



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 同济大学

参赛队号 19102470046

- 队员姓名 1. 许思燕  
2. 邹聪聪  
3. 陆心怡

**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      基于多变量的全球气候与极端天气模型的构建与应用

**摘要：**

尽管许多监测数据和研究表明全球气候存在明显的变暖趋势，但由于全球气候变化是气候尺度上的全球问题，其影响因素和趋势极为复杂。极端天气尤其是极寒天气的发生，使其与全球气候变暖存在表观上的矛盾性。因此，寻找求证影响气候变化的因素，解释极端天气与全球气候变化间的关系，对于理解和认识全球气候变化的态势有着重要意义。

基于此，本文旨在研究全球气候变化的影响因素和相关性程度，构建简化的全球气候变化模型，并分析极端天气的形成原因及其与全球气候变暖之间的关系，构建极端天气模型予以分析，最后尝试提出新的概念来替代“全球变暖”。然后将各因素的历史数据与全球温度之间建立多变量回归模型，构建了三种全球气候模型对未来 25 年的气候变化进行预测；运用 Python 和 SPSS 软件量化全球温度分布数据，建立极端天气数据与全球气候的关联性模型；运用关联性模型量化极端天气与全球局部地区的异常气候现象之间的关联程度，以此提出新概念揭示气候变化的趋势和复杂性。

针对问题一，需要对加拿大各地历史天气变化以及海洋表面温度的历史数据进行分析，研究其时空变化趋势。考虑到加拿大国土面积较大，监测站点较多，研究中选取 4 个典型地点，基于其历史天气数据分析温度对时间的倾向率和温度变化的平稳性与周期性。对于海洋表面温度的分析，首先采用 python 对最新的全球海洋表面温度数据集 ERSST.V5 进行数据提取和分析，将海洋表面温度数据分为高中低三个纬度范围分别进行分析，以规避全球平均处理后的局部温度变化特征。

针对问题二，要求构建刻画气候变化的模型，对未来 25 年的全球气候作出预测。为此，寻找并求证影响气候变化的因素，从而建立气候变化的多变量预测模型。研究中结合文献调研，初步确定影响全球气候变化的因素后，搜集整理了大量的全球气候数据和各要素的历史数据集，运用 SPSS 和 Python 软件对这些历史数据的时空变化趋势等特性做统计分析。然后分析各因素的历史数据与全球温度的关联性，研究中构建了三种全球气候模型：全球气候变化多变量线性回归模型、基于随机森林算法的多变量非线性模型和基于 ARIMA 的时间序列模型。采用这三种全球气候模型对未来 25 年的气候变化进行预测。并

针结果预测结果分析三种模型的优缺点及推广可行性。模型的应用中，发现多变量线性回归模型结果揭示了影响全球气候变化的影响因素及影响程度，基于随机森林算法的多变量非线性模型的拟合程度很高，ARIMA 模型则对于未来 25 年的全球温度预测结果合理性最高。

针对问题三，基于全球范围内各经纬度网格的温度数据集 GISTEMP，将极端天气进行界定并量化，采用极端天气的离群频次作为量化指标，构建极端天气与全球气候的关联性模型。同时，研究基于统计学对极端天气的态势与全球气候变化的关系进行分析，改进性地提出极端天气态势的三种模式：A 模式、B 模式和 C 模式。模型结果揭示全球极端天气的变化趋势接近 C 模式，即全球平均温度升高，极热天气频次增加，极寒天气频次有所降低。结果还表明局地极热频次与全球温度距平存在显著正相关关系，极寒频次与其存在明显负相关。这表明全球气候逐渐变暖，极寒频次有所降低，但这并不意味极寒天气不会发生，或发生时的程度有所降低。

针对问题四，需要解释局地极寒与全球变暖的关系，并提出新概念来替代“全球变暖”。在问题三对于极端天气的量化数据和关联性模型的基础上，本研究进一步尝试探究局地极寒天气等极端气象的形成机理，在搜集整理大量气象要素的历史数据基础上，建立了局地极端天气的多因子关联性模型。从关联性模型运行结果得到的各因素的相关性系数，找出影响全球极端天气的因素并量化影响程度。基于前述研究，本文提出了全球气候变暖的两个可替代性新概念：“全球气候异常”和“全球气候振荡热移”。

本文对于所构建模型均进行了检验与评价，结合模型应用的结果，对于模型优缺点进行评价，提出各模型相应的推广可行性及进一步改进的建议。

**关键词：** 全球气候变化；多变量回归模型；关联性分析；随机森林算法；ARIMA 模型极端天气；局地极寒；量化分析

# 目录

一. 问题重述 .....	5
1.1 问题背景 .....	5
1.2 问题提出 .....	6
二. 模型假设及符号设定 .....	7
2.1 模型假设 .....	7
2.2 符号设定 .....	7
三. 历史天气数据分析 .....	9
3.1 问题一的分析与思路 .....	9
3.2 历史天气数据的特征分析方法 .....	9
3.2.1 气候倾向率分析 .....	9
3.2.2 时间序列的平稳性及周期性检验 .....	10
3.3 问题一中的加拿大历史天气变化分析 .....	10
3.3.1 加拿大点位选取 .....	10
3.3.2 加拿大历史气候特征分析 .....	11
3.3.3 加拿大历史天气数据的平稳性及周期性检验 .....	13
3.4 问题一中的海洋表面温度特征分析 .....	16
3.4.1 海洋表面温度的数据提取与处理 .....	16
3.4.2 海洋表面温度的特征分析 .....	16
3.4.3 海洋表面温度数据的平稳性及周期性检验 .....	17
四. 全球气候变化因素分析及预测模型 .....	20
4.1 问题二的分析与思路 .....	20
4.2 全球气候变化特征 .....	22
4.2.1 全球气候变化特征分析 .....	22
4.2.2 全球气候变化平稳性与周期性分析 .....	22
4.3 全球气候变化的影响因素分析 .....	25
4.3.1 温室气体 .....	25
4.3.2 太阳辐射 .....	26
4.3.3 海洋作用 .....	27
4.3.4 土地利用 .....	28
4.3.5 臭氧层 .....	29
4.3.6 大气气溶胶 .....	31
4.3.7 空气污染 .....	32
4.3.8 化石燃料 .....	34
4.3.9 城市化 .....	35
4.4 全球气候变化多变量线性回归模型 .....	35
4.4.1 模型的构建思路 .....	35
4.4.2 数据集的搜集及变量含义设定 .....	36
4.4.3 数据缺失值处理及标准化 .....	37
4.4.4 因子的关联性分析 .....	39
4.4.5 模型的构建与求解 .....	47
4.4.6 模型预测结果及检验 .....	50

4.5 基于随机森林算法的多变量非线性回归模型 .....	52
4.5.1 随机森林算法的概念及过程 .....	52
4.5.2 模型的构建与结果 .....	52
4.6 基于 ARIMA 时间序列模型的预测 .....	53
4.6.1 ARIMA 模型 .....	53
4.6.2 全球气候的时间序列预测模型 .....	54
4.6.3 时间序列模型的预测结果 .....	60
4.7 三种模型的优缺点及推广可行性 .....	61
4.7.1 模型 1：全球气候变化多变量线性回归模型 .....	61
4.7.2 模型 2：基于随机森林算法的多变量非线性回归模型 .....	61
4.7.3 模型 3：基于 ARIMA 的时间序列模型 .....	61
五. 极端天气模型的构建与应用 .....	62
5.1 问题三的分析与求解思路 .....	62
5.2 极端天气模型的建立与应用 .....	63
5.2.1 本模型中极端天气的界定 .....	63
5.2.2 极端天气的数据处理与分析 .....	63
5.2.3 基于统计学的极端天气与全球气候间的关系分析 .....	65
5.2.4 极端天气与全球气候的关联分析模型应用 .....	66
5.3 模型结果解释及问题三的解决 .....	67
六. 对极端天气与气候变化的进一步探讨 .....	69
6.1 问题四的分析 .....	69
6.2 局地天气极寒的概念及原因分析 .....	69
6.3 局地极端天气的多因子关联性模型 .....	70
6.3.1 各因子对应的数据集处理 .....	70
6.3.2 各因子相关性关系 .....	71
6.3.3 关联性模型结果的解释及应用价值 .....	73
6.4 全球气候变暖的可替代性新概念 .....	73
6.4.1 “全球气候异常” .....	73
6.4.2 “全球气候振荡热移” .....	73
八. 参考文献 .....	74
九. 附录 .....	75
附录 1 netCDF 的 python 提取与处理代码 .....	75
附录 2 对全球海洋表面的数据集的处理代码 .....	77
附录 3 运用 Python 对时间序列的数据处理与可视化分析代码 .....	79
附录 4 运用 Python 构建及应用 ARIMA 模型的代码 .....	80
附录 5 基于随机森林的多变量回归模型的代码 .....	84
附录 6 极端天气的量化数据提取及多因子关联性分析 .....	87

## 一. 问题重述

### 1.1 问题背景

人类活动例如化石燃料的燃烧和森林的砍伐，会产生大量的二氧化碳等温室气体。自太阳辐射中的可见光透过大气中的温室气体到达地球表面被地表吸收，待地表冷却时这部分吸收的光能以红外辐射形式被大地再辐射到空中，具有长波能量的红外线被大气中温室气体所吸收。这样地球大气中的温室气体如同温室的玻璃一样，阳光可以射进温室，温室中的热量却无法散发出去，造成“温室效应”。大气温室气体浓度增加引起的“温室效应”是全球气候变暖的起因**错误!未找到引用源。**。全球气候变暖将引发一系列自然灾害，包括海平面上升、传染病蔓延、飓风次数增多程度增加、洪水和干旱日益频繁，这些自然灾害将对人类生存造成巨大破坏[2]。虽然 21 世纪以来，温室气体浓度不断持续快速上升，但是全球气温并未出现明显增长的情况，甚至出现略微的下降趋势，呈现全球变暖“停滞”（Hiatus）现象**错误!未找到引用源。**。2019 年 1 月 30 日，美国中西部遭遇异常强大的寒流带来的极寒天气，造成多人伤亡，社会秩序瘫痪。这些极端寒冷天气的报道加上全球变暖“停滞”现象，更使公众对于全球气候变暖产生了怀疑。

产生分歧的原因是考虑问题时时间尺度上的不同。气候变化是指气候平均状态统计学意义上的巨大改变，或者持续 10 年或更长时间的气候变动**错误!未找到引用源。**。不同于天气的冷热带给人们的当下的直观感受，气候的变化需要大时间跨度的数据去体现。北京大学大气科学系的学者，以联合国世界气象组织（WMO）与环境问题科学委员会（SCOPE）联合发布的“气候变化科学纲要”为基础，进行数据分析。发现 1999-2008 年全球气温上升趋势几乎为零，出现全球变暖“停滞”现象，但是这十年仍是 1975-2008 年气温最高的十年。**错误!未找到引用源。**也就是说从长期来看，温室气体浓度增加、全球气温上升是真实存在的，只是从短期来看可能出现变暖的减缓。而极寒气候很有可能是全球气候变暖带来的极端天气负面影响。南京信息工程大学的学者，综合美国 NASA、瑞士苏黎世大学、日本东京大学等国内外研究人员的研究成果，将全球变暖“停滞”现象的成因归纳为两种观点。一种认为外强迫导致了该现象，即太阳活动减弱、火山活动产生气溶胶、平流层水汽减少以及人类活动产生气溶胶这四个因素共同作用；另一种认为自然变率造成了变暖停滞，即海洋吸收了人类活动产生的热量。**错误!未找到引用源。**除了人类活动产生温室气体、气溶胶，以及破坏植被、城市化等会造成全球气候变暖外，太阳辐射变化、火山活动、海洋与大气之间的相互作用等自然因素也会对全球气候造成很大影响，这些复杂的影响因素使得研究全球温度变化变得十分困难。

为了有利于非专业人士了解、理解全球气候变化的知识，我们需要建立一个简化的气候模型和极端天气模型，对未来气候变化做出合理的预测，对极端气象与全球变暖之间的关系做出解释。向人们展示全球气候变化的发展趋势及其影响，解释极端天气的发生原因，增强人们保护环境的意识以及危险预警的能力，督促决策者针对全球气候问题做出相应的政策规定。因此，如何准确、有效的建立气候模型和极端天气模型，对于环境保护和人民生命财产安全具有重要意义。

## 1.2 问题提出

问题一：根据加拿大各地区不同时间天气变化的历史数据，分析该地区温度的时空变化趋势。根据各地区历史天气的数据，进行特征分析，包括从气候倾向率进行分析，并且对时间序列模型的平稳性以及周期性进行检验。另外，从海洋表面温度历史数据出发，对海洋表面温度进行特征分析，阐述其中所蕴含的规律性内容。

问题二：为了了解地球的气候，本文试图调查和综合当前关于气候模型层次结构的一些思想，给出了一些关于预测模型的正式描述，并调查了预测模型用于生成，测试和确认假设的可行方式，结合了当代气候模型的一些特点，用于预测实际过程中的温度变化问题。根据全球气候变化的特征以及全球气候变化的影响因素（包括地球的吸热、散热、海洋的温度变化、空气污染等要素）进行分析，建立相关气候模型，并对未来 25 年的气候变化进行预测。

问题三：通过对气候模型进行模拟和观测分析，即使在 21 世纪温度变暖的情况下，极端寒冷事件也可能持续到每个陆地大陆。通过相对于过去的再分析数据和未来的多模型协议的偏差分数来评估模型模拟的冷极值的可信度。分析不同年度，不同地区的极寒天气与气候变化之间的关系。利用题目所提供的加拿大各地天气历史数据、海洋表面温度数据和其它相关数据，建立相应的气候模型和极端天气模型，用以判断全球变暖和局部地区极寒天气的出现是否存在矛盾关系。

问题四：请用自己的文字解释“全球变暖了，某地今年的冬天特别冷”，阐述全球变暖与局部地区冬季天气寒冷之间的关系；请基于气候变化的复杂性，建立一个新概念代替“全球变暖”并予以解释，降低群众面对严寒天气时对“全球变暖”的误解，并要求能反映气候变化的趋势。

## 二. 模型假设及符号设定

### 2.1 模型假设

- (1) 假设各地区相对湿度恒定、粒子蒸发速度相当;
- (2) 假设各地区天气采样数据真实有效;
- (3) 假设采集数据具有一定的相关性，具有明显的周期性;
- (4) 假设不考虑地形、地面状况等其它因素对气候的影响。

### 2.2 符号设定

本文中用到的变量符号及其含义如表 2-1 所示。

表 2-1 不同变量的数据集来源及含义

变量	含义
L_O_Temp_Ano_Nosmooth	温度的距平
L_O_Temp_Ano_Lowess	温度的距平（平滑处理后的值）
L_Mean_Temp	全球陆地平均温度
L_Mean_Temp_Uncer	全球陆地平均温度的不确定性
L_O_Mean_Temp	全球陆地和海洋表面平均温度
L_O_Mean_Temp_Uncer	全球陆地和海洋表面平均温度的不确定性
G_Temp_F_mean	全球平均温度（华氏度）
G_Temp_C_mean	全球平均温度（摄氏度）
Solar_Flux_mean	太阳辐射通量
Urbanization	城市化程度
Emiss_of_Ozone	臭氧消耗物质的排放量
Ozone_hole	臭氧洞区
PM2.5	空气颗粒物
SO2	二氧化硫
Land_Built_up	土地利用指标：建成区面积
Land_Graze	土地利用指标：放牧区面积
Land_Crop	土地利用指标：耕地区面积
annual_CO2	每年二氧化碳排放量
cumu_CO2	累计二氧化碳排放量
CFCsTG_OBS	氟氯碳化合物（数据来源于OBS）
CH4_OBS	甲烷（数据来源于OBS）
CO2_OBS	二氧化碳（数据来源于OBS）
N2O_OBS	一氧化氮（数据来源于OBS）
Aero_SA	平流层气溶胶
Aero_BC	化石燃料黑碳气溶胶
Aero_RA	对流层气溶胶反射作用

Aero_AIE	气溶胶间接作用
SST_high_lat	海洋表面温度（高纬度）
SST_mid_lat	海洋表面温度（中纬度）
SST_low_lat	海洋表面温度（低纬度）
Extreme_L_T_2C	全球范围内极端低温（2°C基准）的年总频次
Extreme_H_T_2C	全球范围内极端高温（2°C基准）的年总频次
Extreme_L_T_3s	全球范围内极端低温（ $3\sigma$ 基准）的年总频次
Extreme_H_T_3s	全球范围内极端高温（ $3\sigma$ 基准）的年总频次
Global_T_Ano	全球平均温度距平
Monsoon_SW_US	美国西南部季风
Monsoon_Mid_Ind	印度中部季风降水
ACE_Atlantic	大西洋累计气旋强度
ACE_EP	东太平洋累计气旋强度
O_AAO_20CR	南极涛动
O_AO_20CR	北极涛动
O_NAO	北大西洋涛动
O_EP	东太平洋涛动
DO_PDO	太平洋代际振荡
O_SOI	南方涛动指数
Hurricane_Total	每月飓风总数
Hurricane_MDR_SST	飓风主要区域内的海洋表面温度
S_I_extent_NH	北半球的海冰范围
S_I_extent_SH	南半球的海冰范围
Nino_1_2_SST	厄尔尼诺1和2
Nino_3_4_SST	厄尔尼诺3和4
Nino_3_SST	厄尔尼诺3
Nino4_SST	厄尔尼诺4

### 三. 历史天气数据分析

#### 3.1 问题一的分析与思路

针对加拿大各地天气变化以及海洋表面温度历史数据，本章通过分析温度对时间的倾向率、温度变化的平稳性及周期性获得各地温度的时间变化规律，通过对不同地区的温度变化趋势，获取温度的空间分布图。

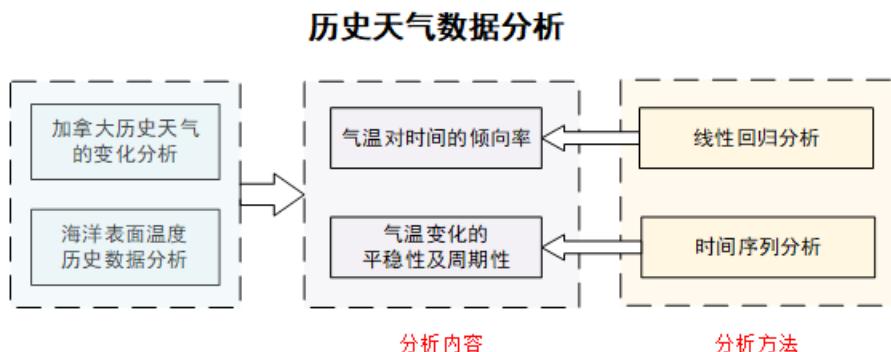


图 3-1 问题一的解决思路

针对问题一中加拿大历史天气的变化，首先，我们对加拿大点位进行选取，以加拿大各地的部分城市作为代表，分析加拿大历史气候的特征。我们对加拿大不同地理位置的气候倾向以及气候异常做了重点分析。气候倾向率的分析包括加拿大北部气候、东海岸气候、内陆气候和南部气候的分析，并对各地区气候趋势图做了详细分析。对于气候异常的分析，主要针对加拿大各地区异常寒冷发生次数和异常寒冷发生频率进行了详细分析。随后我们对加拿大历史天气数据的平稳性和周期性进行检验，建立加拿大各地区历史天气数据的时间序列。

针对问题一中海洋表面温度的特征分析，首先，我们收集了海洋表面温度的历史数据，并对数据进行了提取和处理。然后对海洋表面温度进行了特征分析，并对海洋表面温度进行了平稳性和周期性的检验，将逐月数据分析和逐年数据分析进行对比。

#### 3.2 历史天气数据的特征分析方法

##### 3.2.1 气候倾向率分析

###### (1) 气候倾向率

气候倾向率反应气候要素的变化趋势，包括气温倾向率（ $^{\circ}\text{C}/10\text{ 年}$ ）和降水量倾向率（ $\text{mm}/10\text{ 年}$ ）。本文中气候倾向率均指气温倾向率，即，温度随时间的变化趋势。一般采用一次线性方程来定量描述，即：

$$T_i = b + k \times t_i$$

趋势变化率方程：

$$\frac{dT_i}{dt} = k \quad (2)$$

式中： $T_i$  为气温，单位为 $^{\circ}\text{C}$ ； $t_i$  为时间，单位为年； $k$  为线性趋势项， $k \times 10$  称做气候倾向率，

其单位为 $^{\circ}\text{C}/10$  年。

### (2) 气温异常分析法

在统计各地近 50 年的极端气温时，以  $T$  表示平均气温， $T_p$  表示多年气温平均值， $S$  表示平均气温标准差，参照冷暖等级标准，分析气温的异常特征。有两种极端情况： $T \geq T_p + 2.0S$ ，为异常偏暖； $T_p - 2.0S \leq T$ ，为异常偏冷。

## 3.2.2 时间序列的平稳性及周期性检验

时间序列是按时间进行排序的一组随机变量，通常是在一定的时间间隔内以一定的采样率对某潜在过程进行观测或采样的结果。时间序列数据本质上反映了随机变量随时间的变化趋势。因此通过对加拿大各地以及海洋表面温度的时间序列进行研究，可以获取温度随时间变化的规律。

### (1) 时间序列平稳性

某随机变量的时间序列具有平稳性表示该时间序列的本质特征能延续到未来，即该随机变量的历史和现状特征具有代表性和可延续性，因此可以基于历史和现状预测随机变量的未来趋势。

### (2) 时间序列周期性

某随机变量由于外部影响随自然季节的交替变化出现波峰与波谷的规律。

## 3.3 问题一中的加拿大历史天气变化分析

### 3.3.1 加拿大点位选取

由于加拿大国土面积过于广阔，近 1000 万平方公里，难以分析加拿大各个州市的气候变化，因此在加拿大北部（近北冰洋）、东海岸（近大西洋）、内陆、西南部（近太平洋）各选取一个较为典型城市代表该地区的气候变化。如下图所示。

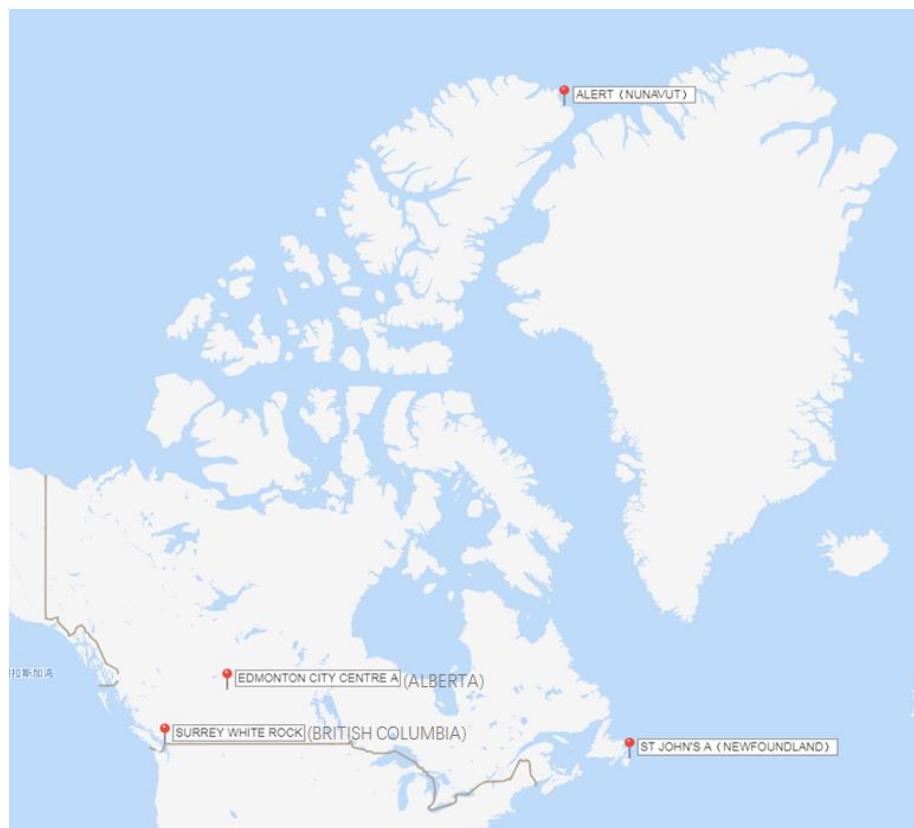


图 3-1 加拿大地图

### 3.3.2 加拿大历史气候特征分析

#### (1) 气候倾向率分析

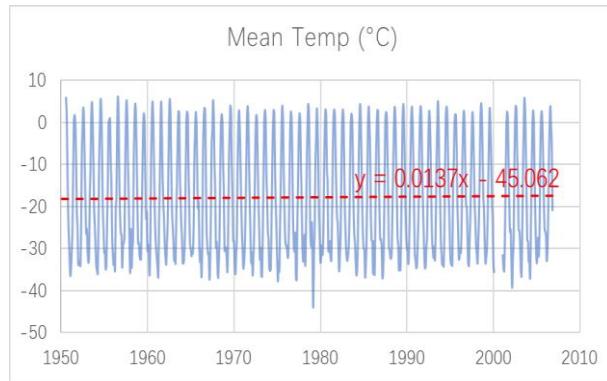


图 3-2a 加拿大北部气候趋势图

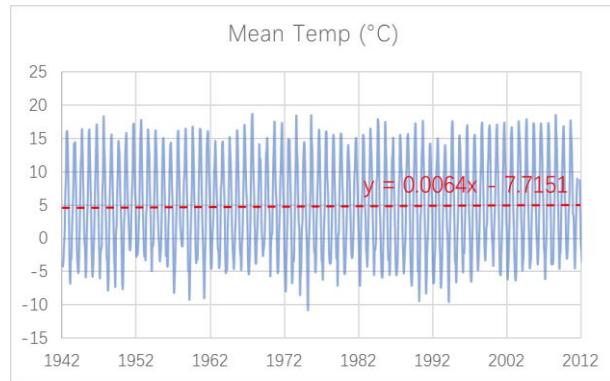


图 3-2b 加拿大东海岸气候趋势图

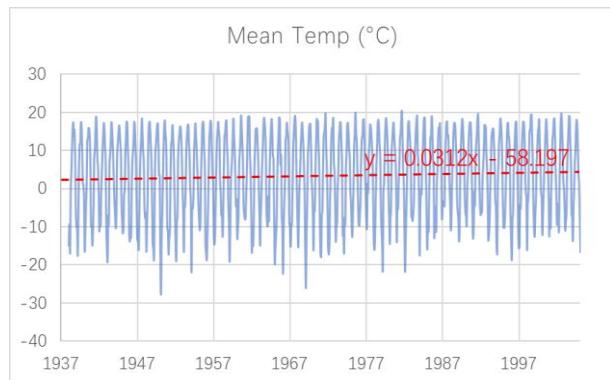


图 3-2c 加拿大内陆气候趋势图

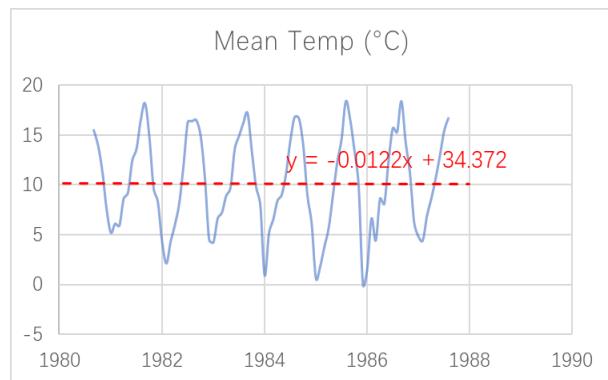


图 3-2d 加拿大西南部气候趋势图

通过对比分析可知，加拿大北部、东海岸、内陆平均气温随年份增加均有上升趋势，西南部气温随年份增长则有略微下降趋势；其中内部升温趋势较为明显，其气候倾向率达 $0.3^{\circ}\text{C}/10$  年，即每十年加拿大内陆气温将升高 $0.3^{\circ}\text{C}$ 。

## (2) 气候异常分析

经统计分析，加拿大四个地区近年来均未发生异常偏暖现象，而异常偏冷现象出现较为频繁，其中内陆和西南部发生异常偏寒的频率达 1.2%，即在加拿大内陆和西南部地区每八年将有一个月发生异常偏冷现象。

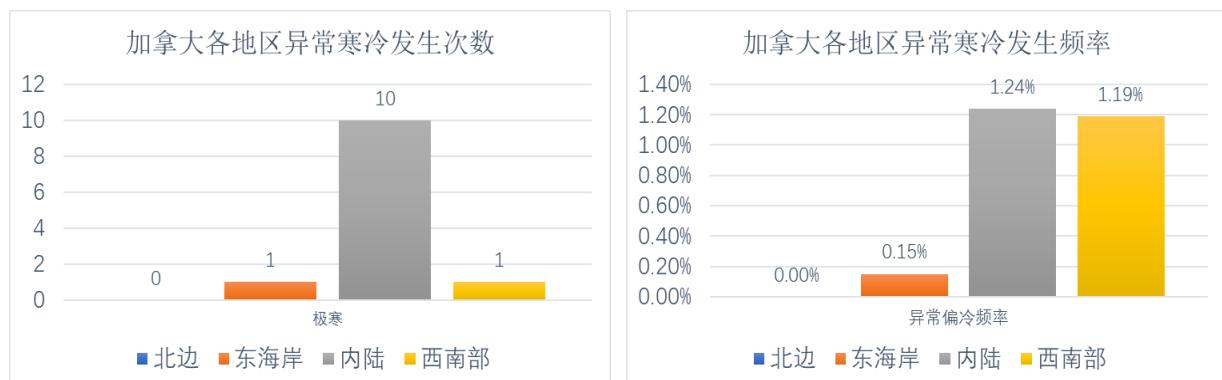


图 3-3 加拿大各地区发生异常寒冷的次数与频率

表 3-1 加拿大内陆地区异常寒冷发生年月

Year	Month	Mean Temp (°C)	加拿大内陆平均气温
1943	1	-19	
1950	1	-27.8	
1952	1	-18.9	
1954	1	-22	
1959	1	-19.2	
1964	12	-19.5	
1966	1	-22.4	
1969	1	-26.1	
1979	2	-21.4	
1982	1	-21.8	3.358°C

可以发现，加拿大内陆地区异常寒冷现象主要发生在 1 月份左右，基本每隔 3-8 年就发

生一次异常寒冷。

### 3.3.3 加拿大历史天气数据的平稳性及周期性检验

#### (1) 历史天气数据的时间序列建立

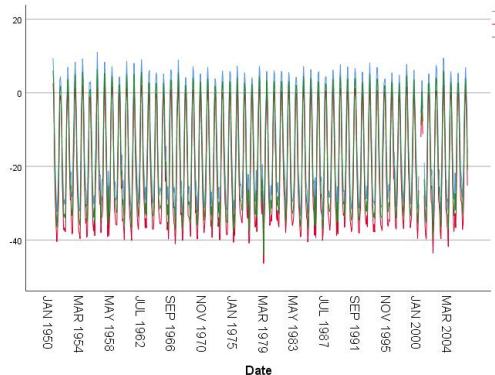


图 3-4a 加拿大北部温度时间序列

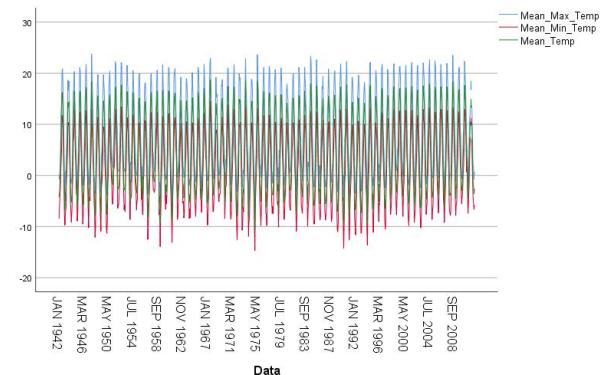


图 3-4b 加拿大东海岸温度时间序列

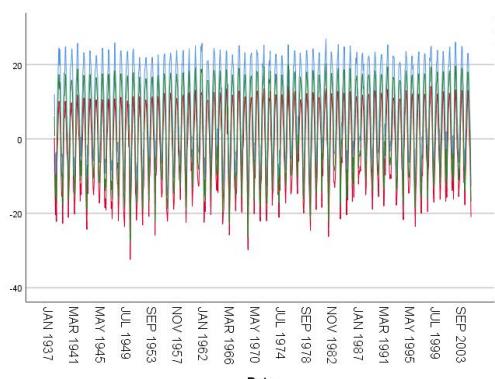


图 3-4c 加拿大内陆温度时间序列

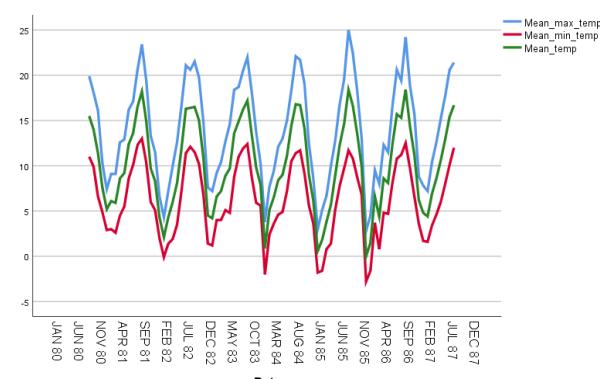


图 3-4d 加拿大西南部温度时间序列

#### (2) 时间序列的平稳性及周期性检验

##### a) 平稳性

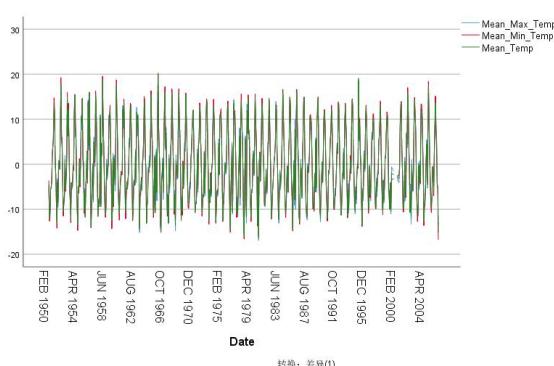


图 3-5a 加拿大北部温度时间序列 1 阶差分

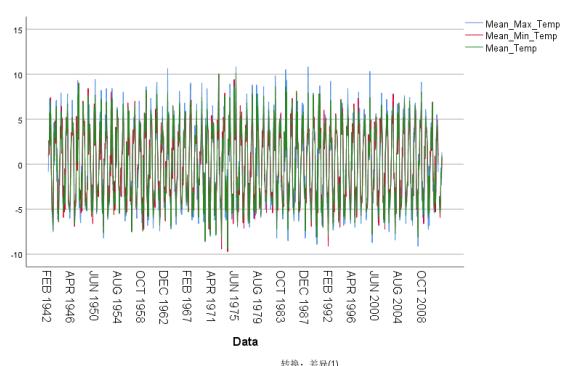


图 3-5b 加拿大东海岸温度时间序列 1 阶差分

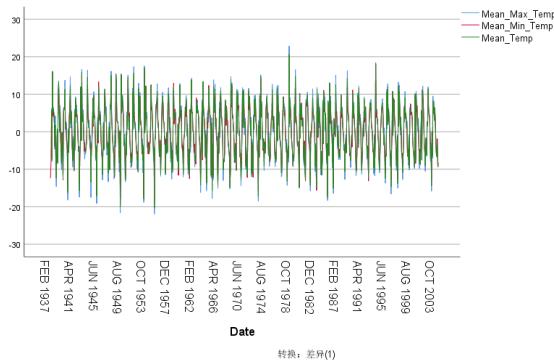


图 3-5c 加拿大内陆温度时间序列 1 阶差分

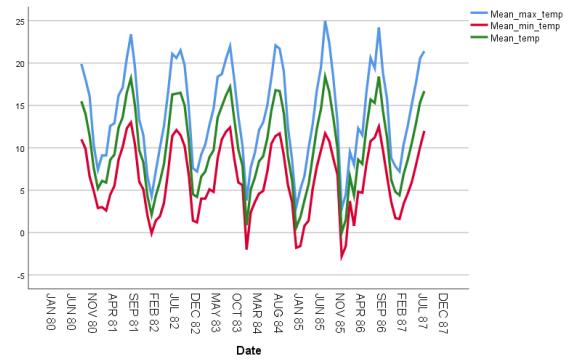


图 3-5d 加拿大西南部温度时间序列 1 阶差分

### b) 周期性

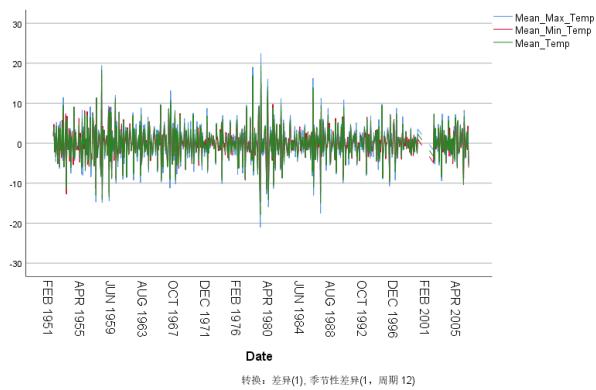


图 3-6a 加拿大北部温度时间序列季节差分

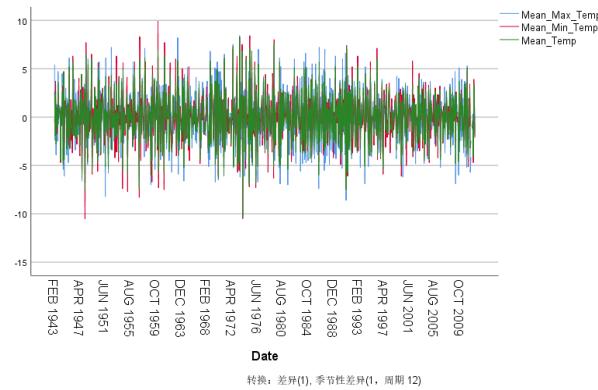


图 3-6b 加拿大东海岸温度时间序列季节差分

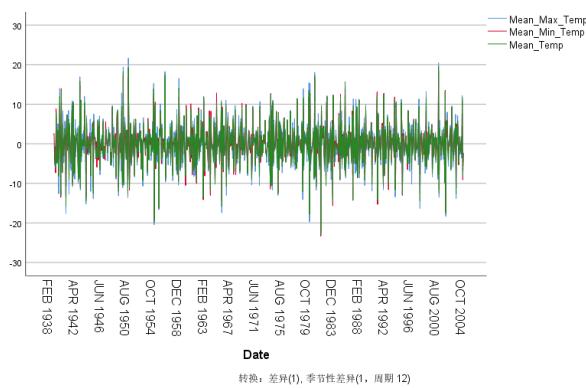


图 3-6c 加拿大内陆温度时间序列季节差分

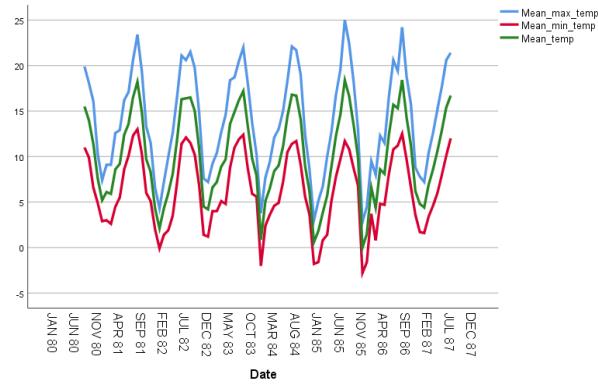


图 3-7c 加拿大西南部温度时间序列季节差分

### c) 自相关性与偏自相关性

相关函数表现出时间序列中任意两个值之间的相关性是如何随着时间间隔而改变的。自相关函数刻画了时间序列相邻变量之间的相关性，偏相关函数则是排除了其它中间变量的影响，真实地反映两个变量之间的相关性，并且二者紧密相连。同时两个相关图所反映的信息在时间序列分析各个方面发挥着关键作用。

