

Статья опубликована в журнале «Заводская лаборатория. Диагностика материалов» 2011. т.77. № 6. С. 60-68.

ВНУТРИЛАБОРАТОРНЫЙ КОНТРОЛЬ КАЧЕСТВА. ОРГАНИЗАЦИЯ КОНТРОЛЯ СТАБИЛЬНОСТИ И ОЦЕНКА ХАРАКТЕРИСТИК ПРОЦЕССА ИЗМЕРЕНИЙ

Дворкин В.И.

Институт нефтехимического синтеза РАН, Москва.

E-mail: dvorkin@ips.ac.ru

В [1] была рассмотрена основная модель процесса измерений и указано, что основным, наиболее эффективным инструментом при внутрилабораторном контроле является контроль стабильности с использованием контрольных карт. Достаточно сказать, что такой контроль (наряду с проверкой приемлемости) рекомендован базовым международным документом в этой области [2]. В настоящей статье рассмотрено, как именно организуется такой контроль в различных ситуациях, а также алгоритмы оценки состояния измерений перед началом собственно контроля. Как и в [1], в основном говорится о контроле качества химического анализа. При этом следует иметь в виду, что, как и в [2], все сказанное может быть приложено любым измерениям с небольшими оговорками – например, при контроле стабильности измерительных приборов в качестве эквивалента контрольного материала может выступать мера.

Рассмотрим, как организуется контроль стабильности *при наличии контрольного материала*. Под контрольным материалом мы понимаем однородный, стабильный и адекватный рутинным пробам материал, однако в отличие от образцов сравнения и стандартных образцов он может быть не аттестован. Подробнее о контрольных материалах см. ниже.

Общая схема контроля стабильности следующая. Для каждого вида химического анализа (измерений):

1. Выбирают контрольные материалы.
2. Определяют, какие определяемые показатели подлежат контролю.
3. Определяют, что является серией измерений (аналитической серией).
4. Составляют схему (план) контроля, в которой задаются природа и количество контрольных материалов, определяемых веществ и число измерений для контрольного материала в одной аналитической серии. Отметим, что в дальнейшем число измерений для данного контрольного материала и определяемого показателя нельзя изменить без нового набора статистики или, по крайней мере, нового расчета статистических характеристик (при уменьшении числа измерений).
5. В пределах некоторого количества аналитических серий проводят анализ (исследование) контрольного материала (материалов) для оценки реальных статистических характеристик аналитической методики. Мы будем называть эту стадию "период оценки характеристик".
6. По окончании периода оценки характеристик в случае соответствия погрешностей анализа предъявляемым к нему требованиям переходят к собственно контролю. При этом, на-

сколько это возможно, раздельно контролируют точность в целом и повторяемость (сходимость) результатов анализа¹.

Рассмотрим требования, предъявляемые к каждому из этих этапов, и наиболее эффективные алгоритмы контроля. Здесь мы будем рассматривать способы контроля стабильности, в которых результаты определения каждого показателя в каждом контрольном материале оцениваются отдельно².

Планирование процедур контроля стабильности

План эксперимента при внутрилабораторном контроле определяется как чисто метрологическими, так и экономическими факторами.

Выбор контрольных материалов. Общие требования к контрольным материалам – стабильность их в течение срока применения, однородность и адекватность рутинным пробам. Наилучшими с точки зрения адекватности являются материалы одной природы с рутинными пробами (при анализе какого-либо сплава – это сплав близкого состава, при анализе сыворотки крови человека – стабилизированная лиофилизацией или замораживанием сыворотка крови человека и т.д.). Таким образом, общие требования к контрольным материалам близки к требованиям, предъявляемым к материалам, подготовленным для аттестации в качестве стандартных образцов состава [3].

С позиций обеспечения надежности контроля количество контрольных материалов с разной концентрацией определяемого вещества должно быть по возможности велико, причем эти концентрации должны перекрывать всю область, в которой "работает" контролируемая методика химического анализа. Однако в реальности лаборатории ограничены экономическими соображениями и ограниченным ассортиментом материалов.

Строгое решение задачи об оптимальном выборе числа контрольных материалов и набора концентраций определяемых веществ в них сопряжено со специальными дорогостоящими исследованиями. Цель таких исследований (которые могли бы проводиться, но обычно не делаются из-за их дороговизны) можно сформулировать так: надо найти такой минимальный набор материалов, чтобы стабильность, продемонстрированная при их анализе, гарантировала бы стабильность во всем диапазоне определяемых по данной методике концентраций.

На практике контрольные материалы выбирают исходя из априорных знаний о применяемом методе (методике) анализа, а также с учетом важности того или иного диапазона концентраций для конкретных областей аналитических работ. При этом если концентрацию невозможно варьи-

¹ Здесь мы сталкиваемся с терминологической проблемой. Дело в том, что "повторяемость (сходимость)" согласно [2] определена как степень близости между независимыми результатами измерений, полученными при помощи одного и того же метода и идентичного анализируемого материала в одинаковых условиях, т.е. является характеристикой не методики химического анализа (измерения) вообще или ее реализации в конкретной лаборатории, а характеристикой результатов анализа конкретных проб. Получается, что в случае одного измерения каждой пробы такой характеристики, как повторяемость (сходимость), как бы нет. Это явное недоразумение - характеристика методики, описывающая разброс результатов внутри серии, существует объективно, независимо от числа параллельных измерений. Поэтому будем употреблять общепринятый термин "повторяемость (сходимость)" в смысле "внутрисерийная прецизионность анализа". Естественно, что для оценки этой величины необходимо провести параллельные измерения, к результатам которых термин "повторяемость (сходимость)" относится безусловно.

