

Приложение 1

**ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИЙ КОМПЛЕКС
«Управление электропотреблением техноценоза»**

Исследование техноценоза начинается со сбора статистической информации о его объектах. Сама по себе эта операция чрезвычайно трудоемкая и требующая многолетних систематических усилий целых научных коллективов. Однако после завершения сбора информации наступает не менее сложный этап глубокой статистической обработки полученных данных. Современное программное обеспечение позволяет осуществлять эффективную обработку данных практически любых объемов и сложности. По адресу: http://gnatukvi.ru/zip_files/task_mcd.zip можно скачать расчетные mathcad-программы с подробными комментариями, предназначенные для статистической обработки данных и динамического моделирования процесса управления электропотреблением техноценоза. В качестве примера здесь используется взаимосвязанный комплекс расчетно-графических модулей, написанных в среде математического пакета Mathcad.

После сбора статистической информации о техноценозе создается база данных, которая представляет собой неупорядоченную совокупность значений электропотребления объектов техноценоза. Рекомендуется базу данных создавать в виде двух файлов Microsoft Excel. В первом файле данные могут быть представлены в любом удобном для исследователя виде с необходимыми пояснениями и комментариями. Во втором файле должны содержаться исключительно числовые значения электропотребления объектов (единицы измерения – кВт·ч за временной промежуток), выстроенные последовательно друг за другом (в соответствии с тем порядком, в котором они следуют в первом файле) в ячейках электронной таблицы без каких-либо текстовых записей (этот файл должен состоять только из цифр). Данные формируются в виде двумерной таблицы, строки которой соответствуют временным интервалам, в течение которых исследовался техноценоз (как правило, это часы, дни, месяцы или годы), а столбцы – объектам техноценоза. Если рассматривается состояние техноценоза только на фиксированный момент времени, таблица состоит лишь из одной строки. В любом случае, в каждой ячейке таблицы содержится только одно число, соответствующее электропотреблению одного объекта на одном временном интервале. Файлы должны быть определенным образом названы и помещены в директорию [c:\mathcad_dat], которая должна быть благововременно создана в корневом каталоге диска «с:\». Рекомендации о том, как следует называть файлы, имеются в каждой из программ. Вместе с программами в архиве http://gnatukvi.ru/zip_files/task_mcd.zip приводится директория [mathcad_dat], содержащая подготовленные для расчетов дан-

ные по электропотреблению одного из реально существующих техноценозов, расположенных в Калининградской области. Ее можно скопировать в корневой каталог диска «с:\» и использовать в качестве примера в ходе демонстрации работы программ. Следует учитывать тот факт, что если директорию [с:\mathcad_dat] не создать и не поместить в нее все требуемые файлы, то расчетные программы работать не будут. Возможно размещение директории и в другом месте дискового пространства, а файлы с исходными данными можно назвать как-либо по-своему. Однако это неизбежно потребует соответствующего переименования файлов внутри каждой из расчетных программ везде, где выполняются операции импорта или экспорта данных. Для работы комплекса на компьютере пользователя должны быть установлены: ОС Windows, а также приложения Mathcad и Excel.

Если в данный момент времени по какой-либо причине в распоряжении пользователя комплекса нет удовлетворительной параметрической базы данных техноценоза, то ее можно искусственно сгенерировать. Для этого предназначена программа «Генератор негауссовой выборки техноценологического типа», которая имеется в файле «task_00.xmcd» архива.

Ниже в данном приложении приводятся тексты отдельных программ информационно-аналитического комплекса с подробными инструкциями и комментариями, которые в 2003 – 2011 годах публиковались автором совместно с учениками А.Е. Севериным, С.Н. Гринкевичем, Д.В. Луценко, А.М. Дубовиком, О.Р. Кивчуном и А.А. Шейниным в журнале «Электрика». Аналитическое описание и результаты реализации программ можно найти в ряде работ [83,85,86,90,94-96,102-104,106,108,110,111,114-119].

Перечень расчетных программ комплекса

- Первичная обработка статистической информации по техноценозу.
- Аппроксимация ранговых распределений.
- Интервальное оценивание объектов техноценоза.
- Прогнозирование электропотребления в техноценозе.
- Нормирование электропотребления объектами техноценоза.
- Проверка данных на соответствие критериям H -распределения.
- Оценка потенциала энергосбережения техноценоза.
- G -метод прогнозирования электропотребления в техноценозе.
- Верификация исходной базы данных по электропотреблению.
- Оценка адекватности работы динамической адаптивной модели.
- Прогнозирование электропотребления методом АГК.
- Обработка ранговой параметрической поверхности методом SSA.
- GZ -анализ рангового параметрического распределения.
- Классификация техноценоза по электропотреблению.
- Генератор негауссовой выборки техноценологического типа.

[на содержание]

ПЕРВИЧНАЯ ОБРАБОТКА СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ ПО ТЕХНОЦЕНОЗУ

Подготовка данных

Прежде всего, программными средствами, осуществляется импорт данных из файла «data.xls» в тело программы Mathcad и задается начало отсчета. Делается это средствами импорта следующим образом:

$$V := \text{Импортировать}(\text{C:\mathcad_dat\data.xls})$$

$$\text{ORIGIN} := 1$$

Сформированная матрица V содержит информацию об исследуемом техноценозе. Причем, каждая строка соответствует определенному году, а каждый столбец – объекту. Для иллюстрации в данной программе использовались собранные за шесть лет данные по электропотреблению техноценоза (69 объектов), расположенного в Калининградской области.

Далее матрицу можно вывести и просмотреть в теле программы.

$$V =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 1 | $1.81 \cdot 10^6$ | $1.268 \cdot 10^6$ | $2.328 \cdot 10^5$ | $5.817 \cdot 10^5$ | $4.386 \cdot 10^5$ | $7.438 \cdot 10^5$ | $2.891 \cdot 10^4$ |
| 2 | $1.447 \cdot 10^6$ | $1.332 \cdot 10^6$ | $8.567 \cdot 10^5$ | $5.244 \cdot 10^5$ | $4.027 \cdot 10^5$ | $6.371 \cdot 10^5$ | $2.636 \cdot 10^4$ |
| 3 | $1.474 \cdot 10^6$ | $1.613 \cdot 10^6$ | $8.373 \cdot 10^5$ | $5.784 \cdot 10^5$ | $5.179 \cdot 10^5$ | $7.094 \cdot 10^5$ | $2.724 \cdot 10^4$ |
| 4 | $1.351 \cdot 10^6$ | $1.614 \cdot 10^6$ | $3.494 \cdot 10^5$ | $4.025 \cdot 10^5$ | $5.673 \cdot 10^5$ | $3.652 \cdot 10^5$ | $3.26 \cdot 10^4$ |
| 5 | $1.162 \cdot 10^6$ | $1.304 \cdot 10^6$ | $3.971 \cdot 10^5$ | $3.274 \cdot 10^5$ | $4.366 \cdot 10^5$ | $8.78 \cdot 10^4$ | $2.701 \cdot 10^4$ |
| 6 | $1.628 \cdot 10^6$ | $1.775 \cdot 10^6$ | $4.101 \cdot 10^5$ | $3.439 \cdot 10^5$ | $3.379 \cdot 10^5$ | $7.25 \cdot 10^4$ | $3.02 \cdot 10^4$ |

Для упрощения дальнейшей работы в Mathcad матрицу можно транспонировать, чтобы ее колонки являлись векторами параметров:

$$W := V^T$$

Получение табулированного рангового распределения

Для получения табулированного рангового параметрического распределения необходимо имеющиеся неупорядоченные фактические данные проранжировать. Ранжирование располагает данные в порядке уменьшения значений исследуемого параметра. Задача решается применением

средств программирования Mathcad. Приведенная ниже подпрограмма позволяет обработать данные любого объема, используя оператор цикла.

$$\text{Zipf}(Y) := \left| \begin{array}{l} \text{for } i \in 1.. \text{cols}(Y) \\ \quad \left| \begin{array}{l} c \leftarrow \text{sort}(Y^{\langle i \rangle}) \\ b^{\langle i \rangle} \leftarrow \text{reverse}(c) \end{array} \right. \\ \quad b \end{array} \right.$$

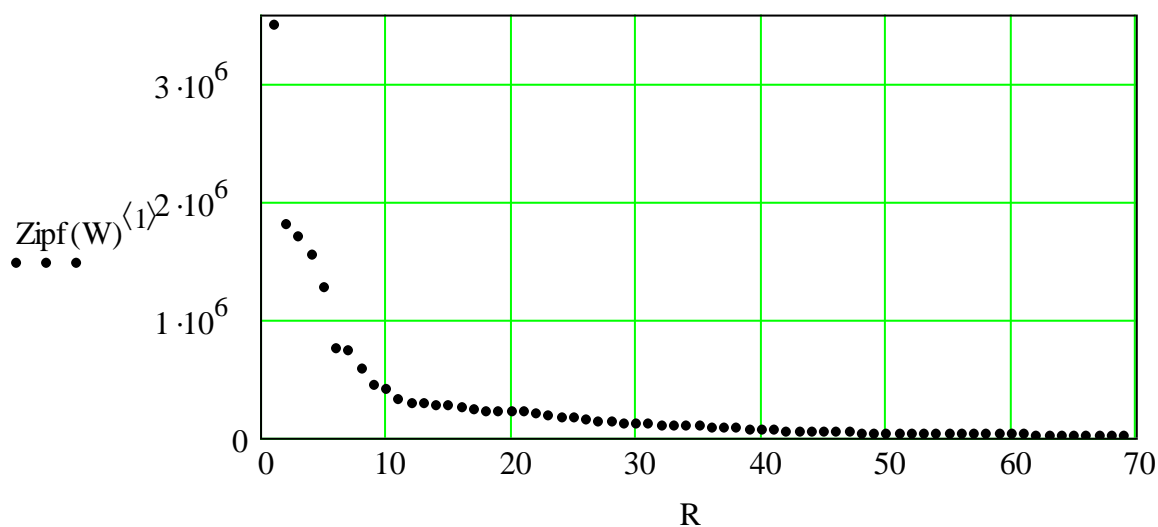
Функция Zipf позволяет сформировать матрицу, столбцы которой являются векторами, представляющими собой, по сути, табулированное ранговое параметрическое распределение по электропотреблению объектов техноценоза на отдельных временных интервалах.

С целью подготовки данных для дальнейшей работы программы определяется количество объектов и формируется вектор рангов:

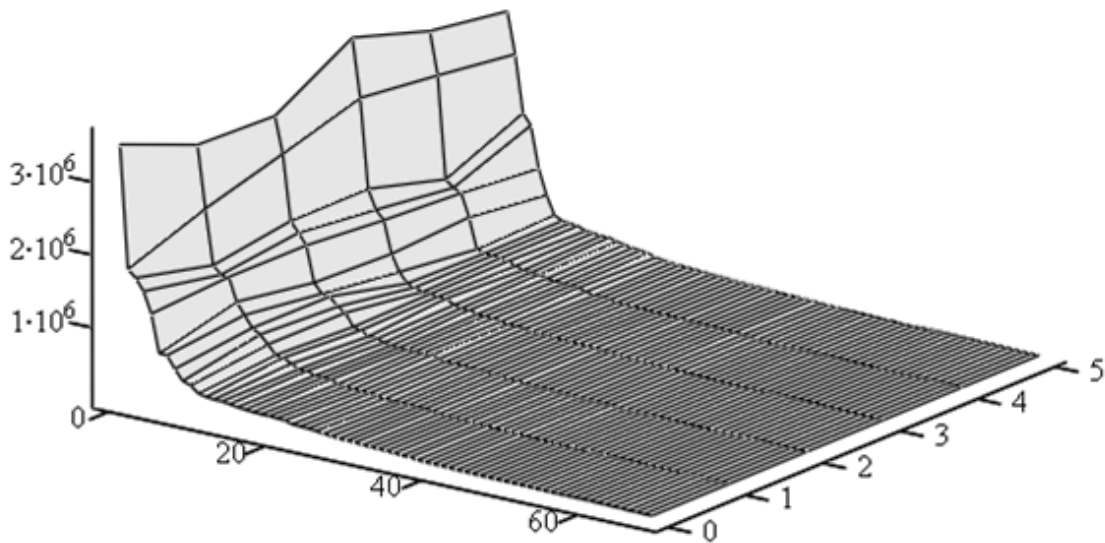
$$\begin{aligned} n &:= \text{length}(\text{Zipf}(W)^{\langle 1 \rangle}) & n &= 69 \\ r &:= 1..n \\ R_r &:= r \end{aligned}$$

Графическое представление данных

Ранговое параметрическое распределение техноценоза по электропотреблению может быть изображено в графической форме. При этом рекомендуется использовать двумерные и трехмерные графики (см. ниже).



Ранговое параметрическое распределение техноценоза (на первый год): абсцисса – ранг объекта; ордината – электропотребление, кВт·ч



Zipf(W)

Трехмерная ранговая параметрическая поверхность техноценоза по электропотреблению: абсцисса – ранг объекта; ордината – временной интервал (в данном случае – номер года исследования); аппликата – электропотребление, кВт·ч; Zipf(W) – наименование изображаемой функции

Определение рангов для каждого объекта

Матрица рангов техноценоза необходима для оценки динамики движения объектов по ранговой поверхности параметрического распределения. Матрица определяется с помощью следующей подпрограммы:

```

Rang(W) := | for j ∈ 1..cols(W)
            |   i ← 1
            |   a ← W(j)
            |   c ← sort(a)
            |   b ← reverse(c)
            |   for m ∈ 1..rows(a)
            |     | for n ∈ 1..rows(b)
            |       |   Li,j ← n if am = bn
            |       |   i ← i + 1
            |   L

```

$$\text{Rang} := \text{Rang}(V^T)^T$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|----|----|----|----|----|----|---|----|----|----|----|----|----|----|----|----|----|----|
| 1 | 2 | 5 | 17 | 8 | 9 | 7 | 53 | 15 | 1 | 4 | 11 | 21 | 69 | 3 | 28 | 19 | 20 | 23 | 13 | 16 |
| 2 | 3 | 5 | 7 | 10 | 11 | 9 | 53 | 18 | 1 | 4 | 14 | 16 | 69 | 2 | 43 | 6 | 13 | 20 | 22 | 15 |
| 3 | 4 | 3 | 6 | 9 | 10 | 7 | 52 | 15 | 1 | 5 | 13 | 18 | 69 | 2 | 42 | 23 | 16 | 20 | 19 | 12 |
| 4 | 4 | 3 | 12 | 9 | 7 | 10 | 52 | 15 | 1 | 5 | 22 | 18 | 69 | 2 | 49 | 27 | 16 | 14 | 11 | 13 |
| 5 | 5 | 3 | 9 | 12 | 7 | 34 | 53 | 19 | 1 | 4 | 15 | 18 | 68 | 2 | 40 | 30 | 17 | 14 | 8 | 13 |
| 6 | 4 | 3 | 8 | 11 | 12 | 40 | 54 | 18 | 1 | 5 | 16 | 15 | 69 | 2 | 24 | 52 | 21 | 25 | 9 | 10 |

Вычисление коэффициента конкордации

Коэффициент конкордации, определенный для совокупности ранговых параметрических распределений, характеризует степень взаимосвязанности техноценоза. Он показывает согласованность перемещения объектов по ранговой поверхности при переходе от одного временного интервала к последующему. Для определения коэффициента конкордации средствами Mathcad выполняются следующие операции:

- определение суммы рангов для каждого объекта:

$$i := 1.. \text{rows}(\text{Rang})$$

$$j := 1.. \text{cols}(\text{Rang})$$

$$S_j := \sum_i \text{Rang}_{i,j}$$

$$S^T = \begin{array}{|c|c|c|c|c|c|c|c|c|c|} \hline & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ \hline 1 & 22 & 22 & 59 & 59 & 56 & 107 & 317 & 100 & 6 & 27 \\ \hline \end{array}$$

- определение общей суммы рангов техноценоза:

$$SS := \sum_{i=1}^n S_i$$

$$SS = 1.449 \times 10^4$$

- вычисление среднего для рангов техноценоза:

$$SR := \frac{SS}{\text{rows}(S)}$$

$$SR = 210$$

- определение отклонения и квадрата отклонения сумм рангов для каждого объекта от средней рангов техноценоза:

$$D := S - SR$$

$$D^T = \begin{array}{|c|c|c|c|c|c|c|c|c|c|} \hline & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ \hline 1 & -188 & -188 & -151 & -151 & -154 & -103 & 107 & -110 & -204 & -183 \\ \hline \end{array}$$

$$DKV := D^2$$

$$DKV^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|--------------------|--------------------|-------------------|-------------------|--------------------|--------------------|--------------------|
| 1 | $3.534 \cdot 10^4$ | $3.534 \cdot 10^4$ | $2.28 \cdot 10^4$ | $2.28 \cdot 10^4$ | $2.372 \cdot 10^4$ | $1.061 \cdot 10^4$ | $1.145 \cdot 10^4$ |

- определение общей суммы квадратов отклонений рангов от средней для рангов техноценоза, а также количества распределений:

$$SDKV := \sum_{j=1}^n DKV_j$$

$$SDKV = 9.536 \times 10^5$$

$$m := \text{rows}(\text{Rang})$$

$$m = 6$$

- вычисление коэффициента конкордации K:

$$K := \frac{12 \cdot SDKV}{m^2 \cdot (n^3 - n)}$$

$$K = 0.968$$

Для автоматизации оценки результатов вычислений используется специальная подпрограмма. Функция RESULT возвращает значение «yes», если коэффициент конкордации значим, «no», – если данный коэффициент не значим и «error», – если в вычислениях допущена ошибка.

$$\text{RESULT}(y) := \begin{cases} \text{return "no"} & \text{if } 0 \leq y < 0.5 \\ \text{return "yes"} & \text{if } 0.5 \leq y \leq 1 \\ \text{return "error"} & \text{otherwise} \end{cases}$$

$$\text{RESULT}(K) = \text{"yes"}$$

Характерно, что для совокупности данных, используемых в качестве примера, коэффициент конкордации значим, что свидетельствует о взаимосвязанности исследуемого техноценоза (во всяком случае, по параметру электропотребления). Данный вывод позволяет использовать созданную базу данных для интервального оценивания, нормирования и прогнозирования электропотребления объектов техноценоза.

Для сохранения данных вне программы (с целью их последующего использования другими программами) их рекомендуется экспортировать в файлы. Реализуется операция стандартными средствами экспорта данных Mathcad. При этом автоматически в директории «с:\mathcad_dat» создаются соответствующие рабочие файлы с расширением «md».

$$\text{WRITEPRN}(\text{"c:\mathcad_dat\Zipf.md"}) := \text{Zipf}(W)$$

$$\text{WRITEPRN}(\text{"c:\mathcad_dat\R.md"}) := R$$

$$\text{WRITEPRN}(\text{"c:\mathcad_dat\Rang.md"}) := \text{Rang}$$

[на содержание]

АППРОКСИМАЦИЯ РАНГОВЫХ РАСПРЕДЕЛЕНИЙ

С точки зрения последующей статистической обработки данных большое значение имеет аппроксимация эмпирических ранговых распределений. Ее задача заключается в подборе аналитической зависимости, наилучшим образом описывающей совокупность точек. Мы задаем в качестве стандартной двухпараметрическую гиперболическую форму. Выбор формы объясняется традиционно сложившимся подходом среди исследователей, занимающихся ранговым анализом. Безусловно, данная форма далеко не самая совершенная, однако она обладает неоспоримым достоинством – сводит задачу аппроксимации к определению всего двух параметров. Аппроксимация может осуществляться различными методами, каждый из которых обладает как достоинствами, так и недостатками. Учитывая высокую скорость расчетов в среде Mathcad, в данной программе аппроксимация будет осуществляться сразу несколькими методами. После этого автоматически будет выбираться наиболее корректный вариант.

Аппроксимация рангового распределения проводится после первичной обработки статистической информации по техноценозу, которая должна быть осуществлена программными средствами Mathcad. Следует напомнить, что по результатам первичной обработки формируется ряд файлов с расширением «*md*», размещенных в директории «*c:\mathcad_dat*», которая должна быть создана в корневом каталоге диска «*c:*».

Метод наименьших модулей

В качестве критерия близости искомой функции в данном методе используется минимум суммы модулей разностей эмпирических и теоретических данных, рассчитанных по уравнению регрессии.

В качестве первой операции программы задается начало отсчета и считывается из рабочих файлов первичная исходная информация. При этом формируются матрицы рангов R и табулированного рангового распределения объектов техноценоза по электропотреблению V .

```
ORIGIN := 1
```

```
R := READPRN("c:\mathcad_dat\R.md")  
V := READPRN("c:\mathcad_dat\Zipf.md")
```

Далее осуществляется ряд подготовительных операций, включающих индексацию и транспонирование полученных выше матриц.


```

n := rows(V)
n = 69
i := 1..n
Y := V<1>

```

После этого задается аппроксимационная форма (в данном случае – двухпараметрическая) и определяется целевая оптимизационная функция метода наименьших модулей с начальными приближениями.

$$f(R, w1, b) := \frac{w1}{R^b}$$

$$\Phi(w1, b) := \sum_i |(Y)_i - f(R_i, w1, b)|$$

```
w1 := Y1
```

```
b := 1
```

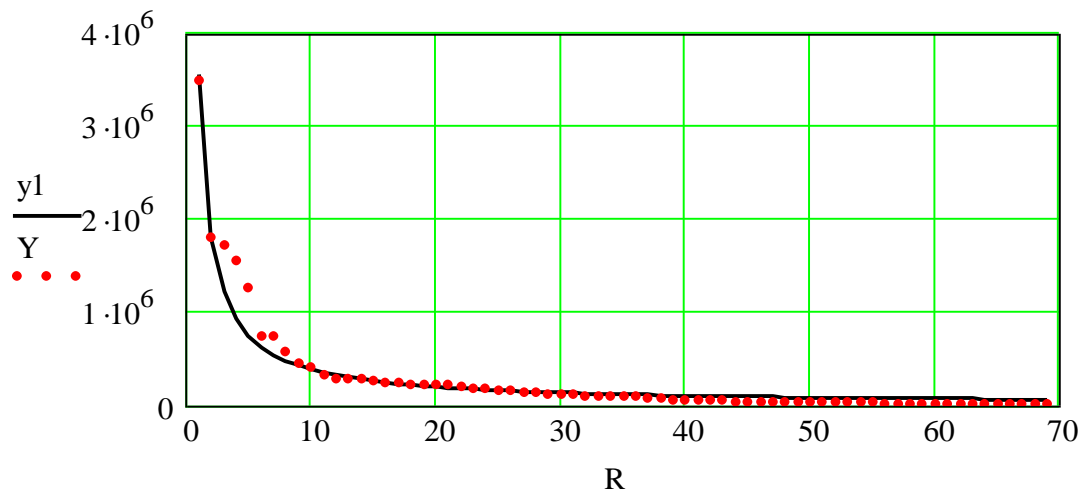
Поиск минимума функции «Ф» осуществляется с использованием встроенной функции Minimize, в которой реализуются алгоритмы оптимизации, основывающиеся на итерационных вычислениях с использованием квазиньютоновского метода или метода сопряженных градиентов (выбор осуществляется в контекстном меню программы). Оба итерационных метода основываются на последовательном приближении к точке минимума. По результатам оптимизации определяется вектор m (параметры аппроксимационного выражения). Следует отметить, что в данной программе в качестве примера используются данные об электропотреблении за шесть лет одного из техноценозов, размещенных на территории Калининградской области (данный пример использовался в предыдущей программе). При этом аппроксимация ранговых распределений осуществляется отдельно по каждому временному интервалу (повекторно).

```
m := Minimize(Φ, w1, b)
```

$$m = \begin{pmatrix} 3.552 \times 10^6 \\ 0.973 \end{pmatrix}$$

$$y1 := \frac{m_1}{R^{m_2}}$$

Далее появляется возможность наглядно вывести результаты аппроксимации в графической форме (см. рис. ниже).



Ранговое параметрическое распределение техноценоза: абсцисса – ранг объекта; ордината – электропотребление, кВт·ч; точки – эмпирические данные; сплошная линия – аппроксимационная кривая, полученная методом наименьших модулей

Линейный метод наименьших квадратов

Известно, что функция может быть линеаризована посредством логарифмирования. Это позволяет при аппроксимации эффективно применять линейные регрессионные методы, которые продемонстрированы в теле программы ниже. Вводим формальные параметры, логарифмируя вектор рангов и эмпирические данные по электропотреблению:

$$P_i := 1$$

$$\begin{aligned} \overrightarrow{RL} &:= \ln(R) \\ \overrightarrow{YL} &:= \ln(Y) \end{aligned}$$

Вычисляем коэффициенты линейной регрессии при помощи метода наименьших квадратов матричным способом:

$$\begin{aligned} X &:= \text{augment}(P, RL) \\ B &:= (X^T \cdot X)^{-1} \cdot (X^T \cdot YL) \end{aligned}$$

$$B = \begin{pmatrix} 16.53 \\ -1.585 \end{pmatrix}$$

Получив значения линейных коэффициентов регрессии, вычисляем их для исходной нелинейной функции рангового параметрического распределения и восстанавливаем первоначальную зависимость:

$$w11 := \exp(B_1)$$

$$\beta 1 := -B_2$$

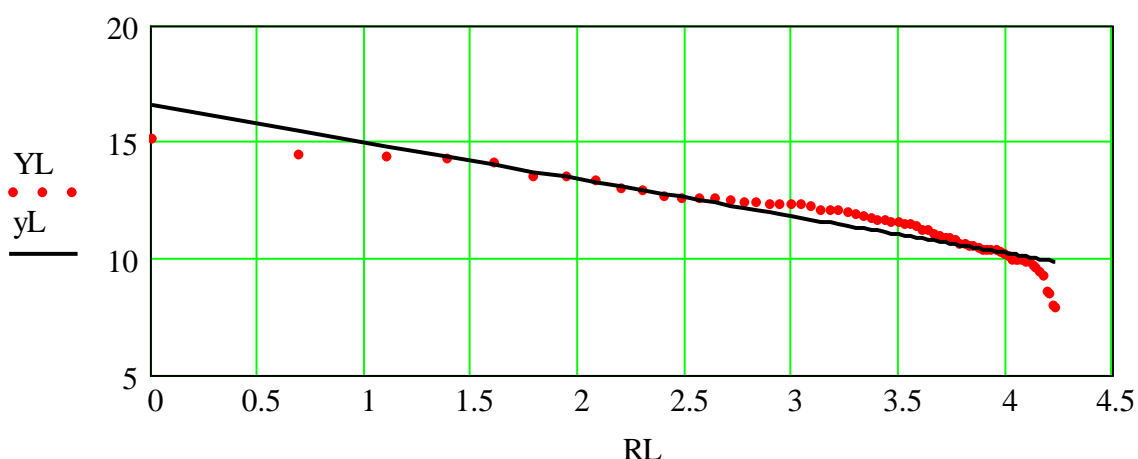
$$w11 := 1.2 \cdot 10^7$$

$$\beta 1 = 1.585$$

$$yL := B_1 + RL \cdot B_2$$

$$y2 := \frac{w11}{R^{\beta 1}}$$

Строим линейризованные графики:



Линейризованное ранговое распределение техноценоза: абсцисса – логарифм ранга объекта; ордината – логарифм электропотребления, $\ln(\text{кВт}\cdot\text{ч})$; точки – эмпирические данные; сплошная линия – аппроксимационная прямая, полученная линейным методом наименьших квадратов

Метод наименьших квадратов

Суть метода заключается в отыскании параметров аналитической зависимости, которые минимизируют сумму квадратов отклонений эмпирических значений (реально полученных в ходе рангового анализа техноценоза) от значений, рассчитанных по аппроксимационной зависимости. Ниже приводится текст подпрограммы, которая вычисляет коэффициенты

регрессии. Входными аргументами функции «Т» являются вектора рангового параметрического распределения по электропотреблению V (рассчитанные для одного временного интервала) и рангов R .

$$T(W, r) := \left| \begin{array}{l} n \leftarrow \text{length}(W) \\ \beta \leftarrow \frac{\sum_{i=1}^n \ln(W_i) \cdot \sum_{i=1}^n \ln(r_i) - n \cdot \sum_{i=1}^n \ln(W_i) \cdot \ln(r_i)}{\left(\sum_{i=1}^n \ln(r_i) \right)^2 - n \cdot \sum_{i=1}^n (\ln(r_i))^2} \\ W1 \leftarrow \exp \left[\frac{1}{n} \cdot \left(\sum_{i=1}^n \ln(W_i) + \beta \cdot \sum_{i=1}^n \ln(r_i) \right) \right] \\ WW \leftarrow \text{stack}(\beta, W1) \\ WW \end{array} \right.$$

$$T(V^{(1)}, R) = \begin{pmatrix} 1.585 \\ 1.509 \times 10^7 \end{pmatrix}$$

$$T(V^{(2)}, R) = \begin{pmatrix} 1.629 \\ 1.699 \times 10^7 \end{pmatrix}$$

При необходимости вычислить коэффициенты регрессии для всех ранговых параметрических распределений за рассматриваемый интервал времени, можно применить следующую подпрограмму:

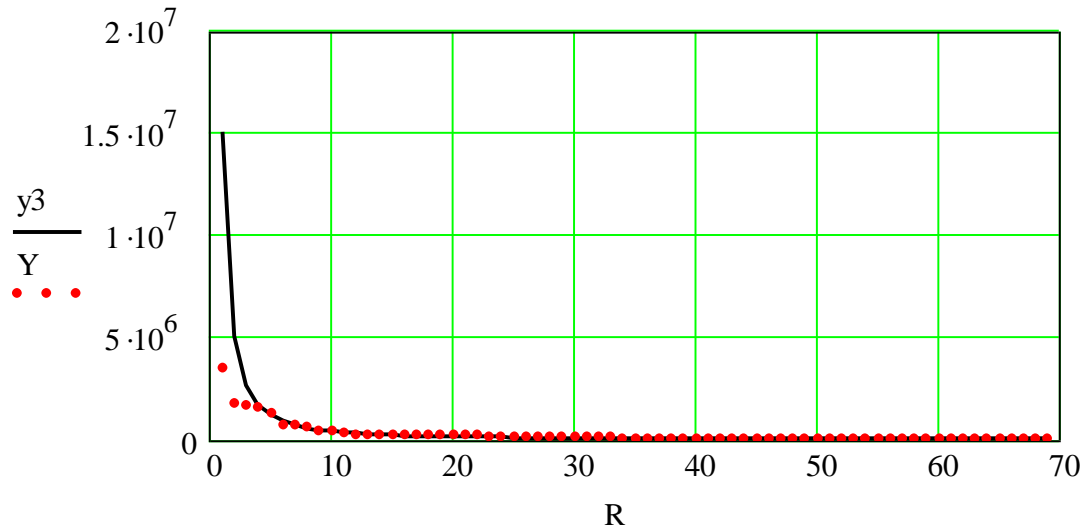
$$H(Z) := \left| \begin{array}{l} j \leftarrow 1 \\ \text{while } j < \text{cols}(V) + 1 \\ \quad \left| \begin{array}{l} X^{(j)} \leftarrow Z(V^{(j)}, R) \\ j \leftarrow j + 1 \end{array} \right. \\ X \end{array} \right.$$

$$\beta := (H(T)^T)^{(1)}$$

$$W1 := (H(T)^T)^{(2)}$$

$$y_3 := \frac{W1_1}{R^{\beta_1}}$$

Далее следует визуализация результатов аппроксимации.



Ранговое параметрическое распределение техноценоза: абсцисса – ранг объекта; ордината – электропотребление, кВт·ч; точки – эмпирические данные; сплошная линия – аппроксимационная кривая, полученная методом наименьших квадратов

Примечательно, что оба метода наименьших квадратов (линейный и нелинейный) дали очень близкие результаты аппроксимационных параметров. Это позволяет в дальнейшем при сравнении с методом наименьших модулей использовать только нелинейный метод наименьших квадратов (как более корректный). Полученные данные необходимо сохранить.

`WRITEPRN("c:\mathcad_dat\Coef.md") := H(T)T`

Оценивание результатов аппроксимации

С целью выбора наиболее эффективного метода аппроксимации ранговых параметрических распределений полученные выше результаты статистически сравниваются. Для этого оценивается истинная ошибка:

$$\varepsilon_1 := Y - y_1 \qquad \varepsilon_2 := Y - y_2$$

$$\max\left(\overrightarrow{|\varepsilon_1|}\right) = 6.307 \times 10^5$$

$$\max(|\overrightarrow{\varepsilon 2}|) = 8.5 \times 10^6$$

$$\sum_{i=1}^n (|\overrightarrow{\varepsilon 1}|)_i = 4.14 \times 10^6$$

$$\sum_{i=1}^n (|\overrightarrow{\varepsilon 2}|)_i = 1.45 \times 10^7$$

$$\varepsilon_{otn1} := \overrightarrow{[\varepsilon 1 \cdot (Y^{-1})]}$$

$$\varepsilon_{otn2} := \overrightarrow{[\varepsilon 2 \cdot (Y^{-1})]}$$

$$\text{МНМ} := \sum_{i=1}^n (|\overrightarrow{\varepsilon_{otn1}}|)_i$$

$$\text{МНК} := \sum_{i=1}^n (|\overrightarrow{\varepsilon_{otn2}}|)_i$$

$$\text{МНМ} = 123.397$$

$$\text{МНК} = 37.64$$

$$\text{corr}(Y, y1) = 0.976$$

$$\text{corr}(Y, y2) = 0.9$$

В данном случае есть основания признать более корректным метод наименьших квадратов (значение переменной МНК существенно меньше МНМ). Результирующая аппроксимационная зависимость определяется средствами Mathcad с использованием подпрограммы:

$$\text{REZ}(x, y, z, c) := \begin{cases} z & \text{if } x < y \\ c & \text{otherwise} \end{cases}$$

$$W := \text{REZ}(\text{МНМ}, \text{МНК}, m_1, W1_1)$$

$$\beta := \text{REZ}(\text{МНМ}, \text{МНК}, m_2, \beta_1)$$

Таким образом, получаем наилучшие параметры для аппроксимационной зависимости рангового параметрического распределения:

$$W = 1.509 \times 10^7$$

$$W(r) := \frac{W}{r^\beta}$$

$$\beta = 1.585$$

Ниже изображен ее график.

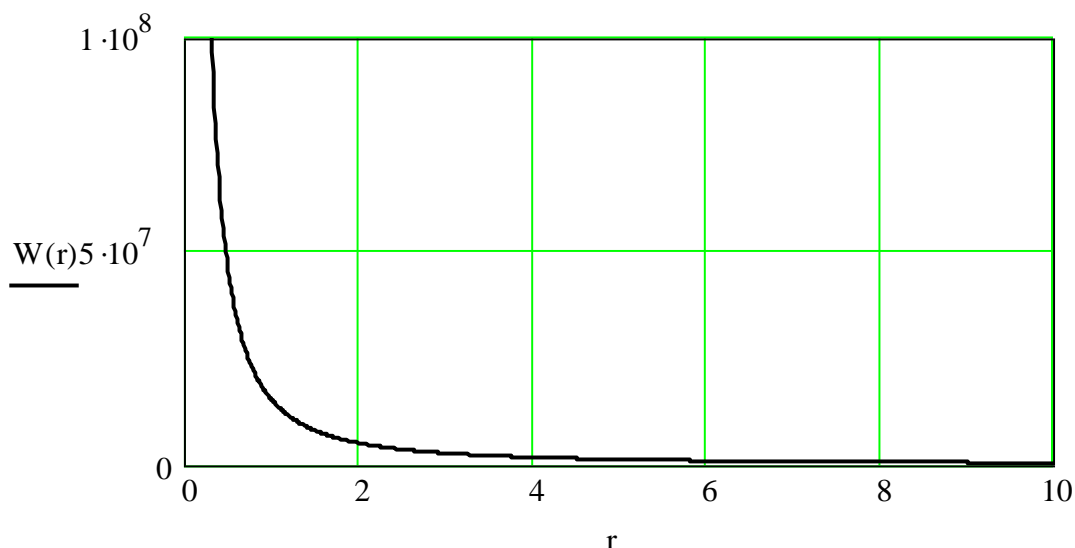


График результирующей аппроксимационной кривой: абсцисса – ранг объекта; ордината – электропотребление, кВт·ч

[на содержание]

ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ ОБЪЕКТОВ ТЕХНОЦЕНОЗА

Одной из процедур рангового анализа является интервальное оценивание. Оно позволяет определить, какие из объектов техноценоза потребляют ресурс аномально. Применительно к электропотреблению, если точка на ранговом распределении входит в доверительный интервал, то в пределах гауссового разброса параметров можно судить, что данный объект потребляет электроэнергию нормально. Если точка находится ниже интервала, то это, как правило, свидетельствует о нарушении нормального технологического процесса электропотребления на данном объекте (частые отключения электроэнергии, неплатежи и т.п.). Если точка находится выше интервала, то на объекте имеет место аномально большое потребление электроэнергии. Именно на эти объекты в первую очередь должно нацеливаться углубленное энергетическое обследование (энергоаудит). Последовательная реализация методологии позволит каждый раз целенаправленно воздействовать на наиболее «слабые» объекты. При этом средства для

проведения энергетических обследований, будут расходоваться эффективно, а общее электропотребление будет постоянно снижаться.

Подготовка и линеаризация данных

Задаем начало отсчета и считываем исходные данные (r – вектор рангов объектов; Y – вектор электропотребления, кВт·ч; n – количество объектов, $Time$ – рассматриваемый временной интервал):

```
ORIGIN := 1
r := READPRN("c:\mathcad_dat\R.md")
M := READPRN("c:\mathcad_dat\Zipf.md")
Rang := READPRN("c:\mathcad_dat\Rang.md")
coef := READPRN("c:\mathcad_dat\Coef.md")
```

```
n := rows(M)
n = 69
i := 1..n
Time := 6
Y := M<Time>
```

Линеаризуем и логарифмируем вектора эмпирических данных Y и рангов r . Создаем вектор единиц « p » и объединяем его с вектором r :

```
r1 := ln(r)
Y1 := ln(Y)
pi := 1
X := augment(p, r1)
```

Считываем коэффициенты регрессии, вычисленные методом наименьших квадратов для Y , и получаем результирующее уравнение аппроксимационной кривой рангового параметрического распределения:

```
 $\beta := \text{coef}_{Time, 1}$ 
 $\beta = 1.506$ 
 $W1 := \text{coef}_{Time, 2}$ 
 $W1 = 1.247 \times 10^7$ 
 $y_m := \frac{W1}{r^\beta}$ 
```


Определение доверительного интервала распределения

Вычислим остаточное среднее квадратическое отклонение эмпирических значений рассматриваемой выборки и коэффициент Стьюдента:

$$S := \sqrt{\frac{1}{n-3} \cdot \sum_i (Y1_i - \ln(ym_i))^2}$$

$$S = 0.535$$

$$t := qt(0.95, n - 3)$$

$$t = 1.668$$

Рассчитаем ковариационную матрицу для линеаризованной модели:

$$C := (X^T \cdot X)^{-1}$$

$$C = \begin{pmatrix} 0.206 & -0.058 \\ -0.058 & 0.018 \end{pmatrix}$$

$$D := X \cdot C$$

$$K := D \cdot X^T$$

Извлечем диагональные элементы из квадратной матрицы:

$$D(M) := \left| \begin{array}{l} n \leftarrow \text{rows}(M) \\ \text{for } i \in 1.. \text{rows}(M) \\ \quad \text{for } j \in 1.. \text{rows}(M) \\ \quad \quad W_i \leftarrow M_{i,j} \text{ if } i = j \\ \quad \\ \quad W \end{array} \right.$$

$$V := D(K)$$

$$V1 := V^T$$

Построим доверительные границы для расчетных данных (линеаризованный и нелинейный графики представлены на рисунках ниже):

$$V := \sqrt{V1^T}$$

$$d := t \cdot S \cdot V$$

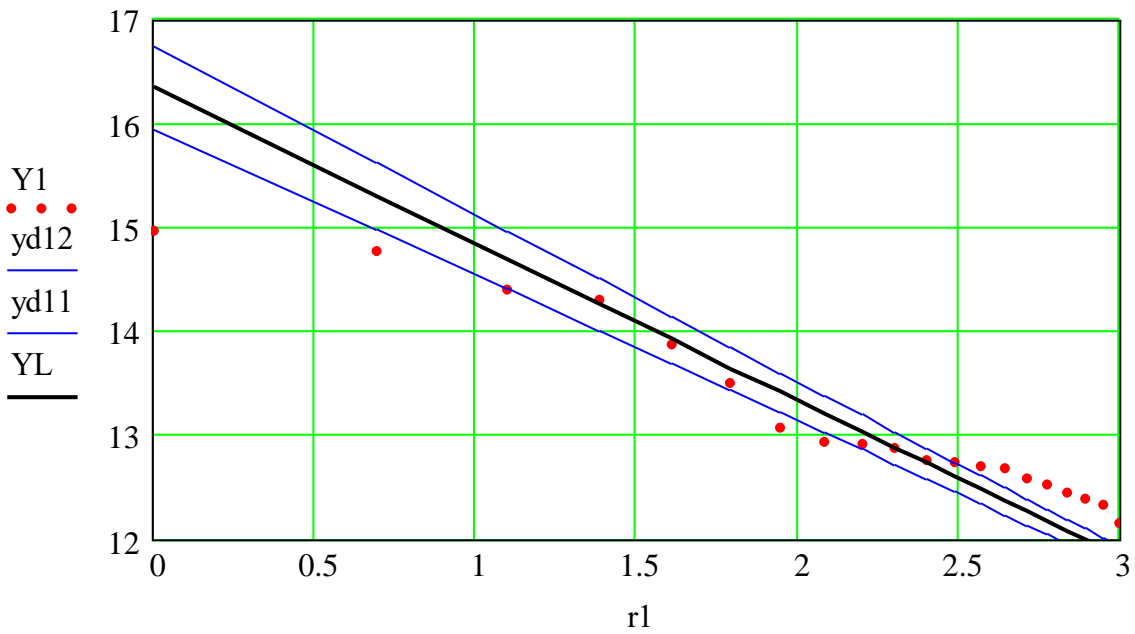
$$yd12 := YL + d$$

$$yd11 := YL - d$$

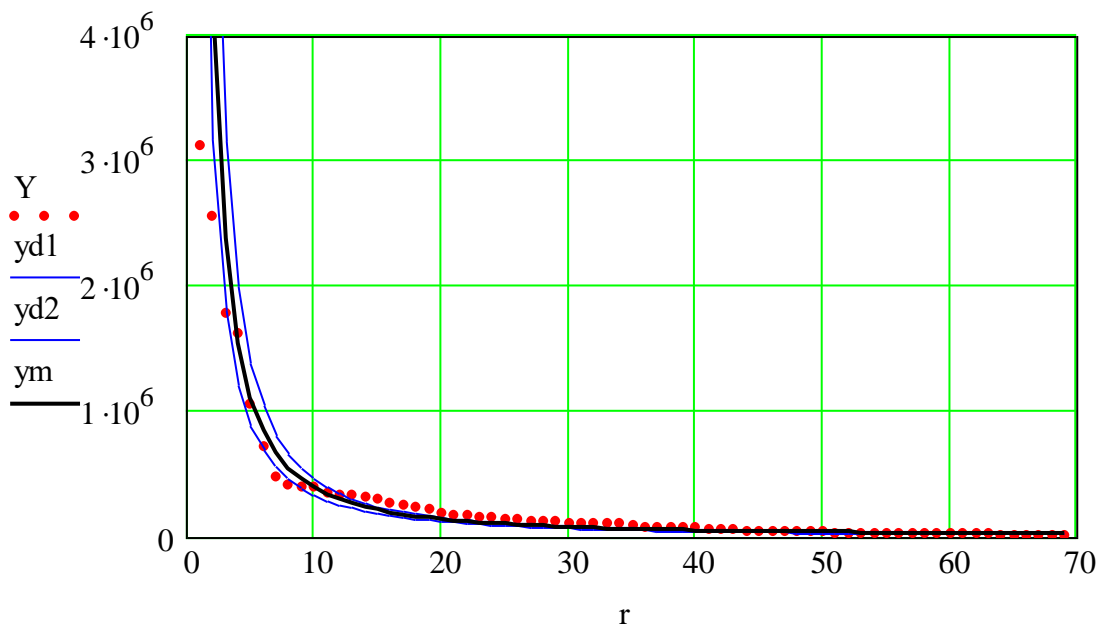
$$yd2 := \exp(yd12)$$

$$YL := \ln(ym)$$

$$yd1 := \exp(yd11)$$



Доверительный интервал для линейризованного рангового параметрического распределения техноценоза (фрагмент): абсцисса – логарифм ранга объекта; ордината – логарифм электропотребления (кВт·ч); точки – эмпирические данные; сплошные линии – аппроксимационная прямая, верхняя и нижняя доверительные границы



Доверительный интервал для полного рангового параметрического распределения в линейных осях: абсцисса – ранг объекта; ордината – электропотребление, кВт·ч; точки – эмпирические данные; сплошные линии – аппроксимационная кривая, верхняя и нижняя доверительные границы

Выявление объектов с аномальным электропотреблением

Определим количество точек, находящихся выше и ниже доверительного интервала (аномальных), а также попавших в него:

$$I_i := \text{if}(Y_i > yd2_i, 1, \text{if}(Y_i < yd1_i, -1, 0))$$

$$IN := \sum_i (I_i = 0)$$

- выше интервала:

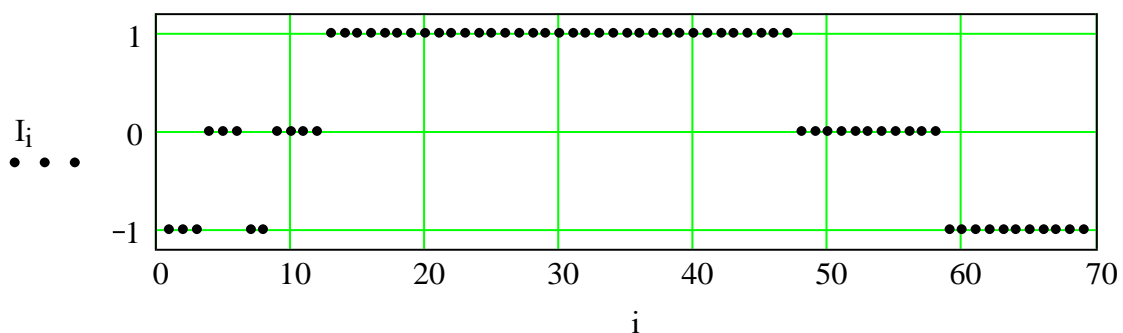
$$\sum_i I_i \cdot (I_i > 0) = 35$$

- ниже интервала:

$$\sum_i I_i \cdot (I_i < 0) = -16$$

- внутри интервала:

$$IN = 18$$



График, иллюстрирующий попадание точек в доверительный интервал: абсцисса – ранг объекта; ордината – индикатор, который принимает значение 0, 1 или -1, если точки соответственно лежат: внутри, выше или ниже доверительного интервала

Чтобы получить список очередности проведения энергоаудита, определим относительную величину отклонения точек от границ доверительного интервала, а затем проранжируем объекты по этому параметру:

$$z := Y - y_m$$

$$Q_i := \text{if}(Y_i > yd2_i, Y_i - yd2_i, \text{if}(Y_i < yd1_i, Y_i - yd1_i, 0))$$

$$OTN := \overrightarrow{(Q \cdot y_m^{-1})}$$

$$OTNM := \overrightarrow{|OTN|}$$

$$R(a,r) := \left| \begin{array}{l} c \leftarrow \text{sort}(a) \\ b \leftarrow \text{reverse}(c) \\ \text{for } m \in 1.. \text{rows}(a) \\ \quad \text{for } n \in 1.. \text{rows}(b) \\ \quad \quad L_n \leftarrow r_m \text{ if } a_m = b_n \\ L \end{array} \right.$$

```
SPR := R(OTNM,r)
ns := n - IN
ns = 51
SPRANG := submatrix(SPR, 1, ns, 1, 1)
```

$$SPRANG^T =$$

| | | | | | | | | | | |
|---|----|----|----|----|----|----|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 69 | 68 | 67 | 66 | 34 | 35 | 1 | 33 | 32 | 19 |

$$Ran := (Rang^T)^{\langle Time \rangle}$$

Используя приведенную ниже подпрограмму, определим номер объекта (столбца) в начальной базе данных техноценоза по электропотреблению (созданная ранее в программе MS Excel матрица V). В итоге получаем список очередности проведения энергоаудита на «аномальных» объектах (в полученном результирующем векторе указан их номер):

$$Unrang(a,b) := \left| \begin{array}{l} L \leftarrow 0 \\ i \leftarrow 1 \\ \text{for } m \in 1.. \text{rows}(a) \\ \quad \left| \begin{array}{l} \text{for } n \in 1.. \text{rows}(b) \\ \quad L_i \leftarrow n \text{ if } a_m = b_n \\ \quad i \leftarrow i + 1 \end{array} \right. \\ L \end{array} \right.$$

```
SPISOK := Unrang(SPRANG, Ran)
```

$$SPISOK^T =$$

| | | | | | | | | | | | | | | | | | | | | |
|---|----|----|----|----|----|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 1 | 13 | 48 | 67 | 58 | 26 | 34 | 9 | 64 | 28 | 53 | 24 | 36 | 27 | 6 | 66 | 8 | 29 | 50 | 61 | 23 |

[на содержание]

ПРОГНОЗИРОВАНИЕ ЭЛЕКТРОПОТРЕБЛЕНИЯ В ТЕХНОЦЕНОЗЕ

В целях корректного управления техноценозом часто возникает необходимость в прогнозе потребления ресурсов на следующий временной интервал. Хорошие результаты прогнозирования получают с использованием рангового анализа. Для примера в данной программе использованы статистические данные за шесть лет об электропотреблении одного из инфраструктурных объектов, расположенных на территории Калининградской области. Методика прогнозирования ресурсов техноценоза (в частности, электропотребления) основывается на известной теории структурно-топологической устойчивости ранговых параметрических распределений. При этом отдельно рассматриваются процедуры прогнозирования объектов, относящихся к ноевым, пойнтер- и саранчовым кастам.

