Including Conceptual Model Information when Kriging Hydraulic Heads

M. Rivest, D. Marcotte and P. Pasquier

Abstract A reliable hydraulic head field is a key element to many hydrogeological, environmental or geotechnical studies. It enables quick identification of areas where high hydraulic gradients could threaten an earth dam's integrity. It also highlights probable contaminant flow paths and determines wells' influence areas. Furthermore, some inversion algorithms (direct methods) require an initial estimation of the entire head field to compute hydraulic conductivity. Interpolation techniques, such as kriging, have the advantage of reproducing the observed values. However, the shape of the interpolated head fields often lacks realism particularly near pumping wells, boundaries and lithological contrasts. In these cases, the flow equation is poorly reproduced by interpolation. On the other hand, numerical modeling can easily integrate the hydrogeological conceptual model and generate realistic head fields. Unfortunately, the numerical model is based on uncertain hard data which poorly reproduce the head observations.

We propose an approach based on kriging that uses the "shape" information present in a numerical conceptual model as an external drift. The performance of the method is first investigated using a 2D synthetic aquifer. In this case, several numerical head fields are used in the external drift to account separately for different aspects of the phenomenon (principle of superposition). A stepwise procedure is used to select the best set of numerical head fields. Kriging with external drift (KED) shows marked improvement over ordinary kriging and universal kriging with first order polynomials. The approach is also applied to the study of two large earth dams in which monitoring data is available. Cross-validation shows again the good performance of KED compared to ordinary kriging or universal kriging with first order polynomials. The approach can be used for 3D head field estimation.

M. Rivest
École Polytechnique de Montréal, Département des génies civil, géologique et des mines, C.P. 6079, Succ. Centre-ville, Montréal, Qc, Canada, H3C 3A7
e-mail: martine.rivest@polymtl.ca

A. Soares et al. (eds.), *geoENV VI – Geostatistics for Environmental Applications*, 141–151

1 Introduction

Many hydrogeological, environmental, and geotechnical studies require a reliable estimate of the hydraulic head field for design or monitoring purposes. A head map can be used directly to determine the zone of influence of a well or high gradients in an earth dam. It is also required as an initial solution for transient state problems, a first guess for non-linear numerical problems or a starting point in some inversion algorithms (Pasquier and Marcotte, 2005, 2006; Sagar et al., 1975) Head fields are typically obtained either by direct interpolation (e.g. kriging) or by numerical modelling. The latter requires knowledge of the geometry, the boundary conditions, and the hydrogeological parameters of the field under study. Heads obtained by a flow simulator respect the diffusion equation; however, uncertainties regarding the model's characteristics lead to important differences between computed and observed heads at measurement points. Model calibration helps to improve the head fit but can be time-consuming if done manually. Furthermore, since calibration is an ill-defined problem, many solutions are possible.

Direct head estimation, such as kriging, ensures exact interpolation at a data point, thus including measurement error. When measurement error is present, it is customary to filter out this component by avoiding estimating data points. Contrarily to numerical modelling, consistency of the resulting head field with the phenomenon is not guaranteed. In fact, kriging often fails to properly reproduce important features such as no-flow and constant head boundaries, as well as the influence of extracting wells (drawdowns and shape). Spurious head minimum or maximum could also be present on the kriged head field. Several techniques are helpful in improving the realism of kriged head fields. No-flow boundaries can be taken into account by the introduction of double points in the kriging system (Brochu and Marcotte, 2003; Delhomme, 1979). Analytical equations describing radial flow in a perfectly homogeneous and infinite medium define drift functions to further constrain the estimation and account for the presence of wells (Brochu and Marcotte, 2003; Tonkin and Larson, 2002). However, this approach remains limited to two-dimensional cases where analytical solutions are available.

