

---

# Geophysical data integration, stochastic simulation and significance analysis of groundwater responses using ANOVA in the Chicot Aquifer system, Louisiana, USA

Ashoka Rahman · Frank T-C. Tsai ·  
Christopher D. White · Douglas A. Carlson ·  
Clinton S. Willson

**Abstract** Data integration is challenging where there are different levels of support between primary and secondary data that need to be correlated in various ways. A geostatistical method is described, which integrates the hydraulic conductivity ( $K$ ) measurements and electrical resistivity data to better estimate the  $K$  distribution in the Upper Chicot Aquifer of southwestern Louisiana, USA. The  $K$  measurements were obtained from pumping tests and represent the primary (hard) data. Borehole electrical

resistivity data from electrical logs were regarded as the secondary (soft) data, and were used to infer  $K$  values through Archie's law and the Kozeny-Carman equation. A pseudo cross-semivariogram was developed to cope with the resistivity data non-collocation. Uncertainties in the auto-semivariograms and pseudo cross-semivariogram were quantified. The groundwater flow model responses by the regionalized and coregionalized models of  $K$  were compared using analysis of variance (ANOVA). The results indicate that non-located secondary data may improve estimates of  $K$  and affect groundwater flow responses of practical interest, including specific capacity and drawdown.

---

Received: 27 June 2007 / Accepted: 27 November 2007  
Published online: 5 March 2008

© Springer-Verlag 2007

---

A. Rahman (✉) · D. A. Carlson  
Louisiana Geological Survey,  
Louisiana State University,  
3079 ECE Building,  
Baton Rouge, LA 70803, USA  
e-mail: arahma1@lsu.edu  
Tel.: +1-225-5785879  
Fax: +1-225-5783662  
e-mail: dcarlson@lsu.edu

F. T-C. Tsai  
Department of Civil and Environmental Engineering,  
Louisiana State University,  
3418G Patrick F. Taylor Hall,  
Baton Rouge, LA 70803, USA  
e-mail: ftsai@lsu.edu

C. D. White  
Department of Petroleum Engineering,  
Louisiana State University,  
3520B Patrick F. Taylor Hall,  
Baton Rouge, LA 70803, USA  
e-mail: cdwhite@lsu.edu

C. S. Willson  
Department of Civil and Environmental Engineering,  
Louisiana State University,  
3513D Patrick F. Taylor Hall,  
Baton Rouge, LA 70803, USA  
e-mail: cwillson@lsu.edu

**Résumé** L'intégration de données entre en jeu lorsque plusieurs niveaux intermédiaires d'assistance sont nécessaires pour corrélérer données primaires et secondaires de diverses manières. Le présent article décrit une méthode géostatistique qui intègre les mesures de conductivité hydraulique ( $K$ ) et les données de résistivité électrique, afin d'estimer plus efficacement la distribution de  $K$  dans l'Aquifère Supérieur de Chicot, au sud-ouest de la Louisiane (Etats-Unis). Les mesures de  $K$  sont issues des pompages d'essai et représentent les données primaires ("dures"). Les données des diagraphies de résistivité électrique ont été considérées comme des données secondaires ("molles"), à partir desquelles les valeurs de  $K$  ont été déduites, par la loi d'Archie et l'équation de Kozeny-Carman. Un pseudo semi-variogramme croisé a été développé afin de pallier à l'absence de colocalisation des données de résistivité. Les incertitudes sur les semi-variogrammes automatiques et sur les pseudo semi-variogrammes croisés ont été quantifiées. Les réponses du modèle d'écoulement des eaux souterraines aux modèles régionalisés et co-régionalisés de  $K$  ont été comparés, par les analyses de variance (ANOVA). Les résultats montrent que les données secondaires non-colocalisées peuvent améliorer les estimations de  $K$ , et affecter efficacement les réponses des écoulements souterrains, y compris les débits spécifiques et les rabattements.

**Resumen** La integración de datos es un gran desafío cuando existen diferentes niveles de apoyo entre datos primarios y secundarios que es necesario correlacionar de varias maneras. Se describe un método geoestadístico el cual integra mediciones de conductividad hidráulica ( $K$ ) y datos de resistividad eléctrica para tener una mejor estimación de la distribución de  $K$  en el Acuífero Chicot Superior del suroeste de Luisiana, Estados Unidos de América. Las mediciones de  $K$  se obtuvieron de pruebas de bombeo y representan los datos primarios (duros). Los datos de sondeos de resistividad eléctrica se consideraron como datos secundarios (suaves) y se usaron para inferir valores de  $K$  a través de la ley de Archie y la ecuación de Carman-Kozeny. Se desarrolló un pseudo semivariograma cruzado para enfrentar la falta de colocación de datos de resistividad. Se cuantificaron las incertidumbres en los auto-semivariogramas y en los semivariogramas cruzados. Las respuestas del modelo de flujo de agua subterránea por los modelos coregionalizados y regionalizados de  $K$  se compararon usando el análisis de varianza (ANOVA). Los resultados indican que los datos secundarios no colocados pueden mejorar los estimados de  $K$  y afectar las respuestas de flujo de agua subterránea de interés práctico, incluyendo capacidad específica y descenso.

**Keywords** Groundwater management · Geostatistics · Variogram uncertainty · ANOVA · USA

## Introduction

Groundwater flow models have been used in many regulatory applications such as well pumping, aquifer sustainability, and capture zone analysis. Simulated drawdown from a groundwater flow model has also been used to determine well specific capacity. Specific capacities reflect hydraulic conductivity or potential of the aquifer, which affects groundwater resource management. For example, management guidelines may stipulate that the average water level for a region should not drop below specified levels. Therefore, uncertainty in model responses leads to uncertainty in the management field.

Groundwater modeling requires hydrogeological parameters as well as other information, e.g., initial/boundary conditions and discharge/recharge fluxes, to constitute a model structure. Although mathematical problems associated with model structure account for all components in the groundwater model, hydraulic conductivity is one of the most essential aquifer parameters in the inverse problem of groundwater hydrology (McLaughlin and Townley 1996). The importance of studying hydraulic conductivity arises from its wide spatial variability (Koltermann and Gorelick 1996) and its significant influence on dispersion and mass transport (Rehfeldt et al. 1992; Neuman 1990; Poeter and Gaylord 1990).

A number of studies have attempted to evaluate and compare various techniques for estimating hydraulic

conductivity fields. Ritzi et al. (1994) compared three indicator-based geostatistical methods to predict zones of higher hydraulic conductivity. Eggleston et al. (1996) compared the hydraulic conductivity field and its sensitivity by using two estimation methods (i.e., kriging and conditional mean) and two simulation methods (i.e., sequential Gaussian and simulated annealing) and found that simulation methods are better at reproducing local contrasts and large-scale features. Boman et al. (1995) studied the response of a transport model to four schemes (three deterministic and one fractal-based stochastic) to interpolate field-measured hydraulic conductivity data, concluding that kriging and fractal interpolation were not significantly better than simpler methods; this study was based on densely sampled measurements which may not be typical of groundwater studies.

In groundwater studies, measurements of hydraulic conductivity are typically sparse. Conductivity fields estimated using regionalized (univariate) models such as kriging or simulation are therefore highly uncertain, as these methods depend on data density. Eggleston et al. (1996) studied a heavily sampled aquifer and reported that the number of data points significantly affects permeability field models. Therefore, data integration from various sources is essential. Hydraulic conductivity estimation with hydrologic data (e.g., hydraulic heads) has been extensively studied through the Bayesian maximum likelihood estimation (Carrera and Neuman 1986) and geostatistics of cokriging (Kitanidis and Vomvoris 1983). Many types of geophysical data have been incorporated in hydraulic conductivity estimation due to their relatively low cost, abundant field measurements, and high correlation with hydraulic conductivity. Geophysical data integration through geostatistics improves the estimation of conductivity (Gloaguen et al. 2001).

In a hydrogeological framework, the spatial structures are constructed from field-measured data. However, field measurements are always limited. Uncertainties associated with the field measurements lead to uncertainties in the correlation structure. Kitanidis (1986) examined the effect of semivariogram model uncertainty in a Bayesian framework. Feyen et al. (2003) examined semivariogram model uncertainty for capture zones in a Bayesian framework, concluding that predictions based on fitted models do not reflect field variance.

This research focuses on a real-world case study, in the estimation of hydraulic conductivity of the Upper Chicot Aquifer, Acadia Parish, in southwestern Louisiana using hydrogeological data and geophysical data. The case study presents many challenges that are commonly encountered in regional groundwater modeling. First of all, the measured hydraulic conductivity data (primary data) from pumping tests and the borehole electrical resistivity data (secondary data) are integrated through a geostatistical approach to better estimate the hydraulic conductivity field. Borehole resistivity data correlate well with the effective porosity of a saturated formation (Archie 1942),

which consequently infers the hydraulic conductivity. Second, a pseudo cross-semivariogram is introduced in the geostatistical framework to cope with the estimation challenge stemming from non-collocation of the resistivity logs and pumping test sites. This study adopts the approach suggested in Clark et al. (1989) to calculate the pseudo cross-semivariogram of these two data sets. Details of the pseudo cross-semivariogram technique are given in Myers (1991). Third, the semivariogram uncertainty due to data scarcity is quantified in terms of mean and variance of the semivariogram at each lag distance using an approach suggested by Ortiz and Deutsch (2002). Assuming lognormal distribution in auto-semivariograms and pseudo cross-semivariograms, the upper and lower limits of the semivariograms, given a confidence interval, are obtained with the requirement of positive definiteness (Goovaerts 1997). The study uses the mean, upper limit and lower limit semivariograms to examine the impact of semivariogram uncertainty on the aquifer hydraulic responses. Fourth, this study develops a groundwater flow model to examine the impact of regionalized and coregionalized hydraulic conductivity fields, semivariogram uncertainty, and simulation processes in a groundwater system. Two regionalized models (kriging and sequential Gaussian simulation) and two coregionalized models (cokriging and cosimulation) are compared. Analysis of variance (ANOVA) assesses the significance of secondary data, interpolation method, and semivariogram uncertainty on the groundwater flow behavior. The relevant methods are discussed first. Data and modeling results for the parish-scale study (25 km<sup>2</sup>) are then described. Finally, the results are interpreted statistically and implications for model choice and interpretation are presented.

### Hydrogeologic interpretation of borehole electrical resistivity data

Borehole electrical resistivity readings reflect the combination of geologic formation and water ionic content with depth. In general, when fresh groundwater is present, electrical resistivity identifies the interfaces between the clay layers (low resistivity value) and sand aquifers (high resistivity value) due to high clay porosity and conductive minerals on the clay surface. Another strong influence on electrical resistivity is the presence of fluid salinity; erratic electrical resistivity readings are often found in sand aquifers when salt water is present.

