

Comment on “Modeling groundwater flow with random hydraulic conductivity using radial basis function partition of unity method” by Shile et al. (2024)

Nicolae Suciu

Tiberiu Popoviciu Institute of Numerical Analysis, Romanian Academy, Cluj-Napoca, Romania

ARTICLE INFO

Dataset link: [Comment on \[Shile et al., 2024\] \(Original data\)](#)

MSC:

76S05

65C05

65M50

Keywords:

Heterogeneous aquifers

Flow benchmark

Monte Carlo

Data fabrication

ABSTRACT

The goal of this article is to explain some striking discrepancies between the Monte Carlo inferences for flow in heterogeneous aquifers presented in (Shile et al., 2024) and reliable and verifiable results previously published. Comparisons with statistical inferences done within the same numerical setup and using the same benchmark codes demonstrate that quantifiable aspects, such as a too small departure from the linear theory and the decrease of the standard deviations with the increase of the aquifer heterogeneity, cannot be produced with the tools used by the authors.

1. Introduction

Shile et al. [1] propose a radial basis function (RBF) collocation method to simulate steady-state flows in heterogeneous aquifers. The RBF approach is tested on benchmark problems proposed in [2], using the Matlab codes published in [3]. The hydraulic-head solution is implemented with the codes posted at [4]. Note that these codes are not properly cited, the reference [15] cited by Shile et al. [1], containing a sample Matlab code, does not provide a link to the codes and functions used to solve the RBF interpolation problems presented in their paper. Also, the Ref. [3] is not cited as a distinct publication, a link to the Git repository being instead added to Ref. [2], so that the reader cannot know where different elements of the benchmark were taken from.

The benchmark study of Alecsa et al. [2] is two-fold. First, code verification tests, with manufactured solutions, were conducted for a finite difference method (FDM), a global random walk algorithm (GRW), a finite element method (FEM), and a discontinuous Galerkin method (DGM), as well as for a collocation spectral method (CSM). The realizations of the hydraulic conductivity were provided by a randomization method as analytic space functions. The manufactured solutions were constructed for a fixed realization of the hydraulic conductivity as smooth space functions. These two ingredients favor the

CSM, which, using analytical differentiation of the hydraulic conductivity and relatively small numbers of collocations points, yields solutions with an accuracy that is orders of magnitude higher than that of the classical discretization approaches FDM, FEM, GRW, and DGM.

The second part of the study presented in [2] consists of statistical inferences on ensembles of Monte Carlo simulations of groundwater flow computed for increasing variance σ^2 of the log-hydraulic conductivity. This setup, representative for realistic situations, favors the classical discretization methods. Instead, the feasibility of the CSM is limited to small values σ^2 , due to the need to increase the number of collocation points as the heterogeneity of the hydraulic conductivity increases.

The RBF approach proposed by Shile et al. [1] passes the code verification tests, with errors with respect to the manufactured solutions smaller than those of the classical discretization methods and much larger than the errors of the CSM method presented in [2]. Instead, the statistical inferences results show striking discrepancies with results obtained within the same numerical setup. As shown in the following, the results from [1] cannot be explained by numerical artifacts and there is a strong evidence that the Monte Carlo results are fabricated and the comparisons with previous studies are falsified.

E-mail address: nsuciu@ictp.acad.ro.

<https://doi.org/10.1016/j.enganabound.2024.105963>

Received 30 June 2024; Received in revised form 10 September 2024; Accepted 12 September 2024

Available online 16 September 2024

0955-7997/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Table 1Deviation from linear theory of velocity variances for $\sigma^2 = 1.5$.

	RBF	DGM	Refs. [5–7]	Eq. (3)
$\epsilon_{V_x}^2$	2.14	14.87	12.50	9.19
$\epsilon_{V_y}^2$	4.81	27.03	25.00	41.60

Table 2

Deviation from linear theory of RBF variances [1].

	Gauss				Exp			
σ^2	0.5	1	1.5	2	0.5	1	1.5	2
$\epsilon_{V_x}^2$	0.00	0.80	1.2456	0.93	1.07	1.07	2.14	1.87
$\epsilon_{V_y}^2$	0.00	0.80	2.1390	3.20	0.48	3.20	4.81	4.40

Table 3

Deviation from linear theory of DGM results [2].

