US 20030028350A1

(19) United States

(12) Patent Application Publication

(10) Pub. No.: US 2003/0028350 A1

(43) Pub. Date: Feb. 6, 2003

Seem

(54) METHOD OF INTELLIGENT DATA ANALYSIS TO DETECT ABNORMAL USE OF UTILITIES IN BUILDINGS

(76) Inventor: **John E. Seem**, Glendale, WI (US)

Correspondence Address:
Joseph N. Ziebert
Foley & Lardner
777 East Wisconsin Ave.
Milwaukee, WI 53202 (US)

(21) Appl. No.: **09/910,371**

(22) Filed: **Jul. 20, 2001**

Related U.S. Application Data

(63) Continuation-in-part of application No. 09/886,920, filed on Jun. 21, 2001, now abandoned.

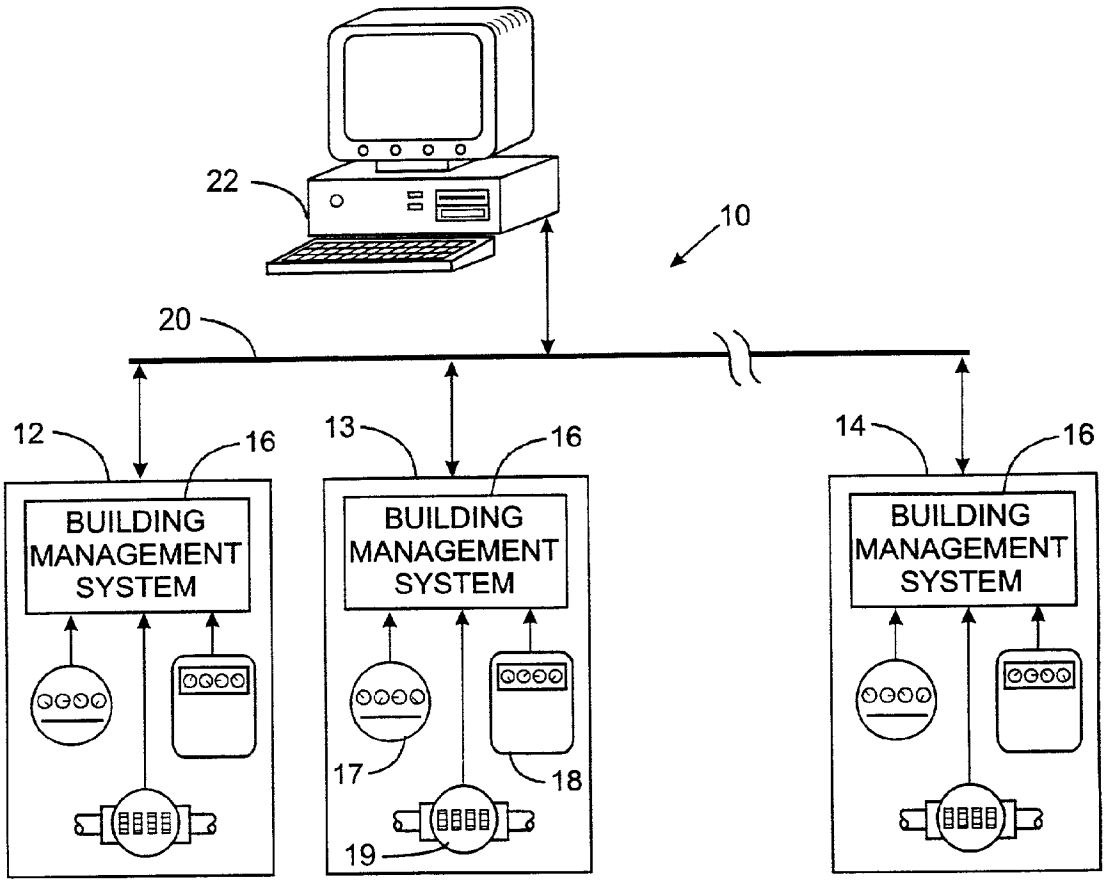
Publication Classification

(51) **Int. Cl.**[7] **G06F 15/00**; G06F 17/18; G06F 101/14

(52) **U.S. Cl.** **702/179**

(57) **ABSTRACT**

Outlier identification is employed to detect abnormally high or low energy use in a building. The utility use is measured periodically throughout each day and the measurements are grouped according to days that have similar average utility consumption levels. The data in each group is statistically analyzed using the Generalized Extreme Studentized Deviate (GESD) method. That method identifies outliers which are data samples that vary significantly from the majority of the data. The degree to which each outlier deviates from the remainder of the data indicates the severity of the abnormal utility consumption denoted by that outlier. The resultant outlier information is readily discernable by the building operators in accessing whether the cause of a particular occurrence of abnormal utility usage requires further investigation.



