Regression estimation of gas concentration in closed-loop control ventilation systems

ALEXANDER V. ZORIN
Saint Petersburg State University
Department of Mathematics and Mechanics
28 Universitetsky pr.,
198504 Saint Petersburg
RUSSIA

Abstract: The analysis of the gas concentration is an important problem in a wide area of industry and several home applications. It concern medicine, chemical industry, mountain, gas industry. The important significance such analysis has in building and construction engineering, housing et cetera.

Now different systems are applied to a control of concentration, including the closed-loop ventilation schemes. In a lot of them sensors with high accuracy are applied. But, as a rule, applied effective techniques of estimation of concentration in a feedback control are rarely.

In this paper presented new closed-loop scheme of monitoring and control of gas concentration. It uses performance, reliable approach of regression function evaluation that suits well for several of estimation problems and applications. In paper described application of the least-squares method for solving the estimation of concentration problem. Cases of nonlinear dependence regression function of time and estimation approaches for such cases considered. Means of obtaining the interval bounds for derived regression function are in details described. Both the cases of a linear dependence concentration of time and nonlinear dependence explained. Choices of nonlinear regression function linearization considered. Influence of confidence intervals on action of ventilation control system, its reliability and reaction under sudden perturbations investigated closely. Methods of definition upper and lower regression bounds depending on sensors instrumental errors are given. These bounds determine the prognosis of regression function and system behavior in general and used as a condition for closed-loop control.

Implementations of the approach and simulation results are reported. Modeling of action the real system with this method presented. Comparison with the similar ventilation control schemes without applying regression estimation is given. Described method provides high reliability and robustness with computational simplicity using standard precision sensors and closed-loop clarity and minimalism.

Key-Words: Confidence interval, Control system, Gas concentration, Mean square value estimates, Measurements, Observation error, Robustness, Sensor, Standardized normal distribution, Ventilation

1 Introduction

Consider a system, composed of pollution gas source, indoor closed-circuit ventilation scheme, sensors and closed-loop control module, as shown in the Figure 1.

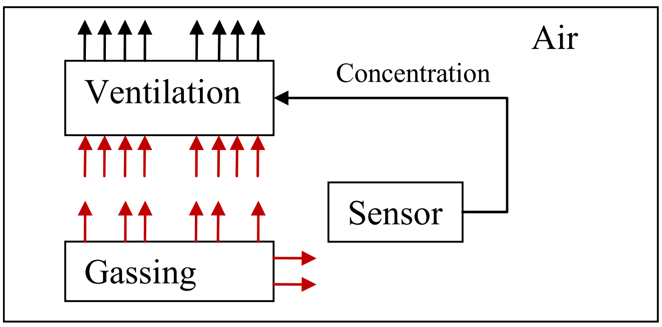

Fig.1: Ventilation system with direct control.

2 Problem Formulation

2.1 Problem statement

Problem of such scheme is the impossibility of instantaneous ventilation control in allowable values caused by inertance of the ventilation system and delay of Gassing source–Sensor–Ventilation control communication. Also there is a limitation on permanent action of ventilation. It can be caused by a number of reasons like energy saving, concentration bounds constraints and so on.

Thus, the lag effect in such control schemes makes a problem. There is a necessity of control system delay correction. It can be realized by prediction of concentration behavior with function of time. Generally, the time dependence of concentration is unknown. Problem is also definition of confidence intervals for the prognosis function. As additional condition we should have the value of concentration strictly in given bounds in all time of system action.

2.2 Proposed Method

The effective solution of these problems proposed. Suggested scheme uses estimation for gas concentration (Figure 2).

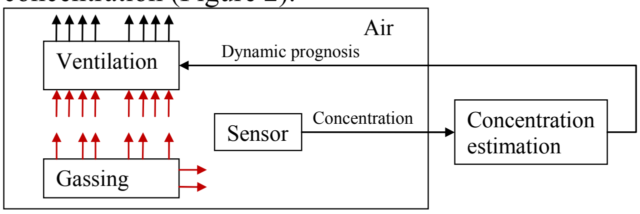

Fig.2: Ventilation system based on estimation of gas concentration.

