# Extracting coherent regional information from local measurements with Karhunen-Loève transform: Case study of an alluvial aquifer (Rhine valley, France and Germany)

Laurent Longuevergne,[1] Nicolas Florsch,[1] and Philippe Elsass[2]

[1] We investigate the ability of combining the Karhunen-Loève transform (KLT) with the kriging method to extract regional information from a set of point measurements. This method was applied to a set of 195 piezometric head time series over a period of 17 years from observation wells distributed within the French and German area of the Rhine valley alluvial groundwater body. Piezometric head time series are analyzed with KLT in order to highlight characteristic temporal signals, classified from the most energetic (global) to the least energetic (local) signals. The first five signals amount to 80% of the global variance of the system and are inferred to represent different hydrological contributions (exchanges with rivers and rainfall), but they also represent a significant anthropogenic component. Kriging is then used to regionalize the signals and to build a reconstruction model of the behavior of the whole aquifer containing only filtered information coming from identified source signals.

## 1. Introduction

[2] The Rhine valley aquifer is a privileged object of study for several reasons: (1) The aquifer is a huge alluvial reservoir containing the largest European groundwater resource. (2) The aquifer is well known, instrumented, and monitored; many hydrological studies have already been carried out, providing abundant additional data. (3) the alluvial aquifer is not bound to a single river, but receives various inflows: Periodic rainfall due to the semicontinental climate of the region, summer inflow from the Rhine River due to snow melting in the Alps and spring river floods from the Vosges and the Black Forest mountains.

[3] We needed to extract smoothed regional information from these point data in order to solve other problems such as the calculation of the geodetic effect of hydrological loading. We therefore decide to find out the main common source signals, which combine differently in the local measurements. The smoothing problem is then expressed as finding a mean behavior for the local combinations of these source signals. In this paper, the variation of the piezometric head of an aquifer $\Delta H$ is considered as the linear superposition of several contributions $H_k$ due to various sources with associated weight $\alpha_k$: $\Delta H = \Sigma_k \alpha_k H_k$. The blind separation of the sources refers to the problem of reproducing the contribution of source signals from piezometric head values, without making any assumption or exploiting any external knowledge, except for some math-

ematical considerations. "Blind" means that the statistics of the source signals and the way they mix are unknown. Several mathematical or physical models are used to mix contributions, for different purposes: e.g., immediate, convolution-based, or spectral-based models.

[4] In this work, the temporal contributions $H_k$ and associated weights $\alpha_k$ are determined using immediate mix model. As a consequence, the method neglect the propagation effects of the physical processes involved.

[5] The source processes are represented by a set of piezometric head time series sampled at various locations within the system. Here, we use a statistical multivariate analysis, the Karhunen-Loève transform (KLT), to sum up the observations into a set of a minimum number of temporal characteristic signals.

[6] The principle of KLT has frequently been used under different names depending on the applications and fields of study: singular value decomposition (SVD) in geomagnetics to search for particular temporal signals [*Pereira*, 2004] and empirical orthogonal function (EOF) analysis in meteorology in order to analyze spatial structure of atmospheric fields [*Grimmer*, 1963]. The link between both methods is given by [*Gerbrands*, 1981]. The method has also been applied on hydrological purpose, for example, by *Gottschalk* [1985], for the interpolation of water balance elements. An evolution of KLT method, multichannel singular spectrum analysis (MSSA) has also been used by *Shun and Duffy* [1999] in order to highlight space-time patterns of precipitation, temperature and runoff.

[7] In this work, the Karhunen-Loève transform is first used to separate space and time and define a new orthogonal basis where each observed piezometric head time series can be expressed as the sum of a small number of characteristic temporal signals $H_k$ calculated for the whole aquifer and

[1] UMR Sisyphe, Université Pierre et Marie Curie, Paris, France.
[2] Service géologique régional Alsace, Bureau de recherches géologiques et minières, Lingolsheim, France.
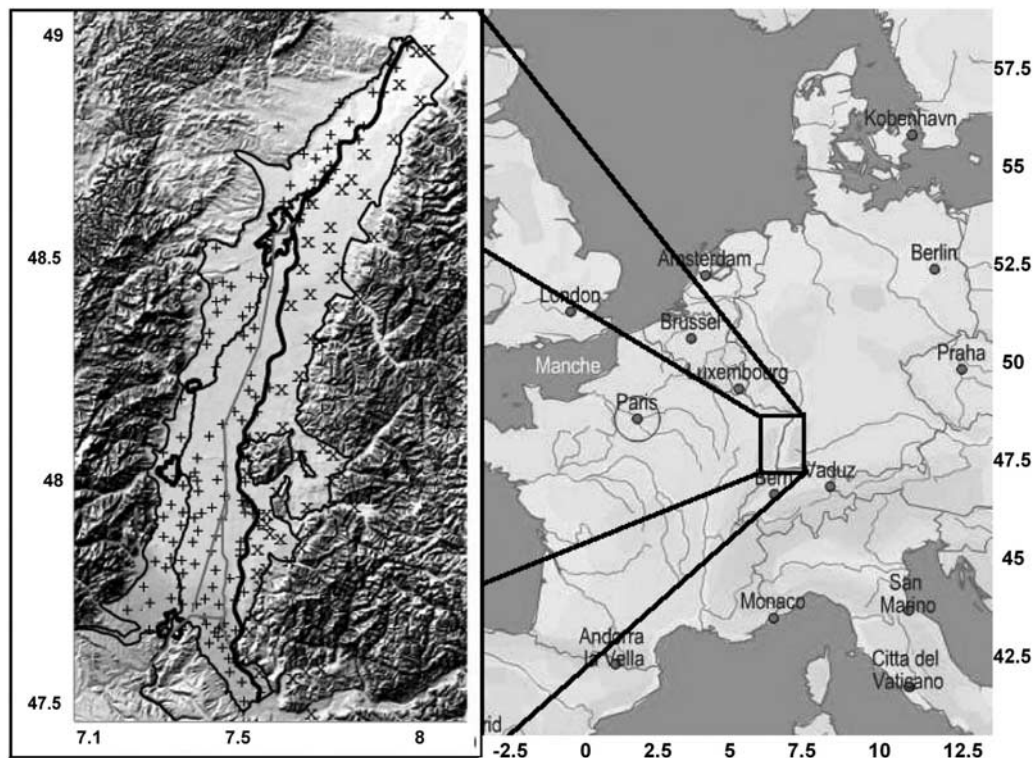
**Figure 1.** (left) Sampling area. Rhenan aquifer body is delimited in the west by Vosges mountains, in the east by the Black Forest, and in the south by Sundgau. Note the great number of rivers flowing down to the valley. Cities, from north to south, are Strasbourg, Selestat, Colmar, and Mulhouse. Major rivers are the Ill River (the western one) and the Rhine River (the eastern one). Note the straight canals between these two rivers. Pluses indicate French piezometers; crosses indicate the German piezometers. Coordinates are indicated in decimal degrees.

associated with spatial information $a_k$ for each piezometric head. This spatial information is then interpreted and regionalized by geostatistical methods in order to make a statistical reconstruction model of the model of the aquifer containing only sound information.

## 2. Data Description

### 2.1. Sampling Area

[8] The study area (see Figure 1) encompasses the southern part of the Upper Rhine valley between the Swiss city of Basel to the south and the German city of Karlsruhe to the north. The left bank of the Rhine, in the west, belongs to the Alsace region in France, the right bank in the east to the German land of Baden-Württemberg. The shallow groundwater body is contained in a mighty sand and gravel alluvial aquifer, up to 200 m thick, deposited in the Rhine graben since the end of the Tertiary by the Rhine and its tributaries [*Sittler*, 1969; *Düringer*, 1988; *Villemin et al.*, 1986]. The aquifer lies in the plain between the Vosges mountain to the west and the Black Forest mountain to the east and is up to 40 km in width.

[9] The north-south trending Vosges mountain lies perpendicular to the dominant atmospheric flow coming from the west and receives an important winter rainfall which is collected by the Vosges and Black Forest Mountain catchments. This also creates high contrasts in the distribution of rainfall in the plain.

[10] The rivers originating in the Vosges mountain generally trend west-east down to the plain, where they are captured by the Ill River south of Strasbourg and by the Rhine to the north. The Ill and Rhine rivers run almost parallel and follow the flow direction of the groundwater, which is N20E.

[11] Several canals have been laid out between the Ill and Rhine for navigation, irrigation, or hydroelectric energy. Some of them also play a role in sustaining groundwater levels in summer.

### 2.2. Sampling and Analytical Procedures

[12] The APRONA (Association for groundwater protection in Alsace) monitors 200 observation wells covering the French side of the Rhine valley alluvial groundwater body. The corresponding section in Germany is covered by 59 reference observation wells, managed by the LfU (the environment protection authority of Baden-Württemberg). The sampling rate is one measurement a week and more for automated measuring systems.

[13] The time frame for this study has been chosen to cover the period from July 1986 to December 2003, i.e., 908 weeks. Several corrections have been performed on the time series: anomalous spikes and jumps are corrected, the short gaps are interpolated, and unreliable series are deleted. In the end, only 195 observation wells have been retained for this study. In order to identify the global contributions of piezometric head variations, each piezometric head must

have the same weight in the analysis. So all time series were scaled to variance unit, in order to avoid the analysis to focus on the piezometric head recording large variations. The mean of the time series is also removed.

