WASABY

Project WASABY: WAter and Soil contamination and Awareness on Breast cancer risk in Young women

2nd Steering Committee

-

16.12.2020

5a

LITERATURE REVIEW on articles studying relation between water & soil pollution and breast cancer risk

Literature review on the main persistent environmental contaminants related to breast cancer

Aims

One of the **main objectives of WP7** is **to analyze the relationship between environmental data and the incidence of breast cancer** focusing on the following aim: identify, through a scientific literature review, persistent contaminants in the aqueous matrix (especially in deep waters) , assessing environmental risk factors and exposure to contaminants, in relation to breast cancer.

Several papers have been found to correlate exposure to organic contaminants - mainly but not only POPs (Persistent Organic Pollutants) – detected in the two environmental matrices (water and soil) and the risk of breast cancer for women.

Literature review on the main persistent environmental contaminants related to breast cancer

Wasaby literature results

We found different scientific articles regarding population studies (case-control, cohort, etc.), which revealed an association between exposure to the main persistent contaminants (mainly POPs) and the risk of breast cancer.

We researched and analyzed **11 IARC monographs** on specific contaminants (PCBs, DDT, dioxins, atrazines, PAHs, PFOA, cadmium, trihalomethanes and its metabolites).

We also reviewed **24 scientific reviews**, for a total of **130 scientific study articles.**

WASABY

Deliverable concluded on 31st March 2019

Tab. 1 – Scientific Articles researched for different persistent pollutant family and breast cancer risk.

Contaminants	number of scientific articles	NO ASSOCIATION scientific articles	POSITIVE scientific articles
PCBs (209 congeners)	38	20	18
DDT, DDD, DDE and principal Organachlorines Compounds	40	33	7
Dioxins (TCDD)	13	8	5
PAHs	14	8	6
PFAA (PFOS, PFOSA and PFOA)	6	3	3
Triazine	8	7	1
Cadmium (Heavy Metal)	6	3	3
Trihalomethanes (THMs)	5	2	3
TOTAL Scientific Articles	130	84	46

Source: http://wasabysite.it/wp7_list.html

Literature review on the main persistent environmental contaminants related to breast cancer

ANNEX 5 – References Perfluoroalkyl substances (PFASs, PFOS, PFOSA and PFOA) articles.

N°	Name of researchers / scientific journal	Place and study design	Exposure contaminants	Exposure assessment procedure	Time of sample collecting	Years of the Study interest	Exposure category or level / Risk estimate (95% CI)	Relevant Comment	Methods and possible covariates
1	**Bonefeld-Jorgensen EC**, Long M, Bossi R, et al. **Environmental Health**, 10:88, 2011.	Case control study. **Greenland**	PFOA, PFOS	Blood Samples	After diagnosis	2000 - 2003	PFOS (ng/ml) OR = 1.03 (1.001; 1.07) *p = 0,05*	A case–control study of Greenland Inuit women found a positive statistically significant association between PFOS exposure and breast cancer but not statistically significant association between PFOA exposure and breast cancer.	Age, BMI, pregnancies, and cotinine;
2	**Ghisari M**, Eiberg H, Long M, et al. **Environ Health**. Mar 16;13(1):19. 2014.	Case–control study. Greenlandic Inuit women **North Canadian**	PFOA, PFOS	Blood Samples	After diagnosis	2000 - 2003	Furthermore, an increased BC risk was observed for women with high serum levels of perfluorooctane sulfonate (PFOS) and perfluorooctanoic acid (PFOA)	The BRCA1 founder mutation and polymorphisms in CYP1A1 (Val) and CYP17 (A1) can increase the BC risk among Inuit women and the risk increases with higher serum levels of PFOS and PFOA	Age, BMI, smoking, menopause status, information on serum levels of PFOA / PFOS.
3	**Bonefeld-Jorgensen EC**, Long M, Fredslund SO, Bet al. **Cancer Causes Control**. 25(11):1439-48. 2014.	Case-cohort study. **Denmark.**	PFOA, PFOS	Blood Samples	15 years Before the diagnosis	1996-2002	PFOSA in the 5th quintile RR=2.40, 95% CI = 1.20, 4.83	The results of this study suggest an association between BC and the measured PFOSA	Age, BMI, smoking, menopause status, information on serum levels of PFOA / PFOS. Questionnaires on lifestyle and environmental exposure (including diet, height, weight, diseases in the family, smoking, and alcohol intake)

WASABY

WASABY Starting indicators

Specific Objective	Review the evidence of polluting agents and their impact on breast cancer	
Process Indicator(s)	**Target**	**Result**
Number of published scientific articles considered	>=50	130
Output Indicator(s)	**Target**	**Result**
Number of water/soil pollutants studied in relation with breast cancer	>=15	40
Outcome/Impact Indicator(s)	**Target**	**Result**
Number of reports on literature review	>=1	1

5b

ANALYSIS of available databases with water & soil pollutants (Waterbase, FOREGS, Lucas, E-PRTR, IPCHEM)

Environmental Database – WP 7.2

Aims

The specific aim of the Wasaby report D7.2 **is to evaluate the availability and the usefulness of environmental electronic databases**, to estimate the exposure of people to pollutants, in this case young women.

WP7.2 is focused on the identification of specific databases useful for epidemiological analysis containing European environmental monitoring data of the main environmental contaminants that persist in water and soil matrices, in particular POPs and heavy metals.

Deliverable concluded on 31st October 2019

Tab. 1 – Principal Environmental Databases

Name	Argument	Web address	Organization	Countries* included	Years covered
Waterbase - Water Quality	Water quality data	https://www.eea.europa.eu/data-and-maps/data/waterbase-water-quality-2	EEA's databases	38	2000 - 2016
FOREGS Geochemical Atlas of Europe	Soil quality data	http://weppi.gtk.fi/publ/foregsatlas/	FOREGS	26	1998 - 2002
LUCAS TOPSOIL	Soil quality data	https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data	JRC	28	2009-2012
E-PRTR	Emission Data of Industrial sites	https://prtr.eea.europa.eu/#/home	EEA's databases	33	2007-2017
IPCHEM Information Platform for Chemical Monitoring	Air, Water, soil, quality data	https://ipchem.jrc.ec.europa.eu/RDSIdiscovery/ipchem/index.html	Different EEA's databases and other databases	38	1980- 2017

*All the Wasaby countries (France, Germany, Italy, Lithuania, Poland, Portugal, Spain, Northern Ireland, Slovenia) are covered by the different databases.

Source:

http://wasabysite.it/material/D7.2_WASABY_DB_Report_for_spatial_analysis.pdf

Environmental database: water quality

- Waterbase water quality: 56464 geo-coded water quality monitoring sites in Europe (water wells).
- 420 different measured parameters.
- Collection of 2000-2012 data.

