Characterization of groundwater nitrate long term trends at regional scale using statistical and geostatistical tools

Bernard Bourgine[1], Benjamin Lopez[2] and Nicole Baran[3]

Abstract European Water Framework Directive requires the Member States to characterize the quality of their ground waters, to identify trends and to take measures to achieve a good chemical and quantitative status by 2015. This is why, together with the Seine-Normandy Basin water agency, the BRGM (France) conducted a study on the whole Seine-Normandy Basin (95,300 km²) in order to understand the current status of groundwater nitrate contamination and describe trends at punctual and regional scales. Time variograms of water table level and nitrate concentrations (around 500 time series collected from the national database in each basin for each variable) were automatically computed and fitted and a hierarchical ascendant classification (HAC) of the variogram models was performed. Results were combined with hydro-geological map, to obtain "reference sectors" for spatial trend detection. A CUSUM algorithm was applied to determine the date of the major change in statistic properties of nitrate concentrations time series. The Man Kendall and the Kendall Regional non parametric tests, coupled with the Sen slope estimator procedure, were used to estimate the "post-change" trends both at punctual and regional scales. At last, cross covariances between water table level and nitrate concentration time series were computed in order to i) determine correlations, ii) detect possible time shift between time series and iii) identify any spatial distribution in this shift. The application of statistical and geostatistical semi-automated procedures appears relevant to describe groundwater nitrate concentration trends at different scales. The proposed methodology is robust, allows the rapid processing of a large number of data and helps stakeholders to achieve the requirements of the Water Framework Directive.

[1] BRGM, (French geological survey) BP 6009, 45060 Orléans Cedex 2 - France b.bourgine@brgm.fr

[2] BRGM, (French geological survey) BP 6009, 45060 Orléans Cedex 2 - France b.lopez@brgm.fr

[3] BRGM, (French geological survey) BP 6009, 45060 Orléans Cedex 2 - France n.baran@brgm.fr

1 Introduction

European Framework Directives (91/676/EEC; 2000/60/CE; 2006/118/CE) require the Member States of European Union to take actions to achieve a “good” chemical and quantitative status of their groundwater resources by 2015. More specifically, European countries have to determine a nitrate (NO_3) threshold and to identify ground waters where this threshold is exceeded or will be exceeded if no actions is undertaken (for instance France lowered down the NO_3 threshold to 40 mg/L). In case of failure to meet these requirements, the concerned countries will have to justify.

Together with the Loire-Brittany Water Agency and with the Seine-Normandy Water Agency, the BRGM conducted an exploratory study on the whole Loire-Brittany basin (157,000 km²) and on the whole Seine-Normandy Basin (95,300 km²) [1]. The main objective of this study was to collect information about the age of the water and to characterize past evolutions of ground waters quality, in order to understand the current state of contamination and to predict (model) its possible evolution.

This work involved (i) the determination of water resources age, (ii) collection of nitrate and water table level time series, the homogenization and the verification of these data, (iii) the use of statistical and geostatistical tools to determine the state of water resources and the spatial or temporal evolutions, (iv) the definition of the typical behavior and evolution, by comparison of piezometric and NO_3 signals with available data such as hydrogeological context and agricultural practices, (v) the comparison and the explanation of the whole set of information and data available.

This paper focuses on the Seine-Normandy basin and on the statistical and geostatistical aspects, from data analysis to results.

2 Methodology

The work was divided in three main tasks. Firstly the basin was divided in “homogeneous” sectors in terms of hydrodynamic behavior and lithology. This analysis was performed using water table time series and geological maps. Secondly, in each of these sectors, the historic nitrate concentration evolutions were studied in order to characterize the presence of annual or multi annual cycles, the possible change in the statistical properties of time series, trends and the possible correlation between nitrate concentration and water table level. Thirdly explanatory factors were reviewed in order to help interpreting the results: water dating, agricultural load and climatic conditions in particular. This later task will not be detailed here.

