Data Processing Approaches for the Measurements of Steam Pipe Networks in Iron and Steel Enterprises

Luo Xianxi, Yuan Mingzhe, Wang Hong and Li Yuezhong

Additional information is available at the end of the chapter

1. Introduction

Steam is important secondary energy resource media in iron& steel enterprises, amounting to nearly 10% of the whole energy consumption. When an enterprise is running, if all the produced steam can meet all the demands and no steam is bled, the overall energy efficiency can be effectively improve. Thus the complex networks of steam pipes and steam production scheduling systems were set up. Obviously, steam scheduling has to depend on the real time measured data from the steam pipe networks. Accurately measuring the variables of pressure, temperature and flow rate is essential to secure the safety and economic efficiency. It's also necessary for accumulating the amount of steam production or the consumption to calculate the energy cost for each working procedure.

With the help of Energy Management System (EMS), all the data collected from the distributed instruments. In practice, the measurements of pressure and temperature are usually accurate enough for the application except that the sensors or the transducers fail. However, the measurements of mass flow rate are not so accurate because of the complex nature of the steam itself, lacking of high precision measuring instruments, the impact of interference and information transmitting network failures and other reasons. The reliability of the measurement of mass flow rate is poor.

When the steam mass flow measurement values deviate from the actual values to a certain extent, the automatic control system may largely deviate from the process requirements substantially. Even worse steam bleeding or accident could happen [1]. Therefore, it is not a satisfactory to decide or adjust the production process according to the data from the flow rate meters [2]. In energy management, the accumulation differences between production and consumption make it difficult to calculate the energy costs, analyze the segments of

irrational use of steam, and find the weakness in the management links. Therefore, improving the reliability the steam flow rate measurement data is essential for the normal production and energy conservation in iron& steel enterprises.

The objective of the work may be depicted by figure 1. The real time data (mainly referring to the mass flow rate variables) would be mainly processed by three approaches. By fault data detection and reconstruction, the fault data would be picked out and the real value would be reconstructed or estimated. By gross error detection, the data with gross error would be discovered and re-estimated. By data reconciliation, the random error would be decreased and the quality of the data would be further improved.
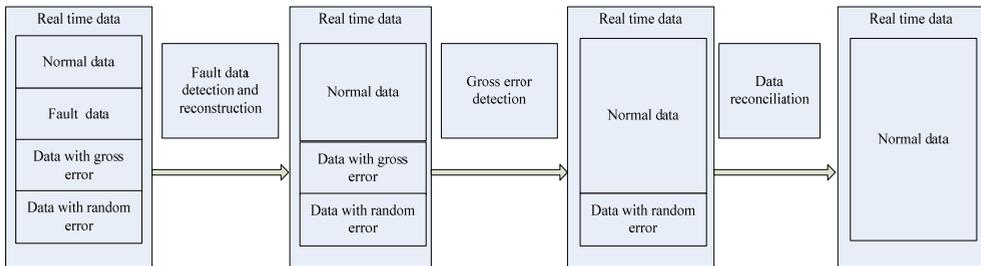

Figure 1. The objective of the work

For fault data detection and reconstruction, the reasons of the low accuracy of steam flow rate data are introduced. By applying the statistical process control theory[3] to determine univariate and multivariate process control limit to monitor the abnormal data online. The approach to calculate the real data (mass flow rates) through the thermal and hydraulic mathematical models of the steam pipe networks is proposed.

For the section of gross error detection, the definition of the problem is introduced. And the two basic gross error detection approaches, the Measurement Test (MT) method and the Method of pseudonodes (MP), are demonstrated.

For the section of data reconciliation, the constrained least-squares problems stated in the section of gross error detection is discussed in detail, including the assumptions for the application,the constraint equations and the selection of the weighted parameter matrix.

The presented approaches of data processing can be programmed for computers to determine the abnormality and improve the precision of the mass flow rate data.

2. Fault data detection and reconstruction

2.1. The reasons of the poor accuracy of steam flow rate measurement

There are many reasons for the poor accuracy of steam measurement data. The most important reasons can be concluded as follows:

1.   The working conditions of the instrument deviating from the designed conditions

At present, the mass flow rates are mostly deduced by the volume flow rates and the density. However, the changes of temperature and pressure in the transmission process would lead to the density of steam deviate from the original designed value [3]. The measurement errors would be very large [1]. Any more, some superheated steam would change into a vapor-liquid two-phase medium, which making precision worse.

2. The Complexity of Steam Characteristics
    With the ambient temperature changing, the total amount of the condensate water in the transmission process will be different. That makes difference between the amount of the production and the consumption of steam. In addition, the steam pipe leakage will add to the difference. Therefore, the accumulated readings are always doubtful.

3. The Occurrence of wearing or damage to the Key Components
    As the orifice differential pressure flow rate meter being in use for long time, the size of aperture would differ from the original size because of the adhered foreign bodies, or erosion by the durative high temperature steam flow. As the parameters can not be adjusted on time, and it is hard to calibrate the instrument, the measurement errors will accumulate.

4. The external Interference or the Failure of Data Transmission Channel
    The instrument interfered by the disturbances or failure happening in data transmission channels will make significant errors to the data derived in the control center.

2.2. Determining control limits for fault data detecting

The abnormal data from certain sensors (including temperature, pressure and flow rate sensors) may be characterized as fluctuation quickly with a large magnitude (induced by the poor contact in the instruments), keeping a certain value without any varying (induced by the failure of the data sampling systems) or being outside the normal variable range. The first two cases are not discussed here because they are easy to be discovered. As for the last case, the statistical process control approaches to determine the control limits for data monitoring is applied. When the values of the controlled variables (or functions) exceed the limits, it shows there are abnormal data except that the process is actually abnormal. The statistical process control includes two types--univariate and multivariate control.

2.2.1. Determining the control limits for single variable

The statistical process control (SPC) and control chats were first proposed by Shewhart in quality prediction. Traditional SPC mainly treated single variables. When the value locates outside the normal range, the system would output an alarming signal to notify the operators to check and make sure whether the state is abnormal. In the work, if the process state is actually normally operating, the values of the variables can be judged as fault data. Reasonably determined control limits can reduce the probability of false alarm. Many researches focus on the problem [4-6]. In the work, empirical distribution function combined with the principle of "3σ"[3]is applied to determine the control limits of different variables.

Denote $X \in R^{n \times 1}$ as one of the measurement data matrix, n is the number of the samples, each row of X is a measured sample. If $X = (x_1, x_2, ..., x_n)^T$, the range of the data matrix is:

$$R(X) = \max(x_i) - \min(x_i), (i \le i \le n) \tag{1}$$

Divide the range into N intervals with the same length, each length of the interval is:

$$c = \frac{R(X)}{N} \tag{2}$$

The corresponding intervals are marked as $c_1, c_2, ..., c_N$. So each element in X belongs to one of the intervals and these data are divided into N groups. The probability of the sampled variable value locates in each interval can be estimated as:

$$f_r = \frac{v_r}{N} \tag{3}$$

$r = 1, 2, ... N, v_i$ is the number of the data locate in the ith interval. As the data are independent to each other (usually the hypothesis is reasonable). The probability defined:

$$P_r \{X_i \in c_r\} \approx f_r \tag{4}$$

The average probability density of each interval can be written as:

$$p_r \{X_i \in c_r\} \approx \frac{f_r}{R(X_i)/N} = \frac{N f_r}{R(X_i)} \tag{5}$$

The empirical distribution of X can be got with the lengths of intervals as bottom and p_r as the height. When n and N are large enough, the bar graph would be more similar to the global distribution of X. According to the central limit theorem, when the system has only one stable state, the measured data would follow normal distribution as the sample size grows. So the global distribution function of X can be written as normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{6}$$

Figure 2 shows the distribution graph. The probability of data locating in the range of μ±3σis 99.73%. When the value of X is outside the range, it's reasonable to suspect the value being abnormal (otherwise, the process doesn't normally function). That's the principle of "3σ". In practice, the outside of "3σ"occupying about 0.25% probability is named "red zone", and the range inside μ±2σis named "green zone"(about 95% probability). The intervals between them are "yellow zone". Four control limits are determined by the method.

