MODELLING AQUIFER VULNERABILITY TO NITRATES UNDER THE ASSUMPTION OF VARYING SPATIAL SUPPORT OF WATER WELL DISTRIBUTION

Fabbri A. G.[1] and Patera A.[2]

[1]Università di Milano-Bicocca, Milan, Italy; andrea.fabbri@unimib.it
[2]Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy; antonio.patera@ingv.it

ABSTRACT

A previously analysed database is revisited to generate different classifications of an area of study into vulnerability ranks. The distribution of 305 water wells around the city of Milan in northern Italy, covering an area of about 2000 km^2, measured nitrate concentration in groundwater. The wells are separated into 133 with NO_3^- value clearly above a threshold of 25 mg/l (impacted wells) and 172 below that (non-impacted wells). Square neighbourhoods of dimensions 20x20 m, 60x60 m, 180x180 m and 1020x1020 m around the 305 wells are used to delimit four training areas of different sizes. Over the neighbourhoods as well as over the 2000 km^2 area, nine natural and anthropogenic map data are assumed, as indirect supporting patterns of the modelling, to reflect both the potential source of nitrates and the relative ease in which nitrates may migrate in ground water. In the training areas comprising impacted and non-impacted well neighbourhoods those of the impacted wells are used as direct supporting patterns for mapping the predicted vulnerability ranks. It is done by a mathematical model that computes spatial relationships between the direct and indirect supporting patterns based on empirical likelihood ratios. The relationships are integrated into *prediction patterns* and, by iterative cross-validations, into *target* and *uncertainty patterns*. These will be then extended over the remaining much larger study areas for analysis and visualization. The analytical procedure proposed is considered applicable to situations in which a spatial database contains just a sampling of measured sites (e.g., drill-hole data), as direct supporting pattern, instead of the natural distribution of hazardous sites such as trigger areas of landslides or of sink holes, as is common in spatial prediction modelling of natural hazard.

Keywords: aquifer vulnerability, nitrate pollution, empirical likelihood ratios, spatial support, prediction patterns, uncertainty patterns, prediction-rate curves

1 INTRODUCTION

A spatial database constructed for the prediction of ground water vulnerability is reanalysed to explore the effects of different spatial support of point data of nitrate concentration. The study area of about 2000 km^2 is located around the city of Milan, in northern Italy and the corresponding database was studied by Masetti *et al.* (2007) [1] as a refinement of earlier works [2, 3], who thoroughly discussed the study area, its groundwater contamination problems and the database they constructed. Furthermore, those authors considered new analyses with different threshold values of nitrate concentration [4, 5]; the reliability of different vulnerability classification schemes; [6]; and compared positive and negative weights for multiclass generalizations [7]. In those works the weight-of-evidence model, WoE, was applied for vulnerability assessment.

Methodological discussions on some of those works [1-4] led to the generous provision of the database by the original authors for complementary modelling applications in a joint contribution [8]. In it a different modelling framework and analytical strategy were preferred: the empirical likelihood ratio function, ELR, and cross-validation for uncertainty assessment of *prediction patterns*. Use was made of the immediate vicinity of 305 available water wells measuring the concentration of NO_3^- in mg/l in a training area within the area of study. Of the wells, 133 were considered as "impacted" by nitrate pollution.

As a follow up to their analyses, this contribution focuses narrowly on the effect of spatial support of the sampling points used to express the presence or absence of NO_3^- pollution in the water table. In practice, the Milano area database becomes an opportunity to point at a very general prediction modelling problem in which the basic direct evidence of a process is obtained by sampling a study area with point like measurement values, as the ones from drill holes or water wells. Main questions are: "What is the acceptable spatial support for the modelling?" and "What happens if we assume broader spatial support?"

The next section offers a brief summary of the database and its initial purpose. The favourability modelling framework is then discussed along with the proposed strategies for characterizing, visualizing and cross-validating. Experiments follow on the four training areas to obtain prediction-rate curves, *prediction*, *target*, *uncertainty* and *combination patterns* of the respective study areas. Concluding remarks follow with considerations on the importance of assuming realistic spatial support for the water well distribution and values.

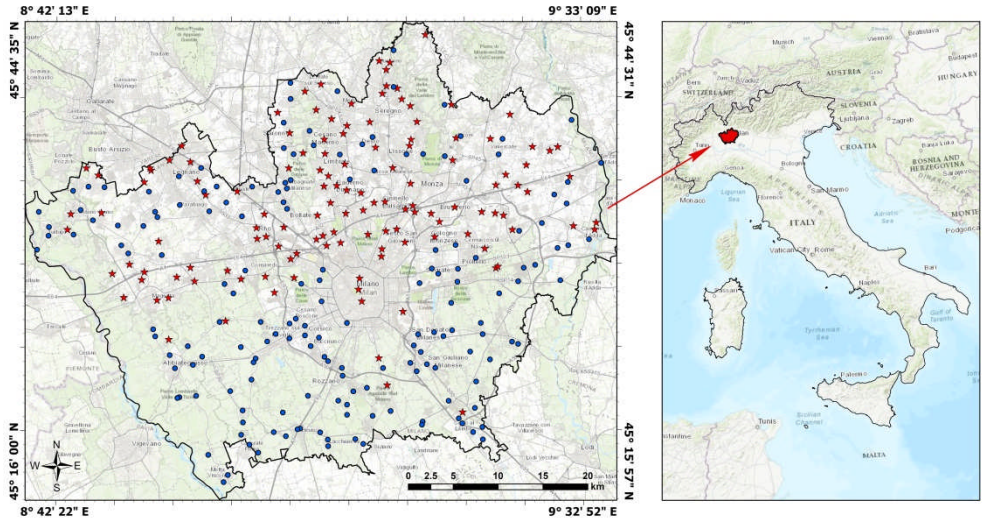

Figure 1: Distribution of 305 water wells in the Milano area of study. Red stars indicate the wells recording ≥ 25 mg/l of NO_3^-; blue circles those recording < 25 mg/l. Sparse red stars among cluster of blue circles can be seen as well as vice versa, indicating a noisy clustered distribution.