[14] The observation matrix $A_{ij} = x_j(t_i)$ is constructed with the selected set of time series sampled at point $x_j, j = 1 \ldots n_{var}$, $n_{var} = 195$, at times $t_i, i = 1 \ldots n_{obs}, n_{obs} = 908$.

## 2.3. Reference Contributions

[15] The statistical signals determined in this work are ultimately confronted to three physical observations, which should be representative for the whole area of study. In this way, the ''mountain river flow'' is calculated as the sum of water flow coming from Vosges and Black Forest mountain rivers and sampled at the entrance of the Rhine aquifer. ''Effective rainfall,'' given by the French Meteorological Institute, is computed as rainfall minus evapotranspiration. Finally, ''Rhine flow'' is Rhine River flow measured at the entrance of the aquifer.

## 3. Processing the Data by KLT and Kriging

[16] In geostatistics the measured value of a parameter $z(x_k)$ at a sampling location $x_k$ within a specified region $D$ is regarded as the outcome of a random mechanism, i.e., one draw from the random variable $Z(x_k)$. This argument can be extended to all other points in the region so that the regionalized variable $z(x), x \in D$ can be viewed as one draw from an infinite set of random variables known as the random function $Z(x), x \in D$.

### 3.1. Karhunen-Loève Transform

[17] This method is used to decompose observed fields into fewer sets of orthogonal, and thereby mathematically uncorrelated, fields. See *Saporta* [1990], *Wackernagel* [1995], *Wilks* [1995], *von Storch and Zwiers* [1999], and *Jolliffe* [2002] for more details. One result is that redundant information is removed.

[18] This method can be considered as a Fourier transform. However, the kernel of the transform is not calculated with complex exponential functions but with the information contained in the measurements (in this case, correlations between time series, and more precisely, the eigenvectors of the correlation matrix), in order to optimally describe the system of study.

[19] The principle of KLT can be approached from a geostatistical point of view. Indeed, the objective consists in deciphering the regionalized variable $Z(X, t)$. We assume here underlying temporal processes $Y_j(t), j = 1 \ldots n_{var}$, also called eigenvectors or characteristic signals that will set up a new orthogonal basis, associated with the spatial information $a_j(X)$ (or projections). So, $Z$ can be written as follows: $Z(X, t) = \sum_{j=1}^{L} a_j(X).Y_j(t) + \epsilon(X, t)$ where $\varepsilon$ is the residue between the regionalized variable and the explained processes $Z^* = \sum_{j=1}^{L} a_j(x)Y_j(t)$. Both spatial information and characteristic signals are determined by minimizing the variance of the estimation error $\epsilon = Z - Z^*$, under the assumption that $\{Y_j\}, j = 1 \ldots n_{var}$ is an orthogonal basis.

[20] KLT is carried out by diagonalization of the correlation-covariance matrix $^tAA$, where $^tA$ denotes the transposed of $A$. This diagonalization exhibits a set of eigenvalues $\{\lambda_j\}, j = 1 \ldots n_{var}$, classified in decreasing order, with associated eigenvectors $\{Y_j\}$. The amount of variance (energy) spanned by each eigenvector depends on the relative value of its eigenvalue with respect to the sum of all eigenvalues. It must be pointed out that this is purely statistical method, and the calculated $Y_j$ have, a priori, no direct physical meaning.

[21] The method exposed here requires the head variations to be approximated by a linear combination of various sources. The Upper Rhine valley aquifer is free; therefore this assumption is only valid in subdomains where piezometric head variations can be considered as negligible with respect to the thickness of the aquifer, so that a constant transmissivity can be defined. This is not true in the region of the aquifer south of Colmar. As for eastern and western limits, the aquifer is hosted in a graben, so the body of the aquifer becomes quickly thick enough considering piezometric head variations. This issue will be discussed later.

[22] In order to reduce the dimensionality of the system and optimally approximate the original data set, the limited number $L$ of characteristic signals is always $L \ll min(n_{var}, n_{obs})$. Some criteria (e.g., the amount of explicated variance and stability of the created subspace) allow to correctly choose $L$. Here we will examine a noise criteria.

[23] KLT has the capacity to extract physical signals from noise. An overview of this issue is described by *Broomhead and King* [1986] and *Elsner and Tsonis* [1996] for white and colored noise. Considering a set of data without noise $\overline{A}$, where time series are arranged in columns, with associated eigenvalues $\{\overline{\lambda}_j\}, j = 1 \ldots n_{var}$ of $^t\overline{A}\overline{A}$. We will use the blank noise matrix $N$, where each column $n_j$ follows a normal law (mean zero, variance $\sigma^2$). So, the observed values can be written as $A = \overline{A} + N$. In the case of no spatial nor temporal signal-noise correlation, the noise alters the original eigenvalues with a bias, and the new eigenvalues can be written as follows: $\lambda_j^2 = \overline{\lambda}_j^2 + M\sigma^2$, $M = max(n_{obs}, n_{var})$. As the series $\{\overline{\lambda}_j\}, j = 1 \ldots n_{var}$ is decreasing and converging to 0, the new eigenvalues converge to $\sigma\sqrt{M}$. In practice, such a level can never be observed, but this value can be used as a criteria to choose the size $L$ of the series, considering that the eigenvectors are correctly dissociated from noise when their associated eigenvalue satisfies $\lambda > \sigma\sqrt{M}$.

[24] A difficulty in the interpretation of the eigenvector $Y_k(t)$ may arise if two or more eigenvectors have the same eigenvalue (degenerate multiplet [see *North et al.*, 1982]). In this case, it can be demonstrated that any linear combination of the degenerated eigenvectors is a solution in the diagonalization of the covariance matrix.

[25] The investigation of the degeneracy eigenvalues requires a measure of uncertainty in the determination of the eigenvectors due to sampling. An estimation of this perturbation has been developed by *North et al.* [1982]: (1) The uncertainty on the determination of the eigenvalue $\Delta\lambda_k \approx \lambda_k(\frac{2}{N})^{1/2}$ where $N$ is the number of realization of $Z(X, t)$ (in this case, $N = n_{var}$). (2) Sampling errors on the eigenvector are of the order of the parameter $\frac{\Delta\lambda_k}{\lambda_j - \lambda_k}$, where $\lambda_j$ is the closer eigenvalue to $\lambda_k$. So eigenvectors have small sampling errors if eigenvalues are well separated ($\lambda_j - \lambda_k$ large) and correctly sampled ($N$ large, so $\Delta\lambda_k$ small).

### 3.2. Kriging

[26] The objective of kriging is to make use of the measured values at a series of neighboring locations $x_k$, to provide statistically sound estimation of the values of the

**Table 1.** Classified Eigenvalues, Amount of Explained Variance, and Associated Sampling Error

| Classified Eigenvalue[a] | Amount of Explained Variance[b] | Cumulative Sum of Explained Variance |
|---|---|---|
| 1 | $44 \pm 3.5\%$ | 44% |
| 2 | $20 \pm 1.5\%$ | 64% |
| 3 | $9 \pm 0.7\%$ | 75% |
| 4 | $4 \pm 0.4\%$ | 80% |
| 5 | $4 \pm 0.3\%$ | 84% |
| 6 | $1 \pm 0.2\%$ | 85% |
| 7 | $1 \pm 0.2\%$ | 86% |
| …195 | 0.000003% | 100% |

[a]Referring to eigenvectors.
[b]Eigenvalue divided by the cumulative sum of eigenvalues.

parameter at locations $x$, where no measurement has been made.

[27] An important tool in geostatistics is the experimental variogram $\gamma^*(\vec{h})$, which quantifies the spatial variability of the parameter under study. It is defined as a function of the point separation vector $\vec{h}$ (vector in geographical space), $\gamma^*(h) = \frac{1}{2} Var(z(x_k + \vec{h}) - z(x_k)) = \frac{1}{2n}\Sigma_{x_k \in D} (z(x_k + \vec{h}) - z(x_k))^2$ and measures the dissimilarity (variance) between all the pairs of the regionalized variables at sampling location $x_k \in D$ with respect to their spatial separation $\vec{h}$. Some tolerance concerning the length and the orientation of the vector $\vec{h}$ may be introduced. Although the distribution of projections is anisotropic, the problem can be reduced to an isotropic problem by a linear transformation of the coordinates, so that the variogram depends only on the modulus of $\vec{h}$ [see *Journel and Huijbregts*, 1978]. This leads to a function describing the spatial structure of the parameter under study.

[28] The experimental variogram needs to be approximated by a theoretical function $\gamma(\vec{h})$, which allows to estimate the variogram analytically for any distance $\vec{h}$; this function is called the variogram model. In practice, this modeling step is mainly interactive and this is one of the major advantages of kriging: representing, interrogating the data and analyzing the spatial structure of the stochastic process under study before performing the estimation.

[29] Once the variogram model has been fitted to the experimental variogram, it is possible to perform the estimation by kriging. A linear estimator is used $Z^*(x) = \Sigma\lambda_k Z(x_k)$, where the unknowns are the weights $\lambda_k$. The values of these weights are determined by minimizing the variance of the estimation error $(Z^* - Z)$, which leads to a system of linear equations (containing explicitly the theoretical variogram function) under some assumptions about the random function $Z$. A complete overview of kriging is given by *Wackernagel* [1995].

