Heterogeneity definition of a deltaic aquifer system using multiple-point statistics

Martin Blouin
INRS-ETE, Québec, Qc, Canada
Richard Martel & Erwan Gloaguen
INRS-ETE, Québec, Qc, Canada

ABSTRACT
This paper is about using multiple-point geostatistic simulations to define the heterogeneity of a deltaic aquifer. A 3-D geological model is built for further groundwater flow investigation. Descriptive information on about 120 boreholes logs, GPR surveys and surface geology interpretations are used to build the model. Multiple-point statistics simulations require a training image as an input. Hence, a conceptual geological model is created. More than 50 simulated models are built and results demonstrate well the deltaic structure of the studied system.

RÉSUMÉ
Cet article traite de l'utilisation de simulations géostatistiques multipoints pour modéliser un système deltaïque hétérogène. Ce modèle géologique est réalisé dans le but d'obtenir une base pour un modèle d'écoulement souterrain. Des données provenant de description de près de 120 forages et de levées géoradar ont été utilisées ainsi que l'information provenant d'interprétations de surface. Les statistiques multipoints nécessitant une image d'entraînement, un modèle conceptuel de la géologie de sous-surface a été produit. Un peu plus de 50 modèles simulés ont été réalisées et les résultats mettent bien en évidence la structure deltaïque de l'environnement à l'étude.

1 INTRODUCTION

In hydrogeology, characterization of Quaternary deposits heterogeneity is important in order to define the groundwater flow and the transport of contaminant. The spatial structure of detritic aquifers is mainly controlled by depositional and diagenetic processes. In such systems, the hydraulic properties can vary within a few meters by orders of magnitude (Koltermann et Gorelick 1996). Paleo-deltaic environments show significant variations in stratigraphic facies continuity. Hence, groundwater flow and mass transport in such systems are controlled by the distribution and the connectivity of high hydraulic conductivity paths (Strebelle 2002).

In order to obtain an efficient groundwater flow system, building a stratigraphic model is necessary to reproduce grain-size heterogeneities and hydraulic preferential paths (Falivene et al. 2006, Cabello et al. 2007). Descriptive drilling information on the study site gives the knowledge on the nature of the sediments and the stratigraphic sequence. However, in a paleo-deltaic system, the lateral variation of geological units cannot be assessed adequately by connecting punctual hard data points. The approach proposed in this paper uses geostatistical simulations, a method that is increasingly used in heterogeneous systems modeling, to generate multiple statistically equivalent high resolution geological models. Such simulations allow time-efficient building of 3-D geological models, which are relevant of the facies distribution and fit all the measured data (Falivene et al. 2006).

1.1 Study site

The study site is located on Defence Research and Development Canada (DRDC-Valcartier) tests range about 30 km north-west of Quebec City in Valcartier, Qc. Since the end of World War II, the Canadian Forces have been using the area of the Valcartier Base for military research and development purposes. Figure 1 shows the location of the Canadian Force Base (CFB- Valcartier) while figure 2 presents an aerial view of the tests range. The site is bounded to the east by the Jacques-Cartier River at an elevation of about 170 masl (meters above sea level) and to the north-west by outcropping rock at more than 300 masl. The study area covers 6 km^2 and its geology consists of quaternary deposits overlaying Canadian Shield gneiss. The system is made of glacial and deltaic sediments put in place by the Champlain Sea.

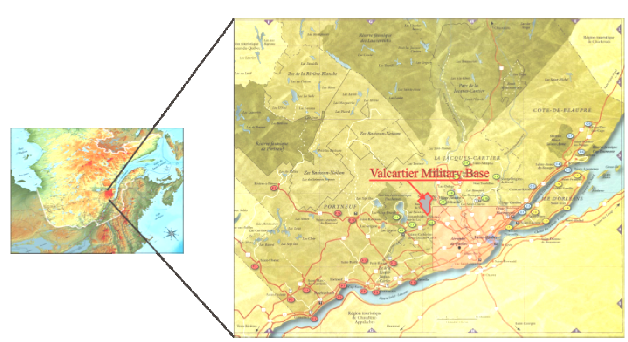

Figure 1. Location of CFB-Valcartier in the province of Quebec

A surface geology interpretation (Michaud et al., 1999) at the studied site revealed the presence of 5 major geological units at the ground surface:

