

Article

A New Time-Series Fluctuation Study Method Applied to Flow and Pressure Data in a Heating Network

Shuai Zhao ^{1,2}, Huizhe Cao ^{1,2,*} , Jiguang Zhu ³, Jinxiang Chen ⁴ and Chein-Chi Chang ^{1,5}¹ School of Architecture, Harbin Institute of Technology, Harbin 150090, China² Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology, Harbin Institute of Technology, Harbin 150090, China³ Flow Measurement Research Center, Harbin Institute of Metrology, Harbin 150036, China⁴ School of Management, Harbin Institute of Technology, Harbin 150090, China⁵ Department of Chemical, Biochemical, and Environmental Engineering, University of Maryland, Baltimore, MD 21250, USA

* Correspondence: caohz@hit.edu.cn

Abstract: The key to achieving smart heating is the rational use of large amounts of data from the heating network. However, many current relevant studies based on generalized mathematical methods are unable to accurately describe the physical relationships between pipe network variables. In order to solve this problem, this paper proposes a new time-series fluctuation research method, which can be applied to the measured data of the hot water heating pipe network. This method is a new approach to identifying step data. Then, we propose the concept of time-series disturbance to quantify the degree of data anomaly. Finally, the results of a case study demonstrate the transfer process of a significant disturbance in the pipe network from the supply end to the return end. The time-series fluctuation method in this paper precisely describes two physical relationships between heating system variables and provides a feasible and convenient new research idea for self-perception and self-analysis of smart heating.

Keywords: hot water heating networks; smart heating; identification of step data; time-series data



Citation: Zhao, S.; Cao, H.; Zhu, J.; Chen, J.; Chang, C.-C. A New Time-Series Fluctuation Study Method Applied to Flow and Pressure Data in a Heating Network. *Energies* **2023**, *16*, 2709. <https://doi.org/10.3390/en16062709>

Academic Editor: Sergey M. Frolov

Received: 22 November 2022

Revised: 5 March 2023

Accepted: 9 March 2023

Published: 14 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Of China's total energy emissions, the proportion of building-related carbon emissions is 51.2%. In particular, the proportion of carbon emissions from the operational phase of buildings is 21.9% [1]. Heating is one of the major causes of energy consumption in the operational phase of buildings, especially in northern China. It is important to reduce heating energy consumption and improve heating energy efficiency under indoor temperatures to meet the needs of the normal lives of residents [2–4]. Smart heating has greater advantages than traditional heating systems in terms of energy savings, regulation, and fault diagnosis, and has huge development potential and marketing prospects [5]. Nowadays, temperature, pressure, and flow sensors are usually installed at every important monitoring point in a heating system. Through a monitoring and data-acquisition system, operational data can be monitored and recorded in real time to support the operational maintenance of a heating system [6,7]. The heating system accumulates a large amount of time-series data, which represent the state changes of the pipe network during the operation of the heating system. In the context of the big-data era, identifying the step data caused by disturbances and analyzing their processes and characteristics is the focus of temporal analysis of data.

Data collected by a heating system over a certain period of time may suddenly fluctuate from a stable state for some reason, then return to the stable state. Data with a step phenomenon between the two stable states are called step data. Step data indicate that the hydraulic status of a pipe network changes over time; therefore, identifying step data

involves determining the time point and the numerical change of the hydraulic state of the pipe network. The reasons for the change in fluid states can be divided into two categories—normal pipe network regulation and faults in the pipe network system.

There are three main physical quantities that operate in a heating system: temperature, pressure, and flow rate. The precise regulation of temperature is easier to achieve than the regulation of pressure or flow. The main reason for temperature regulation is a change in building thermal load; therefore, many people construct prediction models based on a building's energy load or other related time-series data to achieve the goal of improving the accuracy of load prediction. A complete set of temperature regulation logic is then established through high accuracy load prediction [8,9]. Because a heating system has a complex pipe network structure, it is very difficult to accurately analyze the hydraulic state of the pipe network operation; the physical quantities that indicate the hydraulic state of the pipe network are flow and pressure.

A regulation system is an important part of smart heating [10]. The regulation system senses abnormal pressure data and flow data and immediately takes appropriate adjustment measures, such as adjusting valves, to bring the flow and pressure back to normal. However, since each regulation takes a different time to restore the state of the network, and the system cannot distinguish whether the cause of abnormal data is normal operation or network failure, the system will frequently issue regulation commands that result in oscillations in the network. The main cause of this problem occurs during the analysis stage of a smart heating control system, which can sense abnormal data but cannot analyze the physical relationships of step data, so the system cannot diagnose the cause of the step data, and then issues wrong control orders.

The failure of a heating system refers to the partial or total loss of the system's ability to transport the heating medium. The occurrence of failure can reduce the operating efficiency of the system or even cause efficiency to disappear completely [11]. Many people build corresponding detection models based on wavelet analysis and the threshold method; such models can identify fault data in a set of data [12–14]. HVAC systems have complex structures, so the direct analysis of an HVAC system's raw data is difficult. The processing of raw data by wavelet analysis can obtain characteristic data of an HVAC system, which provide the main operational information about the system and diagnose system faults [15]. Data in many heating systems is monitored and uploaded through automated equipment. The threshold method can identify fault data in such automatically uploaded data and classify the fault data into three groups: an improper heat load pattern, a low annual average temperature difference, and poor equipment control. Faults from the first two fault groups are relatively easy to solve, but faults from the third group are difficult to solve [16]. However, as universal mathematic methods, both wavelet analysis and the threshold method are applied to real measurement data—they cannot describe the physical relationships between variables in HVAC systems well.

Wavelet analysis has a high accuracy rate when the difference between faulty data and normal data is large. However, small fault data differs very little from the normal data. Because the physical relationships between variables in an HVAC system cannot be clearly described, this leads to a low success rate for wavelet analysis in the diagnosis of small faults. Although the threshold method can classify the fault data of a heating system, it has difficulty solving faults from the equipment fault group, because there are many devices in the system and each device may cause a fault. In addition, the threshold method does not accurately describe the physical relationships between variables in the system, so the cause and location of a failure cannot be accurately analyzed; thus, the device that generated the failure cannot be analyzed.

The current data processing of the existing operating conditions of a heating network is mainly applied in two ways. One way is the identification of abnormal data caused by the failure of the network; the other way is the prediction of the building heat load. The mechanisms of both the identification algorithm and the prediction algorithm can be attributed to the processing of discrete data via statistics. Such processing ignores the

characteristics of heating network operation data over time, and its results ignore the inherent time-series coupling relationship between the flow and pressure of each pipeline of supply water and return water. The actual pipe network operation processes consist of a non-constant flow, and the flow and pressure data fluctuate with time. Because the fluctuation process of pipe network data must be time-dependent, the characteristics of data changes with time are lost when the discrete data processing method is applied to actual heating pipe network data. Therefore, it is clear that the realization of smart heating must solve the problem that arises because the physical relationship between variables in a heating system cannot be precisely described [17,18].

The aim of the research in this paper is to calculate the time-series fluctuation of multiple variable data sets, based on data correlation. The time-series fluctuation calculation identifies the change characteristics of multiple variable data sets over time. Therefore, this paper can describe the physical relationships of multiple variables in a pipe network with time, based on the time-series change characteristics.

The measured data of a heating system can be continuously affected by many factors. For example, the factors affecting the supply water flow data of a primary network include heat sources, pipes, valves, pumps, and other pipe network equipment. The result of a combination of factors is that the measured data do not match the physical relationships between the variables in a system. When an anomaly is detected in pipe network data, it indicates that the effect of a certain factor on the data is suddenly strengthened. This factor dominates the changes in the data, such as the start of the relay pumping station, leading to an increase in the water flow and pressure of the pipe network. At this point, it is necessary only to analyze the influence of this factor on the data to understand the actual operating status of the pipe network. In the pipe network's actual measurement data, a sudden change of data from the time-series perspective is mainly manifested in two ways: an increase or a decrease of data values and an acceleration of the data change rate [19,20].

This paper proposes the application of the time-series fluctuation research method to identify abnormal data with sudden changes. Based on the numerical value of the data and the rate of change of the data—that is, the average value of the data and the degree of data fluctuation—this paper proposes the concept of time-series disturbance, through which the physical relationships among the variables of a heating system and the transmission process of the disturbance causing data anomalies in a pipe network are demonstrated. In addition, two problems are solved by time-series disturbance: first, the physical relationships among the variables of a heating system are precisely described; second, the transmission process of the disturbance that causes a data anomaly in the pipe network is discovered.