[30] By construction, kriging is an unbiased estimator, i.e., the estimation error has zero expectation in the probabilistic model. This mean that, when considering the regionalized variable, the average error on point estimations is zero over a large area.

### 3.3. Kriging in KLT Space

[31] KLT is a linear transform of a correlated time series from a region into a set of orthogonal and thus uncorrelated functions. This results in a set of series $\{Y_j(t)\}$ describing the temporal variation of the original data set, and in a set of series $a_j(X)$ describing the spatial variability in the region from which the original series are collected.

[32] These projections (or coordinates) $a_j(X)$ are calculated for each observation and for each eigenvector. Variography and kriging are carried out on these coordinates $a_j(X)$ of the orthogonal basis $\{Y_j(t)\}$ in order to regionalize them. One advantage of using KLT before the estimation process lies in determining the projections $a_j(X)$ for each separated physical process, and so in interpolating a smooth parameter (compared to piezometric head level for example). The major advantage of kriging is to respect the spatial structure (quantified by the variogram) of the parameter when interpolating.

[33] The combination of both mathematical methods has already been explored by *Biau et al.* [1998] as a downscaling tool of North Atlantic sea level pressure in order to reconstruct monthly winter precipitations. The authors emphasize the ability of kriging to reproduce correctly the mean of a phenomenon, and its smoothing effect (by underestimation of the variance). However, in their work they build up spatial eigenvectors (a pattern of spatial behavior), and therefore use kriging as a temporal interpolator in the subspace created by the two first eigenvectors. In our case, temporal processes, driven by external source signals in the aquifer, are taken to be the main signals that should be identified. Then, spatially speaking, we try to highlight for each process an average behavior of the groundwater body and its large-scale variations.

[34] As the projections to be kriged have an indirect physical meaning, it is difficult to justify the required underlying hypothesis. A priori, the intrinsic property is satisfied when dealing with coordinates. A posteriori, variograms of the projections have a spherical behavior, so these projections are likely to be kriged.

[35] Here KLT is used to separate space and time. Once global underlying temporal processes $\{Y_j(t)\}$ have been determined, it is possible to return to local information through the regionalization of the associated projections $a_j(X)$ thanks to kriging. So, time series can be reconstructed at any location where no measurement has been made. In this way, a statistical model of the aquifer is obtained coupling KLT and kriging.

## 4. Results and Discussion

### 4.1. Eigenvalues of the System

[36] The amount of variance explained by each eigenvector depends on the relative value of its eigenvalue with respect to the total sum of eigenvalues. In Table 1 the most energetic signals of the piezometric variations are identified. We decided to take into account the first five eigenvectors (see Figure 2) to model the behavior of the aquifer for three reasons: (1) More than 80% of the variance of the system is explained. (2) The noise criterion shows that eigenvectors explaining less than 1% of the variance are not well separated from noise. (3) The step between the 5th and the 6th eigenvalues ensures the stability of the subspace generated by these five first eigenvectors.

[37] The first three eigenvectors have eigenvalues that are well separated (see plus/minus error in Table 1), so these eigenvectors are correctly defined by the method. The forth and the fifth eigenvectors are not correctly resolved because
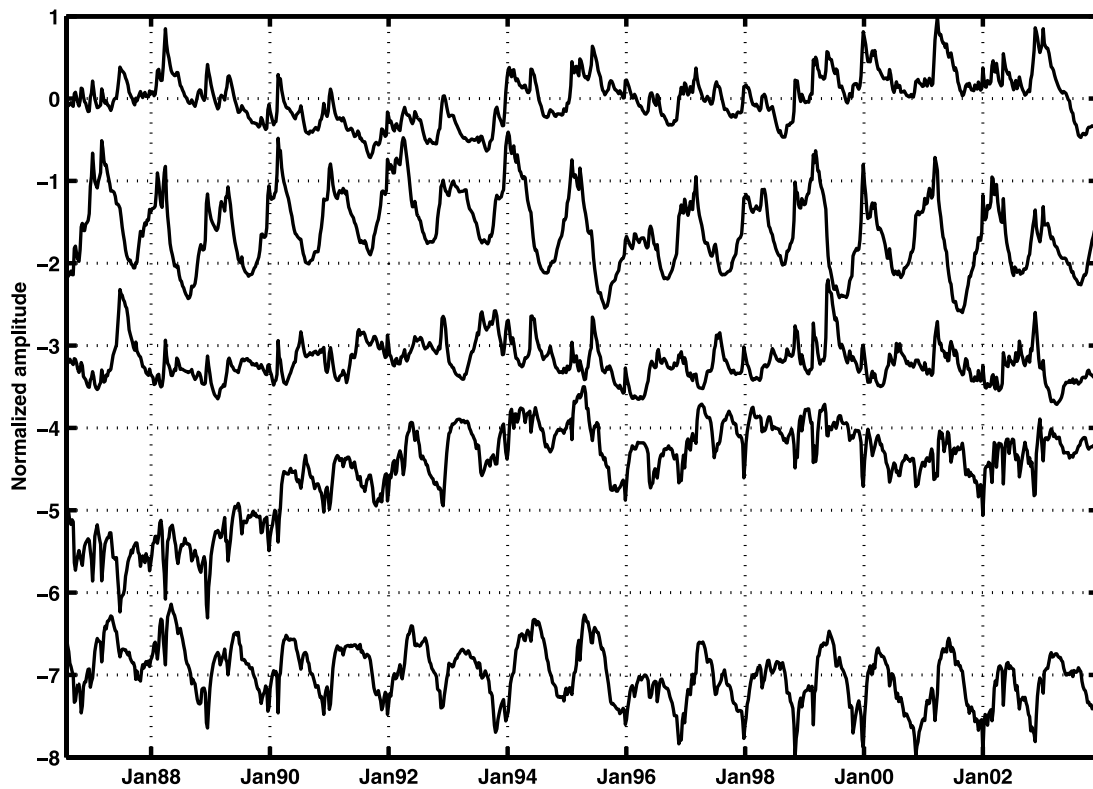
**Figure 2.** Eigenvectors $Y_k(t)$ obtained after KLT, classified in descending order. Note that the means of the eigenvectors are modified for more clarity.

their eigenvalues are very close, so the direct interpretation of these two eigenvectors would be misleading. However, two main characteristics must be highlighted because any linear combination will not remove these characteristics: a long-term evolution for the forth eigenvector and a strong annual term for the fifth (see Figure 2).

[38] So an optimal approximative representation of the original anomalies of the multivariate field $Z$ is obtained by projecting it onto a subspace of dimension five. Each normalized piezometric head variation $\Delta H(X, t)$ can be written as a linear combination of these five temporal eigenvectors $Y_j(t)$ and associated spatial information $a_j(X)$: $\Delta H(X, t) = \Sigma_{j=1}^{5} a_j(X)Y_j(t)$. The spatial information $a_j(X)$ is determined by projecting each normalized piezometric head variation on the temporal basis $\{Y_j\}$, $j = 1..5$.

### 4.2. Physical Interpretation of Eigenvectors

[39] The KLT method is built on the assumption that the set of eigenvectors $\{Y_j(t)\}$ is an orthogonal basis. This property constitutes, however, a strong constraint which puts limits to the physical interpretability of individual eigenvectors since physical signals tend generally to be nonorthogonal [*Simmons et al.*, 1983]. As a consequence, physical interpretability of eigenvectors can be controversial [see, e.g., *Dommenget and Latif*, 2002; *Swadhin et al.*, 2002].

[40] The source signals that are likely to induce a piezometric head variation in the Upper Rhine aquifer as a whole are well known: precipitations, mountain river contribution, Rhine River contribution that have been determined in the reference contributions section. These reference contributions are however correlated (river floods are generated by

rainfalls, almost all environmental signal have a predominant annual term), and a simple assignment of the eigenvector to a physical contribution through time series analysis would be inaccurate. The quantitative spatial information is also used to check whether the temporal assignment corresponds to a spatial reality. This spatial information should confirm that mountain river have a more important contribution at mountain piedmont, etc.

[41] So, the three first eigenvectors $Y_j(t)$ are confronted to these signals of physical meaning for both short-term and long-term variations (respectively periods lower than 10 months and greater than 18 months). The coherence of the eventual assignment is confirmed by the interpretation of the spatial distribution. Discussion is finally conducted to determine whether the eigenvectors have a physical meaning or not.