3 Studied area and used data

The studied area is the Seine-Normandy Basin. This basin mainly includes sedimentary formation from Cenozoic and Mesozoic (Figure 1). The following data were collected from the French national ground waters database “ADES”:

1. 1,750,431 water table level data spread across 558 time series, collected from 1948 to 2010 on daily or weekly basis.
2. 77,978 NO_3 concentrations, from 5962 nitrate time series, from 1971 to 2010. Nitrate concentrations are available at a sampling rate of 1 to 6 months.

Lastly 1/1,000,000 geological maps, completed with information on the karstic nature of aquifers, were used to achieve the sectors delimitation.

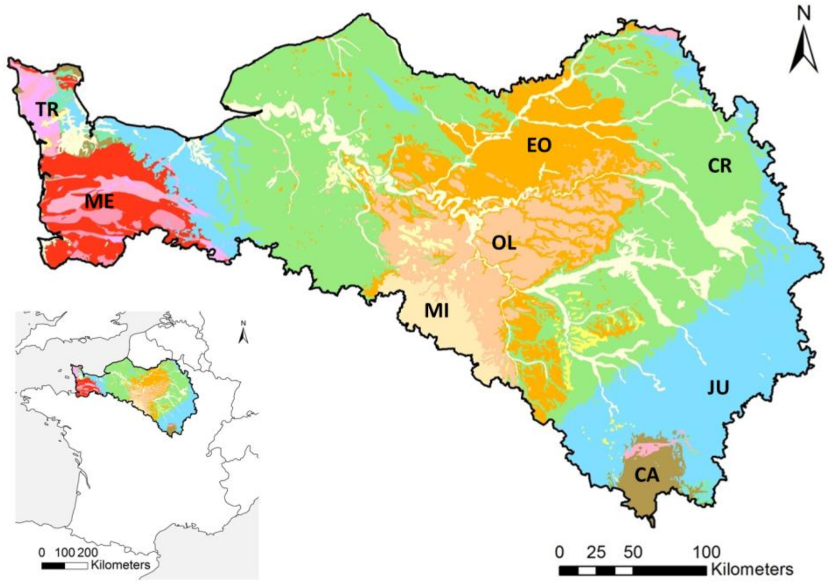

Figure 1 : Studied area. EO: Eocene and Paleocene MI: Miocene, OL: Oligocene; TR: Triassic, JU: Jurassic, CR: Cretaceous). Metamorphic rocks (ME) are also present in the west border, as well as Granites from the Cambrien in the south (CA).

4 Classification of aquifer hydrodynamic behavior using temporal variograms of water table level

In the studied area, hydrogeologists already know the main behavior of aquifers at least from a qualitative perspective. However, there are local variations and the quantitative description of hydrodynamic behavior is not known or available everywhere. For example the amplitude of annual cycles compared to that of multi annual variations on every piezometer is not known. This is why we computed the temporal variogram of each piezometer, and tried to group piezometers into homogeneous classes displaying the same variogram behavior.

After removing "incorrect" and "uncertain" data (this information being given by the ADES database), time series were regularized at the step of 15 days by averaging water table levels in a non-overlapping moving window. Temporal variograms were then computed on time series with at least 4 years of measurements. A total of 556 variograms was obtained. Figure 2 shows four typical behaviors that can be observed: (a): annual cycle with no or little drift (approx. 15% of the cases); (b) annual cycle and multi-annual cycle (approx. 20% of the cases), (c) moderate annual cycle and multi annual drift (approx. 35% of the cases); (d) no annual cycle but long term drift or cycle (approx. 15% of the cases).

