



(51) International Patent Classification:
Not classified

(21) International Application Number:
PCT/IB2020/000149

(22) International Filing Date:
30 January 2020 (30.01.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/798,580 30 January 2019 (30.01.2019) US

(71) Applicant: **AEROQUAL LTD.** [NZ/NZ]; 460 Rosebank Road, Avondale, Auckland, 1026 (NZ).

(72) Inventors: **HENSHAW, Geoffrey, Stephen**; 24 Alberton Avenue, Mt Albert, Auckland, 1025 (NZ). **WILLIAMS, David, Edward**; 93 Doves Bay Road, Kerikeri, 0294 (NZ). **MILES, Elaine**; Apartment 15, 20 Central Road, Kingsland, Auckland, 1021 (NZ). **MISKELL, Georgia**; 6 Hart Street, Mount Maunganui, 3116 (NZ). **WEISSERT, Lena**; 2a Seaton Road, Auckland, 0630 (NZ).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: METHOD FOR CALIBRATING NETWORKS OF ENVIRONMENTAL SENSORS

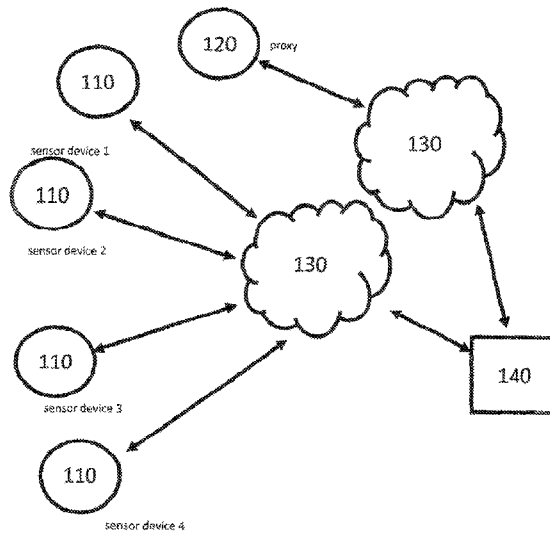


Figure 2

(57) Abstract: Multiple low cost individual sensors communicate with a server. A proxy sensor communicates with the server. The server periodically compares information from the individual sensors with information from the proxy sensor. The server validates and accepts information from individual sensors. The server changes gain and offset values for individual sensors providing information that is not validated by the comparing.



Published:

- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

Method for Calibrating Networks of Environmental Sensors

Air quality regulatory monitors require frequent, rigorous calibration using standard reference materials in order to provide accurate and traceable environmental data. Such a calibration method is a significant operational cost and for that reason not economically feasible for low cost environmental sensor networks. A calibration method that reduces such costs is required in order to achieve useful data from high spatial density networks of low cost environmental sensors.

BACKGROUND OF THE INVENTION

One approach to sensor calibration is periodic co-location with a higher quality data source (reference) and to calculate calibration parameters such as slope and intercept using linear regression. However this approach has problems such as a) periods of missing data where the sensor has been removed for calibration, b) high costs due to intensive fieldwork requirements in regularly moving sensors for co-location, and c) inability to adapt to changes over time such as calibration drift. A modification of this approach is to make the reference mobile so that it visits each of the sensors in the network for a period of time to calibrate them. However this is not reliable since the pollutant levels must vary over a wide enough range during the time of the visit in order for the calibration to be reliable.

Another approach is to use a remote calibration technique based on cross-correlations between network locations often

described as 'blind' calibration because this calibration method relies on no ground truth data to guide expectations. This method is not however robust against missing data and involves training a network using verified data to correct for extraneous influences. A practical limitation can be providing a sufficiently large training set of sufficiently good data.

Another approach is data fusion, which combines pollutant measurements with an atmospheric model. This approach is not traceable and assumes the stability of the sensor response which is not always the case.

Needs exist for improved calibrations of environmental sensors.

SUMMARY OF THE INVENTION

This invention provides a new approach that is economical to use, does not assume sensor stability, does not rely on training data and can update in real-time. The new calibration method adjusts the gain and offset of the sensor measurement to match the probability distribution of the sensor result to that of a proxy reference.

