

ПРИМЕНЕНИЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ В МАТЕМАТИЧЕСКОМ МОДЕЛИРОВАНИИ ЭКОСИСТЕМ

Васечкина Е.Ф., Ярин В.Д.
МГИ АН Украины,
Севастополь, Капитанская, 2
evasvyar@bios.iuf.net

Введение. Рассмотрены два направления практического использования ГА в задачах экологического моделирования. Первое связано с подготовкой данных наблюдений для построения статистических моделей и последующей их верификации. Второе связано с собственно моделированием функциональных зависимостей между параметрами экосистемы с помощью нейронной сети. Интерполяция и экстраполяция данных океанографических наблюдений является по-прежнему актуальной задачей, поскольку при значительном проценте пропущенных данных до сих пор не найдено удовлетворительного способа их восстановления. В работе предлагается использовать для решения этой проблемы аппарат эмпирических ортогональных функций в сочетании с генетическим алгоритмом поиска коэффициентов разложения. При разработке математических моделей морских экосистем возникает проблема построения уравнений для гидрохимического и биологического блоков, т.е. определения функциональных зависимостей между параметрами экосистемы. Как правило, точные количественные соотношения между переменными этих блоков неизвестны, или они слишком сложны для практического использования. Поэтому для описания этих переменных при моделировании исследователи вынуждены использовать некоторые полумэмпирические соотношения, полученные путем анализа данных наблюдений [1]. Вид зависимости при этом обычно задается исследователем, а соответствующие коэффициенты оцениваются по данным. Субъективность такого подхода очевидна. Потому

в данной работе предлагается подход, основанный на применении аппарата нейронных сетей и генетических алгоритмов [2,3,4], который не требует априорных знаний о виде функции и позволяет выявлять ее по данным наблюдений.

Описание генетического алгоритма.

Генетические алгоритмы относятся к группе так называемых эволюционных алгоритмов, в основе которых лежит идея поиска решения задачи путем имитации его «биологической» эволюции. Алгоритм генерирует и поддерживает «популяцию» потенциальных решений исследуемой проблемы. Некоторые из этих решений используются для создания новых решений путем применения к ним специальных операторов, имитирующих процессы реальной биологической эволюции. Это операторы селекции, скрещивания (кроссовера) и мутации. Таким образом, в алгоритме реализуется дарвиновская триада движущих сил эволюции: отбор, наследственность, изменчивость. Потенциальные решения для оператора скрещивания отбираются на основе оценки их функции качества согласно введенным критериям, определяемым конкретной задачей. Новые генерации популяции потенциальных решений рассчитываются последовательно до выполнения алгоритмом некоторого условия, обычно связанного с глубиной найденного экстремума функции качества.

Эффективность алгоритма зависит от многих аспектов, среди которых важное место занимает формализация задачи, т.е. математическое представление ее потенциального решения – генотип экземпляра популяции. Решение этого вопроса зависит от специфики задачи, структуры имеющихся для ее решения данных и опыта исследователя. Выбранное математическое описание генотипа определяет в дальнейшем алгоритмические особенности применения эволюци-

онных операторов кроссовера и мутаций. Существует большое количество работ, в которых рассматриваются различные аспекты формализации задачи и связанные с этим проблемы. Однако есть и готовые рецепты, применимые в большинстве наиболее широко распространенных задач оптимизации. Это так называемое бинарное кодирование, кодирование в абсолютных значениях искомым величин и перестановочное кодирование. Бинарное кодирование формирует генотип экземпляра популяции в виде строки двоичных символов, в которую могут быть преобразованы как последовательность чисел с плавающей точкой, так и другие объекты, представляющие собой возможное решение задачи. В строке могут быть выделены специальные маркировочные поля, отделяющие разные объекты друг от друга и несущие дополнительную информацию о структуре потенциального решения. При кодировании в абсолютных значениях в строку генотипа записываются искомые переменные в исходном виде, это могут быть числа с плавающей запятой, строки произвольных символов или более сложные объекты. Перестановочное кодирование используется в том случае, когда необходимо решить задачу оптимизации последовательности некоторых действий. Генотип экземпляра популяции в этом случае представляет собой строку, в которую записаны порядковые номера каких-либо действий. Наиболее известные примеры из этой серии – задача коммивояжера или составление расписаний. Преобразование генотипа в фенотип с последующей оценкой его функции качества производится с помощью процедуры, определяемой каждой конкретной задачей.

