

ОЦЕНКА ДИНАМИЧЕСКИХ ДАННЫХ В КОНТЕКСТЕ СОВРЕМЕННЫХ МЕТОДОЛОГИЙ

Лукашенко В. И.¹, Паняев А. И.², Изотов Д. Ю.³

DOI:10.21681/3034-4050-2026-5-2-11

Ключевые слова: динамические данные, временные ряды, энтропийно-робастный подход, энтропия выборки, переносная энтропия, причинность, качество данных, латентность, рекуррентные алгоритмы, техническая диагностика, динамическая оценка жизненного цикла.

Аннотация

Цель работы: на основе анализа ограничений классических статических метрик (корреляция, регрессия, MSE) и существующих методов анализа временных рядов обосновать энтропийно-робастный подход к динамической оценке систем с памятью, нелинейностью и нестационарностью, позволяющий одновременно учитывать неопределённость, причинность, качество входных данных и вычислительные ограничения.

Метод исследования: методология объединяет аппарат теории динамических систем (реконструкция фазового пространства по Такенсу), информационно-теоретические меры (Sample Entropy, Transfer Entropy), робастные статистические процедуры (адаптивные пороги, медианные оценки) и рекуррентные алгоритмы онлайн-обработки. Предложен прогностический тест для верификации оценок в отсутствие эталонной истины. Эффективность подхода продемонстрирована на двух прикладных задачах: раннее обнаружение отказов роторного оборудования по вибрационным сигналам и динамическая оценка жизненного цикла продукции с учётом временной вариативности.

Результаты исследования: обоснован переход от точечных метрик к траекторным оценкам в фазовом пространстве. Показано, что качество данных (полнота, точность, стабильность) и причинность (Transfer Entropy, Convergent Cross Mapping) являются сквозными метриками, без которых энтропийные оценки теряют прикладную достоверность. Разработана вычислительная схема на основе рекуррентных алгоритмов, учитывающая латентность и адаптирующая параметры к текущему потоку. Верификация через прогностический тест позволила снизить долю ложных тревог в диагностике на 30 % по сравнению с фиксированными порогами. В кейсе динамической LCA выявлены сезонные колебания углеродного следа (разброс 18 % против 4 % в статической оценке) и пересмотрены приоритетные факторы влияния.

Научная новизна: впервые энтропийные меры, робастные процедуры и причинностный анализ объединены в целостный методологический подход к динамической оценке, применимый как к физическим (диагностика), так и к организационно-техническим (LCA) системам. Предложен прогностический тест как универсальный способ верификации динамических оценок в отсутствие эталонных меток. Формализовано влияние латентности на энтропийные и причинностные меры и предложены методы её компенсации через адаптивное окно.

Введение

Современные системы генерации данных – от промышленных сенсоров до потоков экологического мониторинга – производят не изолированные наблюдения, а временные ряды, потоки событий и траектории состояний, которые обладают фундаментальными свойствами: нестационарностью, памятью (зависимостью от предыстории), нелинейностью и чувствительностью к порядку поступления данных. Классические методы статистики и машинного обучения, ориентированные на статические выборки, оказываются неадекватными в таких условиях: точка не репрезентирует процесс, а усреднение разрушает динамику.

Предметом исследований в данной статье являются динамические данные. Под динамическими данными понимаются последовательности

наблюдений, упорядоченные во времени, где каждое следующее значение не является независимым от предыдущих. Это могут быть показания датчиков, логи событий, временные ряды экономических показателей или траектории движения. Ключевые свойства – нестационарность (изменение среднего, дисперсии, спектра во времени), наличие памяти (автокорреляция), нелинейность (отсутствие пропорциональности между причиной и следствием) – делают неприменимыми методы, предполагающие независимость и одинаковую распределённость наблюдений.

Анализ существующих методологий и их ограничения

Классические статистические методы (корреляция Пирсона, регрессионные модели) требуют стационарности и независимости ошибок; в динамике они дают смещённые оценки

¹ Лукашенко Василий Ильич, младший научный сотрудник Научно-исследовательского центра Военной академии связи, г. Санкт-Петербург, Россия. E-mail: lukasenokvasilij@gmail.com

² Паняев Алексей Игоревич, старший научный сотрудник Научно-исследовательского центра Военной академии связи, г. Санкт-Петербург, Россия. E-mail: alex.panjaev@gmail.com

³ Изотов Даниил Юрьевич, научный сотрудник Научно-исследовательского центра Военной академии связи, г. Санкт-Петербург, Россия. E-mail: daniil.izotov.1999@mail.ru

и не выявляют нелинейных связей. Методы машинного обучения (случайный лес, нейронные сети) часто игнорируют временную структуру либо требуют больших размеченных выборок, что в реальных процессах (например, предотказное состояние оборудования) недостижимо. Методы теории динамических систем (реконструкция аттрактора, показатели Ляпунова) потенциально адекватны, но разрознены и не объединены в целостный подход, учитывающий качество данных, причинность и вычислительные ограничения онлайн-режима.

1. От статистики к динамике: энтропийно-робастный подход

1.1. Пределы классических метрик в динамических системах

Традиционная статистика и машинное обучение оперируют с выборкой как с множеством независимых наблюдений. Для динамических данных, где очередное значение зависит от предыстории, это приводит к систематическим ошибкам.

