



中国研究生创新实践系列大赛
“华为杯”第十六届中国研究生
数学建模竞赛

学 校 浙江工商大学

参赛队号 19103530011

1. 陈向楠

队员姓名 2. 陈佳佳

3. 李振涛

中国研究生创新实践系列大赛

“华为杯”第十六届中国研究生

数学建模竞赛

题目 对气温变化数据的量化分析与预测

摘 要：

本文主要针对加拿大气候变化数据进行量化分析，进而对未来全球平均气温进行预测，探索全球变暖与极端天气之间的相互关系，构建相应数学模型，并提出了减缓全球变暖趋势的建议。文章综合采用了模糊综合评价法、熵权法、时间序列模型、灰度关联度分析、基于 LSTM-ARIMA 的气温组合预测模型、复杂网络模型等经典方法来研究气温数据的量化分析相关问题。

针对问题一，在第一问中，依据地区将加拿大划分为五个地区，基于以上划分采用定量与定性相结合方法，探索其历史气温数据的时空变化趋势。首先绘制气温时间序列图，并采用 BP 突变点检验分析异常点；接下来构建 ARMA 模型对气温数据进行平稳性检验；最后运用模糊评价法对指标项打分，熵权法获得气温影响因子评分，计算系统权重值得出各个因素对于气温变化趋势的影响。结果表明在全球变暖的背景下，包含北极圈在内的全球多地气温普遍上升，出现极端高温天气，其中加拿大变暖速度显著高于全球平均水平，气候变化引发极端事件频发；对于第二问，为完成海洋表面温度的规律分析，采用复杂网络建模构建全球海洋气候的非线性和线性复杂网络模型并通过拓扑仿真方法对比两者的拓扑性质，利用度分布、聚类系数和介数等统计特征可视化海洋不同海域的能量传递等现象。结果表明高纬度地区的海表面温变化的复杂度低于低纬度地区。因为高纬度地区海表面温度的季节性更加明显，序列的规则性更强。相对来说季节性因素对低纬度地区海表面温度的影响不大。四季网络之间的演化从某种角度上体现了海温系统随季节动态变化的特征。

针对问题二，通过研究 2008-2019 年间全球平均气温值以及影响因素数据，运用 LSTM-ARIMA 气温组合预测算法，计算不同的气温影响因素权重，将碳排放量、地球散热、降水、海洋温度变化、大气环流这 5 个最高影响因素作为输入变量，全球平均气温值作为输出变量进行模型训练；通过组合模型预测方法确定参数，最后利用组合模型预测未来 25 年全球平均气温变化，结果表明未来全球变暖现象还将持续，

但可能会出现短暂的局部降温情况，如在 2027 年出现一个小高峰，2022 年和 2033 年出现小低谷，说明在这两个时间段全球平均气温可能会出现异常。

针对问题三，在第一问中，构建复杂网络模型，将全球各地区极端天气表示为复杂网络中的异常节点，通过查看异常节点中的网络性质是否正常对全球气候变化和极端气象之间的关系进行分析，探索气候变化是否对极端气象发生的严重程度及其发生概率和不确定性产生影响。结果表明**极端气象的出现与气候变化相关，且全球气候变暖进一步增加了极端气象的出现**；第二问以美国为局地极寒地区的典型代表，通过建立神经网络模型对是否发生局地极寒现象进行分析，对第一问结论做进一步验证，结果表明全球平均气温的上升对美国降雪量变化有正向影响，气温越高，降雪量越大。而全球平均气温对美国冬季平均气温又负向影响，即全球平均气温越高，美国冬季气温越低。综上，**全球变暖和局地极寒现象并不矛盾，相反，两者之间恰恰有很强的相关性。**

针对问题四，在第一问中，通过研究局地极寒事件发现，**全球变暖造成北极升温和北极海冰融化，北极增温使极地和中纬度地区温差减少，极地高压对极涡的控制能力减弱，极涡较以往更容易携带冷空气南下。**同时，北极海冰融化会加速北极升温，北极秋季海冰面积减少也会造成西风急流和经向风减弱，最终扩大了极地冷空气的影响。考虑到北极增温带来的“炸弹气旋”能够通过能量频散的形式影响欧亚大陆的环流形式，强大的西伯利亚高压给东亚带来更强劲的冷空气，结果导致中国遭遇强劲的寒潮。第二问中，本文以“热效应下的气候紊乱”来代替“全球变暖”概念。首先全球变暖是一个趋势，虽然太阳活动的减弱造成了短暂的全球变暖停滞状态，但改变不了全球升温的长期时间趋势，当前全球气温状况远比我们看到的升温速率快，“全球变热”仍是当代人急需解决的问题。全球变暖引发了地球气候系统的失衡，造成全球气候变化更加紊乱。

**关键词：全球变暖 局部极寒 时间序列 复杂网络算法 拓扑仿真
LSTM-ARIMA 气温组合预测算法**

目录

1 问题重述	5
1.1 问题背景	5
1.2 问题解析	5
2 模型假设	8
3 符号说明	9
4 问题一建模与求解	10
4.1 第一问问题分析	10
4.1.1 方法一：定量分析	10
4.1.2 方法二：定性分析 1	15
4.1.3 方法三：定性分析 2	23
4.2 第一问结论	26
4.3 第二问问题分析	26
4.4 第二问模型建立	27
4.4.1 模型假设	27
4.4.2 网络拓扑结构	27
4.4.3 相似性度量	27
4.5 第二问模型结论	29
5 问题二建模与求解	30
5.1 问题分析	30
5.2 模型建立	30
5.2.1 LSTM 模型	30
5.2.2 ARIMA 模型	32
5.3 模型求解	32
5.3.1 数据准备	32
5.3.2 特征提取	34
5.3.3 模型训练	34
5.4 问题二模型评价	37
6 问题三建模与求解	39
6.1 第一问问题分析	39
6.2 第一问复杂网络模型建立	39
6.3 第一问复杂网络模型求解	41
6.4 第二问问题分析	42

6.5 第二问神经网络模型建立	42
6.6 第二问神经网络模型求解	44
6.6.1 算法说明	44
6.6.2 结果分析	45
7 问题四建模与求解	47
7.1 第一问问题分析	47
7.2 第一问问题结论	47
7.3 第二问新概念提出	51
7.4 第二问新概念解释	51
参考文献	53
附录	57

1 问题重述

1.1 问题背景

近百年来全球气候正经历着以全球变暖为主要特征的显著变化，在上个世纪后半叶，全球温室气体浓度持续快速升高，全球平均表面气温整体保持上升趋势，伴随而来的是海平面升高、高原雪盖融化、极地海冰消退、海表面温度不断上升等一系列气候问题。到了 21 世纪初期，人类活动排放的温室气体会导致全球变暖几乎已经是一个不再具有争议的事实。

在全球变暖的背景之下，1998 年后的全球平均表面气温出现了变暖趋势变缓。科学数据显示自 21 世纪以来，近十年间全球表面气温几乎没有再继续升高过，而与此同时全球温室气体浓度却不断创下新高。这种全球大气温室气体的浓度持续快速上升但同一时期全球表面气温却出现了变暖速率趋缓的现象被称为全球变暖“停滞”。现有的气候学理论不能够完全解释全球变暖趋缓现象的出现，所以这种现象自出现起就立即同时受到了广泛关注，并引发了科学讨论，由此也导致公众和政府对于全球变暖的关注与疑惑。加上极端天气气候事件频发，给全球各地的经济、社会和环境带来了严重的负面影响，制约了社会和经济的发展，因而对全球气候的准确分析与预测引起了各国政府以及国际机构的普遍关注。

1.2 问题解析

问题一：

第一问：从加拿大各地天气变化的历史数据中挖掘出该地区温度的时空变化趋势。首先对从气候区、地形地貌、地区位置将加拿大整体划分为五个区域，接着在此基础上选取具有代表性的城市进行详细分析。本题的解题思路是将原来的加拿大各地区气温数据分解为四部分来看——趋势、周期、时期和不稳定因素，然后综合这些因素分析归纳出加拿大各地区温度的时空变化趋势。本文采用定性与定量分析相结合的方法来解题，具体步骤如下：

步骤一：首先从题目所给出的网址中下载加拿大天气变化历史数据，并对数据进行预处理，具体包括数据清洗、去噪等，对数据集进行特征提取，并结合天气状况的多变性和复杂性，对气温变化的相关影响因素进行整理。

步骤二：针对单一气候区，结合经济、人口、社会、环境等多方面因素选取具有代表性的城市地区，从时间性、周期性、季节性等多个角度研究天气变化规律。

步骤三：利用随机过程理论和数理统计方法绘制序列散点图，并添加趋势线以反映地区多年来气温长期变化趋势。然后利用 BP 突变点检验分析异常点，计算气温时序的变化速度，从而说明加拿大各地气温的时空变化规律。

步骤四：在步骤三的基础上，为防止出现“伪回归”的问题，对气温数据进行纯随机性和平稳性检验；然后利用 ARIMA 模型进行建模，接着计算模型自相关系数、偏自相关系数以及 AIC、BIC 指标值对模型参数进行定阶，最后对模型做出评价。

步骤五：针对以上分析结果，利用熵权法对影响气温的各种因素权重进行计算，并分析各个因素对于气温变化趋势的影响。

第二问：从海洋表面温度历史数据探寻规律。首先利用复杂网络理论构建了全球海洋气候的非线性和线性复杂网络模型，仿真对比了两者的拓扑性质。再利用度分布、聚类系数和介数等统计特征可视化海洋不同海域的能量传递等现象，分析和探索海洋气候的系统稳定性和季节性差异等规律。本题的解题步骤分四步：

步骤一：获取数据。利用网络爬虫从国家海洋信息中心上爬取 2019 年的全球日均海表面温度数据。

步骤二：数据预处理。将全球海洋进行固定尺度网格划分，均值化每个网格内的海表面温度。

步骤三：复杂网络建模。根据相似性方法来度量全球海洋不同海域之间的相似度，得到全球相似度矩阵，然后设定不同的阈值来构建网络，并做详细的拓扑分析。

步骤四：仿真实验与可视化分析。基于互信息和皮尔逊相关系数的网络节点度可视化 SST 数据在地球上的离散分布情况。利用度分布、聚类系数和介数等统计特征分析 SST 差异。

问题二：

建立一个刻画气候变化的模型，并对未来 25 年的气候变化进行预测。对于问题二，可以分为两步来进行操作，首先利用基于 LSTM-ARIMA 的气温组合预测模型对加拿大地区各地的气温变化进行分析，其次，借助以上分析以及建立相应的模型对全球未来 25 年的气候变化进行预测。本题的解题步骤分为六步：

步骤一：获取所需的气温变化数据，对数据进行基本统计和可视化展示；

步骤二：对数据集的影响因素进行特征提取，并将数据分为训练集和测试集两部分；

步骤三：根据问题一得到的分析结果，对训练集数据采用五折交叉验证，以避免过拟合现象。将特征提取出的地球吸热、散热、降水、大气环流以及海洋温度变化这 5 个因素作为输入变量（输入特征），当前时刻 t 的温度作为输出变量（标签）；

步骤四：建立 LSTM 模型和 ARIMA 模型，先利用 ARIMA 模型对数据集进行预测，再利用 LSTM 网络依次对 ARIMA 模型的残差序列进行预测；

步骤五：将 ARIMA 模型预测结果与 LSTM 对残差的预测结果进行相加，调整参数，得到 LSTM-ARIMA 组合模型的最终预测结果。

步骤六：综上，对组合模型进行评价。

问题三：

第一问：查询和收集网上相关信息以及查阅相关文献，解释极端天气现象的产生，并探讨其与气候变化的关系。问题三第一问的主旨是探索极端气象与气候变化之间的相关性，本题的解题思路分两步：

步骤一：根据题目所给规则：因为“极寒天气”是某地的天气现象，并没有给定指定地区，故实验以历史最热月份、历史最热日、历史最干燥 5 天、历史最潮湿 5 天等因素作为评价是否发生局地极端气候事件指标，获取全球不同地区气温数据和极端天气相关数据。

步骤二：构建复杂网络模型，将极端天气表示为复杂网络中的异常节点，探索全球不同地区气温变化与极端气象出现之间的关系，探索气候变化是否对极端气象发生的严重程度及其发生概率和不确定性产生影响。

第二问：根据我们上述收据的数据合理建立数学模型，解释全球变暖和局部极寒现象的出现是否矛盾。问题三第二问的主旨是探究全球变暖与局地极寒现象是否矛盾，是对第一问数据的进一步运用。本题的解题思路分两步：

步骤一：根据问题一中得到的猜想：极端气象的出现与气候变化相关，本文以具体城市为例对以上猜想做进一步验证。

步骤二：考虑到美国频繁遭受极寒袭击，本文决定以美国为案例，以年降雪量和冬季平均气温作为是否发生局地极寒现象指标，探究其与气温之间的关系，并提出两种假设：

- a. 全球年平均气温对美国年降雪量有正向影响；
- b. 全球年平均气温对美国冬季平均气温有负向影响。

问题四：

第一问：用通俗易懂的文字解释：“全球变暖了，某地今年的冬天特别冷”之间的关系。本文以局地极寒事件、全球变暖为切入点，建立全球变暖而局地极寒的理论模型。本小问的解题步骤可以分为三步：

步骤一：查询和收集网上相关信息以及查阅相关文献，寻找全球变暖对地球造成了哪些影响，尤其是引发了哪些天气事件以及如何引发的；

步骤二：查找资料，分析全球局地极寒事件的直接诱因、共同原因；

步骤三：由前节研究结论可知，极寒天气与全球变暖存在统计学意义上的关联，结合现有理论进一步深入挖掘全球变暖与极寒天气的关联性。

第二问：给“全球变暖”下个新定义，从而更好地反映气候变化的趋势和复杂性。本小问的解题步骤可分为三步

步骤一：搜寻资料，寻找目前是否有学者提出替代“全球变暖”的新概念，以提供借鉴和激发思考；

步骤二：深入分析并总结出全球变暖的趋势和特性。

步骤三：为了更好地反映气候变化的趋势性和复杂性，在现有理论的基础之上，用一个全新的概念“热效应下的气候紊乱”替代“全球变暖”。

2 模型假设

在建模的过程中，我们在不影响模型的可靠性和有效性的前提下，为了使模型更加简单明确。做出以下假设：

(1) 通过查阅资料将加拿大整体划分为五个区域，分别为大西洋地区、中部地区、大草原地区、太平洋沿岸地区、北方地区；

(2) 对于同一个城市的不同气温观测点，观测到的数据差别较小，故将其计作统一数据；

(3) 在假设 2 的基础上，对于缺失的温度数据，缺测 1-2 天数据，用相邻日的数据平均值进行插补；缺测 3 天以上的数据，利用同期数据资料的多年平均值进行插补。

(4) 所有的温度数据都是真实准确的，不存在错误和虚假数据。

(5) 考虑到实际情况，假设海洋构建为无向无权的简单网络，这种网络的任何节点之间最多只能有一条边，并且不存在自环路。

(6) 海洋网络中的边表征了不同海域之间温度变化在时间上的相似程度。

(7) 海洋网络的节点代表海洋上的某块区域。

3 符号说明

f: 忘记门

i: 输入门

o: 输出门

c: cell 状态激活向量

σ : sigmoid 激活函数

tanh: 双曲正切函数

W: 权重矩阵, 例如: W_{xf} 是忘记门权重矩阵, W_{xo} 是输出门权重矩阵

y_t : 时间序列 y 的当期值

y_{t-1} : y_t 前一期的值

$\phi_1, \phi_2, \phi_3, \dots, \phi_p$: 自回归系数

$\theta_1, \theta_2, \theta_3, \dots, \theta_q$: 移动平均系数

$\{\varepsilon_t\}$: 白噪声序列

γ_k : 自协方差函数

ACF: 自相关函数

l_{st} : 从节点 V_s 到节点 V_t 的所有最短路径的数目

l_{st}^i : 从节点 V_s 到节点 V_t 的所有最短路径中经过 V_i 的最短路径的数目

$G=(V,E,F)$: 网络拓扑结构三元组定义

$p(x,y)$: 两个随机变量 X 和 Y 的联合概率密度函数

$p(x)$: 随机变量 X 的边际概率密度函数

$I(X;Y)$: 联合分布 $p(x,y)$ 与乘积分布 $p(x)p(y)$ 之间的相对熵

4 问题一建模与求解

4.1 第一问问题分析

对于一个特定的区域来说，纬度、地形地势、海拔高度、下垫面都是恒定的常数，变动的只有天气状况和季节，当按季节进行分类之后，变动因素就只有天气状况。在按季节、按月、按季进行分析研究时，可最大限度减小不确定因素天气状况带来的影响，从而寻找到可用的潜在规律。

因此，本题的解题思路分为五步，主旨是定量方法与定性方法相结合，如图 4-1 所示。

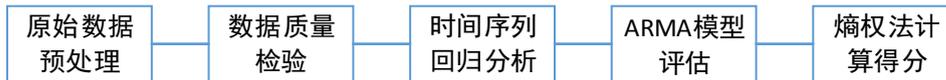


图 4-1 问题一的建模思路

步骤 1: 基于气象观测站点的均一性、资料年代的完整性等原则，综合考虑研究时段以及站点缺测和台站迁移等因素，本文选取 2010-2019 年 783 个站点资料的最高气温、最低气温和平均气温分析加拿大各地气温变化特征，并利用 Python 软件进行数据清洗、去噪预处理，根据天气状况的多变性和复杂性，本文参考世界气象组织和《第五次 IPCC 评估报告》对气温变化的相关影响因子进行整理。

步骤 2: 为控制数据质量，在 SPSS 软件中对气象数据均进行单样本 T 检验，且均通过检验 ($P < 0.05$)。在基准日气象数据基础上，运用多种统计方法分析加拿大地区多年来气候变化趋势，并运用累积距平法及独立样本 t 检验分析加拿大地区气温突变现象。

步骤 3: 利用随机过程理论和数理统计方法绘制序列散点图，并添加趋势线以反映地区多年来气温长期变化趋势。然后利用 BP 突变点检验分析异常点，计算气温时序的变化速度，从而说明加拿大各地气温的时空变化规律。

步骤 4: 在 3 的基础上，为防止出现“伪回归”的问题，对气温数据进行纯随机性和平稳性检验；然后利用 ARMA 模型进行建模，接着计算模型自相关系数以及 AIC 和 BIC 指标值、是否差分、模型定阶，最终对模型进行评价。

步骤 5: 针对以上分析结果，利用熵权法对影响气温的各种因素权重进行计算，并分析各个因素对于气温变化趋势的影响。

4.1.1 方法一：定量分析

4.1.1.1 时间序列模型建立

本题的建模思路主要由三部分组成定量评估和定量评估，本文首先结合人口、经济、社会、环境等综合因素考量，分别选取大西洋省份中的圣约翰城市、中部地区的多伦多、大草原地区的埃德蒙顿、太平洋沿岸地区的维多利亚、北方地区的黄刀等城市为例[1]，分析加拿大各地区不同环境下气温时空变化趋势。

(1) 局部多项式拟合

首先采用局部多项式的非参数方法拟合气温变化的长期趋势。多项式拟合是用一个多项式展开去拟合包含数个分析格点的一小块分析区域中的所有观测点，得到观测数据的客观分析场。目的是将离散点数据拟合成连续函数（曲线），从而分析

数据变化趋势。给定一组数据 $(x_i, y_i) (i=0,1,2,\dots)$, 求 n 次多项式

$$p_n(x) = a_0 + a_1x + \dots + a_nx^n = \sum_{k=0}^n a_k X^k, \quad (4-1)$$

使其满足:

$$F = \sum_{i=0}^n p_n(X_i + Y_i)^2 = \sum_{i=0}^n \left(\sum_{k=0}^n a_k x_i^k - y_i \right)^2 = \min \quad (4-2)$$

模型的突出特点:不用提前设定模型,避免了模型设定误差,以数据自身变化走势拟合数据,可以更好的反映气温变化趋势。虽然局部多项式回归拟合序列可以反应气温序列随时间的变化趋势,但从趋势线判断序列的突变点主观色彩较强,因此使用统计检验判断序列的突变点更加客观准确。

