

Московский государственный университет имени М. В. Ломоносова



Факультет Вычислительной Математики и Кибернетики
Кафедра Математических Методов Прогнозирования

ДИПЛОМНАЯ РАБОТА

**«Анализ пространственно-временного распределения
хлорофилла в Азовском море по данным дистанционного
зондирования Земли»**

Выполнила:

студентка 5 курса 517 группы

Огнева Дарья Сергеевна

Научный руководитель:

д.т.н., профессор

Местецкий Леонид Моисеевич

Москва, 2014

Оглавление

1	Введение	3
2	Постановка задачи	4
3	Обзор аналогов	5
4	Описание входных данных	7
4.1	Коллекция снимков.	7
4.2	Один снимок.	7
5	Визуализация данных	8
5.1	Зоны видимости	9
5.2	Пространственное распределение	9
5.3	Временное распределение	9
5.4	Выводы	17
6	Оценка адекватности данных	18
6.1	Гипотеза пространственно-временной компактности	19
6.2	Временная компактность	19
6.3	Пространственная компактность	23
7	Поиск нелинейного тренда	26
7.1	Локальное решение	27
7.2	Качественный анализ	29
7.3	Количественный анализ	31
8	Заключение	35

Аннотация

В работе рассматривается задача поиска нелинейного тренда в пространственно-временном распределении концентрации хлорофилла по данным дистанционного зондирования Земли.

Анализ данных происходит в три глобальных этапа:

1. описание входных данных и их визуализация
2. проверка на адекватность
3. аппроксимация универсальным трендом

Описание входных данных и визуализация дают общее представление о распределении данных и интуитивное понимание закономерностей. Более того, частота обсервации, введенная в этой части, дает дополнительные сведения о погоде над регионом интереса — Азовским морем.

В проверке на адекватность по аналогии с метрическими алгоритмами классификации сформулирована гипотеза компактности по пространству и времени. Впоследствии эта гипотеза выступает в качестве критерия адекватности данных.

Нелинейным трендом рассматривается многочлен пятой степени, определенный на периоде времени в один год с марта по ноябрь, находятся коэффициенты данного многочлена, обосновывается его универсальность. Результатом аппроксимации данных трендом является опровержение утверждения ученых-океанологов о виде годового распределения концентрации хлорофилла.

1 Введение

Анализ данных является наиболее востребованной в век информации областью математики и информатики. Обладая общими методами и инструментами, он позволяет абстрагироваться от узкой специальности данных, будь то физика и биология или психология и социология, и сделать выводы, основываясь только на полученных данных и некоторых ограничениях, сделанных в рамках модели предметной области.

Из предположения существования вероятностной модели в основе данных анализ сводится к прикладной статистике, поэтому существенным моментом является механизм порождения данных.

Базой для анализа в данной работе служит набор полей концентрации в Азовском море хлорофилла «а», впоследствии называемого просто хлорофиллом, - активного фотосинтезирующего пигмента в фитопланктоне. Данные поля являются результатом применения некоторой эвристической формулы к снимкам с космических аппаратов дистанционного зондирования Земли. Таким образом, распределение концентрации варьируется в зависимости от двух параметров: пространства (поверхность моря) и времени снимка. Рассмотрение аналогичных пространственно-временных распределений имеет особое значение в биологии и географии (океанологии, сейсмологии).

Само распределение концентрации хлорофилла является важной компонентой в океанологии, т. к. показатель его уровня является основным в оценке биопродуктивности морских акваторий, а значит, и контроля экологического состояния. Поэтому активно совершенствуются способы и алгоритмы мониторинга и анализа данной компоненты ([1]–[14]).

Существует значительное количество статей и диссертаций по анализу концентрации хлорофилла, более того, такие институты и университеты, как Институт океанологии имени П.П. Ширшова РАН, Институт космических исследований РАН, Московский государственный университет имени М.В. Ломоносова, Южный Федеральный Университет, активно развивают данное научное направление. Но возможностей развития существенно больше, они активно расширяются в связи с переходом от судовых замеров концентрации к автоматизированному мониторингу посредством дистанционного зондирования Земли и

необходимостью качественно улучшать алгоритмы, вычисляющие значение концентрации.

Данная работа начинается с постановки задачи анализа распределения концентрации хлорофилла, затем коротко описывается релевантная литература. В последующих главах последовательно решается поставленная задача: описываются и визуализируются данные, оцениваются на адекватность и аппроксимируются нелинейным трендом.

2 Постановка задачи

Изначально данные по количеству хлорофилла собирались с помощью судовых измерений: организовывалась экспедиция, которая выходила в благоприятную погоду в море и брала несколько проб морской воды.

С освоением космического пространства и развитием технологий появилась возможность получать замеры не только с меньшими временными и ресурсными затратами, но и по большей части моря взамен ограниченного числа точек. Данная альтернатива вычисляет значение полей концентрации по снимку, полученному в результате дистанционного зондирования Земли. Этот способ только развивается, поэтому имеет неточные алгоритмы преобразования спектра снимка в распределение значений концентрации.

Задача анализа пространственно–временного распределения хлорофилла была предложена доктором географических наук Бердниковым Сергеем Владимировичем.

Она сводилась к двум ключевым подзадачам:

- оценка адекватности входных данных, т. к. формулы для вычисления значения концентрации хлорофилла по снимкам являются эвристическими;
- поиск нелинейного тренда во временном распределении, подтверждающего или опровергающего теоретическое мнение экспертов-океанологов:

распределение концентрации хлорофилла в связи с особенностью биологической активности в разные периоды годового цикла имеет два максимума — весенний (глобальный) и осенний (локальный), и общий вид рис. 1.

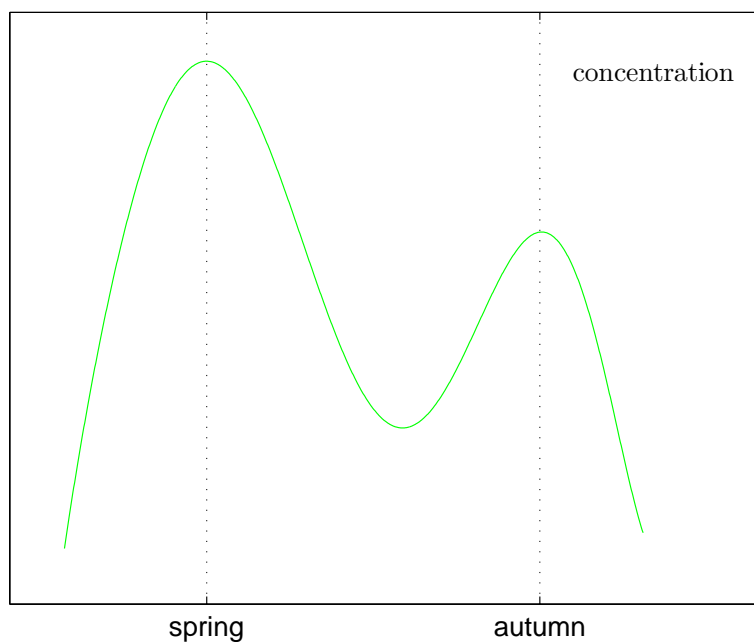


Рис. 1: Вид годового распределения хлорофилла с точки зрения ученых-океанологов.

3 Обзор аналогов

Первая подзадача оценки адекватности данных не является оригинальной: появление алгоритмов вычисления полей концентрации естественным образом повлекло необходимость верификации получаемых данных и совершенствования методов. В результате написано большое количество работ, в которых сравнительный анализ производился с использованием результатов судовых измерений как эталонных.

Например, в работе [1] оценивается погрешность спектрометрического метода, основанного на инфракрасном излучении, или в работе [2] сравнивается качество данных по распределению хлорофилла для двух видов спектрометров. Подобным образом опорой для верификации данных могут служить и методы применимые на акваториях других морей: применимость модифицированного трехполосного алгоритма на западном побережье Флориды [3], изучение и коррекция спутниковых оценок концентрации хлорофилла для Эгейского [4], Японского и Охотского морей [5]. Более того, существуют работы посвященные непосредственно тематике верификации: в [6] по натурным измерениям в оптически сложных водах Азовского моря оцениваются алгоритмы подсчета

концентрация хлорофилла, использующие данные спутниковых сканеров цвета океана; в [7] анализируются основные погрешности, связанные с неточностью задания моделей, и систематические искажения.

Вторая подзадача также многогранно рассматривается в публикациях. Для ее решения используется множество различных техник статистического анализа и преобразования изображений:

- генетические алгоритмы, моделирующие фитопланктон на основе разностного уравнения [8],
- анализ временных рядов прибрежных районов под влияние речных оттоков [9],
- нелинейный анализ [10], разработанный для дифференциации хаотичной динамики объектов и внешнего шума во временных последовательностях, рассмотренный на примере Южного океана,
- мультифрактальный анализ океанических карт хлорофилла [11],
- Фурье-анализ и вейвлет-спектры пространственных и временных сечений распределения хлорофилла в поверхностном слое Черного моря по данным дистанционного зондирования [12],
- модели [13], учитывающие распределения хлорофилла по глубине на основе уравнения «реакция-диффузия»,
- статистический имитационный метод для оценки зависящих от времени параметров в трехмерной модели океана [14].

К вышеизложенным, адаптированным под поля концентрации хлорофилла методам можно добавить методы анализа временных рядов с помощью модели экспоненциального сглаживания и авторегрессионной модели процесса Бокса и Дженкинса [15] и скрытых марковских моделей [16].

Таким образом,

- в публикациях освещено множество способов и методик решения как первой, так и второй подзадачи,
- оценки по качеству получаемых спутниками карт хлорофилла являются удовлетворительными и улучшаются за счет совершенствования алгоритмов,

- каждый адаптированный для полей хлорофилла способ анализа является самодостаточным и не нуждающимся в значительной доработке, потому данная часть задачи требует нового подхода,
- отличительными аспектами нашей задачи является отсутствие судовых замеров и проверка мнения ученых-океанологов.

4 Описание входных данных

4.1 Коллекция снимков.

Исходные данные представлены 984 изображениями в формате tif. Первые 695 изображений сгенерированы центром УК-РАС, остальные 289 — центром Е-РАС. Дата съемки первого подмножества коллекции лежит в диапазоне с 29.05.2003 по 25.01.2012, второго — с 30.04.2003 по 13.06.2008. Время замеров на всех кадрах приходится на промежуток после 7:30 и до 8:45 утра. Интервалы между замерами составляют от 0 до 63 суток, в среднем — 3.25 суток.

4.2 Один снимок.

Название изображения. Прототип имени файла:

MER_FR__2PNEPA20051229_075547_000000982043_00436_20029_1764.tif

MER_XXX_YZpGGGyyyyymmdd_HHMMSS_tttttttPccc_OOOOO_aaaaa_QQQQ.tif

1. *MER_XXX_YZ* — название продукта MERIS (в нашем случае это *MER_FR__2P* — Full Resolution Geophysical Product for Ocean, Land);
2. *p* — флаг его стадии обработки (в нашем случае это *N* — Near Real Time product);
3. *GGG* — идентификатор центра, сгенерировавшего этот файл (два варианта ЕРА (Е-РАС) и УРА (УК-РАС));
4. *yyyyymmdd* — дата начала съемки изображения в формате (год месяц день);
5. *HHMMSS* — время начала съемки изображения в формате (час минуты секунды);
6. *ttttttt* — продолжительность съемки (в секундах);

7. P — identifies the phase of the mission;
8. ccc — is the number of the cycle in the mission phase;
9. OOOOO — относительный номер орбиты внутри цикла (cycle);
10. aaaaa — абсолютный номер орбиты;
11. QQQQ — счетчик для быстрой идентификации файла. Для каждого типа продукта увеличивается на 1 при подготовке нового экземпляра продукта.

Изображение. Снимок имеет размеры 1193×802 . Каждый пиксель изображения содержит неотрицательное значение концентрации хлорофилла либо -9999, если точка не содержится в Азовском море или по ней нет данных.

Примеры полученных файлов представлены на рис.2. Черным цветом на данных изображениях выделены пиксели, в которых значение концентрации -9999, остальные оттенки серого от черного до белого равномерно отображают возрастание концентрации хлорофилла от 0 до максимального значения соответственно.

5 Визуализация данных

Перед непосредственным решением сформулированных подзадач визуализируем имеющиеся данные.

Рассмотрим пространственно-временные колебания концентрации хлорофилла как случайный процесс с дискретными состояниями. Обозначим его за $\psi(X, t)$, где $X = (x, y) \in \mathbb{N}^2, x = 1, \dots, 1193, y = 1, \dots, 802$ — координаты пикселя, $t \in \mathbb{N}$ — время, число дней от 1 января 2003 года до даты снимка.

Визуализация компонент реализации случайного процесса в формате исходных данных $\psi(X, t)$ (рис. 3, 5, 7, 9, 12–19) выявляет выраженные пики концентрации, не сочетающиеся с окружающими значениями и не оправданные идеологическими соображениями, поэтому сгладим все полученные распределения для устранения шума. В пространственном сглаживании используем медианный фильтр с шириной окна 11. Временное сглаживание представим комбинацией двух фильтров. Первый является средним арифметическим с шириной окна 7 по каждому снимку. Более того, в этом случае сглаживания считаем пиксель неопределенным, если $\frac{2}{7}$ его окрестности неопределены. Второй тип

фильтра вычисляется для каждого момента времени по 8 окружающим моментам (9 считая с текущим) как центр тяжести с массами–концентрациями и радиус-векторами–моментами времени.

