102315

WATER RESOURCES RESEARCH, VOL. 40, W04603, doi:10.1029/2003WR002876, 2004

Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology

Francesco Ballio and Alberto Guadagnini

Dipartimento di Ingegneria Idraulica, Ambientale, Infrastrutture Viarie, Rilevamento, Politecnico di Milano, Milan, Italy

Received 10 November 2003; accepted 4 February 2004; published 21 April 2004.

[1] Numerical Monte Carlo simulation is considered to be one of the main tools to be used in groundwater hydrology (1) to quantify the uncertainty in the flow predictions due to imperfect knowledge of aquifer architecture, hydraulic parameters, and forcing terms or (2) to assess the reliability of approximated moment-based equations for flow and/or transport. While the Monte Carlo framework is conceptually straightforward and very flexible, it is recognized as lacking well-established convergence criteria. Here we propose a methodology for convergence analysis of Monte Carlo simulations and therefore for the reliability assessment of the inferred statistical moments. The methodology, based on simple rules of statistical inference, is described with reference to a typical groundwater flow problem and can be extended to different application fields. *INDEX TERMS:* 1829 Hydrology: Groundwater hydrology; 1869 Hydrology: Stochastic processes; 2753 Magnetospheric Physics: Numerical modeling; *KEYWORDS:* convergence, groundwater, Monte Carlo simulations, random, heterogeneity, stochastic

Citation: Ballio, F., and A. Guadagnini (2004), Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology, *Water Resour. Res.*, *40*, W04603, doi:10.1029/2003WR002876.

1. Introduction

[2] Numerical Monte Carlo (MC) simulation is considered to be the main tool to be used in groundwater hydrology and in the oil industry to quantify the uncertainty in flow and transport predictions due to uncertain aquifer architecture, hydraulic parameters, and forcing terms. Results from MC simulations, based on a limited number of synthetic realizations of the random parameters of interest, are also used as ground truth against which to assess the reliability of approximated moment-based equations for groundwater flow and/or transport [e.g., *Hassan et al.*, 1998; *Riva et al.*, 2001; *Guadagnini et al.*, 2003; *Li et al.*, 2003]. It is also well recognized that MC-based simulations still lack well-established convergence criteria. Because of the interest and practical use of Monte Carlo simulations, we believe there are strong theoretical and pragmatic reasons to propose a simple and effective method to analyze the convergence of Monte Carlo-based ensemble moments and to identify a criterion to reliably terminate the Monte Carlo process at a desired level of accuracy.

[3] Stability of the computed moments is usually assessed by plotting the sample moments of the target variable evaluated at specific points as a function of the number of Monte Carlo simulations performed. These graphs are usually presented in natural scale and very rarely show more than a few hundreds of realizations [*Johnson*, 1992; *Riva et al.*, 1999; *Li et al.*, 2003]. Convergence of results is then assessed qualitatively on the basis of the fact that the sample moment curve has reached a more or less flat region. *Bellin et al.* [1994] propose a convergence criterion based on the quantity $\delta^2 X_n = (X_n - X_{n-1})^2$, where X_n is a target sample moment at a point, and n is the sample size, i.e., the number of Monte Carlo simulations. The authors consider the process to have reached convergence when the spatial average of $\delta^2 X_n$ over the whole computational domain vanishes. Once the maximum acceptable average value for $\delta^2 X_n$ is defined, this procedure is a quantitative transposition of the visual judgment of the stabilization of the average sample moment curves.

[4] Here we propose a methodology for convergence analysis of Monte Carlo simulations and therefore for the reliability assessment of the computed statistical moments. The methodology will be described with reference to the first two (ensemble) moments (mean and variance) of hydraulic head in a typical steady state groundwater flow problem, but can be extended to other statistical quantities (for example to global indicators) as well as to different fields where Monte Carlo simulations are used, such as river hydraulics, sediment transport, hydraulic structures [e.g., *Johnson*, 1992; *Yen and Tung*, 1993; *van Vuren et al.*, 2002].

