UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI INGEGNERIA CIVILE, EDILE E AMBIENTALE
Department of Civil, Environmental and Architectural Engineering

Corso di Laurea Magistrale in Environmental Engineering

TESI DI LAUREA

SPATIO-TEMPORAL VARIABILITY OF WATER QUALITY IN THE VENETIAN PLAIN BETWEEN ASTICO AND AGNO RIVERS

Relatore:
Chiar.mo PROF. MATTEO CAMPORESE

Laureando: FRANCESCO MARCON
Matricola: 1242596

ANNO ACCADEMICO 2021-2022

INDEX

1 ABSTRACT 1

2 INTRODUCTION 3

3 AREA 5

3.1 The undifferentiated aquifer 10

3.2 The multi-layer aquifer 11

3.3 Agno valley 13

4 DATASET 15

5 METHODS 20

5.1 Basic concepts of geostatistics 21

5.2 The semivariogram 22

5.3 Ordinary kriging 30

5.4 Lognormal kriging 33

5.5 Semivariograms and kriging with anisotropy 35

5.6 Leave-one-out cross validation (LOOCV) 38

5.7 Preliminary analysis and basic statistical indicators 39

6 ANALYZED POLLUTANTS 42

6.1 Total solvents 45

6.2 Nitrates 47

6.3 Pesticides 49

6.3.1 Triazines: atrazine (ATR), terbutilazine (TBR), simazine and their metabolites 52

6.3.2 Metolachlor 54

7 RESULTS AND DISCUSSION 55

7.1 Case study: TOTAL SOLVENTS 56

7.1.1 Geostatistical analysis 63

7.1.2 Discussion 77

7.2 Case study: NITRATE 81
7.2.1 Geostatistical analysis 86
7.2.2 Discussion 98
7.3 Case study: PESTICIDES 101
7.3.1 DACT 104
7.3.2 Geostatistical analysis 106
7.3.3 Discussion 114
8 CONCLUSIONS 116
9 Bibliography 118

INDEX OF FIGURES

Fig. 1: domain of interest 5
Fig. 2: isophreatic lines, paleo riverbeds and main groundwater flows direction in the domain 7
Fig. 3: domain 3D 8
Fig. 4: Section AB of the domain (conceptual representation) 8
Fig. 5: land use map 9
Fig. 6: stratigraphy of the domain (SOURCE: Altissimo et. Al 1990, "carico inquinante degli acquiferi dell'alto vicentino") 10
Fig. 7: springs and clayey lens 11
Fig. 8: logarithm of hydraulic conductivity (SOURCE: Rinaldo et al. 2007) 12
Fig. 9: Agno valley domain 13
Fig. 10: Agno valley iso phreatic lines - 5m (source: Antonelli, Mari "Carta della vulnerabilità naturale" – 1983) 14
Fig. 11: section of Agno valley (source: Arzignano municipality) 14
Fig. 12: sampled wells 17
Fig. 13: sampled wells with only one measure (CIN) 17
Fig. 14: wells with known stratigraphy - 46 points (sources: Sinergeo, Viacqua, Acegasaps) 18
Fig. 15: points available in Agno valley 19
Fig. 16: example of semivariogram cloud 23
Fig. 17: semivariogram cloud and lag averages 23
Fig. 18: searching neighborhood for semivariogram modelling 24
Fig. 19: example of semivariogram (source: Yasojima et al, 2019) 24
Fig. 20: most frequently utilized model variograms 26
Fig. 21: example of linear model with no sill 28
Fig. 22: example of log transformation of dataset 33
Fig. 23: isotropic (left) VS anisotropic model (right) 35
Fig. 24: azimuth tollerance and bandwith 36
Fig. 25: geometric (left) and zonal (right) anisotropy 36
Fig. 26: types of skewed distributions 39
Fig. 27: boxplot meaning 40
Fig. 28: fate of pesticides 50
Fig: 29: atrazine main metabolites 53

Fig. 30: metabolites of triazines and the pathway to DACT 53
Fig. 31: available points and analysis per year (solvents) 56
Fig. 32: histogram and boxplot of total solvents (complete dataset) 57
Fig. 33: overall mean of total solvents (domain and Agno valley) 57
Fig. 34: total solvents in Torrebelvicino – Schio area 58
Fig. 35: scatterplot of total solvents 59
Fig. 36: Pozzo Ancignano total solvents time series 60
Fig. 37: percentage of above law limit values (10 microg/L) per year (total solvents) 60
Fig. 38: annual solvents mean for different aquifers 61
Fig. 39: histograms of raw data Vs logtransformed data (SOLVENTS 1991) 64
Fig. 40: boxplots (median black dot, mean void dot) of annual mean of solvents 1988-2021. Raw data (left) vs lograsformed data (right) 65
Fig. 41: example of scatterplots of solvents (microg/L) Vs coords (1993 66
Fig. 42: omnidirectional variogram of the first analyzed year (solvents 88) 66
Fig. 43: semivariograms (solvents 1993) in different directions: from 0° (top left) to 135° (top right) 67
Fig. 44: directional semivariograms (2003, 2012) 68
Fig. 45: isotropic omnidirectional variograms fitted by spherical model (logtransformed Solvents 1995 on the left, logtransformed solvents 2015 on the right) 69
Fig. 46: LOOCV of isotropic kriging (solvents 1991 in the left, solvents 2018 right) 71
Fig. 47:Upper plain solvents concentrations 1988-1995: lognormal kriging with anisotropy (left) and isotropic (right) 72
Fig. 48: upper plain solvents concentration 1999: lognormal kriging with anisotropy (left) and isotropic (right) 73
Fig. 49: upper plain kriging prediction standard deviations 1991-1999 73
Fig. 50: solvents concentrations 2003-2009: lognormal kriging with anisotropy (left) and isotropic (right) 74
Fig. 51: solvents concentration 2012 – 2018: lognormal kriging with anisotropy (left) and isotropic (right) 75
Fig. 52: solvents concentration 2021: lognormal kriging with anisotropy (left) and isotropic (right) 76
Fig. 53: standard deviations of kriging predictions solvents 2009-2018 (anisotropy left side, isotropic right side) 76

Fig. 54: solvents concentrations in the 80's in Vicenza 78
Fig. 55: number of available points and measures per year (nitrate) 81
Fig. 56: overall mean of Nitrates 81
Fig. 57: scatterplot of Nitrate 82
Fig. 58: municipalities with high nitrogen per unit area (source: Rinaldo et al 2007) 83
Fig. 59: nitrate concentration in some contaminated wells in the upper plain (Zugliano, Thiene, Sarcedo, Santorso) 84
Fig. 60: nitrate concentration in some wells in Malo, Villaverla 84
Fig. 61: nitrate annual means for different aquifers 85
Fig. 62: histogram of nitrate 1995; difference with a log transformation 87
Fig. 63: histogram of nitrate 2016-18 87
Fig. 64: boxplot of nitrates 1988-1999 in the upper plain (original values on the right and log transformed on the left) 88
Fig. 65: boxplot of nitrates 2001-2021 88
Fig. 66: scatterplots of nitrates and coordinates 88
Fig. 67: examples of omnidirectional semivariograms nitrate (1995 logtransf, 2013-15 raw data) .. 89
Fig. 68: directional variograms of log transformed data (nitrates 1988-1999) 90
Fig. 69: directional variograms (nitrates) 90
Fig. 70: examples of semivariogram fit, sample variance in dot line (from top left to bottom right: nitrates 1991 spherical fit (log), nitrates 1999 exponential (log), nitrates 08-09 gauss, nitrates 19-21 spherical) 91
Fig. 71: leave-one-out cross validation: log transformed Nitrate 1988 (left), Nitrate 16-18 (right) .. 94
Fig. 72: nitrate concentration and standard deviation of lognormal kriging prediction 1988 – 199595
Figura 73: nitrate concentration and standard deviation of kriging prediction 1999 to 2004-06 96
Fig. 74: nitrate concentration and standard deviation of kriging prediction 2007-09 to 2019-21 97
Fig. 75: example of nitrate concentration map with overlayed springs and rivers 99
Fig. 76: scatterplot of DACT 105
Fig. 77: histogram and boxplot of DACT 105
Fig. 78: DACT time series of some wells of Vicenza 106
Fig: 79: DACT (ng/L) vs coords 107
Fig. 80: Omnidirectional variogram DACT 107
Fig. 81: directional semivariogram DACT 108
Fig. 82: Isotropic variogram fitting with Gaussian (left) and linear (right) model – DACT 2020 .. 108

Fig. 83: exponential model LOOCV – DACT 2020 109
Fig. 84: ordinary kriging (Gauss, linear, exponential) DACT 19 111
Fig: 85: ordinary Kriging (Gauss, linear, exponential) DACT 20 112
Fig. 86: ordinary kriging (Gauss, linear, exponential) DACT 21 113
Fig. 87: LOOCV: differences between Gaussian (left) and linear (right) models – DACT 21 114

Index of tables

Tab. 1: Overall dataset properties 16
Tab. 2: ranges of solubility (source: Ronald Ney, “Fate and transport of Organic chemicals in the environment”, 1995) 43
Tab. 3: Koc mobility class ranges. Source: FAO (recommended for use by the US EPA) 43
Tab. 4: ranges of GUS. SOURCE: Gustafson, 1989. “GUS: a simple method for assessing pesticide leachability” 44
Tab. 5: solvents main properties for assessing underground mobility(sources: PubChem database, US Agency for Toxic Substances and Disease Registry database, Roberts et al. 1982) 46
Tab. 6: general informations on the considered pesticides (TR = triazinic group, ME = metabolite of a parent pesticide, CA = cloroacetanilide, AM = acetamide) 51
Tab. 7: pesticides main properties for assessing underground mobility (sources: PubChem database, USDA – ARS pesticides properties database) 52
Tab. 8: main properties of the analyzed periods (total solvents) 63
Tab. 9: ordinary kriging (anisotropic) parameters and LOOCV RMSE (total solvents) 69
Tab 10: ordinary kriging (isotropic) parameters and LOOCV RMSE (total solvents) 70
Tab. 11: main statistical indicators for the complete dataset of nitrate 82
Tab. 12: main properties of the analyzed periods (Nitrates) 86
Tab. 13: ordinary kriging (anisotropic) parameters and LOOCV RMSE (nitrate) 92
Tab. 14: ordinary kriging (isotropic) parameters and LOOCV RMSE (nitrate) 92
Tab. 15: analysis of pesticides dataset (from 1981 – 2021) 102
Tab. 16: analysis of pesticides database (from 2000 – 2012) 102
Tab. 17: analysis of pesticides database (from 2012 – 2021) 102
Tab. 18: number of available points and analysis - DACT 104
Tab. 19: main statistical indicators - DACT 104
Tab. 20: distance matrix of DACT (2019-2021) 106
Tab. 21: DACT semivariogram fitting (Linear model) and LOOCV results 109
Tab 22: DACT semivariogram fitting (Gaus, exponential models) and LOOCV results 109

1 ABSTRACT

This document is the product of an internship at *centro RIVE (Risorse Idriche Venete)*, a research center on water resources made up by Viacqua and Etra. The present vision of the team is the study of the hydrological structures in the venetian plain (mainly in the Vicenza province) and its relationship with the external factors on the area, the monitoring of water quality and the support for the development of water safety plans.

The aim of this thesis is to recognize some underground water pollution episodes in the venetian plane between Astico and Agno river using some geostatistical tools; this area has been exploited for drinkable water by a large number of wells. As observed in many other plains characterized by a widespread presence of industries, the water quality of aquifers is often a target for various chemicals and compounds that may pose danger for human health. For this reason, a clear understanding of the past contamination events, as well as their impact on aquifer water quality, is very important for a management scenario; it is also a tool for evaluating the possible fate of future pollution cases. In this context, geostatistical analysis of existing data may give useful information. Unfortunately, various limitations can alter the quality of the results obtained; this uncertainty, coming from insufficient data, uneven distribution of the observed points and overall measurement variance is always present when analyzing "real world" data and the goal is only to reduce it.

A great number of parameters regarding water quality is present; obviously this study cannot comprehend the analysis on the 20+ compounds present in typical laboratory test results. For this reason, thanks to past publications and the historical memory of people involved in water quality analysis, some compounds like solvents, nitrates and various pesticides are known to be concerning and therefore are selected to be studied. Furthermore, nitrates and solvents are known to be the main historical contaminations in the area characterized by industries and intensive agriculture; pesticides are different because their pattern of contamination is linked also to their metabolites. This aspect is also linked to the framework of the emerging contaminants (or contaminants of emerging concern); some compounds, if not all of them, are naturally evolving and sometimes the by-products can be as dangerous as the original compound.

The methodology used for evaluation of the contamination episodes follows a logical pattern which is briefly described: the first step of the process to reach the goal is to gather and organize data. This thesis comprehends water quality analysis from 1981 to 2021 coming from various entities: Viacqua, AcegasAPS, ATO (Ambito Territoriale Ottimale) Bacchiglione, CIN (Centro Idrico di Novoledo), SPV (Strada Pedemontana Veneta), ULSS (Unità Locale Socio-Sanitaria). Then, some descriptive

statistics and plots of time series are presented to recognize the overall behavior of the analyzed contaminant and the time evolution of it. After that, geostatistical tools like variograms and kriging are implemented: the overall understanding of the pollution episode can be achieved only by observing all the steps.

2 INTRODUCTION

The goal of this thesis is to study the evolution of some aquifer contaminations of the past decades. The considered area is located north of Vicenza, mainly between Astico and Agno rivers, where an important aquifer system has been exploited for drinking water, serving more than 600'000 people (*Sottani et al. 1982*). Due to the fundamental strategic role of water supply in this area, the aquifer's water quality has been studied by many authors in different periods: the aim of this document is to reconstruct the historical contaminations using geostatistical tools (like variograms, kriging etc.) and to expand the study on the last years in order to understand the evolution of past contaminations and estimate future scenarios. A great number of compounds have been systematically measured over the last decades and the examination of all of them is not a goal of this thesis, for this reason only some pollutants of great interest will be analyzed: solvents, nitrates, pesticides.

Due to the great complexity typical of an aquifer contamination, the understanding of the phenomena comes from various steps: for this reason, is important to remember that geostatistical tools, like kriging, are the final step of an analysis and not the whole study. Furthermore, like all methods, geostatistical analysis is based on simplifications and assumptions of the real world.

The term *kriging*, often used as a synonym for geostatistics, comes from the name of a south African engineer, Danie Krige. In the 50's Krige was interested in precious mineral mining: his goal was to estimate the most likely distribution of gold based on samples from a few boreholes. This aspect is the fundamental idea of kriging method: estimate a prediction in an unsampled point using data from observed points. The mathematical rigorous development of this method came in the 60's from Matheron, a French mathematician. After that a rich literature regarding kriging and its applications on natural sciences (like geology, underground hydrology, geochemistry, biology etc.) has been developed. Is also interesting to note that literature on the topic is not completely aligned on all aspects of the method. A more complete description of the method is present in the appropriate paragraph.

The structure of the thesis reflects the logical steps to understand the development of a contamination event. For this reason, the frame is composed mainly by the following components:

- A description of the area to better understand the structure and the properties of the aquifer system and its relationship with human activities.
- A brief overview of the pollutants, their use on the area and the legal framework of these compounds.
- A description of the mathematical structure of the geostatistical tools (kriging) and the main properties of the formulations.

- An overall description of the various contamination phenomena by using descriptive statistics and observing past papers on the subject.
- The application of geostatistical tools on the available data and the presentation of the results obtained, with critical observations to underline negative and positive aspects of the various methods.

Unfortunately, various limitations can alter the quality of the results obtained; this uncertainty, coming from insufficient data, uneven distribution of measurement points and overall measurement variance is always present when analyzing "real world" data and the goal is only to reduce it. Therefore, various simplifying assumptions are made in order to obtain some indications on the overall situation of water quality.

To obtain the results, data coming from different origins has been gathered; for this reason, a unique database containing all available measurements on water quality is the building block for the subsequent elaborations. This thesis in fact contains data coming from different stakeholders on water quality in the area like: Viacqua (Vicenza province water manager), AcegasAPS (Padova province water managers), ATO Bacchiglione, ULSS (national health service) and others. This aspect also underlines the importance of collaboration between the various entities interested in the conservation of environmental quality: a common vision on environmental protection should be the driving force for future improvements.

For the development of the various geostatistical methods, the visual representation of inputs and outputs and the overall analysis of the data various tools from different software have been used. The preliminary elaboration of the data as well as the construction of the unique database has been developed in Excel. For visual inspection of the georeferenced data and maps QGIS is the chosen software, and finally SAGA, R (gstat package) and ArcGIS PRO (geostatistical analyst package) have been used for the implementation of the kriging method and other geostatistical tools.

3 AREA

In this chapter a description of the hydrogeological properties of the area of interest is given. The domain of interest, located in an area between the Astico and Bacchiglione river in the Venetian plain, is one of the most exploited areas for water supply in Veneto: unfortunately, this area has also been interested by various episodes of groundwater contamination. The analyzed area, showed in figure 1, ranges from the venetian Prealps to the urban area of Vicenza and from the last reliefs of the Lessini mountains to the Astico river.

Within the area, the terrain is substantially flat, with decreasing slopes from NW to SE. The plain is interrupted by the hills of Montecchio Prec.no, Zugliano and Sarcedo.

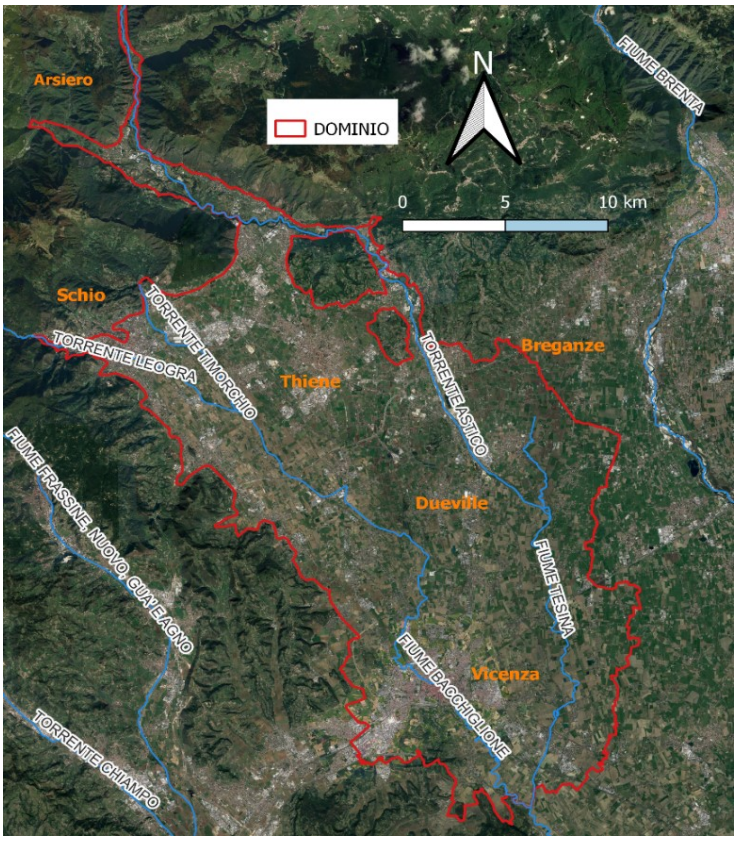


Fig. 1: domain of interest

The anthropogenic (industrial) influence is very strong throughout the survey area, which has over 460,000 residents. The Alto Vicentino is characterized by the existence of over 10,000 production and service activities mainly of small and medium size scattered throughout the territory, but with a greater concentration around the historic industrial centers of Schio, Thiene and Marano Vicentino. Manufacturing companies in the mechanical, textile-clothing, food and furniture sectors prevail.

In agriculture, the various areas are mainly used for forage and pasture, for corn crops and finally for wheat and barley crops. From west to east, from south to north there is a progressive reduction of the areas covered with maize, with a simultaneous increase in permanent meadows and pastures (*Altissimo et al., 1990*). The areas in which agriculture is heavier in terms of utilization of soil can be visually observed in figure 5.

The analyzed area is part of the venetian plain unit, which spread over a wide strip of territory located at the foot of the pre-alpine reliefs, characterized from the hydrographic point of view by the presence of a series of watercourses with a subparallel course which come out of the mountain valleys, cross the plain territory, and finally pour into the sea. These waterways (e.g., the rivers Bacchiglione, Astico etc) are responsible for the deposition of massive quantities of loose materials of fluvioglacial origin which, accumulated in strong thicknesses, gave rise to the subsoil of the high plain, also contributing to the existence of different hydrogeological structures present in the medium and lowlands plains. The area is characterized by 3 main underground structures:

- High plain: These are extensive fan-like structures deposited by rivers at different times when their regime was different from the current one (much higher flow rates) due to the melting of the glaciers. This structure, with undifferentiated gravelly mattress, varies from 5 to over 20 km starting from the foot of Prealps reliefs. The height of the mattress, definable as undifferentiated aquifer, can reach a few hundred meters depth: this stratum, composed mainly by coarse material, is interested by high velocity underground flows (high hydraulic conductivity). The overall thickness of the gravel decreases progressively. This zone in the area ranges roughly from Arsiero to Villaverla – Caldogno – Dueville.

- Middle plain: it is placed immediately downstream of the previous strip, for a width of 5 to 10km, and is characterized by the presence of perennial springs, which are the "spillways" of the undifferentiated aquifer which has its final part in this zone. In the medium plain gravel levels decrease by number, thickness, and grain size; the stratigraphic structure starts to be composed by alternations of gravel mattresses and silty-clayey lens. This zone in the area corresponds roughly to the Caldogno – Dueville municipality area.

- Lower plain: extend from the middle plain to the Venetian lagoon and its subsoil is composed by alternating levels of aquifers (sandy soils) and aquitards (silty – clayey soils). This zone in the area corresponds roughly to the Vicenza municipality.

The main directions of groundwater flows can be observed in figure 2: the flux direction is perpendicular to the isophreatic line (isolines of 5 meters). Is clear that flows from the upper plain converge in the middle plain (springs strip) and then continue towards S-E direction; the hydraulic gradient and therefore the velocity of groundwater flows is much higher in the upper plain.

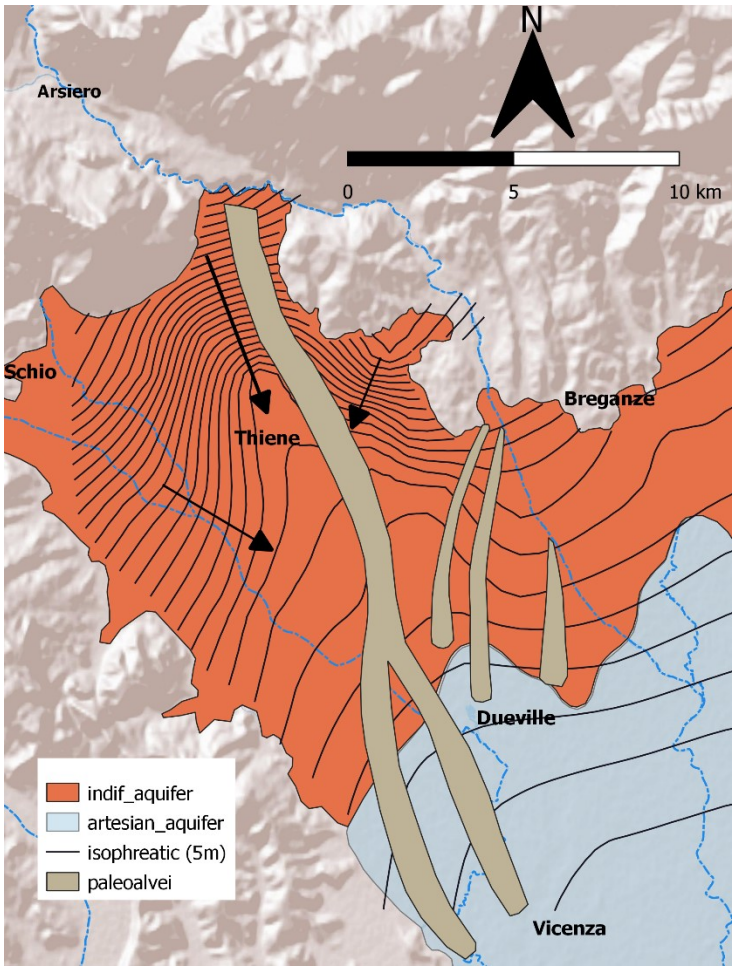


Fig. 2: isophreatic lines, paleo riverbeds and main groundwater flows direction in the domain

The description just presented is complicated by the presence of a series of faults (Schio-Vicenza faults and others) which give rise to a "stepped" structure also influencing the alluvial mattress of the plain itself. In fact, starting from the western limits and progressively moving towards the east, an initial increase in the aquifer is observed (with a maximum of power in the Thiene area), with a subsequent reduction with a minimum on the right bank of the Brenta (outside the domain of interest). The point of minimum depth represents a natural "watershed" that separates the underground flows of two sub-basins (Brenta and Astico). This limit, along the Breganze – Sandrigo axis, is indicated (fig 6) as a dynamic watershed as its location is not fixed but is affected by the different water supplies of the main water courses (*Favero, 1997*).

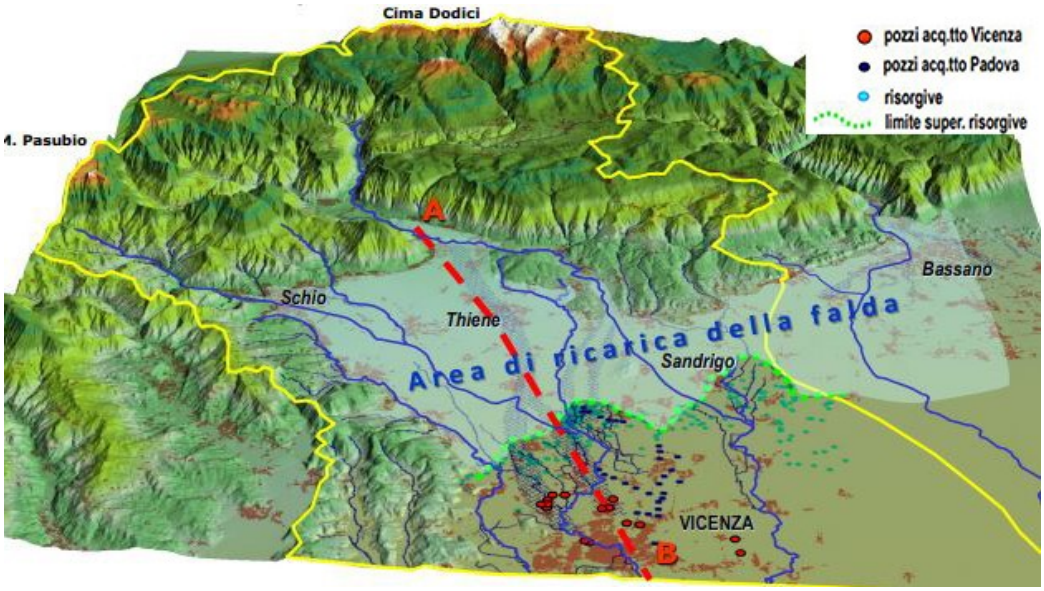


Fig. 3: domain 3D

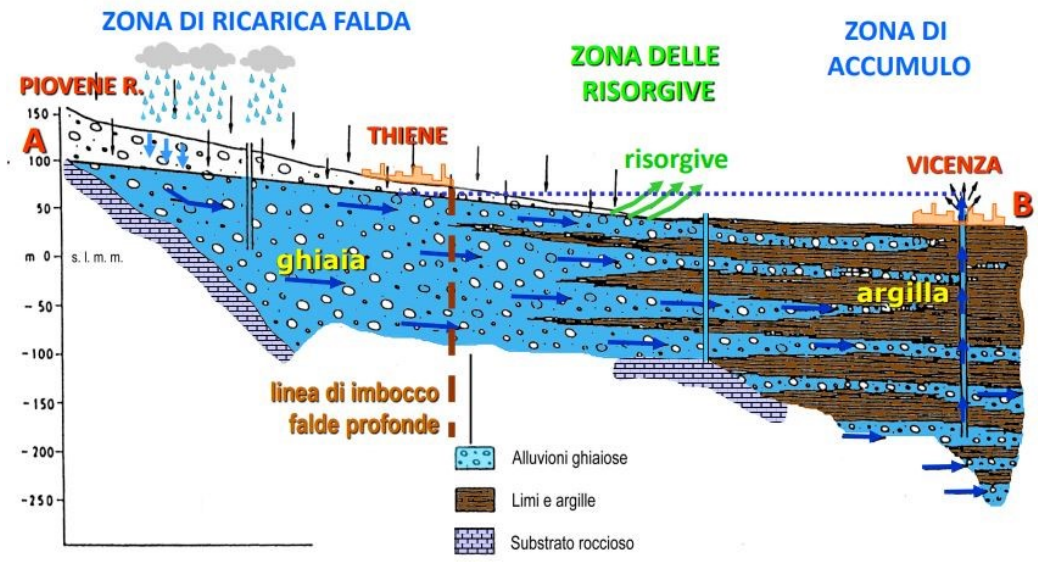


Fig. 4: Section AB of the domain (conceptual representation)

The hydrological balance of groundwater in the domain is difficult to obtain, due to the complexity of the hydrography: in the analyzed area basins of various rivers are present (Astico, Leogra), and for deep aquifers also the Brenta river may represent a recharging contribute. A balance for Astico sub-basin, which is the main component in the area, has been proposed (*Sottani et al. 1982*) referred to the period 1979-1981. The inflows are: effective precipitation (34,1 %), dispersions of natural watercourses (53,1%), dispersions from irrigation practices (12,1%). The dispersions from natural watercourses are the main recharge of groundwater; the contribute of effective precipitation may be

declining from effects related from climate change: the increasing rarity and intensity of precipitation events dictate a lower fraction of effective (infiltrating) rain.

The main outflow of the system is represented by the springs (74,2%) followed by the exploitations through wells (23%).

The flows feeding the aquifer come mainly from the surface (rainwater, irrigation and natural water systems) (*Sottani et al., 1982*) and for this reason the opportunity of percolation of contaminants on the topsoil is triggered. The relative limited speed of movement in the water table, the presence of clayey layers in the transition belt and the considerable flow from the springs suggests that the exchange of water is considerably faster in the superficial part of the aquifer than in depth (*Favero 1997*).

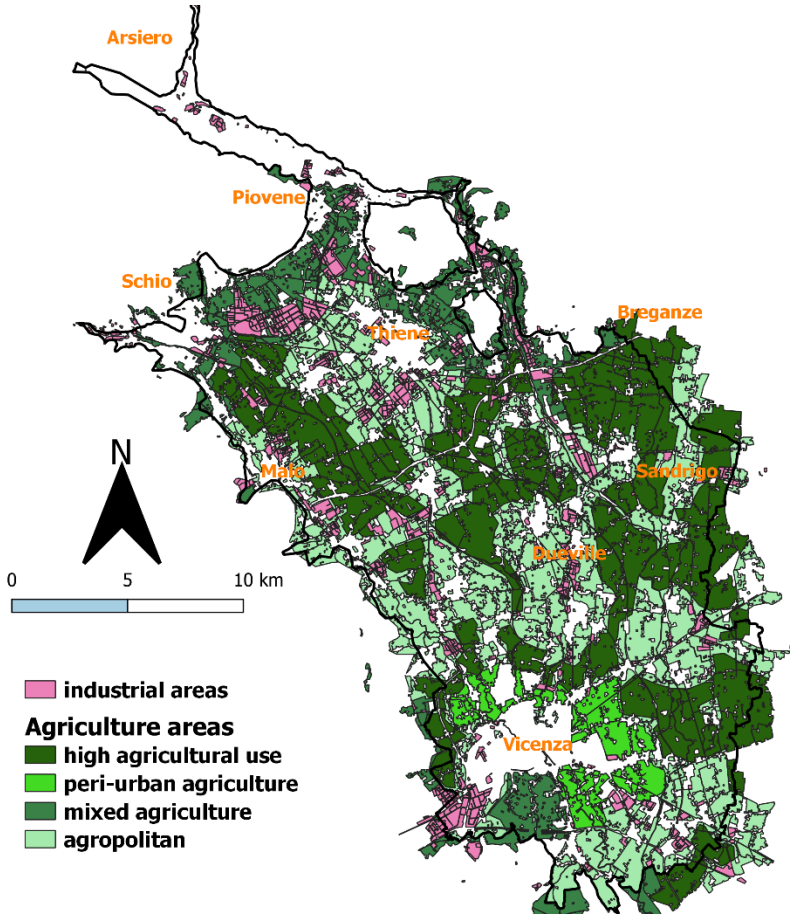


Fig. 5: land use map

3.1 The undifferentiated aquifer

As already mentioned, in the northern part of the domain there is a deep (200-300m) gravelly mattress which forms an undifferentiated aquifer: for this reason, in this area the character of the aquifer is phreatic. The subsoil of this zone has a great abundance of underground water; the geology in fact determines the presence in the subsoil of large volumes of gravel materials with high permeability, which form powerful underground reservoirs. The hydraulic conditions also provide effective recharging mechanisms, allowing the infiltration of significant water flows into the subsoil and the continuous supply of underground reservoirs. The same mechanism also determines the recharge of the southern part of the domain in which the multi aquifer system is present: the balance of the overall groundwater system is regulated by the springs in the middle plain, which naturally act as spillways. The real groundwater velocities are not well defined and show great variability in the upper plain; nevertheless, a range 1 m/day to 10 m/day has been identified.

In this area there is also the presence of paleo-river beds, which determines a preferential direction for underground flows.

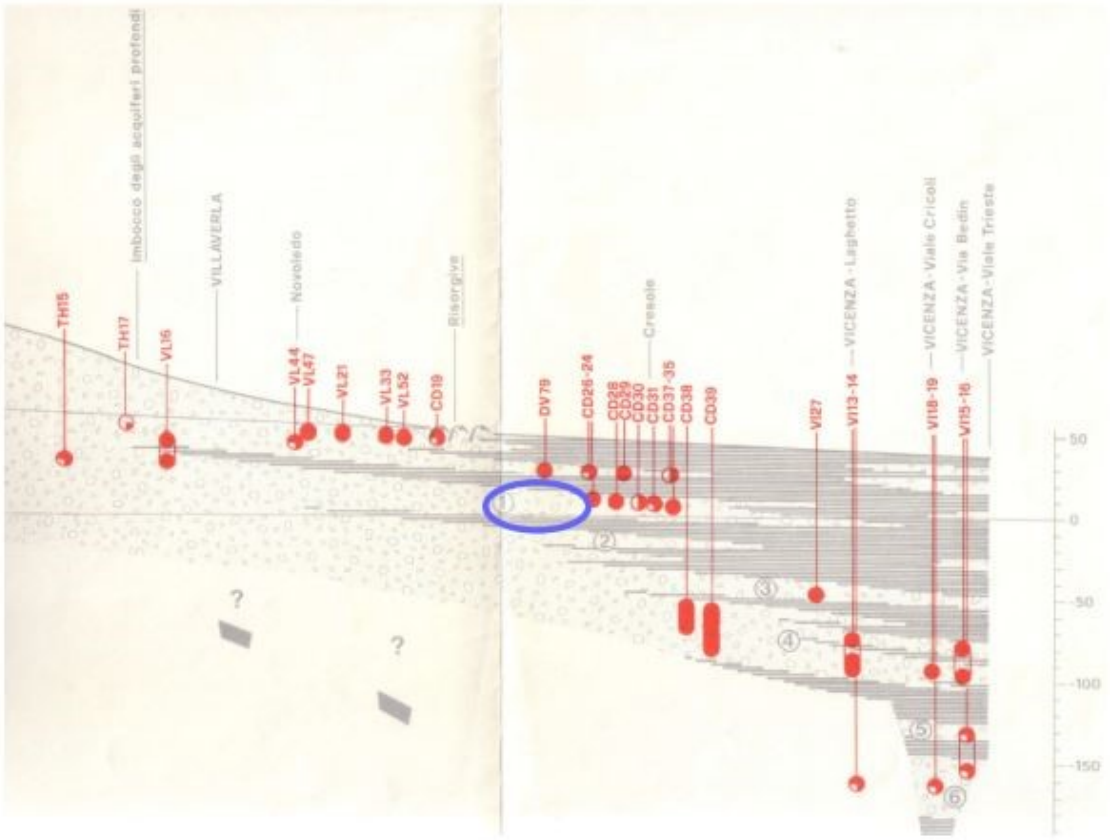


Fig. 6: stratigraphy of the domain (SOURCE: Altissimo et. Al 1990, "carico inquinante degli acquiferi dell'alto vicentino")

3.2 The multi-layer aquifer

Downstream of the undifferentiated aquifer, lens of fine materials, characterized by low hydraulic permeability are found in the subsoil. These layers (aquitards) show increasing thickness in the direction N-S and E-W and determine the hydraulic separation of various pressurized aquifers. The power of this aquifers can reach depths of about 10 meters and the majority of them are artesian, presenting pressures of about 5 meters above the terrain, and naturally giving flows of dozens of liters per second.

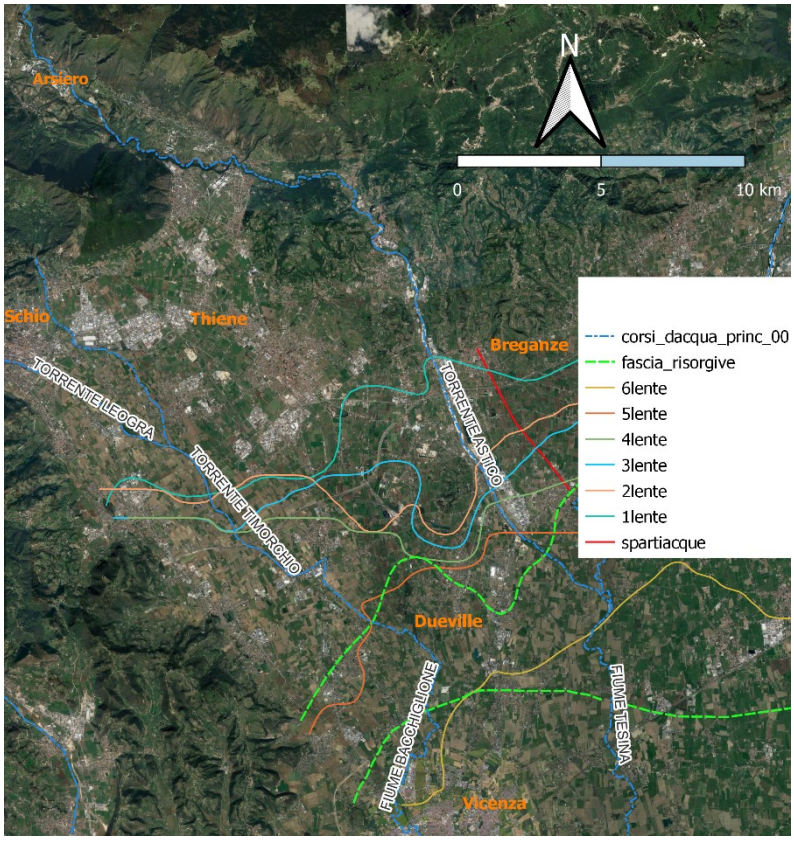


Fig. 7: springs and clayey lens

The water supply of the aquifer layers is given by the undifferentiated aquifer upstream to which they are connected; the presence of ancient large alluvial dejection (mainly linked to ancient positions of the Astico river) pose as effective and relevant axes of recharge of deep aquifers.

In this part of the domain, within the maximum depth of about 200 meters of the loose soil, seven principal structures have been recognized (courtesy of Sinergeo):

- Phreatic superficial aquifer
- 1° pressurized aquifer (mean depth 35 m)
- 2° pressurized aquifer (mean depth 45 m)

- 3° pressurized aquifer (mean depth 80 m)
- 4° pressurized aquifer (mean depth 130 m)
- 5° pressurized aquifer (mean depth 160 m)
- 6° pressurized aquifer (mean depth 200 m)

The principal aquifers, in terms of utilization, are the 3° and 4°, which have thickness of 20-25 meters. The first two layered aquifers (1° and 2°) seem to be differentiated by a semi-permeable layer, so there is not a clear hydraulic delimitation between the two. Authors have also suggested the possibility of the contribution of Brenta's deep underground flows in the 5° and 6° aquifers (*Sottani et. al, 1982*).

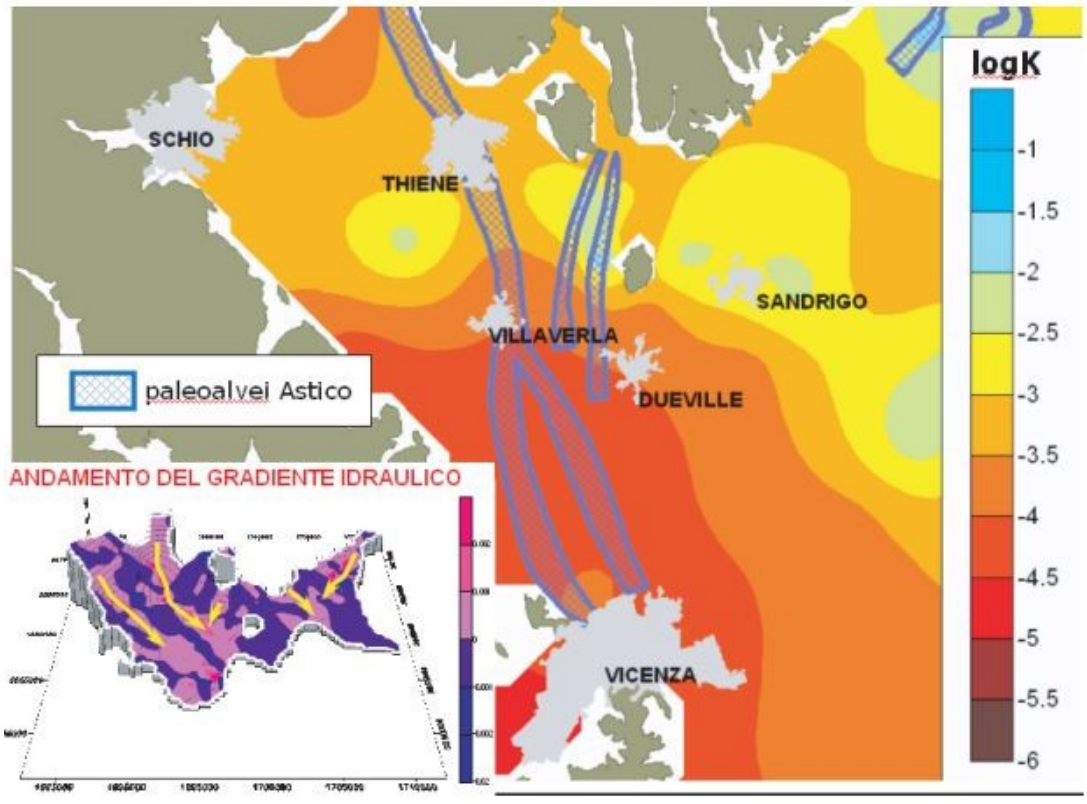


Fig. 8: logarithm of hydraulic conductivity (SOURCE: Rinaldo et al. 2007)

3.3 Agno valley

In this thesis also a minor area, separated from the previously described domain by the hills of the last reliefs of Lessini mountains, is analyzed. This valley, formed by the alluvial deposits of Agno river, is characterized by a subsoil of coarse materials, with occasional small lens of fine granulometry. The thickness of the alluvial mattress is variable and in the central part of the valley reaches 100 meters of depth; the aquifer is therefore undifferentiated and is exploited for drinking purposes and others. Obviously the phreatimetry follows the natural slope directed N-S from the upper valley to the end towards the plain; values of hydraulic conductivity, extrapolated from pumping tests in some wells in Arzignano, at the end of the valley, ranges from 10^{-4} m/s to 10^{-5} m/s with transmissivity of about 10^{-2} m^2/s.

