

О МЕТОДЕ ВЫДЕЛЕНИЯ ОПТИМАЛЬНЫХ ЧАСТОТНЫХ КОМПОНЕНТ НАБОРОВ ДАННЫХ НА ОСНОВЕ КОСИНУС-ЧАСТОТНЫХ ПРЕДСТАВЛЕНИЙ

chernomorets@bsu.edu.ru

Черноморец А.А., канд. техн. наук, доц.,

Петина М.А., канд. геогр. наук.,

Лебедева М.Г., канд. геогр. наук, доц.,

Болгова Е.В.,

Зайцева Н.О.,

Коваленко А.Н.

*Белгородский государственный национальный
исследовательский университет*

Аннотация. В работе рассмотрены проблемы выделения периодических закономерностей, присущих наборам зарегистрированных данных о загрязненности малых рек региона. Разработан и исследован метод выделения оптимальных частотных компонент наборов данных на основе косинусного преобразования Фурье.

Ключевые слова: периодичность, набор данных, косинусное преобразование Фурье, аддитивная частотная компонента, значимые тенденции

В настоящее время одной из значимых проблем региона является борьба с загрязнением малых рек. В сфере производства актуальной задачей является мониторинг, регистрация и анализ наборов зарегистрированных данных о значениях показателей загрязненности. Неравномерное распределение водных ресурсов во времени и пространстве при одновременном увеличении их потребления, а также возрастающее антропогенное воздействие на естественный водный режим приводит к загрязнению и истощению водных ресурсов, что требует разработки и составления схем рационального использования и охраны водных ресурсов региона, создания постоянно действующей системы контроля их загрязненности [1]. Таким образом, использование современной технологии создания, хранения и обработки наборов данных о гидрологической информации стало актуальной проблемой.

Для принятия эффективных управленческих решений, которые направлены на снижение загрязнения химическими соединениями,

попадающими в реки, необходим эффективный инструмент мониторинга и анализа показателей качества вод и выявления объектов с высоким уровнем загрязнения [2].

В большинстве случаев результаты регистрации уровней загрязненности представляют в виде временных рядов. В работе были проанализированы различные подходы к прогнозированию значений загрязненности водных объектов. Методы прогнозирования временных рядов в большинстве случаев делят на две группы: прогнозирование в частотной области и прогнозирование во временной области. Были проанализированы следующие методы, основанные на построении моделей прогнозирования во временной области: регрессионные модели; авторегрессионные модели; модели экспоненциального сглаживания; модели основанные на искусственных нейронных сетях; модели на базе цепей Маркова; модели на базе классификационно-регрессионных деревьев, а также методы прогнозирования в частотной области: спектральный анализ; вейвлет-анализ; кросс-спектральный анализ и др. [3,4,5].

Проведенный анализ показывает, что одними из наиболее перспективных являются методы, учитывающие периодичность (частоту) появления тех или иных явлений.

Большинство явлений, происходящих в окружающей среде, обладают периодичностью, повторяемостью, в том числе и многие процессы, загрязняющие малые реки региона. Так, естественные процессы, такие как таяние снега в весенний период, сильные ливневые дожди в летний период периодически повышают степень загрязненности рек, величину повышения которой можно проследить в течение нескольких лет. Загрязнения, которые имеют техногенную природу, также могут иметь периодический характер – выбросы загрязняющих веществ в дневной и ночной периоды, сбросы отходов при завершении различных стадий производственной деятельности и др.

Таким образом, можно утверждать, что в окружающей среде существует относительная устойчивость выявленных закономерностей изменения геоэкологического состояния малых рек в течение определенного промежутка времени, выявив которые, имеется возможность построить прогноз о будущих значениях показателей загрязненности на основании анализа их предварительных значений.

Для частотного анализа данных в настоящее время широко используются дискретное преобразование Фурье (ДПФ), дискретное косинусное преобразование Фурье (ДКП), вейвлет-преобразование и др.

[6]. Однако, ДКП имеет ряд преимуществ перед ДПФ в способе представления результатов преобразований и объеме их хранения, так как преобразование Фурье отображает вещественные данные в комплексные, что требует дополнительных вычислительных ресурсов и объемов памяти. ДКП также имеет преимущество в области преобразований (частотной области), так как его коэффициенты имеют высокую концентрацию больших значений в области низких частот.

Таким образом, при анализе периодических закономерностей в наборах зарегистрированных данных о геоэкологическом состоянии малых рек адекватной основой является их описание (математическая модель) на основе косинусного преобразования Фурье.

