

## A Review on Data-Driven Modelling and Data Rectification for Steam Pipe Networks in Iron& Steel Enterprises

Xianxi Luo<sup>a</sup>, Wei Wang<sup>b</sup>, Hongwei Deng<sup>c</sup>, Zehao Wu<sup>d</sup>

Jiangxi Province Engineering Research Center of New Energy Technology and Equipment (East China University of Technology), Nanchang, Jiangxi, China

<sup>a</sup>Xianxi\_luo@yahoo.com, <sup>b</sup>543881900@qq.com, <sup>c</sup>1650427313@qq.com, <sup>d</sup>654852846@qq.com

**Keywords:** data driven modelling, fault detection, data rectification, steam pipe network, iron&steel enterprises.

**Abstract:** Steam pipe network is important facility in the iron& steel enterprises to recycling the waste heat energy and reduce the comprehensive energy consumption. In this paper, we analyze the necessity of data-driven modelling and data reconciliation for the schedule and operation of steam system through EMS. The related technologies, including the steam pipe network modelling, fault data detection, process monitoring, and data rectification are reviewed in detail. We also investigate the current application condition of data rectification in industry and present the potentials and the problems for the application in steam pipe networks. The probable solutions for different problems and research directions are suggested for practical application at the end.

### Introduction

The Chinese iron& steel industry developed dramatically in recent years and the crude steel production has exceeded 700 million tons per year since 2012, which is undoubtedly the largest one in the world [1]. However, the average comprehensive energy consumption per ton of steel in China is much higher than that of technology-advanced countries. Collecting and reusing the waste heat energy with steam facilities are the main and effective methods to reduce the energy consumption of the production. It is estimated that the consumption of steam account to more than 10 percent of the overall energy consumption in iron& steel plants. About 7 of these percentiles are from the recovery of the waste heat energy produced in the process of iron& steel production. Because steam pipe networks extend to every production section of iron& steel plants, their geographical distribution are wide, their scale are large, it's quite a challenge to monitor the states of so many steam facilities and keep optimal performances.

As the technological progress of industrialization and informatization integration[2], iron& steel enterprises began to setup Energy Management System (EMS). EMS interacts with many instruments, controllers, monitoring devices and servers through multilevel network, and remotely monitors and manages[3] the energy media, forming a typical networked control system. The steam facilities and pipe network are important objects of the network control system. The enterprises apply the EMS platform to monitor the dynamic variation and schedule the production process to use the steam efficiently.

The schedule and operation of the steam system through EMS platform must base on the data with high quality. The quality of data refers to the accuracy, completeness, consistency and timeliness[4] of the data. However, the actual quality of the original data is affected by many factors. Firstly, the amount of data representing the temperature, pressure and flow rate of the detecting point on steam pipe network are tremendous, and data delay and packet loss are inevitable. Secondly, there are many links from sensors and transmitters through data acquisition units and network devices to EMS platform. Once any of the link fails, real-time measurement information cannot be uploaded. In addition, the complexity of steam properties affects accuracies of the instruments. These situations could influence the performances of the control systems and result in steam emission, even lead to dangerous production problems. Hardware redundancy can be helpful to improve data quality, but it means a higher investment and more difficulties in maintenance and repair.

Through establishing the mathematical models to describe the behavior of steam transmission in the pipe network, the detected data are turned to be inter-connected rather than to be independent. With the models, the occurrence of fault data could be discovered, or even be rectified with specific algorithms (Gross Errors detection and Data Reconciliation). The synthetic technology is known as data rectification. The technology applied in the steam pipe networks is supposed to monitor the state of the networks, detect fault data and try to correct the errors. Fig.1[5] shows the main idea and motivation of such researches. The method of modelling the steam network, monitoring the behavior of the steam and discovering the abnormalities are targeting effectively improving the accuracy of measurement, the completeness, consistency and timeliness of the data which represent the necessary conditions for remotely scheduling and operation of the steam pipe networks. These researches are meaningful for the iron & steel enterprise to utilize the energy resources efficiently and reduce the comprehensive energy consumption per ton of iron & steel enterprises.