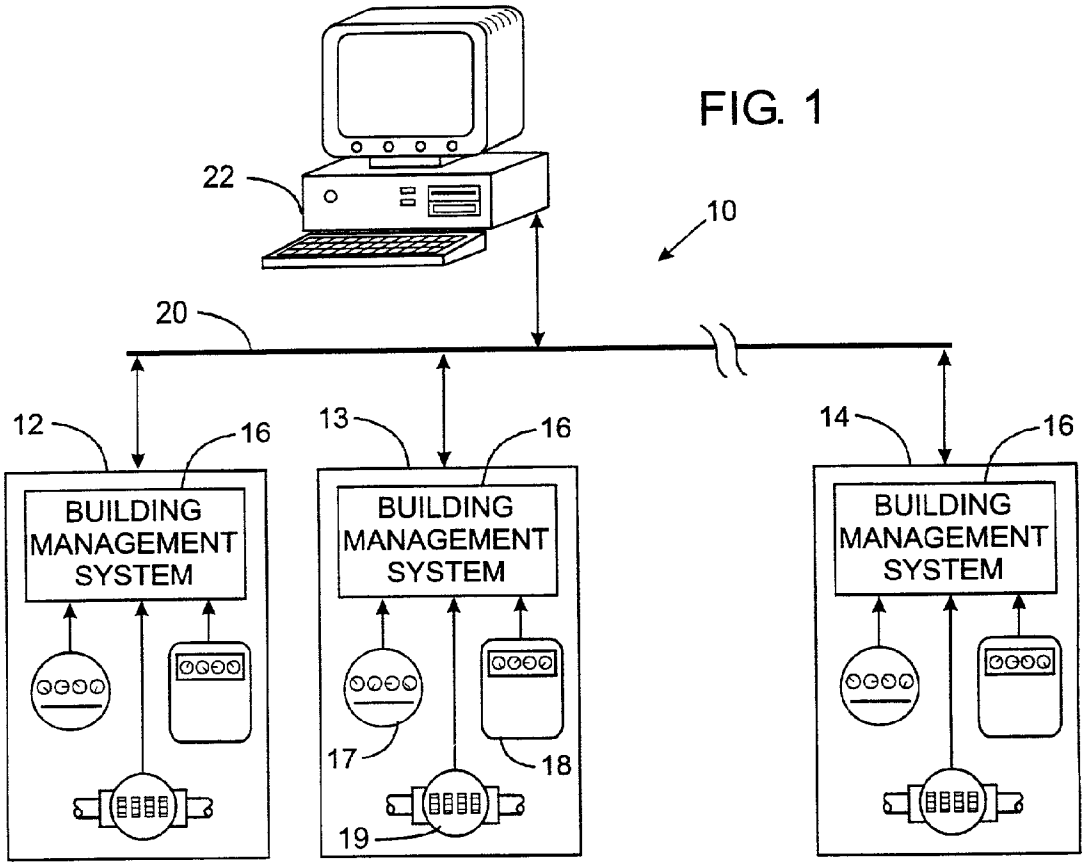

22
10
20
12
16
13
16
14
16
BUILDING MANAGEMENT SYSTEM
BUILDING MANAGEMENT SYSTEM
BUILDING MANAGEMENT SYSTEM
17
18
19

FIG. 1

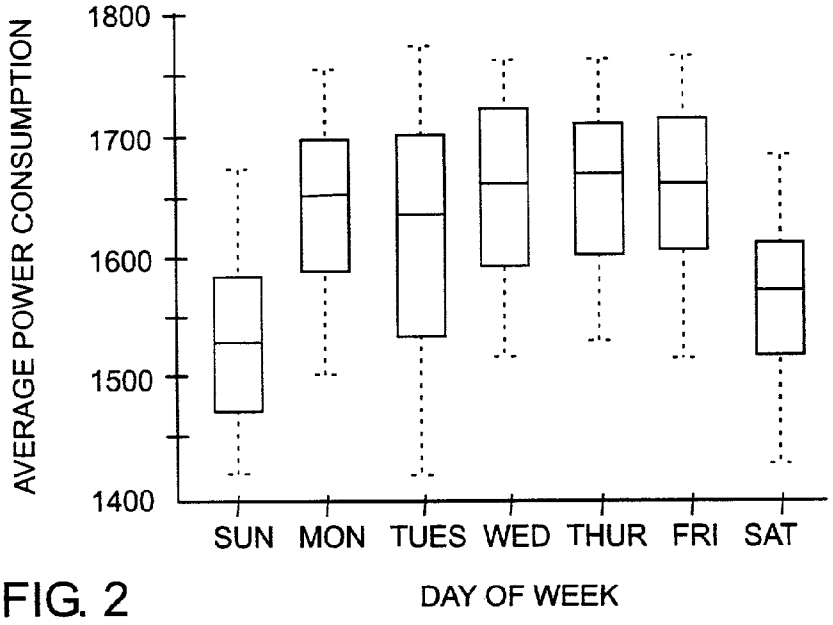