² Один контрольный материал может использоваться для контроля качества нескольких показателей. Поскольку результаты контроля по каждому показателю оцениваются отдельно, все нижеказанное справедливо для определения каждого показателя.

ровать произвольно, контрольные материалы приходится выбирать из имеющихся в наличии. Так, при экологических исследованиях используют либо один материал с концентрацией вблизи ПДК, либо два контрольных материала с концентрациями вблизи и существенно выше ПДК; такой же выбор чаще всего оптимален в случае сертификации пищевых продуктов. Если весь диапазон концентраций одинаково важен, используют либо один материал с концентрацией вблизи середины диапазона определяемых концентраций, либо два или даже три контрольных материала (в последнем случае с содержанием аналита, находящимся на краях и в середине диапазона определяемых концентраций (медицина)).

Для контроля стабильности не обязательны (хотя и желательны) материалы с точно известным содержанием определяемых веществ. Хотя очень часто для этих целей используют аттестованные материалы (ГСО, стандартные образцы предприятий, аттестованные смеси и т.д.), в случае их отсутствия либо из экономии берут не аттестованные – лишь бы они были стабильны, однородны и адекватны объектам анализа. Типичный пример – определение концентрации различных веществ в моче: для контроля стабильности часто используют слитую мочу, стабилизированную антимикробными добавками. В сфере клинической лабораторной диагностики, где контроль стабильности является обязательным, фирмы – производители стандартных образцов в больших количествах выпускают не аттестованные контрольные материалы – цена их в разы ниже цены аналогичных стандартных образцов.

Отдельно надо рассмотреть ситуацию, когда не удастся приобрести или изготовить контрольный материал, адекватный исследуемым пробам, то есть невозможно вести контроль стабильности методики в целом. При этом, конечно, остается возможность вести контроль без контрольного материала (такие варианты будут рассмотрены впоследствии). Однако во многих случаях в подобных ситуациях возможен частичный контроль стабильности – с материалами, имитирующими вещества, получающиеся после каких-либо этапов выполнения методики. Например, при определении концентрации тяжелых металлов в пылевых частицах в воздухе рабочей зоны изготовление контрольного материала, адекватного исследуемому объекту, практически невозможно. При этом легко изготовить стабильный раствор, похожий на те, которые получают после растворения пыли из аспиратора, и вести по нему контроль стабильности. Конечно, в таком случае не будут контролироваться стадии пробоотбора и перевода в раствор. А что делать? Приходится мириться с неполным контролем – все же это лучше, чем ничего.

Определение перечня контролируемых показателей. Конечно, если по данной методике определяется лишь один показатель, вопрос о выборе не стоит. Однако в последние годы появляется все больше методов, в которых в ходе одного анализа определяется несколько (иногда до нескольких десятков) показателей: газовая хроматография, эмиссионный спектральный анализ, масс-спектрометрия и т.д. Конечно, можно контролировать стабильность по всем показателям, однако из соображений экономии их число во многих случаях можно ограничить. Делается это на основе знаний о природе метода и собственного опыта. Так, в газовой хроматографии факторы, влияющие на стабильность, практически одинаковы для разных определяемых веществ: воспроизводимость ввода пробы, неизменность состояния колонки и детектора и условий их работы и т.д. В подобных случаях можно выбрать один или два определяемых показателя (обычно самых «капризных») и контролировать по ним стабильность измерений, полагая, что если для выбранных веществ все стабильно, то и для остальных это тоже так. Так, в лаборатории автора стабильность одновременного определения четырех элементов: C, H, N, S на автоанализаторе ф. Carlo-Erba (это газовый хроматограф с предварительным полным сжиганием пробы) контроль стабильности ведется только по результатам определения водорода в ацетанилиде. Из опыта известно, что водород – самый нестабильный показатель из определяемых элементов, а чистый ацетанилид – превосходный контрольный материал с точно известным содержанием водорода.

Определение серии измерений (аналитической серии). Серия измерений - совокупность образцов, исследуемых одним и тем же методом, в одной и той же лаборатории, при одной и той же калибровке, одним и тем же оператором, с использованием одного и того же оборудования, в пределах короткого промежутка времени (в условиях повторяемости (сходимости) согласно [2]). В большинстве случаев в реальной лаборатории определение серии не вызывает трудностей – это некоторое количество проб, исследуемых (анализируемых) за один раз. Например, при определении концентрации тяжелых металлов в водных пробах методом эмиссионной спектроскопии с индукционно-связанной плазмой лаборант включает прибор, выполняет градуировку и «прогоняет» пробы, поступившие в лабораторию, одна за другой – это серия. Число рутинных образцов в серии («исследованных в условиях повторяемости (сходимости)» согласно [2]), ограничено сверху как реальными потребностями, так и неизменностью условий измерений.

Смысл понятия «серия измерений» в том, что мы должны быть уверены: все факторы (время, градуировка, лаборант и прибор) одинаковы для всех проб в серии. Пробы другой серии исследуются в других условиях; именно поэтому *контроль стабильности ведут в каждой серии* – иначе нельзя гарантировать качество результатов (а вдруг серия, в которой не было контрольных проб, была «плохой»?).