The identification of step data is important for fault diagnosis and for the precise regulation of a pipeline network. Because flow and pressure are the two most important physical quantities for describing the hydraulic regime of a pipe network, the time-series data that this paper focuses on are flow and pressure data. This paper aims to discover the value of step data and proposes a new method to identify abnormal data. The main contents of the proposed method are summarized as follows:

- (1) Time-series fluctuation is proposed to represent the evolution of time-series data over time.
- (2) A method for the cyclic identification of step data is proposed, because different sets of data have different data characteristics.
- (3) The time interval of step data is classified to judge the data relationships among different data sets.
- (4) The concept of time-series disturbance is proposed to quantify the degree of data anomalies and identify the transmission processes of significant disturbance in a pipeline network.

The rest of this paper is organized as follows. For ease of understanding, Section 2 introduces the time-series fluctuation research method in detail. Section 3 analyzes the results of this research and its application in smart heating. Section 4 provides the main conclusions of this study.

2. Theory

An outline of the research related to time-series fluctuation, as discussed in this paper, is presented in Figure 1. The outline consists mainly of four phases: the calculation of time-series fluctuation, the identification of step data points, the classification of time intervals, and the study of time-series disturbances. Each step provides support for the next stage.

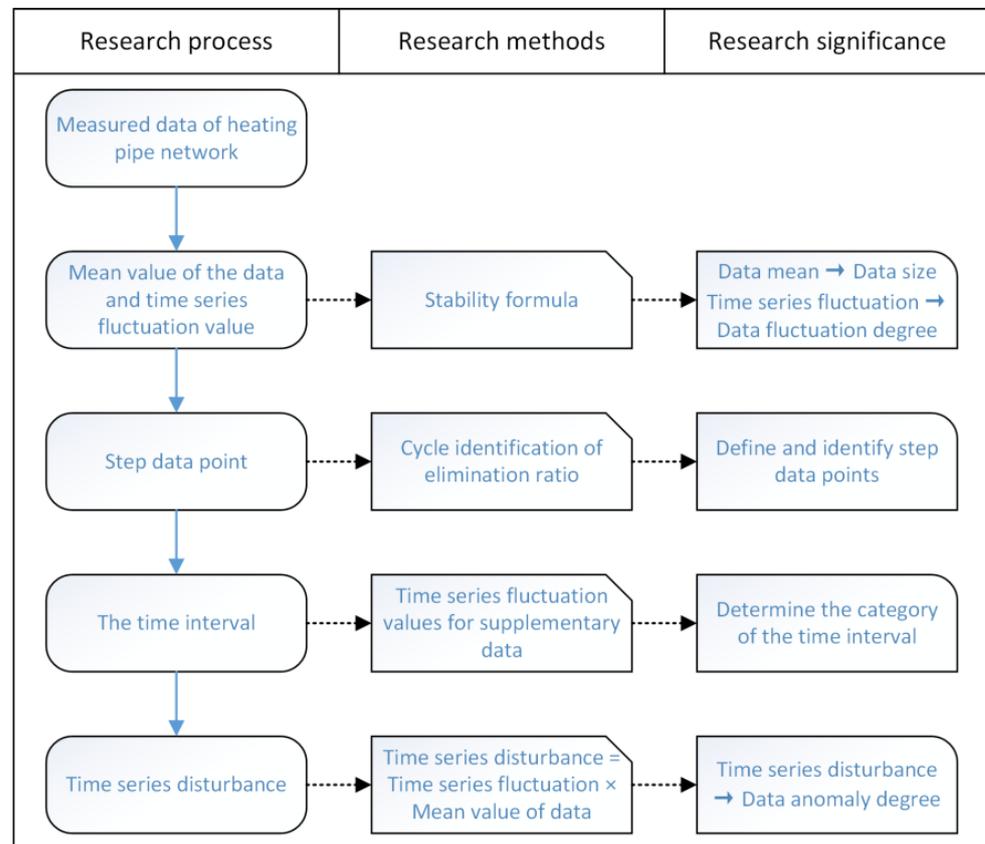


Figure 1. An outline of this research.

2.1. Time-Series Fluctuation

Flow stability E_q is used to measure the stabilization capability of a flow standard device [21,22]. Under experimental conditions, E_q is calculated by continuously obtaining a series of fluid data. The smaller the E_q , the more stable the flow standard device. In fact, the E_q can be understood as the degree of dispersion in a set of fluid data. In this paper, the time-series fluctuation calculation formula is based on stability, as proposed by the ISO standard to measure the stability capability of flow standard devices [23]; accordingly, the calculation method for stability has been validated via the field of flow metering. This paper applies stability to the processing of time-series fluctuation data pertaining to flow and pressure in heating networks, which is among this paper’s innovations.

2.1.1. Calculation Equations

The equations for calculating continuous data for a given sample size n are as follows:

$$q = \frac{1}{n} \sum_{i=1}^n q_i \tag{1}$$

where q_i is the water system data and q is the mean value of the data;

$$E_i = \frac{q_i - q}{q} \times 100\% \tag{2}$$

where E_i is the relative error;

$$R_j = \frac{1}{n-j} \sum_{i=1}^{n-j} E_i E_{i+j} \quad (3)$$

where R_j is the correlation function;

$$r_j = \frac{R_j}{R_0} \quad (4)$$

where r_j is the correlation function ratio—when $r_j < 0.1$, that is, R_j and R_0 are not in the same order of size, $j_{min} = j - 1$;

$$E_q = k \sqrt{\frac{2}{n} \sum_{j=0}^{j_{min}} |R_j|} \times 100\% \quad (5)$$

where the E_q is the flow stability, k is the coverage factor, $k = t_p(v)$, and $t_p(v)$ is the t -distribution function with a confidence level of 95%. The larger the E_q , the more volatile the set of data; the smaller the E_q , the more stable the set of data.

2.1.2. Basic Information about the Example Used for Calculation

The heating network in the calculation is located in Heilongjiang Province, a severe cold zone in northeast China. The heating network has a heating area of about 34 million square meters and serves nearly 260,000 heat users. A schematic diagram of the heating network used in the example used for the calculation is shown in Figure 2.

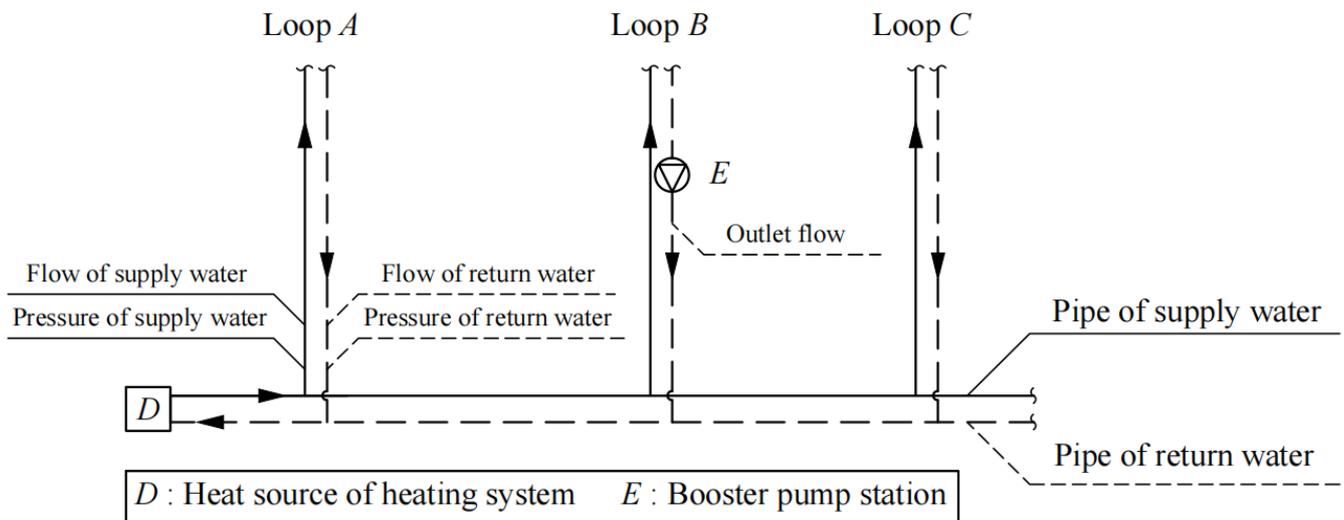


Figure 2. Schematic diagram of the heating network.