AVERAGE POWER CONSUMPTION
1800
1700
1600
1500
1400
SUN
MON
TUES
WED
THUR
FRI
SAT
DAY OF WEEK

FIG. 2

FIG. 3

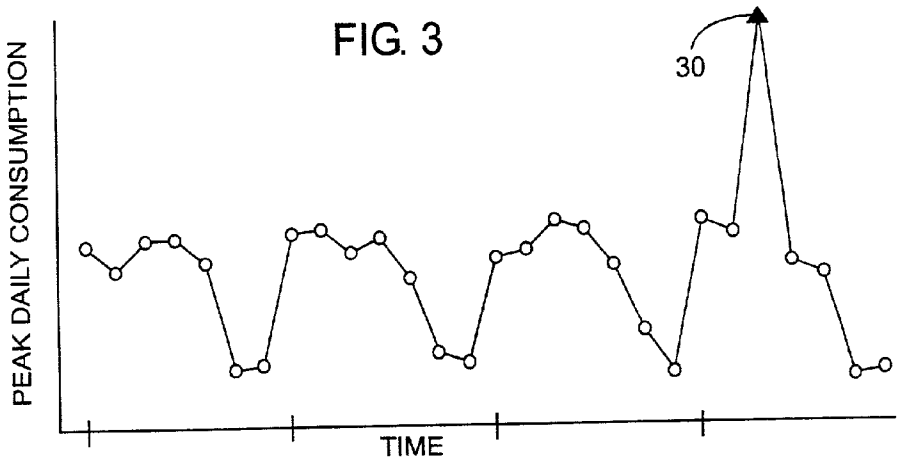
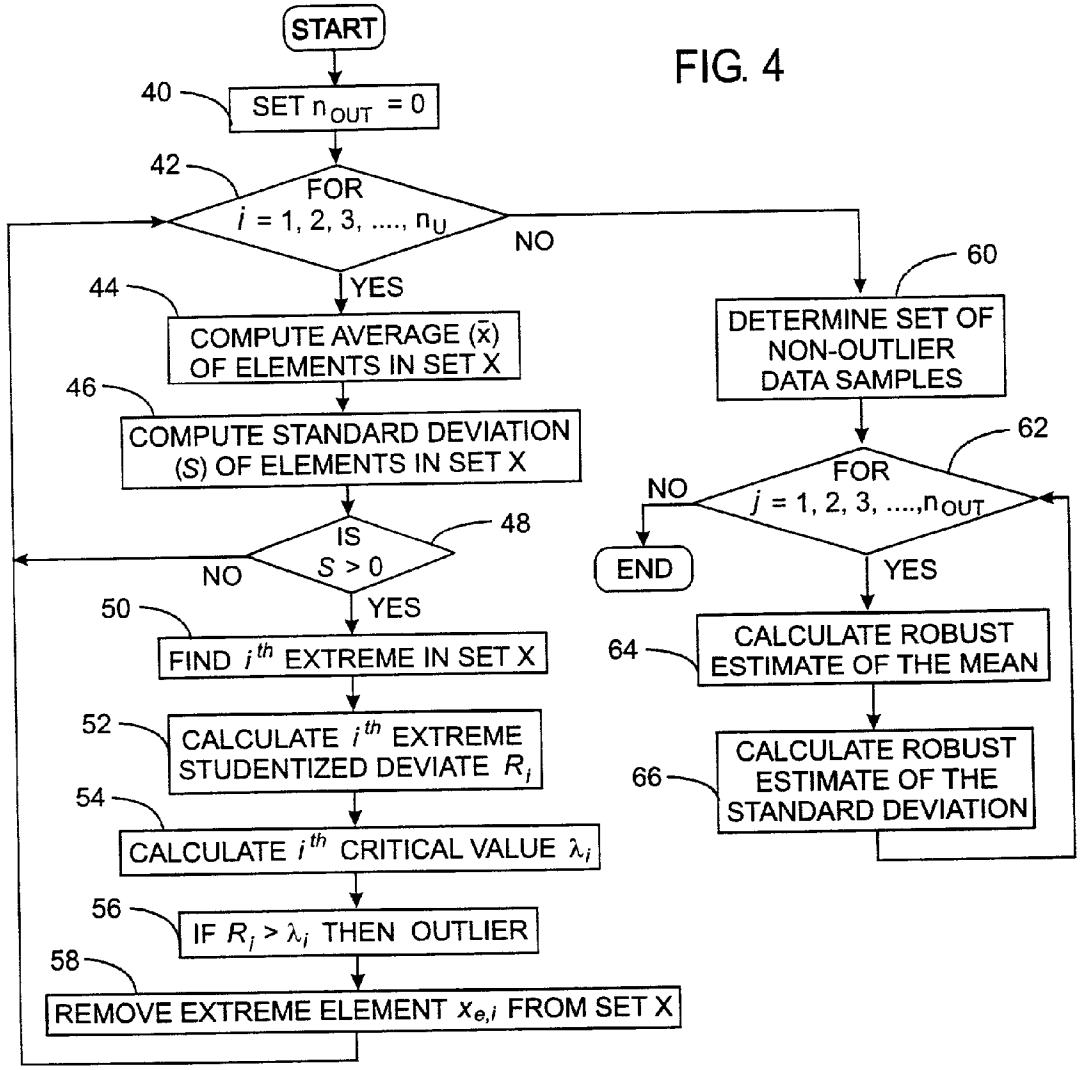