In this article, we propose to combine the advantage of external drift kriging (KED) (Dehomme, 1979) to the capabilities of numerical flow simulation. The idea is to define the drift with numerical head fields, instead of using analytical equations. The performance of this approach is tested with two case studies. The first case consists of a synthetic two-dimensional aquifer composed of two contrasting homogeneous zones under the influence of pumping wells. In the second case, the method is applied to the mapping of hydraulic heads in two monitored earth dams. For each case, cross-validation is used to assess the precision of the method and to compare it with ordinary kriging (OK) and universal kriging with linear drift (UK). The realism of the kriged head fields is discussed. It is shown that KED with numerical drift significantly improves the head field estimation compared to OK and UK.

We emphasize that the proposed approach is not an inverse method. Its purpose is to obtain a realistic head map. This map, if desired, can then be used with inverse methods like the one presented in Pasquier and Marcotte (2006) to estimate the hydraulic conductivity field.

2 Methodology

In practice, site investigation provides the hydrogeologist with useful qualitative and quantitative information about the hydrogeological system. This information is synthesized into a conceptual model that is a simplified representation of the real field. Nevertheless, the conceptual model, when integrated into a flow simulator, yields valuable insight about the *shape* of the head field. This suggests the use of such a field as secondary variable in KED. The following sections present the theory pertaining to the method.

2.1 Kriging with an External Drift

Kriging with an external drift allows the use of secondary information to account for the spatial variation of the primary variable's local mean (non-stationarity of the mean). The secondary variable is chosen for its strong correlation with the variable of interest. Observed values of this variable should be available at every data point and every estimation point (Ahmed and de Marsily, 1987; Galli and Meunier, 1987). The secondary variable should vary more smoothly than the primary variable since it aims at representing the latter's expectation (Castelier, 1993). In KED, the secondary variable appears as additional unbiasedness constraints. For example, when two secondary variables are used, the unbiasedness constraints are:

$$\sum_{i=1}^{n} \lambda_i = 1 \quad ; \quad \sum_{i=1}^{n} \lambda_i f(x_i) = f(x_0) \quad \text{and} \quad \sum_{i=1}^{n} \lambda_i g(x_i) = g(x_0) \tag{2.1}$$

where λ_i are kriging weights, $f(x_i)$ and $g(x_i)$ are values of the secondary variables at data points, and $f(x_0)$ and $g(x_0)$ are values of the secondary variables at the estimation point. Note that the first equation is the usual ordinary kriging unbiasedness constraint. In matrix notation, the kriging system becomes:

$$\begin{bmatrix} \mathbf{K} & \mathbf{F} \\ \mathbf{F}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} \mathbf{k} \\ \mathbf{f} \end{bmatrix} \tag{2.2}$$

where $\mathbf{K}$ is the $n \times n$ covariance matrix (observation-observation), $\mathbf{F}$ is the $n \times p$ drift matrix, λ is the $n \times 1$ kriging weights vector, μ is the $p \times 1$ vector of Lagrange multipliers corresponding to the unbiasedness conditions, $\mathbf{k}$ is the $n \times 1$ covariance vector (observation-estimation), and $\mathbf{f}$ is the $p \times 1$ drift functions evaluated at the estimation point. It is worth expressing KED in its dual form:

$$h^*(x_0) = \begin{bmatrix} \mathbf{b}' & \mathbf{c}' \end{bmatrix} \begin{bmatrix} \mathbf{k}' \\ \mathbf{f}' \end{bmatrix} \text{ where } \begin{bmatrix} \mathbf{b}' \\ \mathbf{c}' \end{bmatrix} = \begin{bmatrix} \mathbf{K}' & \mathbf{F}' \\ \mathbf{F}' & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{h} \\ \mathbf{0} \end{bmatrix} \tag{2.3}$$

where **b** and **c** are the dual weights, **h** is the vector of observed heads, and $h^*(x_0)$ is the estimated head at x_0. The estimation of the hydraulic head h^* combines both a stochastic contribution ($\mathbf{b}'\mathbf{k}$) and a deterministic drift contribution ($\mathbf{c}'\mathbf{f}$). Note that the drift functions are individually weighted by vector **c**. This shows the advantage of splitting the conceptual model into simpler individual components to gain greater flexibility for the drift modelling.

