

校园供水系统智能管理

摘 要

利用校园内各智能水表的实时供水系统运行数据,通过数学建模和数据挖掘及时发现供水系统中存在的问题,并用科学有效的手段解决这些问题,以提高校园服务和管理水平。

针对问题一,通过对各个水表数据的重新分类统计了每日用水量,并利用模型计算出学生宿舍楼区、教学楼区和食堂区的日最大小时用水量、日平均小时用水量和小时变系数,还进一步分析了小时变化系数的统计特征,如:学生宿舍楼区(均值 2.9876,标准差 1.4334)、教学楼区(均值 2.652,标准差 1.631)、食堂区(均值 21.4525,标准差 23.4736),总的用水趋势,体现出寒暑假、用水时段、用水大小和集中度上,不同功能区都有所区别。

针对问题二,由于各级表的关系是确定的,所以可以对各子网供水系统进行多元线性回归模型的建立,其中以 403X 为例,建立 11 个随机自变量的多元线性回归,利用数据计算出了回归系数和回归置信区间,又通过最小二乘法、残差分析,剔除异常点等处理,最终得到最优拟合模型($R^2=0.9687$),并对该模型进行了检验,随机误差符合正态分布。

针对问题三,对 13 个一级表数据进行分析,并将 419T-406T 合并,从而精简为 7 个一级表,并通过建立 2σ 标准差提取异常数据模型,计算出漏水量,计算出各子网供水系统和整个学校的漏损情况。整个学校平均漏损比率随季度分别为:4.92%、7.01%、4.82%、4.54%,全年为 5.31%。

针对问题四,供水网络错综复杂,但网络具有层级性,因此,通过建立 BP 神经网络模型智能识别出漏损位置,通过对全网各节点的智能识别,发现明显的漏损数量,第一季度 16 个,第二季度 21 个,第三季度 31 个,第四季度 6 个。智能识别对突发性,爆裂性漏损非常灵敏,对恒定微弱的漏损不能很好的辨别,因此,再通过引入小时变化系数,进一步对恒定微弱漏损的情况进行识别。如:K 酒店、L 馆、第九宿舍等。

针对问题五,漏损情况不维修,则会因为大量漏水而增大用水成本,若维修,则会增加维修成本,但漏损成本会随之减少,可见漏损成本与维修成本存在反比非线性关系,由于现实中不可能为了将漏损降为零而无限的投入维修费用,所以,只要将总的漏损比率控制在一定范围即可。按照贵州水价 3.3 元/ m^3 计算,全年漏损成本为 71109.984 元,并计算得到,要在原有维修成本的基础上,增加 4207.1832 元,才能将漏损率控制在 5% 以下。

关键词: 统计特征; 多元线性回归模型; BP 神经网络模型; 智能诊断

一、问题重述

本题主要是针对校园供水系统的数据分析和挖掘，以发现和解决问题，为了保障校园供水系统的正常运行而提供有力科学依据，从而提高校园服务和管理水平。

第一问：统计、分析各个水表数据的变化规律，并给出校园内不同功能区（宿舍、教学楼、办公楼、食堂等）的用水特征。

第二问：结合校区水表层级关系，建立水表数据之间的关系模型，并利用已有数据分析模型精确度。

第三问：输水管网的漏损是一个严重问题。资料显示，在维护良好的公共供水网络中，平均失水在 5% 左右；而在比较老旧的管网中，失水则会更多。利用提供的数据，建立数学模型，分析该校园供水管网的漏损情况。

第四问：地下水管暗漏不容易被发现，需要花费大量人力对供水管道的漏损进行检测及定位，通过建立一个适当的模型和利用水表的实时数据及时发现并确定发生漏损的位置。

第五问：管网维修需要一定的人工费和材料费，但同时可以降低管网漏损程度。请根据以上结果和你了解的水价及维修成本确定管网漏损的最优维修决策方案。

二、问题分析

问题一，通过各级表的实时数据（15 分钟），统计为 1 小时、1 日、一季度的数据，并根据供水领域专业的相关分析，如：日用水量、日平均小时用水量、小时变化系数等分析用水特征，以及从统计学角度，分析数据的统计特征。

问题二，根据水表层次关系，初步估计应该符合多元线性回归，因此可以尝试建立多元线性回归模型，带入数据，将回归系数计算出来，并用真实数据进行检验，如果不是最优，可以通过最小二乘法和误差分析，寻找最优模型。

问题三，根据数据，分析漏损的特征，通过这些特征提取异常数据（即漏水量），就可以计算出学校平均漏损比率和各子网系统漏损比率。

问题四，由于供水系统呈网状，具有层级性，非常适合 BP 神经网络模型的使用，通过将某段相对稳定的数据作为训练数据去预测其他时间的数据，若是预测数据与真实数据相差过大，那么该位置就很大可能是漏损。漏损主要呈现两种状态，一种是较恒定微弱的漏损，另外就是突发性较猛烈的漏损。神经网络对于第二种识别是很敏感的，因此再通过小时变化系数来判断第一种，这样就基本上解决了绝大多数漏损识别问题。

问题五，漏损成本与维修成本应该存在反比非线性关系，由于现实中不可能为了将漏损降为零而无限的投入维修费用，所以，只要将总的漏损比率控制在一定范围即可。另外，根据漏损情况集中出现的时段，更有针对性的提前检查预防。

三、模型假设

- 1、供水系统整体运行稳定；
- 2、只要用水，水表就如实记录数据；
- 3、数据较为全面的反映了学校的用水情况，没有遗漏。

四、符号说明

符号	说明
Q_h	最大小时用水量
Q_p	平均小时用水量
K_h	小时变化系数
SSE_j	掉 x_j 后残差平方和
c_{jj}	是 $c = (x^T x)^{-1}$ 对角线 $j = (j = 0, 1, \dots, k)$ 上第 j 个元素
SSR	回归平方和
SSE	残差平方和

五、模型的建立与求解

5.1 问题一的模型建立与求解

5.1.1 问题一的模型建立

(1) 小时变化系数^[2]

$$K_h = Q_h \div Q_p \quad (1)$$

Q_h : 最大小时用水量;

Q_p : 平均小时用水量;

K_h : 小时变化系数;

(2) T 检验模型

$$\text{样本均值: } \bar{X} = \frac{1}{n} \sum X_i \quad (2)$$

$$\text{样本方差: } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (3)$$

$$\text{样本标准差: } S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (4)$$

S 已知, μ 未知, 可用 T 分布检验数据, 且参数 μ 的置信区间为:

$$\left[\bar{X} - \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1), \quad \bar{X} + \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1) \right] \quad (5)$$

5.1.2 问题一的模型求解

对指标进行分类，较为确定的有三类（学生宿舍区、教学区、食堂区），现就针对这三类进行数据分析。

5.1.2.1 学生宿舍区特征分析

学生宿舍区包括：第一、二、三、四、五、七、八、九宿舍楼，数据是每隔 15 分钟记录一次，现在通过 MATLAB 编程，将学生宿舍区的数据转化为 365 天。并作出学生宿舍区及各学生宿舍楼日用水走势图（图 1、图 2）。