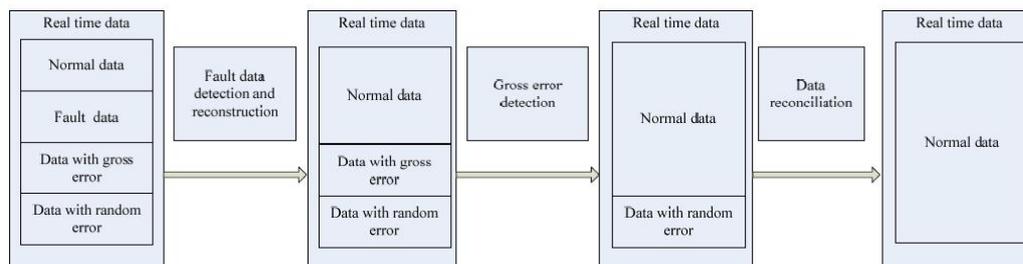


Fig. 1. The motivation of the research on data-driven modelling and data rectification

This paper presents the progress of the researches on modelling of steam networks, the application of fault data detection and process monitoring, data rectification. The potential research problems and directions are also discussed. The paper is organized as follows. The researches on modelling are described in Section 2. The methods of fault detection are presented in Section 3. The methods of data rectification are presented in Section 4. Based on the state of the art, the actual status of the application and the discussion on the future research are presented in Section 5 and Section 6, followed by the final remarks and conclusions.

### Modeling of Steam Pipe Network

W. Wagner published IAPWS-IF97, additional equations of characteristics and data table, which improved the speed and accuracy of steam thermodynamic calculation. The IF-97 equations have become the foundation of steam pipe network modeling and measurement.

In the modeling of steam pipe network, the research results are very rare. Sylvie Lorente and Wisanuruk Wechsathol compared the tree, loop and square type steam pipe networks. They concluded that the best structure is the tree type pipe network [6], but considering the safety of the network, loop type pipe network structures are widely applied. Reference[7] applies mass, momentum and energy conservation principles to construct the relationship of the variables of temperature, pressure and flowrate of the steam in the pipe. Dynamic hydraulic and thermodynamic calculation model of steam pipe network was first setup by[8]. These are useful explorations for the mechanism modeling of steam pipe network, but they did not consider the problems of the state change of the pipe network and the condensate water which induce deviation between the model and the actual condition.

As for the model solution, because the hydraulic and thermal dynamic models of steam network are nonlinear, and many variables are involved, the problems are usually solved by iteration. Song [9]simplified the equations of dynamic pipe network, demonstrated the convergence of nonlinear iteration under certain conditions, but improper initial value may lead to computation failure. Reference [10] used the finite element method to preset the values for variables and iterative calculation, but did not demonstrate the convergence. Luo[11] researched the network with multiple steam sources and proposed the method to solve the flowrate distribution.

As far as we known, the proposed mathematical models of steam pipe network are mainly static models, which cannot reflect the transient of state variables when disturbances happen. These models cannot be applied for model validation and dynamic data rectification directly.

### **Fault data detection and process monitoring**

Statistical Process Control (SPC) is one of the data-driven technologies applied for fault data detection and process monitoring. The traditional statistical process control is mainly based on the single variable, and it is the key to determine the control limit of the statistical process. After years of development, researchers have proposed several types of control charts and the method to determine the control limits. At present, the focus is to determine the control limit in the non-normal distribution. For example, Monte Carlo proposed the Bootstrap method to directly determine the control limit based on empirical distribution control limit[12]. Kuo compared the problem of control limit in symmetric and asymmetric way[13]. It is concluded that the asymmetric control limit has strong robustness when the graph characteristic is offset. However, the selection of quantitative and the way to determine asymmetric limits were not clearly pointed out.

Multivariate Hotelling's  $T^2$  and squared prediction error SPE (Squared Prediction Error, SPE) are used as monitoring indicators, which is a unified multivariable monitoring index and effectively reduces operator's "Information Overload" pressure. The traditional PCA (Principal Component Analysis, PCA) model assumes that the system is stable, and the variables follow normal distribution. Hasebe proposed the way to improve the PCA control limit based on ICA (Independent Component Analysis, ICA), dynamic PCA model, recursive PCA and local linearization method of external analysis[14]. All these methods are based on the normal distribution or approximate normal distribution. However, the problem of fault data detection and process monitoring for the flowrate data of steam pipe network, shows non-normal distribution characteristics. Obviously, the mentioned methods could not fully meet the requirement of this case.

