

О метрологических характеристиках исторических данных

Михаил Михайлович Елисейкин^{1✉}, Валерий Фёдорович Очков²

^{1,2}Московский энергетический институт (национальный исследовательский университет), Москва, Россия

¹muxa@muxa.ru ✉, <https://orcid.org/0000-0001-8754-0699>

²OchkovVF@mpei.ru

Аннотация. Вследствие широкого применения искусственного интеллекта и машинного обучения в промышленности проблема метрологического дефицита в больших данных приобретает особое значение. По мере накопления данных о технологических процессах будет возрастать важность информации о том, какими средствами измерений, по каким методикам, с какой точностью и в каких единицах измерений были получены значения физических величин, используемых при их анализе. Однако сама по себе проблема отсутствия метрологической совместимости у накопленных данных не нова. По сути, получение результатов измерений с требуемой точностью и воспроизводимостью — одна из главных проблем современной науки и техники, которую позволяет решать метрология. История знает множество примеров того, что научную ценность могут иметь не только данные, накопленные за несколько лет, но и данные, собиравшиеся десятилетиями и даже веками. Приведены примеры, показывающие, что утрата метрологических характеристик может сделать собранные данные бесполезными, а своевременное вмешательство и коррекция могут вернуть данным метрологическую совместимость и сделать их пригодными для ретроспективного анализа. Приведённые примеры охватывают период с 18-го по 21-й века и различные научные и прикладные направления.

Ключевые слова: история метрологии, история науки, метрологический дефицит

Финансирование: отсутствует.

Для цитирования: Елисейкин М. М., Очков В. Ф. О метрологических характеристиках исторических данных // Законодательная и прикладная метрология. 2024. № 5. С. 47–51. <https://doi.org/10.32446/2782-5418.2024-5-47-51>

About metrological characteristics of historical data

Mikhail M. Eliseikin^{1✉}, Valerii F. Ochkov²

^{1,2} Moscow Power Engineering Institute (National Research University), Moscow, Russia

¹muxa@muxa.ru ✉, <https://orcid.org/0000-0001-8754-0699>

²OchkovVF@mpei.ru

Abstract. Due to the widespread application of machine learning technologies in industry, the issue of metrological deficit in big data has gained particular significance. As more data on technological processes accumulate, the importance of knowing the tools, methodologies, precision, and units of measurement used to obtain the physical quantities analyzed will increase. However, the problem of metrological incompatibility in accumulated data is not new. Essentially, obtaining measurements with the required and reproducible accuracy is one of the main challenges of modern science on the way of its formation, and metrology, as a separate science, exists precisely to address this issue. History provides numerous examples where data collected not only over several years but also over decades or even centuries can hold significant scientific value. This article gathers several examples showing that the loss of metrological characteristics can make collected data useless, while timely intervention and correction can restore their metrological compatibility and make them suitable for retrospective analysis. The examples span the period from the 18th to the 21st centuries and cover various scientific and applied fields.

Keywords: history of metrology, history of science, metrological deficit

Financial Support: none.

For citation: Eliseikin M. M., Ochkov V. F. About metrological characteristics of historical data. *Legal and applied metrology*. 2024, no. 5, pp. 47–51. <https://doi.org/10.32446/2782-5418.2024-5-47-51>

Введение. Поднятый в статье [1] вопрос метрологического дефицита в накапливаемых данных становится особенно актуальным в настоящее время в связи тем, что исторические данные могут быть основой процессов управления промышленными киберфизическими системами.

Однако следует заметить, что проблема отсутствия метрологической совместимости данных, полученных с большим временным разрывом, была всегда одной из важнейших задач, решаемых метрологией на всём пути её развития. В статье приводится ряд примеров, в которых будут показаны последствия утраты метрологических характеристик в исторических данных и возможность их корректировки для достижения метрологической совместимости собранных данных.

Для метрологов в приведённых примерах вряд ли будет что-то новое, однако внедрение практики устранения метрологического дефицита зависит не только от метрологов, но и от людей, далёких от метрологии – создателей систем мониторинга производства и тех, кто будет создавать программное обеспечение для этих систем.

