На правах рукописи

Исаева Елена Валерьевна

Восстановление функции плотности и оценивание параметров регрессионных зависимостей на основе вейвлет-анализа

Специальность 05.13.17 –

«Теоретические основы информатики»

Автореферат диссертации на соискание учёной степени

кандидата технических наук

Новосибирск – 2022

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего образования «Новосибирский государственный технический университет».

Научный руководитель:	**Тимофеев Владимир Семенович**, доктор технических наук, доцент
Официальные оппоненты:	**Войтишек Антон Вацлавович**, доктор физико-математических наук, профессор, Федеральное государственное бюджетное учреждение науки «Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук», лаборатория стохастических задач, ведущий научный сотрудник;
	Осипов Александр Леонидович, кандидат технических наук, доцент, Федеральное государственное бюджетное образовательное учреждение высшего образования «Новосибирский государственный университет экономики и управления "НИНХ"», кафедра Информационных технологий, доцент
Ведущая организация:	Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский Томский государственный университет»

Защита диссертации состоится 09 сентября 2022 г. в 13:00 часов на заседании диссертационного совета Д 212.173.06 при Федеральном государственном бюджетном образовательном учреждении высшего образования «Новосибирский государственный технический университет» по адресу: 630073, Новосибирск, пр. К. Маркса, 20, I корпус, конференц-зал.

С диссертацией можно ознакомиться в библиотеке Новосибирского государственного технического университета и на сайте http://www.nstu.ru.

Автореферат разослан «___» июля 2022 года.

Ученый секретарь
диссертационного совета, к.т.н., доцент Фаддеенков Андрей Владимирович

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы и степень ее разработанности. Методы теоретической и прикладной статистики нашли широкое применение в различных сферах, среди которых можно выделить оптимизацию сложных технологических процессов, сертификацию технических систем и изделий, геофизические, биологические, генетические и социологические исследования. Возникающие в реальной жизни задачи нередко заставляют исследователя сталкиваться с необходимостью поиска зависимости между входными данными, задающими условия функционирования, и выходными данными, которые характеризуют изучаемый объект. Решение таких задач может быть выполнено путем построения регрессионных моделей, и один из этапов состоит в оценивании неизвестных параметров. Такой подход позволяет не просто восстанавливать исходную зависимость, но и выполнять прогнозирование поведения изучаемого объекта.

Классические методы оценивания неизвестных параметров регрессионных зависимостей позволяют получать достаточно корректные и качественные результаты только при условии, что имеются достоверные предположения о свойствах случайной компоненты. Одним из них является метод максимального правдоподобия. Этот метод основан на предположении о том, что вся информация о статистической выборке содержится в функции правдоподобия. Метод максимального правдоподобия был рекомендован и значительно популяризирован Р. Фишером между 1912 и 1922 годом, хотя ранее он был использован К.Ф. Гауссом и П.С. Лапласом. Применение данного метода возможно при условии, что имеется достоверная информация о виде распределения случайных ошибок наблюдения. Предположение о нормальности распределения ошибок наблюдения позволяет применить метод наименьших квадратов и тем самым упростить поиск оценок. Метод был впервые применен К.Ф. Гауссом в 1795 году, а американский математик Р.А. Эдрейн в 1808-м рассмотрел его теоретико-вероятностные приложения. Работы А.А. Маркова в начале XX века позволили включить метод наименьших квадратов в теорию прикладной математической статистики, в которой он является важной и естественной частью. На практике в большинстве случаев распределение случайной ошибки нельзя считать нормальным, и оценки, полученные таким образом, не позволяют сделать статистически корректные выводы и результаты. Это обстоятельство заставляет исследователей быть не только осторожными при использовании таких оценок, но и искать другие подходы к решению вопроса оценивания неизвестных параметров регрессионной зависимости. В этом случае на помощь исследователю приходят адаптивные методы оценивания параметров регрессионных зависимостей. В данной области можно отметить работы R.V. Hogg, R.V. Lenth, В.И. Мудров, В.И. Денисова. Огромное разнообразие встречающихся распределений случайной ошибки привело к идее восстановления неизвестной функции плотности на основе теории вейвлетов и преобразования Фурье с последующим применением метода максимального правдоподобия. Выбор в пользу такого подхода обусловлен широким использованием вейвлет-анализа и преобразования Фурье для аппроксимации различных функций. Основателем анализа Фурье является французский математик Жан Батист Жозеф Фурье, который сформулировал основы этой теории к 1807 году. С появлением все большего количества экспериментальных данных, которые на тот момент времени обрабатывались преобразованием Фурье, стало понятно: метод имеет ограниченные возможности в поиске закономерностей в этих данных, что привело к возникновению теории вейвлетов. Вейвлет-анализ является сравнительно новым направлением развития в прикладной математике. В 80- х годах Александр Гроссман и Жан Морле в ходе анализа сейсмических и акустических сигналов столкнулись с необходимостью введения нового термина – «вейвлет». До появления термина Альфред Хаар работал над системой базисных функций, которая обладает определенными свойствами и специфичными для вейвлетов признаками, а именно: локальной областью, ортогональностью, единичной нормой, нулевым средним, автомодельностью. В развитии теории вейвлет-анализа можно выделить работы И Добеши, К Чуи, Н.М. Астафьевой, О.В. Нагорнова.

Цель исследования. Целью настоящего исследования является разработка математического и алгоритмического обеспечения для восстановления функции плотности с использованием вейвлет-анализа и преобразования Фурье и адаптивного оценивания параметров линейно-параметризованных регрессионных моделей. Для достижения этой цели были поставлены и решены следующие **задачи:**

1) вычислить нормы для следующих материнских вейвлетов: LITTLEWOOD & PALEY, Морле, DOG, «Мексиканская шляпа» и построить системы базисных функций;

2) разработать и исследовать алгоритмы восстановления функции плотности с использованием ортогональных и неортогональных вейвлетов, а также на основе преобразования Фурье сформулировать рекомендации относительно их использования;

3) разработать и исследовать алгоритмы адаптивного оценивания параметров линейно-параметризованных регрессионных моделей с использованием ортогональных и неортогональных вейвлетов и преобразования Фурье;

4) разработать программный комплекс для адаптивного оценивания параметров линейно-параметризованных регрессионных моделей и восстановления функции плотности распределения и использовать его для решения задачи мониторинга температурных изменений состояния грунта.