As the expectation and the deviation of the X global distribution are unknown, and the estimated values with the sample are not accurate especially when the empirical distribution isn't similar to normal distribution, the principle of "3σ" can not be applied directly.
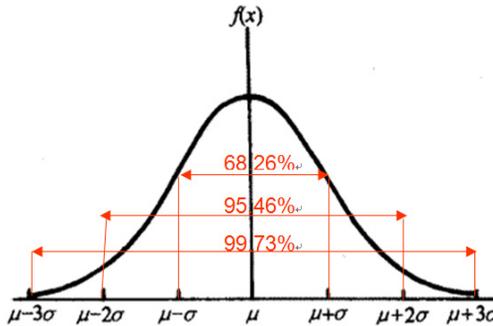
Figure 2. The probability density distribution graph of normal distribution

However, according the principle of "3σ" and the empirical distribution of X, the control limits can be deduced reversely by searching the respective probability of limits.

Two kinds of special instances have to be noticed:

1. No crossing point between empirical distribution and the testing level.
   It shows the corresponding sample value is not appeared in the significance test level range. To determine the limits, the empirical distribution can be linearly extended to outer range or estimated with experience by comparing with the real process performance state.

2. Two more crossing points between empirical distribution and the testing level at one side.
   It shows there are disturbances in the data, which makes the empirical distribution fluctuate at one lateral side. According to principle of "3σ" , find the highest probability interval and search from it to the two laterals for the points first amounting to the probability $(1-\alpha)/2$.

Figure 3-5 show the control limits (4 lines) and monitoring result in a plant. The interval between the two green lines is the "green zone" showing normally working. The intervals outside the red lines are the "red zones" indicating the fault data or process abnormality. The other two intervals are the "yellow zones" which reminds the operators to notice the changes of the data. Figure 4 and 5, in the right side the red and green line are coincident. It can be explained that the outer equipment constraints the variable freely moving.

By determining the limits of single variable, the wild values considered as fault data can easily be discovered. However, two kinds of inherent errors can not be avoided.

The approach of Univariate Statistical Process Control (USPC) can roughly judge fault data. The approach is also appropriate to monitor the temperature and pressure variables.

2.2.2. Determining the control limits of multivariate system based on PCA

The shortage of USPC is the operators being prone to "information overload"when the number of the variables is large. Anymore, it's too rough to judge the normalty of data just by monitor alarming. In practice, the variables maybe constrained by certain inherent functions. It

shows the existence of fault data when the functions are unsatisfied. Principal Component Analysis (PCA) is one of the approaches to discover the relationship among the variables by means of statistics. By determining the control limits of two guidelines, Hotelling's T^2 and SPE (Squared Predictive Error, Q), the multivariate process can be monitored conveniently.
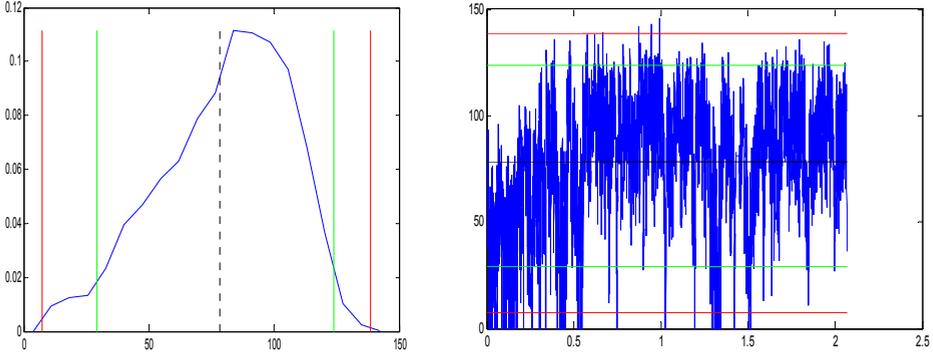

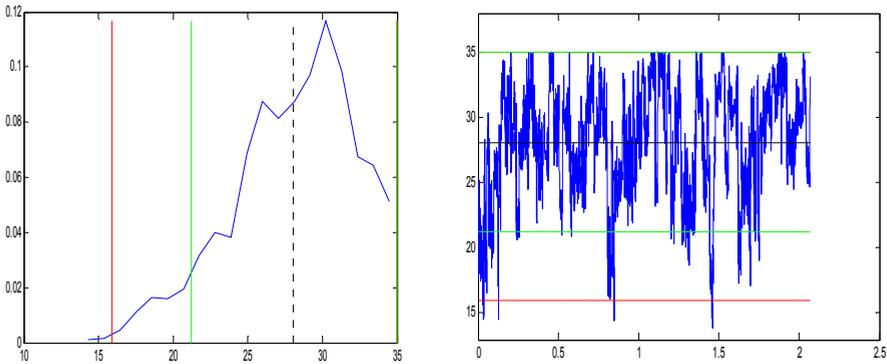**Figure 3.** Monitoring the Steam Flow Rate of the Steel Making Process

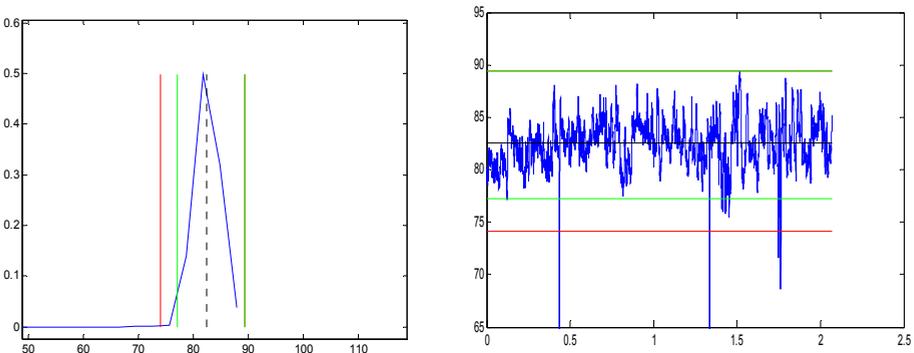**Figure 4.** Monitoring the Flow Rate of Start Steam Boiler

Figure 5. Monitoring the Steam Flow Rata of CDQ (Coke Dry Quenching)

The proposed approach is to distinguish different stable states of the process by empirical distribution function, and determine the control limits for each stable state.

i. Differentiate the stable states and group the samples

Denote the measurement matrix as $X \in R^{n \times m}$, n is the size of samples, m is the number of the monitored variables. The rows of X are the serial samples in time order.

Denote $X_i = (x_{1i}, x_{2i}, ..., x_{ni})$ (i=1,2,…,m) as the ith column. The empirical distribution of X_i can be derived as just mentioned. When the process has two or more stable states, the distribution bar graph would characterize as multi peaks.

Figure 6 is the curve of flow rate and the distribution bar graph of a chemical process in a period. The two peaks represent different demands for steam in different states.

Without loss of generality, the two peaks distribution is to be discussed. If the control limits are determined without considering the different states, the sensitivity of detecting fault data will be too low for the multivariate process. Apply two normal distributions to fit the two peak sections. If the averages and deviations are signed as $X_{i.s1}, \sigma_{i.s1}, X_{i.s2}, \sigma_{i.s2}$, the rule to divide the samples into different groups (different states) is listed as equation (7):
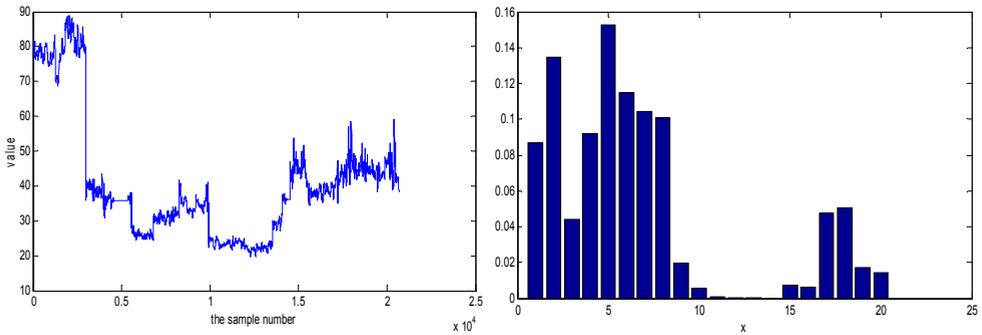

Figure 6. The curve and distribution bar graph of flow rate for a chemical process

$$S = \begin{cases} S_1 & x_{ik} \in [\min(X_i), X_{i.s1} + \gamma\sigma_{i.s1}] \\ S_2 & x_{ik} \in [X_{i.s2} - \gamma\sigma_{i.s2}, \max(X_i)] \end{cases} \quad k = 1, 2, ..., n \quad (7)$$

γ can be selected in the interval of (1, 3), and the total state region should cover the whole range of the sampled data. Ideally, the state regions should not have folded region or the folded region is narrow and with low probability. Or, the variable have no necessity for state differentiating, which is not the case discussed here.