- R: Bedrock = Metamorphic igneous rocks from the Canadian Shield.
- Tv: Unconsolidated till = Unsorted glacial sediment. Heterogeneous mix of clay, silt, sand, gravel and boulders.
- Gx: Glaciofluvial sediments = Sand, gravel and boulders. Possible presence of till and/or diamictite.
- Md: Deltaic sediments = Sand, gravely sand and gravel, stratified and sorted.
- At: Low terrace alluvium; Sand, silty sand, gravely sand and gravel stratified, low presence of organic material.

And one buried:

- Mi: Prodeltaïc sediments = Silty sand, silt, silty clay, stratified.

Figure 2. Aerial view of the studied area (Tests range at DRDC-Valcartier)

1.2 Study objectives

Conventional geological numerical modeling consists in interpolating by hand the different units intercepted at sampled wells. This approach is time consuming and, also, very user dependent. Also, because of these reasons, usually, only one smooth model is generated. Smoothing of the models implies that the heterogeneity of the ground is not reproduced leading to local over or under estimation of mass transport depending. The global effect of such smoothing cannot be estimated. Hence, the main objective of the study is to build multiple statistically equivalent 3-D geological models that will help characterizing the aquifer heterogeneity and evaluate the uncertainty on the groundwater flow and mass transport modeling.

2 THE MULTIPLE-POINT APPROACH

2.1 Motivations

Irregularly shaped heterogeneities that constitute paleo-deltaic systems yields to incapacity in using a classic horizontal layers geological model. Also, preferential flow paths may not be reproduced by conventional two-points geostatistical methods (Guardiano and Srivastava, 1993). When flow is governed by high conductivity channels, the use of multiple-point statistics has found to be efficient (Strebelle 2002).

2.2 MPG general principle

Multiple point geostatictics (MPG) is a pixel-based stochastic modeling method used to generate categorical realizations (Guardiano and Srivastava 1993; Strebelle 2002). The process used by the simulator is based on the sequential simulation method. It randomly visits every empty pixel of the model in order to assign a value based on the proximity and values of the neighboring data. Contrary to most simulation methods which use kriging and require the production of a variogram, MPG necessitates a conceptual model (training image). The probability density function is directly obtained from the conceptual model. The simulator then reproduces in the simulated models the structures identified in the training image as it automatically fit all the measured data (in our case by the drilling logs facies information). This geostatistical method is considered to be a structure-imitating process (Koltermann and Gorelick 1996).

2.3 The snesim algorithm

The *snesim* algorithm or single normal equation simulation (Strebelle 2002) is frequently used in geology for heterogeneous systems modeling and simulation. This algorithm is specially designed for categorical data, but can be used with continuous data grouped in classes. At our studied site, it is not necessary to proceed to a data discretization since both hard data and conceptual model are evaluated in terms of stratigraphic facies. Figure 3 shows a scheme of principle concept of MPG and the relation between hard data and training image.

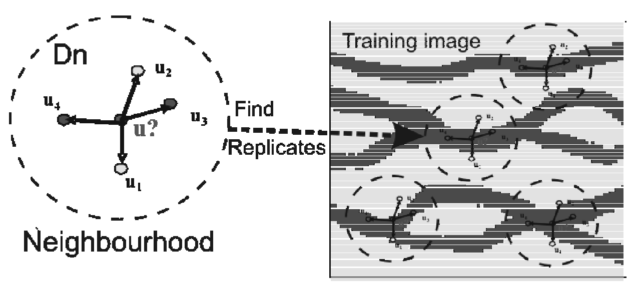

Figure 3. General principle of MPG adapted from Strebelle, 2002

As shown above, the algorithm adapts simulated values to hard data (left) to replicate structures in the training image (right).

2.3.1 Snesim theory

This section presents the probability functions relevant for one simulated data as developed by Guardiano & Srivastava (1993) and Strebelle (2002).

Let the local probability distribution of a variable S(u) conditioned by *n* hard data (neighbourhood). In MPG, the *n* data are combined in one single event and the conditional probability distribution is calculated as *n* +1 points.