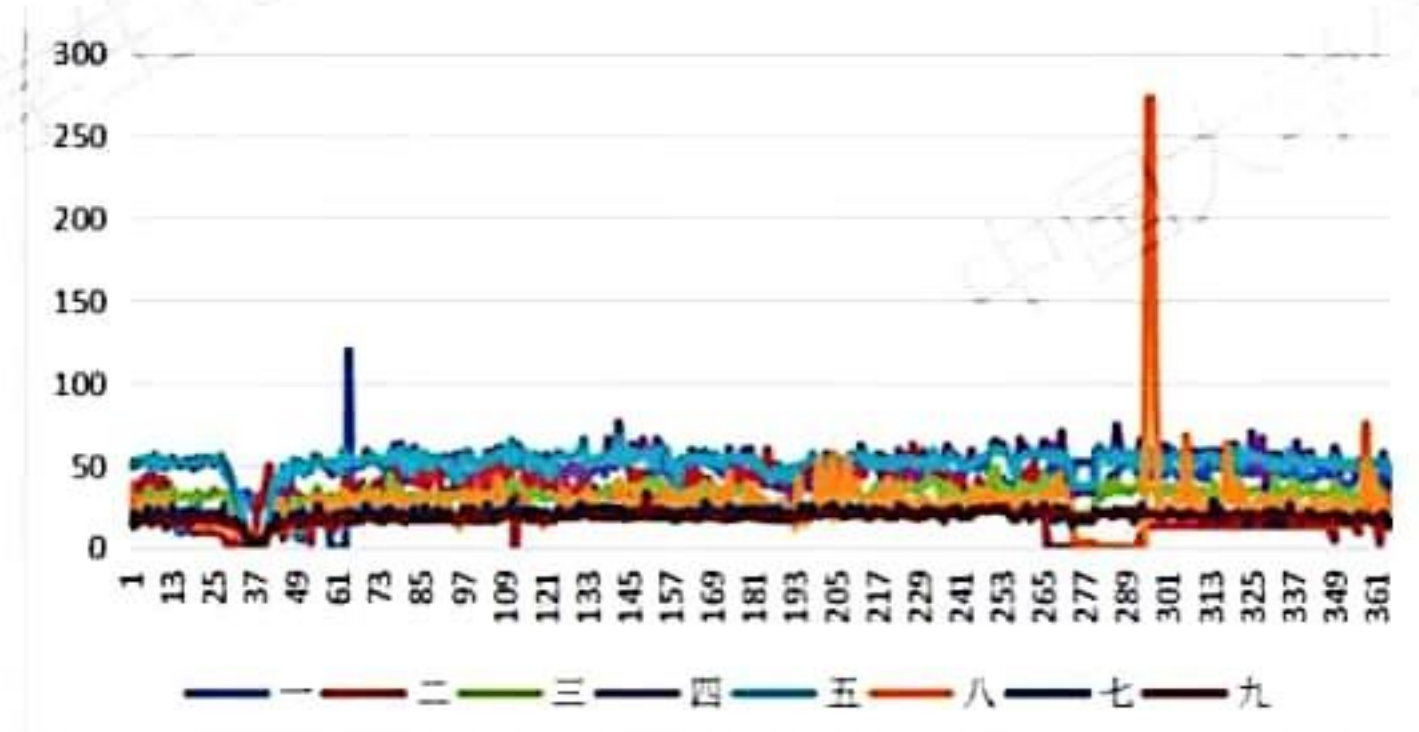


图 1、各学生宿舍楼日用水走势图

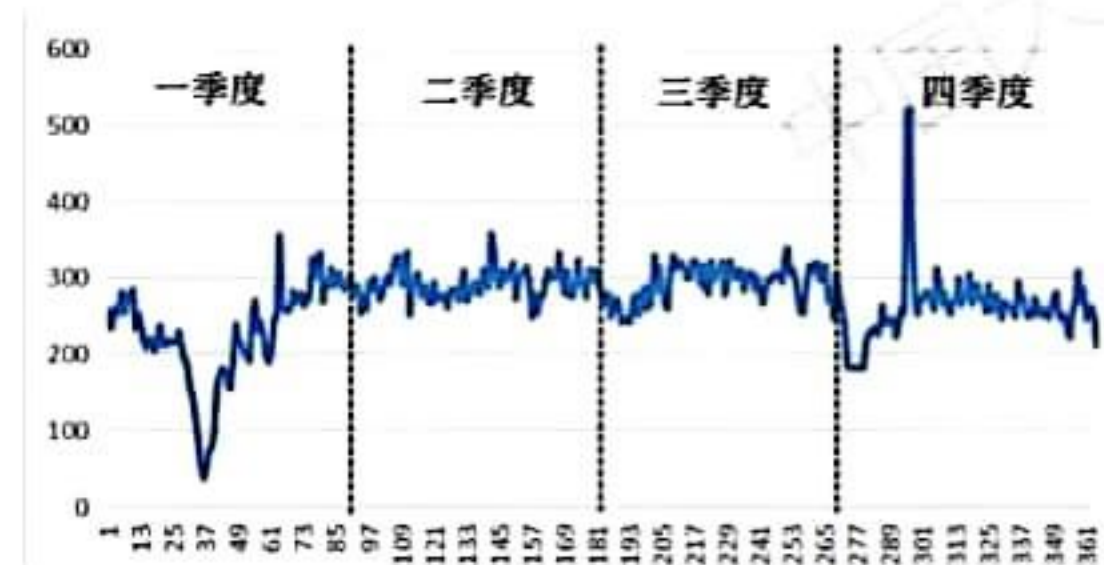


图 2、学生宿舍区日用水走势图

再作出各宿舍楼日用水量统计直方图（图 3、图 4），并初步判断和指定异常用水点（红线圈）。

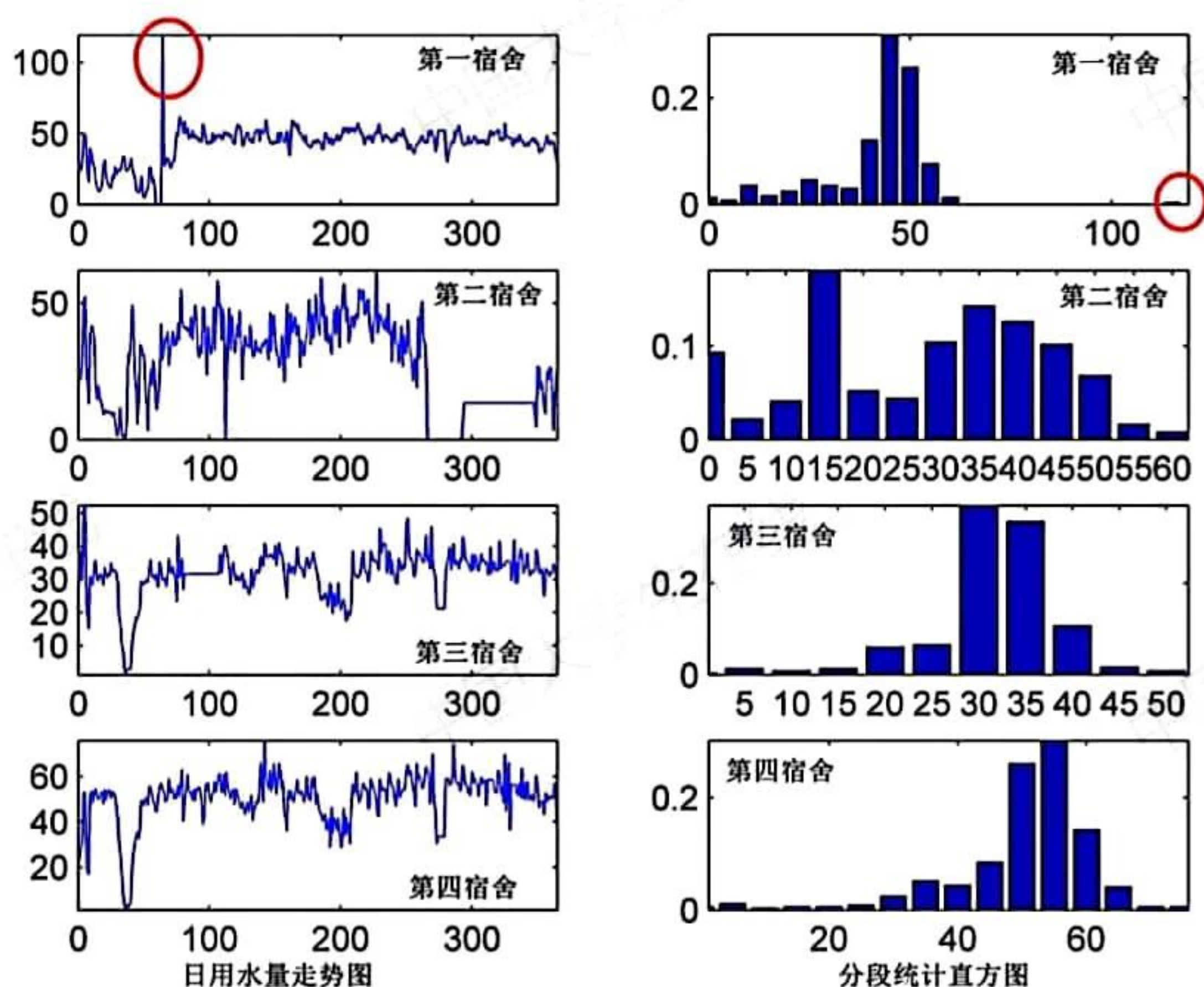


图 3、第一、二、三、四宿舍楼统计直方图

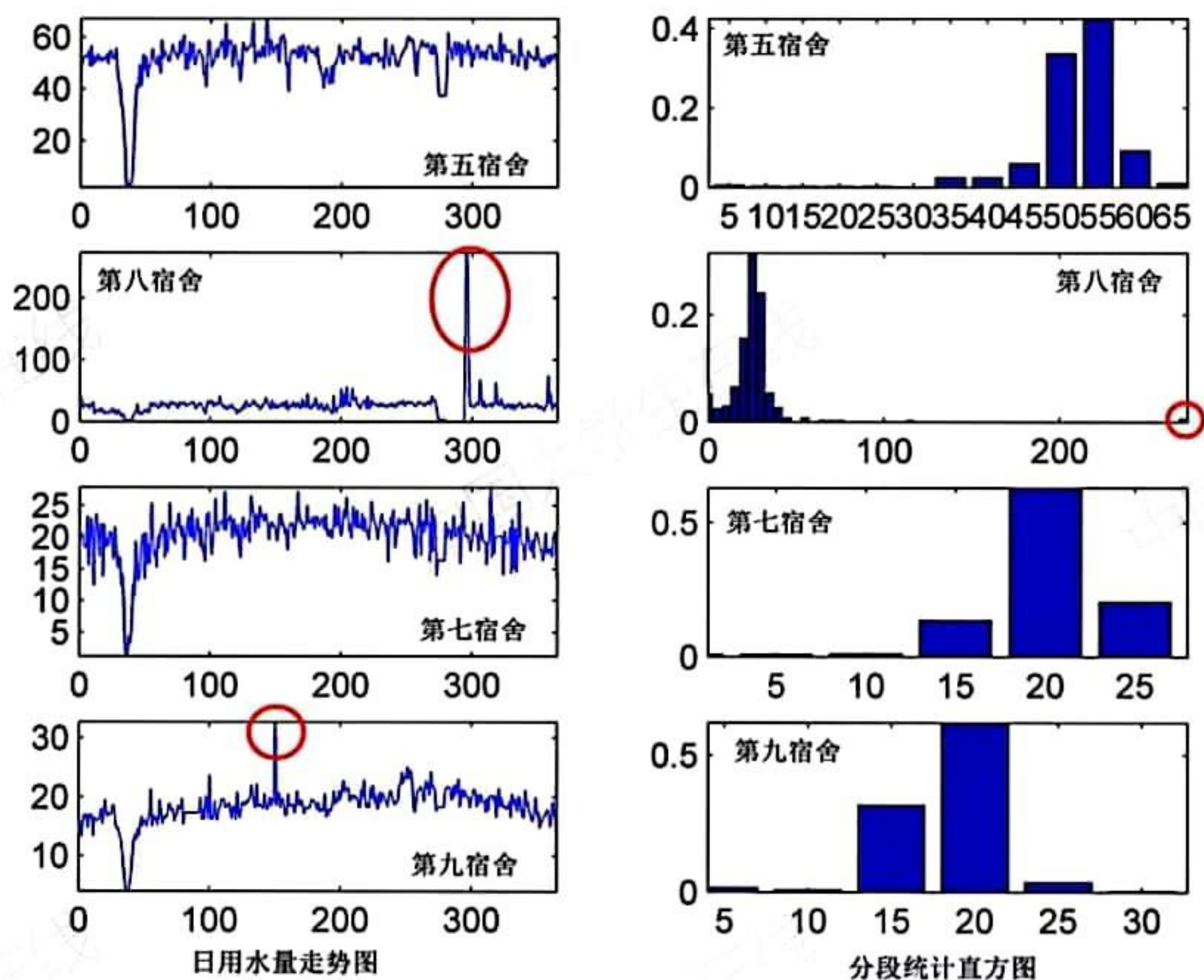


图 4、第五、七、八、九宿舍楼统计直方图

通过 5.1.1 的模型，计算出各宿舍楼的小时变化系数统计特征：

表 1、小时变化系数统计特征（学生宿舍区）

宿舍楼	样本均值 \bar{X}	样本方差 $D(x)$	样本标准差 $SD(x)$	置信区间 (95%) 自由度 $n=6$		置信区间 (95%) 自由度 $n=11$	
一	2.313	1.484	1.218	[1.034	3.592]	[1.539	3.087]
二	2.751	6.059	2.462	[0.167	5.334]	[1.187	4.315]
修正				(1.000	5.334]		
三	2.954	0.997	0.999	[1.906	4.002]	[2.319	3.588]
四	3.412	1.270	1.127	[2.229	4.595]	[2.696	4.128]
五	3.124	1.170	1.081	[1.989	4.259]	[2.437	3.811]
八	3.497	2.826	1.681	[1.732	5.261]	[2.429	4.565]
七	3.338	1.026	1.013	[2.275	4.402]	[2.695	3.982]
九	2.512	0.321	0.567	[1.917	3.106]	[2.152	2.872]

根据小时变化系数公式，可以知道该系数不能小于 1，因此，第二宿舍楼，当自由度 $n=6$ 时，在置信度为 95% 的情况下，其置信区间的左端应该是大于 1，所以将 [0.167, 5.334] 修正为 (1.000, 5.334]，从小时变化系数公式来判断，若该系数为 1，则有可能与恒定微弱漏损有关，因此该系数可以作为识别漏损的一个工具。

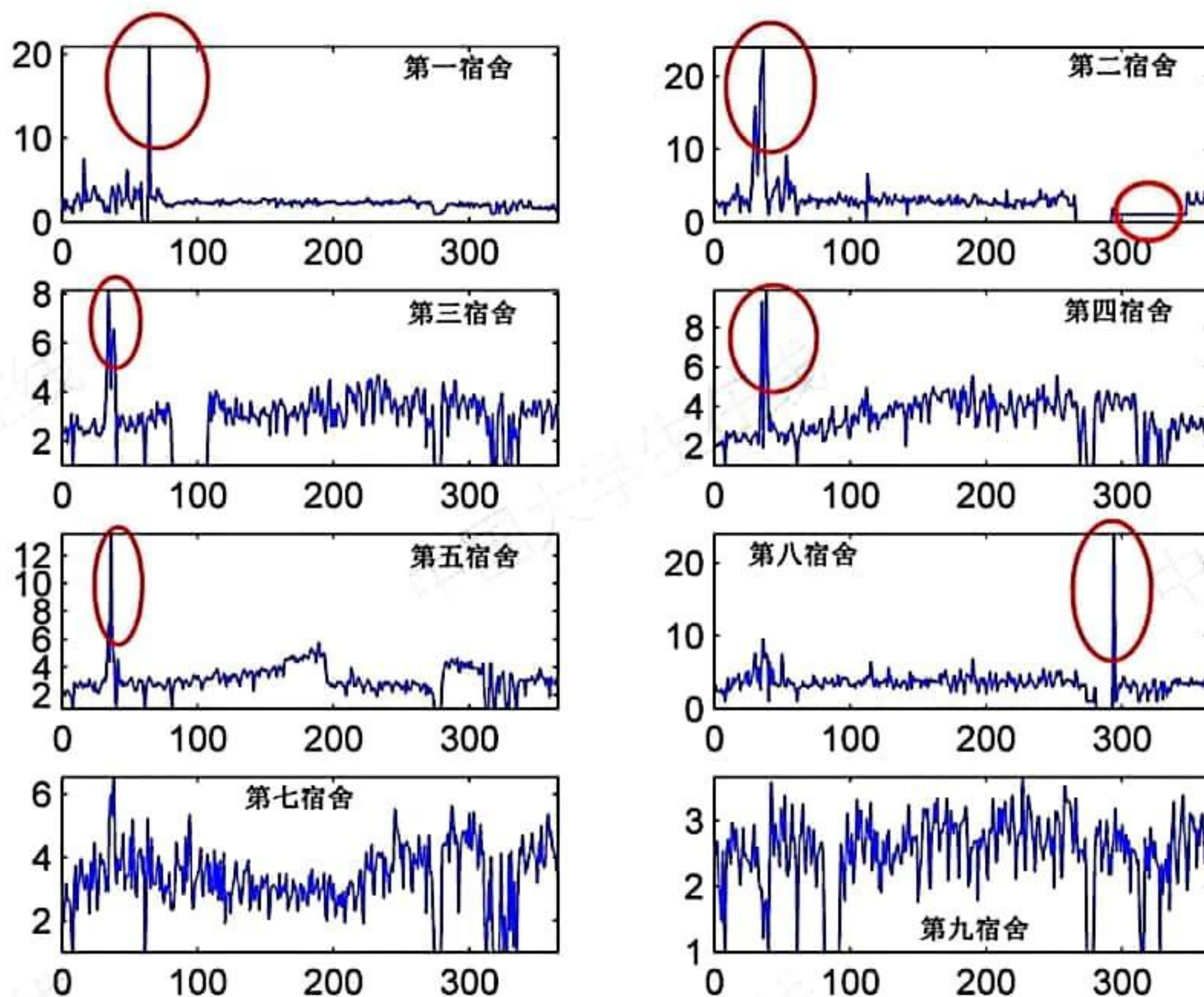


图 5、各宿舍楼小时变化系数走势图

根据以上结果可以分析得出，第一季度可能是由于学生放假的缘故，用水量均有所减少，每个宿舍楼的用水都相对稳定在一定的值，并且初步分析发现，第一、八、九宿舍楼可能出现过水管爆裂导致的漏损（日用水量过大），而第二宿舍楼可能出现过较为稳定微弱的漏损（小时变化系数一段时间内保持 1 数值不变）。

5.1.2.2 食堂区特征分析

食堂区包括：第一、二、五食堂，数据是每隔 15 分钟记录一次，现在通过 MATLAB 编程，将食堂区的数据转化为 365 天。并通过 5.1.1 的模型，计算出各食堂的日最大小时用水量、日平均小时用水量、小时变化系数统计特征：

表 2、小时变化系数统计特征（食堂区）

宿舍楼	样本均值 \bar{X}	样本方差 $D(x)$	样本标准差 $SD(x)$	置信区间 (95%) 自由度 $n=6$		置信区间 (95%) 自由度 $n=11$	
一	0.015	30.878	5.557	[-5.818	5.847]	[-3.516	3.545]
修正				(1.000	5.847]	(1.000	3.545]
五	0.019	14.915	3.862	[-4.035	4.072]	[-2.435	2.472]
修正				(1.000	4.072]	(1.000	2.472]
二	0.012	7.957	2.821	[-2.948	2.973]	[-1.780	1.805]
修正				(1.000	2.973]	(1.000	1.805]

根据小时变化系数公式，可以知道该系数不能小于 1，因此，第一、二、五食堂，当自由度 $n=6$ 和 $n=11$ 时，在置信度为 95% 的情况下，其置信区间的左端小于 1 的都修正为 1，但不取 1。

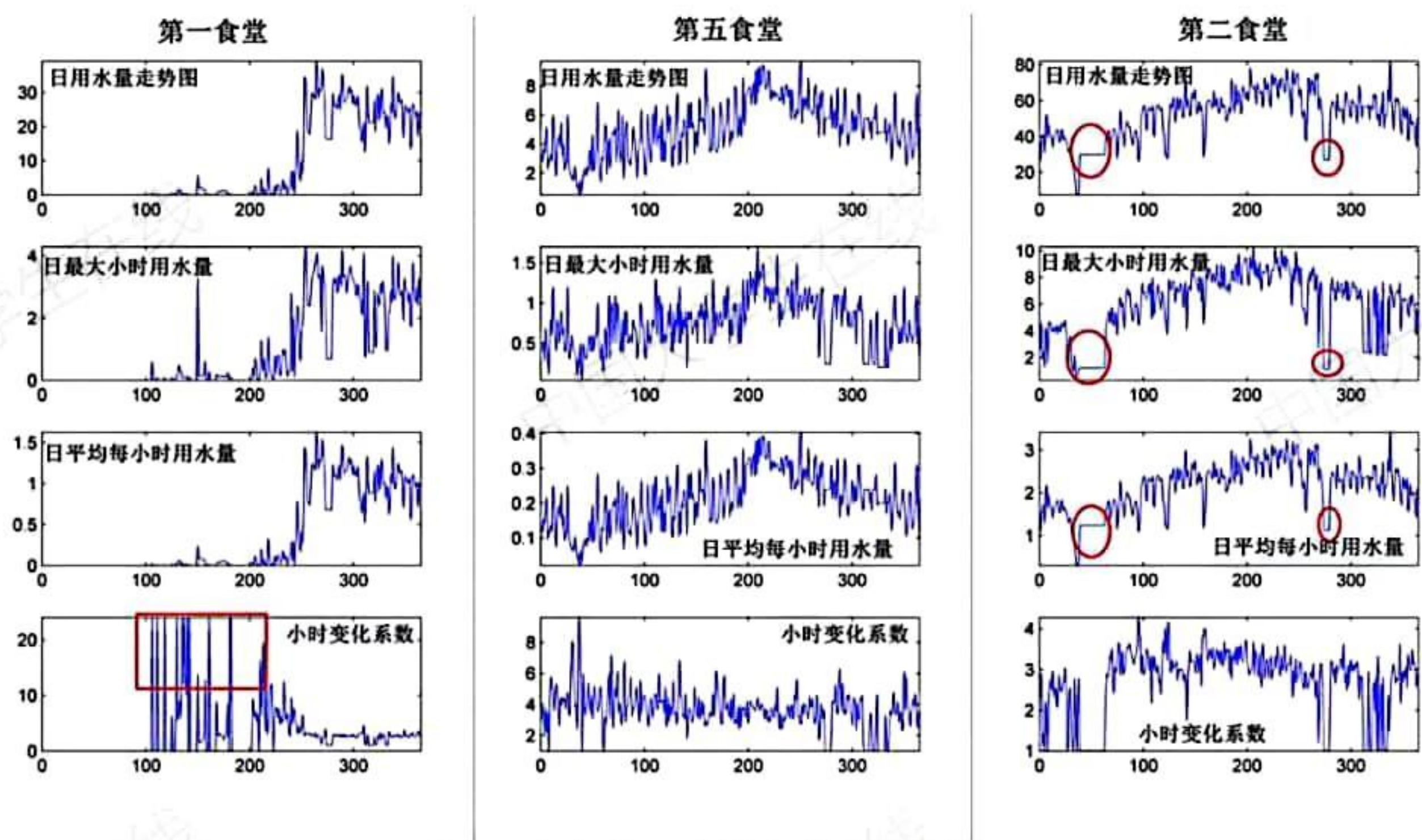


图 6、各食堂四个特征数值变化情况

从图 6 的四个特征指标的数值变化，容易获得以下结论：

表 3、四个特征指标的关系

日用水量	日最大小时用水量	日平均小时用水量	小时变化系数	初步结论
高	高	高	正常	漏损时间长 漏损较稳定 漏损强度大
高	高	高	高	漏损时间短 漏损较稳定 漏损强度大
低	低	低	高	用水集中 漏损时间短
高	高	高	高	漏损突发 漏损强度大

