

МЕТОДЫ ЭКСПЕРИМЕНТАЛЬНОЙ ФИЗИКИ

УДК 537.86; 519.25

ПРИМЕНЕНИЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ
ДЛЯ ФИЛЬТРАЦИИ ШУМОВ ВО ВРЕМЕННЫХ
СИГНАЛАХ КОНЦЕНТРАЦИЙ ДИОКСИДА СЕРЫ
В АТМОСФЕРНОМ ВОЗДУХЕ

В. А. Иванов^{1,2}, Н. Н. Скворцова^{1,3}, Е. А. Позднякова², Е. С. Жигачева²

Представлен подход к обработке сигналов мониторинга загрязняющих веществ с применением вейвлет-фильтрации многолетних данных на базе вейвлетов типа Добеши 4. Вейвлет-преобразование использовано для фильтрации помех во временных рядах данных концентраций диоксида серы в атмосферном воздухе. Показана некорректность предварительной фильтрации рядов данных от шума универсальным пороговым ограничением по методу Донохо–Джонстона и целесообразность применения ручного порога 0.45 от робастного стандартного отклонения для сохранения полезного сигнала.

Ключевые слова: вейвлет-фильтрация, атмосферный воздух, зашумленный сигнал, стохастические погрешности, нестационарность временных рядов.

Введение. Диоксид серы (SO_2) является одним из ключевых загрязняющих веществ атмосферного воздуха [1]. Анализ многолетних трендов загрязнений критически важен для оценки и отслеживания величины экологических рисков. Длительный непрерывный мониторинг атмосферного воздуха сталкивается с проблемой нестационарности временных рядов, накопленных стохастических погрешностей и случайных ошибок. Временные ряды данных, полученные на сети Международной совместной программы мониторинга и оценки дальних переносов атмосферных загрязняющих веществ в

¹ МИРЭА, 119454 Россия, Москва; e-mail: dunkelheit2015@yandex.ru.

² Институт глобального климата и экологии имени академика Ю. А. Израэля, 107258 Россия, Москва.

³ ИОФ РАН, 119991 Россия, Москва, ул. Вавилова, 38.

Европе (ЕМЕП), обладают достаточной физической достоверностью. Однако они содержат в себе шумы, вызванные различными факторами: аппаратурной флуктуацией аналитического прибора, суточной метеорологической вариабильностью, человеческим фактором.

Вейвлет-анализ широко применяется для практических аспектов к интерпретации сигналов различной природы [2]. Особенностью рядов данных в мониторинге атмосферного воздуха является их неоднородность, вызванная наличием экстремально высоких концентраций, сезонностью, отсутствием части значений, изменением методов и подходов к наблюдению. Одним из основных требований к станциям наблюдения ЕМЕП является их расположение на удалении от крупных городов и промышленных центров на более чем 50–100 км (фоновые станции). Поэтому, получаемые на станциях ЕМЕП ряды данных могут содержать в своем составе значения, приближенные к наименьшему уровню детектирования прибора или даже ниже предела определения (НПО). Это увеличивает неоднородность данных, так как подобные значения нельзя полностью охарактеризовать, как достоверные или присвоить им нулевые значения. Накопленные за многолетний период наблюдения шумы и “случайные выбросы” во временных рядах искажают результаты оценивания, выявления трендов [3]. Для более точного прогнозирования тренда необходима предварительная подготовка данных, то есть проведение процедуры сглаживания временного ряда [4].

В настоящей статье разработан подход вейвлет-анализа для очистки шумов в рядах данных концентраций диоксида серы в атмосферном воздухе, полученных в рамках выполнения программы ЕМЕП на основе многолетних баз данных по содержанию SO_2 в атмосферном воздухе на станции Приокско-Террасного государственного заповедника.

Методика вейвлет-фильтрации рядов данных концентраций SO_2 в атмосферном воздухе на территории Приокско-Террасного государственного заповедника. Мониторинг атмосферного воздуха на территории Приокско-Террасного государственного заповедника по программе ЕМЕП проводится более 20 лет. В данной работе был выбран период наблюдений с 2009 по 2024 год.