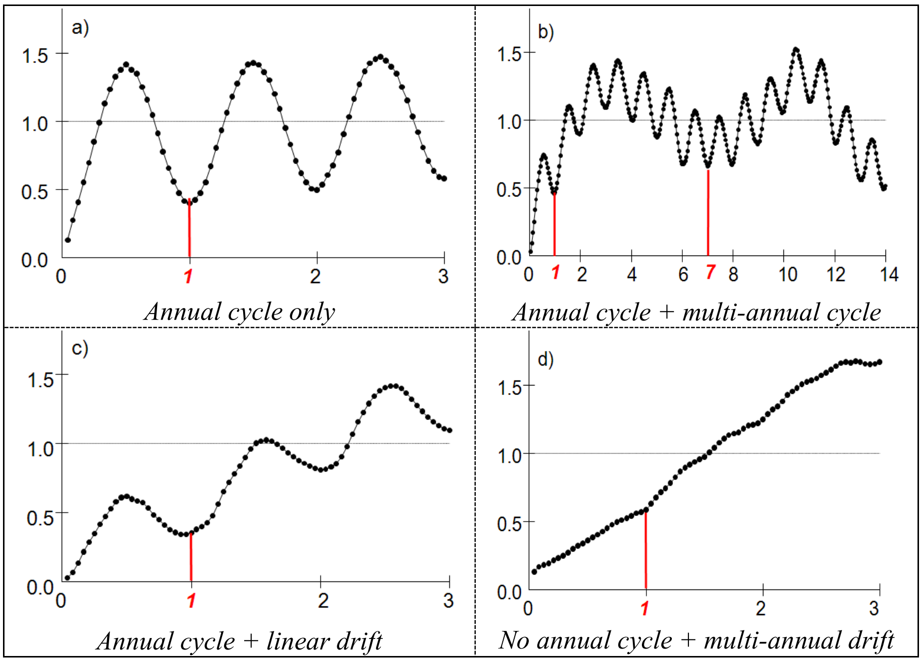

Figure 2: Typical temporal variograms of water table level (X axis = time in year, variogram axis normalized by variance of data).

In a second step these variograms were normalized by the variance of data and fitted by least squares using an automatic procedure. The fitted model was as described by Equation (1):

$$\gamma\ t\ = C_0 + C_1 Sph\ a_1, t\ + C_2(1 - Cos\ \omega_2 t\ ) + k_3\, t. \quad (1)$$

In this Equation, C_0 is the nugget effect, *Sph* a spherical component, *Cos* a cosine component. The spherical component has a range a_1 which was allowed to vary within interval 185 days +/- 35 days. This component is used to model the short term behaviour of the water table levels. The second component is a cosine component. It is used to model the annual cycle and has a parameter $\omega_2 = 2\pi/365$. C_1 and C_2 are the sill of the components and were allowed to vary independently between 0 and 100% of the variance. The last component is a linear one, with a slope k_3. It is used to model the multi-annual drift or the first years of multi-annual cycles. The objective was mainly to split the behavior in i) a short term component, ii) an annual component and iii) a multi-annual component, the multi-annual behavior being synthetized by a unique linear drift term fitted to the first 2 years. Therefore, even if the experimental variogram was computed for time distances greater than two years the model was fitted using only the first two years of the experimental variogram. 441 variograms were fitted among the 546 experimental variograms (those corresponding to time series of four years or more, i.e. to an experimental variogram up to two years).

The fitted parameters C_0, C_1, C_2 and k_3 were then entered in a HAC (hierarchical ascendant classification) analysis in order to obtain a classification into homogeneous sets in terms of variogram behavior. The automatic classification was then checked and modified manually to account for experimental variograms which were poorly fitted and to better account for the long term behavior.

The result of this work was a classification of time series in 7 classes regarding hydrodynamic behavior (Figure 3).

This classification agrees with the knowledge of the various aquifers of the basin. For example the hydrodynamic of the Miocene aquifer is characterized by no annual cycle and long term multi-annual components, whereas Cretaceous formations at the east of the basin show strong annual cycles with no long term component (Figure 3). Moreover, results highlight the difference in hydrodynamic behavior between the east side and the west side Cretaceous aquifers: the east side shows annual cycles whereas the west side is dominated by multi-annual cycles. In some aquifers as in the Oligocene and in the Miocene, spatial variations are clearly shown which allows splitting some aquifers in two parts according to their behavior.

