

АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ И ОКОНТУРИВАНИЕ ОБЛАСТЕЙ ЦВЕТЕНИЯ *LEPIDODINIUM CHLOROPHORUM* ПО ДАННЫМ СКАНИРУЮЩЕЙ АППАРАТУРЫ

© 2011 г. Е. А. Морозов^{1,2*}, О. Б. Анискина^{1,2}, Д. В. Поздняков^{1,3}, Л. Х. Петтерссон³, В. И. Сычев², Х. Грассл⁴

¹Научный фонд “Международный центр по дистанционному зондированию и окружающей среде им. Хансена”, Санкт-Петербург

²Российский государственный гидрометеорологический университет, Санкт-Петербург

³Центр по дистанционному зондированию и окружающей среде им. Хансена, Берген, Норвегия

⁴Метеорологический институт им. Макса Планка, Гамбург, Германия

*E-mail: evgeny.morozov@niersc.spb.ru

Поступила в редакцию 03.02.2010 г.

Для автоматического распознавания и оконтуривания областей цветения динофлагеллята *Lepidodinium chlorophorum* разработаны два методологически независимых алгоритма, основанные на нейронно-сетевом методе и нечеткой логике *k*-средних. Обучение нейронно- сетевого алгоритма и валидация результатов применения обоих алгоритмов к данным спутникового датчика цвета океана MODIS/Aqua производились с использованием данных измерений *in situ* в водах Атлантического побережья Франции. Выявленные локализации цветения указанной водоросли обоими алгоритмами оказались весьма близкими/перекрывающимися, что свидетельствует о возможности их ансамблевого применения. Разработанные алгоритмы не специфичны в отношении типа вод и могут быть применены к различным морским акваториям.

Ключевые слова: MODIS, Бискайский залив, нейронная сеть, цветение вредных водорослей, *Lepidodinium chlorophorum*

ВВЕДЕНИЕ

Из современных достижений спутниковой гидроэкологии можно выделить в той или иной мере успешное оперативное слежение за возникновением и пространственно-временной динамикой цветения фитопланктона-сообщества водных (по преимуществу автотрофных) микроорганизмов растительного происхождения. Само по себе — это, безусловно, важный этап в развитии дистанционного зондирования (ДЗ) Мирового Океана из космоса в силу нетривиальности этой задачи, учитывая все многообразие оптических свойств природных вод (Platt et al., 2008; Zibordi et al., 2009).

Однако в настоящее время все более настойчиво начинает выдвигаться задача по идентификации существующими спутниковыми датчиками цвета океана конкретной культуры (или комплекса культур) фитопланктона, обусловившей/обусловивших резкий рост биомассы наблюдаемого из космоса цветения (ICES-IOC Working Group, 2003).

Если для спутникового мониторинга зон цветения фитопланктона достаточно уметь реги-

стрировать лишь концентрацию хлорофилла-*a* (хл-*a*) — пигмента, универсально присутствующего в клетках всех без исключения автотрофных организмов на Земле и являющегося индикатором их биомассы, то решение идентификации типов фитопланктона оказывается существенно сложнее. В этом случае возникает необходимость в выделении в дистанционно-регистрируемом сигнале дополнительных индикаторов, однозначно характеризующих конкретный таксон. В качестве таковых могут выступать дополнительные к хл-*a* пигменты — другие хлорофиллы, а также каротиноиды и билипротеины. Однако для выделения сигналов от них в общем сигнале, регистрируемом с орбиты, спектральное разрешение функционирующих в настоящее время датчиков цвета океана, информация с которых является общедоступной, оказывается явно недостаточным, — что существенно усложняет практическое решение этой проблемы.

Между тем из основных причин, по которым задача дистанционной идентификации видов фитопланктона является весьма важной в водной экологии, можно выделить весьма тревожный

феномен нарастающе-частых и прогрессирующе-обширных цветений так называемых вредных микроводорослей (harmful algal blooms – HABs) (ICES-IOC Working Group, 2003). При этом под вредностью водорослей понимается довольно широкий спектр их негативного воздействия на среду обитания и ее биоту, а по цепи трофических взаимодействий – и на более широкий круг объектов биосферы Земли, включая и человека. HABs могут выделять в окружающую водную среду токсины различного поражающего действия (нервно-паралитического, диарического, амнезического и др.) или, при их высокой концентрации, приводить к механическому гемолитическому поражению тканей, дисфункции дыхательных органов водных организмов и истощению запасов растворенного кислорода в воде (Hallegraeff et al., 1995). Даже этот далеко неполный перечень возможных поражающих эффектов свидетельствует о важности проблемы своевременной идентификации HABs и мониторинга их пространственно-временной динамики. Понятно, что спутниковый мониторинг HABs в принципе является наиболее предпочтительным способом достижения этой цели в силу своей оперативности и большого одновременного охвата территории зондируемых вод. И с этой точки зрения исследование возможностей идентификации HABs существующими спутниковыми системами регистрации цвета океана в сочетании с разработкой соответствующих методологических подходов представляет несомненный интерес (Cracknell et al., 2001; Babin et al., 2008).

Среди типичных представителей HABs обычно выделяют некоторые виды таких водорослей, как *Diatoms*, *Prymnesiophytes*, *Raphidophytes*, *Cyanobacterium* и *Dinoflagellates* (Hallegraeff et al., 1995).

Однако такой динофлагеллят, как *Lepidodinium chlorophorum* (*L. chlorophorum*), еще относительно недавно не упоминался в литературе как экологически вредная водоросль. Отчасти это объясняется тем, что *L. chlorophorum* была впервые описана лишь в 1996 г., а более корректное ее описание появилось всего несколько лет назад, в 2007 г. Выяснилось, что хотя *L. chlorophorum* не является токсической водорослью, массовое цветение этого вида планктона приводит, вследствие выделения значительного количества слизи, к аноксии вод и к гибели моллюсков.

Таким образом, районы массового развития *L. chlorophorum* должны рассматриваться как HABs, а их мониторинг важен как с точки зрения изучения водно-экологической динамики в районе цветения, так и с практической, связанной с состоянием/сохранением ресурсов морепродуктов.

В настоящем исследовании нами было разработано два типа алгоритмов для идентификации, районирования и отслеживания динамики зон

цветения *L. chlorophorum* с использованием спутникового сенсора MODIS/Aqua.

