

К ВОПРОСУ ОБ ОГРАНИЧЕННОСТИ ПОГРЕШНОСТЕЙ ИЗМЕРЕНИЯ ИЛИ МОДЕЛИРОВАНИЯ

Дементьев В.А.

Институт геохимии и аналитической химии им. В.И.Вернадского РАН, г.Москва

d_vasily@mail.ru

Предлагается считать, что физические измерения следуют закону: В условиях, когда физическая величина стабильна, когда устранены систематические ошибки, случайные погрешности измерений всегда ограничены. Предпринята попытка показать, что это может быть объяснено на основе фундаментальных физических законов. Сделан ряд выводов общего характера, имеющих важное практическое значение для оценки точности прогноза поведения сложных физических моделей.

Пусть случайная величина x распределена предположительно нормально с параметрами θ , σ . Такая ситуация соответствует, например, эксперименту, в котором гистограмма многочисленных центрированных результатов измерения по форме очень близка к кривой нормального распределения. Однако в многочисленных экспериментах замечено, что отклонения x от среднего никогда не выходят за пределы σ .

Описанная ситуация является типичной для современной техники эксперимента, когда исследователь получает огромный массив однородных данных, но никогда не наблюдает неограниченно больших отклонений от среднего, как предсказывает закон нормального распределения. Возникает подозрение, что результаты многочисленных измерений физически стабильной случайной величины x , проводимых в одинаковых условиях, подчиняются скорее не нормальному закону распределения, а какому-то иному. Обзор таких наблюдений приведен в докторской диссертации О.Е. Родионовой [1]. Там же высказано предположение, что ограниченность погрешностей измерений представляет собой некую природную закономерность, которую необходимо как-то обосновать, чтобы превратить в закон хеометрики, а также необходимо продумать процедуры учета этой закономерности при построении статистических и хеометрических методик обработки данных большого объема и высокой сложности.

В данной работе предлагается считать, что многочисленные наблюдения за поведением результатов измерений с помощью физических приборов позволяют сформулировать следующий закон.

В условиях, когда физическая величина по природе стабильна, когда устранены систематические ошибки прибора, случайные погрешности измерений всегда ограничены.

Предпринята попытка показать, что данный закон либо является фундаментальным, либо может быть объяснен на основе фундаментальных физических законов. Это позволит с высокой надежностью сделать ряд выводов общего характера, имеющих важное практическое значение для оценки точности прогноза поведения сложных физических моделей. Предлагается называть приведенное выше утверждение законом ROxY (как сокращение от Rodionova Oхана Yevgenievna).

Методы вывода статистических следствий из закона ROxY

Прежде всего, зададимся следующим вопросом. Как изменятся статистические свойства случайных величин, связанных с x , если заменить в рассуждениях и расчетах модельное нормальное распределение более реалистичным распределением, учитывающим ограниченность отклонений x ?

Можно поступить несколькими различными способами. Например.

1. Реализовать предложение О.Е. Родионовой, обрезав область определения x :

$$-\sigma \leq x \leq \sigma. \quad (1)$$

Тогда распределение перестанет быть нормальным, и могут возникнуть трудности, связанные с его плохой аналитичностью.

2. Оставить модельное распределение нормальным, но сузить его так, чтобы практически исключить выход x за указанные пределы. Для этого достаточно просто пользоваться нормальным распределением с параметрами

$$a = 0, \sigma_1 = m\sigma, m < 1. \quad (2)$$

Тогда все формулы статистики окажутся неизменными.

3. Отказаться от вывода аналитических выражений для оценки статистических свойств любых функций от результатов измерений. Вместо этого предлагается непосредственно оценивать статистические характеристики таких функций с помощью имитационных вычислительных экспериментов. Для этого следует использовать те же самые вычислительные средства, с помощью которых вычисляются сами значения функций. Эти средства следует лишь дополнить генератором случайных погрешностей, действующим по предполагаемому закону распределения погрешностей исходных величин.