At last these classes were combined with the geological map in order to get reference sectors within which the lithology is the same and the hydrodynamic behavior is homogeneous. This led to divide the whole basin in 54 sectors.

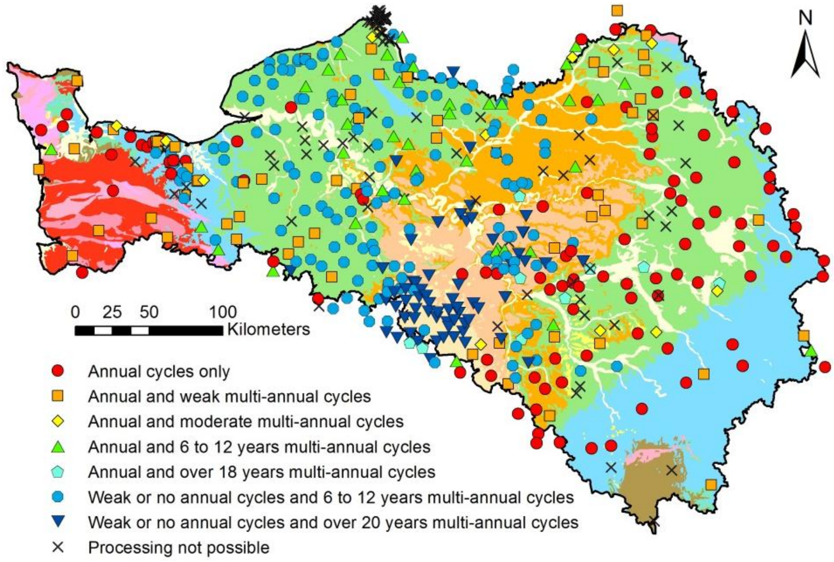

Figure 3: Hydrodynamic behavior of water table time series.

5 Classification of nitrate concentration time series using time variograms

In a similar way as for water table level, the temporal variograms of nitrate concentrations were computed for each of the 542 time series counting at least 35 measures and monitored during at least 2 years. Nitrate concentrations being generally measured every 2 to 6 months, the variograms were computed with a lag of 3 months.

A model was fitted on each experimental variogram, according to Equation (2):

$$\gamma\ t\ = C_0 + C_1 Sph\ a_1, t\ + C_2(1 - Cos\ \omega_2 t\ ) + C_3 Sph\ a_3, t + k_4\, t^{\alpha}. \quad (2)$$

The expression of the fitted model is similar to that of Equation (1), excepted that the drift can be a power model with exponent α, to account for the fact that non linear drift can be observed in nitrate concentration. The third component is a spherical model with a range that can reach 6 years.

By the way 481 time series were fitted (see examples Figure 4). A HAC on the fitted parameters (C_i, a_i, ω_2, a_3, k_4, α) was attempted but did not give consistent results, probably because it is difficult to distinguish the contribution of the two last components in case of a drift when fitting the model only on the first two years (it was impossible to fit on larger time period due to the fact that nitrate time series are generally shorter that water table time series). Instead of working with model parameters we used the values of the model at characteristic times. The following times were selected: 10 days, 6 months, 1 year, 1.5 year and 5.5 years (Figure 4b). More precisely, we worked on the 4 following parameters:

- P1 = γ(10 days)
- P2 = γ(1 year) - γ(0.5 year)
- P3 = γ(5.5 year) - γ(1.5 year)
- P4 = P3 – [γ(1.5 year) - γ(0.5 year)]

P1 corresponds to the nugget effect, P2 to the amplitude of the annual cycle, P3 to the amplitude of multi-annual trend, and P4 to the difference between "long term" trend and annual fluctuations.