Область исследования. Содержание диссертации соответствует п. 5 области исследований «Разработка и исследование моделей и алгоритмов анализа данных, обнаружение закономерностей в данных и их извлечениях, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений» паспорта специальности 05.13.17 – «Теоретические основы информатики» (в области технических наук).

Методы исследования. Для решения поставленных задач использовались базовые понятия математического анализа, линейной алгебры, вычислительной математики, а также теория вероятностей, математическая статистика, регрессионный анализ, вейвлет-анализ, Фурье-анализ, методы оптимизации и методы статистического моделирования.

Достоверность и обоснованность научных положений, выводов и рекомендаций обеспечивается корректными применением аналитических методов исследования, соответствием выводов и результатов вычислительных экспериментов известным теоретическим положениям.

Научная новизна работы состоит в следующем:

1) сформулирован и доказан ряд утверждений, обеспечивающих построение системы базисных функций для восстановления функции плотности распределения;

2) предложены новые алгоритмы, приводящие к повышению качества восстановления функции плотности, на основе ортогональных и неортогональных вейвлетов и преобразования Фурье, а также определены наилучшие значения параметра сглаживания для каждого из материнских вейвлетов;

3) разработаны алгоритмы, повышающие точность оценивания параметров линейно-параметризованных регрессионных моделей с использованием вейвлет-анализа и преобразования Фурье; получены выражения для логарифмической функции правдоподобия и ее производных на основе предложенных оценок функции плотности.

Теоретическая значимость состоит в развитии теории вейвлет-анализа с целью обеспечения возможности использования ненормированных вейвлетов для построения системы базисных функций, а также расширения возможностей регрессионного анализа за счет использования новых оценок функции плотности. В частности, получены выражения для логарифмической функции правдоподобия и ее производных. Такой подход позволяет восстанавливать регрессионные зависимости при условии отсутствия достоверной информации о виде распределения случайных ошибок наблюдения.

Практическая значимость работы заключается в разработанных алгоритмах оценивания параметров линейно-параметризованных регрессионных моделей, когда функция плотности восстанавливается на основе вейвлетов или преобразования Фурье. Разработан программный комплекс, в котором реализованы предложенные алгоритмы оценивания параметров регрессионных моделей и восстановления функции плотности, и зарегистрирован в виде объекта интеллектуальной собственности как программа для ЭВМ (№ гос. рег. 2022613412 от 14.03.2022) [49]. Применение программного комплекса позволило решить прикладную задачу технического характера о прогнозе влияния геокриологических последствий глобального потепления на устойчивость и долговечность жилых зданий и сооружений.

Реализация результатов работы. Результаты диссертационной работы нашли свое применение в ООО «Мерзлотный инженерно-строительный центр», в учебном процессе НГТУ и МАОУ «Инженерный лицей НГТУ», о чем имеются соответствующие акты внедрения.

Положения, выносимые на защиту. На защиту вынесены следующие положения:

1) утверждения об ортонормированности вейвлетов LITTLEWOOD & PALEY, Морле, DOG, «Мексиканская шляпа», позволяющие выполнять построение системы базисных функций;

2) алгоритмы восстановления функции плотности на основе различных ортогональных и неортогональных вейвлетов и преобразования Фурье, а также результаты исследования качества оценивания функции плотности;

3) алгоритмы адаптивного оценивания параметров линейно-параметризованных регрессионных моделей с использованием восстановленных функций плотности, включая выражения логарифмической функции правдоподобия и ее производных, а также результаты их сравнительного анализа;

4) программный комплекс адаптивного оценивания параметров линейно-параметризованных регрессионных моделей и восстановления функции плотности распределения, а также результаты решения задачи о прогнозе влияния геокриологических последствий глобального потепления климата на устойчивость и долговечность жилых зданий и сооружений.

Личный вклад автора в совместных публикациях заключается в:

1) доказательстве утверждений, касающихся свойств ортонормированности материнских вейвлетов и построении на их основе различных систем базисных функций;

2) построении и исследовании алгоритмов восстановления функции плотности распределения случайной величины на основе различных ортогональных и неортогональных вейвлетов;

3) разработке и исследовании алгоритмов адаптивного оценивания параметров линейно-параметризованных регрессионных моделей с использованием вейвлетов: Морле, LITTLEWOOD & PALEY, DOG, «Мексиканская шляпа» и на основе преобразования Фурье;

4) реализации программного комплекса WTiRM V1.0, позволяющего восстанавливать регрессионные зависимости и выполнять оценивание функции плотности, и его использовании для решения прикладных задач.

Апробация работы. Основные результаты работы докладывались и обсуждались на Всероссийской научной конференции молодых ученых «Наука. Технология. Инновации», г. Новосибирск, 2001 г.,2003 г., на Российской научно-технической конференции «Обработка информации и математическое моделирование», г. Новосибирск, 2020 г., 2022 г., на Международной конференции «Информационные технологии в бизнесе и производстве» (ITBI 2020), г. Новосибирск, 2020 г.

Публикации. Основные результаты исследований по теме диссертации опубликованы в 10 печатных работах [1 – 10], из них 3 работы опубликованы в изданиях, рекомендованных ВАК [1 – 3], 1 работа опубликована в издании, индексируемом в базе данных Scopus [4], получено свидетельство о государственной регистрации программы для ЭВМ [10].

Структура и объём работы. По структуре настоящая работа состоит из введения, пяти глав основного содержания, заключения, списка литературы, состоящего из 107 источников, четырёх приложений. Диссертация изложена на 165 страницах основного текста, содержит 41 рисунок и 27 таблиц.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертации, сформулированы цель и задачи исследования, определены научная новизна, теоретическая и практическая значимость работы, дано краткое содержание работы по главам.