ОБЩАЯ ХАРАКТЕРИСТИКА *LEPIDODINIUM CHLOROPHORUM*

Фитопланктон *L. chlorophorum* – ярко-зеленые эндосимбионты филума *Dinoflagellata*, класса *Dinophyceae*, отряда *Gymnodiniales*, классификационная идентификация которых впервые была произведена Эльбрахтером и Шнепфом (Elbrächter, Schnepf, 1996), а затем скорректирована Хансенем, Ботесом и Де-Салесом (Hansen et al., 2007). Полагают, что *L. chlorophorum* эволюционно происходит из класса *Prasinophytes* отдела *Chlorophyta* – примитивных эукариотов – морских зеленых микроводорослей. Это родство, по-видимому, и обусловило наличие в клетках *L. chlorophorum* зеленых перидинин-содержащих пластиидов с хлорофильным комплексом хл-*a* и хлорофилл-*b* (хл-*b*) (Takishita et al., 2008). Содержание пигментов в клетках *L. chlorophorum* таково, что ее интегральный удельный коэффициент поглощения, $a_{L,chl}^*$ довольно высок [$3.32 \cdot 10^{-2} \text{ м}^2 \text{ мгхл-}a^{-1}$] по сравнению, например, с некоторыми другими морскими водорослями – представителями филумов *Dinophyta* и *Haptophyta* (Claquin et al., 2008). Сведения о спектральных значениях $a_{L,chl}^*$, а также удельного коэффициента обратного рассеяния, $b_{L,chl}^*$ по-видимому, в доступной/рецензируемой литературе отсутствуют.

В отличие от многих других представителей динофлагеллят, которые образуют токсические цветения, имеют симбиотические отношения (в т.ч. с кораллами) и имеют гетеротрофный тип питания, *L. chlorophorum* преимущественно фототроф (хотя и сохраняет способность к гетеротрофии (Morden, Sherwood, 2002)) и вместо токсичных веществ выделяет в окружающую водную среду полисахариды в виде прозрачных коллоидных биополимеров, которые затем коагулируют с образованием TEPs (transparent exopolymeric particles). При этом пропорция выделяемого в водную среду углерода к органическому углероду, фотосинтезируемого клеткой *L. Chlorophorum*, может достигать ~71% (в то время как у других таксонов этот показатель варьирует в пределах ~7–30%). TEPs сложным образом вовлечены в биогеохимические процессы в морской среде (Passow, 2002a; Passow, 2002b), в том числе и в процессы седиментации клеток *L. chlorophorum*. Таким образом, время цветения *L. chlorophorum*, в особенности в условиях ветрового перемешивания, ощутимо сокращается. TEPs влияют на скорости выедания фитопланктона зоопланктоном, удерживают/концентрируют бактерии и вирусы, ухудшая тем самым патогенный фон в зоне цветения.

Поскольку у *L. chlorophorum* отношение E/E_k (экспоненциально возрастающей освещенности к параметру светового насыщения) увеличивается с понижением температуры воды (Claquin et al., 2008), этот таксон обладает в холодных/прохладных водах повышенной устойчивостью к эффекту фотоингибирования первичной продуктивности, что объясняет его способность вегетировать в средних и нижних широтах, часто в зонах апвеллингов, в условиях высоких уровней инсоляции, особенно в летний период. Регулярные цветения *L. chlorophorum* отмечаются у побережья Южной Африки и Австралии, вдоль французского атлантического побережья и в Северном море (куда эта водоросль, по-видимому, заносится береговыми течениями), а также во многих других регионах Мирового Океана (Hansen et al., 2007; Elbrachter, 1998).

Не являясь токсигенной, *L. chlorophorum*, благодаря своей способности обильно выделять ТЕРPs, образует в периоды своего массового развития, в зоне цветения, слизистые поля и скопления, которые вызывают аноксию, и, в особенности на мелководье, являются причиной многочисленных дисфункций у ракообразных и моллюсков как за счет дефицита растворенного кислорода, так и в результате обволакивания этих организмов.

В связи с этим как со стороны морских биологов/экологов, так и производителей морепродуктов усилилось внимание к феномену цветения *L. chlorophorum*, в том числе и в контексте организации мониторинга возникновения и пространственно-временной динамики этого явления.

МЕТОДОЛОГИЯ СОЗДАНИЯ АЛГОРИТМОВ

Не ставя здесь перед собой задачу исчерпывающего обзора работ, посвященных методологии дистанционной идентификации типов фитопланктона, и, более специфично, НАВs, отметим лишь, что к настоящему времени опробован целый ряд подходов (Kutser, 2009).

Изначально, методология строилась просто на выявлении областей высокой концентрации хл-*a* в предположении, что обнаруженные таким образом области большой биомассы фитопланктона принадлежат монотаксонному цветению, видовой принадлежность которого подлежала определению *in situ* (если это предварительно уже не было сделано) (Stumpf et al., 2003). В последующем широкое распространение получили алгоритмы, основанные на отношениях сигналов в одной или нескольких парах каналов спутникового датчика (Kutser, 2009). Комбинации каналов подбирались по принципу наибольшей чувствительности к спектральным сигнатурам пигментов таксона, подлежащего мониторингу.

Наряду с этим в лабораторных, а также в полевых условиях (с судов и самолетов), с применением спектрометров высокого разрешения с непрерывной регистрацией спектра восходящего излучения, изучались возможности идентификации планктонных таксонов по полосам поглощения их характерных пигментов, в том числе и с использованием метода дифференциальной спектроскопии. В частности, этот подход был реализован с применением таких спутниковых приборов, как Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) (Brache et al., 2009) и гиперспектральный визуализирующий спектрометр Hyperion. К гиперспектральным данным могут быть успешно применены различные приемы дискриминационного анализа (Weizman, Goldberger, 2009).

Однако для сенсоров цвета океана типа SeaWiFS, MODIS и MERIS, обладающих лишь несколькими полосами в видимом диапазоне спектра и относительно низким спектральным разрешением, метод спектроскопии высокого разрешения нереализуем.

При наличии гидрооптической модели зондируемых вод, т.е. табулированных значений спектральных удельных коэффициентов поглощения и обратного рассеивания всех основных цветформирующих компонент, включая и водоросль, обусловившую цветение, весьма эффективными оказываются методы многомерной оптимизации (см., например, (Korosov et al., 2009)).

В случае наличия большого числа синхронизированных с пролетом спутникового датчика цвета океана таксономических данных *in situ* возможна разработка эмпирических идентификационных алгоритмов на основе многомерных корреляционных методов или метода нейронных сетей (МНС).

Существуют, однако, методы анализа, основанные на нечеткой логике (Zimmermann, 2001), позволяющие на основе регистрируемых дистанционным датчиком спектров световых сигналов производить объективное выделение классов, характеризующееся теми или иными спектральными сигнатурами.

