

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
3 May 2001 (03.05.2001)

PCT

(10) International Publication Number  
**WO 01/31514 A2**

(51) International Patent Classification<sup>7</sup>: **G06F 17/40**

(21) International Application Number: **PCT/US00/29864**

(22) International Filing Date: **30 October 2000 (30.10.2000)**

(25) Filing Language: **English**

(26) Publication Language: **English**

(30) Priority Data:  
60/162,048 28 October 1999 (28.10.1999) **US**

(71) Applicant: **GENERAL ELECTRIC COMPANY**  
[US/US]; Carl A. Rowold, Esquire, Bldg. 14-522, 2901  
East Lake Road, Erie, PA 16531 (US).

(72) Inventors: **CARDELLA, Aidan; - VARNER, Evelyn; -**

(74) Agent: **SANKS, Terry, M.**; Holland & Knight LLP, P.O.  
Box 1526, Orlando, FL 32802-1526 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

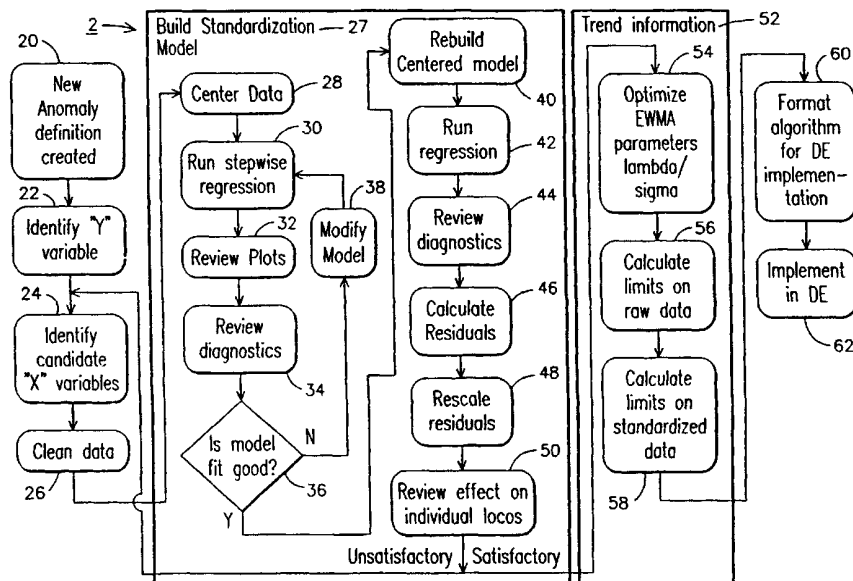
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **A PROCESS FOR THE MONITORING AND DIAGNOSTICS OF DATA FROM A REMOTE ASSET**



(57) Abstract: A method for processing monitored data from a remote asset (5) to optimize maintenance and operation schedules, the method comprising the steps of collecting data from the remote asset, building a data set based on the data collected (20, 22, 24, 26), applying statistical scripts to the data set to create a statistical model (28, 30, 32, 34, 36), comparing the statistical model to the data set, (40, 42, 44, 46), creating a standardization model from the compared statistical model and the data set (48, 50), applying a trending algorithm to the data (52), deriving statistical based control limits, (54, 56, 58) and applying the control limits to a new set of collected data (60, 62).

## **A PROCESS FOR THE MONITORING AND DIAGNOSTICS OF DATA FROM A REMOTE ASSET**

### **BACKGROUND OF THE INVENTION**

This invention relates to diagnostics and monitoring systems, and more particularly to a method and system for processing incoming monitored data from a remote asset, such as a locomotive, where the system automatically, or with limited user interface, evaluates and determines whether the monitored data is within  
5 predetermined operating thresholds.

A locomotive is a complex system comprised of many subsystems. Many of the subsystems need fluids, such as oil, coolant, fuel, and other lubricating fluids for operation. If any of these fluids become contaminated or the fluid levels drop below acceptable operational limits, any one of these fluids may result in engine failure  
10 during operation or may reduce the reliability of the engine not failing before its next scheduled maintenance. Having less than adequate fluid could result in components running hot or operating in a mode that is not considered optimum for the given component.

