

ПРОГНОЗИРОВАНИЕ ЧИСЛЕННОСТИ *MICROCYSTIS AERUGINOSA* НА ОСНОВЕ ПРАВИЛ НЕЧЕТКОЙ ЛОГИКИ И НЕЧЕТКИХ НЕЙРОННЫХ СЕТЕЙ

А.О. Гаязова, С.М. Абдуллаев

Описана процедура поиска потенциальных предикторов и создания прогнозных правил нечеткой логики и нечетких нейронных сетей для последующего прогнозирования вспышек численности синезеленой водоросли *M. aeruginosa*. В результате натурных наблюдений ряда биотических и абиотических параметров водной среды, проведенных на озере Смолино (г. Челябинск) за теплый период 2009 и 2011 года получены временные ряды численности *M. aeruginosa* и значений сопутствующих параметров. С помощью кросс-корреляционного анализа данных установлено, что потенциальными предикторами квазипериодических колебаний численности *M. aeruginosa* с периодом 12-20 дней могут выступать численность водоросли *P. duplex*, температура воды и концентрация нитрат-иона. По результатам кросс-корреляционного анализа заданы прогнозны правила и функции принадлежности в диапазоне изменений предиктанта и предиктора от нуля до 1. Для «автоматического» задания прогнозных правил и функций принадлежности с помощью специально написанной программы произведено обучение нечеткой нейронной сети на данных о значениях предиктанта и отобранных в ходе предварительного анализа параметров-предикторов. Для сравнения результатов дополнительно осуществлена линейная экстраполяция данных о численности предиктанта. Выявлено, что экстраполяционный прогноз хорошо работает на квазилинейных интервалах изменения численности, а алгоритмы нечеткой логики потенциально способны определить время наступления интенсивных вспышек численности предиктанта.

Ключевые слова: *M. aeruginosa*, прогнозирование вспышек, квазипериодические колебания, линейная экстраполяция, нечеткая логика, нечеткие нейронные сети

Введение

В мониторинге окружающей среды все большую актуальность приобретают краткосрочные и среднесрочные прогнозы массового развития макро- и микроводорослей в морских и пресноводных экосистемах. Например, прогноз численности синезеленых водорослей сроком на несколько дней необходим для оптимизации работ систем водоочистки в период массового развития водорослей в водоемах — источниках питьевого водоснабжения. Как показывает опыт [1, 2] и многих других работ наиболее адекватными являются методы прогноза, основанные на математической обработке временных рядов численности прогнозируемой водоросли с учетом ряда гидрологических и биотических факторов, характеризующих текущее состояние водоема.

Специфические особенности экосистем, ограниченность набора первичной информации о параметрах среды, очевидно, ставят вопрос о выборе оптимального метода прогноза в условиях конкретного региона. В частности в работе [1] для включения дополнительных факторов используются правила нечеткой логики, а в [2] поиск прогнозных правил производится с помощью искусственных нейронных сетей. В настоящей работе оценивается сравнительная результативность методов линейной экстраполяции, правил нечеткой логики и нечетких нейронных сетей для краткосрочного прогнозирования численности синезеленой водоросли *Microcystis (M.) aeruginosa* на основе данных натурных наблюдений за изменением значений ее численности и оптимального набора сопутствующих физико-химических и биотических параметров.

1. Данные и методы их обработки

В качестве данных для прогноза используются данные натуральных наблюдений, проведенных первым автором в летне-осенний период 2009 и 2011 годов в прибрежной части озера Смолино, г. Челябинск. В ходе мониторинга одновременно с отбором проб на фитопланктон, определялись гидрохимические показатели, фиксировались локальные гидрометеорологические условия. Методы наблюдений, отбора и обработки проб описаны в предыдущей работе авторов [3, 4]. В результате наблюдений были получены ряды данных общей длиной в 138-159 дней с разрешением во времени от 2 до 7 дней. В целях удобства математического моделирования из исходных нерегулярных рядов посредством кусочно-линейной интерполяции получены регулярные временные серии с разрешением в один день. Кроме этого, поскольку различные данные имеют весьма широкий диапазон изменений величин (от долей единиц для физико-химических параметров водной среды до сотен тысяч для численностей водорослей) поэтому предварительно проведено масштабирование рядов с нормировкой значений на максимальную величину наблюдаемую в ряде.

