

Транспортные сооружения / Russian Journal of Transport Engineering <https://t-s.today>

2026, Том 13, № 1 / 2026, Vol. 13, Iss. 1 <https://t-s.today/issue-1-2026.html>

URL статьи: <https://t-s.today/PDF/30SATS126.pdf>

DOI: 10.15862/30SATS126 (<https://doi.org/10.15862/30SATS126>)

2.1.9. Строительная механика (технические науки)

Ссылка для цитирования этой статьи:

Хайруллин, В. А. Теория и метод пороговых значений в строительстве и различных инженерных задачах / В. А. Хайруллин, И. Г. Терехов // Транспортные сооружения. — 2026. — Т. 13. — № 1. — URL: <https://t-s.today/PDF/30SATS126.pdf>. DOI: 10.15862/30SATS126.

For citation:

Khairullin V.A., Terekhov I.G. Theory and method of thresholds in construction and various engineering tasks. *Russian Journal of Transport Engineering*. 2026;13(1): 30SATS126. Available at: <https://t-s.today/PDF/30SATS126.pdf>. DOI: 10.15862/30SATS126. (In Russ., abstract in Eng.).

УДК 69.04

Хайруллин Виталий Агзамович

ФГБОУ ВО «Уфимский государственный нефтяной технический университет», Уфа, Россия
Научно-исследовательский проектный институт архитектуры и строительства
Ведущий научный консультант
E-mail: Vitalik000@yandex.ru
ORCID: <https://orcid.org/0000-0002-3854-2193>
РИНЦ: https://elibrary.ru/author_profile.asp?id=671334
ResearchGate: https://www.researchgate.net/profile/Vitaliy-Khayrullin?ev=hdr_xprf

Терехов Иван Геннадьевич

ФГБОУ ВО «Уфимский государственный нефтяной технический университет», Уфа, Россия
Архитектурно-строительный институт
Исполняющий обязанности директора, доцент
Кандидат технических наук, доцент
E-mail: iv_98_04@mail.ru
РИНЦ: https://elibrary.ru/author_profile.asp?id=528528

Теория и метод пороговых значений в строительстве и различных инженерных задачах

Аннотация. Данное исследование логически продолжает предыдущую работу авторов, посвященную Теории экстремальных значений в оценке долговечности строительных конструкций. Объект исследования — теория экстремальных значений. Предмет исследования — теория пороговых значений. Цель исследования — представить прикладное применение Теории пороговых значений в строительстве и в различных инженерных задачах. В этом исследовании мы рассмотрим ключевые работы предметной области. Также мы ответим на вопрос о допустимости минимального размера выборки $n = 30$ в предыдущем исследовании — это важный прикладной момент, который некоторыми исследователями был интерпретирован как ошибка расчёта. Мы подробно ответили на этот вопрос. Сформулированы ключевые методологические различия между экстремальными значениями (метод блочных максимумов) и пороговыми значениями (все превышения заданного уровня). Предложена адаптированная для инженерных задач (прежде всего в строительстве) пошаговая процедура Метода пороговых значений, включающая выбор порога, оценивание параметров обобщённого распределения Парето, проверка адекватности модели и расчёт вероятностных прогнозов с возвратными уровнями. Практическая апробация метода выполнена на реальных данных вертикальных перемещений узлов сливо-наливной эстакады. Авторами продемонстрировано, что даже при

ограниченном объёме исходной информации (123 наблюдения) Метод пороговых значений позволяет: (1) обоснованно выбрать порог $u = 2,0$ мм по графику среднего превышения; (2) получить релевантные оценки параметров GPD ($\hat{\xi} = 0,58$, $\hat{\sigma} = 2,94$ мм), указывающие на тяжёлый хвост распределения; (3) подтвердить адекватность модели с помощью QQ-графика и расстояния Колмогорова ($d = 0,069$); (4) рассчитать годовые вероятности превышения нормативного прогиба (30 мм) и кумулятивные риски прогибов выше нормативного значения за 5 и 10 лет (61 % и 85 % соответственно), а также возвратные уровни для периодов повторяемости 5, 10 и 20 лет. Также мы в очередной раз рассмотрели важное прикладное затруднение — интерпретация временной шкалы, которая играет ключевую роль для эффективных вероятностных прогнозов технического состояния инженерной конструкции.

Ключевые слова: теория пороговых значений; обобщённое распределение парето; теория экстремальных значений; метод превышений порога; возвратные уровни; теорема Гнеденко; теорема Пикандса-Балкемы-де Хаана; теория Караматы

Постановка проблемы

В предыдущем исследовании — «Теория экстремальных значений в оценке долговечности объекта строительства» [1] — мы писали: «Следующий этап качественного развития теории экстремальных значений — это работа 1974 года «Остаточное время жизни в преклонном возрасте» [8] голландских математиков Гуса Балкемы и Лоренса де Хаана... Теорема 2 в работе Г. Балкемы и Л. де Хаана известна в настоящее время как теорема Пикандса-Балкемы-де Хаана. Как видно из названия теоремы, впереди появляется фамилия Пикандса, и в отличие от теоремы Гнеденко (в англоязычных источниках — теорема Фишера-Типпета-Гнеденко), где первые две фамилии вызывают вопросы, в этом случае подобное наименование теоремы полностью оправдано. В 1975 году Джеймс Пикандс, американский математик, опубликовал работу «Статистические выводы с использованием статистики крайних порядков» [9], в которой он, используя результаты Г. Балкемы и Л. де Хаана, существенно развил предметную область, предложив практический метод оценки хвостов распределений, который лег в основу современных подходов в оценке пороговых значений в теории экстремальных значений» [1, с. 6–7]. Из этого отрывка у исследователей, глубоко не погруженных в предметную область, может сложиться неверное понимание, что единственными создателями теории пороговых значений являются нидерландские математики: Гуус Балкема (англ. A.A. Balkema), Лауренс де Хаан (англ. Laurens de Haan) и американский математик Джеймс Пикандс (англ. James Pickands). Это правда, но не вся. Поясним. Так как предметная область предыдущего исследования — Теория экстремальных значений, то о Теории пороговых значений (дальнейшей эволюции Теории экстремальных значений) мы сказали «вскользь». Наше упущение в предыдущем исследовании заключается в том, что мы не уточнили, что Г. Балкема, Л. де Хаан и Д. Пикандс опирались, конечно, на фундаментальные исследования, в том числе советских учёных. Теорема Пикандса — Балкемы — Де Хаана, известная как Вторая теорема экстремальных значений — это больше синтетическая конструкция. Это синтез идей и популяризация того, что сделано многими учёными. Здесь следует пояснить. Балкема и Де Хаан исследовали остаточное время жизни в преклонном возрасте людей (работа 1974 года «Остаточное время жизни в преклонном возрасте» [2]). В ней они ссылаются на работу 1943 года ключевого создателя Теории экстремальных значений — академика Б.В. Гнеденко «О предельном распределении максимального члена случайной последовательности» [3]. Важный методологический момент — Балкема и Де Хаан ни о каких «пороговых значениях» не рассуждали (мы разберём их работу далее). Сделал это американец Джеймс Пикандс в работе 1975 года «Статистические выводы с использованием статистики крайних порядков» [4]. При этом Пикандс на странице 131, ссылаясь на работу академика

Гнеденко, указал неверную дату публикации: «*GNEDENKO, B. V. (1973). Sur la distribution limite du terme maximum d'une serie aleatoire. Ann. of Math. 44 423-453*» [4, с. 131]. Он указывает 1973 год, но это неправильно, так как работа была опубликована в 1943 году. Далее в начале исследования Пикандс пишет: «*Мы говорим, что $A(x)$ является «экстремальной функцией распределения», а $F(x)$ лежит в её «области притяжения». Функция распределения $A(x)$ должна принадлежать одному из трёх параметрических семейств. Это теорема о трёх типах экстремальных значений (тройственная теорема экстремальных значений). См. Гнеденко [3], или Гумбель [5], Глава 4. Области притяжения были охарактеризованы рядом авторов. Самая недавняя и наиболее полная работа — это книга де Хаана [2]. Большинство «стандартных» (из учебников) непрерывных семейств функций распределения обсуждаются в [5]. Все они лежат в области притяжения некоторой экстремальной функции распределения*» [4, с. 119]. В этой цитате наглядно видна особенность изложения в англосаксонской академической традиции. Пикандс ссылается на два учебника, при этом пишет: «... Области притяжения были охарактеризованы рядом авторов. Самая недавняя и наиболее полная работа — это книга де Хаана...» [4, с. 131]. Но это не совсем верно. Первую теорему экстремальных значений, как и области притяжения, создал и охарактеризовал академик Б.В. Гнеденко. Вклад в развитие теории экстремальных значений нидерландского математика Лоуренса де Хаана и немецкого математика Эмиля Юлиуса Гумбеля неоспорим, но не они сформулировали и доказали Первую теорему экстремальных значений. Д. Пикандс ссылается на их учебники, в которых излагаются вопросы предметной области и обобщены результаты многих учёных. Пикандс, ссылаясь на чужие учебники, синтезировал чужие идеи, при этом вклад самого Пикандса неоспорим. Вообще, ссылки на учебники за авторством одних учёных, которые освещают вопросы предметной области и обобщают результаты других учёных, искажают историю вопроса, нарушая фактологическую точность. Что и произошло в работе Д. Пикандса. Если открыть её, то складывается впечатление, что ключевые фигуры Лоуренс де Хаан и сам Пикандс. Тем не менее ключевые фигуры, способствующие созданию Теории пороговых значений, были, в том числе, и наши соотечественники. И это, прежде всего, Академик Б.В. Гнеденко; Математик, член-корреспондент АН СССР (1960) Николай Васильевич Смирнов (1900–1966); академик Андрей Николаевич Колмогоров (1903–1987), предложенная им статистика максимального расхождения между эмпирическим и теоретическим распределениями (расстояние Колмогорова) случайной величины впоследствии нашла применение в методе POT для проверки согласия эмпирической функции распределения превышений $\hat{F}_M(y)$ с подобранной GPD и для выбора оптимального порога. Также пусть не прямой, но всё равно важный вклад внесли-ученик академика А.Н. Колмогорова — профессор Юрий Константинович Беляев (1932–2025); советский и российский математик, академик РАН РФ (1990) Александр Алексеевич Боровков.

Наша основная задача — воспроизвести историю вопроса, показать эволюцию и преемственность идей предметной области. Настоящее исследование — это логическое продолжение нашего первого исследования по теории экстремальных значений. Работа состоит из двух основных частей: первая часть — это общие вопросы теории, история вопроса, методология и вторая часть — это прикладные вопросы и практика реализации теории пороговых значений в строительстве.

Некоторые соображения по предыдущему исследованию «Теория экстремальных значений в оценке долговечности объекта строительства»

Ряд коллег нам поставили в упрёк, что на странице 18 в «*таблица 3 Минимальные значения прочности бетона (по всей выборке), Мпа*» [1, с. 18] мы представили слишком малую выборку (30 значений). И что это ошибка, так как в выборках экстремальных значений должно

быть минимум 50 значений, согласно «правилу большого пальца». Действительно существует так называемое «правило большого пальца»¹ (англ. rule of thumb), которое в Теории экстремальных значений многие неправильно интерпретируют, следующим образом: минимальный размер выборки экстремальных значений должен быть не менее 50 значений. Здесь следует пояснить. Правило действительно существует, но оно отнюдь не устанавливает минимальный размер выборки — 50 значений. Откуда взялось заблуждение? Это, прежде всего, работа 1985 года Дж.Р.М. Хоскинга, Дж.Р Уоллиса, Э.Ф. Вуда «Оценивание обобщённого распределения экстремальных значений методом вероятностно-взвешенных моментов» [5], в которой они пишут следующее: *«Оценка экстремальных нижних квантилей, как правило, менее важна на практике, чем оценка верхних квантилей, поэтому результаты моделирования для этого случая не приводятся подробно. Все три метода дают сравнимые результаты при $n \geq 50$, но для малых выборок оценки РВМ имеют наименьшее стандартное отклонение и небольшое или умеренное смещение и, как правило, являются предпочтительными. Все методы оценивания квантилей очень неточны при оценивании экстремальных квантилей в малых выборках с $k < 0$. Разумеется, следует ожидать, что квантиль $x(F)$ не может быть надёжно оценён по выборке объёма n , если $F > 1 - 1/n$. Тем не менее иногда можно получить полезные оценки экстремальных квантилей по коротким наборам данных, комбинируя информацию из нескольких независимых наборов данных»* [5, с. 257]. Выражение из цитаты — $n \geq 50$ многими исследователями были понято как «жёсткое» правило минимального размера выборки, но это не так. И это видно из самой цитаты, в которой не утверждается, что выборка обязательно должна быть равна или быть больше пятидесяти наблюдений (значений). Хоскинг и др. лишь констатируют, что при $n \geq 50$ различия между методами сглаживаются, и все они работают примерно одинаково хорошо — это эмпирический факт, полученный методом Монте-Карло. Для выборок меньше 50 значений (малых выборок) точность оценок падает, и метод вероятностно-взвешенных моментов (англ. РВМ) оказывается более устойчивым. Сама по себе возможность работы с выборками меньше 50 значений не отрицается, просто нужно более ответственно выбирать метод оценивания. Что мы и сделали. Мы намеренно привели сложный пример с малой выборкой, потому что как раз там и возникают прикладные затруднения у специалистов. И мы показали, как их решать. Фраза *«...нельзя надёжно оценить квантиль $x(F)$ по выборке объёма n , если $F > 1 - 1/n$ »* формулирует известное правило: нельзя экстраполировать дальше, чем позволяет объём данных, и это общее соображение, а не конкретное требование к размеру выборки. При этом сами авторы указывают вполне конкретно (в цитате выделено полужирным начертанием), что можно получать вполне полезные оценки по малым наборам данных. И ровно это мы и показали. Размер выборки, конечно, имеет значение, и чем он больше, тем более достоверную оценку на большем перспективном периоде вы получите. Прикладная проблема с экстремальными значениями состоит в том, что их мало, как бы странно это ни звучало — их мало и потому они экстремальны, из-за этого они естественно редки. Мы показали именно прикладную и практическую проблему малых выборок экстремальных значений и как её решать. Отсюда утверждение, что мы допустили ошибку в размере выборки, — ложно. Мы не допускали никакой ошибки. Наоборот, целенаправленно и на реальном примере взяли сложный случай, когда размер выборок минимален. Именно для того, чтобы наглядно показать, как действовать, чтобы достоверно установить закон распределения при минимальном размере выборки экстремальных значений.