加拿大各地平均温度的自相关性与偏自相关性汇总如下：

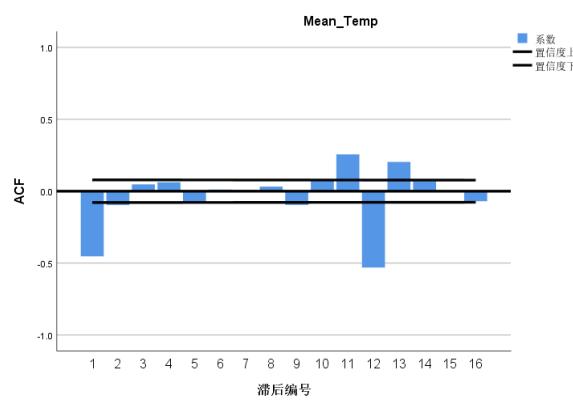


图 3-7a 加拿大北部温度自相关性

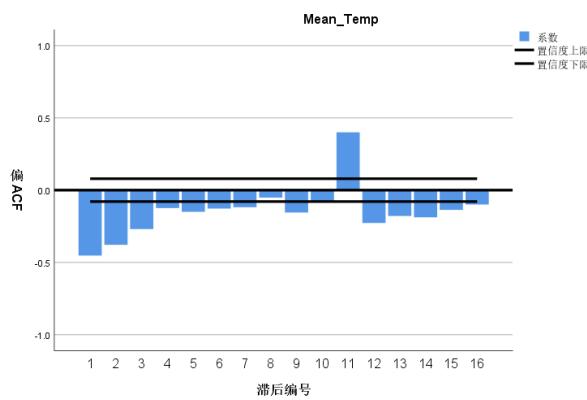


图 3-7b 加拿大北部温度偏自相关性

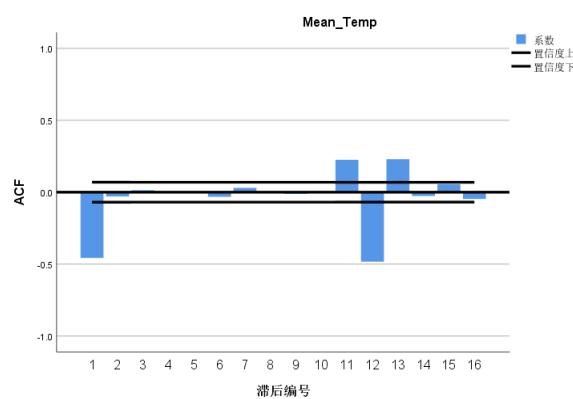


图 3-7c 加拿大东海岸温度自相关性

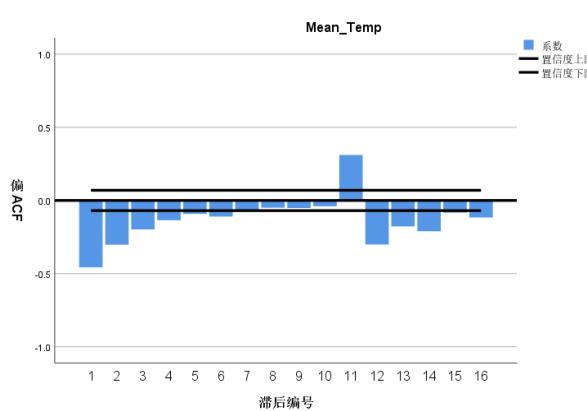


图 3-7d 加拿大东海岸温度偏自相关性

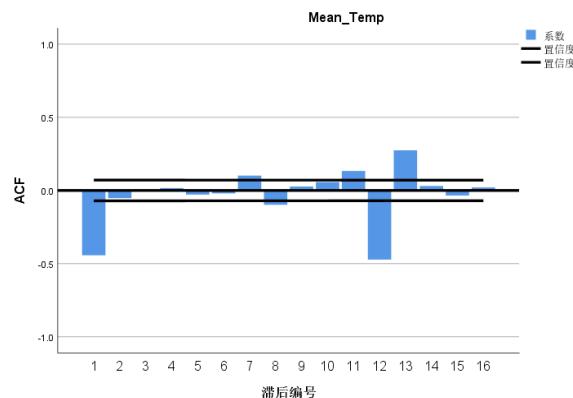


图 3-7e 加拿大内陆温度自相关性

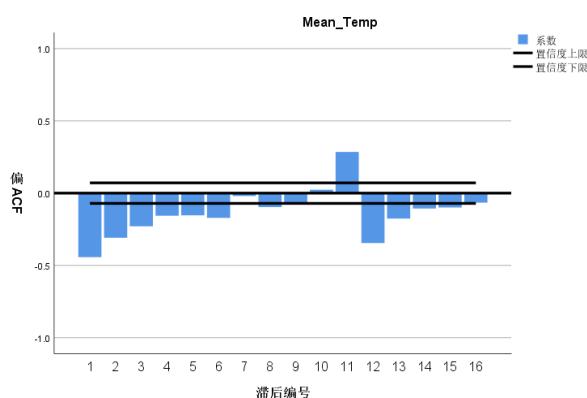


图 3-7f 加拿大内陆温度偏自相关性

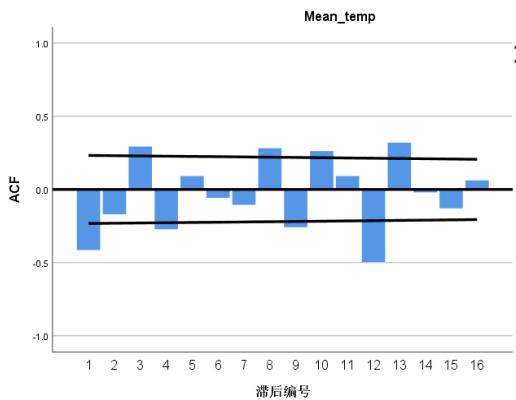


图 3-7g 加拿大西南部温度自相关性

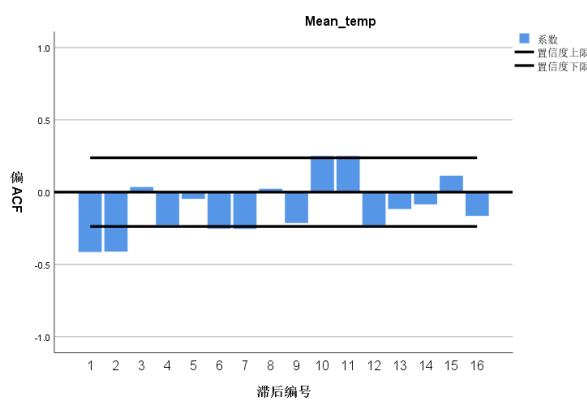


图 3-7h 加拿大西南部温度偏自相关性

自相关性和偏自相关性的分析有助于后续时间序列模型参数的调优提供指导。

## 3.4 问题一中的海洋表面温度特征分析

### 3.4.1 海洋表面温度的数据提取与处理

全球海洋面积约为 3.6 亿平方公里，太过庞大，因此本文对高纬度 ( $60^{\circ}$ - $90^{\circ}$ )、中纬度 ( $30^{\circ}$ - $60^{\circ}$ )、低纬度 ( $0^{\circ}$ - $30^{\circ}$ ) 三个纬度区域内的海洋表面温度分别做平均值统计。

### 3.4.2 海洋表面温度的特征分析

对三个纬度范围内的海洋表面温度做气候倾向率分析如下图所示。

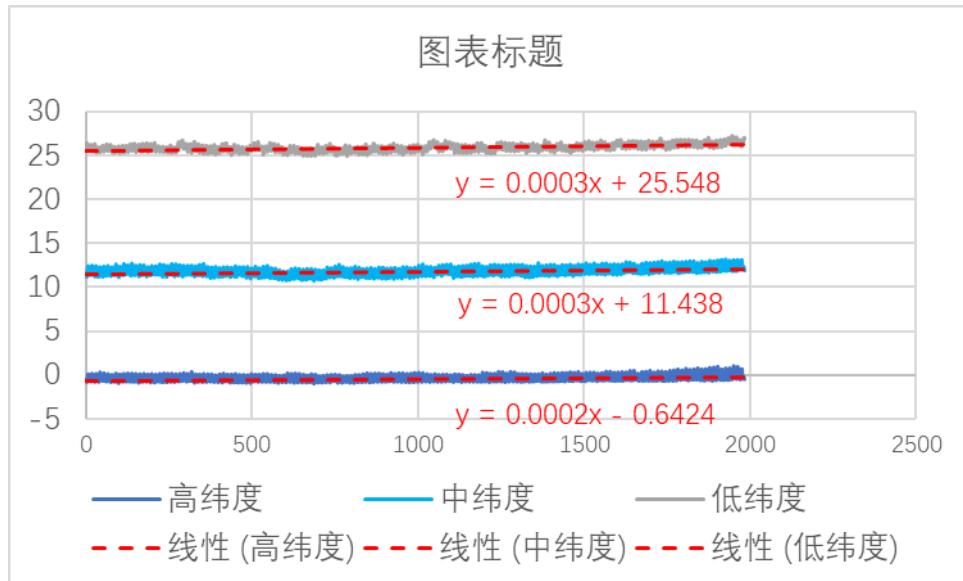


图 3-8 海洋表面温度对时间的倾向率分析

### 3.4.3 海洋表面温度数据的平稳性及周期性检验

#### (1) 基于逐月数据的分析

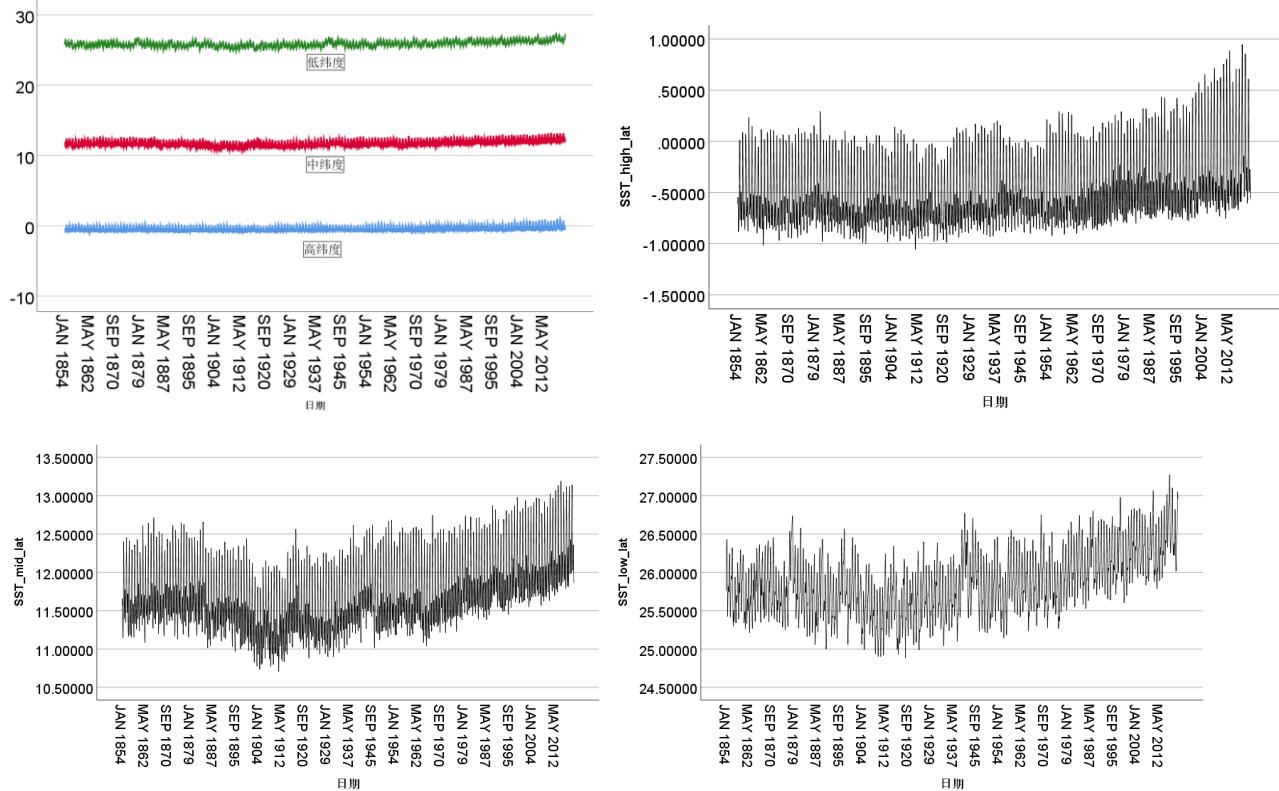


图 3-9 基于逐月数据的海洋表面温度分析（分别为总图和高、中、低纬度的图）

#### a) 平稳性分析

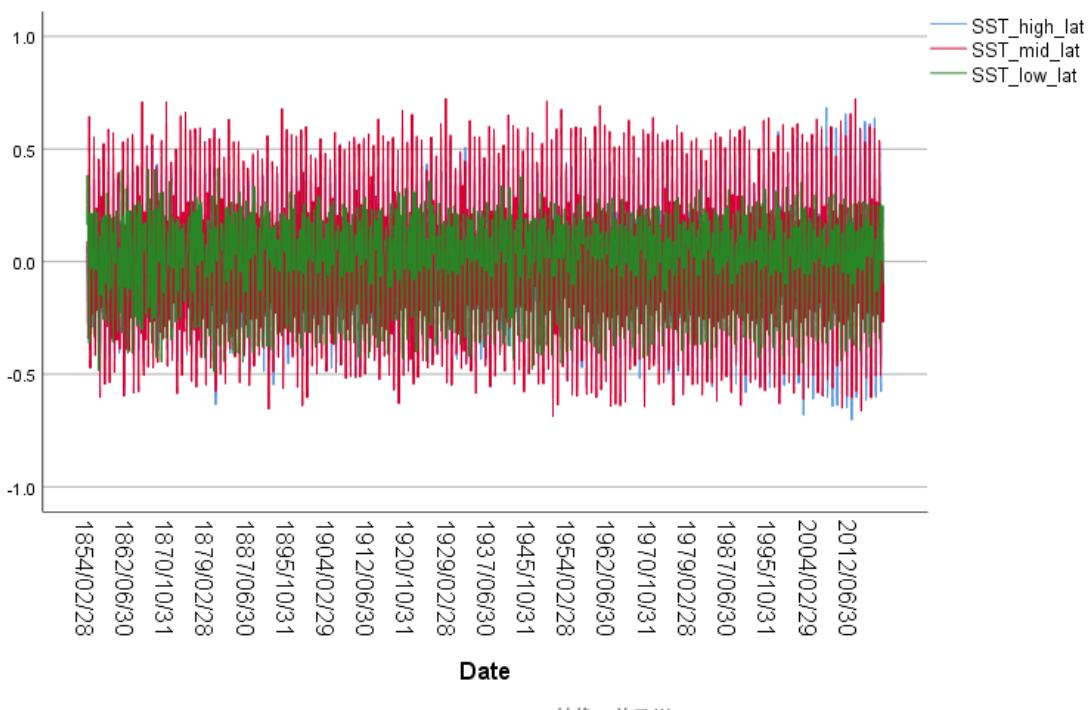


图 3-10 基于逐月数据的海洋表面温度序列

b) 周期性

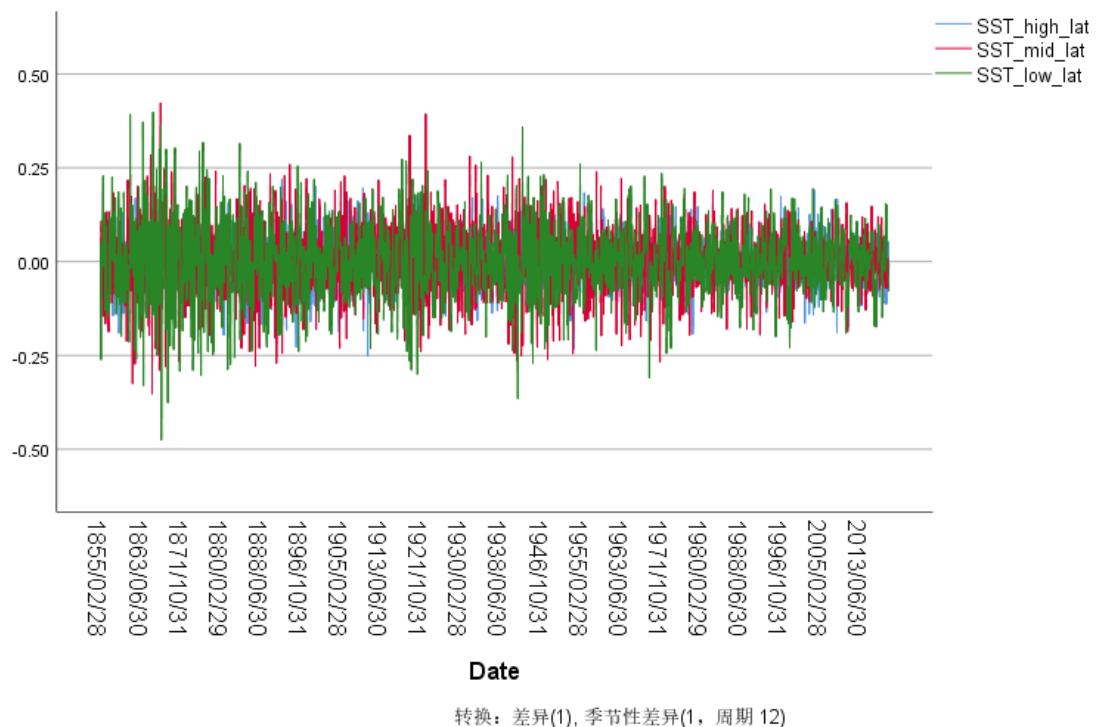


图 3-11 基于逐月数据的海洋表面温度经过差分后的序列

## (2) 基于逐年数据的分析

a) 平稳性

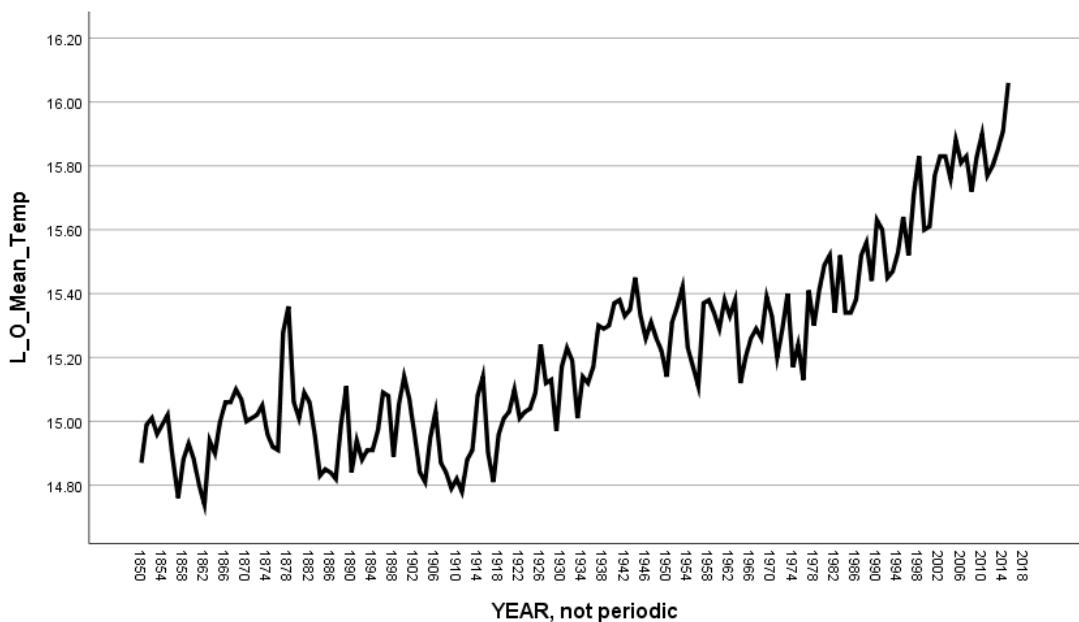
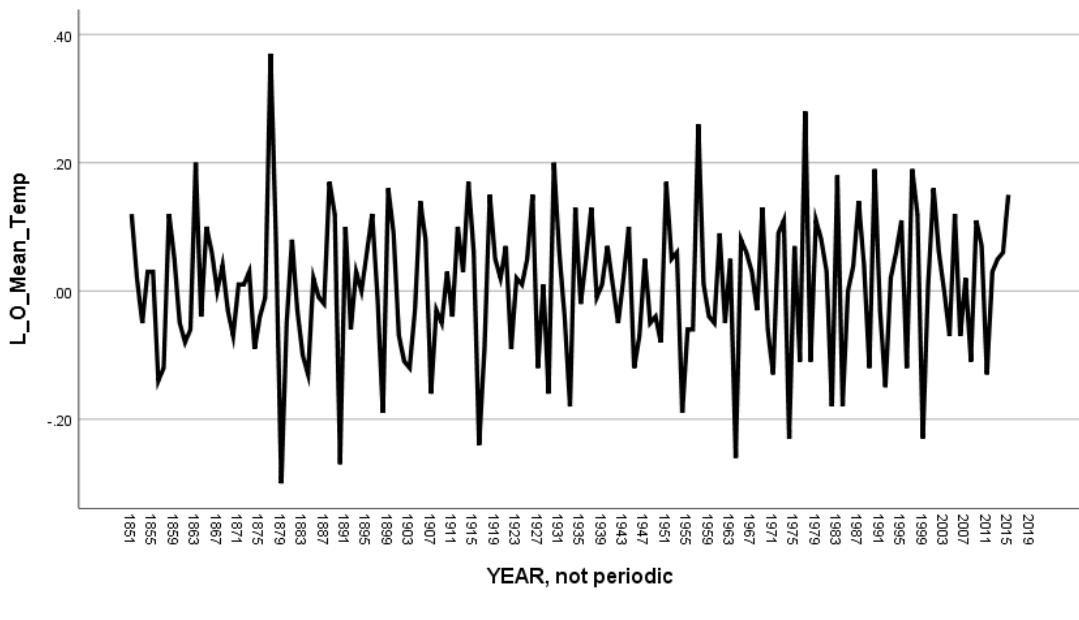


图 3-12 基于逐年数据的海洋表面温度序列



b) 自相关性与偏自相关性

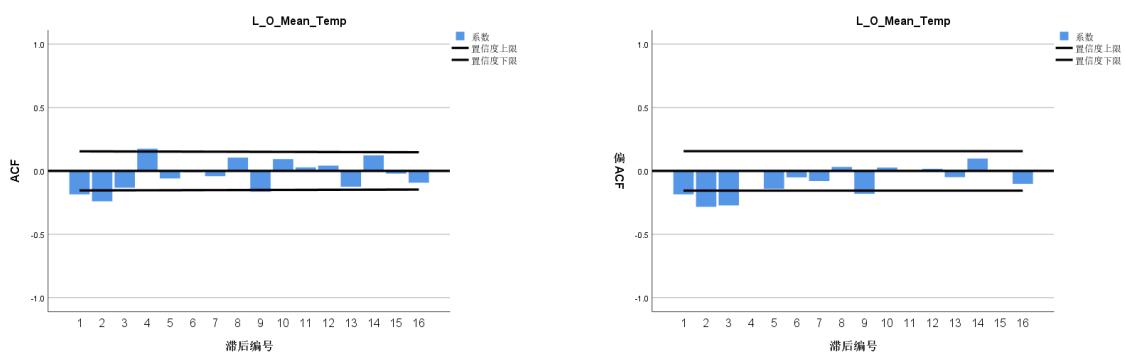


图 3-14 海洋表面温度数据序列的自相关性和偏自相关性分析

## 四. 全球气候变化因素分析及预测模型

全球气候变化是在宏观的空间和时间尺度上的变化。作为系统性的变量，其受到多种因素的影响。寻找、求证全球气候变化的影响因素，为全球气候变化模型的构建及全球未来气候的预测提供重要的理论支持。基于此目的，本章尝试对全球气候数据进行特性分析，从机理和数理统计两方面角度，寻找并求证影响全球气候变化的影响因素及其相关性程度。在此基础上尝试构建全球气候变化模型，并对未来 25 年的气候变化作出预测。

### 4.1 问题二的分析与思路

本题的要求是根据全球气候变化的特征以及全球气候变化的影响因素（包括地球的吸热、散热、海洋的温度变化、空气污染等要素）进行分析，建立相关气候模型，并对未来 25 年的气候变化进行预测。为了了解地球的气候，本文试图调查和综合当前关于气候模型层次结构的一些思想，给出了一些关于预测模型的正式描述，并调查了预测模型用于生成，测试和确认假设的可行方式，结合了当代气候模型的一些特点，用于预测实际过程中的温度变化问题。本题的解决思路是基于历年全球温度数据，分析其特征和变化规律，着重于气候对时间的倾向率、气候变化的平稳性与周期性。探讨影响气候变化的多种可能因素，比如温室气体、太阳辐射、土地利用、臭氧、大气气溶胶、空气污染等等。收集并整理上述因素对气候变化影响的数据，进行数据预处理，分析这些因素的影响程度。然后建立多因素的多元回归气候变化预测模型，设定好参数之后将模型应用于实际问题，检验模型的正确性。最后分别利用 ARIMA 时间序列模型与基于神经网络的预测模型，对未来 25 年全球气候进行预测。