(2) BP 突变点检验

利用时间序列断点检验分析年度平均最高、最低气温及温差、季度平均最高、最低气温及温差,月度平均最高、最低气温及温差序列的突变点。BP 结构断点检验采用的是线性回归模型,模型的时间长度为 T ,并假定该模型中存在 m 个结构突变断点。假定存在线性回归模型:

$$\begin{aligned} w_t &= x_t' \alpha + z_t' \beta_1 + \psi_t, t=1, \dots, T_1 \\ w_t &= x_t' \alpha + z_t' \beta_2 + \psi_t, t=T_1+1, \dots, T_2 \\ &\dots \\ w_t &= x_t' \alpha + z_t' \beta_m + \psi_t, t=T_m+1, \dots, T_{m+1} \end{aligned} \quad (4-3)$$

其中, ψ_t 为残差项, T_1, T_2, \dots, T_m 为突变时间点。模型的向量表达式为: $W = X\alpha + Z\beta + \psi$ 。BP 结构断点检验法就是利用最小二乘法求解未知系数 $(\alpha, \beta_j, T_1, \dots, T_m)$ 使得模型中残差平方和最小。根据最小二乘法可得 α, β_j 的估计值为:

$$\min (W - X\alpha + Z\beta)' (W - X\alpha + Z\beta) = \sum_{i=1}^{m+1} \sum_{T_{i-1}+1}^{T_i} \left[w_t - x_t' \alpha - z_t' \beta_j \right]^2 \quad (4-4)$$

时间序列断点检验方法不需要设定结构突变点的个数,使用最小二乘法来判断检验模型是否存在结构突变点,可以准确的识别时间序列中存在的多个突变点,因此采用 BP 检验判断加拿大气温序列的突变点。最后根据突变点检验计算序列的变化速度,分析加拿大气温变化情况。

4.1.1.2 时间序列模型求解

首先利用编程软件运用局部多项式拟合的非参数方法计算选取点地区的年度平均最高、最低气温及年度平均气温、月度平均最高、最低气温及月度平均气温等,再借助 BP 突变点检验及相对变率等指标和方法,从时间尺度和空间尺度两个方面系统评判加拿大地区气温的时空变化特征。问题基于上述模型与方法,经过 Python 语言进行实现(代码见附件),最终分析结果如下:

(1) 圣约翰年度气温变化趋势分析

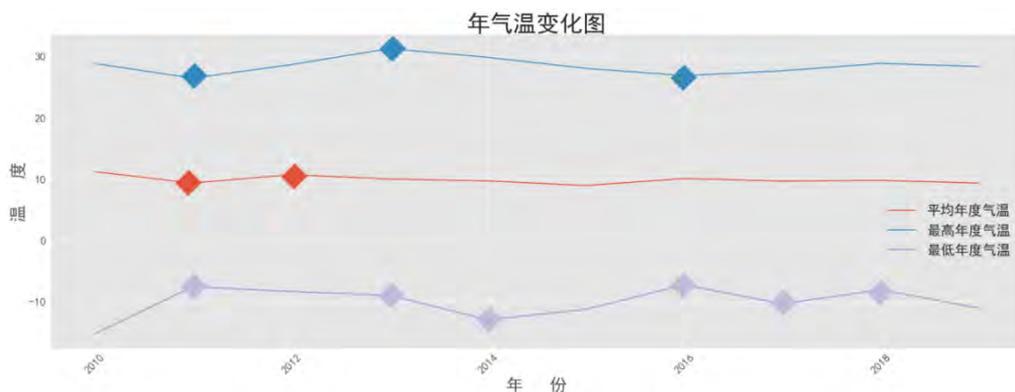


图 4-2 圣约翰 2010-2019 年年度气温变化图

①年度平均最高气温趋势分析：如图 4-2 所示，从 2010-2019 年圣约翰年度平均最高气温的长期变化趋势可以看出，圣约翰最高年度气温大体呈先升高后降低再升高的变化趋势。根据施瓦茨准则最小原则，可知 2010-2019 年圣约翰年度平均最高气温有 5 个突变点。其中：2014 年、2016 年分别是圣约翰市年度平均最高气温的转折突变点，2011 年、2012 年、2018 年是三个变率突变点。将局部多项式趋势拟合结果和 PB 突变点结果综合可知，2010-2019 年间，圣约翰年度平均最高气温演变经历了 4 个阶段：2010-2014 年，年平均最高气温逐年上升，平均每年上升 0.75°C ，其中 2012-2014 年，年平均最高气温变化上升速度减缓，平均每年上升 0.45°C ；2014 年是圣约翰最特殊的一年，达到近 10 年间的最高平均温度年。2014-2016 年，年平均最高气温呈下降趋势，平均每年下降 1.375°C ；2016-2018 年，年平均最高气温转而上升，平均每年上升 1.05°C ；2018 年以后，年平均最高气温略微下降，稳定在 28°C 左右。

总体来看，2010-2019 年，圣约翰年度平均最高气温经历了暖-冷-暖的演变过程，且下降速度快于上升速度。2014-2016 年圣约翰年平均最高气温降低最快，2016-2018 年圣约翰年平均最高气温升高最快。

②年度平均最低气温趋势分析：从 2010-2019 年圣约翰年度平均最低气温的长期变化趋势可以看出，圣约翰最低年度气温大体呈先降低后升高再降低的变化趋势。从 BP 检验可知，2010-2019 年圣约翰年平均最低气温有 4 个突变点：2014 年、2016 年、2017 年、2018 年且都是转折突变点，波动比较明显。将局部多项式趋势拟合结果和 PB 突变点结果综合可知，圣约翰年平均最低气温的变化大体呈现五个不同的阶段。2010-2014 年，年平均最低气温大幅度下降，平均每年下降 2.5°C ，其中 2013-2014 年间下降最快；2014-2016 年，年平均最低气温变化忽而上升，平均每年上升 2.75°C ；2016-2018 年，圣约翰年平均最低气温变化呈现先下降后上升的势态，但是变化幅度不大，最低温差在 1°C 左右。

整体来说，2010-2019 年，圣约翰年平均最低气温呈下降趋势。2013-2014 年是圣约翰年平均最低气温下降速度最快的阶段，其次是 2018-2019 年。与年平均最高气温相对比，年平均最低低温变化趋势相对复杂，变化幅度较大。由此可以得出，圣约翰 2010-2019 年间，年度平均气温维持在 10°C 左右，其中，2014 年是比较特殊的一年，这一年圣约翰平均最高温度达到 30°C ，平均最低温度达到 -15°C 左右，平均温

差达到十年间最大。研究 2010-2019 年圣约翰的气温变化趋势可以看出，年度平均最高气温与年度平均最低气温变化刚好呈相反态势。

(2) 多伦多年度气温变化趋势分析

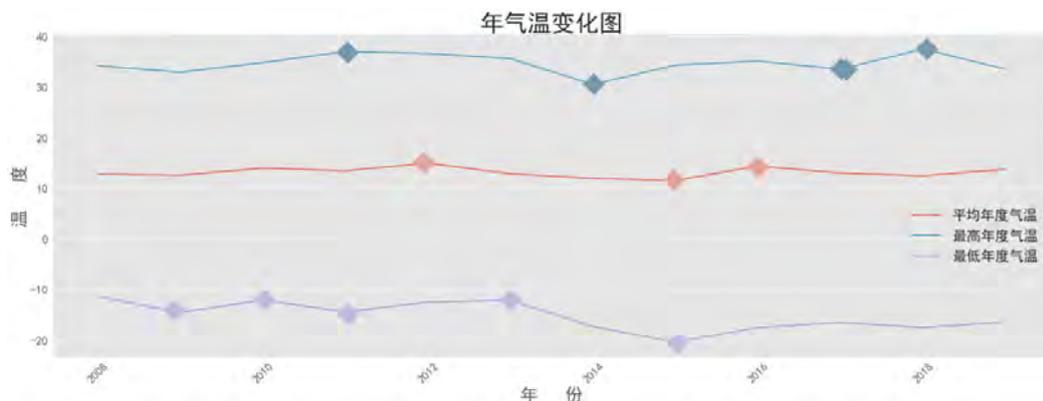


图 4-3 多伦多 2008-2019 年年度气温变化图

①年度平均最高气温趋势分析：如图 4-3 所示，从 2008-2019 年多伦多年度平均最高气温的长期变化趋势可以看出，多伦多最高年度气温大体呈先平稳升高后下降再逐步升高最后骤然下降的变化趋势。根据施瓦茨准则最小原则，可知 2008-2019 年多伦多年度平均最高气温有 7 个突变点。其中：2015 年、2016 年、2018 年分别是多伦多年度平均最高气温的转折突变点，2007 年、2011 年、2013 年、2017 年是三个变率突变点。将局部多项式趋势拟合结果和 PB 突变点结果综合可知，2008-2019 年间，多伦多年度平均最高气温演变经历了五个阶段：2008-2011 年，年平均最高气温总体上升，平均每年上升 0.65°C ，其中 2009 年稍微下降 1°C 左右，之后快速回温。2011-2015 年，年平均最高气温逐步下降，平均每年下降 1.95°C 左右；2015-2016 年年平均最高气温转而上升，上升速率非常快，平均 7.5°C 左右；2016-2018 年，年平均最高气温呈下降趋势，平均每年下降 8°C ；2018 年之后，年平均最高气温呈快速上升趋势。

总体来看，2008-2019 年，多伦多年度平均最高气温经历了暖-冷-暖-冷-暖的演变过程，且下降速度略快于上升速度。2017-2018 年圣约翰年平均最高气温降低最快，2018-2019 年圣约翰年平均最高气温升高最快。2018 是多伦多 10 多年间最反常的一年，年平均最高温度低至 18°C 左右，打破历年最高平均温度在 30°C 左右的界限。

②年度平均最低气温趋势分析：如图 4-3 所示，从 2008-2019 年多伦多年度平均最低气温的长期变化趋势可以看出，多伦多最低年度气温大体呈先下降后升高再下降的变化趋势。根据施瓦茨准则最小原则，可知 2008-2019 年多伦多年度平均最低气温有 5 个突变点：2009 年、2011 年、2013 年、2015 年、2018 年，且都是多伦多年度平均最低气温的转折突变点。将局部多项式趋势拟合结果和 PB 突变点结果综合可知，2008-2019 年间，多伦多年度平均最低气温演变经历了四个阶段：2008-2013 年，年平均最低气温稳定在 -15.5°C 左右，变化幅度均不大；2013-2015 年，年平均最低气温快速下降，平均每年下降 4.5°C 左右；2015-2018 年年平均最低气温转而上升，上升速率很快，平均 4.85°C 左右；2018 年之后，年平均最低气温呈快速下降趋势。

总体来看，2008-2019 年，多伦多年度平均最低气温经历了冷-暖-冷的演变过程，

且上升速度快于下降速度。2013-2015 年圣约翰年平均最低气温降低较快，2017-2018 年圣约翰年平均最低气温升高最快。2018 年，多伦多年平均最低温度达到-6℃左右，达到多年来最暖的平均最低温度。由此可以得出，多伦多 2008-2019 年间，年度平均气温大多维持在 10℃以上，其中，2015 年和 2018 年平均温度低至 5.5℃。研究 2008-2019 年多伦多的气温变化趋势可以看出，2018 年是多伦多 10 多年来最高年均温度和最低年均温度相差最小的一年，此年多伦多天气温度一直维持在-7.5℃~18℃，平均温度在 5℃左右。从图中还可以得出，在年平均最高温度、年平均最低温度相对稳定的情况下，多伦多天气一直维持在 10℃以上，温差波动较大的情况下，年均最低温度和年均最高温度呈相反趋势，即随着年最低温度的上升，年最高温度下降。这与多伦多地处加拿大的南部，属于温带大陆性湿润气候的事实相吻合。

(3) 埃德蒙顿年度气温变化趋势分析

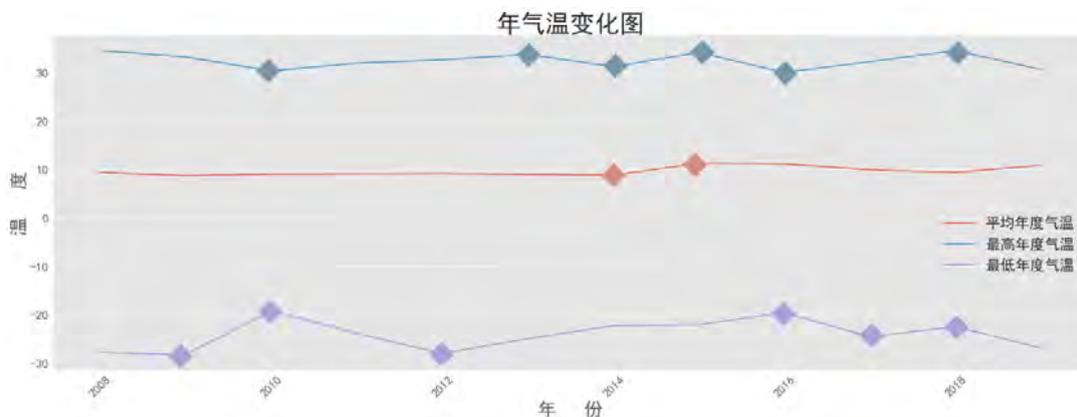


图 4-4 埃德蒙顿 2008-2019 年年度气温变化图

如图 4-4 所示，从 2008-2019 年埃德蒙顿年度平均气温变化图的长期变化趋势可以看出，埃德蒙顿年均最高气温整体上呈平稳趋势，变化幅度不大，基本维持在 30℃~35℃左右；年均最低气温大体呈上升趋势，根据施瓦茨准则最小原则，可知 2008-2019 年多伦多年度平均最低气温有 6 个突变点：其中 2009 年为变率突变点；2010 年、2012 年 2016 年、2017 年和 2018 年是多伦多年度平均最低气温的转折突变点。将局部多项式趋势拟合结果和 PB 突变点结果综合可知，2008-2019 年间，埃德蒙顿年度平均最低气温演变经历了四个阶段：年平均最低气温在 2008-2010 年逐年上升；接着在 2010-2012 年转而下降，平均每年下降 4.83℃左右，在 2012-2016 年间又逐步上升；在 2016 年之后总体呈下降趋势，其中在 2017 年有上升波动。总体来看，埃德蒙顿平均年度气温维持在 10℃左右，最高与最低温差相对稳定，在-26℃~31℃之间波动。

(4) 维多利亚年度气温变化趋势分析

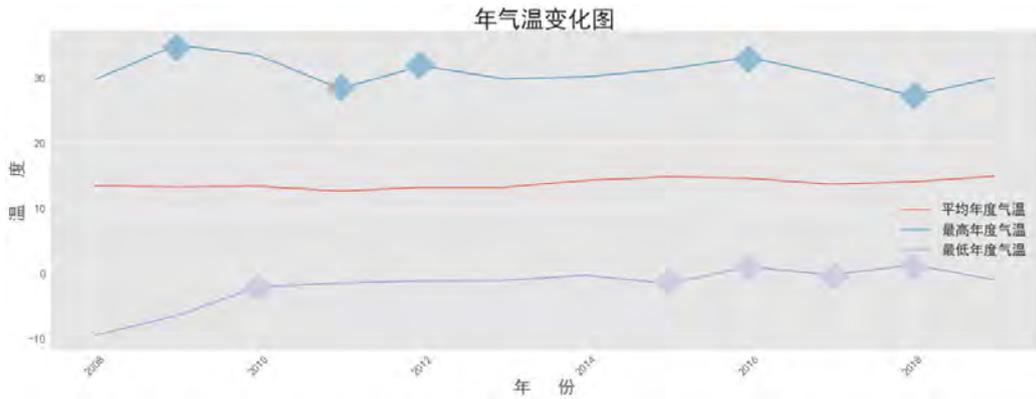


图 4-5 维多利亚 2008-2019 年年度气温变化图

如图 4-5 所示，从 2008-2019 年维多利亚年度平均气温变化图的长期变化趋势可以看出，维多利亚年均最低气温整体上呈上升趋势，2018-2019 年上升速率很快，平均上升 3.75°C 左右，在 2010 年之后上升速率趋于缓慢，最终年平均最低气温维持在 0°C 左右。相比于年平均最低气温，年均最高气温变化较频繁，将局部多项式趋势拟合结果和 PB 突变点结果综合可知，2008-2019 年间，维多利亚年度平均最高气温经历了暖-冷-暖-冷-暖的演变过程，且上升速度慢于下降速度，2018 年达到最低温度 27.6°C。由此得出，维多利亚常年均温度维持在 13°C 左右，气候温和。

(5) 耶罗纳夫年度气温变化趋势分析

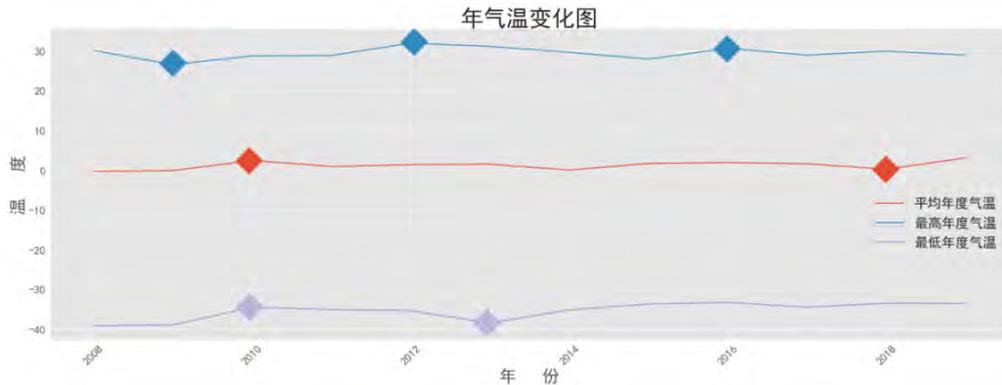


图 4-6 耶罗纳夫 2008-2019 年年度气温变化图

如图 4-6 所示，从 2008-2019 年耶罗纳夫年度平均气温变化图的长期变化趋势可以看出，维多利亚年均气温整体上呈稳定趋势，变化幅度不大。将局部多项式趋势拟合结果和 PB 突变点结果综合可知，2008-2019 年间，耶罗纳夫年度平均最高气温在 30°C 左右，其中 2009 年下降到 27.6°C；年度平均最低气温有两个转折突变点，总体上逐步上升，平均每年上升 0.87°C 左右，但均在 -30°C 以下。

4.1.2 方法二：定性分析一

4.1.2.1 ARMA 算法描述

(1) 平稳时间序列的定义

满足如下条件的序列称为严平稳序列：

对 \forall 正整数 m ， $\forall t_1, t_2, \dots, t_m \in T$ ， \forall 正整数 τ ，有：

$$F_{t_1, t_2, \dots, t_m}(y_1, y_2, \dots, y_m) = F_{t_1 + \tau, t_2 + \tau, \dots, t_m + \tau}(y_1, y_2, \dots, y_m) \quad (4-5)$$

如果 y_t 满足如下三个条件:

- a. 任取 $t \in T, E y_t^2 < \infty$;
- b. 任取 $t \in T, E y_t = \mu$, μ 为常数;
- c. 任取 $k \in T, E[(y_{t+k} - \mu)(y_t - \mu)] = \gamma_k$;

则称 y_t 为宽平稳时间序列, γ_k 称为自协方差函数 (Auto-covariance Function), 其中:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (4-6)$$

称为自相关函数 (Auto-covariance Function), 简称 ACF。

(2) ARMA 模型

具有如下结构的模型称为自回归滑动平均 (moving average) 模型, 简记为 ARMA(p,q), 如下所示

$$\begin{cases} y_t = \phi_0 + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0, \phi_q \neq 0 \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E y_s \varepsilon_t = 0, \forall s < t \end{cases} \quad (4-7)$$

中心化 ARMA(p,q) 模型, 为

$$\begin{cases} y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0, \phi_q \neq 0 \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E y_s \varepsilon_t = 0, \forall s < t \end{cases} \quad (4-8)$$