Общие вопросы теории

Прежде чем приступить к истории вопроса, рассмотрим, что такое Теория пороговых значений и чем она отличается от Теории экстремальных значений. Для этого нам надо провести различие между понятиями: «экстремальное значение» и «пороговое значение».

¹ Эмпирическая закономерность.

Экстремальные значения

В широком смысле экстремальные значения — это наибольшие (или наименьшие) наблюдения в выборках фиксированного объёма. В теории экстремальных значений (англ. EVT) изучается поведение максимумов (или минимумов) независимых одинаково распределённых случайных величин, если разбить временной ряд на блоки равной длины (например, годовые интервалы) и в каждом блоке выбрать максимальное значение, то полученная совокупность максимумов подчиняется (при определённых условиях) обобщённому распределению экстремальных значений (GEV).

Пример: ежедневные замеры скорости ветра за 50 лет, мы формируем 50 годовых максимумов (самый сильный порыв ветра за каждый год) — это экстремальные значения (в данном случае — годовые максимумы). Теорема Гнеденко (первая теорема экстремальных значений или в англосаксонской традиции: теорема Фишера-Типетта-Гнеденко): «Класс предельных законов для $F(a_n x + b_n)$, где $a > 0$ и b_n — подходящим образом выбранные константы, содержит только законы типов $\Phi_a(x)$, $\Psi_a(x)$, $A(x)$ » [3, с. 431]. Объясним более простым языком.

Теорема 3 Гнеденко объясняет, что если мы берём максимумы (самые большие значения за год, за месяц и т. д.) и приводим их к единому масштабу, то эти максимумы могут вести себя только тремя способами и никаких других вариантов природа не предусмотрела.

Вот эти три способа поведения (типа):

Тип I — распределение Гумбеля — это когда хвост распределения случайной величины убывает быстро, как экспонента (хвост распределения случайной величины — это такие области на графике², где находятся самые большие (правый хвост) или самые маленькие (левый хвост) значения (когда пишут просто «хвост», обычно подразумевают правый хвост (для максимумов) — это общепринятая терминология, но для строгости можно уточнять: «правый хвост распределения»).³ Приведём простой пример — рост людей. Максимальный рост в выборке из десятков тысяч людей почти наверняка не превысит 2,5 метра, хотя теоретически может быть любым, но при этом вероятность экстремальных (запредельных) значений падает очень резко.

Тип II — распределение Фреше — это когда хвост распределения случайной величины убывает медленно (степенным образом). Например, размер страховых выплат при природных катаклизмах, когда вероятность катастрофически больших убытков хоть и мала, но как бы «тянется» далеко. Так как бывают ураганы, которые в несколько раз сильнее обычных, катастрофические цунами, сильнейшие землетрясения и они хоть и редки, но отнюдь не исключаемы. Распределение Фреше — это своего рода редкие экстремальные значения.

Тип III — распределение Вейбулла, когда хвост обрывается, так как существует верхняя граница значений, выше которой значения просто невозможны. Пример: прочность бетона на сжатие. Сколько ни улучшай технологию, бесконечно прочным бетон быть не может, потому что есть физический предел, и максимум всегда будет ниже этого предела. То же самое с долговечностью или прочностью материала — бесконечно эксплуатировать или бесконечно прочным материал быть не может.

² Рисунок 5 на странице 11 в исследовании «Теория экстремальных значений в оценке долговечности объекта строительства» [1].

³ Подробно про хвосты распределения случайной величины мы объяснили в нашем первом исследовании «Теория экстремальных значений в оценке долговечности объекта строительства» [1] на страницах 11–13.

Пороговые значения (англ. Peaks Over Threshold)

Метод пороговых значений (англ. POT — Peaks Over Threshold) рассматривает все значения, превышающие некоторый достаточно высокий порог u . При этом изучается условное распределение величины превышения $Y = X - u$ при условии $X > u$. Согласно теореме Пикандса-Балкема-де Хаана, при стремлении порога к границе области определения исходного распределения, это условное распределение сходится к обобщённому распределению Парето (англ. GPD). Пороговый подход позволяет использовать все экстремальные события (а не только одно за блок), что повышает эффективность оценок при ограниченном объёме данных. Например, все те же ежедневные замеры скорости ветра. Мы задаём порог $u = 25$ м/с. Все дни, когда скорость ветра превышала 25 м/с, дают значения превышений Y (сколько именно метров в секунду сверх порога), эти превышения и есть «пороговые значения». Их количество обычно намного больше 30. Соответственно они позволяют точнее оценить хвост распределения нагрузок. То есть пороговые значения — это такие значения, которые либо превысили, либо стали ниже конкретного значения (порога). И это существенно отличается от экстремальных значений, где вы берёте только максимумы (или минимумы) по периоду или выборке. В практическом смысле Теория пороговых значений имеет большую прикладную применимость элементарно за счёт большего размера выборки. Вернёмся к случаю с замерами прочности бетона (из статьи «Теория экстремальных значений в оценке долговечности объекта строительства» [1]). Там мы располагали данными испытаний множества образцов (615), и мы выбрали 30 самых низких значений прочности — это были экстремальные минимумы. Но метод POT предлагает иной путь — мы устанавливаем порог u (например, 31,5 МПа) и рассматриваем все образцы, чья прочность оказалась ниже этого порога. Для левого хвоста (минимумов) это будет аналогия «превышений снизу».

Ключевое различие между экстремальными и пороговыми значениями можно свести к следующему:

- Экстремальные значения — это редкие события, отобранные по принципу «одно самое большое (или самое маленькое) значение за период времени (блок)» и их количество равно числу блоков. Поясним про «блок», мы не сделали этого в предыдущем исследовании. Метод блочных максимумов — это способ подготовки данных для анализа экстремальных значений, при котором исходный ряд наблюдений разбивают на равные промежутки времени (блоки, например, годы), а затем из каждого такого промежутка берут только одно максимальное значение. Полученная совокупность этих максимумов и служит материалом для дальнейшего статистического анализа. Этот подход удобен своей простотой и широко применяется в гидрологии, климатологии и других областях, где есть естественные периоды (год, сезон), но у него есть недостаток — если в каком-то году случилось несколько сильных событий, в выборку попадёт только самое сильное, а остальные будут потеряны. В нашем предыдущем исследовании мы использовали другой принцип (классический подход Фреше, Гнеденко) — брали не по одному значению за период, а отобрали самые низкие значения прочности из всего массива данных, что также является корректным способом получить экстремальные минимумы, но это не метод блочных максимумов.

- Пороговые значения (метод POT) — это все события, которые перешли определённую границу u . Их количество может быть любым и зависит от того, насколько часто процесс «заходит» за эту границу. Почему метод POT считается более эффективным? Главное преимущество POT заключается в том, что он не «выбрасывает» информации. В Теории экстремальных значений, и особенно в Методе блочных максимумов, мы берём только один максимум за год и теряем все остальные сильные изменения, случившиеся в этом же периоде. А в методе POT мы учитываем все превышения порога, что особенно важно, когда данных мало, а с экстремальными значениями это бывает практически всегда. Но если мы используем

Метод пороговых значений, то в некоторых случаях мы практически гарантировано избавляемся от нехватки данных и можем получить более достоверную картину.

Теперь, когда мы разобрались с основными понятиями, перейдём к истории вопроса и посмотрим, как и кем на самом деле закладывались основы теории пороговых значений.

ИСТОРИЯ ВОПРОСА

Работа Г. Балкемы и Л. де Хаана

«Остаточное время жизни в преклонном возрасте» [2] 1974 год

Статья посвящена асимптотическому поведению остаточного времени жизни (англ. residual life time) при больших значениях возраста t людей. Если X — время жизни (положительная случайная величина) человека с функцией распределения $F(x)$ и хвостом $R(x) = 1 - F(x) > 0$ для всех x , то остаточное время жизни человека в момент t определяется условным распределением:

$$R_t(x) = P\{X - t > x \mid X > t\} = \frac{R(x+t)}{R(t)}, x \geq 0. \quad (1)$$

Авторы исследуют, к каким предельным распределениям может сходиться (после подходящей нормировки) $R_t(x)$ при $t \rightarrow \infty$. Полученные результаты тесно связаны с теорией экстремальных значений и, в частности, они ложатся в основу теоремы Пикандса-Балкемы-де Хаана о распределении превышений над высоким порогом.

Во введении авторы приводят релевантный пример с лампочкой: после того как она проработала t часов, остаточное время жизни имеет распределение $R_t(x)$. Естественно спросить, какие предельные распределения возможны для $R_t(x)$ при $t \rightarrow \infty$ и при подходящем выборе масштабных параметров: «Рассмотрим электрическую лампочку. Она имеет некоторый срок службы X , который является случайной величиной с распределением вероятностей F . Предположим, что хвост распределения $R(x) = 1 - F(x) = P\{X > x\}$ положителен для всех x . После того как лампочка проработала t часов, остаётся остаточный срок службы с хвостом распределения R_t , определяемым как

$$R_t(x) = 1 - F_t(x) = P\{X - t > x \mid X > t\}.$$

Представляет значительный интерес знать асимптотическое поведение этих распределений остаточного срока службы при $t \rightarrow \infty$. Естественные вопросы: каковы возможные типы предельных распределений? Для каждого такого предельного распределения G какова его область притяжения $D_r(G)$ (т. е. множество всех функций распределения F , таких что F_t при надлежащей нормировке сходится к G)? Какова скорость сходимости? Настоящая статья даёт достаточно полные ответы на все эти вопросы. Большинство доказательств имеют довольно технический аналитический характер» [2, с. 792]. Далее авторы сразу отмечают, что если хвост экспоненциальный ($R(x) = e^{-\lambda x}$), то $R_t(x) = e^{-\lambda x}$ для всех t , т. е. распределение не меняется. Это хорошо известное характеристическое свойство экспоненциального закона. И потом Балкема и Де Хаан формулируют ключевую идею, что если разрешить только масштабирование, то есть предположить существование положительной функции $a(t)$ такой,

$$R_t(a(t)x) = \frac{R(t + a(t)x)}{R(t)} \rightarrow S(x), \quad (2)$$

то возможные непрерывные пределы $S(x)$ оказываются только двух типов: экспоненциальный $P(x) = e^{-x}$ и Паретовский $\Gamma_\alpha(x) = (1+x)^{-\alpha}$ («теорема 2» [2, с. 123]). Если же дополнительно разрешить сдвиг, т. е. рассматривать

$$\frac{R(b(t) + a(t)x)}{R(t)} \rightarrow S(x), \quad (3)$$

то помимо этих непрерывных распределений, появляются также дискретные предельные типы $\Pi_{p,c}$ и $\Gamma_{\gamma,\alpha}$ («теорема 1» [2, с. 123]). В последнем случае нормировка включает и перенос $b(t)$, что позволяет учесть ситуацию, когда распределение исходной случайной величины сосредоточено на геометрической прогрессии. В работе Балкемы и де Хаана получены предельные распределения для остаточного времени жизни, которые включают как экспоненциальное ($\xi = 0$), так и паретовские распределения с $\xi > 0$ и $\xi < 0$ (последнее соответствует ограниченному хвосту), но (и это важно!), они рассматривали именно остаточное время, а не превышения над произвольным порогом, и не ввели единую параметризацию GPD. Пикандс в 1975 году перенёс именно эти результаты на задачу о превышениях и предложил удобную трёхпараметрическую форму GPD, объединяющую все три случая. Если же разрешить ещё и сдвиг, т. е. рассматривать $R_t(b(t) + a(t)x)$, то появляются также дискретные предельные распределения (эти предельные типы будут выведены у Пикандса в разделе 1). Становится видна аналогия с теорией экстремальных значений, где для максимумов существуют три типа предельных распределений (Фреше, Вейбулл, Гумбель), а Г. Балкема и Л. Де Хаан подчёркивают эту связь: «Мы отмечаем, что эти предельные распределения имеют некоторое сходство с предельными распределениями Φ_α и Λ в теории экстремальных значений (см. раздел 2). Это сходство обусловлено тем, что обе теории имеют общую основу — теорию регулярного варьирования Караматы.⁴ В разделе 5 мы покажем, как теоремы Караматы

⁴ Этот момент в цитате следует пояснить. Так как он значим, и абсолютно справедливо Балкема и Де Хаан на него обратили внимание. Что такое Теория регулярного варьирования Караматы и почему она важна. Теория разработана сербским математиком Йованом Караматой в 1930-х годах [6; 7] (наиболее полно изложена в книге английских исследователей «Регулярное варьирование» [8] 1985 года). Теория Караматы стала фундаментом для многих областей теории вероятностей, включая Теорию экстремальных значений и Теорию пороговых значений. Приведём простой пример, в котором наглядно объясним, насколько важна эта теория. Представьте, что вы проектируете высотное здание в регионе, где возможны ураганы. Ваша задача — рассчитать так называемую расчётную ветровую нагрузку — максимально возможную скорость ветра, которую здание должно выдержать с вероятностью, скажем, 99,9 % за весь срок службы (например, за 100 лет) У вас есть данные метеонаблюдений за последние 50 лет — ежедневные замеры максимальной скорости ветра. Вы выбрали 50 годовых максимумов (самые сильные порывы за каждый год) и по этим 50 значениям вам нужно предсказать, какой ветер может случиться раз в 100 лет (квантиль уровня 0,99) или даже раз в 200 или 300 лет (для особо ответственных сооружений, например АЭС). И тут возникает серьёзная прикладная проблема: как экстраполировать за пределы наблюдений? Так как ветров сильнее, чем за эти 50 лет, ещё не было, но при этом они возможны. Просто взять и продолжить гистограмму «на глаз» нельзя, нужно математически обоснованное поведение хвоста распределения. Что даёт прикладникам и практикам теория регулярного варьирования Караматы? Она объясняет инженеру: смотри не на всю гистограмму, а только на хвост — на самые большие значения. Так как согласно Теории Караматы, для очень многих природных явлений (скорость ветра, высота волн, интенсивность осадков) хвост ведёт себя не как экспонента (быстрое убывание), а как степенная функция: $P(X > x) \approx C \cdot x^{-\alpha}$, для больших x , где $\alpha > 0$ — параметр, характеризующий «тяжесть» хвоста. Это и есть регулярное варьирование. Теперь у инженера появляется мощный инструмент: оценка параметра α по имеющимся данным, например, по 50 максимумам, (есть специальные статистические методы (оценка Хилла и др.), которые используют именно теорию Караматы). Экстраполяция — зная α , можно вычислить ветер 100-летней повторяемости: $x_{100} \approx C' \cdot (100)^{1/\alpha}$. То есть мы можем предсказать событие, которое ещё ни разу не наблюдалось, но с математической точки зрения вполне возможно. Теория Караматы говорит: если хвост регулярно меняется (степенной), то максимумы будут сходиться к распределению Фреше, а превышения над высоким порогом к обобщённому распределению Парето (GPD). Отсюда инженер точно знает, какую модель применять, и может рассчитать не только среднюю нагрузку, но и доверительные интервалы для своей оценки.