Рассмотрим визуализированные компоненты подробнее.

5.1 Зоны видимости

В связи с погодными условиями (облачность, помутнение воды, вызванное штормом и др.) концентрация хлорофилла была доступна только на части акватории Азовского моря, поэтому каждое изображение дает лишь фрагментарное представление о данном распределении.

Рассмотрим функцию частоты обсервации каждого пикселя (рис.3–4). Данная функция зависит от двух параметров — координаты вдоль длины изображения (оси абсцисс) и координаты вдоль ширины (оси ординат). Значением функции является количество снимков из 984, в которых указанный пиксель имеет неотрицательное значение концентрации.

Данная функция имеет смысл частоты реализации ясной погоды в определенном участке акватории моря, соответствующем некоторому пикселю.

5.2 Пространственное распределение

Визуализируем несколько характеристик рассматриваемого случайного процесса.

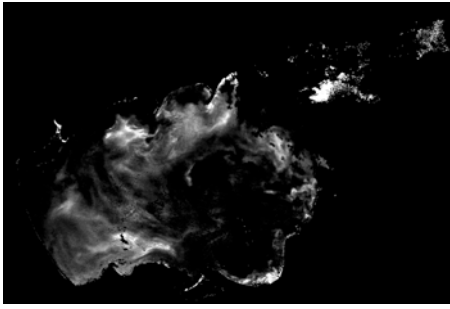
Рассмотренные центральные выборочные моменты(рис.5–8) и функции от них(рис.9–10) представляют собой некоторые пространственные распределения концентрации хлорофилла:

1. среднее значение (рис.5-6),
2. разброс (рис.7-8),
3. стандартное отклонение (рис.9-10)

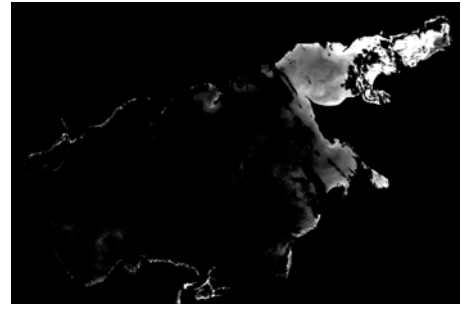
концентрации хлорофилла по коллекции изображений.

5.3 Временное распределение

Продолжим визуализировать компоненты случайного процесса $\psi(X, t)$.



Изображение № 1.



Изображение № 2.

Рис. 2: Примеры

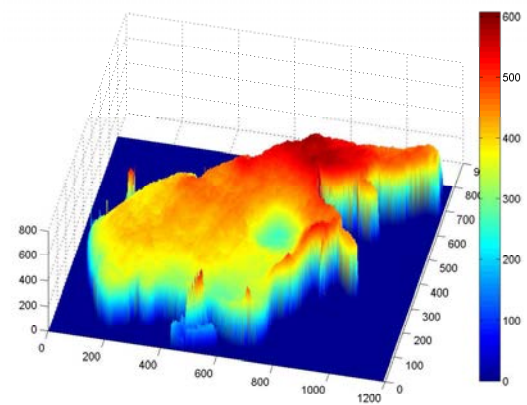
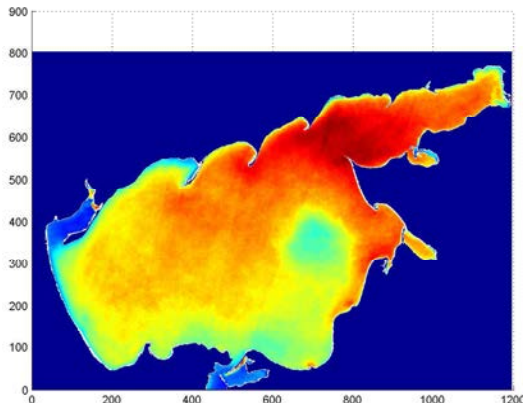


Рис. 3: Частота наблюдения пикселя: исходные данные.

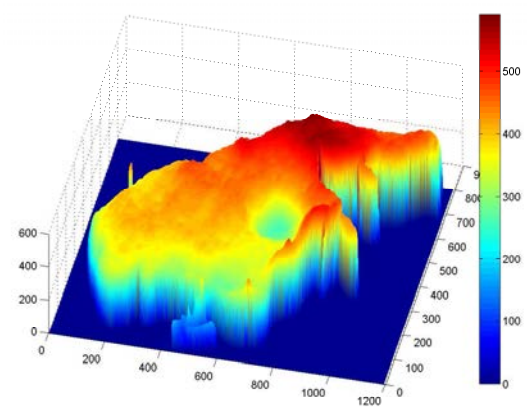
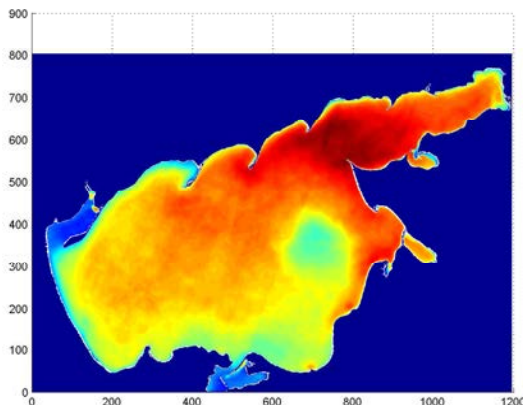


Рис. 4: Частота наблюдения пикселя: сглаженные данные.

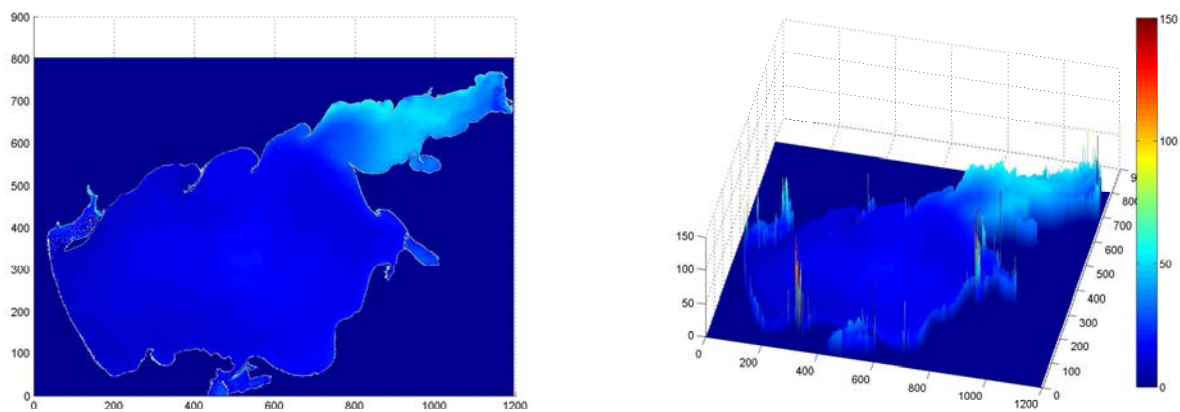


Рис. 5: $\mathbb{E}_t \psi(X, t) : X$ —исходные данные.

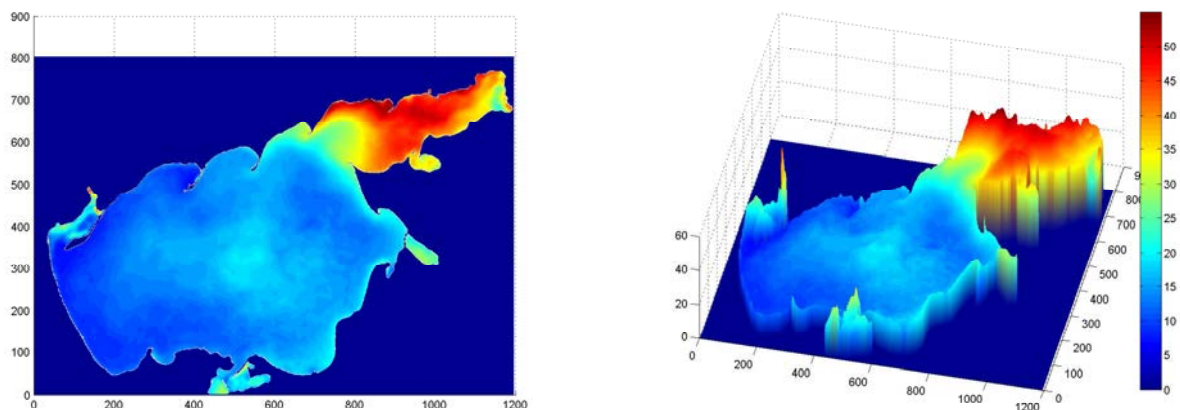


Рис. 6: $\mathbb{E}_t \psi(X, t) : X$ —сглаженные данные.

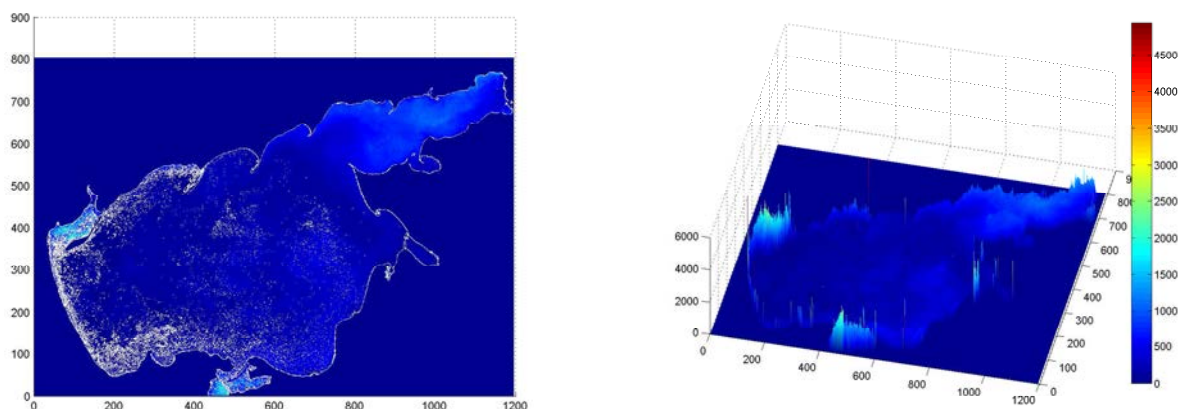


Рис. 7: $\mathbb{D}_t \psi(X, t) : X$ —исходные данные.

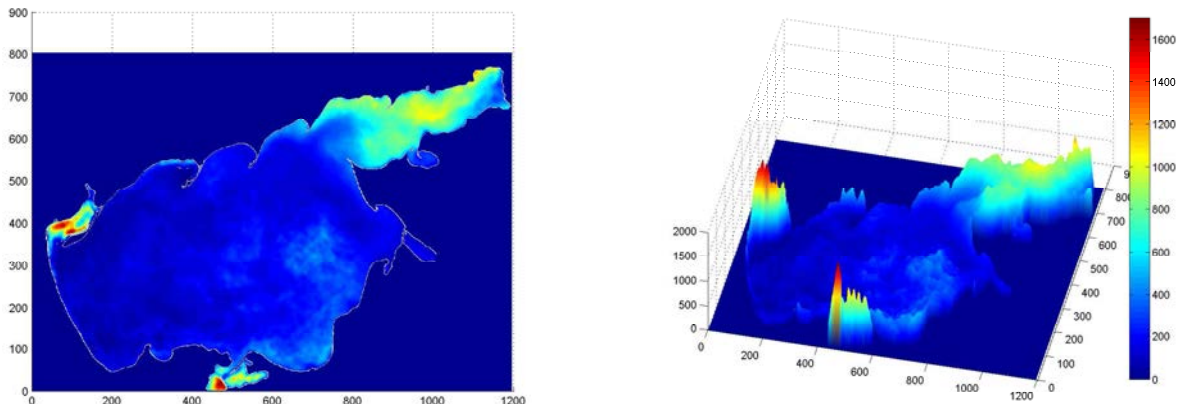


Рис. 8: $\mathbb{D}_t\psi(X, t)$: X — сглаженные данные.

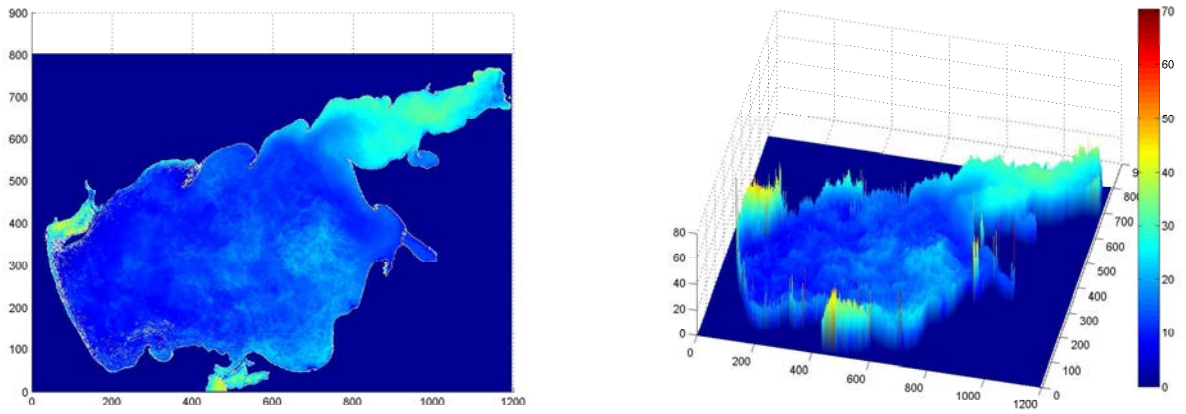


Рис. 9: $\sqrt{\mathbb{D}_t\psi(X, t)}$: X —исходные данные.