Typical resistivity data for the Chicot Aquifer are shown in Fig. 1. The resistivity readings distinguish the Upper Chicot Aquifer and Lower Chicot Aquifer, separated by a thin clay layer with low resistivity around a depth of 122 m (400 ft). The erratic resistivity readings in Fig. 1b also show the salt-water intrusion in deep aquifers due to heavy pumping activity. The areas affected by salt water will not be considered in this study.

Borehole resistivity in the saturated sand aquifer has a high correlation to the formation porosity and water resistivity, which is often expressed as Archie's law (Archie 1942):

$$R_0 = a \Phi^{-m} R_w \quad (1)$$

where  $R_0$  is the resistivity of the saturated formation,  $R_w$  is the formation water resistivity,  $\Phi$  is the sand porosity. Two parameters  $a$  and  $m$  in Eq. (1) represent the pore geometry coefficient and the cementation factor, respectively. Typically, the pore geometry coefficient  $a$  varies between 0.62 and 2.45, and the value of the cementation factor  $m$  has a range between 1.08 and 2.15 depending on the formation type (Asquith and Gibson 1982). The formation water resistivity varies with local geochemical and ionic content. However, when fresh groundwater is considered, the variation is not significant. Assuming constant formation water resistivity is reasonable.

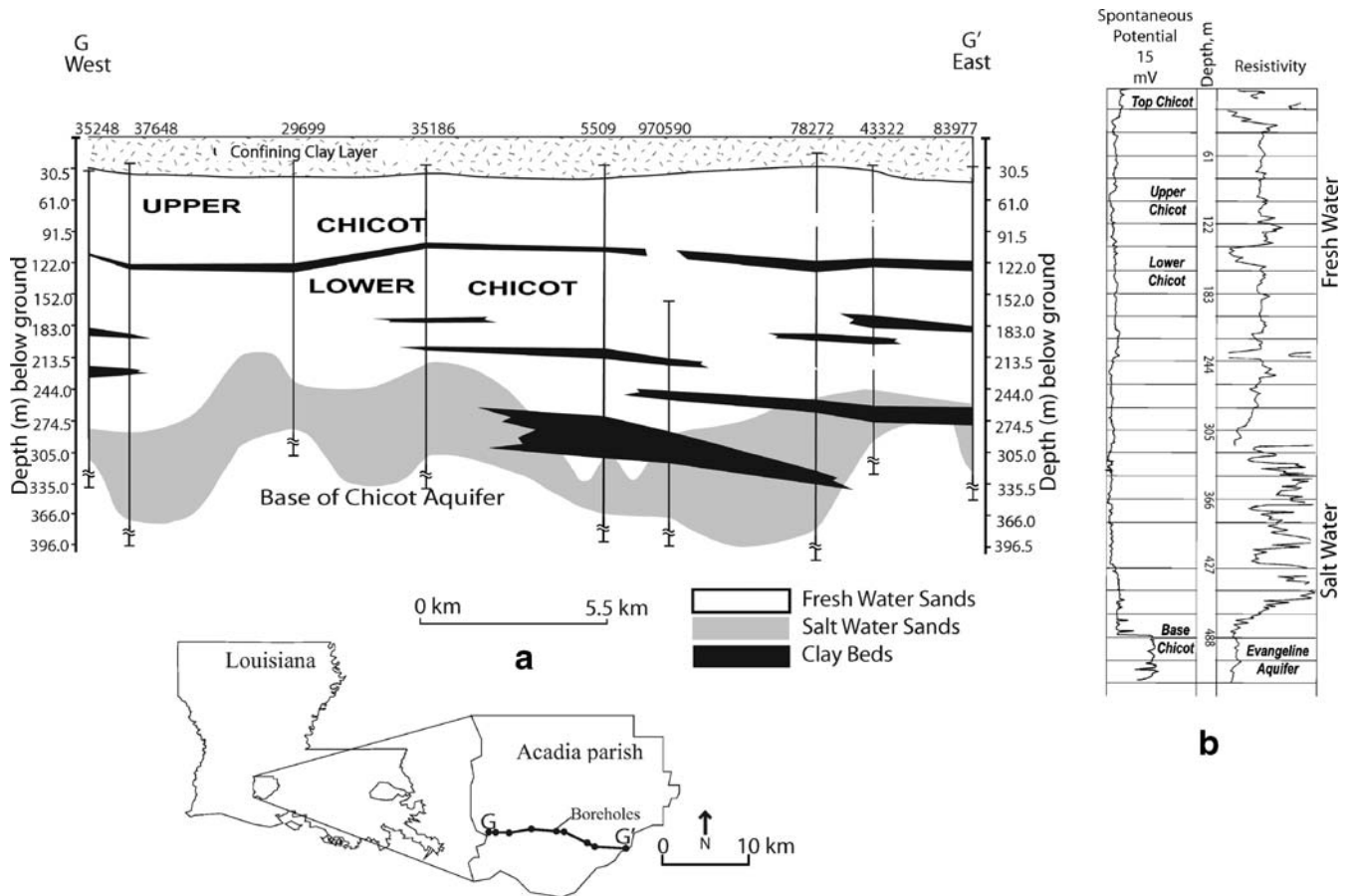
The formulas for determining hydraulic conductivity from particle-size distribution are reduced to the following generalized formula (Vukovic and Soro 1992).

$$K = \frac{\gamma_w}{\mu_w} CG(\Phi)d_e^2 \quad (2)$$

where  $K$  is the hydraulic conductivity,  $\gamma_w$  is the specific weight of water,  $\mu_w$  is the dynamic viscosity of water,  $C$  is a dimensionless parameter,  $G(\Phi)$  is the function of porosity, and  $d_e$  is the effective grain diameter. The term  $CG(\Phi)d_e^2$  in Eq. (2) refers to the soil permeability. In Eq. (2), the hydraulic conductivity can be evaluated with the resistivity data and Archie's law. In this study, the Kozeny-Carman equation (Carman 1956) is adopted, where  $C=1/180$  and  $G(\Phi) = \Phi^3(1 - \Phi)^{-2}$ . Substituting Eq. (1) into Eq. (2), the hydraulic conductivity relating to formation/water resistivity results and is denoted as  $K_F$ :

$$K_F = \frac{\gamma_w}{180\mu_w} \frac{(R_0/R_w)^{-1/m} a^{3/m}}{[(R_0/R_w)^{1/m} - a^{1/m}]^2} d_e^2 \quad (3)$$

where the ratio  $R_0/R_w$  is defined as the formation factor (Archie 1942). The formation factor obtained from borehole resistivity logs (Asquith and Gibson 1982) is a relatively common and low-cost measurement, particularly compared to pumping tests. Kelly (1977) and Kosinski and Kelly (1981) studied site-specific electrical resistivity and hydraulic conductivity and established empirical relationships between the formation factor and hydraulic conductivity. However, their studies did not investigate the theoretical basis for the correlation and the empirical relationship is not necessarily applicable to other sites. The study reported here interprets the hydraulic conductivity using the geophysical resistivity data through Archie's law (1942) and the Kozeny-Carman equation (Carman 1956) for sand units.



**Fig. 1** a Cross section G-G' through southern Acadia Parish, (location also shown in Fig. 2) b geophysical log illustrating the various horizons in the Chicot Aquifer, Acadia Parish, Louisiana (well No. 078272; after Hanson et al. 2001). The wavy lines in boreholes indicate that they continue beyond the depth shown in the figure

**Data integration and simulation via geostatistics**

**Data non-collocation and pseudo cross-semivariogram**

Data integration of the resistivity-driven hydraulic conductivity and the measured hydraulic conductivity by pumping tests is possible through a geostatistical approach. Commonly, the measured hydraulic conductivity denoted as  $K_Y$  is considered as the primary data because the hydraulic conductivity is measured directly through the groundwater responses. The measured and resistivity-driven hydraulic conductivity data are considered as the random fields and are transformed to a normal distribution (the hydraulic conductivity is usually considered as a lognormal distribution) such that these two types of data are classified as second-order stationary. Let  $Y$  and  $F$  be the data transform of  $K_Y$  and  $K_F$ , respectively. If both  $Y$  and  $F$  have sample values at the same locations, the cross-semivariogram of  $Y$  and  $F$  can be obtained for the collocated data

where  $Y_i = Y(\mathbf{x}_i)$ ;  $Y_{i+h} = Y(\mathbf{x}_i + h)$ ;  $h$  is the lag distance between a pair of sample data;  $n_c$  is the number of pairs of  $Y$  or  $F$  at a distance  $h$  apart. However, in practice the electrical resistivity log sites rarely coincide with the pumping test sites. Equation (4) fails for non-collocated data and a cross-semivariogram between  $Y$  and  $F$  cannot be obtained. To cope with this problem, a pseudo cross-semivariogram has to be introduced to replace Eq. (4). For this study, the pseudo cross-semivariogram introduced by Clark et al. (1989) was used, where the two random functions must have the same unconditional mean and unconditional variance. The easiest way to achieve this requirement is to transform the primary and secondary data into the standard normal distribution (normal score) that has zero mean and unity variance (Deutsch and Journel 1998). In this study, the  $K_Y$  and  $K_F$  are transformed to the normal scores  $y$  and  $f$ , respectively, with zero means and unity variances. The pseudo cross-semivariogram is given by the expected squared difference between  $y$  and  $f$  measured at different locations,

$$\hat{\gamma}_{YF}(h) = \frac{1}{2n_c(h)} \sum_{i=1}^{n_c(h)} (Y_i - Y_{i+h})(F_i - F_{i+h}) \quad (4)$$

$$\tilde{\gamma}_{yf}(h) = \frac{1}{2n_{yf}(h)} \sum_{j=1}^{n_{yf}(h)} (y_j - f_{j+h})^2 \quad (5)$$



where  $\tilde{\gamma}_{yf}$  is the pseudo cross-semivariogram; and  $n_{yf}$  is the number of pairs of  $y$  and  $f$  at a distance  $h$  apart. It can be shown that  $\tilde{\gamma}_{yf}(h) = \tilde{\gamma}_{fy}(-h)$ . The auto-semivariograms for  $y$  and  $f$  are:

$$\hat{\gamma}_{yy}(h) = \frac{1}{2n_{yy}(h)} \sum_{j=1}^{n_{yy}(h)} (y_j - y_{j+h})^2 \quad (6)$$

and

$$\hat{\gamma}_{ff}(h) = \frac{1}{2n_{ff}(h)} \sum_{j=1}^{n_{ff}(h)} (f_j - f_{j+h})^2 \quad (7)$$

Details of the pseudo cross-semivariogram technique can be found in Myers (1991). Due to the normalization, the sills for  $\hat{\gamma}_{yy}$ ,  $\hat{\gamma}_{ff}$ , and  $\tilde{\gamma}_{yf}$  are unity as the unconditional variance. Moreover, the Cauchy-Schwarz inequality  $\tilde{\gamma}_{yf} \leq \sqrt{\hat{\gamma}_{yy}\hat{\gamma}_{ff}}$  is required to ensure the positive definiteness in the following cokriging method. The Cauchy-Schwarz inequality will be used to determine the lower and upper limits of the auto- and pseudo cross-semivariograms when semivariogram uncertainty is considered in the section **Significance analysis using ANOVA**.