Одной из особенностей исследуемого ряда данных является то, что они содержат в себе пропуски различного масштаба (сутки, недели, месяцы, годы). При работе с имеющейся базой данных было выявлено, что максимальная длина пропусков подряд оставляет 429 дней. Также имеются суточные, недельные и месячные пропуски. Поскольку алгоритмы дискретного вейвлет-анализа требуют непрерывности сетки, то был применен двухэтапный (гибридный) подход, основывающийся на структурном восстанов-

лении макропропусков и нелинейном заполнении микропропусков. Пропуски значений длиной до 7 дней, заполнялись по методу нелинейной интерполяции с сезонной декомпозицией при помощи алгоритма `na_seadec` [6]. Пропуски более 7 дней заполнялись с помощью непараметрического алгоритма случайного леса (`Randomforest`), реализованного в пакете `missRanger` (`num.trees=1000`, `pmm.k=5`). Для корректной работы календарные даты были преобразованы в тригонометрические функции (синус и косинус) для замыкания годового цикла. Непрерывный временной индекс был заменен на категориальный фактор года для фиксации межгодовых трендов и исключения ложной экстраполяции. Моделирование проводилось на логарифмированных данных, после чего результаты были возвращены в исходную шкалу с применением экспонирования, что позволило исключить ошибочные отрицательные значения концентраций SO_2 [7, 8].

После восстановления микро- и макропропусков, для очистки временного ряда от стохастического шума было применено максимальное перекрывающееся дискретное вейвлет-преобразование (`MODWT`). В качестве материнского вейвлета был выбран фильтр Добеши 4-го порядка с уровнем разложения 3 [9].

Глубина разложения 3-го уровня позволяет изолировать и подвергнуть процедуре жесткой пороговой фильтрации высокочастотные флуктуации и приборный шум масштабом от 2 до 16 дней, что позволяет не исказить при этом сезонные и многолетние тренды. Очистка рядов от шума осуществлялась методом жесткого порогового ограничения (`hardThresholding`) с расчетом универсального порога по методу Донохо–Джонстона (применяемого адресно к первому масштабу деталей) и фиксированным ручным порогом в $0.45 \times \text{MAD}$ для всех уровней детализации (`d1-d3`).

Выбор жесткого порога фильтрации, а не мягкого (`softthresholding`) обусловлен спецификой данных экологического мониторинга фоновых территорий. Жесткий порог убирает только мелкие колебания, которые не превышают уровень приборного и стохастического шума, оставляя реальные концентрации. Данный метод позволяет избежать искусственного занижения амплитуды высоких концентраций и сохранить достоверность временного ряда [10]. Универсальный порог для каждого уровня разложения рассчитывался по формуле:

$$\lambda = \bar{\sigma} \sqrt{2 \ln(N)}, \quad (1)$$

где N – длина временного ряда, σ – робастная оценка стандартного отклонения (`MAD`) вейвлет-коэффициентов на текущем уровне.

Особенностью вейвлет-функций является их колебательный характер вокруг нулевой оси, что может приводить к появлению отрицательных математических значений в

восстановленном ряде данных. Для соблюдения физического закона сохранения массы авторами работы было применено жесткое нелинейное ограничение к итоговому очищенному ряду – все расчетные значения меньше нуля приравнялись к минимальному значению концентрации за 15 лет наблюдений с 2009 по 2024 год.

Работа была выполнена в Rstudio на языке *R* с использованием библиотек: (*readxl*), (*waveslim*), (*ggplot2*), (*dplyr*) (*missRanger*), (*imputeTS*), (*writexl*), (*gridExtra*).