Напомним, что по результатам первичной обработки статистической информации по электропотреблению объектов техноценоза сформирован ряд рабочих файлов с расширением «md», размещенных в директории «с:\mathcad_dat», которые будут использованы в данной программе.

Подготовка данных для прогнозирования

Считываем из файлов исходные данные и задаем начало отсчета:

```
V := 
C:\mathcad_dat\data.xls
```

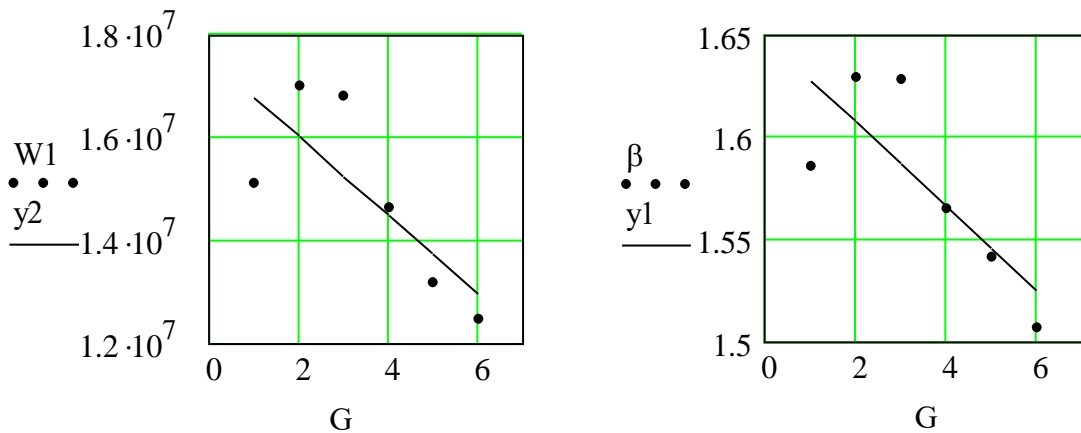
```
ORIGIN := 1
R := READPRN("c:\mathcad_dat\R.md")
Zipf := READPRN("c:\mathcad_dat\Zipf.md")
coef := READPRN("c:\mathcad_dat\Coef.md")
Rang := READPRN("c:\mathcad_dat\Rang.md")
```

Считываем значения коэффициентов регрессии за известные года предыстории для стандартной двухпараметрической гиперболической аппроксимационной формы, описанной в ряде наших работ:

$$\beta := (\text{coef}^{\langle 1 \rangle}) \qquad W1 := (\text{coef}^{\langle 2 \rangle})$$

Далее определяем прогнозные значения рангового коэффициента и первой точки рангового параметрического распределения (см. рис. ниже):

$$\begin{aligned}
 i &:= 1 \dots \text{rows}(\beta) & G_i &:= i & P_i &:= 1 \\
 X &:= \text{augment}(P, G) & B &:= (X^T \cdot X)^{-1} \cdot (X^T \cdot \beta) & B &= \begin{pmatrix} 1.648 \\ -0.021 \end{pmatrix} \\
 y1 &:= B_1 + G \cdot B_2 & \beta_{pr} &:= B_2 \cdot (\text{rows}(\beta) + 1) + B_1 & \beta_{pr} &= 1.503 \\
 A &:= (X^T \cdot X)^{-1} \cdot (X^T \cdot W1) \\
 A &= \begin{pmatrix} 1.753 \times 10^7 \\ -7.66 \times 10^5 \end{pmatrix} & y2 &:= A_1 + G \cdot A_2 \\
 W1pr &:= A_2 \cdot (\text{rows}(W1) + 1) + A_1 \\
 W1pr &= 1.217 \times 10^7
 \end{aligned}$$



Прогнозные кривые параметров рангового распределения: абсциссы – временной интервал (год); ординаты – значения параметров W1 (кВт·ч) и β; точки – эмпирические значения параметров; сплошные линии – аппроксимационные кривые, полученные методом наименьших квадратов

Определяем границы кастовых зон рангового параметрического распределения (ноевой, пойнтер- и саранчовой). При этом используется критерий равного распределения ресурсов между кастовыми зонами, являющийся следствием закона оптимального построения техноценозов. Вычисляем суммарный энергетический ресурс техноценоза:

$$\begin{aligned}
 Y(x) &:= \frac{W1pr}{x^{\beta_{pr}}} \\
 n &:= \text{length}(R) & n &= 69
 \end{aligned}$$

$$S := \frac{\int_1^n Y(x) dx}{3} \qquad S = 7.102 \times 10^6$$

Приведенная ниже подпрограмма позволяет вычислить нижнюю «a» и верхнюю «b» границы пойнтер-касты рангового распределения. При этом точность (время расчетов) задается шагом итерации «h».

```
GR(S) :=
  a ← 1
  b ← a
  h ← 0.001
  S1 ← 0
  S2 ← 0
  while S1 < S
    S1 ← ∫1a Y(x) dx
    a ← a + h
  while S2 < S
    S2 ← ∫a-hb Y(x) dx
    b ← b + h
  augment(a - h, b - h)
```

$$a := (GR(S)^{\langle 1 \rangle})_1$$

$$b := (GR(S)^{\langle 2 \rangle})_1$$

$$a = 1.996$$

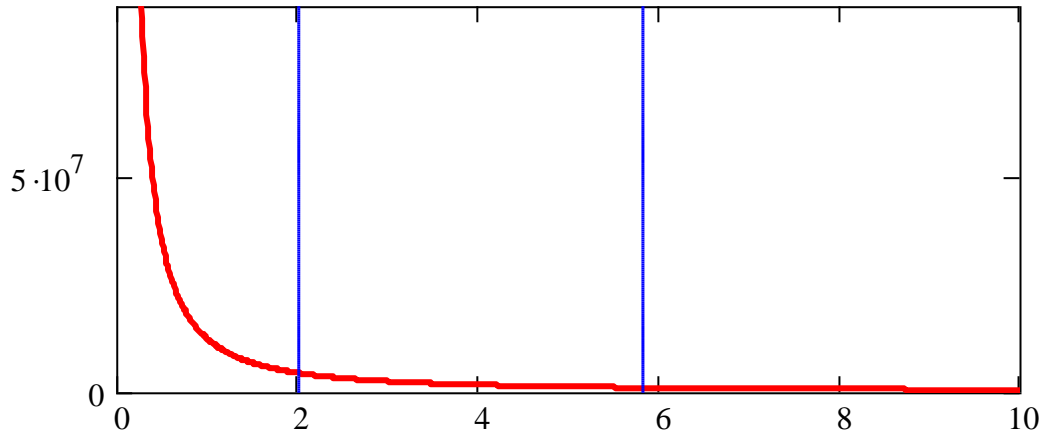
$$b = 5.811$$

$$\int_1^a Y(x) dx = 7.104 \times 10^6$$

$$\int_a^b Y(x) dx = 7.102 \times 10^6$$

$$\int_b^n Y(x) dx = 7.099 \times 10^6$$

Покажем кастовые зоны на распределении графически:



Деление рангового параметрического распределения техноценоза на кастовые зоны: абсцисса – ранг объекта; ордината – электропотребление, кВт·ч; вертикальные линии – кастовые границы (левая соответствует значению «а», правая – «b»)

Прогнозирование электропотребления объектами ноевой касты

В основе принятой методики прогнозирования лежит подбор многочлена, который наилучшим образом описывает тенденцию процесса электропотребления. Выявим ранги точек, попавших в ноеву касту:

$Ank := \text{submatrix}(R, 1, \text{floor}(a), 1, 1)$

$Ank = (1)$

Используя приведенную ниже подпрограмму, определяем по рангу номер объекта техноценоза (столбца) в начальной базе данных по электропотреблению (матрица V , созданная ранее в MS Excel):

```
Unrang(a,b) :=
  L ← 0
  i ← 1
  for m ∈ 1..rows(a)
    for n ∈ 1..rows(b)
      Li ← n if am = bn
    i ← i + 1
  L ← augment(a, L)
```

$Un := \text{Unrang} \left[Ank, \left(Rang^T \right)^{\langle \text{rows}(Rang) \rangle} \right]$

$Un = (1 \ 9)$

По номерам объектов из начальной базы данных формируем матрицу их электропотребления по годам (объекту соответствует столбец):

$$FD(a, b) := \left| \begin{array}{l} n \leftarrow \text{length}(b) \\ \text{for } i \in 1..n \\ \quad L^{(i)} \leftarrow a^{(b_i)} \\ L \end{array} \right. \quad NK := FD(v, U_n^{(2)})$$

Формируем матрицу базы данных за последний год.

$$NK1 := \text{submatrix}(NK, 1, \text{rows}(NK) - 1, 1, \text{cols}(NK))$$

Для прогнозирования временных рядов электропотребления объектами ноевой касты используется полином, который в зависимости от его степени «k» может быть прямой, квадратичной (и т.д.) параболой. Используя встроенные функции regress и submatrix получаем коэффициенты регрессии (вектор «с»). В целях упрощения задачи в качестве критерия оптимальности степени полинома используется минимум квадратического отклонения прогноза на последний известный год от проверочного значения за этот же год. Для осуществления такого прогноза применяется матрица NK1 (см. выше). Получаем матрицу квадратического отклонения для различных степеней полинома и рассматриваемых объектов:

$$EPR(Y, t, p) := \left| \begin{array}{l} n \leftarrow \text{rows}(Y) \\ q \leftarrow 1 \\ \text{for } j \in 1.. \text{cols}(Y) \\ \quad \left| \begin{array}{l} YY \leftarrow Y^{(j)} \\ \text{for } k \in 1..n-1 \\ \quad \left| \begin{array}{l} z \leftarrow \text{regress}(t, YY, k) \\ c \leftarrow \text{submatrix}(z, 4, \text{length}(z), 1, 1) \\ Y_{prj} \leftarrow c_1 + \sum_{i=2}^{k+1} c_i \cdot (\max(t) + 1)^{i-1} \\ E_{k,j} \leftarrow (Y_{prj} - p_j)^2 \end{array} \right. \end{array} \right. \\ E \end{array} \right.$$

$$n := \text{rows}(NK)$$

$$u := 1..n$$

```

tu := u
u1 := 1..n1
E := EPR[ NK1 , t1 , (NKT)<rows(NK)> ]
n1 := rows(NK) - 1
t1u1 := u1
    
```

Определяем оптимальные степени полинома для каждого из объектов ноевой касты рангового параметрического распределения, позволяющие осуществить прогноз с минимальной ошибкой:

```

PR(E) := | for j ∈ 1..cols(E)
          |   | mj ← min(E<j>)
          |   | E1 ← E<j>
          |   | for n ∈ 1..rows(E)
          |   |   | Lj ← n if mj = E1n
          |   |   | j ← j + 1
          |   | L
k := PR(E)
k = (1)
    
```

Используя полученный выше вектор степеней полинома k осуществляем прогноз электропотребления на следующий временной шаг:

```

REG(Y , t , k) := | for j ∈ 1..cols(Y)
                  |   | k1 ← kj
                  |   | Y1 ← Y<j>
                  |   | z ← regress(t , Y1 , k1)
                  |   | c ← submatrix(z , 4 , length(z) , 1 , 1)
                  |   | Yprj ← c1 + ∑i=2k1+1 ci · (max(t) + 1)i-1
                  |   | Ypr
    
```

```

Ynkpr := REG(NK , t , k)
Ynkpr = ( 3.22 × 106 )
    
```

Прогнозирование электропотребления объектами поинтер-касты

Метод прогноза в данном случае основан на экстраполяции временных рядов коэффициентов регрессии параметров гиперболической формы рангового параметрического распределения (двухпараметрической). Методом наименьших квадратов определяем коэффициенты:

$W := \text{submatrix}(\text{Zipf}, \text{ceil}(a), \text{floor}(b), 1, \text{cols}(\text{Zipf}))$

$T(W, r) :=$

$$\left. \begin{array}{l} n \leftarrow \text{length}(W) \\ \beta \leftarrow \frac{\sum_{i=1}^n \ln(W_i) \cdot \sum_{i=1}^n \ln(r_i) - n \cdot \sum_{i=1}^n \ln(W_i) \cdot \ln(r_i)}{\left(\sum_{i=1}^n \ln(r_i) \right)^2 - n \cdot \sum_{i=1}^n (\ln(r_i))^2} \\ W1 \leftarrow \exp \left[\frac{1}{n} \cdot \left(\sum_{i=1}^n \ln(W_i) + \beta \cdot \sum_{i=1}^n \ln(r_i) \right) \right] \\ WW \leftarrow \text{stack}(\beta, W1) \\ WW \end{array} \right|$$

$H(Z) :=$

$$\left. \begin{array}{l} j \leftarrow 1 \\ \text{while } j < \text{cols}(W) + 1 \\ \quad \left| X^{(j)} \leftarrow Z(W^{(j)}, R) \right. \\ \quad \left| j \leftarrow j + 1 \right. \\ X \end{array} \right|$$

$$\beta_{pk} := (H(T)^T)^{(1)}$$

$$n2 := \text{cols}(W) \quad n2 = 6$$

$$G_i := i \quad P_i := 1$$

$$B_{pk} := (X^T \cdot X)^{-1} \cdot (X^T \cdot \beta_{pk})$$

$$B_{pk} = \begin{pmatrix} 0.218 \\ 0.074 \end{pmatrix}$$

$$W1_{pk} := (H(T)^T)^{(2)}$$

$$i := 1 .. \text{cols}(W)$$

$$X := \text{augment}(P, G)$$

$$y1_{pk} := B_{pk1} + G \cdot B_{pk2}$$

Используя экстраполяцию, получаем прогнозные значения рангового коэффициента и первой точки распределения (см. рис. ниже):

$$\beta_{prpk} := W_{pk2} \cdot (n_2 + 1) + W_{pk1}$$

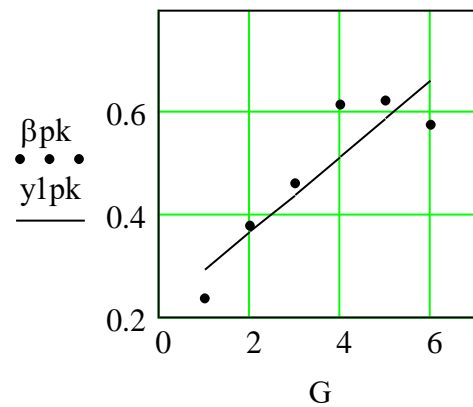
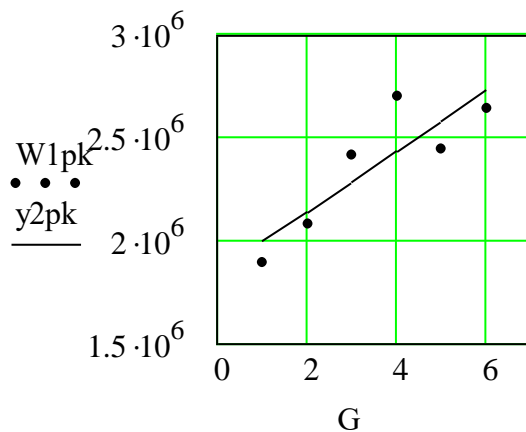
$$\beta_{prpk} = 0.738$$

$$A_{pk} := (X^T \cdot X)^{-1} \cdot (X^T \cdot W_{1pk})$$

$$y_{2pk} := A_{pk1} + G \cdot A_{pk2}$$

$$W_{1prpk} := A_{pk2} \cdot (n_2 + 1) + A_{pk1}$$

$$W_{1prpk} = 2.873 \times 10^6$$



Прогнозные кривые параметров пойнтер-касты: абсциссы – временной интервал (год); ординаты – параметры W1 (кВт·ч) и β; точки – эмпирические значения параметров; сплошные линии – аппроксимационные кривые, полученные методом наименьших квадратов

Используя полученные данные, определяем расчетные ранги и прогнозируемое электропотребление объектов техноценоза:

$$i := 1.. \text{rows}(W)$$

$$r_i := \exp \left[\frac{\ln(W_{1pk_{n2}}) - \ln \left[(W^{(n2)})_i \right]}{\beta_{pk_{n2}}} \right]$$

$$r^T = (1.055 \quad 1.998 \quad 2.323 \quad 4.902)$$

$$Y_{pkpr} := \frac{W_{1prpk}}{r \cdot \beta_{prpk}}$$

$$Y_{pkpr}^T = (2.762 \times 10^6 \quad 1.725 \times 10^6 \quad 1.543 \times 10^6 \quad 8.895 \times 10^5)$$

Прогнозирование электропотребления объектами саранчовой касты

Метод аналогичен тому, который был применен для прогнозирования электропотребления объектов пойнтер-касты (см. рис. ниже).

```

W := submatrix(Zipf, ceil(b), rows(Zipf), 1, b)
T(W, r) :=
  n ← length(W)
  β ← 
$$\frac{\sum_{i=1}^n \ln(W_i) \cdot \sum_{i=1}^n \ln(r_i) - n \cdot \sum_{i=1}^n \ln(W_i) \cdot \ln(r_i)}{\left(\sum_{i=1}^n \ln(r_i)\right)^2 - n \cdot \sum_{i=1}^n (\ln(r_i))^2}$$

  W1 ← 
$$\exp\left[\frac{1}{n} \cdot \left(\sum_{i=1}^n \ln(W_i) + \beta \cdot \sum_{i=1}^n \ln(r_i)\right)\right]$$

  WW ← stack(β, W1)
  WW
  
```

```

H(Z) :=
  j ← 1
  while j < cols(W) + 1
    X⟨j⟩ ← Z(W⟨j⟩, R)
    j ← j + 1
  X
  
```

```

βsk := (H(T)T)⟨1⟩
W1sk := (H(T)T)⟨2⟩
i := 1..cols(W)
Gi := i
Pi := 1
X := augment(P, G)
Bsk := (XT · X)-1 · (XT · βsk)
  
```

$$B_{sk} = \begin{pmatrix} 1.385 \\ -0.03 \end{pmatrix}$$

$$y1_{sk} := B_{sk1} + G \cdot B_{sk2}$$

$$\beta_{prsk} := B_{sk2} \cdot (n2 + 1) + B_{sk1}$$

$$\beta_{prsk} = 1.176$$

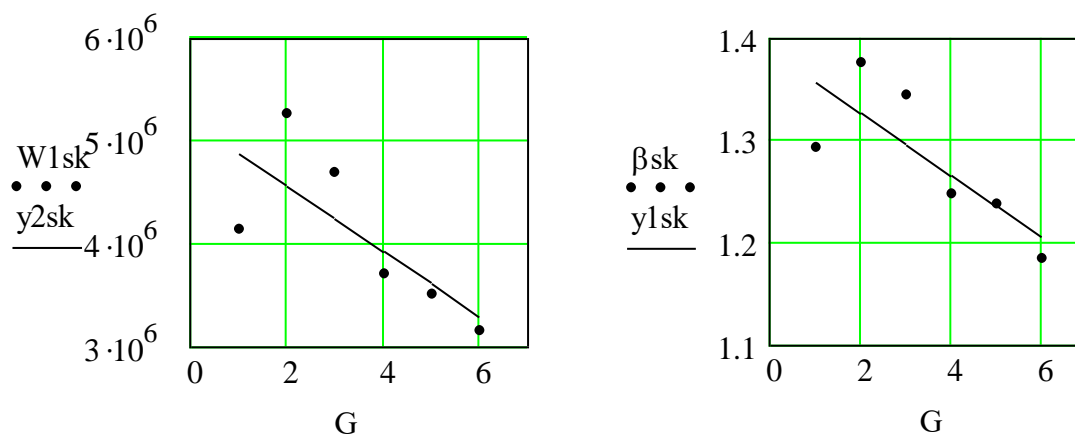
$$Ask := (X^T \cdot X)^{-1} \cdot (X^T \cdot W1_{sk})$$

$$Ask = \begin{pmatrix} 5.19 \times 10^6 \\ -3.189 \times 10^5 \end{pmatrix}$$

$$W1_{prsk} := Ask2 \cdot (n2 + 1) + Ask1$$

$$W1_{prsk} = 2.957 \times 10^6$$

$$y2_{sk} := Ask1 + G \cdot Ask2$$



Прогнозные кривые параметров саранчовой касты рангового распределения: абсциссы – временной интервал (год); ординаты – значения параметров; точки – эмпирические значения; сплошные линии – аппроксимационные кривые, полученные методом наименьших квадратов

Используя полученные данные, определяем расчетные ранги и прогнозируемое электропотребление объектов техноценоза:

$$i := 1.. \text{rows}(W)$$

$$rsk_i := \exp \left[\frac{\ln(W1_{sk_{n2}}) - \ln \left[\left(W^{(n2)} \right)_i \right]}{\beta_{sk_{n2}}} \right]$$

$$Y_{skpr} := \frac{W1_{prsk}}{rsk \cdot \beta_{prsk}}$$

$$Y_{skpr}^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 1 | $6.822 \cdot 10^5$ | $4.454 \cdot 10^5$ | $3.909 \cdot 10^5$ | $3.823 \cdot 10^5$ | $3.693 \cdot 10^5$ | $3.282 \cdot 10^5$ | $3.225 \cdot 10^5$ |

Прогноз электропотребления техноценозом в целом

Используя полученные результаты, осуществляем вывод итогового прогноза электропотребления на следующий временной интервал (год):

- по каждому объекту (кВт·ч):

$$PR := \text{stack}(Y_{nkpr}, Y_{pkpr}, Y_{skpr})$$

$$PR^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|-------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 1 | $3.22 \cdot 10^6$ | $2.762 \cdot 10^6$ | $1.725 \cdot 10^6$ | $1.543 \cdot 10^6$ | $8.895 \cdot 10^5$ | $6.822 \cdot 10^5$ | $4.454 \cdot 10^5$ |

- в целом по инфраструктуре (кВт·ч):

$$SPR := \sum_{i=1}^{\text{length}(PR)} PR_i$$

$$SPR = 1.811 \times 10^7$$

- в графической форме:

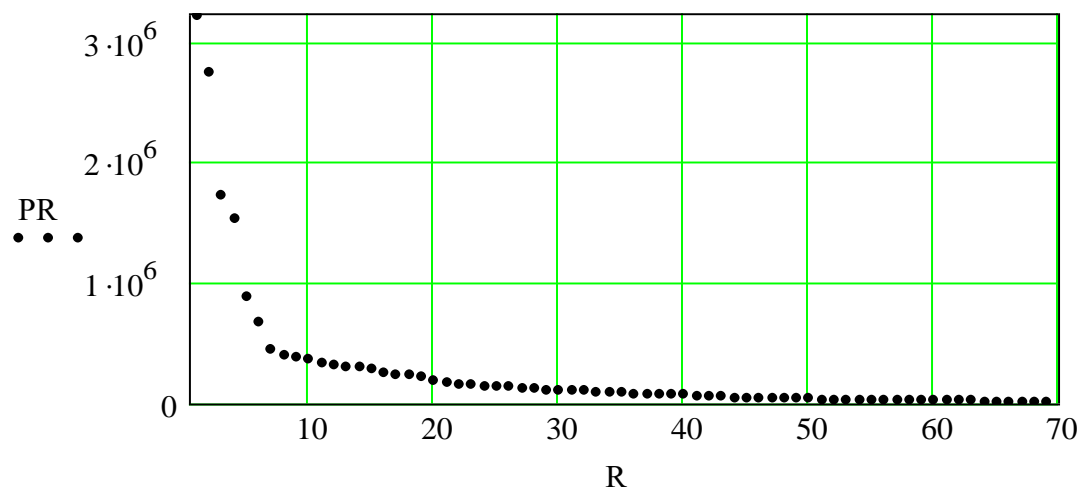


График прогноза электропотребления на следующий временной интервал (год): абсцисса – ранг объекта; ордината – электропотребление, кВт·ч

[на содержание]

НОРМИРОВАНИЕ ЭЛЕКТРОПОТРЕБЛЕНИЯ ОБЪЕКТАМИ ТЕХНОЦЕНОЗА

Задача нормирования потребления ресурсов в техноценозе может быть эффективно решена средствами рангового анализа. В настоящей программе в качестве нормируемого параметра рассматривается электропотребление объектами техноценоза. Ранее данный техноценоз был использован в качестве примера для программной реализации предыдущих этапов рангового анализа. Научно обоснованные нормы электропотребления являются эффективным инструментом энергосбережения. С целью нормирования объекты инфраструктуры разбиваются на группы со сходным электропотреблением, что реализуется посредством кластер-анализа (в частности – с помощью метода «ближнего соседа»).

Напомним, что по результатам первичной обработки статистической информации по электропотреблению исследуемого техноценоза сформирован ряд рабочих файлов с расширением «md», размещенных в директории «c:\mathcad_dat», которые будут использованы в данной программе.

Подготовка данных

Задаем начало отсчета и считываем исходные данные (R – вектор рангов объектов; W – матрица упорядоченных данных по электропотреблению за рассматриваемый период (6 лет), кВт·ч; n – количество объектов; $Rang$ – матрица рангов, соответствующих величине электропотребления каждого объекта за определенный временной интервал):

```
ORIGIN := 1
```

```
W := READPRN("c:\mathcad_dat\Zipf.md")
```

```
R := READPRN("c:\mathcad_dat\R.md")
```

```
Rang := READPRN("c:\mathcad_dat\Rang.md")
```

```
n := rows(W)
```

```
n = 69
```

Кластерный анализ объектов по электропотреблению

Для решения задачи был использован метод усредненного связывания для одномерных данных. В ходе реализации метода на каждом шаге

применен объединяющий алгоритм, основанный на оценке расстояний между точками, выявлении пары ближайших друг к другу точек и замены их средним. Процесс продолжается до тех пор, пока не останется единственное объединение (кластер). Создается дерево, которое представляет собой многоуровневую иерархию, в которой кластеры одного уровня образуют кластеры следующего более высокого уровня. Это позволяет принимать решение о наиболее приемлемом уровне кластеризации. Количество кластеров «С» задается априорно, что позволяет на определенном шаге прекращать работу подпрограммы. «С» обычно составляет пятую – шестую часть от количества объектов. Матрица «N» возвращает номера точек с минимальным расстоянием до следующей точки по годам:

C := 13

```

Dist(W, C) :=
  n ← rows(W)
  for t ∈ 1.. cols(W)
    B ← W<t>
    for j ∈ 1.. n - C
      A ← B<j>
      for k ∈ 1.. n - 1
        dk ←
          | dk ← ∞ if Ak+1 = 0
          | |(Ak - Ak+1)| otherwise
      for i ∈ 1.. n - 1
        if min(d) = di
          | Lj,t ← i
          | Ai ←  $\frac{A_{L_{j,t}} + A_{L_{j,t}+1}}{2}$ 
      Z1 ← submatrix(A, 1, Lj,t, 1, 1)
      Z2 ← submatrix(A, Lj,t + 2, n, 1, 1) if Lj,t + 1 ≠ n
      B<j+1> ←
        | stack(Z1, 0) if Lj,t + 1 = n
        | stack(Z1, Z2, 0) otherwise
    L
N := Dist(W, C)

```

$$N^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1 | 51 | 50 | 58 | 54 | 62 | 63 | 55 | 37 | 43 | 52 | 47 | 51 | 30 | 6 | 50 | 43 | 46 | 44 | 39 | 49 | 47 |
| 2 | 44 | 37 | 58 | 48 | 29 | 14 | 43 | 59 | 60 | 37 | 50 | 42 | 36 | 49 | 39 | 49 | 43 | 33 | 25 | 30 | 43 |
| 3 | 60 | 58 | 64 | 50 | 35 | 53 | 62 | 54 | 31 | 47 | 45 | 47 | 39 | 47 | 27 | 17 | 48 | 38 | 31 | 35 | 43 |
| 4 | 58 | 33 | 50 | 60 | 52 | 17 | 56 | 46 | 43 | 51 | 24 | 57 | 43 | 33 | 26 | 53 | 41 | 45 | 43 | 46 | 32 |
| 5 | 68 | 47 | 54 | 50 | 39 | 32 | 29 | 36 | 53 | 43 | 47 | 49 | 50 | 32 | 29 | 53 | 33 | 36 | 43 | 18 | 37 |
| 6 | 59 | 45 | 54 | 52 | 27 | 52 | 54 | 47 | 44 | 35 | 57 | 32 | 45 | 41 | 13 | 37 | 52 | 44 | 42 | 44 | 40 |

Далее определяем, к какому кластеру относится точка (по годам):

```

CL(B, A) :=
  for t ∈ 1.. cols(A)
  |
  | N ← A<sup>t</sup>
  | R ← B
  | for j ∈ 1.. rows(N)
  |   for i ∈ 1.. rows(R)
  |     Ri ←
  |     | Ri - 1 if Ri > Nj
  |     | Nj if Ri = Nj + 1
  |     | Ri otherwise
  |   L<sup>t</sup> ← R
  | L
Cluster := CL(R, N)
    
```

$$Cluster^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|
| 1 | 1 | 2 | 3 | 4 | 5 | 6 | 6 | 7 | 8 | 8 | 9 | 9 | 9 | 9 | 9 | 10 | 10 | 10 | 10 |
| 2 | 1 | 2 | 3 | 4 | 4 | 5 | 6 | 7 | 7 | 8 | 9 | 9 | 10 | 11 | 11 | 11 | 11 | 11 | 12 |
| 3 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 7 | 8 | 9 | 10 | 10 | 11 | 11 | 11 | 11 | 11 | 11 | 11 |
| 4 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 8 | 9 | 9 | 9 | 9 | 10 | 10 | 10 | 11 | 11 | 11 |
| 5 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 8 | 8 | 8 | 9 | 9 | 9 | 9 | 10 | 10 | 10 | 10 |
| 6 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 8 | 8 | 9 | 9 | 9 | 9 | 10 | 10 | 10 | 10 | 10 |

Нормирование объектов в кластерах

По результатам расчетов объекты инфраструктуры разбиваются на кластеры по электропотреблению. Кластеризация осуществляется отдельно для каждого временного интервала (года). После этого возникает возможность определения норм потребления ресурсов (электропотребления) для каждого объекта. Норма представляет собой среднее и эмпирический стандарт для каждой из групп на каждом временном интервале. Делается это следующим образом. Вычисляем среднее для кластеров:

```

SR(W, Cl, C) :=
  n ← rows(Cl)
  k ← 1
  for t ∈ 1.. cols(W)
    X ← W⟨t⟩
    Cluster ← Cl⟨t⟩
    for j ∈ 1.. C
      for i ∈ 1.. n
        Ai,j ←  $\begin{cases} X_i & \text{if Cluster}_i = j \\ 0 & \text{otherwise} \end{cases}$ 
        ki,j ←  $\begin{cases} 1 & \text{if } A_{i,j} \neq 0 \\ 0 & \text{otherwise} \end{cases}$ 
        mj ←  $\sum_{i=1}^n k_{i,j}$ 
        Sj ←  $\frac{\sum_{i=1}^n A_{i,j}}{m_j}$ 
      for j ∈ 1.. C
        for i ∈ 1.. n
          σj ←  $\begin{cases} 0 & \text{if } m_j = 1 \\ \sqrt{\frac{1}{m_j - 1} \cdot \sum_{i=1}^n (S_j - A_{i,j})^2 \cdot k_{i,j}} & \text{otherwise} \end{cases}$ 
        σ1⟨t⟩ ← σ
        S1⟨t⟩ ← S
        m1⟨t⟩ ← m
     $\begin{pmatrix} \sigma 1 \\ S 1 \\ m 1 \end{pmatrix}$ 

```

$$\sigma := SR(W, Cluster, C)_1$$

$$Sr := SR(W, Cluster, C)_2$$

$$OC := SR(W, Cluster, C)_3$$

Количество объектов в кластерах:

$$OC^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|----|----|----|----|
| 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 5 | 7 | 9 | 7 | 31 |
| 2 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 5 | 7 | 44 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 9 | 11 | 37 |
| 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 4 | 3 | 6 | 10 | 37 |
| 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 4 | 6 | 5 | 14 | 29 |
| 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 4 | 5 | 2 | 19 | 29 |

Среднее для каждого кластера:

$$Sr^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 1 | $3.5 \cdot 10^6$ | $1.81 \cdot 10^6$ | $1.712 \cdot 10^6$ | $1.553 \cdot 10^6$ | $1.268 \cdot 10^6$ | $7.444 \cdot 10^5$ | $5.817 \cdot 10^5$ |
| 2 | $3.055 \cdot 10^6$ | $2.2 \cdot 10^6$ | $1.447 \cdot 10^6$ | $1.336 \cdot 10^6$ | $1.015 \cdot 10^6$ | $8.567 \cdot 10^5$ | $6.702 \cdot 10^5$ |
| 3 | $3.01 \cdot 10^6$ | $2.512 \cdot 10^6$ | $1.614 \cdot 10^6$ | $1.474 \cdot 10^6$ | $1.325 \cdot 10^6$ | $8.373 \cdot 10^5$ | $7.013 \cdot 10^5$ |
| 4 | $3.645 \cdot 10^6$ | $2.84 \cdot 10^6$ | $1.614 \cdot 10^6$ | $1.351 \cdot 10^6$ | $1.231 \cdot 10^6$ | $7.921 \cdot 10^5$ | $5.673 \cdot 10^5$ |
| 5 | $3.295 \cdot 10^6$ | $2.708 \cdot 10^6$ | $1.304 \cdot 10^6$ | $1.211 \cdot 10^6$ | $1.162 \cdot 10^6$ | $8.534 \cdot 10^5$ | $4.366 \cdot 10^5$ |
| 6 | $3.125 \cdot 10^6$ | $2.559 \cdot 10^6$ | $1.775 \cdot 10^6$ | $1.628 \cdot 10^6$ | $1.061 \cdot 10^6$ | $7.188 \cdot 10^5$ | $4.678 \cdot 10^5$ |

Стандарт для каждого кластера:

$$\sigma^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|--------------------|---|---------|--------------------|
| 1 | 0 | 0 | 0 | 0 | 0 | 919.239 | 0 |
| 2 | 0 | 0 | 0 | $5.657 \cdot 10^3$ | 0 | 0 | $4.674 \cdot 10^4$ |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | $1.146 \cdot 10^4$ |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Визуализация результатов кластеризации (см. рис. ниже):

$$BC(BCL, Cluster) := \left| \begin{array}{l} n \leftarrow \text{length}(Cluster) \\ \text{for } i \in 1..n \\ \quad L_i \leftarrow BCL_{Cluster_i} \\ L \end{array} \right.$$

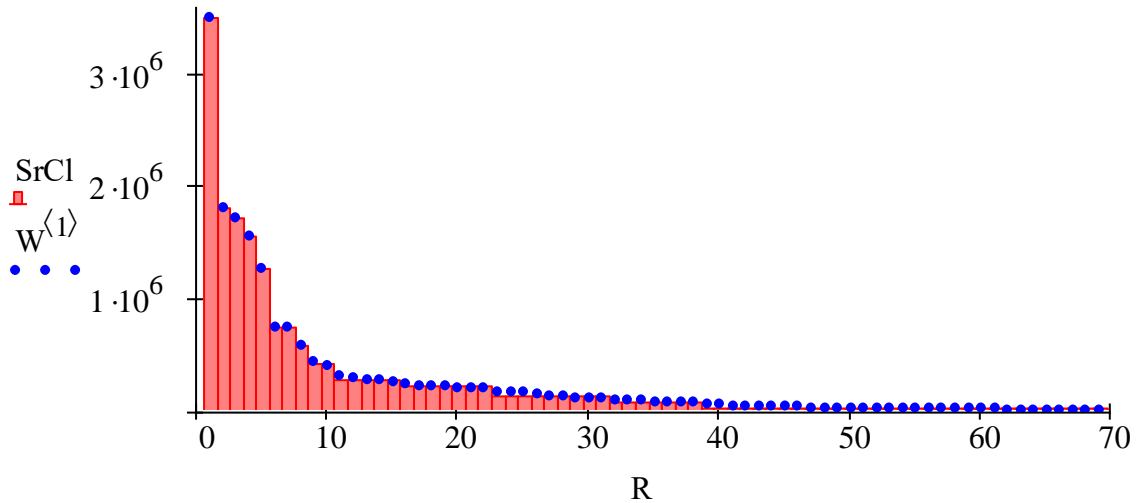
$$low := Sr - \sigma$$

$$h_i := S_r + \sigma$$

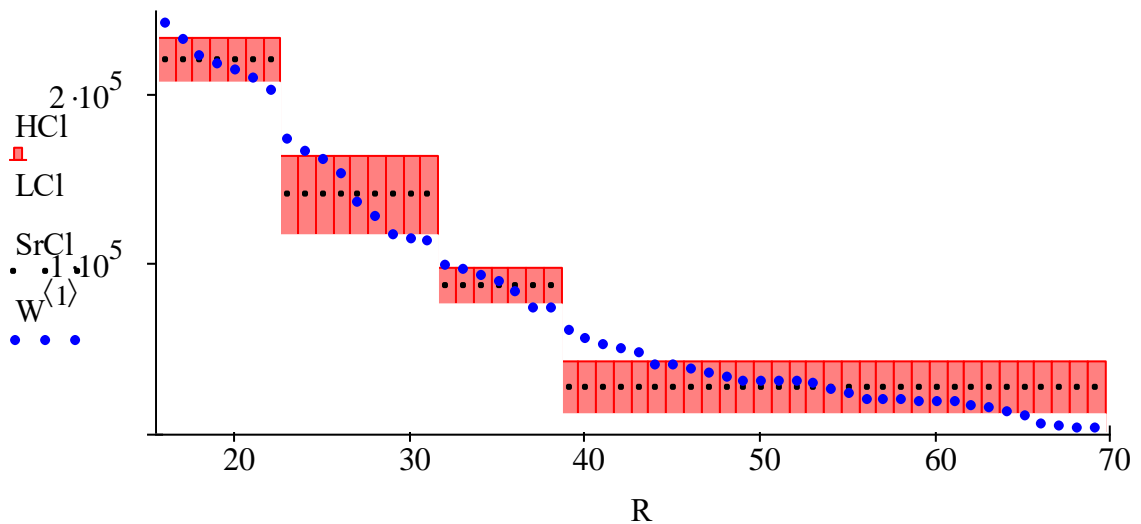
$$SrCl := BC(S_r^{(1)}, Cluster^{(1)})$$

$$LCl := BC(low^{(1)}, Cluster^{(1)})$$

$$HCl := BC(h_i^{(1)}, Cluster^{(1)})$$



Результаты кластерного анализа техноценоза по электропотреблению за один год (первый): абсцисса – ранг объекта; ордината – электропотребление, кВт·ч; точки – эмпирические данные; гистограмма – кластеры, полученные методом усредненного связывания



Статистическое описание норм кластеров (фрагмент): абсцисса – ранг объекта; ордината – электропотребление, кВт·ч; точки – эмпирические данные; горизонтальный пунктир – среднее нормы; горизонтальная красная сегментированная полоса – стандарт нормы

Определение прогнозируемых норм для объектов

Итоговая (прогнозируемая) норма электропотребления объекта техноценоза определяется экстраполяцией на следующий временной интервал норм, полученных для данного конкретного объекта на всех временных интервалах предыстории процесса электропотребления.