Исходя из анализа объема и характера базы дистанционных данных и измерений *in situ*, которой мы располагали, для идентификации *L. chlorophorum* нами были выбраны МНС и метод, основанный на нечеткой логике k -среднего (Dembele, Kastner, 2003). В случае успеха, параллельное применение двух “генетически” независимых алгоритмов позволяло предложить ансамблевый подход, обеспечивающий большую надежность обнаружения и оконтуривания района цветения *L. chlorophorum*, с последующим слежением за его пространственно-временной динамикой.

Нейронно-сетевой алгоритм

Из существующих типов искусственных нейронных сетей наиболее частое применение для целей ДЗ нашли многослойные персептроны (МП) (Найкин, 1998). В общем случае МП состоит из трех слоев: входного, внутреннего, или промежуточного, и выходного. Входной слой содержит нейроны, отражающие свойства входного параметра. В нашем случае это могут быть спектральные значения коэффициента отражения под поверхностью воды для ДЗ, R_{rsw} (по определению R_{rsw} – яркость восходящего света непосредственно под поверхностью воды, нормированная на исходящую освещенность, создаваемую падающим светом на том же уровне (Jerome et al., 1996)). Выходной слой содержит нейроны, число которых соответствует числу искомым параметров (в нашем случае это один нейрон, величина выходного сигнала которого указывает на наличие или отсутствие цветения). Число нейронов во внутреннем слое подбирается экспериментально и зависит от характера связи между входными и выходными параметрами. Каждый нейрон в нейронной сети связывается со всеми нейронами предыдущего и последующего слоя посредством придания передаточной функции индивидуальных весов.

Входные сигналы передаются от нейронов предыдущего слоя к нейронам последующего слоя (так называемый “feed-forward”-способ). При прохождении сигнала от входного слоя к выходному происходит его трансформация соответствующими индивидуальными весами связей между нейронами. Каждый принимающий нейрон суммирует взвешенные сигналы, поступающие к нему от нейронов предыдущего слоя

$$net_j = \sum_{i=1}^N w_{ji} o_i, \quad (1)$$

где w_{ji} – вес связи между нейроном i и нейроном j , а o_i – выходной сигнал нейрона i . Тогда выходной сигнал нейрона j может быть выражен как

$$o_j = f(net_j). \quad (2)$$

Обычно передаточная функция f представляет из себя нелинейную сигмоидную функцию. Эта функция преобразует взвешенную сумму входных сигналов для каждого нейрона для последующей передачи выходного сигнала этого нейрона на нейроны последующего слоя. Достижение входным сигналом выходного слоя знаменует фактическое создание нейронной сети. Созданная таким образом нейронная сеть подлежит тренировке с тем, чтобы придать ей способность к генерализации, т.е способность к интерполяции и

корректному решению обратной задачи с входными данными отличными от тех, что составляли пул тренировочных данных. Значения входного параметра (в нашем случае это спектральные величины $R_{rsw}(\lambda)$) пропускаются через сеть, и сигналы в выходных нейронах сравниваются с фактическими/реальными значениями искомым параметров (в нашем случае – проверяется наличие или отсутствие цветения). Ошибка восстановления пропускается по нейронам в обратном направлении для коррекции весов

$$\Delta w_{ji}(n+1) = \eta(\delta_j o_i) + \alpha \Delta w_{ji}(n), \quad (3)$$

где η – параметр обучения; δ_j – индекс скорости изменения величины ошибки; α – величина момента. Процесс тренировки продолжается до достижения искомой точности восстановления. Сеть, натренированная таким образом, затем тестируется с использованием дополнительного пула данных входных и выходных значений параметров, не вошедших в тренировочный пул.

При реализации нейронно-сетевого алгоритма (НСА) нами использовалось программное обеспечение, созданное в Штутгартском университете (<ftp://ftp.informatik.uni-stuttgart.de/pub/SNNS/>). Количество нейронов во входном слое было выбрано равным шести, по числу спектральных каналов спутникового датчика MODIS/Aqua. В поисках оптимальной архитектуры сети, методом перебора количества нейронов в промежуточном слое нейронной сети число нейронов в промежуточном слое было принято равным шести. Выходной слой содержал один нейрон. На вход сети подавались значения R_{rsw} на длинах волн 412; 443; 488; 531; 551 и 667 нм. Эти входные сигналы инвертировались сетью таким образом, что на выходе результирующее число варьировало между 0 и 1 : 0 или 1 соответствовали соответственно, отсутствию или наличию цветения *L. chlorophorum*, а промежуточные значения интерпретировались как соответствующие пикселям перехода от цветения к его отсутствию.

Алгоритм, основанный на нечеткой логике k -среднего (АКС)

В основе алгоритма заложен метод кластеризации – объединения объектов в группы (кластеры) на основе схожести признаков для объектов одной группы и отличий между группами. Большинство алгоритмов кластеризации не опираются на традиционные для статистических методов допущения о законе распределения данных наблюдений и могут использоваться в условиях почти полного отсутствия статистической инфор-

мации. Если исходной информацией для кластеризации является матрица наблюдений

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & & & \\ x_{M1} & x_{M2} & \dots & x_{Mn} \end{bmatrix}, \quad (4)$$

каждая строчка которой представляет собой результат измерения n признаков одного из M объектов кластеризации, тогда задача кластеризации состоит в разбиении объектов из матрицы X на несколько подмножеств (кластеров), в которых объекты более схожи между собой, чем с объектами из других кластеров. “Схожесть” определяется через расстояние, которое может рассчитываться как между исходными объектами (строчками матрицы X), так и от этих объектов к прототипу кластеров. Обычно (как и в нашем случае) координаты прототипов кластеров заранее неизвестны – они находятся одновременно с разбиением данных.

Разработанные к настоящему времени методы кластеризации можно разделить на четкие (Олдендерфер, Блэшфилд, 1989) и нечеткие (Ross, 1995; Штовба, интернет-ресурс). Четкие методы кластеризации разбивают исходное множество объектов X на несколько непересекающихся подмножеств. Нечеткие методы кластеризации позволяют одному и тому же объекту принадлежать одновременно нескольким (или даже всем) кластерам, но с различной степенью принадлежности. Нечеткая кластеризация во многих ситуациях более “естественна”, чем четкая, например, для объектов, расположенных на границе кластеров.

Методы кластеризации также классифицируются по тому, определено количество кластеров заранее или нет. В последнем случае количество кластеров определяется в ходе выполнения алгоритма на основе распределения исходных данных. В нашей задаче число кластеров известно заранее, поэтому использовался алгоритм k -средних, разбивающий данные на наперед заданное число кластеров.