PEAK DAILY CONSUMPTION
TIME
30

FIG. 4


START
40
SET n_{OUT} = 0
42
FOR i = 1, 2, 3, ..., n_U
NO
YES
44
COMPUTE AVERAGE ($\bar{x}$) OF ELEMENTS IN SET X
46
COMPUTE STANDARD DEVIATION (S) OF ELEMENTS IN SET X
48
IS $S > 0$
NO
YES
50
FIND i^{th} EXTREME IN SET X
52
CALCULATE i^{th} EXTREME STUDENTIZED DEVIATE R_i
54
CALCULATE i^{th} CRITICAL VALUE λ_i
56
IF $R_i > \lambda_i$ THEN OUTLIER
58
REMOVE EXTREME ELEMENT $x_{e,i}$ FROM SET X
60
DETERMINE SET OF NON-OUTLIER DATA SAMPLES
62
FOR j = 1, 2, 3, ...,n_{OUT}
NO
END
YES
64
CALCULATE ROBUST ESTIMATE OF THE MEAN
66
CALCULATE ROBUST ESTIMATE OF THE STANDARD DEVIATION

METHOD OF INTELLIGENT DATA ANALYSIS TO DETECT ABNORMAL USE OF UTILITIES IN BUILDINGS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This is a continuation in part of U.S. patent application Ser. No. 09/886,920 filed on Jun. 21, 2001.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] Not Applicable

BACKGROUND OF THE INVENTION

[0003] 1. Field of the Invention

[0004] The present invention relates managing consumption of utilities, such as electricity, natural gas and water; and more particularly to detecting the occurrence of abnormal usage.

[0005] 2. Description of the Related Art

[0006] Large buildings often incorporate computerized control systems which manage the operation of different subsystems, such that for heating, ventilation and air conditioning. In addition to ensuring that the subsystem performs as desired, the control system operates the associated equipment in as efficiently as possible.

[0007] A large entity may have numerous buildings under common management, such as on a university campus or a chain of store located in different cities. To accomplish this, the controllers in each building gather data regarding performance of the building subsystems which data can be analyzed at the central monitoring location.

[0008] With the cost of energy increasing, building owners are looking for ways to conserve utility consumption. In addition, the cost of electricity for large consumers may be based on the peak use during a billing period. Thus high consumption of electricity during a single day can affect the rate at which the service is billed during an entire month. In addition, certain preferential rate plans require a customer to reduce consumption upon the request of the utility company, such as on days of large service demand throughout the entire utility distribution system. Failure to comply with the request usually results in stiff monetary penalties which raises the energy cost significantly above that for an unrestricted rate plan. Therefore, a consumer has to analyze the energy usage in order to determine the best rate plan and implement processes to ensure that operation of the facility does not inappropriately cause an increase in utility costs.

[0009] In addition, abnormal energy or other utility consumption may indicate malfunctioning equipment or other problems in the building. Therefore, monitoring utility usage and detecting abnormal consumption levels can indicate when maintenance or replacement of the machinery is required.

[0010] As a consequence, sensors are being incorporated into building management systems to measure utility usage for the entire building, as well as specific subsystems such as heating air conditioning and ventilation equipment. These management systems collect and store massive quantities of utility use data which is overwhelming to the facility operator when attempting to analyze that data in an effort to detect anomalies.

[0011] Alarm and warning systems and data visualization programs often are provided to assist in deriving meaning information from the gathered data. However, human operators must select the thresholds for alarms and warnings, which is a daunting task. If the thresholds are too tight, then numerous of false alarms are issued, and if the thresholds are too loose, equipment or system failures can go undetected. The data visualization programs can help building operators detect and diagnose problems, but a large amount time can be spent detecting problems. Also, the expertise of building operators varies greatly. New or inexperienced operators may have difficulty detecting faults and the performance of an operator may vary with the time of day or day of the week.