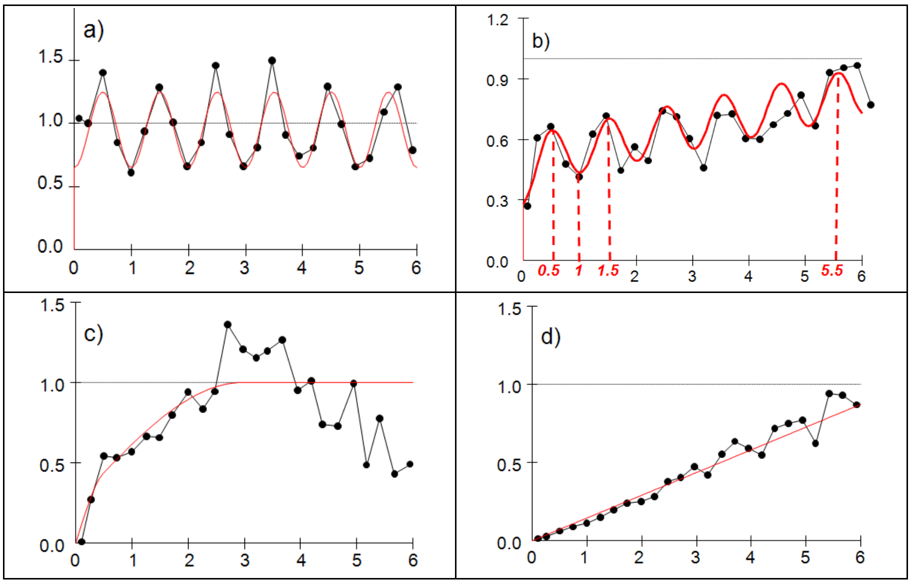

Figure 4: Examples of variogram of nitrate concentration and proposed fits in red (X axis in year, variogram axis normalized)

A HAC was performed on the 4 Pi parameters and modified manually as for the water table level case. This led to the definition of 7 classes, based on the relative amplitudes of the different components (annual, multi-annual, drift). Figure 4 shows four typical behaviors that can be observed: (a): annual cycle with no or little drift; (b) annual cycle and long term drift, (c) no annual cycle, no long term drift; (d) long term drift or cycle.

As it can be seen on Figure 5, variograms of the nitrate time series are not uniformly distributed on the entire Seine-Normandy basin. Due to insufficient data, it was not possible to calculate variograms in some zones. Thus, nitrate concentration behavior is not described in the tertiary aquifers on the center of the basin or in the chalk aquifers on the west side for instance. It is also to notice that points close together show sometimes different behaviors though the nitrate concentration evolution is homogeneous in the whole aquifer. In fact the length and the sampling rate of nitrate time series vary greatly and are lower than that shown by water table time series. As a result, nearly 50% of time series were classified as “no annual but over 6 years multi-annual cycle or drift”. This result is probably biased because many time series are measured with a 6 month frequency, which makes impossible the identification of an annual cycle.

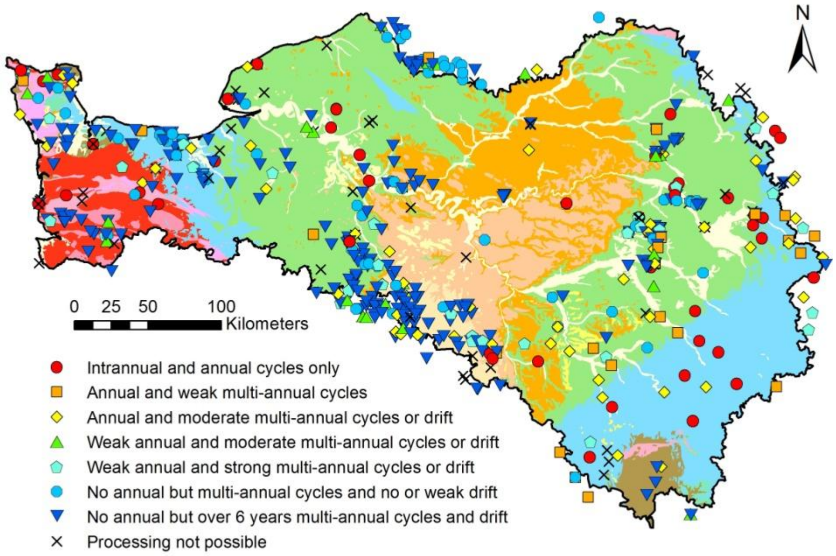

Figure 5: Groundwater nitrate concentration behavior.