### **Theory and application of data rectification**

The technology of estimating the unmeasured data, or the data lost by transmission channels is called parameter estimation. Recognition and elimination of gross errors are called gross error detection. According to the process models, estimating the optional values of the measurements is called data reconciliation. The three technologies are referred to as data rectification. Although data rectification technology has been researched for many years, but the applications in industry just began in recent years.

#### **Gross error detection**

Statistical hypothesis testing is the most common and effective method to detect gross errors. Typical methods include Global Test (GT), Measurement Test (MT), node test (Nodal Test, NT), generalized likelihood method etc. There are many innovated methods to improve the performance of these algorithms. Zhang [15] proposed a method to remove gross errors based on redundancy analysis. Congli[16]combined NT and MT methods and applied sequence compensation to detect significant errors. These mentioned methods belong to the static gross error detection for linear constraints, and have essential limitations in practice.

Aiming at the nonlinear constraints and the presence of unmeasured variables, the researchers propose the iterative measurement and detection, pollution Gauss distribution method [17], robust estimation and Akaike Information Criterion(AIC) to adjust the parameters of robust estimation [18].

The problem of dynamic gross error (outlier) detection attracted more attention. The traditional methods include Kalman filtering dynamic gross error detection criterion, generalized likelihood ratio method, sliding window method, unbiased estimation technique etc. Reference [19]proposed the method to detect the fix deviations and the outliers of the instruments simultaneously. Chen et al. [20] used particle filter to solve the problem of dynamic data reconciliation and outlier detection, and compared it with the Kalman filtering method[21]. Because the nonlinear dynamic data rectification

problem is affected by the errors from various sources such as model error, algorithm error and measurement error, the theory still needs to go through further research and exploration.

#### Data reconciliation algorithm

Data Reconciliation (DR) is the method to optimize the estimation of the actual values and reduce the influence of random errors.

Kueh first proposed the concept of data reconciliation in the field of chemical industry, and transformed this problem into an optimal estimation problem subject to constraints. Mah and Crowe applied the matrix projection technology to decompose the data reconciliation problem into the least square estimation with linear constraints and the estimation of unmeasured variables. Narasimhan et al. Discussed the method of data coordination under the boundary conditions of the coordination value and the significant error. These theories and applications provide the basis for data rectification.

Knepper was the first proposed to be solved by the nonlinear regression parameter estimation method, and the projection matrix technique was extended to the nonlinear situation by Crowe and so on. Zhang Ping and Rong gang[22]used the scheduling information to simplify the solution of the bilinear data reconciliation problem for the actual problems of the material balance in refinery, which obtained good effect. Document [23]develops a maximum likelihood correction method and constructs a probabilistic statistical framework for data reconciliation. The introduction of robust estimation and detection statistics makes the data reconciliation process insensitive to deviating from the ideal probability distribution and outliers. In document[24], the virtual observation equation, penalty function, linearization and equivalent weight method were applied to simplify the robust estimation and reduce the amount of computation.

The dynamic data coordination (Dynamic Data Reconciliation, DDR) is more practical for actual process because of the dynamic in the production process. By Kalman filter or the improved Kalman filtering, the dynamic balance of the stock and flow can be estimated, and the scope of the method expanded. Bai et al[25] defined this process as dynamic data reconciliation filter, combining the estimationg the covariance and Bayesian theory, and established the rigid objective function of dynamic data reconciliation.

It is usually necessary to consider both dynamic data reconciliation and parameter estimation at the same time. In document[18], a three-segment Hempel truncation estimation and Akaike Information Criterion were used to set the parameters. Iterative nonlinear dynamic data reconciliation [26] took into account the upper and lower limit of variables and other algebraic constraints, and overcome the disadvantages of sliding window method. The Untraced Kalman filter [27]improves the previously proposed RNDDR method to reduce the errors caused by the parameter adjustment of Kalman filter. The wavelet technology [28-30] also became one of the tools for data coordination. Although [30]systematically summarized the method of wavelet analysis for dynamic data reconciliation, there is no ideal way to select wavelet basis functions and resolutions in practical applications currently.