В алгоритме нечетких k -средних нечеткие кластеры описываются матрицей нечеткого разбиения

$$F = [\mu_{ki}], \quad \mu_{ki} \in [0, 1], \quad (5)$$

$$k = 1, \dots, M, \quad i = 1, \dots, c,$$

в которой k -я строчка содержит степени принадлежности объекта $(x_{k1}, x_{k2}, \dots, x_{kn})$ к кластерам A_1, A_2, \dots, A_c . Матрица F описывает степень принадлежности объекта к кластеру, и, при нечетком разбиении, степень принадлежности объекта к кластеру принимает значения из интервала $[0, 1]$, а при четком – из двухэлементного множества

$\{0, 1\}$. Условия для матрицы нечеткого разбиения записываются следующим образом:

$$\sum_{i=1}^c \mu_{ki} = 1, \quad k = 1, \dots, M, \quad (6)$$

$$0 < \sum_{k=1}^M \mu_{ki} < N, \quad i = 1, \dots, c. \quad (7)$$

Нечеткое разбиение позволяет просто решить проблему объектов, расположенных на границе двух кластеров, – им назначают степени принадлежности, равные 0.5. Недостаток нечеткого разбиения проявляется при работе с объектами, удаленными от центров всех кластеров. Удаленные объекты имеют мало общего с любым из кластеров, поэтому, казалось бы, для них должны быть назначены малые степени принадлежности. Однако, по условию (6), сумма их степеней принадлежности такая же, как и для объектов, близких к центрам кластеров, т.е. равна единице. Для устранения этого недостатка можно использовать возможностное разбиение, которое требует только, чтобы произвольный объект из X принадлежал хотя бы одному кластеру. Возможностное разбиение получается следующим ослаблением условия (6): $\exists i, \mu_{ki} > 0, \forall k$.

Для оценки качества нечеткого разбиения используется следующий критерий разброса (Штовба, интернет-ресурс):

$$\sum_{i=1}^c \sum_{k=1}^N (\mu_{ki})^m \|V_i - X_k\|^2, \quad (8)$$

где $V_i = \frac{\sum_{k=1}^N (\mu_{ki})^m x_k}{\sum_{k=1}^N (\mu_{ki})^m}$ – центры нечетких кластеров;

$m \in [1, \infty)$ – экспоненциальный вес, определяющий нечеткость/размазанность кластера.

Существует множество алгоритмов нечеткой кластеризации, основанных на минимизации критерия (8). Нахождение матрицы нечеткого разбиения F с минимальным значением критерия (8) представляет собой задачу *нелинейной* оптимизации, которая может быть решена разными методами. В данной работе был применен метод неопределенных множителей Лагранжа (Штовба, интернет-ресурс; Silvey, 1959; Петрушко и др., 2002).

Для повышения эффективности процедуры кластеризации использовались не сами значения $R_{rsw}(\lambda)$, а соответствующие им главные компоненты (в количестве шести), полученные методом главных компонент (МГК). Таким образом, реа-

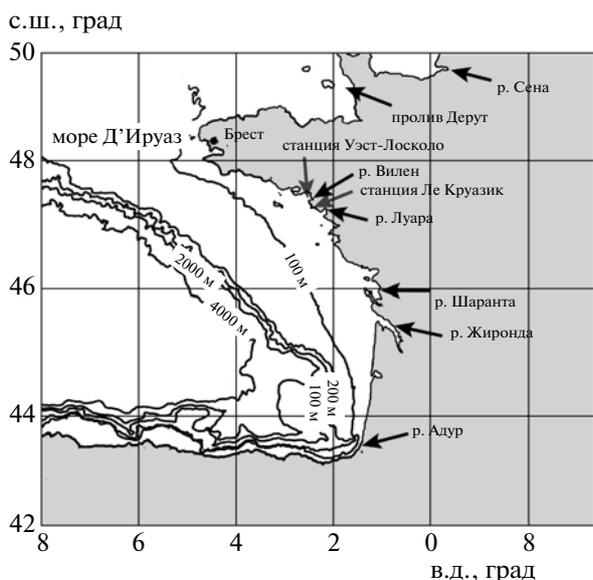


Рис. 1. Батиметрия Бискайского залива, индикация некоторых ориентиров и расположение станций наблюдений.

лизация алгоритма кластеризации по методу нечетких k -средних сводится к выполнению следующих шагов:

– на шаге 1 устанавливаются параметры алгоритма: c – количество кластеров, m – экспоненциальный вес, ε – параметр сходимости итерационного алгоритма;

– шаг 2: случайным образом генерируется матрица нечеткого разбиения F , удовлетворяющая условиям (6), (7);

– на шаге 3 рассчитываются центры кластеров:

$$V_i = \frac{\sum_{k=1}^c (\mu_{ki})^m x_k}{\sum_{k=1}^c (\mu_{ki})^m}, \quad i = 1, \dots, c;$$

– на шаге 4 рассчитываются расстояния между объектами из X и центрами кластеров:

$$D_{ki} = \sqrt{\|X_k - V_i\|^2}, \quad k = 1, \dots, M, \quad i = 1, \dots, c;$$

– шаг 5: пересчитываются элементы матрицы нечеткого разбиения ($k = 1, \dots, M, i = 1, \dots, c$)

$$\mu_{kj} = \frac{1}{\left(D_{ik}^2 \sum_{j=1}^c \frac{1}{D_{jk}^2} \right)^{\frac{1}{m-1}}}, \quad \text{если } D_{ki} > 0,$$

$$\mu_{kj} = \begin{cases} 1, & j = i, \\ 0, & j \neq i \end{cases}, \quad \text{если } D_{ki} = 0, \quad j = 1, \dots, c;$$

– на шаге 6 проверяется выполнение условия $\|F - F\| < \varepsilon$, где F – матрица нечеткого разбиения на предыдущей итерации алгоритма. Если условие выполняется, то осуществляется переход к шагу 7 – разбиение на кластеры состоялось, в противном случае – к шагу 3 – к следующей итерации;

– шаг 7 – окончание процедуры.

Самым важным параметром в алгоритме кластеризации является количество кластеров. В нашем случае это количество определено априори постановкой задачи и равно двум.

Вторым параметром алгоритма кластеризации является экспоненциальный вес (m). Чем больше m , тем конечная матрица нечеткого разбиения F становится более “размазанной”, и при $m \rightarrow \infty$ она принимает вид $F = [1/c]$, что является очень плохим решением, т.к. все объекты оказываются принадлежащими ко всем кластерам с одной и той же степенью. Кроме того, экспоненциальный вес позволяет при формировании координат центров кластеров усилить влияние объектов с большими значениями степеней принадлежности и уменьшить влияние объектов с малыми значениями степеней принадлежности. На сегодня не существует теоретически обоснованного правила выбора значения экспоненциального веса. Обычно устанавливают $m = 2$.