Let A_k, a random binary variable associated with the occurrence of state s_k happening at point *u*:

$$A^k = \begin{cases} 1 & if\ S(u) = s_k \\ 0 & else \end{cases} \qquad (1)$$

The same way, random variable D associated with the event probability d_n from *n* conditioning data:

$$S(u_\alpha) = s_\alpha, \alpha = 1, \ldots, n \qquad (2)$$

$$D = \begin{cases} 1 & if\ S(u_\alpha) = s_\alpha, \forall\, \alpha \in [1, n] \\ 0 & else \end{cases} \qquad (3)$$

If the necessary statistics are available for A_k and the corresponding D, the conditional probability is given by the kriging equation:

$$Prob\{A_k = 1 \mid D = 1\} = E\{A_k\} + \lambda[1 - E\{D\}] \qquad (4)$$

Where D=1: observed event (source data)

E{D} = Prob{D=1} : probability for conditioning data event.

$E\{A_k\}=Prob\{S(u)=s_k\}$: probability for unknown state at *u*.

Hence, by replacing in the kriging equation:

$$Prob\{A_k = 1 \mid D = 1\} = \frac{Prob\{A_k = 1, D = 1\}}{Prob\{D = 1\}} \qquad (5)$$

The equation represents the exact solution of Bayes' theorem definition for **conditional probability**. In fact, Bayes' theorem describes the way the observations of A_k are conditioned by the observation D.

So, the algorithm visits every cell to be simulated only once and assigns a value according to the neighbourhood and the structure identified in the training image. Once the simulated value is assigned to a cell it becomes a hard data conditioning the rest of the model not yet visited.

The definition of the neighborhood considered by the algorithm is crucial because it regulates the precision of the model as well as calculation time. Indeed, in MPG the more data are used in the neighborhood, the more realistic are the simulated values. On the other hand, for a neighborhood of n data which can take K categorical values, the simulator has to evaluate K^n possibilities, so there is a compromise to be made for calculation time.

No explicit constraint assures the conditioning data histogram reproduction. For some simulations, it is possible to end up with a simulated values histogram quite different from the distribution of source data. This difference can be explained by the distribution of training image data (Strebelle, 2002).

The conditioning steps and the *snesim* algorithm process are presented as a scheme in figure 4.

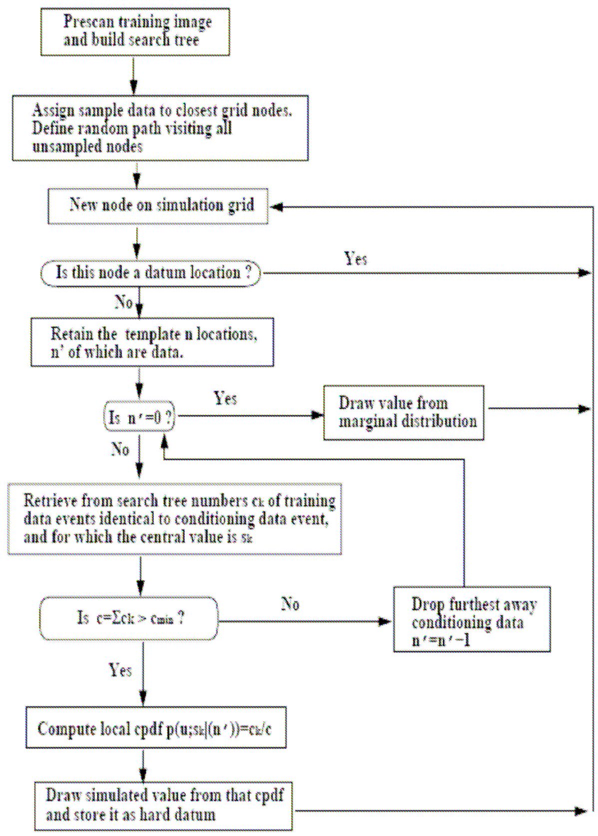

Figure 4. Decision tree of the *snesim* process adapted from Strebelle, 2002

3 METHODOLOGY

This section describes the processing of the raw data leading to the production of the conceptual model used as a training image in multiple-point simulations. The full process of structuring and integrating field data to the MPG simulation software is presented. The simulated models generated using MPG are also discussed.

