

封一

答卷编号（参赛学校填写）：

答卷编号（竞赛组委会填写）：

论文题目：（标明 A、B、C、D、E、F 之一）

F

组 别：（填写研究生、本科生、专科生或中学生）

研究生

参赛队员信息（必填）：

	姓 名	学 号	联系电话
参赛队员 1	韩梦容	21733002	18342787879
参赛队员 2	程晓磊	31733003	18109871270
参赛队员 3	解俊文	31733005	13050850902

参赛学校：

大连理工大学

封二

答卷编号（参赛学校填写）：

答卷编号（竞赛组委会填写）：

评阅情况（学校评阅专家填写）：

学校评阅 1.

学校评阅 2.

学校评阅 3.

评阅情况（联赛评阅专家填写）：

联赛评阅 1.

联赛评阅 2.

联赛评阅 3.

城市供水管网片区用水异常模式识别

摘要 随着城市的快速发展,供水企业对用水异常模式识别是城市供水管网片区漏损检测和运行管理的重要课题。本文主要研究利用流量仪表瞬时流量和压力信息来监测 DMA 区域的用水情况,通过对信息数据处理得到典型用水和漏水模式,进而基于典型用水模式来监测异常模式,对异常模式进行识别。首先利用高斯拟合分析方法得到 DMA 区瞬时流量的拟合曲线,辨识出典型的用水模式,在对每天,每周做相关性分析来验证其典型性。然后通过对原始数据进行盲源分离,采用 FastICA 算法得到典型的漏水模式,最后基于已经得到的典型用水模式,利用局部离群因子算法对异常模式进行识别。

本文针对问题一的用水模式的分析,采用从整体到个体、特殊到一般的数学逻辑思维,对瞬时流量数据进行了以下四方面的分析:1、图像分析 2014.4.15 至 2014.6.12 整体的数据;2、高斯拟合分析方法反映均值瞬时流量的一天时序变化;3、特例分析一周内瞬时流量以一天为单位的变化趋势,来表明用水模式在工作日与非工作日的区别;4、利用相关性分析来得到一天内的用水模式的一般性。通过四方面的模型分析可得典型的用水模式,即上午 8-9 点(小)和 11-12 点(大)出现了两个用水高峰期,在晚上 19-20 点(小)和 21-23 点(大)左右也出现了两个用水高峰期;而在凌晨 2 点至 5 点之间用水量几乎为零;5 点到 8 点和下午 15 点至 19 点两个时间段中用水量是呈上升阶段,12 点至下午 15 点和 23 点至 2 点这两个时间段用水量是逐渐下降趋势的,且在节假日期间用水量整体出现下降趋势。

对问题二的漏水模式分析,根据瞬时流量和压力的数据,采用盲源分离的方法来将夜间的瞬时流量分为夜间基本用水量和漏损量。采用 FastICA 算法利用 MATLAB 和 R 软件分析得到实时量化的漏损量,来辨识出该分区的漏水量模式,即漏损量大致上是维持稳定的,其值约为 23.32503093512465(l/s) 左右。但在某些时间段上会出现波动,可能是周期性的停水导致或者是爆管后的停水检修。

问题三的基于典型用水模式的异常模式识别,通过初步的瞬时流量在每天的时序变化图像分析,粗略得到出现疑似异常的时间段,进一步基于无监督学习的局部离群因子(Local Outlier Factor,简称 LOF)的移动均线标准差偏移图像,结合瞬时流量时间序列移动平均线来综合确定异常点。最后根据异常点的分布来对异常模式进行识别,得到结论:5 月 21 日、5 月 25 日、5 月 27 日这三个时间段异常模式为特定时刻的水量异常增加(短期的,如爆管);5 月 28 日和 6 月 6 日这两个时间段异常模式为节律变化(用水高峰期不符合生活或生产节律)(如偷用生活用水做商业用途)。

关键词 用水模式;高斯拟合;盲源分离漏损量化;局部离群因子;异常检测

目录

1、问题重述.....	3
1.1 问题背景.....	3
1.2 需要解决的问题.....	3
2.问题分析.....	3
2.1 文献检索评述.....	3
2.2 问题 1 的分析.....	4
2.3 问题 2 的分析.....	4
2.4 问题 3 的分析.....	4
3.模型假设.....	4
4.符号说明.....	4
4.1 问题二的符号说明.....	4
4.2 问题三的符号说明.....	5
5.模型建立、结果及分析.....	5
5.1 问题 1：基于高斯拟合和相关性分析的数据分析方法.....	5
5.1.1 建立基于高斯拟合和相关性分析的数据分析模型.....	5
5.1.2 模型结论.....	12
5.2 问题二：盲源分离漏损量化指标来分析漏水量模式.....	12
5.2.1 盲源分离漏损量化模型的建立.....	13
5.2.3 基于盲源分离的漏损量化计算和分析.....	14
5.2.2 模型结果.....	17
5.3 问题三：根据局部离群因子（LOF）进行异常模式识别.....	17
5.3.1 数据分析.....	18
5.3.2 局部离群因子分析模型的建立.....	19
5.3.3 模型结果.....	25
6、模型评价与展望.....	25
6.1 创新点和模型改进.....	25
7 参考文献.....	26
附录.....	27

1、问题重述

1.1 问题背景

掌握每个片区管网用水情况是自来水公司实施科学化运行管理的依据。自来水公司对城市供水管网进行划分，构成多个独立计量区域(DMA)，在每个DMA区域入口处布设流量仪表(如图1所示)，监测DMA区域的用水情况。这种方法是管网漏损控制管理及运行管理的重要手段。

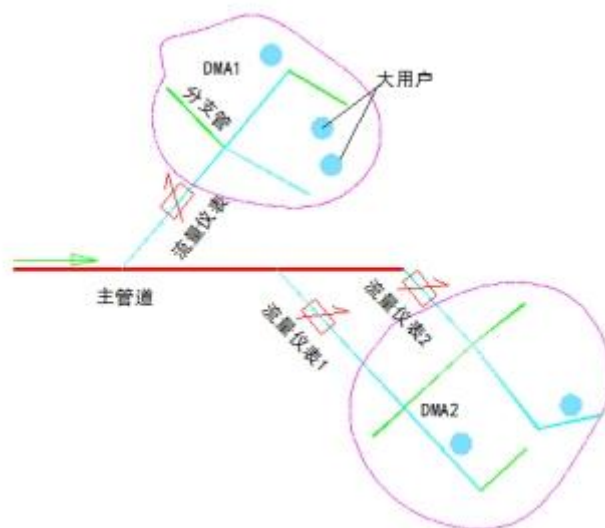


图1 分区计量示意图

在DMA分区入口流量仪表测得的数据序列中(瞬时流量)，一般认为包含两部分水量：1)DMA区域内各用户用水量之和；2)DMA区域内管网漏失的水量。

1.2 需要解决的问题

围绕城市供水管网片区用水异常模式识别的任务，本文依次解决如下问题：

问题1：根据给定的测量结果，辨识出该区域的典型用水模式(如居民小区、工厂、机关事业单位的用水量在时间上有显著不同的分布)；

问题2：辨识该分区的漏水量模式；

问题3：利用辨识出的典型用水模式，对异常模式进行识别；异常模式包括特定时刻的水量异常增加(短期的，如爆管)，节律变化(用水高峰期不符合生活或生产节律)(如偷用生活用水做商业用途)；

2. 问题分析

2.1 文献检索评述

本题是有关城市供水管网片区用户用水模式、漏水模式和用水异常模式识别问题，涉及到数据分析和故障预警两方面。

随着城市的快速发展和给水管网的延伸和资源紧缺困境日益突出，供水企业对给水管网漏损控制和用水异常模式识别愈来愈引起重视。但是大多数供水企业现今仍采用人工经验运行模式，能源消耗较大。如果采用科学调度管理模式，使得调度时供水量尽可能地接近城市时段用水量，并对用水异常模式进行识别，进而对异常情况进行控制与管理。可以节约大量的供水能量消耗。用水模式指用户

用水量的时序变化过程，与用户的用水性质、天气条件和节假日安排等因素密切相关。对于用水模式方面的研究分析，国内外已有许多相关的文献，用到的方法一般有聚类分析法、时间序列三角函数分析法、不确定度量分析法、动态模型等。而漏水量化方法有水平衡法、夜间最小流量法、盲源分离及卡尔曼滤波法。这些方法都有各自的特点和不足。本文根据所提供的数据和问题的重点来对问题进行具体分析进行建模。

2.2 问题 1 的分析

用水模式指用户用水量的时序变化过程，与用户的用水性质、天气条件和节假日安排等因素密切相关。根据所提供的数据，可以发现测量数据时间跨度为 58 天（2014.4.15—2014.6.11）（2014.6.12 的数据为不完整的一天，这里不给予考虑），且每组测量值时间间隔为 15 分钟，数据值为水管的瞬时流量（包含两部分水量：1.DMA 区域内各用户用水量之和；2.DMA 区域内管网漏失的水量）。这里假设用水管道存在一定的漏水量，即夜间 2:00—5:00 的水量可以用来约等于恒定的管道漏损量。因此可对瞬时流量数据进行层层递进的四方面分析来反映该区典型的用水模式，即①考虑整体数据的分布情况来总结大体趋势；②基于一天的数据来进行高斯拟合，从得到的拟合函数中分析用水量模式；③通过一周的数据来反应工作日和周末的用水量区别；④对数据进行以一周为单位的相关性分析，说明用水模式的一般性。

2.3 问题 2 的分析

漏损是城市地下供水管网面临的普遍现象，且城市漏损率呈上升趋势，远超过供水行业的指标要求，漏损量及漏损率是衡量供水管网运行状况好坏的重要指标，由于城市地下供水管网复杂、漏损不可视化等原因，漏损量难以准确计算。可以采用夜间最小流量法来量化，但是其加入了估计项，在进行控漏前不能很好的计算区域漏损程度，只能估计每日漏损总量。因此这里采用一种利用信号处理方法：盲源分离法，来将夜间用水量混合信号进行分离，得到夜间基本用水量信号和漏损量信号，从而实现实时监控漏损量数值，得到该分区的漏水量模式的目标。

2.4 问题 3 的分析

由于瞬时流量数据的周期性，辨识出的典型用水模式反应在了以天为单位的用水的时序规律。这里考虑以天为单位做出瞬时流量分布图，从图中初步粗略地得到疑似异常的日期，进一步采用非监督学习方法的局部离群因子参数来量化地分析出异常模式。异常模式包括特定时刻的水量异常增加（短期的，如爆管），即反应为局部出现一个离群点现象；节律变化（用水高峰期不符合生活或生产节律）（如偷用生活用水做商业用途），即出现连续的几个非零离群点现象。

3. 模型假设

1. 所有的数据无误，无非正常因素对数据产生扰动影响；
2. 假设数据中的异常点对用水模式研究的影响可以忽略不计；
3. 问题一中假设夜间 2:00—5:00 的水量可以用来约等于恒定的管道漏损量；
4. 问题二中假设漏水量来自环状管网单点漏损。

4. 符号说明

4.1 问题二的符号说明

符号	说明
$x(t)$	N 维混合信号, $x(t)=[x_1(t), x_2(t), \dots, x_N(t)]^T$
A	混合矩阵, $n \times n$ 维矩阵
$s(t)$	M 维源信号, $s(t)=[s_1(t), s_2(t), \dots, s_M(t)]^T$
$n(t)$	N 维噪声信号
$y(t)$	分离得到的源信号
W	分离矩阵
\bar{x}	总流量信号均值 (l/s)
N	数据个数
σ	标准差
$y_q(t), l_q(t)$	去均值后的分离用水量, 漏损量信号, 无量纲
$y(t), l(t)$	分离用水量, 漏损量信号, 无量纲
$\bar{y}_q(t), \bar{l}_q(t)$	分离用水量, 漏损量信号均值, 无量纲
$Q_y(t)$	管网实际用水量 (l/s)
$Q_l(t)$	管网实际漏损量 (l/s)
t_1, \dots, t_n	时间 s
$Q_s(t_i)$	在时刻 t_i 管网实际总供水量 (l/s)

4.2 问题三的符号说明

符号	说明
$LOF_k(p)$	对象 p 的局部异常因子
$k-distance(p)$	p 的 k -距离
$reach-dist_k(p, o)$	对象 p 相对于对象 o 的可达距离