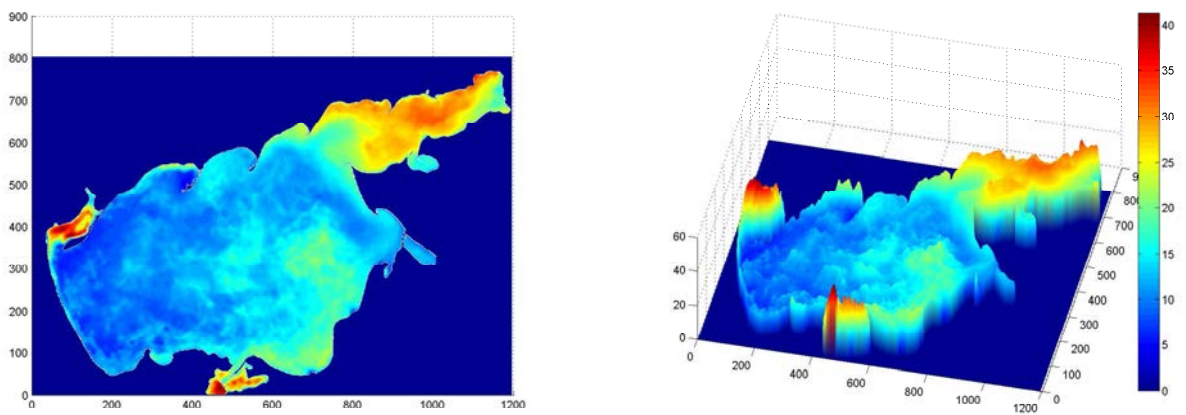


Рис. 10: $\sqrt{\mathbb{D}_t\psi(X, t)}$: X — сглаженные данные.

Рассмотрим несколько сечений случайного процесса по переменной X : зафиксируем 8 точек из областей акватории Азовского моря (рис.11), максимально полно представляющих разнообразие регионов по уже полученным сведениям о пространственном распределении и его характеристиках.

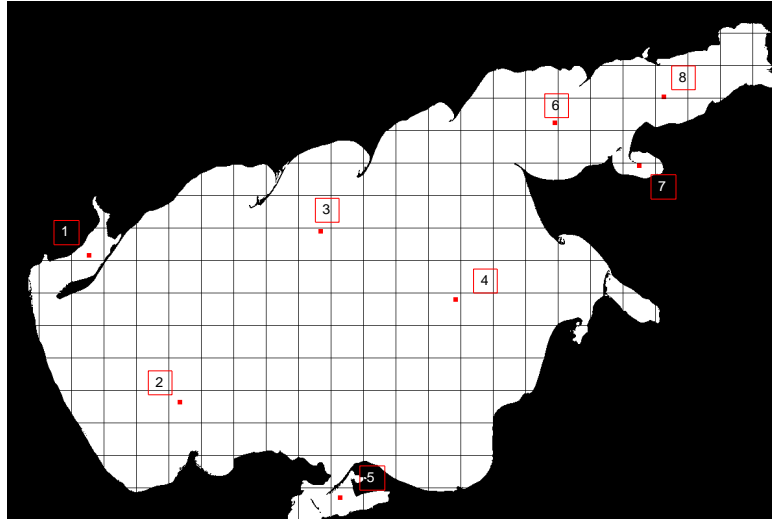


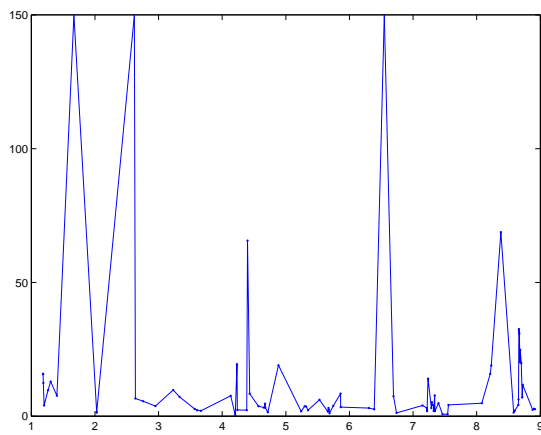
Рис. 11: Выделение и нумерация выбранных точек(красным) на акватории Азовского моря с наложенной решеткой (черной), шаг решетки 50 пикселей.

Координаты точек в регионах интереса структурированы в виде таблицы:

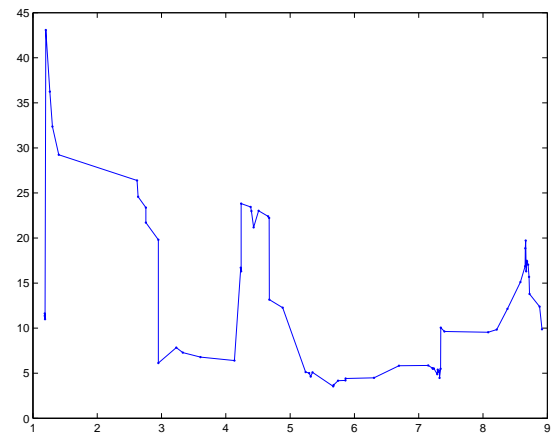
точка	x	y
X_1	127	392
X_2	267	618
X_3	484	355
X_4	692	460
X_5	514	765
X_6	845	188
X_7	1013	148
X_8	975	254

Построим графики концентрации хлорофилла в выбранных точках(рис.12–19).

По оси абсцисс отложено число лет с 2003 года, по оси ординат — значение концентрации в рассматриваемой точке. График представляет собой функцию концентрации, построенную по известным t .