На рисунке показаны два предлагаемых в пунктах 1 и 2 модельных распределения. Ограниченный закон распределения получен из нормального при $\sigma = 1$ и $s = 2.5$ с последующей перенормировкой исходного распределения. Для этого модельного распределения может быть выведено аналитическое выражение, справедливое в области (1).

$$\rho_s(x, 0, \sigma) = \frac{1}{\sqrt{2\pi}\sigma \operatorname{erf}\left(\frac{s}{\sqrt{2}}\right)} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (3)$$

При сравнительно больших величинах s значение интеграла вероятностей $\operatorname{erf}(s/\sqrt{2})$ очень близко к 1, следовательно, обрезание далеких крыльев нормального распределения никак не должно сказаться на всех статистических результатах, полученных с использованием модельного распределения (3). Показанное на рисунке при $s = 2.5$ модельное распределение в области (1) буквально сливается и исходным, неусеченным распределением. Например, для $x = 0$ формула (3) дает 0.404, а исходное распределение дает 0.3989.

Однако, несмотря на близость кривых в области определения (1), интегральные характеристики распределения (3) отличаются от характеристик исходного нормального распределения. Например, в нашем случае стандартное отклонение величины x из области определения (1) составляет не 1, а 0.9602. Эта оценка получена численным интегрированием. Вывод же аналитического выражения для дисперсии распределения (3) уже является довольно громоздким, поскольку приходится учитывать параметр s , входящий в нормировочный множитель.

Отсюда видно, что исследование свойств распределения (3) и получение формул для всех других распределений, являющихся производными от (3), затрудняет работу с таким способом учета ограниченности области определения случайной величины.

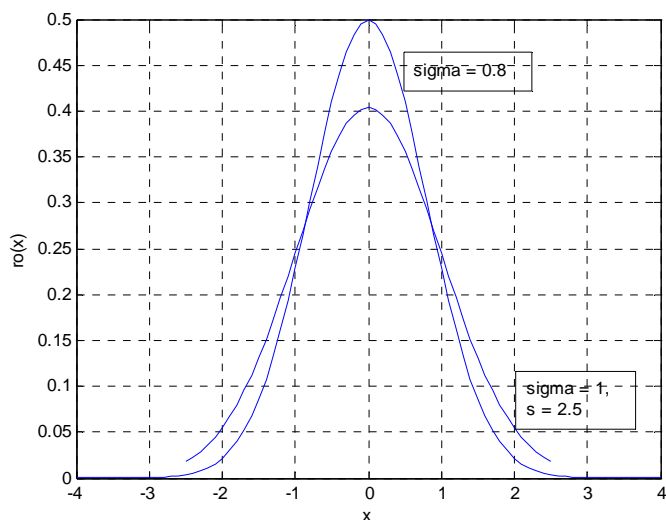


Рис. 1. Ограниченное «нормальное» распределение в сравнении с зауженным нормальным распределением.

Второй способ, упомянутый выше, и сводящийся к изменению параметра σ модельного нормального распределения на меньший параметр σ_1 в соответствии с (2), представляется более легким в исполнении. Он является и более перспективным при его использовании в задачах многомерной статистики и хемометрики.

На рисунке показано нормальное распределение с параметром $\sigma_1 = 0.8$. Оно дает очень мало шансов получить в эксперименте значения x , выходящие за пределы области (1). То есть вполне удовлетворительно отражает наблюдаемое в эксперименте поведение случайной величины x . В то же время оно обладает всеми привычными свойствами нормального распределения. При его интегрировании по области определения (1) в данном примере получается 0.9602, что не очень отличается от 1. Стандартное отклонение случайной величины x теперь хорошо известно. Оно равно 0.8, что опять-таки не очень сильно нарушает экспериментальные представления о разбросе измеренных значений величины x .

Конечно, зауженный закон нормального распределения, как хорошо видно на рисунке, дает завышенный прогноз получения в эксперименте малых отклонений и заниженный прогноз больших отклонений величины x . Но зато он не дает и такого неадекватного предсказания, как закон (3) – вероятность получить в эксперименте отклонение от -2.5 до -2.4 равна 0.002, а вероятность получения любого значения, меньшего -2.5, в точности равна нулю. Это не только непонятный резкий скачок в поведении непрерывной случайной величины. Так в эксперименте могут вести себя только дискретные величины. Но это и неприятное неаналитическое поведение плотности распределения (3), не позволяющее теоретически осмыслить все статистические следствия использования в последующих работах данных о поведении величины x с ограниченной областью определения (1). А такая задача характерна для случая использования результатов измерений в качестве параметров многомерных моделей, создаваемых для прогнозирования сложных явлений.