2 THE MILANO AREA OF STUDY DATABASE

Agricultural practices and industrial activities characterize the area of study around Milan in northern Italy, located as shown in Figure 1. It covers nearly 2000 km^2 and its groundwater system has a complex hydrogeological setting with interaction of three aquifers [2]. The subsoil sediments represent important water resources. The studies by Masetti *et al.* (2007) [1] focused on the vulnerability of an unconfined aquifer that represents the most affected by contaminants from the surface activities. It is termed Traditional Aquifer and consists of Pliocene-Pleistocene sediments. Transmissivity ranges from $5x10^{-2}$ to $1x10^{-3}$ m^2/s, with permeability between $5x10^{-3}$ and $1x10^{-8}$ m^2/s. Thickness ranges from 60 to 120 m. Components are gravel and sands and clay-silt layers that increase southward. Groundwater depth averages 30 m to the north of the area reducing to 5 m to the south.

Table 1: Wells, natural and anthropogenic factors in the Milano study area database (modified after [8]). Note the one-digit short names that will be used to identify the ISPs used for analysis.

Water well data, Area of study and Direct supporting patterns, DSPs			
Factor map name	Short names	Data range	Description
Impacted wells	**133**	1-133	Index $\geq$ 25 mg/l NO_3^-
Non impacted wells	**172**	1-172	Index $\leq$ 24 mg/l NO_3^-
Area of study	**AS**	1-0	Area & out-of-area indicator
Categorical natural and anthropogenic factors			
Factor map name	Short name	Data range	Description
Ground water recharge	**gwr, R**	classes 6-15	Combination of **raf** & **mai** x a function of **spc** as infiltration coefficient
Land use	**ldu, L**	classes 1-3	Urban, agricultural & woods
Soil protection capacity	**spc, S**	classes 1-3	Low, moderate & high
Continuous natural and anthropogenic factors			
Factor map name	Short name	Data range	Description
Ground water depth	**gwd, d**	1-51	m
Ground water velocity	**gwv, v**	112-181	originally 10-20-ln (m/s)
Main annual irrigation	**mai, i**	0.1-1531.0	mm
Nitrogen fertilizer loads	**nfl, n**	0-428	kg/h/y
Population density	**pod, p**	43-7933	inhabitants/km^2
Rainfall	**raf, r**	808-1253	mm/y

Information on nitrate concentration was collected from over 300 water wells, unevenly distributed throughout the whole area, as shown in Figure 1, to monitor four times a year the nitrate concentration that appears not sensitive to seasonality. Concentration varies between 10 mg/l to the south and 70 mg/l to the north, with median value around 20 mg/l. European Community Standard [9] set a guide value in soil of 25 mg/l.

Alberti *et al.* (2001) [2] and Masetti *et al.* (2007) [1] provided a detailed account of the absence of temporal trends and the differences between the northern and the southern parts of the area of study. They employed statistical analyses in their study of regional

groundwater vulnerability to spatially relate measured contaminant locations with the distribution of natural and man-induced factor maps. For this they have constructed a database consisting of impacted and non-impacted water wells selecting the study area and the following natural and anthropogenic factor maps listed in Table 1: groundwater recharge, land use, soil protection capacity, groundwater depth, groundwater velocity, main annual irrigation, nitrogen fertilizer loading, population density and rainfall. Table 1 also lists their short name abbreviations and value ranges.

In essence, the database for the Milano area of study used here consists of a set of 10 digital images contained within a raster of 3300 pixels by 2665 lines. Each pixel corresponds to a square of 20x20 m on the ground. Of the 8,794,500 pixels in the rectangular raster, only 4,908,305 cover the area of study, and 3,886,195 are falling out of it. The distribution of the 305 water wells, shown in Figure 1, is represented as an image with 305 pixels with the values of NO_3^-, ranging from a maximum of 71.0 mg/l to a minimum of 10.9 mg/l. Of those, the 133 with value $\geq$ 25 mg/l were termed "impacted wells" while the 172 with value < 25 mg/l were termed "non-impacted wells." Together the sets of impacted and non-impacted wells represent all we know about nitrate concentration in the groundwater in the area of study so that their location distribution can be used to define training areas for the modelling. Within the training areas, the distribution of sequentially numbered impacted wells is converted into a direct supporting pattern, DSP, and used to establish spatial relationships with the images of natural and anthropogenic factors, converted into indirect supporting patterns, ISP. The relationships are then extended to the remaining study areas.

3 FAVOURABILITY FUNCTION MODELLING

The term "favourability function" was proposed [10] to refer to spatial modelling within a unified mathematical framework. Examples of interpretations that were considered are: Bayesian Probability, Certainty Factor, Dempster-Shafer Belief function and Fuzzy Logic. Their implicit assumptions were discussed along with their computations under different database conditions. In particular, integration rules for the models were discussed by Chung and Moon (1991) [11]. The ELR model has been thoroughly discussed by Chung (2006) [12] and it will be used here.

The modelling with the ELR function generates an image with integrated values ranging from 0 to infinity for each pixel, a *prediction image*. The array of relative integrated values, however, is difficult to interpret as such, so that a transformation is conveniently made of it into a *prediction pattern*. In this transformed version all values are ordered from highest to lowest and equal-area ranks are replacing the ratios for each pixel. "Pattern" is to refer to an artificial construct, i.e., a particular way, of interpreting and displaying the results of modelling. "Prediction" is implying that it indicates areas in which future occurrences are likely to be found. DSP and ISP have already been defined.