2.2 Selecting Secondary Variable

As shown in Eq. 2.3, KED optimizes the weighting of each drift component.. For example, consider the aquifer presented in the synthetic case study (section 1, Fig. 1a). It is important to know the position of the contact between units and the transmissivity contrast in order to obtain the right shape of the drift with a single conceptual model. With two conceptual models, one corresponding to a homogeneous aquifer and the other to the two-unit aquifer, only the position of the contact is required. KED will adjust the weights to implicitly retrieve the transmissivity contrast compatible with the observed data.

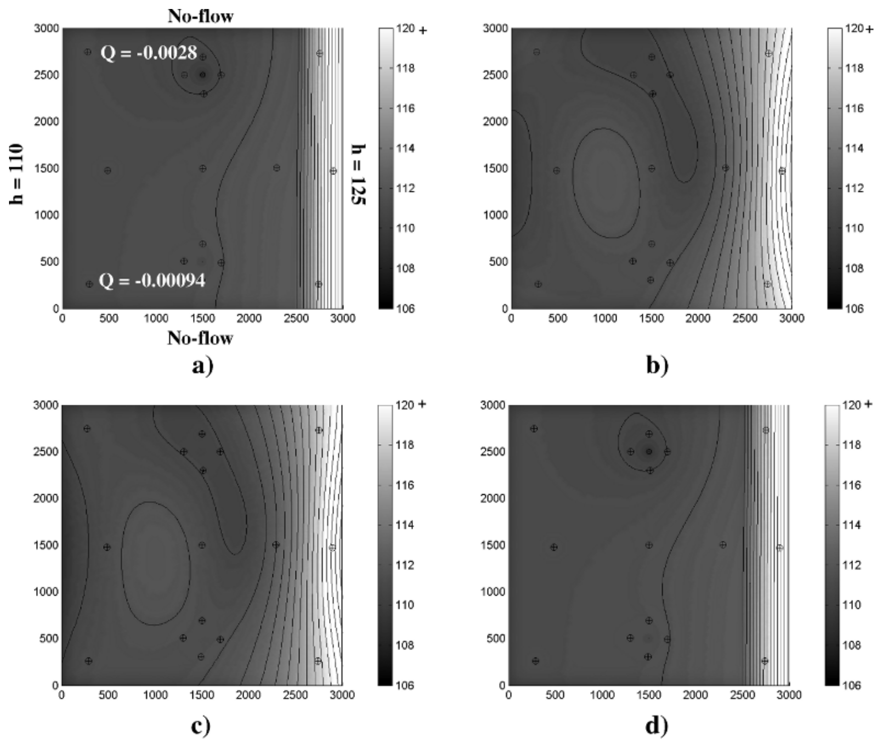

Fig. 1 Synthetic case study a) Reference head field, b) Head field obtained by OK, c) Head field obtained by UK, d) Head field obtained by KDE

It is important to limit the number of drift variables to avoid over-parameterization. Moreover, the p vectors of secondary constraints at data points must be linearly independent to ensure non-singularity of the kriging matrix (left hand side of Eq. 2.2). A stepwise selection procedure (Draper and Smith, 1981) of drift variables is used for this purpose. The approach proceeds sequentially by forward inclusion of drift variables. The candidate variable maximizes the coefficient of determination (R^2). It is retained if the variable yields a significant increase in R^2 (partial-F test). Then, the test determines whether or not the chosen drift variables can be removed (decrease in R^2 not significant). The procedure stops when no more variables can be included in or excluded from the selected set. The stepwise procedure ensures that the retained drift functions are linearly independent. Note that the partial F-test assumes statistical independence of the residuals. The test is only approximate since the residual head data is expected to be spatially correlated. The constant drift function is always included in the kriging.

2.3 Covariance Function

Hydraulic head is a non-stationary variable due to the physics of groundwater flow. The determination of its covariance function is difficult. The head covariance function must be differentiable at least four times because of the double differentiability of head implied by the diffusion equation. The Cauchy gravimetric covariance function has this property (Brochu and Marcotte, 2003). The range parameter and ratios C_0/C (ratio between nugget effect and sill) are chosen by cross-validation (Marcotte, 1995) using KED with the same drift functions used for the final head estimation.