By the rule of (7), the samples are separated into two groups. By the same way, the two groups can be further grouped with other variables. However, it's not suggested to divide the sample into many groups with too few elements in each group. Assume

there are only two groups of samples are derived. They are marked as $X_a \in R^{n_1 \times m}$ and $X_b \in R^{n_2 \times m}$ ($n_1 + n_2 \leq n$, some samples may be discarded because they don't belong to any state region).

ii.   Determining the control limits of different state with PCA

Through grouping, the sample data is more convergence. First standardize X_a, denote:

$$X_a = (x_1, x_2, ..., x_m) \tag{8}$$

In it,, $x_1, x_2, ..., x_m \in R^{n_1 \times 1}$. Calculate each average and standard deviation:

$$M_a = (m_1, m_2, ..., m_m) \in R^{1 \times n_1}. \tag{9}$$

$$S_a = (s_1, s_2, ..., s_m) \in R^{1 \times n_1} \tag{10}$$

The measured data matrix is standardized as:

$$\tilde{X}_a = [X_a - (1 \quad 1 \quad ... \quad 1)^T M_a] diag(\frac{1}{s_1}, \frac{1}{s_2}, ... \frac{1}{s_m}) \tag{11}$$

Derive the eigenvalues of the covariance matrix $C = \frac{1}{n-1} \tilde{X}_a^T \tilde{X}_a$ and orthogonalize the matrix, then the model of PCA can be written as:

$$\tilde{X}_a = \hat{X}_a + E = T_a P^T + E \tag{12}$$

$$T_a = \tilde{X}_a P \tag{13}$$

In the equations, $P \in R^{m \times A}$ is the load matrix. If arrange the eigenvalues of C in descendent order, the former A elements are noted as $\lambda_a = (\lambda_1, \lambda_2, ..., \lambda_A)$. The corresponding unitized eigenvectors form the matrix of P, $P = (P_1, P_2, ..., P_A)$. A is the number of the selected components. $T_a \in R^{n_1 \times A}$ is the scored matrix and E is the error matrix..

The MSPC based on PCA defined two statistical variables: Hotelling's T2 and SPE (Q).

If the jth measurement vector is signed as $\tilde{X}_{aj}$ (row vector), the corresponding scored vector is signed as T_{aj} (ie. The jth row of matrix T_a), then T_{aj}^2 is defined as[7]:

$$T_{aj}^2 = T_{aj} \lambda_a^{-1} T_{aj}^{\ T} \tag{14}$$

When the testing level is α, its control limit can be calculated according F distribution:

$$T_{A,n_1,\alpha}^2 = \frac{A(n_1 - 1)}{n_1 - A} F_{A,n_1-1,\alpha} \tag{15}$$

$F_{A,n_1-1,\alpha}$ is the critical value corresponding to testing level α, degree of freedom A,n1-1.

SPE(also called statistical variable Q)is defined as:

$$Q_{aj} = e_{aj}e_{aj}^T = \tilde{X}_{aj}(I - PP^T)\tilde{X}_{aj}^{\ T}$$ (16)

The control limit can be calculated by:

$$\delta_\alpha^2 = \theta_1(\frac{c_\alpha\sqrt{2\theta_2 h_0^{\ 2}}}{\theta_1} + 1 + \frac{\theta_2 h_0(h_0 - 1)}{\theta_1^2})^{1/h_0}$$ (17)

In the equation $\theta_i = \sum_{j=A+1}^m \lambda_j^i (i = 1,2,3)$, $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$, c_α is the threshold value under the testing level α of normal distribution.When there's no fault data in the sample or the process normally functions, the two statistical variables satisfy the following inequalities.

$$T_{aj}^2 \le T_{A,n_1,\alpha}^2$$ (18)

$$Q_{aj} \le \delta_\alpha^2$$ (19)

the second data matrix X_b can also be processed to derive the control inequalities:

$$T_{bj}^2 \le T_{A,n_2,\alpha}'^2$$ (20)

$$Q_{bj} \le \delta_\alpha'^2$$ (21)

iii.   The monitoring results for the experimental data

Three variables are considered and some of the data are modified on the original data. The no. 167-172 and no.350-355 samples show the transition procedure of two states. The no.360-365 samples are modified data which deviate from the normal constraint. Figure 7 shows the changing curves of the three variables and the scatter diagram in 3-dimention space.

Two experiments are designed to testify the advantage of differentiating states applied in this work. Figure 8 is the case no differentiating states, whilst figure 9 is the differentiating case. As the figures shown, the approach of differentiating states and determining the control limits respectively is more sensitive to the transition of states and fault data.

iv.   Locate and isolate the fault data

In the case of MSPC, the fault data can be located and isolated by contribution diagram. The variable contribute to SPE and T_i^2 can be written as (22)(23). ξ_i is the ith column of unit matrix. The fault data is judged as the variable with larger contribution to SPE and T^2.

$$Cont_i^{SPE} = (\xi_i^T(I - PP^T)x)^2$$ (22)

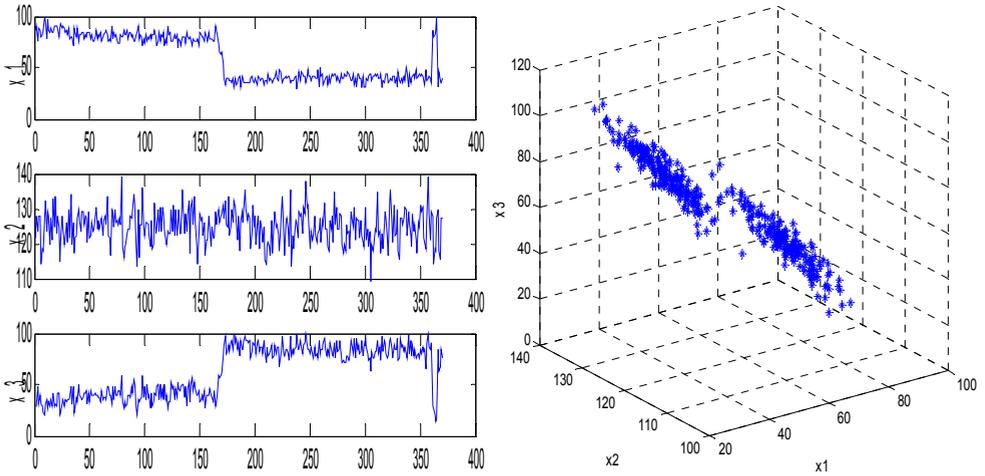$$Cont_i^{T2} = x^T P^T \Lambda^{-1} P \xi_i \xi_i^T x \tag{23}$$

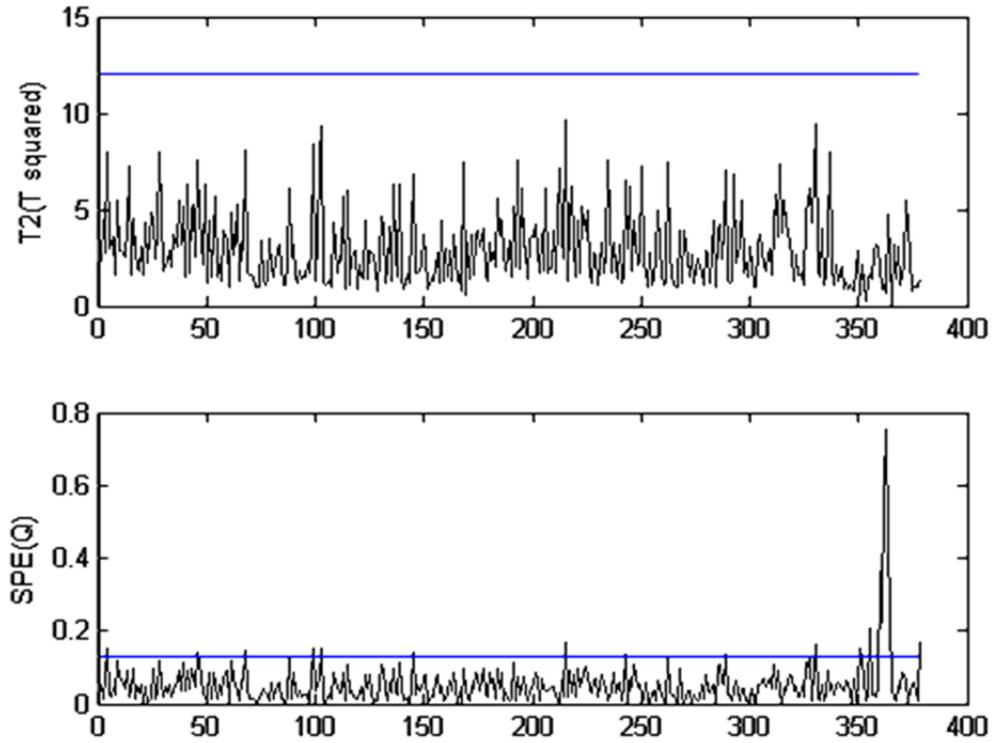**Figure 7.** The changing curves and scatter diagram of three variables