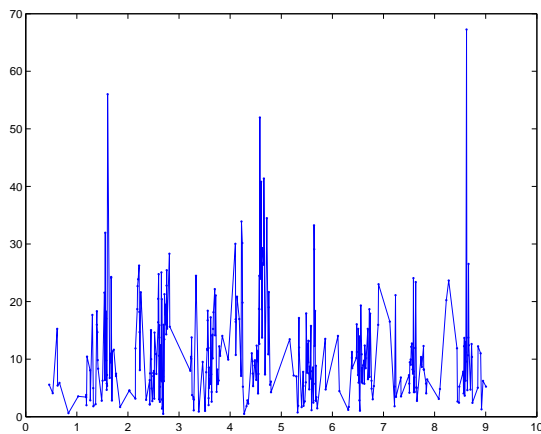


исходные данные

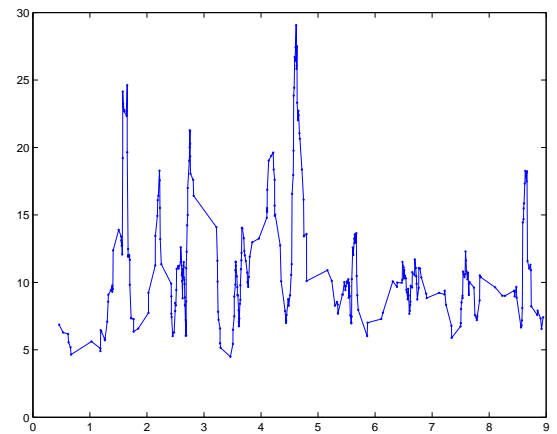


сглаженные данные

Рис. 12: График $\psi(X_1, t)$

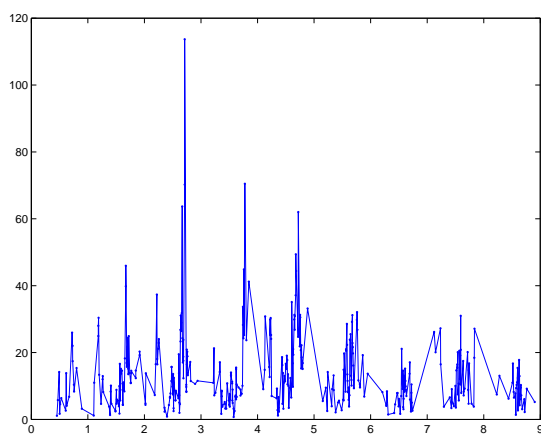


исходные данные

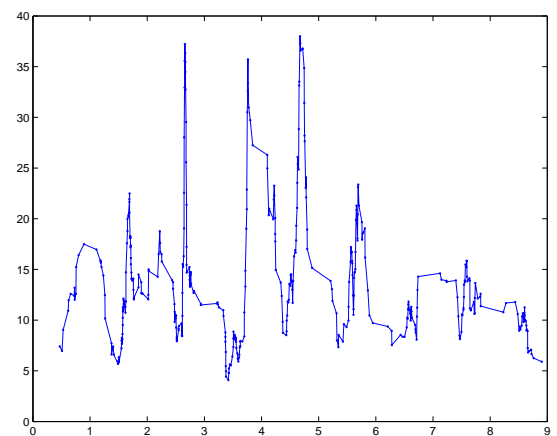


сглаженные данные

Рис. 13: График $\psi(X_2, t)$

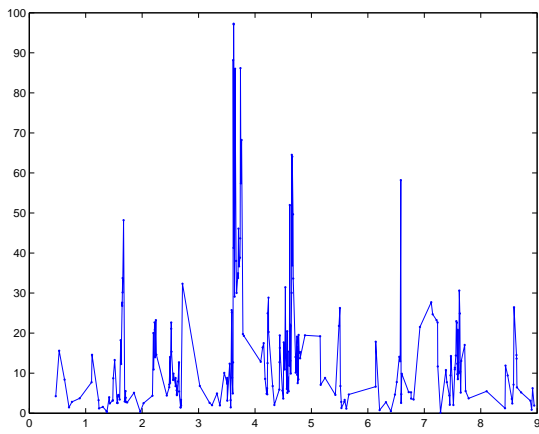


исходные данные

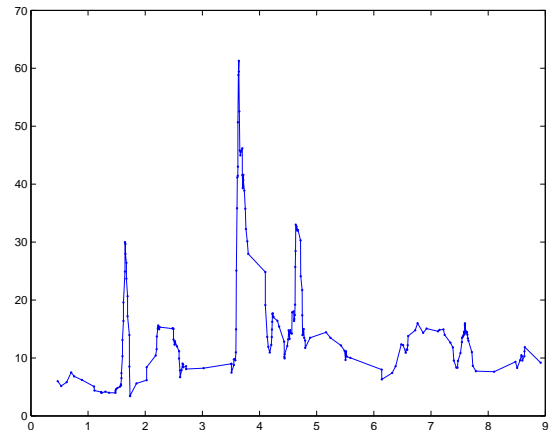


сглаженные данные

Рис. 14: График $\psi(X_3, t)$

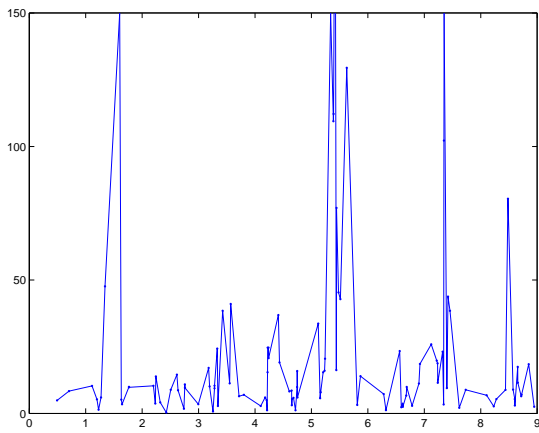


исходные данные

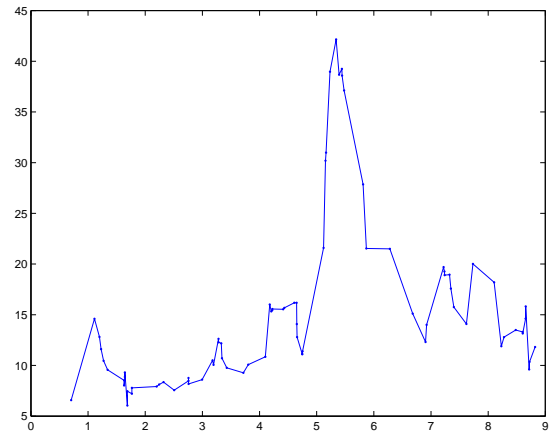


сглаженные данные

Рис. 15: График $\psi(X_4, t)$

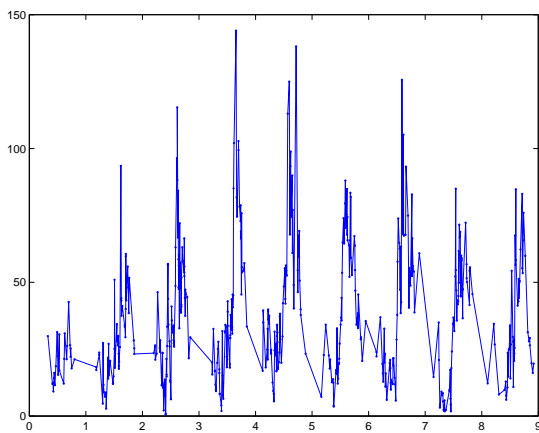


исходные данные

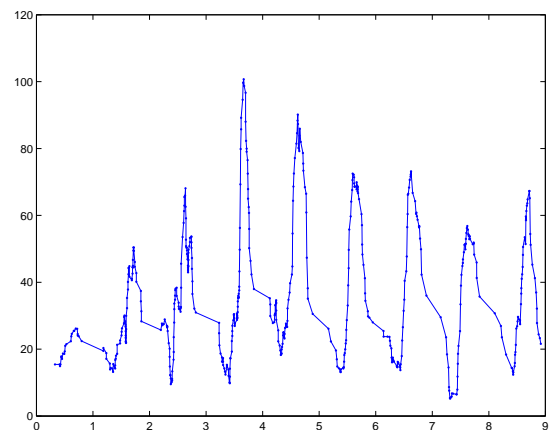


сглаженные данные

Рис. 16: График $\psi(X_5, t)$

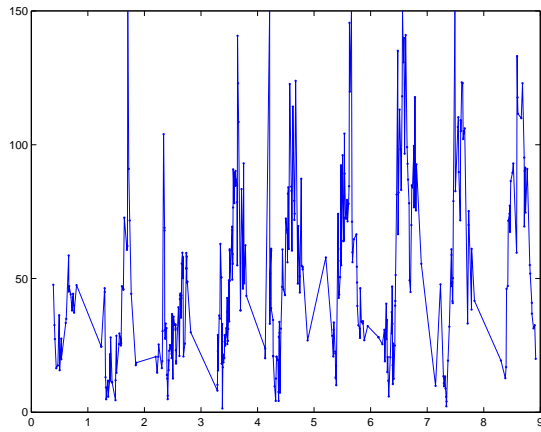


исходные данные

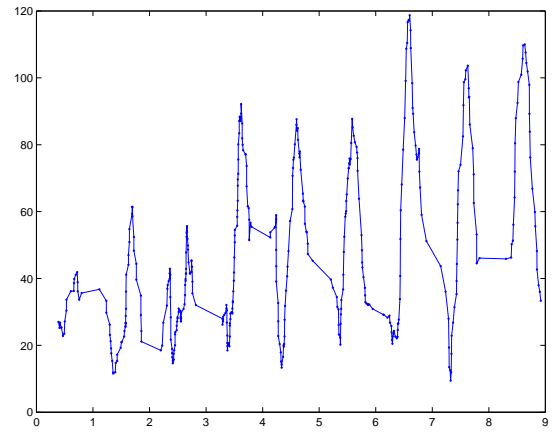


сглаженные данные

Рис. 17: График $\psi(X_6, t)$

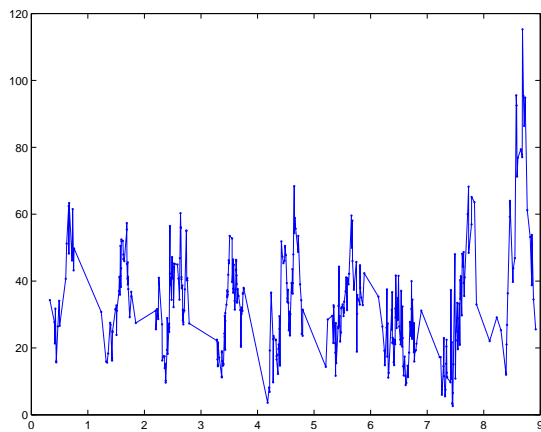


исходные данные

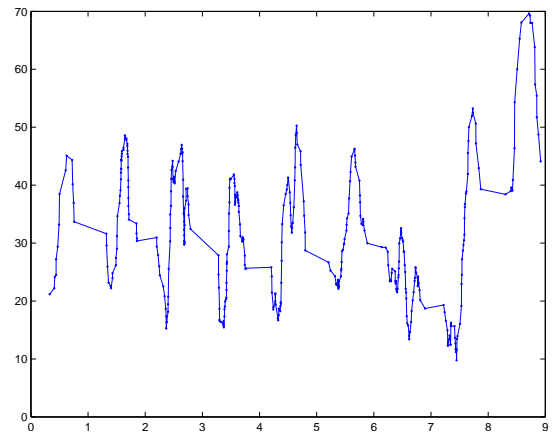


сглаженные данные

Рис. 18: График $\psi(X_7, t)$



исходные данные



сглаженные данные

Рис. 19: График $\psi(X_8, t)$

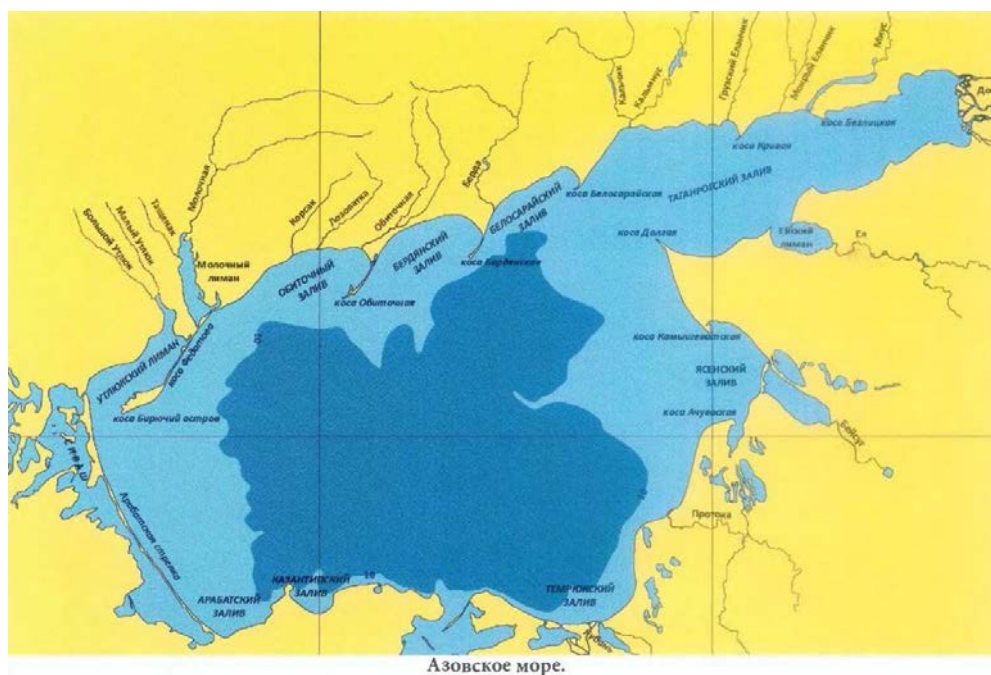


Рис. 20: Карта Азовского моря.

5.4 Выводы

1. Всю акваторию Азовского моря можно кластеризовать в зависимости от пространственных характеристик на Таганрогский залив, северный + центральный + западно-центральный + крайне восточный район, Темрюкский залив + крайне южный + восточно-центральный район моря, Бейсугский лиман, Утлюкский лиман + Керченский пролив (рис.20).
2. В Таганрогском заливе, в северной и крайне восточной части преобладает безоблачная погода (частота наблюдения имеет большие показатели). Про оставшуюся часть акватории нельзя утверждать точно, т.к. не только облака влияют на низкие показатели концентрации хлорофилла.
3. Наибольшие средние показатели концентрации хлорофилла также реализуются в Таганрогском заливе, на фоне почти нулевых значений на большей акватории Азова выделяются более высокими значениями центр и прибрежные районы, особенно Бейсугский лиман, Утлюкский лиман, Керченский пролив.

4. Максимальный разброс концентрации виден в районах Таганрогского залива, Утлюкского лимана и Керченского пролива, минимальный — на западной части Азовского моря.
5. Вертикальные линии во временных графиках показывают, что пики концентрации являются неточностью программы по вычислению концентрации по снимку, или выбросами, т.к. в один день в промежуток времени не более часа значение концентрации в одной точке существенно отличаться не может. Но реализованным фильтром все шумы не устраняются.
6. К сожалению, есть кластеры (Утлюкский лиман + Керченский пролив), количество данных по которым не позволяет установить факт существования закономерности (точки №1 и №5 соответственно).
7. В Таганрогском заливе (точки №6, №7 и №8) колебания концентрации имеют годичный период — рост с апреля по июль и спад с августа по февраль.
8. Временные графики с достаточным количеством информации на остальной территории не позволяют точно утверждать о существовании аналогичной закономерности.

Выводы показывают, что есть некоторая закономерность в распределении рассматриваемой случайной величины. Что позволяет продолжить ее исследование, в частности — изучение свойств и проверку гипотез.

Более того, периодичность во временных распределениях позволяет изменить периодизацию и рассматривать пространственно-временные закономерности посезонно.

6 Оценка адекватности данных

Существенным отличием перечисленных в обзоре работ [1]–[7] от данной является наличие только карт хлорофилла, сгенерированных спутником. Таким образом, оценка не может опираться на значение функционала ошибки от значений судовых измерений.

6.1 Гипотеза пространственно-временной компактности

При визуализации данных мы пользовались разумным предположением об ограниченной разности значений концентрации в соседних пикселях и в соседние моменты времени и классифицировали существенные отклонения как выбросы.

Другими словами, мы априорно предполагаем верность гипотезы пространственной и временной компактности:

- концентрация в смежных пикселях и/или в подряд идущие дни, как правило, должна иметь схожие значения.

Исходя из сформулированной гипотезы оценим адекватность данных.

6.2 Временная компактность

Зафиксируем произвольную точку \tilde{X} акватории Азовского моря.

Пусть $T = \{t_0, \dots, t_n\} \subset \mathbb{N}$ — подмножество значений второго аргумента $\psi(X, t)$, тогда каждое t_i определяет окрестность радиусом δ и множество моментов времени в этой окрестности $T_\delta(i) = \{t \in T, |t - t_i| = \delta\}$. Фиксируем первый параметр распределения концентрации хлорофилла и переобозначим $\psi(\tilde{X}, t) \rightarrow K(t)$. Таким образом, $K(t)$ — временной ряд, зависящий от зафиксированной точки пространства, $K_i = K(t_i)$.

Рассмотрим всевозможные разности $\Delta_a K(i, j) = |K_i - K_j|, i \neq j$. Т. к. область значений $\Delta_a K(i, j)$ вещественна, то дискретизируем ее путем выбора равномерного разбиения от минимального c_0^a до максимального c_k^a значения этой функции с шагом ε . Пусть $\{c^a(\varepsilon)\} = \{c_0^a, c_1^a = c_0^a + \varepsilon, c_2^a = c_1^a + \varepsilon = c_0^a + 2\varepsilon, \dots, c_k^a = c_{k-1}^a + \varepsilon = c_0^a + k\varepsilon\}$ — рассматриваемое разбиение, тогда число его точек определяется из формулы $k = \frac{c_k^a - c_0^a}{\varepsilon}$.

Введем функцию

$$\nu_a(\delta, c_i^a(\varepsilon)) = \left\| (t_i, t_j) : t_j \in T_\delta(i), \Delta_a K(i, j) \in [c_{i-1}^a, c_i^a], \forall \tilde{X} \right\|$$

— мощность множества пар моментов времени, разность между которыми равна δ , а разность значений концентрации в которых лежит в промежутке $[c_{i-1}^a, c_i^a]$.

Аналогично рассмотрим всевозможные относительные разности $\Delta_r K(i, j) = \frac{|K_i - K_j|}{\max(K_i, K_j)}$, $K_i \neq 0, K_j \neq 0, i \neq j$. Данное определение корректно, т. к. если значение концентрации хлорофилла в фиксированной точке акватории в заданный момент времени определено, то оно отлично от 0. Область значений $\Delta_r K(i, j)$ вещественна, поэтому дискретизируем ее путем выбора равномерного разбиения от минимального c_0^r до максимального c_q^r значения этой функции с шагом ε . Пусть $\{c^r(\varepsilon)\} = \{c_0^r, c_1^r = c_0^r + \varepsilon, c_2^r = c_1^r + \varepsilon = c_0^r + 2\varepsilon, \dots, c_q^r = c_{q-1}^r + \varepsilon = c_0^r + q\varepsilon\}$ — рассматриваемое разбиение, тогда число его точек определяется из формулы $q = \frac{c_q^r - c_0^r}{\varepsilon}$.

Введем функцию

$$\nu_r(\delta, c_d^r(\varepsilon)) = \|(t_i, t_j) : t_j \in T_\delta(i), \Delta_r K(i, j) \in [c_{d-1}^r, c_d^r], \forall \tilde{X}\|$$

— мощность множества пар моментов времени, разность между которыми равна δ , а относительная разность значений концентрации в которых лежит в промежутке $[c_{d-1}^r, c_d^r]$.

Данные две функции $\nu_a(\delta, c_l^a)$ и $\nu_r(\delta, c_d^r)$ имеют смысл распределений абсолютного и относительного отклонения концентрации с фиксированной разницей между днями, в которые были сделаны соответствующие измерения.

