

中国研究生创新实践系列大赛  
“华为杯”第十七届中国研究生  
数学建模竞赛

学 校 浙江财经大学

---

参赛队号 20114820008

---

1.赵宇峰

---

队员姓名 2.吴心思

---

3.琚悦琦

---

# 中国研究生创新实践系列大赛

## “华为杯”第十七届中国研究生

### 数学建模竞赛

题 目      大雾天气下的能见度估计与预测问题

---

#### 摘            要：

能见度对于航空安全和道路交通安全有着重要的影响，准确把握不同大雾情况下能见度的演化规律以及预测大雾消散的时间至关重要。针对该问题，本文运用不同的模型与思路进行分析、评估与预测，包括分位数非线性回归模型、基于卷积神经网络和迁移学习的深度学习模型、基于梯度导向滤波技术的暗通道先验算法、自回归移动平均模型（ARMA（3,3）模型）以及基于粒子群优化算法的非线性灰色伯努利模型（PSO-NGBM（1,1）模型）等。首先，分析了机场能见度与地面气象因素之间的关系并提供了能见度与影响因素的非线性关系式；其次，通过构建深度学习模型，结合机场视频数据，对能见度进行精度评估；接着，基于不依赖于能见度仪的观测数据，采用能见度估计算法并绘制高速公路能见度时间序列曲线，分析了高速公路能见度随时间变化规律；最后，通过不同的模型预测大雾变化趋势及消散时间。

对于问题一，我们采用了 Pearson 相关性矩阵、热力图和分位数非线性回归等方法，对能见度数据与气象观测（温度、湿度和风速等）变量数据进行相关性分析，筛选有效变量，从多个分位点角度，刻画了能见度与地面气象观测因素的非线性关系，并在调整  $R^2$  和各变量检验等统计评价指标最优和符合经济意义的多重目标下，选择分位数非线性回归的最优分位点，最终导出机场能见度的具体非线性关系式。

对于问题二，鉴于传统图像识别方法存在准确率低、泛化能力差、且耗费时间长等问题，本文采取一种基于卷积神经网络和迁移学习的深度学习模型，以实现对机场视频图像的识别与估计，并对估计的能见度进行精度评估。为突出迁移学习方法的优越性，我们分别建立了深度学习模型和机器学习模型。在传统图像识别方法上，识别图像 11 维信息特征，建立了 BP 神经网络模型与评估精度；在深度学习模型上，我们首先将机场视频与能见度数据进行匹配，通过预训练模型确定各参数，接着将图像与数据随机以 7:3 的比例划分为训练集与测试集，冻结前层网络参数训练并微调深度学习模型，迭代 100 次以后，训练集、测试集准确率分别达到了 80.48%和 79.63%。结果表明基于卷积神经网络和迁移学习的深度学习模型确实具有更高的效率以及识别精度。

对于问题三，我们首先推导了基于暗通道先验理论的能见度与透射率、消光系数和场景深度间的数理关系。其次，使用基于梯度导向滤波技术的暗通道先验算法对题目提供的高速公路图片求解暗通道图像、采用梯度导向滤波对获得的粗透射率进行细化、求解全局大气光照等获得去雾后的清晰图片，获得图像的透射率数据。基于透射率数据计算得到图像消光系数，结合相机标定法确定场景深度距离，最终测算得到图像的能见度数据。同时，绘制了大雾情形下能见度随时间变化曲线。结果显示能见度序列在 89.28 米到 94.51 米范围内小幅波动，能见度较低。

对于问题四，根据问题三估计的能见度数据，我们建立了基于粒子群优化算法的非线性灰色伯努利模型（PSO-NGBM（1,1）模型）和时间序列中的 ARMA（3,3）模型，以预测大雾未来变化趋势与消散时间。结果表明 PSO-NGBM（1,1）模型与 ARMA（3,3）的平均绝对百分比误差（MAPE）分别 1.21%和 0.63%。最终本文选取拟合精度较高的 ARMA（3,3）模型，进行样本外预测，预估在 2016 年 4 月 15 日上午 7 时以后，MOR 能见度将达到 200 米，雾将散去。

关键词：能见度；分位数非线性回归；迁移学习；卷积神经网络；暗通道先验算法；时间序列模型

# 目 录

|       |                                       |    |
|-------|---------------------------------------|----|
| 1     | 问题重述 .....                            | 5  |
| 1.1   | 研究背景.....                             | 5  |
| 1.2   | 问题提出.....                             | 5  |
| 1.3   | 全文思路框架图.....                          | 6  |
| 2     | 模型假设 .....                            | 7  |
| 3     | 符号说明 .....                            | 8  |
| 4     | 问题一模型建立与求解.....                       | 9  |
| 4.1   | 问题分析.....                             | 9  |
| 4.2   | 数据来源与指标说明.....                        | 10 |
| 4.3   | 能见度与各地面气象要素的变化散点图.....                | 11 |
| 4.3.1 | 能见度（RVR）与各气象要素之间的变化散点图.....           | 11 |
| 4.3.2 | 能见度（MOR）与各气象要素之间的变化散点图.....           | 13 |
| 4.3   | 能见度与地面气象要素相关关系分析.....                 | 15 |
| 4.4   | 分位数非线性回归模型介绍.....                     | 17 |
| 4.5   | 最优分位点选择.....                          | 18 |
| 4.6   | 对比模型：传统的多元线性回归、传统的分位数回归和分位数非线性模型..... | 20 |
| 4.7   | 回归结果分析.....                           | 22 |
| 5     | 问题二模型建立与求解.....                       | 25 |
| 5.1   | 问题分析.....                             | 25 |
| 5.2   | 图像提取说明.....                           | 26 |
| 5.3   | 卷积神经网络与迁移学习.....                      | 27 |
| 5.3.1 | 卷积神经网络.....                           | 27 |
| 5.3.2 | 迁移学习.....                             | 27 |
| 5.4   | 基于图像识别的机器学习模型.....                    | 28 |
| 5.4.1 | 图像处理.....                             | 28 |
| 5.4.2 | BP 神经网络.....                          | 29 |
| 5.5   | 实验设计与结果分析.....                        | 30 |
| 5.5.1 | 实验环境及方案.....                          | 30 |
| 5.5.2 | 模型参数设置.....                           | 30 |
| 5.5.3 | 实验结果分析.....                           | 31 |
| 6     | 问题三模型建立与求解.....                       | 34 |
| 6.1   | 问题分析.....                             | 34 |
| 6.2   | 能见度检测理论模型.....                        | 34 |
| 6.3   | 基于暗通道先验理论的能见度检测.....                  | 36 |
| 6.3.1 | 暗通道先验理论.....                          | 36 |
| 6.3.2 | 大气透射率获取.....                          | 37 |
| 6.3.3 | 基于导向滤波的透射率细化.....                     | 38 |
| 6.3.4 | 能见度的消光系数测算.....                       | 39 |
| 6.3.5 | 基于题目推导的能见度测算式.....                    | 40 |
| 6.4   | 基于相机标定法的距离估计.....                     | 40 |
| 6.4.1 | 相机标定法介绍.....                          | 40 |
| 6.4.2 | 相机标定法原理.....                          | 40 |

|            |                             |    |
|------------|-----------------------------|----|
| 6.4        | 基于高速公路图像信息的能见度分析.....       | 42 |
| 6.4.1      | 高速公路图像说明.....               | 42 |
| 6.4.2      | 高速公路图片能见度变化趋势曲线.....        | 43 |
| 7          | 问题四模型建立与求解.....             | 44 |
| 7.1        | 问题分析.....                   | 44 |
| 7.2        | 基于粒子群优化算法的非线性伯努利模型预测.....   | 44 |
| 7.2.1      | 非线性灰色伯努利模型.....             | 44 |
| 7.2.2      | 基于 PSO 算法的参数优化 .....        | 46 |
| 7.2.3      | PSO-NGBM (1,1) 模型预测结果 ..... | 47 |
| 7.2        | ARMA (3,3) 模型.....          | 48 |
| 7.2.1      | ARMA 模型.....                | 48 |
| 7.2.2      | ARMA 模型预测结果.....            | 48 |
| 7.3        | 能见度划分依据.....                | 51 |
| 7.4        | 各模型预测结果.....                | 51 |
| 8          | 模型的总结与评价.....               | 53 |
| 8.1        | 模型的结论.....                  | 53 |
| 8.2        | 模型的优点.....                  | 53 |
| 8.3        | 模型的不足之处.....                | 54 |
| 参考文献 ..... |                             | 55 |
| 附录 A.....  |                             | 57 |
| 附录 B.....  |                             | 65 |

# 1 问题重述

## 1.1 研究背景

大气能见度是重要的气象观测要素，是指视力正常的人能将目标物从背景中识别出来的最大距离，所谓“能见”，在白天是指能看到和辨认出目标物的轮廓和形体，在夜间是指能清楚看到目标灯的发光点。能见度制约着航空、航海、陆地交通以及军事活动的正常进行，而影响能见度的因素主要是雾和霾。随着我国经济的快速发展，城市人口不断扩张、能源迅速消耗，各地雾霾天气频发，由此导致的航班延误以及交通事故历年均有发生。因此，能见度预测是高速公路管理部门和航空公司十分关注的问题。

1924 年，柯西米德（Koschmieder）建立了大气能见度测量的基本理论，奠定了大气能见度测量研究的基础。随着能见度测量在各个领域发挥着越来越重要的作用，相关研究方法也不断发展。最早，人们采用人工目测法测量大气能见度，规定观测者所能看到的最远的目标物所对应的水平距离即为大气能见度，原始的人工目测法依赖于观测者的视觉分辨以及人为主观判断，存在较大的误差。随着科技的发展与集成电路的开发，科学的能见度测量仪应运而生，虽然这类仪器相对而言具有更高的测量精度，但由于其价格昂贵、维护成本高，无法广泛投入使用。

1994 年，Thomas Legal 等人用数字化摄像机进行了实时测量能见度的简单试验。1998 年，中国科学院周秀骥院士提出了用数码摄像机模仿人眼拍摄图像测算能见度的方法，称为数字摄像法。该方法是将大气光学分析与图像处理及人工智能技术结合，通过对视频图像的分析处理，建立视频图像与真实场景之间的关系，再根据图像特征的变化，间接计算出能见度数值。这类基于视频图像识别的能见度检测方法操作方便、价格低廉且实用性强，在一定程度上克服了激光能见度仪的不足，被国内外专家学者广泛使用研究。

本项目重点关注不同大雾情况下的能见度估计与预测问题，由于大雾的形成和消散有其自身的规律，通常与近地层的气象因素有关。而视频资料包含了丰富的信息，特别是涵盖了大雾的变化过程信息。充分利用这些信息，不仅可以提高能见度估计精度，也可以对大雾的消散进行预测。

## 1.2 问题提出

为了估计不同大雾情况下对应的能见度以及预测大雾的消散，请回答以下问题：

（1）众所周知，雾与近地面的气象因素有关。建立模型描述能见度与地面气象观测（温度、湿度和风速等）之间的关系，并针对题目所提供的数据（机场 AMOS 观测.zip）导出具体的关系式；

（2）根据题目提供的某机场视频数据（机场视频.zip）和能见度数据（机场 AMOS 观测.zip），建立基于视频数据的能见度估计深度学习模型，并对估计的能见度进行精度评估；

（3）高速公路某路段只有监控视频数据，建立不依赖能见度仪观测数据的能见度估计算法（提示：事实上，在有雾的情形可以估计视频中物体的景深。反过来，理论上也可以利用视频中不同景深的物体，在不同能见度下的亮度差异估计能见度），讨论相关算法实现过程，并针对题目提供的一段视频（高速公路视频截图.zip）绘制该时间段这段高速公路能见度随时间变化曲线；

(4) 利用问题三得到的能见度随时间变化规律，建立数学模型预测大雾变化趋势（加重或减弱）、何时散去（达到指定的能见度，比如 MOR=150m）？

### 1.3 全文思路框架图

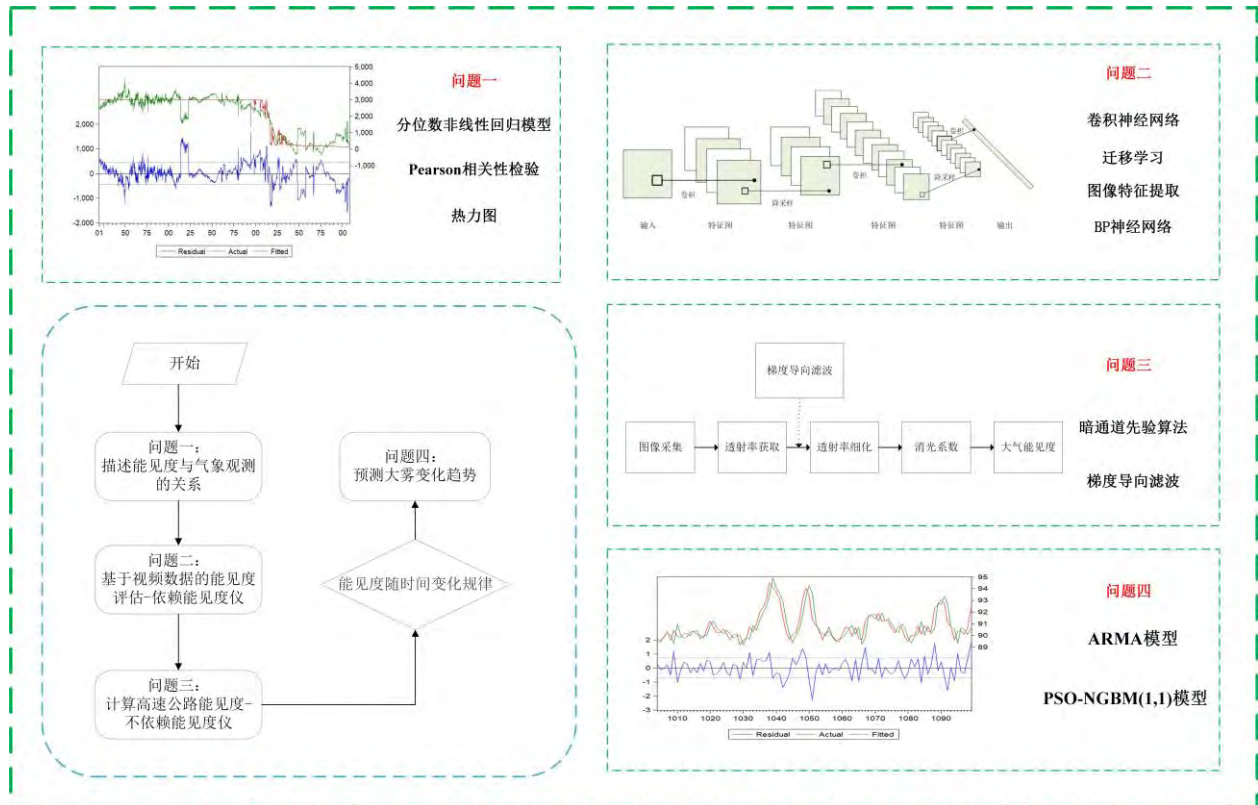


图 1 全文思路框架图

## 2 模型假设

- (1) 假设雾景中的大气微粒是均匀分布的，故散射系数只与波长有关；
- (2) 假设大气亮度分布是各向同性的；
- (3) 假设对于波长较短的可见光而言，因大气微粒吸收产生的反射光能量衰减可忽略不计；
- (4) 假设对某场景点的反射光束来说，它被散射到其他方向的部分不会再进入该反射光束中来；
- (5) 假设当场景点距离摄像头较远时，其反射光束可以近似为平行光束处理；
- (6) 假设场景点的亮度分布是各向同性的；
- (7) 假设提供的公路视频中的公路路面是光滑的水平路面。



### 3 符号说明

| 指标符号 | 指标说明                           |
|------|--------------------------------|
| RVR  | 跑道视程 (RVR≤2000 米) (能见度指标 1)    |
| MOR  | 气象光学视程 (MOR≤10000 米) (能见度指标 2) |
| PAIN | 本站气压                           |
| QNH  | 修正海平面气压                        |
| QFE  | 飞机着陆地区最高点气压                    |
| TEMP | 温度                             |
| RH   | 相对湿度                           |
| DEW  | 露点温度                           |
| LIG  | 灯光数据                           |
| WS2  | 2 分钟平均风速                       |
| WD2  | 2 分钟平均风向                       |
| CW2  | 2 分钟平均垂直风速                     |

## 4 问题一模型建立与求解

### 4.1 问题分析

题目一要求我们解决两个问题：（1）建立模型描述能见度与地面气象观测（温度、湿度和风速等）之间的关系；（2）针对题目所提供的数据导出具体的关系式。

对于该问题，我们首先对题目提供的数据进行预处理，包括指标提取、异常值识别、缺失值处理以及数据合并；其次对数据进行描述性分析，重点对能见度数据与气象观测数据进行相关性分析并绘制相关图表；而后运用统计分析软件构建分位数非线性回归模型，并将其拟合结果与传统多元线性回归、传统非线性回归以及分位数线性回归进行比较，在此基础上筛选变量指标、进行模型形式选择，最终确定最优分位点并采用分位数非线性回归模型探究能见度与地面气象观测之间的关系，最终导出具体的关系式、进行结果分析。

图 4-1 为问题一的思路框架图。

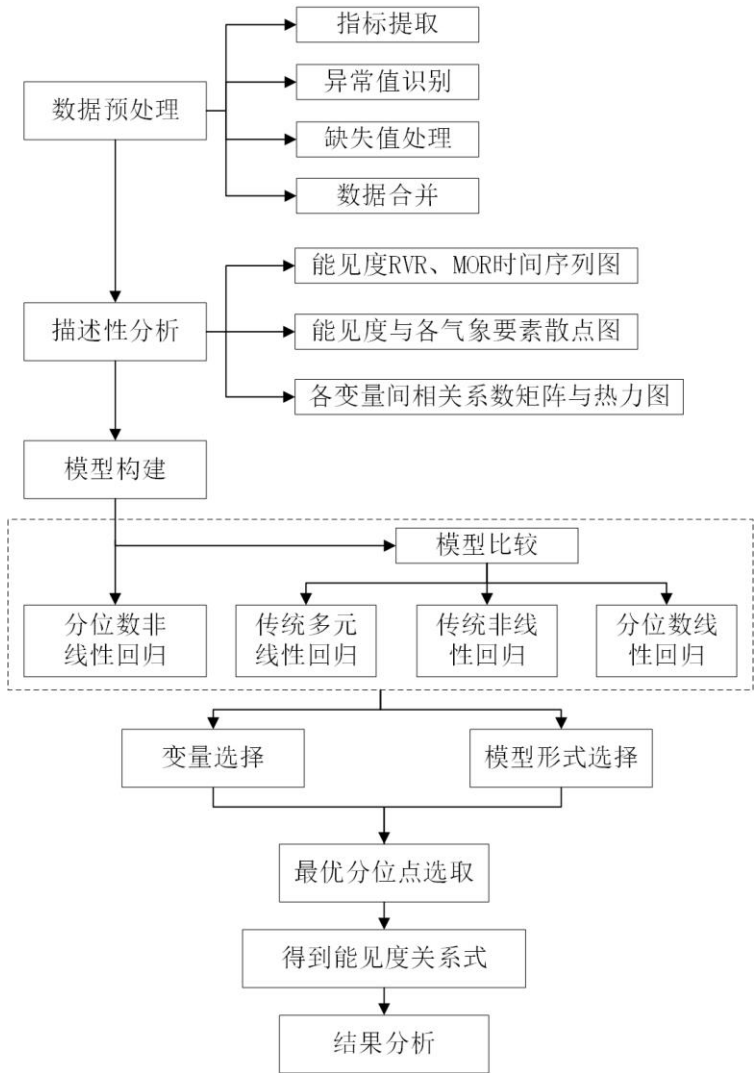


图 4-1 问题一思路框架图

4.2 数据来源与指标说明

能见度数据与气象观测数据均来源于机场 AMOS 观测数据，其中数据时段分别为 2019 年 12 月 15 日 8:00-2019 年 12 月 16 日 7:59 以及 2020 年 3 月 12 日 8:00-2020 年 3 月 13 日 7:59（北京时间）。其中，气压、气温、湿度以及露点温度等数据每分钟监测一次（8:00、8:01、8:02、8:03……），能见度、灯光数据、风速和风向等数据每 15 秒监测一次（8:00:00、8:00:15、8:00:30、8:00:45、8:01:00……），为保持数据的一致性，选取每分钟整的数据作为研究样本数据。各指标数据如表 4-1 所示。

表 4-1 各指标符号及说明

| 指标符号 | 指标说明                         |
|------|------------------------------|
| RVR  | 跑道视程（RVR≤3000 米）（能见度指标 1）    |
| MOR  | 气象光学视程（MOR≤10000 米）（能见度指标 2） |
| PAIN | 本站气压                         |
| QNH  | 修正海平面气压                      |
| QFE  | 飞机着陆地区最高点气压                  |
| TEMP | 温度                           |
| RH   | 相对湿度                         |
| DEW  | 露点温度                         |
| LIG  | 灯光数据                         |
| WS2  | 2 分钟平均风速                     |
| WD2  | 2 分钟平均风向                     |
| CW2  | 2 分钟平均垂直风速                   |

RVR 为跑道视程，是指在跑道中线上的航空器上的飞行员能看到跑道面上的标志或跑道边界灯或中线灯的距离。MOR 为气象光学视程，是指色温为 2700K 的白炽灯发出的平行光束的光通量在大气中衰减降低到它的起始值的 5%距离<sup>[1]</sup>，由于人眼识别以及机器辨识均存在一定的局限性，故两者均存在观测上限值，分别为 3000 米以及 10000 米。图 4-2 展示了 RVR 和 MOR 随时间变化的趋势。



图 4-2 RVR 和 MOR 随时间变化趋势图

总体来看，随时间推延，2020 年 3 月 12 日 8: 00-13 日 8: 00 的能见度呈先上升后下降趋势，下降的第一个拐点大约在下午 16: 30。由图像可知，数据样本中 RVR 存在一段时间（8:00-23:00）数值恒为 3000 米，MOR 同样存在一段时间（12:50-16:30）数值恒为 10000 米，合理推测该两段时间实际能见度大于或等于所填列的测量值。

### 4.3 能见度与各地面气象要素的变化散点图

为直观分析能见度与各气象要素（气压、温度、湿度等）之间的关系，本节基于数据样本，分别绘制了能见度与各要素之间的变化散点图，需要说明的是，为更好地观察能见度与相关要素之间的关系，在绘制散点图时，剔除了 RVR 以及 MOR 为 3000 米以及 10000 米的观测值。

#### 4.3.1 能见度（RVR）与各气象要素之间的变化散点图

图 4-3 绘制了能见度（RVR）与 PAIN（本站气压）之间的关系（由于三个气压指标绘制结果十分接近，故此处仅选取其中之一进行简单分析）。从图中可以看到，在同一天同一地区的气压变化较小，但总体来看能见度与气压呈现出一定的负相关关系，可见随着气压的增加，能见度呈下降趋势。

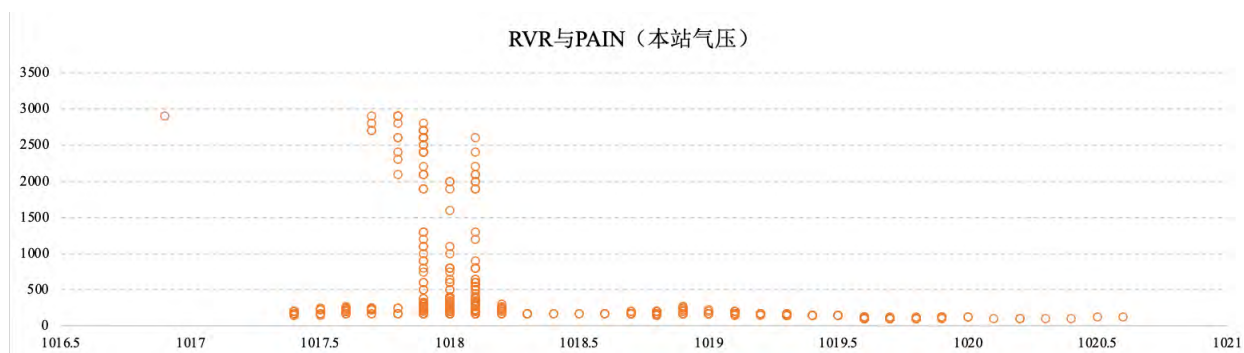


图 4-3 能见度（RVR）及 PAIN（本站气压）的散点图

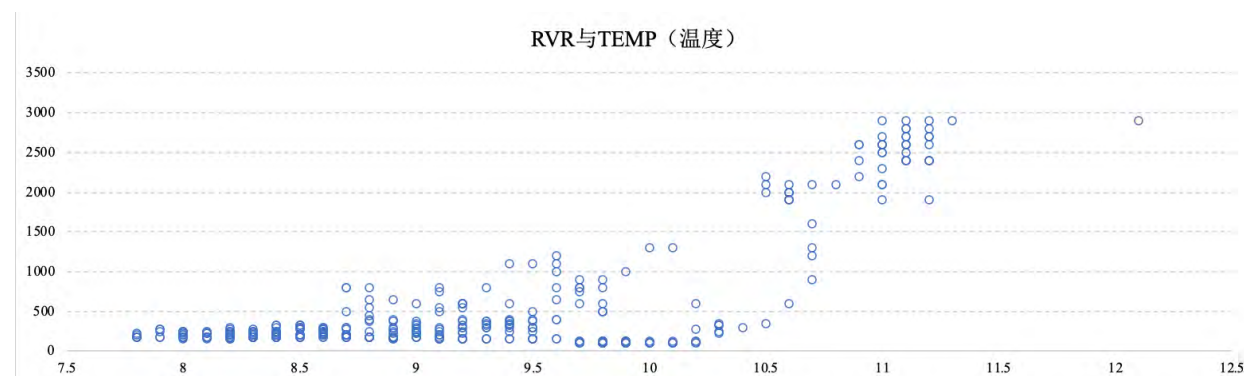


图 4-4 能见度（RVR）与 TEMP（温度）的散点图

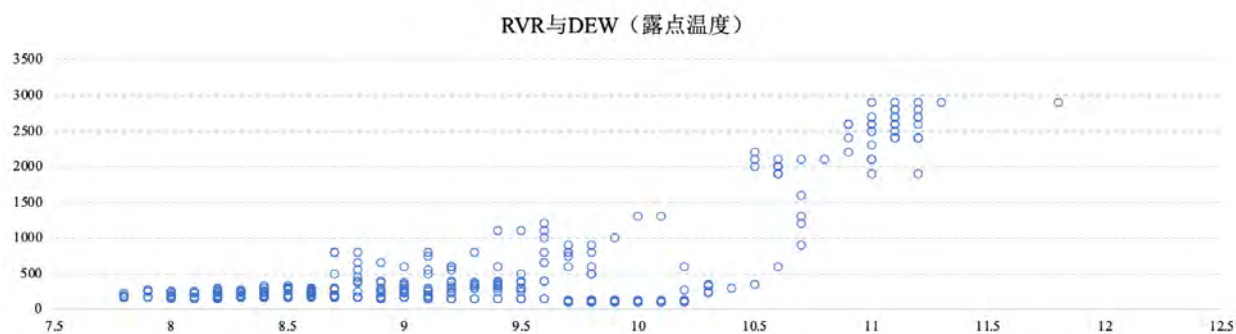


图 4-5 能见度 (RVR) 与 DEW (露点温度) 的散点图

由图 4-4 和图 4-5 不难发现, 能见度 (RVR) 与 TEMP (温度) 及 DEW (露点温度) 之间的图像几乎完全一致, 观察数据可以发现, 在剔除 RVR=3000 米的数据后, 温度与露点温度基本不存在差距。总体来看, 能见度与温度之间的关系呈现出正相关关系, 且随着温度的增加, 能见度的增加速率显著上升, 二者之间可能存在着一定的指数关系。

