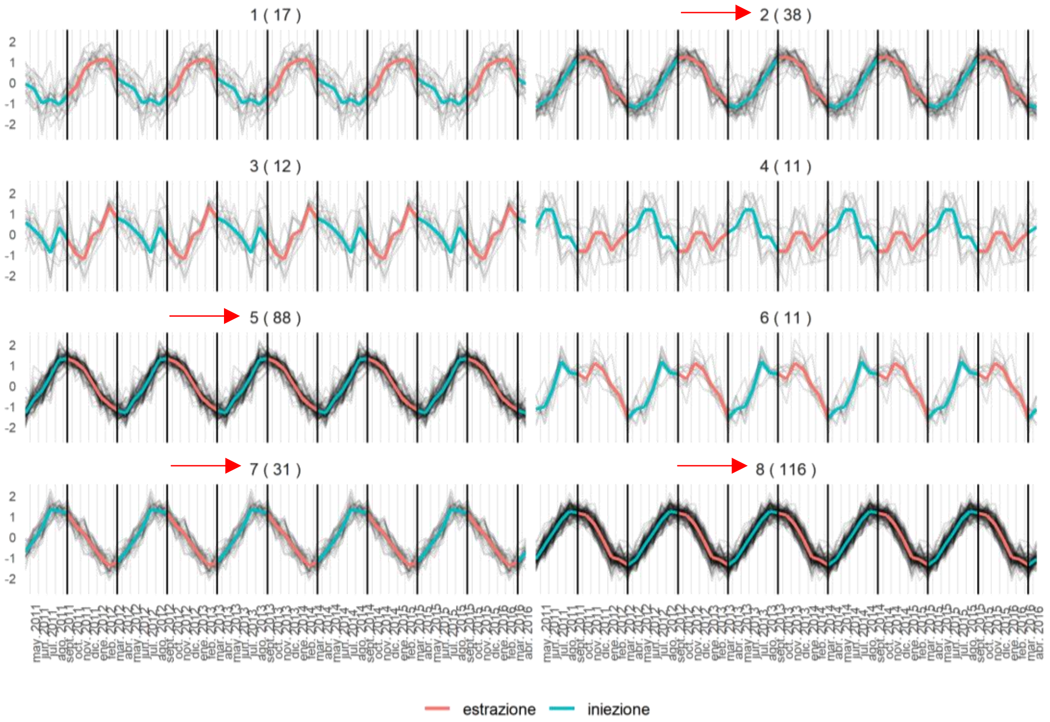


Fig. 104 - Areal Level (324 MP) - Hierarchical Clustering (8 Classes) seasonal component, z-normalized values clustered and plotted. Author.

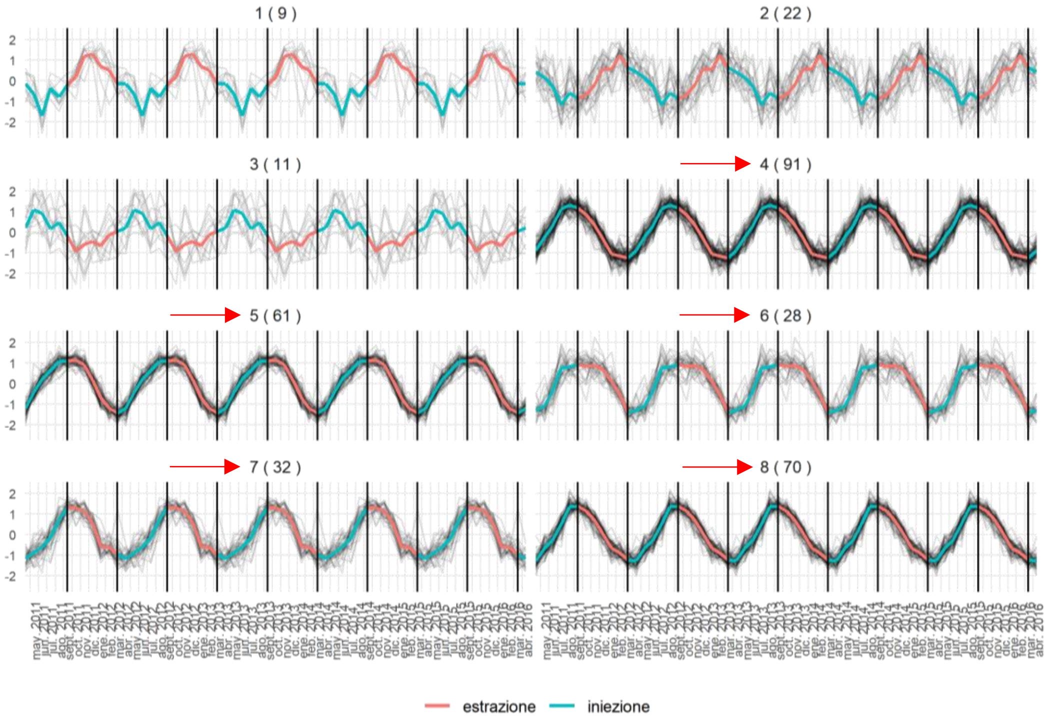


Fig. 105 - Areal Level (324 MP) - Kmean Clustering (8 Classes) seasonal component, z-normalized values clustered and plotted. Author.

MP distribution pie chart by clustering type

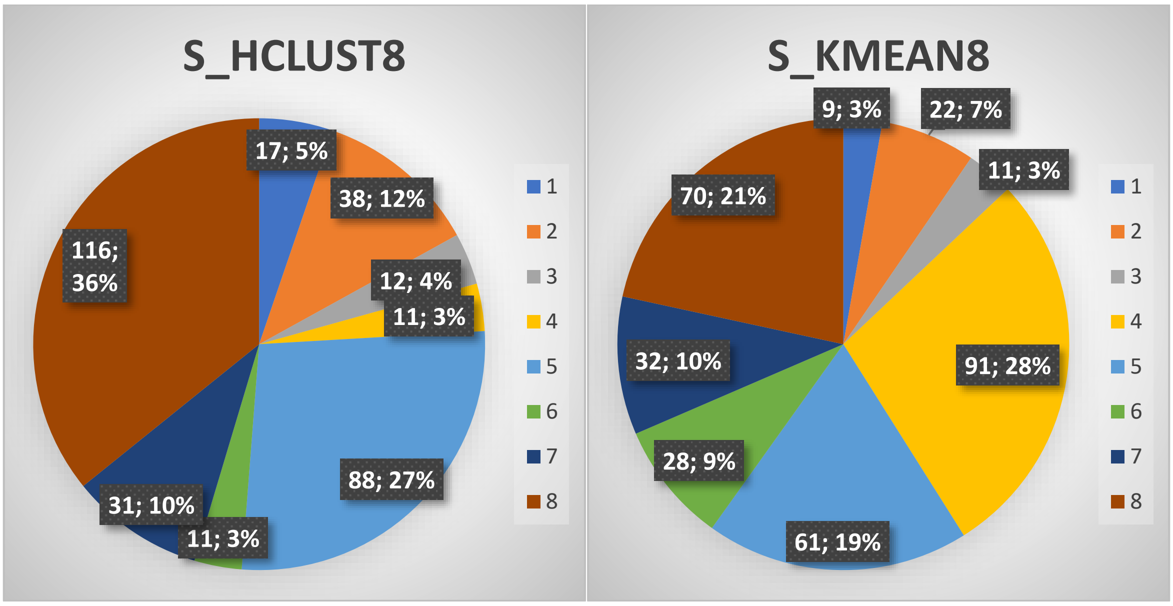


Fig. 106 - Areal X-Field, 8 Classes Hierarchical and Kmean Clustering (normalized data), seasonal component, MP distribution and relative percentages for each class. Author.

6.3.5 8 Classes clustering results (absolute values plotted):

Seasonal component time series clustering charts

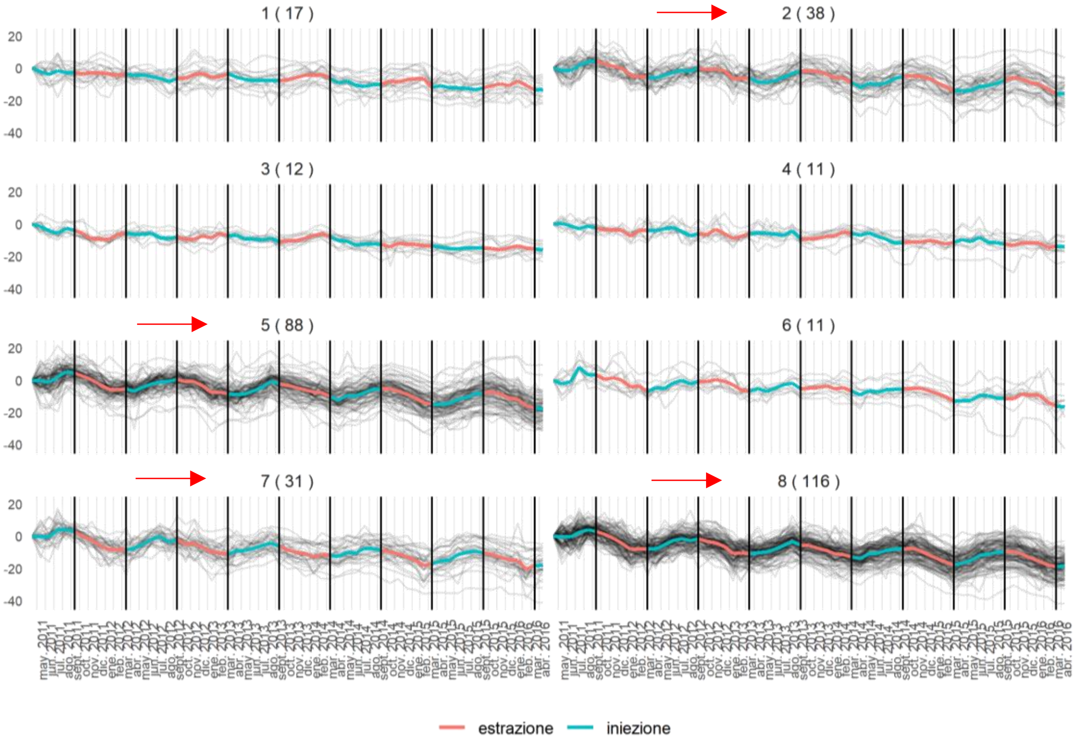


Fig. 107 - Areal Level (324 MP) - Hierarchical Clustering (8 Classes) z-normalized values clustered (by seasonal component) and resampled absolute vertical movement values plotted. Author.

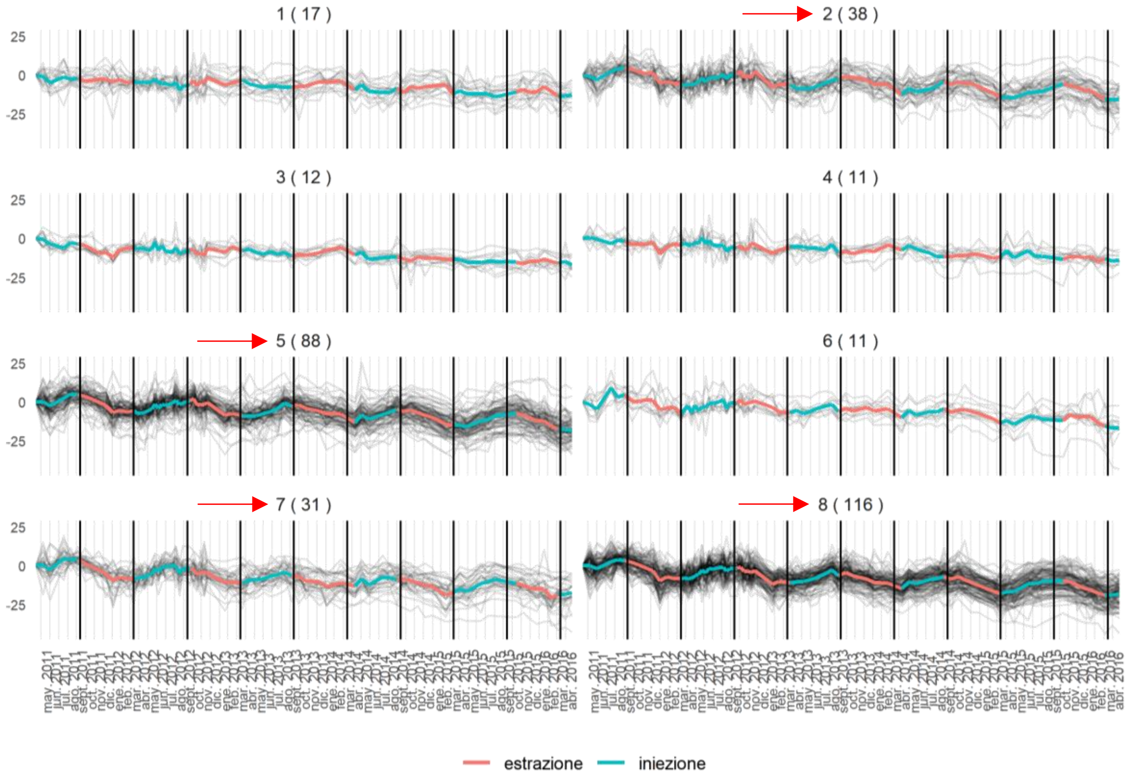


Fig. 108 - Areal Level (324 MP) - Hierarchical Clustering (8 Classes) z-normalized values clustered (by seasonal component) and absolute vertical movement values plotted. Author.

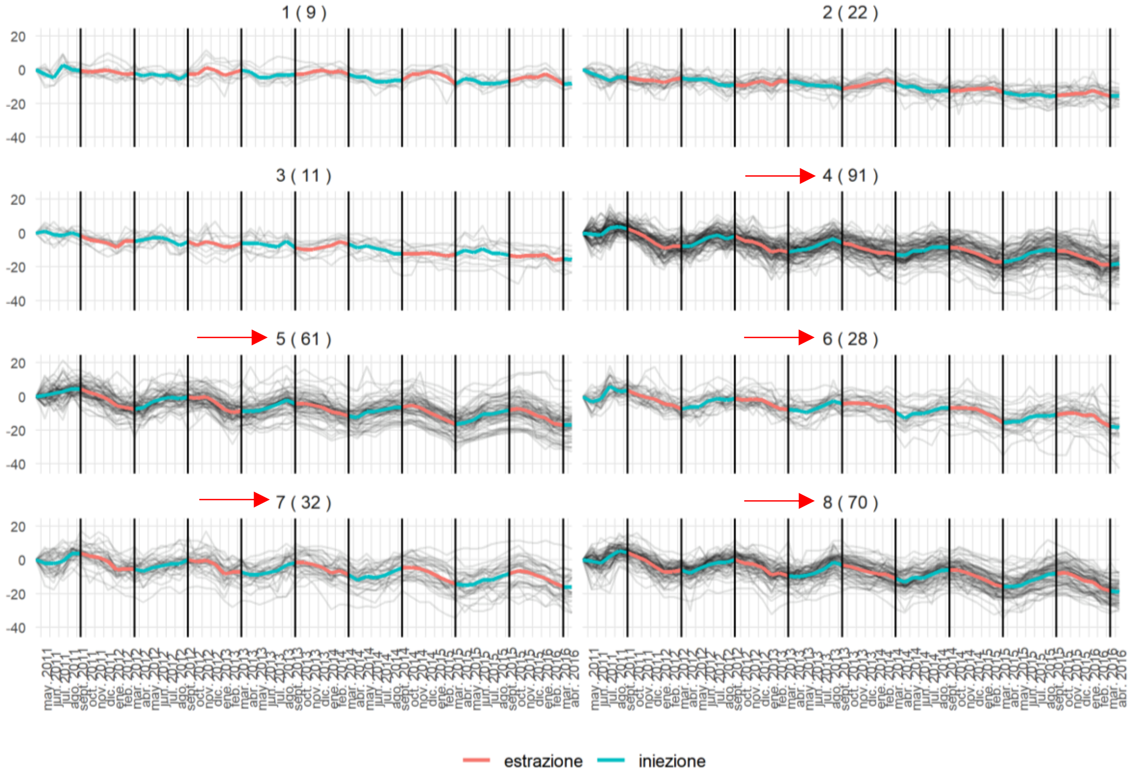


Fig. 109 - Areal Level (324 MP) - Kmean Clustering (8 Classes), z-normalized values clustered (by seasonal component) and resampled absolute vertical movement values plotted. Author.

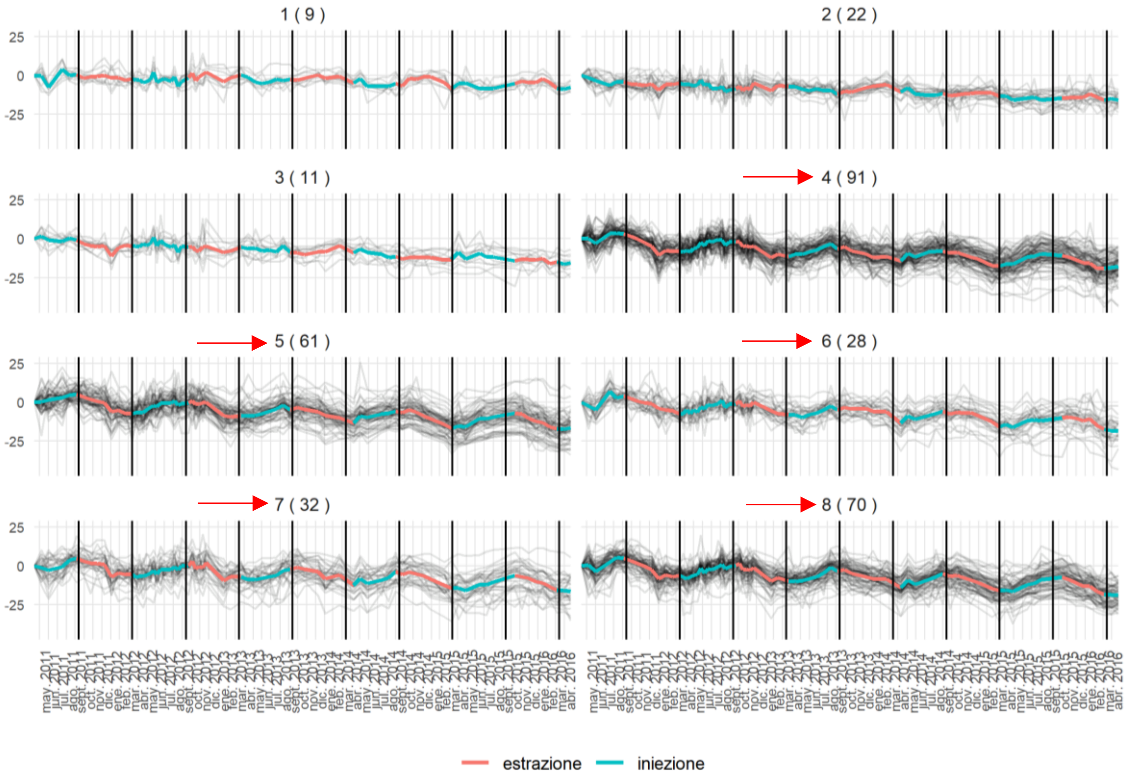


Fig. 110 - Areal Level (324 MP) - Kmean Clustering (8 Classes), z-normalized values clustered (by seasonal component) and absolute vertical movement values plotted. Author.