Для осуществления прогнозирования определяем принадлежность объектов к кластерам и соответствующие им нормы:

$$\text{ObC}(Cl, S, Rg, \sigma) := \left| \begin{array}{l} n \leftarrow \text{rows}(Rg) \\ \text{for } t \in 1.. \text{cols}(S) \\ \quad \left| \begin{array}{l} Sr \leftarrow S^{\langle t \rangle} \\ \sigma 2 \leftarrow \sigma^{\langle t \rangle} \\ Ran \leftarrow Rg^{\langle t \rangle} \\ Cla \leftarrow Cl^{\langle t \rangle} \\ \text{for } i \in 1.. n \\ \quad \left| \begin{array}{l} OS_i \leftarrow Sr(Cla_{Ran_i}) \\ O\sigma_i \leftarrow \sigma 2(Cla_{Ran_i}) \end{array} \right. \\ S1^{\langle t \rangle} \leftarrow OS \\ \sigma 1^{\langle t \rangle} \leftarrow O\sigma \end{array} \right. \\ \left(\begin{array}{l} S1 \\ \sigma 1 \end{array} \right) \end{array} \right. \\ \text{ObC}(Cluster, Sr, Rang^T, \sigma) = \left(\begin{array}{l} \{69,6\} \\ \{69,6\} \end{array} \right)$$

$$OS := \text{ObC}(Cluster, Sr, Rang^T, \sigma)_1^T$$

$$O\sigma := \text{ObC}(Cluster, Sr, Rang^T, \sigma)_2^T$$

Норма для объектов в исходной базе данных:

$$OS =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 1 | $1.81 \cdot 10^6$ | $1.268 \cdot 10^6$ | $2.206 \cdot 10^5$ | $5.817 \cdot 10^5$ | $4.203 \cdot 10^5$ | $7.444 \cdot 10^5$ | $2.699 \cdot 10^4$ |
| 2 | $1.447 \cdot 10^6$ | $1.336 \cdot 10^6$ | $8.567 \cdot 10^5$ | $5.244 \cdot 10^5$ | $3.962 \cdot 10^5$ | $6.702 \cdot 10^5$ | $4.623 \cdot 10^4$ |
| 3 | $1.474 \cdot 10^6$ | $1.614 \cdot 10^6$ | $8.373 \cdot 10^5$ | $5.784 \cdot 10^5$ | $5.179 \cdot 10^5$ | $7.013 \cdot 10^5$ | $3.476 \cdot 10^4$ |
| 4 | $1.351 \cdot 10^6$ | $1.614 \cdot 10^6$ | $3.553 \cdot 10^5$ | $4.176 \cdot 10^5$ | $5.673 \cdot 10^5$ | $3.553 \cdot 10^5$ | $3.912 \cdot 10^4$ |
| 5 | $1.162 \cdot 10^6$ | $1.304 \cdot 10^6$ | $3.888 \cdot 10^5$ | $2.975 \cdot 10^5$ | $4.366 \cdot 10^5$ | $9.7 \cdot 10^4$ | $2.673 \cdot 10^4$ |
| 6 | $1.628 \cdot 10^6$ | $1.775 \cdot 10^6$ | $3.995 \cdot 10^5$ | $3.317 \cdot 10^5$ | $3.317 \cdot 10^5$ | $1.111 \cdot 10^5$ | $2.936 \cdot 10^4$ |

Стандарт для объектов в исходной базе данных:

$$O\sigma =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 1 | 0 | 0 | $1.36 \cdot 10^4$ | 0 | $2.588 \cdot 10^4$ | 919.239 | $1.581 \cdot 10^4$ |
| 2 | 0 | $5.657 \cdot 10^3$ | 0 | 0 | $9.263 \cdot 10^3$ | $4.674 \cdot 10^4$ | $3.506 \cdot 10^4$ |
| 3 | 0 | 0 | 0 | 0 | 0 | $1.146 \cdot 10^4$ | $2.46 \cdot 10^4$ |
| 4 | 0 | 0 | $1.002 \cdot 10^4$ | $2.128 \cdot 10^4$ | 0 | $1.002 \cdot 10^4$ | $2.764 \cdot 10^4$ |
| 5 | 0 | 0 | $1.559 \cdot 10^4$ | $2.064 \cdot 10^4$ | 0 | $1.578 \cdot 10^4$ | $1.628 \cdot 10^4$ |
| 6 | 0 | 0 | $1.148 \cdot 10^4$ | $1.094 \cdot 10^4$ | $1.094 \cdot 10^4$ | $2.912 \cdot 10^4$ | $1.548 \cdot 10^4$ |

Рассчитаем нормы электропотребления на один шаг вперед (экстраполяция с использованием встроенной функции predict):

$$P(Z) := \begin{cases} j \leftarrow 1 \\ \text{while } j < \text{cols}(Z) + 1 \\ \quad \left| \begin{array}{l} p^{(j)} \leftarrow \text{predict}(Z^{(j)}, 1, 1) \\ j \leftarrow j + 1 \end{array} \right. \\ p^T \end{cases}$$

Норма для объектов на следующем временном шаге:

$$SS := P(OS)$$

$$SS^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|-------------------|
| 1 | $1.597 \cdot 10^6$ | $1.743 \cdot 10^6$ | $3.292 \cdot 10^5$ | $3.245 \cdot 10^5$ | $3.246 \cdot 10^5$ | $1.032 \cdot 10^5$ | $2.78 \cdot 10^4$ |

Стандарт для объектов на следующем временном шаге:

$$\sigma\sigma := P(O\sigma)$$

$$\sigma\sigma^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|--------------------|-------------------|--------------------|--------------------|--------------------|
| 1 | 0 | 0 | $7.665 \cdot 10^3$ | $7.75 \cdot 10^3$ | $5.457 \cdot 10^3$ | $1.236 \cdot 10^4$ | $1.394 \cdot 10^4$ |

Верхняя граница нормы:

$$H := SS + \sigma\sigma$$

$$H^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 1 | $1.597 \cdot 10^6$ | $1.743 \cdot 10^6$ | $3.369 \cdot 10^5$ | $3.322 \cdot 10^5$ | $3.301 \cdot 10^5$ | $1.155 \cdot 10^5$ | $4.173 \cdot 10^4$ |

Нижняя граница нормы:

$$L := SS - \sigma\sigma$$

$$L^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 1 | $1.597 \cdot 10^6$ | $1.743 \cdot 10^6$ | $3.216 \cdot 10^5$ | $3.167 \cdot 10^5$ | $3.192 \cdot 10^5$ | $9.082 \cdot 10^4$ | $1.386 \cdot 10^4$ |

Отранжируем результирующие нормы:

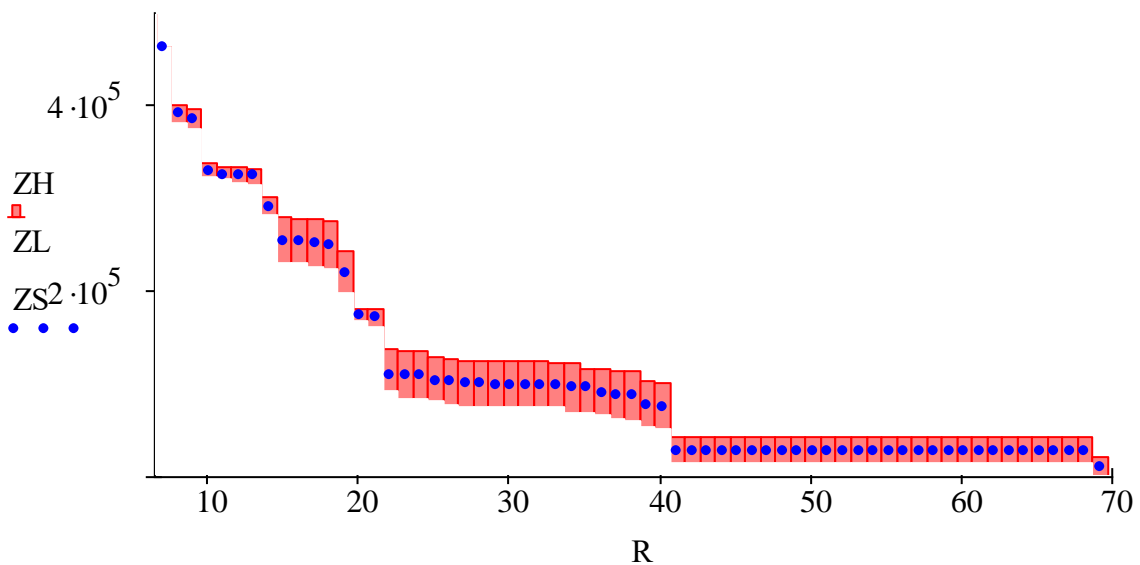
$$\text{Zipf}(Y) := \left| \begin{array}{l} \text{for } i \in 1.. \text{cols}(Y) \\ \left| \begin{array}{l} c \leftarrow \text{sort}(Y^{\langle i \rangle}) \\ b^{\langle i \rangle} \leftarrow \text{reverse}(c) \end{array} \right. \\ b \end{array} \right.$$

$$ZS := \text{Zipf}(SS)$$

$$ZH := \text{Zipf}(H)$$

$$ZL := \text{Zipf}(L)$$

Покажем нормы графически:



Отранжированные прогнозные нормы на следующий временной шаг: абсцисса – ранг объекта; ордината – электропотребление, кВт·ч; точки – нормы; горизонтальные красные полосы – стандарты норм

[на содержание]

ПРОВЕРКА ДАННЫХ НА СООТВЕТСТВИЕ КРИТЕРИЯМ Н-РАСПРЕДЕЛЕНИЯ

С точки зрения последующей статистической обработки данных, ключевое значение имеет проверка на соответствие критериям Н-распределения. Ее задача заключается в глубокой статистической обработке с целью проверки совместного выполнения двух гипотез. Во-первых, совокупность данных не подчиняется нормальному закону и, во-вторых, данные являются значимо взаимосвязанными. В случае если обе гипотезы выполняются, появляется возможность утверждать, что исследуемый объект является техноценозом, а данные по его электропотреблению могут обрабатываться методами рангового анализа. Проверка может осуществляться различными методами, каждый из которых обладает как достоинствами, так и недостатками. Учитывая высокую скорость расчетов в среде Mathcad, в данной программе проверка будет осуществляться сразу несколькими методами. После этого автоматически будет делаться вывод.

Процедуры проверки реализуются после первичной обработки статистической информации по техноценозу, которая должна быть осуществлена программными средствами Mathcad. Следует напомнить, что по результатам первичной обработки формируются рабочие файлы с расширением «*md*», размещенные в директории «*c:\mathcad_dat*», которая должна быть заблаговременно создана в корневом каталоге диска «*c:*».

Подготовка данных

Подготовка исходных данных по техноценозу включает процедуры считывания информации по электропотреблению объектов из внешнего файла, задания начала отсчета и определения вектора рангов.

V := 
C:\mathcad_dat\data.xls

ORIGIN := 1

Y := (V^T)⁽¹⁾

Y^T =

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | 1.81·10 ⁶ | 1.268·10 ⁶ | 2.328·10 ⁵ | 5.817·10 ⁵ | 4.386·10 ⁵ | 7.438·10 ⁵ | 2.891·10 ⁴ |

n1 := length(Y)

n1 = 69

j := 1..n1

$$\text{Rang} := \text{READPRN}("c:\text{mathcad_dat}\text{Rang.md}")$$

$$\text{Rang} =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|----|----|----|----|----|----|---|----|----|----|----|----|----|----|----|----|
| 1 | 2 | 5 | 17 | 8 | 9 | 7 | 53 | 15 | 1 | 4 | 11 | 21 | 69 | 3 | 28 | 19 | 20 | 23 |
| 2 | 3 | 5 | 7 | 10 | 11 | 9 | 53 | 18 | 1 | 4 | 14 | 16 | 69 | 2 | 43 | 6 | 13 | 20 |
| 3 | 4 | 3 | 6 | 9 | 10 | 7 | 52 | 15 | 1 | 5 | 13 | 18 | 69 | 2 | 42 | 23 | 16 | 20 |
| 4 | 4 | 3 | 12 | 9 | 7 | 10 | 52 | 15 | 1 | 5 | 22 | 18 | 69 | 2 | 49 | 27 | 16 | 14 |
| 5 | 5 | 3 | 9 | 12 | 7 | 34 | 53 | 19 | 1 | 4 | 15 | 18 | 68 | 2 | 40 | 30 | 17 | 14 |
| 6 | 4 | 3 | 8 | 11 | 12 | 40 | 54 | 18 | 1 | 5 | 16 | 15 | 69 | 2 | 24 | 52 | 21 | 25 |

$$\text{R} := \text{READPRN}("c:\text{mathcad_dat}\text{R.md}")$$

Проверка гипотезы о несоответствии генеральной совокупности данных по электропотреблению нормальному распределению при помощи критерия Пирсона

Статистический критерий Пирсона позволяет определить значимость гипотезы о принадлежности генеральной совокупности данных к области нормальных распределений. В случае положительного вывода мы будем вынуждены заключить, что исследуемый инфраструктурный объект не является техноценозом. И как следствие, что данные по его электропотреблению нельзя обрабатывать методами рангового анализа.

$$n := \text{ceil}(\sqrt{n1})$$

$$n = 9$$

$$i := 1..n$$

$$h := \frac{\max(Y) - \min(Y)}{n}$$

$$h = 3.886 \times 10^5$$

$$x_i := \max(Y) - \frac{h}{2} - (i - 1) \cdot h$$

$$\text{fr}(Y, x, h) := \left| \begin{array}{l} g \leftarrow 0 \\ \text{for } i \in 1.. \text{length}(x) \\ \quad \text{for } j \in 1.. \text{length}(Y) \\ \qquad g_{i,j} \leftarrow \begin{cases} 1 & \text{if } |Y_j - x_i| \leq \frac{h}{2} \\ 0 & \text{otherwise} \end{cases} \\ g \end{array} \right.$$

$$fr(Y, x, h) =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 6 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 |

$$g_i := \sum_j fr(Y, x, h)_{i,j}$$

$$g^T = (1 \ 0 \ 0 \ 0 \ 2 \ 2 \ 0 \ 5 \ 59)$$

$$\mu := \frac{\sum_{i=1}^n x_i \cdot g_i}{n1}$$

$$\sigma = 4.871 \times 10^5$$

$$\mu = 3.488 \times 10^5$$

$$u := \frac{x - \mu}{\sigma}$$

$$\sigma := \sqrt{\frac{\sum_{i=1}^n g_i \cdot (x_i - \mu)^2}{n1}}$$

$$\phi(u) := \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{u^2}{2}}$$

$$g1 := \frac{n1 \cdot h}{\sigma} \cdot \phi(u)$$

$$u = \begin{pmatrix} 6.069 \\ 5.272 \\ 4.474 \\ 3.676 \\ 2.879 \\ 2.081 \\ 1.283 \\ 0.486 \\ -0.312 \end{pmatrix}$$

$$g1 = \begin{pmatrix} 2.199 \times 10^{-7} \\ 2.027 \times 10^{-5} \\ 9.883 \times 10^{-4} \\ 0.026 \\ 0.348 \\ 2.519 \\ 9.639 \\ 19.517 \\ 20.914 \end{pmatrix}$$

$$\phi(u) = \begin{pmatrix} 3.995 \times 10^{-9} \\ 3.682 \times 10^{-7} \\ 1.796 \times 10^{-5} \\ 4.635 \times 10^{-4} \\ 6.331 \times 10^{-3} \\ 0.046 \\ 0.175 \\ 0.355 \\ 0.38 \end{pmatrix}$$

$$\chi^2 := \sum_{i=1}^n \frac{(g_i - g1_i)^2}{g1_i}$$

$$k := n - 3$$

$$\chi^2_{kp} = 12.592$$

$$\chi^2 = 4.547 \times 10^6$$

$$k = 6$$

$$Kr_1 := \chi^2 > \chi^2_{kp}$$

$$\alpha := 0.05$$

$$\chi^2_{kp} := qchisq(1 - \alpha, k)$$

$$Kr_1 = 1$$

Вывод: Теоретические и эмпирические частоты (χ^2 и $\chi^2_{кр}$) различаются значительно, следовательно, гипотеза о нормальном распределении генеральной совокупности отвергается (выборка негауссова).

Проверка гипотезы о несоответствии распределения нормальному закону методом спрямленных диаграмм

Метод спрямленных диаграмм также позволяет определить значимость гипотезы о принадлежности генеральной совокупности данных к области нормальных распределений (визуализация – см. рис. ниже). В случае положительного вывода мы будем вынуждены заключить, что исследуемый инфраструктурный объект не является техноценозом.

$f_j := 1$ $N_j := j - 1 + 0.5$

$N^T =$

| | | | | | | | | | | |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 0.5 | 1.5 | 2.5 | 3.5 | 4.5 | 5.5 | 6.5 | 7.5 | 8.5 | 9.5 |

 $F := \frac{N}{n1}$

$F^T =$

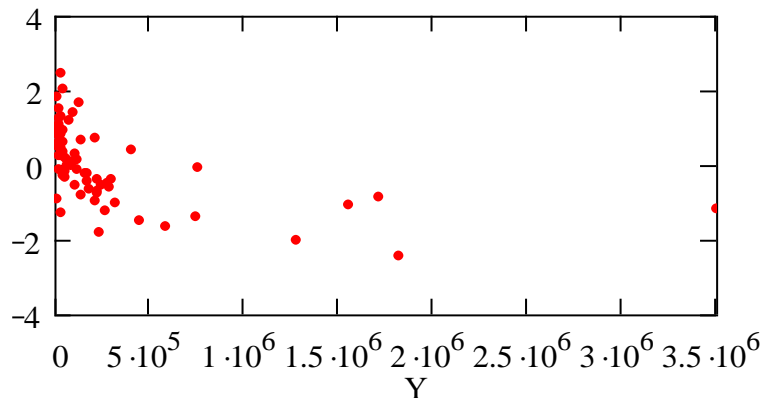
| | | | | | | |
|---|-----------------------|-------|-------|-------|-------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | $7.246 \cdot 10^{-3}$ | 0.022 | 0.036 | 0.051 | 0.065 | 0.08 |

$K(F) :=$

| | |
|--|-------------|
| for $j \in 1.. \text{length}(F)$ | |
| $k_j \leftarrow \text{qnorm}(F_j, 0, 1)$ | |
| k | $k := K(F)$ |

$k^T =$

| | | | | | | | | | |
|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | -2.445 | -2.019 | -1.796 | -1.638 | -1.512 | -1.407 | -1.315 | -1.233 | -1.159 |



Визуальное распределение квантилей электропотребления: абсцисса – электропотребление, кВт·ч; ордината – квантили

Вывод: Из графика видно, что точки не лежат вблизи какой-либо прямой, следовательно, гипотеза о нормальном распределении генеральной совокупности отвергается (выборка является негауссовой).

Исследование взаимосвязанности техноценоза с помощью коэффициента конкордации

Коэффициент конкордации, определенный для совокупности ранговых параметрических распределений по электропотреблению, характеризует степень взаимосвязанности техноценоза. Он показывает согласованность перемещения объектов по ранговой поверхности при переходе от предыдущего временного интервала к последующему.

$$i := 1.. \text{rows}(\text{Rang}) \quad j := 1.. \text{cols}(\text{Rang}) \quad S_j := \sum_i \text{Rang}_{i,j}$$

$$S^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|----|----|----|----|----|-----|-----|-----|---|----|
| 1 | 22 | 22 | 59 | 59 | 56 | 107 | 317 | 100 | 6 | 27 |

$$SS := \sum_{i=1}^n S_i$$

$$SS = 748$$

$$SR := \frac{SS}{\text{rows}(S)}$$

$$SR = 10.841$$

$$D := S - SR$$

$$DKV := D^2$$

$$D^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|--------|--------|--------|--------|--------|--------|---------|--------|--------|
| 1 | 11.159 | 11.159 | 48.159 | 48.159 | 45.159 | 96.159 | 306.159 | 89.159 | -4.841 |

$$DKV^T =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---------|---------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | 124.533 | 124.533 | 2.319·10 ³ | 2.319·10 ³ | 2.039·10 ³ | 9.247·10 ³ | 9.373·10 ⁴ |

$$SDKV := \sum_{j=1}^n DKV_j$$

$$SDKV = 1.179 \times 10^5$$

$$m := \text{rows}(\text{Rang})$$

$$m = 6$$

$$K := \frac{12 \cdot SDKV}{m^2 \cdot (n^3 - n)}$$

$$K = 54.574$$

$$\chi^2_r := \frac{SDKV}{\frac{1}{12} \cdot m \cdot n \cdot (n+1)}$$

$$\chi^2_r = 2.62 \times 10^3$$

$$v := n - 1$$

$$v = 8$$

$$\alpha := 0.05$$

$$\chi^2 := \text{qchisq}(1 - \alpha, v)$$

$$\chi^2 = 15.507$$

$$Kr_2 := \text{if}(\chi^2 \leq \chi^2_r, 1, 0) \quad Kr_2 = 1$$

Вывод: Для исследуемой совокупности данных коэффициент конкордации значим (выполняется условие $\chi^2 \leq \chi^2_r$), что свидетельствует о

взаимосвязанности техноценоза. Данный вывод позволяет использовать созданную базу данных для процедур рангового анализа техноценоза.

Исследование взаимосвязанности техноценоза с помощью выборочного коэффициента ранговой корреляции Кендалла

Выборочный коэффициент ранговой корреляции Кендалла, определенный для пары ранговых параметрических распределений, характеризует степень их взаимосвязанности. Он показывает согласованность перемещения объектов по ранговой поверхности во времени.

$$\text{Kendall}(A, R, t) := \left(\begin{array}{l} N \leftarrow \text{cols}(A) \\ \text{for } q \in 1.. \text{rows}(A) \\ \quad \left| \begin{array}{l} a \leftarrow (A^T)^{\langle q \rangle} \\ \text{for } o \in 1.. \text{rows}(A) \\ \quad \left| \begin{array}{l} b \leftarrow (A^T)^{\langle o \rangle} \\ \text{for } m \in 1.. N \\ \quad \text{for } n \in 1.. N \\ \quad \quad L_m \leftarrow b_n \text{ if } R_m = a_n \\ \text{for } i \in 1.. N \\ \quad \text{for } j \in 1.. N \\ \quad \quad p_{i,j} \leftarrow \begin{cases} 1 & \text{if } L_i < L_j \wedge i < j \\ 0 & \text{otherwise} \end{cases} \\ P_{q,o} \leftarrow \sum_{i=1}^N \sum_{j=1}^N p_{i,j} \\ \tau \leftarrow \frac{2 \cdot P}{\frac{N}{2} \cdot (N-1)} - 1 \\ D_{q,o} \leftarrow \begin{cases} 1 & \text{if } |\tau_{q,o}| > t \\ 0 & \text{otherwise} \end{cases} \end{array} \right. \\ \left. \begin{array}{l} \tau \\ D \end{array} \right)$$

$$\alpha := 0.05$$

$$z\alpha := \left| \text{qnorm}\left(\frac{\alpha}{2}, 0, 1\right) \right|$$

$$z\alpha = 1.96$$

$$T_{kp} := z\alpha \cdot \sqrt{\frac{2 \cdot (2 \cdot n1 + 5)}{9 \cdot n1 \cdot (n1 - 1)}}$$

$$T_{kp} = 0.161$$

$$\tau := \text{Kendall}(\text{Rang}, R, T_{kp})_1 \quad \tau = \begin{pmatrix} 1 & 0.907 & 0.921 & 0.888 & 0.851 & 0.81 \\ 0.907 & 1 & 0.93 & 0.896 & 0.865 & 0.802 \\ 0.921 & 0.93 & 1 & 0.925 & 0.886 & 0.831 \\ 0.888 & 0.896 & 0.925 & 1 & 0.903 & 0.84 \\ 0.851 & 0.865 & 0.886 & 0.903 & 1 & 0.889 \\ 0.81 & 0.802 & 0.831 & 0.84 & 0.889 & 1 \end{pmatrix}$$

$$KK := \text{Kendall}(\text{Rang}, R, T_{kp})_2 \quad KK = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$Kr_3 := \text{rows}(\text{Rang})^2 = \sum_{i=1}^{\text{rows}(\text{Rang})} \sum_{j=1}^{\text{rows}(\text{Rang})} KK_{i,j} \quad Kr_3 = 1$$

Вывод: Выборочные коэффициенты ранговой корреляции значимы (матрица «KK» заполнена единицами), следовательно, ранговые распределения взаимосвязаны, и их можно использовать в процессе дальнейшей статистической обработки данных по электропотреблению.

Исследование взаимосвязанности техноценоза с помощью выборочного коэффициента ранговой корреляции Спирмена

Выборочный коэффициент ранговой корреляции Спирмена, определенный для пары ранговых параметрических распределений, характеризует степень их взаимосвязанности. Он также показывает согласованность перемещения объектов по ранговой поверхности во времени.

$$\alpha := 0.05$$

$$k := n1 - 2$$

$$k = 67$$

$$t := qt\left(1 - \frac{\alpha}{2}, k\right)$$

$$t = 1.996$$

$$Sp(R, n, t) := \left(\begin{array}{l} \text{for } j \in 1.. \text{rows}(R) \\ \quad \text{for } i \in 1.. \text{rows}(R) \\ \quad \quad \left[\begin{array}{l} 6 \cdot \sum_{q=1}^n \left[\left[(R^T)^{\langle i \rangle} \right]_q - \left[(R^T)^{\langle j \rangle} \right]_q \right]^2 \\ \rho_{i,j} \leftarrow 1 - \frac{\quad}{n^3 - n} \\ T_{i,j} \leftarrow t \cdot \sqrt{\frac{1 - (\rho_{i,j})^2}{n - 2}} \\ Z_{i,j} \leftarrow \begin{cases} 1 & \text{if } |\rho_{i,j}| > T_{i,j} \\ 0 & \text{otherwise} \end{cases} \end{array} \right. \\ \left(\begin{array}{l} \rho \\ T \\ Z \end{array} \right) \end{array} \right)$$

$$Sp(\text{Rang}, n1, t)_1 = \left(\begin{array}{cccccc} 1 & 0.978 & 0.984 & 0.974 & 0.953 & 0.924 \\ 0.978 & 1 & 0.987 & 0.978 & 0.956 & 0.907 \\ 0.984 & 0.987 & 1 & 0.989 & 0.968 & 0.936 \\ 0.974 & 0.978 & 0.989 & 1 & 0.973 & 0.939 \\ 0.953 & 0.956 & 0.968 & 0.973 & 1 & 0.972 \\ 0.924 & 0.907 & 0.936 & 0.939 & 0.972 & 1 \end{array} \right)$$

$$Sp(\text{Rang}, n1, t)_2 = \left(\begin{array}{cccccc} 0 & 0.051 & 0.044 & 0.055 & 0.074 & 0.093 \\ 0.051 & 0 & 0.039 & 0.051 & 0.072 & 0.103 \\ 0.044 & 0.039 & 0 & 0.035 & 0.061 & 0.086 \\ 0.055 & 0.051 & 0.035 & 0 & 0.056 & 0.084 \\ 0.074 & 0.072 & 0.061 & 0.056 & 0 & 0.057 \\ 0.093 & 0.103 & 0.086 & 0.084 & 0.057 & 0 \end{array} \right)$$

$$KS := \text{Sp}(\text{Rang}, n1, t)_3$$

$$Kr_4 := \text{rows}(\text{Rang})^2 = \sum_{i=1}^{\text{rows}(\text{Rang})} \sum_{j=1}^{\text{rows}(\text{Rang})} KS_{i,j} \quad Kr_4 = 1$$

Вывод: Выборочные коэффициенты ранговой корреляции значимы (матрица «KS» заполнена единицами), следовательно, ранговые распределения взаимосвязаны, и их можно использовать в процессе дальнейшей статистической обработки данных по электропотреблению.

Исследование взаимосвязанности техноценоза с помощью выборочного коэффициента линейной корреляции

Выборочный коэффициент линейной корреляции, определенный для пары распределений, характеризует степень их взаимосвязанности.

$$k := n1 - 2$$

$$k = 67$$

$$\alpha := 0.05$$

$$t := \text{qt}\left(1 - \frac{\alpha}{2}, k\right) \quad t = 1.996$$

$$\text{CORR}(V, t) := \begin{array}{l} n \leftarrow \text{cols}(V) \\ \text{for } i \in 1..n \\ \quad \text{for } j \in 1..n \\ \quad \quad C_{i,j} \leftarrow \text{corr}(V^{(i)}, V^{(j)}) \\ \quad \quad r_{i,j} \leftarrow \begin{cases} C_{i,j} & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \\ \quad \quad T_{i,j} \leftarrow r_{i,j} \cdot \frac{\sqrt{n-2}}{\sqrt{1 - (r_{i,j})^2}} \\ \quad \quad H_{i,j} \leftarrow \begin{cases} 1 & \text{if } T_{i,j} > t \\ -1 & \text{if } T_{i,j} < -t \\ 0 & \text{otherwise} \end{cases} \end{array} \\ \begin{pmatrix} C \\ H \\ T \end{pmatrix}$$

Выведем на экран корреляционные матрицы:

$$\text{CORR}(V, t)_1 =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|--------|--------|--------|--------|--------|--------|--------|
| 1 | 1 | 0.07 | -0.213 | 0.538 | -0.328 | 0.466 | 0.157 |
| 2 | 0.07 | 1 | 0.071 | -0.314 | 0.041 | -0.355 | 0.551 |
| 3 | -0.213 | 0.071 | 1 | 0.41 | 0.026 | 0.396 | -0.616 |
| 4 | 0.538 | -0.314 | 0.41 | 1 | 0.245 | 0.987 | -0.344 |
| 5 | -0.328 | 0.041 | 0.026 | 0.245 | 1 | 0.341 | 0.331 |
| 6 | 0.466 | -0.355 | 0.396 | 0.987 | 0.341 | 1 | -0.287 |

$$\text{CORR}(V, t)_3 =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|--------|--------|--------|--------|--------|--------|--------|
| 1 | 0 | 0.571 | -1.783 | 5.222 | -2.844 | 4.312 | 1.298 |
| 2 | 0.571 | 0 | 0.586 | -2.702 | 0.334 | -3.109 | 5.408 |
| 3 | -1.783 | 0.586 | 0 | 3.677 | 0.215 | 3.529 | -6.401 |
| 4 | 5.222 | -2.702 | 3.677 | 0 | 2.072 | 49.645 | -3 |
| 5 | -2.844 | 0.334 | 0.215 | 2.072 | 0 | 2.966 | 2.874 |
| 6 | 4.312 | -3.109 | 3.529 | 49.645 | 2.966 | 0 | -2.451 |

$$\text{CORR}(V, t)_2 =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|----|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1 | 0 | 0 | 0 | 1 | -1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | -1 | 1 | 0 |
| 2 | 0 | 0 | 0 | -1 | 0 | -1 | 1 | 0 | 0 | -1 | -1 | 1 | 0 | 1 | 0 | -1 |
| 3 | 0 | 0 | 0 | 1 | 0 | 1 | -1 | -1 | -1 | 0 | 0 | 0 | -1 | 0 | -1 | 1 |
| 4 | 1 | -1 | 1 | 0 | 1 | 1 | -1 | 0 | 0 | 1 | 1 | -1 | -1 | -1 | 0 | 1 |
| 5 | -1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | -1 | -1 | -1 | 1 | -1 | 0 |
| 6 | 1 | -1 | 1 | 1 | 1 | 0 | -1 | 0 | 0 | 1 | 0 | -1 | -1 | -1 | -1 | 1 |

Запишем в файл и считаем в тело программы данные:

```
WRITEPRN("c:\mathcad_dat\Korr.md") := CORR(V, t)2
H := READPRN("c:\mathcad_dat\Korr.md")
```

Зададим начальную индексацию и произведем основные расчеты, используя данные, хранящиеся в корреляционной матрице:

```
j := 1..n1
i := 1..n1
Sj := ∑i |Hi,j|
```

Выведем полученные значения на экран:

$$S^T =$$

| | | | | | | | | | | | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 1 | 37 | 52 | 49 | 60 | 49 | 63 | 49 | 34 | 31 | 59 | 47 | 57 | 54 | 50 | 53 | 50 | 54 | 37 | 56 | 53 |

$$SS := \sum_{i=1}^{n1} S_i$$

$$SS = 3.624 \times 10^3$$

$$R := \frac{SS}{n1^2 - n1} \cdot 100$$

Рассчитаем ключевое процентное соотношение:

$$R = 77.238$$

Выведем значения корреляционных критериев:

$$Kr_5 := R > 50$$

$$Kr_5 = 1$$

Вывод: Величины электропотребления объектов более чем в 77 % случаев статистически значимо коррелируют (выполняется критериальное условие $R \geq 50$), что указывает на существующую связь по параметру электропотребления между элементами исследуемого техноценоза.

Результаты расчетов

Гипотеза о принадлежности исследуемой генеральной совокупности данных по электропотреблению к статистике техноценологического типа верна в случае значимости всех вычисленных выше коэффициентов, что ниже иллюстрируется критериальным выражением вывода.

$$\text{if} \left(5 = \sum Kr, \text{"Tehnocenosis"}, \text{"No"} \right) = \text{"Tehnocenosis"}$$

Вывод: Все коэффициенты статистически значимы (все элементы вектора «Kr» равны единице). Очевидно, что в данном случае мы имеем дело с ярко выраженным техноценозом. Подобный вывод позволяет в дальнейшем при обработке статистических данных по электропотреблению использовать методологию рангового анализа. В частности, после аппроксимации ранговых параметрических распределений могут быть реализованы процедуры интервального оценивания, прогнозирования и нормирования электропотребления объектов техноценоза.

[на содержание]

ОЦЕНКА ПОТЕНЦИАЛА ЭНЕРГОСБЕРЕЖЕНИЯ ТЕХНОЦЕНОЗА

Оценка эффективности методологии оптимального управления электропотреблением, а также потенциала энергосбережения техноценоза осложняется необходимостью осуществления практической внедренческой работы, растягивающейся на несколько лет. А ответ, как правило, необходимо иметь еще до принятия решения о внедрении методологии. Для решения этой задачи используется программа, позволяющая моделировать процесс реализации методики на глубину 5 – 7 лет и более. Программа здесь реализована применительно к примеру, использовавшемуся в предыдущих программах. Для упрощения расчетов сделано допущение, что каждый объект в будущем будет потреблять электроэнергию по прогнозируемой стохастической норме с учетом гауссового разброса параметров внутри доверительного интервала соответствующего кластера техноценоза.

Подготовка данных

Задаем начало отсчета и считываем исходные данные (R – вектор рангов объектов; W – матрица упорядоченных реальных данных по электропотреблению за рассматриваемый период (6 лет), кВт·ч; n – количество объектов; $Rang$ – матрица рангов, соответствующих величине электропотребления каждого объекта за временной интервал).

```
ORIGIN := 1
```

```
W := READPRN("c:\mathcad_dat\Zipf.md")
```

```
R := READPRN("c:\mathcad_dat\R.md")
```

```
Rang := READPRN("c:\mathcad_dat\Rang.md")
```

```
n := rows(W)
```

```
n = 69
```

```
t := cols(W)
```

```
t = 6
```

Интерполяция норм электропотребления

Первоначально с использованием модифицированных программ прогнозирования, интервального оценивания и нормирования осуществляется кластерный анализ и интерполяция норм электропотребления объектов техноценоза на моделируемый единичный временной интервал.

C := 15

```

Dist(W,C,H) :=
  n ← rows(W)
  for t ∈ 1..cols(W)
    B ← W<sup>t</sup>
    U ← H<sup>t</sup>
    for j ∈ 1..n - C
      A ← B<sup>j</sup>
      Q ← U<sup>j</sup>
      for k ∈ 1..n - 1
        dk ←  $\begin{cases} d_k \leftarrow \infty & \text{if } A_{k+1} = 0 \\ |(A_k - A_{k+1})| & \text{otherwise} \end{cases}$ 
      for i ∈ 1..n - 1
        if min(d) = di
          
$$\begin{cases} L_{j,t} \leftarrow i \\ Q_i \leftarrow Q_{L_{j,t}} + Q_{L_{j,t}+1} \\ A_i \leftarrow \frac{A_{L_{j,t}} \cdot Q_i + A_{L_{j,t}+1}}{Q_i + 1} \end{cases}$$

          Z1 ← submatrix(A,1,Lj,t,1,1)
          Z2 ← submatrix(A,Lj,t+2,n,1,1) if Lj,t+1 ≠ n
          Z3 ← submatrix(Q,1,Lj,t,1,1)
          Z4 ← submatrix(Q,Lj,t+2,n,1,1) if Lj,t+1 ≠ n
          B<sup>j+1</sup> ←  $\begin{cases} \text{stack}(Z1,0) & \text{if } L_{j,t}+1 = n \\ \text{stack}(Z1,Z2,0) & \text{otherwise} \end{cases}$ 
          U<sup>j+1</sup> ←  $\begin{cases} \text{stack}(Z3,0) & \text{if } L_{j,t}+1 = n \\ \text{stack}(Z3,Z4,0) & \text{otherwise} \end{cases}$ 
  L

```

Определим исходную индексацию данных:

$q1 := 1..rows(W)$

$q2 := 1..cols(W)$

Сформируем расчетную матрицу:

$H_{q1, q2} := 1$

$N := Dist(W, C, H)$

Произведем расчеты с помощью подпрограммы:

```

CL(B, A) :=
  for t ∈ 1..cols(A)
    N ← A⟨t⟩
    R ← B
    for j ∈ 1..rows(N)
      for i ∈ 1..rows(R)
        Ri ←
          Ri - 1 if Ri > Nj
          Nj if Ri = Nj + 1
          Ri otherwise
    L⟨t⟩ ← R
  L
Cluster := CL(R, N)
    
```

В основу расчетов здесь положен алгоритм имитационного моделирования с динамическими обратными связями. В качестве стандартных процедур, отражающих процесс управления электропотреблением объектов техноценоза, используются модифицированные методики интервального оценивания, прогнозирования и нормирования.

Реализация конкретных значений электропотребления объектов техноценоза в процессе моделирования осуществляется с использованием соответствующих преобразующих функций, основанных на законах распределения нормальном и Вейбулла – Гнеденко. Возможен вариант моделирования с использованием только нормального распределения. Собственно оценка эффективности и потенциала энергосбережения техноценоза выполняется по критериальным выражениям и системе ограничений, основанным на законе оптимального построения техноценозов.

```

SR(W, Cl, C) :=
  n ← rows(Cl)
  k ← 1
  for t ∈ 1.. cols(W)
    X ← W<t>
    Cluster ← Cl<t>
    for j ∈ 1.. C
      for i ∈ 1.. n
        Ai,j ←  $\begin{cases} X_i & \text{if Cluster}_i = j \\ 0 & \text{otherwise} \end{cases}$ 
        ki,j ←  $\begin{cases} 1 & \text{if } A_{i,j} \neq 0 \\ 0 & \text{otherwise} \end{cases}$ 
        mj ←  $\sum_{i=1}^n k_{i,j}$ 
        Sj ←  $\frac{\sum_{i=1}^n A_{i,j}}{m_j}$ 
      for j ∈ 1.. C
        for i ∈ 1.. n
          σj ←  $\begin{cases} 0 & \text{if } m_j = 1 \\ \sqrt{\frac{1}{m_j - 1} \cdot \sum_{i=1}^n (S_j - A_{i,j})^2 \cdot k_{i,j}} & \text{otherwise} \end{cases}$ 
        σ1<t> ← σ
        S1<t> ← S
        m1<t> ← m
     $\begin{pmatrix} \sigma 1 \\ S 1 \\ m 1 \end{pmatrix}$ 

```

$$OC := SR(W, Cluster, C)_3$$

$$Sr := SR(W, Cluster, C)_2$$

$$\sigma_1 := SR(W, Cluster, C)_1$$

$$r := 1..rows(\sigma_1)$$

$$c := 1..cols(\sigma_1)$$

$$Ssr_r := \sum_c Sr_{r,c}$$

$$sr := \frac{Ssr}{cols(\sigma_1)}$$

$$\text{Sigma}(\sigma, Sr, sr) := \left| \begin{array}{l} \text{for } j \in 1..rows(\sigma) \\ \left| \begin{array}{l} a\sigma \leftarrow (\sigma^T)^{\langle j \rangle} \\ \text{for } i \in 1..cols(\sigma) \\ \sigma_{1j,i} \leftarrow \left| \begin{array}{l} a\sigma_i \text{ if } a\sigma_i \neq 0 \\ \sqrt{\frac{1}{rows(\sigma) - 1} \cdot \sum_{i=1}^{cols(\sigma)} (sr_j - Sr_{j,i})^2} \text{ otherwise} \end{array} \right. \end{array} \right. \end{array} \right|_{\sigma_1}$$

$$\sigma := \text{Sigma}(\sigma_1, Sr, sr)$$

$$\text{ObC}(Cl, S, Rg, \sigma) := \left| \begin{array}{l} n \leftarrow rows(Rg) \\ \text{for } t \in 1..cols(S) \\ \left| \begin{array}{l} Sr \leftarrow S^{\langle t \rangle} \\ \sigma_2 \leftarrow \sigma^{\langle t \rangle} \\ Ran \leftarrow Rg^{\langle t \rangle} \\ Cla \leftarrow Cl^{\langle t \rangle} \\ \text{for } i \in 1..n \\ \left| \begin{array}{l} OS_i \leftarrow Sr(Cla_{Ran_i}) \\ O\sigma_i \leftarrow \sigma_2(Cla_{Ran_i}) \end{array} \right. \\ S1^{\langle t \rangle} \leftarrow OS \\ \sigma_1^{\langle t \rangle} \leftarrow O\sigma \end{array} \right. \end{array} \right| \left(\begin{array}{l} S1 \\ \sigma_1 \end{array} \right)$$

$$OS := \text{ObC}(\text{Cluster}, \text{Sr}, \text{Rang}^T, \sigma)_1^T$$

$$O\sigma := \text{ObC}(\text{Cluster}, \text{Sr}, \text{Rang}^T, \sigma)_2^T$$

$$P(Z) := \left| \begin{array}{l} j \leftarrow 1 \\ \text{while } j < \text{cols}(Z) + 1 \\ \quad \left| \begin{array}{l} p^{\langle j \rangle} \leftarrow \text{predict}(Z^{\langle j \rangle}, 1, 1) \\ j \leftarrow j + 1 \end{array} \right. \\ p^T \end{array} \right.$$

$$i := 1..n$$

$$SS_i := \text{if} \left[P(OS)_i > 0, P(OS)_i, (OS^{\langle \text{rows}(OS) \rangle})_i \right]$$

$$\sigma\sigma := P(O\sigma)$$

$$H := SS + \sigma\sigma$$

$$L1 := SS - \sigma\sigma$$

$$L_i := \text{if}(L1_i > 0, L1_i, 1)$$

$$\text{Zipf}(Y) := \left| \begin{array}{l} \text{for } i \in 1.. \text{cols}(Y) \\ \quad \left| \begin{array}{l} c \leftarrow \text{sort}(Y^{\langle i \rangle}) \\ b^{\langle i \rangle} \leftarrow \text{reverse}(c) \end{array} \right. \\ b \end{array} \right.$$

$$ZS := \text{Zipf}(SS)$$

$$ZH := \text{Zipf}(H)$$

$$ZL := \text{Zipf}(L)$$

$$N := \text{augment}(L, SS, H)$$

$$z := N^{\langle 2 \rangle} - N^{\langle 1 \rangle}$$

Модельная реализация стохастических значений электропотребления

Ввиду недостаточной исследованности ряда динамических параметров, описывающих процесс электропотребления техноценоза, в данной программе приняты два допущения. Во-первых, при отсутствии в системе управления техноценоза стимулирующих воздействий, направленных на энергосбережение, в качестве математического ожидания и стандарта при-

нимаются соответствующие параметры нормы, вычисленной для кластера. В противном случае математическое ожидание уменьшается в k_1 , а стандарт – в k_2 раз. Во-вторых, все результаты электропотребления, превышающие норму, заменяются ее максимальным значением для данного объекта. В случае же если электропотребление объекта при моделировании окажется меньше нижней границы нормы, то в качестве электропотребления объекта принимается минимальное значение нормы.


$$PN(z, ZS) := \left| \begin{array}{l} \text{for } i \in 1.. \text{rows}(ZS) \\ S^{(i)} \leftarrow \text{norm}(1, ZS_i, z_i) \\ S \end{array} \right.$$


```
k1 := 0.8          k2 := 0.8
S := PN(z, SS)
S1 := PN(z·k2, SS·k1)
```

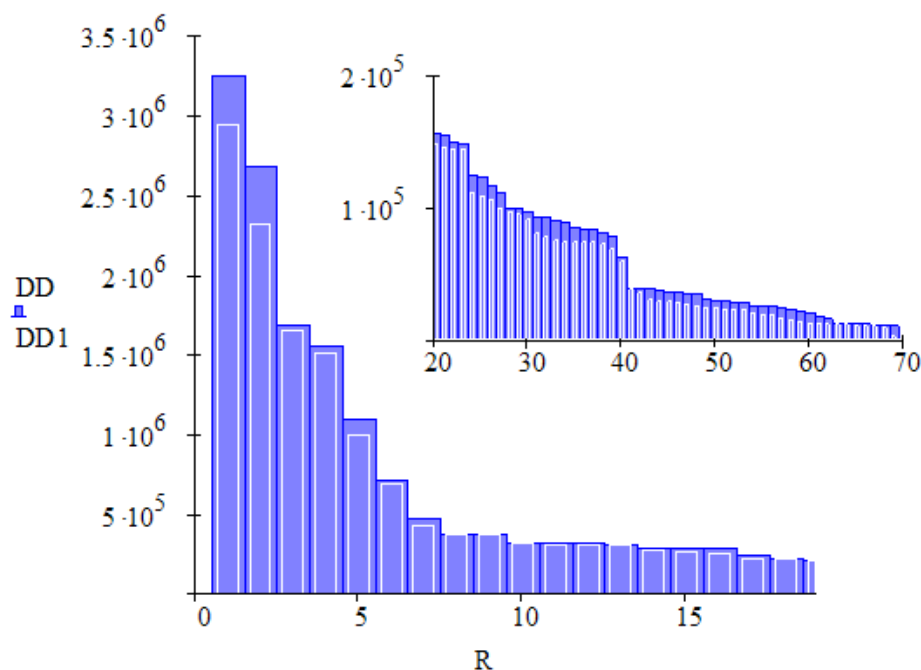
$$KU(S, m_i, m_a) := \left| \begin{array}{l} \text{for } i \in 1.. \text{rows}(S) \\ K_i \leftarrow \left| \begin{array}{l} m_i \text{ if } S_i < m_i \\ m_a \text{ if } S_i > m_a \\ S_i \text{ otherwise} \end{array} \right. \\ K \end{array} \right.$$

```
D := ceil(KU(S^T, N^{(1)}, N^{(3)}))          D1 := ceil(KU(S1^T, N^{(1)}, N^{(3)}))
```

Далее запишем полученные данные в файлы и визуализируем результаты расчетов. При этом в файл «C:\mathcad_dat\Dinam_1.xls» записываются моделируемые результаты электропотребления объектов техноценоза в будущем году при отсутствии мероприятий по энергосбережению, а в файл «C:\mathcad_dat\Dinam_2.xls» – моделируемые результаты электропотребления объектов техноценоза в будущем году в условиях проведения мероприятий по энергосбережению (см. рис. ниже).

```
 C:\mathcad_dat\Dinam_1.xls          DD := reverse(sort(D))
D^T
```

```
 C:\mathcad_dat\Dinam_2.xls          DD1 := reverse(sort(D1))
D1^T
```



Сравнение электропотребления техноценоза к исходу цикла моделирования для двух вариантов: темная синяя гистограмма – электропотребление объектов при отсутствии мероприятий по энергосбережению; светлая гистограмма из белых линий на фоне темной – электропотребление в условиях проведения мероприятий по энергосбережению

Программа позволяет моделировать процесс электропотребления на один год вперед. После этого все результаты, ранее записанные в файлы «Dinam_1.xls» и «Dinam_2.xls», должны в интерактивном режиме вне программы добавляться к исходной базе данных. При этом в отдельных файлах с одним и тем же именем «data.xls», но размещенных в разных папках, параллельно формируются две базы данных: одна с электропотреблением объектов при отсутствии мероприятий по энергосбережению, а вторая – с электропотреблением в условиях проведения мероприятий по энергосбережению. Далее отдельно для каждой из баз данных должны быть реализованы все расчетные программы, включая настоящую. После получения новых результатов моделирования (на второй год) и добавления к базам данных расчеты повторяются в циклическом режиме. Количество прогонов модели соответствует требуемой глубине прогноза. Очевидно, что в процессе моделирования имеет смысл содержать обе базы данных отдельно от рабочих файлов, размещенных в директории «C:\mathcad_dat». После каждого прогона модели и интерактивной модификации баз данных эти файлы, замещая предыдущие, должны экспортироваться в данную директорию как исходные для последующих расчетов. Естественно, при этом должен быть сохранен исходный файл, содержащий данные по реальному электропотреблению объектов техноценоза за предыдущие годы.