5. 模型建立、结果及分析

5.1 问题 1: 基于高斯拟合和相关性分析的数据分析方法

5.1.1 建立基于高斯拟合和相关性分析的数据分析模型

① 利用 MATLAB 工具对整体的数据作折线图 (如图 1.1) 来反应整体数据

的分布情况；具体的作图代码详见附件“Model1_1.m”。

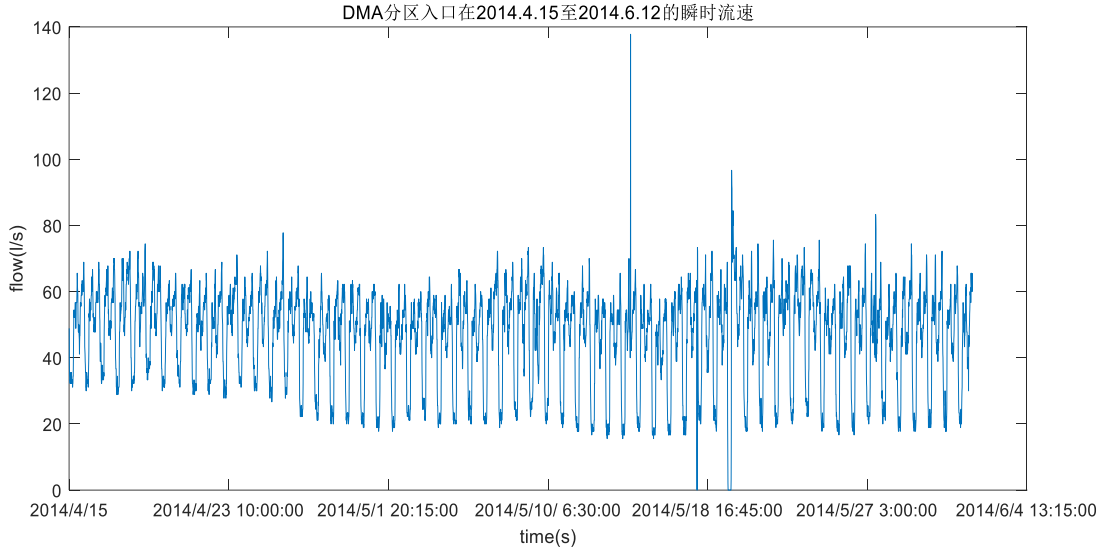


图 1.1 一个 DMA 分区入口在 2014.4. 15 至 2014.6.12 的瞬时流速

从图 1.1 中可以得到：以一天为单位，大体上，整体的用水量具有周期性。且在一个周期内存在两个峰区的波形趋势。但也存在异常的情况，如：2014.5.21 的 15:00:00 时瞬时流量达到了 137.7800(l/s)，可能是出现了爆管等异常情形；2014.5.25 的 23:00:00 至 23:15:00 和 2014.5.27 的 23:45:00 至 2014.5.28 的 4:15:00 这两个时间段瞬时流量为 0(l/s)，可能是出现的短期的偷水、停水等异常情况。为了进一步分析出 DMA 分区的用户用水模式，需要以一天为例来对数据进行二次高斯拟合分析。

② 首先对瞬时流量数据以一天为单位做散点图（见图 1.2），进一步由于要求分析出该区域的典型用水模式，则需要考虑水管的瞬时流量中哪一部分为用水量。这里的用水量为瞬时流量减去管道漏损量，考虑将夜间 2:00—5:00 的水量的均值作为恒定的管道漏损量。此外对于出现的异常情况，可以忽略其影响。因为异常数据的频数很小，从而通过求均值来弱化其影响，可忽略不计。因此对处理过的均值数据进行四次高斯拟合。

这里利用 MATLAB 工具来对数据进行拟合。按照设计思路，观察图 1.2 中每天的瞬时流量图像，发现整体的趋势是呈现两个峰区（每个峰有一大一小两个峰）的波形趋势。因此对数据进行均值处理，得到用水量（瞬时流量）均值随一天时间变化的散点图（见图 1.3），对数据用 MATLAB 中自带的函数拟合工具箱（cftool）程序见 GMM2.sfit 文件，进行粗略地拟合，并得到相关的参数，拟合的结果图像及相应参数见图 1.4。思路就是用四个 Gauss 分布的叠加去近似，即

$$f(t) = a_1 \cdot \exp(-((t-b_1)/c_1)^2) + a_2 \cdot \exp(-((t-b_2)/c_2)^2) \\ + a_3 \cdot \exp(-((t-b_3)/c_3)^2) + a_4 \cdot \exp(-((t-b_4)/c_4)^2)$$

其中 $a_1, b_1, c_1, a_2, b_2, c_2, a_3, b_3, c_3, a_4, b_4, c_4$ 为模型参数。详细代码见附件“Model1_2.m”

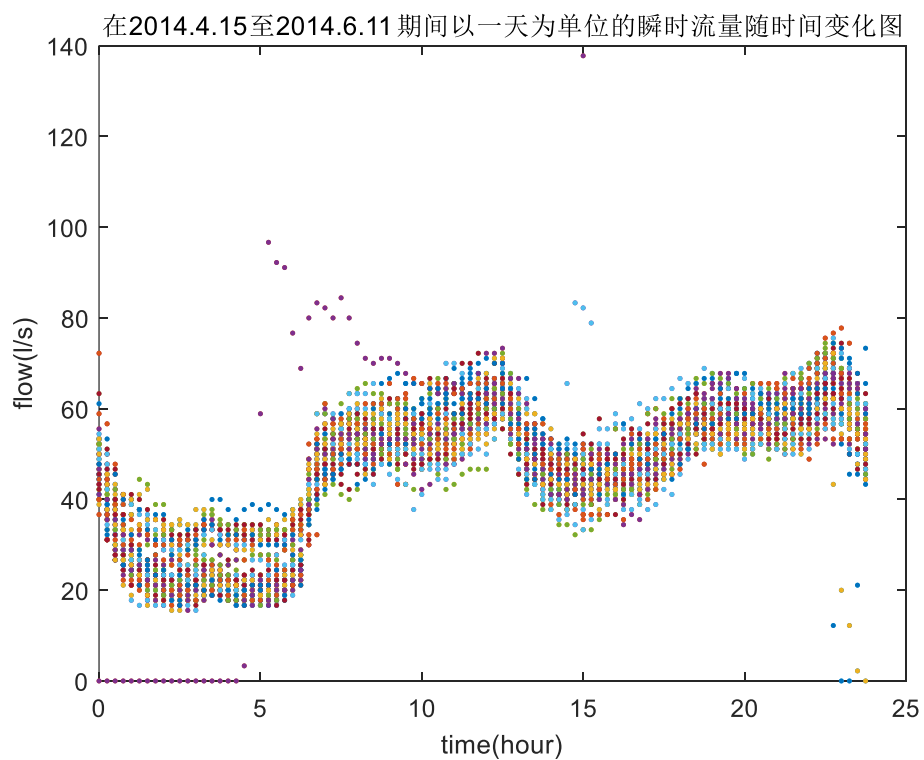


图 1.2 在 2014.4.15 至 2014.6.11 期间，以一天为单位的瞬时流量随时间变化

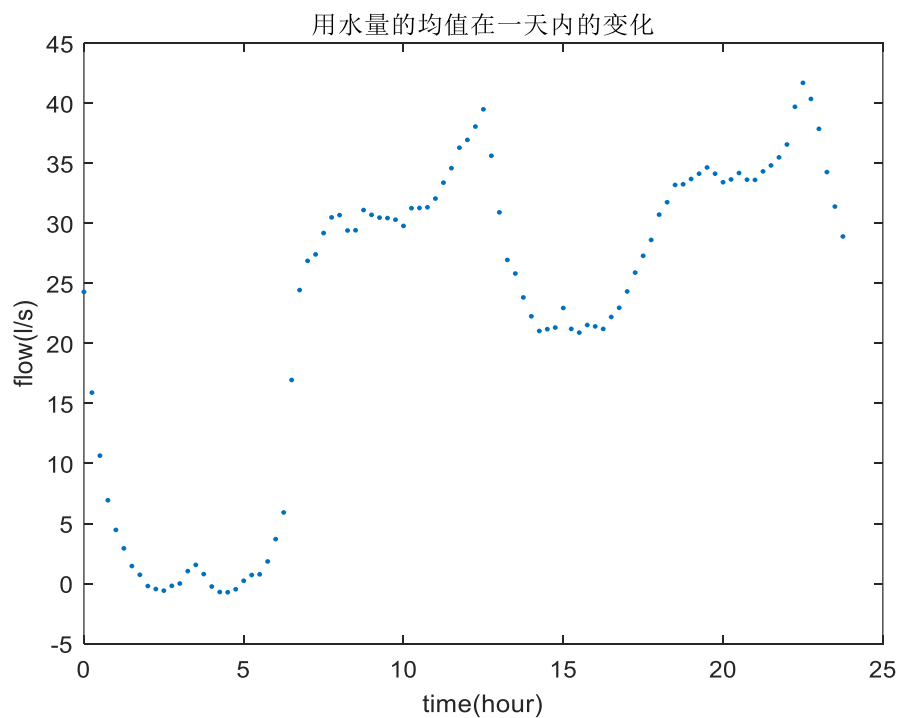


图 1.3 用水量（瞬时流量）均值随一天时间变化的散点图

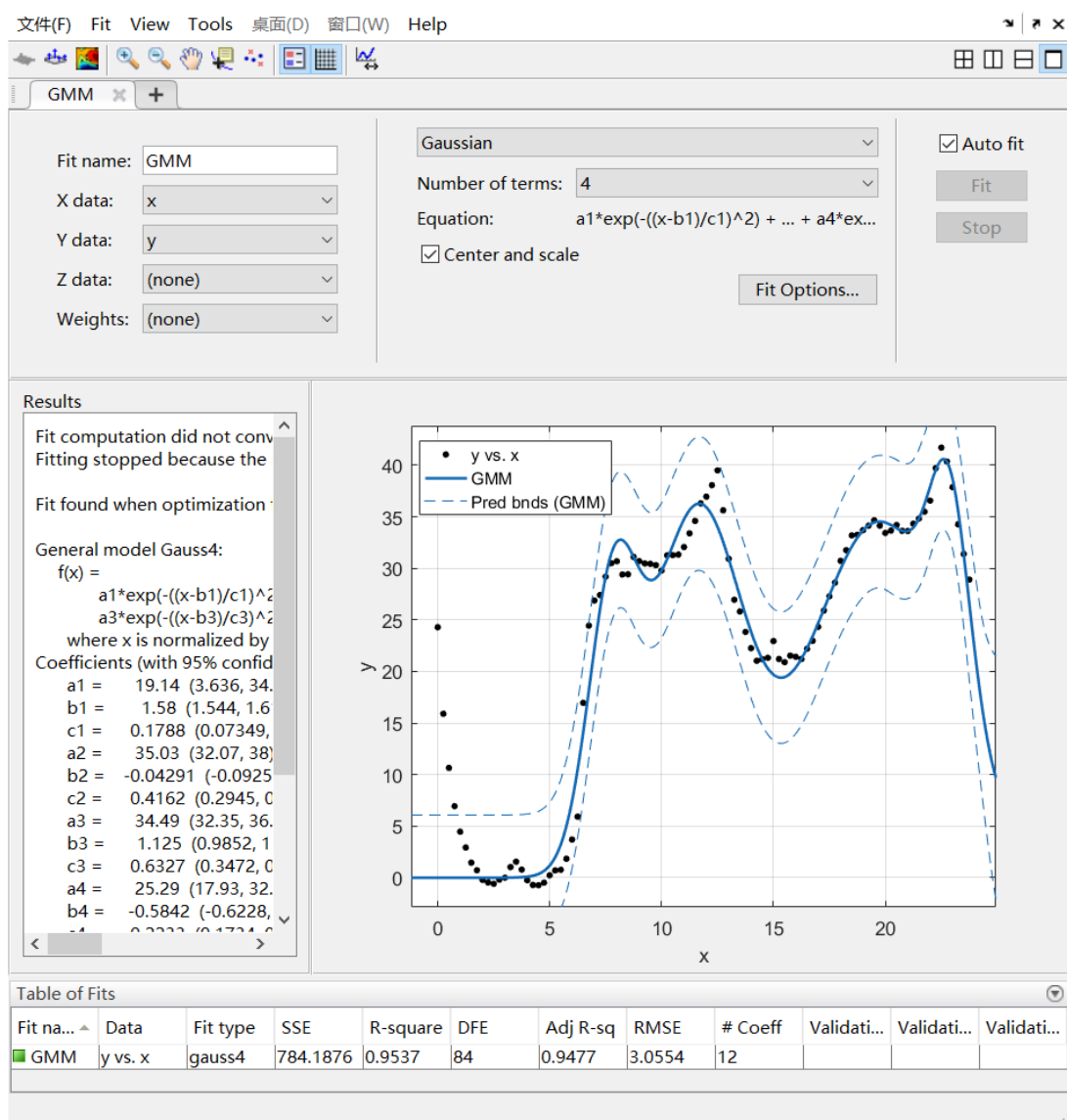


图 1.4 函数拟合工具箱的高斯拟合结果图

通过图 1.4 可以看出，在 95%的置信水平下，

$$a_1 = 19.14(3.636, 34.64), \quad b_1 = 1.58(1.544, 1.617), \quad c_1 = 0.1788(0.07349, 0.2841),$$

$$a_2 = 35.03(32.07, 38), \quad b_2 = -0.04291(-0.0925, 0.00672), \quad c_2 = 0.4162(0.2945, 0.5379),$$

$$a_3 = 34.49(32.35, 36.64), \quad b_3 = 1.125(0.9852, 1.264), \quad c_3 = 0.6327(0.3472, 0.9182),$$

$$a_4 = 25.29(17.93, 32.65), \quad b_4 = -0.5842(-0.6228, -0.5456), \quad c_4 = 0.2233(0.1734, 0.2731),$$