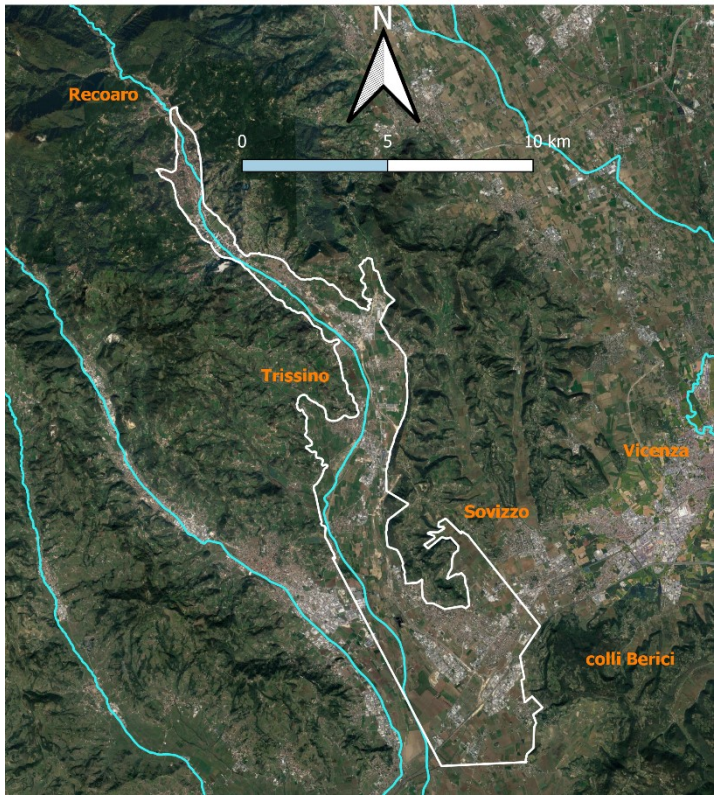


Fig. 9: Agno valley domain

Agno river has an erratic behavior, typical of torrents, linked to meteorology: important and destructive flood waves followed by dry periods are observed. The river is subject to periods of low flux, or even dry seasons during summer, as consequence of withdrawals (for irrigation and others) and the very permeable subsoil. In its higher parts, corresponding to Recoaro municipality and surrounded by the mountains of the "Piccole Dolomiti", the valley is very strict and with important

slopes; towards Valdagno municipality the valley starts to enlarge and the housing density, as well as the presence of human industrial activities, increase. Important industrial and artisanal centers are present in Trissino, Cornedo, Brogliano etc.

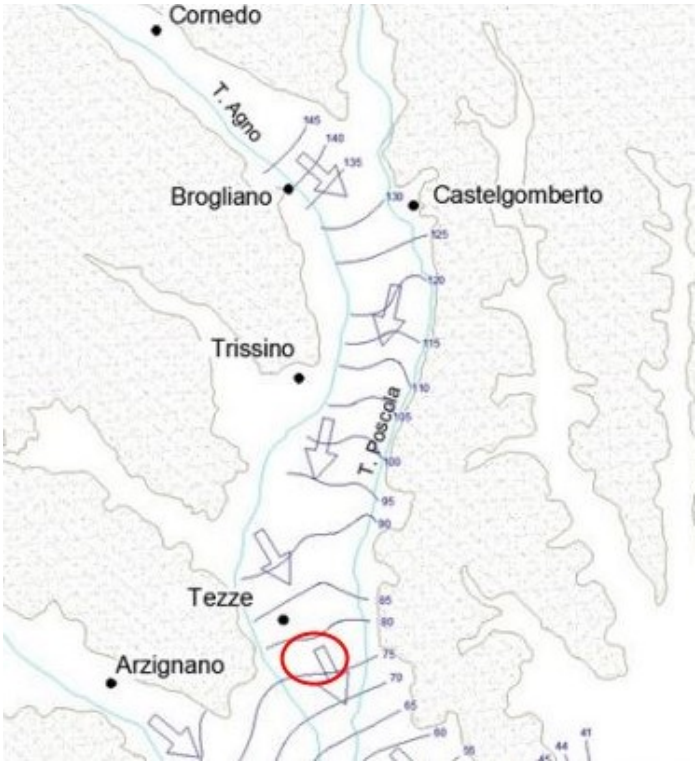


Fig. 10: Agno valley iso phreatic lines - 5m (source: Antonelli, Mari "Carta della vulnerabilità naturale" – 1983)

Unfortunately, this valley has recently experienced a serious case of PFAS contamination, which is being extensively studied and for which many legal cases are being decided. This contamination episode is not included in this study.

The valley can potentially communicate hydraulically with the previous area through a channel, between the Berici hills and the hills of Sovizzo; the direction of the flux in this part is function of the oscillations of phreatic levels (no data on this topic).

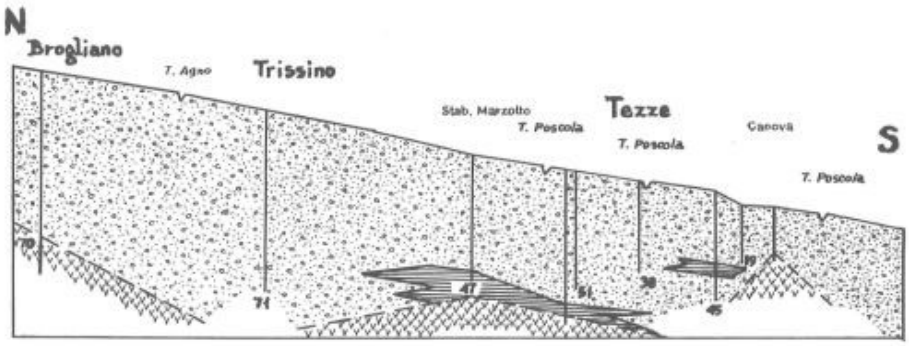


Fig. 11: section of Agno valley (source: Arzignano municipality)

4 DATASET

The analyzed dataset is composed by measurements of various chemicals in water; various institutions are involved in water quality analysis for different reasons. In this thesis, data coming from various organizations are gathered to compose a database of water quality parameters. The temporal span of the measures goes from 1981 to 2021 with no interruptions, but the number of available analysis and points (wells) has great variations through the years. For each well there are approximately two or three measures per year, although is not rare to observe only one measure: this fact has been considered for the choice of the temporal scale of the various geostatistical analysis.
The entities which provided measures of water quality are:

- VIACQUA: the water manager serving 68 municipalities in the Vicenza province. The main sources of water are in the upper plain from mountain springs and in artesian wells in the middle and lower plain.
- ACEGASAPS: the water manager in Padova province, which draws part of the water from some wells in the middle plain of Vicenza.
- CIN (centro idrico di Novoledo): a service company which provides measurements and studies of hydrogeological and biochemical parameters that characterize the Astico – Bacchiglione hydrological system (used for drinking water supply in Vicenza and Padova provinces).
- ATO Bacchiglione: is the entity to which the Veneto region has given the task of supervising the integrated water cycle for the territory of its competence (optimal territorial area) consisting in 136 municipalities between Padova and Vicenza provinces.
- ULSS: public health company operating in Veneto.
- SPV: a private company, concessionaire for the design, construction, and management of the Pedemontana Veneta highway.

Of the presented entities, the most important in terms of quantity of data is Viacqua, the registered office of the internship done at Centro RIVE which resulted in this thesis. Viacqua and ACEGASAPS, being water suppliers, are naturally interested in checking the quality of the water that they provide: for this reason, a great number of analyses are carried out throughout the whole year. It is important to note that this thesis is focused only on the aquifers and for this reason only the analysis of the “raw” water coming out of wells or springs, without any treatment like filtrations, chlorination etc. Water quality in the supplying net is affected by operations to improve some properties for the

optimal human consumption and therefore the group of analysis on other parts of the water supplying chain are not considered.

Another great contribution in terms of quantity of data is given by the CIN, especially in years prior to the 2000's; this entity provides water quality analysis in private (domestic) and industrial wells. CIN provides a great quantity of analyzed wells: most of them are private wells and therefore are sampled only one time in the entire period (fig. 13). For this reason, those measure can increase the spatial density of the measures, especially in the artesian zone, but have the downside of not having historicity (only one measure and therefore no indications on the temporal behavior of water quality). Data provided by ATO consists in the results of several monitoring campaigns to inspect the behavior of various parameters in the water cycle of the interested optimal territorial area. SPV measures have been done in the context of construction of the highway and for this reason follow the spatial path of the construction; samples are usually collected in superficial wells. Minor contributions are given by ULSS providing a limited dataset.

Tab. 1: Overall dataset properties

PROVIDER	FIRST MEASURE	LAST MEASURE	N° OF SAMPLED WELLS
Viacqua	1981	2021	115
AcegasAps	2003	2021	30
CIN	1985	2020	561
ATO Bacchiglione	2017	2020	82
ULSS	2018	2020	21
SPV	2015	2020	67

In years prior to the 2000's the dataset is composed only by the measures of CIN and Viacqua. Furthermore, is observed in the 90's a slightly preferential sampling towards the upper plain, where many industrial and agricultural environmental pressure points are present. This fact is probably due to a higher concern for contaminations, resulting in a higher density of analyzed wells in the upper plain: for this reason in this study, the geostatistical analysis for water quality before the 2000's is focused only in the upper plain.

The complete dataset, with measures coming from all the different entities can be observed from 2015 to 2020, years in which the maximum density of sampled points is reached. The last observed year,

which is 2021, is unfortunately made up only by the samples collected by the water suppliers (Viacqua, AcegasAps).

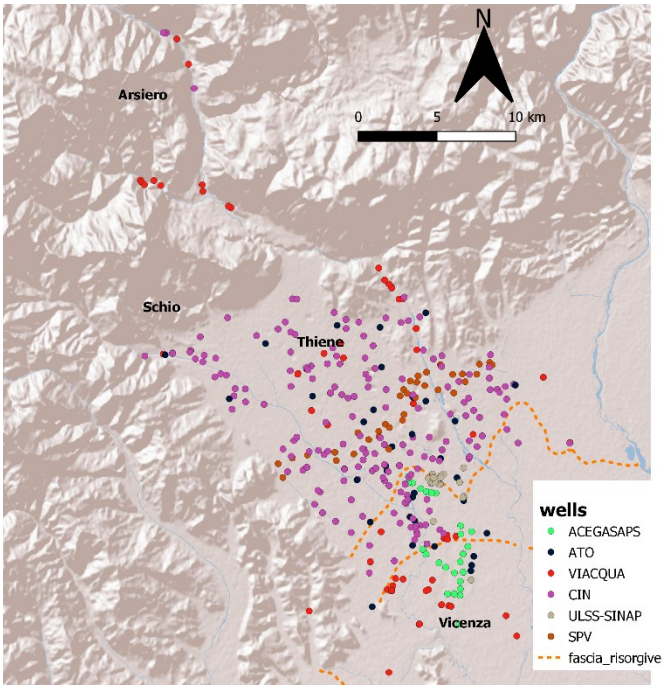


Fig. 12: sampled wells

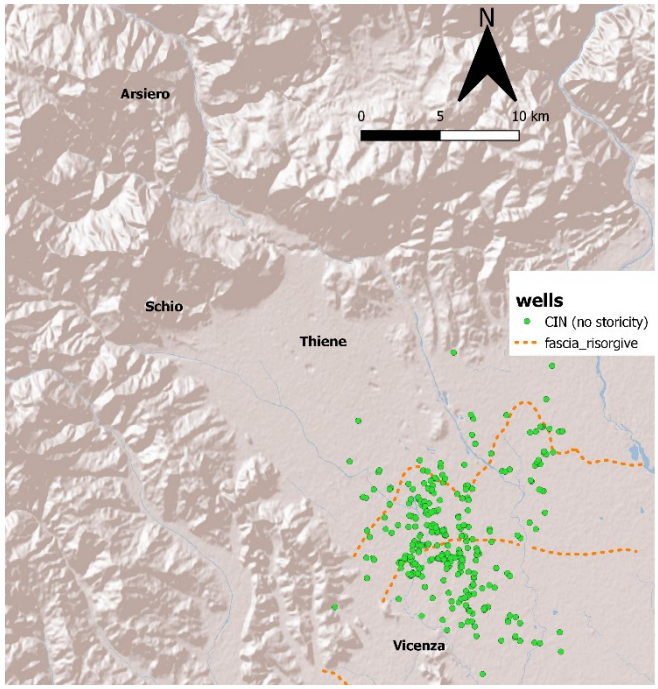


Fig. 13: sampled wells with only one measure (CIN)

Regarding the spatial distribution of the analyzed wells, is clear that there are zones in which the density is lower (like the S-E part of Vicenza) and areas in which is higher (like the middle plain). The situation presented in figure 12 and 13 is an overall view; spatial distributions of different periods are different one another and for this reason should be inspected when analyzing a given time frame. Another important information in the analysis of water quality in the aquifers is related to the vertical dimension: wells can reach different depths and therefore the analysis can be affected by this factor. Different depths, in a hydrogeological view, means different travel times, different hydraulic factors (conductivity, transmissivity etc.) and therefore different conditions for the pollutant to arise. This differentiation is especially important in the artesian zone, where each aquifer is separated from the other.

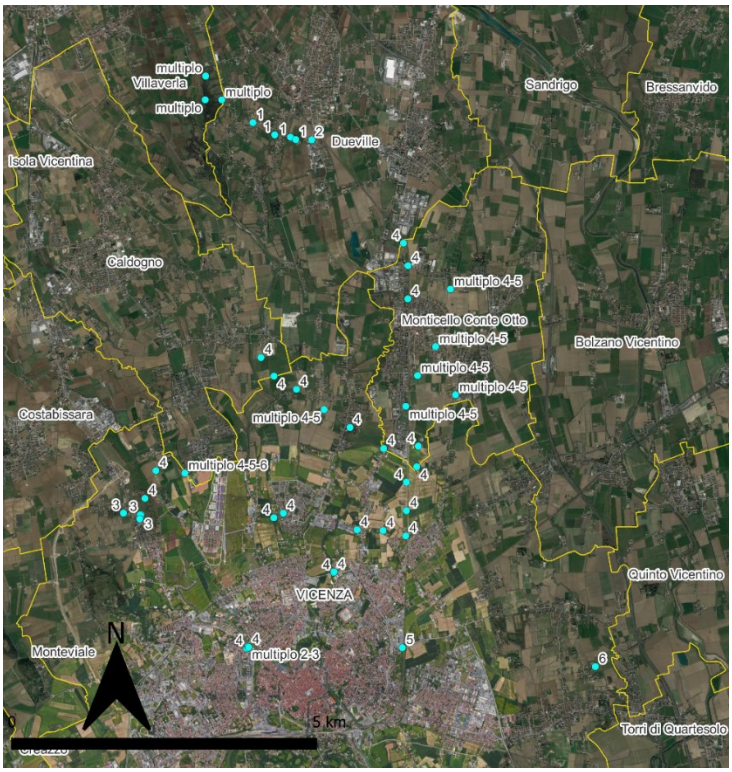


Fig. 14: wells with known stratigraphy - 46 points (sources: Sinergeo, Viacqua, Acegasaps)

As already mentioned in the dedicated paragraph the recharge zone for the artesian zone is very similar for all aquifers but the deeper ones (5° and 6°) can be affected by dispersion of the Brenta river, which can alter the water quality in comparison to the other aquifers which do not receive this dispersion contribution. Another problem related to the stratigraphy is the direct contamination which can occur in the shallower aquifers (1° and 2°) which seem to have only a semi-permeable layer as "cover": for this reason, the percolation from the superficial terrain, even if probably very slow to reach the first aquifers due to the low permeability, could be altering the water quality from the deeper aquifers. For these reasons, a separation of wells based on the depths they reach could be the optimal

strategy to obtain precise results: unfortunately, information on the stratigraphy of the wells is available only on a restricted number of points (46 wells, courtesy of Sinergeo, Viacqua and Acegas APS), where some video inspections have been done or some stratigraphy plots are present.

A lot of wells have been perforated many years ago and a lot of information have been lost. The wells can also intercept various aquifers because filters are found at multiple depths; most of the wells intercept the 3° and 4° aquifer which are the most productive ones. For this reason, as a first approximation, it can be stated that for a well with unknown depth the probability of interception of 3° or 4° aquifer is very high: this fact can be explained by the optimal balance between productivity (high transmissivity and storativity of the aquifer) and reasonable depths (not too shallow nor too deep). In later paragraphs some methods to remedy the problem of treating data with unknown stratigraphy are investigated.

Finally, some points are also present in the Agno valley; the first measurements available trace back in 1999, coming from:

- SPV: 31 points
- Viacqua: 17 points

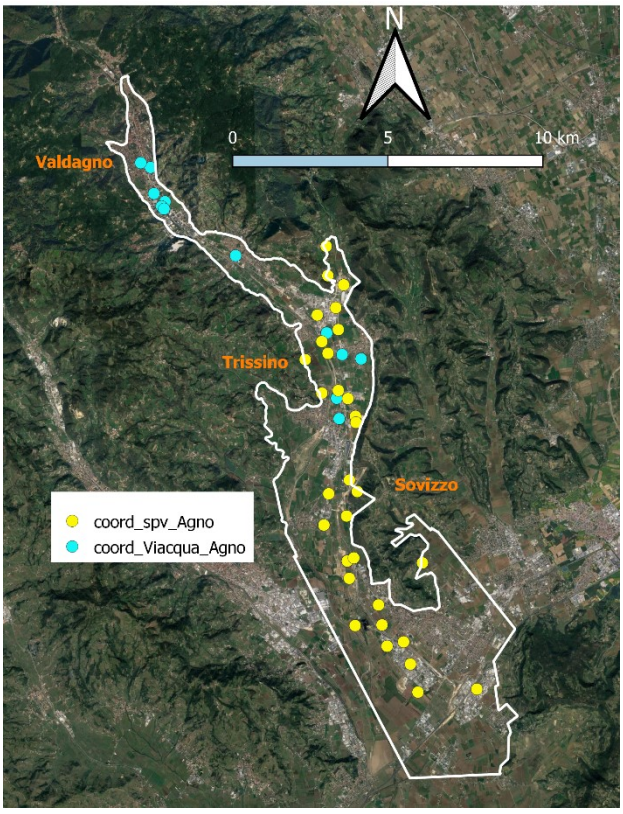


Fig. 15: points available in Agno valley

5 METHODS

In this chapter a description of the various tools used for the analysis of the contaminations is given. There are different ways to study and characterize a water quality pollution phenomenon, and in general ecological problems, where many systems (physical, biological, hydrogeological) interact. All the different approaches can be classified in two groups:

1) Deterministic models: they are based on mathematical formulations describing the physical behavior of the problem. The implementation of these models requires a detailed knowledge about the physical quantities of the underlying problem. Because of that, these kinds of models are often difficult to apply to real world data at large scales.
2) Probabilistic (statistical) models: they are not based on any kind of physical mechanism and are only considering the statistical properties of the problem. This kind of approach may seem to have underlying deficiencies, but in reality, the properties of a problem often present some kind of spatial-temporal correlation, and with the right mathematical tools this relationship between measured points can be exploited to obtain good models. Another positive aspect of this approach is that can give an estimation of the prevision error (this is not true for all statistical models).

Starting from the 50's a new branch of studies, called geostatistics, was developed for environmental type problems after the works of Daniel Krige (a South-African engineer) and Herbert Sichel (a South-African statistic). The goal of Krige was to study the distribution of gold in some mining fields; to do that he developed a geostatistical method, now known as *kriging*. The thought process of Krige for his scope is the basic idea of geostatistical methods: estimate some property of the real world starting from some information collected in points on the field. After the works of Krige and colleagues, a rigorous mathematical formulation of the method was obtained by Matheron, a French mathematician (1961). After that a rich literature on applied cases and on further development of the mathematical structure of this kind of models started. In the following pages a brief description of kriging method is given, a more complete and rigorous description can be found in various papers and books (Isaaks and Srivastava 1990, Delfiner and Chils 2012, Hengl 2009 etc.).

The chosen approach to study the contamination of water quality in the analyzed part of the Venetian plain is in the framework of geostatistics: like any other model, there are some limitations, and the results should be observed by understanding that.

5.1 Basic concepts of geostatistics

Considering a spatial phenomenon, which in the case of this study is the chemical contamination of aquifers, a new aleatory variable Z, which is the concentration in water of a pollutant, can be defined as a function of the position s.

The fundamental definition, at the base of every geostatistical model, is the following:

$$Z(s) = Z^*(s) + \varepsilon(s) \tag{1}$$

Where the value of the spatial aleatory variable Z is the composition of $Z^*(s)$, also called trend surface, which describes the structural (global) characteristics of the phenomenon and is a deterministic component and $\varepsilon(s)$ which accounts for random spatial correlation.

One of the assumptions is related to the stationarity of the process; in particular there are many types of stationarity: strong, second order and intrinsic.

The strong stationarity is satisfied when the probability distribution of $Z(s)$ does not depend on the position inside the domain, and so measurements on different locations can be considered as realizations of the same aleatory variable.

The second order stationarity assumes that the expected value of the aleatory variable Z(s) inside the domain is a constant:

$$E[Z(s)] = m \tag{2}$$

and the covariance function inside the domain, also called covariogram, $C(h)$ of two casual variables separated by a distance h is depending only on the distance itself:

$$C(h) = E[(Z(s+h) - m) \cdot (Z(s) - m)] \tag{3}$$

Note that if the distance h is zero, then the value of $C(0)$ is simply the variance of the sample composed by the realizations of the aleatory variable $Z(s)$. As a consequence of that, this type of stationarity requires the expected value to be constant inside the domain and a finite value of variance which does not depend on the position. The intrinsic stationarity does not have this requirement and for this reason is considered "weaker". This type of stationarity includes equation 2 (expected value is costant inside the domain) and allows the variance to be function of distance h:

$$\frac{1}{2}Var[Z(s+h) - Z(s)] = \frac{1}{2}E[(Z(s+h) - Z(s))^2] = \gamma(h) = C(0) - C(h) \tag{4}$$

When dividing by 2 the obtained function is called semivariogram, which is utilized in geostatistics to describe spatial correlation, and is denoted by γ. It can be demonstrated that if the second order stationarity is satisfied than the intrinsic stationarity is satisfied, whereas the other way around is not true.

5.2 The semivariogram

The spatial auto correlation of an aleatory variable, such as the concentration in water of a chemical, can be inspected visually with the semivariogram. The generic formula for the calculation of the semivariance, as a function of a certain distance h, is:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N} (Z(x_i + h) - z(x_i))^2 \quad (5)$$

For a group of known points N, the distance matrix:

$$D = \begin{bmatrix} d_{i,j} & \dots & d_{i,N} \\ \dots & d_{i+1,j+1} & d_{i+1,N} \\ \dots & \dots & d_{N,N} \end{bmatrix}$$

With:

- $i, j = 1, 2, \dots, N$
- $d_{i,j} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2}$

So every element of the matrix is the Euclidean distance between two points (X,Y are UTM metric coordinates) and the distance matrix contains all the possible distances of the group of known points N. For every distance is possible to calculate the value of the corresponding semivariance:

$$\gamma_{i,j} = \frac{1}{2}(z_i - z_j)^2 \quad (6)$$

With z_i and $z_j$the values of the variable measured at two distinct generic points: by plotting the distance matrix and the corresponding value of semivariance calculated in every single point, the so called "semivariogram cloud" of figure 16 is obtained.

Is clear that is almost impossible to investigate any type of spatial correlation from the semivariogram cloud: for this reason, is necessary to split the cloud in a number of classes with constant width (called *lag*) and assign to each class the mean value of semivariance composed by the points within the considered lag h.

$$\widehat{\gamma_k} = \frac{1}{N_k} \sum_{i,j}^{N} \gamma_{i,j} \quad (7)$$

With $k = 1, \dots, n° \, of \, classses$ and N_k the number of points inside one class. It is easy to see that in different lag classes there could be different numbers of points: at least 30 points in a lag class is considered to be a reasonable number to obtain a meaningful average (*Journel et al. 1978*).

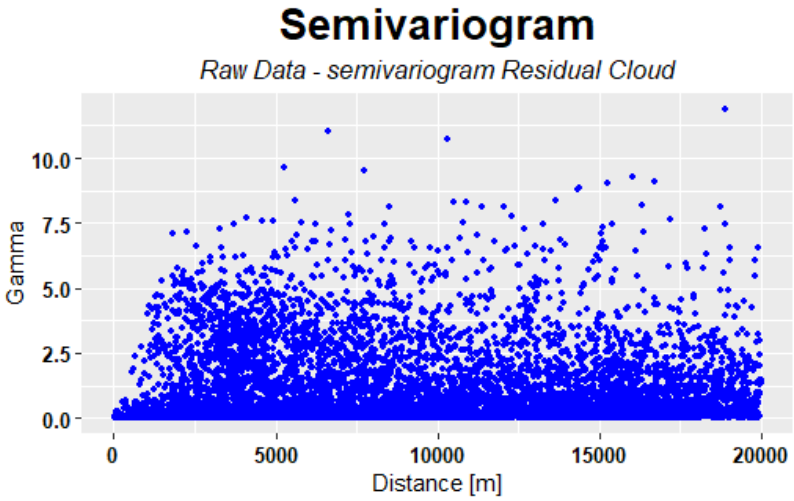


Fig. 16: example of semivariogram cloud

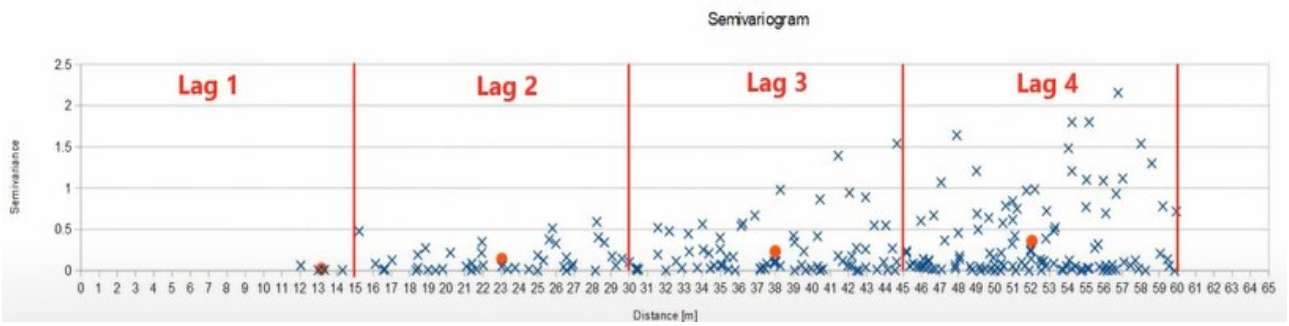


Fig. 17: semivariogram cloud and lag averages

The number of combinations of points, forming the cloud, which will be then grouped inside a class of distance (lag) and averaged to obtain one experimental point of the variogram, with n the number of points inside a range of distance and $k = 2$ (considering pairs), is:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{7}$$

The number of points n is higher if the considered range is higher; for this reason another parameter to set for variogram modelling is the value of this "searching" range. This is, for isotropic variograms, corresponding in calculating semivariances of points falling inside a circle of a certain ray (see figure). This is usually set at half the maximum distance between two points observed (authors also

indicate 1/3 of this distance). In other words, all the experimental points after this distance, called "max distance", are not considered for variogram modelling, due to the fact that they do not add any significant information regarding spatial correlation. Usually, after a certain distance points are simply not correlated.

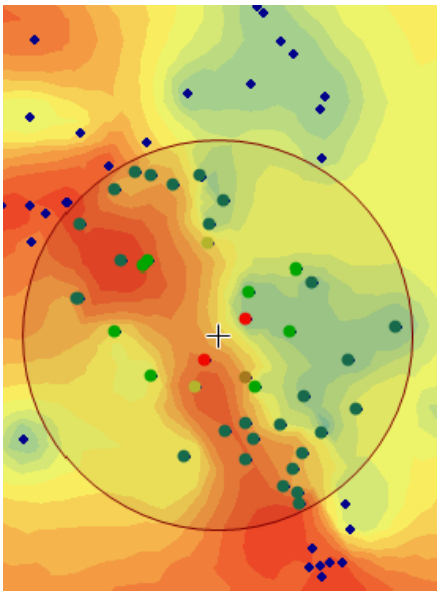


Fig. 18: searching neighborhood for semivariogram modelling

The obtained graph, considering the lag averages, is called experimental variogram. From the fit of the points with a continuous function, which can be achieved with various method, the theoretical variogram is obtained. From the behavior of it, is possible to infer some properties of the analyzed phenomenon.

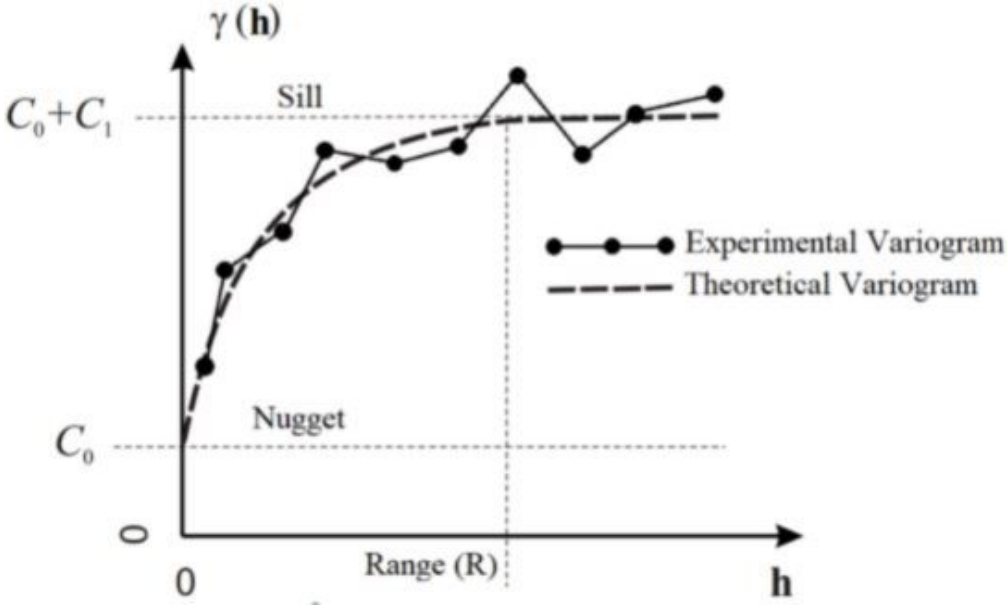


Fig. 19: example of semivariogram (source: Yasojima et al, 2019)

By definition, the value of $\gamma(h)$ is equal to zero in the origin, $\gamma(0) = 0$, although often the variogram does not tend to zero when approaching the origin. For this reason, a discontinuity is often observed at the origin where the variogram is equal to zero in the origin and abruptly goes to a positive value just outside: this discontinuity is called *nugget.* The presence of the nugget effect may be caused by:

- poor spacing of the measurements, resulting in the impossibility to catch the micro-scale variability of the natural phenomenon
- uncertainty in the measurements and poor sensibility of instruments
- accidental errors in measures

Another series of information can be obtained by investigating the behavior of the variogram at large distances. The value at which the semivariance stabilizes is called *sill:* the value should approximately be equal to the sample variance. If the second order stationarity is not satisfied, and the intrinsic one is respected, the variogram appears without a well-defined sill. The distance at which the variogram reaches the sill is called *range*: this value of distance ideally separates the portion of correlated and uncorrelated data. For this reason, the range indicates the dimension for which the measures can be linked by some kind of relationship, whereas at distances greater than the range the measures are independent one another. Empirical variograms are not directly applicable in interpolation techniques. For this purpose, theoretical variograms (also called model variograms), i.e., mathematical functions that fit the experimental points, are used. Model variograms must respect some mathematical properties of the semivariance function:

- is a function of positive values only, i.e.

$$\gamma(h) \geq 0 \tag{8}$$

- the value in the origin is always zero, even if a nugget effect is present, i.e.

$$\gamma(0) = 0 \tag{9}$$

- is an even function, i.e.

$$\gamma(h) = \gamma(-h) \tag{10}$$

- at h approaching infinite, the function grows slower than h^2: if that does not happen it can be demonstrated that the intrinsic stationarity is violated. For this aspect the rigorous demonstration is skipped.

$$\lim_{h\to\infty} \left(\frac{\gamma(h)}{h^2}\right) = 0 \tag{11}$$

There are many models that satisfy the conditions just exposed: the main ones, which will be used for the analysis, are reported hereafter. The nugget is from now on indicated by C_0 , the sill is composed by $C_0 + C_1$ where C_1 is also called partial sill and r is the range.

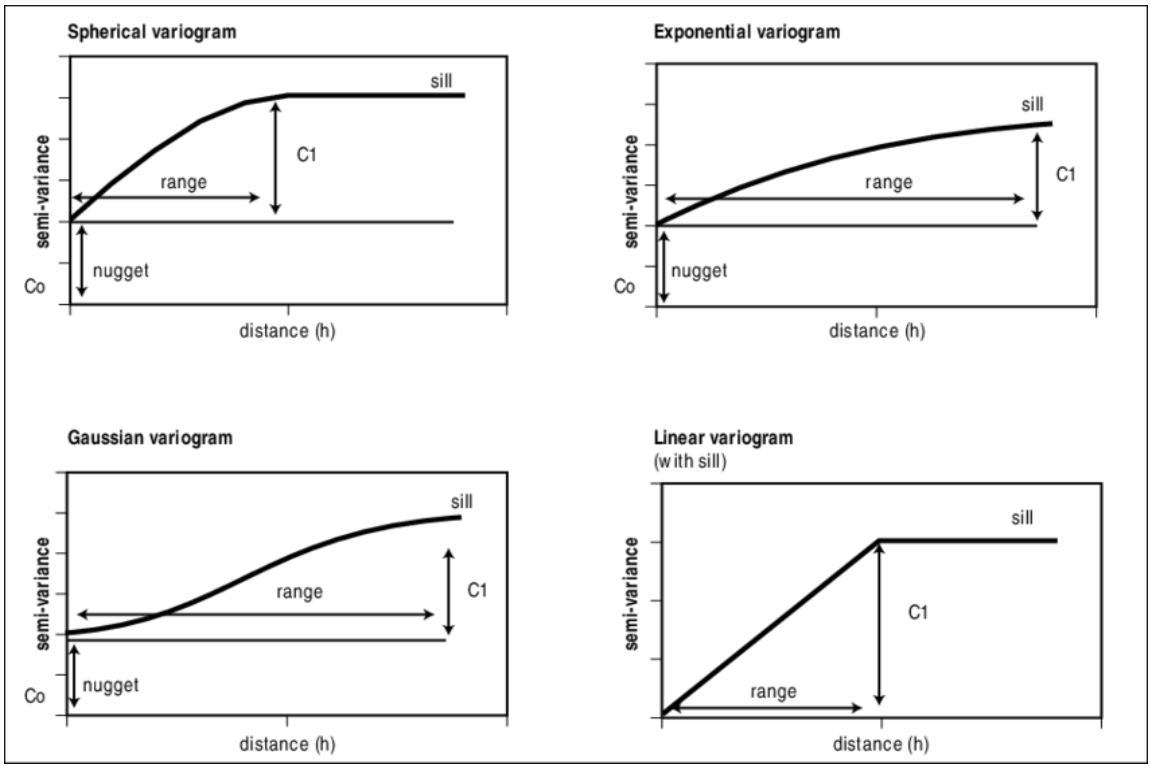


Fig. 20: most frequently utilized model variograms

SPHERICAL MODEL

It is the most utilized in environmental problems. It has a linear behavior close to the origin and has a range. The formulation is:

$$\gamma(h) = \begin{cases} 0 & if\ h = 0 \\ C_0 + C_1\left[\frac{3h}{2r} - \frac{1}{2}\left(\frac{h}{r}\right)^3\right] & if\ 0 < h < r \\ C_0 + C_1 & if\ h > r \end{cases}$$

Just as empirical observation, this model is the most well-performing among the possible ones and visually well approximates a wide number of cases.

EXPONENTIAL MODEL

Like the spherical model, the behavior near the origin is linear. For this model the sill is only reached asymptotically and for this reason theoretically there is no range.

$$\gamma(h) = \begin{cases} 0 & if\ h = 0 \\ C_0 + C_1\left[1 - \exp\left(\frac{-h}{r}\right)\right] & if\ h > 0 \end{cases}$$

For practical purposes, an effective range is defined at the point where the 95% of the sill is reached: to achieve this is sufficient to add a constant equal to 3 in the model, resulting in:

$$\gamma(h) = C(1 - \exp(-3)) = 0.95C$$

GAUSSIAN MODEL

Also in this case the sill is only reached asymptotically; for this reason the same procedure of the exponential model is adopted, defining a practical range and sill, resulting in:

$$\gamma(h) = \begin{cases} 0 & if\ h = 0 \\ C_0 + C_1\left[1 - \exp\left(\frac{-h}{r}\right)\right] & if\ h > 0 \end{cases}$$

The difference between this model and the exponential and spherical, is the behavior near the origin: in this case is not linear but parabolic.

LINEAR MODEL

For this model there are two possible solutions: one considering a sill (like in fig 20) and one without. Due to the fact that with a well-defined sill the previous models outperforms the linear one, only the formulation with no sill is considered. As already mentioned, if the semivariance does not stabilizes around a sill, the stationarity of second order is not satisfied and only the intrinsic one is respected. The model is very simple, and has no sill or range:

$$\gamma(h) = C_0 + bh$$

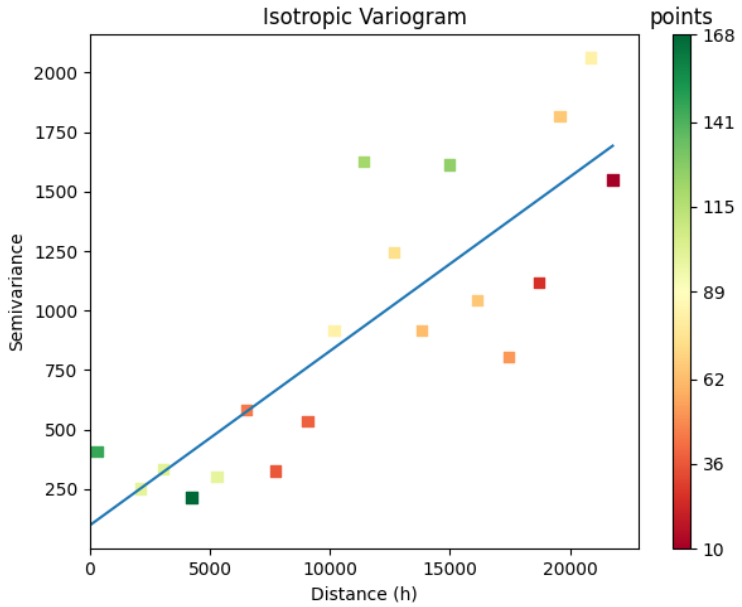


Fig. 21: example of linear model with no sill

VARIOGRAM FIT

Given the shape of the experimental variogram, the set of sill, range and nugget has to be set in order to obtain the best theoretical fit. This operation can be achieved in many ways: visually, ordinary least squares, weighted least squares and others. In this thesis the function fit.variogram of R has been used: it is based on weighted least squares. The method has the positive aspect of giving more importance to the lag classes with more points and is considered by Cressie (1985) a right compromise between simplicity of implementation and accuracy.

The estimate of the set of parameters Θ of the variogram is obtained, with n lag classes, $N(k)$the number of points in every lag class, and h the lag distance, by minimization of the following argument:

$$\sum_{k=1}^{n} \frac{N(k)}{h_i^2} \left(\frac{\bar{\gamma}(h_i)}{\gamma(h, \Theta)} - 1 \right)^2$$

Note that weights (the first fraction of the sum) give more importance to the lag classes containing more points. The minimization is achieved iteratively, and an initial guess of the values of nugget, sill and range is required. These initial values are estimated by visual inspection of the experimental variogram.

5.3 Ordinary kriging

Kriging is a geostatistical interpolation technique that can provide prediction of a quantity in non-sampled points. Another positive aspect of this method is the ability to produce an estimation of the error of prevision.

There are many types of kriging (ordinary, universal, co-kriging etc) which derives from different assumptions on the analyzed phenomenon; in this study the ordinary kriging is the used technique. The main hypothesis of this method is the satisfaction of the second order stationarity (previously described), and the existence of a meaningful spatial correlation. There are no real statistical ways to check the stationarity, but some tools can give information:

- insert the X,Y coordinates and measured values of the inspected variable on a scatterplot: if there is some kind of meaningful trend, for example a clear linear increasing trend in N-S direction, probably is a sign that the expected value of the variable is not constant inside the domain (second order stationarity is violated). In these cases, *universal kriging* can be used, which permits to specify a systematic spatial structure of the variable (for example a quadratic behavior along a coordinate).
- If the variogram does not stabilize around the sample variance, second order stationarity may be violated (already described).

Another aspect is related to the normality of the input dataset: although is not a mathematical requirement, normal distributions of known values provide the best prediction, whereas skewed distributions may produce unsatisfying results. For this reason, is important to perform exploratory data analysis on the input data and consider transformations before any interpolation.