Единицы измерений. Одной из основных проблем измерений является существование разных систем единиц измерений. Эта проблема ассоциируется с разбившимся 25 лет назад космическим аппаратом Mars Climate Orbiter и с аварийной ситуацией рейса AIR CANADA 143, произошедшей более 40 лет назад. Однако это не единственные примеры, которыми можно проиллюстрировать проблемы, связанные с единицами измерений.

В конце 17-го — начале 18-го века в Европе предпринимались попытки определить точные параметры эллипсоида вращения, соответствующего реальной форме планеты Земля. Для этого на разных широтах осуществлялись измерения длины дуги, соответствующей 1° широты. Сопоставление длин дуг 1° на разных широтах должно было показать, насколько различаются полуоси эллипса.

Важнейшим событием тех исследований стали две «градусные экспедиции», одна из которых поехала в Перу, на экватор, а другая — в Лапландию, в приполярные широты. Часто уточняется, что данные экспедиции при отправке получили абсолютно идентичные эталоны туаза (устаревшей французской меры длины, равной 1,949 м). Однако именно это уточнение показывает, что в других случаях градусные измерения могли производиться с иными эталонами туаза, не совпадающими с эталонами Перуанской и Лапландской экспедиций. А это значит, что у нас есть примерно 10 градусных измерений, проделанных французскими исследователями в 18-м веке, но мы не можем быть

уверенными в том, что во всех случаях используемые эталоны туаза имели одну и ту же физическую длину.

Проблема отсутствия метрологической совместимости градусных измерений была понятна давно, и уже в конце 19-го века видный русский геодезист и топограф Василий Васильевич Витковский, в статье «Градусные измерения» написанной для Энциклопедического словаря Брокгауза и Ефрона [2], отмечал, что геодезические градусные измерения начала 18-го века недостаточно точны, а более ранние градусные измерения использовали неверные методики.

Да, сейчас у нас есть точные параметры земного эллипсоида и системы спутниковой навигации, чьё функционирование подтверждает эти параметры, но это не значит, что старые градусные измерения перестали быть нужными.

Имея градусные измерения за 300 лет, мы могли бы отследить изменения параметров земного эллипсоида, что дало бы нам полезную информацию о нашей планете. Однако мы не можем этого сделать, потому что не знаем, чему именно равнялся туаз в каждом конкретном случае. У нас нет даже теоретической возможности перевести результаты тех измерений в используемые сейчас единицы измерений и использовать их для анализа.

К началу 20-го века развитие метрологии и промышленности привело к тому, что появилась возможность задавать точные размеры деталей, произведённых на разных заводах, в разных странах, на разных континентах. Эталоны и методы измерений достигли такого уровня точности, что мы можем быть уверены в том, что 1 мм в 1924 году это такой же 1 мм, что и в 2024 году. И хотя в разных странах могли использоваться разные единицы измерений, было известно их точное соотношение, что позволяло привести разные значения к общим единицам измерений.

Однако точность измерений или изготовления и последующая точность перевода величин в другие единицы измерений не гарантируют того, что в 2024 году мы можем получить точные исторические данные из 1924 года. Дело в том, что в дошедших до нас записях могут присутствовать не точные исходные значения, полученные в результате измерений или изготовления, а округлённые значения, полученные в результате перевода исходных значений в другие системы измерений.

Так, например, если посмотреть на характеристики двигателей внутреннего сгорания, представленные в первой половине 20-го века в США и странах Европы, то можно столкнуться с размерами, изначально представленными в миллиметрах и пере-

ведёнными в дюймы или наоборот. При этом миллиметры округлялись до 1 мм, а дюймы — до 0,1 дюйма.

В результате, при вычислении рабочего объёма двигателя из геометрических размеров поршней и цилиндров, один и тот же двигатель мог иметь в характеристиках разный рабочий объём. И хотя разница в несколько кубических сантиметров может казаться незначительной, она может повлиять на оценку ассортимента техники (две различных модели вместо одной) или её классификацию (499 куб. см и 501 куб. см могут быть в разных классах).

Таким образом, при невозможности провести повторные измерения, возрастает важность информации о том, какие именно единицы измерений были у первоначально полученного значения величины, и не была ли величина подвергнута переводу в другую систему единиц с последующим округлением.