Как видно из рисунка 1 численность *M. aeruginosa* за 2009 (2011) год обнаруживает значительные внутрисезонные колебания амплитуды с периодом 10 (20) дней. Для оценки наличия связей между численностью водорослей и факторами среды, был произведен кросскорреляционный анализ временных рядов, предварительно избавленных от смещения путем вычитания трехточечного скользящего среднего с равными весами. Найдено, что динамика численности *M. aeruginosa* в 2009 году относительно тесно связана с двумя факторами. Первый фактор — это численность водоросли *Pediastrum (P.) duplex* с коэффициентом кросс-корреляции $k \approx 0,9$ при временном сдвиге между рядами на шесть дней. Вторым значимым фактором — минерализация воды со значением $k \approx 0,6$ при нулевом временном сдвиге. Анализ связей между численностью *M. aeruginosa* и факторами среды в 2011 году показал, что для этой выборки характерны высокие корреляции ($k \approx 0,6$) предиктанта с концентрацией нитрат-иона, температурой воды и численностью *P. duplex* с различными временными сдвигами от 20 до 23 дней. Результаты этого анализа использовались для выработки двух видов нечеткого прогноза. Программная реализация всех алгоритмов прогноза осуществлялась в среде MATLAB. Для экстраполяции на срок 2–6 дней использовалось приближение данных полиномом первой степени по методу наименьших квадратов с помощью функции «polyfit». Алгоритм нечеткой логики осуществлен с помощью редактора правил нечеткой логики «fuzzy». Согласно [5] модели гибридных сетей, объединяют достоинства нейронных сетей и систем нечеткого вывода и используют процедуры обучения нейронных сетей на выборках данных с тем, чтоб определить наилучшие параметры функций принадлежности, соответствующие определенной системе нечеткого вывода. Для реализации этого метода разработана специальная программа.

2. Результаты

Результаты прогнозирования тремя методами представлены на рис. 1. Для оценки успешности прогнозов рассчитывали среднюю абсолютную и относительную ошибки методом, изложенным в руководстве по краткосрочным прогнозам погоды [6].