«Замечательный» результат Караматы, про который пишут Балкема и Де Хаан (приведём в интерпретации для инженеров) следующий: оказывается, сходимость одного единственного условного момента (например, среднего превышения над порогом) эквивалентна тому, что распределение вообще сходится к правильному предельному закону. На практике это значит следующее — чтобы убедиться в том, что ваша модель хвоста работает, достаточно проверить, стабилизируется ли с ростом порога некая простая статистика (например, отношение среднего превышения к порогу), и это действительно колоссально упрощает жизнь проектировщику, так как не нужно строить сложные многопараметрические модели, можно ограничиться простыми проверками.

преобразуются в замечательную теорему о моментах: сходимость положительного момента остаточного времени жизни при $t \rightarrow \infty$ эквивалентна слабой сходимости (надлежащим образом нормированных) распределений остаточного времени жизни F_t » [2, с. 793].

В разделе 1 авторы Балкема и Де Хаан доказывают, что если для некоторой функции распределения F с хвостом $R(x) > 0$ существуют нормирующие константы $a(t) > 0$ и $b(t)$ такие, что

$$P\left(\frac{X - b(t)}{a(t)} > x \mid X > t\right) \rightarrow S(x) \text{ слабо при } t \rightarrow \infty \quad (4)$$

и $I - S$ — невырожденная функция распределения, то S может быть только одного из трёх типов (с учётом сдвига и масштаба):

Экспоненциальный тип (обозначается Π):

$$\Pi(x) = e^{-x}, x \geq 0. \quad (5)$$

Парето-тип (обозначается Γ_α , $\alpha > 0$):

$$\Gamma_\alpha(x) = (1 + x)^{-\alpha}, x \geq 0. \quad (6)$$

Дискретные типы $\Pi_{p,c}$ и $\Gamma_{\gamma,\alpha}$, которые появляются при наличии сдвига. Они соответствуют дискретным распределениям, сосредоточенным на геометрической прогрессии.

Приводя различные случаи ($A(y) = I$ или $> I$), авторы получают все возможные типы. Приведём дословную формулировку из статьи: «Теперь мы готовы доказать, что слабые предельные распределения остаточных времён жизни являются в точности теми, которые упоминались во введении».

Теорема 1. Пусть F — функция распределения такая, что $F(x) < I$ для всех x . Пусть нормированные распределения остаточного времени жизни $F_t(b(t) + xa(t))$ с F_t , определёнными в (1), сходятся слабо к невырожденной функции распределения $G(x)$ при $t \rightarrow \infty$. Тогда G принадлежит одному из типов Π , Π_γ , Γ_α или $\Gamma_{\alpha,r}$, определённых во введении» [2, с. 796]. Здесь Π — экспоненциальное, Π_γ — дискретное экспоненциальное (сосредоточенное на точках, кратных γ), Γ_α — непрерывное Паретовское, $\Gamma_{\gamma,\alpha}$ — дискретное Паретовское. То есть все возможные предельные распределения для остаточного времени жизни (с масштабированием и сдвигом) — это либо непрерывные (экспонента и Парето), либо их дискретные аналоги.

Во втором разделе авторы устанавливают важнейшую концептуальную и методическую связь: области притяжения для остаточного времени жизни совпадают с областями притяжения для максимумов (теоремы Гнеденко) — это ключевой момент, показывающий единство теории. «Теорема 3 $D_r(\Pi) = D(\Lambda) \cap D_0$ » [2, с. 798]. Здесь $D_r(\Pi)$ — область притяжения для остаточного времени жизни к экспоненциальному пределу, $D(\Lambda)$ — область притяжения к распределению Гумбеля в теории экстремальных значений, D_0 — множество распределений с бесконечным носителем. Теорема утверждает, что распределение принадлежит области притяжения экспоненциального остаточного времени тогда и только тогда, когда оно принадлежит области притяжения закона Гумбеля для максимумов. «Теорема 4 $D_r(\Gamma_\alpha) = D(\Phi_\alpha)$ для всех $\alpha > 0$ » [2, с. 798]. Здесь $D(\Phi_\alpha)$ — область притяжения распределения Фреше (тяжёлый хвост). То есть остаточное время жизни сходится к Паретовскому пределу тогда и только тогда, когда максимумы сходятся к закону Фреше с тем же параметром α . Если хвост исходного распределения убывает как степенная функция (тяжёлый хвост), то и остаточное время жизни, и максимумы ведут себя одинаково асимптотически. Если хвост убывает быстро (например, экспоненциально), то остаточное время жизни стремится к экспоненте, а максимумы к закону Гумбеля. Можно сказать, что теория остаточного времени жизни и теория экстремальных значений — это две стороны одной медали.

Раздел 3 посвящён дискретным предельным распределениям. В нём Балкема и де Хаан показывают, что предельные распределения возникают, если исходное распределение сосредоточено на точках, образующих геометрическую прогрессию. Распределение F принадлежит области притяжения дискретного предела $\Pi_{p,c}$ тогда и только тогда, когда оно в хвосте эквивалентно дискретному распределению с точками роста t_n , для которых

$$\frac{t_{n+1} - t_n}{t_n - t_{n-1}} \rightarrow e^{pc} \quad (7) \quad \frac{R(t_{n+1})}{R(t_n)} \rightarrow e^{-p}. \quad (8)$$

Хотя эти результаты менее важны для приложений, они завершают полную картину возможных пределов. Сама теорема Пикандса-Балкема-де Хаана обычно формулируется так: для случайной величины X с функцией распределения F и для высокого порога u условное распределение превышений $Y = X - u$ при $X > u$ сходится (при $u \rightarrow x^*$) к обобщённому распределению Парето (GPD)

$$G(y; \xi, \sigma) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{-\frac{1}{\xi}}, y \geq 0. \quad (9)$$

где $\xi \in \mathbb{R}, \sigma > 0$.

В работе Балкемы и де Хаана получен тот же результат для остаточного времени жизни (которое есть частный случай превышений при $u = t$ и $Y = X - t$). Они показали, что возможные пределы — это экспонента ($\xi = 0$), и паретовские распределения ($\xi > 0$). Случай $\xi < 0$ (распределения с конечной верхней границей) в их работе не рассматривается, потому что остаточное время жизни определено для всех t только при условии, что хвост бесконечен (иначе при t близком к верхней границе нормировки становятся иными). Позже, в 1975 году, Джеймс Пикандс распространил этот результат на случай произвольного порога и включил третий тип ($\xi < 0$), объединив все три в GPD.

Ключевая работа Д. Пикандса «Статистический вывод с использованием экстремальных порядковых статистик» [4] 1975 год

В статье Джеймса Пикандса вводится метод оценки хвоста распределения по нескольким самым большим наблюдениям (экстремальным порядковым статистикам). Его главная идея — условное распределение превышений над достаточно высоким порогом хорошо аппроксимируется обобщённым распределением Парето (GPD). На самом деле это та же идея, что и у Балкемы и де Хаана в работе 1974 года, но Пикандс переносит её на практическую задачу статистического вывода и предлагает конкретные оценки параметров.

На какие вопросы Пикандс ответил:

1. Что делаем?

Берём выборку из n наблюдений (например, ежедневные максимумы ветра за много лет). Выбираем порог u — достаточно большое значение, чтобы за ним осталось, скажем, k наблюдений (Пикандс предлагает брать $k = 4M$ и использовать M как параметр). Рассматриваем превышения $Y = X - u$ для тех $X > u$.

2. Какая модель?

На страницах 120–121 Пикандс утверждает, что для достаточно высокого порога u условное распределение превышений $Y = X - u$ (то есть распределение случайной величины X при условии $X > u$) с хорошей точностью описывается обобщённым распределением Парето

(GPD), которое имеет три формы в зависимости от знака параметра формы c^5 (в современных обозначениях ξ):

- I. $c > 0$ (тяжёлый хвост): $1 - G(x) = (1 + cx/a)^{-1/c}$ (распределение Парето);
- II. $c = 0$ (показательный хвост): $1 - G(x) = e^{-x/a}$ (экспонента);
- III. $c < 0$ (ограниченный хвост): $1 - G(x) = (1 - |c| x/a)^{1/|c|}$ при $0 \leq x \leq a/|c|$, и 0 при больших x .

Параметр c (в современных обозначениях ξ) — тот же, что в GEV.

3. Как происходит оценивание?

Пикандс предлагает оценивать c и a по квантилям 50 % и 75 % эмпирического распределения превышений. Пусть $x_{0.5}$ и $x_{0.75}$ — выборочные квантили (для отобранных превышений). Тогда оценка параметра формы:

$$\hat{c} = \frac{1}{\ln 2} \ln \left(\frac{x_{0.75} - x_{0.5}}{x_{0.5}} \right). \quad (10)$$

Оценка масштаба a выражается через $x_{0.5}$ и \hat{c} .

4. Как осуществляется выбор порога?

Это наиболее важный вопрос, так как неправильно выбранный порог делает бессмысленными результаты оценивания. Поясним на простом примере. Представьте, что вы инженер-строитель и вам нужно рассчитать, какую ветровую нагрузку должно выдерживать здание, чтобы гарантировать его безопасность в течение следующих 100 лет эксплуатации. У вас есть данные наблюдений за скоростью ветра за последние 50 лет. Допустим, вы решили анализировать все дни, когда скорость ветра превышала 10 м/с (порог выбрали 10 м/с⁶, но таких дней окажется очень много (например, несколько тысяч) и среди них будет множество обычных, неопасных ветров, которые не несут угрозы зданию. Поэтому если вы построите модель по этим данным, она как бы «размажется» и будет описывать не столько редкие ураганы, сколько повседневную погоду. В результате оценка ветра 100-летней повторяемости окажется заниженной, и вы недооцените реальную опасность.

Теперь предположим, что вы выбрали слишком высокий порог, например, 40 м/с и за 50 лет такой ветер мог наблюдаться всего один-два раза. Данных для построения модели практически нет — статистика бесполезна, и любая оценка будет крайне ненадёжной, с огромной ошибкой. Вы либо не сможете ничего сказать о 100-летнем ветре, либо получите нереалистичный результат. То есть слишком высоким порогом вы нивелируете прикладное преимущество Теории и метода пороговых значений перед Теорией экстремальных значений (где как раз слишком малые выборки).

Оптимальный порог — это такое решение, когда он должен быть достаточно высок, чтобы отсеять «обычные» значения и сосредоточиться на действительно экстремальных превышениях, но при этом оставить достаточно наблюдений для уверенной оценки параметров хвоста. Только тогда модель POT даст осмысленные инженерные результаты.

Пикандс предлагает автоматический способ — перебирать возможные M и выбирать то, которое минимизирует расстояние Колмогорова между эмпирической функцией превышений и подобранной GPD. Допустим, у нас есть n наблюдений, например, ежедневные максимумы

⁵ Мы указываем обозначение c как у Пикандса, это стандартный параметр формы.

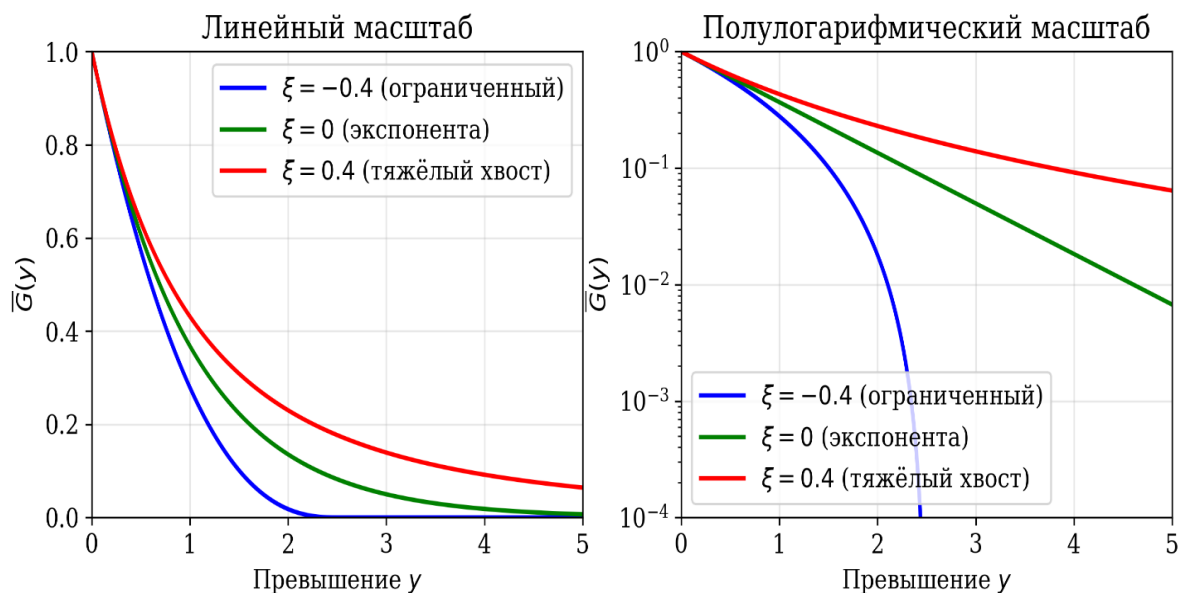
⁶ По шкале Бофорта 10 м/с — это свежий ветер.