Из рисунка видно, что зауженное нормальное распределение более реалистично предсказывает поведение случайной величины на границе области наблюдаемых в эксперименте значений. Например, вероятность получить в эксперименте отклонение от -2.5 до -2.4 равна $4.6588 \cdot 10^{-4}$. Для экспериментатора это, возможно, исчезающе малая величина. В соседний интервал от -2.5 до -2.6 результат измерения может попасть с вероятностью $3.1571 \cdot 10^{-4}$. Тоже исчезающе малая вероятность. Но и вблизи границы области определения, для интервала от -2.4 до -2.3 закон предсказывает разумную очень малую вероятность $6.7687 \cdot 10^{-4}$. Дальше, с приближением к центру области, мы получаем все более резкое

возрастание вероятности получения результатов. Это более похоже на привычное естествоиспытателям поведение случайных непрерывных величин. А привычный аналитический вид предлагаемого модельного зауженного нормального распределения позволит исследователю, при построении сложных моделей, воспользоваться всеми готовыми результатами математической статистики, в том числе, многомерной.

Предложение, описанное в пункте 3, по-видимому, наиболее пригодно для разработки конкретных аналитических методик, когда достаточно один раз оценить точность прогноза поведения функций от измеряемых в эксперименте величин. В таких методических работах нужно получить именно численные оценки точности результата. Поэтому можно вообще обойтись без сложных выводов выражений для статистических характеристик вычисляемых величин.

Физические основы закона ROxY

Можно привести ряд соображений в пользу предположения, что закон ROxY является следствием фундаментальных законов физики.

В теории информации показано, что создание в природе любой единицы информации связано с затратами энергии. Если мы предположим, что измерительный процесс порождает неограниченное множество различных экземпляров погрешности измерения, то нужно будет сделать вывод о бесконечно больших затратах энергии на генерацию таких реальных событий. Это противоречит закону сохранения энергии, который приложим к сравнительно замкнутой системе объект - измерительный прибор.

Предположим, что физический прибор измеряет скорости молекул, вылетающих из малого отверстия в печи, заполненной химически однородным газом. Будем считать прибор идеальным. Тогда разброс в значении скоростей молекул будет связан со свойствами распределения Максвелла. Аналитическое выражение для этого закона дает неограниченное распределение абсолютных значений скоростей молекул. Но это абстракция, некое приближение к реальной действительности. Ни при какой высокой температуре скорость молекулы не может достигать скорости света. Следовательно, реальное, фиксируемое в эксперименте, распределение скоростей молекул получится ограниченным.

Линейный размер нанокристаллита некоего минерала по результатам измерений с помощью любого микроскопа не может выйти за определенные пределы, поскольку вдоль любого кристаллографического направления укладывается целое число элементарных ячеек. Если же учесть возможные дефекты, то можно говорить о целом числе молекул или атомов, уложенных вдоль этого направления.

При распаде радиоактивного препарата число регистрируемых распадов за какой-то промежуток времени не может быть произвольно большим. Это число никак не может превзойти общее число неустойчивых ядер в препарате.

Из этих примеров видно, что, наряду с общими законами природы, в конкретном процессе измерения всегда действует определенный физический механизм, ограничивающий разнообразие результатов измерений. Конечно, на конкретных примерах невозможно доказать справедливость предполагаемого закона природы, как принципиально нельзя логически доказать любой фундаментальный закон. Однако приведенные примеры показывают, почему мы можем считать закон ROxY весьма универсальным свойством любого процесса измерения.

Моделирование и прогнозирование сложных природных явлений в свете закона ROxY

Конечной целью любой науки является построение процедур, позволяющих прогнозировать сложные природные и технические явления. Это избавляет общество от

больших материальных и человеческих затрат на приобретение эмпирических знаний и умений.

- «Учись, мой сын – наука сокращает...»

Так умный А. Пушкин показал нам умного Б. Годунова, не затратив на это и десяти слов.

Известно, что научный прогноз всегда содержит в себе некоторую ошибку. Эта ошибка складывается из нескольких составляющих.