Environmental database: water quality

Parameter Water Body Category: ground water, river water, lake water

CAS number: a unique numerical identifier assigned to every chemical substance

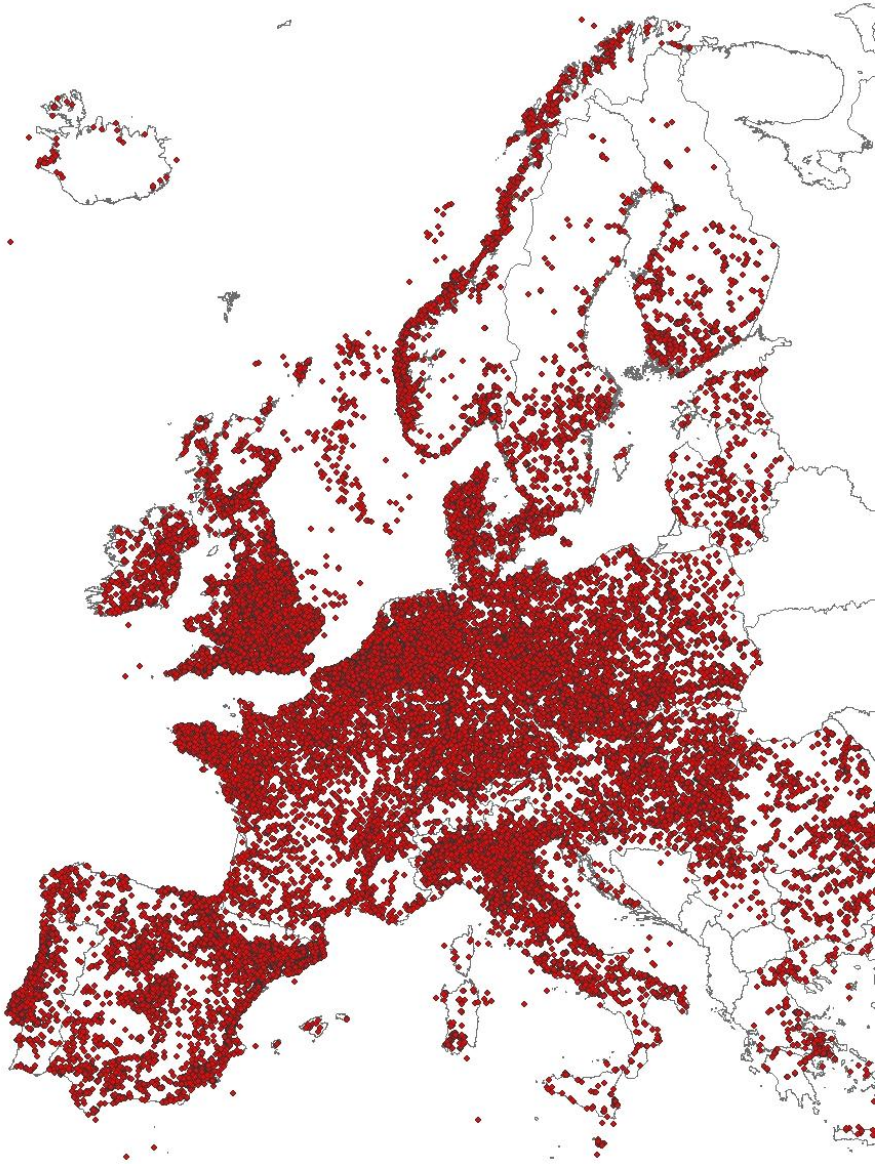
Monitoring Site Identifier

monitoringSite	parameterV	observedPropertyDeter	resultUom	resultObser
SKS001204	GW	CAS_75-27-4	ug/L	0,2
SKS001204	GW	CAS_75-09-2	ug/L	0,2
SKS001204	GW	CAS_87-68-3	ug/L	0,2
SKS001204	GW	CAS_127-18-4	ug/L	0,2
SKS001204	GW	CAS_56-23-5	ug/L	0,2
SKS001204	GW	CAS_79-01-6	ug/L	0,2
SKS001204	GW	CAS_75-01-4	ug/L	0,2
SKS001204	GW	CAS_67-66-3	ug/L	0,2
UKSC017156	GW	EEA_3161-02-2	mg{N}/L	0,148
UKSC017156	GW	CAS_7440-43-9	ug/L	0,015
UKSC017156	GW	CAS_7439-92-1	ug/L	0,35
UKSC017156	GW	CAS_7440-66-6	ug/L	1,1
UKSC012903	GW	CAS_14797-55-8	mg{NO3}/L	0,148
UKSC012903	GW	EEA_3161-02-2	mg{N}/L	0,148
IT06TS04	RW	CAS_7440-66-6	ug/L	50000
IT06TS07	RW	CAS_7440-66-6	ug/L	50000
IT06TS06	RW	CAS_7440-66-6	ug/L	50000
IT06TS04	RW	CAS_7440-66-6	ug/L	50000
IT06TS04	RW	CAS_7440-66-6	ug/L	50000
IT06TS06	RW	CAS_7440-66-6	ug/L	50000
IT06UD63	RW	CAS_1582-09-8	ug/L	3300
UKSC016454	GW	CAS_7439-92-1	ug/L	0,2
IT05143	RW	EEA_32-26-8	ug/L	0,86
IT020031VA1	RW	CAS_107-06-2	ug/L	0,05
IT020031VA1	RW	CAS_107-06-2	ug/L	0,05

Record: 25 di 1048575 | Non filtrato | Cerca

Environmental database: emission from production

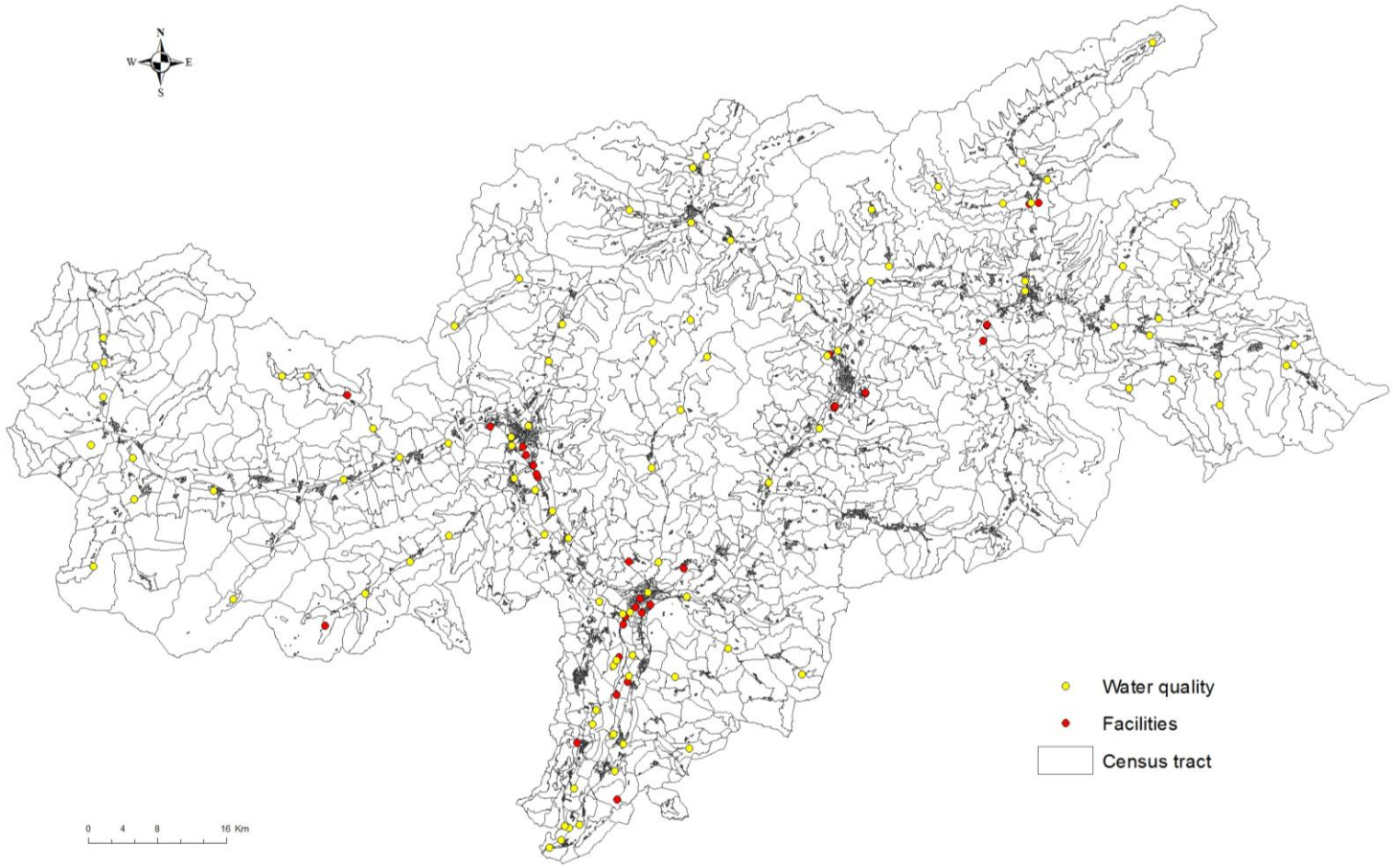
- E-PRTR: pollutants emission data from industrial sites in Europe by productive sector.