后面所对应的区间的置信区间。且拟合的残差平方值为 0.9537。由拟合结果可以看出，在上午 8-9 点（小）和 11-12 点（大）出现了两个用水高峰期，在晚上 19-20 点（小）和 21-23 点（大）左右也出现了两个用水高峰期；而在凌晨 2 点至 5 点之间用水量几乎为零；5 点到 8 点和下午 15 点至 19 点两个时间段中用水量是呈上升阶段，12 点至下午 15 点和 23 点至 2 点这两个时间段用水量是逐渐下降趋

势的。这就是 DMA 分区的用户一天内的用水模式。

③ 进一步通过分析一周 7 天的用水模式曲线，来反应工作日和周末的用水量区别。通过对 2014.5.21 至 2014.5.27 这 7 天的瞬时流量随时间的变化趋势，利用 MATLAB 作图工具来做出其变化趋势散点图，如图 1.5。

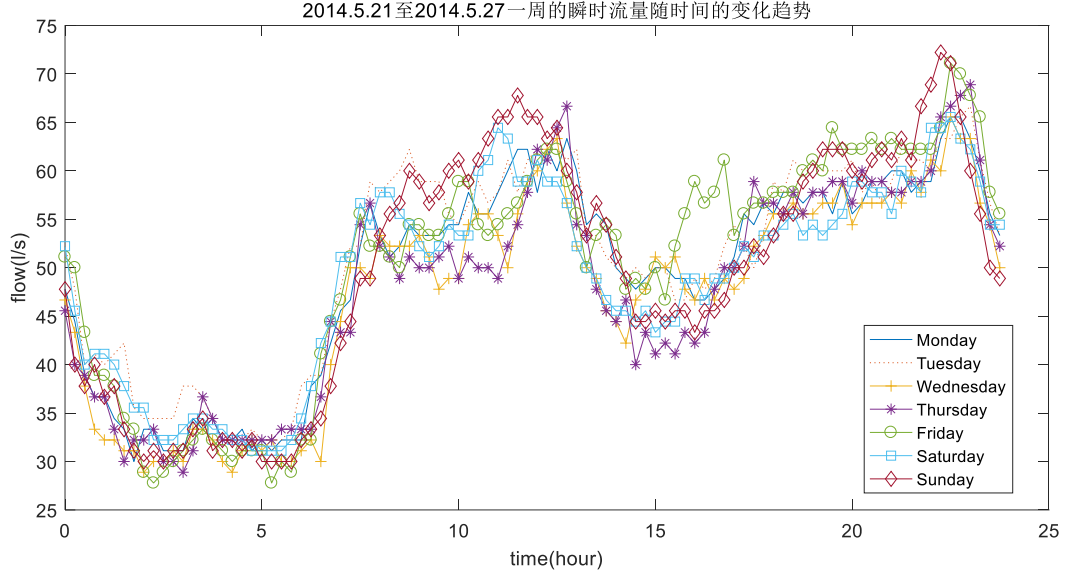


图 1.5 2014.5.21 至 2014.5.27 一周的瞬时流量随时间的变化趋势图

由图 1.5 可知，在一周之内的每一天中 DMA 分区的入口瞬时流量随时间的变化趋势是大体一致的。在考虑只有固定的管道漏损量的情况下，一周内的周末和工作日的用水量没有显著的差别，其用水趋势也是符合②的二次高斯函数的用水模式。

④ 由于②是对以天为单位，通过求 58 天数据的均值来进行分析的，具有整体性，但缺乏个体性。且③是对 58 天数据中任意抽取了一周进行考虑的，具有偶然性。因此，为了证明每天都有一般的用水模式，我们对此更深层次地选择对 58 天过程中八周的每一天都进行相关性分析，即对这八周两两进行相关性检验，来得到该区用户用水模式的一般性。

这里，我们运用 R 语言对八周的瞬时流量数据进行相关性分析，即对每周的瞬时流量都与其他周的瞬时流量进行相关性分析。采用 R 语言的 `corr.test()` 函数进行分析，具体的代码参见文件 Model1_3.R. 得到了八周的瞬时流量数据的相关性矩阵 F 和对应相关性检验的 p-value 组成的矩阵 P。

$$F = \begin{bmatrix} 1.00 & -0.46 & 0.02 & -0.42 & 0.90 & -0.50 & 0.11 & -0.49 \\ -0.46 & 1.00 & -0.53 & 0.07 & -0.44 & 0.74 & -0.55 & 0.10 \\ 0.02 & -0.53 & 1.00 & -0.47 & 0.04 & -0.32 & 0.94 & -0.42 \\ -0.42 & 0.07 & -0.47 & 1.00 & -0.46 & 0.07 & -0.52 & 0.92 \\ 0.90 & -0.44 & 0.04 & -0.46 & 1.00 & -0.49 & 0.11 & -0.52 \\ -0.50 & 0.74 & -0.32 & 0.07 & -0.49 & 1.00 & -0.33 & 0.11 \\ 0.11 & -0.55 & 0.94 & -0.52 & 0.11 & -0.33 & 1.00 & -0.47 \\ -0.49 & 0.10 & -0.42 & 0.92 & -0.52 & 0.11 & -0.47 & 1.00 \end{bmatrix}$$

$$P = \begin{bmatrix} 0.00 & 0.00 & 0.57 & 0.00 & 0.00 & 0.00 & 0.02 & 0.00 \\ 0.52 & 0.00 & 0.00 & 0.28 & 0.00 & 0.00 & 0.00 & 0.03 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.57 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.07 & 0.00 & 0.00 & 0.00 & 0.28 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.29 & 0.00 & 0.00 & 0.00 & 0.03 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.08 & 0.00 & 0.00 & 0.00 & 0.02 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix}$$

由相关性检验的基本原理，我们可以由 P 矩阵中的元素看出，对于 p -value 小于 0.05，认为是显著相关的，即对于绝大部分的周用水量是相似的，每天的用水量呈现出与典型用水模式近似的情况。因此我们可以认为该区用户用水模式是具有代表性的，分析每天的用水量可以得出该区用户用水量的典型的模式。而对于 p -value 大于 0.05 的，如：第一周与第三周，第二周与第四周，第三周与第五周，第四周与第六周，下面我们用 MATLAB 软件分别做出相应的对比图来进一步具体分析，具体的作图代码见 Model1_1.m.

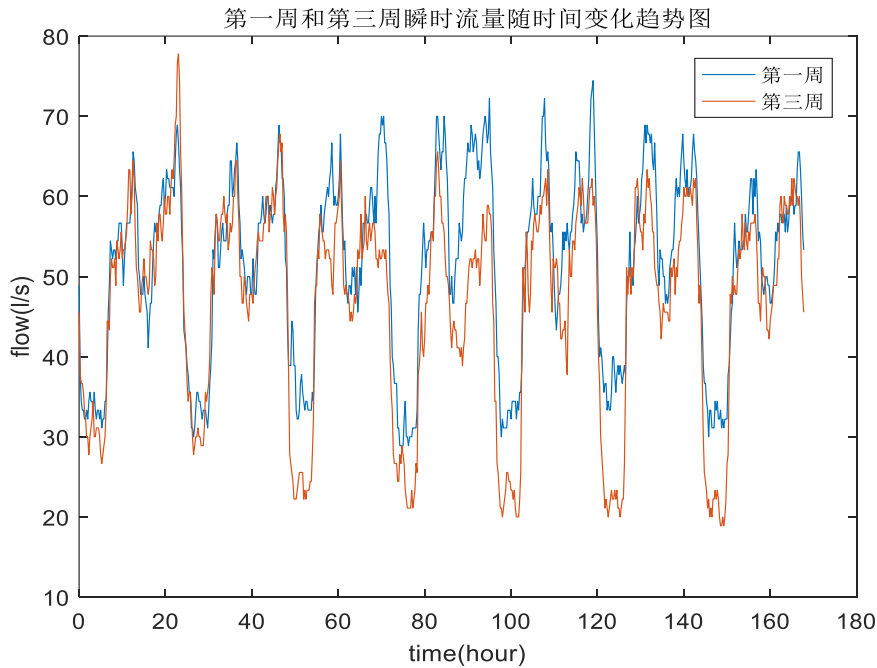


图 1.6 第一周和第三周瞬时流量随时间变化比较图

由瞬时流量数据做出第一周与第三周用水量的图像，如图 1.6.我们可以看出该区用户用水量在第三周内的 70-140 小时的图像一直在第一周内的 70-140 小时图像下方，且有明显的间距，对第一周与第三周的具体日期进行查询可以看出，第三周的 70-140 小时对应的日期为 2014.4.29—2014.5.1，为五一劳动节假日，因此在这三天显示出与第一周相对应日期的用水量有显著差异，可以说明在节假日用水量是有显著变化的，同样的情况出现在第五周的相同天数的用水量上，如图 1.7，进一步验证了假日用水量的显著变化。

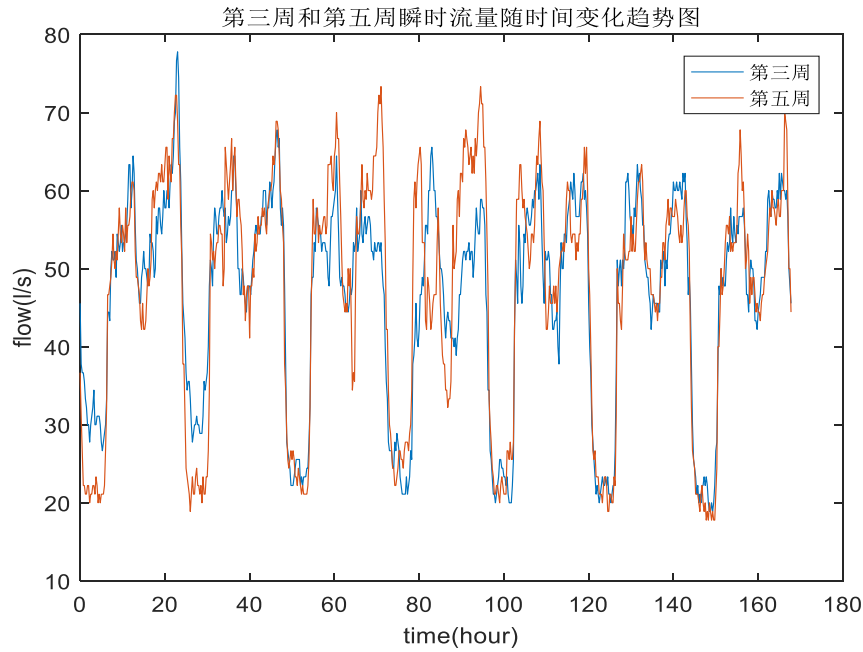


图 1.7 第三周和第五周用水量随时间变化的差异图

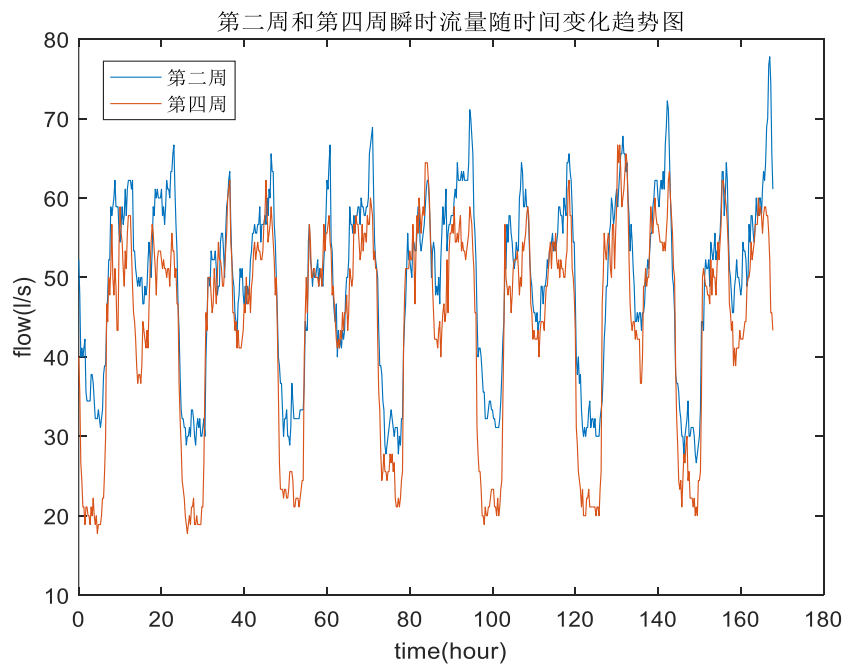


图 1.8 第二周和第四周用水量随时间变化比较图

由瞬时流量数据画出第二周与第四周用水量的图像，如图 1.8，我们可以看出这两周的用水量的波峰上有近似的波动趋势，而主要的差异出现在波谷上，即每天的夜间用水量和漏损水量呈现出显著差异，因为这一周的波谷都有相同的差异，我们可以认为是这一周部分地区停水或者其他情况，这样的异常点，我们会在问题 3 中进行分析。