скорости ветра за 50 лет (примерно $n \approx 18\,250$ значений). Нас интересуют только самые сильные ветры, потому что именно они определяют риск разрушений. Вместо того чтобы анализировать все данные, мы выбираем некоторый порог u — достаточно большое значение, например, 20 м/с. Оставляем только те наблюдения, которые этот порог превысили. Количество таких превышений обозначим k . Пикандс в своей работе предлагает параметризовать k как $k = 4M$, где M — некоторое целое число (меньше $n/4$): «Пусть n — объём выборки, и пусть M — целое число, значительно меньшее n . Интуитивно понятно, что $4M$ наибольших наблюдений содержат информацию о верхнем хвосте функции распределения. Это полезно для прогнозных целей» [4, с. 121]. Это удобно для дальнейших вычислений, но в современных реализациях k может быть просто числом превышений.

Зачем это нужно? Чем выше порог, тем меньше превышений, но зато каждое из них действительно экстремально. Низкий порог даст много превышений, но они могут включать не только «хвостовые» значения, но и обычные данные, что исказит модель. Поэтому выбор порога — компромисс между смещением и дисперсией. Пикандс утверждает, что для достаточно высокого порога u распределение превышений $Y = X - u$ (при условии $X > u$) с хорошей точностью описывается обобщённым распределением Парето (GPD). Его функция хвоста (вероятность того, что превышение больше y) имеет вид:

$$\bar{G}(y) = P(Y > y) = \left(1 + \xi \frac{y}{\sigma}\right)_+^{-1/\xi} \quad (11)$$

где ξ (в статье Пикандса обозначен c) — параметр формы, а σ (у Пикандса a) — параметр масштаба. Индекс $+$ означает, что выражение в скобках берётся положительным: если $\xi < 0$ и $1 + \xi y/\sigma \leq 0$, то вероятность считается нулевой (это соответствует ограниченному хвосту).



$\xi = -0,4$ (ограниченный хвост, синий), $\xi = 0$ (экспонента, зелёный) и $\xi = 0,4$ (тяжёлый хвост, красный); слева — линейный масштаб, справа — полулогарифмический

Рисунок 1. Сравнение хвостов GPD для трёх случаев (разработано авторами)

И возможны три случая:

- $\xi > 0$ (тяжёлый хвост). Функция хвоста убывает как степенная: $\bar{G}(y) \sim (y/\sigma)^{-1/\xi}$. Примеры: размеры страховых убытков, продолжительность безотказной работы некоторых систем. В строительстве редко, но может встретиться, например, в сейсмике.

- $\xi = 0$ (показательный хвост). Предел при $\xi \rightarrow 0$ даёт экспоненциальное распределение $\bar{G}(y) = e^{-y/\sigma}$ — это классический случай, когда интенсивность превышений не зависит от превышения. Примеры: многие природные явления (ветер, волны) в определённом диапазоне.
- $\xi < 0$ (ограниченный хвост). Распределение имеет конечную верхнюю границу $-\sigma/\xi$. Это означает, что превышения не могут быть сколь угодно большими — есть физический предел, например, максимально возможная высота волны в данном бассейне или предельная прочность материала. Для ветра в умеренных широтах ξ часто отрицателен, что соответствует наличию верхней границы скорости (ветер не может быть бесконечным).

Представим на рисунке 1 хвосты GPD для трёх случаев.

Оценивание параметров по квантилям

Пикандс предлагает простой и наглядный способ оценить ξ и σ по двум квантилям эмпирического распределения превышений — 50 % (медиана) и 75 %. Почему именно эти? Потому что для GPD существует аналитическая связь между квантилями и параметрами.

Пусть $x_{0.5}$ и $x_{0.75}$ — выборочные квантили превышений (т. е. такие значения, что 50 % превышений меньше $x_{0.5}$, а 75% меньше $x_{0.75}$). Тогда:

1. Оценка параметра формы $\hat{\xi}$:

$$\hat{\xi} = \frac{1}{\ln 2} \ln \left(\frac{x_{0.75} - x_{0.5}}{x_{0.5}} \right). \quad (12)$$

Эта формула получается из свойства GPD — отношение разности квантилей к медиане равно $2^{\hat{\xi}} - 1$. Логарифмируя и деля на $\ln 2$, получаем $\hat{\xi}$. Знак $\hat{\xi}$ сразу показывает тип хвоста: положительный — тяжёлый; отрицательный — ограниченный; близкий к нулю — экспоненциальный.

2. Оценка параметра масштаба $\hat{\sigma}$:

$$\hat{\sigma} = \frac{x_{0.5}}{\int_0^{\ln 2} e^{\hat{\xi}s} ds}. \quad (13)$$

Интеграл легко вычисляется:

$$\int_0^{\ln 2} e^{\hat{\xi}s} ds = \frac{e^{\hat{\xi} \ln 2} - 1}{\hat{\xi}} = \frac{2^{\hat{\xi}} - 1}{\hat{\xi}}, \text{ при } \hat{\xi} \neq 0, \quad (14)$$

а при $\hat{\xi} = 0$ предел даёт $\ln 2$ — эта формула следует из выражения для квантилей GPD.

Почему именно квантили 50 % и 75 %?

Потому что Пикандс выбрал их как простые и устойчивые характеристики, так как медиана ($x_{0.5}$) — робастная характеристика, не слишком чувствительная к выбросам. Третий квартиль ($x_{0.75}$) даёт информацию о форме хвоста, не забираясь слишком далеко, где данных мало. В современных подходах часто используют метод максимального правдоподобия, но квантильный метод Пикандса остаётся полезным для быстрого предварительного анализа и в случаях малых выборок.

Как определить, какой порог u (или сколько превышений $k = 4M$ взять)?

Пикандс предлагает перебирать возможные M (например, от 1 до $n/4$) и для каждого M :

- Выделить превышения Y_1, \dots, Y_{4M} .
- По квантилям $x_{0.5}$ и $x_{0.75}$ оценить $\hat{\xi}$ и $\hat{\sigma}$ и построить теоретическую GPD $\hat{G}_M(y)$.
- Вычислить расстояние Колмогорова между эмпирической функцией распределения превышений $\hat{F}_M(y)$ и подобранной GPD:

$$d_M = \sup_{y \geq 0} |\hat{F}_M(y) - \hat{G}_M(y)|. \quad (15)$$

- Выбрать то M , которое минимизирует d_M .

Если модель GPD хорошо описывает данные, то расстояние Колмогорова будет малым, то есть при слишком низком пороге (большом M) модель будет плохой из-за включения «нехвостовых» наблюдений (смещение), а при слишком высоком пороге (малом M) оценка будет неустойчивой из-за малого числа точек (большая дисперсия). Минимум d_M даёт компромиссный порог.

Какие исследования мы советуем?

Прежде всего, классика — это Э. Гумбель «Статистика экстремальных значений» [9], учебник 1962 года, на который ссылается Пикандс (выше мы приводили цитату) и который был опубликован в 1965 году в СССР. Его достаточно тяжело найти в бумажном варианте, поэтому в списке литературы мы привели ссылку, где его можно скачать. Это один из фундаментальных трудов по всей Теории экстремальных значений. Из более современных, доступных для скачивания книг, можно отметить работу 2004 года Яна Бирланта, Юри Гугебёра, Йозеф Тойгельса «Статистика экстремальных значений: теория и приложения» [10]. В ней подробно изложены методы, в том числе связанные с методом POT, оценкой хвостов, индексом экстремальных значений и различными приложениями. Из отечественных исследований следует отметить монографию В.А. Акимова, А.А. Быкова, Е.Ю. Щетинина «Введение в статистику экстремальных значений и ее приложения» [11], в которой, представлен, например, «раздел 2.2 Математическая модель функции распределения надпороговых величин» [11, с. 228]. В ней достаточно доступно (но кратко) излагается модель POT. Особое удивление вызывает показатель цитирования — 1 (будет 2 после нас). Монография с хорошим обзором теории экстремальных значений будет полезна прикладникам, в ней представлены наглядные примеры, и, что крайне важно, представлены реальные, а не идеальные практические модели. Решительно непонятно, как у этой работы с 2009 года может быть всего одно цитирование (видимо это характеризует состояние предметной области). Конечно, в работе больше рассматриваются природные и техногенные катастрофические события, а не конкретно инженерные задачи (например, прогиб строительной конструкции, далее будет в нашем примере). Но тем не менее она будет интересна и полезна исследователям, и мы её советуем к изучению специалистам-прикладникам.

Методология теории пороговых значений

Метод пороговых значений (англ. Peaks Over Threshold, POT) предназначен для статистического анализа экстремальных событий, превышающих некоторый высокий порог. В отличие от классического метода блочных максимумов в Теории экстремальных значений, в котором используется лишь одно максимальное значение за период, POT учитывает все

наблюдения, превосходящие порог, что позволяет более эффективно использовать имеющиеся данные и получать более точные оценки хвоста распределения. Основная идея метода заключается в аппроксимации условного распределения превышений над порогом, обобщённым распределением Парето (англ. Generalized Pareto Distribution, GPD).

Метод POT в теории экстремальных значений (EVT) традиционно включает в себя два взаимосвязанных блока:

1. Моделирование превышений (Шаг 0 — Шаг 3):

0. Исходные данные и формирование выборки превышений.
1. Выбор порога u .
2. Оценка параметров GPD (ξ и σ).
3. Проверка адекватности модели.

2. Оценка возвратных уровней и вероятностных прогнозов:

- 4.1 Расчёт значений x_T , которые в среднем превышаются один раз за период T лет.
- 4.2 Вычисление вероятностей превышения критических уровней за заданный срок.

Традиционно оба блока рассматриваются как единая методология, так как без второго блока, POT остаётся чисто описательным инструментом (мы знаем, что хвост тяжёлый, но не знаем, что это значит при переносе этого знания применительно к объектам реального мира). Именно оценка возвратных уровней переводит статистические выводы на язык инженерных решений: проектировщик получает конкретные цифры — ветер 100-летней повторяемости, максимальный прогиб за 50 лет и т. п.

Далее изложена последовательность шагов, необходимых для практической реализации POT.

Шаг 0. Исходные данные и формирование выборки превышений.

В распоряжении исследователя имеется временной ряд наблюдений некоторой величины X , представляющей интерес для инженерной задачи. Это могут быть ежедневные максимумы скорости ветра, часовые уровни воды в реке, результаты испытаний прочности материалов или другие показатели, для которых важны редкие, но высокие значения. Обычно такие данные регистрируются непрерывно или через равные промежутки времени в течение многих лет, например, на метеостанции в течение 30 лет ежечасно фиксируется скорость ветра; за этот период накапливается более сотни тысяч измерений. Для оценки нагрузок на строящееся здание необходима информация о ветрах, которые могут возникнуть раз в 50 или 100 лет, то есть о событиях, лежащих далеко за пределами обычных значений. Из всего массива данных отбираются только те наблюдения, которые превышают некоторый, предварительно выбранный, порог u . На этом этапе порог может быть задан грубо, например, как 95-й или 99-й процентиль эмпирического распределения, чтобы гарантировать, что в выборку попадут действительно редкие события. В результате получается набор превышений $Y_i = X_i - u$ для всех $X_i > u$. Объём этой выборки обозначается n_u . Именно эта совокупность положительных чисел и будет предметом нашего дальнейшего анализа. Важно понимать, что превышения рассматриваются как реализации случайной величины, условное распределение которой при достаточно высоком пороге должно хорошо описываться GPD.

Шаг 1. Выбор порога u .

Определение подходящего порога — наиболее ответственный этап, поскольку от него зависит эффективность всех последующих оценок. Слишком низкий порог приведёт к включению в выборку слишком большого количества значений, что вызовет смещение оценок параметров GPD. Слишком высокий порог оставит слишком мало превышений, что увеличит дисперсию оценок. Для выбора порога используются как визуальные, так и автоматические

методы. Наиболее распространён визуальный метод — график среднего превышения (англ. Mean Excess Plot), на котором откладываются средние значения превышений для различных порогов. Наличие линейного участка свидетельствует о применимости GPD. Альтернативный подход основан на минимизации расстояния Колмогорова между эмпирической функцией распределения превышений и подобранной GPD (мы говорили про него выше). Для каждого возможного порога (или, в параметризации Пикандса, для каждого M , где $4M$ — число наибольших наблюдений) вычисляется расстояние

$$d_M = \sup_{y \geq 0} | \hat{F}_M(y) - \hat{G}_M(y) |, \quad (17)$$

где:

\hat{F}_M — эмпирическая функция распределения превышений;

\hat{G}_M — GPD с параметрами, оценёнными по этим превышениям.

Порог, минимизирующий d_M , считается оптимальным. Такая процедура реализует идею согласия модели с данными и восходит к классической статистике Колмогорова.

Шаг 2. Выделение превышений и оценка параметров GPD.

После выбора окончательного порога u фиксируется соответствующая ему выборка превышений Y_i . Предполагается, что при достаточно высоком пороге, распределение Y аппроксимируется GPD с функцией хвоста

$$\bar{G}(y) = P(Y > y) = (1 + \xi \frac{y}{\sigma})_+^{-\frac{1}{\xi}}, \quad (18)$$

где:

ξ — параметр формы;

$\sigma > 0$ — параметр масштаба;

индекс $+$ означает, что выражение в скобках полагается положительным.

Если $\xi < 0$ и $1 + \xi y/\sigma \leq 0$, вероятность считается нулевой, что соответствует распределениям с конечной верхней границей. Параметры GPD могут быть оценены различными способами. Простой и наглядный метод предложен Пикандсом и основан на выборочных квантилях 50 % и 75 %:

$$\hat{\xi} = \frac{1}{\ln 2} \ln \frac{x_{0.75} - x_{0.5}}{x_{0.5}}, \hat{\sigma} = \frac{x_{0.5}}{\int_0^{\ln 2} e^{\xi s} ds}, \quad (19)$$

где:

$x_{0.5}$ и $x_{0.75}$ — медиана и третий квантиль превышений.