Фильтрация данных и восстановление пропущенных значений. Как известно, при тестировании экологических моделей остро стоит вопрос о недостаточности данных наблюдений при кажущемся их обилии. Это связано с нере-

гулярностью и фрагментарностью измерений, в то время как модель требует определенного единообразия представления информации для ее верификации. Адекватно восстановить пропущенные данные при условии достаточности их общего количества можно, используя аппарат эмпирических ортогональных функций (ЭОФ). Ряд работ [5-8] был посвящен этой проблеме, и полученные в них результаты говорят о перспективности такого подхода. Одновременно с восстановлением пропущенной информации, использование разложения по ЭОФ позволяет естественным образом осуществлять фильтрацию данных, а также упрощает проблему экстраполяции. Упрощение связано с тем, что при разложении прогнозируемой переменной в ряд по ЭОФ происходит разделение пространственной и временной изменчивости, т.е. задача прогноза динамики двумерного поля сводится к задаче прогноза временных рядов коэффициентов разложения. Произвольная реализация наблюдаемого поля $f(\mathbf{x}, t)$ может быть представлена в виде:

$$f(\mathbf{x}, t) = f_0(\mathbf{x}, t) + \sum_k a_k(t) \psi_k(\mathbf{x}) \quad (1)$$

где ψ_k – собственные функции ковариационной матрицы

$$P(\mathbf{x}, \mathbf{y}) = \sum_i f(\mathbf{x}, t) f(\mathbf{y}, t).$$

Коэффициенты разложения $a_k(t)$ для каждой из реализаций рассчитываются с помощью свертки этой реализации с набором ЭОФ:

$$a_k(t) = \int f(\mathbf{x}, t) \psi_k(\mathbf{x}) d\mathbf{x} \quad (2)$$

Предлагается использовать ГА для поиска необходимого набора коэффициентов разложения при нескольких первых ЭОФ для реализаций, где количество данных оказывается недостаточным для получения их по формуле (2). Применяется бинарное кодирование, при котором искомые коэффициенты записываются последовательно друг за другом в двоичном виде в строку генотипа.

В начальный момент популяция генерируется как набор случайных бинарных строк заданной длины. Преобразование генотипа в фенотип (реализацию искомой характеристики морской среды) осуществляется по формуле (1). Рассмотрим здесь конкретный пример восстановления профилей различных характеристик морской среды на станциях с большим количеством пропущенных горизонтов. Отметим, что таким применением обсуждаемая технология далеко не ограничивается, в качестве «реализации» могут рассматриваться также двумерные и трехмерные поля. Основной проблемой был правильный выбор функции качества для оценки потенциальных решений. Наряду с требованием минимума среднеквадратичной ошибки аппроксимации и гладкости получаемого профиля, при решении этого вопроса мы использовали имеющуюся априорную информацию об особенностях вертикального распределения восстанавливаемых характеристик морской среды. Например, при восстановлении профилей температуры в деятельном слое в функцию качества был добавлен член, имеющий смысл теплосодержания в слое 40-80м.

Генотип (набор коэффициентов), обеспечивающий меньшее теплосодержание фенотипа в выделенном слое (наличие ХПС, характерного для Черного моря) при минимуме среднеквадратичной ошибки в имеющихся точках, имел лучшую оценку функции качества и соответственно большую вероятность быть отобранным для образования потомков в новой популяции потенциальных решений.

На рис. 1 показан типичный профиль температуры в слое 0-200 метров $T(z)$ (сплошная линия), имеющиеся для восстановления профиля данные (показаны точками), а также результаты восстановления профиля линейной и сплайновой интерполяцией и по формуле (1) с коэффициентами, рассчитанными с помощью ГА. Как видим, правильный выбор функции качества для оценивания вари-

антов решения приводит к тому, что подобранные коэффициенты разложения содержат информации о перегибе профиля температуры (наличие ХПС), которой не было в данных, использованных для интерполяции.