### **The application of data rectification technology**

Data rectification technology have been applied in practice[31] [32]. A lot of commercial softwares also appeared, such as the Aspen Operation Reconciliation and Accounting of AspenTech company, Production Balance in the Business FLEX suite of the Honeywell company, ESP-Supplant DataPro of Supcon company and so on. These software adopted the relevant technology of data rectification, and gradually popularized and applied [33]in the petroleum and chemical industry.

However, there are few reports on the application in steam pipe network modeling and data rectification. At present, the most advanced steel enterprises in China mainly operate the system by combining the data of measurement and the experienced manual confirmation.

### **Analysis of the potentials and problems for the application in steam pipe networks**

For the section of steam networks in EMS, the procedure of data processing is essential. Due to the complex environment (widely distributed and electro-magnetic interferences) of data transmission in iron&steel plants, the “3i” problems (inaccurateness, incompleteness, inconsistency) inevitably exist.

It is supposed that the quality of the data would be significantly improved with such a process of data rectification. The operation and schedule for the production would be executed based on relatively complete, accurate, consistent and timely data. The performance and safety of the production could be secured effectively.

However, the actual condition is the EMS could only provide low redundancy in data space, low accuracy and poor consistency in the measured data, and cannot perform data rectification directly. Therefore, the commercial software cannot be applied in the iron& steel enterprises directly. Based on the related research progress, combined with the technology demand of steam and pipeline network, the future research problems of data rectification could be as follows.

For the problem of modelling for the steam networks, the hydraulic and thermodynamic equations can provide a series of certain basis for the network measurement data rectification (i.e., constraint equations), the rectified data results should be more convincing when the precision of the model is higher than the precision of the actual measuring instrument[34]. But the precision of the model will usually be challenged in the real applications[35] [36]. Compared with mechanism modeling, data-driven modelling doesn't start with physical laws. The method directly uses historical data or online data to analyze the internal relationship between the data from various instruments. If the method is properly applied, the constraint relation between variables can be quickly obtained. The modeling theory based on data driven and expert knowledge has become a hot research topic in recent years[37]. Data driven modelling of steam pipe network in iron and steel enterprises will be a new exploration and could be the start of the solution of data rectification for EMS.

For the problem of fault data detection and process monitoring, there are many methods based on the rigid mathematical models. But for a large-scale system as the complex steam network, it's unfeasible to monitor so many variables simultaneously. The abnormal data can be discovered with data-driven monitoring technology, in a univariable and multivariable SPC. The core idea is to establish a statistical model through the process, and to analyze abnormality according to the statistical model. The data from the actual application usually follow non-normal distribution, therefore, the problems of non-normal distribution, univariate and multivariate statistical process control become the research direction. The reference[38] showed that dynamic fault diagnosis based on data driven is also a research direction.

For the problem of data rectification applied in the steam networks, due to the complexity of the network structure and the properties of steam, the nonlinearity and dynamics should be considered as the usual condition. These lead to the limitation of the application of data rectification theory and algorithm. If the temporary stable state of the production section can be determined correctly, to estimate the fault data because of the sensor failure, communication network problems becomes possible. The judgment of the process state and fault data recognition should be included in the scope of research in this field. The detection and removal of gross errors are statistical meaning. The probability of the occurrence of errors always exists and is non-zero. The current gross error detection methods are mainly innovated from the two basic test methods of MT and NT. The method of detecting gross errors in dynamic operation of steam networks is still worth further research. In the other hand, it is the main direction to combine statistical methods and data reconciliation in nonlinear and dynamic situations to adapt the characteristics of steam pipe networks. The methods of synthetic test, Akaike Information Criterion(AIC) information criterion and mixed integer linear programming, unscented Kalman filter, particle swarm optimization and nonlinear programming can be tried as the method to minimize of objective function and realize dynamic data reconciliation.

## Conclusion

Steam is an important energy medium in iron& steel enterprises. The schedule and operation performance of the steam network affect the waste heat recycling and the overall energy consumption of th enterprises. Steam networks are remotely controlled via EMS platform. The schedule and operation command should be based on the data with high quality. This paper reviews the related researches which can improve the process data quality including steam network modelling, fault data detection and process monitoring, gross errors detection and data reconciliation.