Starting values for C_0/C cross-validation were chosen such as to provide reasonable values for the measurement error variance and head residual variance. Cross-validation parameters were checked by visual assessment of head maps produced by KED. Results were proven very robust to the choice of the model parameters.

For simplicity, we assume the covariance is isotropic. Alternatively, an anisotropic head covariance model could be used. However, our trials with anisotropic models have shown almost no difference with the isotropic model. This shows the strong conditioning imposed by the choice of the characteristics of the conceptual model (first order effect) compared to the fine tuning of the head residual covariance model (second-order effect).

2.4 Performance Criteria

For both synthetic and real case studies, cross-validation is used to quantify the precision of different types of kriging (OK, UK and KED). The statistic retained for comparison is the cross-validation mean absolute error (mae_{cv}), computed from the differences between the observations and the estimates. For the synthetic case, the entire reference head field of the aquifer is known. Therefore, direct comparison with the kriged field allows for the definition of a second statistic: the mean absolute error (mae_d), which is the difference between reference and kriged fields over the whole domain. In this case, hydraulic heads are interpolated at the center of each element and the size of the elements is used to compute the weighted average of the absolute differences between the reference and kriged values. Thus,

$$mae_d(h^K, h^R) = \sum_{j=1}^{n_{ele}} \left| h_j^K - h_j^R \right| \frac{A_j}{A_\Omega} \tag{2.4}$$

where h^K and h^R correspond respectively, to the kriged value and the reference value at the centre of jth element, A_j is the area of jth element, and A_Ω is the area of the entire field.

3 Results

3.1 Synthetic Aquifer

The proposed approach is tested with a synthetic aquifer obtained using the finite element solver Femlab 3.1 (Comsol, 2004). The study area is 3000 × 3000 [L^2]. A contact between two hydrogeological units is found at x = 2500 [L]. Transmissivity in the left-hand unit is 1×10^{-3} [L^2/T] and 4.5×10^{-5} [L^2/T] in the other unit. The right and left boundaries have constant heads (125 m and 110 m, respectively) while the upper and lower boundaries are impervious (no-flow). Two wells are present in the permeable unit yielding local radial flow ($Q_{upper} = -2.8 \times 10^{-3}$ [L^3/T], $Q_{lower} = -9.4 \times 10^{-4}$ [L^3/T]). The numerical solution for this aquifer yields the reference or "real" head field, which is then sampled at 16 locations (Fig. 1a).

When kriging, we use a ratio of C_0/C=0.001 and a range of 2090 m. We assume the following: the existence of the two pumping wells, the presence of two different hydrogeological units (the left one being more permeable than the right one), the location of the contact between units, and the presence of impervious boundaries along the upper and lower edges of the field. This information is split into four simple models:

1. A uniform aquifer with a homogeneous transmissivity (5×10^{-4} [L^2/T]) on the entire field;
2. A two-unit aquifer in which the position of the contact is known (offset = 0);
3. Same as 2) with an outflow in the upper well only;
4. Same as 2), with an outflow in the lower well only.

In all four models, heads imposed on the right and left boundaries are different from those of the reference model and produce stronger gradients. Similarly, the transmissivities of the two units are intentionally chosen to produce a contrast 100 times larger than that of the reference model. The above decomposition is far from unique. We have checked that other decompositions of the conceptual model provided equivalent results for this test case. The four head fields obtained are retained by the stepwise procedure (Partial F-test at $\alpha = 5\%$). Thus, KED is performed with

Table 1 Comparison of OK, UK and KED. The contact's location is specified in the conceptual model at the exact location minus offset

	OK	UK linear drift	KED offset 0	KED offset 25	KED offset 100
mae_{cv} (n = 16)	1.3	1.9	1.1×10^{-3}	7.1×10^{-2}	2.5×10^{-1}
mae_d	7.1×10^{-1}	6.9×10^{-1}	2×10^{-3}	6.3×10^{-2}	2.3×10^{-1}

the four head fields used as secondary variables. The resulting field is presented in Fig. 1d. The corresponding mae_{cv} and mae_d are respectively 1.1×10^{-3} and 2×10^{-3}. Such results are within numerical precision. Moreover, the estimates obtained by KED show major improvements over the estimates obtained by OK or UK (first 3 columns of Table 1, Fig. 1b and 1c). UK and OK both strongly underestimate the gradient in the right unit and create a large artificial maximum at coordinates (1000, 1250). In addition, only KED presents a head field compatible with the impervious boundaries (upper-lower) and the constant head boundaries (right-left).