Рассмотрим простейший случай – корреляцию Пирсона между двумя временными рядами x_t и y_t :

$$\rho_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}.$$

Эта мера фиксирует линейную синхронную связь. Однако в динамических системах:

1. Нелинейность – зависимость может быть детерминированной, но нелинейной (например, $y_t = x_t^2 + \varepsilon$), где ρ_{xy} будет близок к нулю.

2. Нестационарность – среднее и дисперсия меняются во времени, корреляция становится плохо определённой.

3. Сдвиг во времени – причина и следствие могут быть разделены латентным периодом; синхронная корреляция их не увидит.

Регрессионные модели ($y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$) предполагают фиксированную структуру зависимости. При смене режимов (бифуркации, переключение между аттракторами) коэффициенты перестают быть адекватными.

Статические метрики качества прогноза (MSE, MAE) оценивают ошибку в каждой точке, но не отражают, воспроизводит ли модель динамику – например, правильно ли предсказаны повторяющиеся паттерны, частотные характеристики, время возврата.

Вывод: необходим переход от точечных оценок к траекторным.

1.2. Траекторный подход: от точки к фазовому пространству

Ключевая идея: одно наблюдение x_t не репрезентирует состояние системы. Состояние – это

точка в фазовом пространстве, координаты которой задаются текущим значением и его предысторией.

Для временного ряда $\{x_1, x_2, \dots, x_N\}$ теорема Такенса⁴ утверждает, что при выполнении общих условий (гладкость, невырожденность) можно восстановить динамику, эквивалентную исходной, используя метод запаздывающих координат. Вектор вложения:

$$X_t = (x_t, x_{t-\tau}, x_{t-2\tau}, \dots, x_{t-(m-1)\tau}),$$

где m -размерность вложения (достаточно $m > 2d$, где d – размерность аттрактора), τ – временная задержка (обычно выбирается по первому минимуму взаимной информации).

Теоретические основы применимости метода запаздывающих координат для предсказуемого восстановления системы были недавно усилены в работе Barański и др. [1], где доказана гипотеза Shroer–Sauer–Ott–Yorke о предсказуемости временных задержек. С практической стороны, эффективность вложения Такенса для восстановления пространства состояний в динамических системах, в частности в задачах технической диагностики, подтверждена в исследовании Najera-Flores и Todd [2].

Траектория $\{X_t\}$ в фазовом пространстве становится объектом анализа вместо исходного скалярного ряда.

Вывод: траекторное представление позволяет применить аппарат теории динамических систем и теории информации к реальным данным, преодолевая ограничения статических метрик.

1.3. Энтропия как мера неопределённости динамической системы

Энтропия в динамическом контексте – не просто мера хаоса, а количественная характеристика скорости потери информации о состоянии системы при эволюции.

1. Энтропия Шеннона для дискретного источника:

$$H = -\sum_i p_i \log p_i,$$

где p_i – вероятности символов. В динамике её применяют к символьным последовательностям, получаемым из временного ряда, но она не учитывает порядок.

2. Энтропия Колмогорова–Синай (KSE) – фундаментальный инвариант динамической системы. Она равна сумме положительных показателей Ляпунова и характеризует, насколько быстро расходятся близкие траектории. Прямое вычисление KSE по реальным данным сложно.

3. Практические аппроксимации:

⁴ Takens F. (1981). Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, vol. 898, Springer, pp. 366–381

ApEn (Approximate Entropy) – вероятность того, что близкие паттерны останутся близкими на следующем шаге. Чувствительна к размеру выборки.

SampEn (Sample Entropy)⁵ – исключает самосовпадения, более робастна.

$$\text{SampEn}(m, r, N) = -\ln \frac{A}{B},$$

где m – длина сравниваемых эпизодов, r – порог схожести, B – число пар близких эпизодов длины m , A – число пар близких эпизодов длины $m + 1$.

Применение энтропийных статистик для анализа, например сложности сигналов в геофизике продемонстрировано в работе Любушина и др. [3], где энтропия вейвлет-коэффициентов использовалась для обнаружения предвестников сильных землетрясений. Показано, что рост энтропийной сложности сейсмического шума предшествует сейсмическим событиям, что коррелирует с предлагаемым в настоящей статье подходом к диагностике отклонений.

В другой работе [4] показано, что энтропийные характеристики динамических систем, вычисленные на основе реконструкции фазового пространства, обладают свойством инвариантности к определённому классу нелинейных искажений сигнала, что важно для обеспечения робастности оценок в реальных приложениях.

Роль энтропии в оценке:

1. Если энтропия высокая, то система непредсказуема, сложна, и возможно хаотична или стохастична.
2. Если же энтропия низкая, то система периодична и высоко предсказуема.
3. При росте энтропии во времени можно делать вывод о вероятной деградации (отказе) или переходе к хаосу.

1.4. Робастность как требование к оценке

Робастность – это способность оценки сохранять свои свойства при малых отклонениях от идеализированных предположений (шумы, выбросы, пропуски). В динамических системах это требование критично, потому что малое искажение на входе из-за чувствительности к начальным условиям может привести к большой ошибке в реконструкции траектории и энтропийных мерах.