This interpolation technique (ordinary kriging) derives a prediction Z^* in an unsampled point x with a linear combination of the N known values of the analyzed aleatory variable Z (in this study Z is a chemical concentration in water):

$$Z^*(x) = \sum_{i=1}^{N} \lambda_i Z(x_i) \tag{12}$$

The coefficients λ_i of this linear combination are called weights. Kriging can be schematized as a two-step process: the first one is to "catch" the spatial correlation structure of the sampled points by fitting a variogram, and second, weights derived from this covariance structure are used to interpolate unsampled points in the domain.

The main properties of the method are:

- Is an exact interpolator: in the observation points x_i the value of the prediction z^* is equal to the actual measure z available in the point.
- The values of weights are determined by the shape of the variogram and the distance from observed points.
- Kriging weights are calculated such that point nearby to the location of interest are given more weight than those far away.
- The sum of the weights is equal to 1 (demonstration hereafter).

The error associated with the prediction of eq 12 is:

$$Z^*(x_0) - \sum_{i=1}^{N} \lambda_i Z(x_i) \tag{13}$$

The estimator can't be distorted, and so the mean error has to be zero:

$$E[Z^*(x_0) - \sum_{i=1}^{N} \lambda_i Z(x_i)] = 0 \tag{14}$$

And recalling the condition of second order stationarity, $E[Z(x)] = m$, the previous equation is:

$$m[1 - \sum_{i=1}^{N} \lambda_i Z(x_i)] = 0 \tag{15}$$

And so

$$\sum_{i=1}^{N} \lambda_i = 1. \tag{16}$$

The variance of the prevision error, which is the variance of estimation, is the second order moment of the difference between estimated and measured values, and is:

$$\sigma_k^2(x) = Var[Z(x) - Z^*(x)] = E\left[\left(z(x_0) - \sum_{i=1}^{N} \lambda_i Z(x_i)\right)^2\right] =$$

$$= E\left[Z(x_0)^2 - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j Z\left(x_j\right) - 2\sum_{i=1}^{N} \lambda_i Z(x_i) Z(x_0)\right] =$$

$$= C(0) + \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j C\left(x_i - x_j\right) - 2\sum_{i=1}^{N} \lambda_i C(x_i - x_0) \tag{17}$$

Even though using covariances or semivariance give the same results, the previous equation is rewritten considering semivariance function γ, remembering that $\gamma(h) = C(0) - C(h)$:

$$\sigma_k^2(x) = 2\sum_{i=1}^{N} \lambda_i \gamma(x_i - x_0) - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j \gamma\left(x_i - x_j\right) \tag{18}$$

In order to obtain the maximum precision, is necessary to minimize the variance; the Lagrangian optimization (Lagrange multiplier μ) is used to find the minimum of σ_k^2, calculating the partial derivatives for weights and equaling to zero:

$$\frac{\partial \sigma_k^2}{\partial \lambda_i} = \frac{\partial}{\partial \lambda_i}\left[2\sum_{i=1}^{N} \lambda_i \gamma(x_i - x_0) - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j \gamma\left(x_i - x_j\right) + 2\mu(1 - \sum_{i=1}^{N} \lambda_i)\right]=0$$

$$= 2\gamma(x_i - x_0) - 2\sum_{j=1}^{N} \lambda_j \gamma\left(x_i - x_j\right) - 2\mu = 0 \tag{19}$$

And considering that the sum of weights is 1, the following system is obtained:

$$\begin{cases} \sum_{j=1}^{N} \lambda_j \gamma\left(x_i - x_j\right) + \mu = \gamma(x_i - x_0) & i = 1, \dots, N \\ \sum_{i=1}^{N} \lambda_i = 1 \end{cases} \tag{20}$$

Which written in matricial form is:

$$\begin{bmatrix} \gamma(0) & \gamma(x_1 - x_N) & 1 \\ \gamma(x_N - x_1) & \gamma(0) & 1 \\ 1 & 1 & 0 \end{bmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \lambda_N \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma(x_1 - x_0) \\ \gamma(x_N - x_0) \\ 1 \end{pmatrix} \tag{21}$$

The system represents a linear combination of N+1 equations in N+1 unknowns: semivariogram is an even function, coefficients matrix is symmetric, positive, and consequently invertible. Kriging weights can be calculated by inverting the matrix and multiplying it with the vector of known values which applied to the basic linear combination of kriging, $Z^*(x) = \sum_{i=1}^{N} \lambda_i Z(x_i)$,gives the prediction value at unknown locations.

In addition to estimation of values at new locations, the prediction variance is obtained:

$$\sigma_{k0}^2 = \sum_{i=1}^{N} \lambda_j \gamma\left(x_i - x_j\right) + \mu \tag{22}$$

5.4 Lognormal kriging

Lognormal kriging is basically an ordinary kriging except that the variogram is computed and modelled after transforming the data with a logarithm (*Webster and Oliver, 2001*). The motivation behind this operation on the input data is to cope with skewed data: kriging offers the best unbiased predictions when operating on normal-like distributions. The logarithmic transformation reduces data variance and can give a dataset which is more suitable to obtain meaningful kriging estimates; in case the logarithmic transformation does not produce a normal distribution, the kriging method may not be the most appropriate geostatistical tool.

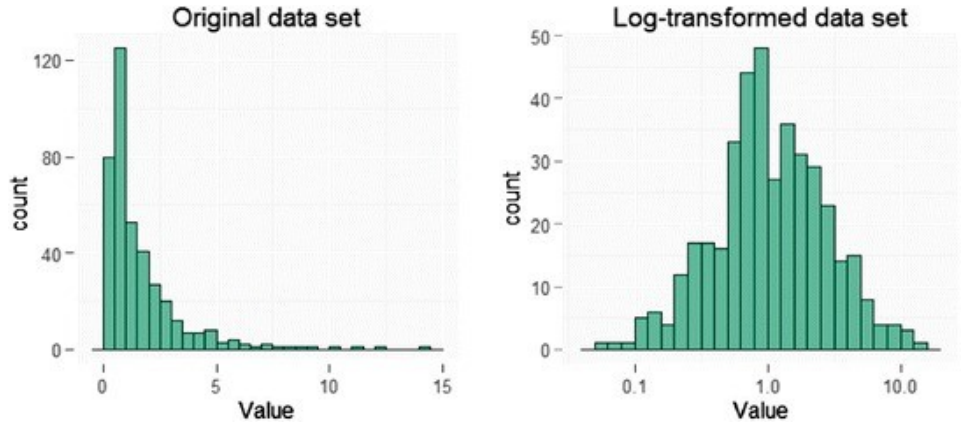


Fig. 22: example of log transformation of dataset

One problem with logarithmic transformations is what to do with the zero values, the main ideas are:

- Add a constant equal to 1 in order to avoid negative values, $x' = \ln(1 + x)$
- Substitute the zeros with the next minimum value recorded in the dataset; this operation may produce negative values if the minimum value is <1

Kriging estimates and prediction variances may be transformed back to the original scale: if a simple inverse operation is done on the values, i.e. exponentiation, the obtained back-transformed values have some level of bias. For this reason, various authors (Laurent, Cressie, Journel, Yamamoto) proposed ways to correct this problem. In this study the approach utilized is the one given by Cressie in the Journal of Mathematical Geoscience (2006). The formula is:

$$y = \exp\left[\ln(y_{OK}) + \frac{1}{2}\sigma_{OK}^2 - \mu\right] \tag{23}$$

Where μ is the Lagrange multiplier and σ_{OK}^2 is the prediction variance.

The prediction variances are also biased and therefore some back-transformations have been proposed to cope with this problem. Unfortunately, the approach of Cressie (2006) showed bias in applied studies also after back-transformation; in particular the back-transformation of variances is susceptible to outliers and overall can furnish poor results. Many papers (for example Yamamoto, "*on unbiased back-transformation of lognormal kriging estimate*", 2007) proposed other ways to correct bias, but this thesis is not a literature review of the many possible methods to apply kriging with logarithmic transformation. For this reason, in this study the back-transformation of variance is simply not performed, and the obtained results of prediction variances are presented in logarithmic scale.

The methodology used by Kishnè et al. in "*Comparison of Ordinary and lognormal kriging on skewed data of Total Cadmium in forest soils of Sweden* (2003)" has been used: for datasets presenting a skewness coefficient greater than 1, the lognormal kriging is adopted, otherwise ordinary kriging without any transformation is preferred. Is important to note that no assumptions on the normality of the transformed dataset are made; non-normal distribution can have skewness lower than 1 too.

5.5 Semivariograms and kriging with anisotropy

So far, the nature of the aleatory variable, which is in this case the concentration in water, has implicitly been considered isotropic. In many natural settings, the phenomenon is instead anisotropic: the variable shows higher autocorrelation in one direction than another. When using a kriging method and choose to account for anisotropy, the empirical semivariogram will show a different spatial relationship for each direction. The analysis is equal to the one described in the paragraph 5.3, but the value of spatial variable h, which was treated only as a module, is now considered as a vector in which also the direction is present. The formula for the directional semivariogram is:

$$\gamma(\vec{h}) = \frac{1}{2N(\vec{h})} \sum_{i=1}^{N} (Z(\vec{x_i} + \vec{h}) - z(\vec{x_i}))^2$$

A practical way to inspect anisotropy is to observe the directional variogram and evaluate if the values of sill and range are the same in all directions: if meaningful and evident differences emerge, anisotropy should be considered. The isotropic model is the same in all directions, whereas the anisotropic model reaches the sill more rapidly in some directions than others.

The length of the longer axis to reach the sill is called the major range, and the length of the shorter axis to reach the sill is called minor range. In other words, the "searching neighborhood" to calculate semivariances (and will be used to assign the kriging weights) is no longer a circle but is instead an ellipsis.

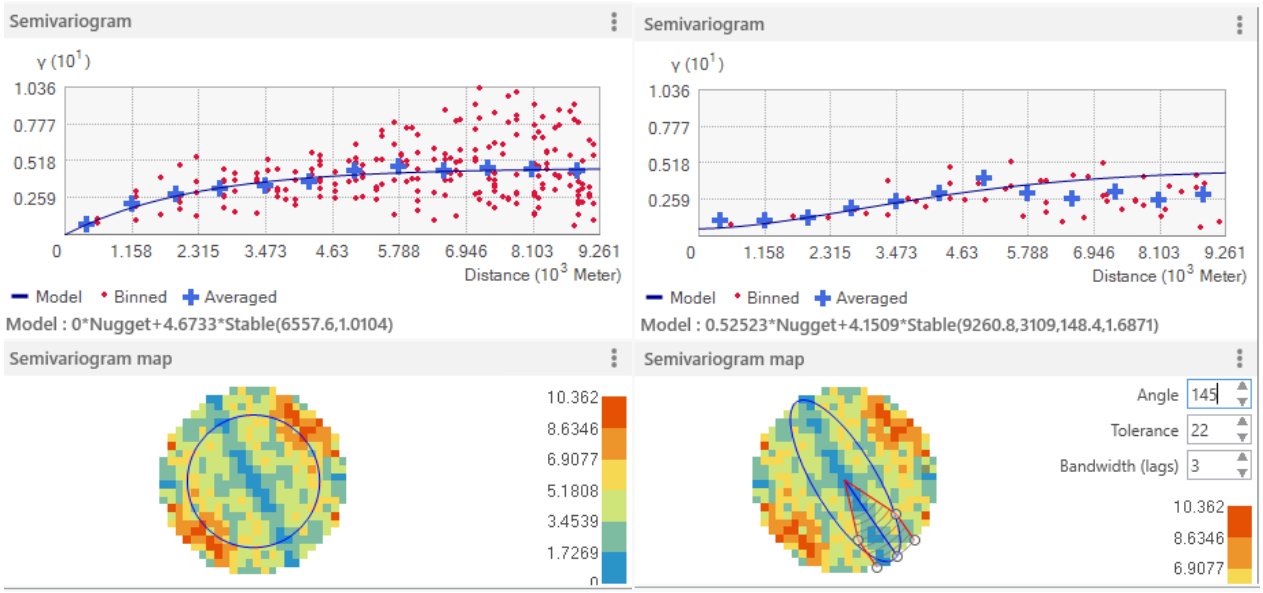


Fig. 23: isotropic (left) VS anisotropic model (right)

To calculate directional variograms the following parameters are needed:

- Azimuth angle θ: is the angle the major axis forms with the North direction. So for example $\theta = 90°$ indicates that the major axis is directed in the direction East-West.
- Azimuth tolerance: an offset from the azimuth angle in order to include more points. If the searching direction is a line, the chances that points are perfectly aligned in one direction is very low. This parameter is usually set at 22°
- Bandwidth: a limitation on the sides of the bin. If this is not specified, azimuth tolerance would have the effect of considering more and more points at higher distances. This is usually set at 3 times the value of the lag.

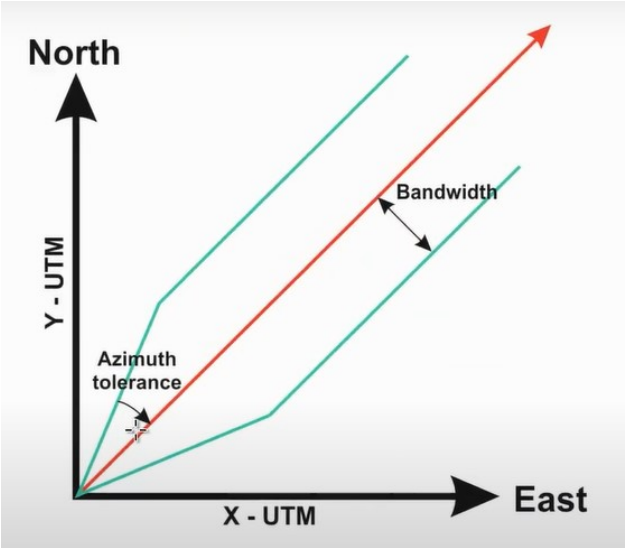


Fig. 24: azimuth tollerance and bandwith

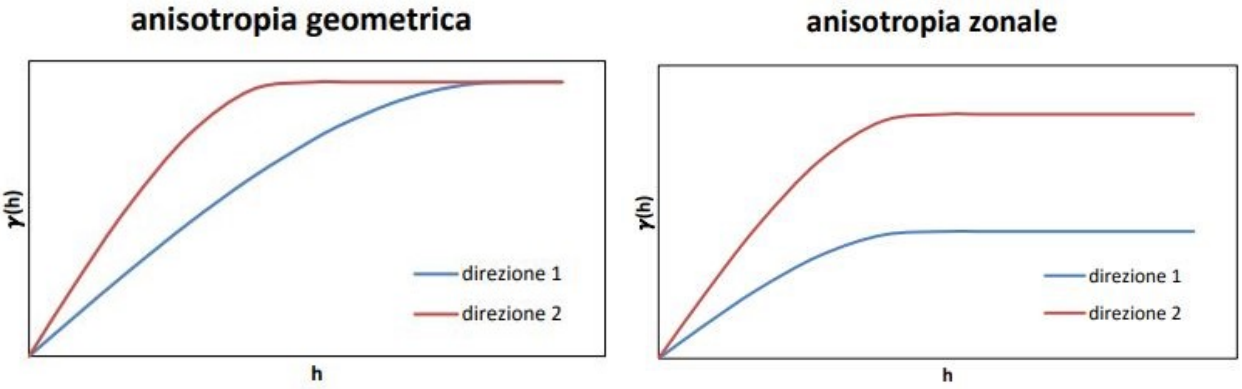


Fig. 25: geometric (left) and zonal (right) anisotropy

There are two types of anisotropy: geometric and zonal. The geometric type is when directional semivariograms have constant sill but different ranges. It is then possible to recognize an ellipse of anisotropy, whose radius varies according to the direction and coincides with the range of the

variogram. The other type, the zonal, is when in different directions the reached sill is different. This kind of anisotropy produce a series of theoretical difficulties for its application and therefore is not considered.

Incorporating geometric anisotropy in the ordinary Kriging method is simply a matter of applying a transformation to the distances. An affine transformation keeps point distances in one direction unchanged and stretches distances in the direction perpendicular to it. The final step of the transformation will give the value of variogram (which is anisotropic in UTM coordinates), in a transformed reference system in which the variogram, with range equal to the major one, is isotropic. The procedure to reach this goal is as follows:

- The first step is a rotation of the axis of the original reference system (X, Y) to a position parallel to the major axis of correlation (major range).

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{pmatrix} X \\ Y \end{pmatrix}$$

- The second step is the transformation of the ellipsis in a circle of ray equal to the major range.

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \phi \end{bmatrix} \cdot \begin{pmatrix} X \\ Y \end{pmatrix}$$

  With ϕ, called anisotropy ratio, equal to the ratio of major and minor axis.

- Rotation, in the opposite direction of the one in the firs step, of the new axis to replace them in the original position.

$$\begin{bmatrix} u'' \\ v'' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{pmatrix} u' \\ v' \end{pmatrix}$$

- The final step is calculating the variogram in the transformed reference system:

$$\gamma(X, Y) = \gamma\left(\sqrt{u'' + v''}\right)$$

Apart from the transformed distances used for defining the search radius and the modified variogram function, the Anisotropic Kriging algorithm is equal to the Ordinary Kriging operation.

5.6 Leave-one-out cross validation (LOOCV)

The accuracy of the model essentially depends on the capacity of the variogram to really model the spatial variability of the variable. To check the performance of the kriging interpolation the leave-one-out cross validation is adopted. The method consists in:

- A known point is iteratively removed from the set of measures and its value is estimated by performing kriging interpolation on the remaining N-1 values.
- The obtained predictions are then compared with the observed values.

Various measures of error are possible to quantify how much the predicted values differ from the real ones:

- Mean error: is simply the mean of difference between measured values and predicted values. For this reason, this indicator is affected by the order of magnitude of the variable. If there are no systematic errors on the predictions, the value of this indicator should be close to zero. The formula is:

$$ME = \frac{1}{N}\sum_i^N z^*(x_i) - z(x_i) \tag{24}$$

- Mean standardized error: solves one of the problem of the mean error, which is its dependance on the order of magnitude of the variable, by standardizing the mean with the variance. The formula is:

$$MSE = \frac{1}{N}\sum_i^N \frac{z^*(x_i) - z(x_i)}{\sigma^2} \tag{25}$$

- Root mean squared error: is one of the most widely used, it has the same unit of measure of the interpolated variable and consequently comparisons are easy to make. The formula is:

$$RMSE = \sqrt{\frac{1}{N}\sum_i^N (z^*(x_i) - z(x_i))^2} \tag{26}$$

The RMSE is the standard indicator of accuracy of the model in this study.

5.7 Preliminary analysis and basic statistical indicators

The preliminary analysis of the dataset allows a more precise and aware geostatistical exploration. As already discussed, the best estimate given by kriging is for input data that is normally distributed and does not present outliers. Unfortunately, very often when dealing with cases of contamination, the distributions are highly skewed and with many outliers: as a consequence of that the errors (variance of prediction) can present skewed distributions. The logarithmic transformation can solve, or at least attenuate, this problem. To quantify and recognize patterns and behavior of the input data, many statistical indicators can be used, starting from the classical statistics like mean and standard deviation. To evaluate the variability of the wells time series, the coefficient of variation has been used:

$$CV = \frac{\sigma}{\mu}$$

With σ the standard deviation and μ the mean. This simple coefficient gives an idea of how much the data spread around the mean, giving a readily quantification of the variability: a threshold to discriminate highly fluctuating distributions and more or less stable ones is set to 0.3 (30%).

Another important indicator is the skewness coefficient g, which quantifies the asymmetry in the distributions: as already mentioned, if the coefficient is greater than 1 a logarithmic transformation is operated. The formula of the skewness coefficient, with n observed values of a variable z, with mean $\bar{z}$ and standard deviation σ is:

$$g = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \frac{(z_i - \bar{z})}{\sigma^3}$$

If g is equal to zero, the distribution is perfectly symmetric.

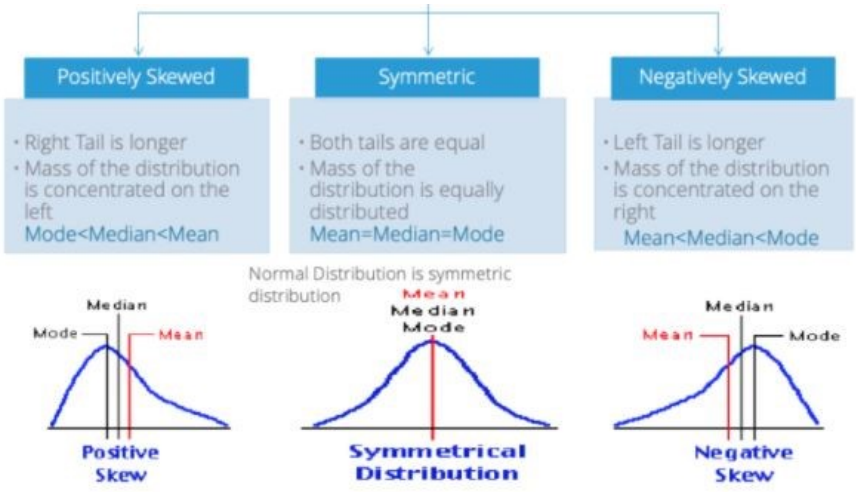


Fig. 26: types of skewed distributions

To visualize the asymmetry of the distribution is also useful to plot histograms, and depending on the shape of the graph, three types of skew are defined:

- Negative skewed distribution, or skewed to the left, left tailed: the left tail is longer, and the mass of the distribution is concentrated more on the right side of the graph.
- Positive skewed distribution, or skewed to the right, right tailed: the right tail is longer, and the mass of the distribution is concentrated more on the left side of the graph.
- Symmetric: both tails are equal.

The identification of outliers, which can substantially impact the performance of kriging method, is performed by visual inspection of the boxplot. The graph is composed by various items:

- Left edge of the box: is the first quartile $q1$ (the value greater than 25% of dataset measures)
- Right edge of the box: is the third quartile $q3$ (the value greater than 75% of dataset)
- A line inside the box which is the median
- Whiskers: they extend from the left or right edge of the box until the points of "minimum" and "maximum".
- Minimum is defined as: $min = q1 - 1.5(q3 - q1)$
- Maximum is defined as: $max = q3 + 1.5(q3 - q1)$

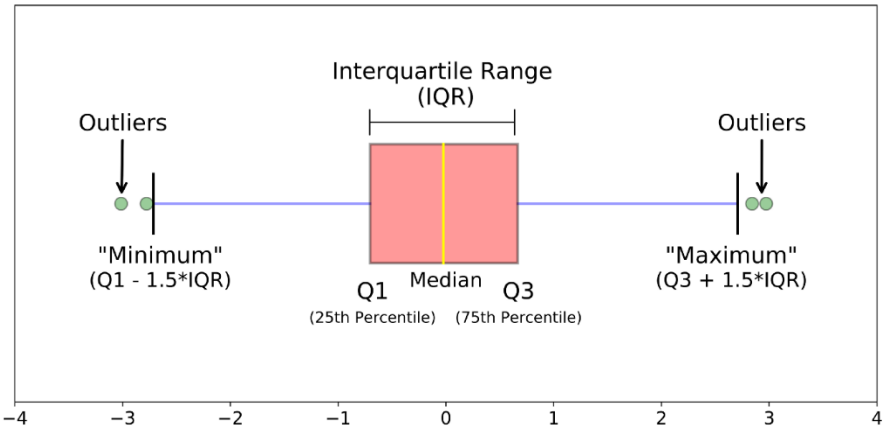


Fig. 27: boxplot meaning

Every point falling outside the points of minimum and maximum is considered an outlier. As shown by Kerry and Oliver (2007) the presence even of few outliers (1.25% of the dataset) can affect greatly the variogram model (they observed that adding few outlies provokes a nugget effect and a variation of the sill) and therefore the kriging method. As a consequence of that many authors approach the

kriging interpolation by preventively removing outliers; in this thesis outliers are not removed. Even though the performance of kriging interpolations will be affected by outliers, the presence of them is indicative of real problems of important contamination: the goal of this thesis is not to obtain the perfect kriging model but is to recognize and characterize various cases of contamination, and the geostatistical analysis is only the final part of the study.

6 ANALYZED POLLUTANTS

In this chapter, a brief description of the analyzed pollutants is presented. Water quality assessment is usually inspected by laboratory analysis on samples: the measured parameters are then compared with the maximum acceptable concentration (MAC) prescribed by law. In Italy, the main regulation nowadays regarding drinking water quality is the dlgs 31/2001. Previous regulations were d.p.r 236/88 and 319/76 (also known as "legge Merli"); other indications on maximum concentrations on water can be extracted from various World Health Organization (WHO) guidelines.
The set of analyzed parameters in an ordinary laboratory analysis comprehends many chemicals: for this reason, a preliminary screening phase is done in order to focus only on few of the most concerning parameters. The rationality of this screening is to inspect the most dangerous pollutants for human health which are consistently found in the area. To quantify the danger of pollutants, in a water safety plan framework, the Veneto region, with the help of many experts on the effects on human health of water quality, produced a classification to rank pollutants. The danger of a contaminant, called gravity, is on a scale to 1 (no concerns) to 5 (maximum concern). The classification is still in a preliminary phase and so the weights are not "official". In the preliminary screening phase of this thesis, contaminants with gravity equal to 4 and 5 were inspected: if the maximum values recorded were below more or less half of the law limit, they were automatically discarded from the analysis. Many contaminants presented occasional values over the law limit; obviously a complete analysis is very time consuming and for this reason only 3 pollutants (total solvents, nitrate and pesticides) were chosen among the possible ones; the choice was made also by consulting past publications and by taking into account historical memory of the workers of Viacqua.

Various properties are inspected to better understand the fate of the contaminant once it enters in the environmental groundwater cycle; some other aspects like its danger to human health, the limit in drinkable water and the commercial use of the products are analyzed. The main parameters used in this thesis for the assessment of the fate of contaminants (especially pesticides) in the environment are:

- **Solubility**: is the weight of a solute dissolved in one liter of solvent (water) when reaching a saturated solution. The higher the solubility, the higher the tendency of a solute to dissolve in water; as solubility increases contaminant's molecules tend to bond with water and travel with it. A less soluble contaminant is more likely to stick to the soil for longer periods of time.

Tab. 2: ranges of solubility (source: Ronald Ney, "Fate and transport of Organic chemicals in the environment", 1995)

Low water solubility	Less than 10 mg/L
Moderate water solubility	From 10 to 1000 mg/L
High water solubility	More than 1000 mg/L

- **Log K_{ow}**: octanol/water partition coefficient is defined as the ratio of the concentration of a chemical in octanol (which mimic a lipid organic phase) and water at equilibrium at a specified temperature.

$$Log(K_{OW}) = log\ \frac{concentration\ in\ octanol}{concentration\ in\ water}$$

This parameter is an indicator of the tendency of an organic compound to partition itself between an organic phase (biomass, organic content in the soil, fish etc) to soil and living organisms. Substances with high Log Kow values tend to adsorb more readily to organic matter in soils or sediments because of their low affinity for water. Chemicals with very high Log Kow values (i.e, >4.5) are of greater concern because they may have the potential to bio-concentrate in living organisms.

- **K_{OC}:** this parameter (organic carbon/water partition coefficient) indicates the mobility of a substance in the soil. Its value depends on the fraction of organic carbon present in the soil and on the soil adsorption coefficient, which measure the amount of chemical substance adsorbed onto the soil per amount of water. The formulation of the organic carbon/water partition coefficient is: $K_{OC} = \frac{K_D * 100}{\%OC}$

Tab. 3: Koc mobility class ranges. Source: FAO (recommended for use by the US EPA)

KOC (mL/g or L/kg)	Log KOC (mL/g or L/kg)	Mobility Class
< 10	< 1	Highly Mobile
10-100	1 - 2	Mobile
100-1,000	2 - 3	Moderately Mobile
1,000 - 10,000	3 - 4	Slightly Mobile
10,000 - 100,000	4 - 5	Hardly Mobile
> 100 ,000	> 5	Immobile

Where K_D is the soil adsorption coefficient (measure of the fraction of substance adsorbed in soils per amount of water) and OC is the percentage of mass of organic carbon of the soil. High value of K_{OC} means that the pollutant is strongly adsorbed into the soil and so movement throughout the soil is difficult.

Being a function of the adsorption coefficient, K_{OC} is dependent on a great number of factors like pH, particle size distribution, surface area of adsorbent etc., for this reason this parameter has a great variability on the field.

- **Half-life:** is the time it takes for a contaminant to be reduced by half. This occurs as it dissipates or breaks down in the environment after physical, chemical and biological processes. Higher values of half-life indicate the tendency of pollutants to accumulate in the environment, whereas lower values lead to quick elimination of the pollutant.

  Compounds can have different half-life in different environmental condition (i.e the half-life of a pollutant in sunny, hot periods can be different from the one for a glacial environment). Because of this fact ranges of half-life in literature can vary greatly; according to the *National Pesticide Information Center (USA)* values are considered "low" if the half-life is below 16 days, moderate if ranging from 16-59 days, high if greater than 60 days.

- **GUS (*groundwater ubiquity score*) index:** it's a parameter for a simple assessment of likelihood of leaching for a pollutant. This index is calculated as:

$$\boldsymbol{GUS = \log(half\ life) * (4 - \log K_{OC})}$$

  As the parameter increases, the potential for movement towards the groundwater increases. In other words, as the half-life is higher and tendency of the pollutant to stick to the soil (K_{OC}) is lower, the chance of a pollutant to leach into groundwater is higher.

  As a consequence of that a pollutant can be regarded as a "leacher", "transitional" and "non leacher".

Tab. 4: ranges of GUS. SOURCE: Gustafson, 1989. "GUS: a simple method for assessing pesticide leachability"

LEACHER	GUS > 2.8
TRANSITIONAL	1.8 < GUS < 2.8
NON – LEACHER	GUS < 1.8

6.1 Total solvents

In this part the parameter inspected is not a single compound but is instead the sum of various components. The name "total solvents" indicates the sum of three molecules:

- TCE, trichloroethylene (or also 1,1,2-Trichloroethene) is a halocarbon commonly used as an industrial solvent for metals, oils, colors and others; it can also be used for analgesic purposes.
- PCE, tetrachloroethylene is a chlorocarbon commonly used as a "dry cleaner fluid"
- MC, 1,1,1-trichloroethane, also known as methyl chloroform, is a chloroalkane

Often MC and PCE are mistaken for one another, due to their similarity in name and structure: in the database the values of MC are almost always zeros, so the parameter "total solvents" is really the sum of PCE and TCE only. Furthermore, the Montreal Protocol in 1996 banned the use of MC for its effect on ozone depletion; for this reason, from 1996 onwards the MC production and use has rapidly decreased to almost zero.

The analyzed solvents (PCE, TCE) are colorless, moderately soluble in water (water solubility is equal 1.1 to 1.5 g/L), non-flammable but, being similar to other organic solvents, if they react with other molecules, they can form flammable and explosive mixtures of vapors. They are very volatile, therefore they have an affinity for the atmosphere, where however they are problematic and very polluting (they also contribute to the reduction of the presence of ozone in the atmosphere). Due to this characteristic, a great quantity of solvents evaporates during their utilization: this phenomenon can also cause pollution by fall – out effect (*Bortolami et al, 1988*).

Their movement in water is regulated also by density and kinematic viscosity; for the solvents the first is greater than water, while the second is lower; these properties favor the vertical movement of the solvent from the ground towards the saturated area. Schwille ("*migration of organic fluid immiscibile with water in the unsaturated zone", 1984*) analyzed the method of propagation on the soil of solvents, concluding that the fate of the tracer, function mainly of quantity and retention capacity of soil, has two possible fates:

1) If the tracer does not reach the saturated zone, the fluid remains in the unsaturated zone and it is leached by infiltrating water; the contaminated zone tend to widen in the zones of lower permeability in which the contaminant will progressively move downwards. Due to the fact that solvents density is only slightly greater than water, the tendency to reach the deepest groundwater levels is limited.
2) If the tracer penetrates in a saturated zone (overcoming the retention forces of the unsaturated zone), it will sink, eventually reaching a low permeability layer (aquitard). In this case

solvents move mainly by following the groundwater flow, with a limited tendency to move transversally.

Other considerations regarding the fate of contaminant are linked to its relationship with environment; when these molecules reach the water table they begin to react in many ways. Solvents do not readily undergo hydrolysis or photolysis but are very slowly biodegraded by microorganisms in anaerobic conditions (typical of underground soils) to dichloroethene, vinyl chloride, and ethene.

Tab. 5: solvents main properties for assessing underground mobility (sources: PubChem database, US Agency for Toxic Substances and Disease Registry database, Roberts et al. 1982)

solvent type	solubility at 20°C (mg/L)	density (kg/m3)	log kow	koc	half-life in soil (days)	GUS index
PCE	206	1623	3.4	200	347	4.3 (L)
TCE	1.28	1464	2.6	205	230	4.0 (L)
MC	1.29	1310	2.4	107	300	4.9 (L)

The metabolites can be dangerous too; for example, vinyl chloride is very toxic. In underground water solvents cannot volatize, and with a K_{OC} around 200, is expected to be moderately mobile in soils. Unfortunately, these chemicals have very high half lives in the soil, and for this reason have plenty of time to reach the groundwater in great quantity. Solvents do not appear to bioaccumulate in animals or food-chains (US department of Health and Human services, 1993). All these properties dictate a great persistency of solvents in the underground environment where the volatilization is almost impossible.

Often the groundwater of industrialized cities is polluted by these compounds due to incorrect ways of disposing of industrial waste, accidental spills, leaks from storage tanks or sewage pipes.

The law limit in drinkable water fixed by dlgs 31/2001 is 10 microg/L; the limit is set as "sum of concentrations of PCE and TCE". Given that the law limit is clear and indicates a threshold on PCE plus TCE concentration, the present study does not focus on the singular component but only in the sum of the two.

The threshold on solvents in water for human consumption, like other components, used to be higher than the present limit; the previous law values were 30 microg/L (d.p.r 236/88) and 1000 microg/L before that (319/76). The guide value established by the World Health Organization (WHO) for drinking water is 40 microg/L. The gravity of total solvents, as sum of PCE and TCE, is set to 5 (the maximum). The effects on human health of solvents, the main exposure to danger is given by respiration of vapors (due to volatilization); the exposure trough water and food pose low health

concerns due to the speed by which solvents evaporate from water and by the small tendency to bioaccumulate. Nevertheless, these compounds are dangerous anyway; PCE and TCE are classified as 2A and 2B carcinogens and have acute effects on the central nervous system.

6.2 Nitrates

Nitrogen is a vital nutrient that helps plants and crop grow, but at high concentration can be harmful for people and nature. In the last decades, nitrate (NO_3^-) and his salts have been exploited in agriculture to help increase productivity: for this reason, excess nitrogen from agricultural sources is one of the main causes of water pollution in Europe (91/676/EEC). In the domain of interest, agricultural areas are present and constitute one of the most relevant sources of income, so an analysis on nitrates is important to recognize the more vulnerable areas. Furthermore, a concentration of Nitrate above 9 mg/L in underground water usually indicates presence of zootechnical areas and a massive use of fertilizers (WHO 2011).

The main sources of release of nitrates in water are: precipitations (acid rains have a higher nitrate concentration), synthetic fertilizers and manure spreading in agricultural fields, waste water and others.

Following the nitrogen cycle, the fate of the nitrate is related to biochemical conditions; for this reason this compound is not stable and the evolution of its concentration in water is influenced by many factors. Without going into details on the various biochemical processes, generally biological and chemical denitrification (evolution of nitrate into molecular nitrogen under anaerobic conditions) are important in many aquifers for removing nitrate: these reactions are related mainly to redox conditions for the chemical denitrification and to available carbon and presence of denitrifier populations for biological denitrification. Other processes like nitrate reduction, dilution and mineralization are potentially important in varying the mass or the concentration of nitrate in groundwater. The anaerobic conditions of the deeper aquifers, as well as the presence of appropriate redox conditions, can result in denitrification processes that can reduce nitrate concentrations, often resulting in an inverse proportional behavior of nitrate concentrations and depth. For this reason, the accumulation of nitrate is more likely to occur in the shallow aquifers.

The situation for shallow, phreatic aquifers is different and is influenced by many factors, resulting in a complex cycle to study. Obviously, this zone is sensible to direct contamination due to percolation from the upper soil as well as biological nitrification-denitrification occurring in the organic-rich

subsoil, and also by the type of cover: the principal action which avoids nitrate releases in deeper soil is the adsorption of nitrogen in the root zone by the plants. Agricultural land that is not completely and permanently covered allows downwards losses of nitrate greater than an agrarian rotation with long vegetative periods (*Juergens e Gschwind, 1989*). The non-covered soil leaves the assimilation of nitrate by plants to zero (there are no roots) and also favors the erosion of soil due to the effects of wind and water. For this reason, the adoption of intercropping culture between the harvest of the previous crop and the subsequential sowing is a considerable anti-percolation measure.
A phreatic aquifer with high (close to the soil) water table, and significant oscillations, has great potential for nitrate elimination thanks to denitrification and roots uptake: the periodic rise of the water table favors the percolation towards deeper layers in which conditions for denitrification are optimal.

The nitrate concentration in superficial aquifers, for the reasons above, and for the level oscillations due to meteorological factors, shows great temporal variability.

Case studies of regional aquifer systems have demonstrated that nitrate disappearance occurs in contaminated aquifers although numerous interpretations of removal mechanisms are possible. Most studies used indirect measurements to indicate that biological and chemical denitrification can reduce nitrate levels in ground water (R.R Lowrance, developments in agricultural and managed forest ecology, 1989).
For the reasons above, nitrate contamination has not a great transfer potential and plume travelling is unlikely to occur; nitrate pollution is for this reason of higher danger if a diffuse pollution is observed.
Veneto region assigned a gravity of 5 (the maximum) for this parameter, due to its effect on human health which includes *methemoglobinemia* (a condition for which oxygen transportation in blood is reduced) and cancer (not enough epidemiological data yet); in particular it has been demonstrated that the real harmful compound is Nitrite which forms in human body after ingestion of Nitrate (AIRC).
The maximum acceptable concentration of nitrate in water for human consumption is currently set by dlgs 31/2001 at 50 mg/L. The same value was identified by dpr 236/88.

6.3 Pesticides

The term "pesticide" is a composite term that includes all chemicals that are used to kill or control pests. In agriculture, this includes herbicides (weeds), insecticides (insects), fungicides (fungi), nematocides (nematodes), rodenticides (vertebrate poisons) and others. In general, pesticides can be classified also by their chemical structure: organochlorines, carbamates and triazines are the main ones.

The use of pesticides coincides with the "chemical age" which has transformed society since the 1950s. In areas where intensive monoculture is practiced, pesticides were used as a standard method for pest control. Unfortunately, with the benefits of chemistry have also come disbenefits, some so serious that they now threaten the long-term survival of major ecosystems by disruption of predator-prey relationships and loss of biodiversity. Also, pesticides can have significant human health consequences (FAO). Pesticides can reach the topsoil by direct spreading on agricultural fields and also by water runoff; after that they can reach the underground environment by leaching.

The behavior of these contaminants when they enter in the environment is influenced by a great number of properties of the soil and the pesticide itself; for this reason, a deterministic fate of the compounds is very difficult to obtain, due to the great number of discontinuities typical of environmental systems. Nevertheless, some general aspects affecting the movement of pesticides in the environment have been recognized in literature and are briefly reported below:

- **Solubility:** highly soluble pesticides have a greater tendency to move by runoff or leaching from the point of application.
- **Volatilization:** Highly volatile chemicals are more likely lost to the atmosphere than to water supplies. However, highly volatile compounds may contaminate water if they are also highly soluble through “wash-out” effect.
- **Degradation and adsorption:** a fraction of the pesticide is reduced (transformed) by plants uptake, microbial degradation (breakdown of chemicals by microorganisms), abiotic degradation (chemical breakdown) and adsorption (binding of chemicals to soil particles). These processes are functions of sunlight, soil and water pH, microbial activity and other soil characteristics.
- **Weather:** rainfall (and irrigation) can cause pesticides movement from the soil into the underground; also the soil moisture profile and elevation of the water table can affect the vertical movement of the chemical. In particular, a water table close to the ground offers less opportunities for the pesticide to being adsorbed or degraded (less soil to travel). For this reason a high water table is of greater concern for contamination.

- **Soil texture and organic matter:** the proportions of sand, clay and silt influence the capability of the soil to retain water and the velocity of water movement through its permeability. Soils with high clay and organic matter fractions tend to hold more water and dissolved chemicals longer; if this kind of soil is on ground level it provides a greater chance to be adsorbed and uptake by roots, leading to degradation of a great fraction of the pesticide. Coarse soils on the other hand offers a greater chance for chemicals to percolate and reach deeper groundwater systems; this phenomenon known as leaching can occur by *matrix flow*, a slow movement of water and chemicals towards lower potential zones, or by *preferential flow,* a quicker movement through cracks, wormholes, root channels and large structural voids in the soil.

Summing up, pesticides fate is controlled by the capability of the soil to retain water, which gives plants and microorganism the opportunity to degrade chemicals; the fraction of compound which is not uptake is then transported by water and the movement velocity of the pesticide is function of its properties (mainly its solubility), and of soil's conditions (mainly its permeability and moisture distribution). Once the pesticide leaches past the root zone, the main degradation process become abiotic breakdown because microbial populations in deeper soils are generally less reactive and less adapted to these chemicals. For these reasons, pesticides seem to be more persistent in deep aquifers but the fraction of chemicals reaching this zone is greatly reduced by processes occurring in the root zone.

