

С.В.Алёмов*, С.А.Бабичев**, В.И.Литвиненко**

*Институт биологии южных морей НАН Украины, г.Севастополь

**Херсонский Национальный Технический Университет, г.Херсон

АНАЛИЗ ДАННЫХ МНОГОЛЕТНЕГО МОНИТОРИНГА ДОННЫХ СООБЩЕСТВ И НЕФТЯНОГО ЗАГРЯЗНЕНИЯ ГРУНТА В ПРИБРЕЖНОЙ ЗОНЕ Г.СЕВАСТОПОЛЯ НА ОСНОВЕ СЕТИ БАЙЕСА

Представлена модель влияния степени загрязнения окружающей среды на биологические показатели макрозообентоса акватории Севастопольских бухт, обработка информации в которой осуществлялась сетью Байеса. В качестве исследуемых показателей использовались влажность грунта, уровень содержания битумоида, число видов зообентоса, их общая численность и биомасса. Экспериментально показана значительная чувствительность всех биологических показателей к содержанию битумоидов в донных осадках.

КЛЮЧЕВЫЕ СЛОВА: *сеть Байеса, донные осадки, макрозообентос, нефтяные углеводороды.*

Актуальность работы определяется современным экологическим состоянием прибрежных морских акваторий, которое характеризуется высокой степенью загрязненностью по причине беспорядочного сброса отходов действующими промышленными предприятиями. Комплексная характеристика экологического состояния любой прибрежной акватории объединяет два основных направления мониторинговых исследований: химическое и биологическое. Первое дает представление о пространственном и количественном распределении того или иного вида загрязнения в границах выделенного района. Задача биологического направления состоит в оценке степени поражающего воздействия конкретного загрязнителя или их комплекса на биообъекты и отклика последних на различных организационных уровнях. Макрозообентос (сообщества донных организмов) является одним из узловых звеньев биотического круговорота веществ и энергии в прибрежных экологических системах. Пространственная стабильность и относительно долготие макробентосных организмов и сообществ делает их наиболее удобными объектами в исследовании долгосрочных изменений морской среды под воздействием загрязнения.

Постановка проблемы. В [1, 2] показано, что число видов донных организмов, их численность и биомасса определяют как биологическое разнообразие, так и интенсивность потока вещества и энергии через донные сообщества. Относительное долготие макробентосных организмов и сообществ делает их удобными объектами в исследовании долговременных изменений морской среды под воздействием загрязнения [3]. В [3 – 5] авторами показано, что при описании состояния водоёмов желательнее предоставлять "абсолютные" исходные данные, не заменяя их только расчётными показателями. Поэтому, исходя из принципа приоритета первичных данных, основным результатом гидробиологического мониторинга являются: плотность видов, плотность организмов, плотность биомассы [6]. Каждый из

© С.В.Алёмов, С.А.Бабичев, В.И.Литвиненко, 2012

перечисленных показателей или их различные комбинации являются основой для построения многих теорий, критериев и методов оценок качества водоёмов. Создание системы прогнозирования состояния экологически опасных объектов обусловлено следующими обстоятельствами:

- накоплением и обработкой больших объемов статистической и аналитической информации;
- комплексным анализом возможных источников загрязнения водоемов;
- анализом влияния структуры экономики государства и отдельного региона (развития промышленности) на экологическое состояние водных объектов;
- факторный анализ возможных последствий загрязнения;
- моделированием сценариев возможного экологически опасного выброса.

Влияние факторов носит вероятностный и неопределенный характер, поэтому для расчета степени адекватности гипотезы об экологическом состоянии гидробиологической среды целесообразно использовать байесовский подход.

В [7, 8] описана методика применения байесовских сетей для анализа санитарного состояния водных объектов.

Существуют два способа обучения байесовских сетей на основе классификации: уточнение параметров сети, если структура сети известна, и выбор из множества моделей, используя введенную метрику ко всей базе классов. Гекерман [9] отмечает четыре преимущества байесовских сетей как средства получения данных:

- поскольку в модели определяются зависимости между всеми переменными, то легко обрабатываются ситуации, когда значения некоторых переменных неизвестны;
- построенные байесовские сети просто интерпретируются и позволяют на этапе прогнозного моделирования легко выполнять анализ по сценарию "что если ...";
- подход позволяет естественным образом совмещать закономерности, выведенные из данных и фоновые знания, полученные в явном виде, например, от экспертов;
- использование байесовских сетей позволяет избежать проблемы переобучения (*overfitting*), т.е. чрезмерного усложнения модели, чем страдают многие методы (например, деревья решений и индукция правил) при слишком буквальном следовании распределению зашумленных данных.

