

# СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ХАРАКТЕРИСТИК ПЕРКОЛЯЦИОННОГО КЛАСТЕРА

П. В. Москалев\*, К. В. Гребенников\*\*, В. В. Шитов\*\*

\* Воронежский государственный аграрный университет им. К.Д. Глинки

\*\* Воронежская государственная технологическая академия

Поступила в редакцию 18.04. 2011 г.

**Аннотация:** В данной работе рассматриваются статистические методы оценивания характеристик математической модели перколяции узлов. Преимущества предлагаемой методики демонстрируются на примере вычисления интервальной оценки массовой фрактальной размерности по выборочной совокупности перколяционных кластеров, формируемой с помощью метода статистических испытаний.

**Ключевые слова:** математическое моделирование, метод статистических испытаний, перколяция узлов, перколяционный кластер, массовая фрактальная размерность.

**Annotation:** In this paper we study statistical methods of parameters estimation of the site percolation model. Advantages of the proposed method is demonstrated for the computing of the confidence interval of mass fractal dimension of a percolation clusters sampling, formed by the Monte Carlo method.

**Key words:** mathematical modeling, Monte Carlo methods, site percolation, percolation cluster, mass fractal dimension

## 1. ВВЕДЕНИЕ

В классическом определении критического явления основной акцент делается на описании поведения вещества в окрестности точки фазового перехода. При этом, если объем новой фазы увеличивается постепенно, то говорят о фазовом переходе первого рода, если же сосуществование двух фаз исключено и процесс имеет скачкообразный характер, то речь идет о фазовом переходе второго рода [1]. С более общих позиций критические явления характеризуют процесс перехода некоторой системы из одного устойчивого состояния в другое, а о принадлежности этого перехода к первому или второму роду можно говорить в зависимости от того, как быстро он происходит. Одной из важных особенностей фазовых переходов второго рода является связь процесса формирования новой фазы с индивидуальными флуктуациями, которые в окрестности критической точки приобретают коррелированный характер с радиусом корреляции, сопоставимым с размером системы.

Математическая теория перколяции занимается изучением так называемых геометрических

фазовых переходов на взвешенных ориентированных однородных графах (в простейшем случае на решетках). Основной особенностью перколяционных решеток является зависимость весовых коэффициентов вершин и/или дуг от пространственного и вероятностного распределения псевдослучайных чисел. В соответствии с принятой в теории перколяции терминологией, ребра графа называют связями, его вершины – узлами, а основными объектами изучения являются связанные подмножества узлов решетки, называемые кластерами. В подобной системе фазовый переход второго рода соответствует появлению перколяционного кластера, размер которого потенциально бесконечен. В ограниченной области такой кластер идентифицируют по размеру, соответствующему размеру всей перколяционной решетки.

Для реализации всех описанных в данной статье алгоритмов была использована свободная система статистической обработки данных и программирования GNU/R [2], возникшая в 1993 году как свободная альтернатива языку S, разработанного в конце 1970-х годов в компании Bell Labs специально для решения задач вычислительной статистики. По замыслу ее создателей, система R должна была обладать легко расширяемой модульной архитектурой

при сохранении быстродействия, присущего программам, написанным на FORTRAN.

В настоящее время сборки системы R существуют для наиболее распространенных семейств операционных систем: GNU/Linux, Apple Mac OS X и Microsoft Windows, а в распределенных хранилищах системы R по состоянию на середину апреля 2011 года доступны для свободной загрузки почти 3000 пакетов расширения, ориентированных на специфические задачи обработки данных, возникающие в финансовом анализе, генетике, экологии, фармацевтике и многих других прикладных областях [3]. Из учебной литературы на русском языке по использованию системы R можно указать на пособие [4], вышедшее в прошлом году.

