1. Introduction

1.1. Brief overview of methodology

The study deals with the assessment of regional groundwater vulnerability coupling the Weights of Evidence model (Bonham-Carter G.F. et al., 1989; Bates L.E. et al., 1996) with geographical information system. A series of predictor maps, representing the spatial distribution of factors influencing the groundwater vulnerability, are derived for the investigated area and a statistic-probabilistic methodology is adopted and implemented in a GIS for the analysis of the relationships between available data sets. The spatial data analysis for groundwater vulnerability consists of the following steps:

♦ Data collection;
♦ Pre-processing of available data;
♦ Construction of GIS database containing all causal factors (classified themes and other supporting information);
♦ Construction of the prediction model, including the creation of unique condition subareas, computation of probability tables, construction of prediction maps and representation of preliminary prediction results. The model expresses the relative susceptibility of groundwater to contamination in terms of probabilities and combines them by the bayesian combination formulas under the conditional independence assumption.
♦ Validation of prediction results (time robustness and space robustness) by the comparison with spatial distribution of nitrate in different years in the area;
♦ Visualization of prediction results.

Potential problems related to conditional independence of the predictor patterns with respect to the occurrence of the response event are also considered.

1.2. Description of selected area

The study area is about 1989 Km$^2$ and is located in the district of Milan (Northern Italy) where both agricultural and industrial activities are extensively present. This area has a complex hydro-geological setting with the presence of many aquifers with different degree of mutual interaction. Conditions of superficial recharge are complex too, because of the presence of an extensive irrigation net and a great number of soil and land use types.

The Milan plain subsoil is characterized by Pliocene-Pleistocene sediments, considered remarkable water resources. The upper units of this stratigraphic sequence are of alluvial and/or alluvial-glacial origin (composition ranges from gravels to sands with few layers of silts and clays) whereas the deeper ones (mainly silts and clays) are of sea origin. In the end of Pleistocene the changing of the sedimentation mechanisms, due to the Alpine orogenic processes, promoted the sea regression and created a delta-lagoon environment, characterized by fine sediments alternated with coarse materials. During Quaternary, the rivers flowing out the southern borders of the glacial moraines, deposited the coarse sediments which now constitute the upper units of the stratigraphic sequence. The increasing of clays and silts occurs not only with the depth but also from the north to the south of Milan plain because the quaternary rivers, far from the glacier, progressively lost their transport energy.

In the Milan plain three main aquifers can be distinguished:
The “Traditional Aquifer” (TR): its water has been mainly used for public and industrial supplies since the beginning of XX century. The TR is a unconfined aquifer which has a good transmissivity (from $5 \times 10^{-2}$ to $1 \times 10^{-3}$ m$^2$/s) and permeability (from $5 \times 10^{-3}$ to $1 \times 10^{-4}$ m/s) characteristics. Its thickness ranges from 60 to 120 m. The aquifer is characterized by quaternary sediments: principally gravels and sands although, from north to south, the presence of clay-silt layers increases. From about the middle of Milan city to the south, a shallow unconfined aquifer (called TRa) and a half-confined aquifer (called TRb), separated by a discontinuous aquitard (5-10 m thick) constituted by clayey-silt, can be distinguished. The regional flow goes from north to south.

The “Continental Aquifer” (C): it is located into the Pleistocene sediments of delta-lagoon origin. The separation from the upper aquifer (TR), is given by continuous and thick layers of clay. The aquifer could represent an important supply water resource given the quality of its water but the low permeability and transmissivity limit its exploitation. At a regional scale the Continental Aquifer is considered a unique aquifer even if, at a more detailed scale, it can be subdivided in many different aquifers. The (C) becomes thicker from north to south and its base has a slope of about 1.4 %, greater than the slope of the topographic surface. The water resource of the aquifer is named 2° Falda and flows from north to south.

The “Sea Aquifer” (M): it’s the deeper one; in fact, in Milan its top can be found between - 230 and - 300 m from the topographic surface. For this reason, few wells reach and exploit it, so there are few information about its hydrogeological parameters.

The study only focuses on the vulnerability of the unconfined aquifer (TR/TRa) which is the one most affected by the contaminant load coming from the surface.

### 1.3. Assumption and limitations

1. All maps, final one included, must always be analyzed and interpreted in relation to the scale of the input data. So these maps should always be considered as useful tool for the understanding of the importance of factors influencing the vulnerability at a regional scale and for the identification of the areas that need further and more detailed investigations. They have not to be directly applied to small localized areas where vulnerability can greatly be affected by local changes in some parameters. This consideration is particularly relevant in each GIS application because of the ease with which maps can be produced at customized scale, regardless of the detail of the original data (original capture scale).

