

УДК 550.3+519.246.8+53.088+004.622

PACS: 05.45.Tr, 06.20.F-, 07.05.Kf

ФРАГМЕНТАЦИЯ ВРЕМЕННЫХ РЯДОВ – НЕ АНОМАЛИЯ, А НОРМА

© 2024 г. М.М. Елисейкин, В.Ф. Очков

*Национальный исследовательский университет “Московский энергетический институт”,
г. Москва, Россия*

Автор для переписки: М.М. Елисейкин, e-mail: muxa@muxa.ru

Поступила в редакцию 08.11.2024 г.; после доработки 17.12.2024 г.

Принята к публикации 18.12.2024 г.

Аннотация. Современные технологии сделали сбор и обработку данных научных наблюдений простыми и доступными. Существующие решения позволяют быстро проводить предварительную обработку собранных данных, очищая их от выбросов, связанных с ошибками измерения, и заполняя пробелы в фрагментированных временных рядах. Однако лёгкость, с которой это делается, создаёт риски бездумного использования таких решений. В результате могут быть отброшены данные, свидетельствующие о реальной физической аномалии, а при объединении фрагментов временного ряда могут возникнуть данные, не соответствующие наблюдаемым процессам и явлениям. Подобная ситуация была оправдана раньше, когда был дефицит вычислительных мощностей: отбрасывание заведомо недостоверных значений и заполнение пробелов упрощали и ускоряли анализ данных. Сейчас же, когда вычислительные мощности есть в достаточном количестве, можно начинать искать закономерности также и в том, что раньше считалось ошибкой наблюдений и отбрасывалось. Кроме того, объём накапливаемых данных может позволить начать рассматривать фрагменты временных рядов как части закономерного процесса, не заполняя при этом пробелы искусственными данными, созданными на основании наших предположений о характере наблюдаемых процессов и явлений. Всё это ставит вопрос о необходимости адаптации подходов, используемых при сборе и анализе результатов наблюдений, к тем возможностям, которые дают нам новые вычислительные инструменты.

Ключевые слова: долговременные наблюдения, качество данных, первичная обработка, метрологический дефицит, метрологическое обеспечение

Цитируйте эту статью как: *Елисейкин М.М., Очков В.Ф.* Фрагментация временных рядов – не аномалия, а норма // Наука и технологические разработки. 2024. Т. 103, № 4. С.39–46. <https://doi.org/10.21455/std2024.4-3>

Введение

Статья “Проблема качества данных при режимном геофизическом мониторинге: кто виноват и что делать?” [Децеровский, 2024], открывшая дискуссию, частью которой является и настоящая работа, вышла одновременно с серией статей “Метрологический дефицит в промышленных “больших данных” и “О метрологических характеристиках исторических данных” [Елисейкин, Очков, 2024а,б], что свидетельствует об очевидной актуальности проблемы. Парадоксальность данной ситуации состоит в том, что при её обсуждении используются примеры возрастом в несколько десятилетий, да и сама проблема метрологической совместимости результатов измерений существует столетиями и представляет собой одну из главных проблем науки в течение всего времени её существования.

Так почему же именно сейчас несколько авторов пытаются ответить на вопрос “что делать в сложившейся ситуации”, если ситуация не нова? Почему мы считаем актуальной и острой проблему, о которой говорят уже десятилетиями?

Ответ прост: отдельные составляющие проблемы были известны и ранее, но они считались незначительными или трудноразрешимыми. Значительную часть времени за последние два века научно-технического прогресса люди вручную обсчитывали данные, полученные при неавтоматизированном снятии показаний с циферблатов приборов. Это накладывало ограничения как на возможности анализа временных рядов, так и на точность результатов, получаемых при их анализе. Развитие вычислительной техники изменило ситуацию, но сначала и оно имело организационные ограничения: компьютерные вычисления долгое время были дефицитным и сложным в использовании ресурсом.

Сейчас же, в связи с увеличением доступности вычислительных мощностей, стало возможным делать трудновыполнимые ранее анализы с недоступной ранее скоростью и точностью. Эти новые возможности уже активно используются в различных областях науки и техники, и их распространение будет только увеличиваться.

Именно поэтому перед нами сейчас стоят два новых и актуальных вопроса:

1) Как анализировать результаты наблюдений, проводившихся десятилетия назад по методам, не предполагавшим реальность 2024 года?