5.1.2.3 教学区特征分析

教学区只包括教学楼总表，数据是每隔 15 分钟记录一次，现在通过 MATLAB 编程，将教学区的数据转化为 365 天。并通过 5.1.1 的模型，计算出教学区的日最大小时用水量、日平均小时用水量、小时变化系数统计特征：

表 3、小时变化系数统计特征（教学区）

	样本均值 \bar{X}	样本方差 $D(x)$	样本标准差 $SD(x)$	置信区间（95%） 自由度 $n=6$		置信区间（95%） 自由度 $n=11$	
教学楼	2.652	1.631	1.277	[1.312	3.993]	[1.841	3.464]

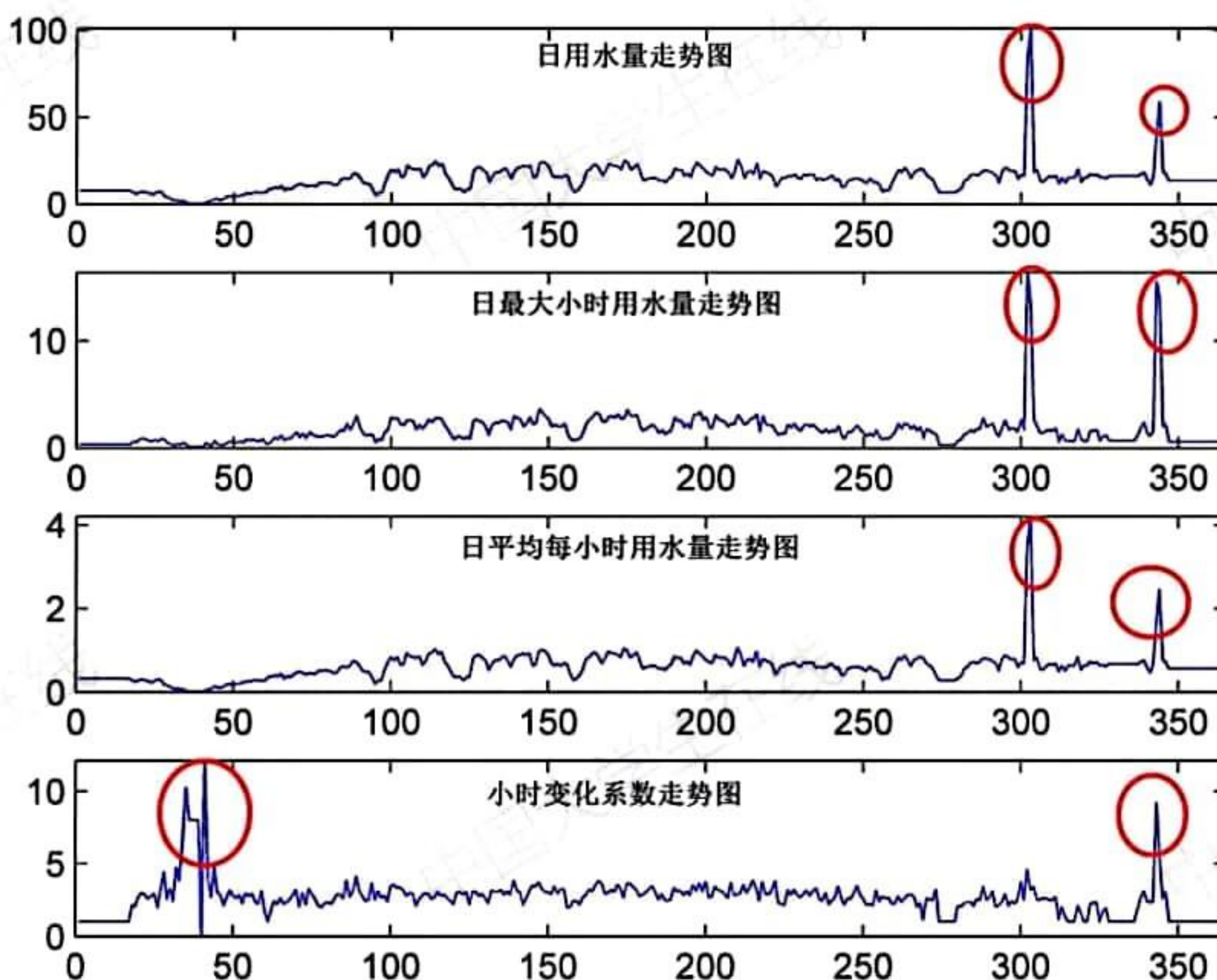


图 7、教学区四个特征数值变化情况

从图 7 中可以初步判断出，教学区在第四季度出现两次管道破损，漏损较大。其他功能区均可以用此模型来进行分析。

5.2 问题二的模型建立与求解

5.2.1 问题二的模型建立

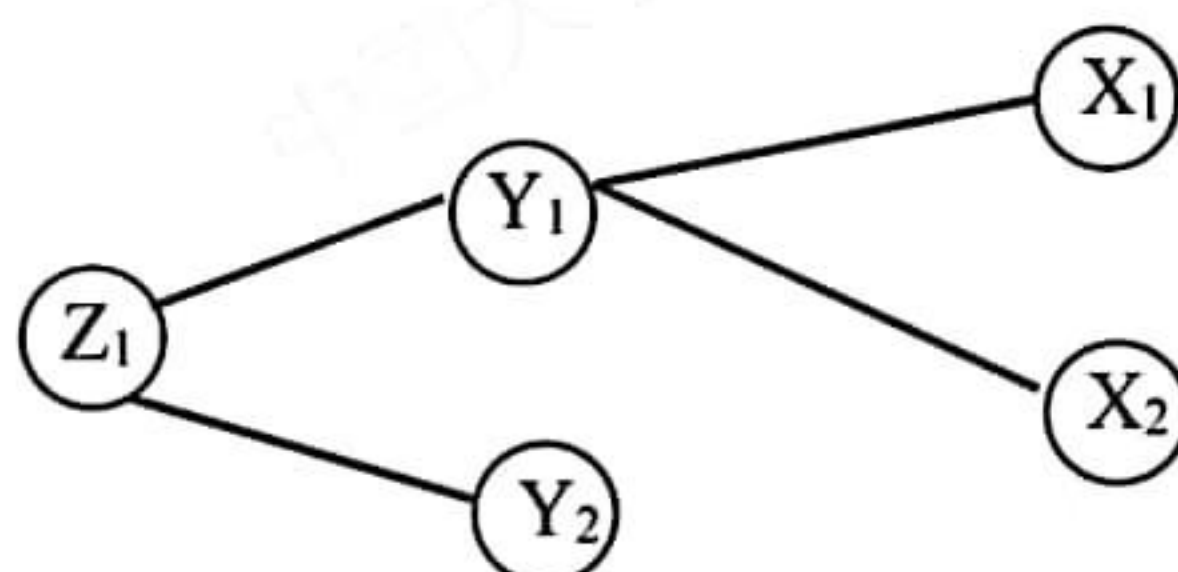


图 8、各级水表视为变量

(1) 建立多元线性回归模型

假设 $Z_1 = b_1 Y_1 + b_2 Y_2 + \xi_1$; $Y_1 = a_1 X_1 + a_2 X_2 + \xi_2$
 $Z_1 = b_1(a_1 X_1 + a_2 X_2 + \xi_2) + b_2 Y_2 + \xi_1$
 推出 $= b_1 a_1 X_1 + b_1 a_2 X_2 + b_2 Y_2 + b_1 \xi_2 + \xi_1$
 $= c_1 X_1 + c_2 X_2 + c_3 Y_2 + \xi_3$
 建立 $Y_i = \beta + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \xi_i (i=1,2,3,\dots)$ (6)
 X_1, \dots, X_k 是随机自变量; Y 是随机因变量;
 β_0, \dots, β_k 是回归系数; ξ 是随机误差项。
 注: 推导结果说明, 用各末端水表数据就可实现回归模型。

(2) 回归方程的显著性检验

原假设 $H_0: \beta_1 = \dots = \beta_k$, 备择假设 $H_1: \beta_1, \dots, \beta_k$ 不全为 0, 当假设成立时, 检验统计量:

$$F = \frac{SSR/k}{SSE/(n-k-1)} \sim F(k, n-k-1) \quad (7)$$

回归平方和: $SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$

残差平方和: $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$

对于给定的显著水平 α , 检验的拒绝域 $F > F_\alpha(k, n-k-1)$ 。

(3) 回归系数的显著性检验

原假设 $H_0: \beta_j = 0$, 备择假设 $H_1: \beta_j \neq 0 (j=0,1,\dots,k)$, 当原假设成立时, 检验统计量:

$$F_j = \frac{SSE_j - SSE}{SSE/(n-k-1)} \sim F(1, n-k-1) \quad (8)$$

掉 x_j 后残差平方和: SSE_j

对于给定的显著水平 α , 检验的拒绝域为 $F_j > F_\alpha(1, n-k-1)$ 。

也可以用检验统计量:

$$t_j = \frac{\hat{\beta}_j}{\hat{\sigma} \sqrt{c_{jj}}} \sim t(n-k-1) \quad (9)$$

c_{jj} 是 $c = (x^T x)^{-1}$ 对角线 $j = (j=0,1,\dots,k)$ 上第 j 个元素。对于给定的显著性水平 α , 检验的拒绝域为 $|t_j| > t_{\alpha/2}(n-k-1)$ 。

5.2.2 问题二的模型求解

选取 403X 子网供水系统的数据, 该系统包括 11 个末端水表, 所以建立 11 个随机自变量的多元线性回归模型。将数据先处理为每日用水量, 然后选取 40 个数据作为回归系数的计算依据, 后 40 个数据作为对比数据, 以进一步评估模型的优劣。

通过 MATLAB 计算得到以下结果：

表 4、回归系数值及统计特征

回归系数	回归系数估计值	回归系数置信区间	
β_0	-105.4976	[-346.3217	135.3265]
β_1	12.9912	[-63.8778	89.8602]
β_2	-6.1822	[-15.4588	3.0944]
β_3	17.3709	[-84.7264	119.4682]
β_4	8.7463	[-18.8519	36.3446]
β_5	-14.6269	[-95.8830	66.6292]
β_6	3.2512	[1.8540	4.6484]
β_7	-0.0140	[-1.2339	1.2060]
β_8	0.0075	[-0.2081	0.2231]
β_9	5.1333	[1.1912	9.0753]
β_{10}	1.2123	[-2.0336	4.4583]
β_{11}	-0.2879	[-1.2887	0.7130]
β_{12}	0.6335	[-2.7854	4.0523]
$R^2=0.8128$		$F=5.0659$	$p>0.0001$ $s^2=482.8741$

从上表 4 可见，拟合效果还比较理想，但还不够完美，此时再进一步完善，将异常数据剔除，再进行回归拟合。

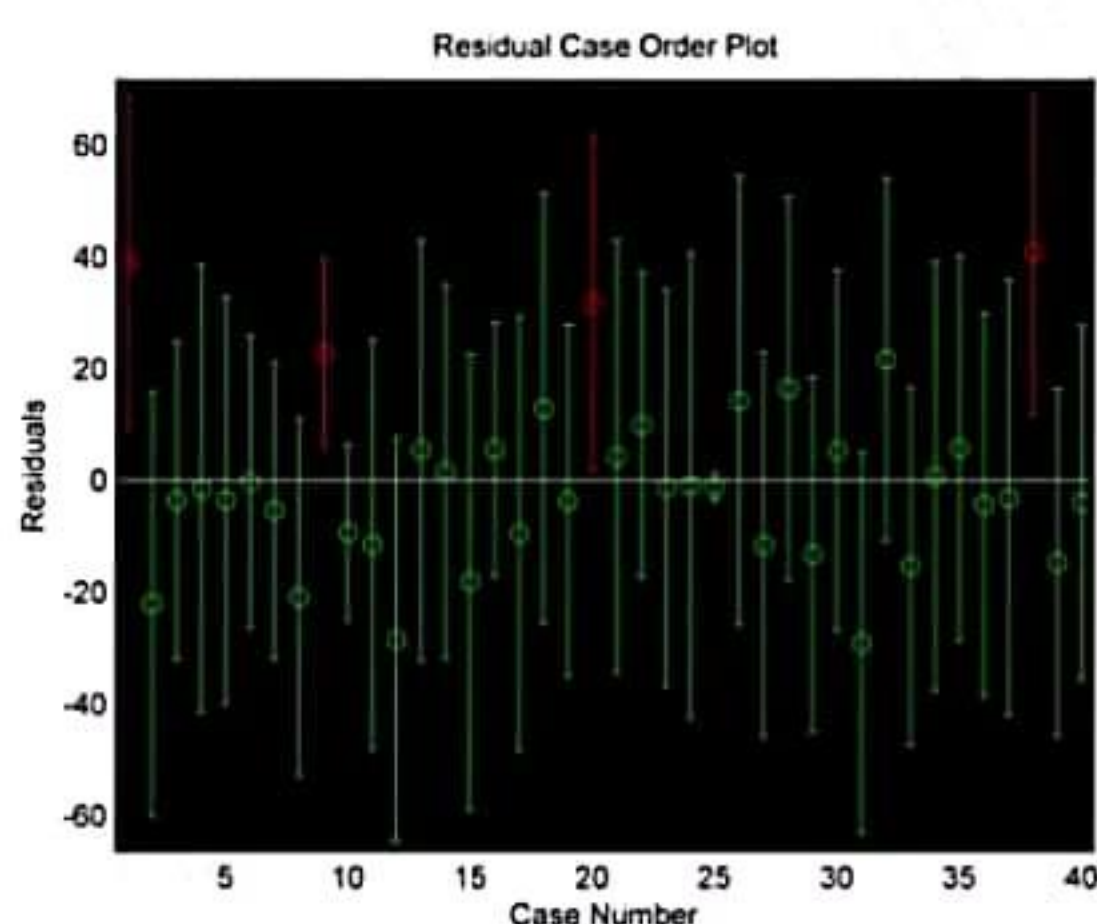


图 9、原始数据

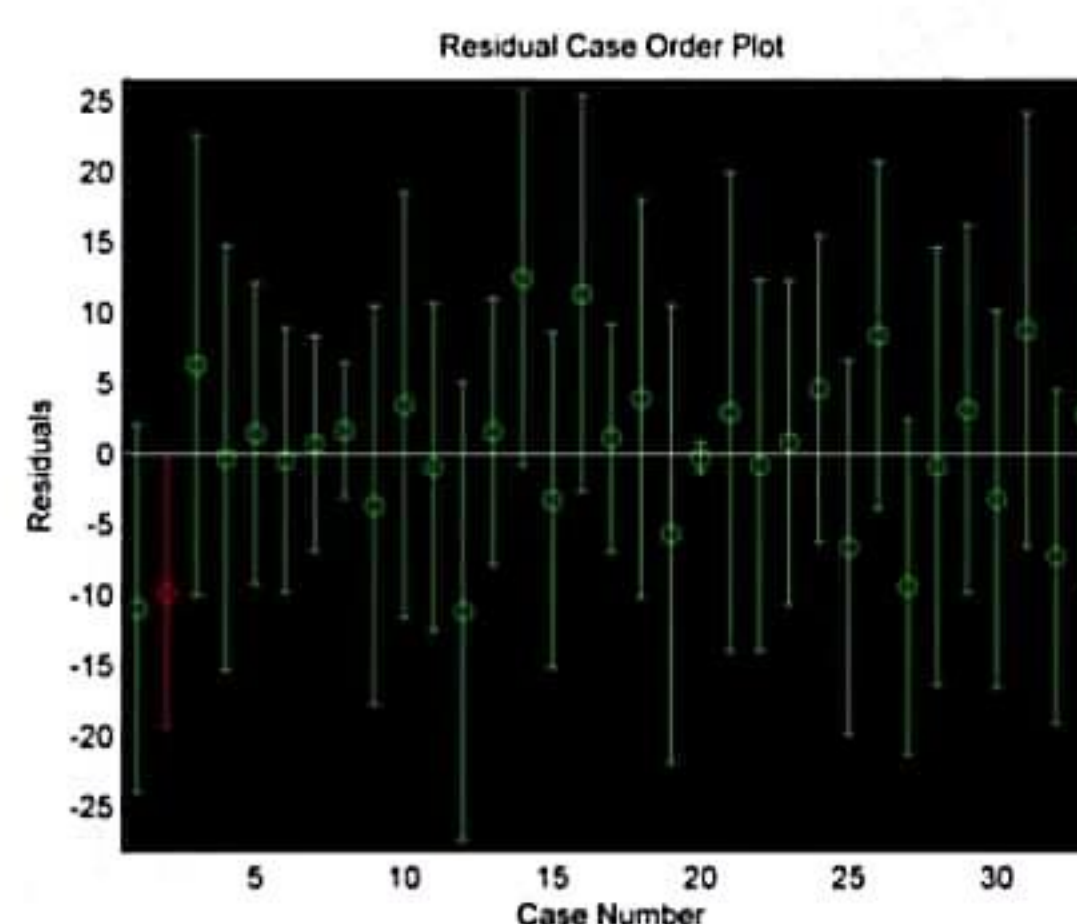


图 10、两次剔除异常数据后

通过两次剔除异常数据后（第一次剔除 1、9、20、38。第二次剔除 10、16、25），再进行多元回归分析，计算得到以下结果：

表 5、回归系数值及统计特征（剔除异常点）

回归系数	回归系数估计值	回归系数置信区间	
β_0	-52.9021	[-190.0877	84.2836]
β_1	46.2210	[9.2407	83.2013]

