

# GZ-АНАЛИЗ РАНГОВОГО ПАРАМЕТРИЧЕСКОГО РАСПРЕДЕЛЕНИЯ ПО ЭЛЕКТРОПОТРЕБЛЕНИЮ

В.И. Гнатюк

Оптимальное управление электропотреблением объектов техноценоза [1] на системном уровне осуществляется в рамках связанной методики, включающей ряд этапов. На этапе статистического анализа осуществляется обработка данных по электропотреблению, которая включает взаимосвязанные процедуры рангового анализа (формирования базы данных, интервального оценивания, прогнозирования и нормирования) (рис. 1) [2-6].

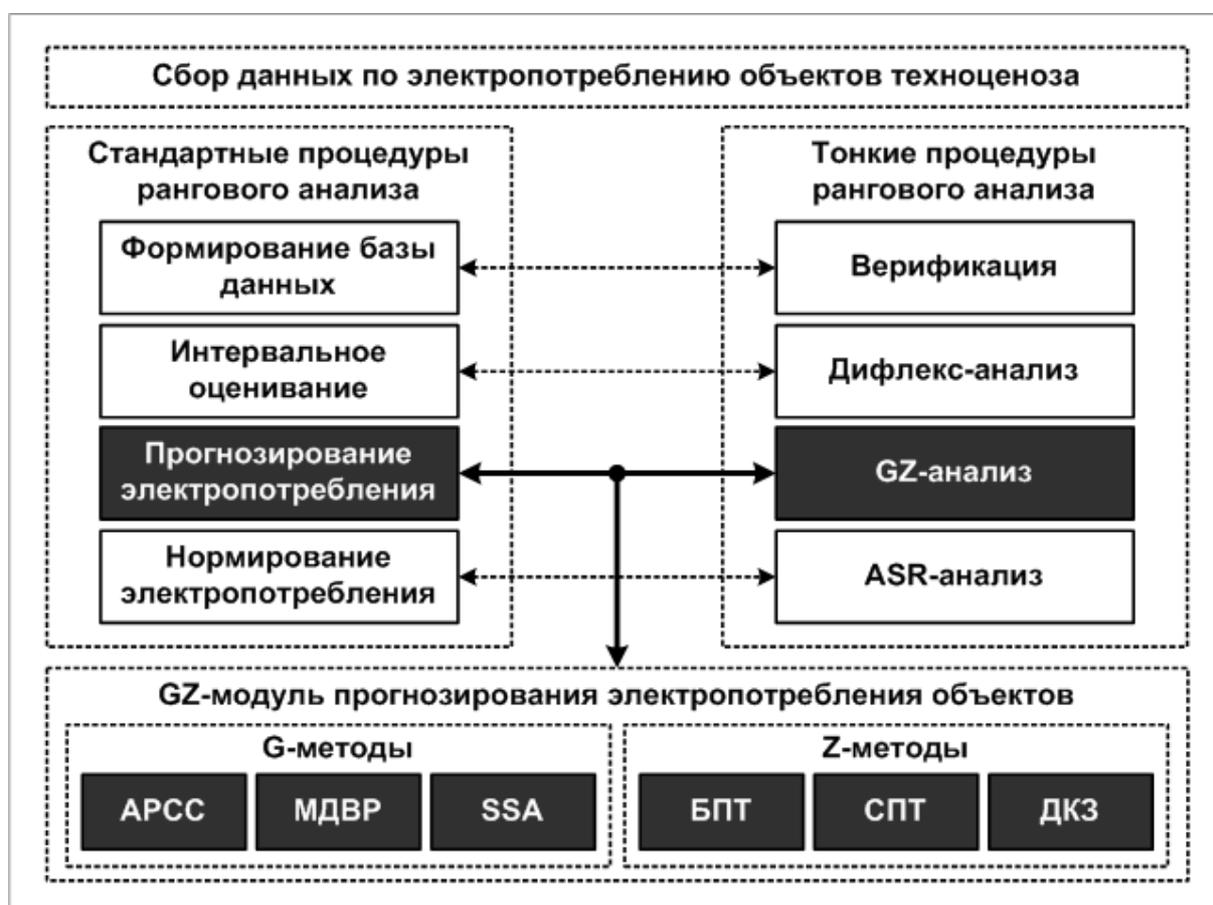


Рис. 1. Стандартные и тонкие процедуры рангового анализа:  
ARCC – модель авторегрессионного скользящего среднего;  
MDVP – модель декомпозиции временного ряда;  
SSA – модель анализа сингулярного спектра временного ряда;  
BPT – модель без фиксированной первой точки;  
SPT – модель с фиксированной первой точкой;  
DKZ – модель с делением на кастовые зоны

С целью повышения точности расчетов стандартные процедуры рангового анализа дополняются соответствующими тонкими процедурами: верификацией базы данных, а также дифлекс-, GZ- и ASR-анализом рангового параметрического распределения (рис. 1) [4-7]. Рассмотрим центральную тонкую процедуру рангового анализа, существенно уточняющую процедуру прогнозирования – GZ-анализ (рис. 2).



Рис. 2. Варианты реализации GZ-анализа

Прогнозирование – процедура оптимального управления ресурсами техноценоза, заключающаяся в определении вероятных значений функциональных параметров в будущем. Прогнозирование может выполняться на основе статической модели, отражающей процесс электропотребления на год вперед. Динамическое стохастическое моделирование позволяет осуществлять прогноз на среднесрочную перспективу (5 – 7 лет) [5-7]. Применительно к техноценозу прогнозирование может осуществляться G-методами (Gauss-методами, основанными на гауссовой математической статистике), Z-методами (Zipf-методами, основанными на ципфовой математической статистике) и синтетическими GZ-методами, органично сочетающими их достоинства. Последние предполагают выполнение предварительной тонкой процедуры GZ-анализа (Gauss-Zipf analysis).

GZ-анализ может реализовываться в двух вариантах, первый из которых является упрощенным эвристическим, а второй – основным критериальным (рис. 2). Эвристический вариант как основной метод прогнозирования применяется только для оценочных расчетов, причем на сравнительно небольших базах данных. Кроме того, с помощью него на предварительном этапе синтезируется так называемая GZ-матрица методов прогнозирования, необходимая для процедуры юстировки ключевых параметров критериального варианта GZ-анализа.