2. Methodology

[5] We start by briefly recalling some basic definitions and properties of sample moments of a random variable, $\Re$, which will be then used in the analysis of corresponding sample moments resulting from MC simulations. It is well known [e.g., *Mood et al.*, 1974] that the sample mean, $\overline{\Re}_n = \frac{1}{n}\sum_{i=1,n} \Re_i$, computed on the basis of n realizations $\Re_i$ ($i = 1, \ldots, n$) of the random variable $\Re$ with probability density function $f()$, in general does not coincide with the ensemble mean, $\mu = \langle \Re \rangle$, of the process. The same holds for the sample variance, $S_n^2 = \frac{1}{n-1}\sum_{i=1,n} (\Re_i - \overline{\Re}_n)^2$, and sample kurtosis, $M_{4,n} = \frac{1}{n}\sum_{i=1,n} (\Re_i - \overline{\Re}_n)^4$, which do not coincide with the process variance, $\sigma^2 = \langle (\Re - \mu)^2 \rangle$, and the process

Copyright 2004 by the American Geophysical Union.
0043-1397/04/2003WR002876$09.00

31113

kurtosis, $\mu_4 = \langle(\Re - \mu)^4\rangle$, respectively. It is therefore common to interpret the sample moments as random quantities and to relate them to their ensemble counterparts.

[6] The sample mean of $\Re$ has the following properties:

$$\langle\overline{\Re}_n\rangle = \mu; \qquad \mathrm{var}\left[\overline{\Re}_n\right] = \frac{1}{n}\sigma^2 \tag{1}$$

Furthermore, if $\Re$ is normal, i.e., $\Re \sim$ N(), then

$$\frac{\overline{\Re}_n - \mu}{\sigma/\sqrt{n}} \sim N_{0,1}() \quad \text{and} \quad \frac{\overline{\Re}_n - \mu}{S_n/\sqrt{n}} \sim \mathrm{t}_{n-1}() \qquad \forall n \tag{2}$$

where $N_{0,1}()$ is the standard normal distribution and $t_{n-1}()$ is the Student distribution with $(n-1)$ degrees of freedom. Equation (1) tells us that we can take $\overline{\Re}_n$ as an estimator for μ; $\mathrm{var}[\overline{\Re}_n]$ is an indicator of the uncertainty of such an estimation. A more precise indication of the uncertainty is given by the evaluation of confidence intervals; these can be derived from the second of (2) whenever is normal:

$$\Pr\left[\overline{\Re}_n - t_{n-1}\left(1-\frac{\alpha}{2}\right)\frac{S_n}{\sqrt{n}} \le \mu \le \overline{\Re}_n + t_{n-1}\left(1-\frac{\alpha}{2}\right)\frac{S_n}{\sqrt{n}}\right] = 1-\alpha \tag{3}$$

where 1-α is the probability that the value of the process mean μ lies within the confidence interval around the sample mean $\overline{\Re}_n$. More in general, if $\Re$ follows a generic (unknown) probability distribution, confidence intervals for $\overline{\Re}_n$ can be (over)estimated via Chebyshev inequality:

$$\Pr\left[\overline{\Re}_n - a\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{\Re}_n + a\frac{\sigma}{\sqrt{n}}\right] \ge 1 - \frac{1}{a^2} \tag{4}$$

where $(1 - 1/a^2)$ is an (under)estimation of the probability that the value of the process mean μ lies within the confidence interval around the sample mean $\overline{\Re}_n$. Equations (3) and (4) show that $S_n/\sqrt{n}$, which is an estimator for $\sigma/\sqrt{n}$, is the scaling factor of the confidence intervals amplitude of the mean for any given probability value.

[7] The sample variance of $\Re$ has the following properties:

$$\langle S_n^2\rangle = \sigma^2; \qquad \mathrm{var}\left[S_n^2\right] = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right) \tag{5}$$

Furthermore, if $\Re \sim$ N(), then

$$\mathrm{var}\left[S_n^2\right] = \frac{2}{n-1}\sigma^4 \tag{6}$$

$$\begin{gathered}\sum_{i=1,n}\frac{(\Re_i - \mu)^2}{\sigma^2} \sim \chi_n^2() \quad \text{and} \\ \sum_{i=1,n}\frac{\left(\Re_i - \overline{\Re}_n\right)^2}{\sigma^2} = (n-1)\frac{S_n^2}{\sigma^2} \sim \chi_{n-1}^2() \qquad \forall n\end{gathered} \tag{7}$$

where $\chi_n^2()$ is the chi-square distribution with n degrees of freedom. It follows that

$$\Pr\left[\frac{n-1}{\chi_{n-1}^2(1-\alpha/2)}S_n^2 \le \sigma^2 \le \frac{n-1}{\chi_{n-1}^2(\alpha/2)}S_n^2\right] = 1-\alpha \tag{8}$$