2. Even if input data have been critically reviewed and accurately selected in order to obtain a homogenous spatial distribution for all of them, some parameters are inevitably concentrated in some areas, especially urban areas.

3. Another important point is the choice of the response variable. From an ideal point of view this should be represented by a chemical compound which is ubiquitously present in all the area in relation to homogeneously distributed dispersed sources, that is actually impossible. This fact will be analyzed further in Chapter 4.

### 2. Choice of predictors

The most important factors affecting the vulnerability have been collected and analyzed to derive predictor maps, trying to define them as much as possible in a quantitative way. Only type of soils and land use has been expressed qualitatively, but using a simple and a worldwide known classification, avoiding subjective interpretations.

At the end, the chosen predictor maps has been:
1. rainfall;
2. irrigation;
3. type of soil;
4. land use;
5. hydraulic conductivity of vadose zone;
6. groundwater depth;
7. groundwater velocity;

The predictor maps derived from selected data have been classified and converted into binary form to produce a series of mutually exclusive map patterns for each predictor variable.

Evaluation of factors and the determination of their spatial distribution is described in detail in the following paragraphs.

### 2.1. Rainfall

Data used for the construction of a rainfall map have been taken from the Hydrographic Service of Po River. Data have been checked to identify stations with record longer than twenty years, and mean annual data have been calculated as input for the map.

Rainfall stations and their mean annual value are reported in table 2.1.1.

Isohyets, (figure 2.1.1.) show a decreasing trend in rainfall height from NW to SE according with altimetrical changes in the same direction in the area. In fact a very localized maximum of 1250 mm/y can be observed at the NW boundary of the area and a minimum of 900 mm/y at the SE boundary, with a difference in percentage of 25%.

<table>
<thead>
<tr>
<th>Station</th>
<th>Mean annual rainfall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Busto Arsizio</td>
<td>1246 mm</td>
</tr>
<tr>
<td>Bereguardo</td>
<td>780 mm</td>
</tr>
<tr>
<td>Cernusco</td>
<td>1014 mm</td>
</tr>
<tr>
<td>Milano</td>
<td>1085 mm</td>
</tr>
<tr>
<td>Miorina</td>
<td>1351 mm</td>
</tr>
<tr>
<td>Paullo</td>
<td>980 mm</td>
</tr>
<tr>
<td>Treviglio</td>
<td>936 mm</td>
</tr>
</tbody>
</table>

*Table 2.1.1.: Mean annual rainfall at the selected recording stations*

*Figure 2.1.1.: Mean annual precipitation*
2.2. Irrigation

More than half of the investigated area (1231 Km² of 1989 Km²) is interested by an extensive irrigation net. This net is mainly located in the central-southern part of the district; the limit which divides it from the northern area is given by the northern border of the Villoresi Canal Consortium, which crosses from west to east the investigated area, going from the Ticino to the Adda river.

Irrigation plays an important role in the recharge of groundwater, increasing very much the amount of water that can infiltrate in soils. Results derived from a numerical model of groundwater flow in the district of Milan (ALBERTI & al. 2000) show that irrigation has a great influence on the seasonal changes of piezometric levels.

Calculation of the mean annual irrigation has been done from the average incoming and outcoming discharge recorded by each Consortium which regulates the availability of water in all the canals in the area. The main characteristics of canals are summarized in table 2.2.1. and their spatial distribution is represented in figure 2.2.1.