### 4.3. Projections on Each Eigenvector

#### 4.3.1. First Characteristic Signal: Mountain River Contribution (Except Rhine)

[42] The most energetic signal (44% of explained variance) could be interpreted as the mountain river contribution into the aquifer (Rhine River excepted). The set of floods observed in winter and in spring is modulated by an annual period, which is maximal in winter and minimal in summer. In fact, winter rainfall on the Vosges and the Black Forest are largely collected by the groundwater body through the rivers which infiltrate into the aquifer. The statistical temporal signal is consistent with the physical signals of the rivers, both in the low- and high-frequency bands (see Figures 3 and 4).
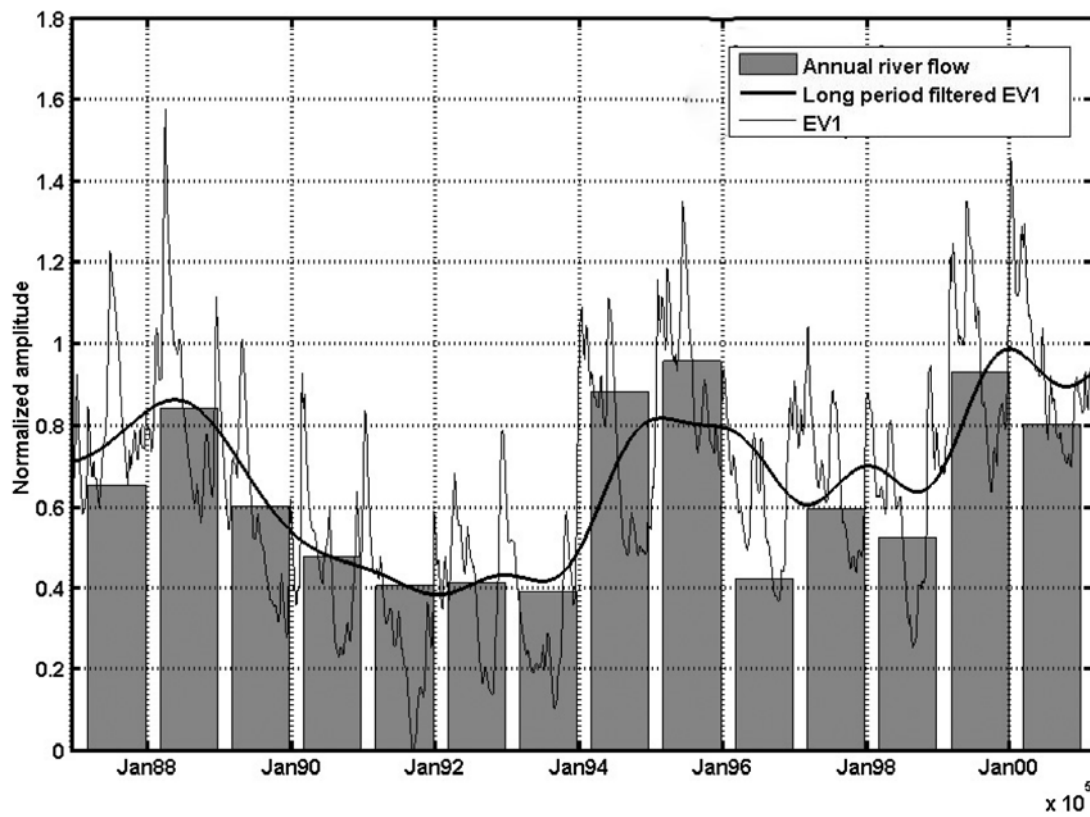
**Figure 3.** Annual river flow at the entrance of the aquifer (except Rhine) toward long-period filtered first eigenvector EV1.
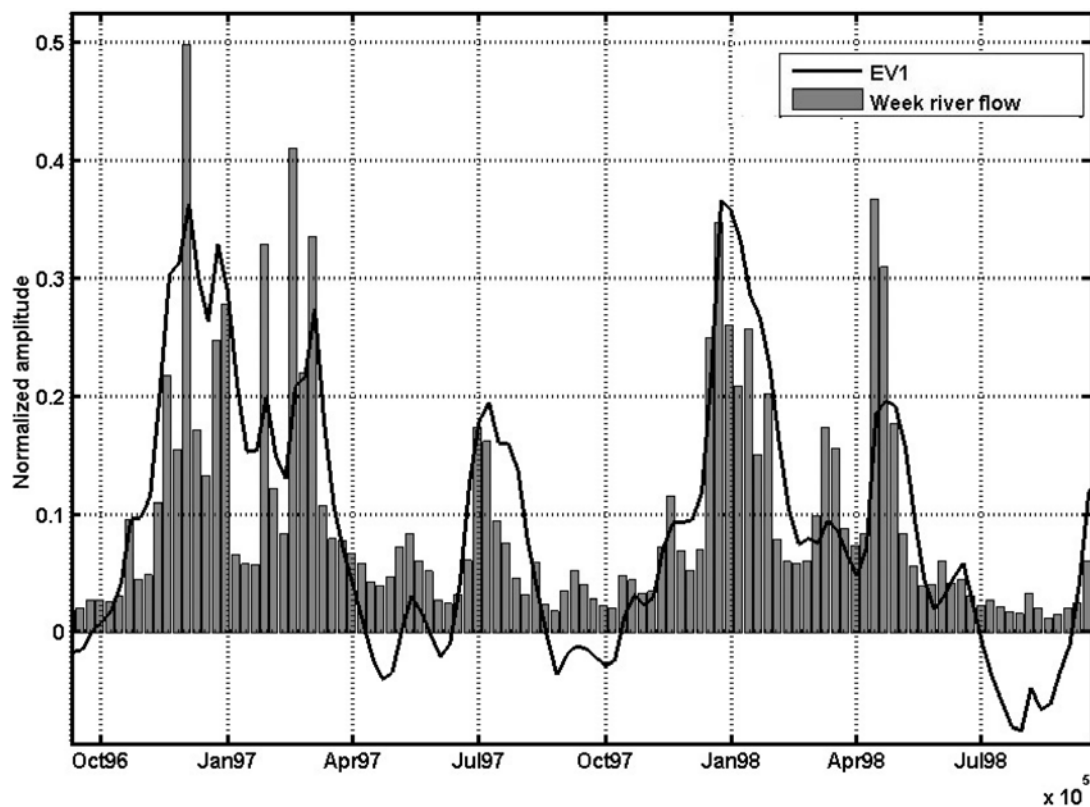


**Figure 4.** Week river flow at the entrance of the aquifer (except Rhine) toward first eigenvector EV1.
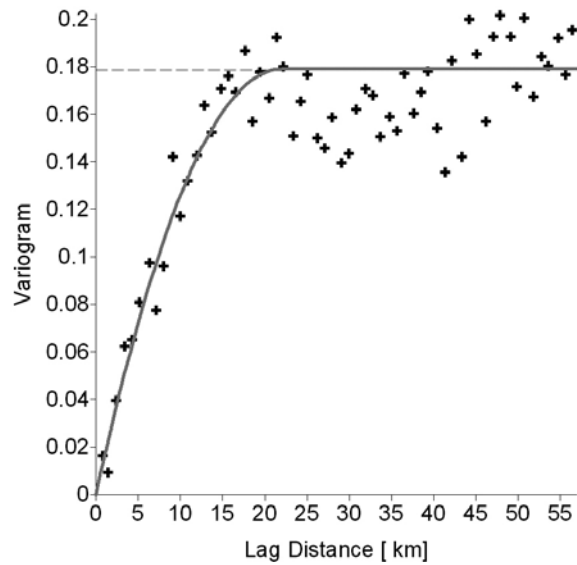
**Figure 5.** Experimental variogram (crosses) with associated spherical model (curve) of the projections $a_1(X)$ on the first eigenvector EV1. Data variance is also plotted (dashed line).

[43] The variogram of the projections (see Figure 5) shows a well-defined structure, with a range of approximately 20 km and a continuity inherited from the diffusive properties of the aquifer. The variogram also highlights anisotropy in the N20E direction, which is consistent with the mean flow direction in the aquifer.

[44] The experimental variogram is modeled by a spherical model (see Figure 5), defined as

$$\gamma_s(h) = \begin{cases} C\left(\frac{3}{2}\frac{|h|}{r} - \frac{1}{2}\frac{|h|^3}{r^3}\right) & \text{if } 0 \le |h| \le r \\ C & \text{if } |h| > r \end{cases}$$

where $r$ is the range, or correlation length of the spatial variation of behavior. $C$ is the scale and converges toward the global variance of the data.

[45] The distribution of the projections (Figure 6) of mountain river contribution observed here is consistent with the geographic configuration of the aquifer. The most sensitive part of the distribution is located in the mountain foothills and in the vicinity of the Ill River. Closer to the Rhine River, the groundwater is less sensitive to mountain river inflow. Note that this eigenvector is the only one that has a spatial average that is nonzero, i.e., the only one which brings a positive water balance to the aquifer.

### 4.3.2. Second Characteristic Signal: Effective Rainfalls Contribution (Immediate Reaction)

[46] The second characteristic signal is much smoother and represents 20% of the global behavior. It can be interpreted as effective rainfall, which is maximal in winter (Figures 7 and 8). Since phase differences are not managed by the method, the infiltration signal is associated with short percolation times (in this case, percolation times are lower than 2 weeks, and the aquifer is highly sensitive to rainfalls [*Cloots-Hirsch*, 1990; *Perez et al.*, 1999]). The signal also reveals the summer water loss, i.e., exfiltration or drainage. Indeed, groundwater is generally very shallow in the Rhine

aquifer north of Colmar, generating the wetlands known as "Ried". As in the case of the first eigenvector, the variogram shows a strong continuity and a 20 km structure as well as a N20E directional anisotropy. It has been modeled by a spherical structure.