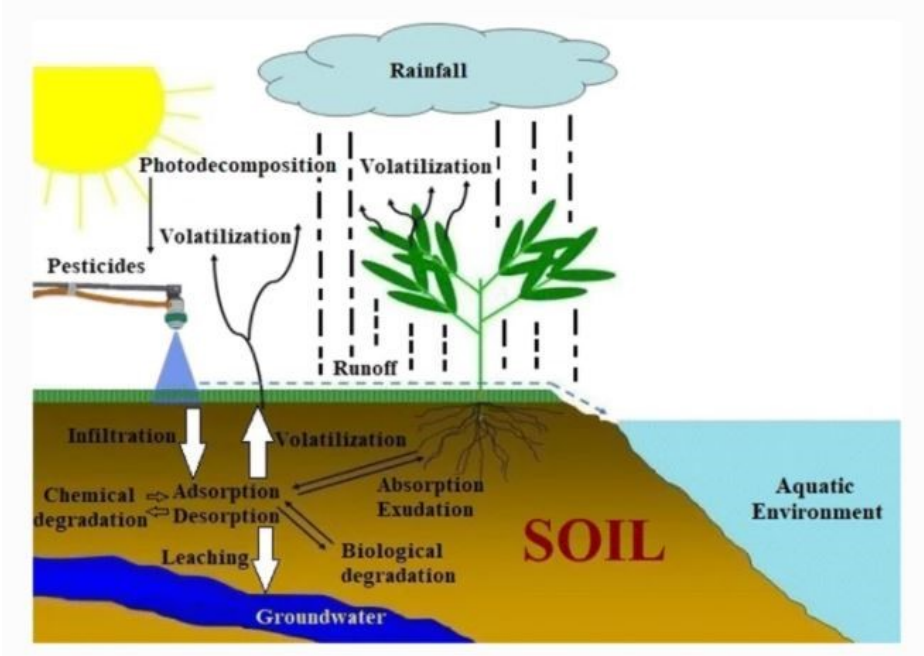


Fig. 28: fate of pesticides

The name "pesticides" indicates a large number of different chemicals: the American chemical society in 1993 recognized 13 million different chemicals with around 500'000 being added annually. Obviously, this thesis focuses only on a restrict subset of pesticides; for this reason the list of compounds present in the database is reported in table 6 which follows the path of a paper regarding pesticides in the area of interest (*Ghirardelli et al., 2021*) in which the same pollutants were studied; later on some steps of the previous work will be reported in an updated version. Note that the value reported in table 7 is the median value of a provided range (source: USDA – ARS database). Also, a brief description of the most interesting (for presence and concentrations in water in the analyzed area) pesticides, which are triazines and metolachlor, is given. The effects on human health of every single pesticide are not reported; the goal of this thesis is not a medical meta-analysis on the concerns of the various parameters.

Regarding the limit value in water for human health, the dlgs 31/2001 specify that the term "pesticide", for which the MAC is set to 0.1 microg/L, refers to: herbicides, insecticides, fungicides, nematicides, rodenticides, algaecide, acaricides, anti-mold organic products and all the metabolites and by-product chemicals of the various pesticides. The gravity assigned to every single pesticide is set to 4. For some pesticides (which are not considered in this study) the limit is lowered to 0.03 microg/L. Moreover, there is also a limit on the "total pesticides" which indicates the sum of all singular pesticides; this value is set to 0.5 microg/L.

Tab. 6: general informations on the considered pesticides (TR = triazinic group, ME = metabolite of a parent pesticide, CA = cloroacetanilide, AM = acetamide)

pesticide	group	EU ban	WHO guide value (microg/L)	dlgs 31/01 limit (microg/L)
ATRAZINE	TR	1992	2	0.1
DEA	ME	\	\	0.1
DET	ME	\	\	0.1
TERBUTILAZINE	TR	\	7	0.1
DACT	ME	\	\	0.1
SIMAZINE	TR	2004	2	0.1
ALACHLOR	CA	2006	20	0.1
METOLACHLOR	CA	2003	10	0.1

Tab. 7: pesticides main properties for assessing underground mobility (sources: PubChem database, USDA – ARS pesticides properties database)

pesticide	solubility at 20°C (mg/L)	density (kg/m3)	log kow	koc	half life in soil (days)	GUS index
ATRAZINE	33	1230	2.6	147	173	4.10 (L)
DEA	NA	-	-	-	-	-
DET	NA	-	-	-	-	-
TERBUTILAZINE	5	1113	3.4	230	60	2.95 (T)
DACT	NA	-	-	-	-	-
SIMAZINE	6	1302	2.1	140	89	3.61 (L)
ALACHLOR	240	1133	3.5	124	27	2.73 (T)
METOLACHLOR	530	1120	3.1	70	141	4.63 (L)

6.3.1 Triazines: atrazine (ATR), terbutilazine (TBR), simazine and their metabolites

Triazine is a class of nitrogen-containing heterocycles compounds; this group is the base for many pesticides formulations. In particular, atrazine, terbutilazine and simazine have been extensively used for herbicide purposes in agriculture. In the study area of this work, atrazine and terbutilazine are the main herbicides used for treatment of maize and vineyards, which are the main cultivations of the area (*Otto et al, 2007; Zanin et al, 1993*). These compounds can pose danger to human health in a variety of ways. A problem of these pollutants is posed by their metabolites which show similar effects of the parent molecule on human health (*WHO, 2010*).

In Italy the ban of atrazine entered into force in 1982, although its use had been constantly declining since 1986 (*Sbriscia et al 1998*). This chemical was one of the most used pesticides in the Vicenza area and overall in Europe; for this reason atrazine is one of the most studied pesticide. Its effect on human health comprehends damages to the reproductive system, liver and heart (the last two have been only observed on rats so far). Many epidemiologic studies conducted on manufacturing workers, pesticide applicators, and farming families do not indicate that triazines are carcinogenic in these populations. A rat-specific hormonal mechanism for mammary tumors has now been accepted by US

EPA, International Agency for Research on Cancer, and the European Union. Chlorotriazines do influence endocrine responses, but their potential impact on humans appears to be primarily on reproduction and development and is not related to carcinogenesis (*Jova and Howd, 2011).*

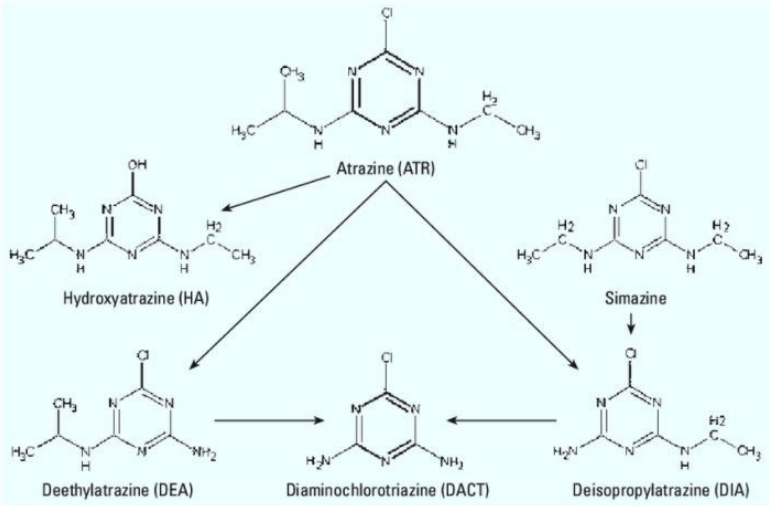


Fig: 29: atrazine main metabolites

Focusing on atrazine, its degradation can be physiochemical or biochemical, and more than 15 metabolites have been identified (Jensen, 1982). Atrazine is catabolized by plants, microbes, and animals to yield four major metabolites (Figure 29): the biodegradation is slow and leads to hydroxylated derivatives (hydroxyatrazine HA), and a series of dealkylated derivatives such as diaminochlorotriazine DACT, deisopropylatrazine DIA and de-ethylatrazine DEA.

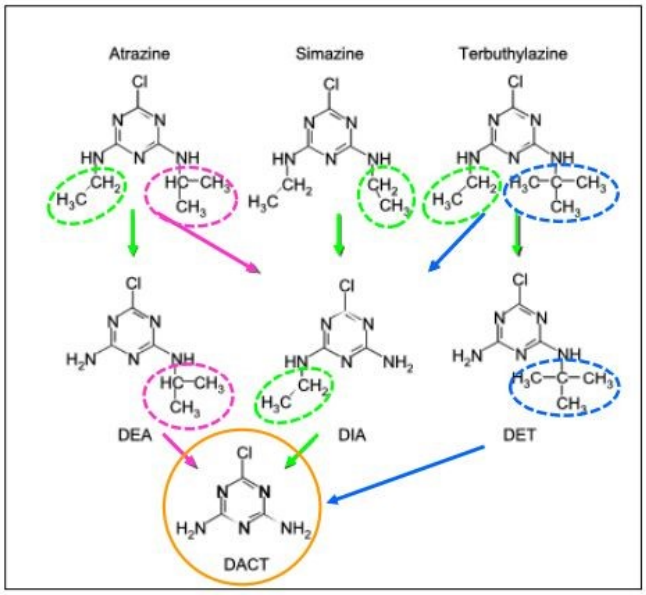


Fig. 30: metabolites of triazines and the pathway to DACT

These metabolites seem to have similar (milder) effects on organism as the parent molecule but show a longer persistency in the environment (*Price, Herbicides, Physiology of action and safety, 2015*).

Other chlorotriazines used in addition to or instead of ATR in the field, such as propazine and simazine, are similarly metabolized (*Enoch et al, 2007*).

This thesis focuses only on a restrict number of those (the ones which are the most observed in the area), in particular DEA (deethylatrazine) and DACT (di-amino-chloro triazine). The relative order observed in a study of concentrations in rural wells of USA was generally as follows: atrazine ~ DEA ~ DACT > DIA > hydroxyatrazine (*atrazine and its metabolites in drinking water, WHO 2011*). This is also another reason for concentrating only on the first 3 compounds.

Moreover, DACT seems to be the by-product of the multiple degradation of atrazine, TBR and simazine (figure 30).

6.3.2 Metolachlor

Metolachlor is an odourless liquid used as selective herbicide for corn, soy, sugar beet, sunflower and tobacco. Metolachlor has a moderate mobility on soils, is soluble and has a long half-life in the environment: for these reasons, leaching of this chemical is possible. This substance has been widely used throughout the Venetian plain and has been detected in surface water as well as groundwater. Metolachlor exposure results in a myriad of health effects that include eye and skin irritation, stomach cramps, shortness of breath, weakness, sweating, diarrhea, dizziness, and nausea. Exposure can also result in anemia, convulsions, and jaundice (DHSS, 2010). Metolachlor has been classified as a possible human carcinogen by the U.S Environmental Protection Agency (EPA). This classification was based on a study that detected increased liver tumors in female rats and a study that replicated these findings (EPA, 1995a).

Metolachlor was withdrawn in Europe in 2003 and was replaced by s – Metolachlor (which is still used): the difference between the two substances is only of the ratio between S – isomer (biologically active) and R – isomer. In Metolachlor the ratio S/R is 1:1 whereas in s – Metolachlor is 9:1. The differentiation of the two substances in laboratory is not easy and for this reason many labs cannot distinguish the two. In the database of this thesis the differentiation between Metolachlor and s – Metolachlor is possible only from 2019 onwards.

7 RESULTS AND DISCUSSION

In this chapter the various contaminants are analyzed with the tools described in previous pages; the knowledge of the properties of the different pollutants in the environment are also of fundamental importance for a better understanding of the phenomena.

The analysis is composed of the following steps:

1) Quantification of the available data for the analysis and general properties of the dataset.
2) Presentation of observations of previous papers on the subject to recognize the general pattern of contamination and the relationship with the activities in the area.
3) Implementation of geostatistical tools (variograms, kriging): this part is preceded by the application of basic statistic indicators in order to characterize the properties (distribution, presence of outliers etc.) of the input dataset for geostatistics.
4) Presentation of the concentration maps obtained.
5) Observations on the obtained results, considering all the previous steps.

The analyzed pollutants are solvents, nitrate and pesticides, which have been described in the previous chapter. For pesticides, only one specific contaminant (di-amino-chloro-triazine DACT), which is particularly interesting, is analyzed with geostatistical tools.

7.1 Case study: TOTAL SOLVENTS

The database available, which is the product of various datasets coming from different sources, is at first analyzed in quantitative way: is clear that in the first years the observations are scarce both in terms of available measures and in terms of spatial completeness (very few points). From 1985 onwards the number of available points increases and from 2000 there are always more than one hundred points per year. The number of points and the number of analysis present are pretty much equal in the first years (one measure per year available) and grows in the later years: although for a good portion of points more from 2000 onwards more than 1 measure/year is present, inspecting the time variability of solvents in the domain on short time scales (<1 year) is impossible due to lack of observations in a short time frame.

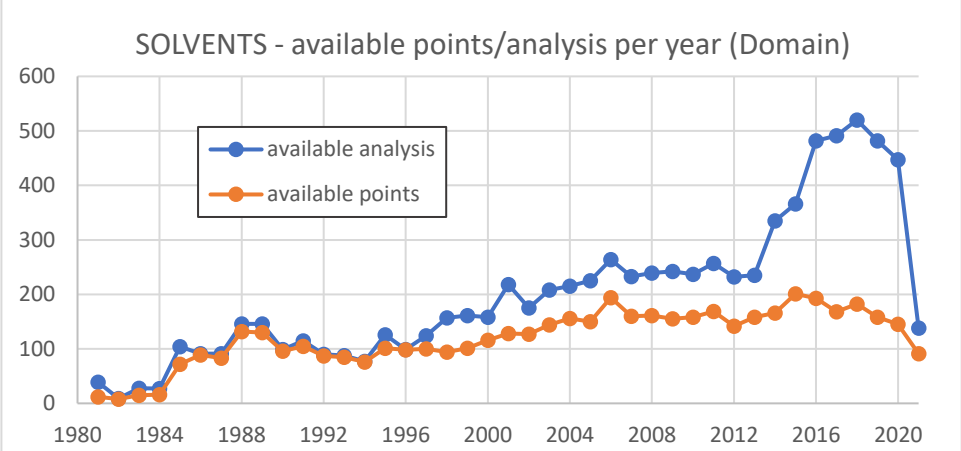


Fig. 31: available points and analysis per year (solvents)

Regarding the situation in the Agno valley, the number of points and measurements available is not enough to provide any significant base for geostatistical tools implementation except for years 2015 – 2019; the threshold of 30 points is considered the absolute minimum for a kriging interpolation by many authors. Furthermore, the solvents presence in the Agno valley area is very different from the one in the plain between Astico and Leogra – Timonchio rivers. This fact is clear by observing figure 31 in which the overall means of the areas are plotted.

Variable	Total Count	Mean	SE Mean	StDev	CoefVar	Q1	Median	Q3	Range	IQR
SOLV(microg/L)	7275	6.471	0.155	13.258	204.89	0.000	1.400	7.200	382.000	7.200

Mode	N for Mode	Skewness	Kurtosis
0	2103	7.25	120.73

The main statistical indicators of the whole dataset are reported in the table; is clear that the data is skewed and not normally distributed. Even if the mode is equal to zero, the great difference between

the median (1.4 microg/L) and the mean (6.4 microg/L) indicates the presence of a great number of outliers (clear signs of contamination episodes). The overall behavior of the measurements can also be visualized in the boxplot and the histogram.

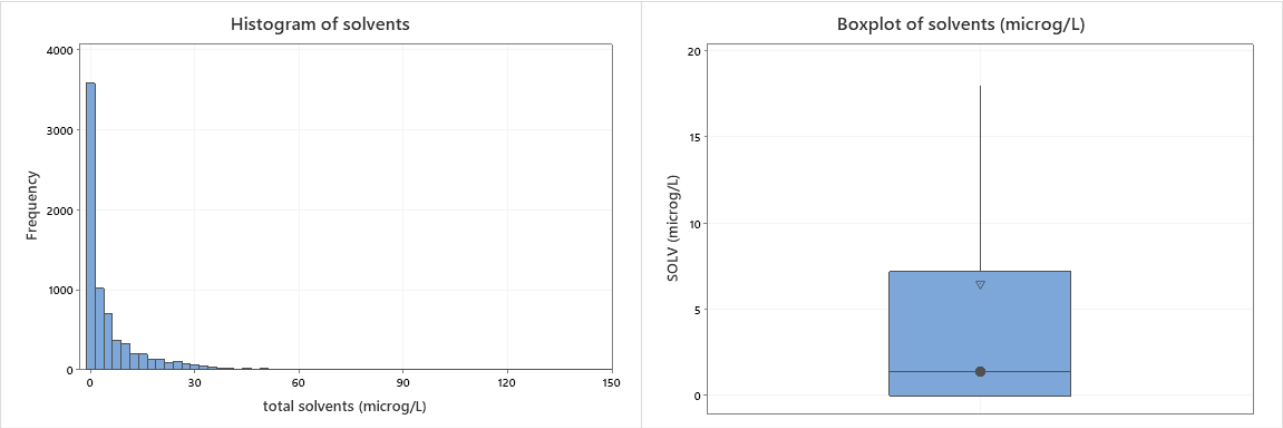


Fig. 32: histogram and boxplot of total solvents (complete dataset)

The situation of the contamination in the principal domain presents a clear decreasing trend although the peak on the initial decade is characterized by lower number of values. Although less significant by means of number of observations, the peak on the years 1981 onwards is due to a known situation in the domain: in the 70's a solvents contamination was observed for the first time in the Schio – Torrebelvicino as well as in the Thiene – Marano - Zanè area, and the migration of the contaminant has gradually interested a larger zone. The overall annual mean is below the actual law limit of 10 microg/L from 1994 to present days: this underlines the high presence of solvents in the area caused by anthropic releases in the past.

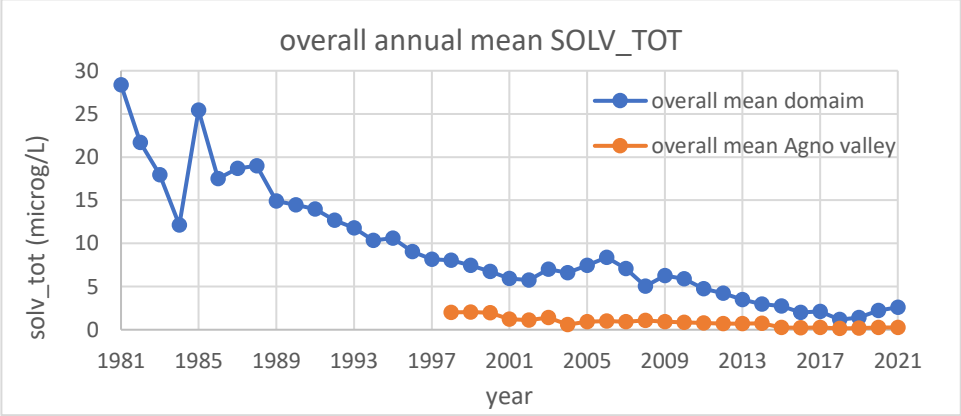


Fig. 33: overall mean of total solvents (domain and Agno valley)

Various authors studied this contamination: data on the major users of solvents has also been collected to better understand the situation.

In a study of *Altissimo et al. 1990* regarding the pollution load on the area of interest, it was found that the most important points of use (quantities of solvents greater than 10 tons per year) are:

- At the mouth of the waterways in the high plain, as is the case with the Leogra nearby Schio
- Along or near the main disperding stretches of the Astico river, linked also to ancient riverdbeds, now presenting high underground flows.

The origin of the pollution of the area of Schio and surroundings cannot be recognized to a singular point with the available data; this is clear by observing the plot below in which various points, which are one downstream the other in the area between Torrebelvicino and Schio, don't present a "waterfall" effect suggesting the fact that the whole area is variably contaminated. Moreover, the point which is the most upstream (VITB-1) is the one with the lower levels of solvents, even if it exceeds the present law limit value in 1991 and 1995 which indicates that the background level is affected by anthropic activities even in the upper parts of the domain close to the mountains. Values above 40 microg/L were observed in the 80's; this observation can be explained by the fact that the plume of contamination, of which the first measures are of the 70's, and of which the temporal origin can't be traced back due to lack of data, already reached in the 80's the lower plain. Another possibility is that the use and consequential mishandling of solvents was in place also in areas of middle-lower plain.

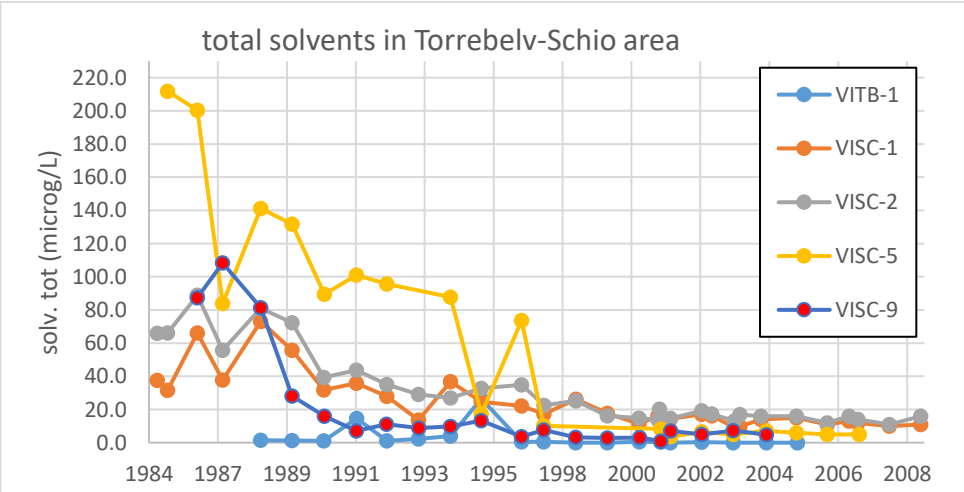


Fig. 34: total solvents in Torrebelvicino – Schio area

A series of initiatives by ULSS 6 and others (like forming industries for a correct stocking of solvents, ban of the products for certain activities and so on) effectively reduced by 50% the overall load of the potential pollutant with respect to 1986. The concentrations of solvents in the area are overall very high in the first years of observations even if the consumption in the area in late 90's was not as big as it was previously. This fact can be explained by the retention time of the soil and the probable

continuous release of contaminant of the porous media as a function of its water content, linked to piezometric oscillations. Due to this phenomenon, as correctly predicted by *Altissimo et al 1990,* the concentration in water of solvents in this area has been stable above law limit for a long time even if the production and use in the Schio area has been greatly reduced. Concluding, the measured values in the area of Schio – Torrebelvicino are very probably related to continuous release of solvents in the area by various industries and the effect of this contamination has existed for decades underlying the great importance of prevention and sensibilization on the topic. Another observation is that all the points have an overall decreasing trend confirming the gradual, even if slow, depletion of the past contamination.

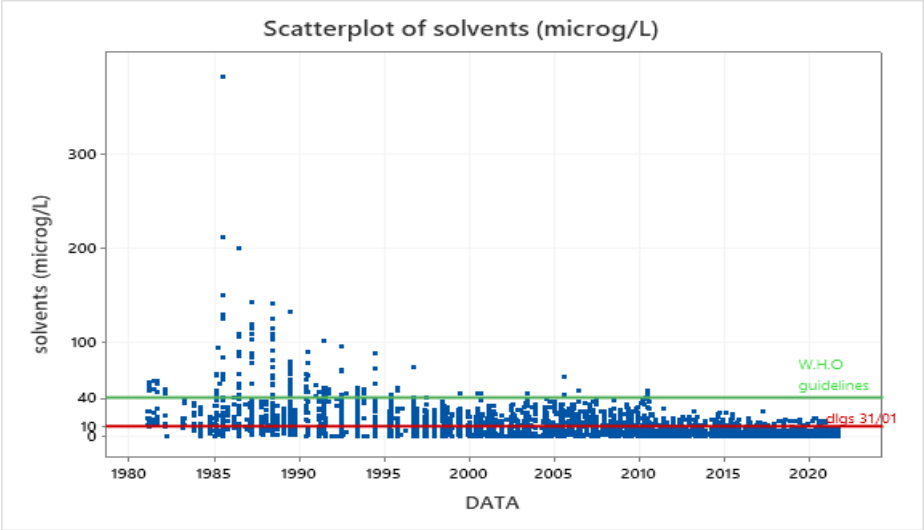


Fig. 35: scatterplot of total solvents

Solvent's pollution is also present in the area on the left side of Astico, interesting various municipalities like Fara Vicentino, Breganze, Sandrigo, Bressanvido. The origin of the contamination trace back to the early 80's once again.

One episode of pollution was observed in Fara Vicentino, and a study of Pagello and Perdon (1984) reconstructed the contamination evolution; the main conclusions are reported below. The pollution origin is punctual and instantaneous; the rupture of a tank containing solvents occurred in a period of dry weather with low phreatic levels, resulting in the fact that not all the contamination reached the water table, leaving a part of the original quantity of solvents to consolidate on the sediments. In this portion then a period of heavy precipitations had a wash out effect. The result of this dynamic is the exhaustion of the contamination in 70 days, in which after the precipitation the solvents concentration drastically decreased. After this period the polluted area was observed to change with the variations of the flow direction. In this case solvents quickly reached the aquifer, whereas in other episodes a big quantity of solvents remained in the unsaturated zone for years and it is not mobilized by

precipitation but only by the rise of the phreatic surface, resulting in a much longer period of contamination. This pollution episode contributed to an overall contamination of the area in which many other industries using solvents are present. It is possible and very probable that the contamination of the area is the sum of various origins.

Another known serious episode of contamination was observed in 2009 (fig 36): in this case the origin is known to be from a factory in Marostica. This contamination spread in the aquifer area of Breganze-Bressanvido-Ancignano (*Altissimo, Passadore 2010*).

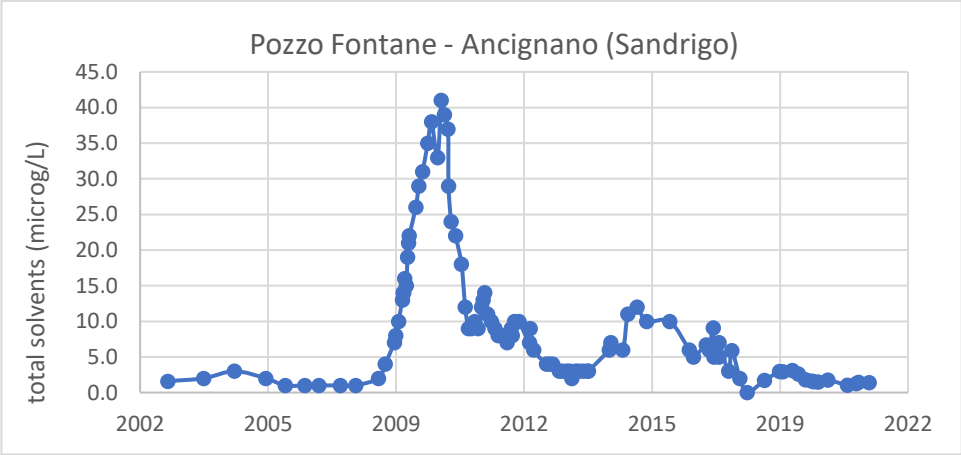


Fig. 36: Pozzo Ancignano total solvents time series

The values recorded in the area are well above the law limit for total solvents (doubling the limit and more) suggesting a very serious case: fortunately, the contamination origin has been readily removed. In later years the situation is naturally becoming less concerning even if the dispersion of solvents downstream could be problematic: this phenomenon cannot be studied with the data available in this script because the plume is likely to migrate S-E outside the domain of interest. This episode is clearly not correlated with the one in the Schio- Thiene area.

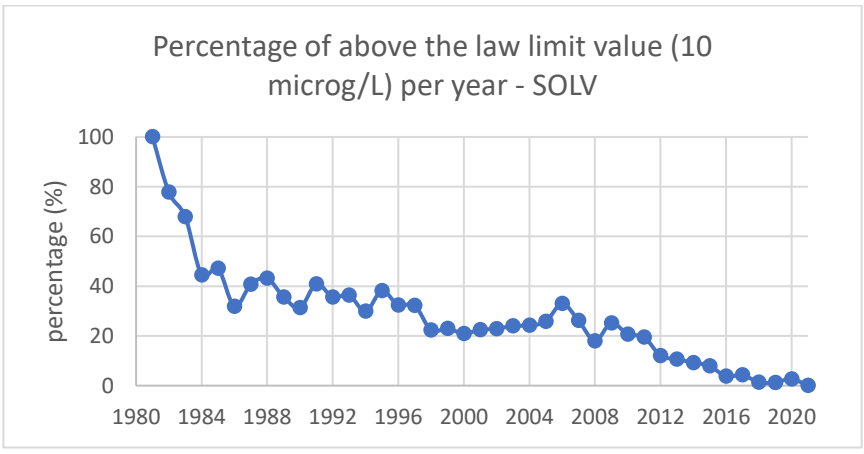


Fig. 37: percentage of above law limit values (10 microg/L) per year (total solvents)

For the reasons above, the solvents are the most frequent reported cases of values above the law limit in the domain; by retroactively fix the present law limit (dlgs 31/2001) in the whole period the

percentage of values above law limit (number of measures > 10 microg/L on number of total measures in a given year) figure 37 has been obtained. The overall situation for solvents contamination is quite bad, although in recent years seems to turn in a better scenario (in the last years of observations the percentage of above the limit values is very close to zero); this fact can also suggest the gradual depletion of past issues or the migration of the plume of concentration outside the domain. Another reassuring fact is that in 2021 the maximum measured value is equal to 8.6 microg/L, standing below the law limit.

Regarding the situation in the multi aquifer system, even if the knowledge of stratigraphy for the majority of the wells is lacking, some observations can be done in the limited dataset available:

- The concentrations of solvents in the 3° and 4° aquifers are very correlated (r = 0.9) and follow a decreasing trend.
- The multifilters which include the 4° aquifer are very correlated with the 3° and 4° aquifer.
- In the 1°, 2°, 5° and 6° aquifers the recorded values are almost always equal to zero.

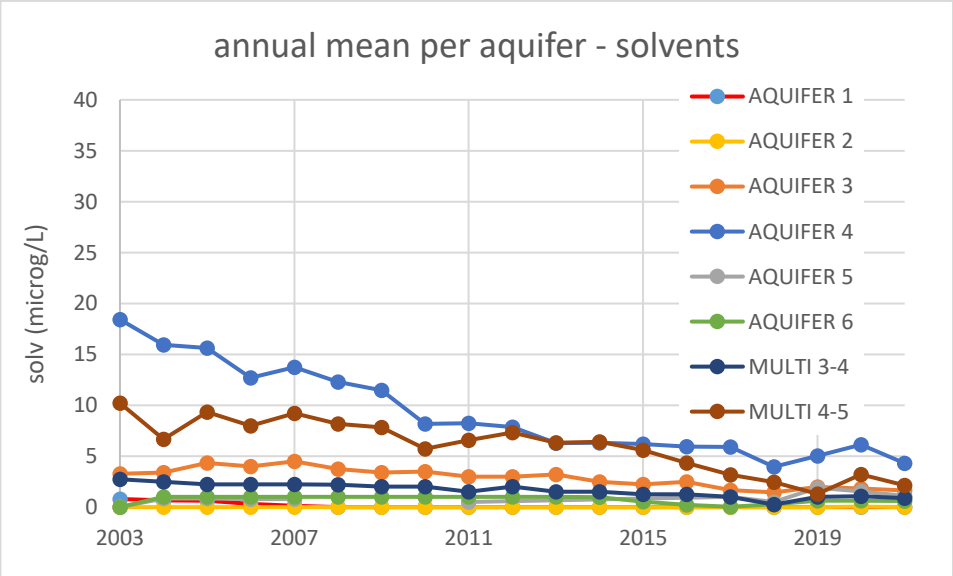


Fig. 38: annual solvents mean for different aquifers

As a consequence of the observations reported above, is clear that values of solvents coming from different aquifers are comparable only if considering the 3° and 4°. For this reason, the analysis for solvents is limited to only the third and fourth aquifer; values measured in different aquifers are eliminated from the dataset. Unfortunately, only a fraction of the wells in the artesian zone of the domain have known stratigraphy and consequently the screening operation can only be achieved "visually". Values in the artesian zone have been eliminated if observed to be very different from the

ones measured in the wells of 3° and 4° aquifer. For example, if the solvent concentration in the well X (known to be in the 4° aquifer) is equal to 20 microg/L and a value at close distance Y (few km from an unknown stratigraphy) is equal to 0 microg/L, value Y is removed from the analysis under the hypothesis that value Y is measured in an aquifer different from 3° or 4° aquifer. Overall more or less 5 wells per year have been eliminated which is a very small fraction (5 - 10 % circa) of the observed wells in the artesian zone: this fact is not a surprise given the fact that the 3° and 4° aquifer are the most productive layers.

This screening phase, although lacking rigorous method, can be considered conservative: the maximum values recorded in the artesian zone are all from 3° and 4° aquifer and for this reason the obtained maps are referred to the worst possible situation in the multi aquifer system. The cause of the enhanced pollution in these two layers can be identified by the lowered piezometric level as a consequence of heavy exploitation, as well as the high hydraulic permeability; these conditions offer a “recall” effect for pollutants in the 3° and 4° aquifers (*Altissimo et al. 1990*).

7.1.1 Geostatistical analysis

As already mentioned, the number of points in the dataset is not sufficient for geostatistical interpolation in years prior to 1988 and the overall mean exhibits a clear decreasing trend. For this reason, the first analyzed year is 1988 and the investigation goes on from there by studying the annual mean of wells: operating on multi-year averages could be misleading due to the presence of the trend. Years are analyzed with a step of more or less 3 years, a right compromise between time consumption and completeness of examination. A table containing the properties of the studied years is given below.

Tab. 8: main properties of the analyzed periods (total solvents)

YEAR	N° OF POINTS	MEAN [microg/L]	ST. DEV. [microg/L]	SKEWNESS	$DIST_{MEAN}$ [m]
1988 (upper plain)	89	15.44	24.51	2.5	743
1991 (upper plain)	88	12.66	17.75	1.9	773
1993 (upper plain)	72	11.62	15.66	1.33	837
1995 (upper plain)	63	10.14	13.55	1.3	877
1999 (upper plain)	64	5.03	9.02	2.4	959
2003 (right Astico)	123	7.43	10.42	1.5	654
2006 (right Astico)	175	9.36	10.67	1.1	535
2009 (right Astico)	141	5.66	8.30	1.7	616
2012 (right Astico)	130	3.76	5.36	1.7	592
2015 (right Astico)	179	3.14	4.99	2.2	468
2018 (right Astico)	159	1.24	2.48	3.0	528
2021	91	1.99	2.62	1.24	553

Data show great variability with standard deviations reaching high values; as a consequence of that the coefficient of variation CV is higher than 0.3 for the majority (80%) of wells. This fact can be explained by the presence of the decreasing trend which can increase the variability of time series; as already mentioned to avoid consequences of this problem, the time scale of the singular geostatistical analysis is the smallest possible with the available data (annual).

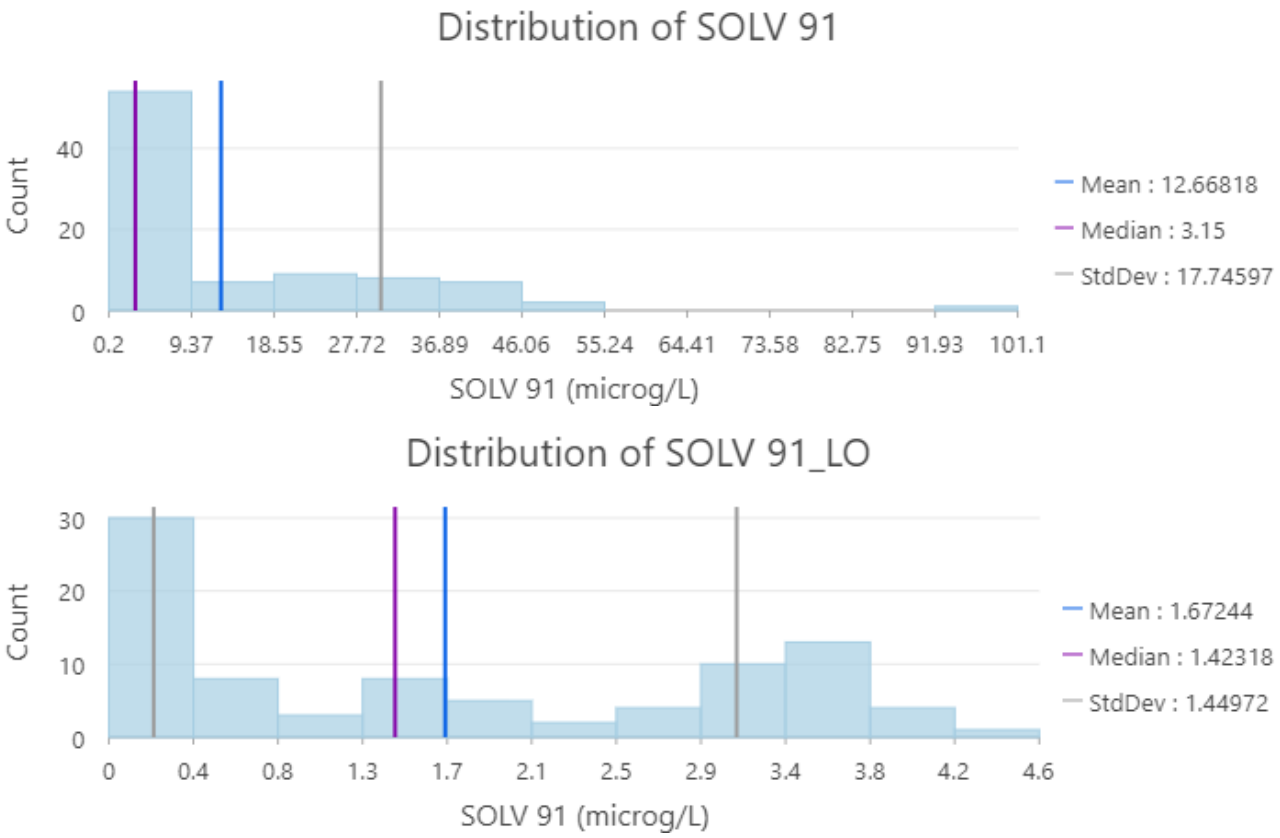


Fig. 39: histograms of raw data Vs logtransformed data (SOLVENTS 1991)

Another observation is that the solvents concentration distribution is highly skewed, with the majority of values close to zero and a high number of outliers, reflecting the contamination phenomena. **For this reason, a logarithmic transformation of the dataset, for all analyzed years, is operated:** although the obtained distribution is once again far from normal, at least the skewness coefficient on the transformed data is lower than 1. The log transformation also has the effect of "shrinking" extreme values of which the dataset is rich (see boxplots). As explained in chapter 5.4 in presence of distributions which are far from normal even after logarithmic transformation, kriging estimation may

be not the most appropriate tool to treat such data. Nevertheless, models outside of kriging go beyond the scope of this thesis and are therefore not considered.

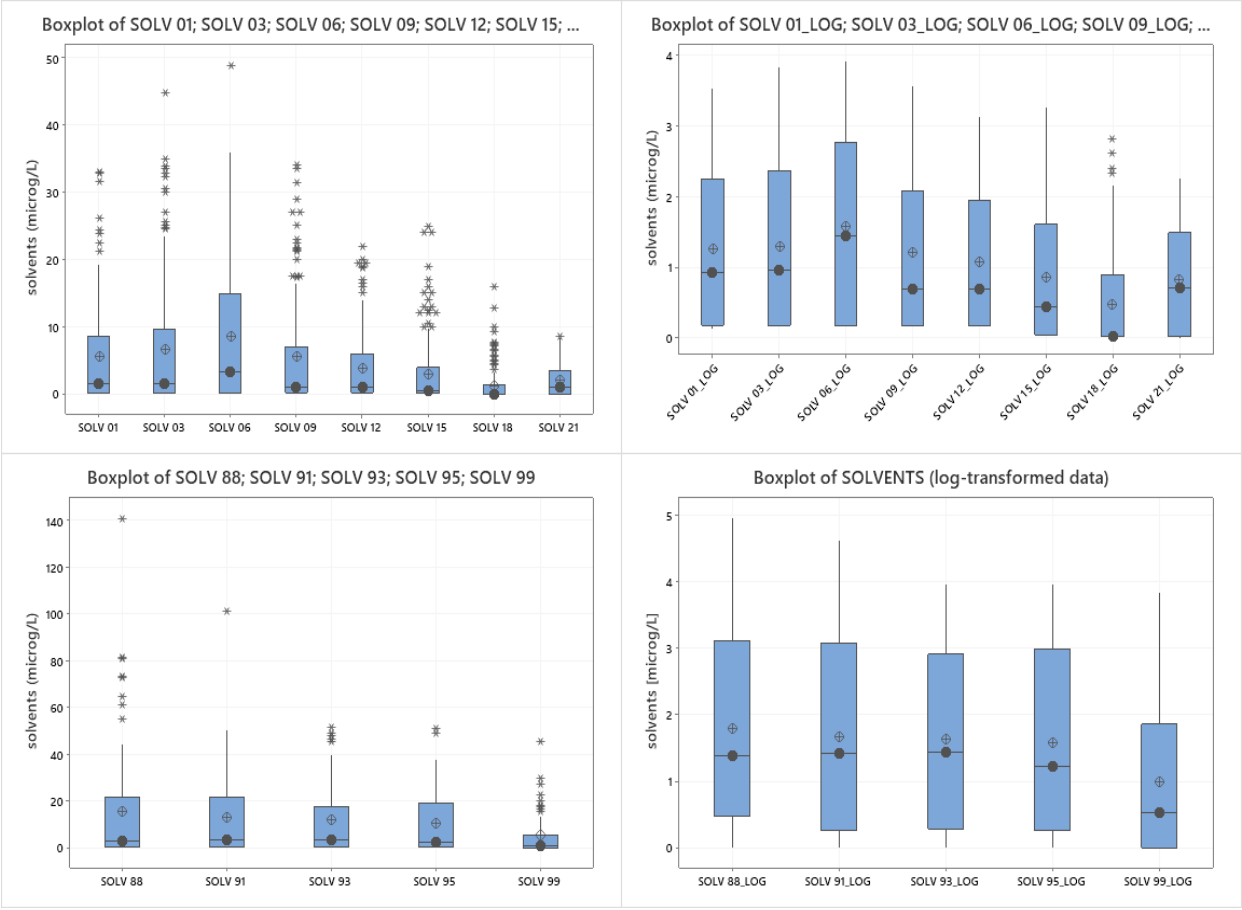


Fig. 40: boxplots (median black dot, mean void dot) of annual mean of solvents 1988-2021. Raw data (left) vs lograsformed data (right)

Regarding the spatial distribution of the points, is observed that in years prior to the 2000's the spacing between the points do not allow a meaningful interpolation in the whole domain: for this reason, in the first part of the analysis only the upper plain portion of the domain is studied, in which the density of points is compatible with kriging interpolation. Focusing only in this restricted area is also a useful operation, since the contamination seems to have its origin in the upper plain; eventually the solvents migrated in the artesian zone following the recharge of aquifer mechanism. As previously mentioned, there are at least two events of solvents contamination in the Breganze – Bressanvido area: unfortunately the number of points in the left side of Astico river is very limited and therefore is not possible to obtain a meaningful interpolation to reconstruct the contamination. For this reason points on the left Astico are not considered in the geostatistical analysis of solvents; a focus on this area only likely requires a dedicated study.

The distribution of points in the area is not ideal and present some clusters. A clear important increasing or decreasing overall spatial trend (different from anisotropy) is not observed, although a minor trend can be observed in the N-S direction; this trend is not considered important for the analysis and therefore data is assumed to be stationary for ordinary kriging interpolation. This fact can also be observed by looking at the scatterplots below: is clear that important trends are not present.