Несмотря на простоту, скорость и интерпретацию результатов, байесовского алгоритм имеет недостатки:

- перемножать условные вероятности корректно только тогда, когда все входные переменные действительно статистически независимы; допущение этой независимости и обуславливает приставку "наивно-" в названии алгоритма, хотя, по приведенным в [10] примерам, он показывает неплохие практические результаты даже при несоблюдении условия статистической независимости; корректно данная ситуация обрабатывается только более сложными методами, основанными на обучении байесовских сетей [9,11];
- невозможна непосредственная обработка непрерывных переменных - их требуется разбивать на множество интервалов, чтобы атрибуты были дискретными; такое разбиение в ряде случаев приводит к потере значимых закономерностей [10];

– “наивно-байесовский” подход учитывает только индивидуальное влияние входных переменных на результат классификации, не принимая во внимание комбинированного влияния пар или троек значений разных атрибутов [11], что было бы полезно с точки зрения прогностической точности, но значительно увеличило бы количество проверяемых комбинаций.

Сети Байеса активно использовались для формализации знаний экспертов в экспертных системах, но с недавних пор их стали применять для получения знаний из наборов данных.

Целью статьи является разработка системы оценки влияния химических загрязнений на биологические показатели, характеризующие состояние гидробиологических объектов, на основе сети Байеса.

Материалы и методы. Пусть Ω – выборочное пространство событий (или множество событий) случайных экспериментов. Это выборочное пространство содержит все возможные значения случайной переменной. Предположим, что есть две переменные E и H , которые некоторым образом связаны между собой. Если мы имеем конкретное значение H , т.е. имеет место конкретное событие, то имеет интерес, какой будет при этом вероятность события E . Вероятность события E , при условии наступления события H , называется условной вероятностью. Условная вероятность события определяется выражением:

$$p(E|H_k) = \frac{p(E \cap H_k)}{p(H_k)}, \quad (1)$$

где \cap – операция пересечения множеств. Если E_1, E_2, \dots, E_n – такие взаимоисключающие события, что $\cup_{i=1}^n E_i = \Omega$, то говорят, что события E_i формируют полное (исчерпывающее) множество. Две переменные не связаны (не пересекаются), если они не имеют одинаковых значений. Если две переменные являются исчерпывающими и несвязанными, то можно записать, что

$$E = \cup_i (E \cap H_i), \quad (E \cap H_i) \cap (E \cap H_j) = \emptyset, \quad i \neq j. \quad (2)$$

Теория построения Байесовских сетей основана на предположении, что события являются исчерпывающими и не пересекаются. Если это условие не выполняется, то результаты применения сети будут неконсистентными (т.е. неточными). В случае если события являются исчерпывающими и не пересекаются, то вероятность события можно вычислить с помощью условных вероятностей

$$p(E) = \sum_{i=1}^n p(E \cap H_i) = \sum_{i=1}^n p(E|H_i) \cdot p(H_i). \quad (3)$$

Используя формулу (3), сумму пересечений событий E и H можно выразить следующим образом:

$$p(E \cap H_k) = p(E|H_k) \cdot p(H_k) = p(H_k|E) \cdot p(E). \quad (4)$$

С равенства $p(E|H_k) \cdot p(H_k) = p(H_k|E) \cdot p(E)$ имеем:

$$p(H_k|E) = \frac{p(E|H_k) \cdot p(H_k)}{p(E)},$$

а с учетом (3) получаем выражение:

$$p(H_k|E) = \frac{p(E|H_k) \cdot p(H_k)}{\sum_{i=1}^n p(E|H_i) \cdot p(H_i)}, \quad (5)$$

представляющее собой формулу Байеса. На основе этой формулы строятся сети Байеса.

В (5) H_k означает любую гипотезу из n возможных. Вероятности $p(E|H_k)$ задаются экспертами априорно. То есть, их можно рассматривать как ответ на вопрос: "Какой будет вероятность некоторого события, если известно, какая гипотеза была реализована?". Вероятности $p(E|H_k)$ очень полезны, поскольку, как правило, легче найти вероятность последовательности событий типа причина-следствие, чем наоборот. Значения $p(H_k)$ называют априорными вероятностями, они определяют начальные вероятности для всех гипотез. Сила Байесовская метода заключается в том, что априорные вероятности можно уточнять (обновлять) в соответствии с реалиями протекания исследуемого процесса. Это позволяет уточнять вероятности событий при поступлении дополнительной информации. Знаменатель выражения (5) можно рассматривать как нормирующий член, который устанавливает значение вероятности в промежутке между 0 и 1.