Сложности с определением, что такое серия, возникают, когда измерения ведутся без разделения проб на группы. Реальный пример: в заводской лаборатории химического производства необходимо постоянно и оперативно контролировать вязкость полупродуктов. Пробу отбирают, приносят в лабораторию и свободный лаборант сразу ее исследует на любом свободном вискозиметре. В подобных ситуациях ищут компромисс: не следуя строго определению серии, понимают под ней пробы, проанализированные в одних условиях. В нашем примере это, скорее всего все пробы, исследованные в течение одной смены – можно условиться, что в одну смену условия одинаковы, а в другую смену они будут другими (хотя возможны и варианты – например, считать, что серия – это пробы, полученные одним лаборантом в одну смену, что сложнее).

Другая непростая ситуация – непрерывно работающий автоматический прибор, выполняющий периодические (например, раз в 10 минут) измерения без участия человека. Пример - круглосуточный контроль концентрации в химическом реакторе автоматическим хроматографом. Размер серии в этом случае определяют исходя из соотношения скорости возможных изменений статистических характеристик анализа (дрейфов) и общей погрешности исследования таким образом, чтобы этими изменениями можно было бы заведомо пренебречь. Другими словами, это - типичный период, в течение которого вероятность выхода процесса измерений из-под контроля достаточно мала.

Если есть возможность планировать размер серии, необходимо принимать во внимание и экономические соображения. Так, в случае "длинных" серий (30–40 и более рутинных проб) доля затрат на контрольные материалы и их анализ значительно меньше, чем при "коротких" сериях, что дает исследователю бóльшую степень свободы.

Выбор числа измерений для контрольного материала в одной серии. Когда контрольный материал (материалы) для контроля данной МВИ выбраны и определены показатели, по которым будет проводиться контроль, остается определить число измерений, которые будут выполняться для каждого контрольного материала в каждой серии, и расположение контрольных проб в сериях. Отметим, что в дальнейшем число измерений для данного контрольного материала и определяемого показателя нельзя изменить без нового набора статистики или, по крайней мере, нового расчета статистических характеристик (при уменьшении числа измерений).

Во многих случаях число измерений для контрольного материала выбирают таким же, как и при анализе рутинных проб. На первый взгляд это кажется естественным (и даже включено в нормативные документы [4]). На самом деле желательно иметь больше параллельных измерений для контрольного материала, чем для рутинных проб. Это наиболее очевидно для тех ситуаций, когда при поточном рутинном анализе из экономических соображений проводят только одно измерение (анализ почв, клинический анализ, геологические исследования и т.д.). В этом случае на результат измерений влияет как межсерийный разброс (промежуточная прецизионность), так и случайный разброс внутри аналитической серии (повторяемость (сходимость)), причем эти факторы имеют разную природу [5, 6]. Учитывая, что целью контроля стабильности является не только выявление неудовлетворительных результатов, но и устранение причин возникшего сбоя, естественно контролировать оба компонента погрешности – и повторяемость (сходимость), и промежуточную прецизионность.

Дополнительным аргументом, заставляющим стремиться к возможно большему числу параллельных измерений, является тот факт, что, согласно центральной предельной теореме [5, 6] даже при существенном отклонении распределения результатов измерений от нормального среднее из нескольких результатов приближается к нормальному (то есть чем больше результатов усредняется, тем меньше распределение средних отличается от нормального). Известно, что при распределениях, реально встречающихся при анализе вещества, группировка с усреднением приводит к практически нормальным распределениям средних значений уже при числе измерений в группе 4–5, а часто и меньше. А практически все методы внутрилабораторного контроля основаны на предположении о нормальном распределении результатов измерений.

Понятно, что экономические факторы в огромной степени влияют на организацию рутинных измерений и контроля их качества. Исследование контрольных материалов может вносить существенный вклад в стоимость каждого измерения для рутинных проб, и этот вклад прямо зависит от числа измерений для контрольных материалов. Минимальное число измерений для одного контрольного материала, позволяющее контролировать повторяемость (сходимость) по контрольной карте размахов – два. Именно поэтому чаще всего при контроле стабильности проводят два измерения для каждого контрольного материала. То же мы видим в основополагающем документе [2], где во всех примерах фигурирует по два контрольных измерения.

Важную роль играет расстановка контрольных проб в серии. Она выбирается из того соображения, что результаты контроля должны отражать качество исследования рутинных проб. Поскольку случайный характер погрешностей в пределах серии не гарантирован (возможны дрейфы и другие систематические эффекты, которые также надо выявлять), рекомендуется случайная расстановка. Идеально использовать таблицы случайных чисел. Часто применяемое, особенно при работе с автоматическими пробоотборниками, расположение контрольных проб на фиксированных позициях может привести к некорректным оценкам повторяемости (сходимости). В нашей практике были случаи, когда из-за дефекта пробоотборника результаты измерений в некоторых позициях были неправильными (а в остальных – нормальными). При случайной расстановке контрольных проб такая погрешность рано или поздно будет выявлена (когда контрольная проба окажется в дефектной позиции). Если же контрольные пробы всегда располагаются на одних и тех же позициях, этого может не произойти никогда.

