Записки Ленинградского ордена Ленина, ордена Октябрьской Революции
и ордена Трудового Красного Знамени горного института им. Г.В.Плеханова,
т. LXXI, 1980 г.

УДК 556.36

О МЕТОДАХ ГЕНЕРИРОВАНИЯ СИСТЕМ НЕЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН В ИМИТАЦИОННОМ МОДЕЛИРОВАНИИ ДЛЯ РЕШЕНИЯ ЗАДАЧ ОХРАНЫ ПОДЗЕМНЫХ ВОД

А.В.ПОСПЕЛОВ

Исследование проблемы охраны и рационального использования подземных вод в горнодобывающих районах связано с необходимостью количественного анализа сложных механико-математических моделей, которые объединяют в себе многообразную информацию, полученную при разведке и освоении месторождений. Переработка этой информации с точки зрения требований гидрогеологической схематизации, а также поиски эффективных путей повышения полноты и качества исходных гидрогеологических данных становятся невозможными без имитационного моделирования изучаемых гидрогеологических систем.

Широкое применение имитационных моделей позволит не только решить сложные конкретные проблемы, но (и это, пожалуй, наиболее важно) сделать также возможным исследование еще не решенных проблем гидрогеологии; одной из важнейших здесь может считаться проблема обоснования целесообразных объемов опытных полевых работ при разведке месторождений полезных ископаемых. Общий подход к ее исследованию с помощью имитационных моделей изложен в работе [3].

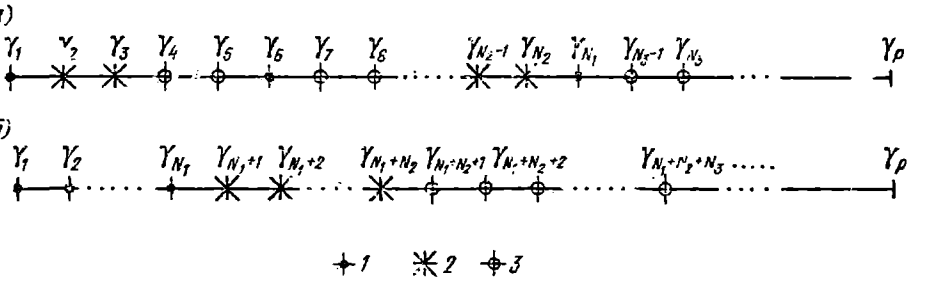
Первостепенное значение имеет внимательный учет фильтрационной неоднородности изучаемых водоносных пластов: ею определяется и соотношение объемов опытных работ разных видов, и целесообразная методика исследований, и последующий выбор расчетных параметров по совокупности локальных определений. Применительно к изучению процессов загрязнения подземных вод влияние фильтрационной неоднородности является особенно важным, так как при этом часто оказывается недопустимым осредняющий подход, общепринятый при анализе фильтрационных процессов.

Проблема моделирования неоднородных фильтрационных полей на имитационных моделях требует внимательного изучения. Прямолинейный статистический подход к этому вопросу привел, в конечном счете, к принципиально неверным выводам о роли неоднородности в процессах фильтрации и массопереноса. Поэтому кафедрой гидрогеологии ЛГИ начаты работы, имеющие целью разработку эффективных стохастических алгоритмов, позволяющих всесторонне учесть в имитационных моделях вероятностную природу фильтрационной

неоднородности водоносных систем[x]. Первые шаги в этом направлении требуют разработки корректной методики формирования систем независимых случайных величин, представляющих фильтрационные параметры водоносных пластов, при условии многократного повторения решения на имитационной модели.

При реализации машинных имитационных экспериментов часто требуется одновременно моделировать несколько независимых случайных величин с заданными законами распределения вероятностей. Наиболее естественным и очевидным способом решения этой задачи является получение каждой случайной величины с помощью отдельного базового датчика псевдослучайных чисел [5, 6], по которым обычно понимается программный датчик, порождающий последовательность $\Gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_n\}$ псевдослучайных чисел, равномерно распределенных на отрезке $[0, 1]$. Случайные величины с другими законами распределения получают, как правило, путем преобразования элементов базовой последовательности Γ [1, 2, 4, 5, 6].