В ходе апробации вейвлет-фильтрации (универсальный порог Донохо–Джонстона и ручной порог $0.45 \times \text{MAD}$) рассчитаны 4 метрики качества (табл. 1). Задавались следующие значения: N – общее количество дней (наблюдений) во временном ряду; X_t – исходный (заполненный после пропусков, но зашумленный) ряд концентраций SO_2 в момент времени t ; $\hat{X}$ – очищенный с помощью вейвлет-преобразования MODWT ряд концентраций SO_2 в момент времени t . В табл. 2 приведены рассчитанные метрики качества.

Т а б л и ц а 1

Выражения для вычисления метрик качества

RMSE (среднеквадратическая ошибка) [11]	$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (X_t - \hat{X}_t)^2}$
SNR (Отношение сигнал/шум) [12]	$\text{SNR} = 10 \cdot \log_{10} \left(\frac{\sum_{t=1}^N \hat{x}_t^2}{\sum_{t=1}^N (X_t - \hat{X}_t)^2} \right)$
NRR (Подавление шума в %)	$\text{NR} = \left(1 - \frac{\text{VAR}(\hat{X}_T)}{\text{VAR}(X_T)} \right) \cdot 100\%$
NRE (коэффициент снижения шума на основе L_1 нормы) [13]	$\begin{aligned} \text{NRE} = \\ = \left(1 - \frac{\sum_{t=1}^N \hat{X}_t - \hat{X}_{t-1} }{\sum_{t=1}^N X_t - X_{t-1} } \right) \cdot 100\% \end{aligned}$

Для обоснования выбора ручного порога на уровне $0.45 \times \text{MAD}$ был проведен пошаговый перебор в диапазоне от 0.10 до 1.00 с шагом в 0.05. При ручном пороге в 0.10 был достигнут максимальный показатель SNR – 45.0157. Расчет метрик проводился между исходным рядом и очищенным заполненным рядом, при этом сравнение происходило только с днями, где были реальные измерения.

Таблица 2

Метрики качества вейвлет-фильтрации сигналов SO_2 при различных конфигурациях

Параметр	RMSE	SNR(dB)	NR	NRE
(1) n-level 3 Универсальный порог Донохо–Джонстона, примененный изолированно к первому уровню детализации (d1)	0.295	5.77	45.02	53.33
(2) n-level 3 сквозной адаптивный ручной порог ($0.45 \times MAD$) для всех уровней детализации (d1-d3)	0.040	25.41	2.24	2.49

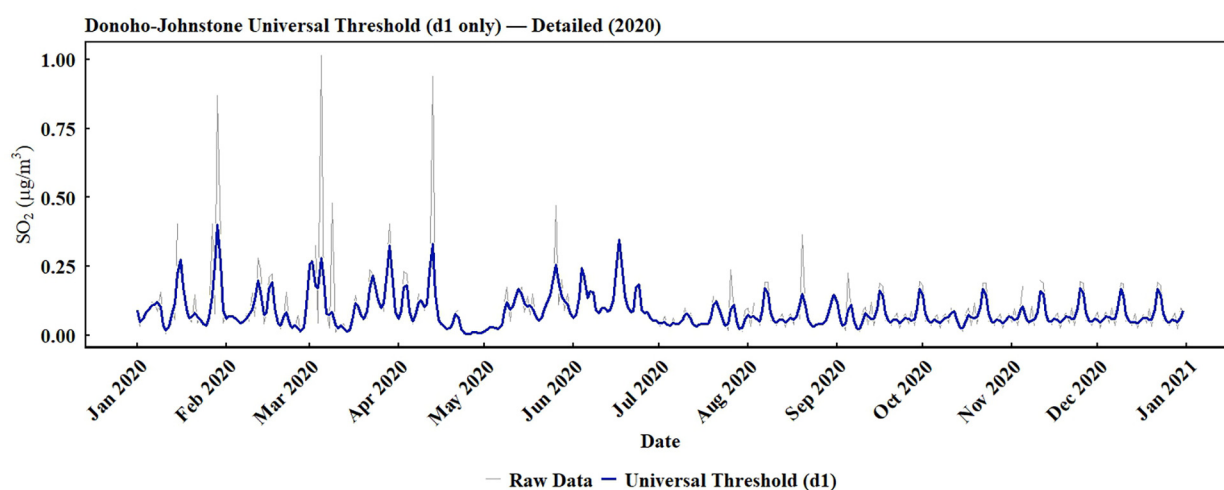


Рис. 1: Результаты вейвлет-фильтрации ряда данных содержания SO_2 в атмосферном воздухе на станции Приокско-Террасного государственного заповедника по программе ЕМЕП за 2020 год методом жесткого порогового ограничения с универсальным порогом.