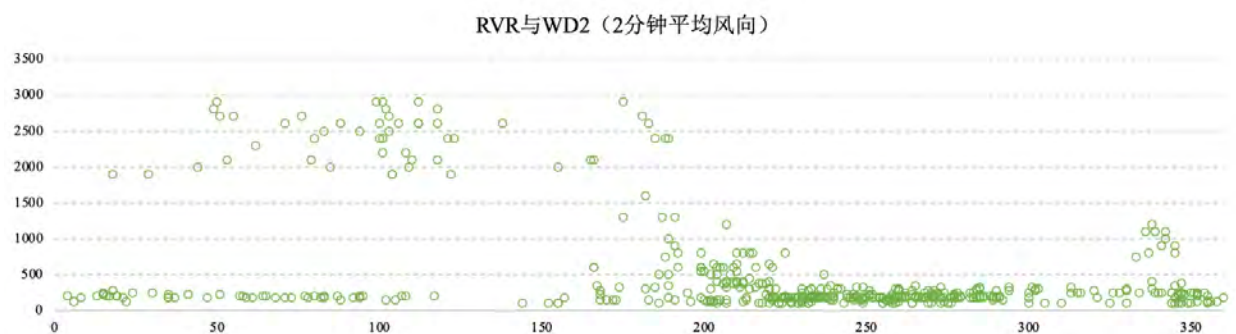
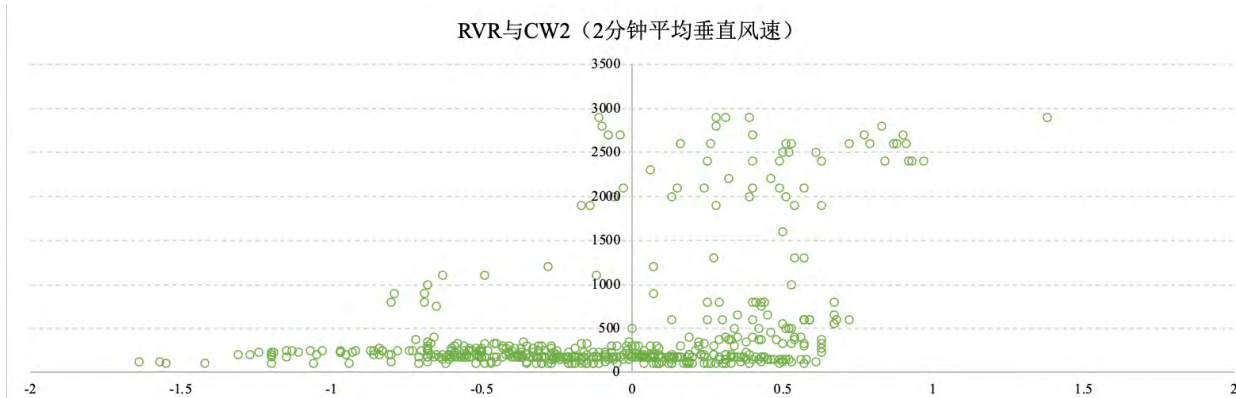
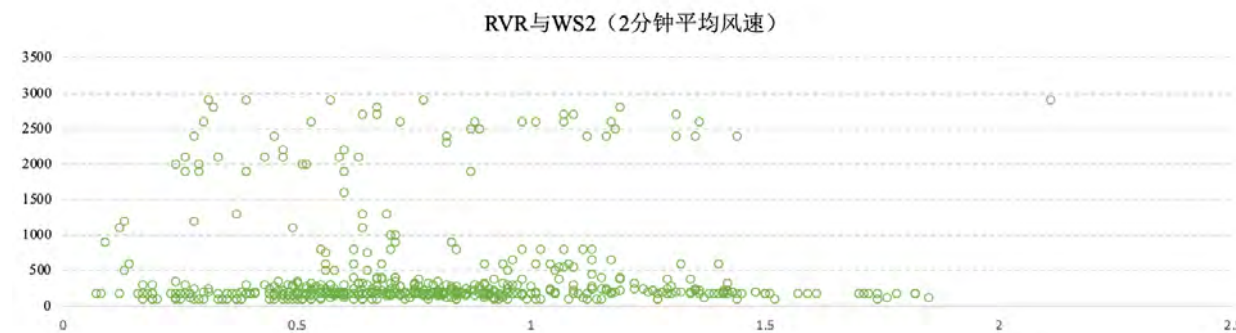


图 4-6 能见度 (RVR) 与 WS2 (2 分钟平均风速)、CW2 (2 分钟平均垂直风速) 及 WD2 (2 分钟平均风向) 的散点图

由图 4-6 观察可得，能见度（RVR）与风速之间呈现出大致的正相关关系，与风向之间呈现出负相关关系。

另外，由于在剔除 RVR=3000 米的数据后，RH(相对数值)为 100%的比例占到了 99.8%，故此处不对其与能见度数据之间的关系进行绘图；另外，灯光数据仅有 10、100 两类观测数据，同样不做绘制。

4.3.2 能见度（MOR）与各气象要素之间的变化散点图

图 4-7 展示了能见度（MOR）及 PAIN（本站气压）的散点图，总体而言，能见度（MOR）与气压之间同样呈现负相关关系，当气压大于 1018 时，能见度降到 1000 米以下的水平。

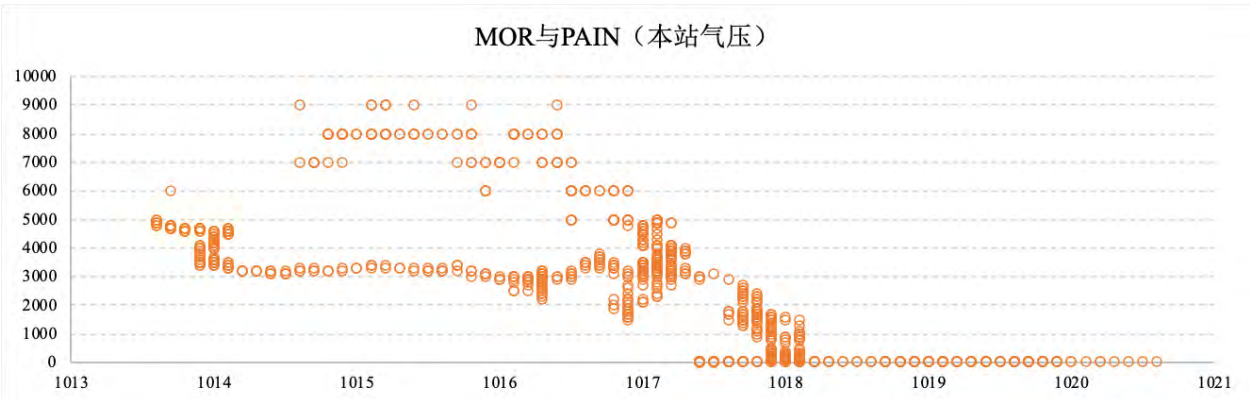


图 4-7 能见度（MOR）与 PAIN（本站气压）的散点图

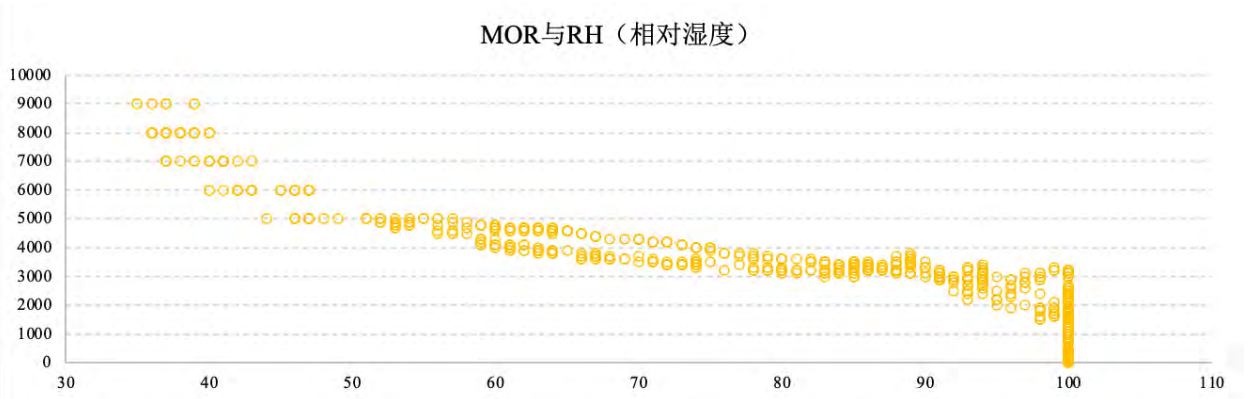


图 4-8 能见度（MOR）与 RH（相对湿度）的散点图

根据图 4-8，能见度（MOR）与 RH（相对湿度）之间呈现出显著的负相关关系，且随着相对湿度的增加，能见度下降的趋势逐渐减小，由于当相对湿度达到一定水平时，空气中的气溶胶粒子吸湿能力达到饱和，即便相对湿度继续增加，其对能见度的影响有限，故能见度下降趋势渐缓。二者关系接近于对数函数。



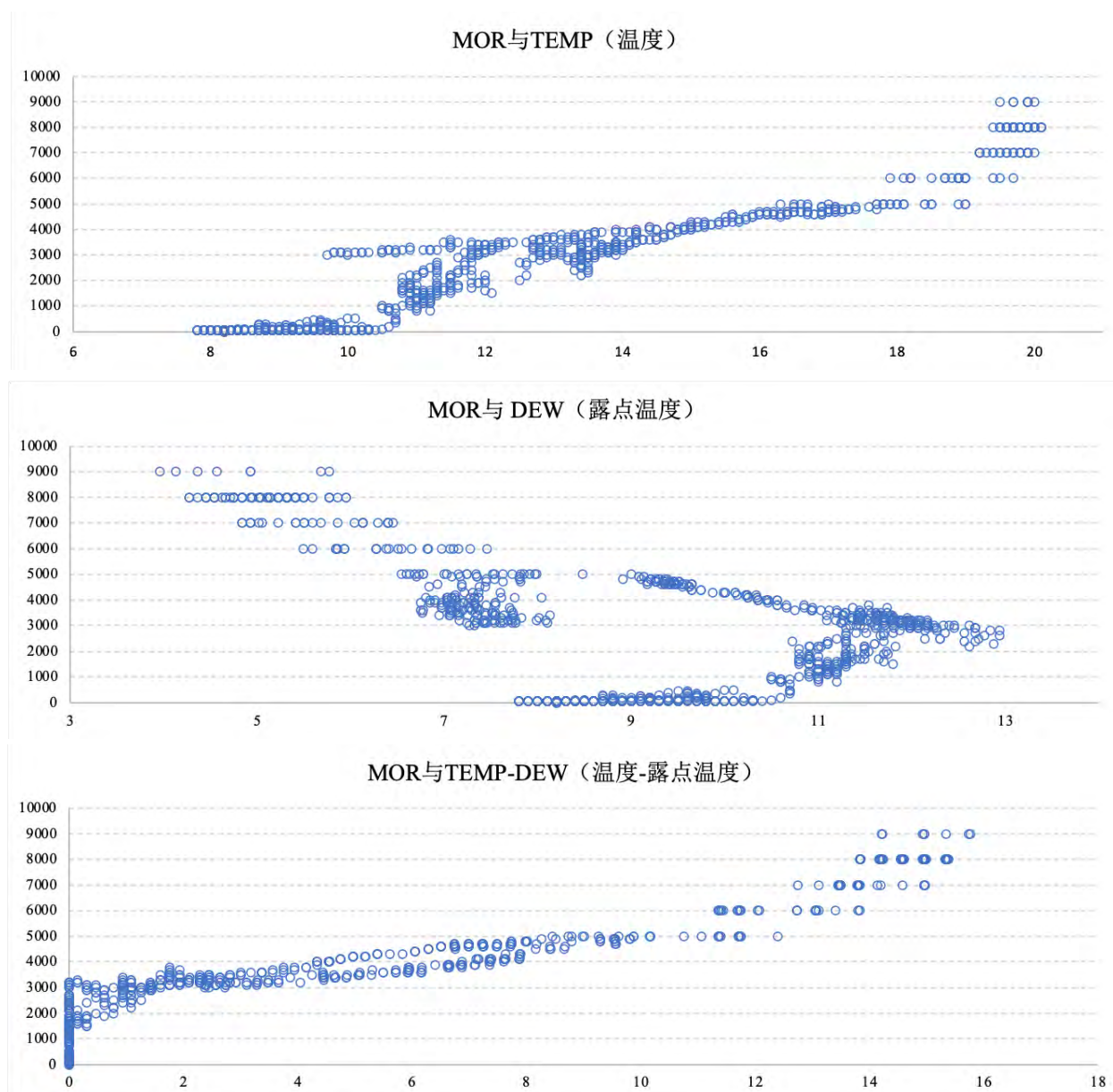


图 4-9 能见度 (MOR) 与 TEMP (温度)、DEW (露点温度) 及 TEMP-DEW (温度-露点温度) 的散点图

从图 4-9 中可以发现，能见度与温度呈现出正相关关系，能见度与露点温度的关系相对不明显。根据相关资料显示，空气中气态水液化成小液滴的必要条件是温度需要达到露点温度，故将二者作差，观察能见度与差值之间的关系。如图所示，能见度与二者之差呈现出显著正相关关系，且随着温差的升高，能见度的变化速率逐渐增大，呈现出一定的指数函数关系。

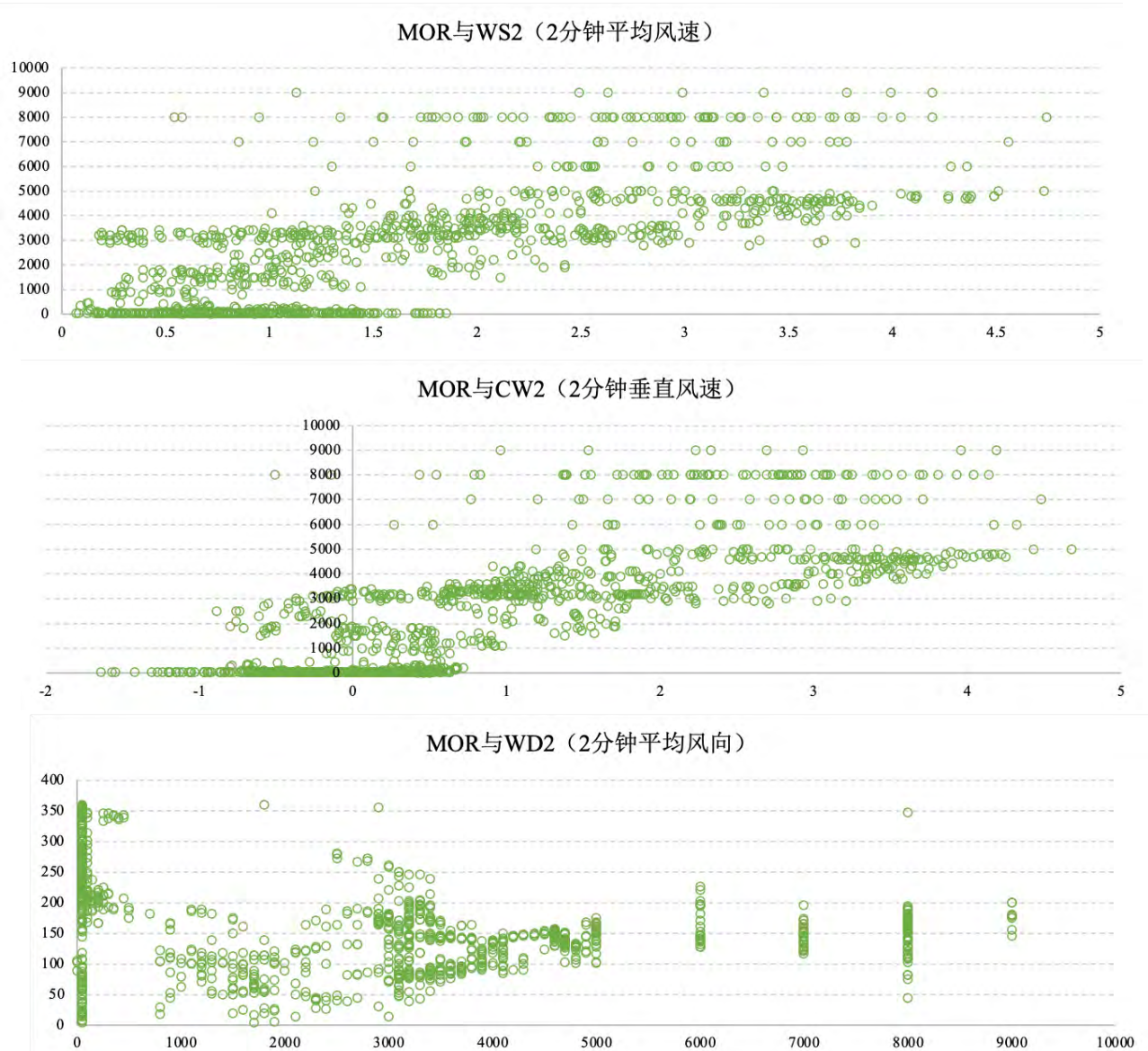


图 4-10 能见度 (MOR) 与 WS2 (2 分钟平均风速)、CW2 (2 分钟平均垂直风速) 及 WD2 (2 分钟平均风向) 的散点图

由图 4-10 观察可得，能见度 (MOR) 与风速之间呈现出大致的正相关关系，与风向之间呈现出负相关关系。

### 4.3 能见度与地面气象要素相关关系分析

一般而言，大气污染物和气象条件是影响大气能见度的主要因素<sup>[2]</sup>，本项目主要考虑地面气象要素与能见度之间的关系。为确定二者之间的关系并筛选后续用于回归的变量，本节运用统计分析软件 R 对大气能见度各时期数据（分别包括 RVR 以及 MOR）与同期各气象要素数据进行相关性分析。结果如图 4-11 及图 4-12 所示。



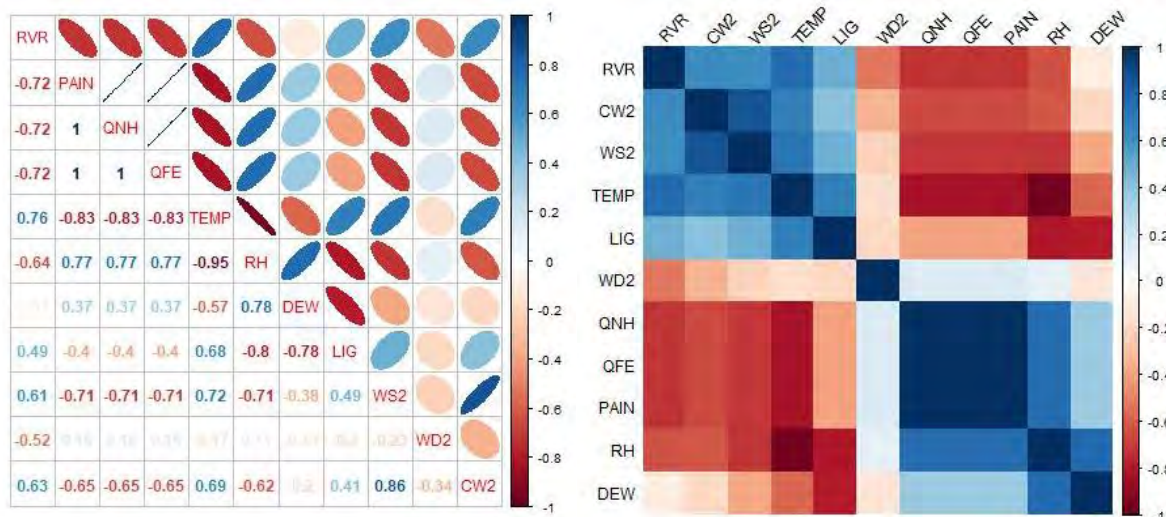


图 4-11 能见度（RVR）与气象因素的相关系数矩阵图及热力图

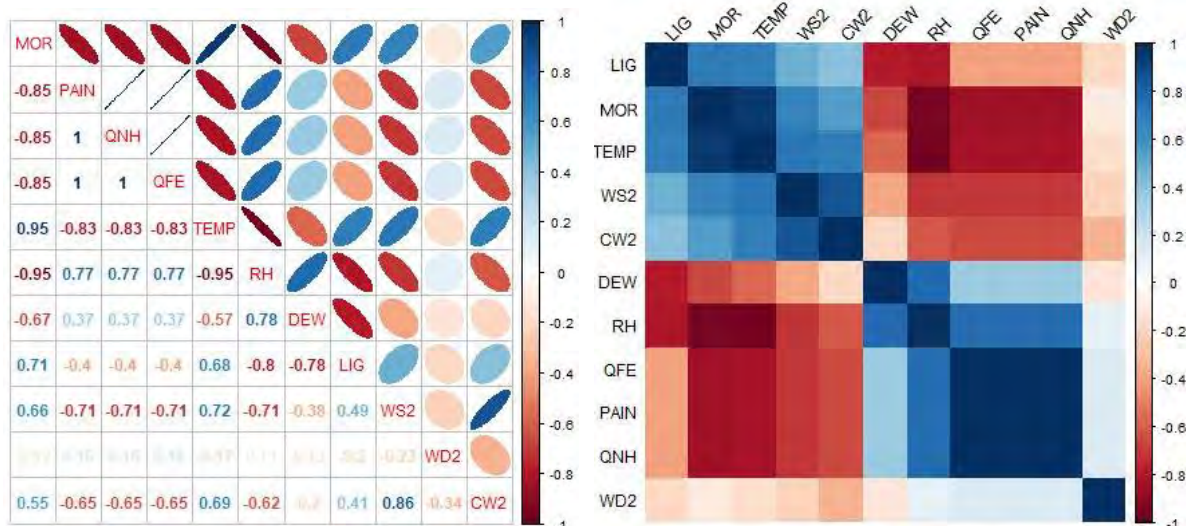


图 4-12 能见度（MOR）与气象因素的相关系数矩阵图及热力图

从相关系数矩阵图以及热力图可以发现，能见度与三个气压指标均呈现明显的负相关关系，由于三个气压指标之间的相关系数均为 1，为避免多重共线性对回归结果的影响，故在实际建模过程中仅保留气压指标 QFE；能见度与湿度呈现负相关关系，这与相关研究结果一致，即相对湿度的增加会带来大气可见度的下降，其原因在于空气中相对湿度的增高会带来大气颗粒物的吸湿增长，造成颗粒物粒径增长，使得消光系数增加，进而导致能见度降低<sup>[3]</sup>；能见度与温度呈正相关（相关系数分别为 0.76 和 0.95），与露点温度呈负相关（相关系数分别为-0.64 和-0.67），露点温度是指在空气中水汽含量不变，气压一定的情况下，使空气冷却达到饱和时的温度，露点温度作为空气湿度的一种表示方式，再次表明能见度与湿度呈现负相关。气温和露点温度的差值越小，表示空气越接近于饱和，相对湿度越大，越容易形成雾，能见度越低；能见度与灯光数据呈正相关关系（相关系数分别为 0.49 和 0.71），即前者随后者增加而增加；能见度与风速呈正相关关系，一般而言风速越大，空气污染物、雾气等易消散，故能见度大幅提升。

根据相关性分析结果，本章节最终对气象因素进行了筛选，由于三个气压指标强相关，故仅保留 1 个气压变量 QFE（飞机着陆地区最高点气压）；对于温度指标，由于 TEMP（温

度)与 RH(相对湿度)强相关,故仅保留 DEWPOINT(露点温度),剔除 TEMP(温度)指标;对于风速指标,仅保留 WS2(2 分钟平均风速),剔除 CW2(2 分钟平均垂直风速)。

#### 4.4 分位数非线性回归模型介绍

分位数回归技术能够更加全面的描述变量条件分布的全貌,而不像 OLS 那样仅分析变量的条件期望。不同分位数下的估计量和预测结果也不同。由于分位数回归对误差项并不要求很强的假设条件,因此当分布是重尾或存在强影响点,以及异常点时,基于分位数回归的预测结果更加稳健。在应用研究中,该分位数回归技术的强稳健性已被不同领域的学者广泛验证,例如交通安全、生态工程、风险价值、薪酬结构等。分位数回归的优点是能够更加全面的描述被解释变量条件分布的全貌,而不是仅仅分析被解释变量的条件期望(均值),也可以分析解释变量如何影响被解释变量的中位数、分位数等。不同分位数下的回归系数估计量常常不同,即解释变量对不同水平被解释变量的影响不同。

设随机变量  $X$  的分布函数为:

$$F(x) = P(X \leq x) \quad (4-1)$$

则随机变量  $X$  的分位数定义为:

$$Q_\tau(x) = \arg \inf \{x \in \mathfrak{R}; F(x) \geq \tau\} (0 < \tau < 1) \quad (4-2)$$