Некоторые организационные аспекты. В целом можно выделить три составляющих внутрилабораторного контроля: персонал, который собственно выполняет контроль; контрольные материалы (если контроль ведется с их применением, как чаще всего бывает) и обработка и использование его результатов – ведение контрольных карт, выполнение расчетов и т.д. Часто создается иллюзия, что наиболее сложная компонента – последняя, несколько проще решить вопрос с кон-

трольными материалами, а люди, конечно, будут делать все так, как надо. На самом деле все обстоит с точностью до наоборот: проще всего наладить расчеты и ведение контрольных карт, во многих случаях приходится помучиться с выбором контрольных материалов, а самое сложное – добиться, чтобы персонал честно и корректно выполнял нужные процедуры.

Дело в следующем. Все, что делается в рамках внутрилабораторного контроля (исследование контрольных материалов, обработка их результатов и т.д.), с позиции лаборанта³ может показаться дополнительной и ненужной работой. Соответственно они часто делают эту работу спустя рукава, а иногда даже не делают ее вовсе, фальсифицируя результаты контроля. Чтобы этого не происходило, необходимо максимально удобно организовать контроль на рабочем месте и добиться понимания важности этих процедур всеми сотрудниками лаборатории. Это в основном определяет успех всех мероприятий по внутрилабораторному контролю.

Удобство выполнения контроля – естественное эргономическое требование. Конечно, если контрольный материал просто стоит на полке над рабочим столом, никаких проблем не возникает. Однако контрольные материалы часто отличаются от рутинных проб и требуют специальных условий хранения (например, в холодильнике или в запаянной ампуле). Часто при их исследовании необходимы специальные подготовительные процедуры, не нужные для рутинных проб (вскрытие ампул, размораживание, разбавление концентрата и т.д.). Эти процедуры должны доставлять лаборанту минимум неудобств (так, холодильник с контрольными пробами должен находиться в том же помещении, в котором происходят измерения, а не на другом этаже). То же касается и компьютерной обработки результатов измерений – компьютер либо терминал должен находиться рядом с рабочим местом лаборанта (благо сейчас это недорого). Все это кажется очевидным, но часто не делается – сказывается некоторое пренебрежение контрольными процедурами. Между тем если выполнение контроля требует серьезных дополнительных усилий, его могут просто не делать.

Возможно, еще более важным является психологический аспект. Прямым путем к добросовестному и качественному выполнению процедур внутрилабораторного контроля является убеждение лаборантов и вообще всех сотрудников лаборатории в том, что их благосостояние в настоящем и будущем зависят от качества выполняемых исследований, а это качество невозможно гарантировать без внутрилабораторного контроля. На самом деле это действительно так. Для коммерческих лабораторий, работающих на свободном рынке, доходы (а значит, зарплата сотрудников и условия их работы) напрямую зависят от качества измерений – это очевидно. Однако и для заводских лабораторий выдача некачественных результатов чревата ухудшением отношения со стороны руководства и, как следствие, уменьшением стимулирующих выплат и закупок оборудования и, возможно, кадровыми решениями. Даже один ошибочный результат измерения (если ошибка будет выявлен, а это происходит не так уж редко) потребители результатов – технологи, экологи и т.д. – будут вспоминать этот случай годами!

Эта мысль должна быть доведена до всех сотрудников лаборатории, а особенно – до лаборантов. Сделать это непросто, требуется индивидуальный подход. Однако можно дать общий совет – ни в коем случае нельзя ругать лаборанта или (тем более) не наказывать его материально в случае неудовлетворительных результатов внутрилабораторного контроля: это будет тем, что психологи называют «отрицательным стимулированием». После «разноса» лаборант будет делать все возможное, чтобы он не повторился. А поскольку сбои в процессе измерений неизбежны и являются нормальным явлением, избежать их полностью можно только одним способом – подделывать результаты в случае выхода процесса из-под контроля. Конечно, картина при этом будет благодная, но вряд ли это улучшит реальное качество измерений.

³ Под «лаборантом» мы понимаем любого сотрудника, выполняющего рутинные измерения.

Следует иметь в виду, что на самом деле низкое качество – следствие неудовлетворительного руководства; система руководства есть причина большинства проблем в этой области. Ключ к хорошему качеству – уважение, признание людей [7]. Поэтому нормальная реакция на выявленный выход процесса измерений из-под контроля – не претензии и угрозы, а изъяснение удовлетворения по типу «как хорошо, что мы вовремя заметили этот сбой и неправильный результат не пойдет заказчику. Теперь нам надо разобраться, из-за чего это произошло» и т.д. Такой образ действий создает максимально комфортную атмосферу в лаборатории.

В настоящее время во многих отечественных лабораториях психологические проблемы решены и внутрилабораторный контроль проводится реально и в полном объеме.

«Юридическая» проблема. Одна из сложностей, возникающих при внедрении современных процедур внутрилабораторного контроля в нашей стране, состоит в том, что практически во всех методиках в соответствии с ГОСТом [8] прописаны процедуры контроля качества, и для большинства методик они безнадежно устарели. Общее число методик на МВИ оценивается в 30-35 тысяч и многие были разработаны еще в 70-е – 80-е годы прошлого века. Правда, приказ о внедрении документа [2] содержал требование приведения в соответствие с ним всех нормативных документов (в том числе и документов на методики), однако оно не было выполнено. Конечно, было бы правильным централизованно внести изменения в методики (и узаконить их). Однако ясно, что этого не будет – объем работы огромен, да и многих организаций-разработчиков методик уже не существует.

Получается, что формально использование современных методов внутрилабораторного контроля (контроль стабильности, проверка приемлемости) во многих случаях противоречит тексту методик.