ХАРАКТЕРИСТИКА ДАННЫХ *IN SITU*

Собранные в разные месяцы с 2001 по 2008 г. и предоставленные нам коллегами из IFREMER (Франция) данные *in situ* в количестве 47 измерений относились к двум станциям – Уэст-Лосколо (47.46° с.ш., 2.54° в.д.) и Ле Круазик (47.3° с.ш., 2.51° в.д.). Обе станции расположены у побережья Бискайского залива в районе дельт р. Вилен и р. Луара (рис. 1).

Станционные данные включали в себя идентификацию таксона *L. chlorophorum* и количество клеток в единице объема водной пробы. При этом лишь 12 из них по времени совпадали с пролетами MODIS/Aqua (синхронные и квазисинхронные с пролетом спутника станционные данные (С-КС-данные)).

С целью увеличения объема данных для статистически обеспеченной тренировки нейронного-сетевого алгоритма, была применена следующая процедура: для каждого района, содержащего С-КС-данные создавалось RGB-изображение (путем сочетания в определенных пропорциях сигналов в красной, зеленой и синей областях спектра). Пикселы, непосредственно окружающие пиксел с С-КС-данными, принимались как области цветения *L. chlorophorum*, если имели одинаковый с содержащим С-КС-данные пиксе-

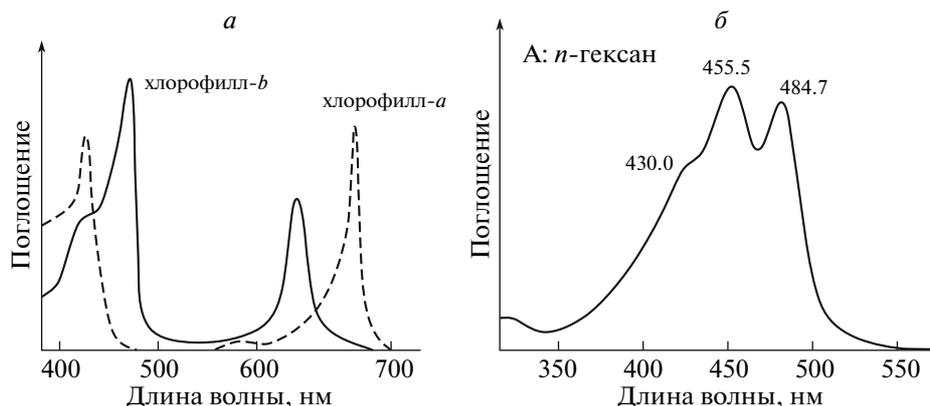


Рис. 2. Спектры поглощения: а – хл-а и хл-б; б – перидинина в стандартных растворах (Jeffrey et al., 1997).

лом цвет и идентичные/близкие по интенсивности и спектральному ходу значения $R_{rsw}(\lambda)$. Спектры в пикселах, не отвечающие этому критерию, рассматривались как статистическая база, характеризующая цветения иные, чем *L. chlorophorum*. В результате была создана база данных, состоящая из 6365 спектров $R_{rsw}(\lambda)$, из которых для тренировки сети и ее тестирования было соответственно 4240 и 2125 спектров $R_{rsw}(\lambda)$, из них 1688 (26.5%) полагались принадлежащими *L. chlorophorum* и 4677 – другим таксонам фитопланктона.

Тренировка сети осложнялась тем, что, судя по С-КС-данным, во многих случаях цветения фитопланктона не были монотаксонными и, наряду с *L. chlorophorum*, включали и цветения других водорослей, что, естественно, отражалось на спектральном составе регистрируемого $R_{rsw}(\lambda)$. Для преодоления этой сложности обучающая и тестирующая базы данных по $R_{rsw}(\lambda)$ подвергались предварительной обработке, суть которой описана в следующем разделе.

ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА СПУТНИКОВЫХ ДАННЫХ

Как отмечалось в разделе, посвященном общей характеристике *L. chlorophorum*, характерной/отличительной особенностью пигментной системы этого планктонного таксона является наличие перидинина, а в хлорофилльном комплексе, помимо хл-а, также и хл-б, спектры поглощения которых в растворах приведены на рис. 2а и б.

Учитывая, что центры спектральных каналов сканера MODIS/Aqua расположены при длинах волн 412, 443, 488, 531, 551 и 667 нм, следует ожидать, что в случае визирования области с преимущественным цветением *L. chlorophorum*, в спектре $R_{rsw}(\lambda)$ на канале 3 (488 нм) должен наблюдаться локальный минимум умеренной глубины, обусловленный длинноволновым крылом полосы

поглощения хл-б (*in vivo*, ее максимум располагается при ~470 нм (Soohoo et al., 1986; Hoepffner, Sathyendranath, 1993)), а также длинноволновой полосой поглощения перидинина (~485 нм) (рис. 2). В случае, если в цветении в визируемом пикселе *L. chlorophorum* присутствует, но ее доля не является доминирующей, то, по-видимому, в спектре $R_{rsw}(\lambda)$ на канале 3 можно ожидать, по крайней мере, отсутствие локального перегиба, при котором частная производная $\partial R_{rsw}(\lambda)/\partial \lambda$ инвариантна, или ее значение не уменьшается при переходе через канал 3. Именно такие спектры $R_{rsw}(\lambda)$ часто наблюдаются для Бискайского залива при цветении водорослей, не содержащих хл-б.

Основываясь на этом предположении, все спектральные данные по $R_{rsw}(\lambda)$ предварительно фильтровались с исключением пикселей, не удовлетворяющих указанному критерию. Это было сделано как для последующего обучения и тестирования НСА, так и алгоритма, основанного на нечеткой логике *k*-среднего. В последнем случае это было также оправдано, поскольку мы искусственно задавали разделение только на принадлежность двум классам: 1) *L. chlorophorum* и 2) не *L. chlorophorum*, – в то время как воды Бискайского залива в оптическом отношении отличаются весьма значительной сложностью и объективно разделение должно было бы производиться на большее число классов/кластеров.