## 2. ИЗОТРОПНАЯ ПЕРКОЛЯЦИЯ УЗЛОВ

Одна из простейших моделей некоррелированной перколяции получается на изотропной квадратной решетке  $[0, 1]^2$  при взвешивании ее узлов стандартной равномерно распределенной последовательностью псевдослучайных чисел  $u_{xy} \sim U[0, 1]$ . Состав подмножества узлов кластера определяется итерационной процедурой сравнения псевдослучайных чисел  $u_{xy} < p$ , соответствующих узлам  $(x, y)$  из единичной окрестности текущего подмножества  $\{(x, y)\}$ , с фиксированным числом  $0 < p \leq 1$ , соответствующим относительной доле достижимых узлов перколяционной решетки. Удовлетворяющие указанному неравенству узлы формируют на решетке кластеры, представляющие собой реализации псевдослучайного перколяционного процесса.

На рис. 1 хорошо видно, что при доле достижимых узлов, меньшей некоторого критического значения  $p = p_c - 0.02$ , где  $p_c = 0.5927\dots$ , перколяционный процесс является ограниченным, а при доле достижимых узлов, большей критического значения  $p = p_c + 0.02$ , – распространяется по всей решетке.

Реализация алгоритма построения подобных кластеров для плоской квадратной изотропной решетки показан в листинге 1.

**Листинг 1.** Построение кластера узлов на квадратной изотропной решетке

```
psi20 <- function(x=129,
p=0.592746,
sst=(x*x+1)/2, trg=c(), label=2)
{
  acc <- array(runif(x*x),
c(x,x))
  acc[sst] <- label
  acc[c(1,x),] <- acc[,c(1,x)] <-
label + 0.001
  repeat {
    acc[sst <- unique(c(
sst[acc[sst - x] < p] - x,
sst[acc[sst - 1] < p] - 1,
sst[acc[sst + 1] < p] + 1,
sst[acc[sst + x] < p] + x))] <-
label
    if (length(sst) < 1) break }
  return(acc) }
```

В качестве аргументов функции "psi20()" используются следующие переменные: "x" – линейный размер решетки; "p" – доля достижимых узлов решетки;

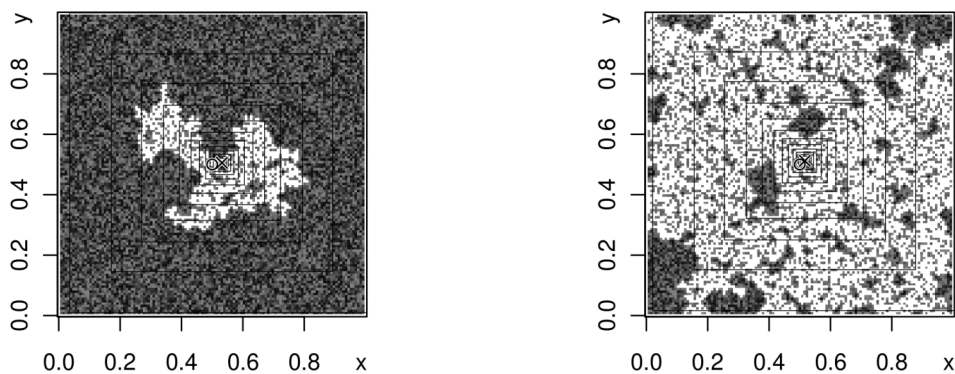


Рис. 1. Типичные реализации кластеров при перколяции из  $(0.5, 0.5)$  на квадратной изотропной решетке при  $p = p_c - 0.02$  (слева) и  $p = p_c + 0.02$  (справа)

"sst[]", "trg[]" – векторы индексов для стартового и целевого подмножеств; "label" – кластерная метка. В процессе выполнения функции "psi20()" формируется двумерный массив "acc[]", изначально заполняемый псевдослучайными числами  $u_{xy}$  со стандартным равномерным распределением, а элементы, соответствующие стартовому подмножеству, помечаются меткой "label". Затем, на базе цикла с постусловием организуется итерационный процесс в ходе которого узлы, принадлежащие единичной окрестности текущего стартового подмножества и удовлетворяющие условию  $u_{xy} < p$ , помечаются заданной меткой "label" и формируют новое стартовое подмножество для следующей итерации. Тогда, появление на очередной итерации вектора "sst[]" нулевой длины означает выход из цикла и возврат массива "acc[]" в качестве результата функции "psi20()".