В случае, когда анализируемый набор данных имеет существенную периодическую закономерность, данный факт проявляется в значительных величинах коэффициентов косинусного преобразования Фурье, соответствующих отдельной частотной подобласти, – набору частот, характеризующих повторяемость данной закономерности. Тем самым, интегральные оценки косинусного преобразования Фурье в заданной частотной области позволяют выделить значимые закономерности, определяющие значения, которые соответствуют анализируемым показателям в будущие периоды времени, что позволяет утверждать об адекватности предлагаемого подхода на основе косинус-частотных представлений задаче анализа и прогнозирования данных о состоянии малых рек.

В большинстве работ по анализу состояния рек используется одномерное представление значений зарегистрированных данных. Во многих случаях, одномерное представление зарегистрированных данных не позволяет представить целостную картинку изменения характеристик состояния реки в зависимости от времени и местоположения проведения измерений [7].

Учитывая потребность регистрации данных о загрязненности рек как в различные моменты времени, так и в различных точках русла реки, в работе предлагается использовать модель анализа двумерных наборов данных о геоэкологическом состоянии малых рек, представленных в виде матрицы $\Phi = (f_{ik})$, $i = 1, 2, \dots, N_1$, $k = 1, 2, \dots, N_2$. Двумерная модель позволяет осуществлять одновременный анализ изменения состояния рек как в пространстве (первая координата матрицы), так и во времени (вторая координата матрицы).

При осуществлении прогнозирования показателей загрязненности малых рек предлагается использовать свойства аддитивных частотных

компонент наборов зарегистрированных данных об их состоянии, соответствующие различным подобластям пространственных частот, получаемых в результате разбиения частотной области, что обеспечит учет сохранения периодичности, повторяемости явлений и факторов, которые определяют показатели загрязненности рек.

Задачи анализа двумерных наборов зарегистрированных данных можно решать, используя разбиение области определения косинус-преобразования Фурье,

$$0 \leq u < \pi, \quad 0 \leq v < \pi,$$

на ряд частотных подобластей $\Delta_{r_1 r_2}$, $r_1 = 1, 2, \dots, R_1$, $r_2 = 1, 2, \dots, R_2$, следующего вида

$$\Delta_{r_1 r_2} = \{(u, v) \mid u_1^{r_1} \leq u < u_2^{r_1}, \quad v_1^{r_2} \leq v < v_2^{r_2}\}, \quad (1)$$

$$u_1^{r_1} = (r_1 - 1) \frac{\pi}{R_1}, \quad u_2^{r_1} = r_1 \frac{\pi}{R_1}, \quad r_1 = 1, 2, \dots, R_1,$$

$$v_1^{r_2} = (r_2 - 1) \frac{\pi}{R_2}, \quad v_2^{r_2} = r_2 \frac{\pi}{R_2}, \quad r_2 = 1, 2, \dots, R_2.$$

Для описания метода выделения оптимальных частотных компонент наборов зарегистрированных данных, соответствующих заданным подобластям частот, введем следующие обозначения. Обозначим, $F_C^\Phi(u, v)$, $F_C^Y(u, v)$ – результаты косинусного преобразования Фурье исходного набора зарегистрированных данных $\Phi = (f_{ik})$, $i = 1, 2, \dots, N_1$, $k = 1, 2, \dots, N_2$, и некоторого набора данных $Y = (y_{ik})$, $i = 1, 2, \dots, N_1$, $k = 1, 2, \dots, N_2$:

$$F_C^\Phi(u, v) = \frac{2}{\pi} \sum_{i=1}^{N_1} \sum_{k=1}^{N_2} f_{ik} \cos(u(i - \frac{1}{2})) \cos(v(k - \frac{1}{2})), \quad (2)$$

$$F_C^Y(u, v) = \frac{2}{\pi} \sum_{i=1}^{N_1} \sum_{k=1}^{N_2} y_{ik} \cos(u(i - \frac{1}{2})) \cos(v(k - \frac{1}{2})).$$

Сформулируем задачу разделения набора зарегистрированных данных Φ на аддитивные компоненты: выделяемая в заданной частотной подобласти $\Delta_{r_1 r_2}$ вида (1) частотная компонента Y должна иметь косинус-преобразование Фурье, которое в «идеальном» случае удовлетворяет условию

$$F_C^Y(u, v) = F_C^\Phi(u, v), \quad (u, v) \in \Delta_{r_1 r_2},$$

$$F_C^Y(u, v) = 0, \quad (u, v) \notin \Delta_{r_1 r_2}.$$

Для оптимального приближения к «идеальному» случаю при решении сформулированной задачи разделения наборов зарегистрированных данных на аддитивные составляющие в работе предложено в качестве критерия оптимальности использовать меру погрешности данного приближения. Так, функционал вида

$$I(Y) = \iint_{(u,v) \in \Delta_{r_1 r_2}} (F_C^\Phi(u,v) - F_C^Y(u,v))^2 dudv + \iint_{(u,v) \notin \Delta_{r_1 r_2}} (F_C^Y(u,v))^2 dudv, \quad (3)$$

может служить в качестве меры отклонения косинусного преобразования выделяемой компоненты Y от косинусного преобразования исходного набора данных Φ в заданной частотной подобласти $\Delta_{r_1 r_2}$ и отклонения от нуля вне этой подобласти.