β_2	-3.7585	[-8.4207	0.9037]
β_3	24.4646	[-27.6850	76.6141]
β_4	14.1476	[0.0334	28.2618]
β_5	10.2162	[-31.8248	52.2572]
β_6	3.0964	[2.4409	3.7519]
β_7	-0.6418	[-1.7949	0.5112]
β_8	0.1085	[0.0100	0.2070]
β_9	3.6908	[1.1379	6.2436]
β_{10}	3.7356	[1.9567	5.5146]
β_{11}	-0.5347	[-1.0398	-0.0296]
β_{12}	-0.9017	[-2.5775	0.7741]
$R^2=0.9687 \quad F=24.109 \quad p<0.0001 \quad s^2=82.2312$			

从表 5 可见，此时的 $R^2=0.9687$, $p<0.0001$ ，十分显著，说明拟合效果非常好。现在用该回归模型预测后 40 个数据，观察吻合度。

回归模型如下（系数对照表 5）：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} \quad (10)$$

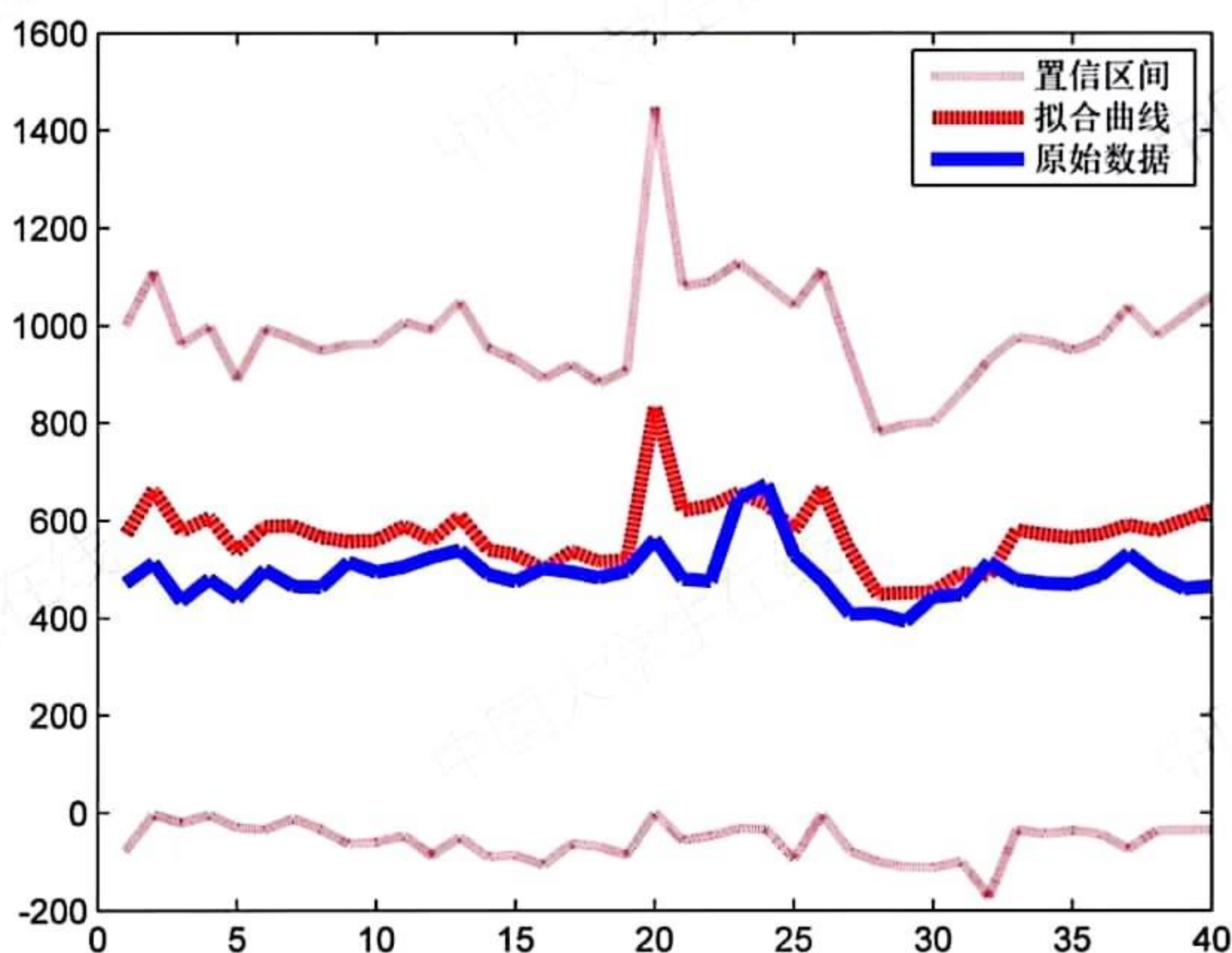


图 11、预测值与实际值比较

从图 11 上看，预测值变化趋势与实际值变化趋势基本吻合，但是细心发现，预测值稍微偏高，因此可以继续将模型优化，通过最小二乘法，获得最优拟合模型。通过计算，在原来回归模型（10）的式子右端减去 82 即可。

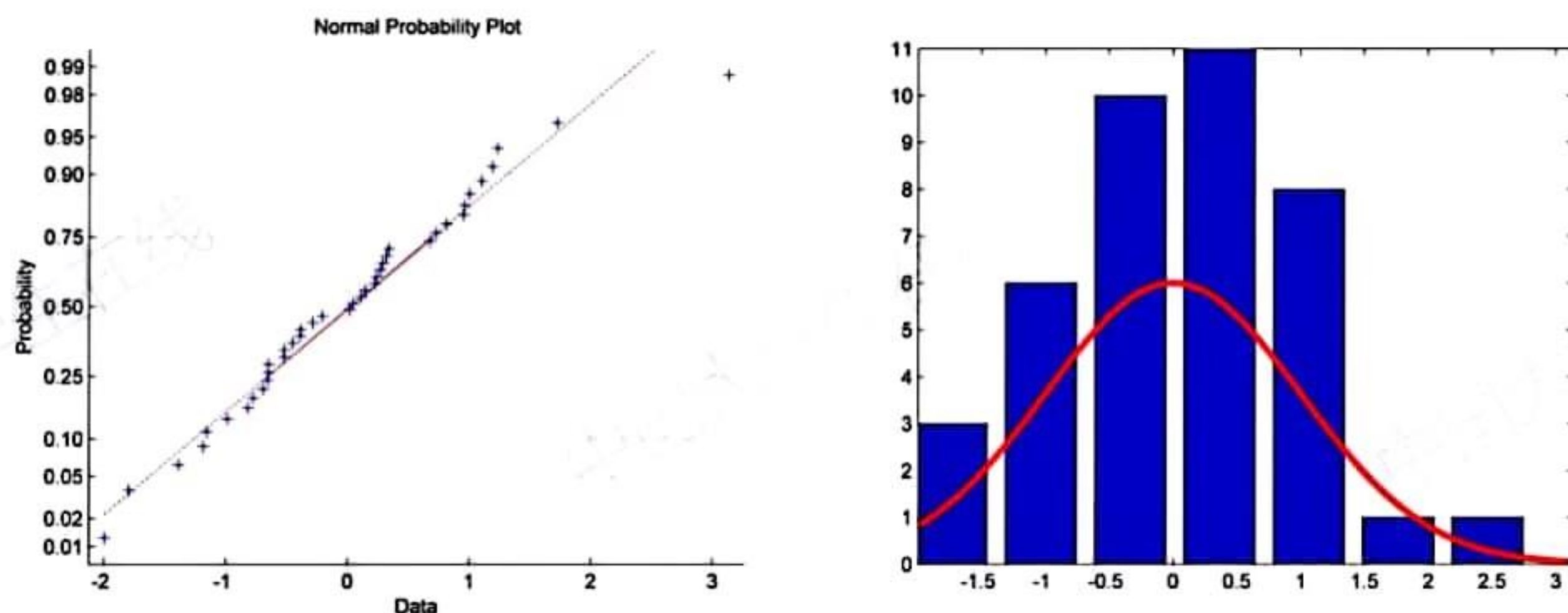


图 12、残差正态性检验

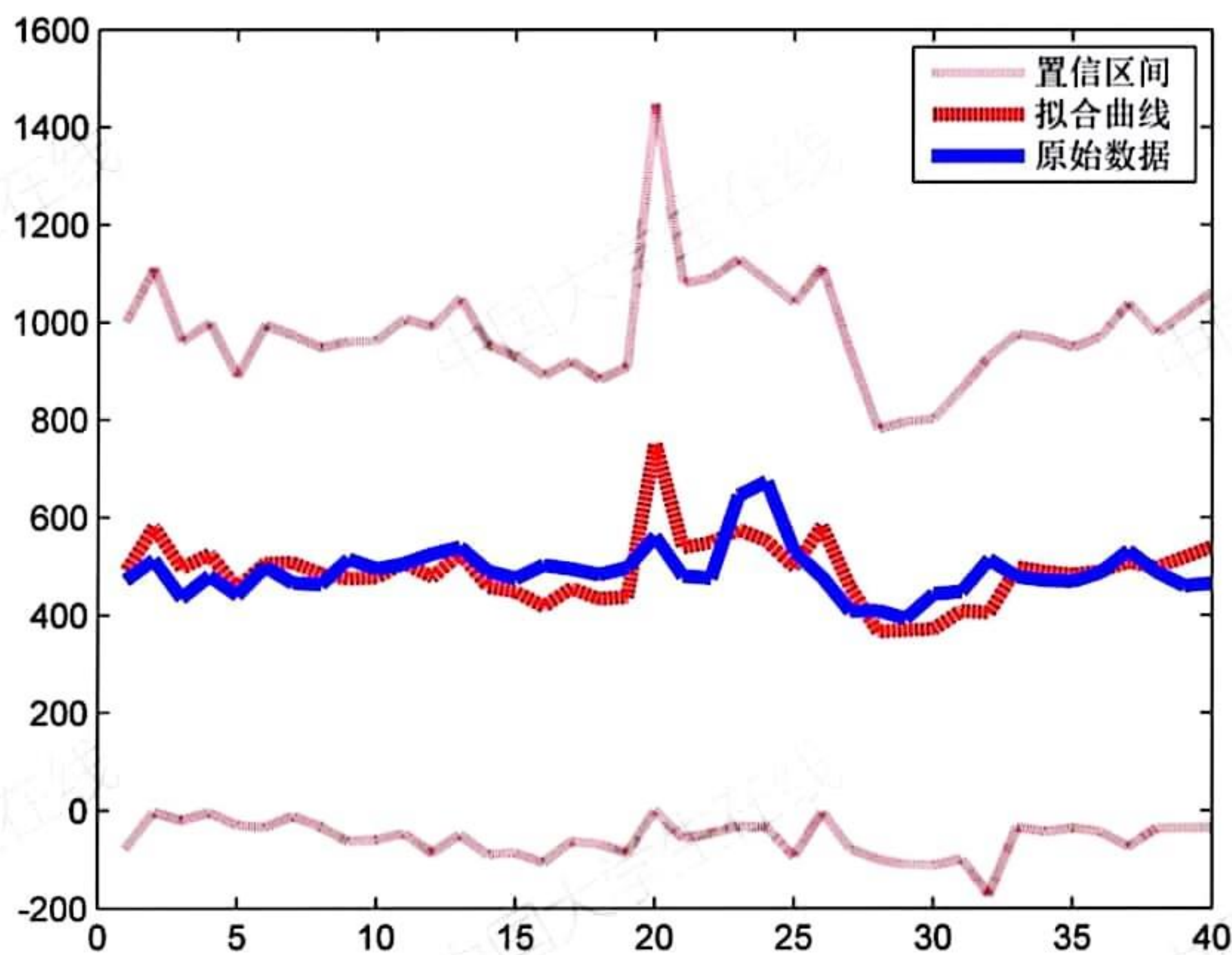


图 13、修正后的回归模型拟合效果

从上图 12、图 13 中的拟合图像，以及残差分析，可知该回归模型是最优模型。

5.3 问题三的模型建立与求解

5.3.1 问题三的模型建立

(1) 建立三倍标准差剔除异常值模型

在正态分布中 σ 代表标准差, μ 代表均值 $X = \mu$ 即为图像的对称轴 3σ 原则。

$$\bar{X} = \frac{\sum X_i}{n} \quad (11)$$

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}} \quad (12)$$

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (13)$$

正常值为 $(\mu - 3\sigma, \mu + 3\sigma)$ 内的数值, 超过者为异常值。

变量取值在区间 $[(\mu - \sigma, \mu + \sigma)]$ 之间的概率: $P(\mu - \sigma \leq \xi \leq \mu + \sigma) = 0.6827$

变量取值在区间 $[(\mu - 2\sigma, \mu + 2\sigma)]$ 之间的概率: $P(\mu - 2\sigma \leq \xi \leq \mu + 2\sigma) = 0.9545$

变量取值在区间 $[(\mu - 3\sigma, \mu + 3\sigma)]$ 之间的概率: $P(\mu - 3\sigma \leq \xi \leq \mu + 3\sigma) = 0.9973$

在本题中, 为了更好的将异常数据剔除 (即剔除异常用水点), 采用两倍标准差剔除数据。

5.3.2 问题三的模型求解

考察学校整体的漏损情况, 需要对 13 个一级表数据进行分析, 由于 419T-406T 没有分表, 所以可以将他们合并为一个一级表, 从而精简为 7 个一级表。现在分别对 7 个指标进行求解。

表 6、第一季度各子供水网漏损情况 (2 σ 标准下)

子网系统	416X	405X	404X	403X	402X	401X	419T-406T	合计
总用水量	5479.31	7047.71	4213.96	33809.49	7632.71	8453.75	3486.71	70123.64
漏损水量	0.00	1914.45	0.00	1088.23	419.53	30.19	0.00	3452.40
漏损比率	0.00%	27.16%	0.00%	3.22%	5.50%	0.36%	0.00%	4.92%