缺省默认条件, ARMA(p,q) 模型可以简写为

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (4-9)$$

显然:

当 $q=0$ 时, ARMA(p,q) 模型就退化成了 AR(p) 模型;

当 $p=0$ 时, ARMA(p,q) 模型就退化成了 MA(q) 模型;

所以, AR(p) 模型和 MA(q) 模型实际上是 ARMA(p,q) 模型的特例, 它们都统称为 ARMA 模型。

(3) 数据序列检验

结合已知数据, 笔者首先对数据序列平稳性进行检测, 平稳性的检验方法如下:

a. 时序图检验是指根据平稳时间序列的均值和方差都为常数的性质, 平稳序列的时序图显示该序列值式中在一个常数附近波动, 并且波动的范围有界。倘若有明显的趋势性或周期性, 那他通常不是平稳序列。

b. 自相关图检验: 平稳序列具有短期相关, 说明对平稳序列而言通常近期的序列值对现有值影响较大, 间隔越远的过去值对现有值影响较小。随着延迟期数 k 的增

加，平稳序列的自相关系数会比较快的衰减趋向于 0，在 0 处波动，而非平稳序列则衰减速度较慢，这就是利用自相关图进行平稳性检验的标准。

c.单位根检验。单位根检验是指检验序列中是否存在单位根，如果存在单位根就是非平稳时间序列了。

4.1.2.2 ARMA 模型定阶

ARMA 模型参数定阶是指依据系统的输入输出数据建立模型结构的过程，包括模型的定阶与参数估计。由于客观现象存在不确定性和获取观测数据时受条件与环境的制约，从数据到模型往往要经过多次反复探索，因而必须采取合理的建模过程，才能得到满意的结果，参数识别步骤如图 10。本题中模型识别过程的具体步骤如下：

Step1：平稳性检测。首先画出加拿大各地区的时间序列变化图，观察变化是否稳定，均值是否保持稳定；

Step2：计算自相关系数。若自相关系数长期不等于 0，说明序列具有很强的长期相关性。再检验序列是否非平稳序列。本文使用图检验方法，判断（平方根检验统计量对应的） p 值是否大于 0.05，可判断序列是否非平稳序列。然后进行白噪声检验，（白噪声检验统计量对应的） p 值远小于 0.05，说明该序列属于非白噪声序列。

Step3：如果不是非平稳序列，就对序列进行差分运算，再根据第 1 步进行平稳性验证，不满足就再进行差分运算，直至得到平稳序列。

Step4：针对平稳非白噪声序列，利用 ARMA 模型进行分析。确定 p 、 q 值。

Step5：相对最优模型识别。计算 ARMA(p,q)，当 p 和 q 均小于等于 3 的所有组合的 BIC 信息量，取其中 BIC 信息量达到最小的模型阶数。BIC 矩阵中最小值的索引(a,b), $p=a,q=b$ ，如最小 BIC 值的索引为(0,1)，表示 $p=0,q=1$ ，即 p 、 q 定阶完成。

Step6：综上 n 阶差分后确定的模型为 ARIMA(p,n,q)。

Step7：针对建立好的 ARIMA 模型检验残差(保证为白噪声序列，认为是无偏的)。

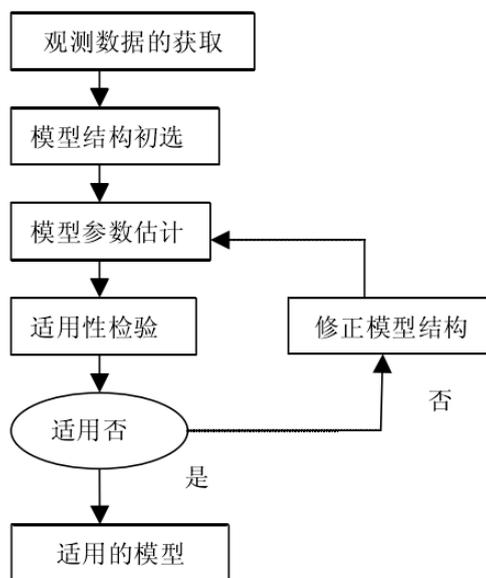


图 4-7 模型参数识别的过程图

4.1.2.3 ARMA 模型求解

结合上述 4.3 可知，加拿大各地气温时间序列为平稳非白噪声序列，符合 ARMA 模型建模要求。因此我们可以采用 ARIMA 模型解题，具体分析步骤如下：

Step1: 求出该观察值序列的样本自相关系数(ACF)和样本偏自相关系数(PACF)的值。

Step2: 根据样本自相关系数和偏自相关系数的性质，选择适当的 ARMA (p,q) 模型进行拟合。

Step3: 估计模型中位置参数的值。

Step4: 检验模型的有效性。如果模型不通过检验，转向 Step2，重新选择模型再拟合。

Step5: 模型优化。如果拟合模型通过检验，仍然转向 Step2，充分考虑各种情况，建立多个拟合模型，从所有通过检验的拟合模型中选择最优模型。

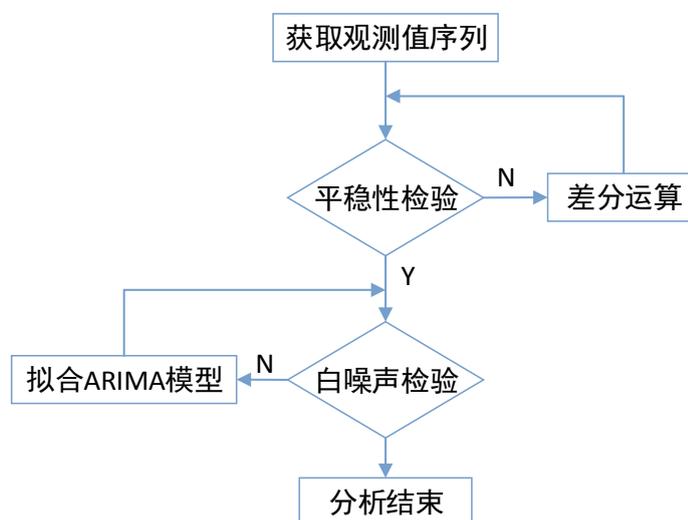


图 4-8 ARMA 模型过程图

4.1.2.4 ARMA 模型分析

(1) 圣约翰年度气温变化趋势分析

首先对数据进行分析，绘制圣约翰气温数据图，如图 4-9 所示。由图 4-9 可以看出，数据具有周期性，且周期为 12 个月，图中黑色点表示该地区 2010-2019 年间日最高气温，蓝色点为该地区日最低气温，橙色点为日平均气温。

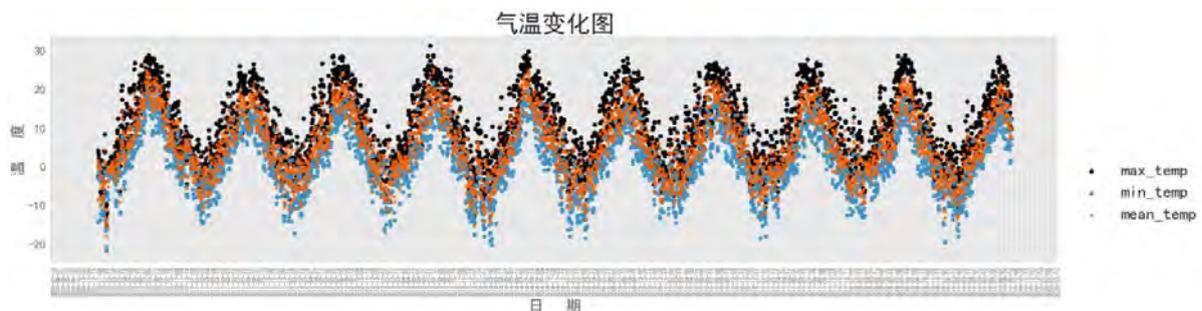


图 4-9 2010-2019 年圣约翰年度气温序列变化图

通过对时间序列数据进行分析发现其为非平稳数据，为了避免“伪回归”问题先对数据进行差分处理，具体去除周期性的方法是用本次数据和相差为周期数的上期数据相减，作为新的数据值。公式为 $y_t = x_t - x_{t-12}$ 。由于时间序列具有增长趋势，选择差分运算。由于 ARMA 模型只能分析 0-1 均值化的时间序列，而计算出序列的均值为-0.0041，表明对序列的影响很小，故忽略不计，认为数据是均值化的。根据 ARMA 模型的自相关和偏相关分析法编写程序，确定模型的阶数，进行模型识别，经过二阶差分运算后，圣约翰气温序列趋于平稳序列，结果如图 4-10 所示。

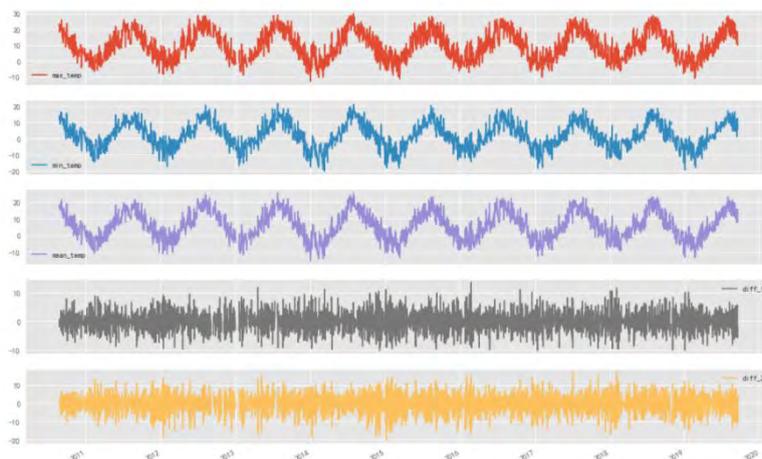


图 4-10 2010-2019 年圣约翰气温差分运算后转化为平稳序列数据图
自相关函数和偏自相关函数分析见图 4-11。

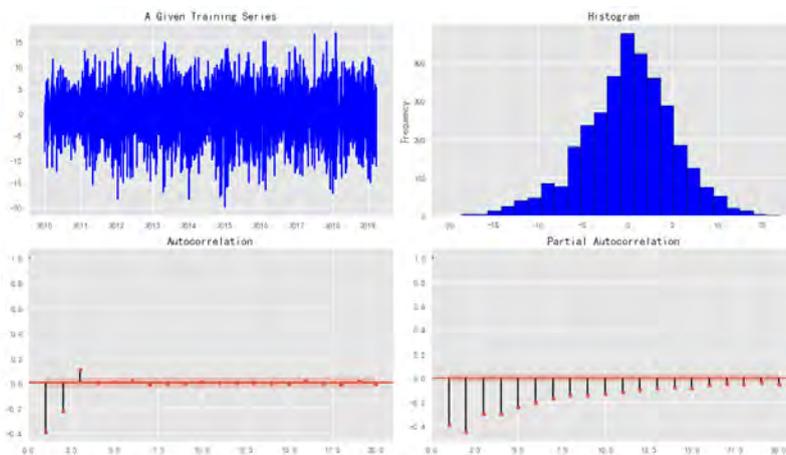


图 4-11 平稳序列的自相关系数和偏自相关系数

经过自相关图和偏自相关图分析，可见序列的自相关函数和偏自相关函数都拖尾，所以尝试用 ARIMA(1,2,1)模型进行拟合。最后使用赤池信息准则 AIC 和贝叶斯信息准则 BIC 对 ARIMA 模型模型进行评估,评估结果见图 4-12。



图 4-12 BIC 模型评估图

(2) 多伦多年度气温变化趋势分析

如图 4-13 为多伦多地区年度气温序列散点图。图中黑色点表示该地区 2010-2019 年间日最高气温，蓝色点为该地区日最低气温，橙色点为日平均气温。

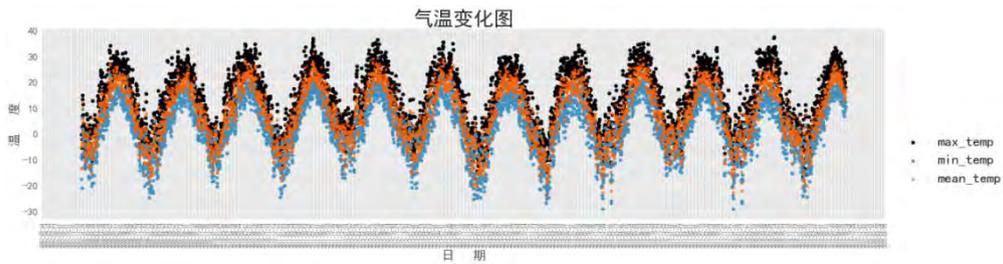


图 4-13 2010-2019 年多伦多年度气温序列变化图

根据 ARMA 模型的自相关和偏相关分析法编写程序，确定模型的阶数，进行模型识别，如图 4-14 所示，经过二阶差分运算后，多伦多气温序列趋于平稳序列。经过自相关图和偏自相关图分析，可见序列的自相关函数和偏自相关函数都拖尾，所以尝试用 ARIMA(2,2,1)模型进行拟合。最后使用赤池信息准则 AIC 和贝叶斯信息准则 BIC 对 ARIMA 模型模型进行评估。

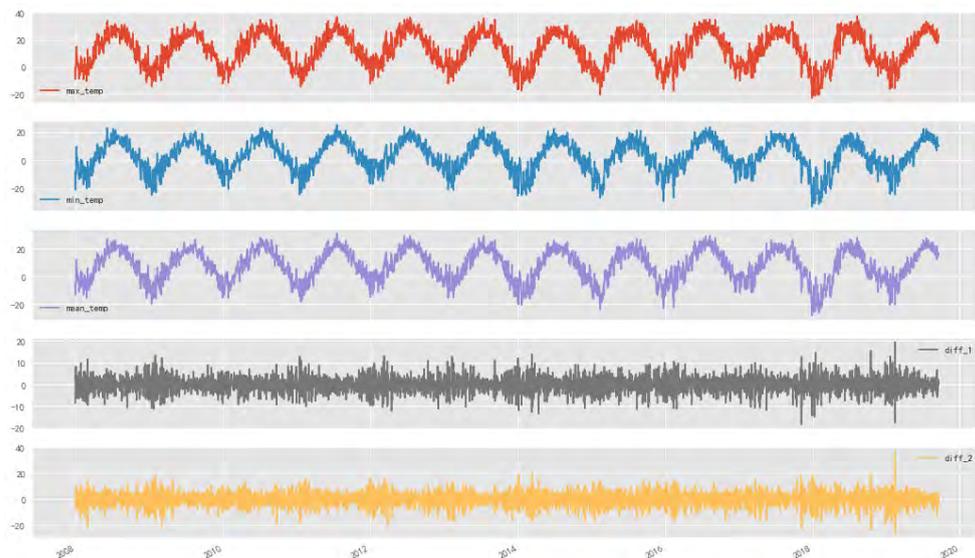


图 4-14 2010-2019 年多伦多气温差分运算后转化为平稳序列数据图

(3) 埃德蒙顿年度气温变化趋势分析

如图 4-15 为埃德蒙顿地区年度气温序列散点图。图中黑色点表示该地区 2010-2019 年间日最高气温，蓝色点为该地区日最低气温，橙色点为日平均气温。

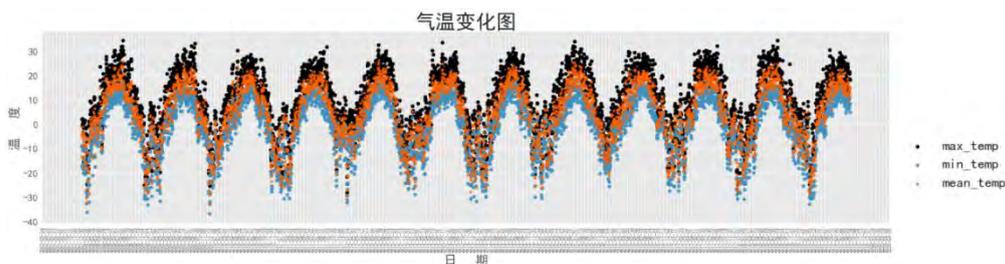


图 4-15 2010-2019 年埃德蒙顿年度气温序列变化图

根据 ARMA 模型的自相关和偏相关分析法编写程序，确定模型的阶数，进行模型识别，如图 4-15 所示，经过二阶差分运算后，埃德蒙顿气温序列趋于平稳序列。如图 4-16，经过自相关图和偏自相关图分析，可见序列的自相关函数和偏自相关函数都拖尾，所以尝试用 ARIMA(1,2,2)模型进行拟合。最后使用赤池信息准则 AIC 和贝叶斯信息准则 BIC 对 ARIMA 模型模型进行评估。

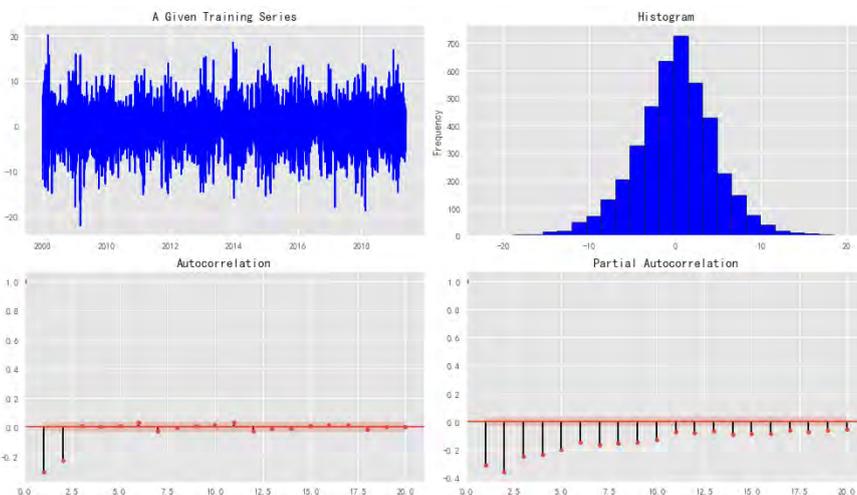


图 4-16 平稳序列的自相关系数和偏自相关系数

(4) 维多利亚年度气温变化趋势分析

如图 4-17 为埃德蒙顿地区年度气温序列散点图。图中黑色点表示该地区 2010-2019 年间日最高气温，蓝色点为该地区日最低气温，橙色点为日平均气温。

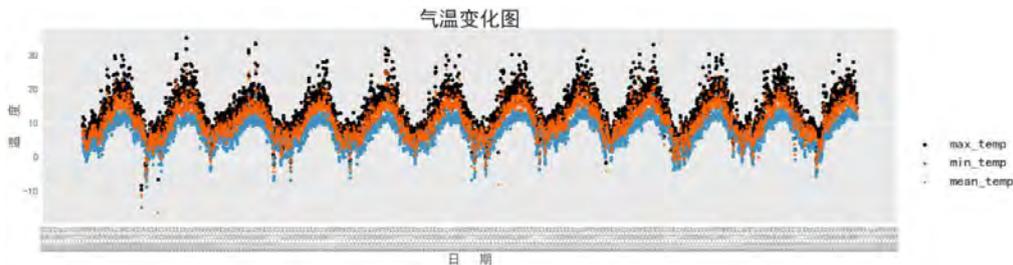


图 4-17 2010-2019 年维多利亚年度气温序列变化图

根据 ARMA 模型的自相关和偏相关分析法编写程序，确定模型的阶数，进行模型识别，如图 4-17 所示，经过二阶差分运算后，如图 4-18，维多利亚气温序列趋于平稳序列。经过自相关图和偏自相关图分析，可见序列的自相关函数和偏自相关函数都拖尾，所以尝试用 ARIMA(3,2,1)模型进行拟合。最后使用赤池信息准则 AIC 和贝叶斯信息准则 BIC 对 ARIMA 模型模型进行评估。