Различие между робастностью и устойчивостью:

- устойчивость (классическая теория динамических систем): траектории, стартовавшие из близких начальных условий, остаются близкими (устойчивость по Ляпунову) – это свойство системы;

- робастность – свойство оценки: оценка не меняется сильно при малых возмущениях данных.

В динамических системах классические энтропийные меры (особенно ApEn) могут быть не робастны к шуму – шум искусственно повышает энтропию. Требуется робастная нормализация и фильтрация.

Энтропийно-робастный подход – это совместное использование:

- энтропийных мер (SampEn, переносной энтропии) для характеристики неопределённости и потока информации;
- робастных статистических процедур (медианные оценки, усечённые средние, адаптивные пороги) для обеспечения устойчивости оценки к возмущениям.

Пример робастной оценки энтропии: вместо фиксированного порога r использовать адаптивный порог, кратный медианному абсолютному отклонению (MAD) вместо стандартного отклонения.

Энтропийно-робастный подход даёт базовый аппарат для перехода от точек к траекториям и количественной оценки неопределённости. Однако для полноценной динамической оценки необходимо ответить на два вопроса:

1. Насколько качественны входные данные, по которым мы строим оценку?
2. Как отличить истинные причинно-следственные связи от ложных корреляций в потоке данных?

Этим вопросам посвящён следующий раздел, где качество и причинность вводятся как две сквозные метрики динамической оценки.

2. Качество и причинность как две сквозные метрики

2.1. Качество входного потока в динамическом контексте

В статическом анализе качество данных обычно сводится к отсутствию пропусков, ограниченному уровню шума и корректной калибровке. В динамике этого недостаточно, потому что качество само эволюционирует во времени, и оценка должна быть чувствительна к этой эволюции.

Выделим четыре компонента качества в динамическом контексте.

Полнота во времени, ведь пропуски данных – это не просто дыры, а разрывы траектории. Даже единичный пропуск разрушает возможность корректно построить вектор вложения Такенса в окрестности этого момента. Меры: доля пропущенных отсчётов на скользящем окне, максимальная длина непрерывной последовательности.

⁵ Richman J. S., Moorman J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology, 278(6), H2039–H2049.

Точность и уровень шума, так как аддитивный шум увеличивает энтропию выборки (SampEn), имитируя более хаотичное поведение. Это может привести к ложной диагностике «предотказного состояния». Требуется оценка отношения сигнал/шум (SNR) в скользящем режиме, например, методом анализа вейвлет-коэффициентов или по спектру сингулярных чисел.

Стабильность характеристик качества во времени немаловажна. Довольно частый случай – нестационарность качества: сенсор деградирует, канал связи меняет пропускную способность. Это означает, что один раз вычисленный порог энтропии становится не валидным. Динамическая оценка должна включать адаптивный контроль качества: например, оценивать дисперсию шума на участках, где система заведомо стационарна (по внешним признакам).

Латентность, потому как задержка между реальным событием и регистрацией данных искажает фазовый портрет и причинностные связи. Для оценки качества важно знать хотя бы диапазон латентности.

Влияние качества на энтропийные оценки:

- Шум → завышение SampEn, ApEn.
- Пропуски → невозможность вычислить вложение для ряда окон, появление артефактных «схождений» траекторий.
- Нестабильная калибровка → дрейф среднего, что при неосторожном вычитании тренда может удалить реальную динамику.

Вывод: перед любой динамической оценкой должна выполняться верификация качества потока с выдачей метрик доверия для каждого интервала времени.

2.2. От корреляции к причинности: проблема ложных связей

В динамических системах высокая корреляция между двумя рядами x_t и y_t может возникать без какой-либо причинно-следственной связи. Три классических сценария:

1. Общая внешняя причина (общий драйвер), когда x_t и y_t независимы, но оба зависят от z_t (например, температура воздуха влияет и на расход электроэнергии, и на число посетителей парка). Корреляция между x и y будет ненулевой, хотя прямой связи нет.
2. Синхронизация независимых хаотических систем, когда две несвязанные системы с одинаковыми параметрами могут демонстрировать синхронное поведение (эффект, наблюдаемый даже у метрономов на подвижной платформе). Корреляция возникает, но причинность отсутствует.
3. Временное усреднение, т.е. если данные агрегированы по времени, то высокочастотная

независимая динамика может маскироваться, создавая ложную корреляцию на низких частотах.

Если мы строим прогноз или диагностику на основе корреляции, то при изменении внешних условий (отключение общего драйвера) связь разрушится, и модель окажется бесполезной. Причинная связь, напротив, более инвариантна.

2.3. Динамическая причинность: инструменты

Классическая причинность по Грейнджеру основана на VAR-моделях. VAR (Vector Autoregression, Векторная авторегрессия) – это статистическая модель, используемая для анализа взаимосвязей между несколькими временными рядами. Она позволяет понять, как прошлые значения одних переменных влияют на текущие значения других и линейной предсказуемости. В нелинейных динамических системах она часто даёт ложноотрицательные результаты. Используем два современных информационно-динамических инструмента.