Задача состоит в нахождении набора данных, минимизирующего функционал $I(Y)$ (3)

$$I(Y) \Rightarrow \min_Y, \quad (4)$$

В работе доказано, что набор данных в виде матрицы $Y_{r_1 r_2}$, являющейся решением оптимизационной задачи (4), определяется следующим соотношением:

$$Y_{r_1 r_2} = G_{r_1} \Phi G_{r_2}. \quad (5)$$

где $\Phi = (f_{ik})$, $i = 1, 2, \dots, N_1$, $k = 1, 2, \dots, N_2$, – матрица исходного набора данных, $G_{r_1} = (g_{i_1 i_2}^{r_1})$, $i_1, i_2 = 1, 2, \dots, N_1$, и $G_{r_2} = (g_{k_1 k_2}^{r_2})$, $k_1, k_2 = 1, 2, \dots, N_2$, – матрицы, соответствующие заданной подобласти частот $\Delta_{r_1 r_2}$ вида (1), значения элементов которых вычисляются на основании следующих соотношений,

$$\begin{aligned} g_{i_1 i_2}^{r_1} &= a_{i_1 i_2}^{r_1} + h_{i_1 i_2}^{r_1}, & (6) \\ g_{k_1 k_2}^{r_2} &= a_{k_1 k_2}^{r_2} + h_{k_1 k_2}^{r_2}, \\ a_{i_1 i_2}^{r_1} &= \begin{cases} \frac{\sin(\sigma_1 r_1 (i_1 - i_2)) - \sin(\sigma_1 (r_1 - 1)(i_1 - i_2))}{\pi(i_1 - i_2)}, & i_1 - i_2 \neq 0, \\ \frac{\sigma_1}{\pi}, & i_1 - i_2 = 0, \end{cases} \\ h_{i_1 i_2}^{r_1} &= \frac{\sin(\sigma_1 r_1 (i_1 + i_2 - 1)) - \sin(\sigma_1 (r_1 - 1)(i_1 + i_2 - 1))}{\pi(i_1 + i_2 - 1)}, \end{aligned}$$

$$a_{k_1 k_2}^{r_2} = \begin{cases} \frac{\sin(\sigma_2 r_2 (k_1 - k_2)) - \sin(\sigma_2 (r_2 - 1)(k_1 - k_2))}{\pi(k_1 - k_2)}, & k_1 - k_2 \neq 0, \\ \frac{\sigma_2}{\pi}, & k_1 - k_2 = 0, \end{cases}$$

$$h_{k_1 k_2}^{r_2} = \frac{\sin(\sigma_2 r_2 (k_1 + k_2 - 1)) - \sin(\sigma_2 (r_2 - 1)(k_1 + k_2 - 1))}{\pi(k_1 + k_2 - 1)},$$

$$\sigma_1 = \frac{\pi}{R_1}, \quad \sigma_2 = \frac{\pi}{R_2}.$$

Соотношение между элементами матриц, используемыми в выражении (6), можно записать в матричном виде:

$$G = A + H.$$

В дальнейшем предлагается матрицу G называть G -субполосной, а матрицу H – квазисубполосной, матрица A – известная субполосная матрица [8].

Следует отметить, что применение разработанного метода выделения оптимальных частотных компонент наборов зарегистрированных данных, соответствующих заданным подобластям частот, позволяет находить частотные компоненты, не вычисляя при этом косинусное преобразование Фурье.

Указанные частотные компоненты полностью определяются результатами косинус-преобразования Фурье в выбранной частотной подобласти. Поэтому корректное выделение этих компонент позволяет адекватно судить об их свойствах в пространственно-временной области. В частности, частотные компоненты могут проявлять характер периодичности вдоль некоторого направления, например, вдоль одной из осей координат: время или пространство.