Для методик, используемых вне сферы государственного надзора, особой проблемы нет – легко внести соответствующие изменения в методику, узаконить их в рамках предприятия и согласовать с заказчиками исследований, тем более что в большинстве случаев контроль стабильности «поглощает» другие методы контроля. Согласно Закону о техническом регулировании [9] обязательными являются только технические регламенты – выполнение требований всех остальных документов, включая методики выполнения измерений (если они не вошли в какой-либо регламент) является добровольным и во главу угла ставятся требования Заказчика. Поэтому лаборатории могут самостоятельно изменить методику в части процедур контроля качества измерений, согласовав эти изменения с заказчиками исследований и, при необходимости, утвердив эти изменения в рамках своего предприятия (стандарт предприятия). Конечно, все это требует внесения соответствующих изменений в Руководство по качеству лаборатории.

Сложнее обстоит дело с работами, лежащими в сфере государственного контроля и надзора. Статьей 15 Закона [10] установлен государственный надзор за соблюдением аттестованных методик измерений, относящихся к сфере государственного регулирования обеспечения единства измерений. Формально (а это стандартный подход надзорных органов) изменение процедур контроля может быть расценено, как нарушение требований методики. Понятно, что при аккредитации орган по аккредитации может закрыть на это глаза или даже указать, что процедуры контроля, записанные в методике, являются недостаточными, но неприятности со стороны надзорных органов вполне возможны. Остается надеяться на здравый смысл проверяющих и на то, что процесс обновления методик постепенно выведет устаревшие, неэффективные способы контроля из обращения.

Оценка подконтрольности измерений за период оценки характеристик

Как указано выше, этот период состоит в анализе контрольного материала (материалов) в течение нескольких серий измерений для оценки⁴ реальных статистических характеристик методики. Естественно, первый вопрос – какова продолжительность этого периода. Выбор продолжительности периода оценки характеристик анализа, как и при оценке метрологических характеристик методики количественного химического анализа при ее аттестации и формы распределения вообще, представляет собой компромисс между желанием получить возможно более точные оценки и необходимостью быстрее перейти к собственно контролю. Точные рекомендации дать невозможно, однако в большинстве случаев считается достаточным иметь результаты 20 серий измерений; во всяком случае это никак не меньше 15 серий.

Результатом должна стать оценка метрологических характеристик и соответствия их предъявляемым требованиям. Все это иногда называют «валидацией».

Оценка повторяемости (сходимости)

Наиболее естественный и не требующий дополнительных измерений способ оценки повторяемости (сходимости) данном случае следующий.

Обозначим результаты, полученные в период оценки характеристик, как X_{ji} , где $j = 1, 2, \dots, M$ – номер серии, $i = 1, 2, \dots, N$ – номер измерения. Как указано выше, обычно $M = 20 - 25, N = 2$ (Табл. 1). Точки в нижней строке таблицы говорят о том, что после набора M серий, то есть окончания периода оценки характеристик, начинается собственно контроль стабильности.

Таблица 1. Результаты, полученные в период оценки характеристик

Номер серии	Номер измерения					
	1	2	...	i	...	N
1	X_{11}	X_{12}	...	X_{1i}	...	X_{1N}
2	X_{21}	X_{22}	...	X_{2i}	...	X_{2N}
•
j	X_{j1}	X_{j2}	...	X_{ji}	...	X_{jN}
•
M	X_{M1}	X_{M2}	...	X_{Mi}	...	X_{MN}
•

В этом случае стандартное отклонение повторяемости s_r рассчитывают по формуле

$$s_r = \sqrt{\frac{\sum_{j=1}^M \sum_{i=1}^N (X_{ji} - \bar{X}_j)^2}{M(N-1)}}, \quad (2)$$

где $\bar{X}_j = \sum_{i=1}^N X_{j,i} / N$. При $N = 2$ расчеты можно проводить по формуле

⁴ в метрологической литературе используют термин "оценивание".

$$s_r = \sqrt{\frac{\sum (X_{j,1} - X_{j,2})^2}{2M}}, \quad (3)$$

где $X_{j,1}$ и $X_{j,2}$ - результаты, полученные в j -й серии.

Полученную величину s_r сравнивают с нормативом σ_r - приписанным значением стандартного отклонения повторяемости (сходимости), почти всегда приводимым в методике измерений⁵, по формуле

$$\frac{s_r^2}{\sigma_r^2} < \chi^2(\bar{P}, f) / f, \quad (4)$$

где $\chi^2(\bar{P}, f)$ есть $\bar{P} = 1 - \alpha$ квантиль χ^2 -распределения с числом степеней свободы f (доверительная вероятность). В отсутствие специальных требований α принимают равной 0.05. При расчетах по формулам (2, 3) число степеней свободы равно $f = M(N - 1)$.

Описанная оценка повторяемости (сходимости) проводится, как правило, при работе с контрольными картами - на этапе оценки статистических характеристик. Однако такая оценка может выполняться за любой промежуток времени, в который выполнялись соответствующие измерения.

В [2] рекомендован другой способ оценки повторяемости (сходимости), при котором оценка проводится в одной серии измерений. Для такой оценки необходим однородный, стабильный в течение времени измерений и адекватный материал в достаточном количестве. Это может быть, например, достаточно однородная либо тщательно перемешанная рутинная проба; гомогенизированная смесь нескольких рутинных проб; стандартный образец состава; и так далее. Согласно [2] этот материал анализируют не менее чем $N = 8 \div 10$ раз в одной аналитической серии и находят оценку стандартного отклонения сходимости s_r как среднеквадратичное отклонение полученных результатов $X_1, X_2, \dots, X_i, \dots, X_N$

$$s_r = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N-1}}, \quad (5)$$

⁵ Часто в методиках приводится величина предела повторяемости (сходимости) r , в более старых методиках обозначаемая как d . До сих пор иногда это делается формулировкой «разность двух результатов, получаемых в одной лаборатории, не превышает ...». В таких случаях величину σ_r рассчитывают как $\sigma_r = r/2.77$.