After combining these classes with the former classification based on hydrodynamic behavior and lithology, a new classification into 64 zones was obtained (see Figure 6).

6 Characterisation of temporal trends in nitrate concentrations

The next step of the study was the detection of “recent” trends in nitrate concentration evolutions, in each sector. This task was divided in 2 parts: (1) to detect the date of change in nitrate evolution behavior and thus identify the last parts of time series that show homogeneous statistical properties, if any, and (2) to compute the trend after this date at punctual and spatial scale. This work was done for each time series and also at the scale of the 64 sectors.

The CUSUM test [2] was used to detect change in statistic properties of time series. 5862 time series were processed, among which 1987 showed a significant change. 80% of the changes in statistical behavior occur within the period

February 1993 - June 2004. This could be related to the implementation of both the agro-environmental measures and the European common agricultural politics but also with climate conditions, the main possible factors that can explain the observed nitrate evolutions.

Once the date of change of statistical properties of time series determined, the non-parametric Mann-Kendall (MK) test, a robust statistical trend detection test [3], was used to estimate the "recent" trends for each time series from the date of change to 2010 or from 1993 (1st decile value of the whole date of change) to 2010 if there was no significant change.

Following the Water Framework Directives, a "regional" trend was also computed. The Kendall Regional test (KR) [4] was applied to assess the regional trend. The KR test consists in the creation of a virtual time series defined as the grouping of all the time series contained in the same zone. In our case the zones were the 64 sectors defined in section 5.

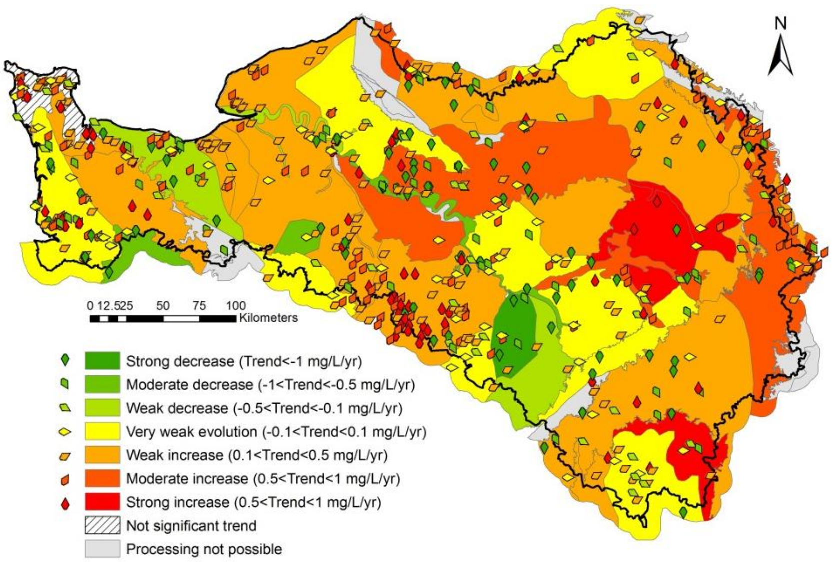

Figure 6: "Recent" nitrate concentration trends calculated at punctual scale by the Mann-Kendall test (symbols) and at the 64 zones scale by the Kendall Regional test (colored areas).

Results in Figure 6 show that the two tests lead to similar result when all the times series within a given zone have the same sampling rate. However, both tests can differ when within a given zone, some time series show an increase of nitrate concentration and others a decrease, both having similar variogram. Differences can also be explained by the higher number of nitrate time series included in the Kendall Regional test. Some nitrate time series that could not be computed by the Mann-Kendall test (not enough data) or that do not show significant trend when treated in isolation, could in fact be taken into account in the Kendall Regional test.