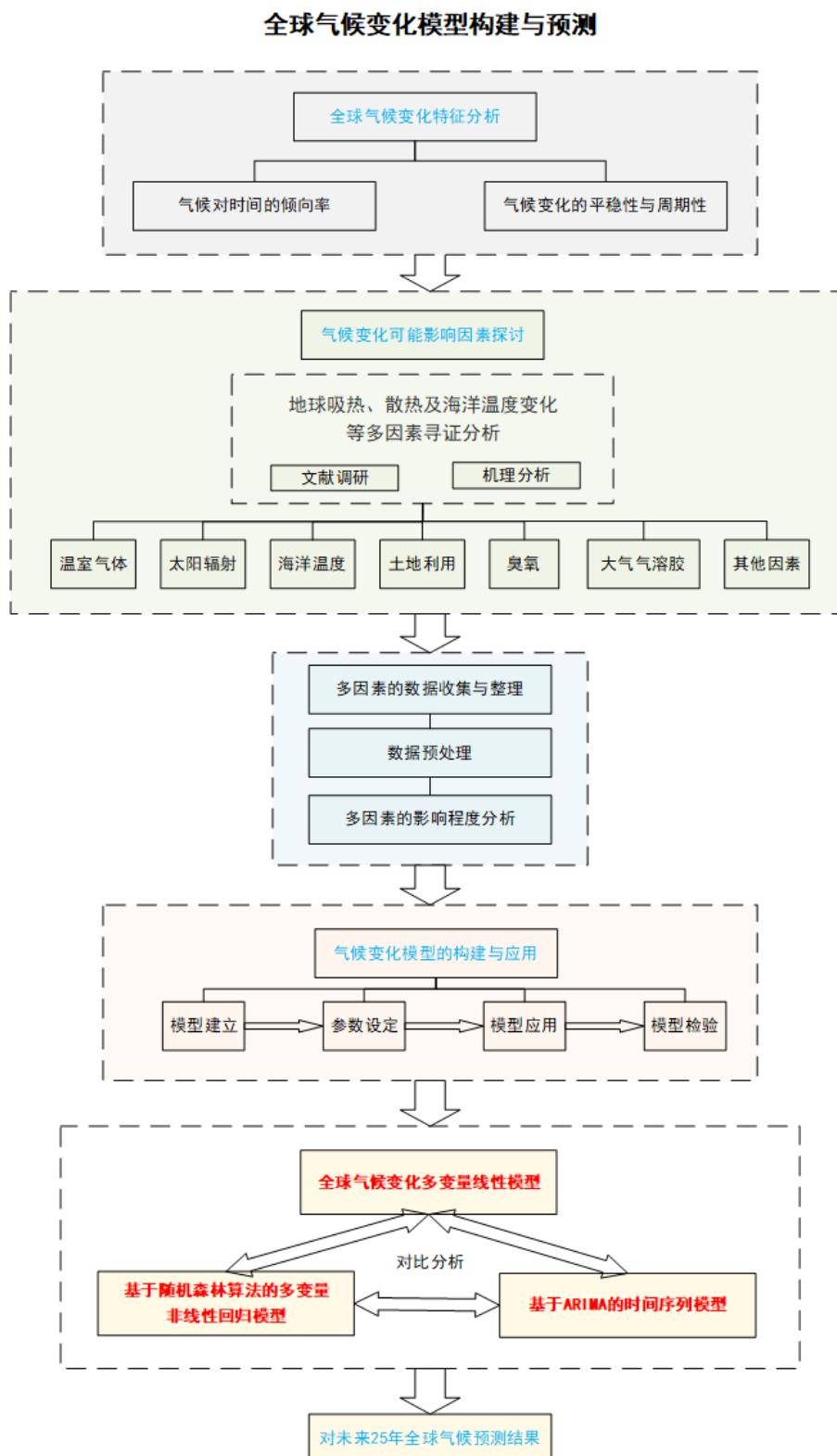


图 4-1 问题二的解决流程

## 4.2 全球气候变化特征

### 4.2.1 全球气候变化特征分析

对全球温度数据做气候倾向率分析图 4-2 所示。

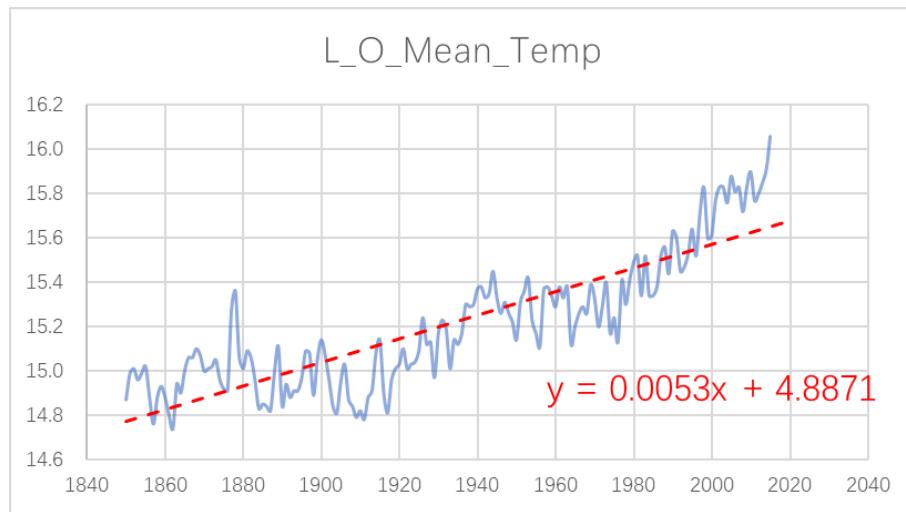


图 4-2 全球温度的气候倾向率

### 4.2.2 全球气候变化平稳性与周期性分析

#### (1) 全球温度的逐年数据平稳性与周期性分析

将全球温度的逐年数据做如下图 4-3 的序列图，可看出序列具有逐渐上升的趋势。

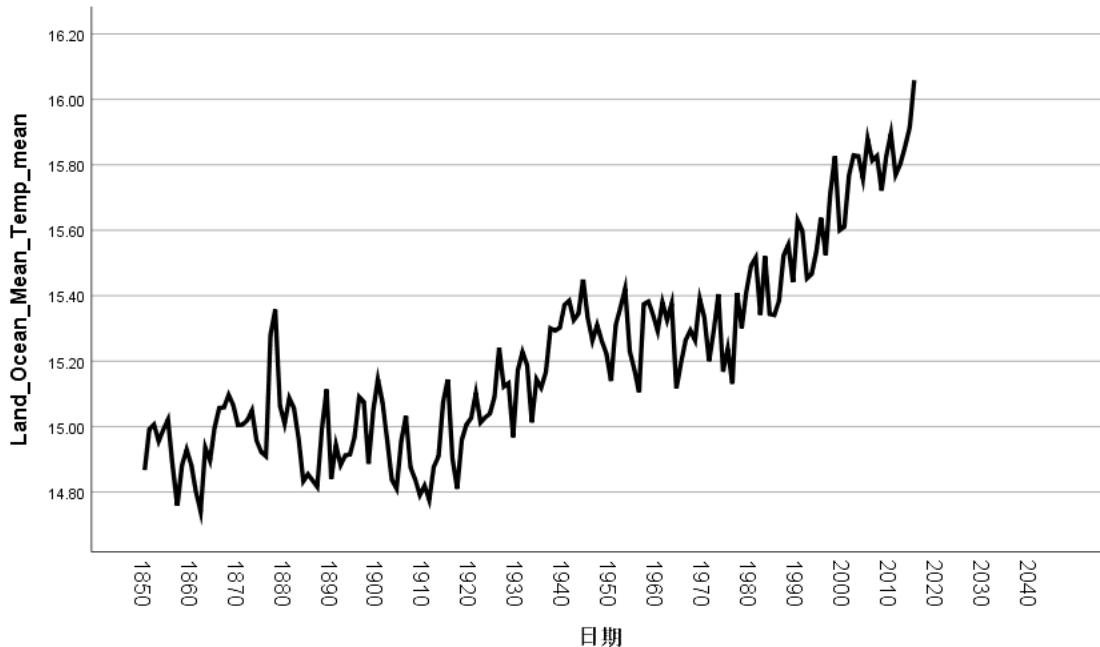


图 4-3 全球温度的逐年数据序列图

将逐年数据做一阶差分处理后（如图 4-3 所示），得到的序列非常平稳。

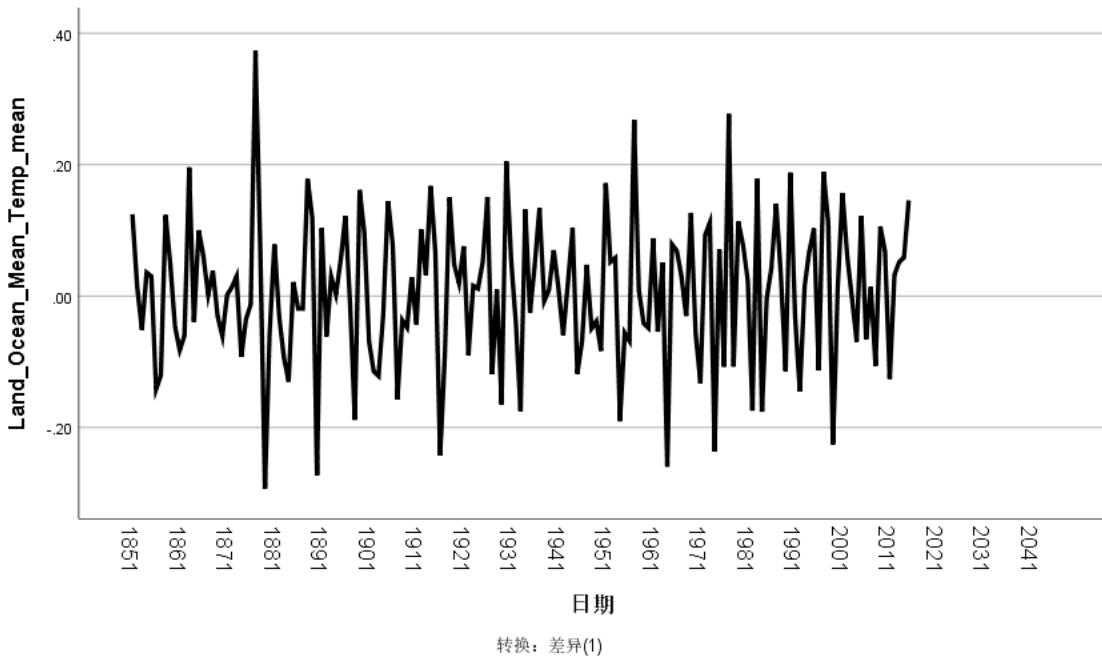


图 4-4 全球温度的逐年数据经一阶差分处理后的序列图

继续对逐年数据序列做自相关性和偏自相关性分析，如图 4-5 所示。自相关性和偏自相关性分析的结果可有助于后续对全球温度的进一步建模和参数调优。

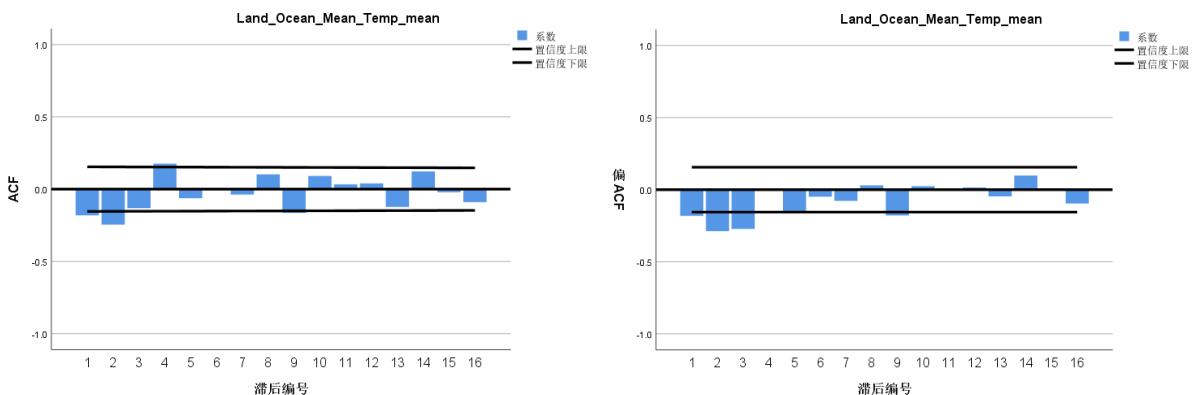


图 4-5 全球温度的逐年数据序列的自相关性和偏自相关性分析

### (1) 全球温度的逐月数据平稳性与周期性分析

将全球温度的逐月数据做如下图 4-6 的序列图，可看出序列具有逐渐上升的趋势。

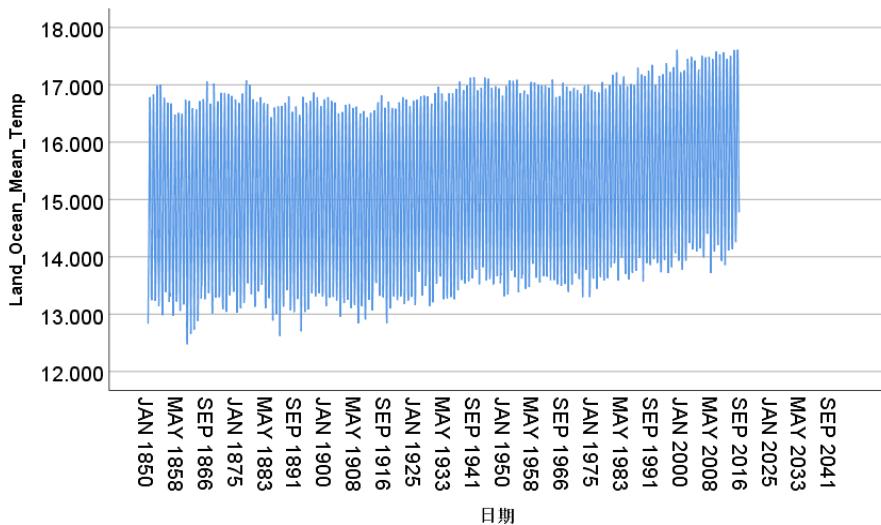


图 4-6 全球温度的逐年数据序列图

将逐年数据做一阶差分及季节性差分处理后（如图 4-7 所示），得到的序列非常平稳。表明逐月数据序列的周期性非常明显。

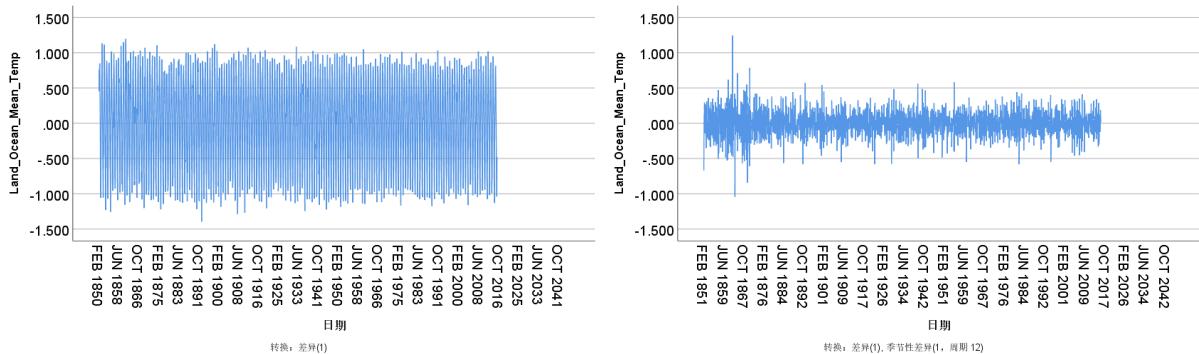


图 4-7 全球温度的逐年数据经一阶差分处理后的序列图

继续对逐年数据序列做自相关性和偏自相关性分析，如图 4-XXX 所示。自相关性和偏自相关性分析的结果可有助于后续对全球温度的进一步建模和参数调优。

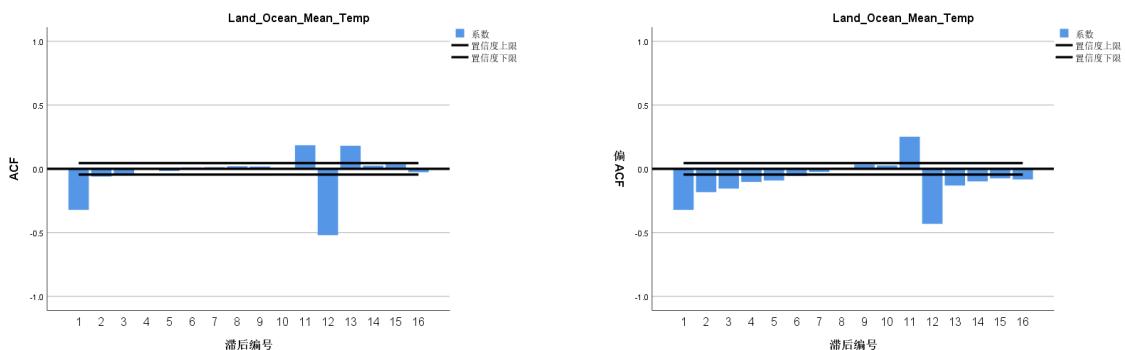


图 4-8 全球温度的逐年数据序列的自相关性和偏自相关性分析

## 4.3 全球气候变化的影响因素分析

影响全球气候变化的因素很多，结合大量的文献调研可以发现，影响因素主要包括温室气体、太阳辐射、海洋作用、土地利用、臭氧、大气气溶胶、空气污染、化石燃料、城市化等多种因素。

### 4.3.1 温室气体

温室气体，即地球大气中起温室效果的气体称之为温室气体，主要有  $\text{CO}_2$ 、 $\text{N}_2\text{O}$ 、 $\text{CH}_4$ 、氟利昂以及水蒸气等。全球气候变暖是由于“温室效应”不断累积造成的。科研人员通过检测发现，没有大气时，地表平均温度大约为  $-23^{\circ}\text{C}$ ，而现实生活中地表平均温度为  $15^{\circ}\text{C}$ ，也就是说，大气的威力强大，温室气体能阻止地球热量的流失，从而使地表温度直接提高。

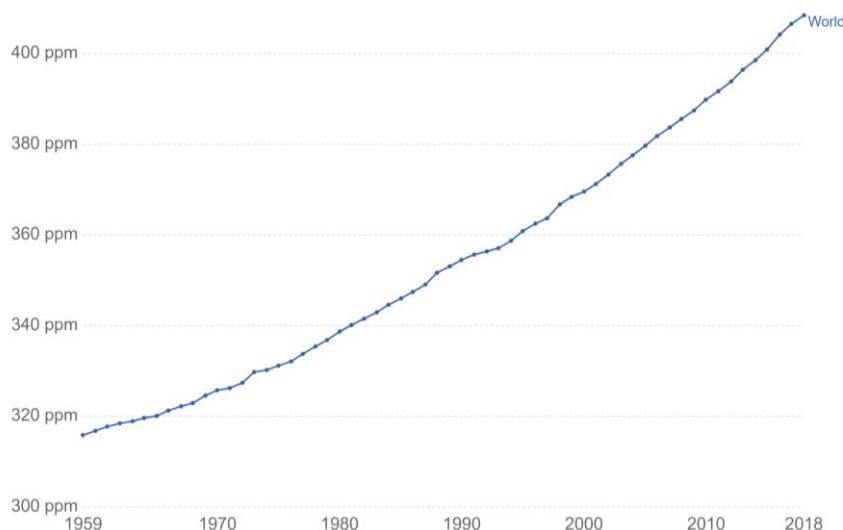


图 4-9 全球不同时间段  $\text{CO}_2$  浓度变化

在冰川时期，空气中的  $\text{CO}_2$  的浓度大约是 200 ppm。在比较温暖的间冰期大概是 280 ppm。从上面的图可以看出，目前大气中  $\text{CO}_2$  比过去 40 万年的任何时期都要高出很多。自工业革命以来，由于人类活动所造成的  $\text{CO}_2$  浓度的释放呈现出井喷式上升趋势。而据夏威夷莫纳罗亚气象台的传感器数据显示，2019 年 5 月 11 日，大气中的  $\text{CO}_2$  浓度达到了 415.26 ppm，这在人类历史上是首次发生。

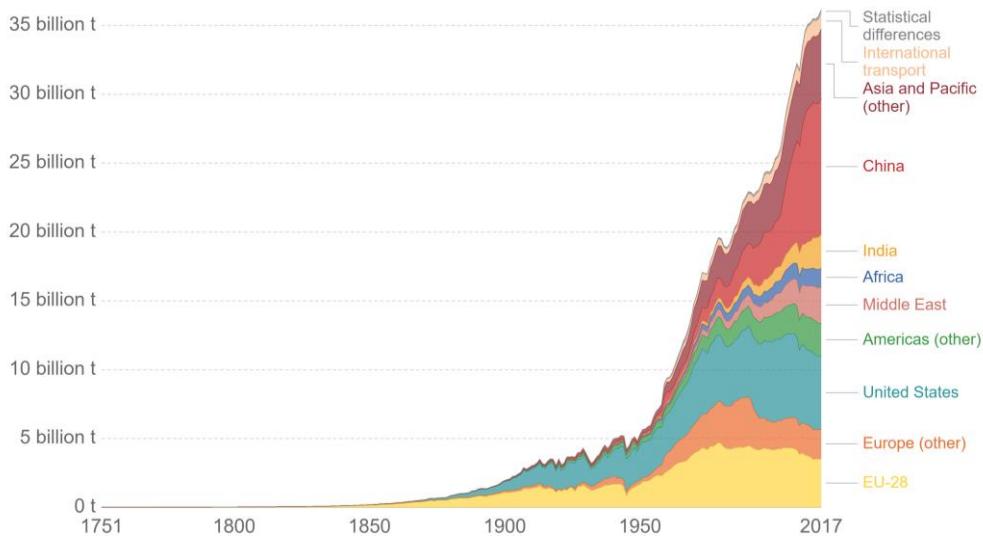


图 4-10 全球各地区不同时段 CO<sub>2</sub> 排放量

温室气体造成的影响就如同多米诺骨牌一样，形成了一种恶性循环，导致全球温度上升。

(1) 随着两极冰盖的加速融化，缺少了冰层的反射，就会有更多的阳光到达地表，南北两极冰川融化。

(2) 对水轮回的影响主要体现在全球降雨量可能会增加，而区域性降雨量的改变仍为可知，某些地区可能会出现持续降雨的现象，降雨量增多；有些地区则降雨量消减。另外，温度的升高会增加水分的蒸发，从而对地面上水源的运用带来压力。

(3) 对海洋生态的影响主要体现在海洋生物的减少，特别是贝壳类生物数目的减少。河口水水质变咸会消减淡水鱼品种的数目，相反该地区海洋鱼类的品种数目也可能会增多。

### 4.3.2 太阳辐射

太阳辐射是地球表面最终的能量来源，与气候变化具有紧密的联系。太阳辐射量到达地面，但可能由于多种因素的共同作用，如云的散射和反射、气溶胶的吸收和散射、大气分子的吸收和散射等使太阳辐射被削弱，从而在不同程度上引起到达地面太阳辐射的变化，对地球的生态系统及气候造成了深渊影响。

近 50 年来，地面太阳辐射的变化特征：

在 1950-1990 年期间，全球大部分地区的地面太阳辐射量呈现下降的趋势。Liepert 等给出的太阳辐射量平均减少幅度为每 10 年 1.3% 或  $7 \text{ W/m}^2$ ；Stanhill 等认为每 10 年 2.7% 或  $5 \text{ W/m}^2$ 。但是，从 1980-2000 年，北半球、南半球等地均出现太阳辐射量逐渐增加的现象，根据 BSRN 资料显示，太阳辐射的增加量平均每年大约为  $0.66 \text{ W/m}^2$ 。[2]

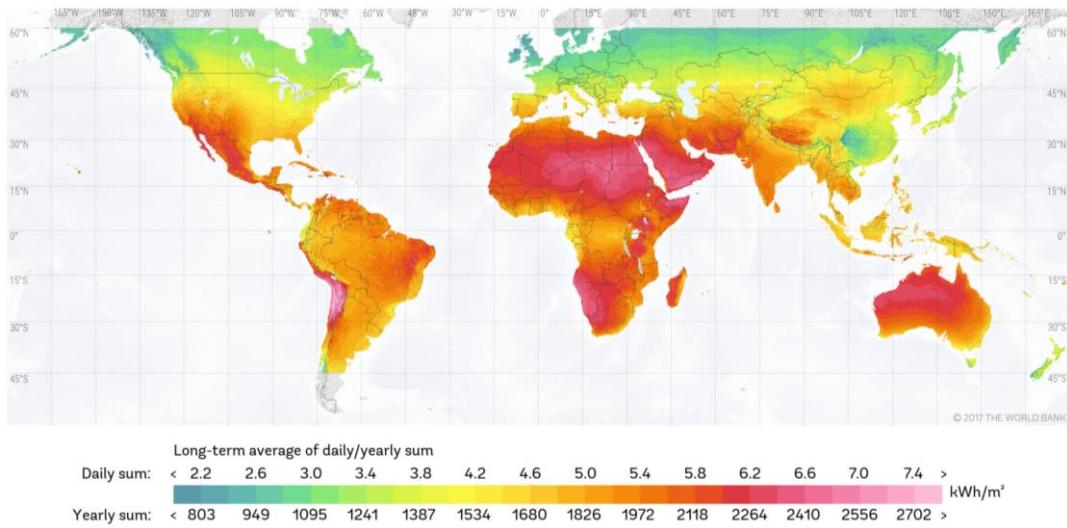


图 4-11 全球水平辐射的全球地图

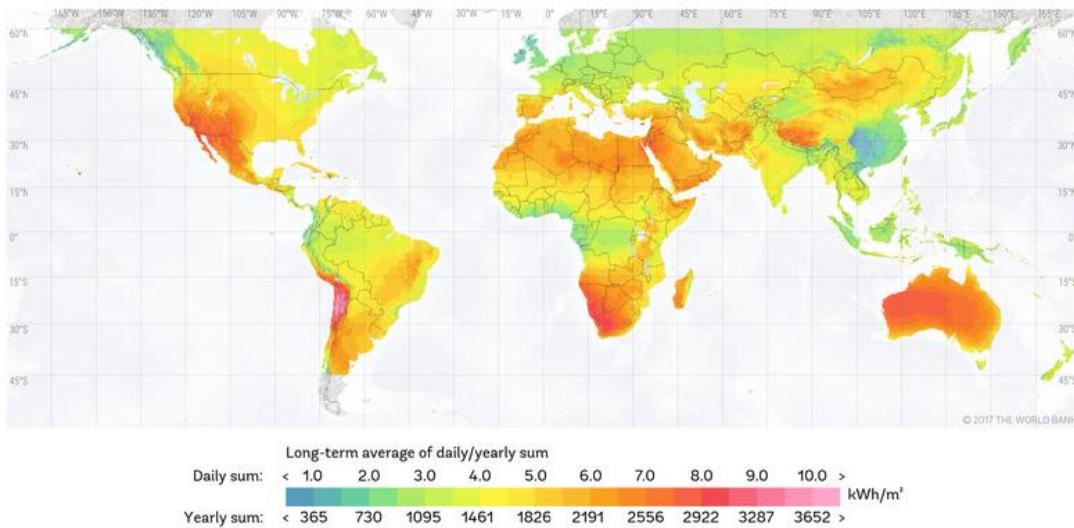


图 4-12 直接法线辐射的全球地图

### 4.3.3 海洋作用

海洋是气候系统的最重要组成部分之一。海洋可以吸收大部分的温室气体，而储存在气候系统中的多余能量，则会导致热膨胀以及海平面的上升。因此，海洋系统在地球的能量平衡中起着至关重要的作用。

海洋在全球变暖中的作用主要体现在，海洋作为气候系统的重要因子，参与全球变暖的整个过程，并且在其中发挥调节作用。随着全球变暖的加剧，海洋的温度也在逐渐上升。同时，升温也会导致海冰的融化。自 1979 年以来，北极海冰已经减少了大约 20%，海冰的消融也将不可避免地为海洋注入淡水，在海洋表层形成“淡水帽”，从而促进热盐环流的减弱。一旦热盐环流减弱或崩溃，将对全球气候造成不可预估的影响。[10]

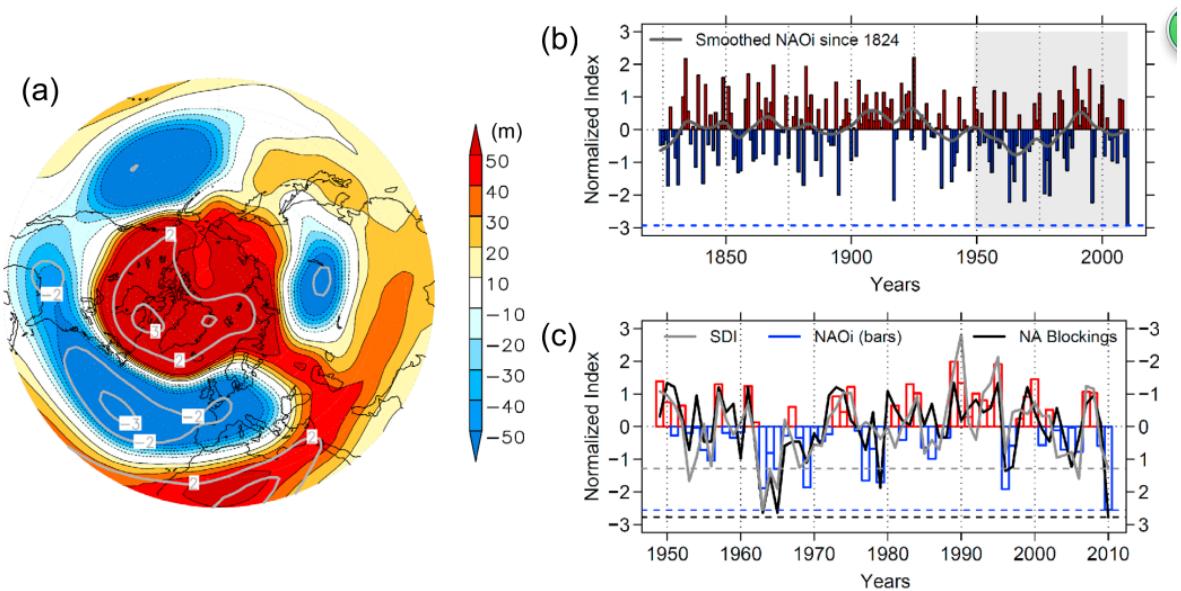


图 4-13 海洋表面温度

在 20 世纪 90 年代以前，绝大多数的海洋测量值都不超过 700m。然而，从 2006 年以来，浮标全部部署深度达 2000m。研究发现全球 2000m 以上的海洋在此期间吸收的热量为  $0.30 \pm 0.06 \text{ W/m}^2$ , 2000m 以下为  $0.028 \pm 0.026 \text{ W/m}^2$ , 年代波动较大。自 1871 年以来的总 OHC(海洋热量含量)变化估计为  $436 \pm 91 \times 10^{21} \text{ J}$ ，且在 1921-1946 ( $145 \pm 62 \times 10^{21} \text{ J}$ ) 年期间增加，与 1990-2015 年期间一样大。通过与直接估算进行比较，推断在 1955-2017 年期间，由于海洋运输变化引起的热量传输，导致了一半的大西洋变暖以及低纬度到中纬度的热膨胀海平面上升。

#### 4.3.4 土地利用

土地利用主要是指人类活动作用于自然环境，是土地覆被变化的最主要的驱动因子。土地覆被的变化，如城市化建设、农业活动、森林的采伐、土地退化等，都能引起温室气体排放吸收、地面反射率以及蒸发作用的变化，从而引起了全球尺度上气候的变化。

下面的图表显示了长期的人类土地利用（自公元前 10000 年以来），并详细说明了耕地、牧场和城市用地的总土地面积（公顷）的变化。

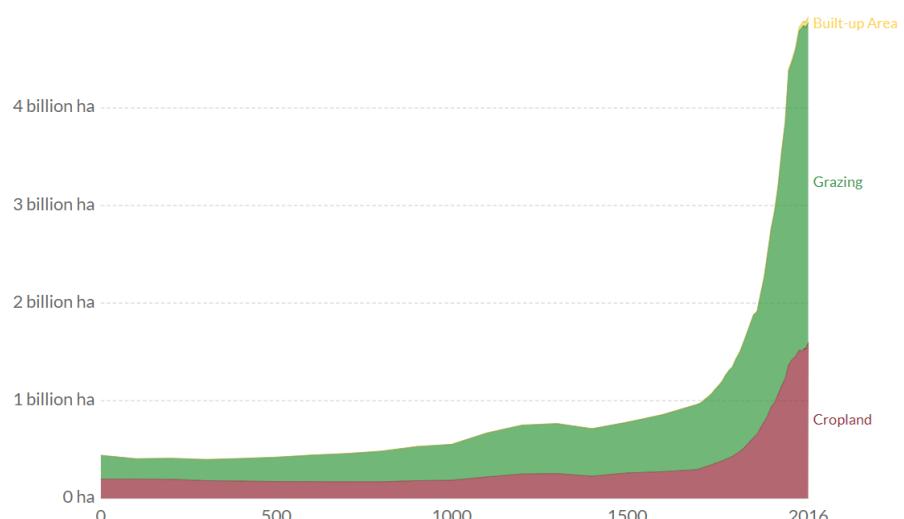


图 4-14 世界范围内的长期土地利用

土地利用对全球气候造成的影响：

(1) 土地利用与温室效应之间具有紧密的关系。目前，土地利用的气体释放量占了全球温室气体释放总量的三分之一以上，大约是四分之三的 CH<sub>4</sub> 释放总量。土地利用主要是通过改变全球温室气体的收支平衡，主要体现温室气体排放量增加。

(2) 土地利用和水热平衡之间的关系。土地利用首先是引起了土地覆被的变化，从而引起了地表反照率、植被叶面积指数、粗糙度以及地表植物覆盖度发生了明显的改变。地面是大气的主要加热源，一旦地面热状况发生变化，必将破坏大气原来的热量分布平衡。

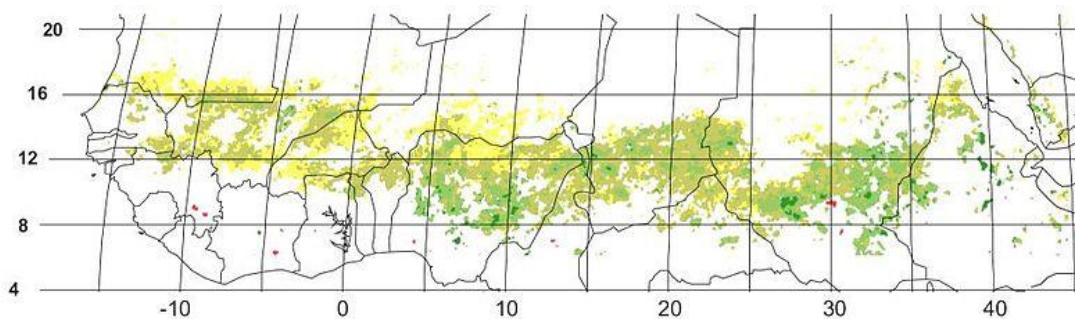


图 4-15 萨赫勒绿化

### 4.3.5 臭氧层

臭氧消耗物质（ODS）是含有氯和（或）溴的卤素气体，它有可能分解平流层中的臭氧，导致臭氧层的消耗。现有的臭氧消耗物质有很多种，包括氯氟烃（CFCs）、氢氟氯烃（HCFCs）、氯甲烷和溴化物以及哈龙。臭氧消耗物质有两类排放源，分别是自然排放源和人为排放源。从下图可以看到 1960 年以后臭氧消耗物质的排放量。从 1960 年到 1980 年代末，排放量迅速增加（增加了三倍以上），随后的几十年里，排放量也同样迅速减少。到 2010 年，排放量已恢复到 1960 年的水平。

下图所示为自然排放水平（在此期间大致一致），以及自然和人为排放总量之和。在这里，我们看到消耗臭氧层的排放量出现了明显的增长峰值减少趋势，从 1960 年到 1980 年代末，排放量迅速增加（增加了三倍以上），随后的几十年里，排放量迅速减少。到 2010 年，排放量已恢复到 1960 年的水平。

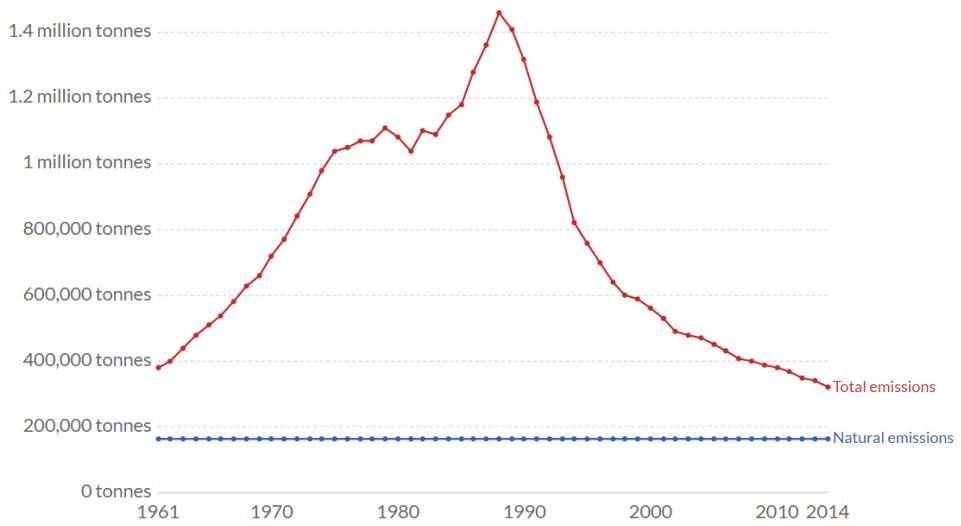


图 4-16 全球消耗臭氧层物质排放量

下图所示为全球消耗臭氧层物质消费量的变化曲线，可以看出自 1986 年以来消耗臭氧层物质消费量下降幅度之大。在 1987 年增长 28% 之后，全球消耗臭氧层物质消费量快速下降，到 1995 年下降了 60% 以上；到 2000 年下降了 80%；到 2016 年下降了 98-99%。

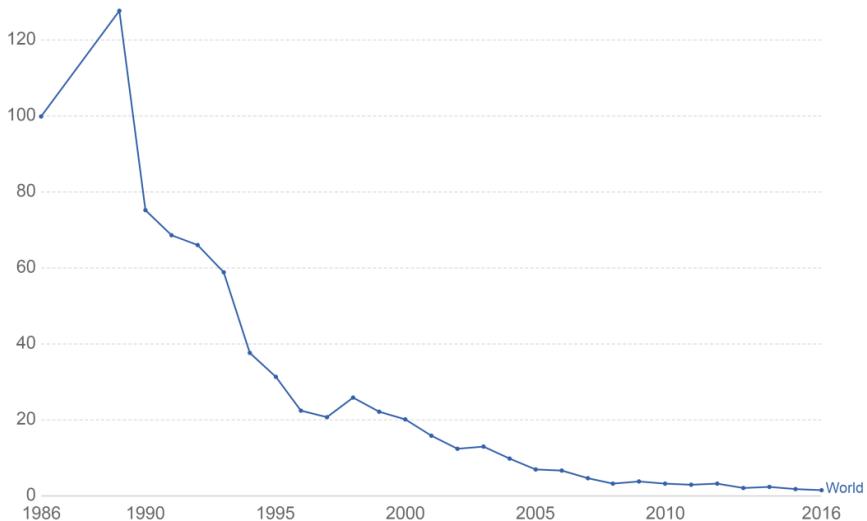


图 4-17 不同时间段臭氧消耗物质的消费量

从下图我们可以看到 1979 年至 2017 年南半球平流层臭氧的平均浓度，进而分析人造臭氧消耗物质的排放对平流层臭氧浓度的影响。从 1979 年到 90 年代初，南半球的平流层臭氧浓度下降到了 100 DU 的“臭氧洞”水平。自 20 世纪 90 年代以来的几十年，浓度持续接近 100 DU 左右（或低于 100 DU）。然而，自 2010 年以来的最近几年，臭氧浓度开始缓慢上升。

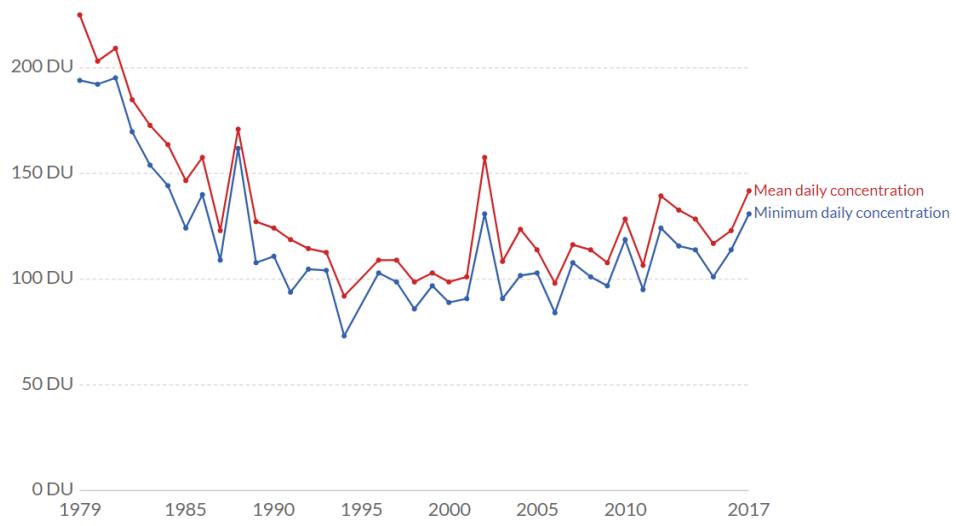


图 4-18 南半球平流层臭氧浓度

下图所示为南极上空臭氧空洞的最大面积和平均面积变化曲线，单位为平方公里( $\text{km}^2$ )。从图中可以得到，自 1979 年以来，南极臭氧空洞面积明显增加，在 21 世纪初达到最大值 3000 万平方公里。但自 20 世纪 90 年代末以来，臭氧空洞面积基本稳定在 2000 万至 2500 万平方公里之间。

