**Моделирование динамики микробиома оз. Байкал**

В соответствии с основной целью проекта – разработка математической модели микробиома оз.Байкал – на этапе 2017 года разработана процедура идентификации структуры динамической модели взаимодействия организмов в исследуемых сообществах альго-бактерий. Для поддержки исследований по этому направлению спроектирован прототип программной системы, организующий процесс исследований: предоставляется графический пользовательский интерфейс для визуального отображения всех этапов построения и расчета математической модели, результаты каждого этапа моделирования представлены в виде соответствующих визуальных структур, обеспечивается интеграция вычислительных подсистем в рамках одного исследовательского приложения.

Различные виды организмов тесно связанны между собой различными механизмами взаимодействия. В экосистеме озера Байкал большая часть первичной продукции производиться фитопланктоном. Продуктивность фитопланктона в свою очередь связанна с рядом внешних факторов (температура, освещенность, концентрация биогенных элементов). Первичная продукция, полученная в ходе фотосинтеза, утилизируется, в том числе и бактериями, входящими в сложную систему альго-бактериальных взаимодействий. Одним из путей прогнозирования поведения такой сложной системы является ее представление в виде динамической модели, позволяющей отслеживать эволюцию и поведение системы во времени. Математическое моделирование системы осуществляется с применением математического аппарата дифференциальных уравнений. Этап идентификации модели микробиома проводится в три этапа. Первый этап - статический корреляционный анализ имеющиеся информации по численности и биомассе видов, входящих в состав сообщества, концентрации биогенных элементов в воде озера. В результате корреляционного анализа определяются виды взаимодействий в рамках сообщества, их численность, биомасса видов, а также их связь с концентрациями биогенов в воде озера. Второй этап - разработка концептуальной сети взаимодействия видов друг с другом и с окружающей средой, представления сети взаимодействий в виде размеченного графа. Третий этап - создание на основе этого графа системы специального варианта системы дифференциальных уравнений, позволяющей прогнозировать поведение системы с течением времени.

Для поиска попарных взаимосвязей между представленностью видов бактериопланктона, в пробах; представленностью видов сообщества микроорганизмов; биомассой исследуемых доминирующих видов диатомовых в пробах; физико-химическими показателями воды (в число которых вошла общая биомасса фитопланктона) был использован корреляционный анализ с применением коэффициента корреляции Cпирмена *r* (Hollander M., Wolfe D.A., 1973). Выбор способа расчета коэффициента корреляции основан на предварительном анализе выборок с помощью критерия Шапио – Уилка (Patrick Royston 1995), который показал, что не все рассматриваемые выборки представленности видов в пробах, выборки физико-химических показателей воды в пробах и биомассы рассматриваемых видов диатомовых водорослей распределены по нормальному закону. Все рассчитанные коэффициенты корреляции объединены в корреляционную матрицу. Корреляционная матрица визуализировалась с помощью *тепловой карты* полученной средствами пакета “gplot” (Gregory R. et al 2015) языка программирования R. Пакет “gplot” в данном случае использовался не только для визуализации корреляционной матрицы, но и для параллельной классификации показателей численности видов и физико-химических характеристик среды по схожести коэффициентов корреляции, с помощью кластерного анализа.

Проведен ряд расчетов по предложенной схеме идентификации. В качестве примера анализа попарных взаимодействий приведем результаты корреляционного анализа для определения связи между биотическим и абиотическими показателями среды и представленностью (численность) некоторых видов бактериопланктона, доминирующего в фотическом слое пелагиали озера Байкал. Представленность (численность) видов бактериопланктона оценивалась с помощью метагеномного анализа ампликонов гена 16S рибосомальной РНК.