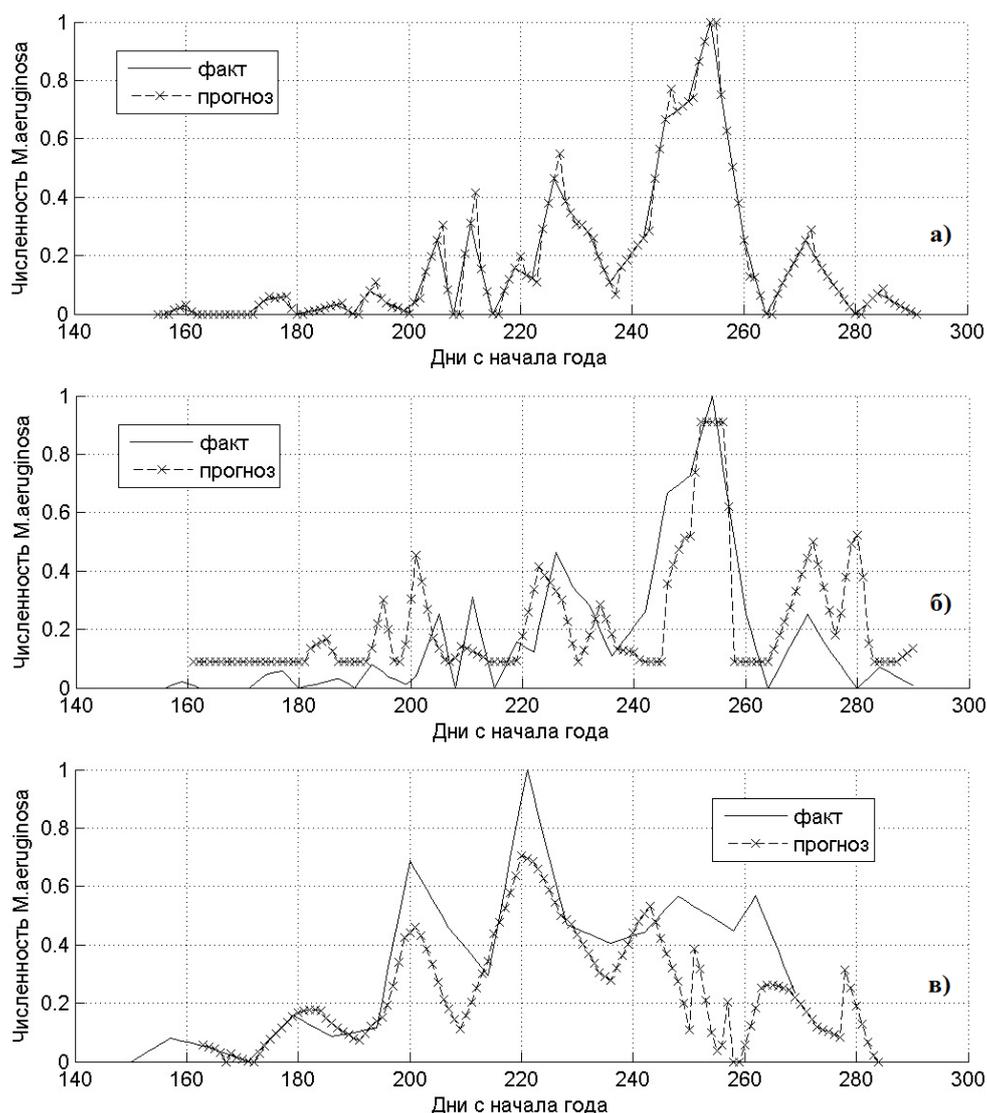


Рис. 1. Результаты прогнозирования численности *M. aeruginosa* различными методами: результаты линейной экстраполяции данных 2009 года (а), прогнозирования по правилам нечеткой логики для 2009 года (б) и (в) прогноза, осуществленного самообучающейся нечеткой нейронной сетью по данным 2011 года

Экстраполяционный прогноз. Экстраполяция ряда численности является самым простым методом прогноза, для применения которого достаточно данных о численности прогнозируемой водоросли и предположения о наличии некоторой инерционности процессов ее увеличения или снижения. Результат линейной экстраполяции данных 2009 года, представлен на рисунке 1 (а). Средняя абсолютная ошибка экстраполяции с использованием последних двух наблюдений на 1 день вперед за два года была в диапазоне 0,6–1,2 %, на 2 дня 3,6–3,7 %, на 3 дня 3,6–7,2 %. Несмотря на то, что результаты экстраполяционного прогноза, судя по значениям абсолютных ошибок кажутся вполне удовлетворительным, очевидно, что при экстраполяции далеко за крайний узел ошибка будет велика. При этом с увеличением числа точек использованных для интерполяции, происходит ухудшение качества прогноза. Так при ис-

пользовании последних 6 наблюдений для экстраполяции на 1 день ошибка достигает 5,3 %, а на 3 дня уже 11 %. Очевидно, что в точках значительного изменения скорости роста численности, и особенно в момент изменения ее знака, экстраполяционный прогноз будет наименее успешным и должен корректироваться. Вопрос заключается в том, возможно ли предугадать момент времени, когда возможен перелом в эволюции численности.

Прогноз с использованием правил нечеткой логики. Выявленные в ходе кросс-корреляционного анализа зависимости, позволили составить правила нечеткой логики, которые легли в основу прогноза численности *M. aeruginosa* для 2009 и 2011 годов. При этом срок прогноза задавался значением сдвига предиктора, определенным в ходе кросс-корреляционного анализа. Так, прогнозирование по данным 2009 года осуществлялось с помощью одного правила: если численность *P. duplex* сегодня «высокая» («низкая»), то численность *M. aeruginosa* через 6 дней будет «высокая» («низкая»). Лингвистические оценки «низкая», «высокая» задавались с помощью редактора функций принадлежности на отрезке [0, 1] и представляют собой множества значений переменных «численность *P. duplex*» и «численность *M. aeruginosa*». Таким образом, заблаговременность прогноза с учетом данных наблюдений за последние сутки достигает 5 дней. Результаты прогнозирования (рис. 1б) демонстрируют угадывание «основных» пиков нормированной численности водоросли. Средняя абсолютная ошибка прогноза для этой модели в 2009 году составила 13 %.

Для выборки 2011 года были использованы три правила: 1) если температура воды сегодня «высокая» («низкая»), то численность *M. aeruginosa* через 20 дней будет высокая («низкая»); 2) Если численность *P. duplex* сегодня «высокая» («низкая»), то численность *M. aeruginosa* через 23 дня будет «высокая» («низкая»); 3) если концентрация нитрат-иона сегодня «высокая» («низкая»), то численность *M. aeruginosa* через 20 дней будет «высокая» («низкая»). Средняя абсолютная ошибка нечеткого прогноза для данных 2011 года составила 15 %.