The data collection period ran from 16 March 2021 to 18 March 2021, with a 30s data collection frequency and a total of 8640 pieces in each data set. This work analyzes flow and pressure data; the data monitoring points are depicted in Figure 3. This paper contains six data sets, five of which are measured and one of which is calculated. The measured data includes the flow of supply water, the flow of return water, the pressure of supply water, and the pressure of return water of loop A, and the outlet flow of booster pump station E. The calculated data show the difference between the pressure of supply water and return water in loop A.

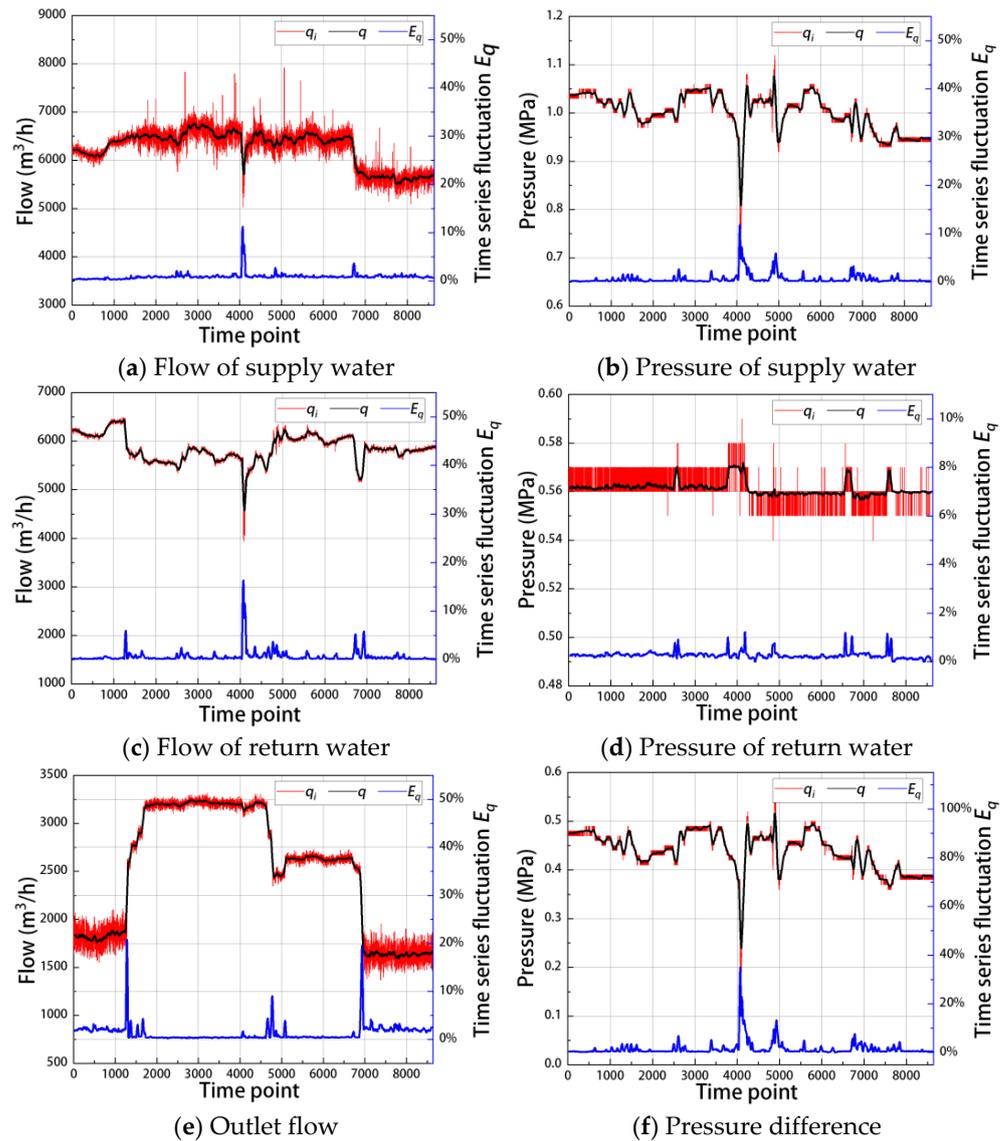


Figure 3. The time-series fluctuation of raw data: (a) flow of return water; (b) pressure of supply water; (c) flow of return water; (d) pressure of return water; (e) outlet flow; (f) pressure difference.

2.1.3. Time-Series Fluctuation Calculation

The volume of circulating water of the entire network is huge, making the heating system a massive inert system. As the response time of the flow change is relatively long, 0.5 h is used as the observation window length. The first to the sixtieth data points in a flow data set represent an observation window, from which the time-series fluctuation value is calculated. Table 1 displays the data from 1 to 60.

Equations (1) and (2) can be used to calculate q , E_1, E_2, \dots, E_{60} . After the calculations, $j_{min} = 0$, so the correlation function is

$$R_0 = \frac{1}{60} (E_1^2 + E_2^2 + \dots + E_{60}^2) = 6.99 \times 10^{-5}$$

According to the t -distribution table, $t_{95}(60) = 2.000$, so the time-series fluctuation E_q is

$$E_q = 2.000 \times \sqrt{\frac{2}{60} \times 6.99 \times 10^{-5}} \times 100\% = 0.31\%$$

For the first to the sixtieth data, the time point of the observation window and the time-series fluctuation value was set to 31; for the second to the sixty-first data, the time point of the observation window and time-series fluctuation value was set to 32; and so on. The time points of the observation window and timeseries fluctuation value are shown in Table 2.

Table 1. The first to the sixtieth flow data of supply water (m³/h).

1–10	11–20	21–30	31–40	41–50	51–60
6196.11	6143.33	6279.17	6157.50	6196.11	6216.94
6245.00	6250.28	6205.00	6185.56	6255.28	6183.06
6128.89	6277.78	6217.78	6228.61	6228.89	6131.94
6293.61	6375.28	6170.56	6122.50	6211.94	6266.11
6181.67	6179.72	6207.50	6182.50	6150.28	6161.11
6235.28	6284.72	6242.50	6161.39	6228.33	6257.78
6250.00	6244.17	6256.39	6247.22	6194.72	6268.06
6148.89	6223.89	6242.78	6167.22	6175.83	6241.67
6267.78	6140.00	6186.94	6174.44	6201.39	6173.61
6290.56	6191.39	6245.83	6315.56	6115.28	6221.11

Table 2. Time points of observation window and time-series fluctuation value.

Data			Observation Window		
Time Interval	Time Point	Time-Series Fluctuation Value	Time Interval	Time Point	Time-Series Fluctuation Value
[1, 60]	31	0.31%	[6703, 6762]	6733	3.67%
[2, 61]	32	0.31%	[6704, 6763]	6734	3.74%
[3, 62]	33	0.31%	[6705, 6764]	6735	3.63%
...			...		
[4042, 4101]	4072	11.24%	[8579, 8638]	8609	0.6427%
[4043, 4102]	4073	11.32%	[8580, 8639]	8610	0.6480%
[4044, 4103]	4074	11.22%	[8581, 8640]	8611	0.6786%
...					

Within the observation window, the E_q is the stability of the 60 fluid data and represents the degree of dispersion of the data. However, when considering the time significance of a set of data, the E_q is the time-series fluctuation value of the time-series data and represents the degree of fluctuation of the time-series data.

Based on the above equation, the time-series fluctuation value of the observation window at time point 31 is 0.31 percent.

Similarly, the above-mentioned time point correspondence was also applied to the other five sets of data. The time-series fluctuation calculation was performed on the six data sets. The results are shown in Figure 3.

Figure 3 shows the data results obtained by calculating the time-series fluctuation of the heating network data. The red line represents the raw data q_i , the black line represents the average value q of the raw data in the observation window, and the blue line represents the time-series fluctuation E_q .