На рисунке 3 приведена упрощенная структура прогнозной базы данных техноценоза по электропотреблению.

Ранг	Электропотребление объектов по временным интервалам, кВтч										
	...	t-8	t-7	t-6	t-5	t-4	t-3	t-2	t-1	t	t+1
1	...	$W_{18}$	$W_{17}$	$W_{16}$	$W_{15}$	$W_{14}$	$W_{13}$	$W_{12}$	$W_{11}$	$W_{10}$	?
2	...	$W_{28}$	$W_{27}$	$W_{26}$	$W_{25}$	$W_{24}$	$W_{23}$	$W_{22}$	$W_{21}$	$W_{20}$	?
3	...	$W_{38}$	$W_{37}$	$W_{36}$	$W_{35}$	$W_{34}$	$W_{33}$	$W_{32}$	$W_{31}$	$W_{30}$	?
4	...	$W_{48}$	$W_{47}$	$W_{46}$	$W_{45}$	$W_{44}$	$W_{43}$	$W_{42}$	$W_{41}$	$W_{40}$	?
5	...	$W_{58}$	$W_{57}$	$W_{56}$	$W_{55}$	$W_{54}$	$W_{53}$	$W_{52}$	$W_{51}$	$W_{50}$	?
6	...	$W_{68}$	$W_{67}$	$W_{66}$	$W_{65}$	$W_{64}$	$W_{63}$	$W_{62}$	$W_{61}$	$W_{60}$	?
7	...	$W_{78}$	$W_{77}$	$W_{76}$	$W_{75}$	$W_{74}$	$W_{73}$	$W_{72}$	$W_{71}$	$W_{70}$	?
.....	...	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
n-1	...	$W_{(n-1)8}$	$W_{(n-1)7}$	$W_{(n-1)6}$	$W_{(n-1)5}$	$W_{(n-1)4}$	$W_{(n-1)3}$	$W_{(n-1)2}$	$W_{(n-1)1}$	$W_{(n-1)0}$	?
n	...	$W_{n8}$	$W_{n7}$	$W_{n6}$	$W_{n5}$	$W_{n4}$	$W_{n3}$	$W_{n2}$	$W_{n1}$	$W_{n0}$	?

Рис. 3. Структура прогнозной базы данных по электропотреблению:

$W_{km}$  – электропотребление k-го объекта техноценоза на (t – m)-ом временном интервале (месяце, году)

Из базы данных предварительно осуществляется выделение следующих информационных подсистем. Фактические известные данные по электропотреблению на текущем временном интервале составляют «Вектор верификации». Прогнозируемые данные на будущем временном интервале определяются как «Вектор прогнозирования». Все остальные известные данные образуют «Матрицу данных» [6].

Процесс прогнозирования электропотребления объектов техноценоза с помощью эвристического варианта GZ-анализа реализуется следующим образом (рис. 2 – 4). На первом этапе в качестве базы прогнозирования используется матрица данных, применительно к которой реализуются последовательно все имеющиеся в распоряжении методы прогнозирования. Статистическое сравнение полученных прогнозных результатов с соответствующими данными вектора верификации позволяет для каждого из объектов определить наиболее эффективный метод. Затем вектор верификации присоединяется к матрице данных и осуществляется окончательный прогноз электропотребления, причем процедура для каждого объекта осуществляется именно тем методом, который на первом этапе был определен для него как наиболее эффективный. В базе данных прогнозирования могут использоваться самые различные методы (как G-, так и Z-). В любом случае в процессе реализации эвристического варианта GZ-анализа выбирается наиболее эффективный из них [4-7].

**БАЗА ДАННЫХ ПО ЭЛЕКТРОПОТРЕБЛЕНИЮ (кВт·ч) Год**

	1	2	3	4	5	6	7	8	
<b>Ранг</b>	1	2	3	4	5	6	7	8	
W =	1	$3.5 \cdot 10^6$	$3.055 \cdot 10^6$	$3.01 \cdot 10^6$	$3.645 \cdot 10^6$	$3.295 \cdot 10^6$	$3.125 \cdot 10^6$	$3.127 \cdot 10^6$	$3.129 \cdot 10^6$
	2	$1.81 \cdot 10^6$	$2.2 \cdot 10^6$	$2.512 \cdot 10^6$	$2.84 \cdot 10^6$	$2.708 \cdot 10^6$	$2.559 \cdot 10^6$	$2.227 \cdot 10^6$	$2.248 \cdot 10^6$
	3	$1.712 \cdot 10^6$	$1.447 \cdot 10^6$	$1.613 \cdot 10^6$	$1.614 \cdot 10^6$	$1.304 \cdot 10^6$	$1.775 \cdot 10^6$	$1.624 \cdot 10^6$	$1.829 \cdot 10^6$
	4	$1.553 \cdot 10^6$	$1.34 \cdot 10^6$	$1.474 \cdot 10^6$	$1.351 \cdot 10^6$	$1.211 \cdot 10^6$	$1.628 \cdot 10^6$	$1.617 \cdot 10^6$	$1.747 \cdot 10^6$
	5	$1.268 \cdot 10^6$	$1.332 \cdot 10^6$	$1.325 \cdot 10^6$	$1.231 \cdot 10^6$	$1.162 \cdot 10^6$	$1.061 \cdot 10^6$	$1.061 \cdot 10^6$	$1.063 \cdot 10^6$
	6	$7.451 \cdot 10^5$	$1.015 \cdot 10^6$	$8.373 \cdot 10^5$	$7.921 \cdot 10^5$	$8.534 \cdot 10^5$	$7.188 \cdot 10^5$	$7.479 \cdot 10^5$	$7.898 \cdot 10^5$
	7	$7.438 \cdot 10^5$	$8.567 \cdot 10^5$	$7.094 \cdot 10^5$	$5.673 \cdot 10^5$	$4.366 \cdot 10^5$	$4.678 \cdot 10^5$	$4.826 \cdot 10^5$	$4.827 \cdot 10^5$
	8	$5.817 \cdot 10^5$	$7.032 \cdot 10^5$	$6.932 \cdot 10^5$	$4.326 \cdot 10^5$	$4.04 \cdot 10^5$	$4.101 \cdot 10^5$	$4.107 \cdot 10^5$	$4.113 \cdot 10^5$
	9	$4.386 \cdot 10^5$	$6.371 \cdot 10^5$	$5.784 \cdot 10^5$	$4.025 \cdot 10^5$	$3.971 \cdot 10^5$	$4.01 \cdot 10^5$	$3.928 \cdot 10^5$	$3.141 \cdot 10^5$
	10	$4.02 \cdot 10^5$	$5.244 \cdot 10^5$	$5.179 \cdot 10^5$	$3.652 \cdot 10^5$	$3.858 \cdot 10^5$	$3.873 \cdot 10^5$	$3.657 \cdot 10^5$	$2.964 \cdot 10^5$