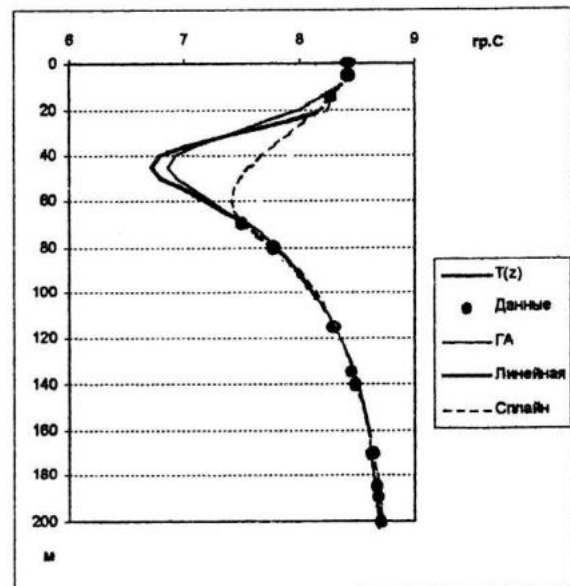


Рис. 1. Пример восстановления профиля температуры воды в деятельном слое Черного моря разными методами.

Тестирование метода производилось с помощью массива профилей температуры и содержания кислорода, в каждый из которых были внесены искусственные пропуски. Подготовленные таким образом данные затем восстанавливались с помощью линейной и сплайновой интерполяции, а также ГА. Для каждой реализации рассчитывалась абсолютная и относительная ошибки интерполяции. Осредненные по ансамблю ошибки заполнения пропусков в профилях двух характеристик морской среды представлены в таблице 1.

Второе направление использования ГА связано с построением архитектуры нейронной сети, с помощью которой строится эмпирическая модель взаимодействия между параметрами экосистемы.

	30%			50%			70%		
	ГА	ЛИ	СИ	ГА	ЛИ	СИ	ГА	ЛИ	СИ
Температура (°C)	0,068	0,081	0,097	0,098	0,155	0,172	0,106	0,575	0,942
	0,062	0,068	0,082	0,092	0,123	0,143	0,105	0,472	0,792
Кислород (мл/л)	0,147	0,211	0,508	0,266	0,301	0,800	0,604	0,935	2,455
	0,093	0,146	0,291	0,150	0,171	0,544	0,341	0,688	1,824

Таблица 1. Результаты тестирования метода восстановления пропущенных данных (абсолютная и относительная ошибки) с помощью ГА, линейной (ЛИ) и сплайновой (СИ) интерполяции. В процентах обозначено количество пропущенных отсчетов.

Построение архитектуры полиномиальной нейронной сети.

Архитектура нейронной сети строится последовательным наращиванием числа слоев, причем для каждого узла конфигурация его связей с другими узлами определяется с помощью генетического алгоритма. В алгоритме используется стандартный ГА, поддерживавший популяцию индивидуумов, каждый из которых представляет собой возможную комбинацию связей одного из узлов в слое, начиная с первого (считая набор входных узлов за нулевой слой), с узлами предыдущего слоя сети. Генотип индивидуума задается в виде битовой строки $L = (001100010...01)$, где "1" обозначает наличие связи данного узла с соответствующим узлом предыдущего слоя, а "0" – ее отсутствие. Схема кодирования связей узлов в сети представлена на рис.2. Для каждого узла сети запускается генетический алгоритм поиска оптимальной конфигурации связей этого узла с нейронами предыдущего слоя. Каждый индивидуум популяции оценивается по некоторой функции качества, характеризующей «удачность» подобранной конфигурации связей.

Для определения этой функции качества представим данные для обучения сети в виде таблицы, столбцы которой представляют собой последовательность соответствующих точек входных и выходных параметров системы, соответствующие друг другу. Это могут быть, например, временные ряды составляющих вектора состояния исследуемой динамической системы. Строку такой таблицы, представляющую собой согласованный

набор значений параметров в один момент времени, будем называть точкой таблицы данных. Обозначим Z_k последовательность значений точного решения задачи, т.е. выходного параметра системы. Соответствующий им вектор значений входных параметров обозначим U_k , где индекс k – номер строки в таблице данных.