1. Прогноз основан на модели, которой оперирует научная теория. Модель всегда отлична от реальности. Теория всегда стремится отбросить несущественные детали описываемого явления. Отсюда возникают ошибки моделирования. Их следовало бы отнести к классу систематических погрешностей. Однако при прогнозировании различных вариантов исследуемого явления эти ошибки варьируют свои величины и знаки, то есть напоминают случайные погрешности. Конечно, моделирование может давать и систематическую погрешность. Такая погрешность выявляется в процессе решения обратных задач, когда прогноз сравнивают с известным результатом эксперимента. Если никакая вариация параметров модели и теории не приближает прогноз к эксперименту, исследователь вынужден изменить систему параметров модели или основные положения теории. Тем самым систематическая погрешность устраняется. Далее мы будем считать, что в рассуждениях не следует учитывать систематические погрешности моделирования.
2. Любая модель содержит некие параметры, значения которых получают из экспериментов путем решения обратных задач. Известно, что обратные задачи являются некорректными и всегда дают неопределенности в своих решениях, даже если опираются на идеальный эксперимент. Таким образом, прогноз основан на несколько неопределенных данных, заложенных в модель. Эта неопределенность усугубляется погрешностями того экспериментального материала, на основе которого решаются обратные задачи ради добычи параметров модели.
3. Вычислительные процедуры, используемые в прогнозе, бывают сложными, используют приближенные методы расчетов. Следовательно, прогнозотягчен вычислительными погрешностями моделирования.

Ошибки, происходящие из всех указанных источников, как-то комбинируются в вычислениях, приводя к ошибке прогноза. Возникает вопрос – могут ли эти комбинации приводить к неограниченной ошибке прогноза?

Можно сразу обрисовать такую ситуацию, когда прогноз приводит к неопределенно большому расхождению с действительностью и становится совершенно несостоятельным. Это ситуация очень далекой экстраполяции. Параметры модели находят в одних физических условиях, а прогнозируют поведение модели в совершенно других условиях, когда сама модель и теория становятся неадекватными реальной действительности. Исключим такую патологическую для современного естествознания ситуацию из рассмотрения. Тогда можно убедиться, анализируя вышеприведенные причины возникновения ошибок прогноза, что при использовании работоспособной модели, адекватной теории для интерполирования и близкого экстраполирования характеристик поведения изучаемого явления ошибки прогноза будут с неизбежностью ограниченными.

Последнее утверждение нельзя доказать в общем случае. Но его справедливость подтверждается всей практикой современного естествознания. Проиллюстрируем данное утверждение примером прогнозирования поведения очень сложной механической системы.

В качестве примера моделирования в вычислительных экспериментах приведем результаты расчетов для кристаллита алмаза [2, 3]. Кристаллит алмаза является удобным объектом для данного исследования, поскольку параметры его элементарной ячейки известны и могут быть заданы в программах с математической точностью. Такая точность сохраняется при произвольном наращивании размеров кристаллита. Следовательно, все возможные ошибки вычислений связаны с машинной неточностью вычисления элементов

матрицы T_p (T_s) и с диагонализацией этой матрицы. Еще данная модель удобна и потому, что колебательные координаты типа растяжений валентных связей и изменения валентных углов полностью описывают нормальные колебания модели, как бы ни наращивались размеры этой модели, лишь бы прирост выражался некоторым целым числом элементарных ячеек. При этом правильно описываются движения как внутренних, так и поверхностных атомов модели.

Одним из преимуществ модели кристаллита алмаза является ее высокая симметрия. Это позволяет для кубической модели автоматически получать представление ее гамильтониана в виде 10 независимых блоков матрицы T_s , а для модели в форме параллелепипеда – в виде 6 блоков. Это позволяет строить модели кристаллитов, содержащих до нескольких сотен атомов, описывать их колебания несколькими тысячами естественных колебательных координат.

Приведем результаты для трех моделей алмазных кристаллитов. Нитевидные модели получали, повторяя элементарную ячейку в направлении одного вектора трансляции. При этом две соседние элементарные ячейки имеют один общий атом. Модели обладают довольно высокой гибкостью, поэтому с ростом длины нити быстро уменьшаются как собственные числа матрицы T_s , так и частоты колебаний. Для данного исследования это наиболее интересная «одномерная» модель, поскольку погрешности вычисления малых собственных чисел матрицы T_s могут оказаться столь заметными, что определяют вычислительный предел, лишая модель физического смысла.