<table>
<thead>
<tr>
<th>Main canals</th>
<th>Secondary canals</th>
<th>Code for figure 2.2.1</th>
<th>Area (m²)</th>
<th>Sum of average monthly discharge (m³/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VILLORESI</td>
<td>Castano I and</td>
<td>1</td>
<td>20.169.473</td>
<td>5,085</td>
</tr>
<tr>
<td></td>
<td>Castano II</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cuggione</td>
<td>2</td>
<td>27.031.155</td>
<td>17,795</td>
</tr>
<tr>
<td></td>
<td>Magenta</td>
<td>3</td>
<td>434.081.848</td>
<td>21,268</td>
</tr>
<tr>
<td></td>
<td>Corbetta</td>
<td>4</td>
<td>60.964.658</td>
<td>25,210</td>
</tr>
<tr>
<td></td>
<td>Parabiago 5+6+7+8+9</td>
<td></td>
<td>18.207.000</td>
<td>68,959</td>
</tr>
<tr>
<td></td>
<td>Rho</td>
<td>10</td>
<td>16.877.000</td>
<td>1,229</td>
</tr>
<tr>
<td></td>
<td>Passirana and Arese</td>
<td>11</td>
<td>55.237.734</td>
<td>9,205</td>
</tr>
<tr>
<td></td>
<td>Garbagnate</td>
<td>12</td>
<td>31.482.300</td>
<td>2,867</td>
</tr>
<tr>
<td></td>
<td>Val Seveso</td>
<td>13</td>
<td>35.234.100</td>
<td>2,319</td>
</tr>
<tr>
<td></td>
<td>Nova and Cinisello</td>
<td>14</td>
<td>31.011.100</td>
<td>0,654</td>
</tr>
<tr>
<td></td>
<td>Brugherio</td>
<td>16</td>
<td>18.921.500</td>
<td>1,605</td>
</tr>
<tr>
<td></td>
<td>Carugate and Cernusco</td>
<td>17</td>
<td>27.653.617</td>
<td>7,914</td>
</tr>
<tr>
<td></td>
<td>Gorgonzola and Pessano</td>
<td>18</td>
<td>26.348.500</td>
<td>3,472</td>
</tr>
<tr>
<td>MARTESANA</td>
<td>19+20+21</td>
<td>224.696.396</td>
<td>148,010</td>
<td></td>
</tr>
<tr>
<td>NAV. GRANDE</td>
<td>22</td>
<td>778.822.422</td>
<td>218,776</td>
<td></td>
</tr>
<tr>
<td>NAV. PAVIA</td>
<td>23</td>
<td>324.400.000</td>
<td>59,957</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2.1.: Main characteristics of canals.
2.3. Type of soil

Soil types have been determined on the basis of Curve Number (CN) classification (USDA-SCS 1972). The choice of this classification is based on the following considerations:

a) it’s worldwide known for its good reliability;
b) it can be easily used both for small and large areas;
c) it fits very well with experimental and field data, normally available.

According to CN model soils can be grouped in four classes (tab. 2.3.1.) with different hydrologic behavior, according to their infiltration capability.

<table>
<thead>
<tr>
<th>HYDROLOGIC GROUP</th>
<th>SOIL DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Soils with high infiltration capacities, even when thoroughly wetted</td>
</tr>
<tr>
<td>B</td>
<td>Soils with moderate infiltration rates when thoroughly wetted</td>
</tr>
<tr>
<td>C</td>
<td>Soils with slow infiltration rates when thoroughly wetted</td>
</tr>
<tr>
<td>D</td>
<td>Soils with very slow infiltration rates when thoroughly wetted</td>
</tr>
</tbody>
</table>

Table 2.3.1. – Characteristics of “hydrologic groups” S.C.S-CN.

Determination of these groups has been done according to the scheme adopted by BORSELLI & al. (1989) based on studies conducted in the Po Plain in Northern Italy. Following a semi quantitative approach, which considers information normally reported on pedologic maps (such as texture, structure, soil depth, soil permeability, permeability reduction with depth), each soil can be attributed to a hydrogeologic group.

A score can be associated to each soil by summing values calculated from given tables (Tab. 2.3.2 and 2.3.3).
Table 2.3.2. – Scores derived from texture and structure characteristics

<table>
<thead>
<tr>
<th>SOIL TEXTURE</th>
<th>WELL DEVELOPED</th>
<th>MEDIUM</th>
<th>POORLY DEVELOPED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Medium</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Fine</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 2.3.3. – Scores derived from soil permeability, permeability reduction with depth and depth of substratum

<table>
<thead>
<tr>
<th>SOIL PERMEABILITY (cm/h)</th>
<th>PERMEABILITY REDUCTION WITH DEPTH</th>
<th>DEPTH OF SUBSTRATUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,5 − 2</td>
<td>low</td>
<td>&lt; 25</td>
</tr>
<tr>
<td></td>
<td>moderate</td>
<td>25 ÷ 50</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>50 ÷ 100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt; 100</td>
</tr>
</tbody>
</table>

Table 2.3.4. – Final score and corresponding hydrologic group

The total score can be classified in four classes, each of which corresponds to one of the four hydrologic groups, according to Table 2.3.4.

Score (Tab. 1 + Tab. 2) | Hydrologic Group
------------------------|------------------|
0 ÷ 5                   | A                |
6 ÷ 10                  | B                |
11 ÷ 15                 | C                |
16 ÷ 19                 | D                |

The Pedologic Maps, realized by the “Office for the agricultural development of Lombardia Region” (E.R.S.A.L.) have been the base of the analysis; according to FAO classification of 1990, 53 different soils can be distinguished in the district of Milan.