GIS visualization Hierarchical clustering

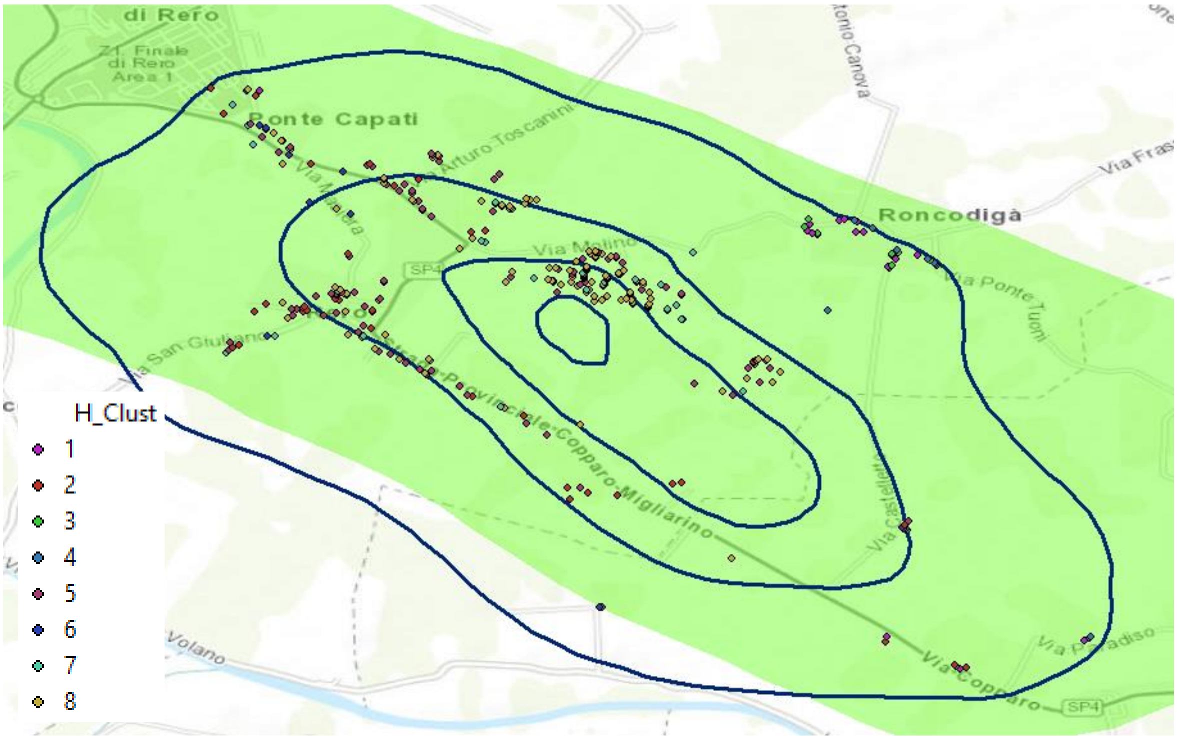


Fig. 111 - Local X-Field, 8 Classes Hierarchical Clustering, seasonal component, z-normalized values clustered and plotted. Author.

GIS visualization Kmean clustering

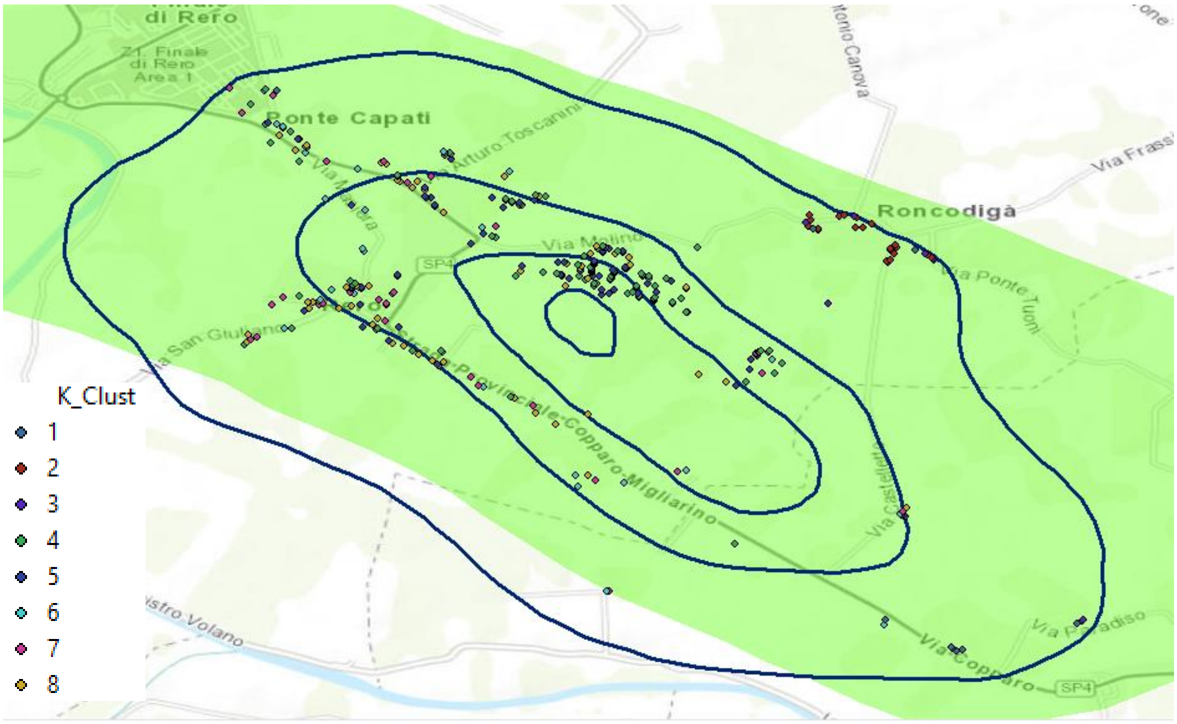


Fig. 112 - Local X-Field, 8 Classes Kmean Clustering, seasonal component, z-normalized values clustered and plotted. Author.

6.3.7 Raw resampled data clusterization comparison: 4 Classes clustering results (absolute values plotted):

Seasonal component time series clustering charts

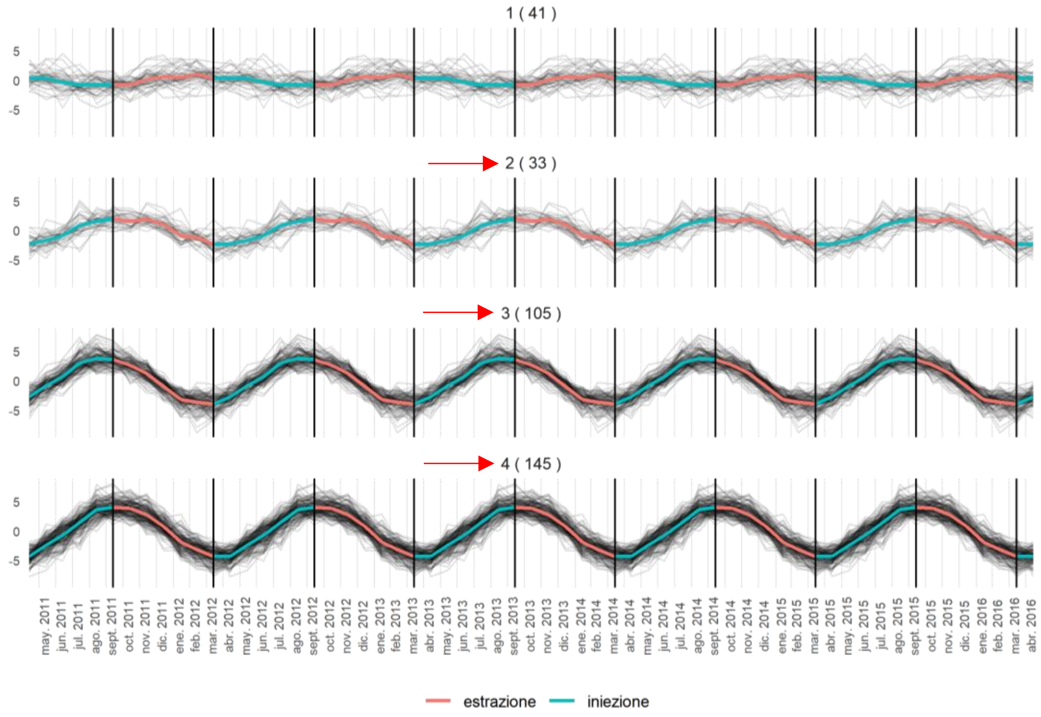


Fig. 113 - Areal Level (324 MP) - Hierarchical Clustering (4 Classes) seasonal component, raw resampled values clustered and plotted. Author.

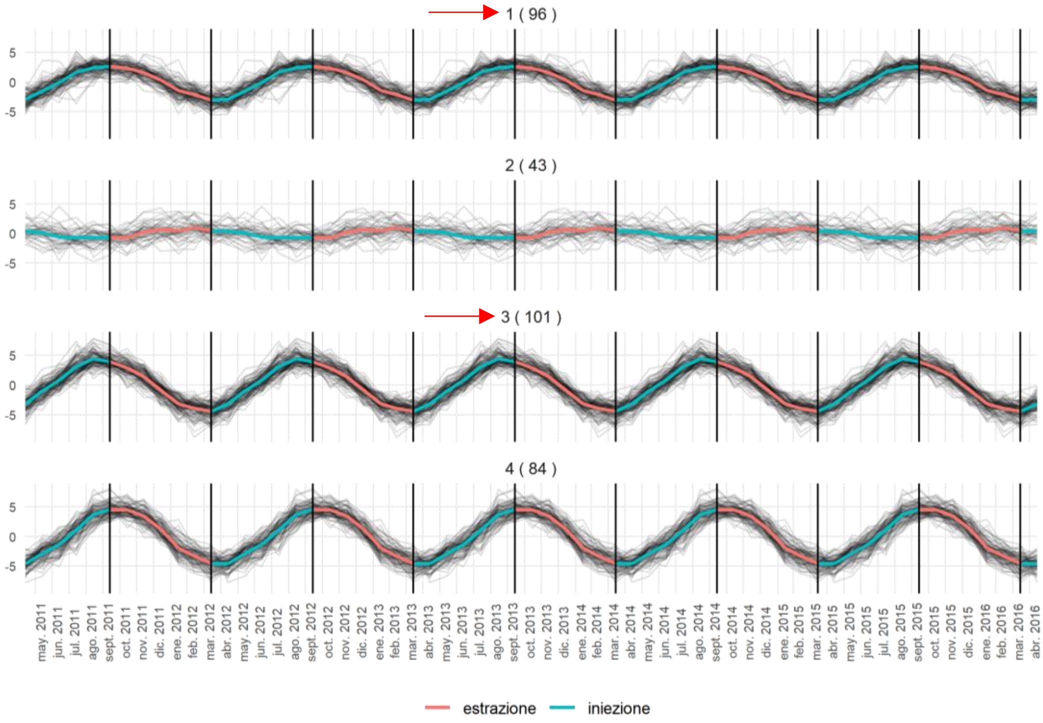


Fig. 114 - Areal Level (324 MP) - Kmean Clustering (4 Classes) seasonal component, raw resampled values clustered and plotted. Author.

MP distribution pie chart by clustering type

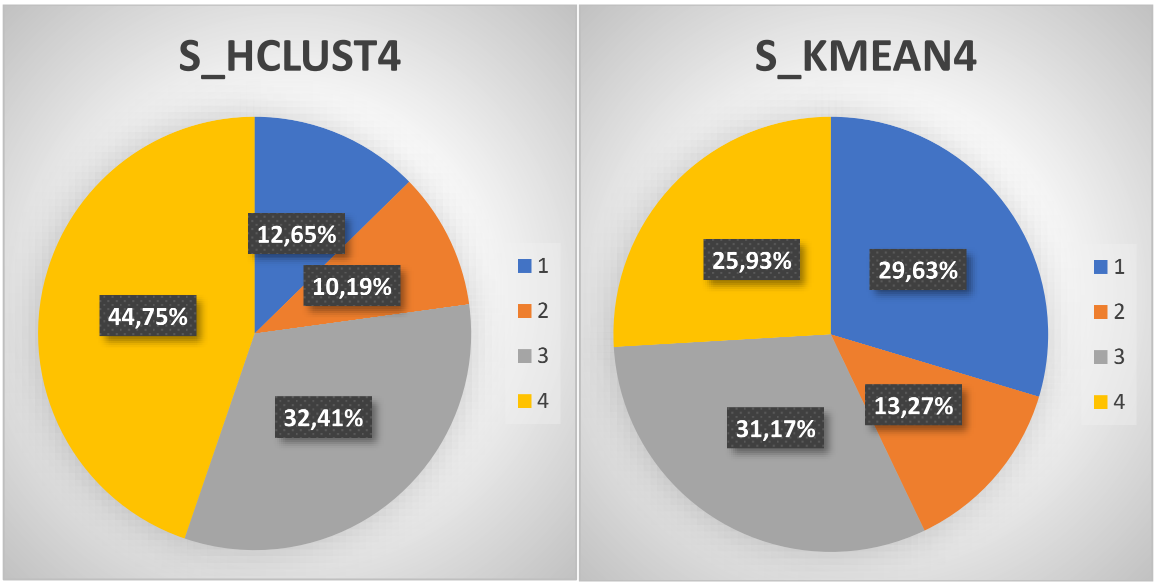


Fig. 115 - Areal X-Field, 4 Classes Hierarchical and Kmean Clustering (resampled absolute data), seasonal component, MP distribution and relative percentages for each class. Author.

GIS visualization Hierarchical clustering

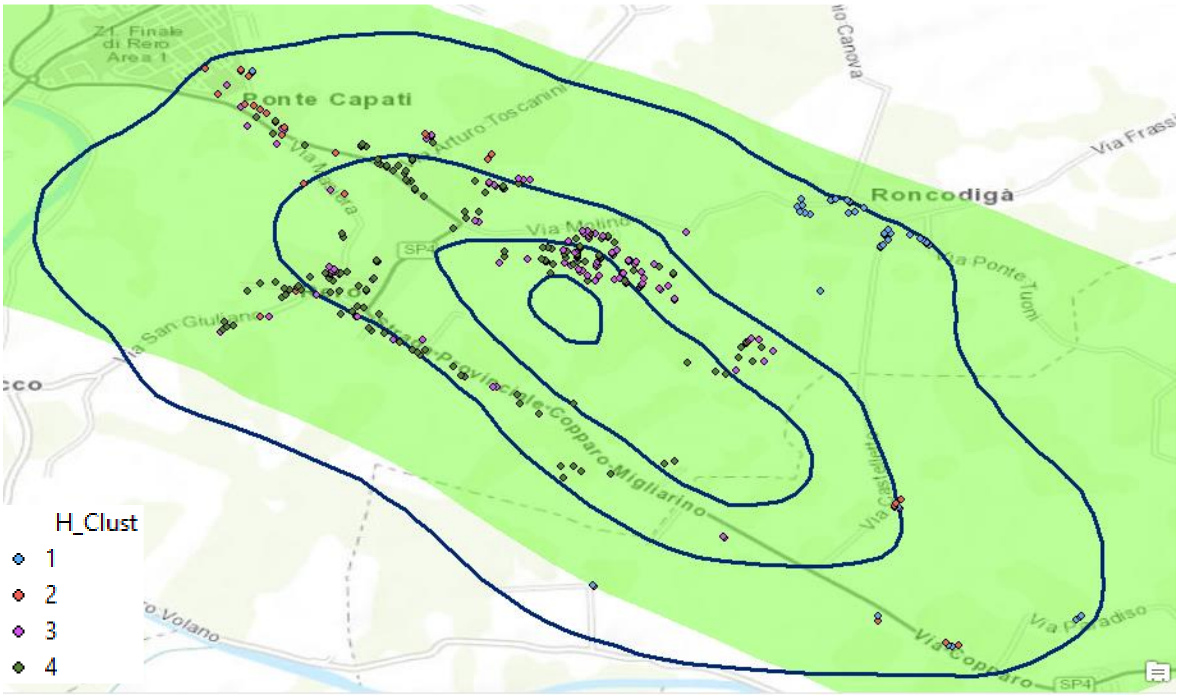


Fig. 116 - Local X-Field, 4 Classes Hierarchical Clustering, seasonal component, raw resampled data clustered and plotted. Author.

GIS visualization Kmean clustering

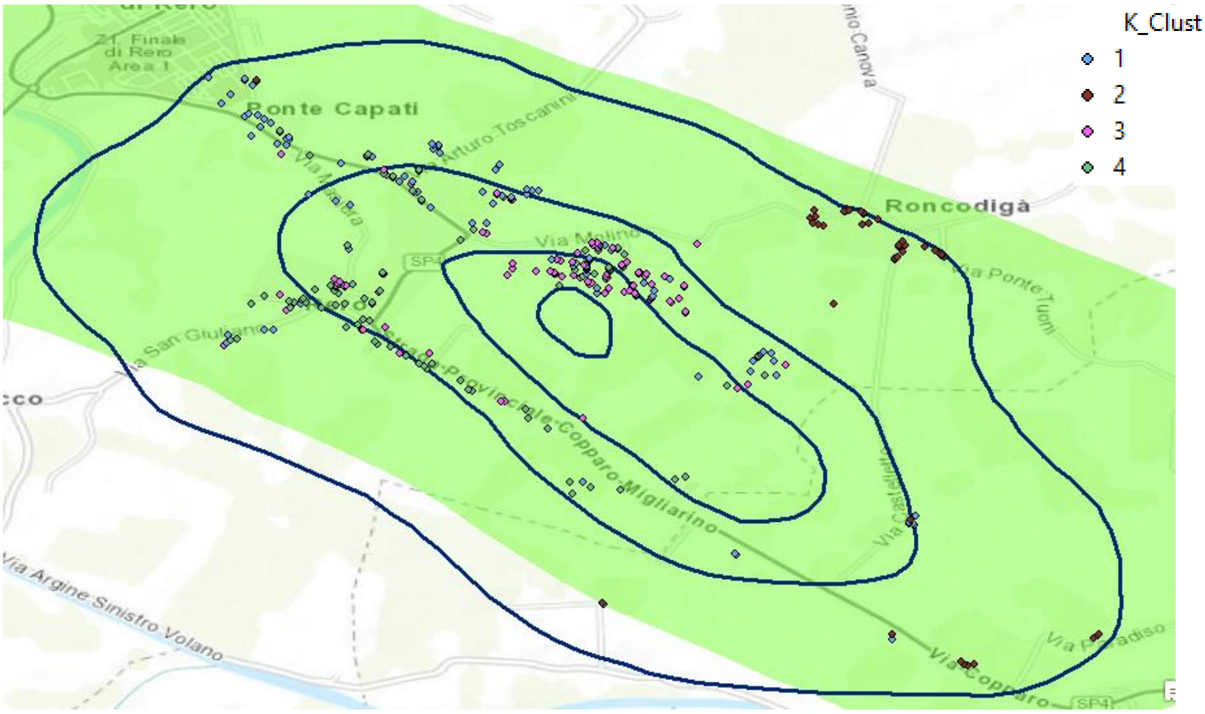


Fig. 117 - Local X-Field, 4 Classes Kmean Clustering, seasonal component, raw resampled data clustered and plotted. Author.

6.3.9 Raw resampled data clusterization comparison: 8 Classes clustering results (absolute values plotted):

Seasonal component time series clustering charts

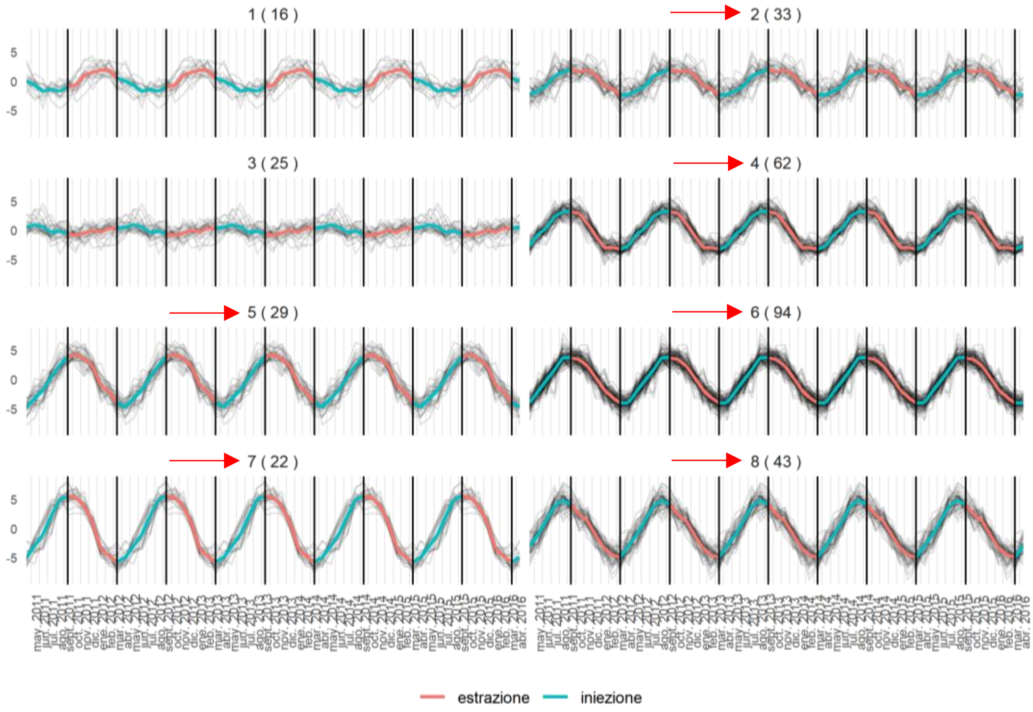


Fig. 118 - Areal Level (324 MP) - Hierarchical Clustering (8 Classes) seasonal component, raw resampled values clustered and plotted. Author.