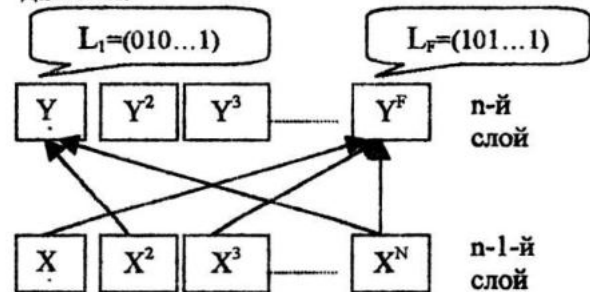


Рис.2. Схема кодирования структуры связей нейронов между слоями в генотипе L.

Рассмотрим последовательность значений на выходе n -го узла сети в слое m как некоторое частное решение нашей задачи

$$Y_{nk} = a_{n0} + \sum_i^F a_{ni} X_{ik} + \sum_i^F \sum_j^F a_{nij} X_{ik} X_{jk} + \dots \quad (3)$$

где F – количество узлов в слое $m-1$;
 a_{n0} – пороговое значение нейрона;
 a_{ni} – весовые коэффициенты входных узлов;

X_{ik} – числовые значения, полученные на выходе нейронов предыдущего слоя сети, с которыми данный узел должен быть связан в соответствии с генотипом L.

Такое представление является достаточно универсальным, поскольку соответствует разложению в ряд Тейлора. Если ограничиться в правой части ли-

нейными членами, получим обычное уравнение множественной регрессии. Учет членов разложения более высокого порядка существенно расширяет пространство поиска (удлиняется генотип L), но появляется возможность учета вторых, третьих и т.д. моментов распределения величин входных параметров при построении модели выходного параметра системы. Практическое применение алгоритма показало, что для многих задач идентификации сложных динамических систем, особенно, природных, линейного представления оказывается недостаточно и требуется более полное описание выходного сигнала нейрона с учетом по крайней мере квадратичных членов.

Для того, чтобы вычислить весовые коэффициенты в формуле (3) потребуем, чтобы значения на выходе нейрона Y_{nk} были как можно ближе к точным значениям искомого параметра Z_k , заданным для обучения сети, по критерию минимума среднеквадратичной ошибки. Подставляя в (3) данные из выборки и осредняя полученные соотношения, получим линейную систему нормальных уравнений. Решением системы нормальных уравнений является набор весовых коэффициентов, соответствующих связям, отмеченным "1" в генотипе L . Заметим, что при оценке коэффициентов первого слоя X_{ik} интерпретируются как числовые значения входных узлов сети U_{ik} , а в последующих слоях X_{ik} - это последовательность значений, полученных на выходе нейронов предыдущего слоя сети. Таким образом, таблица данных для обучения сети при наращивании числа слоев постепенно переписывается с использованием ранее полученных в узлах сети значений. Многослойность алгоритма позволяет использовать рекурсию при построении модели и многократно использовать одни и те же точки выборки данных для все более точной подгонки уравнения множественной регрессии.

Для того чтобы обеспечить хорошие обобщающие свойства построенной сети

(т.е. ее способность к экстраполяции), необходимо оценивать эффективность подбора конфигурации связей нейрона на выборке, которая не использовалась для идентификации весовых коэффициентов. Поэтому вся имеющаяся для обучения сети выборка данных должна быть поделена на три части: обучающую, контрольную и экзаменационную последовательности, причем контрольная должна быть не меньше обучающей. По методу наименьших квадратов на обучающей части выборки определяются весовые коэффициенты, соответствующие "1" в генотипе L . Для этого строится и решается линейная система нормальных уравнений. На контрольной части выборки оценивается функция качества такого частного решения (фенотипа индивидуума популяции) как среднеквадратическая ошибка этого решения относительно известных величин выходных параметров Z_k . Эта процедура повторяется для каждого индивидуума популяции, и в результате применения стандартных операторов кроссовера, мутации и селекции через определенное небольшое число поколений отыскивается некий локальный оптимум конфигурации связей данного узла. Поскольку для каждого из узлов требовалось отыскать лишь локальный оптимум связей, ГА поддерживал небольшую численность популяции (≤ 150) и ограниченное число поколений (≤ 30).