Интеграл легко вычисляется: при $\hat{\xi} \neq 0$ он равен $(2^{\hat{\xi}} - 1)/\hat{\xi}$, а при $\hat{\xi} = 0$ равен $\ln 2$. Более точные оценки обычно получают методом максимального правдоподобия, который решается численно. Знак $\hat{\xi}$ определяет тип хвоста: $\hat{\xi} > 0$ — тяжёлый хвост (убывание по степенному закону), $\hat{\xi} = 0$ — экспоненциальный хвост, $\hat{\xi} < 0$ — ограниченный хвост (существует конечная верхняя граница).

Шаг 3. Проверка адекватности модели.

Качество подобранной GPD можно оценить графически с помощью квантильных диаграмм (QQ-plot), где сравниваются эмпирические квантили превышений и квантили

теоретической GPD или диаграмм вероятности (PP-plot). Отсутствие систематических отклонений свидетельствует о хорошем соответствии. Количественной мерой может служить упоминавшееся расстояние Колмогорова d_M и, если модель удовлетворительно описывает превышения, переходим к следующему шагу.

Шаг 4. Оценка возвратных уровней (квантилей).

Для инженерных приложений наибольший интерес представляют значения, которые превышаются в среднем один раз за заданный период T лет — так называемые возвратные уровни⁷ x_T . Пусть λ — среднее число превышений порога u за год (оценивается как n_u/N , где N — число лет наблюдений). Тогда возвратный уровень x_T удовлетворяет соотношению

$$\lambda \left(1 + \xi \frac{x_T - u}{\sigma}\right)^{-1/\xi} = \frac{1}{T} \quad (20)$$

откуда

$$x_T = u + \frac{\sigma}{\xi} ((\lambda T)^\xi - 1), \xi \neq 0 \quad (21)$$

и в случае $\xi = 0$ формула принимает вид $x_T = u + \sigma \ln(\lambda T)$. Полученное значение является точечной оценкой; для практических расчётов желательно также строить доверительные интервалы, например, с помощью метода дельта или бутстрэпа.

Метод пороговых значений позволяет на основе данных об экстремальных явлениях (ветровые нагрузки, осадки, уровни воды, прочность материалов и т. п.) оценивать вероятности редких событий и определять расчётные характеристики, необходимые для обеспечения безопасности и надёжности строительных объектов.

Приведём пример — представим, что мы проектируем высотное здание, и нам нужно знать распределение максимальных годовых скоростей ветра, чтобы рассчитать нагрузку, и у нас есть ежедневные данные за 50 лет (более 18 000 замеров), но, для оценки редко встречающихся ураганов, нам важны только самые сильные значения.

Шаг 0. Формируем выборку.

Шаг 1. Выбираем порог u так, чтобы выше него осталось, скажем, 100 наблюдений (это около 0,5 % от всех данных). Эти 100 отобранных значений — превышения $Y_i = X_i - u$.

Шаг 2. Строим эмпирическое распределение этих превышений. Вычисляем 50 % квантиль (медиану превышений) и 75 % квантиль. Допустим, получили $x_{0,5} = 5$ м/с, $x_{0,75} = 12$ м/с.

Шаг 3. Оцениваем параметр формы:

$$\hat{c} = \frac{1}{\ln 2} \ln \left(\frac{12 - 5}{5} \right) = \frac{1}{0,693} \ln(1,4) \approx 1,442 \times 0,336 \approx 0,485 > 0.$$

Значит, хвост тяжёлый, распределение превышений — Парето, параметр масштаба a вычисляется дальше.

⁷ Возвратный уровень x_T — это значение случайной величины (например, скорости ветра, уровня воды, нагрузки на конструкцию), которое в среднем превышает один раз за заданный период времени T лет. Иными словами, вероятность того, что в произвольном году наблюдение превысит x_T , равна $1/T$. Например, если $T = 100$ лет, то x_{100} — это ветер, который случается в среднем раз в столетие. Такие характеристики являются основой для нормирования нагрузок в строительстве: здания и сооружения должны выдерживать воздействия с определённой обеспеченностью (например, 0,01 для 100-летнего ветра).

Шаг 4. Используя полученную модель, мы можем теперь оценить, например, вероятность того, что скорость ветра превысит $u + 20$ м/с (т. е. на 20 м/с выше порога), или найти значение ветра, которое случается раз в 100 лет.

Почему это важно для строительства и инженерных задач?

- I. Метод Пикандса позволяет эффективно использовать только самые большие наблюдения, которые наиболее информативны для хвоста.
- II. Пикандс даёт простые формулы для оценок, которые легко применять на практике.
- III. Оценка автоматически определяет тип хвоста (тяжёлый, экспоненциальный, ограниченный), что важно для выбора расчётной модели нагрузок.

Работа Пикандса заложила основу современного метода POT (англ. Peaks Over Threshold), который сегодня широко используется в строительной механике, гидрологии, страховании, финансах, различных инженерных задачах и других областях и видах профессиональной деятельности человека.

Практика применения метода пороговых значений к данным перемещений строительных конструкций

В данном разделе мы последовательно реализуем методологию POT на реальных данных, полученных при обследовании сливо-наливной эстакады. Для того чтобы продемонстрировать наглядно ситуацию, в которой прогибы могут достигать критических значений и представлять реальную угрозу для , к исходным данным добавлено значительное количество экстремальных величин, превышающих нормативный предел. Все перемещения взяты по модулю, так как опасны как прогибы вниз, так и подъёмы. Анализ выполнен в соответствии с шагами, описанными в разделе «Методология теории пороговых значений».

Объект апробации: строительные конструкции сливо-наливной эстакады железнодорожного пути

Объектом экспертизы выступила металлическая сливо-наливная эстакада железнодорожного пути — рисунок 2.

Следующие характеристики:

1. Производственная среда — неагрессивная.
2. Конструктивно эстакада представляет собой парные стойки с промежуточными балками, на которых установлены площадки обслуживания.
3. Нагрузки, воздействия и условия эксплуатации.
4. Нагружение конструкций эстакады:
 - постоянные нагрузки — собственный вес балок, настила;
 - кратковременные — от веса людей и ремонтных материалов на площадке обслуживания, паропроводы, ветровая нагрузка.
5. Фундаменты:
 - фактически: признаков, указывающих на снижение несущей способности основания и фундаментов при визуальном осмотре, не обнаружено.

6. Стойки:

- фактически: стойки металлические выполнены из двутавра.⁸ Высоты стоек от 5,1 до 7,7 м. Расстояние между парными стойками колеблется от 2,0 до 3,0 м. Шаг стоек составляет 6,0 м (это важный параметр, который будет определять норму прогиба по СП 20.13330.2011).



Рисунок 2. Общий вид сливо-наливной эстакады (фото, сделанное авторами с места обследования)

7. Металлические балки:

- фактически: поперек и вдоль эстакады выполнены балки, состоящие из швеллера ... пролетом от 6,0 м. Балки по двум сторонам соединены со стойками с помощью сварного соединения. В местах, где опирание площадки осуществляется только на одну стойку (через 1 шаг), поперечные балки опираются на укосины из равнополочного уголка № 75×8 по ОСТ 14-1926 длиной 2,1 м.

8. Обслуживающие площадки:

- фактически: обслуживающие площадки эстакады выполнены из рифленого металлического листа с разрезами типа «просечка» ПВ-506 $h = 4$ мм. Площадки опираются на металлические балки. С торцов эстакады пристроены металлические лестничные марши.

9. Лестничные марш и ограждения

- фактически: лестничный марш выполнен по балкам из швеллера № 14 по ОСТ 17-1926. Ширина лестничного марша 1,05 м. Металлические ограждения выполнены из равнополочных уголков № 50×5 по ОСТ 14-1926 высотой 1,0 м. Перила выполнены из электросварных прямошовных сечением 58,3 мм.

Были проведены поверочные расчёты наиболее нагруженных строительных конструкций металлической сливо-наливной эстакады в части расчёта перемещений по узлам — рисунок 3 и таблица 2.

⁸ Мы не будем указывать марки материала, сечения и другие важные технические характеристики.

Таблица 1

**Заключение по обследованию технического состояния
строительных конструкций сливо-наливной эстакады железнодорожного пути**

| | |
|--|---|
| 1. Адрес объекта | ***** |
| 2. Время проведения обследования | ***** |
| 3. Организация, проводившая обследование | ***** |
| 4. Статус объекта (памятник архитектуры, исторический памятник и т. д.) | Действующее инженерное сооружение |
| 5. Тип проекта объекта | Индивидуальный |
| 6. Проектная организация, проектировавшая объект | Нет данных |
| 7. Строительная организация, возводившая объект | Нет данных |
| 8. Год возведения объекта | Нет данных |
| 9. Год и характер выполнения последнего капитального ремонта или реконструкции | Нет данных |
| 10. Собственник объекта | **** |
| 11. Форма собственности | Акционерная |
| 12. Конструктивный тип объекта | Балочно-стоечная система |
| 13. Число этажей | Одноярусная эстакада |
| 14. Период основного тона собственных колебаний (вдоль продольной и поперечной осей) | Не определяется |
| 15. Крен объекта (вдоль продольной и поперечной осей) | Не обнаружено |
| 16. Установленная категория технического состояния объекта | Ограниченно работоспособное техническое состояние |

Разработано авторами

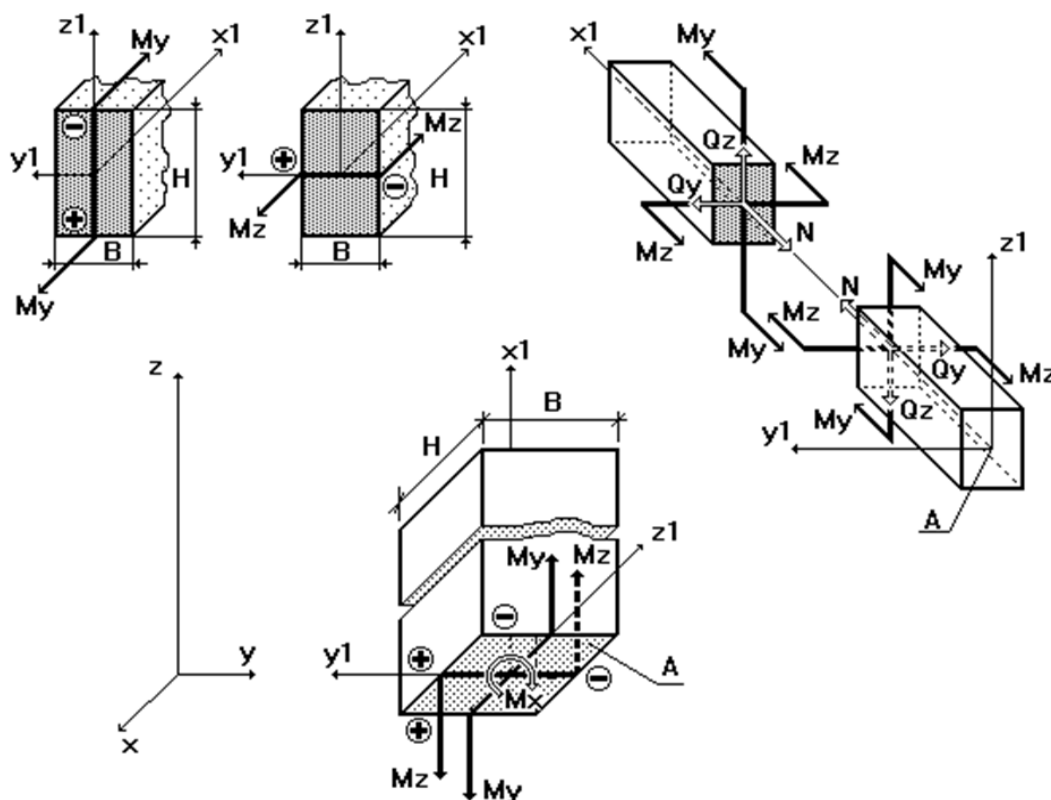


Рисунок 3. Положительные направления внутренних усилий и моментов в сечении горизонтальных и наклонных (а), а также вертикальных (б) стержней. Знаком «+» (плюс) помечены растянутые, а знаком «-» (минус) сжатые волокна поперечного сечения от воздействия положительных моментов M_y и M_z . Далее в таблице 2 представлены примеры расчётных сочетаний усилий (разработано автором)