Environmental database: linkage with CR data

- Using ArcGIS software we performed a spatial join between water quality data from waterbase database, emission pollutants data from E-PRTR database and Census Tract layer of every CR area.
- The record linkage variable was the Census Tract (or the equivalent geographic unit).
- By this procedure, the information from every CR and the corresponding socio-economic information were linked to the potential pollution/emission sources in the same area.
- Check the summarizing Excel file at **http://www.wasabysite.it/material/CR_Reports.rar** for the already linked areas.

Environmental database - Example of emission sources distribution in Alto Adige CR area



Alto Adige Cancer Registry - Italy

Next Step - The Pilot Environmental Study

• The area/s interested by the pilot environmental study should provide both water and soil pollutants; therefore an environmental monitoring of the area must be present and the corresponding data available and affordable.

• A first area has been individuated in Italy: the Varese province, why ?

• With almost 900,000 inhabitants, it is one of the most populous provinces of Lombardy making it the fifth province in Italy for population density.

From an orographic point of view, the province of Varese is very varied and is divided into at least three sectors:

1) the mountain portion, the northern area of Varese, formed by reliefs above 600 mts. a.s.l., occupies 32% of the territory;
2) The hilly area (altitude between 200 mts. a.s.l. and 600 mts. a.s.l.) occupies the central area and constitutes 46% of the territory;
3) The plain (altitude below 200 m s.l.m.), which extends only to the extreme south of the province and constitutes 22 % of the territory.

Next Step - The Pilot Environmental Study

The Province of Varese is also an area geographically rich in lakes (such as Lake Maggiore, Varese, Comabbio, Monate, Delio, Ganna, Ghirla) and rivers (Ticino, Olona and Tresa) and in parks in the pre-Alpine area.
It also presents in its territory from north to south different categories and sectors of large production activities, ranging from the mechanical sector to that of plastics, to the textile and energy, chemical, pharmaceutical sectors, also the presence of landfills, cement factories and an incinerator and the main Italian international airport (Malpensa) are present.

In Summary
This area seems to have all the characteristics needed to be the pilot area for experiencing the model of analyses to be shared with the community of the Cancer Registries:

- a long-lasting Cancer Registries with high-quality geo-coded data, covering all the province;
- an important distribution and differentiation of productive activities in a densely populated area, with a well-differentiated orographic;
- the possibility of requesting local environmental databases and local monitoring of the pollution and its sources through the Regional Agency for the Protection of the Environment of Lombardy (ARPA Lombardy) available for research purposes.

WASABY Starting indicators

Specific Objective	Environmental study on correlation between soil and water contamination and breast cancer risk	
Process Indicator(s)	**Target**	**Result**
Number of European databases with available data on water/soil pollutants	>=2	2
Number of national databases with available data on water/soil pollutants	>=3	4
Number of pollutant indicators available in different countries	>=5	40

6a

LITERATURE REVIEW on methods to study relation between water & soil (W&S) pollution and outcomes

The methodological literature review

The combined work of WP4 and WP7 produced a literature review about the methods to evaluate the effects of soil and water pollution on cancer outcomes at different territorial level (small areas, individuals, etc.), considering the socio-economic characteristics.

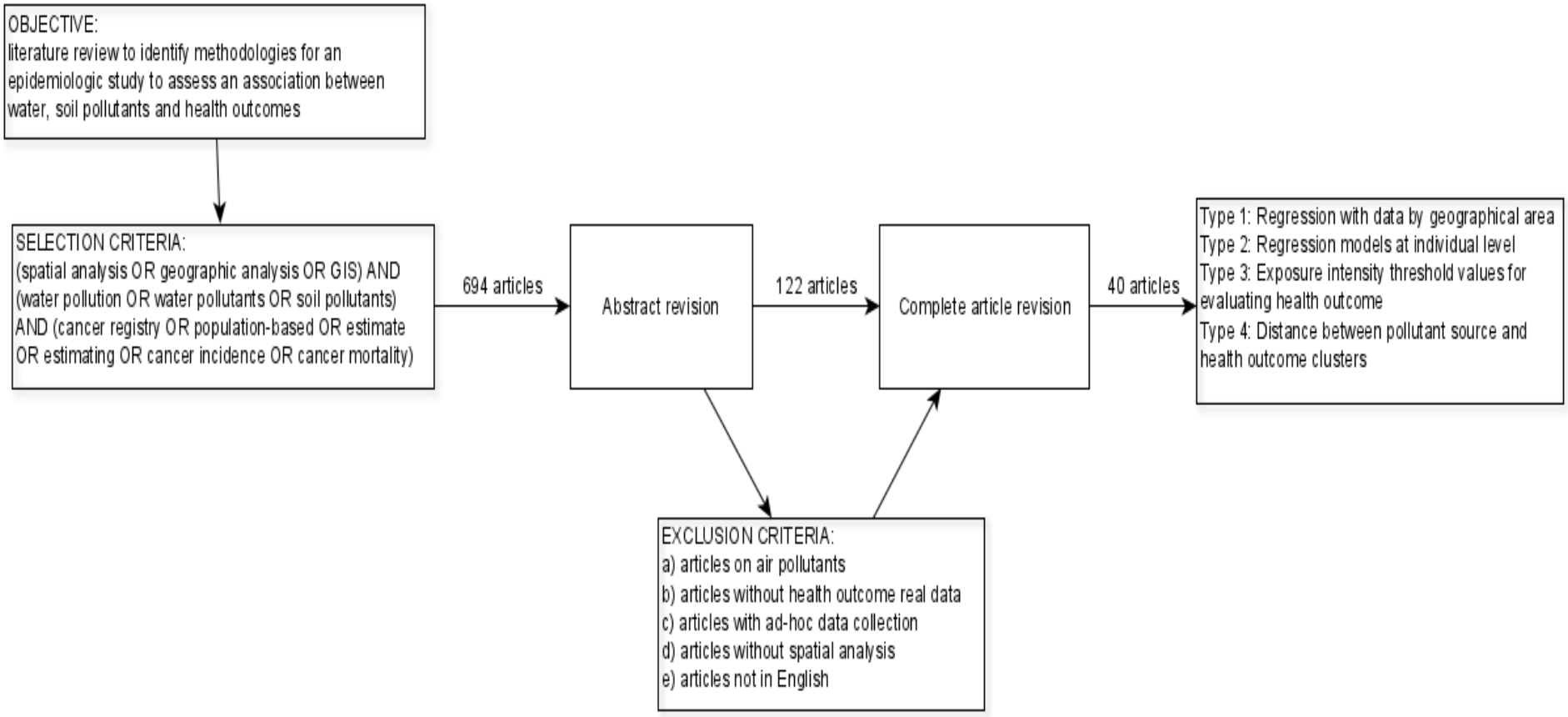
After various steps of evaluation, 40 articles were accepted for review and classified in 4 categories, according to the methods of analysis.

