

ПРОБЛЕМА ВАЛИДАЦИИ ИМИТАЦИОННОЙ МОДЕЛИ И ЕЕ ВОЗМОЖНЫЕ РЕШЕНИЯ

И. В. Яцкив (Рига)

Введение

Процесс проверки адекватности имитационной модели имеет сегодня в англоязычной литературе различные обозначения: VV&A, VV&T, VV&C. Остановимся на последнем обозначении VV&C, которое дословно расшифровывается: верификация, валидация и заслуживающая доверие. *Верификация* – процесс контроля корректности трансляции концептуальной модели в имитационную программу. *Валидация* – процесс проверки является ли модель, допустимым представлением реальной системы, основываясь на целевой направленности модели (Fishman&Kiviat (1968)), т. е. процесс, который называется в русской литературе по моделированию, – проверка адекватности. Если модель и ее результаты приняты пользователем и используются для принятия решений, то модель считается *заслуживающей доверие (credible)* (Carson (1986)). Многочисленные имитационные программы облегчают сегодня процесс программирования и верификации, а вот успехи валидации не столь очевидны. Еще не редко в настоящее время сложные имитационные модели вообще не валидируются.

В докладе рассмотрены основные технологии проверки адекватности модели. Особый акцент делается на статистические технологии при валидации результатов. Симуляция, прежде всего, означает экспериментирование с моделью, а любое экспериментирование влечет статистический анализ. Такой анализ является лишь частью валидационного процесса, но его преимущество в наибольшей формализации.

1. Обзор литературы

В качестве монографий, включающих обширные разделы по данному вопросу, выделим Banks, Carson, Nelson, Nicol (2000), Law&Kelton (2000), Kleijnen&Groenendaal (1992). Среди множества статей на эту тему отметим:

- Sargent (1984a), (1984b) – обсуждаются вопросы анализа чувствительности;
- Sargent (1996) – дается суммарный обзор технологий проверки адекватности;
- Kleijnen (1995a) – обсуждаются основные технологии верификации и валидации;
- Kleijnen (1995b) – рассматривается технология анализа чувствительности;
- Kleijnen (1996) – впервые представляется novel тест;
- Kleijnen (1999) – рассматриваются варианты статистических технологий в зависимости от имеющихся данных по реальной системе;
- Fosset (1991) – представляются 14 факторов, позволяющих признать модель заслуживающей доверие;
- Balci (1994) – обзорная статья, включающая 102 ссылки на литературу по валидации модели.

Одна из ранних работ по этой проблеме (Naylor and Finger (1967)) представляет *трехэтапный подход к проверке адекватности* имитационной модели. На первом этапе (валидация внешнего представления) проверяется насколько модель выглядит адекватной с точки зрения специалистов, которые с ней будут работать. В процессе этапа требуется постоянный контакт с заказчиком модели, дискуссии с экспертами по системе. На втором этапе (эмпирическое тестирование допущений модели) осуществляется графическое представление данных, проверка гипотез о распределениях, анализ чувствительности и др. Третий этап – статистические сравнения между откликами реальной системы и модели. Методика применения статистических технологий зависит от доступности данных по реальной системе.

Другой вариант классификации процедур проверки адекватности предусматривает три группы тестов: концептуальные, операционные и основанные на выходных данных. Концептуальные тесты проводятся на этапе построения концептуальной модели и включают проверку: постановки задачи моделирования, входных данных, структурных допущений, логическую валидацию (Советов&Яковлев (2001)). Операционные тесты (тесты на функционирование) анализируют адекватность поведения модели, включают тесты на непрерывность, анализ чувствительности, анализ вырождения, анализ анимации. Тесты, основанные на выходах, могут быть применены при наличии реальных выходных данных.

Рассмотрим применение статистических технологий для анализа результатов в зависимости от доступности данных: реальные данные недоступны; реальные выходные данные доступны (без соответствующих входов); доступны реальные выходные данные с соответствующими входами или хорошо известной трассировкой.