С точки зрения реализации машинного имитационного эксперимента, гораздо удобнее иметь в программе один базовый датчик [2, 4, 5, 6]. Тогда "независимые" случайные величины $\xi_1, \xi_2 \ldots \xi_K$ моделируются с помощью последовательностей $\Gamma_1, \Gamma_2, \ldots, \Gamma_K$, являющихся подпоследовательностями Γ ($\Gamma_k < \Gamma \quad k = 1,2,\ldots,K$).



Последовательности, образованные по методу произвольной (а) и интервальной (б) выборки
1-3 – элементы Γ, образующие Γ_1 (1), Γ_2 (2) и Γ_3 (3)

В настоящее время в силу простоты реализации для формирования $\Gamma_1, \Gamma_2, \ldots, \Gamma_K$, предназначенных для моделирования ξ по алгоритму $\xi = \varphi(\Gamma_1, \Gamma_2, \ldots, \Gamma_K)$, используется только метод произвольных последовательностей (или метод произвольной выборки). Очень часто этот метод применяется и для формирования $\Gamma_1, \Gamma_2, \ldots, \Gamma_K$, предназначенных для моделирования системы независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_K$ (см. рисунок, а).

---

[x] Аналогичные алгоритмы потребуются и при решении многих других задач имитационного моделирования, связанных с изучением влияния различных природных и технических факторов на качество гидрогеологической информации.

Предположим, что Γ предназначена для моделирования случайных величин ξ_1, равномерно распределенных на отрезке $[a,b]$ Γ_2 – для моделирования нормально распределенных случайных величин ξ_2, по методу Марсальи – Брея [4], Γ_3 – для моделирования случайных величин ξ_3, отвечающей распределению Вейбулла. Моделирование вейбулловского распределения обычно производят по алгоритму $\xi_3 = f(\eta_1, \eta_2)$, где η_1 и η_2 – независимые, нормально распределенные случайные величины.

Очевидно, что применение метода произвольной выборки в имитационном моделировании, включающем в себя неоднократное повторение решения, практически основывается на априорном предположении о репрезентативности любой совокупности подпоследовательностей $\Gamma_1, \Gamma_2, \ldots, \Gamma_K$, состоящей из каких угодно элементов Γ. Хорошо известно, что практически любому базовому датчику в той или иной степени присущи отклонения от случайности по некоторым его подпоследовательностям Γ_K [1, 5]. Поэтому ни один датчик псевдослучайных чисел не может удовлетворить всем приложениям сразу. И пригодность того или иного датчика для решения конкретного класса задач должна устанавливаться на основе специализированных тестов. В пользу разработки и применения таких тестов свидетельствует и то, что для установления локальных отклонений подпоследовательностей $\Gamma_1, \Gamma_2, \ldots, \Gamma_K$, образованных по методу произвольной выборки, нет соответствующих стандартных тестов.

Покажем на основе специализированного теста, представляющего собой аналог некоторых задач, встречающихся в имитационном моделировании гидрогеологических процессов, что ни один из рассмотренных нами базовых датчиков не удовлетворит приложениям, основанным на методе произвольной выборки.

Построим тест-аналог по следующей схеме. Предположим, что имеется водоносный пласт, водопроводимость которого T требуется определить с помощью опробования. Допустим, что искомый параметр опробуемого объекта отвечает модели случайной величины, и результатом опробования является совокупность значений параметра, которая может рассматриваться как репрезентативная выборка. Целью опробования является определение некоторого среднего значения параметра T_{CP}[x].

Допустим, что опробование заключается в проведении нескольких одиночных откачек, совокупность которых дает вполне определенного объема выборку, используемую в дальнейшем для вычисления T_{CP}. Предположим, что с целью установления влияния на величину T_{CP} количества одиночных откачек (или что то же самое – объема выборки), было произведено I опробований, каждое из которых дало выборку объема N_i ($i = 1,2,\ldots,I$), т.е. при первом опробовании было сделано N_1 одиночных откачек, при втором – N_2 и т.д.

---

[x] Если T есть нормально распределенная случайная величина, то таким параметром будет среднее арифметическое, в случае логарифмически нормального распределения – среднее геометрическое, при вейбулловском распределении – среднее гармоническое. В некоторых случаях, например, при расчете водозаборов, соответствующим образом вычисленный средний параметр может являться и эффективным параметром [3].

Затем все I опробований были воспроизведены на J ($j = 1,2,\ldots,J$) участках исследуемого объекта S, что позволяет установить влияние расположения скважин на величину $T_{ср}$.