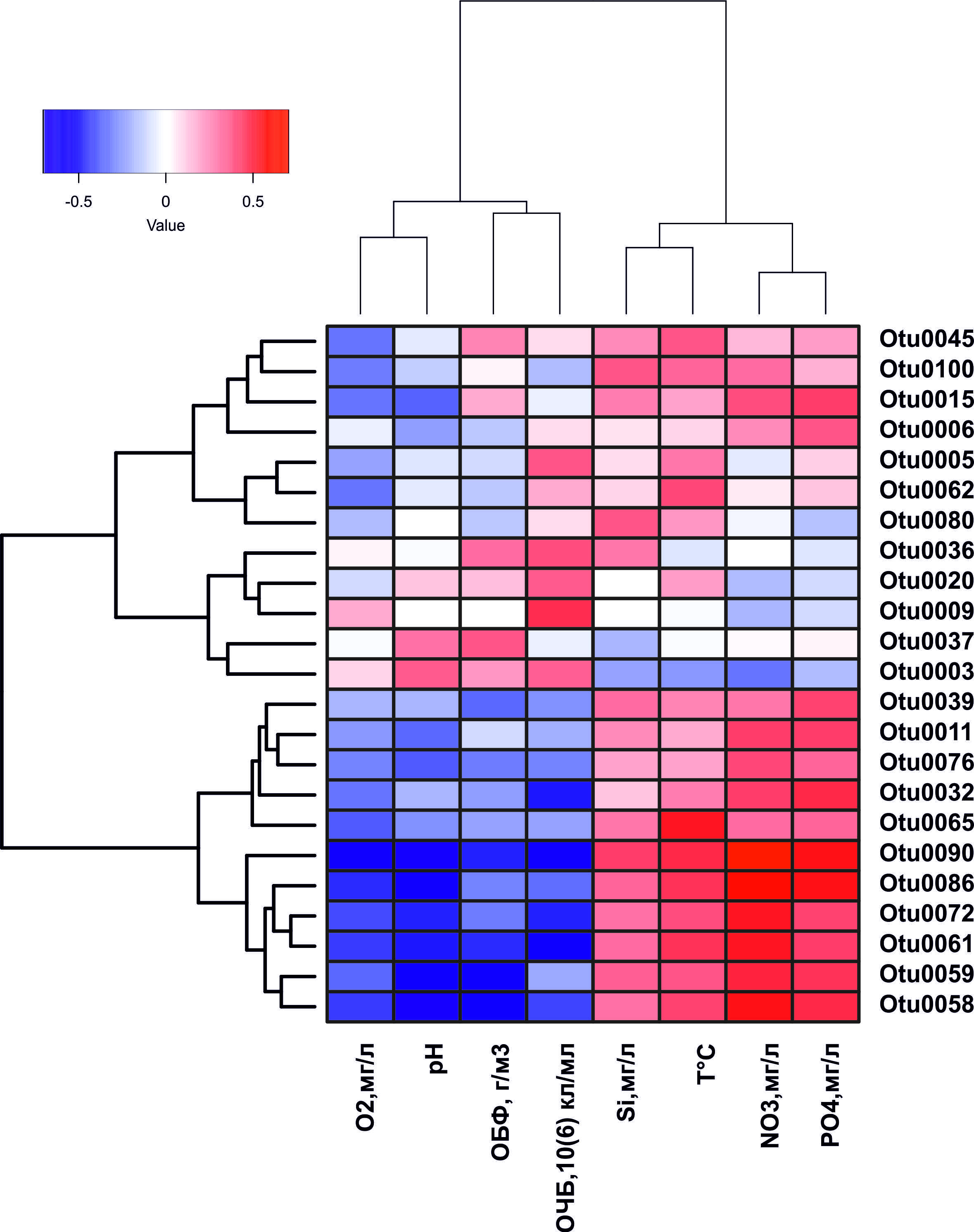


Рис.1. Тепловая карта, отражающая корреляционную зависимость между параметрами среды, в число которых входит общая биомасса фитопланктона с численностью доминирующих в пробе видов бактерий. Виды бактерий отмечены как OTU, каждый со своим номером.

Анализ корреляции между бактериальными видами (OTU) и биомассой видов фитопланктона показал, что OTU 40 и OTU 30 (*Synechococcus*) отрицательно коррелируют с биомассой видов весеннего фитопланктона (Рис. 1). Небольшое количество OTU положительно коррелировало с биомассой видов фитопланктона. С биомассой *S. acus* subsp. *radians* положительно коррелировало OTU 37 *Cryomorphaceae* (r=0.47); с *A. baicalensis* – OTU 89 *Chitinophagaceae* (r=0.43), OTU 53 *Actinomycetales* (r=0.38). С *G. baicalense* положительно коррелировали OTU 33, 19 (*Ilumatobacter*), OTU 32, 155 (*Verrucomicrobia* Subdivision3), OTU 81 Acidimicrobineae, OTU 127 Unclassified *Bacteria*, и отрицательно коррелировали OTU, принадлежащие *Bacteroidetes* и *Betaproteobacteria* (Рис. 1).