All soils can be classified in the Hydrologic Group A, B and C. In the area there are no soils belonging to the Group D, the one having lower attitude to infiltration.
2.4. Land use

Available thematic maps have been used to create the land use map. In particular this map distinguishes urban, agricultural and woodland areas which together cover more or less the totality of the area in the district of Milan. More detailed distinctions into these areas have not been considered because of the working scale and the lack of more detailed original data.

A sketch of the map reporting the land use is represented in figure 2.4.1.

2.5. Hydraulic conductivity of vadose zone

Determination of hydraulic conductivity of vadose zone has been realized according to the following steps:
- collection and critical review of well boring logs;
- selection of 594 wells according to their spatial distribution and the thoroughness of boring logs description.
Then for each well has been accomplished:

- the identification of the vadose zone on the basis of the mean annual piezometric level;
- the subdivision of the vadose zone in layers characterized by the same type of granulometry;
- the attribution of an hydraulic conductivity value for each layer, using values in table 2.5.1.

<table>
<thead>
<tr>
<th>GRANULOMETRY</th>
<th>HYDRAULIC CONDUCTIVITY (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gravel, cobbles</td>
<td>4.00E-03</td>
</tr>
<tr>
<td>Coarse sand, small gravel, highly weathered conglomerate</td>
<td>2.00E-03</td>
</tr>
<tr>
<td>Sand and weathered conglomerated</td>
<td>8.00E-04</td>
</tr>
<tr>
<td>Fine sand</td>
<td>4.00E-04</td>
</tr>
<tr>
<td>Silt, very dense sand, slightly fractured conglomerated</td>
<td>8.00E-05</td>
</tr>
<tr>
<td>Clay</td>
<td>8.00E-06</td>
</tr>
</tbody>
</table>

Table 2.5.1. Hydraulic conductivity (m/s) adopted for the different granulometry.

- The calculation of hydraulic conductivity of the vadose zone, $k_{eqv}$. Because of the flux in the vadose zone is mainly vertical and it’s very sensible to low permeability layers, the following formula has been used:

$$
 k_{eqv} = S \frac{k_1k_2k_3...k_n}{S_1k_2k_3...k_n + S_2k_1k_3...k_n + \ldots\ldots + S_nk_1k_2...k_{n-1}}
$$

where $S_1$ and $S_n$ = depth of first and $n$-th layer, $S$ = total depth and $k_1$ and $k_n$= hydraulic conductivity of first and $n$-th layer.

![Fig. 2.5.1 – Distribution of hydraulic conductivity (m/s) of vadose zone; highest values are in red, lowest in blue.](image)

The classified map related to hydraulic conductivity of vadose zone has been derived through the interpolation of 594 obtained values (figure 2.5.1.). An evident variability of the spatial value distribution can be observed both in the central and in the northern sector of the Milan district area, while a more uniform distribution of low values of the analyzed parameter can be seen in the southern part of the investigated area. In general, a decreasing trend from north to south can be observed. This distribution is in complete agreement with the evolution of sedimentary events that have characterized this sector during the Quaternary age. In fact, because of the reduction of the deposition energy southward, there is a general increase of fine fraction in this sector and a consequent decrease of hydraulic conductivity of soils.
2.6. Groundwater depth

Piezometric levels have been continuously monitored by the Province of Milan by means of 364 observation wells, realizing four measures every year, whose spatial distribution is represented in figure 2.6.1.

Figure 2.6.1. Distributions of wells used to derive groundwater depth map

As shown in figure 2.6.2, groundwater depth, which is here subdivided in three classes, generally decreases from north to south, ranging from values higher than 25 meters to values lower than 5 meters.

Figure 2.6.2.: Spatial distribution of groundwater depth classes.
2.7. Groundwater velocity

Groundwater velocity data have been calculated by means of the application of a numerical model which has simulated the groundwater flow (ALBERTI & al., 2000). The numerical model has been applied to about 1200 km$^2$ and realized with a finite difference code (Modflow, USGS). The model has been implemented to simulate and predict the variation of groundwater levels in the Milan district area; for this reason internal conditions, such as water well abstraction and recharge (rain and irrigation) have been considered and analyzed. So, the obtained groundwater velocity data have been aggregated with other available data (for those areas outside the model) and then interpolated in order to obtain a spatial distribution of the analyzed parameter over the entire study area.

Figure 2.7.1.: Spatial distribution of groundwater velocity

There are two zone of high velocity located in the central northern and in the central southern sectors. Low values are mainly concentrated in the northwestern and the southeastern sectors. Difference in velocity can be equally attributed to hydraulic conductivity and hydraulic gradient changes even if the last are generally higher in the northeast quadrant. A sketch of the map is represented in fig. 2.7.1, where the high velocity area in the central-northern sector can be seen.
3. Weights of evidence and its application to aquifer vulnerability mapping

The ultimate purpose of this GIS project is to combine together spatial data, derived from different sources of information, in order to find, describe and analyze interactions, to make predictions with models and to provide support for decision-makers.