### Conditional estimation using hydrogeological data and resistivity data

The conditional estimation of the normalized hydraulic conductivity is made through a linear-weighted interpolation form, which honors the normalized primary and secondary data:

$$\hat{y}_{CoK} = \sum_{j=1}^{n_y} \lambda_j y_j + \sum_{\ell=1}^{n_f} \nu_\ell f_\ell \quad (8)$$

where  $\hat{y}$  is the estimate of  $y$ ;  $\lambda_j$  and  $\nu_\ell$  are the weighting coefficients that need to be determined.  $n_y$  is the number of primary data; and  $n_f$  is the number of the secondary data. The conditional estimate of hydraulic conductivity is obtained via the back transformation of the normal score. By minimizing the variance of estimation error  $[\hat{y}(x_0) - y(x_0)]$  at an unsampled site  $x_0$ , the optimal weighting coefficients  $\lambda_j$  and  $\nu_\ell$  are obtained by the following set of linear equations,

$$\begin{cases} \sum_{j=1}^{n_y} \lambda_j \hat{\gamma}_{yy}(\mathbf{x}_k, \mathbf{x}_j) + \sum_{\ell=1}^{n_f} \nu_\ell \tilde{\gamma}_{yf}(\mathbf{x}_k, \mathbf{x}_\ell) + \mu = \hat{\gamma}_{yy}(\mathbf{x}_0, \mathbf{x}_k), & k = 1, 2, \dots, n_y \\ \sum_{j=1}^{n_y} \lambda_j \tilde{\gamma}_{fy}(\mathbf{x}_i, \mathbf{x}_j) + \sum_{\ell=1}^{n_f} \nu_\ell \hat{\gamma}_{ff}(\mathbf{x}_i, \mathbf{x}_\ell) + \mu = \tilde{\gamma}_{fy}(\mathbf{x}_0, \mathbf{x}_i), & i = 1, 2, \dots, n_f \\ \sum_{j=1}^{n_y} \lambda_j + \sum_{\ell=1}^{n_f} \nu_\ell = 1 \end{cases} \quad (9)$$

where  $\mu$  is the Lagrange multiplier. Equation (9) involves auto-semivariograms of  $y$  and  $f$  as well as the pseudo cross-semivariogram for non-located data fusion.

If primary data are sparse and more densely sampled secondary data are strongly correlated to the primary variable, geostatistics of cokriging can help improve estimates of the primary variable. In this work, the secondary data (the borehole resistivity logs) are non-located with the pumping test sites. Therefore, a pseudo cross-semivariogram,  $\tilde{\gamma}_{yf}$  has to be used to replace the cross-semivariogram. For the comparison purpose, the kriging conditional estimation, which uses only the measured hydraulic conductivity data, is also considered

$$\hat{y}_{OK} = \sum_{j=1}^{n_y} \lambda_j y_j \quad (10)$$

Therefore, Eq. (9) is reduced to the following for  $y$ ,

$$\begin{cases} \sum_{j=1}^{n_y} \lambda_j \hat{\gamma}_{yy}(\mathbf{x}_i, \mathbf{x}_j) + \mu = \hat{\gamma}_{yy}(\mathbf{x}_0, \mathbf{x}_i), & i = 1, 2, \dots, n_y \\ \sum_{j=1}^{n_y} \lambda_j = 1 \end{cases} \quad (11)$$

### Conditional simulation and cosimulation of hydraulic conductivity

The conditional estimation of hydraulic conductivity obtained by the kriging and cokriging methods is a smoothed distribution over the region, which does not reveal the actual variability of the hydraulic conductivity. However, the  $K$  variability has a significant influence on the groundwater responses when a groundwater flow model is adopted (Fogg 1986). In this study, conditional simulation and conditional cosimulation will be conducted to study the  $K$  variability in relation to groundwater model responses in the real-world case study. Moreover, the significance analysis using the conditional cosimulation against conditional simulation in groundwater responses will be conducted using analysis of variance (ANOVA), and is discussed later. Conditional simulation and cosimulation techniques simulate the estimation error (kriging error) as a correlated random process using the semivariogram and cross-semivariogram. For non-located data, a pseudo cross-semivariogram is adopted in the cosimulation process. This section provides the information and procedure for conducting cosimulation of the measured hydraulic conductivity and the resistivity-derived hydraulic conductivity in the study. The  $K$  simulation using measured hydraulic conductivity can be inferred by reducing the cosimulation procedure for a univariate model.

A cosimulated hydraulic conductivity field is generated by using a full linear model of cross covariance. The steps followed in the cosimulation algorithm include (1) modeling  $\hat{\gamma}_{yy}$ ,  $\hat{\gamma}_{ff}$ , and  $\hat{\gamma}_{yf}$  for ensuring positive definiteness; (2) obtaining a cokriged estimate of the secondary data  $f_{CoK}$  at computation grids; and (3) calculating hydraulic conductivity estimates (using SGSIM\_FC subroutine; C.V. Deutsch, University of Alberta, Canada, personal communication, 2003). In the study area, the number of borehole resistivity data points is approximately the same number as that of the measured hydraulic conductivity data points. Therefore, the information from the primary data ( $y_j$ ) for the secondary data estimation ( $\hat{f}_{CoK}$ ) is important. This was verified by comparing the kriged estimate against the cokriged estimate of  $f_{CoK}$ . A hybrid method was used to impose joint correlation and the correct variability in the cosimulation procedure:

- Step 1. Obtain a cokriged estimation of secondary data  $\hat{f}_{CoK}$  over the field.
- Step 2. Obtain a single-variable simulation of the correlated residual of the secondary variable only, using the semivariogram for the secondary data and zero at the secondary data locations.
- Step 3. Sum the simulated result in step 2 and the cokriged results in step 1.
- Step 4. Use SGSIM\_FC with the normal score transform of the conductivity data and results from step 3 in place of the secondary variable simulation.
- Step 5. Back-transform from normal score values of the primary data to the hydraulic conductivity values.

The aim here is to get the primary variable structure into the secondary variable via the cokriging step, with the assumption that the mean carries most of the information; the secondary residual is approximated as uncorrelated with the primary data.

Due to scarcity of measured hydraulic conductivity data and electrical-log data, one needs to investigate the uncertainty embedded in the semivariograms  $\hat{\gamma}_{yy}$ ,  $\hat{\gamma}_{ff}$ , and  $\hat{\gamma}_{yf}$ . From Eq. (5), the mean or expected pseudo cross-semivariogram is given by

$$E[\tilde{\gamma}_{yf}(h)] = \frac{1}{2n_{yf}(h)} \sum_{j=1}^{n_{yf}(h)} E(y_j - f_{j+h})^2 \quad (12)$$

The variance of the pseudo cross-semivariogram is given by

$$\begin{aligned} \sigma^2[\tilde{\gamma}_{yf}(h)] &= \frac{1}{4n_{yf}^2(h)} \sum_{i=1}^{n_{yf}(h)} \sum_{j=1}^{n_{yf}(h)} E[(y_i - f_{i+h})^2 (y_j - f_{j+h})^2] \\ &\quad - \{E[\tilde{\gamma}_{yf}(h)]\}^2 \end{aligned} \quad (13)$$

The covariance of  $(y_i - f_{i+h})^2$  and  $(y_j - f_{j+h})^2$  is

$$\begin{aligned} \tilde{C}_{yf}(\mathbf{x}_i, \mathbf{x}_j|h) &= E[(y_i - f_{i+h})^2 (y_j - f_{j+h})^2] \\ &\quad - E[(f_i - f_{i+h})^2] E[(f_j - f_{j+h})^2] \end{aligned} \quad (14)$$

Substituting Eq. (11) into Eq. (10), the pseudo cross-semivariogram variance is

$$\sigma^2[\tilde{\gamma}_{yf}(h)] = \frac{1}{4n_{yf}^2(h)} \sum_{i=1}^{n_{yf}(h)} \sum_{j=1}^{n_{yf}(h)} \tilde{C}_{yf}(\mathbf{x}_i, \mathbf{x}_j|h) \quad (15)$$

Similarly, the auto-semivariogram variances for  $y$  and  $f$  are

$$\sigma_{\gamma_{yy}(h)}^2 = \frac{1}{4n_{yy}^2(h)} \sum_{i=1}^{n_{yy}(h)} \sum_{j=1}^{n_{yy}(h)} \hat{C}_{yy}(\mathbf{x}_i, \mathbf{x}_j|h) \quad (16)$$

and

$$\sigma_{\gamma_{ff}(h)}^2 = \frac{1}{4n_{ff}^2(h)} \sum_{i=1}^{n_{ff}(h)} \sum_{j=1}^{n_{ff}(h)} \hat{C}_{ff}(\mathbf{x}_i, \mathbf{x}_j|h) \quad (17)$$

where  $\hat{C}_{yy}$  is the covariance of  $(y_i - y_{i+h})^2$  and  $(y_j - y_{j+h})^2$ ; and  $\hat{C}_{ff}$  is the covariance of  $(f_i - f_{i+h})^2$  and  $(f_j - f_{j+h})^2$ . Equations (16) and (17) represent the variances of the semivariograms for a given lag distance, which are the average covariances between pairs of pairs used to calculate the semivariograms. If the data in  $y_i$  and  $f_j$  are Gaussian distributions,  $\tilde{C}_{yf}$ ,  $\hat{C}_{yy}$  and  $\hat{C}_{ff}$  can be evaluated by the methods described in Ortiz and Deutsch (2002).

As for the semivariogram uncertainty analysis, the specific lower-limit and upper-limit semivariograms using the semivariogram variances for the auto-semivariogram models and pseudo cross-semivariogram models will be determined according to the real data. The lower and upper limit semivariograms along with the mean semivariograms will be used for the significance analysis using ANOVA.

### Significance analysis using ANOVA

A single estimate of hydraulic conductivity cannot simulate the actual heterogeneity and complexity. In order to better understand the influence of the coregionalized model, ANOVA (analysis of variance) was adopted to differentiate the statistical significance among different approaches on generating  $K$  distributions. ANOVA is able to make inferences about mean differences by comparing variances among different groups (between-group vari-

ance) to the variance within the groups (within-group variance) using  $F$ -test (Fisher 1935). Donnelly-Makowicka and Moore (1999) used a combination of ANOVA and jackknife to test the statistical significance of differences in a hydrologic model performance using different factors. Rong (2002) used ANOVA to compare MTBE (methyl tertiary-butyl ether) concentrations in sand/gravel and fine-grained soils. Ikem et al. (2002) used ANOVA to compare upgradient and downgradient contaminant concentrations near a waste site to find the impact of the waste site on the groundwater quality.