Повторив эту процедуру для каждого из узлов, количество которых ограничивается некоторым максимальным числом, получим конфигурацию потенциальных связей данного слоя с предыдущим. То же самое повторяется в последующих слоях сети, число которых также ограничивается максимально допустимым. Для слоя в целом вычисляется критерий качества, состоящий из двух величин: среднее значение функции качества в узлах и ее минимальное значение в слое. В последнем слое сети имеется по-прежнему набор узлов, каждый из которых представляет собой вариант

решения задачи с соответствующей ему сетевой архитектурой. Окончательный выбор решения осуществляется на экзаменационной последовательности, которая до этого момента никак не использовалась в процедуре поиска. Последовательность Y_{mk} , дающая минимальную ошибку относительно Z_k на экзаменационной последовательности, определяет выбор выходного узла с номером m , а вместе с ним и окончательный выбор архитектуры сети.

Для запуска процедуры самоорганизации структуры нейронной сети необходимо определить следующие параметры:

- число входных узлов и соответствующую им таблицу данных, на которой будет обучаться сеть;
- максимальные размеры сети, т.е. максимально допустимое число слоев и нейронов в каждом слое;
- численность популяции и максимально допустимое число поколений генетического алгоритма.

Процесс самоорганизации продолжается до достижения минимума среднеквадратичной ошибки, полученной в узлах слоя, либо некоторого ее постоянно не уменьшающегося значения.

Результаты практического применения алгоритма. Техника, аналогичная описанному выше алгоритму, но в упрощенном виде, уже применялась авторами ранее для решения задач идентификации моделей морских экологических систем. Полученные результаты были достаточно интересными, чтобы продолжить разработки в этом направлении. Здесь мы также приведем два примера применения алгоритма в задачах, связанных с идентификацией зависимостей между различными параметрами морской среды. Первый пример – это тестовая задача по идентификации моделей двух параметров экосистемы озера Эри [9]. Вторая задача связана с

оценкой среднего содержания в 100-метровом слое Черного моря кислорода, биогенных элементов и хлорофилла по имеющимся данным на поверхности.

Идентификация моделей компонентов экосистемы озера Эри.

Эта тестовая задача была взята из базы тестовых задач, собранных на сайте Интернет (www.esat.kuleuven.ac.be) [9] для сравнительного анализа различных алгоритмов. Входные параметры представляли собой временные ряды среднемесячных значений температуры воды, теплопроводности, щелочности, содержания NO_3 и жесткости воды. На выходе сеть должна была генерировать значения растворенного в воде кислорода и биомассы водорослей. Выборка данных этой тестовой задачи содержала ряды точных значений перечисленных параметров и зашумленные ряды с 10%, 20% и 30% уровнем шума. Таким образом, на этой задаче можно было протестировать алгоритм на помехоустойчивость.

В сформированную для обучения сети таблицу данных были добавлены ряды входных параметров с временными сдвигами до -5. Таким образом, таблица данных состояла из 30 входных параметров и двух выходных. Все ряды были нормализованы и приведены к интервалу (0,1). Длина выборки составляла 52 точки, которые распределялись следующим образом: по 23 точки на обучающую и контрольную и 6 точек на экзаменационную последовательности. С точными рядами данных достаточно хорошо работал линейный вариант алгоритма. Результаты применения алгоритма суммированы в таблице 2. Функция качества оценивалась как относительная среднеквадратичная ошибка аппроксимации искомого параметра на экзаменационной последовательности. В таблице 2 указаны значения параметров нейронной сети, полученные осреднением по всем выполненным расчетам.

Компоненты экосистемы	Число слоев			Число нейронов			Число связей			Функция качества
	макс.	мин.	сред.	макс.	мин.	сред.	макс.	мин.	сред.	
Содержание кислорода	5	4	4,2	48	27	39,2	263	81	205,8	0,08
Биомасса водорослей	4	3	3,6	66	33	47,2	514	57	266,2	0,10

Таблица 2. Средние значения параметров архитектуры нейронной сети, полученные для тестовой задачи идентификации переменных экосистемы озера Эри.