设  $F(x)$  的逆函数为:

$$F_X^{-1}(\tau) = \inf \{y \in \mathfrak{R}; F(y) \geq \tau\} \quad (4-3)$$

则有:

$$Q_\tau(X) = F_X^{-1}(\tau) \quad (4-4)$$

设  $X$  在给定  $Z = z$  的条件下的累积分布函数为  $F_{X|Z=z}(x|z)$ , 则  $X|Z=z$  的  $\tau$  条件分位数定义为:

$$Q_\tau(X|Z=z) = \arg \inf \{X \in \mathfrak{R}; F(X|z) \geq \tau\} (0 < \tau < 1) \quad (4-5)$$

设样本序列满足回归模型  $x(k) = \mathbf{z}(k)\boldsymbol{\beta}^T + \varepsilon(k)$ ,  $\{k=1, \dots, m\}$ ,  $x$  表示有  $m$  维观测序列,  $\mathbf{z}$  表示  $m \times n$  维自变量函数矩阵,  $\boldsymbol{\beta}$  表示  $1 \times n$  行向量回归系数, 误差项  $\varepsilon_k$  为独立同分布的序列, 且分布情况未知, 则  $x$  的  $\tau$  条件分位数:

$$m_\tau(x) = \arg \min \{\rho_\tau(x - \hat{X}) | \mathbf{z}\} \quad (4-6)$$

其中:

$$R(\boldsymbol{\beta}_\tau(k)) = \sum_k^m \rho_\tau(\mu_\tau(k)) \quad (4-7)$$

为分位数回归模型的损失函数, 其中  $\rho_\tau(\mu(k))$  定义如下:

$$\rho_\tau(\mu(k)) = \int (\tau - I(\mu(k) \leq 0)) \cdot \mu(k) dk \quad (4-8)$$

式 (4-8) 中,  $I(\cdot)$  为示性函数, 损失函数本质上是对误差为正的值赋以权重  $\tau$ , 对误差为负的值赋以权重  $(1-\tau)$ 。在最小化损失函数准则下, 随着  $\tau \rightarrow 1$ , 误差为正的部分会逐渐减少, 分位数回归曲线会逐渐向上移动, 因此对应的回归系数会相应变化。

连续变量  $X^{(0)}$  的概率密度函数用  $f(x)$  表示, 其第  $\tau$  分位数用  $x_{(\tau)}$  表示, 则我们希望以  $E(\rho_\tau | x - \hat{X} |)$  最小:

$$E(\rho_\tau | x - \hat{X} |) = \int_{-\infty}^{\hat{X}} (\tau - 1)(x - \hat{X})f(x)dx + \int_{\hat{X}}^{\infty} \tau(x - \hat{X})f(x)dx \quad (4-9)$$

则令  $E(\rho_\tau | x - \hat{X} |) = 0$ , 即  $(\tau - 1)\int_{-\infty}^{\hat{X}} dF(x) - \tau\int_{\hat{X}}^{\infty} dF(x) = F(\hat{X}) - \tau = 0$ , 我们得到

$$F(\hat{X}) = \tau \quad (4-10)$$

用样本经验分布函数代替  $F(x)$  即  $F(x) = \frac{\sum_{k=1}^m I(X(k) \leq x(k))}{m}$ , 则

$$\int \rho_\tau(x - \hat{X})dF_m(x) = m^{-1} \sum_{k=1}^m \rho_\tau(x(k) - \hat{X}) \quad (4-11)$$

与公式 (4-11) 等价的表述是  $m^{-1} \sum_{k=1}^m \rho_\tau(x(k) - \hat{X})$  以  $\hat{X} = x_{(\tau)}$  时为最小。因此, 分位数回归估计量可以通过加权绝对离差法估计:

$$\min_{\beta_\tau} \left( \sum_{x(k) \geq z(k)\beta_\tau^T} \tau |x(k) - z(k)\beta_\tau^T| + \sum_{x(k) < z(k)\beta_\tau^T} (1-\tau) |x(k) - z(k)\beta_\tau^T| \right) \quad (4-12)$$

由于式 (4-12) 目标函数不可微, 采用传统的目标函数求导方法不再适用, 因此本文选择一种对参数直接估计的方法, 改进的单纯形算法。

#### 4.5 最优分位点选择

为确保回归结果的准确性与科学性, 本文在符合经济意义和统计检验通过的前提下, 试图寻找该回归模型的最优分位点。图 4-13 和图 4-14 分别展示了基于能见度 (RVR) 和能见度 (MOR) 对分位数回归方程进行系列分位数回归的系数估计值结果。

通过综合比较不同分位点下的各变量 P 值、模型的 F 统计量、P 值以及调整 R<sup>2</sup> (由于篇幅原因, 在此不展示不同分位点下的模型结果, 附录 A 和附录 B 分别展示了以 RVR 和 MOR 为例的分位点非线性回归的分位点选择), 本文基于能见度 (RVR) 的非线性分位数回归最优分位点取 0.5, 基于能见度 (MOR) 的非线性分位数回归最优分位点取 0.6。

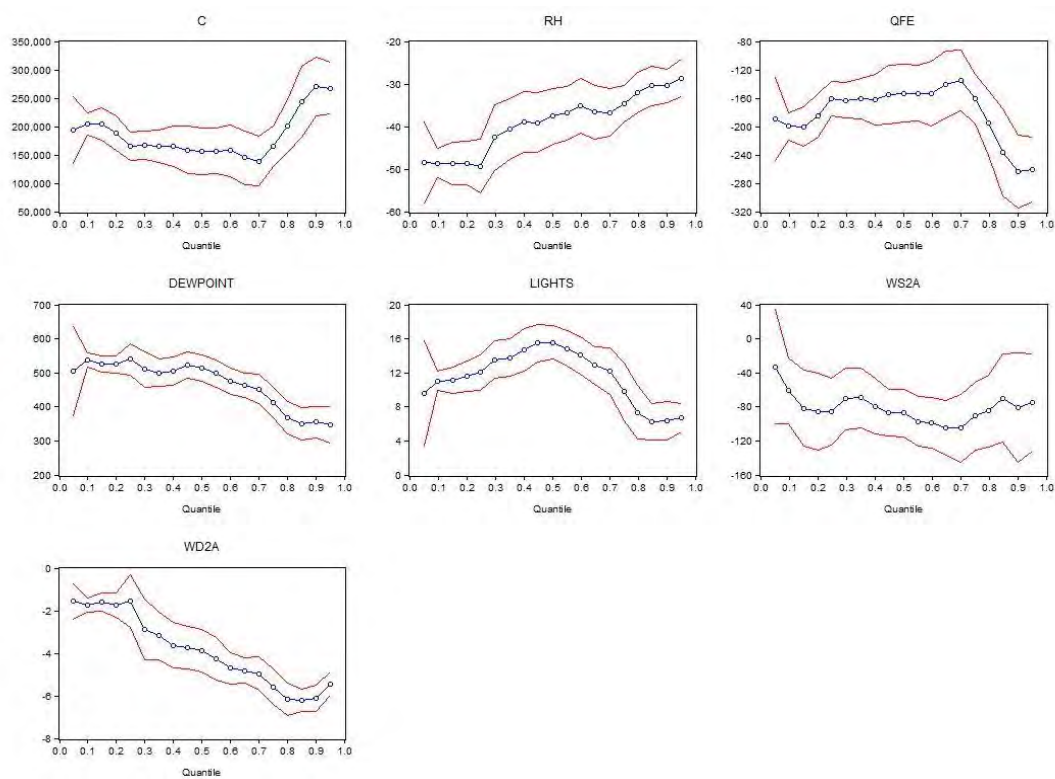


图 4-13 基于能见度（RVR）的系列分位数回归方程系数估计值及置信区间（95%）

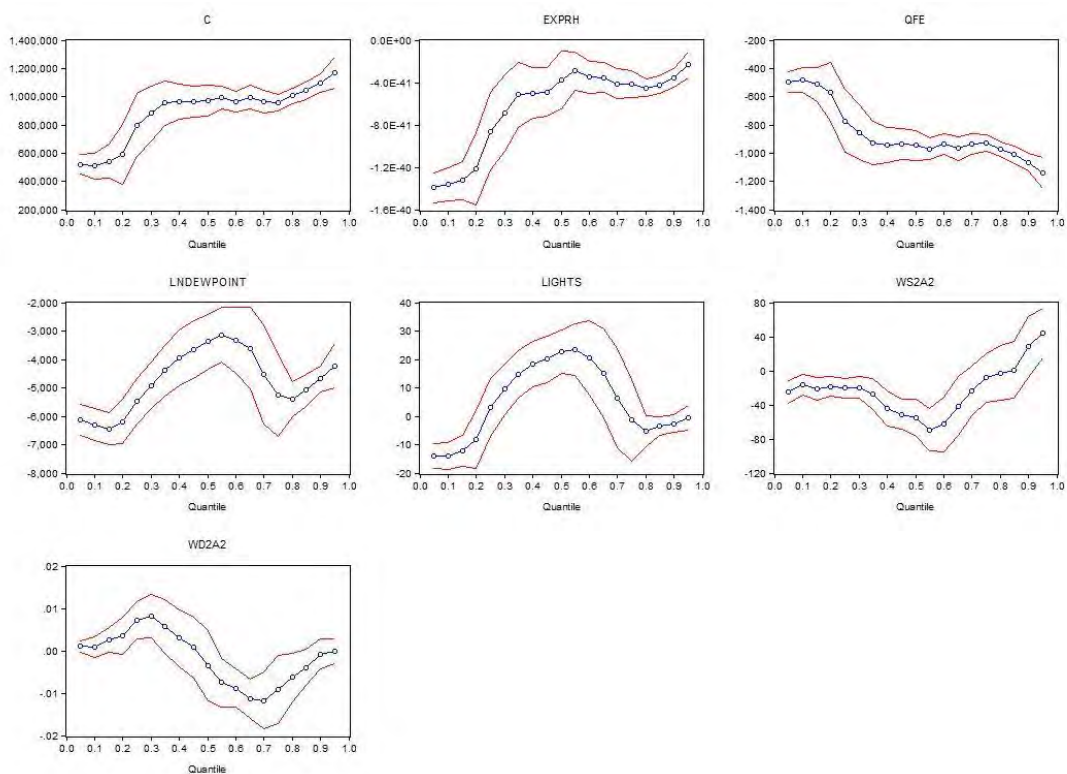


图 4-14 基于能见度（MOR）的系列分位数回归方程系数估计值及置信区间（95%）

#### 4.6 对比模型：传统的多元线性回归、传统的分位数回归和分位数非线性模型

本文还进行了基于不同能见度的传统多元线性回归、传统分位数回归等模型比较。在调整  $R^2$  和各变量检验等统计评价指标最优和符合经济意义的多重目标下，我们认为本文提出的基于最优分位点的分位数非线性回归效果最优，因此选择了分位数非线性回归模型。

表 4-4 基于能见度（RVR）的传统多元线性回归

| Variable           | Coefficient | Std. Error            | t-Statistic | Prob.  |
|--------------------|-------------|-----------------------|-------------|--------|
| C                  | 171927.2    | 12230.43              | 14.05733    | 0.0000 |
| RH                 | -42.58954   | 1.638004              | -26.00088   | 0.0000 |
| QFE                | -167.9291   | 12.10204              | -13.87610   | 0.0000 |
| DEWPOINT           | 513.1262    | 10.73045              | 47.81963    | 0.0000 |
| LIGHTS             | 12.71337    | 0.596149              | 21.32584    | 0.0000 |
| WS2A               | -81.18738   | 16.70179              | -4.861000   | 0.0000 |
| WD2A               | -3.451859   | 0.180674              | -19.10547   | 0.0000 |
| R-squared          | 0.889458    | Mean dependent var    | 2105.629    |        |
| Adjusted R-squared | 0.888994    | S.D. dependent var    | 1274.443    |        |
| S.E. of regression | 424.6128    | Akaike info criterion | 14.94508    |        |
| Sum squared resid  | 2.58E+08    | Schwarz criterion     | 14.97073    |        |
| Log likelihood     | -10745.99   | Hannan-Quinn criter.  | 14.95466    |        |
| F-statistic        | 1920.383    | Durbin-Watson stat    | 0.145314    |        |
| Prob(F-statistic)  | 0.000000    |                       |             |        |

表 4-5 基于能见度（RVR）的分位数线性回归

| Variable               | Coefficient | Std. Error         | t-Statistic | Prob.  |
|------------------------|-------------|--------------------|-------------|--------|
| C                      | 156391.9    | 21024.05           | 7.438716    | 0.0000 |
| RH                     | -37.60279   | 3.346207           | -11.23744   | 0.0000 |
| QFE                    | -153.0675   | 20.78902           | -7.362900   | 0.0000 |
| DEWPOINT               | 514.2836    | 20.15624           | 25.51485    | 0.0000 |
| LIGHTS                 | 15.57985    | 0.985016           | 15.81685    | 0.0000 |
| WS2A                   | -87.27139   | 14.23022           | -6.132820   | 0.0000 |
| WD2A                   | -3.880644   | 0.518317           | -7.487007   | 0.0000 |
| Pseudo R-squared       | 0.643707    | Mean dependent var | 2105.629    |        |
| Adjusted R-squared     | 0.642214    | S.D. dependent var | 1274.443    |        |
| S.E. of regression     | 438.2859    | Objective          | 229274.7    |        |
| Quantile dependent var | 3000.000    | Restr. objective   | 643500.0    |        |
| Sparsity               | 794.4363    | Quasi-LR statistic | 4171.262    |        |
| Prob(Quasi-LR stat)    | 0.000000    |                    |             |        |

表 4-6 基于能见度（MOR）的传统多元线性回归

| Variable           | Coefficient | Std. Error            | t-Statistic | Prob.  |
|--------------------|-------------|-----------------------|-------------|--------|
| C                  | 613732.3    | 24720.27              | 24.82708    | 0.0000 |
| RH                 | -114.0023   | 3.310751              | -34.43396   | 0.0000 |
| QFE                | -591.9606   | 24.46078              | -24.20040   | 0.0000 |
| DEWPOINT           | 118.1176    | 21.68850              | 5.446093    | 0.0000 |
| LIGHTS             | 4.011734    | 1.204942              | 3.329399    | 0.0009 |
| WS2A               | -475.1012   | 33.75783              | -14.07381   | 0.0000 |
| WD2A               | 0.510168    | 0.365180              | 1.397032    | 0.1626 |
| R-squared          | 0.940490    | Mean dependent var    | 3659.521    |        |
| Adjusted R-squared | 0.940241    | S.D. dependent var    | 3510.776    |        |
| S.E. of regression | 858.2319    | Akaike info criterion | 16.35248    |        |
| Sum squared resid  | 1.05E+09    | Schwarz criterion     | 16.37812    |        |
| Log likelihood     | -11758.61   | Hannan-Quinn criter   | 16.36205    |        |
| F-statistic        | 3771.888    | Durbin-Watson stat    | 0.083267    |        |
| Prob(F-statistic)  | 0.000000    |                       |             |        |

表 4-7 基于能见度（MOR）的分位数线性回归

| Variable               | Coefficient | Std. Error         | t-Statistic | Prob.  |
|------------------------|-------------|--------------------|-------------|--------|
| C                      | 428909.1    | 42001.22           | 10.21183    | 0.0000 |
| RH                     | -131.8817   | 9.276155           | -14.21728   | 0.0000 |
| QFE                    | -409.3185   | 41.28028           | -9.915594   | 0.0000 |
| DEWPOINT               | 185.8910    | 40.64902           | 4.573073    | 0.0000 |
| LIGHTS                 | 6.894982    | 5.980997           | 1.152815    | 0.2492 |
| WS2A                   | -578.9874   | 77.64411           | -7.456939   | 0.0000 |
| WD2A                   | 0.271041    | 0.352090           | 0.769805    | 0.4415 |
| Pseudo R-squared       | 0.750889    | Mean dependent var | 3659.521    |        |
| Adjusted R-squared     | 0.749845    | S.D. dependent var | 3510.776    |        |
| S.E. of regression     | 905.5287    | Objective          | 503023.1    |        |
| Quantile dependent var | 3200.000    | Restr. objective   | 2019275.    |        |
| Sparsity               | 1998.520    | Quasi-LR statistic | 6069.500    |        |
| Prob(Quasi-LR stat)    | 0.000000    |                    |             |        |



4.7 回归结果分析

根据上文选定的最优分位点，本文选用统计分析软件 R 对能见度与筛选后的气象指标进行分位数非线性回归，基于能见度（RVR）和能见度（MOR）的模型回归结果分别如表 4-8 和表 4-8 所示，能见度观测值与回归方程计算值时间序列折线图分别如图 4-15 和图 4-16 所示。

表 4-8 基于能见度（RVR）的最优分位点非线性回归结果

| Variable               | Coefficient | Std. Error         | t-Statistic | Prob.  |
|------------------------|-------------|--------------------|-------------|--------|
| C                      | 156391.9    | 21024.05           | 7.438716    | 0.0000 |
| RH                     | -37.60279   | 3.346207           | -11.23744   | 0.0000 |
| QFE                    | -153.0675   | 20.78902           | -7.362900   | 0.0000 |
| DEW                    | 514.2836    | 20.15624           | 25.51485    | 0.0000 |
| LIG                    | 15.57985    | 0.985016           | 15.81685    | 0.0000 |
| WS2                    | -87.27139   | 14.23022           | -6.132820   | 0.0000 |
| WD2                    | -3.880644   | 0.518317           | -7.487007   | 0.0000 |
| Pseudo R-squared       | 0.643707    | Mean dependent var | 2105.629    |        |
| Adjusted R-squared     | 0.642214    | S.D. dependent var | 1274.443    |        |
| S.E. of regression     | 438.2859    | Objective          | 229274.7    |        |
| Quantile dependent var |             |                    |             |        |
| var                    | 3000.000    | Restr. objective   | 643500.0    |        |
| Sparsity               | 794.4363    | Quasi-LR statistic | 4171.262    |        |
| Prob(Quasi-LR stat)    | 0.000000    |                    |             |        |

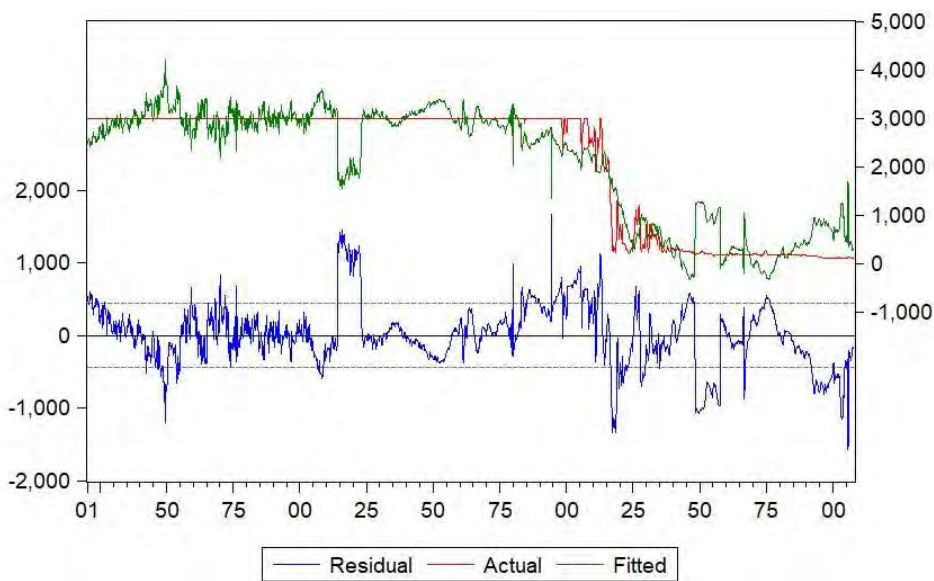


表 4-15 能见度（RVR）观测值与回归方程计算值时间序列折线图



模型的调整后  $R^2$  在 60% 以上，P 值为 0.00，可以认为模型通过了统计检验，具有较高的显著性和可信度。根据结果所示，能见度（RVR）与各气象因素之间的关系式为：

$$RVR_{(0.5)} = 156391.9 - 37.60 \times RH - 153.07 \times QFE + 514.28 \times DEW + 15.58 \times LIG - 87.27 \times WS2 - 3.88 \times WD2$$

由此可得，RH（相对湿度）每增加一个单位，RVR 将下降 37.7 个单位，根据相关文献<sup>[4]</sup>，不同相对湿度情况下，PM2.5 浓度对能见度的影响不同，当相对湿度较高时，较小的 PM2.5 浓度就能够导致低能见度的出现，可以推断，随着相对湿度的增加，气溶胶吸湿性增加显著，这会导致其具备更强的散射能力，导致能见度下降明显，另外，相对湿度越高，空气中的饱和水蒸气越容易凝结生成大量的小液滴，进一步导致能见度下降；QFE（飞机着陆地区最高点气压）每增加一个单位，RVR 下降 153.07 个单位，可见气压对能见度有显著的负影响，这与我们的日常感受相一致，往往在气压较低的时候，空气显得更为透明；DEW（露点温度）每增加一个单位，RVR 将上升 514.28 个单位，二者呈现正相关关系；LIG（灯光数据）每增加一个单位，RVR 将上升 15.58 个单位，机场灯光能够提高能见度水平，保障飞机起飞与着陆的安全性；风速和风向与能见度之间均 呈现出负相关关系，其中风速的影响大于风向，表现为系数更高，由于地面风场一般是通过稀释大气污染物和改变大气的稳定度来影响大气能见度的，风速较大时，有利于大气污染物的稀释扩散，能见度随之升高。

表 4-9 基于能见度（MOR）的最优分位点非线性回归结果

| Variable               | Coefficient | Std. Error         | t-Statistic | Prob.  |
|------------------------|-------------|--------------------|-------------|--------|
| C                      | 963911.5    | 38614.10           | 24.96268    | 0.0000 |
| EXPRH                  | -3.49E-41   | 7.58E-42           | -4.601648   | 0.0000 |
| QFE                    | -937.3770   | 37.28329           | -25.14201   | 0.0000 |
| LNDEW                  | -3333.424   | 585.7150           | -5.691205   | 0.0000 |
| LIG                    | 20.60510    | 6.671845           | 3.088367    | 0.0021 |
| WS2 <sup>2</sup>       | -62.98533   | 16.11602           | -3.908244   | 0.0001 |
| WD2 <sup>2</sup>       | -0.008795   | 0.002251           | -3.906926   | 0.0001 |
| Pseudo R-squared       | 0.698698    | Mean dependent var | 3659.521    |        |
| Adjusted R-squared     | 0.697436    | S.D. dependent var | 3510.776    |        |
| S.E. of regression     | 1164.794    | Objective          | 622019.1    |        |
| Quantile dependent var | 3600.000    | Restr. objective   | 2064440.    |        |
| Sparsity               | 2974.361    | Quasi-LR statistic | 4041.263    |        |
| Prob(Quasi-LR stat)    | 0.000000    |                    |             |        |

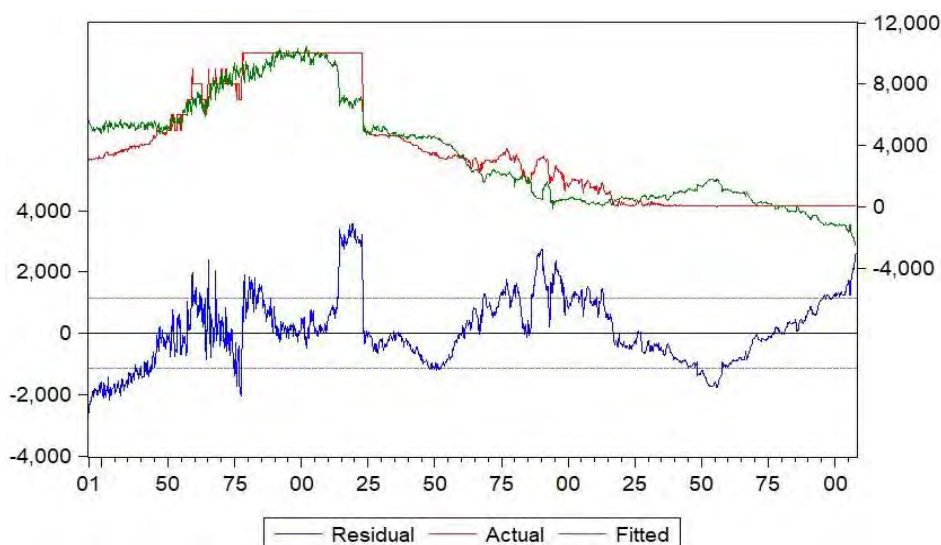


表 4-16 能见度（MOR）观测值与回归方程计算值时间序列折线图

模型的调整后  $R^2$  接近 70%，P 值为 0.00，可以认为模型通过了统计检验，具有较高的显著性和可信度。根据结果所示，能见度（MOR）与各气象因素之间的关系式为：

$$MOR_{(0.6)} = 963911.5 - 3.49 \times 10^{-41} e^{RH} - 937.3770 \times QFE - 3333.424 \times \ln(DEW) + 20.605 \times LIG - 62.985 \times WS2^2 - 0.009 \times WD2^2$$

由此可得，MOR 与 RH（相对湿度）之间为指数关系，当相对湿度低于某值时，能见度随相对湿度下降程度较小，当相对湿度达到某水平时，能见度随相对湿度下降而下降的幅度明显增大；MOR 与 QFE（飞机着陆地区最高点气压）以及 LIG（灯光数据）之间的关系与 RVR 相似；MOR 与 DEW（露点温度）之间为对数关系，可见 MOR 随 DEW 变化的速率会随 DEW 的增加而下降；MOR 与 WS2（风速）的平方呈现显著负关系。

## 5 问题二模型建立与求解

### 5.1 问题分析

题目要求我们根据题目提供的某机场视频数据（机场视频.zip）和能见度数据（机场AMOS 观测.zip），建立基于视频数据的能见度估计深度学习模型，并对估计的能见度进行精度评估。

传统的图像识别方法主要通过对图像提取例如纹理、颜色等明显的特征，然后训练分类以达到识别的目的，但这种方法准确率低、泛化能力差、且耗费时间长，故为了提高能见度的预测性能，本文采取一种基于卷积神经网络模型的迁移学习方法，以实现机场视频图像的识别，并对估计的能见度进行精度评估。为对比该深度学习模型与传统图像识别方法之间的差异，本文同时采用传统的机器学习方法对图像进行纹理特征、颜色特征的识别，针对每张图片提取 11 个维度的指标（包括灰度共生矩阵（GLCM）、颜色矩、颜色直方图等），并对以上指标以及能见度数据进行相关性分析，建立深度学习模型并进行精度评估。通过两种方法的对比，表明基于卷积神经网络模型的迁移学习方法确实具有更高的效率以及识别精度。图 5-1 为本问题的思路框架图。