The algorithm of ventilation system operation includes stages: (**Algorithm 1**)

1. The ventilation is off, the probable growth of gas concentration is predicted. Regression with its bounds is calculated. The system operates in the tracking mode.
2. If regression upper bound exceeds defined upper boundary, then ventilation turns on. The system operates in the ventilation mode.
3. If regression lower bound reaches defined lower boundary, then ventilation turns off. The system passes in the tracking mode again.

Such algorithm of operation brings economic benefit. It allows applying more power ventilation components [1,5,7] without additional expenditures that increases reliability. Mathematical technique sets up required certainty and durability.

Key phases of the method:

1. Find gas concentration dependencies of time when ventilation turned on and turned off.
2. Calculations of regression function for obtaining point estimations of concentration.
3. Computations of interval bound estimations of regression with defined confidence band.
4. Forming set of control actions based on estimated regression function and its bounds.

As a measure of method efficiency there is matching its implementation with existing ventilation control systems. The purpose is to clear up facility and applicability of assumed method.

2.3 Main assumptions

Consider that is regression dependence of one random variable Y – indoor gas concentration from other random (or nonrandom values) variables $X_1,\ldots,X_k$. Variables $X_1,\ldots,X_k$ mean the input gas flow concentrations, speed of gas flow and other real characteristics. Random variable Y is named output response, and $X_1,\ldots,X_k$ – input factors.

Regression model of observational data shows statistical dependence between variables. Such model considered below.

Suppose, that observed response value is sum:

$$y = f(x) + \varepsilon, \qquad (1)$$

where $f(x)$ is function of input factors $x = (x_1,\ldots,x_k)$, and ε – stochastic variable is called regression model error. The stochastic error ε means influence of some random factors, response variability and measurements error.

The problem statement in terms of regression analysis is *to produce* estimation of function $\hat{y} = \hat{f}(x)$, by stochastic experimental data sample (x_i, y_i): $y_i = f(x_i) + \varepsilon_i, i = 1,\ldots,n$

Assume, that regression function can be described by unknown parameters $\beta_1,\ldots,\beta_l$ or its vector

$$\beta = (\beta_1,\ldots,\beta_l): y_i = f(x_i,\beta) + \varepsilon_i, i = 1,\ldots,n \qquad (2)$$

In this case estimation of regression leads to evaluation of unknown parameters: $\hat{y} = \hat{f}(x,\hat{\beta})$

For example, in linear regression model regression function can be represented as:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1,\ldots,n, \qquad (3)$$

where β is two-element vector parameter (β_0, β_l).

Such function is regression dependence $f(x,\beta)$ linear in relation to unknown vector parameter β.

In most cases gas concentration dependence of time is linear regression function, but response y dependence of factors x may be not.

3 Problem Solution

3.1 Definition the variation functions of gas concentration

When ventilation is off and gas flow is almost invariable the concentration of gas in internal air varies directly as exponential function. Since in an index of an exponent is a slow variable then it is possible to ignore curvature of an exponential function on some considered interval. Thus there is a possibility of linear approximation of varying concentration function.

When ventilation is on it is also possible to suppose linear estimation of concentration varying.

For sudden gas emissions it is difficult to forecast precisely concentration dependence on time, but in most cases it can be defined by the second-order function.

The laws defining gas concentration dynamics in closed space:

$$\begin{aligned} \dot{c}v + \dot{v}c &= c_1F_1 + c_2F_2 - c(F + F_B) \\ \dot{v}c &= c(F_1 + F_2 - F - F_B) \end{aligned} \qquad (4)$$

where c_1, c_2 – gas concentration in exuding gas mixture and absorbing air ($c_2 = 0$ – absorbing air is pure),

F_1, F_2 – flows of gas mixture and ventilation air, F, F_B – flows of input and output ventilation air, v – volume of enclosed space.