2) Как проводить наблюдения в 2024 году, чтобы спустя годы их результаты можно было эффективно использовать?

Ответы на эти вопросы кроются в понимании сути произошедших изменений.

Социальные корни проблемы

Раньше анализ данных был сложной и медленной задачей, которой занимался специалист, владеющий предметной областью. Во время медленного анализа этот специалист думал, строил гипотезы и принимал решение о целесообразности использования того или иного инструмента. Даже если расчёты делегировались другому человеку или выполнялись на компьютере, необходимо было принять решение о том, какие методы и алгоритмы будут использованы.

Сейчас же сложные вычислительные задачи полностью переложены на компьютер и типовые статистические алгоритмы, которые работают быстрее, чем человек нажимает кнопки на клавиатуре. При этом человек, нажимающий кнопки, может не понимать смысла анализируемых данных и не отдавать себе отчёта в том, можно ли использовать те или иные алгоритмы в конкретной ситуации. Более того, вся суть прогресса состоит в том, что потребителю нет необходимости понимать низкоуровневые процессы: можно ездить на автомобилях, не зная, как работает двигатель, можно пользоваться компьютером, не зная, как работают микросхемы, можно получать результаты работы компьютерной программы, не зная, какой алгоритм она использует.

Точно так же, как у нерадивых студентов возникает соблазн использовать правило Лопиталья при любом раскрытии неопределённости без учёта его ограничений, у людей, анализирующих временные ряды, возникает желание использовать общепринятые процедуры предварительной очистки временных рядов без учёта “границ допустимого” [Дещеревский, 2024, с. 21]. При этом сам по себе вопрос об аномалиях в собранном временном ряде может послужить поводом для обвинения исследователя в неумении проводить наблюдения [Дещеревский, 2024, с. 4], что снижает желание разбираться в каждом конкретном случае. Иными словами, сообщение об аномалии в результатах наблюдений социально порицаемо, а анализ аномалий экономически нецелесообразен, поскольку высока вероятность того, что время исследователя будет

потрачено на получение бесполезного ответа: “Технический сбой при проведении наблюдений”.

Таким образом, актуальность данной темы состоит не в самих аномалиях или несовпадении метрологических свойств собираемых данных, а в том, что именно они и не рассматриваются в качестве проблемы, требующей внимания.

Из этого вывода следует один из ответов на вопрос “что делать”: анализ аномалий должен стать нормой и даже, возможно, обязательной частью исследований. Вывод же о том, что аномалия не является проявлением неучтённого явления или процесса, должен быть обоснован и изложен в явном виде.

К сожалению, это может создать новую проблему: если поощрять обнаружение аномалий в собранных данных, то недобросовестные исследователи могут начать фальсифицировать аномалии в необходимом им количестве.

Особенности возникновения временных рядов

При изучении проблем, связанных с временными рядами, полезно рассмотреть механизм возникновения этих рядов в различных областях. Можно сравнить три варианта получения данных: лабораторные эксперименты, мониторинг промышленного оборудования и полевые геофизические наблюдения.

Продолжительность наблюдений

Лабораторные эксперименты скоротечны и однородны в плане методик и используемых инструментов. Даже если речь идёт о длительной подготовке к эксперименту и длительной обработке полученных данных, сам эксперимент будет занимать сравнительно мало времени. Так как одним из важнейших требований к лабораторным экспериментам является их воспроизводимость в другой лаборатории и другим человеком, такие эксперименты в принципе не могут длиться десятилетиями – от первоначального результата до его воспроизведения при этом будет проходить слишком много времени. По сути, отдельные лабораторные эксперименты дают очень точные, но точечные результаты, которые можно и нужно сравнивать между собой, но не всегда нужно выстраивать в некий временной ряд.

При **мониторинге промышленного оборудования** данные могут собираться годами и десятилетиями. За это время могут заменяться измерительные инструменты, но сам исследуемый объект остаётся тем же самым, и это задаёт определённые временные рамки: например, минимальные нормативные сроки службы паровых турбин для теплоэлектростанций составляют от 25 до 40 лет в зависимости от типа турбины [ГОСТ..., 2019]. Тем самым, задача мониторинга промышленного оборудования – это именно создание временного ряда наблюдений длительностью в весь срок службы оборудования.