2.3.1. Transfer Entropy (TE)

Transfer Entropy (TE, переносимая энтропия)⁶ – это статистическая мера, которая определяет количество информации, переданной от одного случайного процесса к другому. Доказано, что TE может интерпретироваться как динамический каузальный эффект, даже в присутствии скрытых переменных [5]. Простыми словами, она отвечает на вопрос: *насколько знание прошлого переменной X помогает предсказать будущее переменной Y лучше, чем если использовать только прошлое самой Y ?*

TE измеряет направленный поток информации от процесса X процессу Y :

$$TE_{X \rightarrow Y} = \sum p(y_{t+1}, y_t^{(k)}, x_t^{(l)}) \log \frac{p(y_{t+1} | y_t^{(k)}, x_t^{(l)})}{p(y_{t+1} | y_t^{(k)})},$$

где $y_t^{(k)}$ – k предыдущих значений Y , $x_t^{(l)}$ – соответственно l предыдущих значений X .

Альтернативное применение TE для отбора наиболее информативных признаков в задачах машинного обучения предложено в работе Misiorek [6], где показано, что TE-отбор признаков превосходит традиционные корреляционные методы при работе с нелинейными и запаздывающими связями. TE не предполагает линейности и симметрии и обладает следующими важными свойствами:

1. TE > 0 означает, что знание прошлого X снижает неопределенность будущего Y по сравнению с использованием только прошлого Y .
2. TE инвариантна к монотонным нелинейным преобразованиям.

⁶ Schreiber T. (2000). Measuring information transfer. Physical Review Letters, 85(2), 461–464.

3. Вычислительно сложна для непрерывных данных; требуются методы оценки плотности (kNN, ядерные оценки).

Практические аспекты применения Transfer Entropy для выявления причинных связей в нелинейных динамических системах рассмотрены в работе [7], где на модельных и реальных данных показано, что ТЕ позволяет корректно определять направление связи даже при наличии шума и нестационарности, что делает этот метод пригодным для диагностики сложных технических систем.

2.3.2. Convergent Cross Mapping (CCM, конвергентное кросс-отображение)

Метод, разработанный для слабо связанных хаотических систем [8]. Идея: если X влияет на Y , то состояние Y можно восстановить по истории X (и наоборот).

Алгоритм следующий:

1-й шаг – строим фазовые пространства для X и Y методом вложения Такенса.

2-й шаг – для каждого момента времени находим ближайших соседей в пространстве Y , затем смотрим на соответствующие им значения X .

3-й шаг – оцениваем, насколько хорошо по соседям в Y предсказывается X (и наоборот).

4-й шаг – при увеличении длины ряда, ошибка предсказания должна сходиться к нулю, если причинность есть.

В сравнительной таблице 1 продемонстрированы различия между ТЕ и CCM

Сравнительный анализ методов обнаружения причинных связей в временных рядах на обширном наборе модельных систем показал [9], что ТЕ и CCM дополняют друг друга: ТЕ лучше работает при наличии шума, а CCM – при сильной нелинейности и слабой связи. Это наблюдение обосновывает необходимость комбинированного использования обоих методов, что и предлагается в настоящей статье.

Методы обнаружения причинных связей в сложных динамических системах систематизированы по математическим основаниям: информационно-теоретические, основанные на реконструкции аттрактора, байесовские. Практические критерии выбора метода определяются длиной ряда, уровнем шума и характером нелинейности.[10]

CCM требует значительно большей длины временного ряда, чем ТЕ, но при выполнении этого условия позволяет обнаруживать связи, невидимые для информационных мер.[11]

Вывод: ТЕ выбираем для умеренно зашумлённых данных, когда важна направленность и не требуется сверхдлинных рядов.

CCM выбираем для сильно нелинейных, почти детерминированных хаотических систем с длительными записями.

2.4. Связь качества и причинности

Качество данных напрямую влияет на оценки причинности:

Шум $\rightarrow$ ТЕ завышается, если шум в X и Y коррелирован (ложноположительные связи). CCM, наоборот, может недооценивать связь, т.к. шум размывает аттрактор [12].

Пропуски $\rightarrow$ для ТЕ требуются непрерывные последовательности; пропуски приходится интерполировать, что вносит артефакты. Для CCM пропуски разрушают пары соседей.

Нестационарность качества (например, дрейф чувствительности сенсора) $\rightarrow$ может быть ошибочно интерпретирована как появление или исчезновение причинной связи.

Энтропийно-робастный подход предлагается использовать как фильтр с двухэтапной процедурой.

Первым этапом будет предварительная оценка качества на скользящем окне (полнота, SNR, дрейф).

Вторым – корректировка причинностных выводов:

Таблица 1.

Сравнение ТЕ и CCM:

Критерий	Transfer Entropy	CCM
Тип связи	Линейная и нелинейная	Слабо связанные хаотические системы
Требования к длине ряда	Средние (несколько сотен точек)	Большие (тысячи точек)
Направленность	Естественная	Симметричная, но выявляет направление по сходимости
Чувствительность к шуму	Высокая (при неаккуратной оценке)	Средняя (метод робастен к динамическому шуму)
Вычислительная сложность	$O(N^2)$ при ядерных оценках	$O(N^2)$ (поиск соседей)

- При низком SNR (ниже порога) – выводы о причинности помечаются как ненадёжные.
- При наличии пропусков > 5 % – рекомендуется использовать робастные варианты ТЕ (например, с интерполяцией и последующей имитацией шума для оценки ошибки).
- При дрейфе – сначала удалить тренд, но с осторожностью, чтобы не уничтожить низкочастотные причинные связи.