2. Отсутствие реальных данных

Если данные по реальной системе отсутствуют, то полноценную валидацию провести нельзя. Однако могут быть использованы экспертные знания. Применяется анализ чувствительности, включающий проектирование эксперимента и регрессионные метамоделли. В литературе нет четкого и единого определения термина «анализ чувствительности». Kleijnen (1999) определяет его как систематическое исследование реакции откликов моделирования на экстремальные значения входов модели или на кардинальные изменения в структуре модели. Например, что случится со средним временем ожидания клиентов, когда изменится интенсивность их поступления (приоритет обслуживания)?

Анализ чувствительности, прежде всего, выявляет значимые факторы в системе. Если возможно, должна быть собрана информация относительно этих факторов. Если существенные факторы управляемы пользователями, то анализ чувствительности показывает, как изменить (заменить) эти факторы, чтобы оптимизировать реальную систему. Разработано несколько вариантов технологии проведения анализа чувствительности (Fossett (1991)). Анализ чувствительности модели требует выполнения набора прогонов, при этом факторы в течение прогона моделирования остаются постоянными и изменяются от прогона к прогону. Каждый фактор имеет, по крайней мере, два уровня в эксперименте. Фактор может быть качественным (например, приоритетные правила).

Центральная проблема проектирования эксперимента – выбор ограниченного набора комбинаций факторных уровней, которые будут фактически моделироваться при экспериментировании с моделью (Хинчин (1997), (Советов&Яковлев (2001))). Популярный тип проекта – 2^{k-p} проект: k факторов изменяются в эксперименте; каждый фактор имеет два уровня; только доля (а именно 2^{-p} где $p = 0, 1, \dots$) из комбинаций 2^k фактически моделируется. После выбора комбинаций факторных уровней выполняется имитационная программа. Затем данные (вход/выход) эксперимента анализируются на основе дисперсионного или регрессионного анализа. Производится собственно анализ чувствительности (выявляются статистически значимые факторы). В области моделирования такие регрессионные модели называются метамоделями (модель поведения ввода–вывода имитационной модели) (Friedman (1996), Kleijnen (1999)). Как правило, метамоделли использует полином аппроксимации без и со взаимодействиями между факторами. Для выбора степени полинома аппроксимации и проверки корректности метамоделли используют множественный коэффициент корреляции R^2 или усовершенствованные процедуры выбора (на основе технологии последовательного проектирова-

ния эксперимента с перекрестной проверкой и F-критерием) (Kleijnen, Cheng, Feelders (1998)).

Недостаток проектирования эксперимента с метамоделями состоит в том, что имитационная модель рассматривается как черный ящик: наблюдаются вход-выход модели и оцениваются эффекты факторов в метамоделях, но не используется знание структуры модели. Преимущество – в том, что может применяться ко всем имитационным моделям: детерминированным или стохастическим, с дискретными событиями или непрерывными.

С анализом чувствительности связан анализ риска, который также требует выполнения имитационной модели для различных комбинаций факторных уровней. Основное различие между анализом чувствительности и анализом риска в том, что последний пытается ответить на вопрос: какова вероятность редкого события (бедствия)? Это может быть ядерный инцидент, экологический крах, финансовая потеря инвестиции и т. д. При анализе чувствительности рассматривались повторяющиеся события (ожидание клиента). Следовательно, валидация при анализе риска значительно сложнее. При анализе риска, на основе применения некоторых статистических методов пробуют улучшить допустимость рассматриваемой модели (Fossett, Harrison, Weintrob, Gass (1991)). Например, для обнаружения факторов, имеющих существенные эффекты, применяют регрессионный анализ; затем, используя экспертов, пробуют объяснить важность этих факторов.

3. Наличие реальных выходных данных

Если моделируемая система существует и доступны выходные данные, то для сравнения измерений на реальной системе и результатов экспериментов можно применить различного рода классические тесты: хи-квадрат, Колмогорова-Смирнова и др. Однако ряды значений отклика модели и системы почти всегда нестационарны и с автокорреляцией. Поэтому применение данных тестов не всегда корректно.

Рассмотрим для примера применение классической статистики Стьюдента для двух выборок. Предположим, что реальный отклик x – 90% квантиль интересующей характеристики w_t , измеренной за день в реальной системе. Смоделированный отклик y – 90% квантиль величины v_t за день в модели. Пусть наблюдаются n дней в системе и моделируются m дней. Предположим, что каждый реальный и моделируемый день дает независимое и тождественно распределенное наблюдение (отсутствует сезонность).