Obtain set of equations:

$$\begin{cases} \dot{c}v = (c_1 - c)F_1 + (c_2 - c)F_2 \\ \dot{v} = F_1 + F_2 - (F + F_B) \end{cases} \qquad (5)$$

3.1.1 Conditions of concentration dynamics in the system:

With automatic ventilation control:

$c(0) = c(t_1)$, where t_1 – time of ventilation start.

$v(0) = v_0, F_2 = F_B$

$c_1 = \begin{cases} c_{10} \\ c_{10}t \end{cases}$, $c_2 = 0$, the possibility of sudden gas ejection is considered (fast variation of gas concentration – $c_{10}t$ in input exuding gas flow), where c_{10} – gas concentration in exuding gas mixture.

$\dot{c}v = (c_1 - c)F_1 + cF_2$, $v = v_0 = const$

$\dot{c} = (c_1 - c)\frac{1}{T_1} - c\frac{1}{T_2}$, where T_1, T_2 – rates of concentrations changes in flows, $T_1 = \frac{v_0}{F_1}$, $T_2 = \frac{v_0}{F_2}$

$$\dot{c} = -(\frac{1}{T_1} + \frac{1}{T_2})c + \frac{c_1}{T_1} \qquad (6)$$

Transfer function for the system:

$$w_{aut}(p) = \frac{T_2}{T_1T_2p + (T_1 + T_2)}. \qquad (7)$$

$c(\infty) = \frac{c_{10}T_2}{T_1 + T_2}$ – steady-state value of gas concentration in enclosed space air.

Without ventilation:

$F_2 = F_B = 0$, $F_1 = F$, $c_1 \geq c$, $c_1(t) = c_{10}t$ – there is permanent gas ejection into enclosed space air with concentration of gas in mixture $c_{10} = const$.

$$\dot{c} = (c_1 - c)\frac{F_1}{v_0}, \ \dot{c} = -\frac{1}{T_1}c + c_1\frac{1}{T_1}. \qquad (8)$$

Transfer function:

$$w(p) = \frac{1}{T_1p + 1} \qquad (9)$$

$c(\infty) = c_{10}$ – steady-state value of gas concentration in enclosed space air is equal to gas concentration in ejection gas mixture.

3.2 Solving the regression problem for obtaining point estimations of concentration

For obtaining point estimations of linear variation of concentration the simple linear regression will be used. Regression model of stochastic data definition is named simple linear regression, if, as (3)

$$y_i = \beta_0 + \beta_1x_i + \varepsilon_i, i = 1,...,n, \qquad (10)$$

where $x_1,...,x_n$ - defined values of factors; $y_1,...,y_n$ - observed response values; $\varepsilon_1,...,\varepsilon_n$ - independent, normally distributed unobservable variation values with zero average of distribution and identical (unknown) dispersion $N(0,\sigma)$; $\beta_0,...,\beta_1$ - unknown parameters subject to estimation.

In assumed model response y depends on factor x only and all dispersion of experimental points induced by observation errors (measurement result) of response y only. The observation errors of x in this model supposed much less than observation errors of y, so it can be ignored.

These assumptions suppose to evaluate confidence intervals for linear regression estimations exactly.

Estimation of regression parameters is generated by the least-squares method.

Proposed approach is to enter a criterion of error between response and regression. The regression parameters estimations defined so that to minimize this error. Such criterion choice provides simple obtaining of the calculated formulas for estimations:

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n}\varepsilon^2 = \sum_{i=1}^{n}(y_i - \beta_0 + \beta_1x_i)^2 \to \min(\beta_0, \beta_1) \qquad (11)$$

In addition to simplicity of calculations the parameters estimations, obtained using least-squares method, have a number of advantageous statistical properties.

Least-squares estimations of β_0 and β_1 are gained on condition that $Q(\beta_0, \beta_1)$ is minimum.