[на содержание]

G-МЕТОД ПРОГНОЗИРОВАНИЯ ЭЛЕКТРОПОТРЕБЛЕНИЯ В ТЕХНОЦЕНОЗЕ

Для корректного управления техноценозом часто возникает необходимость в прогнозе электропотребления на следующий временной интервал. Хорошие результаты прогнозирования получают с использованием GZ-методологии. G-методы основываются на классической гауссовой математической статистике и теории временных рядов. Это позволяет в большей степени учесть индивидуальные свойства функционирующих объектов. Z-методы прогнозирования основываются на ципфовой математической статистике и теории структурно-топологической устойчивости ранговых параметрических распределений. В данном случае, в процессе функционирования элемента техноценоза, учитывается в большей степени проявление системных свойств. Таким образом, реализация синтетической GZ-методологии позволяет органично учесть, как индивидуальные, так и системные свойства объектов техноценоза. Z-методы прогнозирования ранее неоднократно публиковались. Здесь рассмотрим порядок реализации в пакете Mathcad G-метода прогнозирования электропотребления.

После сбора статистической информации по электропотреблению техноценоза следующим этапом является анализ собранной информации и прогнозирование на очередной временной шаг. Собранная статистическая информация заносится в файл Microsoft Excel, который должен быть назван «EP.xls» и помещен в директорию «с:\mathcad_dat», заранее созданную в корневом каталоге диска «с:\». Полученный файл представляет собой таблицу данных по электропотреблению рассматриваемого техноценоза, каждый столбец которой соответствует определенному объекту, а строка – значению электропотребления, зафиксированному в очередной месяц текущего года. В данном примере рассматривается помесечное (за 8 лет предыстории) электропотребление реального техноценоза, расположенного на территории Калининградской области и состоящего из 69 объектов. Программа реализуется отдельно для каждого объекта.

Подготовка данных

Осуществляем импорт исходных данных из файла «EP.xls» в тело программы, задаем начало отсчета, извлекаем столбец из матрицы (в данном конкретном случае рассматривается объект техноценоза № 1), осуществляем логарифмирование данных, выделяем тренд методом наименьших квадратов и производим его удаление (см. рис. ниже).

M := 
C:\mathcad_dat\EP.xls

ORIGIN := 1

i := 1..rows(M)

x1 := M⁽¹⁾

x2 := $\overrightarrow{\ln(x1)}$

t_i := i

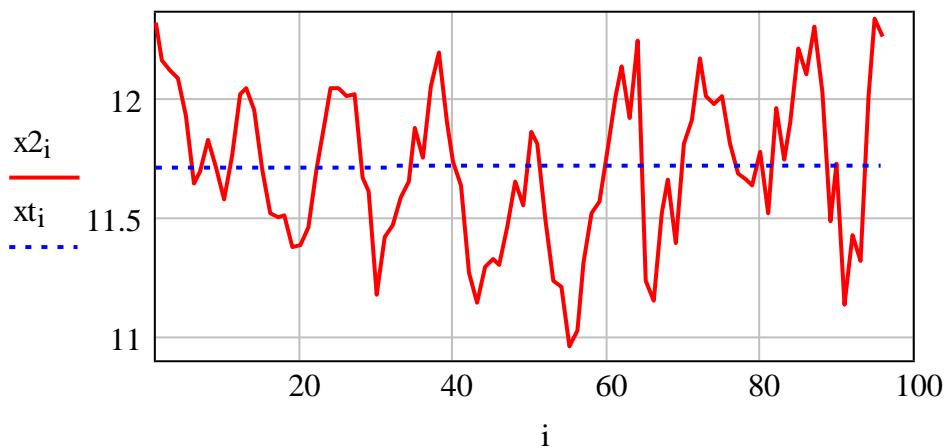
P_i := 1

X := augment(P, t)

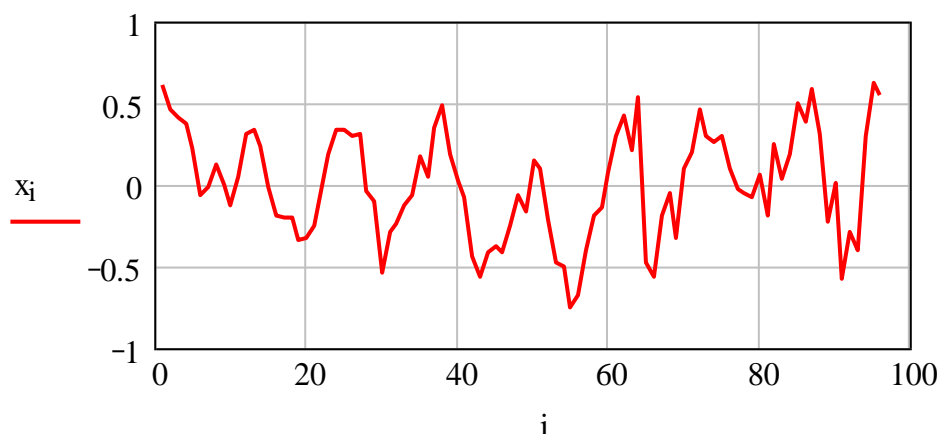
A := $(X^T \cdot X)^{-1} \cdot (X^T \cdot x2)$

xt := A₁ + A₂·t

x := x2 - xt



Анализ данных по электропотреблению объекта после их преобразования: абсцисса – номер месяца; ордината – электропотребление в относительных единицах по логарифмической шкале; ломанная кривая – электропотребление за известный период времени; штриховая прямая – тренд



Электропотребление объекта после удаления тренда: абсцисса – номер месяца; ордината – электропотребление объекта техноценоза в относительных единицах по логарифмической шкале

Анализ сезонности данных

Группируем данные (см. рис. ниже) и выделяем зависимость электропотребления объекта от времени года (сезонности).

$$j := 1..12$$

$$y1_j := x_j$$

$$y2_j := x_{(12+j)}$$

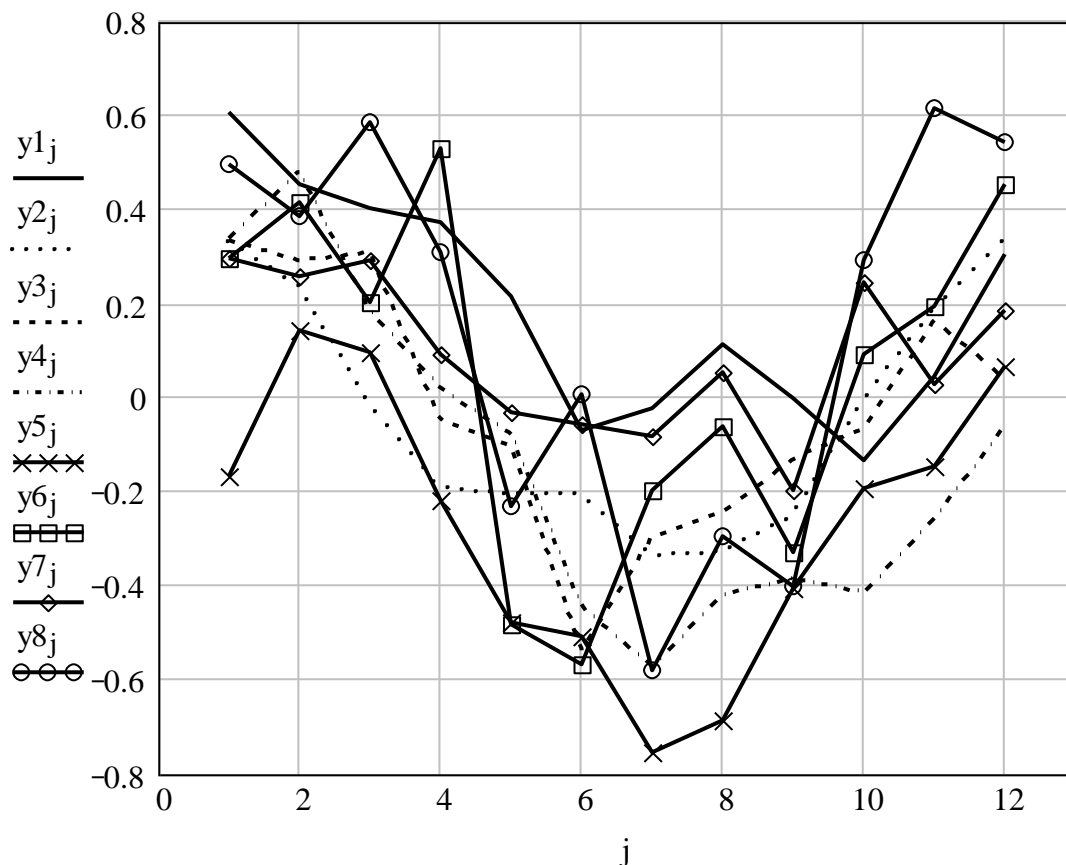
$$y3_j := x_{(24+j)}$$

$$y4_j := x_{(36+j)}$$

$$y5_j := x_{(48+j)}$$

$$y6_j := x_{(60+j)}$$

$$y7_j := x_{(72+j)}$$

$$y8_j := x_{(84+j)}$$


Изменение месячного электропотребления объекта в различные годы исследуемого периода: абсцисса – номер месяца года; ордината – электропотребление в относительных единицах; кривые $y1_j$ – $y8_j$ – электропотребление соответственно с 1995 по 2002 годы

Из характера изменений зависимостей на рисунке можно сделать вывод, что в процессе электропотребления объекта явно прослеживается сезонность: зимой электропотребление выше, летом – ниже.

Определение статистических характеристик электропотребления

Для дальнейшей обработки данных необходимо выделить ежемесячные медианные и минимальные уровни электропотребления. С этой целью задаем матрицу сгруппированных по годам значений электропотребления и вычисляем медианный, средний и минимальный уровни. Сглаживаем данные по методу Гаусса в зависимости от ширины гауссовского окна и времени. Визуализируем полученные результаты (см. рис. ниже).

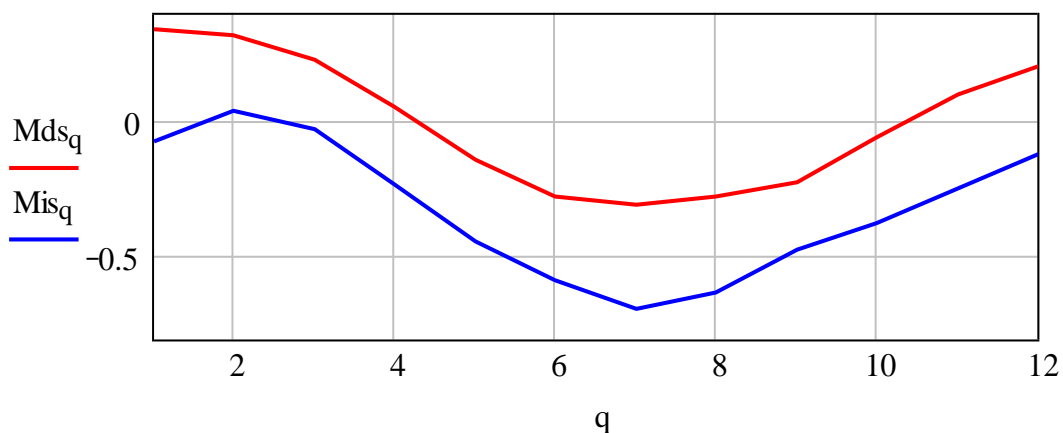
$$MY := \text{augment}(y1, y2, y3, y4, y5, y6, y7, y8)$$

$$M1 := MY^T \quad p := 1.. \text{rows}(M1) \quad q := 1.. \text{cols}(M1)$$

$$\text{rows}(M) = 96 \quad Md_q := \text{median}(M1^{\langle q \rangle}) \quad Me_q := \text{mean}(M1^{\langle q \rangle})$$

$$Mi_q := \text{min}(M1^{\langle q \rangle}) \quad b := 2 \quad \tau_q := q$$

$$Mds := \text{ksmooth}(\tau, Md, b)$$

$$Mis := \text{ksmooth}(\tau, Mi, b)$$


Среднемесячные характеристики электропотребления: абсцисса – номер месяца в году; ордината – электропотребление в относительных единицах; верхняя кривая – среднемесячные медианные значения электропотребления; нижняя кривая – среднемесячные минимальные значения

Вычисляем среднемесячные значения дисперсии, для чего задаемся доверительным интервалом в две сигмы (достаточно для данного случая) и соответствующей шириной гауссовского окна для функции сглаживания.

$$\Delta := 2.0$$

$$b1 := 2$$

$$S_q := \sqrt{\frac{1}{(\text{cols}(M1) - 1)} \cdot \sum_p [M1_{(p,q)} - Me_q]^2}$$

Для того чтобы не вычислять каждое среднемесячное значение дисперсии за все годы предыстории, проверяем гипотезу об однородности дисперсий электропотребления соответствующего объекта по месяцам за все годы наблюдений с помощью критерия Кохрена.

$$D_s := S^2$$

$$Q := \frac{\max(D_s)}{\sum_q D_{sq}}$$

$$Q = 0.121$$

$$Q_{kr} := 0.23$$

Так как $Q < Q_{kr}$, то нулевая гипотеза принимается. Дисперсии однородны и в расчетах может использоваться их среднее:

$$S_m := \text{mean}(S)$$

Верхняя и нижняя границы доверительного интервала:

$$E_{pi_q} := M_{ds_q} + \Delta \cdot S_m$$

$$E_{si_q} := M_{ds_q} - \Delta \cdot S_m$$

Прогнозирование электропотребления на следующий временной шаг

Задаем необходимый горизонт прогноза электропотребления, тренд, минимальные и медианные параметры, верхний и нижний 95 %-ые пороги. Объединяем и визуализируем данные (см. рис. ниже).

$$G := 12$$

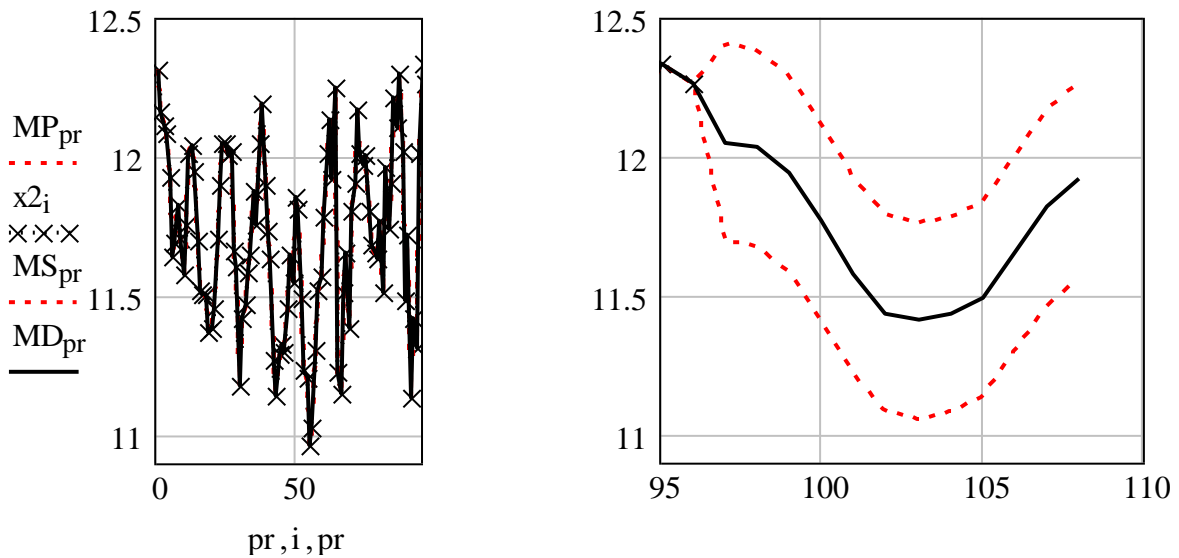
$$pr := 1.. \text{rows}(M) + G$$

$$tpr_{pr} := pr$$

$$xpr := A_1 + A_2 \cdot tpr$$

$PR_q := xpr_{rows}(M)+q$
 $Emi := Mis + PR$
 $Emd := Mds + PR$
 $Epr := Epi + PR$
 $Eps := Esi + PR$

$MD := stack(x2, Emd)$
 $MP := stack(x2, Epr)$
 $MS := stack(x2, Eps)$

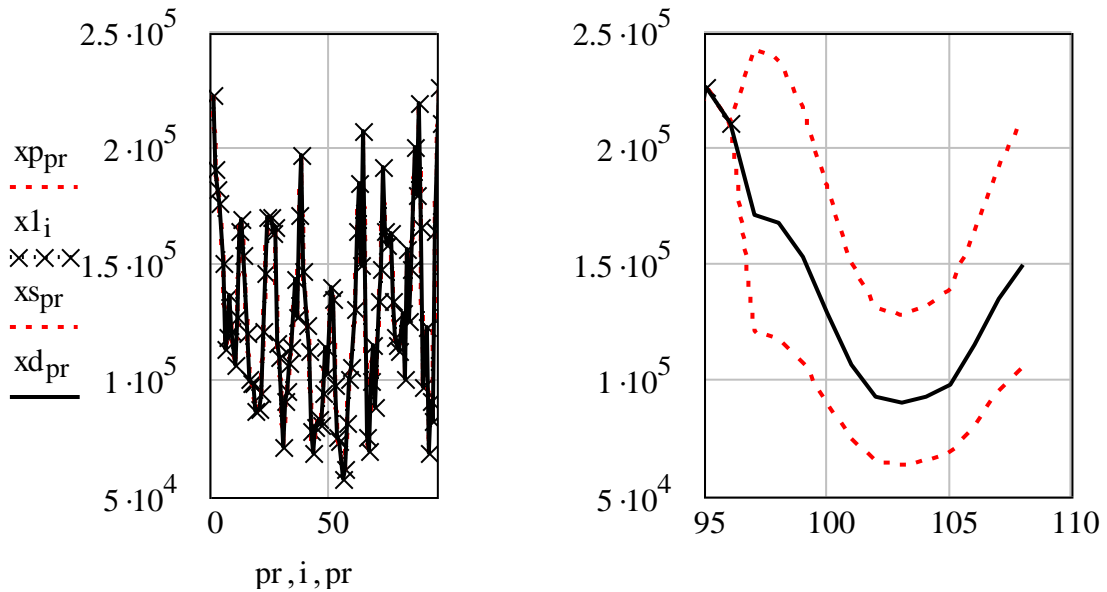


Электропотребление объекта на последующих временных шагах: абсцисса – номер месяца; ордината – электропотребление в относительных единицах; левая часть графика – электропотребление за известный период времени с реальными данными (крестики); правая часть – прогнозное значение электропотребления с 95 %-ми доверительными границами (штриховые линии)

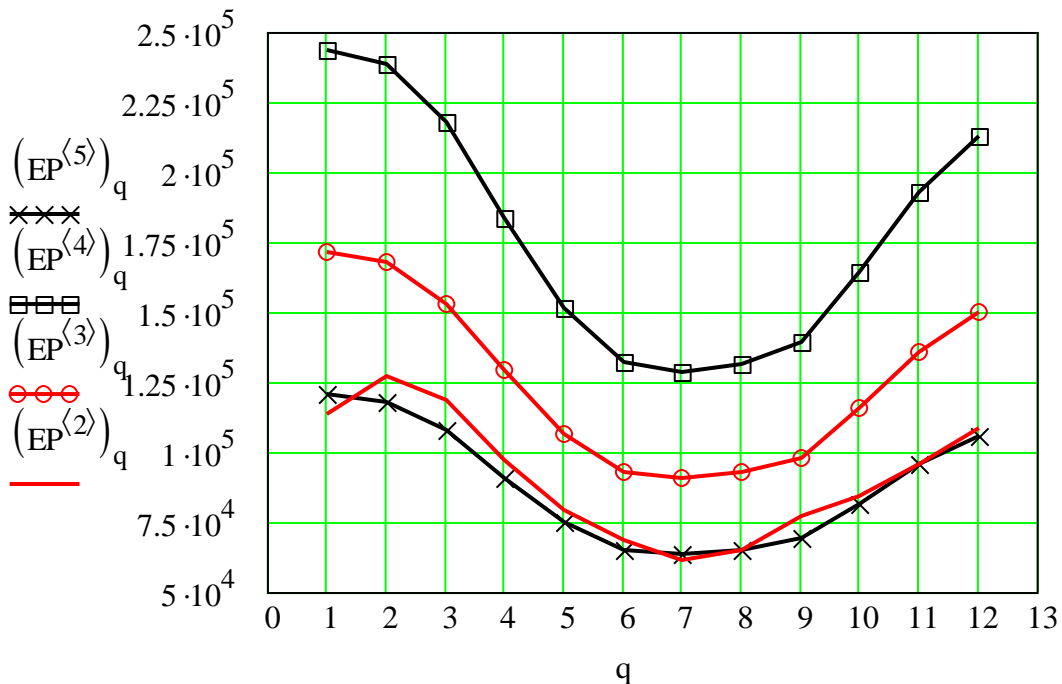
Осуществим обратное преобразование данных и получим прогнозные значения электропотребления объекта в кВт·ч (см. рис. ниже). Для получения прогнозного значения электропотребления формируется матрица, включающая преобразованные минимальный, медианный и 95 %-ый доверительный интервал с учетом тренда на 12 месяцев (см. рис ниже).

$\xrightarrow{\quad}$
 $xp := \exp(MP)$
 $\xrightarrow{\quad}$
 $xd := \exp(MD)$
 $\xrightarrow{\quad}$
 $xs := \exp(MS)$

$EP := augment\left(\tau, \xrightarrow{\quad} \exp(Emi), \xrightarrow{\quad} \exp(Emd), \xrightarrow{\quad} \exp(Epr), \xrightarrow{\quad} \exp(Eps)\right)$



Электронепотребление объекта на следующий временной шаг, кВт·ч (расшифровка та же, что и к предыдущему рисунку)



Прогноз электронепотребления объекта на год: абсцисса – номер месяца; ордината – электронепотребление, кВт·ч; EP(2) – EP(3) – прогнозный минимальный – медианный интервал с учетом тренда на год; EP(5) – EP(4) – прогнозный доверительный 95 %-ый интервал с учетом тренда на год

По результатам расчетов строится таблица прогнозных значений электронепотребления рассматриваемого объекта (кВт·ч), в которой по столбцам отражается: 1 – номер месяца; 2 – минимальные значения; 3 – медианные значения; 4 и 5 – доверительные 95 %-ые границы.

| | 1 | 2 | 3 | 4 | 5 |
|------|---|--------------------|--------------------|--------------------|--------------------|
| EP = | 1 | $1.133 \cdot 10^5$ | $1.708 \cdot 10^5$ | $2.429 \cdot 10^5$ | $1.2 \cdot 10^5$ |
| | 2 | $1.266 \cdot 10^5$ | $1.675 \cdot 10^5$ | $2.383 \cdot 10^5$ | $1.178 \cdot 10^5$ |
| | 3 | $1.183 \cdot 10^5$ | $1.527 \cdot 10^5$ | $2.172 \cdot 10^5$ | $1.073 \cdot 10^5$ |
| | 4 | $9.702 \cdot 10^4$ | $1.286 \cdot 10^5$ | $1.83 \cdot 10^5$ | $9.043 \cdot 10^4$ |
| | 5 | $7.868 \cdot 10^4$ | $1.06 \cdot 10^5$ | $1.508 \cdot 10^5$ | $7.455 \cdot 10^4$ |
| | 6 | $6.813 \cdot 10^4$ | $9.243 \cdot 10^4$ | $1.315 \cdot 10^5$ | $6.498 \cdot 10^4$ |

В качестве прогнозного электропотребления объекта на последующий год берется сумма медианных прогнозных значений:

$$EPP := \sum_q \overrightarrow{\exp(Emd_q)}$$

$$EPP = 1.498 \times 10^6$$

Анализ результатов прогноза на гауссовость

Завершающим этапом в процедуре прогнозирования является анализ остатков на соответствие нормальному закону (см. рис. ниже).

$$MO1_{p,q} := M1_{p,q} - Me_q$$

$$MO := MO1^T$$

$$O := \text{stack}(MO^{(1)}, MO^{(2)}, MO^{(3)}, MO^{(4)}, MO^{(5)}, MO^{(6)}, MO^{(7)}, MO^{(8)})$$

$$\text{ORIGIN} := 0$$

$$m := \text{round}(\sqrt{\text{length}(O)})$$

$$h := \frac{\max(O) - \min(O)}{m}$$

$$l := 0..m$$

$$v_l := \min(O) + h \cdot l$$

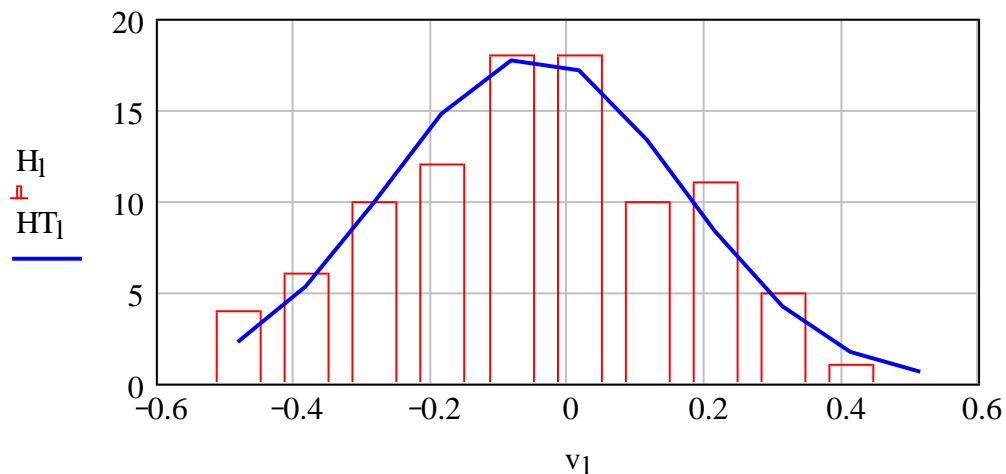
$$H := \text{hist}(v, O)$$

$$H_{m+1} := 0$$

$$v_{m+1} := v_m + h \quad \sum_1 H_l = 95$$

$$HT_1 := \text{length}(O) \cdot \int_{v_1}^{v_{l+1}} \text{dnorm}(u, \text{mean}(O), \text{stdev}(O)) du$$

$$\sum_1 HT_1 = 94.753$$



Проверка на соответствие распределения прогнозных значений электропотребления нормальному закону: гистограмма – расчетные значения; кривая – теоретические значения

$$\chi_i := \sum_1 \frac{(H_1 - HT_1)^2}{HT_1}$$

$$\chi = 18.307$$

$$\chi := \text{qchisq}(0.95, 10)$$

$$\chi_i = 4.637$$

Так как $\chi_i < \chi$, то гипотеза принимается: остатки распределены по нормальному закону. Это, в свою очередь, свидетельствует о корректности прогноза G-методом для данного объекта техноценоза.

[\[на содержание\]](#)

ВЕРИФИКАЦИЯ ИСХОДНОЙ БАЗЫ ДАННЫХ ПО ЭЛЕКТРОПОТРЕБЛЕНИЮ

После сбора информации о техноценозе создается электронная база данных, которая представляет собой неупорядоченную совокупность значений исследуемых параметров электропотребления объектов техноценоза. Далее, уже программными средствами, осуществляется импорт исходных данных из файла «data_d.xls» в тело программы Mathcad и задается начало отсчета. Делается это следующим образом:

V := 
C:\mathcad_dat\data_d.xls

ORIGIN := 1

Сформированная таким образом матрица «V» содержит информацию об электропотреблении исследуемого техноценоза. Причем, каждая строка соответствует определенному году, а каждый столбец – объекту. Для иллюстрации в данной программе использовались собранные за десять лет данные по электропотреблению одного из реально существующих техноценозов, расположенных на территории Калининградской области (всего объектов – 69; диапазон электропотребления – 3 ... 3000 МВт·ч/год; суммарное потребление – 18 ГВт·ч/год; тип – устойчивый техноценоз). Далее полученную матрицу можно вывести и просмотреть в теле программы.

$$V =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 1 | $1.81 \cdot 10^6$ | $1.268 \cdot 10^6$ | $2.328 \cdot 10^5$ | $5.817 \cdot 10^5$ | $4.386 \cdot 10^5$ | $7.438 \cdot 10^5$ | $2.891 \cdot 10^4$ |
| 2 | $1.447 \cdot 10^6$ | $1.332 \cdot 10^6$ | $8.567 \cdot 10^5$ | $5.244 \cdot 10^5$ | $4.027 \cdot 10^5$ | $6.371 \cdot 10^5$ | $2.636 \cdot 10^4$ |
| 3 | $1.474 \cdot 10^6$ | $1.613 \cdot 10^6$ | $8.373 \cdot 10^5$ | $5.784 \cdot 10^5$ | $5.179 \cdot 10^5$ | $7.094 \cdot 10^5$ | $2.724 \cdot 10^4$ |
| 4 | $1.351 \cdot 10^6$ | $1.614 \cdot 10^6$ | $3.494 \cdot 10^5$ | $4.025 \cdot 10^5$ | $5.673 \cdot 10^5$ | $3.652 \cdot 10^5$ | $3.26 \cdot 10^4$ |
| 5 | $1.162 \cdot 10^6$ | $1.304 \cdot 10^6$ | $3.971 \cdot 10^5$ | $3.274 \cdot 10^5$ | $4.366 \cdot 10^5$ | $8.78 \cdot 10^4$ | $2.701 \cdot 10^4$ |
| 6 | $1.628 \cdot 10^6$ | $1.775 \cdot 10^6$ | $4.101 \cdot 10^5$ | $3.439 \cdot 10^5$ | $3.379 \cdot 10^5$ | $7.25 \cdot 10^4$ | $3.02 \cdot 10^4$ |
| 7 | $1.624 \cdot 10^6$ | $1.617 \cdot 10^6$ | $4.107 \cdot 10^5$ | $3.509 \cdot 10^5$ | $3.144 \cdot 10^5$ | $9.39 \cdot 10^4$ | $2.898 \cdot 10^4$ |

Как показывает опыт, далеко не всегда исходные данные, используемые для процедур управления ресурсами техноценоза, оказываются вполне корректными, что в значительной степени снижает достоверность интервального оценивания, прогнозирования и нормирования. Следовательно, требуется предварительная верификация базы данных, которая включает следующие основные статистические процедуры:

- устранение нулевых данных;
- устранение явно ошибочных данных (выбросов);
- устранение абсолютно равных данных;
- восстановление утерянных данных.

Нулевые и абсолютно равные данные являются первым признаком некорректности базы, что очевидно даже с точки зрения физического смысла. Дело в том, что реально существующий объект никак не может потреблять нулевое количество электроэнергии, равно как и два разных объекта не могут потреблять абсолютно равное количество энергии. Кроме того, подобные данные очень плохо обрабатываются компьютерными программами. Выбросы в данных являются следствием грубых ошибок при фиксации параметров обслуживающим персоналом либо сбоев в работе технических средств измерений. Наконец, утеря данных может произойти как по вине персонала, так и по причине тяжелых сбоев в работе серверов, на которых хранятся базы. В ряде случаев требуется просто корректное наращивание базы данных на несколько лет «назад». Необходимо отме-

туть, что верификация не является обязательной процедурой, однако она всегда должна применяться в том случае, если есть хотя бы малейшее сомнение в корректности исходных данных. Итак, рассмотрим процедуры верификации на примере имеющейся базы по электропотреблению.

Устранение нулевых данных

Нулевые значения электропотребления, находящиеся внутри базы данных, восстанавливаются при помощи сплайн-интерполяции, а на границе базы данных – применением функции экстраполяции.

Устраним нулевые данные внутри базы данных методом сплайн-интерполяции. Для этого сформируем вектор времени.

```
i := 1..rows(V)
Qi := i
```

Входными параметрами программы являются исходная матрица данных по электропотреблению техноценоза и вектор времени.

```
H1(A,Q) := | V ← A
            | for i ∈ 1..rows(V)
            |   for j ∈ 1..cols(V)
            |     if Vi,j = 0 ∧ i ≠ 1 ∧ i ≠ rows(V)
            |       | X ←  $\begin{pmatrix} Q_{i-2} \\ Q_{i-1} \\ Q_{i+1} \end{pmatrix}$ 
            |       | Y ←  $\begin{bmatrix} (V^{<j>})_{i-2} \\ (V^{<j>})_{i-1} \\ (V^{<j>})_{i+1} \end{bmatrix}$ 
            |       | S ← cspline(X, Y)
            |       | Wi,j ← interp[S, X, Y, (V^{<1>})i]
            |       | Wi,j ← Vi,j otherwise
            |     W
```

Интерполяционные процедуры связаны с нахождением функциональной зависимости, наилучшим образом описывающей имеющиеся эмпирические данные по электропотреблению. При этом могут быть применены критерии, вытекающие из методов наименьших квадратов или наименьших модулей (последний способ менее предпочтителен, т.к. трудно реализуем на практике). Полученная функциональная зависимость называется интерполяционной. После ее нахождения определяются значения функции, соответствующие рангам объектов, для которых в базе данных имеются нули или абсолютно равные данные.

$$H(A, Q) := \left| \begin{array}{l} V \leftarrow A \\ \text{for } i \in 1.. \text{rows}(V) \\ \quad \text{for } j \in 1.. \text{cols}(V) \\ \quad \quad \text{if } V_{i,j} = 0 \wedge i \neq 1 \wedge i \neq \text{rows}(V) \\ \quad \quad \quad \left| \begin{array}{l} X \leftarrow \begin{pmatrix} Q_{i-1} \\ Q_{i+1} \\ Q_{i+2} \end{pmatrix} \\ Y \leftarrow \begin{bmatrix} (V^{(j)})_{i-1} \\ (V^{(j)})_{i+1} \\ (V^{(j)})_{i+2} \end{bmatrix} \\ S \leftarrow \text{cspline}(X, Y) \\ W_{i,j} \leftarrow \text{interp}(S, X, Y, Q_i) \\ W_{i,j} \leftarrow V_{i,j} \text{ otherwise} \end{array} \right. \\ \quad \quad \quad W \end{array} \right.$$

$$F := H(V, Q)$$

Алгоритм программы имеет одну существенную особенность: для устранения нулевых данных необходимо, чтобы одно значение до нуля и два после не были равны нулю. Это получается далеко не всегда. При необходимости исследователь может адаптировать алгоритм программы в зависимости от положения нулевых данных, для чего надо целенаправленно изменить значения элементов векторов «X» и «Y», используемых основной функцией сплайн-интерполяции «cspline».

```

H1(A,Q) := | V ← A
            | for i ∈ 1..length(V)
            |   | if Vi = 0 ∧ i ≠ 1 ∧ i ≠ rows(V)
            |   |   | X ←  $\begin{pmatrix} Q_{i-1} \\ Q_{i+2} \\ Q_{i+3} \end{pmatrix}$ 
            |   |   | Y ←  $\begin{pmatrix} V_{i-1} \\ V_{i+2} \\ V_{i+3} \end{pmatrix}$ 
            |   |   | S ← cspline(X, Y)
            |   |   | Wi ← interp(S, X, Y, Qi)
            |   | Wi ← Vi otherwise
            | W
    
```

Устраним нулевые данные методом экстраполяции.

```

FF(A) := | for i ∈ 1..rows(A)
          |   for j ∈ 1..cols(A)
          |     | if Ai,j = 0 ∧ i = 1
          |     |   | F ← reverse(A⟨j⟩)
          |     |   | F1 ← submatrix(F, 1, rows(F) - 1, 1, 1)
          |     |   | Wi,j ← |predict(F1, 3, 1)|
          |     | Wi,j ← Ai,j otherwise
          |   for k ∈ 1..rows(W)
          |     for l ∈ 1..cols(W)
          |       | if Wk,1 = 0 ∧ k = rows(W)
          |       |   | F2 ← submatrix(W⟨l⟩, 1, rows(W⟨l⟩) - 1, 1, 1)
          |       |   | W1k,1 ← |predict(F2, 3, 1)|
          |       | W1k,1 ← Wk,1 otherwise
          | W1
    
```

W := FF(F)

Устранение явно ошибочных данных (выбросов)

Выбросы – это нетипичные данные, которые существенно (без корректного объяснения) отклоняются от центра распределения. Подобные данные связаны с грубыми ошибками измерения электропотребления и обязательно должны подвергаться процедуре верификации. Рассмотрим, как это может быть осуществлено встроенными программными средствами Mathcad (визуализация процедуры – см. рис. ниже).

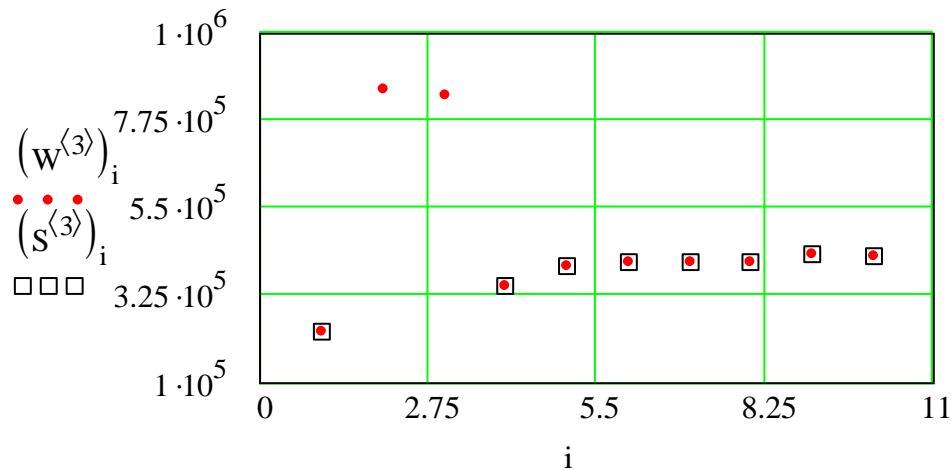
```

DR(G) := | S1 ← G
          | for j ∈ 1..cols(S1)
          |   | n ← length(S1<sup>j</sup>)
          |   | y ← sort(S1<sup>j</sup>)
          |   | H1 ← (0.25·n + 0.25)
          |   | H2 ← (0.75·n + 0.75)
          |   | Q1 ← H1 – floor(H1)
          |   | Q2 ← H2 – floor(H2)
          |   | p1 ← yfloor(H1) + Q1·(yceil(H1) – yfloor(H1))
          |   | p2 ← yfloor(H2) + Q2·(yceil(H2) – yfloor(H2))
          |   | L ← p2 – p1
          |   | for i ∈ 1..n
          |   |   | qi,j ← S1i,j if p1 – 1.5·L < S1i,j < p2 + 1.5·L
          |   |   | qi,j ← 0 otherwise
          | q
    
```

DR(W) =

| | 1 | 2 | 3 | 4 |
|---|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | 1.81·10 ⁶ | 1.268·10 ⁶ | 2.328·10 ⁵ | 5.817·10 ⁵ |
| 2 | 1.447·10 ⁶ | 1.332·10 ⁶ | 0 | 5.244·10 ⁵ |
| 3 | 1.474·10 ⁶ | 1.613·10 ⁶ | 0 | 5.784·10 ⁵ |
| 4 | 1.351·10 ⁶ | 1.614·10 ⁶ | 3.494·10 ⁵ | 4.025·10 ⁵ |
| 5 | 1.162·10 ⁶ | 1.304·10 ⁶ | 3.971·10 ⁵ | 3.274·10 ⁵ |
| 6 | 1.628·10 ⁶ | 1.775·10 ⁶ | 4.101·10 ⁵ | 3.439·10 ⁵ |
| 7 | 1.624·10 ⁶ | 1.617·10 ⁶ | 4.107·10 ⁵ | 3.509·10 ⁵ |

S := DR(W)



Временной ряд электропотребления одного из объектов техноценоза (номер объекта в исходной базе данных – 3): абсцисса – непрерывное время (в данном случае – годы); ордината – электропотребление, кВт·ч; точки с квадратами – нормальные значения; точки без квадратов – выбросы

Анализируя выбросы значений электропотребления, исследователь должен выявить их природу и принять решение, использовать их для дальнейшей работы либо произвести замену сплайн-интерполяцией или экстраполяцией. Ниже приведен пример повекторного восстановления.