The article was accepted for publication on the Reviews of Environmental Contamination and Toxicology (Impact Factor: 5.767).

"**Water and soil pollution: ecological environmental study methodologies useful for public health projects. A literature review**"

WASABY

The methodological literature review



WASABY

Article accepted for publication (not yet received proof)

	Environmental factor data	Health outcome data	Analysis	Number of articles
Type 1	Data by geographical areas	Data by geographical areas	Regression models using data by geographical areas	20
Type 2	Data at individual level	Data at individual level	Regression models using data at individual level	4
Type 3	Data by geographical areas	Data by geographical areas	Threshold values for exposure intensity are computed, in order to define cut-off points for evaluating trends in the health outcome variable influenced by the environmental factor	9
Type 4	Environmental pollution geographic clusters obtained by considering environmental factors and their potential emission sources	Clusters of areas or people generated by the analysis of the considered health outcomes	The two different kinds of clusters were identified separately. Comparisons between health outcomes geographic clusters and environmental pollution geographic clusters by considering the distance between them	7

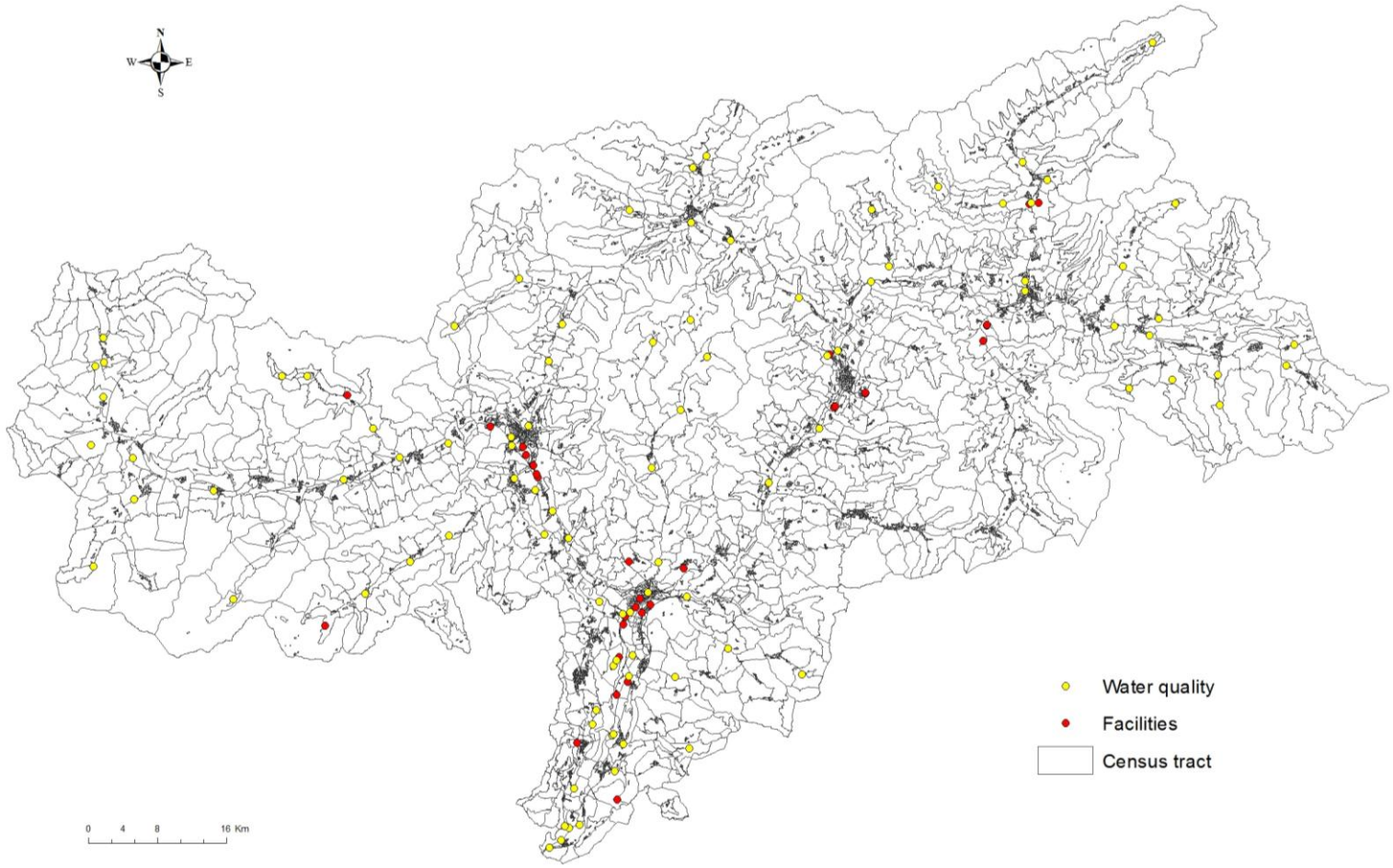
The article was accepted for publication on the Reviews of Environmental Contamination and Toxicology