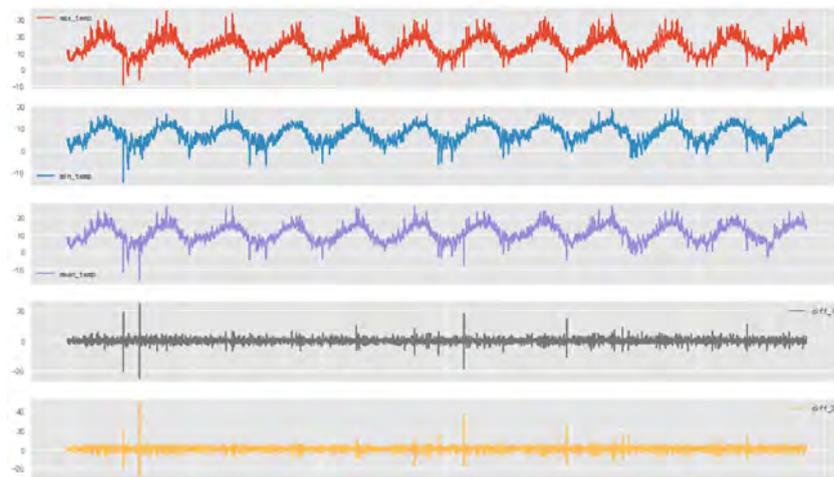


图 4-18 2010-2019 年维多利亚气温差分运算后转化为平稳序列数据图

(5) 耶罗纳夫年度气温变化趋势分析

如图 4-19 为耶罗纳夫地区年度气温序列散点图。图中黑色点表示该地区 2010-2019 年间日最高气温，蓝色点为该地区日最低气温，橙色点为日平均气温。

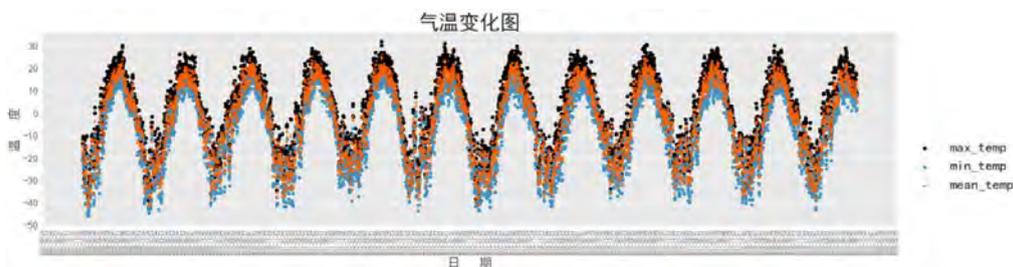


图 4-19 2010-2019 年耶罗纳夫年度气温序列变化图

根据 ARMA 模型的自相关和偏相关分析法编写程序，确定模型的阶数，进行模型识别，如图 4-19 所示，经过二阶差分运算后，耶罗纳夫气温序列趋于平稳序列。如图 4-20，经过自相关图和偏自相关图分析，可见序列的自相关函数和偏自相关函

数都拖尾，所以尝试用 ARIMA(2,2,3)模型进行拟合。最后使用赤池信息准则 AIC 和贝叶斯信息准则 BIC 对 ARIMA 模型模型进行评估。

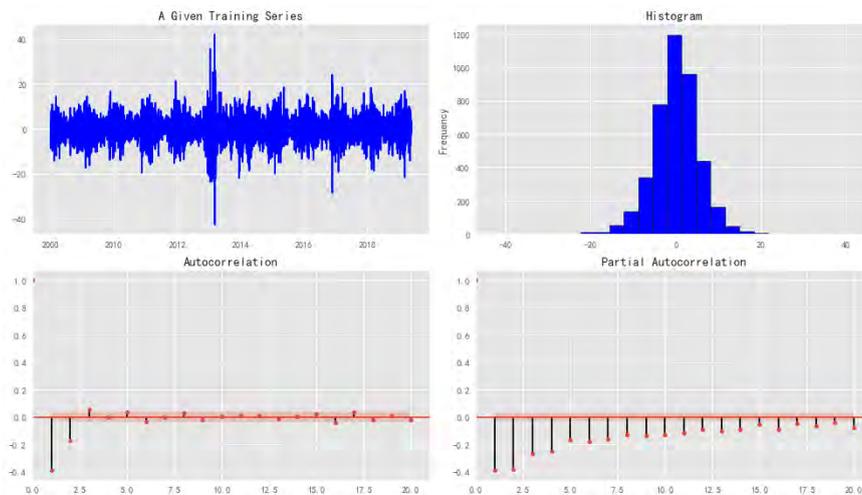


图 4-20 平稳序列的自相关系数和偏自相关系数

4.1.3 方法三：定性分析二

经查询和收集网上相关信息以及查阅相关文献，我们发现影响加拿大的气候因素主要包括太阳辐射、下垫面状况、纬度位置、洋流因素、大气环流、人类活动等，各因素间互相影响，而且这些因素变化的趋势不一致，变化的幅度也不一致，所以在评价各气候因子对加拿大气温的影响时，需要分析各气候因子对加拿大气温的重要性程度，即确定各气候因子的权重。本文首先在分析过去已发生的白灾事件基础上，分析出评价加拿大气温的气候指标体系，并分析各气候因子的变化趋势，然后利用熵权法确定各因子的权重，从而建立模型评价气候变化对研究区气温的影响。

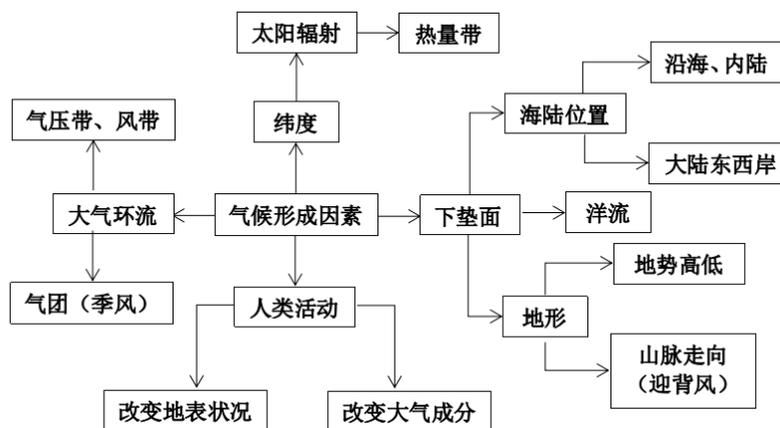


图 4-21 影响气候变化因素

4.1.3.1 熵权法模型建立

权系数的确定有多种方法，通常可分为两类：主观法和客观法。前者是由决策者根据主观评断来确定权重系数的一类方法，常用的有二项系数加权法、循环打分法、层次分析法等；后者则是根据指标值反映的客观信息确定权系数的一类方法，常用的有均方差法、最优权法、熵权法等。其中熵权法可以在决策者预先确立的偏好系数基础上，将决策者的主观判断和指标客观信息结合起来，可以得到更加精确的估计。

熵(entropy)的概念源于热力学, 后由香农 (Shannon) 引入信息论, 其基本思想是根据指标的信息熵大小来确定评分的权重系数, 根据信息论, 如果某项指标的信息熵越小, 则该指标的变异程度越大, 携带的信息也就越重要, 反之则说明该指标的变异程度小, 能够提供的信息量小。因此对于提供信息量小的指标赋予小的权重, 该指标得分在最终的综合评价中影响也就越小。

利用熵权法确定各指标的权数步骤如下:

Step1: 构建各因素各评价指标的判断矩阵;

Step2: 将判断矩阵进行归一化处理, 得到归一化判断矩阵:

$$R = (r_{ij})_{m \times n} \quad (4-10)$$

Step3: 由标准化后的决策矩阵计算 p_{ij} , 其中:

$$p_{ij} = \frac{r_{ij}}{\sum_{i=1}^m r_{ij}} \quad (i=1,2,\dots,m; j=1,2,\dots,n) \quad (4-11)$$

Step4: 求第 j 个指标输出的熵:

$$E_j = -K \sum_{i=1}^m P_{ij} \ln P_{ij} \quad (j=1,2,\dots,n; K = (\ln m) - 1) \quad (4-12)$$

由于 $0 \leq P_i \leq 1$, 所以 $0 \leq E_j = K \sum_{i=1}^m P_{ij} \ln P_{ij} \leq \ln m$, 由此可知: $0 \leq E_j \leq 1 \quad j=1,2,\dots,n$;

Step5: 第 j 个指标的熵权 d_j 定义为:

$$d_j = (1 - E_j) / \left(n - \sum_{j=1}^n E_j \right) \quad (4-13)$$

Step6: 求各指标的综合熵权。假设决策者已有一偏好的权系数 η_j , 则按下式确定各指标的综合熵权:

$$w_j = \lambda_j d_j / \sum_{j=1}^n \lambda_j d_j \quad (4-14)$$

Step7: 计算系统的权重值, 得出结论。

4.1.3.2 模型求解

本题选择的研究区域是加拿大, 地处北纬 $41^\circ\text{C} \sim 83^\circ\text{C}$ 、西经 $52^\circ\text{C} \sim 141^\circ\text{C}$ 之间。研究区因受西风影响, 大部分地区属大陆性温带针叶林气候。东部气温稍低, 南部气候适中, 西部气候温和湿润, 北部为寒带苔原气候。北极群岛终年严寒。中西部最高气温达 40°C 以上, 北部最低气温低至 -60°C 。加拿大领土面积为 998.467 万平方公里, 人口主要集中在南部五大湖沿岸, 是一个地广人稀的国家。

2020-2050预计总降水量变化

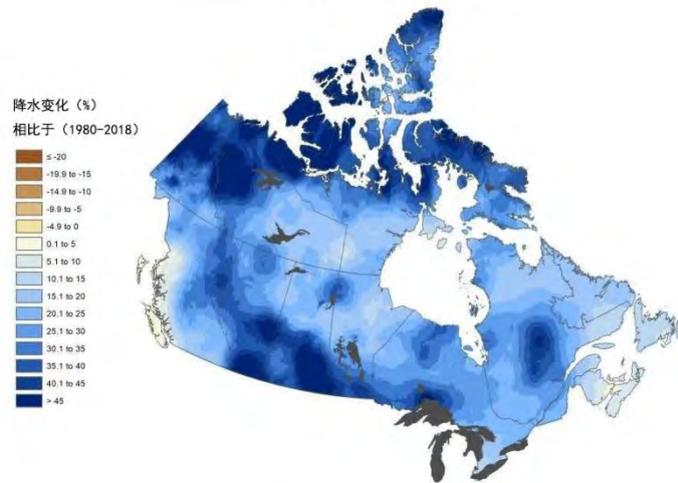


图 4-22 预计未来 30 年加拿大预计降水量变化图

本题利用国家气候中心提供的地面气象资料月报，选取加拿大气象站 1970—2019 年 50 年间的月降水资料；1989—2019 年 30 年间逐日的气温、风速、积雪数据，同时利用世界气象组织提供的 1980—2018 年 38 年间逐日的气温、降水、风速、积雪数据，如图。根据选取的评价指标，碳排放设为 X_1 ；积雪持续时间设为 X_2 ；低温且容易负积温设为 X_3 ；五日温差 $\geq 12^\circ\text{C}$ 日数设为 X_4 ；积雪期大风日数 X_5 ，降水量设为 X_6 ，得到决策矩阵 $X=(X_1 X_2 X_3 X_4 X_5 X_6)$ 。

首先按照公式对决策矩阵进行标准化，然后按照熵权法步骤计算得出各指标熵权 d_j ，见表 1。通过咨询 10 位专家意见，给出两两指标之间相对重要性程度，采用相对比较法[2]，确定较为主观的偏好系数 η_j ，见表 1。最终得到各指标的综合权重 w_j ，见表 1。

表 1 评价加拿大的气候指标权重计算结果

	碳排放	积雪持续 时间	低温且容 易负积温	五日温差 \geq 12 $^\circ\text{C}$ 日数	积雪期大风 日数	降水量
E_j	0.939	0.925	0.967	0.944	0.887	0.962
d_j	0.163	0.198	0.091	0.147	0.300	0.102
η_j	0.214	0.214	0.203	0.124	0.145	0.101
w_j	0.21	0.25	0.11	0.11	0.26	0.06

如前节分析，加拿大大部分地区处于极地，季节性温度变化范围相当大，冬天长期严寒、夏天短暂、春季和秋季极短。低温且容易负积温增加有利于评估加拿大气温变化，权重应为正值；而碳排放和积雪期大风日数增多有加重数据拟合的趋势，其权重应赋为负值，得 $W'=(w_1 w_3 w_5)=(-0.21 \ 0.11 \ -0.26)$ ，计算综合评价值：

$T = B'W'^T = 1.43$ ， $T > 0$ 说明在碳排放、容易负积温和积雪期大风日数综合作用下，加

拿大天气随现有气候变化有明显加重趋势。

4.2 第一问结论

本题在分析近几十年加拿大天气变化的基础上，结合专家评价法发现加拿大的气温 1948 年以来已经升高了 1.7°C ，是同期全球平均增温 0.8°C 的两倍以上，得出了评价加拿大的气候指标体系，并在分析各指标的变化趋势的基础上，引入了熵权法确定各指标的权重，从而建立了加拿大地区随气候变化评价方法，最后得出如下结论：

1. 北极地区是加拿大最寒冷的地区，平均气温 -5°C - 6°C [3]，一年中气温在冰点以上的时间只有几个月；

2. 北方地区由草原省份的北部以及安大略省、魁北克省和大西洋沿岸省份的大部分组成，气候特点是冬季漫长且凛冽，夏天仅两个月左右，降雨量适中，大多在夏季；

3. 草原地区属大陆性气候，四季分明，冬天寒冷，而夏天可能炎热异常；

4. 山脉地区，天气多变，大部分时间风雪交加，气温随着地势增高而下降；

5. 太平洋沿岸地区包括不列颠哥伦比亚的沿岸地区以及该省的沿岸岛屿，气候温和宜人，常年吹西风，夏季凉爽、干燥，冬季多云、潮湿却暖和，降雨丰沛，年降雨量为 2000-3000 厘米，各个季节降水量比较均匀。其中加拿大最温暖的的城市温哥华就位于这个地区，年平均气温 9°C - 8°C ；

6. 东南部地区，冬暖夏凉，属于加拿大人口最多的地区。

在全球变暖的大背景下，由于人类活动导致二氧化碳的过量排放，加强了冰川和海冰消融的反馈回路——变暖之后冰雪消融，裸露出的深色地表和海面吸收了更多的太阳辐射，使温度进一步升高，因此又进一步加速冰雪消融，让更多的地表和海面裸露出来，使得加拿大的变暖速度显著高于全球平均水平。

4.3 第二问问题分析

海洋系统是一个动态的复杂系统，在演化过程中，海洋系统动力学特征往往呈现出复杂性和非线性。海洋表面温度(SST)是一个重要的海洋环境参数[4]，几乎所有的海洋过程，特别是海洋动力过程都直接或间接地与温度有关。SST 时间序列具有明显的季节性，长记忆性行为是复杂系统内在机制的综合表现，是探索时空耦合系统的深层动力机制的有效工具。因此对 SST 时间序列的复杂性及相关性进行研究，对揭示海洋的规律和特性具有非常重大的意义。

对于第二问从海洋表面温度历史数据探寻规律，可以利用复杂网络理论构建了全球海洋气候的非线性和线性复杂网络模型，仿真对比了两者的拓扑性质。再利用度分布、聚类系数和介数等统计特征可视化海洋不同海域的能量传递等现象，分析和探索海洋气候的系统稳定性和季节性差异等规律。

如图 4-23 所示，本题的解题步骤分四步：

步骤一：获取数据。利用网络爬虫从国家海洋信息中心上爬取 2019 年的全球日均海表面温度数据[5]。

步骤二：数据预处理。将全球海洋进行固定尺度网格划分，均值化每个网格内的海表面温度。原始数据的大小为 3600×7200 个像素点，首先根据经纬度将全球划分为 90×180 个网格，每个网格由 40×40 个像素点组成，球面网格经纬度的分辨率均为 2° 。除去噪声数据，最终网络的有效节点数为 11689。

步骤三：复杂网络建模。根据相似性方法来度量全球海洋不同海域之间的相似度，得到全球相似度矩阵，然后设定不同的阈值来构建网络，并做详细的拓扑分析。

步骤四：仿真实验与可视化分析。基于互信息和皮尔逊相关系数的网络节点度可视化 SST 数据在地球上的离散分布情况。

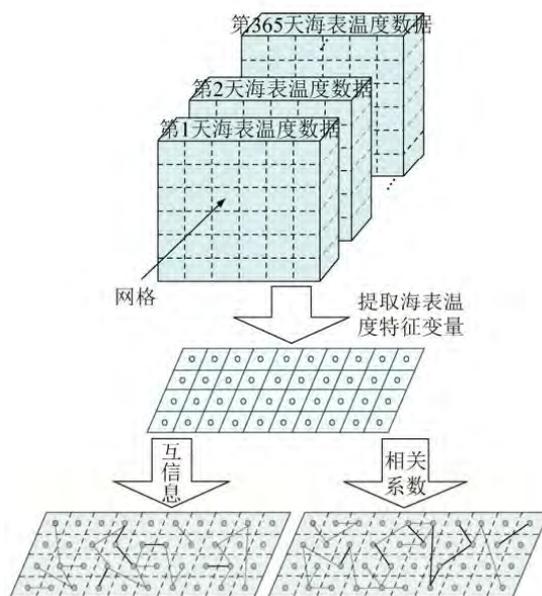


图 4-23 基于海洋表面温度历史数据的复杂网络建模

4.4 第二问模型建立

4.4.1 模型假设

假设 1：考虑到实际情况，假设海洋构建为无向无权的简单网络，这种网络的任何节点之间最多只能有一条边，并且不存在自环路。

假设 2：海洋网络中的边表征了不同海域之间温度变化在时间上的相似程度。

假设 3：海洋网络的节点代表海洋上的某块区域。

4.4.2 网络拓扑结构

网络拓扑可以定义为一个三元组 $G = (V, E, F)$ ，其中集合 $V = (V_1, V_2, \dots, V_M)$ 称为节点集， $E = (e_1, e_2, \dots, e_N)$ 称为边集， F 是边集 E 到节点集 V 的一个映射。节点是网络中最基本、最重要的组成元素，不同性质的网络中节点的含义也不同，海洋网络的节点一般表示海洋上的某块区域。

4.4.3 相似性度量

(1) 互信息

互信息是指一个随机变量由于已知另一个随机变量而减少的不肯定性。它是信息论里一种有用的信息度量。考虑两个随机变量 X 和 Y ，它们的联合概率密度函

数为 $p(x,y)$,其边缘概率密度函数分别是 $p(x)$ 和 $p(y)$ 。互信息 $I(X;Y)$ 为联合分布 $p(x,y)$ 与乘积分布 $p(x)p(y)$ 之间的相对熵,即

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (4-15)$$

从互信息的定义中可以看到,当变量 X 和 Y 完全统计独立时,它们的互信息为零,这意味着两个变量之间不存在共同的信息;反之,当两个变量之间依赖程度越高时,互信息 $I(X;Y)$ 的值就越大,所共享的信息含量越大。在互信息的基础上,可以得到归一化互信息,两者转换关系如下:

$$NMI(X;Y) = \frac{2 \times I(X;Y)}{H(X) + H(Y)} \quad (4-16)$$

(2) 皮尔逊相关系数

皮尔逊相关系数是气候网络构建中度量相似度的常用方法[6,14]。变量 A 、 B 之间的皮尔逊相关系数[6]定义为两个变量之间的协方差和标准差的商。相关系数的计算公式如下:

$$\rho_{A,B} = \frac{\text{cov}(A,B)}{\sigma_A \sigma_B} = \frac{E[(A - \mu_A)(B - \mu_B)]}{\sigma_A \sigma_B} \quad (4-17)$$

基于样本对协方差和方差进行估计,可以得到样本的皮尔逊相关系数,一般表示成 r ,公式为:

$$r(A;B) = \frac{\sum_{i=1}^t (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^t (a_i - \bar{a})^2 \sum_{i=1}^t (b_i - \bar{b})^2}} \quad (4-18)$$