Проблема округления при переводе значений из одних единиц измерений в другие имеет значение не только при анализе относительно отдалённых явлений, зафиксированных в исторических данных, но и при создании современных экспертных и рекомендательных систем.

В 2014 году на медицинской конференции [3] была представлена работа, посвящённая проблеме использования единиц измерений в медицинских данных. В основе работы лежали записи о 536 уникальных пациентах, собранные в период с 1986 по 1989 годы. Эти записи содержали данные с разными единицами измерений (метры и сантиметры), в разных системах единиц измерений (метрической и имперской) или вообще без единиц измерений. В работе была показана возможность автоматического приведения значений величин к одним и тем же единицам измерений для целей последующего автоматического диагностирования состояния пациентов.

Отдельно авторы акцентировали внимание на том, что после перевода в другие единицы измерений с последующим округлением, существует риск постановки диагноза, отличного от того, который был поставлен врачом в 1980-х. Речь шла об анализе индекса массы тела и, связанных с ним, рисков для здоровья пациента.

Проверка показала, что выводы, сделанные описанной автоматической системой, совпали с диагнозами, поставленными врачами. Однако, несмотря на отсутствие проблемы в данном конкретном случае, по мнению авторов упомянутой работы, эти риски стоит учитывать при дальнейшей разработке системы автоматической конвертации величин между разными единицами измерений, и осуществлять округления с учётом рекомендаций врачей.

То есть исторические данные, которые будет анализировать экспертная система, могут не просто быть конвертированы и округлены исходя из правил математики, но и подвергнуты некой коррекции. И было бы полезно сохранять изначальные данные в тех единицах измерений и с той точностью, которые были в момент их возникновения.

Точность средств измерений (СИ). Другой проблемой исторических данных является то, что мы не всегда знаем о том, какие СИ были использованы и насколько они точны.

Сейчас мы имеем дело с различными СИ, имеющими различные метрологические характеристики, но в каждом конкретном случае известно, о какой точности идёт речь. Однако так было не всегда.

Мы иногда слышим о погодных рекордах за всю историю наблюдения и чаще всего эта история наблюдений начинается в 19-м веке. Но несмотря на то, что результаты тех измерений могли быть записаны в известных единицах измерений, мы не можем быть уверены в их числовых значениях. Потому что, говоря об измерениях, сделанных в то время, мы можем иметь дело с измерениями, сделанными самодельными СИ.

Так, в инструкции 1841 года [4], составленной академиком А. Я. Купфером для горных офицеров, заведующих магнетическими обсерваториями, разделы, посвящённые тем или иным измерениям, начинаются с описания того, как изготовить измерительный прибор. Например, в разделе «Метеорологические наблюдения. Термометрические наблюдения» содержится подробное описание вспомогательного оборудования для определения температуры таяния льда и температуры кипения воды. Описывается, какое положение должно быть у термометра и как избежать влияния размеров термометра на процесс его поверки; как отводить пар от кипящей воды, чтобы он не мешал разметке шкалы; приведена таблица зависимости температуры кипения воды от атмосферного давления. Отдельно есть приложение с инструкцией о том, как изготовить прибор в случае, если термометрическая трубка имеет не цилиндрическую, а коническую форму.

Наверняка горный офицер, заведующий обсерваторией, имел необходимое образование и навыки, а упомянутая выше инструкция в точности соответствовала физической основе измерений соответствующих величин. То есть точность измерений зависела от навыков и добросовестности сотрудника, на месте изготавливающего для самого себя измерительный прибор. Это значит, что метрологически несовместимыми могут оказаться не только исторические

и современные данные, но и исторические данные, сделанные в одно и то же время.

К концу 19-го века в России стали предпринимать первые шаги по обеспечению единства измерений. Депо образцовых мер и весов было преобразовано в Главную палату мер и весов, вокруг которой выстраивалась единая государственная метрологическая система. Так же в 1893–1894 годах были приняты решения об обязательной поверке СИ в Главной физической обсерватории.

Может показаться, что проблема метрологической совместимости измерений осталась в прошлом, но это не так. Потому что только требований и обязательных процедур по поверке СИ недостаточно, требования и процедуры надо ещё и выполнять.