6b

Correlation between clusters and pollution sources

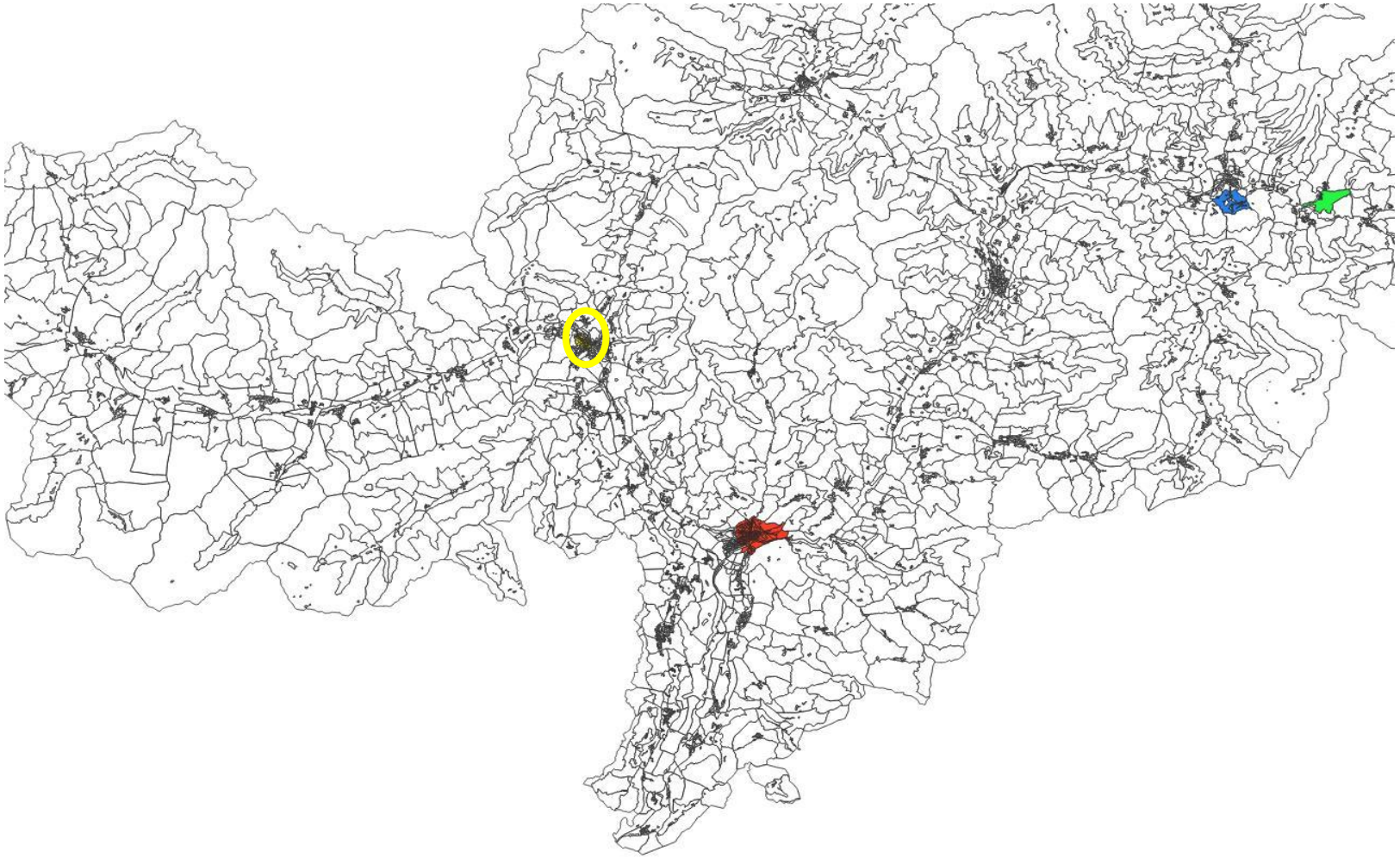
Summary for each cancer registry

Environmental database - Example of emission sources distribution in Alto Adige CR area



Alto Adige Cancer Registry - Italy

ALREADY OBTAINED RESULTS - EXAMPLES ON ALTO ADIGE CR
SatScan clusters



WASABY

Connection beteween clusters, industrial facilities and water wells

In order to evaluate the probability of connection between water wells and facilities with their positioning in the cluster, a $\chi 2$ test ($p<0.05$) was performed between these geo-referenced by CTs sources and their belonging or not to a cluster. The result is the following:

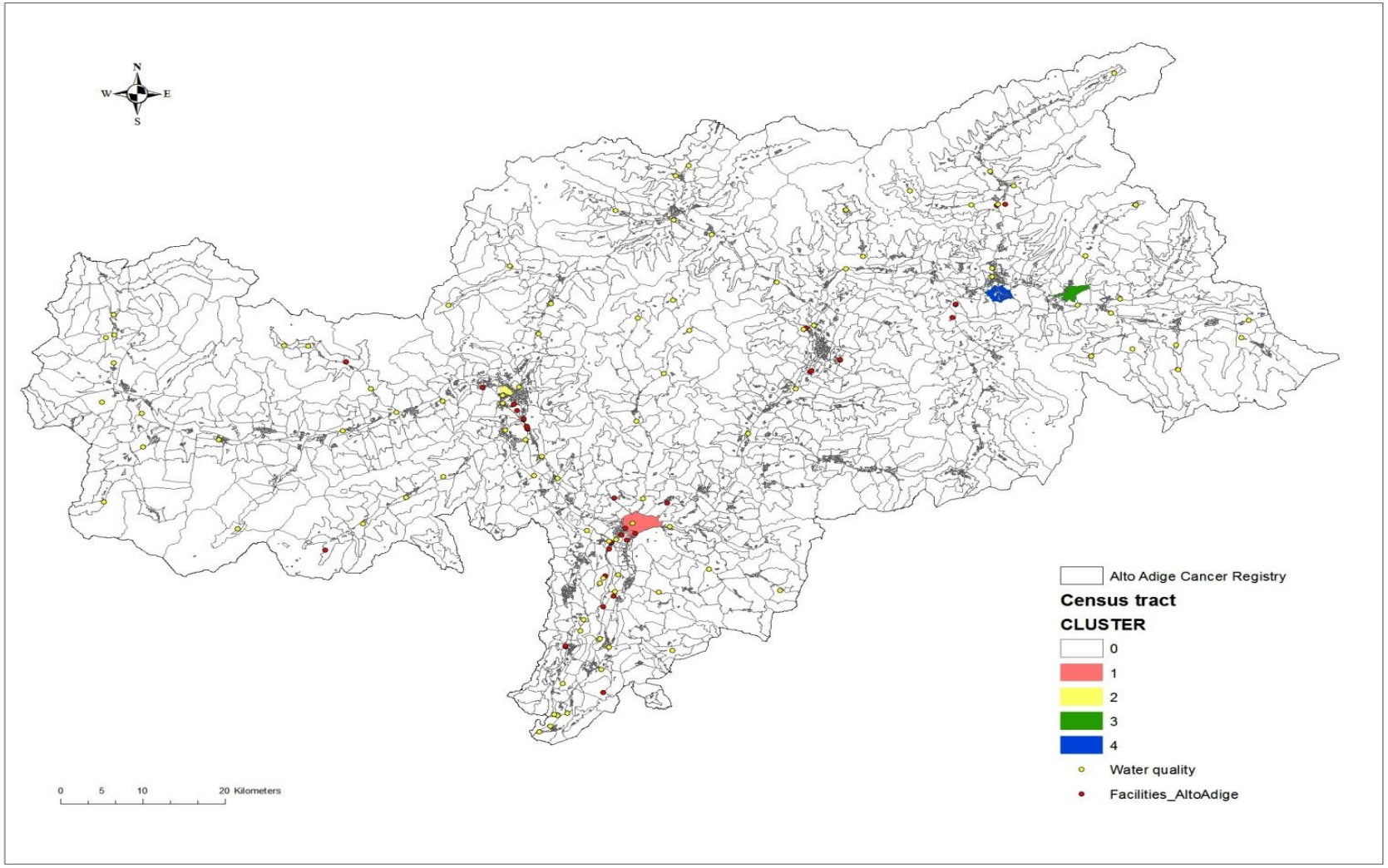
Pearson chi2(3) = 7.2483 **Pr = 0.050 (both standardization)**

There is a borderline statistical significance between the presence of the sources and their belonging to the clusters, therefore further investigation should be performed.

In the following maps, the estimated grouped SIRs by CT and the clusters identified for Alto Adige CR are respectively combined with the industrial and other emitting activities facilities registered by E-RPTR (red points) and the geo-referenced ground water wells from Waterbase Water Quality (yellow points).

WASABY

Connection beteween clusters, industrial facilities and water wells



8. Identification of industrial facilities and water wells

Country	Number of CRs	Analysis performed
France	17	To be defined
Germany	1	Local analyses
Italy	10	10
Lithuania	1	0
Poland	5	Local analyses
Portugal	2	2
Slovenia	1	Local analyses
Spain	6	6
Northern Ireland	1	0
Total	**44**	**18 (+ French CRs)**

WASABY

7

A pilot study on correlation between W&S and BC risk using regression and Kriging interpolation

Interpolation method

A specific method was applied to estimate the distribution of the pollutants on the considered areas, starting from fixed sources of water and pollution.

After an accurate international literature review (in press) on the methods used for such studies from the Seventies to today, the kriging interpolation method was chosen.

More specifically, the Empirical Bayesian Kriging was chosen, due to its advantages and to the peculiarities of the available environmental data.

The considered software were ArcGIS 10.4 and QGIS 3.14.16.

Kriging interpolation method

Kriging is one of several methods that use a limited set of sampled data points to estimate the value of a variable over a continuous spatial field.

It differs from simpler methods, such as Inverse Distance Weighted Interpolation, Linear Regression, or Gaussian decays in that it uses the spatial correlation between sampled points to interpolate the values in the spatial field: the interpolation is based on the spatial arrangement of the empirical observations, rather than on a presumed model of spatial distribution.

Kriging also generates estimates of the uncertainty surrounding each interpolated value.

Kriging interpolation method

In a general sense, the kriging weights are calculated such that points nearby to the location of interest are given more weight than those farther away.

Clustering of points is also taken into account, so that clusters of points are weighted less heavily (in effect, they contain less information than single points). This helps to reduce bias in the predictions.