[47] In specific areas (Figure 9), such as in the Hardt forest and in the south of Strasbourg, the projections are negative. As a consequence, the characteristic signal is opposite to the piezometric head time series. These negative spatial projections are generated by the fact that two phenomenons are out of phase, they are therefore inserted in the same eigenvector with either positive or negative projection. Indeed, negative values reflects the impact of summer contribution from the Rhine River and from the Hardt canal (into which Rhine water is piped in the summer season for irrigation needs), that is temporally opposite to the evapotranspiration contribution.

### 4.3.3. Third Characteristic Signal: Rhine Contribution

[48] The third eigenvector can be associated with Rhine inflow into the aquifer. The high-frequency part (Figure 10) of the signal matches with the observed flow of the Rhine. The low-frequency component (Figure 11) of the characteristic signal, however, does not correlate well with annual Rhine flow. This discrepancy can be explained by the projection distribution: the signal describing the Rhine flow
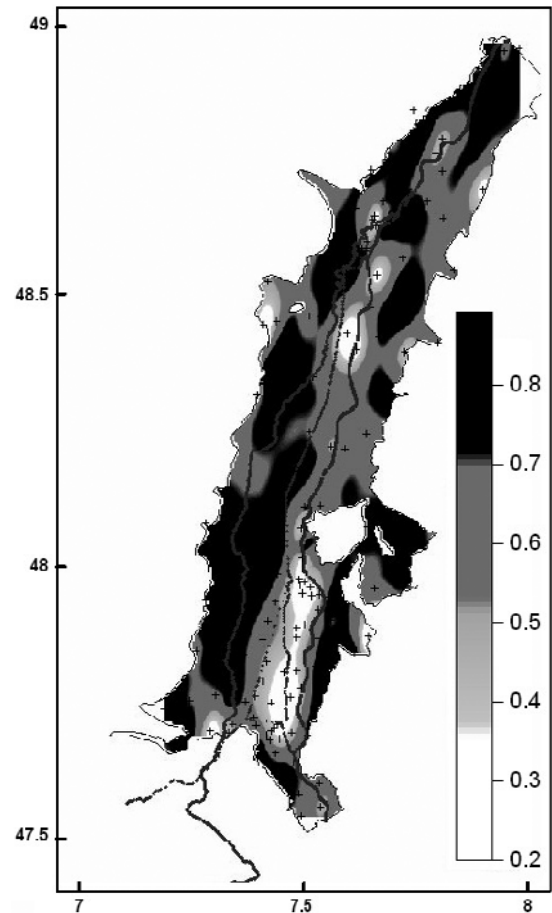


**Figure 6.** Regionalization of the projections $a_1(X)$ on the first eigenvector EV1, describing relative importance of river contribution to aquifer.
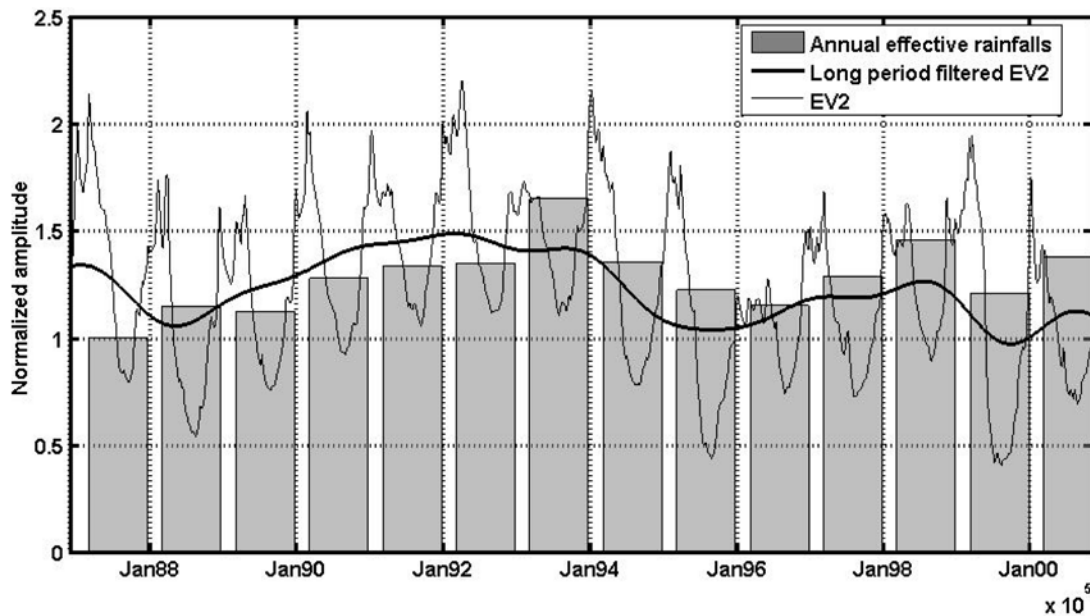
**Figure 7.** Annual effective rainfalls toward long-period filtered second eigenvector EV2.

at high frequencies is modulated by a low-frequency contribution from the Sundgau area (where projections are negatives). Here is an example of the "construction effect" generated by the orthogonality required by the mathematical method [*Simmons et al.*, 1983]. The distribution of the positive projections (Figure 12) concentrates along the course of the Rhine river and reveals the main exchange areas between the Rhine River and the groundwater body, such as in the south of Strasbourg and Strasbourg harbor, where the Rhine and groundwater levels are equilibrated by the docking basins. The projections are almost equal to zero on the Vosges and Black Forest mountains foothills except

in the northeast of Strasbourg where Rhine water is canalized for irrigation.

### 4.3.4.　Fourth Characteristic Signal

[49] The fourth characteristic signal representing 4% of the global variance is the first correction signal. Interpretation is delicate, but from a temporal point of view, a water "stocking-destocking" effect should be highlighted (see Figure 2). This eigenvector is the only vector whose variogram shows a linear behavior with a nugget effect at 2/3 of the variance. It therefore does not reveal any structure. Generally, the projections are equal to zero (Figure 13) but pointing out a few specific areas such as the area between Mulhouse and Colmar where the Ill River infil-
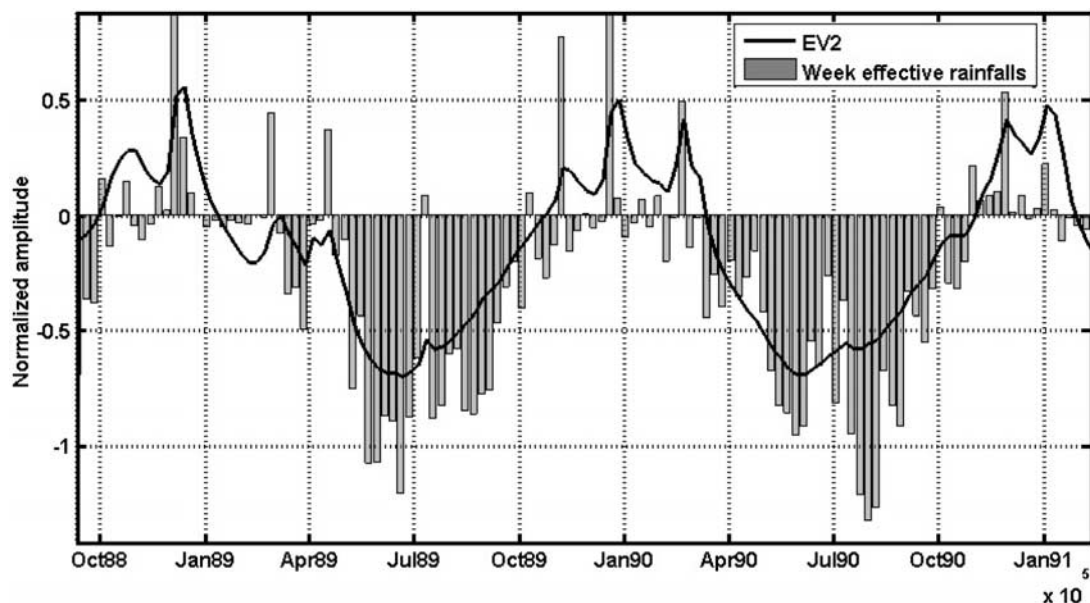


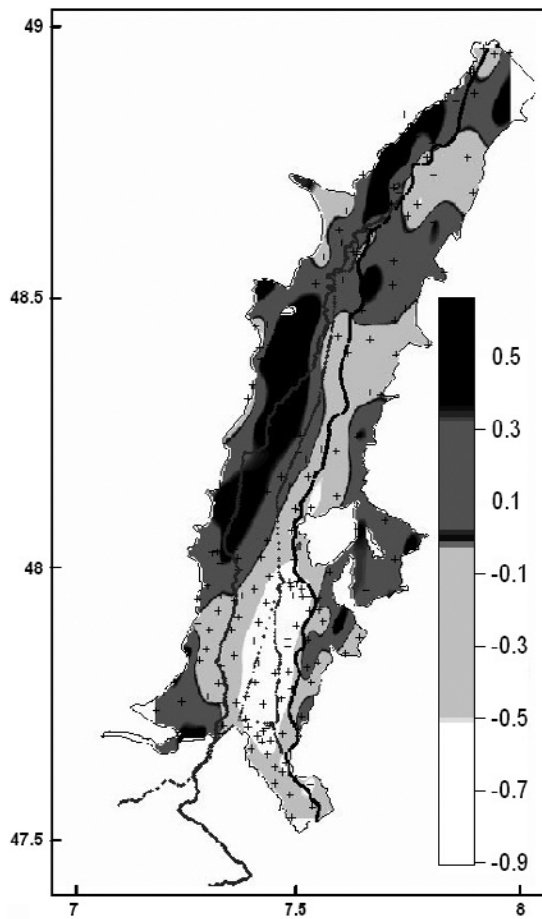**Figure 8.** Week effective rainfalls toward second eigenvector EV2.