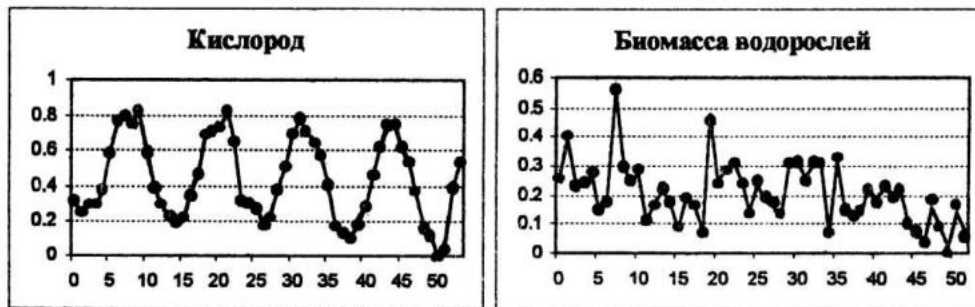


Рис. 3. Результаты аппроксимации двух параметров экосистемы озера Эри (точками показаны точные данные, линией – математическая модель).

Лучшие из полученных вариантов эмпирических зависимостей показаны на рис.3. Относительная среднеквадратическая ошибка составляет 7,5% для содержания кислорода и 9,8% для биомассы водорослей. Ошибка аппроксимации с помощью нейронной сети, обученной на зашумленных данных, была значительно выше и достигала 15-20% при уровне шума в данных 10% и 25-50% при зашумленности данных 30%. При использовании зашумленных данных лучшие результаты достигались при включении в частные решения квадратичных членов разложения (3).

Численный эксперимент с натурными данными в деятельном слое Черного моря.

Целью численного эксперимента был поиск эмпирических зависимостей между поверхностными и средними значениями гидрохимических и биологических переменных в деятельном слое Черного моря. В расчетах использовались данные натурных измерений, выполненных в разные годы судами МГИ и ИНБЮМ НАН Украины. По банку дан-

ных МГИ НАН Украины была сделана выборка гидрологических станций, на которых измерялись одновременно гидрологические, химические и биологические параметры морской среды. Также привлекались данные метеорологических наблюдений на судах, выполненных в то же время. Черное море условно делилось на три района: северо-западный шельф, ограниченный параллелью $44^{\circ}36'$ с.ш., западный и восточный районы, разделенные меридианом 35° . Отобранные станции группировались по этим условным районам и сезонам года для формирования статистического ансамбля реализаций. Под реализацией здесь понимается одномоментное вертикальное распределение какой-либо из переменных. Сформированный таким образом ансамбль реализаций использовался для расчета первого и второго моментов распределения вертикальных профилей различных компонент морской среды. По ковариационной матрице затем рассчитывались эмпирические ортогональные функции, которые использовались для сглаживания и восстановления пропущенных значений в данных.

Подготовленные таким образом вертикальные профили переменных осреднялись в слое 0-100м, полученные величины включались в вектор состояния системы, который в данном случае определялся как набор переменных, условно подразделяющихся на входные и выходные. Под входными подразумевались переменные, которые можно легко измерить (значения на поверхности) либо рассчитать по гидродинамической модели, а также метеорологические параметры. Выходными или искомыми переменными были средние в слое 0-100м значения химических и биологических компонент: содержания растворенного кислорода, фосфатов, хлорофилла "а". Таблица данных для обучения НС формировалась из столбцов, представляющих собой значения компонент вектора состояния системы. Строка таблицы объединяла значения этих компонент на одной станции, включенной в статистический ансамбль.

Применение описанного выше алгоритма позволило выделить некоторые неочевидные функциональные зависимости между компонентами вектора состояния. Для примера выпишем полученное уравнение для среднего в деятельном слое содержания растворенного кислорода, дающее на экзаменационной последовательности абсолютную среднеквадратическую ошибку 0,23мл/л и относительную – 22%:

$$O_{cp} = 27.6550 - 0.0275T_{cp}^2 - 0.0588S_{cp}^2 - 6.88 \cdot 10^{-4}T_n S_{cp} - 0.65 \cdot 10^{-4}V^3 - 1.42 \cdot 10^{-4}V^2(T_a - T_w)^2 \quad (4)$$

где O_{cp} – среднее содержание кислорода в слое 0-100м, T_n , S_n , T_{cp} , S_{cp} – температура и соленость на поверхности и средние в слое соответственно, V – скорость ветра, $(T_a - T_w)$ – разница температур воздуха и воды.