表 7、第二季度各子供水网漏损情况 (2 σ 标准下)

子网系统	416X	405X	404X	403X	402X	401X	419T-406T	合计
总用水量	5310.92	13514.08	5106.01	45385.75	11336.60	12909.83	6457.12	100020.31
漏损水量	11.68	348.50	0.00	4392.30	972.71	84.74	1203.72	7013.65
漏损比率	0.22%	2.58%	0.00%	9.68%	8.58%	0.66%	18.64%	7.01%

表 7、第三季度各子供水网漏损情况 (2 σ 标准下)

子网系统	416X	405X	404X	403X	402X	401X	419T-406T	合计
总用水量	6908.49	13827.18	5874.30	64573.18	27114.24	14167.05	7755.57	140220.01

漏损水量	0.00	213.86	0.00	5649.64	788.53	108.82	0.00	6760.85
漏损比率	0.00%	1.55%	0.00%	8.75%	2.91%	0.77%	0.00%	4.82%

表 8、第四季度各子供水网漏损情况（2 σ 标准下）

子网系统	416X	405X	404X	403X	402X	401X	419T-406T	合计
总用水量	6446.68	12950.17	5355.32	42220.50	7726.54	15090.33	5318.02	95107.56
漏损水量	0.00	435.48	0.00	3667.93	0.00	218.17	0.00	4321.58
漏损比率	0.00%	3.36%	0.00%	8.69%	0.00%	1.45%	0.00%	4.54%

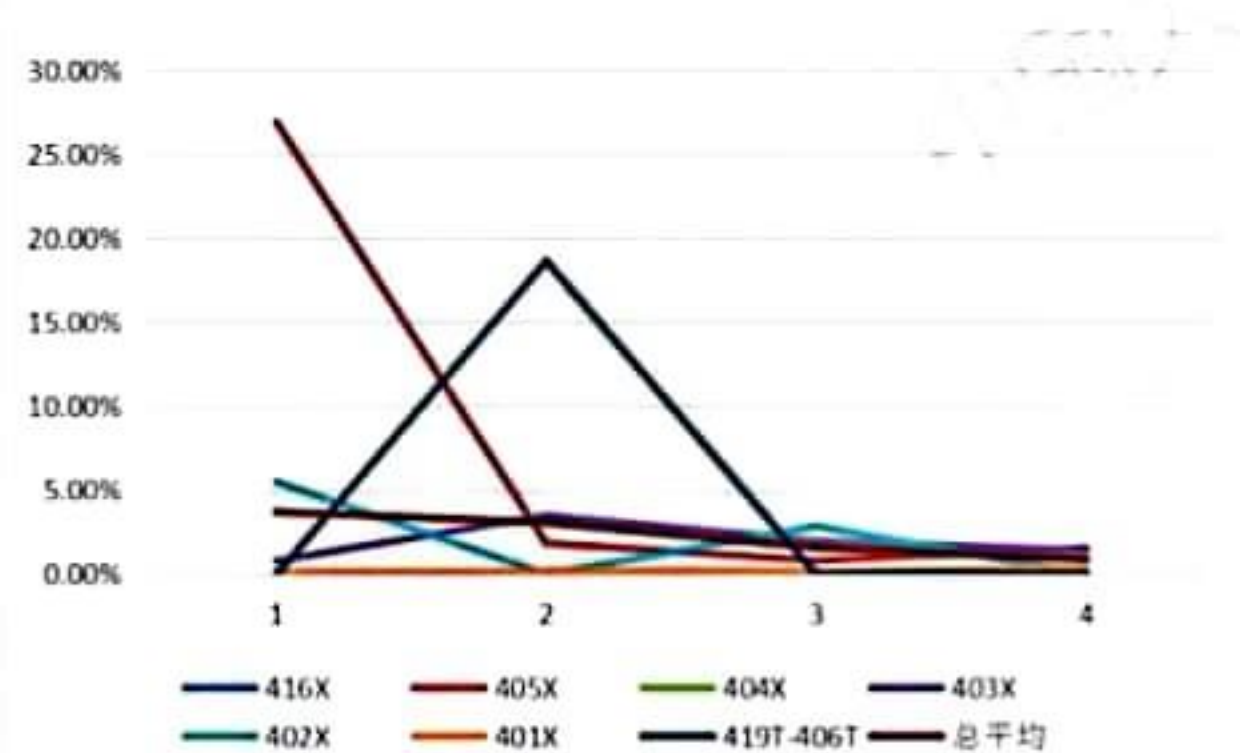
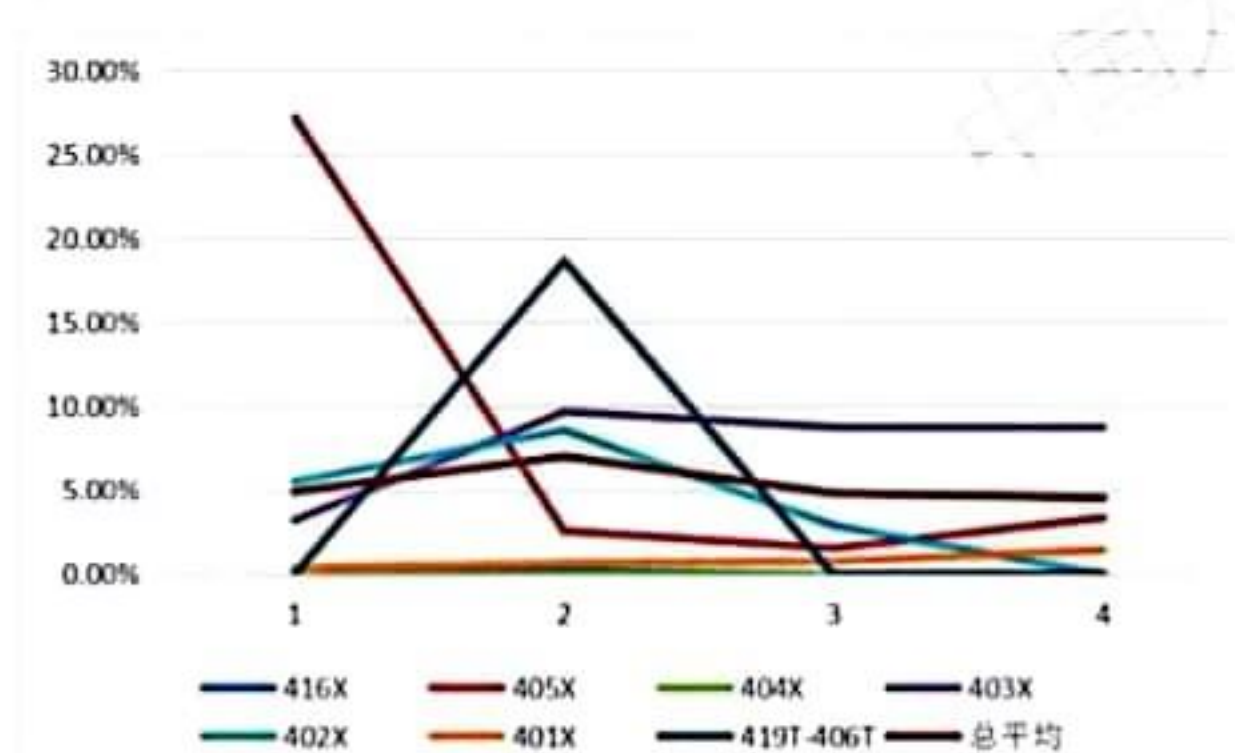


图 14、各子网漏损比率随季度走势图（2 σ ） 图 15、各子网漏损比率随季度走势图（3 σ ）

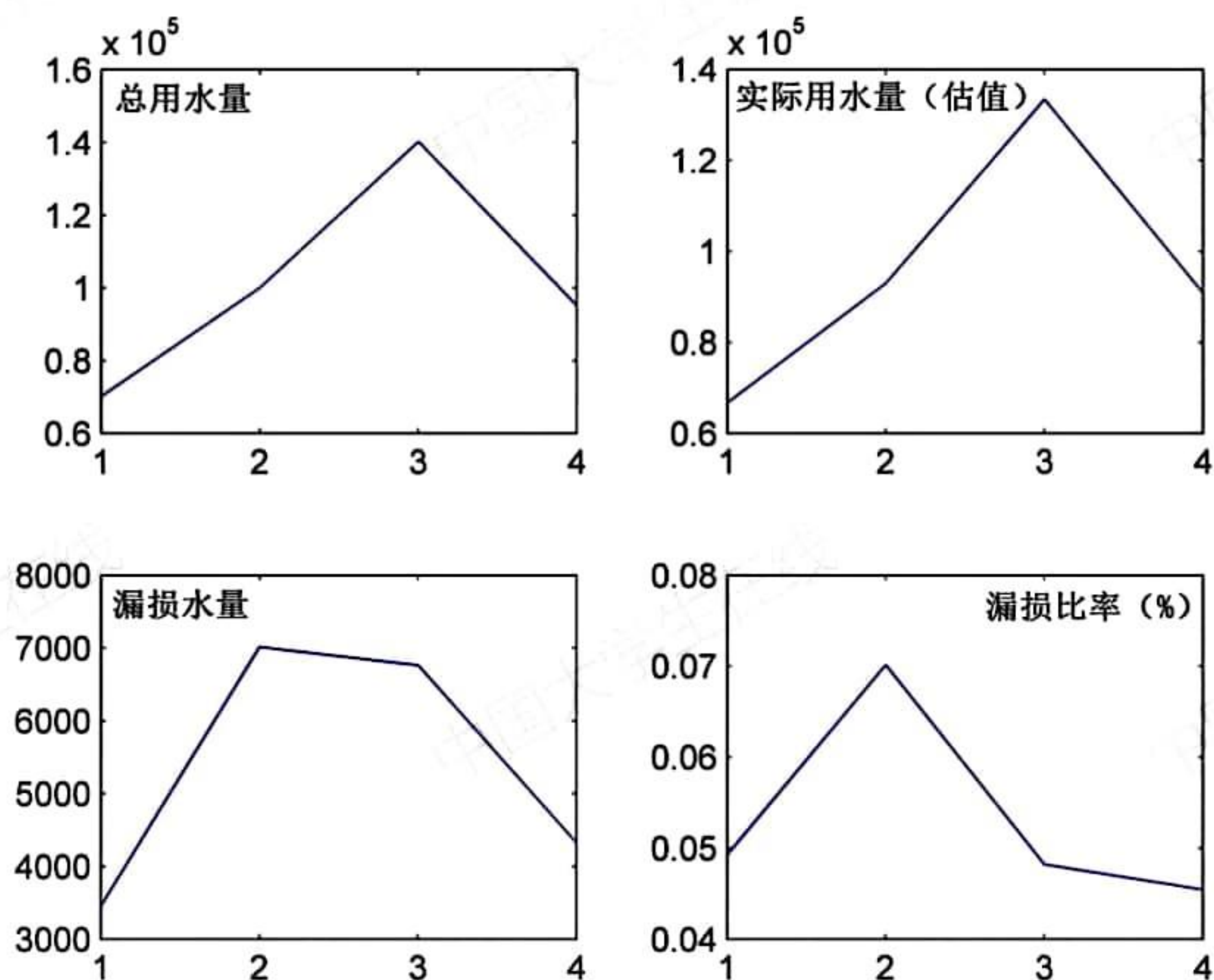


图 16、该校区漏损随季度变化情况

通过计算得到以上结果，并将 7 个指标漏损比率和平均漏损比率归纳到下表中：

表 9、各子系统漏损情况

一级表 \ 季度	一季度	二季度	三季度	四季度
416X	0.00%	0.22%	0.00%	0.00%
405X	27.16%	2.58%	1.55%	3.36%
404X	0.00%	0.00%	0.00%	0.00%
403X	3.22%	9.68%	8.75%	8.69%
402X	5.50%	8.58%	2.91%	0.00%
401X	0.36%	0.66%	0.77%	1.45%
419T-406T	0.00%	18.64%	0.00%	0.00%
总平均	4.92%	7.01%	4.82%	4.54%

整个学校随季度的漏损比率为：4.92%、7.01%、4.82%、4.54%。第二季度较高，其原因是 419T-406T 出现了较大的漏损，导致第二季度较高。但总体来说都在 5% 左右，与公共用水平均漏损率相当，说明，用该模型求解的正确性和可靠性。

5.4 问题四的模型建立与求解

5.4.1 问题四的模型建立

(1) 建立 BP 神经网络模型^[1]

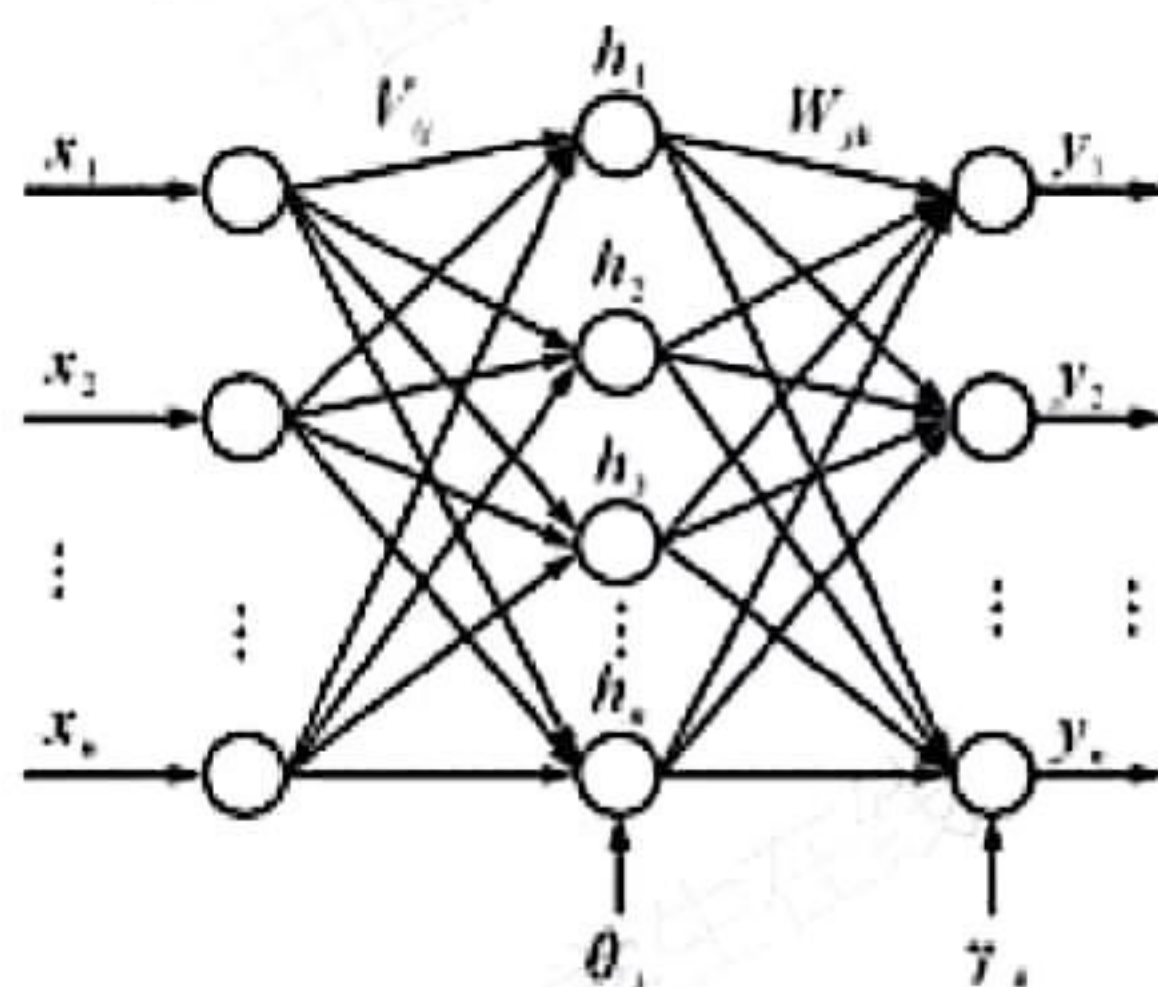


图 17、BP 神经网络结构

标准的 BP 神经网络算法内容和步骤描述如下，先定义一下变量和自变量：

输入层向量： $X = (x_1, x_2, \dots, x_i, \dots, x_n)$;