### 3. СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ХАРАКТЕРИСТИК РЕАЛИЗАЦИИ КЛАСТЕРОВ

Статистически самоаффинным называют множество, допускающее разбиение на конечное число непересекающихся подмножеств, обладающих статистическими характеристиками, идентичными полному множеству с точностью до некоторого аффинного преобразования. Одной из важнейших структурных характеристик подобного множества [5] является массовая фрактальная размерность  $d_{c1}$ , определяемая методом наименьших квадратов для единичной реализации перколяционного кластера как коэффициент линейной регрессии логарифма числа узлов кластера  $n_i$ , покрываемых квадратом текущего размера  $r_i$ , к логарифму этого размера ( $\ln r_i, \ln n_i$ ):

$$\ln n_i = d_{c1} \ln r_i + d_{c0} + e_{ci}, \sum (e_{ci})^2 \rightarrow \min, \quad (1)$$

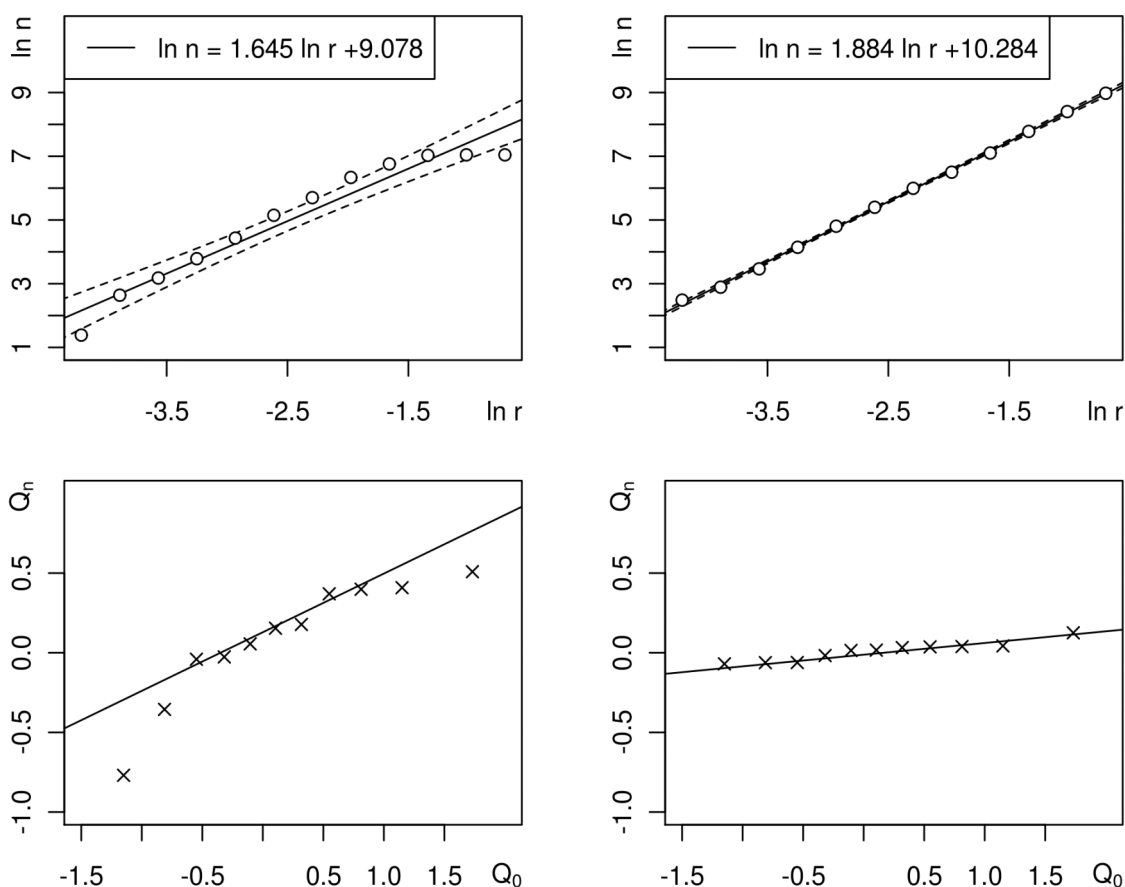


Рис. 2. Линейные модели (вверху) и Q–Q графики остатков  $e_{ci}$  (внизу) к оценке массовой фрактальной размерности  $d_{c1}$  для показанных на рис. 1 реализаций кластеров при  $p = p_c - 0.02$  (слева) и  $p = p_c + 0.02$  (справа)

где  $e_{ci}$  – минимизируемые методом наименьших квадратов отклонения эмпирических точек от линии регрессии. Центром симметрии покрывающего множества является центр масс анализируемого кластера, определяемый по выборочным средним координатам его узлов. Для показанных на рис. 1 кластеров стартовое подмножество расположено в точке (0.5, 0.5) и обозначено символом “о”, центры масс кластеров расположены в точках (0.532, 0.506) и (0.517, 0.512) и обозначены символами “х”.

Решение задачи линейной регрессии (1) для статистического оценивания массовой фрактальной размерности  $d_{c1}$  анализируемых кластеров показано на рис. 2, а соответствующие  $p_c - 0.02$  и  $p_c + 0.02$  долям достижимых узлов точечные оценки  $d_{c1}$  равны 1.645 и 1.884 при коэффициентах детерминации  $R^2$ , близких к единице: 0.940 и 0.999.

Символами “о” на рис. 2 вверху показаны эмпирические точки ( $\ln r_i, \ln n_i$ ), сплошными линиями – линии регрессии, а штриховыми – их 0.95-доверительные интервалы, методика построения которых описана, например, в [4].