	Gauss				Exp			
σ^2	0.5	1	1.5	2	0.5	1	1.5	2
$\epsilon_{V_x}^2$	1.78	3.35	4.70	5.97	7.73	11.07	14.87	19.07
$\epsilon_{V_y}^2$	10.18	15.27	19.59	23.45	10.22	18.56	27.03	35.71

2. Discrepancy with literature results

The numerical procedures used in Monte Carlo simulations of transport in groundwater are validated through comparisons with linear theory and similar numerical approaches (see [2,5–8]). One of the mostly used validation test in two-dimensional simulations consists of comparisons of the variances of the longitudinal velocity V_x and of the transverse velocity V_y , defined by

$$\sigma_{V_x}^2 = \langle V_x^2 \rangle - \langle V_x \rangle^2, \quad \sigma_{V_y}^2 = \langle V_y^2 \rangle - \langle V_y \rangle^2, \quad (1)$$

where $\langle \cdot \rangle$ stands for the average over the ensemble of realizations.

The studies cited above indicate a similar, monotonic increase of the velocity variances with the increase of the variance σ^2 of the log-hydraulic conductivity above the variances predicted by the linear theory, which, for unit mean longitudinal velocity considered in [1,2], are given by

$$\sigma_{lin,V_x}^2 = \frac{3}{8}\sigma^2, \quad \sigma_{lin,V_y}^2 = \frac{1}{8}\sigma^2. \quad (2)$$

The departure from the linear approximation can be illustrated by the semi-analytical second-order approximation, calibrated by Monte Carlo simulations, obtained by Gotovac et al. [8]:

$$\sigma_{V_x}^2 = \frac{3}{8}\sigma^2 \left(1 + \frac{376}{875}\sigma^2\right)^{0.175}, \quad \sigma_{V_y}^2 = \frac{1}{8}\sigma^2 \left(1 + \frac{328}{535}\sigma^2\right)^{0.535}. \quad (3)$$

The percentage deviation from the linear approximation (2) will be evaluated in the following by

$$\epsilon_{V_x}^2 = \frac{\sigma_{V_x}^2 - \sigma_{lin,V_x}^2}{\sigma_{lin,V_x}^2} 100, \quad \epsilon_{V_y}^2 = \frac{\sigma_{V_y}^2 - \sigma_{lin,V_y}^2}{\sigma_{lin,V_y}^2} 100. \quad (4)$$

In Table 1 the results of Shile et al. [1, Table 9] for $\sigma^2 = 1.5$ and exponential correlation of the log-hydraulic conductivity are compared via (4) to the DGM results of Alecsa et al. [2], to the results from [5–7], which are close at less than 1.5% to each other [7], and to the second-order approximation (3). The deviation from linear theory of the RBF results shown in Table 1 is one order of magnitude smaller than in earlier published studies.

The same striking discrepancy is shown in Table 2 for the results of Shile et al. [1] for the whole range σ^2 and for both Gaussian and exponential correlations.

Tables 3 and 4 show deviations from the linear theory of the DGM results [2] and of the second-order approximation (3), respectively, which are again one order of magnitude larger than those

Table 4

Deviation from linear theory of the second-order approximation (3) for exponential correlation of the log-hydraulic conductivity.

σ^2	0.5	1	1.5	2
$\epsilon_{V_x}^2$	3.74	6.46	9.19	11.47
$\epsilon_{V_y}^2$	15.38	29.15	42.13	53.45

Table 5

Deviation of the RBF results from the second-order approximation (3) for exponential correlation of the log-hydraulic conductivity.

σ^2	0.5	1	1.5	2
$\epsilon_{V_x}^2$	-2.58	-5.06	-6.46	-8.61
$\epsilon_{V_y}^2$	-12.91	-20.09	-26.25	-31.96

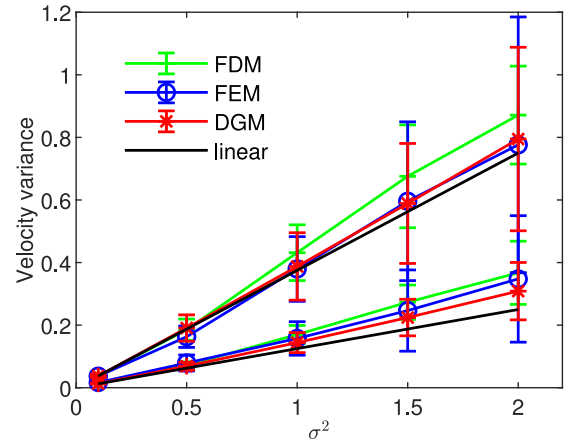


Fig. 1. $\sigma_{V_x}^2$ (upper branch) and $\sigma_{V_y}^2$ (lower branch) from Alecsa et al. [2] - Gaussian correlation.

