

# Применение квазислучайных множеств при планировании многокомпонентного градуировочного эксперимента

А.С. Мананков

Самарский государственный технический университет, Самара, Россия

**Обоснование.** Планирование эксперимента является междисциплинарным разделом науки. Его основной задачей является достижение поставленных целей исследования минимальным числом измерений и образцов, а значит, экономия средств и времени.

В химическом спектральном анализе часто исследуются сложные смеси, включающие два и более компонента. На основании спектральных данных строится градуировочная модель, с помощью которой можно количественно определять концентрации веществ в образце. Модель строится на основании некой обучающей выборки (набора). Имея обученные модели, можно определить концентрации всех анализируемых компонентов из единичного спектра. Можно выделить три основных подхода к формированию градуировочного набора образцов: ручной, систематический и случайный.

Для создания градуировочного набора образцов нужны четкие критерии для его оценки. Критерии оценки качества обучающего набора для создания нескольких градуировочных моделей на одних и тех же спектральных данных ранее не были четко сформулированы. Как правило, принимают во внимание только отсутствие корреляции между концентрациями компонентов. Однако этого явно недостаточно, и в работе [1] была показана важность еще двух параметров: полноты и равномерности заполнения пространства эксперимента (ПЭ). Под ПЭ понимается квадрат, куб или гиперкуб, образованный диапазонами, огранивающими концентрации компонентов (факторов). Набор образцов, таким образом, представляется координатами точек в данном пространстве. Вручную выбрать оптимальный обучающий набор градуировочных образцов трудно даже для двухкомпонентной смеси, тем более для трех и более компонентов. Для этого используют специальные алгоритмы.

Для оценки качества обучающих наборов и их сравнения между собой требуются численные критерии. Наиболее неоднозначной является мера равномерности заполнения ПЭ. В качестве такого критерия в настоящей работе использована функция размаха [1]. Ее преимуществами по сравнению с другими критериями [2, 3] являются скорость расчета и четкая интерпретация полученных значений с точки зрения качества данных [4].

В данном исследовании рассматривается случайный подход к формированию градуировочного набора образцов. Использование случайных значений часто встречается при задании концентраций компонентов в обучающем наборе для градуировочного эксперимента в спектральном анализе. При случайном подходе обычно используется какой-либо программный генератор псевдослучайных чисел. Однако, как было показано в работе [5], создание качественного градуировочного набора этим методом требует относительно большого числа «случайных» образцов, что делает его неоптимальным.

В настоящей работе предлагается использование квазислучайных множеств [6], то есть множеств точек распределенных более равномерно, чем случайное множество, но не образующих строгую периодическую структуру. В спектральном анализе такой подход ранее не использовался.

**Цель** — найти двумерное квазислучайное множество, позволяющее достигнуть требуемых критериев качества градуировочного набора, используя минимальное число точек (образцов).

**Методы.** Рассматривался диапазон от 3 до 30 точек, где при каждом формировании набора высчитывались три критерия качества: коэффициент корреляции, функция размаха и незаполненность ПЭ. Последняя величина выражается как единица минус отношение площади образованного точками выпуклого многоугольника к общей площади. Хорошими будем считать следующие диапазоны критериев: 0–0,15 по корреляции и незаполненности, 0–0,2 — по функции размаха; допустимыми: 0,15–0,3 по корреляции и незаполненности, 0,2–0,4 — по функции размаха. Более высокие значения критериев считаются недопустимыми. Были исследованы четыре метода создания последовательностей: Соболя, Халтона, Хаммерсли и R2.