3.1 Borehole data

Since 1995, more than 180 boreholes have been drilled by consulting firms and INRS-ETE team in the studied area. Boreholes logs have been assigned in the first place by an alphanumerical CGQ code developed by *Centre Géoscientifique de Québec* (Parent et al. 2003). This classification is based mainly on sediment origin and on its grain size. Hence, this CGQ code is highly correlated to geological materials hydraulic characteristics.

3.2 GPR surveys

Twenty GPR profiles were made on the study site to increase the resolution of the stratigraphic contacts and piezometry surface,. According to boreholes data, fine sediments horizons are located at a depth of 20 to 30 meters. The system used is a pulse Ekko 100 with 50 MHz antennas and a 1000V transmitter. Figure 5 shows the location of GPR profiles on CFB-Valcartier tests site while figure 6 presents profile #2 interpretations. The interpretation of the GPR profiles was made according to boreholes logs information and from piezometric surveys.

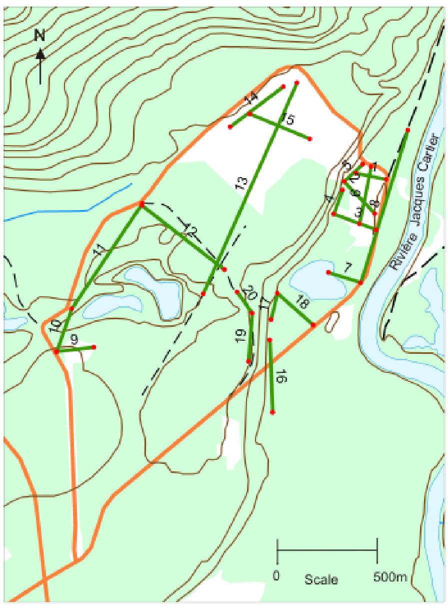

Figure 5. Location of GPR surveys on the study site

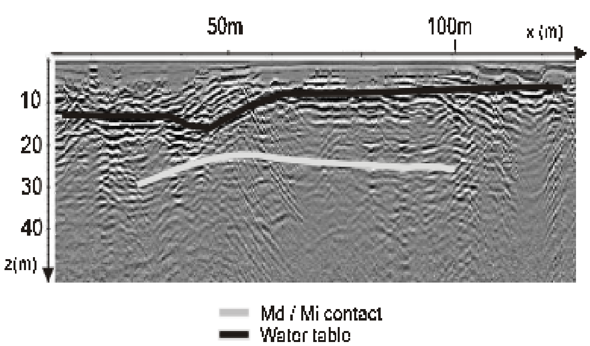

Figure 6. Interpretation of GPR profile #2

3.3 Construction of the 3-D conceptual geological model

The 3-D conceptual geological model was built using the Gocad® software. This Paradigm® society product allows building and visualizing of stratigraphic units. Three-dimensional data points containing alphanumeric codes as geological unit property were gathered in the modeling software. This type of geological information is the main element to guide the conceptual model construction. It also helps delineating stratigraphic units and making the interpretation of geophysical data such as GPR survey.

A preliminary version of the sub-surface geology in the study area was made according to the information gathered in the Gocad® and the surface geology interpretation by Michaud et al. 1999. Figure 7A shows a 3-D view of the conceptual model while figure 7B presents 2D cross-sections oriented to illustrate geological units distribution. In order to use the conceptual model as a training image for the multiple point simulations, each layer of the model was assigned a facies number:

Facies 1 = Bedrock (R)
Facies 2 = Unconsolidated till (Tv)
Facies 3 = Glaciofluvial sediments (Gx)
Facies 4 = Fine sediments (Mi + Ma)
Facies 5 = Deltaic sediments (Md)
Facies 6 = Low terrace alluvium (At)