Equation (5) states that S_n^2 is a possible estimator for σ^2. If the sample variance S_n^2 and kurtosis $M_{4,n}$ are used in equation (5) instead of their "exact" counterparts, σ^2 and μ_4, one can obtain an indicator of the uncertainty of the estimation of σ^2 by means of S_n^2. For the case of normally distributed variables it can easily be shown from equation (8) that the width of the confidence intervals of the variance, for any α value, scales with S_n^2/n for $n\to\infty$. Again, Chebyshev inequality can be used to estimate the confidence interval for S_n^2 when $\Re$ is not normally distributed.

[8] For easier evaluation of the confidence intervals (3) and (8), it is useful to point out that both the $t_n()$ and the $\chi_n^2()$ distributions can be approximated via the normal distribution when n is sufficiently large (say, $n \ge 30$), being $t_n() \cong N_{0,1}()$ and $\sqrt{2\chi_n^2()} - \sqrt{2n-1} \cong N_{0,1}()$.

[9] Equations (1)–(8) express well known properties of statistical inference. Their application to the evaluation of the degree of convergence of Monte Carlo results is straightforward. As MC realizations are used to estimate ensemble moments, the uncertainties of these estimates can be evaluated as a function of the values of the sample moments and of the size of the sample, i.e., of the number of process realizations.

3. Application Example

[10] The methodology is here applied to a simple two-dimensional convergent flow example taken from the literature [*Riva et al.*, 2001]. We consider steady state incompressible horizontal flow to a well of zero radius, located at the center of a circular (random, scalar) transmissivity domain having radius L. The well pumps at a constant deterministic rate, Q, and is assumed as the origin of a Cartesian coordinate system (x, y). The head at the outer circular boundary remains at a constant, deterministic value, H_L.

[11] We performed numerical Monte Carlo simulations using the same code which was employed by *Riva et al.* [2001] for steady state flow in a square domain with 100 rows and 100 columns of uniform size ($\Delta x = \Delta y = \Delta$). A circular boundary of radius $L = 50\Delta$ was defined about the well by designating all cells outside it as inactive. We modeled the log hydraulic transmissivity, $Y = \ln T$, as a statistically homogeneous and isotropic random function of space with Gaussian covariance. The hydraulic head, H_L, along the circular boundary was set equal to 80 (in consistent units). A pumping well at a constant rate $Q = 15$ was placed at the central node of the grid. Gaussian sequential simulation (based on the code SGSIM [*Deutsch and Journel*, 1998]) was used to generate random realizations of Y on the above defined two-dimensional grid. Each realization constituted a sample from a multivariate Gaussian, statistically homogeneous field, Y, of zero mean, $\langle Y\rangle = 0$, and with the target isotropic Gaussian covariance of unit variance, σ_Y^2, and unit correlation length, λ.

[12] An extensive suite of 200,000 Monte Carlo flow simulations was performed, to obtain relatively stable values of the sample moments of interest and verify the method of investigation. In the following we focus on the analysis of the first and second moments of the hydraulic head, h, at a point with coordinates ($x/L = -0.44$, $y/L = 0$). In order to better grasp the relative uncertainties, we will consider the head drawdown, $\Delta h = H_L - h$, from the undisturbed condition.

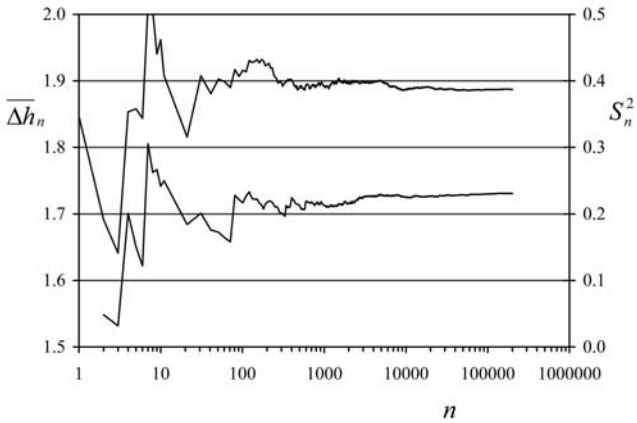

Figure 1. Convergence of the sample mean and variance of head at the monitoring point as functions of the number of Monte Carlo simulations.