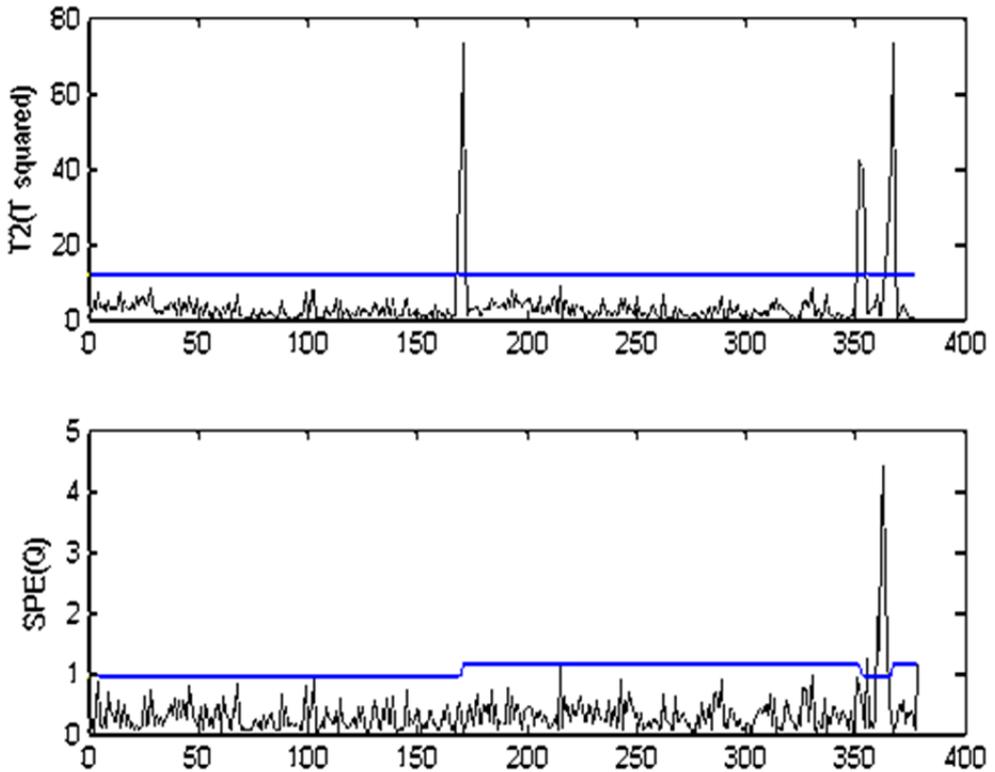**Figure 8.** The result of monitoring no differentiating states

Figure 9. The result of monitoring no differentiating states

2.3. The steam pipe network modeling and calculating the fault data based on the model

After the fault data are detected, the next step is to estimate the true values of the variables. As the experience shown, the measured temperature and pressure values are usually accurate. So the object of the work is to calculate the flow rate value with the detected temperatures and pressures according to the steam network model.

2.3.1. The static model of steam pipe network

To deduce the model, the real conditions are simplified as: 1) The steam in the pipes is axial-direction one dimensional flow; 2) The network is composed of nodes and branches(pipes). 3) There is no condensate or secondary steam and the effect is neglected.

i.   The hydraulic and thermal model of single pipe.
   -   The hydraulic model of single pipe

      By the law of momentum conservation the hydraulic model can be written as [8]:

$$P_1^2 - P_2^2 = 1.25 \times 10^8 \frac{\lambda q^2 P_1 T_2 Z_2 L}{D^5 \rho_m T_1 Z_1} \tag{24}$$

P_1, P_2 — the input and output pressure of the steam in the pipe, Mpa;

T_1, T_2 — the input and output absolute temperature, K;

Z_1, Z_2 — compressibility factor of the input and output steam;

λ — frictional factor;

q — the mass flow rate, t/h;

L — the length of the pipe, m;

D — the inner diameter of pipe, mm;

ρ_m — the weighted mean density of steam, kg/m3.

Usually, there's little difference between the products of compressibility factors and temperatures in the same pipe. Considering the elbows, reducer extenders and other friction factors, the equivalent coefficient η is added to the equation. Equation (24) changes to (25).

$$P_1^2 - P_2^2 = 1.25 \times 10^8 \frac{\lambda q^2 P_1 (1+\eta) L}{D^5 \rho_m} \tag{25}$$

Thus:

$$q = \frac{D^5 \rho_m}{1.25 \times 10^8 \lambda q P_1 (1+\eta) L} P_1^2 - P_2^2 = C_P (P_1^2 - P_2^2) \tag{26}$$

$$C_P = \frac{D^5 \rho_m}{1.25 \times 10^8 \lambda q P_1 (1+\eta) L} \tag{27}$$

In accordance with Альтшуль equation, frictional factor λ :

$$\lambda = 0.11 \times (\frac{\Delta}{D} + \frac{68}{\mathrm{Re}})^{0.25} \tag{28}$$

Δ — the equivalent absolute roughness, mm; Re — the Reynolds number of the steam.

After transferring the units of D, q to mm and t/h (ton per hour), Reynolds number is :

$$\mathrm{Re} = \frac{D u \rho_m}{\mu} = 354 \frac{q}{D\mu} \tag{29}$$

In the equation, u — the characteristic velocity of the steam in this pipe, m/s; μ — the mean dynamic viscosity coefficient. It's determined by equation (33)[9].

Denote ρ_1, ρ_2 as the input and output steam densities. $\rho_1, \rho_2, \mu, \rho_m$ can be calculated with following equations (31),(32),(33). The unclaimed symbols were defined in reference[9].

$$\rho_1 = \frac{P_1}{\pi(r_\pi^0 + r_\pi^r)RT_1} \tag{30}$$

$$\rho_2 = \frac{P_2}{\pi(r_\pi^0 + r_\pi^r)RT_2} \tag{31}$$

$$\rho_m = \frac{\rho_1}{3} + \frac{2\rho_2}{3} \tag{32}$$

$$\mu = \Psi(\delta,\tau)\eta^* \tag{33}$$

- The thermal model of single pipe

By the law of energy conversation, the static thermal model can be written as follows:

$$T_1 - T_2 = \frac{(1+\beta)q_l L}{1000 c_p q \times \frac{1000}{3600}} = \frac{(1+\beta)q_l L}{278 c_p q} \tag{34}$$

$$q = \frac{278 c_p q^2}{(1+\beta)q_l L}(T_1 - T_2) = C_T(T_1 - T_2) \tag{35}$$

$$C_T = \frac{278 c_p q^2}{(1+\beta)q_l L} \tag{36}$$

The definitions and units of T_1, T_2, L, q are the same as eqation (24).(note: changing equation (34) to (35) is to make thermal model has the same pattern with equation (26)).

β —the appending heat loss coefficient for the appendix of pipe, valve, and support, the value can be 0.15-0.25 depending on pipe laying methods;
c_p —specific heat capacity at constant pressure, kJ/kg.K;
q_l —the amount of heat loss along unit pipe length, W/m.