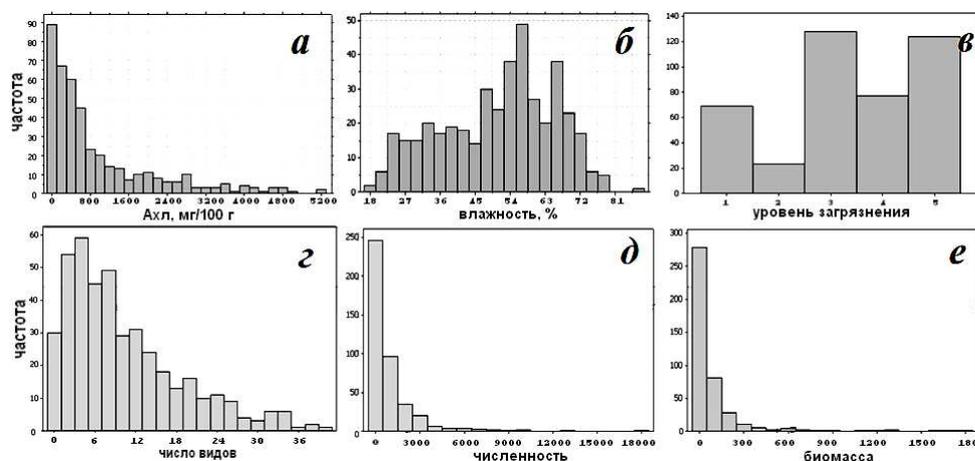
Сеть Байеса, формируемая на основании формулы (5), представляет собой направленный ациклический граф, где каждый узел представляет собой переменную, а каждая дуга – вероятностную зависимость, определяемую количественно использованием условного распределения вероятностей для каждого узла. В состав сети Байеса входят следующие компоненты:

- множество узлов, определяющих компоненты системы;
- множество направленных связей между компонентами системы.

Для построения сети использовались данные Института биологии южных морей им. А.А. Ковалевского по характеристикам донных осадков и макрозообентоса, полученные в ходе проведения многолетних мониторинговых исследований отделом морской санитарной гидробиологии [12]. Пробы донных осадков и макрозообентоса отбирали одновременно по стандартной сетке станций на глубинах 4 – 20 м в летний период. На каждой станции делали подъем одной пробы для определения натуральной влажности донного осадка и содержания ХЭВ по стандартным методикам. Одновременно проводили отбор проб макрозообентоса дночерпателем с площадью захвата или 0,038 м² (в трех повторностях). Донный осадок промывали через сито с диаметром отверстий 1 мм и фиксировали этанолом (96°). Определяли сырой вес фиксированных организмов макрозообентоса, данные перечисляли на 1 м².

Исследовалось влияние показателей влажности донных осадков и содержания битумоида в донном осадке на общий уровень загрязнения, а также на такие биологические показатели как число видов, общая численность и общая биомасса. Всего для шести показателей было сделано 421 наблюдение.

Результаты. Для моделирования исследуемых процессов была построена статическая байесовская цепь, целевым узлом в которой был выбран уровень загрязнения окружающей среды. Дискретизация исследуемых параметров была осуществлена с учетом мнения экспертов в данной предметной области следующим образом: диапазон значений битумоида при минимуме 1 и максимуме 5700 был разбит на 29 интервалов; значение влажности грунта при минимуме 18 % и максимуме 86 % было разбито на 24 интервала; число видов при минимуме 0 и максимуме 40 было разбито на 21 интервал; общая численность видов при минимуме 0 и максимуме 18330 была



Р и с . 1 . Частотная гистограмма распределения содержания битумоида в донном осадке (а), влажности грунта (б), уровня загрязнения (в), числа видов макробентоса (г), численности (д) и биомассы (е) макрозообентоса.

разбита на 20 интервалов; биомасса при минимуме 0 и максимуме 1770 была разбита на 19 интервалов. Уровень загрязнения определялся заранее расчетными методами на основе интегрированных показателей с помощью соответствующих методик и был представлен в виде 5-ти дискретных показателей. Частотные распределения исследуемых показателей представлены на рис.1.