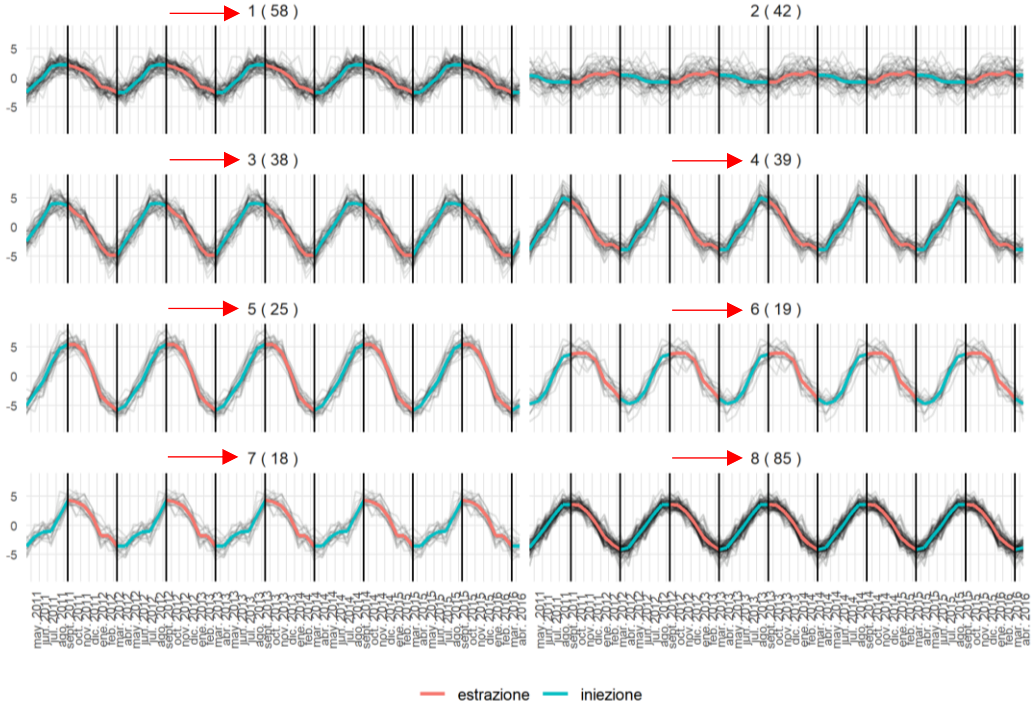


Fig. 119 - Areal Level (324 MP) - Kmean Clustering (8 Classes) seasonal component, raw resampled values clustered and plotted. Author.

MP distribution pie chart by clustering type

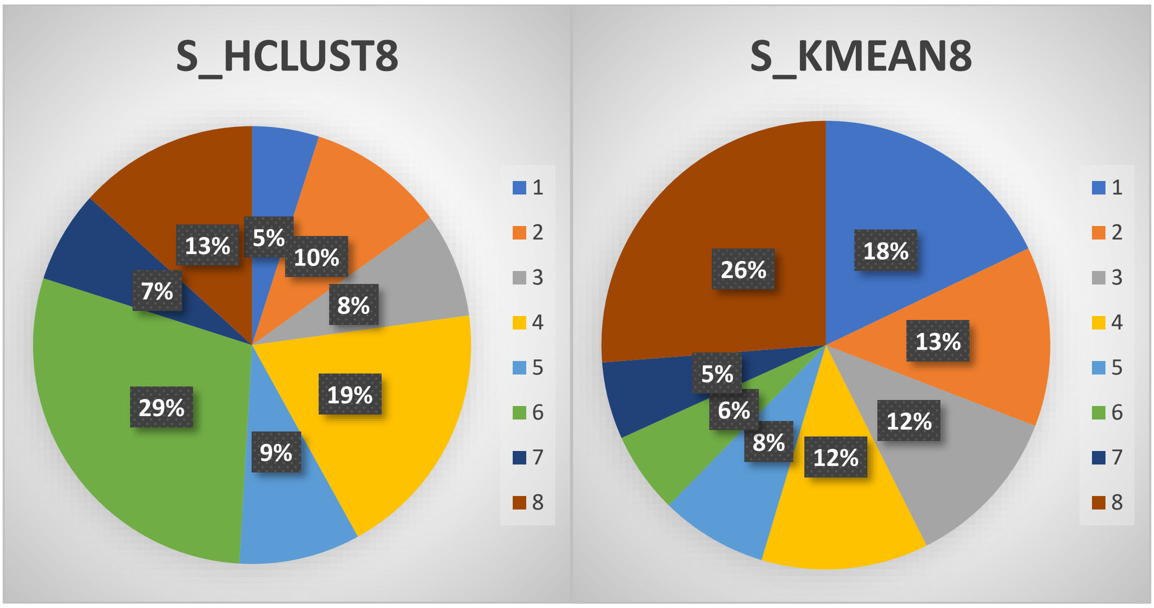


Fig. 120 - Areal X-Field, 8 Classes Hierarchical and Kmean Clustering (resampled absolute data), seasonal component, MP distribution and relative percentages for each class. Author.

GIS visualization Hierarchical clustering

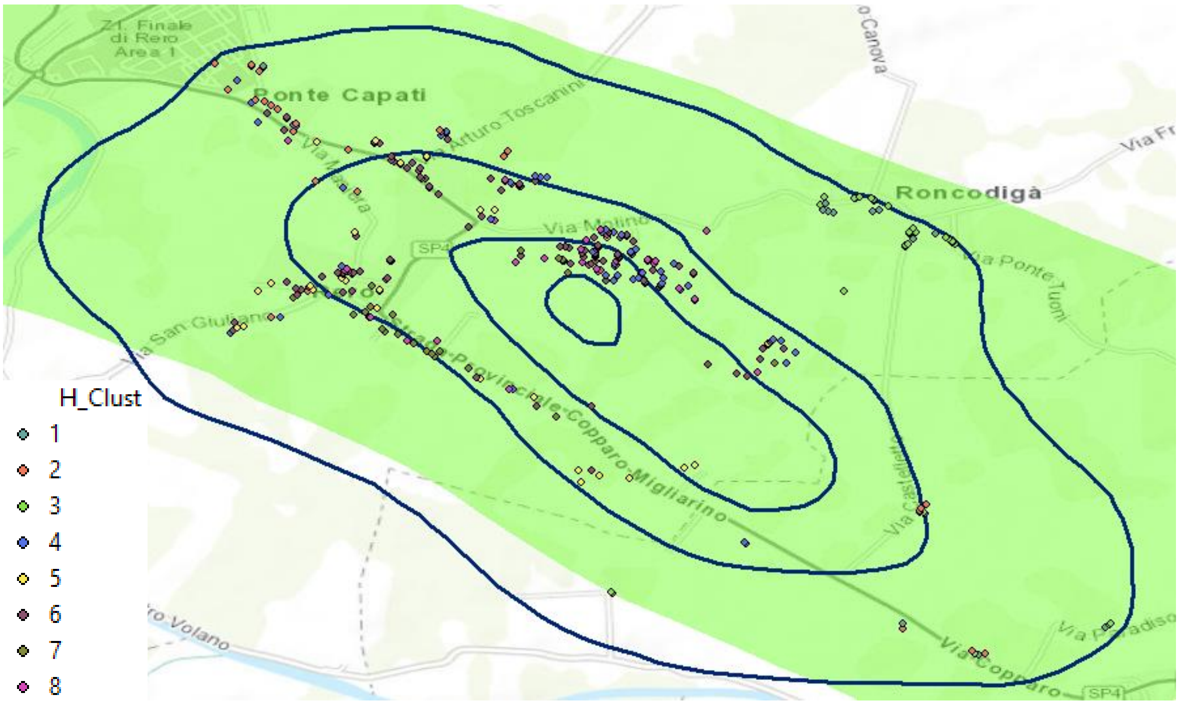


Fig. 121 - Local X-Field, 8 Classes Hierarchical Clustering, seasonal component, raw resampled data clustered and plotted. Author.

GIS visualization Kmean clustering

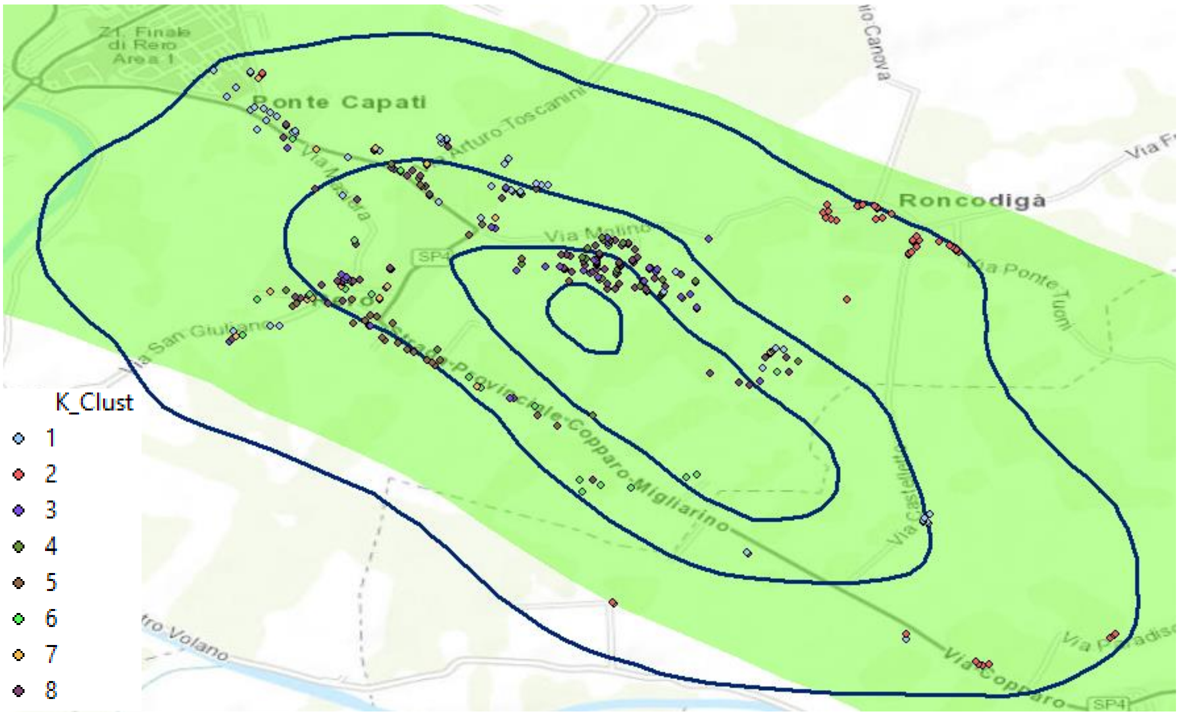
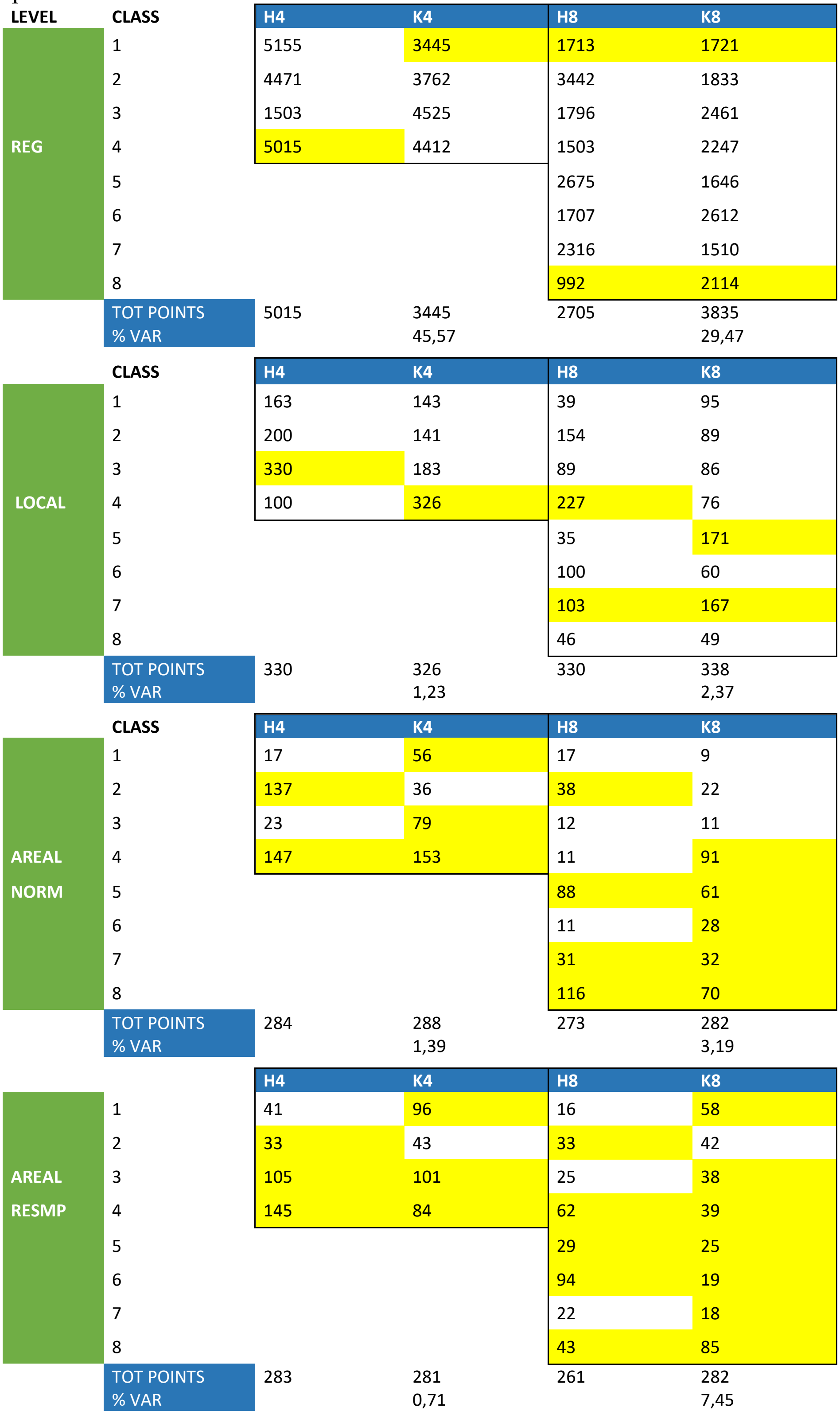


Fig. 122 - Local X-Field, 8 Classes Kmean Clustering, seasonal component, raw resampled data clustered and plotted. Author.

6.4 Clustering match comparison in-between clustered classes

On yellow the classes inside the n-4 and n-8 clusters (regarding the seasonal component clustering) that better portrait the subsidence behavior for the UGS-related activities.

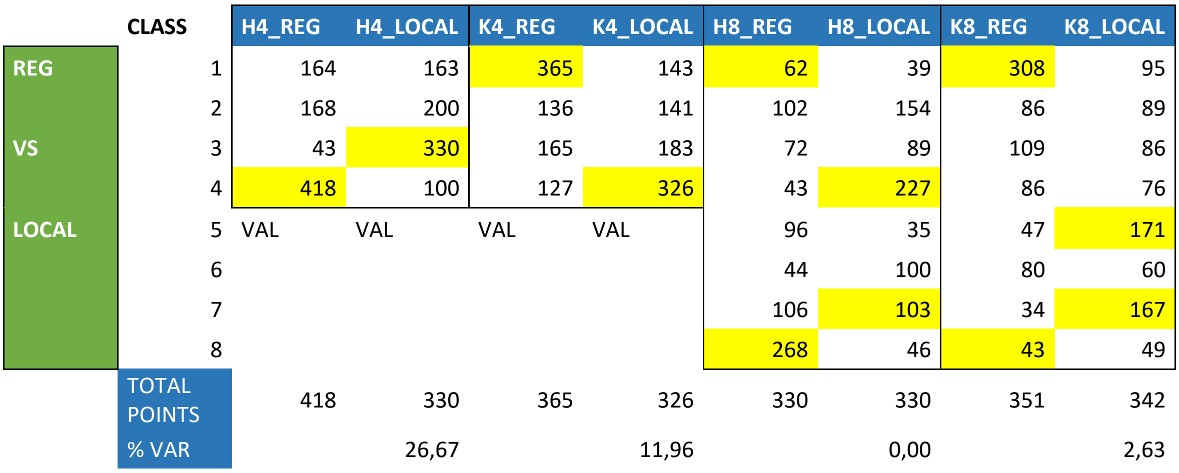


LEVEL	CLASS	H4	K4	H8	K8
REG	1	5155	3445	1713	1721
	2	4471	3762	3442	1833
	3	1503	4525	1796	2461
	4	5015	4412	1503	2247
	5			2675	1646
	6			1707	2612
	7			2316	1510
	8			992	2114
	TOT POINTS	5015	3445	2705	3835
	% VAR		45,57		29,47
	CLASS	H4	K4	H8	K8
LOCAL	1	163	143	39	95
	2	200	141	154	89
	3	330	183	89	86
	4	100	326	227	76
	5			35	171
	6			100	60
	7			103	167
	8			46	49
	TOT POINTS	330	326	330	338
	% VAR		1,23		2,37
	CLASS	H4	K4	H8	K8
AREAL NORM	1	17	56	17	9
	2	137	36	38	22
	3	23	79	12	11
	4	147	153	11	91
	5			88	61
	6			11	28
	7			31	32
	8			116	70
	TOT POINTS	284	288	273	282
	% VAR		1,39		3,19
		H4	K4	H8	K8
AREAL RESMP	1	41	96	16	58
	2	33	43	33	42
	3	105	101	25	38
	4	145	84	62	39
	5			29	25
	6			94	19
	7			22	18
	8			43	85
	TOT POINTS	283	281	261	282
	% VAR		0,71		7,45

Table 9 - Match of subsidence behavior for clustered classes, nested and unnested approach at the same level. Author.

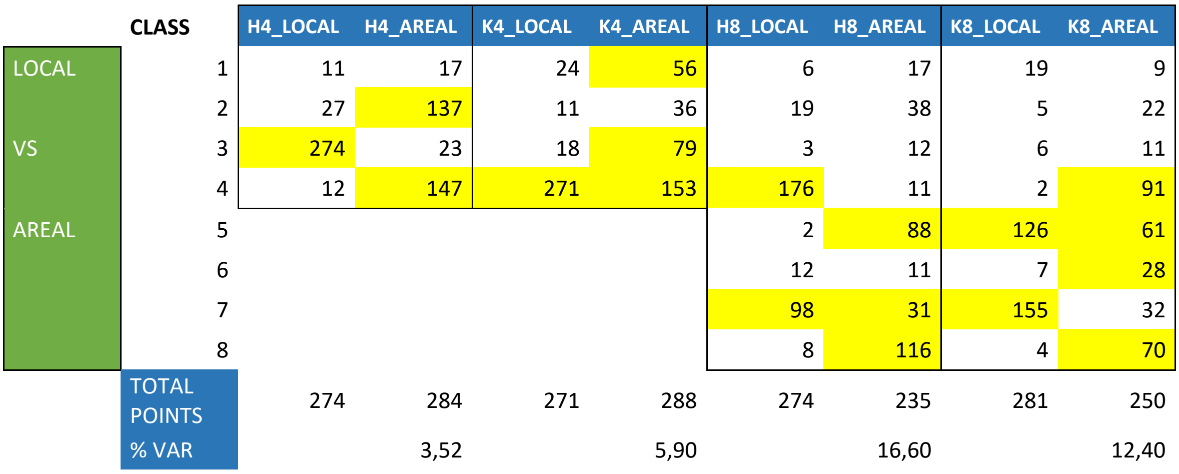
6.5 Clustering class matching comparison between levels of analysis

On yellow the next four tables highlight the clusters (regarding the seasonal component clustering) that better portrait the subsidence behavior for the UGS-related activities among the different levels of analysis that match spatially and in shape (partially).