In term of equation IF-97, The unclaimed symbols are defined in the reference[9].

$$c_p = -R\tau^2(\gamma_{\tau\tau}^0 + \gamma_{\tau\tau}^\gamma) \tag{37}$$

The amount of heat loss along unit pipe length can be calculated using equation (38)

$$q_l = \frac{T_0 - T_a}{\frac{1}{2\pi\varepsilon}\ln\frac{D_o}{D_i} + \frac{1}{\pi D_0 a_w}} \qquad (38)$$

T_0, T_a —the out surface temperature of the pipe and the environment temperature, K;
D_0, D_i —the outer and inner diameters of the steam pipe, mm;
ε —the coefficient of thermal conductivity of the heat insulated material, W/m.K;

In equation (38), a_w is determined by:

$$a_w = 11.6 + 7\sqrt{v} \qquad (39)$$

v —the velocity of environmental air, m/s;

ii. Steam pipe network synthetic model

- Incidence matrix of pipe network

Draw out pipe network map, number the nodes in the order of steam sources, users, three-way nodes, and number the pipes in the order of leaf pipes, branch pipes. Suppose there are m_1 intermediate nodes and m_2 outer nodes, there are total $m = m_1 + m_2$ nodes and $p = m - 1$ pipes. Set the element of the ith row and the jth column in $A \in R^{m \times p}$ as a_{ij}, it's defined as:

$$a_{ij} = \begin{cases} 1 \ (flow \ in \ from \ node \ i) \\ -1(flow \ out \ from \ node \ i) \\ 0 \ (unrelated) \end{cases} \qquad (40)$$

- The flow rate balance equation of the pipe network

Denote: $P = (P_1^2, P_2^2, ..., P_m^2)^T$, $P_i^2, (i = 1, ..., m)$ the square of the ith node's pressure, MPa²;

$T = (T_1, T_2, ..., T_m)^T$, $T_i, (i = 1, ..., m)$ the ith node temperature, °C;

$\vec{q} = (q_1, q_2, ..., q_p)^T$, $q_j, (j = 1, ..., p)$ the flow rate of the jth pipe, t/h;

$Q = (Q_1, Q_2, ..., Q_m)^T$, $Q_i, (i = 1, ..., m)$ the flow rate of the ith node, unit t/h; For steam souce the value is minus q_j, and for the user the value is positive q_j, otherwise, the value is 0.

$C_p^* = diag(C_{P1}, C_{P2}, ..., C_{Pp}), C_{Pj}, (j = 1, ..., p)$ parameters are determined by equation (27);

$C_T^* = diag(C_{T1}, C_{T2}, ..., C_{Tp}), C_{Tj}, (j = 1, ..., p)$ the parameter are determined by equation(36).

In term of mass conservation law, the total flow rate of each node should be 0, that is:

$$A\vec{q} + Q = 0 \qquad (41)$$

According to the hydraulic and thermal equations, for the pipe network equations:

$$\vec{q} = C_P^* A^T P = C_T^* A^T T \qquad (42)$$

Substitute (42) into (41):

$$AC_P^* A^T P + Q = 0 \qquad (43)$$

$$AC_T^* A^T T + Q = 0 \qquad (44)$$

Equations (26), (27), (35), (36), (42), (43), (44) comprise the static model.

2.3.2. Hydraulic and thermal calculation based on searching

Industrial networks are equipped with temperature, pressure and flow meters except intermediate nodes. The proposed algorithm tries to calculate the flow rate of each pipe with the given condition. The flow rate meter readings are applied to evaluate of the algorithm.

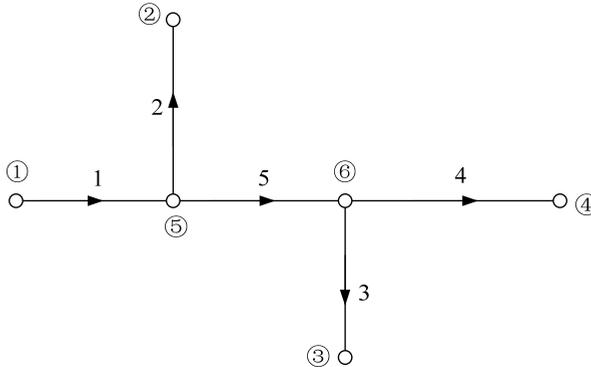

Figure 10. Layout of steam pipe network

Set figure 10 as an example. the index of the nodes and pipes are shown in the figure:

Step 1.   Determine the elements of matrices A, Q as the preliminarily defined.

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & -1 \end{pmatrix} \qquad (45)$$

$$Q = (-q_1, q_2, q_3, q_4, 0, 0)^T \tag{46}$$

As the previous definitions for equation (41)(42), $P_5, P_6, T_5, T_6, \vec{q}$ are unknown. From equation (27), it's easy to discover friction factor and density are relative with P_5, P_6, T_5, T_6. Only by presetting $\vec{q}$ can equations (42) (43) be applied for hydraulic iterative calculation. However, in thermal calculation only c_p relates with P_5, P_6, T_5, T_6. Changing equation (34) to (35), $\vec{q}$ can be calculated directly.

$$q_i = \frac{(1+\beta)q_{li}L_i}{278c_p \Delta T_i}, (i = 1, 2, ..., 5) \tag{47}$$

In equation (47), i —the pipe index, ΔT_i —the temperature difference of input and output.

Step 2. Thermal calculation

It can be inferred by the flow direction of steam that:

$$P_2 < P_5 < P_1, \max(P_3, P_4) < P_6 < P_5 \tag{48}$$

$$T_2 < T_5 < T_1, \max(T_3, T_4) < T_6 < T_5 \tag{49}$$

Presume:

$$P_5 = \frac{1}{2}(P_2 + P_1); P_6 = \frac{1}{2}(P_5 + \max(P_3, P_4)) \tag{50}$$

Set the searching start value for T_5, T_6 as T_1 and $\max(T_3, T_4)$ respectively, and set the searching step size as h then begin searching T_5, T_6 in the range (49). For each searching step size, calculate c_p (equation (37)) and flow rate of each pipe (equation (49),(50)), and test whether $\|A\vec{q} + Q\| \leq \xi_1$ is satisfied. If it is not satisfied, update T_5, T_6 with one step to continue calculation, or retain the temperature value and the corresponding flow rates.

Step 3. Hydraulic calculation

Just as the thermal calculation, first presume the temperature of the intermediate nodes as the value searched by the former step, and set the start point and step size for the two intermediate pressures. Then begin searching in the region determined by inequality (48) with the step size to calculate the density (equation (32)), frictional factor ((28), (29), (33)), and flow rate ((26), (42)) of each pipe. When calculating the flow rate, presetting the initial values as the result of step 2 will reduce the iterative calculation time. Test whether $\|A\vec{q} + Q\| \leq \xi_2$ is satisfied, If the inequality is not satisfied, update P_5, P_6 with one step size to continue calculation, or retain the pressure values and the corresponding flow rates.

Step 4. Thermal model Verification

Substitute the intermediate nodes' temperature (determined by step 2), pressure and flow rate values (determined by step3) into thermal model to verify the inequality

$\left\| AC_T^* A^T T + Q \right\| \leq \xi_3$. If it is satisfied, the calculation ends with the result of step 3. If not, presume the intermediate pressure values as the result of step 3, reduce step size and go to step 2 and start another calculation cycle.

In the algorithm, the value ξ_1, ξ_2, ξ_3 can be set as 0.05-0.3 depending on the demand for errors and calculation stability. The calculation flow chart is depicted with figure 11.

2.3.3. Comparisons of calculation results and measurements

The specifications of each pipe are shown in table 1. The parameters in the model and algorithm are initially set as: $\eta = 0.2$, $\Delta = 0.2$, $\beta = 0.15$, $\varepsilon = 0.035$, $\xi_{1,2,3} = 0.3$.

Substitute the specification data in Table 1 and temperature and pressure data in Table 2 to the algorithm calculating flow rate values.The comparision results are listed in Table 2. It shows the largest difference is less than 6%. Actually, the neglected factors in model, the parameter errors and measurement errors add to the difference.

Pipe No.	Length(m)	Inner Diameter (mm)	Outer Diameter (mm)	Heat Insulation Layer Thickness (mm)
1	600	700	720	110
2	2300	600	630	105
3	700	400	426	95
4	1300	500	529	100
5	1200	600	630	105

Table 1. The Specifications of the Pipes

Outer node No.	Pressure(MPa)	Temperature (°C)	Measured q(t/h)	Calculated q (t/h)	the relative difference
1	0.705	286	117	121	3.4%
2	0.679	195	19	20	5.2%
3	0.563	246	19	19	0
4	0.445	256	79	82	3.8%

Table 2. The Comparisons of the Measured Data and the Calculated Data.

Configuration of pipe network,
specification data

Number each pipe and node,
Determine A,Q

Analyze the intervals of intermediate
nodes 'temperature and pressure value,
set step size for searching

Presume the intermediate pressure value,
set the start points of searching for
intermediate temperature values

Thermal calculation:
Calculate (14),(15),(24),(19) respectively

Update the
temperature values
with a step size

$$\left\| A\vec{q} + Q \right\| \leq \xi_1$$ N

Reduce step size

Presume the intermediate nodes
'temperature value as the searched
result, set the start points of searching
for intermediate pressure values.