Заметим, что из-за ограниченности перколяционного процесса при докритических значениях доли достижимых узлов  $p = p_c - 0.02$  средний радиус доверительного интервала для получаемой линии регрессии оказывается значительно более широким, а распределение отклонений эмпирических значений от теоретических, судя по всему, слабо соответствует гипотезе о нормальном распределении ошибок.

Последний вывод подтверждается при сравнительном анализе квантиль–квантильных (Q–Q) графиков остатков  $e_{ci}$ , показанных на рис. 2 внизу. Символами “х” показаны соответствующие  $e_{ci}$  теоретические и эмпирические квантили, а сплошной линией – теоретическая кривая нормального распределения остатков  $F_0(e)$ .

На Q–Q графиках хорошо видно, что при  $p < p_c$  эмпирическое распределение остатков  $e_{ci}$  будет сильно отличаться от нормального, что существенно снижает достоверность оценок фрактальной размерности реализаций кластеров  $d_{c1}$  на этих режимах, не взирая на высокие значения коэффициента детерминации  $R^2$ .

#### 4. СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ХАРАКТЕРИСТИК ВЫБОРКИ КЛАСТЕРОВ

Из курса математической статистики хорошо известно, что повышение достоверности получаемых оценок  $d_{c1}$  достигается за счет построения репрезентативной выборочной совокупности, что приводит к повышению вычислительной сложности задачи, пропорциональному объему используемой выборки.

Как показывают наши недавние исследования [6], характеристики перколяционного процесса зависят не только от размера и типа решетки, но и от взаимного расположения на решетке стартового и целевого подмножеств. Иными словами речь идет о статистической связи распределения относительных частот узлов перколяционной решетки с характеристиками выборочной совокупности кластеров.

Тогда, для нахождения оценки массовой фрактальной размерности по выборке перколяционных кластеров можно воспользоваться методом, предложенным в несколько другом контексте в работе [1]. Этот метод использует относительные частоты узлов в качестве весовых коэффициентов, что позволяет находить статистическую оценку размерности перколяционного кластера сразу для всей их совокупности.