The contact's position was shifted 25 units left on the external drift models (Table 1) to show the effect of uncertainty on this information. This was repeated for a 100 m shift. Even with an offset, the estimates obtained by KED are much more precise than those obtained by OK and UK.

Note that mae_{cv} and mae_d behave alike (Table 1). This suggests that the differences observed at data points by cross-validation (mae_{cv}) are a good estimate of the differences occurring over the whole field (mae_d).

3.2 Real Earth Dams

In this section, the method is applied on two real earth dams (location undisclosed for confidentiality reason). The dams were built using local borrow material: glacial till for the impervious core and various sands for the filters. Dam instrumentation consists of vibrating wire piezometers and of few observation wells. The Precision of measurements is reported to be of the order of magnitude of a meter.

3.2.1 Boundary Conditions

The boundary conditions used in the numerical model are based on a few assumptions: 1) the head loss between the reservoir and the upstream part of the core can be neglected; 2) the foundation is impervious due to grouting; 3) the hydraulic conductivity (K) ratio between the central core and the downstream filter is large enough to produce a seepage face at the interface. As a result, only the core needs to be considered in the model. The position of the seepage point is found numerically using an iterative method described by Chapuis and Aubertin (2001). No-flow conditions prevail above the seepage point, while the head below the seepage point is equal to the free surface elevation.

3.2.2 Kriging

Due to the simplicity of the problem, we use a single external drift model having homogeneous K and the boundary conditions described in 3.2.1.

Dummy values equal to the water level in the reservoir are included on the upstream side of the core and used in all krigings to enforce the constant head condition. These points were not considered in the computation of mae_{cv}. A value of 0.1 was chosen for C_0/C after a few cross-validation trials. This higher ratio compared to the one used with synthetic data reflects measurement error present in real data. Range parameters for Dam 1 and 2 are 23 m and 15 m, respectively.

Head fields for both dams are depicted in Fig. 2 and 3. All mae_{cv} values are compiled in Table 2. Note that kriging for Dam 1 was performed using 15 observations, whereas 12 observations were used for Dam 2.

The mae_{cv} shows strong improvement of KED and UK over OK for both dams. OK solutions (Fig. 2b and 3b) appear totally unrealistic. In both cases, the flow is converging toward a point in the lower part of the core. Head contours suggest flow through the foundation and also through the top of the core. Moreover, the head gradient at the toe is less with OK than with KED.

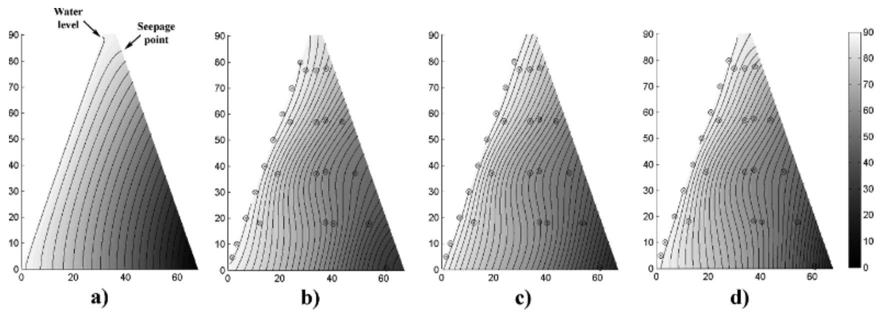

Fig. 2 Results for Dam 1 a) Numerical head field used as a secondary variable, b) Head field obtained by OK, c) Head field obtained by UK (linear drift), d) Head field obtained by KDE (one variable). Circles on the upstream (left) edge are dummy points with head equal to the reservoir level