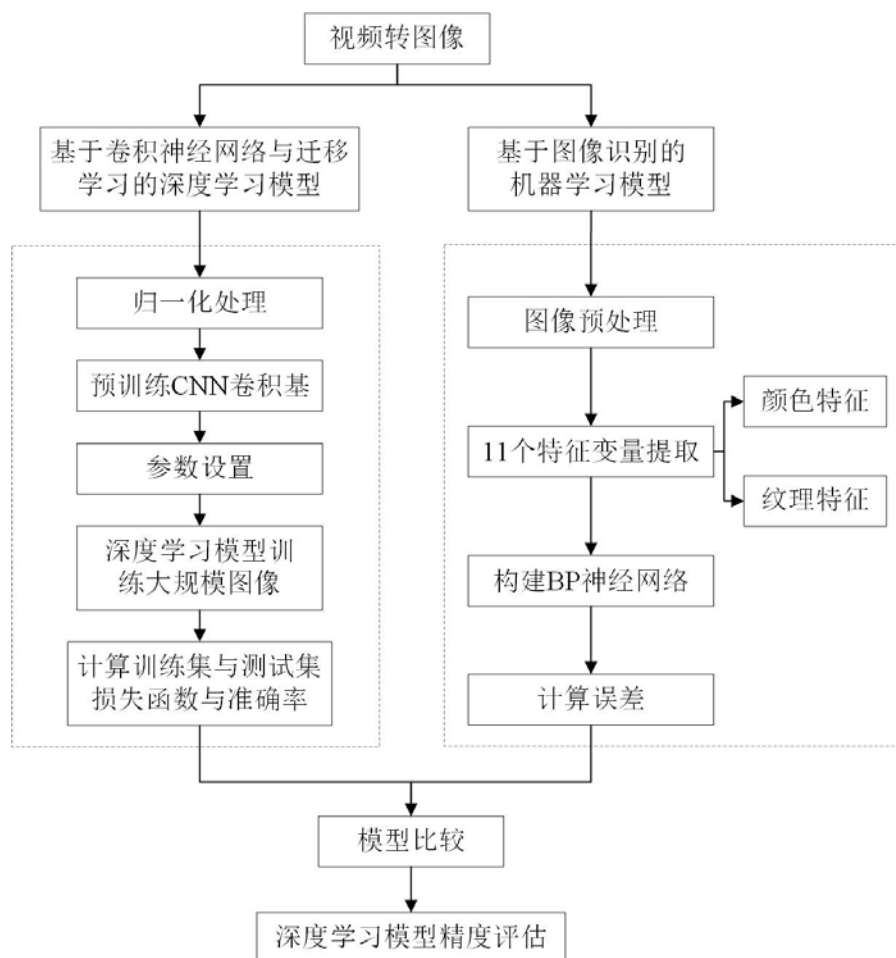


图 5-1 问题二思路框架图

根据附件中提供的机场视频与题中的能见度数据匹配筛选后截取的 1858 张图片，训练集数据占全部样本的百分之七十，测试集数据占百分之三十。将训练好的模型在机场能见度图像数据集上进行迁移训练，保留预训练模型所有卷积层的参数，冻结前层网络参数并精调神经网络模型，使其能适应机场视频图像的识别。结果表明，迁移学习后，神经网络模型预测精度得到了有效提高。

## 5.2 图像提取说明

由于机场视频的时间范围为 2020 年 3 月 13 日 0:00-11:47，而题目所提供的能见度数据与之的重合时间段为 2020 年 3 月 13 日 0:00-7:59，且时间间隔为 15 秒。为保持数据的一致性及对应关系，本文利用视频编辑软件 Adobe Premiere Pro 对机场视频（25fps）进行处理，按照视频左上角所标定的实际时间，每隔 15 秒导出一张图片。需要说明的是，于视频时间 00:47 处，视频中左上角标定实际时间为 00:47 时发生了中断，而后转至 00:52，再迅速转至 01:01，导致这段时间所对应的机场视频图片有所缺失。最后基于视频导出 1858 张图片用于下一步的提取与分析，图片像素为 1280\*720。

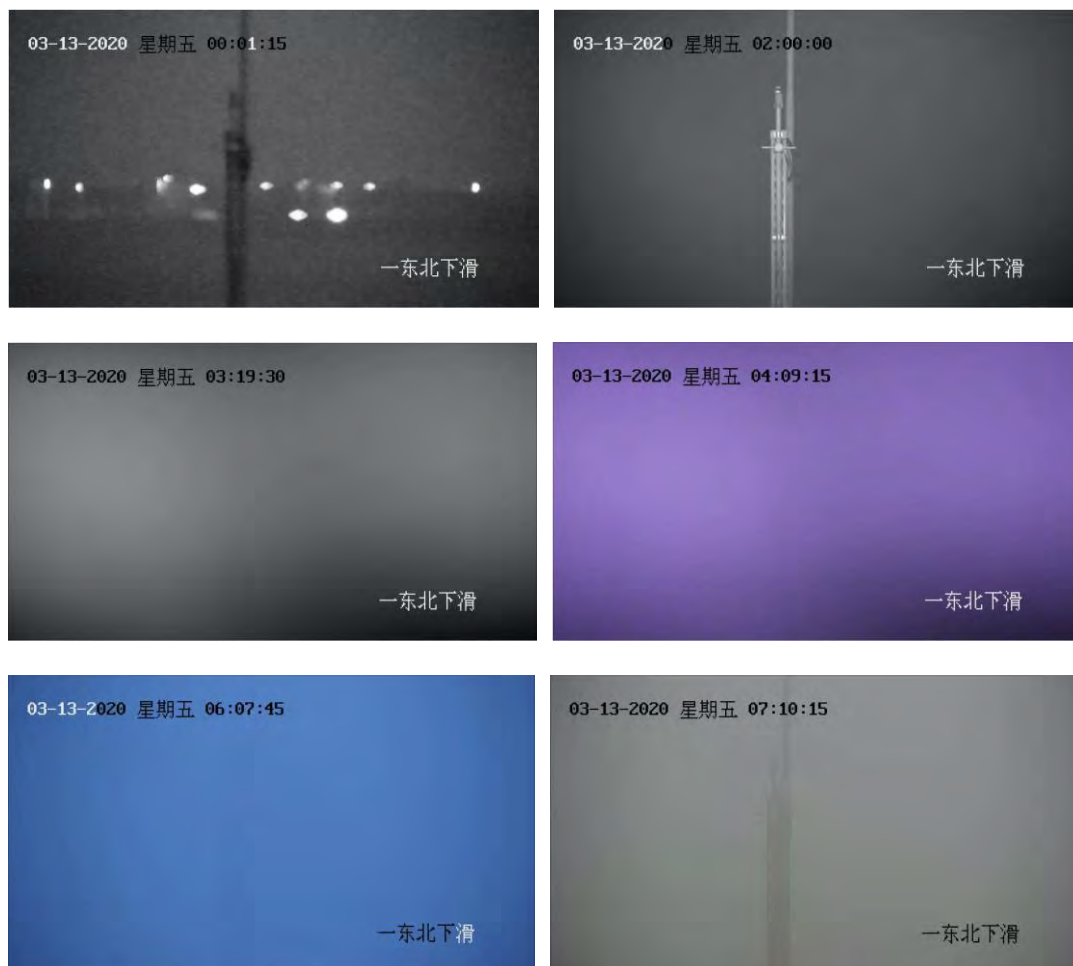


图 5-2 基于机场视频导出的部分图片

## 5.3 卷积神经网络与迁移学习

### 5.3.1 卷积神经网络

近年来，深度学习（Deep Learning, DL）已成为机器学习领域中的重要发展方向，随着相关研究的推进，使用卷积神经网络（Convolutional Neural Networks, CNN）解决图像分类与识别问题也逐渐成为主流，从 LeNet、AlexNet、GoogLeNet 到 ResNet，网络结构不断复杂化，图像识别分类的精度也逐步提高。

卷积神经网络是一种特殊的多层感知器或前馈神经网络，其将人工神经网络与深度学习方法相结合，使用一种基于梯度的改进反向传播算法来训练网络中的权重，实现了深度学习的算法<sup>[5]</sup>，其结构一般如图 5-3 所示，输入层读取规则化的图片，每一层的单元再将前一层中的一组小的局部近邻的单元作为输入。

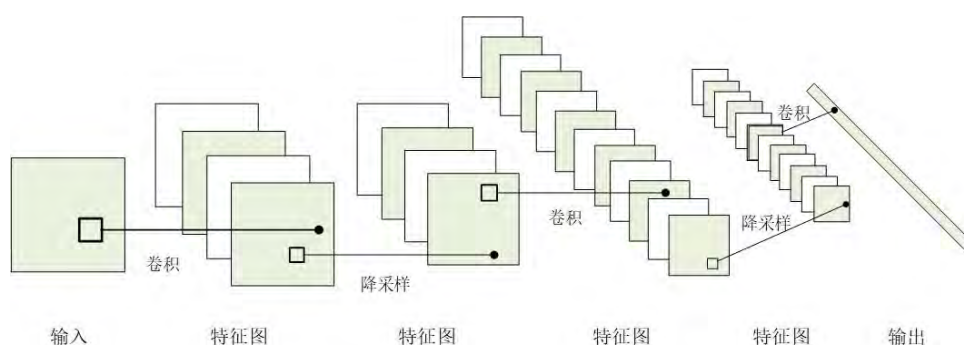


图 5-3 卷积神经网络结构

卷积过程如下式所示：

$$X_j^L = f \left( \sum_{i \in M_j} X_i^{L-1} \cdot K_{ij}^L + b_j^L \right) \quad (5-1)$$

其中， $l$  表示层数， $X_j^l$  表示  $l$  层第  $j$  个特征图， $M_j$  表示感受野， $K$  表示卷积核， $b$  代表偏置， $f$  代表激活函数。

卷积神经网络利用一组分布于图像不同位置但权值向量相同的单元以获取图像特征，进而构成一幅特征图，不同位置上，来自不同特征图的单元获得不同类型的特征。而一个卷积层中通常包括多个具有不同权值向量的特征图，使得同一个位置能够获得多种不同特征，采样层则能够使模型保留全局主要特征并减少模型参数<sup>[5]</sup>。

### 5.3.2 迁移学习

近年来，迁移学习引起了广泛的关注与研究，其主要是运用已有的知识对其他相关领域的问题进行求解，是一种全新的机器学习方法。传统机器学习中存在两个基本假设：（1）用于学习的训练样本与测试样本满足独立同分布；（2）需要足够可利用的训练样本以学习得到好的分类模型<sup>[6]</sup>，而迁移学习则放宽了以上假设，能够在存在差异却又存在联系的

两个领域挖掘出二者不变的本质特征和结构，实现跨领域的非结构化数据的分类和预测问题<sup>[7]</sup>。

基于卷积神经网络的迁移学习就是将已经训练好的模型与其参数在新的识别任务上再次进行训练，而分类器前的网络相当于传统特征提取器，以特征代表图片从而进行分类识别。

## 5.4 基于图像识别的机器学习模型

### 5.4.1 图像处理

本文根据机场视频与题中所给的能见度数据匹配筛选后截取了 1858 张图片，首先进行图像模式识别，提取的图像特征主要有颜色特征、纹理特征。图像识别的一般步骤为如下图所示：

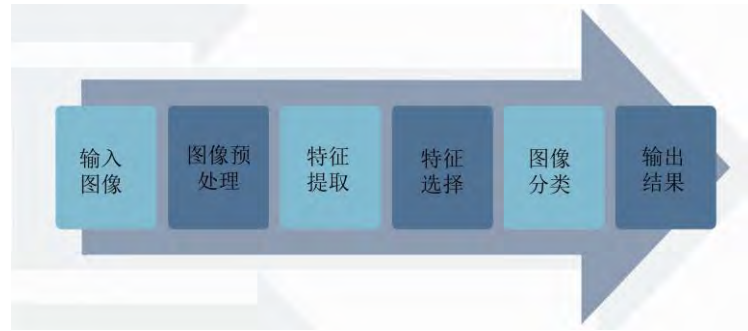


图 5-4 图像识别的一般步骤

(1) 颜色特征。颜色特征属于全局特征，较为直观。它是基于图像中每个像素点来提取出来的特征，且由于其对于图像内区域或像素点的大小、方向变化等不敏感，因此颜色特征的稳定性和抗干扰能力较强。但是颜色特征属于全局特征，不能很好的反映图像中的局部信息，表征图像局部变化情况。颜色信息主要分布在低阶矩中，所以本文选用一阶矩、二阶矩和三阶矩来表达图像的颜色分布。该方法的优点是：不需要颜色空间量化，特征向量维数低。一阶、二阶、三阶颜色矩的数学定义为：

$$\mu_i = \frac{1}{N} \sum_{j=1}^N p_{i,j} \quad (5-2)$$

$$\sigma_i = \left( \frac{1}{N} \sum_{j=1}^N (p_{i,j} - \mu_i)^2 \right)^{\frac{1}{2}} \quad (5-3)$$

$$s_i = \left( \frac{1}{N} \sum_{j=1}^N (p_{i,j} - \mu_i)^3 \right)^{\frac{1}{3}} \quad (5-4)$$

其中， $p_{i,j}$  表示彩色图像第  $j$  个像素的第  $i$  个颜色分量， $N$  表示图像中的像素个数。

(2) 纹理特征。纹理特征反映的是图像中同质现象的视觉特征，体现物体表面的具有缓慢变换或周期性变化的表面组织结构排列属性。图像纹理通过像素及其周围空间邻域的灰度分布来表现，即局部纹理信息。另外，局部纹理信息不同程度上的重复性，就是全



局纹理信息。本文选取灰度共生矩阵（Gray-level Co-occurrence Matrix，GLCM）作为纹理特征的提取与匹配方法，它主要依赖于能量、角二阶矩、熵和相关性四个参数。与颜色特征相比，纹理特征对于噪声的抗干扰能力增强。

1. 能量（Energy）是灰度共生矩阵各元素值的平方和，是对图像纹理的灰度变化稳定程度的度量，反应了图像灰度分布均匀程度和纹理粗细度。能量值大表明当前纹理是一种规则变化较为稳定的纹理。

2. 熵（Entropy）测量图像纹理的随机性（强度分布）。熵是图像包含信息量的随机性度量，表现了图像的复杂程度。当共生矩阵中所有值均相等或者像素值表现出最大的随机性时，熵最大；因此熵值表明了图像灰度分布的复杂程度，熵值越大，图像越复杂。

3. 角二阶矩（ASM）用来描述图像灰度分布的均匀程度和纹理的粗细程度。如果 GLCM 的所有值都非常接近，则 ASM 值较小；如果矩阵元素取值差别较大，则 ASM 值较大。

4. 相关性（Correlation）。它度量空间灰度共生矩阵元素在行或列方向上的相似程度，因此，相关值大小反映了图像中局部灰度相关性。当矩阵元素值均匀相等时，相关值就大；相反，如果矩阵像元值相差很大则相关值小。

| 表 5-1 图像特征提取指标 |        |
|----------------|--------|
| 特征             | 指标     |
| 颜色特征           | 图像一阶矩  |
|                | 图像二阶矩  |
|                | 图像三阶矩  |
|                | 能量-均值  |
|                | 能量方差   |
| 纹理特征           | 熵均值    |
|                | 熵方差    |
|                | 角二阶矩均值 |
|                | 角二阶矩方差 |
|                | 相关性均值  |
|                | 相关性方差  |

5.4.2 BP 神经网络

BP 神经网络是 1986 年由 Rumelhart 和 McClelland 为首的科学家提出的概念，是一种按照误差逆向传播算法训练的多层前馈神经网络，是应用最广泛的神经网络。它的基本思想是梯度下降法，利用梯度搜索技术，以期使网络的实际输出值和期望输出值的误差均方差为最小。

BP 神经网络由输入层、隐含层、输出层三大层组成，其中隐含层可以为一层或多层。隐含层内的神经元与外界没有直接的联系，但其状态的改变能影响输入和输出之间的关系。

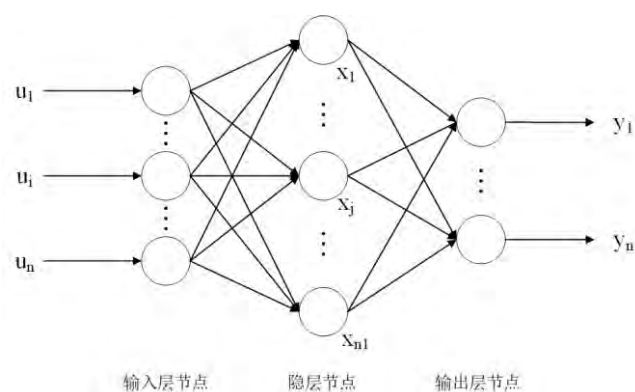


图 5-5 BP 神经网络系统结构

## 5.5 实验设计与结果分析

### 5.5.1 实验环境及方案

实验是在 Windows 10 专业版系统安装 Anaconda 3，基于谷歌研发的第二代人工智能学习系统 Tensorflow 及 Keras 深度学习框架开发环境，完成实验训练及验证结果。

实验首先将原始提取得到的机场图像进行归一化处理，然后经过预训练卷积网络的卷积基，利用大规模数据集上预训练模型的参数权重初始化网络，保存输出图像的特征，将保存数据输入进行密集分类层训练验证，保存参数，输出识别结果。

具体研究流程如图 5-6 所示。

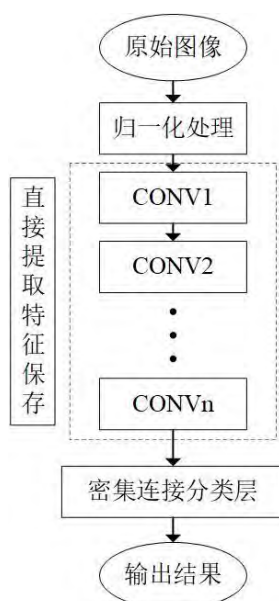


图 5-6 实验流程图

### 5.5.2 模型参数设置

(1) 学习率  $lr$  是深度学习模型里最重要的参数之一，不同的学习率，模型的效果可能会是天壤之别。一般而言， $lr$  可先从一个较大的值开始，如 0.1（或者 loss 为非 Nan 时的最大值），这样也可以实现快速调试。

(2) 动量参数方法主要是为了解决 Hessian 矩阵病态条件问题的，同时也可以加速学习，一般将参数设为 0.5、0.9 或者 0.99，分别表示最大速度 2 倍、10 倍、100 倍于 SGD 的算法。

(3) Gamma 变换是对输入图像灰度值进行的非线性操作，使输出图像灰度值与输入图像灰度值呈指数关系：这个指数即为 Gamma。Gamma 变换就是用来图像增强，其提升了暗部细节，简单来说就是通过非线性变换，让图像从曝光强度的线性响应变得更接近人眼感受的响应，即将漂白（相机曝光）或过暗（曝光不足）的图片，进行矫正。本文设定为 0.1。

(4) Step\_size 表示每运行一个 step，更新一次参数权重，即进行一次学习，每一次更新参数需要训练的样本数来进行运算学习，本文为 1298 个样本。根据运算结果调整更新一次参数。

最终，本文模型参数设置初始学习率为 0.1，优化后的学习率为 0.001，迭代次数为 100 次，Gamma 为 0.1，step\_size 为 7。

### 5.5.3 实验结果分析

(1) 基于卷积神经网络与迁移学习的深度学习模型

基于以上模型设定，实验分以下步骤进行：

步骤 1，创建机场大雾视频图像库，该数据库包括训练样本集和测试样本集，根据题中所给的机场能见度数据分别进行分类并进行标签化处理，分为 22 类。得到了用于训练的包含 1858 张图片的训练数据集，测试数据集无标签，共有 558 张图片；

步骤 2，采用卷积神经网络提取训练样本集里预处理后的雾霾图像的最远能见度边缘特征，得到多个特征图；将特征图通过前向传播至配置好的卷积神经网络进行训练，并通过反向传播算法进行卷积神经网络中各层间权重的调整，反复迭代求取得用于雾霾图像分类的卷积神经网络模型，再通过预处理后的交叉验证样本集对卷积神经网络模型进行优化，最终获得一个用于对雾霾图像分类的能见度检测模型；

步骤 3，利用步骤 2 中得到的能见度检测模型对于机场摄像机拍摄的图片进行分类判断，将判读的结果与题中所给的能见度数据进行对比，从而实现一个可以对雾霾状况的实时检测的快速分类模型。模型在训练和测试过程中的损失函数和准确率函数如图 5-7 和图

5-8 所示， $accuracy = \frac{TP + TN}{TP + TN + FP + FN}$ ，误差如图 5-9 和图 5-10 所示。

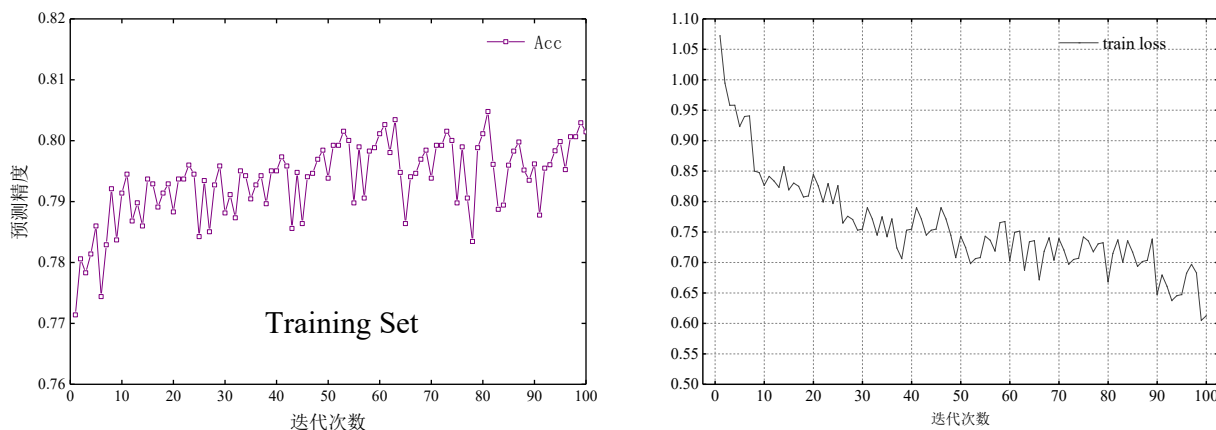


图 5-7 训练集预测精度和 loss 函数变化表

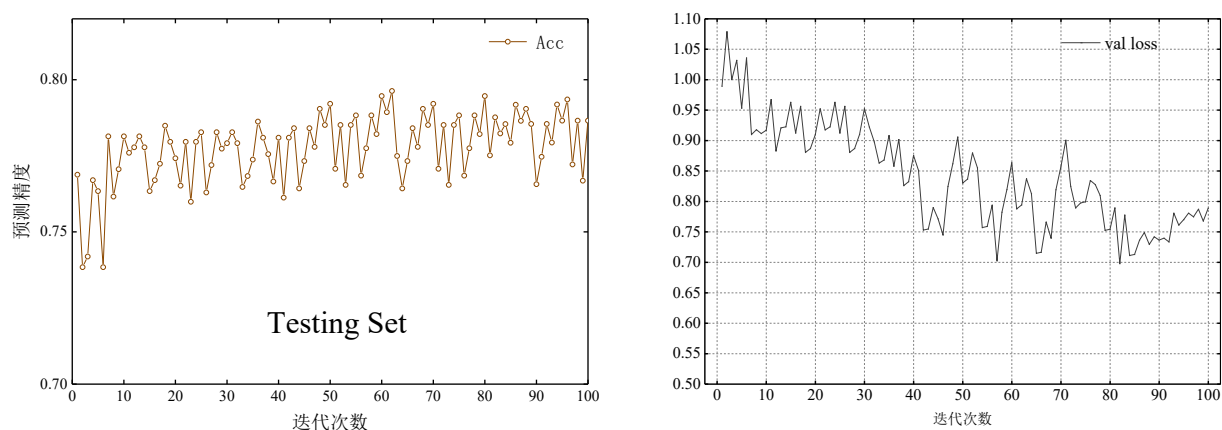


图 5-8 测试集预测精度和 loss 函数变化表

由图 5-7 和图 5-8 可知，本文所采用基于卷积神经网络和迁移学习的深度学习模型，能够获得较好的预测精度，在训练集准确率最高达到了 80.48%，在测试集中准确率最高达到了 79.63%。从 loss 函数变化图来看，训练集和测试集的 loss 函数呈现波动下降的趋势，表明随着迭代次数的增加，本文所采用的模型精度在不断提高。从误差图来看，模型的绝对误差大部分集中在 0 中心线，误差较小。