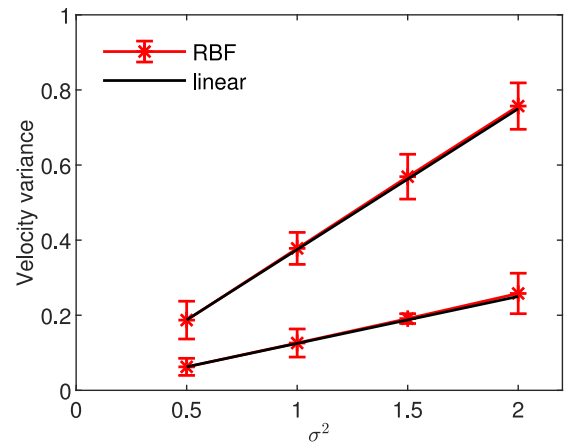


Fig. 2. $\sigma_{V_x}^2$ (upper branch) and $\sigma_{V_y}^2$ (lower branch) from Shile et al. [1] - Gaussian correlation.

of Shile et al. [1]. And, finally, Table 5 shows that the RBF variances results [1] are systematically smaller than those given by the second-order approximation (3) and the discrepancy increases with σ^2 .

A visual comparison of the results of Shile et al. [1] and earlier results from Alecsa et al. [2] and Gotovac et al. [8], computed with Eq. (3), is presented in Figs. 1–4. Here again one can see that the results presented by Shile et al. [1, Table 9] largely underestimate the

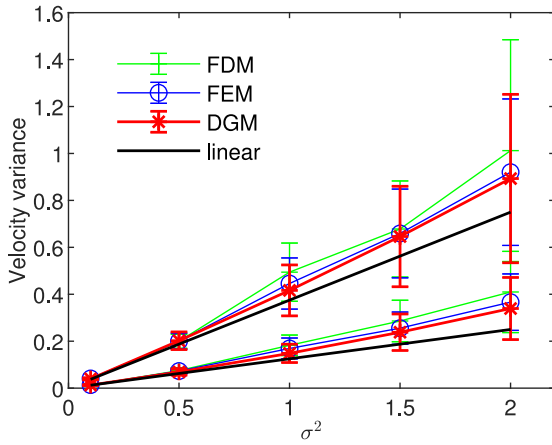


Fig. 3. $\sigma_{V_x}^2$ (upper branch) and $\sigma_{V_y}^2$ (lower branch) from Alecsa et al. [2] - exponential correlation.

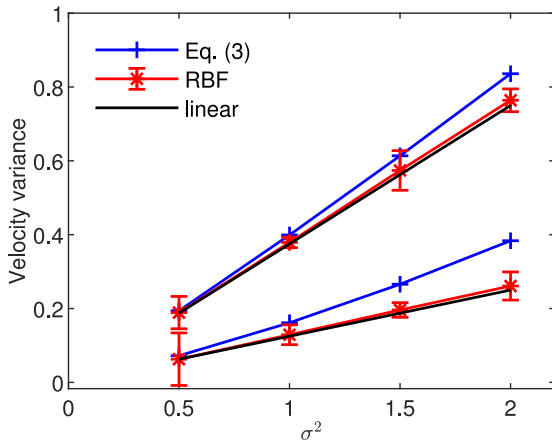


Fig. 4. $\sigma_{V_x}^2$ (upper branch) and $\sigma_{V_y}^2$ (lower branch) from Shile et al. [1] - exponential correlation.

departure of the velocity variances from the prediction of the linear theory.

3. Made up results

The second striking discrepancy is the non-monotonous behavior of the standard deviations in Shile et al. [1, Table 9], indicated by error bars in Figs. 2 and 4, as compared with monotonically increasing standard deviations in Figs. 1 and 3 which present results from [2], that are similar to those published by Bellin et al. [5, Figs. 6a and 6b].