**Figure 9.** Regionalization of the projections $a_2(X)$ on the second eigenvector EV2, describing relative importance of effective rainfall contribution to aquifer (positive values). The out-of-phase signal (negative values) describes the summer contribution from Rhine and associated canals.

trates into the aquifer. In 1990, a strong flood led to the decolmation of the river bed of the Ill River, which caused higher infiltration rates and depletion of the river flows in summer that had to be compensated by piping water from the Rhine River through the Huningue Canal in Mulhouse in order to sustain its course [*Jaillard*, 2003].

### 4.3.5. Fifth Characteristic Signal

[50] The fifth and last eigenvector (Figure 14) is the second correction signal, and precautions should be taken for interpretation. From a temporal point of view (see Figure 2), the strong annual variation of this eigenvector is noticeable, with a maximum in winter and sharp minima corresponding to winter rainfalls and associated river floods.

[51] The variogram shows a continuous phenomenon with a structure of around 10 km. Because of the sparse sampling in the north of the aquifer compared to the correlation length, spatial interpretation could be problematic. Generally speaking, the dark colors (negative projections, i.e., amplification of winter rainfalls and river floods in high water situations and amplification of drainage in shallow water) are concentrated in the center of the valley, whereas light colors (positive projections, diminution of effect of winter rainfalls and river floods in high water situations and diminution of drainage in shallow water) are distributed on the borders. When focusing on the center of the aquifer, where sampling is tighter, a few elongated areas even have stronger negative projections. This means that annual variations have a much more significant amplitude than described by the two first characteristic signals.

[52] This eigenvector may be interpreted as the correction of the nonlinearity of the diffusive equation due to the small thickness of the aquifer toward piezometric head variations (light colors are distributed on the eastern and western borders, and especially in the south). In fact, the mathematical method strives to describe at best the energy of the piezometric head variations of the system as a whole. As a consequence, a global "linear" behavior is determined for the aquifer as a whole. A correction of the signal is then
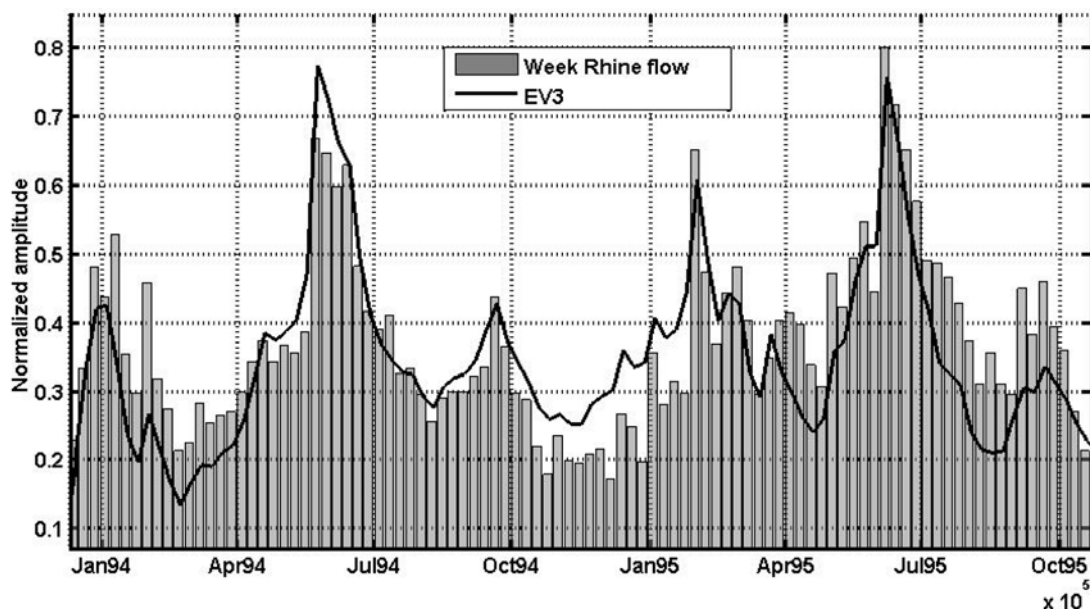


**Figure 10.** Weekly Rhine flow at the entrance of the aquifer toward the third eigenvector EV3.
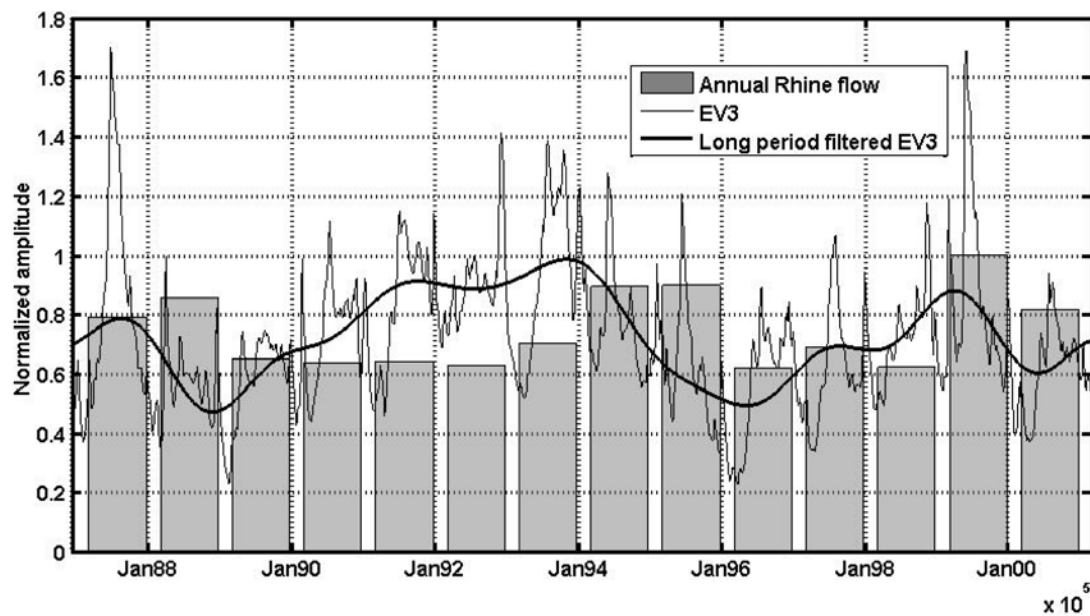
**Figure 11.** Annual Rhine flow at the entrance of the aquifer toward long-period filtered third eigenvector EV3.
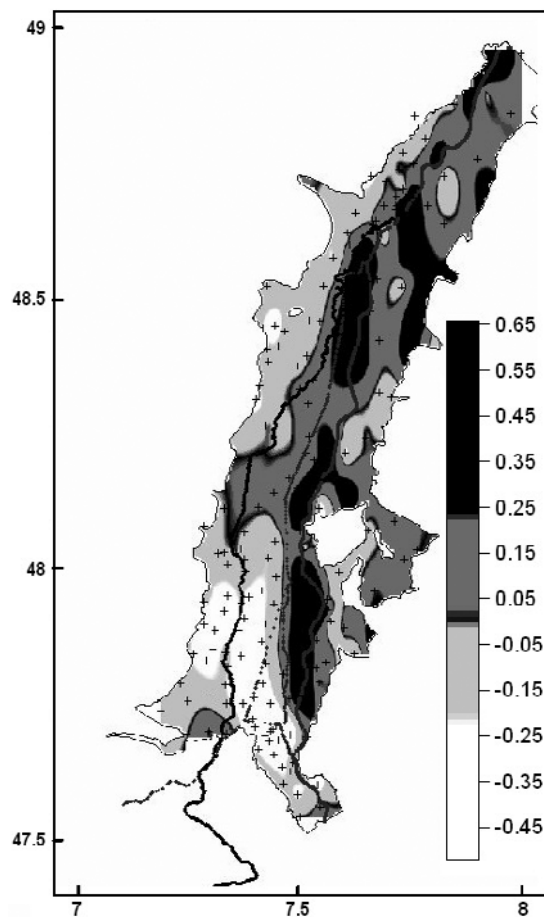