Наибольший интерес представляет минимальная временная разница в один день, т. к. с одной стороны является базовой в утверждении гипотезы, с другой стороны оценивает любую другую разницу. Поэтому построим функции $\nu_a(1, c_l^a(1))$ и $\nu_r(1, c_d^r(0.05))$, дополнив их функцией совместного распределения

$$\nu(\delta, c_l^a(\varepsilon_a), c_d^r(\varepsilon_r)) = \|(t_i, t_j) : t_j \in T_\delta(i), \Delta_a K(i, j) \in [c_{l-1}^a, c_l^a], \Delta_r K(i, j) \in [c_{d-1}^r, c_d^r], \forall \tilde{X}\|.$$

Областью определения на рис. 21–22 являются точки разбиения абсолютного или соответственно относительного отклонения концентрации, на рис. 23 — их декартово произведение.

Рис. 21–23 показывают резкий спад количества событий, имеющих фиксированное отклонение концентрации, при увеличении последнего. Совместно с экспертом (Бердниковым Сергеем Владимировичем) были установлены пороговые значения для шумов: $\Delta_a K(i, j) > p_a = 20$, $\Delta_r K(i, j) > p_r = 0.5$.

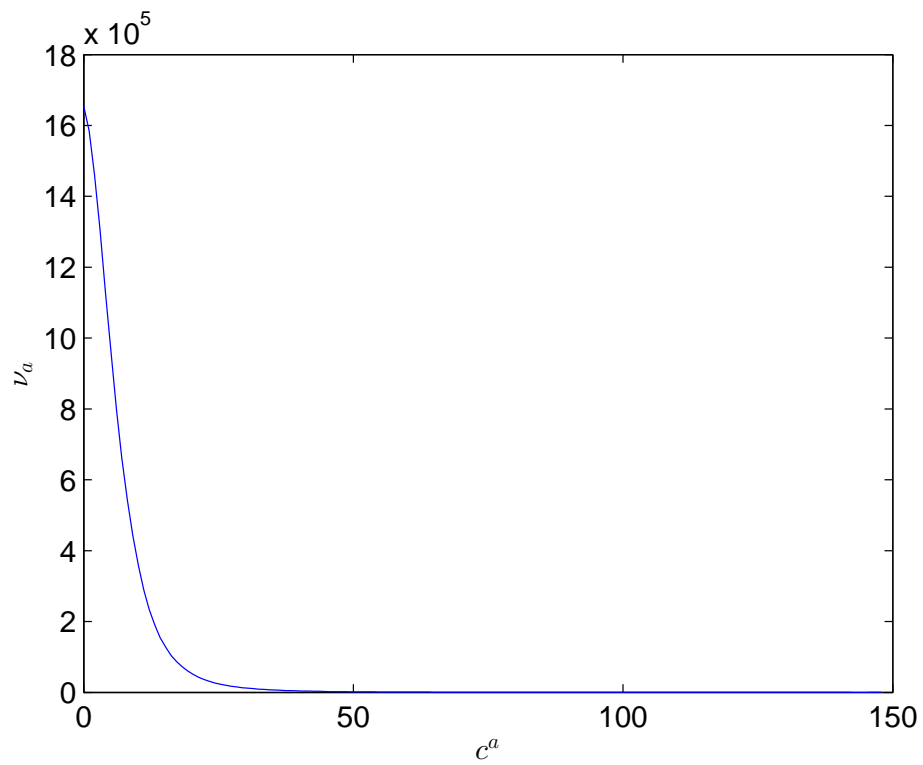


Рис. 21: $\nu_a(1, c_l^a(1))$

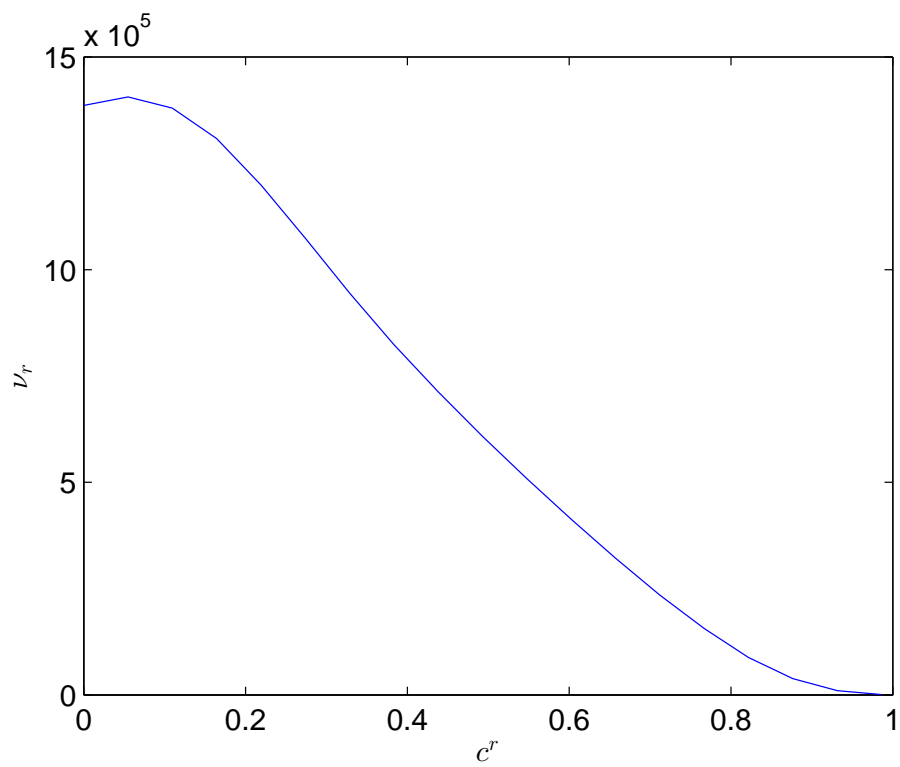


Рис. 22: $\nu_r(1, c_d^r(0.05))$

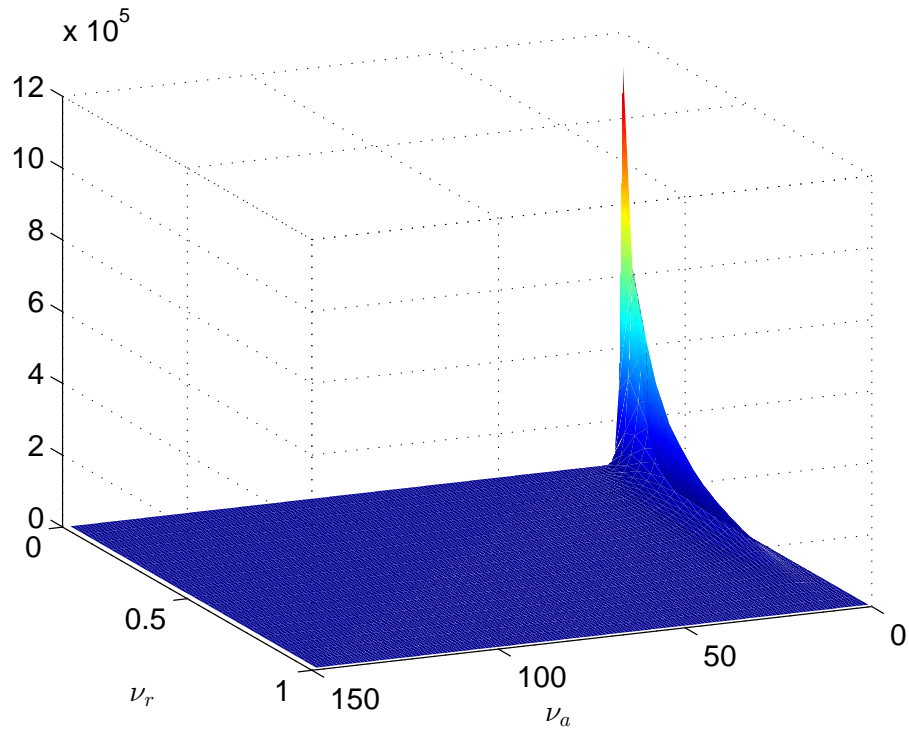


Рис. 23: $\nu(1, c_t^a(1), c_a^r(0.05))$

Для заданных значений подсчитаем процент случаев, попадающих под категорию выбросов:

- для абсолютного отклонения — 3.24%,
- для относительного отклонения — 15.75%,
- для относительного отклонения с дополнительным ограничением на знаменатель $\Delta_r K: \max(K_i, K_j) > 20$ — 4.81%,
- для совокупности условий (ограничиваются оба типа отклонений) — 2.08%.

Для абсолютного отклонения процент выбросов приемлем, несмотря на то, что для относительного отклонения является достаточно большим, с дополнительным условием на абсолютное значение концентрации в точке становится удовлетворительным. Т.к. все три типа ограничений, взаимно дополняя друг друга, отображают условие временной компактности, то можно утверждать, что для решения задачи поиска тренда данные являются адекватными.

6.3 Пространственная компактность

Для анализа пространственной изменчивости фиксируем момент времени \tilde{t} и повторим шаги предыдущего пункта

1. $\check{X} = \{X_0, \dots, X_m\} \subset \mathbb{N}^2$ — подмножество значений первого аргумента $\psi(X, t)$,
2. $\rho(X_i, X_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ — евклидово расстояние между точками X_i и X_j ,
3. X_i определяет окрестность радиусом δ : $U_\delta(i) = \{X \in \check{X}, \rho(X, X_i) = \delta\}$,
4. $M(X) = \psi(X, \tilde{t})$, $M_i = M(X_i)$,
5. $\Delta_a M(i, j) = |M_i - M_j|, i \neq j$,
6. дискретизация $\Delta_a M$: $\{w^a(\varepsilon)\} = \{w_0^a, w_1^a = w_0^a + \varepsilon, w_2^a = w_1^a + \varepsilon = w_0^a + 2\varepsilon, \dots, w_s^a = w_{s-1}^a + \varepsilon = w_0^a + s\varepsilon\}$,
7. $s = \frac{w_s^a - w_0^a}{\varepsilon}$,
8. $\mu_a(\delta, w_l^a(\varepsilon)) = \|(X_i, X_j) : X_j \in U_\delta(i), \Delta_a M(i, j) \in [w_{l-1}^a, w_l^a], \forall \tilde{t}\|$ — распределение абсолютного отклонения концентрации с фиксированным расстоянием между точками $X_j \in U_\delta(i)$, в которых были сделаны соответствующие измерения;
9. $\Delta_r M(i, j) = \frac{|M_i - M_j|}{\max(M_i, M_j)}, M_i \neq 0, M_j \neq 0, i \neq j$,
10. дискретизация $\Delta_r M$: $\{w^r(\varepsilon)\} = \{w_0^r, w_1^r = w_0^r + \varepsilon, w_2^r = w_1^r + \varepsilon = w_0^r + 2\varepsilon, \dots, w_g^r = w_{g-1}^r + \varepsilon = w_0^r + g\varepsilon\}$,
11. $g = \frac{w_g^r - w_0^r}{\varepsilon}$,
12. $\mu_r(\delta, w_d^r(\varepsilon)) = \|(X_i, X_j) : X_j \in U_\delta(i), \Delta_r M(i, j) \in [w_{d-1}^r, w_d^r], \forall \tilde{t}\|$ — распределение относительного отклонения концентрации с фиксированным расстоянием между точками $X_j \in U_\delta(i)$, в которых были сделаны соответствующие измерения;
13. $\mu(\delta, w_l^a(\varepsilon_a), w_d^r(\varepsilon_r)) = \|(X_i, X_j) : X_j \in U_\delta(i), \Delta_a M(i, j) \in [w_{l-1}^a, w_l^a], \Delta_r M(i, j) \in [w_{d-1}^r, w_d^r], \forall \tilde{t}\|$ — совместное распределение абсолютного и относительного отклонения с фиксированным расстоянием между точками $X_j \in U_\delta(i)$.