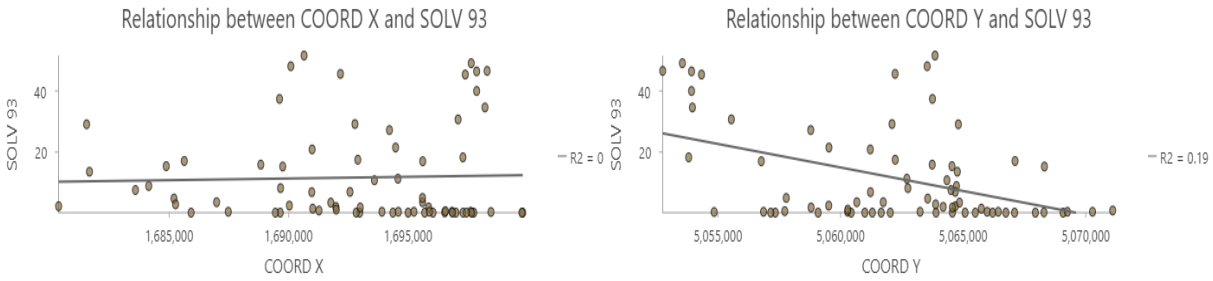


Fig. 41: example of scatterplots of solvents (microg/L) Vs coords (1993

The average distance between points ranges from 600 m to 1,5 km: as a consequence of that a lag distance of 1 km is adopted for the variogram inspection. The maximum distance is set at roughly half of the observed maximum distance between points: this corresponds to 15 km for the 90's and to 20 km for the later years. The obtained semivariogram for the first observed year stabilizes roughly at the sample variance and show a typical spherical behavior.

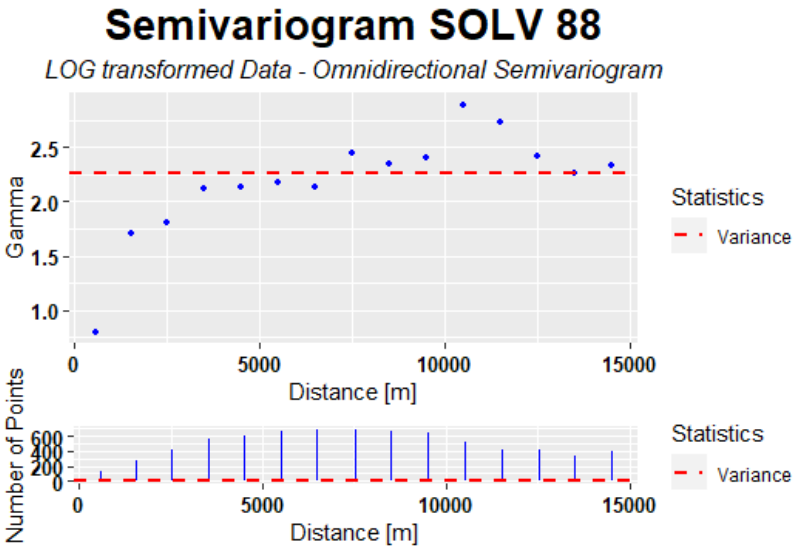


Fig. 42: omnidirectional variogram of the first analyzed year (solvents 88)

Thanks to past publications, is known that the historical contamination of solvents formed a plume which travelled in the subsoil; for this reason, the directional behavior of the problem has been analyzed. At first glance, a preferential path, roughly in N-SE direction is observed. This anisotropy determines a greater range (distance of correlation) at 135° direction (N-SE) circa. This property is

observed for almost all the analyzed years: this plume of concentration, migrating from the upper plain to the Vicenza area, was also recognized by many authors in different years (*Altissimo et al 1990, Passadore at al 2009*).

The following image is an example to recognize the anisotropy in the analysis.

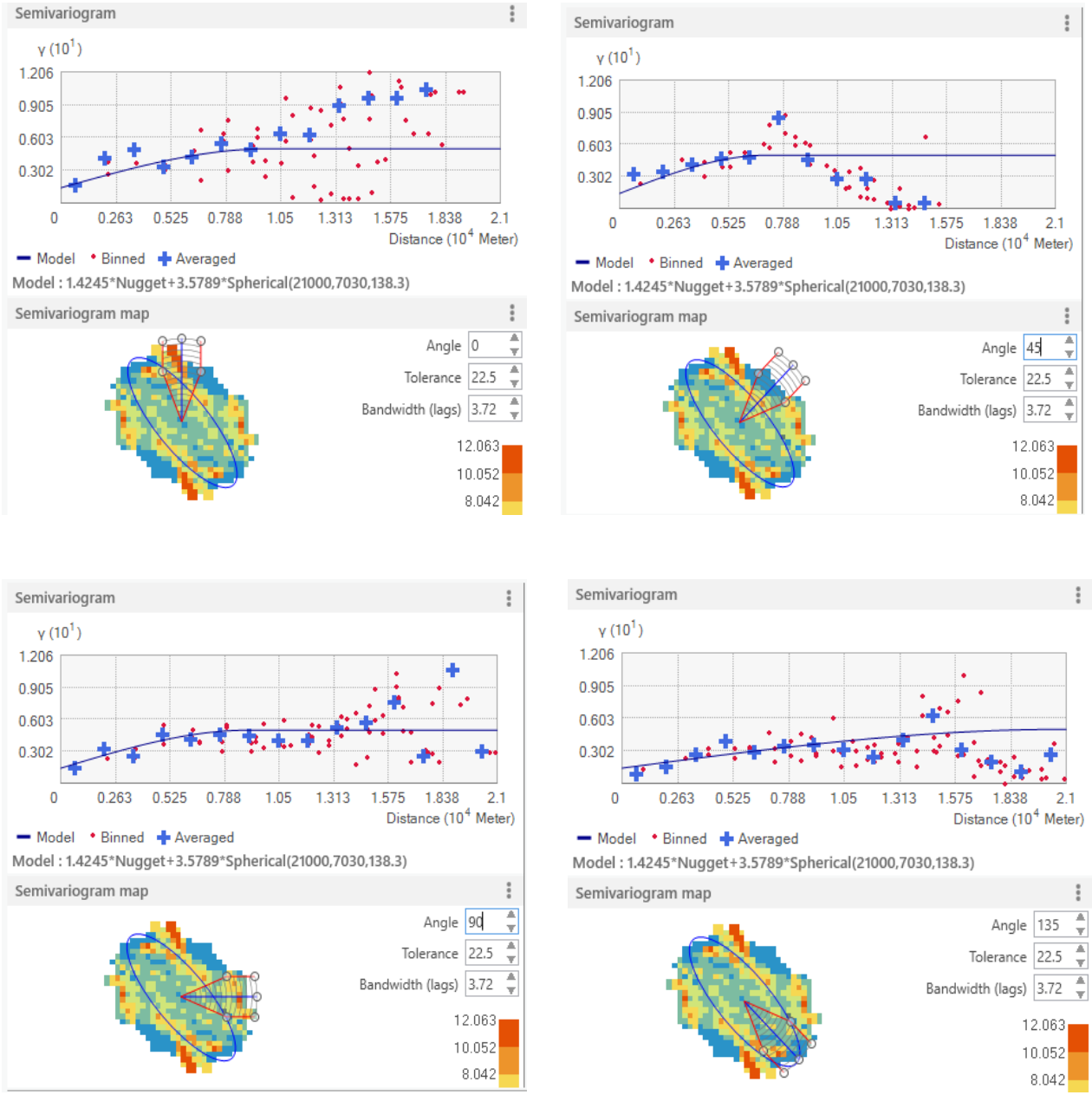


Fig. 43: semivariograms (solvents 1993) in different directions: from 0° (top left) to 135° (top right)

Is clear that in the 0° direction semivariance goes to higher values at very short distance; this is also true for 90° direction although the range is higher. An interesting behavior is observed at 45° direction where data seems to have no correlation from the start but at higher distances some very low values of semivariances are observed: this peculiar phenomenon can be easily explained by the shear fact

that there is not a high number of points in the lags composing the last values of the semivariogram, therefore the "weight" of the points is not very indicative. Finally, in 135° (S-E) direction solvents are correlated basically for the entire length of analysis, suggesting a clear preferential path, probably due to the joint effects of the presence of ancient riverbeds and the main flux direction imposed by phreatimetry in the upper plain portion of the domain. The presented structure of directional semivariograms (fig. 43), referred to 1993, is replicated with very similar properties for all the analyzed years before the 2000's.

For the later years in the analysis the directional semivariograms present minor differences, stabilizing at different ranges and with lower values of sills: this behavior is due to the fact that more points in the analysis are present and for the gradual depletion of the contamination.

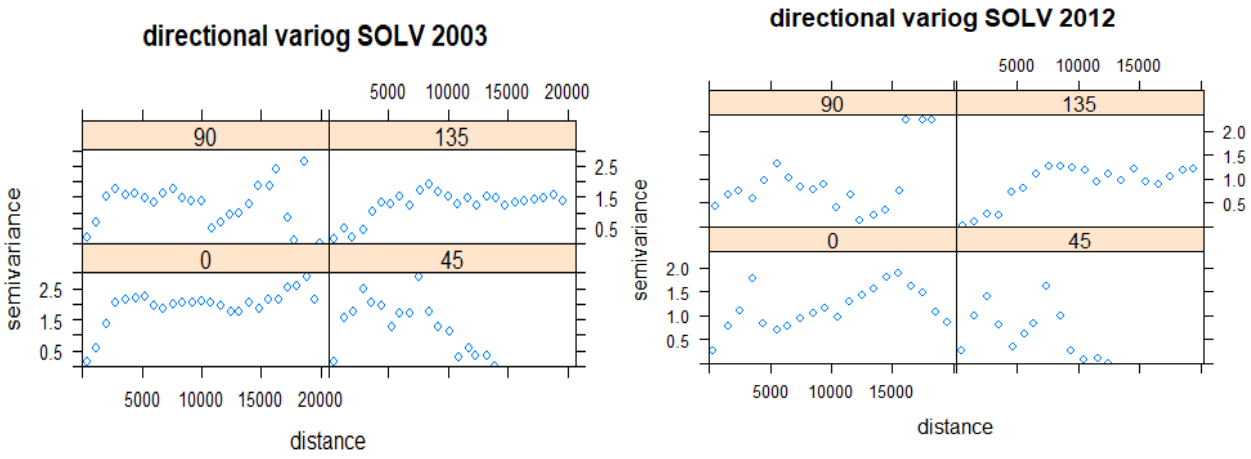


Fig. 44: directional semivariograms (2003, 2012)

The values of semivariance at which the data stabilize (the sill) is roughly the same in all direction except the 45° one where there are less points; for this reason, the anisotropy can be considered a geometrical type one. It is easy to observe that the spherical model can fit properly the variograms; for this reason is the chosen theoretical model. Other models like gaussian and exponential have been tested but the spherical model obtained the best results in terms of errors; the obtained maps with exponential and spherical models are very similar, whereas the gaussian model can produce some minor differences. Another observation regards the values of nugget, sill and ranges; unfortunately the nugget in the first years is different from zero. This behavior can be explained by a preferential sampling campaign which can produce high correlations at short distances or by an overall poor spacing between points. In years with small gaps between sill and nugget the errors are higher. Values of ranges reflect the observed anisotropy of the contamination, with a major range which is

significantly higher than the minor one. The properties of the various models are organized in table 9.

Omnidirectional kriging, where the phenomena is considered isotropic, has also been tested to make comparisons: similar (or even better) performances are obtained for years in which the mean of solvents is low, and the number of points is high (2015-2021).

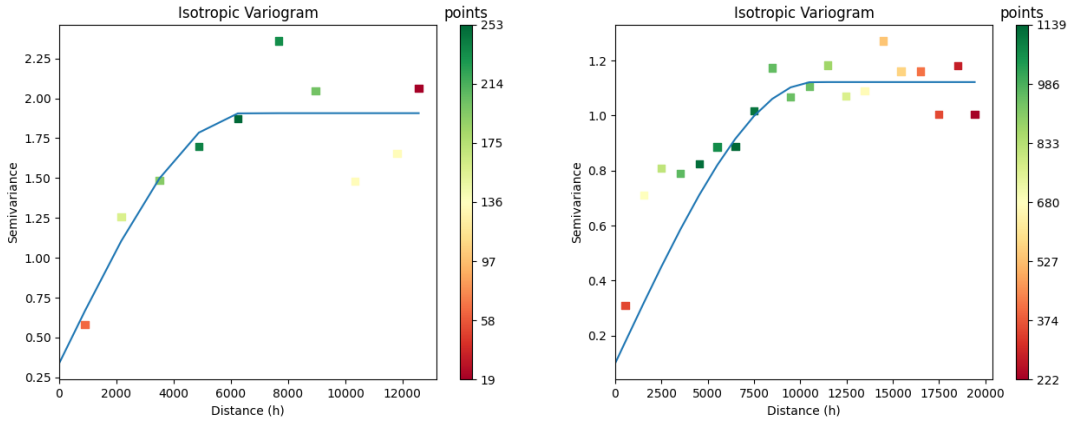


Fig. 45: isotropic omnidirectional variograms fitted by spherical model (logtransformed Solvents 1995 on the left, logtransformed solvents 2015 on the right)

To better understand the kriging maps, a table containing the details of the models is given below.

Tab. 9: ordinary kriging (anisotropic) parameters and LOOCV RMSE (total solvents)

Year	Model	N° of points	LAG [m]	Max dist [km]	Nugget	Sill	Major range [m]	Minor range [m]	angle	LOOCV RMSE [microg/L]
1988 (up.plain)	Sph	89	1000	15	0.62	2.46	15k	5018	147°	23.4
1991 (up.plain)	Sph	88	1000	15	0.52	2.40	15k	5026	147°	17.0
1993 (up.plain)	Sph	72	1000	15	0.13	2.08	9054	3027	123°	11.5
1995 (up.plain)	Sph	63	1000	15	0.23	2.18	14854	4999	149°	12.41
1999 (up.plain)	Sph	64	1000	15	0	1.50	6876	2303	138°	6.57

2003	Sph	123	1000	20	0	1.48	11840	3949	142°	5.75
2006	Sph	175	1000	20	0.03	1.78	13605	4549	127°	7.13
2009	Sph	141	1000	20	0.02	1.32	10370	3470	137°	4.41
2012	Sph	130	1000	20	0.02	1.00	8407	2802	137°	3.31
2015	Sph	179	1000	20	0.07	1.05	7292	2441	136°	3.29
2018	Sph	159	1000	20	0.02	0.48	9706	3253	130°	3.31
2021	Sph	91	1000	20	0.05	0.55	11648	3884	129°	1.01

Tab 10: ordinary kriging (isotropic) parameters and LOOCV RMSE (total solvents)

Year	Model	N° of points	LAG [m]	Max dist [km]	Nugget	Sill	Range [m]	LOOCV rmse
1988 (up.plain)	Sph	89	1000	15	0.60	2.12	6272	23.4
1991 (up.plain)	Sph	88	1000	15	0.51	2.22	6082	18.3
1993 (up.plain)	Sph	72	1000	15	0.52	2.07	5605	15.20
1995 (up.plain)	Sph	63	1000	15	0.20	2.04	6500	11.68
1999 (up.plain)	Sph	64	1000	15	0	1.51	4032	6.57
2003	Sph	123	1000	20	0	1.43	3758	6.93
2006	Sph	175	1000	20	0.03	1.75	6651	8.01
2009	Sph	141	1000	20	0.03	1.32	4888	4.62
2012	Sph	130	1000	20	0.07	1.00	3605	3.81
2015	Sph	179	1000	20	0.40	1.12	10845	3.29
2018	Sph	159	1000	20	0.03	0.45	3960	2.51
2021	Sph	91	1000	20	0.05	0.52	4821	1.06

The errors in the first years are very high and progressively go down in the later years. Presence of a large number of outliers, the highly skewed distributions and the overall non-normality, even after

the logarithmic transformation of the input dataset, are the cause of this. Accounting for anisotropy has the beneficial effect of reducing the RMSE, and maps obtained with this method are superior and more realistic. The difference between the two methods is lower in the latest periods of analysis: this fact may be caused by the depletion of the existing plume. The strong anisotropy is confirmed by the ratio of the ranges, which is consistently more or less equal to 3; even if data from the lower plain is not considered in the 90's for geostatistical interpolation, is safe to assume that the major range could also be higher in those years.

Values of ranges and sills, which systematically decrease along the years, suggests that contamination intensity has declined: lower values of sills and ranges basically mean that solvents concentrations are spatially correlated for less distances, determining a reduction of the contaminated zone, and with lower concentrations (lower sill). One exception to this behavior is represented by the year 2015, which presents a high range and nugget.

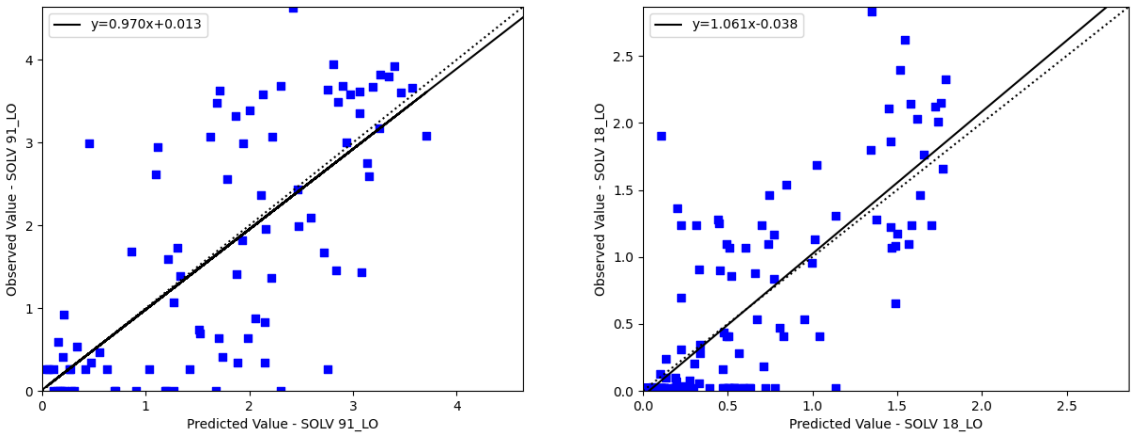


Fig. 46: LOOCV of isotropic kriging (solvents 1991 in the left, solvents 2018 right)

Hereafter are reported the obtained concentration maps: the dark red is set at the dlgs 31/2001 maximum acceptable concentration of 10 microg/L to quickly recognize the most polluted areas. In order to draw comparisons, for every analyzed year are reported the maps obtained considering anisotropy and without. The small black dots in the maps are the observation points.

Standard deviation maps are also present: since they are similar each other, only a few examples, which are representative of the periods, are reported. As described in chapter 5.4 the issue of back transformation of lognormal prediction variances is complicated and hereafter are reported the standard deviation in logarithmic scale, without any back-transformation. For this reason, is not possible to precisely quantify the prediction variances and the reported maps should be considered as a first approximation; a more rigorous and in-depth study on the issue of back-transformation of lognormal kriging variances is needed for a precise analysis.

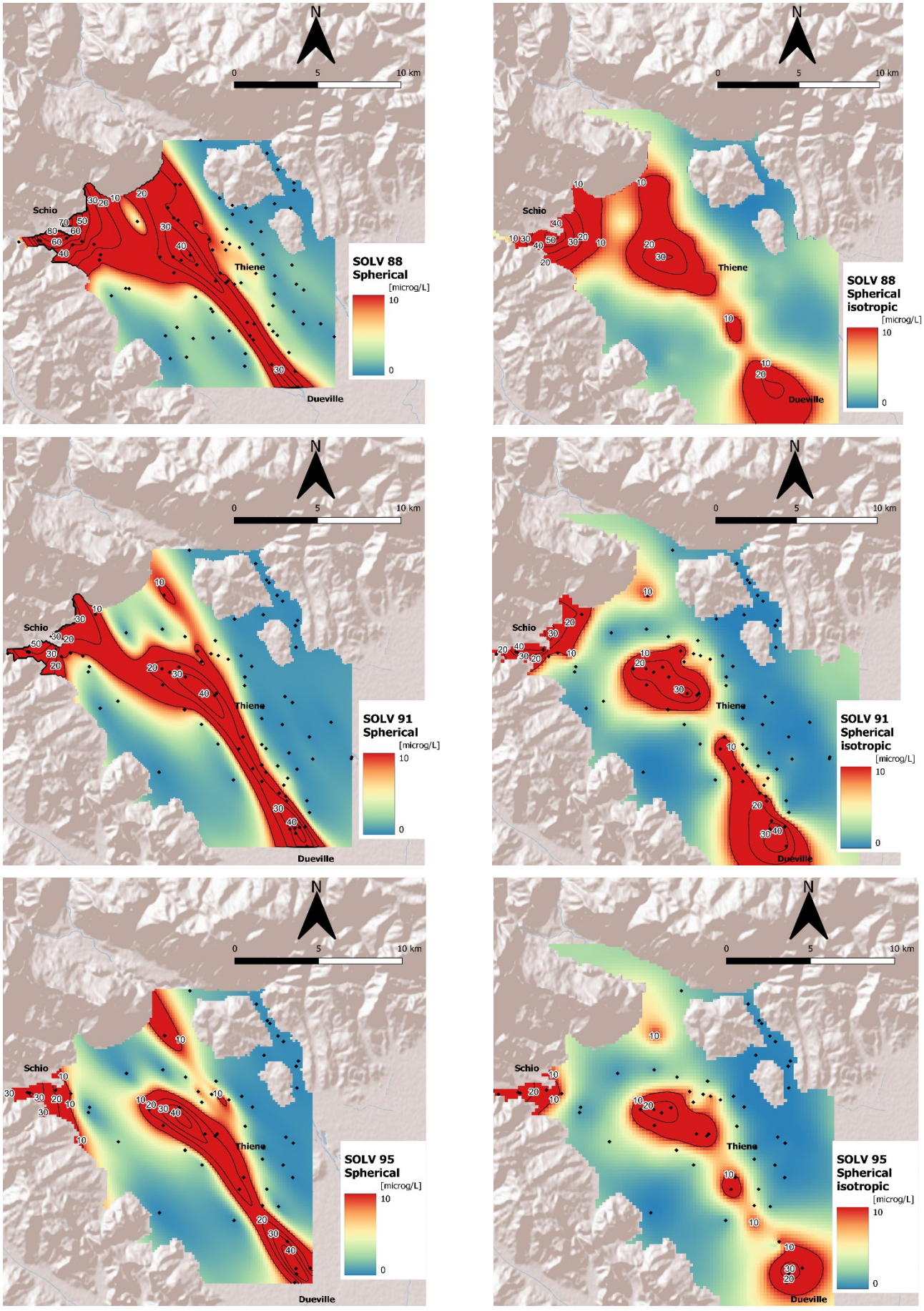
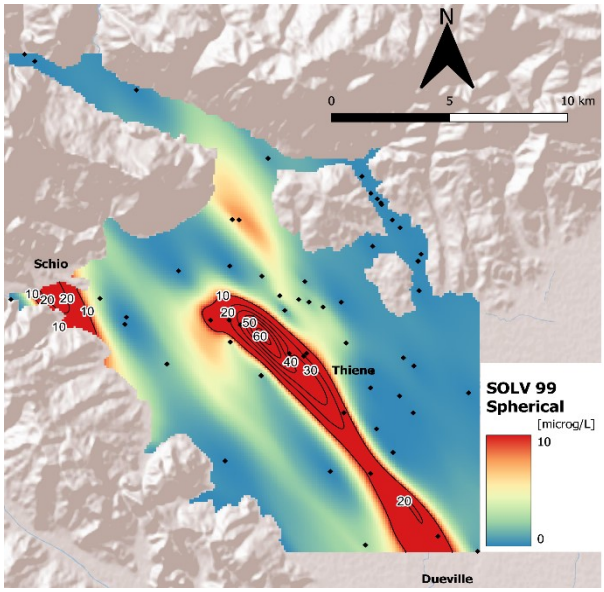


Fig. 47:Upper plain solvents concentrations 1988-1995: lognormal kriging with anisotropy (left) and isotropic (right)



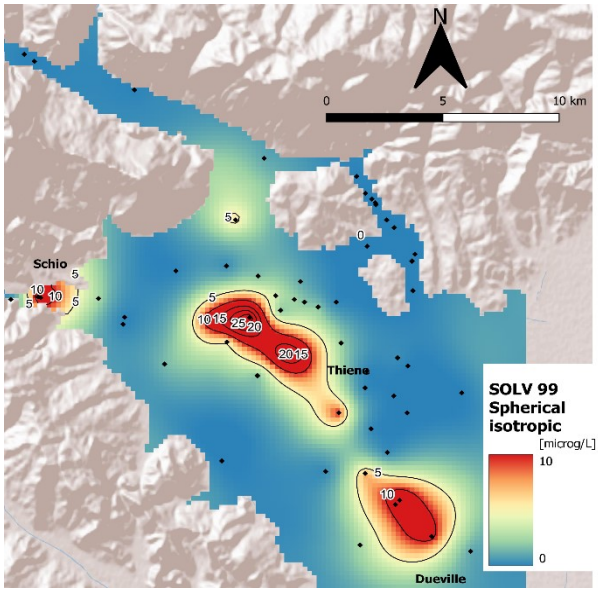
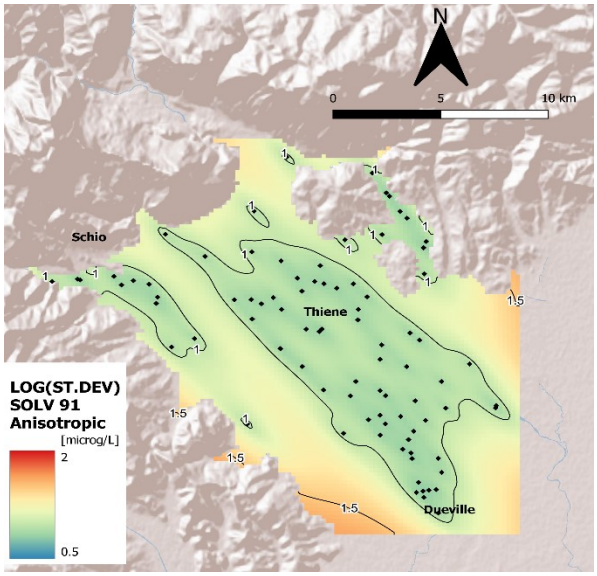
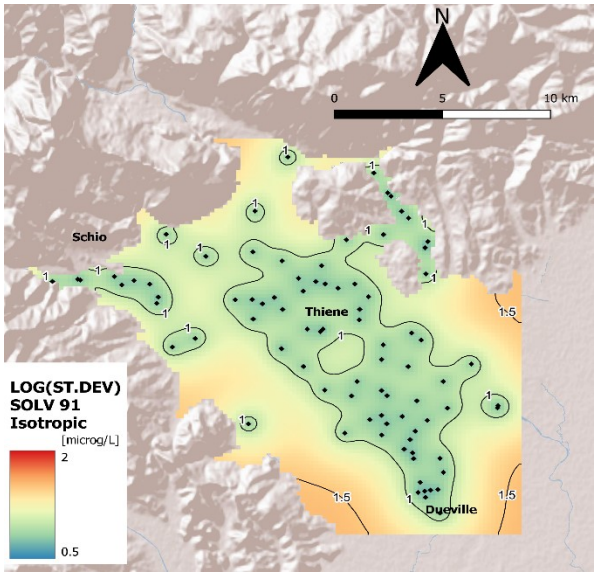
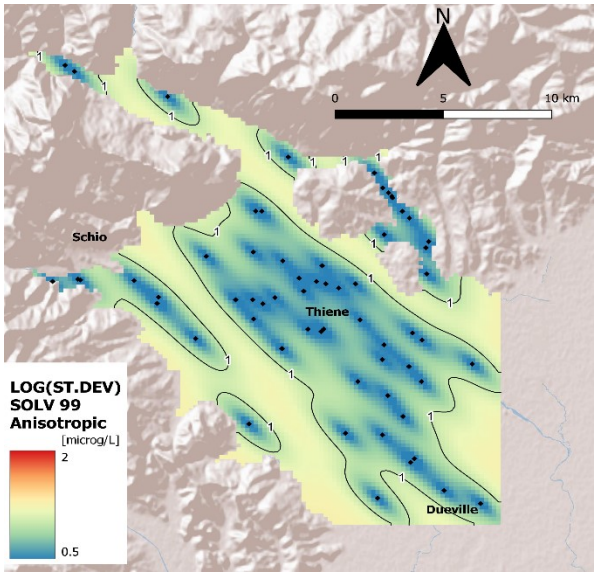


Fig. 48: upper plain solvents concentration 1999: lognormal kriging with anisotropy (left) and isotropic (right)







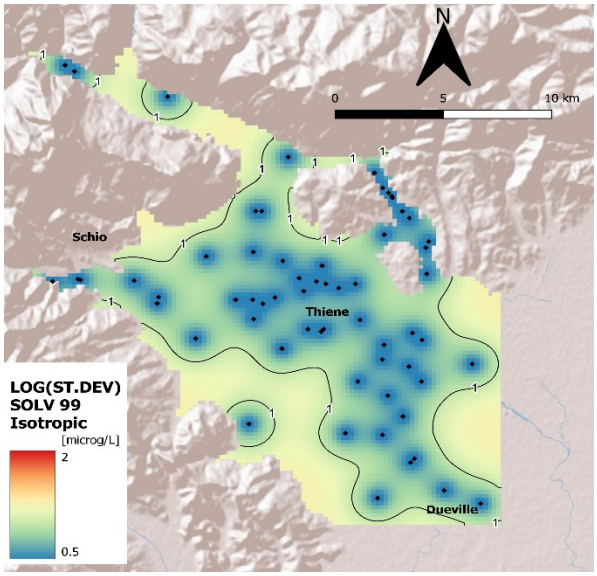


Fig. 49: upper plain kriging prediction standard deviations 1991-1999

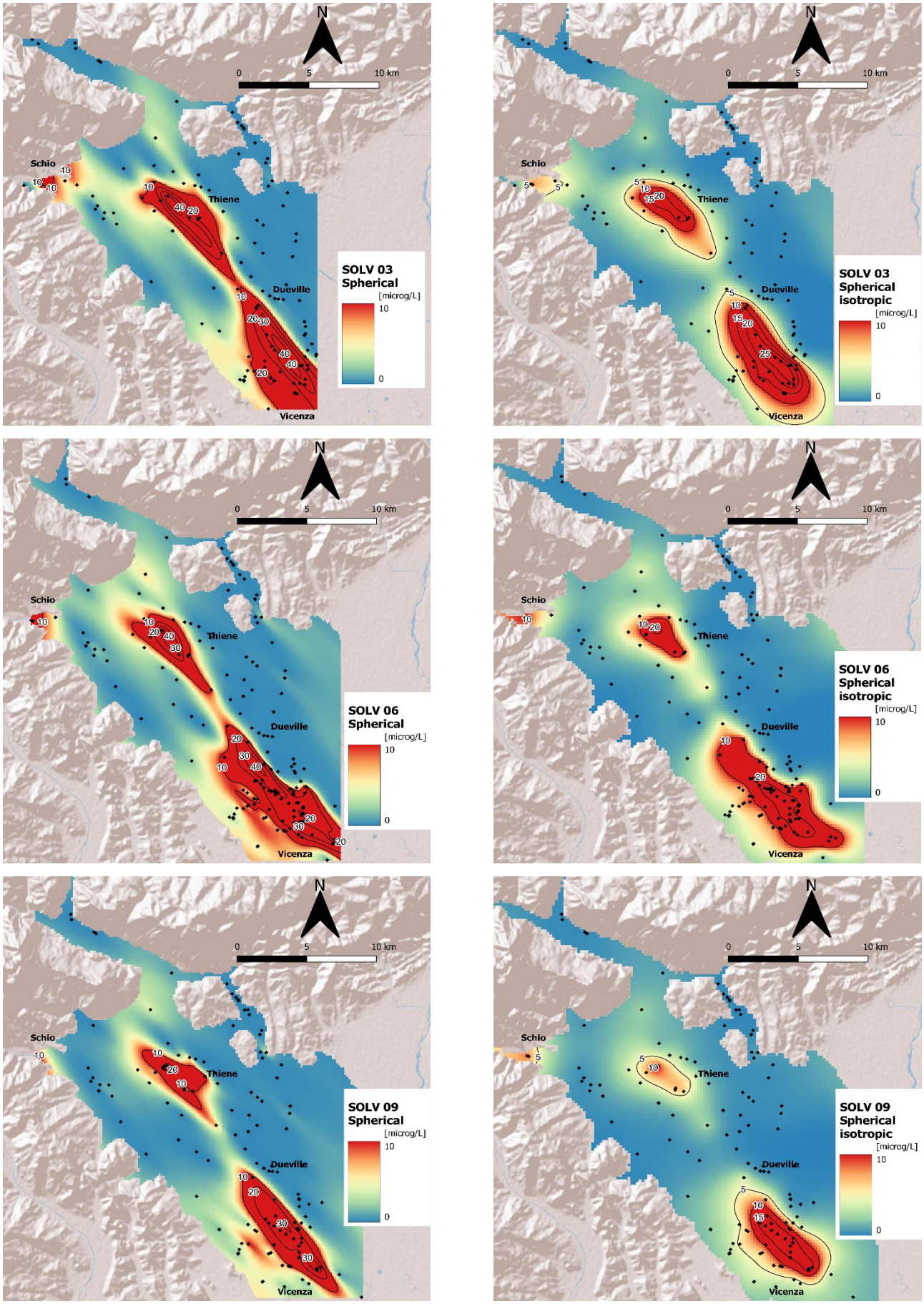


Fig. 50: solvents concentrations 2003-2009: lognormal kriging with anisotropy (left) and isotropic (right)

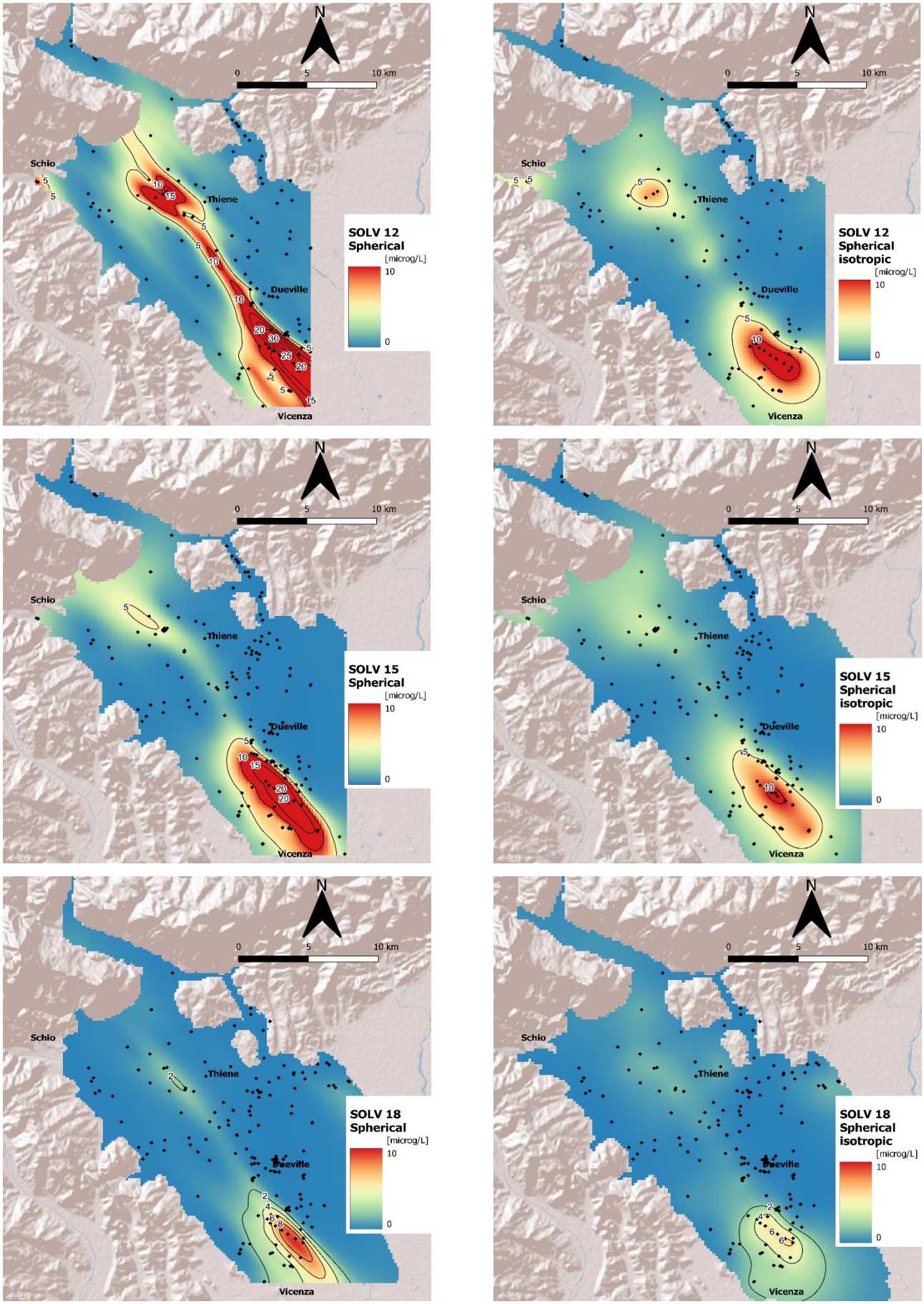


Fig. 51: solvents concentration 2012 – 2018: lognormal kriging with anisotropy (left) and isotropic (right)

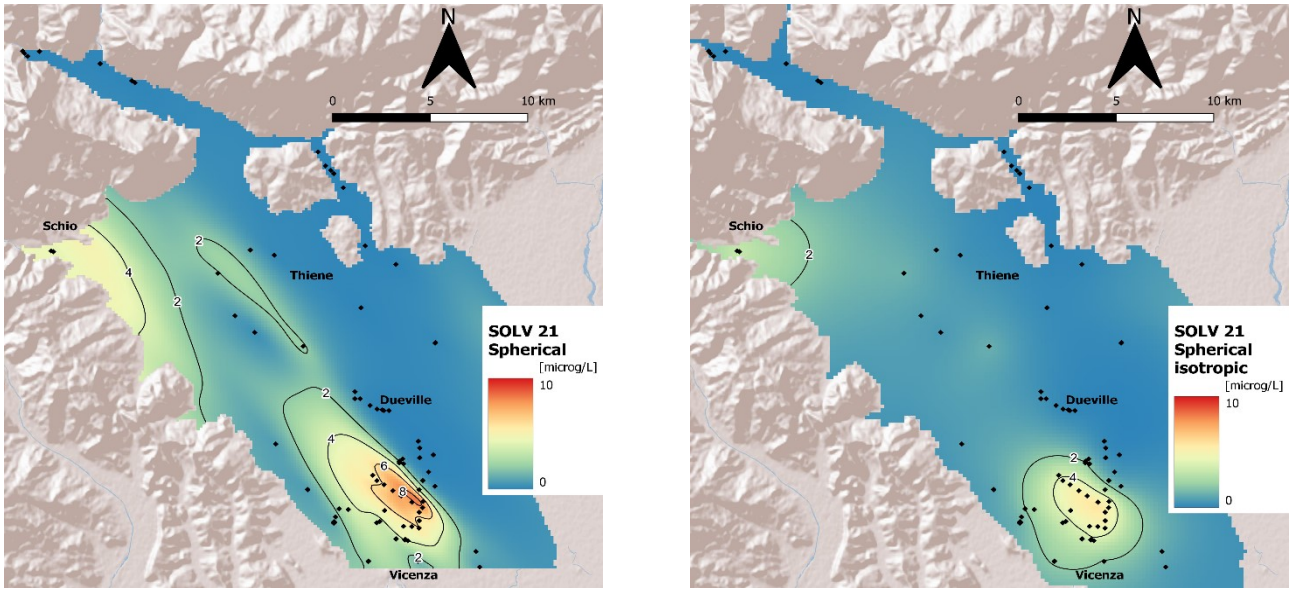


Fig. 52: solvents concentration 2021: lognormal kriging with anisotropy (left) and isotropic (right)

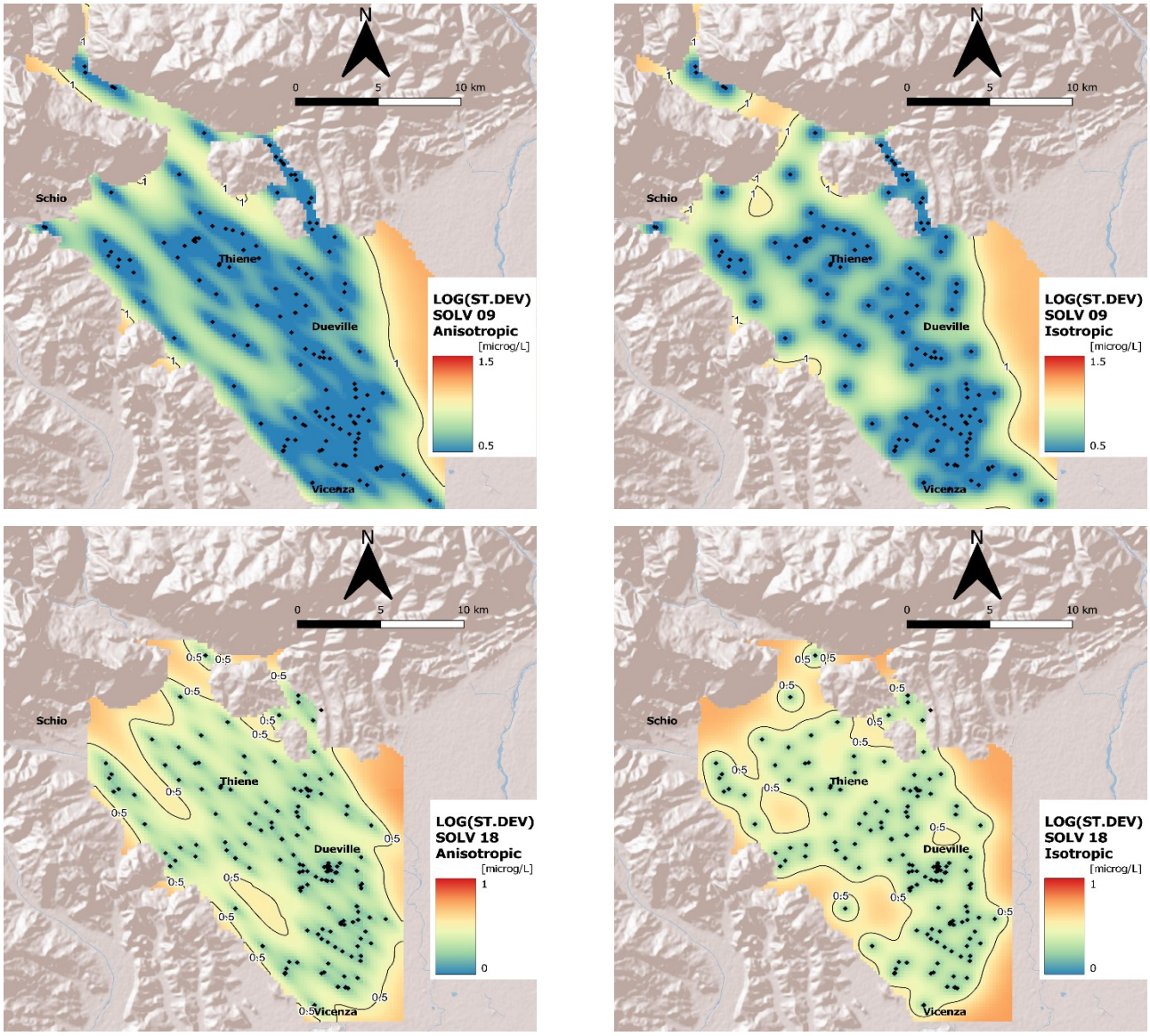


Fig. 53: standard deviations of kriging predictions solvents 2009-2018 (anisotropy left side, isotropic right side)

7.1.2 Discussion

Observing the presented maps, many conclusions regarding the spatio-temporal evolution of the solvent's contamination are drawn:

- As already mentioned, the errors in the first years of simulations are very high (RMSE is higher than the overall mean) and therefore a reasonable uncertainty must be considered when observing the maps. This fact, as already discussed, is caused by a low "quality" of the input data, presenting a high number of outliers and whose distribution is far from normal, even if considering a logarithmic transformation. Nevertheless, obtained maps are useful at least as first approximation for more advanced geostatistical tools, eventually capable in dealing with this kind of data. The errors are lower for later years in which more points are present, and the contamination intensity decreased.