Figure 3a depicts the relationship between the rate of change of the flow data and the time-series fluctuation for the flow of the supply water. When the flow data change quickly, the corresponding time-series fluctuation value increases. For example, the peak of time-series fluctuation at time point 4073 was 11.32 percent due to significant changes in data caused by flow fluctuations; at time point 6734, the peak was 3.74 percent due to drastic changes in data caused by a decrease in the total average flow.

2.2. Step Data Points

Whenever the hydraulic state of a heating network changes, the data produce numerical changes rapidly. The higher the time-series fluctuation value of the data, the faster

the data change. The time-series fluctuation values are sorted in order from low to high. According to the preset elimination ratio, the data with the highest time-series fluctuation value was first identified. Specifically, we used the identified raw data as the step data point of the elimination ratio and used the continuous time points corresponding to the continuous step data points as the time interval of the step data points.

2.2.1. Identification of Step Data Points

In general, different sets of data in the same physical quantity data present different data characteristics. Since a single elimination calculation based on the same elimination ratio for different sets of data will not satisfy all the calculation requirements, the data must be eliminated iteratively. Figure 4 depicts the step data point identification process.

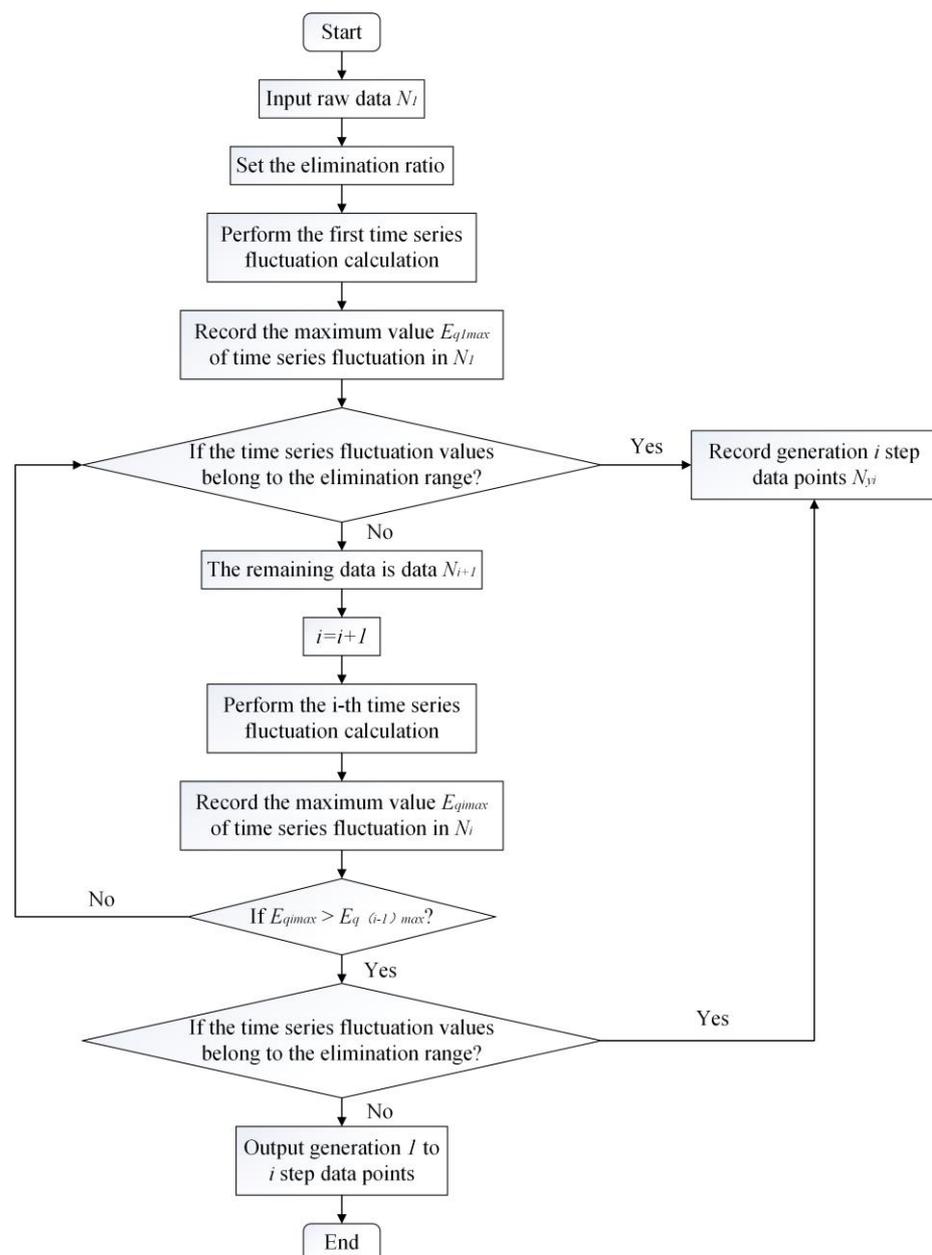


Figure 4. Identification flow chart of step data points.

In this study, N_1 represents the raw data set and E_{q1max} is the highest value of time-series fluctuation value in N_1 . When the first elimination calculation is performed with N_1 , the first generation of step data points can be obtained. Thus, N_{y1} is the set of the first

generation of step data points. Similarly, N_2 represents the remaining data, and E_{q2max} is the highest value of the time-series fluctuation value in N_2 . When the second elimination calculation is performed with N_2 , the second generation of step data points can be obtained. Hence, N_{y2} is the set of second-generation step data points, and so on.

When $E_{qimax} > E_{q(i-1)max}$, this means that after deleting step data points, the maximum value of the remaining data's time-series fluctuation value increases, as does the degree of data fluctuation. The value of N_{yi} should then be recorded, and the elimination calculation should come to an end.

The elimination ratios were set as 0.5%, 1.0%, 1.5%, 2.0%, and 5%, respectively, to generate a variety of step data points to compare so that the best elimination range could be found. To ensure that the data set had a complete amount of data, it was necessary to supplement the data of the excluded step data points. The supplemental data could be calculated using the following equation if the time interval of the step data points was (n, m) ($m > n$).

$$q_i = q_n + \frac{q_m - q_n}{m - n} \times (i - n) \quad (6)$$

where $i = n, n + 1, \dots, m$.

Using the flow data of water supply for reference, we first computed the time-series fluctuation on the original data to determine the number of step data points with various elimination ratios and then computed the time-series fluctuation on the supplementary data to determine the maximum value of time-series fluctuation with various elimination ratios. The results of our calculations are depicted in Figure 5.

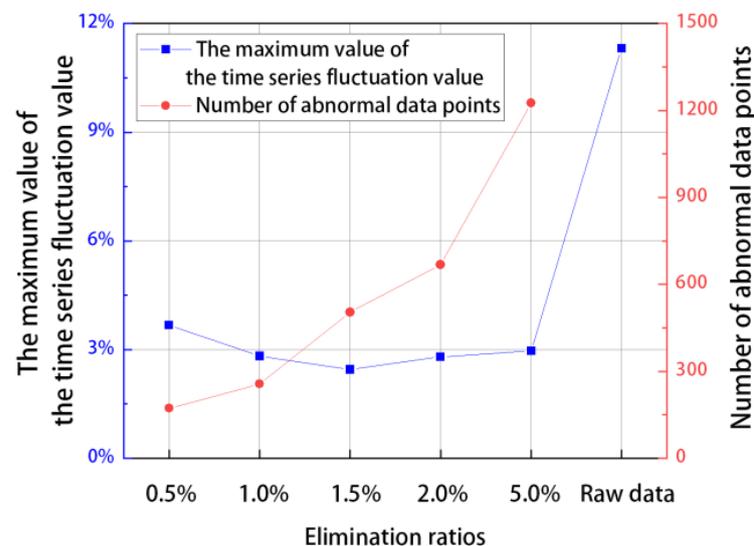


Figure 5. Calculation results of different elimination ratios.

The results show that as the number of step data points increases, so does the task volume of supplementing the data and sorting out the step data points. As shown in Figure 5, the maximum value of the time-series fluctuation value represents the most unstable degree in a set of data, so the smaller the maximum value, the more stable the data. When considering the number of step data points and the maximum value of time-series fluctuation, the elimination ratio of 1.0 percent is preferred.