Результаты вейвлет-фильтрации данных мониторинга SO_2 в атмосферном воздухе на территории Приокско-Террасного государственного заповедника. На рис. 1 и 2 изображен фрагмент оцениваемого ряда данных с заполненными пропущенными значениями (полностью отсутствовали данные за июнь, а также фиксировались эпизодические суточные пропуски). Показано, что заполнение проведено по определенной для данных цикличности, с высокими значениями в зимний период и низкими в летний. Полученные значения не могут интерпретироваться как истинные, так как не представляют собой реальные измерения и являются смоделированными. Тем не менее такой подход позволил проводить вейвлет-фильтрацию данных.

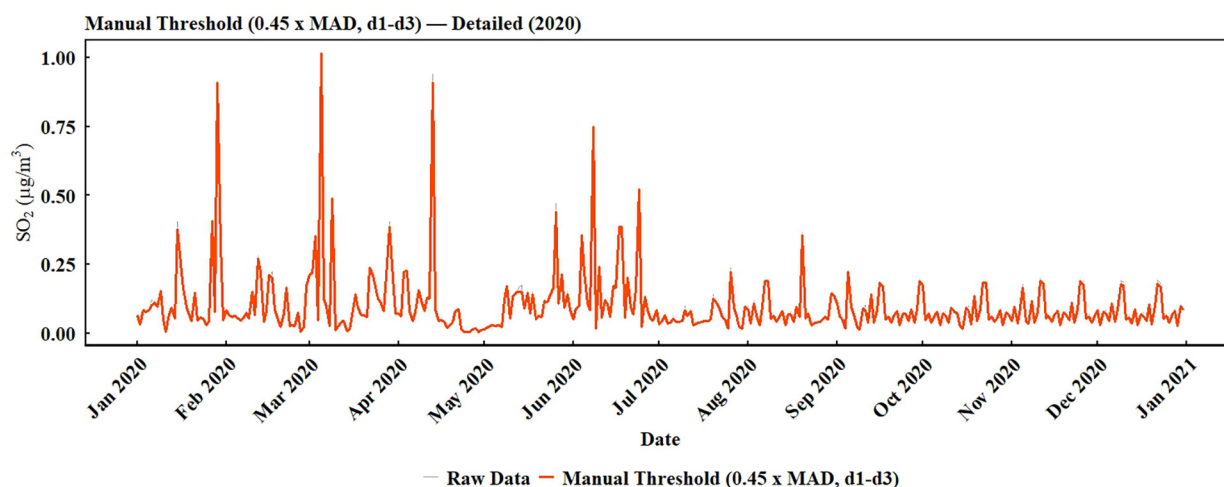


Рис. 2: Результаты вейвлет-фильтрации ряда данных по содержанию SO_2 в атмосферном воздухе на станции Приокско-Тerrasного государственного заповедника по программе ЕМЕП за 2020 год с использованием фиксированного ручного порога.