Пленочные модели получали, повторяя элементарную ячейку в направлении двух векторов трансляции. При этом соседние элементарные ячейки имеют несколько общих атомов. Модели обладают менее высокой гибкостью, чем нитевидные. Тем не менее, даже при небольших размерах моделей наблюдаются низкочастотные колебания, охватывающие всю «двумерную» пленку. Поэтому и для таких моделей можно было ожидать выхода на вычислительный предел, лишая модель физического смысла.

«Трехмерные» модели получали, повторяя элементарную ячейку в направлении трех векторов трансляции. Эти модели имеют высокую жесткость, поэтому в проведенных экспериментах трудно было ожидать выхода на стремящиеся к нулю частоты колебаний, принадлежащих акустическим ветвям. Тем не менее, и для этих моделей было интересно проследить за поведением собственных чисел матрицы T_s .

Компьютерные эксперименты с вышеописанными моделями ставились следующим образом. Для каждой модели вводились все координаты растяжений валентных связей между ближайшими атомами, а также все координаты изменений углов между всеми валентными связями. Эта система обобщенных колебательных координат полностью соответствует химическим представлениям об атомных взаимодействиях в кристаллите алмаза. Система таких координат содержит много «лишних» координат, поскольку модель включает в себя множество циклов со структурой кольца циклогексана. Если кинематическая часть гамильтониана (матрица T_s) составлена с математической точностью, то некоторое количество собственных чисел этой матрицы будут нулевыми. Однако все вычислительные процедуры выполняются с ограниченной точностью, поэтому линейные зависимости в какой-то мере нарушаются. В результате ни одно из собственных чисел матрицы T_s может не получиться нулевым. Возникает задача подобрать некий дискриминант и с его помощью отбросить те малые ненулевые собственные числа, которые соответствуют искаженным линейным зависимостям между колебательными координатами. Это можно сделать, потребовав, чтобы на выходе программы диагонализации получалось теоретически ожидаемое число нормальных колебаний m , равное $m = N - 6$, где N – число атомов в модели.

В результате расчетов можно отметить верхнюю и нижнюю границы «нулевых» собственных чисел, отвечающих зависимостям между естественными колебательными координатами модели, и нижнюю границу тех собственных чисел, которые соответствуют

нормальным колебаниям модели. Может случиться, что для некоторых моделей эти границы соприкоснутся, либо пересекутся, и тогда будет трудно назначить дискриминант для правильного решения указанной проблемы. Такое сближение будет более характерным для «одномерных», нитевидных моделей, где из-за высокой гибкости алмазной нити, толщиной в одну элементарную ячейку, с ростом длины нити будут появляться очень низкие по частоте нормальные колебания. Это явление могло быть характерным и для «двумерных», пленочных моделей толщиной в одну элементарную ячейку. Для «трехмерных» моделей алмаза, где изгибы всего кристаллита затруднены, такое явление должно быть менее характерным.

Ниже приведены результаты численных экспериментов. Отметим, прежде всего, что решения колебательных задач с учетом и без учета симметрии получаются одинаковыми. Отсюда следует, что при работе с нерегулярными молекулярными средами получатся сходные результаты.

В результате проведенных экспериментов показано, что для всех исследованных моделей дискриминант, равный $1 \cdot 10^{-6}$, обеспечивает удачный поиск и исключение «нулевых» собственных чисел кинематической части молекулярного гамильтониана. Подробные данные об исследованных моделях приведены в таблице 1.

Таблица 1. Характеристики моделей и результаты расчетов при значении дискриминанта, равном $1 \cdot 10^{-6}$

Модель алмаза, число ячеек	Атомов	Координат	Частот	Наименьшее собственное число матрицы T_s	Отбросить числа в диапазоне min, max	Наименьшая частота, см^{-1}
2	27	96	75	1.8e-3	-4e-8, 7e-8	54.6
4	53	196	153	3e-4	-6e-8, 5e-8	21.7
8	105	396	309	3e-5	-7e-8, 6e-8	6.4
16	209	796	621	1.8e-6	-7e-8, 7e-8	1.7
32	417	1596	1245	1.23e-7	-7e-8, 7e-8	4.5e-1
2x2	50	202	144	3e-3	-8e-8, 1e-7	73
4x4	182	850	540	7e-4	-8e-8, 9e-8	33
8x8	686	3490	2052	7.4e-5	-9e-8, 9e-8	11
4x4x2	327	1812	975	2e-3	9e-8, 1e-7	73.5
4x4x4	617	3736	1845	2.5e-3	-2e-7, 2e-7	103

Более подробную информацию о поведении моделей дают гистограммы собственных чисел матриц T_s исследованных моделей. На рисунке 2 приведены гистограммы двух множеств собственных чисел матрицы T_s модели алмаза размером 4x4x4 элементарных ячейки.