Далее, для получения $T_{ср}$, не искаженного случайными отклонениями единичного опробования, все I опробований на всех J участках были проделаны по K раз. Другими словами, на всех уровнях факторов N и S было произведено по K параллельных опытов.

Будем считать, что T есть нормально распределенная случайная величина с математическим ожиданием μ и средним квадратичным отклонением σ. Как будет показано ниже, результаты машинного эксперимента, имитирующего опробование пласта в рамках сделанных предложений, не зависят от конкретных числовых значений μ и σ.

Таким образом, опробование пласта по описанной процедуре дало $T_{ср\,11k}, \ldots, T_{ср\,ijk}, \ldots, T_{ср\,IJK}$, являющихся средними арифметическими каждой ijk-й выборки, по которым требуется определить величину расхождений между значениями среднего параметра. Если расхождения окажутся значительными, то необходимо установить влияние факторов N и S.

Машинный эксперимент, имитирующий исследование объекта в рамках сделанных предположений, построим по следующей схеме. Опробование пласта имитируется получением выборок из нормально распределенной генеральной совокупности $\eta(\mu,\sigma)$, порождаемой датчиком псевдослучайных чисел. Элементами выборки являются подвергнувшиеся преобразованию члены базовой последовательности с номерами $kjn+\Delta$, где $n = 1,2,\ldots,I$; Δ – случайный аддитивный член, с равными вероятностями принимающий значения 0,1,2,3,4 и преобразующий случайную величину, равномерно распределенную на интервале [0,1] в нормально распределенную случайную величину в соответствии с методом Марсальи – Брея. Такая организация выборки позволяет реализовать метод произвольных подпоследовательностей, причем не в самом крайнем его варианте.

Опробование по различным участкам объекта S_j имитируется применением различных датчиков G_1, G_2, G_3 [x]. Датчик G_1 работает по методу "перемешивания" [7]. Датчик G_2 – конгруэнтный смешанный датчик, работающий в соответствии с алгоритмом: $X_{i+1} = (3141592653\,X_i + 2718281829)^x \bmod 2^{35}$, $\gamma_i = X_i/2^{35}$ [1]. Датчик G_3 – конгруэнтный мультипликативный датчик, работающий в соответствии с алгоритмом $X_{i+1} = (3141592653\,X_i) \bmod 2^{35}$ [1]. Датчики типа G_2 и G_3 в настоящее время являются самыми распространенными источниками псевдослучайных чисел в стохастическом моделировании.

Напомним, что на всех уровнях факторов N и G проводится по K параллельных опытов.

В рассматриваемом машинном эксперименте результатом (или наблюдением) каждого ijk-го опыта будет $T_{ср\,ijk}$, являющееся оценкой μ, вычисленной по нормальной выборке i-го объема, порождаемой j-м датчиком в k-м параллельном опыте.

---

[x] Все три датчика реализованы для ЭВМ „НАИРИ-К".

В силу центральной предельной теоремы совокупности параллельных наблюдений ($T_{cp_{ij1}}, T_{cp_{ij2}}, \ldots T_{cp_{ijk}}$) будут распределены нормально с параметрами μ_{ij} μ и σ_{ij} $\sigma/\sqrt{N_i}$ σ_i Таким образом, наблюдения на различных уровнях фактора N разнородны по дисперсии, что препятствует обработке результатов методами дисперсионного анализа.

Преобразуем $T_{cp_{ijk}}$ в стандартную нормальную форму

$$Z_{ijk} = (T_{cp_{ijk}} - \mu)/\sigma_i \qquad (1)$$

Наблюдения, подвергнувшиеся преобразованию (I), образуют выборку из генеральной совокупности, отвечающей нормальному распределению с $M(Z) =$ и $\sigma(Z) = 1$. Отклонения Z_{ijk} от математического ожидания, обусловленные конечными размерами выборки, $e_{ijk} = Z_{ijk} - M(Z)$ также будут распределены нормально с $M(e) = 0$ и $\sigma(e) \neq 0$.

Важно отметить, что в рамках этой модели соблюдены все необходимые предпосылки дисперсионного анализа, что в итоге позволит сделать весьма надежные выводы.