	CLASS	H4_REG	H4_LOCAL	K4_REG	K4_LOCAL	H8_REG	H8_LOCAL	K8_REG	K8_LOCAL
REG	1	164	163	365	143	62	39	308	95
	2	168	200	136	141	102	154	86	89
VS	3	43	330	165	183	72	89	109	86
	4	418	100	127	326	43	227	86	76
LOCAL	5	VAL	VAL	VAL	VAL	96	35	47	171
	6					44	100	80	60
	7					106	103	34	167
	8					268	46	43	49
	TOTAL POINTS	418	330	365	326	330	330	351	342
	% VAR		26,67		11,96		0,00		2,63

Table 10 - Cluster class spatial/shape matching, regional vs local level, normalized data. Author.



	CLASS	H4_LOCAL	H4_AREAL	K4_LOCAL	K4_AREAL	H8_LOCAL	H8_AREAL	K8_LOCAL	K8_AREAL
LOCAL	1	11	17	24	56	6	17	19	9
	2	27	137	11	36	19	38	5	22
VS	3	274	23	18	79	3	12	6	11
	4	12	147	271	153	176	11	2	91
AREAL	5					2	88	126	61
	6					12	11	7	28
	7					98	31	155	32
	8					8	116	4	70
	TOTAL POINTS	274	284	271	288	274	235	281	250
	% VAR		3,52		5,90		16,60		12,40

Table 11 - Cluster class spatial/shape matching, local vs areal level, normalized data. Author.

	CLASS	H4_REG	H4_LOCAL	K4_REG	K4_LOCAL	H8_REG	H8_LOCAL	K8_EXT	K8_LOCAL
REG	1	36	17	280	56	27	17	265	9
	2	23	137	21	36	9	38	14	22
VS	3	5	23	14	79	10	12	11	11
	4	260	147	9	153	5	11	3	91
LOCAL	5					13	88	3	61
	6					4	11	13	28
	7					11	31	3	32
	8					245	116	12	70
	TOTAL POINTS	260	284	280	288	272	235	277	250
	% VAR		8,45		2,78		15,74		10,80

Table 12 - Cluster class spatial/shape matching, regional vs areal level, normalized data. Author.

	CLASS	H4_NORM	H4_RAW	K4_NORM	K4_RAW	H8_NORM	H8_RAW	K8_NORM	K8_RAW
LOCAL	1	17	41	56	96	17	16	9	58
NORM	2	137	33	36	43	38	33	22	42
	3	23	105	79	101	12	25	11	38
VS	4	147	145	153	84	11	62	91	39
	5					88	29	61	25
LOCAL	6					11	94	28	19
RAW	7					31	22	32	18
	8					116	43	70	85
	TOTAL POINTS	284	283	288	281	235	261	250	264
	% VAR		0,35		2,49		9,96		5,30

Table 13 - Cluster class spatial/shape matching, areal level: normalized vs raw resampled clustered data. Author.

6.6 Trend component analysis

6.6.1 Regional Level

Normalized values clustering and plotting

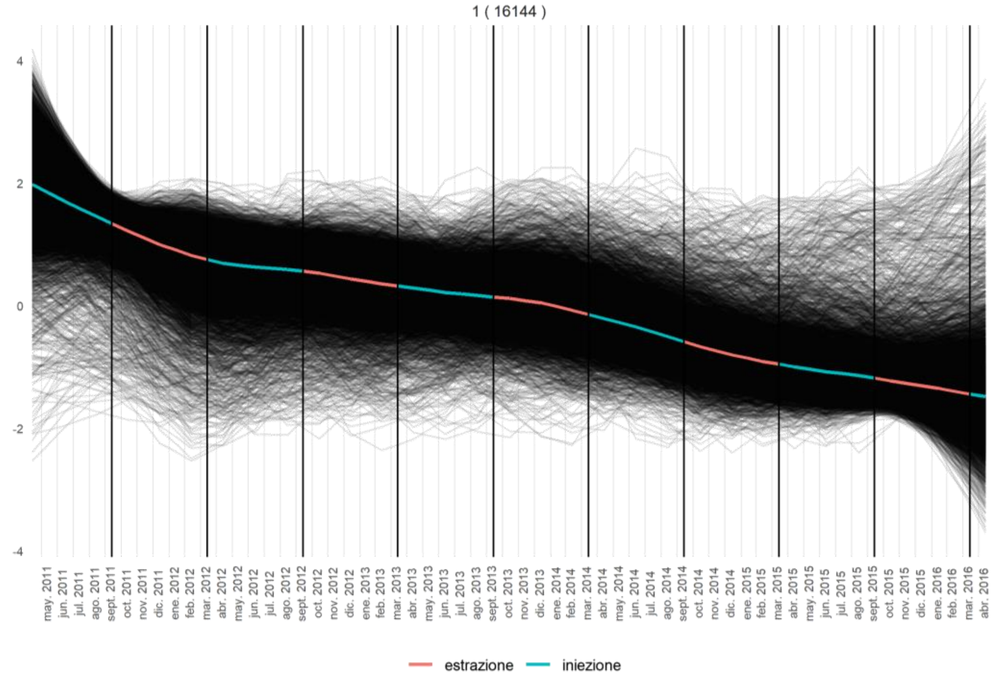


Fig. 123 - Regional X-Field, 1 Class Hierarchical Clustering, trend component (normalized values). Author.

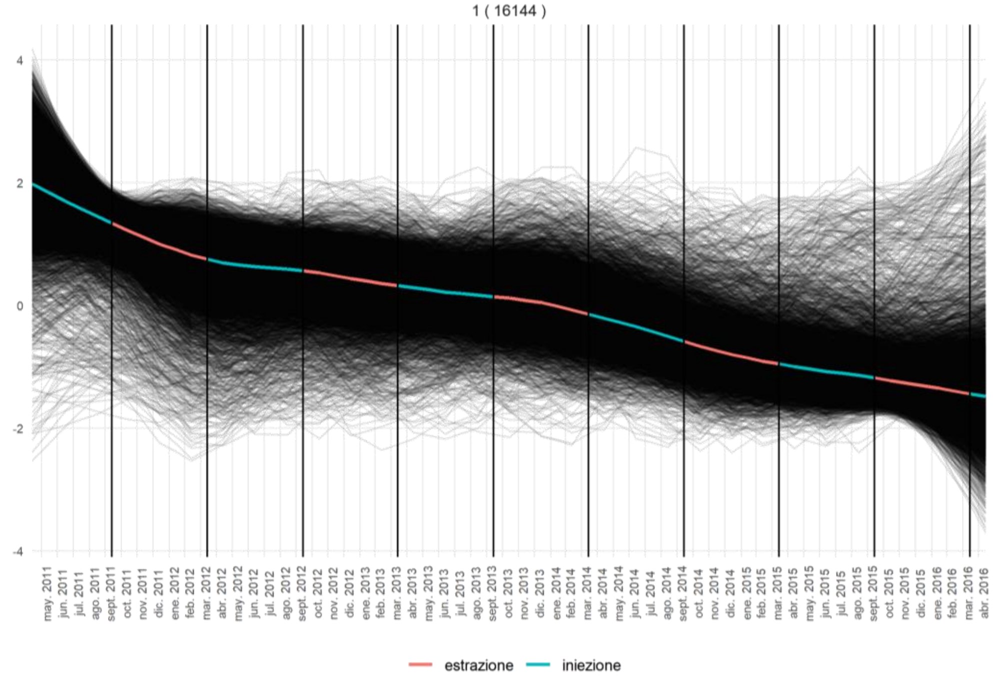


Fig. 124 - Regional X-Field, 1 Class Kmean Clustering, trend component (normalized values). Author.

Normalized values clustering and absolute values plot

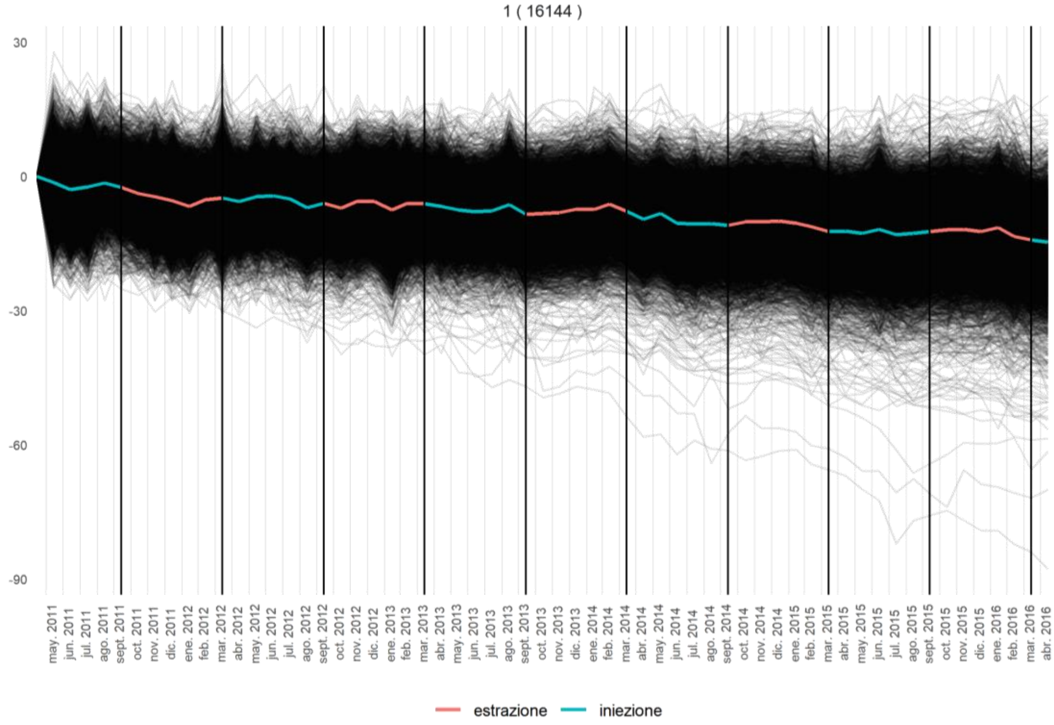


Fig. 125 - Regional X-Field, 1 Class Hierarchical Clustering, trend component (resampled absolute values). Author.

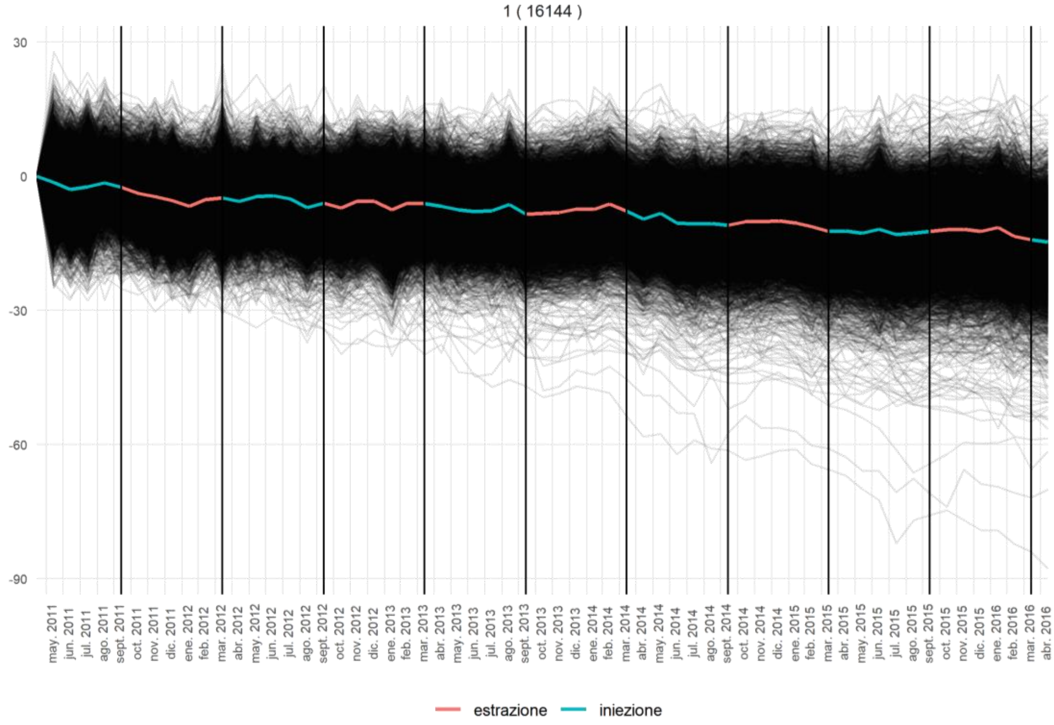


Fig. 126 - Regional X-Field, 1 Class Kmean Clustering trend component (resampled absolute values). Author.

6.6.2 Local Level

Normalized values clustering and plotting

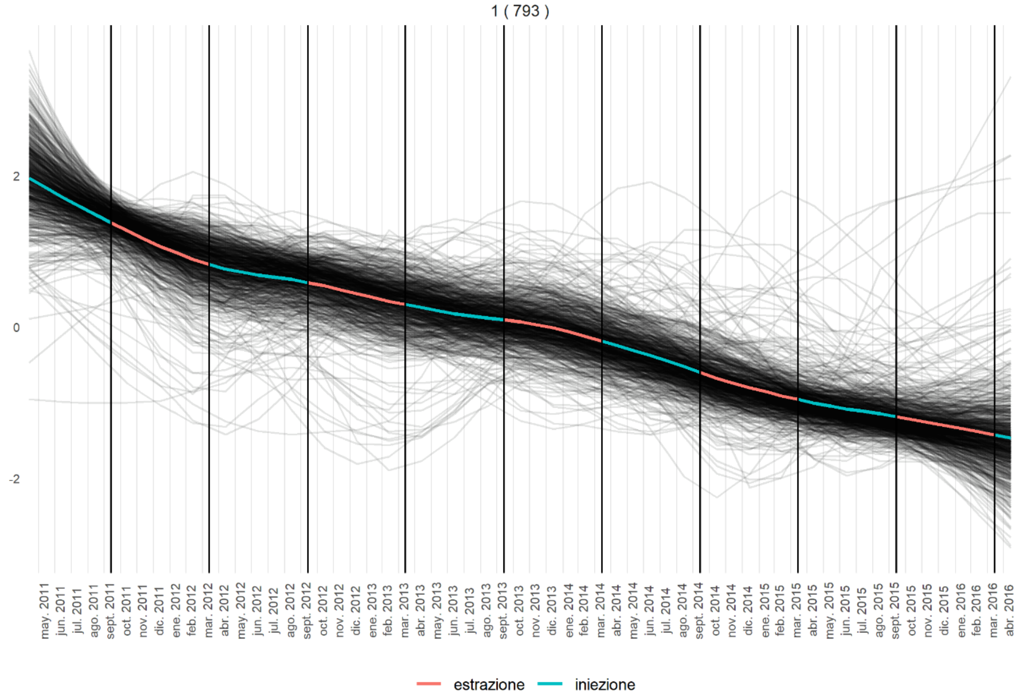


Fig. 127 - Local X-Field, 1 Class Hierarchical Clustering, trend component (normalized values). Author.

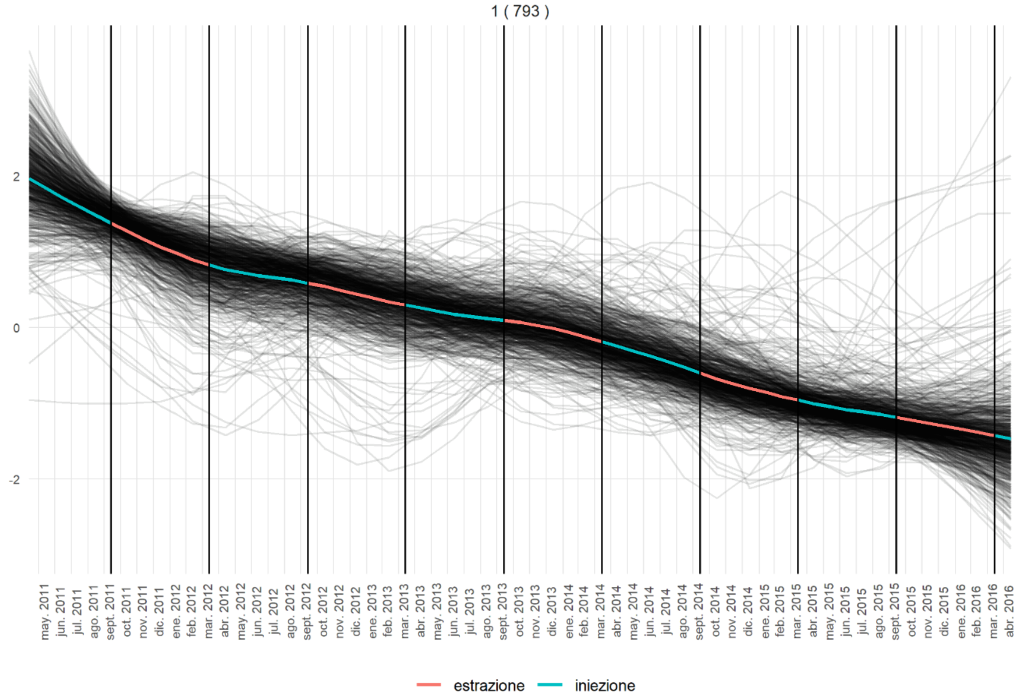


Fig. 128 - Local X-Field, 1 Class Kmean Clustering, trend component (normalized values). Author.

Normalized values clustering and absolute values plotted

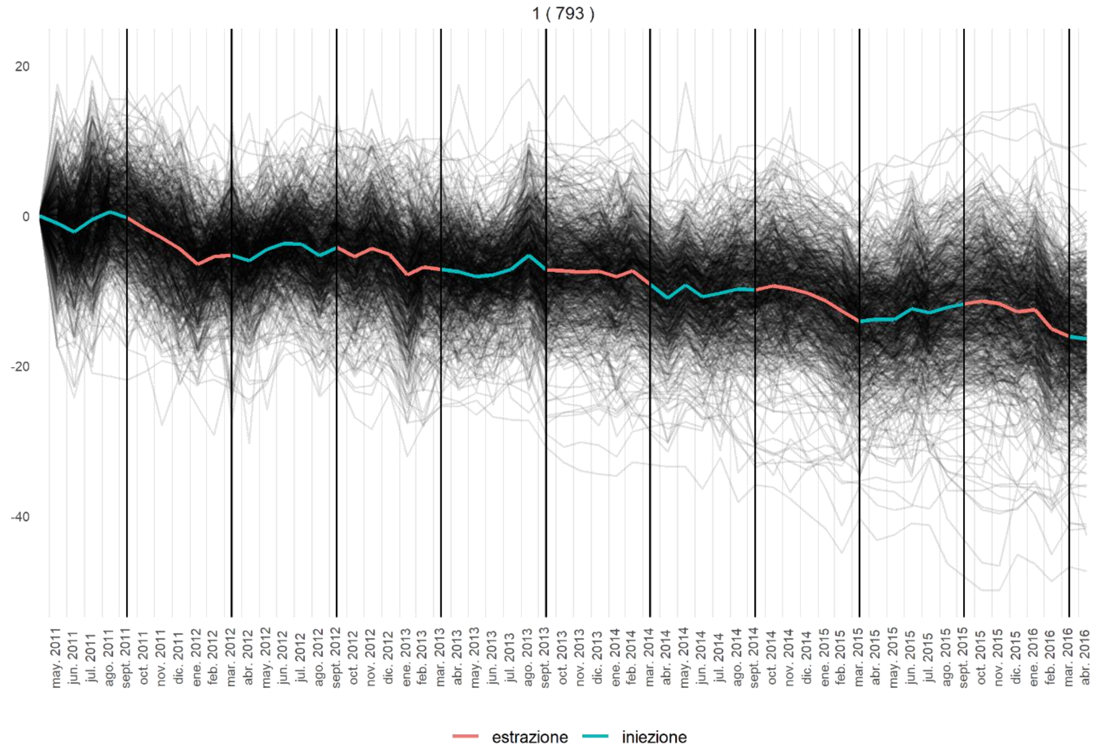


Fig. 129 - Local X-Field, 1 Class Hierarchical Clustering, trend component (resampled absolute values). Author.