Systems exist which are capable of measuring the quality and level of an  
15 engine's various fluids. For example, it is believed that U.S. Patent No. 5,964,318 discloses a system for measuring the quality and level of lubricants in an engine, specifically in an lubricant reservoir, wherein new lubricant is injected as needed. The state of the lubricant is then communicated to a remote site through a data link.

However even if a system can detect that a lubricant level is low and replace  
20 the missing lubricant, it does not appear to determine how the less than optimum fluid level has effected the engine, nor has it determined whether the lubricant has been contaminated.

## SUMMARY OF THE INVENTION

Towards this end, there is a particular need to a method or process to automatically, or with limited user interface, detect anomalous conditions of a remote asset. This method or process can be achieved with a series of algorithms to build a data set based on historical data and then apply statistical scripts to the data set to perform several iterations of statistical analysis on the data collected from a remote asset which are eventually applied to current data collected. The process enables a more accurate and precise monitoring and diagnosis of a remote asset's anomalous condition via a reduction in the false alarm rate associated with the input thresholds or limits. This is accomplished by a reduction in the variability of the input signal via application of a standardization algorithm and optimum choice of trending parameters.

The method comprises collecting data from the remote asset, building a data set based on the data collected, applying statistical scripts to the data set to create a statistical model, comparing the statistical model to the data set, and creating a standardization model from the compared statistical model and the data set, applying a trending algorithm to the data, deriving statistical based control limits, and applying the control limits to a new set of data.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is an exemplary example of components used for a remote diagnostics and monitoring system of a remote asset.

Figure 2 illustrates the steps taken in processing incoming monitored parameter data to aid in monitoring and diagnosis of a remote asset.

## DETAILED DESCRIPTION OF THE INVENTION

Figure 1 is an exemplary example of components used for a remote diagnostics and monitoring system of a remote asset. The remote asset, or locomotive, 5 has an on-board monitor system 10 to monitor such items as fluid temperatures, fluid levels, component temperatures, and current levels and voltage outputs. The system can also monitor the location of the locomotive via a global positioning system 12. Once the monitored data is collected, it is sent, via either a

satellite link 14, or a direct line connection, not shown, to a monitoring and diagnostic service center 16 which includes a respective transceiver, not shown, at each location. In another embodiment, the monitoring and diagnostic service center 16 uses a processor 11 to process 2, as shown in Figure 2, the data. In one embodiment, the monitoring and diagnostic service center 16 has the ability to communicate either the collected data or processed results to a customer facility 17 as well as repair depots 18. Communication of the data or results is also possible between the customer facility 17 and repair depots 18. Communication with and between the customer facility 17 and the repair depots 18 can be by either a direct line connection or by a satellite link where each location has a respective transceiver.

Figure 2 illustrates the overall process 2 which comprises the steps, algorithms or sub-processes, taken in processing incoming monitored parameter data to aid in monitoring and diagnosis of a remote asset. In one embodiment, the steps of Figure 2 are implemented and executed by a central processing unit located at the monitoring and diagnostic service center 16, shown in Figure 1. The steps delineate a process 2 by which a series of algorithms, or sub-processes, 20, 22, 24, 26 are used to build a data set and then utilize an array of statistical scripts 28, 30, 32, 34, 36, 40, 42, 44, 46, 48, 50 to perform several iterations of statistical analysis on the incoming monitored parameter data 27. The preliminary output of the process is a statistical model which is applied to the monitored parameter data in order to eliminate the effects of extraneous variables and obtain a standardized signal.

Trending 52 the resulting standardized signal next occurs. In trending, control chart type limits, boundaries, are placed around the data to indicate whether the data are within predetermined control limits. The control chart type limits are derived from the data using a time series modeling optimization technique, such as an Exponentially Weighted Moving Average (EWMA) technique 54. In one embodiment, an Auto-Regressive Integrated Moving Average (ARIMA) technique is used to optimize a value used in calculating the EWMA chart. With the EWMA chart, limits on the data and subsequent standardized data are determined 56, 58. The resultant standardization model and data thresholds are then formatted recorded 60 for subsequent implementation 62 in a completely automated monitoring and diagnostic system where newly collected data is compared to the resultant standardization model

and data thresholds to determine anomalous readings from the data. As illustrated in Figure 1, these results can then be communicated to a customer facility 17 as well as repair depots 18.