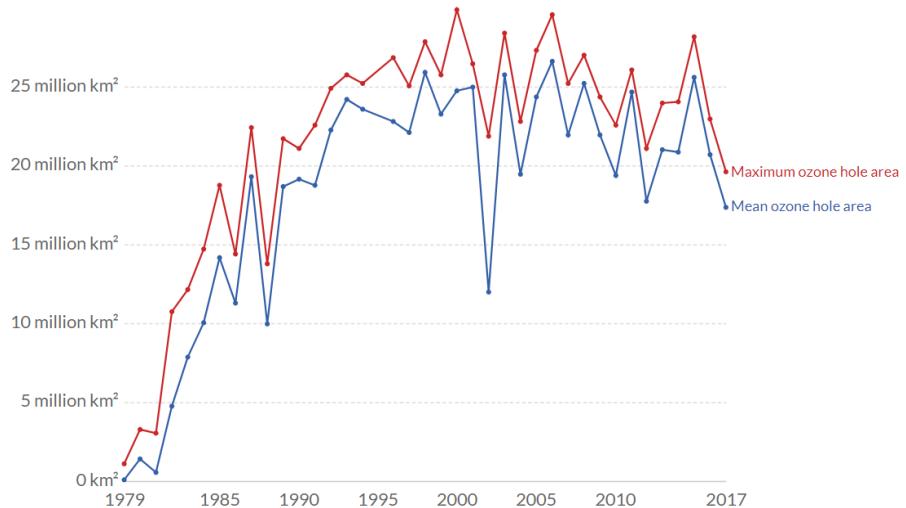


图 4-19 南极臭氧空洞面积

#### 4.3.6 大气气溶胶

导致全球变冷的主要因素是大气气溶胶。进入地球大气的能量，可以在大气内部被散射、吸收，反射，由于大气气溶胶的存在，引起的这些过程的改变叫做气溶胶辐射强迫。其中，黑碳气溶胶可产生  $0.10[\pm 0.10]\text{W / m}^2$  的辐射强迫。其它气溶胶粒子，例如：硝酸盐、硫酸盐、矿物沙尘等，总直接辐射强迫和间接辐射强迫分别为  $-0.50[\pm 0.40]\text{W / m}^2$  和  $-0.70[-1.1, +0.4]\text{W / m}^2$ ，二者总计达到  $-1.2\text{ W / m}^2$

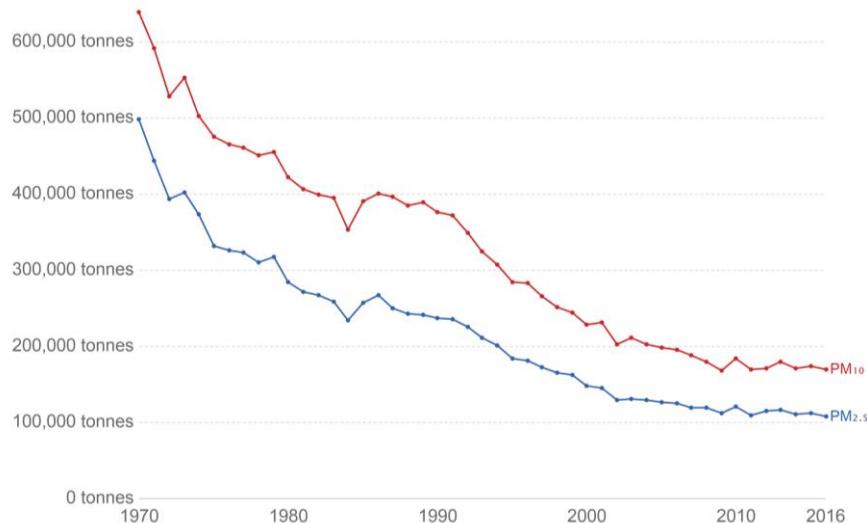


图 4-20 颗粒污染物的排放

#### 4.3.7 空气污染

空气污染可以被定义为有害性物质向大气的排放。广泛意义上的污染物包括以下几种：二氧化硫（SO<sub>2</sub>），氮氧化物（NO<sub>x</sub>），臭氧（O<sub>3</sub>），颗粒物（不同尺寸的微小悬浮颗粒），一氧化碳（CO）和挥发性有机化合物（VOCs）。各种污染物的来源各不相同，大多数的污染物都是作为燃料燃烧和工业活动的副产品被排放。空气污染有很多负面影响，包括破坏生态系统、粮食作物和建筑环境，甚至损害人体健康。世界卫生组织曾发表声明，称空气污染是威胁人类身体健康程度最大的环境风险。

追溯人为空气污染的历史源头时，我们通常以 SO<sub>2</sub> 为指标进行数据查询。根据煤炭这一化石燃料的生产、进出口和工业生产加工等数据，对历史上 SO<sub>2</sub> 排放量可以进行大致估计。由下图可以看出，工业化过程伴随大量含硫燃料的燃烧，导致 SO<sub>2</sub> 排放量出现一个关键转折点。北美的二氧化硫排放量在 1970 年达到高峰，欧洲在 1980 年达到高峰，南美在 1990 年达到高峰。世界其他地区由于工业化起步较晚，二氧化硫排放量的上升推迟到二十世纪。

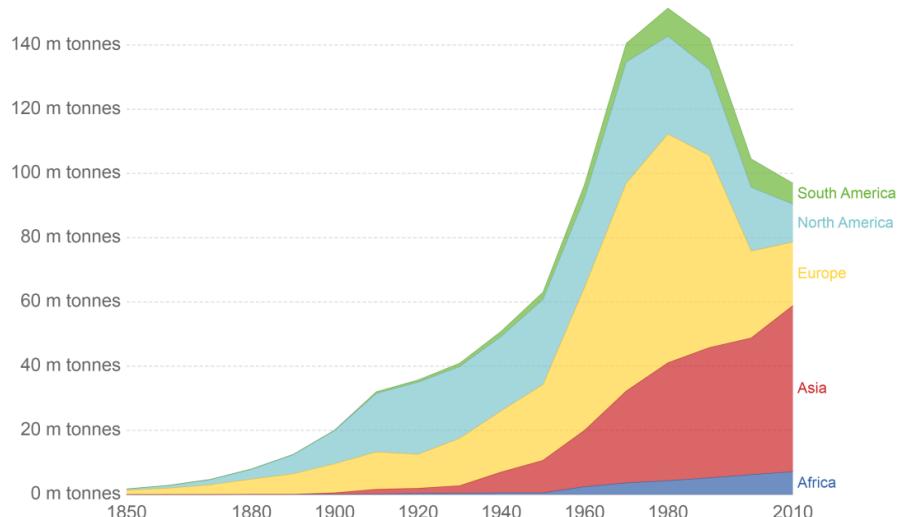


图 4-21 全球各地区不同时间段 SO<sub>2</sub> 排放量

在研究空气污染物对人体的伤害程度时，我们通常关注 PM2.5 的作用，特别是直径小于 2.5 微米的小颗粒物。这种尺寸的微小颗粒物被人体吸入后可以穿透肺部，从而对人体呼吸系统的健康造成不良影响。**错误!未找到引用源。** 环境 PM2.5 污染的人口加权暴露定义为一个国家人口暴露于空气动力学直径小于 2.5 微米的悬浮颗粒浓度的平均暴露水平。由下图可以看出，不同国家的暴露量之间有很大差别。印度和中国的 PM2.5 人口加权平均暴露浓度居高不下，尤其印度还处于上升趋势，在 2014 年之后甚至超过  $70 \mu\text{g}/\text{m}^3$ 。而工业化起步较早的一些发达国家例如英国、美国，其 PM2.5 暴露浓度较低且总体处于下降趋势。

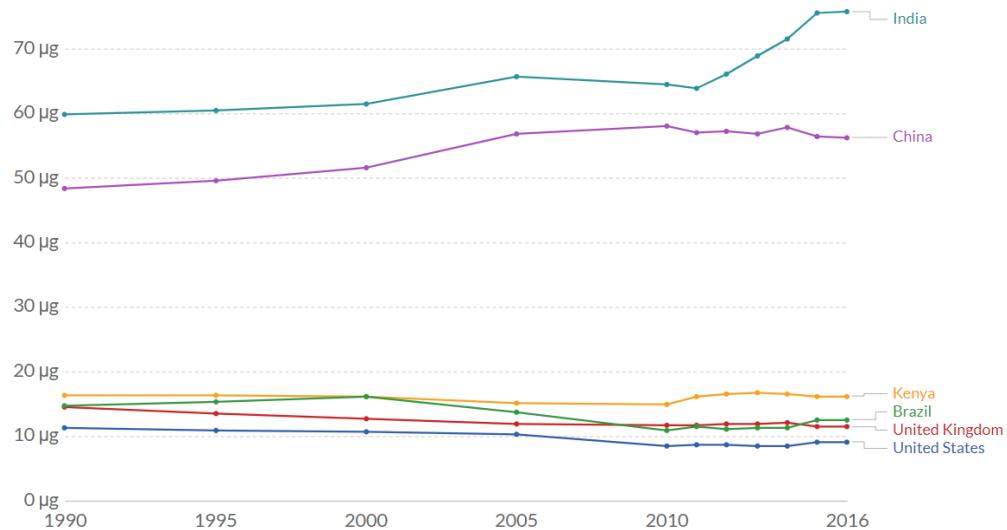


图 4-22 部分国家不同时段颗粒物（PM2.5）人口加权平均暴露浓度

污染造成的死亡主要有三种原因：室内燃烧固体燃料（家庭污染）、暴露于室外环境的臭氧（O<sub>3</sub>）和室外环境颗粒物（PM）污染。下图展示了部分国家/地区归因于室外空气污染的死亡人数绝对值。与 PM2.5 暴露浓度相对应的，中国和印度两国由空气污染导致的死亡人数较高。自 1990 年以来的一段时期，中国的与污染相关的死亡人数增长似乎在放缓，仅在 2010 年出现小幅增长。相反，印度的由室外空气污染导致的死亡率仍在持续上升。[11]

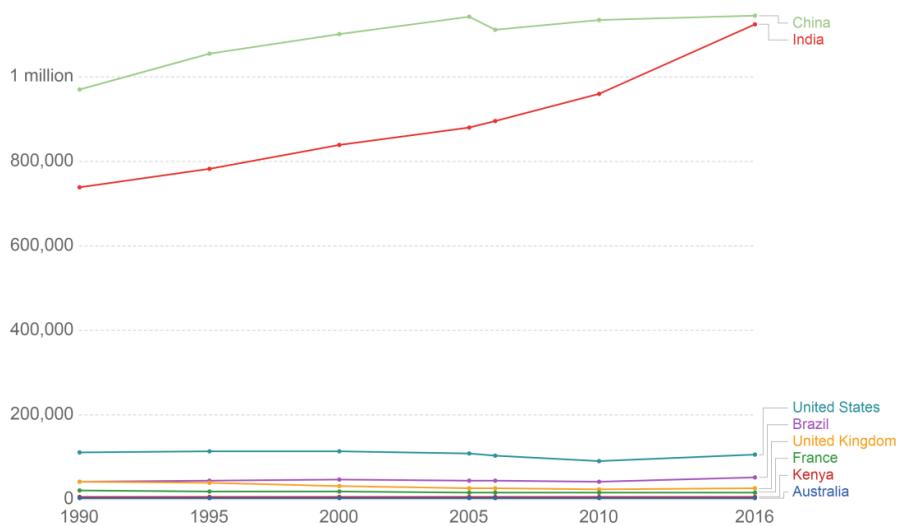


图 4-23 部分国家不同时段室外空气污染导致的死亡人数绝对值

### 4.3.8 化石燃料

化石燃料包括煤、石油和天然气，化石能源是工业革命的根本动力。然而，化石燃料也有负面影响，它是空气污染物的主要来源，也是二氧化碳和其他温室气体的排放来源。下图所示为1800年以后全球化石燃料的消费量，包括煤炭、石油和天然气。总体上看，在此期间全球化石能源消费量增长了1300多倍。直到1860年代开始消耗原油，煤炭是第一个也是唯一的化石来源。天然气的生产开始于1880-90年代。20世纪化石能源消费出现了巨大的多样化，煤炭占总产量的比例从1900年的96%下降到2000年的不足30%。如今，原油是最大的能源，约占化石能源的39%，其次是煤炭和天然气，分别占33%和28%。

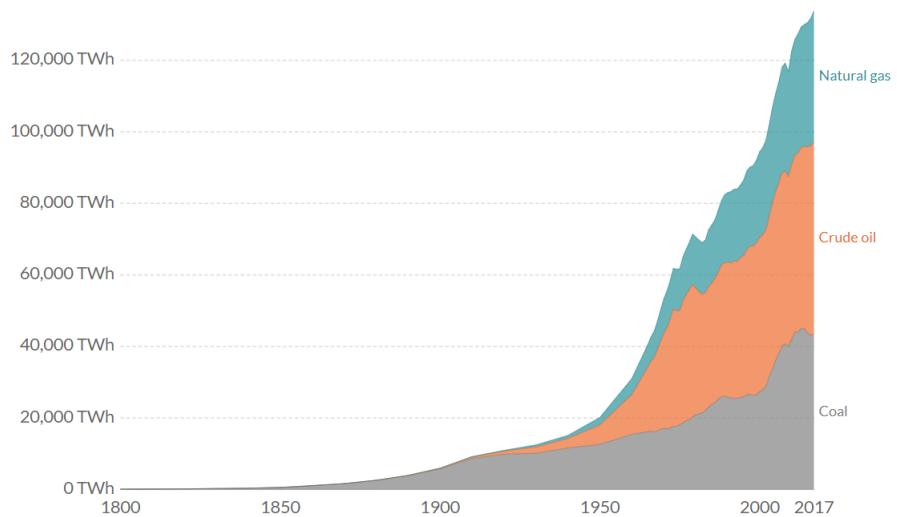


图 4-24 全球化石燃料消耗量

人们对燃料储备提出的一个常见的问题：我们什么时候会用完？当然，化石燃料是有限的资源，如果无限期地消耗，我们最终会耗尽化石燃料。根据目前已知的储量，对化石燃料剩余年数的最佳估计如下图所示。尽管我们经常担心全球燃料资源太少，导致资源稀缺危机，但从气候变化的角度来看，我们实际上拥有太多资源了。如果我们想要达到将全球平均气温上升幅度保持在低于2°C的目标，我们就必须把大部分（高达80%）的化石燃料留在地里。

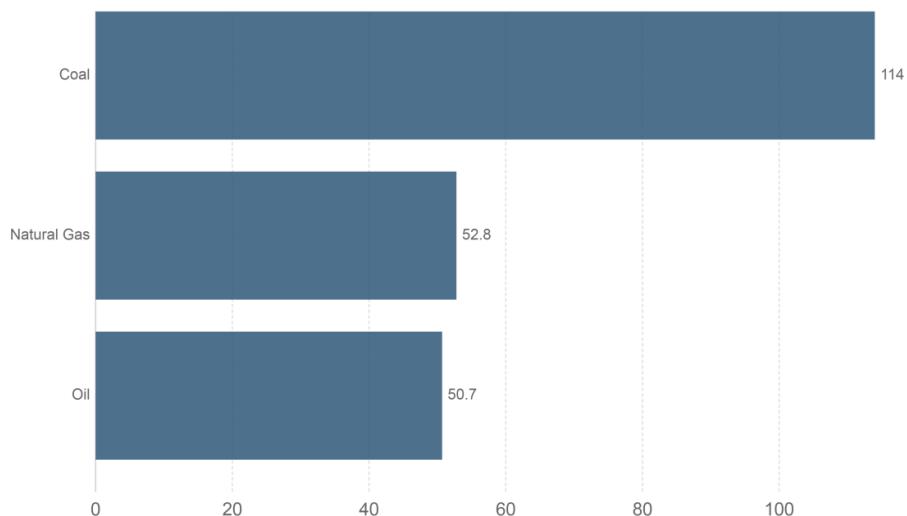


图 4-25 化石燃料剩余年数

### 4.3.9 城市化

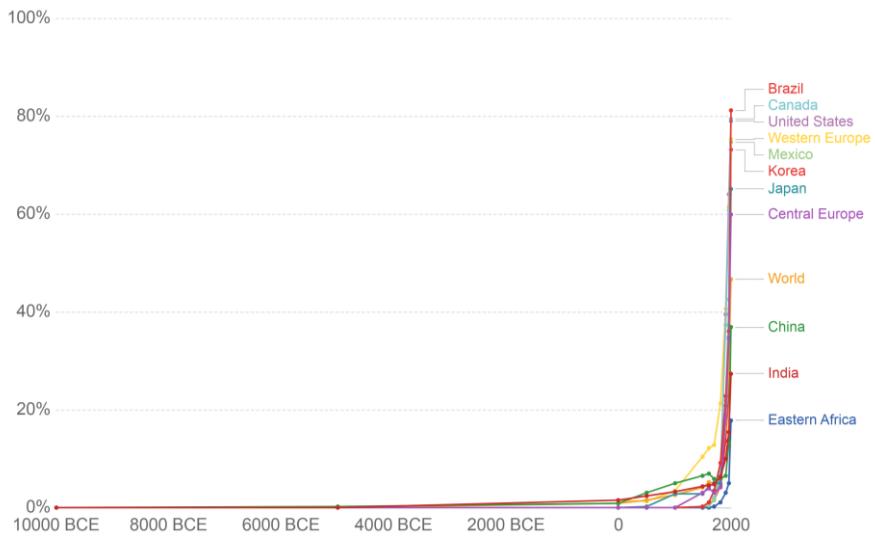


图 4-26 居住在城市地区的人口比例

从上图中可以看到各地区在不同时间内生活在城市化地区的人口比例，居民地扩展对地区温度影响的最明显例子是城市的热岛效应，也有可能从局部影响到全球区域。在某些地区（如北京和芝加哥），由于气候变化，增加了热岛强度。

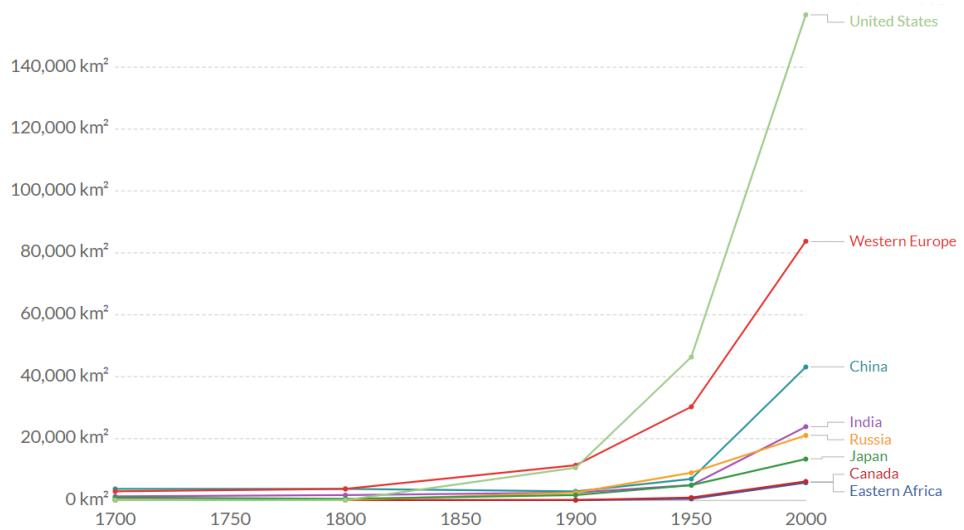


图 4-27 世界上各地区的城市化总面积

从上面的图可以看出，从 1700-2000 年以来，世界上各地区的城市化总面积在不断增加。城市化带来的各种问题也在不断引起气候的变化，例如汽车尾气排放量增加、光污染等因素均会导致全球变暖。

## 4.4 全球气候变化多变量线性回归模型

### 4.4.1 模型的构建思路

根据上述分析及相关研究，多种因素影响着全球气候变化。为了能够定量的分析各种因

素对于全球气候变化的影响程度，并能够根据这些因素的情况对宏观上的全球气候进行预测，本文尝试构建全球气候变化的多变量回归模型。

为构建多变量回归模型，需要搜集并处理各因素的数据集，通过 SPSS 软件和 Python 软件对数据进行清洗、标准化及缺失值替换等相应处理，使得数据具有可用性。对于处理后的数据，可对全球温度及各因素进行关联性分析，得到各因素之间的关联性及其与全球温度之间的关联性，从而可以寻找和求证影响全球气候变化的因素。

为了能够对未来 25 年的天气进行预测，将各因素的历史数据与全球温度之间建立多变量回归模型，构建的模型将尝试应用于未来全球气候变化的预测。

#### 4.4.2 数据集的搜集及变量含义设定

为了分析各因素与全球气候变化的影响关系，需搜集相关数据集。由于不同要素的资料统计部门不同，下表 4-1 给出本章所采用的数据集来源及变量的含义。

表 4-1 不同变量的数据集来源及含义

变量	数据集来源	含义
L_O_Temp_Ano_Nosmooth	NASA数据集	温度的距平
L_O_Temp_Ano_Lowess	NASA数据集	温度的距平（平滑处理后的值）
L_Mean_Temp	Berkeley Earth	全球陆地平均温度
L_Mean_Temp_Uncer	Berkeley Earth	全球陆地平均温度的不确定性
L_O_Mean_Temp	Berkeley Earth	全球陆地和海洋表面平均温度
L_O_Mean_Temp_Uncer	Berkeley Earth	全球陆地和海洋表面平均温度的不确定性
G_Temp_F_mean	Berkeley Earth	全球平均温度（华氏度）
G_Temp_C_mean	Berkeley Earth	全球平均温度（摄氏度）
Solar_Flux_mean	Ourworldindata网站 <sup>1</sup>	太阳辐射通量
Urbanization	Ourworldindata网站 <sup>1</sup>	城市化程度
Emiss_of_Ozone	Ourworldindata网站 <sup>1</sup>	臭氧消耗物质的排放量
Ozone_hole	Ourworldindata网站 <sup>1</sup>	臭氧洞区
PM2.5	Ourworldindata网站 <sup>1</sup>	空气颗粒物
SO2	Ourworldindata网站 <sup>1</sup>	二氧化硫
Land_Built_up	Ourworldindata网站 <sup>1</sup>	土地利用指标：建成区面积
Land_Graze	Ourworldindata网站 <sup>1</sup>	土地利用指标：放牧区面积
Land_Crop	Ourworldindata网站 <sup>1</sup>	土地利用指标：耕地区面积
annual_CO2	Ourworldindata网站 <sup>1</sup>	每年二氧化碳排放量
cumu_CO2	Ourworldindata网站 <sup>1</sup>	累计二氧化碳排放量
CFCsTG_OBS	Ourworldindata网站 <sup>1</sup>	氟氯碳化合物（数据来源于OBS）
CH4_OBS	Ourworldindata网站 <sup>1</sup>	甲烷（数据来源于OBS）
CO2_OBS	Ourworldindata网站 <sup>1</sup>	二氧化碳（数据来源于OBS）
N2O_OBS	Ourworldindata网站 <sup>1</sup>	一氧化氮（数据来源于OBS）
Aero_SA	学位论文[16]	平流层气溶胶
Aero_BC	学位论文[16]	化石燃料黑碳气溶胶
Aero_RA	学位论文[16]	对流层气溶胶反射作用

Aero_AIE	学位论文[16]	气溶胶间接作用
SST_high_lat	NOAA数据集	海洋表面温度（高纬度）
SST_mid_lat	NOAA数据集	海洋表面温度（中纬度）
SST_low_lat	NOAA数据集	海洋表面温度（低纬度）

注：1. 网址：<https://ourworldindata.org/#entries>

#### 4.4.3 数据缺失值处理及标准化

##### (1) 数据缺失值的处理

由于数据记录部门及周期不同，不同数据集的数据量不尽相同，为了使各变量的数据集具有可用性，对数据集中的缺失值做相应处理。本文对这些数据集中的缺失值处理方法主要包括以下两种：

a) 均值插补：采用序列的均值来替换缺失值。

b) 基于序列趋势的插值：对于序列中间的缺失值，如一段历史数据中某几个年份缺乏记录，首先分析该序列的趋势，用最佳的回归模型进行拟合或者采用时间序列模型进行拟合，从而计算出相应的估计值，替换缺失值。对于序列末尾的缺失值，通过对序列趋势及特性的分析，采用回归模型或时间序列模型来拟合预测，得到对应的预测值以替换缺失值。对于序列首段的缺失值，可将序列先倒序后再采用前述进行逆时间预测，得到对应的预测值从而替换缺失值。

##### (2) 各变量的数据标准化

由于各变量的单位和数量级并不相同，为了便于构建多变量回归模型，可将各变量的数据进行标准化。本文采用 z-score 标准化对数据进行处理。标准分数（standard score）也叫 z 分数（z-score），是将变量值与平均数的差再除以标准差的过程。可采用下列公式进行表示：

$$z = (x - \mu) / \sigma$$

其中  $x$  为变量的某一具体指， $\mu$  为数据序列的平均数， $\sigma$  为数据序列的标准差。

标准化得到的 Z 值的量代表着变量值和数据序列的平均值之间的距离，是以标准差为单位计算。在变量的原始值低于平均值时 Z 值为负数，反之则为正数。

本章将变量标准化后在变量名前加 Z 进行表示，从而与变量的原始值区分开来，其含义为该变量数据的标准化分数。如 Solar\_Flux\_mean 表示太阳辐射通量平均值变量，那么 ZSolar\_Flux\_mean 则表示对该变量数据的标准化结果。其它变量名的含义类似，不再赘述。

##### (3) 标准化后的数据集分析

可采用 Python 软件对各变量的数据集特征进行分析，从而了解各变量的数据特征。



图 4-28 各变量数据集的特征统计分析

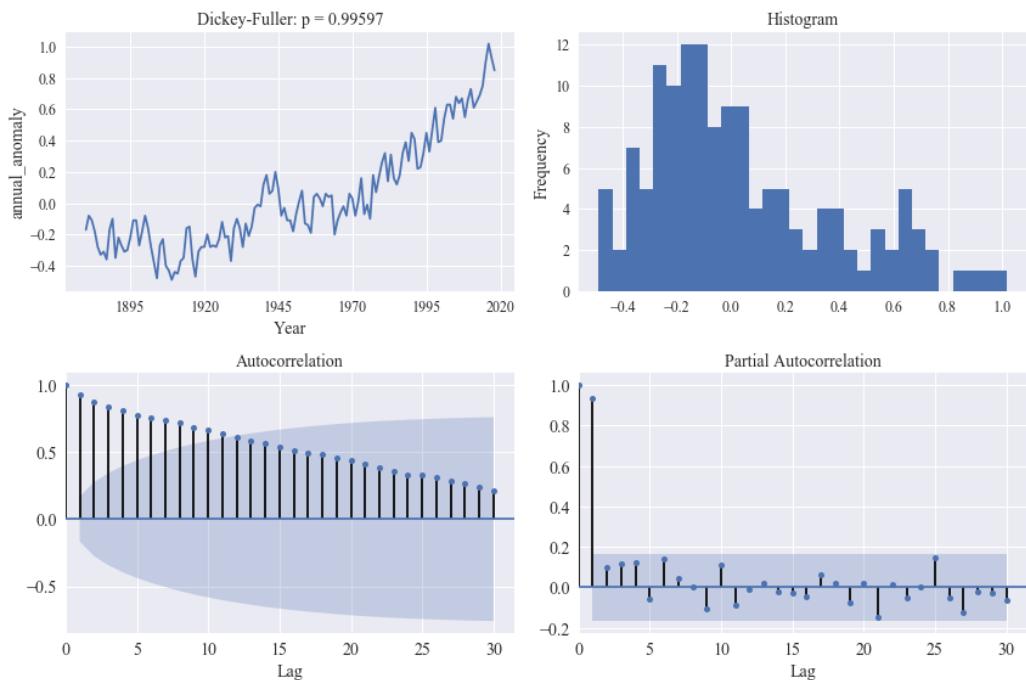


图 4-29 全球温度距平的逐年数据的可视化分析

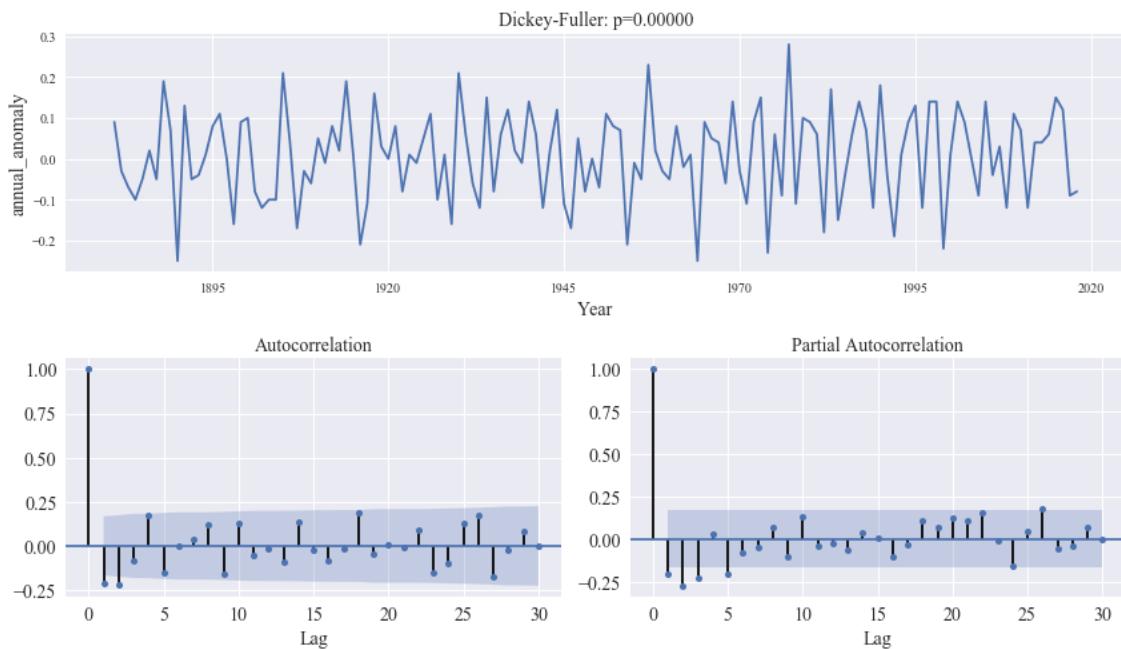


图 4-30 全球温度距平的逐年数据一次差分可视化分析

#### 4.4.4 因子的关联性分析

为分析各变量之间的关联性及其对全球温度的影响，可采用 SPSS 和 Python 软件对各因子进行关联性分析，分析后可以得到相应地结果，如表 4-2 所示。

表 4-2 各因子间的 Pearson 相关性系数

相关性	L_O_Temp_Ano_Nos_mooth	G_Temp_C_mean	Solar_Flux_C_mean	Urbanization	Emiss_of_Ozone	Ozone_hole	PM2.5	SO2	Land_Built_up	Land_Graze	Land_Crop	annual_CO2	cumu_CO2	CFCs_TG_OBS	CH4_OBS	CO2_OBS	N2O_OBS	Aero_SA	Aero_BC	Aero_RA	Aero_AIE	SST_high_lat	SST_mid_lat	SST_low_lat			
L_O_Temp_Ano_Nos_mooth	1	.999**	-0.141	.920**	-.345*	.509**	.850**	.672**	.927**	.790**	.805**	.920**	.939**	.856**	.873**	.895**	.899**	0.019	.844**	-.849*	-.828*	.925**	.925**	.941**			
G_Temp_C_mean	.999**	1	-0.107	.934**	-.339*	.515**	.837**	0.083	.943**	.767**	.820**	.921**	.952**	.900**	.859**	.917**	.915**	-0.025	.856**	-.838*	-.826*	.912**	.942**	.933**			
Solar_Flux_mean	-0.141	-0.107	1	-0.177	.292*	-.413*	*	-0.191	0.007	-0.176	-0.112	-0.137	-0.198	-0.196	0.008	-0.013	0.003	0.003	0.113	-0.033	0.03	0.04	-0.189	-0.116	-0.13		
Urbanization	.920**	.934**	-0.177	1	-.382*	*	.667**	.802**	.838**	.975**	.950**	.967**	.967**	.956**	.874**	.984**	.949**	.967**	-.260*	.962**	-.978*	-.994*	.898**	.945**	.841**		
Emiss_of_Ozone	-.345*	-.339*	.292*	-.382*	*	1	-.557*	-.831*	.879**	-.387*	*	-0.092	-0.102	-.304*	-.388*	0.149	0.259	0.01	0.027	0.004	.426**	-.532*	-.502*	-0.258	-0.25	-0.203	
Ozone_hole	.509**	.515**	-.413*	*	.667**	-.557*	*	1	0.027	-.726*	*	.663**	.902**	.718**	.542**	.615**	.903**	.911**	.787**	.821**	-0.024	.950**	-.952*	-.962*	.368*	.504**	.371*
PM2.5	.850**	.837**	-0.191	.802**	-.831*	*	0.027	1	-.892*	*	.815**	-0.3	.897**	.771**	.822**	.789**	.827**	.969**	.954**	0.481	.	-.868*	-.863*	*	.772**	.773**	.825**
SO2	.672**	0.083	0.007	.838**	.879**	-.726*	-.892*	*	1	.856**	.951**	.946**	.856**	.789**	.719**	.933**	.861**	.874**	-0.086	.910**	-.941*	-.953*	*	.651**	.553**	.546**	
Land_Built_up	.927**	.943**	-0.176	.975**	-.387*	*	.663**	.815**	.856**	1	.901**	.907**	.992**	.989**	.916**	.992**	.996**	.997**	-0.059	.996**	-.994*	-.984*	*	.879**	.767**	.748**	
Land_Graze	.790**	.767**	-0.112	.950**	-0.092	.902**	-0.3	.951**	.901**	1	.997**	.868**	.833**	.725**	.948**	.908**	.915**	0.014	.925**	-.942*	-.976*	*	.702**	.601**	.607**		

Land_Crop	.805**	.820**	-0.137	.967**	-0.102	.718**	.897**	.946**	.907**	.997**	1	.873**	.838**	.717**	.942**	.907**	.911**	0.015	.932**	-.947*	-.981*	.704**	.586**	.602**	
annual_CO2	.920**	.921**	-0.198	.967**	-.304*	.542**	.771**	.856**	.992**	.868**	.873**	1	.991**	.938**	.987**	.981**	.986**	-0.105	.983**	-.990*	-.965*	.896**	.804**	.772**	
cumu_CO2	.939**	.952**	-0.196	.956**	-.388*	*.615**	.822**	.789**	.989**	.833**	.838**	.991**	1	.964**	.975**	.990**	.994**	-0.091	.986**	-.982*	-.954*	.912**	.820**	.788**	
CFCsTG_OBS	.856**	.900**	0.008	.874**	0.149	.903**	.789**	.719**	.916**	.725**	.717**	.938**	.964**	1	.902**	.934**	.936**	-0.154	.924**	-.908*	-.847*	.831**	.740**	.695**	
CH4_OBS	.873**	.859**	-0.013	.984**	0.259	.911**	.827**	.933**	.992**	.948**	.942**	.987**	.975**	.902**	1	.983**	.989**	-0.076	.993**	-.999*	-.989*	.811**	.677**	.659**	
CO2_OBS	.895**	.917**	0.003	.949**	0.01	.787**	.969**	.861**	.996**	.908**	.907**	.981**	.990**	.934**	.983**	1	.998**	-0.063	.991**	-.984*	-.964*	.819**	.663**	.659**	
N2O_OBS	.899**	.915**	0.003	.967**	0.027	.821**	.954**	.874**	.997**	.915**	.911**	.986**	.994**	.936**	.989**	.998**	1	-0.071	.992**	-.989*	-.971*	.829**	.695**	.676**	
Aero_SA	0.019	-0.025	0.113	-.260*	*	0.004	-0.024	0.481	-0.086	-0.059	0.014	0.015	-0.105	-0.091	-0.154	-0.076	-0.063	-0.071	1	-0.069	0.081	0.022	-0.067	-0.052	-0.067
Aero_BC	.844**	.856**	-0.033	.962**	.426**	.950**	.*.	.910**	.996**	.925**	.932**	.983**	.986**	.924**	.993**	.991**	.992**	-0.069	1	-.995*	-.983*	.849**	.782**	.689**	
Aero_RA	-.849*	-.838*	0.03	-.978*	-.532*	-.952*	-.868*	-.941*	-.994*	-.942*	-.947*	-.990*	-.982*	-.908*	-.999*	-.984*	-.989*	0.081	-.995*	1	.989**	-.861*	-.805*	-.700*	
Aero_AIE	-.828*	-.826*	0.04	-.994*	-.502*	-.962*	-.863*	-.953*	-.984*	-.976*	-.981*	-.965*	-.954*	-.847*	-.989*	-.964*	-.971*	0.022	-.983*	.*.	.989**	1	-.840*	-.770*	-.672*
SST_high_lat	.925**	.912**	-0.189	.898**	-0.258	.368*	.772**	.651**	.879**	.702**	.704**	.896**	.912**	.831**	.811**	.819**	.829**	-0.067	.849**	-.861*	-.840*	1	.831**	.797**	
SST_mid_lat	.925**	.942**	-0.116	.945**	-0.25	.504**	.773**	.553**	.767**	.601**	.586**	.804**	.820**	.740**	.677**	.663**	.695**	-0.052	.782**	-.805*	-.770*	.831**	1	.807**	
SST_low_lat	.941**	.933**	-0.13	.841**	-0.203	.371*	.825**	.546**	.748**	.607**	.602**	.772**	.788**	.695**	.659**	.659**	.676**	-0.067	.689**	-.700*	-.672*	.797**	.807**	1	