The kriging predictor is an “optimal linear predictor” and an exact interpolator, meaning that each interpolated value is calculated to minimize the prediction error for that point. The value that is generated from the kriging process for any actually sampled location will be equal to the observed value at this point, and all the interpolated values will be the Best Linear Unbiased Predictors (BLUPs).

Kriging interpolation method

Kriging will in general not be more effective than simpler methods of interpolation if there is little spatial autocorrelation among the sampled data points (that is, if the values do not co-vary in space).

If there is at least moderate spatial autocorrelation, however, kriging can be a helpful method to preserve spatial variability that would be lost using a simpler method.

Due to the peculiarities of the available data in WASABY, the chosen method among the various kriging techniques is the Empirical Bayesian Kriging Interpolation.

Empirical Bayesiam Kriging interpolation method

Empirical Bayesian Kriging (EBK) is a geostatistical interpolation method that automates the most difficult aspects of building a valid kriging model.

Empirical Bayesian Kriging also differs from other kriging methods by accounting for the error introduced by estimating the underlying semivariogram.

Other kriging methods calculate the semivariogram from known data locations and use this single semivariogram to make predictions at unknown locations; this process implicitly assumes that the estimated semivariogram is the true semivariogram for the interpolation region.

By not taking the uncertainty of semivariogram estimation into account, other kriging methods underestimate the standard errors of prediction.

ALREADY OBTAINED RESULTS - EXAMPLES ON ALTO ADIGE CR
Estimated distribution by Empirical Bayesian kriging

As an example of application of the kriging technique, the case of Alto Adige is presented.

We consider the molecules and other elements and characteristics in the water of the province.

The information comes from the various water wells for water inspection existing in the area, as collected by the EEA Water Quality Database.

All the water wells were geo-coded in terms of X&Y coordinates and attributed to the corresponding Census Tract.

According to the examined literature, also the water wells in a 10 kms. buffer were considered for interpolation.

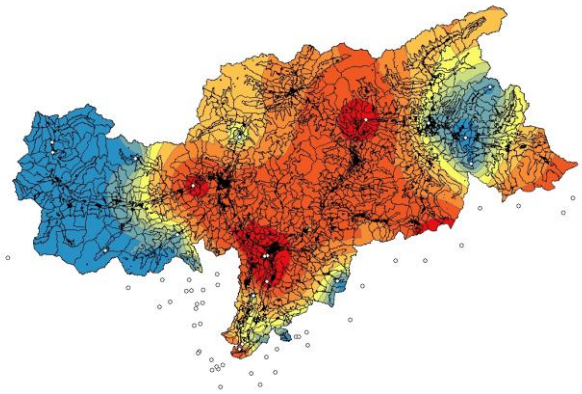
ALREADY OBTAINED RESULTS - EXAMPLES ON ALTO ADIGE CR
Estimated distribution by Empirical Bayesian kriging

Seventy elements were found (e.g., molecules, water ph, etc.) and 15 of them were considered for this example:

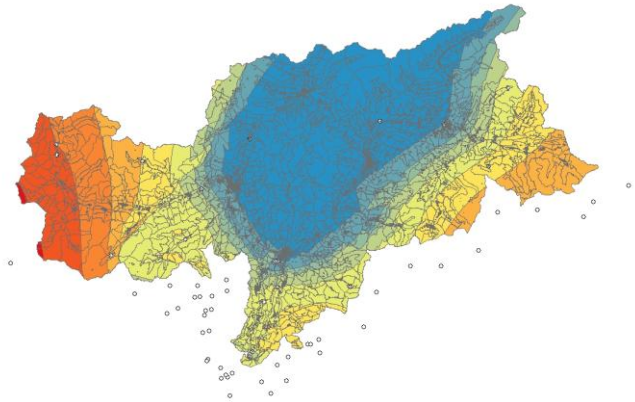
- Aldrin
- Arsenic and its compounds
- Atrazine
- Cadmium and its compounds
- Chlorpyrifos
- Chromium and its compounds
- DDTpp1
- Dieldrin
- Endrin
- Lead and its compounds
- Mercury and its compounds
- Nickel and its compounds
- Simazine
- Trichloroethylene
- Trichloromethane

The following maps report the distribution of the interpolated estimation of the quantity of the 15 elements (in deciles) in the Alto Adige province (that is the CR's area).

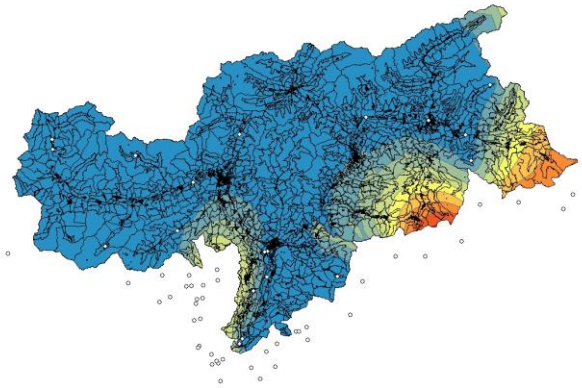
ALREADY OBTAINED RESULTS - EXAMPLES ON ALTO ADIGE CR
Estimated distribution by Empirical Bayesian kriging



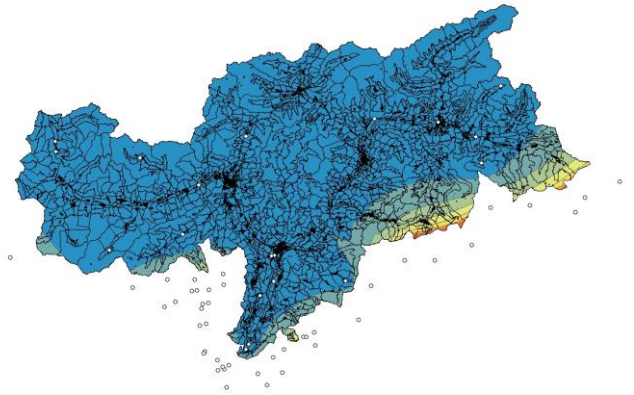
Aldrin



Arsenic and its compounds



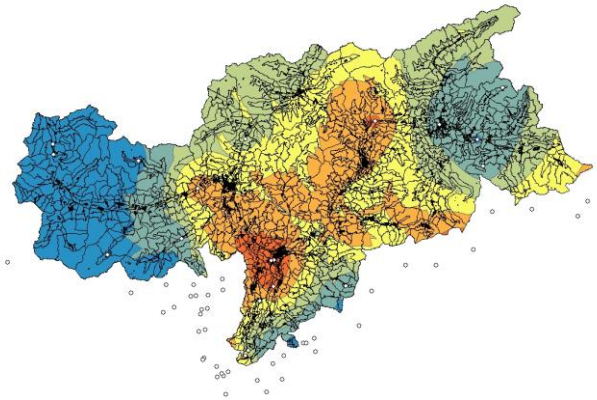
Atrazine



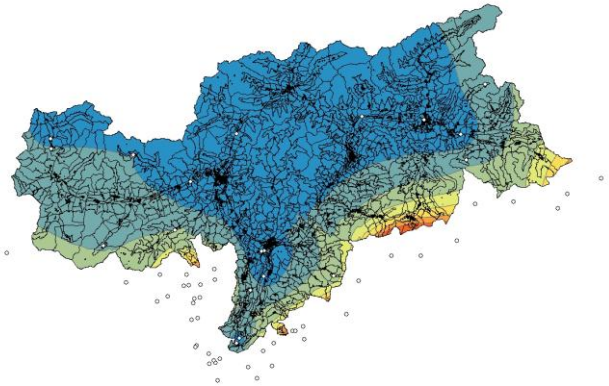
Cadmium and its compounds

WASABY

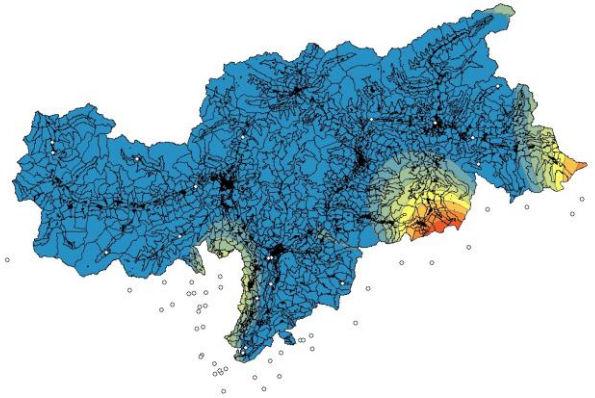
ALREADY OBTAINED RESULTS - EXAMPLES ON ALTO ADIGE CR
Estimated distribution by Empirical Bayesian kriging