To apply a model we must use all the known occurrences first, in our case all the 133 impacted sequentially numbered well locations, for instance. This should generate the most informed pattern. However, for interpreting it we need to study the pattern's ability to "predict" future occurrences, i.e. their location or distribution. We then pretend not to know the location of some of the occurrences ("younger"), apply again the model using the remaining occurrences ("older") as DSP, generate a new *prediction pattern* and then verify where in it the excluded occurrences are located: hopefully in the higher ranks of the pattern. Of course there are many ways to perform such a cross-validation, as we have termed it. Convenient iterative strategies are of excluding sequentially a few occurrences

repeating modelling and cross-validation a number of times. Alternatively sequential selection or random selection can be preferred. Clearly, the strategy can be tailored to the peculiarities of the available data.

Another critical aspect of spatial prediction modelling is the selection of a training area in which to establish the spatial relationships between DSP and ISP. This is because in it the relationships are considered either more accurate or more easily measurable. From the training area the relationships computed can then be extended to the remaining part of the area of study, termed study area.

The visual expression of a *prediction pattern* can be generated by conveniently grouping the 200 equal area ranks into fixed and recognizable classes: for instance broader classes for lower ranks of lesser concern, and narrower classes for higher ranks. Furthermore, for facilitating comparisons between predictions, the classes must remain the same for all subsequent results of the iterative cross-validations process. Iterations allow computing *target patterns*, *uncertainty patterns* and their *combination patterns*. For instance, using robust statistics such as median and range, we can obtain a *target pattern* with pixel values of the median rank of the *prediction patterns* generated by iterations. The values of the *uncertainty patterns* correspond to the rank of the ranges around the median of the *target pattern*.

The spatial relationships and their integration by a model do imply assumptions as to the data type available and to the specific mathematical model being used. Examples of data and model assumptions have been discussed in a previous work by Fabbri *et al*. (2010) [8] who used the very same Milano database. They do not need repeating here. We can just point at the following: (i) the assumption that the future occurrences of pollution will take place under conditions similar to the ones represented in the database, and (ii) that the indirect spatial support consists of conditionally independent factor maps. Assumption (i) represents our hope given that we consider as satisfactory, for the past and the future, the information collected in the database. Assumptions (2) are related with the representation and integration rules specific to many mathematical models that were initially formulated for non-spatial factors like medical symptoms for the prescription of medications or identification of diseases. In the geosciences, however, most frequently different thematic maps over the same area are hardly conditionally independent. It becomes of importance then to verify the dependence effects on the modelling.

4 EXPERIMENTS ON TRAINING AND STUDY AREAS

We can now formulate assumptions on the spatial support of the water wells: for instance, a pixel area of 20x20 m (1x1 pixels), of 60x60 m (3x3 pixels), of 180x180 m (9x9 pixels) or of 1020 x1020 m (51x51 pixels). Having analysed the distribution of the 305 wells and their values, we have found that it is “dispersed”, i.e., the average distance of the wells is greater than a hypothetical random distribution. Distances range from 242 m to 3942 m. The 133 wells are “clustered” and show a hot-spot to the NNE and a cold-spot to the SSW of the area of study, as visible in Figure 1. There are numerous low-high outliers and a few high-low ones.

However, when gridded into 20x20 m pixels and used to identify target areas of 305 pixel neighbourhoods containing 133 impacted pixel neighbourhoods, we have to discover what the spatial relationships within the database do contribute to the spatial modelling results as *prediction patterns*. To do that, four sets of training areas and study areas were computed from the database. They were termed **Ta1**, **Ta3**, **Ta9** and **Ta51**, and the respective remaining parts, the areas of study, as **Sa1**, **Sa3**, **Sa9** and **Sa51**. The training

areas cover, respectively, the following number of pixels of 20 m resolution: 305, 2,745, 24,706 and 761,161. The corresponding study areas cover complementary numbers.

The database consists of three types of factor maps: (i) well locations, (ii) categorical natural and anthropogenic factors, and (iii) continuous ones. They are described and short named in Table 1.

For each training area, the modelling applied consists of the steps described in the following sub-sections: (1) calculation of empirical likelihood ratios; (2) generation of prediction patterns later extended to the study areas; (3) computations of iterative cross-validations; and (4) generation of **target**, *uncertainty* and *combination patterns* later extended to the study areas

Table 2: Empirical likelihood ratios for **ISP**s are listed for the different training areas. Mostly ratios > 1.5 are listed and when ≥ **2** they are in bold. Upper case letters with subscripts indicate the individual categorical map units. Numbers in italics indicate ranges of continuous field values with bracketed maximum value and ratio reached. Note that for **spc** all ratios are well below 1.5.

ISP Ta	**ELRs** (≥**2** and > 1.5)
gwr Ta1	$R_8$1.78, **R_9 3.20**;
ldu	L_2 1.29;
spc	< 1.50, S_1, S_2, S_3;
gwd	≥ **2** *21.65-24.55* (*23.15* max **2.43**); *30.05-48.50* (*39.95* max **4.03**);
gwv *10	≥ **2** *120.59-127.40* (125.88 max **2.10**); *163.02-164.00* (*164.00* max **5.60**);
mai *10	≥ 1 *0.00-1424.74* (1.00 max 1.69); ≥ **2** *6430.80-11528.62* (*11298.99* max **2.28**);
nfl *10	≥ 1 *684.80-1264.62* (*894.91* max 1.58); ≥ **2** *2166.82-2235.80* (*2199.93* max **2.12**); *2519.98-2602.65* (*2553.09* max **2.16**);
pod	≥ **2** *2178.94-2546.76* (*2320.91* max **2.17**); *2946.83-5734.54* (4934.35 max **322.85**); *6244.34-6496.01* (*6496.01* max **3.05**);
raf	≥ **2** *1057.81- 1128.09* (1100.14 max **4.93**).
gwr Ta3	R_8 1.84, **R_9 3.04**;
ldu	$L_2$1.51;
spc	< 1.50, S_1, S_2, S_3;
gwd	≥ **2** *21.75-24.45* (*23.10* max **2.35**); *30.40-48.40* (*40.15* max **4.10**);
gwv *10	≥ 1 *105.21-133.26* (*117.45* max 1.60);
mai *10	≥ 1 *0.00-1424.74* (1.00 max 1.69); ≥ **2** *6430.80-11528.62* (*11222.45* max **2.29**);
nfl *10	≥ 1 *2127.17-2931.82* (*2461.01* max 1.63);
pod	≥ **2** *3054.22-5688.01* (*4942.29* max **35.10**); *6251.26-7933.08* (*7933.08* max **7030.60**);
raf	≥ **2** *1086.14- 1129.78* (1107.34 max **2.16**).