Вычислим индексы векторов, имеющих нулевые данные, а затем извлечем эти вектора из базы данных по электропотреблению.

```

Ind(S) := | D ← 0
          | for j ∈ 1.. cols(S)
          |   for i ∈ 1.. rows(S)
          |     D ← stack(D,j) if S(i,j) = 0
          |   D1 ← submatrix(D,2,length(D),1,1)
          | D1T
    
```

Ind(S) = (3 3 16 24 32 35 54 57 62)

$$S^{(3)} =$$

| | |
|---|-------------------------|
| | 1 |
| 1 | 2.328 · 10 ⁵ |
| 2 | 0 |
| 3 | 0 |
| 4 | 3.494 · 10 ⁵ |
| 5 | 3.971 · 10 ⁵ |
| 6 | 4.101 · 10 ⁵ |

$$S^{(16)} =$$

| | |
|---|-------------------------|
| | 1 |
| 1 | 2.177 · 10 ⁵ |
| 2 | 0 |
| 3 | 1.593 · 10 ⁵ |
| 4 | 1.376 · 10 ⁵ |
| 5 | 1.12 · 10 ⁵ |
| 6 | 3.132 · 10 ⁴ |

$$S^{(24)} =$$

| | |
|---|-------------------------|
| | 1 |
| 1 | 0 |
| 2 | 1.155 · 10 ⁵ |
| 3 | 1.378 · 10 ⁵ |
| 4 | 1.514 · 10 ⁵ |
| 5 | 1.512 · 10 ⁵ |
| 6 | 1.235 · 10 ⁵ |

$$S^{(32)} =$$

| | 1 |
|---|--------------------|
| 1 | $1.915 \cdot 10^4$ |
| 2 | $1.726 \cdot 10^4$ |
| 3 | $1.756 \cdot 10^4$ |
| 4 | $2.131 \cdot 10^4$ |
| 5 | $2.126 \cdot 10^4$ |
| 6 | $3.602 \cdot 10^4$ |

$$S^{(35)} =$$

| | 1 |
|---|--------------------|
| 1 | $6.02 \cdot 10^4$ |
| 2 | $5.659 \cdot 10^4$ |
| 3 | $5.609 \cdot 10^4$ |
| 4 | $6.358 \cdot 10^4$ |
| 5 | $4.958 \cdot 10^4$ |
| 6 | $5.762 \cdot 10^4$ |

$$S^{(54)} =$$

| | 1 |
|---|--------------------|
| 1 | $1.222 \cdot 10^4$ |
| 2 | $1.184 \cdot 10^4$ |
| 3 | $1.062 \cdot 10^4$ |
| 4 | $1.374 \cdot 10^4$ |
| 5 | $1.193 \cdot 10^4$ |
| 6 | $1.557 \cdot 10^4$ |

$$S^{(57)} =$$

| | 1 |
|---|--------------------|
| 1 | $3.514 \cdot 10^4$ |
| 2 | $3.365 \cdot 10^4$ |
| 3 | $3.059 \cdot 10^4$ |
| 4 | $3.813 \cdot 10^4$ |
| 5 | $3.975 \cdot 10^4$ |
| 6 | $4.705 \cdot 10^4$ |

$$S^{(62)} =$$

| | 1 |
|---|--------------------|
| 1 | $1.877 \cdot 10^4$ |
| 2 | $1.734 \cdot 10^4$ |
| 3 | $1.617 \cdot 10^4$ |
| 4 | $1.583 \cdot 10^4$ |
| 5 | $1.625 \cdot 10^4$ |
| 6 | $2.062 \cdot 10^4$ |

Как видно, структура нулевых данных различна, поэтому восстанавливаем вектора по отдельности, а затем производим их замену в базе данных техноценоза по электропотреблению (см. рис. ниже).

$$Q1 := H1(S^{(3)}, Q)$$

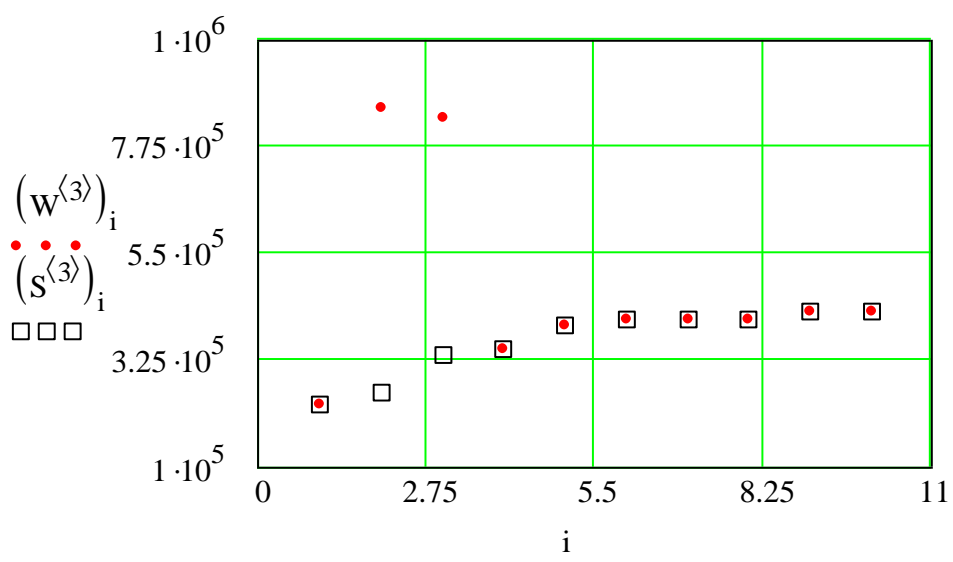
$$S^{(3)} := Q1$$

```

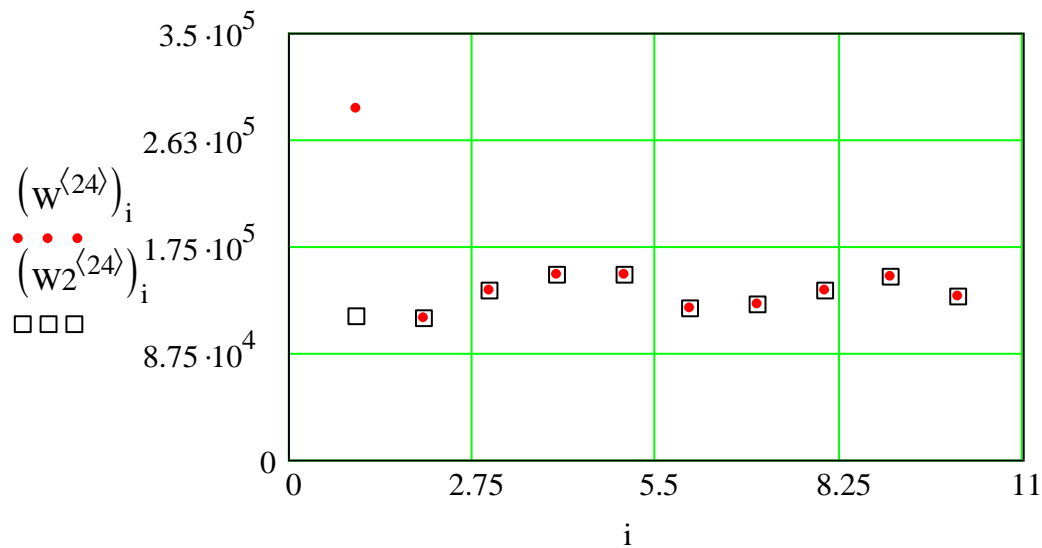
H2(A,Q) :=
  V ← A
  for i ∈ 1..length(V)
    if Vi = 0 ∧ i ≠ 1 ∧ i ≠ rows(V)
      X ←  $\begin{pmatrix} Q_{i-1} \\ Q_{i+1} \\ Q_{i+2} \end{pmatrix}$ 
      Y ←  $\begin{pmatrix} V_{i-1} \\ V_{i+1} \\ V_{i+2} \end{pmatrix}$ 
      S ← cspline(X,Y)
      Wi ← interp(S,X,Y,Qi)
    Wi ← Vi otherwise
  W
  
```

```

H3(A,Q) := | V ← A
            | for i ∈ 1..length(V)
            |   | if Vi = 0 ∧ i ≠ 1 ∧ i ≠ rows(V)
            |   |   | X ←  $\begin{pmatrix} Q_{i-2} \\ Q_{i-1} \\ Q_{i+1} \end{pmatrix}$ 
            |   |   | Y ←  $\begin{pmatrix} V_{i-2} \\ V_{i-1} \\ V_{i+1} \end{pmatrix}$ 
            |   |   | S ← cspline(X,Y)
            |   |   | Wi ← interp(S,X,Y,Qi)
            |   |   | Wi ← Vi otherwise
            |   | W
            | W
Q3 := H2(S<16>,Q)   S<16> := Q3   Q4 := H3(S<32>,Q)   S<32> := Q4
Q5 := H3(S<35>,Q)   S<35> := Q5   Q6 := H3(S<62>,Q)   S<62> := Q6
W2 := FF(S)
    
```



Временной ряд электропотребления одного из объектов техноценоза (номер объекта в базе данных – 3): абсцисса – непрерывное время (годы); ордината – электропотребление, кВт·ч; точки без квадратов – выбросы; квадраты без точек – верифицированные данные



Временной ряд электропотребления одного из объектов техноценоза (номер данного объекта в исследуемой базе данных – 24): абсцисса – непрерывное время (годы); ордината – параметр электропотребления, кВт·ч; точка без квадрата – аномальный выброс; квадрат без точки – верифицированное значение электропотребления объекта

Устранение абсолютно равных данных

Абсолютно равные данные устраняется путем прибавления к исходным данным относительно малых значений случайной величины, сгенерированной по нормальному закону. При этом в качестве генератора псевдослучайного числа используется стандартная функция «rnorm».

$$P(T) := \begin{cases} \text{for } j \in 1.. \text{cols}(T) \\ a^{\langle j \rangle} \leftarrow T^{\langle j \rangle} + \text{rnorm}(\text{rows}(T), 2, 1.1) \\ a \end{cases}$$

$$A := P(W2)$$

Восстановление утерянных данных

Восстановление утерянной части базы данных техноценоза может быть осуществлено экстраполяцией по каждому объекту.

Прежде всего, исследователь должен в теле программы задать переменную «int» – количество лет, на которое необходимо восстановить базу данных техноценоза по электропотреблению (в данном случае – 2 года).

int := 2

$$USS(A, int) := \left| \begin{array}{l} \text{for } j \in 1.. \text{cols}(A) \\ \quad \left| \begin{array}{l} S \leftarrow \text{reverse}(A^{(j)}) \\ A1^{(j)} \leftarrow \text{stack}(\text{predict}(S, 3, int), A^{(j)}) \end{array} \right. \\ A1 \end{array} \right.$$

S := USS(A, int)

Верифицированную базу данных техноценоза по электропотреблению далее можно просмотреть в теле программы как матрицу (в данном случае показан лишь ее фрагмент размерностью семь на семь).

$$S = \begin{array}{|c|c|c|c|c|c|c|c|} \hline & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \hline 1 & 1.62 \cdot 10^6 & 1.402 \cdot 10^6 & 1.723 \cdot 10^5 & 5.69 \cdot 10^5 & 5.014 \cdot 10^5 & 6.93 \cdot 10^5 & 2.7 \cdot 10^4 \\ \hline 2 & 1.689 \cdot 10^6 & 1.333 \cdot 10^6 & 1.46 \cdot 10^5 & 5.629 \cdot 10^5 & 5.012 \cdot 10^5 & 6.051 \cdot 10^5 & 2.737 \cdot 10^4 \\ \hline 3 & 1.81 \cdot 10^6 & 1.268 \cdot 10^6 & 2.328 \cdot 10^5 & 5.817 \cdot 10^5 & 4.386 \cdot 10^5 & 7.438 \cdot 10^5 & 2.891 \cdot 10^4 \\ \hline 4 & 1.447 \cdot 10^6 & 1.332 \cdot 10^6 & 2.565 \cdot 10^5 & 5.244 \cdot 10^5 & 4.027 \cdot 10^5 & 6.371 \cdot 10^5 & 2.636 \cdot 10^4 \\ \hline 5 & 1.474 \cdot 10^6 & 1.614 \cdot 10^6 & 3.37 \cdot 10^5 & 5.784 \cdot 10^5 & 5.179 \cdot 10^5 & 7.094 \cdot 10^5 & 2.724 \cdot 10^4 \\ \hline 6 & 1.351 \cdot 10^6 & 1.614 \cdot 10^6 & 3.494 \cdot 10^5 & 4.025 \cdot 10^5 & 5.673 \cdot 10^5 & 3.652 \cdot 10^5 & 3.26 \cdot 10^4 \\ \hline 7 & 1.162 \cdot 10^6 & 1.304 \cdot 10^6 & 3.971 \cdot 10^5 & 3.274 \cdot 10^5 & 4.366 \cdot 10^5 & 8.78 \cdot 10^4 & 2.701 \cdot 10^4 \\ \hline \end{array}$$

Кроме того, визуализируем результаты произведенных расчетов, построив итоговую верифицированную ранговую параметрическую поверхность техноценоза по электропотреблению (см. рис. ниже).

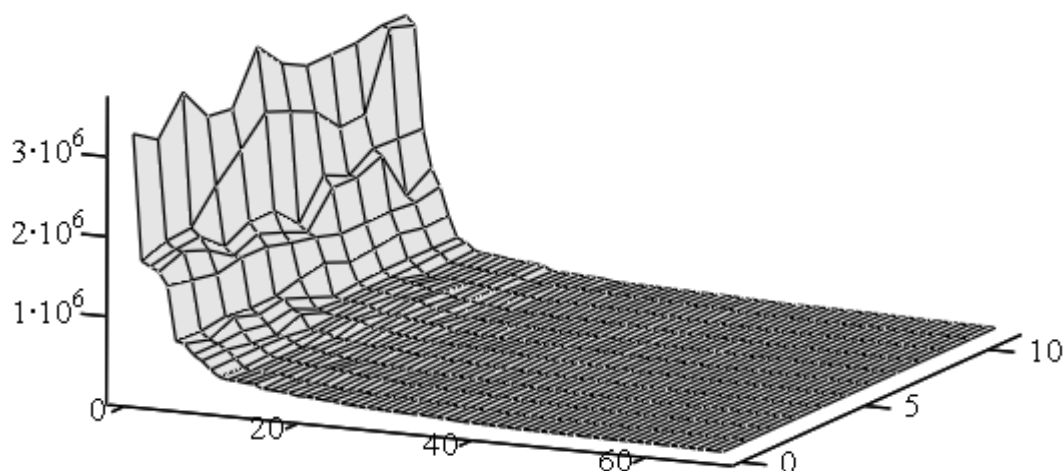
$$\text{Zipf}(S) := \left| \begin{array}{l} W1 \leftarrow S^T \\ \text{for } i \in 1.. \text{cols}(W1) \\ \quad \left| \begin{array}{l} c \leftarrow \text{sort}(W1^{(i)}) \\ b^{(i)} \leftarrow \text{reverse}(c) \end{array} \right. \\ b \end{array} \right.$$

В заключение сохраним верифицированную базу данных техноценоза по электропотреблению во внешний файл «data_dv.xls».



C:\mathcad_dat\data_dv.xls

S



Zipf (S)

Верифицированная трехмерная ранговая поверхность техноценоза по электропотреблению: абсцисса – ранг объекта; ордината – непрерывное время (годы); аппликата – электропотребление, кВт·ч

[на содержание]

ОЦЕНКА АДЕКВАТНОСТИ РАБОТЫ ДИНАМИЧЕСКОЙ АДАПТИВНОЙ МОДЕЛИ

Оценка адекватности динамической адаптивной модели управления процессом электропотребления техноценоза включает две основные процедуры. Первая заключается в оценке ошибок (погрешностей) прогнозирования, т.е. собственно проверке качества процесса моделирования. Вторая процедура предполагает анализ остатков моделирования на наличие полезной информации и неучтенных моделью закономерностей.

Замысел оценки адекватности заключается в следующем. Из имеющихся данных по электропотреблению объектов техноценоза последние пять лет резервируются в качестве матрицы верификации, а по оставшимся данным осуществляется перспективный прогноз на глубину матрицы верификации. Затем полученные прогнозные значения электропотребления техноценоза сравниваются с соответствующими эмпирическими данными с использованием двух процедур оценки адекватности. Отрицательный вывод по первой процедуре свидетельствует либо о существенных недостатках в расчетной методологии, применяемой в модели, либо о неправильной оценке ее прогностических возможностей. Вторая процедура служит индикатором полноты и корректности алгоритма модели. В любом

случае, отрицательный вывод хотя бы по одной из процедур не позволяет применять проверяемую модель в практических расчетах.

Результаты эмпирических и прогнозных значений электропотребления заносятся в файл MS Excel. Файл должен быть назван «Prognoz.xls» и помещен в директорию «с:\mathcad_dat», которая должна быть заблаговременно создана в корневом каталоге диска «с:\». Файл должен состоять из одних цифр и иметь следующую структуру: столбцы должны соответствовать объектам техноценоза, а строки – годам. Причем первые пять строк заполняются прогнозными значениями, полученными в результате моделирования, а последующие пять – эмпирическими (реально зафиксированными) данными для соответствующих объектов техноценоза в те же годы. Таким образом, первая прогнозная строка должна соответствовать шестой эмпирической, вторая – седьмой, третья – восьмой и т.д.

Здесь используются результаты расчетов, осуществленных применительно к реально существующему в Калининградской области техноценозу, состоящему из рассредоточенных по всей территории региона 69 объектов. Расчеты производились с помощью динамической адаптивной модели электропотребления. При этом использовалась база данных техноценоза по электропотреблению за 10 лет в период 1995 – 2004 гг.

Подготовка данных

Задаем начало отсчета, импортируем данные для обработки, а полученную матрицу транспонируем и извлекаем из нее столбцы, где:

x1 – x5 – полученные с помощью модели (с использованием данных за 1995 – 1999 гг.) прогнозные значения на 2000 – 2004 гг.;

x6 – x10 – эмпирические значения электропотребления за 2000 – 2004 гг.

ORIGIN := 1

V :=  C:\.\Prognoz.xls

W := V^T

x1 := W^{<1>} x2 := W^{<2>} x3 := W^{<3>} x4 := W^{<4>} x5 := W^{<5>}
 x6 := W^{<6>} x7 := W^{<7>} x8 := W^{<8>} x9 := W^{<9>} x10 := W^{<10>}

Определяем объем выборки и задаем натуральный ряд:

n := rows(W)

n = 69 i := 1 .. n

Оценка ошибок прогнозирования (первая процедура оценки адекватности модели)

Для оценки точности моделирования полученные результаты статистически сравниваются. Прежде всего, вычисляются абсолютные ошибки прогнозирования электропотребления объектов техноценоза:

$$\begin{aligned}\varepsilon 1 &:= x6 - x1 & \varepsilon 2 &:= x7 - x2 \\ \varepsilon 3 &:= x8 - x3 & \varepsilon 4 &:= x9 - x4 \\ \varepsilon 5 &:= x10 - x5\end{aligned}$$

Максимальные абсолютные ошибки по годам:

$$\begin{aligned}\max(\overrightarrow{|\varepsilon 1|}) &= 1.863 \times 10^4 & \max(\overrightarrow{|\varepsilon 2|}) &= 4.687 \times 10^4 \\ \max(\overrightarrow{|\varepsilon 3|}) &= 4.226 \times 10^4 & \max(\overrightarrow{|\varepsilon 4|}) &= 1.865 \times 10^4 \\ \max(\overrightarrow{|\varepsilon 5|}) &= 8.545 \times 10^4\end{aligned}$$

Суммарные абсолютные ошибки:

$$\begin{aligned}\sum_{i=1}^n (\overrightarrow{|\varepsilon 1|})_i &= 3.294 \times 10^5 & \sum_{i=1}^n (\overrightarrow{|\varepsilon 2|})_i &= 4.421 \times 10^5 \\ \sum_{i=1}^n (\overrightarrow{|\varepsilon 3|})_i &= 5.316 \times 10^5 & \sum_{i=1}^n (\overrightarrow{|\varepsilon 4|})_i &= 5.094 \times 10^5 \\ \sum_{i=1}^n (\overrightarrow{|\varepsilon 5|})_i &= 7.348 \times 10^5\end{aligned}$$

Абсолютные ошибки моделирования, рассчитанные для исследуемого техноценоза, взятого в целом, на каждом временном шаге:

$$xf1 := \sum_{i=1}^n x6_i \quad xf1 = 1.819 \times 10^7 \quad xf2 := \sum_{i=1}^n x7_i \quad xf2 = 1.757 \times 10^7$$

$$xf3 := \sum_{i=1}^n x8_i \quad xf3 = 1.781 \times 10^7 \quad xf4 := \sum_{i=1}^n x9_i \quad xf4 = 1.826 \times 10^7$$

$$xf5 := \sum_{i=1}^n x10_i \quad xf5 = 1.81 \times 10^7 \quad xd1 := \sum_{i=1}^n x1_i \quad xd1 = 1.809 \times 10^7$$

$$xd2 := \sum_{i=1}^n x2_i \quad xd2 = 1.773 \times 10^7 \quad xd3 := \sum_{i=1}^n x3_i \quad xd3 = 1.803 \times 10^7$$

$$xd4 := \sum_{i=1}^n x4_i \quad xd4 = 1.86 \times 10^7 \quad xd5 := \sum_{i=1}^n x5_i \quad xd5 = 1.848 \times 10^7$$

$$E1 := |xf1 - xd1| \quad E1 = 1.017 \times 10^5$$

$$E2 := |xf2 - xd2| \quad E2 = 1.589 \times 10^5$$

$$E3 := |xf3 - xd3| \quad E3 = 2.16 \times 10^5$$

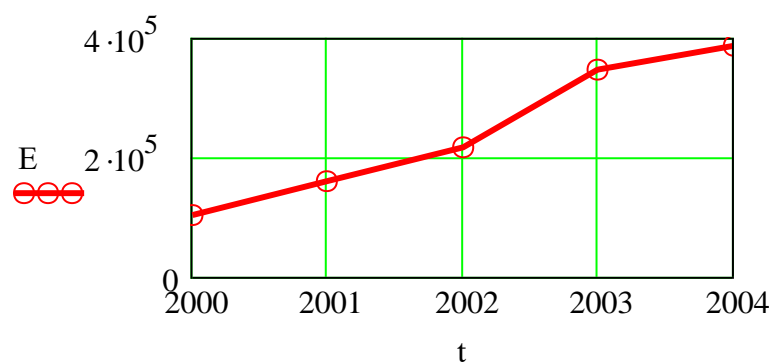
$$E4 := |xf4 - xd4| \quad E4 = 3.44 \times 10^5$$

$$E5 := |xf5 - xd5| \quad E5 = 3.851 \times 10^5$$

Группировка данных для вывода на график (см. рис. ниже):

t := stack(2000, 2001, 2002, 2003, 2004)

E := stack(E1, E2, E3, E4, E5)



Динамика изменения абсолютной ошибки моделирования техноценоза: абсцисса – годы исследования; ордината – электропотребление, кВт·ч

Коэффициент вариации, который является мерой рассеяния прогнозных значений электропотребления объектов техноценоза:

$$K1 := \frac{\text{stdev}(\varepsilon1)}{\text{mean}(x6)} \cdot 100 \quad K1 = 2.263$$

$$K2 := \frac{\text{stdev}(\varepsilon2)}{\text{mean}(x7)} \cdot 100 \quad K2 = 3.561$$

$$K3 := \frac{\text{stdev}(\varepsilon3)}{\text{mean}(x8)} \cdot 100 \quad K3 = 3.888$$

$$K4 := \frac{\text{stdev}(\varepsilon4)}{\text{mean}(x9)} \cdot 100 \quad K4 = 2.545$$

$$K5 := \frac{\text{stdev}(\varepsilon5)}{\text{mean}(x10)} \cdot 100 \quad K5 = 5.713$$

Группировка данных для вывода на график (см. рис.):

$K := \text{stack}(K1, K2, K3, K4, K5)$



Динамика изменения коэффициента вариации: абсцисса – годы исследования; ордината – значение коэффициента вариации, %

Значения коэффициента вариации в период 2000 – 2004 гг. находятся в пределах 2 – 6 %. Это показывает довольно небольшое отклонение прогнозных значений электропотребления техноценоза от эмпирических.

Относительные ошибки на каждом временном шаге:

$$\varepsilon_{otn1} := \overrightarrow{\left[\varepsilon1 \cdot (x6^{-1}) \right]}$$

$$\varepsilon_{otn2} := \overrightarrow{\left[\varepsilon2 \cdot (x7^{-1}) \right]}$$

$$\varepsilon_{otn3} := \overrightarrow{\left[\varepsilon3 \cdot (x8^{-1}) \right]}$$

$$\varepsilon_{otn4} := \overrightarrow{\left[\varepsilon4 \cdot (x9^{-1}) \right]}$$

$$\varepsilon_{otn5} := \overrightarrow{\left[\varepsilon5 \cdot (x10^{-1}) \right]}$$

Суммарные относительные ошибки:

$$\varepsilon_{otn1_Σ} := \sum_{i=1}^n \left(\overrightarrow{|\varepsilon_{otn1}|} \right)_i \quad \varepsilon_{otn1_Σ} = 5.114$$

$$\varepsilon_{otn2_Σ} := \sum_{i=1}^n \left(\overrightarrow{|\varepsilon_{otn2}|} \right)_i \quad \varepsilon_{otn2_Σ} = 6.844$$

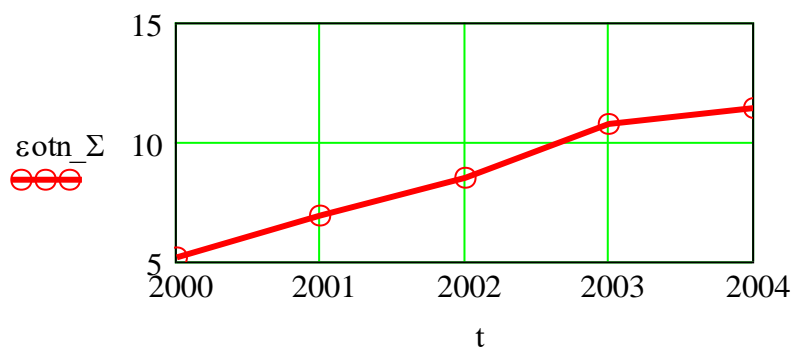
$$\varepsilon_{otn3_Σ} := \sum_{i=1}^n \left(\overrightarrow{|\varepsilon_{otn3}|} \right)_i \quad \varepsilon_{otn3_Σ} = 8.498$$

$$\varepsilon_{otn4_Σ} := \sum_{i=1}^n \left(\overrightarrow{|\varepsilon_{otn4}|} \right)_i \quad \varepsilon_{otn4_Σ} = 10.679$$

$$\varepsilon_{otn5_Σ} := \sum_{i=1}^n \left(\overrightarrow{|\varepsilon_{otn5}|} \right)_i \quad \varepsilon_{otn5_Σ} = 11.387$$

Группировка данных для вывода на график (см. рис.):

$$\varepsilon_{otn_Σ} := \text{stack}(\varepsilon_{otn1_Σ}, \varepsilon_{otn2_Σ}, \varepsilon_{otn3_Σ}, \varepsilon_{otn4_Σ}, \varepsilon_{otn5_Σ})$$



Динамика изменения суммарной относительной ошибки моделирования процесса электропотребления объектов техноценоза: абсцисса – годы исследования; ордината – относительные единицы

Средние относительные ошибки, определяемые в процентах на основе относительных ошибок по всем объектам техноценоза:

$$\varepsilon_{sr_otn1} := \frac{\sum_{i=1}^n \left| \overrightarrow{(\varepsilon_{otn1}_i)} \cdot 100 \right|}{n}$$

$$\varepsilon_{sr_otn1} = 7.412$$

$$\varepsilon_{sr_otn2} := \frac{\sum_{i=1}^n \left| \overrightarrow{(\varepsilon_{otn2}_i)} \cdot 100 \right|}{n}$$

$$\varepsilon_{sr_otn2} = 9.919$$

$$\varepsilon_{sr_otn3} := \frac{\sum_{i=1}^n \left| \overrightarrow{(\varepsilon_{otn3}_i)} \cdot 100 \right|}{n}$$

$$\varepsilon_{sr_otn3} = 12.315$$

$$\varepsilon_{sr_otn4} := \frac{\sum_{i=1}^n \left| \overrightarrow{(\varepsilon_{otn4}_i)} \cdot 100 \right|}{n}$$

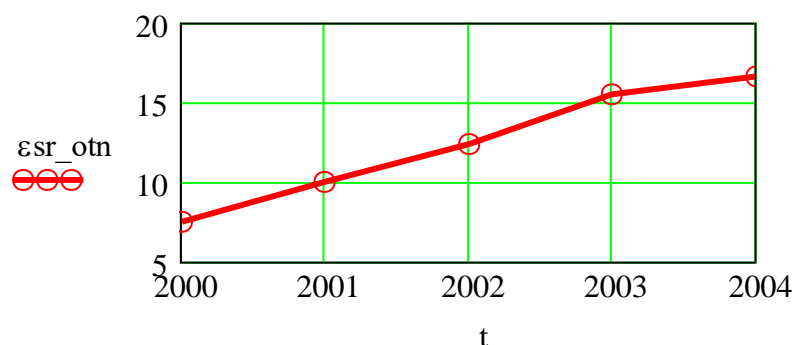
$$\varepsilon_{sr_otn4} = 15.476$$

$$\varepsilon_{sr_otn5} := \frac{\sum_{i=1}^n \left| \overrightarrow{(\varepsilon_{otn5}_i)} \cdot 100 \right|}{n}$$

$$\varepsilon_{sr_otn5} = 16.503$$

Группировка данных для вывода на график (см. рис.):

$$\varepsilon_{sr_otn} := \text{stack}(\varepsilon_{sr_otn1}, \varepsilon_{sr_otn2}, \varepsilon_{sr_otn3}, \varepsilon_{sr_otn4}, \varepsilon_{sr_otn5})$$



Динамика изменения средней относительной ошибки моделирования: абсцисса – годы исследования; ордината – относительные единицы

Относительные ошибки для техноценоза в целом:

$$OE1 := \frac{E1}{xf1} \cdot 100$$

$$OE1 = 0.559$$

$$OE2 := \frac{E2}{xf2} \cdot 100$$

$$OE2 = 0.904$$

$$OE3 := \frac{E3}{xf3} \cdot 100$$

$$OE3 = 1.212$$

$$OE4 := \frac{E4}{xf4} \cdot 100$$

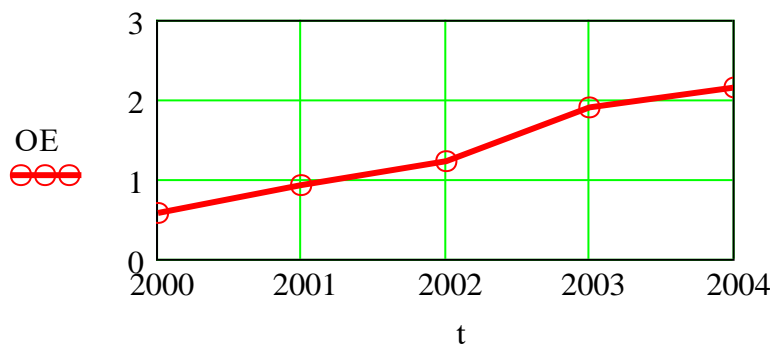
$$OE4 = 1.884$$

$$OE5 := \frac{E5}{xf5} \cdot 100$$

$$OE5 = 2.128$$

Группировка данных для вывода на график (см. рис.):

$$OE := \text{stack}(OE1, OE2, OE3, OE4, OE5)$$



Динамика изменения относительной ошибки для техноценоза в целом: абсцисса – годы исследования; ордината – относительные единицы

Главным показателем, характеризующим точность расчетов, является относительная ошибка моделирования для всего техноценоза. В данном случае ее значение растет с увеличением глубины прогноза, принимая максимальное значение на последнем году – 2,1 %. Подобная точность перспективного прогноза электропотребления техноценоза может быть признана хорошей. Как представляется, граничным значением относительной ошибки моделирования на пятом году является 4 – 5 %.

Анализ остатков моделирования (вторая процедура оценки адекватности модели)

При корректном выполнении моделирования остатки должны являться результатом случайного рассеяния, а не доминирующего действия какого-либо неконтролируемого и неуправляемого воздействия. Если это условие выполняется, то согласно центральной предельной теореме, при увеличении числа опытов, распределение остатков будет подчиняться закону Гаусса. В рамках второй процедуры оценки адекватности модели в первую очередь определяются стандартизированные остатки моделирования электропотребления по каждому прогнозируемому году:

$$se1 := \frac{\varepsilon 1 - \text{mean}(\varepsilon 1)}{\text{stdev}(\varepsilon 1)}$$

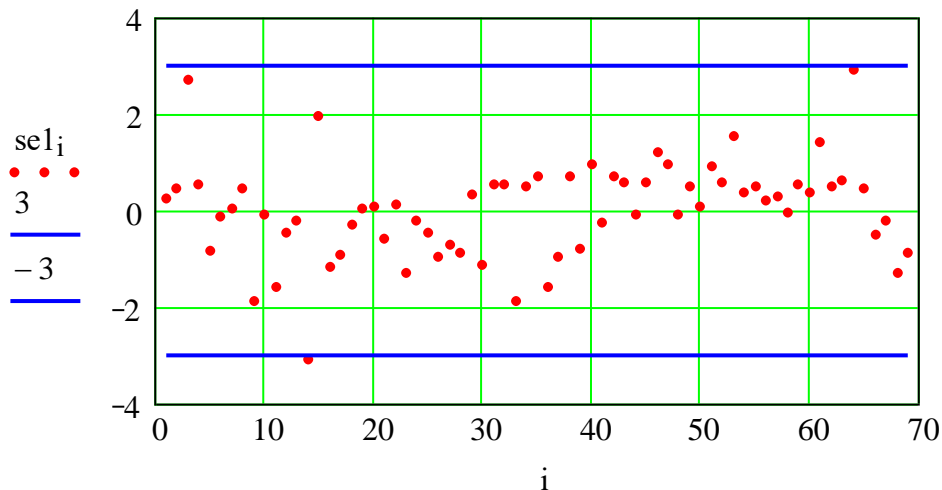
$$se3 := \frac{\varepsilon 3 - \text{mean}(\varepsilon 3)}{\text{stdev}(\varepsilon 3)}$$

$$se5 := \frac{\varepsilon 5 - \text{mean}(\varepsilon 5)}{\text{stdev}(\varepsilon 5)}$$

$$se2 := \frac{\varepsilon 2 - \text{mean}(\varepsilon 2)}{\text{stdev}(\varepsilon 2)}$$

$$se4 := \frac{\varepsilon 4 - \text{mean}(\varepsilon 4)}{\text{stdev}(\varepsilon 4)}$$

В качестве примера приведена визуализация распределения остатков моделирования электропотребления для 2000 года:



Распределение стандартизированных остатков моделирования электропотребления на 2000 год: абсцисса – номер объекта; ордината – относительные единицы; точки – стандартизированные остатки; сплошные линии – границы доверительного интервала

Таким образом, стандартизированные остатки моделирования лежат на прямой без существенных отклонений (выбросов) и представляют собой белый шум (это будет дополнительно проверено ниже).

Строгая проверка гипотезы о соответствии генеральной совокупности нормальному распределению осуществляется по известному статистическому критерию Пирсона. Здесь приведены расчеты для значений электропотребления объектов техноценоза в 2001 году.

Выполним переименование вектора стандартизированных остатков моделирования и сформируем соответствующий интервал:

$$Y := se2$$

$$m := \text{round}(\sqrt{\text{length}(Y)})$$

$$m = 8$$

$$l := 1..m$$

$$h := \frac{\max(Y) - \min(Y)}{m}$$

$$v_l := \min(Y) + h \cdot l$$

Гистограмма абсолютных частот:

$$H := \text{hist}(v, Y)$$

$$H_{m+1} := 0$$

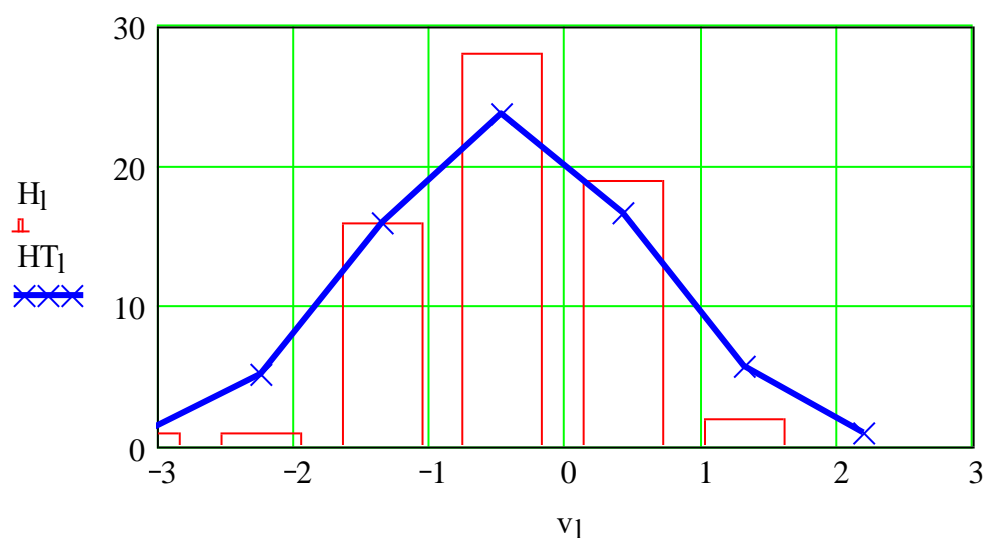
$$\sum_1 H_l = 67$$

$$v_{m+1} := v_m + h$$

Далее рассчитаем теоретические значения частоты попадания в интервал в предположении, что закон распределения нормальный, а также осуществим визуализацию результатов расчетов (см. рис. ниже):

$$HT_l := \text{length}(Y) \cdot \int_{v_l}^{v_{l+1}} \text{dnorm}(u, \text{mean}(Y), \text{stdev}(Y)) du$$

$$\sum_1 HT_l = 68.927$$



Соответствие распределения стандартизованных остатков нормальному закону: абсцисса – границы интервалов; ордината – частота попадания точек в интервал; гистограмма – эмпирические частоты; кривая линия – теоретические частоты

Функция суммы квадратов отклонений экспериментальных значений от теоретических (наблюдаемое значение критерия):

$$\chi^2 := \sum_1 \frac{(H_1 - HT_1)^2}{HT_1} \quad \chi^2 = 7.805$$

Критическая точка правосторонней критической области:

$$\chi^2_{kr} := qchisq(0.95, m - 3) \quad \chi^2_{kr} = 11.07$$

Гипотеза о распределении генеральной совокупности по нормальному закону принимается, если следующий критерий выполняется (как видно, в данном случае неравенство и критерий выполняются):

$$K := \text{if}(\chi^2 < \chi^2_{kr}, 1, 0) \quad K = 1$$

Для остальных лет получаем:

$$2000 \text{ год: } \chi^2 = 10.681$$

$$2002 \text{ год: } \chi^2 = 6.393$$

2003 год: $\chi^2 = 9.347$

2004 год: $\chi^2 = 10.819$

Таким образом, теоретические и эмпирические частоты различаются не значимо, следовательно остатки подчиняются нормальному закону.

Далее необходимо проверить однородность дисперсий остатков моделирования с помощью критерия Кохрена. Исправленные выборочные дисперсии стандартизированных остатков для всех прогнозных лет:

$$d1 := \frac{n}{n-1} \cdot \text{var}(se1)$$

$$d2 := \frac{n}{n-1} \cdot \text{var}(se2)$$

$$d3 := \frac{n}{n-1} \cdot \text{var}(se3)$$

$$d4 := \frac{n}{n-1} \cdot \text{var}(se4)$$

$$d5 := \frac{n}{n-1} \cdot \text{var}(se5)$$

$D := \text{stack}(d1, d2, d3, d4, d5)$

$k := 1.. \text{rows}(D)$

Фактическое значение критерия Кохрена:

$$Q := \frac{\max(D)}{\sum_k D_k}$$

$$Q = 0.2$$

Критическое значение критерия Кохрена:

$$Q_{kr} := 0.2754$$

Гипотеза об однородности дисперсий принимается, если выполняется следующий критерий (в данном случае – выполняется):

$$Kk := Q < Q_{kr}$$

$$Kk = 1$$

Таким образом, рассматриваемые выборки имеют однородные дисперсии, что является доказательством случайного рассеяния остатков.

Не менее важной является проверка на наличие сериальной корреляции остатков результатов моделирования по критерию Дарбина – Уотсона. Вычислим значения d-статистики для исследуемых лет:

$$de1 := \frac{\sum_{i=2}^n (se1_i - se1_{i-1})^2}{\sum_{i=1}^n (se1_i)^2} \quad de1 = 1.866$$

$$de2 := \frac{\sum_{i=2}^n (se2_i - se2_{i-1})^2}{\sum_{i=1}^n (se2_i)^2} \quad de2 = 2.011$$

$$de3 := \frac{\sum_{i=2}^n (se3_i - se3_{i-1})^2}{\sum_{i=1}^n (se3_i)^2} \quad de3 = 1.433$$

$$de4 := \frac{\sum_{i=2}^n (se4_i - se4_{i-1})^2}{\sum_{i=1}^n (se4_i)^2} \quad de4 = 1.449$$

$$de5 := \frac{\sum_{i=2}^n (se5_i - se5_{i-1})^2}{\sum_{i=1}^n (se5_i)^2} \quad de5 = 2.154$$

Группировка полученных результатов:

DE := stack(de1, de2, de3, de4, de5)

Оценивание двухстороннего критерия Дарбина – Уотсона с равными хвостами. Критические точки, характеризующие критическую область:

$$dL := 1.46$$

$$dU := 1.77$$

Для реализации критерия применяются две специальные подпрограммы (см. ниже). По результатам работы каждой из подпрограмм выводится вектор из пяти цифр, которые являются индикаторами d-статистики для соответствующих лет. Если хотя бы в одном из векторов для данного года индикатор равен «1», то нулевая гипотеза принимается.

$$\text{RES1}(DE, dL, dU) := \begin{array}{l} \text{for } i \in 1.. \text{rows}(DE) \\ K_i \leftarrow \begin{array}{l} 1 \text{ if } DE_i > dU \\ 0 \text{ if } DE_i < dL \\ -1 \text{ otherwise} \end{array} \\ K \end{array}$$

$$\text{RES1}(DE, dL, dU)^T = (1 \ 1 \ 0 \ 0 \ 1)$$

$$\text{RES2}(DE, dL, dU) := \begin{array}{l} \text{for } i \in 1.. \text{rows}(DE) \\ K_i \leftarrow \begin{array}{l} 1 \text{ if } 4 - DE_i > dU \\ 0 \text{ if } 4 - DE_i < dL \\ -1 \text{ otherwise} \end{array} \\ K \end{array}$$

$$\text{RES2}(DE, dL, dU)^T = (1 \ 1 \ 1 \ 1 \ 1)$$

Таким образом, для рассматриваемого примера, сравнение вычисленных значений d-статистики с критическими точками позволяет заключить, что d незначимо, и гипотеза об отсутствии сериальной корреляции остатков принимается для всех пяти лет моделирования.