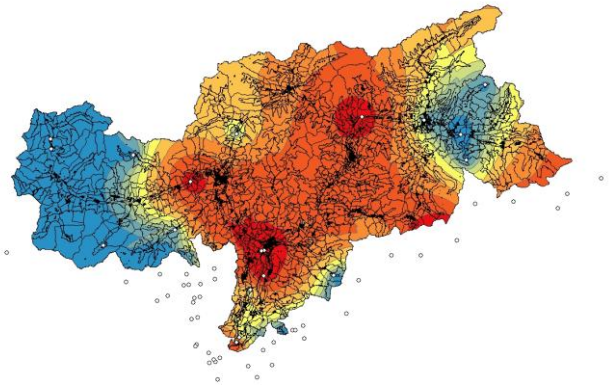
Chlorpyrifos



Chromium and its compounds



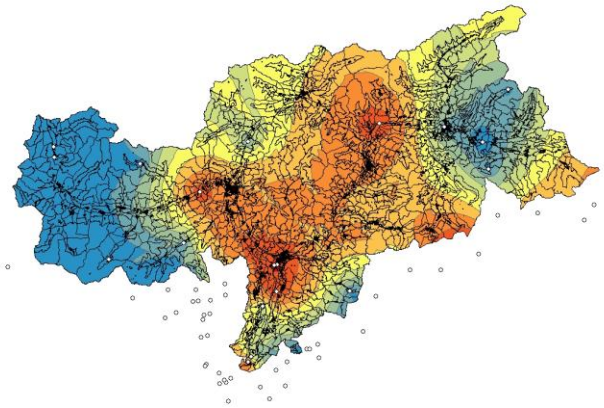
DDTpp1



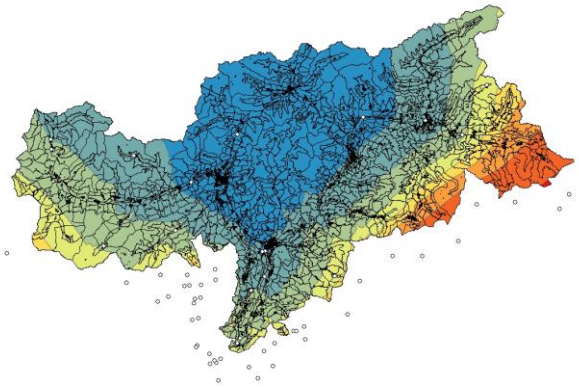
Dieldrin

WASABY

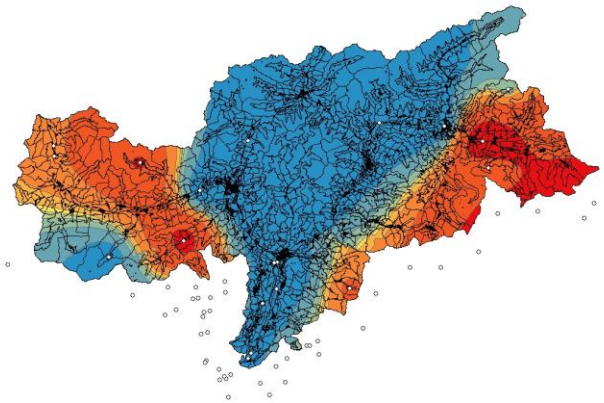
ALREADY OBTAINED RESULTS - EXAMPLES ON ALTO ADIGE CR
Estimated distribution by Empirical Bayesian kriging



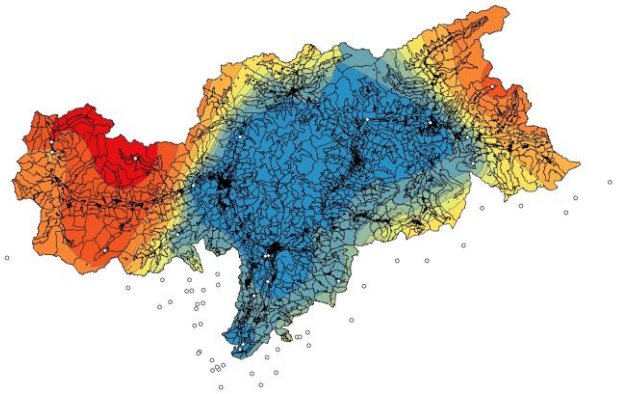
Endrin



Lead and its compounds



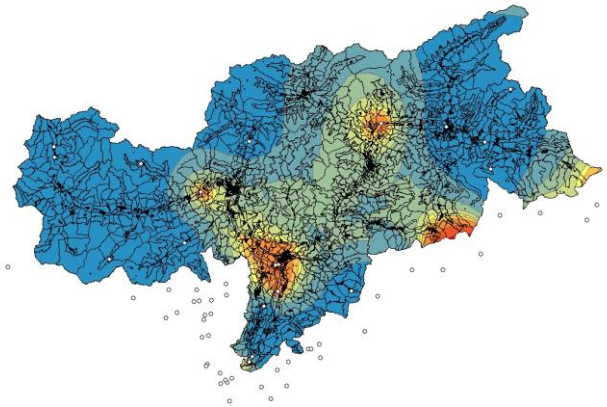
Mercury and its compounds



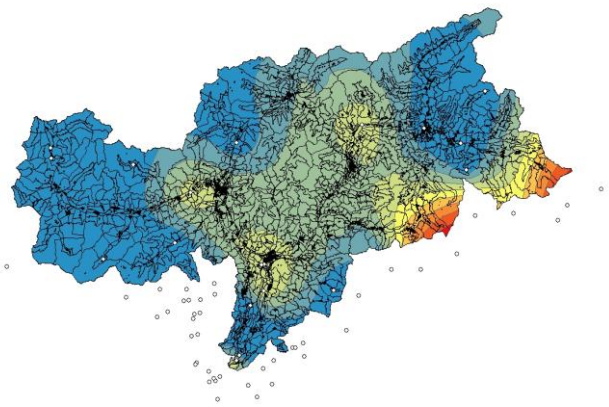
Nickel and its compounds

WASABY

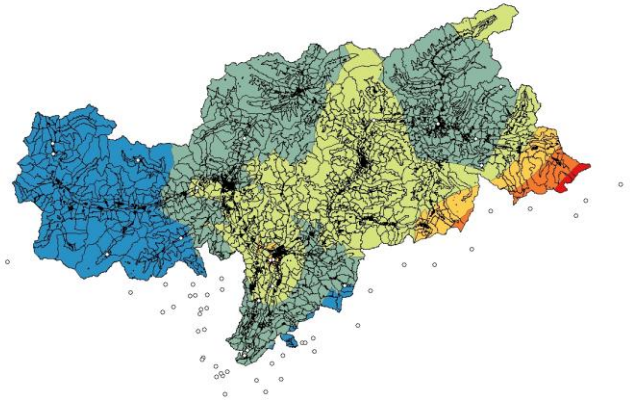
ALREADY OBTAINED RESULTS - EXAMPLES ON ALTO ADIGE CR
Estimated distribution by Empirical Bayesian kriging