Первая глава является обзорной. В разделе 1.1 описываются базовые определения преобразования Фурье и вейвлет-анализа. Обозначим через $L^2(0,2\pi)$ пространство всех измеримых функций $f(t)$, определенных на интервале $(0,2\pi)$, и таких, что выполняется следующее условие $\int_0^{2\pi} |f(t)|^2 dt < \infty$. Функции из $L^2(0,2\pi)$ периодически продолжаемы на всю вещественную ось $R = (-\infty, \infty)$, а именно $f(t) = f(t-2\pi)$ для всех t. Для конструирования базиса пространства $L^2(R)$ измеримых функций $f(t)$, определенных на всей числовой оси, необходимо определить простые функции $\psi(t)$, которые называют вейвлетами. В разделе 1.1.2 представлены наиболее популярные материнские вейвлеты, а именно вейвлет Хаара, LITTLEWOOD & PALEY, Морле, DOG и «Мексиканская шляпа». Функция $\psi(t)$ может быть использована для порождения всего пространства $L^2(R)$ путем целочисленных сдвигов, а именно $\psi(t-k),\ k \in Z$. Базис пространства $L^2(R)$, построенный с помощью непрерывных масштабных преобразований и переносов вейвлета $\psi(t)$ с произвольными значениями базисных параметров, описывается следующим соотношением:

$$\psi_{ab}(t) = |a|^{-1/2}\,\psi\left(\frac{t-b}{a}\right),$$

где $a,b \in R,\ a \neq 0,\ \psi(t) \in L^2(R)$. Тогда прямое вейвлет-преобразование будет иметь вид

$$W(a,b) = \int_{-\infty}^{+\infty} f(t)\,\psi_{ab}(t)\,dt.$$

В дискретном представлении базис пространства $L^2(R)$ выглядит так

$$\psi_{m,n}(t) = a_0^{-\frac{m}{2}}\psi\left(a_0^{-m}t - n\right),$$

где $m, n \in Z,\ a_0 > 1,\ \psi(t) \in L^2(R)$. В этом случае любая функция $f(t) \in L^2(R)$ выражается разложением в ряд

$$f(t)=\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}C_{m,n}\psi_{m,n}(t),$$

где $C_{m,n}=\int\limits_{-\infty}^{+\infty}f(t)\psi_{m,n}(t)dt$.

В разделе 1.2 отмечено, что оценка функции плотности распределения может быть получена как с помощью параметрических методов оценивания, так и не непараметрических. Параметрические методы, такие как метод максимального правдоподобия или метод моментов требуют задания закона распределения случайной величины, что не всегда возможно на практике. Среди непараметрических методов наибольшую известность получила ядерная оценка Розенблатта – Парзена. Однако ее качество сильно зависит от значения параметра сглаживания, неоптимальный выбор которого может приводить к резким флуктуациям графика восстанавливаемой функции плотности, что особенно критично на малых выборках и делает невозможным её дальнейшее использование при решении задач, связанных с оцениванием неизвестных параметров регрессионных моделей. Это обстоятельство заставляет исследователей быть осторожными при использовании ядерных оценок. В разделе 1.3 представлена задача регрессионного анализа и рассмотрены методы оценивания вектора неизвестных параметров уравнения регрессионной зависимости, которые пользуются наибольшей популярностью. В разделе 1.3.1 определена регрессионная модель, считающаяся истинной моделью исследуемого объекта:

$$y=X\theta+\varepsilon, \tag{1}$$

где $X=\begin{bmatrix}\varphi_1(x_{11}) & \dots & \varphi_q(x_{1q})\\ \dots & \dots & \dots\\ \varphi_1(x_{n1}) & \dots & \varphi_q(x_{nq})\end{bmatrix}$ – матрица значений действительных функций, $rg(X)=q$; $\theta=(\theta_1,\theta_2,...,\theta_q)^T$ - оцениваемые неизвестные параметры модели; $\varphi_i(x)$ - известные вещественные функции; x_{ij} – значения входных факторов в n наблюдениях; $y=(y_1,y_2,...,y_n)^T$ – вектор значений отклика; $\varepsilon=(\varepsilon_1,...,\varepsilon_n)^T$ – вектор случайных ошибок, q – число неизвестных параметров; n – число проведенных экспериментов.

Будем считать, что вектор ошибок наблюдений ε состоит из независимых одинаково распределенных случайных величин с функцией плотности $h(t)$, для которых выполняются условия

$$E(\varepsilon_i)=0,\ D(\varepsilon_i)=\sigma^2<\infty. \qquad (2)$$

Требуется по имеющимся значениям отклика и входных факторов выполнить наиболее точное оценивание вектора неизвестных параметров регрессионного уравнения (1).

В пунктах 1.3.2 и 1.3.3 рассматриваются классические методы оценивания параметров регрессионных моделей такие, как метод максимального правдоподобия и метод наименьших квадратов (МНК). В разделах 1.4 и 1.5 представлены устойчивые и адаптивные методы оценивания параметров регрессионных уравнений. В разделе 1.6 проводится анализ существующих программных решений для оценивания неизвестных параметров регрессионных зависимостей и восстановления функции плотности.

Во второй главе рассматривается задача построения алгоритмов для оценивания функции плотности распределения на основе различных ортонормированных базисов, таких как тригонометрическая система функций и система функций на основе вейвлетов. Для построения базисных функций был сформулирован и доказан ряд утверждений и вычислены нормы для вейвлетов Морле, LITTLEWOOD & PALEY, DOG, «Мексиканская шляпа» [2, 7, 8].

Утверждение 1. Пусть $\psi(t)=\dfrac{\sin 2\pi t-\sin \pi t}{\pi t}$ – материнский вейвлет LITTLEWOOD & PALEY, порождающий систему функций $\psi_i(t)\in L^2(R)$, которая принимает вид $\psi_i(t)=2^{\frac{k}{2}}\left(\dfrac{\sin 2\pi(2^k t-(j-1))-\sin \pi(2^k t-(j-1))}{\pi(2^k t-(j-1))}\right)$, где $k\geq 0$, $1\leq j\leq 2^k$, такие что $i=2^k+j$. Тогда для любых i, k, j выполняется $\|\psi_i\|=1$.

