

УДК 574.9:57.045

ПРИМЕНЕНИЕ МЕТОДОВ ПРОБИТ-АНАЛИЗА ПРИ ИЗУЧЕНИИ БИОРАЗНООБРАЗИЯ НА ПРИМЕРЕ МЫШЕОБРАЗНЫХ

Звягинцева А.В., Перелыгина Н.В.

Донецкий национальный технический университет, Донецк

На основе данных многолетних наблюдений за видовым разнообразием мышеобразных и обработки картографической информации ареалов их распространении разработана база данных. Предложена методика исследования имеющейся в базе данных информации с использованием метода пробит-анализа.

Введение

Исследования и мониторинг биоразнообразия проводятся с целью сохранения и рационального использования биологических ресурсов и играют важную роль в сохранении разнообразия мира. Биоразнообразие представляет собой уникальную особенность живой природы, именно благодаря ему создается структурно-функциональная организация экологических систем. Подотряд мышеобразных отряда грызунов является одной из самых крупных таксономических единиц среди семейств млекопитающих, он включает 10 современных семейств (сони, полёвки, мышовки, тушканчики, слепыши, крысы, мыши, хомяки и др.) и 3 вымерших. Грызуны распространены по всему миру, за исключением, Антарктиды, и встречаются почти во всех наземных биотопах. Подотряд мышеобразных удобен как объект моделирования для анализа разнообразия и оценки условий обитания животных. Жизненные популяции грызунов можно рассматривать как индикатор состояния окружающей среды. Мышеобразные являются частью уникального генофонда земной поверхности, играют важную роль в трофических цепях и в функционировании природных экосистем. Они рассматриваются как ценная биологическая модель, а также представляют практический интерес в эпидемиологических исследованиях и занимают важное место в цепях питания. Тем не менее, распространение мелких грызунов в природе в полной мере не исследовано, структура их сообществ недостаточно изучена [1]. Актуальность исследования видового разнообразия мышеобразных заключается в необходимости изучения ареалов обитания грызунов, их экологических и морфологических особенностей, закономерностей распространения видов в зависимости от разных показателей. Приоритетным считается изучение структурных особенностей сообществ в конкретных природных условиях с учетом их видового состава, численности, продолжительности жизни и других параметров. Эти сведения имеют несомненную прикладную ценность, так как изменения в структуре сообществ часто служат индикаторами природных нарушений.

Постановка задачи

На основе собранных данных формируются банки данных. При накоплении достаточного объема данных они используются для проведения анализа

информации. Очень часто получаемые из различных источников данные хранятся в неупорядоченном, разрозненном виде, в связи с чем исследователь практически не имеет систематизированной и полной информации об интересующем его объекте. Поэтому на основе имеющихся данных необходимо создавать структурированные базы данных (далее, БД), хранящие информацию в соответствии с некоторым упорядочиванием, сортировкой.

В связи с этим **целью** работы является сбор и систематизация накопленной информации, а также создание на ее основе общей БД, позволяющей применить методы интеллектуального анализа данных для поиска закономерностей биоразнообразия.

Выполненные работы и возникшие трудности

На основе имеющихся БД построена структурированная БД, охватывающая множество численных показателей и содержащая графические изображения ареалов распространения мышеобразных и фотографии их внешнего вида. Численные показатели включают в себя информацию, характеризующую среднюю продолжительность жизни, температуру тела, вес, размеры особи, количество выводков, интервал между выводками, продолжительность беременности, период лактации, интенсивность метаболизма и др. по каждому из 300 видов подсемейства мыши. Общее количество внесенных численных значений в БД – более 7 тыс. Следует отметить, что в процессе исследования возникали трудности, связанные, прежде всего, с недостаточной изученностью отдельных видов мышеобразных и соответственно недостаточным количеством данных по таким показателям, как количество особей в группе, продолжительность жизни в природе, послеродовые темпы роста и т.п.