Применительно к данной задаче массовая фрактальная размерность выборки кластеров  $d_{b1}$  будет определяться как коэффициент линейной регрессии логарифма относительных частот узлов  $v_i$ , покрываемых квадратом текущего размера  $r_i$ , к логарифму этого размера ( $\ln r_i, \ln v_i$ ):

$$\ln v_i = d_{b1} \ln r_i + d_{b0} + e_{bi}, \sum (e_{bi})^2 \rightarrow \min, \quad (2)$$

где  $e_{bi}$  – минимизируемые методом наименьших квадратов отклонения эмпирических точек от линии регрессии. Подобный переход вполне правомерен, так как при равных объемах выборки средние значения абсолютных частот узлов перколяционной решетки  $\langle n_i \rangle$  соответствуют их относительным частотам  $v_i$ , что в пределе приводит к идентичности математических ожиданий оценок параметров  $d_{c1}$  и  $d_{b1}$  моделей (1) и (2).

На рис. 3 показаны распределения относительных частот узлов решетки  $v_i$  для выборочных совокупностей объемом 500 кластеров, стартовое подмножество которых расположено

в центральной точке области  $(0.5, 0.5)$  и обозначено символом « $\circ$ ».

С учетом изотропной постановки задачи центры симметрии покрывающих множеств, обозначенные на рис. 3 символами « $\times$ », будут совпадать со стартовым множеством.

Решение задачи линейной регрессии для статистического оценивания массовой фрактальной размерности  $d_{b1}$  анализируемых кластеров показано в верхнем ряду на рис. 4, а соответствующие  $p_c - 0.02$  и  $p_c + 0.02$  долям точечные оценки  $d_{b1}$  равны 1.397 и 1.827 при коэффици-

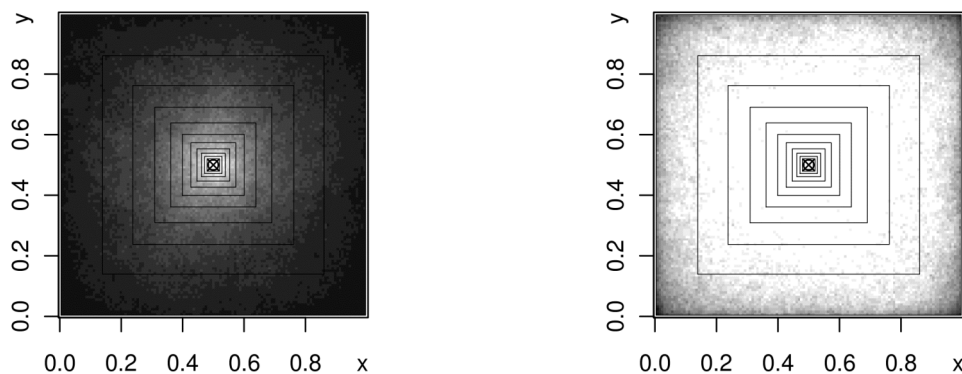


Рис. 3. Типичные выборки кластеров при перколяции из  $(0.5, 0.5)$  на квадратной изотропной решетке при  $p = p_c - 0.02$  (слева) и  $p = p_c + 0.02$  (справа)