\*\*. 在 .01 水平（双侧）上显著相关。

\*. 在 0.05 水平（双侧）上显著相关。

c. 因为至少有一个变量为常量，所以无法进行计算。

表 4-3 Kendall 检验得出的各因子间的相关性系数

		L_O_Temp_Index_Nosmooth	Solar_Flux_mean	Urbanization	Emissons_of_Ozone	Ozone_hole	PM2.5	SO2	Land_Built_up	Land_Graze	Land_Crop	annual_CO2	CFCs_STG_OB_S	cumulative_CO2	CH4_OB_S	CO2_OB_S	N2O_OB_S	Aero_S_A	Aero_B_C	Aero_R_A	Global_L_O_Mean_Temp	
Kendall 的 tau_b	L_O_Temp_Index_Nosmooth	1.00 0	-0.0 70	.762 **	-.25 8**	.301 **	.740 **	.510 **	.697 **	.663 **	.692 **	.691 **	.663 **	.701 **	.647 **	.650 **	.647 **	0.01 5	.625 **	-.62 4**	.894 **	
	Solar_Flux_mean	-0.0 70	1.00 0	-0.1 01	0.17 0	-0.1 71	-0.1 03	0.05 0	-0.1 01	-0.0 84	-0.0 87	-0.1 06	-0.0 05	-0.1 16	-0.0 03	-0.0 05	-0.0 05	.212 *30	-0.0 7	0.03 6	-0.0 03	
	Urbanization	.762 **	-0.1 01	1.00 0	-.21 5*	.464 **	.907 **	.727 **	1.00 0**	.940 **	.978 **	.962 **	.994 **	1.00 0**	.999 **	.988 **	1.00 0**	-.15 1*	.995 **	-1.0 00**	-1.0 00**	.747 **
	Emissions_of_Ozone	-.25 8**	0.17 0	-.21 5*	1.00 0	-.29 8*	-.89 1**	.706 **	-.21 5*	0.00 4	-0.1 13	-.20 9*	0.12 1	-.21 5*	0.12 8	0.12 1	0.12 1	0.01 0	.464 **	-.38 4**	-.37 6**	-.27 9**
	Ozone_hole	.301 **	-0.1 71	.464 **	-.29 8*	1.00 0	-0.0 65	-.58 1**	.464 **	.606 **	.401 **	.376 **	.686 **	.416 **	.688 **	.686 **	.686 **	0.12 4	.790 **	-.81 7**	-.80 0**	.384 **
	PM2.5	.740 **	-0.1 03	.907 **	-.89 1**	-0.0 65	1.00 0	-.85 7**	.907 **	-.2 40	.673 **	.913 **	.750 **	.907 **	.706 **	.750 **	.750 **	.624 **	-.49 6*	-0.4 55	.634 **	
	SO2	.510	0.05	.727	.706	-.58	-.85	1.00	.871	.884	.882	.856	.874	.871	.902	.896	.901	-0.0	.902	-.89	-.89	.561

		**	0	**	**	1**	7**	0	**	**	**	**	**	**	**	**	26	**	1**	0**	**
Land_Built_up	.697 **	-0.1 01	1.00 0**	-.21 5*	.464 **	.907 **	.871 0	1.00 **	.971 0	.989 **	.981 **	.975 **	1.00 0**	1.00 0**	.995 **	1.00 0**	0.00 1	.996 **	-1.0 00**	-1.0 00**	.682 **
Land_Graze	.663 **	-0.0 84	.940 **	0.00 4	.606 **	-0.2 40	.884 **	.971 0	1.00 **	.964 **	.952 **	.972 **	.971 **	.997 **	.991 **	.996 **	0.00 1	.996 **	-1.0 00**	-1.0 00**	.667 **
Land_Crop	.692 **	-0.0 87	.978 **	-0.1 13	.401 **	.673 **	.882 **	.989 **	.964 **	1.00 0	.972 **	.965 **	.989 **	.990 **	.985 **	.990 **	-0.0 10	.996 **	-.98 5**	-.98 4**	.679 **
annual_CO2	.691 **	-0.1 06	.962 **	-.20 9*	.376 **	.913 **	.856 **	.981 **	.952 **	.972 **	1.00 0	.956 **	.981 **	.978 **	.976 **	.978 **	0.00 4	.961 **	-.96 4**	-.96 4**	.676 **
CFCsTG_OBS	.663 **	-0.0 05	.994 **	0.12 1	.686 **	.750 **	.874 **	.975 **	.972 **	.965 **	.956 **	1.00 0	.975 **	.975 **	.970 **	.976 **	0.00 0	.982 **	-.98 6**	-.98 6**	.666 **
cumulative_CO2	.701 **	-0.1 16	1.00 0**	-.21 5*	.416 **	.907 **	.871 **	1.00 0**	.971 **	.989 **	.981 **	.975 **	1.00 0	1.00 0**	.995 **	1.00 0**	0.00 1	.996 **	-1.0 00**	-1.0 00**	.682 **
CH4_OBS	.647 **	-0.0 03	.999 **	0.12 8	.688 **	.706 **	.902 **	1.00 0**	.997 **	.990 **	.978 **	.975 **	1.00 0**	1.00 0	.994 **	.999 **	0.00 1	.996 **	-1.0 00**	-1.0 00**	.645 **
CO2_OBS	.650 **	-0.0 05	.988 **	0.12 1	.686 **	.750 **	.896 **	.995 **	.991 **	.985 **	.976 **	.970 **	.995 **	.994 **	1.00 0	.994 **	-0.0 03	.987 **	-.99 1**	-.99 1**	.647 **
N2O_OBS	.647 **	-0.0 05	1.00 0**	0.12 1	.686 **	.750 **	.901 **	1.00 0**	.996 **	.990 **	.978 **	.976 **	1.00 0**	.999 **	.994 **	1.00 0	0.00 1	.996 **	-.99 9**	-1.0 00**	.646 **
Aero_SA	0.01 5	.212 *	-.15 1*	0.01 0	0.12 4	.624 **	-0.0 26	0.00 1	0.00 1	-0.0 10	0.00 4	0.00 0	0.00 1	0.00 1	-0.0 03	0.00 1	1.00 0	-0.0 04	-0.0 01	-0.0 01	0.05 9
Aero_BC	.625 **	-0.0 30	.995 **	.464 **	.790 **	.	.902 **	.996 **	.996 **	.996 **	.961 **	.982 **	.996 **	.996 **	.987 **	.996 **	-0.0 04	1.00 0	-.99 6**	-.99 6**	.681 **
Aero_RA	-.62 4**	0.03 7	-1.0 00**	-.38 4**	-.81 7**	-.49 6*	-.89 1**	-1.0 00**	-1.0 00**	-.98 5**	-.96 4**	-.98 6**	-.1.0 00**	-.1.0 00**	-.99 1**	-.99 9**	-0.0 01	-.99 6**	1.00 0	1.00 0**	-.68 1**
Aero_AIE	-.62 4**	0.03 6	-1.0 00**	-.37 6**	-.80 0**	-.0.4 55	-.89 0**	-1.0 00**	-1.0 00**	-.98 4**	-.96 4**	-.98 6**	-.1.0 00**	-.1.0 00**	-.99 1**	-.99 00**	-0.0 01	-.99 6**	1.00 0	1.00 0**	-.68 0**

	Global_L_O_Mean_Temp	.894 ** 03	-0.0 ** 9**	.747 ** 9**	-.27 **	.384 **	.634 **	.561 **	.682 **	.667 **	.679 **	.676 **	.666 **	.682 **	.645 **	.647 **	.646 **	0.05 9	.681 **	-.68 1**	-.68 0**	1.00 0
--	----------------------	------------------	-------------------	-------------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	-----------	------------	-------------	-------------	-----------

\*\*. 在 .01 水平(双侧)上显著相关。

\*. 在 0.05 水平(双侧)上显著相关。

c. 因为至少有一个变量为常量，所以无法进行计算。

表 4-4 Spearman 检验得出的各因子间的相关性系数

	L_O_Temp_Index_Nosmooth	Solar_Flux_mean	Urbanization	Emissions_of_Ozone	PM2.5	SO2	Land_Built_up	Lan_d_Graze	Lan_d_Crop	annual_CO2	CFCs_TG_OB	cumulative_CO2	CH4_OB_S	CO2_OB_S	N2O_A	Aero_S	Aero_B	Aero_R	Aero_A	Glo bal_L_O_Mean_Temp
L_O_Temp_Index_Nosmooth	1.00 0	-0.1 05	.914 ** 9**	-.37 **	.446 **	.896 **	.749 **	.870 **	.862 **	.869 **	.867 **	.840 **	.873 **	.834 **	.835 **	.833 **	0.02 4	.817 ** 7**	-.81 7**	.979 **
Solar_Flux_mean	-0.1 05	1.00 0	-0.1 42	0.25 8	-0.2 93	-0.1 59	0.06 1	-0.1 42	-0.1 11	-0.1 24	-0.1 51	0.00 7	-0.1 63	0.00 8	0.00 7	0.00 7	.324 * 31	-0.0 2	0.04 1	-0.0 02
Urbanization	.914 ** 42	-0.1 0	1.00 6*	-.34 **	.618 **	.978 **	.866 **	1.00 0**	.986 **	.998 **	.995 **	1.00 0**	1.00 0**	1.00 0**	.998 **	1.00 0**	-.25 6**	.999 ** 00**	-1.0 00**	.904 **
Emissions_of_Ozone	-.37 9** 8	0.25 6*	-.34 0	1.00 9**	-.51 2**	-.97 **	.876 6*	-.34 97	-0.1 6*	-.28 2*	-.34 8	0.08 6*	-.34 9	0.08 8	0.08 8	0.08 7	.477 ** 1**	-.41 0**	-.41 9**	-.41 9**
Ozone_hole	.446 ** 93	-0.2 **	.618 9**	-.51 0	1.00 04	-.01 6**	-.74 **	.618 **	.753 **	.550 **	.544 **	.832 **	.561 **	.832 **	.832 **	.832 3	0.16 ** 1**	.905 ** 1**	-.94 8**	.556 **

PM2.5	.896 ** 59	-0.1 ** 2**	.978 ** 2**	-.97 04	-0.1 0	1.00 -.95	-.95 5**	.978 ** 84	-0.1 ** 84	.859 ** 84	.980 ** 84	.897 ** 84	.978 ** 84	.887 ** 84	.897 ** 84	.897 ** 84	.815 ** 84	.	-.69 4*	-.68 2*	.812 **
SO2	.749 ** 1	0.06 **	.866 **	.876 **	-.74 6**	-.95 5**	1.00 0	.956 **	.958 **	.958 **	.954 **	.966 **	.956 **	.970 **	.969 **	.970 **	-0.0 72	.965 **	-.96 3**	-.96 3**	.788 **
Land_Buil t_up	.870 ** 42	-0.1 0**	1.00 6*	-.34 **	.618 **	.978 **	.956 **	1.00 0	.995 **	.999 **	.998 **	.997 **	1.00 0**	1.00 0**	1.00 0**	1.00 0**	-0.0 26	1.00 0**	-1.0 00**	-1.0 00**	.862 **
Land_Gra ze	.862 ** 11	-0.1 **	.986 97	-0.1 **	.753 **	-.01 84	.958 **	.995 **	1.00 0	.994 **	.993 **	.997 **	.995 **	1.00 0**	.999 **	1.00 0**	-0.0 26	1.00 0**	-1.0 00**	-1.0 00**	.859 **
Land_Cro p	.869 ** 24	-0.1 **	.998 6*	-.28 **	.550 **	.859 **	.958 **	.999 **	.994 **	1.00 0	.998 **	.996 **	.999 **	.999 **	.999 **	.999 **	-0.0 42	1.00 0**	-.99 8**	-.99 8**	.861 **
annual_C O2	.867 ** 51	-0.1 **	.995 2*	-.34 **	.544 **	.980 **	.954 **	.998 **	.993 **	.998 **	1.00 0	.995 **	.998 **	.998 **	.998 **	.998 **	-0.0 21	.995 **	-.99 5**	-.99 5**	.857 **
CFCsTG_ OBS	.840 ** 7	0.00 0**	1.00 8	0.08 **	.832 **	.897 **	.966 **	.997 **	.997 **	.996 **	.995 **	1.00 0	.997 **	.997 **	.996 **	.997 **	-0.0 18	.998 **	-.99 9**	-.99 9**	.841 **
cumulativ e_CO2	.873 ** 63	-0.1 0**	1.00 6*	-.34 **	.561 **	.978 **	.956 **	1.00 0**	.995 **	.999 **	.998 **	.997 **	1.00 0	1.00 0**	1.00 0**	1.00 0**	-0.0 26	1.00 0**	-1.0 00**	-1.0 00**	.862 **
CH4_OBS	.834 ** 8	0.00 0**	1.00 9	0.08 **	.832 **	.887 **	.970 **	1.00 0**	1.00 0**	.999 **	.998 **	.997 **	1.00 0	1.00 0**	1.00 0**	1.00 0**	-0.0 26	1.00 0**	-1.0 00**	-1.0 00**	.834 **
CO2_OBS	.835 ** 7	0.00 **	.998 8	0.08 **	.832 **	.897 **	.969 **	1.00 0**	.999 **	.999 **	.998 **	.996 **	1.00 0**	1.00 0**	1.00 0**	1.00 0**	-0.0 29	.999 **	-.99 9**	-.99 9**	.834 **
N2O_OBS	.833 ** 7	0.00 0**	1.00 8	0.08 **	.832 **	.897 **	.970 **	1.00 0**	1.00 0**	.999 **	.998 **	.997 **	1.00 0**	1.00 0**	1.00 0**	1.00 0	-0.0 26	1.00 0**	-1.0 00**	-1.0 00**	.834 **
Aero_SA	0.02 4	.324 *	-.25 6**	0.02 7	0.16 3	.815 **	-0.0 72	-0.0 26	-0.0 26	-0.0 42	-0.0 21	-0.0 18	-0.0 26	-0.0 26	-0.0 29	-0.0 26	1.00 0	-0.0 34	0.02 6	0.02 6	0.08 4
Aero_BC	.817 ** 31	-0.0 **	.999 **	.477 **	.905 **	.	.965 **	1.00 0**	1.00 0**	1.00 0**	.995 **	.998 **	1.00 0**	1.00 0**	.999 **	1.00 0**	-0.0 34	1.00 0	-1.0 00**	-1.0 00**	.872 **
Aero_RA	-.81 7**	0.04 2	-1.0 00**	-.41 1**	-.94 1**	-.69 4*	-.96 3**	-1.0 00**	-1.0 00**	-.99 8**	-.99 5**	-.99 9**	-1.0 00**	-1.0 00**	-.99 9**	-1.0 00**	0.02 6	-1.0 00**	1.00 0	1.00 0	-.87 2**

Aero_AIE	-.81 7**	0.04 1	-1.0 00**	-.41 0**	-.93 8**	-.68 2*	-.96 3**	-1.0 00**	-1.0 00**	-.99 8**	-.99 5**	-.99 9**	-1.0 00**	-1.0 00**	-.99 9**	-1.0 00**	0.02 6	-1.0 00**	1.00 0**	1.00 0	-.87 2**
Global_L_O_Mean_Temp	.979 **	-0.0 02	.904 **	-.41 9**	.556 **	.812 **	.788 **	.862 **	.859 **	.861 **	.857 **	.841 **	.862 **	.834 **	.834 **	.08 4	.872 **	-.87 2**	-.87 2**	1.00 0	

\*\*. 在 .01 水平(双侧)上显著相关。

\*. 在 0.05 水平(双侧)上显著相关。

c. 因为至少有一个变量为常量，所以无法进行计算。

采用 Python 软件对各因子间的关联性程度做进一步的可视化处理，将相关性系数绘制成立热力图来展现，如下图 4-31 所示。

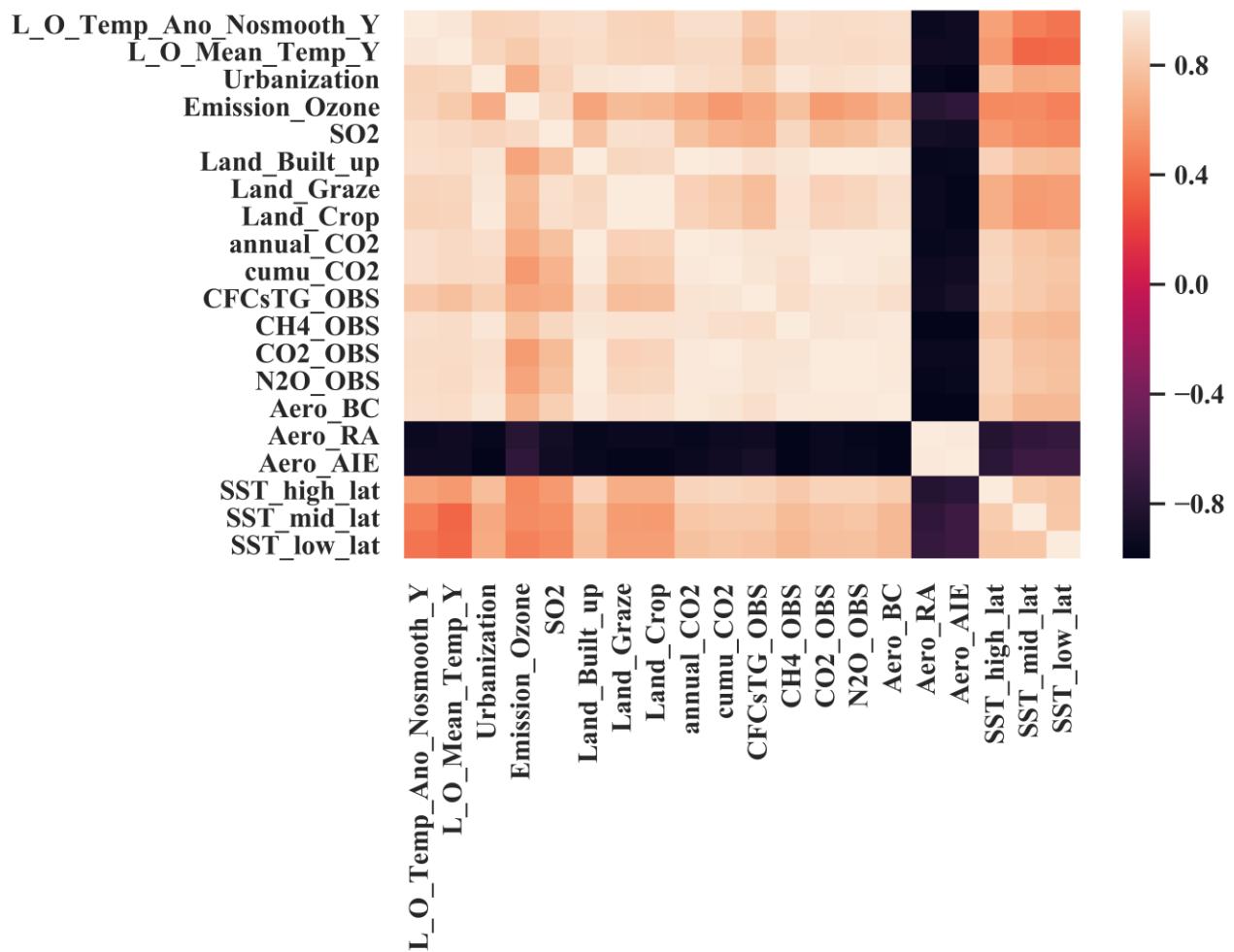


图 4-31 相关性系数的热力图

#### 4.4.5 模型的构建与求解

运用 SPSS 软件构建得到多变量线性回归模型，模型的系数如下表 4-5 所示。

表 4-5 多变量回归模型的系数

因变量	参数	B	标准误差	t	Sig.	95% 置信区间		观测到的 幂 <sup>a</sup>
						下限	上限	
L_O_Temp_Ano_Nosmo oth_A	截距	.683	.345	1.978	.050	-.001	1.367	.501
	ZSolar_Flux_mean	-.002	.009	-.269	.788	-.020	.015	.058
	ZUrbanization	1.390	.771	1.804	.074	-.136	2.917	.432
	ZEmiss_of_Ozone	-.178	.145	-1.225	.223	-.465	.109	.229
	ZSO2	-.539	.328	-1.641	.103	-1.189	.111	.370
	ZLand_Built_up	3.815	1.203	3.171	.002	1.433	6.196	.882

	ZLand_Graze	-.063	1.100	-.057	.954	-2.242	2.115	.050
	ZLand_Crop	-2.092	1.603	-1.305	.195	-5.266	1.083	.253
	Zannual_CO2	.116	.417	.278	.781	-.709	.941	.059
	Zcumu_CO2	.326	2.729	.119	.905	-5.078	5.730	.052
	ZCFCsTG_OBS	-1.238	.716	-1.730	.086	-2.655	.179	.404
	ZCH4_OBS	-2.550	1.379	-1.848	.067	-5.281	.182	.450
	ZCO2_OBS	-.039	.840	-.046	.963	-1.701	1.624	.050
	ZN2O_OBS	-.326	1.400	-.233	.816	-3.098	2.446	.056
	ZAero_SA	.022	.011	1.999	.048	.000	.044	.509
	ZAero_BC	-1.660	.584	-2.845	.005	-2.815	-.505	.806
	ZAero_RA	-4.675	1.683	-2.777	.006	-8.008	-1.342	.787
	ZAero_AIE	.502	.510	.983	.328	-.509	1.512	.164
	ZSST_high_lat	-.002	.029	-.076	.939	-.059	.055	.051
	ZSST_mid_lat	.051	.028	1.818	.072	-.005	.107	.438
	ZSST_low_lat	-.006	.017	-.343	.732	-.040	.028	.063
L_O_Mean_Temp_A	截距	15.852	.363	43.636	.000	15.133	16.571	1.000
	ZSolar_Flux_mean	-.004	.009	-.449	.654	-.023	.014	.073
	ZUrbanization	.800	.811	.987	.326	-.805	2.404	.165
	ZEmiss_of_Ozone	-.157	.152	-1.031	.305	-.459	.145	.176
	ZSO2	-.447	.345	-1.296	.198	-1.131	.236	.251
	ZLand_Built_up	2.477	1.265	1.958	.052	-.027	4.981	.493
	ZLand_Graze	-.820	1.157	-.709	.480	-3.111	1.470	.108
	ZLand_Crop	-.596	1.686	-.353	.725	-3.933	2.742	.064
	Zannual_CO2	.107	.438	.243	.808	-.761	.975	.057
	Zcumu_CO2	1.307	2.870	.456	.650	-4.375	6.990	.074
	ZCFCsTG_OBS	-.848	.753	-1.127	.262	-2.338	.642	.201
	ZCH4_OBS	-2.299	1.450	-1.585	.116	-5.171	.573	.349
	ZCO2_OBS	-.121	.883	-.137	.892	-1.869	1.628	.052
	ZN2O_OBS	-.810	1.472	-.550	.583	-3.724	2.105	.085
	ZAero_SA	.018	.012	1.527	.129	-.005	.041	.328
	ZAero_BC	-1.311	.614	-2.137	.035	-2.526	-.097	.564
	ZAero_RA	-3.227	1.770	-1.823	.071	-6.732	.278	.440
	ZAero_AIE	-.520	.537	-.970	.334	-1.583	.542	.161
	ZSST_high_lat	.026	.030	.853	.395	-.034	.085	.135
	ZSST_mid_lat	.049	.030	1.660	.100	-.009	.108	.377
	ZSST_low_lat	.004	.018	.237	.813	-.032	.040	.056

基于系数表，可将全球气候变化变量线性回归模型表述如下：

$$\begin{aligned}
 L_O_Temp\_Ano\_Nosmooth\_A = & 0.683 - 0.002 * ZSolar\_Flux\_mean + 1.390 * ZUrbanization \\
 & - 0.178 * ZEmmiss\_of\_Ozone - 0.0539 * ZSO2 + 3.815 * ZLand\_Built\_up - 0.063 * \\
 & ZLand\_Graze - 2.092 * ZLand\_Crop + 0.116 * Zannual\_CO2 + 0.326 * Zcumu\_CO2 \\
 & - 1.238 * ZCFCsTG\_OBS - 2.550 * ZCH4\_OBS - 0.039 * ZCO2\_OBS - 0.326 * ZN2O\_OBS \\
 & + 0.022 * ZAero\_SA - 1.660 * ZAero\_BC - 4.675 * ZAero\_RA + 0.502 * ZAero\_AIE - 0.002 * \\
 & ZSST\_high\_lat + 0.051 * ZSST\_mid\_lat - 0.006 * ZSST\_low\_lat
 \end{aligned}$$

$$\begin{aligned}
 L_O_Mean\_Temp\_A = & 15.852 - 0.004 * ZSolar\_Flux\_mean + 0.800 * ZUrbanization - 0.157 \\
 & * ZEmmiss\_of\_Ozone - 0.447 * ZSO2 + 2.477 * ZLand\_Built\_up - 0.820 * ZLand\_Graze \\
 & - 0.596 * ZLand\_Crop + 0.107 * Zannual\_CO2 + 1.307 * Zcumu\_CO2 - 0.848 * \\
 & ZCFCsTG\_OBS - 2.299 * ZCH4\_OBS - 1.210 * ZCO2\_OBS - 0.810 * ZN2O\_OBS + 0.018 \\
 & * ZAero\_SA - 1.311 * ZAero\_BC - 3.227 * ZAero\_RA - 0.520 * ZAero\_AIE + 0.026 * \\
 & ZSST\_high\_lat + 0.049 * ZSST\_mid\_lat + 0.004 * ZSST\_low\_lat
 \end{aligned}$$

表 4-6 未经标准化处理得出的系数估计

因变量	参数	B	标准误差	t	Sig.	95% 置信区间	
						下限	上限
L_O_Temp_Anosmooth_A	截距	12.345	27.102	.455	.650	-41.315	66.005
	Solar_Flux_mean	-4.844E-006	1.801E-005	-.269	.788	-4.050E-005	3.082E-005
	Urbanization	.095	.053	1.804	.074	-.009	.199
	Emiss_of_Ozone	-4.475E-007	3.653E-007	-1.225	.223	-1.171E-006	2.758E-007
	SO2	-.011	.007	-1.641	.103	-.024	.002
	Land_Built_up	2.400E-007	7.569E-008	3.171	.002	9.019E-008	3.899E-007
	Land_Graze	-8.052E-011	1.404E-009	-.057	.954	-2.860E-009	2.699E-009
	Land_Crop	-6.245E-009	4.786E-009	-1.305	.195	-1.572E-008	3.232E-009
	annual_CO2	1.078E-011	3.873E-011	.278	.781	-6.590E-011	8.745E-011
	cumu_CO2	7.426E-013	6.222E-012	.119	.905	-1.158E-011	1.306E-011
	CFCsTG_OBS	-2.701	1.561	-1.730	.086	-5.793	.390
	CH4_OBS	-7.353	3.978	-1.848	.067	-15.229	.524
	CO2_OBS	-.001	.026	-.046	.963	-.053	.050
	N2O_OBS	-21.450	92.178	-.233	.816	-203.955	161.055
	Aero_SA	.048	.024	1.999	.048	.000	.095
	Aero_BC	-10.130	3.561	-2.845	.005	-17.180	-3.080
	Aero_RA	-11.838	4.263	-2.777	.006	-20.277	-3.398
	Aero_AIE	1.761	1.792	.983	.328	-1.787	5.310
	SST_high_lat	-.018	.241	-.076	.939	-.496	.459
	SST_mid_lat	.211	.116	1.818	.072	-.019	.441
	SST_low_lat	-.020	.059	-.343	.732	-.138	.097
L_O_Mean_	截距	34.884	28.498	1.224	.223	-21.539	91.308

Temp_A	Solar_Flux_mean	-8.511E-006	1.894E-005	-.449	.654	-4.601E-005	2.899E-005
	Urbanization	.055	.055	.987	.326	-.055	.164
	Emiss_of_Ozone	-3.960E-007	3.841E-007	-1.031	.305	-1.157E-006	3.645E-007
	SO2	-.009	.007	-1.296	.198	-.023	.005
	Land_Built_up	1.559E-007	7.959E-008	1.958	.052	-1.706E-009	3.134E-007
	Land_Graze	-1.047E-009	1.476E-009	-.709	.480	-3.970E-009	1.876E-009
	Land_Crop	-1.778E-009	5.033E-009	-.353	.725	-1.174E-008	8.187E-009
	annual_CO2	9.909E-012	4.072E-011	.243	.808	-7.072E-011	9.053E-011
	cumu_CO2	2.980E-012	6.542E-012	.456	.650	-9.973E-012	1.593E-011
	CFCsTG_OBS	-1.850	1.642	-1.127	.262	-5.100	1.401
	CH4_OBS	-6.631	4.183	-1.585	.116	-14.913	1.651
	CO2_OBS	-.004	.027	-.137	.892	-.058	.050
	N2O_OBS	-53.320	96.925	-.550	.583	-245.224	138.584
	Aero_SA	.038	.025	1.527	.129	-.011	.088
	Aero_BC	-8.002	3.744	-2.137	.035	-15.415	-.589
	Aero_RA	-8.171	4.482	-1.823	.071	-17.045	.704
	Aero_AIE	-1.827	1.885	-.970	.334	-5.559	1.904
	SST_high_lat	.216	.254	.853	.395	-.286	.719
	SST_mid_lat	.203	.122	1.660	.100	-.039	.444
	SST_low_lat	.015	.062	.237	.813	-.109	.138

#### 4.4.6 模型预测结果及检验

应用构建的多变量回归模型对全球天气进行拟合和预测，模型的拟合情况如下图 4-XXX 所示。模型对于历史数据的拟合程度较好，拟合值的趋势与真实趋势相同。预测得到的未来 25 年的全球平均温度和全球温度距平如表 4-7 所示。

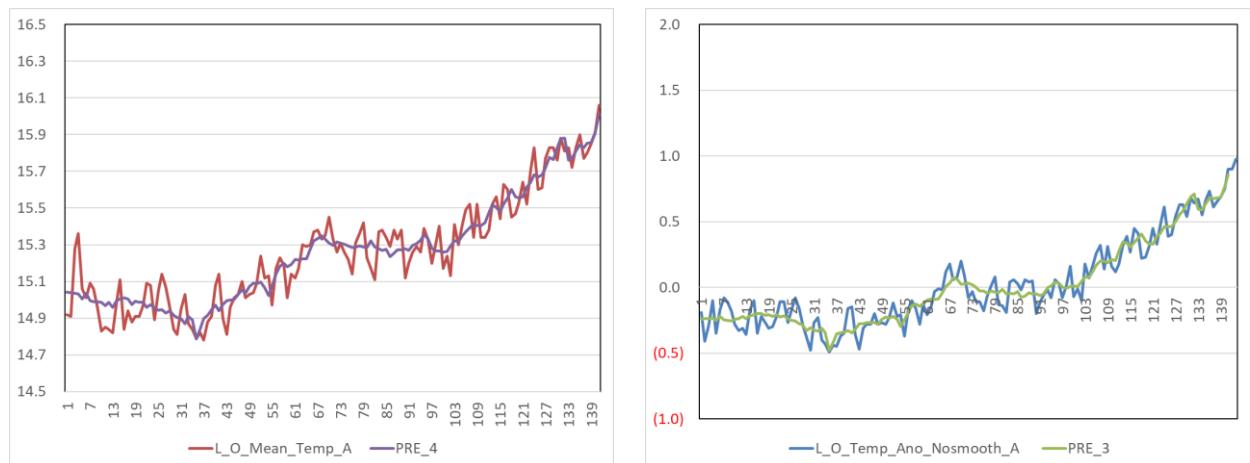
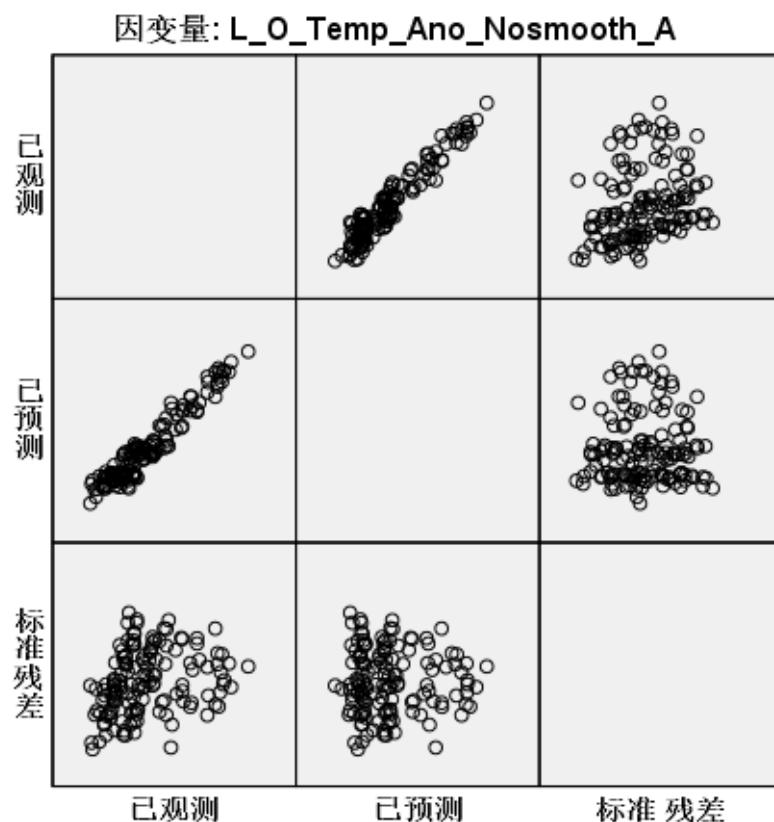


图 4-32 对于全球温度和距平的拟合预测结果

表 4-7 对于未来 25 年的全球温度预测结果

年份	2020	2021	2022	2023	2024
全球温度距平(°C)	1.78	2.03	2.27	2.47	2.66
全球平均温度(°C)	16.77	16.95	17.16	17.32	17.44
年份	2025	2026	2027	2028	2029
全球温度距平(°C)	2.91	3.07	3.26	3.40	3.56
全球平均温度(°C)	17.65	17.83	17.99	18.15	18.33
年份	2030	2031	2032	2033	2034
全球温度距平(°C)	3.81	4.05	4.33	4.57	4.77
全球平均温度(°C)	18.54	18.75	18.97	19.17	19.38
年份	2035	2036	2037	2038	2039
全球温度距平(°C)	5.11	5.31	5.49	5.79	6.00
全球平均温度(°C)	19.67	19.88	20.11	20.35	20.60
年份	2040	2041	2042	2043	2044
全球温度距平(°C)	6.26	6.47	6.70	6.98	7.14
全球平均温度(°C)	20.85	21.13	21.37	21.61	21.75



模型 : 截距 + Solar\_Flux\_mean + Urbanization + Emiss\_of\_Ozone + SO2 + Land\_Built\_up + Land\_Graze + Land\_Crop + annual\_CO2 + cumu\_CO2 + CFCsTG\_OBS + CH4\_OBS + CO2\_OBS + N2O\_OBS + Aero\_SA + Aero\_BC + Aero\_RA + Aero\_AIE + SST\_high\_lat + SST\_mid\_lat + SST\_low\_lat

图 4-33 多变量回归模型结果的残差图

回归模型的残差图反映了模型的线性拟合度较好。

## 4.5 基于随机森林算法的多变量非线性回归模型

### 4.5.1 随机森林算法的概念及过程

随机森林算法是一种较新的高度灵活的一种机器学习算法。该算法的基本单元是决策树，但其本质属于机器学习中的方法。

随机森林算法主要过程：

(1) 样本集的选择

假设原始样本集总共有  $N$  个样例，则每轮从原始样本集中通过 Bootstrapping（有放回抽样）的方式抽取  $N$  个样例，得到一个大小为  $N$  的训练集。在原始样本集的抽取过程中，可能有被重复抽取的样例，也可能有一次都没有被抽到的样例。

