Advances in Water Resources 111 (2018) 20–30

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Analysis and generation of groundwater concentration time series

Maria Crăciun, Călin Vamoş, Nicolae Suciu*

Tiberiu Popoviciu Institute of Numerical Analysis, Romanian Academy, Fantanele 57, Cluj-Napoca 400320, Romania

ARTICLE INFO

Keywords:
Time series analysis
Monte Carlo simulations
Global random walk
Groundwater

MSC:
37M10
65C05
60G50
86A05

ABSTRACT

Concentration time series are provided by simulated concentrations of a nonreactive solute transported in groundwater, integrated over the transverse direction of a two-dimensional computational domain and recorded at the plume center of mass. The analysis of a statistical ensemble of time series reveals subtle features that are not captured by the first two moments which characterize the approximate Gaussian distribution of the two-dimensional concentration fields. The concentration time series exhibit a complex preasymptotic behavior driven by a nonstationary trend and correlated fluctuations with time-variable amplitude. Time series with almost the same statistics are generated by successively adding to a time-dependent trend a sum of linear regression terms, accounting for correlations between fluctuations around the trend and their increments in time, and terms of an amplitude modulated autoregressive noise of order one with time-varying parameter. The algorithm generalizes mixing models used in probability density function approaches. The well-known interaction by exchange with the mean mixing model is a special case consisting of a linear regression with constant coefficients.

1. Introduction

Time series are ubiquitous in all fields of hydrology. Monitoring strategies and sampling procedures in various water management contexts, as for example in monitoring of nutrients (Geer et al., 2016) or of nitrate transport in groundwater (Turkeltaub et al., 2016), are based on analyses of concentration time series obtained from geophysical observations. The most common examples of groundwater concentration time series are the breakthrough curves obtained from concentration measurements at reference planes in aquifers. Also often used in subsurface hydrology literature are series indexed, instead of time, by a one-dimensional spatial variable (see e.g. Meerschaert et al., 2013). Examples of such series are, among others, measured data on porosity, permeability, hydraulic conductivity, electrical resistivity (see Riva et al., 2015 and references therein). The complexity of the hydrological time series drew the attention of the scientific community long time ago. The analysis of long-run hydrological and other geophysical time series (Mandelbrot and Wallis, 1968, 1969) lead to new concepts extending the classical Gaussian models, such as Gaussian fractional noise and fractional Brownian motion (Mandelbrot and Van Ness, 1968). During the last two decades, such mathematical objects were used in subsurface hydrology, notably to model the complex structure of the hydraulic conductivity (Liu et al., 2009; Meerschaert et al., 2013; Molz et al., 1997). These modeling approaches are based on both measured and synthetic time series. Synthetic time series obtained from numerical simulations can be used to investigate the behavior of some quantities which are not accessible to direct measurements. For instance, it was shown, by using an automatic algorithm to decompose time series into intrinsic components, that in some conditions the transverse component of the trajectory of the classical advection-dispersion model of transport in heterogeneous aquifers with randomly distributed hydraulic conductivity behaves as a fractional Brownian motion or even as a multifractal (Vamoş et al., 2015).

Concentration time series recorded at regular time intervals at given points of the aquifer system supply observational data for monitoring and risk assessments of groundwater quality. In the context of stochastic subsurface hydrology, the heterogeneity of the groundwater flow domain is modeled as a random environment. Consequently, predictions on the behavior of the contaminant concentration are uncertain and have to be modeled as random functions. In discrete representations, a random function is equivalent to an ensemble of random time series recorded at different points in the domain. Risk assessments are based on the probability density function (PDF) of the random concentration. Concentration PDFs can be inferred, in case of moderate heterogeneity, if one assumes a Gaussian shape of the concentration with random spatial moments estimated from the statistics of the random hydraulic conductivity (Dentz and Tartakovsky, 2010), or by assuming the shape of the concentration PDF itself (Pope, 1985). In recent years there were several attempts to apply the PDF approach from turbulence modeling (Pope, 1985, 2000) to problems of stochastic

* Corresponding author.
E-mail address: suciu@math.fau.de (N. Suciu).

http://dx.doi.org/10.1016/j.advwatres.2017.10.039
Received 7 April 2017; Received in revised form 26 October 2017; Accepted 30 October 2017
Available online 31 October 2017

subsurface hydrology (Meyer et al., 2010; Sanchez-Vila et al., 2009; Suciu et al., 2015, 2016).