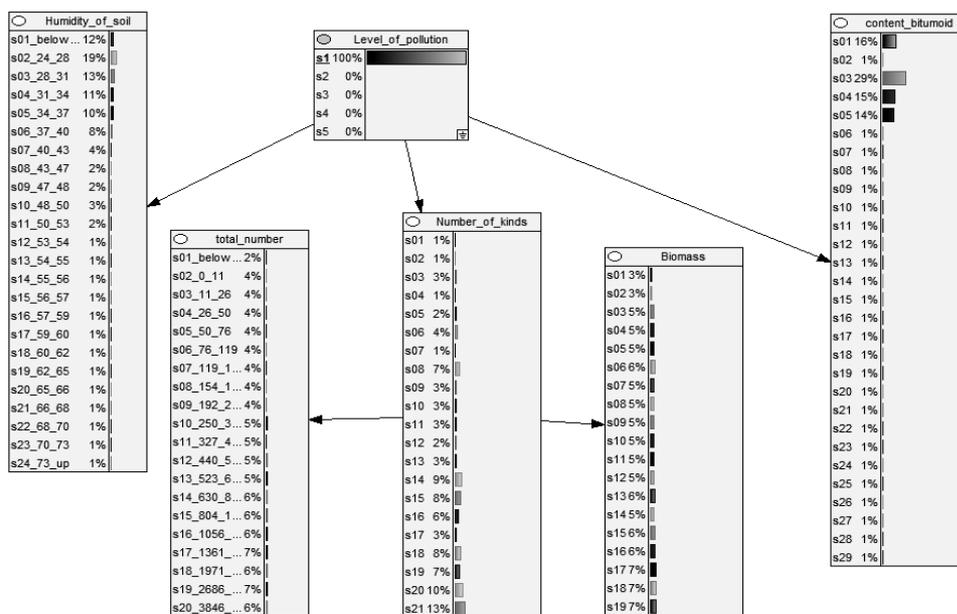
Модель построенной байесовской сети при низком и высоком уровне загрязнения окружающей среды представлена на рис.2.

Анализ результатов построенных моделей показал, что показатель влажности фактически не реагирует на уровень загрязнения окружающей среды, что, однако, не исключает влияние влажности на другие биологические и физико-химические показатели, что может быть объяснено недостаточной сложностью модели. Значения содержания битумоидов в донных осадках весьма чувствительно к изменению уровня загрязнения. Так, например, при низком уровне загрязнения высокие значения имеют дискретные значения 1 – 5, отражающие вероятность дискретных значений с низким значением содержания битумоидов, а при высоком уровне загрязнения высокие значения имеют дискретные значения 21 – 29, отражающие вероятность дискретных значений с высоким значением содержания битумоидов. Количество видов и общее количество живых организмов возрастает при уменьшении и уменьшается при возрастании уровня загрязнения окружающей среды, что согласуется с мнением экспертов в данной предметной области. Биомасса также возрастает при уменьшении уровня загрязнения окружающей среды и уменьшается при его возрастании, что также согласуется с мнением экспертов. Таким образом, экспериментально установлена значительная чувствительность всех видов биологических показателей на содержание битумоидов.

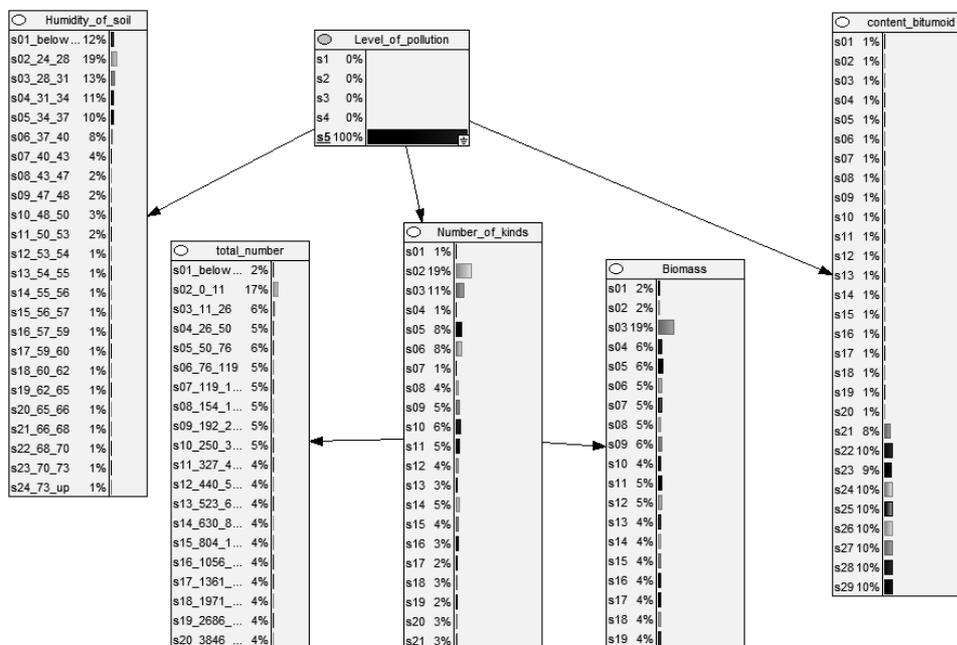
Вывод. В работе представлена модель влияния степени загрязнения окружающей среды на биологические показатели макрозообентоса акватории Севастопольских бухт, обработка информации в которой осуществлялась сетью Байеса. В качестве исследуемых показателей использовались показатель влажности грунта, уровень содержания битумоида, число видов зоо-

бентоса, их общая численность и биомасса.