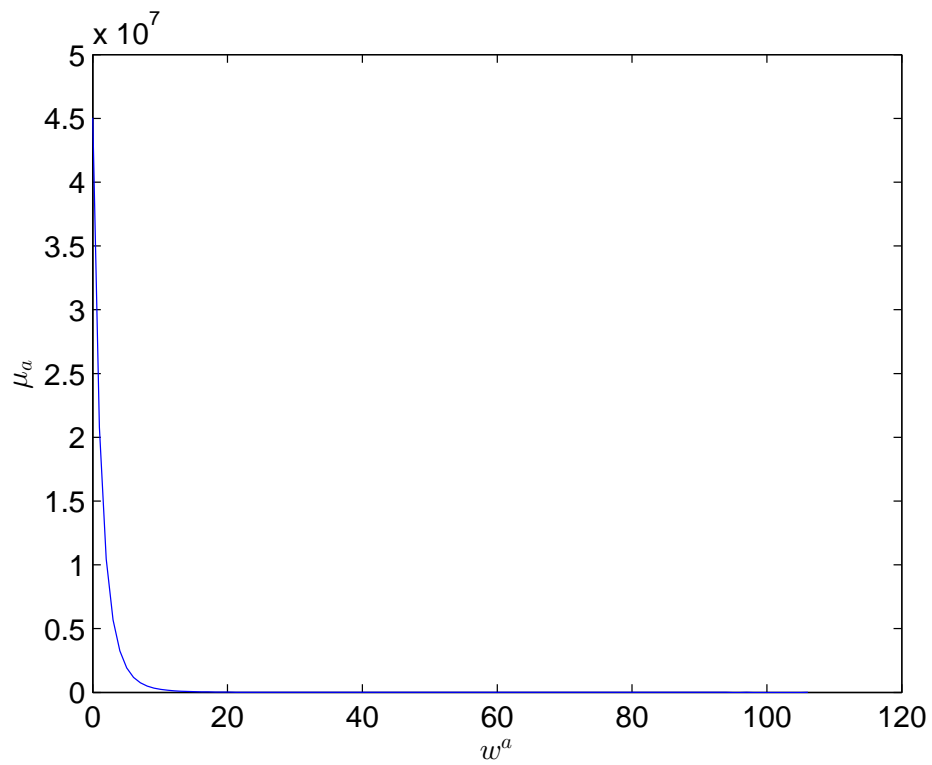


Рис. 24: $\mu_a(1, w_l^a(1))$

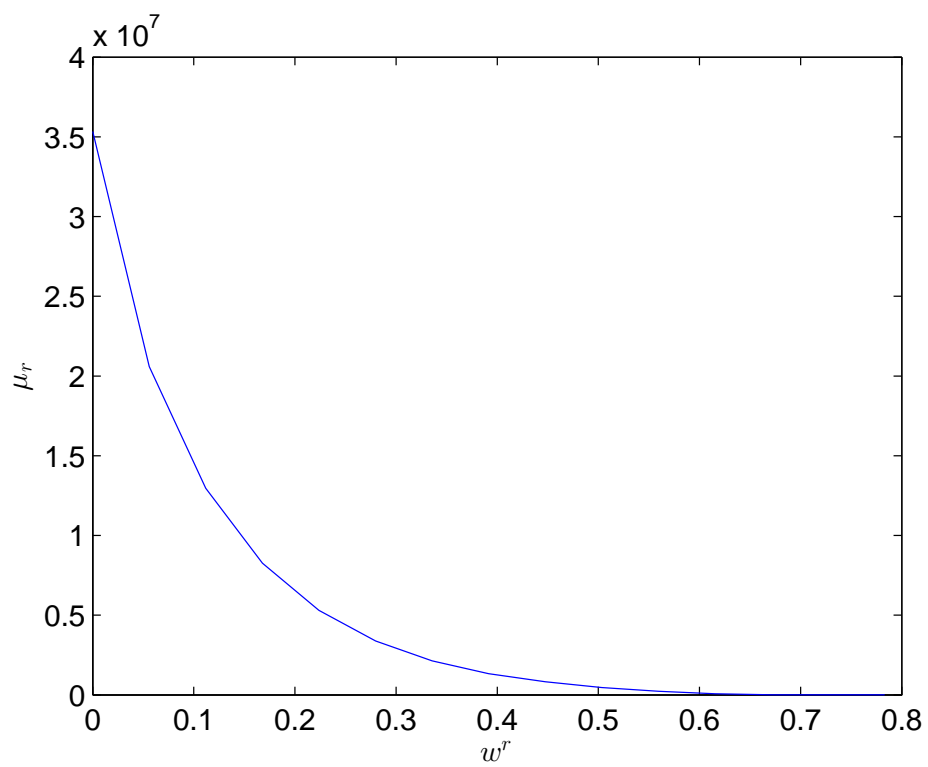


Рис. 25: $\mu_r(1, w_d^r(0.05))$

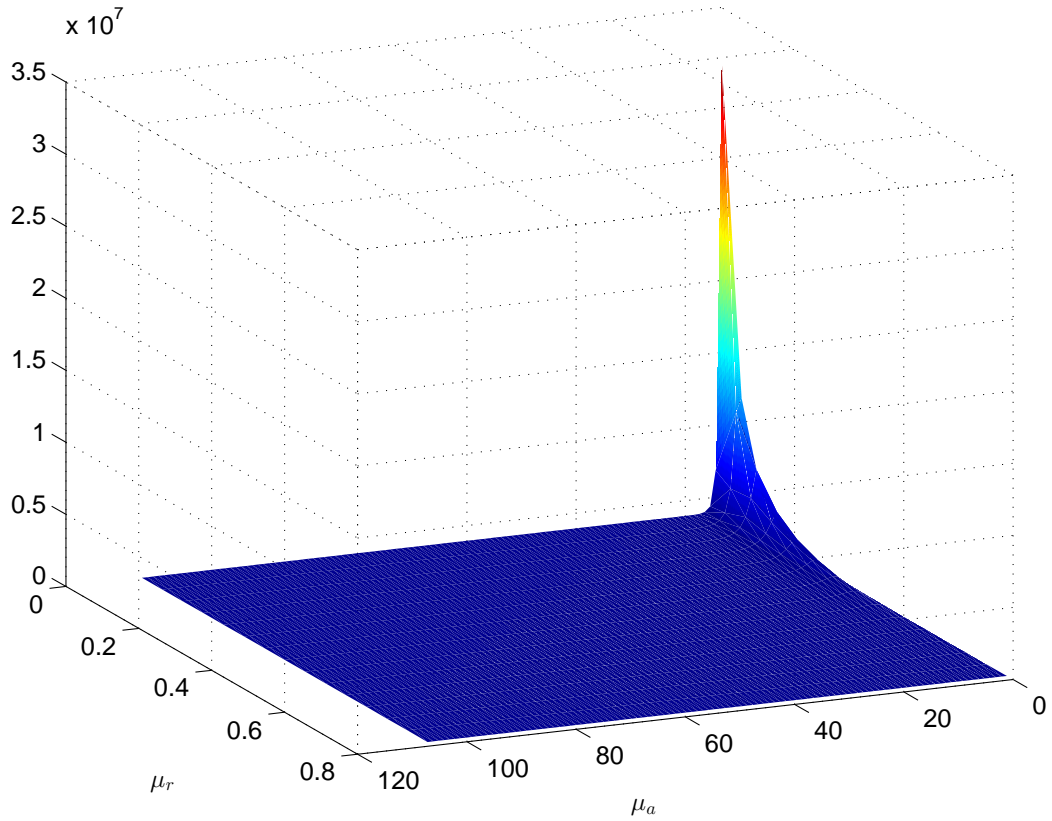


Рис. 26: $\mu(1, w_l^a(1), w_d^r(0.05))$

Т.к. минимальное расстояние, которое сформулировано в гипотезе, равно 1 и способно оценить любое другое расстояние, то для оценки адекватности данных достаточно визуализировать введенные характеристики $\mu_a(1, w_l^a(1))$, $\mu_r(1, w_d^r(0.05))$ и $\mu(\delta, w_l^a(\varepsilon_a), w_d^r(\varepsilon_r))$ для соседних точек.

Как и рис. 21–23, рис. 24–26 показывают резкий спад количества событий, имеющих фиксированное отклонение концентрации, при увеличении последнего. Рассмотрим пороговые значения для шумов: $\Delta_a M(i, j) > p_a^1 = p_a = 20$, $\Delta_r M(i, j) > p_r^1 = p_r = 0.5$; $\Delta_a M(i, j) > p_a^2 = 10$, $\Delta_r M(i, j) > p_r^2 = 0.25$.

Тогда процент случаев, попадающих под категорию выбросов:

1. для p_a^1 — 0.12%,
2. для p_a^2 — 1.08%,
3. для p_r^1 — 0.52%,
4. для p_r^2 — 10.32%,
5. для p_r^1 с дополнительным ограничением на знаменатель $\Delta_r M: \max(M_i, M_j) > p_a^1$ — 0.028%,

6. для p_r^2 с дополнительным ограничением на знаменатель $\Delta_r M: \max(M_i, M_j) > p_a^2 - 4.82\%$,
7. для совокупности ограничений p_a^1 и $p_r^1 - 0.0055\%$.
8. для совокупности ограничений p_a^2 и $p_r^2 - 0.57\%$.

Для всех типов порогов проценты выбросов существенно ниже аналогичного для временных распределений, но распределены аналогичным образом: определенных большим относительным отклонением выбросов существенно больше, чем определенных абсолютным. Т.к. все три типа ограничений, взаимно дополняя друг друга, отображают условие пространственной компактности, то можно утверждать, что для решения задачи поиска тренда данные являются адекватными.

7 Поиск нелинейного тренда

По визуализированным данным (рис.3,4,12–19) можно сделать выводы

- Частота обсервации неравномерна по всей области Азовского моря
- В большинстве случаев числа измерений недостаточно для выявления тренда:
 - измерения проведены редко,
 - полученные временные ряды для фиксированных точек акватории при визуальной оценке представляются случайным шумом без какого-либо тренда, что можно наблюдать на рис.12–16,
- Таганрогский залив (17–19) является обширной областью с достаточным числом измерений.

В результате только область Таганрогского залива из всей акватории Азовского моря является репрезентативной, поэтому сузим задачу поиска тренда на данную область. Более того, достаточно рассмотреть период одного года с марта по ноябрь, т.к. видимый тренд является годичным и в зимние месяцы в связи с непогодой данные нерепрезентативны.

7.1 Локальное решение

Данные хорошо интерполируются полиномами (рис. 27), поэтому будем искать нелинейный тренд в виде полинома пятой степени $Q(z) = b_0 + b_1z + b_2z^2 + b_3z^3 + b_4z^4 + b_5z^5$, где z — номер дня в году в выбранном периоде.

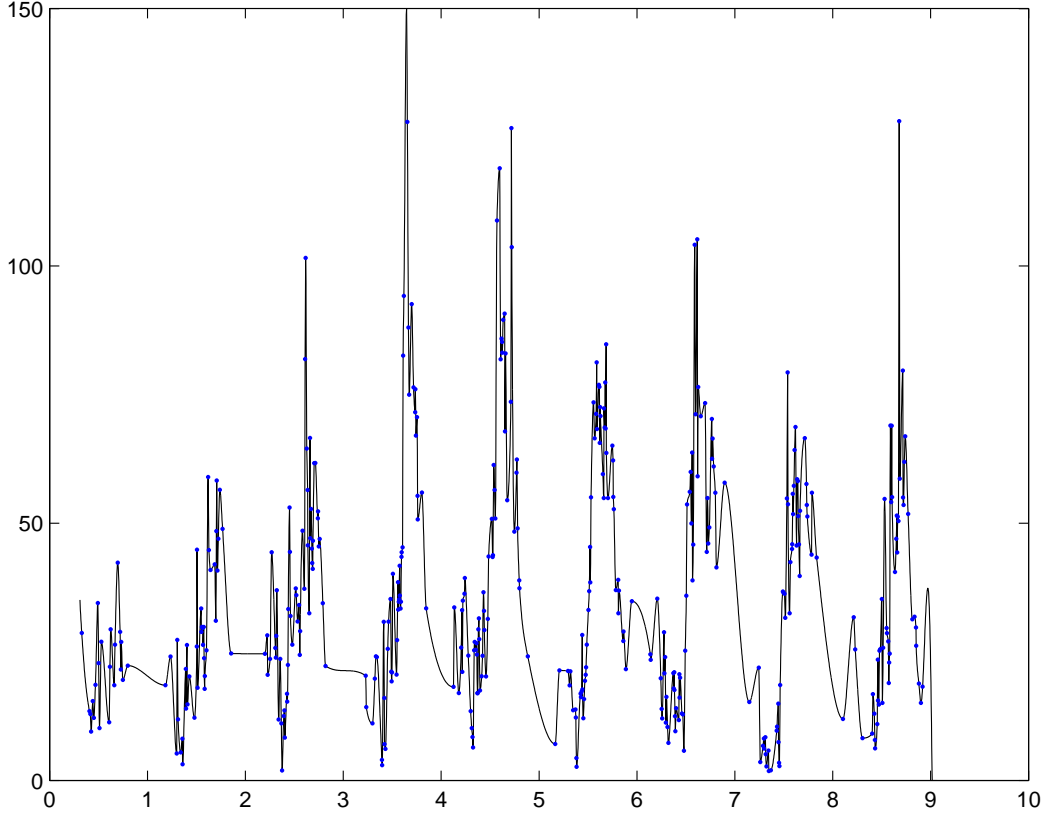


Рис. 27: График $\psi(X_6, t)$, интерполированный полиномами на $t \in \mathbb{R}$

Фиксируем точку акватории \tilde{X} и фиксируем изучаемый год. Обозначим $\gamma = t - z$ — число дней, прошедших с первого года съемки Азовского моря до текущего фиксированного года, $\bar{b} = (b_0 \ b_1 \ b_2 \ b_3 \ b_4 \ b_5)^T$ — вектор искомых коэффициентов многочлена. Тогда в силу ранее введенных обозначений $z_0, \dots, z_N \leftrightarrow t_0, \dots, t_N \leftrightarrow K_0, \dots, K_N$.