Когда же мы говорим о **полевых геофизических наблюдениях**, надо понимать, что они по самой своей сути предназначены для создания гораздо более продолжительных временных рядов, чем в предыдущих случаях. Даже если каждое отдельное наблюдение будет непродолжительным, оно подразумевает, что аналогичное наблюдение может быть выполнено через годы, десятилетия и даже века, с тем чтобы создать временной ряд и проанализировать динамику изменений параметров наблюдаемого явления. Имея продолжительность систематических геофизических наблюдений примерно в 200 лет (начиная с XIX в.), мы должны понимать, что и получаемые сейчас данные также могут быть использованы спустя века. Для этого случая есть хороший пример наблюдений, хоть и не относящийся к геофизике, – анализ солнечной активности с использованием зарисовок, сделанных Кеплером в XVII в. [Hayakawa et al., 2024].

Таким образом, учитывая тот факт, что самый продолжительный лабораторный эксперимент длится около 100 лет [Johnston, 2013], и это считается чем-то экстраординарным, может быть нецелесообразно делать все практики и стандарты лабораторных экспериментов ориентиром для полевых наблюдений.

В этом плане геофизические наблюдения близки к мониторингу промышленного оборудования, но есть отличие в продолжительности жизненного цикла наблюдаемого объекта. Промышленное оборудование функционирует в течение нескольких десятилетий (крайне редко – 1–2 столетия), в то время как наблюдаемая геофизиками планета существует гораздо дольше, и наблюдаемые явления могут длиться тысячи и даже миллионы лет.

Целенаправленность наблюдений

Другим важным отличием полевых геофизических наблюдений от лабораторных экспериментов и мониторинга промышленного оборудования является отсутствие конкретной цели. Конечно же, сами по себе геофизические наблюдения имеют цель, состоящую в отслеживании изменений конкретных природных явлений, но в них нет цели, на достижение которой надо ориентироваться.

При проведении лабораторных экспериментов сперва задаются проверяемые гипотезы и предполагаются ожидаемые результаты. Данные, получаемые в результате эксперимента, должны совпасть или не совпасть с гипотезой. Экспериментатор знает, что он ищет в результатах эксперимента, и может однозначно понять, нашёл он это или нет.

Мониторинг промышленного оборудования также имеет своей целью сравнение измеренных значений с ожидаемыми, потому что само оборудование было создано для решения конкретной задачи и исходя из известных физических закономерностей. Тем самым, исследователь знает, какие именно результаты он должен получить в идеальном случае, и должен забить тревогу в случае, если измеренные значения не совпадают с расчётными.

Таким образом, и в лабораторных экспериментах, и в промышленном мониторинге есть идеальная модель, с которой исследователь сверяется при анализе полученных данных. Если результат анализа не совпадает с ожидаемым, мы можем проанализировать возникшую аномалию и попытаться определить её природу.

Однако если речь идёт о геофизических наблюдениях, то далеко не всегда есть идеальная модель, с которой можно сравнивать результаты наблюдений. Исследователь просто наблюдает в расчёте на то, что сумеет обнаружить в результатах наблюдений какую-то новую закономерность. В этом случае у него нет никаких подсказок, указывающих на характер той или иной аномалии в результатах наблюдений; он даже не знает, является ли она техническим сбоем в процессе измерений или же редким, но закономерным явлением.

Именно отсутствие ожидаемого идеального результата в случае геофизических наблюдений указывает нам на возможный вариант решения проблемы. Суть в том, что сами по себе наблюдения – это одновременно и мониторинг промышленного оборудования, которым является измерительный прибор. Как показано в примере с кондуктометром KB-101 [Децереvский, 2024, с. 18], техногенные аномалии, связанные со сбоями измерительных приборов, могут быть закономерными и зависеть от конструкции измерительного прибора и физического принципа, положенного в основу измерений. При этом такие закономерные аномалии связаны с самим измерительным прибором, а не с областью его применения, т.е. при их анализе может быть использован

опыт использования аналогичных приборов, не ограничивающийся лишь геофизическими наблюдениями.

Фрагментация временных рядов

Из факта большой продолжительности временных рядов, возникающих при мониторинге промышленного оборудования и геофизических наблюдениях, можно сделать полезный вывод: так как измерительные приборы выходят из строя, а методики измерений меняются, то при достаточной продолжительности процесса сбора данных неизбежно наступит ситуация, когда измерение одного и того же явления в одной и той же точке пространства будет остановлено и начато с другими метрологическими характеристиками. Это может быть другой прибор, другого поколения, с другой точностью и, что не исключено, на другом физическом принципе. Однако даже если это будет тот же самый прибор, метрологические изменения произойдут из-за самого факта его остановки и повторного запуска [Дещеревский, 2024, с. 12].