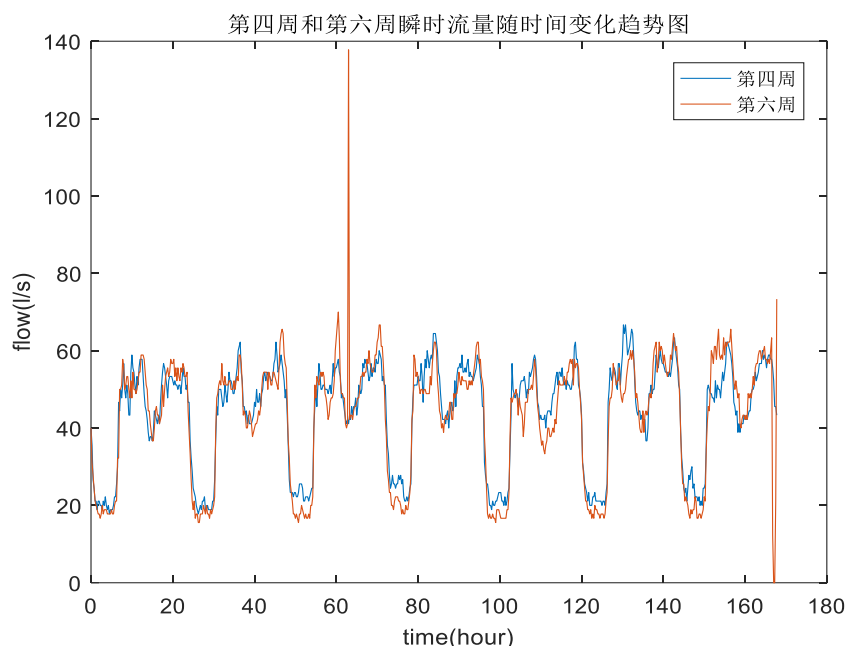


图 1.9 第四周和第六周用水量随时间变化的差异

由瞬时流量数据画出第四周与第五周用水量的图像，如图 1.9，我们可以看出这两周的用水量在周五上显著差异，也是相关性不显著的主要来源，可以看出在周五用水量异常值出现，体现在用水剧增，我们猜测是爆管导致，这样的异常点，也会在问题三中进行处理。

对这些相关性不显著的点进行分析有助于我们把握整体的典型用水量，首先对每天的用水量分析可以得到典型用水量的模式，这样不同周，不同星期，不同天该区域的典型用水模式具有典型性，其次该区域的典型用水模式具有节律性，最后对于异常的分析在问题三种展开。

5.1.2 模型结论

通过基于高斯拟合和相关性分析的数据分析方法，我们可以得到该区用户典型的用水模式：一天中，上午 8-9 点（小）和 11-12 点（大）出现了两个用水高峰期，在晚上 19-20 点（小）和 21-23 点（大）左右也出现了两个用水高峰期；而在凌晨 2 点至 5 点之间用水量几乎为零；5 点到 8 点和下午 15 点至 19 点两个时间段中用水量是呈上升阶段，12 点至下午 15 点和 23 点至 2 点这两个时间段用水量是逐渐下降趋势的；且在节假日期间用水量整体出现下降趋势。

5.2 问题二：盲源分离漏损量化指标来分析漏水模式

对于供水管网漏损量化方法上，有夜间最小流量法和盲源分离等方法，但是夜间最小流量法是以人为估算为准的漏损量化方法，误差大。而本文采用的盲源分析可以快速、精准的计算漏损量，并且为漏损控制项目的可行性分析提供可靠数据支持。

夜间最小流量法需要建立在独立计量区域（DMA）的基础上，通过对夜间流量进行分析，利用差值法减去独立计量区域夜间真实用水量，近似得到独立计量区域的实际漏损量，进而评估该区域的实际漏损情况。夜间流量只有漏损量是小概率事件，几乎不发生。一般情况下，夜间最小流量可能包括区域的工业用水、居民用水、背景漏损、管道漏损、非法用水和其他合法用水，视区域供水性质而定。当夜间最小流量得到时，只能凭借经验和对历史数据的分析来确定每个组分所占的百分比，从而确定独立计量区域的实际漏损量，存在一定的局限性。

盲源分离法是在供水管网安装流量计和压力计的情况下，使用 FastICA 算法计算漏损量，在独立计量区域（DMA）中将实际漏损量从流量计用水量分离出来，取得很好的结果。

5.2.1 盲源分离漏损量化模型的建立

盲源分离是从供水管网入口流量和入口压力信号中分离得到漏损量。采用线性瞬时混合方式建立盲源分离漏损量化混合模型如下：

$$x(t) = As(t) + n(t)$$

供水管网入口流量和管网入口压力信号是管网用水量和漏损量两个信号经过混合矩阵 A 混合得到的，用水量和漏损量是未知变量，混合矩阵 A 为未知矩阵，观测到的管网入口流量和管网入口压力为已知量。盲源分离漏损量化分解模型如下式：

$$y(t) = Wx(t)$$

供水管网盲源分离漏损量化模式，如图 2.1：

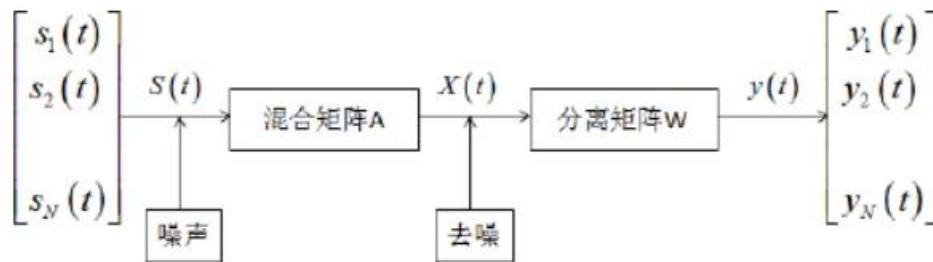


图 2.1 盲源分离漏损量化模式图

盲源分离就是在约束条件下从观测信号中得到分离矩阵 W ，并恢复出源信号的过程。

5.2.1.1 供水管网应用盲源分离的依据及验证

盲源分离之前的已知条件仅有混合信号或是观测信号，如果没有分析信号的先验知识，盲分离问题通常无解。盲源信号分析时，通常做如下四种假设，以使盲源分离有解。

- 1) 混合信号的各源信号均值均为零，并且是实随机变量，源信号之间相互独立；
- 2) 源信号的个数不应大于混合信号或观测信号的个数，一般取观测信号的个数等于源信号个数；
- 3) 混合矩阵 A 是可逆的；
- 4) 各源信号中最多有一个服从高斯分布，多个高斯信号的叠加仍然是高斯信号。

5.2.1.2 供水管网应用盲源分离的可行性条件

城市供水管网可以应用盲源分离的依据如下

- (1) 用水量和漏损量相互独立供水管网中的节点流量包含用水量和漏损量两部分。根据已知文献对漏损量与压力的关系及用水量与压力的关系可知，节点压力大于临界压力时管网正常供水，用水量和压力不存在函数关系，漏损量和压力呈函数关系。因此，供水管网的用水量和漏损量是相互独立的。
- (2) 非高斯性验证源信号中至多一个服从高斯分布，即用水量和漏损量中至少有一个服从非高斯分布。数据的非高斯性度量准则有峭度和负熵。峭度是归一化的

4 阶中心矩。独立计量区域（DMA）下用水量是否服从高斯性采用峭度法，峭度公式如下：

$$K = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^4 = \frac{1}{\sigma^4 N} \sum_{i=1}^N (x_i - \bar{x})^4$$

当 $K=3$ 时，曲线为零峭度，说明用水量服从高斯分布；当 $K>3$ 时，曲线为正峭度； $K<3$ ，曲线为负峭度。 K 值越大，曲线分布的最高处越高于高斯分布曲线，否则，越接近于非高斯分布。

5.2.2.3 供水管网用水量的高斯性验证

分别对城市供水管网的独立计量区域（DMA）中用水量和漏损量数据验证是否是高斯性问题。

由于漏水量数据未知，我们采用夜间最小流量法进行粗略的估算，采取每天夜间 2 点到 5 点数据进行估算。得到的结果是： $K_1 = 0.725287548134487$ 。

对于用水量数据也是未知的情况，我们采用在仪表测得的数据减去漏水量估算的均值作为用水量数据进行验证，结果是： $K_2 = 2.835087868107973$ 。

有这两个结果可以看出：供水管的用水量是更加近似服从高斯分布的，而漏水量是服从非高斯分布，符合盲源分离的条件。

5.2.2.4 盲源分离漏损量化算法

针对建立的盲源分离漏损量化模型选择快速独立分量分析算法，采用 R 语言编制算法求解程序（具体程序见附录 Model2_1.m 文件），对算法中涉及的参数在求解中进行优化，求得供水管网漏损量。独立分量分析算法是基于源信号相互独立情况下的寻优算法，当盲源分离用水量和盲源分离漏损量相互独立时，算法目标函数达到最大值或最小值。

5.2.3 基于盲源分离的漏损量化计算和分析

首先对于 DMA 分区入口流量仪表测得的分别是管网入口流量和入口压力，如图 2.2 和 2.3。

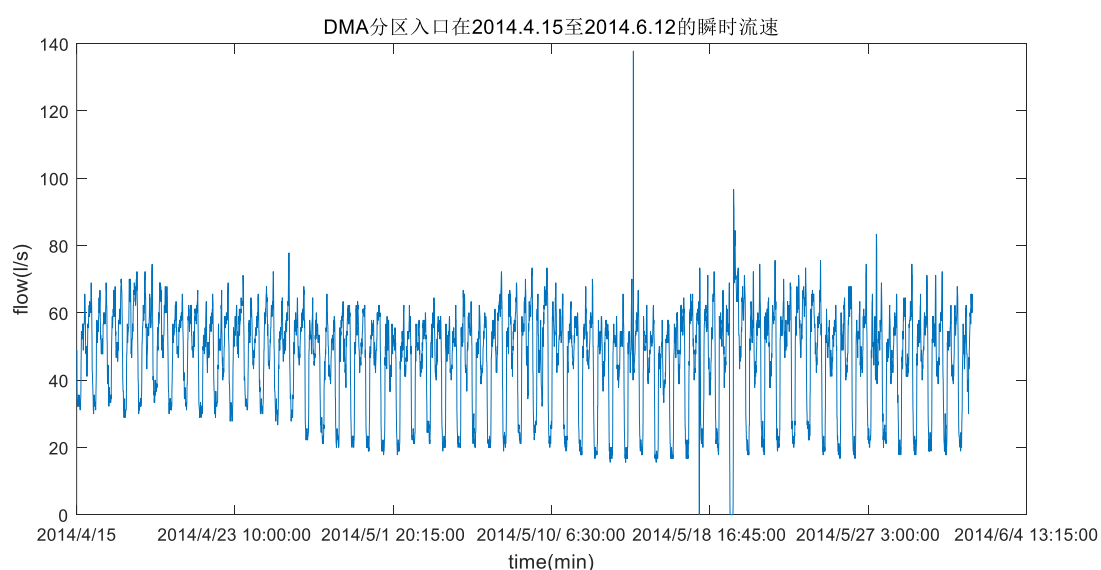


图 2.2 管网入口流量示意图

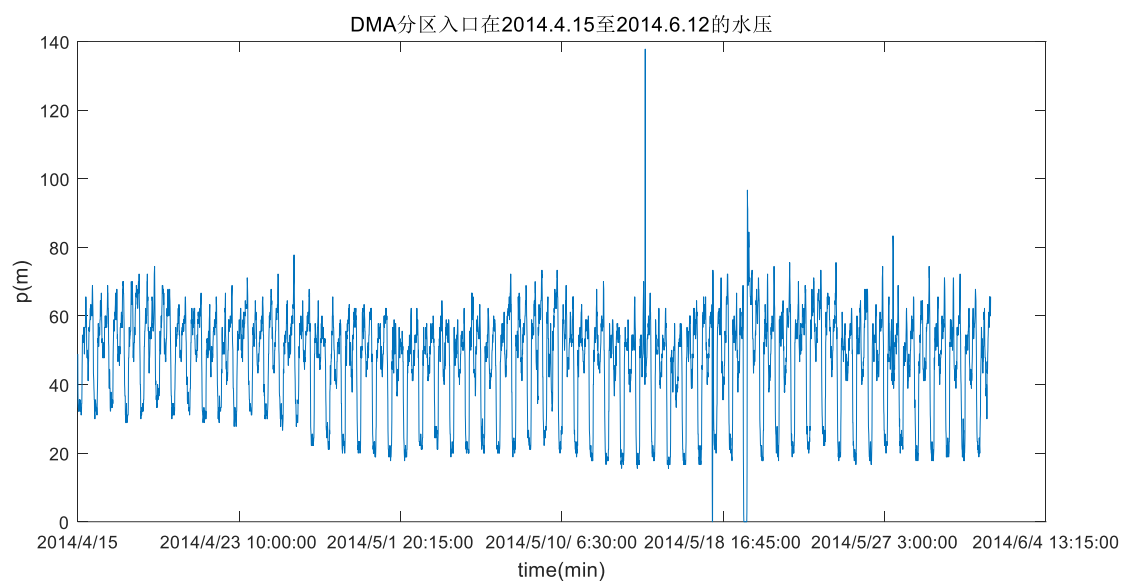


图 2.3 管网入口压力示意图

利用盲源分离独立分量分析算法从管网入口流量和入口压力两个混合信号中分离得到分离信号 1 和分离信号 2 的变化趋势，如下图 2.4 和 2.5.