4.1 Empirical Likelihood Ratios

Table 2 lists the empirical likelihood ratio values, ELR, for the nine ISPs corresponding to the different training areas. Values are shown when > 1.5 for at least one ISP in one or more areas for comparison. In bold fonts are all values ≥ 2, a tentative value to threshold the ratios. A value of 1 indicates a frequency in the presence of impacted well identical to that in their absence within the training area. A value of 2 indicates a frequency twice that in the absence of an impacted well. The table filters the essential characteristics of likelihood ratio histograms for the three categorical ISPs and ratio functions for the six continuous field ISPs. We can consider the set of ratios as "signatures" of the training areas. The ELR model integrates them for each pixel of a training area the empirical likelihood values of the nine ISPs. This generates a *prediction image* to be transformed into, and interpreted as, a *prediction pattern*.

Comparing the ratios for the four training areas we can observe the following similarity and differences: (1) $\mathbf{R_9}$ is high, > 3 in all areas; (2) $\mathbf{L_2}$ is > 2 only for **Ta9**; (3) **spc** is low, < 1.5 in all areas; (4) **gwd**, **mai**, **pod** and **raf** are > 2 in all areas, however, **raf** is higher, **4.93** in **Ta1**; and (5) **gwv** is > 2 only in **Ta1**.

Table 2: Continued.

ISP Ta	**ELRs** (≥**2** and > 1.5)
gwr Ta9 **ldu** **spc** **gwd** **gwv** *10 **mai** *10 **nfl** *10 **pod** **raf**	R_8 1.75, $\mathbf{R_9}$ **3.17**; $\mathbf{L_2}$ **2.09**; < 1.50, S_1, S_2, S_3; ≥ **2** *21.80-24.25* (*24.05* max **2.29**); *30.05-48.10* (*39.95* max **4.13**); ≥ 1 *105.06-133.27* (117.44 max 1.59); ≥ 1 *0.00-1423.83* (15.31 max 1.67); ≥ **2** *6430.22-11528.36* (7838.75 max **2.30**); ≥ 1 *2127.17-2931.82* (*2461.01* max 1.63); ≥ **2** *3078.02-5688.01* (4942.29 max **34.89**); *6251.26-7933.08* (*7933.08* max **7032.39**); ≥ **2** *1085.75-1129.44* (*1106.97* max **2.15**).
gwr Ta51 **ldu** **spc** **gwd** **gwv** *10 **mai** *10 **nfl** *10 **pod** **raf**	R_8 1.62, $\mathbf{R_9}$ **3.38**; L_2 1.34; < 1.50, S_1, S_2, S_3; ≥ 2 *21.35-23.90* (*22.55* max **2.16**); *29.85-45.55* (*38.25* max **3.69**); ≥ 1 *1.00-57.88* (max 1.70); *102.34-133.30* (116.92 max1.57); ≥ 1 *0.00-1224.00* (*229.65* max 1.62); *6169.95-6445.53*; ≥ 2 *6460.84-11513.05* (7348.82 max **2.20**); ≥ 1 *2110.05-2863.34* (*2362.57* max 1.82); ≥ **2** *3268.41-5703.87* (4942.29 max **41.06**); *6338.52-7903.08* (*7903.08* max **4881.51**); ≥ 1 *0.00-411.58 (248.95* max 1.70); **≥ 2** *1085.86-1129.64* (*1107.13* max **2.16**).

These differences in ELR values appear as minor and we may expect similar modelling results. The likelihood values are characteristic properties of the spatial database, measuring the spatial relationships between DSP and ISPs in the training areas. We can consider the ratios in Table 2 as database signatures. The mathematical modelling, in our case by the ELR function model, will integrate the ratios for each point or pixel in the database, under a number of assumptions and following specific combination rules. Recall that the ratios range in values between zero and infinity and that the values are difficult to interpret. Having the signatures of the training areas we can now try to see their effects on the modelling of the respective *prediction patterns*.