Прогноз с использованием нечеткой нейронной сети. В отличие от правил нечеткой логики, параметры функций принадлежности в данном случае не задавались «вручную», а определялись автоматически в результате обучения нейронной сети на заранее проанализированных данных. Для обучения нейронной сети использовались пары «вход-выход». В качестве «входа» использовали данные о значении отобранных в ходе кросс-корреляционного анализа параметров с соответствующими сдвигами во времени, а в качестве выхода — данные о численности *M. aeruginosa* на момент прогноза. При этом необходимо заметить, что сдвиги параметров определяют срок прогноза.

Для 2009 года в ходе обучения получены две модели — с учетом минерализации и с учетом численности *P. duplex*. Для 2011 года в ходе обучения получены четыре модели, дающие удовлетворительные результаты. Три однопараметрические с учетом концентрации нитрат-иона; с учетом температуры воды; с учетом численности *P. duplex* и одна двухпараметрическая с учетом численности *P. duplex* и температуры воды. Лучшие результаты со средней ошибкой прогноза 13 % дает модель с двумя параметрами (см. рис. 1в). Как видно из рис. 1в, модель на основе нейро-нечеткой сети не позволяет точно оценить численность водорослей, но достаточно уверенно определяет время и интенсивность максимальной вспышки численности с заблаговременностью до 20 дней. Очевидно, что относительно невысокие результа-

ты прогноза с использованием самообучающейся нейро-нечеткой сети обусловлены короткой длиной выборкой, недостаточной для ее полноценного обучения. С другой стороны, модель требует очень тонкой настройки и тщательного выбора индикаторных факторов. По мере накопления новых данных о состоянии среды и численности водорослей, некоторые факторы будут отбраковываться, другие проявляться.

3. Заключение

Проведенная оценка результатов краткосрочного прогнозирования численности *M. aeruginosa*, методами экстраполяции и методами основанными на правилах нечеткой логики и нейронных сетях показала, что каждый из них обладает своими достоинствами и недостатками. Оценка качества прогнозов на один-два дня вперед выполненных различными методами, показала, что экстраполяция хорошо работает на квазилинейных интервалах изменения численности. С другой стороны, при увеличении средних ошибок прогноза на малые сроки методы с использованием прогнозных правил с большей заблаговременностью предсказывают времена и амплитуды вспышек численности. Очевидно, что искомые правила нечетких методов о связи численности предиктанта с параметрами окружающей среды должны быть уточнены и дополнены после накопления дополнительных натуральных наблюдений. Однако уже на данном этапе это не исключает применения этого метода для оптимизации работы очистных сооружений водопровода в летний период, когда ежедневно производится анализ целого спектра биологических и гидрохимических показателей, необходимых для настройки методов.

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации в рамках федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы (Соглашение № 14.В37.21.0613).

Литература

1. Marsili-Libelli, S. Fuzzy Prediction of the Algal Blooms in the Orbetello Lagoon / S. Marsili-Libelli, G. Pacini, C. Barresi // Environmental Modeling & Software 19. – 2004. – P. 799–808.
2. Oh, H.-M. Community Patterning and Identification of Predominant Factors in Algal Bloom in Daechung Reservoir (Korea) Using Artificial Neural Networks / H.-M. Oh, Ch.-Y. Ahn, J.-W. Lee, T.-S. Chon, K.H. Choi, Y.-S. Park // Ecological modeling. – 2007. – Vol. 203. – P. 109–118.
3. Гаязова, А.О. Исследование сезонной динамики массовых видов фитопланктона озера Смолино / А.О. Гаязова, С.М. Абдуллаев // Вестник Челябинского государственного университета. Экология и природопользование. – 2011. – № 5(220). Вып. 5. – С. 70–75.
4. Гаязова, А.О. Оценка факторов среды, определяющих массовое развитие *Microcystis aeruginosa* в озере Смолино (г. Челябинск) / А.О. Гаязова // Биоиндикация в мониторинге пресноводных экосистем II. Сборник материалов международной конференции. – СПб.: Любавич, 2011. – С. 114–118.

5. Леоненков, А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH / А.В. Леоненков. – СПб.: БХВ-Петербург, 2005. – 736 с.
6. Руководство по краткосрочным прогнозам погоды. / под. ред. С.Л. Белоусова, А.А. Васильева, Н.Ф. Вельтищева и др. – Л.: Гидрометеоиздат, 1986. – Ч. I. – 702 с.