分离信号1

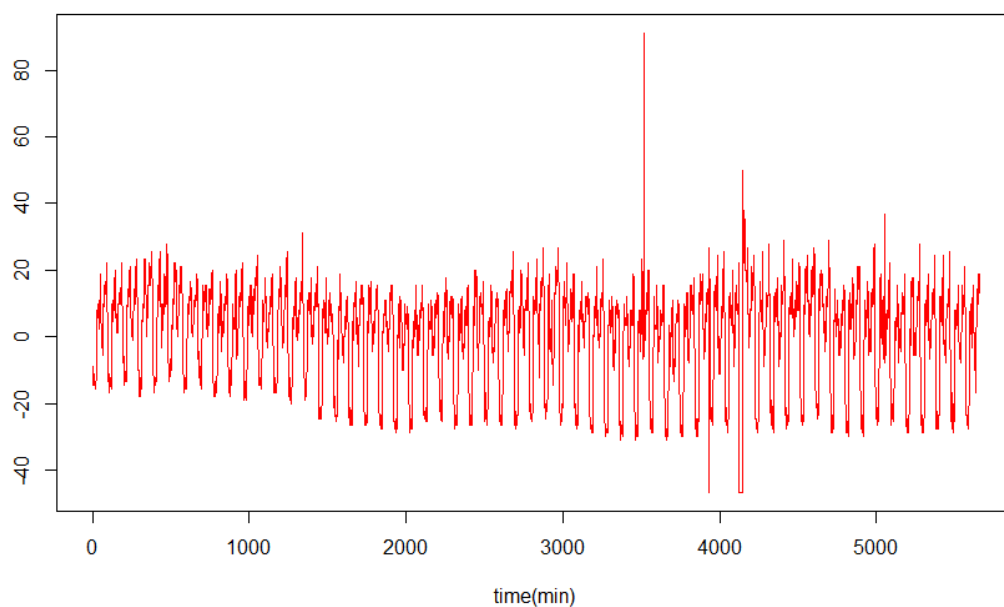


图 2.4 分离得到分离信号 1

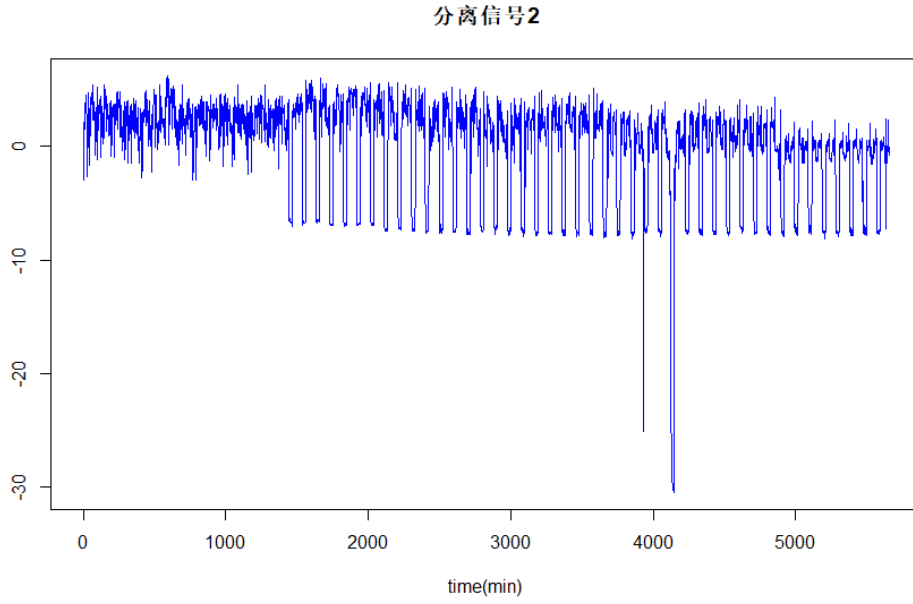


图 2.5 分离得到分离信号 2

并且我们得到了盲源分离的混合矩阵 A 和分离矩阵 W ：

$$A = \begin{bmatrix} 12.88781 & 4.160153 \\ -6.94783 & 1.143649 \end{bmatrix} \quad W = \begin{bmatrix} -0.8998631 & 0.4361725 \\ 0.4361725 & 0.8998631 \end{bmatrix}$$

由于供水管网入口流量信号和入口压力信号中都包含用水量和漏损量信号，根据用水量和总供水量趋势相似性，认为分离信号 1 为盲源分离用水量信号；考虑到算例管网流量信号和压力信号中无系统噪音和环境噪音，分离信号 2 为盲源分离漏损量。

假设盲源分离得到的用水量信号、漏损量信号变化趋势与供水管网实际用水量与漏损量变化趋势及相对幅度具有一致性。那么，对分离得到的用水量和漏损量信号做如下式处理：

$$\begin{cases} y_q(t) = y(t) - \bar{y}(t) \\ l_q(t) = l(t) - \bar{l}(t) \end{cases}$$

设供水管网实际用水量均值为 μ_y ，标准差为 σ_y ，实际漏损量均值为 μ_l ，标准差为 σ_l ，那么下式成立：

$$\begin{cases} Q_y(t) = y_q(t) \cdot \sigma_y + \mu_y \\ Q_l(t) = l_q(t) \cdot \sigma_l + \mu_l \end{cases} \quad (*)$$

将实际用水量的均值 μ_y 、标准差 σ_y 及实际漏损量的均值 μ_l 、标准差 σ_l 看作是未知数，由流量平衡关系得每一时刻供水管网流量满足方程：

$$\begin{cases} Q_s(t_1) = y_q(t_1) \cdot \sigma_y + \mu_y + l_q(t_1) \cdot \sigma_l + \mu_l \\ Q_s(t_2) = y_q(t_2) \cdot \sigma_y + \mu_y + l_q(t_2) \cdot \sigma_l + \mu_l \\ \dots \\ Q_s(t_n) = y_q(t_n) \cdot \sigma_y + \mu_y + l_q(t_n) \cdot \sigma_l + \mu_l \end{cases}$$

管网实际总供水量、去均值后的盲源分离用水量及去均值后的漏损量为已知量，根据最小二乘法利用 MATLAB 软件编程解上述超定方程组可以解决盲源分离用水量和漏损量幅值不确定问题。通过解超定方程组得实际用水量均值、标准差及实际漏损量均值、标准差的最优解。分别为：

$$\sigma_y = 1.000000000000002, \quad \mu_y = 23.32503093512463,$$

$$\sigma_l = -1.149087487086998e-14, \quad \mu_l = 23.325030935124637。$$

将最优解带入式（*）得到解决幅值问题后得到环状管网单点漏损的盲源分离用水量和盲源分离漏损量。如下图 2.6 所示：

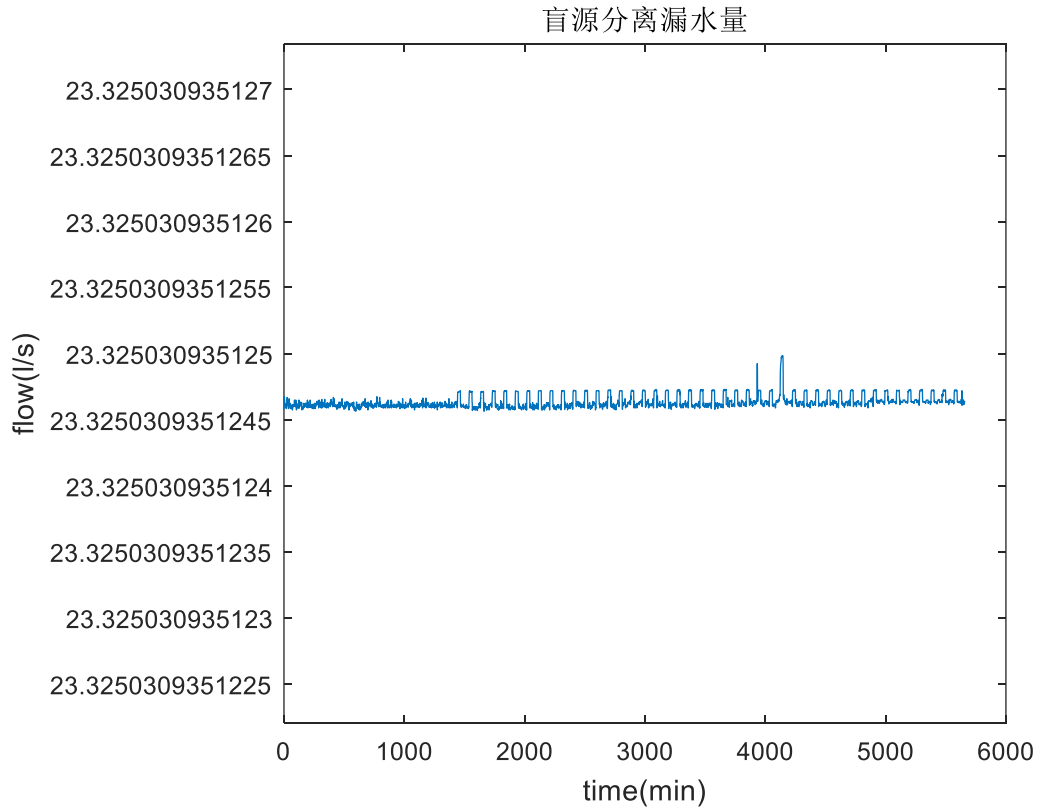


图 2.6 环状管网单点漏损的盲源分离用水量和盲源分离漏损量

5.2.2 模型结果

对漏损量的图像分析可知，在此独立计量区域（DMA）中漏损量几乎是恒定的，但是在后面时间段上出现了波动，可能是周期性的停水导致，还有在某些时间段上出现异常，可能是爆管后的停水检修，对于这些异常我们会在接下来讨论。

5.3 问题三：根据局部离群因子（LOF）进行异常模式识别

5.3.1 数据分析

(1) 数据分段处理

为了简明、针对性分析异常模式，进而针对性控制管理。我们对瞬时流量数据进行了分段预处理，根据第一问的分析过程，得到典型用水模式具有周期性，且在工作日和非工作日没有明显的区别。因此我们以一天为周期来分析异常现象。在一张图上表示出每天的瞬时流量随时间的变化（见图 3.1）。具体代码见附录文件 Model1_2.m.

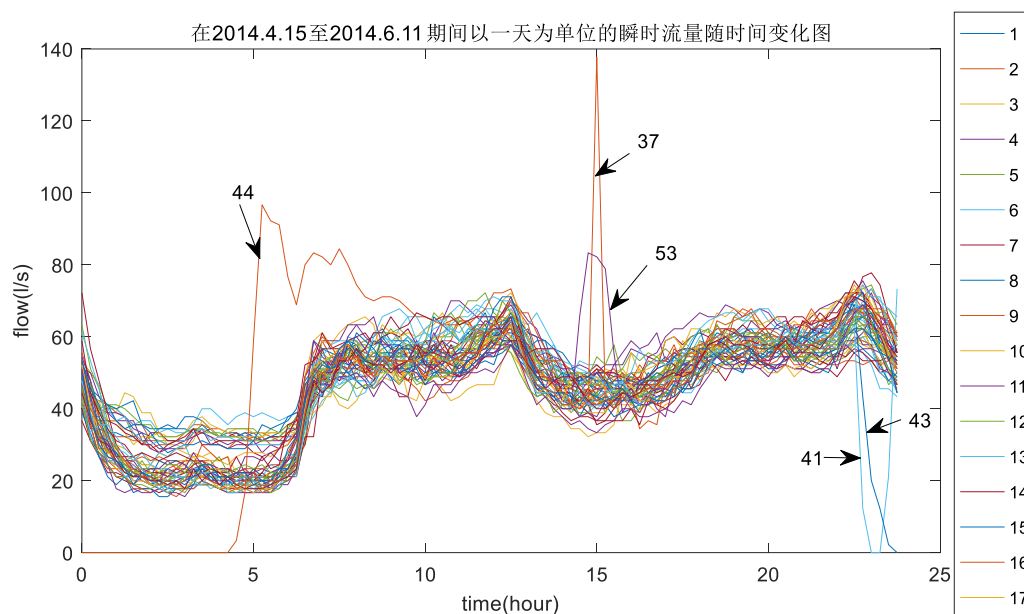


图 3.1 每天的瞬时流量随时间的变化图

从上图中，我们可以粗略地得到疑似异常现象出现的时间，即 5 月 21 日(37)、5 月 25 日(41)、5 月 27 日(43)、5 月 28 日(44) 和 6 月 6 日(53) 六个时间段。