The PDF approach consists of solving a PDF evolution equation, which can be derived from the advection–dispersion–reaction equation verified by the random concentration (Pope, 1985). Evolution equations with the same structure are obtained in the filtered density function (FDF) approach where, instead of ensemble averages, spatial filtering procedures are used to infer statistical quantities (Suciu et al., 2016). Such equations have to be parameterized by using closure hypotheses and models. While the transport of the concentration PDF in the physical space can be modeled by known upscaling procedures, building a "mixing model" which describes the transport of the PDF in the concentration space is still a challenging issue (Suciu et al., 2016). A possible way to search for appropriate mixing models is to use the numerical approach based on the Fokker–Planck equation describing the evolution of a suitably weighted concentration PDF (Suciu et al., 2015). In this approach, the mixing model for nonreactive transport is given by an ordinary stochastic differential equation describing the evolution of the concentration on trajectories of an advection-dispersion process. In case of reactive transport with species-independent diffusion coefficients, chemical reactions are modeled by including in the stochastic differential equation drift terms given by reaction rates (Suciu et al., 2016, Eq. (5)). In this study, we investigate the structure of an ensemble of synthetic random time series of concentration values recorded on the trajectory of the center of mass of the solute plume and we derive a stochastic process which generates time series with almost the same statistics. This process reveals the structure and provides the parameters of the mixing model which closes the PDF equation for the concentration at the plume center of mass.

Time series associated to the classical model of nonreactive transport were recently constructed by integrating over the transverse direction of a two-dimensional domain the concentration fields computed with the global random walk (GRW) algorithm (Vamoş et al., 2003) at successive longitudinal coordinates of the plume center of mass. An ensemble of such time series was computed with independent realizations of the random velocity field and a PDF approach was developed to simulate the PDF of the random concentration (Suciu et al., 2015, 2016). The mechanism responsible for the transport of the concentration PDF in concentration space is, in this case, the stochastic process which generates the ensemble of concentration time series. A simple mixing model inferred from the ensemble of simulated time series generates time series by summing up increments of a time-dependent trend and Gaussian white noise terms with decaying amplitude (Suciu et al., 2016). Though this PDF approach provides results close to reference Monte Carlo (MC) simulations at early times, its performance deteriorates at larger times. This can be attributed to the inability of the mixing model to reduce the spreading of the realizations of the process generating the time series around their ensemble average, necessary to produce the asymptotical narrowing of the concentration PDF shown by MC simulations.

Keeping in mind the utility of the concentration time series in designing mixing models for PDF approaches, we look for a stochastic model for the concentration time series that can be used to generate statistical ensembles having similar features with the initial ensemble of time series obtained from GRW-MC simulations.

In this paper, the simple model previously used in Suciu et al. (2016) is refined and modified by considering more complex time series increments. The correct asymptotic behavior is ensured by a linear dependence between distances of time series realizations to their ensemble average and their increments in time. In this way, the larger the distances from the ensemble average, the stronger the increments tend to reduce them. The residuals obtained after removing the linear regression are correlated and, as a first approximation, we model them with an amplitude modulated autoregressive process of order one (AR(1)) with time-varying parameter. Our modeling approach uses methods of time series theory (Brockwell and Davis, 1987; Hamilton, 1994), where statistical inferences are obtained from individual time series, as well as methods of statistical physics, based on statistical ensembles (Landau and Lifshitz, 1984).