- As expected, the situation in the upper plain in the 90's is particularly concerning and reflects the observations of past publications: a diffuse pollution is present in the zone Schio-Thiene propagating towards SE direction, with solvents concentrations in the order of 30 microg/L. In the upper parts of the valley of Astico river the values are close to zero, but in some points at the foot of the mountains in Carrè – Piovene occasional values above 10 microg/L are observed, suggesting a "fork" pollution shape with one rod directed mainly towards east direction from Schio to Thiene and the second one directed South from Piovene to Thiene (with lower values). This structure is particularly evident in 1991. The values of concentrations in the hotspots of the upper plain remain basically constant throughout the 90's confirming the widespread use of these products in the area as well as the high persistency on the field of the pollutant (their use was declining in the 90's). The values obtained in the interpolations for the 90's can probably be considered as a lower bound for solvents concentrations in the 80's: although measurements are available only in a very restricted number of points, observed values on the area (especially towards Schio) were in the order of 100 microg/L.

- As already mentioned, a meaningful interpolation in the artesian zone for years prior to 2000's is not feasible for the low density (and poor spacing) of points in the area. Nevertheless, measurements tracing back to 1981 in the artesian wells of Vicenza show high values (>10 microg/L) of solvents even in this aquifer. This fact indicates that in the 80's the plume had already reached the most productive layers of the artesian aquifer: for this reason is probable

that the solvents contamination has its origin decades prior to the 80's. The contamination of the deep layers (for example 3° and 4°) of the artesian aquifer also is somewhat of a contradiction of some past publications which stated that the tendency of solvents to reach deep aquifers should be considered limited (*Schwille, 1984*). In this case is clear that solvents are capable, in long times, of reaching high depths: this phenomenon can also be explained by the lower piezometric levels caused by heavy exploitation of deep layers of the artesian aquifer.

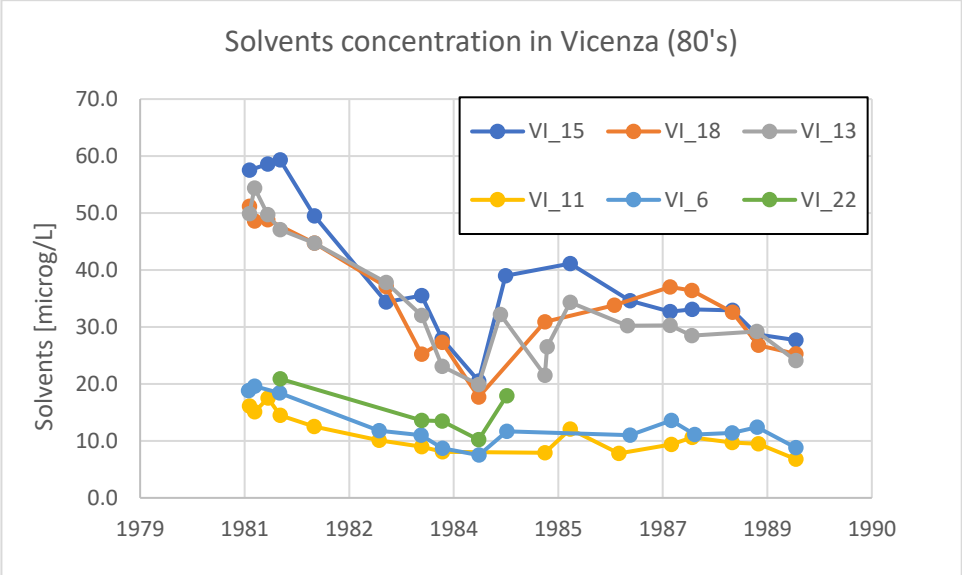


Fig. 54: solvents concentrations in the 80's in Vicenza

- The temporal evolution of the contamination is characterized by the gradual shrinking of the areas interested by high concentrations of solvents; is clear that throughout the years in the polluted areas concentrations gradually decreased. In the upper plain, in the zone of Thiene-Marano vicentino, maximum concentration passes from 40 microg/L throughout the 90's to 20 microg/L in 2009 and then to 5 microg/L in 2015. In the Schio area concentration also decreased along the years and the size of the contaminated area eventually reduced to few points. Similar fate is observed in the artesian zone, although concentrations of 30 microg/L lasted longer than in the upper plain (until 2012); this fact can be explained by the lower values of hydraulic conductivity of the lower plain which may retain chemicals for longer periods.

- The area interested by concentrations above 10 mg/L seems to be larger in the artesian zone (Vicenza area) than in the upper plain: this can be explained by the fact that in the lower plain

the velocities in the aquifer are small compared to the upper plain and this phenomenon may favor the spreading of the contaminant in the transversal direction. Another possible explanation may be given by the fact that the hydraulic gradients are lower and don't determine a strong principal direction of flux as it happens in the upper plain.

- The "link" between the hypothesized contamination origin in the upper plain and the polluted area in Vicenza, is given by the plume which travels through the middle plain. In these areas the contamination is strongly anisotropic, following a clear preferential path, probably related to the joint effect of the presence of ancient riverbeds and the main flux direction given by phreatimetry, forming a "corridor" of high solvents concentrations. The plume propagates N-SE, which reflects the angle of anisotropy (angle formed with the North) of about 145°. This phenomenon is particularly evident when comparing maps obtained with and without the effect of anisotropy: the isotropic maps show interruptions of the plume. This difference between maps is particularly evident observing the 2012 analysis. This effect is caused by points, which are close to contaminated wells, but fall outside the plume, that recorded values close to zero, affecting in this way the kriging weights. Accounting for anisotropy solves this problem, reflecting in a reduction of RMSE, and the obtained plume dimension is therefore more realistic; unfortunately there are wells that insist over this plume and are therefore highly polluted.

- The depletion of the contamination, suggested by the overall decrease of mean values, is observed to occur; the situation in the last years (2018,2021) only shows a residue of pollution in Vicenza area with concentrations below the law limit. The upper plain, in the zones interested by decades of heavy pollution, is now only presenting a slightly increased background concentration of about 2 microg/L.

At the end of the geostatistical analysis, although characterized by some limits, is clear that the solvents contamination has been, and could potentially be in the future, one of the most concerning in terms of concentrations reached. This kind of pollution showed great persistency, for the reasons described in chapter 6.2: values above law limit have been recorded from 1981 until today, even if the use of this product in the area basically ended in the late 90's. For this reason, the persistency of the contamination is in the order of 10-20 years which is also may be underestimated: there are no data for years prior to late 70's and for this reason is difficult to estimate. In the early 80's wells of

Vicenza were already contaminated, is clear that the contamination's origin trace back at least to the early 70's, considering the slow groundwater velocities (in the order of few meters per day), and the hypothesis for which the origin was located somewhere in the upper plain. The depletion of the plume has been a gradual process, interesting both the dimensions of polluted areas, and the concentrations exhibited. The "end" of the pollution at a level that is considered concerning (i.e. with meaningful frequencies of concentrations above law limit of 10 microg/L) is recognized to have been occurred somewhere around 2019; this conclusion is limited to the dataset available, and some occasional values above 10 microg/L can for sure be found today. The only reassuring fact is that at least the contamination no longer interests a large area but could be still present in some unfortunate wells: for example, in a campaign of winter 2021 (data not already available for the thesis) a value above law limit has been recorded in Schio. This fact is also a reminder that kriging predictions are affected by the spatial density of the input points and so it is not possible to inspect spatial discontinuities happening at very small spatial scales, unless an appropriate net of points is present in the micro-area. For this reason, is important to proceed in monitoring for the years to come.

Finally, a very important lesson can be taken from this case: errors in the past may pose serious problems for decades, and the ability to recognize, forecast future scenarios and operate is of vital importance to contain dangerous effects.

7.2 Case study: NITRATE

Considering the whole dataset, some preliminary considerations can be done: the time series starts in 1981; in the first years very few measurements are present and the number of points and analysis available increase in later years.

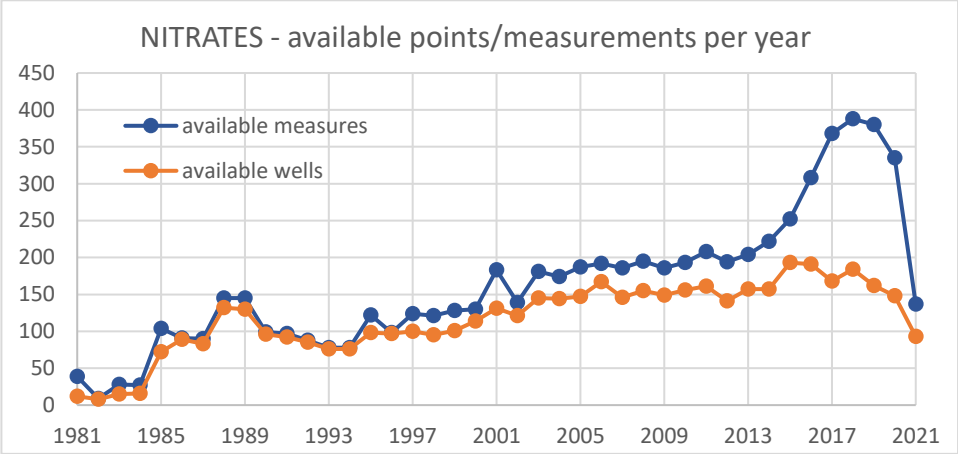


Fig. 55: number of available points and measures per year (nitrate)

The overall mean does not change dramatically over the years after 2001 and is independent from the amount of analysis on the year; this could indicate an overall stable concentration of this pollutant in the macro area considered, although local significant oscillations can be present. The great variability of the first decade of observations (1981-1991) is explained by the fact that the points analyzed are located in small areas: for example, the points present for years 1981-1983 are all coming from Vicenza municipality whereas the dataset of years 1985-1986 is composed mainly by measurements taken in the upper part of the domain. This fact also indicates that in those years nitrate concentrations were higher in the northern part of the domain.

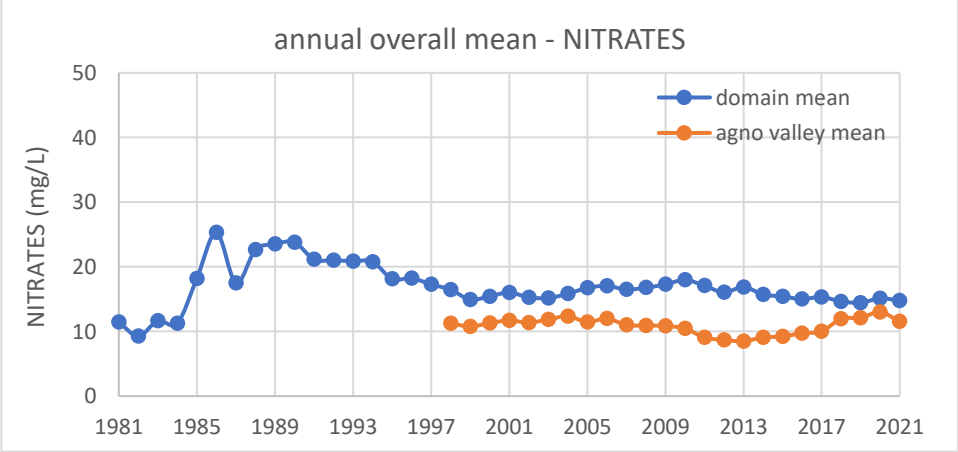


Fig. 56: overall mean of Nitrates

Without considering the part of figure 56 previous to 1988, which for the reasons above have great uncertainty, the overall mean exhibits a slow decreasing trend at first, and then an overall constant value, indicating perhaps a stabilization of nitrate concentration in water.

The situation in the Agno valley is very stable and comparable with the one in the domain; the value of nitrate is around 10-15 mg/L, which is below half of the law limit of 50 mg/L. An analysis of the dataset for this area also suggests that the great majority of the nitrate concentrations recorded are below 25 mg/L with only some occasional values above (for example 52 mg/L in 2017): this situation is considered fairly safe, and evidence of serious contamination is not present. For this reason, the situation in the Agno valley is not further examined.

The background level, indicated by the overall mean, is not describing the various zonal (as well as temporal) discontinuities typical of consequences of contaminations: for this reason, the scatterplot below can furnish a useful tool to recognize possible temporal hotspots.

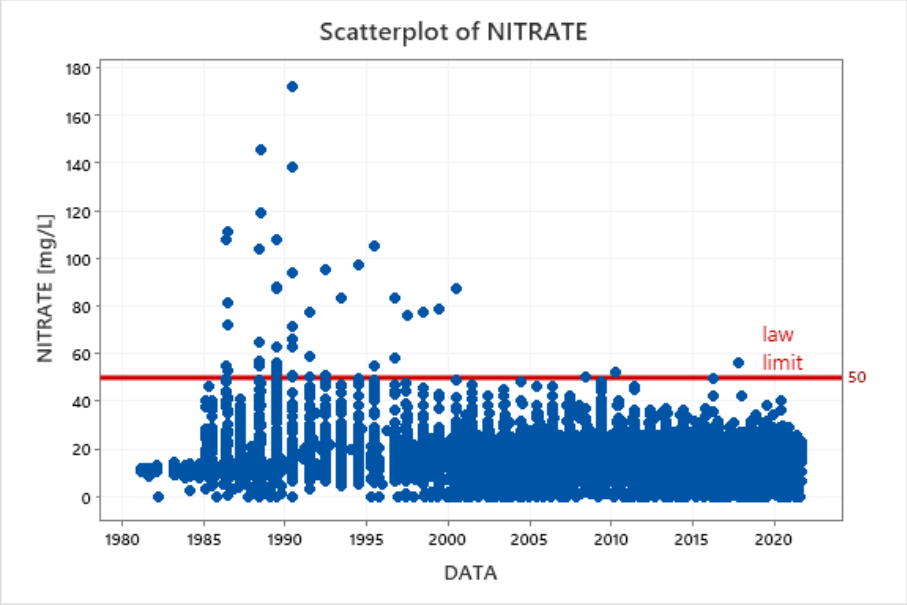


Fig. 57: scatterplot of Nitrate

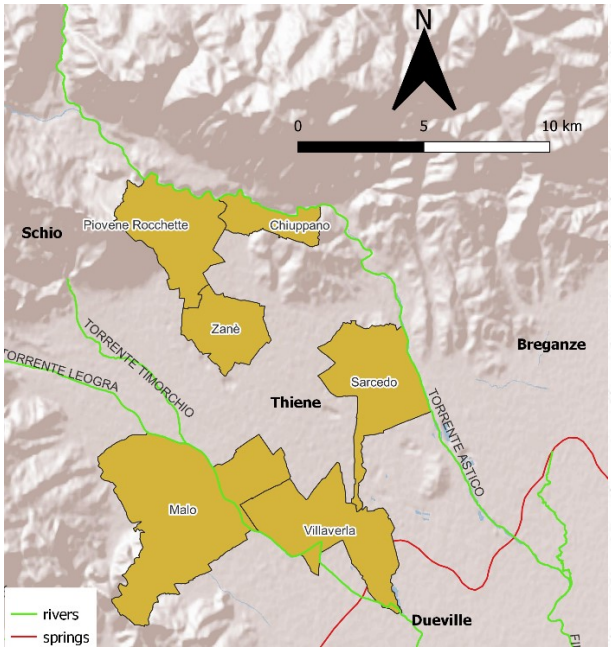
Tab. 11: main statistical indicators for the complete dataset of nitrate

Variable	Mean	SE Mean	StDev	Q1	Median	Q3	Range	Skewness	Kurtosis
VALORE	16.700	0.123	10.243	9.500	16.000	22.000	172.000	2.73	23.66

Is clear that the majority of values above the law limit (50 mg/L) were recorded in years prior to 2000, reaching concentrations of 100 mg/l and above; the last episode is from 2017. As already discussed, the values in the dataset from 1980-1984 are very low and have been recorded in the Vicenza area:

this fact suggests that in the artesian zone there is no significant contamination from the 80's. In particular the maximum values were around 30 mg/L (*Sottani et al, 1982*)

The most polluted areas are in the upper part of the domain, where important uses of nitrate are present: in figure 58 are reported the municipalities, which are all in the undifferentiated aquifer zone, with the highest quantity of nitrogen for unit of agricultural area, due to manure spreading for agricultural production. This kind of fertilization technique is common practice for maize and forage cultivations. The main zones interested by heavy utilization of nitrogen for agriculture are in the group of municipalities Sarcedo-Chiuppano-Zanè located in the upper plain and Malo-Villaverla which are closer to the springs area. Another area (Breganze), in the left part of Astico river is interested by great agricultural production. In these municipalities the agriculture industry plays a fundamental role in the economical texture and is one of the main sources of income for residents. Another meaningful point of pressure could be represented by the wastewater treatment operations, which can be sources of point pollution: important plants are present in Thiene and Schio. This kind of contamination can be considered less of a concern than the spread type one represented by agricultural practices.



Comune	Azoto (kg/ha)
Chiuppano	495
Fara Vicentino	230
Malo	401
Monte di Malo	181
Piovene Rocchette	304
Santorso	180
Sarcedo	453
Thiene	196
Torrebelvicino	214
Villaverla	215
Zanè	331
Zugliano	195

Fig. 58: municipalities with high nitrogen per unit area (source: Rinaldo et al 2007)

For the reasons reported in paragraph 6.2, the nitrate contamination is unlikely to travel for long distances; the most polluted areas are expected to be in the immediate surroundings of the municipalities in figure 58. Some time series plots of wells presenting measured values above 50 mg/L in those area show an oscillating behavior around a value of concentration. Authors (*Altissimo*

et al 1990) concluded that a good correlation is present between release of nitrogen per area and presence of nitrate in the aquifer. This fact suggests that in the areas of figure 58 there may be some hotspots for prolonged periods, related to nitrate spreading through agriculture. Nitrate contamination is a diffused type of pollution.

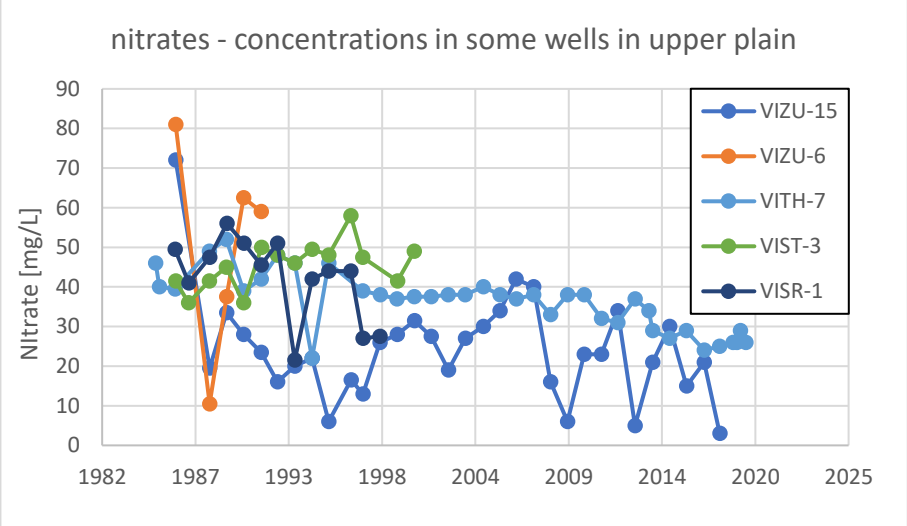


Fig. 59: nitrate concentration in some contaminated wells in the upper plain (Zugliano, Thiene, Sarcedo, Santorso)

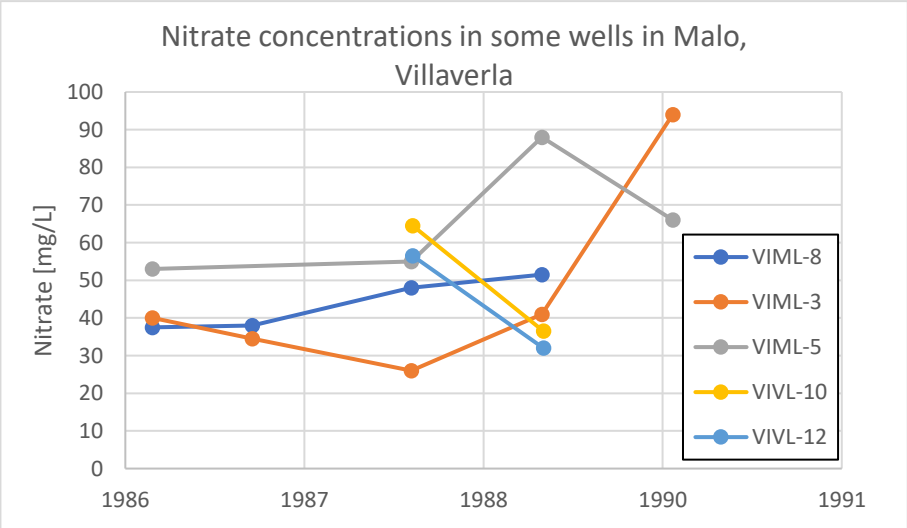


Fig. 60: nitrate concentration in some wells in Malo, Villaverla

Unfortunately, for some of these interesting wells the time series is very limited. The overall maximum values (above 100 mg/L) are all coming from wells of Zugliano and Thiene, which are located in the upper plain, in the period from 1985-1995.

Another interesting area, for which there is no data for nitrogen use per area, is the one in the right side of Astico river, corresponding to Breganze-Sandrigo-Bressanvido municipalities. Although less

concerning than the upper plain in left Astico, in this zone some episodes (one in 1988, one in 2010 and the last in 2017) above 50 mg/L have been recorded, suggesting a possible minor contamination. Unfortunately, is not possible to indicate a correlation between agricultural practices and nitrate concentrations in water due to lack of data on the use of nitrogen in the area.

Overall, the situation of nitrate contamination seems to evolve entirely in the undifferentiated aquifer, where agricultural practices are more prevalent. The artesian zone is less concerning, showing values that rarely surpass half of the law limit values (25 mg/L) suggesting a stable condition, which is for sure influenced by anthropic activities (values above 9 mg/L indicate use of nitrogen above natural levels, WHO 2011), but is not posing a threat for serious health's problems. The difference in nitrate concentration in the various layers, shown in figure below, suggests a fairly similar condition for all the layers, except for the 5° and 6° confined aquifers which always show values very close to 0 mg/L. The first two aquifers seem to fluctuate more whereas the others are very stable; overall the values of nitrate concentrations are not concerning.

Except for the first two layers, aquifers show great correlation.

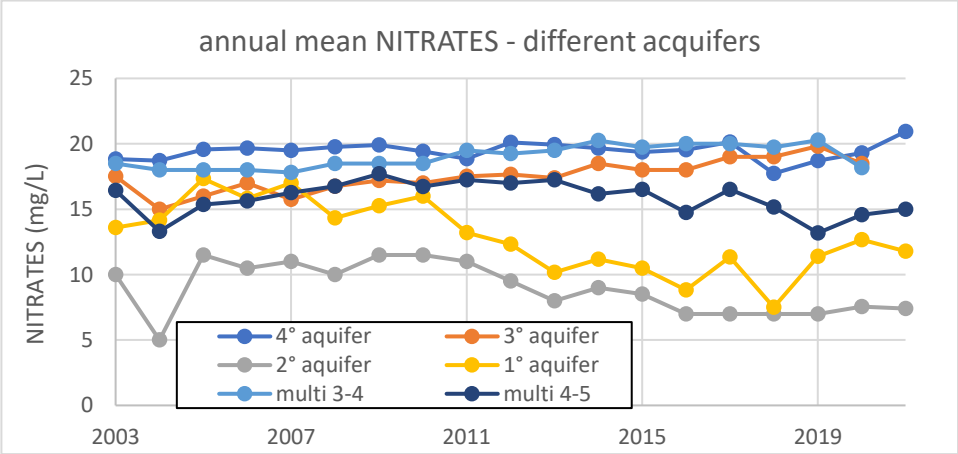


Fig. 61: nitrate annual means for different aquifers

For the reasons above, the analysis is not differentiated for various aquifers: differences in nitrate concentrations among aquifers are not considered significant and worthy of separated studies.

7.2.1 Geostatistical analysis

For the geostatistical analysis, years prior to 1988 are disregarded because the number of points, as well as their distribution, is not satisfactory. For years prior to 2000's the only meaningful area is the upper plain: the southern parts of the domain lack points to obtain meaningful interpolations. Since in the first years of observations the fluctuations regarding the number of points are significant, it is decided to study single years (annual means) with a step of more or less 3 years. For the situation from 2001 onwards, the number of points, as well as the distribution of nitrate concentrations, allows to enlarge the temporal scale, and so it is decided to study triennial averages: this methodology was also adopted by *Altissimo, Passadore* in 2010 ("Pressioni ambientali esercitate sui corpi idrici sotterranei dagli scarichi di reflui civili ed industriali nell'alto Vicentino").
To study the distribution of nitrate concentrations, the mean, standard deviation, and skewness are reported for the analyzed periods.

Tab. 12: main properties of the analyzed periods (Nitrates)

YEARS	N° OF POINTS	MEAN [mg/L]	ST. DEV. [mg/L]	SKEWNESS	$DIST_{MEAN}$ [m]
1988 (up. Plain)	116	25.27	21.42	2.82	691
1991 (up. Plain)	93	21.26	12.30	1.63	814
1995 (up. Plain)	78	20.86	14.83	2.76	899
1999 (up. Plain)	80	15.89	11.33	2.49	938
2001-03	199	15.44	8.84	0.34	660
2004-06	208	17.19	8.51	0.15	569
2007-09	187	16.85	9.23	0.25	581
2010-12	186	17.04	9.02	0.13	533
2013-15	243	15.96	8.67	0.02	490
2016-18	258	14.70	7.10	-0.04	452
2019-21	188	14.23	7.15	0.09	525

In the first years the number of points is lower, and the mean, standard deviation and skewness are higher. In particular, from 1988 to 1999 the skewness is higher than one and many outliers are present (see boxplots); for this reason, a logarithmic transformation is operated before kriging and the obtained distributions have all skewness lower than 1. For the other periods the dataset is studied as it is; the distributions are fairly similar to normal, with mean very close to median, and with only occasional outliers. As example representing the different periods the following histograms show the distribution of data.

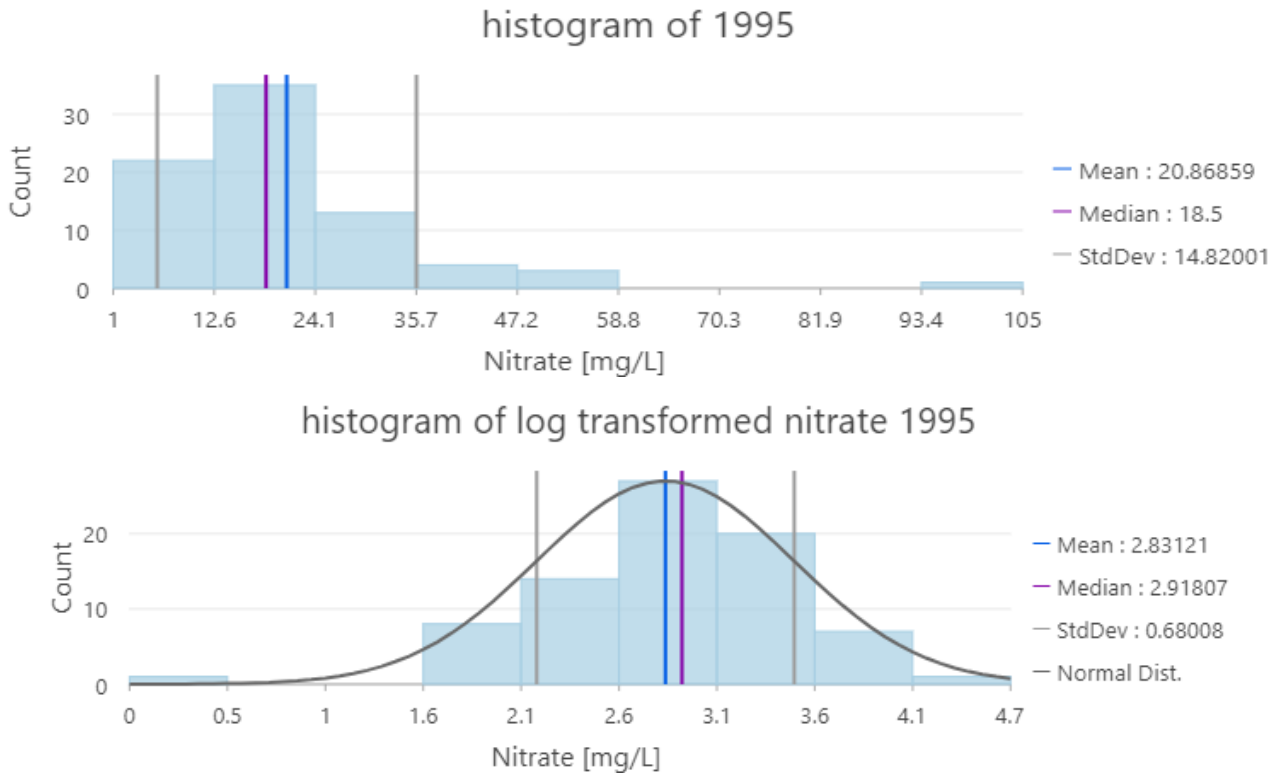


Fig. 62: histogram of nitrate 1995; difference with a log transformation

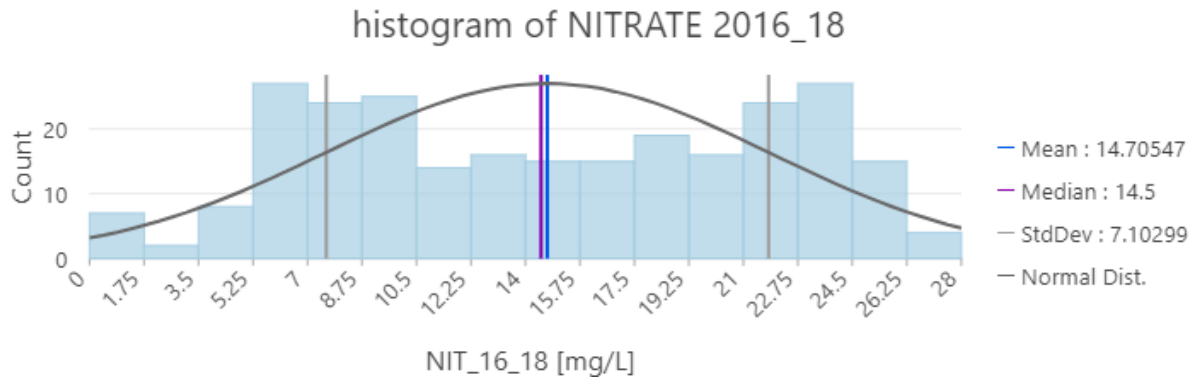


Fig. 63: histogram of nitrate 2016-18

Overall, the time series of nitrate concentrations are fairly stable, with the majority of wells indicating a coefficient of variance lower than 0.3.

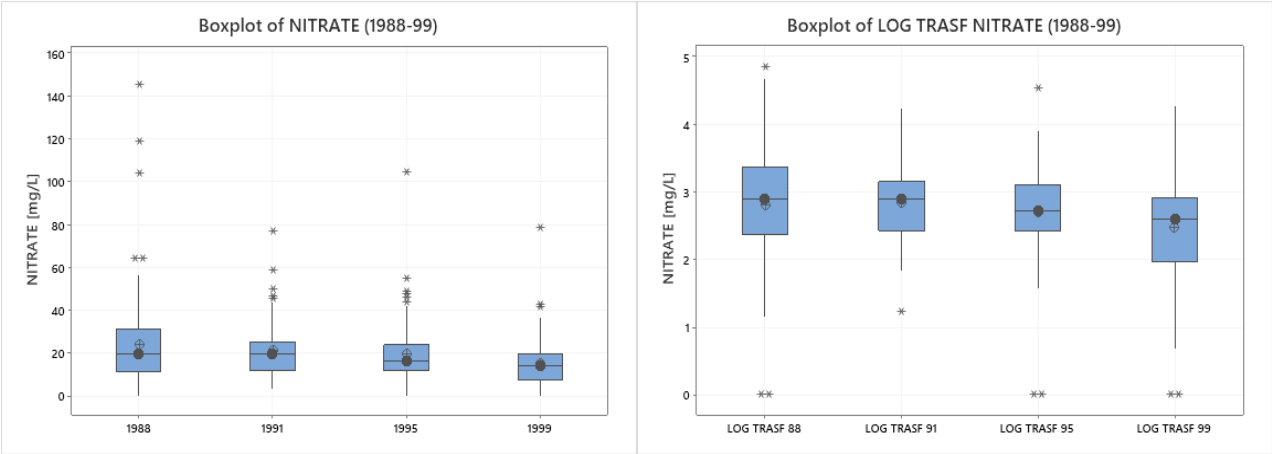


Fig. 64: boxplot of nitrates 1988-1999 in the upper plain (original values on the right and log transformed on the left)

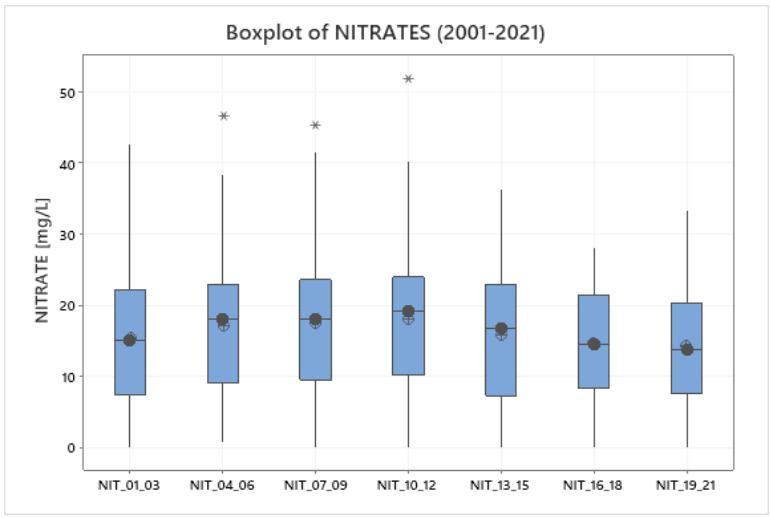


Fig. 65: boxplot of nitrates 2001-2021

Data do not present any significant spatial trend, for this reason is considered second order stationary and considered appropriate for ordinary kriging interpolation. Following scatterplots are representative of the whole analyzed period.

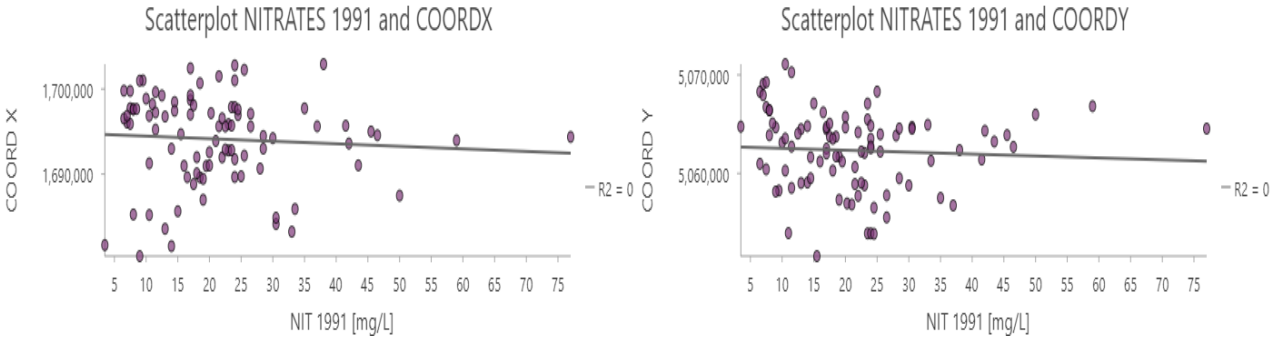


Fig. 66: scatterplots of nitrates and coordinates

For variogram inspection, a lag of 1500 meters is set for the analysis before 2001, whereas for the other period a lag of 1000 meters is chosen as the average distance between points is lower. Omnidirectional (isotropic) variograms stabilize at a certain sill for all analyzed periods and this is another confirm for second order stationarity; the considered models for variograms fitting are the Spherical, Exponential and Gaussian. As examples the following semivariograms are presented, the others obtained variograms are similar; years prior to 2000 are log transformed.

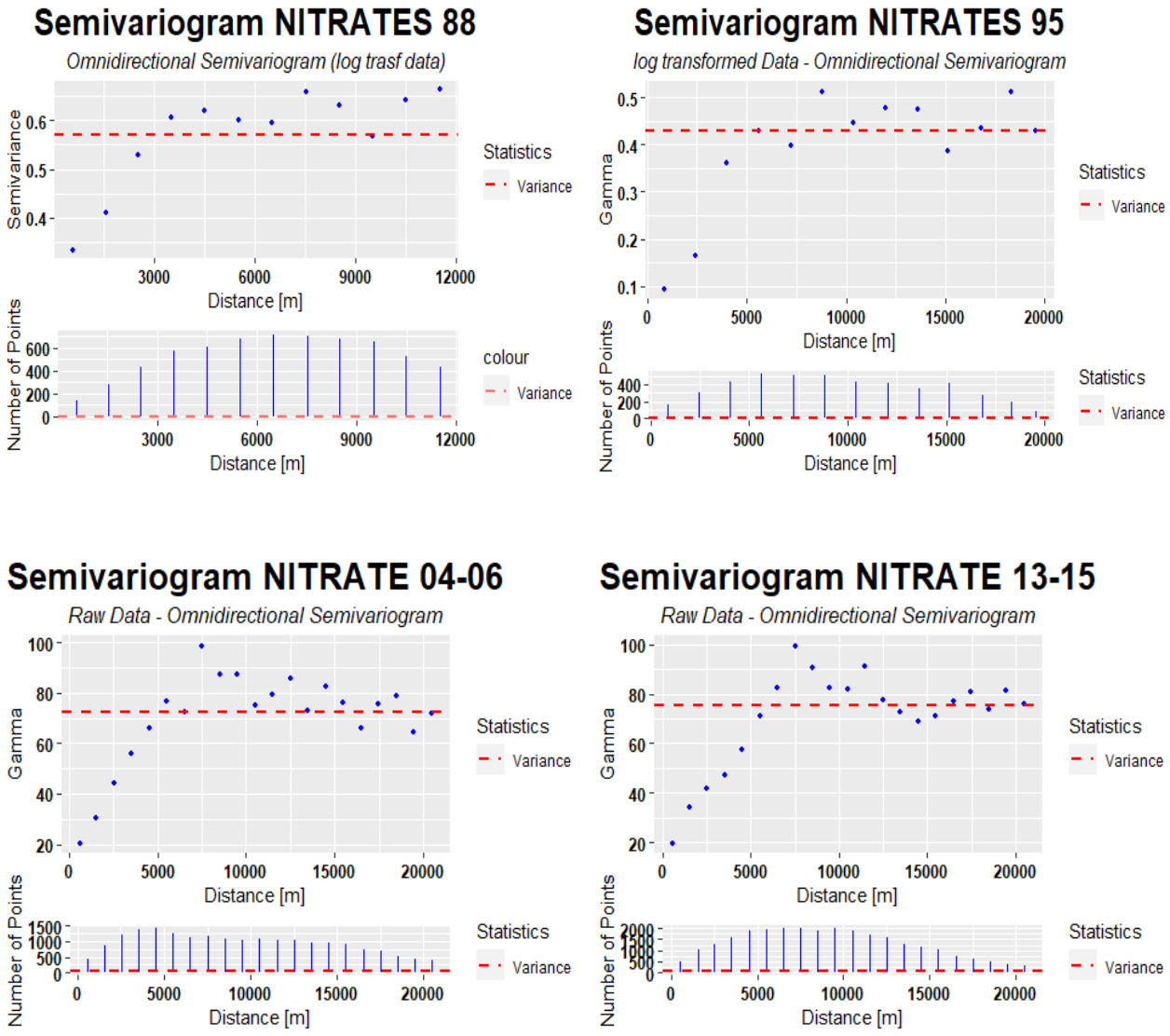


Fig. 67: examples of omnidirectional semivariograms nitrate (1995 logtransf, 2013-15 raw data)

Regarding the possible anisotropy of data, the following directional variograms, obtained with a tolerance of 22° are inspected:

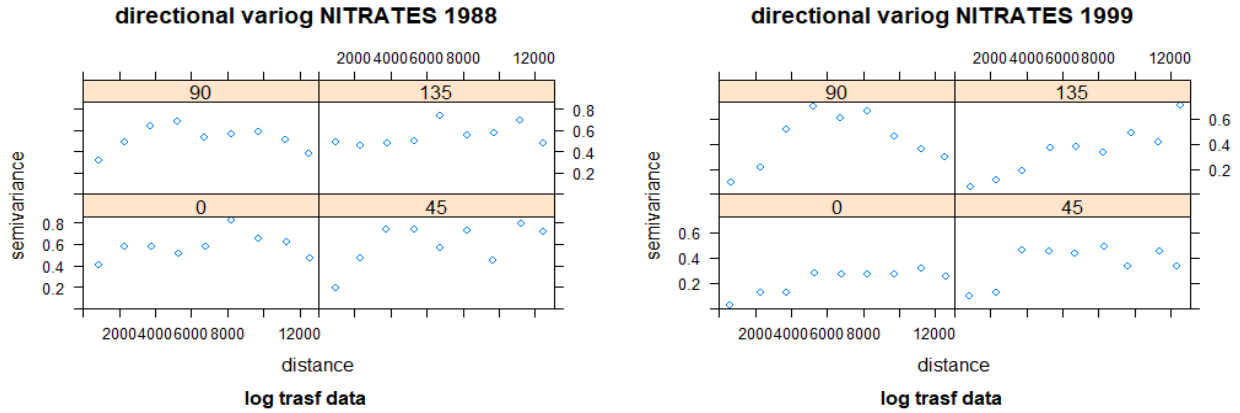


Fig. 68: directional variograms of log transformed data (nitrates 1988-1999)

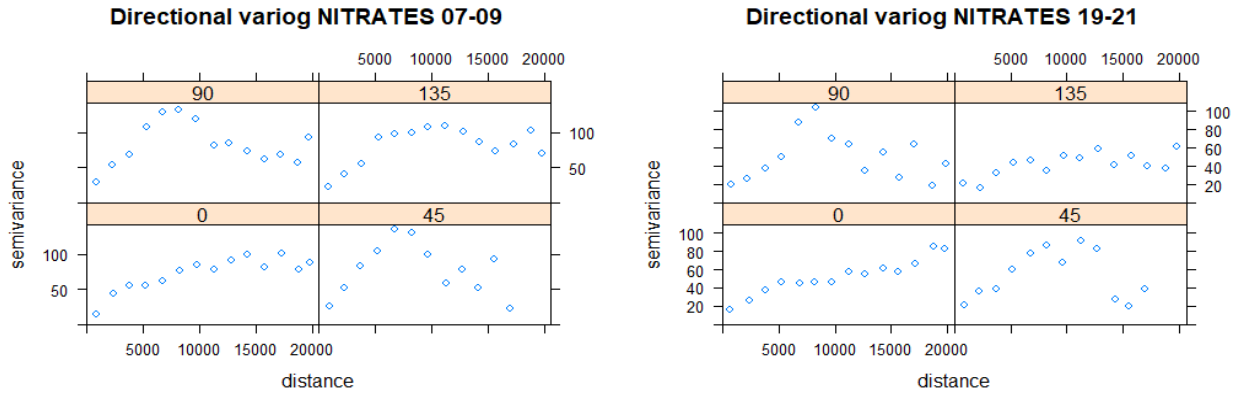


Fig. 69: directional variograms (nitrates)

In the directional semivariogram of 1988, directions 0° and 135° show a behavior similar to a pure nugget effect, whereas the other two directions present higher correlation at short distances (range around 4 km), following a spherical type variogram; all directions stabilize at the same sill value and a mild anisotropy could play a role, especially at short distances. For this reason, is taken into account for the first years; the graph for 1999 show a preferential direction at 0° (N-S direction) in which data remain correlated basically for the entire distance. Is clear that in later periods (observe directional variogram 2007-2009) there is no significant anisotropy, the variogram looks almost the same in every direction, stabilizing at similar sills and with comparable ranges (around 8 km). The isotropic hypothesis is considered accurate enough for interpolation of years after 2003.