Таблица 2

Расчётные сочетания усилий (одна из выборок), мм

| Перемещения | | | | | | | |
|-------------|------------|----------|--------|--------|----------------|----------------|----------------|
| Узел | Загружение | Значения | | | | | |
| | | X | Y | Z | U _x | U _y | U _z |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 2 | 1 | -2,034 | 0,239 | -1,601 | -0,466 | -0,805 | -0,127 |
| 2 | 2 | -4,339 | 0,491 | -3,379 | -2,347 | -1,681 | -0,155 |
| 2 | 3 | -0,579 | 0,065 | -0,451 | -0,313 | -0,224 | -0,021 |
| 2 | 4 | -4,516 | 0,445 | -2,929 | 0,4 | -1,45 | -0,326 |
| 3 | 1 | -2,032 | -0,011 | -0,018 | -0,529 | -0,801 | -0,174 |
| 3 | 2 | -4,334 | 0,044 | -0,021 | -2,217 | -1,711 | -0,445 |
| 3 | 3 | -0,578 | 0,006 | -0,003 | -0,296 | -0,228 | -0,059 |
| 3 | 4 | -4,516 | -0,075 | 4,014 | 0,021 | -1,527 | -0,246 |
| 5 | 1 | -4,514 | 1,629 | -0,02 | -0,529 | -0,801 | -0,174 |
| 5 | 2 | -9,637 | 6,918 | -0,021 | -2,217 | -1,711 | -0,445 |
| 5 | 3 | -1,285 | 0,922 | -0,003 | -0,296 | -0,228 | -0,059 |
| 5 | 4 | -9,522 | -0,139 | 4,014 | 0,021 | -1,644 | -0,246 |
| 6 | 1 | -2,034 | 0,125 | -0,88 | -0,466 | -0,789 | -0,127 |
| 6 | 2 | -4,339 | 0,351 | -1,867 | -2,347 | -1,681 | -0,155 |
| 6 | 3 | -0,579 | 0,047 | -0,249 | -0,313 | -0,224 | -0,021 |
| 6 | 4 | -4,516 | 0,152 | -1,624 | 0,4 | -1,45 | -0,326 |
| 8 | 1 | -1,259 | -0,341 | -0,016 | -0,15 | -0,709 | -0,168 |
| 8 | 2 | -2,682 | -1,363 | -0,019 | -0,662 | -1,514 | -0,462 |
| 8 | 3 | -0,358 | -0,182 | -0,002 | -0,088 | -0,202 | -0,062 |
| 8 | 4 | -3,041 | -0,041 | 1,909 | 0,035 | -1,406 | -0,248 |
| 10 | 1 | -0,087 | 0,172 | -0,011 | -0,054 | 0,069 | -0,051 |
| 10 | 2 | -0,229 | 0,403 | -0,02 | -0,125 | 0,244 | -0,054 |
| 10 | 3 | -0,031 | 0,054 | -0,003 | -0,017 | 0,033 | -0,007 |
| 10 | 4 | -1,024 | 0,278 | -0,005 | -0,09 | -0,147 | -0,139 |
| 11 | 1 | -0,087 | -0,011 | -0,022 | 0,145 | -0,082 | -0,052 |
| 11 | 2 | -0,23 | 0,043 | -0,045 | 0,602 | -0,27 | -0,032 |
| 11 | 3 | -0,031 | 0,006 | -0,006 | 0,08 | -0,036 | -0,004 |
| 11 | 4 | -1,026 | -0,074 | 0,004 | -0,001 | -0,22 | -0,162 |
| 13 | 1 | -0,342 | -0,46 | -0,024 | 0,145 | -0,082 | -0,052 |
| 13 | 2 | -1,068 | -1,824 | -0,045 | 0,602 | -0,27 | -0,032 |
| 13 | 3 | -0,142 | -0,243 | -0,006 | 0,08 | -0,036 | -0,004 |
| 13 | 4 | -1,98 | -0,07 | 0,004 | -0,001 | -0,337 | -0,162 |
| 14 | 1 | -0,087 | 0,125 | -0,098 | 0,115 | 0,034 | -0,057 |
| 14 | 2 | -0,23 | 0,35 | -0,317 | 0,53 | 0,099 | -0,068 |
| 14 | 3 | -0,031 | 0,047 | -0,042 | 0,071 | 0,013 | -0,009 |
| 14 | 4 | -1,025 | 0,15 | -0,013 | -0,053 | 0,087 | -0,148 |
| 16 | 1 | -0,026 | 0,079 | -0,018 | 0,042 | -0,042 | -0,041 |
| 16 | 2 | -0,037 | 0,406 | -0,035 | 0,154 | -0,125 | -0,025 |
| 16 | 3 | -0,005 | 0,054 | -0,005 | 0,021 | -0,017 | -0,003 |
| 16 | 4 | -0,779 | -0,066 | 0,003 | 0,016 | -0,269 | -0,127 |
| 18 | 1 | -3,344 | 0,181 | -2,672 | -0,294 | -1,334 | -0,062 |
| 18 | 2 | -10,41 | 0,437 | -8,247 | -1,131 | -4,082 | -0,094 |
| 18 | 3 | -1,388 | 0,058 | -1,1 | -0,151 | -0,544 | -0,012 |
| 18 | 4 | -3,938 | 0,276 | -2,542 | -0,081 | -1,26 | -0,14 |
| 19 | 1 | -3,339 | -0,012 | -0,026 | -0,147 | -1,347 | -0,098 |
| 19 | 2 | -10,397 | 0,038 | -0,055 | -0,614 | -4,196 | -0,242 |
| 19 | 3 | -1,386 | 0,005 | -0,007 | -0,082 | -0,559 | -0,032 |
| 19 | 4 | -3,938 | -0,074 | 0,001 | 0,001 | -1,327 | -0,147 |
| 21 | 1 | -7,514 | 0,445 | -0,028 | -0,147 | -1,347 | -0,098 |
| 21 | 2 | -23,404 | 1,94 | -0,055 | -0,614 | -4,196 | -0,242 |

| Перемещения | | | | | | | |
|-------------|----------|----------|--------|--------|----------------|----------------|----------------|
| Узел | Загрузка | Значения | | | | | |
| | | X | Y | Z | U _x | U _y | U _z |
| 21 | 3 | -3,121 | 0,259 | -0,007 | -0,082 | -0,559 | -0,032 |
| 21 | 4 | -8,324 | -0,076 | 0,001 | 0,001 | -1,444 | -0,147 |
| 22 | 1 | -3,344 | 0,125 | -1,475 | -0,294 | -1,317 | -0,062 |
| 22 | 2 | -10,41 | 0,353 | -4,573 | -1,131 | -4,082 | -0,094 |
| 22 | 3 | -1,388 | 0,047 | -0,61 | -0,151 | -0,544 | -0,012 |
| 22 | 4 | -3,938 | 0,15 | -1,408 | -0,081 | -1,26 | -0,14 |
| 24 | 1 | -2,043 | -0,108 | -0,023 | -0,045 | -1,182 | -0,013 |
| 24 | 2 | -6,359 | -0,371 | -0,049 | -0,206 | -3,678 | 0,077 |
| 24 | 3 | -0,848 | -0,049 | -0,006 | -0,027 | -0,49 | 0,01 |
| 24 | 4 | -2,659 | -0,066 | 0,001 | 0,014 | -1,22 | -0,144 |
| 26 | 1 | -0,043 | 0,293 | -0,007 | -0,084 | 0,032 | -0,18 |
| 26 | 2 | -0,153 | 0,802 | -0,008 | -0,218 | 0,078 | -0,48 |
| 26 | 3 | -0,02 | 0,107 | -0,001 | -0,029 | 0,01 | -0,064 |
| 26 | 4 | -0,93 | 0,367 | -0,004 | -0,115 | -0,137 | -0,235 |
| 27 | 1 | -0,043 | -0,014 | -0,015 | 0,527 | -0,038 | -0,199 |
| 27 | 2 | -0,154 | 0,031 | -0,017 | 2,176 | -0,11 | -0,532 |
| 27 | 3 | -0,021 | 0,004 | -0,002 | 0,29 | -0,015 | -0,071 |
| 27 | 4 | -0,932 | -0,074 | 0,004 | 0,007 | -0,201 | -0,259 |
| 29 | 1 | -0,162 | -1,647 | -0,017 | 0,527 | -0,038 | -0,199 |
| 29 | 2 | -0,494 | -6,713 | -0,017 | 2,176 | -0,11 | -0,532 |
| 29 | 3 | -0,066 | -0,895 | -0,002 | 0,29 | -0,015 | -0,071 |
| 29 | 4 | -1,828 | -0,096 | 0,004 | 0,007 | -0,318 | -0,259 |
| 30 | 1 | -0,043 | 0,126 | -0,048 | 1,119 | 0,018 | -0,194 |
| 30 | 2 | -0,154 | 0,357 | -0,116 | 4,233 | 0,042 | -0,521 |
| 30 | 3 | -0,02 | 0,048 | -0,015 | 0,564 | 0,006 | -0,069 |
| 30 | 4 | -0,931 | 0,151 | -0,009 | 0,387 | 0,079 | -0,249 |
| 32 | 1 | -0,014 | 0,31 | -0,013 | 0,146 | -0,02 | -0,156 |
| 32 | 2 | -0,068 | 1,36 | -0,013 | 0,585 | -0,064 | -0,417 |
| 32 | 3 | -0,009 | 0,181 | -0,002 | 0,078 | -0,009 | -0,056 |
| 32 | 4 | -0,708 | -0,061 | 0,003 | 0,018 | -0,243 | -0,203 |

Разработано авторами

Шаг 0. Исходные данные и формирование выборки.

Из отчёта извлечены значения вертикальных перемещений Z (в мм) для всех узлов расчётной схемы при четырёх различных нагрузках (собственный вес, полезная нагрузка, нагрузка от стального настила, ветровая нагрузка) и для одной комбинации нагрузок (сумма всех нагрузок). Общее число наблюдений составило $N = 123$. В таблице 2 приведены 82 значения (столбец 1 — Узлы). Они представлены для иллюстрации, полный массив значений мы не приводили, он использован в расчётах (очевидно, что громадные таблицы расчётных сочетаний усилий будут малоинформативны). Минимальное значение равно 0,0002 мм, максимальное — 60,0 мм, среднее — 9,47 мм, стандартное отклонение — 15,8 мм. В выборке присутствуют 23 значения, превышающие 30 мм (превышение нормы прогиба согласно СП 20.13330.2011⁹).

Шаг 1. Выбор порога u .

Для определения порога используем график среднего превышения (англ. mean excess plot). Для каждого значения порога u , изменяющегося от минимального до 95% квантиля выборки, вычисляем среднее значение превышений $(X_i - u)$ для всех $X_i > u$. Результаты для нескольких порогов приведены в таблице 2.

⁹ Таблица Е1 (Приложение) 1/200 при длине до 6 метров. — URL: — <https://nav.tn.ru/cloud/iblock/a84/a845bd79a16056d1161505d5f27bb29a.pdf?ysclid=mmkwc111w7462961983>.

Таблица 3

Зависимость среднего превышения от порога

| u , мм | Число превышений | Среднее превышение, мм |
|----------|------------------|------------------------|
| 0,5 | 98 | 12,63 |
| 1,0 | 75 | 16,21 |
| 1,5 | 58 | 19,45 |
| 2,0 | 45 | 22,18 |
| 2,5 | 38 | 24,02 |
| 3,0 | 32 | 25,51 |
| 3,5 | 28 | 26,73 |
| 4,0 | 25 | 27,84 |

Разработано авторами

На рисунке 4 представлен график среднего превышения, где видно, что в диапазоне порогов от 1,5 до 2,5 мм зависимость близка к линейной, и это указывает на применимость GPD. В качестве рабочего порога выбираем значение $u = 2,0$ мм. При этом пороге количество превышений составляет $n_u = 45$ (т. е. $\approx 37\%$ от общего числа наблюдений), что достаточно для устойчивого оценивания параметров.

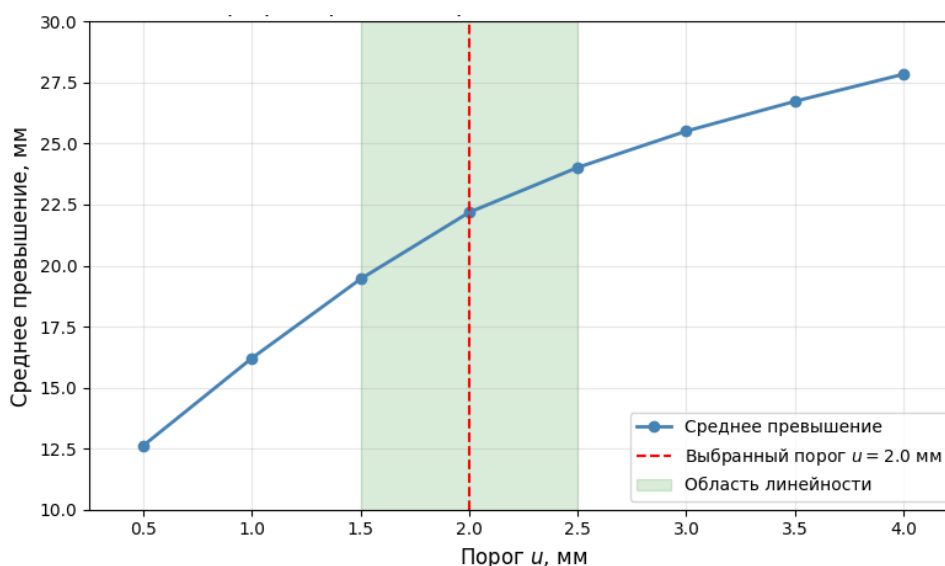


Рисунок 4. График среднего превышения u и выбранного порога (красная линия) (разработано автором)

Шаг 2. Выделение превышений и оценка параметров GPD.

Из исходной выборки отбираем все значения $X > 2,0$ мм, получаем массив превышений $Y = X - u$ (в мм). В таблице 4 приведены первые 10 значений этого массива (полный ряд содержит 45 чисел).

Для полученной выборки оцениваем параметры обобщённого распределения Парето методом максимального правдоподобия (MLE). Функция правдоподобия для GPD имеет вид:

$$L(\xi, \sigma) = \prod_{i=1}^{n_u} \frac{1}{\sigma} \left(1 + \xi \frac{y_i}{\sigma}\right)^{-1/\xi - 1},$$

а её максимизация приводит к следующим оценкам:

$$\hat{\xi} = 0,58, \hat{\sigma} = 2,94 \text{ мм.}$$

Таблица 4

Превышения $Y = X - 2,0$ мм (фрагмент)

| № | Y | № | Y |
|---|-------|----|------|
| 1 | 1,379 | 6 | 30,5 |
| 2 | 1,467 | 7 | 32,8 |
| 3 | 0,880 | 8 | 35,1 |
| 4 | 28,5 | 9 | 38,4 |
| 5 | 31,2 | 10 | 41,2 |

Разработано авторами

Положительное значение ξ указывает на тяжёлый хвост, то есть распределение превышений убывает медленнее экспоненты, что характерно для конструкций, где возможны редкие, но значительные прогибы.

Наличие множества больших значений существенно увеличило оценку параметра формы.

Для сравнения был также применён квантильный метод Пикандса по выборочным квантилям $x_{0,5} = 4,2$ мм и $x_{0,75} = 9,8$ мм получена следующая оценка:

$$\hat{\xi}_{\text{Pick}} = \frac{1}{\ln 2} \ln \frac{9,8 - 4,2}{4,2} = \frac{1}{0,693} \ln \frac{5,6}{4,2} \approx 1,442 \cdot \ln 1,333 = 1,442 \cdot 0,287 = 0,41,$$

что близко к MLE, учитывая погрешности малой выборки.