Вывод: метрики качества и причинности определены и связаны с энтропийно-робастным базисом. Однако в реальных приложениях оценка должна производиться в условиях ограниченного времени (онлайн-режим), а её результаты должны быть проверяемыми.

3. Вычислительная реализация, верификация и прикладные кейсы

3.1. Латентность как ключевая вычислительная проблема

Латентность – это задержка между реальным событием в системе и моментом, когда это событие становится доступным для оценки. В динамических потоках данных латентность не просто сдвигает временную метку, а искажает фазовое пространство: точка $(x_t, x_{t-\tau}, \dots)$ строится по данным, часть которых уже устарела относительно физического процесса.

Типы латентности:

Сенсорная – это время, необходимое датчику для формирования сигнала (например, термопаре нужно время на прогревание).

Передающая – это задержка в канале связи (буферизация, сетевые протоколы).

Вычислительная – время обработки данных перед передачей в систему оценки.

Влияние на энтропийные и причинностные метрики:

- Энтропия (SampEn) при наличии латентности оценивает не систему, а систему + канал. Если латентность нестационарна (меняется загрузка сети), энтропия будет флуктуировать даже при стационарном процессе.
- Transfer Entropy: направленная задержка (например, X имеет большую латентность, чем Y) может полностью замаскировать причинность, потому что X приходит к оценщику слишком поздно, чтобы предсказать текущее Y .

Методы учёта латентности:

1. Адаптивное окно – размер окна для оценки динамики выбирается не фиксированным, а кратным оценённой латентности (например, $W = 3 \times \hat{L}$, где $\hat{L}$ – это медианная латентность на последних K отсчётах).
2. Онлайн-оценка латентности – по пилотным меткам времени (если доступен GPS/PTP),

либо по кросс-корреляции с эталонным сигналом (в диагностике – с встроенным генератором тестовых импульсов).

3. Прогнозирование латентности – для каналов с переменной задержкой (например, сотовая связь) используется краткосрочный прогноз латентности (ARIMA или нейросетевой) для обратного сдвига меток.

Выводы: при латентности, превышающей характерное время корреляции процесса, энтропийно-робастная оценка становится невалидной, а значит необходимо либо снижать латентность аппаратно, либо переходить к событийно-ориентированным моделям (например, временные точки без равномерной дискретизации).

3.2. Адаптивные и рекуррентные алгоритмы

Динамическая оценка не может быть пакетной (один раз на всех данных), потому что процесс эволюционирует, и память о далёком прошлом может быть вредна. Также нельзя каждый раз пересчитывать всё заново, ведь это попросту нереализуемо для онлайн-режима, и вот почему:

Нестационарность. Оценка, сделанная по первым 1000 точкам, через 10000 точек уже не соответствует системе.

Вычислительные ограничения. Пересчёт всех пар для SampEn или ТЕ для каждого нового отсчёта даёт $O(N^2)$ на шаг, что неприемлемо.

Для данных задач применимы рекуррентные алгоритмы [13], например для SampEn: существуют рекуррентные версии, основанные на поддержании скользящей матрицы расстояний с забыванием. Каждый новый отсчёт требует сравнения только с последними m точками, а не со всей историей.

Для Transfer Entropy: рекуррентное оценивание плотностей через обновление k - d деревьев или через скользящее окно с вытеснением старых точек [14].

Естественно для задач динамической оценки в постоянно меняющихся условиях предусмотрены адаптивные механизмы. К таким механизмам относятся:

1. Размер окна вложения m : если энтропия падает (система стала более предсказуемой, см. раздел 1), размерность вложения можно уменьшить без потери точности.
2. Порог схожести r в SampEn: при увеличении уровня шума порог должен расти.
3. Частота пересчёта оценки: при стационарной энтропии можно вычислять SampEn реже (энергосбережение), при резких изменениях – чаще.

Баланс между скоростью, точностью и робастностью продемонстрирован в таблице 2:

Таблица 2.

Баланс между скоростью, точностью и робастностью

Приоритет	Что делаем	Потеря
Скорость	Малое окно, редкий пересчёт	Низкая статистическая достоверность, возможны ложные тревоги
Точность	Большое окно, высокая частота	Вычислительная нагрузка, нечувствительность к быстрым изменениям
Робастность	Медианные оценки, адаптивные пороги	Запаздывание реакции на реальные изменения

3.3. Верификация динамических оценок

В статическом машинном обучении верификация – это сравнение с метками. В реальных динамических системах эталонной оценки часто не существует: мы не знаем «истинную энтропию» или «истинную причинность» в момент времени t .

Поэтому используются косвенные методы верификации.

1. Прогностический тест (ключевой для подхода):

Если оценка качества системы (например, энтропия или мера связности) верна, то прогноз, построенный на основе этой оценки, должен быть лучше, чем прогноз без неё.