Гаязова Анна Олеговна, гидробиолог, муниципальное предприятие «Производственное объединение водоснабжения и водоотведения», why.ann@mail.ru.

Абдуллаев Санжар Муталович, доктор географических наук, профессор кафедры вычислительной математики, факультет Вычислительной математики и информатики, Южно-Уральский государственный университет, sanjar@mail.ru.

FORECASTING OF MICROCYSTIS AERUGINOSA SEASONAL DYNAMICS USING THE FUZZY LOGIC AND FUZZY NEURAL NETWORKS

A.O. Gayazova, municipal enterprise “Production association of water supply and distribution” (Chelyabinsk, Russian Federation),

S.M. Abdullaev, South Ural State University (Chelyabinsk, Russian Federation)

The retrieval of potential predictors of blue-green algae *M. aeruginosa* blooms and bloom prediction using fuzzy logic rules and fuzzy neural networks is discussed. Time series of seasonal dynamics of *M. aeruginosa* quantities and parameter values were obtained through field observations of biotic and abiotic parameters of the water environment held at Lake Smolino (Chelyabinsk) in the warm season of 2009 and 2011. The cross-correlation analysis of the data revealed the potential predictors of *M. aeruginosa* abundance quasi-periodic oscillations with a period of 12-20 days: algae *P. duplex* abundance, water temperature and the concentration of nitrates. According to the results of cross-correlation analysis a number of rules and membership functions in a range of changes from zero to one is set forward. Specially written program was used to train the fuzzy neural network in data on the values of predictant and selected predictors to apply the predictive rules and membership functions automatically. To compare the results additionally performed a linear extrapolation of the predictant abundance. Seasonal development of *M. aeruginosa* was well predicted by the extrapolation forecast only on quasilinear intervals of *M. aeruginosa* abundance evolution, whereas fuzzy logic theory was good to predict the *M. aeruginosa* intense outbreaks.

Keywords: *M. aeruginosa*, algae bloom forecasting, quasi-periodic oscillations, linear extrapolation, fuzzy logic, fuzzy artificial neural networks.

References

1. Marsili-Libelli S., Pacini G., Barresi. C. Fuzzy Prediction of the Algal Blooms in the Orbetello Lagoon. Environmental Modeling & Software, 2004. Vol. 19, P. 799–808.
2. Oh H.-M., Ahn Ch.-Y., Lee J.-W., Chon T.-S., Choi K.H., Park Y.-S. Community Patterning and Identification of Predominant Factors in Algal Bloom in Daechung Reservoir (Korea) Using Artificial Neural Networks. Ecological modeling. 2007. Vol. 203, P. 109–118

3. Gayazova A.O., Abdullaev S.M. Issledovanie sezonnoj dinamiki massovyh vidov fitoplanktona ozera Smolino [Study of Dominant Algae Species Seasonal Abundance in Smolino Lake]. Vestnik Cheliabinskogo gosudarstvennogo universiteta. Seriya: Ekologiya i Prirodopolzovanie [Bulletin of Chelyabinsk State University. Series: Ecology and Nature management]. 2011. Vol. 220, No. 5. P. 70–75.
4. Gayazova A.O. Otsenka factorov sredi opredeliaushih massovoe razvitie Microcystis aeruginosa v ozere Smolino (Chelyabinsk) [Evaluation of Environmental Factors that Determine the Mass Development of Microcystis aeruginosa in Lake Smolino (Chelyabinsk)]. Bioindikatsia v monitoringe presnovodnyh ekosistem II. Sbornik materialov mezhdunarodnoj konferentsii [Bioindication in Monitoring of Fresh Water Ecosystems II. Book of Papers of the International Conference]. St. Petersburg, "Lubavich", 2011. P. 114–118.
5. Leonenkov A.V. Nechetkoe modelirovanie v srede MATLAB i fuzzyTECH [Fuzzy Modeling with MATLAB and FuzzyTECH]. St. Petersburg, BHV-Petersburg, 2005. 736 p.
6. Rukovodstvo po kratkosrochnym prognozam pogody. Chast 1. [Guidance on Short-term Weather Forecasts. Part 1]. Edited by S.L. Belousov, A.A. Vasilyev, N.F. Veltishev et al. Leningrad, Gidrometeoizdat, 1986. 702 p.

Поступила в редакцию 5 ноября 2012 г.