Следует подчеркнуть, однако, что этот критерий – $(R_{rsw,3} - R_{rsw,2})/(\lambda_3 - \lambda_2) < (R_{rsw,4} - R_{rsw,3})/(\lambda_4 - \lambda_3)$ – не является самодостаточным хотя бы потому, что такая локальная спектральная особенность присуща и таким водорослям, как *Chlorophyta* и *Euglenophyta*, пигментная система которых также содержит *chl-b* (Kirk, 1983). Правда, в случае Бискайского залива подобной неопределенности не возникает за отсутствием таких водорослей в его водах. Аналитическим достоинством обоих алгоритмов, основанных как на нейронно-

Таблица 1. Сравнение определений *in situ* по двум станциям с результатами идентификации водоросли *L. chlorophorum* двумя алгоритмами НСА и АКС

Станция	Дата	Концентрация <i>L. chlorophorum</i> (кл/мл), определенная <i>in situ</i>	Алгоритм	
			НСА	АКС
Уест-Лосколо	09.09.2002	7200	0	0
То же	24.09.2002	380000	1	1
»	07.06.2004	34000	1	1
»	19.09.2005	400000	1	1
»	03.10.2005	200000	1	1
»	07.11.2006	66000	0	0
»	29.05.2007	8600	0	0
»	12.06.2007	160000	1	1
»	24.07.2007	900000	1	1
Ле-Крузик	09.07.2007	7450000	1	0
То же	06.08.2007	28800	1	1
»	15.07.2008	6200	0	1

Примечание. 1 и 0 обозначают, что водоросль идентифицирована и не идентифицирована соответственно.

сетевом методе, так и нечеткой логике k -среднего, является то обстоятельство, что, в дополнение к указанному критерию, они оперируют со всеми спектральными особенностями $R_{rsw}(\lambda)$, регистрируемыми в имеющихся каналах дистанционного датчика, чем и обеспечивается эффективность их работы.

РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ АЛГОРИТМОВ К СПУТНИКОВЫМ ДАННЫМ: ВНУТРИГОДОВАЯ И МЕЖГОДОВАЯ ИЗМЕНЧИВОСТЬ ЦВЕТЕНИЯ *L. CHLOROPHORUM*

“Точечные” идентификации

В табл. 1 приведены результаты сравнения дистанционной идентификации *L. chlorophorum* с данными *in situ* на станциях Уест-Лосколо и ЛеКрузик с применением НСА и АКС. Как видно, во всех случаях высоких концентраций клеток *L. chlorophorum* оба алгоритма безошибочно идентифицируют присутствие этого таксона. В случаях, когда концентрация этой водоросли невысока или относительно невысока (и *L. chlorophorum*, по-видимому, является маргинальной по оптическому влиянию водорослью в фитопланктонном комплексе), точность идентификации снижается, что естественно объяснить подавлением спектральных сигнатур *L. chlorophorum* сигнатурами таксонов-доминантов в анализируемом спектре $R_{rsw}(\lambda)$.

Таким образом, можно ожидать, что в случае, когда *L. chlorophorum* оказывается доминантной водорослью в области цветения, оба алгоритма — НСА и АКС — должны с большой долей вероят-

ности идентифицировать этот таксон и аналогичным образом оконтуривать область цветения. Это предположение полностью подтвердилось в ходе дальнейших исследований.

Площадные идентификации

Разработанные НСА и АКС были применены к обработке изображений MODIS/Aqua для ряда районов западного атлантического побережья Франции на моменты, или месяцы, когда сообщалось о факте цветения *L. chlorophorum* (Gohin, 2009). В качестве примера на рис. 3 приведены результаты идентификации и оконтуривания цветения *L. chlorophorum* в море Д’Ируаз (ориентир — г. Брест на рис. 1). Выбор района и времени (сентябрь 2006 г.) был продиктован сообщениями в прессе о мощном цветении этой водоросли (“Le Télégramme”, 28 September 2006; в статье Stepane Jézéquel “Mer d’Iroise/ Elle vire ou vert!” (Море Д’Ируаз. Оно становится зеленым!). Применение ранее разработанного нами алгоритма восстановления концентрации хл-*a* (Morozov et al., 2010) выявляет довольно сложную картину пространственного распределения фитопланктона на 24.09.2006 в море Д’Ируаз (рис. 3а). Включение НСА и АКС для обработки этого изображения однозначно выявляет (рис. 3б) наличие довольно обширной области цветения *L. chlorophorum*, о котором сообщалось в прессе.

Рамки настоящей публикации не позволяют проиллюстрировать другие примеры применения НСА и АКС. Можно лишь отметить, что нами были выявлены и цветения в некоторых районах за пределами Бискайского залива, например, в авгу-

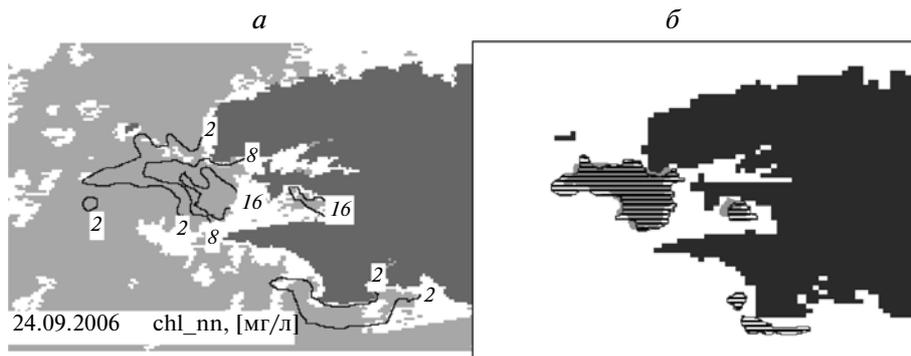


Рис. 3. Результаты обработки данных MODIS/Aqua: *a* – пространственное распределение хл-*a*; *b* – область цветения *L. chlorophorum* в море Д’Ируаз на 26.10.2006 г., выделенная с применением НСА (черная штриховка) и АКС (темно-серая область).

сте 2009 г. в районе дельты р. Сены, которые получили подтверждения в устных сообщениях (Gohin, 2009).

В свете полученных результатов и в отсутствии данных *in situ* о действительных границах цветения исследуемой водоросли представляется целесообразным производить усреднение контуров цветения, определенных обоими алгоритмами. Можно надеется, что такой конечный результат будет более адекватен реальному пространственному распределению цветения *L. chlorophorum*.

Ретроспективный анализ частотности и пространственной протяженности цветений L. chlorophorum в период с 2002 по 2009 г.

Нами были проанализированы результаты применения разработанных алгоритмов к данным MODIS/Aqua в период 2002–2009 гг. к району исследования. В результате было выявлено семь районов, в которых идентифицировалась *L. chlorophorum* (табл. 2). За исключением моря Д’Ируаз и южной части пролива Дерут (рис. 1), остальные области оказались приуроченными к дельтам рек

и прилегающим к ним районам. Условно эти районы названы по впадающим в них рекам.