Покажем, что Z_{ijk} не зависят от μ и σ случайных величин T. Как следует из выражения (I),

$$Z_{ijk} = (T_{cp_{ijk}} - \mu)/\sigma_i = \left\{\left[\sum_{i=1}^{N_i}(\mu + \eta_{ijk}\sigma)\right]/N_i - \mu\right\}/(\sigma/\sqrt{N_i}) =$$

$$= \left[\mu - \mu + \sigma\left(\sum_{i=1}^{N_i}\eta_{ijk}\right)/N_i\right]/\left[\sigma/\sqrt{N_i}\right] =$$

$$= \left[\left(\sum_{i=1}^{N_i}\eta_{ijk}\right)/N_i\right]\sqrt{N_i}\ , \qquad (2)$$

где η_{ijk} – i-я реализация нормально распределенной случайной величины с $M(\eta) = 0$ и $\sigma(\eta) = 1$, полученная от j-го датчика в k-м параллельном опыте.

Нетрудно видеть, что результаты эксперимента полностью определяются свойствами рассматриваемых датчиков, способом получения выборки и ее объемом (см. таблицу).

Обработка результатов эксперимента, представленных в таблице, была произведена по схеме двухфакторного дисперсионного анализа с одинаковым количеством наблюдений в ячейках [7]. Отношения дисперсий в эксперименте по методу произвольной выборки следующие: $ES_N/ES_e = 1,4$; $ES_G/ES_e =$ 3,24 и ES_{NG}/ES_e 1,33 (здесь FS_N, ES_G – дисперсия по фактору N G соответственно; ES_e – дисперсия ошибки e; ES_{NG} – дисперсия по взаимодействию факторов N и G).

Проверим гипотезу об отсутствии влияния факторов N и G, а также их взаимодействия на уровне значимости $\alpha = 0,99$, который позволяет снести к минимуму вероятность отбрасывания правильной гипотезы, что вполне оправдано применительно к нашему эксперименту, удовлетворяющему всем предпосылкам дисперсионного анализа.

Сравнивая отношение дисперсий с критическими значениями F-критерия при соответствующих степенях свободы, нетрудно видеть, что различия

N_i	G_j			Z_i (среднее по строкам)
	G_1	G_2	G_3	
		Произвольная выборка		
$N_1 = 10$	$Z_{11} = -0,396$	$Z_{12} = -0,391$	$Z_{13} = -0,040$	$Z_{1..} = -0,276$
$N_2 = 20$	$Z_{21} = -0,219$	$Z_{22} = -0,837$	$Z_{23} = -0,193$	$Z_{2..} = -0,416$
$N_3 = 40$	$Z_{31} = 0,486$	$Z_{32} = -0,788$	$Z_{33} = -0,163$	$Z_{3..} = -0,155$
$N_4 = 70$	$Z_{41} = -0,262$	$Z_{42} = -0,0497$	$Z_{43} = -0,652$	$Z_{4..} = -0,321$
$N_5 = 100$	$Z_{51} = -0,779$	$Z_{52} = -1,179$	$Z_{53} = -0,813$	$Z_{5..} = -0,724$
(среднее по столбцам)	$Z_{.1} = -0,154$	$Z_{.2} = -0,649$	$Z_{.3} = -0,322$	$Z_{...} = -0,378$
		Интервальная выборка		
N_1 10	$Z_{11} = -0,369$	$Z_{12} = 0,265$	$Z_{13} = 0,135$	$Z_{1..}$ 0,010
$N_2 = 20$	$Z_{21} = -0,472$	$Z_{22} = 0,054$	$Z_{23} = -0,008$	$Z_{2..}$ -0,141
$N_3 = 40$	$Z_{31} = -0,025$	$Z_{32} = 0,199$	$Z_{33} = -0,148$	$Z_{3..}$ 0,0089
$N_4 = 70$	$Z_{41} = 0,208$	$Z_{42} = 0,127$	$Z_{43} = -0,023$	$Z_{4..} = 0,102$
$N_5 = 100$	$Z_{51} = -0,135$	$Z_{52} = -0,066$	$Z_{53} = -0,151$	$Z_{5..} = -0,117$
$Z_{.j.}$	-0,1595	0,1159	-0,03811	$Z_{...} = -0,0272$

Примечание. Точка вместо соответствующего индекса означает, что по этому индексу произведено осреднение, например $Z_{...} = (IJK)^{-1} \sum_{1}^{I} \sum_{1}^{J} \sum_{1}^{K} Z_{ijk}$

между Z_{ijk} незначительны. Следовательно, все полученные оценки среднего параметра либо одинаково хороши, либо одинаково плохи (с точки зрения статистики). Причем на качество Z_{ijk} не влияет ни объем выборки, ни применяемый датчик.