- The process in Figure 2 enables creating a new anomaly definition, step 20.
- 5 Creating a new anomaly definition is simply creating a rule established for detecting specific anomalous conditions that may be indicative of a forthcoming problem with a remote asset. Once done, Step 22 is to identify a variable, "Y". An example of a Y variable include, but is not limited to, lube oil temperature and manifold air pressure. The Y variable may be a monitored parameter or some quantity derived from one or
  - 10 more monitored parameters as specified in the anomaly definition. Step 24 is to identify candidate "X" variables. Candidate X variables are those factors, monitored parameters or quantities derived from monitored parameters, that are believed to partially explain the variability observed in the Y variable. Examples of X variables include, but is not limited to engine coolant temperature, engine speed, ambient
  - 15 temperature and barometric pressure. Choice of candidate X variables can be experienced-based, engineering knowledge-based, or data-based. Step 26 involves "cleaning" the data. In this step, unacceptable or invalid data are removed from further evaluation. For example, if the information collected is outside a range of what has been determined as acceptable data as specified in the anomaly definition,
  - 20 this data is deleted during this step. Though not a conclusive list, such bad data may include instances when a locomotive's gross horse power is less than 5800; when engine speed is less than 1045 revolutions per minute (RPM); or when missing values are detected.

- After the data is cleansed, a standardization model is built 27. The first step
- 25 28 is to center the data. In this statistical script, the "X" variables are centered at 0 by subtracting the mean of each variable from individual observations. Specifically, if  $X = (x_1, x_2, x_3, \dots, x_n)$  and then, centered  $X = (x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x})$ . The next step, 30 is to run a regression algorithm, or technique, such a stepwise

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

regression algorithm on the data set. An appropriate computer package, such as

SAS or S Plus, can be used to run this step. Using such an algorithm will determine a degree of linear association between the data collected from the remote asset and the parameters included in the new anomaly definition. Running a regression algorithm will also assist in determining what may be causing a anomalous reading. For  
5 example, if the lubrication temperature is running high, the process must determine whether it is running high because of ambient temperature or barometric pressure. The stepwise regression algorithm allows the process to compensate the data for effects that are caused by environmental parameters.

The process next reviews information that is collected. Specifically data plots,  
10 step 32, are reviewed and evaluated. Plots of raw data (e.g. Y variable vs. time,  $x_i$  vs. time, and cross correlation plots) for unusual observations, relationships amongst the X variables and indications of additional potential X variables are examined. The diagnostics data, step 34, is next reviewed. For example, the diagnostic data plots and metrics are reviewed. Once the data is reviewed, a decision gate, step 36, is  
15 reached. Here the process makes an assessment of whether the data, or model, fits established parameters. This assessment is made from the information gathered from the raw data plots, diagnostic plots and diagnostic numerical outputs to evaluate the adequacy of the model. If the model is not acceptable, the system modifies the model, step 38, by returning to step 30 to rerun the stepwise regression and then proceeding  
20 through steps 32, 34, 36 again.

If the model is acceptable, the next step is to rebuild a centered model, step 40. This is done by un-centering the X variables where the mean for each variable is added back to individual observations. The next step is to run a regression algorithm on rebuilt un-centered data, step 42. At this phase of the process, diagnostics are  
25 again reviewed, step 44, by examining plots of raw data (e.g. Y variable vs. time,  $x_i$  vs. time, and cross correlation plots) for unusual observations, relationships amongst the X variables and indications of additional potential x variables and metrics data. Within this step, the system will again assess whether the model fits predetermined parameters given the appearance of raw data plots, diagnostic plots and diagnostic  
30 numerical outputs.

Next, the system will calculate residuals at step 46 where a residual is calculated as

Residual = Actual Y – Predicted Y,

or in other words, the difference of information remaining from the actual Y variable when compared to the predicted Y variable. The process will then re-scale the residuals, or build a standardized variable at step 48. The standardized variable is  
5 known as the residual plus target. A standardized Y variable value is calculated by adding the residual to the target value to calculate a standardized Y variable value. The target value is the mean of the Y variable as calculated in steady state.