Динамическая модель включает, в простейшем случае, следующие характерные компоненты альго-бактериального сообщества: 1) меняющаяся во времени концентрация растворенных биогенов воде, обозначаемая r1; 2) меняющаяся во времени численность какого либо вида фитопланктона h1; 3) масса первичного органического вещества m; численность определенного вида фитопланктона, потребителя первичного органического вещества b1. Количество компонентов в системе определяется количеством взаимодействующих видов и количеством первичных ресурсов, доступных для этих видов. В общем виде рассмотренная четырехкомпонентная система описывается следующим уравнением:

(1)

В этом уравнении параметры имеют следующий биологический смысл: c(t) – приток биогена в озеро, – скорость элиминации биогена из окружающей среды фитопланктоном, – скорость элиминации биогена из окружающей среды бактериопланктоно, - скорость размножения фитопланктона, – коэффициент, определяющий интенсивность вымирания фитопланктона в единицу времени из-за конкуренции за общие ресурсы, - скорость размножения бактериопланктона и - коэффициент, определяющий интенсивность вымирания бактериопланктона в единицу времени из-за конкуренции за общие ресурсы. В уравнении фигурирует кадратичная смертность из-за конкуренции за общие ресурсы и рождаемость по экспонециальному закону.

В случае если в системе фигурирует *n* видов первичного ресурса (биогенов), *l* видов фитопланктона и *d* видов бактериопланктона, то система уравнений примет следующий вид:

(2)

В данном уравнении параметры имеют тот же биологический смысл, что и в предыдущей системе. Произошло добавление следующих параметров: – параметр, определяющий интенсивность взаимодействия между соответствующими видами фитопланктона; – параметр, определяющий интенсивность взаимодействия между соответствующими видам фитопланктона и бактерипланктона; – параметр, определяющий интенсивность взаимодействия между соответствующими видами бактерипланктона; – параметр, определяющий интенсивность взаимодействия между соответствующими видам бактерипланктона и фитопланктона. Значения равны 0, если нет соответствующей корреляционной связи, больше 0, если связь положительна, и меньше 0, если связь отрицательна.

Разрабатываемые модели визуализируются в виде графов взаимодействий (Рис. 2). Для построения таких графов их тепловых карт разработано программное обеспечение для поддержки научных исследований Model Studio. Тепловая карта анализируется по видам бактерипланктона и на первом этапе представители видов распределяются по классам и близким OTU. Аналогично распределяются взаимодействия. Подобное представление отображает естественную структуру взаимодействия составляющих динамической модели близкую к полисистемному расслоению.

Каждый слой полисистемы содержит объекты и взаимодействия, характеризующиеся общими свойствами. В данном случае выделены четыре слоя: внешние факторы (солнечная радиация, поступление углекислого газа): водоросли, рост которых в основном определяется внешними факторами; бактерии, численность которых зависит в основном от жизнедеятельности водорослей; состояние водной среды, оказывающей воздействие на слои водорослей и бактерий.

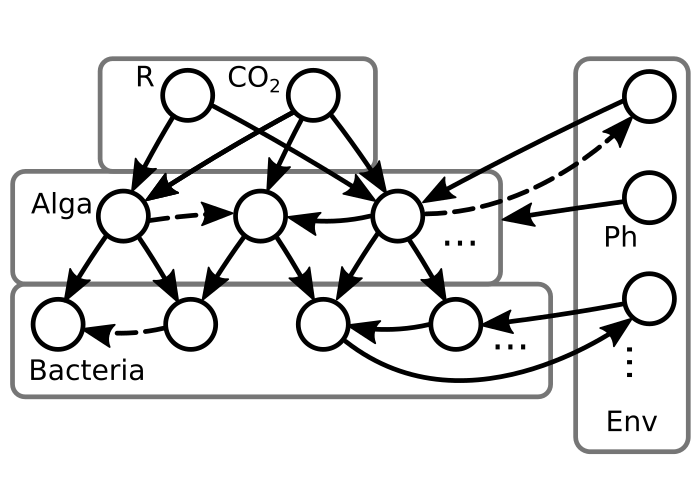


Рис.2. Результат визуализации графа взаимодействий (пример)

В программе Model Studio реализуются модули для представления динамической модели сообществ в виде а) тепловой карты (Рис. 1) б) общего графа (Рис. 2), в) графа data flow, аналогичного представлению вычислительного процесса в RapidMiner, г) дифференциального уравнения, д) перечня анализируемых критериев модели, е) результатов моделирования. На Рис. 3 приведен пример интерфейса программы, отображающий тепловую карту.