隐含层输入向量： $H = (h_1, h_2, \dots, h_j, \dots, h_m)$;

输出层输出向量： $Y = (y_1, y_2, \dots, y_k, \dots, y_l)$;

期望值输出向量： $D = (d_1, d_2, \dots, d_k, \dots, d_l)$;

输入层到隐含层之间的权值连接矩阵： $V = (V_1, V_2, \dots, V_j, \dots, V_m)$;

隐含层到输出层之间的权值连接矩阵： $W = (W_1, W_2, \dots, W_k, \dots, W_l)$ 。

BP 神经网络具体实现步骤如下：

网络初始化 W, V 矩阵，赋值期间由激活函数值域决定。确定最大训练次数 M 和学习精度值 ε ，选择激活函数 $f(x)$ ，通常选用单极限 sigmoid 函数：

$$f(x) = 1 / (1 + e^{-x}) \quad (14)$$

数据预处理，选择样本数据输入，得到隐含层 h_j 和输出层 y_k 的输出：

$$h_j = f(T_j^T X^T), j = 1, 2, 3, \dots, m \quad (15)$$

$$y_k = f(W_k^T H^T), k = 1, 2, 3, \dots, l \quad (16)$$

利用网络的实际输出值 y_k 和期望输出值 d_k 计算误差：

$$e = 1/2 \sum_{k=1}^l (d_k - y_k)^2 \quad (17)$$

分别计算误差函数对隐含层和输出层个神经元的偏导数 δ_j^h 和 δ_k^y ：

$$\delta_j^h = (\sum_{k=1}^l \delta_k^y W_{jk}) * h_j * (1 - h_j), j = 1, 2, 3, \dots \quad (18)$$

$$\delta_k^y = (d_k - y_k) * y_k * (1 - y_k), k = 1, 2, 3, \dots \quad (19)$$

利用误差信号调整各层的连接权值；隐含层到输出层权值 w_{jk}^{N+1} 和输入层到隐含层权值 v_{ij}^{N+1} ：

$$w_{jk}^{N+1} = w_{jk}^N + \varphi \delta_k^y h_j, k = 1, 2, \dots, l \quad j = 0.1, \dots, m \quad (20)$$

$$v_{ij}^{N+1} = v_{ij}^N + \varphi \delta_j^h x_i, j = 1, 2, \dots, m \quad i = 0.1, \dots, n \quad (21)$$

计算全局误差 E ：

$$E = 1/2 \sum_{p=1}^P \sum_{k=1}^l (d_k - y_k)^2 \quad (22)$$

5.4.2 问题四的模型求解

由于供水系统具有的层级网络特征，再加之神经网络的“万能逼近”，对于这类数据，有较好的预测效果，以预测值与实际值的误差作为识别漏损的依据，若是误差过大，那么就说明有漏损情况。由于全网涉及的水表太多，篇幅有限，现只展示四个被识别出来的漏损情况：

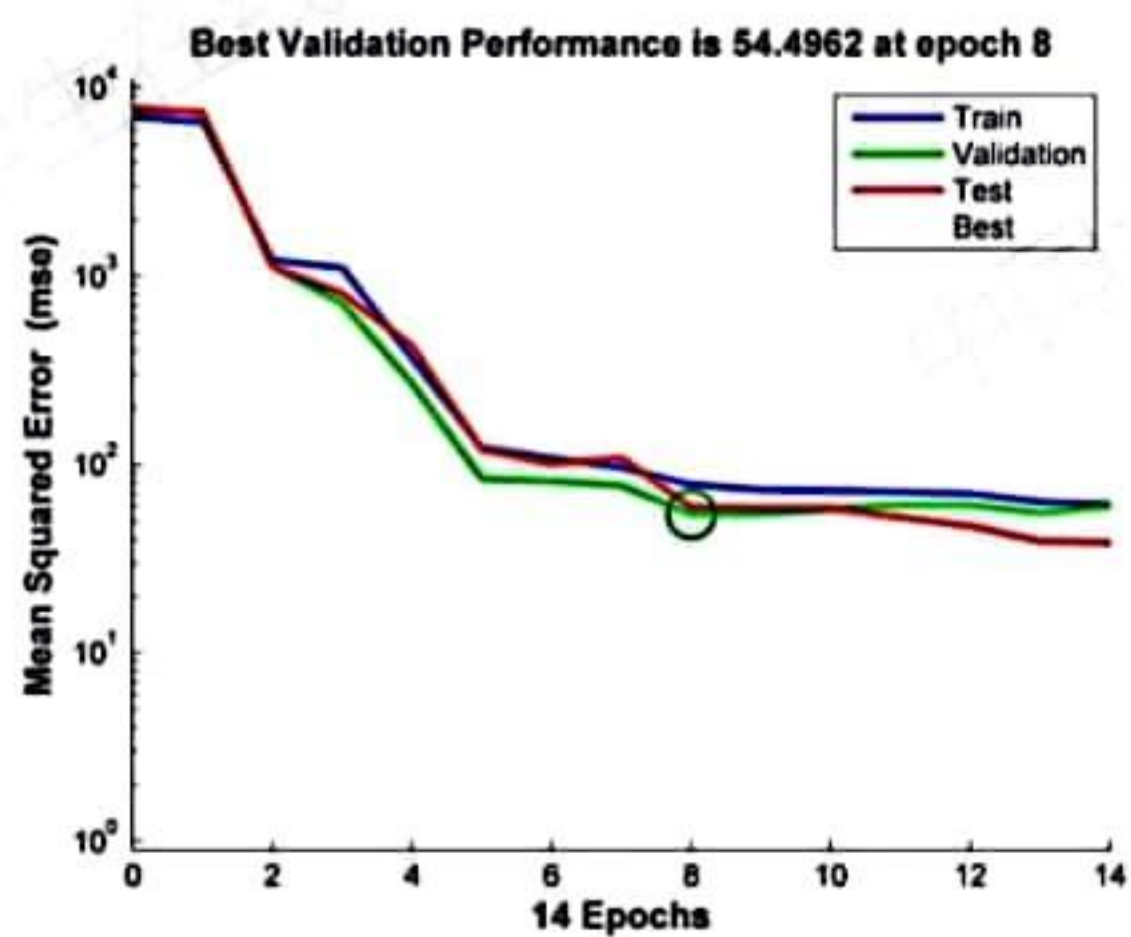
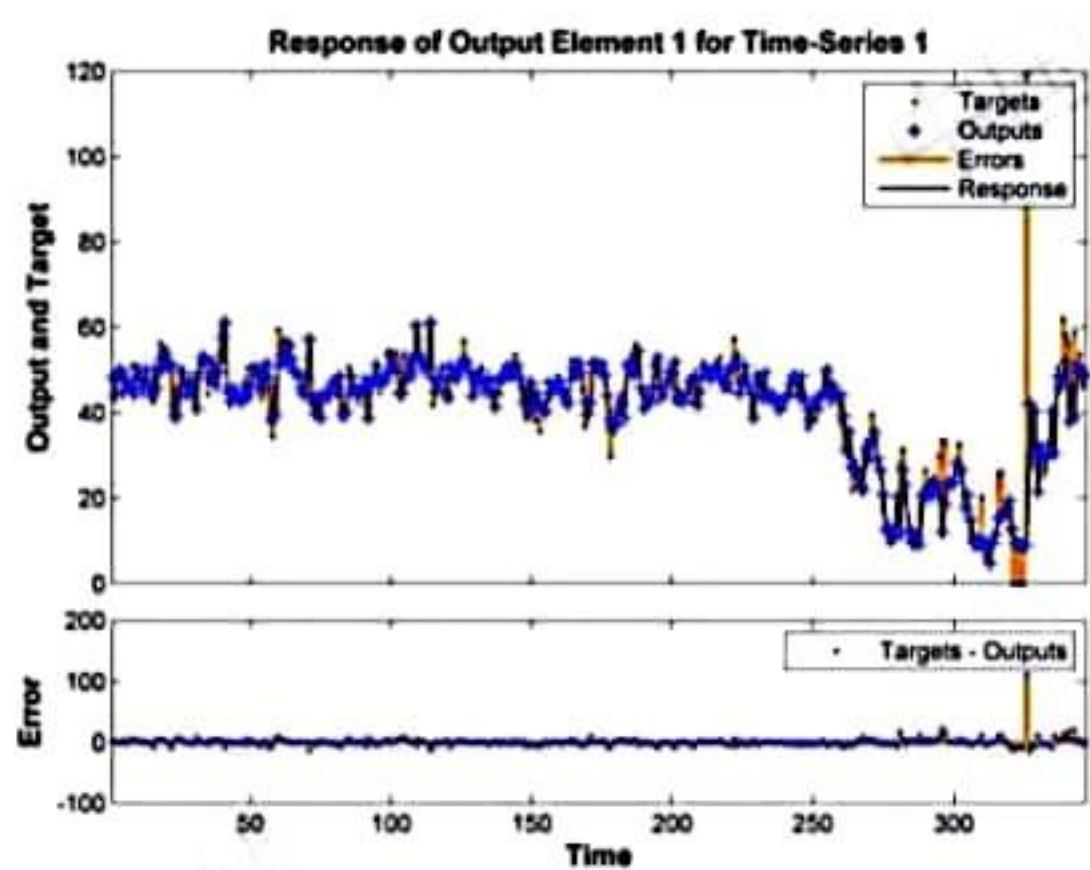
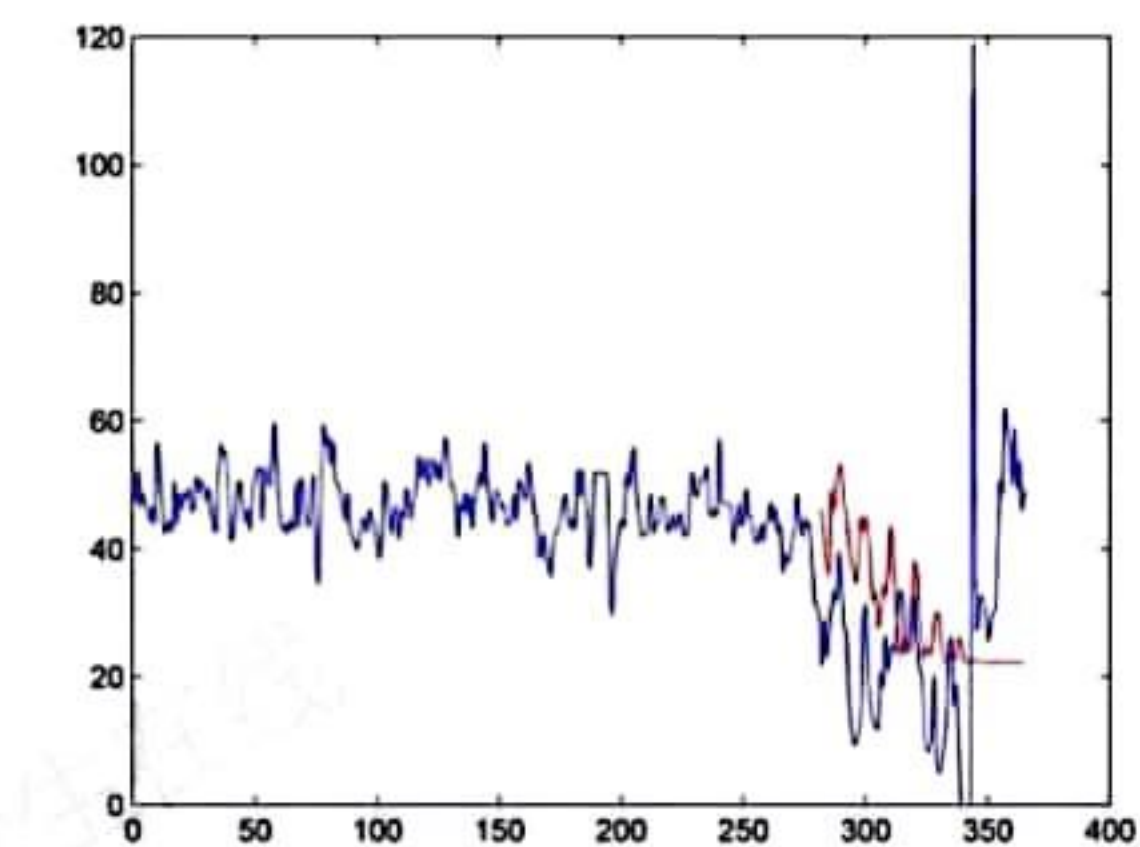