Тест остатков моделирования на белый шум проводится с помощью критерия Бартлетта. При этом с помощью быстрого преобразования Фурье определяются выборочные значения кумулятивного спектра стандартизированных остатков (здесь приведен пример для 2000 г.). Вычислим длину вектора, зададим натуральный ряд и определим частоту спектра:

$$Sa := \text{cfft}(se1) \quad N := \text{last}(Sa) \quad k := 1..N \quad Pa := \overrightarrow{(Sa \cdot Sa)} \quad f_k := \frac{k}{n}$$

Теоретически нормированный накопленный спектр белого шума, верхняя и нижняя доверительные границы спектра:

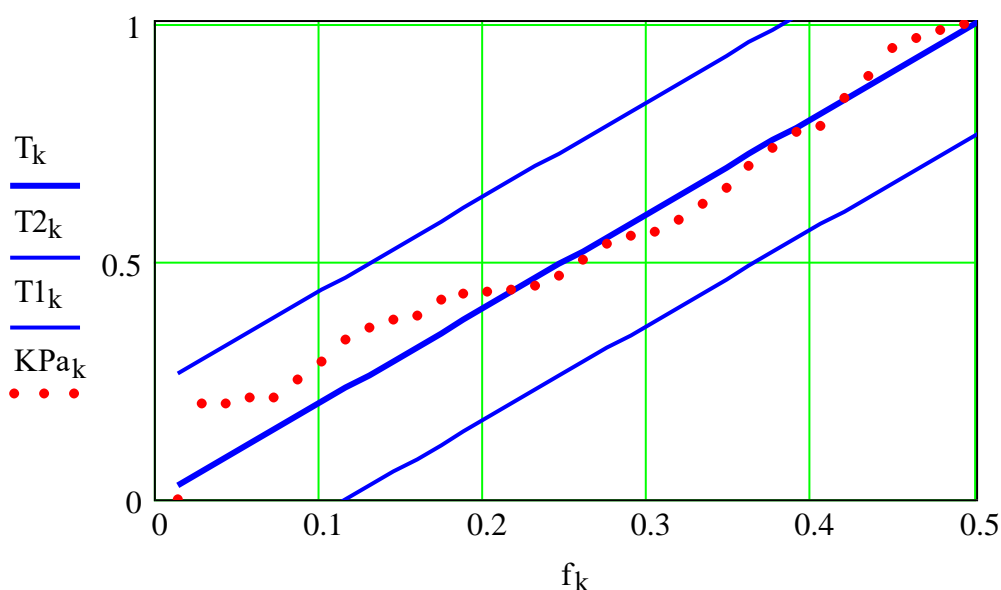
$$T_k := 2 \cdot f_k$$

$$T1_k := T_k + 1.36 \cdot \frac{\sqrt{2}}{\sqrt{n-2}}$$

$$T2_k := T_k - 1.36 \cdot \frac{\sqrt{2}}{\sqrt{n-2}}$$

Определим накопленный кумулятивный спектр (измерен в относительных единицах) и визуализируем результаты расчетов и сравнения:

$$KPa_k := \frac{2}{n \cdot \text{var}(se1)} \cdot \sum_{k=1}^k Pa_k$$



Спектр остатков с доверительными границами: абсцисса – частота спектра; ордината – накопленный спектр в относительных единицах; верхняя и нижняя линии – доверительные границы кумулятивного спектра стандартизованных остатков; средняя линия – теоретический нормированный кумулятивный спектр стандартизованных остатков; красные точки – кумулятивный спектр стандартизованных остатков

Таким образом, накопленный спектр стандартизованных остатков моделирования для 2000 года лежит внутри доверительных границ, т.е. является белым шумом, что подтверждает адекватность разработанной модели. Данная процедура должна выполняться для всех прогнозных лет.

Критерий Фишера позволяет проверить, является ли дисперсия, обусловленная регрессией, значимо больше дисперсии остатков. Он предполагает вычисление выборочных дисперсий остатков (абсолютных ошибок моделирования электропотребления) на исследуемые годы:

$$d\varepsilon_1 := \frac{n}{n-1} \cdot \text{var}(\varepsilon_1)$$

$$d\varepsilon_1 = 3.612 \times 10^7$$

$$d\varepsilon_2 := \frac{n}{n-1} \cdot \text{var}(\varepsilon_2)$$

$$d\varepsilon_2 = 8.344 \times 10^7$$

$$d\varepsilon_3 := \frac{n}{n-1} \cdot \text{var}(\varepsilon_3)$$

$$d\varepsilon_3 = 1.022 \times 10^8$$

$$d\varepsilon_4 := \frac{n}{n-1} \cdot \text{var}(\varepsilon_4)$$

$$d\varepsilon_4 = 4.601 \times 10^7$$

$$d\varepsilon_5 := \frac{n}{n-1} \cdot \text{var}(\varepsilon_5)$$

$$d\varepsilon_5 = 2.278 \times 10^8$$

Выборочные дисперсии регрессионной модели на исследуемые годы:

$$S_1 := \frac{1}{n-5} \cdot \sum_{i=1}^n (x_{1i} - \text{mean}(x_1))^2$$

$$S_1 = 3.253 \times 10^{11}$$

$$S_2 := \frac{1}{n-5} \cdot \sum_{i=1}^n (x_{2i} - \text{mean}(x_2))^2$$

$$S_2 = 2.963 \times 10^{11}$$

$$S_3 := \frac{1}{n-5} \cdot \sum_{i=1}^n (x_{3i} - \text{mean}(x_3))^2$$

$$S_3 = 3.139 \times 10^{11}$$

$$S_4 := \frac{1}{n-5} \cdot \sum_{i=1}^n (x_{4i} - \text{mean}(x_4))^2$$

$$S_4 = 3.624 \times 10^{11}$$

$$S_5 := \frac{1}{n-5} \cdot \sum_{i=1}^n (x_{5i} - \text{mean}(x_5))^2$$

$$S_5 = 3.632 \times 10^{11}$$

Критерий Фишера:

$$\begin{array}{ll}
 F1 := \frac{S1}{d\varepsilon1} & F1 = 9.006 \times 10^3 \\
 F2 := \frac{S2}{d\varepsilon2} & F2 = 3.551 \times 10^3 \\
 F3 := \frac{S3}{d\varepsilon3} & F3 = 3.071 \times 10^3 \\
 F4 := \frac{S4}{d\varepsilon4} & F4 = 7.878 \times 10^3 \\
 F5 := \frac{S5}{d\varepsilon5} & F5 = 1.594 \times 10^3
 \end{array}$$

$$F := \text{stack}(F1, F2, F3, F4, F5)$$

Критическое значение критерия Фишера:

$$F_{cr} := qF(0.95, n - 5, 1) \quad F_{cr} = 252.328$$

Сравнение вычисленного значения F-критерия с его критическим значением осуществляется автоматически при помощи специальной подпрограммы, по результатам работы которой выводится вектор, состоящий из пяти цифр-индикаторов. Значение индикатора, равное «1», соответствует выводу об адекватности модели в данном году, «0» – нет.

$$\text{RESULT}(F, F_{cr}) := \left| \begin{array}{l} \text{for } i \in 1.. \text{rows}(F) \\ K_i \leftarrow \left| \begin{array}{l} 1 \text{ if } F_i > F_{cr} \\ 0 \text{ if } F_i < F_{cr} \end{array} \right. \\ K \end{array} \right.$$

$$\text{RESULT}(F, F_{cr})^T = (1 \ 1 \ 1 \ 1 \ 1)$$

Следовательно, для всех лет моделирования процесса электропотребления дисперсия, обусловленная регрессией, значимо больше дисперсии остатков, что позволяет сформулировать положительный ответ относительно адекватности рассматриваемой модели по критерию Фишера.

Таким образом, оценка адекватности работы динамической адаптивной модели управления процессом электропотребления объектов исследуемого техноценоза показала следующее. Относительная ошибка моделирования для всей инфраструктуры, несколько возрастающая с увеличением глубины прогноза, на последнем (пятом) году не превышает 2,1 %. Следует отметить, что погрешность прогноза статической модели, описанной в ряде работ, уже на первом году моделирования достигает 2 %, а на последующих годах резко возрастает. Анализ остатков динамической модели электропотребления показал, что предъявляемые к ним требования выпол-

няются: 1) у стандартизированных остатков отсутствуют ярко выраженные выбросы, и они распределены по нормальному закону с однородной дисперсией; 2) отсутствуют сериальные корреляции остатков, которые являются белым шумом, а дисперсия, обусловленная регрессией, значимо больше дисперсии остатков. Это является подтверждением адекватности модели, которая может эффективно применяться для моделирования электропотребления объектов техноценоза на глубину до 5 – 7 лет.

[на содержание]

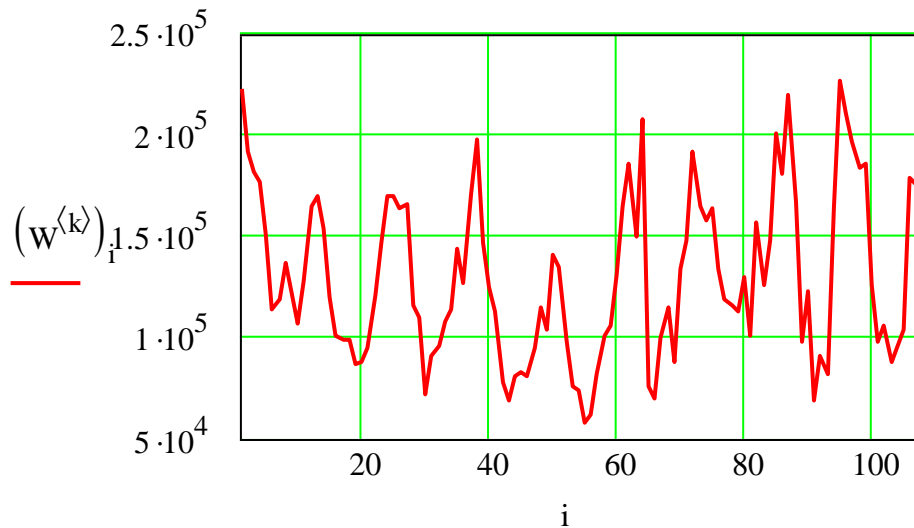
ПРОГНОЗИРОВАНИЕ ЭЛЕКТРОПОТРЕБЛЕНИЯ МЕТОДОМ АНАЛИЗА ГЛАВНЫХ КОМПОНЕНТ

В ряде работ показана возможность анализа и синтеза временных рядов с помощью метода анализа главных компонент (АГК), получившего в англоязычной версии название Singular Spectrum Analysis (SSA). Как показано ранее данный метод, предназначенный для исследования структуры временных рядов, совмещает в себе достоинства многих других методов, в частности, анализа Фурье и регрессионного анализа. Одновременно он отличается простотой и наглядностью в управлении. Базовый вариант метода состоит в преобразовании одномерного ряда в многомерный с помощью однопараметрической сдвиговой процедуры, исследовании полученной многомерной траектории с помощью анализа главных компонент (сингулярного разложения) и восстановлении (аппроксимации) ряда по выбранным главным компонентам. Таким образом, результатом применения метода является разложение временного ряда на простые компоненты: медленные тренды, сезонные и другие периодические или колебательные составляющие, а также шумовые компоненты. Полученное разложение может служить основой корректного прогнозирования, как самого параметрического ряда, так и его отдельных составляющих.

Используем этот метод для прогнозирования электропотребления объектов техноценоза. После сбора информации о месячном электропотреблении (в кВт·ч) создается электронная база данных. Рекомендуется базу данных формировать в виде файла Microsoft Excel, в котором должны содержаться только числовые значения, при этом столбцы должны соответствовать объектам техноценоза, а строки – месяцам. Перед началом обработки данные необходимо подвергнуть верификации и сохранить под именем «data_lu.xls» в директории «с:\mathcad_dat», которая должна быть заблаговременно создана в корневом каталоге диска «с:\». Далее, уже программными средствами, осуществляется импорт исходных данных из внешнего файла «data_lu.xls» в тело mathcad-программы. В настоящем примере используются десятилетние данные по электропотреблению реально существующего в Калининградской области техноценоза, состоящего из рассредоточенных по всей территории региона 69 объектов.

ние параметра k прямо в теле программы позволяет произвести мгновенный пересчет применительно к любому интересующему объекту.

```
i := 1..rows(W)
k := 1
```



Случайный процесс электропотребления k -го объекта: абсцисса – месяцы; ордината – электропотребление, кВт·ч

Развертка одномерного ряда в многомерный

Подпрограмма $P1(x, \tau)$ преобразует отдельную реализацию значений X длиной n в матрицу с параметром τ . Определим параметр Φ , который задает количество членов исходной реализации, составляющих строку конечной матрицы (данный параметр равен числу ее столбцов).

```
 $\tau := 32$ 
 $\Phi := \text{rows}(W) - \tau + 1$ 
 $\Phi = 77$ 
```

```
 $P1(x, \tau) := \left| \begin{array}{l} N \leftarrow \text{length}(x) \\ n \leftarrow N - \tau \\ \text{for } k \in 1.. \tau \\ \quad c^{(k)} \leftarrow \text{submatrix}[x, k, (n + k), 1, 1] \\ \quad c^T \end{array} \right.$ 
```

Подпрограмма SSA(W) осуществляет преобразование по всей матрице данных по электропотреблению. Результатом выполнения программы является вектор, элементы которого – вложенные матрицы.

$$SSA(W) := \begin{cases} \text{for } i \in 1.. \text{cols}(W) \\ \quad E_i \leftarrow P1(W^{(i)}, \tau) \\ E \end{cases}$$

$$SSA(W) =$$

| | |
|---|---------|
| | 1 |
| 1 | [32,77] |
| 2 | [32,77] |
| 3 | [32,77] |
| 4 | [32,77] |
| 5 | [32,77] |
| 6 | [32,77] |
| 7 | [32,77] |
| 8 | [32,77] |

Возможен просмотр преобразованной матрицы для k-го объекта техноценоза (размер выводимой на экран области данных регулируется).

$$SSA(W)_k =$$

| | | | |
|---|--------------------|--------------------|--------------------|
| | 1 | 2 | 3 |
| 1 | $2.228 \cdot 10^5$ | $1.912 \cdot 10^5$ | $1.82 \cdot 10^5$ |
| 2 | $1.912 \cdot 10^5$ | $1.82 \cdot 10^5$ | $1.761 \cdot 10^5$ |
| 3 | $1.82 \cdot 10^5$ | $1.761 \cdot 10^5$ | $1.505 \cdot 10^5$ |
| 4 | $1.761 \cdot 10^5$ | $1.505 \cdot 10^5$ | $1.134 \cdot 10^5$ |
| 5 | $1.505 \cdot 10^5$ | $1.134 \cdot 10^5$ | $1.189 \cdot 10^5$ |
| 6 | $1.134 \cdot 10^5$ | $1.189 \cdot 10^5$ | $1.364 \cdot 10^5$ |
| 7 | $1.189 \cdot 10^5$ | $1.364 \cdot 10^5$ | $1.211 \cdot 10^5$ |
| 8 | $1.364 \cdot 10^5$ | $1.211 \cdot 10^5$ | $1.062 \cdot 10^5$ |
| 9 | $1.211 \cdot 10^5$ | $1.062 \cdot 10^5$ | $1.271 \cdot 10^5$ |

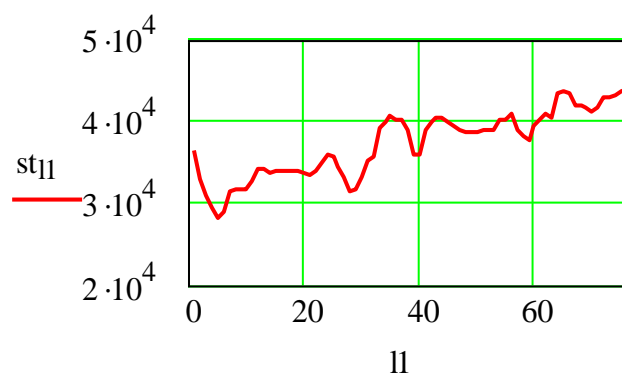
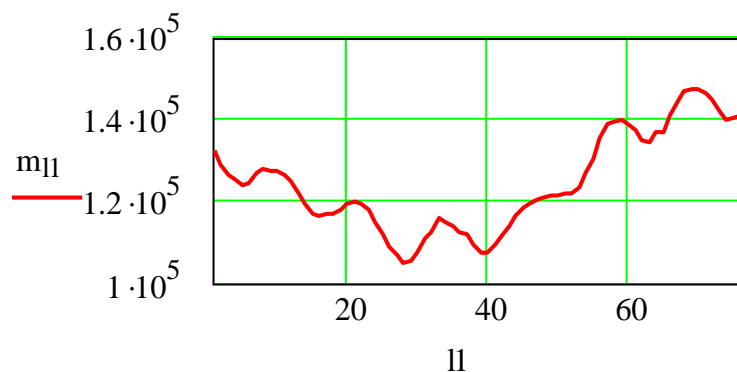
$$AA := SSA(W)$$

Проанализируем изменение выборочных средних и средних квадратичных отклонений (СКО) в зависимости от номера столбца преобразованной матрицы k-го объекта техноценоза (см. рис. ниже).

$$l1 := 1.. \text{cols}(AA_k)$$

$$m_{l1} := \text{mean} \left[(AA_k)^{\langle l1 \rangle} \right]$$

$$st_{l1} := \text{stdev} \left[(AA_k)^{\langle l1 \rangle} \right]$$



Динамика изменения выборочного среднего и СКО для k -го объекта техноценоза: абсцисса – месяцы; ордината – СКО, кВт·ч

Анализ ковариационной матрицы для каждого объекта

Зададим натуральный ряд, длина которого соответствует количеству объектов исследуемого техноценоза, и сформируем матрицу:

$j := 1.. \text{length}(AA)$

$$CX_j := \frac{1}{\text{rows}(AA_j)} \cdot AA_j \cdot AA_j^T$$

Выведем вектор, элементы которого представляют собой ковариационные матрицы, сформированные на предыдущем шаге.

| | |
|---|---------|
| | 1 |
| 1 | [32,32] |
| 2 | [32,32] |
| 3 | [32,32] |
| 4 | [32,32] |
| 5 | [32,32] |
| 6 | [32,32] |

$CX =$

$$CX_k =$$

| | 1 | 2 | 3 |
|---|-----------------------|-----------------------|-----------------------|
| 1 | $4.101 \cdot 10^{10}$ | $3.95 \cdot 10^{10}$ | $3.828 \cdot 10^{10}$ |
| 2 | $3.95 \cdot 10^{10}$ | $3.987 \cdot 10^{10}$ | $3.857 \cdot 10^{10}$ |
| 3 | $3.828 \cdot 10^{10}$ | $3.857 \cdot 10^{10}$ | $3.912 \cdot 10^{10}$ |
| 4 | $3.711 \cdot 10^{10}$ | $3.748 \cdot 10^{10}$ | $3.794 \cdot 10^{10}$ |
| 5 | $3.586 \cdot 10^{10}$ | $3.624 \cdot 10^{10}$ | $3.677 \cdot 10^{10}$ |
| 6 | $3.502 \cdot 10^{10}$ | $3.538 \cdot 10^{10}$ | $3.589 \cdot 10^{10}$ |
| 7 | $3.499 \cdot 10^{10}$ | $3.469 \cdot 10^{10}$ | $3.514 \cdot 10^{10}$ |
| 8 | $3.524 \cdot 10^{10}$ | $3.469 \cdot 10^{10}$ | $3.449 \cdot 10^{10}$ |

Определим собственные числа ковариационной матрицы и зададим вектор ее собственных чисел, отсортированных по убыванию.

$$VX_j := \text{reverse}(\text{sort}(\text{eigenvals}(CX_j)))$$

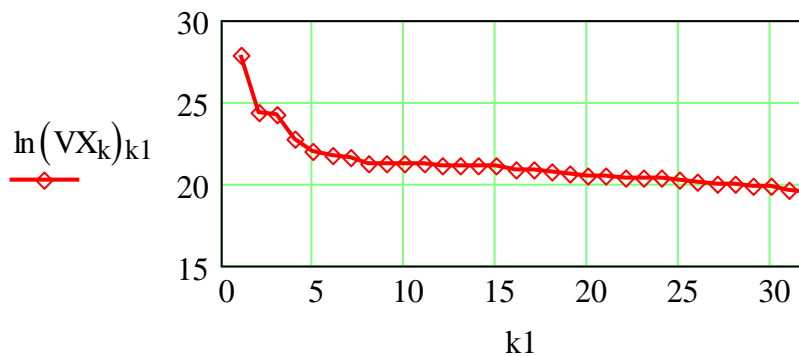
$$VX =$$

| | 1 |
|---|--------|
| 1 | [32,1] |
| 2 | [32,1] |
| 3 | [32,1] |
| 4 | [32,1] |
| 5 | [32,1] |
| 6 | [32,1] |
| 7 | [32,1] |
| 8 | [32,1] |
| 9 | [32,1] |

$$VX_k =$$

| | 1 |
|---|-----------------------|
| 1 | $1.205 \cdot 10^{12}$ |
| 2 | $3.682 \cdot 10^{10}$ |
| 3 | $3.308 \cdot 10^{10}$ |
| 4 | $7.096 \cdot 10^9$ |
| 5 | $3.317 \cdot 10^9$ |
| 6 | $2.599 \cdot 10^9$ |
| 7 | $2.42 \cdot 10^9$ |

$$k1 := 1.. \text{length}(VX_k)$$



Кривая собственных чисел k -го объекта: абсцисса – номера собственных чисел; ордината – логарифмы собственных чисел

Визуальный анализ рисунка показывает, что для k-го объекта техноценоза первые четыре собственных числа имеют наибольшее значение, а значит, способны «объяснить» основную долю дисперсии.

Определим количество собственных чисел, «объясняющих» заданную долю дисперсии. Предлагаемый алгоритм использует одно из свойств собственных чисел: чем больше доля собственного числа в сумме элементов вектора собственных чисел, тем большую долю информации «способна объяснить» соответствующая ему главная компонента. Для выполнения подпрограмм необходимо задать долю дисперсии (в процентах), которую необходимо «объяснить» (в данном случае задается 99 %).

m := 99

```

NS(R,m) :=
  S ← 0
  SO ← ∑ R
  for i ∈ 1..length(R)
    S ← S +  $\frac{R_i}{SO} \cdot 100$ 
    break if S ≥ m
    k ← i
  k
    
```

```

D(VX,m) :=
  for i ∈ 1..length(VX)
    a ← VXi
    di ← NS(a,m)
  d
    
```

Получаем вектор, элементы которого представляют собой количество собственных чисел, объясняющих 99 % дисперсии параметра электропотребления k-го объекта исследуемого техноценоза.

Q := D(VX,m)

$$Q^T =$$

| | | | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | 13 | 25 | 20 | 22 | 18 | 21 | 20 | 18 | 15 | 14 | 19 | 16 |

Найдем собственные вектора.

```

TT(CX, VX) :=
  for j ∈ 1..length(VX)
  |
  |   Sch ← VXj
  |   X ← CXj
  |   for i ∈ 1..length(Sch)
  |     a⟨i⟩ ← eigenvec(X, Schj)
  |   sj ← a
  |
  s
    
```

Сформируем вектор, элементы которого представляют собой матрицы собственных векторов k-го объекта исследуемого техноценоза.

AQ := TT(CX, VX)

| | |
|---|---------|
| | 1 |
| 1 | [32,32] |
| 2 | [32,32] |
| 3 | [32,32] |
| 4 | [32,32] |
| 5 | [32,32] |
| 6 | [32,32] |
| 7 | [32,32] |

| | | |
|----|-------|--------|
| | 1 | 2 |
| 1 | 0.176 | 0.232 |
| 2 | 0.175 | 0.225 |
| 3 | 0.173 | 0.163 |
| 4 | 0.173 | 0.057 |
| 5 | 0.171 | -0.056 |
| 6 | 0.171 | -0.157 |
| 7 | 0.171 | -0.202 |
| 8 | 0.172 | -0.197 |
| 9 | 0.173 | -0.146 |
| 10 | 0.174 | -0.049 |

Прогнозирование электропотребления объектов техноценоза методом АГК

Будем полагать, что структура ряда для зарезервированных и обработанных данных остается неизменной. Следовательно, ковариационная матрица, собственные числа и собственные вектора остаются неизменными также и для второй части временного ряда электропотребления. С учетом этого допущения составим подпрограмму прогнозирования временно-

го ряда на основе следующих входных аргументов: AQ – матрица собственных векторов, x – обрабатываемая часть ряда, n – горизонт прогноза, τ – параметр преобразования ряда в матрицу, L – задаваемое число собственных векторов. Если число собственных векторов максимальное, т.е. равно параметру τ , то на каждом шаге прогноза приходится решать системы линейных уравнений, в которых число неизвестных в точности равно числу уравнений. Как известно, эти системы имеют единственное решение. Если число учитываемых собственных векторов меньше параметра τ , то системы линейных уравнений становятся избыточными, и для их решения следует применять метод наименьших квадратов. В любом случае свободными членами систем являются члены ряда, дополняемые прогнозируемыми значениями. В связи с чем необходимо на каждом шаге формировать вектор свободных членов из ряда переменной длины. Для этого составим вспомогательную подпрограмму R , аргументами которой являются: обрабатываемая часть ряда x , параметр преобразования ряда τ , а также номер шага реализуемого прогноза m .

$$R(x, \tau, m) := \left| \begin{array}{l} N \leftarrow \text{length}(x) \\ \text{for } i \in 1.. \tau - 2 - m + 1 \\ \quad y_i \leftarrow x_{N - \tau + 1 + m + i} \\ y \end{array} \right.$$

Подпрограмма PR служит для прогнозирования отдельной реализации процесса электропотребления объекта техноценоза.

$$PR(T, u, n, \tau, L) := \left| \begin{array}{l} TR \leftarrow \text{submatrix}(T, 1, \tau - 1, 1, L - 1) \\ T1R \leftarrow \text{submatrix}(T, \tau, \tau, 1, L - 1) \\ k \leftarrow 1 \\ xp \leftarrow R(u, \tau, 0) \\ \text{while } k \leq n \\ \quad \left| \begin{array}{l} L1 \leftarrow (TR^T \cdot TR)^{-1} \cdot (TR^T \cdot xp) \\ xr_k \leftarrow |T1R \cdot L1| \\ k \leftarrow k + 1 \\ x1 \leftarrow R(u, \tau, k - 1) \\ xp \leftarrow \text{stack}(x1, xp) \end{array} \right. \\ xr \end{array} \right.$$

В свою очередь, подпрограмма Prognoz служит для прогнозирования по всей матрице данных (аргументами которой являются параметр преобразования ряда τ и матрица собственных векторов AQ).

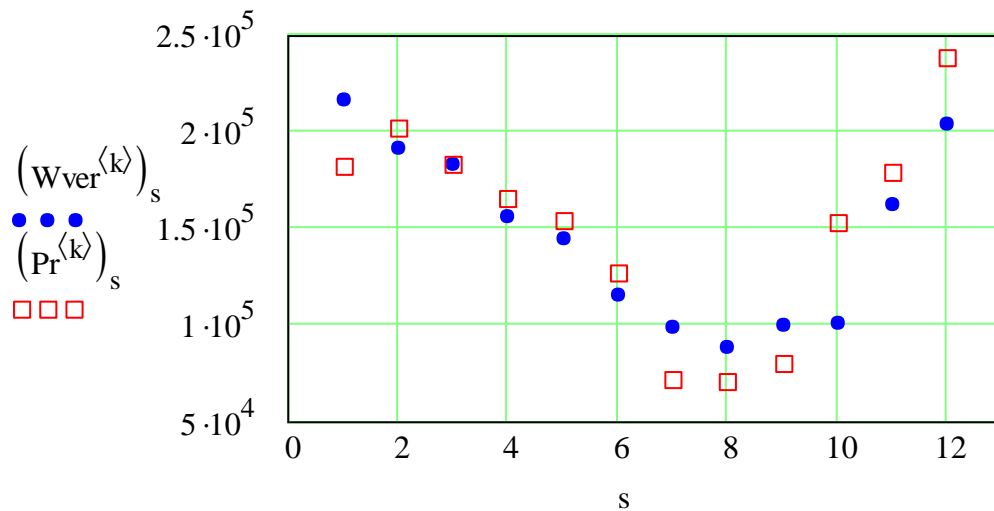
```

Prognoz(W, AQ, Q, n,  $\tau$ ) :=
  for j  $\in$  1..length(AQ)
    x  $\leftarrow$  W $\langle j \rangle$ 
    T  $\leftarrow$  AQj
    L  $\leftarrow$  Qj
    d  $\leftarrow$  PR(T, x, n,  $\tau$ , L)
    a $\langle j \rangle$   $\leftarrow$  d
  a
    
```

Визуализируем результаты расчетов для глубины прогноза, задаваемого параметром s (в данном случае s – месяцы).

```

Pr := Prognoz(W, AQ, Q, n,  $\tau$ )
s := 1..n
    
```



Сравнение прогнозных значений с реальными данными для k-го объекта: абсцисса – месяцы; ордината – электропотребление, кВт·ч; точки – реальные данные; квадраты – прогнозные значения

Вычислим относительную среднеквадратическую ошибку прогноза электропотребления по каждому объекту техноценоза.

$$v1_j := \frac{\sqrt{\frac{\sum_s \left[\left(W_{ver}^{(j)} \right)_s - \left(Pr^{(j)} \right)_s \right]^2}{n}}}{\text{mean} \left(W_{ver}^{(j)} \right)}$$

$$v1 =$$

| | 1 |
|---|--------|
| 1 | 16.665 |
| 2 | 50.826 |
| 3 | 10.421 |
| 4 | 50.482 |
| 5 | 42.438 |
| 6 | 34.659 |
| 7 | 44.049 |
| 8 | 38.7 |
| 9 | 19.239 |

%

Относительная годовая ошибка прогноза электропотребления по каждому объекту исследуемого техноценоза:

$$\text{error1}_j := \left| \frac{\sum_s \left(W_{ver}^{(j)} \right)_s - \sum_s \left(Pr^{(j)} \right)_s}{\sum_s \left(W_{ver}^{(j)} \right)_s} \right| \cdot 100$$

$$\text{error1} =$$

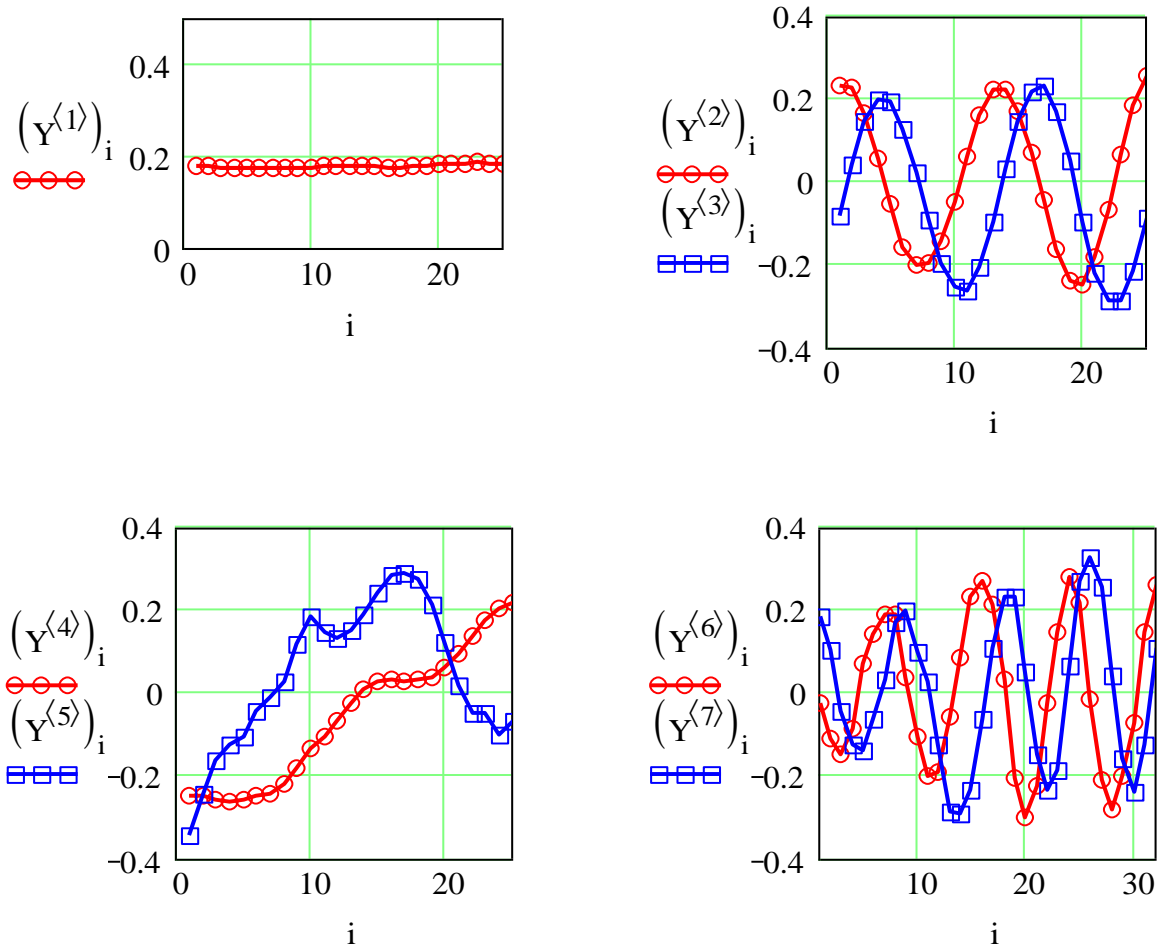
| | 1 |
|----|--------|
| 1 | 2.581 |
| 2 | 6.271 |
| 3 | 2.068 |
| 4 | 3.967 |
| 5 | 7.083 |
| 6 | 27.437 |
| 7 | 10.897 |
| 8 | 0.409 |
| 9 | 7.607 |
| 10 | 16.455 |

Анализируя ошибки прогноза электропотребления, можно сделать вывод о том, насколько структура обработанных данных соответствует ранее зарезервированным. При значительных ошибках, как правило, превышающих 10 – 15 %, объекты следует подвергать дополнительному обследованию с целью выявления причин структурных изменений. Прогнозирование электропотребления таких объектов должно производиться другими методами (при этом, достаточно хорошие результаты дают, так называемые, системные Z-методы или синтетические GZ-методы).

Анализ собственных векторов временного ряда

При необходимости можно осуществить по объектный анализ собственных векторов с целью выявления информации о структуре процесса электропотребления и составляющих его факторных слагаемых. Для этого выполним анализ собственных векторов (на примере k-го объекта). На рисунке 6 приведены графики, иллюстрирующие главные компоненты (номера векторов показаны по оси ординат графиков).

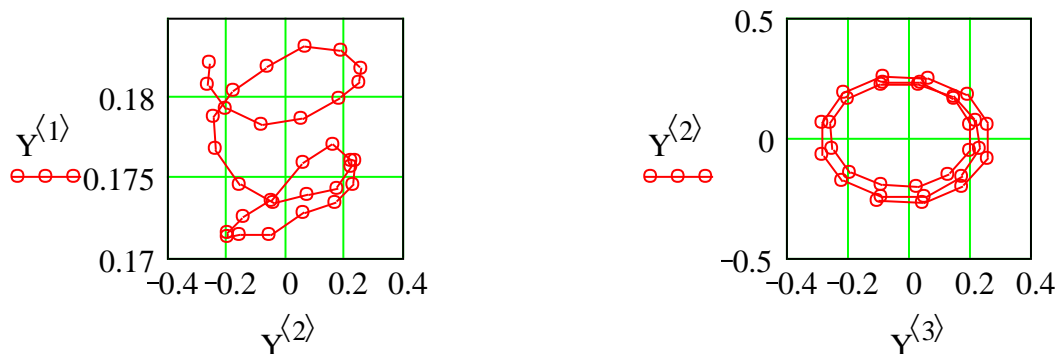
$$Y := AQ_k$$

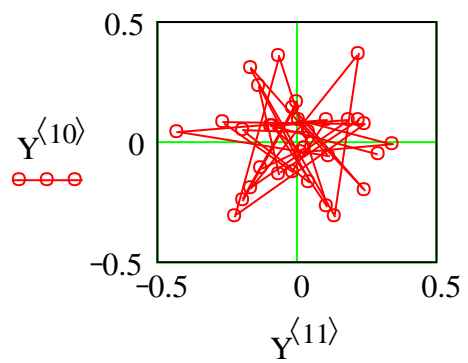
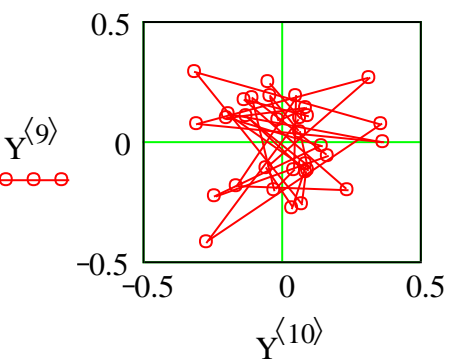
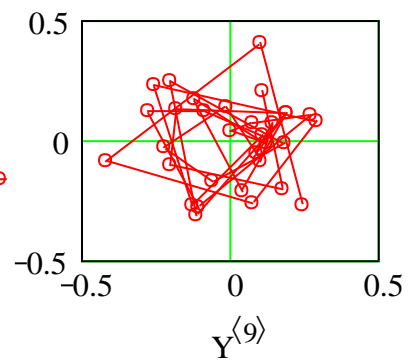
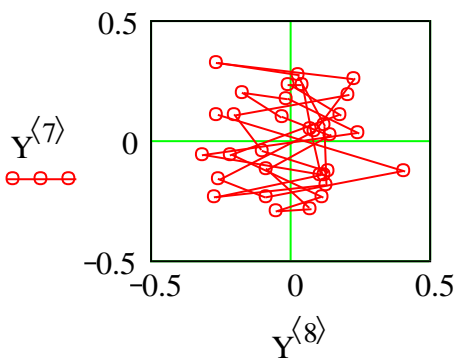
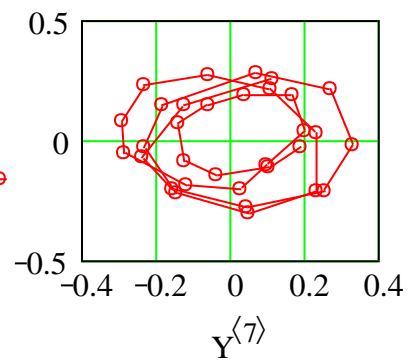
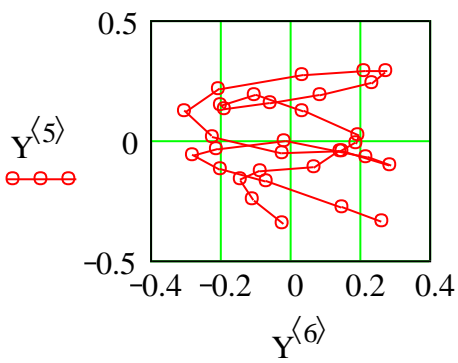
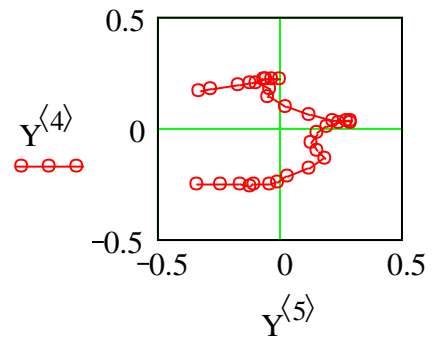
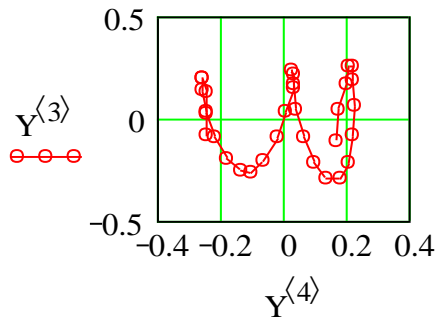


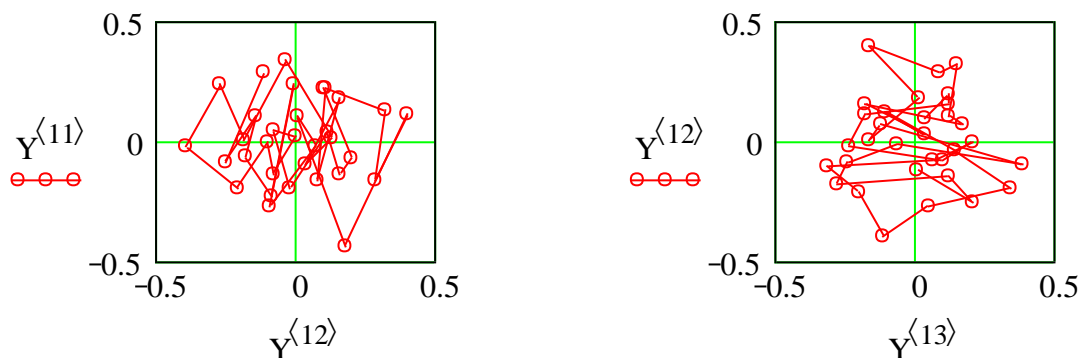
Главные компоненты временного ряда электропотребления: абсцисса – месяцы; ордината – собственные вектора

Анализ рисунка показывает, что первый собственный вектор является составляющей тренда и уровня процесса, второй и третий, а также шестой и седьмой – относятся к низкочастотной составляющей. Однако четвертый и пятый вектора полезной информации о процессе не несут.

Ниже на рисунке показаны двумерные проекции собственных векторов k-го объекта (номера проецируемых векторов показаны по осям).







Проекции собственных векторов временного ряда электропотребления k-го объекта: номера проецируемых векторов показаны по осям

Анализ рисунка позволяет по k-му объекту сделать следующие выводы. Второй и третий собственные вектора отражают годовые циклы, шестой и седьмой – восьмимесячные, восьмой и девятый – трехмесячные циклы. Остальные двумерные проекции собственных векторов полезной информации о процессе электропотребления не несут. Также можно сделать общие выводы по k-му объекту: сравнительно невысокая относительная годовая ошибка прогноза электропотребления (2,581 %) в целом свидетельствует об адекватности выполненных расчетов; можно предположить, что за исследуемый период отсутствовали существенные структурные изменения внутри самого объекта; тенденция на увеличение электропотребления отсутствует; в процессе функционирования объекта достаточно ярко выражены годовые, восьмимесячные и квартальные циклы электропотребления, которые необходимо учитывать при планировании ремонта и технического обслуживания оборудования.

[на содержание]

ОБРАБОТКА РАНГОВОЙ ПАРАМЕТРИЧЕСКОЙ ПОВЕРХНОСТИ МЕТОДОМ «SINGULAR SPECTRUM ANALYSIS»

Для эффективной реализации тонких процедур рангового анализа техноценоза (дифлекс-, GZ- и ASR-анализа), подробно описанных в ряде работ, зачастую возникает необходимость исследования ранговой параметрической поверхности, которая представляет собой двумерное поле рангового параметрического распределения во времени. Очевидно, что на техноценоз в процессе функционирования оказывают влияние несистематические, случайные факторы, искажающие форму ранговой поверхности. Такие искажения могут в значительной степени снизить достоверность тонких процедур рангового анализа. Процедура выделения и устранения несистематических компонент осуществляется при помощи метода Singular Spectrum Analysis (SSA), обладающего простотой и наглядностью.

Алгоритм метода SSA для двумерного поля состоит из следующих шагов: преобразования исходной матрицы данных с помощью сдвиговой процедуры в траекторную матрицу; дальнейшего ее исследования с помощью анализа главных компонент; восстановления исходной матрицы по выбранным главным компонентам. В конечном итоге, результатом применения метода является разложение матрицы (двумерного поля) на составляющие: систематические (определяемые закономерностью развития процесса во времени) и несистематические («шум») с последующим обязательным удалением «шумовой» составляющей.

Используем этот метод для рангового анализа электропотребления объектов техноценоза. После сбора информации о годовом электропотреблении (в кВт·ч) создается электронная база данных. Рекомендуется базу данных формировать в виде файла Microsoft Excel, в котором должны содержаться только числовые значения, при этом столбцы должны соответствовать объектам техноценоза, а строки – годам процесса электропотребления. Перед началом программной обработки данные следует подвергнуть процедуре верификации и сохранить под именем «data_d.xls» в директории «с:\mathcad_dat», которая должна быть заблаговременно создана в корневом каталоге диска «с:\». Далее, уже программными средствами, осуществляется импорт данных из файла «data_d.xls» в тело mathcad-программы. Здесь используются десятилетние данные по электропотреблению реально существующего в Калининградской области техноценоза, состоящего из рассредоточенных по территории 69 объектов.

Подготовка данных

С целью упорядочения индексации в матрицах и векторах, используемых в программе, зададим начало отсчета и импортируем данные.