The standard deviations of the ensemble average velocity variances (1) are computed by space averages according to

$$SD(\sigma_{V_x}^2) = \sqrt{\langle (\sigma_{V_x}^2)^2 \rangle_s - \langle \sigma_{V_x}^2 \rangle_s^2}, \quad SD(\sigma_{V_y}^2) = \sqrt{\langle (\sigma_{V_y}^2)^2 \rangle_s - \langle \sigma_{V_y}^2 \rangle_s^2}, \quad (5)$$

where $\langle \cdot \rangle_s$ stands for the spatial average over the inner core region of the computational domain not affected by the boundary conditions [2]. If the random velocity is statistically homogeneous $\sigma_{V_x}^2$ and $\sigma_{V_y}^2$ do not depend on the spatial variable and their standard deviations defined by (5) vanish. However, since the numerically generated fields are not strictly stationary, the ensemble averages vary from point to point, resulting in a consistent increase of the standard deviations (5) with the increase of σ^2 , reported in many studies published since the pioneering article of Bellin et al. [5].

Shile et al. [1] used the codes from [3] to generate the random hydraulic conductivity and claim [1, Sect. 4.4] that they used the same numerical setup as Alecsa et al. [2] and the definition (5) to compute standard deviations of spatially variable velocity variances. Moreover, the code verification tests with two-dimensional manufactured solutions [1, Tables 5 and 6] show that, for $\sigma^2 \leq 2$, their RBF solver is at least as accurate as the DGM code used in [2, Tables 1 and 3]. In these conditions the results obtained with the two approaches should be, in principle, almost identical.

On the other side, the manufactured solutions used in code verification tests are very smooth, with slow oscillations in space (see [1, Fig. 12]), and can be approximated by relatively small numbers of collocation points used in the RBF code, as well as in the CSM code tested in [2]. In these circumstances, the two collocation methods mentioned above benefit from the explicit analytical expression of the hydraulic conductivity used in the benchmark problems [2, Eq. (A3)] to compute its derivatives. Therefore, they outperform the classical approaches which discretize the derivatives (see discussion in [2, Sect. 6]). Instead, the solutions of the homogeneous flow equation used in Monte Carlo simulations are highly heterogeneous and in a first-order approximation follow the shape of the hydraulic conductivity shown in [1, Fig. 2]. Accurate representations of such solutions require more collocation points than in case of the smooth manufactured solutions.

We have thus two possible situations. First, if the same number of collocation points as in code verification tests allows accurate representations of the solutions, as the authors claim in [1, Sect. 4.4], then we are in the case described above, when the RBF and DGM results should be almost identical. This is not the case. The second possibility is that the number of collocation points is not large enough. Then, the representation of the solution is inaccurate and there are realizations with large deviations from the expected almost-laminar behavior (see Supplementary material 2. in [2] which shows “bad” realizations of the CSM collocation method). In this case, larger spatial variability yields larger standard deviations of the velocity variances than the DGM code. This also is not the case (compare Figs. 1 and 2, for Gaussian correlation, and Figs. 3 and 4, for exponential correlation of the $\ln K$ field). A first conclusion of this analysis is that the standard deviations presented by Shile et al. [1, Table 9] cannot be produced by their RBF code.

To elucidate the issue of non-monotonous standard deviations shown in Figs. 2 and 4, I computed ensembles of 100 realizations of the $\ln K$ field, for Gaussian and exponential correlations, by using the same parameters and generated with the same codes as in [1,2], posted in the new folder “Comment on [Shile et al. 2024]” of the Git repository [3]. The estimated mean and standard deviation of variance $\sigma_{\ln K}^2$ are plotted in Fig. 5. As expected, the standard deviation $SD(\sigma_{\ln K}^2)$ increases monotonically with the nominal parameter σ^2 . With these, the nominal variance σ^2 used in the Monte Carlo simulations is approximated with a 95% confidence level by

$$\sigma^2 \approx \overline{\sigma_{\ln K}^2} \pm SD(\sigma_{\ln K}^2). \quad (6)$$

From (2) and (6) we get the following linear approximation of the velocity variances,

$$\sigma_{\ln V_x}^2 \approx \frac{3}{8} \overline{\sigma_{\ln K}^2} \pm \frac{3}{8} SD(\sigma_{\ln K}^2), \quad \sigma_{\ln V_y}^2 \approx \frac{1}{8} \overline{\sigma_{\ln K}^2} \pm \frac{1}{8} SD(\sigma_{\ln K}^2). \quad (7)$$

If the number of random modes used in the field generator and the number of realizations go to infinity, one approaches a statistical homogeneous $\ln K$ field, $SD(\sigma_{\ln K}^2) \rightarrow 0$, and (7) reduces to (2).