Дальнейшим направлением исследований авторов является построение динамической байесовской сети, которая позволит повысить точность и чувствительность работы модели в исследуемой предметной области.



a



b

Р и с . 2 . Модель байесовской сети при низком (а) и высоком (б) уровнях загрязнения окружающей среды.

СПИСОК ЛІТЕРАТУРИ

1. *Мальцев В. И.* О возможности применения показателя функционального обилия для структурных исследований зооценозов // Гидробиологический журнал.– 1990.– 26, № 1.– С.87-89.
2. *Rybarczyk H., Elkaim B., Ochs L., Loquet N.* Analysis of the trophic network of a macrotidal ecosystem: the Bay of Somme // *Estuar. Coast. Shelf Sci.*– 2003.– 58.– P.405-421.
3. *Шуйский В.Ф., Максимова Т.В., Петров Д.С.* Изоболический метод оценки и нормирования многофакторных антропогенных воздействий на пресноводные экосистемы по состоянию макрозообентоса.– СПб: МАНЭБ, 2004.– 304 с.
4. *Алёмов С.В.* Некоторые методические аспекты расчёта «индекса биологического качества» (BQI) // Междунар. науч. конф. «Современные проблемы гидробиологии. Перспективы, пути и методы исследований». 24-27 июля 2006 г.– Херсон, 2006.– С.6-9.
5. *Петров А.Н.* Реакция прибрежных макробентических сообществ Чёрного моря на органическое обогащение донных отложений // *Экология моря.*– 2000.– вып.51.– С.45-51.
6. *Шитиков В. К., Розенберг Г. С., Зинченко Т. Д.* Количественная гидроэкология: методы системной идентификации.– Тольятти: ИЭВБ РАН, 2003.– 463 с.
7. *Алёмов С.В., Бардачев Ю.Н., Ливиненко В.И., Фефелов А.А.* Использование байесовских сетей для оценки состояния водных объектов // *Моделирование и управление состоянием эколого-экономических систем региона.*– 2006.– вып.3.– С.14-26.
8. *Алёмов С.В., Бардачев Ю.Н., Гожий А.П., Литвиненко В.И., Миронов О.Г.* Ситуационное моделирование санитарного состояния гидробиологических систем с использованием байесовских сетей // Вторая междунар. конф. «Современные проблемы гидробиологии. Перспективы, пути и методы решений».– Херсон, 2008.– С.242-252.
9. *Heckerman D.* "Bayesian Networks for Data Mining" // *Data Mining and Knowledge Discovery.*– 1997.– № 1.– P.79-119.
10. *Brand E., Gerritsen R.* Naive-Bayes and Nearest Neighbor // *DBMS.*– 1998.– № 7.– P.131-165.
11. *Friedman N., Geiger D., Goldszmidt M., etc.* Bayesian Network Classifiers // *Machine Learning.*– 1997.– 29.– P.131-161.
12. *Миронов О.Г., Кирюхина Л.Н., Алёмов С.В.* Санитарно-биологические аспекты экологии Севастопольских бухт в XX веке.– Севастополь, 2003.– 185 с.

Матеріал поступив в редакцію 15.10.2012 г.

АНОТАЦІЯ. Представлена модель впливу міри забруднення довкілля на біологічні показники макрозообентосу акваторії Севастопольських бухт, обробка інформації в якій здійснювалася мережею Байеса. Як досліджувані показники використовувалися вологість ґрунту, рівень змісту бітумоїду, число видів зообентосу, їх загальна чисельність і біомаса. Експериментально показана значна чутливість усіх біологічних показників до змісту бітумоїдів в донних відкладеннях.

ABSTRACT. The model of influence of degree of environmental contamination on the biological indexes of macrozoobenthos of the Sevastopol bays aquatorium is presented. Treatment of information came true by the Bayesian Network. As the investigated indexes humidity of sediments, level of, number of species, general abundance and biomass of zoobenthos was used. The considerable sensitiveness of all biological indexes to bituminoid concentration is experimentally shown.