Эта функциональная зависимость была найдена при обучении НС на данных за весенний период, полученных в

западном глубоководном районе Черного моря. Таблица данных состояла из 97 точек, 20 из них использовались как экзаменационная последовательность.

При попытке использовать это соотношение для другого сезона года (осени) выяснилось, что модель хорошо работает, но требует небольшой подгонки коэффициентов. На рис. 4 слева показаны результаты аппроксимации точек экзаменационной последовательности, т.е. части данных за весенний период. Справа – аппроксимация реальных данных за осенний период. По оси абсцисс отложены реальные данные, по оси ординат – модельные. Сплошной линией обозначено геометрическое место точек с нулевой ошибкой аппроксимации. Выборка осенних данных имела длину 50 точек, среднеквадратическая ошибка аппроксимации составила 30% от стандартного отклонения искомого ряда.

При обучении НС на выборке осеннего периода было получено отличающееся от (4) соотношение:

$$O_{cp} = 48.1743 - 2.1528S_{cp} - 0.0017T_n^2 + 0.0057T_n T_{cp} + 0.0072T_n O_n - 0.0407T_{cp} \sigma_n \quad (5)$$

где σ_n и O_n – условная плотность воды и содержание кислорода на поверхности.

При использовании этого соотношения для аппроксимации данных за весенний период также потребовалась небольшая коррекция коэффициентов, среднеквадратическая ошибка составила 34% на осенних данных и 33% на весенних.

Приблизительное равенство ошибок свидетельствует о хороших обобщающих возможностях полученной математической модели. Неоднозначность полученных решений (4) и (5), видимо, объясняется тесной нелинейной зависимостью между переменными вектора состояния, что приводит к их взаимозаменяемости, а также тем, что в разные сезоны года меняется приоритет факторов, воздействующих на содержание кислорода в воде.

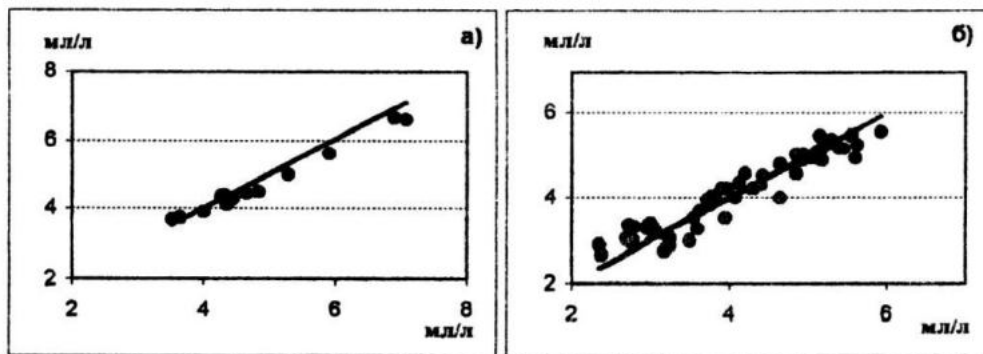


Рис. 4. Результаты математического моделирования содержания растворенного кислорода: а) аппроксимация данных на экзаменационной последовательности (весна); б) аппроксимация данных за осенний период с помощью соотношения, полученного на выборке за весенний период.

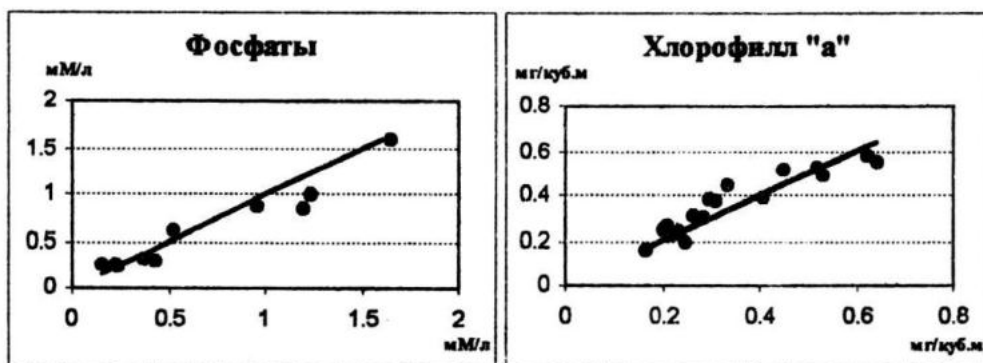


Рис. 5. Результаты математического моделирования содержания фосфатов и хлорофилла "а" на экзаменационной последовательности (весенний период).