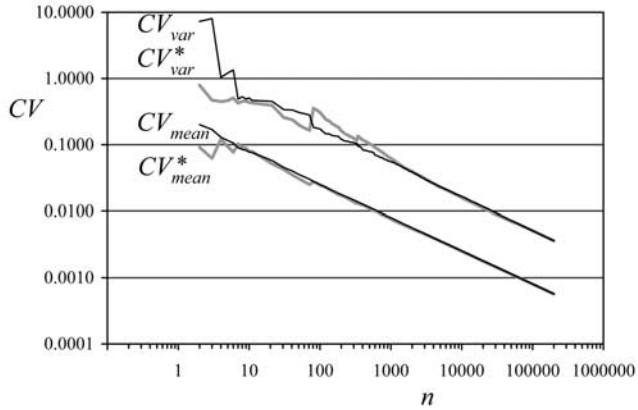

Figure 2. Coefficients of variation for the sample mean and variance of head at the monitoring point as functions of the number of Monte Carlo simulations.

[13] Figure 1 depicts the sample mean, $\overline{\Delta h_n}$, and variance, S_n^2, of head drawdown computed at the monitoring point. We consider a log n-axis to be preferred to a linear axis for a visual assessment of convergence. As previously discussed, the root mean square (RMS) of the estimators for the mean and the variance scales with $1/\sqrt{n}$. This implies that, for a number of Monte Carlo simulations $n = 100$, one needs to add about 300 more realizations in order to halve uncertainty. However, 3000 additional realizations are needed for the same reduction when analyzing moments obtained by $n = 1000$. This behavior is automatically reproduced by the log scale, while the natural scale stretches oscillations and trends, thus leading to an overestimation of convergence rates. It can be noted that increasing the number of MC simulations from 1000 to 2000 to obtain a better stabilization of the results, as it is often done in the literature, has basically no effect on the overall convergence of the process, as one should change order of magnitude to clearly see significant effects.

[14] The qualitative convergence analysis is then followed by the quantification of the uncertainty associated with the computed sample moments. This is accomplished by analyzing their coefficients of variation, $CV_{mean} = \mathrm{RMS}[\overline{\Delta h_n}]/\overline{\Delta h_n}$ and $CV_{var} = \mathrm{RMS}[S_n^2]/S_n^2$, respectively. When the numerators of the two coefficients of variation are estimated upon substitution of the sample second and fourth moments in the second of (1) and (5) we obtain the corresponding estimated values CV^*_{mean} and CV^*_{var}. The coefficients of variation are depicted in Figure 2 as a function of the number of Monte Carlo simulations, at the chosen monitoring point. Figure 2 also reports the "exact" values CV_{mean} and CV_{var}, computed by using the sample moments from the whole set of 200,000 MC realizations. As n increases, values of CV^* become increasingly accurate estimates of the corresponding CV. These graphs can provide information about the order of magnitude of the uncertainty associated to moments estimated by a given number of Monte Carlo simulations, even though this cannot be strictly translated into a confidence interval. This can also provide an estimate about the number of simulations required to refine the accuracy of our predictions by a given percentage and allows one to effectively plan the number of Monte Carlo needed.

[15] These aspects are also emphasized in Figures 3a and 3b, depicting profiles along the axis $y = 0$ of differences between the mean head drawdowns computed after 100 and 2,000 MC simulations and what we assumed to be the ground truth, i.e., the mean head computed after $n = 200{,}000$. The corresponding $\pm 2\ S_n/\sqrt{n}$ intervals are also reported. We can clearly see how the $2\ S_n/\sqrt{n}$ intervals vary with distance from the well and that their width decreases as