If $X_i(t)$ denotes the true data value at location i and time t , then the sensor data, $Y_i(t)$, if the sensor response is linear over a range of expected measurements, will satisfy:

$$X_i(t) = a_0 + a_1 Y_i(t)$$

{1}

The new method then defines a reference data set from one or more regulatory or non-regulatory monitors located some distance from the sensor, which are a good estimate for the true concentration. It is assumed that the frequency distribution of measurements from a sensor and a suitable reference averaged over a time that captures the diurnal variations are similar. For air pollutants, such distributions are log-normal. Slope and offset estimates can then be derived by matching the mean and variance of the sensor to the reference distribution Z , which is an estimate of the true distribution X .

Let $Z_i(t-t_d:t)$ denote the reference data and $Y_i(t-t_d:t)$ denote the sensor data for site i over the time interval $(t-t_d:t)$, then the slope, $\hat{\alpha}_1$ and offset, $\hat{\alpha}_0$, estimates for the corrected data are:

$$\hat{\alpha}_1 = \sqrt{\frac{\text{var}\langle Z(t-t_d:t) \rangle}{\text{var}\langle Y(t-t_d:t) \rangle}}$$

{2}

$$\hat{\alpha}_0 = E\langle Z(t-t_d:t) \rangle - \hat{\alpha}_1 E\langle Y(t-t_d:t) \rangle$$

{3}

Where $E\langle \rangle$ denotes the arithmetic mean evaluated over the time period $(t-t_d:t)$ and $\text{var}\langle \rangle$ denotes the arithmetic variance about the mean.

Then the estimate of X_i at time t , $\hat{X}_{i,t}$:

$$\hat{X}_{i,t} = \hat{a}_0 + \hat{a}_1 Y_{i,t}$$

{4}

According to {2-4}, $E\langle\hat{X}(t - t_d: t)\rangle = E\langle Z(t - t_d: t)\rangle$ and $\text{var}\langle\hat{X}(t - t_d: t)\rangle = \text{var}\langle Z(t - t_d: t)\rangle$. Thus, if the distributions are characterized by only two parameters, then the site distribution over sample time t_d is constrained to be the same as the reference distribution. More generally, the site distribution is constrained to be similar to the reference distribution for values around the mean. In practice this is not a limitation: the prior averaging time period, t_d , imposes an averaging filter on the data. It is empirically chosen to obtain a reasonable estimate of the distributions of Z and Y but is also chosen to average the short-term fluctuations and emphasize the longer-term, regular component of the concentration variation within the distribution of values. It is found that the frequency distribution of this longer term, regular component of variations of air pollutant concentration does indeed tend to be well correlated over extensive spatial regions, even though the variations might occur at different times of the day, or though local concentrations might show rather large excursions from the mean for relatively short times compared with the averaging time. If short-term fluctuations are particularly large, then data truncation can be used: the mean and variance of the test and reference distributions are simply matched over a data range that excludes the extreme fluctuations. The adjusted test data will then capture these fluctuations reliably.

The new approach can be combined with a rule-based framework that places constraints on the adjustment of the parameters \hat{a}_0 and \hat{a}_1 . For example, the time series of \hat{a}_0 and \hat{a}_1 can be smoothed and the smoothed trend values used in equation 4. Other parameters can be used to determine whether the reference data set is reliable for particular conditions: for example particular wind directions might transport pollutants from other parts of the region being measured and bias the distributions with respect to one another. If such a condition occurred, it is a simple matter to implement a rule that uses the previous long-term trend value of the parameters \hat{a}_0 and \hat{a}_1 , only updating when the reference condition returns to that where the distribution similarity is known to be valid.

If the sensor measurement comprises a linear combination of two parameters then the method would be modified thus.

$$C_{\text{pollutant}} = b_1 * C_{\text{sensor1}} + b_2 * C_{\text{sensor2}} + b_0$$

Where $C_{\text{pollutant}}$ is the pollutant measurement derived from sensor1 and sensor2 and b_1 , b_2 and b_0 are parameters determined by the calibration method. The sensor is calibrated by matching the $C_{\text{pollutant}}$ measurements to that of a proxy ($P_{\text{pollutant}}$) and calculating the b_1 , b_2 and b_0 values so that the Bayesian formulation defined below is satisfied.

$$\mathbb{P}(C_{\text{pollutant}} | C_{\text{sensor1}}, C_{\text{sensor2}}, \hat{b}_j) = \mathbb{P}(P_{\text{pollutant}})$$

The values of b_1 , b_2 and b_0 are determined by using the mean ($E(x)$), standard deviation ($SD(x)$) and a third moment, ($RT[x]$) of $C_{\text{pollutant}}$ and $P_{\text{pollutant}}$, where $RT[x]$ is defined below.