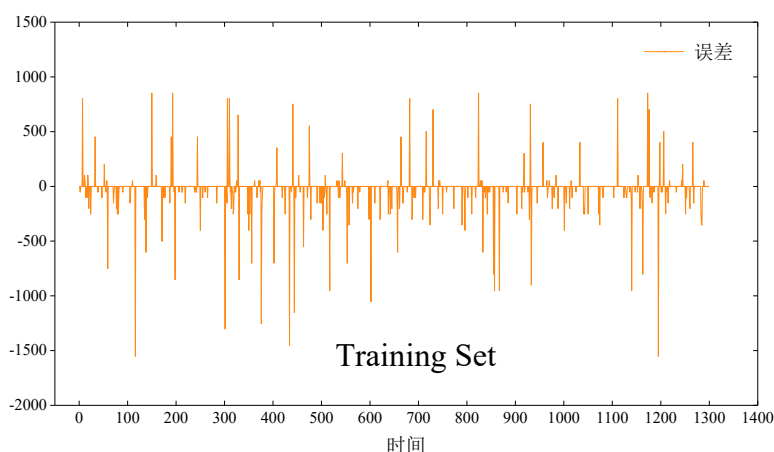


图 5-9 训练集误差

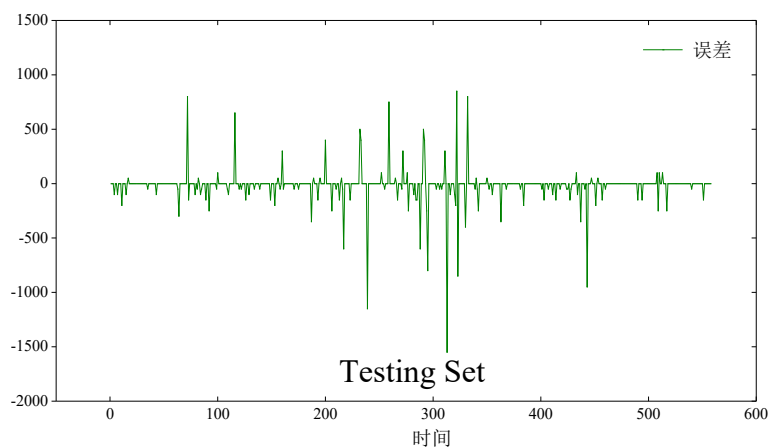


图 5-10 测试集误差

## （2）基于图像识别的机器学习模型

本文依据 5.4.1 节的图像识别数据，提取了 11 个指标变量，包括了图像的颜色特征和纹理特征，同时利用 Pearson 相关性分析剔除了相关性较强的 3 个指标，最终利用了 10 个指标进行预测。经过多次训练比较，当隐含层点数为 12 时，网络训练速度最快、性能最好。BP 神经网络隐含层的激活函数采用 `logsig` 函数，输出层的传递函数采用 `tansig` 函数。网络训练采用 Levenberg-Marquardt 算法进行学习，其训练函数为 `trainlm`，最大训练次数设为 1000，期望误差设为 0.01，预测值与实际值之间的误差采用  $R$  来表示。对 1298 个图片处理数据按上述模型进行训练，在训练过程中不断调整权重和阈值，使误差最小，直至达到预先要求。BP 神经网络训练过程中，在 198 步迭代后期望误差  $R$  稳定 0.792 在左右。预测精度明显弱于基于卷积神经网络与迁移学习的深度学习模型。图 5-11 是采用 BP 神经网络模型得到的预测值和实际值比较结果图。

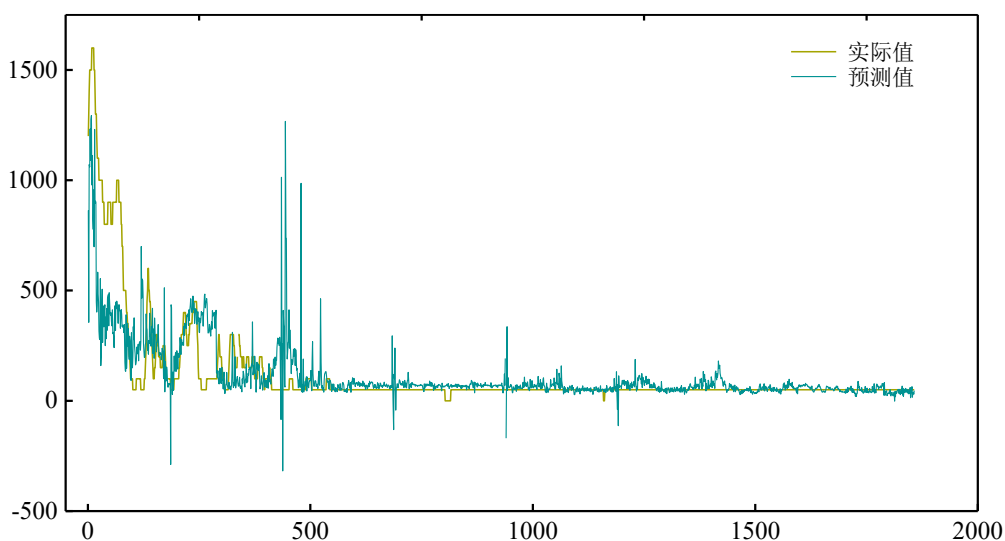


图 5-11 采用 BP 神经网络预测值和实际值比较

## 6 问题三模型建立与求解

### 6.1 问题分析

题目三要求我们建立不依赖能见度仪观测数据的能见度估计算法（提示：事实上，在有雾的情形可以估计视频中物体的景深。反过来，理论上也可以利用视频中不同景深的物体，在不同能见度下的亮度差异估计能见度），讨论相关算法实现过程，并针对题目提供的一段视频（高速公路视频截图.zip）绘制该时间段这段高速公路能见度随时间变化曲线。

为解决该问题，我们首先基于暗通道先验理论对能见度与透射率、消光系数之间的数理关系进行推导，获取能见度的测算式，其次利用 MATLAB 对题目提供的高速公路图片求解暗通道图像、采用梯度导向滤波对获得的粗透射率进行细化、求解全局大气光照等获得去雾后的清晰图片，同时获得图像的透射率数据，最后基于透射率数据计算图像消光系数，进而测算得到图像的能见度数据、绘制能见度随时间变化曲线。图 6-1 为本问题的思路框架图。

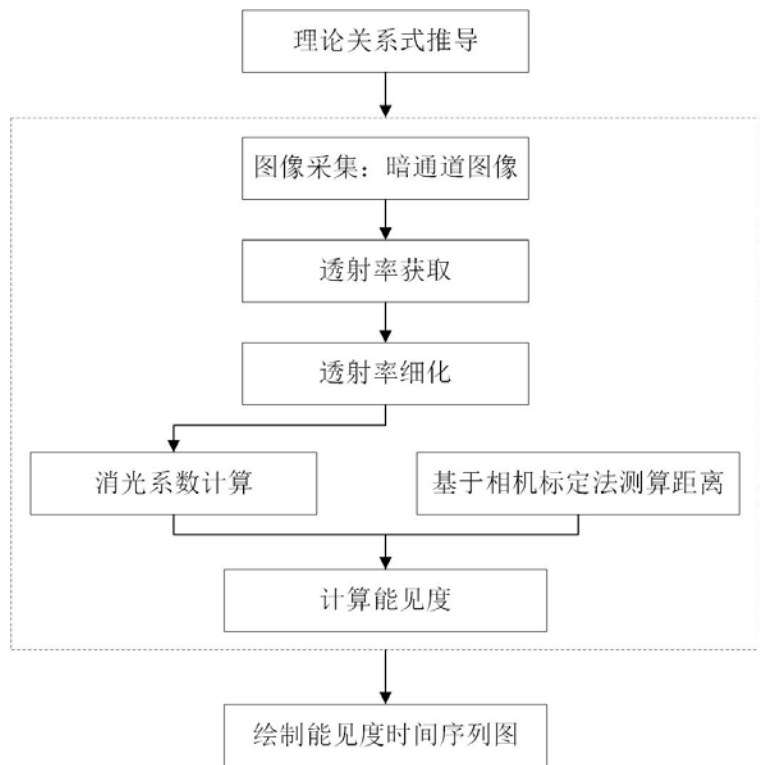


图 6-1 问题三思路框架图

### 6.2 能见度检测理论模型

能见度的概念最早是由气象部门提出的，当人们沿水平方向观察某物时，空气中的气溶胶粒子会产生散射形成漫散光现象，导致人们在观察目标物时产生一定的遮挡感，在这



种情况下，间接缩短了人们观察目标物的距离，因此产生了“远近”的差距，亦即人们所说的能见度<sup>[8]</sup>。

由于大气中存在的各类悬浮物、大气分子及气溶胶粒子，外界物体的亮度在进入人眼时会发生能量的衰减。当这些大气环境中的大气分子和气溶胶粒子对光线产生了吸收和散射时，光线的传播偏离了原有的方向，从而导致观察到的图像质量下降、不清晰，而相关因素（如气温、相对湿度等）的影响则会加剧该过程，进一步影响能见度。

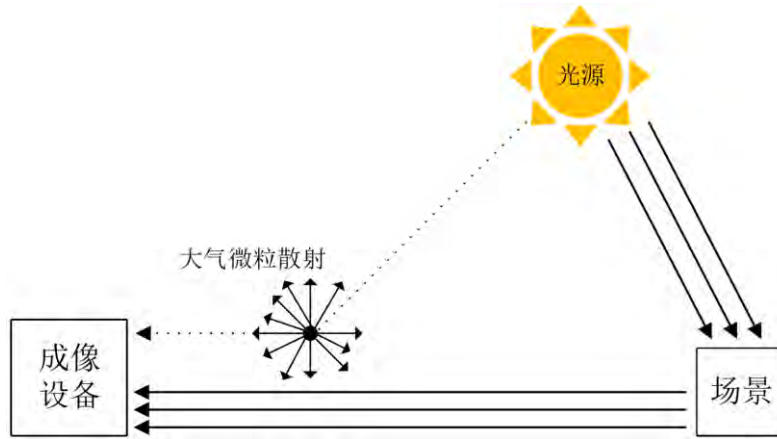


图 6-2 大气散射物理模型

为了从物理模型角度对有雾的图像进行清晰化处理，首先需要了解大气散射模型。图 6-2 展示了大气在空气中的散射物理模型，其主要包含两部分，第一部分称为直接衰减项（Direct Attenuation），也称直接传播，第二部分称为大气光照（Airlight）项。直接衰减项解释了大气光从周围环境中的一个点传播到成像设备（观察者眼睛）的路径中产生的衰减，该衰减是由大气微粒对大气光的散射导致，故随着传播距离的增加，成像设备所接收到的光照强度降低趋势呈指数形式；大气光照项则解释了环境中存在的其他天空光经过散射后到达观测点的光强度，随传播距离增加，环境光强度逐步增强。

20 世纪初 Koschmieder 基于光学视程的定义研究了大气亮度衰减问题，推导出大气消光系数的等式，该等式在计算机视觉和图像学领域得到了广泛的使用。

$$L = L_0 e^{\beta d} + L_\infty (1 - e^{\beta d}) \quad (6-1)$$

式（6-1）中， $L$  表示目标物的固有亮度， $L_0$  表示周围环境亮度， $\beta$  表示大气消光系数， $d$  表示目标物的距离， $L_\infty$  表示天空大气的亮度。

在 Koschmieder 的研究基础上，Dumtley 基于大气衰减模型进一步提出了雾天环境下的对比衰减定律：

$$C = \frac{e^{\beta d} (L_0 - L_\infty)}{L_\infty} \quad (6-2)$$

式（6-2）中， $C$  表示亮度对比度。

根据国际照明委员会对能见度的定义：人类的肉眼在对比度阈值  $C/C_0$  的数值为 0.05 的情况下所能观测到的规定尺寸的黑色目标物（ $C_0=1$ ）所对应的最大的距离<sup>[9]</sup>。根据该定义及式（6-2）的转换式，推导出能见度检测的表达式：

$$V = -\frac{1}{\beta} \ln(0.05) = \frac{3}{\beta} \quad (6-3)$$

式（6-3）中， $V$  表示能见度，由该公式我们可以将测量公路能见度的问题转化为测量图像大气消光系数的问题。

### 6.3 基于暗通道先验理论的能见度检测

本文采用的基于暗通道先验理论的能见度检测算法将通过图像处理计算能见度的问题转化为计算大气消光系数的问题，检验算法的基本框架如下图 6-3 所示。

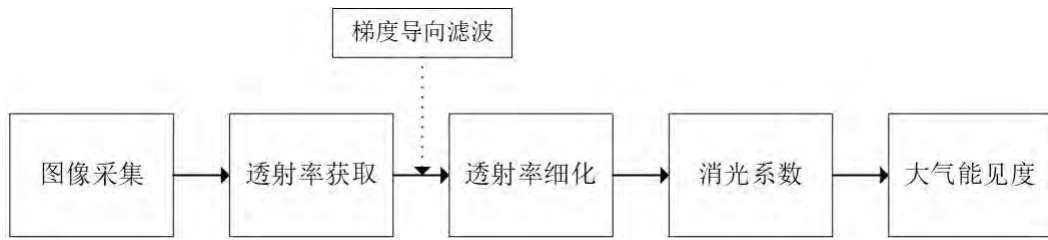


图 6-3 能见度检测算法基本框架

#### 6.3.1 暗通道先验理论

He 等<sup>[10]</sup>通过对大量在户外拍摄的自然景物图片进行统计分析，得出一个结论：在绝大多数非天空的局部区域里，某一些像素总会（R、G、B 中至少一个颜色通道）具有较小的值，由此提出暗通道先验理论（Dark Channel Prior）。该理论指出，对于一张户外无雾的彩色图像时，我们可以认为其暗通道数值除了天空区域以外的其他部分趋向于 0。基于该现象规律，对于任意一幅图像而言，我们都有以下关于其对应暗通道的计算表达式：

$$J^{dark}(x) = \min_{y \in \Omega(x)} \left( \min_{c \in \{r, g, b\}} J^c(y) \right) \quad (6-4)$$

在式（6-4）中， $J^{dark}(x)$  表示经操作之后的暗通道图像， $J^c$  表示的是图像中 R、G、B 三个颜色通道中的某一个颜色通道， $\Omega_x$  表示的是以  $x$  为中心的一个邻域。 $\min_{c \in \{r, g, b\}}$  表示的是对图像中的某个像素进行的最小化操作， $\min_{y \in \Omega(x)}$  表示的是在像素  $x$  邻域范围内存在的像素最小值，一个暗通道操作过程中包含着上述两次最小化操作。

由于户外无雾彩色图像的构成大多数为深色物体、彩色物体以及阴影，因此图像暗通道值趋向于 0，所对应的暗通道图像也是昏暗的。相比之下，雾霾天气下的图片亮度由于大气光成分的存在变得更为明亮一些，故暗通道图像的亮度会明显更高，其透射率相对也

会较小。换言之，含有雾霾的图像其对应的暗通道亮度与其雾霾信息的大小呈现正比例关系，利用该特性可估计题目所给予的公路图片中的大气透射率。

### 6.3.2 大气透射率获取

关于雾天图像成像原理的描述，在计算机视觉和计算机图像学领域广泛使用的数学模型如下<sup>[11]</sup>：

$$I(x) = J(x)t(x) + A(1-t(x)) \quad (6-5)$$

其中， $I$  表示的是检测到的有雾图像， $J$  表示的是目标物所在场景的反射光强度， $A$  表示的是环境中大气光线的强度， $t$  表示的是所处环境中传播介质的透射率大小，即场景反射光通过大气中介质到达成像设备的比率， $x$  表示的是图像像素点的坐标。

基于先验知识假设研究中大气光的数值  $A$  为已知，分别对每个通道进行通  $A$  正规化操作，将式 (6-5) 标准化得到下式 (6-6)：

$$\frac{I^c(x)}{A^c} = t(x) \frac{J^c(x)}{A^c} + 1 - t(x) \quad (6-6)$$

进而假设在研究中数字图像的透射率在  $\Omega(x)$  上保持恒定不变，将其记作  $\tilde{t}(x)$ 。然后对式 (6-6) 进行暗通道处理得到下式 (6-7)：

$$\min_{y \in \Omega(x)} \left( \min_c \frac{I^c(x)}{A^c} \right) = \tilde{t}(x) \min_{y \in \Omega(x)} \left( \min_c \frac{J^c(x)}{A^c} \right) + 1 - \tilde{t}(x) \quad (6-7)$$

目标物所在的场景强度图  $J(x)$  即为要恢复的无雾图像，因此  $J(x)$  满足暗通道先验理论，其暗通道值是接近于 0 的，有如下公式：

$$J^{dark}(x) = \min_{y \in \Omega(x)} \left( \min_c J^c(y) \right) = 0 \quad (6-8)$$

又式 (6-7) 中的  $A^c$  恒为正值，故可以得到如下式的结果：

$$\min_{y \in \Omega(x)} \left( \min_c \frac{J^c(y)}{A^c} \right) = 0 \quad (6-9)$$

由式 (6-7) 和式 (6-9) 可以得到：

$$\tilde{t}(x) = 1 - \min_{y \in \Omega(x)} \left( \min_c \frac{I^c(y)}{A^c} \right) \quad (6-10)$$

式 (6-9) 中， $\min_{y \in \Omega(x)} \left( \min_c \frac{I^c(y)}{A^c} \right)$  表示的是标准化雾霾图像  $\frac{I^c(y)}{A^c}$  的暗通道操作，

上述步骤即为基于暗通道先验理论获取大气透射率的基本原理。

需要注意的是，由于天空区域在理论上并不符合暗通道先验理论的原理，但在雾霾图像中普遍存在着大气光强  $A$  无限接近于天空区域的颜色，对天空区域的暗通道处理如下式所示：

$$\min_{y \in \Omega(x)} \left( \min_c \frac{I^c(y)}{A^c} \right) \rightarrow 1 \quad (6-11)$$

由于在实际生活中，即使在晴天图像中，景深较远的场景中依然会存在少量的雾气，因此我们引入一个权重参量  $\omega$  ( $0 < \omega < 1$ ) 减小去雾程度，式(6-10)可转变为以下式(6-12)：

$$\tilde{t}(x) = 1 - \omega \min_{y \in \Omega(x)} \left( \min_c \frac{I^c(y)}{A^c} \right) \quad (6-12)$$

上式中， $\omega$  作为一个经验值常量依靠实验经验设置，取值 0.95。

### 6.3.3 基于导向滤波的透射率细化

通过上述的操作步骤得到的透射率存在一定的误差，在图像上则表现为模糊效应，因此需要对透射率进行细化处理。He 等<sup>[10]</sup>曾采用软抠图 (Soft Matting) 的方式对得到的粗透射率进行细化，但由于该方法时间和内存花费较大，本文采用基于导向滤波的透射率细化，相对而言具有明显的优势。

导向滤波是一种能够保存边缘信息的图像滤波方法，运用该方法细化透射率的操作可以看作是类似于双边带滤波的边缘信息保留操作，其核心是假设指导图像与滤波后输出图像之间呈现局部线性关系。

设定  $p$  表示输入图像， $q$  表示输出图像， $I$  表示指导图像，那么  $I$  在以某个像素  $k$  为中心的窗口  $\omega_k$  中通过指导滤波的局部线性模型进行线性变换后输出的图像  $q$  的表达式为：

$$q_i = \alpha_k I_i + b_k, \forall i \in \omega_k \quad (6-13)$$

上式中， $(a_k, b_k)$  表示的矩形窗  $\omega_k$  中保持不变的线性系数，由于  $\nabla q = \alpha \nabla I$ ，所以它们之间的关系是线性的，因此只有在指导图像  $I$  存在边缘时滤波输出图像  $q$  才会存在边缘信息。

通过对目标图像  $p$  进行约束，可以得到  $(a_k, b_k)$  的实际数值，将  $p$  中的噪声以及纹理信息 ( $n$ ) 去除后，能够得到输出图像  $q$ ：

$$q_i = p_i - n_i \quad (6-14)$$

在保持上述公式所示的线性关系的同时，将输入图像  $p$  与输出图像  $q$  之间的差别降至最低，即为导向滤波法的核心目的。

对于一个窗口  $\omega_k$ ，考虑以下代价函数：

$$E(a_k, b_k) = \sum_{i \in \omega_k} \left( (a_k I_i + b_k - p_i)^2 + \varepsilon \alpha_k^2 \right) \quad (6-15)$$

其中,  $\varepsilon$  是为防止  $\alpha_k$  过大设置的参数。

利用线性回归求解上式 (6-15) 可得:

$$\begin{cases} a_k = \frac{\frac{1}{|\omega|} \sum_{i \in \omega_k} I_i p_i - \mu_k \bar{p}_k}{\sigma_k^2 + \varepsilon} \\ b_k = \bar{p}_k - \alpha_k \mu_k \end{cases} \quad (6-16)$$

上述公式中,  $\mu_k$  表示的是所选窗口  $\omega_k$  的均值,  $\sigma_k^2$  表示所选窗口  $\omega_k$  的方差,  $|\omega|$  表示的是所选窗口  $\omega_k$  内所包含的像素个数,  $\bar{p}$  表示的是输入图像  $p$  在窗口  $\omega_k$  内的均值。

对于输入图像包含的所有像素点的所有局部窗口均应用上述线性模型, 将某个像素点  $i$  的输出  $q_i$  设定为其所对应的每个窗口输出的平均值, 由此得到滤波输出结果为:

$$q_i = \frac{1}{|\omega|} \sum_{k, i \in \omega_k} (\alpha_k I_i + b_k) = \bar{\alpha}_i I_i + \bar{b}_i \quad (6-17)$$

上式中,  $\bar{\alpha}_i$  表示的是  $\alpha$  在窗口  $\omega_k$  内的均值, 且  $\bar{\alpha}_i = \frac{1}{|\omega|} \sum_{k \in \omega_i} \alpha_k$ ;  $\bar{b}_i$  表示的是  $b$  在窗口  $\omega_k$  内的均值, 且  $\bar{b}_i = \frac{1}{|\omega|} \sum_{k \in \omega_i} b_k$ ;  $(\bar{\alpha}_i, \bar{b}_i)$  表示的是平均滤波的输出, 由于其梯度很小, 故满足  $\nabla q \approx \bar{\alpha} \nabla I$ , 表明输出图像  $q$  的局部特征与指导图像  $I$  相似且整体特征与输入图像  $p$  相似, 能够达到较好的透射率细化效果。

#### 6.3.4 能见度的消光系数测算

根据视频提供的公路路面图片, 假设公路路面是光滑的水平平面, 测算图像的消光系数可以通过以下公式计算得到:

$$\beta(i, j) = \frac{|\ln t(i, j)|}{L} \quad (6-18)$$

$$\beta = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \beta(i, j) \quad (6-19)$$

式 (6-18) 中,  $\beta(i, j)$  表示的是图像中任意某点像素  $(i, j)$  所对应的消光系数,  $L$  表示的是摄像头到目标点的距离, 由于目标点分布较广, 故  $L$  选取摄像头距离中心点的距离, 其具体测算方法于下节叙述。

式（6-19）中， $\beta$ 表示的是图像的消光系数， $N$ 表示的是检测图像中所检测的像素点的数量。

### 6.3.5 基于题目推导的能见度测算式

根据题目所提供的能见度测量基本方程：

$$F = F_0 e^{-\sigma z} \quad (6-20)$$

其中， $F$ 和 $F_0$ 分别表示观测和入射的光照强度，参数 $\sigma$ 称为衰减指数，亦即前文测算的消光系数 $\beta$ ，得到气象光学视程即能见度的测算公式为：

$$MOR = z = \frac{\ln(F / F_0)}{-\sigma} = \frac{\ln(0.05)}{-\sigma} \quad (6-21)$$

该式与式（6-3）一致，即为能见度与消光系数（衰减系数）之间的关系式。

## 6.4 基于相机标定法的距离估计

根据式（6-18），为计算图像中任意某点像素 $(i, j)$ 所对应的消光系数，需要首先测量摄像头到目标点的距离 $L$ 。由于题目未提供此数据，且按照相关资料主观估计该距离存在较大的误差，故本文主要介绍采用相机标定法对视频中的距离 $L$ 进行估计。

### 6.4.1 相机标定法介绍

相机标定法的核心在于从物体或场景的二维图像中获得物体或场景的三维性质，随着机器视觉的不断发展，其已经日渐成为三维重建、工业检测以及航空航天等领域的关键技术之一。目前，主要有三种标定方法，分别是自标定法、主动视觉标定法和传统标定法<sup>[12]</sup>。自标定法不依赖于标定参照物，操作复杂且对实时性要求高；主动视觉标定法需要使用高精度的主动视觉平台进行标定，成本较高故使用并不广泛；传统标定法则是利用尺寸已知的标定物，利用数学变化以及相关计算从而获得所需参数，该方法虽然略复杂，但由于具备较高的精度且成本相对较低，故而受到多数人的青睐。在传统的标定方法中，张正友标定法是目前求解相机内参和畸变系数的经典方法，也是最受欢迎的方法。

### 6.4.2 相机标定法原理

相机标定中有四个不同层次的坐标系：世界坐标系、相机坐标系和图像坐标系以及计算机坐标系，一般的相机标定系统坐标系如图 6-4 所示。

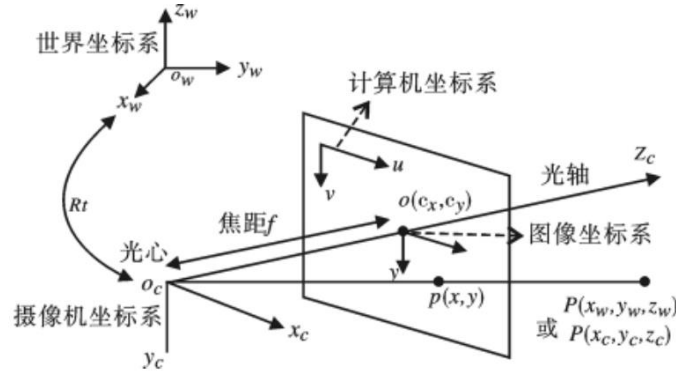


图 6-4 相机标定法坐标系

世界坐标系是指真实世界坐标系，用  $x_w y_w z_w$  表示，它是客观世界的绝对坐标；摄像机坐标系是以相机光心为中心建立的坐标系，用  $x_c y_c z_c$  表示，其中一般取相机的光轴为  $z_c$ ；图像坐标系是指在相机内形成的像平面二维坐标系，一般取像平面与相机坐标系平面平行，用  $xy$  表示；计算机坐标系是在相机内形成的坐标系，一般取像平面左上角为原点，用  $uv$  表示。

为了描述相机的成像原理，可以采用针孔模型（又称线性模型）对相机进行数学建模，根据针孔成像原理可以得到从相机坐标系到图像坐标系的变换：

$$\begin{cases} x = f \frac{x_c}{z_c} \\ y = f \frac{y_c}{z_c} \end{cases} \Rightarrow z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \quad (6-22)$$

从图像坐标系到计算机坐标系的变换：

$$\begin{cases} u = \frac{x}{dx} + c_x \\ v = \frac{y}{dy} + c_y \end{cases} \Rightarrow \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & c_x \\ 0 & \frac{1}{dy} & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (6-23)$$

其中， $dx$  和  $dy$  是每个像素在  $x$  和  $y$  方向上的物理尺寸， $(c_x, c_y)$  是图像的中心坐标。

从世界坐标系到相机坐标系的变换：

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = t \Rightarrow \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (6-24)$$



由式（6-22）、式（6-23）、式（6-24）可得到世界坐标系与其投影点的计算机图像坐标 $(u,v)$ 的变换关系：

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \mathbf{M}_1 \mathbf{M}_2 \mathbf{x}_w = \mathbf{M} \mathbf{x}_w \quad (6-25)$$