(2) 非节假日和节假日数据处理

由问题一的分析过程可知：节假日(五一)和非节假日的用水模式存在差别，我们绘制节假日的日用水量与非节假日的日用水量平均值的图像，来判断差异性如下所示（见图 3.2），程序见附录文件 Model1_2.m:

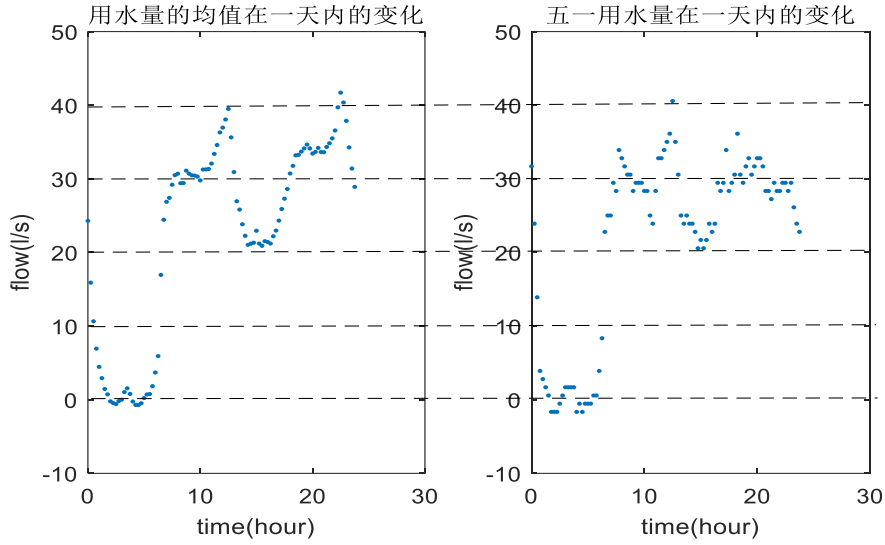


图 3.2 节假日（五一）和非节假日的时间-用水量散点图

通过上图初步分析，我们发现节假日（五一）和非节假日的每一时刻用水量大致上趋势一样，整体上节假日用水量较非节假日用水量少些。但不存在明显的差异。因此，为了方便统计与参数分析，在下面分析过程中，不对节假日（五一）和非节假日的数据进行差异处理，考虑节假日属于正常用水模式。

5.3.2 局部离群因子分析模型的建立

(1) 理论基础

LOF(对象的局部异常因子)算法中每个数据都被分配一个局部异常因子，局部异常因子愈大，就认为它更可能是一种异常；反之则可能性小。计算局部异常因子，先产生所有数据点的 k -邻域(同时得到 k -距离)，并计算到其中每个点的距离。算法流程如下：

a、计算对象 p 的 k -距离

对任意的自然数 k ，定义 p 的 k -距离（ $k-distance(p)$ ），为 p 和某个对象 o 之间的距离，这里的 o 满足：至少存在 k 个对象 $o' \in D \setminus \{p\}$ ，使得 $d(p, o') \leq d(p, o)$ ，并且至多存在 $k-1$ 个对象 $o' \in D \setminus \{p\}$ ，使得 $d(p, o') < d(p, o)$ 。

b、计算对象 p 的 k -距离邻域（ $N_{k-distance}$ ）

$$N_{k-distance}(p) = \{q \mid d(p, q) \leq k-distance(p)\}$$

c、计算对象 p 相对于对象 o 的可达距离

给定自然数 k ，对象 p 相对于对象 o 的可达距离为：

$$reach-dist_k(p, o) = \max\{k-distance(o), d(p, o)\}$$

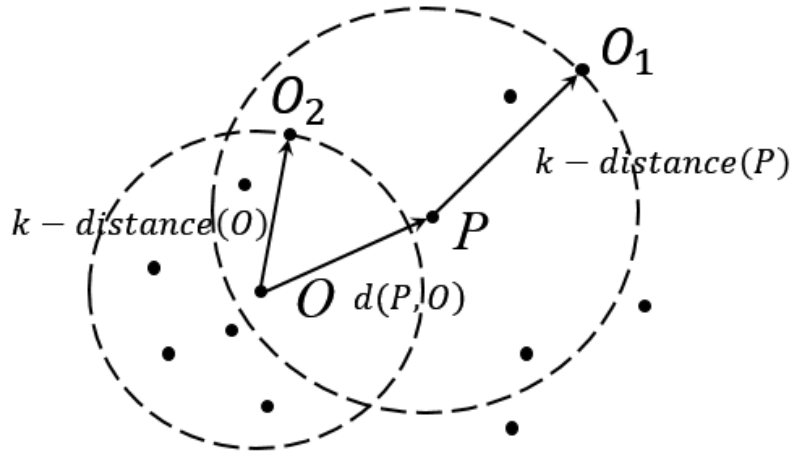


图3.3 可达距离示意图

d、计算对象 p 的局部可达密度(Local Reachable Distance)

对象 p 的局部可达密度为对象 p 与它的 k -邻域的平均可达距离的倒数。

$$lrd_k(p) = 1 / \left[\frac{\sum reach-dist_k(p, o)}{|N_k(p)|} \right]$$

e、计算对象 p 的局部异常因子

$$LOF_k(p) = \sum_{o \in N_k(p)} \frac{lrd_k(o)}{lrd_k(p)} / |N_k(p)|$$

LOF提取（详细代码参见附录LOF.m文件）后，对于下一问的故障判断，我们还会采用移动平均线结合标准差的办法加强模型稳定性。

(2) 用水量异常模式检测模型

在分区用户用水瞬时流量分析中，我们初步得到疑似异常现象出现的时间，即5月21日（37）、5月25日（41）、5月27日（43）、5月28日（44）和6月6日（53）六个时间段，作图代码参见附录Model3_1.m文件。

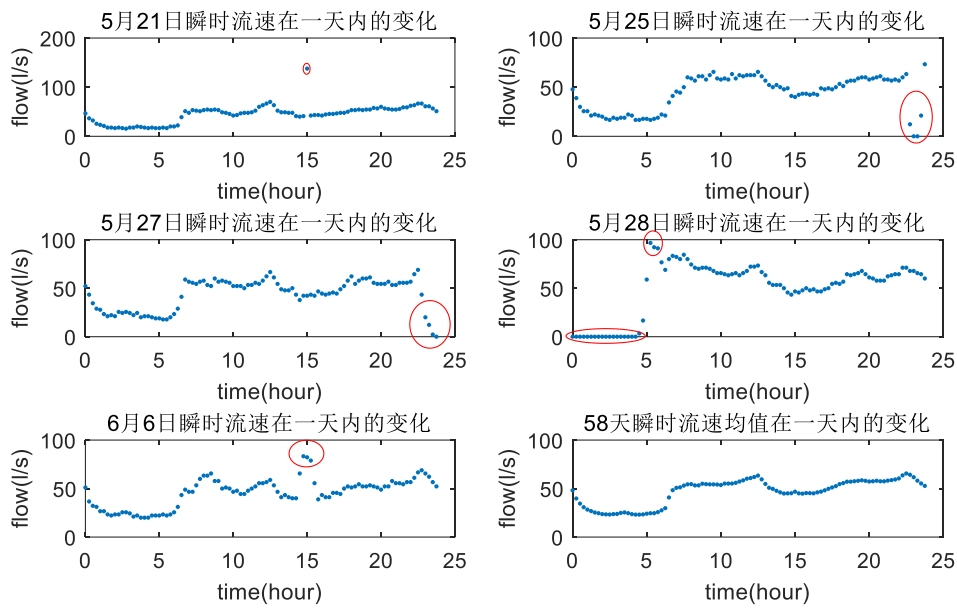


图3.4 疑似异常现象出现的时间及瞬时流速均值在一天内的变化

LOF检验离群点需要确定 k 值， k 值取太小会使得平缓部分的一部分疑似异常点被当成正常点，如图3.5左图所示。 k 值取太大又会使得上升下降阶段时LOF值较敏感，如图3.5右图所示，具体代码参见附录Model3_1.m文件：

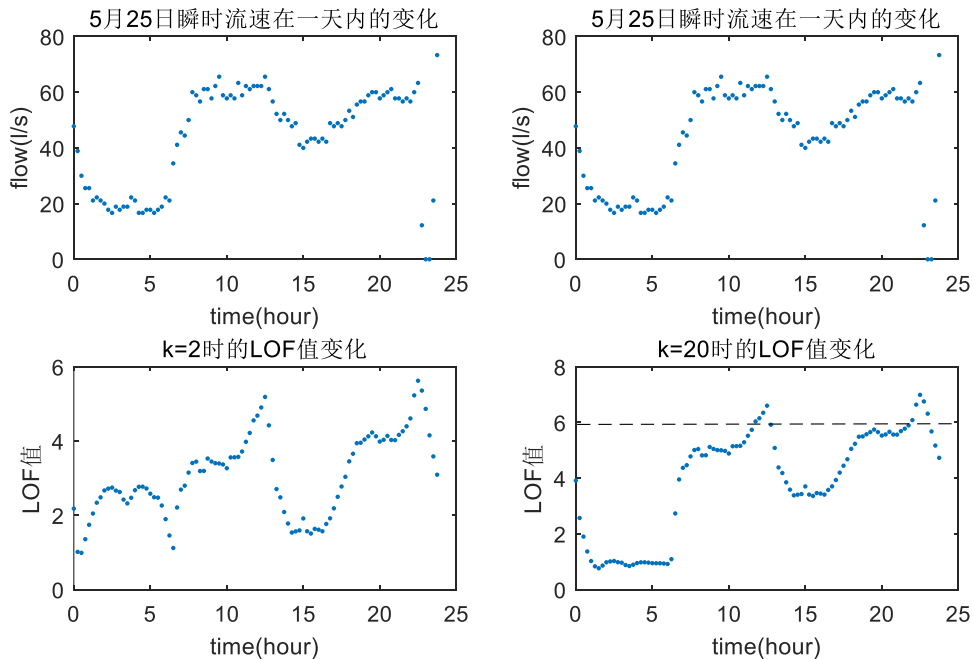


图3.5 不同 k 值下LOF的变化图

经过 k 取值的分析对比和本题的数据分布，我们采用 $k=20$ 作为邻近对象。

一般情况下，我们认为异常因子接近1的点，说明它和周围点的密度一致，判定为正常；异常因子越小说明它和周围点的密度相差越大，因此，成为异常点的可能性也就越大。由于本题的数据分布特点，我们发现不能直接定下一个LOF阈值（如 $LOF > 6$ 算作疑似异常点）作为判断依据，如图3.5， $k=20$ ，我们选取LOF

值等于6作为阈值观察情况。从上图中我们可以看到，虽然在23小时之后时间段也有一部分离群点，但是对于 $LOF > 6$ 所对应的时间段的离群点来说LOF值较小。我们如果直接利用阈值划分来判断是否为离群点，会将部分离群点直接忽略。因此我们引入移动平均+标准差的阈值划分方法，针对不同时间段的分布特点能够分别进行判断。

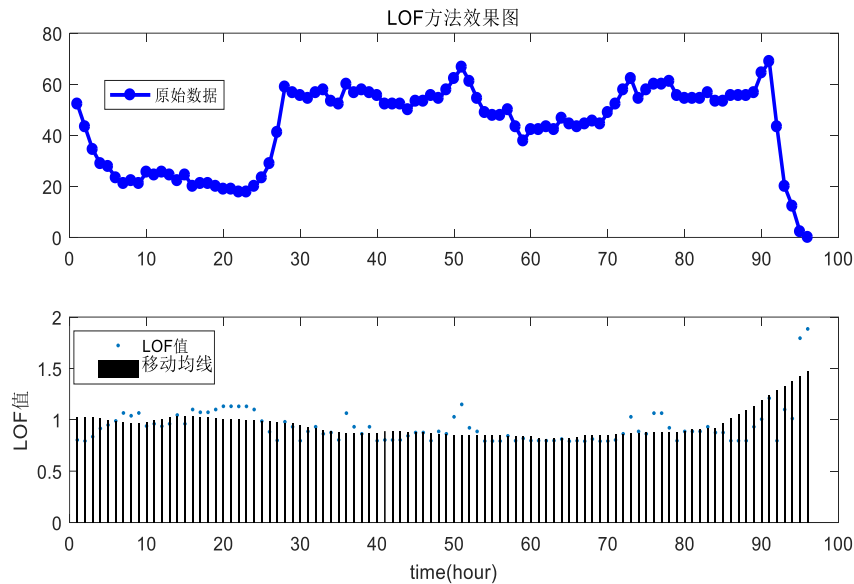


图3.6 移动平均处理后的图像

移动平均法(moving average method)是根据时间序列,逐项推移,依次计算包含一定项数的序时平均数,以此进行预测的方法。如图3.6所示；移动均线与标准差的结合为根据数据呈现的时间序列，计算一段项数的实时标准差，加上序时平均数得到“均值+标准差”图像。如图3.7所示。