Как следует из табл. 2, в двух районах (пролив Дерут и море Д’Ируаз, в которые не поступают речные воды) цветение *L. chlorophorum* происходило преимущественно в последние годы. Спутниковые данные свидетельствуют о наличии двух отчетливых периодов – весеннего и летне-осеннего – с продолжительностью не более месяца. При этом наиболее значительные площади наблюдались в летне-осенний период. В областях-реципиентах речных вод, исследуемое явление наблюдается ежегодно, при этом период цветения практически непрерывен в среднем с апреля-мая по сентябрь–октябрь (см. также табл. 1). Здесь наблюдается очень сложная динамика размеров площади развития *L. chlorophorum* с незначительными пиками в указанный период. Примечательно, что размеры цветения вдоль побережья Бискайского залива сокращаются с севера к югу. Это особенно проявляется в районе р. Адур. Однако и для этого района, начиная с 2008 г., наблюдается явная динамика в сторону расширения области цветения этой водоросли и увеличения периода ее раз-

Таблица 2. Выявленные по спутниковым данным районы и годы развития *L. chlorophorum* вдоль западного побережья Франции, представленного на рис. 1

Район	Год							
	2002	2003	2004	2005	2006	2007	2008	2009
р. Сена	++	++	++	++	++	++	++	++
пролив Дерут	–	++	–	–	–	+	+	+
море Д’Ируаз	–	–	–	–	++	++	++	–
р. Луара	+	+	+	+	+	+	+	+
р. Шаранта	++	+	+	++	+	++	++	++
р. Жиронда	+	+	+	+	+	+	+	+
р. Адур	–	+	+	+	+	+	++	++

Примечание. + и – означают наличие или отсутствие развития популяции исследуемой водоросли соответственно; ++ – обширное цветение.

вития. Анализ совокупности полученных рядов позволяет говорить об общем постепенном усилении исследуемого явления к 2007–2009 гг. Следует при этом оговориться, что речь идет лишь о расширении с годами площади развития этой водоросли в каждом из исследованных районов, но не интенсивности цветения, поскольку разработанные нами алгоритмы предназначены исключительно для идентификации и оконтуривания *L. chlorophorum*, а не для определения ее концентрации. Тот факт, что наличие этого динофлагеллята определяется нами ежегодно в областях-реципиентах речных вод и в течение практически всего вегетационного периода, свидетельствует лишь, что *L. chlorophorum* входит в состав фитопланктонного комплекса в этом морском регионе и не более того. Косвенно неперемное присутствие этого таксона в фитокомплексе Бискайского залива подтверждается и литературными данными: при практически круглогодичном доминировании диатомовых также неизменно определяются и динофлагелляты (без указания конкретных таксонов); их концентрации, существенно варьируя в течение года, наиболее значительны с мая по сентябрь (Gohin et al., 2003; Lavender et al., 2008).

ЗАКЛЮЧЕНИЕ

Таким образом, можно констатировать, что разработанные нами алгоритмы идентификации и оконтуривания областей цветения *L. chlorophorum* продемонстрировали свою эффективность в условиях Атлантического побережья Франции — Бискайского залива и далее на север вплоть до дельты р. Сены. Поскольку оба алгоритма, НСА и АКС, основаны на отыскании в спектре дистанционно регистрируемого сигнала (R_{rsw}) сигнатур, специфически характерных для этой конкретной водоросли, можно ожидать, что разработанный подход может быть применен и для других водных объектов, в которых возникают массовые цветения *L. chlorophorum*. Поскольку оба алгоритма обеспечивают весьма близкие результаты, появляется возможность применения ансамблевого подхода к обработке данных MODIS/Aqua с целью повышения надежности идентификации и оконтуривания областей цветения *L. chlorophorum*.

Проведенный нами ретроспективный анализ спутниковых данных свидетельствует об определенном нарастании частотности возникновения и размеров площади цветения этой водоросли. Об устойчивости этой тенденции можно будет говорить более уверенно с дальнейшим продолжением временного ряда спутниковых наблюдений.

В заключение выражаем свою благодарность сотрудникам IFREMER в лице д-ра Франсиса Гоина (Francis Gohin) за предоставленные данные

in situ и сообщения во французской прессе, касающиеся цветений *L. chlorophorum*.

СПИСОК ЛИТЕРАТУРЫ

- Олдендерфер М.С., Блэшфилд Р.К. Факторный, дискриминантный и кластерный анализ / Под. ред. И.С. Енюкова. М.: Финансы и статистика, 1989. 215 с.
- Петрушко И.М., Кузнецов Л.А., Прохоренко В.И., Сафонов В.Ф. Курс высшей математики: Интегральное исчисление. Функции нескольких переменных. Дифференциальные уравнения. М.: МЭИ, 2002. 340 с.
- Штовба С.Д. Введение в теорию нечетких множеств и нечеткую логику. (<http://matlab.exponenta.ru/fuzzylogic/book1/12.php>).
- Babin M., Roesler C.S, Cullen J.J. Real-time coastal observing systems for marine ecosystem dynamics and harmful algal blooms: Theory, instrumentation and modelling // Oceanographic Methodology series. Paris: UNESCO Publishing, 2008. 807 p.
- Bracher A., Vountas M., Dinter T.J., Burrows P., Ottgers R.R. Peeken I. Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data // Biogeosciences. 2009. V. 6. P. 751–764.
- Claquin P., Probert I., Lefebvre S., Veron B. Effects of temperature on photosynthetic parameters and TEP production in eight species of marine microalgae // Aquatic Microbial Ecology. 2008. V. 5. P. 1–11. doi: 10.3354/ame01187.
- Cracknell A.P., Newcombe S.K., Black A.F., Kirby N.F. The ABDMAP (Algal Bloom Detection, Monitoring and Prediction) concerted action // Int. J. Rem. Sens. 2001. V. 22. P. 205–247.
- Dembele D., Kastner, P. Fuzzy c-means method for clustering microarray data // Bioinformatics. 2003. V. 19. P. 973–980.
- Elbrächter M., Schnepf E. Gymnodinium chlorophorum, a new, green bloom forming dinoflagellate (Gymnodiniales, Dinophyceae) with avestigial prasinophyte endosymbiont // Phycologia. 1996. V. 35. P. 381–393.
- Elbrachter M. Exotic flagellates of coastal North Sea waters // Helgoland Marine Res. 1998. V. 52. № 3–4. P. 235–242.
- Gohin F., Lampert L., Guillaud J-P., Herbland A., Nezan E. Satellite and in situ observations of a late winter phytoplankton bloom in the northern Bay of Biscay // Cont. Shelf Res. 2003. V. 23. № 11–13. P. 1117–1141.
- Gohin F. 2009. Personal communication.
- Hallegraeff G.M., Anderson D.M., Cembella A.D. Manual on Harmful Marine Microalgae // IOC Manuals and Guides. № 33. Paris: UNESCO Publ., 1995. 430 p.
- Hansen G., Botes L., De Sallas M. Ultrastructure and large subunit rDNA sequences of *Lepidodinium viride* reveal a close relationship to *Lepidodinium chlorophorum* comb. nov. (*Gymnodinium chlorophorum*) // Phycological Res. 2007. V. 55. № 1. P. 25–41.
- Haykin S. Neural Networks. A Comprehensive Foundation. New Jersey: Prentice Hall Publ., 1998. 1103 p.
- Hoepffner N., Sathyendranath S. Determination of the major groups of phytoplankton pigments from the absorption