Идеальная имитационная модель имеет эмпирическую функцию распределения для отклика F_y идентичную распределению для реальной системы F_x . На практике заказчик модели часто заинтересован не в распределении F_x , а в частных характеристиках, например, в среднем $E(x) = \mu_x$. Пусть, 90% квантиль изменяется каждый день, и за критерий управления принято его ожидаемое значение.

Определим различие в средних $\mu_d = \mu_x - \mu_y$. Используя n и m наблюдений на реальной и моделируемой системе, получим классические оценки \bar{x}, \bar{y}, s_x^2 и s_y^2 средних и дисперсий x и y . Тогда статистика Стьюдента с $(n + m - 2)$ степенями свободы равна:

$$t_{n+m-2} = \frac{(\bar{x} - \bar{y}) - \mu_d}{\left[\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2} \right]^{\frac{1}{2}}} \cdot \frac{[(n+m-2)nm]^{\frac{1}{2}}}{(n+m)^{\frac{1}{2}}}. \quad (1)$$

Нулевая гипотеза состоит в том, что различие в средних равно нулю $H_0: \mu_d = 0$. Мощность этого критерия увеличивается, если в уравнении (1) разность $|\mu_d|$, увеличивается (большие различия проще обнаружить); n или m увеличиваются; σ_x или σ_y уменьшаются. Могут иметь место ошибки II рода, если моделируются только несколько дней или присутствует много помех: значимая разность может быть не обнаружена (незначимый t). Возможна также ошибка I рода – при большом количестве доступных данных даже незначимое различие между моделируемым и реальным откликом может давать значимый критерий t .

При допущении о независимости и одинаковом нормальном распределении отклика тест Стьюдента является все-таки не очень чувствительным к типу распределения. Но модели могут давать результаты (например, оценки квантилей), существенно отличающиеся от нормального закона. Тогда альтернативными к нему могут быть:

- модифицированная статистика Стьюдента, включающая оценку асимметрии распределения выхода (Johnson (1978));
- класс свободных от распределения тестов (например, Уилкоксона) (Айвазян (2000));
- jackknifing – устойчивая процедура, требующая больше компьютерного времени для анализа результатов экспериментов на модели (Эфрон (1988));
- bootstrapping – разновидность метода Монте-Карло (Efron&Tibshirani (1993)).

4. Доступны реальные данные по входу-выходу системы

При наличии помимо реальных откликов системы еще и входов может быть применен подход, который называется *trace-driven* моделирование Kleijnen (1996). *Trace-driven* моделирование означает, что на вход имитационной программы в исторической последовательности подаются реальные входные данные. После выполнения имитационной программы, сравнивают временной ряд результатов экспериментов с моделью с историческим временным рядом реальных выходов системы. Проблема опять состоит в процедуре сравнения (обе системы изучаются при похожих сценариях). Могут быть использованы те же классические тесты, что в и пункте 3. Рассмотрим процедуры, специально разработанные для валидации имитационных моделей.

- Тест Туринга (Schruben-Turing). Основная идея в перемешивании одинакового количества выходов реальной системы и модели и демонстрация их эксперту для определения разницы между ними. Полностью зависит от уровня эксперта.
- Тест по прогнозу поведения системы. Основываясь на откликах модели, построить прогноз реакции системы на изменение ее структуры. Затем провести изменение структуры на реальной системе, измерить отклик и сравнить с прогнозом. Идеальный тест, но трудный в реализации.
- Наивный (naïve) тест адекватности. Для x и y – реальных и моделируемых откликов при одних и тех же входных данных подберем линию регрессии $y = \beta_0 + \beta_1 x$. Задача состоит в проверке гипотезы: $\beta_0 = 0$ и $\beta_1 = 1$. Однако этот тест имеет тенденцию слишком часто отвергать адекватную модель. Предположим, что имитационная модель адекватна, а именно: реальный и моделируемый выходы имеют одинаковые средние $\mu_x = \mu_y (= \mu)$ и дисперсию $\sigma_x^2 = \sigma_y^2 (= \sigma^2)$. Предположим, что среднее положительно ($\mu > 0$) – типично при моделировании систем с очередями – и что имитационная модель не идеальна ($\rho_{xy} < 1$). Для регрессионной модели $y = \beta_0 + \beta_1 x$

имеем $\beta_1 = \frac{\beta_{xy} \epsilon_y}{\epsilon_x}$ и $\beta_0 = \epsilon_y - \beta_1 \epsilon_x$. Следовательно, адекватная имитационная модель приводит к следующим соотношениям для оценок $0 < \beta_1 < 1$ и $0 < \beta_0 < \mu$. Тогда при проверке условия $\beta_1 = 1$ и $\beta_0 = 0$ вероятно отклонение адекватной модели.