The quantities $\frac{3}{8} SD(\sigma_{\ln K}^2)$ and $\frac{1}{8} SD(\sigma_{\ln K}^2)$ in (7) are confidence half-intervals of the mean variances $\frac{3}{8} \overline{\sigma_{\ln K}^2}$ and $\frac{1}{8} \overline{\sigma_{\ln K}^2}$ induced by the statistical inhomogeneity of the $\ln K$ field. The same repository also contains ensembles of linear approximations of the corresponding velocity components generated by the Kraichnan method described in [9, Appendix C.3], without the need to solve the flow equation. The comparison presented in Table 6 shows that the confidence intervals

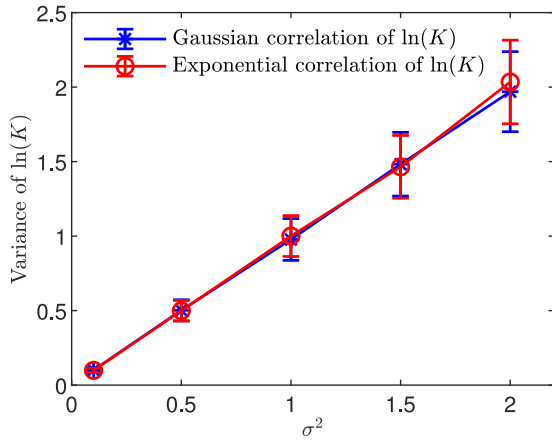


Fig. 5. Standard deviations of $\ln K$ computed in a first order approximation with the Kraichnan method.

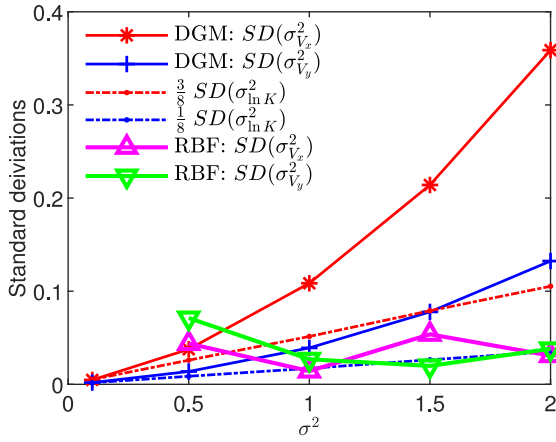


Fig. 6. Standard deviations for DGM [2] and RBF [1] compared with the first-order estimates - exponential correlation of $\ln K$.

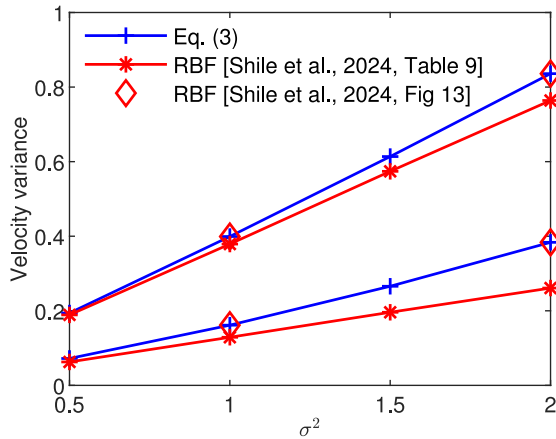


Fig. 7. Analysis of Fig. 13 in [1] - detail for $\sigma^2 \leq 2$.

from (7) estimate the standard deviations of the velocity variances in a first-order approximation. Since numerical methods to solve the flow equation are prone to errors which increase the spatial inhomogeneity, the estimates from Table 6, computed without solving the flow equation, represent lower bounds of the standard deviations.