(4-17)式中: a_i 代表变量 A 中第 a 个元素值; \bar{a} 为整个时序变量中所有元素的均值。从公式中可以得到皮尔逊相关系数的取值范围为 $(-1, 1)$,当两个变量的线性关系增强时,值趋于 1 或 -1 。当一个变量随另一个变量的增大而增大时,表明它们之间是正相关的,相关系数大于 0 ;反之,表明它们之间是负相关的相关系数小于 0 ;若相关系数等于 0 的话,表明它们之间不存在线性相关关系。考虑到负相关同样意味着两个变量存在线性相关,因此我们取 r 的绝对值作为衡量相似度大小的标准[3]。

(3) 介数

假设信息或能量在从节点 V_i 向非邻居节点 V_t 传播的过程中,总是沿着这两个节点之间的最短路径传播。如果某个中间节点 V_j 被许多条最短路径穿过,则该节点在传播过程中就很重要。介数常用来表示这种特性,它反映了节点在整个动态过程中的控制力。节点 V_i 的介数定义为:

$$BC(i) = \sum_{i \neq s, i \neq t, s \neq t} \frac{l_{st}^i}{l_{st}} \quad (4-19)$$

式中: l_{st} 为从节点 V_s 到节点 V_t 的所有最短路径的数目; l_{st}^i 是从节点 V_s 到节点

V_i 的所有最短路径中经过 V_i 的最短路径的数目。介数有助于发现网络中信息和能量传递的显著性特征，因此其应用比较广泛。

4.5 第二问模型结论

本题根据前述相似性方法来度量全球海洋不同海域之间的相似度，得到了大小为 11789×11789 的全球相似度矩阵，然后设定阈值作为保留或移除边的依据。为了深入研究海表温度网络，通过运用互信息和皮尔逊相关系数两种不同的度量方式来分析不同海域 SST 的相关性，基于复杂网络的角度构建海表温度网络拓扑模型，通过设定不同的阈值进行拓扑建模，最后可视化分析解释海洋规律。

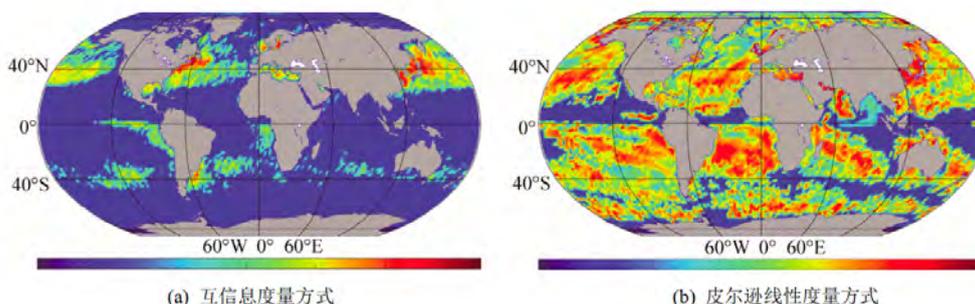


图 4-24 节点介数在地理上的离散分布情况

如图 4-24 所示，通过对比在不同边密度下的节点度分布，分析了在非线性和线性度量方式下两种海洋网络的拓扑特征，发现虽然不同度量方式下所构建的网络结构差异比较明显，但是对于某一特定类型的网络而言，网络的静态特征在地理上的分布并没有较大的改变，从而体现了海表温度环境的稳定性。由于海洋节点的度分布呈现出幂律特性，因此说明海洋系统具有脆弱性。如图 4-25 可视化了不同季节之间节点介数在地理上的离散分布情况，从可视化介数的地理分布图中可以看到介数分布的变化在一定程度上体现了能量和热量随季风传播的特点。

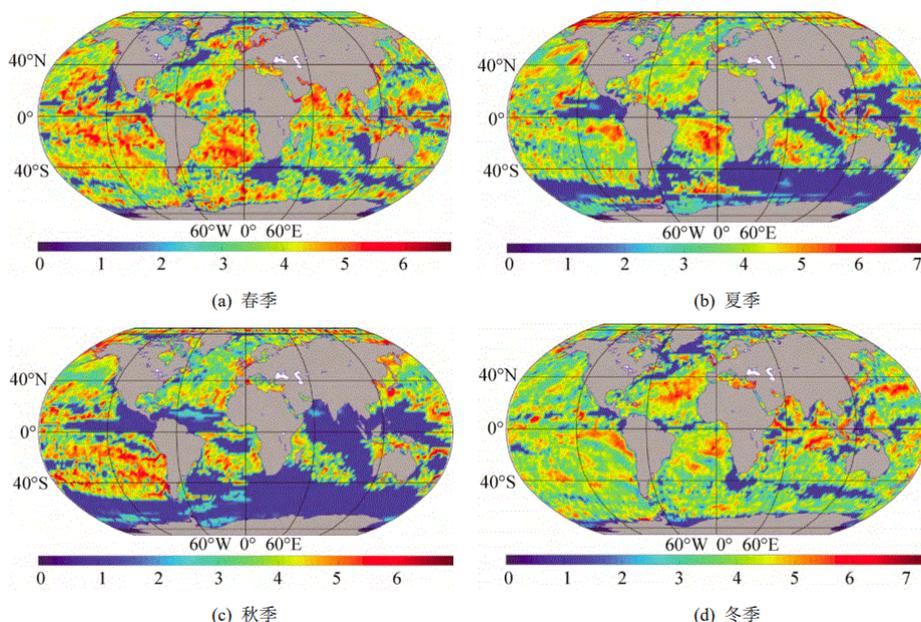


图 4-25 季节性网络的节点介数在地理上的离散分布情况

5 问题二建模与求解

5.1 问题分析

近百年来全球气候正经历一次以变暖为主要特征的显著变化，全球变暖幅度自 1990 年以来明显加速，1880 年至 2012 年间，全球地表温度大约升高了 0.85°C ，全球平均气温的改变不仅对生态和环境有重大影响，且涉及到社会经济与社会生产等诸多方面。因此准确预测未来全球平均气温变化情况引起了各国政府以及国际机构的普遍关注。

对于问题二，可以分为两步来进行操作，首先利用基于 LSTM-ARIMA 的气温组合预测模型对加拿大地区各地的气温变化进行分析，其次，借助以上分析以及建立相应的模型对全球未来 25 年的气候变化进行预测。

本题的思路可以分为六步：

步骤一：获取所需的气温变化数据，对数据进行基本统计和可视化展示；

步骤二：对数据集的影响因素进行特征提取，并将数据分为训练集和测试集两部分；

步骤三：根据问题一得到的分析结果，对训练集数据采用五折交叉验证，以避免过拟合现象。将特征提取出的碳排放量、地球散热、降水、大气环流以及海洋温度变化这 5 个因素作为输入变量（输入特征），当前时刻 t 的温度作为输出变量（标签）；

步骤四：建立 LSTM 模型和 ARIMA 模型，先利用 ARIMA 模型对数据集进行预测，再利用 LSTM 网络依次对 ARIMA 模型的残差序列进行预测；

步骤五：将 ARIMA 模型预测结果与 LSTM 对残差的预测结果进行相加，调整参数，得到 LSTM-ARIMA 组合模型的最终预测结果。

步骤六：综上，对组合模型进行评价。

5.2 模型建立

统计学和机器学习的预测方法主要有三类，一类是依据和响应变量相关的协变量预测响应变量。主要有多元回归法、逐步回归法等，机器学习的方法主要有各类神经网络、SVM 等；第二类是利用长时间序列前期观测值建立复杂的预测模型对序列未来数值进行预测[8]，包括 AR 模型、ARMA 模型等；三类是组合预测法。针对不同的实际问题，将不同种预测方法加以组合规避单一模型对数据拟合效果不佳的后果以达到更高的预测精度。ARMA 模型[9]是最经典、最常用于温度预测的一种时间序列模型，但预测精度不够理想，故本文采用组合预测法构建基于 LSTM-ARIMA 的气温组合预测模型对全球未来气温进行预测，使模型预测准确率提高。

通过以上对问题二的分析，我们提出一个可以预测气温变化的基于机器学习的时间序列预测分析模型，对加拿大乃至全球未来 25 年的气温变化做出合理预测。优化目标包括：结合地球吸热、散热以及海洋温度变化等因素，选取最优的预测模型，使准确率和召回率尽可能最高。

5.2.1 LSTM 模型

时间序列预测分析就是利用过去一段时间内某事件时间的特征来预测未来一段时间内该事件的特征，通常使用 RNN-LSTM 模型来进行分析。RNN 模型是指递归

神经网络模型，可以从先前输入数据的整个历史映射到每个输出，故有利于处理顺序数据，简单的 RNN 构架图如图 5-1 所示。图中的每一个节点都表示网络单元层，其中有三个重要的连接：输入到隐藏层，隐藏到隐藏层，隐藏层到输出层。

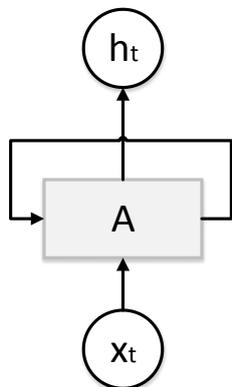


图 5-1 RNN 模型

LSTM 模型是 RNN 模型的衍生，相较 RNN 模型更为复杂。LSTM 模型的主要贡献是存储单元，其包含在步骤中获得的观察的时间步处的信息，计算方式可以表示为：

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (5-1)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (5-2)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (5-3)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (5-4)$$

$$h_t = o_t \tanh(c_t) \quad (5-5)$$

其中， f ， i ， o 和 c 分别是忘记门，输入门，输出门和 cell 状态激活向量，这些向量与隐藏输出向量 h 的大小相同， σ 是 sigmoid 激活函数， \tanh 为双曲正切函数， W 是权重矩阵，例如： W_{xf} 是忘记门权重矩阵， W_{xo} 是输出门权重矩阵。

本文对于时空特性中的时间采用了 RNN-LSTM 模型来解决。在这个模型当中每个存储块包含至少一个自连接的累加器单元和若干乘法单元，例如输入门遗忘门和输出门。这三个门允许通过分配诸如 0 和 1 的计数器来存储和访问信息 LSTM 的存储块结构如图 5-2 所示。

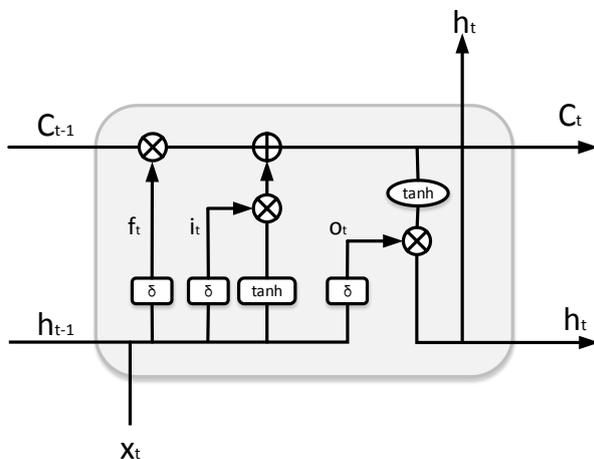


图 5-2 RNN-LSTM 模型

5.2.2 ARIMA 模型

ARIMA(p,d,q)模型叫差分自回归移动平均模型，其中 AR 是自回归，p 是自回归项，MA 是移动平均，q 为移动平均项，d 为时间序列成为平稳时所需做的差分次数（阶数）。ARIMA 模型是指将非平稳时间序列转化为平稳时间序列，然后将因变量仅对它的滞后值以及随机误差项的现值和滞后值进行回归所建立的模型。ARIMA 模型的通用表达式为：

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (5-6)$$

其中 y_t 为时间序列 y 的当期值， y_{t-1} 为 y_t 前一期的值， y_{t-2} 则为 y_{t-1} 前一期的值，依次类推。 $\phi_1, \phi_2, \phi_3, \dots, \phi_p$ 是自回归系数，p 是自回归阶数， $\theta_1, \theta_2, \theta_3, \dots, \theta_q$ 是移动平均系数，q 是移动平均阶数， $\{\varepsilon_t\}$ 是白噪声序列。

5.3 模型求解

5.3.1 数据准备

通过查找全球近十年平均气温统计数据，接入百度地图 API 接口，制作了全球气温热力图，来显示全球平均气温变化情况，如图 5-3 所示：

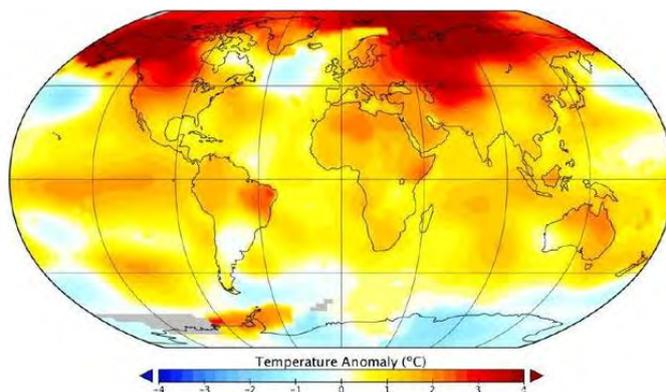


图 5-3 全球平均气温热力图

利用 Python 对数据进行分析 and 可视化展示，按平均温度给国家分类，并绘制水平主图得到如图 5-4 所示，其中俄罗斯和加拿大都有平均温度较低的值，最低温度出现在格陵兰地区，最热的国家自然在非洲赤道区域。

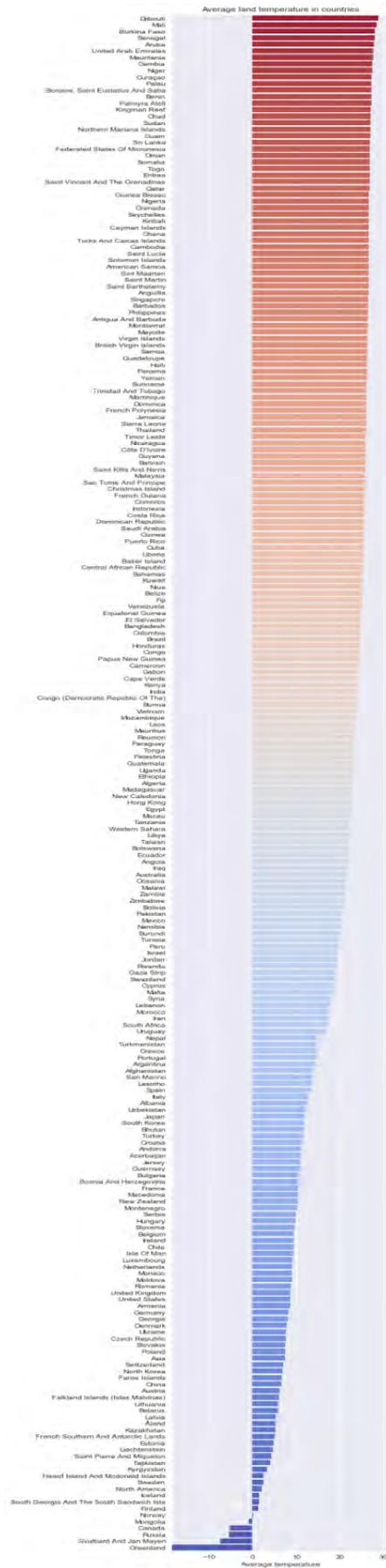


图 5-4 全球各城市平均气温

5.3.2 特征提取

影响全球气温变化的相关因素有很多，表 5-1 中，A-G 表示全球平均气温变化的相关影响因素，为了提高预测的准确性本文将获得的数据通过灰度关联度分析，得出各影响因素与全球气温变化的关系如图 5-5 所示。本文选取了关联度大于 0.6 的影响因素作为预测的特征项，分别为 A-E，剔除了 F-G 两个与全球平均气温变化关联度小于 0.6 的因素。

表 5-1 全球平均气温变化的相关影响因素

影响因素符号	影响因素
A	地球散热
B	碳排放量
C	降水
D	海洋温度变化
E	大气环流
F	地球吸热
G	人类活动

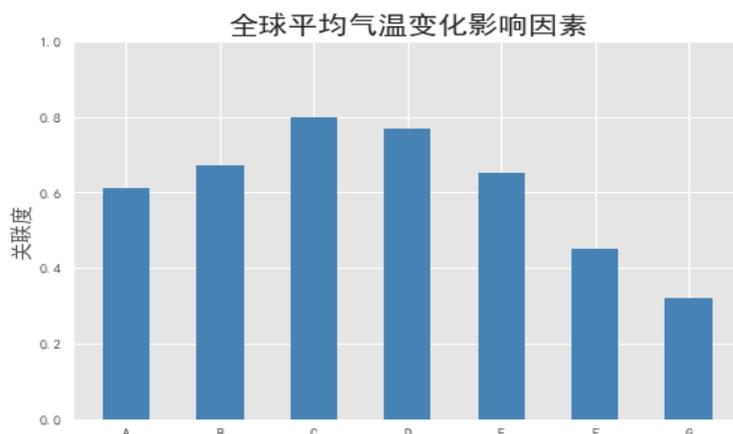


图 5-5 全球平均气温变化影响因素关联分析图

5.3.3 模型训练

首先基于上一节内容对全球气温的影响因素进行特征提取，实验数据选取每天为时间间隔，为期十年的历史记录，其中的 90%作为训练数据，10%作为测试数据，并进行归一化处理。

(1) ARIMA 模型的建立过程

①序列平稳性的检测

本文选取了 2008-2018 年间全球平均气温变化数据集，通过构建时间序列折线图和 ADF 单位根检验进行平稳性测试。假设数据存在单位根，即不平稳；且当显著性水平为 1%时严格拒绝原假设，为 5%时拒绝原假设，10%类推。通过查看 P 值和显著性水平 α 的大小对显著性水平做出评价，p 值越小，小于显著性水平的话，就拒绝原假设，认为序列是平稳的；大于的话，不能拒绝，认为是不平稳的。结果如图 5-6 所示，由图可看出全球平均气温变化的序列折线图趋于平稳并且 ADF 检测得到的检