其中 $\mathbf{M}_1$ 为内参数矩阵， $f_x$ 、 $f_y$ 分别是 $u$ 轴和 $v$ 轴的有效焦距， $u_0$ 、 $v_0$ 是光学中心； $\mathbf{M}_2$ 为外参数矩阵， $\mathbf{R}$ 和 $\mathbf{t}$ 是旋转矩阵和平移向量；对于非线性模型的内部参数，还包括径向畸变参数 $k_1$ 、 $k_2$ 和切向畸变参数 $p_1$ 、 $p_2$ 。

## 6.4 基于高速公路图像信息的能见度分析

### 6.4.1 高速公路图像说明

根据题目所提供的视频截图，图片起始点为2016年4月14日6点30分26秒，结束点为2016年4月14日7点39分11秒，每张相邻图间隔42秒。

本文利用MATLAB对输入的有雾图像首先求解暗通道图像，然后使用梯度导向滤波对得到的粗透射率进行细化，最终获得去雾后的清晰图像，如下图所示。

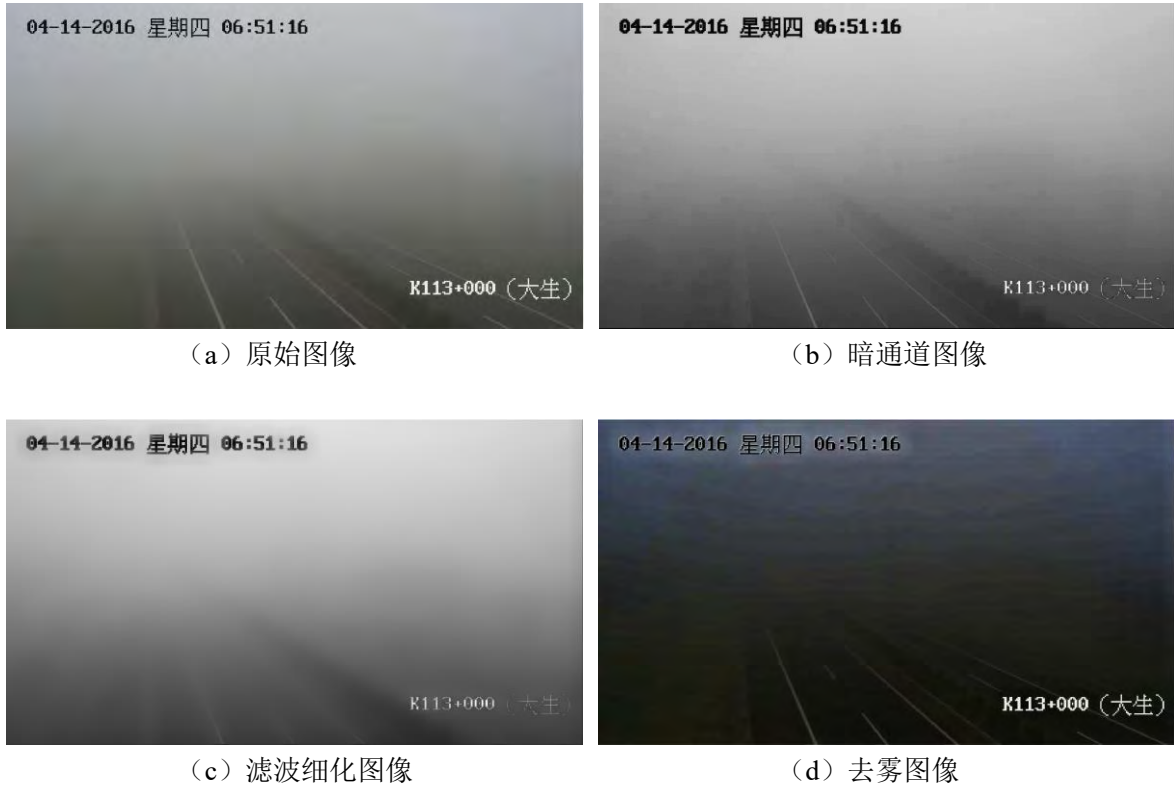


图 6-5 高速公路某时间原始图像及透射率细化后获得的去雾图像

在去雾过程中，同时测算得到图片的透射率，根据透射率以及前文推导所得公式即可测算得到最终的能见度数值。

#### 6.4.2 高速公路图片能见度变化趋势曲线

本文使用 MATLAB 软件，通过各个图像中估计的细化透射率，即可计算出每张图的消光系数，进而估计每张图的能见度。基于每张图像的能见度值，该序列有 100 个样本点（具体值见附件）。本文最终得到高速公路能见度随时间变化趋势如下图所示：

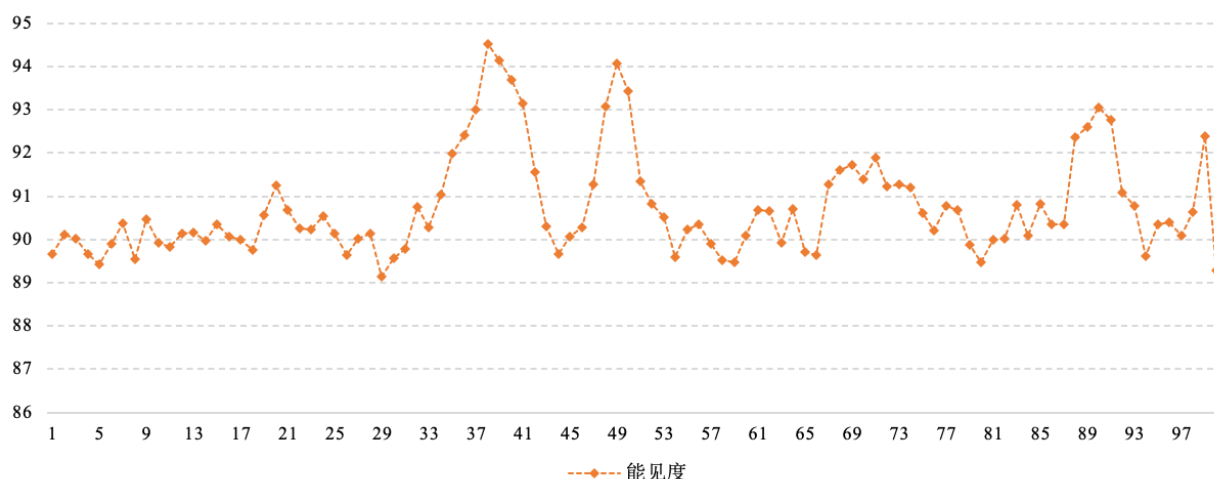


图 6-6 高速公路能见度时间序列图

根据题目所给的高速公路每 100 帧 BMP 图片，我们估计出对应的自 2016 年 4 月 16 日 06:30:26 至 2016 年 4 月 16 日 07:39:11 这个时间段高速公路能见度随时间变化数据。该折线图横坐标序号与时间存在一一对应关系。在这 1 小时多的时间段内，该序列最高峰达 94.51 米，最低谷为 89.28 米，存在一定的小幅波动，整体波动幅度不大，未超过 100 米，能见度较低，平均能见度为 90.44 米。

## 7 问题四模型建立与求解

### 7.1 问题分析

题目要求我们利用问题三得到的能见度随时间变化规律，建立数学模型预测大雾变化趋势（加重或减弱）、何时散去（达到指定的能见度，比如 MOR=150m）。

该问题主要目的是基于问题三所获得的能见度数据与时间变化规律，通过不同的模型来预测大雾情形下能见度的未来变化趋势。为解决该问题，本文构建了基于粒子群优化算法的伯努利模型（PSO-NGBM 模型）和时间序列中的 ARMA 模型，通过对比分析确定大雾未来变化趋势与消散时间。图 7-1 为本问题的思路框架图。

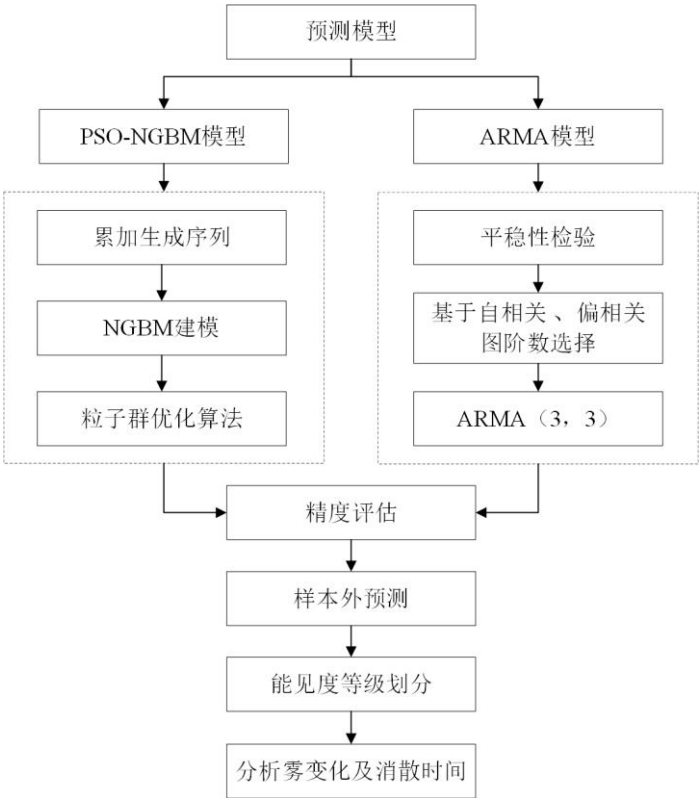


图 7-1 问题四思路框架图

### 7.2 基于粒子群优化算法的非线性伯努利模型预测

#### 7.2.1 非线性灰色伯努利模型

记

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)} = b \left[ x^{(1)} \right]^{\gamma} \tag{7-1}$$

是  $NGBM(1,1)$  模型的白化方程，则求解过程为

令  $w^{(1)}(t) = [x^{(1)}(t)]^{1-\gamma_{m,m}}$ ，则

$$\frac{dw^{(1)}(t)}{dt} = (1-\gamma)[x^{(1)}(t)]^{-\gamma} \frac{dx^{(1)}(t)}{dt} \quad (7-2)$$

对公式 (7-1) 基于公式 (7-2) 变换得

$$\hat{x}^{(1)}(t) = \left\{ e^{\int (1-\gamma)\alpha dt} \left( (1-\gamma) \int \beta e^{\int (1-\gamma)\alpha dt} dt + C \right) \right\}^{\frac{1}{1-\gamma}}, \hat{x}^{(1)}(1) = x^{(1)}(1) \quad (7-3)$$

为求解参数  $\alpha$ 、 $\beta$ ，采用最小二乘法估计参数。灰色预测模型中白化方程转化为灰微分方程，其中  $x^{(0)}(t)$  可表示为：

$$Y = B\hat{\alpha} \quad (7-4)$$

其中

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, B = \begin{bmatrix} -[z^{(1)}(2)]^\gamma & 1 \\ -[z^{(1)}(3)]^\gamma & 1 \\ \vdots & \vdots \\ -[z^{(1)}(n)]^\gamma & 1 \end{bmatrix} \quad (7-5)$$

对其展开求解

$$x^{(0)}(k) = -\alpha z^{(1)}(k) + \beta [z^{(1)}(k)]^\gamma + \varepsilon(k) \quad (7-6)$$

其中  $\varepsilon(k)$  指的是模型中的误差项，基于最小二乘法的残差平方和形式如下：

$$\begin{aligned} S(\alpha) &= \varepsilon^T \varepsilon = [Y - B\hat{\alpha}]^T [Y - B\hat{\alpha}] \\ &= \sum_{k=2}^n \left[ x^{(0)}(k) + \alpha z^{(1)}(k) - \beta (z^{(1)}(k))^\gamma \right]^2 \end{aligned} \quad (7-7)$$

运用拉格朗日乘数法对公式 (7-7) 求导

$$\begin{cases} \frac{\partial S(\alpha)}{\partial \alpha} = -2 \sum_{k=2}^n \left[ x^{(0)}(k) + \alpha z^{(1)}(k) - \beta (z^{(1)}(k))^\gamma \right] z^{(1)}(k) = 0, \\ \frac{\partial S(\alpha)}{\partial \beta} = -2 \sum_{k=2}^n \left[ x^{(0)}(k) + \alpha z^{(1)}(k) - \beta (z^{(1)}(k))^\gamma \right] (z^{(1)}(k))^\gamma = 0, \\ \frac{\partial S(\alpha)}{\partial \gamma} = -2 \beta \sum_{k=2}^n \left[ x^{(0)}(k) + \alpha z^{(1)}(k) - \beta (z^{(1)}(k))^\gamma \right] (z^{(1)}(k))^\gamma \ln(z^{(1)}(k)) = 0. \end{cases} \quad (7-8)$$

因此，参数求解为：

$$\alpha = \frac{\frac{1}{n-1} \sum_{k=2}^n x^{(0)}(k) \cdot \sum_{k=2}^n [z^{(1)}(k)]^\gamma - \sum_{k=2}^n x^{(0)}(k) \cdot [z^{(1)}(k)]^\gamma}{\sum_{k=2}^n [z^{(1)}(k)]^{2\gamma} - \frac{1}{n-1} \left[ \sum_{k=2}^n z^{(1)}(k) \right]^{2\gamma}} \quad (7-9)$$

$$\beta = \frac{1}{n-1} \left[ \sum_{k=2}^n x^{(0)}(k) + \alpha \sum_{k=2}^n [z^{(1)}(k)]^\gamma \right] \quad (7-10)$$

结合粒子群算法寻找最优 $\gamma$ ，由此求解得到参数序列为 $\hat{\delta} = [\alpha, \beta, \gamma]^T$ 。

根据公式(7-1) -式(7-7)求解得时间响应函数：

$$\hat{x}^{(1)}(k+1) = \left[ (x^{(0)}(1))^{(1-\gamma)} - \frac{\beta}{\alpha} e^{-\alpha(1-\gamma)k} + \frac{\beta}{\alpha} \right]^{\frac{1}{(1-\gamma)}}, \hat{x}^{(1)}(1) = x^{(1)}(1) \quad (7-11)$$

其中， $\gamma \neq 1$ ， $k=1,2,\dots$ ，并对 $\hat{x}^{(1)}(k+1)$ 还原，可得原始序列的拟合值

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1), k=2,3,\dots,n \quad (7-12)$$

### 7.2.2 基于 PSO 算法的参数优化

为了得到 NSGM(1,m) 模型参数的全局最优解，并在最大程度上提高模型的预测精度，本文采用粒子群优化 (PSO) 算法求解模型中的非线性参数。PSO 算法是计算智能领域中一种群体智能的优化算法，源于对鸟类捕食行为的研究。鸟类捕食最简单有效的策略是搜寻当前距离食物最近的鸟的周围区域<sup>[13]</sup>。PSO 算法中每一个粒子代表极值优化问题中的一个潜在解，粒子的指标特征分别是位置、速度和适应度值，其中每一粒子对应一个由适应度函数决定的适应值，粒子的速度决定粒子移动方向和距离，速度随自身及其他粒子的移动进行动态调整，从而实现个体在可解空间中的最优。

假设在一个  $D$  维搜索空间中，由  $N$  个粒子组成的种群  $X = (X_1, X_2, \dots, X_N)$ ，其中第  $i$  个粒子表示一个  $D$  维的向量  $X_i = (x_i^1, x_i^2, \dots, x_i^D)$ ，代表第  $i$  个粒子在  $D$  维搜索空间中的位置，根据目标函数即可计算出每个粒子位置  $X_i$  对应的适应度值。第  $i$  个粒子的速度为  $V_i = (V_{i1}, V_{i2}, \dots, V_{iD})^T$ ，其个体极值为  $P_i = (P_{i1}, P_{i2}, \dots, P_{iD})^T$ ，种群的群体极值为  $P_g = (P_{g1}, P_{g2}, \dots, P_{gD})^T$ 。在每次迭代过程中，粒子通过个体极值和群体极值更新自身的速度和位置，即：

$$V_{id}^{k+1} = \omega V_{id}^k + c_1 r_1 (P_{id}^k - X_{id}^k) + c_2 r_2 (P_{gd}^k - X_{id}^k) \quad (7-13)$$

$$X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1} \quad (7-14)$$

其中， $\omega$  为惯性权重， $d=1,2,\dots,D$ ， $i=1,2,\dots,N$ ， $k$  为当前迭代次数， $V_{id}$  为粒子的速度， $c_1$  和  $c_2$  是非负常数的加速度因子（一般  $c_1 = c_2 = 2$ ）， $r_1$  和  $r_2$  是分布于  $[0,1]$  区间的随机数。

本文使用粒子群算法求解如下规划模型：

$$\underset{\gamma}{Min} avg(e(k)) = \frac{1}{n-1} \sum_{k=2}^n \frac{|\hat{x}^{(0)}(k) - x^{(0)}(k)|}{x^{(0)}(k)}, \quad k=2,3,\dots,n \quad (7-15)$$

$$s.t. \begin{cases} \hat{x}^{(1)}(k) = \left[ (x^{(0)}(1))^{(1-\gamma)} - \frac{b}{a} e^{-a(1-\gamma)(k-1)} + \frac{b}{a} \right]^{\frac{1}{1-\gamma}}, \\ \hat{\delta} = (B^T B)^{-1} B^T Y = [a, b]^T, \\ \hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1), k=2,3,\dots,n. \end{cases} \quad (7-16)$$

### 7.2.3 PSO-NGBM (1,1) 模型预测结果

本文构建了 PSO-NGBM (1,1) 模型，通过对问题三得到的 100 个样本序列进行拟合。基于粒子群优化算法，以平均绝对百分比误差为目标函数，寻找最优幂指数，最优参数取 1.21。拟合与预测结果如图所示，最终得到拟合误差 (MAPE) 为 1.21%。由于样本外预测结果变化较大，本文考虑另一种时间序列模型以作比较。

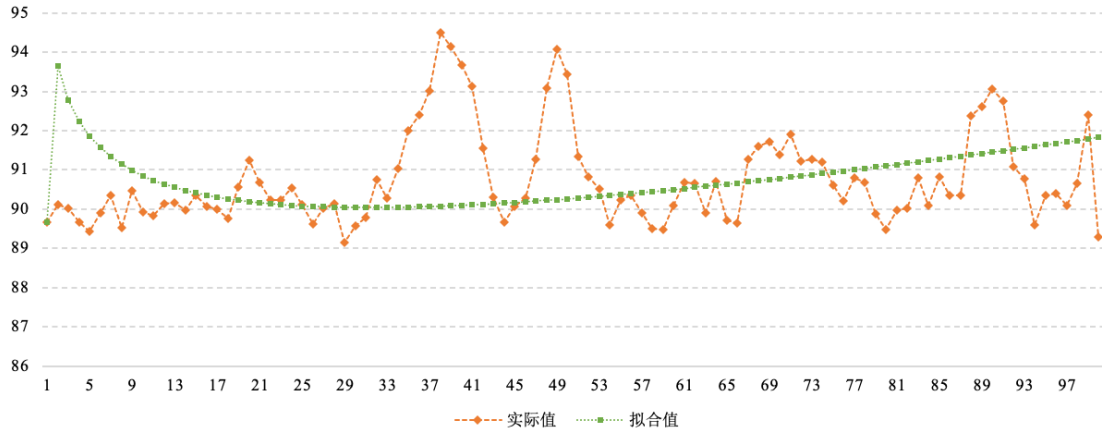


图 7-2 PSO-NGBM 模型拟合过程

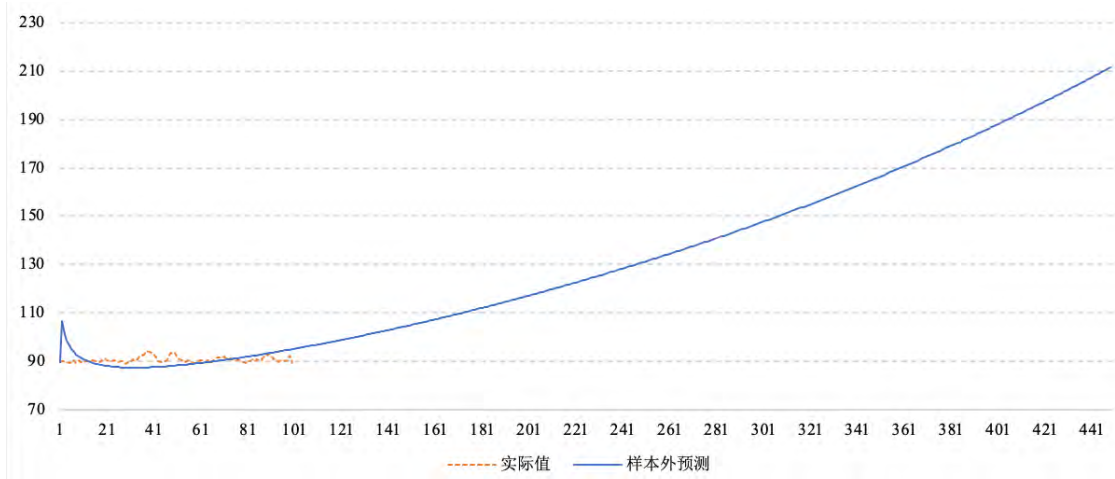


图 7-3 PSO-NGBM 模型预测过程

## 7.2 ARMA (3,3) 模型

### 7.2.1 ARMA 模型

若序列为平稳的非随机序列数据，则该序列包含相关信息。我们可以基于这些相关信息，采用时间序列模型拟合数据，其中自回归模型（AR）、移动平均模型（MA）和自回归移动平均模型（ARMA）最为典型。首先基于一组稳定序列为  $\{x_t, t \in T\}$ ，若给定序列  $\{x_t, t \in T\}$ ，满足以下结构的模型为  $p$  阶自回归模型（简记为  $AR(p)$ ）：

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t \quad (7-17)$$

其中， $\phi_0, \phi_1, \phi_2, \dots, \phi_p$  为  $p+1$  个固定常数，且  $\phi_p \neq 0$ ； $\varepsilon_t$  为零均值的白噪声序列，且与  $x_{t-i}$ ， $i=1,2,\dots$  无关。此外，白噪声序列  $\{\varepsilon_t\}$  满足：

$$\begin{cases} E\varepsilon_t = 0, \\ Var(\varepsilon_t) = \sigma^2 \\ E(\varepsilon_t \varepsilon_s) = 0, \forall t > s \end{cases} \quad (7-18)$$

若给定序列  $\{x_t, t \in T\}$ ，满足以下结构的模型为  $q$  阶移动平均模型（简记为  $MA(q)$ ）

$$x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_t \quad (7-19)$$

其中， $\theta_q \neq 0$ ； $\varepsilon_t$  为零均值的白噪声序列。

由自回归和移动平均两部分共同构成的随机过程称为自回归移动平均过程，记为  $ARMA(p, q)$ ，其中  $p, q$  分别表示自回归和移动平均部分的最大阶数。 $ARMA(p, q)$  的一般表达式是：

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_t \quad (7-20)$$

其中， $\phi_p \neq 0, \theta_q \neq 0$ ； $\varepsilon_t$  为均值为 0 的白噪声序列，且  $\varepsilon_t$  与  $x_{t-i}$ ， $i=1,2,\dots$  无关。应用延迟算子，简记为：

$$\Phi(B)x_t = \Theta(B)\varepsilon_t \quad (7-21)$$

当  $q=0$  时， $ARMA(p, q)$  即为  $AR(p)$  模型；当  $p=0$  时， $ARMA(p, q)$  即为  $MA(q)$  模型。综上三个模型可统称为  $ARMA(p, q)$  模型。

### 7.2.2 ARMA 模型预测结果

首先判断该序列具有平稳性特征，接着绘制自相关图和偏相关图，如下表所示，3 阶以内的自相关系数均不截尾，2 阶以内的偏相关系数不截尾， $p$  值检验均为 0，通过了显著



性检验，故尝试选用  $ARMA(2,3)$  模型。

表7-1  $ARMA(p,q)$  阶数选择

| Autocorrelation                                                                     | Partial Correlation                                                                 | AC   | PAC      | Q-Sta... | Prob   |       |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|------|----------|----------|--------|-------|
|    |    | 1    | 0.793    | 0.793    | 64.096 | 0.000 |
|    |    | 2    | 0.549    | -0.21... | 95.179 | 0.000 |
|    |    | 3    | 0.316    | -0.12... | 105.57 | 0.000 |
|    |    | 4    | 0.086    | -0.17... | 106.36 | 0.000 |
|    |    | 5    | -0.06... | 0.007    | 106.86 | 0.000 |
|    |    | 6    | -0.11... | 0.118    | 108.21 | 0.000 |
|    |    | 7    | -0.07... | 0.086    | 108.89 | 0.000 |
|    |    | 8    | -0.04... | -0.07... | 109.12 | 0.000 |
|    |    | 9    | -0.02... | -0.08... | 109.20 | 0.000 |
|    |    | 1... | 0.030    | 0.135    | 109.31 | 0.000 |
|    |    | 1... | 0.068    | 0.028    | 109.83 | 0.000 |
|    |    | 1... | 0.051    | -0.07... | 110.13 | 0.000 |
|    |    | 1... | 0.029    | -0.03... | 110.23 | 0.000 |
|    |    | 1... | 0.012    | 0.008    | 110.25 | 0.000 |
|    |    | 1... | -0.04... | -0.06... | 110.46 | 0.000 |
|    |    | 1... | -0.07... | 0.036    | 111.21 | 0.000 |
|    |    | 1... | -0.08... | 0.005    | 112.04 | 0.000 |
|    |    | 1... | -0.06... | -0.01... | 112.63 | 0.000 |
|    |    | 1... | -0.07... | -0.06... | 113.31 | 0.000 |
|   |   | 2... | -0.05... | 0.032    | 113.74 | 0.000 |
|  |  | 2... | -0.01... | 0.069    | 113.75 | 0.000 |
|  |  | 2... | -0.04... | -0.21... | 114.06 | 0.000 |
|  |  | 2... | -0.11... | -0.09... | 115.79 | 0.000 |
|  |  | 2... | -0.14... | 0.026    | 118.61 | 0.000 |
|  |  | 2... | -0.15... | 0.085    | 121.65 | 0.000 |
|  |  | 2... | -0.13... | 0.024    | 124.29 | 0.000 |
|  |  | 2... | -0.11... | -0.10... | 126.25 | 0.000 |
|  |  | 2... | -0.04... | 0.057    | 126.49 | 0.000 |
|  |  | 2... | -0.02... | -0.10... | 126.54 | 0.000 |
|  |  | 3... | -0.06... | -0.08... | 127.14 | 0.000 |
|  |  | 3... | -0.10... | -0.05... | 128.76 | 0.000 |
|  |  | 3... | -0.09... | 0.111    | 130.23 | 0.000 |
|  |  | 3... | -0.11... | -0.02... | 132.08 | 0.000 |
|  |  | 3... | -0.11... | -0.05... | 134.24 | 0.000 |
|  |  | 3... | -0.09... | -0.03... | 135.62 | 0.000 |
|  |  | 3... | -0.09... | -0.08... | 137.07 | 0.000 |