Проблемы с метрологическим обеспечением производства возникли в 1930-х годах в СССР, когда шло увеличение числа заводов и был большой дефицит квалифицированных рабочих кадров. В результате встречались ситуации, распространённые в 19-м веке, когда серийные машины и инструменты были несовместимы между собой в плане взаимозаменяемости запчастей.

Одним из путей решения проблемы стало принятие в 1938 году постановления Совета Народных Комиссаров «Об упорядочении измерительного хозяйства в Союзе ССР» [5]. Выстраивалась общесоюзная система по производству, поверке и ремонту измерительных инструментов, вводились меры по стимулированию работников.

Однако, если эти меры и дали ожидаемый результат, остался временной период с высоким риском недостоверности результатов измерений. И в случае ретроспективного анализа тех или иных результатов измерений, было бы полезно это учитывать. Даже когда есть качественное оборудование и квалифицированный персонал, по-прежнему есть вероятность возникновения проблем с достоверностью измерений.

В книге с описанием истории метеостанции в парке Blue Hills, находящемся недалеко от Бостона (США) [6], рассказывается о том, что её сотрудникам приходится сталкиваться с проблемой неопределённости измерений. Так, в 1959 году барометры метеостанции переехали с первого этажа на второй, а в 1977 году оказалось, что это изменило их показания на 0,3 миллибара. В результате соответствующее изменение (вместе с другой поправкой) было внесено во все записи с момента переезда на второй этаж в 1959 году.

В другой момент времени произошла замена типов СИ и методов, используемых для измерений скорости ветра. Последовавшая за этим проверка записей

показала, что часть сведений о скорости ветра можно считать неверными, а часть просто отсутствует. Для исправления этой ситуации исторические данные за несколько лет были заменены на новые, полученные путём пересчёта данных Бостонской метеостанции. Таким образом, исторические данные, с которыми приходится работать, могут быть не только неточно измеренными, но и уже скорректированными или синтезированными на основании каких-то других данных.

В данном случае может сложиться ситуация, когда при ретроспективном анализе будет обнаружена чёткая зависимость между скоростью ветра в Бостоне и в парке Blue Hills, которая, на самом деле возникла из-за того, что данные Blue Hills были синтезированы на основании идеи о существовании такой зависимости.

Знание о том, с помощью какого метода были получены данные, могло бы помочь избежать неверных выводов при анализе.

Методика измерений. Методика измерений может влиять не только на точность результатов измерения, но и на физический смысл полученных результатов.

Возможна ситуация, когда квалифицированный персонал с помощью точных СИ произвёл измерения в полном соответствии с методикой, но полученные данные будут несовместимы с аналогичными историческими данными, потому что эти данные, несмотря на одинаковое название величин и единицы измерения, будут иметь разный физический смысл.

Ярким примером такого несоответствия является ситуация с мощностью двигателей внутреннего сгорания, измеряемой по различным методикам. Так, в 2006 году японская компания Honda объявила, что в дальнейшем будет указывать мощность двигателей в соответствии со стандартом SAE J1349. Сравнение старой (японской) и новой (европейской) мощностей показало, что новое значение будет на 20 % — 30 % меньше прежнего. И если есть накопленные данные о работе предприятия, на основании которых можно выявить взаимосвязь между суммарной мощностью имеющихся двигателей и объёмом выполненных работ или произведённой продукции, то может возникнуть риск ошибочной оценки ситуации и неверных управленческих решений.

Учёт изменения производительности на единицу мощности и сопоставление с другими факторами позволяет планировать расходы и объём выполняемых работ исходя из доступных мощностей. Учитывать сохранённую информацию о том, по какой методике

проводились измерения, необходимо для анализа информации и планирования в будущем.

Выводы. В быту мы часто используем СИ с неустановленной погрешностью и считаем его показания достоверными. Часы отстают — можно выходить из дома раньше. Термометр завышает температуру — можно одеваться с поправкой на эту погрешность.

Но в промышленности приходится иметь дело с ситуациями, когда СИ много и они периодически заменяются, а персонал, ориентирующийся на их показания, может не знать об их индивидуальных особенностях.

В настоящее время мы находимся в ситуации, когда анализом данных и принятием управленческих решений начинают заниматься не люди, которые помнили, когда какой прибор ставили и какие у него особенности, а компьютерные программы, которые просто совершают вычисления на основании доступных им данных.