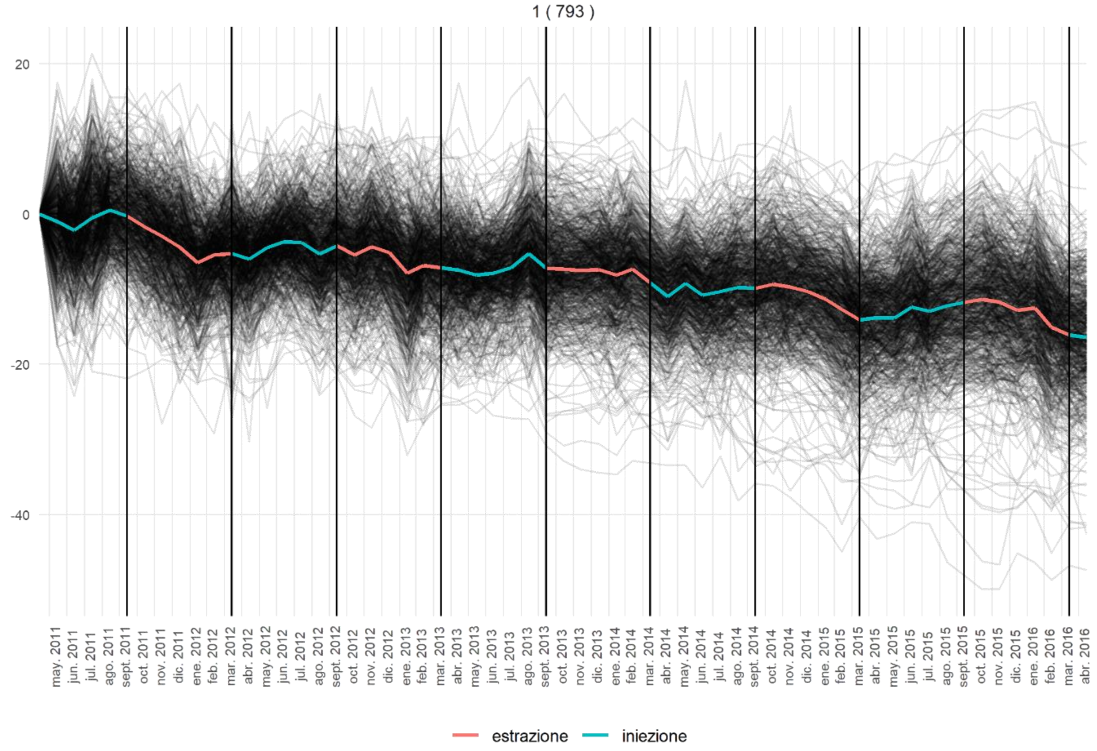


Fig. 130 - Local X-Field, 1 Class Kmean Clustering trend component (resampled absolute values). Author.

6.6.3 Areal Level

Normalized values clustering and plotting

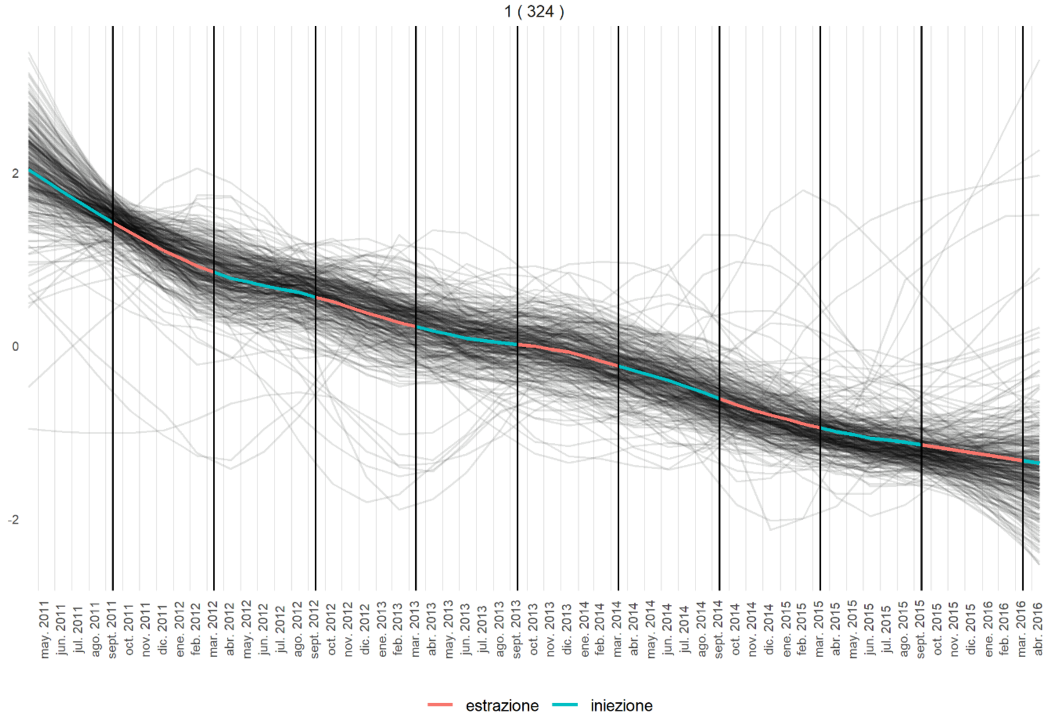


Fig. 131 - Areal X-Field, 1 Class Hierarchical Clustering, trend component (normalized values). Author.

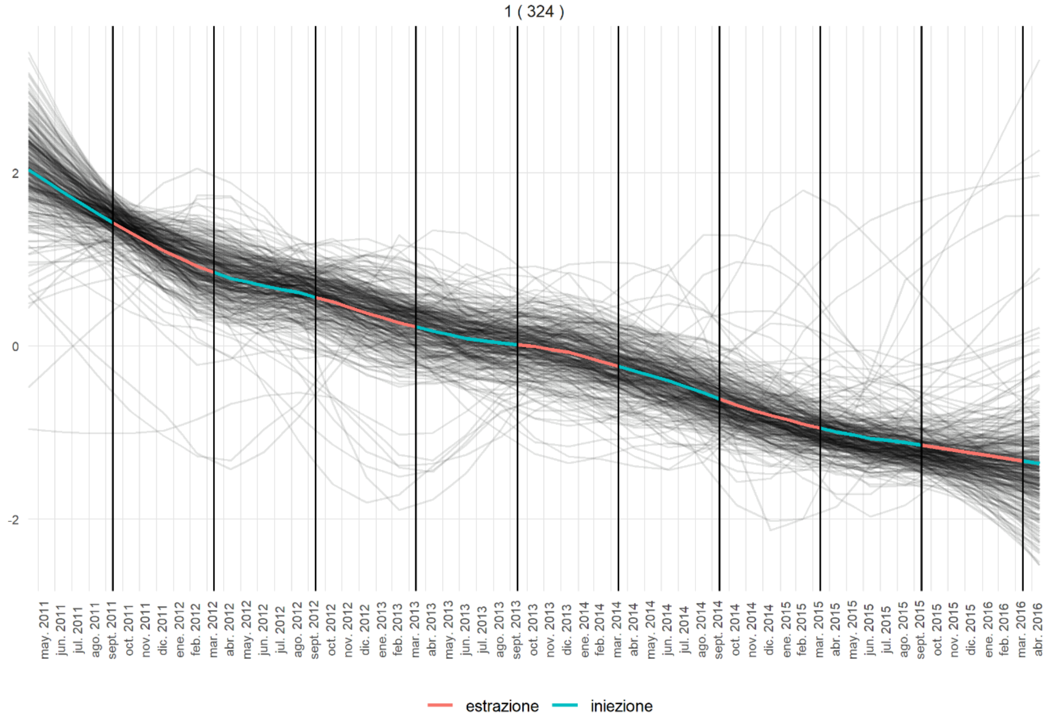


Fig. 132 - Areal X-Field, 1 Class Kmean Clustering, trend component (normalized values). Author.

Normalized values clustering and absolute values plotted

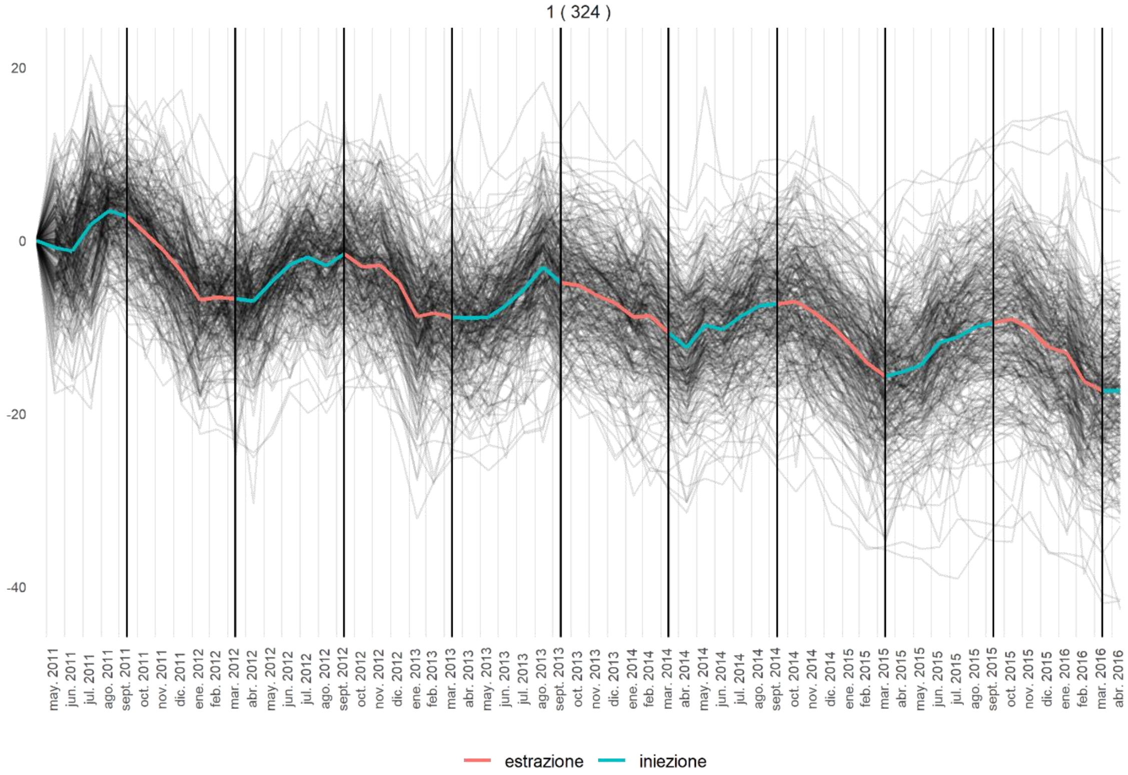


Fig. 133 - Areal X-Field, 1 Class Hierarchical Clustering, trend component (resampled absolute values). Author.

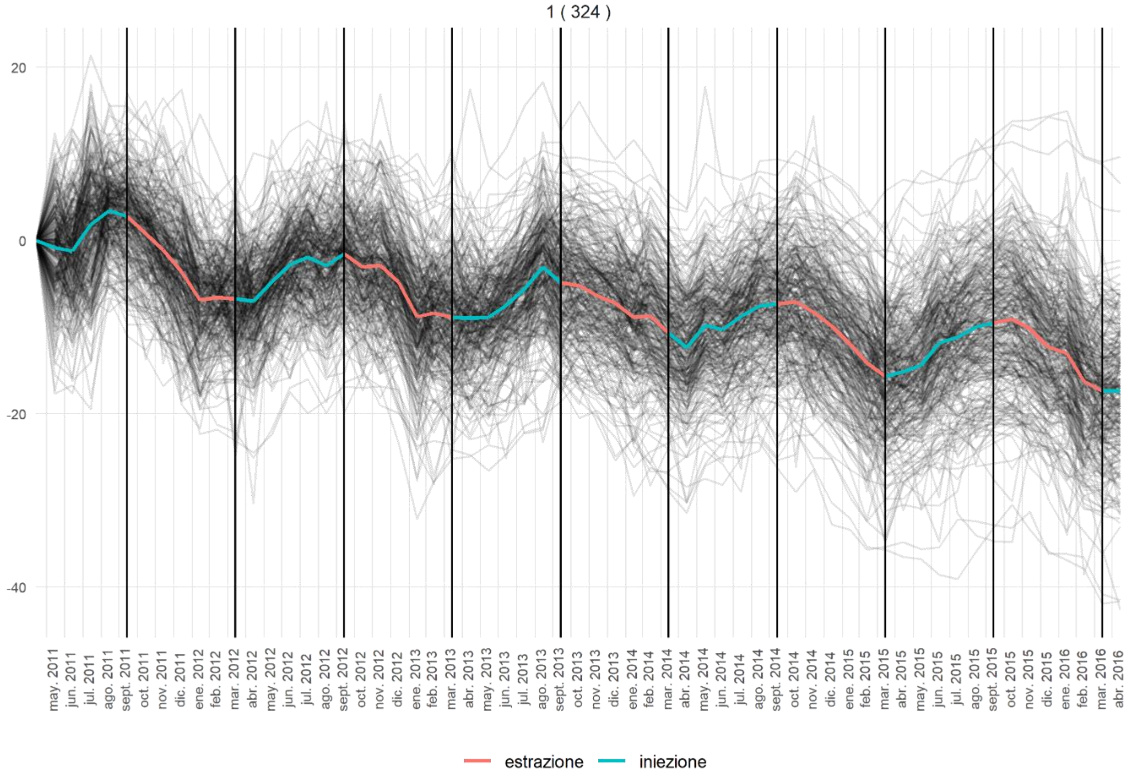


Fig. 134 - Areal X-Field, 1 Class Kmean Clustering trend component (resampled absolute values). Author.

7 RESULTS ANALYSIS

The subsection is always divided according to n-classes for clustering: 4 and 8 which were found to better depict, in terms of precision and usefulness, the subsidence seasonal patterns. Even though the suggested n-optimal algorithms like NbClust() and fviz_nbclust() with methods as wss, silhouette or gap_stat, always suggested 2 or 3 clusters for both types of this prototype based clustering approach, data showed that to many details were missing if only considered that reduced amount of clusters. Furthermore, at the same time at regional level a greater number of clusters dispersed too much the information among too many clusters of small dimensions, adding noise to the clustered obtained, as in the case of 8 clusters for the areal level analysis.

To keep analytical and geographical perspective, all the polygons considered were always presented over the map (when possible). The red arrows located on the time series clustered charts represent the classes identified as better representative of the subsidence phenomenon due to coherence between the seasonal fluid movement for the underground gas storage activities and the ground movement: uplift in summer (March - September: injection period), subsidence in winter (October-April: production period).

7.1 Regional level Seasonal analysis (16.144 MP ≈ 580 km2)

Regarding the 4 classes analysis both cases, hclust and kmean algorithm, reflects a seasonal periodicity compatible with the shape of the storage behavior, moreover the deviation of the processed data with respect to the average values rarely exceeds 2σ.

Noteworthy that the shape of the kmean clusters is smoother in relation with the hclust, and better portrait the phenomenon, as its class number 1 demonstrates grouping near 22 percent of the points (3445 MP), on which non only the shape but the time coherence is reflected, hclust class 4 with 28 percent (5015 MP) share a pseudo sinusoidal form but in both cases, there are important time forward shifts. The reason to select the number 4 as representative is given by the fact that according to GIS rendering it better encompasses the well-known UGS activity at areal level, indicating that even though the correct data is present further discrimination is needed. A 46% variation regarding the quantity of points present on these representative clusters confirms this hypothesis.

In relation to the 8 classes analysis, classes 1 and 8 both for the hclust and kmean algorithm correctly portrait the expected subsidence seasonality, both in time and space, but on the first case class 8 better depicts the correct time dependency and 1 does it for the second case. In terms of percentages hclust class 1 groups 1713 points (11%) and kmean class number 8 with 2114 measures (13%) showing a better compromise between cluster numbers and phenomenon representation, with a reduction in MP grouped of nearly 30% and refined GIS mapping.

Worthy of mention the behavior of various collection of aligned points that seems to retrieve some highway traffic-related pattern: at X-field south area classes 1 and 4 (both 4-hclust and 4-kmean), and again mainly by 1 and 2 (8-hclust) 5 and 8 but with lower point concentration. At X-field west area classes 2 and 4 (4-hclust) 2 and 3 (4-kmean), seen again by class 5 and 7 (8-hclust) and mainly 2,3 and 6 (8- kmean). For the south area case, classes 2 and 5 (8-hclust) exhibit and opposite trend to the UGS activity. In the west area case classes 2 and 3 (4-kmean) and 5 and 7 (8-hclust) share the same reverse behavior.

For the case of the class number 6 (8-hclust) there's a peak during the autumn winter period followed by an interesting nearly flat curve for the spring and summer time, as almost mirroring cluster 4 trend, indicating possible seasonal activity yet to be correlated but that shows clear impact on the subsidence phenomenon.

Finally, it is important to notice that the algorithm retrieves and clusters with success (both hclust and kmeans) the seasonal behavior affecting the area for the Local X-Field and even the iso-variation lines at the Areal X-Field which is a very strong indicator of the validity of the approach. However, it also groups into these same clusters other areas on which no stocking activity is conducted, but are affected by other seasonal activities such as for example water withdrawal. As a consequence, a crossed analysis with the integration of different information, such as InSAR data and the location and extension of UGS concession (public data available on the UNMIG website), is mandatory for a correct data interpretation. It is worthy to be mention that the location and extension of the UGS concession allows a first and roughly identification of the area in which subsidence takes place but is not the subsidence area itself.

7.2 Local level Seasonal analysis (793 MP ≈ 21 km2)

The first impression that arises is the fact that the curves are better shaped and more defined in terms of sinusoidal behavior, reflecting a better understanding of the seasonal subsidence effects for the points inside the polygon.

The 4 classes nested approach shows class number 3 as the more accurate both in terms of time and space, fully capturing the monthly variation of the injection and production cycles and the entire X-Field at mapping areal level. Same behavior is founded on class 4 of the unnested algorithm. Hierarchical class 3 number 330 MP (42%) and class 4 326 (41%) showing between them a variation of just 1,23 percent. Maps presented show, overall speaking, a well-defined grouping of data around the X-Field area, both for hclust and kmean algorithm making these the better results obtained for the present study.

The 8 classes nested approach for this case shows two main data groups of remarkable coherence both temporal and spatial: classes 4 and 7 (330 MP: 227 MP and 29% + 103 MP and 13%) being 7 the most outstanding one perfectly fitted into a tight zone over the map inside the X-Field area. For the unnested algorithm classes 5 and 7 (338MP: 171 MP and 22% + 167 MP and 21%) having again 7 the better fitting into a tight zone over the map over the X-Field area. In general terms a relative variation of 2,37% among both cluster in terms of points reflecting the UGS-related behavior.

The local case was used to test the reliability of the developed methodology. In an investigated area limited by the UGS concession boundary, the algorithm is successfully able to differentiate and cluster the points that are directly affected by the UGS operations: these points are inside the polygon defined by Codegone et al. (2016).