$$RT[x] = \sqrt[3]{\frac{\sum(x-E[x])^3}{n}}$$

Using initial values of $b_1 = b_2 = 1$; $b_0 = 0$, the parameters are optimised to minimise the objective function, $f(P_{\text{pollutant}}, C_{\text{pollutant}})$ defined as:

$$f(P_{\text{pollutant}}, C_{\text{pollutant}}) = (E[P_{\text{pollutant}}] - E[C_{\text{pollutant}}])^2 + (SD[P_{\text{pollutant}}] - SD[C_{\text{pollutant}}])^2 + (RT[P_{\text{pollutant}}] - RT[C_{\text{pollutant}}])^2$$

In principle, any number of sensors could be employed such that the result is a linear combination of the values given by each sensor. In that case, the number of moments of the distributions to be determined and matched using the objective function should equal the number of coefficients to be determined.

These and further and other objects and features of the invention are apparent in the disclosure, which includes the above and ongoing written specification, with the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the hourly-averaged raw data from the analyzers versus raw sensor data and calibrated sensor data using the new method.

Figure 2 is a diagram showing an example representation of a sensor network able to be calibrated by the new method.

Figure 3 is a process flow diagram showing how a sensor is calibrated.

Figure 4 is a process flow diagram showing how data is compared.

Figure 5 is a process flow diagram showing how gain and offset values for a sensor are calculated.

DETAILED DESCRIPTION

We used a running averaging time $t_d = 72$ hours and examined a network of O_3 and No_2 sensors. Figure 1 shows as an example the hourly-averaged raw data from the analyzers versus raw sensor data and calibrated sensor data using the new method. The method was successful in calibrating and routinely updating correct coefficient values for a poor sensor without any initial co-location requirement, despite the corrections being large.

Figure 1 shows scatterplot of analyzer and semi-blind corrected sensor hourly-averaged data (earlier weeks = lighter triangles and later weeks = darker triangles). B is similar to A using raw sensor data. C is stepping \widehat{a}_0 , and D is stepping \widehat{a}_1 .

The method of deriving calibration coefficients by matching the mean and standard deviation of the data to that of a proxy has been shown to be a robust means of correcting data from drifting or mis-

calibrated environmental sensors. The method is based on the idea that running over a time that is sufficiently long to remove the influence of short-term fluctuations but sufficiently short that results can be obtained in a practically useful time whilst still preserving the regular diurnal variations, the mean and standard deviation of measurements are highly correlated given an appropriate choice of reference. Reference choice made on the basis of distance or land-use similarity has been demonstrated to be effective. A running time of 72 hr is appropriate for diurnal air pollutants but this interval may be longer or shorter for different environmental measurements in different conditions. Sensor data corrected using this method measure reliably for data averaged over intervals from 1 minute. Use of data truncation in the proxy matching identified where the proxy and sensor data distributions differed and could be used to determine the reliability of the results.

Figure 2 is a diagram showing an example representation of a sensor network able to be calibrated by the method. The sensor network can include a sensor device 110 and a proxy 120 in communication with a server device 140 over a network 130. The network 130 can include various types of networks, such as the Internet, a virtual private network (VPN), a wide area network (WAN), a local area network (LAN or VLAN), etc. The proxy 120 can include various types of reference instruments such as a regulatory air quality station, a calibrated gas instrument, a mobile instrument, a satellite or a computer model. The sensor devices 110 can transmit sensor data

wirelessly to the server device 140 over the network 130. The proxy 120 can transmit proxy data wirelessly to the server device 140 over a network 130. The server device 140 can include a local storage device to store data associated with sensor data received from the sensor devices and proxy. The server device determines whether each sensor device 110 needs to be calibrated by comparing its data with data from the proxy 120. If it does, then the server device needs to be calibrated, calculates new gain and offset data using the method described. The sensor device can then send these gain and offsets back to the sensor device 110 over network 130 and adjust the sensor data stored on the server device 140.

Figure 3 is a process flow diagram showing how a sensor is calibrated. The server 140 receives sensor data 210 and proxy data 220 and compares it 230 to see if the sensor data has drifted. If it has drifted server 140 calculates 240 new gain and offsets using the method described and stores them in the server. The gain and offset can be uploaded 260 to the sensor 110. If the sensor data has not drifted, the sensor data is accepted, and no new gain and offset is calculated or uploaded 250. The output of the process is validated sensor data 270 in the server 140.