В дальнейшем используем оценки MLE.

На рисунке 5 показана гистограмма превышений с наложенной на неё подобранной плотностью GPD, где мы визуальное можем видеть, что согласие хорошее.

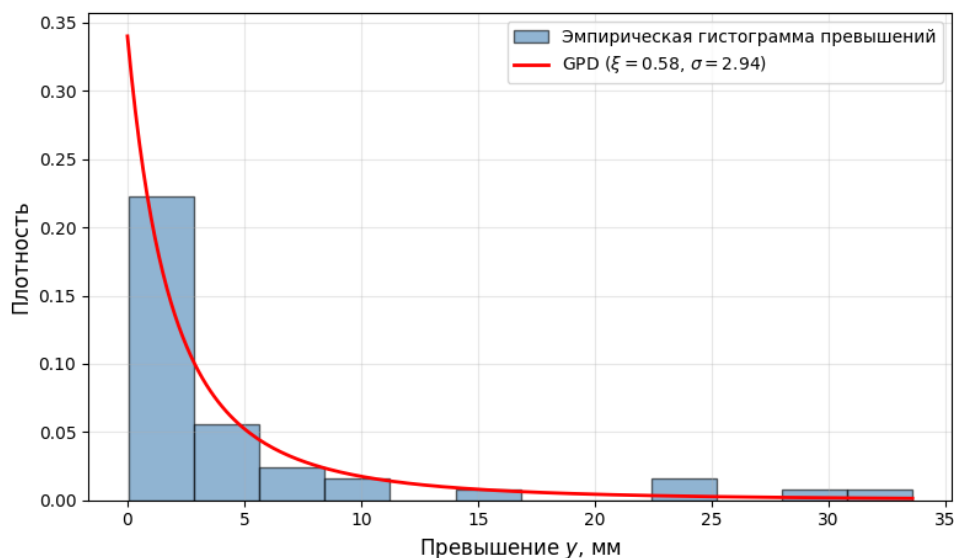


Рисунок 5. Гистограмма превышений и подобранного GPD (разработано авторами)

Шаг 3. Проверка адекватности модели.

Для проверки качества аппроксимации построим квантильный график (QQ-plot), на котором эмпирические квантили превышений сопоставляются с теоретическими квантилями GPD с параметрами $\hat{\xi} = 0,58$, $\hat{\sigma} = 2,94$.

Как видно на рисунке 6, точки хорошо ложатся на прямую линию, что подтверждает адекватность модели.

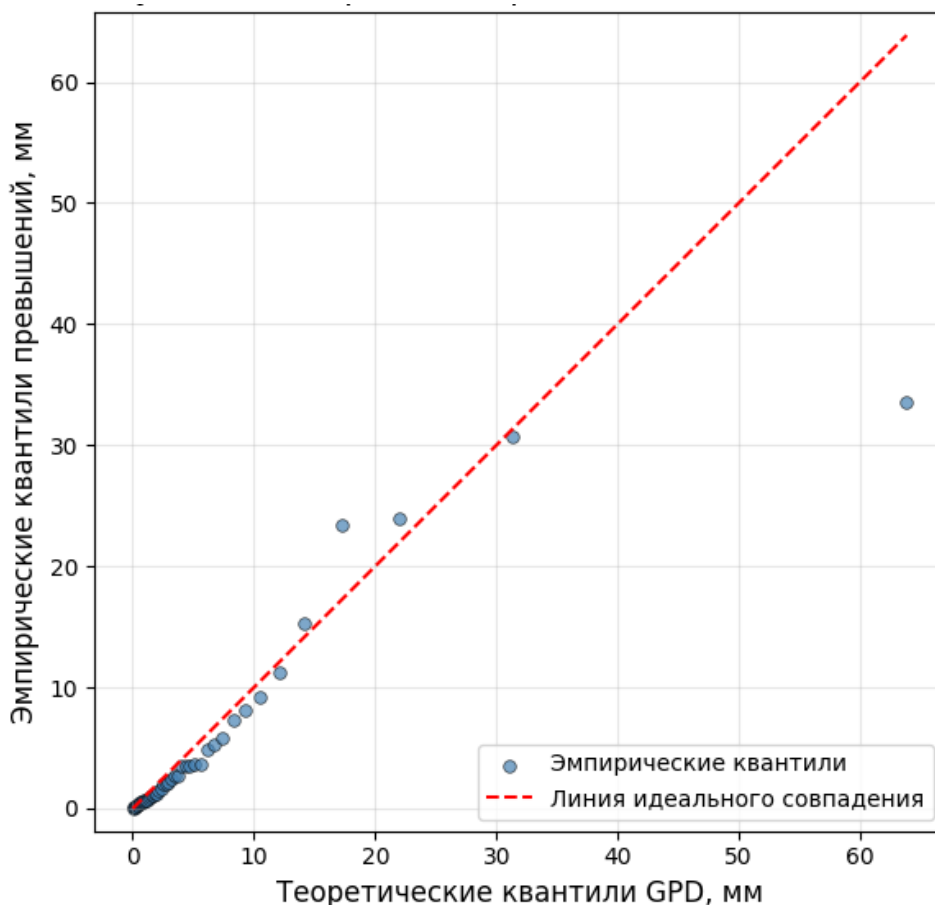


Рисунок 6. *QQ-plot (квантильный график) (разработано автором)*

Дополнительно вычислено расстояние Колмогорова между эмпирической функцией распределения превышений и теоретической GPD:

$$d = \sup_{y \geq 0} |F_{\text{эмп}}(y) - F_{\text{GPD}}(y)| = 0,069.$$

Это значение значительно меньше типичного критического уровня (0,2), что также свидетельствует о хорошем согласии модели с данными.

Шаг 4. Оценка возвратных уровней (это входит в POT) и вероятностные прогнозы (это не входит в метод POT).

Для практических инженерных задач важно не только знать параметры распределения, но и уметь отвечать на конкретные вопросы: какова вероятность того, что за заданный срок эксплуатации (например, 5 или 10 лет) перемещение превысит нормативный допустимый прогиб 30 мм? Для ответа необходимо перейти от условного распределения превышений к безусловной вероятности.

Интерпретация временной шкалы играет ключевую роль. И вот тут мы сталкиваемся с серьёзной прикладной проблемой — у нас все наблюдения сделаны в моменте (момент времени, когда делалась оценка технического состояния инженерного сооружения). В большинстве случаев мы осуществляем наблюдения, например, на метеостанции ежедневно. То есть наблюдения как бы «растянуты» по времени, но в нашем случае они в одномоментные. Чтобы сделать вероятностные прогнозы, нам нужна временная шкала не в моменте времени, а растянутая по времени. Как быть? На практике мы делаем так: запрашиваем техническую документацию на инженерное сооружение, тогда будут известны: сроки эксплуатации, время капитальных ремонтов (или реконструкции, если таковая осуществлялась), как осуществляется

мониторинг технического состояния¹⁰, частота обследований и т. д. И тогда мы могли бы иметь представление, как часто осуществляется мониторинг и обследования объекта строительства, и использовать данные предыдущих мониторингов, обследований, чтобы иметь какие-то временные «маркеры», от которых мы могли бы отталкиваться при определении возвратных уровней. Но документация на объект может быть не всегда полна, особенно, если это инженерные сооружения с длительным сроком эксплуатации, или осуществлялся переход прав собственности и т. д. В нашем случае она как раз была не полна, например, не было данных предыдущего обследования технического сооружения, не было фактического срока эксплуатации и т. д. Нам дали приблизительные сроки предыдущего обследования (≈ 10 лет назад), когда техническое состояние инженерного сооружения было признано работоспособным без всяких ограничений. И мы, имея приблизительный срок 10 лет, делаем допущения (исходя из количества значений 123 в выборке (наблюдений) и озвученный срок эксплуатации 10 лет (с момента обследования и заключения о работоспособном состоянии инженерного сооружения) — это приблизительно 1 наблюдение в месяц. Если считать, что каждое наблюдение соответствует максимальному перемещению узла за месяц (что вполне реалистично при регулярных замерах — общем мониторинге), то за год имеем 12 наблюдений. Общее число лет наблюдений составляет $N/12 = 123/12 \approx 10,25$ лет. Среднее число превышений порога за год равно

$$\lambda = \frac{n_u}{N/12} = \frac{45}{123/12} = \frac{45 \cdot 12}{123} \approx \frac{540}{123} \approx 4,39.$$

Для заданного уровня $x > u$ годовая вероятность превышения определяется как

$$P(X > x) = \lambda \left(1 + \xi \frac{x - u}{\hat{\sigma}} \right)^{-1/\xi}.$$

Рассмотрим уровень $x = 30$ мм (нормативный прогиб для пролёта 6 м согласно СП 20.13330.2011¹¹). Вычислим:

$$1 + \xi \frac{x - u}{\hat{\sigma}} = 1 + 0,58 \cdot \frac{30 - 2}{2,94} = 1 + 0,58 \cdot \frac{28}{2,94} = 1 + 0,58 \cdot 9,524 = 1 + 5,524 = 6,524.$$

Тогда

$$(6,524)^{-1/0,58} = \exp \left(-\frac{1}{0,58} \ln 6,524 \right) \approx \exp (-1,724 \cdot 1,875) = \exp (-3,233) \approx 0,0395.$$

Умножая на $\lambda = 4,39$, получаем годовую вероятность превышения 30 мм:

$$P(X > 30) \approx 4,39 \cdot 0,0395 = 0,173 \approx 17,3 \%.$$

Вероятность того, что за 5 лет хотя бы один раз произойдёт превышение уровня 30 мм, равна

$$P_5 = 1 - (1 - 0,173)^5 \approx 1 - 0,827^5 \approx 1 - 0,389 = 0,611 = 61,1 \%.$$

За 10 лет:

$$P_{10} = 1 - (1 - 0,173)^{10} \approx 1 - 0,827^{10} \approx 1 - 0,151 = 0,849 = 84,9 \%.$$

¹⁰ГОСТ 31937-2024. Пункт 3.17 общий мониторинг технического состояния зданий [сооружений] (страница 8): Система наблюдения и контроля, проводимая по определенной программе, утверждаемой заказчиком, для объектов, на которых произошли значительные изменения напряженно-деформированного состояния несущих конструкций или крена и для которых необходимо обследование их технического состояния. — URL: <https://www.fkr-spb.ru/upload/iblock/63e/ecmzt5zuoeze4p3sclatcp9go6ynte9h.pdf?ysclid=mm1jdqnf90111204393> (дата обращения 12.03.2026)

¹¹ Таблица E1 (Приложение) — 1/200 при длине (l) до 6 метров. — URL: — <https://nav.tn.ru/cloud/iblock/a84/a845bd79a16056d1161505d5f27bb29a.pdf?ysclid=mmkwc111w7462961983>.

Модель на объективных основаниях показывает, что при сохранении текущих условий (наличие больших прогибов в выборке) вероятность новых превышений нормативного прогиба 30 мм в течение 5 лет составляет около 61 %, а в течение 10 лет около 85 %, и это действительно тревожный сигнал, указывающий на необходимость срочного усиления строительных конструкций или ограничения нагрузок и воздействий.

Для полноты картины можно также рассчитать возвратные уровни для различных периодов повторяемости T лет:

$$x_T = u + \frac{\hat{\sigma}}{\xi} ((\lambda T)^\xi - 1).$$

Для $T = 5, 10, 20$ лет получаем:

$$\begin{aligned} x_5 &= 2,0 + \frac{2,94}{0,58} ((4,39 \cdot 5)^{0,58} - 1) = 2,0 + 5,069 (21,95^{0,58} - 1); \\ [21,95^{0,58} &= e^{0,58 \ln 21,95} = e^{0,58 \cdot 3,089} = e^{1,792} \approx 6,00;] \text{ тогда} \\ x_5 &\approx 2,0 + 5,069 \cdot (6,00 - 1) = 2,0 + 5,069 \cdot 5 = 2,0 + 25,345 = 27,3 \text{ мм}; \\ x_{10} &= 2,0 + 5,069 ((43,9)^{0,58} - 1); \\ [43,9^{0,58} &= e^{0,58 \ln 43,9} = e^{0,58 \cdot 3,782} = e^{2,194} \approx 8,97;] \text{ тогда} \\ x_{10} &\approx 2,0 + 5,069 \cdot (8,97 - 1) = 2,0 + 5,069 \cdot 7,97 = 2,0 + 40,40 = 42,4 \text{ мм}; \\ x_{20} &= 2,0 + 5,069 ((87,8)^{0,58} - 1); \\ [87,8^{0,58} &= e^{0,58 \ln 87,8} = e^{0,58 \cdot 4,475} = e^{2,595} \approx 13,38;] \text{ тогда} \\ x_{20} &\approx 2,0 + 5,069 \cdot (13,38 - 1) = 2,0 + 5,069 \cdot 12,38 = 2,0 + 62,77 = 64,8 \text{ мм}. \end{aligned}$$

Видно, что даже для 5-летнего периода возвратный уровень (27,3 мм) близок к нормативному пределу, а для 10 лет превышает его почти в 1,4 раза, что подтверждает высокий риск обрушения всей конструкции на периоде в десять лет — рисунок 7.