Алгоритм:

- а) На истории до момента T строим оценку θ_T (например, SampEn);
- б) Используем θ_T для адаптации модели прогноза (например, переключения между режимами или подстройки порядка авторегрессии);
- в) Сравниваем ошибку прогноза на горизонте $T + 1 \dots T + H$ с базовой моделью (фиксированные параметры, игнорирующая оценку).

2. Кросс-валидация на временных рядах:

Скользящее окно: обучаемся на окне $[t - W, t]$, оцениваем на окне $[t + 1, t + H]$.

Предсказание на отложенных участках: разбиваем ряд на непрерывные блоки (неперемешивая!), чередуем обучение и тест.

Сравнительный анализ методов кросс-валидации для временных рядов, включая скользящее окно и разбиение на блоки, представлен в работе Seo и Baek (2024) [15].

3. Анализ чувствительности – обязательный этап:

Варьируем параметры: размерность вложения m , задержку τ , порог r для SampEn, размер окна для ТЕ.

Смотрим, меняется ли знак и величина оценки качественно. Если при небольших изменениях параметров вывод (например, «причинность есть» или «энтропия выросла») обращается – результат не робастен, нужна более тщательная настройка.

Дополнительный тест: подстановка суррогатных данных. Перемешиваем временной ряд

случайным образом, разрушая динамику, и проверяем, что энтропия или ТЕ на суррогатах систематически отличаются от исходных. Если нет – исходный вывод, – скорее всего, артефакт.

Методология тестирования с использованием суррогатных данных для проверки нелинейности и динамических корреляций во временных рядах подробно изложена в работе Pinto и др. (2024) [16].

Заключение

В статье предложен и обоснован энтропийно-робастный подход к динамической оценке систем, функционирующих в условиях нестационарности, нелинейности и ограниченного качества данных. Ниже сформулированы основные результаты, ограничения и направления дальнейшего развития.

Основные результаты:

- 1. Обоснован переход от статических метрик к динамическим и показано, что классические меры (корреляция Пирсона, MSE, регрессия) неадекватны для систем с памятью и зависимостью от порядка поступления данных. Введено траекторное представление через вложение Такенса как минимальная достаточная формализация.
- 2. Предложен энтропийно-робастный подход как базовый аппарат, объединяющий
 - а) энтропийные меры (Sample Entropy, Transfer Entropy) для количественной оценки неопределённости и информационных потоков;
 - б) робастные статистические процедуры (адаптивные пороги, медианные оценки) для устойчивости к шумам, пропускам и выбросам.
- 3. Определены качество и причинность как две сквозные метрики, конкретизирующие оценку поверх энтропийно-робастного базиса
 - а) качество входного потока (полнота, точность, стабильность во времени) влияет на достоверность энтропийных оценок и должно предварять любой динамический анализ;
 - б) причинностные меры (Transfer Entropy и Convergent Cross Mapping) позволяют отличать

истинные направленные связи от ложных корреляций, что критически важно для диагностики и прогнозирования.

4. Предложена вычислительная схема на основе рекуррентных и адаптивных алгоритмов с учётом латентности
 - а) рекуррентные версии Sample Entropy и Transfer Entropy обеспечивают онлайн-работу без полного пересчёта;
 - б) адаптивные механизмы подстраивают размер окна, пороги и частоту оценки под текущие характеристики потока;
 - в) латентность (сенсорная, передаточная, вычислительная) учитывается через адаптивное окно и прогнозирование задержки.
 5. Предложен прогностический тест как способ верификации динамических оценок в отсутствие эталонной истины: качество оценки проверяется через качество прогноза, построенного на её основе. Дополнительно используются кросс-валидация на временных рядах, анализ чувствительности и тест с сюррогатными данными.
- Ограничения предложенного подхода
- Требование достаточной длины временного ряда для реконструкции фазового пространства

(эмпирически не менее 50–100 точек для низкоразмерных систем, для хаотических – тысячи точек).

- Чувствительность к выбору параметров встраивания (m , τ) и порогов энтропии (r); автоматический подбор параметров остаётся открытой задачей.
- Вычислительная сложность при высоких размерностях системы (большое m) и при использовании Transfer Entropy на непрерывных данных с ядерными оценками плотности.
- Не всегда применим для сверхкоротких рядов (менее 50 точек) – в этом случае динамическая оценка уступает статическим методам или требует привлечения байесовских подходов с сильными априорными распределениями.

Предложенный энтропийно-робастный подход формирует целостную методологию динамической оценки, которая преодолевает ограничения статических метрик, учитывает качество данных и причинность, и может быть реализована вычислительно. Ограничения подхода не носят принципиального характера и открывают конкретные направления для дальнейших исследований.