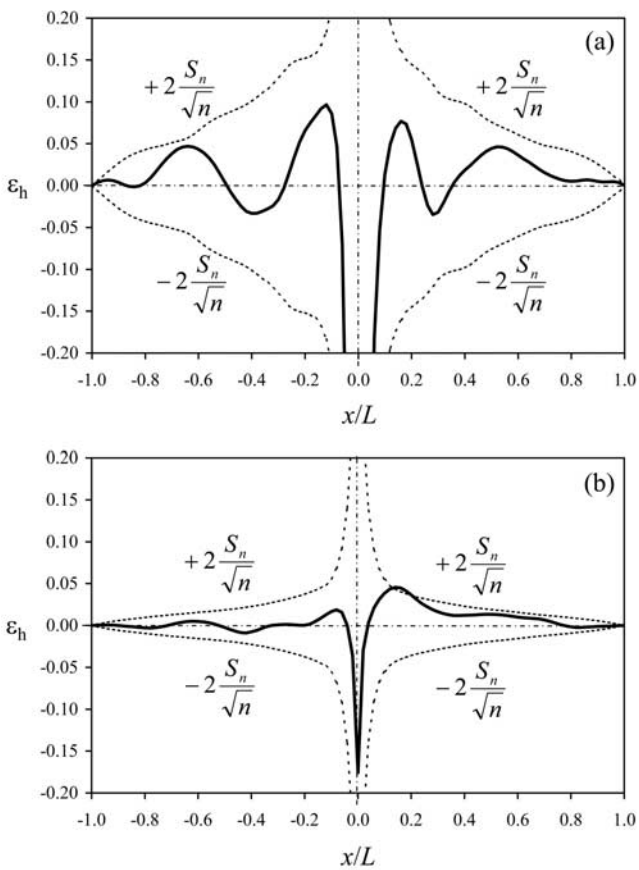

Figure 3. Differences between the mean head drawdowns, $\varepsilon_h = h(n) - h(n = 200{,}000)$. Dashed curves indicate the $\pm 2S_n(x/L, y/L = 0)/\sqrt{n}$ intervals. (a) $n = 100$; (b) $n = 2{,}000$.

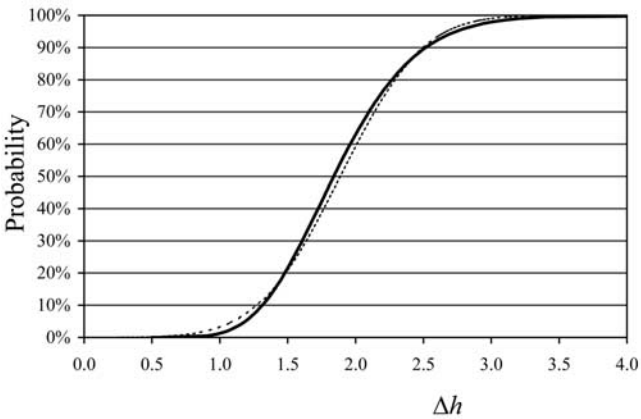

Figure 4. Cumulative distribution of hydraulic head drawdown at the monitoring point (solid line) compared to the corresponding normal distribution for the same mean and variance values (dashed line).

the number of MC simulations increases. The width of the radial oscillations of differences is coherent with the local scale of the 2 $S_n/\sqrt{n}$ bounds.

[16] Values of estimator variance can be related to confidence intervals if the shape of the probability distribution of the estimator itself is known. This is the case if the hydraulic head is normally distributed (equations (2) and (7)). Figure 4 depicts the cumulative distribution of hydraulic head computed at the monitoring point, after n = 60,000. The corresponding normal distribution is reported, computed with the same sample mean and variance. Shapiro-Wilkins test of the null hypothesis that these 60,000 Monte Carlo-generated heads are Gaussian was negative at a significance level of 5%. The result of the test shows that the hydraulic head cannot be considered normally distributed from a statistical point of view; however, Figure 4 shows that the departure from normality is not too large. We can therefore expect equations (3) and (8) to give reasonable approximations for confidence intervals. The significance of such approximation can be judged from Figure 5a, depicting the stabilization analysis of the mean hydraulic head drawdown $\overline{\Delta h}_n$ at the monitoring point. We have here reported (1) the set of curves one would obtain upon breaking down the sample of n = 200,000 into 100 subsets, each one comprising 2,000 simulations, (2) the stabilization curve obtained for the complete set of 200,000 simulations (the same as in Figure 1), and (3) the 95% estimated confidence intervals (α = 0.05). The latter have been computed on the basis of the standard deviation obtained after n = 200,000. In order to verify the approximation of the computed confidence intervals, Figure 5b depicts the cumulative distribution of the normalized variable $(\overline{\Delta h}_n - \mu)/(S_n/\sqrt{n})$, as obtained