- Novel тест адекватности описан Kleijnen, Bettonvil, Groenendaal (1996,1998). Вычисляется не только n разностей d_i , ($n = m$), но и n сумм $q_i = x_i + y_i$. Затем подбирается линия $d = \gamma_0 + \gamma_1 q$ к этим n парам (d_i, q_i) и формулируется нулевая гипотеза $H_0 : \gamma_0 = 0$ и $\gamma_1 = 0$. Очевидно, эта гипотеза влечет $\mu_d = 0$ или $\mu_x = \mu_y$. Принимая нормальность для x и y , просто доказать, что $\gamma_1 = 0$ влечет равенство дисперсий: $\sigma_x^2 = \sigma_y^2$. Для проверки гипотезы, можно использовать стандартную программу регрессионного анализа (F-критерий). Kleijnen (1998), применяя оба теста к системе M/M/1, делает следующие выводы:
 - *naive* тест отклоняет адекватную имитационную модель чаще, чем *novel* тест;
 - *naive* тест показывает «ошибочное» поведение в определенной области; то есть чем хуже имитационная модель в этой области, тем выше вероятность ее принятия;
 - *novel* тест не отклоняет верную модель слишком часто, отклоняет с вероятностью α , если выводы преобразованы логарифмически (для реализации нормальности).

И *naive*, и *novel* тесты предполагают, что реальный и моделируемый выходы независимо и одинаково распределены (по нормальному закону).

- Подход, основанный на доверительном интервале. Предположим, что наблюдаются в реальной системе n дней и моделируются m дней. Обозначим X_1, X_2, \dots, X_n - средние значения интересующего показателя за n дней и Y_1, Y_2, \dots, Y_m - соответственно за m дней имитационной модели. Пусть $\mu_x = E(X_j)$ и $\mu_y = E(Y_j)$. Задача сводится к построению доверительного интервала для $\xi = \mu_x - \mu_y$.

Возьмем $n = m$ и $Z_j = X_j - Y_j$ для $j = 1, \dots, n$. $\{Z_j\}$ - последовательность независимых одинаково распределенных величин. Оценки среднего и дисперсии равны

$$\bar{Z}(n) = \frac{\sum_{j=1}^n Z_j}{n} \quad \text{и} \quad \text{Var}[\bar{Z}(n)] = \frac{\sum_{j=1}^n [Z_j - \bar{Z}(n)]^2}{n(n-1)}.$$

Тогда доверительный $100(1-\alpha)\%$ ин-

тервал равен $[\bar{Z}(n) \pm t_{n-1, 1-\alpha/2} \sqrt{\text{Var}[\bar{Z}(n)]}]$. Если интервал не включает ноль, то наблюдаемая разница между μ_x и μ_y статистически значима на уровне α . Считается наиболее надежным подходом, но для применения требует большого числа данных.

- Подход, не требующий нормальности распределения. Проверка адекватности модели с ненормальными откликами может быть успешно реализована на базе бутстреп-технологии (Efron&Tibshirani (1993)). Kleijnen (1996) продемонстрировал, что лучший тест - бутстрепирование разности между средними моделируемых и реальных откликов - дает корректную вероятность ошибки I рода и имеет высокую мощность. Andronov (2000) рассматривает следующий вариант применения технологии бутстрепирования. Пусть имеется исторический набор входов, который многократно подается на вход модели, по выходным данным модели методом процен-

тилей строится доверительный интервал. Если отклик реальной системы попадает в интервал, то делается вывод об адекватности модели.