2.2.2. Classification of Step Data Points

We conducted causal analysis for data anomalies in this section. Because data anomalies can be caused by a variety of factors and can be manifested in a variety of step data points, we categorized all the step data points in order to investigate the potential causes of data anomalies.

The E_q is the value of the raw data time-series fluctuation, and E_q' is the value of the supplementary data time-series fluctuation. c is defined as the difference percentage between E_q and E_q' . It can be expressed as

$$c = \frac{E_q - E_q'}{E_q} \times 100\% \quad (7)$$

In the time interval of step data points, the mean of the time-series fluctuation value represents the average degree of fluctuation of all data in the interval. Therefore, c can be used to indicate how much the data fluctuation is reduced after supplementation.

Tables 3–5 illustrate the statistical step data points and the related calculation results for the three sets of flow data.

Table 3. Step data points and associated calculation results for the flow of supply water.

Time Interval	Step Data Points		The Mean of the Time-Series Fluctuation Values		
	Quantity	Quantity Percentage	Raw Data	Supplementary Data	Difference Percentage
2493–2495	3	1.17%	2.13%	2.11%	1.04%
2497–2500	4	1.56%	2.12%	2.09%	1.46%
2764	1	0.39%	2.05%	2.08%	−1.36%
4015–4151	137	53.52%	5.19%	0.87%	83.32%
4834–4837	4	1.56%	1.57%	1.54%	2.15%
4840	1	0.39%	1.96%	1.97%	−0.38%
4847–4861	15	5.86%	2.66%	2.58%	3.17%
6685–6775	91	35.55%	2.23%	1.90%	15.07%

Table 4. Step data points and associated calculation results for the flow of return water.

Time Interval	Step Data Points		The Mean of the Time-Series Fluctuation Values		
	Quantity	Quantity Percentage	Raw Data	Supplementary Data	Difference Percentage
1261–1314	54	21.09%	4.14%	3.31%	20.08%
4045–4135	91	35.55%	10.99%	1.66%	84.89%
6705–6753	49	19.14%	4.16%	3.35%	19.47%
6895–6956	62	24.22%	5.25%	4.53%	13.77%

Table 5. Step data points and associated calculation results for the outlet flow.

Time Interval	Step Data Points		The Mean of the Time-Series Fluctuation Values		
	Quantity	Quantity Percentage	Raw Data	Supplementary Data	Difference Percentage
1243–1323	81	47.37%	10.33%	6.80%	34.21%
6888–6977	90	52.63%	10.57%	8.67%	17.96%

We defined the short interval of step data (less than or equal to 10% of the quantity percentage) and the long interval of step data (greater than 10% of the quantity percentage) based on 10% of the total amount of step data. Tables 3–5 present comparison results between raw and supplementary data. Most notably, the differences in time-series fluctuation value are small (the difference percentage $\leq 5\%$), and fluctuations in time-series fluctuation value can be up or down in the short intervals. Because of the small number of step data points, the time-series fluctuation value of supplementary data did not decrease

significantly in the short interval and had little effect on the overall data. Additionally, it was discovered that in the long interval, the mean value of time-series fluctuation values decreased by more than 10%. The change suggested that data supplementation reduces data fluctuation in long intervals.

For a more precise analysis of step data in long intervals, we divided the long intervals into two categories. The fluctuation of the data of the heating network is a continuous process, as its medium flow is a continuous process. There are two types of first-kind discontinuities in continuous functions: removable discontinuity and jump discontinuity. Similarly, the long intervals are divided into the removable interval and the jump interval.

The removable interval is a long interval of step data points with a difference percentage greater than 80%, and the data points within this interval are removable data points. There is a relatively small difference $q_m - q_n$ between the two ends of the raw data in the removable interval, indicating that something is causing the data to destabilize, resulting in an increase in the time-series fluctuation value, but through data supplementation, the data returns to a relatively stable state—roughly to its previous level.

The jump interval, on the other hand, is a long interval of step data points with a difference percentage range from 10% to 35%, and the data points within this interval are the jump data points. There is a large difference $q_m - q_n$ between the two ends of the original data in the jump interval, indicating that there is a cause that disturbs the stability of the data and leads to an increase in the time-series fluctuation value; however, the reduction is limited by data supplementation.

2.3. Long Intervals

This section examines four long intervals of flow data to investigate the relationship between time-series fluctuation values of different data sets in the long interval.

2.3.1. Identification of Step Data Points

The research on the flow data yielded four long intervals. According to the network operation record for this period, the switching on and off of the booster pump at time points 1230 and 6938, respectively, caused anomalies in the data of the return flow and the outlet flow in the time intervals [1240, 1320] and [6888, 6977], but the flow of supply water was unaffected. The flow data for supply water and return water were anomalous in the time intervals [4015, 4151] and [6685, 6775], but the outlet flow was unaffected. Table 6 depicts the classification of different flow data for the four long intervals.

Table 6. The long interval classification of flow data.

ID	Range	Supply Water	Return Water	Outlet
I	[1240, 1320]	-	Jump interval	Jump interval
II	[6888, 6977]	-	Jump interval	Jump interval
III	[6685, 6775]	Jump interval	Jump interval	-
IV	[4015, 4151]	Removable interval	Removable interval	-

The time interval of the step data identified in this paper is consistent with the time point of the actual working condition change, which verifies that this method can be applied to the actual heat network model, and future research will be focused on verifying this with more data. As shown in Table 6, each long interval has two sets of different flow data, and the long interval types for the two sets of flow data are the same. To determine whether the reasons for the anomalies of the two sets of different flow data are the same, the variation patterns of the time-series fluctuation values of step data points in the long interval must be studied.

2.3.2. Variation Patterns of the Time-Series Fluctuation Value

As shown in Figure 6, the time-series fluctuation values of step data points were recorded at different long intervals, and the time-series fluctuation values of different flow data were compared.

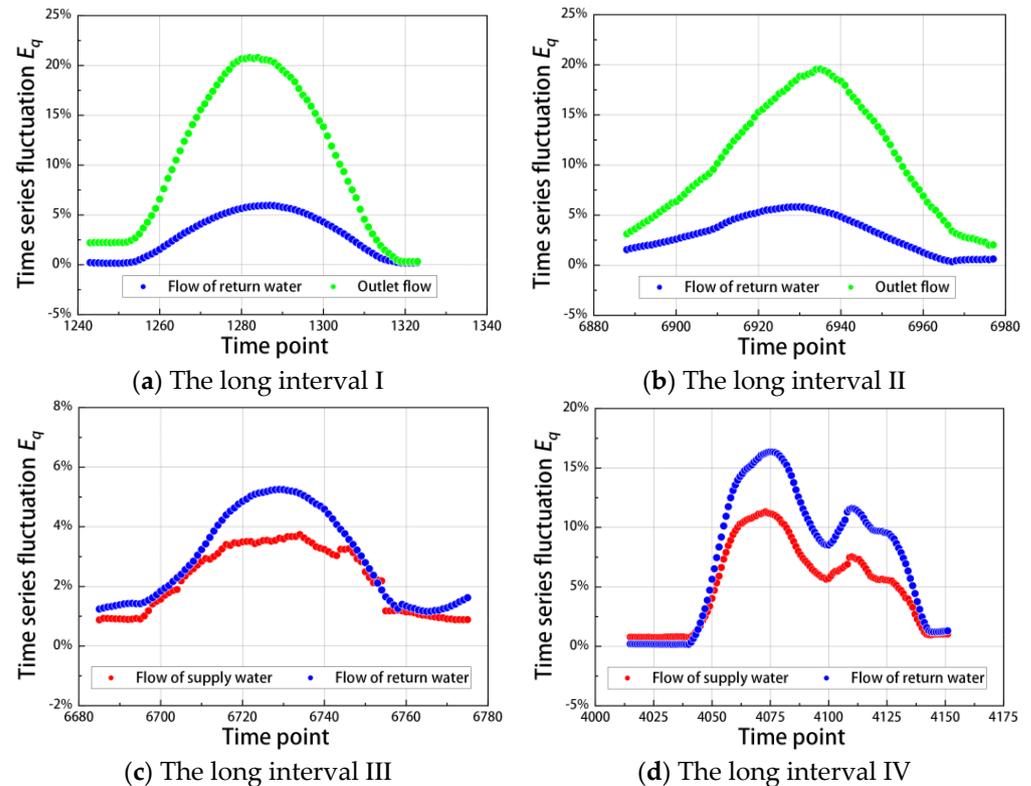


Figure 6. Time-series fluctuation values of step data points: (a) the long interval I; (b) the long interval II; (c) the long interval III; (d) the long interval IV.