Hydraulic calculation:
Calculate equation (3) to (10),(19). For each
flow rate q, the iterative initial value set as the
result of thermal calculation

Update the
pressure values
with a step size

$$\left\| A\vec{q} + Q \right\| \leq \xi_2$$ N

Substitute the searching result:
intermediate nodes 'temperature,
pressure and flow rate values into
thermal model

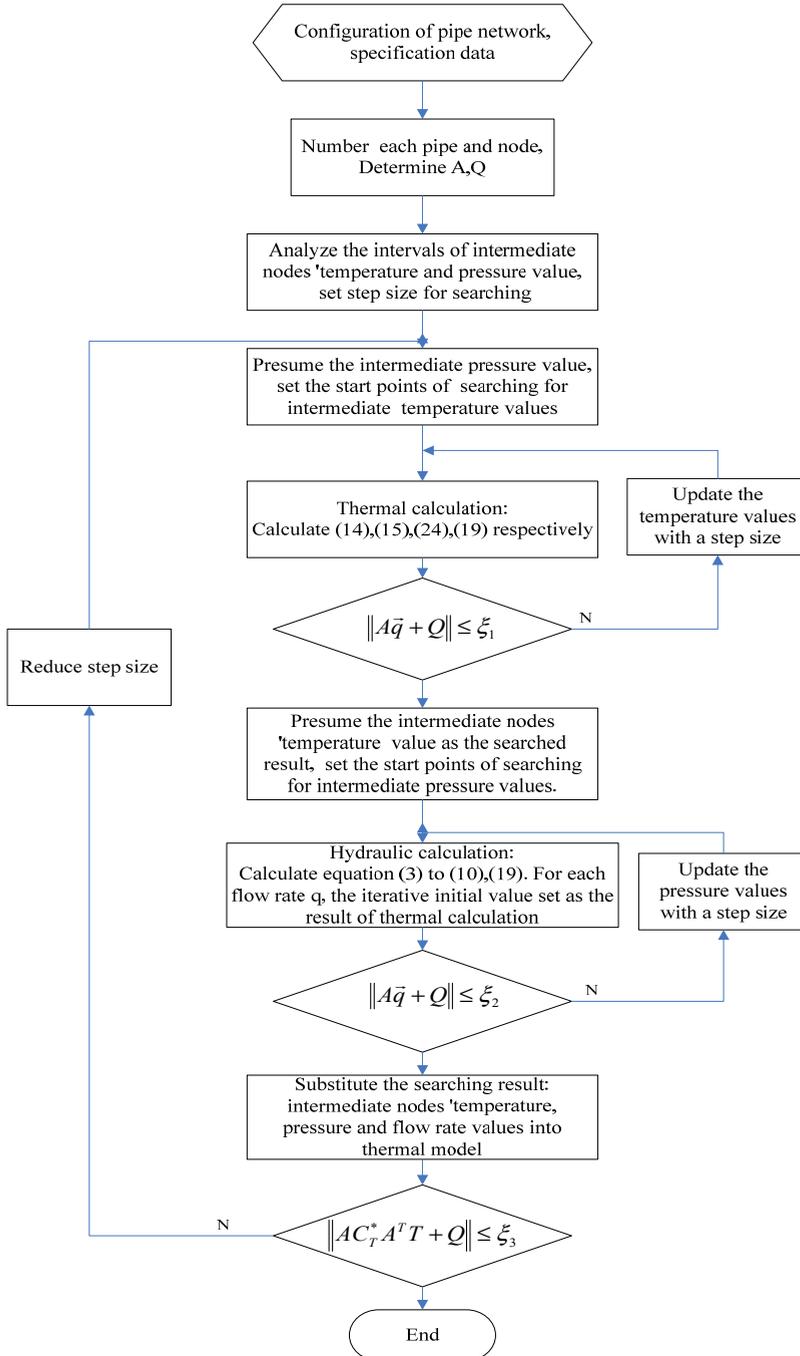N $$\left\| AC_T^* A^T T + Q \right\| \leq \xi_3$$

End

Figure 11. The flow chat of flow rate calculation

The results show the validation of the model and effectiveness of the algorithm, and the proposed model and algorithm can be applied to simulate the running of static steam pipe network, reconstruct the discovered fault data of the mass flow rates.

As for the larger scale steam network with more than 3 intermediate nodes, it's difficult to apply the algorithm directly. However, the pipe network can be divided into several smaller networks from the nodes with known temperature and pressure.

3. Gross errors detection

3.1. Problem definition

A section of the steam network named "S2" for an iron &steel plant in China is shown in figure 12. In the figure, N1-N7 represent the different production processes. The arrows point to the direction of steam flow, the variables Xi or Xij on behalf of the real steam mass flow. The electric valves are remotely controlled by the operators. Many industrial steam systems are similar in structure to this system, but the scale is much larger.
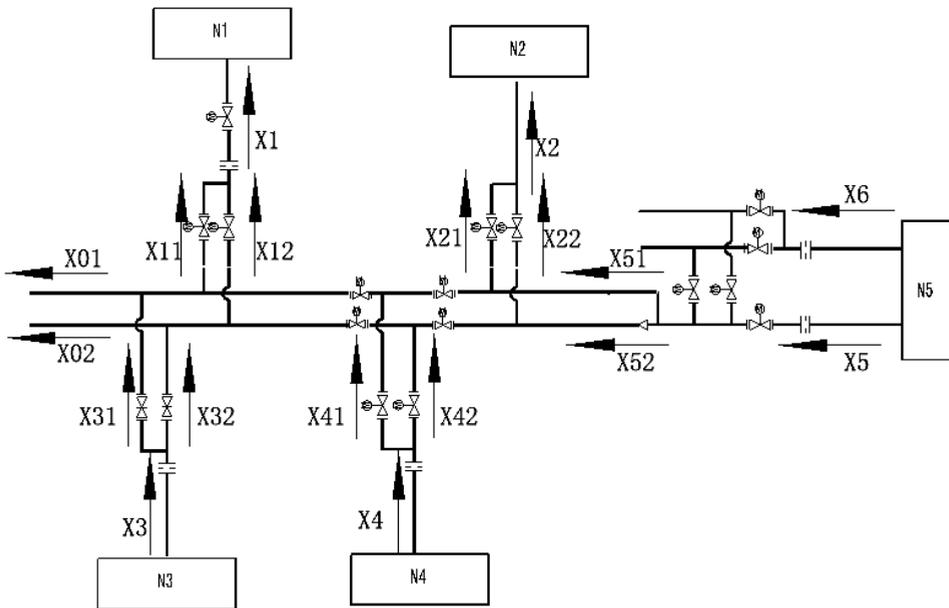

Figure 12. A section diagram of steam network

If all of the variables above have been measured, suppose that the electric valves are fixed at a certain position, and the pipeline leakage and the amount of condensate water can be neglected, the constraint equations can be written out on the basis of the mass balance.

$$\begin{cases} X1 - X11 - X12 = 0 \\ X2 - X21 - X22 = 0 \\ X3 - X31 - X32 = 0 \\ X4 - X41 - X42 = 0 \\ X5 + X6 - X51 - X52 = 0 \end{cases} \tag{51}$$

The overall balance equation can also be written out as:

$$X5 + X6 + X4 + X3 - X1 - X2 - X01 - X02 = 0 \tag{52}$$

If we remark

$$X = (X01, X02, X1, X11, X12, X2, X21, X22, X3, X31, X32, X4, X41, X42, X5, X6)^T = (X_1, X_2, ..., X_{18})^T \tag{53}$$

as the vector of mass flow rates in the steam network.(51) (52) can be abbreviated as:

$$AX = 0 \tag{54}$$

In (54), A is the incidence matrix. The matrix is composed by the elements of 1, -1, and 0.

If Y represents the vector of measured flow rates and X is the vector of true flow rates , then:

$$Y = X + W + \varepsilon \tag{55}$$

In the equation, W is the vector of the measurement gross errors (or systematic errors), and ε is the vector of random measurement errors with each element being normally distributed with zero means and known covariance matrix, Q. The approaches to detect and remove the gross error from the measurements are to be discussed in this section.