In this study, available data useful to produce a series of maps depicting the potential regional groundwater intrinsic vulnerability (defined as the relative susceptibility of groundwater to contamination due to anthropogenic activities), have been selected.

The potential regional groundwater vulnerability has been carried out using an empirical model, based on statistical relationships between the available predictors (events) and the vulnerability of aquifer to pollution, in terms of nitrate concentrations (the response), in each precise geographic location: the events (predictors) will be used to explain or predict the likelihood of another event (the response). In fact, the physical and chemical principles governing the relative vulnerability of aquifer to pollution on a regional basis are, for the most part, too complex for direct prediction from a mathematically expressed theory. So the prediction must rely mostly on empirical relationships with the aid of a “vulnerability model”.

The study has led to the recognition of a large number of relationships between the predictors and the response. Each of these relationships will be accurately defined and described in order to guide to the search for new similar situations of the same type in the investigated area. These descriptions include, as far as possible, an indirect evaluation of the chemical and physical processes that control aquifer vulnerability, but the model itself cannot be expressed in purely mathematical terms. From this point of view the “vulnerability model” plays an important role both in the selection of maps that are likely to be good predictors and, consequently, in the assignment of weights to the variable classes of each predictor map. This assignment of weights will be carried out using statistical criteria which are able to quantitatively estimate the spatial relationships (strength) between predictor maps and response map through the use of conditional probability statements.

A series of predictor maps, representing the spatial distribution of factors influencing groundwater vulnerability, are derived for the investigated area and a statistic-probabilistic methodology is adopted and implemented in a GIS for the analysis of the relationships between available data sets.

The spatial data analysis for groundwater vulnerability consists of the following steps (fig. 3.1):

- **Data collection.** Available data have been selected from an existing GIS project concerning both the definition of the hydrogeologic setting of the Milan district area and the description and analysis of the water resources of the investigated area. Particular attention has been given to the data useful for the production of maps concerning the spatial distribution of nitrate concentrations (fig. 3.2). These maps are based on chemical analysis performed on samples derived from public and private wells, located in Milan district, which are generally screened in different aquifer. The comparison between well boring logs, representing screen levels, and the bottom level of TR/TRa aquifer, allowed the implementation of a chemical database for each aquifer of Milan plain. Selection of response variable has been enough problematic. As reported in chapter 2, a chemical compound which is ubiquitously present in all the investigated area, because of the presence of homogeneously distributed dispersed sources, should represent the ideal response variable. Given that this is not the case, what can be done is to select a parameter whose presence can be attributed to conditions as much as possible similar to the ideal situation.

- Nitrate concentrations, as a chemical compound which is ubiquitously present in all the investigated area, are selected as response variable. Nitrate concentrations are one of the most common contaminants present in groundwater in the area since many years. They are almost ubiquitously distributed and continuously monitored in a great number of wells.
As a rule of thumb, the concentration of nitrates that can be considered as normally present in groundwater (Provincia di Milano, 1999) are estimated in about 15 mg/l. Higher concentrations should be attributed to human activities.

- The most common sources of nitrates are considered sewer systems or urban areas in general, agriculture and breeding. The district of Milan can be roughly divided in two subareas: the north one prevalently urban and the south one prevalently agricultural; breeding is present but circumscribed to small areas. In these subareas the two activities are so developed that contaminant sources can be considered as dispersed sources for the adopted working scale.

- Pre-processing of available data and construction of GIS database containing all causal factors (classified themes and other supporting information). The pre-processing step has involved both the classification of the available data and the definition of their spatial distribution.

![Maps concerning the spatial distribution of nitrate concentration, measured on 1983. Values range from 0 to 49 mg/l. The black lines mapped in the figures represent the nitrate concentration limit of 25 mg/l (left) and 30 mg/l (right) respectively.](image)

Some parameters, in fact, are already mapped with a complete spatial distribution covering the whole study area (for example, the land use, soil type, etc.); others are represented by means of maps obtained from interpolation of original point data (rainfall, hydraulic conductivity, etc.) using geostatistical analysis. A variogram has been firstly calculated for every data set and the interpolation has been made using Kriging algorithm. At the end, the seven predictor maps have been obtained. Moreover two other maps, available in the database, have been used in this study: the map concerning the nitrate concentration related to 1983, used for the construction of the prediction model; the map concerning the nitrate concentration related to 1997, used for the validation of the prediction results.