Figure 4 is a process flow diagram showing how data is compared 230. Taking the sensor data 210, a window of time of a specified length and that includes at least one unvalidated sensor observation is selected 320. The sensor data and proxy data 220 are sampled over the window 340. Comparative statistics between the

sampled sensor and proxy data are calculated 350. The statistics are compared against threshold values, setting off warnings if they are exceeded 360. The number of consecutive warnings is counted 370, and if it equals a specified threshold, an alarm is triggered. If the alarm is triggered, the system enters the recalibration process 240. If the alarm is not triggered, the process ends 250.

Figure 5 is a process flow diagram showing how gain and offset values for a sensor are calculated 240. A sample of uncalibrated sensor data 210 and proxy data 220 is collected. The gain and offset values are set to default values 430 and are applied to generate new sensor data 440. The new sensor data is assessed against the proxy data using an objective function 450. If the objective function is satisfied the offset and gain values are accepted 460 and the process ends. If the objective function is not satisfied, then new input parameters are generated for the objective function which generates new gain and offset values 470. New sensor data is calculated and comparison process is repeated 440.

While the invention has been described with reference to specific embodiments, modifications and variations of the invention may be constructed without departing from the scope of the invention.

Claims

1. A method comprising calibrating a first sensor in a network of sensors further comprising collecting data from the first sensor and collecting data from a proxy reference which represents a good estimate for the true concentration at the site of the first sensor and adjusting gain and offset of the first sensor so that over a selected period of time first probability distributions of adjusted first sensor measurements closely matches second probability distribution of the proxy reference.
2. The method as in claim 1 wherein the selected period of time over which the first and second probability distributions are matched is from 1 hour to 1000 hours.
3. The method as in claim 1 wherein the first sensor measures air pollutants including ozone, nitrogen dioxide, sulfur dioxide, carbon monoxide, benzene, particulate matter, hydrogen sulfide, ammonia, nitric oxide, carbon dioxide, methane.
4. Method as in claim 1 wherein the site of the first sensor includes outdoor and indoor environments.
5. Method as in claim 1 where the proxy reference is a regulatory air quality station, an air quality instrument with a traceable calibration, a collection of sensors, a satellite based instrument, a mobile air quality instrument or a computer model.
6. Method as in claim 1 where the proxy reference second probability distribution is a weighted combination of different data sources, including a weighted combination of data from a regulatory

air quality station, an air quality instrument with a traceable calibration, a collection of sensors, a satellite based instrument, a mobile air quality instrument and a computer model.

7. Method as in claim 1 where the proxy reference is selected on the basis of similarity in land use to the site of the sensor.

8. Method as in claim 1 where the proxy reference is a mobile instrument on an unmanned aerial vehicle.

9. Method as in claim 1 where the gain and offset values calculated from matching first probability distributions of the first sensor and the proxy reference second probability distributions are stored locally on the first sensor and remotely on a server.

10. Method as in claim 1 wherein long-term trend values of gain and offset are employed.

11. Method as in claim 1 wherein a set of rules is used to determine whether the correction adjusting of the gain and offset should be corrected.

12. Method as in claim 11 where the set of rules includes wind direction, wind speed, rate of change of humidity, or any combination of these.

13. Apparatus comprising a server, multiple individual sensors communicating with the server, a proxy sensor communicating with the server wherein the server is adapted to obtain proxy sensor data from the proxy sensor and sensor data from each of the individual sensors, wherein the server is adapted to compare the sensor data from each of the individual sensors with the

proxy sensor data, and wherein the server is configured to accept the data from each individual sensor and separately validate incoming and future data from each individual sensors from which data has been accepted.

14. The apparatus of claim 13 wherein the server is adapted to calculate new gain and offset for each individual sensor having data that has not been accepted and to upload new gain and offset into each individual sensor having data that has not been accepted.

15. The apparatus of claim 13 wherein the server is adapted to select windows of data to be received from each individual sensor and to compare the data received during each window wherein the server is adapted to calculate statistics to compare statistics against a threshold and to count a number of process warnings, to accept data if the threshold is met and if the number of process warnings is low and to trigger a recalibration process for any individual sensor from which data is received when the threshold is not met and the number of process warnings is high.