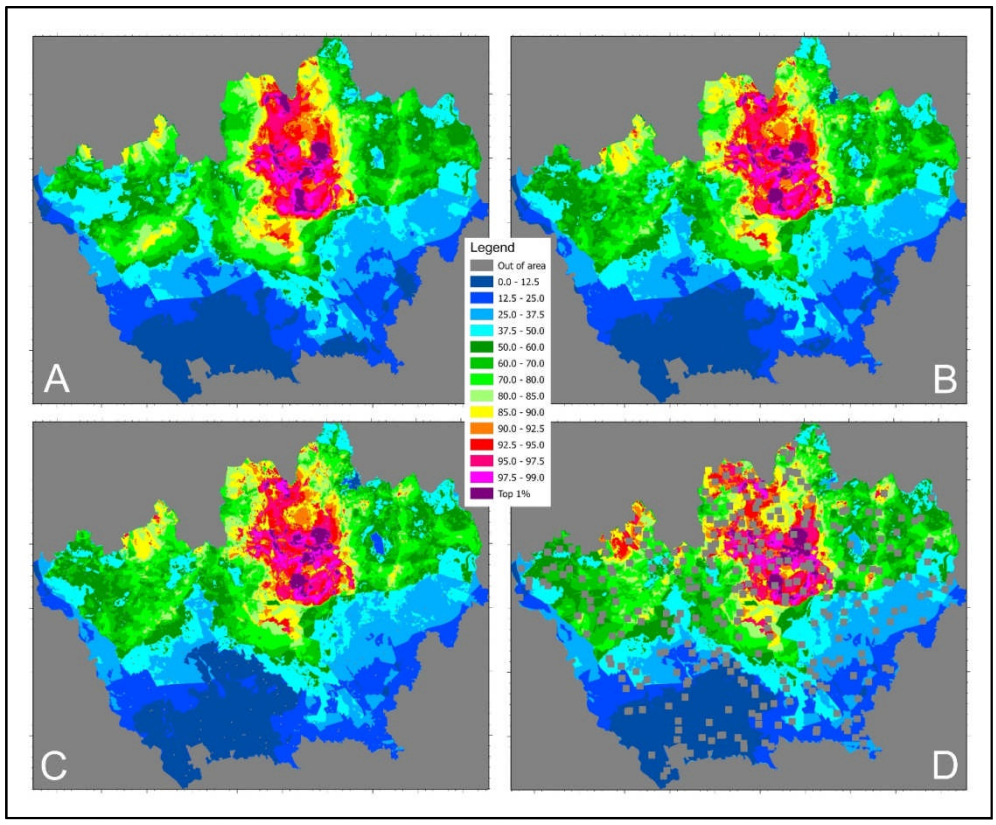

Figure 2: *Prediction patterns* for study areas. XLR *prediction pattern* for **Sa1** are in (A); for **Sa3** in (B), for **Sa9** in (C); and for **Sa51** (D). Explanation is in text.

4.2 Prediction Patterns

ELR *prediction patterns* were obtained for the four training areas: **ELR_Ta1_133_RLS_dvinpr** to **ELR_Ta51_133_RLS_dvinpr**. They are named using the abbreviation of the mathematical model, ELR, followed by the identification of the training

area, the DSP of the impacted wells, and the list of categorical and continuous ISPs, as upper case and lower case abbreviations, respectively.

Because the images of the training areas consist of groups of single pixels or of small pixel neighbourhoods, their colour display is not informative, except for **Ta51** where the neighbourhoods are larger, as done in Figure 4A and 4D. However, when using the modelling statistics from the training areas to extend it to the corresponding study areas, different *prediction patterns* are generated as follows (**X** indicates extension, and **Sa** the study areas): **XLR_Sa1_133_RLS_dvinpr** to **XLR_Sa51_133_RLS_dvinpr**. They are displayed in Figure 2. There we can see the similarity of selected ranks in the patterns and also some minor differences of particular relevance for the 10% highest classes of ranks. Note that the legend's ranked classes are wider for lower ranks and narrower for higher ranks. The pseudo-colouring scheme goes from cold to warm colours but the class boundaries remain fixed in order to facilitate recognition and allow comparison.

We can observe the following characteristics in Figure 2: (1) greater compactness of colours in the prediction pattern in Figure 2A; (2) strong similarities of higher classes for the top 1% (purple) and top 5% (red to purple); (3) patches of high value colours to the west and the east in Figures 2B, C and D (yellow to red); (4) altogether rather similar prediction patterns for the four study areas with a large hot area to the NNE of the city of Milan, the industrial zone.

At this point it becomes instructive to ask: How good are the *prediction patterns* as predictor of areas of future impacted wells? How stable and how certain? Clearly it has to depend on how good predictors are the patterns generated from the respective training areas, **Ta1** to **Ta51**. So far we can only consider fitting rates of the impacted wells within the ranks generated using them as DSP. They do not provide any information on predictive capability. To obtain some measure of effectiveness in predicting we can use strategies of blind testing via iterative cross-validation of *prediction patterns*.

4.3 Iterative cross-validations

A number of strategies can be formulated all based on pretending to ignore some of the sample points available. In our case they are the 133 impacted wells (pixels or pixel neighbourhoods) and are critical for establishing the spatial relationships between DSP and ISPs. First we use all the 133 wells to generate the best or most informed *prediction patterns*, as shown in Figure 2. Then we repeat the analyses by pretending not to know some relevant numbers of impacted wells.

For instance, in case of only a few tens of wells available, we can sequentially exclude one and use the remaining n-1 for modelling new *prediction patterns*. Then we validate them with the prediction rates corresponding to the excluded well locations. We can iterate the process n times to obtain n prediction rates, one per excluded impacted well. The distribution of the rates as ranks throughout those of the *prediction pattern* is obtained as a table and a corresponding prediction-rate histogram or cumulative curve. In Figure 3, for instance, we have used the strategy of excluding 8 wells from the 133, using the remaining 125 for modelling in the iterative cross-validation process. This generates 16 prediction patterns each validated by 8 prediction rates. Another strategy used was of selecting at random 93 wells out of the 133 (about the 70%) and repeating the modelling 16 times for the four training areas. That did generate results very similar to the ones in Figure 3.

In the illustration the prediction-rate curves were calculated for the four training areas, **Ta1** to **Ta51**. The diagram in Figure 3A shows the relative proportion of training areas

ranked as vulnerable in decreasing order on the horizontal axis and the corresponding cumulative proportion of impacted wells in the class on the vertical axis. Immediately it can be seen that the **Ta1** prediction-rate curve is very steep. For **Ta3** to **Ta51** the curves are increasingly shallower. In addition, the histograms in Figure 3B, that consider just the top 20% ranks in classes of 4% of training areas, show a monotonically increasing histogram for **Ta1**, red columns, representing an acceptable classification of vulnerable areas. The blue columns, instead, show for **Ta3** a non-increasing histogram, as well as for the corresponding histograms for **Ta9** and **Ta51,** not shown here.