共进行  $k$  轮的抽取，则每轮抽取的训练集分别为  $T_1, T_2, \dots, T_k$ 。

(2) 决策树的生成

假如特征空间共有  $D$  个特征，则在每一轮生成决策树的过程中，从  $D$  个特征中随机选择其中的  $d$  个特征 ( $d < D$ ) 组成一个新的特征集，通过使用新的特征集来生成决策树。

在  $k$  轮中共生成  $k$  个决策树，由于这  $k$  个决策树在训练集的选择和特征的选择上都是随机的，因为这  $k$  个决策树之间是相互独立的。

(3) 模型的组合

由于生成的  $k$  个决策树之间是相互独立的，每个决策树的重要性是相等的，因而在将它们进行组合时，无需考虑它们的权值，或者可以认为它们具有相同的权值。对于分类问题，最终的分类结果使用所有的决策树投票来确定最终分类结果；对于回归问题，使用所有决策时输出的均值来作为最终的输出结果。

(4) 模型的验证

模型的验证需要验证集，但不需要额外专门再获取验证集，只需要从原始样本集中选择没有被使用过的样例即可。

在从原始样本中选择训练集时，存在部分样例一次都没有被选中过，在进行特征选择时，也可能存在部分特征未被使用的情况，我们只需将这些未被使用的数据拿来验证最终的模型即可。

随机森林算法主要有以下 2 个特性：

(1) 有两个随机抽取过程：分别为从原始样本集中随机抽取训练集，和决策树的特征选择上随机抽取部分特征生成决策树。

(2) 各决策树之间的相互独立性。因为相互独立，故而在决策树的生成过程可以并行进行，大大提高了算法的时间效率。

### 4.5.2 模型的构建与结果

采用 Python 建立基于随机森林的多变量非线性回归模型，模型的拟合与预测结果如图 4-34 所示。

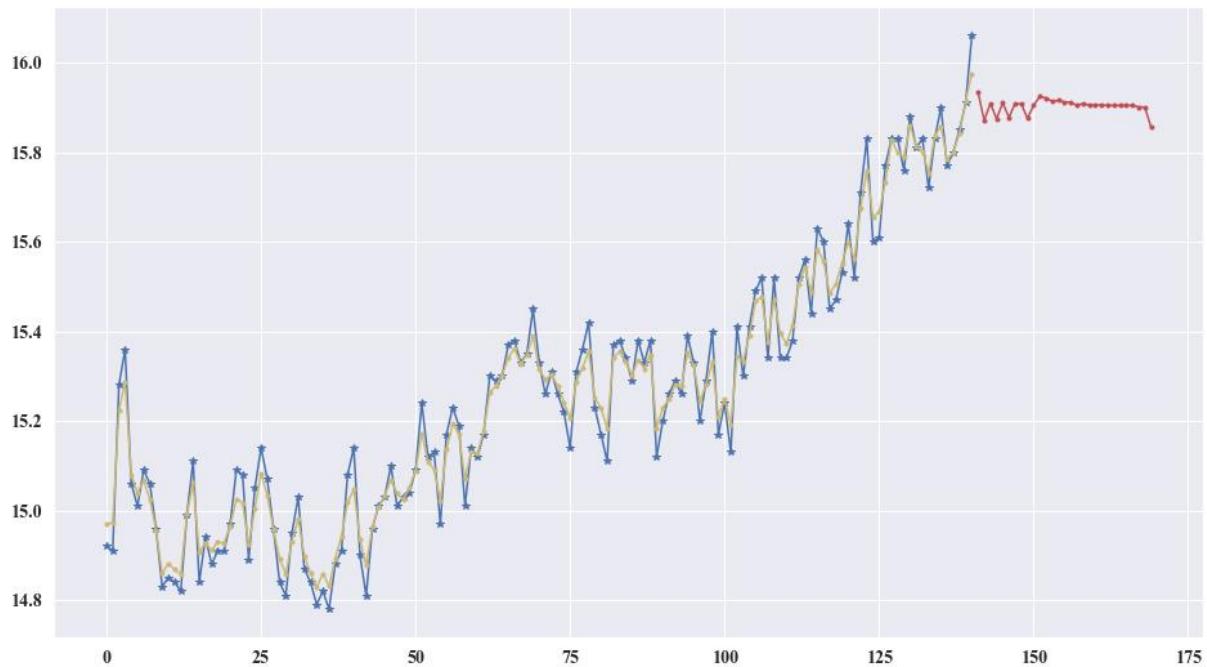


图 4-34 随机森林算法的拟合与预测结果

## 4.6 基于 ARIMA 时间序列模型的预测

### 4.6.1 ARIMA 模型

ARIMA 全称为自回归积分滑动平均模型(Autoregressive Integrated Moving Average Model, 简记 ARIMA), ARIMA(p,d,q)模型是针对非平稳时间序列所建立的模型。ARIMA 的含义包含 3 个部分, 即 AR、I、MA。

其中:

AR 表示 auto regression, 即自回归模型;

I——表示 integration, 即单整阶数, 时间序列模型必须是平稳性序列才能建立计量模型, ARIMA 模型作为时间序列模型也不例外, 因此首先要对时间序列进行单位根检验, 如果是非平稳序列, 就要通过差分来转化为平稳序列, 经过几次差分转化为平稳序列, 就称为几阶单整;

MA——表示 moving average, 即移动平均模型。可见, ARIMA 模型实际上是 AR 模型和 MA 模型的组合。

相应的, 有三个参数:  $p, d, q$ 。其中:

$p$ —代表预测模型中采用的时序数据本身的滞后数(lags), 也叫做 AR/Auto-Regressive 项

$d$ —代表时序数据需要进行几阶差分化, 才是稳定的, 也叫 Integrated 项。

$q$ —代表预测模型中采用的预测误差的滞后数(lags), 也叫做 MA/Moving Average 项

因而, 用数学形式表示为(假设  $p, q, d$  已知):

$$\hat{y}_t = \mu + \phi_1 * y_{t-1} + \dots + \phi_p * y_{t-p} + \theta_1 * e_{t-1} + \dots + \theta_q * e_{t-q}$$

其中,  $\phi$  表示 AR 的系数,  $\theta$  表示 MA 的系数

## 4.6.2 全球气候的时间序列预测模型

### (1) 逐年和逐月数据的序列构建

由于可获取自 1850 年至今的逐月的全球温度数据，数据量充足，对数据进行年汇总可得到逐年的全球温度数据。逐月的全球温度数据和逐年的全球温度数据对应的序列图分别如图 4-35 和 4-36 所示。从逐月的数据图可以分析得到全球温度具有周期性，逐年数据图可以分析出全球温度有逐年上升的趋势。

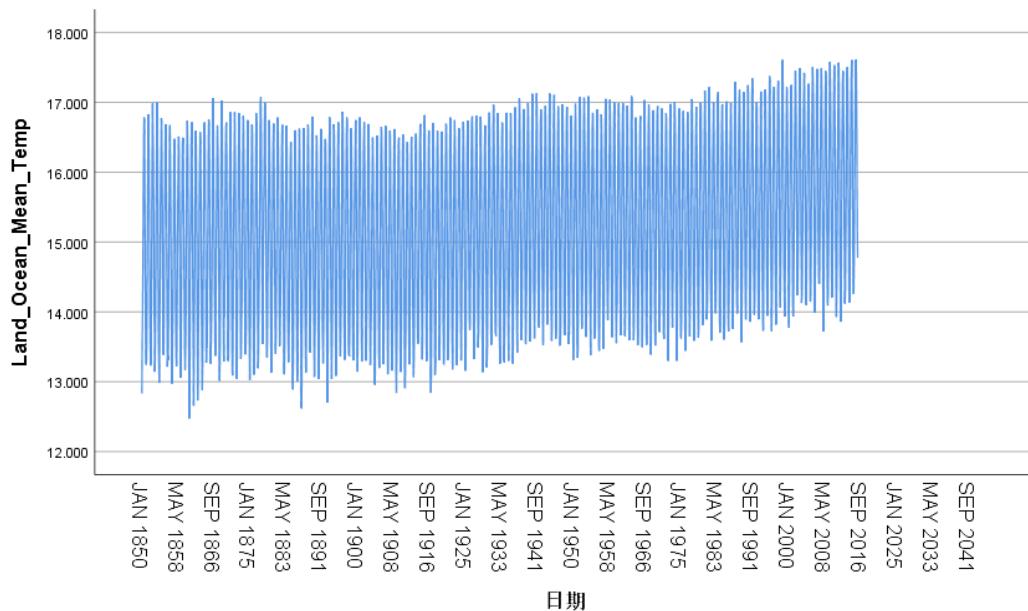


图 4-35 逐月的全球陆地与海洋的平均温度

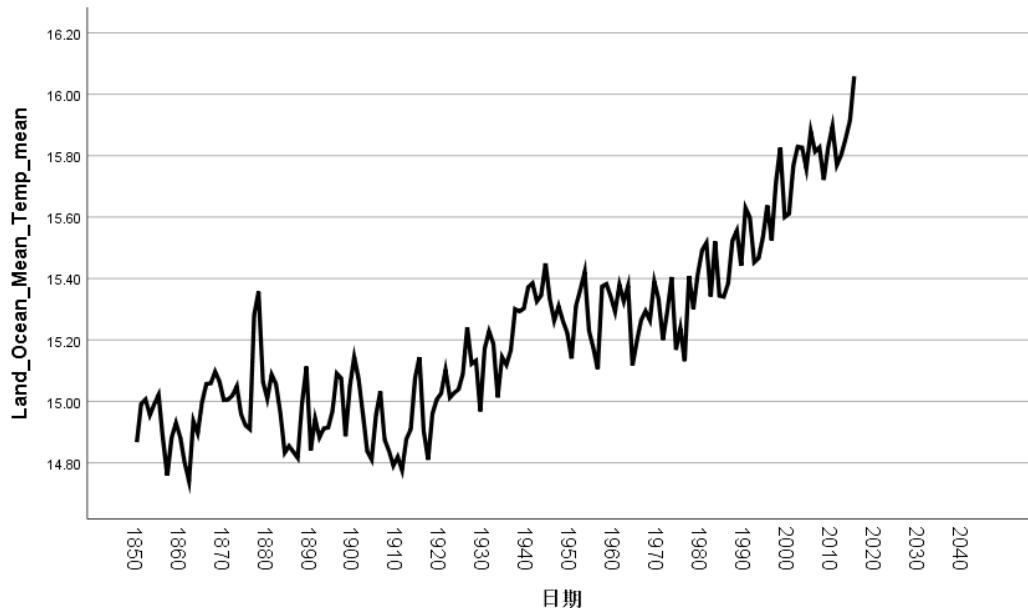


图 4-36 逐月的全球陆地与海洋的平均温度

## (2) 基于逐月数据的时间序列模型预测

对于逐月数据的序列进行一次差分后如下图 4-37 所示。

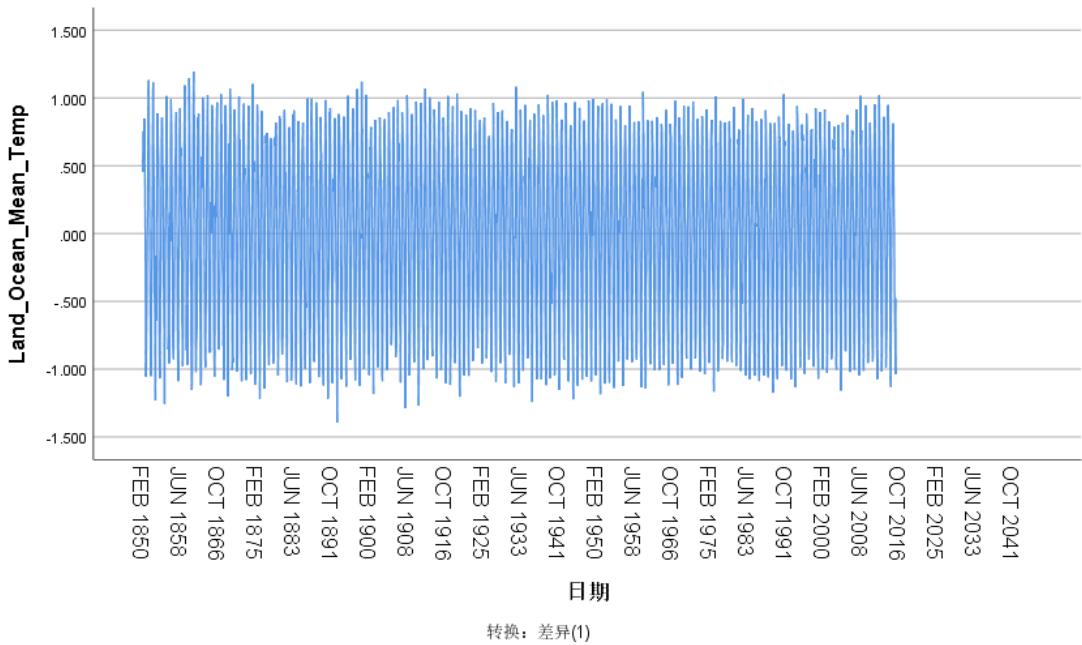


图 4-37 逐月数据进行一次差分后的结果

对于逐月数据进行平稳性、周期性、自相关性和偏相关性分析结果，为ARIMA模型的构建和参数优化提供指导。经过分析和多次参数调优后，对逐月数据采用ARIMA(1, 0, 3)(0, 1, 1)模型进行拟合和预测时，模型的拟合预测效果最佳。

模型统计如表4-8所示。

表 4-8 对于逐月数据的时间序列模型的参数统计

模型	模型拟合度统计		杨-博克斯 Q(18)			离群值数
	平稳 R 方	R 方	统计	DF	显著性	
Land_Ocean_Mean_Temp-模型_1	.641	.992	15.019	13	.306	0

该模型的R方达到0.992，模型结果具有很好的可信度。

模型的残差 ACF 和 PACF 图如下图 4-38 所示。

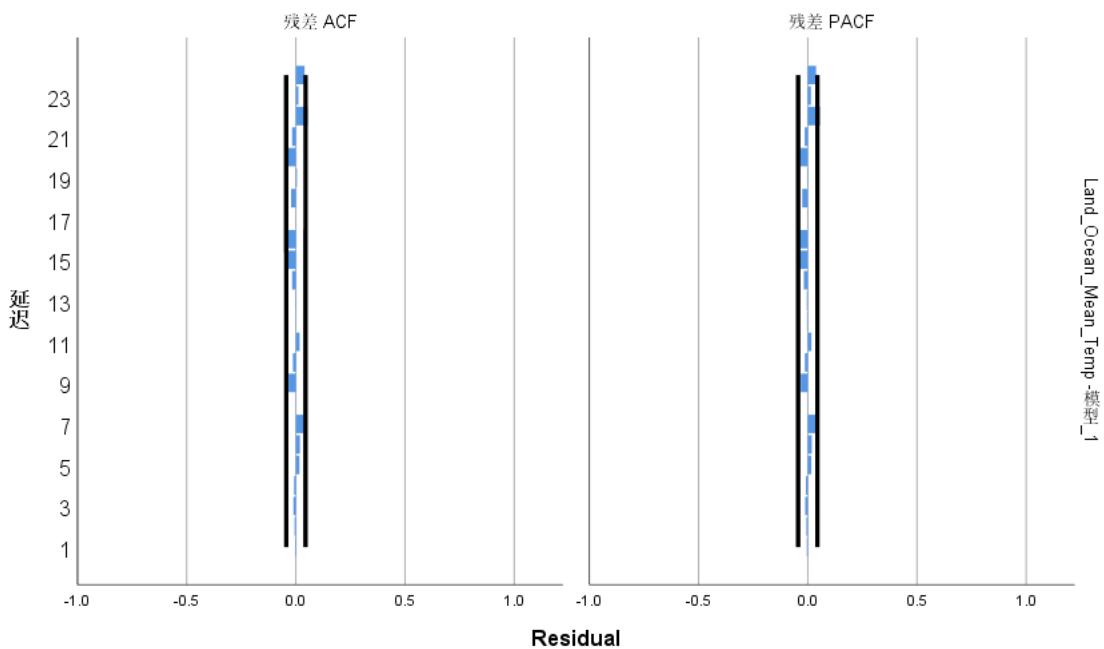


图 4-38 基于逐月数据的时间序列模型的残差图

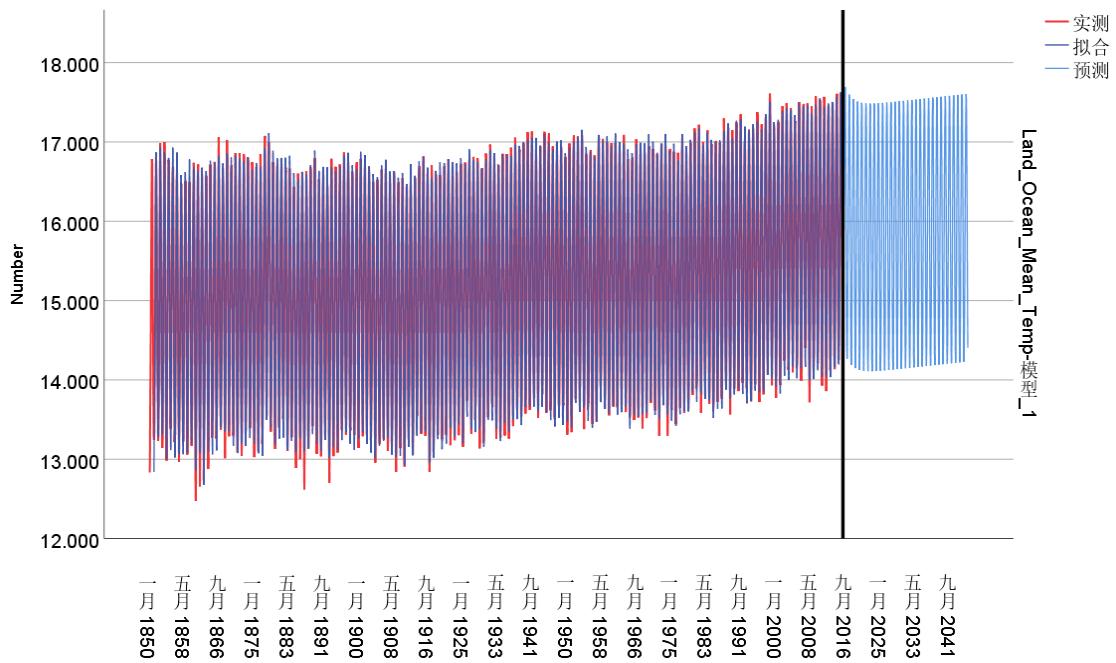


图 4-39 基于逐月数据的时间序列模型的拟合与预测结果

## (2) 基于逐年数据的时间序列模型预测

对于逐年数据的序列进行一次差分后如下图 4-40 所示。

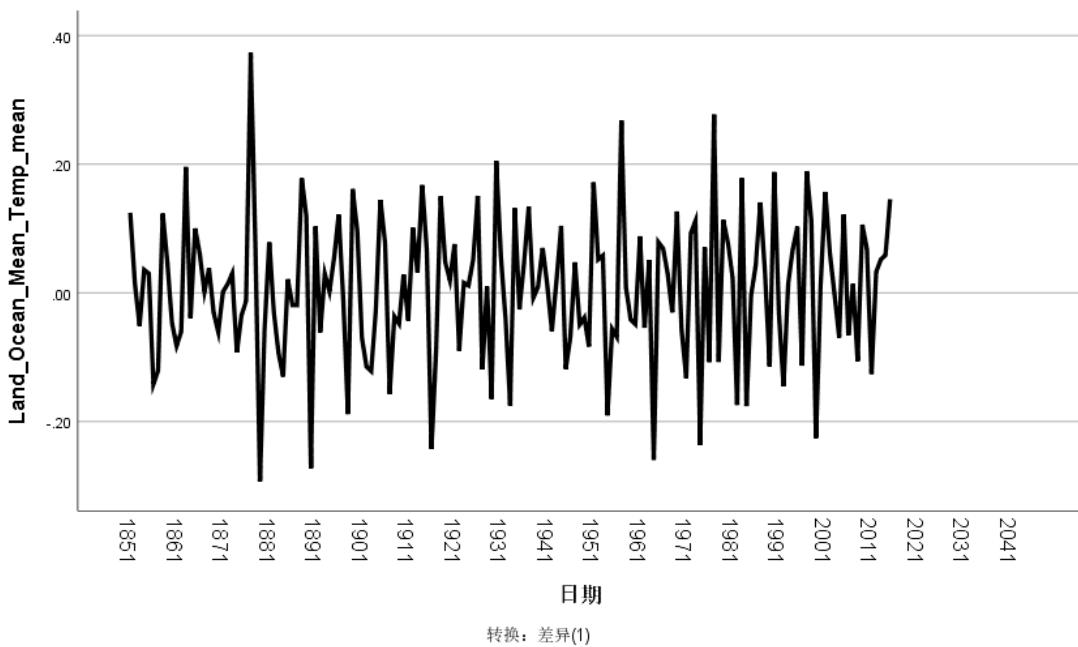


图 4-40 逐年数据进行一次差分后的结果

一次差分后的序列图平稳性有显著的提高，进一步印证全球陆地-海洋的平均温度具有逐渐上升的趋势。

类似地，对于逐年数据进行平稳性、周期性、自相关性和偏相关性分析结果，为ARIMA模型的构建和参数优化提供指导。经过分析和多次参数调优后，对逐月数据采用ARIMA(3,1,0)模型进行拟合和预测时，模型的拟合预测效果最佳。

模型统计如表4-9所示。

表 4-9 对于逐年数据的时间序列模型的参数统计

模型	模型拟合度统计		杨-博克斯 Q(18)			离群值数
	平稳 R 方	R 方	统计	DF	显著性	
Land_Ocean_Mean_Tem p_mean-模型_1	0.172	0.884	18.498	15	.237	0

该模型的R方为0.884，模型结果具有很好的可信度。

模型的残差ACF和PACF图如下图4-41所示。

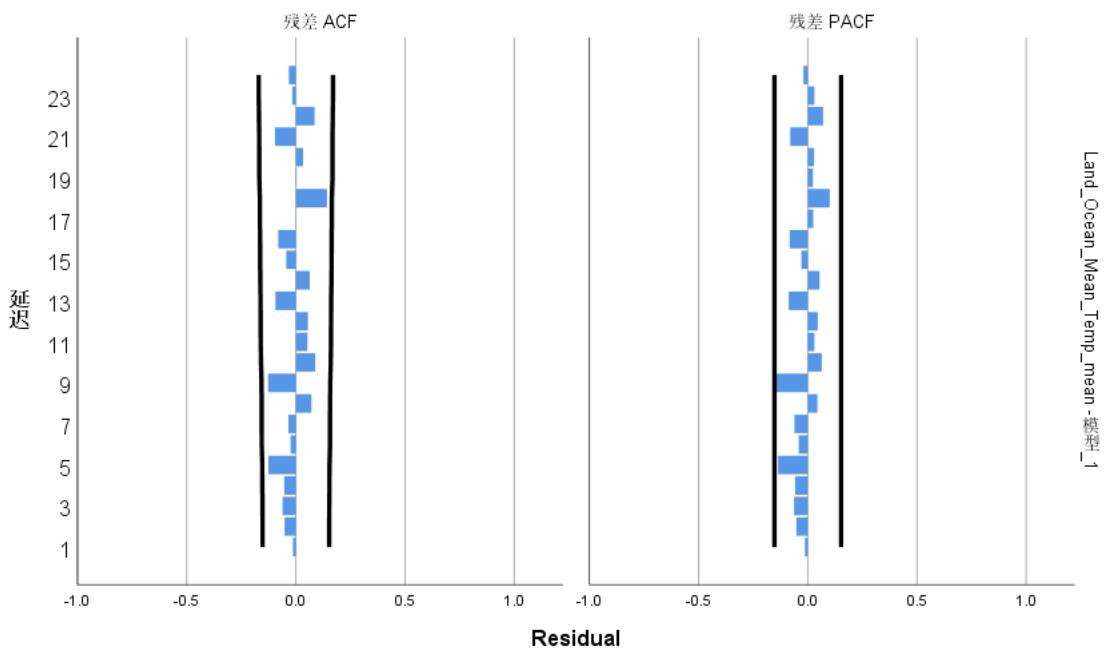


图 4-41 基于逐年数据的时间序列模型的残差图

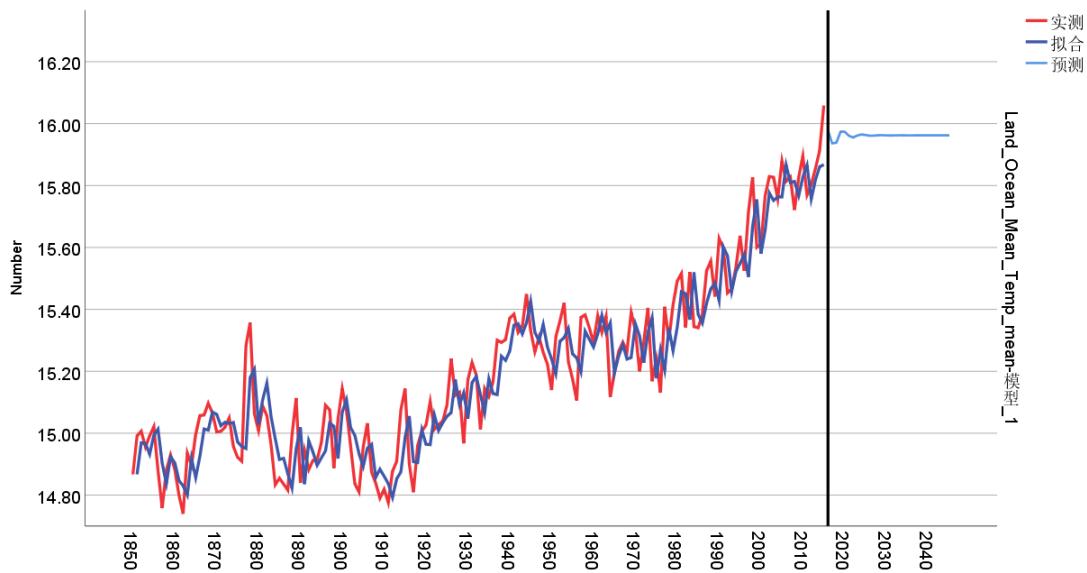


图 4-42 基于逐年数据的时间序列模型的拟合与预测结果

#### (4) 对于全球温度距平的时间序列预测

基于与上述相同的思路及方法，对全球温度距平的历史数据进行分析，构建时间序列及模型，对全球温度距平进行预测，采用 python 运行得到模型的预测及拟合情况如下图所示。

```

Statespace Model Results
=====
Dep. Variable: anomaly No. Observations: 139
Model: SARIMAX(2, 1, 4) Log Likelihood 122.480
Date: Sat, 21 Sep 2019 AIC -230.960
Time: 14:20:07 BIC -210.469
Sample: 12-31-1880 HQIC -222.633
- 12-31-2018
Covariance Type: opg
=====
              coef    std err      z   P>|z|   [0.025   0.975]
-----
ar.L1        0.0166   0.073   0.229   0.819   -0.126   0.159
ar.L2        0.9448   0.075  12.672   0.000    0.799   1.091
ma.L1       -0.3106   0.118  -2.628   0.009   -0.542  -0.079
ma.L2       -1.2318   0.109 -11.301   0.000   -1.445  -1.018
ma.L3        0.2507   0.096   2.608   0.009    0.062   0.439
ma.L4        0.3466   0.083   4.169   0.000    0.184   0.510
sigma2       0.0098   0.002   6.278   0.000    0.007   0.013
-----
Ljung-Box (Q): 41.25 Jarque-Bera (JB): 2.83
Prob(Q): 0.42 Prob(JB): 0.24
Heteroskedasticity (H): 1.21 Skew: -0.09
Prob(H) (two-sided): 0.53 Kurtosis: 2.32
-----
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

```

图 4-43 基于全球温度距平的逐年数据的 ARIMA 模型拟合结果报表

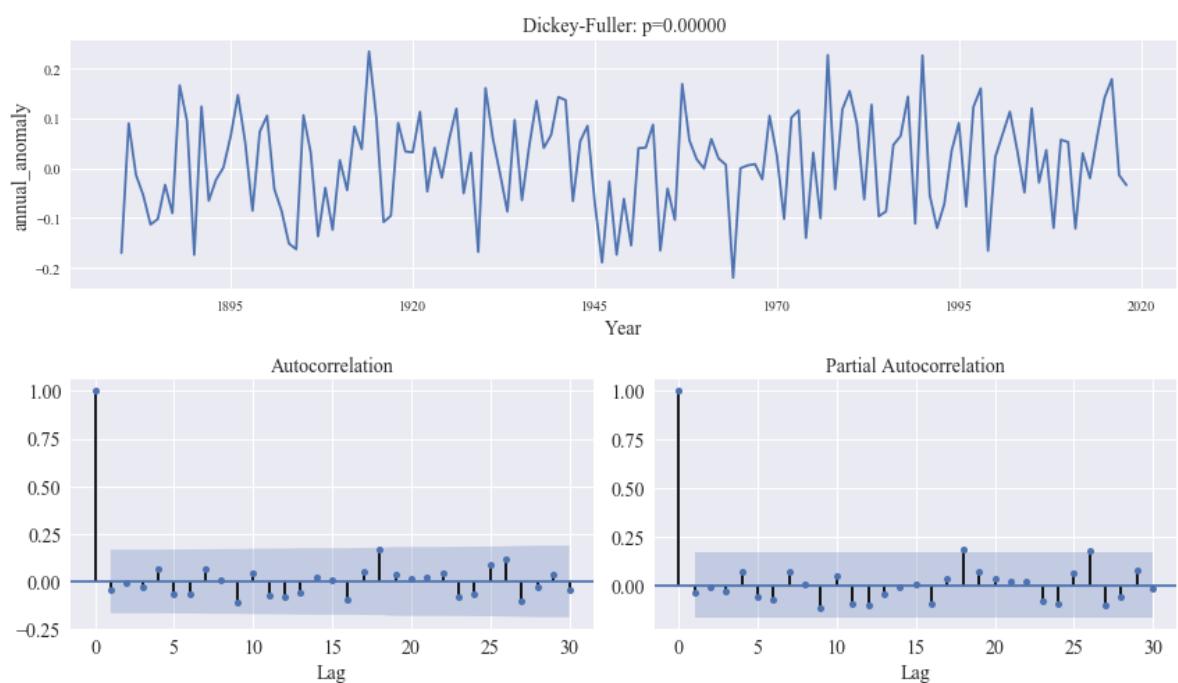


图 4-44 基于全球温度距平的逐年数据的 ARIMA 模型的残差可视化分析

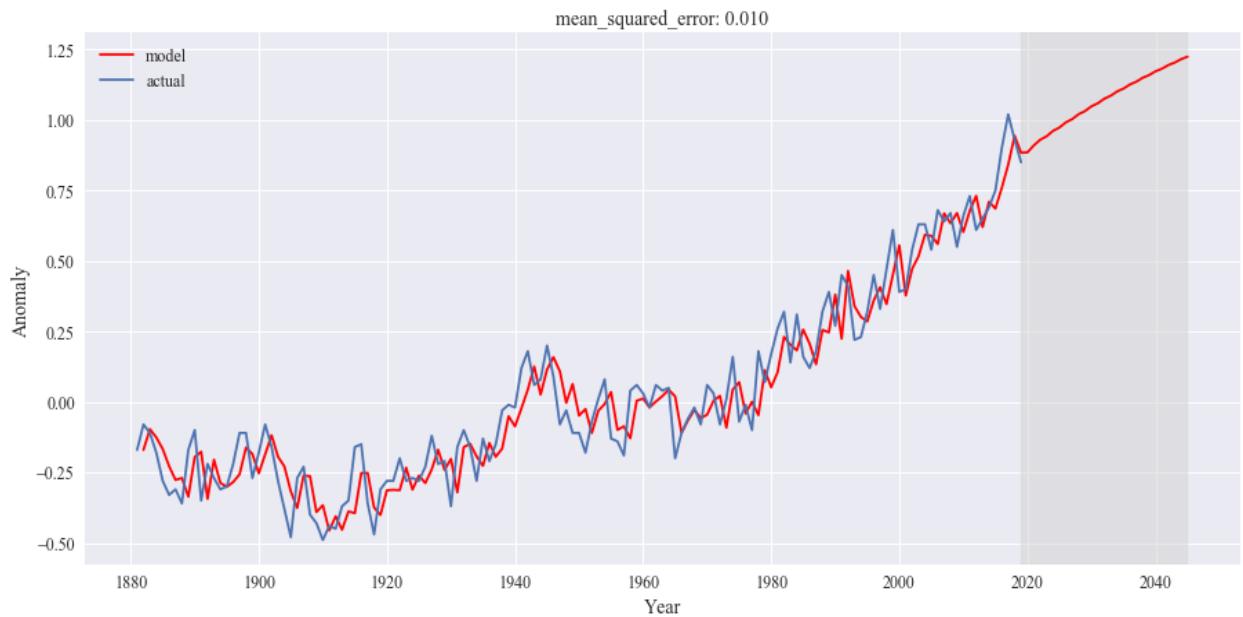


图 4-45 对全球温度距平的 ARIMA 模型的拟合与预测结果

### 4.6.3 时间序列模型的预测结果

对上述时间序列模型的预测未来 25 年的全球温度的结果整合如下表 4-10 所示。

表 4-10 对于未来 25 年全球气候的模型预测结果

年份	2020	2021	2022	2023	2024	2025	2026
全球平均温度 Tm(°C)	15.84	15.84	15.83	15.83	15.84	15.84	15.85
全球平均温度 Ty(°C)	15.97	15.96	15.96	15.96	15.97	15.96	15.96
年份	2027	2028	2029	2030	2031	2032	2033
全球平均温度 Tm(°C)	15.85	15.86	15.86	15.87	15.87	15.88	15.88
全球平均温度 Ty(°C)	15.96	15.96	15.96	15.96	15.96	15.96	15.96
年份	2034	2035	2036	2037	2038	2039	2040
全球平均温度 Tm(°C)	15.89	15.9	15.9	15.91	15.91	15.92	15.92
全球平均温度 Ty(°C)	15.96	15.96	15.96	15.96	15.96	15.96	15.96
年份	2041	2042	2043	2044	2045		
全球平均温度 Tm(°C)	15.93	15.94	15.94	15.95	15.95		
全球平均温度 Ty(°C)	15.96	15.96	15.96	15.96	15.96		

对于未来 25 年的全球温度距平的预测结果如表 4-8 所示。

表 4-8 对于未来 25 年的全球温度距平的预测结果

年份	2019	2020	2021	2022	2023	2024	2025
全球温度距平 (°C)	0.879	0.911	0.928	0.946	0.963	0.980	0.996
年份	2026	2027	2028	2029	2030	2031	2032
全球温度距平 (°C)	1.013	1.029	1.045	1.061	1.077	1.092	1.107
年份	2033	2034	2035	2036	2037	2038	2039

全球温度距平 (°C)	1.122	1.137	1.152	1.167	1.181	1.195	1.209
年份	2040	2041	2042	2043	2044		
全球温度距平 (°C)	1.223	1.237	1.250	1.264	1.277		

## 4.7 三种模型的优缺点及推广可行性

本章对于全球气候变化的数据进行了特征分析，并搜集整理可能与全球气候变化存在关联性的各种因子的数据集。构建了三种全球气候模型，下面对三种模型进行比较分析，讨论其推广应用的可行性。

### 4.7.1 模型 1：全球气候变化多变量线性回归模型

该模型的机理较为简单，建立多变量与因变量之间的关联性，从而可以直接判断求证出对因变量存在影响的因子及影响程度。模型 1 的结果也很好地展现了这一优点，可以较为科学合理地从数据上得出全球气候变化的影响因素及其影响程度。

模型 1 在做全球气候变化的未来预测时，需要将历史数据与未来的因变量之间建立线性关系，虽然得到的线性较好，但这种预测方法存在一定的滞后性，也与预测数据集的选取有较大关系，从而会使得预测结果受较近历史数据趋势影响，较大程度上损失数据集的特征，可能造成预测结果的偏差。

本章中采用模型 1 的预测结果即表明了这一特性，预测出的未来 25 年天气数据延展了近几年的趋势，导致存在一定程度上的偏离。

### 4.7.2 模型 2：基于随机森林算法的多变量非线性回归模型

该模型作为一种较新的机器学习算法为基础的模型，其灵活程度较高，可以从多变量的历史数据集中选取一定的数据作为训练样本。经过参数调优后，拟合值与训练样本的实际值的拟合程度非常高，模型应用的结果也体现了这一点。