Simazine



Trichloroethylene



Trichloromethane

WASABY

ALREADY OBTAINED RESULTS - EXAMPLES ON ALTO ADIGE CR
Connection between pollutant variables and cluster

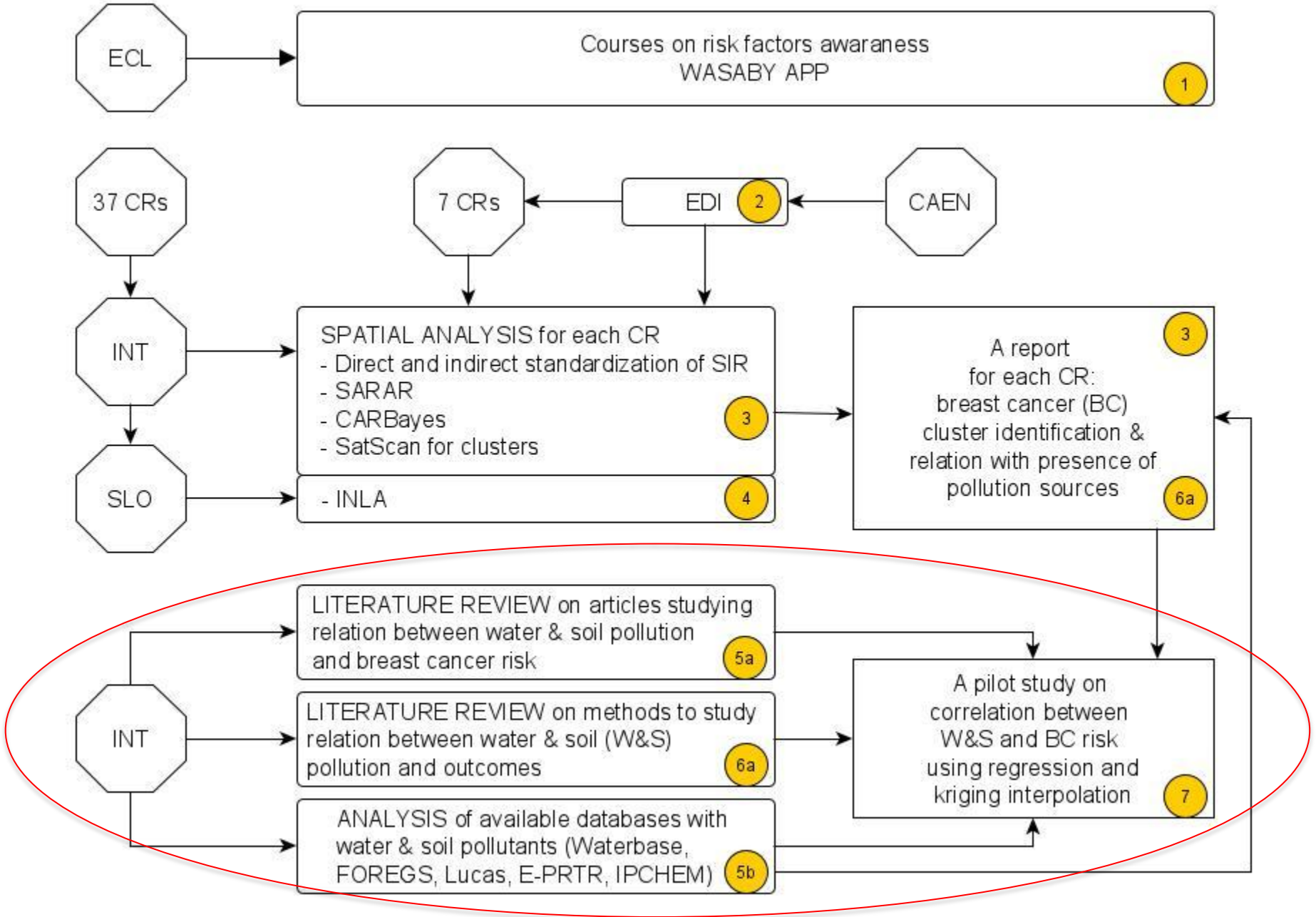
The possible connection between pollutant variables and clusters was evaluated by logistic regression model (Relative Risk Ratios, $p<0.05$).
The table reports the elements statistically significant interacting with the clusters.

CLUSTER	Elements	RRR	*p*	[95% Conf. Interval]	
No cluster		*ref.*			
Cluster 1	Arsenic and its compounds	0.000000434	*0.000*	0.00	0.000024
	Chlorpyrifos	8.60	*0.000*	1.95	78.00
	Endrin	1.04	*0.000*	2.12	5.13
	Simazine	3.26	*0.000*	1.33	32.00
	Trichloroethylene	2.19	*0.000*	1.40	11.90
	Trichloromethane	6.61	*0.000*	3.99	10.90
Cluster 2	Aldrin	1.82	*0.000*	1.24	2.67
	Arsenic and its compounds	0.000000303	*0.000*	0.00	0.000479
	Dieldrin	2.02	*0.000*	1.35	3.01
	Endrin	8.03	*0.000*	3.29	19.60
	Simazine	1.66	*0.000*	1.23	12.10
	Trichloroethylene	7.30	*0.000*	2.31	23.10
	Trichloromethane	0.000000121	*0.000*	0.00	0.0007
Cluster 3	*The clusters 3 and 4 were not statistically significant defined; no connection with elements was found*				
Cluster 4					

WASABY

DISCUSSION

Summary of the present agenda



Next Step - The Pilot Environmental Study

• The area/s interested by the pilot environmental study should provide both water and soil pollutants; therefore an environmental monitoring of the area must be present and the corresponding data available and affordable.

• A first area has been individuated in Italy: the Varese province, why ?

• With almost 900,000 inhabitants, it is one of the most populous provinces of Lombardy making it the fifth province in Italy for population density.

From an orographic point of view, the province of Varese is very varied and is divided into at least three sectors:

1) the mountain portion, the northern area of Varese, formed by reliefs above 600 m a.s.l., occupies 32% of the territory;
2) The hilly area (altitude between 200 m s.l.m. and 600 m s.l.m.) occupies the central area and constitutes 46% of the territory;
3) The plain (altitude below 200 m s.l.m.), which extends only to the extreme south of the province and constitutes 22 % of the territory

Next Step - The Pilot Environmental Study

The Province of Varese is also an area geographically rich in lakes (such as Lake Maggiore, Varese, Comabbio, Monate, Delio, Ganna, Ghirla) and rivers (Ticino, Olona and Tresa) and in parks in the pre-Alpine area.

It also presents in its territory from north to south different categories and sectors of large production activities, ranging from the mechanical sector to that of plastics, to the textile and energy, chemical, pharmaceutical sectors, with the presence of landfills, cement factories and an incinerator.

In Summary

This area seems to have all the characteristics needed to be the pilot area for experiencing the model of analyses to be shared with the community of the Cancer Registries:

- a long-lasting Cancer Registries with high-quality geo-coded data, covering all the province;
- an important distribution and differentiation of productive activities in a densely populated area, with a well-differentiated orographic;
- the possibility of requesting local environmental databases and local monitoring of the pollution and its sources through the Regional Agency for the Protection of the Environment of Lombardy (ARPA Lombardy) available for research purposes.

Points for discussion

- Decision of the pilot study area
  - Involvement of experts of the area
  - Availability of water & soil pollutants
  - Presence of clusters
  - Connection with local agencies on environmental protection