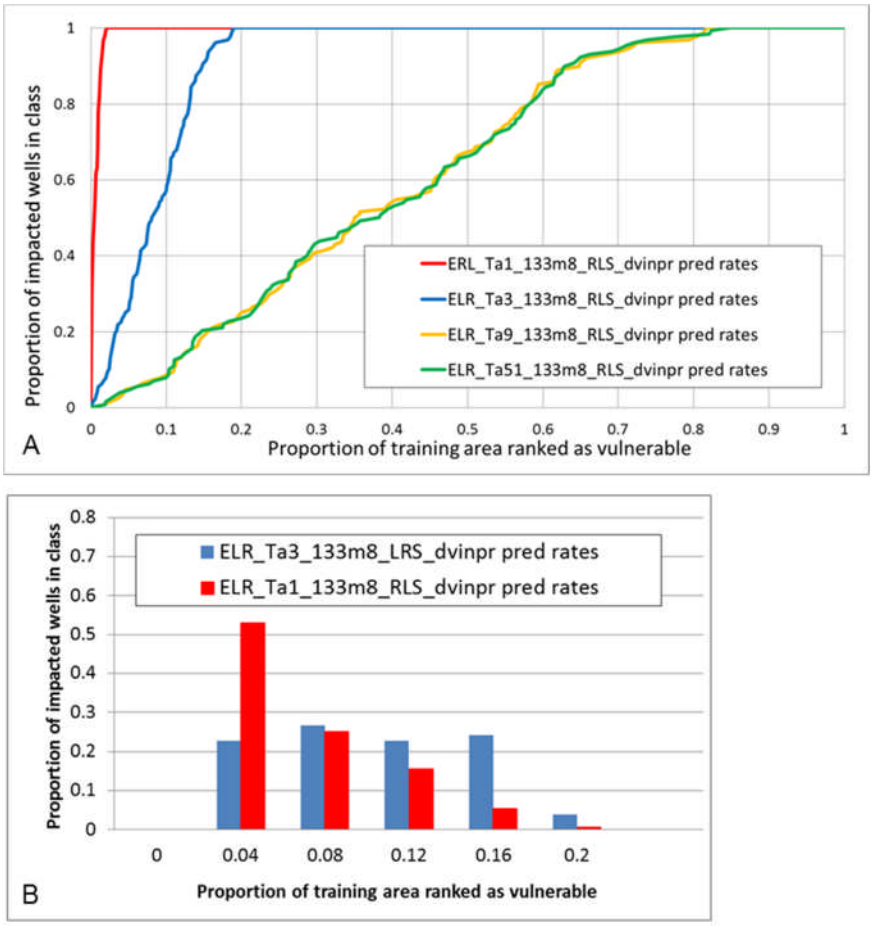

Figure 3: ELR prediction-rate curves and histograms. In (A) for the four training areas **Ta1** to **Ta51**, obtained using the iterative cross-validation strategy of sequential exclusion of 8 impacted wells out of 133. The process generates 16 *prediction patterns* each cross-validated by eight excluded wells. The curves for **Ta9** and **Ta51** reflect nearly random distribution of ranks. In (B) are the histograms for **Ta1**, monotonically increasing, and **Ta3**, non-increasing.

We can observe that the 100% of impacted wells (vertical axis) are ranked within the top 2% for **Ta1**, the top 19% for **Ta3**, the top 82% for **Ta9** and the top 85% for **Ta51**

(horizontal axis). The histogram for **Ta1**, in Figure 3B, indicates a good classification, with the higher equal area classes in 4% intervals and monotonically increasing columns towards higher ranks.

Which curve is the one representing a better prediction pattern? Is that because it contains more impacted wells at higher ranks? Or is it because the *prediction pattern* has less uncertainty associated?