16. The apparatus of claim 15 further comprising wherein the server is further adapted to set gain and offset values of each individual sensor to default values, wherein the server and is adapted to receive the data with the default values from each individual sensor and to compare that data with the data from the proxy sensor using an objective function, is adapted to generate new parameters for the gain objective and offset as each individual sensor is the function and is not satisfied for each individual sensor, and the server is adapted to

recalculate gain and offset values for each individual sensor, and wherein the server is adapted to compare the data from each individual sensor with the data from the proxy sensor and to accept the individual sensor if the objective function is satisfied and if the objective function is not satisfied then new input parameters are generated for new gain and offset values for that individual sensor.

17. The apparatus of claim 16 wherein the server is adapted to repeat the comparing process for the new data with the new gain and offset for that individual sensor.

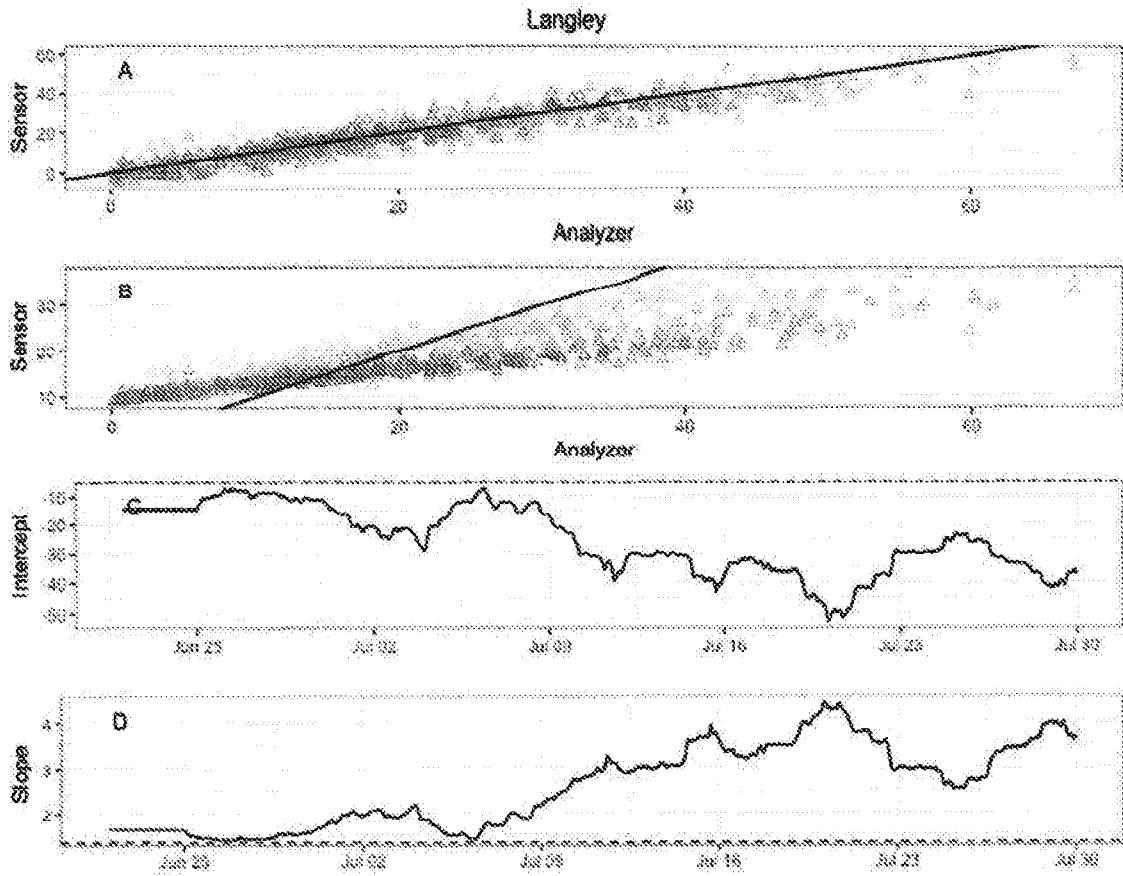


Figure 1

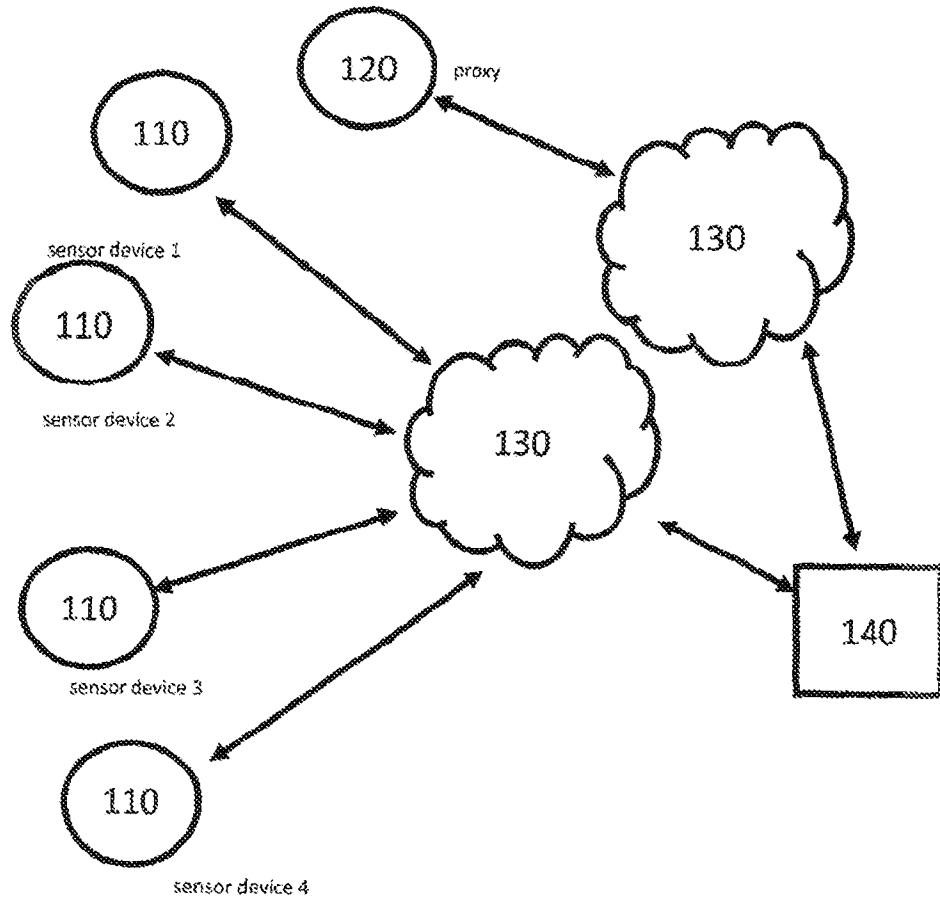


Figure 2

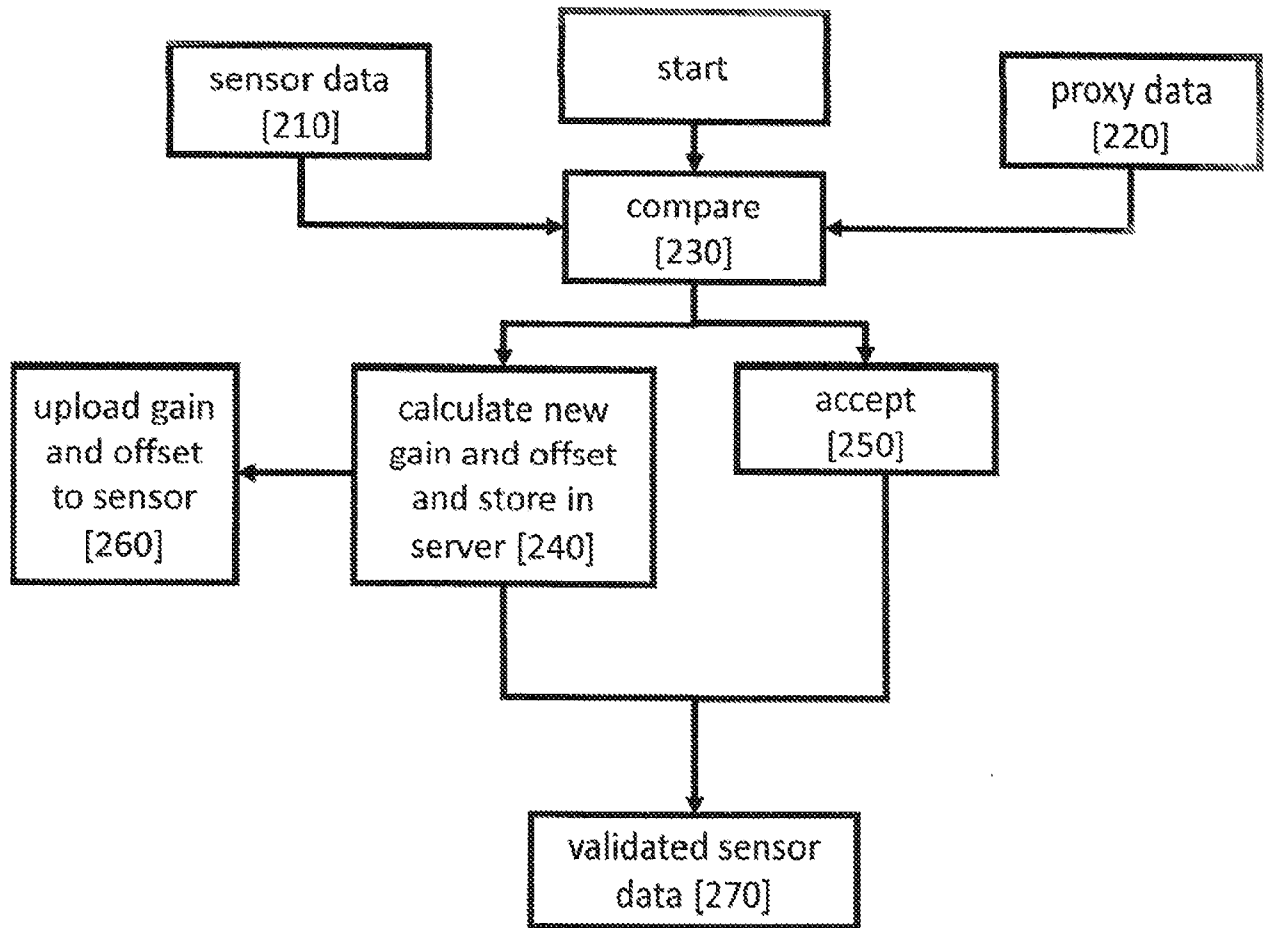


Figure 3

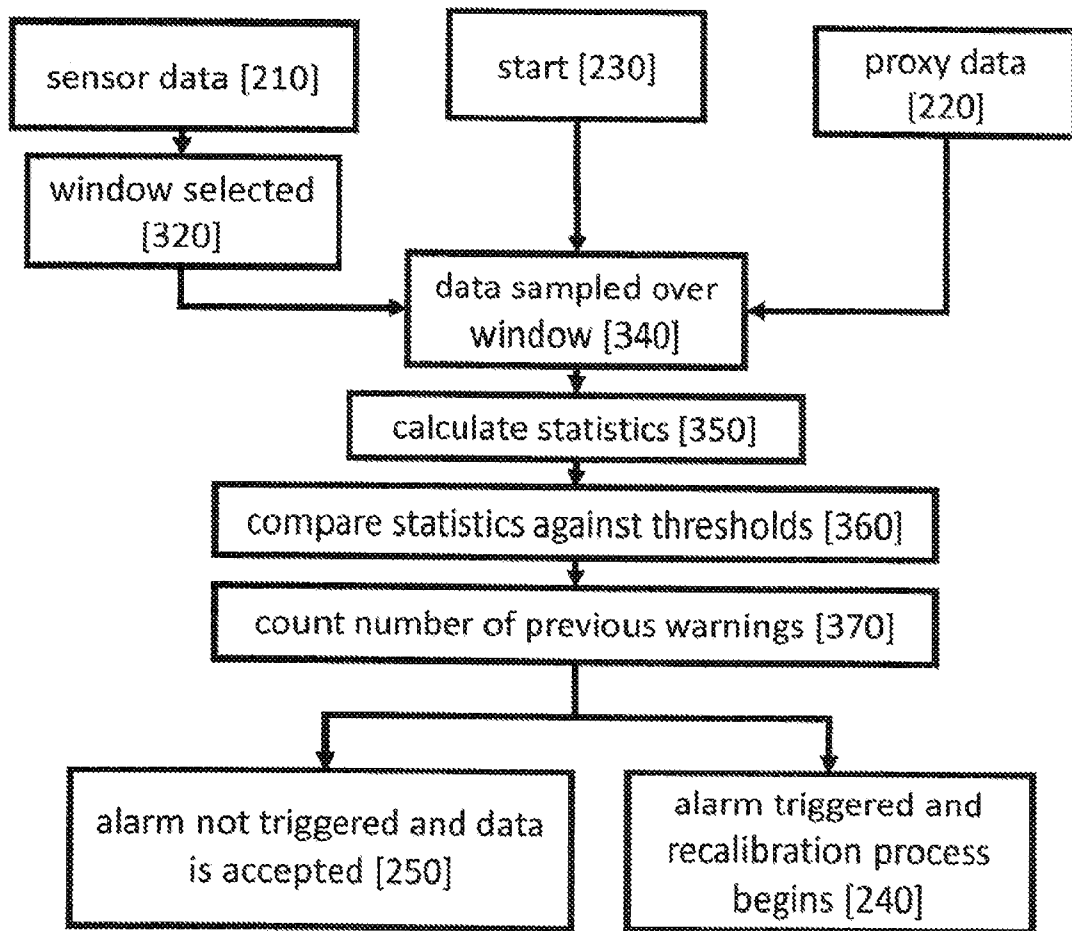


Figure 4

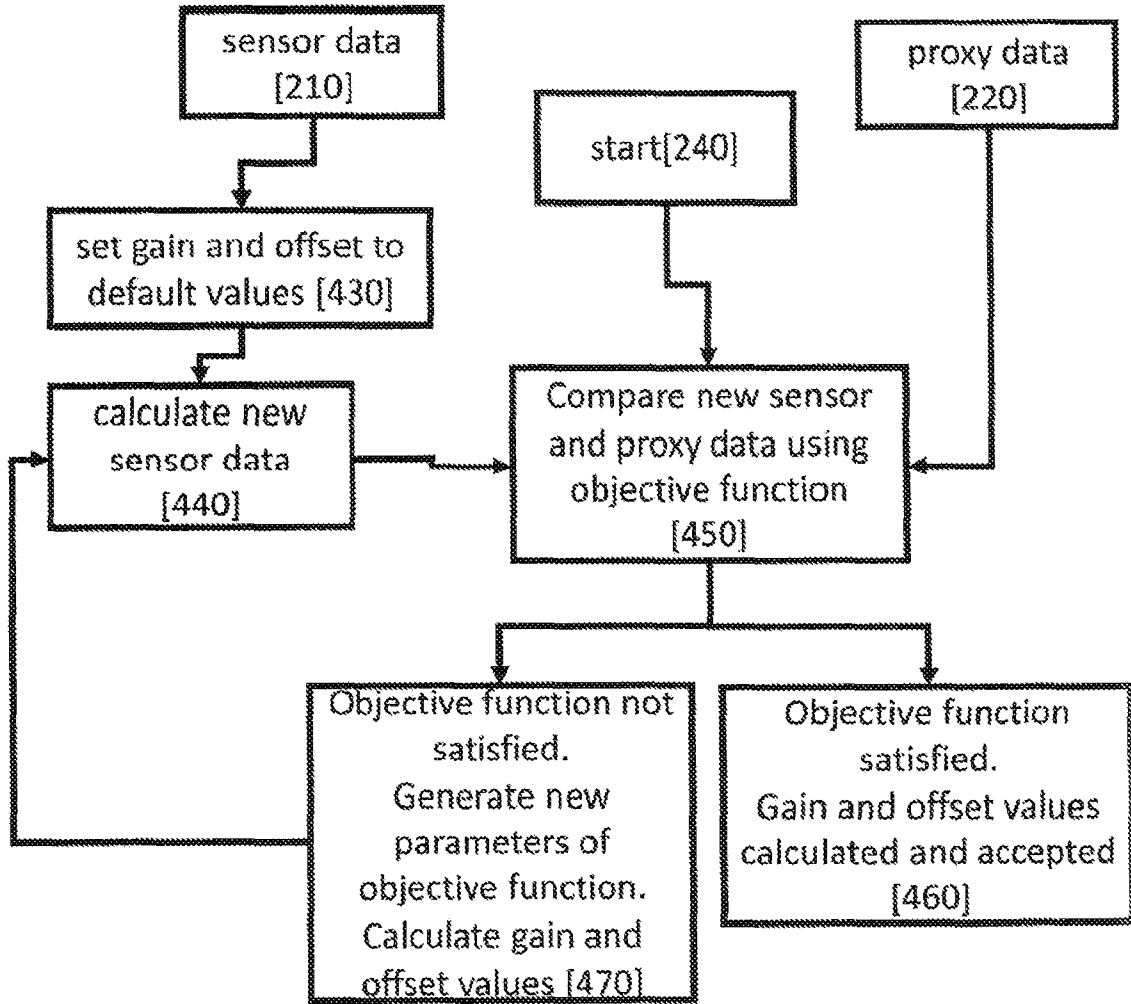


Figure 5