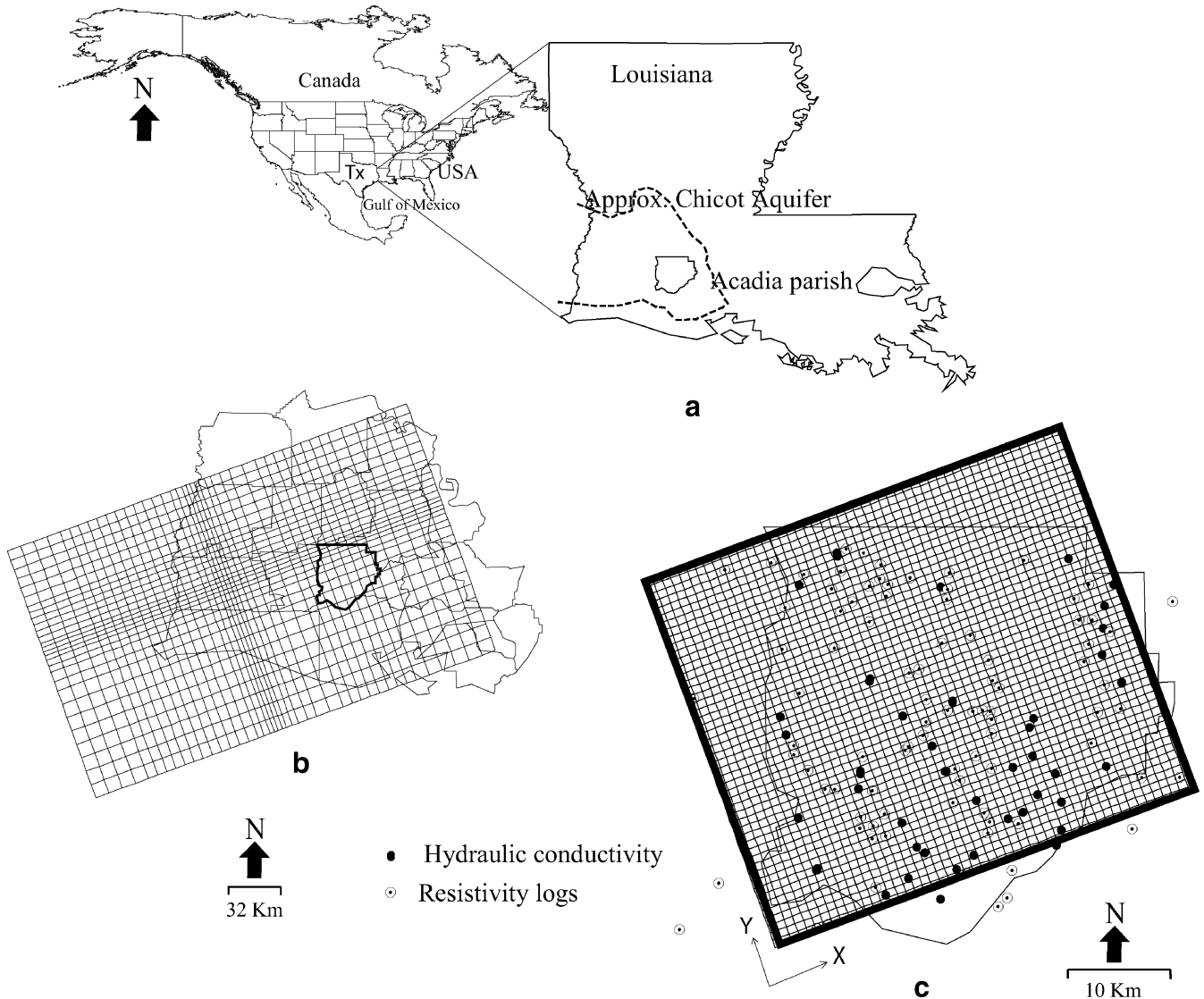
In this study, four groups of simulations, each using an alternative hydraulic conductivity field created by kriging, cokriging, simulation, and cosimulation, were run with variogram means, upper and lower bounds. Statistics from various flow responses are calculated and the significance of the (1) process (kriging versus simulation and cokriging versus cosimulation); (2) variables (kriging versus cokriging and simulation versus cosimulation); and (3) vario-

gram variance (mean, 95% upper and lower bounds) are tested using a multifactor ANOVA test.

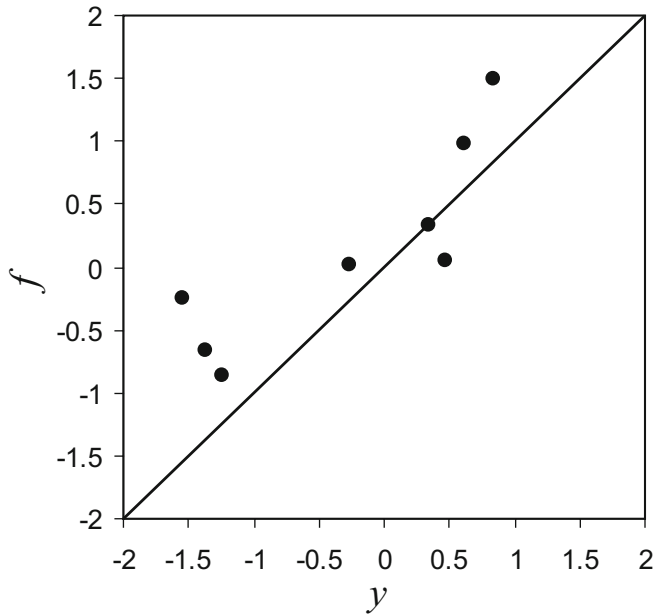
## Case study

### Semivariogram estimation and uncertainty in the upper Chicot Aquifer

The Chicot aquifer system is the principal groundwater source for southwestern Louisiana, which underlies 15 parishes. It is also the most heavily pumped aquifer in Louisiana. Rice irrigation accounts for about 85% of the groundwater pumped from the aquifer in Acadia Parish (Sargent 2002). The Chicot system in Acadia Parish is divided into Upper and Lower Chicot units (sandy aquifers) separated by clay lenses referred to as the Upper/Lower confining zone (Fig. 1). This study focuses on the hydraulic conductivity estimation in the Upper Chicot Aquifer in Acadia Parish.



**Fig. 2** a Study area in Louisiana State; adjacent to Texas (Tx), b Chicot Aquifer model grid with parish boundaries, c Acadia Parish model grid with location of hydraulic conductivity data points and geophysical logs



**Fig. 3** Cross plot between primary ( $y$ ) and secondary ( $f$ ) data at short distances. Both variables have normal score transformed values

The local characterization of the Chicot Aquifer in Acadia Parish was developed using geophysical resistivity logs obtained from 74 oil and gas exploration wells. Note that not all the geophysical logs could be used; only 53 of them were used in the cokriging method as some of them did not cover the Upper Chicot Aquifer. The geology is divided into: confining clay layer, Upper Chicot Aquifer, dividing layer, Lower Chicot Aquifer. Figure 1a shows a west-east cross section of the local model.

*Field K measurements*

There are 42 hydraulic conductivity values (Fig. 2c) determined from specific capacity tests according to Eq. (15) (Bradbury and Rothschild 1985),

$$K_Y = \frac{Q}{4\pi b(s - CQ^2)} \left[ \ln \left( \frac{2.25Kt}{r_w^2 S_s} \right) + 2 \frac{1 - \ell_s/b}{\ell_s/b} \left( \ln \frac{b}{r_w} - G \right) \right] \tag{18}$$

where  $Q$  is the pumping rate;  $s$  is drawdown at time  $t$ ;  $b$  is the Upper Chicot Aquifer thickness;  $\ell_s$  is the length of screen interval;  $r_w$  is the well radius; and  $S_s$  is the specific storage. The term  $CQ^2$  represents the well loss, where  $C$  is the well loss constant. The second term in the bracket

**Table 1** The  $t$ -test statistic on the trend detection in X- and Y-directions using the bilinear model

Data	Bilinear model	$t$ -test	
		$\bar{X}$ , $Pr > t$	$Y$ , $Pr > t$
$y$	$y \sim X + Y$	0.06	0.82
$f$	$f \sim X + Y$	0.67	0.35

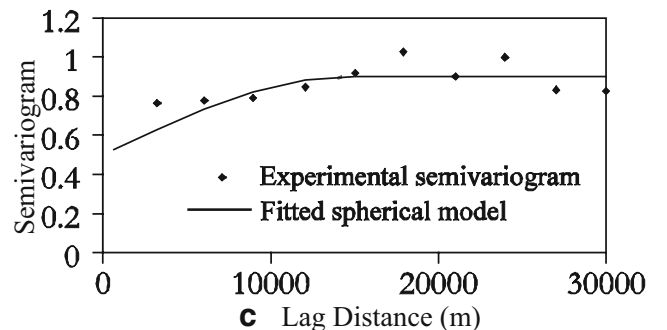
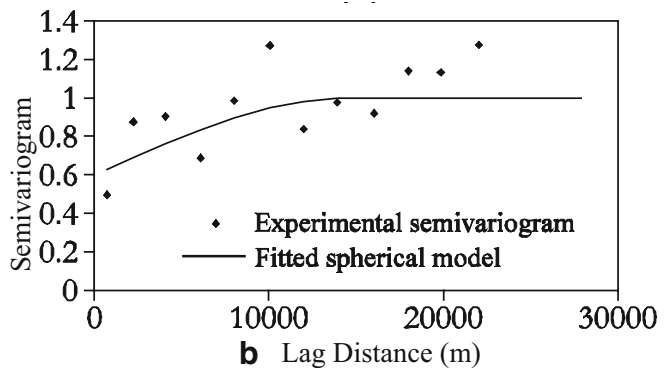
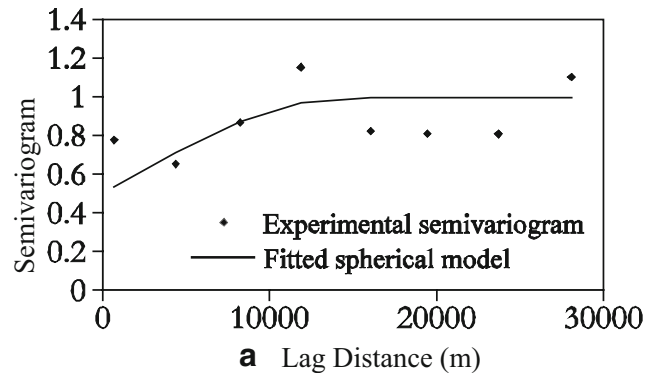
$Pr$  probability

corrects the  $K_Y$  calculation due to partial penetration of the pumping well. The term  $G$  is the polynomial function of the ratio of the screen length to the aquifer thickness,

$$G = 2.948 - 7.363(\ell_s/b) + 11.447(\ell_s/b)^2 - 4.675(\ell_s/b)^3 \tag{19}$$

The technique solves Eq. (15) in an iterative fashion by taking other parameters as an input to Eq. (15). The Upper Chicot Aquifer thicknesses ( $b$ ) were determined by analyzing well log data (Carlson et al. 2003).