Утверждение 2. Пусть $\psi(t)=e^{-t^2/2}\cos 5t$ – материнский вейвлет Морле, порождающий систему функций $\psi_i(t)\in L^2(R)$, которая принимает вид $\psi_i(t)=2^{\frac{k}{2}}e^{\frac{-(2^k t-(j-1))^2}{2}}\cos 5(2^k t-(j-1))$, где $k\geq 0$, $1\leq j\leq 2^k$, такие что $i=2^k+j$. Тогда для любых i, k, j выполняется $\|\psi_i\|=\dfrac{\sqrt[4]{\pi}}{\sqrt{2}}\left(1+e^{-25}\right)^{1/2}$.

Утверждение 3. Пусть $\psi(t)=e^{-\frac{t^2}{2}}-\dfrac{1}{2}e^{-\frac{t^2}{8}}$ – материнский вейвлет DOG, порождающий систему функций $\psi_i(t)\in L^2(R)$, которая принимает вид

$\psi_i(t) = 2^{\frac{k}{2}}\left(e^{-\frac{1}{2}(2^k t-(j-1))^2} - \frac{1}{2}e^{-\frac{1}{8}(2^k t-(j-1))^2}\right)$, где $k \geq 0$, $1 \leq j \leq 2^k$, такие что $i = 2^k + j$. Тогда для любых i, k, j выполняется $\|\psi_i\| = \left(\frac{3}{2}\sqrt{\pi} - \frac{2}{5}\sqrt{10\pi}\right)^{\frac{1}{2}}$.

Утверждение 4. Пусть $\psi(t) = (1-t^2)e^{-\frac{t^2}{2}}$ – материнский вейвлет «Мексиканская шляпа», порождающий систему функций $\psi_i(t) \in L^2(R)$, которая принимает вид $\psi_i(t) = 2^{\frac{k}{2}}\left(1-(2^k t-(j-1))^2\right)e^{\frac{-(2^k t-(j-1))^2}{2}}$, где $k \geq 0$, $1 \leq j \leq 2^k$, такие что $i = 2^k + j$. Тогда для любых i, k, j выполняется $\|\psi_i(t)\| = \frac{\sqrt{3}}{2}\pi^{\frac{1}{4}}$.

Доказательства представленных выше утверждений опубликованы в работах [2, 7, 8].

Оценка функции плотности $\hat{f}_n(t)$ случайной величины на произвольном отрезке $[c,d]$ выражается

$$\hat{f}_n(t) = \sum_{i=1}^{N} \hat{c}_i \psi_i(t), \tag{3}$$

где N – число членов ряда (параметр сглаживания), $\hat{c}_i$ – оценки коэффициентов разложения по данному базису, которые определяются по имеющимся статистическим данным

$$\hat{c}_i = \frac{1}{n}\sum_{j=1}^{n} \psi_i(x_j).$$

В качестве ортонормированной системы $\psi_i(t)$ рассматривался набор базисных функций, ортонормированных на $[0,1]$:

$$\psi_i(t) = 2^{\frac{k}{2}}\psi(2^k t-(j-1)), \tag{4}$$

где $k \geq 0$, $1 \leq j \leq 2^k$, такие что $i = 2^k + j$, $\psi(t)$ – материнский вейвлет. Если порождающий вейвлет $\psi(t)$ не является нормированным, то при конструировании системы функций $\psi_i(t)$ необходимо добавить нормировочный множитель $z = \frac{1}{\|\psi_i(t)\|}$.

Базисные функции (4) с учетом доказанных утверждений принимают вид (см. таблица 1) [1, 2, 3 – 9]:

Таблица 1 – Базисные функции на основе различных материнских вейвлетов

Материнский вейвлет	Базисные функции, где $\tau=\frac{t-c}{d-c}$, $z=\frac{1}{\lVert\psi_i(t)\rVert}$
LITTLEWOOD & PALEY	$\psi_i(t)=\frac{1}{\sqrt{d-c}}2^{\frac{k}{2}}\left(\frac{\sin 2\pi(2^k\tau-(j-1))-\sin\pi(2^k\tau-(j-1))}{\pi(2^k\tau-(j-1))}\right)$,
Морле	$\psi_i(t)=\frac{z}{\sqrt{d-c}}2^{\frac{k}{2}}e^{\frac{-(2^k\tau-(j-1))^2}{2}}\cos 5(2^k\tau-(j-1))$,
DOG	$\psi_i(t)=\frac{z2^{\frac{k}{2}}}{\sqrt{d-c}}\left(e^{-\frac{(2^k\tau-(j-1))^2}{2}}-\frac{1}{2}e^{-\frac{(2^k\tau-(j-1))^2}{8}}\right)$
«Мексиканская шляпа»	$\psi_i(t)=\frac{z2^{\frac{k}{2}}}{\sqrt{d-c}}(1-(2^k\tau-(j-1))^2)e^{-\frac{(2^k\tau-(j-1))^2}{2}}$

Отсутствие ортогональности системы базисных функций может негативно сказаться на качестве восстановления функции плотности, поэтому для вейвлетов DOG и «Мексиканская шляпа» было предложено ввести нормировочный множитель $\tilde{z}$, который является более общей поправкой. Выражение для $\tilde{z}$ получено в утверждениях 4 и 5, доказательство которых представлено в работах [2, 4].

Утверждение 5. Пусть $\psi(t)=(1-t^2)e^{-\frac{t^2}{2}}$ – материнский вейвлет «Мексиканская шляпа», порождающий систему функций $\psi_i(t)\in L^2(R)$, которая принимает вид $\psi_i(\tau)=\tilde{z}2^{\frac{k}{2}}(1-\tau^2)e^{-\frac{\tau^2}{2}}$, где $\tau=\frac{2^k}{d-c}(t-c)-(j-1)$, $k\geq 0$, $1\leq j\leq 2^k$, $i=2^k+j$. Тогда для любых i, k, j нормировочный множитель выражается соотношением $\tilde{z}=\left(\frac{\sqrt{\pi}}{n}(2-\sqrt{2})\sum_{j=1}^{n}\sum_{i=1}^{N}2^{-\frac{k}{2}}\psi_i(x_j)\right)^{-\frac{1}{2}}$.