Результаты удаления шумов методом жесткого порогового ограничения с универсальным порогом (фрагмент за 2020 год) представлены на рис. 1, а с фиксированным ручным порогом на рис. 2. Показано, что применяемые методы позволили сохранить уровни концентраций, но при этом был удален шум, что подтверждается расчетными метриками из табл. 2. Однако при сравнении методов пороговой фильтрации дискретного вейвлет-преобразования MODWT ($n\text{-level}=3$) отмечено преимущество использования фиксированного ручного порога в $0.45 \times \text{MAD}$ перед универсальным порогом Донохо–Джонстана, ограниченного первым уровнем разложения. При ручном пороге показатель SNR достиг значения 25.41 против 5.77 при универсальном пороге. Напротив, RMSE при пороге Донохо–Джонстана в 7 раз выше, чем при ручном пороге $0.45 \times \text{MAD}$. То есть при универсальном пороге Донохо–Джонстана теряется информация о высоких суточных концентрациях SO_2 в атмосферном воздухе. Ручной метод отфильтровал 2.24% избыточной шумовой энергии и сохранил естественную динамику, снизив общую вариативность ряда на 2.49%.

Заключение. В работе показано применение вейвлет-фильтрации на базе вейвлетов типа Добеши 4 к данным программы ЕМЕП – фоновым многолетним суточным концентрациям SO_2 в атмосферном воздухе на территории Приокско-Тerrasного государственного заповедника. Продемонстрирован гибридный способ заполнения отсутствующих значений, учитывающих особенности временного ряда – низкие концентрации на

уровне предела обнаружения аналитического прибора, отсутствующие значения. Опробовано несколько способов удаления шумов из ряда данных. Показана целесообразность применения ручного порога во избежание чрезмерного сглаживания и сохранения полезного сигнала в данных, получаемых на фоновых территориях. Разработанный подход вейвлет-фильтрации предположительно может быть применен для других рядов фонового мониторинга атмосферного воздуха. Полученные временные ряды после вейвлет-фильтрации могут быть использованы для анализа многолетних трендов загрязнений в целях оценки и отслеживания критически важных экологических рисков.

Исследование частично выполнено в рамках темы 4.5. Развитие и модернизация методов и технологий комплексного фонового мониторинга и комплексной оценки состояния и загрязнения окружающей среды РФ ее динамики (по интегрированным результатам сетей мониторинга Росгидромета) (Регистрационный номер 125031703848-8).

Л И Т Е Р А Т У Р А

- [1] Н. Fagerli, E. C. R. Aas, A. Benedictowet, et al., *Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components, EMEP Status Report 1/2025* (Norwegian Meteorological Institute, 2025).
- [2] П. Г. Фрик, Д. Д. Соколов, Р. А. Степанов, УФН **192**, 69 (2022). DOI: 10.3367/UFNr.2020.10.038859.
- [3] А. А. Белов, А. Ю. Проскуряков, Проектирование и технология электронных средств **1**, 27 (2010).
- [4] В. А. Иванов, Е. А. Позднякова, С. А. Громов, Прикладная физика и математика **3**, 32 (2021).
- [5] A. Prakash, U. Kumar, K. Kumar, V. K. Jain, Environmental Modeling & Assessment **16**(5), 503 (2011). DOI: 10.1007/s10666-011-9270-6.
- [6] S. Moritz, T. Bartz-Beielstein, "ImputeTS: Time Series Missing Value Imputation in R" R Journal **9**(1), 207(2017).
- [7] D. J. Stekhoven, P. Bühlmann, Bioinformatics **28**(1), 112 (2012). DOI: 10.1093/bioinformatics/btr597.
- [8] И. Добеши, *Десять лекций по вейвлетам* (Регулярная и хаотическая динамика, Ижевск, 2001).
- [9] D. L. Donoho, I. M. Johnstone, Journal of the American Statistical Association **90**(432), 1200 (1995).
- [10] J. R. Taylor, *An Introduction to Error Analysis* (University Science Books, 1997) ed. 2.

- [11] B. Wieland, K. Urban, S. Funken, *Speech Signal Noise Reduction with Wavelets* (Verlag nichtermittelbar, 2009).
- [12] B. Wang, Q. Jiang, P. Jiang, *Journal of environmental management* **246**, 299 (2019). DOI: 0.1016/j.jenvman.2019.05.074.

Поступила в редакцию 11 августа 2026 г.

После доработки 27 августа 2026 г.

Принята к публикации 10 сентября 2026 г.