где $\bar{X} = \sum_{i=1}^N X_i / N$. Далее найденную величину s_r сравнивают с метрологической характеристикой σ_r по формуле (4). Число степеней свободы при расчете s_r по формуле (5) равно $f = N - 1$.

Как указано выше, такой способ плохо сочетается с контролем стабильности, при котором, контрольный материал исследуется в нескольких сериях и в каждой серии выполняется немного (обычно 2) измерения. Естественно и менее трудоемко выполнять оценку без «лишних» измерений. Конечно, оценка в одной, специально организованной серии неизбежна, если контроль стабильности не ведется вообще либо при контроле выполняется лишь одно измерение в серии.

При использовании обоих способов оценки стандартного отклонения повторяемости (стабильности) это - проверка гипотезы о равенстве двух дисперсий – оцененной s_r^2 и заданной нормативным документом σ_r^2 . Интересно, что такая процедура предполагает «презумпцию невиновности» лаборатории - величина s_r , несколько превышающая норматив σ_r признается удовлетворительной.

Если в методике приведено значение предела повторяемости r , то σ_r находят по формуле $\sigma_r = r / 2.77$. Если в методике, предусматривающей расчет окончательно приводимого результата как среднего из n параллельных измерений, дано стандартное отклонение повторяемости среднего значения $\sigma_{r, \bar{X}}$, то σ_r находят по формуле $\sigma_r = \sigma_{r, \bar{X}} / \sqrt{n}$. Если же для такой методики приведен предел повторяемости для среднего значения $r_{\bar{X}}$, то σ_r находят по формуле $\sigma_r = r_{\bar{X}} / 2.77 \sqrt{n}$.

Оценка промежуточной прецизионности

Имея данные, приведенные в Табл. 1, стандартное отклонение промежуточной прецизионности s_{IP} можно рассчитать по формуле

$$s_{IP} = \sqrt{\frac{\sum (\bar{X}_j - \bar{\bar{X}})^2}{M - 1} + \left(1 - \frac{1}{N}\right) s_r^2}, \quad (6)$$

где $\bar{X}_j = \sum_{i=1}^N X_{ji} / N$, $\bar{\bar{X}} = \sum_{j=1}^M \bar{X}_j / M$.

Норматив для промежуточной прецизионности в методиках практически всегда отсутствует. Можно, однако, предложить полуколичественный тест, а именно: если $s_{IP} > 1.5 s_r$, то с высокой вероятностью разброс результатов от серии к серии (который, собственно, и характеризуется промежуточной прецизионностью) чрезмерно велик и желательно принять меры по его уменьшению. И уж, конечно, должно выполняться условие $s_{IP} < \sigma_R$, где σ_R - стандартное отклонение

повторяемости, заданное в методике. Если это соотношение не выполняется, безусловно, необходимо принять меры к снижению промежуточной прецизионности.

Оценка правильности

Как и оценка повторяемости (сходимости), правильность может быть оценена в нескольких сериях (в ходе контроля стабильности) и в одной, специально организованной серии. Если ведется контроль стабильности, правильность естественно оценивать в нескольких сериях (данные из Табл. 1). Такая оценка может проводиться только в случае, когда для контроля используется контрольный материал с известным принятым опорным значением: стандартный образец либо аттестованная смесь; образец, предварительно исследованный в авторитетной лаборатории и т.д. Мы рассмотрим способы с использованием стандартного образца (СО) и образца, предварительно проанализированного другой (авторитетной) лабораторией.

Оценка правильности в ходе контроля стабильности с использованием СО либо образца сравнения. Проведя измерения в M сериях, проверяют гипотезу об отсутствии систематической погрешности следующим образом. Рассчитывают величину

$$t = \frac{|\bar{X} - X_a|}{\sqrt{\frac{s_{\bar{X}_j}^2}{M} + \frac{(\Delta X_a)^2}{3}}}, \quad (7)$$

где X_a и ΔX_a - соответственно, принятое для СО опорное значение и его погрешность⁶,

$s_{\bar{X}_j} = \sqrt{\frac{\sum (\bar{X}_j - \bar{X})^2}{M - 1}}$. Если $t > t(\bar{P}, f)$, при выбранном уровне доверительной вероятности \bar{P}

(обычно 0.975 для двустороннего критерия) и числе степеней свободы $f = M - 1$, то гипотезу отвергают, что означает наличие систематической составляющей погрешности. В этом случае можно оценить полную систематической погрешности $\Delta_{c,n} = \bar{X} - X_a$ и нижнюю границу доверительного интервала для нее

$$\Delta_{c,n} = \left| \bar{X} - X_a \right| - \frac{s_{\bar{X}_j} t(\bar{P}, f)}{\sqrt{M}} - \Delta X_a. \quad (8)$$

Если в методике указана максимально допустимая величина систематической погрешности $\Delta_{c,max}$, то необходимо выполнение неравенства

$$\Delta_{c,n} < \Delta_{c,max}. \quad (9)$$

⁶ Конечно, всегда лучше использовать такое СО, погрешностью аттестации которого согласно принципу пренебрежения малыми погрешностями можно пренебречь. Однако это не всегда возможно.