The hydraulic conductivity values of this study were determined from specific capacity tests associated with the development of high-capacity wells usually for irrigation, public supply or industrial use. Because of the intended use of these wells, the average test discharge  $Q$  was  $7,850 \pm 6,210 \text{ m}^3/\text{day}$ . The dimensions of the wells are



**Fig. 4** Experimental semivariograms and spherical semivariogram models **a** auto-semivariogram of  $y$ , **b** auto-semivariogram of  $f$ , and **c** pseudo cross-semivariogram of  $y$  and  $f$



**Table 2** Semivariogram model parameters

Semivariogram models		Nugget ( $c_0$ )	Relative sill ( $c_1$ )	Range ( $\times 10^3$ m) ( $c_2$ )	Total sill ( $c_0+c_1$ )
$\hat{\gamma}_{yy}$	Mean	0.5	0.5	1.5	1.0
	Lower limit	0.2	0.7	4.2	1.0
	Upper limit	1.1	4.2	4.5	5.3
$\hat{\gamma}_{ff}$	Mean	0.6	0.4	1.5	1.0
	Lower limit	0.4	1.3	4.2	1.7
	Upper limit	2.0	2.6	4.5	4.6
$\hat{\gamma}_{yf}$	Mean	0.5	0.4	1.5	0.9
	Lower limit	0.0	1.0	4.2	1.0
	Upper limit	1.4	3.3	4.5	4.7

large in response to the intended large demand for water: average radius is  $9.8 \pm 3.1$  cm, and average screen length is  $17.4 \pm 7.1$  m. The specific capacity tests lasted on average  $9.8 \pm 9$  h. These wells are clearly partially penetrating wells in that the average sand thickness surrounding a well screen is  $45.1 \pm 15.9$  m. The average specific capacity test drawdown is  $7.9 \pm 5.0$  m.

From previous studies, the storativity values for the Chicot Aquifer range from  $1.1 \times 10^{-4}$  to  $3 \times 10^{-3}$  with an average of  $6.7 \times 10^{-4}$ . These values are a clear indicator that the Chicot is a confined aquifer (Weight and Sonderegger 2001). The confining clay layer above the Upper Chicot in Acadia Parish has an average thickness of  $28.9 \pm 10.0$  m as determined from examination of 2,648 well log reports.

The well loss constant value used for this study is  $2.66 \times 10^{-8}$   $\text{day}^2/\text{m}^5$ , which is significantly smaller than the one listed in Bradbury and Rothschild (1985), which is  $5.46 \times 10^{-6}$   $\text{day}^2/\text{m}^5$  after unit conversion for  $Q$ . However, a smaller value of well loss constant is used to avoid well loss head exceeding drawdown for this study's high capacity tests were the average  $Q$  is  $7,850 \pm 6,210$   $\text{m}^3/\text{day}$  compared to the smaller tests of  $Q=10$  gallons per meter ( $54.5$   $\text{m}^3/\text{day}$ ) noted in Bradbury and Rothschild (1985).

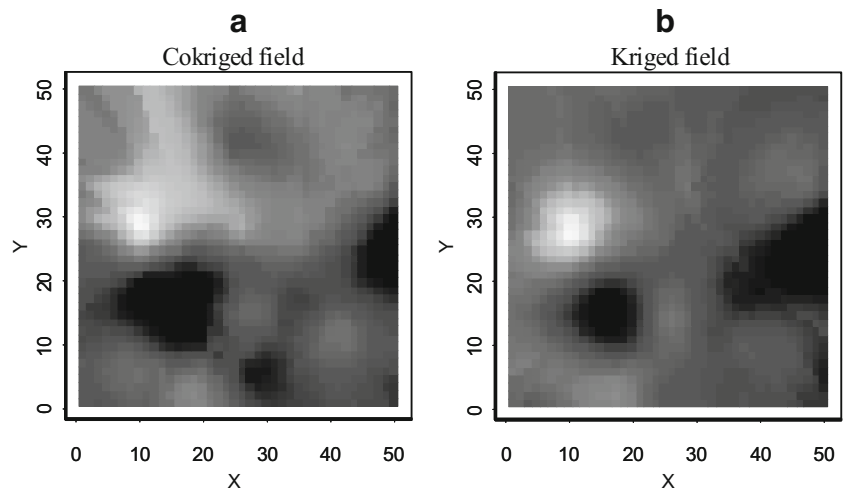
The capacity test data set (42 values) shows that the measured hydraulic conductivity values vary from 5 to 700 m/day with a geometric mean  $\bar{K}_Y = 88$  m/day and a standard deviation  $\sigma_{K_Y} = 115$  m/day.

#### Electrical-log data

A total of 53 oil and gas geophysical resistivity logs (Fig. 2c) were obtained from Louisiana Department of Natural Resources, Office of Conservation (OC), Well Log Library. To obtain the average  $R_0$  for each log, the resistivity readings are divided into 3-m (10-ft) sections between the top and bottom elevation of Upper Chicot Aquifer. Again, it is reasonable to assume constant formation water resistivity because the study area is small and only freshwater is present in the Upper Chicot unit. The groundwater temperature is reported as  $25^\circ\text{C}$ . The formation water resistivity  $R_w=12.7$  ohm-m, water specific weight  $\gamma_w=9.8$  KN/ $\text{m}^3$  and dynamic viscosity are used in this study. The variance in the sand formation resistivity factor ( $R_0/R_w$ ) is therefore the result of differences in sand/clay ratio and porosity. The effective saturated formation resistivity over the depth of the Upper Chicot is the average of the formation resistivity ( $R_0$ ) values calculated at the 3 m (10 ft) intervals. The 53 borehole resistivity logs in the Upper Chicot Aquifer show that the formation resistivity value varies from 43 to 100 ohm-m with a mean of 75 ohm-m and a standard deviation 16 ohm-m. The average effective particle diameter  $d_e$  is 0.42 mm (1.22 in phi units) calculated from US Geological Survey unpublished sieve test data files (US Geological Survey, unpublished data, 2003).

As for the pore geometry coefficient and cementation factor in the Archie's equation, the maximization of the correlation over a short lag distance between the measured hydraulic conductivity ( $y_i$ ) and resistivity-driven hydraulic conductivity  $f_j(a, m)$  in the normal scale was conducted to obtain the optimal values of for  $a$  and  $m$ . This is equivalent to finding the best  $a$  and  $m$  values

**Fig. 5** The conditional estimates of hydraulic conductivity by **a** cokriging and **b** kriging using mean semivariogram models. The  $K$  varies between  $K=5.18$  m/day (black color) to  $K=701$  m/day (white color)



such that the deviation between  $y_i$  and  $f_j(a, m)$  is minimized:

$$\min_{a,m} \sum_{(i,j)} [y_i - f_j(a, m)]^2 \quad (20)$$

Eight pairs of  $y_i$  and  $f_j$  were selected with a lag distance less than 300 m. The bounds of  $a$  and  $m$  were  $0.62 \leq a \leq 2.45$  and  $1.08 \leq m \leq 2.15$ , respectively, according to Asquith and Gibson (1982). It has been noted that the zero-lag distance should be used for collocated data. However, a subjective selection of a short lag distance (300 m) is necessary for non-collocated data. Equation (20) is solved by a gradient-based nonlinear optimization method and the optimal values of  $a$  and  $m$  are found to be 1.76 and 1.64. Figure 3 shows the cross plot for the eight pairs of  $y_i$  and  $f_j$  with optimal  $a$  and  $m$  values, which indicates a nugget between the measured hydraulic conductivity and resistivity-driven hydraulic conductivity data.

The electrical-log data set (53 values) shows that the resistivity-driven hydraulic conductivity values vary from 4 to 66 m/day with a geometric mean  $\bar{K}_F = 14$  m/day and a standard deviation  $\bar{\sigma}_F = 13.7$  m/day.

#### Trend analysis and normal score transformation

The measured hydraulic conductivity data and resistivity-driven hydraulic conductivity data are tested for trend analysis using  $t$ -statistic from a bilinear model. The bilinear model is fit in the X-direction and Y-directions (Fig. 2c) to detect if a trend exists. With a 10% level of significance, Table 1 concludes that the conductivity data have no significant trend in the X or Y direction. Moreover, the measured hydraulic conductivity and resistivity-driven hydraulic conductivity show positive skewness, which indicates that one can assume hydraulic conductivity to be log-normally distributed (Domenico and Schwartz 1990). In this study, the data are standardized to zero mean and unit variance and transformed to normal scores (Deutsch and Journel 1998). The normal transform facilitated computation of the Archie coefficients ( $a$  and  $m$ ) and the pseudo cross-semivariogram, which requires both random functions to have the same mean and variance.

#### Experimental semivariograms and pseudo cross-semivariogram

The 42 measured hydraulic conductivity values and 53 resistivity-driven hydraulic conductivity values are used in conditional estimation of the hydraulic conductivity field for both the regionalized model (i.e.,  $\hat{y}$  in Eq. (10)) and coregionalized model—i.e.,  $\hat{y}$  in Eq. (8). Again, as none of the data in this study area are collocated, a pseudo cross-semivariogram is used instead of the cross-semivariogram. The experimental auto-semivariograms ( $\hat{\gamma}_{yy}$  and  $\hat{\gamma}_{ff}$ ) and pseudo cross-semivariograms ( $\hat{\gamma}_{yf}$ ) are computed using Eqs. (5)–(7) and shown in Fig. 4.