Корректность модели рассмотренного эксперимента позволяет все же установить, к какой категории следует отнести полученные результаты. Все средние (см. таблицу) являются реализациями нормально распределенных случайных величин с математическими ожиданиями, равными нулю и соответствующими средними квадратичными отклонениями, которые нетрудно установить, пользуясь центральной предельной теоремой. Так $Z_{...}$ есть реализация случайной величины, отвечающей нормальному распределению с $\sigma = I/\sqrt{IJK} =$ $= 0,08$. Среднее квадратичное отклонение случайной величины $\sigma(Z_{i..}) =$ $= I/\sqrt{JK} = 0,18$; $\sigma(Z_{.j.} = I/\sqrt{IK} = 0,133$. Итак, $Z_{...}$, а также $Z_{5..}$ и $Z_{.2.}$ выходят за пределы соответствующих 3σ, что свидетельствует о систематическом отклонении от среднего.

Каждая отдельная выборка верно отражает свойство генеральной совокупности, т.е. является репрезентативной (см. таблицу). Однако все выборки в целом дают недопустимо большое отклонение от среднего в меньшую сторону, что в случае многократных прогонов имитационной модели приводит к систематическому занижению ее параметров, а следовательно, и к искажению выхода модели.

Относительная средняя ошибка при многократном повторении решения на имитационной модели (в процентах)

$$\delta = m\frac{\sigma_i}{\mu}100 \qquad (3)$$

где $m = Z\ldots$

Естественной альтернативой методу произвольной выборки является метод интервальной выборки, также позволяющий моделировать систему независимых случайных величин с помощью одного базового датчика псевдослучайных чисел (см. рисунок, б).

Подпоследовательности $\Gamma_1, \Gamma_2, \ldots, \Gamma_K$ в случае применения метода интервальной выборки являются непересекающимися интервалами базовой последовательности. Реализация данного метода более сложна, однако это полностью компенсируется отсутствием систематических отклонений от среднего. Этот факт нетрудно установить с помощью рассмотренного выше теста-аналога, который был произведен по схеме, примененной для проверки произвольных подпоследовательностей (см. таблицу). Отношения дисперсий при интервальной выборке следующие: ES_N / ES_e 0,287; $ES_e / ES_\cdot = 0{,}893$ и $ES_{N\Phi} / ES_e = 0{,}349$.

Сравнивая отношения дисперсий с критическими значениями F-критерия при $\alpha = 0{,}99$, нетрудно видеть, что расхождения между средними незначимы. Следовательно, влияние факторов N и G на результаты эксперимента отсутствуют, как и в предыдущем случае. Однако, в отличие от него, совокупность интервальных выборок не дает систематических отклонений от среднего, что легко проверить, применяя к полученным оценкам правило 3σ.

Выводы

1. Широко распространенный в имитационном моделировании метод произвольной выборки дает систематическую погрешность при многократных повторениях решения на модели. Относительную ошибку, обусловленную применением этого метода, можно вычислить по формуле (3).

2. В качестве основного метода формирования системы независимых случайных величин следует использовать способ интервальной выборки, который, в отличие от метода произвольной выборки, не дает систематических отклонений и, следовательно, не искажает результата моделирования.

3. Полученные результаты, в случае применения датчиков, подобных базовым датчикам, рассмотренным в данной работе, могут быть распространены на имитационные модели для других двоичных ЭВМ, например ЕС ЭВМ.

ЛИТЕРАТУРА

Кнут [illegible] усство программирования для ЭВМ. Т. 2, М., Мир, 1977. 724 с.

2. Клейнен Д. Статистические методы в имитационном моделировании. М., Мир, 1978. Т. 1. 222 с.

3. Мирошенко В.А., Шестаков В.М. Теория и методы интерпретации опытно-фильтрационных работ. М., Недра, 1976. 326 с.

4. Нейлор Т. Машинные имитационные эксперименты с моделями экономических систем. М., Мир, 1975. 500 с.

5. Соболь И.М. Численные методы Монте-Карло. М., Наука, 1975. 311 с.

6. Шеннон Р. Имитационное моделирование систем - искусство и наука. М., Мир, 1975. 418 с.

7. Шеффе Г. Дисперсионный анализ. М., Гос. изд-во физ.-мат. лит-ры, 1963. 625 с.