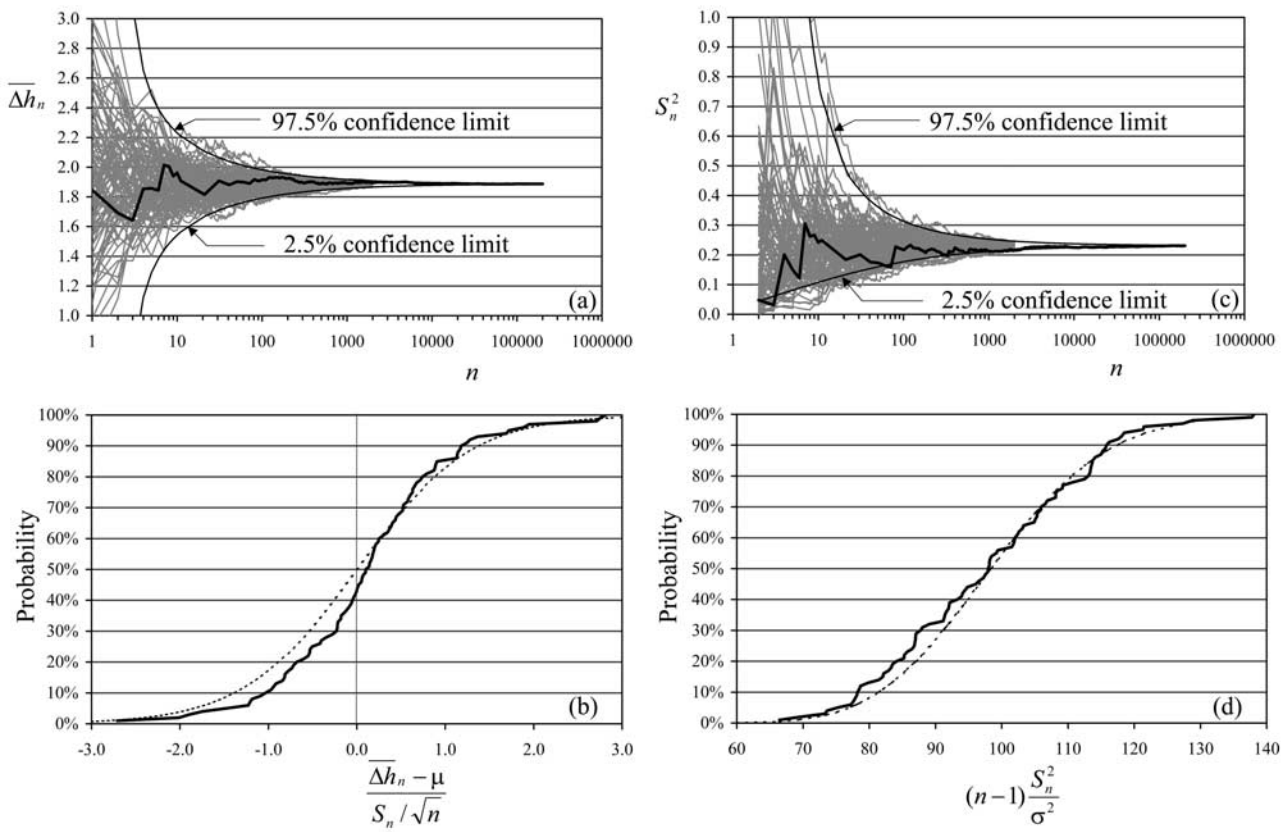

Figure 5. Dependence of convergence of (a) sample mean and (c) variance of head on the number of Monte Carlo simulations. The data set of n = 200,000 (thick solid line) is subdivided into 100 series of n = 2,000 each (shaded lines). Cumulative distributions (for the 100 subsets) of the estimates for (b) mean and (d) variance of head drawdown for n = 100. Corresponding Student and χ^2 distributions are shown (dashed lines).

from the 100 subsets created, and the corresponding $t_{n-1}(\ )$ distribution, at $n = 100$. The corresponding depictions for the hydraulic head variance, S_n^2, at the same location are reported in Figures 5c and 5d. In Figure 5d, for the sake of comparison, the $\chi^2_{n-1}(\ )$ distribution has been juxtaposed to the empirical distribution of $(n - 1)\ S_n^2/\sigma^2$. The values for μ and σ^2 are those computed after $n = 200{,}000$. Visual inspection of Figure 5 reveals that the computed confidence intervals are indeed meaningful, in that they provide the order of magnitude of the uncertainty associated with the first and second moments one evaluates on the basis of a finite number of Monte Carlo realizations.