[0012] Therefore there is a need for robust data analysis methods to automatically determine if the current energy use is significantly different than previous energy patterns and if so, alert the building operator or mechanics to investigate and correct the problem.

SUMMARY OF THE INVENTION

[0013] Abnormal utility usage by a building or a particular apparatus in the building can be determined by repeatedly measuring the level of use of the utility thereby producing a plurality of utility measurements. A Generalized Extreme Studentized Deviate (GESD) statistical procedure is applied to the plurality of utility measurements to identify any measurement outliers. The measurement outliers denote times when unusual utility consumption occurred, thereby indicating times during which operation of the building or the particular apparatus should be investigated.

[0014] In the preferred embodiment, a severity of abnormal utility usage can be established by determining a degree to which the associated outlier deviates from the norm. This can be accomplished by calculating robust estimates of the mean ($\bar{x}_{robust}$) and the standard deviation (s_{robust}) of each outlier

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] **FIG. 1** is a block diagram of a distributed facility management system which incorporates the present invention;

[0016] **FIG. 2** is a box plot of average electrical power consumption for a building;

[0017] **FIG. 3** is a graph depicting the energy consumption for a building; and

[0018] **FIG. 4** is a flowchart of the algorithm that analyzes the energy consumption data for the building.

DETAILED DESCRIPTION OF THE INVENTION

[0019] With reference to **FIG. 1, a** distributed facility management system **10** supervises the operation of systems in a plurality of buildings **12**, **13** and **14**. Each building contains its own building management system **16** which is a computer that governs the operation of various subsystems within the building. Each building management system **16**

PETITIONER'S EXHIBIT 1007
US PATENT NO. 8,909,482

also is connected to numerous sensors throughout the building that monitor consumption of different utility services at various points of interest. For example, the building management system **16** in building **13** is connected to a main electric meter **17**, the central gas meter **18** and the main water meter **19**. In addition, individual meters for electricity, gas, water and other utilities can be attached at the supply connection to specific pieces of equipment to measure their consumption. For example, water drawn into a cooling tower of an air conditioning system may be monitored, as well as the electric consumption of the pumps for that unit.

[**0020**] Periodically the building management system **16** gathers data from the sensors and stores that information in a database contained within the memory of the computer for the building management system. The frequency at which the data is gathered is determined by the operator of the building based on the type of the data and the associated building function. The utility consumption for functions with relatively steady state operation can be sampled less frequently, as compared to equipment large variations in utility consumption.

[**0021**] The gathered data can analyzed either locally by the building management system **16** or forwarded via a communication link **20** for analysis by a centralized computer **22**. For example, the communication link **20** can be a wide area computer network extending among buildings in an office park or a university campus, or the communication link may comprise telephone lines extending between individual stores and the principal office of a large retailer.

[**0022**] The present invention relates to a process by which the data acquired from a given building is analyzed to determine abnormal usages of a particular utility service. This is accomplished by reviewing the data for a given utility service to detect outliers, data samples that vary significantly from the majority of the data. The data related to that service is separated from all the data gathered by the associated building management system. That relevant data then is categorized based on the time periods during which the data was gathered. Utility consumption can vary widely from one day of the week to another. For example, a typical office building has relatively high utility consumption Monday through Friday when most workers are present, and significantly lower consumption on weekends. In contrast, a manufacturing facility that operates seven days a week may have similar utility consumption every day. However, different manufacturing operations may be scheduled on different days of the week, thereby varying the level of utility consumption on a daily basis.

[**0023**] Therefore, prior to implementing the outlier analysis, the building operator defines one or more groups of days having similar utility consumption. That grouping can be based on a knowledge of the building use, or from data regarding daily average or peak utility consumption. For example, **FIG. 2** is a box plot of the average daily electrical power consumption for an exemplary building. A similar box plot can be generated for the peak electrical power consumption. It is apparent from an examination of this graph that consumption during weekdays (Monday through Friday) is similar, i.e. the normal consumption of electricity falls within one range of levels (A), and weekend periods (Saturday and Sunday) also have similar consumption levels that fall within a second range (B). Therefore, separate utility consumption analyses would be performed on data from two groups of days, weekdays and weekends. However, different day groups would apply to a manufacturing plant in which high utility consuming equipment is run only on Tuesdays, Thursdays and Saturdays. In this latter example, Tuesdays, Thursdays and Saturdays would be placed into one analysis group with the remaining days of the week into a second group.

[**0024**] **FIG. 3** depicts the peak daily consumption for this building over a period of four weeks. The weekday peaks are significantly greater than the peak consumption on the weekends. Point **30** represents a day when peak consumption of electricity was abnormally high. This may have been caused by a large piece of equipment turning on unexpectedly, for example an additional chiller of an air conditioning system activating on a single very hot day. The data value for this abnormally high level is referred to as an "outlier" and building operators are interested in finding such outliers and learning their cause. Outliers often result from equipment of system control malfunctions which require correction.