Рис. 4. Данные по электропотреблению, кВт·ч (фрагмент из [6])

Для прогнозирования электропотребления объектов G-методами в основном используются модели авторегрессионного скользящего среднего (см. рис. 5 для примера, показанного на рис. 4), декомпозиции временного ряда, а также различные вариации методов на основе анализа сингулярного спектра траекторной матрицы временного ряда [8]. В процессе прогнозирования электропотребления техноценоза Z-методами должны учитываться техноценологические свойства, сводящиеся в конечном итоге к понятию устойчивости гиперболических ранговых параметрических распределений (см. рис. 6 для примера, показанного на рис. 4). Здесь, как правило, находят применение методы без фиксированной первой точки, с фиксированной первой точкой, а также с делением на кастовые зоны. Полная совокупность методов составляет GZ-модуль прогнозирования [4-7].

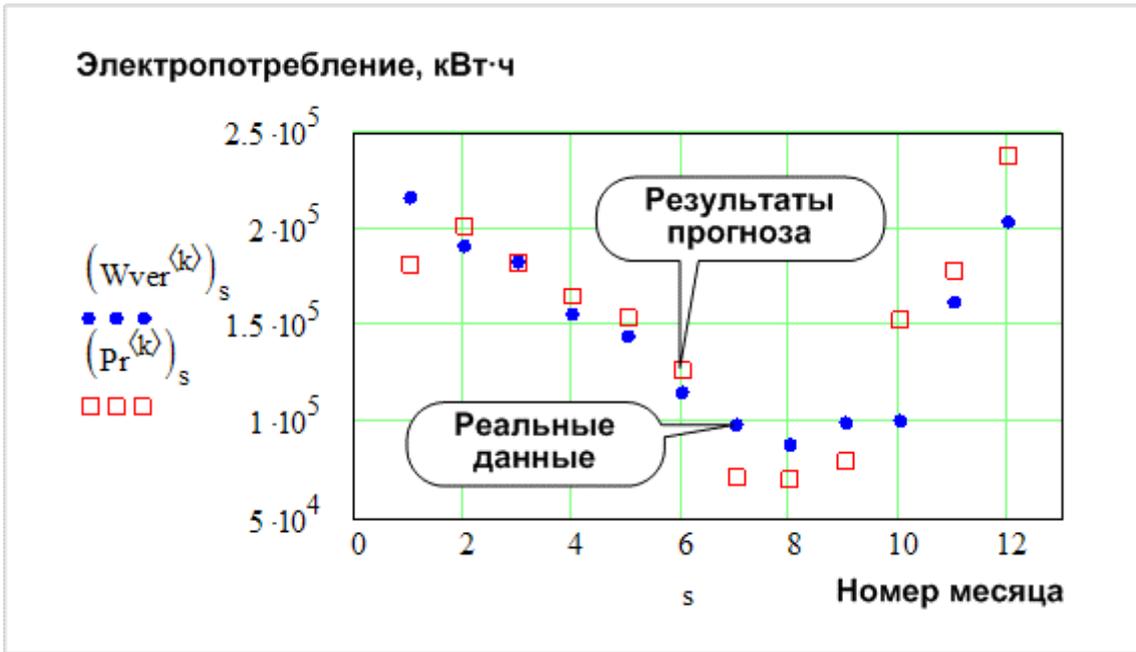


Рис. 5. Пример осуществления процедуры прогнозирования одного из объектов техноценоза G-методом [6]