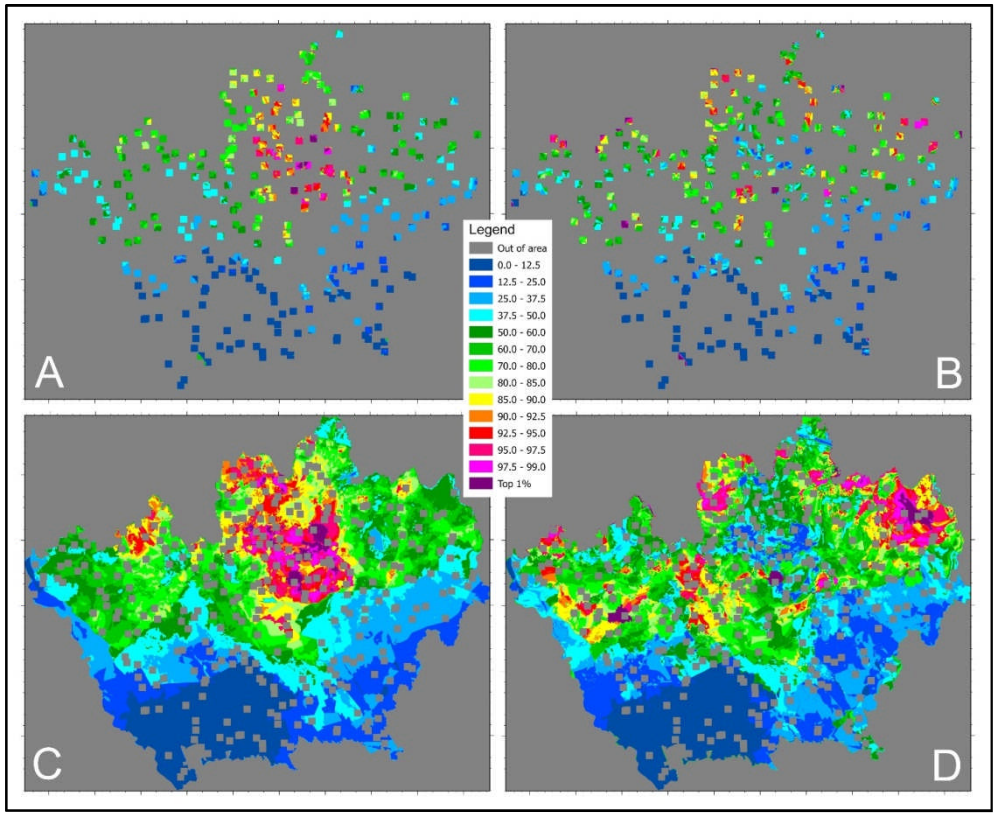

Figure 4: *Target* and *uncertainty patterns* for training and study areas. ELR *target pattern* for **Ta51** is in (A); *uncertainty pattern* for **Ta51** in (B), XLR *target pattern* for **Sa51** in (C); and *uncertainty pattern* for **Sa51** in (D). Explanation is in text.

4.4 *Target* and *uncertainty patterns*

To answer this type of question we can proceed to generate the 16 *prediction patterns* out of the 133 minus 8 x16 strategies for the four training areas to generate *target* and *uncertainty patterns* as done in Figures 4A and 4B for **Ta51**. In Figure 4C and 4D for **Sa51** we have extended the statistics obtained from **Ta51** to **Sa51**. The *target patterns* in the illustration have been obtained from the set of 16 *prediction patterns* for **Ta51**. The median of the 16 ranks (of the 16 patterns) is selected for each pixel and becomes the rank of the *target pattern*, as shown in Figures 4A and 4C.

Visually, the *target pattern* is very similar to the *prediction pattern* (compare Figure 4C and Figure 2D). However, the range of the 16 ranks represents an estimation of the uncertainty in the ranking. The wider is the range the more uncertain can the *target patterns* be considered and consequently also the initial *prediction pattern*. The same legend is being used for the illustrations of *target* and *uncertainty patterns* in Figure 4. Obviously, the significance of the uncertainty ranks, in Figure 4B and 4D, is the reverse the one of the target ranks, in Figure 4A and 4C For instance, we have selected the 50% the lowest values from the *uncertainty pattern* to identify all the ranks in the *target pattern* (or the *prediction pattern*) corresponding to lower uncertainty. This produced the *combination patterns* shown in Figure 5. Observe in Figure 5D the 50% *combination pattern* for **Sa51**, obtained combining the patterns in Figure 4D, *uncertainty*, with the one in 4C, *target pattern*. We have applied median and range statistics here, due to its robustness, differently from the previous work [8] where the more sensitive mean and variance were tentatively used.

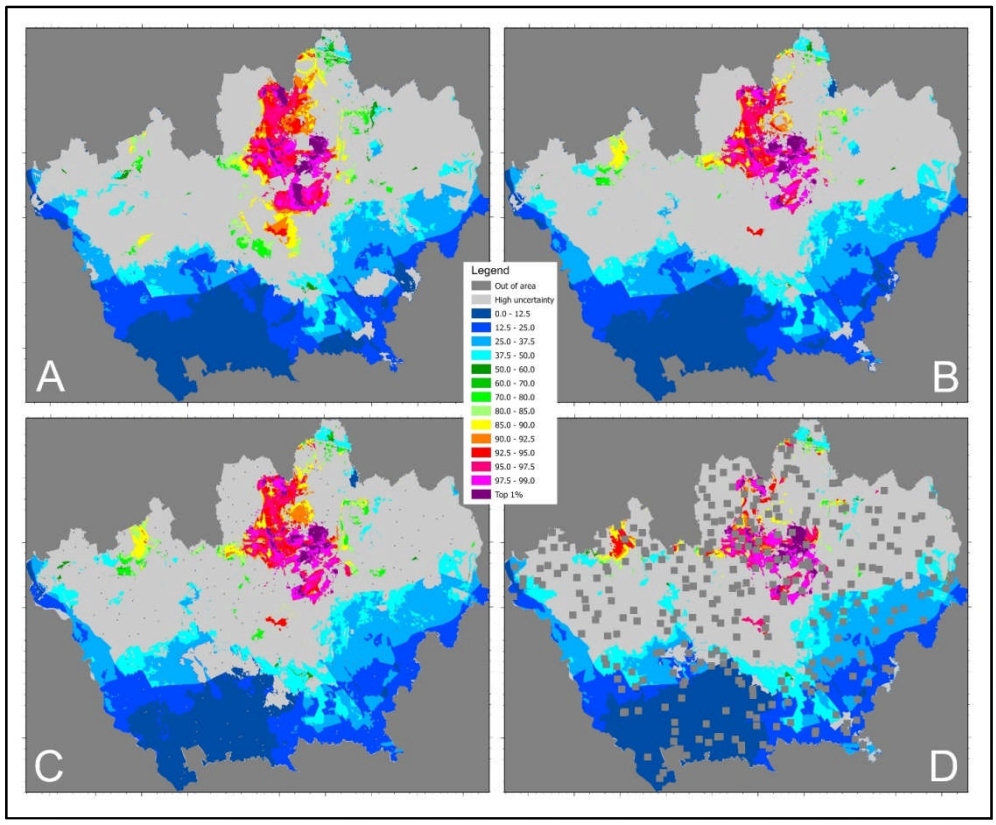

Figure 5: XLR 50% *combination patterns* of *uncertainty* and *target patterns* for four study areas. In (A) is for **Sa1**, in (B) for **Sa3**, in (C) for **Sa9** and in (D) for **Sa51**. Explanation is in text.