图 17、第一宿舍漏损预测

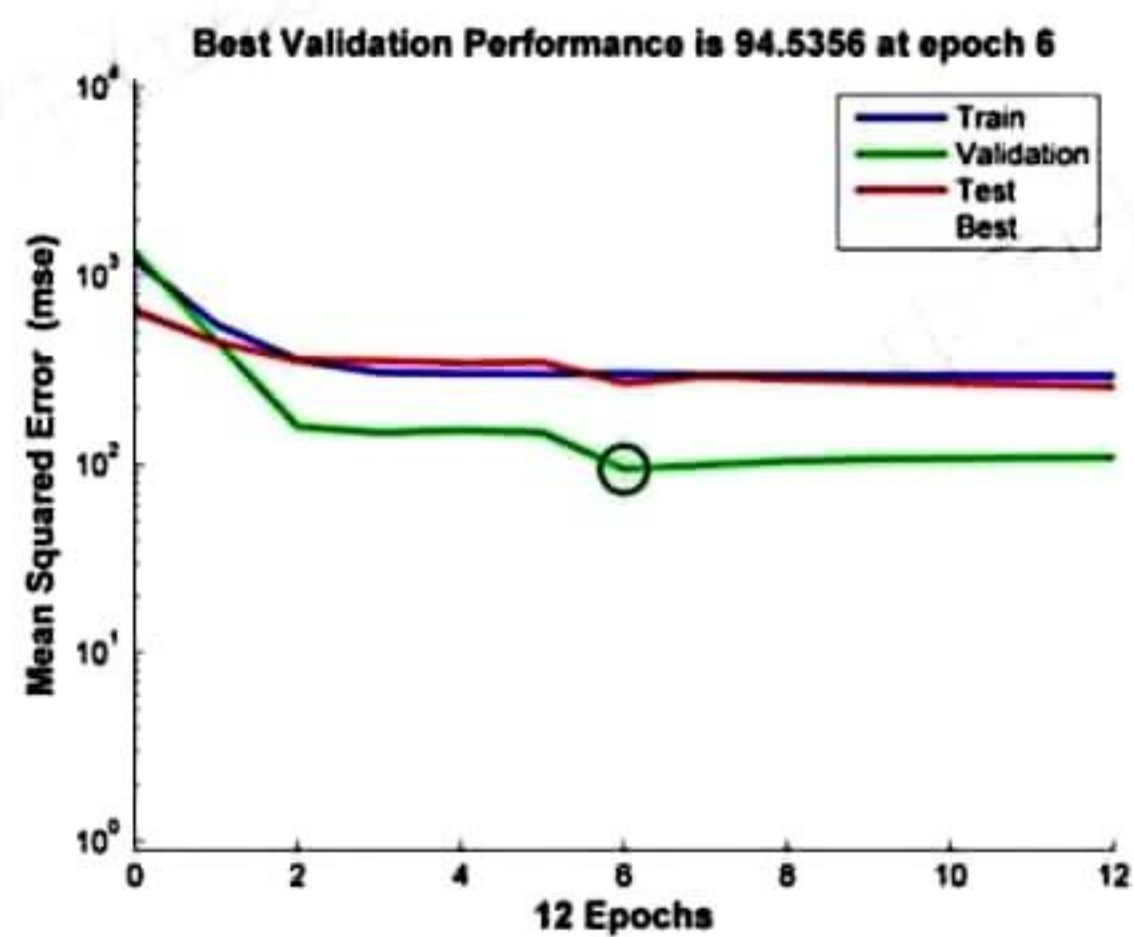
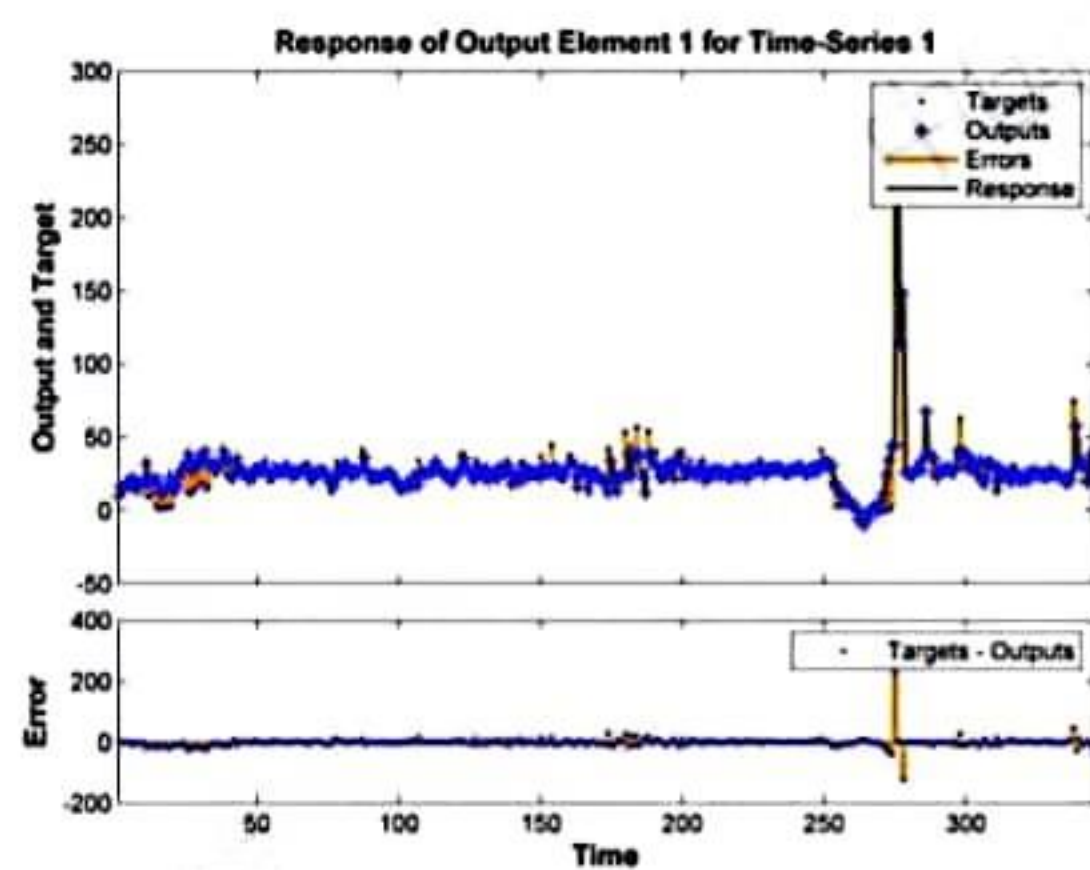
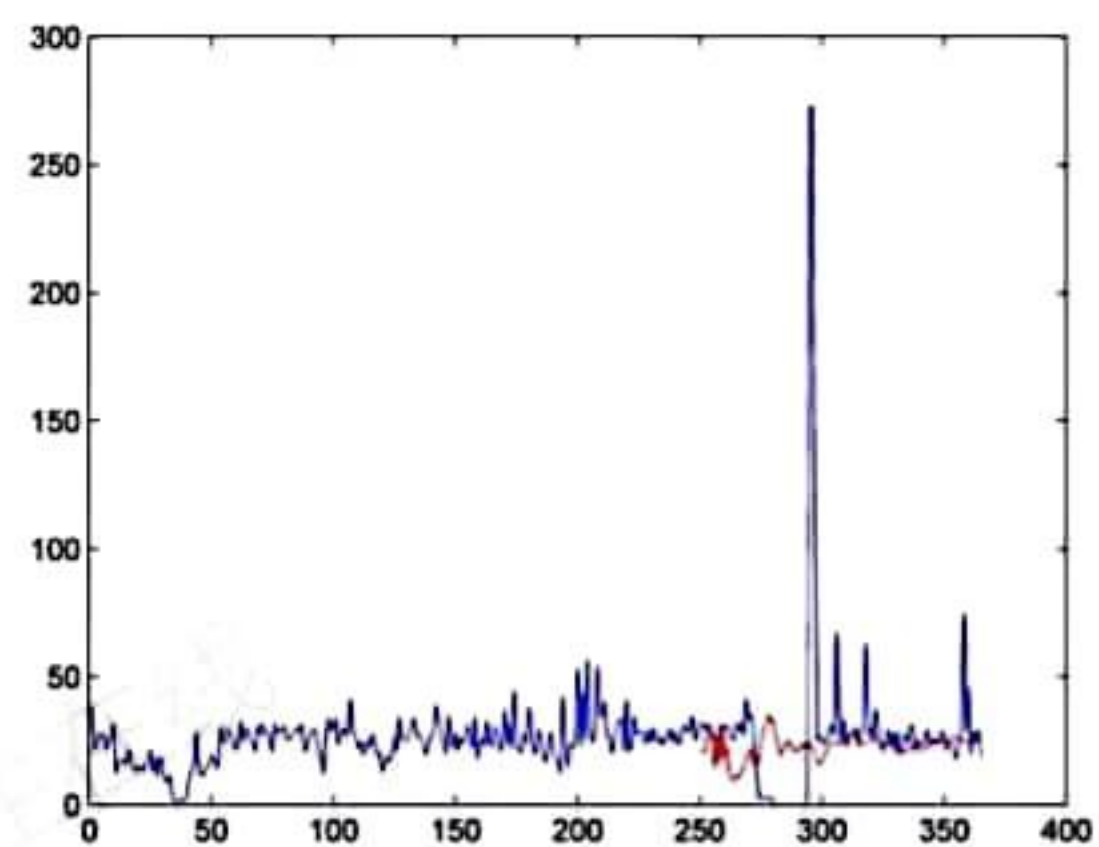
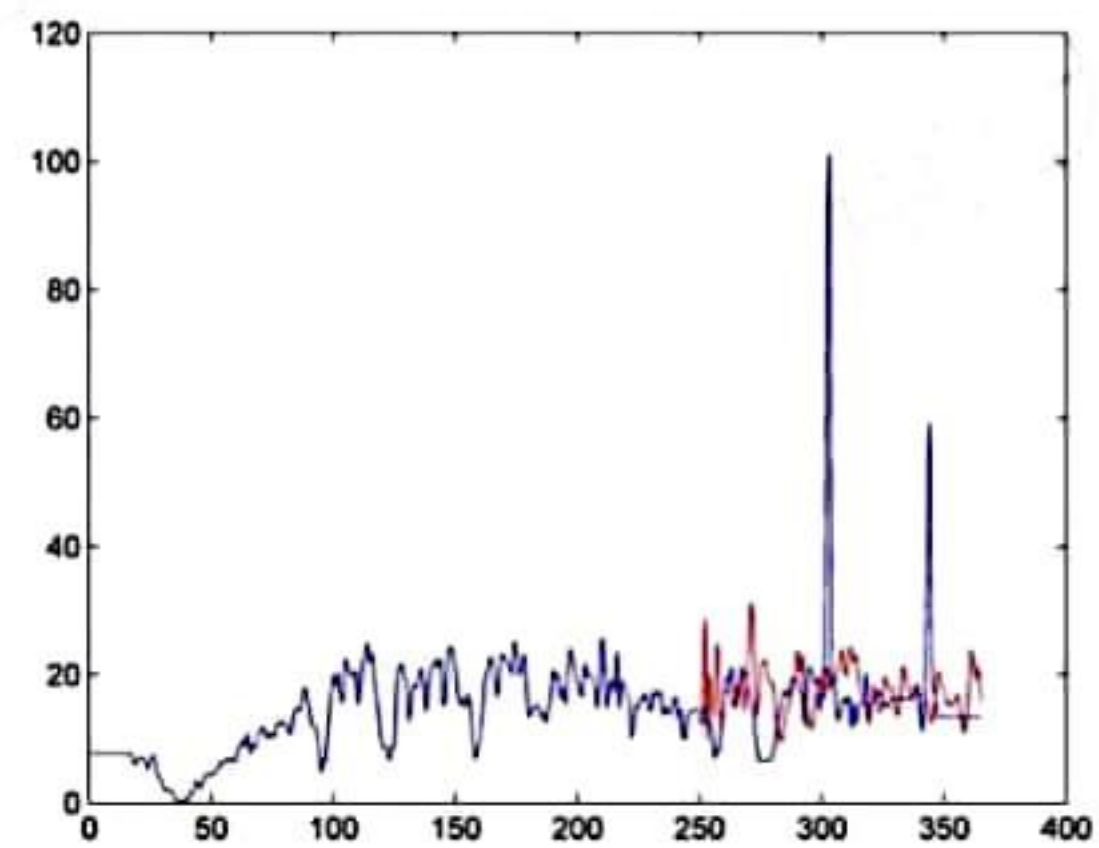
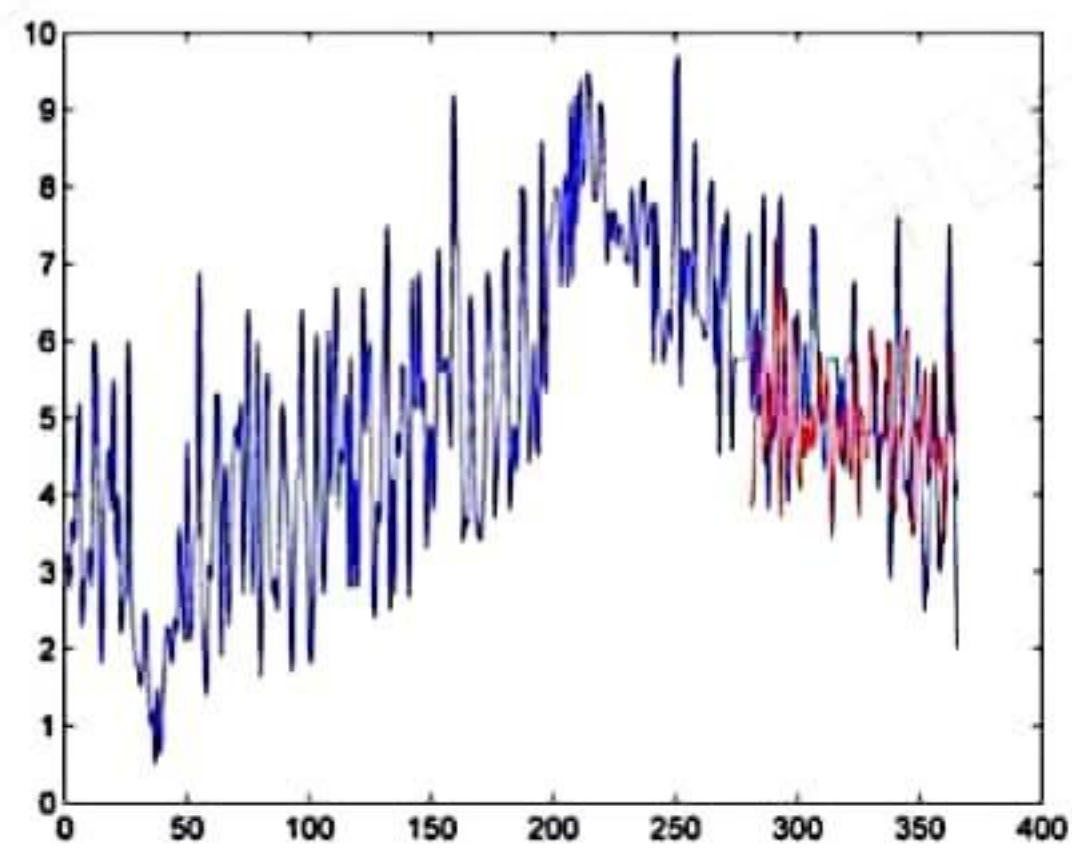