В существующих на сегодняшний день БД площади ареалов распространения большинства мышеобразных отсутствуют в цифровом виде, а также не имеется в наличии общей БД карт с нанесенными ареалами обитания видов [2]. В процессе сбора пространственных данных осуществлялся поиск наиболее популярными поисковыми системами Google и Yandex картографической информации с изображениями ареалов распространения исследуемых видов. Найденные растровые изображения были представлены в виде материков или отдельных их участков с нанесенными на них ареалами распространения тех или иных видов, которые визуально выделялись на фоне карты. Собранная картографическая информация оцифрована и занесена в БД в виде численного значения площади для каждого из ареалов распространения всех видов мышей в отдельности и общей площади распространения каждого из видов. Также в БД помещена векторная карта мира с множеством слоев со всеми нанесенными на нее ареалами распространения всех анализируемых видов мышеобразных. Фотографии с изображениями внешнего вида мышеобразных и растровые изображения ареалов их распространения хранятся в виде файлов в формате *.png. с достаточно хорошим сжатием изображения и минимальной потерей качества.

С помощью картографического метода осуществлена пошаговая оцифровка и обработка растровых изображений, а именно определение границ ареалов распространения видов и площадей соответствующих полигонов.

Для создания БД, статистической и графической обработки информации в работе

использована система *MapInfo*, совмещающая преимущества обработки информации, которыми обладают БД (в том числе язык запросов *SQL*), и наглядность карт, списков и графиков. Это позволило спроектировать БД мышьеобразных и наполнить её текстовыми, графическими и численными данными с возможностью определения параметров графических объектов (установление границ и определение площади ареала обитания).

Используемая методика обработки графических данных представляет собой пошаговый процесс оцифровки растровых изображений и их обработку: определение границ областей и площадей полигонов. После привязки крайних точек растровой карты к координатам на векторной карте мира растровое изображение становилось оцифрованным. Далее обозначались границы ареала распространения грызунов, что позволило определить площадь данного ареала, населенную тем или иным видом (рис. 1).