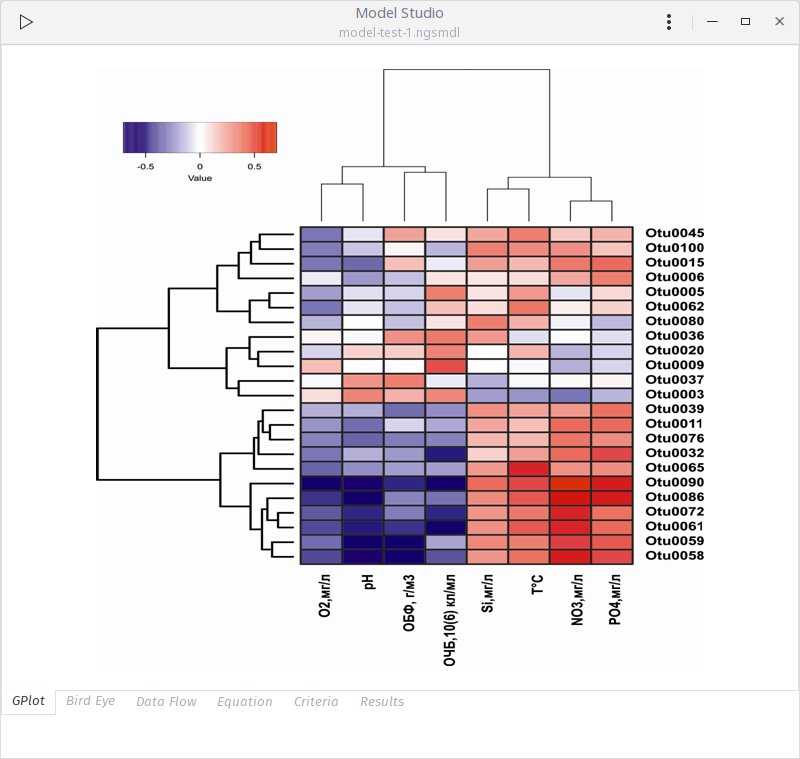


Рис. 3. Представление тепловой карты в Model Studio

Model Studio реализуется при помощи ряда систем программирования:

1. Python позволяет эффективно разрабатывать комплексные системы (склеивать компоненты) из различных гетерогенных подсистем, а также выступать в качестве удобного языка программного управления приложением; кроме того, использование этого языка позволяет в дельнейшем обеспечить взаимодействие с разрабатываемым в проекте хранилищем данных;
2. R для проведения этапа интеллектуального анализа данных, в этой среде реализованы процедуры анализа результатов секвенирования проб и построения корреляционных диаграмм;
3. Библиотека matplotlib, позволяющая строить графики и диаграммы и встраивать их в интерфейс пользователя, а также проводить экспорт в форматы, используемые в публикации данных;
4. SWI Prolog использована для формализации знаний этапа анализа структуры тепловой диаграммы и построения графа, а также для задания качественных критериев оценки результатов модели;
5. Библиотеки graphviz, использованной для раскладки графа на плоскости в процессе построения его визуального представления, кроме этой библиотеки для задачи визуализации задействованы библиотеки python-igraph, позволяющей обрабатывать большие по объему графы, NetworkX, предназначенной для изучения структур динамических сетей в биологии, социуме и технике, и graph-tool, поддерживающая многоядерные вычислительные архитектуры, фильтрацию данных, стандартные форматы данных, оценку статистических параметров графов, топологические алгоритмы, а также, в некоторой степени, возможности логического вывода на статических данных;
6. Языка Vala, подобный Java/С++/С#, использован для реализации основного окна приложения и вычислительных процедур, язык представляет возможность разработки вычислительных процедур, т.к. в отличие от своих прототипов является компилируемым;
7. Среда разработки интерфейсов пользователя GTK+, которая позволяет интегрировать подсистемы приложения на уровне пользовательского интерфейса, а также разрабатывать кросс-платформенные приложения.

Таким образом, на первом этапе проекта разработана первая версия системы, интегрирующей различное программное обеспечение в рамках одного приложения. Реализован ряд функций визуализации промежуточных данных и результатов. Полученные результаты позволяют перейти к разработке подсистем, обеспечивающих исследователя возможностью проведения экспериментов с разрабатываемыми моделями, т.е. задачам следующего года исследований. Разработанные алгоритмы опробованы на задачах, носящий технический характер, т.е. направленных на тестирование алгоритмов и систем приложения.

***Список литературы***

1. Patrick Royston (1995) Remark AS R94: A remark on Algorithm AS 181: The *W* test for normality. Applied Statistics, **44**, 547–551.
2. Gregory R. Warnes, Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz and Bill Venables (2015). gplots: Various R Programming Tools for Plotting Data. R package version 2.17.0. <http://CRAN.R-project.org/package=gplots>
3. Hollander M., Wolfe D.A. (1973), Nonparametric Statistical Methods. New York: John Wiley & Sons. Pages 185–194 (Kendall and Spearman tests).