[**0025**] The daily usage pattern for each type of utility service can be different. For example, the electricity use in a manufacturing facility may be relatively uniform every day of the week, but a special gas furnace is operated only on certain days of the week. The grouping of days for analyzing electricity use in this facility will be different than the day groups for gas consumption. As a consequence, each utility being monitored is configured and analyzed independently.

[**0026**] Focusing on one type of utility service, such as electricity use for the entire building, acquisition of periodic electric power measurements from the main electric meter **17** produces a set X of n data samples where $X \in \{x_1, x_2, x_3, \ldots, x_n\}$. The analysis will find the elements in set X that are outliers, i.e., statistically significantly different than most of the data samples. This determination uses a form of the Generalized Extreme Studentized Deviate (GESD) statistical procedure described by B. Rosner, in "Percentage Points for a Generalized ESD Many-Outlier Procedure"*Technometrics,* Vol. 25, No. 2, pp. 165-172, May 1983.

[**0027**] Prior to the analysis the user needs to specify the probability α of incorrectly declaring one or more outliers when no outliers exist and an upper bound (n_u) on the number of potential outliers. The probability α defines the sensitivity of the process and is redefined periodically based on the number of false warnings that are produced by the system finding outliers. In other words the probability is adjusted so that the number of outliers found results in an acceptable level of warnings of abnormal utility consumption within the given reporting period, recognizing that false warnings can not be eliminated entirely and still have an effective evaluation technique. The upper bound (n_u) specifies a maximum number of data samples in set X that can be considered to be outliers. This number must be less that fifty percent of the total number of data samples, since by definition a majority the data samples can not be outliers, i.e., $n_u \leqq 0.5(n-1)$. For example, a upper bound (n_u) of thirty percent can be employed for electricity consumption analysis.

[**0028**] The data analysis commences at step **40** by setting the initial value n_{out} for number of outliers to zero. Then at step **42** a FOR loop is defined in which the program

execution loops through steps **44-58** processing each data sample specified by the upper bound n_u, i.e. samples x_i, where i=1, 2, 3, . . . , n_u. The arithmetic mean ($\bar{x}$) of all the elements in set X is calculated at the first step **44** of this loop. Then at step **46**, the standard deviation (s) of the elements in set X is calculated.

[**0029**] If the standard deviation is not greater than zero (s>0), i.e. the samples of utility usage are substantially the same as may occur in rare cases, then the pass through the loop terminates at step **48** by returning to step **42**. Otherwise the execution of the algorithm advances to step **50** at which the ith extreme member in set X is located. That extreme element $x_{e,1}$ is the element in set X that is farthest from the mean $\bar{x}$. Using that extreme element $x_{e,i}$ the computer **22** calculates the ith extreme studentized deviate R_i at step **52** according to the expression:

$$R_i = \frac{|x_{e,i} - \bar{x}|}{s} \tag{1}$$

[**0030**] The ith 100α percent critical value λ_i then is calculated at step **54** using the equation:

$$\lambda_i = \frac{(n-i)t_{n-i-1,p}}{\sqrt{(n-i+1)(n-i-1+t^2_{n-i-1,p})}} \tag{2}$$

[**0031**] where $t_{n-i-1,p}$ is the student's t-distribution with (n−i−1) degrees of freedom, and a percentile p is determined from:

$$p = 1 - \left(\frac{\alpha}{2(n-i+1)}\right) \tag{3}$$

[**0032**] Abramowitz and Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables,* Dover Publications, Inc., New York, 1970, provides an process for determining the student's t-distribution $t_{v,p}$ for the pth percentile of a t-distribution with v degrees of freedom. This determination begins by estimating the standardized normal deviate f at the pth percentile, according to:

$$f = \sqrt{\ln\left(\frac{1}{(1-p^2)}\right)} \tag{4}$$

$$z_p \cong f - \left(\frac{2.515517 + 0.802853f + 0.010328f^2}{1 + 1.432788f + 0.189269f^2 + 0.001308f^3}\right) \tag{5}$$

[**0033**] The student's t-distribution $t_{v,p}$ is estimated from z_p and the degrees of freedom v using the following expressions:

$$g_1 = \frac{1}{4}(z_p^3 + z_p) \tag{6}$$

$$g_2 = \frac{1}{96}(5z_p^5 + 16z_p^3 + 3z_p) \tag{7}$$

-continued

$$g_3 = \frac{1}{384}(3z_p^7 + 19z_p^5 + 17z_p^3 - 15z_p) \tag{8}$$

$$g_4 = \frac{1}{92160}(79z_p^9 + 776z_p^7 + 1482z_p^5 - 1920z_p^3 - 945z_p) \tag{9}$$

$$t_{v,p} \cong z_p + \frac{g_1}{v} + \frac{g_2}{v^2} + \frac{g_3}{v^3} + \frac{g_4}{v^4} \tag{10}$$

[**0034**] Upon solving equations (1) and (2), if at step **56** the ith extreme studentized deviate R_i is greater than the ith 100α percent critical value λ_i ($R_i > \lambda_i$), then the ith extreme data sample $x_{e,i}$ is an outlier and the number of outliers equals i.