通过参数估计与检验结果可知，Adj  $R^2$  为 0.6066，通过调整各阶数，对模型做进一步的修正。本文继续尝试  $ARMA(3,1)$ ， $ARMA(3,2)$ ， $ARMA(3,3)$ ， $ARMA(2,1)$ ， $ARMA(2,2)$ ， $ARMA(2,3)$  等模型对时间序列进行模拟，检验结果如下表所示。当  $p=3$ ， $q=3$  时，AIC 和 SC 值相对较大，Adj  $R^2$  的值也相对较大，最终确定该序列的拟合模型为  $ARMA(3,3)$ 。

表 7-2 各模型检验结果

|             | AIC    | SC     | $R^2$  | Adj $R^2$ |
|-------------|--------|--------|--------|-----------|
| $ARMA(3,1)$ | 2.2723 | 2.3792 | 0.6243 | 0.6243    |
| $ARMA(3,2)$ | 2.2503 | 2.3838 | 0.6400 | 0.6242    |
| $ARMA(3,3)$ | 2.2391 | 2.3993 | 0.6514 | 0.6320    |
| $ARMA(2,1)$ | 2.2514 | 2.3310 | 0.6218 | 0.6137    |
| $ARMA(2,2)$ | 2.2717 | 2.3779 | 0.6219 | 0.6097    |
| $ARMA(2,3)$ | 2.2894 | 2.4221 | 0.6230 | 0.6066    |

综上所述，ARMA（3,3）模型各参数绝大部分通过了 t 检验，检验均比前面几个模型效果好，因此最终确定模型的表达式为：

$$x_t = 1.122AR(1) - 0.924AR(2) + 0.802AR(3) + 0.036MA(1) + 0.876MA(2) + 0.277MA(3)$$

表7-3 ARMA（3,3）模型拟合结果

| Variable           | Coefficient | Std. Error            | t-Statistic | Prob     |
|--------------------|-------------|-----------------------|-------------|----------|
| AR(1)              | 1.122202    | 0.085948              | 13.05671    | 0.0000   |
| AR(2)              | -0.923579   | 0.116979              | -7.895223   | 0.0000   |
| AR(3)              | 0.802004    | 0.074230              | 10.80436    | 0.0000   |
| MA(1)              | 0.036408    | 0.131536              | 0.276793    | 0.7826   |
| MA(2)              | 0.875634    | 0.041892              | 20.90199    | 0.0000   |
| MA(3)              | 0.276763    | 0.116970              | 2.366101    | 0.0201   |
| R-squared          | 0.651381    | Mean dependent var    |             | 90.78253 |
| Adjusted R-squared | 0.632013    | S.D. dependent var    |             | 1.185556 |
| S.E. of regression | 0.719181    | Akaike info criterion |             | 2.239053 |
| Sum squared resid  | 46.54987    | Schwarz criterion     |             | 2.399325 |
| Log likelihood     | -101.4745   | Hannan-Quinn criter.  |             | 2.303838 |
| Durbin-Watson stat | 1.923936    |                       |             |          |

此外，本文还对残差序列进行了白噪声检验，该检验通过，建立的 ARMA（3,3）模型合理。如下图所示，实际值使用红线标记，ARMA（3,3）拟合值使用绿色标记，两者之差为残差，左纵坐标为残差刻度，在[-1,1]范围内波动。由残差结果 MAPE 为 0.63%可知，ARMA（3,3）模型为最佳模型。

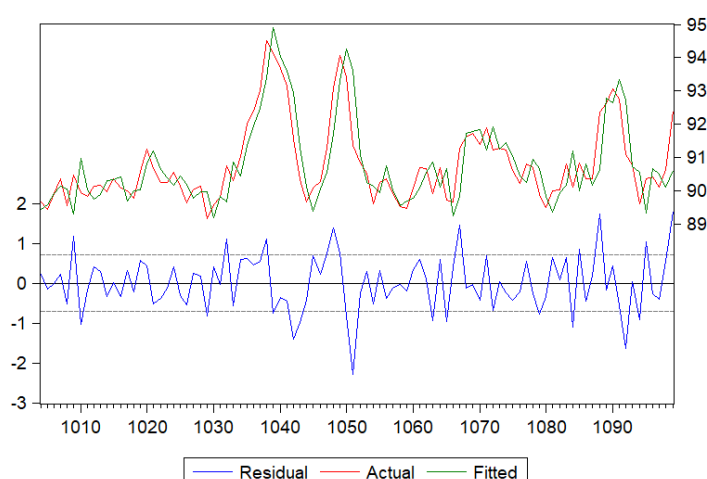


图7-4 ARMA（3,3）模型拟合过程

本文对序列进行样本外动态预测，得到未来能见度值的预测趋势，如下图所示。从第

101 个样本点对应的时间为 2016 年 4 月 14 日 07:39:11，其能见度随时间呈一定的向上趋势。本文预估在第 2158 个样本点对应的时间为 2016 年 4 月 15 日 07:39:47，MOR 能见度达到 200 米。往后样本点仍呈现缓慢上升趋势。

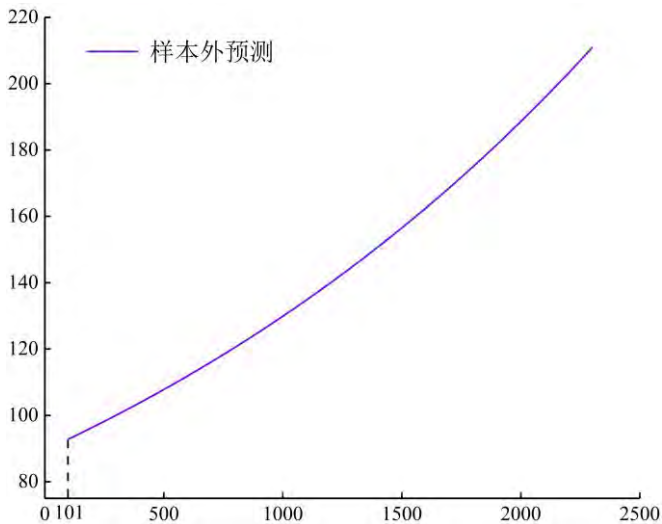


图7-5 ARMA (3,3) 模型预测过程

7.3 能见度划分依据

根据题目三最后绘制的能见度随时间变化趋势图，可以发现能见度呈现出缓慢上升趋势，即未来高速公路的能见度将不断增加，大雾将会逐渐散去。为具体判定大雾何时将会散去，本节将基于相关资料对高速公路能见度进行划分。

基于高速公路能见度等级划分和低能见度下高速公路应急管理措施，本文对能见度进行了多模式划分<sup>[14-15]</sup>，具体如表 7-1 所示。

表 7-1 高速公路能见度等级划分表

| 能见度等级 | 能见度范围     | 高速公路应急管理措施                     |
|-------|-----------|--------------------------------|
| 0     | 500<V     | 雾消散——无需采取措施                    |
| 1     | 200<V<500 | 时速不超过 80km，车距 150m 以上          |
| 2     | 100<V<200 | 时速不超过 40km，车距 100m 以上，局部封闭高速公路 |
| 3     | 50<V<100  | 时速不超过 40km，车距 100m 以上，局部封闭高速公路 |
| 4     | V<50      | 驶离雾区，最高时速不超过 20km              |

基于问题四要求预测大雾何时散去，本文结合文献认为序列超过 200m 阈值的时间点即为雾散去的时间点。

7.4 各模型预测结果

根据问题三中估计并绘制的能见度随时间变化曲线图，本章通过构建 PSO-NGBM 和 ARMA 两种模型，预测大雾变化程度和散去的时间。鉴于 ARMA 模型在拟合结果上具有较好的效果且通过了各项检验，本文认为使用 ARMA (3,3) 模型预测该序列未来趋势较

为合理。

根据高速公路能见度等级划分，200 米阈值为关键的衡量雾散去的能见度值。根据上文 ARMA 模型样本外预测结果，我们预估 2016 年 4 月 15 日 07:39:47 以后，MOR 能见度将达到 200 米，大雾散去。

综上，本文回答了第四个问题，建立 ARMA (3,3) 模型预测大雾变化趋势，并将其与 PSO-NGBM (1,1) 模型进行对比，预测出能见度趋势在缓慢上升，大雾在缓慢散去，将于第二天早上 7 点左右散去（具体时间为 2016 年 4 月 15 日 07:39:47）。

## 8 模型的总结与评价

### 8.1 模型的结论

本文紧密围绕能见度与影响因素关系及预测问题，构建了分位数非线性回归模型、基于卷积神经网络模型的迁移学习方法、暗通道先验算法和时间序列等多种模型，由点及面，针对问题中的机场视频数据、高速公路图片等数据分析了机场能见度与地面气象因素（温度、湿度和风速等）间的关系并提供了能见度非线性关系式；其次，通过构建深度学习模型，对机场能见度进行精度评估；接着，基于不依赖能见度仪的观测数据建立能见度估计算法并绘制高速公路能见度时间序列曲线；最后，进一步分析了高速公路能见度随时间变化规律并预测大雾变化趋势及散去时间。

对于问题一，我们采用了 Pearson 相关性矩阵和热力图剔除了不合理的变量，选择分位数非线性回归等方法，刻画了能见度与地面气象观测因素的非线性关系。本文基于能见度（RVR）的非线性分位数回归最优分位点取 0.5，基于能见度（MOR）的非线性分位数回归最优分位点取 0.6，其最优结果下的调整  $R^2$  分别为 0.6422 和 0.6974，各变量  $p$  值均小于 0.05，通过了显著性检验。在统计评价指标最优和符合经济意义的多重目标下，我们最终导出了机场能见度的具体非线性关系式。

对于问题二，鉴于传统图像识别方法存在准确率低、泛化能力差等问题，我们采取一种基于卷积神经网络模型的迁移学习方法，通过预训练模型确定各参数，将图像与数据随机以 7:3 的比例划分为训练集与测试集，冻结前层网络参数训练并微调深度学习模型，迭代 100 次以后，训练集、测试集准确率分别达到了 80.48% 和 79.63%。而传统图像识别方法上，我们识别了图像 11 维信息特征，建立了 BP 神经网络模型并评估了精度。结果表明基于卷积神经网络模型的迁移学习方法确实具有更高的效率以及识别精度。

对于问题三，我们推导了能见度与透射率、消光系数间的数理关系，并使用基于梯度导向滤波技术的暗通道先验算法对题目提供的高速公路图片求解暗通道图像、采用梯度导向滤波对获得的粗透射率进行细化、求解全局大气光照等获得去雾后的清晰图片，同时获得图像的透射率数据，最后基于透射率数据计算图像消光系数，进而测算得到图像的能见度数据，绘制了能见度随时间变化曲线。结果显示能见度序列在 89.28 米到 94.51 米范围内小幅波动，能见度较低。

对于问题四，根据问题三估计的能见度数据，我们建立了基于粒子群优化算法的伯努利模型（PSO-NGBM（1,1）模型）和时间序列中的 ARMA（3,3）模型，以预测大雾未来变化趋势与消散时间。结果表明 PSO-NGBM（1,1）模型与 ARMA（3,3）的平均绝对百分比误差（MAPE）分别 1.21% 和 0.63%，最终本文基于 ARMA（3,3）模型，进行样本外预测，我们预估 2016 年 4 月 15 日 7 点以后，MOR 能见度将达到 200 米，雾将散去。

### 8.2 模型的优点

（1）本文采用多种模型和思路对大雾情况下机场和高速公路的能见度进行影响因素分析、评估、计算、预测，环环相扣，较好地反映了能见度的演化规律和变化趋势。

(2) 问题一中采用的分数位非线性回归模型全面的描述了气象因素影响下的能见度条件分布的全貌，温度、湿度、风向等与能见度的非线性关系得到了更好地刻画，且基于分位数回归的预测结果更加稳健。

(3) 鉴于有限的能见度数据量，本文将最前沿的基于迁移学习和卷积神经网络的深度学习模型应用于该实际问题中，评估精度和实际效果得到了验证。

(4) 能见度计算逻辑清晰。从图像采集-暗通道算法-梯度导向滤波改进-透射率获取-消光系数计算-相机标定法-能见度计算，各环节步骤紧密联系，成功把握能见度的时间演化规律。

(5) 根据问题三计算的能见度数据，对比分析 ARMA 模型和基于粒子群算法的非线性灰色伯努利模型，预测大雾情况下能见度的未来变化趋势。

### 8.3 模型的不足之处

(1) 时间有限，训练成本较高。深度学习模型训练和图像识别过程需要较长时间且依赖于强大的计算机硬件资源，这可能导致模型评估或计算结果不够理想。

(2) 问题四中基于高速公路视频的能见度数据样本量较小，且由问题一结论，能见度与气象因素密切相关，相关因素数据缺失，因此本文采用的模型可能并不能精确的确定大雾消散时间。

## 参考文献

- [1]刘敏,赵普洋.气象光学视程(MOR)在民用航空地面气象观测中的应用[J].气象水文海洋仪器,2012,29(01):78-80.
- [2]宋明,韩素芹,张敏,姚青,朱彬.天津大气能见度与相对湿度和PM<sub>10</sub>及PM<sub>2.5</sub>的关系[J].气象与环境学报,2013,29(02):34-41.
- [3]Nessler R, Weingartner E, Baltensperger U. Effect of humidity on aerosol light absorption and its implications for extinction and the single scattering albedo illustrated for a site in the lower free troposphere[J]. Journal of Aerosol Science, 2005, 36(8): 958-972.
- [4]郭智亮,高聪晖,谢文锋,陈淑敏,黎伟标. 2005—2017 年白云机场能见度变化特征及其与影响因子关系研究[J]. 气象科技进展,2019,9(06):40-43.
- [5]许可. 卷积神经网络在图像识别上的应用的研究[D].浙江大学,2012.
- [6]庄福振,罗平,何清,史忠植.迁移学习研究进展[J].软件学报,2015,26(01):26-39.
- [7]龙明盛. 迁移学习问题与方法研究[D].清华大学,2014.
- [8]倪爱伟. 基于双目立体视觉的三维重建技术研究[D].南京航空航天大学,2009.
- [9]Hautière N, Tarel J P, Lavenant J, et al. Automatic fog detection and estimation of visibility distance through use of an onboard camera[J]. Machine Vision and Applications, 2006, 17(1): 8-20.
- [10]He K, Sun J, Tang X. Single image haze removal using dark channel prior[C]// Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009:1956-1963.
- [11]Narasimhan S G, Nayar S K. Contrast restoration of weather degraded images [J].IEEE Trans. Pattern Anal. Mach. Intell. 2003, 25 (6): 713-724.
- [12]杨明,王海晖,陈君,万妮.双目标定系统精度提高的方法[J].武汉工程大学学报,2012,34(01):69-73.
- [13]Shi Y, Eberhart R C. Parameter selection in particle swarm optimization[C]//International conference on evolutionary programming. Springer, Berlin, Heidelberg, 1998: 591-600.
- [14]QX/T 47-2007.地面气象观测规范[S],北京:气象出版社,2007.
- [15]中华人民共和国公安部.关于加强低能见度气象条件下高速公路交通管理的通告[Z].1998-12-31.
- [16]S. K. Nayar, S. G. Narasimhan, Vision in bad weather, ICCV'99
- [17]R. T. Tan, Visibility in bad weather from a single image, CVPR, 2008
- [18]N. Hautiere, J-P Tarel, J Lavenant D. Aubert, Automatic fog detection and estimation of visibility distance

through use of onboard camera, Machine Vision and Application, 2006, 17 (1) : 8-20

[19]C. Sakaridis, D. Dai, L. V. Gool, Semantic foggy scene understanding with synthetic data, International J. Computer Vision, 2018, 3



## 附录 A

表 A-1 基于 RVR 的分位数非线性回归的分位点选择

|    | Quantile | Coefficient | Std. Error | t-Statistic | Prob.  |
|----|----------|-------------|------------|-------------|--------|
| C  | 0.050    | 193548.1    | 29812.32   | 6.492219    | 0.0000 |
|    | 0.100    | 203861.2    | 9885.399   | 20.62246    | 0.0000 |
|    | 0.150    | 204519.2    | 14201.74   | 14.40100    | 0.0000 |
|    | 0.200    | 188683.3    | 15685.92   | 12.02883    | 0.0000 |
|    | 0.250    | 164647.1    | 12607.85   | 13.05909    | 0.0000 |
|    | 0.300    | 166625.1    | 12734.97   | 13.08406    | 0.0000 |
|    | 0.350    | 164837.0    | 14597.55   | 11.29211    | 0.0000 |
|    | 0.400    | 165618.8    | 18293.67   | 9.053335    | 0.0000 |
|    | 0.450    | 158372.3    | 21442.86   | 7.385783    | 0.0000 |
|    | 0.500    | 156391.9    | 21129.97   | 7.401429    | 0.0000 |
|    | 0.550    | 156769.9    | 20375.29   | 7.694122    | 0.0000 |
|    | 0.600    | 157669.8    | 23407.92   | 6.735745    | 0.0000 |
|    | 0.650    | 144684.8    | 24312.11   | 5.951140    | 0.0000 |
|    | 0.700    | 138870.3    | 22236.43   | 6.245173    | 0.0000 |
|    | 0.750    | 165845.0    | 18211.77   | 9.106477    | 0.0000 |
|    | 0.800    | 201160.0    | 23857.90   | 8.431588    | 0.0000 |
|    | 0.850    | 243219.8    | 31760.28   | 7.657987    | 0.0000 |
|    | 0.900    | 270848.5    | 26187.36   | 10.34272    | 0.0000 |
|    | 0.950    | 267822.1    | 23600.66   | 11.34808    | 0.0000 |
| RH | 0.050    | -48.39342   | 4.864645   | -9.947986   | 0.0000 |
|    | 0.100    | -48.59468   | 1.776446   | -27.35501   | 0.0000 |
|    | 0.150    | -48.72114   | 2.515519   | -19.36822   | 0.0000 |
|    | 0.200    | -48.58809   | 2.641628   | -18.39324   | 0.0000 |
|    | 0.250    | -49.34258   | 3.242469   | -15.21760   | 0.0000 |
|    | 0.300    | -42.60300   | 3.901226   | -10.92041   | 0.0000 |
|    | 0.350    | -40.63155   | 3.617886   | -11.23074   | 0.0000 |
|    | 0.400    | -38.99133   | 3.618180   | -10.77651   | 0.0000 |
|    | 0.450    | -39.13306   | 3.594169   | -10.88793   | 0.0000 |
|    | 0.500    | -37.60279   | 3.363065   | -11.18111   | 0.0000 |
|    | 0.550    | -36.83476   | 3.198995   | -11.51448   | 0.0000 |
|    | 0.600    | -35.20593   | 3.295463   | -10.68315   | 0.0000 |
|    | 0.650    | -36.63691   | 3.241216   | -11.30345   | 0.0000 |
|    | 0.700    | -36.67933   | 2.877244   | -12.74808   | 0.0000 |
|    | 0.750    | -34.69501   | 2.171384   | -15.97830   | 0.0000 |
|    | 0.800    | -32.10700   | 2.438056   | -13.16910   | 0.0000 |
|    | 0.850    | -30.44857   | 2.376523   | -12.81223   | 0.0000 |
|    | 0.900    | -30.44906   | 2.036822   | -14.94929   | 0.0000 |

|     |       |           |          |           |        |
|-----|-------|-----------|----------|-----------|--------|
|     | 0.950 | -28.57439 | 2.223527 | -12.85093 | 0.0000 |
| QFE | 0.050 | -189.5377 | 29.96859 | -6.324546 | 0.0000 |
|     | 0.100 | -199.8610 | 9.737657 | -20.52455 | 0.0000 |
|     | 0.150 | -200.3267 | 14.06818 | -14.23970 | 0.0000 |
|     | 0.200 | -184.6730 | 15.54968 | -11.87632 | 0.0000 |
|     | 0.250 | -161.0738 | 12.49079 | -12.89540 | 0.0000 |
|     | 0.300 | -163.0398 | 12.57975 | -12.96050 | 0.0000 |
|     | 0.350 | -161.2581 | 14.49123 | -11.12798 | 0.0000 |
|     | 0.400 | -162.0784 | 18.16549 | -8.922327 | 0.0000 |
|     | 0.450 | -155.0479 | 21.25248 | -7.295520 | 0.0000 |
|     | 0.500 | -153.0675 | 20.89375 | -7.325993 | 0.0000 |
|     | 0.550 | -153.2046 | 20.13268 | -7.609745 | 0.0000 |
|     | 0.600 | -153.8578 | 23.17104 | -6.640091 | 0.0000 |
|     | 0.650 | -140.7523 | 24.07156 | -5.847244 | 0.0000 |
|     | 0.700 | -134.8392 | 21.91933 | -6.151608 | 0.0000 |
|     | 0.750 | -160.9336 | 17.86816 | -9.006724 | 0.0000 |
|     | 0.800 | -195.2186 | 23.48223 | -8.313463 | 0.0000 |
|     | 0.850 | -236.4590 | 31.29450 | -7.555927 | 0.0000 |
|     | 0.900 | -263.5963 | 25.97643 | -10.14752 | 0.0000 |
|     | 0.950 | -260.7496 | 23.32968 | -11.17673 | 0.0000 |
| DEW | 0.050 | 504.5570  | 67.15558 | 7.513255  | 0.0000 |
|     | 0.100 | 537.2931  | 10.72604 | 50.09237  | 0.0000 |
|     | 0.150 | 525.3391  | 12.20402 | 43.04640  | 0.0000 |
|     | 0.200 | 525.1297  | 13.26402 | 39.59054  | 0.0000 |
|     | 0.250 | 540.3388  | 23.58909 | 22.90630  | 0.0000 |
|     | 0.300 | 509.6654  | 26.22508 | 19.43427  | 0.0000 |
|     | 0.350 | 500.1168  | 20.76597 | 24.08348  | 0.0000 |
|     | 0.400 | 504.3642  | 21.23814 | 23.74804  | 0.0000 |
|     | 0.450 | 522.9201  | 19.59997 | 26.67964  | 0.0000 |
|     | 0.500 | 514.2836  | 20.25778 | 25.38696  | 0.0000 |
|     | 0.550 | 498.2389  | 20.81557 | 23.93588  | 0.0000 |
|     | 0.600 | 475.1253  | 19.63972 | 24.19206  | 0.0000 |
|     | 0.650 | 463.2715  | 18.51602 | 25.02004  | 0.0000 |
|     | 0.700 | 452.5985  | 22.29343 | 20.30188  | 0.0000 |
|     | 0.750 | 412.7595  | 23.29660 | 17.71758  | 0.0000 |
|     | 0.800 | 367.1070  | 24.20185 | 15.16855  | 0.0000 |
|     | 0.850 | 350.3781  | 23.58204 | 14.85783  | 0.0000 |
|     | 0.900 | 355.1974  | 23.41638 | 15.16875  | 0.0000 |
|     | 0.950 | 348.1980  | 27.28662 | 12.76076  | 0.0000 |
| LIG | 0.050 | 9.566810  | 3.155715 | 3.031582  | 0.0025 |
|     | 0.100 | 11.06762  | 0.585104 | 18.91565  | 0.0000 |
|     | 0.150 | 11.11002  | 0.788737 | 14.08585  | 0.0000 |
|     | 0.200 | 11.60080  | 0.894822 | 12.96436  | 0.0000 |
|     | 0.250 | 12.08093  | 1.062464 | 11.37067  | 0.0000 |

|     |       |           |          |           |        |
|-----|-------|-----------|----------|-----------|--------|
|     | 0.300 | 13.54983  | 1.114610 | 12.15657  | 0.0000 |
|     | 0.350 | 13.78608  | 1.123700 | 12.26847  | 0.0000 |
|     | 0.400 | 14.67146  | 1.267391 | 11.57611  | 0.0000 |
|     | 0.450 | 15.52574  | 1.121664 | 13.84170  | 0.0000 |
|     | 0.500 | 15.57985  | 0.989978 | 15.73757  | 0.0000 |
|     | 0.550 | 14.86339  | 1.073182 | 13.84983  | 0.0000 |
|     | 0.600 | 14.06116  | 1.140869 | 12.32496  | 0.0000 |
|     | 0.650 | 12.87692  | 1.133363 | 11.36169  | 0.0000 |
|     | 0.700 | 12.19372  | 1.372024 | 8.887398  | 0.0000 |
|     | 0.750 | 9.769973  | 1.737960 | 5.621517  | 0.0000 |
|     | 0.800 | 7.377458  | 1.633925 | 4.515176  | 0.0000 |
|     | 0.850 | 6.255931  | 1.086519 | 5.757773  | 0.0000 |
|     | 0.900 | 6.326231  | 1.151917 | 5.491917  | 0.0000 |
|     | 0.950 | 6.710103  | 0.864827 | 7.758899  | 0.0000 |
| WS2 | 0.050 | -32.86525 | 34.14557 | -0.962504 | 0.3360 |
|     | 0.100 | -61.18095 | 19.71189 | -3.103759 | 0.0019 |
|     | 0.150 | -81.65115 | 22.72467 | -3.593062 | 0.0003 |
|     | 0.200 | -85.59365 | 22.91592 | -3.735117 | 0.0002 |
|     | 0.250 | -85.28663 | 20.14360 | -4.233932 | 0.0000 |
|     | 0.300 | -70.62526 | 18.60959 | -3.795099 | 0.0002 |
|     | 0.350 | -69.09370 | 17.92234 | -3.855172 | 0.0001 |
|     | 0.400 | -79.18011 | 16.93493 | -4.675550 | 0.0000 |
|     | 0.450 | -86.46185 | 13.91246 | -6.214708 | 0.0000 |
|     | 0.500 | -87.27139 | 14.30191 | -6.102078 | 0.0000 |
|     | 0.550 | -97.02011 | 14.91932 | -6.502986 | 0.0000 |
|     | 0.600 | -98.70446 | 15.09754 | -6.537785 | 0.0000 |
|     | 0.650 | -104.8542 | 16.20991 | -6.468528 | 0.0000 |
|     | 0.700 | -105.2220 | 20.61629 | -5.103826 | 0.0000 |
|     | 0.750 | -90.68129 | 20.23199 | -4.482074 | 0.0000 |
|     | 0.800 | -84.79819 | 21.55244 | -3.934505 | 0.0001 |
|     | 0.850 | -69.54675 | 26.19824 | -2.654635 | 0.0080 |
|     | 0.900 | -81.28823 | 32.76545 | -2.480913 | 0.0132 |
|     | 0.950 | -74.88063 | 29.00298 | -2.581826 | 0.0099 |
| WD2 | 0.050 | -1.568656 | 0.415986 | -3.770931 | 0.0002 |
|     | 0.100 | -1.731872 | 0.163812 | -10.57229 | 0.0000 |
|     | 0.150 | -1.602031 | 0.217453 | -7.367248 | 0.0000 |
|     | 0.200 | -1.734511 | 0.284267 | -6.101687 | 0.0000 |
|     | 0.250 | -1.544778 | 0.623090 | -2.479223 | 0.0133 |
|     | 0.300 | -2.861541 | 0.727429 | -3.933774 | 0.0001 |
|     | 0.350 | -3.189706 | 0.581346 | -5.486763 | 0.0000 |
|     | 0.400 | -3.622429 | 0.551886 | -6.563724 | 0.0000 |
|     | 0.450 | -3.718846 | 0.510319 | -7.287298 | 0.0000 |
|     | 0.500 | -3.880644 | 0.520928 | -7.449478 | 0.0000 |
|     | 0.550 | -4.271790 | 0.503472 | -8.484664 | 0.0000 |

|       |           |          |           |        |
|-------|-----------|----------|-----------|--------|
| 0.600 | -4.704472 | 0.382433 | -12.30143 | 0.0000 |
| 0.650 | -4.813401 | 0.302111 | -15.93256 | 0.0000 |
| 0.700 | -4.958193 | 0.397695 | -12.46732 | 0.0000 |
| 0.750 | -5.580387 | 0.432469 | -12.90354 | 0.0000 |
| 0.800 | -6.175700 | 0.395905 | -15.59896 | 0.0000 |
| 0.850 | -6.220222 | 0.270634 | -22.98388 | 0.0000 |
| 0.900 | -6.124067 | 0.324289 | -18.88460 | 0.0000 |
| 0.950 | -5.434476 | 0.284635 | -19.09280 | 0.0000 |