- Construction of the prediction model, including the creation of unique condition subareas, computation of weight probability values, construction of prediction maps and representation of preliminary prediction results. The adopted methodology expresses the relative susceptibility of groundwater to contamination in terms of probabilities and combines them by means of the bayesian combination formulas, under the conditional independence assumption. The Weights of Evidence method (Bonham-Carter, 1989) is the log-linear version of the general bayesian model, normally applied where the evidence is binary. The bayesian approach to the problem of combining data sets uses a probability framework based on the idea of prior and posterior probability. The prior probability is the probability that an event will re-occur only based on existing data (the distribution of nitrates, as mapped in 1983), without any additional information (predictor maps). The availability of new evidences (land use, irrigation, etc.) will permit the calculation of the posterior probability which will increase or decrease the prior probability. In other words, the posterior probability can be expressed in terms of the prior probability and a multiplication factor that can be greater or lesser than one. A full description of the Weights of Evidence method is provided in Bonham-Carter (1989). So, the main idea is that the prior probability can be successively updated with the addition of new evidences (derived from multiple data sources) in order to realize posterior probability maps where the “favourability” for potential aquifer vulnerability is greater or smaller than the average (prior probability). In this approach each variable class is treated as being binary, either present or absent. The method is fundamentally based on the calculation of the terms \( P(V_i|T)/P(V_i|T^-) \), known as sufficiency ratio and \( P(V_i^-|T)/P(V_i^-|T^-) \), known as necessity ratio, where \( V_i \) represents the \( i \)-th variable class, \( T \) represents the response variable (target) and the symbol ‘-’ represents the absence of the variable class or the absence of the response variable. In the weights of evidence, the natural logarithm of both these ratios determines the positive and the negative weights (\( W^+ \) e \( W^- \)), respectively, which can be calculated from the available data (tab. 3.1).
Table 3.1: Weights values obtained considering as response variable a nitrate concentration higher than 25 mg/l (above) and 30 mg/l (below). (CAN = irrigation; SOG = groundwater depth; VEL = groundwater velocity; KPO = hydraulic conductivity of vadose zone; PLU = rainfall; SOI = type of soil; LAN = land use; CPL = classes resulting from the intersection (AND boolean operator) of irrigation, rainfall and land use; CPS = classes resulting from the intersection (AND boolean operator) of CPL and soil type).
Defining all expressions in odds form (odds are defined as a ratio of the probability that an event will occur to the probability that the same event will not occur), it’s possible to obtain:

\[
\log_e O(T \mid V_j) = \log_e O(T) + W^* \quad [3.1] \\
\log_e O(T \mid V_i^+) = \log_e O(T) + W^- \quad [3.2]
\]

where \(\log_e O(T)\) is the natural logarithm of prior odds, derived from the prior probability of having a unit cell (a pixel), chosen at random, which contains the target, where no other information is available. In this case, the occurrence of the response variable together with each variable class (predictor pattern) is analyzed independently to produce a pair of weights. So, if the \(W^*\) is positive and \(W^-\) is negative, there is a positive correlation between the response variable and the binary pattern, indicating more points on the pattern than would be expected just due to the chance. Conversely, if \(W^*\) is negative and \(W^-\) is positive, there is the case where the pattern is negatively correlated and fewer points occur on the pattern than would be expected just due to the chance. If \(W^* = W^- = 0\), (and the posterior probability is equal to the prior probability) the response variable is independent of whether the pattern is present or not (and the probability of finding the response variable would be unaffected by the presence or absence of the binary pattern). Therefore the contrast \(C = W^* - W^-\) gives a useful measure of the correlation between the predictors and the response, in each precise geographic location, becoming zero when a predictor has a distribution which is spatially independent in relation to the response. After obtaining the weight values, the next step consists of the combination of the available predictor maps, provided that the variable classes constituting each predictor map are conditionally independent with respect to response variable:

\[
\log_e O(T \mid V_1 \cap V_2 \cap \ldots \cap V_n) = \log_e O(T) + \sum_{k=1}^{n} W
\]

where \(V_i\) represents the \(i\)-th variable class, \(T\) represents the response variable (target). After calculating the preliminary probability map, a sensitivity analysis has been performed in order to optimize the correspondence between the number of pixels calculated “potentially vulnerable” and the number of pixels really observed and mapped in 1983 with a concentration of nitrates higher than 25/30 mg/l. There are some assumptions which govern the application of this methodology:

1. The prior probability is assumed to be constant over the entire study area, and this is not always the case. In this study the prior probability is initially determined from the spatial distribution of nitrate concentrations higher than 25 mg/l and 30 mg/l, without defining sub-areas with different values of prior probabilities. The prior probability values are equal to 0.45 and 0.29 for a nitrate concentration higher than 25 and 30 mg/l respectively (as response variable).