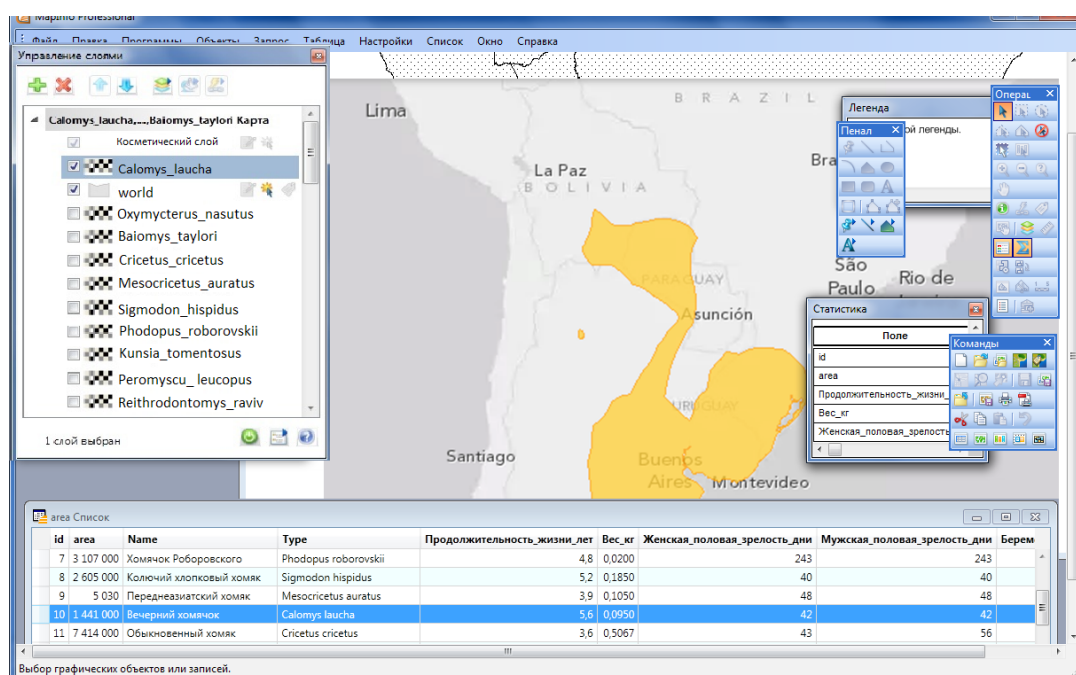


Рисунок 1. Результат оцифровки растрового изображения

Всего обработано порядка 300 карт, характеризующих распространение мышьеобразных на Земной поверхности. Наполнение БД происходило в следующей последовательности. Вначале выбиралось в виде полигона оцифрованное изображение ареала распространения вида, к которому впоследствии привязывалась атрибутивная информация (рис. 2), вносимая вручную. При выборе точки на пересечении ареалов распространения мышьеобразных, обитающих на территории с определенной широтой и долготой, отображался список площадей ареалов их распространения (рис. 3).

Полученное векторное изображение, находящееся в созданной БД, представляет собой один из слоев изображений проекта. Для выбора отображаемых слоев необходимо инструментом управления слоями указать на необходимый слой.

Направления дальнейших исследований

На основе собранных данных можно оценить вероятности распространения вида. Для этого предлагаем использовать отношение площади ареала распространения каждого из видов к площади сухопутной части земной поверхности. Далее следует

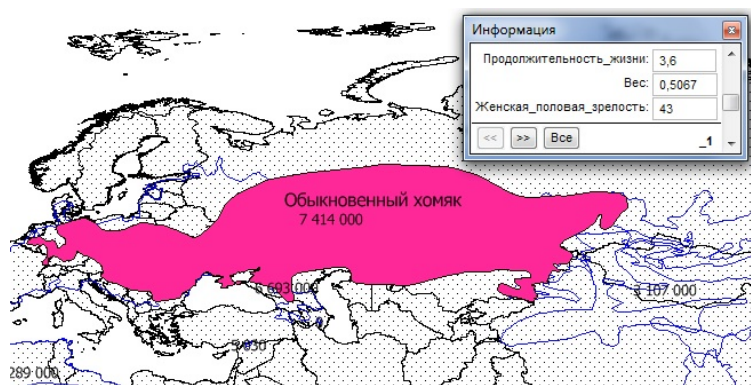


Рисунок 2. Привязка атрибутивной информации

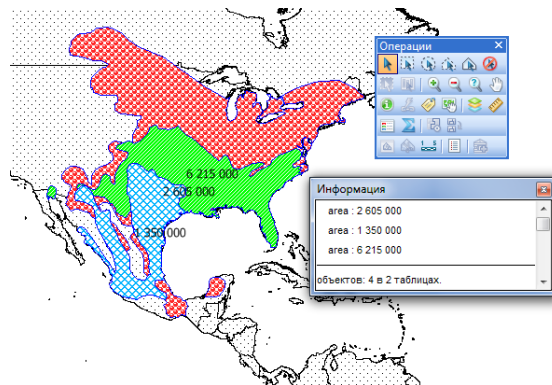


Рисунок 3. Выбор точки на пересечении ареалов распространения мышеобразных

искать зависимости между вероятностями распространения видов и биологическими параметрами мышеобразных. Для этого можно попытаться применить типовую методику оценки вероятности наступления событий по результатам полученных данных, которая широко применяется в биологических науках. В основе данной методики лежат методы пробит-анализа, разработанные в XX веке известным энтомологом Честером Блиссом [3]. Многочисленные опытные данные, полученные в токсикологии, радиобиологии, энтомологии, микробиологии, фармакологии, экологии и т.д. показывают, что зависимость между долей особей, у которых наблюдаются некоторые эффекты, к примеру, негативные, и количеством воздействия, например, дозой, выражается вероятностной кривой, имеющей S-образную форму. Обычно для трансформации этой кривой в прямую линию на оси абсцисс откладываются логарифмы доз, а по оси ординат – вероятностные единицы, так называемые пробиты. В биологических науках и безопасности систем пробит определяется согласно уравнения

$$Pr = \alpha_0 + \alpha_1 \cdot \ln z_1 + \alpha_2 \cdot \ln z_2 + \dots + \alpha_n \cdot \ln z_n \quad (1)$$

где α_i – const, z_i – биологические параметры или характеристики окружающей среды.