Поэтому для того, чтобы эти программы могли сделать свою работу, надо предоставить им информацию о том, как именно возникли анализируемые данные. Были ли это одни и те же СИ или череда нескольких? Каковы принципы работы СИ и методики измерений? В каких единицах измерений и с какой точностью были получены данные? Подвергались ли они корректировке? Данные получены в результате прямых измерений или косвенных?

Как видно из приведённых выше примеров, эти вопросы были актуальны всегда, но раньше ответы на них могли быть не важны, потому что их было невозможно проанализировать и из них было сложно извлечь полезную информацию.

Сейчас же данные, собиравшиеся в течение 5 лет, становятся ценными и незаменимыми. Мы можем извлечь из них пользу, но сделать это можно только в том случае, если 5 лет назад мы начали их собирать.

Поэтому «метрологический дефицит», то есть отсутствие информации о метрологических характеристиках, собранных и анализируемых данных, надо начинать устранять уже сейчас.

Список источников

1. Елисейкин М.М., Очков В.Ф. Метрологический дефицит в промышленных «больших данных» // Законодательная и прикладная метрология. 2024. № 4. С. 19–24. <https://doi.org/10.32446/2782-5418.2024-4-19-24>

2. Энциклопедический словарь. Т. 9а. Гравилат — Давенант. Под ред. проф. И. Е. Андреевского. Санкт-Петербург: Ф. А. Брокгауз, И. А. Ефрон, 1890–1907. 1893. С. 475–974.

3. Samadian S., McManus B., Wilkinson M. D. Automatic detection and resolution of measurement-unit conflicts in aggregated data // *BMC Med Genomics*. 2014; 7 (Suppl 1). S12. <https://doi.org/10.1186/1755-8794-7-S1-S12>

4. Купфер А. Я. Руководство к деланию магнетических и метеорологических наблюдений, составленное для горных офицеров, заведующих магнетическими обсерваториями, академиком А. Купфером. Санкт-Петербург: типография Экспедиции заготовления гос. бумаг, 1841. 153 с.

5. Собрание постановлений и распоряжений правительства СССР за 1938 г. № 29–58. Отдел первый. Совет народных комиссаров Союза ССР. М.: Б. и., б.г. 296 с.

6. Conover J. H. *The Blue Hill Meteorological Observatory: The First 100 Years, 1885–1985*. American Meteorological Society, Boston, Massachusetts. 1990 514 p.

References

1. Eliseykin MM, Ochkov V. F. Metrological deficit in industrial “big data”. *Legal and applied metrology*. 2024;(4):19–24. (In Russ.)

2. Andreevskii I. E. (ed.) *Entsiklopedicheskii slovar'*. Vol 9a. Gravitat — Davenant. Saint Petersburg: F. A. Brokgauz, I. A. Efron, 1890–1907 Publ., 1893, pp. 475–974. (In Russ.)

3. Samadian S, McManus B, Wilkinson MD. Automatic detection and resolution of measurement-unit conflicts in aggregated data. *BMC Med Genomics*. 2014;7(1): S12.

4. Kupfer AYa. *Rukovodstvo k delaniyu magneticheskikh i meteorologicheskikh nablyudenii, sostavlennoe dlya gornyx ofitserov, zaveduyushchikh magneticheskimi observatoriyami, akademikom A. Kupferom* [Manual for making magnetical and meteorological observations, compiled for mountain officers in charge of magnetical observatories by Academician A. Kupfer]. Saint Petersburg: tipografiya Ekspeditsii zagotovleniya gos. bumag Publ., 1841, 153 p. (In Russ.)

5. *Sobranie postanovlenii i rasporyazhenii pravitel'stva SSSR za 1938 g. № 29–58* [Collection of Resolutions and Orders of the Government of the USSR for 1938]. Otdel pervyi / Sovet narodnykh komissarov Soyuzo SSR. Moscow: [unpublished], 296 p. (In Russ.)

6. Conover JH. *The Blue Hill Meteorological Observatory: The First 100 Years, 1885–1985*. Boston, Massachusetts, American Meteorological Society, 1990, 514 p.

Дата принятия 10.09.2024