When there is no gross error in the measurement vector, finding a set of adjustments to the measured flow rates to satisfy equation (54) is the problem of data reconciliation. Denote the adjustment vector as a , and the adjusted flow rate vector as $\hat{X}$, we get:

$$\hat{X} = Y + a \tag{56}$$

Applying Least squares method, it can be stated as the constrained least-squares problems:

$$\min a^T Q^{-1} a \text{ subject to } A\hat{X} = 0$$

Q is the weighed parameter matrix. Usually, Q is selected as:

$$Q = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_{18}^2) \tag{57}$$

$\sigma_i (1 \le i \le 18)$ is the deviation of the measurement variable $X_i (1 \le i \le 18)$.

The solution $\hat{X}^*$ to the problem can be obtained by Largrange multipliers[10]

$$\hat{X}^* = Y - QA^T (AQA^T)^{-1} AY \tag{58}$$

The vector of residuals e

$$e = Y - \hat{X}^* = QA^T(AQA^T)^{-1}AY \tag{59}$$

By gross error detection, the systematic errors in the measurements can be removed, and by data reconciliation the random errors can be reduced. However, if the gross errors are not removed, the gross errors in some variables will propagate to other accurately measured variables in the procedure of data reconciliation. So gross error detection had better be adopted in advance.There are two basic types of gross error detection methods, they are based on measurement test (MT) and nodal imbalance test (NT).

3.2. Gross error detection based on test of residuals

Detection based on test of residuals belongs to the statistical test and has been termed and evaluated [11]. Measurement test[12] is the basic algorithm.

Step 1.   Apply the least-squares routine by using equation (58) (59) to compute $\hat{X}^*$ and e.

Step 2.   Compute the variable for each pipe (or stream)

$$z_j = e_j / \sqrt{v_{jj}} \tag{60}$$

In the equation,

$$V = QA^T(AQA^T)AQ \tag{61}$$

On the hypothesis that the measured value in the jth stream doesn't contain gross error, z_j follows standard normal distribution.

Step 3.   Compare z_j with a critical test value z_c. If $|z_j| > z_c$, denote stream j as a bad stream. z_c recommended[12] $z_c = z_{1-\beta/2}$, the $1 - \beta / 2$ point of the standard normal distribution.

$$\beta = 1 - (1 - \alpha)^{1/n} \tag{62}$$

Where n is the number of measurements tested (currently, n=18), α (the recommended value is 0.05) is the overall probability of a type I error for all tests, and β is the probability of a type I error for each individual test. Denote by S the set of bad streams found by the above procedure. The measurement $y_j, j \in S$ is considered to contain gross error.

Step 4.   If S is empty, proceed to step 7. Otherwise, remove the streams contained in S and aggregate the nodes connected with the stream. This process yields a system of lower dimension with compressed incidence matrix A', measurement vector Y', and weighed matrix Q'. Denote T as the set of streams with the measurement data in Y'.

Step 5. Replace A, Y, and Q with A', Y', and Q' respectively and Compute the least-squares estimates in T applying equation (58).

Step 6. Solve equation (54) to Compute the rectified values of the streams in S by substituting the estimated values computed in step 5 with the data of the steams in T. The original measured data are used for the streams in the set $R = U - (S \cup T)$, where U is the all stream set.

Step 7. The vector comprised of the results from steps 5, 6 and the original measured data for the streams in R is the rectified measurement vector. If S is empty, then $\hat{Y} = \hat{X}^*$, the rectification of the measured data is completed in step 1.

Notations to the algorithm:

1. In step 4, it's possible to remove the good streams into S in some cases.

2. The equation (62) and critical level value provides a conservative test since the residuals are generally not independent. It's not always applied [13].

3. The different significance levels will induce different results and effects.

4. For the shortage of MT that it tends to spread the gross errors over all the measurement and obtain unreasonable (negative or absurd) reconciled data, The Iterative Measurement Test (IMT) method and Modified IMT (MIMT) method were proposed. However, the main frames are the same with MT.

3.3. Gross error detection based on nodal imbalance test

The algorithms of gross error detection based on statistical test of nodal imbalance are mainly based on the work of literature [14]. Applying nodal imbalance test to each node and the aggregation node (ie. pseudonode) to locate and remove the gross errors. The basic algorithm named Method of pseudo nodes (MP) is listed as:

Step 1. Compute the nodal imbalances vector r and the statistical testing variable vector z. On the assumption that no systematic error exits, the defined variable z_i are standard normal distributed.

$$r = AY \tag{63}$$

$$z_i = r_i / \sqrt{g_{ii}} \tag{64}$$

Where

$$G = AQA^T \tag{65}$$

Step 2. Compare each z_i with a critical test value $z_c = z_{1-\alpha/2}$, which corresponding to the point of the significance testing level $(1 - \alpha/2)$. For the instance a test at the 95% significance level, $\alpha = 0.05$ and $z_c = 1.96$. If $|z_i| \le z_c$ is satisfied, the ith node is regarded as a good node and denote all streams connected to the node i as good streams.

Step 3.   If no bad nodes are detected in step 2, proceed to step5. Otherwise, repeat steps 1 ,2 (changing the matrices and vectors accordingly) for pseudonodes containing 2, 3… m nodes.

Step 4.   Denote the set of all streams not denoted as good in the previous steps by S. The measurements $y_i, i \in S$ are considered containing gross errors.

Steps 5-8. The procedure is the same as the steps 4-7 of the MT algorithm.

Notations to the algorithm:

1.   The principle assumption is that the errors in two or more measurements do not cancel.

2.   In step 3, m is chosen by the effect of locating the gross errors. If the increasing of m doesn't gain any improvement, the step can be stopped.

3.   By applying graph-theoretical rules[14] some streams can be determined as bad measured streams. The additional identification may be useful for the procedure of MP.

4.   Equation (62) is not applied in the algorithm to control type I error. The probability of a type I error in the nodal imbalance test is not necessarily equal to the probability of rejecting a good measurement in MT [15].

5.   If the set S obtained in step 4 is empty, but there is one or more nodes are truly bad. Usually the instances happen when there's leak in the network system or the measured errors canceling each other. To solve the problem, the Modified MP (MMP), MT-NT combined methods [16] are proposed.

As the above two methods shown, gross data detection and data reconciliation are inherently combined together. Gross errors detection is proceeded with the help of the least-square routine, which actually can obtained the optimal estimates.

4. Data reconciliation

4.1. The basic problems of data reconciliation

Just as mentioned in section 3, the data reconciliation problem is expressed as the solution of the constrained least-squares problems.However, the assumptions, the constraint equations, and the weighted parameter matrix have to be discussed.

4.1.1. On the assumptions of data reconciliation in the present application

For the application of the steam network, there are four latent assumptions. They are:

1.   The process is at the steady state or approximately steady state.
     Suppose that the electric valves are fixed at a certain position, and the mass flow rate in each steam has been close to a constant for a period of time. So the constraint equation (54) can be written out on the basis of the balance for all the nodes (including real nodes and pseudonode) and the solution of the problem can be in accordance with the actual state.

2. The measurement data are serial uncorrelated. The assumption makes it easier to estimate the deviations of the measurements. Although this is usually not the real case, the serial data can be processed to reduce the confliction [17].

3. The Gross Errors have been detected and removed. If there's gross error in the measurement data, the procedure of data reconciliation will propagate the errors.

4. The constrained equations are linear. The style of equation (54) is linear; however the true constrained equation will be related with many other environmental variables and obviously not linear. So the assumption is to be approximately satisfied.
   Only these four assumptions are nearly satisfied, can the solution of the problem (equation (58)) be closer to the true value.

4.1.2. On the constraint equation

Need to note equations (51) and (52) are based on the assumptions of steady state and linear constraints, no pipeline leakage or condensate water loss. But it cannot avoid the condition of the pipeline leakage and condensate water. So the equations (51)(52)should be written as:

$$\begin{cases} X1 - X11 - X12 - \delta1 = 0 \\ X2 - X21 - X22 - \delta2 = 0 \\ X3 - X31 - X32 - \delta3 = 0 \\ X4 - X41 - X42 - \delta4 = 0 \\ X5 + X6 - X51 - X52 - \delta5 = 0 \end{cases} \tag{66}$$

$$X5 + X6 + X4 + X3 - X1 - X2 - X01 - X02 - \delta0 = 0 \tag{67}$$

Here:

$$\delta i = f_i(\tau, D_i, l_i, T_i, P_i, X_i) \qquad (i = 0, 1, \ldots, 5) \tag{68}$$

It represents the vector for the condensate water and leakage loss amount of each constraint equation. Each element has relations with the environmental temperature τ, pipe diameter D, pipe length l, steam temperature T, pressure P and the flow rates of main pipes.