Таким образом, фрагментация происходит неизбежно, в том числе из-за действий, направленных на повышение точности измерений. Если принять этот факт, то нам открывается возможность анализа повторяющихся закономерных процессов – самих процессов длительных измерений. Хотя из-за условности отметки нуля мы не всегда можем “склеить” два временных ряда в один, мы можем исследовать однотипные переходные процессы в начале каждого фрагмента и возможный дрейф датчика в процессе его работы.

Другим выводом, следующим из принятия неизбежности фрагментации временных рядов, является потребность в использовании алгоритмов, которые учитывают эту фрагментацию. Необходима разработка и пропаганда использования алгоритмов, учитывающих то, что крайние точки соседних фрагментов временного ряда могут иметь разную неопределённость.

Таким образом, ещё один ответ на вопрос “что делать?” может быть сформулирован так: сохранять информацию не только о самих результатах наблюдений, но и об их метрологических характеристиках [Елисейкин, Очков, 2024а]. Как минимум, это информация о модели прибора, с помощью которого было проведено измерение, что позволит сопоставить изменение во временном ряде с фактом замены этого прибора или процессом его деградации.

Будущие риски

Если говорить о том, что борьба за точность может привести к ухудшению ситуации, то необходимо затронуть и ещё один важный момент. Проблема дрейфа датчиков и их заведомой неточности не нова, и в этом направлении ведутся работы: например, в некоторые датчики уже встраиваются механизмы автокомпенсации дрейфа. Учитывая то, что нам становятся доступны всё большие вычислительные мощности и всё более сложные алгоритмы, это может создать новые риски.

Так, будущие датчики могут быть компьютерами, которые не просто преобразуют физические явления в показания, а трактуют их по какому-то алгоритму. Это очень удобно и позволяет вносить в показания те или иные корректировки, устраняя нелинейность точности датчика, компенсируя дрейф, старение и т.д. Как следствие, мы снова имеем дело с парадоксальной проблемой: программа, корректирующая показания датчика, может быть обновлена в течение его жизненного цикла, что приведёт к изменению метрологических характеристик датчика. И даже если это будет улучшение работы измерительного прибора, оно тоже приведёт к фрагментации временного ряда.

Значит, следует сохранять не только информацию о модели датчика, его серийном номере и о том, что с ним производились физические манипуляции, – придётся сохранять информацию о версии программного обеспечения (ПО) датчика и времени его обновления. Это позволит сопоставить аномалию в данных с моментом обновления ПО.

Заключение

Как уже было сказано, сейчас мы находимся в процессе изменения практик, связанных с наблюдениями, измерениями, мониторингом и т.п. Меняются методы сбора и анализа данных. Меняются стандарты связанных с этими процессами профессий.

Цель настоящей статьи – показать, что такие изменения должны происходить не только пассивно, как механические последствия технических изменений измерительных приборов и вычислительной техники, но и активно – как результат адаптации исследователей к изменениям доступной им материальной части.

Надо принять, что часть из того, что мы считаем случайным сбоем, на самом деле является закономерным и неизбежным процессом.

Надо менять культуру анализа данных и использовать возросшие вычислительные возможности не только для облегчения предварительной обработки данных, но и для поиска закономерностей в том, что раньше считалось случайными выбросами.

Надо принять то, что временные ряды всегда фрагментированы, и развивать алгоритмы, которые учитывали бы фрагментированность временных рядов и метрологическую неоднородность анализируемых данных.

И, конечно же, надо бережно хранить исходные данные, не прошедшие нормализацию и чистку от выбросов, чтобы иметь возможность вернуться к ним в будущем с новыми подходами и инструментами.