7 Correlation between water table level fluctuations and nitrate concentrations

Many authors report a correlation between water table and nitrate concentration, which could be explained by processes of dissolution/remobilization of nitrate stocks in soil and in the unsaturated zone. In this study, the water table and nitrate time series are never available at the same point. It is therefore impossible to compute directly the correlation between the two signals. This is why we tried to characterize a possible correlation by the computation of cross covariances.

The first step of this work consisted in creating pairs of time series (water table level / nitrate concentration) that were not too far one from each other – a maximum distance of 10 km was chosen, that were surveying the same aquifer, and that were not separated by any major hydrogeological discontinuity. A total of 2157 pairs was obtained. However, due to uncertainties about the aquifer assigned to the time series, it is not absolutely sure that all pairs are relevant.

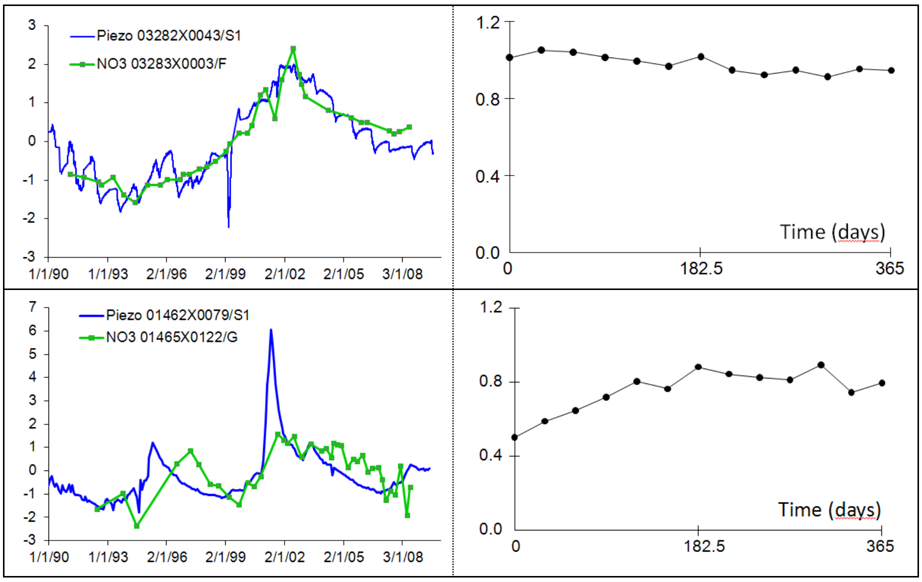

Figure 7: Example of two cross covariances between water table level and nitrate concentration. Left: centered-normalized time series, right cross covariance.

The next step consisted in a data preprocessing in order to obtain measurements at the first day of each month. For the water table level time series, data were averaged in a window +/- 15 days around the first day of each month. For nitrate time series, which sampling rate is around 1 month in the best case, the date of measurement was shifted to the nearest beginning of month. Time series were centered and normalized.

Cross covariances where then calculated. The time at which the maximum of cross covariance was reached was determined automatically and analyzed. Results were only considered when the average number of pairs of measurements involved in the cross covariance calculation was greater than 10, i. e. for 1850 pairs of time series. Among these pairs 40% appears as uncorrelated (maximum cross covariance lower than 0.4), 30% as synchronous, and the remaining 30% shows a time shift of few months to 10 months (examples in Figure 7). Type of correlation (positive or negative) and time shift were classified, plotted on a map and analyzed (Figure 8).