В противном случае систематическая погрешность превышает допустимую, и необходимо принять меры по ее уменьшению. То же необходимо сделать, если величина $\Delta_{c,\max}$ в методике не приведена.

В случае контроля стабильности с помощью образца, предварительно проанализированного в другой (авторитетной) лаборатории исследуют этот образец, так же, как СО в предыдущем случае.

Обозначим результаты, полученные при исследовании образца в другой лаборатории, как $Y_1, Y_2, \dots, Y_i, \dots, Y_K$.

Гипотезу об отсутствии систематической погрешности в этом случае проверяют следующим образом. Рассчитывают величину

$$t = \frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{s_X^2}{M} + \frac{s_Y^2}{K}}}, \quad (10)$$

Если $t > t(\bar{P}, f)$, при уровне доверительной вероятности \bar{P} (обычно 0.975 для двустороннего критерия) и числе степеней свободы $f = M + K - 2$, то гипотезу отвергают, что означает наличие систематической составляющей погрешности. В этом случае оценка полной систематической погрешности $\Delta_c = \bar{X} - \bar{Y}$, а нижняя граница доверительного интервала для нее

$$\Delta_{c,n} = |\bar{X} - \bar{Y}| - \Delta X - \Delta Y, \quad (11)$$

где $\Delta X = \frac{s_X \cdot j t(\bar{P}, f)}{\sqrt{M}}$, $\Delta Y = \frac{s_Y t(\bar{P}, f_1)}{\sqrt{K}}$, $f_1 = K - 1$. Оценка систематической погрешности

в случае, когда в методике указана допустимая величина систематической погрешности, проводится по формуле (9). Если систематическая погрешность превышает допустимую, необходимо принять меры по ее уменьшению. То же необходимо сделать, если величина $\Delta_{c,\max}$ в методике не приведена.

Оценка правильности в отдельной серии измерений. В случае использования для контроля стабильности не аттестованного контрольного материала оценивать правильность приходится в специально организованном эксперименте. При этом можно использовать разные способы: анализ стандартного образца либо образца, проанализированного в образцовой лаборатории, метод добавок и т.д. Рассмотрим способы с использованием СО и образца, проанализированного другой (авторитетной) лабораторией.

Оценка правильности в одной серии с использованием СО. СО (либо образец сравнения) анализируют несколько раз в одной серии. Далее проверяют гипотезу об отсутствии систематической погрешности следующим образом. Рассчитывают величину

$$t = \frac{|\bar{X} - X_a|}{\sqrt{\frac{s_r^2}{N} + \frac{(\Delta X_a)^2}{3}}}, \quad (12)$$

где X_a и ΔX_a - соответственно, принятое для СО опорное значение и его погрешность. Если $t > t(\bar{P}, f)$, при выбранном уровне доверительной вероятности \bar{P} (обычно 0.975 для двустороннего критерия) и числе степеней свободы $f = N - 1$, то гипотезу отвергают, что означает наличие систематической составляющей погрешности. В этом случае оценивают полную систематической погрешности $\Delta_c = \bar{X} - X_a$ и нижнюю границу доверительного интервала для нее

$$\Delta_{c,n} = |\bar{X} - X_a| - \frac{s_r t(\bar{P}, f)}{\sqrt{N}} - \Delta X_a \quad . \quad (13)$$

Если в методике указана максимально допустимая величина систематической погрешности $\Delta_{c,max}$, то необходимо выполнение неравенства (9). Если оно не выполняется, систематическая погрешность превышает допустимую, и необходимо принять меры по ее уменьшению. То же относится к случаю, когда $\Delta_{c,max}$ в методике не нормируется.

Оценка правильности в одной серии с помощью образца, проанализированного в другой (авторитетной) лаборатории. Такой образец анализируют в одной серии несколько (N) раз.

Обозначим результаты, полученные при исследовании образца в другой лаборатории, как $Y_1, Y_2, \dots, Y_i, \dots, Y_K$, среднее значение $\bar{Y} = \frac{\sum Y_i}{K}$, стандартное отклонение величин Y как

$$s_Y = \sqrt{\frac{\sum (Y_i - \bar{Y})^2}{K - 1}} \quad .$$

Далее проверяют гипотезу об отсутствии систематической погрешности, рассчитывая величину

$$t = |\bar{X} - \bar{Y}| / \sqrt{\frac{s_r^2}{N} + \frac{s_Y^2}{K}} \quad . \quad (14)$$

Если $t > t(\bar{P}, f)$ при выбранном уровне доверительной вероятности \bar{P} (обычно 0.975 для двустороннего критерия) и числе степеней свободы $f = N + K - 2$, то гипотезу отвергают, что означает наличие систематической составляющей погрешности. В этом случае оценка полной систематической погрешности $\Delta_c = \bar{X} - \bar{Y}$, а нижняя граница доверительного интервала для нее

$$\Delta_{c,n} = |\bar{X} - \bar{Y}| - \Delta X - \Delta Y \quad , \quad (15)$$

где $\bar{Y} = \frac{\sum Y_i}{K}$, $\Delta X = \frac{s_r t(\bar{P}, f)}{\sqrt{N}}$, $\Delta Y = \frac{s_{r,Y} t(\bar{P}, f_1)}{\sqrt{K}}$, $f_1 = K - 1$. Оценка системати-

ческой погрешности в случае, когда в методике указана допустимая величина систематической погрешности, проводится по неравенству (9). Если оно не выполняется, систематическая погрешность превышает допустимую, и необходимо принять меры по ее уменьшению. То же относится к случаю, когда $\Delta_{c,max}$ в методике не нормируется.