Для вычисления оптимальных коэффициентов введем квадратичный функционал ошибки:

$$L(\bar{b}) = \sum_{i=1}^N (Q(z_i) - K_i)^2 = \sum_{i=1}^N (b_0 + b_1z_i + b_2z_i^2 + b_3z_i^3 + b_4z_i^4 + b_5z_i^5 - K_i)^2 \quad (1)$$

Тогда искомые коэффициенты представлены решением задачи оптимизации без ограничений:

$$L(\bar{b}) \rightarrow \min_{\bar{b}} \quad (2)$$

Найдем решение аналитически по методу наименьших квадратов:

$$j = 0, \dots, 5: \quad 0 = \frac{\partial L}{\partial b_j} = \sum_{i=1}^N 2(b_0 + b_1 z_i + b_2 z_i^2 + b_3 z_i^3 + b_4 z_i^4 + b_5 z_i^5 - K_i) \cdot z_i^j,$$

$$0 = b_0 \sum_{i=1}^N z_i^j + b_1 \sum_{i=1}^N z_i^{j+1} + b_2 \sum_{i=1}^N z_i^{j+2} + b_3 \sum_{i=1}^N z_i^{j+3} + b_4 \sum_{i=1}^N z_i^{j+4} + b_5 \sum_{i=1}^N z_i^{j+5} - \sum_{i=1}^N K_i z_i^j,$$

Т.к. z_i и K_i — известные числа, то вычислимы суммы

$$j = 0, \dots, 10: \quad S_j = \sum_{i=1}^N z_i^j,$$

$$j = 0, \dots, 5: \quad G_j = \sum_{i=1}^N K_i z_i^j.$$

В новых обозначениях получим систему линейных алгебраических уравнений:

$$j = 0: \quad b_0 S_0 + b_1 S_1 + b_2 S_2 + b_3 S_3 + b_4 S_4 + b_5 S_5 = G_0,$$

$$j = 1: \quad b_0 S_1 + b_1 S_2 + b_2 S_3 + b_3 S_4 + b_4 S_5 + b_5 S_6 = G_1,$$

$$j = 2: \quad b_0 S_2 + b_1 S_3 + b_2 S_4 + b_3 S_5 + b_4 S_6 + b_5 S_7 = G_2,$$

$$j = 3: \quad b_0 S_3 + b_1 S_4 + b_2 S_5 + b_3 S_6 + b_4 S_7 + b_5 S_8 = G_3,$$

$$j = 4: \quad b_0 S_4 + b_1 S_5 + b_2 S_6 + b_3 S_7 + b_4 S_8 + b_5 S_9 = G_4,$$

$$j = 5: \quad b_0 S_5 + b_1 S_6 + b_2 S_7 + b_3 S_8 + b_4 S_9 + b_5 S_{10} = G_5,$$

— с квадратной матрицей $\hat{S} = \begin{pmatrix} S_0 & S_1 & S_2 & S_3 & S_4 & S_5 \\ S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ S_2 & S_3 & S_4 & S_5 & S_6 & S_7 \\ S_3 & S_4 & S_5 & S_6 & S_7 & S_8 \\ S_4 & S_5 & S_6 & S_7 & S_8 & S_9 \\ S_5 & S_6 & S_7 & S_8 & S_9 & S_{10} \end{pmatrix}$

и вектором свободных членов $\hat{G} = (G_0 \ G_1 \ G_2 \ G_3 \ G_4 \ G_5)^T$.

В матричной форме записи система имеет вид

$$\hat{S}\bar{b} = \hat{G}, \tag{3}$$

и единственное решение при невырожденной матрице \hat{S}

$$\bar{b} = \hat{S}^{-1}\hat{G}. \tag{4}$$

7.2 Качественный анализ

Качественный анализ подразумевает когнитивную оценку универсальности полученного тренда. Под универсальностью подразумевается независимость или ограниченная зависимость от фиксированных параметров — координаты \tilde{X} и «годового сдвига» γ .

Для экспериментов выберем область Таганрогского залива с небольшой граничной эрозией, чтобы убрать из рассмотрения шумовые эффекты прибрежных вод.

Рассмотрим 2 серии экспериментов:

1. фиксируем «годовой сдвиг» γ , варьируется координата X (рис. 29),
2. и наоборот: фиксируем X , варьируется γ (рис. 30),

— при этом точки X генерируем случайным образом, года рассматриваем последовательно.



Рис. 28: Выделение и нумерация точек на акватории Таганрогского залива.

В итоге получаем два типа графиков, по оси абсцисс которых отложен номер дня, по оси ординат — соответствующее дню значение концентрации.

Рис. 29–30, полностью согласуясь с многочисленными экспериментами в 1 и 2 сериях, дают представление об общем виде кривой:

1. локальный минимум в конце апреля–начале мая,
2. глобальный максимум в конце августа–начале сентября,

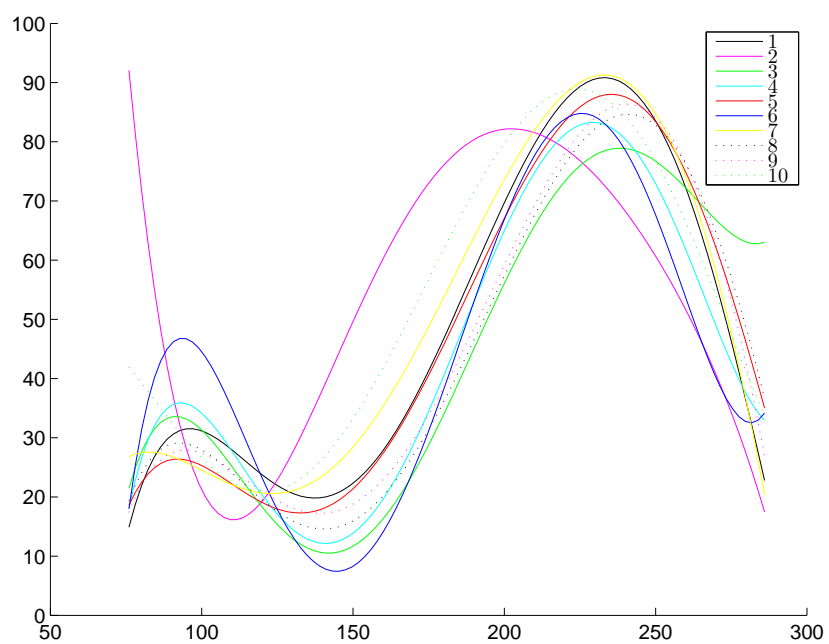


Рис. 29: Аппроксимация на 2008 год временных рядов выбранных точек.

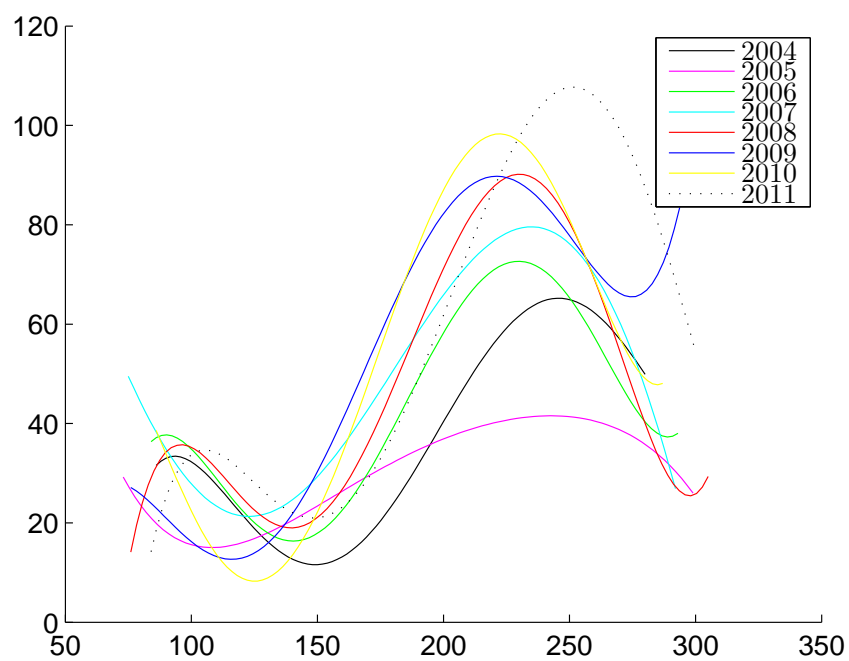


Рис. 30: Аппроксимация в точке 1 временных рядов за 2004-2011 года

3. на большинстве прямых заметен локальный максимум в конце марта–начале апреля,
4. весенний максимум меньше осеннего,
5. графики, дифференцированные по годам, отличаются существенно, чем дифференцированные по координатам,
6. благодаря краевым эффектам, о данных в начале и конце периода ничего определенного утверждать нельзя.

Обратим внимание, что качественный анализ позволяет опровергнуть теоретическую оценку ученых-океанологов о распределении концентрации хлорофилла при достоверности набора карт распределения концентрации хлорофилла. Перейдем к численной оценке.

7.3 Количественный анализ

Пусть z_{min} — номер дня в году, в который функционал $Q(z_i)$ достигает весеннего минимума, и z_{max} — номер дня, в который достигает осеннего максимума, Q_{min} и Q_{max} их соответственные значения. Также при детерминированном весеннем максимуме Q_{lmax} обозначим за z_{lmax} день, в который он достигается.

Воспользуемся методом Монте-Карло для оценки введенных величин:

1. случайным образом выберем на области Таганрогского залива n точек,
2. для каждой найдем соответствующий полином $Q(z)$,
3. вычислим стационарные точки полинома,
4. выберем z_{min} как минимальное значение на периоде с марта по ноябрь, z_{max} — как максимальное значение в этом же периоде,
5. если есть стационарная точка, принадлежащая рассматриваемому периоду, меньшая z_{min} , то она будет являться z_{lmax} , т.к. при существовании весеннего максимума он опережает весенний минимум.
6. для каждой из полученных величин построим гистограмму ее распределения и соответственного значения $Q(z)$.

Критерий Шапиро-Уилка при уровне значимости 0.05 отклоняет гипотезу нормальности для всех представленных распределений, но вид распределений позволяет говорить о наиболее вероятной области значений.