The change patterns with time for the time-series fluctuation value of different traffic data are the same, as shown in Figure 6, indicating that the reasons for the abnormality of different flow data are the same. For example, in the long interval IV, as shown in Figure 6d, the change patterns of the time-series fluctuation values, which record the flow of supply water and return water, are the same, indicating that the reasons for the step data are the same. However, the value of the supply water flow data is greater than that of the return water, and the time-series fluctuation value of the supply water flow data is less than that of the return water, indicating that the greater the data value, the greater the ability to resist data anomalies.

From the patterns in Figure 6 and Table 6, the following conclusions can be drawn:

- (1) The main distinction between the jump interval and the removable interval is whether the data anomaly is ongoing or transient from the time-duration perspective. The continuous impact caused by the overall reduction in flow at the supply water end of loop A, for example, is the cause of the step data in the long interval III. The step data in the long interval IV are the result of a transitory pressure change, causing a transitory flow change at the supply water end of loop A.
- (2) The different types of flow data in the long interval can be used to identify the source of data anomalies from a heating network operation data information perspective. For example, the long interval I contains flow data of return water and outlet flow data, and the reason for the disparity between the two sets of data is that the start of the booster pump affects the return flow data. Likewise, the flow data of supply water and return water are contained in the long interval III, and the reason for the

discrepancy between the two sets of data is the flow change at the supply water end of loop A, which affects the flow data of the entire loop.

2.4. Time-Series Disturbance

As discussed in this section, we attempted to predict the abnormal working conditions of the heating network based on the disturbance in the pipe network. As shown in Figure 6, the time-series fluctuation values of different data sets have the same variation patterns in the long interval of step data points, indicating that the causes of the step data are the same. It was found that the water was in the turbulent roughness zone in this interval, by calculating the Re number. In terms of pipe network hydraulics, the pipeline of loop A in Figure 2 can be abstracted as a long-pipe model, and its flow and pressure drop conform to the following equation.

$$\Delta P = SQ^2 \quad (8)$$

where ΔP is the pressure drop of the system, S is the resistance coefficient of the pipe network, and Q is the flow.

Figure 6d depicts the obvious significant disturbance obtained by the calculation of the time-series fluctuation. The disturbance must complete the spatial and temporal transfer of the fluid through the pipe network, so the transmission method must follow Equation (8) for the hydraulic calculation of long pipes, which indicates that the time-series disturbance of flow is transmitted in a square manner and the time-series disturbance of pressure is transmitted in a linear manner. Therefore, the exact time interval and the numerical value of the significant disturbance can be found via the time-series fluctuation research method; then, the abnormal working conditions of the heating network can be predicted.

2.4.1. Time-Series Disturbance of Flow

The time-series disturbance of flow $D_{t,Q}$ can be calculated with the following equation:

$$D_{t,Q} = q_Q^2 \times E_{q,Q} \quad (9)$$

where q_Q is the mean value of the flow data measured during the observation period and $E_{q,Q}$ is the time-series fluctuation value of the flow data in the same period.

Figure 6d depicts the time-series fluctuation value of two sets of data in the time interval [4015, 4151] and Figure 7 depicts the average values of two sets of data. It may be observed that the larger the data mean, the greater the ability to resist data anomalies, and the smaller the time-series fluctuation value.

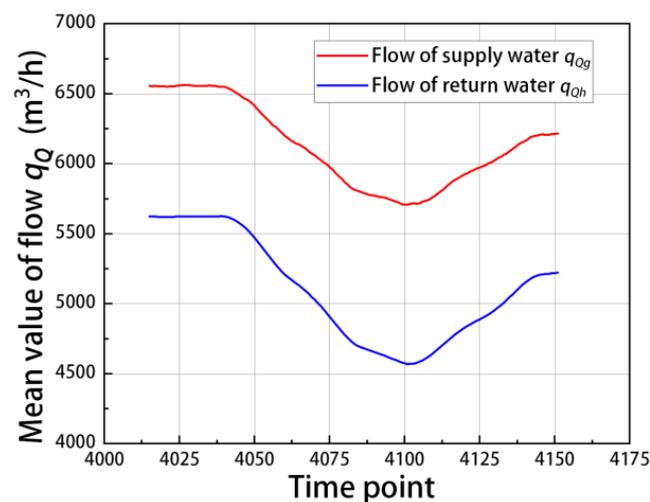


Figure 7. Average value of the flow data.

Here, R_Q was set to the time-series disturbance ratio of the two flow data sets for supply water and return water, which can be calculated as follows:

$$R_Q = \frac{D_{t,Qg}}{D_{t,Qh}} = \frac{q_{Qg}^2 \times E_{q,Qg}}{q_{Qh}^2 \times E_{q,Qh}} \tag{10}$$

where g is the supply water and h is the return water.

Next, the time-series disturbance value and the time-series disturbance ratio were calculated in the time interval [4015, 4151], as shown in Figure 8.

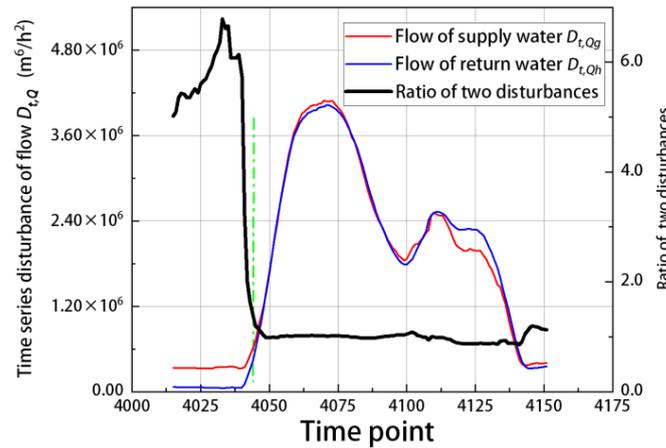


Figure 8. Time-series disturbance value and time-series disturbance ratio of the flow data.

As shown in Figure 8, the left Y axis represents the time-series disturbance of flow and the red and blue lines represent the time-series disturbance of the two, respectively. The right Y axis represents the time-series disturbance ratio and the black line represents the time-series disturbance ratio of the two. The green dash line divides the time interval into two parts. For example, in the time interval [4045, 4151], which is to the right of the green dotted line where the red line and the blue line basically overlap, the time-series disturbance ratio is 0.855~1.200 and the average time-series disturbance ratio is 0.991, indicating that the disturbance is transmitted from the supply water to the return water in loop A. However, in the interval [4015, 4044], which is to the left of the green dotted line, there is a continuous disturbance (non-significant disturbance) in the flow of supply water that is greater than that in the flow of return water. A significant disturbance has not yet formed; thus, the time-series disturbance ratio R_Q in this interval does not reflect transmission from the supply pipeline to the return pipeline.

2.4.2. Time-Series Disturbance of Pressure

The time-series disturbance of pressure $D_{t,P}$ can be calculated with the following equation:

$$D_{t,P} = q_P \times E_{q,P} \tag{11}$$

where q_P is the mean value of the pressure data measured during the observation period and $E_{q,P}$ is the time-series fluctuation value of the pressure data in the same period.

In loop A, where the measurement point of the return water pipeline is close to the point of constant pressure replenishment, the pressure data of the return water are stable and the time-series disturbances are small. In the case of pressure data of supply water and pressure difference data, the time-series disturbance ratio of pressure is described as the ratio of the two.

R_P is set to the time-series disturbance ratio of the two, which can be calculated with the following equation:

$$R_P = \frac{D_{t,Pg}}{D_{t,Pc}} = \frac{q_{Pg} \times E_{q,Pg}}{q_{Pc} \times E_{q,Pc}} \tag{12}$$

where c is the difference.

In the time interval [4032, 4224], the time-series disturbance value and the time-series disturbance ratio are calculated, as shown in Figure 9.