Утверждение 6. Пусть $\psi(t)=e^{-\frac{t^2}{2}}-\frac{1}{2}e^{-\frac{t^2}{8}}$ – материнский вейвлет DOG, порождающий систему функций $\psi_i(t)\in L^2(R)$, которая принимает вид

$\psi_i(\tau) = 2^{\frac{k}{2}} \tilde{z} \left(e^{-\frac{\tau^2}{2}} - \frac{1}{2} e^{-\frac{\tau^2}{8}} \right)$, где $\tau = \frac{2^k}{d-c}(t-c) - (j-1)$, $k \ge 0$, $1 \le j \le 2^k$, $i = 2^k + j$. Тогда для любых i, k, j нормировочный множитель выражается соотношением $\tilde{z} = \left(\frac{\sqrt{\pi}}{n}(2-\sqrt{2}) \sum_{j=1}^{n} \sum_{i=1}^{N} 2^{-\frac{k}{2}} \psi_i(x_j) \right)^{-\frac{1}{2}}$.

Также был предложен другой способ определения нормировочного множителя $\tilde{z}$, который заключается в анализе степени близости истинной функции плотности и ее оценки путем минимизации значения статистики χ^2:

$$\chi^2 = \sum_{r=1}^{T} (x_{r+1} - x_r) \frac{\left(O_r - \sum_{j=1}^{n} \sum_{i=1}^{N} \tilde{z}^2 \psi_i(t_r) \psi_i(x_j) \right)^2}{\sum_{j=1}^{n} \sum_{i=1}^{N} \tilde{z}^2 \psi_i(t_r) \psi_i(x_j)}, \qquad (5)$$

где T – количество интервалов разбиения выборки, полученное по формуле Стерджеса t_r – середина r - го интервала; O_r – фактическое число элементов выборки, попавших в r - й интервал. Тогда искомое значение $\tilde{z}^*$ будет получено как решение следующей оптимизационной задачи: $\tilde{z}^* = \arg\min_{\tilde{z}} \chi^2$. В рамках данной работы процесс минимизации был выполнен методом золотого сечения после предварительной локализации интервала значений $\tilde{z}$. В этом случае исходная функция плотности распределения должна быть известна априорно, что не всегда выполнимо на практике. В данной работе $f(t)$ была заменена на ее эмпирическую оценку, полученную по выборке.

В разделах 2.4 – 2.7 было выполнено исследование качества восстановления функции плотности на основе сконструированных базисных функций. Данное исследование проводилось на основе технологии статистического моделирования. На рисунке 1, *а – г* представлены некоторые из полученных результатов восстановления функции плотности для стандартного нормального распределения $f(t)$ при следующих параметрах сглаживания вейвлетов LITTLEWOOD & PALEY и Морле – это $N = 5$, для вейвлета DOG – это $N = 34$, а для вейвлета «Мексиканская шляпа» - это $N = 8$, объем выборки – 200. Одним из результатов следует считать рекомендации по выбору параметра сглаживания.

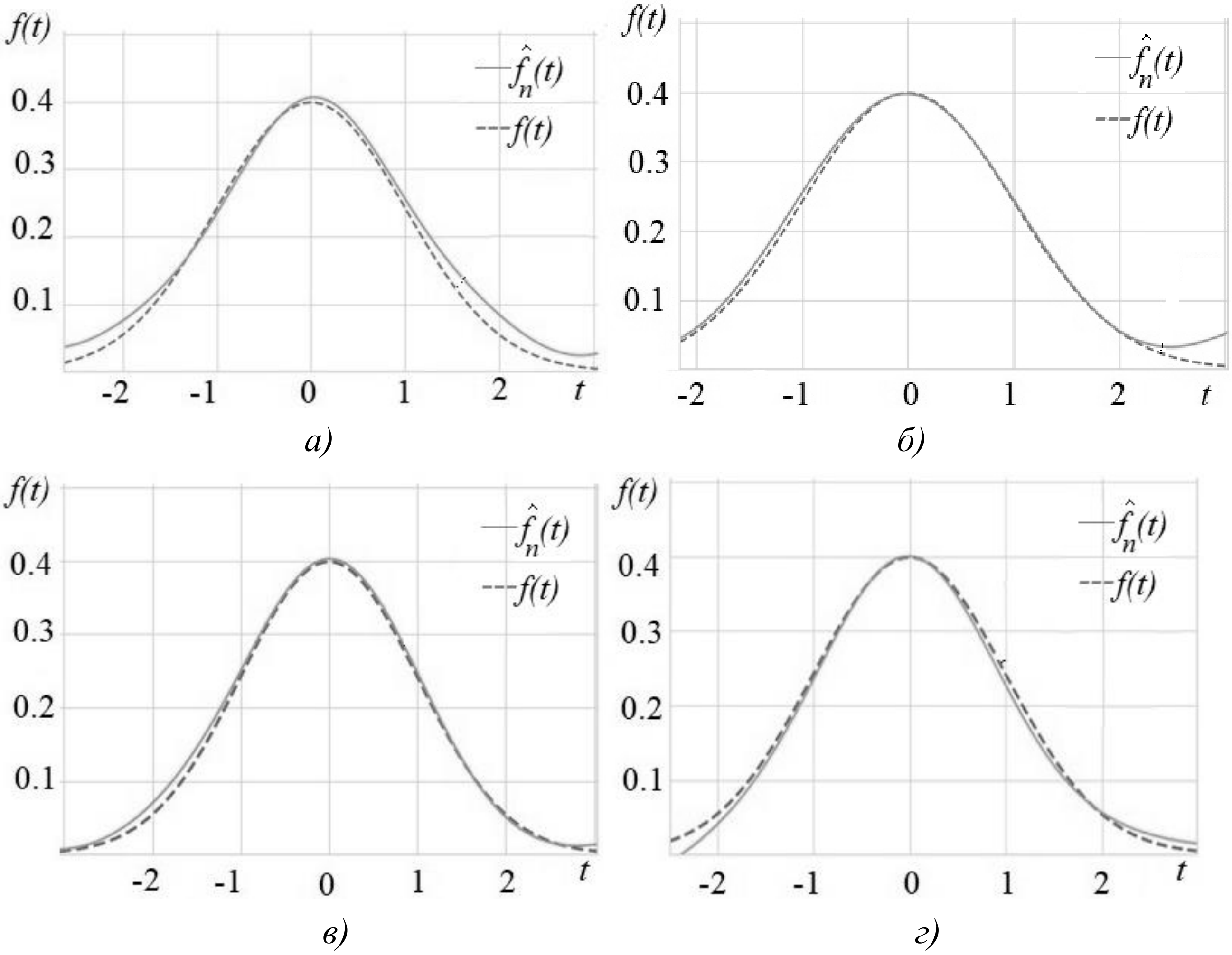

Рисунок 1 – Оценка плотности распределения на основе
а) LITTLEWOOD & PALEY; *б)* Морле; *в)* DOG; *г)* «Мексиканская шляпа»