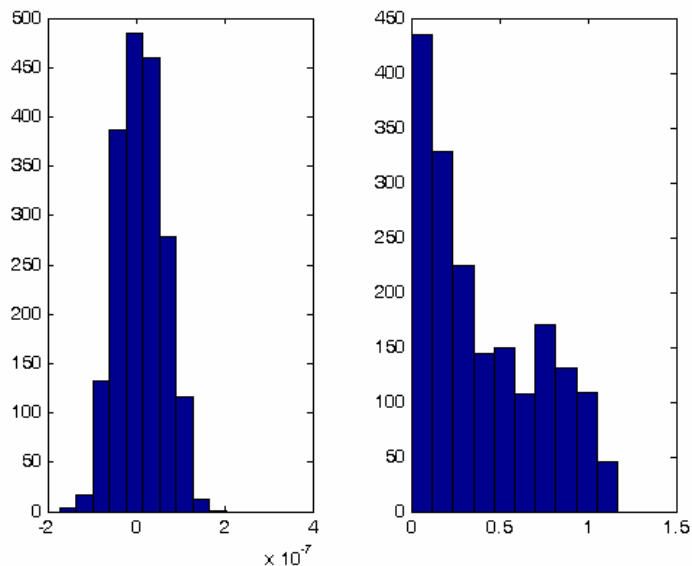


Рис. 2. Слева – гистограмма малых собственных чисел матрицы T_s модели алмаза размером $4 \times 4 \times 4$ элементарных ячейки. Эти числа соответствуют искаженным из-за погрешностей вычислений линейным зависимостям между колебательными координатами модели. Справа – собственные числа матрицы T_s , соответствующие нормальным колебаниям модели

Слева на рисунке дана гистограмма «нулей». Эти числа лежат ниже выбранного значения дискриминанта и отброшены программой на выходе из процедуры диагонализации. Справа на рисунке дана гистограмма тех собственных чисел матрицы T_s , которые соответствуют нормальным колебаниям модели. Аналогичные данные были получены и для других исследованных моделей алмазных кристаллитов. Из этих данных видно, что распределение «нулей» близко к нормальному распределению, причем дисперсия распределения не меняется с ростом размера модели. Еще более важно, что разброс «нулей» не разрастается с ростом размеров модели. То есть, в этом эксперименте проявляет себя нормальный закон распределения с ограничением крыльев. Это весьма обнадеживающий факт. Он позволяет надеяться, что при анализе моделей еще больших размеров не возникнет опасность расплывания искаженных «нулей» и результаты анализа не будут погублены по причине смешивания «нулей» с настоящими собственными числами матрицы T_s .

Из рисунка 2 также видно, что настоящие собственные числа матрицы T_s распределены по собственному закону, определенному структурой этой матрицы. Большое относительное количество малых собственных чисел отражает физические особенности крупной молекулярной модели. Сдвиг нижней границы в сторону уменьшения при росте размеров модели закономерен и необходим. Именно такой сдвиг может обусловить предел применимости описанной методики расчета. Однако имеются другие факты, которые показывают, что этой угрозы можно избежать, разумно выбирая размер модели при решении конкретной исследовательской задачи. Такие факты легче всего увидеть с помощью гистограмм частот колебаний исследуемых моделей. Эти гистограммы приведены на рисунках 3 и 4.