The paper is organized as follows. In Section 2 the statistical ensemble obtained by GRW-MC simulations is described. Section 3 is dedicated to the analysis and modeling of the ensemble average of the time series and of the increments of the centered time series obtained by subtracting the ensemble average from each time series. In Section 4 we introduce the linear regression between the centered time series and their increments and we infer the regression coefficients. The demodulated regression residuals are further modeled as an AR(1) process with time-varying parameter in Section 5. In the last section we present some conclusions and discuss the implications of the new time series model for mixing models in PDF methods. The influence of sampling time, spatial sampling domain, and hydraulic conductivity on the structure of the time series model is analyzed in Appendix A.

2. Statistical ensemble of concentration time series

We consider the two-dimensional problem for nonreactive transport in saturated aquifers, previously used in numerical investigations based on GRW simulations (Suciu, 2014; Suciu et al., 2006, 2009) and PDF approaches (Suciu et al., 2015, 2016). The nonreactive transport was modeled as an advection–dispersion process with a constant isotropic local dispersion coefficient $D = 0.01\ \mathrm{m}^2/\mathrm{d}$ and advection velocity fields given by realizations of a random space function. For a log-normally distributed hydraulic conductivity field K with small variance $\sigma^2_{lnK} = 0.1$ and isotropic Gaussian correlation with correlation length set to $\lambda = 1$ m, the velocity realizations were generated numerically as superpositions of $N_p = 6400$ random periodic modes with the Kraichnan routine (Kraichnan, 1970) by the usual approximation for small variances of $\ln K$ (see details in Suciu et al., 2016, Appendix A). The mean velocity field with an amplitude $U = 1$ m/d was aligned with the x-axis.

The simulations were conducted over 4000 time steps $\delta t = 0.5$ d, which correspond to 2000 days or equivalently, to 2000 advection time scales λ/U. The computational domain was a rectangle with dimensions of 750 m in the longitudinal direction and 300 m in the transversal direction. A constant spatial step of 0.1 m was considered in both directions, so that the GRW lattice contained 22.5 millions of nodes. The computational domain was larger than the maximum extension of the plume and, every time when groups of particles reached the outflow boundary, the domain was displaced in the direction of the mean flow, so that it contained the entire plume at all times. Therefore, with this numerical setting no boundary conditions were necessary. The initial condition consisted of an instantaneous injection of $N = 10^{24}$ particles, uniformly distributed in a transverse slab of 1 m × 100 m. Details on the numerical implementation of the GRW algorithm can be found in Suciu et al. (2006). The cross-section concentration recorded at the x-coordinate of the expected center of mass of the solute plume, $x = Ut$, was obtained by summing the number of particles $n(x, y, t)$ over transverse slabs $\Delta x \times L_y$, where $\Delta x = 1$ m and L_y is the transverse dimension of the two-dimensional domain. For each simulation, one obtains in this way a time series

$$C(t) = C(Ut, t) = \frac{1}{N\Delta x L_y}\int_0^{L_y}\int_{Ut-\Delta x/2}^{Ut+\Delta x/2} n(x', y, t)dx'dy. \tag{1}$$

The concentrations were sampled at intervals $2\delta t$, which correspond to 1 day. A statistical ensemble $\{C^{(s)}(t)\}$, $t = 1, 2, ..., T, s = 1, 2, ..., S$ of time series of length $T = 2000$ was obtained by repeating the simulations for $S = 1000$ independent realizations of the velocity field (see Fig. 1(a)).

In Suciu et al. (2006, Appendix B2) it has been shown that the large values of the parameters S, N, and N_p used in our GRW-MC simulations ensure the reliability of the statistical inferences obtained by averaging over the ensemble $\{C^{(s)}(t)\}$. The simplified advection–dispersion model (constant, isotropic local dispersion coefficient and isotropic hydraulic