Финансирование

Исследование выполнено по инициативе авторов без какого-либо внешнего финансирования.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Литература

- ГОСТ 34484-2018 Турбины паровые стационарные. Нормы расчета на прочность корпусов цилиндров и клапанов. М.: Стандартиформ, 2019. 50 с.
- Децеровский А.В. Проблема качества данных при режимном геофизическом мониторинге: кто виноват и что делать? // Наука и технологические разработки. 2024. Т. 103, № 3. С.3–26. <https://doi.org/10.21455/std2024.3-1>
- Елисейкин М.М., Очков В.Ф. Метрологический дефицит в промышленных “больших данных” // Законодательная и прикладная метрология. 2024а. № 4 (190). С.19–24. <https://doi.org/10.32446/2782-5418.2024-4-19-24>
- Елисейкин М.М., Очков В.Ф. О метрологических свойствах ретроспективных данных // Законодательная и прикладная метрология. 2024б. № 5 (191). С.47–51. <https://doi.org/10.32446/2782-5418.2024-5-47-51>
- Hayakawa H., Murata K., Teague E.T.H., Bechet S., Sôma M. Analyses of Johannes Kepler’s sunspot drawings in 1607: A revised scenario for the solar cycles in the early 17th century // *Astrophys. J. Lett.* 2024. V. 970, N 2. Art. L31. 7 p. <https://doi.org/10.3847/2041-8213/ad57c9>

Johnston R. World's slowest-moving drop caught on camera at last // Nature. 2013. 2 p.
<https://doi.org/10.1038/nature.2013.13418>

Сведения об авторах

ЕЛИСЕЙКИН Михаил Михайлович – аспирант, НИУ “МЭИ”. 111250, Россия, г. Москва, ул. Красноказарменная, д. 14, стр. 1. E-mail: muxa@muxa.ru

ОЧКОВ Валерий Фёдорович – доктор технических наук, профессор, НИУ “МЭИ”. 111250, Россия, г. Москва, ул. Красноказарменная, д. 14, стр. 1. E-mail: OchkovVF@mpei.ru

METADATA IN ENGLISH

About the journal

NAUKA I TEKHNOLOGICHESKIE RAZRABOTKI (SCIENCE AND TECHNOLOGICAL DEVELOPMENTS), ISSN: 2079-5165, eISSN: 2410-7948, DOI: 10.21455/std; https://elibrary.ru/title_about.asp?id=32295; <https://portal.ifz.ru/ntr>. The journal was founded in 1992.

**FRAGMENTATION OF TIME SERIES IS NOT AN ANOMALY,
BUT THE NORM**

M.M. Eliseykin, V.F. Ochkov

Moscow Power Engineering Institute (National Research University), Moscow, Russia

Corresponding author: M.M. Eliseykin, e-mail: muxa@muxa.ru

Abstract. Contemporary technologies have simplified and made accessible the collection and processing of data from scientific observations. Current solutions enable rapid preliminary processing of collected data, cleansing it of outliers related to measurement errors and filling gaps in fragmented time series. However, the ease with which this is achieved presents risks of indiscriminate use of such solutions. Consequently, data indicating real physical anomalies may be discarded, and the amalgamation of time series fragments might result in data that does not correspond to the observed processes and phenomena. Such a situation was justified in the past when there was a scarcity of computational resources. Discarding inherently unreliable values and filling gaps simplified and accelerated data analysis. Now, with sufficient computational power available, it is possible to begin searching for patterns in what was previously considered observational error and discarded. Moreover, the volume of accumulated data may allow for the consideration of fragments of time series as parts of a regular process, without filling the gaps with artificial data created based on our assumptions about the nature of the observed processes and phenomena. This raises the question of the necessity to adapt the approaches used in collecting and analyzing observational results to the possibilities afforded by new computational tools.

Keywords: long-term observations, data quality, primary processing, metrological deficit, metrological support

Cite this article as: Eliseykin M.M., Ochkov V.F. Fragmentation of time series is not an anomaly, but the norm, *Nauka i tekhnologicheskie razrabotki* (Science and Technological Developments), 2024, vol. 103, no. 4, pp. 39–46. [in Russian]. <https://doi.org/10.21455/std2024.4-3>

Financial support

The work was carried out on the authors' initiative without any external funding.

Ethics declarations

The authors declare no conflict of interest.

About the authors

ELISEYKIN Mikhail Mikhailovich – Postgraduate Student, Moscow Power Engineering Institute (National Research University). Russia, 111250, Moscow, Krasnokazarmennaya st. 14, bld. 1. E-mail: muxa@muxa.ru

OCHKOV Valery Fedorovich – Dr. Sc. (Tech. Sci.), Professor, Moscow Power Engineering Institute (National Research University). Russia, 111250, Moscow, Krasnokazarmennaya st. 14, bld. 1. E-mail: OchkovVF@mpei.ru