测统计量 $p\text{-value}=0.0003$, 远小于 0.05 。说明全球平均气温变化时间序列是平稳序列, 无需进行差分变换。

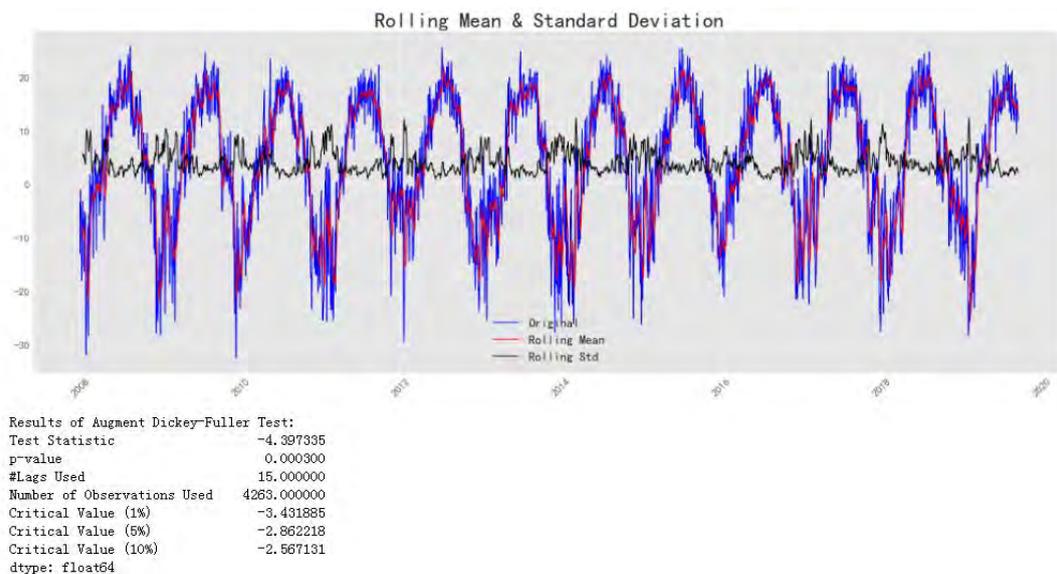


图 5-6 全球平均气温平稳性检测图

对上述平稳序列进行移动平滑窗口检验, 探索变量之间的相关性如图 5-7 所示。

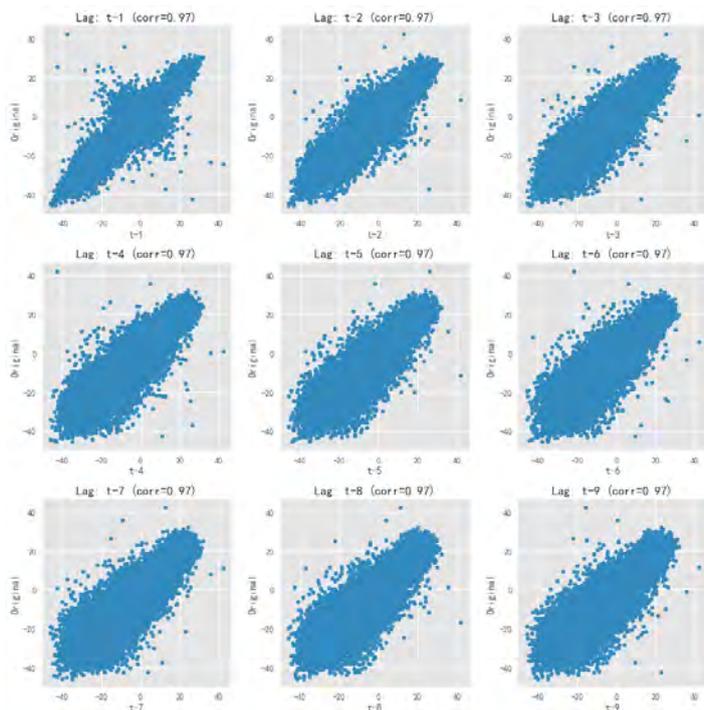


图 5-7 全球平均气温相关性检测

②参数估计及模型检验

通过条件平方和的方法找到初始值, 然后用最大似然法进行各个阶数的系数估计, 因为原始训练数据集较多, 我们引入信息准则, 通过加入模型复杂度的惩罚项来避免过拟合问题, 本文中采用以下两个模型选择方法——赤池信息准则 (AIC) 和贝叶斯信息准则 (BIC)。选取使 AIC、BIC 最小的参数对模型进行定阶, 结果如图 5-8 所示。并利用 ARIMA 模型对全球平均气温数据进行预测。



图 5-8 ARIMA 定阶

(2) LSTM 模型的建立过程

先利用 ARIMA 模型对时间序列进行预测,再利用 LSTM 网络依次对 ARIMA 模型的残差序列进行预测。LSTM 预测模型主要分为三层,分别是输入层、隐藏层和输出层。网络的输入层是将训练数据中碳排放量、地球散热、降水、海洋温度变化、大气环流这 5 个影响因素的特征量和全球平均气温作为输入,记作 X,将 X 输入隐含层,隐含层由 L 个连接的同构 LSTM 细胞构成,经过隐含层后,网络的输出为 Y。网络模型结构为单层的 LSTM 层,损失函数 loss 使用均方误差函数[10],结果如图 5-9 所示,优化器选择 Adam 算法,模型批量处理长度为 300。

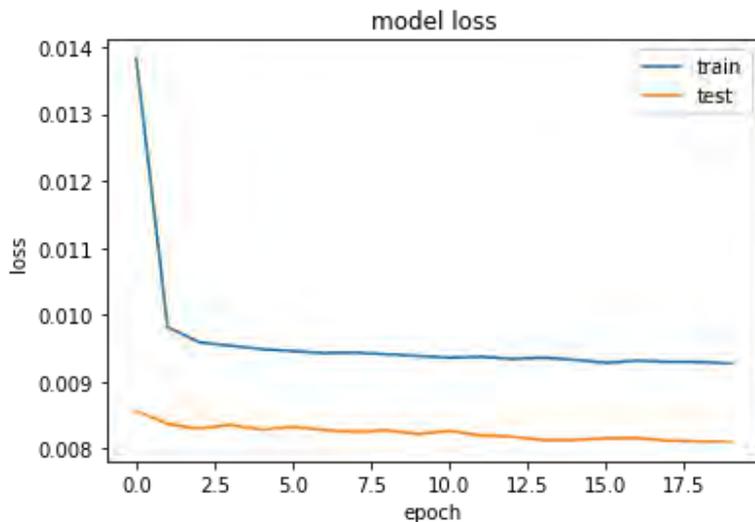


图 5-9 损失函数

(3) LSTM-ARIMA 气温组合预测模型的建立过程

利用上文提出的 ARIMA 模型、LSTM 模型和相关数据,将 ARIMA 模型预测结果与 LSTM 对残差的预测结果进行相加,则得到组合模型的最终预测结果。通过对

测试集数据进行验证，实验得到如下图 5-10 所示结果。其中红色曲线为全球平均气温真实值，蓝色曲线所示为全球平均气温预测值。从图中可以发现，真实值与预测值基本拟合，模型的评价指标得分如下表 5-2 所示，准确性为 0.81321，准确度较高。

表 5-2 模型的评价指标得分

评价指标	误差 ME	绝对误差 MAE	百分比误差 MAPE	均方误差平方和 MSE	准确性 score
分数	0.3429	0.4857	0.0516	0.2612	0.81321

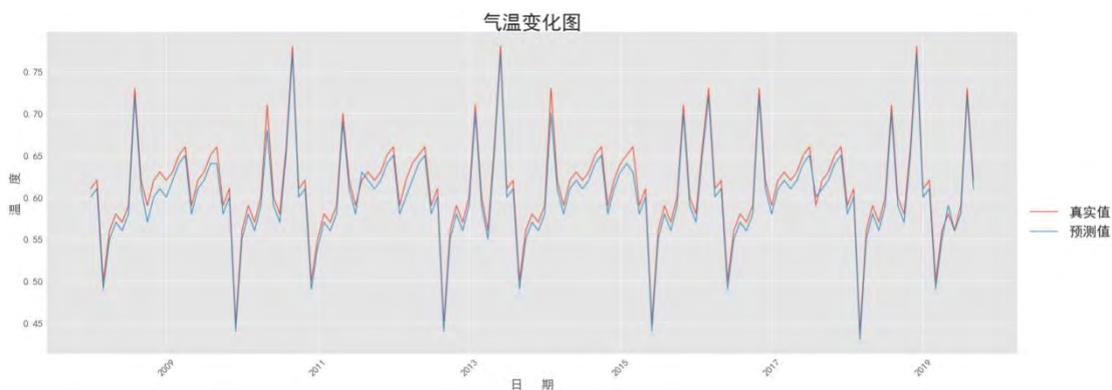


图 5-10 组合模型预测结果

综上，本文利用调参完成的 LSTM-ARIMA 气温组合预测模型对全球未来 25 年的平均气温进行预测，具体结果如图 5-11 所示。结果表明未来全球变暖现象还将持续，但可能会出现短暂的局部降温情况，在 2027 年出现一个小高峰，2022 年和 2033 年出现小低谷，说明在这两个时间段全球平均气温可能会出现异常。



图 5-11 未来全球平均气温预测结果

5.4 问题二模型评价

(1) 本题在预测全球平均气温时选择了基于 LSTM-ARIMA 的气温组合预测模型时间序列分析预测法，该方法将时间因素处于核心位置，突出了时间因素在预测

中的作用，同时也是根据客观事物发展的连续规律性，运用过去的历史数据，通过统计分析，进一步推测全球平均气温的发展趋势。

(2) 时间序列法的预测速度快，对规律变化比较平稳的符合序列有较高的预测精准度，可以反映变化的连续性特点。

(3) 在分析影响全球平均温度变化的各项因素时，在提取特征后采用灰度关联度分析，得出各影响因素与全球气温变化的关系，避免了模型的过拟合，优化了组合模型。

6 问题三建模与求解

6.1 第一问问题分析

全球变暖是一个世界性现象，1880 年至 2012 年间，全球地表温度大约升高了 0.85°C [11]，且陆地比海洋增温快，高纬地区比低纬地区增温快，冬半年比下半年增温明显。在这种全球气候变暖的背景下，极端天气气候事件却频繁发生，给各地的经济、社会和环境带来了严重的负面影响。

问题三第一问的主旨是探索极端气象与气候变化之间的相关性，本题的解题思路分两步：

本题的解题思路分两步：

步骤一：根据题目所给规则：因为“极寒天气”是某地的天气现象，并没有给定指定地区，故实验以历史最热月份、历史最热日、历史最干燥 5 天、历史最潮湿 5 天等因素作为评价是否发生局地极端气候事件指标，获取全球不同地区气温数据和极端天气相关数据。

步骤二：构建复杂网络模型，将极端天气表示为复杂网络中的异常节点，探索全球不同地区气温变化与极端气象出现之间的关系，探索气候变化是否对极端气象发生的严重程度及其发生概率和不确定性产生影响。

6.2 第一问复杂网络模型建立

复杂网络通常使用图 $G(V, E)$ 来标识， V 表示网络中的所有节点， E 表示网络中的所有边， $|V|$ 表示网络中的节点数， $|E|$ 表示网络中的边数。实际上，网络中的每个节点表示现实中的个体，每条连边表示个体间的交互关系[12]。

在建立基于气温数据的回归模型时，若选取连续数据构造训练样本集，则难以覆盖个种类的实际情况，所建模型的泛化能力普遍较差。；若为覆盖各个种类的实际情况而选取大量连续样本作为训练集，又会因冗余而使模型精度下降，且建模过程耗时较长。

针对上述实际情况，本文提出选取连续实际数据来构造样本，采用基于样本构建复杂网络模型完成社区发现，并按规则选取典型样本加入训练集，从而保证训练集尽可能包含各种实际情况，使训练得到的模型具有良好的泛化性能，提高模型精度[13]。对于约简的复杂网络模型，其样本集可表示为：

$$S = \{(x_1(t), x_2(t), \dots, x_m(t), y(t) | t = 0, 1, \dots, N\} \quad (6-1)$$

其中 $(x_1(t), x_2(t), \dots, x_m(t))$ 为输入样本， $y(t)$ 为输出样本。如果复杂网络模型针对的实际问题为时间序列类型，则需进行迭代，即将每次得到的 $y(t)$ 重新加入构造样本输入。本文将每一个训练样本作为一个节点，来构建复杂网络，从而进行社区发现筛选训练样本。选用欧式距离来计算衡量各个节点之间的距离，因为欧式距离可以体现各个变量对距离结果的影响，两节点距离越大，说明两样本的相似性越低，即其来源于不同种类实际的可能性就越大。欧式距离公式如下：

$$m_{ij} = \sqrt{\sum_{i \neq j} (s_{ik} - s_{jk})^2} \quad (6-2)$$

各个数据样本节点间的欧式距离矩阵为：

$$M = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1N} \\ m_{21} & m_{22} & \cdots & m_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N1} & m_{N2} & \cdots & m_{NN} \end{pmatrix} \quad (6-3)$$

其中 $m_{ij} = m_{ji}$ ，对角元 $m_{ii} = 0$ 。该距离矩阵为对角元为 0 的对称阵。因两数据样本节点间的连线无方向性，所以为对称阵；对角元为 0 表明样本节点自身与自身之间无连线。

若两样本所代表的实际情况之间相似度越高，则两样本之间的欧式距离越小。计算得到每两样本间的欧式距离之后，利用距离截断阈值 R ，与每两样本间的欧式距离进行比较，若样本间的距离小于 R 则样本间有边相连，否则无边相连。原始距离矩阵根据阈值处理后，可得表示样本节点连接的矩阵：

$$M' = \begin{pmatrix} m'_{11} & m'_{12} & \cdots & m'_{1N} \\ m'_{21} & m'_{22} & \cdots & m'_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ m'_{N1} & m'_{N2} & \cdots & m'_{NN} \end{pmatrix}, \text{ 其中 } m'_{ij} = \begin{cases} 1, & \text{如果 } i, j \text{ 相连} \\ 0, & \text{否则} \end{cases} \quad (6-4)$$

由此得到的样本节点与表示节点间连接状态的邻接矩阵，共同组成了一个复杂网络。

利用复杂网络模型对样本数据中的节点进行社区发现，网络社区结构描述了节点组成的子网络中彼此间密集连接，而与其他节点的连接非常稀疏，为了确定最优的社区划分和社区数目，本文应用“模块度” [14] 的评价指标：

$$Q = \sum_i (e_{ii} - a_i^2) \quad (6-5)$$

其中 e_{ii} 表示社区 i 内所连边数占整个网络边数的比例， $a = \sum_j e_{ij}$ 表示与社区 i 中节点相连边数所占比例。社区结构越强，其对应的 Q 值越大。利用凝聚分层聚类，初始将每个节点看作一个社区，循环进行社区融合，每次融合成对进行。每次进行循环社区融合时，选择使 Q 值增长最大的社区融合进行。两社区融合后对 Q 值的影响可表示为：

$$\Delta Q = e_{ij} + e_{ji} - 2a_i a_j = 2(e_{ij} - a_i a_j) \quad (6-6)$$

网络初始状态为单个节点独立成为社区，假如节点 i 和 j 有边相连，则 $e_{ij} = 1/2m$ ，否则为 0， $a_i = k_i / 2m$ 。 ΔQ 矩阵初始化为：

$$\Delta Q = \begin{cases} 1/2m - k_i k_j / (2m)^2, & \text{如 } i, j \text{ 相连} \\ 0, & \text{否则} \end{cases} \quad (6-7)$$

该算法的时间复杂度为 $O(n \log^2 n)$ ，速度上有很大提高，使其可以应用于大型的复杂网络社区发现。

6.3 第一问复杂网络模型求解

收集全球不同地区气温数据和极端天气数据，提取数据集中各种统计特征。将极端天气表示为复杂网络中的异常节点，对全局异常检测与异常区域检测两个问题进行了研究，探索全球不同地区气象变化对局地极端气候事件的严重程度及其发生概率和不确定性的影响，实验以历史最热月份、历史最热日、历史最干燥 5 天、历史最潮湿 5 天等因素作为评价是否发生局地极端气候事件指标。

对构建完成的复杂网络模型，计算异常分数[15]，统计异常点和观察值之间的偏差，从而进行异常点分析，探索极端气象与气候变化之间的关系，具体公式如下：

$$OutlierScore_f(G_k) = distance(f(G_k), f(G'_k)) = distance(f(G_k), M(G, t_k)) \quad (6-8)$$

利用复杂网络中的线性回归模型以及修正的迭代线性回归模型，对极端气象这一异常点与气候数据变化情况进行拟合，如图 6-1 所示。异常点检测问题是指给定异常时刻，选出最能代表其网络统计特征的变化的区域。实验结果如图 6-2 所示。

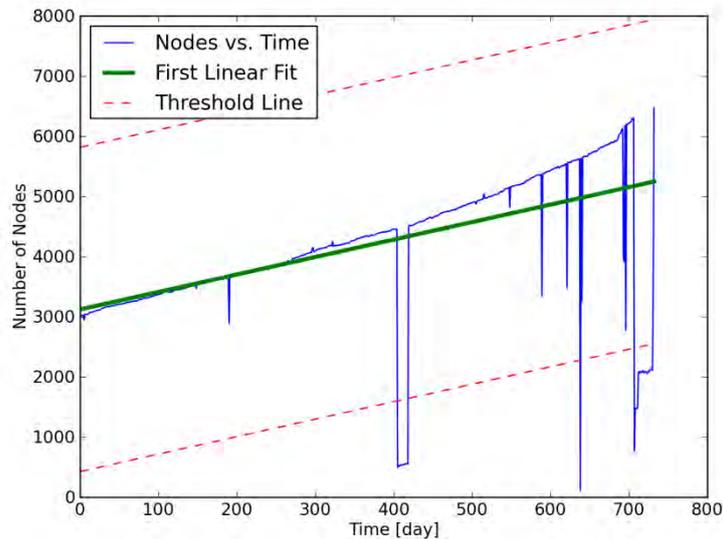


图 6-1 异常点拟合

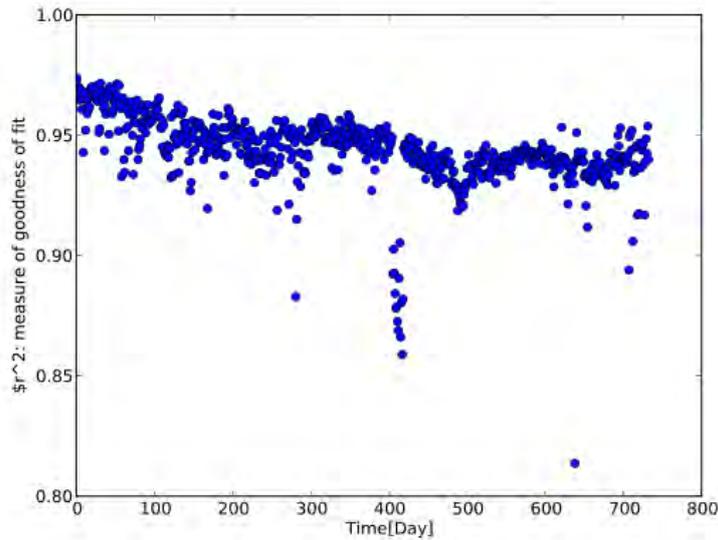


图 6-2 异常点检测

研究结果表明，随着全球气候逐渐变暖，全球 80% 以上的可观测区域历史最热月份和历史最热日发生概率增加，极端事件发生严重性也有所增长；57% 和 41% 观测区域历史最干燥 5 天和历史最潮湿 5 天的发生概率有所增加。对于长时间热浪和干旱事件，热带地区受气候变化影响最为严重，其历史最热月份和历史最干燥年发生的概率分别至少增加到了原来的 4 倍和 2 倍。此外，研究结果还显示，气候变暖增加了北极海冰面积打破低纪录的概率。综上我们做出猜想：极端气象的出现与气候变化相关，且全球气候变暖进一步增加了极端气候发生现象。

6.4 第二问问题分析

问题三第二问的主旨是探究全球变暖与局地极寒现象是否矛盾，是对第一问数据的进一步运用。

本题的解题思路分两步：

步骤一：根据问题一中得到的猜想：极端气象的出现与气候变化相关，本文以具体城市为例对以上猜想做进一步验证。

步骤二：考虑到美国频繁遭受极寒袭击，本文决定以美国为案例，以年降雪量和冬季平均气温作为是否发生局地极寒现象指标[16]，探究其与气温之间的关系，并提出两种假设：

- a. 全球年平均气温对美国年降雪量有正向影响；
- b. 全球年平均气温对美国冬季平均气温有负向影响。

6.5 第二问神经网络模型建立

假设神经网络有 M 层， n 维输出向量 $x_p = (x_{p1}, x_{p2}, x_{p3}, \dots, x_{pn})$ ，用 N_j 表示第 j 个神经元的个数，目标输出为 d 维的向量 $T_p = (T_{p1}, T_{p2}, T_{p3}, \dots, T_{pd})$ ，实际输出 $O_p = (O_{p1}, O_{p2}, O_{p3}, \dots, O_{pd})$ ，BP 神经网络模型如图 6-3 所示。显然 $N_m = d$ ， $N_0 = n$ ， W_{jik} 表示从 $(j - 1)$ 层的第 k 个神经元到第 j 层的第 i 个神经元的输出，这样 $Y_{Mi} = O_{pi}$ ，