This study adopts a spherical (Sph) semivariogram model with nugget as the following:

$$\gamma_{Sph}(h) = c_0 + c_1 \left\{ \frac{3}{2} \frac{h}{c_2} - \frac{1}{2} \left( \frac{h}{c_2} \right)^3 \right\}, h \leq c_2 \quad (21)$$

where  $h$  is the lag distance;  $c_0$  is the nugget;  $c_1$  is the relative sill;  $(c_0 + c_1)$  is the total sill; and  $c_2$  is the

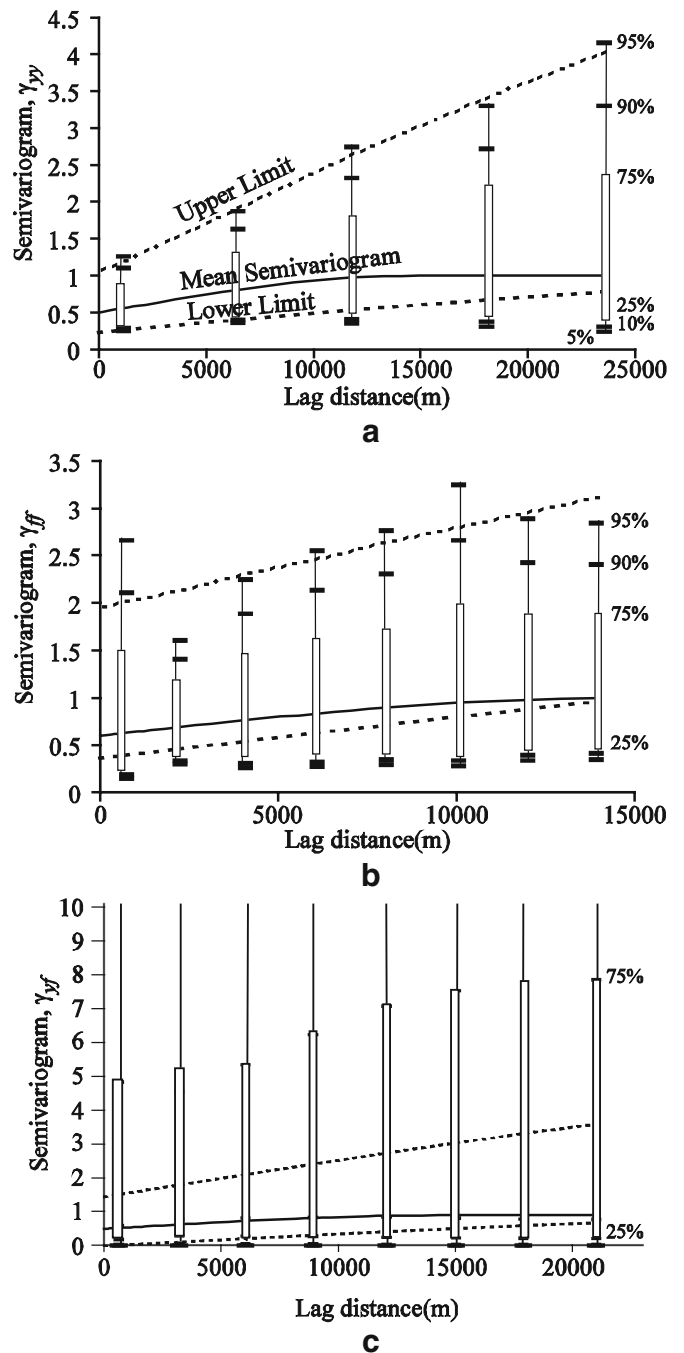


Fig. 6 Semivariograms and their confidence limits for a auto-semivariogram of  $y$ , b auto-semivariogram of  $f$ , and c pseudo cross-semivariogram of  $y$  and  $f$ . The upper hinge of the box indicates the 75th percentile of the data set, and the lower hinge indicates the 25th percentile

semivariogram range. Moreover,  $\gamma_{\text{Sph}}(0)=0$  and  $\gamma_{\text{Sph}} = c_0 + c_1$  for  $h > c_2$ . With the spherical model, the integral scale is

$$I = \int_0^{c_2} \left(1 - \frac{\gamma_{\text{Sph}}(h) - c_0}{c_1}\right) dh = \frac{3}{8} c_2 \quad (22)$$

The parameters (nugget, sill, and range) embedded in the semivariograms are estimated through a weighted normalized least-squares estimation (Cressie 1985):

$$\min_{c_0, c_1, c_2} \sum_{k=1}^n n(h_k) \left\{ \frac{\hat{\gamma}(h_k)}{\gamma_{\text{Sph}}(h_k; c_0, c_1, c_2)} - 1 \right\}^2 \quad (23)$$

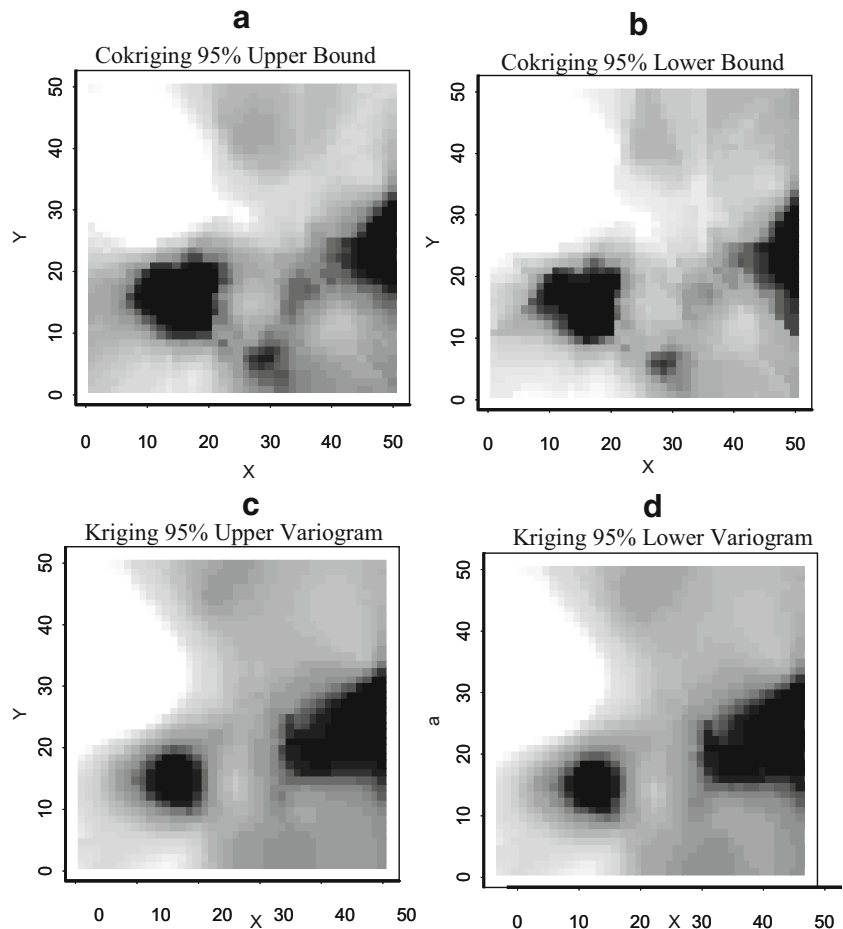
where  $h_k$  is the  $k$ th lag distance;  $\gamma_{\text{Sph}}$  is the spherical semivariogram model; and  $n(h)$  is number of points at each lag representing the weight for the selected lag distance. The minimization problem in Eq. (23) is subjected to the Cauchy-Schwarz inequality conditions to ensure the auto-semivariogram models and pseudo cross-semivariogram model are positive definite (Hohn 1998). The experimental semivariograms in Fig. 4 were fitted by the weighted least-squares estimation method to obtain the optimal values for  $c_0$ ,  $c_1$ , and  $c_2$ . The total sill for the auto-semivariogram is equal to 1, which represents the uncon-

ditional variance of the normal score transformed data. However, the total sill for the pseudo cross-semivariogram is 0.9. Table 2 lists the optimized semivariogram parameter values and shows moderate correlation at short range (as expected when investigating hydraulic conductivity). The resulted semivariogram models are shown in Fig. 4. Especially, the pseudo cross-semivariogram in Fig. 4c shows correlation over a length of approximately 15,000 m, approximately one-half of the length of the study region, which indicates the appropriateness of using cokriging for the correlated primary and secondary data. The semivariogram models determine kriging weights and cokriging weights (see section [Conditional estimation using hydrogeological data and resistivity data](#)), which are used to obtain the conditional estimates and conditional variances. Moreover, these semivariogram models are the means in the following semivariogram uncertainty analysis. The kriged and cokriged hydraulic conductivity distributions using the mean semivariograms are shown in Fig. 5.

#### Semivariogram uncertainty

The experimental auto-semivariograms of  $y$  and  $f$  are the mean semivariograms according to the method described in section [Significance analysis using ANOVA](#). A FOR-

**Fig. 7** The conditional estimates of hydraulic conductivity by cokriging (**a, b**) and kriging (**c, d**) using the upper limit semivariogram model and lower limit semivariogram model. The  $K$  varies between  $K=5.18$  m/day (black color) to  $K=701$  m/day (white color)

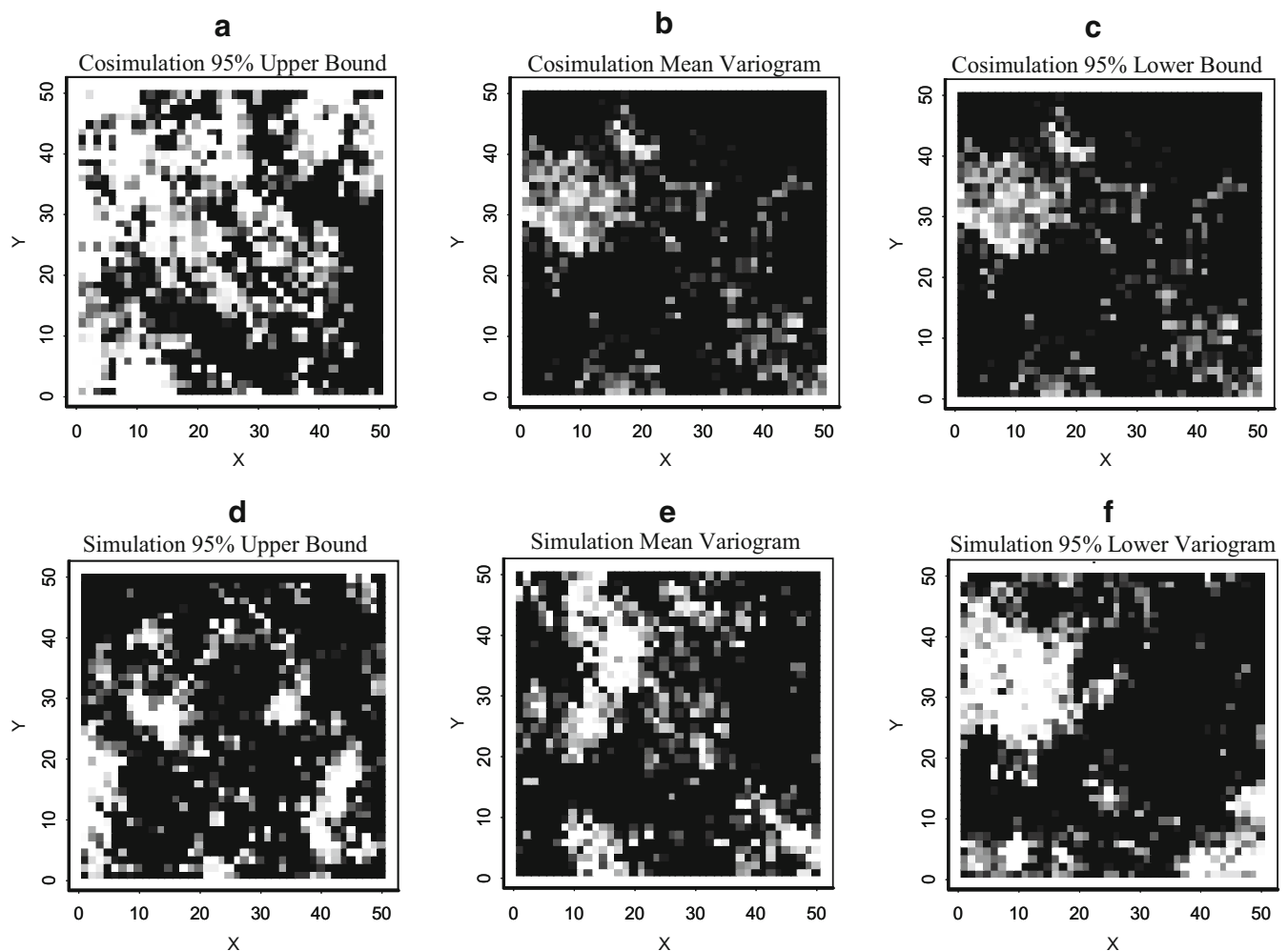


TRAN program (C.V. Deutsch, University of Alberta, Canada, personal communication, 2003) is used to determine the semivariogram variance that allows determination of the lower and upper limits under the assumption that the semivariograms are log-normally distributed at each lag distance. The procedures to determine these limits are as follows.