Фильтрация результатов измерений в ходе оценки характеристик

Часто (хотя и не всегда) перед собственно оценкой характеристик выявляют и отбрасывают грубые промахи. При этом используются два основных статистических теста: ζ – критерий [11] (он же «критерий Граббса» в [2]) и критерий Кохрена.

ζ -критерий (критерий Граббса) предназначен для выявления грубого промаха в совокупности результатов измерений, имеющих простейшую структуру $X_1, X_2, \dots, X_j, \dots, X_M$. В применении к нашим данным (Табл. 1) $X_j = \sum_{i=0}^N X_{ji} / N$. Пусть величины X_j расположены по возрастанию. Тогда ζ -критерий (критерий Граббса) рассчитывается как $\zeta = (X_M - \bar{X})/s$, если подозрительно выделяется X_M , либо как $\zeta = (\bar{X} - X_1)/s$, если подозрительно выделяется X_1 . Если $\zeta > \zeta(\alpha, M)$, где $\zeta(\alpha, M)$ – критическое значение ζ -статистики при вероятности ошибки 1-го рода α , то подозрительно выделяющееся значение считают грубым промахом и удаляют из дальнейших расчетов. В [2] соответствующая статистика обозначена как G . Обычно выбирают $\alpha = 0.05$ или $\alpha = 0.01$ (таблицы критических значений имеются в [2, 11]).

Критерий Кохрена (Кочрена; Cochren) используется для выявления серий, разброс результатов в которых аномально высок. Предполагается, что число измерений в сериях одинаковой (как это и должно быть). Рассчитывают стандартные отклонения s_j для каждой j -й серии и находят максимальное из них s_{\max} . Рассчитывают величину

$$G = \frac{s_{\max}^2}{s_1^2 + s_2^2 + \dots + s_{\max}^2} \quad (16)$$

Найденную величину G сравнивают с процентной точкой распределения $g(\alpha, M, N)$, где α – вероятность ошибки 1-го рода (обычно 0.05). Если $G > g(\alpha, M, N)$, то серию с максимальным стандартным отклонением s_{\max} отбрасывают. Таблицы критических значений для критерия Кохрена имеются, например, в [2, 11] (в [2] критерий обозначен буквой C).

Оба критерия предполагают нормальное распределение результатов как внутри серий, так и между ними. В общем-то, можно проводить фильтрацию последовательно несколько раз, однако надо очень хорошо подумать, прежде чем отбрасывать более чем одну серию из 20 – это может привести к чрезмерному цензурированию выборки и нереальным оценкам метрологических характеристик.

Во многих случаях, если какую-либо серию отбрасывают (отфильтровывают), вместо нее выполняют еще одну.

Если по завершении стадии оценки метрологических характеристик была проведена фильтрация данных по ζ -критерию (критерию Граббса) и критерию Кохрена и затем корректно велся контроль стабильности по контрольным картам средних и расхождений (размахов) (см. ниже), в дальнейшем применение этих критериев обычно не выявляет достоверных промахов.

Проверка соответствия метрологических характеристик, достигнутых лабораторией, нормативным проводится по окончании периода оценки характеристик. Однако понятно, что в реальности опасно дожидаться окончания этого достаточно длительного периода – можно оказаться у разби-

того корыта, если какая-то характеристика окажется неудовлетворительной. Поэтому естественно проверять качество анализа несколько раз по мере набора данных - например, после набора 4-5; 8-10 и т.д. серий - и продолжать работать только при удовлетворительных результатах. В противном случае набор данных останавливают и принимают меры по повышению качества измерений.

После завершения периода оценки характеристик и при удовлетворительных результатах оценки переходят собственно к контролю стабильности, который будет рассмотрен в следующей статье на эту тему.

ЛИТЕРАТУРА

1. *Дворкин В.И. Заводская лаборатория. 2011.*
2. ГОСТ Р ИСО 5725-2002. Точность (правильность и прецизионность) методов и результатов измерений. Ч. 1-6.
3. ГОСТ 8.315-97 "ГСИ. Стандартные образцы состава и свойств и материалов. Основные положения".
4. РМГ 76-2004 ГСИ. Внутренний контроль качества результатов количественного химического анализа. Стандартиформ, 2006.
5. *Чарыков А.К.* Математическая обработка результатов химического анализа. Л.: Химия, 1984. С. 33.
6. *Прохоров Ю.В., Розанов Ю.А.* Теория вероятностей. М.: Наука, 1987. С. 352.
7. *Буйташ П., Кузьмин Н.М., Лейстнер Л.* Обеспечение качества химического анализа. М.: Наука, 1993. 188 с.
8. ГОСТ 8.563-96. "ГСИ. Методики выполнения измерений". – М.: Изд-во стандартов.
9. Федеральный закон о техническом регулировании в редакции от 27.12.2002 № 184-ФЗ (с поправками от 01.05.2007 № 65ФЗ).
10. Федеральный закон Российской Федерации «Об обеспечении единства измерений» N 102-ФЗ от 26 июня 2008 года.
11. *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики. М., Наука. 1983. 416 с.