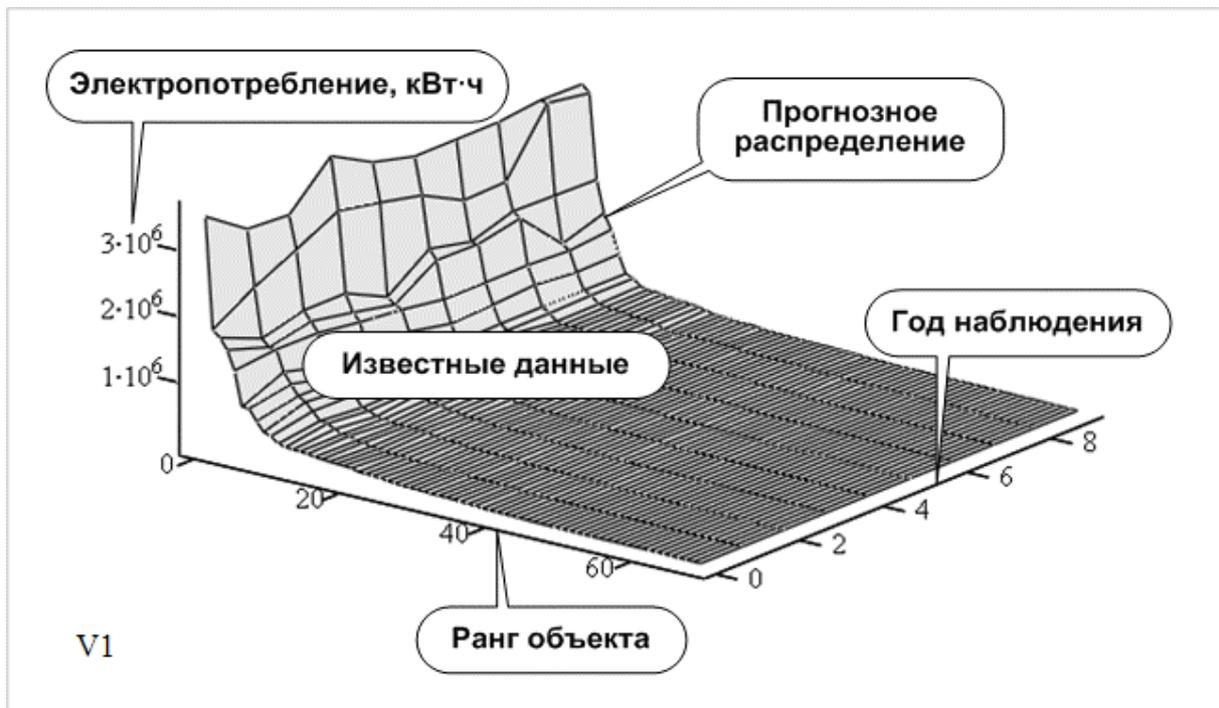


Рис. 6. Пример прогноза электропотребления техноценоза Z-методом: последняя гиперболическая кривая – прогноз на девятый год

Критериальный вариант GZ-анализа (рис. 2) позволяет еще до начала собственно процедуры прогнозирования осуществить выбор G- или Z-методологии, что существенно ускоряет расчеты и повышает их точность (положительный эффект пропорционален размерам базы данных). Как следует из закона оптимального построения техноценозов, в качестве критерия выбора метода следует рассматривать соотношение объемов системного и гауссового ресурсов кластеров техноценоза (рис. 7) [5,6].

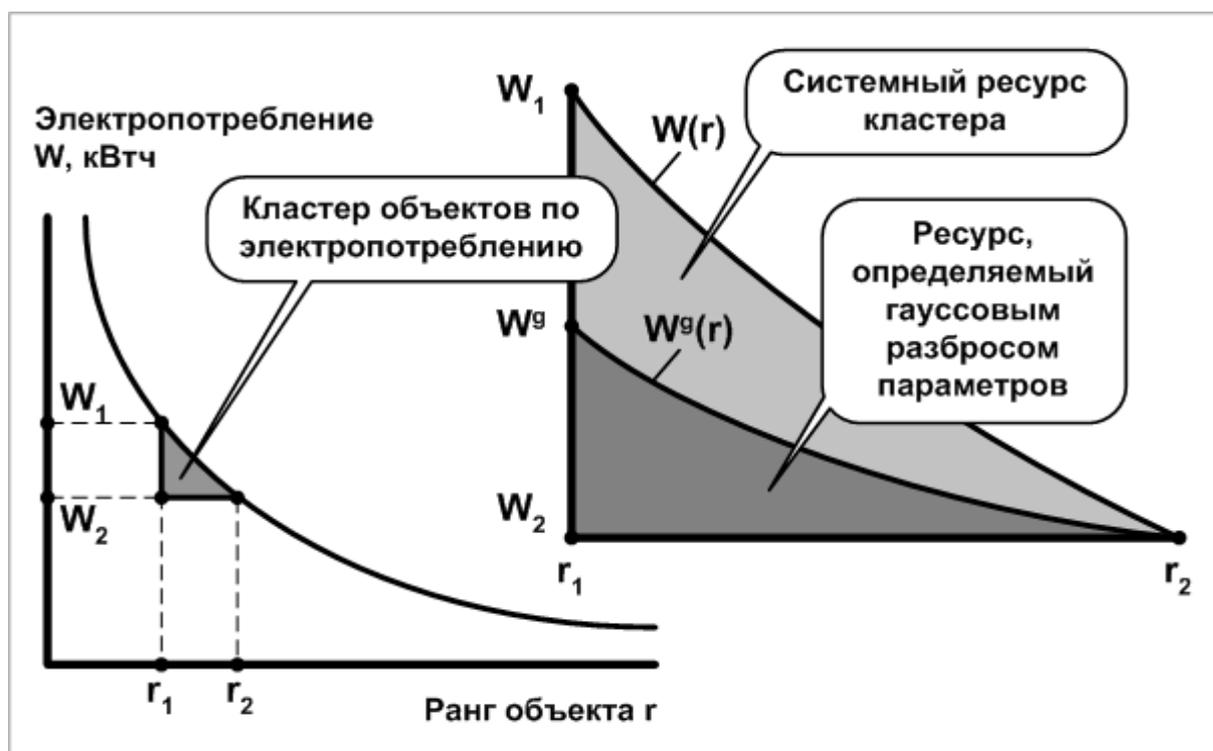


Рис. 7. К вопросу априорного выбора G- или Z-метода прогнозирования по критерию соотношения системного и гауссового ресурсов

Как следует из уравнений закона оптимального построения техноценозов, гауссовый ресурс кластера равен [5,6]:

$$W_G = \left( \int_{r_1}^{r_2} W^g(r) dr \right) - ((r_2 - r_1) W_2), \quad (1)$$

- где
- $W(r)$  – ранговое параметрическое распределение техноценоза по электропотреблению;
  - $W^g(r)$  – гауссовое распределение, соответствующее кластерному распределению параметров в ранговой дифференциальной форме;
  - $W_2$  – значение электропотребления, соответствующее правой ранговой границе кластера  $r_2$ .

При этом системный ресурс кластера техноценоза определяется следующим образом [5,6]:

$$W_Z = \int_{r_1}^{r_2} (W(r) - W^g(r)) dr. \quad (2)$$

Как показывает анализ, процедура кластеризации ранговых параметрических распределений по исследуемому параметру существенно затруднена из-за негауссовости распределений, что неотвратно ведет к негауссовости кластеров. При этом нарушается главный минимаксный критерий кластер-анализа (то, что статистически внутри кластера функциональные параметры объектов должны распределяться по нормальному закону). Выход из этого положения возможен в результате увеличения количества кластеров (стремлении размера кластера к нулю), при этом системный и гауссовый ресурсы кластеров в пределе сводятся соответственно к системному и гауссовому доверительным интервалам объектов и рангов [6].