First, the 90% confidence interval was used to determine the semivariogram limits for  $\hat{\gamma}_{yy}$ . Figure 6a show the 5% percentile, 10% percentile, and so forth for each lag distance. The upper limit semivariogram model is determined using the spherical model to best fit the 95% limits. As for the lower-limit semivariogram model, the spherical model has 5% limit at the first lag distance and gradually moved toward the sill according to Ortiz and Deutsch (2002). Figure 6a shows the lower and upper limit semivariograms, which are positive definite. The upper limit for  $\hat{\gamma}_{ff}$  is also determined by the best fit of the spherical model to the 95% limit at each lag distance as shown in Fig. 6b. However the lower limit of  $\hat{\gamma}_{ff}$  and the

lower and upper limits of  $\hat{\gamma}_{yf}$ , as shown in Fig. 6c, are to be adjusted such that the Cauchy-Schwarz inequality condition  $\hat{\gamma}_{yf} \leq \sqrt{\hat{\gamma}_{yy}\hat{\gamma}_{ff}}$  is satisfied for positive definiteness. Again, these lower and upper limits are spherical models with the model parameters listed in Table 2.

The cokriged hydraulic conductivity fields using the lower and upper limits of the semivariograms in Fig. 7 show that the incorporation of the electrical log data introduces more heterogeneity which cannot be observed in the kriged hydraulic conductivity fields. However, because the cokriged field has less conditional variance, the simulated field using kriging has higher variability in hydraulic conductivity than that using cosimulation as shown in Fig. 8. The overall structure of the cokriging and cosimulation are similar, but the cosimulation has a higher variance. Moreover, the lower limit semivariogram introduces longer correlation length compared with that of the mean semivariogram. The upper limit semivariogram has higher kriging variance than that of the lower limit semivariogram.



**Fig. 8** Realizations using cosimulation (a–c) and simulation (d–f) with the mean semivariogram model, upper limit semivariogram model and lower limit semivariogram model. The  $K$  varies between  $K=5.18$  m/day (black color) to  $K=701$  m/day (white color)



**Significance analysis on groundwater responses**

With the obtained semivariogram models and their associated semivariogram variances, this section investigates the significance of hydraulic conductivity heterogeneity to the groundwater system responses for different hydraulic conductivity distributions using the kriging method, cokriging method, simulation, and cosimulation with the mean, lower limit and upper limit semivariograms. In this section, a groundwater flow model in the study area is developed to conduct ANOVA on groundwater responses that include hydraulic head variability and mean groundwater level.

*Development of groundwater flow model*

A regional-scale model of the Chicot Aquifer system in southwest Louisiana was developed to better understand the groundwater flow system in the region (Hanson et al. 2001). The Chicot groundwater model underlying Acadia Parish is based on MODFLOW (Harbaugh and McDonald 1996) and has five layers and 50 rows and 50 columns, resulting in a grid size of approximately 0.83 km<sup>2</sup>. The grid is oriented 20° counter-clockwise from the geographic coordinate in order to closely align with the flow direction (Fig. 2). A local model in Acadia Parish has been developed to better understand the groundwater flow dynamics in the study area (Rahman 2005). A telescopic mesh refinement technique MODTMR (Leake and Claar 1999) was used to extract the modeling heads from the regional model (Chicot model) and assign the heads to the hydraulic head boundary conditions of the local model. Therefore, time-varied head boundary conditions are specified for all sides of the study domain. The top layer has a constant head-boundary condition; and the bottom layer is the no-flow boundary. The initial condition of the hydraulic head is created from the 1961 water level estimates. Pumping well types and locations were obtained from the Water Well GIS database of Louisiana Department of Transportation and Development (DOTD). A total of 411 water wells (26 industrial, 27 public supply and 358 irrigation wells) are included in the model. Only industrial, irrigation and public supply water wells are used in this study as these three types are responsible for more than 90% of the groundwater withdrawn from the Chicot Aquifer (Sargent 2002). Yearly pumping rates were estimated by linear interpolation from the 5-year water-use reports of US Geological Survey (Sargent 2002). The total rate was evenly distributed to the registered wells by dividing yearly pumping rate by total number of registered wells within each sector. The groundwater flow was simulated from the year of 1961 to the year of 2001 and the results at the final year are used for analyzing the head responses.

*Groundwater responses*

Three responses are derived from the groundwater flow model and investigated for the ANOVA test. The

**Table 3** Groundwater flow responses using different hydraulic conductivity distributions generated by the geostatistical methods and conditional simulation with semivariogram models

Semivariogram models	Processes	Variables	Water level variability ( $\times 10^5 \text{ m}^2$ ) ( $R_1$ )		Specific capacity ( $\text{m}^2/\text{day}$ ) ( $R_2$ )				Average water level elevation (m) ( $R_3$ )				
			Lower 10% CI (t-test) Mean	Upper 10% CI (t-test) Mean	Mean	Upper 10% CI (t-test) Mean	Var	Min	Max	Mean	Var	Min	Max
Upper limit	Cosimulation	Coregionalized model	5.76	28.9	27.8	28.9	7.4	22	33	-28.0	0.7	-32	-25
Mean			4.53	27.7	27.2	27.7	1.4	25	29	-29.0	0.4	-32	-28
Lower limit			4.16	28.9	28.6	28.9	0.6	27	31	-27.7	1.1	-31	-25
Upper limit	Cokriging		3.94		28.7					-27.7			
Mean			3.91		28.8					-27.4			
Lower limit			3.97		28.7					-27.7			
Upper limit	Simulation	Regionalized model	4.67	28.3	27.2	28.3	7.4	19	31	-29.3	2.8	-40	-26
Mean			4.65	27.7	26.9	27.7	2.9	23	30	-29.3	0.9	-34	-27
Lower limit			4.20	28.3	28.0	28.3	0.8	27	31	-28.0	1.4	-29	-27
Upper limit	Kriging		4.10		28.3					-28.0			
Mean			4.12		28.2					-28.0			
Lower limit			4.11		28.3					-28.0			

CI confidence interval; Var variance

responses are (1) hydraulic head variability ( $R_1$ ), (2) specific capacity of a pumping well ( $R_2$ ), and (3) mean groundwater level ( $R_3$ ). The hydraulic head variability (first response,  $R_1$ ) is taken as the mean squares error of the difference between the initial hydraulic head distribution in 1961 and the calculated hydraulic head distribution in the year 2000 obtained using different hydraulic conductivity distributions over 2,500 computational nodes at the Upper Chicot Aquifer:

$$R_1 = \frac{1}{n_r} \sum_{r=1}^{n_r} \sum_{i=1}^{2500} (\phi_{ini,i} - \phi_{r,i})^2 \quad (24)$$

where  $\phi_{ini,i}$  is the initial groundwater head;  $\phi_{r,i}$  is the simulated groundwater head for a particular realization  $r$ ;  $n_r$  is the number of realizations.

The number of realizations needed depends on the uncertainty of the  $K$  field being addressed (Deutsch and Journel 1998). In this study, the number of realizations ( $n_r$ ) is selected based on a simple test. Model responses ( $R_1$ ) for different numbers of hydraulic conductivity realizations are calculated while keeping everything else as constant. Response for 20 realizations is taken as the finest case. The error terms,  $\varepsilon_R = |R_1 - R_{finest}|$ , are plotted against the square root of number of realizations,  $\sqrt{N_{realizations}}$ . Increasing the number of hydraulic conductivity realizations from 8 to 20 (finest case) only accounts for 6% of response uncertainty. Therefore, an average over 20 realizations was used in stochastic cases. When the kriged field and cokriged field are considered, the  $n_r$  is equal to 1.

The specific capacity is also a key factor to assess the groundwater response. A decrease in the specific capacity indicates a decline in the productivity of the well due to the lower effective hydraulic conductivity field in the near-well area. A decrease in the specific capacity decreases the ability of the well to produce water economically. The specific capacity assesses influence of the hydraulic conductivity field on the ability of a well to produce water at a prescribed flow rate. In this study, the second response was chosen to be the specific capacity of an irrigation well (second response,  $R_2$ ) located at row 31 and column 25—561235 E (easting), 3347383 N (northing) in the UTM coordinate system—with the real pumpage  $Q_{real}=757$  m<sup>3</sup>/day (0.2 million gallons per day). The irrigation well located near the center of the study area is chosen as the boundary conditions have least

**Table 4** The  $F$ -test statistic to compare simulation and cosimulation in variance of groundwater level

Semivariogram model	$\sigma_s^2$	$\sigma_{cs}^2$	$F$	Pr> $F$
Upper limit	9.1	2.2	1.33	0.0001 <sup>a</sup>
Mean	2.9	1.2	1.14	0.02 <sup>a</sup>
Lower limit	4.6	3.6	1.17	0.01 <sup>a</sup>

<sup>a</sup> Indicates significant effects

$\sigma_s^2$  = variance of third responses using simulation process

$\sigma_{cs}^2$  = variance of third responses using cosimulation process

**Table 5** The  $t$ -test statistic to test specific capacity ( $R_2$ ) of conditional simulation to the kriging methods

Variable	Mean of 20 realizations	Conditional estimate	$t$ -test	Pr> $t$
Coregionalized	27.22	28.80	-5.9	0.0001 <sup>a</sup>
Regionalized	26.94	28.24	-3.0	0.008 <sup>a</sup>

<sup>a</sup> Significant effects

influence on it. Simulated drawdown ( $s$ ) at the end of last stress period is used to evaluate the specific capacity:

$$R_2 = \frac{1}{n_r} \sum_{r=1}^{n_r} \frac{Q_{real}}{s_r} \quad (25)$$

where  $s_r$  is the simulated drawdown for a particular realization. An average over 20 realizations was considered for the simulation process.