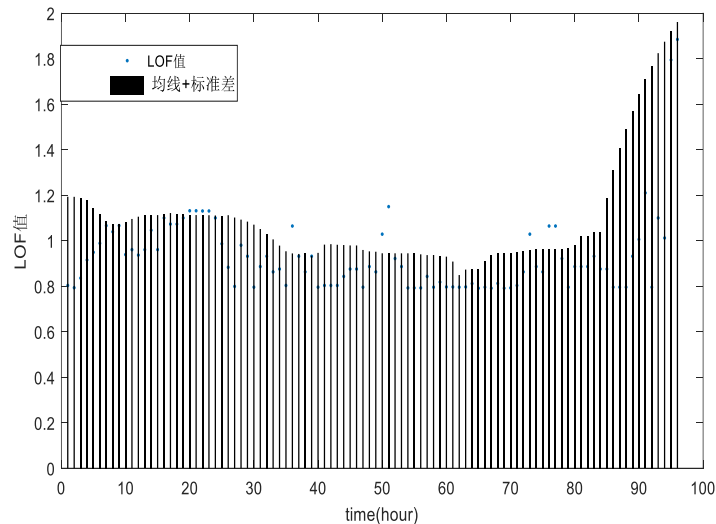


图3.7 移动标准差后的图像

我们还需要考虑到，离群有向上离群和向下离群两种情况，在判断瞬时流量是否异常时，我们还需要考虑瞬时流量随时间序列变化的图像。我们发现用函数拟合的效果不理想，因此我们还是采用移动均线+标准差来筛选离群点，如下图

3.8所示:

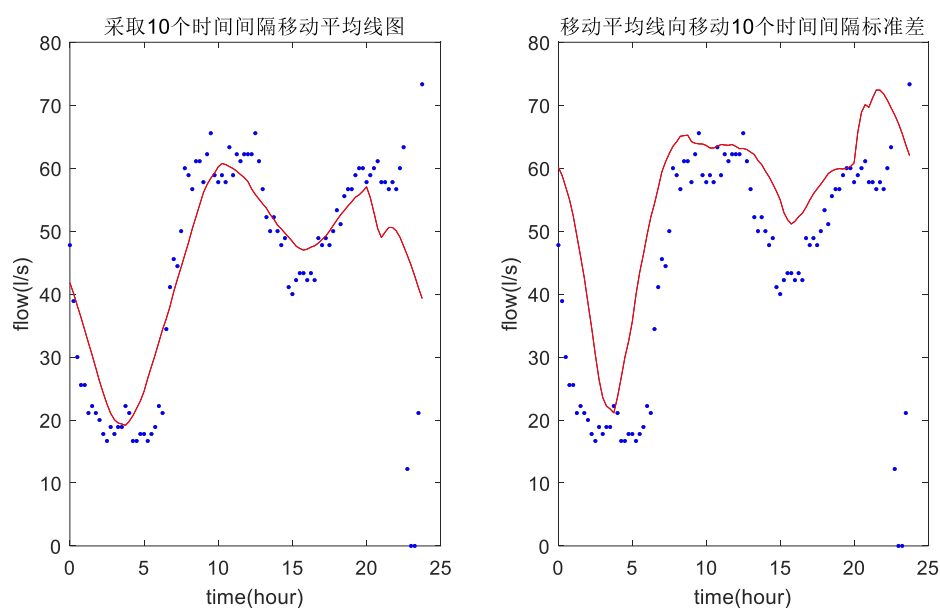


图3.8 瞬时流量随时间序列变化的移动均线及移动均线+标准差图像

同时位于LOF移动均线标准差偏移图像上方以及瞬时流量时间序列移动平均线下方的数据点即为疑似异常点。根据上面的算法实现过程，给出疑似异常模式出现时间：5月21日（37）（见图3.9）、5月25日（41）（见图3.10）、5月27日（43）（见图3.11）、5月28日（44）（见图3.12）和6月6日（见图3.13）六个时间段的最终异常点示意图。具体代码参见附录method_lof.m文件。

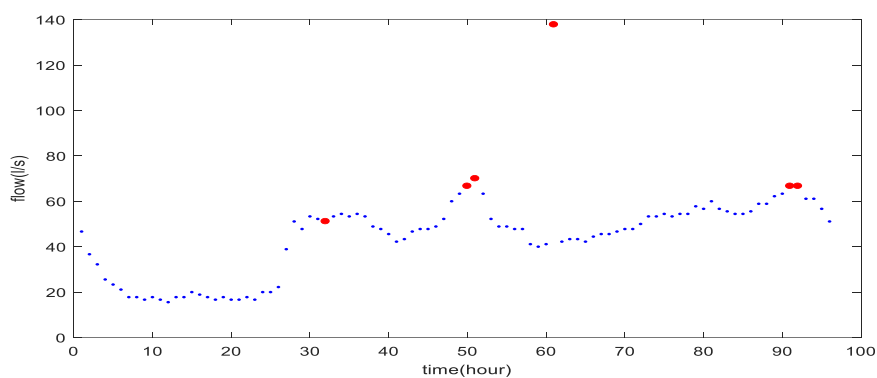


图3.9 5月21日最终异常点（红色标记）

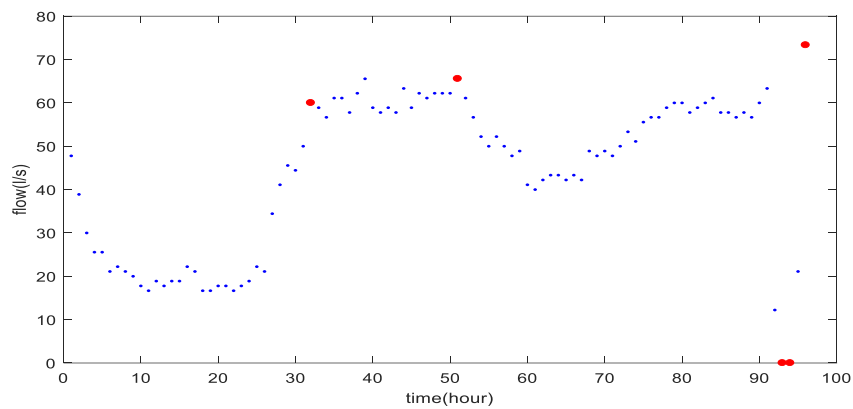


图3.10 5月25日最终异常点（红色标记）

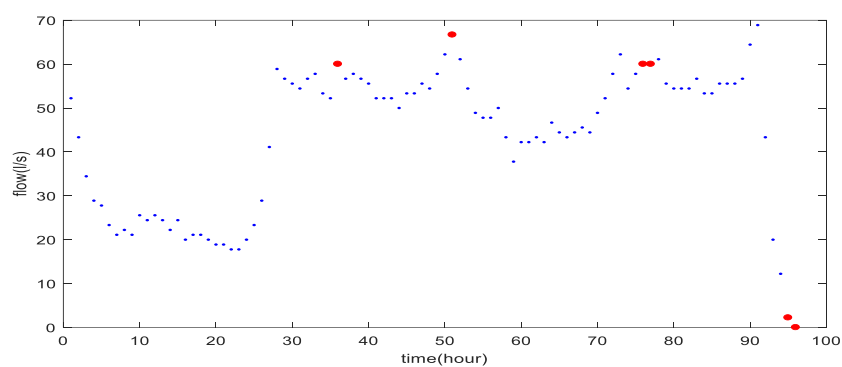


图3.11 5月27日最终异常点（红色标记）

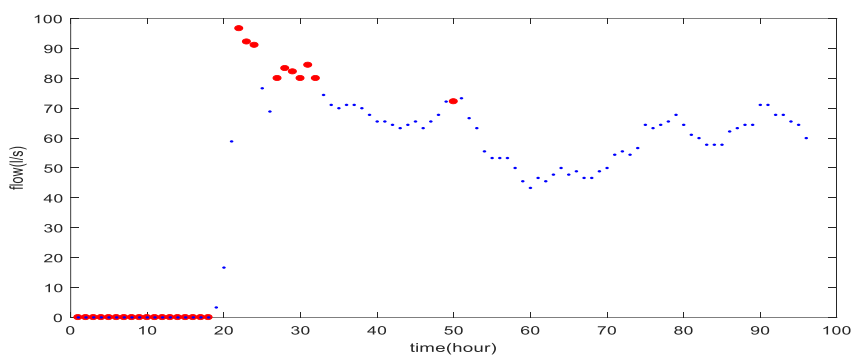


图3.12 5月28日最终异常点（红色标记）

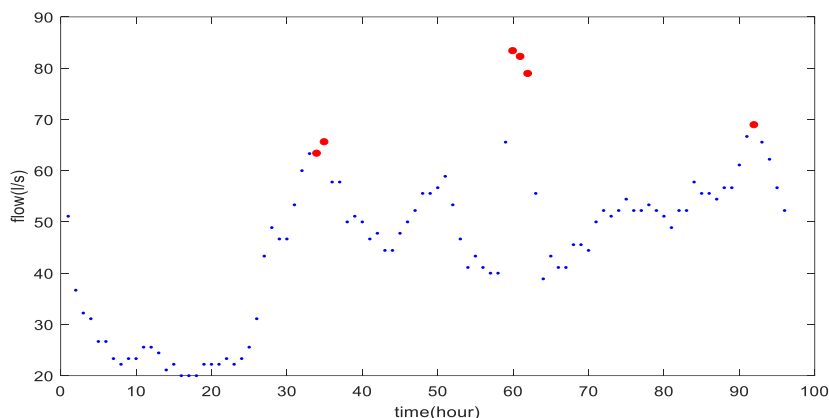


图3.13 6月6日最终异常点（红色标记）

5.3.3 模型结果

从最终异常点示意图可得：由粗略观察得到的疑似异常时间段：5月21日（37）、5月25日（41）、5月27日（43）、5月28日（44）和6月6日（53）确为异常现象发生时间。且可以从中识别出不同的异常模式，即5月21日（37）、5月25日（41）、5月27日（43）这三个时间段异常模式为特定时刻的水量异常增加（短期的，如爆管）；5月28日（44）和6月6日（53）这两个时间段异常模式为节律变化（用水高峰期不符合生活或生产节律）（如偷用生活用水做商业用途）。

6、模型评价与展望

6.1 创新点和模型改进

创新点：

- (1) 采用层层递进、逐步细化的数据分析模式，层次明确、条理清晰且全面地对该区用户用水模式进行分析与建模。
- (2) 利用盲源分离方法将漏水量实现了实时量化的目的，进而更加精准、有效地反映该区用户水管网的漏水模式。
- (3) 考虑利用LOF移动均线标准差偏移图像和瞬时流量时间序列移动平均线来综合反映异常点，进而分局异常点的分布来得到不同的异常模式。

模型改进：

- (1) 对用水量进行的四次高斯拟合，其效果不是很好（拟合的残差平方值为0.9537），可以考虑分段拟合思想来提高拟合效果。
- (2) 对于得到的理想的典型用水模式，可以考虑用监督型学习的模式识别，来更准确反映异常模式。
- (3) 在对漏水模式识别解超定方程组，可以采用奇异值分解的方法得到更可靠的结果。

7 参考文献

- [1] 程桂, 李翠梅, 郜阔. 城市行业用水量变化与用水模式实例研究[J]. 深圳大学学报理工版, 2016,33(1):49-53
- [2] 谢善斌, 金波, 刘遂庆. 基于不确定性度量的给水管网用户的用水模式分析[J]. 中国给水排水, 2011, 27(7) :59-62.
- [3] 马海宁. 给水管网水力建模的用水模式研究[J]. 中国给水排水, 2008.34(1):109-112
- [4] 李娟娟. 基于小波变换的供水管网漏损量化方法研究[D]. 哈尔滨工业大学. 2015.6: 26-28
- [5] Vanzyl J.E,Cassa A.M. Modeling elastically deforming leaks in water distribution pipes[J].American Society of Civil Engineers,2014,140:182-189
- [6] MM BREUNIG, H-P KRIEGEL. LOF: identifying density-based local outliers[C]. Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Boston: ACM Press, 2009, 93-104. 39.
- [7] 储著敏, 顾思远, 赵东阳. ATM交易状态特征分析与异常检测[C]. 2017年“深圳杯”数学建模论文集, 2017,13-18

附录

附录1 Modell_1.m

```
Shuju=xlsread('2018年东北三省数学建模联赛F题附件');
flow = shuju(:,1);
x = (1:5657)';
t = {'2014/4/15','2014/4/23 10:00:00','2014/5/1 20:15:00','2014/5/10/ 6:30:00','2014/5/18
16:45:00','2014/5/27 3:00:00','2014/6/4 13:15:00','2014/6/12 23:30:00'};
plot(x,flow)
set(gca,'XTickLabel',t)
title('DMA分区入口在2014.4.15至2014.6.12的瞬时流量')
xlabel('time(s)')
ylabel('flow(l/s)')
```