Заключение

Проверка адекватности модели включает много процедур, применение которых, прежде всего, зависит от того существует ли моделируемая система или нет, доступны ли данные по ней и какие. Статистические процедуры, применяемые в каждом из вариантов, используют методы регрессионного и дисперсионного анализа, критерии, зависящие и свободные от распределений, анализ чувствительности, методы компьютерной интенсивной статистики. Получение с помощью имитационного моделирования результатов, вызывающих доверие у заказчиков модели, во многом зависит от корректности применения данных статистических процедур на этапе валидации модели.

Литература

1. **Andronov A.M.** (2000) Efficiency Analysis of Queueing Model Validation by Use of Trace-Driven Simulation. In *Abstracts of Memorial Seminar Dedicated to the 60th Birthday of Vladimir Kalashnikov*. September 7–11, 2002, Petrozavodsk, Russia. – 3 pp.
2. **Carson J.S.** (1986) Convincing Users of Model's Validity Is Challenging Aspect of Modeler's Job, *Ind.Eng.*, 18: 74–85.
3. **Balci O.** (1994) Validation, Verification and Testing Techniques Throughout the Life Cycle of a Simulation Study, *Annals of Operation Research*.
4. **Banks J., Carson J.S., Nelson B.L., Nicol D.** (2000) *Discrete-Event System Simulation*. Prentice Hall, New Jersey.
5. **Banks J., Carson J.S., Nelson B.L., Nicol D.** (2000) *Discrete-Event System Simulation*. Prentice Hall, New Jersey.
6. **Efron B., Tibshirani R.** (1993) *Introduction to the Bootstrap*. Chapman&Hall, London.
7. **Fossett C.A., Harrison D., Weintrob H., Gass A.I.** (1991) An Assessment Procedure for Simulation Models: a case study. *Operations Research* 39, pp.710–723.
8. **Friedman L.W.** (1996) *The Simulation Metamodel*. Kluwer, Dordrecht, Netherlands.
9. **Johnson N.J.** (1978) Modified t Tests and Confidence Intervals for Asymmetric Populations. *Journal of the American Statistical Association*. 73, pp.536–544.
10. **Kleijnen J.P., Groenendaal W.** (1992) *Simulation: a statistical perspective*. Wiley, N.Y.
11. **Kleijnen J.P.C.** (1995a) Statistical Validation of Simulation Models, *European Journal of Operational Research* 87, p.21–34.
12. **Kleijnen J.P.C.** (1995b) Verification and Validation of Simulation Models, *European Journal of Operational Research* 82, pp.145–162.
13. **Kleijnen J.P.C., Bettonvil B., Groenendaal W.V.** (1996) Validation of Trace-Driven Simulation Models: Regression Analysis Revisited, *Proceedings of the 1996 Winter Simulation Conference*.
14. **Kleijnen J.P.C., Cheng R.C.H., Felders A.J.** (1998) Bootstrapping and Validation of Metamodels in Simulation. *Proceedings of the 1998 Winter Simulation Conference*, pp.701–706.
15. **Kleijnen J.P.C.** (1999) Validation of models: statistical techniques and data availability.
16. **Law A.M., Kelton W.D.** (2000) *Simulation Modelling and Analysis*. McGraw-Hill.
17. **Naylor T.H., Finger J.M.** (1967) Verification of Computer Simulation Models, *Management Sci.*, 14: pp. 92–101.
18. **Sargent R.G.** (1984a) Simulation Model Validation, Simulation and Model Based Methodologies: An Integrated View, Springer-Verlag, pp.537–535.

19. **Sargent R.G.** (1984b) a Tutorial on Validation and Verification of Simulation Models, *Proceedings of the 1984 Winter Simulation Conference*, pp.115–121.
20. **Sargent R.G.** (1996) Verifying and Validating Simulation Models, *Proceedings of the 1996 Winter Simulation Conference*, pp.133–141.
21. **Айвазян С.А., Мхитарян В.С.** (2000) Прикладная статистика и основы эконометрики. М:ЮНИТИ.
22. **Советов Б.Я., Яковлев С.А.** (2001) Моделирование систем. М.: Высшая школа.
23. **Хинчин И.В.** (1997). Принципы статистического моделирования. Минск: БГУ.
24. **Эфрон Б.** (1988) Нетрадиционные методы многомерного статистического анализа. М: Финансы и статистика.