Decision variable $\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 \cdot x$ pass over experimental points as follows (Fig.3):

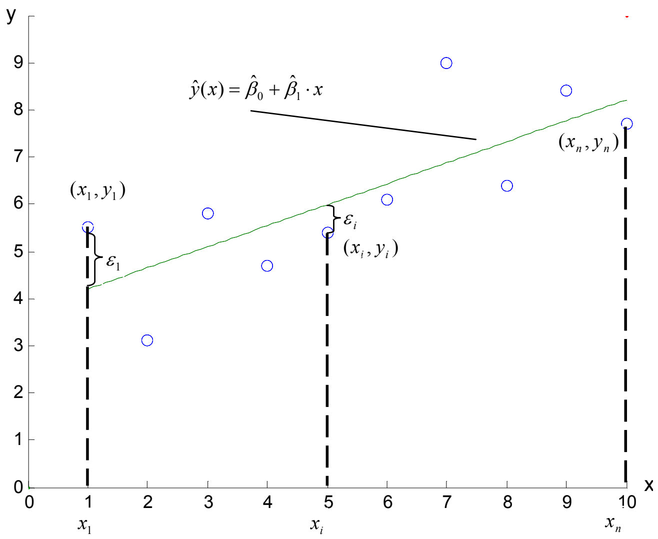

Fig.3: Linear regression estimation of concentration obtained by least squares method.

Model of considered technique with samples of regression problems solutions is implemented in Matlab [13].

3.3 Calculation of interval estimations with determined reliability

Let define confidence interval (interval estimate in general) for average of distribution m_x of normally distributed variation X assumed, that its dispersion σ^2 is known. So obtain point estimation for average of distribution m_x using random sample $(x_1, x_2, ..., x_n)$ of range n: $\bar{x}_B = \frac{1}{n}\sum_{i=1}^{n} x_i$. (12)

Mean and dispersion of sample (12) are equal m_x and σ^2/n accordingly. In conditions of the suppositions distribution of sample means will be normal: $\bar{x}_B \in N(m_x, \sigma^2/n)$.

Assume standardized sample mean as a statistic of

$$S(\Theta, \hat{\Theta}): \; z = \frac{(\bar{x}_B - m_x)\cdot\sqrt{n}}{\sigma}, \qquad (13)$$

which also will be distributed normally with zero mean and unit variance: $z \in N(0,1)$. (14)

Let set a significance level α and define quantiles of the standardized normal distribution (14) $z_{\alpha/2}$ and $z_{1-\alpha/2}$, connected together in this case by relation:

$$z_{1-\alpha/2} = -z_{\alpha/2}; z_{\alpha/2} < 0. \qquad (15)$$

Confidence probability $\gamma = 1-\alpha$ for considered case is defined as probability of presence of a random variable z in a range $[z_{\alpha/2}, z_{1-\alpha/2}]$, i.e.

$$P[z_{\alpha/2} < z < z_{1-\alpha/2}] = F_z(z_{1-\alpha/2}) - F_z(z_{\alpha/2}) = \gamma. \qquad (16)$$

on condition that $[z_{\alpha/2} < z < z_{1-\alpha/2}]$.

Use it for definition of the parameter δ and, thus, of interval estimation of average of distribution.

Statistics (13) is perfectly decreasing function on $\Theta = m_x$, therefore, stated inequality will take the form of $z_{\alpha/2} < z < z_{1-\alpha/2}$, and so

$$\bar{x}_B - \frac{\sigma\cdot z_{1-\alpha/2}}{\sqrt{n}} < m_x < \bar{x}_B + \frac{\sigma\cdot z_{1-\alpha/2}}{\sqrt{n}}. \qquad (17)$$

This relation defines a confidence interval $[\Theta_1, \Theta_2]$ for an estimation m_x with confidence probability $\gamma = 1-\alpha$:

$$[\bar{x}_B - \frac{\sigma\cdot z_{1-\alpha/2}}{\sqrt{n}}, \bar{x}_B + \frac{\sigma\cdot z_{1-\alpha/2}}{\sqrt{n}}]. \qquad (18)$$

On the basis of a relation (15) these confidence interval can be retyped in the form:

$$[\bar{x}_B + \frac{\sigma\cdot z_{\alpha/2}}{\sqrt{n}}, \bar{x}_B - \frac{\sigma\cdot z_{\alpha/2}}{\sqrt{n}}]. \qquad (19)$$