从预测的结果来看，对于未来长周期的预测，仅在短期的未来保留了历史数据及的振荡等特征，在较长期的未来范围内存在偏离，但好于模型 1。然而模型 2 中难以直观地给出多变量或因子与因变量（全球气候变化）的关联性，并且模型 2 的预测准确性容易受数据集的大小与训练样本的影响。

模型 2 的推广应用对于数据集的大小、质量及训练样本的要求较高，当数据集能够满足其要求时，该模型本身的拟合效果较好，可以改善其对于历史数据规律的学习程度，用以指导和提高对于未来预测的准确与合理性。

### 4.7.3 模型 3：基于 ARIMA 的时间序列模型

ARIMA 模型采用自回归的方式，能够仅根据单一变量的历史数据对其未来值做预测。仅从对于未来做预测的层面上来分析，ARIMA 模型的预测结果的合理性要优于模型 1 和模型 2，特别是对全球温度的逐月数据（数据量大）做自回归及预测时，结果的合理性很高。但是 ARIMA 模型也难以给出多变量或因子与因变量（全球气候变化）的关联性。对于 ARIMA 模型的改进与推广，可以从模型 1 中得到启示，可以建立多变量 ARIMA 模型，将多种变量的历史数据与因变量之间建立关系，并采用时间序列模型对其进行预测。

## 五. 极端天气模型的构建与应用

在全球变暖的大趋势下，世界局部地区的极寒天气也继续发生。2019年1月美国2/3的地区变成了一个大冰窖，出现了“几十年一遇”的极度寒冷天气，成为有人怀疑全球变暖的依据之一。但是由于天气和气候的概念并不相同，天气是一定区域短时间内的大气状态及其变化的总称，而气候是长时间内气象要素和天气现象的统计状态。全球气候变暖是从全球范围内长时间的观测和统计，而局地天气极寒等极端天气是在局部地区短时间内发生的现象。

为了揭示极寒天气等极端气象与全球气候变化的关系，本章将尝试界定极端天气将其量化分析，分析其与全球气候的关系，建立极端天气与全球气候的关联分析模型，在模型求解结果的支撑下探讨两者间的关系。

此外，由于目前研究和实际应用中对于极端天气的界定范围较广，其机理也有待研究探讨，本章在解决问题三的基础上，尝试进一步分析极端天气的影响因素，建立多因素的极端天气回归模型，为后续研究提供支撑。

### 5.1 问题三的分析与求解思路

问题三要求分析极寒天气等极端气象的出现与气候变化之间是否存在关系，并需建立相应的模型来分析全球变暖与局地极寒现象的关系。

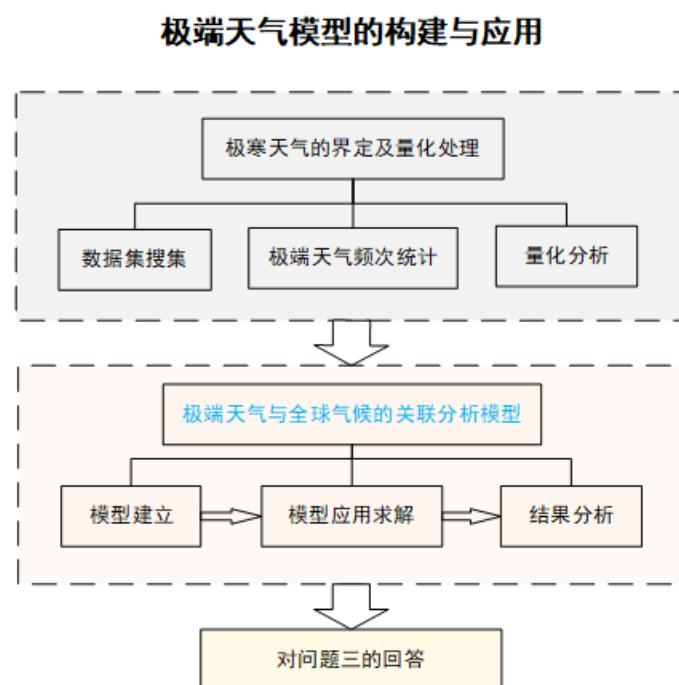


图 5-1 问题三的解决思路

为了分析不同空间范围和时间纬度上的两个概念间的关系，需将两者进行量化并同步到相同的时间和空间范围。

本章先将极端气象用温度来表征，分析温度中的极寒情况，由于没有专门记录全球极端寒冷的数据集，需先搜集世界相关组织提供的全球范围温度的数据集并对其进行提取和分析，提取极端天气的频次数据得到极端天气的量化数据。分析极端天气数据与全球气候

数据的关联性，构建极端天气与全球气候的关联性模型，运用 Python 和 SPSS 软件处理数据及求解模型，依据模型求解结果，讨论并解释极端天气和全球气候间的关系。

## 5.2 极端天气模型的建立与应用

由于已有的数据集可以提供全球范围内各经纬度网格内的温度分布数据，可对逐月及逐年数据进行分析。界定极端天气后尝试将极端天气进行量化，从而提取出其量化数据，将量化后的极端天气数据与全球气候作关联性分析，建立关联性模型。

### 5.2.1 本模型中极端天气的界定

由于全球范围内各经纬度网格的温度数据集是以温度距平进行记录的，因此本模型中采用温度距平对天气进行刻画。

在本模型中，为得到极端天气的量化数据，对极端天气进行界定。通过对全球范围内的温度分布数据进行多次的整合分析后，本模型中采用以下 2 种界定方法将极端天气进行界定并量化。

- a) 统计全球范围内各经纬度网格内温度距平  $T_{\text{Ano}}$  的绝对值超过  $2^{\circ}\text{C}$  时对应极热天气( $T_{\text{Ano}} > 2^{\circ}\text{C}$ )和极寒天气( $T_{\text{Ano}} < -2^{\circ}\text{C}$ )的各自总频次(每年)。
- b) 先统计分析全球范围内的各经纬度网格内温度距平  $T_{\text{Ano}}$  的均值  $\mu$  和标准差  $\sigma$ ，极热天气对应为  $T_{\text{Ano}} > \mu + 3\sigma$ ，极寒天气对应为  $T_{\text{Ano}} < \mu - 3\sigma$ ，分别统计每年即极热天气和极寒天气在全球范围内对应的总频次。

### 5.2.2 极端天气的数据处理与分析

通过 Python 对全球范围内各经纬度网格的温度距平进行分析，并统计极热天气和极寒天气对应的年度总频次。

```
global_temp_avg = np.zeros(temanomaly.shape[0])
global_temp_std = np.zeros(temanomaly.shape[0])
extreme_high_temp_cnt = np.zeros(temanomaly.shape[0])
extreme_low_temp_cnt = np.zeros(temanomaly.shape[0])
sigma3_high_temp_cnt = np.zeros(temanomaly.shape[0])
sigma3_low_temp_cnt = np.zeros(temanomaly.shape[0])
for year, xy_anomaly in enumerate(temanomaly):
    xy_anomaly_dropna = xy_anomaly[np.where(xy_anomaly == 32767, False, True)]

    year_temp_avg = xy_anomaly_dropna.mean()
    year_temp_std = xy_anomaly_dropna.std()
    global_temp_avg[year] = year_temp_avg
    global_temp_std[year] = year_temp_std
    extreme_high_temp_cnt[year] = np.where(xy_anomaly_dropna > 2, 1, 0).sum() # year_temp_avg + 4 * year_temp_std
    extreme_low_temp_cnt[year] = np.where(xy_anomaly_dropna < -2, 1, 0).sum() # year_temp_avg - 4 * year_temp_std
    sigma3_high_temp_cnt[year] = np.where(xy_anomaly_dropna > year_temp_avg + 3 * year_temp_std, 1, 0).sum() #
    sigma3_low_temp_cnt[year] = np.where(xy_anomaly_dropna < year_temp_avg - 3 * year_temp_std, 1, 0).sum() #

global_temp.plot(subplots=True, figsize = (20, 12))
```

图 5-2 运用 Python 对全球全网格温度 nc 文件数据处理的主要代码

具体代码请详见附录 1。

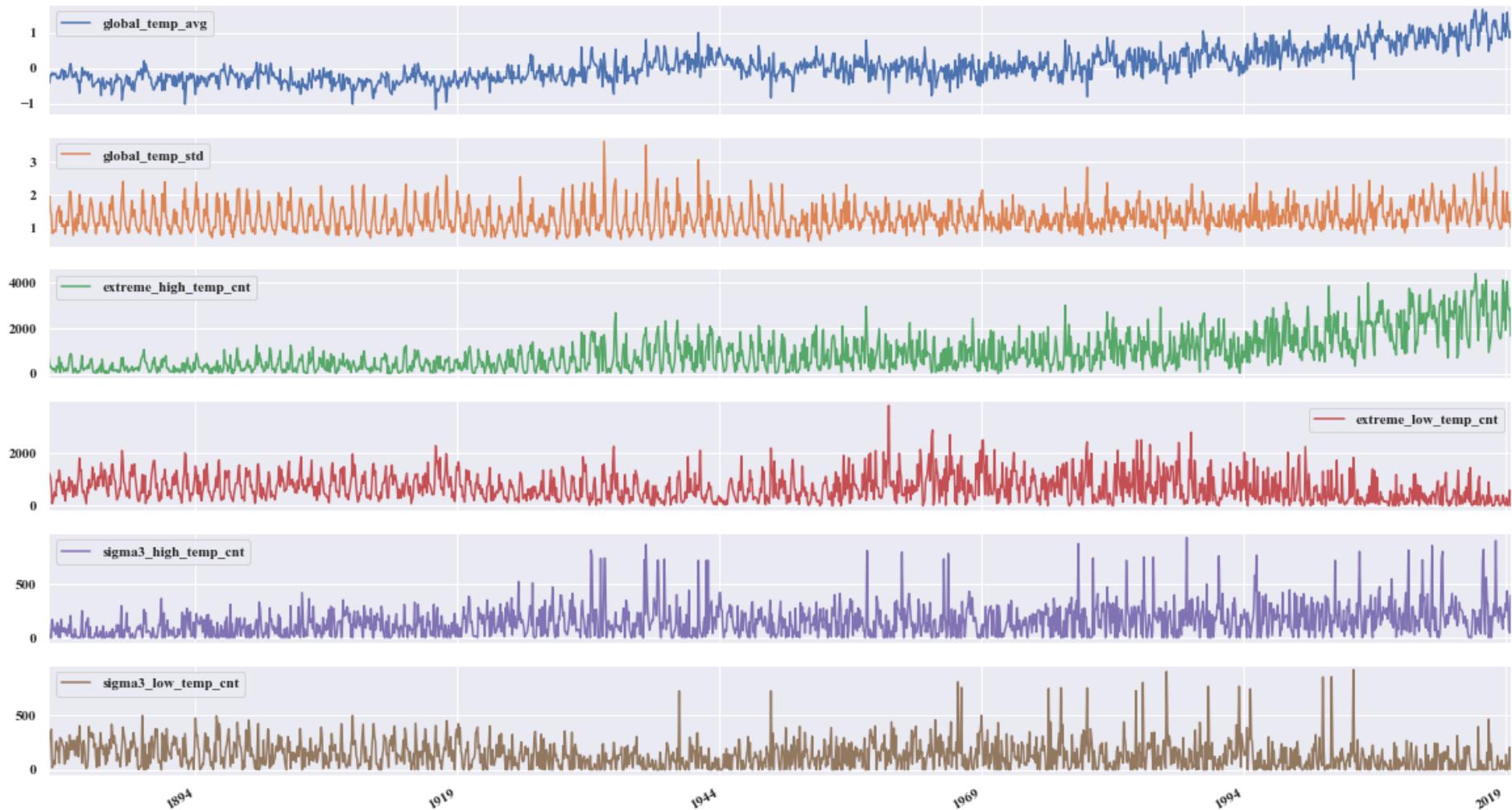


图 5-3 极端天气的量化数据提取分析

### 5.2.3 基于统计学的极端天气与全球气候间的关系分析

为了理解局地极端天气与全球气候间的关系，从统计学的角度，采用以下示意图来尝试解释。

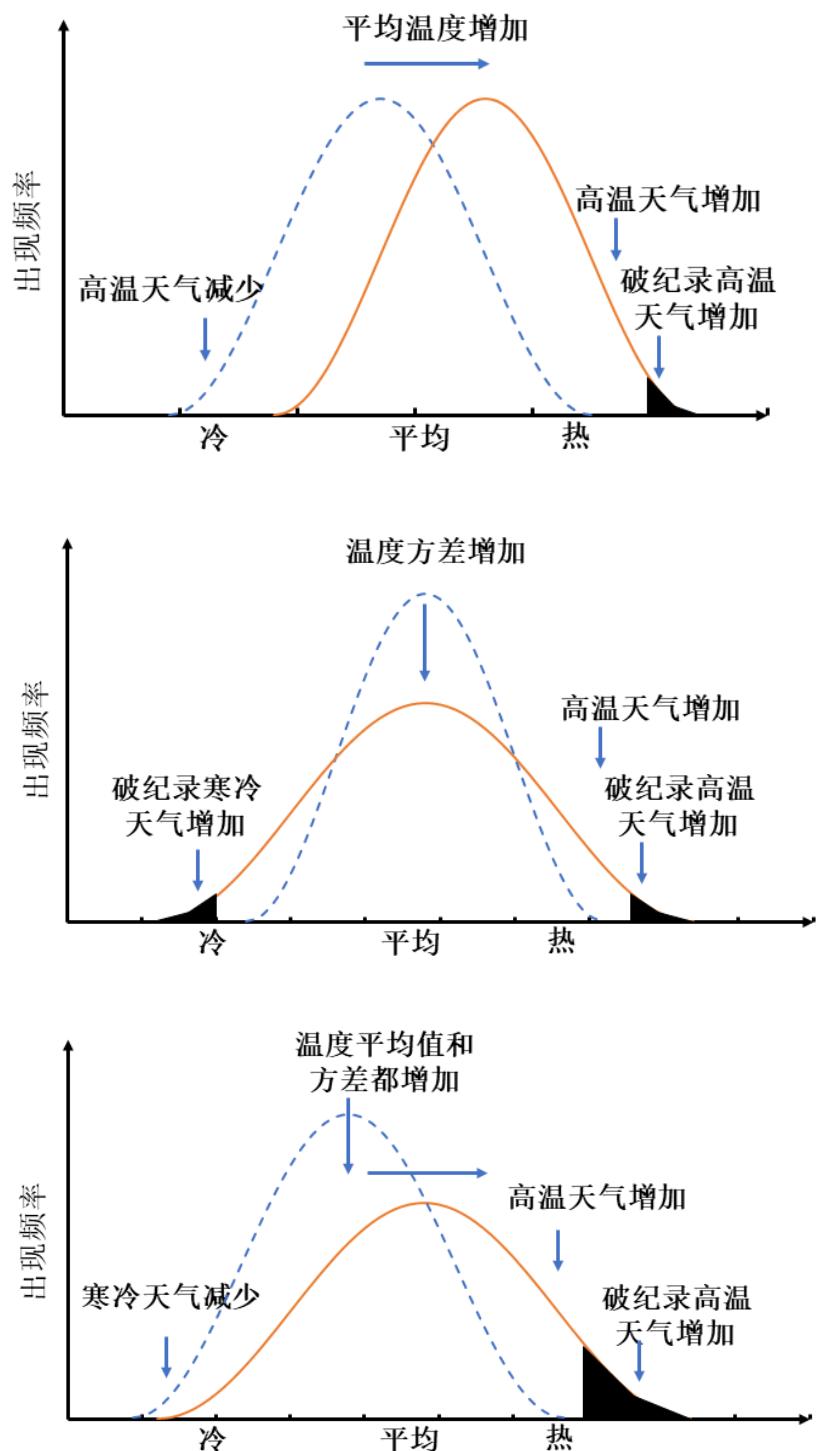


图 5-4 基于统计学的极端天气与全球气候间关系的示意图

从上至下依次对应为：A 模式—平均值增加；B 模式—方差增加；C 模式—两者均增加

图 5-4 表明全球温度数据的平均值增加，即全球平均温度增加（气候变暖）和温度数据的方差（对平均值离差的平均值）增加对极端温度的影响。

全球温度数据经常是呈正态分布的，但这种正态分布具有非定常性或不稳定性，这是由平均值或方差的变化引起的。

采用正态分布图的移动趋势，可以较清晰地解释全球极端气候变化的模式，及其与全球温度分布间的关系。主要可以分为三种模式：A 模式，B 模式和 C 模式。

(1) A 模式：平均值的增加可以导致新的破纪录的高温值（图 5-4 中的顶部示意图），但在这种情况下，方差并不发生变化，由图 1a 可注意到，最冷和最暖温度的范围并不变，只是整个正态分布向高温（极热天气）一方移动，从而比原来平均气候条件下出现更多的极热天气。

(2) B 模式：如果平均值不变，而方差增加，则使极寒天气和极热天气的出现概率和绝对值都会增加（图 5-4 中的中间示意图）。

(3) C 模式：平均值和变率都会发生变化（图 5-4 中的底部示意图），这时可影响极寒天气和极热天气发生的概率。在这种模式下，极热天气将更频繁地发生，而极寒天气的频次或概率则有所降低。平均值和方差都发生变化，可有不同的组合，则导致的气候极端事件出现的结果也不同。

#### 5.2.4 极端天气与全球气候的关联分析模型应用

通过 Python 对极端天气量化后的数据与全球温度距平建立关联分析模型，得出的关联结果如下图 5-5 所示。

从模型结果可看出极端天气量化指标与全球平均温度距平之间有着较为明显的相关关系。极热天气年度总频次与全球平均温度距平存在着较为显著的正相关关系，随着全球平均温度距平的升高，全球发生极热天气的频次升高。极寒天气年度总频次与全球平均温度距平存在着较明显的负相关关系，随着全球平均温度距平的升高，全球发生极热天气的频次有所下降。

结合上一小节中对于极端天气与全球气候的关系里的 C 模式（温度分布逐渐向右移动，温度升高，极热天气频次增加，极寒天气频次有所降低）。

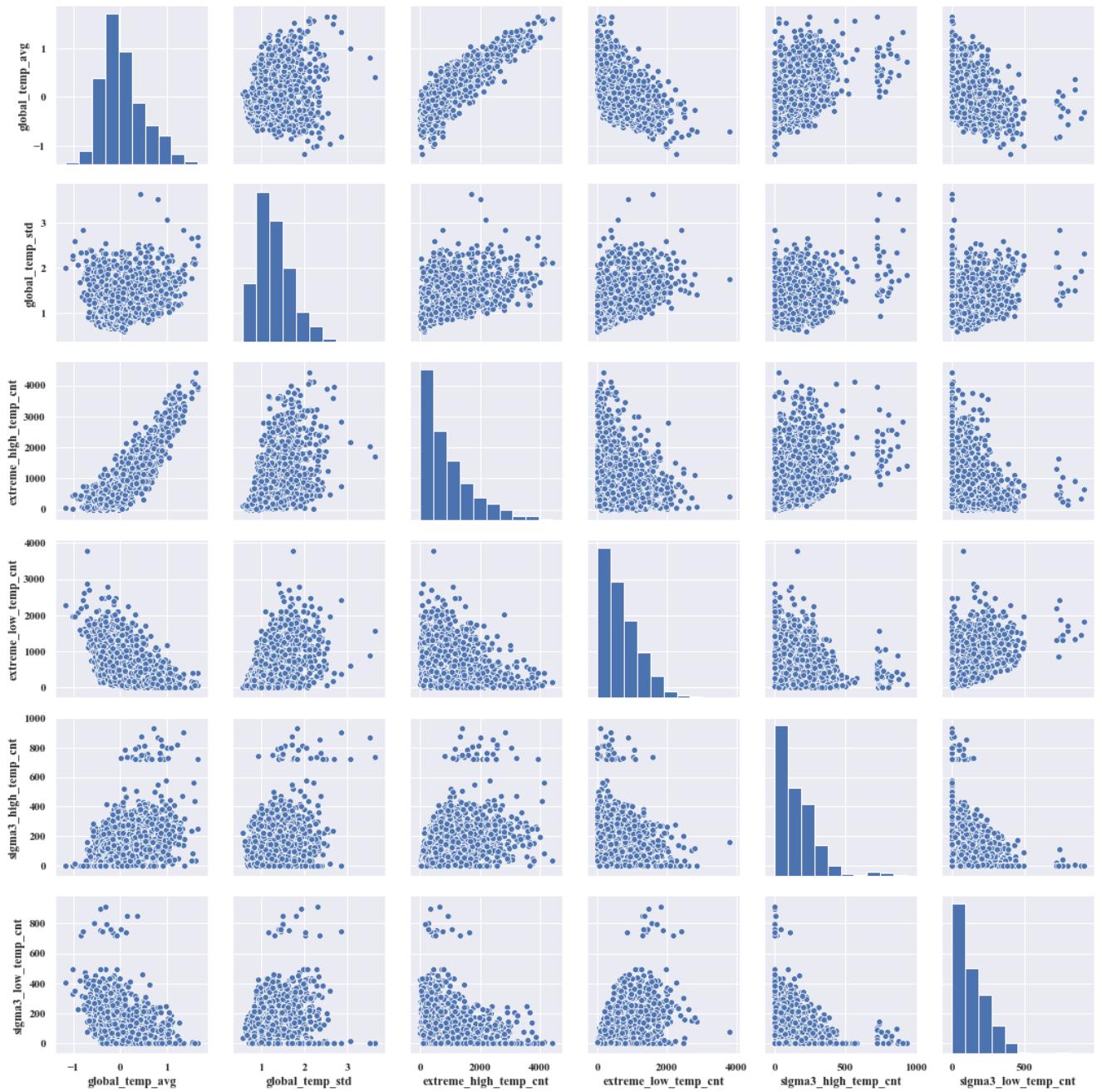


图 5-5 极端天气与全球气候间的关联性

### 5.3 模型结果解释及问题三的解决

上一节中探讨了全球极端天气的变化趋势的三种模式：A 模式、B 模式和 C 模式。模型的计算结果显示，全球极端天气的发生（变化）趋势接近与 C 模式，即逐渐向更热的方向移动，极热天气频次（概率）增加，极寒天气频次（概率）有所降低。

极寒天气频次的降低并不意味着不会发生极寒天气，也不会意味着极寒天气发生时的程度降低，而仅仅是其频次或概率上的降低。

气候模型的分析表明，即使在 21 世纪全球变暖的情况下，极端寒冷事件仍可能在各地持续存在。尽管就全球范围内极端寒冷事件数量而言，模型与预计的极端寒冷事件数量

相一致，但在区域范围内，模型间的变异性和模型性能的差异往往占主导地位。模型的应用结果表明，尽管全球总体呈变暖趋势，但即使到本世纪末，区域极端寒冷事件的发生也不会受到影响。

前节中的模型检验了以下假设：在全球变暖的情况下，极端寒冷程度将变得不那么严重。区域温度升高和热浪恶化。可以通过将模型后预报与过去的观测结果进行比较，并通过未来的多模型协议来对寒冷极端预测的可靠性进行评估。

全球变暖与局地极寒的关系：

- (1) 全局与局部、长期与短期的关系并不影响全球变暖的趋势。
- (2) 正是全球变暖，才会导致了极寒天气。

全球变暖指的是常年的平均气温变暖，在冬季出现一段时间的极寒天气，并不会改变平均气温升高。

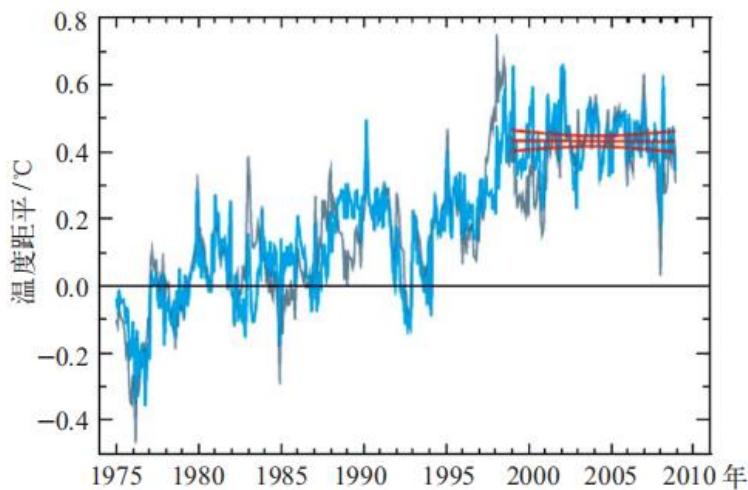


图 5-6 1975-2008 年全球平均温度距平

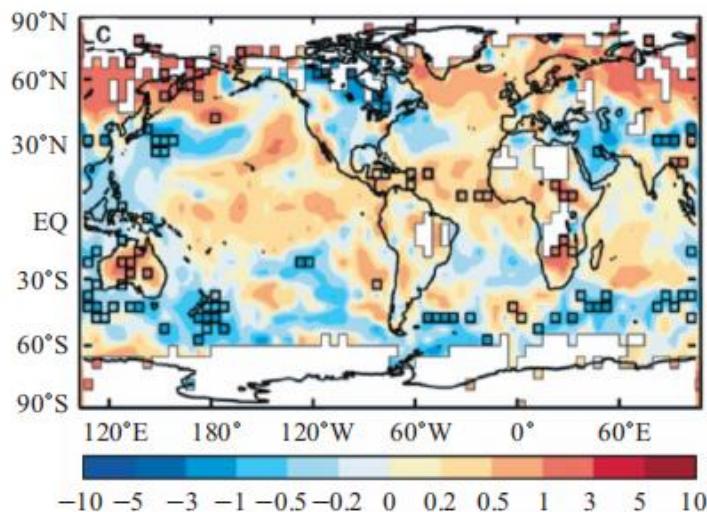


图 5-7 1999-2008 年全球温度增量

## 六. 对极端天气与气候变化的进一步探讨

### 6.1 问题四的分析

目前国内外学术界的主流观点是认同地球正在经历全球变暖的，但如同上文所述，局部地区极寒气象时有发生。2009年跨入2010年的冬季，美国大西洋沿岸中部各州迎来了极寒天气，降雪量异常之大，美国部分地区例如杜勒斯机场、华盛顿特区等迎来了有史以来最大的降雪量。2018年1月，美国和加拿大东岸的多个城市发生极寒气象，气温降到历来最低点，暴风雪导致交通大瘫痪。冬季严寒气象鼓励了否认全球变暖的人和那些反对限制温室气体排放的人，媒体各式各样的报道也使普通群众陷入了对于全球变暖是否真实存在的迷茫。

如上一章所述，产生分歧的原因是考虑问题时时间尺度上的不同。不同于天气的冷热带给人的当下的直观感受，气候的变化需要大时间跨度的数据去体现。人们感受到了极寒天气带来的冷感，但这并不能代表气候上的降温。非专业人士需要一个新的概念替代“全球变暖”，以解决感知上的冲突，直观认识到天气与气候的不同。同时还需要区分全球变暖与气候变化[12]，全球变暖只单有全球平均温度这一指标，由温度的变化可以看到全球变暖现象的发生、暂停或中断、减缓。而气候变化受到多种因素的影响，包括温室气体、太阳辐射、海洋与大气的相互作用等，当然也包括温度这一因素。公众若是把这两个概念混为一谈，就很容易得出全球变暖停滞等结论。这也要求替代“全球变暖”的新概念能反映气候变化的趋势和复杂性。

为了解决这些问题，本章将在上一章对于极端天气的量化数据和关联性模型的基础上，尝试进一步探讨局地极端天气的形成原因与影响因素，将极端天气与全球局部地区的异常气候现象如厄尔尼诺现象、拉尼娜现象、太平洋涛动等相关量，尝试量化它们之间的关联程度，从而进一步揭示局地极寒等极端天气的形成机理。在模型结果的基础上，尝试提出新的概念以代替“全球变暖”，以说明气候变化的趋势与复杂性。

### 6.2 局地天气极寒的概念及原因分析

极端气候事件通常被定义为与概率分布的尾区相关联的气候事件，通常通过给定阈值和超出该阈值的次数进行测量。特别是极寒天气，可以通过各种气候指数来测量，包括阈值超标、霜冻日（ $T_{min} < 0^{\circ}\text{C}$  的天数）、结冰日（ $T_{max} < 0^{\circ}\text{C}$  的天数）和寒冷持续时间指数。根据IPCC（2011年）的估计，极端天气带来的损失难以估计。自1980年以来，每年的损失范围从几十亿美元到超过2,000亿美元。全球与天气有关的灾害损失，难以估价和货币化。

不同地区因其地理环境不同，发生极寒天气现象的原因也各不相同。针对目前北半球出现的极寒天气，中国科学院海洋研究所副研究员陈永利说，国际上对于极寒天气这一现象的产生原因有不同的观点，但主流观点是认为全球变暖现象造成了极寒天气。中国气象局经过研究表示，北极涛动造成了此次北半球的极寒天气[13]。美国哥伦比亚大学的学者对2009年10月冬季在美国中部和欧洲西北部出现异常大降雪这一现象进行分析，得出负的NAO值和厄尔尼诺现象是造成当时冬季天气极寒降暴雪的原因[14]。

## 6.3 局地极端天气的多因子关联性模型

### 6.3.1 各因子对应的数据集处理

美国宇航局戈达德空间研究所（GISS）提供全球表面温度数据集 GISTEMP，该数据集提供了从 1880 年 1 月至 2018 年 3 月全球范围内空间分辨率  $2 \times 2$  度（经纬度）的海洋与陆地的表面温度。它是用于监视全球和区域温度变化和趋势的主要数据集之一。与 MLOST 或 HadCRUT4 相比，GISTEMP 的极地覆盖范围更大，一部分的原因是由于其包含了南极的“READER”站，另一部分原因则是由于采用了插值方法。该数据集的优势是数据范围几乎覆盖全球，尤其是在 1200 km 平滑版本中[15]。

对于全球极端天气的统计将从全球范围内进行，统计每年每个经纬度网格内低于或高于一定基准区间阈值的天气频次，在全球范围内求和即可得到该年度全球极端天气的年度总频次。

全球气候异常现象如美国西南部季风、大西洋累计气旋强度、东太平洋累计气旋强度、太平洋代际振荡、厄尔尼诺现象等的数据集来源于 ESRL 机构。

表 6-1 各因子相应的变量及数据统计性描述

变量	含义	平均值	标准偏差	个案数
Extreme_L_T_2C	全球范围内极端低温（ $2^{\circ}\text{C}$ 基准）的年总频次	657.6058	577.57564	860
Extreme_L_T_3s	全球范围内极端低温（ $3\sigma$ 基准）的年总频次	125.5674	139.11712	860
Global_T_Ano	全球平均温度距平	-.1906	1.41282	860
Monsoon_SW_US	美国西南部季风	1.1368	.81664	756
Monsoon_Mid_Ind	印度中部季风降水	806.0337	1091.81801	624
ACE_Atlantic	大西洋累计气旋强度	101.5072	57.52391	828
ACE_EP	东太平洋累计气旋强度	129.8478	61.14835	552
O_AAO_20CR	南极涛动	-.29341	.985641	780
O_AO_20CR	北极涛动	-.08643	1.008601	780
O_NAO	北大西洋涛动	-.0718	1.09674	836
O_EP	东太平洋涛动	-.0073	1.07369	767
DO_PDO	太平洋代际振荡	-.0899	1.11786	849
O_SOI	南方涛动指数	.2021	1.55272	824
Hurricane_Total	每月飓风总数	1.9955	4.76574	828
Hurricane_MDR_SST	飓风主要区域内的海洋表面温度	.1276	.34385	748
S_I_extent_NH	北半球的海冰范围	11.4589	3.21864	488
S_I_extent_SH	南半球的海冰范围	11.5805	5.58018	488
Nino_1_2_SST	厄尔尼诺1和2	23.0322	2.24986	836
Nino_3_4_SST	厄尔尼诺3和4	26.9031	.98623	836
Nino_3_SST	厄尔尼诺3	25.7780	1.28668	836
Nino4_SST	厄尔尼诺4	28.3832	.71682	836

### 6.3.2 各因子相关性关系

表 6-2 极端天气及各影响因素间的相关性系数

相关性系数	Extreme_L_T_2C	Extreme_L_T_3s	Glob al_T_An	Mon s_S_W_	Mon s_M_id_I	ACE_Atlantic	ACE_EP	O_AAO_20CR	O_AO_20C_R	O_NAO	O_EP	DO_PDO	O_SOI	Hurri_Total	Hurri_DR_SST	S_I_exten_NH	S_I_exten_SS	Nino_1_2_SS	Nino_3_4_SS	Nino_3_SST	Nino_4_SST
Extreme_L_T_2C	1	.397**	-.315**	-.114**	-.27	-0.0	.084	.135	-.10	-.11	-0.0	-0.0	0.00	-.11	-0.0	.194**	-.098*	.119**	-0.0	0.037	-.106**
Extreme_L_T_3s	.397**	1	0.017	0.047	0.063	-.085	.088	0.057	-.007	-.007	0.004	-0.044	-0.054	-.005	-.104	-0.061	.125**	-.114**	0.007	-0.06	0.020
Global_T-Ano	-.315**	0.017	1	.292**	.752**	0.019	0	0.026	.105**	.092**	0.050	.136**	-.0043	.298**	0.019	-.646**	.649**	-.398**	.279**	0.036	.356**
Mons_SW-US	-.114**	0.047	.292**	1	.503**	-0.047	.145**	-.004	0.057	.0907	.138**	.110**	-.128**	.149**	-0.027	-.310**	.243**	-.215**	0.035	-.121**	.144**
Mons_Mid_Ind	-.271**	0.063	.752**	.503**	1	0.002	0.016	-0.027	0.062	.0882	.123**	0.024	0.057	.312**	-0.032	-.672**	.611**	-.507**	-0.053	-.279**	.092*
ACE_Atlantic	-0.065*	-.089	0.0147	-0.002	0.002	1	-.376**	.100**	-.071*	-.080*	-0.067	-.159**	.156**	.232**	.358**	-0.054	0.009	-.109	-.196**	-.156**	-.159**
ACE_EP	.084*	.088*	0	.145**	0.016	-.376**	1	-.0062	.163**	.173**	.111*	.271**	-.295**	-.090*	-.271**	0.023	-0.007	.176**	.268**	.214**	.260**
O_AAO_20CR	.135**	0.057	0.026	-0.04	-0.027	.100**	-0.062	1	0.021	0.013	-0.055	0.023	.127**	0.021	0.036	0	0.002	0.028	-.101**	-0.026	-.078*
O_AO_20C_R	-.101**	-0.069	.105**	0.057	0.062	-.071*	.163**	0.021	1	.634**	-.135**	-.154**	.098**	0.045	-.222**	0.008	0.012	-.081*	-0.045	-.052	-.029
O_NAO	-.119**	-0.036	.092**	.090*	.088*	-.080*	.173**	0.013	.634**	1	0.038	0.047	-0.067	.071*	-.227**	0.046	-.002	-.082	0.022	-.002	.068*