Из рисунка 1 видно, что оценка $\hat{f}_n(t)$ функции плотности распределения и истинная функция $f(t)$ очень близки. На «хвостах» распределения наблюдается некоторое увеличение отклонения $\hat{f}_n(t)$ от истинной функции плотности. Наибольшее отклонение наблюдается для оценок на основе вейвлета Морле и LITTLEWOOD & PALEY, а наименьшее – для вейвлета DOG. В работе было установлено, что качество вейвлет оценки плотности распределения зависит как от выбор материнского вейвлета, так и от значения параметра сглаживания. При малом значении N имеет место существенное отклонение оценки $\hat{f}_n(t)$ от истинной плотности, а при достаточно большом N оценки функции плотности имеют дополнительные экстремумы, которые ухудшают её качество. Количество таких экстремумов зависит от материнского вейвлета, который используется при построении оценки. Исследование качества восстановления $f(t)$ показало, что наилучшей оценкой $\hat{f}_n(t)$ является оценка на основе материнского вейвлета DOG.

В третьей главе были предложены и исследованы алгоритмы оценивания неизвестных параметров регрессионных моделей, основанные на технологии метода максимального правдоподобия. При этом использовался тот факт, что значения остатков $e_i = y_i - X_i\widehat{\theta}$ – независимые случайные величины с плотностью распределения $h(e_i, \theta)$.

Шаг 1. На основе МНК определить начальное значение вектора неизвестных параметров $\widehat{\theta}^0 = (\widehat{\theta}_1^0, \widehat{\theta}_2^0, ..., \widehat{\theta}_q^0)^T$ уравнения (1), при $l = 0$, где l – номер итерации.

Шаг 2. Вычислить значения остатков $e_i = y_i - X_i\widehat{\theta}^l$ регрессионной модели.

Шаг 3. Получить оценку функции плотности распределения с помощью соотношения (3) по какому-либо набору нормированных базисных функций из таблицы 1.

Шаг 4. Определить значение логарифмической функции правдоподобия $\ln L(e_1, ..., e_n, \widehat{\theta}) = \ln\left(\prod_{i=1}^{n} h(e_i, \widehat{\theta})\right) = \sum_{i=1}^{n} \ln\left(h(e_i, \widehat{\theta})\right)$.

Шаг 5. Найти значение оценки вектора неизвестных параметров $\widehat{\theta}^{l+1} = \arg\max_{\theta} L(e_1, ..., e_n, \widehat{\theta}^l)$.

Шаг 6. Итерационный процесс завершается, если $\left\|\widehat{\theta}^{l+1} - \widehat{\theta}^l\right\| < \delta$, где δ – заданная погрешность вычисления. Если же $\left\|\widehat{\theta}^{l+1} - \widehat{\theta}^l\right\| > \delta$. то выполняется переход на шаг 2 при условии $l = l + 1$.

В разделе 3.2 получены логарифмические функции правдоподобия на основе преобразования Фурье, где набор базисных функций $\psi_i(t)$ получен через тригонометрическую систему функций [8]. В разделе 3.3 получены логарифмические функции правдоподобия на основе ортогональных вейвлетов LITTLEWOOD & PALEY и Морле соответственно [3]:

- $\ln(L(e_1, ..., e_n, \widehat{\theta})) = -n\ln n + \sum_{s=1}^{n} \ln(\sum_{j=1}^{n}\sum_{i=1}^{N} \frac{2^k}{d-c} \frac{\sin 2\pi\tau_s - \sin \pi\tau_s}{\pi\tau_s} \frac{\sin 2\pi_j\tau - \sin \pi_j\tau}{\pi\tau_j})$;
- $\ln(L(e_1, ..., e_n, \widehat{\theta})) = -n\ln n + \sum_{s=1}^{n} \ln(\sum_{j=1}^{n}\sum_{i=1}^{N} \frac{z^2 2^k}{d-c} e^{\frac{-\tau^2_s}{2}} e^{\frac{-\tau^2{}_j}{2}} \cos 5\tau_s \cos 5\tau_j)$;

В разделе 3.4 получены логарифмические функции правдоподобия на основе неортогональных вейвлетов DOG и «Мексиканская шляпа» соответственно [3]:

- $\ln\left(L(e_1,...,e_n,\widehat{\theta})\right) = -n\ln n + \sum_{s=1}^{n}\ln(\sum_{j=1}^{n}\sum_{i=1}^{N}\frac{z^2 2^k}{d-c}(e^{\frac{-\tau_s{}^2}{2}} - \frac{1}{2}e^{\frac{-\tau_s{}^2}{8}})(e^{\frac{-\tau_j{}^2}{2}} - \frac{1}{2}e^{\frac{-\tau_j{}^2}{8}}))$;
- $\ln\left(L(e_1,...,e_n,\widehat{\theta})\right) = -n\ln n + \sum_{s=1}^{n}\ln(\sum_{j=1}^{n}\sum_{i=1}^{N}\frac{z^2 2^k}{d-c}(1-\tau_s{}^2)e^{\frac{-\tau_s{}^2}{2}}(1-\tau_j{}^2)e^{\frac{-\tau_j{}^2}{2}})$.