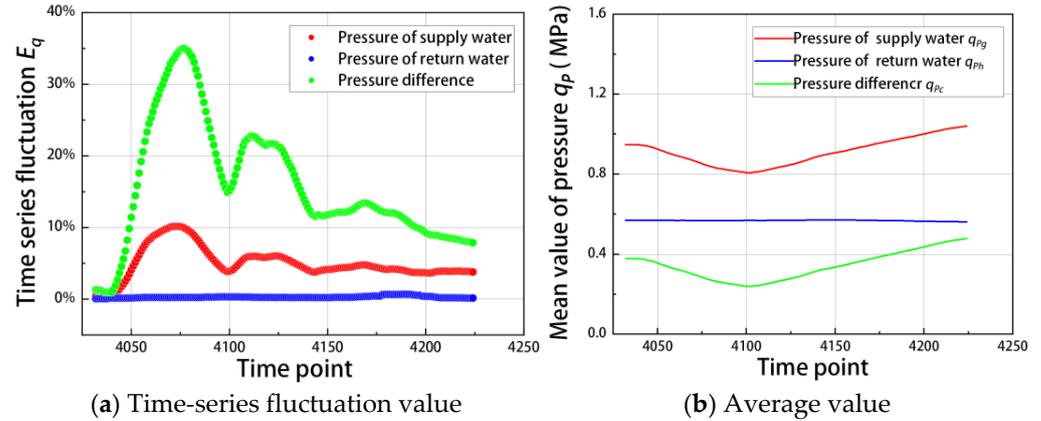


Figure 9. Time-series fluctuation value and average value of pressure. (a) Time-series fluctuation value. (b) Average value.

In Figure 10, the left Y axis represents the time-series disturbance of pressure and the red and green lines represent the time-series disturbance of each; the right Y axis represents the time-series disturbance ratio and the black line represents the time-series disturbance ratio of both. Therefore, in the time interval [4032, 4224], where the red line and the green line basically overlap, the time-series disturbance ratio is 0.857~1.180 and the average time-series disturbance ratio is 0.994, indicating that the strong disturbance is transmitted from the supply pipeline to the return pipeline in the loop A.

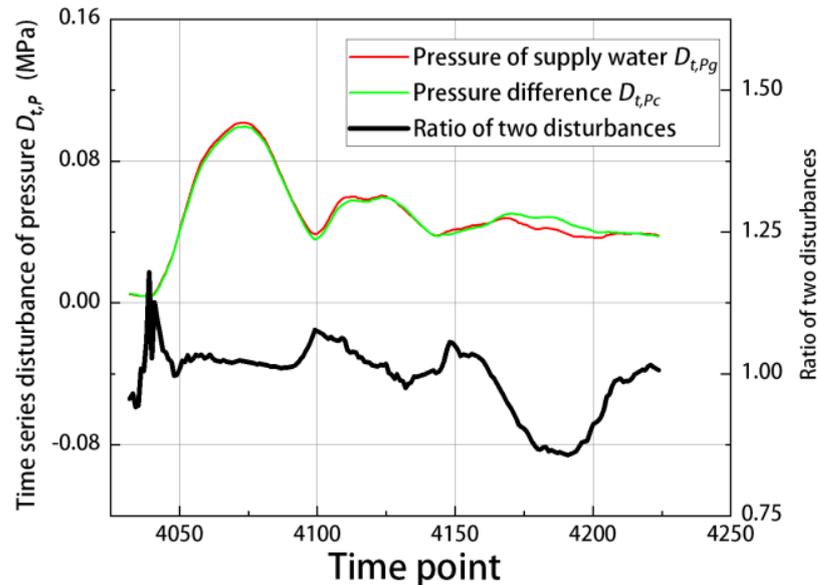


Figure 10. Time-series disturbance value and time-series disturbance ratio of the pressure data.

3. Analysis of Results and Discussion

We calculated the time-series fluctuation value for 6 sets of data, with 8640 pieces in each data set, and identified step data points with time intervals. Then, we selected two typical time intervals to calculate the time-series disturbance of flow and pressure, respectively. Figure 3 shows the calculated time-series fluctuation value for the six sets of data, indicating the basic relationship between the time-series fluctuation values and the fluctuation patterns of the raw data with time. Table 6 illustrates that the categories of

step data are the same for different data sets. In Figure 6, the data monitoring points are different for different data sets; however, the variation pattern of the time-series fluctuation values of the step data is similar, indicating that the disturbance causing the data anomaly completes the spatio-temporal transmission through the fluid in the tube.

The time-series disturbance ratio of the flow data in Figure 8 is 0.855~1.200, and the average time-series disturbance ratio is 0.991, which indicates that the time-series disturbance of the two sets of data is highly fitted. Similarly, the time-series disturbance ratio of the pressure data in Figure 10 is 0.857~1.180 and the average time-series disturbance ratio is 0.994. We concluded that the time-series disturbance to quantify the degree of data anomalies is reasonable, and the high degree of curve fit indicates that the significant perturbation causing data anomalies in different data sets is the same. It is important to note that the significant disturbances demonstrated in this study were transmitted in different data sets, which were the same physical quantity data.

In the long pipe model, the flow rates at both ends of the pipe are equal; however, in the actual pipe network flow data, it is difficult to equalize the supply flow rate and the return flow rate. In addition to the presence of leakage water, this is because the measurement of flow data is affected by many other devices. In the time interval of the step data, this paper demonstrates that the time-series disturbance of supply water flow and return water flow are equal and describes the basic physical relationships among the heating system variables. In the long pipe model, the continuous flow of liquid in the pipe is due to the unequal pressure at the two ends of the pipe, and the liquid flows from the high-pressure end to the low-pressure end. In the actual pipe network pressure data, the supply water pressure is greater than the return water pressure. The pipe network has a constant pressure at the return end, resulting in the return water pressure being almost constant during the time interval of the abnormal data, so the evolution of the pressure difference over time is similar to the supply water pressure. The pressure difference is less than the supply water pressure; however, this paper demonstrates that the time-series disturbance of the supply water pressure and the pressure difference are equal during the time interval of the step data. The pressure difference is what makes the liquid flow in the pipe network, so the disturbance that causes the abnormal pressure data originates from the supply water side. The above analysis describes a new physical relationship among heating system variables. The results show that the time-series fluctuation research proposed in this paper is a reliable method for analyzing fluid data in pipe networks.

Smart heating has five characteristics: self-perception, self-analysis, self-diagnosis, self-decision, and self-learning [24]. Figure 11 interprets smart heating from the perspective of step data.

The prerequisite for the realization of smart heating is the reasonable application of a large amount of measured data, and the focus is on how to accomplish the goal of a “data-driven model”. In this paper, Sections 2.1 and 2.2 proposed a new method of identifying step data, which was applied to the self-perception characteristic; Sections 2.3 and 2.4 proposed the time-series disturbance to analyze the evolution of data over time, which was applied to the self-analysis characteristic. The research carried out in this paper provides a preparatory study for the self-diagnosis of smart heating.

An important contribution of this paper is to extract and retain the time characteristics of the data set from the perspective of time sequence, so as to provide technical support for the future intelligent pipeline network. The way to realize smart heating is to effectively apply time-series data to the heating field. The extraction of the characteristics of time-series step data from the heating network data set is the great significance of this research. Smart heating must be realized through AI or machine learning. The research carried out in this paper provides better technical support for the intelligent application of machine learning in heating. Our next direction is to better realize intelligent heating based on the combination of time-series fluctuation theory and an intelligent algorithm.