The variogram fit, obtained with least squares method, is visually satisfying with theoretical models passing through experimental points.

One possible problem can be given by the presence of a meaningful nugget effect; the ratio between sill and nugget cannot be considered insignificant. The consistent value of nugget suggests that data has meaningful short-scale spatial variability; this fact may arise from the complexity of the nitrogen cycle which can be affected by external conditions of short spatial scales events like cycles of vegetative and "naked" soil conditions due to agriculture, meteorology and others. Another possible explanation is that the density and spatial distribution of points is unable to "catch" the short spatial scale correlation of nitrate concentrations in water.

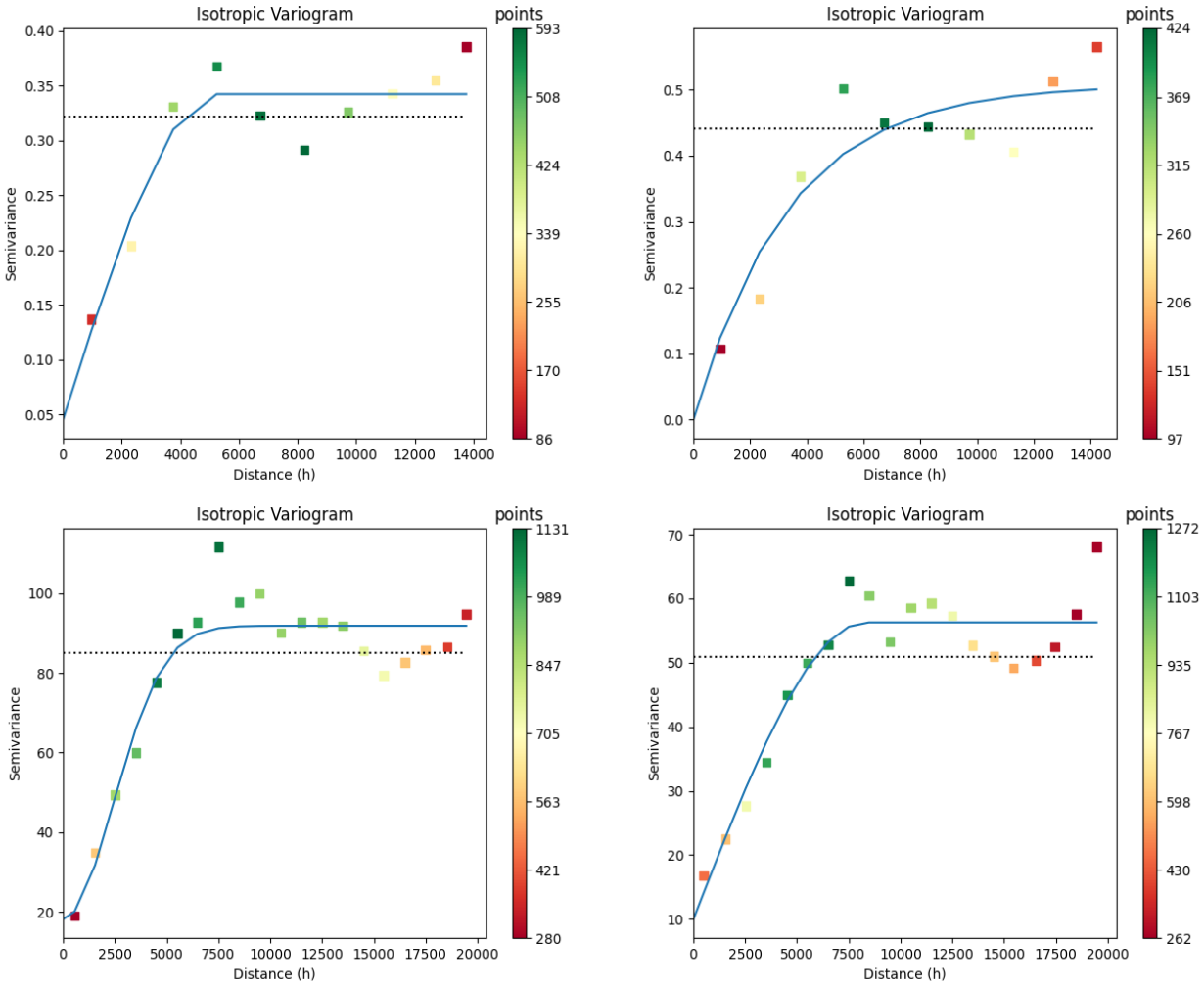


Fig. 70: examples of semivariogram fit, sample variance in dot line (from top left to bottom right: nitrates 1991 spherical fit (log), nitrates 1999 exponential (log), nitrates 08-09 gauss, nitrates 19-21 spherical)

Parameters and leave-one-out cross validation error estimations are reported in the following tables: for years presenting some preferential direction in variograms the anisotropic behavior has been considered.

Tab. 13: ordinary kriging (anisotropic) parameters and LOOCV RMSE (nitrate)

Year	Model	N° of points	LAG [m]	Max dist [km]	Nugget	Sill	Minor range [m]	Major range [m]	angle	LOOCV RMSE [mg/L]
1988 (up. Plain)	SPH (log)	116	1500	14	0.49	0.81	3790	11320	137°	21.91
1991 (up. Plain)	SPH (log)	93	1500	14	0.10	0.36	4201	15000	158°	9.48
1995 (up. Plain)	SPH (log)	78	1500	14	0.02	0.534	4691	15000	158°	10.59
1999 (up. Plain)	SPH (log)	80	1500	14	0	0.48	4762	15000	158°	7.99
2001-03	SPH	199	1000	20	8.22	81.80	5320	15911	167°	4.93

Tab. 14: ordinary kriging (isotropic) parameters and LOOCV RMSE (nitrate)

Year	Model	N° of points	LAG [m]	Max dist [km]	Nugget	Sill	range [m]	LOOCV rmse
1988 (up. Plain)	SPH (log)	116	1500	14	0.36	0.69	4345	21.17
1988 (up. Plain)	EXP (log)	116	1500	14	0.05	0.79	2752	21.79
1991 (up. Plain)	SPH (log)	93	1500	14	0.04	0.34	5245	9.76
1991 (up. Plain)	EXP (log)	93	1500	14	0.03	0.34	5527	9.38
1995 (up. Plain)	SPH (log)	78	1500	14	0.02	0.53	7330	10.86

1995 (up. Plain)	EXP (log)	78	1500	14	0	0.56	10542	10.79
1999 (up. Plain)	SPH (log)	80	1500	14	0	0.475	6586	7.62
1999 (up. Plain)	EXP (log)	80	1500	14	0	0.52	10843	7.67
2001-03	SPH	199	1000	20	8.56	83.82	7061	5.08
2001-03	EXP	199	1000	20	0	83.97	7038	5.11
2001-03	GAU	199	1000	20	19.01	83.73	5887	5.26
2004-06	SPH	208	1000	20	11.15	80.54	7544	5.42
2004-06	EXP	208	1000	20	5.00	81.67	10939	5.25
2004-06	GAU	208	1000	20	18.94	80.34	6082	5.44
2007-09	SPH	187	1000	20	7.39	92.08	7137	5.34
2007-09	EXP	187	1000	20	3.00	92.76	7629	5.01
2007-09	GAU	187	1000	20	18.03	92.06	5943	5.63
2010-12	SPH	186	1000	20	12.28	86.24	7139	6.22
2010-12	EXP	186	1000	20	2.11	86.26	6861	6.22
2010-12	GAU	186	1000	20	21.14	86.17	5889	6.57
2013-15	SPH	243	1000	20	11.09	81.03	8095	5.55
2013-15	EXP	243	1000	20	3.00	81.60	8262	5.65
2013-15	GAU	243	1000	20	20.48	81.15	6891	5.80
2016-18	SPH	258	1000	20	9.85	54.82	7638	3.99
2016-18	EXP	258	1000	20	2.31	55.97	8544	4.02
2016-18	GAU	258	1000	20	14.65	54.59	6047	4.13
2019-21	SPH	188	1000	20	9.98	56.28	8323	4.21
2019-21	EXP	188	1000	20	4.48	57.36	9453	4.23
2019-21	GAU	188	1000	20	15.72	56.38	7029	4.50

The obtained results, compared using leave-one-out cross validation root mean squared error (LOOCV RMSE), are similar for isotropic and anisotropic models. Isotropic interpolations obtained slightly better performances (lower RMSE) than anisotropic except for 1995. For isotropic interpolations, spherical, exponential and gaussian models have been tested: the first two show similar

accuracy in terms of RMSE, whereas the gaussian model is consistently worse. The angle of anisotropy suggests a preferential direction roughly towards N-SE (180° is exactly N-S); this behavior is also confirmed by the maps, which show minor differences between anisotropic and isotropic models. For the first years of analysis, where a logarithmic transformation is operated, only the exponential and spherical model have been tested due to the poor visual fit of the gaussian model.
Overall, the best results are obtained in the later years, when there is a large number of points, and the skewness is very low. Values of sill and range remain fairly stable along the years: this phenomenon is expected when dealing with a diffused contamination where the distance at which the concentrations are uncorrelated does not depend on an accidental punctual release of a pollutant (like the solvents contamination). This is also the reason why the anisotropy does not provide meaningful differences in the maps or better performances: the presence of a diffused similar background level inside the domain makes difficult to observe any predominant plume direction. The presence of some plume-like shapes of high nitrate concentration is likely to occur if in presence of vast areas of high nitrogen use for agriculture.

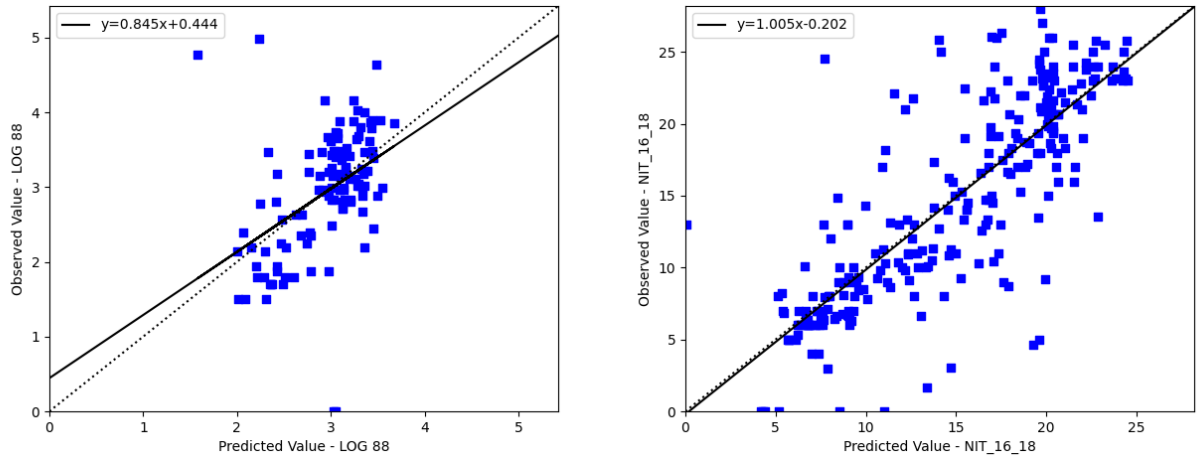


Fig. 71: leave-one-out cross validation: log transformed Nitrate 1988 (left), Nitrate 16-18 (right)

Hereafter are reported the obtained concentration maps: the dark red is set at the dlgs 31/2001 maximum acceptable concentration of 50 mg/L to quickly recognize the most polluted areas.
Standard deviation maps are also present: since they are similar each other, only a few examples, which are representative of the periods, are reported

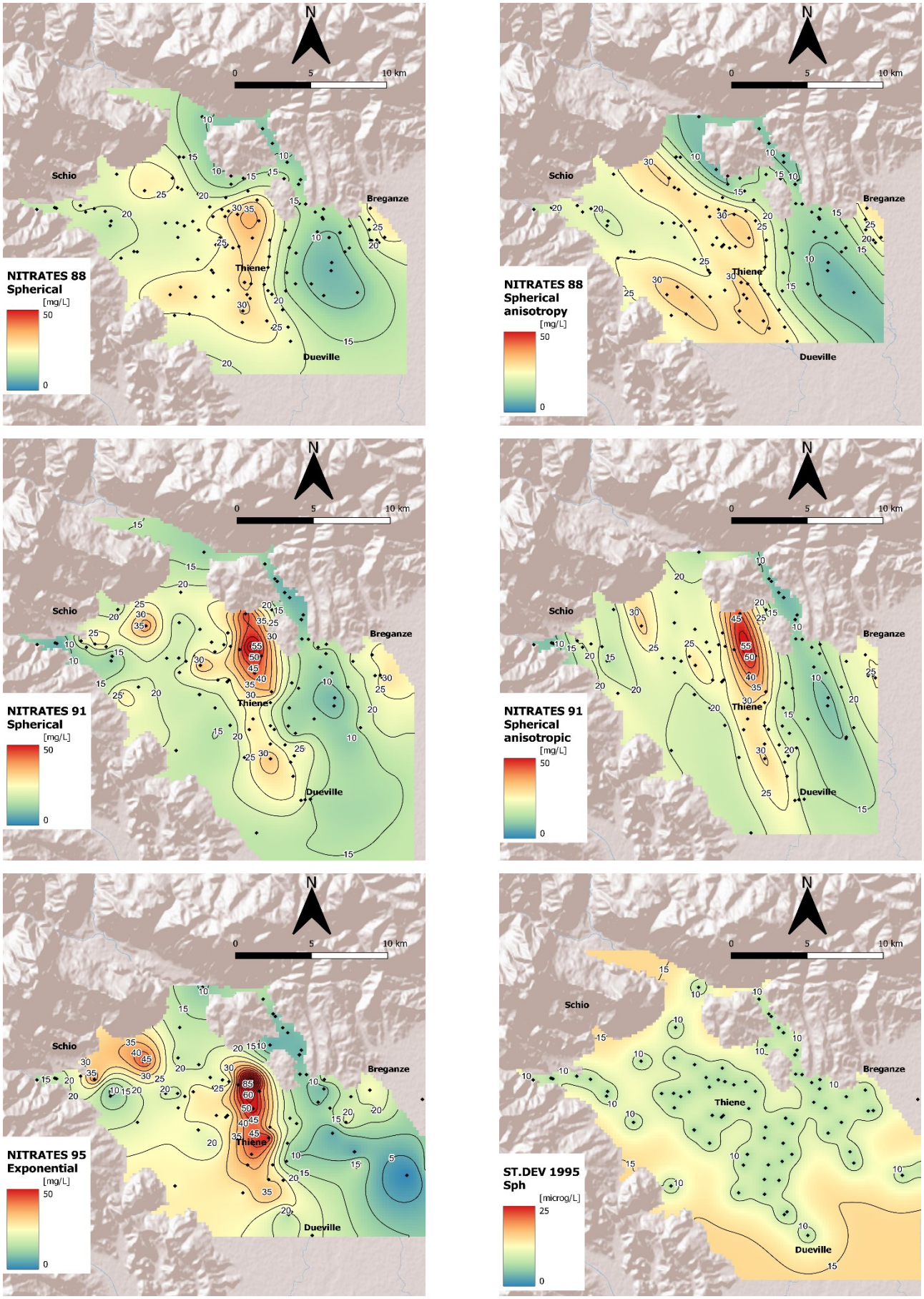


Fig. 72: nitrate concentration and standard deviation of lognormal kriging prediction 1988 – 1995

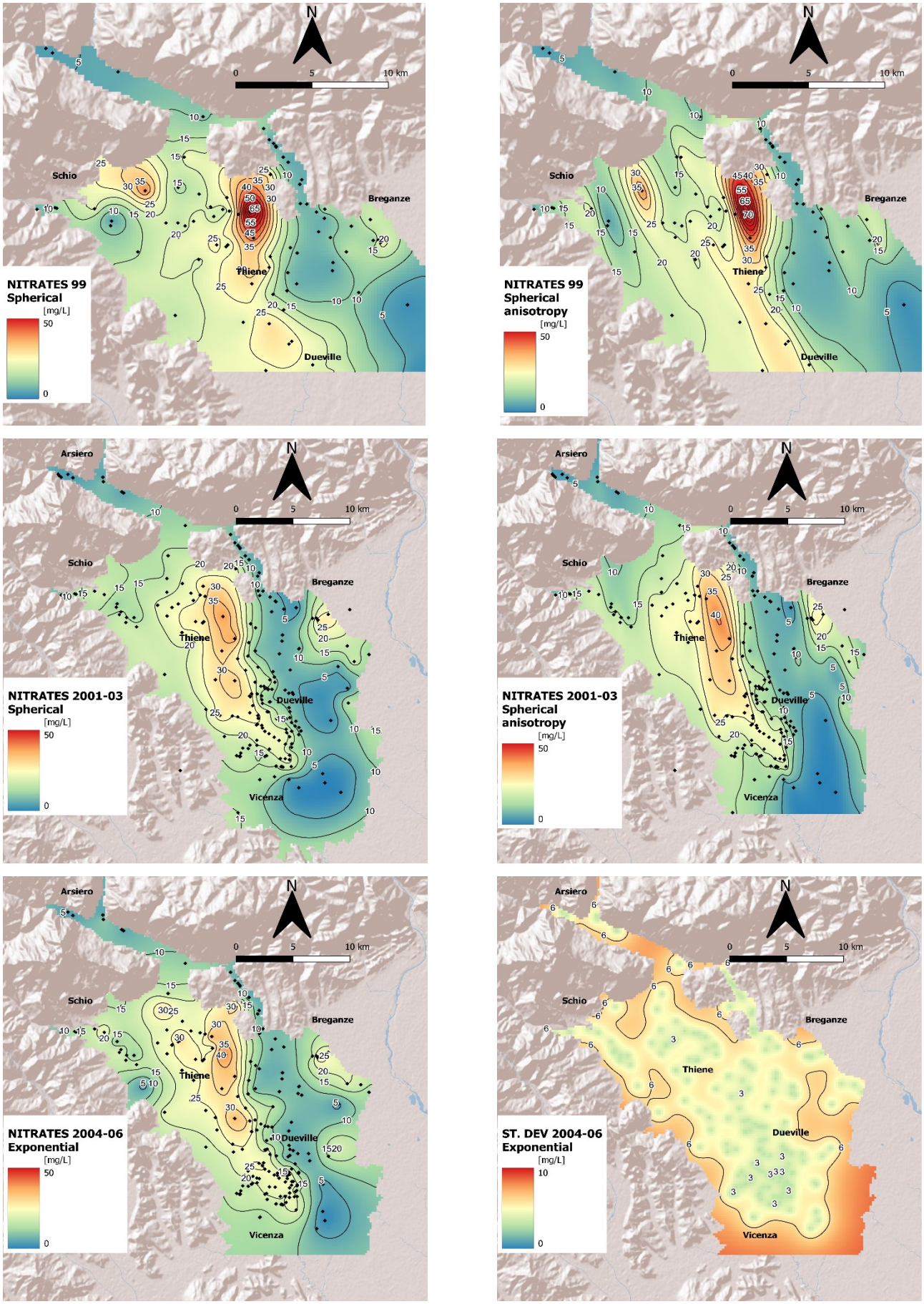


Figura 73: nitrate concentration and standard deviation of kriging prediction 1999 to 2004-06

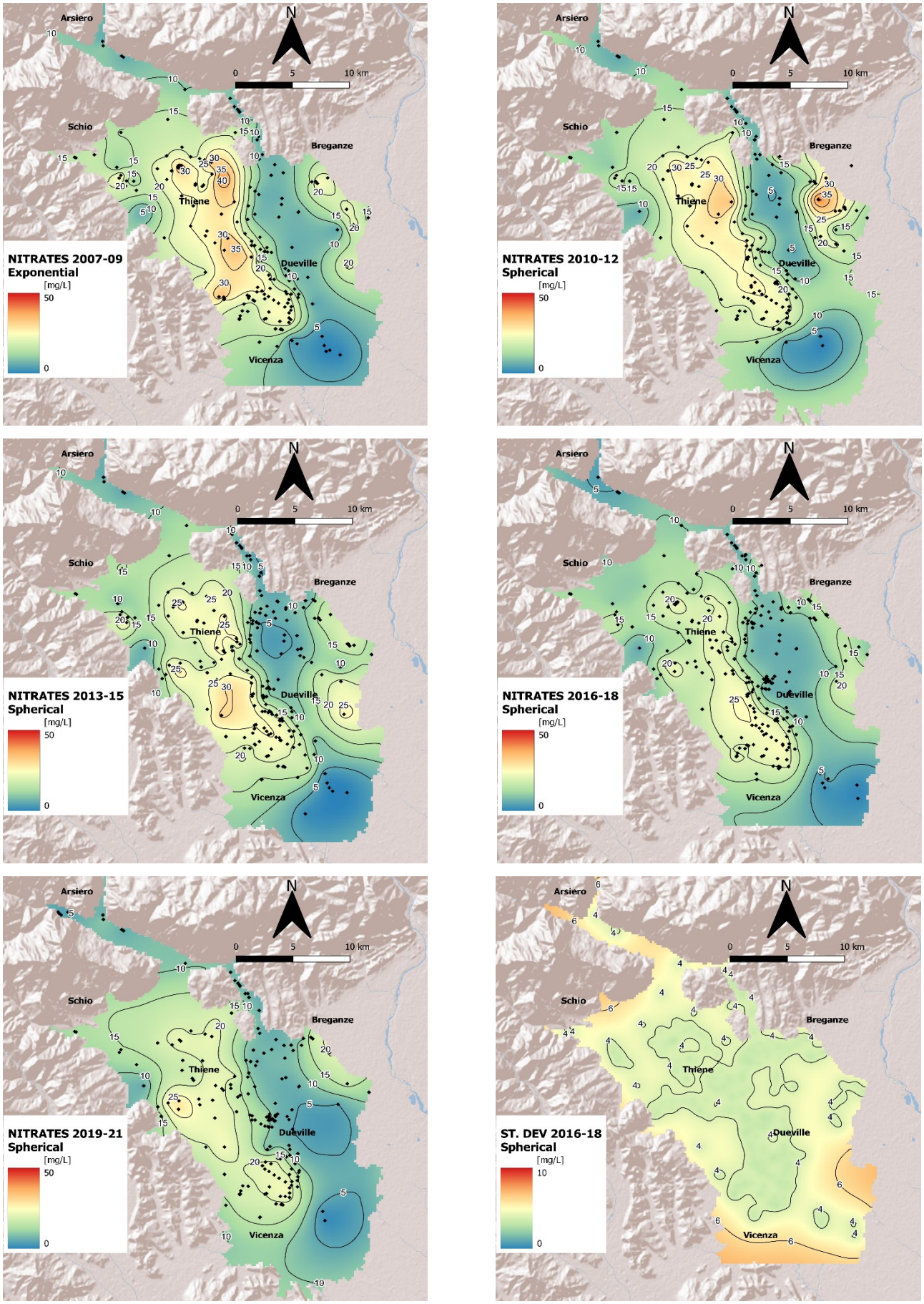


Fig. 74: nitrate concentration and standard deviation of kriging prediction 2007-09 to 2019-21

7.2.2 Discussion

Observing the maps, some observations, regarding general aspects and spatio-temporal evolution of the nitrate contamination along the years can be made:

- The situation in the first analyzed year (1988) show a diffused pollution in the upper-middle plain with an average concentration around 30 mg/L: the mean of nitrate in this year is the highest but is characterized by the presence of many outliers, resulting in a highly skewed distribution.
- Standard deviations maps suggest a mean estimate of prediction error of around 3 to 6 mg/L for maps after 2001 and around 10 mg/L for maps of the upper plain from 1981-1999. The results are considered accurate enough to draw meaningful observations on the predicted nitrate concentrations. This is especially true for the later years: for the first periods variances are higher, this is likely due to less points in the analysis and more skewed distributions. This fact have to be taken into account when making comparisons between periods.

- The situation in the upper plain in the 90's is characterized by some clear hotspots: one in Zugliano-Thiene-Sarcedo and the other in Santorso-Schio. The first one is the most concerning, showing concentrations consistently over the law limit in a contaminated area following the main flux direction towards N-S direction: the analysis considering the anisotropy shows similar results. The other hotspot (Santorso – Schio) interests a smaller area and with concentration approaching the law limit. These two contamination are basically stable for the entire 90's. Minor hotspots can be observed in the Breganze-Bressanvido area, with concentrations around half of the law limit, and towards west in Villaverla-Malo area; in this zone the maps are affected by the number of available points. In 1988 there are many points in the west part of the domain and for this reason a diffused pollution, altough with lower concentrations than the other main hotspot, is recognized whereas in other years with less points the situation seems better. For this reason, the "pulsating" effect of this minor hotspot can be considered a pure reflection of the differrent number of points along the 90's; in reality the situation is likely to be stable and comparable to the one obseved in the maps of 1988.

- The maps show areas that never experienced any significant nitrate contamination: one of these zones corresponds to the urban territory of Vicenza (except the northern part in which levels of 20-25 mg/L are observed occasionally) and overall all the lower plain below the lower (southern) limit of the springs strip. Another "safe" zone is composed by the municipalities in the mountains reliefs, up in the Astico's valley. Other area in which the nitrate concentration in water is stable under 10-15 mg/l is in the immediate surroundings of the Astico river: this phenomenon can be explained by the dilution effect caused by dispersion of Astico's waters. These areas, in which Astico river and spring strips are present are indicated in the graph below. These observations cannot be explained by a lack of points in the areas; these conclusions are reflecting the behavior of real spatial correlation.

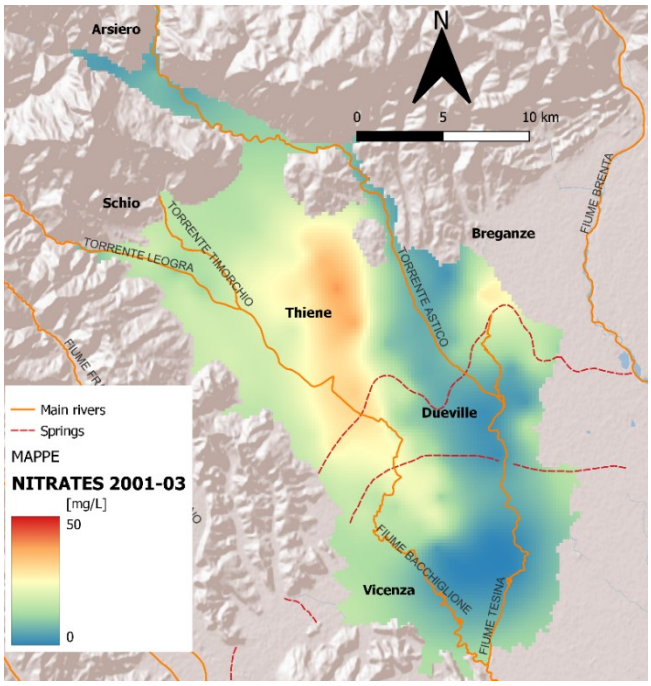


Fig. 75: example of nitrate concentration map with overlayed springs and rivers

- In the maps from 2001 onwards a positive evolution is observed: all the hotspots have lower concentrations respect to the 90's, stabilizing below the law limit of 50 mg/L. In particular the contaminated area of Santorso-Schio is absent, only showing background concentrations of 10-15 mg/L. The only part of the domain in which the nitrate concentrations increase is the area of Breganze-Sandrigo, where the maximum values are reached in 2010-2012. After 2012 also in this area a decrease in nitrate concentrations is observed. The most important contaminated area, in terms of size and concentrations, extending approximately in direction N-S from Zugliano-Sarcedo towards Thiene, is observed to remain fairly stable for a decade (2001-2013). After that, reaches a residual concentration of more or less 20 mg/L.

- As previously mentioned, the two areas in which the load of nitrogen per unit area is higher are Zanè-Sarcedo and surroundings and Malo-Villaverla: an explation for the reason why in the upper zone (Zanè-Sarcedo) the contamination has interested bigger areas and higher concentration of nitrate could be given by the topsoil texture. In the Malo-Villaverla soils the content of clay is higher, resulting in a increased capacity of retention of nitrate (Rinaldo et al. 2007) as well as an higher cationic exchange capacity resulting in better conditions for nitrate removal.

- The background nitrate concentration, outside hotspots, is observed to be from 5 to 15 mg/L: these values are consistent with the presence of antropic activities that can increase the natural background concentration which is extimated to be around 9 mg/L (WHO 2011). Anyway these values are considered completely safe for human consumption.

- From 2013 onwards only some minor hotspots (values around 30 mg/L) are found towards the western part of the domain in the zone of Malo-Villaverla; furthermore in 2019-2021 the only “point” presenting values above 25 mg/L is found in Malo. These observations suggest an overall better condition in recent years for nitrate concentration in water in the analyzed domain, even if monitoring remains a fundamental operation: nitrate contamination is highly dependent on the environmental conditions, and given the velocity at which antropic activities can modify the landscape, nitrate fluctuations can be significant even at small scales.

7.3 Case study: PESTICIDES

In the first part of this chapter a methodology, following the study of Ghirardelli et al. (2021) in "*Thirty-year monitoring of s-triazine herbicide contamination in the aquifer north of Vicenza (north-east Italy)*" is adopted. One of the ideas of the authors consisted in a report of the presence of various pesticides, quantified by a series of simple parameters (given in tables 15-17) that permits an easy overview of the pesticide scenario in the domain. The structure of the table comprehends the max value recorded, the overall mean, the percentage of measures which are different from zero, the percentage of measures above the law limit (n° of measure greater than 0.1 microg/L divided n° of available measures), the years in which the maximum values have been recorded and the municipality in which they have been measured. It has been decided to subdivide the analysis in various periods to inspect also the temporal evolution.

As a first step the whole period (complete dataset) is analyzed in table 15; the main observations are:

- every analyzed pesticide has at least one measure above or equal the law limit (10 microg/L) which indicates an overall presence of these contaminants in the groundwater.
- Some compounds, like metabolites DET and DACT, have their time series starting in later years. This fact is probably due to the complexity surrounding metabolites; is not always easy to know which metabolite can form from the parent molecule (dozens are possible) and how to detect it in laboratory analysis.
- Atrazine has the max concentration detected and its metabolite DEA has the overall greater percentage of values above 0.1 microg/L. Note also that high concentrations of DEA have been found 20 years after the maximum concentration of atrazine, indicating a great persistency of the contamination. The same fact is true for terbutilazine (TBZ) and its metabolite DET although the maximum concentrations of the two are similar and the time series of DET starts only in 2008.
- The maximum values, except from DACT and DET, have all been detected in upper plain municipalities, known to have an important agricultural presence.
- Unfortunately, DACT measures are available in the database only from 2019 onwards. Nevertheless, DACT mean and percentages above zero and law limit are among the most concerning in the analysis.

Tab. 15: analysis of pesticides dataset (from 1981 – 2021)

pesticide	1° measure	MAX (microg/L)	MEAN	% measures > 0	% measures > law limit	years with measures above law limit	year MAX	location MAX
ATRAZINE	1981	9.99	0.016	34.7	0.07	1988,1990	1990	Malo
DEA	1998	0.54	0.025	37.8	7.57	from 93-2003, 2006, 2010	1993	Zugliano
DET	2008	0.1	0.005	20.9	0.08	2017	2017	Caldogno
TERBUTILAZINE	1998	0.1	0.003	16.9	0.09	2002,2008	2002,2008	Malo
SIMAZINE	1981	0.11	0.002	9.6	0.04	1988	1988	Malo
ALACHLOR	1981	4.3	0.003	8.6	0.56	some years between 1988 to 2010	1990	Malo
METOLACHLOR	1992	0.26	0.004	15.4	0.97	various from 2001-2021	2019	Marano vic
DACT	2019	0.13	0.04	67.2	3.02	2019,2020,2021	2020	Vicenza

The analysis continues by inspecting similar tables for two periods, covering the two decades of the 2000's.

Tab. 16: analysis of pesticides database (from 2000 – 2012)

pesticide	period	MAX (microg/L)	MEAN	% measures > 0	% measures > law limit	years with measures above law limit	year MAX	location MAX
ATRAZINE	2000-12	0.07	0.008	40.1	0.0	NA	2001	Marano vic
DEA	2000-12	0.33	0.019	40.3	3.3	2001-2003, 2006, 2010	2003	Zugliano
DET	2008-12	0.07	0.01	37.4	0.0	NA	2017	Caldogno
TERBUTILAZINE	2000-12	0.1	0.004	28.7	0.2	2002,2008	2002,2008	Malo
SIMAZINE	2000-12	0.05	circa 0	17.1	0.0	NA	2001	Thiene
ALACHLOR	2000-12	0.17	0.002	14.1	0.5	2004-2006, 2008, 2010	2004	Thiene
METOLACHLOR	2000-12	0.21	0.006	22.8	1.5	various	2010	Thiene

There aren't recorded values above MAC for atrazine, but DEA have many; also note that DEA mean is twice the mean of Atrazine. This fact indicates that atrazine, following the ban in 1992, is no longer a threat for water quality but its various metabolites still are.

Tab. 17: analysis of pesticides database (from 2012 – 2021)

pesticide	period	MAX (microg/L)	MEAN	% measures > 0	% measures > law limit	years with measures above law limit	year MAX	location MAX
ATRAZINE	2012-21	0.04	0.004	22.3	0.0	NA	2015	Vicenza
DEA	2012-21	0.09	0.006	22.3	0.0	NA	2014,2018	Vicenza
DET	2012-21	0.1	0.002	9.8	0.1	2017	2017	Caldogno
TERBUTILAZINE	2012-21	0.03	0.001	6.4	0.0	NA	2020	Caldogno
SIMAZINE	2012-21	0.02	circa 0	2.8	0.0	NA	2020	Caldogno
ALACHLOR	2012-21	0.02	circa 0	6.1	0.0	NA	2013	Thiene
METOLACHLOR	2012-21	0.26	0.002	9.1	0.4	2016,2019,2020,2021	2019	Marano vic
DACT	2012-21	0.13	0.04	67.2	3.0	2019,2020,2021	2020	Vicenza

By confronting the three tables is clear that:

- Simazine don't represent a problem after the 90s and is rarely detected with values above zero.
- Atrazine, following its ban in 1992, has been constantly decreasing for mean and detection percentage over the years; however its metabolite DEA had a peak on the first decade of the 2000s in which presented the highest mean of the pesticides. In the second decade DEA also shows a great decrease of its mean and percentage of detection, as well as being stable under law limit, suggesting the degradation of this metabolite.
- A similar fate is observed for TBR: although not banned, its use has been decreasing and as a consequence of that after 2008 no more values above law limit are present. As for atrazine, the metabolite DET experienced a rise after the degradation of its parent molecule, reaching its max in 2017.
- Alachlor following its ban in 2006 has been constantly decreasing, with the last episode above law limit in 2010. In the second decade of the 2000s this compound has rarely been detected.
- Metolachlor is one of the most concerning pesticides: this chemical doesn't seem to experience a clear depletion as other compounds. Although it is showing a decrease in mean and percentage of detection, Metolachlor is still found with episodic high values in 2021. This fact suggest that is still been used and its presence in the environment could be a challenge in the years to come.
- DACT, which is thought to be the last metabolite of the triazines, is also a concern for water quality. In the few observed years, DACT shows the highest mean, percentage of above zero values and percentage of above the limit values: for this reason, is considered a contaminant of emerging concern. Is also interesting to note that this metabolite is found with high values in Vicenza, which is located south from the hotspots of the other triazines: this suggests a migration directed south, following the main groundwater direction, of the other triazinic pesticides.

7.3.1 DACT

Unfortunately, the time series for this parameter is short (from 2019 to 2021) and the quantity of data is limited, with around 60 points monitored per year. The area with the most monitored wells is in Vicenza municipality, mainly composed by various pumping field like Moracchino, Bedin and Abbadia Polegge. Other points outside of Vicenza are present, allowing the inspection of the whole domain: the variance of the estimation of the kriging map obtained will be higher in the area with less points. Nevertheless, a map of DACT concentration outside of Vicenza, even if characterized by higher uncertainty, is for sure useful at least as a first approximation.

Regarding the Agno valley, a few points per year are present in the dataset; the recorded values are only zeros. Although it's not possible to conclude that DACT is for sure not a problem in this area (due to the limited length of the dataset) further considerations are impossible to make; in the end the reassuring fact is that there is no DACT presence recorded in the area. For this reason, no more considerations regarding DACT in the Agno Valley are made.

The length of the dataset is almost constant during the years, both for number of wells analyzed and for overall mean, and a table resuming the main information regarding the dataset is reported below.

Tab. 18: number of available points and analysis - DACT

YEAR	n° analysis	n° wells	DACT overall mean (ng/L)	n° of values above law limit	% wells with values above law limit
2019	106	62	46	3	4.8
2020	76	52	47	3	5.7
2021	69	52	42	2	3.8

To obtain further informations on DACT contamination its useful to calculate some descriptive statistics and plot some graphs.