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) [4,6]. Для определения ширины доверительного интервала используется понятие интерквартильного размаха применительно к выборке значений электропотребления, соответствующих данному рангу на протяжении ряда временных интервалов [6]:

$$\Delta W_Z = W_{0,75}^q - W_{0,25}^q, \quad (3)$$

где  $W_{0,75}^q$  – верхний квартиль распределения значений электропотребления (квантиль порядка 0,75);  
 $W_{0,25}^q$  – нижний квартиль (квантиль порядка 0,25).

Квартили распределения определяются как значение электропотребления, при котором функция распределения становится равной соответствующей величине (0,75 или 0,25). Если исходить из того, что распределение значений электропотребления для одного ранга на протяжении ряда временных интервалов является нормальным [7], то ширина интерквартильного размаха может быть определена на основе решения относительно переменной  $\Delta W_Z$  следующего уравнения [6]:

$$\frac{\Delta W_Z / 2}{\sigma} = \Phi^{-1}(p_d / 2), \quad (4)$$

- где  $\Delta W_Z / 2$  – ширина системного доверительного интервала в одну сторону от математического ожидания (в расчетах принимается эмпирическое среднее);
- $\sigma$  – среднеквадратичное отклонение экспериментальных точек от математического ожидания (в расчетах принимается эмпирический стандарт);
- $\Phi^{-1}(\tau)$  – обратная функция Лапласа;
- $p_d$  – априорно принимаемая доверительная вероятность (для получения интерквартильного размаха должна быть принята равной 0,95).

Для моделирования функции распределения нормального распределения в (4) применяется стандартная функция Лапласа:

$$\Phi(\tau) = \frac{1}{2\pi} \int_0^{\tau} e^{-x^2/2} dx, \quad (5)$$

где  $x$  – формальная переменная интегрирования.

Системный доверительный интервал является мерилем системного ресурса объекта, «занимающего» в рассматриваемый момент времени определенный ранг на ранговом параметрическом распределении, и характеризует разброс параметров техноценоза, при котором его функционирование можно считать нормальным (рис. 8). Фактически этот интервал отражает требования или ограничения, которые система выдвигает объектам, чтобы обеспечить свое устойчивое инерционное функционирование. Если эмпирическое значение параметра объекта (в частности – электропотребления) выходит за пределы системного интервала, то это означает, что объект не подчиняется системным требованиям и в этом смысле проявляет индивидуальность. Сама же величина отклонения может рассматриваться как мера индивидуальности данного объекта.

Ширина системного доверительного интервала определяется предысторией развития техноценоза, будучи взятого в целом, и если она невелика, то это свидетельствует о том, что техноценоз развивается стабильно и сбалансировано, все изменения в нем происходят плавно. Напротив, широкий интервал свидетельствует о резких структурных изменениях. Методика определения системного интервала позволяет противопоставить его гауссовому доверительному интервалу, который определяется применительно не к рангу распределения, а к объекту техноценоза.

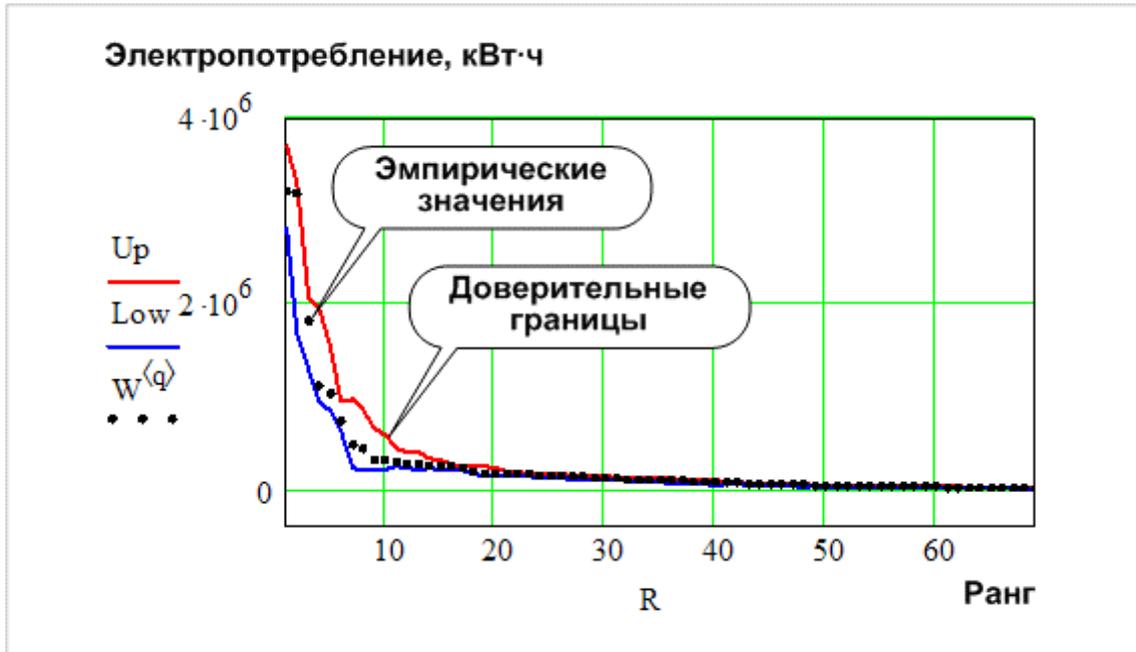


Рис. 8. Пример системного доверительного интервала

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) [4,6]. Для определения ширины гауссового доверительного интервала также используется понятие интерквартильного размаха применительно к выборке значений электропотребления объекта на протяжении ряда временных интервалов [6]:

$$\Delta W_G = W_{0,75}^q - W_{0,25}^q. \quad (6)$$

Если исходить из того, что распределение значений электропотребления для одного объекта на протяжении ряда временных интервалов также является нормальным [7], то ширина интерквартильного размаха может быть определена на основе решения относительно переменной  $\Delta W_G$  следующего уравнения, аналогичного (4) [6]:

$$\frac{\Delta W_G / 2}{\sigma} = \Phi^{-1}(p_d / 2). \quad (7)$$

Гауссовый доверительный интервал является мерилем гауссового ресурса объекта на ранговом параметрическом распределении и характеризует разброс параметров объекта, при которых его функционирование можно считать нормальным (независимо от поведения техноценоза). 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 (осуществлена модернизация оборудования, внедрены новые технологии) либо система предъявила новые требования, для выполнения которых объект вынужден резко изменить свой режим функционирования (рис. 9).