[**0035**] At step **58**, the extreme element $x_{e,i}$ is removed from set X and the number of elements in that set now equals n−i. The algorithm then returns to step **42** to repeat the process and hunt for another outlier. Eventually the set of data samples becomes reduced to the upper bound (n_u) at which point the FOR loop terminates by branching to step **60**. At that point, the outliers have been identified with a set of outliers given by $X_{out} \in \{x_{e,1}, x_{e,2}, \ldots, x_{e,n_{out}}\}$. If no outliers were found in set X, then X_{out} is an empty set.

[**0036**] After the outliers have been identified a robust estimate of the mean ($\bar{x}_{robust}$) and a standard deviation (s_{robust}) for the set of n data samples $X \in \{x_1, x_2, x_3, \ldots, x_n\}$ are calculated at steps **64** and **66**. In essence this determines how far the outliers deviate from the remainder of the data and thus represents the severity of the abnormal utility consumption denoted by each outlier. The process for making this determination commences with the set of outliers X_{out} and the set ($X_{non\text{-}out}$) of the data samples from set X that are not outliers. Specifically:

$$X_{non\text{-}out} = \{x | x \in X \text{ and } x \in X_{out}\} \tag{11}$$

[**0037**] The robust estimate of the mean ($\bar{x}_{robust}$) is the average value of the elements in set $X_{non\text{-}out}$ as given by:

$$\bar{x}_{robust} = \frac{\sum_{j=1}^{n-n_{out}} x_j}{n - n_{out}} \tag{12}$$

[**0038**] where $x_j \in X_{non\text{-}out}$.

[**0039**] The robust estimate of the standard deviation (s_{robust}) is the sample standard deviation of the elements in set $X_{non\text{-}out}$ as defined by the expression:

$$s_{robust} = \sqrt{\frac{\sum_{j=1}^{n-n_{out}} (x_j - \bar{x}_{robust})^2}{n - n_{out} - 1}} \tag{13}$$

[**0040**] The robust estimates of the mean ($\bar{x}_{robust}$) and the standard deviation (s_{robust}) quantify the severity of the abnormal utility usage represented by the corresponding

outlier. These values can be plotted to provide a graphical indication as to that severity by which the building operator is able to determine whether investigation of the cause is warranted.

[0041] For days with abnormal energy consumption, the robust estimates of the mean ($\bar{x}_{robust}$) and the standard deviation (s_{robust}) are used to determine how different the energy use is from the typical day. One measure is a robust estimate of the number of standard deviations from the average value:

$$z_j = \frac{x_{e,j} - \bar{x}_{robust}}{s_{robust}} \tag{14}$$

[0042] where $x_{e,j}$ is the energy consumption for the jth outlier, $\bar{x}_{robust}$ is a robust estimate of the average energy consumption for days of the same day type as outlier j, and s_{robust} is a robust estimate of the standard deviation of energy consumption for days of the same day type.

[0043] The operator can be presented with tables or graphs that show the outliers and the amount of variation for the outliers.

What is claimed is:

1. A method for determining abnormal consumption of a utility by a system:

repeatedly measuring the level of use of the utility thereby producing a plurality of utility measurements;

employing a statistical procedure to identify any outliers in the plurality of utility measurements; and

evaluating performance of the system in response to any outliers identified.

2. The method as recited in claim 1 wherein evaluating performance of the system comprises determining a severity of abnormal utility usage represented by an outlier.

3. The method as recited in claim 2 wherein determining a severity of abnormal utility usage comprises calculating how many standard deviations a given outlier is from the average value for utility usage by using the expression:

$$z_j = \frac{x_{e,j} - \bar{x}_{robust}}{s_{robust}}$$

where $x_{e,j}$ is the energy consumption for the jth outlier, $\bar{x}_{robust}$ is a robust estimate of the average energy consumption for days of the same day type as outlier j, and s_{robust} is a robust estimate of the standard deviation of energy consumption for days of the same day type.

4. The method as recited in claim 2 wherein determining a severity of abnormal utility usage comprises calculating a standard deviation of each outlier.

5. The method as recited in claim 2 wherein determining a severity of abnormal utility usage comprises determining an amount that each outlier deviates from a mean of the plurality of utility measurements which are not identified as outliers.