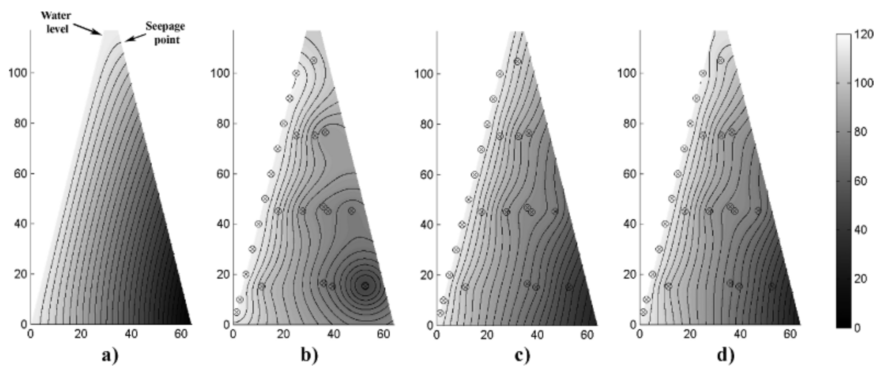

Fig. 3 Results for Dam 2 a) Numerical head field used as a secondary variable, b) Head field obtained by OK, c) Head field obtained by UK, d) Head field obtained by KED. Circles on the upstream (left) edge are dummy points with head equal to the reservoir level

Table 2 Cross-validation results for the dams

Method	Dam 1 - mae_{cv} (m)	Dam 2 - mae_{cv}(m)
OK	3.0	3.9
UK linear drift	2.0	2.2
KED	1.8	2.7

KED and UK provide comparable statistics, especially for the second dam, because the flow simulator solutions (Fig. 2a and 3a) are well approximated by a plane. Comparison of Fig. 2c and 3c to 2d and 3d respectively, also confirms the similarity between the two solutions. However, the no-flow condition is better reproduced by KED than by UK as KED head contours are more perpendicular to the boundaries.

KED does not exactly reproduce the seepage face condition. The head at the bottom downstream extremity of the dam is higher than zero. However, this estimate is consistent with positive pore pressures that are measured at the bottom of the filter, close to the seepage face. This shows that when data is incompatible with the specified conceptual model, KED gives priority to the data but still tries to match the constraints imposed by the model farther from the data (e.g. no-flow at the bottom boundary).

4 Discussion

In the aquifer synthetic case study, KED yielded a very small estimation error. The information required to obtain the numerical fields used in KED is realistic and limited to the main characteristics of the studied field. The values of the Dirichlet boundary conditions, the pumping rate of wells, or the transmissivity of formations need not be known with precision. Rough estimates used in the flow simulation will be implicitly adjusted by KED, as it is known that multiplying the drift function by a constant does not change kriging results. Hence, the choice for constant head values at the left and right boundaries has no impact on the results. In our case study, the transmissivity ratio between the different media could be changed by order of magnitudes without affecting the precision of KED results as long as the left unit is specified as more permeable. The position of the contact between units needs not to be known perfectly either. This demonstrates the great flexibility of the approach and its applicability in routine hydrogeological studies. With the earth dams' cases, we essentially reach the same conclusions as with the synthetic case. In the dams' examples, the UK solution was closer to the KED solution due to the fact that the studied problem possessed a head field close to a plane. As expected, the estimation errors are more important than in the ideal synthetic case. The errors could be due to: measurement errors; incompatibilities with a homogeneous numerical model; information loss in neglecting the effect of the downstream sand filter; assumption of impervious foundations not totally met; lack of information about what is really occurring at the downstream boundary; variations in the geometry of the dam along the third dimension (not modelled here).