It's difficult to set up the mathematical models of these loss amounts. However, by the first order of Taylor Expansion at the point of ($\tau_0, D_{i0}, l_{i0}, T_{i0}, P_{i0}, X_{i0}$):

$$\delta i = f_i(\tau_0, D_{i0}, l_{i0}, T_{i0}, P_{i0}, X_{i0}) + \frac{\partial f_i}{\partial \tau_0}(\tau - \tau_0) + \frac{\partial f_i}{\partial T_0}(T - T_{i0}) + \frac{\partial f_i}{\partial P_{i0}}(P_i - P_{i0}) + \frac{\partial f_i}{\partial X_{i0}}(X_i - X_{i0})$$
$$= \delta_{i0} + k_{i\tau}(\tau - \tau_0) + k_{iT}(T - T_0) + k_{iP}(P_i - P_{i0}) + k_{ix}(X_i - X_{i0}) \tag{69}$$

The constants in equation (69) can be determined by the method of multiple linear regression with a certain number of history data. As the constraint equation changed into:

$$AX = \delta \tag{70}$$

The solution to the least-squares problem of reconciliation is:

$$\hat{X}^* = Y - QA^T(AQA^T)^{-1}(AY - \delta) \tag{71}$$

4.1.3. On selection of the weighted parameter matrix Q

The selection of Q directly influences the result of the data reconciliation. In theory Q is recommended as the equation (57). However, the deviations are usually unknown or may vary as the time of instruments being used getting long. The deviation of each measurement can be estimated by the standard deviation of the sample data.

$$\sigma_i \approx s_i = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(X_{ij} - \overline{X}_i)^2} , (1 \le i \le 8) \tag{72}$$

Where, X_{ij} is the j th sample of the variable X_i , and $\overline{X}_i$ is the average of X_i .

If a small number of high-precision instrument are applied,, the corresponding elements in the matrix of Q with smaller value, then the quality of data will be greatly improved. Though some literatures [18] [19] recommended the methods to determine or adjust the matrix Q, the methods or theories for general application is not available.

4.1.4. Simulation results

The simulation results of data reconciliation for the present application are shown in table 3. The measured data is presumed according to (73)

$$Y = X + 10\% \times X. \times Rand(18,1) \tag{73}$$

The standard deviations are set to be 0.5 percent of the true values. In the simulation, the pipe network loss is not considered. As shown in table 3, most of the rectified data are much closer to the true values. The simulation testifies the efficiency of data reconciliation.

	X01	X02	X1	X11	X12	X2	X21	X22	X3
true value	5	5	50	25	25	60	30	30	30
measured value	5.347	5.159	54.751	25.086	26.097	62.289	32.297	32.386	30.561
variance	0.0006	0.0006	0.0625	0.0156	0.0156	0.0900	0.0225	0.0225	0.0225
rectified value	5.347	5.150	52.228	25.090	25.514	62.178	31.946	31.730	31.174
	X31	X32	X4	X41	X42	X5	X51	X52	X6
true value	15	15	40	20	20	20	25	25	30
measured value	15.735	15.668	42.585	21.419	21.509	20.552	26.699	26.638	32.488

	X01	X02	X1	X11	X12	X2	X21	X22	X3
variance	0.0056	0.0056	0.0400	0.0100	0.0100	0.0100	0.0156	0.0156	0.0225
rectified value	15.582	15.592	42.306	21.340	21.566	20.946	26.086	26.236	31.375

Table 3. The Result of Data Rectification

5. Conclusion

In the chapter, three data processing approaches to improve data quality are demonstrated. For the importance of properly controlling the steam system performance normally, The data obtained from EMS should be accurate and reliable. However, the data may be influenced many outer factors. The approaches proposed in the chapter are to detect the fault data, locate and remove the gross errors and reduce the random errors.

Four main reasons induce the low accuracy of the mass flow rate measurement. Combining the principle of "3σ" and empirical distribution function to determine control limit is proposed for single variable monitoring, and applying PCA to determine the control limits for the multivariate process. With the limits, most of fault data can be identified easily. For the fault data of flow rates, the approach to setup the mathematical model of the steam network and calculate the flow rates is proposed. The simulation and experimental results show the effectiveness of the approaches.

Two approaches, MT and MP, to detect the gross errors are demonstrated. Both are preceded by selecting the statistical variables, which follow standard normal distribution, and applying hypothesis test. Some notations for the two algorithms are stated.

The constrained least-squares problem applied for present application is discussed. The four assumptions are approximately satisfied when the steam network is normally function and the state is nearly static. The pipe network loss can be considered to add to the constraint equations for more accurate results. The weighed parameter matrix has influence on the results of data reconciliation. To estimate the deviations of the instruments online and apply several instruments with high precision will improve the quality of reconciled data.

Author details

Luo Xianxi, Yuan Mingzhe and Wang Hong
Key Laboratory of Industrial Informatics, Shenyang Institute of Automation,
Chinese Academy of Science, Shenyang, China

Luo Xianxi and Li Yuezhong
East China Institute of Technology, Nanchang, China

Luo Xianxi
Graduate School of the Chinese Academy of Sciences, Beijing, China

Yuan Mingzhe and Wang Hong
Shenyang Institute of Automation, Guangzhou, Chinese Academy of Science, China

Acknowledgement

6. References

[1] T. Tong, *The Analysis on the Flow Rate Measurement of Steam* Journal of Hebei Institute of Architectural Engineering, 2002. 6(20): p. 43-45.

[2] D. Zhenjian, On the Methods to Reduce the Steam Measurement Difference along the Transmission Line and the Existing Problems. China Metrology, 2003. 3: p. 67-68.

[3] Zhang J., Y.X.H., Multivariate Statistical Process Control. Peking: Chemical Industry Press (CIP). 2002.

[4] Chand, S.T.S., Estimating the limits for statistical process control charts: A direct method improving upon the bootstrap. European Journal of Operational Research, 2007. (178): p. 472-481.

[5] Kuo, H.C.W.-L., Comparisons of the Symmetric and Asymmetric Control Limits for X and R Charts. Computers & Industrial Engineering, 2010.（59）: p. 903-910.

[6] Chen, T., *On Reducing False Alarms in Multivariate Statistical Process Control.* Chemical Engineering Research and Design, 2010. (88): p. 430-436.

[7] Zhou Donghua, Li Gang, ed. Data Driven Fault Diagnosis Technology for Industrial Process--Based on Principal Component Analysis and Partial Least Squares. Peking: Scince Press. 2011.

[8] Zhang Zenggang, Research on Coupling Hydraulic & Thermal Calculation for Steam Pipe Network Theory & its Application(D). 2008: China University of Petroleum.

[9] W.Wagner, A.K., *Properties of Water and Steam.* 1998: Springer-Verlag Berlin Heiderberg.

[10] Kuhn, D.R., H.Davidson, *Computer Control II:Mathematics of Control.* Chem. Eng.Prog., 1961: p. 57.

[11] Iordache, C., R.S.H.Mah, Performance Studies of the Measurement Test for Detection of Gross Errors in Process Data. AIChE Journal, 1985. 31,1: p. 187.

[12] Mah, R.S.H., A.C.Tamhane, *Detection of Gross Errors in Process Data.* AIChE Journal, 1982. 28: p. 828.

[13] Crowe, C.M., Y.A.Garcia, *Reconciliation of Process Flow Rates By Matrix Projection.* AIChE Journal, 1983. 29: p. 881.

[14] Mah, R.S., G.M.Stanley and D.Downing, *Reconciliation and Rectification of Process Flow and Inventory Data.* IEC Proc. Des. Dev., 1976. 11: p. 197.

[15] Serth, R.W. and W.A. Heenan, Gross Error Detection and Data Reconciliation in Steam-metering Systems. AIChE Journal, 1986. 32(5): p. 733-742.

[16] Mei, C., H. Su, and J. Chu, *An NT-MT Combined Method for Gross Error Detection and Data Reconciliation.* Chinese Journal of Chemical Engineering, 2006. 14(5): p. 592-596.

[17] Kao.C.S, e., A General Prewhitening Procedure for Process and Measurement Noises. Chem. Eng.Prog., 1992. 118: p. 49.

[18] Wu Shengxi, Z.e., Estimation of Measurement Error Variances/Covariance in Data Reconciliation, in the 7th World Congress on Intelligent Control and Automation. 2008: Chongqing, China. p. 714-718.

[19] Narasimhan, S. and S.L. Shah, *Model identification and error covariance matrix estimation from noisy data using PCA.* Control Engineering Practice, 2008. 16(1): p. 146-155.