В разделе 3.5 представлены результаты исследования предложенных алгоритмов оценивания неизвестных параметров θ регрессионной зависимости (1) на основе преобразования Фурье и различных материнских вейвлетов. Для этого были проведены многочисленные вычислительные эксперименты. Была рассмотрена регрессионная зависимость

$$y = \theta_1 + \theta_2 x + \theta_3 x^2 + \varepsilon. \qquad (6)$$

Отметим, что значения входного фактора x определялись из отрезка $[-4,4]$, $q=3$, истинные значения параметров $\theta_1 = 4, \quad \theta_2 = -9, \quad \theta_3 = 2$. Элементы вектора ошибок наблюдений ε_i моделировались независимыми с функцией распределения вида $F(x) = (1-\mu)F_1(x,0,\sigma_1) + \mu F_2(x,0,\sigma_2)$, где $F_i(x,0,\sigma_i)$ – функция нормального распределения с нулевым математическим ожиданием и дисперсией σ_i^2, $i=1,2$, $\mu\in[0,1]$- параметр смеси. В процессе моделирования предполагалось. что $\sigma_2^2 \geq \sigma_1^2$, а значения дисперсий σ_1^2 и σ_2^2 задавались через значения уровня шума, который определяется соотношением $\rho = \frac{\delta}{\delta_c}\cdot 100\%$, где δ- дисперсия ошибки, $\delta_c{}^2 = (n-1)^{-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ - интенсивность сигнала. В качестве показателя точности оценивания параметров использовалось следующее соотношение $\nu_1 = \frac{1}{R}\sum_{i=1}^{R}\sum_{j=1}^{3}\frac{\left|\theta_j - \widehat{\theta}_{ij}\right|}{\left|\theta_j\right|}$, где R- число вычислительных экспериментов, $\widehat{\theta}_{ij}$ - оценка j-го параметра регрессионной зависимости (6) в i-ом вычислительном эксперименте. Количество вычислительных экспериментов для различных комбинаций μ и ρ варьировалось от 100 до 500. Результаты исследования точности оценивания неизвестных параметров уравнения (6) представлены на рисунке 2 для объема выборки 200. Было отмечено, что алгоритмы оценивания, основанные на таких материнских вейвлетах, как Морле, LITTLEWOOD&PALEY, DOG, «Мексиканская шляпа» и преобразовании Фурье показывают хорошую точность оценивания при разных условиях формирования выборок.

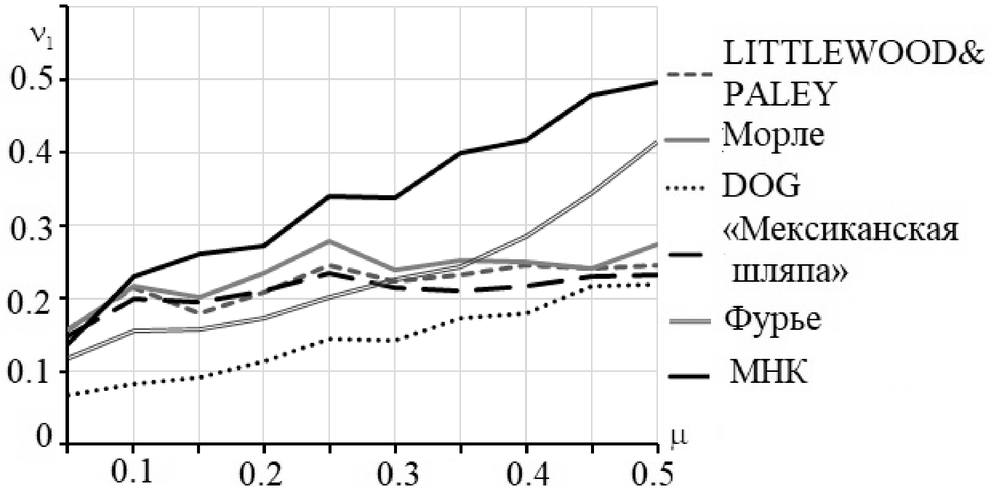

Рисунок 2 – Значение ν_1 в зависимости от доли выбросов

Из рисунка 2 видно, что при значениях $\mu \in [0.05, 0.1]$ алгоритмы на основе LITTLEWOOD & PALEY, Морле и «Мексиканская шляпа» имеют точность, сопоставимую с методом наименьших квадратов. С ростом значений μ, т.е. с ростом числа грубых ошибок наблюдений в исходных данных, алгоритмы на основе вейвлетов и преобразования Фурье дают более точную оценку параметров регрессионной модели в сравнении с МНК. При $\mu = 0.5$ значения показателей ν_1 для алгоритма на основе преобразования Фурье меньше на 14% значения ν_1, полученного при оценивании параметров МНК. Оценки, полученные с помощью алгоритмов на основе вейвлет-анализа, дают более точную оценку неизвестных параметров регрессионных зависимостей в сравнении с алгоритмом на основе преобразования Фурье (см. рисунок 2). Наилучшее качество оценивания неизвестных параметров уравнения (6) получается в результате использования вейвлета DOG в качестве базисного, как при малых значениях параметра μ, так и при близких к 0.5. При $\mu = 0.05$ и при $\mu = 0.5$ значение показателя ν_1, полученного при использовании алгоритма на основе DOG – вейвлета на 43% и 46% соответственно меньше значения ν_1, полученного при оценивании параметров методом наименьших квадратов.

В четвертой главе описаны общие сведения о разработанном программном комплексе оценивания параметров регрессионных моделей и восстановления функции плотности WTiRM V1.0 [10], такие как функциональное назначение, логическая структура, технические характеристики, графический интерфейс.