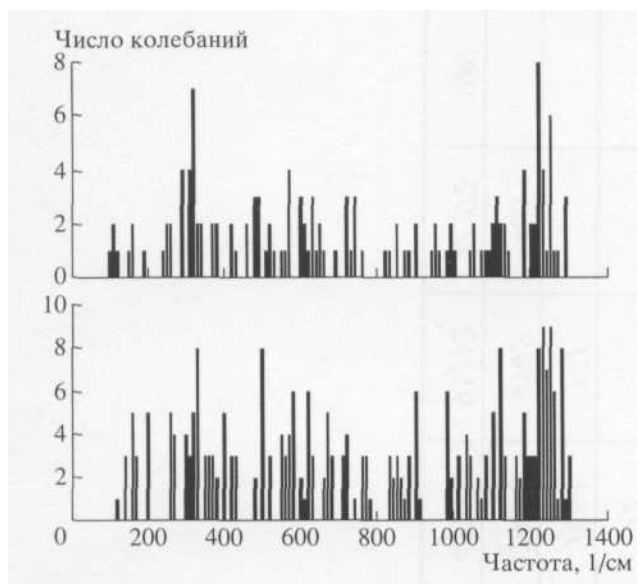


Рис. 3. Гистограммы распределения плотности колебаний на шкале частот для кристаллической модели алмаза размера 2x2x1 и 2x2x2 ячеек. Коэффициент корреляции между гистограммами составляет 0.3351. "Кристаллическое насыщение" еще не проявляется

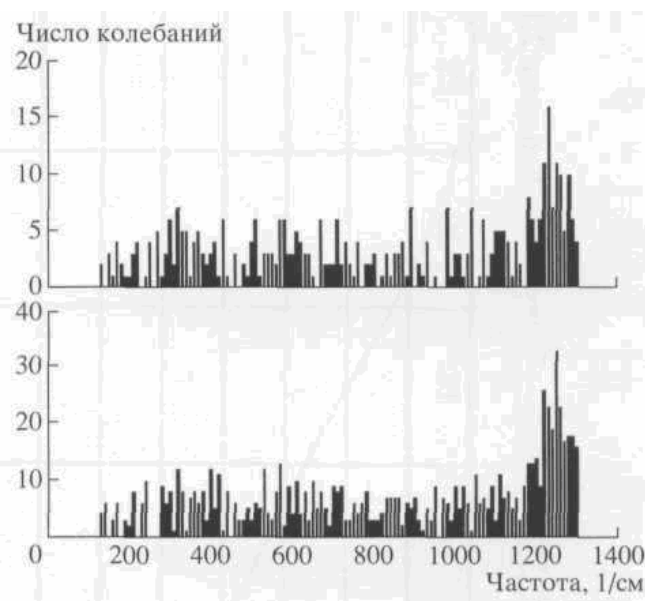


Рис. 4. Гистограммы распределения плотности колебаний на шкале частот для кристаллической модели алмаза размера 3x2x2 и 3x3x3 ячеек. Коэффициент корреляции между гистограммами составляет 0.5897. Сходство гистограмм возрастает.

Заклучение

На основании приведенных соображений и конкретных результатов моделирования поведения сложного физического объекта можно предложить более общую формулировку закона РОХУ.

В условиях, когда физическая величина по природе стабильна, когда устранены систематические ошибки прибора, случайные погрешности измерений всегда

ограничены. Использование результатов экспериментов при построении адекватных моделей и теорий с целью прогнозирования сложных явлений в условиях интерполяции и близкой экстраполяции физических условий приводит к ограничению погрешностей прогноза.

Если верно это утверждение, то мы впервые в естествознании получаем фундаментальный закон, доброжелательный к исследователю. В отличие от, например, законов сохранения, которые предупреждают исследователя – «Ты не можешь произвольно сделать то-то и то-то, поскольку ты не можешь взять ниоткуда энергию и импульс в замкнутой системе». Закон РОхУ поощряет исследователя – «Ты можешь строить удобные тебе адекватные модели и теории, с их помощью прогнозировать то-то и то-то. Твой прогноз будет содержать ошибки, но эти ошибки обязательно будут ограничены какими-то рамками. Будь только добр, определить эти рамки доступными тебе средствами».

Работа выполнена в рамках Программы № 14 Президиума РАН.

Литература

1. *Родионова О.Е.* Докторская диссертация. М., 2007.
2. *Грибов Л.А., Дементьев В.А.* Программный комплекс для прямого расчета колебательных состояний кристаллических структур в молекулярном приближении. Кристаллография. 2004, Том 49, № 5, с. 960-967.
3. *Грибов Л.А., Дементьев В.А.* Поиск и исключение зависимостей между колебательными координатами при моделировании спектров больших молекул. ЖПС. 2006, Том 73, № 5, с. 561-565.