考虑第 j 层第 i 个神经元。

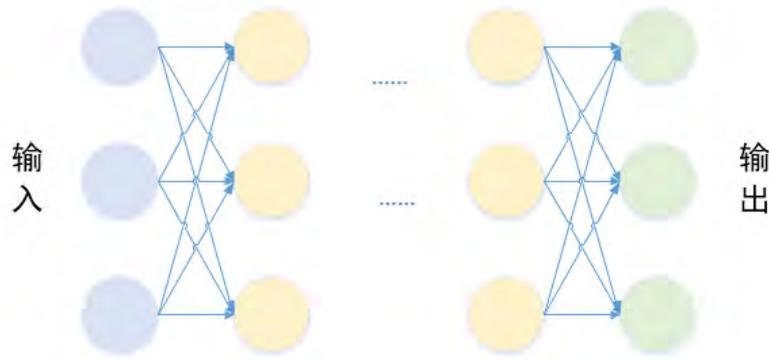


图 6-3 BP 神经网络模型

激活函数为:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (1-1) \quad (6-9)$$

则:

$$f'(x) = f(x) \cdot (1 - f(x)) \quad (6-10)$$

训练样本误差为:

$$E_p = \frac{1}{2} \sum_{j=0}^{N_m} (T_{pj} - O_{pj})^2 \quad (6-11)$$

第 j 层第 i 个神经元输出为:

$$Y_{ji} = f(\text{net}_{ji}) \quad (6-12)$$

其中:

$$\text{net}_{ji} = \sum_{k=1}^{N_{j-1}} Y_{(j-1)k} W_{jik} \quad (6-13)$$

定义局部梯度为:

$$\delta_{ji} = \frac{\partial E_p}{\partial \text{net}_{ji}} \quad (6-14)$$

考虑权重 W_{jik} 对误差的影响, 可以得到:

$$\frac{\partial E_p}{\partial W_{jik}} = \delta_{ji} \frac{\partial \text{net}_{ji}}{\partial W_{jik}} \quad (6-15)$$

$$\Delta w_{jik} = \alpha \frac{\partial E_p}{\partial W_{jik}} = \alpha * \delta_{ji} * Y_{(j-1)k} \quad (6-16)$$

则新的权值为:

$$w_{j i k}^{t+1} = w_{j i k}^t + \Delta w_{j i k}^t \quad (6-17)$$

下面分两种情况计算 δ_{ji} 。

第一种情况当第 j 层是输出层的时候，这时 $j = M$ ，注意 $Y_{ji} = Y_{Mi} = O_{pi}$

$$\delta_{Mi} = \frac{\partial E_p}{\partial net_{ji}} = \frac{\partial E_p}{\partial Y_{Mi}} * \frac{\partial Y_{Mi}}{\partial net_{Mi}} = (T_{pi} - O_{pi}) O_{pi} (1 - O_{pi}) \quad (6-18)$$

第二种情况第 j 层是中间隐藏层的时：

$$\delta_{ji} = Y_{ji} (1 - Y_{ji}) \sum_{\sigma=1}^{N_{j+1}} [\delta_{(j+1)\sigma} W_{(j+1)\sigma i}] \quad (6-19)$$

6.6 第二问神经网络模型求解

6.6.1 算法说明

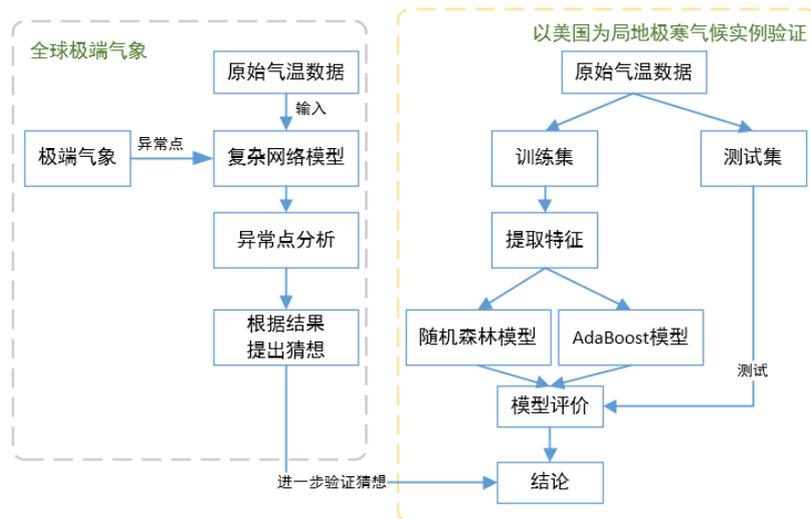


图 6-4 训练流程图

针对问题三中第二问的训练步骤如下，具体流程图见图 6-4：

Step1: 首先对数据进行预处理，包括缺失值填充，数据归一化处理等，将预处理完成的数据分为测试集和训练集两部分。

Step2: 对训练集数据进行特征提取，将提取出的数据特征作为输入变量（输入特征），是否发生局地极寒现象(标签值)作输出变量；

Step3: 将训练集分别输入神经网络模型（随机森林模型、AdaBoost 模型）中，引入两个误差评价指标对模型进行评价[17]，分别为均方误差（MSE）和相对误差值（RE）：

$$MSE = \sqrt{\frac{\sum_{i=1}^n (f(x_i) - y_i)^2}{n}} \quad (6-20)$$

$$RE = \frac{1}{n} \sqrt{\sum_{i=1}^n \frac{(f(x_i) - y_i)^2}{y_i^2}} \quad (6-21)$$

$$OutlierScore_f(G_k) = distance(f(G_k), f(G'_k)) = distance(f(G_k), M(G, t_k))$$

其中 $f(x_i)$ 为预测值， y_i 为真实值， n 为测试样本个数。

Step4: 利用第一问中的分析结果的基础上，对第二问模型结果进行分析，进一步验证第一问的猜想，评价模型，探索全球变暖和局地极寒现象关系，并对可能原因进行分析。

6.6.2 结果分析

为了对上述复杂网络模型的猜想做进一步验证，本文以美国为例，探索全球变暖与局地极寒事件相关性。实验将全球年平均气温、美国年降雪量、美国冬季平均气温数据作为原始数据集，随机选取 90% 的数据作为训练集，其余数据作为测试集。对数据利用 Python 进行可视化展示可以得到图 6-5。

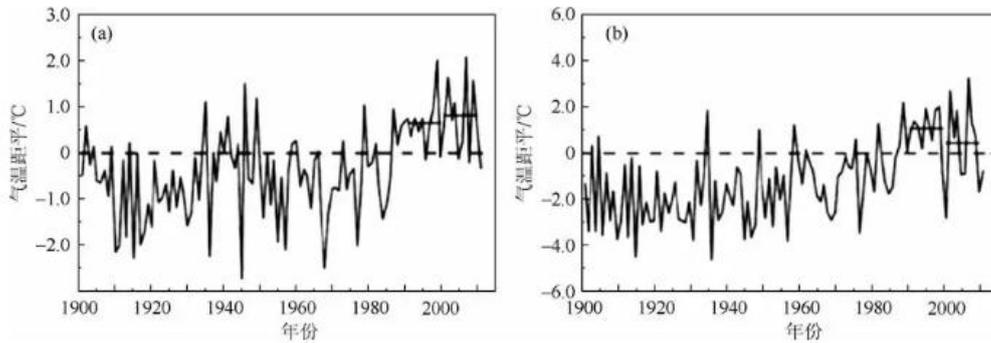


图 6-5 1900-2000 全球与美国气温对比图

图中左侧为全球平均气温，右侧为美国冬季平均气温，由图可看出，近百年全球地表和对流层大气的温度有明显升高。

接下来对数据采用 3 层神经网络模型，对训练集进行十折交叉验证，利用随机森林模型和 AdaBoost 模型在训练集分别训练得到 T 个神经网络回机，表 6-1 给出了当 T 分别为 20, 25, 50, 100 时随机森林模型、AdaBoost 模型以及集成中最佳个体对全球温度的预测效果，其中“最佳个体”表示 T 个神经网络中表现最好的个体。

表 6-1 神经网络模型评价指标

		随机森林模型	AdaBoost 模型	最佳个体
T=20	MSE	0.0617	0.0646	0.0719
	RE	0.0090	0.0094	0.0105
T=25	MSE	0.0707	0.0802	0.0859
	RE	0.0103	0.0118	0.0126
T=50	MSE	0.0609	0.0634	0.0638
	RE	0.0089	0.0092	0.0092
T=100	MSE	0.0649	0.0660	0.0663
	RE	0.0096	0.0097	0.0098

由表可知，两种神经网络模型相比较而言，随机森林模型的误差较小，且曲线平滑、最为稳定，AdaBoost 误差较大，且效果较不稳定，个体各方面表现不佳。所以本文最终选择随机森林模型作为最终验证模型，输入测试集数据得到如表 6-2 所示结果。

表 6-2 随机森林模型得分

自变量	因变量	重要性	标准化的重要性
全球平均气温	美国降雪量	0.92	52.2%
	美国冬季平均气温	-0.86	45.8%

由结果可知，全球平均气温的上升对美国降雪量变化有正向影响，气温越高，降雪量越大。而全球平均气温对美国冬季平均气温又负向影响，即全球平均气温越高，美国冬季气温越低。因此假设成立。从模型来看，全球变暖和局地极寒现象并不矛盾，相反，两者之间恰恰有很强的相关性。

7 问题四建模与求解

7.1 第一问问题分析

问题四属于开放性题目,因为 WMO 中缺省数据较多且有很多极端天气事件的指标,故可挖掘的信息和研究方向较多,如怎么评估全球变暖造成的未知的经济损失,如何根据问题 3 中建立的气候变化模型的进行预测未来未知的气象状况。基于上述研究方向,经过查阅相关参考文献后,本文发现全球气候是一个系统性问题,全球变暖造成的影响在时间、空间上必然会存在差异,且这些影响会受到气候系统变化的扰动。

7.2 第一问问题结论

首先,“全球变暖了,某地今年的冬天特别冷”两者之间并不矛盾。通俗解释如下:全球变暖对于地球来说是一个快速的气候变化,但是对于人类来说是一个缓慢的气候变化,是在年复一年中逐步积累的,并且会潜移默化的产生许多深刻的影响,最终可能会在量变后达到质变。这涉及到天气和气候两个不同的术语概念。天气指一个地方较短时间内各种气象要素的表现状态,气候则是大气物理特征的长期平均状态。全球气候变暖是在长时间序列上观察全球年均温,发现其呈总体上升趋势,并非表明地球上每一个地区气温都在上升,更不排除局部地区会出现严寒天气。因此,某个地方特定一天、一周或是一月内出现的寒冷天气只是暂时的天气现象,并不能用来否定全球气候变暖的趋势。

导致全球变暖的原因可能是有多方面的,其中最为主要的有两方面[18],一是人类在工业革命以来,人类大量使用煤炭、石油等化石燃料,从而向大气中排放大量的二氧化碳,每年人类排放的二氧化碳总量超过 400 亿吨。二是人类对于地球植被的破坏,特别是能够进行光合作用吸收二氧化碳的大量森林的破坏,减少了植被对于二氧化碳的吸收。全球变暖除了带来全球地表平均气温升高,还带来了海平面上升,气候带向极地移动,生物量和生物多样性,地球原有生态系统的改变,极端天气,气候灾害,国际冲突等深刻影响。具体分析如下:

(1) 全球变暖导致海平面上升,淹没沿海低地

全球变暖一方面使得两极地区的大陆冰川冰川融化,大量冰川融水进入海洋;另一方面,全球变暖使得海水受热膨胀,两方面共同促使海平面上升,淹没沿海低地和一些岛屿,比如马尔代夫、瑙鲁、图瓦卢等国家将来都有可能遭受灭顶之灾,目前,图瓦卢已经宣布在对抗海平面上升中失败,打算举国迁徙到新西兰。由于海水比热容较大,90%的全球变暖能量储存在海洋中。因此海洋变暖是全球变暖的一个核心和基本指标。如图 7-1 所示,相比与 1981-2010 年全球平均海表温度平均值,根据问题 3 中气候模型预估,1870~2018 年,全球平均海表温度表现为显著升高趋势。2018 年,全球大部分海域海表温度较常年值偏高,全球平均海表温度为 1870 年以来的第四高值;全球海洋热含量(上层 2000 米)超过 2017 年创下的纪录,2018 年成为有现代海洋观测记录以来海洋最暖的年份。

1980-2017 年，中国沿海海平面呈波动上升趋势，上升速率为 3.3 毫米/年，高于同期全球平均水平。

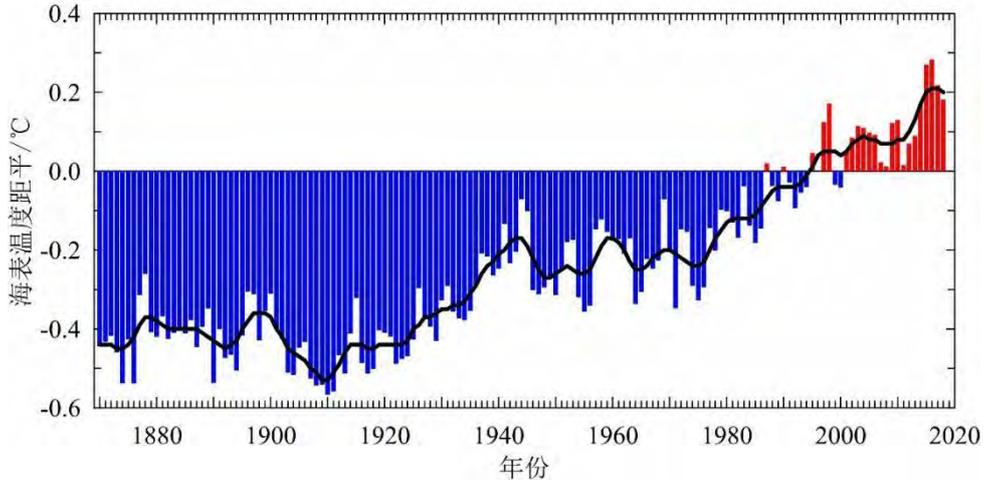


图 7-1 1870~2018 年全球平均海表温度距平

(2) 全球变暖影响大气急流

急流由大气层的活动引起，气候变化可以通过影响大气层的状态，间接改变急流的活动形态。伴随季节变化，急流的纬度也会发生南北摆动。在南北半球的大气层中，各自从低纬到高纬存在着三圈并存的闭合环流——三圈环流，分别称为哈得莱环流、费雷尔环流和极圈环流。急流就被认为是三圈环流的产物（参见示意图 7-2）。

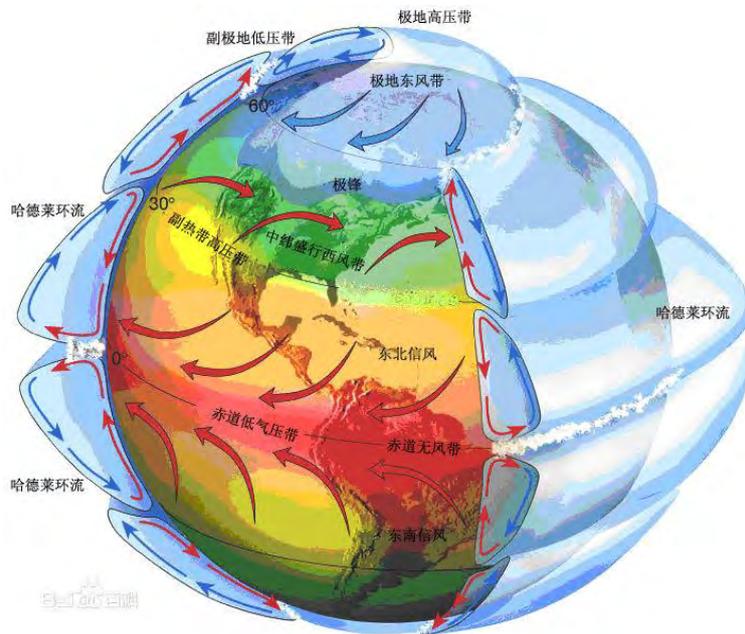


图 7-2 三圈环流示意图

当海冰融化，冰下的海水暴露后，会吸收更多太阳辐射，海洋以及海洋上空的大气开始变暖，从而带来更多热量使得海冰继续融化，形成正反馈机制加速海冰融化。随着海冰逐渐融化，暴露出来的海水会在秋冬时释放出储存的热量，对北极上空大气的基本态产生长达数月、影响明显的扰动。海冰的变化同样也能影

响到大气环流的变化[19]。受北极海冰下降的影响，位势高度场在北极、北太平洋地区明显升高，在中纬度地区却降低。这种分布极易造成北极极涡爆破，导致对流层风场改变。因此，受北极海冰减少的影响，位势高度场和对流层风场的变化，为北极冷空气入侵东亚地区创造了条件，使东亚大陆地区的冬季容易冷冬等极端天气事件。

如果地球持续变暖，在急流中伴随高压脊的地方，高温将会制造更多热浪和干旱；在急流向南弯曲伴随低压槽的地方，海洋上的暖湿空气会随之传播，形成更多暴雨。如果急流持续缓慢移动，并伴随较大振幅的罗斯贝波，还会持续性引发更多的极端天气。

(3) 北极升温和海冰融化导致北半球冷冬频发

随着全球气候变暖，极端天气出现的强度和频率增大，气象灾害增多。如图 7-3 所示，稳定的强极涡（左图）与不稳定的弱极涡（右图）之间的区别：不稳定的弱极涡会导致中纬度地区的寒潮和极寒天气。北极地区和中纬度地区之间气温的不断拉近，导致暖空气侵入北极圈时，减弱了极地高压对极地涡旋的控制能力，进而致使极地涡旋频繁南下，冷空气团向南转移到了欧洲、亚洲和美国。比如，2018 年初袭击美国东北部的“炸弹气旋”，致使气压下降速度创美国近年之最，因此带来了异乎寻常的降雪量。

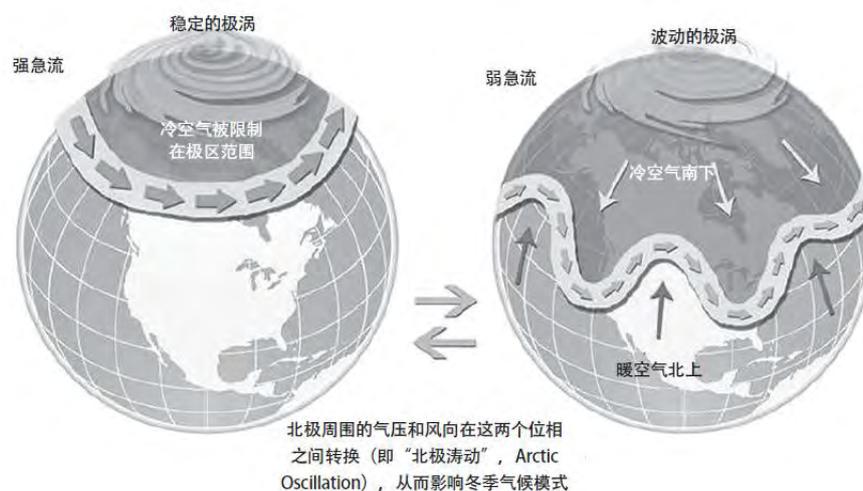


图 7-3 稳定的强极涡与不稳定的弱极涡示意图

(2) 全球变暖会导致地球原有生态系统的改变

自然地理环境具有整体性特征，地形、气候、生物、水文和土壤等自然地理要素相互影响、相互制约，气候条件的变化会导致整个生态系统发生变化，包括一些物种可能会灭绝，比如海洋中的珊瑚虫对海水温度的变化十分敏感，全球变暖已经导致大堡礁中的很多珊瑚虫死亡[20]。此外，全球变暖对人类的主要生产领域产生影响，其中对于农业、林业、牧业和渔业等部门的影响则更为显著。

(5) 极端天气气候事件趋多趋强，气候风险水平呈上升趋势

20 世纪 80 年代以来，中国群发性或区域性极端天气气候事件频次增加，中国气候灾害影响范围逐渐扩大，影响程度日趋严重，直接经济损失不断增加，但

死亡人数持续下降[21]。本文采用机器学习中的逻辑回归理论以及均值回归方法，对近 60 年中国极端天气气候事件进行回归分析。

结果发现，如图 7-4 所示，1961~2018 年间，中国气候风险指数总体呈升高趋势，阶段性变化明显，20 世纪 70 年代末以来波动上升。1991~2018 年中国平均气候风险指数较 1961~1990 年平均值增加了 54%；如图 7-5，中国极端强降水事件呈增多趋势，高温日数和暴雨日数增加；如图 7-6，极端低温频次明显下降，极端低温事件显著减少，极端高温事件在 20 世纪 90 年代中期以来明显增多。