Теоретического обоснования для подобной процедуры статистической обработки данных пока нет, данные методики – это междисциплинарный научный факт, когда используется универсальный метод построения зависимости «доза-эффект», на основе которого и устанавливают связи между вероятностями негативных эффектов и параметрами воздействия.

Методика наших дальнейших исследований предполагает, что мы ищем связи между вероятностью распространения видов и наиболее важными показателями (масса, продолжительность жизни, температура тела и т.д.), используя уравнения (1).

Следующим этапом работы будет разработка моделей, направленных на поиск закономерностей распространения видов мышеобразных в зависимости от климатических параметров местообитания каждого из видов. Для этого предполагается использовать климатические базы данных, которые представляют собой временные ряды для каждой ячейки регулярной широтно-долготной решетки множества климатических показателей, которые доступны из системы Climate Wikience [4, 5], разработанной ДонНТУ. Это возможно при совмещении созданной БД по распространению мышеобразных с БД Всемирного климатического центра [6], где хранятся архивы климатических данных

повторного анализа. Архив данных Всемирного климатического центра позволяет построить осредненные во времени поля климатических показателей для всей планеты. Подобное осреднение можно провести за 15 лет на широтно-долготной сетке земного шара с ячейкой в 1x1 градус. В архиве хранится информация в несколько терабайт по 80 климатическим показателям, в частности, температуре, относительной влажности, скорости и направлению ветра, освещенности, осадкам, высоте над уровнем моря и т.д. С помощью применения регрессионного анализа возможно определение наиболее значимых климатических показателей, которые влияют на ареалы распространения вида. Кроме того из БД может быть взята атрибутивная информация, количественно характеризующая каждый вид. То есть объединение созданной БД с базой данных осредненных показателей Всемирного климатического центра позволит подойти к созданию моделей распространения видов на территории планеты.

Выводы

Таким образом, разработаны ГИС-модели распространения видов мышеобразных и создана БД картографической и атрибутивной информации, включающей в себя среднюю продолжительность жизни, температуру тела, вес, размеры особи, количество выводков, интервал между выводками, продолжительность беременности, период лактации, интенсивность метаболизма и др. по каждому из 300 видов подсемейства мыши. Предложена методика исследования имеющейся в базе данных информации с использованием метода пробит-анализа.

Литература

- [1] База данных: возраст и продолжительность жизни животных / Интернет ресурс. – Режим доступа: <http://genomics.senescence.info/species>.
- [2] Красный список исчезающих видов / Интернет ресурс. – Режим доступа: <http://maps.iucnredlist.org>.
- [3] А.Г. Платонов, М.Я. Ахалая. Применение метода пробит-анализа в радиобиологии. Расчет полулетальной дозы ЛД50. – Режим доступа: http://library.mephi.ru/Data-IRBIS/book-mephi/Platonov_Primerenie_metoda_probit-analiza_v_radiobiologii_2010.pdf.
- [4] Система Climate Wikience [Электронный ресурс] Режим доступа: wikience.donntu.edu.ua.
- [5] Rodrigues Zalipynis R.A., Zapletin E.A., Averin G.V., The Wikience: Community Data Science. Concept and Implementation., Proceedings of the 7th International Scientific-Technical Conference “Informatics and Computer Technologies” (ICT-2011), Vol. 1, pp. 113-117, Donetsk, November 22-23, 2011.
- [6] Более 80 климатических показателей. / Интернет ресурс. – Режим доступа: http://data-portal.ecmwf.int/data/d/interim_full_moda/; <http://nomads.ncdc.noaa.gov/data.php?name=access#CFSR-data>.
- [7] Классификация животных / Интернет ресурс. – Режим доступа: [www/ URL: ttp://www.apus.ru](http://www.apus.ru).