7.3 Areal level Seasonal analysis (324 MP ≈ 9 km2)

For this section two analysis were carried out: one clustering normalized data and one clustering raw sampled data. For the case of hclust 4 classes z-normalized data two groups, 2 and 4 emerge as most indicative of seasonal behavior (284 MP: 137 MP and 43% + 147 MP and 45%). Kmeans three instead: 1, 3 and 4 reveal as most indicative of seasonal behavior (288 MP: 56 MP and 17% + 79 MP and 24% + 153 MP and 47%). A total variation of 1,39 percent among them indicating not only method coherence but also GIS data shows good at spatial correspondence.

Plotting of absolute values, as pedagogic element, makes evident the convenience of normalize data for making subsidence analysis of this kind: even though the same clusters are reported, and the shape of the curves are consistent with UGS vertical movement behave, its less evident and harder to identify that in the normalized plots. Nevertheless, the absolute variation analysis allows discriminating different subsidence areas according to their magnitude, adding a very important piece of information for the standard subsidence analysis. The iso-subsidence lines reported in Codegone et al (2016) were again adopted to verify the reliability of the obtained results. For hclust results, series 2 and 4 and for kmean, series 1 and 4, oscillate approximately between ±15mm for production and ±12mm for injection periods (in agreement with Codegone's findings, see figs. 47 and 48), being 2 (hclust) and 4 (kmean) smoother overall speaking. Series 1 and 3 for hclust and for kmean series 2 and 3, have more reduced ranges of near 5-10mm for the same periods and near to flat shapes indicating less influence by the subsidence phenomena, as expected and in agreement with the iso-variation lines already mentioned.

With respect to the hclust 8 classes z-normalized data four groups (2, 5, 7 and 8) emerge as indicative of seasonal behavior (273 MP: 38 MP and 12% + 88 MP and 27% + 31 MP and 10% + 116 MP and 36%) but a couple of considerations are required. Even though the most coherent, both in time and space are the series 5 and 8, the decision of including both 2 and 7 relies on the fact that their shapes match, even though the class number 2 is a little spiky along the curve, but have great time coherence, and on the contrary 7 presents an observable temporal back shifting on the injection period but better shape fitting, appreciable also for the end of the production valley. As GIS mapping shows the majority of the MP belonging to class 2 are farer from the X-Field center (identified as the most active of the area), for class 7 is distributed between a tight concentration at the center and border areas, and is plausible to attribute these variations to the distance from the epicenter of the phenomenon.

Kmeans three principals instead: 4, 5 and 8 reveal as most indicative of seasonal behavior and adding also 6 and 7 (282 MP: 91 MP and 28% + 61 MP and 19% + 70 MP and 21% + 28 MP 9% + 32 MP 10%), with the exact same considerations as before for 7 and 2, respectively, from the above hclust analysis indicating good coherence between methods and validating the obtained results.

In relative terms, between 8 classes nested and unnested approaches a variation of 3,19 percent arises validating again the method despite the higher difference between UGS-representing clustered MP with respect to the 4-classes analysis.

Plotted shapes of absolute values are consistent with UGS vertical movement behave, in terms of magnitude for hclust results, series 5 and 8 and for kmean series 4, 5 and 8, oscillates approximately between ±15mm for production and ±12mm for injection periods (again in agreement with Codegone's findings), being 5 and 8 (hclust) and 5 and 8 (kmean) smoother overall speaking. Series 2 and 7 for hclust, and for kmean, series 6 and 7, present themselves as spikier and have more reduced ranges of near +-10mm for the same periods, being the rest of all the series appreciable less sinusoidal and near to flat shapes indicating lower influence by the subsidence phenomena, as expected. Special mention to the consistent downward slope for all of the plots, consistent with ARPAE findings for the Ravenna area.

7.4 Trend component analysis

In order to get a comprehensive idea of the behavior of data in the main, a single normalized data cluster was plotted for the three levels of analysis (both z-normalized and raw resampled plots were generated for magnitude quantification and behavior studies). As expected, the results where exactly the same for the hclust and kmean analysis since the centroid of the curves does not depend on the clustering technique and all clusters behaves the same when the class is equal to one on a prototype all-inclusive clustering approach.

At regional level the trend shows an increase on the subsidence phenomena, but all factors considered the slope is smooth and stable with an increment of nearly 10mm in subsidence since the beginning to the end of the studied period.

At local level the slope increases in importance reaching up to 20mm in vertical downward displacement, with a consistent trend but a pseudo seasonal spikier variation.

At areal level the slopes are similar to the local one, reaching again absolute differences of about 20mm in subsidence with even more seasonal noticeable behavior and the presence of palpable valleys related to the UGS phenomenon, all in agreement with the activities over the restricted polygon of the X-Field.

The analysis of trend component shows a great potential to detect natural subsidence trend if a sufficient amount of data could be analyzed, both in terms of time scale and area of investigation

8 CONCLUSIONS

Overall speaking the objectives set by this work of effectively representing and identifying the subsidence phenomena patterns mainly related to UGS activity where achieved, and the clusterization method demonstrated its validity under the form of the prototype based hierarchical and the kmean approaches.

The analysis is carried out at three scales: macro (regional), meso (local), micro (areal) ones. At regional scale, the algorithm groups into these same clusters other areas on which no stocking activity is conducted, but are affected by other seasonal activities such as for example water withdrawal. As a consequence, a crossed analysis with the integration of different information, such as InSAR data and the location and extension of UGS concession (public data available on the UNMIG website), is mandatory for a correct data interpretation. The local case was used to test the reliability of the developed methodology. In an investigated area limited by the UGS concession boundary, the algorithm is successfully able to differentiate and cluster the points that are directly affected by the UGS operations: these points are inside the polygon defined by Codegone et al. (2016). Finally, at areal scale, the algorithm is adopted to discriminate different subsidence sub-areas according to their magnitude, adding a very important piece of information for the standard subsidence analysis.

Some general comments can be addressed. The results show a strong coherence between the clustered seasonal data: comparison of spatial settings responds very well between k-mean versus hierarchical in terms of the amount and position of the grouped MP. In relation to the preference of selection among kmean and hclust further analysis, as no strong criterion came out at the moment to select one or another according to the obtained results.

In terms of cluster-class sensitivity, it is observed that at local and areal level (300-1.000 MP) a higher number of classes doesn't imply better understanding of the phenomenon but small numbers (as 4 on this case) are really useful to analyze the selected MP and find representative patterns for vertical displacement. Incrementing the number of them (as 8 classes viewpoint demonstrates) just augment the dispersion of patterns among forced classes without practical physical meaning, as the cluster math analysis shows. On the other hand, at regional scale (>15.000 MP) 8 classes revealed to be the proper tool to increment the precision and capacity to differentiate between several behaviors that can take places.

The smoothed behavior shown by all of the cases of the kmean algorithms, in comparison with the hclust, could possibly relies on their own nature. Whereas the second is a nested approach on which subclusters are grouped into grater ones by superimposition of data, and subsequently effects, the second determines new centroids and different groups for each n-class configuration showing more refined optimal relation between the distances of the cluster's centroids.

Some temporal shifts are presents, and this is due to the not exactly start of the different periods for injection or production and because usually it takes a little time from the formation to responds in terms of vertical movement to the fluid displacements.

A strong superposition of effects between the investigated UGS phenomena and other anthropogenic causes (such as water withdrawal, highways, activities related to urbanized area) is detected. Even if the UGS shows a very peculiar seasonality, other elements, such as water production, could superimposed on it. The possibility to add other information of the input data (such us production data for both water production and gas injection/production) could be a promising approach for a better identification of the subsidence area.

8.1 Further development of the project

Since the available literature regarding cluster analysis related to subsidence event is almost inexistent up to the present day, it brings the opportunity to pioneer into the exploration of the capabilities and possibilities the method has to offer, paving the way to a new generation of state-of-the-art subsidence studies.

In this sense, the next milestones could be the analysis of the superposition of effect related to water withdrawal data. Also, another new objective is to see if the extended regional analysis over a greater area (order of magnitude of 20 times of the actual MP universe considered) still can be analyzed by the same number of clusters and if those are able to retrieve the activity of another gas storage sites present over the area. Furthermore, the analysis of trend component shows a great potential to detect natural or anthropogenic subsidence trend if a sufficient amount of data could be analyzed, in terms of both time scale and area of investigation.

The work is the first step toward the development of an automated patterns type detection algorithm, offering the possibility to provide end user with possible scenario identification and class clustering grouping into different possible causes, anthropic or not.

9 APPENDIX

9.1 Appendix I – STL() function

Seasonal Decomposition of Time Series by Loess

Package {stats}

Description: Decompose a time series into seasonal, trend and irregular components using loess, acronym STL.

Usage

```
stl(x, s.window, s.degree = 0,
   t.window = NULL, t.degree = 1,
   l.window = nextodd(period), l.degree = t.degree,
   s.jump = ceiling(s.window/10),
   t.jump = ceiling(t.window/10),
   l.jump = ceiling(l.window/10),
   robust = FALSE,
   inner = if(robust)  1 else 2,
   outer = if(robust) 15 else 0,
   na.action = na.fail)
```

Arguments

x: univariate time series to be decomposed. This should be an object of class "ts" with a frequency greater than one.

s.window: either the character string "periodic" or the span (in lags) of the loess window for seasonal extraction, which should be odd and at least 7, according to Cleveland et al. This has no default.

s.degree: degree of locally-fitted polynomial in seasonal extraction. Should be zero or one.

t.window: the span (in lags) of the loess window for trend extraction, which should be odd. If NULL, the default, nextodd(ceiling((1.5*period) / (1-(1.5/s.window)))), is taken.

t.degree: degree of locally-fitted polynomial in trend extraction. Should be zero or one.

l.window: the span (in lags) of the loess window of the low-pass filter used for each subseries. Defaults to the smallest odd integer greater than or equal to frequency(x) which is recommended since it prevents competition between the trend and seasonal components. If not an odd integer its given value is increased to the next odd one.

l.degree: degree of locally-fitted polynomial for the subseries low-pass filter. Must be 0 or 1.

s.jump, t.jump, l.jump: integers at least one to increase speed of the respective smoother. Linear interpolation happens between every *.jumpth value.

robust: logical indicating if robust fitting be used in the loess procedure.

inner: integer; the number of ‘inner’ (backfitting) iterations; usually very few (2) iterations suffice.

outer: integer; the number of ‘outer’ robustness iterations.

na.action: action on missing values.

Details
The seasonal component is found by loess smoothing the seasonal sub-series (the series of all January values, ...); if s.window = "periodic" smoothing is effectively replaced by taking the mean. The seasonal values are removed, and the remainder smoothed to find the trend. The overall level is removed from the seasonal component and added to the trend component. This process is iterated a few times. The remainder component is the residuals from the seasonal plus trend fit.

Several methods for the resulting class "stl" objects, see, plot.stl.

value: stl returns an object of class "stl" with components

time.series: multiple time series with columns seasonal, trend and remainder.

weights: the final robust weights (all one if fitting is not done robustly).

call: the matched call.

win:integer (length 3 vector) with the spans used for the "s", "t", and "l" smoothers.

deg:integer (length 3) vector with the polynomial degrees for these smoothers.

jump:integer (length 3) vector with the ‘jumps’ (skips) used for these smoothers.

ni:number of inner iterations

no:number of outer robustness iterations

Note
This is similar to but not identical to the stl function in S-PLUS. The remainder component given by S-PLUS is the sum of the trend and remainder series from this function.

Author(s)
B.D. Ripley; Fortran code by Cleveland et al (1990) from ‘netlib’.

[Package stats version 4.1.0 Index]

9.2 Appendix II – approxfun() function

Interpolation Functions:

Package {stats}

Description: Return a list of points which linearly interpolate given data points, or a function performing the linear (or constant) interpolation.

Usage

```
approx   (x, y = NULL, xout, method = "linear", n = 50,
        yleft, yright, rule = 1, f = 0, ties = mean, na.rm = TRUE)
approxfun(x, y = NULL,       method = "linear",
        yleft, yright, rule = 1, f = 0, ties = mean, na.rm = TRUE)
```

Arguments

x, y: numeric vectors giving the coordinates of the points to be interpolated. Alternatively a single plotting structure can be specified: see xy.coords.

xout: an optional set of numeric values specifying where interpolation is to take place.

Method: specifies the interpolation method to be used. Choices are "linear" or "constant".

N:If xout is not specified, interpolation takes place at n equally spaced points spanning the interval [min(x), max(x)].

yleft::the value to be returned when input x values are less than min(x). The default is defined by the value of rule given below.

yright: the value to be returned when input x values are greater than max(x). The default is defined by the value of rule given below.

rule: an integer (of length 1 or 2) describing how interpolation is to take place outside the interval [min(x), max(x)]. If rule is 1 then NAs are returned for such points and if it is 2, the value at the closest data extreme is used. Use, e.g., rule = 2:1, if the left and right side extrapolation should differ.

f: for method = "constant" a number between 0 and 1 inclusive, indicating a compromise between left- and right-continuous step functions. If y0 and y1 are the values to the left and right of the point then the value is y0 if f == 0, y1 if f == 1, and y0*(1-f)+y1*f for intermediate values. In this way the result is right-continuous for f == 0 and left-continuous for f == 1, even for non-finite y values.

ties: handling of tied x values. The string "ordered" or a function (or the name of a function) taking a single vector argument and returning a single number or a list of both, e.g., list("ordered", mean), see ‘Details’.

na.rm: logical specifying how missing values (NA's) should be handled. Setting na.rm=FALSE will propagate NA's in y to the interpolated values, also depending on the rule set. Note that in this case, NA's in x are invalid, see also the examples.

Details: The inputs can contain missing values which are deleted (if na.rm is true, i.e., by default), so at least two complete (x, y) pairs are required (for method = "linear", one otherwise). If there are duplicated (tied) x values and ties contains a function it is applied to the y values for each distinct x value to produce (x,y) pairs with unique x. Useful functions in this context include mean, min, and max.

If ties = "ordered" the x values are assumed to be already ordered (and unique) and ties are not checked but kept if present. This is the fastest option for large length(x).

If ties is a list of length two, ties [2] must be a function to be applied to ties, see above, but if ties [1] is identical to "ordered", the x values are assumed to be sorted and are only checked for ties. Consequently, ties = list("ordered", mean) will be slightly more efficient than the default ties = mean in such a case.

The first y value will be used for interpolation to the left and the last one for interpolation to the right. Value: approx returns a list with components x and y, containing n coordinates which interpolate the given data points according to the method (and rule) desired.

The function approxfun returns a function performing (linear or constant) interpolation of the given data points. For a given set of x values, this function will return the corresponding interpolated values. It uses data stored in its environment when it was created, the details of which are subject to change.

Warning: The value returned by approxfun contains references to the code in the current version of R: it is not intended to be saved and loaded into a different R session. This is safer for R >= 3.0.0.

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) The New S Language. Wadsworth & Brooks/Cole.

[Package stats version 4.1.0 Index]