ORIGIN := 1

V :=



C:\mathcad_dat\data_d.xls

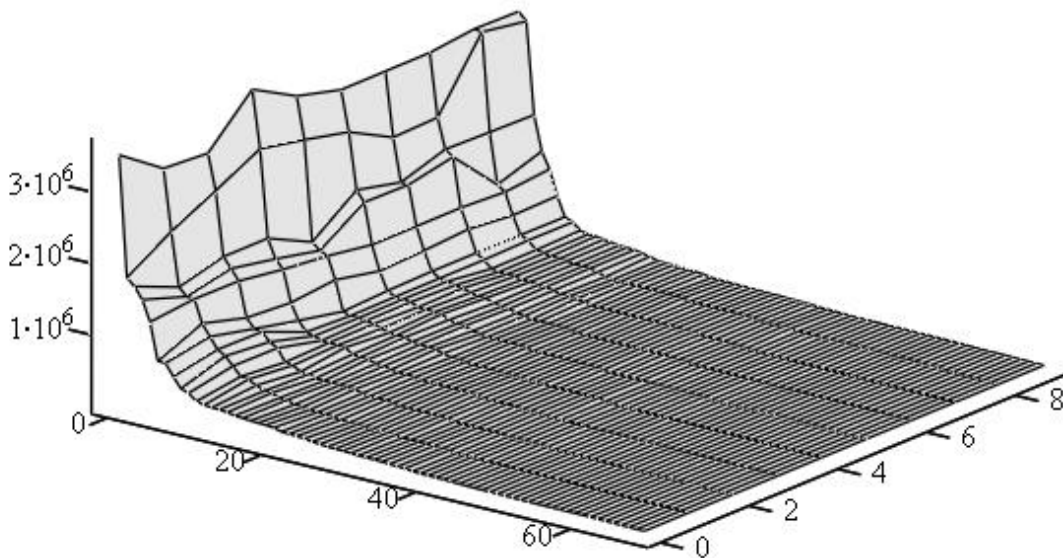
При необходимости можно осуществить просмотр данных по электропотреблению непосредственно в теле программы.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 1 | $1.81 \cdot 10^6$ | $1.268 \cdot 10^6$ | $2.328 \cdot 10^5$ | $5.817 \cdot 10^5$ | $4.386 \cdot 10^5$ | $7.438 \cdot 10^5$ | $2.891 \cdot 10^4$ |
| 2 | $1.447 \cdot 10^6$ | $1.332 \cdot 10^6$ | $8.567 \cdot 10^5$ | $5.244 \cdot 10^5$ | $4.027 \cdot 10^5$ | $6.371 \cdot 10^5$ | $2.636 \cdot 10^4$ |
| 3 | $1.474 \cdot 10^6$ | $1.613 \cdot 10^6$ | $8.373 \cdot 10^5$ | $5.784 \cdot 10^5$ | $5.179 \cdot 10^5$ | $7.094 \cdot 10^5$ | $2.724 \cdot 10^4$ |
| 4 | $1.351 \cdot 10^6$ | $1.614 \cdot 10^6$ | $3.494 \cdot 10^5$ | $4.025 \cdot 10^5$ | $5.673 \cdot 10^5$ | $3.652 \cdot 10^5$ | $3.26 \cdot 10^4$ |
| 5 | $1.162 \cdot 10^6$ | $1.304 \cdot 10^6$ | $3.971 \cdot 10^5$ | $3.274 \cdot 10^5$ | $4.366 \cdot 10^5$ | $8.78 \cdot 10^4$ | $2.701 \cdot 10^4$ |

Построим табулированные ранговые параметрические распределения техноценоза по электропотреблению (в матричной форме).

$$\text{Zipf}(W1) := \left| \begin{array}{l} \text{for } i \in 1.. \text{cols}(W1) \\ \quad \left| \begin{array}{l} c \leftarrow \text{sort}(W1^{\langle i \rangle}) \\ b^{\langle i \rangle} \leftarrow \text{reverse}(c) \end{array} \right. \\ b \end{array} \right. \quad V1 := \text{Zipf}(V^T)$$

Таким образом, столбец матрицы $V1$ представляет собой ранговое параметрическое распределение исследуемого техноценоза на определенном временном интервале (году). Визуализируем данные.



$V1$

Трехмерная ранговая параметрическая поверхность техноценоза: $V1$ – исходная эмпирическая матрица данных; абсцисса – ранг объекта; ордината – временной интервал (год); аппликата – электропотребление, кВт·ч

Преобразование матрицы данных с помощью сдвиговой процедуры в траекторную матрицу

Процесс преобразования матрицы ранговых параметрических распределений техноценоза в траекторную матрицу при помощи матрицы-окна определенного размера подробно описан в ряде работ. Подпрограмма GS предназначена для преобразования матрицы-окна в вектор.

$$GS(a) := \left| \begin{array}{l} c \leftarrow a^T \\ \text{for } i \in 1.. \text{cols}(c) \\ \left| \begin{array}{l} d \leftarrow c^{(i)} \text{ if } i = 1 \\ d \leftarrow \text{stack}(d, c^{(i)}) \text{ otherwise} \end{array} \right. \\ d \end{array} \right.$$

Для того чтобы можно было в дальнейшем восстановить исходную матрицу, создадим специальную матрицу индексов. При этом матрица индексов должна совпадать по размеру с исходной матрицей ранговых параметрических распределений техноценоза по электропотреблению.

$$VAD(a) := \left| \begin{array}{l} \text{for } i \in 1.. \text{rows}(a) \\ \text{for } j \in 1.. \text{cols}(a) \\ c_{i,j} \leftarrow \begin{pmatrix} i \\ j \end{pmatrix} \\ c \end{array} \right. \quad A := VAD(V1)$$

Зададим размер матрицы-окна:

$$\tau1 := 40 \qquad \tau2 := 5$$

$$d(V, \tau1, \tau2) := \left| \begin{array}{l} k \leftarrow 1 \\ \text{for } i \in 1.. \text{rows}(V) - \tau1 + 1 \\ \text{for } j \in 1.. \text{cols}(V) - \tau2 + 1 \\ \left| \begin{array}{l} s \leftarrow \text{submatrix}(V, i, i + \tau1 - 1, j, j + \tau2 - 1) \\ f^{(k)} \leftarrow GS(s) \\ k \leftarrow k + 1 \end{array} \right. \\ f \end{array} \right.$$

Далее преобразуем матрицу ранговых параметрических распределений и матрицу индексов с помощью подпрограммы, приведенной выше.

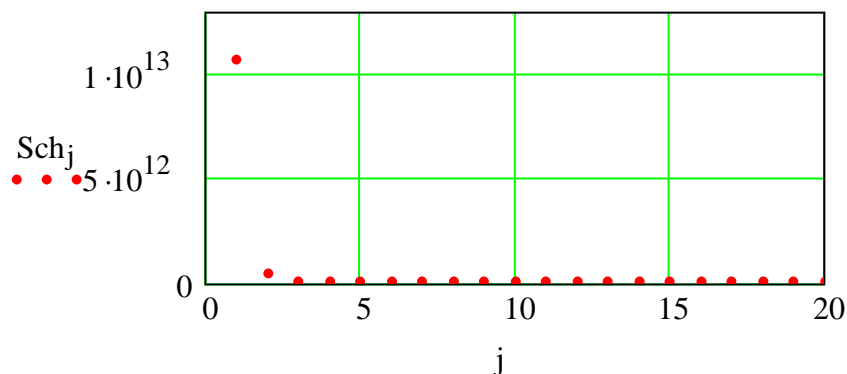
$$X := d(V1, \tau1, \tau2) \qquad A := d(A, \tau1, \tau2)$$

Исследование траекторной матрицы с помощью анализа главных компонент

В первую очередь найдем ковариационную матрицу и ее собственные числа. Далее визуализируем результаты расчетов.

$$CX := \frac{X \cdot X^T}{\text{rows}(X)}$$

Sch := reverse(sort(eigenvals(CX))) j := 1 .. length(Sch)



Собственные числа ковариационной матрицы: абсцисса – порядковый номер числа; ордината – значение собственного числа

Определим количество собственных чисел, исчерпывающим образом объясняющих заранее заданную долю дисперсии.

m := 98.89

```
NS(R, m) :=
  S ← 0
  SO ← ∑ R
  for i ∈ 1 .. length(R)
    S ← S +  $\frac{R_i}{SO} \cdot 100$ 
    break if S ≥ m
  k ← i
  k
```

Таким образом, для рассматриваемого здесь примера количество собственных чисел, объясняющих 98,89 % дисперсии, составляет:

$$NS(Sch, m) = 3$$

$$i := 1..3$$

Найдем собственные вектора по определенному выше количеству собственных чисел. Процедуру нахождения собственных векторов можно рассматривать как задание нового многомерного базиса, на направлениях орт которого дисперсия зафиксирована как максимальная.

$$Svec^{(i)} := \text{eigenvec}(CX, Sch_i)$$

Найдем главные компоненты. При этом осуществляется проецирование (так называемое, «натягивание») исходной многомерной совокупности данных на орты в пространстве меньшей размерности.

$$GK^{(i)} := \left(Svec^{(i)T} \cdot X \right)^T$$

Восстановление исходной матрицы по выбранным главным компонентам

Восстановим траекторные матрицы соответственно по первому, второму и третьему главным компонентам разложения.

$$M1 := \left(GK^{(1)} \cdot Svec^{(1)T} \right)^T$$

$$M2 := \left(GK^{(2)} \cdot Svec^{(2)T} \right)^T$$

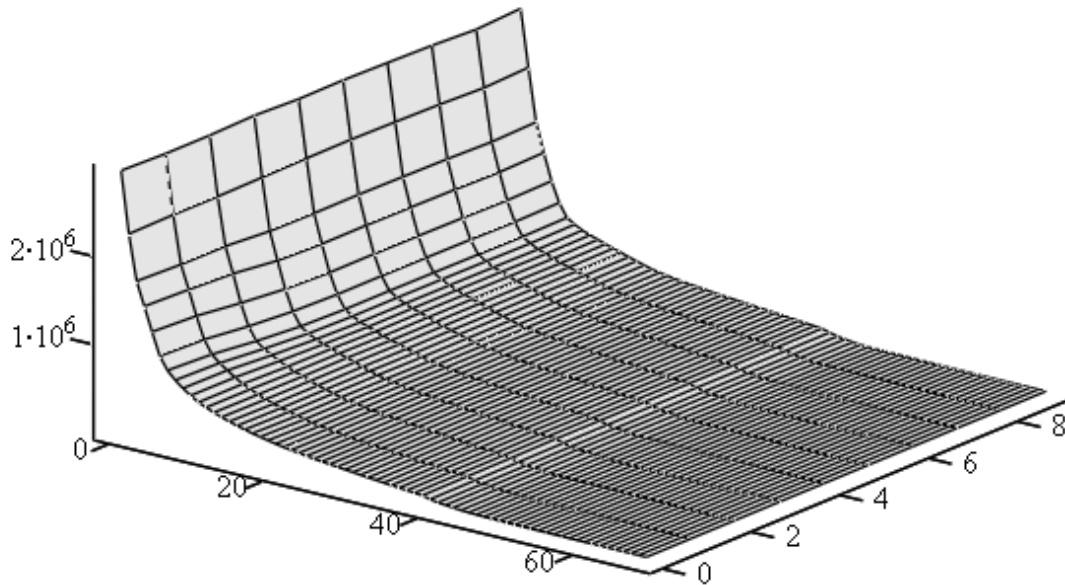
$$M3 := \left(GK^{(3)} \cdot Svec^{(3)T} \right)^T$$

Восстановим матрицу ранговых параметрических распределений.

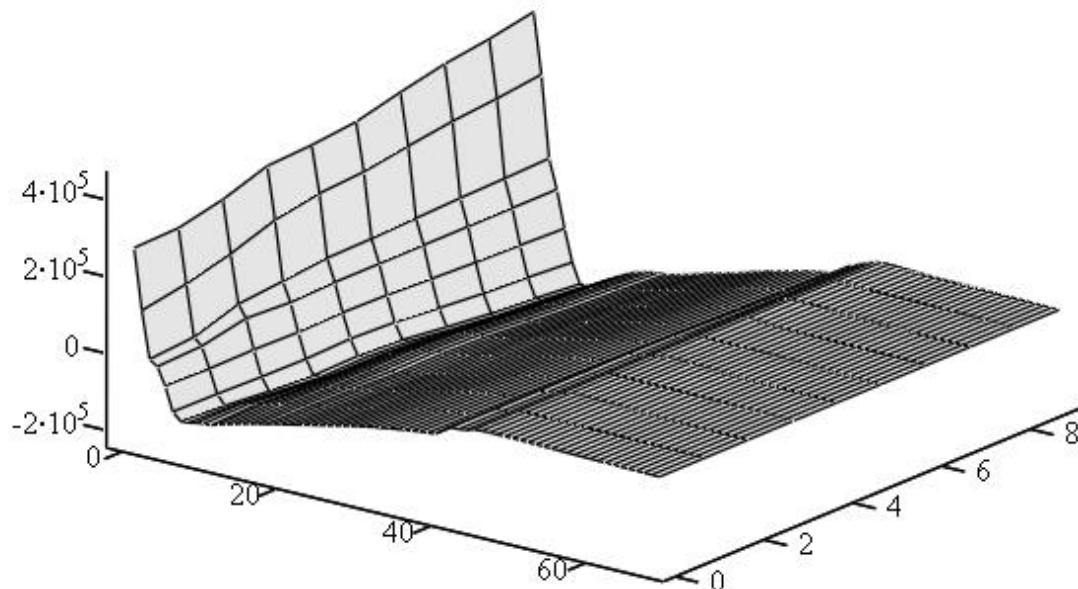
$$Vosst(a, b) := \left| \begin{array}{l} \text{for } i \in 1.. \text{rows}(a) \\ \quad \text{for } j \in 1.. \text{cols}(a) \\ \quad \quad \left| \begin{array}{l} c \leftarrow b_{i,j} \\ k \leftarrow c_1 \\ n \leftarrow c_2 \\ d_{(k,n)} \leftarrow a_{i,j} \end{array} \right. \\ \quad \quad \quad d \end{array} \right.$$

$$W1 := \text{Vosst}(M1, A)$$
$$W2 := \text{Vosst}(M2, A)$$
$$W3 := \text{Vosst}(M3, A)$$

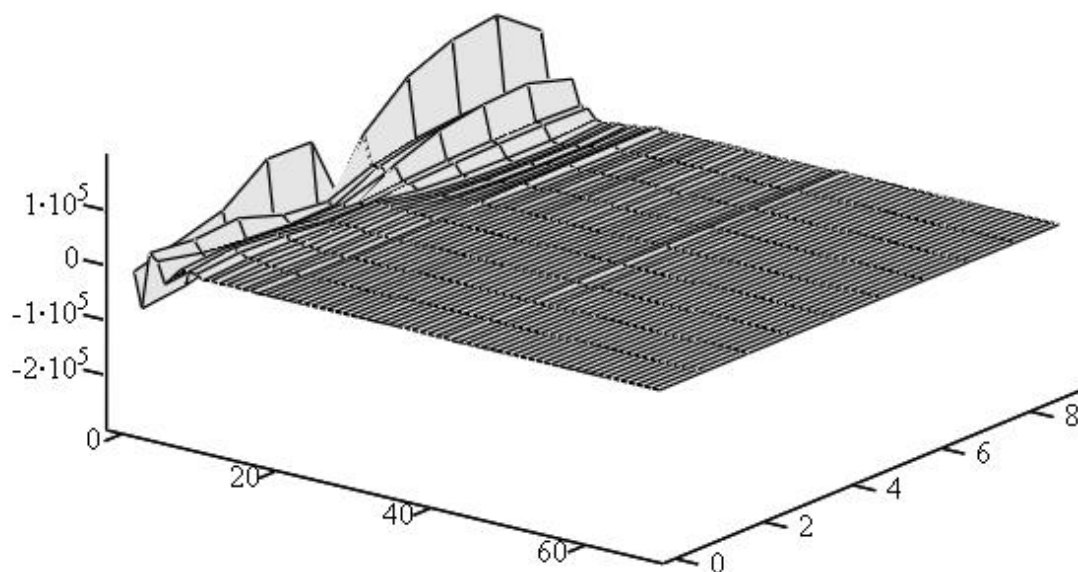
В результате проделанных процедур получаем разложение матрицы ранговых параметрических распределений по электропотреблению.



W1



W2



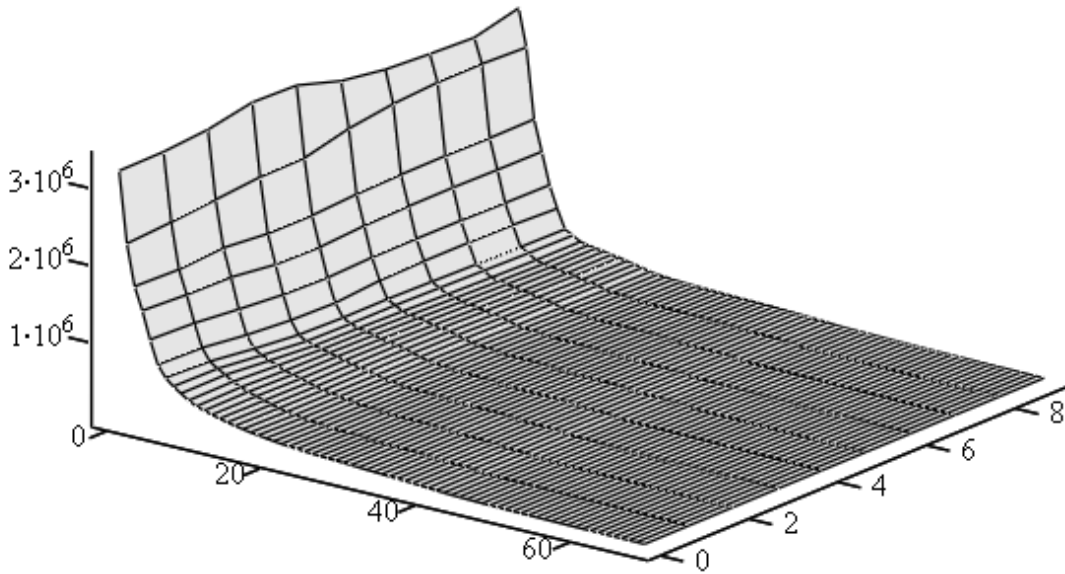
W3

Трехмерные ранговые поверхности техноценоза, восстановленные по главным компонентам: W1, W2 и W3 – матрицы соответственно первой, второй и третьей главных компонент; абсцисса – ранг объекта; ордината – временные интервалы (годы); аппликата – электропотребление (знак учитывает отклонения второй и третьей компонент), кВт·ч

Первая главная компонента является основной в разложении. Она задает структуру ранговой параметрической поверхности техноценоза по электропотреблению. Данная компонента описывает техноценоз в целом. В известном смысле ее можно считать «цифровой компонентой». Вторая главная компонента является вторичной (с точки зрения порядка величин). Она задает тренд процесса электропотребления. Данная компонента имеет отношение в основном к ноевой и пойнтер-кастам техноценоза. Ее можно считать «гауссовой компонентой разложения». Третья главная компонента является третичной (с точки зрения порядка величин). Она задает цикличность процесса электропотребления. Данная компонента имеет отношение в основном к ноевой и старшим пойнтер-кастам техноценоза. Ее также можно считать «гауссовой компонентой разложения».

Очевидно, что для нахождения результирующей матрицы необходимо осуществить суперпозицию отдельных компонент (просуммировать матрицы соответственно первой, второй и третьей главных компонент).

$$W := W1 + W2 + W3$$



W

Трёхмерная ранговая поверхность техноценоза, восстановленная по трем главным компонентам: W – восстановленная матрица данных; абсцисса – ранг объекта; ордината – временные интервалы (годы); аппликата – электропотребление, кВт·ч

Проанализируем отклонения восстановленных значений (матрица W на рис. выше) от исходных эмпирических (матрица V1).

$$DIF(W, V1) := \left| \begin{array}{l} \text{for } j \in 1.. \text{cols}(W) \\ \left| \sum_{i=1}^{\text{length}(W^{(j)})} (W^{(j)})_i - \sum_{i=1}^{\text{length}(V1^{(j)})} (V1^{(j)})_i \right| \\ a_j \leftarrow \frac{\quad}{\sum_{i=1}^{\text{length}(V1^{(j)})} (V1^{(j)})_i} \cdot 100 \end{array} \right|$$

DIF := DIF(W, V1) max(DIF) = 4.691

$$DIF^T = \begin{array}{|c|c|c|c|c|c|c|c|c|c|c|} \hline & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ \hline 1 & 4.514 & 2.547 & 1.231 & 1.937 & 3.717 & 0.815 & 2.52 & 0.563 & 4.691 & 3.749 \\ \hline \end{array}$$

Таким образом, в ходе реализации программы удалось разложить матрицу ранговых параметрических распределений, определить ее главные

компоненты, а также восстановить данные после удаления несистематических составляющих. При этом суммарное максимальное относительное отклонение восстановленных значений от исходных эмпирических на одном временном интервале не превысило 5 %. Следовательно, задача удаления «шумовой» составляющей решена без потери полезной информации.

[на содержание]

GZ-АНАЛИЗ РАНГОВОГО ПАРАМЕТРИЧЕСКОГО РАСПРЕДЕЛЕНИЯ ПО ЭЛЕКТРОПОТРЕБЛЕНИЮ

На этапе статистического анализа и построения эмпирической модели процесса ресурсопотребления осуществляется полномасштабная обработка данных, которая включает взаимосвязанные процедуры рангового анализа. Это позволяет упорядочивать информацию, выявлять в динамике и наглядно представлять объекты с аномальным ресурсопотреблением, эффективно осуществлять прогнозирование и нормирование ресурсопотребления отдельными объектами и техноценозом в целом. С целью повышения точности расчетов стандартные процедуры рангового анализа дополняются соответствующими тонкими процедурами: верификацией базы данных по ресурсопотреблению, а также дифлекс-, GZ-, ASR-анализом рангового параметрического распределения. Как известно, в процедуре прогнозирования существенную сложность составляет операция выбора G- или Z-методологии. Предлагается данный выбор осуществлять на основе предварительного, так называемого, тонкого GZ-анализа параметрического распределения, а в качестве критерия выбора метода рассматривать определенное соотношение объемов системного и гауссового ресурсов кластеров техноценоза, которое в пределе сводится к соотношению системного и гауссового доверительных интервалов объектов.

Под системным доверительным интервалом рангового параметрического распределения техноценоза понимается совокупность верхних и нижних доверительных границ, каждая из которых получается в результате статистической обработки выборки значений параметров, соответствующих данному рангу на протяжении определенного количества временных интервалов (независимо от объектов, которые «проходят» через ранг в процессе функционирования). Системный интервал характеризует разброс параметров техноценоза, при котором его функционирование можно считать нормальным. Фактически этот интервал отражает требования или ограничения, которые система выдвигает объектам, чтобы обеспечить свое устойчивое функционирование. Если эмпирическое значение параметра объекта выходит за пределы системного интервала, то это означает, что объект не подчиняется системным требованиям и проявляет индивидуальность. Сама же величина отклонения может рассматриваться как мера индивидуальности данного объекта. Ширина системного доверительного ин-

тервала определяется предысторией развития техноценоза и если она невелика, то это свидетельствует о том, что техноценоз развивается стабильно и сбалансированно, все изменения в нем происходят плавно. Напротив, широкий интервал свидетельствует о резких структурных изменениях.

Под гауссовым доверительным интервалом рангового параметрического распределения техноценоза понимается совокупность верхних и нижних доверительных границ, каждая из которых получается в результате статистической обработки выборки значений параметров, соответствующих данному объекту на протяжении определенного количества временных интервалов (независимо от рангов, которые он принимает в процессе функционирования). Гауссовый доверительный интервал характеризует разброс параметров объекта, при которых его функционирование можно считать нормальным. Ширина данного интервала задается предысторией развития объекта, и чем ширина меньше, тем стабильней его функционирование. В случае выхода эмпирического значения параметра объекта за границы гауссового интервала, можно говорить, что на самом объекте произошли резкие изменения (модернизация оборудования, новые технологии) либо система предъявила новые требования, для выполнения которых объект вынужден резко изменить режим функционирования.

Таким образом, для каждого объекта можно вычислить отношение системного и гауссового интервалов, которое называется коэффициентом когерентности и показывает степень согласованности поведения объекта по отношению к техноценозу. Если коэффициент когерентности близок к единице, то можно говорить о согласованном поведении. Кроме того, коэффициент когерентности является индикатором для выбора наиболее эффективного метода прогнозирования для рассматриваемого объекта. При сравнительно больших значениях коэффициента когерентности лучше работают G-методы, в противном случае – Z-методы.

Используем GZ-анализ для прогнозирования электропотребления объектов техноценоза. Следует сразу уточнить, что проблема выбора метода актуальна только при прогнозировании электропотребления отдельных объектов. В процессе прогнозирования электропотребления техноценоза в целом всегда лучше работают Z-методы.

После сбора информации о годовом электропотреблении (в кВт·ч) создается электронная база данных. Рекомендуется базу данных формировать в виде файла Microsoft Excel, в котором должны содержаться только числовые значения, при этом столбцы должны соответствовать объектам техноценоза, а строки – годам. Перед началом обработки данные необходимо подвергнуть верификации и сохранить под именем «data_d.xls» в директории «с:\mathcad_dat», которая должна быть заблаговременно создана в корневом каталоге диска «с:\». Далее, уже программными средствами, осуществляется импорт данных из файла «data_d.xls» в тело mathcad-программы. В настоящем примере используются десятилетние данные по

электропотреблению реально существующего в Калининградской области техноценоза, состоящего из рассредоточенных по всей территории региона 69 объектов. Кроме того, в настоящей программе будет осуществлено сравнение полученных коэффициентов когерентности с кодами лучших методов прогнозирования, ранее объединенных в так называемую GZ-матрицу. Для этого в директории «с:\mathcad_dat» должен быть размещен файл с именем «GZ_matrix.xls» (имеется среди файлов примера – см. по адресу: http://gnatukvi.ru/zip_files/task_mcd.zip).

Подготовка данных

Зададим начало отсчета и импортируем данные по электропотреблению объектов техноценоза из внешнего файла.

ORIGIN := 1

V := 
C:\mathcad_dat\data_d.xls

Определим глубину базы данных, используемой для дальнейшей обработки. Здесь необходимо вручную задать число, равное глубине базы данных в годах (в данном случае мы задаем число 9).

q := 9

Сформируем матрицу данных, а также создадим табулированное ранговое параметрическое распределение и вектор рангов.

V := submatrix(V, 1, q, 1, 69)

W1 := V^T

Zipf(Y) :=
$$\left| \begin{array}{l} \text{for } i \in 1.. \text{cols}(Y) \\ \left| \begin{array}{l} c \leftarrow \text{sort}(Y^{\langle i \rangle}) \\ b^{\langle i \rangle} \leftarrow \text{reverse}(c) \end{array} \right. \\ b \end{array} \right.$$

W := Zipf(W1)

i := 1.. rows(W)

R_i := i

Построение системного доверительного интервала

Расчет системного доверительного интервала осуществляется на основе интерквартильного размаха применительно к матрице ранговых параметрических распределений. В данной программе используется размах, соответствующий 95 %-му доверительному интервалу. Подпрограмма *Qvartil* предназначена для нахождения значения случайной величины из заданной выборки по значению функции распределения. Ее аргументы: x – выборка случайной величины, p – значение функции распределения.

$$\text{Qvartil}(x, p) := \left| \begin{array}{l} q \leftarrow p \cdot \text{length}(x) + p \\ a \leftarrow \begin{cases} 0 & \text{if } q \leq 0 \\ q & \text{if } 0 < q < \text{length}(x) - 1 \\ \text{length}(x) - 1 & \text{if } q \geq \text{length}(x) - 1 \end{cases} \\ b \leftarrow a - \text{floor}(a) \\ s \leftarrow \text{sort}(x) \\ q \leftarrow s_{\text{floor}(a)} + b \cdot (s_{\text{ceil}(a)} - s_{\text{floor}(a)}) \\ q \end{array} \right.$$

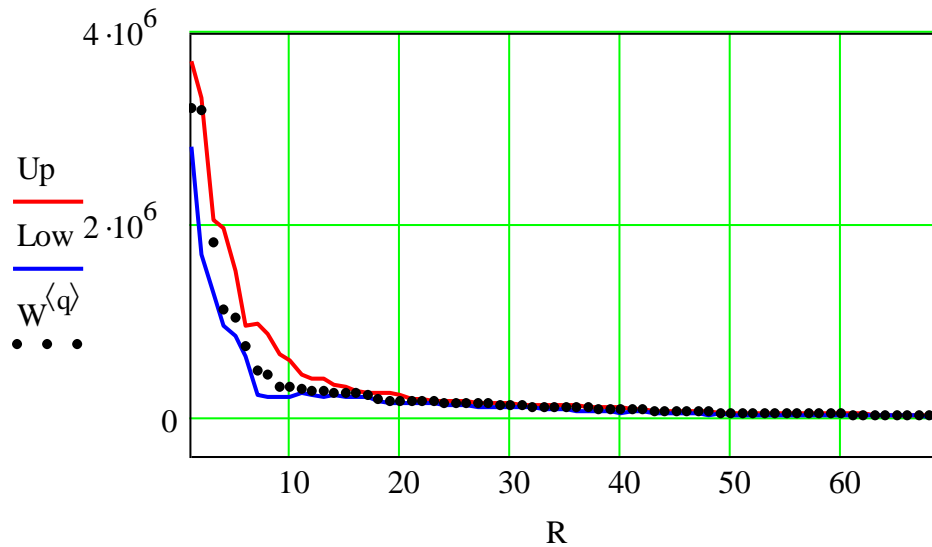
Подпрограмма *RE* предназначена для нахождения нижних и верхних границ доверительного интервала. Входным аргументом здесь является матрица ранговых параметрических распределений. На выходе получается матрица, первый столбец которой есть вектор верхних границ, второй – вектор нижних границ системного доверительного интервала.

$$\text{RE}(W) := \left| \begin{array}{l} W \leftarrow W^T \\ \text{for } i \in 1.. \text{cols}(W) \\ \quad \left| \begin{array}{l} \text{lower} \leftarrow \text{Qvartil}(W^{(i)}, 0.25) \\ \text{up} \leftarrow \text{Qvartil}(W^{(i)}, 0.75) \\ \text{IQ} \leftarrow \text{up} - \text{lower} \\ a_{i,1} \leftarrow \text{up} + 0.95 \cdot \text{IQ} \\ a_{i,2} \leftarrow \text{lower} - 0.95 \cdot \text{IQ} \end{array} \right. \\ a \end{array} \right.$$

Получаем верхнюю и нижнюю границы системного доверительного интервала, а затем визуализируем результаты расчетов.

$$Up := RE(W)^{\langle 1 \rangle}$$

$$Low := RE(W)^{\langle 2 \rangle}$$



Системный доверительный интервал: ордината – электропотребление, кВт·ч; абсцисса – ранг объекта; точки – эмпирические значения; ломаные линии – соответственно нижняя и верхняя границы интервала

Построение гауссового доверительного интервала

Расчет гауссового доверительного интервала осуществляется также на основе интерквартильного размаха применительно к матрице ранговых параметрических распределений техноценоза по электропотреблению. Проведем описанные выше операции только для значений электропотребления не рангов, а объектов. По вектору значений электропотребления каждого объекта находим соответствующее медианное значение.

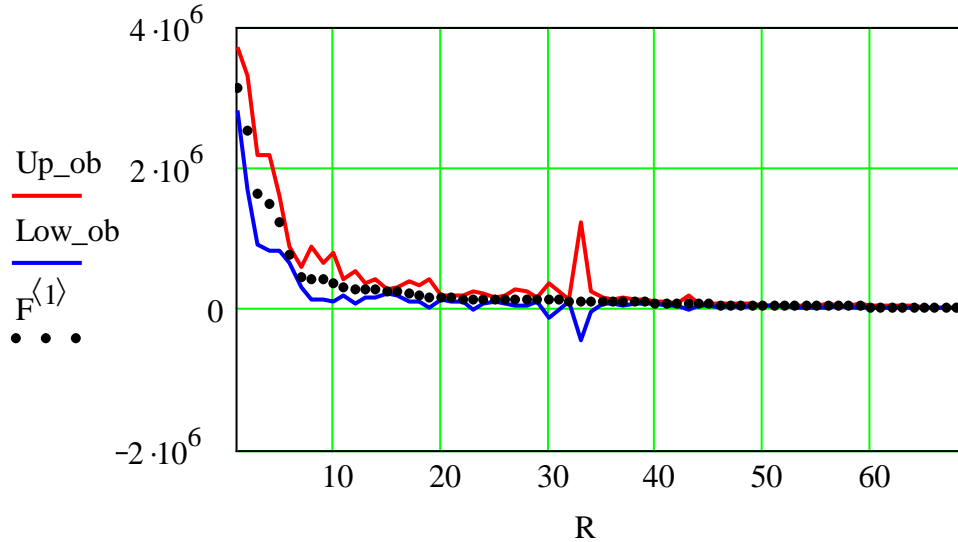
$$Median(a) := \begin{cases} \text{for } i \in 1.. \text{cols}(a) \\ c_i \leftarrow \text{median}(a^{\langle i \rangle}) \\ c \end{cases} \quad Med := Median(V)$$

При помощи подпрограммы RE на основе интерквартильного размаха определим доверительный интервал для каждого объекта, а затем отсортируем его по медианным значениям, которые ранжируются по убыванию. После этого визуализируем результаты расчетов (см. рис. ниже).

$$F := \text{augment}(\text{Med}, \text{RE}(V^T))$$

$$\text{Up_ob} := F^{\langle 2 \rangle}$$

$$F := \text{reverse}(\text{csort}(F, 1))$$

$$\text{Low_ob} := F^{\langle 3 \rangle}$$


Гауссовый доверительный интервал: ордината – электропотребление, кВт·ч; абсцисса – ранг объекта; точки – эмпирические медианные значения; ломаные линии – нижняя и верхняя границы интервала

Рассчитаем ширину системного доверительного интервала для каждого ранга параметрического распределения.

$$\text{Int_rang} := \text{Up} - \text{Low}$$

Далее рассчитаем ширину гауссового доверительного интервала для каждого объекта исследуемого техноценоза.

$$\text{Int_ob} := \text{Up_ob} - \text{Low_ob}$$

Расчет коэффициента когерентности

Вычислим коэффициент когерентности, наращивая глубину базы данных по электропотреблению до 9 лет, прибавляя по году начиная с 5 лет, и сформируем матрицу, отражающую динамику изменения данного коэффициента (глубина базы устанавливается в начале программы присвоением соответствующего значения параметру q). По размеру и способу формирования данная матрица схожа с GZ-матрицей, использовавшейся в ряде работ для оценки динамики изменения методов прогнозирования.

Ниже приведена подпрограмма, при помощи которой определяется значение коэффициента когерентности для каждого объекта техноценоза.

$$\text{TE}(a, b) := \left| \begin{array}{l} \text{for } i \in 1.. \text{rows}(a) \\ \quad d_i \leftarrow \frac{a_i}{b_i} \\ \quad d \end{array} \right.$$

$\text{Ind1} := \text{TE}(\text{Int_rang}, \text{Int_ob})$
 $\text{F1} := \text{augment}(\text{Med}, \text{R})$
 $\text{Kod_ob} := \text{reverse}(\text{csort}(\text{F1}, 1)) \langle 2 \rangle$
 $\text{Ind} := \text{csort}(\text{augment}(\text{Kod_ob}, \text{Ind1}), 1) \langle 2 \rangle$

Следует отметить, что полученный вектор коэффициентов когерентности Ind может быть использован в качестве индикатора при выборе наиболее эффективного метода прогнозирования. Однако для определения критериальных значений данного индикатора необходимо провести его юстировку путем статистического сравнения коэффициентов когерентности с индексами образцовой GZ -матрицы, полученной ранее в результате простого анализа методов прогнозирования техноценоза.

Итак, изменяя глубину базы данных, сформируем матрицу, отражающую динамику изменения коэффициента когерентности. Делается это следующим образом. Глубина базы устанавливается в начале программы присвоением соответствующего значения параметру q (см. первый пункт в начале данной программы). При этом следует сначала установить значение q , равное 5 (т.к. общая глубина базы составляет 10 лет, мы резервируем первые 5 лет в качестве базовой матрицы данных). Затем необходимо осуществить расчет программы и полученный вектор Ind вручную скопировать в первый столбец матрицы G (см. ниже). Далее надо вернуться в начало программы, изменить значение q на 6, а полученный новый вектор Ind скопировать во второй столбец матрицы G . В итоге, изменяя q от 5 до 9, получим матрицу G (размерностью 69 на 5), в которой будут записаны значения коэффициентов когерентности для всех 69 объектов исследуемого техноценоза на протяжении пяти лет функционирования.

| | | | | | |
|--------|------|-------|-------|-------|-------|
| $G :=$ | 0,75 | 0,714 | 1,002 | 0,725 | 0,75 |
| | 0,73 | 0,883 | 0,85 | 0,672 | 0,593 |
| | 0,26 | 0,266 | 0,58 | 0,405 | 0,875 |
| | 1,30 | 1,199 | 0,773 | 0,53 | 0,536 |
| | 1,69 | 1,343 | 1,533 | 1,483 | 0,828 |
| | 0,60 | 0,492 | 0,039 | 0,036 | 0,024 |
| | 1,06 | 1,298 | 1,492 | 1,951 | 1,914 |
| | 0,43 | 0,625 | 0,728 | 1,222 | 1,411 |

Юстировка коэффициента когерентности

Для дальнейшего анализа преобразуем матрицу коэффициентов когерентности в вектор, а также импортируем GZ-матрицу.

$$SX(a) := \begin{cases} \text{for } i \in 1.. \text{cols}(a) \\ \left| \begin{array}{l} s \leftarrow a^{\langle 1 \rangle} \text{ if } i = 1 \\ s \leftarrow \text{stack}(s, a^{\langle i \rangle}) \text{ otherwise} \end{array} \right. \\ s \end{cases} \quad Gv := SX(G)$$

GZ := 
C:\..\GZ_matrix.xls

$$GZ^T = \begin{array}{c|cccccccccccccccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 \\ \hline 1 & 5 & 6 & 5 & 5 & 4 & 4 & 3 & 3 & 6 & 6 & 4 & 5 & 2 & 2 & 5 & 2 & 6 & 2 & 4 \\ 2 & 4 & 3 & 4 & 5 & 4 & 5 & 4 & 3 & 6 & 6 & 2 & 5 & 5 & 3 & 4 & 1 & 1 & 5 & 4 \\ 3 & 5 & 5 & 2 & 3 & 3 & 4 & 4 & 3 & 6 & 6 & 6 & 3 & 3 & 5 & 1 & 5 & 5 & 6 & 1 \\ 4 & 1 & 6 & 2 & 4 & 2 & 4 & 3 & 3 & 3 & 6 & 1 & 2 & 2 & 2 & 6 & 2 & 4 & 4 & 5 \\ 5 & 3 & 3 & 5 & 5 & 3 & 5 & 3 & 3 & 3 & 6 & 5 & 4 & 4 & 5 & 4 & 5 & 2 & 2 & 5 \end{array}$$

Следует отметить, что данная GZ-матрица получена ранее в результате последовательной многократной реализации различных методов прогнозирования применительно к той же базе данных по электропотреблению. Столбцы этой матрицы – объекты, строки – последние пять лет предыстории функционирования техноценоза. Элемент матрицы – код метода, который давал для объекта наименьшую относительную годовую ошибку. Кодировка методов прогнозирования в GZ-матрице следующая:

1 – G-метод прогнозирования на основе модели авторегрессионного скользящего среднего (Garss);

2 – G-метод прогнозирования на основе модели декомпозиции временного ряда (Gmdvr);

3 – G-метод прогнозирования на основе анализа сингулярного спектра траекторной матрицы временного ряда (Gssa);

4 – Z-метод без фиксированной первой точки (Zbpt);

5 – Z-метод прогнозирования с фиксированной первой точкой (Zspt);

6 – Z-метод прогнозирования с делением на кастовые зоны (Zdkz).

Преобразуем значения в GZ-матрице по следующему правилу: если лучшим был G-метод прогнозирования электропотребления, то значению элемента матрицы присваивается 1, если Z-метод прогнозирования – 2. После этого векторизуем полученную матрицу в вектор Kgz.

$$DT(GZ) := \left| \begin{array}{l} \text{for } i \in 1.. \text{rows}(GZ) \\ \quad \text{for } j \in 1.. \text{cols}(GZ) \\ \quad \quad \left| \begin{array}{l} a_{i,j} \leftarrow 1 \text{ if } GZ_{i,j} \leq 3 \\ a_{i,j} \leftarrow 2 \text{ otherwise} \end{array} \right. \\ \quad \quad a \end{array} \right. \quad GZ1 := DT(GZ)$$

$$Kgz := SX(GZ1)$$

Объединим в единую расчетную матрицу вектора, полученные ранее из GZ-матрицы, а также матрицы коэффициентов когерентности.

$$Bn := \text{augment}(Gv, Kgz)$$

Осуществим с помощью стандартной функции Mathcad сортировку элементов по возрастанию значений коэффициента когерентности.

$$Q2 := \text{csort}(Bn, 1)$$

Определим с помощью стандартных функций Mathcad максимальное и минимальное значения коэффициента когерентности.

$$\begin{array}{ll} Mx := \max(Q2^{(1)}) & Mx = 4.493 \times 10^4 \\ Mn := \min(Q2^{(1)}) & Mn = 0.024 \end{array}$$

Исследуем, насколько величина коэффициента когерентности связана с кодами методов прогнозирования. Ниже приведена подпрограмма D, которая по заданной ширине интервала значений коэффициента когерентности извлекает соответствующий вектор с кодами методов прогнозирования. Входные параметры подпрограммы: a – матрица, b и c – левая и правая границы интервала значений коэффициента когерентности. На выходе подпрограммы формируется вектор с кодами методов прогнозирования.

$$D(a, b, c) := \left| \begin{array}{l} k \leftarrow 1 \\ \text{for } i \in 1.. \text{rows}(a) \\ \quad \text{if } b \leq a_{i,1} \leq c \\ \quad \quad \left| \begin{array}{l} d_k \leftarrow a_{i,2} \\ k \leftarrow k + 1 \end{array} \right. \\ \quad \quad d \end{array} \right. \quad M := D(Q2, 1.5, Mx)$$

го сформируем вектор из исходной GZ-матрицы, объединим его с вектором коэффициента когерентности, после чего произведем сортировку.

$KGZ := SX(GZ)$

$B := \text{augment}(Gv, KGZ)$

$B1 := \text{csort}(B, 1)$

$$D(a, b, c) := \left| \begin{array}{l} k \leftarrow 1 \\ \text{for } i \in 1.. \text{rows}(a) \\ \quad \text{if } b \leq a_{i,1} < c \\ \quad \quad \left| \begin{array}{l} d_{k,2} \leftarrow a_{i,2} \\ d_{k,1} \leftarrow a_{i,1} \\ k \leftarrow k + 1 \end{array} \right. \\ \quad \quad \quad d \end{array} \right.$$

$M1 := D(B1, 0.5, 1.5)$

$$S(a, n) := \left| \begin{array}{l} k \leftarrow 1 \\ \text{for } i \in 1.. \text{rows}(a) \\ \quad \text{if } a_{i,2} = n \\ \quad \quad \left| \begin{array}{l} d_{k,1} \leftarrow a_{i,1} \\ d_{k,2} \leftarrow a_{i,2} \\ k \leftarrow k + 1 \end{array} \right. \\ \quad \quad \quad d \end{array} \right.$$

$T := S(M1, 3)$

Рассчитаем частоту встречаемости методов прогнозирования в интервале, а также статистические параметры наиболее часто встречающегося метода (как видим ниже, таковым оказался Gssa).

$$\text{Chastota_Garss} := \frac{\text{rows}(S(M1, 1))}{\text{rows}(M1)} \qquad \text{Chastota_Garss} = 0.104$$

$$\text{Chastota_Gmdvr} := \frac{\text{rows}(S(M1, 2))}{\text{rows}(M1)} \qquad \text{Chastota_Gmdvr} = 0.156$$

$$\text{Chastota_Gssa} := \frac{\text{rows}(S(M1, 3))}{\text{rows}(M1)} \qquad \text{Chastota_Gssa} = 0.223$$

$$\text{Chastota_Zbpt} := \frac{\text{rows}(S(M1, 4))}{\text{rows}(M1)} \qquad \text{Chastota_Zbpt} = 0.185$$

$$\text{Chastota_Zspt} := \frac{\text{rows}(S(M1, 5))}{\text{rows}(M1)} \quad \text{Chastota_Zspt} = 0.199$$

$$\text{Chastota_Zdkz} := \frac{\text{rows}(S(M1, 6))}{\text{rows}(M1)} \quad \text{Chastota_Zdkz} = 0.133$$

$$\text{Sred_kof_kog_SSA} := \text{mean}(S(M1, 3)^{\langle 1 \rangle})$$

$$\text{Sred_kof_kog_SSA} = 0.899$$

$$\text{SKO_kof_kog_SSA} := \text{stdev}(S(M1, 3)^{\langle 1 \rangle})$$

$$\text{SKO_kof_kog_SSA} = 0.245$$

Анализ полученных результатов

Если значение коэффициента когерентности лежит в пределах от минимального до 0.5, то это значит, что объект ведет себя нестабильно. В такой ситуации G-методы, выполняющие прогноз на основе анализа траектории электропотребления, дают значительные ошибки. Выходом может служить использование Z-методов, которые в условиях нестабильности способны осуществлять прогноз параметров объекта через поведение техноценоза, обладающего существенно большей инерционностью.