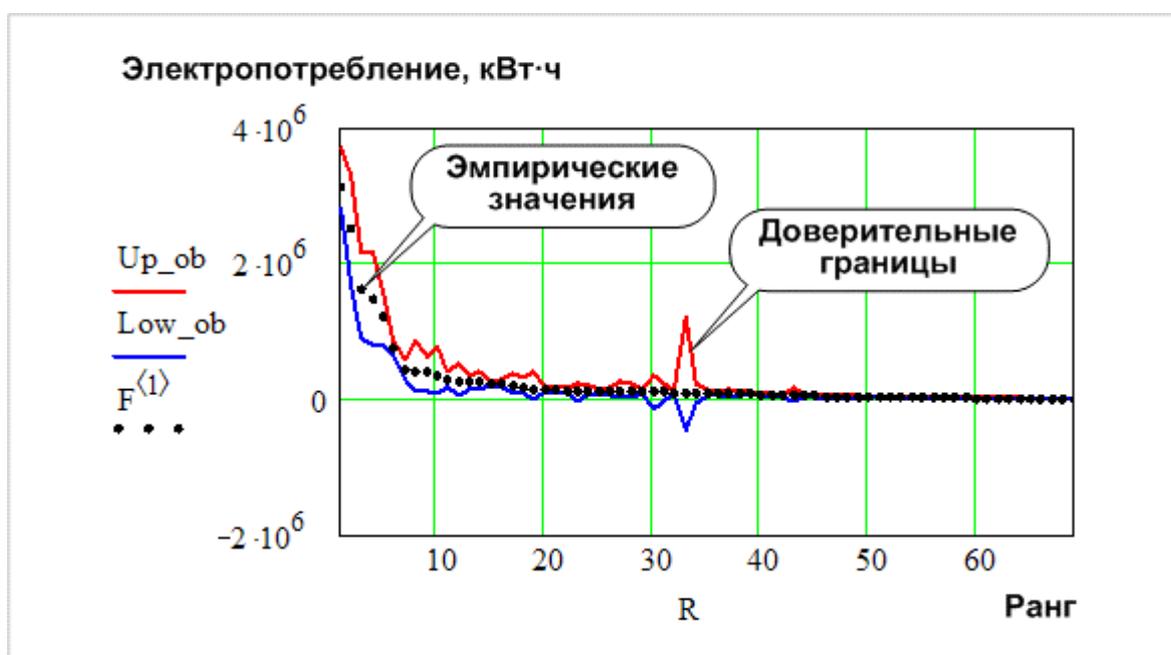


Рис. 9. Пример гауссового доверительного интервала

Таким образом, для каждого объекта можно получить отношение системного и гауссового доверительных интервалов, которое называется коэффициентом когерентности и показывает степень согласованности поведения объекта по отношению к техноценозу [4,6]:

$$K_{GZ} = \lim_{KK \rightarrow KO} \frac{W_Z}{W_G} \cong \frac{\Delta W_Z}{\Delta W_G}, \quad (8)$$

где KK — количество кластеров техноценоза;  
 KO — количество объектов техноценоза.

Итак, теоретически коэффициент когерентности может быть определен как предел отношения системного ресурса кластера (выражение (2)) к

его гауссовому ресурсу (1) при условии сужения ширины кластера до нуля (устремления количества кластеров к общему количеству объектов техноценоза). Это проиллюстрировано на рисунке 7. Эмпирически же, как указано выше, коэффициент когерентности может быть определен как отношение системного доверительного интервала (см. выражение (3)) к гауссовому (выражение (6)). Если коэффициент когерентности близок к единице (то есть системный и гауссовый интервалы примерно равны), то можно говорить о согласованном поведении данного объекта и техноценоза.

Дальнейший более глубокий GZ-анализ техноценоза заключается в определении так называемого кумулятивного когерент-фактора (coherent factor), который равен отношению суммарного системного доверительного интервала всех объектов техноценоза к соответствующему суммарному гауссовому доверительному интервалу:

$$K_{\Sigma} = \frac{\sum_{i=1}^n \Delta W_{Zi}}{\sum_{i=1}^n \Delta W_{Gi}}, \quad (9)$$

где  $n$  – количество объектов техноценоза.

Когерент-фактор техноценоза показывает, в какой степени его системное поведение согласовано с индивидуальным поведением объектов по отдельности (под поведением здесь, безусловно, понимаются свойства параметрических временных рядов объектов и техноценоза). Затем могут быть получены динамические когерент-функции (coherent function):

$$\begin{cases} K_{\Sigma}(t); \\ K_{GZi}(t); \\ i = 1 \dots n, \end{cases} \quad (10)$$

где  $t$  – время функционирования техноценоза.

Когерент-функции позволяют оценивать в любой момент времени и прогнозировать изменение в будущем динамических свойств как техноценоза в целом, так и его объектов в частности.

Когерент-параметры также могут существенно оптимизировать процесс прогнозирования в техноценозе. В частности, как показывают исследования и реализация на практике, коэффициент когерентности является индикатором выбора наиболее эффективного метода прогнозирования для

рассматриваемого объекта. При сравнительно больших значениях коэффициента лучше работают G-методы, в противном случае – Z-методы, а в качестве критерия выбора может применяться альтернатива [6]:

$$\begin{cases} K_{GZ} \in [K_{GZ}^{mn}; K_{GZ}^1) \Rightarrow Z - \text{method}; \\ K_{GZ} \in [K_{GZ}^1; K_{GZ}^2) \Rightarrow GZ - \text{method}; \\ K_{GZ} \in [K_{GZ}^2; K_{GZ}^{mx}] \Rightarrow G - \text{method}, \end{cases} \quad (11)$$

где  $K_{GZ}^1$  и  $K_{GZ}^2$  – соответственно левое и правое критериальные значения коэффициента когерентности;  
 $K_{GZ}^{mn}$  и  $K_{GZ}^{mx}$  – соответственно минимальное и максимальное значения коэффициента когерентности.