Tab. 19: main statistical indicators - DACT

Variable	Total Count	Mean	SE Mean	StDev	CoefVar	Q1	Median	Q3	Range
DACT (ng/L)	235	45.41	2.26	34.72	76.45	0.00	54.53	71.69	125.32

Variable	Skewness	Kurtosis
DACT (ng/L)	-0.04	-1.25

The scatterplot shows no particular temporal trend of the data and also shows the above the law limit episodes. The histogram exhibits a typical bimodal like distribution: the mode is zero but the rest of the graph is like a normal distribution. Also, mean and median are very close each other. These facts are also true when analyzing the single year (the histogram of year 2019,2020, 2021 are almost identical)

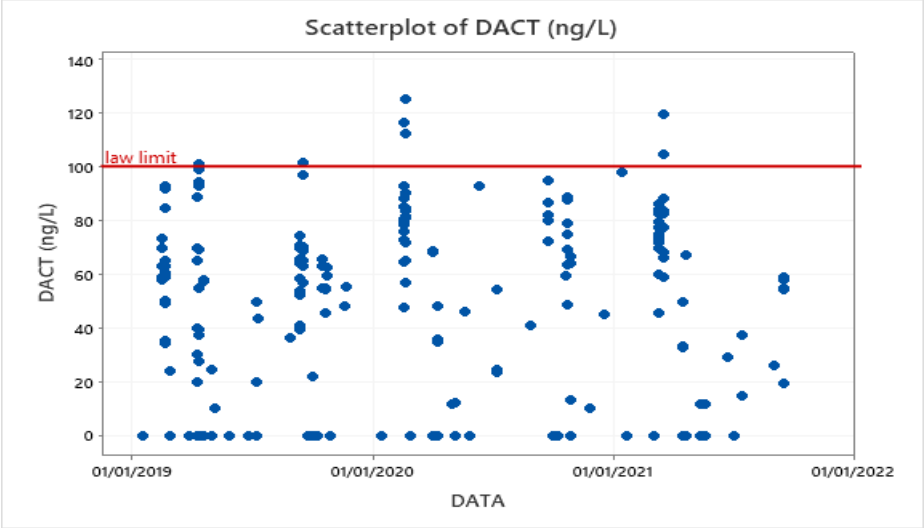


Fig. 76: scatterplot of DACT

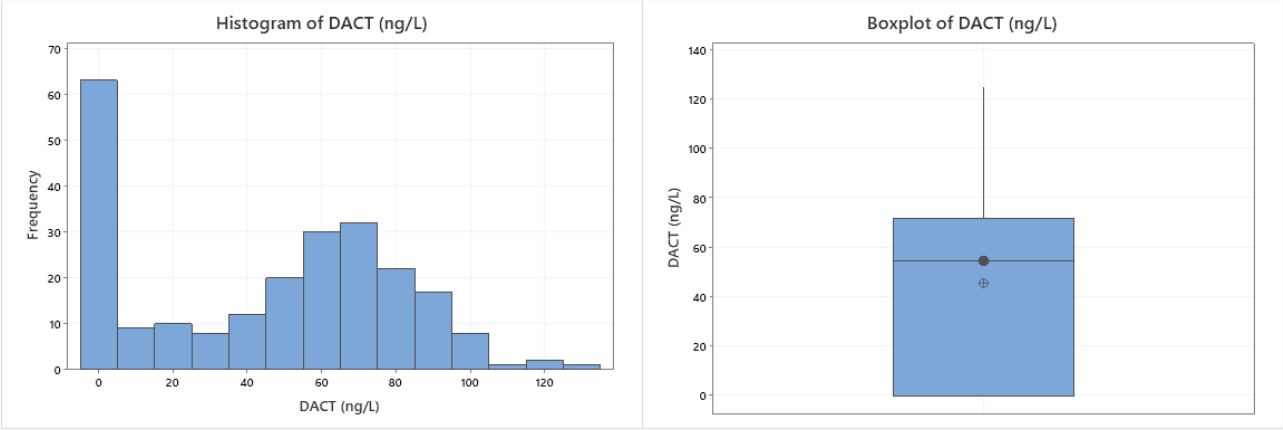


Fig. 77: histogram and boxplot of DACT

Some episodes above the law limit of 0.1 microg/L (100 ng/L) have been recorded: these values are all coming from wells of Vicenza, indicating a hotspot for this contamination in the artesian aquifers. Furthermore, the episodes have all been recorded inside one pumping field.

The time series of the main wells of Vicenza reveals a strong correlation of the DACT concentration. Some of the wells are known to be pumping from the third and fourth aquifer, so is reasonable to assume that the correlation between the two aquifers is very high and also that the other wells (in

which the position of the filters are unknown) are probably located at variable depths between the two aquifers. Unfortunately, there are not records on the first and second aquifers, as well as the fifth and the sixth. There is only one indirect observation that can suggest a better condition of the fifth and sixth aquifers: the well 34, which have filters in 4°,5° and 6° aquifer has slightly lower DACT concentrations although being located very close to the other wells plotted in figure.

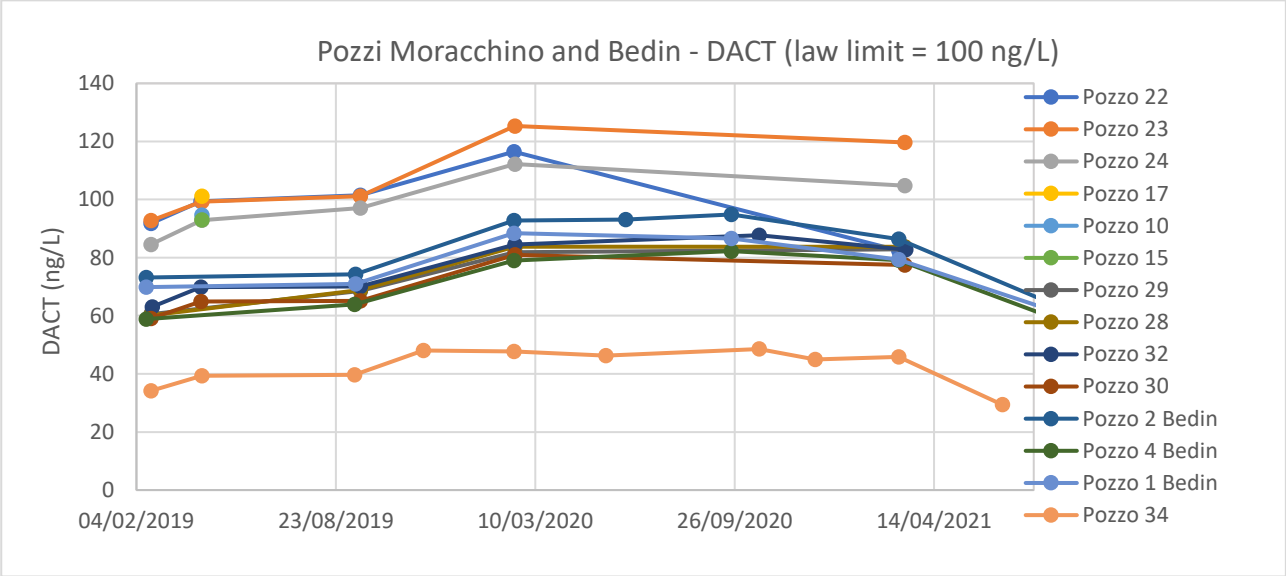


Fig. 78: DACT time series of some wells of Vicenza

7.3.2 Geostatistical analysis

For the geostatistical analysis, the annual mean of DACT concentration of every well is the input; in this way maps of years 2019, 2020 and 2021 are obtained. For kriging method implementation is observed that the overall mean is constant, and the skewness index is much lower than one. For these reasons a transformation of the dataset is not performed.

A distance matrix analysis reveals that:

Tab. 20: distance matrix of DACT (2019-2021)

YEAR	AVERAGE DISTANCE (m)	MAX. DISTANCE (km)
2019	714	42
2020	940	42
2021	741	37

As mentioned before, various authors in literature have suggested that the lag distance should be 1.5 to 2 times the average distance; for this reason, in this case LAG = 1200 m for all three years due to the fact that the average distance are very close one another.

Regarding the spatial distribution of DACT, the scatterplots of figure 79 show correlation both with coordinate X and Y. Furthermore the correlation of DACT with the East coordinate is 0.4 and with North coordinate is -0.8 (strong correlation). This fact suggests the presence of a trend in which Southern points have higher concentration of DACT systematically; for this reason the second order stationarity (described in paragraph 5) is not satisfied and for this reason *universal kriging* may be the appropriate interpolation choice. The same behavior is present for all the three analyzed years; for this reason is presented only the analysis of 2020 (the other years are very similar).

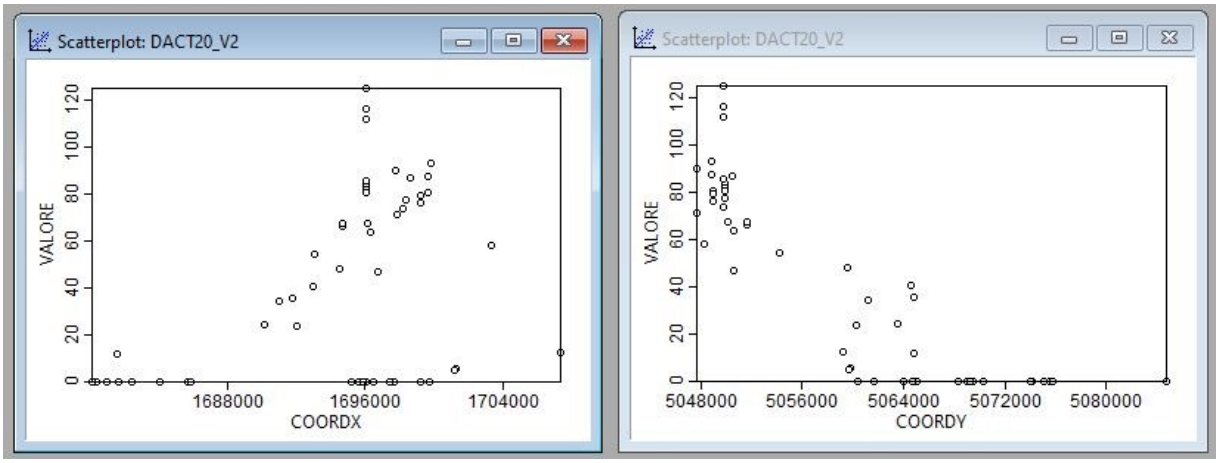


Fig: 79: DACT (ng/L) vs coords

Observing the omnidirectional variogram is clear that there is not a well-defined sill and range and for this reason the gaussian and linear model seems to fit the data better; the spherical model has not been explored in this case due to the poor match with the data. This fact is also confirming that second order stationarity should not be assumed.

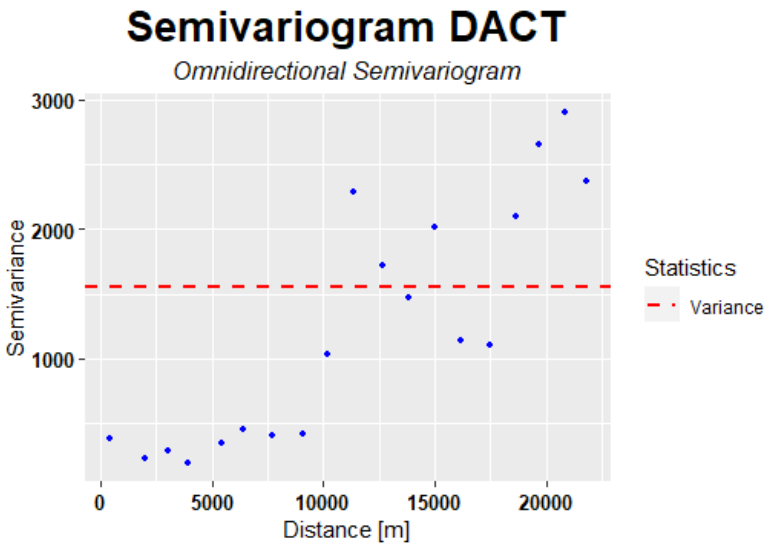


Fig. 80: Omnidirectional variogram DACT

The directional variogram obtained for angles 0, 45, 90, 135° with tolerance 22° reveals that for direction 0° and 135° there is correlation for the whole distance; for the other two directions the obtained semivariograms are similar to the omnidirectional one, with data becoming less correlated after 10 km circa. Overall, the behavior of DACT can be considered isotropic for distances lower than 10 km.

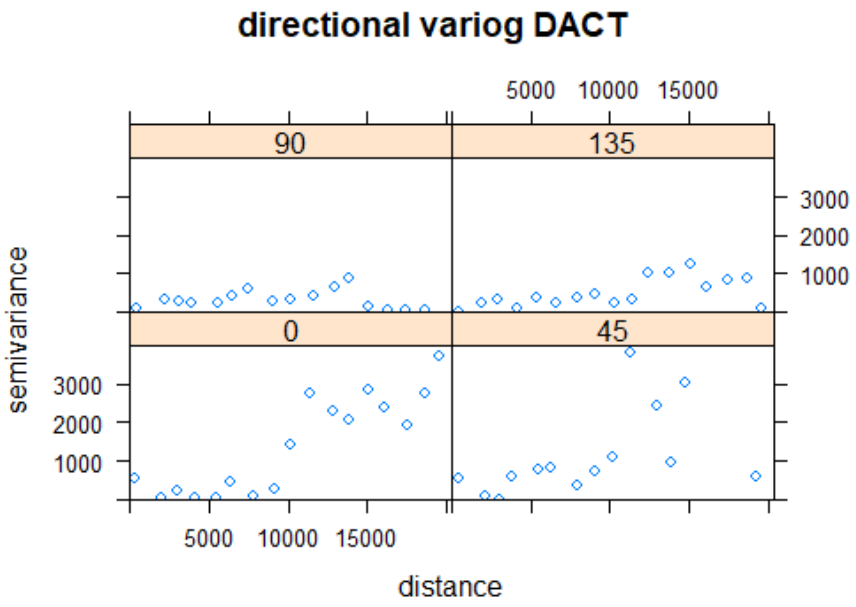


Fig. 81: directional semivariogram DACT

The omnidirectional semivariogram has been modelled, fitted with least squares method, using linear, gaussian and exponential theoretical models. For linear model the values of the coef. of the fitting equation A+Bx are reported, whereas for Gaussian and exponential models the values of nugget, sill and range are present.

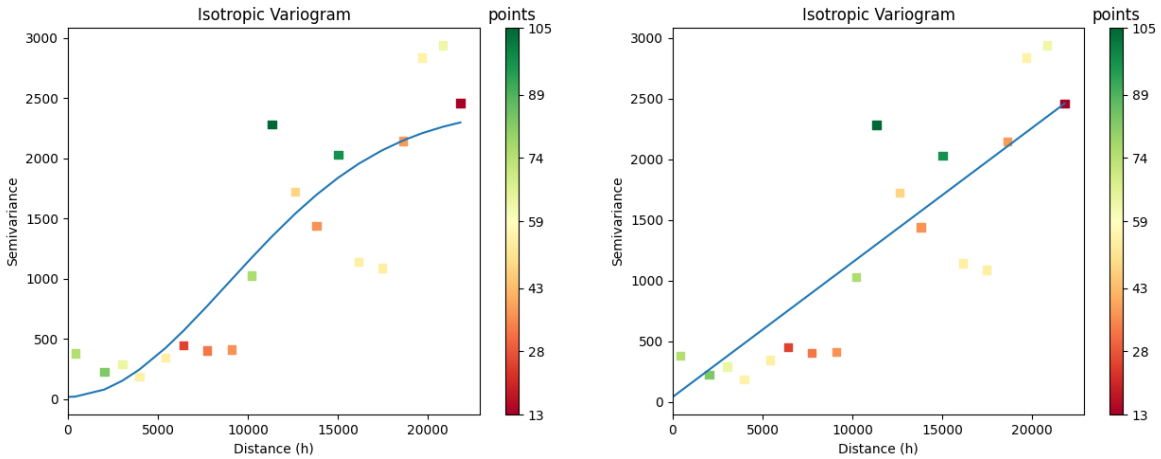


Fig. 82: Isotropic variogram fitting with Gaussian (left) and linear (right) model – DACT 2020

The parameters describing the model are summarized on the tables below where is also reported the evaluation of the results of kriging obtained with leave-one-out cross validation.

Tab. 21: DACT semivariogram fitting (Linear model) and LOOCV results

LINEAR	A	B	LOOCV_RMSE
2019	78	0.0783	14.03
2020	0	0.1113	11.36
2021	0	0.1110	12.31

Tab 22: DACT semivariogram fitting (Gaus, exponential models) and LOOCV results

MODEL	Nugget	Sill	Range (m)	LOOCV_RMSE
GAUS 2019	173	1796	21776	14.65
EXP 2019	10	1407	21776	13.88
GAUS 2020	19	2418	21813	14.18
EXP 2020	0	1922	21813	10.92
GAUS 2021	12	2210	21821	13.63
EXP 2021	0	1751	21821	11.91

The parameters obtained are very close one another except for year 2019: this is easily explained by the fact that in 2020 and 2021 the number of points is exactly the same whereas in 2019 there are ten more points. Is clear that the best model, evaluated by leave one out cross validation, seems to be the exponential one, although being the worst fitting curve for the semivariogram visually. The performances of the model are visually inspected in the graph below: the predicted values of the kriging method (with leave-one-out procedure) are very close to the observed ones.

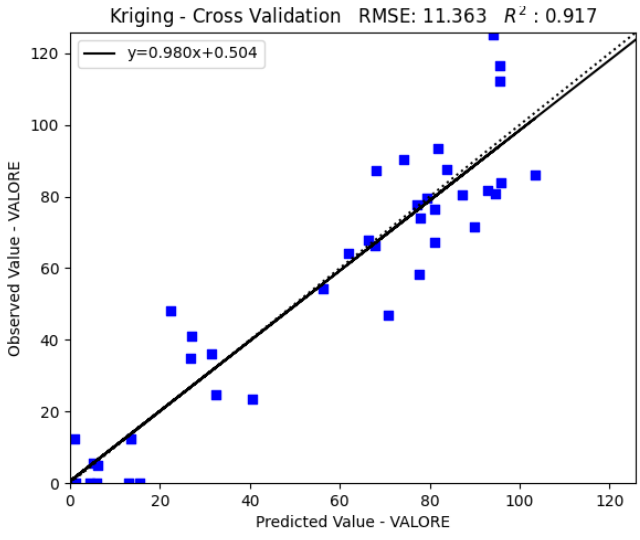


Fig. 83: exponential model LOOCV – DACT 2020

Overall, the results of RMSE are similar for all models, with high performances in all three years. Values of nugget are small compared to the sill, confirming the goodness of the variogram modelling. Values of ranges are very close one another: in this case however looking at the variograms is clear that there is not a real sill because the points do not stabilize at the sample's variance at a certain distance. For this reason, the range is almost equal to the maximum distance considered.

As previously observed, second order stationarity may not be in place for this case; for this reason, a systematic drift of the concentration (universal kriging) may be considered to obtain better performance. Universal kriging with drift modelled as linear combination of X and Y coordinates has been performed, but the obtained results were very similar to the ones obtained with ordinary kriging. For this reason, ordinary kriging is the utilized method: although an external drift may be present, the consideration of it does not seem to add any prevision power. Also, the presence of a trend can be explained by the simple fact that there are not many points for interpolation, and the poor density of points can arise this systematic drift.

The obtained results, concentration and standard deviation maps, are reported hereafter. In the standard deviation maps, are also reported the points of observation; in some maps they are not visible since their presence would make the map too "heavy" from a graphical point of view.
In the concentration maps the dark red is set at the dlgs 31/2001 maximum acceptable concentration of 100 ng/L (0.1 microg/L) to quickly recognize the most polluted areas.

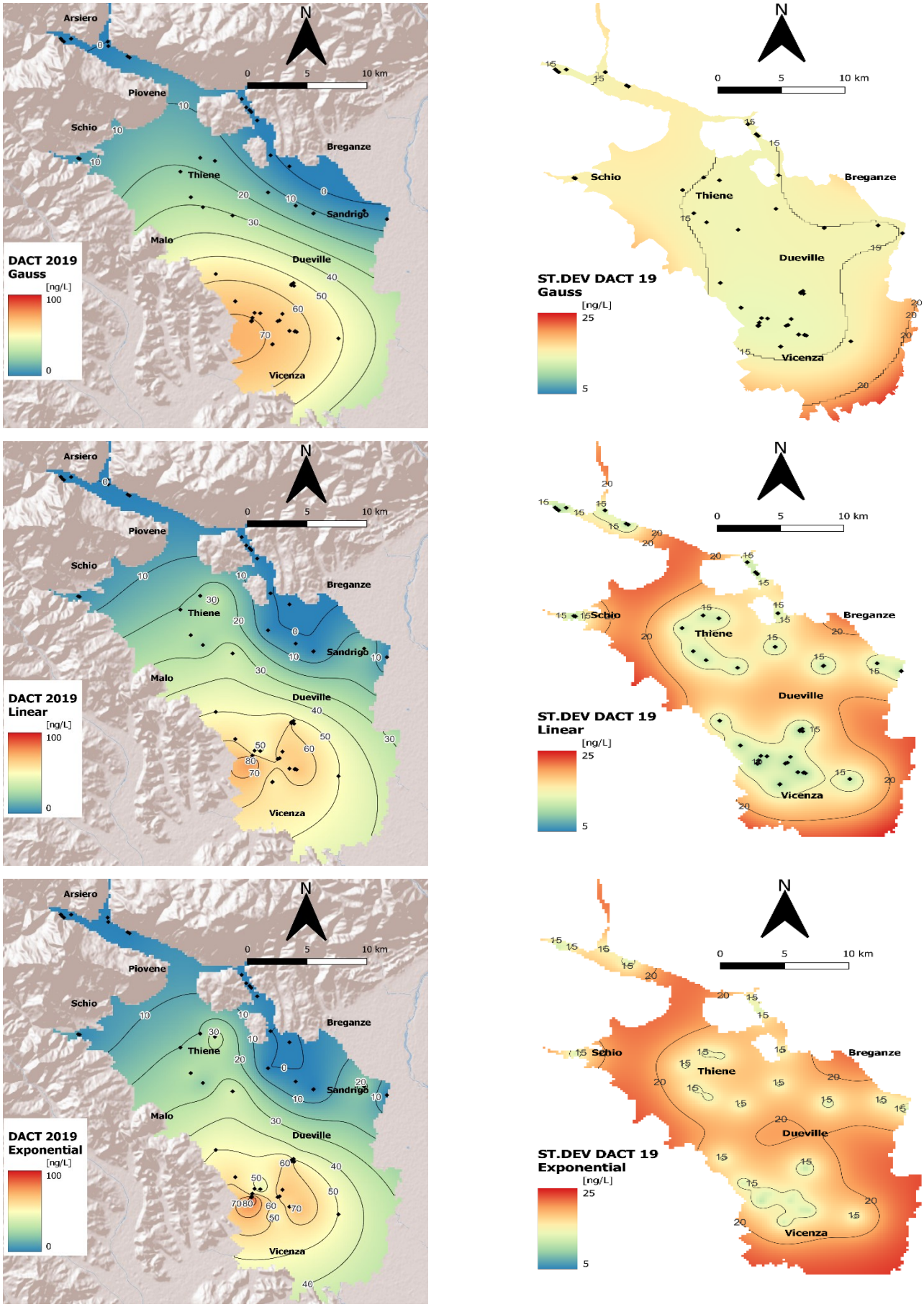


Fig. 84: ordinary kriging (Gauss, linear, exponential) DACT 19

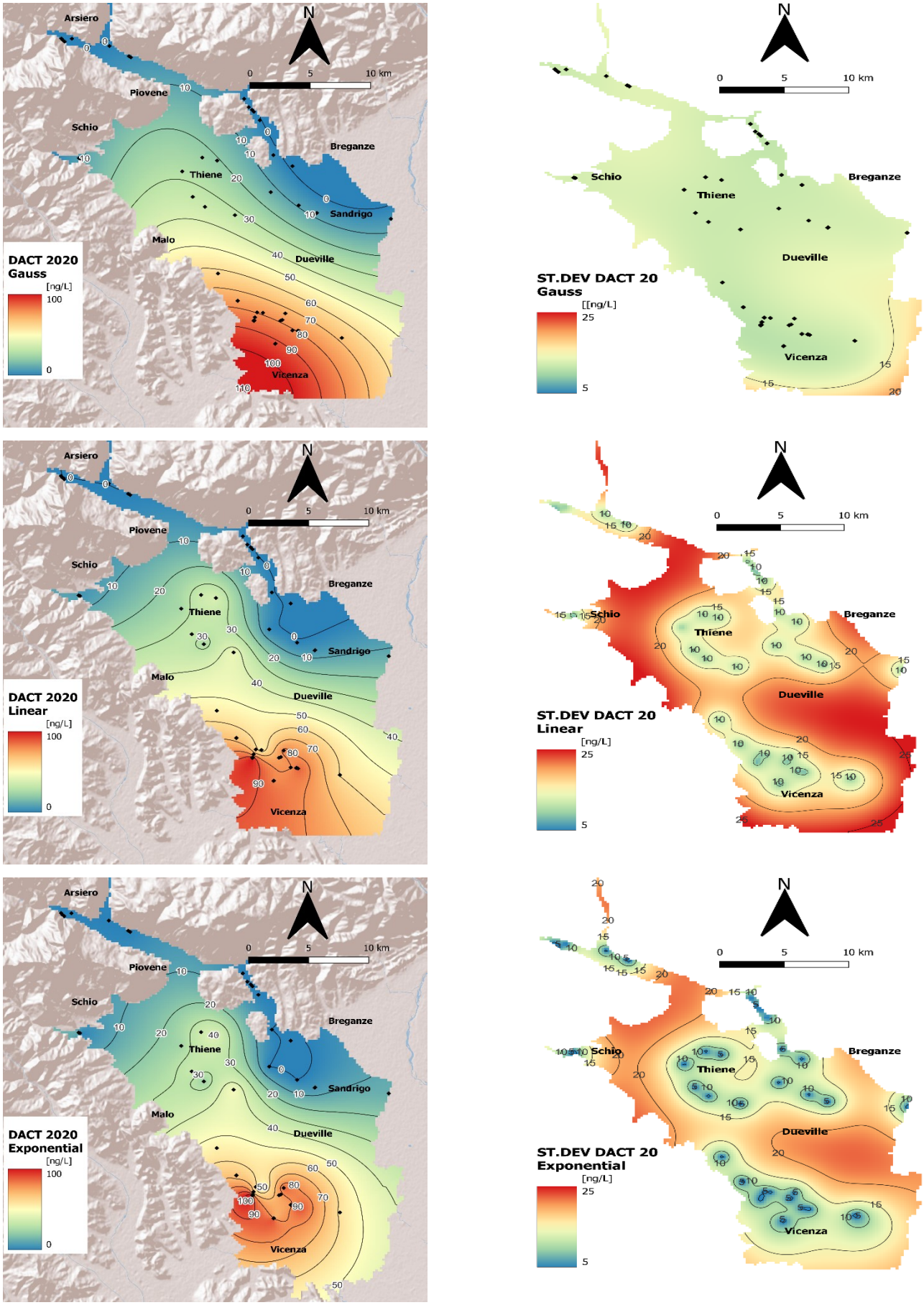


Fig: 85: ordinary Kriging (Gauss, linear, exponential) DACT 20

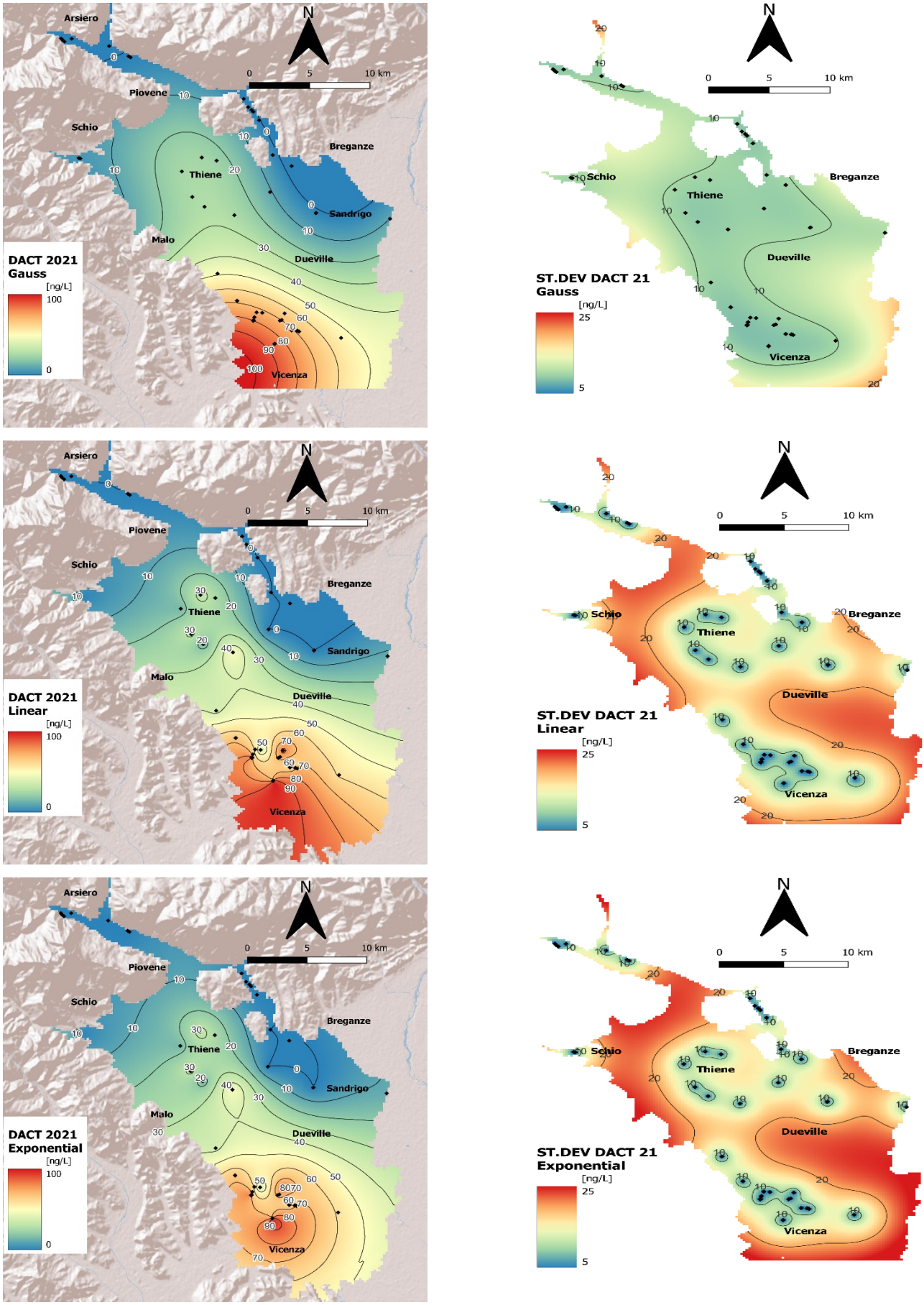


Fig. 86: ordinary kriging (Gauss, linear, exponential) DACT 21

7.3.3 Discussion

Observing the maps, some observations, regarding general aspects and spatio-temporal evolution of the DACT contamination of the observed three years can be made:

- Maps obtained with gaussian modelling are more "smooth" than the ones obtained with exponential and linear models which instead form "islands" of concentration around the points. The difference between the models is also visible in the standard deviation maps where the Gaussian map has an overall quite costant error around 15 – 20 ng/L, whereas the linear and exponential maps have lower errors around the observation points but a higher error distant from the points. These differences can also be inferred looking at the LOOCV graphs: the linear and exponential models predict better than the gaussian model the higher values. In conclusion, the Gaussian maps have an overall smaller standard deviation far from the observation points, but they lose detail around the points.

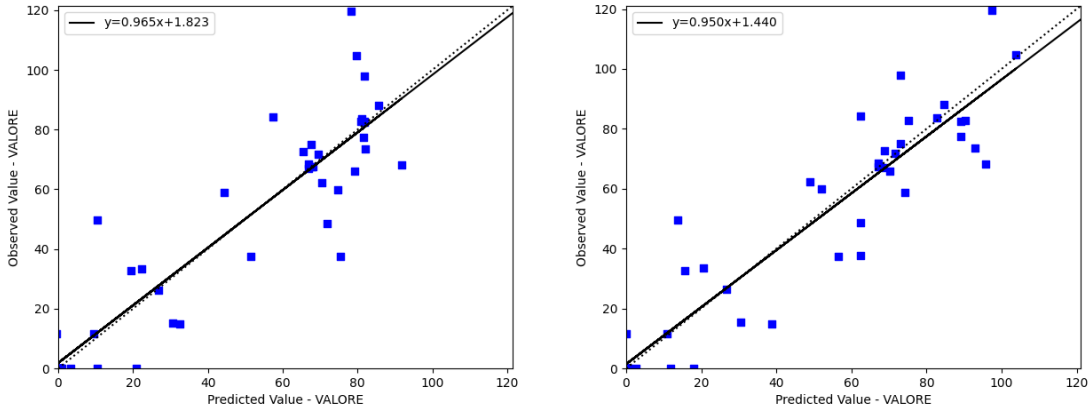


Fig. 87: LOOCV: differences between Gaussian (left) and linear (right) models – DACT 21

- The most polluted area is in Vicenza and all the models confirm that; is also clear that in the eastern and southern parts of the city the prediction values are affected by a greater uncertainty than the central part.
- The temporal evolution of the contamination in the lower plain of the domain follows an initial increase in concentrations from 2019 to 2020, reaching values close to the law limit in the western part of Vicenza (Moracchino wells), and then a slight decrease in 2021 with lower values of DACT in the most contaminated zones. This slow decreasing trend is confirmed by measurements obtained with a new campaign, performed in the winter of 2021, with values that are consistenly lower than the previous ones. In particular, concentrations of DACT sampled in the Vicenza area are around 50 ng/L.

- The upper parts of the domain show low values of DACT concentration; an overall increasing gradient NS is present. This is also confirmed by comparison with the Bano's thesis (2017): in his work DACT concentrations in 2017 (measurements not available in this dataset) were consistently found above 50 ng/L around Thiene, whereas in 2021 this values are found in southern areas like Villaverla and Costabissara.
- Some values above or close to 100 ng/L start to appear from 2020 in Vicenza, especially in NE parts (Moracchino wells) and in central parts (viale Trento): for this reason the measurement campaign for DACT should be replied in the years to come, focusing also in the SE parts of the domain, to have a better understanding of the spatial evolution of the contamination.
- Observing the exponential and linear models, the maximum concentration area (above 90 ng/L) seems to be slightly shifting SE towards the center part of Vicenza, whereas the situation in the upper parts of the domain is stable.

Unfortunately the situation of DACT is unkown in years prior to 2019 in this thesis and for this reason is not easy to estimate how long this contamination could last with high values. Some hypothesis can be made:

- Atrazine, the parent molecule of DACT, showed high persistency (20+ years) as well as DEA, one of its main metabolite. For this reason DACT could be persistent for a period similar to its parent molecules. Unfortunately there are no indications regarding the temporal origin of DACT contamination, therefore is not easy to estimate when this metabolite started to appear. Analyzing the available data, presenting a slow decreasing trend, the most likely scenario seems to be a consistent decrease in values of DACT concentrations across the area, following a slow tail-like distribution.
- Given the fact that atrazine is no longer being used, the situation in the upper parts of the analyzed domain can be considered safe due to the fact that atrazine contamination, and therefore its metabolites, has been migrating south for a long time (40 years circa). For this reason the maximum concentrations reported in 2021 in the northern areas are all below half the law limit; the value of 50 ng/L can be considered a upper bound in these areas.

8 CONCLUSIONS

The obtained results have shown the spatio-temporal evolution of solvents, nitrate and pesticides in the area from the 80's to nowadays. Regarding the situation in the Agno valley is clear that the analyzed contaminants are not concerning. In the domain the most concerning contamination case, the one regarding solvents, has interested the area for decades: tenors higher than the law limit of 10 microg/L have been observed troughout the whole domain. In particular, the contamination interested the aquifer in the upper plain areas, especially in the zone of Schio-Thiene, and the plume of contamination followed a N-SE travel direction downstream, reaching Vicenza municipality; unfortunately is not possible to estimate a certain temporal origin of the contamination, but the first measurements of the 70's suggests that the pollution was already in place. In the 2000's the plume presented a gradual depletion both in terms of concentrations and size of the interested areas: in the last years the contamination seems to have ended and is now presenting only an increased background concentration (lower than 10 microg/L) in the historical interested areas. This conclusion, drawn from the observation of kriging concentration maps, is limited to the macro-scale behavior of the contamination, which is dictated by the density of available observation points, therefore small residual hotspots are still possible.

Another contamination episode regarded nitrate, which presented a diffused pollution, interesting many municipalities in the upper and middle plain like Malo, Villaverla, Sarcedo, Zanè, Santorso, Breganze and others. Kriging concentration maps show good agreement with the data on the nitrogen spread per unit area. Nitrate concentrations above the law limit (50 mg/L) were observed troghout the 90's and with tenors close to the law limit in the decades after; overall the contamination showed a stabilization in the first decade of the 2000's and then a gradual depletion in the last decade where maximum values are around 30 mg/L. A possible improvement of this study on nitrate can be made by implementing co-kriging method, using nitrogen load per unit area as a covariate for nitrate concentration.

Regarding pesticides is observed that atrazine and its various metabolites persisted in the environment for decades, and nowadays the metabolite DACT has the highest concentrations in the domain among the pesticides, eventually exceeding the law limit o 0.1 microg/L; another concernign erbicide is Metolachlor which occasionally also exceed the law limit. Kriging interpolation performed on DACT concentration revealed that the most polluted zone is in the lower plain in Vicenza municipality; in the upper and middle plain concentrations are around half the law limit in the areas in which concentrations are different from zero. Being an atrazine metabolite, DACT can potentially be a challenge for water quality in the years to come, given the great persistency showed by its parent

molecule. Unfortunately, the time series of DACT concentrations is very limited (from 2019 onwards) and the density of points is not ideal; for these reasons a more dense net of sampled wells could help to improve knowledge of the spatial distribution of this contaminant.

One of the main limitations of the study is the knowledge of the stratigraphy of wells; unfortunately a rigourous analysis of the water quality differences between the various layers of the artesian aquifer is not possible, even though some hypotheses can be made observing the limited dataset of known stratigraphies. Is likely that the 3° and 4° aquifers, which are the most productive layers, are more polluted by solvents, and therefore possibly by other contaminants; this phenomena is probably caused by the reduction of piezometric level as a consequence of hevy exploitaion of the layers. An improvement of this study is represented by the possibility to conduct a differentiated analysis of the various aquifer layers.

This thesis also confirms the importance of widespread campaign of water quality sampling in the territory, essential to understand the spatio-temporal evolution of contaminations and therefore to manage actions to improve water quality for human consumption.

9 Bibliography

- Altissimo L., Marcolongo B., Pretto L., Righetto G., Silvestri R. (1990). Carico inquinante degli acquiferi dell'alto vicentino: valutazione della sua potenzialità e riscontro di alcuni valori reali in falda.
- Altissimo L., Passadore G. (2010). Pressioni ambientali esercitate sui corpi idrici sotterranei dagli scarichi di reflui civili ed industriali nell'alto Vicentino. Relazione tecnica vers. 1.
- Calvi G. (2019), Butera I. (relatore). Tesi di laurea: analisi geostatistica di dati di concentrazione nelle acque sotterranee. Corso di laurea in ingegneria civile, politecnico di Torino.
- Cressie N. (1985). Fitting variogram models by weighted least squares. Mathematical Geology 17
- Cressie, N. (2006). Block Kriging for Lognormal Spatial Processes. Journal of Mathematical Geosciences 38
- EEA European Environmental Agency (2018)– Chemicals in European waters, knowledge developments. EEA report no 18/2018.
- Enoch RR, Stanko JP, Greiner SN, Youngblood GL, Rayner JL, Fenton SE (2007). Mammary gland development as a sensitive end point after acute prenatal exposure to an atrazine metabolite mixture in female Long-Evans rats. Environ Health Perspect. 2007
- Favero P. (1997)., Marcomini A. (relatore), Altissimo L., Capodaglio G., Dal Prà A. (correlatori). Tesi di laurea: contaminazione da ione nitrato nelle acque dei fontanili della media pianura veneta. Corso di laurea in scienze ambientali, università Cà Foscari, Venezia.
- Ghirardelli A., Otto S., Masin R., Bano C., Altissimo L., Russo L., Zanin G. (2021). Thirty-year monitoring of s-triazine herbicide contamination in the aquifer north of Vicenza (north-east Italy). Science of The Total Environment, 752
- Gustafson, D.I. (1989) Groundwater Ubiquity Score: A Simple Method for Assessing Pesticide Leachability. Environmental Toxicology and Chemistry, 8
- Hengl T., (2009). A practical guide to Geostatistical mapping.
- Journel A., Huijbrechts C. (1978). Mining Geostatistics. Academic Press, London.
- Juegerns S., Gschwind L., (1989). Ground water nitrates in other developed Countries. Nitrogen management and groundwater protection, R. F. Follett, Amsterdam.
- Kerry, R. and Oliver, M.A. (2007). Determining the Effect of Skewed Data on the Variogram. I. Underlying Asymmetry. Computers & Geosciences.

- Kishne A. (2003). Comparison of Ordinary and lognormal kriging on skewed data of total cadmium in forest soils of Sweden. Environmental monitoring and assessment (July 2003)
- Lewis, K.A., Tzilivakis, J., Warner, D. and Green, A. (2016) An international database for pesticide risk assessments and management. Human and Ecological Risk Assessment: An International Journal
- Lowrance R., (1989). Developments in agricultural and managed forest ecology.
- Otto, S., Altissimo, L., Zanin, G., (2007). Terbuthylazine Contamination of the Aquifer North of Vicenza (North-East Italy).
- Pagello A., Perdon G., (1984). Indagine geologica sulle vie e modalità di diffusione nel sottosuolo dell'alta pianura vicentina di un inquinamento da PCE iniziato in località "Tintoria Astico" in comune di Fara Vicentino.
- Price (2015). Herbicides, Physiology of action and safety
- Rinaldo A., Altissimo L., Putti M., Passadore G., Monego M, Sartori M. (2007). Bacino del Bacchiglione: studi e ricerche idrologiche finalizzati alla messa a punto di modelli matematici per la tutela e la gestione delle risorse idriche sotterranee.
- Ronald Ney (1995). Fate and transport of Organic chemicals in the environment.
- Sbriscia Fioretti, C., Zanin, G., Ferrario, P., Vighi, M., 1998. Chemicals characteristics: the case of herbicides in Italy, in: Swanson, T., Vighi, M. (Eds.), Regulating Chemical Accumulation in the Environment. Cambridge University Press, Cambridge
- Schwille (1984). Migration of organic fluid immiscibile with water in the unsaturated zone.
- SINERGEO (2019). Studio geologico ed idrogeologico dell'area Moracchino per la realizzazione di un nuovo pozzo acquedottistico. Relazione tecnica ed idrogeologica. Committente: Viacqua s.p.a
- Sottani N., Pretto L., Marcolongo B., Viero C. (1982). Gli acquiferi nella pianura a nord di Vicenza: studio del sistema, bilancio idrico e proposte gestionali. A.I.M Vicenza, C.N.R Roma
- Webster, R. and Oliver, M. A.: 2001, Geostatistics for Environmental Scientists, J. Wiley & Sons, Ltd., Chichester, U.K
- WHO (1996). Tetrachloroethene in Drinking-water. Background document for development of WHO Guidelines for Drinking-water Quality
- WHO (2005). Trichlorethene in Drinking-water. Background document for development of WHO Guidelines for Drinking-water Quality

- WHO (2010). Atrazine and Its Metabolites in Drinking-water. Background document for development of WHO Guidelines for Drinking-water Quality
- YAMAMOTO, J.K. (2007). On unbiased backtransform of lognormal kriging estimates. Computers & Geosciences, v. 11, p. 219-234.
- Zanin, G., Borin, M., Altissimo, L., Calamari, D., (1993). Simulation of herbicide contamination of the aquifer north of Vicenza (North-East Italy). Chemosphere 26