Если значение коэффициента когерентности лежит в пределах от 0.5 до 1.5, то это значит, что объект в основном ведет себя согласованно с техноценозом. В данной ситуации можно ожидать в равной степени хорошей работы как G-, так и Z-методов прогнозирования. Однако в ходе углубленного статистического анализа установлено, что здесь лучше всего работает метод на основе анализа сингулярного спектра (Gssa) (среднее значение коэффициента когерентности по всему техноценозу для данного метода составляет 0.899, а среднеквадратичное отклонение – 0.245).

Если значение коэффициента когерентности лежит в пределах от 1.5 до максимального значения, то это значит, что объект ведет себя гораздо стабильней системы, которая начинает претерпевать существенные структурные изменения. В данной ситуации для объектов лучше работают G-методы прогнозирования электропотребления.

Вывод

Таким образом, коэффициент когерентности позволяет уже на этапе предварительной статистической обработки априорно судить о наиболее эффективных методах прогнозирования ресурсопотребления объектов, что существенно повышает скорость обработки данных и повышает точность прогноза (особенно если речь идет о крупных базах данных, формирующихся автоматически и управляемых в реальном времени).

[на содержание]

КЛАССИФИКАЦИЯ ТЕХНОЦЕНОЗА ПО ЭЛЕКТРОПОТРЕБЛЕНИЮ

В ряде процедур рангового анализа (интервальное оценивание, нормирование, ASR-анализ и др.) применяется предварительная процедура классификации объектов техноценоза по электропотреблению. Однако использование стандартных процедур кластер-анализа, как правило, наталкивается на трудности, связанные с негауссовостью техноценозов и, как следствие, существенной нечеткостью границ между кластерами. Подобные трудности в значительной степени снимаются применением классификации, моделирующей известный принцип «Решета Эратосфена».

Данный метод классификации основан на идее формирования кортежа наиболее похожих объектов. Степень похожести определяется мерой близости объектов в параметрическом пространстве (в данном случае рассматривается пространство в координатах «ранг объекта – время – электропотребление, кВт·ч»). Для каждого объекта техноценоза формируется и ставится в соответствие кортеж наиболее близких к нему объектов. После этого осуществляется сортировка кортежей, при этом на первое место ставится кортеж, разброс объектов (внутриклассовая дисперсия) внутри которого относительно его центра минимален. Далее производится процедура формирования классов. В качестве первого класса принимается кортеж с минимальной дисперсией, после чего объекты, попавшие в него, из дальнейшего анализа и классификации исключаются. В качестве второго класса принимается кортеж, не исключенный из процедуры классификации в ходе предыдущего шага и имеющий минимальную дисперсию (в состав класса включаются только объекты, не расклассифицированные на предыдущем шаге). Описанные процедуры продолжаются до тех пор, пока вся выборка объектов не будет расклассифицирована. В качестве функционала, определяющего качество классификации, выбирается отношение суммы параметров, характеризующих межклассовый разброс и суммарную дисперсию классов, к сумме количества элементов в кортеже и числа классов. Оптимальным считается разбиение, при котором описанный выше функционал принимает максимальное значение.

Рассмотрим пример. После сбора информации о годовом электропотреблении (в кВт·ч) создается электронная база данных. Рекомендуется базу данных формировать в виде файла Microsoft Excel, в котором должны содержаться только числовые значения, при этом столбцы должны соответствовать объектам техноценоза, а строки – годам. Перед началом обработки данные необходимо подвергнуть верификации и сохранить под именем «data_d.xls» в директории «с:\mathcad_dat», которая должна быть заблаговременно создана в корневом каталоге диска «с:\». Далее, уже обыч-

ными программными средствами, осуществляется импорт данных из файла «data_d.xls» в тело mathcad-программы. В настоящем примере используются десятилетние данные по электропотреблению реально существующего в Калининградской области техноценоза, состоящего из рассредоточенных по территории региона 69 объектов.

Подготовка данных

Зададим начало отсчета индексов для матриц, векторов и массивов, а затем импортируем данные по электропотреблению техноценоза.

```
ORIGIN := 1
W      := 
        C:\mathcad_data\data_d.xls
```

Сформируем ранговое параметрическое распределение по медианным значениям электропотребления объектов техноценоза.

```
i := 1..cols(W)
Ki := i
```

$$\text{ZIPF}(a) := \begin{cases} \text{for } i \in 1.. \text{cols}(a) \\ b_i \leftarrow \text{median}(a^{\langle i \rangle}) \\ c \leftarrow \text{reverse}(\text{sort}(b)) \\ \text{augment}(c, K) \end{cases}$$

```
Q := ZIPF(W)⟨1⟩
A := ZIPF(W)
qi := (QT)⟨i⟩
```

Определение оптимального количества элементов в кортеже

Ниже последовательно описаны стандартные расчетные модули, которые затем будут использоваться в подпрограмме, предназначенной для нахождения оптимального количества элементов в кортеже.

Модуль для вычисления центра классов:

$$\text{Senter}(I) := \left| \begin{array}{l} N \leftarrow \text{cols}(I) \\ \sum_{i=1}^N I^{\langle i \rangle} \\ b \leftarrow \frac{\quad}{N} \\ b \end{array} \right.$$

Модуль для вычисления внутриклассовой дисперсии:

$$\text{Disp}(I) := \left| \begin{array}{l} N \leftarrow \text{cols}(I) \\ b \leftarrow \text{Senter}(I) \\ \sum_{i=1}^N \left[(I^{\langle i \rangle} - b)^T \cdot (I^{\langle i \rangle} - b) \right] \\ S \leftarrow \frac{\quad}{N} \\ S \end{array} \right.$$

Модуль для определения матрицы, предназначенной для фиксации попарных расстояний между объектами техноценоза:

$$\text{Koff}(a) := \left| \begin{array}{l} \text{for } i \in 1.. \text{length}(a) \\ \quad \text{for } j \in 1.. \text{length}(a) \\ \quad \quad b_{i,j} \leftarrow \left| (a_i - a_j)^T \cdot (a_i - a_j) \right| \\ b \end{array} \right.$$

Модуль для выявления кодов близких элементов (формальные параметры: a – матрица попарных расстояний, b – вектор кодов объектов техноценоза, c – количество элементов в кортеже):

$$\text{Kod}(a, b, c) := \left| \begin{array}{l} \text{for } i \in 1.. \text{cols}(a) \\ \quad \left| \begin{array}{l} d \leftarrow \text{augment}(a^{\langle i \rangle}, b) \\ d \leftarrow \text{csort}(d, 1)^{\langle 2 \rangle} \\ e^{\langle i \rangle} \leftarrow \text{stack}(i, \text{submatrix}(d, 2, c, 1, 1)) \end{array} \right. \\ e \end{array} \right.$$

Модуль для составления кортежей элементов по найденным кодам (формальные параметры: a – вектор с координатами объектов исследуемого техноценоза, b – матрица кодов элементов кортежа):

$$\text{Korteg}(a, b) := \left| \begin{array}{l} \text{for } i \in 1 \dots \text{length}(a) \\ \quad \left| \begin{array}{l} \text{for } k \in 1 \dots \text{rows}(b) \\ \quad \left| \begin{array}{l} c \leftarrow a_{b_{1,i}} \text{ if } k = 1 \\ c \leftarrow \text{augment}[c, a_{(b_{k,i})}] \text{ otherwise} \end{array} \right. \\ \quad d_i \leftarrow c \end{array} \right. \\ d \end{array} \right.$$

Модуль для вычисления дисперсии и сортировки кортежей в порядке возрастания внутриклассовой дисперсии (формальные параметры данного модуля: a – вектор, элемент которого есть кортеж с координатами объектов исследуемого техноценоза, b – вектор кодов элементов кортежа):

$$\text{SORT}(a, b) := \left| \begin{array}{l} \text{for } i \in 1 \dots \text{length}(a) \\ \quad d_i \leftarrow |\text{Disp}(a_i)| \\ \quad e \leftarrow \text{augment}(d, b) \\ \quad \text{csort}(e, 1)^{\langle 2 \rangle} \end{array} \right.$$

На выходе данного модуля получаем упорядоченный вектор, элементом которого является код объекта (на первом месте стоит код объекта, кортеж которого имеет минимальную внутриклассовую дисперсию).

Далее следует модуль, исключающий из выборки объекты после попадания их в один из классов (формальные параметры: b – вектор кодов объектов, «населяющих» кортеж, a – вектор, элемент которого представляет собой код первого объекта кортежа). На выходе модуля (см. ниже) получаем вектор « a », из которого исключены коды объектов, попавших в кортеж « b ». Однако здесь следует остановиться несколько подробнее. Прежде всего, необходимо учитывать следующую особенность алгоритма. Классификация, проводимая с помощью рассматриваемых модулей, в общем случае приводит к покрытию классами, а не разбиению, т.к. таксоны могут пересекаться. Это связано с тем, что хотя элементы выборки на определенном этапе последовательно исключаются из рассмотрения, тем не менее, они оказывают влияние на последующую классификацию, поскольку на начальном этапе они участвовали в упорядочении выборки, и

порядок этот после их исключения не пересматривался. Очевидно, что подобная особенность алгоритма, помимо прочего, позволяет также учесть критерии соответствия Н-распределению (прежде всего, взаимосвязанность объектов в техноценозе посредством слабых связей особой природы). Что же касается перемешивания объектов в окончательной классификации, то этого в техноценозах никогда не происходит благодаря присутствующей параметрической выборке обязательной негауссовости.

$$\text{Trim}(a, b) := \left| \begin{array}{l} n \leftarrow 1 \\ a \leftarrow a^T \\ \text{for } j \in 1.. \text{cols}(a) \\ \quad \left| \begin{array}{l} k \leftarrow 0 \\ \text{for } i \in 1.. \text{length}(b) \\ \quad k \leftarrow k + 1 \text{ if } a_{1,j} \neq b_i \\ \text{if } k = \text{length}(b) \\ \quad \left| \begin{array}{l} d^{\langle n \rangle} \leftarrow a^{\langle j \rangle} \\ n \leftarrow n + 1 \end{array} \right. \\ d \leftarrow a \text{ if } n = 1 \\ d^T \end{array} \right. \end{array} \right.$$

Модуль для формирования класса объектов исследуемого техноценоза (формальные параметры модуля: a – матрица, первый столбец которой является вектором кодов объектов, а второй – вектором координат объектов; b – вектор кодов объектов, «населяющих» кортеж):

$$\text{classes}(a, b) := \left| \begin{array}{l} a \leftarrow a^T \\ n \leftarrow 1 \\ \text{for } j \in 1.. \text{cols}(a) \\ \quad \text{for } i \in 1.. \text{length}(b) \\ \quad \quad \text{if } a_{1,j} = b_i \\ \quad \quad \quad \left| \begin{array}{l} c^{\langle n \rangle} \leftarrow a_{2,j} \\ n \leftarrow n + 1 \end{array} \right. \\ c \end{array} \right.$$

Подпрограмма для формирования классов (формальные параметры подпрограммы: q – вектор с координатами объектов, K – вектор кодов объектов, $w3$ – вектор, элемент которого является кодом первого объекта в кортеже (см. модуль SORT), $w1$ – матрица кортежей, столбцы которой формируются кодами объектов, «населяющих» кортеж):

$$\text{Cluster}(q, K, w3, w1) := \left| \begin{array}{l} s \leftarrow \text{augment}(K, q) \\ k \leftarrow 1 \\ m \leftarrow 0 \\ \text{while } m < \text{length}(q) \\ \quad \left| \begin{array}{l} b \leftarrow w1 \langle (w3_1) \rangle \\ a_k \leftarrow \text{classes}(s, b) \\ m \leftarrow m + \text{cols}(a_k) \\ s \leftarrow \text{Trim}(s, b) \\ w3 \leftarrow \text{Trim}(w3, b) \\ k \leftarrow k + 1 \end{array} \right. \\ a \end{array} \right.$$

Модуль для нахождения межклассового разброса:

$$\text{BetweenCluster}(a) := \left| \begin{array}{l} \text{for } i \in 1.. \text{length}(a) \\ \quad b \langle i \rangle \leftarrow \text{Senter}(a_i) \\ c \leftarrow |\text{Disp}(b)| \\ c \end{array} \right.$$

Модуль для нахождения суммарной внутриклассовой дисперсии:

$$\text{SumDevCluster}(a) := \left| \begin{array}{l} c \leftarrow 0 \\ \text{for } i \in 1.. \text{length}(a) \\ \quad c \leftarrow c + |\text{Disp}(a_i)| \\ c \end{array} \right.$$

Подпрограмма для нахождения значений функционала, характеризующего качество классификации в зависимости от количества объектов,

формирующих кортеж (см. ниже). На каждом шаге подпрограммы осуществляется классификация множества и, в зависимости от объема кортежа, вычисляется функционал. В качестве функционала, определяющего качество классификации, выбирается отношение суммы параметров, характеризующих межклассовый разброс и суммарную дисперсию классов, к сумме количества элементов в кортеже и числа классов. На выходе получаем матрицу, первый столбец которой составляют значения функционала, второй – значения объема кортежа для осуществления классификации.

$$\text{OPTCLUSTER}(q, K) := \left| \begin{array}{l} a \leftarrow \text{Koff}(q) \\ k \leftarrow 1 \\ \text{for } i \in 2.. \text{length}(q) - 1 \\ \quad \left| \begin{array}{l} b \leftarrow \text{Kod}(a, K, i) \\ c \leftarrow \text{Korteg}(q, b) \\ e \leftarrow \text{SORT}(c, K) \\ f \leftarrow \text{Cluster}(q, K, e, b) \\ a1 \leftarrow \text{BetweenCluster}(f) \\ a2 \leftarrow \text{SumDevCluster}(f) \\ a3 \leftarrow \frac{a1 + a2}{i + \text{length}(f)} \\ s^{(k)} \leftarrow \text{stack}(i, a3) \\ k \leftarrow k + 1 \end{array} \right. \\ (s)^T \end{array} \right.$$

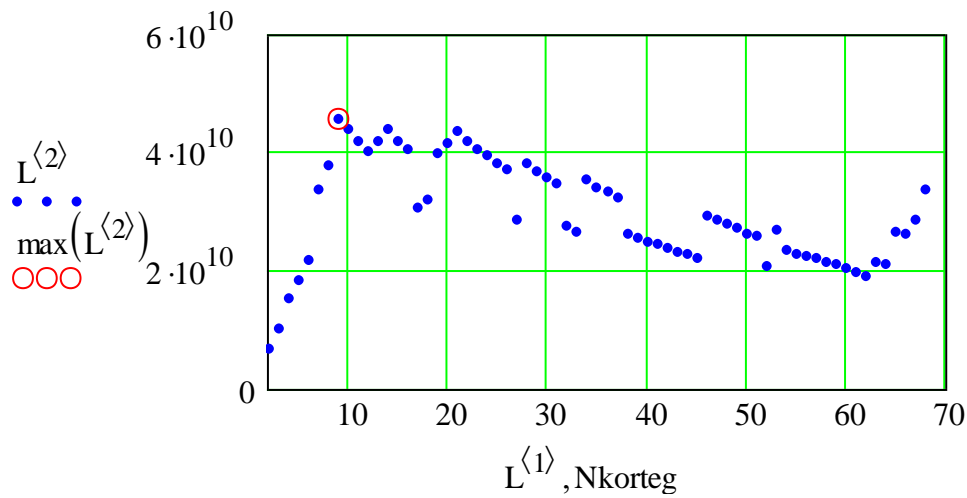
Модуль для нахождения по максимальному значению функционала, определяющего качество классификации, оптимального объема кортежа:

$$\text{Idclas}(a, b) := \left| \begin{array}{l} \text{for } i \in 1.. \text{rows}(a) \\ \quad \text{if } a_i = b \\ \quad \quad \left| \begin{array}{l} c \leftarrow i + 1 \\ \text{break} \end{array} \right. \\ c \end{array} \right.$$

$$L := \text{OPTCLUSTER}(q, K)$$

$$\text{Nkorteg} := \text{Idclas}(L^{(2)}, \max(L^{(2)}))$$

Модуль визуализации оптимизационного процесса:



Функционал качества разбиения на классы: абсцисса – количество элементов в кортеже; ордината – значение функционала качества (оптимальное значение обведено окружностью)

Итак, в рассматриваемом примере оптимальный объем кортежа:

$$Nkorteg = 9$$

Классификация техноценоза

Разобьем исходную совокупность объектов техноценоза с учетом оптимального количества объектов в кортеже (в данном случае – 9).

$$CLUSTER(q, K, n) := \left\{ \begin{array}{l} a \leftarrow Koff(q) \\ b \leftarrow Kod(a, K, n) \\ c \leftarrow Korteg(q, b) \\ e \leftarrow SORT(c, K) \\ f \leftarrow Cluster(q, K, e, b) \\ f \end{array} \right.$$

$$CL := CLUSTER(q, K, Nkorteg)$$

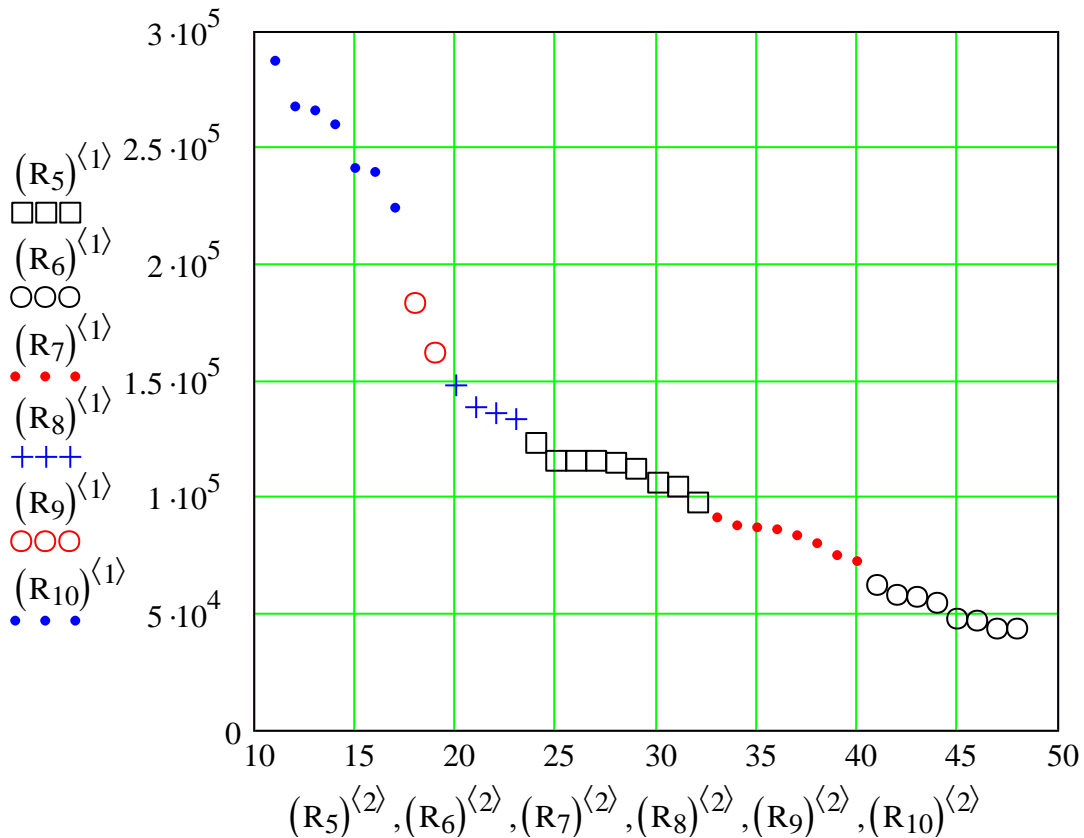
Результирующее количество классов в оптимальном разбиении:

$$length(CL) = 14$$

Визуализируем классы на параметрическом распределении в исходных координатах («ранг – электропотребление, кВт·ч») (см. рис. ниже).

$$M1(d, b) := \left| \begin{array}{l} d \leftarrow d^T \\ \text{for } j \in 1.. \text{length}(d) \\ \quad \text{for } k \in 1.. \text{rows}(b) \\ \quad \quad n^{(j)} \leftarrow \text{stack}(b_{k,1}, b_{k,2}) \text{ if } d_j = b_{k,1} \\ n^T \end{array} \right.$$

$$M2(a, b) := \left| \begin{array}{l} \text{for } i \in 1.. \text{length}(a) \\ \quad d_i \leftarrow M1(a_i, b) \\ d \end{array} \right. \quad R := M2(CL, A)$$



Фрагмент распределения классов по электропотреблению в исходных координатах: абсцисса – ранг объекта; ордината – электропотребление, кВт·ч; классы (с 5-го по 10-й) выделены с помощью различных условных графических обозначений

Кроме того, любой из классов можно оперативно просмотреть прямо в теле программы. Ниже для примера представлены 1-й, 4-й, 12-й и 14-й классы объектов техноценоза по электропотреблению. Следует иметь в виду, что нумерация классов в разбиении не соответствует нумерации объектов в исходном ранговом параметрическом распределении техноценоза (в основном она обратная, но не везде строго по порядку).

$$R = \begin{array}{|c|c|} \hline & 1 \\ \hline 1 & [9,2] \\ \hline 2 & [5,2] \\ \hline 3 & [2,2] \\ \hline 4 & [5,2] \\ \hline 5 & [9,2] \\ \hline 6 & [8,2] \\ \hline 7 & [8,2] \\ \hline 8 & [4,2] \\ \hline 9 & [2,2] \\ \hline 10 & [7,2] \\ \hline 11 & [4,2] \\ \hline 12 & [2,2] \\ \hline 13 & [3,2] \\ \hline 14 & [1,2] \\ \hline \end{array} \quad R_1 = \begin{pmatrix} 2.907 \times 10^4 & 51 \\ 2.867 \times 10^4 & 52 \\ 2.757 \times 10^4 & 53 \\ 2.732 \times 10^4 & 54 \\ 2.634 \times 10^4 & 55 \\ 2.55 \times 10^4 & 56 \\ 2.365 \times 10^4 & 57 \\ 2.3 \times 10^4 & 58 \\ 2.288 \times 10^4 & 59 \end{pmatrix} \quad R_4 = \begin{pmatrix} 1.088 \times 10^4 & 65 \\ 6.914 \times 10^3 & 66 \\ 4.638 \times 10^3 & 67 \\ 3.365 \times 10^3 & 68 \\ 2.737 \times 10^3 & 69 \end{pmatrix} \quad R_{12} = \begin{pmatrix} 1.157 \times 10^6 & 5 \\ 7.419 \times 10^5 & 6 \end{pmatrix} \quad R_{14} = \begin{pmatrix} 3.159 \times 10^6 & 1 \end{pmatrix}$$

Вывод

Таким образом, с помощью предлагаемой программы осуществлена успешная классификация исследуемого техноценоза по электропотреблению со следующими оптимальными параметрами: общее количество классов – 14, оптимальное число элементов в кортеже – 9.

[\[на содержание\]](#)

ГЕНЕРАТОР НЕГАУССОВОЙ ВЫБОРКИ ТЕХНОЦЕНОЛОГИЧЕСКОГО ТИПА

В результате параметрического описания техноценоза формируется трехмерная выборка, отражающая изменение во времени рассматриваемого параметра, описывающего поведение каждого объекта (особи). На основе подобных выборок в рамках процедур рангового анализа строятся и обрабатываются ранговые параметрические распределения (или поверхности). К настоящему времени на весьма обширном эмпирическом материале

множественно показана устойчивость, взаимосвязанность и негауссовость ранговых параметрических распределений техноценозов.

Как правило, на практике исследователь начинает работу с построения статистического распределения, которое возникает при эмпирическом описании техноценоза. С другой стороны, в ряде случаев возникает необходимость искусственной генерации негауссовых выборок техноценологического типа (создания однопараметрической модели). Решить данную задачу можно средствами Mathcad. При этом должна быть случайно сгенерирована двумерная матрица числовых значений, удовлетворяющая требованиям негауссовости, взаимосвязанности и устойчивости во времени. Сами числовые значения матрицы представляют собой параметры, столбцы – это объекты техноценоза, а строки – временные интервалы (месяцы, годы и др.). Кроме того, чтобы сгенерированная матрица отражала реальные процессы, происходящие в техноценозе, в ней должны присутствовать так называемые «аномальные» параметры двух типов. Первый тип – это нулевые данные или ошибочные выбросы. Ко второму типу относятся аномальные параметры, являющиеся результатом значительного рассогласования поведения объекта по сравнению с поведением системы. В качестве исходных данных при генерации выборки на различных этапах задаются: количество объектов в техноценозе, максимальное значение параметра, ранговый коэффициент и количество временных интервалов.

Формирование начальных параметров

Количество объектов «n» моделируемого техноценоза в качестве предварительного шага задается исследователем в теле программы.

$$n := 70$$

Значения первой точки $W1$ и рангового коэффициента β определяются при помощи встроенного генератора случайных чисел $\text{rnd}(x)$ (возвращает равномерно распределенное случайное число в интервале $[0; x]$). При этом базовые значения соответствующих параметров распределения могут быть заданы исследователем в теле программы.

$$Wb1 := 5 \cdot 10^5$$

$$\beta b := 1$$

$$W1 := Wb1 \cdot (1 + \text{rnd}(1))$$

$$\beta := \beta b + \text{rnd}(1)$$

$$W1 = 5.006 \times 10^5$$

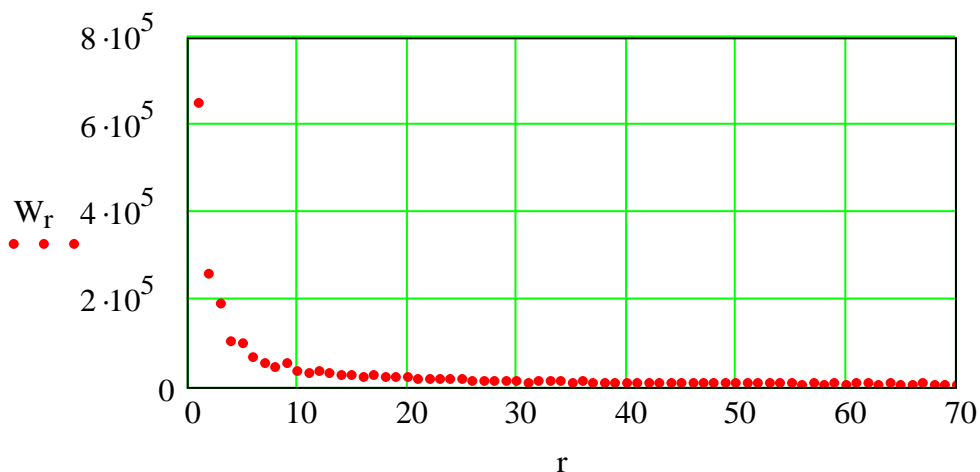
$$\beta = 1.193$$

Генерирование рангового параметрического распределения математических ожиданий

Вектор математических ожиданий параметров играет ключевую роль, так как определяет структуру создаваемой модели. Для его формирования используется подпрограмма, основанная на стандартной двухпараметрической зависимости, входными аргументами которой являются: n – количество объектов, $W1$ – первая точка, β – ранговый коэффициент.

$$F(n, W1, \beta) := \left\{ \begin{array}{l} \text{for } i \in 1..n \\ w_i \leftarrow \text{round} \left[\frac{W1}{i^\beta} \cdot (1 + \text{rnd}(0.5)) \right] \\ w \end{array} \right. \quad \begin{array}{l} W := F(n, W1, \beta) \\ r := 1.. \text{length}(W) \end{array}$$

Визуализируем результаты работы подпрограммы.



Ранговое параметрическое распределение математических ожиданий: абсцисса – ранг объекта; ордината – значение параметра

Моделирование фактора цикличности в генерируемых данных

Наличие цикличности в данных, описывающих изменение параметра во времени, проявляется наличием сильной корреляционной связи между отрезками ряда, разделенными определенным интервалом времени. В реальных техноценозах ширина данного интервала может быть определена периодичностью воздействия сезонных (или каких-либо других) факторов погоды либо особенностью технологического процесса.

Основой для задания циклической сезонности в данных является вектор индексов сезонности, нормированные значения элементов которого отражают среднестатистические изменения параметра внутри интервала сезонности. Ниже задан вектор месячных индексов сезонности, отражающий внутригодовую динамику изменения функционального параметра, характерную для многих техноценозов средней полосы России (см. рис. ниже). Следует иметь в виду, что вектор индексов сезонности, при необходимости, может быть скорректирован самим исследователем. Для этого необходимо осуществлять изменения данных ниже в таблице.

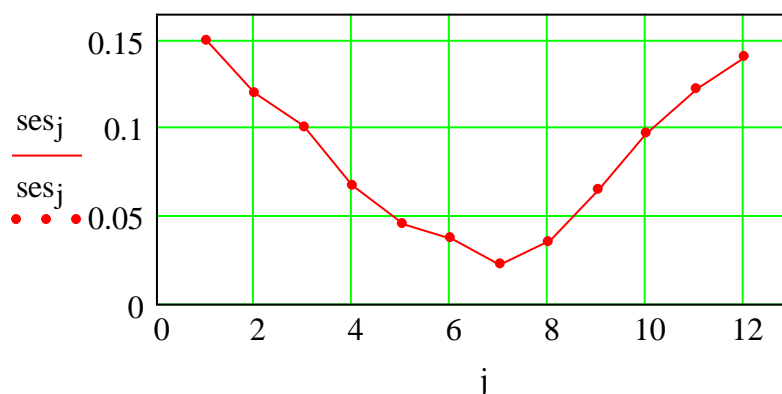
$$S :=$$

| Янв. | Февр. | Март | Апр. | Май | Июнь | Июль | Авг. | Сент. | Окт. | Нояб. | Дек. |
|------|-------|------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| 0,15 | 0,12 | 0,1 | 0,067 | 0,045 | 0,037 | 0,022 | 0,035 | 0,065 | 0,097 | 0,122 | 0,14 |

$$\sum S = 1$$

$$ses := S^T$$

$$j := 1..length(ses)$$



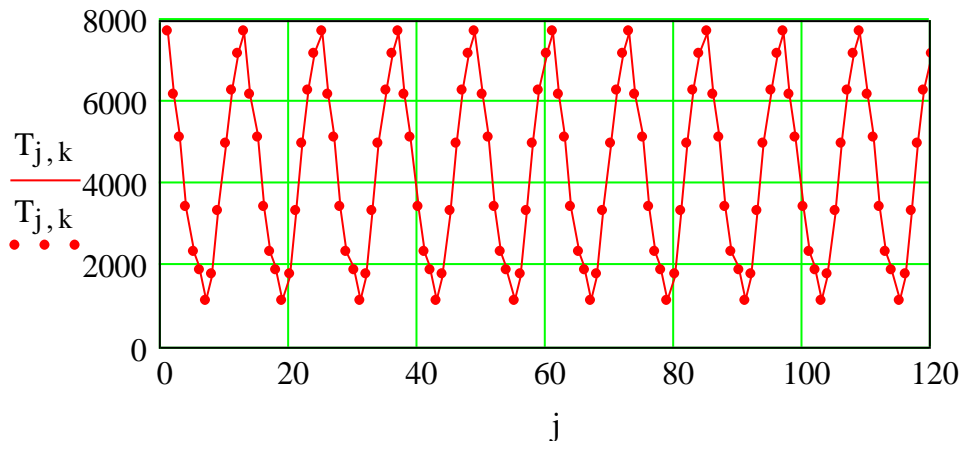
Месячные индексы сезонности электропотребления: абсцисса – номер месяца; ордината – значение индекса

Ниже приведена подпрограмма, формирующая матрицу данных с фактором сезонности путем автоматического рекуррентного переприсвоения значений электропотребления за пределами интервала сезонности (визуализация результатов работы данной подпрограммы – см. рис. ниже). Входные аргументы подпрограммы: W – вектор математических ожиданий, ses – вектор сезонных индексов, gor – горизонт генерируемых данных. Горизонт генерируемых данных представляет собой общее количество месяцев, которое определяется как результат произведения двенадцати (число месяцев в году) на количество лет (в данном случае – 10).

```
t := 10
gor := 12·t
gor = 120
```

```
Ses_vec(W, ses, gor) :=
  q ← ses·WT
  for j ∈ 1..cols(q)
    b ← q⟨j⟩
    for i ∈ 1..gor
      ai ← bi if i ≤ 12
      ai ← ai-12 otherwise
    m⟨j⟩ ← a
  m
```

```
T := Ses_vec(W, ses, gor)
j := 1..rows(T)
k := 7
```



Вектор сезонных данных для k-го объекта: абсцисса – номер месяца на десятилетнем интервале; ордината – сгенерированное значение параметра

Прибавление к генерируемым данным полиномиального тренда со случайными параметрами

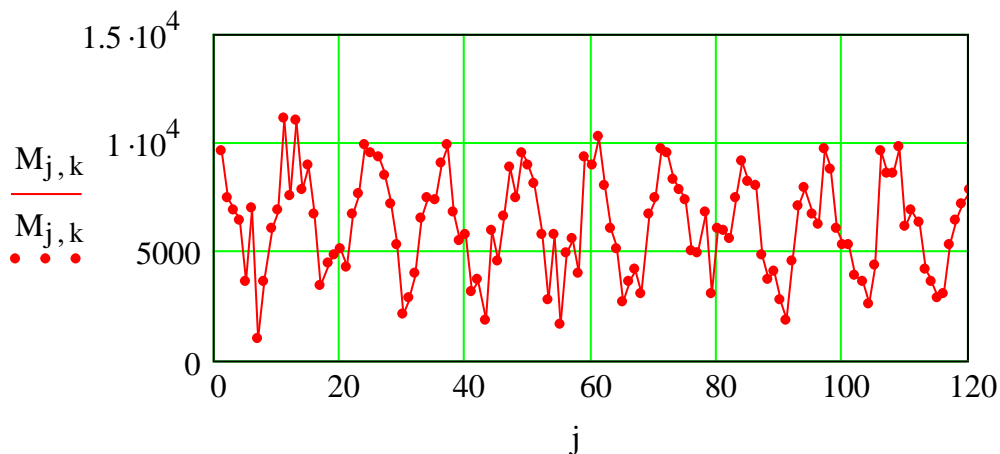
Ниже приведена подпрограмма, при помощи которой к матрице первоначально сгенерированных данных по электропотреблению «Т» прибавляются полиномиальный тренд и шумовая компонента. Тренд формирует-

ся на основе полинома второго порядка, коэффициенты которого определяются с помощью датчика случайных чисел (см. рис. ниже).

```

Trend(T) :=
  for i ∈ 1..cols(T)
    k ← -1
    while k < 0
      a ← T<i>
      b ← 0.02·mean(a)
      b1 ← rnd(b)
      b2 ← rnd(b·0.04)
      for j ∈ 1..rows(T)
        S1j,i ← j·b1 - (j2·b2) + Tj,i
      S<i> ← S1<i> + rnorm(rows(T), b·20, 15·b)
      k ← min(S<i>)
    S
M := Trend(T)

```



Случайный процесс изменения параметра с сезонностью и трендом для k -го объекта: абсцисса – номер месяца; ордината – параметр

Визуализация сгенерированных данных

Ниже приведена подпрограмма, которая путем суммирования внутри каждого интервала сезонности из матрицы месячных данных «М» формирует матрицу годовых данных «Y» (с учетом определенных параметров).

$$\begin{array}{l}
 \text{YR(M)} := \left| \begin{array}{l}
 \text{for } j \in 1.. \text{cols(M)} \\
 \left| \begin{array}{l}
 k \leftarrow 1 \\
 b \leftarrow M^{\langle j \rangle} \\
 \text{for } i \in 0, 12.. \text{length}(b) - 12 \\
 \left| \begin{array}{l}
 d_k \leftarrow \sum \text{submatrix}(b, i + 1, i + 12, 1, 1) \\
 k \leftarrow k + 1
 \end{array} \right. \\
 c^{\langle j \rangle} \leftarrow d
 \end{array} \right. \\
 c
 \end{array} \right.
 \end{array}$$

$$Y := \text{YR(M)}$$

$$Y =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 1 | $1.009 \cdot 10^6$ | $4.072 \cdot 10^5$ | $3.071 \cdot 10^5$ | $1.506 \cdot 10^5$ | $1.429 \cdot 10^5$ | $9.835 \cdot 10^4$ | $6.868 \cdot 10^4$ |
| 2 | $1.186 \cdot 10^6$ | $4.055 \cdot 10^5$ | $3.011 \cdot 10^5$ | $1.731 \cdot 10^5$ | $1.429 \cdot 10^5$ | $1.003 \cdot 10^5$ | $8.379 \cdot 10^4$ |
| 3 | $1.301 \cdot 10^6$ | $4.643 \cdot 10^5$ | $3.523 \cdot 10^5$ | $1.94 \cdot 10^5$ | $1.508 \cdot 10^5$ | $9.26 \cdot 10^4$ | $9.189 \cdot 10^4$ |
| 4 | $1.324 \cdot 10^6$ | $4.465 \cdot 10^5$ | $3.502 \cdot 10^5$ | $2.028 \cdot 10^5$ | $1.614 \cdot 10^5$ | $9.106 \cdot 10^4$ | $8.528 \cdot 10^4$ |
| 5 | $1.367 \cdot 10^6$ | $5.293 \cdot 10^5$ | $3.679 \cdot 10^5$ | $2.113 \cdot 10^5$ | $1.733 \cdot 10^5$ | $8.743 \cdot 10^4$ | $8.417 \cdot 10^4$ |
| 6 | $1.468 \cdot 10^6$ | $4.65 \cdot 10^5$ | $3.876 \cdot 10^5$ | $1.964 \cdot 10^5$ | $1.719 \cdot 10^5$ | $8.475 \cdot 10^4$ | $8.964 \cdot 10^4$ |
| 7 | $1.41 \cdot 10^6$ | $4.7 \cdot 10^5$ | $3.887 \cdot 10^5$ | $1.913 \cdot 10^5$ | $1.643 \cdot 10^5$ | $9.066 \cdot 10^4$ | $8.774 \cdot 10^4$ |
| 8 | $1.392 \cdot 10^6$ | $4.844 \cdot 10^5$ | $3.584 \cdot 10^5$ | $2.102 \cdot 10^5$ | $1.758 \cdot 10^5$ | $8.329 \cdot 10^4$ | $9.258 \cdot 10^4$ |
| 9 | $1.323 \cdot 10^6$ | $4.548 \cdot 10^5$ | $3.275 \cdot 10^5$ | $1.76 \cdot 10^5$ | $1.703 \cdot 10^5$ | $7.957 \cdot 10^4$ | $9.212 \cdot 10^4$ |
| 10 | $1.226 \cdot 10^6$ | $4.617 \cdot 10^5$ | $3.531 \cdot 10^5$ | $1.543 \cdot 10^5$ | $1.741 \cdot 10^5$ | $7.386 \cdot 10^4$ | $8.853 \cdot 10^4$ |

Столбцы полученной матрицы «Y» соответствуют объектам техноценоза, а строки – временным интервалам (здесь – годам).

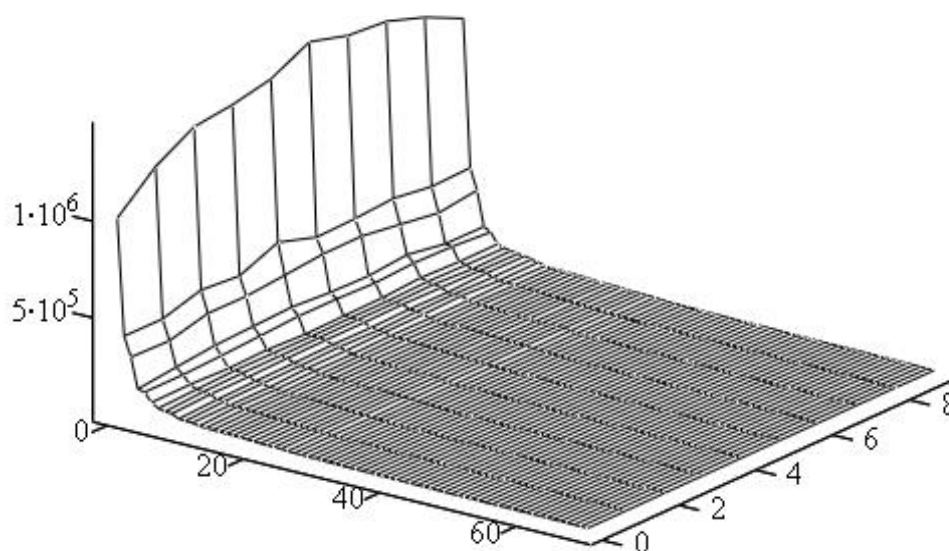
Подпрограмма Zipf предназначена для формирования матрицы табулированных ранговых параметрических распределений техноценоза по параметру электропотребления. При этом для удобства визуализации матрица транспонируется, а на каждом временном интервале объекты техноценоза ранжируются по значению параметра электропотребления.

$$\begin{array}{l}
 \text{Zipf(Y)} := \left| \begin{array}{l}
 \text{for } i \in 1.. \text{cols(Y)} \\
 b^{\langle i \rangle} \leftarrow \text{reverse}(\text{sort}(Y^{\langle i \rangle}))
 \end{array} \right. \\
 b \\
 Z := \text{Zipf}(Y^T)
 \end{array}$$

$$Z =$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 1 | $1.009 \cdot 10^6$ | $1.186 \cdot 10^6$ | $1.301 \cdot 10^6$ | $1.324 \cdot 10^6$ | $1.367 \cdot 10^6$ | $1.468 \cdot 10^6$ | $1.41 \cdot 10^6$ |
| 2 | $4.072 \cdot 10^5$ | $4.055 \cdot 10^5$ | $4.643 \cdot 10^5$ | $4.465 \cdot 10^5$ | $5.293 \cdot 10^5$ | $4.65 \cdot 10^5$ | $4.7 \cdot 10^5$ |
| 3 | $3.071 \cdot 10^5$ | $3.011 \cdot 10^5$ | $3.523 \cdot 10^5$ | $3.502 \cdot 10^5$ | $3.679 \cdot 10^5$ | $3.876 \cdot 10^5$ | $3.887 \cdot 10^5$ |
| 4 | $1.506 \cdot 10^5$ | $1.731 \cdot 10^5$ | $1.94 \cdot 10^5$ | $2.028 \cdot 10^5$ | $2.113 \cdot 10^5$ | $1.964 \cdot 10^5$ | $1.913 \cdot 10^5$ |
| 5 | $1.429 \cdot 10^5$ | $1.429 \cdot 10^5$ | $1.508 \cdot 10^5$ | $1.614 \cdot 10^5$ | $1.733 \cdot 10^5$ | $1.719 \cdot 10^5$ | $1.643 \cdot 10^5$ |
| 6 | $9.835 \cdot 10^4$ | $1.003 \cdot 10^5$ | $9.26 \cdot 10^4$ | $9.81 \cdot 10^4$ | $8.796 \cdot 10^4$ | $9.201 \cdot 10^4$ | $9.483 \cdot 10^4$ |
| 7 | $7.695 \cdot 10^4$ | $8.379 \cdot 10^4$ | $9.189 \cdot 10^4$ | $9.106 \cdot 10^4$ | $8.743 \cdot 10^4$ | $9.139 \cdot 10^4$ | $9.066 \cdot 10^4$ |

На основе данных матрицы табулированных ранговых параметрических распределений по электропотреблению «Z» строится ранговая поверхность, по которой может осуществляться визуальный анализ.



Трехмерная ранговая поверхность, построенная по годовым сгенерированным данным: абсцисса – ранг объекта; ордината – временной интервал (номер года); аппликата – значение параметра

Сгенерированные данные для последующего рангового анализа сохраняются в файлах Excel, расположенных в директории «с:\mathcad_dat»:



C:\mathcad_dat\g_year.xls – годовые данные;

У



C:\mathcad_dat\g_mounth.xls – месячные данные.

М

[на содержание]