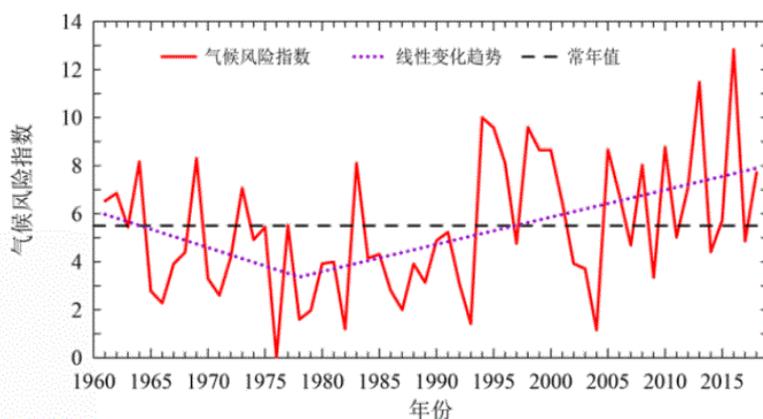


图 7-4 1961~2018 年中国气候风险指数变化示意图

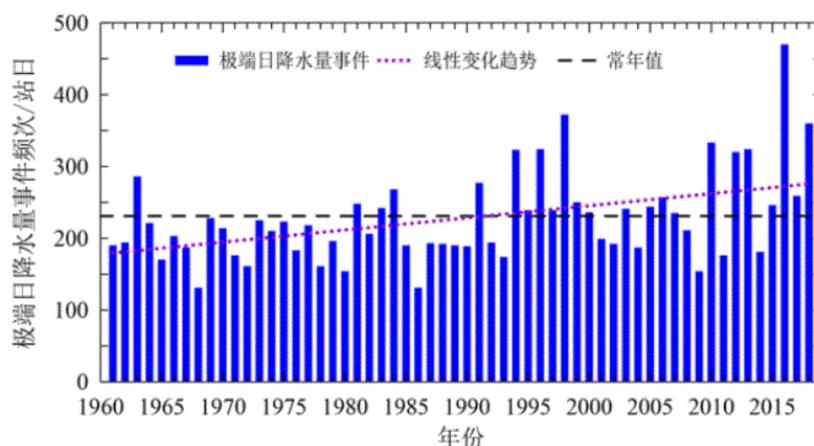


图 7-5 1961~2018 年中国极端日降水量事件趋势图

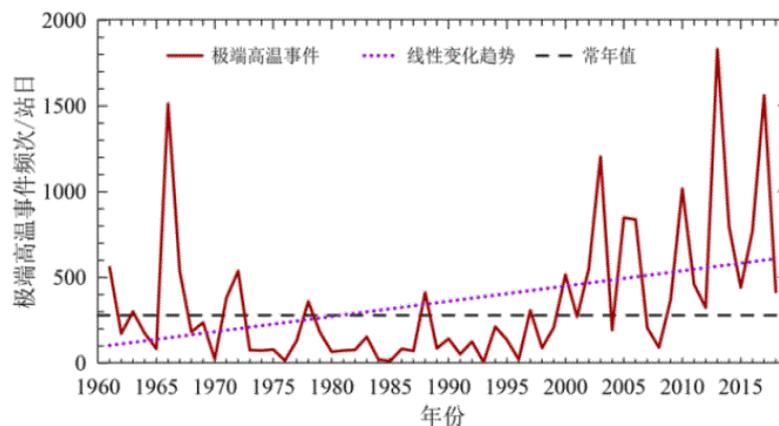


图 7-6 1961~2018 年中国极端高温事件趋势图

经过上述对自 1961~2018 年中国极端高温事件定量分析，结合参阅相关参考文献[22]，本文总结出因全球变暖导致中国遭遇极端天气的原因有以下几点：

- ①当全球变暖导致北冰洋的海冰减少时，北极上空的气压发生变化；
- ②气压变化使北冰洋上空的极锋急流将容易向北侧弯曲摆动；
- ③北极地区的冷空气就会随着大幅摆动的极锋急流涌入西伯利亚地区，使得西伯利亚高压较正常年份更为强大；
- ④强大的西伯利亚高压给东亚带来更强劲的冷空气，结果导致中国遭遇强劲的寒潮。

通俗地说，就是北极冰川的融化改变了大气运动，冷空气就变得更强大。

7.3 第二问新概念提出

为了用更准确的语言描述当下全球面临的环境危机，我们采用“热效应下的气候紊乱”这个新概念来取代全球变暖的概念，主要是通过使用全球气候紧急情况、生态危机、崩溃及“全球变热”等词汇，确保真实反映出当今全球天气变化的混乱状况，而且能够更好地预估未来气候变化的发展趋势。根据图 7-7 中的全球气候类型和洋流图，联合国气候专家表示到 2030 年，碳排放量必须减少一半，才能避免数亿人口面临更严重的干旱和洪涝灾害、食物短缺、极端气候和贫困的威胁[23]。5 月，全球气候专家指出，野生动物的灭绝和地球生物赖以生存的生态系统的破坏，加剧了人类社会的危机。

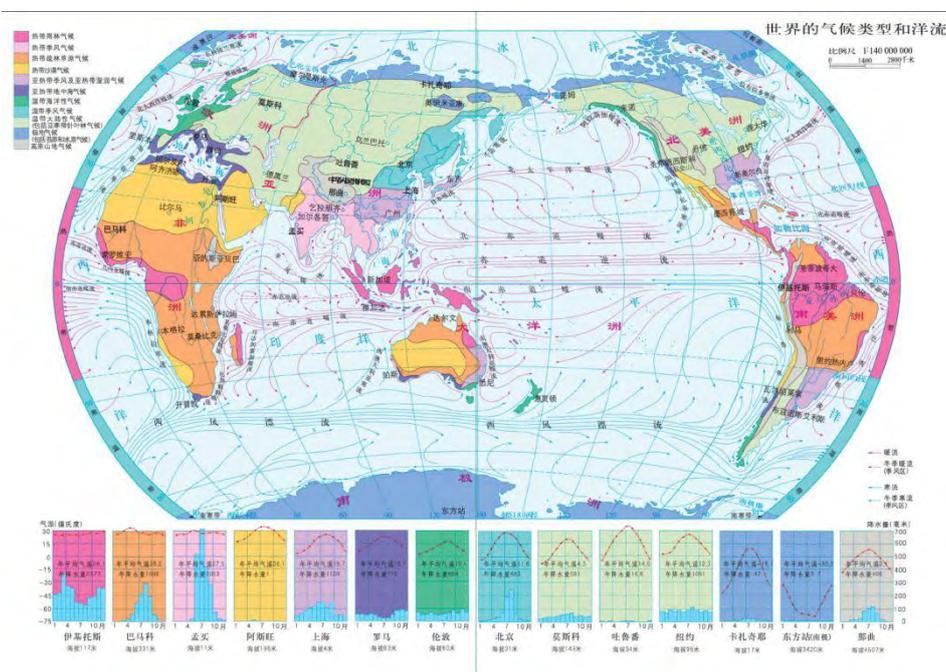


图 7-7 全球气候类型和洋流图

7.4 第二问新概念解释

如图 7-8 所示，自 19 世纪以来，全球平均地面气温大约增加了 1℃，20 世纪全球平均地面气温大约增加 0.75℃，20 世纪后半叶，温室气体造成平均每 10

年增温约 0.13℃。20 世纪以来全球温度的整体变化特征呈现出同温室气体增加类似的增长趋势，然而自从进入 2000 年以来，虽然温室气体的浓度在持续增长，但是全球平均地面气温增温速率却趋于 0（甚至为负），近 10 年间全球全年平均气温上升率仅为 0.03℃，几乎未变化，这种现象叫作 Hiatus（全球变暖停滞状态）。正因为出现全球变暖停滞现象，使公众对全球变暖产生了怀疑。

事实上，全球变暖停滞是由外部强迫和气候系统内部变率共同作用所致。太阳活动的减弱、火山爆发、对流层平流层气溶胶排放的变化以及平流层水汽含量降低会对全球温度变化趋势产生不同程度的影响。从全球温度变化的时间变率及空间特征来看，外部强迫并非全球变暖停滞的主导因子，其对全球变暖停滞的直接贡献远远小于气候系统内部变率。全球变暖停滞更属于气候系统内部变率的一部分，与赤道东太平洋的年代际冷却，信风增强、环流变化以及冷舌模态等的异常变化密切相关。同时，大西洋、印度洋通过海气相互作用对太平洋海洋变率产生影响。

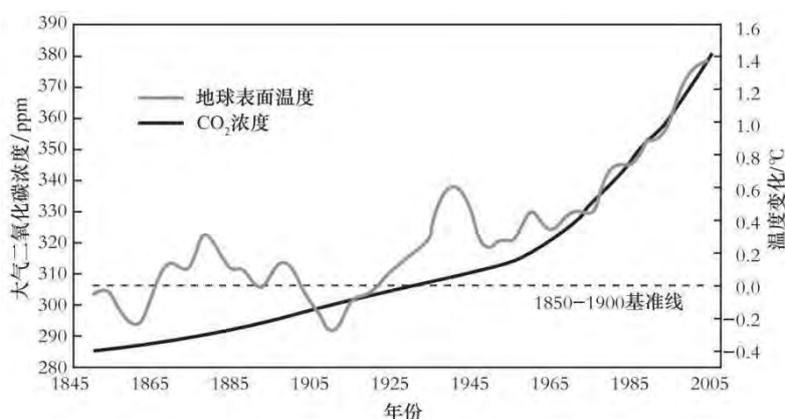


图 7-8 1846-2005 年大气 CO_2 浓度变化趋势图

近年来，全球温度变化又出现明显回升。而这种迅速回升可能是多种因素共同造成的结果，既与人类活动产生的 CO_2 迅速增加有关，也与自然变率有关。人类影响和改变了大气成分和水汽含量，从而影响气候的形成和分布。另外，由于人类活动影响了下垫面性质，从而引起局部地区气候的变化，形成局部地区的小气候，如城市热岛效应等。人类活动对气候的影响越来越大，人类向大气中排放大量的 CO_2 加剧了地球上的温室效应。例如 2014 年以来的全球迅速增温与赤道太平洋地区的持续升温同时发生；2015 年的厄尔尼诺事件能够引起全球 0.02℃ 的增温，同时，热带太平洋通过调控大气热量释放引起全球迅速增温。

综上所述，全球变暖具有复杂性、系统性。影响全球变暖的因素并非是单方面的，而是多种因素综合作用的结果。全球气候是一个系统，全球变暖扰乱、破坏了原有的气候运行方式，但其造成的影响在不同时间、地区有很大差异的。同时，气候系统的变化也会对全球变暖产生影响，放大或减缓全球变暖。当今，全球面临严峻的气候状况和环境危机，气候变化的速率远不止“全球变暖”，应该为“全球变热”。

参考文献

- [1]龚亚亚. 《全球变暖—通俗版》(第二部分)翻译报告[D]. 2016.
- [2]郑彩萍. ARMA 模型的两种参数估计法及残差模型的应用[D]. 燕山大学, 2009.
- [3]张馨月. ARMA 模型的实证应用[J]. 科技信息, 2012(20):151-152.
- [4]安潇潇. ARMA 相关模型及其应用[D]. 燕山大学, 2008.
- [5]郭艳鹂. 多维 ARMA 模型的谱估计及预测方法[D]. 西南交通大学, 2008.
- [6]呼伦贝尔 2017 年冬季极寒天气特征分析_赵洁 (1)
- [7]宋贵彩. 基于 ARMA 模型的结构识别[D]. 重庆大学, 2005.
- [8]韩晓宇. 极寒天气对高速铁路信号设备的影响[J]. 工程建设与设计, 2019(7):131-132.
- [9]卢星宇. 阶数未知的 ARMA 模型 Bootstrap 预测区间构造[D]. 2017.
- [10]赵翠平, 陈岩, 王卫光,等. 黄河流域近 50 a 极端降水指数的时空变化[J]. 人民黄河, 2015, 37(1):18-22.
- [11]苏京志, 温敏, 丁一汇, et al. 全球变暖趋缓研究进展[J]. 大气科学, 2016, 40(6):1143-1153.
- [12]张丽娟. 全球变暖与极端天气频发矛盾吗[J]. 中学地理教学参考, 2018(15).
- [13]任国玉, 陈峪. 全球变暖与中国极端气候事件变化[J]. 科学, 2010, 62(5):30-33.
- [14]赵宗慈[1, 2], 罗勇[1, et al. 全球变暖中的科学问题[J]. 气象与环境学报, 2015(1):1-5.
- [15]Herring S C , Nikolaos C , Andrew H , et al. Explaining Extreme Events of 2016 from a Climate Perspective[J]. Bulletin of the American Meteorological Society, 2018, 99(1):S1-S157.
- [16]崔红艳. 北极海冰变化对北半球气候影响研究[D]. 2014.
- [17]李宏[1], 李建武[2], 宋玉龙[1]. 基于神经网络集成的失业预警方法[J]. 经济与管理研究, 2012(1):89-94.
- [18]张志富[1], 希爽[2], 刘娜[1], et al. 1961-2012 年中国降雪时空变化特征分析 [J]. 资源科学, 2015, 37(9):1765-1773.
- [19]杜岩, 张玉红, 施建成. 海洋表面盐度与海洋环流和气候变化的关系[J]. 中国科学:地球科学, 2019, 49(05):17-29.
- [20]周璞, 徐相明, 黄文娟. 长三角地区平均温度变化特征及其与全球变暖停滞的响应[J]. 安徽农学通报, 2018, 24(19):122-125+129.
- [21]蔡怡, 李海. 全球变暖背景下最近 40 年太平洋海面高度和热容量变化数值模拟[J]. 热带海洋学报, 2006, 25(4):14-19.
- [22]Carey, John. Global Warming: Faster Than Expected?[J]. Scientific American, 2012, 307(5):50-55.

- [23]Corinne Le Quéré, Róisín Moriarty, Andrew R M , et al. Global Carbon Budget 2014[J]. Earth System Science Data, 2015, 7(7):47-85.
- [24]Liu Z . Dynamics of Interdecadal Climate Variability: A Historical Perspective**[J]. Journal of Climate, 2012, 25(6):1963-1995.

附录

附录程序一：

```
%matplotlib inline
import pandas as pd
import pandas_datareader
import datetime
import matplotlib.pyplot as plt
import seaborn as sns
from matplotlib.pyplot import style
from statsmodels.tsa.arima_model import ARIMA
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
import numpy as np
import pandas as pd
from pandas import Series, DataFrame
import statsmodels.api as sm
style.use('ggplot')
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus'] = False

file = r'E:/shumo/大西洋/TEST.xlsx'
fh = pd.read_excel(file, parse_dates=[0])
date=list(fh['date'])
max_temp=list(fh['max_temp'])
min_temp=list(fh['min_temp'])
# mean_temp=list(fh['mean_temp'])

save_file = pd.DataFrame({'真实值':max_temp,'预测值':min_temp})
tem_train=save_file.loc[:]
tem_train.index = date
tem_train.plot(figsize=(35,12))
plt.legend(bbox_to_anchor=(1, 0.5),fontsize= 28.0)#可视化展示
plt.title("气温变化图",size=40)
plt.xlabel("日期",size=25)
plt.ylabel("温度",size=25)
plt.xticks(rotation=45,size=20)
plt.yticks(size=20)
sns.despine()
```

```

plt.savefig(filename="name",format="jpg")
tem_diff = tem_train.diff()
tem_diff = tem_diff.dropna()

plt.figure(figsize=(20,5))#差分序列
plt.plot(tem_diff)
plt.title('一阶差分')
plt.show()
tem_train['diff_1'] =tem_train['mean_temp'].diff(1)
# tem_train['diff_1'] =tem_train['min_temp'].diff(1)
tem_train['diff_2'] = tem_train['diff_1'].diff(1)

tem_train.plot(subplots=True, figsize=(18, 12))
n_sample = fh.shape[0]
# save_file
n_sample = save_file.shape[0]
n_train=int(0.95*n_sample)+1
n_forecast=n_sample-n_train

ts_train =tem_train[:n_train]['diff_2']
ts_test = tem_train[n_train:]['diff_2']

def tsplot(y, lags=None, title='', figsize=(14, 8)):
    fig = plt.figure(figsize=figsize)
    layout = (2, 2)
    ts_ax = plt.subplot2grid(layout, (0, 0))
    hist_ax = plt.subplot2grid(layout, (0, 1))
    acf_ax = plt.subplot2grid(layout, (1, 0))
    pacf_ax = plt.subplot2grid(layout, (1, 1))

    y.plot(ax=ts_ax, color='b')
    ts_ax.set_title(title)
    y.plot(ax=hist_ax, kind='hist', bins=25,color='b',edgecolor = 'black')
    hist_ax.set_title('Histogram')
    smt.graphics.plot_acf(y.fillna(0), lags=lags, ax=acf_ax)
    smt.graphics.plot_pacf(y.fillna(0), lags=lags, ax=pacf_ax)
    [ax.set_xlim(0) for ax in [acf_ax, pacf_ax]]
    sns.despine()
    fig.tight_layout()

```

```
return ts_ax, acf_ax, pacf_ax
```

附录程序二:

```
lags=9
```

```
ncols=3
```

```
nrows=int(np.ceil(lags/ncols))
```

```
fig, axes = plt.subplots(ncols=ncols, nrows=nrows, figsize=(4*ncols, 4*nrows))
```

```
#可视化展示
```

```
for ax, lag in zip(axes.flat, np.arange(1,lags+1, 1)):
```

```
    lag_str = 't-{}'.format(lag)
```

```
    X = (pd.concat([tem_train, tem_train.shift(-lag)], axis=1,
                   keys=['y'] + [lag_str]).dropna())
```

```
    X.plot(ax=ax, kind='scatter', y='y', x=lag_str);
```

```
    corr = X.corr().as_matrix()[0][1]
```

```
    ax.set_ylabel('Original')
```

```
    ax.set_title('Lag: {} (corr={:.2f})'.format(lag_str, corr));
```

```
    ax.set_aspect('equal');
```

```
    sns.despine();
```

```
fig.tight_layout()
```

```
import itertools
```

```
p_min = 0
```

```
d_min = 0
```

```
q_min = 0
```

```
p_max = 4
```

```
d_max = 0
```

```
q_max = 4
```

```
# Initialize a DataFrame to store the results
```

```
results_bic = pd.DataFrame(index=['AR{}'.format(i) for i in
range(p_min,p_max+1)],
```

```
                           columns=['MA{}'.format(i) for i in
range(q_min,q_max+1)])
```

```
for p,d,q in itertools.product(range(p_min,p_max+1),
```

```
                               range(d_min,d_max+1),
```

```
                               range(q_min,q_max+1)):
```

```
    if p==0 and d==0 and q==0:
```

```
        results_bic.loc['AR{}'.format(p), 'MA{}'.format(q)] = np.nan
```

```

        continue

    try:
        model = sm.tsa.SARIMAX(ts_train, order=(p, d, q),
                               enforce_stationarity=False,
                               enforce_invertibility=False,
                               )

        results = model.fit()
        results_bic.loc['AR{}'.format(p), 'MA{}'.format(q)] = results.bic
    except:
        continue
results_bic = results_bic[results_bic.columns].astype(float)

fig, ax = plt.subplots(figsize=(10, 10))
ax = sns.heatmap(results_bic,
                 mask=results_bic.isnull(),
                 ax=ax,
                 annot=True,
                 fmt='.2f',
                 );
ax.set_title('BIC');

train_results = sm.tsa.arma_order_select_ic(tem_train['diff_2'].dropna(),
ic=['aic', 'bic'], trend='nc', max_ar=4, max_ma=4)#平均温度

print('AIC', train_results.aic_min_order)
print('BIC', train_results.bic_min_order)

arma_model =
sm.tsa.ARMA(tem_train['mean_temp'].dropna(),(1,3)).fit(dispatch=-1,maxiter=100)
predict_data = arma_model.predict(start=str(2010), end=str(2010+8), dynamic =
False)
# predict_data = arma_model.predict(start=str(2010), dynamic = False)
print(predict_data)

```