The mathematical model from physical principles (hydraulic and thermal models) are too complex to setup and applied in practical steam network to monitor the system movement and rectify the data from the instruments. By data-driven method, the model could be much easier to obtain and more effective. Data driven modelling of steam pipe network should be the start of the solution of data rectification for EMS.

Fault detection and process monitor based on SPC is more practical than the method based on the rigid mathematical models. The problems that the data from the field usually don't follow normal-distribution, and dynamic fault diagnosis based on data driven should be further researched.

Data rectification includes parameter estimation, gross error detection and data reconciliation. The nonlinearity and dynamics being considered as the usual condition lead to the limitation of the application of data rectification theory and algorithm. Several potential methods are suggested as the probable solution to realize data rectification.

By applying data-driven modelling and data rectification, the data with higher quality can be derived and the remote operation would be more reliable. At the same time, the large amount of investment for traditional hardware redundancy is saved. Therefore, data rectification technology is meaningful for large scale production processes and has great potentials for further development.

### **Acknowledgement**

This work is supported under the grant projects of Chinese National Sciences Foundation (No. 61463003 , 51409047), the Chinese Scholarship Council(No.201508360120) and project from Jiangxi Province (JXNE2017-01) .

### **References**

- [1] Chen, G., An overview of the production and operation of China's iron and steel industry in 2011. *Metallurgical Management*, 2012. (4) : p. 16—19.
- [2] Li, D.Z.G., ed. *Data Driven Fault Detection Technology in Industrial Processes. The Methods Based on PCA and PLS*. 2011, Scientific Press: Peking.
- [3] Zhang, Q., et al., Development of Energy Management System in Integrated Iron and Steel Works, in *Advanced Research on Industry, Information Systems and Material Engineering*, Pts 1-7, H. Zhang, G. Shen, and D. Jin, Editors. 2011. p. 1737-1740.
- [4] Liang, J. and T. Li, Research on constraint based data quality evaluation algorithm. *Science and technology and Engineering*, 2012. 12 (3) : p. 551—554.
- [5] Xianxi, L. and Y. Ming-zhe, *Data Processing Approaches for the Measurements of Steam Pipe Networks in Iron& Steel Enterprises, in Energy Efficiency*. 2012, INTECH: Rijeka,Croatia.
- [6] A. Morris, J.D., M. Kourmpetis, High Temperature Steam Pipelines - Development of the ARCMAC Creep Monitoring System. *Strain*, 2006. 42(3): p. 181-186.
- [7] Tian, Z. and F. Bao, Computer real time simulation of super large heat network. *Journal of Shanghai Jiao Tong University*, 2000. 34(4): p. 486-489.
- [8] Zhang, Z., Study on the theory and application of hydraulic thermo mechanical coupling calculation of steam pipe networks[D]. 2009, Qiangdao:China University of Petroleum,2008.
- [9] Song, Y., F. Cheng, and R. Cai, Modelling and Simulation of Steam Heating Networks. *Journal of Tsinghua University Natural Science Edition*, 2001. 41 (10) .
- [10] Gao, L., Research and software development of hydraulic thermo mechanical coupling calculation method for steam pipe network[D]. 2009, Shandong Jianzhu University,2009.
- [11] Luo, X. and S. Liu. Modeling and Simulation of Steam Pipeline Network with Multiple Supply Sources in Iron& Steel Plants. in *Chinese Control and Decision Conference*. 2016. Yingchuan.
- [12] 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 (2007) : p. 472-481.
- [13] Kuo, H.C.W.-L., Comparisons of the symmetric and asymmetric control limits for X and R charts. *Computers & Industrial Engineering*, 2010. 59 (2010) : p. 903-910.
- [14] Hasebe, M.K.S., Evolution of multivariate statistical process control: application of independent component analysis and external analysis. *Computers and Chemical Engineering*, 2004. 28 (2004) : p. 1157-1166.
- [15] Zhang, P., G. Rong, and Y. Wang, A new method of redundancy analysis in data reconciliation