The system will then review the effect of standardization on individual locomotives at step 50. Here, the system will assess variability decreases due to  
10 standardization by the locomotive by examining box plots, which are graphical representations of the data, and values of standard deviation of individual locomotives before and after standardization. If the results do not meet a set of predetermined factors, the system will return to step 24 to re-identify candidate X variables. However, if the results do meet a set of predetermined factors, the system  
15 will begin to trend the information 52.

In the Trend Information 52 segment of the process 2, the first step, 54, is to optimize parameters lambda/sigma in an Exponentially Weighted Moving Average (EWMA) chart. An Auto-Regressive Integrated Moving Average technique (ARIMA) is used to calculate a value, lambda. ARIMA is a family of time series forecasting models  
20 that rely on a tendency of the next item in some series to relate not just to prior values (auto-regressive), but to a moving average of prior values.

An EWMA chart is a control chart for variables data (data that is both quantitative and continuous in measurement, such as a measured dimension or time). It plots weighted moving average values. A weighting factor is chosen by the user to  
25 determine how older data points affect the mean value compared to more recent ones. Because a EWMA chart uses information from all samples, it detects much smaller process shifts than a normal control chart would.

This includes determining tolerable false alarm rate; determining the size of shift EWMA should detect; and optimizing the value of a trend smoothing constant,  
30 (weight =  $\lambda$ ), and width of control limits, in k sigmas, where k specifies the width of the control charts limits as a multiple of the standard errors of the plotted EWMA's,

given the above mentioned information. EWMA is used to create an exponentially weighted moving average (EWMA) control chart, which is used to determine whether a process is in a state of statistical control and to detect shifts in the process average. Each point on the chart indicates the value of the EWMA for a measured subgroup.

- 5 For example, the EWMA for a given subgroup ( $E_i$ ) is defined recursively as

$$E_i = \text{Lambda}(\bar{X}) + (1 - \text{Lambda})E_{i-1}$$

Where  $\bar{X}$  represents current observation and  $i > 0$ . Within step 56 the EWMA limits on data, or information, are calculated and then limits on standardized data, or information, are calculated, step 58.

- 10 Once the limits for the standardized data are calculated, step 58, the next step, 60, is to format an algorithm to facilitate implementation via a diagnostic compute engine (DE), such as the General Electric Transportation System Remote Monitoring and Diagnostics Service Center Diagnostic Engine (DE). This is the compute engine which takes in the anomaly definitions generated by the process described above and  
15 applies them to incoming observations. Now, the process is implemented in DE, step 62. Supervised verification and validation of the algorithm is performed by passing or using an external field data test set with a known output value in order to assess performance and validate the methodology employed.

- 20 While the invention has been described in what is presently considered to be the preferred embodiment, many variations and modifications will become apparent to those skilled in the art. Accordingly, it is intended that the invention not be limited to the specific illustrative embodiment but be interpreted within the full spirit and scope of the appended claims.

## CLAIMS

## WHAT IS CLAIMED IS:

1. A method for processing monitored data from a remote asset to optimize maintenance and operation schedules, the method comprising the steps of:
  - collecting data from the remote asset;
  - building a data set based on the data collected 20, 22, 24, 26;
  - 5 applying statistical scripts to the data set to create a statistical model 28, 30, 32, 34, 36;
  - comparing the statistical model to the data set 40, 42, 44, 46;
  - creating a standardization model from the compared statistical model and the data set 48, 50;
  - 10 applying a trending algorithm to the data 52;
  - deriving statistical based control limits 54, 56, 58;
  - applying the control limits to a new set of collected data 60, 62.
2. The method of claim 1 wherein the building a data set further  
15 comprises using a series of algorithms to build the data set.
3. The method of claim 2 wherein using a series of algorithms further comprises:
  - creating an anomaly definition 20; and
  - 20 identifying variables to monitor 22, 24.
4. The method of claim 1 wherein the applying statistical scripts further comprises:
  - using a statistical script for centering the data at a predetermined variable 28;
  - 25 using a statistical script for running a stepwise regression script on the centered data 30;
  - using a statistical script for un-centering the data 40.