**Figure 12.** Regionalization of the projections $a_3(X)$ on the third eigenvector EV3, describing relative importance of the Rhine contribution to the aquifer (positive values).
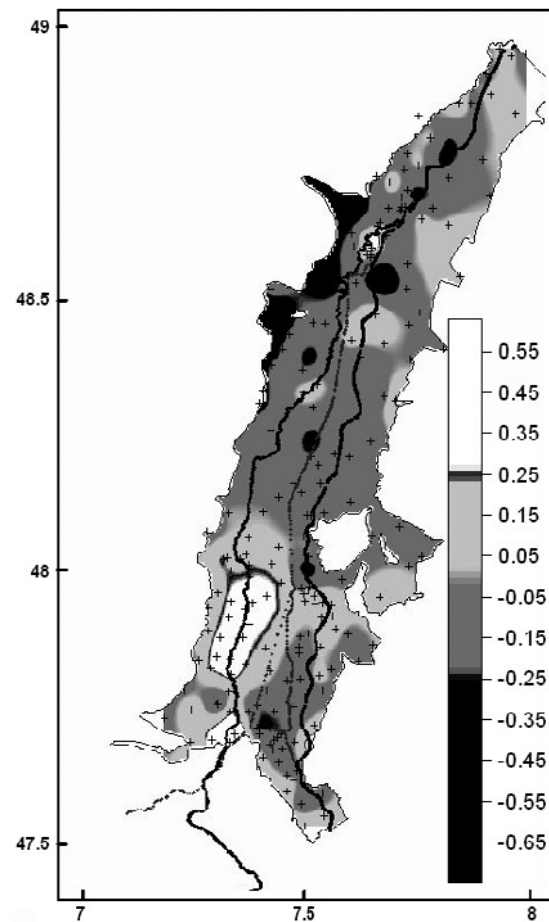


**Figure 13.** Regionalization of the projections $a_4(X)$ on the fourth eigenvector EV4. The long-term variation of EV4 describes the increase of the water table consecutive to the decolmation of the Ill River in 1990 (blank values).
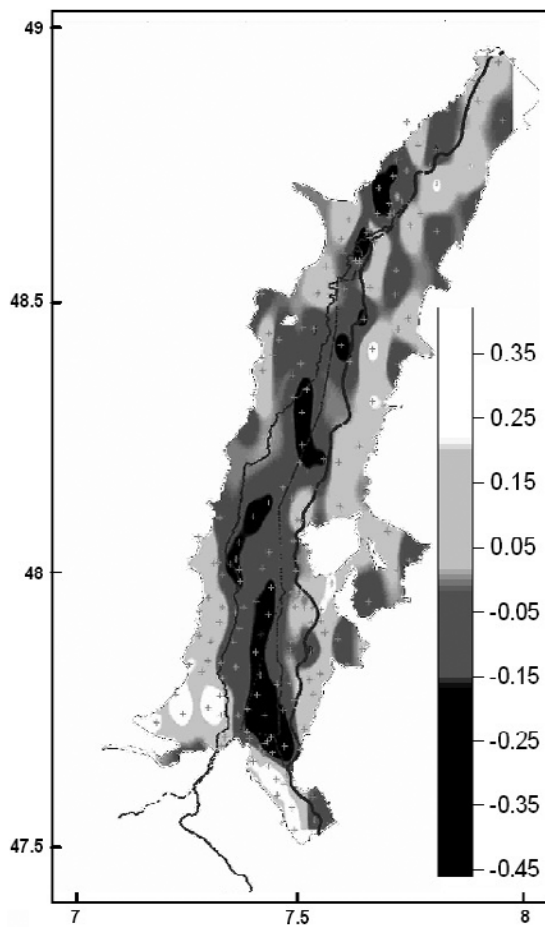
**Figure 14.** Regionalization of the projections $a_5(X)$ on the fifth eigenvector EV5, the annual correction of the first and second eigenvectors. Negative values indicate a stronger influence by rain and river contribution.

necessary to "complete the behavior" more locally in order to adjust the first linear behavior: A nonlinear behavior can be approximated as the sum of two linear behaviors.

[53] In the center of the aquifer, this eigenvector also seems to describe differences in aquifer properties. The shape of these north-south black structures cannot be accurately determined because of the sparse spatial sampling, but it could be interpreted as possible paleochannels. Indeed, branches of the Rhine River have been spreading in the valley, because of the flat topography, drawing a lot of islands and peninsula. From the 17th century, civil works were carried out in order to canalize the Rhine River and drain wetlands [*Descombes*, 1985].

### 4.4. Discussion

#### 4.4.1. Flaws in the Description of the System

[54] The statistical model set up with the regionalization of the first five eigenvectors describes 80% of the variance. The missing 20% are considered as noise by the method, i.e., real noise or processes that are too local to be distinguished from noise. Globally speaking, the main behavior of the aquifer is well described. In Figure 15 the Nash efficiency for the reconstructed piezometric head is shown. Nash efficiency N [*Nash and Sutcliffe*, 1970] is defined as a

function of observed and reconstructed piezometric head variations $H_{observed}$ and $H_{simulated}$, and mean piezometric head variation $E(H_{observed})$ is defined as

$$N = 1 - \frac{\Sigma_t(H_{observed}(t) - H_{simulated}(t))^2}{\Sigma_t(H_{observed}(t) - E(H_{observed}(t)))^2}$$

Sixty percent of the observation wells have a Nash efficiency higher than 0.7, and 80% are higher than 0.5.

[55] The major problems are located in the center of the alluvial plain south of Strasbourg. In this area, a large man-made lake equilibrates the groundwater level with the Rhine level, resulting in very low piezometric head variations. The spatial scarcity of the data should also be noted in this area (four observation wells only). Another problematic area is located in the northeast side of the valley close to the Black Forest mountains, where water coming from the Rhine is used for irrigation.

[56] Temporally speaking, KLT is a model of immediate mix, so propagation effects of the physical processes are neglected, and amplitude errors may be introduced. Better results are obtained when quasi static phenomena are sampled. In our case a weekly sampling is surely optimal
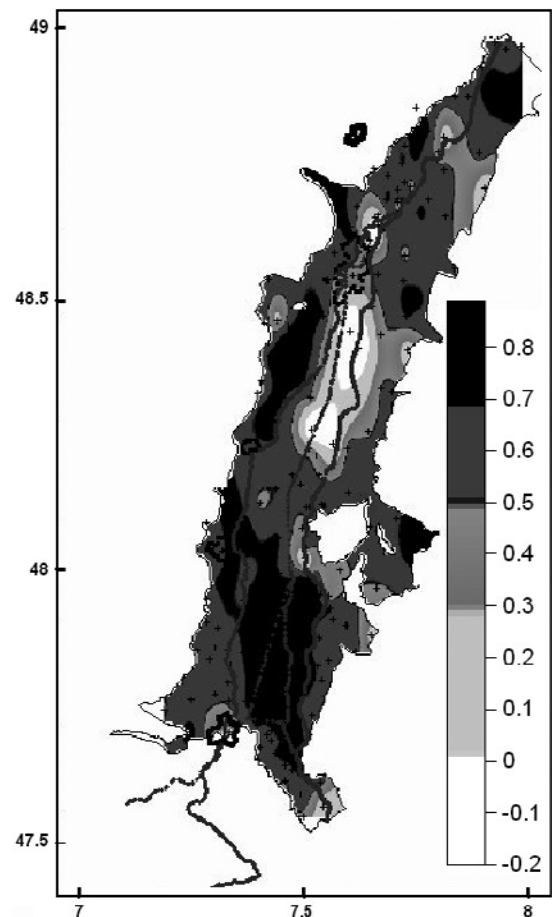


**Figure 15.** Regionalization of Nash-Sutcliffe efficiency of reconstructed piezometric head variations of the first five eigenvectors.
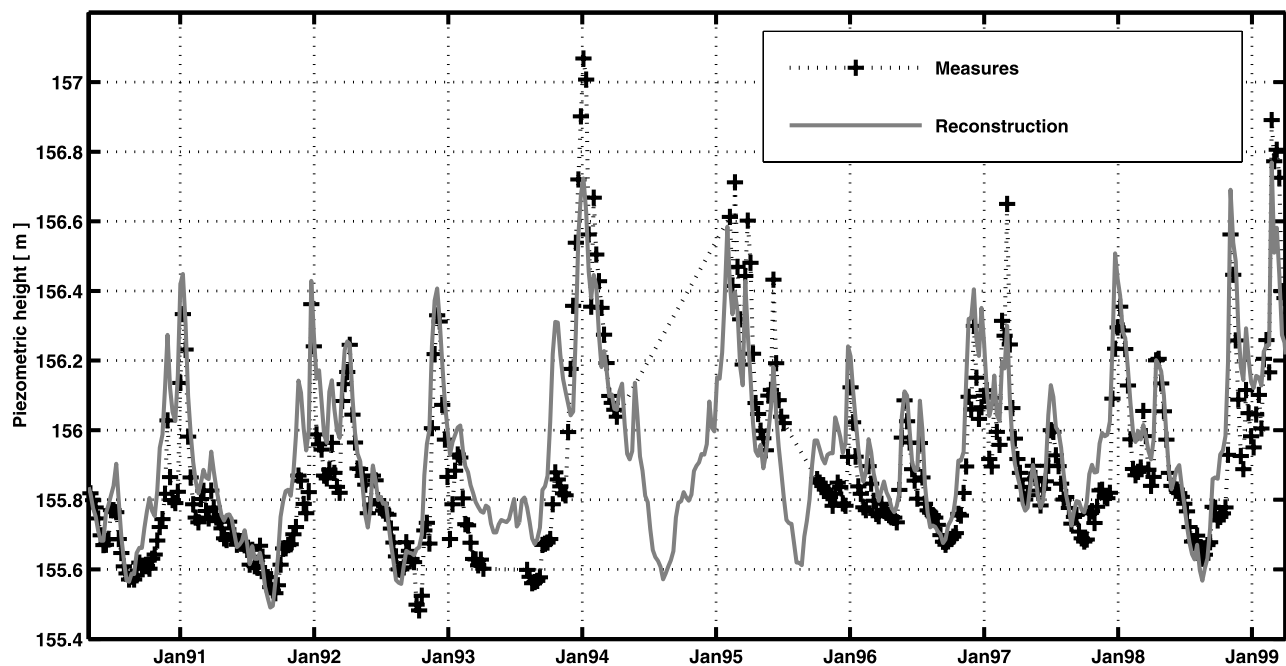
**Figure 16.** Superposition of measured piezometric head not used for analysis and reconstructed time series. Nash criterion is 0.7. Note the capacity of the method to interpolate time gaps.

to separate correctly the source signals into the aquifer and preserve information about river floods.