- and its application. *Computers and Chemical Engineering*, 2001. 25(7-8): p. 941-949.
- [16] Congli, M., S. Hongye, and C. Jian, An NT-MT Combined Method for Gross Error Detection and Data Reconciliation. *Chinese Journal of Chemical Engineering*, 2006. 14(5): p. 592-596.
- [17] Alhaj-dibo, M., D. Maquin, and etal, Data reconciliation: A robust approach using a contaminated distribution. *Control Engineering Practice*, 2008. 16(2): p. 159-170.
- [18] Arora, N. and L.T. Biegler, Redescending estimators for data reconciliation and parameter estimation. *Computers and Chemical Engineering*, 2001. 25(11-12): p. 1585-1595.
- [19] Abu-el-aet, Z.H. and V.M. Becerra, Combined bias and outlier identification in dynamic data reconciliation. *Computers and Chemical Engineering*, 2002. 26(6): p. 921-935.
- [20] Chen, T., J. Morris, and E. Martin, Dynamic Data Rectification Using Particle Filters. *Computers and Chemical Engineering*, 2008. 32(3): p. 451-462.
- [21] Singhal, A. and D.E. Seborg, Dynamic data rectification using the expectation maximization algorithm. *AIChE Journal*, 2000. 46(8): p. 1556-1565.
- [22] Zhang, P. and G. Rong, Steady-state Bilinear Data Reconciliation Dealing with Scheduling, in 15th IFAC World Congress. 2002: Barcelona, Spain.
- [23] Morad, K., B.R. Young, and etal, Rectification of Plant Measurements Using a Statistical Framework. *Computers and Chemical Engineering*, 2005. 29(5): p. 919-940.
- [24] Zhou, L., H. Su, and J. Chu, A New Method to Solve Robust Data Reconciliation in Nonlinear Process. *Chinese Journal of Chemical Engineering*, 2006. 14(3): p. 357-363.
- [25] Bai, S.H., J. Thibault, and etal, Dynamic Data Reconciliation: Alternative to Kalman Filter. *Journal of Process Control*, 2006. 16(5): p. 485-498.
- [26] Vachhani, P. and etal, Recursive estimation in constrained nonlinear dynamical systems. *AIChE Journal*, 2005. 51(3): p. 946-959.
- [27] Vachhani, P. and S. Narasimhan, Robust and reliable estimation via unscented recursive nonlinear dynamic data reconciliation. *Journal of Process Control*, 2006. 16(10): p. 1075-1086.
- [28] Bakshi, B.R. and etal, Multiscale bayesian rectification of data from linear steady-state and dynamic systems without accurate models. *Industrial & Engineering Chemistry Research*, 2001. 40(1): p. 261-274.
- [29] Ungarala, S. and B.R. Bakshi, A multiscale, bayesian and error-in-variables approach for linear dynamic data rectification. *Computers and Chemical Engineering*, 2000. 24(2-7): p. 445-451.
- [30] Tona, R.V. and etal, Dynamic reconciliation based on wavelet trend analysis. *Industrial & Engineering Chemistry Research*, 2005. 44(12): p. 4323-4335.
- [31] Bazin, C., D. Hodouin, and M. Zouadi, Data reconciliation and equilibrium constant estimation: Application to copper solvent extraction. *Hydrometallurgy*, 2005. 80(1-2): p. 43-53.
- [32] Deandrade and L.R.P. Lima, Nonlinear data reconciliation in gold processing industrial utility plant. *Minerals Engineering*, 2006. 19(9): p. 938-951.
- [33] Miao, Y., Research and application of data coordination and significant error detection method(D), Zhejiang University. 2009.
- [34] Rieger, L., et al., Data Reconciliation for Wastewater Treatment Plant Simulation Studies-Planning for High-Quality Data and Typical Sources of Errors. *Water Environment Research*, 2010. 82(5): p. 426-433.
- [35] Mitsas, C.L., Data reconciliation and variable classification by null space methods. *Measurement*, 2009. 43(5): p. 702-707.
- [36] Szega, M., Application of the Data Reconciliation Algorithm in Optimization of Measurements Location for Calculation of Thermal Capacity of a Steam Boiler. *Rynek Energii*, 2010(2): p. 149-153.
- [37] Zhang, J., Research on neural network modeling with expert knowledge and its application in soft sensing. 2013, East China University of Science and Technology.
- [38] Bai, X. and L. Dai, Rapid detection method for instrument failure in process industry. *Chemical automation and instrument*, 2008(01): p. 36-39.