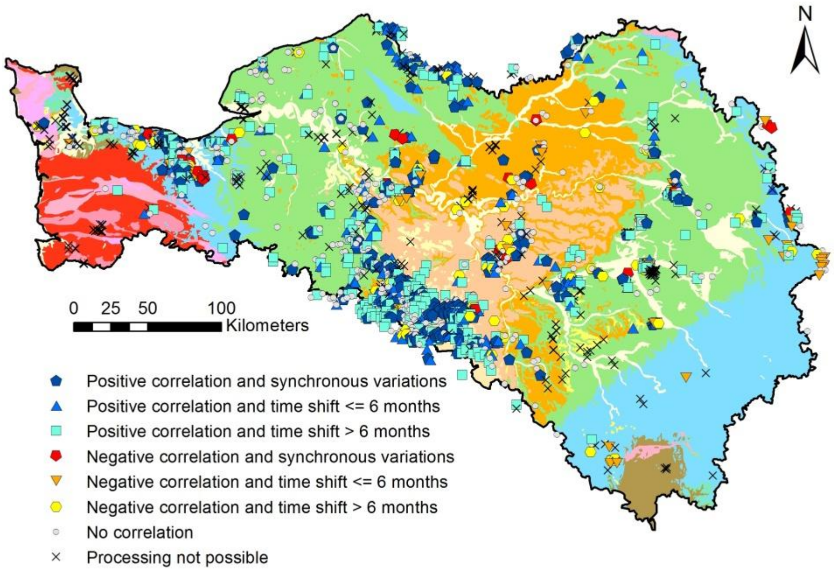

Figure 8: Results of cross covariance calculation for 1850 pairs of time series (water table and nitrate).

Such as the nitrate time series that are in favor of statistical treatments (length and sampling rate), pairs of time series are not uniformly distributed on the basin. Results show that there is no real regionalization of places where signals are correlated and places where they are not. It is also to notice that negative cross covariances were observed. In this study, many water table time series and nitrate time series show periodic behavior (mainly annual cycles). This can lead to negative cross covariances in case of annual cycles with a time shift of 6 months. Consequently, the cross covariance procedure should be adapted and developed to be efficient when applied on ground water quantity and quality data.

8 Conclusions

This work constitutes the first step of a global study on the identification and the understanding of the ground water nitrate long term trends at punctual and spatial scales. Statistical and geostatistical tools appear to be very useful to reach this goal Thanks to the ADES national database, and to the calculation of the variograms of water table and nitrate concentration time series, we have a rapid and accurate understanding of the hydrodynamic and the nitrate behaviors of the ground waters. This procedure helps to divide the whole basin in reference sectors with respect to the lithology, the nitrate contamination and the hydrodynamic behaviors of the aquifers. These new sectors constitute the work units for the spatial trend calculation. The non-parametric statistical trend tests Mann-Kendall (point scale calculation) and Kendall Regional (spatial scale calculation) are appropriated to compute large and heterogeneous ground water quality data set.
Once the nitrate trends identified, the explaining factors are investigated by coupling these information with the age of ground waters, the historical agricultural practice evolutions, the aquifer recharge and the hydrodynamic behaviors. Regarding the later factor, cross covariances appear useful to identify possible correlations between water table and nitrate concentration evolution and to highlight shifts when they occur. Finally, these statistical treatments help to reveal errors in the database such as reversed water table time series or inconsistent nitrate concentration values.

Acknowledgments: The authors thank the Seine-Normandy Water Agency and the BRGM for funding this work.

Bibliography

[1] Lopez B., Baran N., Bourgine B., Brugeron A., Gourcy L. (2012) - Pollution diffuse des aquifères du bassin Seine-Normandie par les nitrates et les produits phytosanitaires : temps de transfert et tendances. Rapport final BRGM/RP-60402-FR ; 319p.

[2] Page, E. S. (1961). Cumulative sum control charts. Technometrics 3, 1-9

[3] Kendall, M. G. and Stuart, A. (1976). The Advanced Theory of Statistics: 3, Design and Analysis, and Time Series. New York: Hafner Press.

[4] Dennis R. Helsel and Lonna M. Frans (2006) Regional Kendall Test for Trend. Environ. Sci. Tec., vol 40 no 13.