附录2 Modell_2.m

```
%% 读取数据
data = xlsread('2018年东北三省数学建模联赛F题附件');
flow = data(:,1); %流量
p2 = data(:,2); %压力
%% 每天的流量
for i = 1:58
    flow_everyday(i,:)=flow((i-1)*96+1:i*96,1);
    plot(0:0.25:23.75,flow_everyday(i,:));
    hold on;
end
hold off;
title('在2014.4.15至2014.6.11期间以一天为单位的瞬时流量随时间变化')
xlabel('time(hour)')
ylabel('flow(l/s)')
c=1:58;
LA=sprintfc('%g',c);
legend(LA)
%% 2014.5.5-2014.5.14七天连续数据
plot(flow(1920:2880,1));
hold on;
plot(p2(1920:2880,1));
hold off;
%% 减去漏水量的用水量
for i = 1:58
    yejian(i,:) = flow((i-1)*96+9:(i-1)*96+21,1);
end
loushui = mean(mean(yejian)); % 夜间平均漏水
x = 0:0.25:23.75; %时间
y = mean(flow_everyday)-loushui; %减去漏水量的用水量
%% 节假日（五一）和非节假日的用水模式存在差别
```

```

subplot(2,2,1)
plot(x,y,'.')
title('用水量均值在一天内的变化')
xlabel('time(hour)')
ylabel('flow(l/s)')
subplot(2,2,2)
y1=flow_everyday(16,:)-loushui;
plot(x,y1,'.')
title('五一用水量在一天内的变化')
xlabel('time(hour)')
ylabel('flow(l/s)')
%% 高斯拟合
% GMM2.sfit    cftool工具箱
%% 4月21日至4月27日七天数据
for i = 1:7
    flow_week(i,:)=flow((i-1)*96+577:i*96+576,1);
end
t=0:0.25:23.75;
plot(t,flow_week(1,:),'-',t,flow_week(2,:),':',t,flow_week(3,:),'-+',t,flow_week(4,:),'-*',t,flow_week(5,:),'-o',t,flow_week(6,:),'-s',t,flow_week(7,:),'-d')
legend('Monday','Tuesday','Wednesday','Thursday','Friday','Saturday','Sunday')
title('2014.5.21至2014.5.27一周的瞬时流量随时间的变化趋势')
xlabel('time(hour)')
ylabel('flow(l/s)')
%% 相关性分析
% 见1.R
%% p>0.05部分的差异图
% 1&3周
x=0:0.25:167.75;
plot(x,flow(1:672,1))
hold on
plot(x,flow(1249:1920,1))
hold off
legend('第一周','第三周')
title('第一周和第三周瞬时流量随时间变化趋势图')
xlabel('time(hour)')
ylabel('flow(l/s)')
% 2&4周
plot(x,flow(673:1344,1))
hold on
plot(x,flow(1921:2592,1))
hold off
legend('第二周','第四周')
title('第二周和第四周瞬时流量随时间变化趋势图')

```

```

xlabel('time(hour)')
ylabel('flow(l/s)')
% 3&5周
plot(x,flow(1249:1920,1))
hold on
plot(x,flow(2593:3264,1))
hold off
legend('第三周','第五周')
title('第三周和第五周瞬时流量随时间变化趋势图')
xlabel('time(hour)')
ylabel('flow(l/s)')
% 4&6周
plot(x,flow(1921:2592,1))
hold on
plot(x,flow(3265:3936,1))
hold off
legend('第四周','第六周')
title('第四周和第六周瞬时流量随时间变化趋势图')
xlabel('time(hour)')
ylabel('flow(l/s)')

```

附录3 Model1_3.R

```

#=====相关分析=====
install.packages("psych")
library(psych)
data <- read.csv('D:\\数学建模\\data\\F\\data.csv',header = F)
B <- as.vector(data$V2)
C <- B[1:5568]
D <- matrix(data = C,ncol = 8,byrow = F)
E <- as.data.frame(D)
corr.test(E)
#=====盲源分离=====
install.packages("fastICA")
library(fastICA)
data <- read.csv('D:\\数学建模\\data\\F\\data.csv',header = F)
F <- as.matrix(data[,2:3])
G <- fastICA(F,2) #盲源分离，列表
X <- G[["X"]] #分离的信号
A <- G[["A"]]
W <- G[["W"]]
X1 <- as.vector(X[,1]) #分离信号1
X2 <- as.vector(X[,2]) #分离信号2
#par(mfrow=c(2,1)) #画图
plot.ts(X1,col = "red",main="分离信号1", ylab="", xlab="time(min)")

```

```
plot.ts(X2,col = "blue",main="分离信号2", ylab="", xlab="time(min)")
write.csv(X,file="D:\\数学建模\\data\\F\\mangyuan.csv")
```

附录4 LOF.m

```
function lof = LOF(x,K,dist)
m=size(dist,1); %m为对象数，dist为两两之间的距离
distance = zeros(m,m);
num = zeros(m,m); %distance和num用来记录排序后的顺序，和对象编号顺序
kdistance = zeros(m,1); %计算每个对象的kdistance
count = zeros(m,1); %k领域的对象数
reachdist = zeros(m,m); %计算两两之间的reachable-distance
lrd = zeros(m,1);
lof = zeros(m,1);
%计算k-distance
for i=1:m
    [distance(i,:),num(i,:)] = sort(dist(i,:), 'ascend');
    kdistance(i) = distance(i,K+1);
    count(i) = -1; %自己的距离为0，故去掉
    for j = 1:m
        if dist(i,j) <= kdistance(i)
            count(i) = count(i)+1;
        end
    end
end
for i = 1:m
    for j=1:m
        reachdist(i,j) = max(dist(i,j),kdistance(j));
        reachdist(j,i) = reachdist(i,j);
    end
end
for i = 1:m
    sum_reachdist=0;
    for j=1:count(i)
        sum_reachdist=sum_reachdist+reachdist(i,num(j+1));
    end
    %计算每个点的lrd
    lrd(i)=count(i)/sum_reachdist;
end
%%得到lof值
for i=1:m
    sumlrd=0;
    for j=1:count(i)
        sumlrd=sumlrd+lrd(num(j+1))/lrd(i);
    end
end
```



```

        lof(i)=sumlrd/count(i);
    end
% %画图
subplot(2,1,1)
plot(1:size(x,1),x,'-b.','linewidth',2,'markersize',14.5);
legend('原始数据');
title('LOF方法效果图');
set(legend,'location','best');
subplot(2,1,2)
plot(1:size(lof,1),lof,'-r.','linewidth',2,'markersize',14.5);
legend('LOF值');
set(legend,'location','best');
end

```

附录5 Model2_1.m

```

%% 盲源分析
% flow_yongshui 峭度计算
flow_yongshui = flow - loushui;
a1=mean(flow_yongshui);
b1=(flow_yongshui-a1).*(flow_yongshui-a1).*(flow_yongshui-a1).*(flow_yongshui-a1);
f1=sum(b1);
c1=f1/5657;
e1=std(flow_yongshui,1).*std(flow_yongshui,1).*std(flow_yongshui,1).*std(flow_yongshui,1);
d1=c1/e1; %接近3， 接近高斯
% p2峭度计算
flow_night = reshape(yejian,[],1); %夜间最小用水量
a2=mean(flow_night);
b2=(flow_night-a2).*(flow_night-a2).*(flow_night-a2).*(flow_night-a2);
f2=sum(b2);
c2=f2/5657;
e2=std(flow_night,1).*std(flow_night,1).*std(flow_night,1).*std(flow_night,1);
d2=c2/e2; %远离3， 非高斯
%分析过程及图
% 见Model1_3.R
% 导入数据
mangyuan = csvread('mangyuan.csv');
yt = mangyuan(:,2);
lt = mangyuan(:,3);
yt_mean = mean(yt);
lt_mean = mean(lt);
yq = yt - yt_mean;
lq = lt - lt_mean;
%解超定方程
one = ones(4,1);

```

```

A=[yq(200:203,:) one lq(200:203,:) one];
Q = flow(200:203,1);
X = pinv(A)*Q;
% 真实值
Qy = yq * X(1,1) + X(2,1);
Ql = lq * X(3,1) + X(4,1);
%漏损量图
plot(Ql);
title('盲源分离漏水量');
xlabel('time(min)');
ylabel('flow(L/s)')

```

附录6 Model3_1.m

```

x = 0:0.25:23.75;
subplot(3,2,1)
day37=flow_everyday(37,:);
plot(x,day37,'.')
title('5月21日瞬时流量在一天内的变化')
xlabel('time(hour)')
ylabel('flow(l/s)')
subplot(3,2,2)
day41=flow_everyday(41,:);
plot(x,day41,'.')
title('5月25日瞬时流量在一天内的变化')
xlabel('time(hour)')
ylabel('flow(l/s)')
subplot(3,2,3)
day43=flow_everyday(43,:);
plot(x,day43,'.')
title('5月27日瞬时流量在一天内的变化')
xlabel('time(hour)')
ylabel('flow(l/s)')
subplot(3,2,4)
day44=flow_everyday(44,:);
plot(x,day44,'.')
title('5月28日瞬时流量在一天内的变化')
xlabel('time(hour)')
ylabel('flow(l/s)')
subplot(3,2,5)
day53=flow_everyday(53,:);
plot(x,day53,'.')
title('6月6日瞬时流量在一天内的变化')
xlabel('time(hour)')
ylabel('flow(l/s)')

```

```

subplot(3,2,6)
y=mean(flow_everyday);
plot(x,y,'.')
title('58天瞬时流量均值在一天内的变化')
xlabel('time(hour)')
ylabel('flow(l/s)')
%%k值分析
subplot(2,2,1)
day41=flow_everyday(41,:);
plot(x,day41,'.')
title('5月25日瞬时流量在一天内的变化')
xlabel('time(hour)')
ylabel('flow(l/s)')
subplot(2,2,2)
day41=flow_everyday(41,:);
plot(x,day41,'.')
title('5月25日瞬时流量在一天内的变化')
xlabel('time(hour)')
ylabel('flow(l/s)')
dist = zeros(size(y,2));
for i = 1:size(y,2)
    for j = 1:size(y,2)
        dist(i,j) = norm(y(:,i)-y(:,j));
    end
end
end
lof1 = LOF(y,2,dist);
subplot(2,2,3)
plot(x,lof1,'.')
title('k=2时的LOF值变化')
xlabel('time(hour)')
ylabel('LOF值')
lof2 = LOF(y,20,dist);
subplot(2,2,4)
plot(x,lof2,'.')
title('k=20时的LOF值变化')
xlabel('time(hour)')
ylabel('LOF值')

```

附录7 method_lof.m

```

function [ lof_judge ] = method_lof( day,range_minute,D_volume,k)
%% lof方法判断
% 输入参数day—第i天
% range_minute—移动均线取前后分钟数
% D_volume—瞬时流量按天划分的数据

```

```

% k-lof中设定的k值
%输出参数lof_judge—移动平均判断结果
%%构造移动平均的数据
%加入前一天夜里与后一天凌晨的数据
if day==58 %数据的最后一天，取该天的凌晨代替后一天的凌晨
A=[D_volume(96-range_minute+1:96,day-1);D_volume(:,day);D_volume(1:range_minute,day)];
    else if day==1 %数据的第一天，取该天的夜里代替前一天的夜里
A=[D_volume(96-range_minute+1:96,day);D_volume(:,day);D_volume(1:range_minute,day+1)];
    else %加入前一天夜里与后一天凌晨的数据
A=[D_volume(96-range_minute+1:96,day-1);D_volume(:,day);D_volume(1:range_minute,day+1)
];
        end
    end
end
dist = zeros(size(A));
for i = 1:size(A,1)
    for j = 1:size(A,1)
        dist(i,j) = norm(A(i,:)-A(j,:));
    end
end
lof = LOF(A(1+range_minute:length(A)-range_minute),k,dist);
for i=1+range_minute:length(lof)-range_minute
    lof_ave(i-range_minute)=mean(lof(i-range_minute:i+range_minute)); %计算平均值
    lof_std(i-range_minute)=std(lof(i-range_minute:i+range_minute)); %计算标准差
end
%去除前一天夜里与后一天凌晨的数据
lof=lof(1+range_minute:length(lof)-range_minute);
lof_judge=zeros(length(lof),1);
for i=1:length(lof)
    if lof(i)>(lof_ave(i)+2*lof_std(i))&&lof(i)>1.5
        lof_judge(i)=1
    end
end
end
A=A(1+range_minute:length(A)-range_minute);
%%画图
%移动均线
plot(lof,'.')
hold on
plot(lof_ave,'o')
figure
plot(lof,'.')
hold on
plot(lof_ave+lof_std,'*')
%标记找出的点
figure

```

```

%% 对于连续零值现象的处理
% if (max(find(A==0))-min(find(A==0))+1)==size(find(A==0),1) && size(find(A==0),1)> k
%     plot(find(A==0),A(find(A==0)),r.','MarkerSize',15)
%     hold on
% end
for i=1:length(lof_judge)
    if lof_judge(i)==1
        %record=[record i];
        plot(i,A(i),r.','MarkerSize',15)
        hold on
    else
        plot(i,A(i),'b.')
        hold on
    end
end
end
%输出lof判断的天数
disp('---lof平均找出---')
sum(lof_judge);
end

```