In our case, formula defining a confidence interval for $y(x) = \beta_0 + \beta_1 x$, determines also bounds of prediction values errors $y(x) = \beta_0 + \beta_1 x$ with confidence probability $1-\alpha$ (where the defined probability of extreme situation is equal α):

These bounds are two hyperbolas defined by the equations:

$$y = \hat{y} \pm t_{1-\alpha/2}(n-2)\cdot s \cdot \frac{\sqrt{1 + \frac{(x-\bar{x})^2}{(s_x)^2}}}{\sqrt{n}}, \qquad (20)$$

where $t_{1-\alpha/2}(n-2)$ – quantile of a random variable distributed by Student deviation with $(n-2)$ degrees of freedom, s – permanent mean-square deviation of measurement errors, n – number of measurement points, $\bar{x}$ – mean value of the factor, $(s_x)^2$ – dispersion of the factor.

Axis of these bounds is the estimation of direct regression $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.

Obviously, that error of estimation $y = \beta_0 + \beta_1 x$ is least at $x = \bar{x}$ and increases in process of moving away $\bar{x}$ (Fig.4).

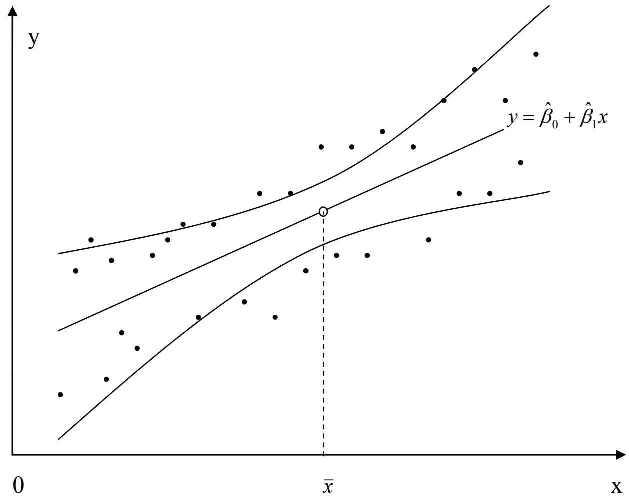

Fig.4: Interval bounds of linear regression estimation of concentration.

3.4 Recursive calculations

For calculation efficiency improvement the recursive calculations are applied.

On addition of new experimental data (iteration step) the probability characteristics are computed using the recurrent formulas.

Computational complexity for iterative formula:

$$y_{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} x_i, i = (1,..,n) \quad (21)$$

for all calculations of mean values $y_{\Sigma j}, j = (1,..,n)$ is $O(n^2)$, and it for all same calculations by recurrent formula:

$$y_i = y_{i-1} + \frac{1}{i}(x_i - y_{i-1}), i = (1,..,n); \quad (22)$$

$$y_j = y_{\Sigma}, j = (1,..,n)$$

is $O(n)$.

The application of recurrent expressions for stochastic characteristics allows to dispose computing resources more rationally and to improve real-time control time.

4 Simulation Results

Assume the scheme shown on Fig.2 with regression estimation by the least-squares method.

The algorithm (**Algorithm 1**) is applied to control ventilation operation.

The following modeling results are obtained (Fig.5) in Matlab application [13]:

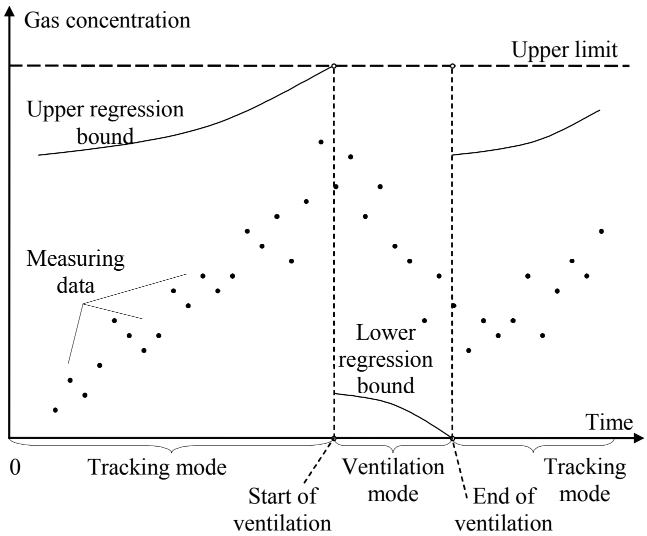

Fig.5: Simulation results of ventilation control based on regression estimation of gas concentration.