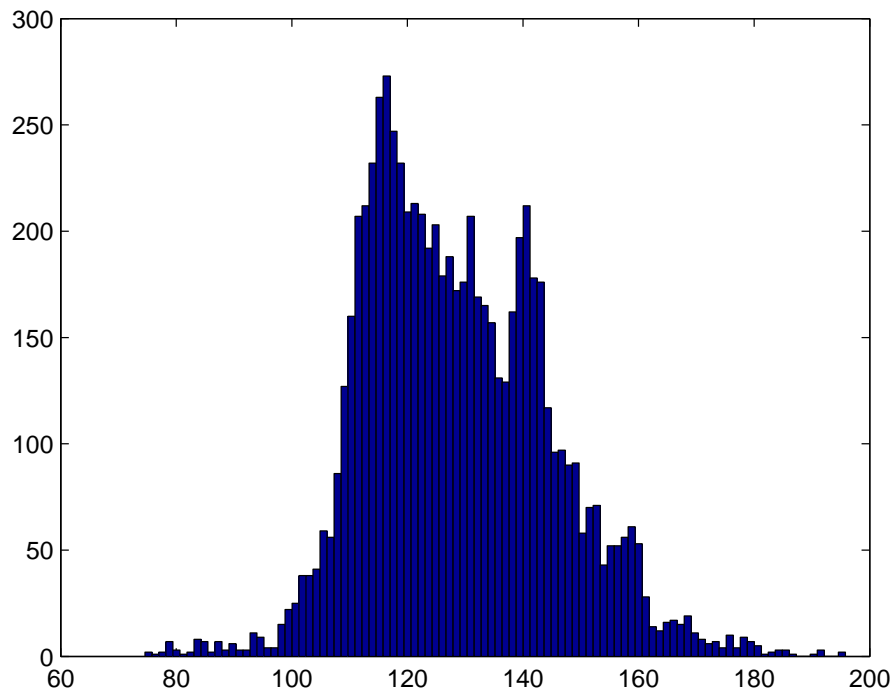


Рис. 31: Гистограмма z_{min} за все годовые периоды, $n = 1000$

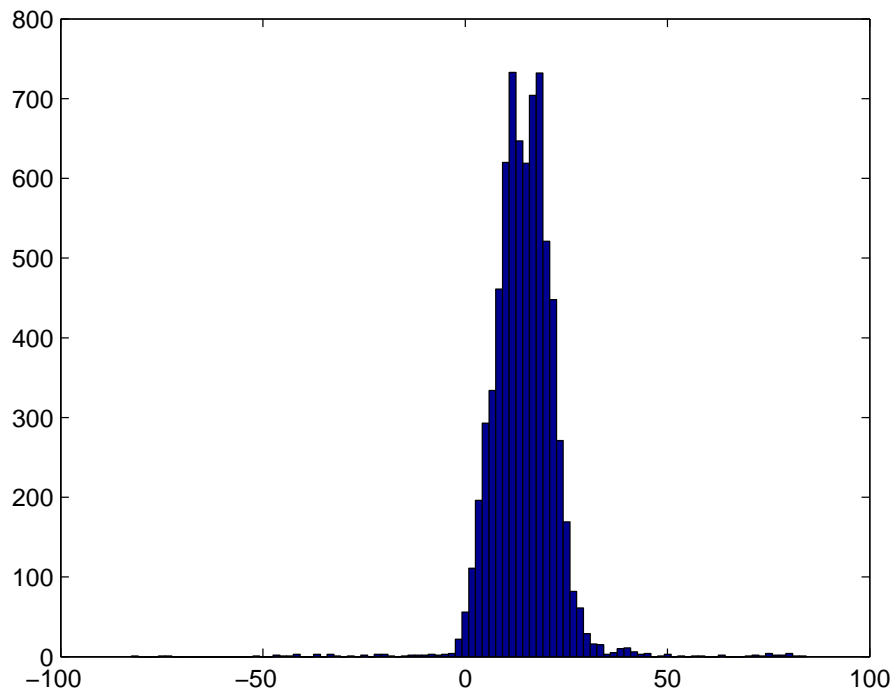


Рис. 32: Гистограмма Q_{min} за все годовые периоды, $n = 1000$

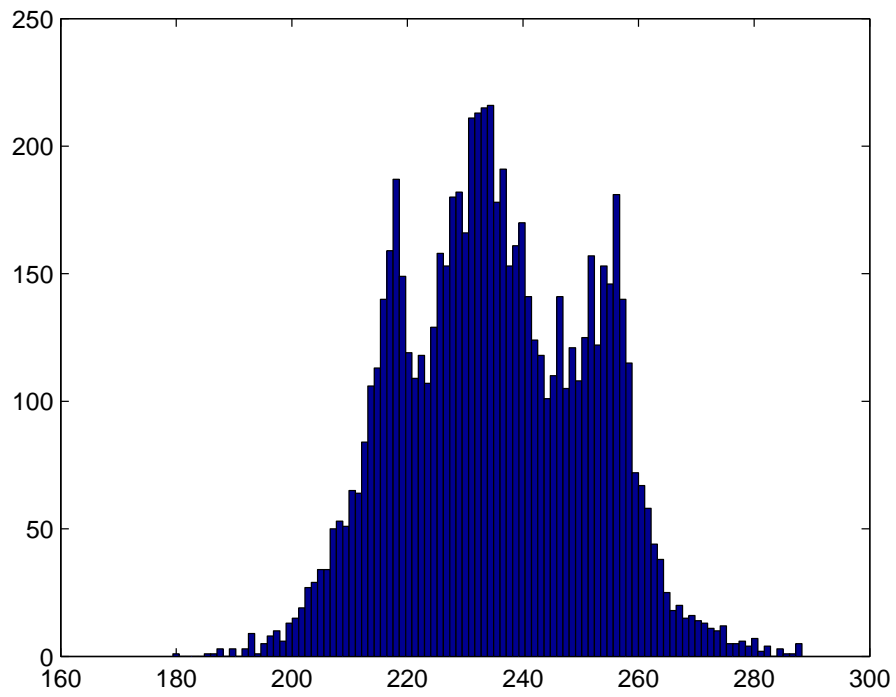


Рис. 33: Гистограмма z_{max} за все годовые периоды, $n = 1000$

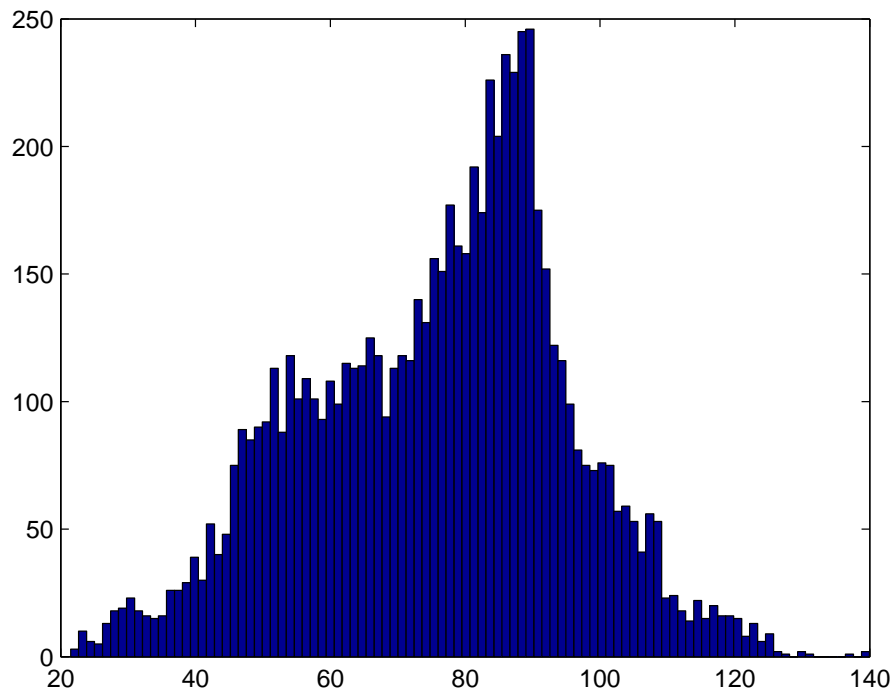


Рис. 34: Гистограмма Q_{max} за все годовые периоды, $n = 1000$

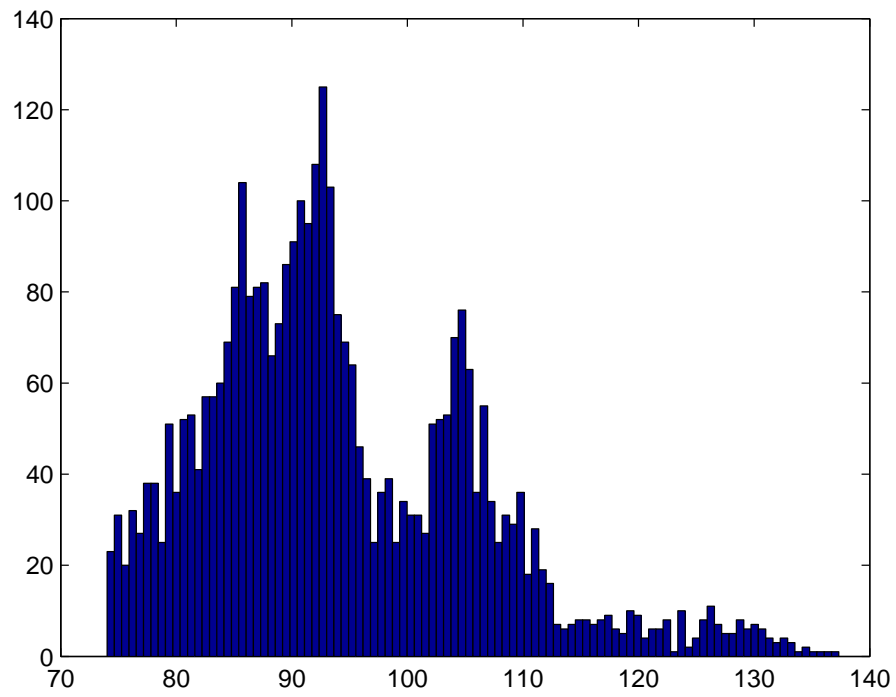


Рис. 35: Гистограмма z_{lmax} за все годовые периоды, $n = 1000$

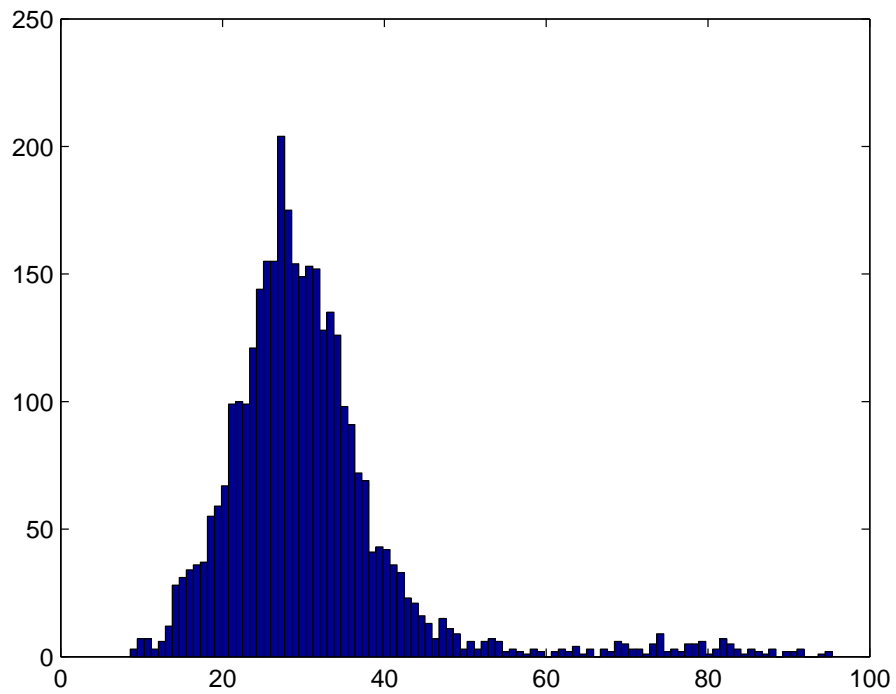


Рис. 36: Гистограмма Q_{lmax} за все годовые периоды, $n = 1000$

Полученные гистограммы численно показывают локализацию моды распределений:

- весеннего минимума в районе конца апреля–начале мая,
- осеннего максимума — в середине августа,
- весеннего максимума, при существовании (47.48%), — в начале апреля,
- минимальное значение концентрации Q_{min} — 10-20,
- максимальное значение Q_{max} — 90,
- локальное максимальное значение Q_{lmax} — 25-30.

Таким образом, несмотря на нестандартность распределений стационарных точек и значений в них, данные числа имеет хорошую степень локализации, подтверждающую общий вид распределения рис. 29–30. Более того, $Q_{lmax} < Q_{max}$ почти всюду, что экспериментально доказывает несостоятельность либо вида распределения концентрации хлорофилла рис. 1, либо способа перевода снимков спутника в карты концентрации хлорофилла.

8 Заключение

В представленной работе был исследован набор карт концентрации хлорофилла. Результатами, в частности, являются:

1. визуализация пространственных характеристик распределения и временных рядов отдельных точек,
2. оценка адекватности данных с опорой на гипотезу пространственно-временной компактности,
3. выделение репрезентативного подмножества Азовского моря — Таганрогского залива,
4. представление тренда временного ряда одного периода (март–ноябрь одного года) в виде многочлена пятой степени от дня периода
5. качественные и количественные подтверждения универсальности данного многочлена относительно координат точки Таганрогского залива и года периода,

6. количественное и качественное доказательство несостоятельности либо мнения экспертов о виде распределения концентрации хлорофилла рис. 1, либо способа перевода снимков спутников в карты концентрации хлорофилла.

Список литературы

- [1] *Moses W. J., Gitelson A. A., Berdnikov S., Saprygin V., Povazhnyi V.* Operational MERIS-based NIR-red algorithms for estimating chlorophyll-a concentrations in coastal waters — The Azov Sea case study // *Remote Sensing of Environment*, 2012. — Pp. 118–124.
- [2] *Moses W. J., Gitelson A. A., Berdnikov S., Povazhnyi V.* Estimation of chlorophyll-a concentration in case II waters using MODIS and MERIS data— Successes and challenges // *Environmental Research Letters*, 2009. — Pp. 8.
- [3] *Chen J., Zhang X., Quan W.* Retrieval chlorophyll-a concentration from coastal waters: three-band semi-analytical algorithms comparison and development // *Optics Express*, 2013. — Vol. 21, No. 7. — Pp. 9024–9042.
- [4] *Артемьев В. А., Буренков В. И., Возняк С. Б.* Подспутниковые измерения цвета океана: натурный эксперимент в Черном и Эгейском морях // *Океанология*, 2000. — Т. 40, No 2. — С. 192–198.
- [5] *Штрайхерт Е. А.* Распределение хлорофилла-А в Японском и Охотском морях по спутниковым и судовым данным // *дис.канд.геогр.наук*, 2008. — С. 160.
- [6] *Сапрыгин В. В.* Изучение распределения хлорофилла а в Азовском море по данным дистанционного зондирования Земли из космоса и результатам судовых измерений // *дис.канд.геогр.наук*, 2011. — С. 125.
- [7] *Буренков В. И., Копелевич О. В., Шеберстов С. В.* Подспутниковые измерения цвета океана: верификация спутниковых данных сканера цвета SeaWiFS // *Океанология*, 2000. — Т. 40, No 3. — С. 357–362.
- [8] *Whigham P. A., Recknagel F.* Predicting chlorophyll-a in freshwater lakes by hybridising process-based models and genetic algorithms // *Ecological*, 2001. — Vol. 146. — Pp. 243–251.
- [9] *Acker J. G., McMahon E., Shen S., Hearty T., Casey N. C.* Time-series analysis of remotely-sensed SeaWiFS chlorophyll in river-influenced coastal regions // *Ecological*, 2009. — Vol. 8, No 2. — Pp. 114–140.

- [10] *Strutton P. G., Mitchell J. G., S-Parslow J.* Non-linear analysis of chlorophyll a transects as a method of quantifying spatial structure // *Journal of Plankton Research*, 1996. — Vol. 18, No 9. — Pp. 1717–1726.
- [11] *Montera L., Jouini M., Verrier S., Thiria S., Crepon M.* Multifractal analysis of oceanic chlorophyll maps // *Ocean Science*, 2011. — Vol. 8, No 1. — Pp. 55–84.
- [12] *Васечкина Е. Ф.* Возможности вейвлет-анализа для изучения пространственно-временной изменчивости содержания хлорофилла «а» в поверхностном слое Черного моря // *Морской гидрофизический журнал*, 2012. — No 2. — С. 50–59.
- [13] *Абакумов А. И., Израильский Ю. Г.* Модельный способ оценки содержания хлорофилла в море на основании спутниковой информации // *Компьютерные исследования и моделирование*, 2013. — Т. 5, No 3. — С. 473–482.
- [14] *Mattern J. P., Fennel K., Dowd M.* Estimating time-dependent parameters for a biological ocean model using an emulator approach // *Journal of marine systems*, 2012. — Vol. 96–97, No 5. — Pp. 32–47.
- [15] *McKenzie E.* General exponential smoothing and the equivalent ARMA Process // *Journal of Forecasting*, 1984. — No 3. — С. 333–344.
- [16] *Bishop C.* *Pattern Recognition and Machine Learning* // Springer, 2006.