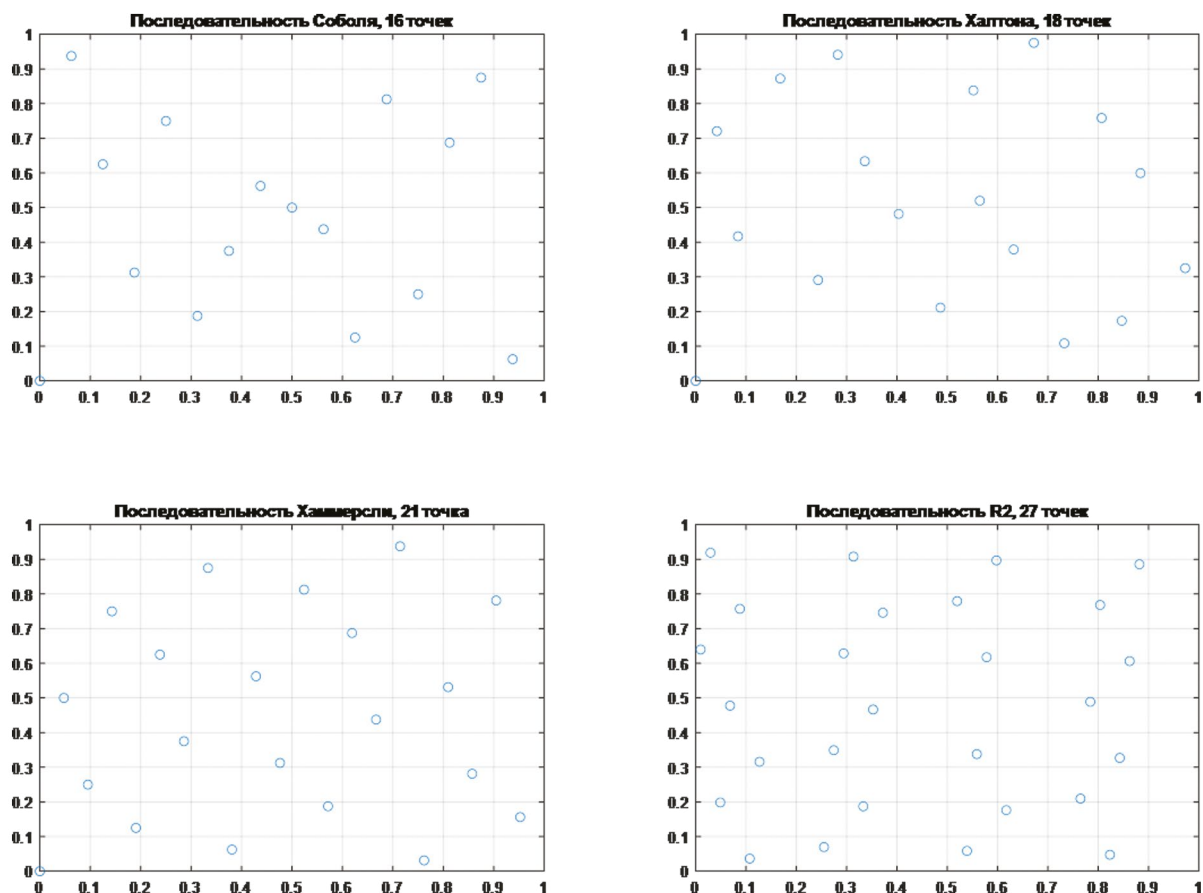


Рис. 1. Графические представления последовательностей

**Результаты.** Рассчитанные значения критериев отражены в табл. 1.

Таблица 1. Последовательности и критерии

Последовательность	Минимальное количество точек	Корреляция	Функция размаха	Незаполненность пространства
Соболя	16	0,012	0,392	0,234
Халтона	18	0,052	0,397	0,284
Хаммерсли	21	0,109	0,360	0,272
R2	27	0,009	0,234	0,296

**Выводы.** Среди квазислучайных последовательностей лучшими оказались последовательности Соболя и Халтона. С применением последовательности Соболя было достигнуто допустимое качество набора для 16 образцов; последовательностью Халтона — для 18 образцов.

**Ключевые слова:** планирование эксперимента; квазислучайные последовательности; равномерное распределение; математическое моделирование.

### Список литературы

1. Bogomolov A. Diagonal designs for a multi-component calibration experiment // *Analytica Chimica Acta*. 2017. Vol. 951. P. 46–57. doi: 10.1016/j.aca.2016.11.038
2. Liang Y.-Z., Fang K.-T., Xu Q.-S. Uniform design and its applications in chemistry and chemical engineering // *Chemom Intell Lab Syst*. 2001. Vol. 58, N 1. P. 43–57. doi: 10.1016/S0169-7439(01)00139-3
3. Kirsanov D., Panchuk V., Agafonova-Moroz M., et al. A sample-effective calibration design for multiple components // *The Analyst*. 2014. Vol. 139, N 17. P. 4303–4309. doi: 10.1039/c4an00227j

4. Leardi R. Experimental design in chemistry: a tutorial // Anal Chim Acta. 2009. Vol. 652, N 1-2. P. 161–172. doi: 10.1016/j.aca.2009.06.015
5. Brown S.D., Tauler R., Walczak B. Comprehensive chemometrics. 2nd edit. Amsterdam: Elsevier, 2024.
6. Кейперс Л., Нидеррайтер Г. Равномерное распределение последовательностей. Москва: Наука, 1985. 408 с.

*Сведения об авторе:*

**Александр Сергеевич Мананков** — аспирант, группа 1-УПНК-1.4.2, химико-технологический факультет; Самарский государственный технический университет, Самара, Россия. E-mail: s90w23.14@mail.ru

*Сведения о научном руководителе:*

**Андрей Юрьевич Богомолов** — доктор химических наук, доцент; Самарский государственный технический университет, Самара, Россия. E-mail: c11b0f5057f5@mail.ru