4. Conclusions

[17] Our work leads to the following major conclusions.

[18] 1. Assessment of convergence of Monte Carlo simulations must be performed both qualitatively and quantitatively. A qualitative analysis is more meaningful when plotting the results in semi-log scale. A local quantitative analysis can be performed by means of simple relationships, based on rules of statistical inference.

[19] 2. It is possible to identify the order of magnitude of the uncertainty associated with the first and second moments of a state variable of interest when these are evaluated on the basis of a finite number of Monte Carlo runs.

[20] 3. From a practical viewpoint, it is possible to obtain an estimate of the number of Monte Carlo iterations required to refine the accuracy of our predictions by a given percentage.

[21] 4. Although the methodology has been described with reference to the first two (ensemble) moments of hydraulic head in a typical groundwater flow problem, it can be extended to other statistical quantities as well as to different application fields where Monte Carlo simulations are used.

[22] **Acknowledgments.** This work was supported by the European Commission under contract EVK1-CT-1999-00041 (W-SAHaRA - Stochastic Analysis of Wellhead Protection and Risk Assessment).

References

Bellin, A., Y. Rubin, and A. Rinaldo (1994), Eulerian-Lagrangian approach for modeling of flow and transport in heterogeneous geological formations, *Water Resour. Res.*, *30*(11), 2913–2924.

Deutsch, C. V., and A. G. Journel (1998), *GSLIB Geostatistical Software Library and User's Guide*, Oxford Univ. Press, New York.

Guadagnini, A., M. Riva, and S. P. Neuman (2003), Three-dimensional Steady state flow to a well in a randomly heterogeneous bounded aquifer, *Water Resour. Res.*, *39*(3), 1048, doi:10.1029/2002WR001443.

Hassan, A. E., J. H. Cushman, and J. W. Delleur (1998), A Monte Carlo assessment of Eulerian flow and transport perturbation models, *Water Resour. Res.*, *34*(5), 1143–1163.

Johnson, P. A. (1992), Reliability-based pier scour engineering, *J. Hydraul. Eng.*, *118*(10), 1344–1358.

Li, L., H. A. Tchelepi, and D. Zhang (2003), Perturbation-based moment equation approach for flow in heterogeneous porous media: Applicability range and analysis of high-order terms, *J. Comput. Phys.*, *188*, 296–317.

Mood, A. M., F. A. Graybill, and D. C. Boes (1974), *Introduction to the Theory of Statistics*, 3rd edition, McGraw-Hill, New York.

Riva, M., A. Guadagnini, and F. Ballio (1999), Time related capture zones for radial flow in two-dimensional randomly heterogeneous media, *Stochastic Environ. Res. Risk Assess.*, *13*(3), 217–230.

Riva, M., A. Guadagnini, S. P. Neuman, and S. Franzetti (2001), Radial flow in a bounded randomly heterogeneous aquifer, *Transp. Porous Media*, *45*, 139–193.

van Vuren, S., H. van der Klils, and H. de Vriend (2002), Large-scale floodplain lowering along the River Waal: A stochastic prediction of morphological impacts, in *River Flow 2002: Proceedings of the International Conference on Fluvial Hydraulics*, edited by D. Bousmar and Y. Zech, pp. 903–912, A. A. Balkema, Brookfield, Vt.

Yen, B. C., and Y. Tung (Eds.) (1993), *Reliability and Uncertainty Analysis in Hydraulic Design*, Am. Soc. of Civ. Eng. Subcomm. on Uncertainty and Reliability Anal. in Design of Hydraul. Structures, Reston, Va.

F. Ballio and A. Guadagnini, Dipartimento di Ingegneria Idraulica, Ambientale, Infrastrutture Viarie, Rilevamento, Politecnico di Milano, Piazza L. Da Vinci 32, 20133 Milano, Italy. (francesco.ballio@polimi.it)