На основании предлагаемого метода была разработана его программная реализация и проведены вычислительные эксперименты по оцениванию величины энергии компонент, полученных на основе предложенного метода и КИХ-фильтрации [6] в заданных частотных подобластях. На рисунке 1 приведены графики средних значений погрешностей, характеризующие степень растекания энергии за пределы выбранной частотной подобласти в зависимости от разбиения частотной области.

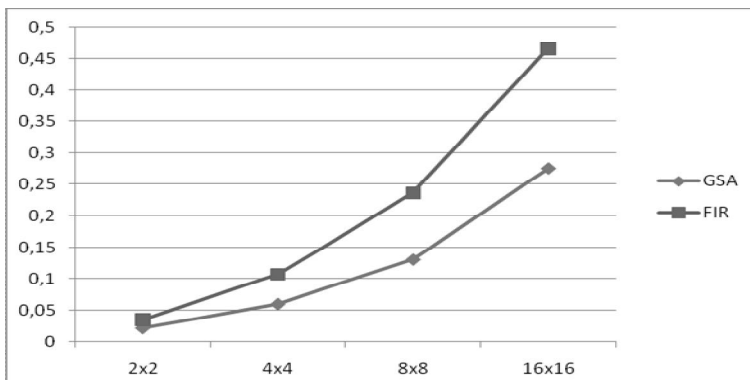


Рисунок 1 – Графики, характеризующие степень растекания энергии за пределы выбранной частотной подобласти в зависимости от разбиения частотной области (*GSA* – разработанный метод, *FIR* – КИХ-фильтрация)

Отметим, что использование КИХ-фильтров, в отличие от применения разработанного метода, не позволяет получить частотные компоненты, которые не зависят от поведения косинус-преобразования Фурье наборов данных в подобластях частот, смежных с заданной частотной подобластью. Причиной этому служит наличие переходных полос в их амплитудно-частотных характеристиках, в пределах которых воздействие соседних частотных подобластей подавляется не полностью. Поэтому при использовании КИХ-фильтров возможны искажения в частотных компонентах, которые могут быть существенными, если в смежных частотных подобластях сосредоточена значительная доля энергии исходного набора зарегистрированных данных, особенно когда она превосходит долю энергии в выбранной частотной подобласти. Данные искажения не возникают при применении разработанного метода.

Результаты вычислительных экспериментов показали, что разработанный метод выделения частотных компонент наборов зарегистрированных данных обладает преимуществами в точности аппроксимации квазиэнергий выбранных наборов данных в пределах выбранной частотной подобласти перед широко используемой в настоящее время КИХ-фильтрацией. Эффективность предложенного метода обусловлена решением лежащей в его основе оптимизационной задачи минимизации евклидовой нормы погрешности аппроксимации значений исходного косинусного преобразования Фурье в заданной частотной подобласти и нуля – вне данной подобласти.

Разработанный метод выделения оптимальных частотных компонент наборов зарегистрированных данных о загрязненности малых рек на основе косинус-частотных представлений позволяет адекватно учесть частотные свойства анализируемых данных, что имеет существенное значение для разработки алгоритмов прогнозирования, учитывающих значимые зависимости (тенденции), присущие зарегистрированным данным.

Исследования выполнены при финансовой поддержке РФФИ и Правительства Белгородской области в рамках проекта №14-47-08052 «р_офи_м».

Список литературы:

1. Руководство по гидробиологическому мониторингу пресноводных экосистем: под ред. В. А. Абакумова. СПб.: Гидрометеиздат, 1992. 318 с.
2. Гидрохимические показатели состояния окружающей среды: справочные материалы / Т. В. Гусева [и др.]. СПб.: Эколайн, 2000. 127 с.
3. Афанасьев В.Н., Юзбашев М.М. Анализ временных рядов и прогнозирование: учебник. М.: Финансы и статистика, ИНФРА-М, 2010. 320 с.
4. Хайкин С. Нейронные сети: полный курс. М.: ООО «И. Д. Вильямс», 2006. 1104 с.
5. Малоземов В. Н., Машарский С. М. Основы дискретного гармонического анализа. СПб.: НИИММ, 2003. 288 с.
6. Сергиенко А. Б. Цифровая обработка сигналов: учеб. пособие для студ. вузов. СПб.: Питер. 2002. 603 с.
7. Анализ и оценка качества поверхностных вод: учебное пособие / А.Н. Петин, М.Г. Лебедева, О.В. Крымская. Белгород: БелГУ, 2006. 252 с.
8. Жилияков Е.Г., Белов С.П., Черноморец А.А. Вариационные методы анализа сигналов на основе частотных представлений // Вопросы радиоэлектроники. 2010. Сер. ЭВТ. Вып. 1. С. 10-25.