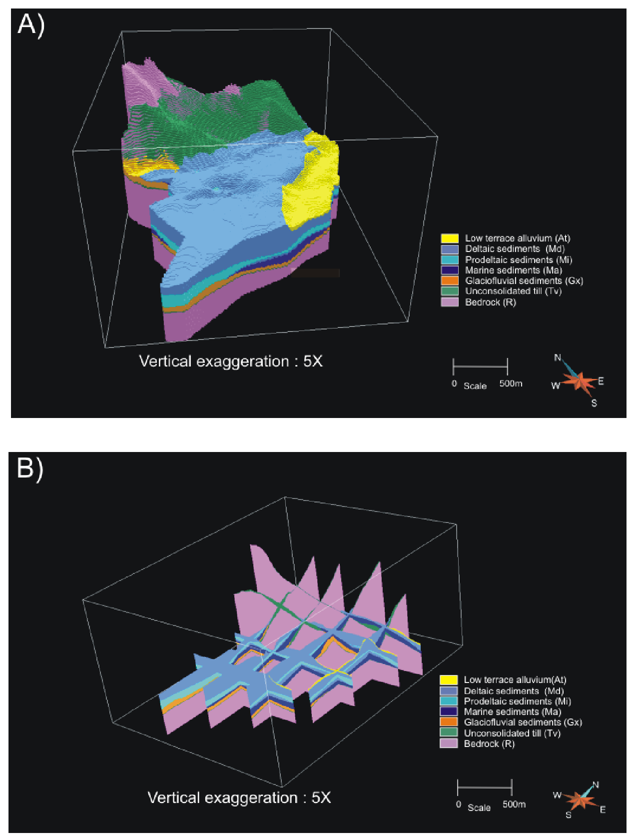

Figure 7. Conceptual geological model

3.4 Structuring hard data

Because conditional multiple-point simulations require hard data (conditioning) as categorical values, a second interpretation of boreholes data had to be made. This new classification of drilling descriptive information was made following the alphanumeric CGQ code previously assigned and the elevation of the data. Each interpretation is considered as a 3-D punctual value. A facies number analog to the conceptual model's geological unit is given. Figure 8 presents an example of the structuring and conversion of boreholes data.

3.5 Isatis® software from Geovariances®

Once the data are structured and the variables converted, it is ready to be integrated in the geostatistics simulator. The conceptual model 3-D grid is used as an output for simulated values of stratigraphic facies. The training image, boreholes and surface geology interpretations are given as an input. The parameter *n* points representing the neighborhood considered by the algorithm to generate simulated values has to be set. Finally, the random number seed and the minimum number of matching structures for statistics need to be determined. The optimal set of parameters was obtained by analysis of the corresponding simulated fields based on geological a priori.

Well_ID	From m	To m	Description	CGQ code
Biogenie_juillet2005_E2_15	0	1.8	Sable graveleux, présence de cailloux.	S,G
Biogenie_juillet2005_E2_15	1.8	7	Sable moyen à fin, traces de silt.	F2
Biogenie_juillet2005_E2_15	7	14.9	Sable moyen à fin silteux, traces d'argile.	S,F1
Biogenie_juillet2005_E2_15	14.9	19.4	Silt sableux, traces d'argile.	F2
Biogenie_juillet2005_E2_15	19.4	27.5	Silt sableux, un peu d'argile.	F2
Biogenie_juillet2005_E2_15	27.5	30.8	Sable moyen à fin.	S
Biogenie_juillet2005_E2_15	30.8	33.6	Sable moyen à fin, traces de silt, un peu de cailloux.	D2

Easting (m)	Northing (m)	Elevation (masl)	Facies #
308936.00	5200349.00	161.10	6
308936.00	5200349.00	157.60	5
308936.00	5200349.00	151.05	5
308936.00	5200349.00	144.85	4
308936.00	5200349.00	138.55	4
308936.00	5200349.00	132.85	5
308936.00	5200349.00	129.80	3

Figure 8. Conversion of boreholes description in punctual 3-D data

4 RESULTS

The first step towards getting a realistic model in the simulation process is the determination of the best-suited parameters.

4.1 Simulation parameters

The bedrock and till units have limited samples (hard data) and must cover an important volume according to the conceptual model. Because of the antagonism between soft and hard data, simulating those two units may lead to undesired and geologically unrealistic models. Hence, the multiple-point simulations are made only on the sedimentary part of the system which represents 4 facies: glaciofluvial sediments, fine sediments (Prodeltaic + marine), deltaic sediments and low terrace alluvium.