Критериальные значения коэффициента когерентности в выражении (11) определяются на основе юстировки базы методов прогнозирования к базе данных по электропотреблению. Здесь находит применение эвристический вариант GZ-анализа, целью которого является определение так называемой GZ-матрицы, представляющей собой таблицу методов прогнозирования, определенных как наиболее эффективных применительно к объектам на соответствующих временных интервалах (рис. 10) [6].

**КОДЫ ЛУЧШИХ МЕТОДОВ ПРОГНОЗИРОВАНИЯ**

Год	Объект																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
$GZ^T =$	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

Рис. 10. Фрагмент транспонированной GZ-матрицы (пример)

Строки GZ-матрицы – объекты, столбцы – последние пять временных интервалов предыстории функционирования. Элемент матрицы – код метода, который дает для объекта наименьшую относительную ошибку. Кодировка методов прогнозирования в GZ-матрице следующая (рис. 1): 1 – G-метод на основе APCC; 2 – G-метод на основе МДВР; 3 – G-метод на основе SSA; 4 – Z-метод БПТ; 5 – Z-метод СПТ; 6 – Z-метод ДКЗ.

Следует отметить, что GZ-матрица получается в результате последовательной многократной реализации различных методов прогнозирования

применительно к базе данных по электропотреблению. При этом глубина матрицы данных меняется прибавлением вектора данных по одному временному интервалу (начиная от минус пятого), и применительно к каждому состоянию матрицы данных выполняются процедуры эвристического варианта GZ-анализа (см. комментарии к рис. 2 и 3) [6].

Параллельно с GZ-матрицей формируется матрица коэффициентов когерентности объектов на последних пяти временных интервалах, имеющая ту же размерность (матрица G, рис. 11). При этом расчет коэффициентов когерентности каждый раз осуществляется на основе статистической обработки выборок значений электропотребления из матрицы данных соответствующей глубины [6]. В итоге каждому объекту на каждом временном интервале можно соотнести два параметра: код наиболее эффективно метода прогнозирования и коэффициент когерентности.

**КОЭФФИЦИЕНТЫ КОГЕРЕНТНОСТИ ОБЪЕКТОВ**

Год		1	2	3	4	5	6	7	8	9	10	11	Объект
$G^T =$	1	0.746	0.727	0.262	1.301	1.691	0.596	1.056	0.429	1	0.588	0.844	
	2	0.714	0.883	0.266	1.199	1.343	0.492	1.298	0.625	1	0.869	0.554	
	3	1.002	0.85	0.58	0.773	1.533	0.039	1.492	0.728	1	0.949	0.581	
	4	0.725	0.672	0.405	0.53	1.483	0.036	1.951	1.222	1	0.909	0.981	
	5	0.75	0.593	0.875	0.536	0.828	0.024	1.914	1.411	1	0.871	0.874	

Рис. 11. Фрагмент транспонированной матрицы коэффициентов когерентности (пример)

Статистическое сравнение GZ-матрицы и матрицы коэффициентов когерентности G позволяет получить параметры для выражения (11). При этом граничные значения коэффициента  $K_{GZ}^{mn}$  и  $K_{GZ}^{mx}$  получаются путем извлечения минимального и максимального элементов из матрицы G:

$$\begin{cases} K_{GZ}^{mn} = \min(G); \\ K_{GZ}^{mx} = \max(G). \end{cases} \quad (12)$$

Значения левого и правого критерияльных значений коэффициента когерентности  $K_{GZ}^1$  и  $K_{GZ}^2$  получаются на основе предварительного анализа чувствительности, в ходе которого методом случайного поиска осуществляется варьирование с целью определения таких значений, при которых частота встречаемости в GZ-матрице кодов G- и Z-методов различа-

лась бы в наибольшей степени. Частота встречаемости кодов определяется путем статистической обработки предварительно векторизованной GZ-матрицы. Поиск осуществляется в разделенной пополам области определения, в качестве которой выступает векторизованная матрица коэффициентов когерентности G, путем численного решения задачи (рис. 12) [6]:

$$\left\{ \begin{array}{l} \frac{|C_G - C_Z|}{C_\Sigma} \xrightarrow{K_{GZ}^1} \max, \{K_{GZ} \in G \mid K_{GZ} = [K_{GZ}^{mn}; (K_{GZ}^{mx} - K_{GZ}^{mn})/2]\}; \\ \frac{|C_G - C_Z|}{C_\Sigma} \xrightarrow{K_{GZ}^2} \max, \{K_{GZ} \in G \mid K_{GZ} = [(K_{GZ}^{mx} - K_{GZ}^{mn})/2; K_{GZ}^{mx}]\}, \end{array} \right. \quad (13)$$

где  $C_G$  – количество элементов GZ-матрицы, имеющих коды, соответствующие G-методам (1 – 3);  
 $C_Z$  – количество элементов GZ-матрицы, имеющих коды, соответствующие Z-методам (4 – 6);  
 $C_\Sigma$  – общее количество элементов GZ-матрицы.

Следует подчеркнуть, что целевая функция (13) формально прописана неоднозначно, т.к. в ее левой части предусмотрены операции над GZ-матрицей, а в правой записаны условия, касающиеся матрицы коэффициентов когерентности G. Однако данная неоднозначность снимается однозначным соответствием между элементами двух матриц.