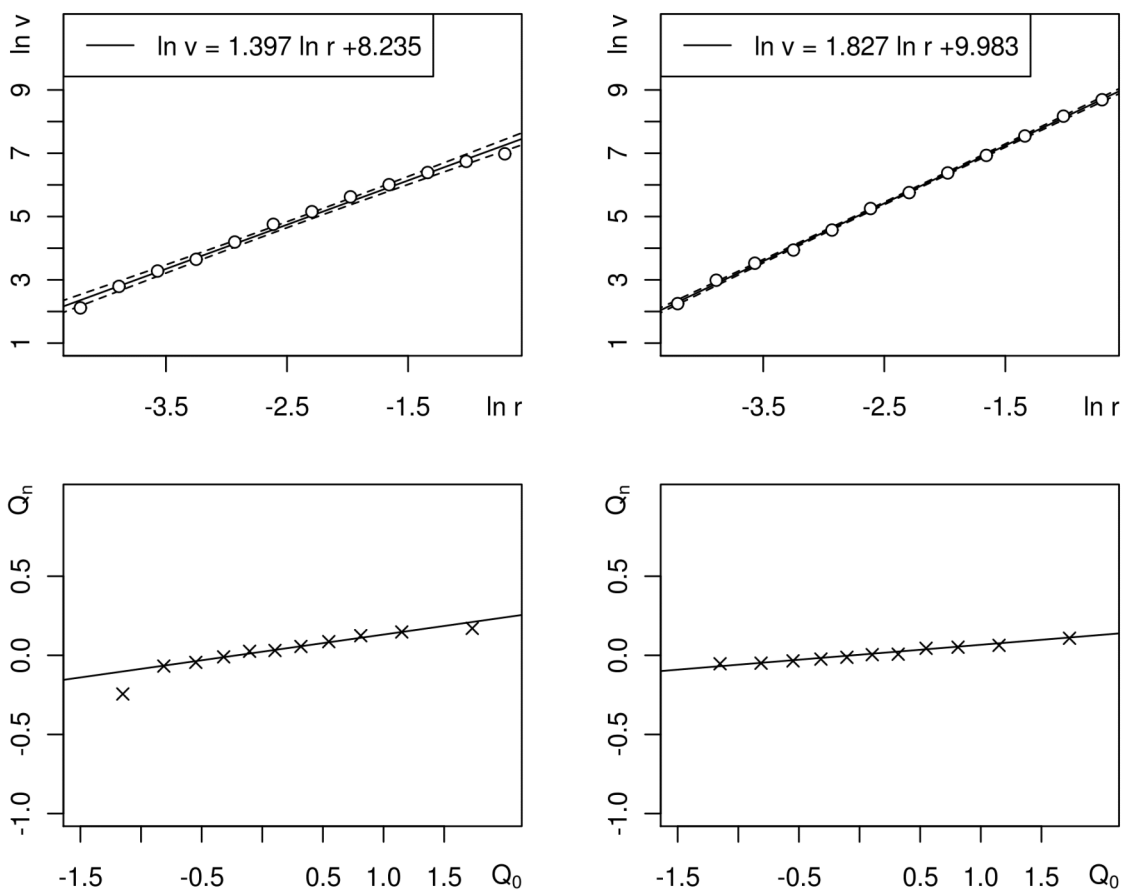


Рис. 4. Линейные модели (вверху) и Q–Q графики остатков  $e_{bi}$  (внизу) к оценке массовой фрактальной размерности  $d_{b1}$  для показанных на рис. 3 выборок кластеров при  $p = p_c - 0.02$  (слева) и  $p = p_c + 0.02$  (справа)

ентах детерминации  $R^2$ , чрезвычайно близких к единице: 0.991 и 0.999. Символами “о” на рис. 4 показаны эмпирические точки ( $\ln r_i$ ,  $\ln v_i$ ), сплошными линиями – линии регрессии, а штриховыми – их 0.95-доверительные интервалы.

Еще одним результатом вышеописанного подхода является построение соответствующих  $p_c - 0.02$  и  $p_c + 0.02$  долям достижимых узлов 0.95-доверительных интервалов массовой фрактальной размерности  $d_{b1}$  выборочных совокупностей  $I_{0.95}(d_{b1})$ , которые будут равны (1.311, 1.484) и (1.791, 1.863).