According to simulation tests, a neighborhood of 10 values is chosen in regards to data quality and computing time. Also, the minimum number of matching patterns is set to 10. Surface geology interpretations are used as points to be included with conditioning data.

The training image is composed of 50 X 50 X 126 cells in a 3-D grid. It is design in a way that elevation cells have a 1 meter increment. This 3-D grid is used as an input for the simulated values. The simulator creates a new variable and the results of the simulations are assigned in the same grid. Figure 9 shows the conceptual geological model used as a training image and reduced to a limited number of 4 stratigraphic facies. The model shown is an adapted version of figure 7A.

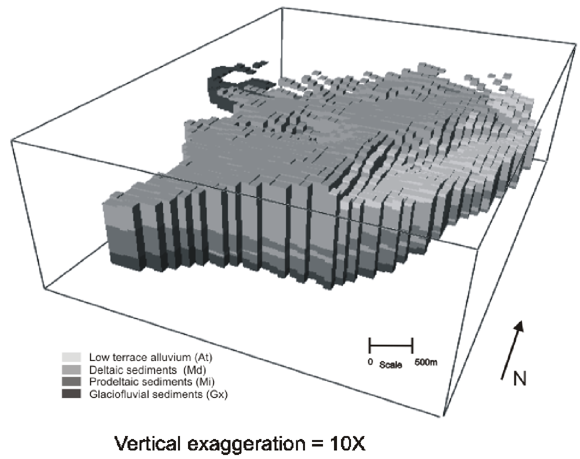

Figure 9. Training image used for multiple-point simulations

4.2 Multiple-point simulations

With the training image illustrated on figure 9, more than fifty simulated models are built, each one taking about 40 minutes of computing time. Figure 10 shows a 3-D view of one of the simulated models.

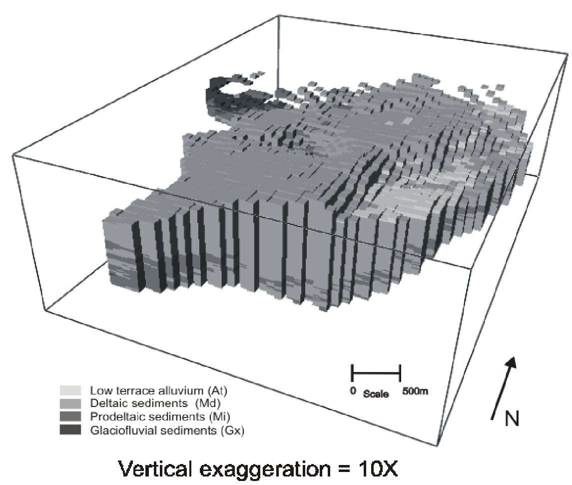

Figure 10. Deltaic system model from multiple-point simulation (one simulated model)

The model shown on figure 10 respects the interpretation of surface geology and the sequence of sediments suggested by the geological conceptual model. It shows the advantage of using MPG for hydrogeological purposes. The principal objective of the study is to reproduce the distribution of fine sediments in deltaic environment and the simulated model gives an interesting possible scenario.

Contrary to the continuous horizontal units proposed by the conceptual geological model, the simulated model presents discontinuous and highly heterogeneous distribution of prodeltaic sediments as expected in this kind of geological setting.

However, the glaciofluvial sediments unit is underestimated by the model. Quaternary interpretation (Michaud et al., 1999) suggests higher presence of this glacial deposit. Figure 9 shows cross-sections of the simulated model that highlight the heterogeneous distribution in the deltaic system while underestimating the Gx unit.

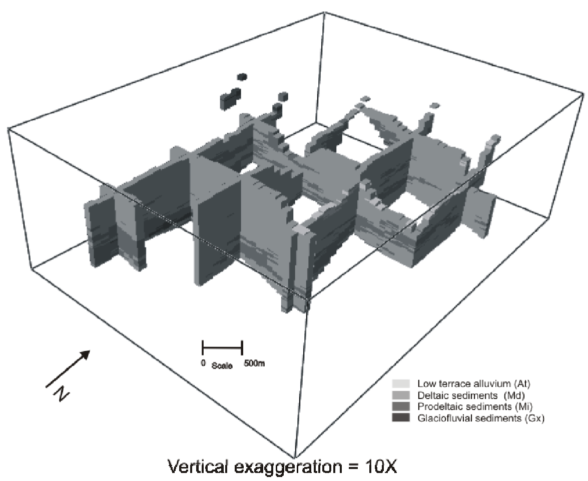

Figure 11. Cross-sections of the simulated model

4.3 Simulated models utility

Such simulated models are very useful when it comes to the integration of hydraulic conductivity data as the distribution of high and low permeability path is confined to specific 3-D areas.