- spectra of total particulate matter // *J. Geophys. Res.* 1993. V. 98. № C12. P. 22789–22803.
- Jerome J.H., Bukata R.P., Miller J.R. Remote sensing reflectance and its relationship to optical properties of natural water // *Int. J. Rem. Sens.* 1996. V. 17. № 1. P. 43–52.
- IOCCG. Why ocean colour? The societal benefits of ocean colour technology. Rep. Int. Ocean Colour Coordinating Group № 7 (IOCCG) / Eds. T. Platt, N. Hoepffner, V. Stuart, C.W. Brown. Dartmouth, Canada: IOCCG Publ., 2008. 141 p.
- Kirk J.T.O. Light and photosynthesis in aquatic ecosystems. L.: Cambridge Press, 1983. 405 p.
- Korosov A.A., Pozdnyakov D.V., Folkestad A., Pettersson L.H., Sorensen K., Shuchman R. Semi-empirical algorithm for the retrieval of ecology-relevant water constituents in various aquatic environments // *Algorithms*. 2009. V. 2. P. 470–497. doi: 10.3390/a2010470.
- Kutser T. Passive optical remote sensing of cyanobacteria and other intense phytoplankton blooms in coastal and inland waters // *Int. J. Rem. Sens.* 2009. V. 30. № 17. P. 4401–4425.
- Lavender S.J., Raitos D. E., Pradhan Y. Variations in the phytoplankton of the north-eastern Atlantic Ocean: from the Irish Sea to the Bay of Biscay. Remote sensing of the European Seas / Eds. V. Barale, M. Gade. Ispa: JRC Publ., 2008. P. 67–76.
- Morden C.W., Sherwood A.R. Continued evolutionary surprises among dinoflagellates // *PNAS*. 2002. V. 99. № 18. P. 11558–11560.
- Morozov E., Korosov A., Pozdnyakov D., Pettersson L., Sychev V. A new area-specific bio-optical algorithm for the Bay of Biscay and assessment of its potential for SeaWiFS and MODIS/Aqua data merging // *Int. J. Rem. Sens.* 2010. (submitted).
- Passow U. Transparent exopolymer particles (TEP) in aquatic environments // *Prog. Oceanogr.* 2002a. V. 55. P. 287–333.
- Passow U. Production of transparent exopolymer particles (TEP) by phyto- and bacterioplankton // *Mar. Ecol. Prog. Ser.* 2002b. V. 236. P. 1–12.
- Phytoplankton pigments in oceanography: guidelines to modern methods / Eds. S.W. Jeffrey, R.F.C. Mantoura, S.W. Wright. Paris: UNESCO Publ., 1997. 615 p.
- Report of the ICES-IOC Working Group on Harmful Algal Bloom Dynamics. 17–20 March, 2003. Aberdeen, UK. Copenhagen: International Council of the Exploration of the Sea Publ., 2003. 37 p.
- Ross T.J. Fuzzy logic with engineering applications. N.Y.: McGraw-Hill Publ., 1995. 456 p.
- Silvey S.D. The Lagrangian Multiplier // *Ann. Math. Statist.* 1959. V. 30. № 2. P. 389–407.
- Soohee J.B., Kiefer D.A., Collins D.J., McDermid I.S. In vivo fluorescence excitation and absorption spectra of marine phytoplankton: I. Taxonomic characteristics and responses to photoadaptation // *J. Plankton Res.* 1986. V. 8. № 1. P. 197–214.
- Stumpf R.P., Culve, M.E., Tester P.A., Tomlinson M., Kirkpatrick G.J., Pederson B.A., Truby E., Ransibrahmanakul V., Soracco M. Monitoring *Karenia brevis* blooms in the Gulf of Mexico using satellite ocean colour imagery and other data // *Harmful Algae*. 2003. V. 2. P. 147–160.
- Takishita K., Kawachi M., Noel M-H., Matsumoto T., Kikizoe N., Watanabe M.M., Inouye I., Ishida K., Hashimoto T., Inagaki Y. Origins of plastids and glyceraldehydes-3-phosphate dehydrogenase genes in the green-colored dinoflagellate *Lepidodinium chlorophorum* // *Gene*. 2008. V. 410. № 1. P. 26–36.
- Weizman L., Goldberger J. Classification of hyperspectral remote-sensing images using discriminative linear projections // *Int. J. Rem. Sens.* 2009. V. 30. № 21. P. 5605–5617.
- Zibordi G., Berthon J-F., Melin F., D'Alimonte D., Kaitala S. Validation of satellite ocean color primary products at optically complex coastal sites: northern Adriatic, northern Baltic Proper and Gulf of Finland // *Rem. Sens. Environ.* 2009. V. 113. P. 2574–2591.
- Zimmermann H.-J. Fuzzy set theory and its application. Boston, Dordrecht, London: Kluwer Academic Publishers, 2001. 514 p.

Automated Identification and Delineation of *Lepidodinium Chlorophorum* Blooms Using Scanning Instruments Data

E. A. Morozov¹, O. B. Aniskina², D. V. Pozdnyakov¹, L. H. Pettersson³, V. I. Sychev, H. Grassl⁴

¹Scientific Foundation “Nansen International Environmental and Remote Sensing Center”, St. Petersburg, Russia

²Russian State Hydrometeorological University, St. Petersburg, Russia

³Nansen Environmental and Remote Sensing Center, Bergen, Norway

⁴Max Planck Institute for Meteorology, Hamburg, Germany

For identification and delineation of *Lepidodinium chlorophorum* blooms, two algorithms based on the neural network and fuzzy logic *k*-means techniques have been developed and successfully applied to MODIS/Aqua images over the western coastline of France. To increase the results reliability, these two algorithms can be employed conjointly as an ensemble tool. Both algorithms are area non-specific and applicable to other marine waters.

Keywords: MODIS, Bay of Biscay, neural network, harmful algal blooms, *Lepidodinium chlorophorum*