В данном случае, в весенний период большее влияние оказывает воздухообмен с атмосферой, тогда как осенью среднее содержание кислорода в воде зависит от свойств воды (ее температуры и солености) и существенно меньше от взаимодействия с атмосферой. В приведенных соотношениях нет "расходной части", т.е. зависимости от компонент системы, влекущих расход кислорода, и это вносит определенную ошибку в математическую модель. Мы сознательно исключили эти переменные из обучающей таблицы данных, поскольку их значения сами определяются содержанием кислорода как лимитирующим фактором.

Для фосфатов и хлорофилла "а" также были рассчитаны эмпирические соотношения. На рис. 5 приведены результаты аппроксимации с их помощью ре-

альных величин на экзаменационной последовательности (среднеквадратическая ошибка составила $0,15\mu\text{M}/\text{л}$ или 34% для фосфатов и $0,05\text{м.л./л}$ или 15% для хлорофилла "а"). Нейронная сеть обучалась на выборке данных за весенний период.

Заключение. Рассмотренная задача восстановления (интерполяции и экстраполяции) вертикальных профилей гидрохимических и гидробиологических элементов по неполным данным с использованием генетического алгоритма показала, что предлагаемый метод обладает существенными преимуществами перед традиционными. Метод поиска функциональных зависимостей между компонентами вектора состояния сложной природной системы даже на ограниченных выборках данных позволяет находить решение с ошибкой, не превы-

шающей 25-35% от среднеквадратического отклонения искомой переменной. Он может с успехом применяться в тех случаях, когда необходимо оценить распределение ненаблюдаемой гидрохимической или биологической переменной состояния по имеющейся гидрометеорологической информации, а также при разработке динамических моделей сложных экологических систем.

ЛИТЕРАТУРА

1. Беляев В.И., Кондуфорова Н.В., Математическое моделирование экологических систем шельфа. - Киев, "Наукова думка", 1990, 240с.

2. Beasley D., Bull D.R., Martin R.R., An Overview of Genetic Algorithms: Part I, Fundamentals, *University Computing*, 1993, 15(2), pp. 58-69.

3. Beasley D., Bull D.R., Martin R.R., An Overview of Genetic Algorithms: Part II, Research Topics, *University Computing*, 1993, 15(4), pp. 170-181.

4. Muller B., Reinhardt J., Neural Networks. An Introduction. - Berlin, Springer-Verlag, 1991, 266с.

5. Тимченко И.Е., Ярин В.Д., Бамба
3. Прогноз температуры поверхности

океана на основе спутниковых измерений. НИЦ Конакри-Рогбане, Конакри, 1991, 42с.

6. Репинская Р.П., Бабич Я.Б. Аппроксимация рядами эмпирических ортогональных функций северополушарных полей облачности по спутниковым данным, Исследование Земли из космоса, 1999, 6, сс. 8-15.

7. Рябинин А.И., Тимченко И.Е., Васечкина Е.Ф., Игумнова Е.М., Ярин В.Д., Шибаева С.А. Прогнозирование распределения растворенного кислорода в Азовском море с применением эмпирических ортогональных функций. - "Метеорология и гидрология", 1996, 11, сс. 64-69.

8. Васечкина Е.Ф., Тимченко И.Е., Ярин В.Д. Восстановление структуры гидрофизических полей по неполной информации. - МГЖ, 1997, 2, сс. 37-44.

9. De Moor B.L.R. (ed.), DaISy: Database for the Identification of Systems, Department of Electrical Engineering, ESAT/SISTA, K.U.Leuven, Belgium, URL: <http://www.esat.kuleuven.ac.be/sista/daisy/>, 8.02.2000. Used dataset: Data of a simulation of the western basin of Lake Erie, Environmental Systems, [96-005].