Also, the production of multiple scenarios gives rapidly an inventory of possible solutions that can be expected from the measured data and the geological training image. The validation and the choice of the most appropriate models among all the generated ones will be assessed through the mass transport and flow modeling on every simulated model. The aspect of transport uncertainty caused by geological modeling uncertainty becomes really interesting for further decision makers and risk management. For example, this approach will yield to probability map to overpass a given contamination thresholds.

5 CONCLUSION

The multiple-point geostatistics algorithm is an adequate pixel-based tool to define deltaic systems heterogeneity. Simulated models built for the studied area help generating multiple scenarios of preferential flow paths in this sedimentary environment. The surface geology interpretation is respected by the simulated models as long as conditioning data are created. Out of the four stratigraphic facies simulated, the simulation was less-effective in the reproduction of the glaciofluvial sediments unit.

MPG are time-efficient and don't require the production of indicator variogram (Chilès and Delfiner, 1999). However, a training image relevant of the geology in place must be made. Further work in the studied area will aim ground flow characterization. Simulated models have to be sorted and only the more realistic will be picked for the remainder of the study. From there, integrating hydraulic conductivity to the model will be done to ultimately get a groundwater model.

6 ACKNOWLEDGEMENT

The authors would like to acknowledge the contribution of Lise Lamarche, a PhD student at INRS-ETE and also the help of Marie-Catherine Poulin-Talbot for data and knowledge transfer of the studied area. The support of Ministry of National Defence for research funding and the collaboration of Nathalie Roy from DRDC-Valcartier is also gratefully acknowledged.

7 REFERENCES

Cabello, P., Cuevas J.L., Ramos E., 3D modelling of grain size distribution in Quaternary deltaic deposits (Llobregat Delta, NE Spain). Geologica Acta, Vol.5, Nº 3, 2007, 231-244.

Chilès, J-P. and Delfiner, P., 1999. Geostatistics: modeling spatial uncertainty. Wiley series in probability and statistics, NY, 695 pp.

Falivene, O., Cabrera, Ll., Muñoz, J.A., Arbués, P., Fernández, O., Sáez, A., 2007a. Statistical grid-based facies reconstruction and modelling of sedimentary bodies. Alluvial-palustrine and turbiditic selected examples. Geologica Acta, 5(3), 199-230.

Guardiano, F., Srivastava, R.M., 1993. Multivariate geostatistics: beyond bivariate moments. In: Soares, A. (ed.). Geostatistics-Troia 1. Dordrecht, Kluwer Academic Publications, 113–114.

Koltermann, C.E., Gorelick, S.M., 1996. Heterogeneity in sedimentary deposits: A review of structure-imitating, processimitating, and descriptive approaches. Water Resources Research, 32, 2617-2658.

Michaud Y, Parent M, Mailloux M, Boisvert É, Lefebvre R, Martel R, Boivin R, Roy N, et Hains S, 1999. Cartographie des formations superficielles et cartographie hydrogéologique de la base des forces canadiennes Valcartier. Rapport interne CGQ soumis à l'Unité de Soutien de Secteur Valcartier, 1 CD-Rom, 2 cartes.

Parent, M., Girard, F., Fagnan, N., Michaud, Y., Boisvert, É., Fortier, R., 2003. Caractérisation géologique des formations superficielles enfouies. In: Michaud, Y., Lefebvre, R., McCormack, R. (Eds.), Guide méthodologique pour la cartographie hydrogéologique régionale des aquifères granulaires de surface. Report prepared by the Quebec Geoscience Centre for the Quebec Environment Ministry, 140pp. (in French).

Strebelle, S., 2002 Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics. Mathematical Geology, 34, 1-22.