Thus, using regression estimation in ventilation control systems predicts dangerous situations safely. Real-time control with minimum time delay and defined confidence interval is provided. As shown, control started earlier than concentration measurements should be equal to gas concentration upper limit. Therefore proposed technique brings safe control response time. Electric power economy also provided by using variable ventilation in contrast to direct ventilation systems, which make fixed energy charges.

5 Conclusion

In this paper proposed new stochastic based approach of controlling gas concentration in building and underground constructions. The approach established on regression estimation and confidence bounds definition for observational data. The least squares method suits well to estimate concentration variation. It is easy to embed the technique in ventilation control systems.

Advantages of the method:

1 Energy saving due to the periodic action of ventilation.

2 To lowering of maintenance charges and other operating costs because of short intervals of operation and possibility of system diagnostics in a down time.

3 It is a rather simple to implement on microcontroller in embedded platform with sensors and control unit.

References:

[1] Grot, R.A., Hodgson, A.T., Daisey, J.M., and Persily, A., "*Indoor Air Quality Evaluation of a*

New Office Building", ASHRAE Journal, September, 1991.

[2] John Lewis, S. Lakshmivarahan, and Sudarshan Dhall. *Dynamic Data Assimilation: A Least Squares Approach*. Cambridge University Press, Cambridge, UK, 2006.

[3] Geir Evensen. Sampling strategies and square root analysis schemes. *Ocean Dynamics*, 54:539–560, 2004.

[4] Y. Bar-Shalom, X.-R. Li, *Estimation and Tracking – Principles, Techniques ans Software*, Artech House, 1993.

[5] Lagus, P.L., Kluge, V., Woods, P., and Pearson, J., "Tracer Gas Testing within the Palo Verde Nuclear Generating Station Unit 3 Auxiliary Building", Proceedings of the *20th NRC/DOE Air Cleaning Conference*, Boston, MA, 1988.

[6] Data Assimilation via Error Subspace Statistical Estimation. P.F.Lermusiaux, A.R. Robinson, *Monthly Weather Review*, Vol. 127, July 1999.

[7] US Nuclear Regulatory Commission, Regulatory Guide 1.197 "Demonstrating Control Room Envelope Integrity at Nuclear Power Reactors", USNRC, Washington, DC, 2003.

[8] Bendat J.S., Piersol A.G. *Random Data: Analysis & Measurement Procedures*, New York: Wiley, 2000, 594 p.

[9] Feustel, H.E., "Mathematical Modeling of Infiltration and Ventilation," Proceedings 10th AIVC Conference *Progress and Trends in Air Infiltration Research*, Air Infiltration and Ventilation Centre, 1989.

[10] Murphy, K.G. and Campe, K.M., "Nuclear Power Plant Control Room Ventilation System Design for Meeting general design criterion 19", Proceedings of the *13th AEC Air Cleaning Conference*, Baltimore, MD,1974.

[11] US Nuclear Regulatory Commission, Generic Letter 2003-01 "Control Room Habitability", USNRC, Washington DC, 2003.

[12] Sherman, M.H., "Analysis of Errors Associated with the Passive Ventilation Measurement Technique," *Building & Environment* 24(2), pp. 131-139, 1989.

[13]www.mathworks.com/matlabcentral/

[14] Zorin A.V. Regression Method of Estimation in Ecological Observations Processing Software // In: Proc. of the *8-th St. Petersburg Conference on Soft Computing and Measurements*, St. Petersburg, 2005.