5. The method of claim 1 further comprising eliminating extraneous variables discovered during comparing the statistical model to the data set 46, 48.

6. The method of claim 1 wherein applying a trending algorithm to the data further comprises determining control chart type limits 54.

7. The method of claim 6 wherein determining control chart type limits comprises applying a time series modeling optimization technique 54 to determine control chart type limits.

8. The method of claim 1 wherein applying the control limits to a new set of data further comprising the steps of:

recording the standardization model and data thresholds for implementation in an automated monitoring and diagnostic system 56, 58, 60; and  
determining a remote asset's maintenance and operation schedule 62.

9. The method of claim 8 further comprising the step of notifying a user of results after determining a remote asset's maintenance and operation schedule.

10. A system for processing monitored data from a remote asset to evaluate and determine a status of the remote asset with minimum user interface, the system comprising:  
a remote asset 5;  
a data gathering module 10 located at the remote asset to collect data about the remote asset;  
a monitoring and diagnostic service center 16;  
respective network interfaces located at the remote asset and the monitoring and diagnostic service center;  
a processor 11 to manage sending, evaluating, and receiving the data;  
a process 2 residing in the processor which uses algorithms to build a data set based on the data gathered 20, 22, 24, 26, uses statistical scripts on the data to determine a statistical model 27, and uses a time series modeling optimization

technique to trend information 52 to determine the remote asset's maintenance and operation schedules.

11. The system of claim 10 wherein the remote asset 5 transmits data collected by the data collection module 10 to the process 2 located at the monitoring and diagnostic service center 16.

12. The system of claim 10 further comprising:  
a customer facility 17;  
10 a repair depot 18;  
respective network interfaces located at said customer facility 17 and repair depot 18;  
wherein the remote asset's maintenance and operation schedules determined by the process 2 are communicated from the monitoring and diagnostic service  
15 center 16 to the customer facility 17 and repair depot 18.

13. The system of claim 10 wherein the process 2 comprises a sub-process to compare the statistical model to the collected data.

20 14. The system of claim 10 wherein the time series modeling optimization technique 54 used in the process is an Exponentially Weighted Moving Average technique.

15 15. The system of claim 14 wherein an Auto-Regressive Integrated Moving Average technique is used to calculate a value in the Exponentially Weighted Moving Average technique 54.

30 16. The system of claim 10 wherein the processor further comprises a diagnostic compute engine to determine whether the data is within a specified operating range as determined by the process 62.

17. The system of claim 10 wherein the results determined by the process are communicated to a user 62.

18. A method 2 for processing monitored data from a remote asset 5, the  
5 method 2 comprising the steps of:  
creating an anomaly definition which comprises acceptable parameters 20;  
identifying a first variable 22 ;  
identifying a variable which is used to monitor the first variable 24;  
deleting data that is outside of the specified parameter 26;  
10 centering data about a specified variable 28;  
running a first stepwise regression algorithm 30;  
determining whether data is within acceptable parameters 32, 34, 36;  
un-centering data 40 if data is within acceptable parameters;  
running a second stepwise regression algorithm on the un-centered data 42;  
15 calculating residuals 46 based on a calculated first variable compared to a  
predicted first variable;  
building a standardized variable 48 based on calculated residuals;  
assessing data 50 before and after standardization to determine whether results  
are within an acceptable limit;  
20 trending the data 52 to determine whether the assessed data are within a  
predetermined limit;  
reporting results to a user 62.

19. The method of claim 18 wherein trending the data 52 further comprises  
25 optimizing an Exponentially Weighted Moving Average technique 54 and applying  
the optimized Exponentially Weighted Moving Average technique to the standardized  
variable 58 and the monitored data 56.

20. The method of claim 18 wherein reporting results to a user further  
30 comprises formatting an algorithm to facilitate implementation with a diagnostic  
compute engine 60 and implementing the diagnostic compute engine 62.