图 18、第八宿舍漏损预测



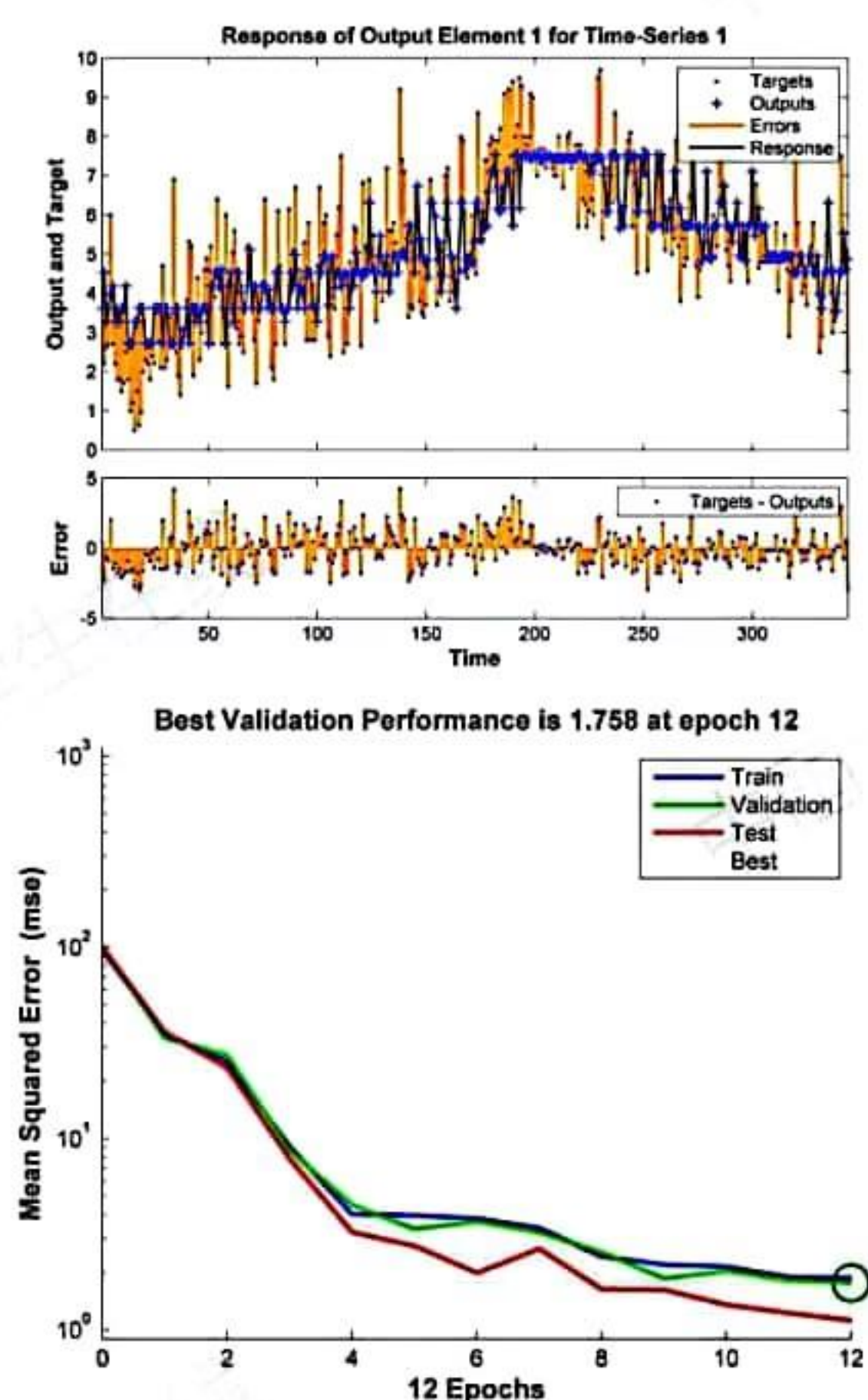


图 19、第五食堂漏损预测

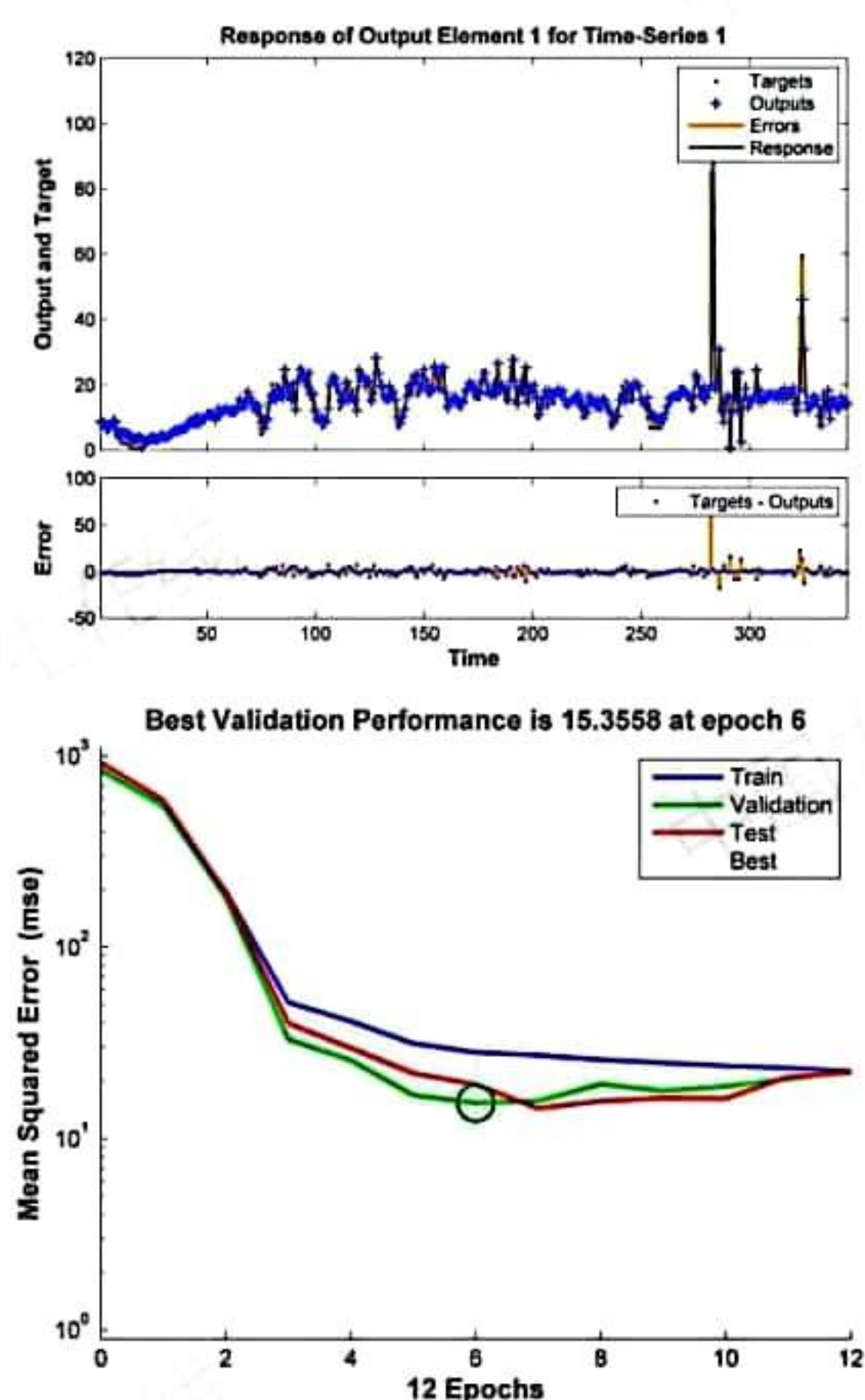


图 20、教学楼漏损预测

通过神经网络智能识别了全网节点，识别出以下漏损情况：

表 10、神经网络智能识别漏损情况

季度	水表节点
一季度	64397 副表、第一学生宿舍、XXX 花圃+、老六楼、东大门传达室、教育超市+、农业试验站大棚、区域 3+、区域 4+、新留学生楼、养鱼组厕所、养殖队 6721 副表、养殖馆+、养殖馆附房保卫处宿舍+、养殖馆附房二楼厕所+、养殖馆附房一楼厕所+
二季度	64397 副表、XXXL 楼、XXXS 宾馆、XXX 第二学生宿舍、XXX 干训楼、XXX 锅炉房、XXX 国际纳米研究所、XXX 航空航天、后勤楼、花圃+、老医务室、体育馆、污水处理、游泳池、中心水池、东大门大棚、离退休活动室、理发店+、区域 2、区域 4、书店
三季度	XXX 副表、舍热泵、M 管、S 管、后勤、成教院 XXX 分院、第二学生宿舍、图书管、校医院、区域 4+、体育馆网球场值班室、物业、消防、校管中心种子技术+、校医院南、新大门传达室、养殖队 67 副表、养殖管、养殖管房二楼厕所、K 酒店严、大楼厕所西、东大楼、游泳池是严重漏损、第八宿舍、车队、东大门温室+、东大门大棚、留学生(新)、区域 3+、养鱼组临工宿舍、第九学生宿舍、第七学生宿舍、第三学生宿舍、第一学生宿舍、第四学生宿舍、第五食堂、第五学生宿舍，第一食堂，L 管、花圃、区域（西）、新留学生楼（新）、浴室
四季度	K 酒店、第一食堂、养殖管、国际纳米、教学大楼、茶叶园、东大门温室、大楼厕所西、花圃、新留学生楼、宾馆、锅炉房

根据神经网络智能识别的节点，发现随着夏季的到来，用水达到高峰，漏损也更加严重。某些漏损还有片状发展趋势，例如第一季度的养殖场区域，成片出现，可能是管道老旧或者管道串联造成的。

5.5 问题五的模型建立与求解

5.5.1 问题五的模型建立

(1) 维修成本与漏损成本的关系

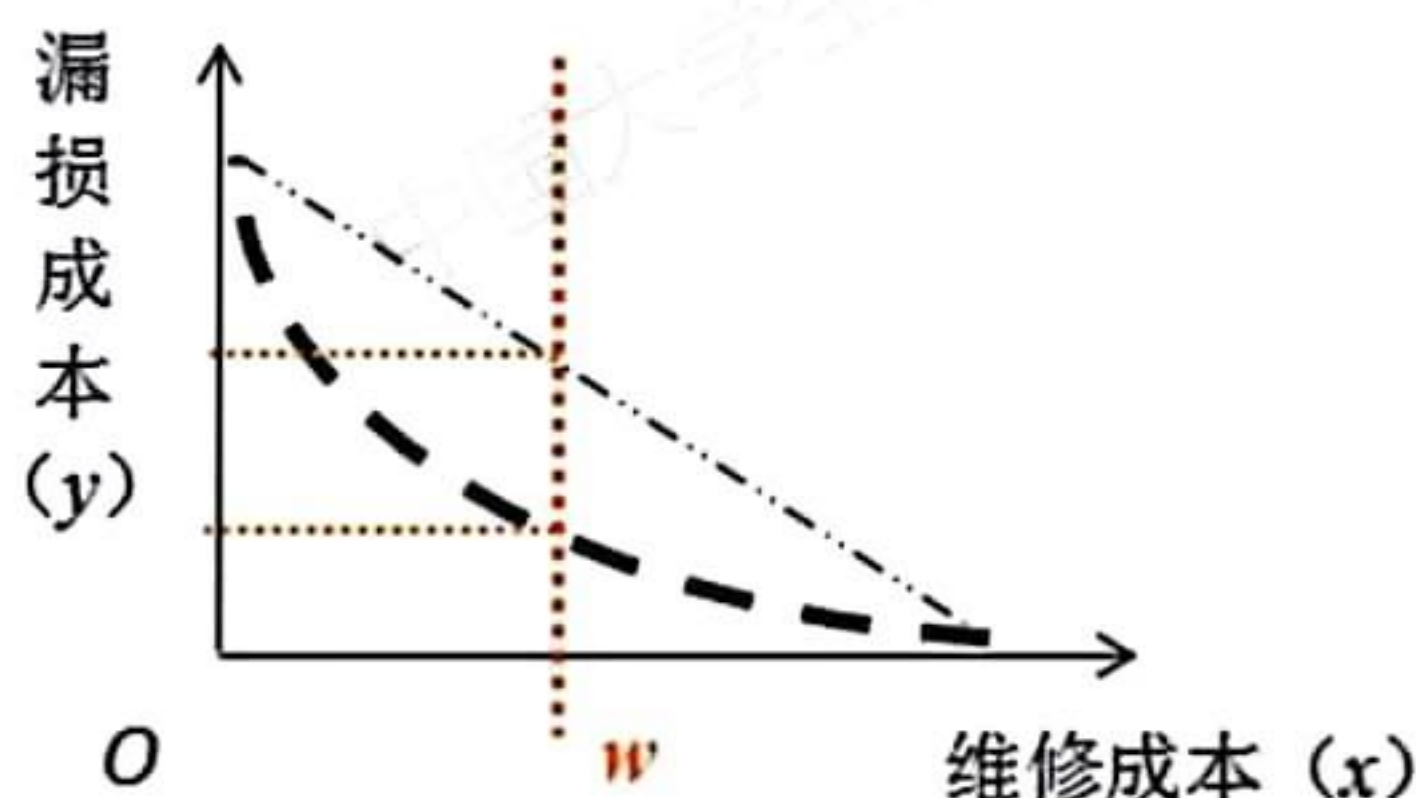


图 21、维修成本与漏损成本的关系

(2) 统计各管道设施设备损坏数据

$$R = \sum_{i=1}^n N_i \cdot R_i \quad (23)$$

R : 维修总成本

N_i : i 种设施设备损坏的个数

R_i : i 种设施设备损坏所需维修费

(3) 漏损比率控制

若漏损比率控制在 5% 以下，保持现在的维修成本不改变，若是升高，则适当提高维修费用。

5.5.2 问题五的模型求解

从 5.3 中计算的结果可知，学校全年漏损率为 5.31%，超过了 5%，需要提高一定的维修成本减低漏损率。全年漏损 $21548.48m^3$ ，根据贵州水价 3.3 元/ m^3 ，全年漏损成本为 71109.984 元，现需要减低漏损成本，则要通过增加维修成本来控制，假设两者为线性关系，相关系数为 -1，那么需要 $21548.48m^3 - 20273.576m^3 = 1274.904m^3$ ，通过水价计算得到 4207.1832 元。也就是在原有维修成本的基础上增加 4207.1832 元，就可以将漏损率控制在 5% 以下。

六、模型的评价

问题一中，计算结果不仅很好的反映出每天、每月、每季度的变化情况，还通过四个指标（日用水量、日最大小时用水量、日平均小时用水量、日小时变化系数）综合反映了各楼、各功能区的用水平稳程度、管网漏损等情况。如果统计数据时间长度都相同，那么会更好的反映出各级表的用水特征。

问题二中，通过多元线性回归模型建立、异常数据的处理和最小二乘法的运用，使得回归模型更加可靠，残差的分析也表明模型是最优模型。

问题三中，由于漏损不能直接通过计算得到，因此采用估算的方式得到漏损的水量，并分为四个季度去考察，使用二倍标准差剔除数据的方法将异常数据（漏水量）累加得到总的漏损量，从而可以计算得到漏损比率。有两个一级表没有数据，因此是通过累加二级表数据的方式获得，如果有该两个一级表的数据，计算结果会更加精确。

问题四中，需要快速有效且智能的方式寻找到漏损位置，因此建立神经网络智能算法去识别，通过智能算法的计算，找到了各级表漏损故障，十分可靠的。如果要对微小的漏损都快速识别，那么需要更多更完整的数据，以及管网的分布，设计，规格等。

问题五中，建立漏损成本与维修成本的线性反比关系函数，通过仅有的数据去粗略估算，只能是作为参考，因为两者的关系应该是非线性的，所以如果能获得学校每月、每季度、每年的维修费用，以及维修各种规格型号的管道及相关设备所需费用，还有这些设施设备的损坏次数等，就可以更加准确的找到两者的关系函数。

参考文献

- [1] 邵圆媛.嵌套 BP/GMS 神经网络模型在供水管网漏损预测中的研究[D].重庆大学.
- [2] 常金秋.III类学生宿舍用水特征分析[J].上海应用技术学院学报(自然科学版).13(3):245-248.

附录 1

```
%神经网络预测代码
x=C;
% 该脚本用来做 NAR 神经网络预测
lag=18; % 自回归阶数
iinput=x; % x 为原始序列（行向量）
n=length(iinput);
%准备输入和输出数据
inputs=zeros(lag,n-lag);
for i=1:n-lag
    inputs(:,i)=iinput(i:i+lag-1)';
end
targets=x(lag+1:end);
%创建网络
hiddenLayerSize = 10; %隐藏层神经元个数
net = fitnet(hiddenLayerSize);
% 避免过拟合，划分训练，测试和验证数据的比例
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;
%训练网络
[net,tr] = train(net,inputs,targets);
%% 根据图表判断拟合好坏
yn=net(inputs);
errors=targets-yn;
%figure, ploterrcorr(errors) %绘制误差的自相关情况（20lags）
%figure, parcorr(errors) %绘制偏相关情况
[h,pValue,stat,cValue]= lbqtest(errors) %Ljung-Box Q 检验（20lags）
figure,plotresponse(con2seq(targets),con2seq(yn)) %看预测的趋势与原趋势
%figure, ploterrhist(errors) %误差直方图
figure, plotperform(tr) %误差下降线
%% 下面预测往后预测几个时间段
fn=85; %预测步数为 fn
f_in=iinput(n-lag+1:end)';
f_out=zeros(1,fn); %预测输出
% 多步预测时，用下面的循环将网络输出重新输入
for i=1:fn
    f_out(i)=net(f_in);
    f_in=[f_in(2:end);f_out(i)];
end
% 画出预测图
figure,plot(1:365,iinput,'b')
hold on
plot(281:365,f_out,'r')
```


附录 2

```
%残差正态性检验代码
normplot(CXE)
CXE=zscore(CAN);
CXE=CXE';
alpha=0.05;
[mu,sigma]=normfit(CXE);
p1=normcdf(CXE,mu,sigma);
[H1,s1]=kstest(CXE,[CXE,p1],alpha);
n=length(CXE);
if H1==0
    disp('该数据服从正态分布。')
else
    disp('该数据不服从正态分布。')
end
```