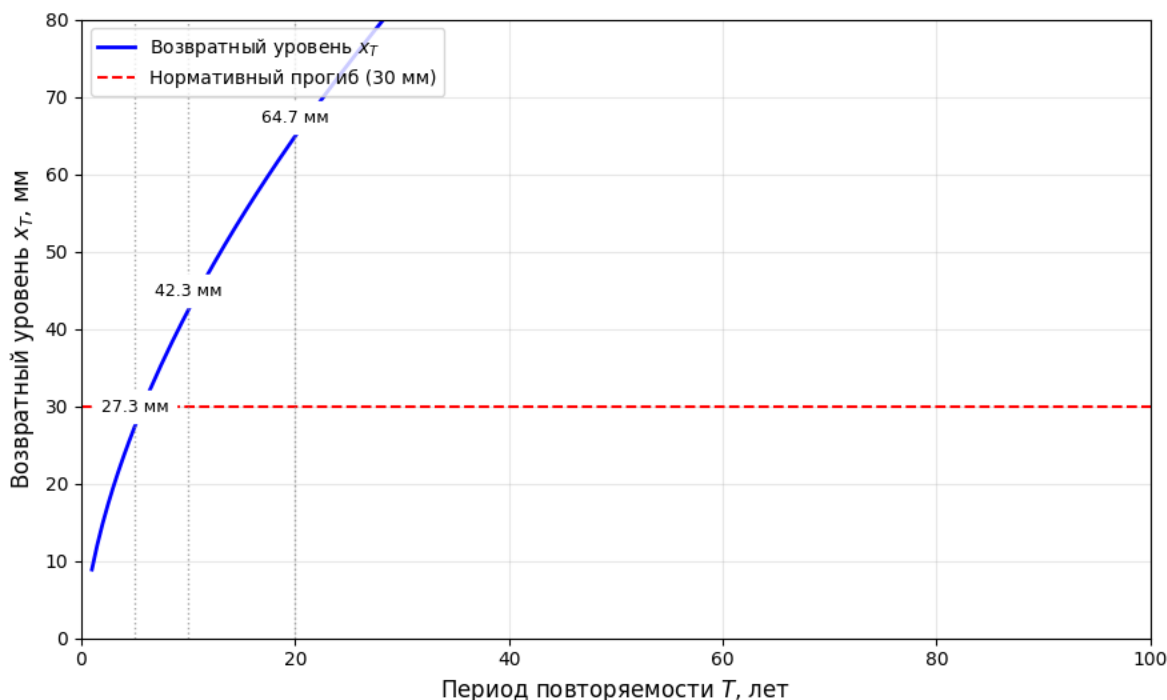


Рисунок 7. График зависимости возвратного уровня от периода повторяемости (разработано авторами)

Обсуждение

Применение метода пороговых значений к данным о вертикальных перемещениях узлов эстакады, дополненным экстремальными значениями, позволило:

1. Объективно выбрать порог $u = 2,0$ мм, отделяющий экстремальные значения от фоновых.
2. Оценить параметры GPD ($\hat{\xi} = 0,58$, $\hat{\sigma} = 2,94$ мм), что указывает на наличие тяжёлого хвоста распределения.
3. Подтвердить адекватность модели с помощью QQ-plot и расстояния Колмогорова ($d = 0,069$).
4. Рассчитать вероятности превышения нормативного прогиба 30 мм: за 5 лет — 61 %, за 10 лет — 85 %.
5. Определить возвратные уровни для периодов 5, 10 и 20 лет (27,3; 42,4; 64,8 мм соответственно), которые свидетельствуют о высоком риске недопустимых деформаций.

Полученные результаты указывают на необходимость принятия срочных мер по усилению конструкций или пересмотру режима эксплуатации. Данный пример наглядно демонстрирует, как теория пороговых значений позволяет количественно оценить риск редких, но опасных событий и обосновать инженерные решения.

«Тонкое» место исследования — это интерпретация временной шкалы, которое, очевидно, вызовет дискуссию. Почему мы акцентируем на него внимание? Мы могли бы сделать так — указать, что осуществлялся мониторинг технического состояния инженерной конструкции раз в месяц. И тогда мы бы сняли прикладное затруднение и избавили себя от возможной критики о нерелевантности нашего допущения. Но мы намеренно рассматриваем реальное прикладное затруднение, как и в случае с предыдущим исследованием (выше мы делали пояснение). Здесь необходимо быть честными, именно в практическом смысле. В реальности — вы приехали на объект, в 95 % случаев, если объект достаточно долго эксплуатируется, достоверную информацию о мониторинге и о его частоте вы получите из нормативных документов, но как он осуществлялся в реальности вы не знаете. И опять же, пусть мониторинг осуществлялся раз в месяц, но тут же возникает вопрос: а фиксируемое отклонение, обнаруженное в конкретном месяце — это в начале месяца, в середине, в конце месяца? Каково распределение отклонений уже в самом периоде. По факту фиксация отклонения в конкретный момент времени в конкретном месяце отнюдь не предполагает, что вы именно в этот самый момент оно и произошло. Вторая проблема состоит в том, что в реальности мониторинг той же сливно — наливной эстакады осуществляется каждый день, но по факту отклонение вы зафиксируете визуально, когда оно станет явным. Вы не будете делать замеры по узлам загружений, и то, на сколько миллиметров произошло отклонение, ни каждый день, ни каждый месяц. Увидите вы это только визуально.

Заключение

Настоящее исследование является логическим продолжением предыдущей работы авторов, которая посвящена Теории экстремальных значений. В данной работе выполнен детальный анализ эволюции теоретических представлений, приведших к формированию современной теории пороговых значений (POT). Показано, что фундамент Теории пороговых значений был заложен в работах советских математиков — Б.В. Гнеденко, Н.В. Смирнова,

А.Н. Колмогорова, чьи идеи (классификация предельных распределений максимумов, анализ числа выходов за высокий уровень, расстояние Колмогорова) получили развитие в трудах Г. Балкемы, Л. де Хаана и Дж. Пикандса. То есть мы постарались устранить распространённый в англоязычной литературе перекокс, сводящий историю метода лишь к работам перечисленных нидерландских и американских математиков. Полученные количественные оценки по конкретной инженерной конструкции, позволяют перейти от констатации факта наличия экстремальных перемещений к объективному инженерному решению о необходимости усиления конструкций или ограничения эксплуатационных нагрузок, что подтверждается математически обоснованными показателями риска обрушения (достижения слишком высоких прогибов). Как видим, Теория пороговых значений представляет собой эффективный и методологически стройный прикладной инструмент для анализа редких событий превышений в строительной механике и смежных инженерных задачах, а её применение позволяет извлекать максимум информации из ограниченных данных о деформациях, нагрузках и других критических параметрах, обеспечивая достоверную основу для проектирования, экспертизы и управления техническим состоянием сооружений. Дальнейшее развитие работы может быть связано с распространением предложенного подхода на другие типы конструкций и виды воздействий, а также с построением доверительных интервалов для прогнозных оценок с использованием «методов бутстрепа» [12; 13] или «асимптотической нормальности» [14; 15].

ЛИТЕРАТУРА

1. Хайруллин В.А. Теория экстремальных значений в оценке долговечности объекта строительства / В.А. Хайруллин, В.Н. Зенцов, Р.Б. Масалимов // Вестник евразийской науки. — 2025. — Т. 17, № 4. — EDN: [ORJGJG](https://doi.org/10.26907/2542-4542.2025.17.4). (дата обращения: 12.12.2024).
2. Balkema A.A. Residual life time at great age / A.A. Balkema, L. de Haan // The Annals of Probability. — 1974. — Т. 2, № 5. — С. 792–804. — DOI: [10.1214/aop/1176996548](https://doi.org/10.1214/aop/1176996548). (дата обращения: 12.12.2024).
3. Gnedenko B. Sur la distribution limite du terme maximum d'une serie aleatoire [О предельном распределении максимального члена случайного ряда] / B. Gnedenko // Annals of Mathematics. — 1943. — Т. 44, № 3. — С. 423–453. — DOI: [10.2307/1968974](https://doi.org/10.2307/1968974). (дата обращения: 12.12.2024).
4. Pickands III J. Statistical inference using extreme order statistics / J. Pickands III // The Annals of Statistics. — 1975. — С. 119–131. — URL: <https://www.jstor.org/stable/2958083>. (дата обращения: 10.05.2025).
5. Hosking J.R.M. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments / J.R.M. Hosking, J.R. Wallis, E.F. Wood // Technometrics. — 1985. — Т. 27, № 3. — С. 251–261. — DOI: [10.1080/00401706.1985.10488049](https://doi.org/10.1080/00401706.1985.10488049). (дата обращения: 12.12.2024).
6. Karamata J. Sur un mode de croissance régulière des fonctions [Об одном режиме регулярного роста функций] / J. Karamata. — 1930. — URL: <http://elibrary.matf.bg.ac.rs/bitstream/handle/123456789/1608/surjka.pdf?sequence=1>. (дата обращения: 10.03.2026).
7. Karamata J. Sur un mode de croissance régulière. Théorèmes fondamentaux [Об одном режиме регулярного роста. Основные теоремы] / J. Karamata // Bulletin de la Société Mathématique de France. — 1933. — Т. 61. — С. 55–62. — URL: <https://www.numdam.org/item/10.24033/bsmf.1196.pdf>. (дата обращения: 10.03.2026).

8. Bingham N.H. Regular variation / N.H. Bingham, C.M. Goldie, J.L. Teugels. — Cambridge: Cambridge University Press, 1989. — Т. 27. — 491 с. — ISBN 0521379431, 9780521379434.
9. Гумбель Э. Статистика экстремальных значений / Э. Гумбель; пер. с англ. В.Ю. Татарского; под ред. Д.М. Чибисова; с предисл. Б.В. Гнеденко. — Москва: Мир, 1965. — 450 с. — URL: <https://libcats.org/book/792024?ysclid=mn21w0w7io185745018>. (дата обращения: 22.03.2026).
10. Beirlant J. Statistics of extremes: theory and applications / J. Beirlant [et al.]. — John Wiley & Sons, 2006. — URL: <https://nzdr.ru/data/media/biblio/edu/math/high/%D0%92%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D1%8C/Beirlant%20J.,%20Goegebeur%20Y.,%20Teugels%20J.,%20Statistics%20of%20Extremes.%20Theory%20and%20Applications.pdf?ysclid=mn23erdv2o123879557>. (дата обращения: 10.03.2026).
11. Акимов В.А. Введение в статистику экстремальных значений и ее приложения / В.А. Акимов, А.А. Быков, Е.Ю. Щетинин; МЧС России; Всероссийский научно-исследовательский институт по проблемам гражданской обороны и чрезвычайных ситуаций МЧС России. — Москва: Всероссийский научно-исследовательский институт по проблемам гражданской обороны и чрезвычайных ситуаций МЧС России, 2009. — 536 с. — ISBN 978-5-93970-037-5. — EDN: [NCHAAR](https://nchaar.ru). (дата обращения: 12.12.2024).
12. Efron B. Bootstrap methods: another look at the jackknife / B. Efron // Breakthroughs in statistics: Methodology and distribution. — New York, NY: Springer New York, 1992. — С. 569–593. — DOI: [10.1007/978-1-4612-4380-9_41](https://doi.org/10.1007/978-1-4612-4380-9_41). (дата обращения: 12.12.2024).
13. Tibshirani R.J. An introduction to the bootstrap / R.J. Tibshirani, B. Efron // Monographs on statistics and applied probability. — 1993. — Т. 57, № 1. — С. 1–436. — URL: https://web.archive.org/web/20170808144050id_/http://cindy.informatik.uni-bremen.de/cosy/teaching/CM_2011/Eval3/pe_efron_93.pdf. (дата обращения: 12.12.2024).
14. Dombry C. Maximum likelihood estimators based on the block maxima method / C. Dombry, A. Ferreira // Bernoulli. — 2019. — Т. 25, № 3. — С. 1690–1723. — DOI: [10.3150/18-BEJ1032](https://doi.org/10.3150/18-BEJ1032). (дата обращения: 12.12.2024).
15. Dombry C. Existence and Consistency of the Maximum Likelihood Estimators for the Extreme Value Index within the Block Maxima Framework / C. Dombry // Bernoulli. — 2015. — Т. 21, № 1. — С. 420–436. — URL: <http://www.jstor.org/stable/43590225>. (дата обращения: 31.03.2026).

Khairullin Vitaly Agzamovich

Ufa State Petroleum Technical University, Ufa, Russia
Scientific Research Design Institute of Architecture and Construction
E-mail: Vitalik000@yandex.ru

ORCID: <https://orcid.org/0000-0002-3854-2193>

RSCI: https://elibrary.ru/author_profile.asp?id=671334

ResearchGate: https://www.researchgate.net/profile/Vitaliy-Khayrullin?ev=hdr_xprf

Terekhov Ivan Gennadievich

Ufa State Petroleum Technological University, Ufa, Russia
Institute of Architecture and Civil Engineering
E-mail: iv_98_04@mail.ru

RSCI: https://elibrary.ru/author_profile.asp?id=528528

Theory and method of thresholds in construction and various engineering tasks

Abstract. This study logically continues the previous work of the authors devoted to the Theory of extreme values in assessing the durability of building structures. The object of research is the theory of extreme values. The subject of the research is the theory of threshold values. The purpose of the study is to present the applied application of Threshold Theory in construction and in various engineering tasks. In this study, we will look at the key works of the subject area. We will also answer the question about the admissibility of the minimum sample size of $n = 30$ in the previous study. This is an important applied point, which some researchers interpreted as a calculation error. We have answered this question in detail. The key methodological differences between extreme values (the block maxima method) and threshold values (all exceedances of a given level) are formulated. A step-by-step threshold method procedure adapted for engineering tasks (primarily in construction) is proposed, including threshold selection, estimation of parameters of the generalized Pareto distribution, model adequacy verification and calculation of probabilistic forecasts with return levels. The practical approbation of the method was performed on real data of vertical movements of the nodes of the draining and filling overpass. The authors have demonstrated that even with a limited amount of initial information (123 observations), the Threshold Value method allows: (1) it is reasonable to choose the threshold $u = 2.0$ mm according to the graph of the average excess; (2) obtain relevant estimates of the GPD parameters ($\xi = 0,58$, $\sigma = 2,94$ mm), indicating a heavy tail of the distribution; (3) confirm the adequacy of the model using the QQ graph and the Kolmogorov distance ($d = 0,069$); (4) calculate the annual probabilities excess of the standard deflection (30 mm) and cumulative risks of deflections above the standard value for 5 and 10 years (61 % and 85 %, respectively), as well as return levels for repeatability periods of 5, 10 and 20 years. We also once again considered an important applied difficulty — the interpretation of the time scale, which plays a key role for effective probabilistic forecasts of the technical condition of an engineering structure.

Keywords: theory of threshold values; generalized Pareto distribution; theory of extreme values; method of threshold exceedances; return levels; Gnedenko's theorem; Pickands-Balkema-de Haan theorem; Karamata theory