21. The method of claim 18 further comprising determining whether the data is within acceptable parameters 36.

5 22. The method of claim 21 further comprising re-running 38 the first stepwise regression 30 if the data is not within acceptable parameters 32, 34, 36.

10 23. The method of claim 18 further comprising identifying a new second variable which is used to monitor the first variable 24 and continuing through a build standardization model 27 when assessed data 50 is not within an acceptable limit.

1/2

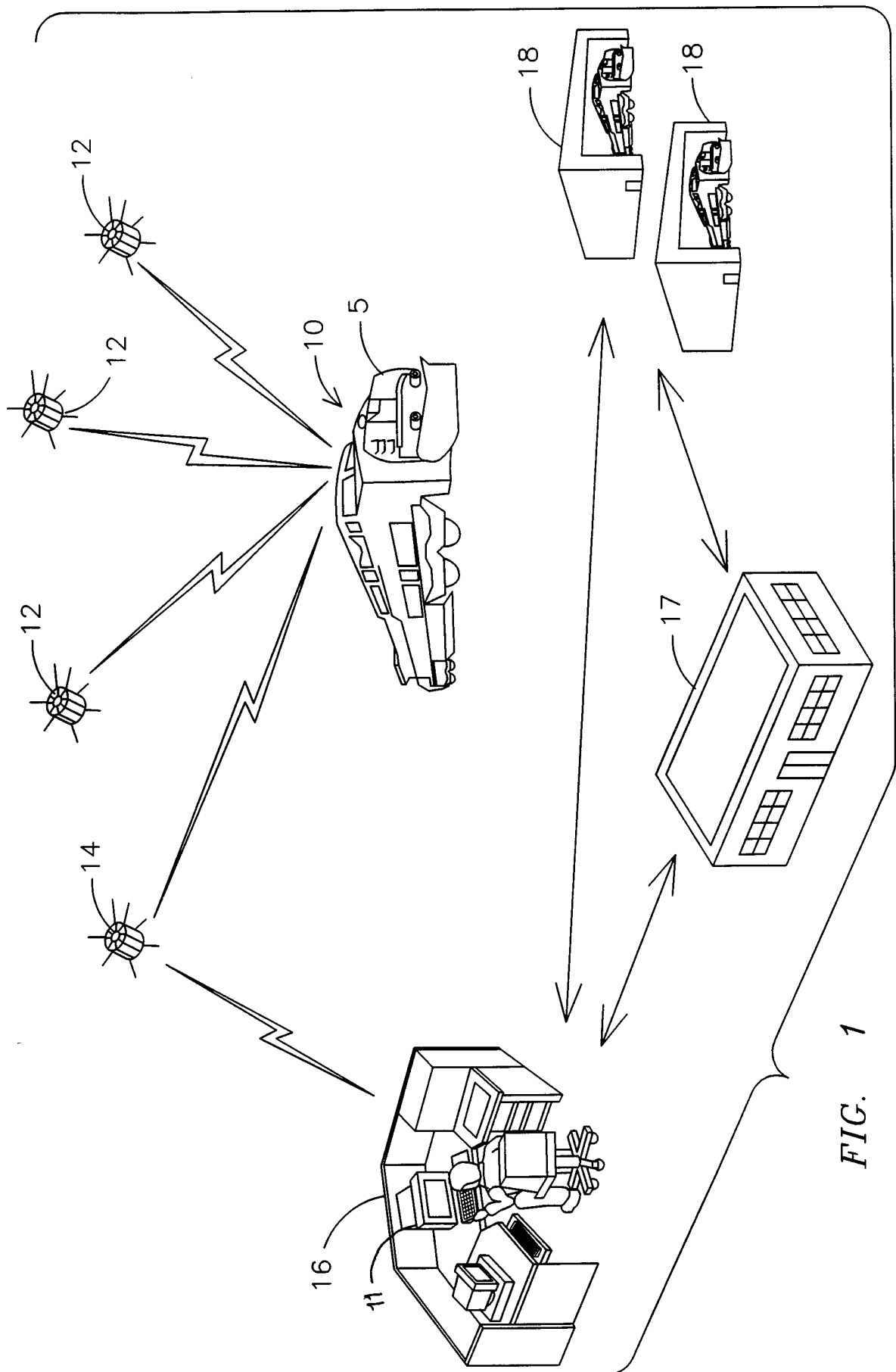


FIG. 1

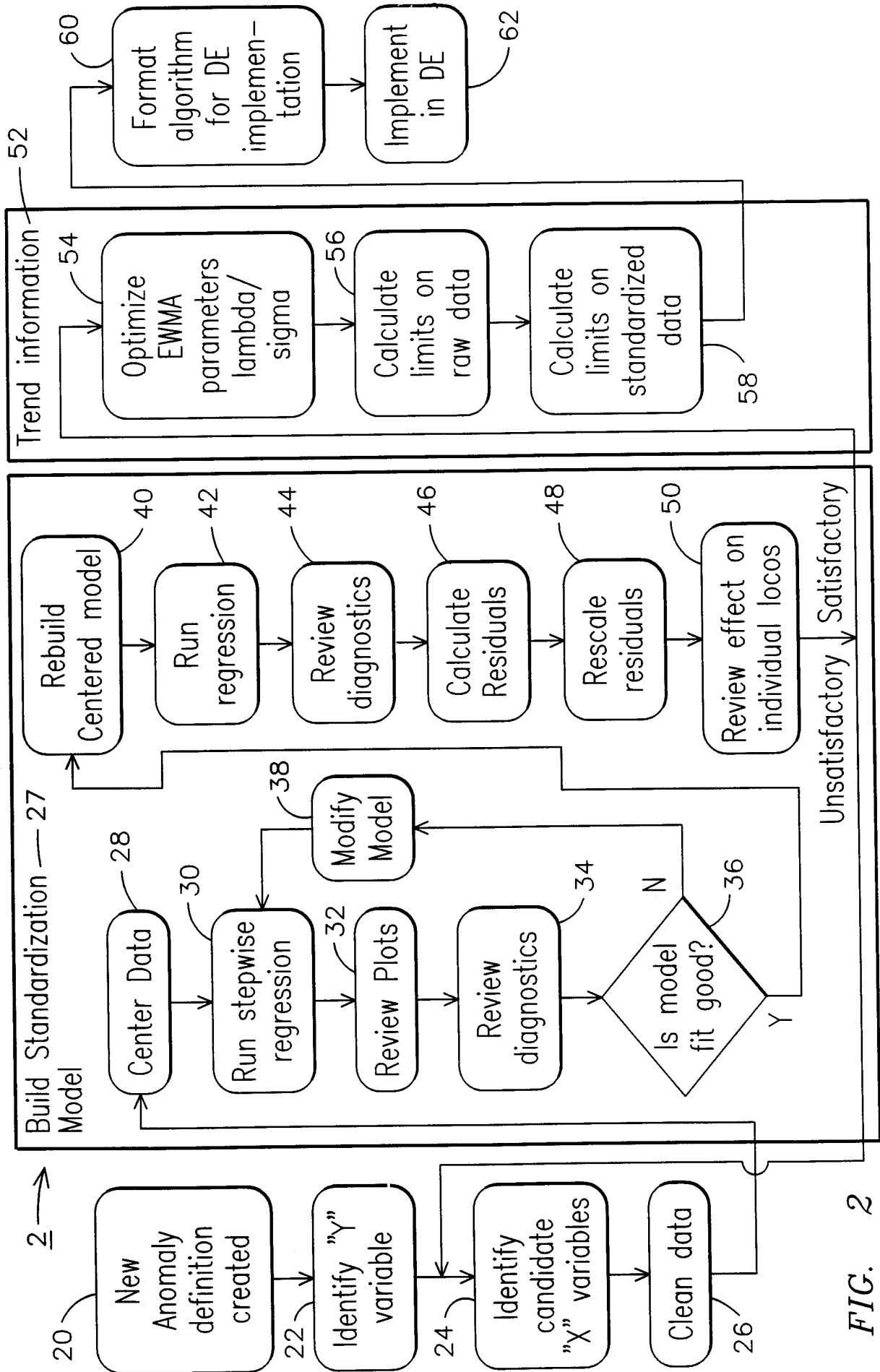


FIG. 2