2. The conditional independence of the predictor patterns with respect to the occurrence of the response events. When the evidences from several maps are combined, the weights are calculated from each map independently and then combined into a single equation. This requires the assumption of conditional independence that, if not verified, leads to a model that, like most models, does not fit the data perfectly. This provides a simplification that, when used carefully, is useful for prediction and gives insight into the relative contributions of the separate sources of evidence. If conditional independence is violated, the predicted vulnerable areas (\(N[V]\)) is much larger (in terms of number of pixels) than the observed vulnerable areas:

\[
N[V]_{calc} = \sum_{k=1}^{m} P_k * N[A]_k 
\]

The predicted number is determined by adding together the product of each posterior probability class (P) times the number of pixels belonging to that posterior probability class, for each \(k\)-th posterior probability class on the map. In practice, the predicted number is always larger than the observed number with weights of evidence, but if the predicted number is more than 10-20% larger than the observed number then a serious check of the pairwise tests and remedial action is in order (Bonham-Carter, 1989).

Under the assumption of conditional independence, the effects of each binary map can be evaluated individually and then combined by adding (in the log-linear case) the factors for several maps together. In practice, probably conditional independence is always violated to some degree even if this assumption can be checked with statistical tests (pairwise tests such as \(\chi^2\)) to show the magnitude of the problem and pinpoint the most problematic maps. These maps can then be rejected from the analysis or modified to reduce the problem. The pairwise tests have to be applied to all possible combinations of variables. In this study, the combination of all 29 variable classes results in an enormous number of combinations. Therefore the test has been done using the most important variable classes obtaining a limited number of combinations.

The method can be summarized in the following two steps: 1) identification of the variable classes conditional dependent; 2) combination of the dependent pattern together with SUM operations or with boolean AND operator. In the second case, only areas where both binary patterns are simultaneously present have been considered. For example,
some combinations of irrigation, rainfall and landuse classes have showed extremely high $\chi^2$ values, proving to be dependent. Thus when the individual combinations are used to construct the final posterior probability map, values for whose combinations which are statistically dependent will be too high. For evaluating the effect of dependent variable classes on the final posterior probability values, three different situations will be examined: a) each predictor map (in binary format) has been evaluated individually and then combined by adding the factors for several maps together; b) three predictor maps (irrigation, rainfall and land use) have been overlaid and summed obtaining a new map reclassified with five new variable classes (CPL). These five classes give us a measure about the amount of the superficial water available to infiltration (tab. 3.3); c) the CPL predictor map (irrigation, rainfall, land use) has been overlaid and intersected (AND boolean operator) with the soil type map, obtaining 15 new variable classes (CPS). These fifteen new classes give us a measure of the amount of water which will pass through the vadose zone (characterized by a specific value of hydraulic conductivity) and reach the groundwater.

<table>
<thead>
<tr>
<th>Irrigation (mm)</th>
<th>Rainfall (mm)</th>
<th>Landuse</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAN1 0,1</td>
<td>PLU1 &lt; 950</td>
<td>LAN1 0,1</td>
</tr>
<tr>
<td>CAN2 246,0</td>
<td>PLU2 950 - 1040</td>
<td>LAN2 0,7</td>
</tr>
<tr>
<td>CAN3 358,4</td>
<td>PLU3 1040 - 1130</td>
<td>LAN3 0,9</td>
</tr>
<tr>
<td>CAN4 473,2</td>
<td>PLU4 &gt; 1130</td>
<td></td>
</tr>
<tr>
<td>CAN5 790,6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAN6 1531,6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tab. 3.3: Results of the summation of three predictor maps (irrigation, rainfall and land use). The new map has been reclassified according to five new variable classes.

- Validation of prediction results (time robustness and space robustness). The validity of the final vulnerability map has been tested through the comparison of the post probability map with spatial distribution of nitrate in different years in the area (comparative analysis, fig.3.3). In fig. 3.4, the 1997 nitrate concentration limit of 30 mg/l has been used to evaluate the prediction results for the whole study area.

**Success rate and prediction rate of the model**

**Sensitivity Analysis (nitrate concentration 1983)**

Refine, modify and adapt model in order to get the best fitting between calculated data and observed data (success rate)

**Comparative Analysis (nitrate concentration 1997)**

Verify the prediction rate of the model through a comparison between the prediction results and the situation really mapped in 1997

Fig. 3.3: Flow chart concerning the principal steps in the definition of the success rate and prediction rate of the model for groundwater vulnerability assessment
4. Results

Different scenarios have been analyzed using all predictor maps in relation to different nitrate concentrations: > 25 mg/l and > 30 mg/l.