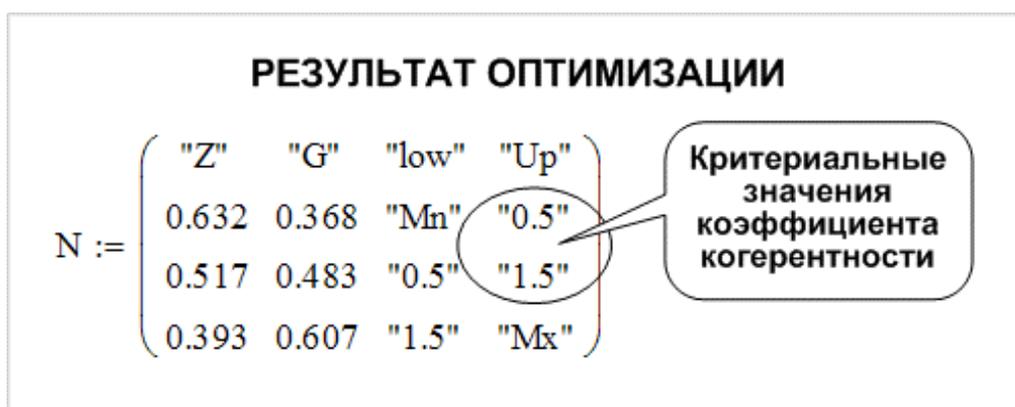


Рис. 12. Результат решения оптимизационной задачи (пример)

Оптимизационная задача (13) неизбежно порождает область неопределенности, в которой G- и Z-методы прогнозирования, как наиболее эффективные, встречаются примерно с одинаковой частотой (см. выражение (11) и рис. 12). В области неопределенности предлагается использовать комбинацию применяемых методов прогнозирования электропотребления.

Прогноз в заданный момент времени вычисляется по отдельности каждым методом с использованием информации, доступной на интервале предыстории. Окончательное прогнозное значение электропотребления получается как билинейная комбинация всех применяемых методов [9]:

$$\left\{ \begin{array}{l} W_{t+1} = \sum_{i=1}^{KF} \left( \varphi_{it+1} \sum_{j=1}^{KM} (\psi_{ijt+1} W_{ijt+1}) \right); \\ \sum_{i=1}^{KF} \varphi_{it+1} = 1; \sum_{j=1}^{KM} \psi_{ijt+1} = 1, \end{array} \right. \quad (14)$$

- где  $W_{t+1}$  – прогнозное значение электропотребления на момент времени (t+1) (см. рис. 3);  
 $KF$  – количество учитываемых факторов;  
 $\varphi_{it+1}$  – весовой коэффициент i-го фактора;  
 $KM$  – количество применяемых методов прогнозирования;  
 $\psi_{ijt+1}$  – весовой коэффициент i-го метода прогнозирования;  
 $W_{ijt+1}$  – прогнозное значение электропотребления, полученное j-м методом с учетом i-го фактора.

Весовые коэффициенты факторов, воздействующих на процесс электропотребления, определяются на основе фактор-анализа [8], а весовые коэффициенты каждого из методов прогнозирования вычисляются как частота встречаемости данного метода прогнозирования как лучшего в области неопределенности GZ-матрицы:

$$\left\{ \begin{array}{l} \psi = \frac{C_{BN}}{C_{\Sigma N}}; \sum_{GZ} C_{BN} = C_{\Sigma N}; \\ K_{GZ} \in G \mid K_{GZ} = [K_{GZ}^1; K_{GZ}^2]; \\ GZ_{i,j} \xleftrightarrow{i,j} G_{i,j}, \end{array} \right. \quad (15)$$

- где  $C_{BN}$  – количество элементов области неопределенности GZ-матрицы, где метод фиксируется как лучший;  
 $C_{\Sigma N}$  – общее количество элементов области неопределенности;  
 $GZ$  – GZ-матрица лучших методов прогнозирования;  
 $G$  – G-матрица коэффициентов когерентности.

К системе (15) необходимо дать следующие важные пояснения. Во-первых, для выполнения нормировочного условия (второе уравнение) должны быть рассмотрены все элементы области неопределенности GZ-матрицы. Во-вторых, границы самой области неопределенности фиксируются по соответствующим критериальным значениям коэффициента когерентности G-матрицы (третье уравнение системы). В-третьих, однозначное соответствие между GZ-матрицей и матрицей коэффициентов когерентности устанавливается по индексам строк и столбцов (четвертое выражение).

Таким образом, на этапе статистического анализа и построения эмпирической модели процесса электропотребления осуществляется глубокая обработка данных по электропотреблению объектов, которая включает интервальное оценивание, прогнозирование и нормирование. Прогнозирование электропотребления объектами осуществляется с использованием гауссовых и ципфовых (G- и Z-) методов. Более тонкий анализ рангового параметрического распределения, осуществляемый с помощью процедур GZ-анализа, позволяет существенно повысить эффективность прогнозирования. Как показано в работах [4-7], погрешность прогнозирования электропотребления с применением процедур критериального GZ-анализа для отдельных объектов может составить 4 – 10 %. При этом погрешность прогноза для техноценоза в целом, как правило, не превышает 1,5 – 2 %.

## Литература

1. Кудрин Б.И. Введение в технетику. – Томск: ТГУ, 1993. – 552 с.
2. Гнатюк В.И., Лагуткин О.Е. Ранговый анализ техноценозов. – Калининград: БНЦ РАЕН – КВИ ФПС РФ, 2000. – 86 с.
3. Гнатюк В.И., Северин А.Е. Ранговый анализ и энергосбережение. – Калининград: КВИ ФПС РФ, 2003. – 120 с.
4. Гнатюк В.И. Интернетсайт «Техника, техносфера, энергосбережение». – М., 2000. – <http://www.gnatukvi.ru>.
5. Гнатюк В.И. Закон оптимального построения техноценозов. – Выпуск 29. Ценологические исследования. – М.: Изд-во ТГУ – Центр системных исследований, 2005. – 384 с.
6. Гнатюк В.И. Закон оптимального построения техноценозов. – Компьютерная версия, перераб. и доп. – М.: Изд-во ТГУ – Центр системных исследований, 2005 – 2012. – <http://gnatukvi.ru/ind.html>.
7. Гнатюк В.И. и др. // – М.: Журнал «Электрика», 2003, №№ 2 – 6; 2004, № 7; 2005, № 2; 2006, №№ 1, 7, 12; 2007, №№ 2, 3, 7, 8, 11, 12.
8. Данилов Д.Л., Жиглявский А.А. Главные компоненты временных рядов: Метод «Гусеница». – СПб.: СПбГУ, 1997. – 308 с.
9. Воронцов К.В., Егорова Е.В. Динамически адаптируемые композиции алгоритмов прогнозирования // – М.: Искусственный Интеллект, 2006. – С. 277 – 280.