В пятой главе решена реальная техническая задача прогнозирования температурного состояния грунтов с помощью разработанных алгоритмов. Выходной признак y – температура грунта. В качестве входных факторов рассмотрены: время (квартал) и дата измерения (год), расположение скважины, шифр сваи, глубина измерения в метрах. Отметим, что только «глубина измерения» является количественным фактором, оставшиеся являются качественными, использование которых в модели потребовало введения некоторого числа фиктивных переменных. В результате искомая модель была записана в форме модели ковариационного анализа, которая является более компактной

$$y_{ijkr} = \mu + \alpha_j + \beta_k + \gamma_r + \theta_1 h_{ijkr} + \theta_2 h_{ijkr}^2 + \theta_3 d_{ijkr} + \varepsilon_{ijkr}, \qquad (7)$$

где $\theta = \left(\mu, \alpha_1, ..., \alpha_4, \beta_1, ..., \beta_p, \gamma_1, ..., \gamma_s, \theta_1, \theta_2, \theta_3\right)^T$ –вектор неизвестных параметров, p - количество уровней фактора «расположение скважины», s – количество уровней фактора «шифр сваи», y_{ijkr}, h_{ijkr}, d_{ijkr}, ε_{ijkr} – значения i – го наблюдения температуры, глубины измерения, «дата измерения» и случайной ошибки для «времени измерения» j, «расположение скважины» k и «шифра» r соответственно. Оценивание параметров модели (7) проводилось после её редукции. Использование разработанных алгоритмов на основе преобразования Фурье и на основе вейвлетов позволило получить оценки параметров модели, описывающей изменения температуры грунтов.

ЗАКЛЮЧЕНИЕ

Основные результаты, полученные в работе, заключаются в следующем:

1. Сформулирован и доказан ряд утверждений, на основе которых получены нормы для материнских вейвлетов LITTLEWOOD & PALEY, Морле, DOG, «Мексиканская шляпа» и построены системы базисных функций на их основе. Сформулирован и доказан ряд утверждений, позволяющих улучшить точность оценивания функции плотности распределения. Предложены три способа вычисления нормировочного коэффициента для построения системы базисных функций.

2. Предложены, разработаны и исследованы алгоритмы восстановления функции плотности с использованием ортогональных и неортогональных вейвлетов, а также на основе преобразования Фурье; сформулированы рекомендации относительно их использования.

3. Разработаны и исследованы алгоритмы адаптивного оценивания параметров линейно-параметризованных регрессионных моделей с использованием ортогональных и неортогональных вейвлетов и преобразования Фурье, получены выражения для логарифмической функции правдоподобия и ее производных.

4. Создан программный комплекс адаптивного оценивания параметров линейно-параметризованных регрессионных моделей и восстановления функции плотности распределения.

5. С использованием предложенных алгоритмов решена задача о прогнозе влияния геокриологических последствий глобального потепления климата на устойчивость и долговечность жилых зданий и сооружений.

ПУБЛИКАЦИИ АВТОРА ПО ТЕМЕ ДИССЕРТАЦИИ

Издания из Перечня ВАК ведущих рецензируемых научных изданий для опубликования основных научных результатов диссертаций

1. *Тимофеев В. С.* Об оценивании функции плотности распределения случайной величины с использованием вейвлетов / В. С. Тимофеев, Е. В. Исаева // Научный вестник НГТУ. – 2019. – № 4 (77). – С. 71–84.

2. *Тимофеев В. С.* Особенности оценивания функции плотности распределения с помощью не ортонормированных вейвлетов / В. С. Тимофеев, Е. В. Исаева // Южно-Сибирский научный вестник. – 2022. – № 2 (42). – С. 87–94. – URL: http://s-sibsb.ru/issues/122-2022-issues/issue-42/1348-12.

3. *Тимофеев В. С.* Об оценивании параметров регрессионных моделей с использованием вейвлетов / В. С. Тимофеев, Е. В. Исаева // Современные наукоемкие технологии. – 2022. – № 4. – С. 114–121.

Международные рецензируемые издания:

4. *Timofeev V. S.* Estimating the distribution density function using a DOG wavelet / V. S. Timofeev, E. V. Isaeva // Journal of Physics: Conference Series. – 2020. – Vol. 1661. – Art. 012084. – DOI 10.1088/1742-6596/1661/1/012084.

Другие издания:

5. *Исаева Е. В.* К вопросу об отслеживании смысловой целостности текста / Е. В. Исаева, В. А. Селезнев // Наука. Техника. Инновации : тезисы докладов : в 5 ч. – Новосибирск, 11-12 дек. 2001 г.. – Ч. 1. – С. 116–117.

6. *Исаева Е. В.* О задаче идентификации смысловой целостности художественных текстов / Е. В. Исаева, В. А. Селезнев // Наука. Технологии. Инновации : труды. – Новосибирск, 2003. – С. 224–225.

7. Оценивание функции плотности распределения с использованием вейвлета Литлвуда – Пэли / В. С. Тимофеев, Е. В. Исаева, Е. Д. Малышкина, А. Э. Слободчикова // Обработка информации и математическое моделирование : материалы Рос. науч.-техн. конф., Новосибирск, 23–24 апр. 2020 г. – Новосибирск, 2020. – С. 165–170.

8. *Исаева Е. В.* Оценивание функции плотности распределения с использованием вейвлета Морле = Estimation of the distribution density function using the Morlet wavelet / Е. В. Исаева. – DOI 10.17513/mjpfi.13353 // Международный журнал прикладных и фундаментальных исследований. – 2022. – № 2. – С. 22–27.

9. Оценивание функции плотности распределения на основе преобразований Фурье = Estimation of the distribution function based on the Fourier transform / В. С. Тимофеев, Е. В. Исаева, А. Э. Слободчикова, Е. Д. Малышкина. // Обработка информации и математическое моделирование : материалы Рос. науч.-техн. конф., Новосибирск, 20–21 апр. 2022 г. – Новосибирск : Изд-во СибГУТИ, 2022. – С. 170–177. – 20 экз. – ISBN 978-5-91434-070-1.

Свидетельства о государственной регистрации права для ЭВМ

10. Программный комплекс для восстановления функции плотности и оценивания параметров регрессионных моделей на основе вейвлет анализа (WTiRM V1.0): свидетельство о гос. регистрации программы для ЭВМ № 2022613412 : заявл. 24.02.2022 : зарег. 14.03.2022 / Исаева Е. В., Малышкина Е. Д., Слободчикова А. Э., Тимофеев В. С. ; правообладатель Новосиб. гос. техн. ун-т. – 1 с.

Отпечатано в типографии
Новосибирского государственного технического университета
630073, г. Новосибирск, пр. К. Маркса, 20, тел. (383) 346-08-56
Формат 60×90/16. Усл. печ. л. 1.1. Тираж 100 экз.
Подписано в печать 07.07.2022. Заказ №1186