6. The method as recited in claim 1 further comprising separating the plurality of utility measurements into groups wherein each group contains utility measurements acquired during days that under normal conditions have similar utility consumption levels; and

wherein the statistical procedure is applied separately to each group.

7. The method as recited in claim 1 further comprising separating the plurality of utility measurements into groups wherein each group contains utility measurements acquired during predefined time periods that under normal conditions have similar utility consumption levels; and

wherein the statistical procedure is applied separately to each group.

8. The method as recited in claim 1 wherein the outliers are identified using a Generalized Extreme Studentized Deviate (GESD) statistical procedure.

9 The method as recited in claim 8 wherein employing a statistical procedure involve utilizing a Generalized Extreme Studentized Deviate (GESD) statistical procedure comprising:

(a) calculating an arithmetic mean ($\bar{x}$) of the plurality of utility measurements;

(b) finding an extreme utility measurement $x_{e,i}$ which is the utility measurement that has a value which is farther numerically from the arithmetic mean ($\bar{x}$) than the other ones of the plurality of utility measurements;

(c) using the extreme utility measurement $x_{e,i}$ to calculate an extreme studentized deviate R_i;

(d) calculating a 100α percent critical value λ_i for the extreme utility measurement $x_{e,i}$;

(e) declaring the extreme utility measurement $x_{e,i}$ to be an outlier when the extreme studentized deviate R_i is greater than the 100α percent critical value λ_i.

10. The method as recited in claim 9 further comprising:

removing the extreme utility measurement $x_{e,i}$ from the plurality of utility measurements to form a new plurality of utility measurements; and

repeating steps (a) through (e) for the new plurality of utility measurements.

11. The method as recited in claim 9 wherein the extreme studentized deviate R_i is calculated according to the expression:

$$R_i = \frac{|x_{e,i} - \bar{x}|}{s}$$

where s is a standard deviation of the plurality of utility measurements.

12. The method as recited in claim 9 wherein the 100α percent critical value λ_i then is calculated using the equation:

$$\lambda_i = \frac{(n-i)t_{n-i-1,p}}{\sqrt{(n-i+1)(n-i-1+t^2_{n-i-1,p})}}$$

where n is the number of utility measurements, i is a number identifying a particular outlier being evaluated, $t_{n-i-1,p}$ is a student's t-distribution with (n−i−1) degrees of freedom, and

p is a value based on the user defined probability α of incorrectly declaring one or more outliers when no outliers exist.

13. The method as recited in claim 12 wherein percentile p is determined from:

$$p = 1 - \left(\frac{\alpha}{2(n-i+1)}\right).$$

14. A method for determining abnormal consumption of a utility by a system:

(a) repeatedly measuring the level of use of the utility, thereby producing a plurality of utility measurements;

(b) forming a group of those of the plurality of utility measurements taken during predefined periods of time;

(c) calculating an arithmetic mean ($\bar{x}$) of the group;

(d) finding an extreme utility measurement $x_{e,i}$ which is the utility measurement having a value that is farthest numerically from the arithmetic mean ($\bar{x}$);

(e) using the extreme utility measurement $x_{e,i}$ to calculate an extreme studentized deviate R_1;

(f) calculating a 100α percent critical value λ_i for the extreme utility measurement $x_{e,i}$;

(g) declaring the extreme utility measurement $x_{e,i}$ to be an outlier indicative of abnormal utility use when the extreme studentized deviate is greater than the 100α percent critical value;

(h) removing the extreme utility measurement $x_{e,i}$ from the group of utility measurements; and

(i) repeating steps (c) through (h) a defined number of times.

15. The method as recited in claim 14 wherein the extreme studentized deviate R_i is calculated according to the expression:

$$R_i = \frac{|x_{e,i} - \bar{x}|}{s}$$

where s is a standard deviation of the plurality of the group of utility measurements.

16. The method as recited in claim 14 wherein the 100α percent critical value λ_i is calculated using the equation:

$$\lambda_i = \frac{(n-i)t_{n-i-1,p}}{\sqrt{(n-i+1)(n-i-1+t^2_{n-i-1,p})}}$$

where n is the number of utility measurements, i is a number identifying a particular outlier being evaluated, $t_{n-i-1,p}$ is the student's t-distribution with (n−i−1) degrees of freedom, and p is a value based on the user defined probability α of incorrectly declaring one or more outliers when no outliers exist.

17. The method as recited in claim 16 wherein percentile p is determined from:

$$p = 1 - \frac{\alpha}{2(n-i+1)}.$$

18. The method as recited in claim 14 further comprising defining periods of time during a plurality of days in which under normal conditions similar utility consumption levels occur during each one of those periods of time.

19. The method as recited in claim 14 further comprising performing maintenance on the system in response to examination of one or more of the outliers.

* * * * *