---

表 A-2 基于 RVR 的分位数非线性回归的分位点选择

|       | Quantile | Coefficient | Std. Error | t-Statistic | Prob.  |
|-------|----------|-------------|------------|-------------|--------|
| C     | 0.050    | 520930.1    | 36376.56   | 14.32049    | 0.0000 |
|       | 0.100    | 508147.8    | 46244.24   | 10.98835    | 0.0000 |
|       | 0.150    | 540951.7    | 61891.71   | 8.740294    | 0.0000 |
|       | 0.200    | 593225.1    | 107487.8   | 5.518997    | 0.0000 |
|       | 0.250    | 799883.1    | 115114.1   | 6.948608    | 0.0000 |
|       | 0.300    | 880659.9    | 100294.6   | 8.780730    | 0.0000 |
|       | 0.350    | 954954.2    | 79695.05   | 11.98260    | 0.0000 |
|       | 0.400    | 966831.4    | 63698.84   | 15.17816    | 0.0000 |
|       | 0.450    | 964501.4    | 55595.03   | 17.34870    | 0.0000 |
|       | 0.500    | 971152.2    | 54561.99   | 17.79906    | 0.0000 |
|       | 0.550    | 996132.6    | 40533.07   | 24.57580    | 0.0000 |
|       | 0.600    | 963911.5    | 38808.63   | 24.83756    | 0.0000 |
|       | 0.650    | 995931.6    | 42966.10   | 23.17948    | 0.0000 |
|       | 0.700    | 965276.7    | 40199.52   | 24.01215    | 0.0000 |
|       | 0.750    | 957723.6    | 30715.47   | 31.18050    | 0.0000 |
|       | 0.800    | 1004882.    | 26373.32   | 38.10221    | 0.0000 |
|       | 0.850    | 1042394.    | 31456.17   | 33.13797    | 0.0000 |
|       | 0.900    | 1096381.    | 32918.97   | 33.30546    | 0.0000 |
|       | 0.950    | 1171360.    | 55677.51   | 21.03829    | 0.0000 |
| EXPRH | 0.050    | -1.39E-40   | 7.15E-42   | -19.48865   | 0.0000 |
|       | 0.100    | -1.36E-40   | 8.12E-42   | -16.76782   | 0.0000 |
|       | 0.150    | -1.32E-40   | 9.31E-42   | -14.22442   | 0.0000 |
|       | 0.200    | -1.21E-40   | 1.74E-41   | -6.961985   | 0.0000 |
|       | 0.250    | -8.62E-41   | 1.87E-41   | -4.616928   | 0.0000 |
|       | 0.300    | -6.89E-41   | 1.85E-41   | -3.719593   | 0.0002 |
|       | 0.350    | -5.15E-41   | 1.56E-41   | -3.302940   | 0.0010 |
|       | 0.400    | -4.99E-41   | 1.22E-41   | -4.090749   | 0.0000 |
|       | 0.450    | -4.90E-41   | 1.18E-41   | -4.152108   | 0.0000 |
|       | 0.500    | -3.73E-41   | 1.40E-41   | -2.663894   | 0.0078 |
|       | 0.550    | -2.91E-41   | 9.31E-42   | -3.123209   | 0.0018 |
|       | 0.600    | -3.49E-41   | 7.61E-42   | -4.578582   | 0.0000 |
|       | 0.650    | -3.54E-41   | 7.31E-42   | -4.839626   | 0.0000 |
|       | 0.700    | -4.11E-41   | 7.29E-42   | -5.628121   | 0.0000 |
|       | 0.750    | -4.15E-41   | 6.33E-42   | -6.549279   | 0.0000 |
|       | 0.800    | -4.50E-41   | 4.16E-42   | -10.82362   | 0.0000 |
|       | 0.850    | -4.20E-41   | 4.25E-42   | -9.897748   | 0.0000 |
|       | 0.900    | -3.60E-41   | 4.49E-42   | -8.008377   | 0.0000 |
|       | 0.950    | -2.33E-41   | 6.39E-42   | -3.646782   | 0.0003 |
| QFE   | 0.050    | -495.4302   | 36.17742   | -13.69446   | 0.0000 |
|       | 0.100    | -482.4317   | 45.85075   | -10.52178   | 0.0000 |

|       |       |           |          |           |        |
|-------|-------|-----------|----------|-----------|--------|
|       | 0.150 | -514.4199 | 61.09585 | -8.419884 | 0.0000 |
|       | 0.200 | -566.6002 | 106.6897 | -5.310731 | 0.0000 |
|       | 0.250 | -772.1467 | 114.1734 | -6.762933 | 0.0000 |
|       | 0.300 | -853.1529 | 99.47878 | -8.576230 | 0.0000 |
|       | 0.350 | -927.4480 | 79.12317 | -11.72157 | 0.0000 |
|       | 0.400 | -939.8658 | 63.39022 | -14.82667 | 0.0000 |
|       | 0.450 | -938.0157 | 55.60181 | -16.87024 | 0.0000 |
|       | 0.500 | -945.0184 | 54.26552 | -17.41471 | 0.0000 |
|       | 0.550 | -969.8496 | 39.76859 | -24.38733 | 0.0000 |
|       | 0.600 | -937.3770 | 37.47112 | -25.01599 | 0.0000 |
|       | 0.650 | -967.9303 | 41.39653 | -23.38192 | 0.0000 |
|       | 0.700 | -935.2831 | 39.07837 | -23.93352 | 0.0000 |
|       | 0.750 | -925.9491 | 29.82256 | -31.04861 | 0.0000 |
|       | 0.800 | -971.7557 | 26.14843 | -37.16306 | 0.0000 |
|       | 0.850 | -1009.415 | 31.33432 | -32.21435 | 0.0000 |
|       | 0.900 | -1063.453 | 32.77271 | -32.44934 | 0.0000 |
|       | 0.950 | -1138.190 | 55.60003 | -20.47103 | 0.0000 |
| LNDEW | 0.050 | -6124.616 | 275.8167 | -22.20538 | 0.0000 |
|       | 0.100 | -6299.172 | 291.7025 | -21.59451 | 0.0000 |
|       | 0.150 | -6437.710 | 295.5767 | -21.78016 | 0.0000 |
|       | 0.200 | -6190.978 | 402.3314 | -15.38776 | 0.0000 |
|       | 0.250 | -5482.782 | 404.1019 | -13.56782 | 0.0000 |
|       | 0.300 | -4916.800 | 413.7698 | -11.88293 | 0.0000 |
|       | 0.350 | -4399.732 | 456.5004 | -9.637957 | 0.0000 |
|       | 0.400 | -3952.054 | 500.3042 | -7.899302 | 0.0000 |
|       | 0.450 | -3659.645 | 517.5823 | -7.070653 | 0.0000 |
|       | 0.500 | -3378.171 | 490.8970 | -6.881630 | 0.0000 |
|       | 0.550 | -3128.795 | 490.4287 | -6.379714 | 0.0000 |
|       | 0.600 | -3333.424 | 588.6657 | -5.662677 | 0.0000 |
|       | 0.650 | -3604.483 | 726.2874 | -4.962888 | 0.0000 |
|       | 0.700 | -4536.076 | 886.4991 | -5.116842 | 0.0000 |
|       | 0.750 | -5265.924 | 728.1734 | -7.231689 | 0.0000 |
|       | 0.800 | -5410.651 | 317.7494 | -17.02805 | 0.0000 |
|       | 0.850 | -5084.294 | 289.0508 | -17.58962 | 0.0000 |
|       | 0.900 | -4681.563 | 234.1768 | -19.99158 | 0.0000 |
|       | 0.950 | -4228.606 | 395.7740 | -10.68440 | 0.0000 |
| LIG   | 0.050 | -14.03903 | 2.198582 | -6.385492 | 0.0000 |
|       | 0.100 | -14.08298 | 2.435399 | -5.782616 | 0.0000 |
|       | 0.150 | -12.27879 | 2.821956 | -4.351162 | 0.0000 |
|       | 0.200 | -8.135725 | 5.147886 | -1.580401 | 0.1142 |
|       | 0.250 | 3.001367  | 5.175678 | 0.579898  | 0.5621 |
|       | 0.300 | 9.492309  | 4.587734 | 2.069063  | 0.0387 |
|       | 0.350 | 14.67633  | 4.218982 | 3.478641  | 0.0005 |
|       | 0.400 | 18.39972  | 4.153315 | 4.430129  | 0.0000 |

|                  |       |           |          |           |        |
|------------------|-------|-----------|----------|-----------|--------|
|                  | 0.450 | 20.08338  | 4.185540 | 4.798278  | 0.0000 |
|                  | 0.500 | 22.61293  | 3.879086 | 5.829447  | 0.0000 |
|                  | 0.550 | 23.41295  | 4.641197 | 5.044594  | 0.0000 |
|                  | 0.600 | 20.60510  | 6.705456 | 3.072886  | 0.0022 |
|                  | 0.650 | 15.10250  | 7.919701 | 1.906954  | 0.0567 |
|                  | 0.700 | 6.418061  | 8.900397 | 0.721098  | 0.4710 |
|                  | 0.750 | -1.427380 | 7.369908 | -0.193677 | 0.8465 |
|                  | 0.800 | -5.277042 | 2.845338 | -1.854627 | 0.0639 |
|                  | 0.850 | -3.500974 | 1.626701 | -2.152193 | 0.0315 |
|                  | 0.900 | -2.545653 | 1.664361 | -1.529508 | 0.1264 |
|                  | 0.950 | -0.564442 | 2.193199 | -0.257360 | 0.7969 |
| WS2 <sup>2</sup> | 0.050 | -24.74732 | 6.600397 | -3.749369 | 0.0002 |
|                  | 0.100 | -16.14691 | 6.281539 | -2.570534 | 0.0103 |
|                  | 0.150 | -20.96436 | 6.643411 | -3.155662 | 0.0016 |
|                  | 0.200 | -18.27323 | 6.069603 | -3.010614 | 0.0027 |
|                  | 0.250 | -20.32316 | 6.018140 | -3.376983 | 0.0008 |
|                  | 0.300 | -19.70112 | 6.523337 | -3.020098 | 0.0026 |
|                  | 0.350 | -27.22092 | 9.097256 | -2.992212 | 0.0028 |
|                  | 0.400 | -44.23555 | 10.27887 | -4.303543 | 0.0000 |
|                  | 0.450 | -50.97475 | 8.923434 | -5.712458 | 0.0000 |
|                  | 0.500 | -55.25912 | 11.00345 | -5.021980 | 0.0000 |
|                  | 0.550 | -69.26274 | 12.74598 | -5.434083 | 0.0000 |
|                  | 0.600 | -62.98533 | 16.19721 | -3.888654 | 0.0001 |
|                  | 0.650 | -41.30976 | 17.75339 | -2.326866 | 0.0201 |
|                  | 0.700 | -23.19782 | 14.88684 | -1.558277 | 0.1194 |
|                  | 0.750 | -8.285124 | 14.64039 | -0.565909 | 0.5715 |
|                  | 0.800 | -2.551873 | 16.24252 | -0.157111 | 0.8752 |
|                  | 0.850 | 1.008259  | 16.97647 | 0.059392  | 0.9526 |
|                  | 0.900 | 27.90422  | 18.15660 | 1.536864  | 0.1245 |
|                  | 0.950 | 44.33377  | 14.91991 | 2.971450  | 0.0030 |
| WD2 <sup>2</sup> | 0.050 | 0.000997  | 0.000702 | 1.421258  | 0.1555 |
|                  | 0.100 | 0.000957  | 0.001236 | 0.774766  | 0.4386 |
|                  | 0.150 | 0.002580  | 0.001511 | 1.707865  | 0.0879 |
|                  | 0.200 | 0.003504  | 0.002242 | 1.562566  | 0.1184 |
|                  | 0.250 | 0.007191  | 0.002182 | 3.295306  | 0.0010 |
|                  | 0.300 | 0.008151  | 0.002611 | 3.121823  | 0.0018 |
|                  | 0.350 | 0.005796  | 0.003211 | 1.804899  | 0.0713 |
|                  | 0.400 | 0.002946  | 0.003464 | 0.850445  | 0.3952 |
|                  | 0.450 | 0.000788  | 0.003642 | 0.216304  | 0.8288 |
|                  | 0.500 | -0.003504 | 0.004241 | -0.826059 | 0.4089 |
|                  | 0.550 | -0.007554 | 0.002899 | -2.605381 | 0.0093 |
|                  | 0.600 | -0.008795 | 0.002263 | -3.887342 | 0.0001 |
|                  | 0.650 | -0.011277 | 0.002379 | -4.741279 | 0.0000 |
|                  | 0.700 | -0.011775 | 0.003437 | -3.425603 | 0.0006 |

|       |           |          |           |        |
|-------|-----------|----------|-----------|--------|
| 0.750 | -0.009160 | 0.004139 | -2.213020 | 0.0271 |
| 0.800 | -0.006265 | 0.002933 | -2.135845 | 0.0329 |
| 0.850 | -0.003919 | 0.002219 | -1.766247 | 0.0776 |
| 0.900 | -0.000736 | 0.001825 | -0.403623 | 0.6866 |
| 0.950 | -3.14E-05 | 0.001507 | -0.020844 | 0.9834 |

---



## 附录 B

### R 软件代码：描述性分析与相关性分析

```
#cor=read.csv("D://rvr.csv",header=T)
#cor=read.csv("D://mor.csv",header=T)
cor=read.csv("D://total.csv",header=T)
head(cor) #r 内置数据，查看数据前 6 行数据
summary(cor) #描述性分析
pairs(cor) #散点图
cor1<-cor(cor) #计算相关系数
cor1
use="complete.obs" #删除缺失值或字符变量
round(cor1,digits=2) #输出并保留两位小数
#####安装 corrplot 程序包
#install.packages("corrplot") #安装特殊绘图的程序包
library(corrplot) #载入程序包
#install.packages("PerformanceAnalytics")
library(PerformanceAnalytics)
#chart.Correlation(cor, histogram=FALSE, pch=19)
#1 热力图
#corrplot(cor1,method="shade",shade.col=NA,tl.col="black",tl.srt=45,order="AOE")
#2 划分上下对角阵之带系数的椭圆
#corrplot.mixed(cor1,upper="ellipse") #带系数的椭圆
#3 划分上下对角阵之三角圆圈图
#corrplot(cor1,method="circle",type="upper",order="hclust")
```

### Eviews 软件操作：点击式的步骤省略，以下为部分非线性分位数回归等模型代码

```
genr lndewpoint=log(dewpoint)
genr exprh=exp(rh)
genr ws2a2=ws2a^2
genr wd2a2=wd2a^2
genr lnwd2a=log(wd2a)
mor_1a c rh qfe dewpoint lights ws2a2 wd2a cw2a temp pains qnh
rvr_1a c rh qfe dewpoint lights ws2a2 wd2a cw2a temp pains qnh
mor_1a c rh qfe dewpoint lights ws2a wd2a
mor_1a c rh qfe lndewpoint lights lnws2a lnwd2a
mor_1a c ln rh qfe lndewpoint lights lnws2a lnwd2a
mor_1a c ln rh qfe lndewpoint lights ws2a2 wd2a2
mor_1a c exprh qfe lndewpoint lights ws2a2 wd2a3
mor_1a c exprh qfe lndewpoint lights lnws2a2 wd2a
mor_1a c exprh qfe lndewpoint lights ws2a2 wd2a2
```

```

rvr_1a c rh qfe dewpoint lights ws2a wd2a
rvr_1a c rh qfe lndewpoint lights lnws2a lnwd2a
rvr_1a c lnrh qfe lndewpoint lights lnws2a lnwd2a
rvr_1a c lnrh qfe lndewpoint lights ws2a2 wd2a2
rvr_1a c exprh qfe lndewpoint lights ws2a2 wd2a3
rvr_1a c exprh qfe lndewpoint lights lnws2a2 wd2a
rvr_1a c exprh qfe lndewpoint lights ws2a2 wd2a2

```

**Matlab 软件：**基于梯度导向滤波技术的暗通道理论去雾技术以及透射率，消光系数和能见度的计算

```

clc;
clear all;
close all;
for ii=1:100
    J = imread([num2str(ii) '.bmp']);
    % J = imread('D:\5.jpg');
    J = double(J);
    J = J ./255 ;
    %figure(1); imshow(J);
    %% 求暗通道图像 Jdark = min(min());
    Jdark = Idark(J);
    %figure(2);imshow(Jdark,[]);
    % 本代码采用梯度导向滤波实现
    Jdark = gradient_guidedfilter(Jdark,Jdark, 0.04);
    %figure(3);imshow(Jdark,[]);
    %% 大气物理模型  $J = I * t + A * (1 - t)$  【直接衰减项】 + 【大气光照】
    % 透射率 t 与深度的关系  $t = \exp(-a * \text{depth})$ 
    w = 0.95; %雾的保留系数
    Jt = 1 - w * Jdark; %求解透射率

    %% 求解全局大气光照
    % 1.首先对输入的有雾图像 I 求解其暗通道图像 Jdark。
    % 2.选择 Jdark 总像素点个数千分之一 (N/1000) 个最亮的像素点，记录像素点 (x,y) 坐标
    % 3.根据点的坐标分别在原图像 J 的三个通道 (r,g,b) 内找到这些像素点并加和得到 (sum_r,sum_g,sum_b) .
    % 4.Ac=[Ar,Ag,Ab]. 其中  $Ar = \text{sum\_r}/N$ ;  $Ag = \text{sum\_g}/N$ ;  $Ab = \text{sum\_b}/N$ .
    [m,n,~] = size(J);
    N = floor( m*n./1000 );
    MaxPos = [0,0]; % 初始化
    for i=1:1:N
        MaxValue = max(max(Jdark));
        [x,y] = find(Jdark==MaxValue);
        Jdack(Jdark==MaxValue) = 0; %最大值置零，寻找下一次次大值
    end
end

```

```

%检查长度
MaxPos = vertcat(MaxPos,[x,y]);
Cnt = length(MaxPos(1));
if Cnt > N
    break;
end
end
MaxPosN = MaxPos(2:N+1,:);

Rsum = 0; Jr = J(:,1);
Gsum = 0; Jg = J(:,2);
Bsum = 0; Jb = J(:,3);
for j=1:1:N
    Rsum = Rsum + Jr(MaxPosN(j,1),MaxPosN(j,2));
    Gsum = Gsum + Jg(MaxPosN(j,1),MaxPosN(j,2));
    Bsum = Bsum + Jb(MaxPosN(j,1),MaxPosN(j,2));
end

Ac = [Rsum/N, Gsum/N, Bsum/N];

%% 求解清晰的图像
% 根据  $J = I*t + A*(1-t)$   $I = (J-A)/Jt + A$ 
Iorg = zeros(m,n,3);
for i = 1:1:m
    for j = 1:1:n
        for k = 1:1:3
            Iorg(i,j,k) = (J(i,j,k)-Ac(k)) ./ Jt(i,j) + Ac(k);
        end
    end
end
end
%figure(4); imshow(Iorg,[]);
total=0;
for i=1:720
    for j=1:1280
        total=total+log(Jt(i,j));
    end
end
end
beta=abs(total/(720*1280*30));
MOR(ii)=log(0.05)/(-beta);
end
%function1
function Jdark = Idark( I )
% output: Jdark = min(min(r),min(g),min(b));
Wnd = 3;

```

```

Ir = I(:,:,1);
Ig = I(:,:,2);
Ib = I(:,:,3);
%% 图像拓展
[m,n,~] = size(I);
Irr = zeros(m+Wnd-1, n+Wnd-1);
Irr((Wnd-1)/2 : m+(Wnd-1)/2-1, (Wnd-1)/2 : n+(Wnd-1)/2-1) = Ir;
Igg = zeros(m+Wnd-1, n+Wnd-1);
Igg((Wnd-1)/2 : m+(Wnd-1)/2-1, (Wnd-1)/2 : n+(Wnd-1)/2-1) = Ig;
Ibb = zeros(m+Wnd-1, n+Wnd-1);
Ibb((Wnd-1)/2 : m+(Wnd-1)/2-1, (Wnd-1)/2 : n+(Wnd-1)/2-1) = Ib;
%% 暗通道
for i=1:1:m
    for j=1:1:n
        Rmin = min(min ( Irr(i:i+Wnd-1, j:j+Wnd-1) ));
        Gmin = min(min ( Igg(i:i+Wnd-1, j:j+Wnd-1) ));
        Bmin = min(min ( Ibb(i:i+Wnd-1, j:j+Wnd-1) ));
        Jdark(i,j) = min(min(Rmin,Gmin),Bmin);
    end
end
end
%function2
function q = gradient_guidedfilter(I, p, eps)
%   GUIDEDFILTER    O(1) time implementation of guided filter.
%   - guidance image: I (should be a gray-scale/single channel image)
%   - filtering input image: p (should be a gray-scale/single channel image)
%   - regularization parameter: eps
r=16;
[hei, wid] = size(I);
N = boxfilter(ones(hei, wid), r); % the size of each local patch; N=(2r+1)^2 except for boundary
pixels.
mean_I = boxfilter(I, r) ./ N;
mean_p = boxfilter(p, r) ./ N;
mean_Ip = boxfilter(I.*p, r) ./ N;
cov_Ip = mean_Ip - mean_I .* mean_p; % this is the covariance of (I, p) in each local patch.
mean_II = boxfilter(I.*I, r) ./ N;
var_I = mean_II - mean_I .* mean_I;
%weight
epsilon=(0.001*(max(p(:))-min(p(:))))^2;
r1=1;
N1 = boxfilter(ones(hei, wid), r1); % the size of each local patch; N=(2r+1)^2 except for
boundary pixels.
mean_I1 = boxfilter(I, r1) ./ N1;
mean_II1 = boxfilter(I.*I, r1) ./ N1;

```

```

var_I1 = mean_I1 - mean_I1 .* mean_I1;
chi_I=sqrt(abs(var_I1.*var_I));
weight=(chi_I+epsilon)/(mean(chi_I(:))+epsilon);
gamma = (4/(mean(chi_I(:))-min(chi_I(:))))*(chi_I-mean(chi_I(:)));
gamma = 1 - 1./(1 + exp(gamma));
%result
a = (cov_Ip + (eps./weight).*gamma) ./ (var_I + (eps./weight));
b = mean_p - a .* mean_I;
mean_a = boxfilter(a, r) ./ N;
mean_b = boxfilter(b, r) ./ N;
q = mean_a .* I + mean_b;
end
%function3
function imDst = boxfilter(imSrc, r)
%   BOXFILTER   O(1) time box filtering using cumulative sum
%
%   - Definition imDst(x, y)=sum(sum(imSrc(x-r:x+r,y-r:y+r)));
%   - Running time independent of r;
%   - Equivalent to the function: colfilt(imSrc, [2*r+1, 2*r+1], 'sliding', @sum);
%   - But much faster.
[hei, wid] = size(imSrc);
imDst = zeros(size(imSrc));
%cumulative sum over Y axis
imCum = cumsum(imSrc, 1);
%difference over Y axis
imDst(1:r+1, :) = imCum(1+r:2*r+1, :);
imDst(r+2:hei-r, :) = imCum(2*r+2:hei, :) - imCum(1:hei-2*r-1, :);
imDst(hei-r+1:hei, :) = repmat(imCum(hei, :), [r, 1]) - imCum(hei-2*r:hei-r-1, :);
%cumulative sum over X axis
imCum = cumsum(imDst, 2);
%difference over X axis
imDst(:, 1:r+1) = imCum(:, 1+r:2*r+1);
imDst(:, r+2:wid-r) = imCum(:, 2*r+2:wid) - imCum(:, 1:wid-2*r-1);
imDst(:, wid-r+1:wid) = repmat(imCum(:, wid), [1, r]) - imCum(:, wid-2*r:wid-r-1);
end

```

（附录仅展示部分代码，其余代码见附件）