Заметим, что при докритическом значении доли достижимых узлов  $p = p_c - 0.02$  радиус доверительного интервала по-прежнему вдвое больший, а распределение отклонений эмпирических точек от линии регрессии хуже соответствует гипотезе о нормальном распределении ошибок, по сравнению со сверхкритическим режимом  $p = p_c + 0.02$ . Однако, сравнительный анализ Q-Q графиков остатков  $e_{bi}$ , приведенных на рис. 4 внизу, показывает, что даже при  $p < p_c$  эмпирическое распределение остатков  $e_{ci}$  оказывается близким к нормальному, что в целом позволяет говорить о достоверности оценок фрактальной размерности выборок кластеров  $d_{b1}$  на всех режимах. Кроме того, из общих соображений ясно, что в ростом доли достижимых узлов  $p \rightarrow 1$  массовая фрактальная размерность  $d_{b1}$  будет возрастать, стремясь к своему предельному значению  $d_{b1} \rightarrow 2$ .

## 5. ЗАКЛЮЧЕНИЕ

Построение репрезентативных векторов ( $\ln r_i$ ,  $\ln v_i$ ) для отдельных реализаций кластеров возможно лишь для критических и сверхкритических долей достижимых узлов  $p \geq p_c$ . Предлагаемый метод использует распределение относительных частот узлов решетки  $v_i$  и обеспе-

чивает корректный учет кластеров малых размеров в выборочной совокупности, что позволяет получать статистические оценки, более адекватные исследуемому явлению.

Еще одной важной особенностью предложенного метода является существенное сокращение трудоемкости вычислений за счет исключения необходимости решения задачи регрессии для каждой реализации кластера. Вполне очевидно, что указанная методика применима для получения достоверных и вычислительно эффективных оценок большинства статистически устойчивых характеристик перколяционных кластеров.

Вышеизложенное позволяет говорить о полезности описанного подхода для повышения эффективности применения перколяционных моделей как в теоретических, так и прикладных исследованиях.

## СПИСОК ЛИТЕРАТУРЫ

1. Москалев П. В. Анализ структуры перколяционного кластера / П. В. Москалев // ЖТФ, 2009. Т. 79. Вып. 6, С. 1–7.
2. The R Project for Statistical Computing // URL: <http://www.r-project.org>.
3. The Comprehensive R Archive Network // URL: <http://www.cran.r-project.org>.
4. Буховец А. Г. Статистический анализ данных в системе R / А. Г. Буховец, П. В. Москалев, В. П. Богатова, Т. Я. Бирючинская. – Воронеж: ВГАУ, 2010. – 124 с.
5. Тарасевич Ю. Ю. Перколяция: теория, приложения, алгоритмы / Ю. Ю. Тарасевич. – М.: Едиториал УРСС, 2002. – 112 с.
6. Москалев П. В. О распределении выборочных частот узлов перколяционного кластера / П. В. Москалев, В. В. Шитов, К. В. Гребенников // Информатика: проблемы, методология, технологии: материалы XI Международной научно-методической конференции. Т. 2. – Воронеж: ИПЦ ВГУ, 2011. – С. 54–58.

**Москалев Павел Валентинович** – кандидат технических наук, доцент кафедры высшей математики и теоретической механики Воронежского государственного аграрного университета им. К. Д. Глинки. Тел.: (473) 253-73-71; e-mail: moskalefff@gmail.com

**Moskalev Pavel Valentinovich** – Ph. D., Associate Professor, Department of Mathematics and Theoretical Mechanics, Voronezh State Agricultural University after K. D. Glinki. Tel.: +7 (473) 253-73-71; e-mail: moskalefff@gmail.com

**Гребенников Константин Владимирович** – аспирант кафедры промышленной энергетики Воронежской государственной технологической академии; e-mail: greb86@mail.ru

**Grebennikov Konstantin Vladimirovich** – Postgraduate Student, Department of Industrial Power Engineering, Voronezh State Technological Academy; e-mail: greb86@mail.ru

**Шитов Виктор Васильевич** – доктор технических наук, профессор, заведующий кафедрой промышленной энергетики Воронежской государственной технологической академии. Тел.: +7 (473) 244-44-66; e-mail: svw@list.ru

**Shitov Viktor Vasiljevich** – Doctor of Engineering Science, Full Professor, Head of Department of Industrial Power Engineering, Voronezh State Technological Academy. Tel.: +7 (473) 244-44-66; e-mail: svw@list.ru