Evaluating and comparing *uncertainty patterns* is a complex issue, worthy of extensive research. The four 50% *combination patterns* in Figure 5 show strong differences in the distribution of the ranks. All our modelling is based on relative measures and on

ranking statistics. What we could do in a rough empirical manner is to compare ranks for a given top set or class. For instance, by arbitrarily selecting the top 10% ranks, we can evaluate the loss of the top target rank area in the 50% *combination patterns*. They mask the part of the *target patterns* with higher uncertain ranks (belonging to the higher 50% of the *uncertainty pattern*). For the patterns in Figures 5A to 5D, we have a relative decrease for the top 10% target ranks as follows: of 29.64%, for **Sa1**, of 46.54% for **Sa3**, of 39.88% for **Sa9**, and finally of 61.33% for **Sa51**. This reveals a relatively greater uncertainty affecting **Ta3** to **Ta51** than that affecting **Ta1**. In other words, it indicates a better ranking for **Ta1** with less relative uncertainty.

5 CONCLUDING REMARKS

This contribution reanalyses a database constructed for assessing aquifer vulnerability to nitrate pollution around the city of Milan. Favourability function modelling with the empirical likelihood ratio is applied to four training areas extracted from the database. The areas simulate increasingly wider spatial support of water well values of nitrate concentration. The statistics resulting from the *prediction patterns* and their cross-validations is extended from the training areas to the surrounding study areas. In this way the *uncertainty patterns* are computed and compared. This is done in parallel with the predictive capability of the *prediction patterns* represented by prediction-rate curves and histograms.

Increasing the extension of spatial support from one training areas to another leads to *prediction patterns* that appear similar but imply worsening of prediction quality, the loss of monotonically increasing character of ranking, and an increase of the uncertainty associated with the *target* and *prediction patterns*. The 20 m spatial support and corresponding training area appear preferable to the wider ones. Such aspects must be considered when applying prediction models to areas for which the vulnerable occurrences, i.e., the impacted wells, consist of point values sampling a process. The training areas being modelled have to be limited to the neighbourhoods of the points, water wells, because their spacing and spatial support are visibly affecting the resulting *prediction patterns*. The many implications of the results of studying the Milano study area make it worth further attention. More generally, research issues worthy of consideration, beside the assessment of spatial support, are: comparisons of *prediction patterns* and prediction-rate curves and comparisons of relative uncertainty of *target patterns*.

REFERENCES

[1] Masetti, M., Poli, S. & Sterlacchini, S, The use of Weights-of-Evidence modeling technique to estimate the vulnerability of groundwater to nitrate contamination. *Natural Resources Research*, **16(**2), pp. 109-119, 2007.

[2] Alberti L, DeAmicis M., Masetti M. and Sterlacchini S., Bayes rule and GIS for evaluating sensitivity of groundwater contamination. *Proceedings of the IAMG 2001 Conference*, 6-12 September, 2001, Cancun, Mexico, 18 pp., 2001, CD, http://www.iamg.org.

[3] Masetti, M., Poli, S. & Sterlacchini, S., Aquifer vulnerability assessment using weights of evidence modelling technique: application to the Province of Milan, northern Italy. *Proceedings of the IAMG 2005 Conference: GIS and Spatial Analysis*, 21-25 August, Toronto, Canada, **1**, pp. 499-504, 2005.

[4] Masetti, M., Sterlacchini, S., Ballabio, C., Sorichetta, A. & Poli, S., Influence of threshold value in the use of statistical methods for groundwater vulnerability assessment. *Science for the Total Environment*, **407**, pp. 3836-3846, 2009.

[5] Masetti, M., Sorichetta, A., Ballabio, C., Sterlacchini, S & Pozzi, M., 2011, Using different thresholds in assessing ground-water vulnerability through statistical methods. *Proceedings of the IAMG 2011 Conference*, Salzburg, Austria, September 5-9, 2011, pp. 959-969, 2011.
[6] Sorichetta, A., Masetti, M., Ballabio, C., Sterlacchini, S. & Beretta, G. P., Reliability of groundwater vulnerability maps obtained through statistical methods. *Journal of Environmental Management*, **92**, pp. 1215-1224, 2011.
[7] Sorichetta, A., Masetti, M., Ballabio, C. & Sterlacchini, S., Aquifer nitrate vulnerability assessment using positive and negative weights of evidence methods, Milan, Italy. *Computers & Geosciences*, **48**, pp. 199-210, 2012.
[8] Fabbri, A. G., Cavallin, A., Masetti, M., Poli, S., Sterlacchini, S. & Chung, C.-J., Spatial uncertainty of groundwater-vulnerability predictions assessed by a cross-validation strategy: an application to nitrate concentrations in the Province of Milan, northern Italy. In, Brebbia C. A. (ed.), *Risk Analysis VII & Brownfields V*, Southampton, WIT Press, PI 497-514, 2010, also WIT Transactions on Information and Communication Technologies, vol. 43, © WIT Press www.witpress.com , ISSN 1743-3517 (on-line) doi: 10.2495/RISK100421.
[9] European Community, Council Directive 91/676/EEC of 12 December 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources, (Nitrate Directive) OJ L 375, 31.12.1991: pp. 1-8, 1991.
[10] Chung, C. F. & Fabbri, A. G., The representation of geoscience information for data integration. *Nonrenewable Resources*, **2**(2), pp. 122-139, 1993.
[11] Chung, C. F. & Moon, W. M., Combination rules of spatial geoscience data for mineral exploration. *Geoinformatics*, **2**, pp. 159-169, 1991.
[12] Chung, C.F., Using likelihood ratio functions for modelling the conditional probability of occurrence of future landslides for risk assessment. *Computers & Geosciences*. **32**, pp. 1025-1065, 2006.