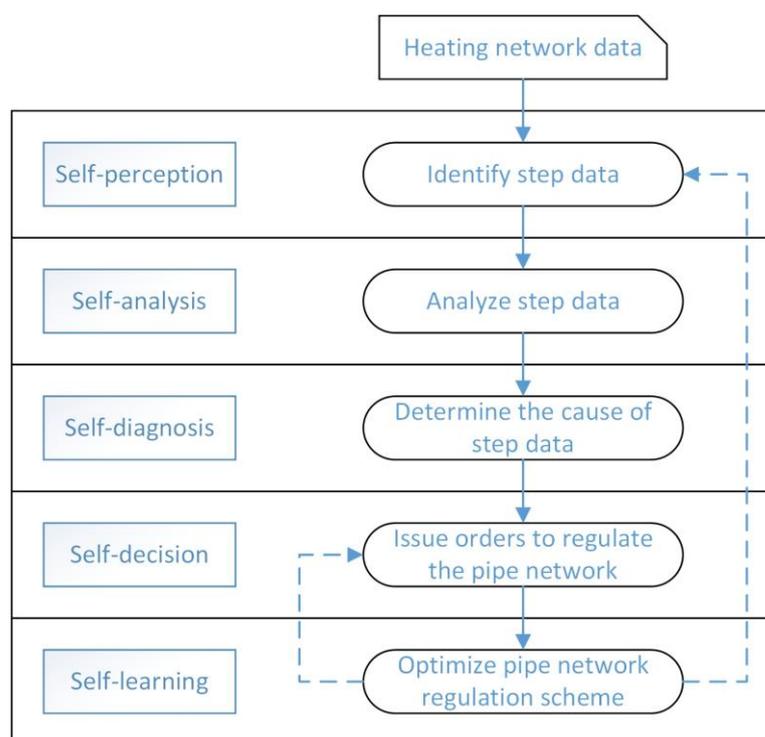


Figure 11. Step data route for smart heating.

4. Conclusions

A prerequisite for realizing smart heating is the reasonable use and analysis of a large amount of data from the heating network, including historical operation data and real-time update data. Currently, many studies on data do not adequately describe the physical relationships among the variables in HVAC systems. Based on the relevant research background, this paper proposed a time-series fluctuation research method that can be applied to the measured data of a hot water heating network, which can identify abnormal data and quantify the degree of data abnormality.

The time-series fluctuation calculation of six sets of data in the study identified step data points and their corresponding time intervals, and the time intervals were divided into long and short intervals. According to the difference percentages and values of the time-series fluctuation between the data at both ends of the intervals, the long intervals were divided into removable intervals and jump intervals. In our research, it was confirmed from the perspective of time duration and heating network operation data information that the anomalies of different data are caused by the same factors based on the flow data categories and the variation patterns between time-series fluctuations. The concepts of time-series disturbance of flow and pressure were proposed, based on the principle of long pipe impedance calculation in hydraulics. The results show that it is reasonable to quantify the data anomalies by temporal perturbations. Based on the analysis of the case study, the time-series disturbance ratio of flow was 0.855~1.200 and the mean value was 0.991, while for the time-series disturbance ratio of pressure, the value was 0.857~1.180 and the mean value was 0.994, indicating that strong disturbances were transmitted from the supply line to the return line. The time-series fluctuation research method applied in this paper provides a feasible and convenient new research idea for self-perception and self-analysis in smart heating.

Author Contributions: Conceptualization, S.Z. and H.C.; data curation, J.C.; formal analysis, S.Z.; funding acquisition, H.C.; investigation, S.Z. and H.C.; methodology, S.Z. and J.Z.; validation, S.Z.; writing—original draft, S.Z.; writing—review and editing, H.C., J.C., and C.-C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Key R&D Program of China (Grant No. 2018YFD1100703).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. CABEE. China Building Energy Consumption Annual Report 2020. *Build. Energy Conserv.* **2021**, *49*, 1–6.
2. Gong, M.; Werner, S. An assessment of district heating research in China. *Renew. Energy* **2015**, *84*, 97–105. [[CrossRef](#)]
3. Zhang, Q.L.; Zhang, L.; Nie, J.Z.; Li, Y.L. Techno-economic analysis of air source heat pump applied for space heating in northern China. *Appl. Energy* **2017**, *207*, 533–542. [[CrossRef](#)]
4. Wu, W.; Wang, B.L.; Shi, W.X.; Li, X.T. Absorption heating technologies: A review and perspective. *Appl. Energy* **2014**, *130*, 51–71. [[CrossRef](#)]
5. Gao, L.; Cui, X.Y.; Ni, J.X.; Lei, W.N.; Huang, T.; Bai, C.; Yang, J.H. Technologies in Smart District Heating System. In Proceedings of the 9th International Conference on Applied Energy, Cardiff, UK, 21–24 August 2017; p. 142.
6. Xue, P.N.; Jiang, Y.; Zhou, Z.G.; Chen, X.; Fang, X.M.; Liu, J. Machine learning-based leakage fault detection for district heating networks. *Energy Build.* **2020**, *223*, 110161. [[CrossRef](#)]
7. Lund, H.; Werner, S.; Wiltshire, R.; Svendsen, S.; Thorsen, J.E.; Hvelplund, F.; Mathiesen, B.V. 4th Generation District Heating (4GDH) Integrating smart thermal grids into future sustainable energy systems. *Energy* **2014**, *68*, 1–11. [[CrossRef](#)]
8. Chen, Z.; Chen, Y.B.; Xiao, T.; Wang, H.L.; Hou, P.W. A novel short-term load forecasting framework based on time-series clustering and early classification algorithm. *Energy Build.* **2021**, *251*, 111375. [[CrossRef](#)]
9. Gadd, H.; Werner, S. Heat load patterns in district heating substations. *Appl. Energy* **2013**, *108*, 176–183. [[CrossRef](#)]
10. Mariano-Hernandez, D.; Hernandez-Callejo, L.; Zorita-Lamadrid, A.; Duque-Perez, O.; Garcia, F.S. A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis. *J. Build. Eng.* **2021**, *33*, 101692.
11. Babiarz, B.; Chudy-Laskowska, K. Forecasting of failures in district heating systems. *Eng. Fail. Anal.* **2015**, *56*, 384–395. [[CrossRef](#)]
12. Guo, J.C.; Lan, Y.; Xue, P.N.; Zhou, Z.G.; Liu, J. Study on Correction Method for Abnormal Operation Data of Heating Network. *Gas Heat* **2020**, *40*, 8–13+41,42. (In Chinese)
13. Yuan, D.; Xue, G.J.; Zhang, H.M. Design of Abnormal Data Recognition Algorithm in Distributed Heating Network. *Comput. Simul.* **2022**, *39*, 494–498. (In Chinese)
14. Li, Y. Network traffic anomaly detection based on time series analysis. *Mod. Electron. Technol.* **2017**, *40*, 85–87+91.
15. Du, Z.M.; Jin, X.Q.; Yang, Y.Y. Fault diagnosis for temperature, flow rate and pressure sensors in VAV systems using wavelet neural network. *Appl. Energy* **2009**, *86*, 1624–1631. [[CrossRef](#)]
16. Gadd, H.; Werner, S. Fault detection in district heating substations. *Appl. Energy* **2015**, *157*, 51–59. [[CrossRef](#)]
17. Zhou, Z.G.; Xue, P.N.; Liu, J.; Fang, X.M.; Zheng, J.F. Construction of intelligent heating engine based on intelligent algorithm and dynamic simulation. *Gas Heat* **2019**, *39*, 12–19+42. (In Chinese)
18. Wu, J.J.; Yang, Y.H.; Cheng, X.; Zuo, H.F.; Cheng, Z. The Development of Digital Twin Technology Review. In Proceedings of the 2020 Chinese Automation Congress (CAC 2020), Shanghai, China, 6–8 November 2020; pp. 4901–4906.
19. Wang, X.; Mueen, A.; Ding, H.; Trajcevski, G.; Scheuermann, P.; Keogh, E. Experimental comparison of representation methods and distance measures for time series data. *Data Min. Knowl. Discov.* **2013**, *26*, 275–309. [[CrossRef](#)]
20. Li, H.L.; Guo, C.H.; Yang, L.B. Fault Detection Algorithm Based on Time Series Data Mining. *J. Data Acquis. Process.* **2016**, *31*, 782–790. (In Chinese)
21. Gareth, J.; Daniela, W.; Trevor, H.; Robert, T. *An Introduction to Statistical Learning*; Springer: New York, NY, USA, 2013.
22. Yu, L.J.; Zhao, W.; Shu, X.; Zhang, S.C.; Tao, D.X. Experimental Research and Analysis on the Stability Calculation Method of Liquid Flow Standard Device. *Meas. Tech.* **2020**, *1*, 28–31.
23. *DIN ISO 9368-1-1993*; Measurement of Liquid Flow in Closed Conduits by the Weighing Method—Procedures for Checking Installations—Static Weighing Systems. ISO: London, UK, 1993.
24. Fang, X.M.; Yang, D.Y.; Zhou, Z.G.; Liu, J. Connotation and Goal of Smart Heating. *Gas Heat* **2019**, *39*, 1–7+41. (In Chinese)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.