相关性系数	Extreme_L_T_-2C	Extreme_L_An_-T_3s	Glob al_T_o	Mon s_S_W_-US	Mon s_M_id_I_nd	ACE_Atlantic	ACE_EP	O_AAO_20CR	O_AO_20CR	O_NAO	O_EP	DO_PDO	O_SOI	Hurri_Total	Hurri_MDR_SST	S_I_exten_t_NH	S_I_exten_t_SS	Nino_1_2_SS	Nino_3_4_SS	Nino_3_SST	Nino_4_SST
O_EP	-0.041	0.004	0.05	.138**	.123**	-0.067	.111*	-0.055	-.135**	0.038	1	.251**	-.171**	0.028	0.013	0.075	-0.041	-0.025	.154**	0.047	.218**
DO_PDO	-0.025	-0.044	.136**	.110**	0.024	-.159**	.271**	0.023	-.154**	0.047	.251**	1	-.387**	-.089*	.174**	.1667**	-.137**	.272**	.474**	.411**	.466**
O_SOI	0.004	-0.0543	-0.043	-.128**	0.057	.156**	-.295**	.127**	.098**	-0.060	-.171**	-.387**	1	0.042	-.109**	-0.014	-0.012	-.190**	-.619**	-.420**	-.640**
Hurri_Total	-.114**	-0.014	.298**	.149**	.312**	.232**	-.090*	0.021	0.045	.071*	0.028	-.089*	0.042	1	.092*	-.594**	.487**	-.455**	-.191**	-.357**	-.0017
Hurri_MD_R_SST	-0.015	-.104**	0.019	-0.027	-0.032	.358**	-.271**	0.036	-.222	-.227	0.013	.174**	-.109**	.0921	-0.083	0.009	.078*	.217**	.146**	.315**	
S_I_extent_NH	.194**	-0.061	-.646**	-.310**	-.672**	-0.054	0.023	00	0.008	0.046	0.075	.166**	-0.014	-.594**	-0.083	1	-.926**	.813**	.131**	.515**	-.239**
S_I_extent_SH	-.098*	.125**	.649**	.243**	.611**	0.009	-0.007	0.002	0.012	-0.032	-0.041	-.137**	-0.012	.487**	0.009	-.926**	1	-.844**	-0.077	-.484**	.262**
Nino_1_2_SST	.119**	-.114**	-.398**	-.215**	-.507**	-.109**	.176**	0.028	-.081*	-.089**	-0.025	.272**	-.190**	-.455**	.078*	.813**	-.844**	1	.455**	.811**	.076*
Nino_3_4_SST	-0.023	0.007	.279**	0.035	-0.053	-.196**	.268**	-.101**	-.0045	0.022	.154**	.474**	-.619**	-.191**	.217**	.131**	-0.077	.455**	1	.859**	.819**
Nino_3_SS_T	0.037	-0.06	0.036	-.121**	-.279**	-.156**	.214**	-.0026	-.0052	-.0029	0.047	.411**	-.420**	-.357**	.146**	.515**	-.484**	.811**	.859**	1	.489**
Nino4_SST	-.106**	0.02	.356**	.144**	.092*	-.159**	.260**	-.078*	-.0029	.068*	.218**	.466**	-.640**	-0.017	.315**	-.239**	.262**	.076*	.819**	.489**	1

\*\*. 在 0.01 级别（双尾），相关性显著。\*. 在 0.05 级别（双尾），相关性显著。

### **6.3.3 关联性模型结果的解释及应用价值**

从关联性模型运行结果得到的相关性系数可以分析出各因子之间的关联性及其与局地极端天气的关联性，于是可以得出影响全球极端天气的因素及其影响程度。

## **6.4 全球气候变暖的可替代性新概念**

全球变暖虽然是全球气候的变化趋势，但其难以反映在变暖环境下的极端天气发生的情况，这容易导致非专业人士可能对此存在的误解。为了提高概念的普及性与准确性，本文尝试在前述研究的基础上提出两个新的概念。

### **6.4.1 “全球气候异常”**

目前，很多研究采用温度距平来刻画全球气候，温度距平的优势在于可以了解全球范围内的温度距离全球平均温度的差值，从而可以刻画全球温度的冷热情况。但是该指标也存在一定的不足，由于是对全球所有区域的平均，极热和极冷有可能存在部分抵消的情况，从而一定程度上削弱了局地气象在全球温度指标上的体现。

为此，本文尝试提出“全球气候异常”这一概念，该概念旨在表示全球的气候不仅仅是变暖，在变暖的同时，由于有多种影响气候的子系统与因素，从而导致在局部区域出现极端天气现象。气候异常概念较“全球变暖”更加宽泛，便于非专业人士能够迅速理解全球气候的趋势与复杂性。

### **6.4.2 “全球气候振荡热移”**

前述的“气候异常”这一概念有所外延，有一定优势，但宽泛后的表述也容易丢失对全球气候的主要特征——变暖的描述。基于此，本文尝试提出第二种概念——“全球气候振荡热移”，试图替代“全球变暖”表述。

振荡是全球气候变化的特征之一。由于全球气候受到温室气体、太阳辐射、海洋作用等多因素的影响，而有些因素如太阳辐射是具有周期性，同时地气自身的调节作用，也会导致全球气候具有一定的振荡性。

热移是全球气候变化的特征之二。其表述了“变暖”这一趋势，如同本文第五章讨论的三种模式，温度的频次或概率分布通常呈正态分布，均值和方差的改变造成温度分布的移动。而模型的结果则显示，全球温度正向着变热的方向移动。

综合以上两点分析，振荡热移可以同时表达全球气候变化的趋势和复杂性，较为准确地反映了全球气候变化的特征。

## 八. 参考文献

- [1] 文继英, 李青霞, 全球变暖问题浅探, 科协论坛, 4: 355-356, 2007。
- [2] 王景琳, 申冰冰, 刘新斌, 全球气候变暖的成因和后果, 长治学院学报, 3: 57-61, 1999。
- [3] 林霄沛, 许丽晓, 李建平, 等, 全球变暖“停滞”现象辨识与机理研究, 地球科学进展, 31(10): 995-1000, 2016。
- [4] 傅桦, 全球气候变暖的成因与影响, 首都师范大学学报(自然科学版), 28(6): 11-15, 21, 2007。
- [5] 王绍武, 罗勇, 唐国利, 等, 近 10 年全球变暖停滞了吗?, 气候变化研究进展, 6(2): 95-99, 2010。
- [6] 宋斌, 智协飞, 胡耀兴, 全球变暖停滞的形成机制研究进展, 大气科学学报, 38(2): 145-154, 2015。
- [7] 百度百科, 时间序列,  
<https://baike.baidu.com/item/%E6%97%B6%E9%97%B4%E5%BA%8F%E5%88%97/1389644?fr=aladdin>, 2019 年 9 月 23 日。
- [8] 于宁莉, 易东云, 涂先勤. 时间序列中自相关与偏相关函数分析[J]. 数学理论与应用, 2007(01):54-57。
- [9] 维基百科, 太阳辐射,[http://wikipedia.moesalih.com/Solar\\_radiation.2019](http://wikipedia.moesalih.com/Solar_radiation.2019)。
- [10] 马浩, 忠辉. 海洋在气候系统中的作用. 中国科技纵横, 267-268, 2010。
- [11] Xing, Y.F., Xu, Y.H., Shi, M.H., Lian, Y.X, The impact of PM2.5 on the human respiratory system, Journal of Thoracic Disease, 8: E69–E74, 2016.
- [12] 王绍武, 罗勇, 赵宗慈, 对变暖停滞的思考, 气候变化研究进展, 10(4):303-306, 2014。
- [13] 百度百科, 极寒天气,  
<https://baike.baidu.com/item/%E6%9E%81%E5%AF%92%E5%A4%A9%E6%B0%94/8225052?fr=aladdin>, 2019 年 9 月 22 日。
- [14] Seager R , Kushnir Y , Nakamura J , et al, Northern Hemisphere winter snow anomalies: ENSO, NAO and the winter of 2009/10, Geophysical Research Letters, 37(14):14703, 2010.
- [15] National Center for Atmospheric Research Staff (Eds). Last modified 09 Jun 2018. "The Climate Data Guide: Global surface temperature data: GISTEMP: NASA Goddard Institute for Space Studies (GISS) Surface Temperature Analysis." Retrieved from <https://climatedataguide.ucar.edu/climate-data/global-surface-temperature-data-gistemp-nasa-goddard-institute-space-studies-giss>, 2019 年 9 月 22 日。
- [16] 李华斌. 全球变暖定量分析及未来全球平均气温变化预测[D]. 浙江大学, 2007.

## 九. 附录

### 附录 1 netCDF 的 python 提取与处理代码

```
# 导入相关的工具包

import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings("ignore")
pd.options.display.max_rows = 100
pd.options.display.max_columns = 20

import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
sns.set()
plt.rcParams["font.family"] = "Times New Roman"

from dateutil.relativedelta import relativedelta # working with dates with style
from scipy.optimize import minimize
import scipy.stats as scs
import statsmodels.api as sm
import statsmodels.formula.api as smf
import statsmodels.tsa.api as smt
from tqdm import tqdm_notebook
from itertools import product

%load_ext autoreload
%autoreload 2

# 读入数据，对数据进行初步整理
annual = pd.read_excel('NASA-Land-Ocean Temperature Index (C).xlsx'
                      , sheet_name = 'NASA-Land-Ocean Temperature Ind'
                      , index_col = 'Year'
                      , usecols = 1
                      )

annual.set_index(pd.date_range('1880', '2019', freq='Y'), inplace=True)
annual.columns = ['year', 'anomaly']
```

```

# 定义方便的绘图函数，用于绘制时序图、直方图、acf图和pacf图

def tsplot(y, lags=None, figsize=(12, 7), style='bmh'):
    """
    Plot time series, its ACF and PACF, calculate Dickey - Fuller test

    y - timeseries
    lags - how many lags to include in ACF, PACF calculation
    """
    if not isinstance(y, pd.Series):
        y = pd.Series(y)

    # with plt.style.context(style):
    fig = plt.figure(figsize=figsize)
    layout = (2, 2)
    ts_ax = plt.subplot2grid(layout, (0, 0), colspan=2)
    acf_ax = plt.subplot2grid(layout, (1, 0))
    pacf_ax = plt.subplot2grid(layout, (1, 1))

    y.plot(ax=ts_ax)
    p_value = sm.tsa.stattools.adfuller(y)[1]
    ts_ax.set_title('Dickey-Fuller: p={0:.5f}'.format(p_value), fontsize = 14)
    ts_ax.set_xlabel('Year', fontsize=14)
    ts_ax.set_ylabel('annual_anomaly', fontsize=14)

    smt.graphics.plot_acf(y, lags=lags, ax=acf_ax)
    acf_ax.set_title('Autocorrelation', fontsize=14)
    acf_ax.set_xlabel('Lag', fontsize=14)
    acf_ax.tick_params(labelsize=14)

    smt.graphics.plot_pacf(y, lags=lags, ax=pacf_ax)
    pacf_ax.set_title('Partial Autocorrelation', fontsize=14)
    pacf_ax.set_xlabel('Lag', fontsize=14)
    pacf_ax.tick_params(labelsize=14)

    plt.tight_layout()

```

## 附录 2 对全球海洋表面的数据集的处理代码

```
from netCDF4 import Dataset

dataset = Dataset('sst.mnmean.nc')

dataset.dimensions.keys()

odict_keys(['lat', 'lon', 'time', 'nbnds'])

pd.date_range('1854', periods=1985, freq = 'M')

DatetimeIndex(['1854-01-31', '1854-02-28', '1854-03-31', '1854-04-30',
               '1854-05-31', '1854-06-30', '1854-07-31', '1854-08-31',
               '1854-09-30', '1854-10-31',
               ...
               '2018-08-31', '2018-09-30', '2018-10-31', '2018-11-30',
               '2018-12-31', '2019-01-31', '2019-02-28', '2019-03-31',
               '2019-04-30', '2019-05-31'],
              dtype='datetime64[ns]', length=1985, freq='M')

dataset.variables
```

```

for e in dataset.variables:
    print(e)
    print(dataset[e].shape)

lat
(89, )
lon
(180, )
time_bnds
(1985, 2)
time
(1985, )
sst
(1985, 89, 180)

for index, e in enumerate(np.array(dataset['lat'])):
    print(index, e)
    ...
pd.DataFrame(np.array(dataset['time'])).to_csv('date_range.csv')

sst = np.array(dataset['sst'])

sst.shape
    ...
np.where(sst>-1e+36, True, False)

fig = plt.figure(figsize=(15, 7))
layout = (3, 1)
ax1 = plt.subplot2grid(layout, (0, 0))
ax2 = plt.subplot2grid(layout, (1, 0))
ax3 = plt.subplot2grid(layout, (2, 0))
sst_df.plot(subplots=True, figsize = (16,12))
# plt.xlabel('Year', fontsize=14)
# plt.xticks(fontsize=12)
# plt.ylabel('Temp', fontsize=14)
# plt.legend(fontsize=12)
# plt.yticks(fontsize=12)
# plt.grid(True);

array([<matplotlib.axes._subplots.AxesSubplot object at 0x000001E308CF5B70>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000001E3086C9CC0>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000001E3086D2518>], dtype=object)

```

The figure displays a single line plot titled "high\_lat\_sst". The x-axis represents time in years, and the y-axis represents temperature. The plot shows a highly oscillatory signal with a gradual upward trend, indicating warming over time.

```
# sst_df.to_csv('sst.csv')

pd.date_range('1854', periods=1985, freq = 'M')

DatetimeIndex(['1854-01-31', '1854-02-28', '1854-03-31', '1854-04-30',
               '1854-05-31', '1854-06-30', '1854-07-31', '1854-08-31',
               '1854-09-30', '1854-10-31',
               ...,
               '2018-08-31', '2018-09-30', '2018-10-31', '2018-11-30',
               '2018-12-31', '2019-01-31', '2019-02-28', '2019-03-31',
               '2019-04-30', '2019-05-31'],
              dtype='datetime64[ns]', length=1985, freq='M')
```

```
avg_sst
```

```
array([ 13.41712284, 13.52676105, 13.60245705, ..., 14.11854935,
       14.0453167 , 13.96929932])
```

### 附录3 运用Python对时间序列的数据处理与可视化分析代码

```
def tsplot2(y, title = 'timeseries', lags=None, figsize=(12, 8), style='bmh'):
    """
        The 4 plots that you always want to make when looking at a particular series
    """
    if not isinstance(y, pd.Series):
        y = pd.Series(y)

    # with plt.style.context(style):
    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, fontsize=12)
    p_value = sm.tsa.stattools.adfuller(y)[1]
    ts_ax.set_title('Dickey-Fuller: p = {:.5f}'.format(p_value),
                    fontsize=14, fontweight='bold')
    ts_ax.set_xlabel('Year', fontsize=14)
    ts_ax.set_ylabel('annual anomaly', fontsize=14)

    y.plot(ax=hist_ax, kind='hist', bins=30, fontsize=12)
    hist_ax.set_title('Histogram', fontsize=14)
    hist_ax.set_ylabel('Frequency', fontsize=14)
```

```

smt.graphics.plot_acf(y, lags = lags, ax = acf_ax)
acf_ax.set_title('Autocorrelation', fontsize=14)
acf_ax.set_xlabel('Lag', fontsize=14)
acf_ax.tick_params(labelsize=14)

smt.graphics.plot_pacf(y, lags = lags, ax = pacf_ax)
pacf_ax.set_title('Partial Autocorrelation', fontsize=14)
pacf_ax.set_xlabel('Lag', fontsize=14)
pacf_ax.tick_params(labelsize=14)
[ax.set_xlim(0) for ax in [acf_ax, pacf_ax]]

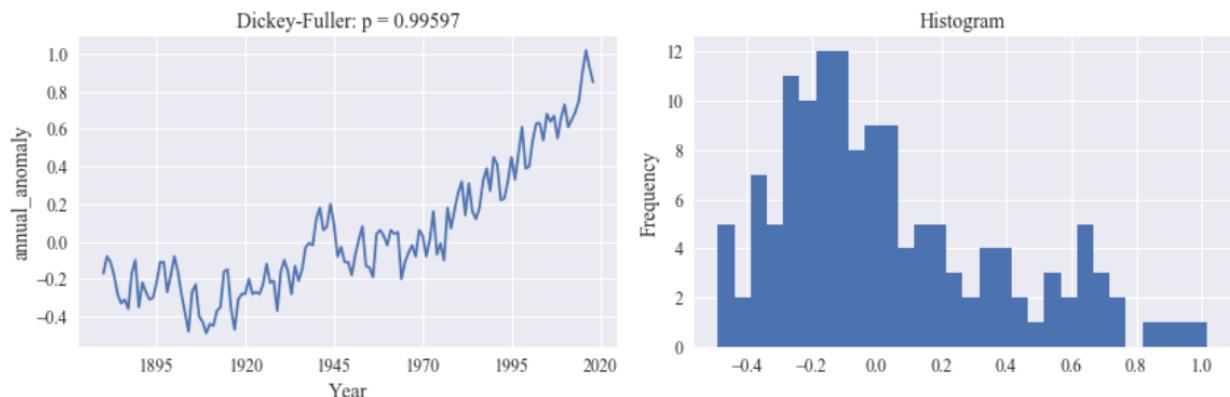
sns.despine()
plt.tight_layout()

return ts_ax, acf_ax, pacf_ax

```

```
tsplot2(annual.anomaly, annual.anomaly.name, lags=30)
```

```
(<matplotlib.axes._subplots.AxesSubplot at 0x1e352345128>,
<matplotlib.axes._subplots.AxesSubplot at 0x1e35239cd68>,
<matplotlib.axes._subplots.AxesSubplot at 0x1e3523c9400>)
```



## 附录 4 运用 Python 构建及应用 ARIMA 模型的代码

```

# 根据acf和pacf的结果，设定SARIMA模型参数的调节范围，进行模型寻优
ps = range(1, 4)
d = 1
qs = range(4, 7)

Ps = range(0, 1)
D = 0
Qs = range(0, 1)

s = 0 # season length

# creating list with all the possible combination of parameters
parameters = product(ps, qs, Ps, Qs)
parameters_list = list(parameters)
print(parameters_list.__len__())

```

```

# 编写模型寻优函数

def optimizeSARIMA(annual.anomaly, parameters_list, d, D, s):
    """
    Return dataframe with parameters and corresponding AIC

    parameters_list - list with (p, q, P, Q) tuples
    d - integration order in ARIMA model
    D - seasonal integration order
    s = length of season
    """

    results = []
    best_aic = float('inf')

    for param in tqdm_notebook(parameters_list):
        print(f'Parameter: {param} .....')
        # we need try_except because on some combinations model fails to converge
        try:
            model = sm.tsa.statespace.SARIMAX(annual.anomaly, order=(param[0], d, param[1]),
                                                seasonal_order=(param[2], D, param[3], s)).fit(disp=-1)
        except:
            continue

        aic = model.aic
        # save the best model, AIC and parameters
        if aic < best_aic:
            best_model = model
            best_aic = aic
            best_param = param

        results.append([param, model.aic])

    result_table = pd.DataFrame(results)
    result_table.columns = ['parameters', 'aic']
    # sorting in ascending order, the lower AIC is the better
    result_table = result_table.sort_values(by=['aic'], ascending=True).reset_index(drop=True)

    return result_table, best_aic, best_model

# 模型寻优
result_table, _, _ = optimizeSARIMA(annual.anomaly, parameters_list, d, D, s)

# 得到最优模型的参数并拟合
p, q, P, Q = result_table.parameters[0]

best_model = sm.tsa.statespace.SARIMAX(annual.anomaly, order=(p, d, q),
                                         seasonal_order=(P, D, Q, s)).fit(disp=-1)
print(best_model.summary())

```

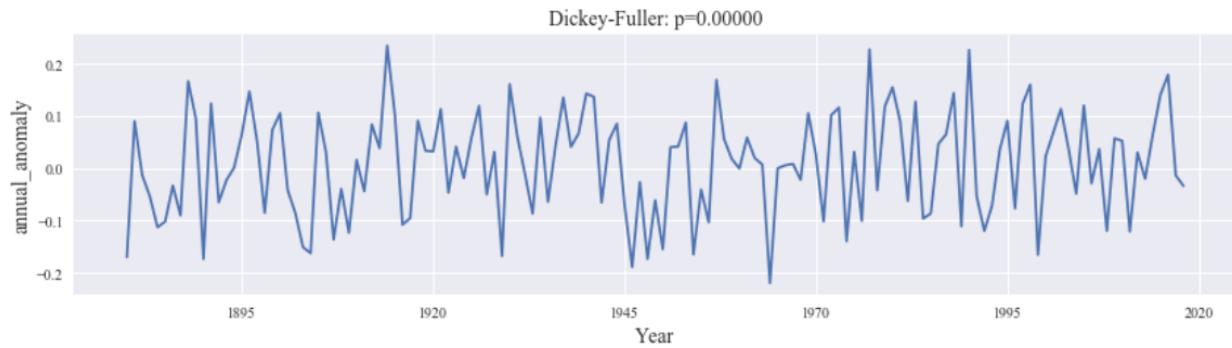
```
# 寻优结果  
result_table
```

	parameters	aic
0	(2, 4, 0, 0)	-230.960061
1	(1, 4, 0, 0)	-230.170935
2	(3, 5, 0, 0)	-229.597340
3	(1, 5, 0, 0)	-228.736433
4	(3, 4, 0, 0)	-227.146639
5	(1, 6, 0, 0)	-226.777519
6	(2, 5, 0, 0)	-225.128195
7	(2, 6, 0, 0)	-224.910065
8	(3, 6, 0, 0)	-224.793632

```
best_model=sm.tsa.statespace.SARIMAX(annual_anomaly, order=(2, 1, 4),  
                                         seasonal_order=(0, 0, 0, 0)).fit(disp=1)  
print(best_model.summary())
```

```
# 使用SARIMA模型(实际为ARIMA模型)做预测  
sarima_predict = best_model.predict(start='1910-12-31', end='2044-12-31')
```

```
tsplot(best_model.resid, lags=30)
```



```
# 引入误差函数
```

```
from sklearn.metrics import r2_score, median_absolute_error, mean_absolute_error  
from sklearn.metrics import median_absolute_error, mean_squared_error, mean_squared_log_error  
  
def mean_absolute_percentage_error(y_true, y_pred):  
    return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
```

```

# 编写预测图像的绘制函数

def plotSARIMA(series, model, n_steps):
    """
        Plots model vs predicted values

        series - dataset with timeseries
        model - fitted SARIMA model
        n_steps - number of steps to predict in the future

    """

    # adding model values
    data = series.copy()
    data.columns = ['actual']
    data['arima_model'] = model.fittedvalues
    # making a shift on s+d steps, because these values were unobserved by the model
    # due to the differentiating
    data['arima_model'][:s+d] = np.NaN

    # forecasting on n_steps forward
    forecast = model.predict(start = data.shape[0], end = data.shape[0]+n_steps)
    forecast = data.arima_model.append(forecast)
    # calculate error, again having shifted on s+d steps from the beginning
    error = mean_squared_error(data['actual'][s+d:], data['arima_model'][s+d:])

    plt.figure(figsize=(15, 7))
    plt.title("mean_squared_error: {:.3f}".format(error), fontsize=14)
    plt.plot(forecast, color='r', label="model",)
    plt.axvspan(data.index[-1], forecast.index[-1], alpha=0.5, color='lightgrey')
    plt.plot(data.actual, label="actual")
    plt.xlabel('Year', fontsize=14)
    plt.xticks(fontsize=12)
    plt.ylabel('Anomaly', fontsize=14)
    plt.legend(fontsize=12)
    plt.yticks(fontsize=12)
    plt.grid(True);


```

---

```
plotSARIMA(annual[['anomaly']], best_model, 25)
```

## 附录 5 基于随机森林的多变量回归模型的代码

```
# 导入相关的工具包

import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings("ignore")
pd.options.display.max_rows = 100
pd.options.display.max_columns = 20

import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
sns.set()
plt.rcParams["font.family"] = "Times New Roman"

from dateutil.relativedelta import relativedelta # working with dates with style
from scipy.optimize import minimize
import scipy.stats as scs
import statsmodels.api as sm
import statsmodels.formula.api as smf
import statsmodels.tsa.api as smt
from tqdm import tqdm_notebook
from itertools import product

%load_ext autoreload
%autoreload 2

# 读入数据, 读取数据的字段
data = pd.read_excel('nasa_lo_avg_temp.xlsx',
                      sheet_name = 'Sheet1')
data.columns
Index(['DATE_Y', 'L_0_Temp_Ano_Nosmooth_Y', 'L_0_Mean_Temp_Y', 'DATE_X',
       'L_0_Temp_Ano_Nosmooth_X', 'L_0_Mean_Temp_X', 'Solar_Flux_mean',
       'Urbanization', 'Emission_Ozone', 'SO2', 'Land_Built_up', 'Land_Graze',
       'Land_Crop', 'annual_CO2', 'cumu_CO2', 'CFCsTG_OBS', 'CH4_OBS',
       'CO2_OBS', 'N2O_OBS', 'Aero_SA', 'Aero_BC', 'Aero_RA', 'Aero_AIE',
       'SST_high_lat', 'SST_mid_lat', 'SST_low_lat'],
      dtype='object')
```

```

# 取若干字段子集，便于根据需要索引相应数据集

subcols = ['L_0_Temp_Ano_Nosmooth_Y', 'L_0_Mean_Temp_Y',
           'Solar_Flux_mean', 'Urbanization', 'Emission_Ozone', 'SO2', 'Land_Built_up', 'Land_Graze',
           'Land_Crop', 'annual_CO2', 'cumu_CO2', 'CFCsTG_OBS', 'CH4_OBS',
           'CO2_OBS', 'N2O_OBS', 'Aero_SA', 'Aero_BC', 'Aero_RA', 'Aero_AIE',
           'SST_high_lat', 'SST_mid_lat', 'SST_low_lat']

subcols_1 = ['L_0_Temp_Ano_Nosmooth_Y',
             'Solar_Flux_mean', 'Urbanization', 'Emission_Ozone', 'SO2', 'Land_Built_up', 'Land_Graze',
             'Land_Crop', 'annual_CO2', 'cumu_CO2', 'CFCsTG_OBS', 'CH4_OBS',
             'CO2_OBS', 'N2O_OBS', 'Aero_SA', 'Aero_BC', 'Aero_RA', 'Aero_AIE',
             'SST_high_lat', 'SST_mid_lat', 'SST_low_lat']

subcols_2 = ['L_0_Mean_Temp_Y',
             'Solar_Flux_mean', 'Urbanization', 'Emission_Ozone', 'SO2', 'Land_Built_up', 'Land_Graze',
             'Land_Crop', 'annual_CO2', 'cumu_CO2', 'CFCsTG_OBS', 'CH4_OBS',
             'CO2_OBS', 'N2O_OBS', 'Aero_SA', 'Aero_BC', 'Aero_RA', 'Aero_AIE',
             'SST_high_lat', 'SST_mid_lat', 'SST_low_lat']

subcols_x = [
    'Solar_Flux_mean', 'Urbanization', 'Emission_Ozone', 'SO2', 'Land_Built_up', 'Land_Graze',
    'Land_Crop', 'annual_CO2', 'cumu_CO2', 'CFCsTG_OBS', 'CH4_OBS',
    'CO2_OBS', 'N2O_OBS', 'Aero_SA', 'Aero_BC', 'Aero_RA', 'Aero_AIE',
    'SST_high_lat', 'SST_mid_lat', 'SST_low_lat']

subcols_3 = ['L_0_Temp_Ano_Nosmooth_Y', 'L_0_Mean_Temp_Y', 'Urbanization', 'Emission_Ozone', 'SO2', 'Land_Built_up', 'Land_Graze',
             'Land_Crop', 'annual_CO2', 'cumu_CO2', 'CFCsTG_OBS', 'CH4_OBS',
             'CO2_OBS', 'N2O_OBS', 'Aero_BC', 'Aero_RA', 'Aero_AIE',
             'SST_high_lat', 'SST_mid_lat', 'SST_low_lat']

```

## # 数据集基本信息

```
data.info()
```

```
...
```

## # 查询数据集的空值

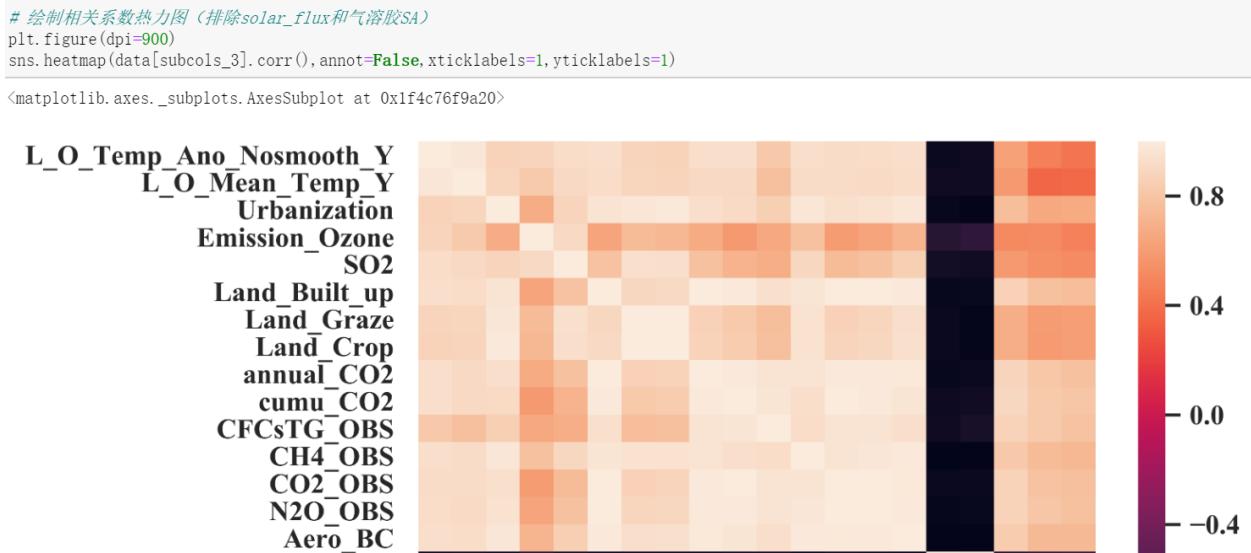
```
data[subcols].isnull().sum()
```

```
...
```

## # 查看所有预测变量（排除solar\_flux和气溶胶SA）的频率分布直方图

```
data[subcols_x].hist(figsize = (20, 20),)
```

```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x000001F4F7482E80>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000001F4F7662CC0>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000001F4E7969C18>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000001F4E7137048>],
      [<matplotlib.axes._subplots.AxesSubplot object at 0x000001F4E75E4470>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000001F48053D0B8>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000001F480560A58>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000001F480599470>],
      [<matplotlib.axes._subplots.AxesSubplot object at 0x000001F4805994A8>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000001F4806027B8>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000001F48063E198>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000001F480669B38>],
```



```
# 绘制相关系数热力图
plt.figure(dpi=900)
sns.heatmap(data[subcols].corr(), annot=False, xticklabels=1, yticklabels=1)
```

## 机器学习预测部分

```
# 利用sklearn机器学习框架构建多元非线性预测模型
# 采用随机森林算法
from sklearn.ensemble import RandomForestRegressor
from sklearn.preprocessing import StandardScaler
```

```
# 取数据集中前140个数据点作为训练数据集
y = data[['L_O_Mean_Temp_Y']].iloc[0:141, :]
X = data[subcols_x].iloc[0:141, :]

# 取数据集中后29个数据点用于预测未来多年的 global temperature
X_test = data[subcols_x].iloc[141:, :]
```

```
# 对输入数据进行标准化
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
X_test_scaled = scaler.transform(X_test)
```

```
# 定义一个随机森林模型并拟合数据
rf = RandomForestRegressor(max_depth = 20, n_estimators=200)
rf.fit(X_scaled, y.values)
```

```
RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=20,
                      max_features='auto', max_leaf_nodes=None,
```

```

# 绘制模型拟合与预测效果图

plt.rcParams['font.family'] = "Times New Roman"
plt.figure(dpi=900)
plt.rcParams['font.size'] = 14
plt.rcParams['axes.labelsize'] = 14
plt.rcParams['axes.labelweight'] = 'bold'
plt.rcParams['axes.titlesize'] = 14
plt.rcParams['xtick.labelszie'] = 14
plt.rcParams['ytick.labelszie'] = 14
plt.rcParams['legend.fontsize'] = 14

plt.figure(figsize=(16,9))
plt.plot(range(X_scaled.shape[0]), y, linestyle='--', marker='*', color='b')
plt.plot(range(X_scaled.shape[0]), rf.predict(X_scaled), linestyle='--', marker='.', color='y')
plt.plot(range(X_scaled.shape[0], X_scaled.shape[0] + X_test_scaled.shape[0]), rf.predict(X_test_scaled),

```

## 附录 6 极端天气的量化数据提取及多因子关联性分析

```

data = Dataset('gistemp1200_GHCNv4_ERSSTv5.nc')

data.variables.keys()
dict_keys(['lat', 'lon', 'time', 'time_bnds', 'tempanomaly'])

for e in data.variables:
    print(e)
    print(data[e].shape)

lat
(90,)
lon
(180,)
time
(1676,)
time_bnds
(1676, 2)
tempanomaly
(1676, 90, 180)

data.variables['tempanomaly']
tempanomaly = np.array(data.variables['tempanomaly']) # _FillValue: 32767
lat = np.array(data.variables['lat'])
lon = np.array(data.variables['lon'])
time = np.array(data.variables['time'])

lat, lon

```

```

tempanomaly.shape
(1676, 90, 180)

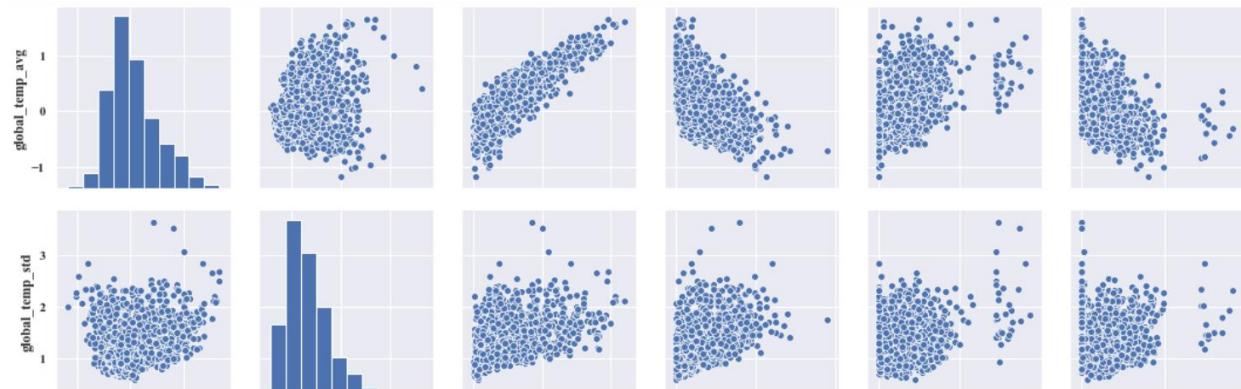
global_temp_avg = np.zeros(tempanomaly.shape[0])
global_temp_std = np.zeros(tempanomaly.shape[0])
extreme_high_temp_cnt = np.zeros(tempanomaly.shape[0])
extreme_low_temp_cnt = np.zeros(tempanomaly.shape[0])
sigma3_high_temp_cnt = np.zeros(tempanomaly.shape[0])
sigma3_low_temp_cnt = np.zeros(tempanomaly.shape[0])
for year, xy_anomaly in enumerate(tempanomaly):
    xy_anomaly_dropna = xy_anomaly[np.where(xy_anomaly == 32767, False, True)]

    year_temp_avg = xy_anomaly_dropna.mean()
    year_temp_std = xy_anomaly_dropna.std()
    global_temp_avg[year] = year_temp_avg
    global_temp_std[year] = year_temp_std
    extreme_high_temp_cnt[year] = np.where(xy_anomaly_dropna > 2, 1, 0).sum() # year_temp_avg + 4 * year_temp_std
    extreme_low_temp_cnt[year] = np.where(xy_anomaly_dropna < -2, 1, 0).sum() # year_temp_avg - 4 * year_temp_std
    sigma3_high_temp_cnt[year] = np.where(xy_anomaly_dropna > year_temp_avg + 3 * year_temp_std, 1, 0).sum() #
    sigma3_low_temp_cnt[year] = np.where(xy_anomaly_dropna < year_temp_avg - 3 * year_temp_std, 1, 0).sum() #

global_temp = pd.DataFrame(data = {'global_temp_avg':global_temp_avg,
                                    'global_temp_std':global_temp_std,
                                    'extreme_high_temp_cnt':extreme_high_temp_cnt,
                                    'extreme_low_temp_cnt':extreme_low_temp_cnt,
                                    'sigma3_high_temp_cnt':sigma3_high_temp_cnt,
                                    'sigma3_low_temp_cnt':sigma3_low_temp_cnt},
                           index = pd.date_range(start = '1880', periods = 1676, freq = 'M'))

sns.pairplot(data = global_temp)

```



```

global_temp.plot(subplots=True, figsize = (20, 12))

array([<matplotlib.axes._subplots.AxesSubplot object at 0x000002AFC0191B00>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000002AFC060B518>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000002AFC063BD30>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000002AFC03D45C0>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000002AFC0403E10>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000002AFC043C6A0>],
      dtype=object)

```