The detailed comparison of standard deviations for exponential correlation from Fig. 6 shows a total mismatch between the RBF and

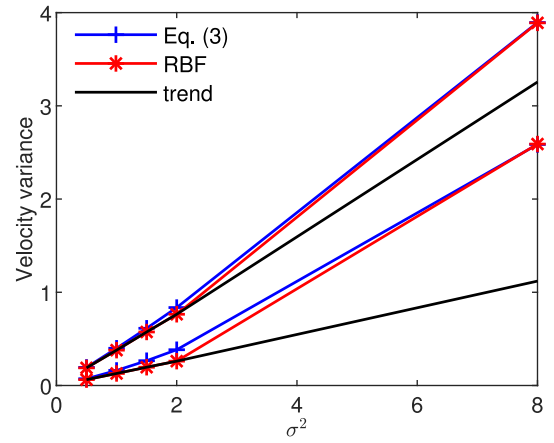


Fig. 8. Analysis of Fig. 13 in [1] - with extrapolated trend from Fig. 7.

Table 6

Standard deviations of velocity variances estimated by (7) compared with Monte Carlo estimates for the linear approximation in the exponential case.

σ^2	0.1	0.5	1	1.5	2
$\frac{3}{8} SD(\sigma^2_{\ln K})$	5.17e-03	2.59e-02	5.14e-02	7.90e-02	1.05e-01
$SD(\sigma^2_{\ln V_x})$	5.07e-03	2.69e-02	5.35e-02	7.82e-02	1.06e-01
$\frac{1}{8} SD(\sigma^2_{\ln K})$	1.72e-03	9.01e-03	1.63e-02	2.96e-02	3.34e-02
$SD(\sigma^2_{\ln V_y})$	1.76e-03	8.26e-03	1.71e-02	2.55e-02	3.41e-02

DGM results. Moreover, the RBF data points show non-monotonous dependence of the standard deviations on σ^2 and values smaller than the lower bounds $\frac{3}{8} SD(\sigma^2_{\ln K})$ and $\frac{1}{8} SD(\sigma^2_{\ln K})$. Since such results cannot be obtained with any numerical method, which was also validated by comparisons with manufactured solutions, the only possible explanation for these flagrant discrepancies is that the results of Shile et al. [1, Table 9] are made up.

Summarizing, the results presented by Shile et al. [1] mimic the general feature that the distance between the numerically estimated variance of the velocity components and their linear approximation increases with the increase of σ^2 , while neglecting three other aspects:

- the magnitude of the distance from the linear approximation, which is almost the same in previously published studies;
- the monotonous increase of the standard deviations of the velocity variances;
- the lower bound of these standard deviations given by the standard deviation of the variance of the $\ln K$ field multiplied by 3/8 and 1/8, for longitudinal and transverse velocity, respectively.

4. Falsified comparison with published results

Based on a comparison plotted in their Fig. 13, Shile et al. [1] conclude the presentation of the statistical inferences by stating that “there is good agreement between all the stochastic solutions, supporting the verification of RBFUM-QR with previously published stochastic results”. However, as we will see in the following, this statement is not supported by the numerical results presented by the authors in [1, Table 9].

The analysis of the results shown in Figs. 7 and 8 demonstrates that the claim “Fig. 13 confirms that for higher values of σ^2 , RBFUM-QR produces results similar to those obtained using the Eqs. (42) and (43)” is false.

First, I compare in Fig. 7 the data points given by Eq. (3) and the RBF data from [1, Table 9] for $\sigma^2 \leq 2$ to the data points from [1, Fig. 13] for $\sigma^2 = 1$ and $\sigma^2 = 2$. The distance of 0.123 between the points corresponding to RBF and to Eq. (3) on the lower branch for $\sigma^2 = 2$ corresponds to about half the height of the diamond marker

Table 7

Data points of the second-order approximation (3) compared to RBF data points from Shile et al. [1, Fig. 13 and Table 9].

σ^2	1		2	
	$\sigma_{V_s}^2$	$\sigma_{V_f}^2$	$\sigma_{V_s}^2$	$\sigma_{V_f}^2$
Eq. (3)	4.00e-01	1.61e-01	8.36e-01	3.84e-01
Shile et al. [1, Fig. 13]	4.00e-01	1.60e-01	8.37e-01	3.82e-01
Shile et al. [1, Table 9]	3.79e-01	1.29e-01	7.64e-01	2.61e-01

in [1, Fig. 13]. As such, this difference should have been visible in [1, Fig. 13] and the marker should have been placed just below the line representing Eq. (43). This observation is reinforced by the RBF data points from [1, Fig. 13] provided by the authors during the review process which, as shown in Table 7, are different from the Monte Carlo estimates presented by Shile et al. [1, Table 9], being instead practically identical to those given by Eq. (3).