This study considers the average groundwater head from a 30×30-inner grid (from row 10 to row 40 and from column 10 to column 40) at the last stress period as the third response,  $R_3$ :

$$R_3 = \frac{1}{900n_r} \sum_{r=1}^{n_r} \sum_{i=1}^{900} \phi_{r,i} \quad (26)$$

The three responses are computed from the groundwater model using the alternative hydraulic conductivity fields (Table 3).

#### Statistical analysis using ANOVA

All three responses are found insensitive to semivariogram uncertainty for the kriged and cokriged models. For the simulation and cosimulation results, increased variance in hydraulic conductivity fields (i.e., upper limit semivariogram) causes more variation in water-head variability and specific capacity. In all levels of semivariogram uncertainty, the coregionalized method has higher water levels than those in the regionalized method.

Results from the cosimulated scenario (Table 3) show that there is a 10% chance that specific capacity is less than the cokriging estimate by at least 11%, and specific capacity from the simulated scenario is less than the

**Table 6** ANOVA results on flow model responses

Class	Levels	Water level variability Pr> $F$	Specific capacity Pr> $F$	Average water level Pr> $F$
Process	2	0.03 <sup>a</sup>	0.01 <sup>a</sup>	0.04 <sup>a</sup>
Variable	2	0.77	0.04 <sup>a</sup>	0.01 <sup>a</sup>
Semivariogram	3	0.27	0.13	0.19

<sup>a</sup> Indicates significant effects

kriging estimate by at least 8%. However, this difference may be too small to be of practical importance.

Table 3 shows that the variance of the average water-level response using simulation is greater than the variance using cosimulation. To verify the significant difference in these two variances, the *F*-tests method was introduced to conduct the significance analysis at all three levels of semivariogram uncertainty. The *F*-test results shown in Table 4 conclude that the variances are different at the 10% significance level. This is because the cosimulation reduces uncertainty in aquifer heterogeneity and aquifer response by introducing the secondary variable in the simulation process.

The *t*-tests in Table 5 show that the mean specific capacity using simulation and cosimulation is significantly different from conditional estimates using kriging and cokriging methods at 10% level of significance. Table 6 shows the ANOVA significance analysis results on the flow model responses in terms of the processes (i.e., kriging versus simulation; cokriging versus cosimulation), variables (i.e., kriging versus cokriging; simulation versus co-simulation); and semivariograms (i.e., lower limit, mean, and upper limit). At 10% level of significance, the ANOVA results show that semivariogram uncertainty does not have a significant effect on any of the flow model responses.

However, the results in the ‘process’ class (Table 6) show that the use of simulation and cosimulation methods has a more significant impact on all of the flow model responses than the use of kriging and cokriging methods. The results in the ‘variable’ class (Table 6) also show that the use of the resistivity data does have a significant effect on the flow model responses.

## Conclusions

Geophysical data such as borehole electrical resistivity have been integrated with the measured hydraulic conductivity under the cokriging framework to improve spatially distributed hydraulic conductivity estimation. A pseudo cross-semivariogram has been evaluated to cope with the non-collocated resistivity data. The improvements in the groundwater model have been investigated by comparing the head levels from different hydraulic conductivity models and the significance of either (1) processes; (2) variables; or (3) semivariograms as determined by ANOVA. The results show that the use of the coregionalized field is statistically significant in the flow-model response compared to the regionalized model. The study also shows that the simulation process can reproduce the aquifer heterogeneity and is statistically significant in the flow-model response; therefore, cosimulation should be used instead of cokriging. The semivariogram uncertainty has been studied in all models, both regionalized and coregionalized, by using the upper and lower-limit semivariograms, assuming that the sills are log-normally distributed. However, results show that the semivariogram uncertainty on the groundwater-flow model response is not significant.

**Acknowledgements** The research was partially supported by Louisiana Department of Transportation and Development; Louisiana Geological Survey; and Louisiana State University Faculty Research Grant Program. Special thanks go to C.V. Deutsch, Dept. of Civil and Environmental Engineering, University of Alberta, Canada for his invaluable discussion. We also acknowledge the help from R. Milner of the Louisiana Geological Survey for resistivity-log data interpretation.

## References

- Archie GE (1942) The electrical resistivity logs as an aid in determining some reservoir characteristics, *AIME Trans* 146:54–62
- Asquith GB, Gibson CR (1982) Basic well log analysis for geologists. American Association of Petroleum Geologists Methods in Exploration Series 3, The American Association of Petroleum Geologists, Tulsa, OK, USA, 216 pp
- Boman GK, Molz FJ, Guven O (1995) An evaluation of interpolation methodologies for generating three dimensional hydraulic property distribution from measured data. *Ground Water* 33(2):247–258
- Bradbury KR, Rothschild ER (1985) A computerized technique for estimating the hydraulic conductivity of aquifers from specific capacity data. *Ground Water* 23(2):240–246
- Carlson D, Milner R, Hanson B (2003) Evaluation of the aquifer capacity to sustain short–long term groundwater withdrawal from point sources in the Chicot Aquifer for southwest Louisiana, part 1. Report of Investigation Series 03–01, 90, Louisiana Geological Survey, Baton Rouge, LA
- Carman PC (1956) Flow of gases through porous media. Butterworths, London
- Carrera J, Neuman SP (1986) Estimation of aquifer parameters under transient and steady-state conditions. 1. Maximum-likelihood method incorporating prior information. *Water Resour Res* 22:199–210
- Clark I, Basinger K, Harper W (1989) MUCK: a novel approach to co-kriging. In: Buxton BE (ed) Proceedings of the Conference on Geostatistical, Sensitivity and Uncertainty: methods for groundwater flow and radionuclide transport modeling. Battelle Press, Columbus, OH, USA, pp 473–494
- Cressie N (1985) Fitting variogram models by weighted least squares. *J Int Assoc Math Geol* 17:563–586
- Deutsch CV, Journel AG (1998) GSLIB: geostatistical software library and users guide, 2nd edn. Oxford University Press, New York, 369 pp
- Domenico PA, Schwartz FW (1990) Physical and chemical hydrogeology. Wiley, New York
- Donnelly-Makowecia LM, Moore RD (1999) Hierarchical testing of three rainfall-runoff models in small forested catchments. *J Hydrol* 219:136–152
- Eggleston JR, Rojstaczer SA, Pierce JJ (1996) Identification of hydraulic conductivity structure in sand and gravel aquifers: Cape Cod data set. *Water Resour Res* 32(5):1209–1222
- Feyen L, Ribeiro PJ, De Smedt F, Diggle PJ et al (2003) Stochastic delineation of capture zones: classical versus Bayesian approach. *J Hydrol* 281(4):313–324
- Fisher RA (1935) The design of experiments. Oliver and Boyd, Edinburgh, UK
- Fogg GE (1986) Groundwater-flow and sand body interconnectedness in a thick, multiple-aquifer system. *Water Resour Res* 22(5):679–694
- Gloaguen E, Chouteau M, Marcotte D, Chapuis R (2001) Estimation of hydraulic conductivity of an unconfined aquifer using cokriging of GPR and hydrostratigraphic data. *J Appl Geophys* 47(2):135–152
- Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York, 483 pp
- Hanson B, Milner R, Willson C, Rahman A, Paulsell R (2001) Evaluation of aquifer capacity to sustain long term ground water

- withdrawal from point sources: a pilot study. Report to the Louisiana Department of Natural Resources, Baton Rouge, 171 pp
- Harbaugh AW, McDonald MG (1996) User's documentation for MODFLOW-96, an update to the U.S. geological survey modular finite-difference ground-water flow model. US Geol Surv Open-File Rep 96-485, 56 pp
- Hohn ME (1998) Geostatistics and petroleum geology, 2nd edn. Kluwer, Dordrecht, The Netherlands, 248 pp
- Ikem A, Osibanjo O, Sridhar MKC, Sobande A (2002) Evaluation of groundwater quality characteristics near two waste sites in Ibadan and Lagos, Nigeria. *Water Air Soil Poll* 140:307-333
- Kelly WE (1977) Geoelectric sounding estimating aquifer hydraulic conductivity. *Ground Water* 15(6):420-425
- Kitanidis PK (1986) Parameter uncertainty in estimation of spatial functions: Bayesian-analysis. *Water Resour Res* 22(4):499-507
- Kitanidis PK, Vomvoris EG (1983) A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. *Water Resour Res* 19(3):677-690
- Koltermann C, Gorelick SM (1996) Creating images of heterogeneity in sedimentary deposits: a review of descriptive, structure-imitating, and process-imitating approaches. *Water Resour Res* 32(9):2617-2658
- Kosinski WK, Kelly WE (1981) Geoelectric soundings for predicting aquifer properties. *Ground Water* 19(2):163-171
- Leake SA, Claar DV (1999) Procedures and computer programs for telescopic mesh refinement using MODFLOW. US Geol Surv Open-File Rep 99-238, 53 pp
- McLaughlin D, Townley LR (1996) A reassessment of the ground-water inverse problem. *Water Resour Res* 32(5):1131-1162
- Myers DE (1991) Pseudo-cross variograms, positive definiteness and cokriging. *Math Geol* 23:805-816
- Neuman SP (1990) Universal scaling of hydraulic conductivities and dispersivities in geologic media. *Water Resour Res* 26(8):1749-1758
- Ortiz CJ, Deutsch CV (2002) Calculation of uncertainty in the variogram. *Math Geol* 34(2):169-183
- Poeter EP, Gaylord DR (1990) Influence of aquifer heterogeneity on contaminant transport at the Hanford site. *Ground Water* 28(6):900-909
- Rahman A (2005) Improvements in groundwater flow modeling through the integration of resistivity logs and hydraulic conductivity and the use of variogram uncertainty, PhD Thesis, Louisiana State University, USA, 104 pp
- Rehfeldt KT, Boggs JM, Gelhar LW (1992) Field-study of dispersion in a heterogeneous aquifer. 3. Geostatistical analysis of hydraulic conductivity. *Water Resour Res* 28(12):3309-3324
- Ritzi RW, Jayne DF, Zahradek AJ, Field AA, Fogg GE (1994) Geostatistical modeling of the heterogeneity in glaciofluvial buried-valley aquifers. *Ground Water* 32(4):666-674
- Rong Y (2002) Groundwater data analysis for MTBE relative to other oxygenates at gasoline-impacted sites 9(4):184-190
- Sargent P (2002) Water use in Louisiana, 2000, Special Report no. 15, Department of Transportation and Development Water Resources, Baton Rouge, LA, USA
- Vukovic M, Soro A (1992) Determination of hydraulic conductivity of porous media from grain-size composition. *Water Resources Publications, Highlands Ranch, CO, USA*, pp 1-83
- Weight WD, Sonderegger JL (2001) Manual of applied field hydrogeology, McGraw-Hill, New York, 608 pp