9.3 Appendix III – Full developed R script

```
######### FILE LOADING -----------------------------------------------------

library(sf)      # load shapefiles
library(readxl)  # read excel
library(ggplot2) # plot
library(plyr)
library(dplyr)   # dataframe arrangement
library(stats)   # hclust function
library(leaflet) # interactive map

######### SELECT STUDY AREA -----------------------------------------------------

#SA<- 1 #"1_Stoccaggio"

SA<- 2 #"2_Sar_Extended"

#SA<- 3 # "15C_-_PROD"

#SA<- 4 #"15A_-_INIEZ"

#SA<- 5 #"Cot_C"

######### TYPE DATA SELECTION FOR CLUSTERING AND PLOTTING

### CLUSTER RAW RESAMPLED DATA / Clusterizza raw resampled data e plotta direttamente valori assoluti
CLUST_RAW_RESAMPLED=0 #1 for SI 0 for NO

### CLUSTER Z NORMALIZED RESAMPLED DATA / Clusterizza valori normalizzati e consente di scegl x plottare
CLUST_Z_NORM_RESAMPLED=1 #1 for SI 0 for NO

  ### PLOT ABSOLUTE UNRESAMPLED VALUES
  PLOT_ABS_VALS=0 #1 for SI 0 for NO

  ### PLOT ABSOLUTE RESAMPLED CLUSTERED VALUES
  PLOT_ABS_VALS_RESAMP=1 #1 for SI 0 for NO

  ### PLOT Z NORMALIZED CLUSTERED DATA
  PLOT_Z_NORM=0 #1 for SI 0 for NO

######### TS DECOMPOSITION -----------------------------------------------------

# chose which component you want to look for "seasonal" or "trend"
#component <- "remainder"
component <- "trend"
#component <- "seasonal"

######### CLUSTER -----------------------------------------------------
# select the number of clusters you want to have and continue the analysis
```

```
N_CLUSTERS <- 1

######### WRITE EXCEL FILE CON PUNTI PER CLUSTER

WRITE_EXCEL=0 #1 for SI 0 for NO

######### PLOT INTERACTIVE MAP

PLOT_MAP=0 #1 for SI 0 for NO

######### CREATE WGS84 shapefile ARCMAP LAYER

CREATE_LAYER=0 #1 for SI 0 for NO

######### FILE LOADING -----------------------------------------------------

########## ######### ALBERTO---------------------------------------------------

############ LOAD DATI SAR PERSORCO ALBERTO

#SAR <- sf::st_read("./Dati Regione Emilia Subsidenza/dati/Dati SAR 2011-
2016/Ravenna/RAVENNA_RSAT_RSAT2_A_MAY2016_CAL_VERT_I10336A2S.shp")

if (SA==1){
# load from excel SAB ALBERTO
 Stoccaggio <- read_excel("./R/Punti_SARGIS_Stoccaggio_Urbani_Extended.xlsx","Sabbioncello")
 Data_Origin <- "Sabbioncello_Stoccaggio"
}

# load from excel SAB EXTENDED
if (SA==2){
Stoccaggio <- read_excel("./R/Punti_SARGIS_Stoccaggio_Urbani_Extended.xlsx","SABBIONCELLO SAR
EXTENDED")
Data_Origin <- "Sab_Sar_Extended"
 }

# load from excel 15C_-_PROD
if (SA==3){
Stoccaggio <- read_excel("./R/Punti_SARGIS_Stoccaggio_Urbani_Extended.xlsx","SABBIO15C - PROD")
Data_Origin <- "15C_-_PROD"
}

# load from excel 15A_-_INIEZ
 if (SA==4){
Stoccaggio <- read_excel("./R/Punti_SARGIS_Stoccaggio_Urbani_Extended.xlsx","SABBIO15A - INIEZ")
Data_Origin <- "15A_-_INIEZ"
}

# load from excel Cot_C
if (SA==5){
  Stoccaggio <- read_excel("./R/Punti_SARGIS_Stoccaggio_Urbani_Extended.xlsx","Cotignola C")
  Data_Origin <- "Cot_C"
}

######### PREPROCESSING -----------------------------------------------------
Stoccaggio_Vector <- Stoccaggio$"Punto SAR"

SAR_Stoccaggio <- dplyr::filter(SAR, SAR$CODE %in% Stoccaggio_Vector)

# remove geometry for dataframe purposes
SAR_Stoccaggio$geometry <- NULL

# create temporary variable to get only interesting variables
tmp <- SAR_Stoccaggio[,c(8:dim(SAR_Stoccaggio)[2])]

# get colnames
date <- colnames(tmp)

# create code column
code <- SAR_Stoccaggio$CODE

# transpose in column
tmp1 <- data.table::transpose(tmp, fill = NA, ignore.empty = FALSE, keep.names = NULL, make.names =
NULL)

# add data column
tmp1$Date <- date

# create vector of columns
colnames(tmp1) <- c(code, "Date")

# create column date from date parse string to date
tmp1$Date <- as.POSIXct(tmp1$Date , format = "D%Y%m%d" , tz = "GMT")

tmp2 <- tidyr::pivot_longer(tmp1, c(1:(dim(tmp1)[2]-1) ), names_to = "Point", values_to = "Values")

head(tmp2)

######### TS DECOMPOSITION -----------------------------------------------------

# initialize empty dataframe
df_decomposed <- data.frame()
approx_ts_df <- data.frame()

for (i in 1:(dim(tmp1)[2]-1)) {

  # # irregular time points at which data was sampled
  # t <- c(5,10,15,25,30,40,50)
  # # measurements
  # y <- c(4.3,1.2,5.4,7.6,3.2,1.2,3.7)
  #
  # f <- approxfun(t,y)
  #
  # # get interpolated values for time points 5, 20, 35, 50
  # f(seq(from=5,to=50,by=15))
  # [1] 4.3 6.5 2.2 3.7

  # get the actual ts
  actual_ts <- tmp1[,i]
  # plot(actual_ts, type = "l") # plot for debug

  # count the ts length
  lunghezza_ts <- length(actual_ts)

  # crete a new regular date that starts at first(tmp1[,"Date"]) and finishes at last(tmp1[,"Date"])
  # we need this to set a regular frequency one for month
  new_date <- seq(from = first(tmp1[,"Date"]) , to = last(tmp1[,"Date"]) , by = "month"  )

  # create a function that resamples the original timeseries on the new_date
  f <- approxfun(tmp1[,"Date"] ,actual_ts, method = "linear")

  approx_ts <- f(new_date)
  # plot(approx_ts, type = "l") # plot for debug

  ts_prova <- ts(approx_ts, frequency = 12, start = c(2011,5,4))
  # plot.ts(ts_prova)  # plot for debug

  # ts_prova_decomposed <- decompose(ts_prova, type = "additive")
  ts_prova_decomposed <- stl(ts_prova,s.window="periodic")

  ### CLUSTER ABSOLUTE VALUES
  if  (CLUST_RAW_RESAMPLED==1){
  # column_to_add <- as.data.frame(ts_prova_decomposed[[component]])
  column_to_add <- as.data.frame(  ts_prova_decomposed$time.series[,component])
  }

  ### CLUSTER Z NORMALIZED DATA
  if  (CLUST_Z_NORM_RESAMPLED==1){
  # column_to_add <- as.data.frame(ts_prova_decomposed[[component]])
  column_to_add <- as.data.frame(  ts_prova_decomposed$time.series[,component])

  # normalize column min max
  #column_to_add <- (column_to_add$x - min(column_to_add$x, na.rm = T) ) / ( max(column_to_add$x,
na.rm = T)  - min(column_to_add$x, na.rm = T) )
  # normalize column z-score
  column_to_add <- (column_to_add$x - mean(column_to_add$x, na.rm = T) ) / sd(column_to_add$x, na.rm
= T)
  column_to_add <- as.data.frame(column_to_add)
  }

  if  (i==1){
    df_decomposed <- column_to_add
    colnames(df_decomposed)[i] <- colnames(tmp1)[i]

  }else{
    df_decomposed <- cbind(df_decomposed,column_to_add )
    colnames(df_decomposed)[i] <- colnames(tmp1)[i]
  }

}

if  (CLUST_Z_NORM_RESAMPLED==1 && PLOT_ABS_VALS_RESAMP==1){

  for (i in 1:(dim(tmp1)[2]-1)) {

    # get the actual ts
    actual_ts <- tmp1[,i]
    # plot(actual_ts, type = "l") # plot for debug

    # count the ts length
    lunghezza_ts <- length(actual_ts)

    # crete a new regular date that starts at first(tmp1[,"Date"]) and finishes at last(tmp1[,"Date"])
    # we need this to set a regular frequency one for month
    new_date <- seq(from = first(tmp1[,"Date"]) , to = last(tmp1[,"Date"]) , by = "month"  )

    # create a function that resamples the original timeseries on the new_date
    f <- approxfun(tmp1[,"Date"] ,actual_ts, method = "linear")

    approx_ts <- f(new_date)
    # plot(approx_ts, type = "l") # plot for debug

    column_add <- as.data.frame(approx_ts)

    if  (i==1){
      approx_ts_df <- column_add
      colnames(approx_ts_df)[i] <- colnames(tmp1)[i]

    }else{
      approx_ts_df <- cbind(approx_ts_df,column_add )
      colnames(approx_ts_df)[i] <- colnames(tmp1)[i]
    }
  }

  # extract new date and add info to df_decomposed
  date <- new_date
  approx_ts_df <- cbind(approx_ts_df, date) %>% na.omit()
  head(approx_ts_df)

  }

# extract new date and add info to df_decomposed
date <- new_date
df_decomposed <- cbind(df_decomposed, date) %>% na.omit()
head(df_decomposed)

# pivot only for plot purpose
```

```
df_pivot_longer_seasonality <- tidyr::pivot_longer(df_decomposed, c(1:(dim(df_decomposed)[2]-1) ),
names_to = "Point", values_to = "Values")


######### CLUSTER ------------------------------------------------------

# riorganizzo il dataframe in orizzontale per effettuare clustering di profilo
spreaded_df <- tidyr::pivot_wider(df_pivot_longer_seasonality, names_from = "date", values_from =
"Values")

# considero solamente le osservzioni (escludo ID prima colonna)
data <- spreaded_df[2:dim(spreaded_df)[2]]

# set row names
rownames(data) <- spreaded_df$Point


############ CLUSTER GERARCHICO
diss_matrix <- dist(data, method = "euclidean")          # matrice dell distanze euclidee

hcl <- hclust(diss_matrix, method = "ward.D2") #cluster gererchico (agglomerativo) con metodo ward Metodi:
ward.D2, single, complete, average

# add hierarchical cluster information to the dataframe

spreaded_df$clusterH <- as.factor(paste(cutree(hcl, N_CLUSTERS)) ) # Diciamo quanti cluster vogliamo
tagliando il dendrogamma (creiamo una nuova feature in df1 con il numero del cluster per ciascun giorno)


############ CLUSTER PARTITIVO KMEANS
set.seed(1) # 1
clustpart <- stats::kmeans(data, N_CLUSTERS, iter.max = 500, nstart=1) #cluster partitivo euclidean

# add hierarchical cluster information to the dataframe
spreaded_df$clusterK <- as.factor(clustpart$cluster)


########## PLOT GGPLOT

# riporta l'informazione al dataset origiunale
df_pivot_longer_seasonality1 <- spreaded_df %>%
  tidyr::pivot_longer(c(2:(dim(spreaded_df)[2]-2)), names_to = "date", values_to = "Values")

#****************WRITE EXCEL WITH ONLY CLUSTER POINT NUMBER***********
if  (WRITE_EXCEL==1){

  if  (CLUST_Z_NORM_RESAMPLED==1 && PLOT_Z_NORM==1){
      ADD <- "Z Norm Resamp Values + Z Norm Values"
  }
  if (CLUST_Z_NORM_RESAMPLED==1 && PLOT_ABS_VALS==1){
      ADD <- "Z Norm Resamp Values + Abs Values"
  }
  if (CLUST_Z_NORM_RESAMPLED==1 && PLOT_ABS_VALS_RESAMP==1){
    ADD <- "Z Norm Resamp Values + Abs Resamp Values"
  }
  if (CLUST_RAW_RESAMPLED==1){
    ADD <- "Raw Resamp Data"
  }

  library("xlsx")
  write.xlsx(df_pivot_longer_seasonality1 %>% distinct(Point, .keep_all = TRUE), file = paste0("./R/Scripts/Stoccaggi Regione/Analisi Clusters Plots/EXCEL DATI CLUSTERS DECOMPOSITIONS/","Punti_",Data_Origin,"_H_K_",N_CLUSTERS,"_Clusters_",component,"_colored_",ADD,".xlsx"), sheetName = paste(N_CLUSTERS,"Clust"), append = FALSE)

  #library("readr")
  #write_csv(df_pivot_longer_seasonality1, path = paste0("./R/Scripts/Stoccaggi Regione/Analisi Clusters Plots/EXCEL DATI CLUSTERS DECOMPOSITIONS/","Punti_",Data_Origin,"_H_K_",N_CLUSTERS,"_Clusters_",component,".csv"))

  ##########
}
#################### PLOTTING RAW OR Z NORM DATA

#by default plotta i valori decomposed clusterizzati, sia raw che norm, qui scegliamo se plottare
#quelli decomposed o gli assoluti non resampled rispettando i clusters

if  (CLUST_RAW_RESAMPLED==1){

  df_plot <- df_pivot_longer_seasonality1 %>% arrange(Point)
  df_ploty <- df_pivot_longer_seasonality1 %>% arrange(Point)
  ADD <- ADD <- "Raw Resamp Values"
  }

if  (CLUST_Z_NORM_RESAMPLED==1 && PLOT_Z_NORM==1){

  df_plot <- df_pivot_longer_seasonality1 %>% arrange(Point)
  df_ploty <- df_pivot_longer_seasonality1 %>% arrange(Point)
    ADD <- "Z Norm Resamp Values + Z Norm Values"
}

if  (CLUST_Z_NORM_RESAMPLED==1 && PLOT_ABS_VALS_RESAMP==1){

  tmprs <- tidyr::pivot_longer(approx_ts_df, c(1:(dim(approx_ts_df)[2]-1) ), names_to = "Point", values_to = "Values")

  df_pivot_longer_seasonalityRedc <- df_pivot_longer_seasonality1 %>% distinct(Point, .keep_all = TRUE)

  tmprs$clusterH <- NA_character_
  tmprs$clusterK <- NA_character_
  for( row in seq_len( nrow( df_pivot_longer_seasonalityRedc ) ) ) {
    tmprs$clusterH[ substr( tmprs$Point, 0, nchar( df_pivot_longer_seasonalityRedc$Point[row] ) ) == df_pivot_longer_seasonalityRedc$Point[row] ] <- df_pivot_longer_seasonalityRedc$clusterH[row]
    #tmprs$clusterK[ substr( tmprs$Point, 0, nchar( df_pivot_longer_seasonalityRedc$Point[row] ) ) ==
df_pivot_longer_seasonalityRedc$Point[row] ] <- df_pivot_longer_seasonalityRedc$clusterK[row]
  }
  for( row in seq_len( nrow( df_pivot_longer_seasonalityRedc ) ) ) {
    #tmprs$clusterH[ substr( tmprs$Point, 0, nchar( df_pivot_longer_seasonalityRedc$Point[row] ) ) ==
df_pivot_longer_seasonalityRedc$Point[row] ] <- df_pivot_longer_seasonalityRedc$clusterH[row]
    tmprs$clusterK[ substr( tmprs$Point, 0, nchar( df_pivot_longer_seasonalityRedc$Point[row] ) ) ==
df_pivot_longer_seasonalityRedc$Point[row] ] <- df_pivot_longer_seasonalityRedc$clusterK[row]
  }
  colnames(tmprs)<-c('date','Point','Values','clusterH','clusterK')

  df_plot <- tmprs %>% arrange(Point)
  df_ploty <- tmprs %>% arrange(Point)
  date <- colnames(data)
  ADD <- "Z Norm Resamp Values + Abs Resamp Values"
}

if (CLUST_Z_NORM_RESAMPLED==1 && PLOT_ABS_VALS==1){

  tmp22 <- tmp2

  df_pivot_longer_seasonalityRedc <- df_pivot_longer_seasonality1 %>% distinct(Point, .keep_all = TRUE)

  tmp22$clusterH <- NA_character_
  tmp22$clusterK <- NA_character_
  for( row in seq_len( nrow( df_pivot_longer_seasonalityRedc ) ) ) {
    tmp22$clusterH[ substr( tmp22$Point, 0, nchar( df_pivot_longer_seasonalityRedc$Point[row] ) ) ==
df_pivot_longer_seasonalityRedc$Point[row] ] <- df_pivot_longer_seasonalityRedc$clusterH[row]
    #tmp22$clusterK[ substr( tmp22$Point, 0, nchar( df_pivot_longer_seasonalityRedc$Point[row] ) ) ==
df_pivot_longer_seasonalityRedc$Point[row] ] <- df_pivot_longer_seasonalityRedc$clusterK[row]
  }
  for( row in seq_len( nrow( df_pivot_longer_seasonalityRedc ) ) ) {
    #tmp22$clusterH[ substr( tmp22$Point, 0, nchar( df_pivot_longer_seasonalityRedc$Point[row] ) ) ==
df_pivot_longer_seasonalityRedc$Point[row] ] <- df_pivot_longer_seasonalityRedc$clusterH[row]
    tmp22$clusterK[ substr( tmp22$Point, 0, nchar( df_pivot_longer_seasonalityRedc$Point[row] ) ) ==
df_pivot_longer_seasonalityRedc$Point[row] ] <- df_pivot_longer_seasonalityRedc$clusterK[row]
  }
    colnames(tmp22)<-c('date','Point','Values','clusterH','clusterK')

    df_plot <- tmp22 %>% arrange(Point)
    df_ploty <- tmp22 %>% arrange(Point)
    date <- colnames(tmp)
    ADD <- "Z Norm Resamp Values + Abs Values"
 }

# add info on how many ts are in the given cluster
countK <- df_plot %>%
  dplyr::group_by(clusterK) %>%
  dplyr::summarise(n_K = n()) %>%
  dplyr::mutate(n_K = n_K/length(date))

# add info on how many ts are in the given cluster
```

```
countH <- df_plot %>%
  dplyr::group_by(clusterH) %>%
  dplyr::summarise(n_H = n()) %>%
  dplyr::mutate(n_H = n_H/length(date))

# merge info to overall dataframe
df_merged1 <- merge.data.frame(countK, df_plot)
df_plot <- merge.data.frame(countH, df_merged1) %>%
  dplyr::mutate(clusterK = paste(clusterK, "(", n_K, ")"),
              clusterH = paste(clusterH, "(", n_H, ")")) %>%
  dplyr::mutate(date = as.POSIXct(date)) %>%
    dplyr::select(-n_K, -n_H)

# centroide gerarchico
df_centroideH <- df_plot %>%
  dplyr::group_by(clusterH, date) %>%
  dplyr::summarise(Values = mean(Values))

df_centroideH$Point <- "centroidH"
df_centroideH <- df_centroideH %>%
  dplyr::mutate(
    month = lubridate::month(date),
    regime = as.factor( if_else(month %in% c(4,5,6,7,8,9), "iniezione", "estrazione") )
  ) %>%
  dplyr::select(Point, clusterH, date, Values, regime)

# centroide partitivo
df_centroideK <- df_plot %>%
  dplyr::group_by(clusterK, date) %>%
  dplyr::summarise(Values = mean(Values))

df_centroideK$Point <- "centroidK"
df_centroideK <- df_centroideK %>%
  dplyr::mutate(
    month = lubridate::month(date),
    regime = as.factor( if_else(month %in% c(4,5,6,7,8,9), "iniezione", "estrazione") )
  ) %>%
  dplyr::select(Point, clusterK, date, Values, regime)

#df_pivot_longer_seasonality2 <- rbind(df_pivot_longer_seasonality1, df_centroide)

if  (N_CLUSTERS<=4){
  NUMCOL<-1
} else {
  NUMCOL<-2
}

dev.new()
ggplot() +
  geom_line(data = df_plot, aes(x = date, y = Values, group = Point), alpha = 0.1 ) +
  geom_line(data = df_centroideK, aes(x = date, y = Values, group = Point, color = regime), size = 1 ) +
  geom_vline(xintercept= c(1318550400, 1334361600, 1350172800, 1365897600, 1381708800, 1397433600,
1413244800, 1428969600, 1444780800, 1460592000))+
  labs(x = NULL, y = NULL, title = paste(Data_Origin,"Kmeans", N_CLUSTERS,"clusters component",
component,ADD) ) +
  facet_wrap(~clusterK, ncol = NUMCOL,  strip.position = "top")  +
  scale_x_datetime(
    breaks = scales::date_breaks("1 month"),                    # specify breaks month
    labels = scales::date_format(("%b %Y") , tz = "Etc/GMT+12"),# specify format of labels messe anno
    expand = c(0,0)                                  # espande l'asse x affinche riempia tutto il box in orizzontale
  ) +
  theme_minimal() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        plot.title = element_text(size = 14,margin = margin(t = 0, r = 0, b = 5, l = 0)),
        plot.subtitle = element_text(size = 12,margin = margin(t = 0, r = 0, b = 10, l = 0)),
        # x axis
        axis.title.x = element_text(size = 9, margin = margin(t = 20, r = 20, b = 0, l = 0)),
        axis.text.x = element_text(size = 7, angle = 90, hjust = 0.5),
        panel.grid.major.x = element_line(color = "gray90", linetype = 1, size = 0.3),
        # y axis
        axis.title.y = element_text(size = 9, margin = margin(t = 20, r = 20, b = 0, l = 0)),
        axis.text.y = element_text(size = 7),
        panel.grid.major.y = element_line(color = "gray90", linetype = 1, size = 0.3),
        # legend
        legend.title = element_blank(),
        legend.position = "bottom"
        )

ggsave( paste0("./R/Scripts/Stoccaggi Regione/Analisi Clusters Plots/PLOTS/",Data_Origin,"_Kmeans-",
N_CLUSTERS,"_Clusters_", component,"_colored_",ADD,".png"), width = 8, height = 6 )
#dev.off()

dev.new()
ggplot() +
  geom_line(data = df_plot, aes(x = date, y = Values, group = Point), alpha = 0.1 ) +
  geom_line(data = df_centroideH, aes(x = date, y = Values, group = Point, color = regime), size = 1 ) +
  geom_vline(xintercept= c(1318550400, 1334361600, 1350172800, 1365897600, 1381708800, 1397433600,
1413244800, 1428969600, 1444780800, 1460592000))+
  labs(x = NULL, y = NULL, title = paste(Data_Origin,"Hierarchical", N_CLUSTERS,"clusters component",
component,ADD) ) +
  facet_wrap(~clusterH, ncol = NUMCOL,  strip.position = "top")  +
  scale_x_datetime(
    breaks = scales::date_breaks("1 month"),                    # specify breaks every month
    labels = scales::date_format(("%b %Y") , tz = "Etc/GMT+12"),# specify format of labels messe anno
    expand = c(0,0)                                  # espande l'asse x affinche riempia tutto il box in orizzontale
  ) +
  theme_minimal() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        plot.title = element_text(size = 14,margin = margin(t = 0, r = 0, b = 5, l = 0)),
        plot.subtitle = element_text(size = 12,margin = margin(t = 0, r = 0, b = 10, l = 0)),
        # x axis
        axis.title.x = element_text(size = 9, margin = margin(t = 20, r = 20, b = 0, l = 0)),
        axis.text.x = element_text(size = 7, angle = 90, hjust = 0.5),
        panel.grid.major.x = element_line(color = "gray90", linetype = 1, size = 0.3),
        # y axis
        axis.title.y = element_text(size = 9, margin = margin(t = 20, r = 20, b = 0, l = 0)),
        axis.text.y = element_text(size = 7),
        panel.grid.major.y = element_blank(),
        # legend
        legend.title = element_blank(),
        legend.position = "bottom"
  )

ggsave( paste0("./R/Scripts/Stoccaggi Regione/Analisi Clusters Plots/PLOTS/",Data_Origin,"_Hierarchical_",
N_CLUSTERS,"_Clusters_", component,"_colored_",ADD,".png"), width = 8, height = 6 )
#dev.off()

--------------------------- MAPPING PLOTTING / LAYERIONG DATA CREATION ---------------
tmppp <- df_plot %>%
  dplyr::mutate(CODE = as.factor(Point)) %>%
  dplyr::select(-date, - Values, - Point) %>%
  unique()

SAR_Stoccaggio_map <- SAR %>%
  dplyr::filter(CODE %in% Stoccaggio_Vector) %>%
  dplyr::mutate(CODE = as.factor(CODE)) %>%
  dplyr::select(CODE)

merged_map <- merge.data.frame(tmppp, SAR_Stoccaggio_map) %>%
  dplyr::select( dplyr::everything(), geometry)

merged_map$lat <- 0
merged_map$lon <- 0
for (i in 1:dim(merged_map)[1]) {
  merged_map$lat[i] <- as.numeric(merged_map$geometry[[i]])[2]
  merged_map$lon[i] <- as.numeric(merged_map$geometry[[i]])[1]
}

#----------------------------------------------------------------------------------

if (PLOT_MAP==1){
########## PLOT ,
# set map default view
default_view <- list(
  center_lat = 45.06402941,
  center_lng = 7.66043533,
  def_zoom = 16,
  def_min_zoom = 15,
  def_max_zoom = 19
)

# tmppp <- df_plot %>%
#   dplyr::mutate(CODE = as.factor(Point)) %>%
#   dplyr::select(-date, - Values, - Point) %>%
#   unique()
#
# SAR_Stoccaggio_map <- SAR %>%
#   dplyr::filter(CODE %in% Stoccaggio_Vector) %>%
#   dplyr::mutate(CODE = as.factor(CODE)) %>%
#   dplyr::select(CODE)
#
# merged_map <- merge.data.frame(tmppp, SAR_Stoccaggio_map) %>%
#   dplyr::select( dplyr::everything(), geometry)

# merged_map$lat <- 0
# merged_map$lon <- 0
# for (i in 1:dim(merged_map)[1]) {
#   merged_map$lat[i] <- as.numeric(merged_map$geometry[[i]])[2]
#   merged_map$lon[i] <- as.numeric(merged_map$geometry[[i]])[1]
# }

# remove geometry
merged_map$geometry <- NULL

paletteH <- colorFactor(c("Dark2"), domain = unique(merged_map$clusterH))

leaflet() %>%
  addProviderTiles(providers$OpenStreetMap,
                   options = list(opacity = 0.5)) %>%
  leaflet.extras::setMapWidgetStyle(list(background = "white"))  %>%
  addCircleMarkers(data = merged_map,
                   lat = ~lat,
                   lng = ~lon,
                   color = ~paletteH(clusterH),
                   stroke = FALSE,
                   fillOpacity = 0.2,
                   label = paste(merged_map$CODE, "(", merged_map$clusterH, ")"),
                   layerId = merged_map$CODE,
                   labelOptions = labelOptions(
                     noHide = F,
                     direction = "bottom",
                     style = list(
                       "color"        = "black",
                       "font-family"  = "helvetica",
                       "box-shadow"   = "3px 3px rgba(0,0,0,0.25)",
                       "font-size"    = "12px",
                       "border-color" = "rgba(0,0,0,0.5)"
                     )
                   )
  )%>%
  addLegend(
    data      = merged_map,
    pal       = paletteH,
    title     = 'Hierarchical',
    values    = ~ (merged_map$clusterH),
    position  = 'topright'
  )

paletteK <- colorFactor(c("Dark2"), domain = unique(merged_map$clusterK))

leaflet() %>%
  addProviderTiles(providers$OpenStreetMap,
                   options = list(opacity = 0.5)) %>%
  leaflet.extras::setMapWidgetStyle(list(background = "white"))  %>%
  addCircleMarkers(data = merged_map,
                   lat = ~lat,
                   lng = ~lon,
                   color = ~paletteK(clusterK),
                   stroke = FALSE,
                   fillOpacity = 0.2,
                   label = paste(merged_map$CODE, "(", merged_map$clusterK, ")"),
                   layerId = merged_map$CODE,
                   labelOptions = labelOptions(
                     noHide = F,
                     direction = "bottom",
                     style = list(
                       "color"        = "black",
                       "font-family"  = "helvetica",
                       "box-shadow"   = "3px 3px rgba(0,0,0,0.25)",
                       "font-size"    = "12px",
                       "border-color" = "rgba(0,0,0,0.5)"
                     )
                   )
  )%>%
  addLegend(
    data      = merged_map,
    pal       = paletteK,
    title     = 'Kmeans',
    values    = ~ (merged_map$clusterK),
    position  = 'topright'
  )
}

if  (CREATE_LAYER==1){

  ############################## LAYERING SHAPEFILE ----------####################

  library(tidyverse)

  df_ploty<- df_ploty %>% distinct(Point, .keep_all = TRUE)

  MAXROW<-nrow(df_ploty)

  df_pivot_longer_seasonality11<- df_ploty %>% slice(1:MAXROW)


  library(sp)
  library(sf)
  library(rgdal)
  library(proj4)

  #------------------------- CREATE DATAFRAME WITH POINTS AND COORDINATES

  merged_map_layer<- data.frame
(merged_map$CODE,merged_map$lon,merged_map$lat,df_pivot_longer_seasonality11$clusterH,df_pivot_l
onger_seasonality11$clusterK)

  colnames(merged_map_layer) <- c("CODE","LON","LAT","H_Clust","K_Clust")

  rownames(merged_map_layer)<-merged_map$CODE

  merged_map_layer_coords<- data.frame (merged_map_layer$LON,merged_map_layer$LAT)

  colnames(merged_map_layer_coords) <- c("LON","LAT")

  #------------------------- ADD SPATIAL INFO

  GCS_WGS_1984 <- CRS("+init=epsg:4326 +proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs")

  coordinates(merged_map_layer_coords)

  SpatialPoints(merged_map_layer_coords)

  merged_layer<- SpatialPointsDataFrame(merged_map_layer_coords, merged_map_layer,proj4string =
GCS_WGS_1984)

  #------------------------- CREATE WGS84 LAYER

  writeOGR(merged_layer, 'D:/Documents/Italia/Politecnico/3-Terso Anno/Tesi/E -
Elaborati/R/Scripts/Stoccaggi Regione/Analisi Clusters Plots/LAYERS COLORATI PER SEASONS',
layer=paste0(Data_Origin,'_H_K_',N_CLUSTERS,"_Clust_", component,"_colored_",ADD), driver="ESRI
Shapefile",overwrite_layer=TRUE)

  #####
}
```