The mismatch between the results presented in Shile et al. [1, Table 9 and Fig. 13] is even more obvious in Fig. 8, where the Eq. (3) and the RBF data points from Fig. 7 are completed with the points computed by (3) for $\sigma^2 = 8$ on which I superposed the RBF points for $\sigma^2 = 8$ which, as seen in [1, Fig. 13] are identical to the points given by (3). The double data points on the upper and lower branches are far away from the trend of RBF points from Fig. 7. This demonstrates that the RBF points cannot be almost identical with the points given by the second-order approximation (3). Hence [1, Fig. 13] falsifies the comparison between the RBF and the previously published results.

5. Conclusions

The article of Shile et al. [1] reproduces, with unnecessary details, the description of the benchmark for flow in saturated aquifers proposed by Alecsa et al. [2] and the principle of the RBF interpolation approach. The codes from [3,4] are further used to conduct code verification tests of the RBF solution. Such tests constitute a necessary preparatory step for applications of the numerical methods to more complex and realistic problems.

The RBF method performs better than classical discretization methods and could be a promising alternative approach in numerical simulations of groundwater flow based on Kraichnan generators of hydraulic conductivity which provide differentiable coefficients of the flow equation. The challenge is to prove that the RBF code can manage realistic spatial heterogeneity conditions which, unlike the smooth manufactured solutions used in code verification tests, require significantly

larger numbers of collocation points. Unfortunately, the results of the Monte Carlo simulations presented by Shile et al. [1] are fabricated and the comparison with similar studies, which is supposed to validate the RBF approach, is falsified.

CRediT authorship contribution statement

Nicolae Suciu: Writing – review & editing, Software, Methodology, Conceptualization.

Declaration of competing interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the link

[Comment on \[Shile et al., 2024\] \(Original data\) \(GitHub\)](#)

References

- [1] Shile F, Ben-Ahmed EH, Sadik M. Modeling groundwater flow with random hydraulic conductivity using radial basis function partition of unity method. *Eng Anal Bound Elem* 2024;163:237–50. <http://dx.doi.org/10.1016/j.enganabound.2024.03.015>.
- [2] Alecsa CD, Boros I, Frank F, Knabner P, Nechita M, Prechtel A, Rupp A, Suciu N. Numerical benchmark study for flow in heterogeneous aquifers. *Adv Water Resour* 2020;138:103558. <http://dx.doi.org/10.1016/j.advwatres.2020.103558>.
- [3] Alecsa CD, Boros I, Frank F, Nechita M, Prechtel A, Rupp A, Suciu N. FlowBenchmark. 2019, <http://dx.doi.org/10.5281/zenodo.3662401>, Git repository.
- [4] Larsson E. Radial basis function interpolation - RBF-QR. 2022, https://www.it.uu.se/research/scientific_computing/software/rbf_qr.
- [5] Bellin A, Salandin P, Rinaldo A. Simulation of dispersion in heterogeneous porous formations: Statistics, first-order theories, convergence of computations. *Water Resour Res* 1992;28(9):2211–27. <http://dx.doi.org/10.1029/92WR00578>.
- [6] Salandin P, V. Fiorotto. Solute transport in highly heterogeneous aquifers. *Water Resour Res* 1998;34(5):949–61. <http://dx.doi.org/10.1029/98WR00219>.
- [7] de Dreuzy J-R, Beaudoin A, Erhel J. Asymptotic dispersion in 2D heterogeneous porous media determined by parallel numerical simulations. *Water Resour Res* 2007;43(W10439). <http://dx.doi.org/10.1029/2006WR005394>.
- [8] Gotovac H, Cvetkovic V, R. Andricevic. Flow and travel time statistics in highly heterogeneous porous media. *Water Resour Res* 2009;45(W07402). <http://dx.doi.org/10.1029/2008WR007168>.
- [9] Suciu N. Diffusion in random fields. In: Applications to transport in groundwater. Cham: Birkhäuser; 2019, <http://dx.doi.org/10.1007/978-3-030-15081-5>.