[57] Spatially speaking, errors may be generated by the smoothing effect of kriging [*Chauvet*, 1994], so difficulties could arise in the case of (1) strong local behavior of an observation well or (2) sparse sampling. In this study, each piezometric time series has been normalized (amplitude divided by temporal variance), meaning that the reconstruction of the time series requires a spatial interpolation of this temporal variance. In a similar study on separate hydrological basins [*Hisdal and Tveito*, 1992] kriging of temporal standard deviation gave unsatisfactory results due to its large spatial variability. Results are more reliable in our case because kriging is applied to temporal variance of the time series (which is an additive field) on a single hydrological unit.

[58] On the whole, when confronting unused data to reconstructed time series at the same point (see Figure 16 for an example), the error is equivalent to the one calculated for the whole aquifer.

### 4.4.2. Robustness of the Method

[59] A second calculation was carried out on only the French part of the Rhine aquifer (149 out of 195 observation wells, i.e., 75% of data), and a third one was carried out on only the German part (25% of observation wells). The first three major eigenvectors are almost identical to those derived from the study of the whole aquifer (see Table 2 for correlation coefficients).

[60] Phases of floods and annual variations correctly described, being only flood amplitudes slightly different between the various calculations. As a consequence, the spatial distribution of the projection is also identical too. This is important because the method succeeds in extracting the major eigenvectors thanks to a limited number of measurements. Adding data helps in separating signals from noise. As the two correction signals reflect more local

behavior, it is natural to have a lower agreement between these eigenvectors.

## 5. Conclusion

[61] The temporal variation of 195 piezometric heads from observation wells has been adequately described by three eigenvectors dealing with mean contribution from the river and hydrometeorological forcings into the aquifer, and two correction signals associated with the nonlinearity of the diffusion equation, environmental shaping and differences in the aquifer properties. These five characteristic signals have been derived for the entire aquifer. The spatial information was also conveniently downscaled using kriging to describe local behaviors. A statistical model of the aquifer has thus been established thanks to the combination of KLT and kriging. Only a few set of measurements is needed to extract the source signals. It is also possible to extract either a spatial or a temporal contribution from each time series, i.e., to (1) discriminate between the various contributions (rainfall, exchanges with rivers) in a piezometric head time series, (2) draw the spatial pattern of the amount of each contribution in piezometric head variation,

**Table 2.** Correlation Coefficient Between the Eigenvectors of the First Study With 195 Observation Wells and the Experiments With 75% of the Observation Wells and Experiments With 25% of the Observation Wells

| Eigenvector | 75% Observation Wells | 25% Observation Wells |
|---|---|---|
| 1 | 99% | 97% |
| 2 | 99% | 95% |
| 3 | 99% | 90% |
| 4 | 95% | 85% |
| 5 | 95% | 65% |

(3) reconstruct piezometric heads anywhere within the aquifer in the time span considered in the study, and (4) draw smoothed piezometric surfaces representing only sound global behavior.

[62] Finally, this method can be seen as a starting point for other problems since major hydrological processes have been quantified via the separation of source signals. Another interesting property of this mathematical decomposition is the separation of space and time that helps in other mathematical mixings, such as convolution in order to calculate the geodetic effects of mass variation in the aquifer.

## References

Biau, G., E. Zorita, H. Von Storch, and H. Wackernagel (1998), Estimation of precipitation by kriging in the EOF space of the level pressure field, *J. Clim.*, *12*, 1070–1084.

Broomhead, D. S., and G. P. King (1986), Extracting qualitative dynamics from experimental data, *Physica D*, *20*, 217–236.

Chauvet, P. (1994), Aide mémoire de géostatistique linéaire,, *Cah. Geostat.*, *2*, 210 pp.

Cloots-Hirsch, A. R. (1990), Une carte de l'extension et des mécanismes des inondations dans le ried central de l'Ill ralisée dans le cadre du PIREN-eau/Alsace, *Rev. Geogr. Est*, *30*(III), 285–288.

Descombes, R. (1985), *Encyclopédie de l'Alsace*, Ed. Publitotal, Strasbourg, France.

Dommenget, D., and M. Latif (2002), A cautionary note on the interpretation of EOFs, *J. Clim.*, *15*, 216–225.

Düringer, P. (1988), Les conglomérats des bordures du rift cénozoïque rhénan. Dynamique sédimentaire et contrôle climatique, Ph.D. thesis, 278 pp., Univ. of Strasbourg, Strasbourg, France.

Elsner, J. B., and A. A. Tsonis (1996), *Singular Spectrum Analysis: A New Tool in Time Series Analysis*, Elsevier, New York.

Gerbrands, J. J. (1981), On the relationship between SVD, KLT and PCA, *Pattern Recognit.*, *14*(1), 375–381.

Gottschalk, L. (1985), Hydrological regionalization of Sweden, *Hydrol. Sci.*, *30*(1), 65–83.

Grimmer, M. (1963), The space-filtering of monthly surface anomaly data in terms of pattern using empirical orthogonal functions, *Q. J. R. Meteorol. Soc.*, *89*, 395–408.

Hisdal, H., and O. E. Tveito (1992), Generation of runoff series using empirical orthogonal functions, *Stochastic Hydrol. Hydraul.*, *6*, 255–269.

Jaillard, L. (2003), BRGM: Elements d'hydrologie et d'hydrogologie entre Mulhouse et Colmar en vue d'une modlisation de la nappe d'Alsace, report, 16 pp., Luc Jaillard Ing.-Conseil, Mulhouse, France.

Jolliffe, I. T. (2002), *Principal Component Analysis*, Springer, New York.

Journel, A. G., and C. J. Huijbregts (1978), *Mining Geostatistics*, 600 pp., Elsevier, New York.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models, 1, A discussion of principles, *J. Hydrol.*, *10*, 282–290.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng (1982), Sampling errors in the estimation of empirical orthogonal functions, *Mon. Weather Rev.*, *110*, 699–706.

Pereira, F. (2004), Analyse spatio-temporelle du champ géomagnétique et des processus d'accélération solaires observés en émission radio, Ph.D. thesis, 178 pp., Univ. of Orléans, Orléans, France.

Perez, J. S., M. Tremolires, N. Takatret, P. Ackerer, A. Eichborn, and G. Maire (1999), Quantification of nitrate removal by a flooded alluvial zone in the Ill floodplain (eastern France), *Hydrobiologia*, *410*, 185–193.

Saporta, G. (1990), *Probabilités, Analyse des Données et Statistique*, 493 pp., Technip, Paris.

Shun, T., and C. J. Duffy (1999), Low-frequency oscillations in precipitation, temperature and runoff on a west facing mountain front: A hydrogeologic interpretation, *Water Resour. Res.*, *35*(1), 191–201.

Simmons, A. J., G. W. Branstator, and J. M. Wallace (1983), Barotropic wave propagation, instability and atmospheric teleconnection patterns, *J. Atmos. Sci.*, *40*, 1363–1392.

Sittler, C. (1969), Le fossé rhénan en Alsace, aspect structural et histoire géologique, *Rev. Geogr. Phys. Geol. Dyn.*, *11*, 465–494.

Swadhin, K. B., A. R. Suryachandra, N. S. Hameed, and Y. Tshio (2002), Comments on "A cautionary note on the interpretation of EOFs," *J. Clim.*, *16*, 1087–1093.

Villemin, T., F. Alavarez, and J. Angelier (1986), The Rhinegraben: Extension, subsidence and shoulder uplift, *Tectonophysics*, *128*, 47–59.

von Storch, H., and F. W. Zwiers (1999), *Statistical Analysis in Climate Research*, Cambridge Univ. Press, New York.

Wackernagel, H. (1995), *Multivariate Geostatistics*, 256 pp., Springer, New York.

Wilks, D. S. (1995), *Statistical Methods in the Atmospheric Sciences*, Elsevier, New York.

————————————

P. Elsass, Service géologique régional Alsace, BRGM, 15 Rue du Tanin, BP 6009, Lingolsheim, F-67834 Tanneries, France.

N. Florsch and L. Longuevergne, UMR Sisyphe, Université Pierre et Marie Curie, 4 place Jussieu, F-75252 Paris, France. (laurent.longuevergne@ccr.jussieu.fr)