Nonetheless, the results obtained by external drift kriging showed noticeable improvements over OK and, to a lesser extent, over UK. New geological information (e.g. foundation permeability; sand filter to consider) could improve the realism of the current numerical model. We stress that, although they are more realistic than OK and UK, the head fields obtained by KED do not totally comply with the conceptual model. As the earth dam case study showed, it depends on the compatibility of the data with the assumed conceptual model(s) used to define the numerical drift(s). In the dam case, data was not compatible with a seepage face along the downstream edge; therefore, the KED solution did not conform to it. However, KED was still able to reproduce faithfully the other boundary conditions. The results of KED could point out some weaknesses of the conceptual model that need improvement. Of course, after such a revision, drift functions could be recomputed and KED re-run.

Note that KED kriging variances provide a (non-local) measure of head uncertainty. Alternatively, head conditional simulations could be done as a way to assess hydraulic conductivity uncertainty in inverse methods, as well as to study the sensitivity to aspects of the conceptual model (e.g. geometry, boundaries) or to covariance parameters.

The approach was applied for the estimation of 3D head fields in the study of synthetic cases (Rivest, 2005). Results obtained (not shown) were in all points similar to the 2D case presented in this work. The proposed method enables fast estimation of entire hydraulic head fields showing a good degree of realism. The use of these fields in direct methods of inversion (e.g. Pasquier and Marcotte (2005, 2006)) is currently investigated.

5 Conclusion

Simple hydrogeological conceptual models, coupled with flow simulation, enable the definition of numerical drift functions to be used in kriging with external drift. The kriged head fields obtained by KED using this approach are more precise and more realistic than those obtained by ordinary kriging and universal kriging with a linear drift. It provides a simple way to transfer most valuable qualitative and semi-quantitative knowledge about the studied field directly in the kriged head maps without using calibration.

Acknowledgments Financial support for this work was provided by NSERC (Canada) research grant. Thanks to Jérémie Gaucher for editing the manuscript. Constructive comments from two anonymous referees helped to improve the manuscript.

References

Ahmed S, de Marsily G (1987) Comparison of geostatistical methods for estimating transmissivity and specific capacity. Water Resour Res 23 (no. 9):1717–1737

Brochu Y, Marcotte D (2003) A simple approach to account for radial flow and boundary conditions when kriging hydraulic head fields for confined aquifers. Math Geol 35 (no. 2): 111–136

Castelier E (1993) Dérive externe et régression linéaire: compte-rendu des journées de géostatistique, Fontainebleau: cahiers de géostatistique, Fascicule 3, pp 47–59

Chapuis RP, Aubertin M (2001) A simplified method to estimate saturated and unsaturated seepage through dikes under steady-state conditions, Can Geotech J 38 (no. 6):1321–1328

Comsol AB (2004) Femlab 3.0 User and reference manual, Stockholm, Sweden

Dehomme JP (1979) Étude de la géométrie du réservoir de Chemery. Internal report, Centre d'informatique géologique, École des Mines de Paris, Fontainebleau

Delhomme JP (1979) Kriging under boundary conditions, Presented at the American Geophysical Union fall meeting, San Francisco

Draper NR, Smith H (1981) Applied regression analysis (2nd edn). Wiley, New York

Galli A, Meunier G (1987) Study of a gas reservoir using the external drift method. In: Matheron G, Armstrong M (eds) Geostatistical case studies, Reidel D, Dordrecht pp 105–120

Marcotte D (1995) Generalized cross-validation for covariance model selection and parameter estimation. Math Geo 27 (no. 6):749–762

Pasquier P, Marcotte D, (2005) Solving the groundwater inverse problem by successive flux estimation. In: Renard et al (eds) GeoEnv 2004: Geostatistics for environmental applications. Springer, Dordrecht, pp 297–308

Pasquier P, Marcotte D, (2006) Steady- and transient-state inversion in hydrogeology by successive flux estimation. Adv Water Resour 29:1934–1952

Rivest M (2005) Rapport de projet de fin d'études, École Polytechnique de Montréal, p 41

Sagar B, Yakowitz S, Duckstein L (1975) A direct method for the identification of the parameters of dynamic nonhomogeneous aquifers, Water Resour Res 11 (no. 4):563–570

Tonkin MJ, Larson SP (2002) Kriging water levels with a regional-linear and point-logarithmic drift, Ground Water 33 (no. 1):185–193