Литература

1. Barański K., Gutman Y., Śpiewak A. On the Shroer–Sauer–Ott–Yorke predictability conjecture for time-delay embeddings // *Communications in Mathematical Physics*. – 2022. – Vol. 391. – Pp. 609–641. DOI: 10.1007/s00220-022-04323-y.
2. Najera-Flores D. A., Todd M. D. State space reconstruction from embeddings of partial observables in structural dynamic systems // *Data Science in Engineering*. – Vol. 10. – Cham: Springer, 2023. – DOI: 10.1007/978-3-031-34946-1_18.
3. Любушин А. А., Копылова Г. Н., Касимова В. А., Таранова Л. Н. Мультифрактальные и энтропийные статистики сейсмического шума на Камчатке в связи с сильнейшими землетрясениями // *Компьютерные исследования и моделирование*. – 2023. – Т. 15. – № 6. – С. 1507–1521. – DOI: 10.20537/2076-7633-2023-15-6-1507-1521.
4. Вершик А. М., Лифшиц М. А. О m энтропии банахова пространства с гауссовской мерой // *Теория вероятностей и ее применения*. – 2023. – Т. 68. – № 3. – С. 532–543. – DOI: 10.4213/tvp5637.
5. Smirnov D. A. Information transfers and flows in Markov chains as dynamical causal effects // *Chaos: An Interdisciplinary Journal of Nonlinear Science*. – 2024. – Vol. 34. – No. 3. – 033130. – DOI: 10.1063/5.0187629.
6. Misiorek A. Causal feature selection via transfer entropy // *IEEE Conference on Computational Intelligence and Games (CIG)*. Yokohama, Japan, 2024.
7. Пономаренко В. И., Прохоров М. Д. Анализ причинных связей по временным рядам на основе переносной энтропии // *Известия вузов. Прикладная нелинейная динамика*. – 2023. – Т. 31. – № 4. – С. 456–472.
8. Zhou Y., Wang H., Gu C., Yang H. Detecting causalities between strongly coupled dynamical systems // *Physica A: Statistical Mechanics and its Applications*. – 2024. – Vol. 652. – 129899. – DOI: 10.1016/j.physa.2024.129899.
9. Osada Y., Ushio M., Kondoh M. Unified understanding of nonparametric causality detection in time series // *bioRxiv*. – 2023. – DOI: 10.1101/2023.04.20.537743.
10. Vlachos I., Kugiumtzis D., Paluš M. Causality detection from coupled time series using similarity of embedding projections // *Chaos*. – 2022. – Vol. 32. – 053111. – DOI: 10.1063/5.0085882.
11. Runge J., et al. Causal inference for time series // *Nature Reviews Earth & Environment*. – 2023. – Vol. 4. – Pp. 487–505. – DOI: 10.1038/s43017-023-00431-y.
12. Krakovská A., Jakubík J., Chvosteková M., Coufal D., Jajcay N., Paluš M. Comparison of six methods for the detection of causality in bivariate time series // *Physical Review E*. – 2018. – Vol. 97. – 042207. – DOI: 10.1103/PhysRevE.97.042207.
13. Mønster D., Fusaroli R., Tylén K., Roepstorff A., Sherson J.F. Inferring causality from noisy time series data: A test of Convergent Cross-Mapping // *arXiv:1603.01155*. 2016.

14. Liu C., et al. Incremental feature selection based on uncertainty measure for dynamic interval-valued data // International Journal of Machine Learning and Cybernetics. – 2023. – Vol. 15. – Pp. 1453–1472.
15. Краевский А. Constructing Early Indicators for Features' Interdependent Shift Detection on Financial Markets via Information Metric's Estimation: выпускная квалификационная работа. М.: НИУ ВШЭ, 2023.
16. Seo Y., Baek C. Comparison of time series cross-validation methods for deep learning // Journal of the Korean Data & Information Science Society. – 2024. – Vol. 35. – No. 3. – Pp. 397–410.
17. Pinto H., Lazic I., Antonacci Y., Pernice R., Gu D., Barà C., Faes L., Rocha A.P. Testing dynamic correlations and nonlinearity in bivariate time series through information measures and surrogate data analysis // Frontiers in Network Physiology. – 2024. – Vol. 4. – 1385421. – DOI: 10.3389/fnetp.2024.1385421.

EVALUATION OF DYNAMIC DATA IN THE CONTEXT OF MODERN METHODOLOGIES

Lukashenok V. I.⁷, Panyayev A. I.⁸, Izotov D. Yu.⁹

Keywords: dynamic data, time series, entropic-robust approach, sample entropy, portable entropy, causality, data quality, latency, recurrent algorithms, technical diagnostics, dynamic life cycle assessment.

Abstract

Objective: based on the analysis of the constraints of classical static metrics (correlation, regression, MSE) and existing methods of time series analysis, to substantiate an entropic-robust approach to the dynamic evaluation of systems with memory, nonlinearity, and nonstationarity, which allows simultaneously taking into account uncertainty, causality, quality of input data, and computational constraints.

Research method: the methodology combines the apparatus of the theory of dynamical systems (phase space reconstruction according to Takens), information-theoretical measures (Sample Entropy, Transfer Entropy), robust statistical procedures (adaptive thresholds, median estimates) and recurrent algorithms for online processing. The effectiveness of the approach is demonstrated on two applied tasks: early detection of rotary equipment failures by vibration signals and dynamic assessment of the product life cycle taking into account time variability.

Results of the study: the transition from point metrics to trajectory estimates in phase space is substantiated. It is shown that data quality (completeness, accuracy, stability) and causality (Transfer Entropy, Convergent Cross Mapping) are end-to-end metrics, without which entropy estimates lose their applied reliability. A computational scheme based on recurrent algorithms was developed, taking into account latency and adapting the parameters to the current flow. Verification through a predictive test made it possible to reduce the share of false alarms in diagnostics by 30 % compared to fixed thresholds.