The choice of comparing predictor maps with nitrate concentration higher than 15 mg/l (as above mentioned) has been done because of the strong presence of contaminant in the area. The use of 15 mg/l widely extends the spatial distribution of the response variable increasing the prior probability values all over the area. In this way, the number of relationships between predictors and response variable increases, resulting in high probability values in almost all the area and determining a reduction in the prediction capability of the method.

Maps representing the spatial distribution of post probability values (fig. 3.4) broadly divide the territory into three areas: the northern sector with high probability values, the southern sector with low probability values and the central sector which represent a transition zone with a wide range of probability values.

However there are many exceptions to this general trend especially in the eastern area where it is relatively easy to find high localized probability values in the southern sectors and low probability values in the northern one.

In general there is a great influence of infiltration process, and this is the reason of low probability values in all the southern sectors which is the one covered by the irrigation net.

The importance of this factor in reducing vulnerability is also showed by the urban area of the city of Milan, located more or less in the center of the district, which constitutes an extension southward of high probability values. In fact in this urban area the extensive impermeable cover at the surface cause a great decrease of infiltration increasing the probability values. The inverse relation between this predictor and vulnerability can be attributed to its strong action of mixing and dilution of contaminants in the vadose zone which consents to maintain low concentrations in nitrate even where there is an extensive agricultural practice and a general low groundwater depth, as in the southern sector of the district.
The comparison between the shape of 30 mg/l nitrate isoline in 1997 and posterior probability map allows some other interesting observations. Posterior probability map predicts that vulnerability of first aquifer is low in sectors II and IV (Fig. 4.1). Likewise, a great difference of concentrations between sector III and II, both localized on Milan city where there is the same recharge rate. That means that in the middle part of Milan District not only recharge rate has a great influence but something else becomes important. The map of vadose zone hydraulic conductivity gives an explanation of this fact. In sectors with low recharge rates where there is low conductivity values (sector II and IV) vulnerability decrease, where conductivity is higher (sector III) vulnerability increases.

The influence of hydraulic conductivity of vadose zone can also be seen in two sectors in the southern part of Milan District (yellow ellipses in fig. 4.1.). In these two sectors, even if partially hidden by the recharge influence, there are some spots showing medium – high probability values, whose location fits with the spatial distribution of high hydraulic conductivity values.

On the basis of this analysis, it’s possible that the 30 mg/l isoline could have a great probability to advance southward in sectors I, III and V with the same condition of contaminant diffusion.

![Maps showing vulnerability assessment](image)

**Fig. 4.1 - Comparison between vadose zone hydraulic conductivity (a) and posterior probability map (b)**

**Conclusion**

In this study we analyze the assessment of regional groundwater vulnerability coupling with the Weights of Evidence model with GIS.

This approach is uncommon in vulnerability studies and its application has required a series of assumptions and difficulties which have been analyzed in the study. The good results obtained after performing the validation of prediction results show that the method here proposed can be a powerful tool to analyze regional groundwater vulnerability.

It represents a useful support for studies concerning contaminant migration and identification of sources areas and provides a guide for planning future data collection.

The mathematical framework, which supports the Weights of Evidence model, makes the reasoning transparent and consistent and permits the incorporation of experts’ knowledge to improve the prediction results.

In particular for the investigated area the following main results have been obtained:

- maps representing the spatial distribution of post probability values broadly divide the territory into three areas;
- recharge is the most important factor in influencing probability values;
- vadose zone hydraulic conductivity also plays an important role in the definition of the vulnerability;
- results obtained in the southern sector show that use of nitrates in agricultural practice should not strongly affect groundwater quality;
- analyses permits to do some hypothesis on the most probable evolution of contamination spreading;
- results of the analyses suggest that some predictors could be combined to improve the final vulnerability map avoiding the problems related to the conditional independence of the predictor patterns with respects to the occurrence of the response event;
- the combination of predictors should be done in agreement with their physical meaning not losing information on the importance of processes affecting vulnerability.
BIBLIOGRAPHY


FRANCANI V., ALBERTI L., PARRI A & TALI P., 1999: “Prospezioni idrogeologiche per la prevenzione e riduzione del rischio di inquinamento nelle aree urbane”, 3° Convegno Nazionale sulla Protezione e Gestione delle Acque Sotterranee per il III Millennio, Quaderni di Geologia Applicata, Vol. 1

PROVINCIA DI MILANO, 1999: “Nitrati in falda” Public Report


USDA 1987: Water, United States Department of Agriculture Yearbook