10 REFERENCES

A. Ferretti, A. Fumagalli, F. Novali, C. Prati, F. Rocca and A. Rucci. (2011) "A New Algorithm for Processing Interferometric Data-Stacks: SqueeSAR," in IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no. 9, pp. 3460-3470, Sept., doi: 10.1109/TGRS.2011.2124465.

Alvarenga R., Manhães A., Mattos G., Regis G. (2016). A mathematical model and a Clustering Search metaheuristic for planning the helicopter transportation of employees to the production platforms of oil and gas. Computers & Industrial Engineering, vol. 101. https://doi.org/10.1016/j.cie.2016.09.006.

ARPA - Agenzia Regionale per la Prevenzione e l'Ambiente – Emilia-Romagna. (2017). Regione Emilia-Romagna. Rilievo della subsidenza nella pianura emiliano-romagnola 2016 - Prima Fase.

ARPA - Agenzia Regionale per la Prevenzione e l'Ambiente – Emilia-Romagna. (2018 Regione Emilia-Romagna. Rilievo della subsidenza nella pianura emiliano-romagnola 2017 - Seconda fase - Relazione finale.

Baldi, P et al. (2006). Approccio multi-disciplinare al problema della subsidenza nella regione Emilia -Romagna. Doi: http://hdl.handle.net/2122/2671

Beckwith, H., Slemmons, B., Weeks, E. (1991) Use of Low-Sun Angle Photography for Identification of Subsidence Induced Earth Fissures in Land Subsidence, in: Land Subsidence, edited by: Johnson, A. I., Proceedings of the Fourth International Symposium on Land Subsidence, 12–17 May 1991, Arvada, CO, Oxfordshire, UK: International Association of Hydrological Sciences Press, IAHS Publication No. 200, 261–269, 1991.

Bhaskaran P., Chennippan M., Subramaniam T. (2020). Future prediction & estimation of faults occurrences in oil pipelines by using data clustering with time series forecasting. Journal of Loss Prevention in the Process Industries, vol. 66. https://doi.org/10.1016/j.jlp.2020.104203.

Biot, M. (1941) General Theory of Three-Dimensional Consolidation. Journal of Applied Physics, Vol 12, Number (2). 155-164 doi:10.1063/1.1712886

Bitelli G., Bonsignore F., Vittuari L. (2008). Il monitoraggio della subsidenza in Emilia-Romagna: risultati recenti. Atti 12a Conferenza Nazionale ASITA, L'Aquila 21–24 ottobre, 453-458.

Boesen T., Haber E., Hoversten M. (2021). Data-driven semi-supervised clustering for oil prediction. Computers & Geosciences, vol. 148. https://doi.org/10.1016/j.cageo.2020.104684.

Carminati E., Di Donato G. (1999). Separating natural and anthropogenic vertical movements in fast subsiding areas: the Po plain (N. Italy) case. Geophysical Research Letters, vol. 26, no. 15, pages 2291-2294. doi: doi:10.1029/1999gl900518

Codegone, G., Rocca, V., Verga, F. et al. Subsidence Modeling Validation Through Back Analysis for an Italian Gas Storage Field. Geotech Geol Eng 34, 1749–1763 (2016). https://doi.org/10.1007/s10706-016-9986-9

Cremaschi S., Shin J., Subramani H. (2015). Data clustering for model-prediction discrepancy reduction – A case study of solids transport in oil/gas pipelines, Computers & Chemical Engineering, vol. 81. https://doi.org/10.1016/j.compchemeng.2015.04.027.

Dake L.P. (1983). Fundamentals of Reservoir Engineering. First edition. Elsevier Sciences.

E. Carminati, G. Martinelli. (2002). Subsidence rates in the Po Plain, northern Italy: the relative impact of natural and anthropogenic causation. Engineering Geology, vol. 66, Issues 3–4, Pages 241-255. DOI: https://doi.org/10.1016/S0013-7952(02)00031-5.

Fabris M. et al. (2014). Valutazione della subsidenza nell'area di Ravenna tramite un approccio integrato InSAR/livellazione classica. ASITA 2014.

Fergason K., Rucker M., Panda B. (2015). Methods for monitoring land subsidence and earth fissures in the Western USA. DOI/ISBN: 10.5194/piahs-372-361-2015

Fish Bureau USA. (1981). An Introduction to the Environmental Literature of the Mississippi Deltaic Plain Region.

Galloway, Devin & Burbey, Thomas. (2011). Review: Regional land subsidence accompanying groundwater extraction. Hydrogeology Journal - HYDROGEOL J. 19. 10.1007/s10040-011-0775-5.

Gambolati, G. (1975). Numerical models in land subsidence control. Computer Methods in Applied Mechanics and Engineering. Volume 5. Issue 2. Pages 227-237. ISSN 0045-7825 Doi: https://doi.org/10.1016/0045-7825(75)90054-7.

Gambolati, G. & Teatini, Pietro & Ferronato, Massimiliano. (2006). Anthropogenic Land Subsidence. Encyclopedia of Hydrological Sciences. John Wiley & Sons, Ltd. DOI: 10.1002/0470848944.hsa164b.

Gambolati, G., Ricceri, G., Bertoni, W., Brighenti, G., and Vuillermin, E. (1991), Mathematical Simulation of the Subsidence of Ravenna, Water Resour. Res., 27(11), 2899– 2918. Doi:10.1029/91WR01567.

Geertsma, J. (1973). Land Subsidence Above Compacting Oil and Gas Reservoirs. J-Pet Technol. Number 25. Pages: 734–744. Doi: https://doi.org/10.2118/3730-PA

Ghielmi M., Minervini M., Nini C., Rogledi S., Rossi M., Vignolo A. (2010) Sedimentary and tectonic evolution in the eastern Po-Plain and northern Adriatic Sea area from Messinian to Middle Pleistocene (Italy). Rend Fis Acc Lincei 21(Suppl 1): S131–S166. doi:10.1007/s12210-010-0101-5

Giani, G., Gotta, A., Marzano, F., Rocca, V. (2017). How to Address Subsidence Evaluation for a Fractured Carbonate Gas Reservoir Through a Multi-disciplinary Approach. Geotechnical and Geological Engineering. 35. 10.1007/s10706-017-0296-7.

Goldstein, R., Zebker, H., and Werner, C. (1988), Satellite radar interferometry: Two-dimensional phase unwrapping, Radio Sci., 23(4), 713– 720, doi:10.1029/RS023i004p00713.

Gongjie Zhang, Shijian Lu and Wei Zhang. (2019). CAD-Net: A Context-Aware Detection Network for Objects in Remote Sensing Imagery. IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 12. doi: 10.1109/tgrs.2019.2930982

Guo, X., Chen, W., Yu, J. (2018). Combined Effect of Vertical and Horizontal Ground Motions on Failure Probability of RC Chimneys. Advances in Civil Engineering. Vol. 2018. Article ID 9327403. Doi: https://doi.org/10.1155/2018/9327403

Hartigan, J., & Wong, M. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1), 100-108. doi:10.2307/2346830). https://doi.org/10.1016/j.petrol.2020.108093.

Land Information New Zealand (LINZ).(2015).Geotechnical information on horizontal land movement due to the Canterbury Earthquake Sequence. Job No: 53708.0000

Laouami, N. (2019). Vertical ground motion prediction equations and vertical-to-horizontal (V/H) ratios of PGA and PSA for Algeria and surrounding region. Bulletin of Earthquake Engineering. 17. Doi: 10.1007/s10518-019-00635-y.

Lee, E., Novotny, J., Wagreich, M. (2019). Subsidence Analysis and Visualization for Sedimentary Basin Analysis and Modelling. Doi: 10.1007/978-3-319-76424-5.

Mancin N, Di Giulio A, Cobianchi M (2009) Tectonic vs. climate forcing in the Cenozoic sedimentary evolution of a foreland basin (Eastern Southalpine system, Italy). Basin

Massoli D., Koyi H., Barchi M. (2006) Structural evolution of a fold and thrust belt generated by multiple de ́collements: analogue models and natural examples from the Northern Apennines (Italy). J Struct Geol 28:185–199. doi: 10.1016/j.jsg.2005.11.002

Merxhani, A. (2016). An introduction to linear poroelasticity. Cornell University. Doi: https://arxiv.org/abs/1607.04274.

Olneva, T., Kuzmin, D., Rasskazova, S., & Timirgalin, A. (2018). Big data approach for geological study of the big region West Siberia. Paper SPE-191726-MS, presented at the SPE Annual Technical Conference and Exhibition, Dallas, Texas, 24-26 Semptember 2018. DOI: https://doi.org/10.2118/191726-MS

Ori G., Friend P. (1984). Sedimentary basins formed and carried piggyback on active thrust sheets. Geology 12(8):475–478. doi:10.1130/0091-7613(1984)12\475:SBFACP

Pieri M., Groppi G. (1981). Subsurface geological structure of the Po Plain, Italy. CNR Prog Fin Geodin 414:1–13. Pointe South Mountain Resort, Phoenix, Arizona, 2004. Res 21:799–823. doi:10.1111/j.1365-2117.2009.00402.x

Rucker M., Greenslade M., Weeks R., Fergason K., Panda B. (2008). Geophysical and Remote Sensing Characterization to Mitigate McMicken Dam, GeoCongress 2008: Geosustainability and Geohazard Mitigation (GSP 178).

Sclater, J., Christie, P. (1980), Continental stretching: An explanation of the Post-Mid-Cretaceous subsidence of the central North Sea Basin, J. Geophys. Res., 85(B7), 3711– 3739, doi:10.1029/JB085iB07p03711.

Tan P., Steinbach M., Karpatne A., Kumar V. (2018). Introduction to Data Mining, 2nd edition. Chapter 7: Cluster Analysis: Basic Concepts and Algorithms. Pearson.

Teatini P., Castelletto N., Ferronato M., Gambolati G., Janna C., Cairo E., Marzorati D., Colombo D., Ferretti A., Bagliani A., Bottazzi F. (2011). Geomechanical response to seasonal gas storage in depleted reservoirs: a case study in the Po River basin, Italy. J Geophys Res. doi:10.1029/2010JF001793

Teatini, P., Ferronato, M., Gambolati, G., and Gonella, M. (2006), Groundwater pumping and land subsidence in the Emilia-Romagna coastland, Italy: Modeling the past occurrence and the future trend, Water Resour. Res., 42, W01406, doi:10.1029/2005WR004242.

Teatini, P., Ferronato, M., Gambolati, G. et al. (2005). A century of land subsidence in Ravenna, Italy. Environmental Geology. 47. 831-846. doi: 10.1007/s00254-004-1215-9.

Tergazhi, K. (1922). Der grundbruch an stauwerken and seine verhiltung. Die Wasserkraft 17. 687445–449.

Terzaghi, K. (1936). The shearing resistance of saturated soils and the angle between the planes of shear. International Conference on Soil Mechanics and Foundation Engineering. Harvard University Press: Cambridge, MA, 54–56

Tomás, R., Romero, R., Mulas, J. et al. Radar interferometry techniques for the study of ground subsidence phenomena: a review of practical issues through cases in Spain. Environ Earth Sci 71, 163–181 (2014). https://doi.org/10.1007/s12665-013-2422-z

Toscani G., Burrato D, Di Bucci D, Seno S, Valensise G (2009). Plio-Quaternary tectonic evolution of the Northern Apennines thrust fronts (Bologna-Ferrara section, Italy): seismotectonic implications. Ital J Geosci 128(2):605–613. doi:10.3301/IJG.2009.128.2.605

Weeks R., Panda B. (2004). Defining Subsidence-Induced Earth Fissure Risk at McMicken Dam, for Association of State Dam Safety Officials Annual Conference, 25–30 September 2004,

Whittaker B., Reddish J. (1989). "Subsidence arising from ground-water withdrawal. oil and gas held activities and underground coal gasification". Subsidence Occurrence, Prediction and Control - Developments in Geotechnical Engineering, vol. 56, chapter 15

Yiping W., Buqing S., Jianjun W., Qing W., Haowu L., Zhanxiang L., Ningning Z, Qingchao C. (2021). An improved multi-view collaborative fuzzy C-means clustering algorithm and its application in overseas oil and gas exploration. Journal of Petroleum Science and Engineering, vol.

10.1 Electronic References

Arpae Emilia-Romagna. At: *https://www.arpae.it/*. (Accessed August 24, 2020).

Bollettino ufficiale degli idrocarburi e delle georisorse. Ufficio Nazionale Minerario per gli Idrocarburi e le Georisorse. At: *https://unmig.mise.gov.it/index.php/it/informazioni/buig.* (Accessed June 20, 2020).

CRAN Project. R documentation. At: *https://www.r-project.org/other-docs.html* (Accessed April 7, 2021).

Extracting Seasonality and Trend from Data: Decomposition Using R. (2015). At: *https://anomaly.io/seasonal-trend-decomposition-in-r/index.html* (Accessed September 2, 2021).

Hyndman et al,.Plot time series decomposition components using ggplot. At: *https://pkg.robjhyndman.com/forecast/reference/autoplot.seas.html.* (Accessed August 8, 2021).

Hayes, A. (2021). What Is a Time Series? At*: https://www.investopedia.com/terms/t/timeseries.asp*. (Accessed October 3, 2021).

Hyndman, R.J., & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. At: *OTexts.com/fpp3*. (Accessed may 20, 2021).

Microsoft documentation. (2019). Normalize Data. At: *https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/normalize-data.* (Accessed October 3, 2021).

National Institute of Standards and Technology.Introduction to Time Series Analysis: Definitions, Applications and Techniques. At: *https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc41.htm*. (Accessed October 5, 2021).

Permessi di ricerca e concessioni di coltivazione. Ufficio Nazionale Minerario per gli Idrocarburi e le Georisorse. At: *https://www.arcgis.com/home/webmap/viewer.html?webmap=550af1c650d8467c b25984d50e668197&extent=9.1024,42.3886,17.2158,45.9221*. (Accessed June 10, 2020).

Prabhakaran S. R tutorial available online. At: *http://r-statistics.co/Loess-Regression-With-R.htm*l (Accessed April 7, 2021).

Ricci, Vito. (2005). Analisi Delle Serie Storiche con R. At: *https://cran.r-project.org/doc/contrib/Ricci-ts-italian.pdf* (Accessed April 5, 2021).

University of the West of England, Faculty of Environment and Technology. Stress in the ground. At: *http://environment.uwe.ac.uk/geocal/soilmech/stresses/stresses.htm.* (Accessed October 5, 2021).

Videpi Project. Ufficio Nazionale Minerario per gli Idrocarburi e le Georisorse. At: https://www.videpi.com/videpi/videpi.asp. (Accessed June 20, 2020).