Scientific novelty: for the first time, entropy measures, robust procedures and causal analysis are combined into a holistic methodological approach to dynamic assessment, applicable to both physical (diagnostics) and organizational and technical (LCA) systems.

References

1. Barański K., Gutman Y., Śpiewak A. On the Shroer–Sauer–Ott–Yorke predictability conjecture for time-delay embeddings // Communications in Mathematical Physics. – 2022. – Vol. 391. – Pp. 609–641. – DOI: 10.1007/s00220-022-04323-y.
2. Najera-Flores D. A., Todd M. D. State space reconstruction from embeddings of partial observables in structural dynamic systems // Data Science in Engineering. – Vol. 10. – Cham: Springer, 2023. – DOI: 10.1007/978-3-031-34946-1_18.
3. Lyubushin A. A., Kopylova G. N., Kasimova V. A., Taranova L. N. Mul'tifraktal'nye i ehntropijnye statistiki sejsmicheskogo shuma na Kamchatke v svyazi s sil'nejshimi zemletryaseniyami // Komp'yuternye issledovaniya i modelirovanie. – 2023. – Vol. 15. – No. 6. – Pp. 1507–1521. – DOI: 10.20537/2076-7633-2023-15-6-1507-1521.
4. Vershik A. M., Lifshic M. A. O mm ehntropii banahova prostranstva s gaussovskoj meroj // Teoriya veroyatnostej i ee primeneniya. – 2023. – Vol. 68. – No. 3. – Pp. 532–543. – DOI: 10.4213/tvp5637.
5. Smirnov D. A. Information transfers and flows in Markov chains as dynamical causal effects // Chaos: An Interdisciplinary Journal of Nonlinear Science. – 2024. – Vol. 34. – No. 3. – 033130. – DOI: 10.1063/5.0187629.
6. Misiorek A. Causal feature selection via transfer entropy // IEEE Conference on Computational Intelligence and Games (CIG). Yokohama, Japan, 2024.
7. Ponomarenko V. I., Prohorov M. D. Analiz prichinnyh svyazey po vremennym ryadam na osnove perenosnoj ehntropii // Izvestiya vuzov. Prikladnaya nelinejnaya dinamika. – 2023. – Vol. 31. – No. 4. – Pp. 456–472.
8. Zhou Y., Wang H., Gu C., Yang H. Detecting causalities between strongly coupled dynamical systems // Physica A: Statistical Mechanics and its Applications. – 2024. – Vol. 652. – 129899. – DOI: 10.1016/j.physa.2024.129899.

⁷ Vasily I. Lukashenok, Junior Researcher, Research Center, Military Academy of Communications, St. Petersburg, Russia. E-mail: lukashenokvasilij@gmail.com

⁸ Alexey I. Panyayev, Senior Researcher, Research Center, Military Academy of Communications, St. Petersburg, Russia. E-mail: alex.panyayev@mail.com

⁹ Daniil Yu. Izotov, Researcher, Research Center, Military Academy of Communications, St. Petersburg, Russia. E-mail: daniil.izotov@mail.ru

9. Osada Y., Ushio M., Kondoh M. Unified understanding of nonparametric causality detection in time series // *bioRxiv*. – 2023. – DOI: 10.1101/2023.04.20.537743.
10. Vlachos I., Kugiumtzis D., Paluš M. Causality detection from coupled time series using similarity of embedding projections // *Chaos*. – 2022. – Vol. 32. – 053111. – DOI: 10.1063/5.0085882.
11. Runge J., et al. Causal inference for time series // *Nature Reviews Earth & Environment*. – 2023. – Vol. 4. – Pp. 487–505. – DOI: 10.1038/s43017-023-00431-y.
12. Krakovská A., Jakubík J., Chvosteková M., Coufal D., Jajcay N., Paluš M. Comparison of six methods for the detection of causality in bivariate time series // *Physical Review E*. – 2018. – Vol. 97. – 042207. – DOI: 10.1103/PhysRevE.97.042207.
13. Mønster D., Fusaroli R., Tylén K., Roepstorff A., Sherson J. F. Inferring causality from noisy time series data: A test of Convergent Cross-Mapping // *arXiv:1603.01155*. – 2016.
14. Liu C., et al. Incremental feature selection based on uncertainty measure for dynamic interval-valued data // *International Journal of Machine Learning and Cybernetics*. – 2023. – Vol. 15. – Pp. 1453–1472.
15. Kraevskij A. Constructing Early Indicators for Features' Interdependent Shift Detection on Financial Markets via Information Metric's Estimation: vypusknaya kvalifikacionnaya rabota. M.: NIU VSHEh, 2023.
16. Seo Y., Baek C. Comparison of time series cross-validation methods for deep learning // *Journal of the Korean Data & Information Science Society*. – 2024. – Vol. 35. – No. 3. – Pp. 397–410.
17. Pinto H., Lazic I., Antonacci Y., Pernice R., Gu D., Barà C., Faes L., Rocha A. P. Testing dynamic correlations and nonlinearity in bivariate time series through information measures and surrogate data analysis // *Frontiers in Network Physiology*. – 2024. – Vol. 4. – 1385421. – DOI: 10.3389/fnetp.2024.1385421.

